



Universidad de Valladolid



**ESCUELA DE INGENIERÍAS
INDUSTRIALES**

UNIVERSIDAD DE VALLADOLID

ESCUELA DE INGENIERIAS INDUSTRIALES

Grado en Ingeniería en Organización Industrial

Implementación de un sistema de Business Intelligence en una corporación industrial

Autor:

Stoyanov Georgiev, Kaloyan

Tutores:

López Paredes, Adolfo

**Dpto. de Organización de Empresas y
Comercialización e Investigación de
Mercados**

Gordaliza Pastor, Alfonso

**Dpto. de Estadística e Investigación
Operativa**

Valladolid, junio 2019



Universidad de Valladolid



ESCUELA DE INGENIERÍAS
INDUSTRIALES



RESUMEN

En este proyecto se presenta un caso de estudio para ilustrar la utilidad para la toma de decisiones de contar Cuadros de Mando interactivos y dinámicos, que incorporen información estructurada y no estructurada de la empresa. Previamente se ha realizado una revisión del concepto de Business Intelligence. Se analiza el contexto histórico y actual, además de los modelos de madurez más utilizados. También se propone una breve introducción al Big Data y un acercamiento a conceptos como el Data Science y el Business Analytics.

Por último, el caso práctico realizado corresponde al de una empresa industrial. En él se presenta un análisis de los datos operativos de la compañía y un análisis de sentimientos de clientes y empleados para conocer su nivel de satisfacción.



Universidad de Valladolid



ESCUELA DE INGENIERÍAS
INDUSTRIALES



Contenido

ÍNDICE DE ILUSTRACIONES	3
ÍNDICE DE TABLAS	7
INTRODUCCIÓN.....	9
OBJETO	9
OBJETIVOS.....	9
1. CONTEXTO.....	11
1.1 ACTUALIDAD	11
1.2 BASES DE DATOS	13
1.2.1 Definición	13
1.2.2 Componentes.....	13
1.2.3 Modelo entidad-relación	15
1.2.4 Modelo relacional	16
1.2.5 Bases de datos multidimensionales	18
1.3 OLTP Y OLAP	20
1.3.1 OLTP (OnLine Transaction Processing)	20
1.3.2 OLAP (OnLine Analytical Processing).....	21
2. BUSINESS INTELLIGENCE.....	27
2.1 DEFINICIÓN E HISTORIA	27
2.2 ELEMENTOS Y ARQUITECTURA DEL BI	30
2.2.1 Ficheros Excel	31
2.2.2 Ficheros externos	31
2.2.3 Datos operacionales.....	31
2.2.4 ERP, CRM.....	31
2.2.5 Bases de datos.....	31
2.2.6 Data Warehouse.....	31
2.2.8 Data Mart.....	32
2.2.9 Cubos multidimensionales (cubos OLAP).....	33
2.2.10 Reporting.....	34
2.2.11 Proceso ETL (Extraction, Transformation, Load)	34
2.3 CARACTERÍSTICAS DEL BUSINESS INTELLIGENCE	36
2.4 VENTAJAS DEL BUSINESS INTELLIGENCE	37
2.5 ESTRATEGIAS DEL BUSINESS INTELLIGENCE	39
2.5.1 Necesidad de un sistema de Business Intelligence.....	40
2.5.2 Desarrollo de la estrategia	40
2.6 MODELOS DE MADUREZ	43
2.6.1 Business Intelligence Maturity Model (BIMM)	43
2.6.2 Forrester BI Maturity Model.....	44
2.6.3 Business Information Maturity Model.....	44
2.6.4 Gartner's Maturity Model for BI and Performance Management	45
2.6.5 Business Intelligence Maturity Hierarchy	46
2.6.6 Errores comunes al implantar BI.....	47



2.7 BREVE INTRODUCCIÓN AL BIG DATA	47
2.7.1 <i>Diferencias entre Business Intelligence y Big Data</i>	51
2.8 OTROS ENFOQUES	52
2.8.1 <i>Data Science</i>	52
2.8.2 <i>Analytics y Business Analytics</i>	54
3. HERRAMIENTAS DE BI	57
3.1 MICROSOFT POWER BI	66
3.2 TABLEAU	68
3.3 QLIK	70
3.4 THOUGHTSPOT	73
3.5 COMPARATIVA	75
3.6 CONCLUSIONES	79
4. DESCRIPCIÓN DEL CASO DE ESTUDIO.....	81
4.1 PROCESO DETALLADO	82
4.2 ESTADO DE LA COMPAÑÍA	83
4.3 OBJETIVOS Y REQUISITOS DEL SISTEMA DE BUSINESS INTELLIGENCE	87
4.4 MODELO DEL SISTEMA	88
4.4.1 <i>Software y arquitectura</i>	88
4.4.2 <i>Obtención de los datos</i>	89
5. ANÁLISIS DE DATOS ESTRUCTURADOS	91
5.1 DATOS	91
5.2 CONSIDERACIONES GENERALES DEL ANÁLISIS	93
5.3 CUADRO DE MANDO: ESTADO GENERAL.....	98
5.4 CUADRO DE MANDO: INVENTARIOS DE PRODUCTOS Y MP.....	101
5.5 CUADRO DE MANDO: COMPRAS DE MP.....	106
5.6 CUADRO DE MANDO: PRODUCCIÓN	110
5.7 CUADRO DE MANDO: OPERARIOS Y LÍNEAS DE FABRICACIÓN.....	115
6. ANÁLISIS DE DATOS NO ESTRUCTURADOS	121
6.1 INTRODUCCIÓN Y PROPÓSITO DEL ANÁLISIS	121
6.2 PREPARACIÓN DE LOS DATOS	126
6.3 PLANTEAMIENTO Y ANÁLISIS	130
6.4 RESULTADOS DEL ANÁLISIS	141
7. CONCLUSIONES Y EXTENSIONES.....	147
7.1 CONCLUSIONES	147
7.2 EXTENSIONES FUTURAS	148
8. BIBLIOGRAFÍA.....	151
9. PÁGINAS WEB CONSULTADAS.....	153



Índice de ilustraciones

ILUSTRACIÓN 1. CRECIMIENTO DE LOS DATOS. FUENTE: IDC (2018)	12
ILUSTRACIÓN 2. DATOS, INFORMACIÓN Y CONOCIMIENTO	13
ILUSTRACIÓN 3 EJEMPLO DE MODELO ENTIDAD-RELACIÓN (RECUPERADO DE: HTTPS://ES.WIKIPEDIA.ORG/WIKI/Modelo_entidad-relaci%C3%B3n)	15
ILUSTRACIÓN 4. EJEMPLO DE MODELO RELACIONAL (RECUPERADO DE: HTTP://WWW.DATAPRIX.COM/DATA-WAREHOUSING-Y-METODOLOGIA-HEFESTO/-METODOLOGIA-HEFESTO/55-PASOS-Y-APLICACION-METODOLOGICA/553-P)	16
ILUSTRACIÓN 5. ESQUEMA EN COPO DE NIEVE (RECUPERADO DE: HTTPS://ES.WIKIPEDIA.ORG/WIKI/Esquema_en_copo_de_nieve)	18
ILUSTRACIÓN 6. EJEMPLO MODELO MULTIDIMENSIONAL. FUENTE: ELABORACIÓN PROPIA.	19
ILUSTRACIÓN 7. COMPARATIVA DE LOS BI. FUENTE: ELABORACIÓN PROPIA	29
ILUSTRACIÓN 8. ARQUITECTURA COMÚN DE BI. FUENTE: ELABORACIÓN PROPIA	30
ILUSTRACIÓN 9. EJEMPLO DE CUBO OLAP. FUENTE: ELABORACIÓN PROPIA	33
ILUSTRACIÓN 10 CÍRCULO VIRTUOSO DE LA INFORMACIÓN. FUENTE: ELABORACIÓN PROPIA	38
ILUSTRACIÓN 11. HABILIDADES DEL BICC. FUENTE: ELABORACIÓN PROPIA	41
ILUSTRACIÓN 12. GARTNER'S MATURITY MODEL FOR BI AND PERFORMANCE MANAGEMENT (RECUPERADO DE: HTTPS://CO.PINTEREST.COM/PIN/372039619207360824/)	46
ILUSTRACIÓN 13. POPULARIDAD DE BIG DATA Y BUSINESS INTELLIGENCE (RECUPERADO DE: HTTPS://TRENDS.GOOGLE.ES/TRENDS/EXPLORE?DATE=ALL&GEO=ES&Q=BIG%20DATA,BUSINESS%20INTELLIGENCE, VISUALIZADO EL 23/05/19)	48
ILUSTRACIÓN 14 CUADRO RESUMEN DE TIPOS DE DATOS. FUENTE: ELABORACIÓN PROPIA	50
ILUSTRACIÓN 15. EJEMPLO DE CMI. FUENTE: ELABORACIÓN PROPIA	60
ILUSTRACIÓN 16. EJEMPLO DE ALGORITMO C4.5. FUENTE: UIAF (2014)	63
ILUSTRACIÓN 17. EJEMPLO MSV. FUENTE: UIAF (2014)	63
ILUSTRACIÓN 18. EJEMPLO DE DASHBOARD EN POWER BI. FUENTE: CEDIDO POR ENTIDAD BPMSAT S.L	65
ILUSTRACIÓN 19. INTERFAZ POWER BI (RECUPERADO DE: HTTPS://POWERBI.MICROSOFT.COM/ES-ES/BLOG/EXPLORING-YOUR-MICROSOFT-DYNAMICS-NAV-DATA-WITH-POWER-BI/)	66
ILUSTRACIÓN 20. INTERFAZ TABLEAU (1) (RECUPERADO DE: HTTPS://WWW.TABLEAU.COM/ES-ES/LEARN/TUTORIALS/ON-DEMAND/TABLEAU-INTERFACE)	69
ILUSTRACIÓN 21. INTERFAZ TABLEAU (2) (RECUPERADO DE: HTTPS://WWW.TABLEAU.COM/ES-ES/PRODUCTS/SERVER)	70
ILUSTRACIÓN 22. INTERFAZ QLIKSENSE (RECUPERADO DE: HTTPS://WWW.QLIK.COM/ES-ES/PRODUCTS/QLIK-SENSE)	71
ILUSTRACIÓN 23. INTERFAZ QLIK VIEW (RECUPERADO DE: HTTPS://WWW.QLIK.COM/ES-ES/PRODUCTS/QLIKVIEW)	72
ILUSTRACIÓN 24. INTERFAZ THOUGHTSPOT (1) (RECUPERADO DE: HTTPS://WWW.THUGHTSPOT.COM/PRODUCT)	74
ILUSTRACIÓN 25. INTERFAZ THOUGHTSPOT (2) (RECUPERADO DE: HTTPS://WWW.THUGHTSPOT.COM/PRODUCT)	75
ILUSTRACIÓN 26. MAGIC QUADRANT 2018 Y 2019. FUENTE: GARTNER INC. (2018), GARTNER INC. (2019).....	80
ILUSTRACIÓN 27. DATOS DE PRODUCTOS Y MP.	84
ILUSTRACIÓN 28. ESQUEMA SIMPLIFICADO DEL FLUJO DE TRABAJO. ELABORACIÓN PROPIA	85
ILUSTRACIÓN 29. ARQUITECTURA DE LA SOLUCIÓN PARA EL CASO PRÁCTICO. ELABORACIÓN PROPIA	89
ILUSTRACIÓN 30. TABLAS PARA EL ANÁLISIS DE DATOS EN POWER BI	91
ILUSTRACIÓN 31. PMP Y MRP PROPUESTOS	92
ILUSTRACIÓN 32. TABLA COMPRAS	94
ILUSTRACIÓN 33. TABLA PRODUCCIÓN	95



ILUSTRACIÓN 34. TABLA VENTAS	96
ILUSTRACIÓN 35. PÁGINA 1: ESTADO GENERAL	97
ILUSTRACIÓN 36. INGRESOS POR CANAL DE VENTA Y PRODUCTO	99
ILUSTRACIÓN 37. INGRESOS DE A MENSUALES	99
ILUSTRACIÓN 38. INGRESOS POR CANAL DE VENTA Y PRODUCTO EN MAYO.....	99
ILUSTRACIÓN 39. INGRESOS POR CANAL DE VENTA Y PRODUCTO EN ABRIL	100
ILUSTRACIÓN 40. INVENTARIOS A ÚLTIMA FECHA DE MOVIMIENTO	100
ILUSTRACIÓN 41. EBITDA MENSUAL	101
ILUSTRACIÓN 42. PÁGINA 2: INVENTARIOS DE LOS PRODUCTOS A, B, C	102
ILUSTRACIÓN 43. PÁGINA 3: INVENTARIOS DE LAS MP X, Y, Z.....	103
ILUSTRACIÓN 44. INVENTARIO DE A. REPOSICIÓN DE STOCK DE SEGURIDAD	104
ILUSTRACIÓN 45. INVENTARIO DE C. REPOSICIÓN DEL STOCK DE SEGURIDAD	104
ILUSTRACIÓN 46. AUMENTO EN LOS INVENTARIOS DE MP	104
ILUSTRACIÓN 47. PÁGINA 4: COMPRAS DE MP	105
ILUSTRACIÓN 48. CANTIDADES DE MP COMPRADAS POR PROVEEDOR CON % DEL GASTO TOTAL.....	107
ILUSTRACIÓN 49. GRÁFICO DE BURBUJAS DE CANTIDADES ADQUIRIDAS Y GASTO EN CADA PROVEEDOR	107
ILUSTRACIÓN 50. COMPARATIVA DEL LEAD TIME TEÓRICO Y REAL DE PROVEEDORES	108
ILUSTRACIÓN 51. DEFECTOS Y RETRASOS DE PROVEEDORES 1 Y 2	108
ILUSTRACIÓN 52. DEFECTOS Y RETRASOS DE PROVEEDORES 3 Y 4	109
ILUSTRACIÓN 53. COMPARATIVA DE GASTO Y PRESUPUESTO MENSUAL.....	110
ILUSTRACIÓN 54. PÁGINA 5: PRODUCCIÓN	111
ILUSTRACIÓN 55. CANTIDADES FABRICADAS DE PRODUCTO CON % DEL GASTO TOTAL	112
ILUSTRACIÓN 56. DESAJUSTE DE MP POR PRODUCTOS	113
ILUSTRACIÓN 57. COMPARATIVA DE TIEMPOS DE FABRICACIÓN TEÓRICOS Y REALES	113
ILUSTRACIÓN 58. TIEMPOS TEÓRICOS Y REALES DE PARADAS Y TIEMPO INVERTIDO.....	114
ILUSTRACIÓN 59. PARADAS MÁS FRECUENTES POR MES.....	114
ILUSTRACIÓN 60. PRODUCTOS DEFECTUOSOS	115
ILUSTRACIÓN 61. PRODUCTOS DEFECTUOSOS DE C EN MARZO	115
ILUSTRACIÓN 62. PÁGINA 6: OPERARIOS Y LÍNEAS DE PRODUCCIÓN.....	116
ILUSTRACIÓN 63. HORAS DE TRABAJO POR OPERARIO	118
ILUSTRACIÓN 64. NO-CALIDADES POR OPERARIO	118
ILUSTRACIÓN 65. PRODUCTIVIDAD (UNIDADES POR HORA) POR OPERARIO	118
ILUSTRACIÓN 66. HORAS TRABAJADAS POR LÍNEA	119
ILUSTRACIÓN 67. RETRASOS EN PEDIDOS POR OPERARIO	120
ILUSTRACIÓN 68. NO-CALIDADES POR LÍNEA	120
ILUSTRACIÓN 69. OBTENCIÓN DE DATOS DE TWITTER EN FORMATO JSON CON REST API. FUENTE: BATRINCA ET AL (2015)	123
ILUSTRACIÓN 70. EJEMPLO DE BÚSQUEDA EN FACEBOOK GRAPH API SEARCH. FUENTE: BATRINCA ET AL, 2015 ..	124
ILUSTRACIÓN 71. ANÁLISIS DE SENTIMIENTOS. CARGA PREVIA	127
ILUSTRACIÓN 72. ANÁLISIS DE SENTIMIENTOS. PURGADO DE DATOS	128
ILUSTRACIÓN 73. ANÁLISIS DE SENTIMIENTOS. FUNCIÓN IDENTIFICA LENGUAJE	128
ILUSTRACIÓN 74. ANÁLISIS DE SENTIMIENTOS. FUNCIÓN PALABRAS CLAVE	129
ILUSTRACIÓN 75. ANÁLISIS DE SENTIMIENTOS. FUNCIÓN PALABRAS CLAVE LISTA	129
ILUSTRACIÓN 76. ANÁLISIS DE SENTIMIENTOS. FUNCIÓN RESULTADOS	129
ILUSTRACIÓN 77. ANÁLISIS DE SENTIMIENTOS. INVOCAR FUNCIONES	130
ILUSTRACIÓN 78. ANÁLISIS DE SENTIMIENTOS. TABLA CON PALABRAS CLAVE EN CADENA	131
ILUSTRACIÓN 79. ANÁLISIS DE SENTIMIENTOS. TABLA CON UNA PALABRA POR FILA (ID REPETIDO)	132
ILUSTRACIÓN 80. ANÁLISIS DE SENTIMIENTOS. NUBE DE PALABRAS PRELIMINAR	133
ILUSTRACIÓN 81. ANÁLISIS DE SENTIMIENTOS. FUNCIÓN CLASIFICADORA	133
ILUSTRACIÓN 82. ANÁLISIS DE SENTIMIENTOS. PÁGINA 1: GENERAL	135



ILUSTRACIÓN 83. ANÁLISIS DE SENTIMIENTOS. PÁGINA 2: DEPARTAMENTOS	137
ILUSTRACIÓN 84. ANÁLISIS DE SENTIMIENTOS. PÁGINA 3: ORÍGENES.....	138
ILUSTRACIÓN 85. ANÁLISIS DE SENTIMIENTOS. PÁGINA 4: PALABRAS.....	139
ILUSTRACIÓN 86. ANÁLISIS DE SENTIMIENTOS. PÁGINA 5: EVOLUCIÓN.....	140
ILUSTRACIÓN 87. ANÁLISIS DE SENTIMIENTOS. VALORACIÓN POR DEPARTAMENTO.....	142
ILUSTRACIÓN 88. ANÁLISIS DE SENTIMIENTOS. VALORACIÓN DEPARTAMENTO POR MES	142
ILUSTRACIÓN 89. ANÁLISIS DE SENTIMIENTOS. PALABRAS MÁS FRECUENTES	143
ILUSTRACIÓN 90. ANÁLISIS DE SENTIMIENTOS. PALABRAS DE PRODUCCIÓN	144
ILUSTRACIÓN 91. ANÁLISIS DE SENTIMIENTOS. EVOLUCIÓN PALABRAS PRODUCCIÓN.....	144
ILUSTRACIÓN 92. ANÁLISIS DE SENTIMIENTOS. PRODUCTO Y DESPERFECTOS EN CALIDAD	145
ILUSTRACIÓN 93. ANÁLISIS DE SENTIMIENTOS. EVOLUCIÓN PALABRAS DE ATENCIÓN AL CLIENTE.	145



Universidad de Valladolid



ESCUELA DE INGENIERÍAS
INDUSTRIALES



Índice de tablas

TABLA 1. EJEMPLO TABLA EN MODELO LÓGICO.....	16
TABLA 2. DIFERENCIAS ENTRE BD RELACIONAL Y MULTIDIMENSIONAL. FUENTE: VARGA (2001).....	20
TABLA 3. VENTAJAS Y DESVENTAJAS DE ROLAP. FUENTE: TAMAYO (2006)	24
TABLA 4. VENTAJAS Y DESVENTAJAS DE MOLAP. FUENTE: TAMAYO (2006)	25
TABLA 5. COMPARACIÓN ANALYTICS, BA, BI. FUENTE: SCHNIEDERJANS ET AL (2014).....	54
TABLA 6. CARACTERÍSTICAS DE LAS PERSPECTIVAS. FUENTE: ELABORACIÓN PROPIA.....	58
TABLA 7. COMPARATIVA HERRAMIENTAS BI (1). FUENTE: ELABORACIÓN PROPIA.....	77
TABLA 8. COMPARATIVA HERRAMIENTAS BI (2). FUENTE: ELABORACIÓN PROPIA.....	78
TABLA 9. DATOS DE PROVEEDORES	82
TABLA 10. LISTA DE MATERIALES DE PRODUCTOS.....	82
TABLA 11. DATOS DE OPERARIOS	83
TABLA 12. DATOS DE LÍNEAS	83
TABLA 13. DATOS DE PARADAS DE LÍNEA	83
TABLA 14. DATOS DE PROVEEDORES (2)	109



Universidad de Valladolid



ESCUELA DE INGENIERÍAS
INDUSTRIALES



INTRODUCCIÓN

OBJETO

El objeto del presente Trabajo de Fin de Grado es mostrar al lector la importancia, características y ventajas de la adopción de uno de los conjuntos de herramientas que más ha influenciado el aumento de la competitividad de las empresas (tanto industriales como de cualquier otro sector) en las últimas décadas: el Business Intelligence.

OBJETIVOS

Se ha decidido que la manera óptima de profundizar en el Business Intelligence es que a través de los capítulos se detallarán conceptos básicos introductorios para comprender en la medida de lo posible un ámbito muy extenso como lo es el Business Intelligence, facilitando al lector la comprensión de la aplicación real. Los objetivos que se proponen son:

1. Justificar mediante el contexto actual de la necesidad primordial de Business Intelligence en las empresas.
2. Conocer el estado del arte del Business Intelligence: definición, conceptos, arquitectura, características...
3. Mostrar las amplias funcionalidades y características de las herramientas y software de los desarrolladores más importantes del mercado.
4. Realizar una comparativa de las principales herramientas de los factores clave a considerar al adquirir una de ellas.
5. Aplicar Business Intelligence en un caso práctico que permita evidenciar las capacidades y potencial de las herramientas mediante el análisis de datos estructurados y no estructurados corporativos.





1. Contexto

1.1 Actualidad

Vivimos en un mundo informatizado y conectado. La gran mayoría de la población posee al menos un dispositivo electrónico informatizado capaz de leer o generar datos. Este a su vez, suele estar conectado a una red, ya sea local o sea Internet. Sin ofrecer ningún dato en primera instancia, y conociendo simplemente que el mundo está poblado por más de 7 mil millones de personas, podemos comenzar a tomar conciencia de la cantidad de datos que generamos o ponemos en disposición solo como individuos. A todo este conjunto se le deben sumar los datos generados por instituciones, aplicaciones informáticas, datos recogidos activamente y pasivamente por complejos sistemas informáticos, datos corporativos generados y una lista sin fin de fuentes.

A su vez, el crecimiento exponencial de la generación de datos se ha visto contribuida con la explosión de avance tecnológico sufrido en una sola década. Para ejemplificar, cualquier PC para uso personal de sobremesa hace 10 años contenía almacenamiento de 250 GB de media y tal vez un par de GB de memoria RAM. En 2019, cualquier PC portátil contiene como mínimo 1 TB de almacenamiento y 4 u 8 GB de memoria RAM por un precio asequible. Hasta los dispositivos de almacenamiento USB han sufrido este gran avance, pasando de tener hace una década hasta 2 GB de almacenamiento, a ahora como mínimo 8 GB.

El avance en la computación y en la reducción de tamaño de todos los componentes tecnológicos permite crear procesadores, almacenamiento y a su vez software, cada vez más potente.

Si deseamos otorgar a esta visión un sentido propiamente objetivo y cuantificable, podemos recurrir a estimaciones de ciertas organizaciones.

Por ejemplo, la consultora IDC ha estimado que la cantidad de datos mundial crecerá desde los 33 Zetabytes (1 Zetabyte = 10^9 Gigabytes) hasta los 175 Zetabytes en 2025 (IDC, 2018) (Ilustración 1). Este dato es incluso un 9% superior la estimación del informe del año anterior. En él, se sostiene que en este periodo hasta 2025 la cantidad de datos en la nube aumentará progresivamente hasta convertirse en la parte mayoritaria (49%), frente a los datos personales y los datos corporativos. Por otra parte, también afirma que los dispositivos IoT (Internet of Things), los cuales se interconectan entre ellos mediante Internet, crearán 90 ZB para 2025.

Por otra parte, expertos reconocidos en el ámbito digital como Philip Evans, consultor senior en Boston Consulting Group ha afirmado que el 90% de los datos ha sido creado en los últimos años y el 99% de ellos está digitalizado (BBVA, 2015).

Aunque la lista de informes, libros y artículos que tratan este tema es interminable, es fácil comprobar la importancia y el crecimiento de los datos. En parte, también, debido que a todos estos datos contienen una información que nos permitirá conocer mejor el ámbito que queremos.

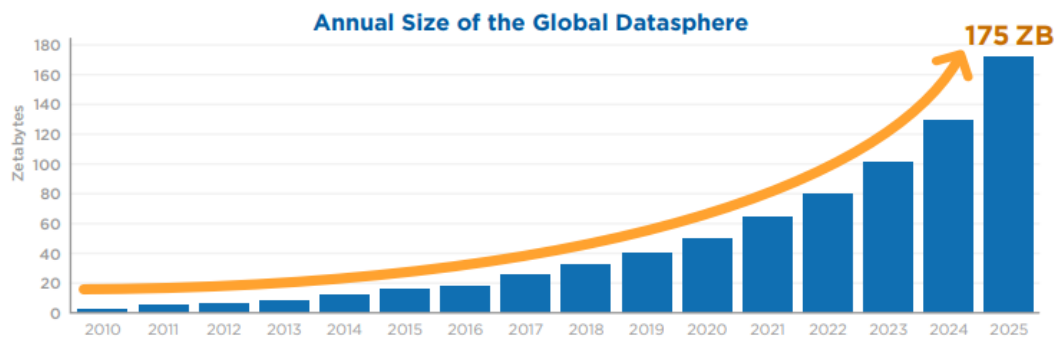


Ilustración 1. Crecimiento de los datos. Fuente: IDC (2018)

Durante la presente introducción se ha usado reiteradamente la palabra ‘datos’, y una única vez la palabra ‘información’. No ha sido casualidad. En nuestra vida cotidiana utilizamos prácticamente de manera indistinta los términos **datos, información y conocimiento**. Sin embargo, en este sentido, es crucial que comprendamos la **diferencia entre ellos**.

- ✓ Un **dato** es un conjunto discreto, de factores objetivos sobre un hecho real. Su mera existencia no facilita la toma de decisiones, es decir, valores que no contienen la explicación de los hechos ni son orientativos. En una organización son los registros del proceso de producción, registros de transacciones, registros de clientes...
- ✓ La **información** tiene un significado, es decir, relevancia y propósito para el receptor en un contexto. La información es útil para la toma de decisiones, ya que disipa dudas e incertidumbres. Los datos se convierten en información añadiéndoles valor:
 - Contextualizando: se conoce su contexto y el propósito con el que se generaron.
 - Categorizando: se conocen sus unidades de medida para su interpretación.
 - Calculando: se obtienen a través de operaciones matemáticas y procesos estadísticos.
 - Corrigiendo: se mejoran a través de la eliminación de errores.
 - Condensando: se resumen de forma más concisa.
- ✓ El **conocimiento** se define como una mezcla fluida de experiencia estructurada, valores, información contextual e internalización experta que proporciona un marco para la evaluación e incorporación de nuevas experiencias e información. (Davenport T. H., Prusak L., 1999). El conocimiento surge en los que conocen el sistema, y queda fijado en la organización mediante costumbres, procesos, prácticas o normas. Está formado por la experiencia, los criterios, las reglas o intuiciones, los valores, las creencias y la complejidad.

En la Ilustración 2 podemos observar un cuadro resumen de los tres conceptos.

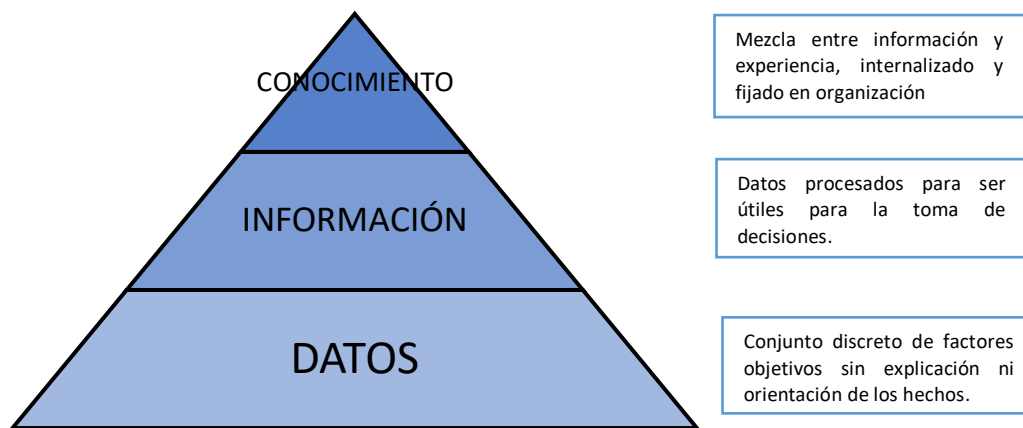


Ilustración 2. Datos, información y conocimiento

1.2 Bases de Datos

Cabe destacar que, aunque existen diferentes tipos de Bases de Datos, en este apartado se está contemplando el tipo relacional.

La generación de conocimiento es directamente producida mediante el tratamiento de datos. Un aspecto fundamental para este proceso es que todos los datos a tratar tengan una estructura definida y estable. De esta manera, es posible que las herramientas utilizadas puedan analizar de la manera deseada dichos datos y proporcionar información.

1.2.1 Definición

Denominaremos como Bases de Datos (a partir de ahora designada como [DB]) a una colección o depósito de datos integrados, almacenados en soporte secundario (no volátil) y con redundancia controlada. Los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de ellos, y su definición (estructura de la base de datos) única y almacenada junto con los datos, se ha de apoyar en un modelo de datos, el cual ha de permitir captar las interrelaciones y restricciones existentes en el mundo real. Los procedimientos de actualización y recuperación, comunes y bien determinados, facilitarán la seguridad del conjunto de los datos (Piattini Velthuis, 2006). Es decir, las principales características de los datos almacenados es que están organizados, relacionados, fácilmente accesibles, no repetidos y aislados de otros programas.

1.2.2 Componentes

Los componentes básicos de una DB son (Piattini Velthuis, 2006):

- **Datos:** Su existencia es evidente, aunque no tan evidente es los que se deben disponer (diferencias entre el ámbito que representan), los que se deben almacenar (la cantidad de datos históricos que se deben almacenar) o la manera en que el ordenador debe entenderlos (si por ejemplo, están encriptados o no).
- **Metadatos:** Dado que una DB puede tener diferentes formas en función de la necesidad que satisfaga e incluso para una misma necesidad presentar



diferentes formas, debe proporcionarse una manera de definir lo que cada dato representa (si el dato es fundamental, si puede ser omitido al introducir datos, su objetivo o la aplicación a la que va destinada). Para ello están los metadatos. Son datos sobre los datos.

- **Sistema Gestor de Bases de Datos (SGBD):** lo definiremos más adelante.
- **Usuarios de la BD:** aunque puedan ser muy variopintos, los usuarios se podrían clasificar en 4 tipos.
 - **Usuarios normales:** no conocen la estructura interna y utiliza la BD a través de aplicaciones visuales desarrolladas por programadores. Utilizan formularios.
 - **Programadores de aplicaciones:** desarrollan los programas que son requeridos por los usuarios.
 - **Usuarios sofisticados:** utilizan la BD mediante lenguaje especializado del SGDB.
 - **Usuarios especializados:** desarrollan aplicaciones de BD especializadas.
 - **Administradores de la DB:** son los responsables del control y gestión de la BD (esquemas, modelos, implementación, seguridad, copias de seguridad, etc.)
- **Elementos de seguridad:** el conocimiento de los elementos de seguridad por parte del administrador es fundamental para su gestión. De esta manera podemos identificar 3 tipos de seguridad sobre las bases de datos.
 - **Información accesible:** es posible la restricción a ciertos elementos por parte de ciertos usuarios. Es decir, que cada usuario pueda acceder exclusivamente a la información que necesita para su tarea.
 - **Operaciones posibles:** al igual que se puede restringir el acceso a información, es posible de la misma manera restringir las operaciones que se pueden realizar sobre la información (cuatro grandes grupos: Altas, Bajas, Modificaciones y Consultas).
 - **Acceso a metadatos y estructura de la DB:** como ya comentamos anteriormente, los metadatos son una parte fundamental de la DB ya que supone una medida para integrar y facilitar el entendimiento de los datos. Es por esta importancia que debe condicionarse su acceso mediante privilegios o acceso personal a ésta y a la estructura de la DB para evitar problemas severos.
- **Lenguajes de Bases de Datos:** la interacción del usuario con la DB se realiza a través del SGDB y es con esta interacción con la que se solucionarán los problemas o se aplicarán los requisitos de la DB. Sin embargo, para conseguir una comunicación efectiva debe de realizarse a través de un lenguaje lo más parecido al lenguaje natural humano. Podemos distinguir dos tipos de lenguajes.
 - **Lenguaje de definición de datos (LDD):** lenguaje utilizado en exclusiva por el administrador y es a través del cual se declaran las sentencias que marcarán las características de la DB. En resumidas cuentas, es a través del cual se crea y se gestiona la DB.
 - **Lenguaje de manipulación de datos (LMD):** el lenguaje de manipulación de datos es el que utilizan los usuarios para tratar, manipular y modificar los datos de la DB. Los usuarios son los que

utilizan este lenguaje, en resumen, para insertar, eliminar, modificar y consultar datos o editar opciones de visualización.

1.2.3 Modelo entidad-relación

Aunque una base de datos representa solamente una porción de la realidad, la relación entre los datos ha de estar completamente definida. Es necesario **inicialmente** el diseño del modelo **entidad-relación**, basado en una colección de objetos básicos llamados entidades y de relaciones entre ellos que es representado por un diagrama (Ilustración 3). Los elementos de dicho modelo son:

- **Entidades:** representan los “objetos” o “cosas” de la porción de realidad, y son claramente diferentes. En el diagrama son representados mediante rectángulos.
- **Atributos:** definen e identifican las características de las entidades. Cada entidad puede tener varios atributos y además, de diferente forma (texto, número, fecha...). Son representados mediante elipses unidos a una entidad.
 - **Simple y compuestos.** Define si pueden ser divididos en subpartes o no. Ejemplo: el atributo “Nombre” puede ser simple si mediante él expresamos el nombre completo o compuesto si “Nombre” lo dividimos en “Nombre de pila”, “Primer apellido” y “Segundo Apellido”.
 - **Monovalorados y multivalorados.** Define si los atributos pueden tomar un único valor o varios. El concepto de monovalorado es común y entendible por todos. Un atributo multivalorado, por ejemplo, es si un empleado tiene 0, 1 o varios números de contacto personales.
 - **Derivados.** Son aquellos que se pueden obtener a partir del valor de otros atributos o entidades relacionados. Por ejemplo, si deseamos el “Número de compras de un cliente”, basta con contar cada una de las “Compras” con el “Nombre” de dicho cliente.
- **Relación:** vinculan la dependencia entre entidades, de manera que las entidades puedan compartir atributos. En el diagrama se representan mediante rombos unidos a las entidades.
- **Relaciones de cardinalidad:** es la restricción que interviene para delimitar cuántos “objetos” o “cosas” pueden intervenir entre una entidad y una relación.
 - **Uno a uno.**
 - **Uno a varios o varios a uno. (1,N), (N,1)**
 - **Varios a varios (N, N)**

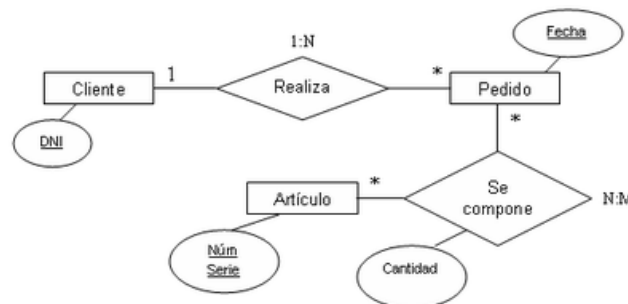


Ilustración 3 Ejemplo de modelo entidad-relación (recuperado de: https://es.wikipedia.org/wiki/Modelo_entidad-relaci%C3%B3n)

El siguiente paso es convertir este modelo entidad-relación en un **modelo lógico o relacional**.

1.2.4 Modelo relacional

Tras haber obtenido el modelo entidad-relación, transformamos nuestro modelo a uno lógico. Conseguiremos un diagrama parecido al anterior pero **con tablas**. En dicha tabla, cada **columna** representa un **atributo**, cada fila una tupla (un valor introducido) y el **nombre** de cada tabla una **relación**. Además, debemos asignar las claves en cada tabla y los dominios de los atributos:

- **Claves:** distinguen a los atributos o aplican un vínculo entre entidades.
 - **Clave primaria (PK):** identifica inequívocamente un solo atributo en una entidad, de manera que **no se pueda repetir**.
 - **Clave foránea (FK):** identifica en una entidad la clave primaria de otra entidad, vinculándolas y relacionándolas.
- **Dominios:** descripción del intervalo de valores que puede tomar un atributo.

Definiremos, por otra parte, a los Sistemas de Gestión de Bases de Datos (SGBD) como las aplicaciones que permiten el almacenamiento, modificación y extracción de información de una BD. Es la capa intermedia entre la BD y el usuario y es de lo que depende el lenguaje de implementación de la BD.

En la Ilustración 4 se muestra un Modelo Relacional típico.

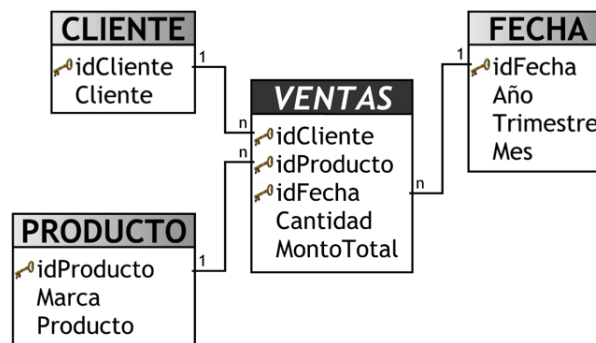


Ilustración 4. Ejemplo de Modelo relacional (recuperado de: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/-metodologia-hefesto/55-pasos-y-aplicacion-metodologica/553-p>)

La tabla Producto de la anterior ilustración podría tener una forma como en la Tabla 1.

idProducto	Marca	Producto
1	Samsung	Smartwatch
2	Samsung	Teléfono móvil
3	Apple	Auriculares
4	Huawei	Teléfono móvil
...

Tabla 1. Ejemplo tabla en modelo lógico

Este modelo lógico es el que se ha de implantar a través de un SGBD con el lenguaje LDD descrito anteriormente. En el ámbito de los SGBD existen muchos software



diferentes, ya sean libres o de pago, con lenguaje SQL. A muestra de resumen podemos mencionar los más populares:

- MariaDB
- MySQL
- SQLite
- PostgreSQL
- Microsoft SQL Server
- ORACLE

El esquema elegido para construir una base de datos relacional es fundamental a la hora de buscar un mejor rendimiento a la hora de implementarla. Es por ello, que en la Ilustración 4 anteriormente presentada, el esquema mostrado no es fruto de la casualidad.

Existen 2 esquemas fundamentales y principales a la hora de hablar de Business Intelligence basado en bases de datos relacionales: esquema en estrella y esquema en copo de nieve.

ESQUEMA EN ESTRELLA

La Ilustración 4 presenta un esquema en estrella. En dichos esquemas existe una tabla central denominada de hechos rodeada de tablas de dimensiones. La particularidad del esquema es que las tablas de dimensiones poseen una clave primaria simple y la tabla de hechos está formada por estas claves primarias de las tablas de dimensiones (actuando como “*Foreign Key, FK*”).

Este esquema es ideal por su simplicidad y la rápida respuesta en las consultas. Permite establecer niveles y jerarquizar los datos para formar agrupaciones.

ESQUEMA EN COPO DE NIEVE

El esquema en copo de nieve es un esquema más complejo que el esquema en estrella, tal y como se aprecia en la Ilustración 5. En este, las dimensiones son implementadas con más de una tabla. Se busca la normalización de los datos (una serie de reglas y formas para evitar la redundancia y proteger la integridad de los datos y disminuir problemas de actualización de los datos en las tablas) y reducir el espacio a la hora de almacenar.

La principal ventaja de este esquema es, como ya hemos comentado, evitar la redundancia y ahorrar espacio. Sin embargo, la adición de más tablas (comparado con el esquema en estrella) produce la necesidad de más relaciones (“*Joins*”) y por tanto, necesidad de un **código más extenso y complejo para realizar consultas**.

Tanto este esquema como el esquema en estrella son utilizados en Data Warehousing, lo que es fundamental para implementar OLAP (particularmente, ROLAP) y en general cualquier técnica de Business Intelligence, para poder ofrecer un análisis multidimensional del mismo modo que lo podría hacer una base de datos multidimensional (utilizadas en MOLAP).

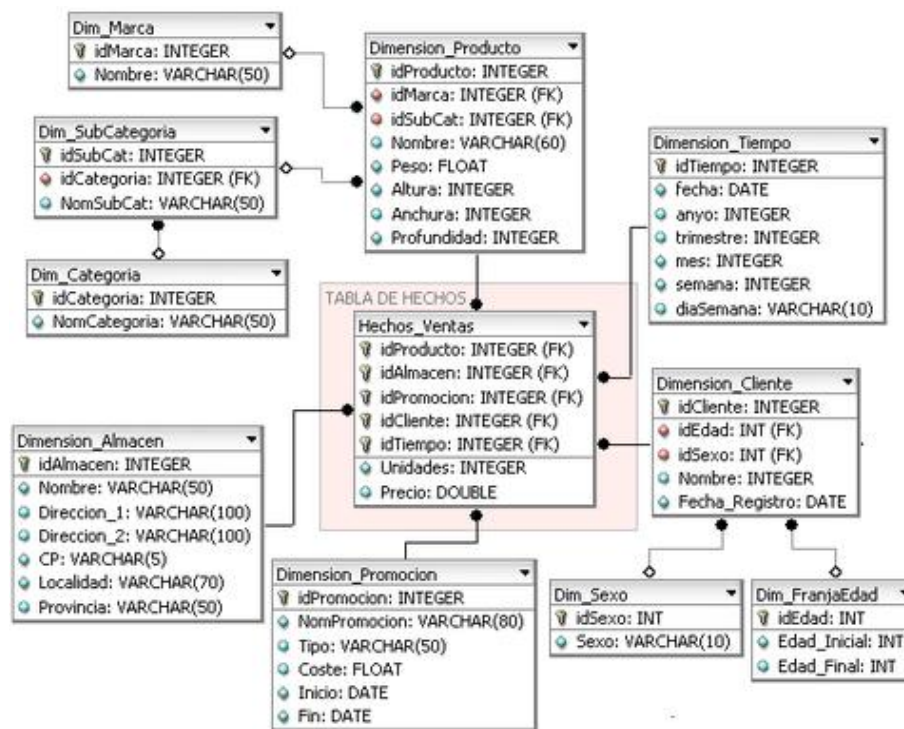


Ilustración 5. Esquema en copo de nieve (recuperado de: https://es.wikipedia.org/wiki/Esquema_en_copo_de_nieve)

1.2.5 Bases de datos multidimensionales

Las bases de datos relacionales, como ya hemos visto, basan su funcionamiento en consultas mediante lenguaje SQL, provocando que sea complicada la capacidad de obtener información más sofisticada. Además, el progreso tecnológico y el aumento de la competitividad industrial han hecho que sea necesario la respuesta de preguntas en ambientes corporativos que mediante lenguaje de SQL es como poco, tedioso. Por ello, surgen los modelos de datos multidimensionales.

La base de datos multidimensional es definida como una variación del modelo relacional que usa estructuras multidimensionales para organizar los datos y mostrar las relaciones entre ellos (O'Brien, Marakas, 2009). Es decir, los datos se almacenan en celdas y cada posición de celda está definida por una serie de variables denominadas dimensiones. El **modelo de datos multidimensional** cuenta con tablas de hechos con atributos cuantitativos y tablas de dimensiones con cualitativos.

Cada tabla almacena registros de la forma:

D₁, D₂, D₃,...M₁, M₂, M₃,...

De manera que cada **D_i** representa una **Dimensión**:

- Describe un ámbito de la empresa.
- Describe la organización de los datos.
- Facilita un método para analizar datos de la empresa.

De manera que cada **M_i** representa una **Medida (hecho)**:

- Cuantitativos (numéricos)
- Presentes en todas las dimensiones en todos los niveles.
- Son indicadores que representan sumas, promedios, totales, porcentajes...

Teniendo en cuenta estos factores, se muestra en la Ilustración 6 un posible modelo multidimensional.

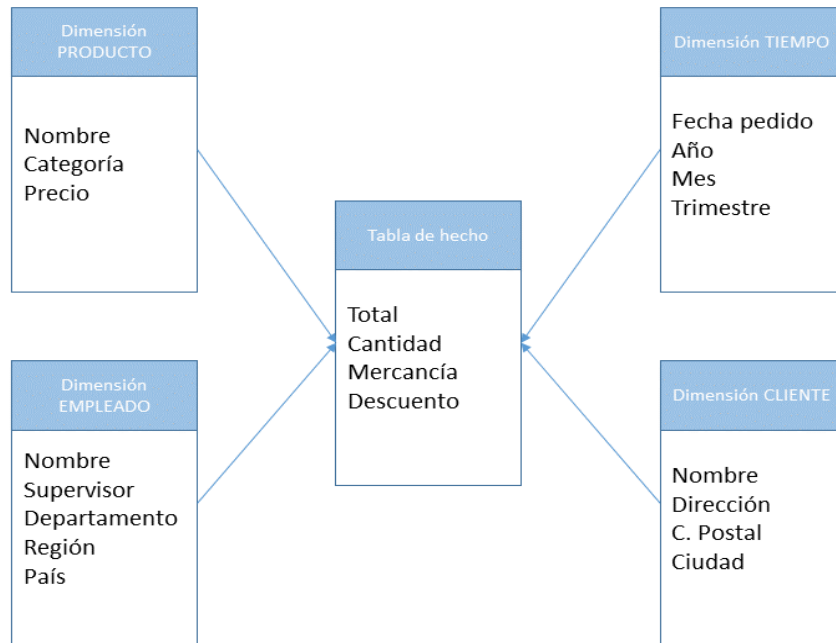


Ilustración 6. Ejemplo Modelo Multidimensional. Fuente: Elaboración propia.

El modelo que observamos está simplificado para su entendimiento (pareciéndose a uno relacional), pero **en la práctica se implanta mediante vectores y matrices multidimensionales en el SGDB pertinente**. Como vector o matriz multidimensional podemos entender lo que en tres dimensiones es un cubo, aunque puede contener más dimensiones (hipercubos). **Desarrollaremos esta idea de cubos en siguientes secciones.**

Además, un concepto fundamental en las bases de datos multidimensionales es la **jerarquía**. Esta jerarquía determina los niveles de agregación y desagregación (qué contiene a que, o qué está contenido en qué) de las **dimensiones**.

Por ejemplo podemos observar que la jerarquía de la dimensión **TIEMPO** es: Año → Trimestre → Mes. Cada una de estas agregaciones se denomina **niveles**.

Es por tanto, que un modelo multidimensional de datos permite una fácil navegación y consulta. Existe en él una menor cantidad de tablas y relaciones respecto al modelo relacional clásico y es más flexible en cuanto al diseño para su implementación. Las bases de datos multidimensionales reúnen, entre otras, las siguientes ventajas:

- Mejora del rendimiento.
- Fácil gestión y mantenimiento.
- Consulta y entendimiento de los datos fácil e intuitiva.
- Facilidad de manipulación de datos mediante interfaces.



Sin embargo, aunque se haya citado que su diseño es más sencillo, la principal desventaja de las bases de datos multidimensionales es que su propia **implementación es compleja y requiere de equipos de expertos** que entiendan y comprendan en profundidad la porción de realidad que se quiere representar (un proceso productivo, por ejemplo).

A modo de resumen, en la Tabla 2 se expone un cuadro comparativo de los modelos relacionales y multidimensionales (Varga, 2001).

	Relacional	Multidimensional
Objetivo/uso	Sistemas OLTP	Sistemas OLAP o DSS
Objeto de análisis	Ejecución del proceso productivo (flujo del proceso)	Efectos del proceso productivo
Interés del análisis	Descubrir entidades importantes en el proceso productivo	Descubrir entidades asociativas que representen los efectos y consecuencias del proceso productivo
Detalles del análisis	Definición de los atributos de estas entidades importantes y las relaciones entre ellas	Definición de indicadores y medidas del proceso productivo (atributos de las entidades asociativas, definición de las dimensiones)
Lenguaje	SQL	MDX

Tabla 2. Diferencias entre BD Relacional y Multidimensional. Fuente: Varga (2001)

Las bases de datos multidimensionales han sido creadas para los Data Warehouses y para aplicaciones OLAP, de los que hablaremos en las siguientes páginas.

1.3 OLTP y OLAP

1.3.1 OLTP (OnLine Transaction Processing)

En un primer momento surgieron las bases de datos como concepto útil para multitud de campos. Las empresas pronto entendieron las capacidades y ventajas de éstas. Sin embargo, las bases de datos en una empresa, debían de soportar el acceso simultáneo por diferentes usuarios sin entorpecerse entre ellos ni poner en riesgo la integridad y seguridad de la DB.

Es por ello que surgieron los sistemas **OLTP** (Online Transaction Processing) para la captura de datos común diaria en cualquier organización. Es decir, son **sistemas que facilitan la administración de entrada y procesamiento de los datos** en las DB. Están adaptados para un registro masivo de grandes volúmenes de datos en DB en tiempo real mediante estructuras de datos optimizadas para la adición, modificación y eliminación (INSERT, UPDATE, DELETE) de nuevos datos (normalmente de uno en uno). En definitiva, los sistemas OLTP definen el comportamiento operacional de cualquier industria con un enfoque de procesamiento de datos, sirviendo de escasa utilidad para la toma de decisiones.



La ventaja de adoptar un sistema OLTP es la unificación de las fuentes de adición de nuevos datos al sistema y la gestión automática de los datos. Además, es un sistema consistente y seguro, proporcionando fiabilidad, disponibilidad y seguridad, ya que mediante simples opciones se pueden regular los accesos y privilegios de usuarios. Su manejo es fácil y su evolución o cambios son nulos, debido a la simplicidad del sistema y a la poca capacidad de innovación de ellos. Sin embargo, como se ha citado, la capacidad de los sistemas OLTP es meramente transaccional, sin posibilidad de análisis u obtención de información y conocimiento directamente relacionado o proporcionado por los datos.

Unos ejemplos típicos de sistemas OLTP son un cajero automático, el cual actualiza los datos de las cuentas bancarias de los clientes aunque se acceda simultáneamente, tal y como puede ocurrir con cuentas conjuntas, o una plataforma de reservas de billetes de avión, la cual actualiza los datos de los asientos ocupados aunque estén reservando varias personas a la vez.

1.3.2 OLAP (OnLine Analytical Processing)

La base y origen de OLAP se remonta a los años 60 con la publicación de Ken Iverson "A Programming Language", el cual traía consigo un lenguaje matemático con operadores de procesamiento y variables multidimensionales. Este lenguaje fue resumido por el propio Iverson a uno basado en notación mediante símbolos griegos, lo que supuso la necesidad de teclados, equipos e impresores especiales para su implementación en ordenadores IBM, convirtiéndolo en uno muy caro y considerado además como complejo e ineficiente, consumiendo demasiados recursos en los equipos.

En décadas más recientes, expertos y fabricantes han comenzado a denotar al OLAP como un conjunto de aplicaciones que deben cumplir una serie de requisitos (tal y como mostraremos más adelante con las 12 normas para un programa OLAP de Ted Codd) e incluso se consideran programas OLAP anteriores a las propias definiciones formales posteriores (véase el programa Express de los años 70, posteriormente adquirido por ORACLE).

La característica diferenciadora del OLAP es la estructuración de las fuentes de datos. El objetivo de las herramientas OLAP es realizar consultas rápidas en grandes cantidades de datos y debido a ello se necesita de bases de datos con modelo dimensional (ya sea relacionales o multidimensionales).

Sin embargo, hasta ahora no hemos dado una **definición explícita** de lo que son las **herramientas OLAP**. Las herramientas OLAP son herramientas implementadas en equipos informáticos (incluidas dentro del ámbito del Business Intelligence) que se caracterizan por permitir un análisis multidimensional. Los datos se han de ser modelizados mediante medidas, dimensiones y hechos. Las consultas OLAP generadas se formulan sin conocer exactamente la estructura o procedencia de los datos, permitiendo un conocimiento completo del proceso. Considerado como un sistema analítico, OLAP permite a usuarios y organizaciones el descubrimiento de causas de hechos ocurridos o información compleja. Un ejemplo:



“¿Cantidad de compras durante el último trimestre de los productos Mesa por departamento y mes?”

Observamos que en dicha consulta existen los elementos:

- **Medida:** IMPORTE
- **Hecho:** COMPRAS
- **Restricciones:** producto Mesa, compras del último trimestre
- **Parámetros de consulta:** por departamento y por mes.

Este tipo de consulta en lenguaje SQL es complejo. OLAP permite esta visualización mediante un lenguaje MDX (*Multidimensional Expressions*) mucho más simple, convirtiendo un código largo en uno de una o dos líneas. Aunque esta diferencia sea abismal respecto de consultas basadas en SQL (directamente sobre la base de datos), la principal y más importante diferencia (y además ventaja) es la posibilidad de “subir o bajar” en los niveles. Tal y como explicamos en las bases de datos multidimensionales (también presentes en esquemas de estrella y copo de nieve), las jerarquías juegan un papel importante en este tipo de bases de datos. Y su importancia es clara en estos ejemplos, ya que fácilmente podremos cambiar la restricción de “último trimestre” por “último año”, y no supondrá para el sistema ningún esfuerzo de procesamiento ni bajadas de rendimiento.

Todas las herramientas OLAP están formadas por un motor y una interfaz. El motor es todo aquello que hasta ahora hemos definido (conceptos y manera de funcionar) y la interfaz es aquella que permite al usuario la consulta y la manipulación de datos.

En definitiva, las herramientas OLAP son útiles para la consulta y obtención de información corporativa, además, algunas de ellas disponen de herramientas para la creación de informes gráficos con dicha información, mediante un método mucho más sencillo en comparación con sistemas tradicionales.

Cualquier herramienta OLAP ha de cumplir las **12 reglas formuladas por Ted Codd** (Codd et al, 1993) en cuanto a funcionalidades de herramientas OLAP:

- **Vista Multidimensional**
Esta vista es fácilmente entendible mediante el concepto de niveles y jerarquías, donde podemos elegir si queremos “mes” o “año”, por ejemplo, lo cual está relacionado con las operaciones posibles en los cubos.
- **Transparencia**
Las herramientas OLAP han de soportar diferentes fuentes de datos, el usuario no debe tener conocimientos ni importarle los detalles específicos acceso a datos o transformaciones y deben poder ser alojadas en cualquier sistema.
- **Accesibilidad**
La herramienta OLAP ha de servir como interfaz de acceso de consulta a los datos, sea cual sea el origen y estructura de estos, sirviendo como paso intermedio para la obtención de información.
- **Reporting consistente**
El rendimiento no ha de verse disminuido al aumentar el número de dimensiones.



- **Arquitectura cliente/servidor**
La herramienta OLAP ha de proporcionar el máximo rendimiento independientemente de la cantidad de clientes que se conecten al servidor, el cual ha de ser capaz de integrar bases de datos de diferentes orígenes.
- **Dimensionalidad genérica**
No debe estar restringida a 3 dimensiones ni a ningún número particular de dimensiones, incluso a 1 sola. Además, todas han de ser iguales, y por tanto, se les puede aplicar las mismas operaciones.
- **Manejo de matrices dispersas dinámicas**
Relacionado con la idea del “NULL” en bases de datos relacionales. La estructura del servidor de la herramienta OLAP ha de manejar de forma óptima las matrices dispersas, es decir, identificar valores nulos o vacíos.
- **Soporte multiusuario**
Debe soportar el acceso de varios usuarios simultáneos, aplicando operaciones a los cubos o vistas individuales de la misma base de datos.
- **Sin restricciones en las operaciones sobre dimensiones**
Las dimensiones son creadas por definición como iguales, por lo que todos los cálculos deben ser posibles sobre todas las dimensiones, no solo sobre las dimensiones de medida (definidas como M_i).
- **Manipulación de datos intuitiva**
Los usuarios han de poder disfrutar de características “drag-and-drop” (arrastrar y soltar) en vez de necesitar operaciones complejas o menús.
- **Reporting flexible**
La herramienta ha de tener la posibilidad de presentar la información como el usuario desee o exclusivamente la información que necesite, mediante opciones de personalización.
- **Dimensiones y niveles ilimitados**
Debe soportar al menos, como mínimo 15 dimensiones. Aunque es recomendable 20 dimensiones.

Anteriormente se comentó que las bases multidimensionales suelen ser una opción óptima para herramientas OLAP, pero su implementación puede ser compleja. Es por ello que pueden, y suelen, ser usados arquitecturas de **bases de datos relacionales** con esquema **en estrella o copo de nieve**.

Es con esta diferencia donde surgen los diferentes tipos de **OLAP: ROLAP, MOLAP y HOLAP**.

ROLAP

Relational OLAP es la versión de OLAP en la que los datos están organizados en bases de datos relacionales. ROLAP está enfocado a un sistema de consultas al alcance de cualquier empresa, renunciando a cierta velocidad a cambio de escalabilidad y ahorro de recursos.

ROLAP utiliza una arquitectura de tres niveles: la base de datos relacional, el nivel de aplicación (el motor OLAP) y una herramienta especializada para la presentación de los datos.



Los datos de sistemas transaccionales han de ser cargados al Data Warehouse (**se explicará en el capítulo 2**) después que su modelo se ha definido. Así pues, los usuarios son capaces de realizar consultas mediante la transformación de la petición por parte del motor OLAP a lenguaje SQL, lo cual se ejecuta en la DB relacional y los resultados son devueltos al usuario.

El motor OLAP se unifica con la herramienta de presentación para ejecutar y visualizar las consultas de los clientes.

ROLAP también brinda la oportunidad de preprocesar los datos y tener estructuras preparadas o incluso carga dinámica de datos, sin embargo, estos procesos no están optimizados para lenguaje SQL. Algunas de las ventajas y desventajas de ROLAP se muestran en la Tabla 3 (Tamayo, 2006).

ROLAP	
VENTAJAS	DESVENTAJAS
-Se puede aprovechar la tecnología relacional, ya que se facilita aprovechar las inversiones realizadas en hardware y en SGBD clásicos relacionales.	-Menor rendimiento frente a MOLAP en cantidades pequeñas de datos y pequeñas DB
-Uso de la seguridad e integridad de los SGBD relacionales.	-Limitación a la hora de realizar consultas complejas.
-Capacidad de almacenar y gestionar grandes cantidades de datos con un rendimiento aceptable.	-Utilización de mucho almacenamiento en disco.
-Existencia de SGDB gratuitos	
-Soporte de un número muy alto de dimensiones.	
-Escalabilidad (capacidad de aumentar dimensiones y cantidad de datos de forma considerable)	

Tabla 3. Ventajas y desventajas de ROLAP. Fuente: Tamayo (2006)

MOLAP

Multidimensional OLAP (MOLAP) es la versión clásica y más pura de OLAP, implementada a través de bases de datos multidimensionales (MDDB), cuyo objetivo es un sistema veloz y eficaz en las consultas.

En este sistema se utiliza una arquitectura de dos niveles: la base de datos multidimensional y el de aplicación (motor analítico). La base de datos es la encargada de gestionar y consultar los datos, mientras que el nivel de aplicación es el encargado de la ejecución de las consultas OLAP y la presentación de la información a través de una interfaz visual.

La facilidad de movernos, anteriormente comentada, a través de los niveles es a la vez +una desventaja en MOLAP. La posibilidad de realizarlo está condicionado a que los



cálculos y consultas en MOLAP están predefinidos y precalculados en ficheros y cualquier nueva combinación de datos, o adición de nuevas dimensiones deben ser recalculadas. OLAP basa su funcionamiento en cubos (o hipercubos), jugando con los niveles y las jerarquías y las operaciones posibles en ellos, y son estos que han de ser recalculados cada vez que deseemos actualizar nuestra base de datos. **Estos cubos serán explicados más adelante en el capítulo 2.**

Algunas de las ventajas y desventajas de MOLAP se muestran en la Tabla 4 (Tamayo, 2006).

MOLAP	
VENTAJAS	DESVENTAJAS
-Ofrece buen rendimiento cuando se trabaja sobre datos agregados, totales y subtotales, series temporales y diversos grados de detalle de los datos.	-Dificultad de asimilación de los conceptos de multidimensionalidad (especialmente a partir de 4 dimensiones, lo que se denominan hipercubos).
-Facilita el estudio a alto nivel de los datos, al ofrecer una mayor flexibilidad y rapidez de acceso para el análisis de los mismos.	-Las MDDDB y sus estructuras, además de la introducción de los datos puede tomar tiempo y esfuerzo significantes.
-Eficiencia en almacenamiento y consultas.	-Limitación de la cantidad de dimensiones existentes para un manejo relativamente asequible.
-La complejidad de la MDDDB se oculta a los usuarios.	-Necesidad de recargar los cubos cada vez que se incluyan nuevos datos.
-Análisis sobre datos precalculados y agregados en diferentes niveles.	-Limitación en la cantidad de datos que se puede utilizar.
-Mayor rendimiento frente a ROLAP en DB pequeñas.	-La mayoría de herramientas que lo soportan son de pago.

Tabla 4. Ventajas y desventajas de MOLAP. Fuente: Tamayo (2006)

HOLAP

Hybrid OLAP es la nomenclatura que reciben los OLAP que poseen una estructura híbrida en la que se combina tanto DB Relacionales como DB Multidimensionales, es decir, ROLAP y MOLAP. En general, se puede resumir que en las DB Relacionales se almacenan las mayores cantidades de datos detallados y las DB Multidimensionales almacenan la cantidad más pequeña de datos poco detallados o directamente agregados.

Y es exactamente ahí donde reside su mayor punto débil: la necesidad de realizar un buen análisis para identificar dónde implantar MOLAP y donde implantar ROLAP.



OTROS TIPOS

Aunque estos tres tipos mencionados son los más destacados y expandidos, existen otros tipos de OLAP tales como:

- **DOLAP (Desktop OLAP):** está orientado a que el usuario pueda acceder a parte de la información que necesita en la DB, descargarla en su equipo (escritorio) y realizar las consultas que necesite. Tiene un coste mucho menor que cualquier otro tipo de OLAP, sin embargo, las funcionalidades y posibilidades de éste son bastante más limitadas.
- **WOLAP (Web OLAP):** basa su funcionamiento en permitir al usuario el acceso a través del navegador web. Difiere del resto de tipos en que no usa una arquitectura cliente/servidor si no una arquitectura de tres niveles: cliente, middleware y una base de datos. La gran ventaja es la accesibilidad desde cualquier dispositivo con conexión a internet y una fácil instalación y uso. Sin embargo, la desventaja de usar un navegador web limita las funcionalidades, visualización e interfaz.
- **IMOLAP (In-Memory OLAP):** consiste en un enfoque moderno por el cual las compañías de desarrollo de software están optando. Pretende construir la estructura multidimensional únicamente a nivel de memoria y almacenar los datos en cierto formato que disminuya su tamaño y potencie el rendimiento. Un ejemplo es utilizar bases de datos de lógica asociativa.
- **SOLAP (Spatial OLAP):** antes de explicar lo que es SOLAP, debemos conocer lo que es GIS (Geographic Information Systems)
 - Geographic Information System es un conjunto de herramientas que pretende una manipulación, integración y análisis de datos que tienen una referencia espacial (mapeado de los datos) para poder incluir información social, cultural o económica que permita una mejor toma de decisiones.

SOLAP pretende la integración de GIS y OLAP en una sola herramienta.



2. Business Intelligence

2.1 Definición e historia

Hasta ahora hemos explicado OLAP y sus diferentes tipos. Sin embargo, también hemos de recordar que OLAP es una tecnología que surge varias décadas atrás. En la última década, con el crecimiento exponencial de la tecnología y la computación, además de los avances en tecnologías como la IA (Inteligencia Artificial), se ha comenzado a escuchar de manera recurrente el término **Business Intelligence** (desde ahora denominado **[BI]** por comodidad), el cual da un paso más allá que el OLAP. El término es conocido, o al menos ha sido oído por cualquier empresa de un tamaño considerable y ciertamente informatizada. Sin embargo, una definición concisa del término es difícil de dar.

Atendiendo a una **definición teórica formal**, se entiende por BI al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización. (Conesa et al, 2010).

Para un concepto más práctico, podemos entender por BI a todos aquellos recursos que permiten que nuestras decisiones empresariales sean basadas en información objetiva.

Para encontrar el origen del término debemos remontarnos a 1958 cuando Hans Peter Luhn cita por primera vez el término y lo define como “la habilidad de aprender las relaciones de hechos presentados de forma que guíen las acciones hacia una meta deseada” (Luhn, 1958). Debido a su carácter tan generalista, el término abarcaba un amplio campo de investigación.

En los años 70 surgieron las primeras (y aunque rudimentarias y costosas) aplicaciones y programas de ayuda a la toma de decisiones, algo muy innovador considerando el carácter plenamente informativo y de control del resto de programas en empresas. Con el paso de los años y el nacimiento de las bases de datos, surgieron nuevas herramientas para soporte a la dirección, de procesamiento analítico en línea (OLAP) o programas de análisis predictivo.

En los años 80 surge el concepto del almacén de datos (conocido popularmente por su anglicismo, **Data Warehouse**, el cual a partir de ahora denominaremos **[DW]**), el cual será básico para la proliferación del BI.

En 1989, Howard Dressner (analista de Gartner Group) dio voz y popularidad al término de Business Intelligence. Todas aquellas herramientas meramente analíticas y otras de nueva creación, destinadas al soporte de la toma de decisiones mediante el acceso a bases de datos y datos estructurados, se comenzarían a denominar como herramientas de Business Intelligence.

Los **años 90** se caracterizan por el surgimiento del **BI 1.0**. Su nacimiento es consecuencia directa de la capacidad de gestión de datos y los DW. Se basa



fuertemente en tecnologías de recolección, extracción y análisis de datos dentro de las organizaciones. La creación de **Data Marts** (a partir de ahora nombrado por **[DM]**) y herramientas de extracción, transformación y carga (conocidas por su anglicismo **ETL** (“*Extraction, Transformation and Load*”) son fundamentales para la integración de datos. Las herramientas de BI 1.0 son responsabilidad directa del departamento de IT y son utilizadas para consultas en base de datos, OLAP y reporting mediante gráficos intuitivos basado en un modelo cliente-servidor. También son utilizadas para la medición del desempeño o rendimiento (**Business Performance Management**, a partir ahora nombrado como **[BPM]**) mediante cuadros de mando que reúnen diferentes métricas que determinen el rendimiento de la empresa en ciertos aspectos, más conocidos como indicadores de desempeño (**Key Performance Indicator**, a partir ahora nombrado como **[KPI]**) (Chen et al, 2012). En resumen, son herramientas dedicadas a analizar datos históricos para evitar la repetición de errores, es decir, un sistema reactivo.

Durante la **primera década de los 2000**, comienza a aparecer el **BI 2.0**. La revolución producida en Internet y en las páginas web con la aparición de motores de búsqueda (Google, Yahoo...) y páginas de comercio electrónico (Amazon, eBay) permitió a las organizaciones presentarse al mundo con sus propias páginas web de ventas. Para mejorar y ofrecer una mejor oferta, se comenzó a analizar el comportamiento de usuarios gracias a las cookies y registros de servidores. En definitiva, el **BI 2.0 se centra en las personas** y es el punto de partida para el **análisis de información no estructurada** como el comportamiento de usuarios (texto, opiniones y sentimientos publicados en RRSS, foros o páginas web) para ofrecer publicidad, ofertas, posicionamiento de productos o diseños de páginas web completamente personalizadas al usuario. Las herramientas de BI 2.0 se convierten a un modelo web donde no es responsabilidad exclusiva del departamento de IT y en el que el objetivo último es la exploración de datos y la predicción.

La **última década** se ha caracterizado por el aumento de la conexión a Internet de los dispositivos. Más allá de móviles o portátiles, vivimos en la **era del ‘IoT’** (“*Internet of Things*”) donde cualquier aparato cotidiano se puede conectar a Internet. Es por ello que el **BI 3.0** está o estará centrado en ellos. En un mundo predominado por aparatos con sensores y conectados a Internet, la nueva fuente de datos son éstos. Por otra parte, las herramientas de BI también se han adaptado fuertemente a datos no estructurados y versiones móviles, ofreciendo un servicio de análisis en tiempo real y en cualquier lugar, convirtiéndose prácticamente en herramientas dedicadas a la anticipación. A su vez han focalizado sus esfuerzos en herramientas intuitivas que no necesiten de alto grado de conocimientos o formación. También se ha externalizado el acceso de los datos de las organizaciones permitiendo que la red de clientes o proveedores pueda acceder a ellos.

Como ayuda al lector, se ha realizado en la Ilustración 7 un resumen comparativo desde el BI 1.0 al BI 3.0.

En la ilustración 7 se ha destacado el carácter de los diferentes BI, que en particular, se refiere al tipo de análisis que ofrece. Es decir, se trata de análisis descriptivo, predictivo y anticipativo. Es la característica más importante, rompedora y

diferenciadora de cada uno de ellos. Haremos hincapié en cada uno de ellos mediante sus definiciones, sus propósitos y ejemplos de aplicación.¹

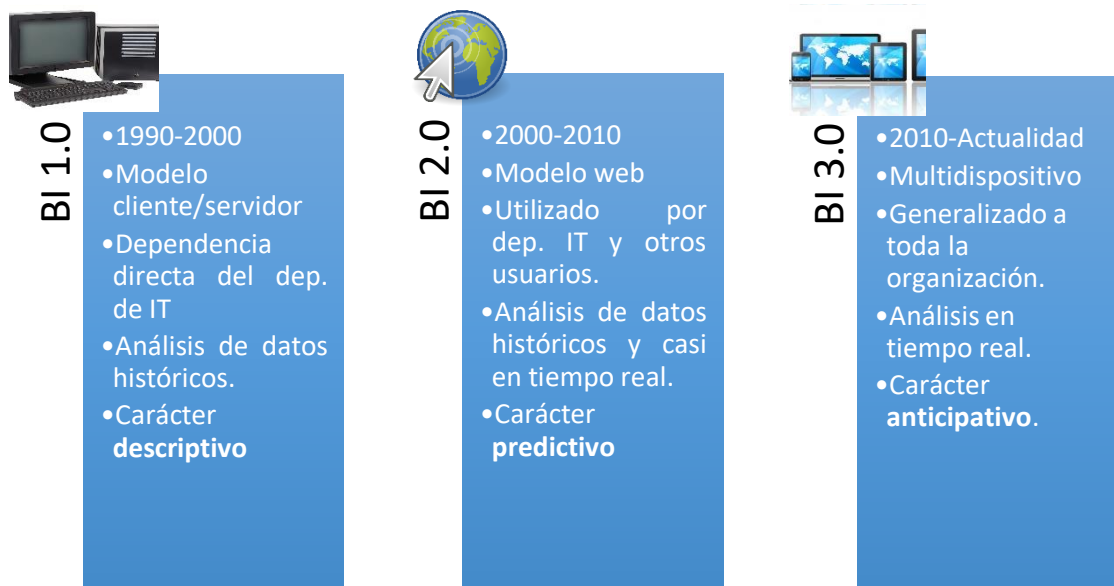


Ilustración 7. Comparativa de los BI. Fuente: elaboración propia

- **Análisis descriptivo**
Supone la aplicación de técnicas estadísticas simples que definen y describen los datos contenidos en una DB. Es decir, pretende la identificación tendencias y patrones en los datos para obtener una visión muy genérica de lo que suponen los datos. Entre los ejemplos típicos se encuentran: media, mediana, moda, medidas de dispersión, gráficas, distribuciones de probabilidad, etc.
- **Análisis predictivo**
En este caso, es la aplicación de técnicas sofisticadas estadísticas, software de información, etc. para la identificar las variables influyentes de los datos para generar un modelo predictivo, que permita identificar relaciones y tendencias que no fueron detectadas durante un análisis descriptivo. Los ejemplos típicos son: regresiones múltiples, ANOVA, minería de datos, clasificaciones...
- **Análisis anticipativo**
Es la aplicación de la teoría de dirección y decisiones junto a metodologías de optimización de procesos para hacer la mejor asignación de recursos para el cumplimiento de objetivos. Pretende asignar los recursos de manera óptima para tomar ventaja de las tendencias o futuras oportunidades descubiertas por el análisis predictivo. Ejemplos: programación lineal.

¹ Adaptado de (Schniederjans et al, 2014)

2.2 Elementos y arquitectura del BI

Los proyectos de BI en cualquier organización han de estar respaldados por un plan de proyecto. En él, es fundamental la creación y formalización de la arquitectura del sistema de BI. Por ello, en este Trabajo de Fin de Grado debemos de ahondar en ello.

Aunque cada plan y caso es único, podemos generalizar una arquitectura básica, común y lógica para la implementación del sistema. En él, se determinan los elementos del sistema y los flujos de datos, además de las transformaciones que sufren.

En un primer lugar plantearemos una arquitectura y posteriormente, se analizarán sus componentes y procesos (Ilustración 8).

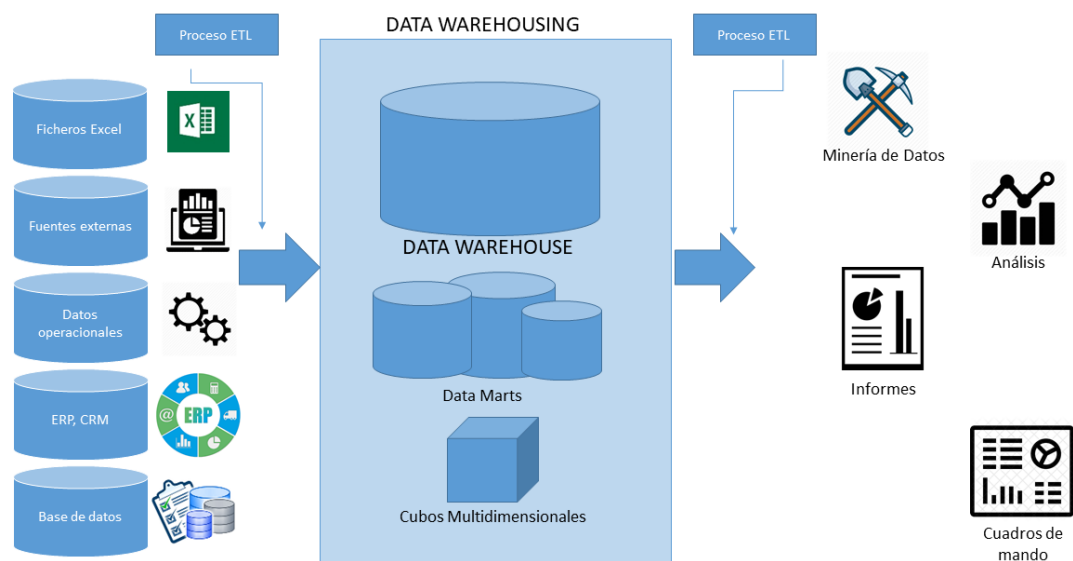


Ilustración 8. Arquitectura común de BI. Fuente: elaboración propia

Tal y como podemos observar en la ilustración, existen tres etapas ('stages') principales de los datos, promovidas por 2 transformaciones.

1. **Fuentes de datos.** Son todos aquellos **datos estructurados**, sin analizar y sin información conocida de ellos.
2. **Data warehousing.** Son los 'almacenes de datos' personales de la entidad. Las diferentes fuentes de datos son unificadas en un solo origen a través del proceso ETL. El proceso de unificación se debe hacer periódicamente, separando dentro del DW los datos utilizados frecuentemente de aquellos que serán usados para fines de ayuda a toma de decisiones. El Datawarehousing se puede realizar de manera física (servidores, sistemas de almacenamiento) o en la nube (servidores o sistemas de almacenamiento externos, propiedad de terceros).
3. **Reporting.** Es el añadido de valor a los datos poseídos. Los recursos obtenidos son resultado directo del uso de herramientas de BI para la obtención de información y posteriormente, conocimiento.



Observando la ilustración proporcionada, debemos explicar en cada uno de los *stages* los elementos que existen.

2.2.1 Ficheros Excel

Abundan en la mayoría de organizaciones de pequeño y mediano tamaño. Suelen propiciar la duplicación o triplicación de información, errores de transcripción, falta de actualización, pérdidas de información, diferentes formatos por departamentos, etc. Suelen ser poco recomendables.

2.2.2 Ficheros externos

Son datos ofrecidos por clientes o proveedores, a través de plataformas de colaboración conjuntas o intercambio de información.

2.2.3 Datos operacionales

Datos extraídos de procesos productivos propios, de los sensores de las máquinas y de otras fuentes internas.

2.2.4 ERP, CRM

ERP (Enterprise Resource Planning). Son software que aúnan datos sobre los diferentes procesos internos como producción, logística, inventario, compras, ventas e incluso de recursos humanos. Automatizan gran parte de los procesos operativos de la empresa. Generalmente tienen una interfaz modular en la que es posible la habilitación e inhabilitación de diferentes adaptándose a las necesidades de cada negocio.

CRM (Customer Relationship Management). Son sistemas orientados a la gestión de la relación con los clientes: gestión comercial, marketing y servicio postventa. Pretenden mejorar la relación con el cliente a través de personalización de contenido, entendiendo sus necesidades y anticipándose a ellas.

Ambos sistemas son sistemas para agrupar datos de fuentes más primitivas. La gran mayoría poseen herramientas de análisis y obtención de información.

2.2.5 Bases de datos

Las bases de datos son creadas para suplir unas necesidades de información en una organización. Sin embargo, puede haber más de una base de datos o alguna que no esté integrada.

2.2.6 Data Warehouse

El término apareció por primera vez a finales de los años 80 por parte de los investigadores de IBM Barry Devlin y Paul Murphy. Sin embargo, el impulsor del término se considera que es William H. Inmon el cual lo definió por una serie de características (H. Inmon, 1992):



- **Orientado a temas:** los datos se agrupan por temas para facilitar su acceso y comprensión. Es decir, organizamos los elementos de datos relativos a un ámbito, evento u objeto queden relacionados entre sí. Ejemplo: agrupar todos los datos de proveedores en una única tabla.
- **Variante en el tiempo:** A diferencia de las DB transaccionales, los DW presentan una perspectiva temporal mucho más amplia (meses o años), por lo que cada uno de los datos deben estar identificados con un periodo de tiempo específico.
- **No volátil:** El DW constituye un sistema separado de los datos operativos. Los datos introducidos en el almacén no se pueden actualizar, por lo que en un ambiente de DW sólo podemos considerar la carga de los datos y la consulta de los datos. Es un mero silo de datos.
- **Integrado.** Se forma a partir de múltiples y diferentes fuentes de datos. A través de técnicas de depurado, integración y transformación, éstos son incluidos en el DW. Los datos contenidos engloban todos los datos operacionales de la organización, de manera que estos se encuentren en una estructura consistente, con diferentes niveles de detalle para adaptarse a las necesidades de los usuarios.

En definitiva, DW es un almacén de datos corporativos en el que se unifican y se depuran los datos provenientes de diferentes fuentes (datos “vírgenes”) para luego ser usados por software DSS (Decision Support Systems) o EIS (Executive Information System) para su agrupación, exploración y análisis. Es decir, la funcionalidad principal de un DW es la **estandarización de los datos corporativos**. La implantación de un DW es el primer gran e importante paso a la hora de comenzar la implementación de una herramienta de BI.

Los DW también permiten la introducción de metadatos, es decir, datos sobre los datos, lo que facilita la automatización e introducción de datos en los sistemas de almacenamiento. En la actualidad, estos DW pueden ser físicos mediante servidores corporativos o en la nube.

Por otra parte, el aumento de la cantidad y variabilidad de datos ha provocado la aparición de los Data Lakes (a partir de ahora designados como [DL]), el cual almacena enormes cantidades de datos estructurados, semi-estructurados y no estructurados en su formato original, sin transformar ni estandarizar como sucede en los DW. Ofrece una gran cantidad de datos para aumentar las capacidades de obtención de información y su fiabilidad, a través de herramientas destinadas a Big Data.

2.2.8 Data Mart

Es un **subconjunto del DW**, orientado principalmente a las consultas antes que el almacenamiento en sí. A pesar de lo que se pueda creer, su implementación es igual de compleja que el de un DW corporativo, ya que se necesita de la misma manera un modelo relacional y una integración de las fuentes de datos. En términos generales son creados para la satisfacción de necesidades especiales de cada departamento. Sin embargo existen algunas otras razones para su creación:

- **Prestaciones:** mejorar el rendimiento de las consultas con un número menor de volumen de datos total.

- Seguridad: reagrupar, separar o restringir accesos a datos sensibles o de carácter confidencial.
- Conveniencia: pasar por alto las restricciones o permisos necesarios para operar en un DW centralizado corporativo.
- Pruebas: aplicación en pequeña escala para demostrar la capacidad de instalar un DW general.
- Otras. Estrategia, política, diseño...

Generalmente, su creación, mantenimiento y control está subordinada al departamento. Debe determinarse y encontrarse la estructura óptima de datos para analizar la información de la manera adecuada.

Las fuentes de datos de un DM normalmente es el propio DW, pero es posible que tenga fuentes de datos del propio departamento.

2.2.9 Cubos multidimensionales (cubos OLAP).

El siguiente paso tras los Data Marts son los cubos OLAP. Los cubos OLAP son estructuras jerárquicas que permiten un análisis aún más rápido a través de las diferentes dimensiones que definen a los datos, como se observa en la Ilustración 9. Los datos analizados pueden provenir de diferentes Data Marts o de otras fuentes. La ventaja principal de los cubos OLAP es la capacidad de unificar e integrar únicamente las dimensiones que deseemos, de los datos que deseemos. Permite profundizar en el análisis de éstos, delimitando el grado de detalle de las dimensiones (por ejemplo: de la dimensión Fecha, poder elegir Día, Mes, Trimestre o Año).

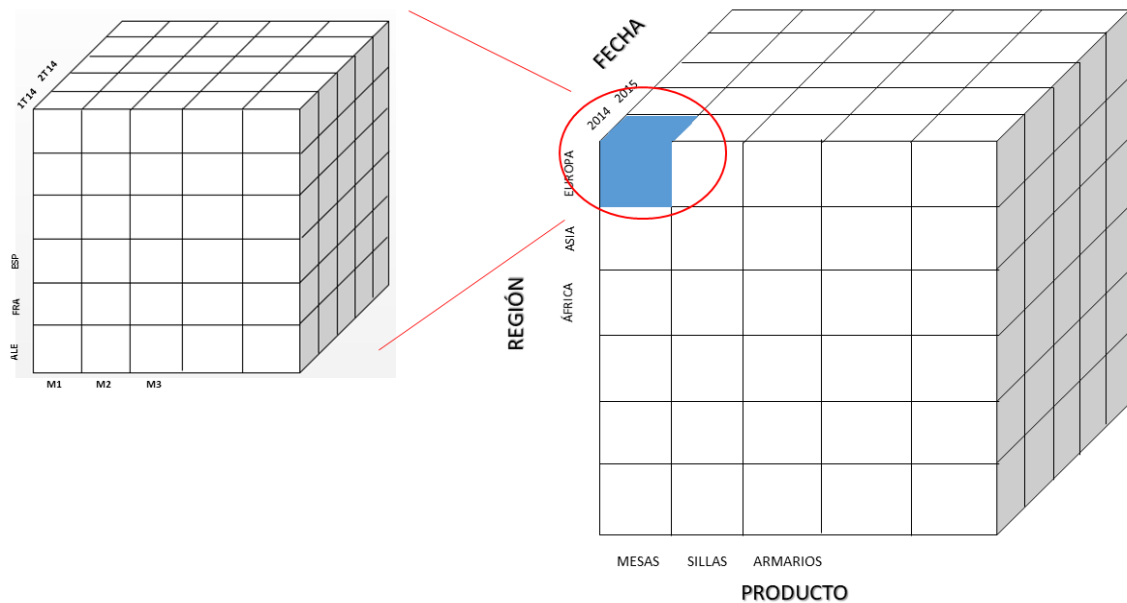


Ilustración 9. Ejemplo de Cubo OLAP. Fuente: elaboración propia

Si bien es cierto que en la ilustración se ha representado lo que generalmente sucede y lo que es gráficamente representable (la existencia de 3 dimensiones) los sistemas



OLAP pueden contener más de 3 dimensiones, y es por ello que también se pueden nombrar como 'hipercubos'.

Sin embargo, otra ventaja muy importante de los cubos OLAP es la rápida capacidad de respuesta. Los cubos OLAP tienen precalculados los ejes y las dimensiones, de manera que cualquiera de las **operaciones posibles en el cubo** no requieren de capacidad de computación. Estas operaciones son:

- **Roll-up:** consiste en la agrupación de una o varias dimensiones. En la ilustración podemos considerarlo como el proceso inicial de agrupar los países de las ventas en REGIONES. Supone 'subir en la jerarquía'.
- **Drill-down:** es el proceso contrario de Roll-up. Consiste en la desagregación de una o varias dimensiones en partes más pequeñas, es decir 'bajar en la jerarquía'. En la ilustración supondría la transformación de FECHA por Años a FECHA por Trimestres.
- **Slice:** es la selección y filtrado de los datos al seleccionar una de los valores de la dimensión, creando un nuevo sub-cubo (en este caso, un plano). Por ejemplo, en la ilustración, sería la cara frontal del cubo al seleccionar FECHA=2014.
- **Dice:** es una operación parecida a Slice. Sin embargo, en ella seleccionas dos o más valores de una o varias dimensiones, formando un sub-cubo más pequeño. Por ejemplo, podríamos elegir solo REGIÓN= EUROPA o ASIA para FECHA= 2014 o 2015 con la dimensión PRODUCTO sin modificar.
- **Pivot:** es la sencilla operación de recolocar los ejes, de manera que las dimensiones ocupen diferentes posiciones para una interpretación más sencilla.

2.2.10 Reporting

La última columna principal que hemos denominado Reporting es, en términos generales, donde se ubica la parte más importante del BI. Es en este 'stage' es donde radican las características principales y diferenciadoras del BI y es en lo que nos centraremos durante los siguientes apartados.

2.2.11 Proceso ETL (Extraction, Transformation, Load)

El paso de un stage a otro siempre va acompañado con un proceso de Extracción, Transformación y Carga. Es un proceso clave de la arquitectura del BI, ya que de él depende, la carga de los datos al DW o a las herramientas de análisis de una manera correcta e íntegra. Vemos qué son cada uno de los términos:

EXTRACCIÓN

Como su nombre indica, se deben extraer los datos desde el sistema de origen. Tanto en el primer proceso ETL como en el segundo de nuestra arquitectura es fácil observar que la extracción proviene de varias fuentes de datos. En el primero, se extraen datos "vírgenes" del proceso productivo, fuentes externas o sistemas OLTP. En el segundo, aunque sean datos propios contenidos en nuestro DW, en algunos casos también han de ser extraídos para ser utilizados por las herramientas de BI. En cualquiera de los datos, lo común es que esta extracción se produzca desde una base de datos (relacional o multidimensional) o ficheros planos.



Además, una práctica común en los procesos de extracción es la comprobación del estado de los datos, es decir, comprobar si son válidos (ya sean los datos en sí o el archivo, el cual puede estar dañado por ejemplo) y en caso contrario rechazarlos.

Este proceso ha de causar un impacto mínimo en la fuente de origen, de manera que no afecte ni su velocidad, ni su integridad ni su rendimiento.

La periodicidad de este proceso puede estar automatizado o realizado manualmente cuando se desee. A continuación resumiremos cada una de las fases del ETL.²

TRANSFORMACIÓN

El proceso de transformación se basa en convertir los datos en datos válidos para el destino, es decir en datos compatibles y congruentes. Aplica una serie de reglas o requisitos preestablecidos para convertirlos en datos que serán cargados. Independientemente del proceso de transformación llevado a cabo (cambios grandes o pequeños) es imprescindible mantener su integridad. Estas transformaciones pueden ser complicadas o simples, pero sin duda, son muy diferentes. Para resumir, podemos decir que las más importantes son:

- **Codificación.**
Consiste en unificar la manera de determinar los datos que designan un estado. Por ejemplo, una luz puede determinar su estado mediante binario (0,1), texto (apagada, encendida) o convencionalismos (off, on). Es donde se convierte en esencial la codificación.
- **Unidades de los atributos**
Las diferentes fuentes de origen pueden utilizar distintas unidades para referirse a un mismo atributo, y todo ello en un mismo negocio. Debemos estandarizar mediante el proceso de transformación a una única unidad que mejor se adapte a los objetivos de análisis (generalmente Sistema Internacional).
- **Normalización de nombramiento.**
Hace referencia a un concepto muy parecido que la codificación. En este caso se refiere a campos de texto que han de ser unificados para un uso y entendimiento óptimo por parte de los analistas. Por ejemplo, si en diferentes fuentes de origen hacemos referencia al número del cliente y en las diferentes fuentes está diferenciado como: id_cliente, num_cliente, numerocliente.... Se debe unificar bajo un único nombre.
- **Duplicación de datos por diferentes fuentes**
Un mismo dato puede provenir de distintas fuentes. Se debe elegir aquella fuente que se considere más fiable y mejor implementada para que dicho dato sea utilizado.

² Adaptado de Bernabéu (2010)



Además de estas operaciones típicas, la transformación de datos se ocupa de una actividad muy importante en el proceso ETL: la **limpieza de datos**.

- **Limpieza de datos.**

Su objetivo es tomar acciones sobre los datos erróneos, inconsistentes o irrelevantes.

Contra **valores anómalos**, lo más habitual es:

- ✓ No tomar acciones (ignorar)
- ✓ Eliminar o filtrar columna/fila.
- ✓ Cambiar el valor
- ✓ Discretización del intervalo de valores (por ejemplo: bajo, medio, alto)

Contra **valores vacíos (puede darse el caso de la inexistencia de datos)**:

- ✓ No tomar acciones (ignorar)
- ✓ Eliminar o filtrar columna/fila
- ✓ Cambiar valor
- ✓ Esperar hasta que dicho valor esté disponible

CARGA

Los datos de la fase de transformación son cargados en el destino. El proceso de carga se ocupa fundamentalmente de dos actividades: carga y mantenimiento de estructura.

- La carga se refiere tanto a la primera carga de datos como a las actualizaciones periódicas pertinentes. La primera carga suele ocupar un tiempo mucho mayor que las actualizaciones periódicas o automáticas (pequeña cantidad de datos cargados).

La forma más común de carga de datos es la **acumulación simple**. Consiste en realizar un resumen de todos los datos (transacciones) de un intervalo de tiempo y mover el resultado como una única transacción hacia el destino.

Por otra parte, también puede usarse una carga **Rolling**. Está destinada en casos en los que se quieren establecer diferentes **niveles** en **jerarquías**. Los datos se consolidan en agrupaciones en función de tiempo u otra clasificación que se quiera y se cargan bajo ese 'título' para su identificación posterior.

- Por otra parte, el proceso de carga ha de asegurar que se mantiene la estructura del DW definida al cargarlos: claves primarias, claves foráneas, relaciones entre dimensiones, etc.

2.3 Características del Business Intelligence

El Business Intelligence como concepto, reúne muchas características comunes a todas las herramientas de BI. Sin embargo, una es la que prevalece sobre las demás: la **ayuda en la toma de las decisiones**. Así pues, las herramientas de BI son incluidas comúnmente en los DSS (Decision Support Systems).

El BI introduce en la empresa una ventaja competitiva respecto del mercado. Pretende obtener información tangible a partir de grandes cantidades de datos que en un primer momento, no tienen valor real. El **análisis de nuestros datos** es lo que **genera** información, y a su vez, este se convierte en **conocimiento**, lo que nos permite tomar mejores decisiones.



Hasta cierto momento de la historia, la mayoría de las decisiones corporativas se tomaban en base a la experiencia pasada. El BI pretende eliminar en todo momento este factor y permitir a los directivos tomar decisiones en base a los datos, de una manera objetiva y eficaz. En cierta parte, pretende que cualquier persona de la organización pueda llegar a la misma decisión o conclusión sin tener la misma experiencia que un directivo. Podemos entender esta ayuda en la toma de decisiones como una **eliminación parcial de los riesgos** que supone tomar una decisión corporativa.

Además de esta característica fundamental, existen otras alrededor de ella ciertamente relacionadas:

- **Sistema completamente integrado en la actividad de la organización.** El sistema de BI es un componente más de los procesos del negocio. Su disponibilidad ha de ser completa para cualquier usuario (aunque es posible la restricción de acceso a datos sensibles o confidenciales) y tener responsables que se ocupen de su gestión y mantenimiento.
- **Compatibilidad completa de análisis de datos estructurados.** El BI debe ser capaz de analizar cualquier tipo de datos estructurados, independientemente de su organización las bases de datos, ya sean relacionales o multidimensionales. Además, en los últimos años también están apareciendo compatibilidades parciales con datos no estructurados.
- **Medición de desempeños.** El uso de BI convierte el negocio en uno perfectamente medible y cuantificable mediante la determinación de medidas que se asocien a departamentos, procesos, rendimientos, objetivos, etc.
- **Análisis de causalidad.** El BI también ha de ser capaz, como explicamos en su momento del OLAP, de permitir al usuario la detección de causas tanto de hechos que hayan perjudicado al negocio como los que lo han beneficiado, tanto para evitar la reaparición como para maximizar su aparición.
- **Carácter predictivo y anticipativo.** La adopción de un sistema de BI pretende, a la vez que analizar causas, predecir lo que sucederá en función de datos históricos. Las más modernas herramientas del mercado realizan esto mediante complejos motores analíticos y algoritmos que pretenden recrear las tendencias de los datos mediante modelos matemáticos precisos. En última instancia, este carácter predictivo ha de convertirse en uno anticipativo, detectando posibles riesgos y evitándolos.

2.4 Ventajas del Business Intelligence

Se pueden enumerar de manera sistemática y lógica ciertos beneficios de implementar Bi en los negocios. Por ejemplo (Conesa et al, 2010):

- Creación del círculo virtuoso de la información (Ilustración 10). Es el círculo que definen los datos transformados en información, los cuales generan un conocimiento para la toma de buenas decisiones, traducidos en unos resultados conformados por datos.
- Integración de la información de manera histórica, consolidada y de calidad.
- Determinar, establecer y seguir los KPI (Key Performance Indicator) y los KGI (Key Goal Indicator). La diferencia básica entre ellos es, como su nombre

indica, que el KPI simplemente es una métrica de rendimiento y en cambio, el KGI es una métrica sobre el objetivo propuesto por la organización.

- Existencia de datos e información actualizada de la organización, tanto jerarquizada como detallada (tal y como vimos en las bases de datos).
- Nivelar y orientar hacia un mismo objetivo al departamento encargado del sistema de BI y a la organización como entidad.
- Mejora de competitividad de la empresa respecto al mercado debido a las nuevas capacidades adquiridas por el sistema de BI: agilidad en toma de decisión, disposición de información actualizada...

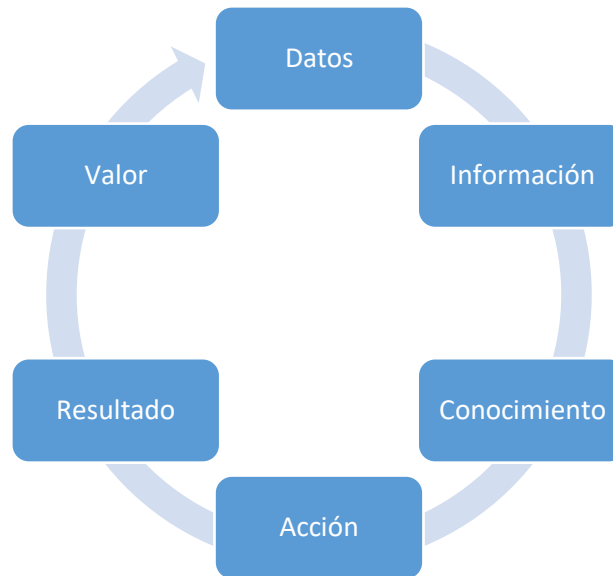


Ilustración 10 Círculo Virtuoso de la Información. Fuente: elaboración propia

Sin embargo, ciertos autores han contemplado la posibilidad de clasificar de una manera sistemática los beneficios aportados por el BI en beneficios tangibles, intangibles y estratégicos (Cano, 2007). A continuación se muestra cada uno de ellos, y además, se ejemplifican.

- **Beneficios tangibles.** Son aquellos que repercuten directamente sobre el rendimiento y rentabilidad del negocio.
 - Generación de ingresos.
 - ✓ Mejorar la captación de clientes y reducir la tasa de abandono de los ya fidelizados, incrementar su fidelidad...
 - ✓ Aumentar ingresos por aumento de ventas (análisis temporal de mejor momento para las ventas).
 - ✓ Aumentar resultados a través de más compras de actuales clientes.
 - ✓ Evitar pérdidas de ventas debido a competidores.
 - ✓ Incrementar rentabilidad gracias a información detallada de clientes y productos.
 - ✓ Geoposicionamiento de ventas para conocer su distribución y cantidad de ventas por lugar.
 - ✓ Aumentar la cuota de mercado.
 - ✓ Reducir tiempos de lanzamiento de productos (diseño, prototipado, producción y comercialización).

- ✓ Mejorar sistemas de recolección de datos corporativos.
- ✓ Mejorar y facilitar orientación de la estrategia.
- ✓ Permitir un acceso 'self-service' a los usuarios del sistema de información.
- ✓ Cuantificación de la efectividad de las campañas de ventas.
- ✓ Optimizar marketing.
- ✓ Crear nuevas oportunidades.
- Reducción de costes.
 - ✓ Identificación de proveedores más competitivos, aumentar capacidad de negociación con los actuales, gestionar descuentos por compras, asignar y automatizar compras...
 - ✓ Reducir mano de obra necesaria para procesos productivos o reasignamiento para aumentar eficiencia.
 - ✓ Incrementar productividad y eliminar ineficiencias del proceso productivo.
 - ✓ Mejorar control de costes y disminuir gastos.
 - ✓ Reducir reclamaciones.
 - ✓ Evitar excesos de producción y orientar a producción JIT (Just In Time) para reducir costes de inventario.
 - ✓ Acortar los periodos entre informes.
 - ✓ Análisis de fallos y reparaciones para detectar y solventar los más frecuentes.
 - ✓ Seguimiento de procesos de resolución y reparación.
 - ✓ Reducir devoluciones de productos.
 - ✓ Analizar y optimizar la productividad de operarios.
 - ✓ Abaratar costes de campañas de marketing.
- **Beneficios intangibles.** Son aquellos que mejoran nuestra posición competitiva indirectamente.
 - ✓ Mejorar atención y satisfacción al cliente.
 - ✓ Mejorar el acceso a los datos y su análisis.
 - ✓ Periodos más cortos de actualización de los datos.
 - ✓ Aumentar precisión de los datos.
 - ✓ Conseguir ventaja competitiva.
 - ✓ Mayor integración de la información.
- **Beneficios estratégicos.**
 - ✓ Mayor habilidad para analizar estrategias de precios.
 - ✓ Mejorar toma de decisiones.
 - ✓ Mayor visibilidad de la gestión.
 - ✓ Dar soporte a las estrategias.
 - ✓ Aumentar valor de mercado.

2.5 Estrategias del Business Intelligence

Hasta ahora, hemos repasado una serie de puntos básicos alrededor del concepto del Business Intelligence, desde su arquitectura, pasando por características, hasta sus beneficios. Sin embargo, no hemos entrado en detalle en la realidad de las implantaciones de sistemas de BI. **En este apartado**, desvelaremos **cuándo** es conveniente **implantarlo** o las **estrategias a seguir** en el proceso (teniendo en cuenta nuestro grado de madurez empresarial).



2.5.1 Necesidad de un sistema de Business Intelligence

Algunas situaciones pueden evidenciar la necesidad de un sistema de BI. A lo largo de este capítulo se ha repetido innumerables veces que el BI ha de habilitar la toma de decisiones basadas en información. Por ello, podemos intuir que una razón de su implementación es que en la organización se siguen **tomando las decisiones mediante intuición**. Por otra parte también se ha comentado la ineficiencia del **almacenamiento de datos en Excel** (provocando silos de información, en la mayoría de casos, duplicada, no integrada, errada y poco sincronizada), por lo que también se puede tomar como una razón para utilizar BI. En términos genéricos, podemos intuir que una razón es la **falta de datos de calidad**.

Estas dos premisas son fundamentales, pero podemos deducir otras tales como la necesidad de cruzar información de forma ágil entre departamentos, es decir, **tener a mano datos actualizados de diferentes departamentos** en cualquier momento y lugar que se requiera en la organización. También es común hacer de un sistema de BI necesario cuando nuestras campañas de publicidad y marketing no ofrecen los resultados esperados en base a los datos que se tenían en un principio, es decir, necesitamos un sistema que proporcione **información y predicciones más fiables**.

Estos casos se encuentran en organizaciones con cierta cultura de digitalización de datos. Sin embargo, podemos encontrar motivos en extremos opuestos. Por un lado, la **necesidad de automatizar la extracción y distribución de datos**, en organizaciones que con poco grado de digitalización o, por otro lado, la necesidad de nuevos métodos para **analizar grandes cantidades de datos** (organizaciones que ya poseían sistemas de análisis pero han alcanzado la masa crítica).

2.5.2 Desarrollo de la estrategia

Las implementaciones de BI en las organizaciones no suelen ser proyectos sencillos y rápidos. La estrategia ha de aunar y coordinar tanto el proyecto en sí como sus objetivos. Satisfacer de manera óptima los factores que han delimitado la necesidad y posterior comienzo de la implementación. El desarrollo de la estrategia es vital en cualquier proyecto de BI. En primer lugar, antes de tratar de delimitar o trazar un camino general para el establecimiento de la estrategia, la opción más fácil es determinar mediante ciertos **indicadores que no existe una estrategia**. Entre otros:

- Resto de empleados consideran que el departamento encargado del BI como el origen de los problemas.
- La directiva considera el BI como un coste más al BI.
- Incapacidad de medir el retorno de la inversión (ROI) del proyecto de BI.
- Se considera como la misma estrategia la del DW que la del BI.
- No existe responsable de la estrategia (o no está bien delimitado o asignado).
- Existencia de diferentes sistemas de BI o parecidos en diferentes departamentos.
- No hay plan de contratación, formación y retención de equipo destinado al BI.
- No existen planes de formación al usuario para el sistema de BI.
- Desconfianza por parte de usuarios de la integridad de los datos almacenados en el DW.

Ahora que se han delimitado ciertos indicadores, podemos indicar otros que en términos generales, remedien los otros y que además constituyan una hoja de ruta como estrategia de BI:

- El primer paso y fundamental para el desarrollo de la estrategia es la **creación del BICC (Business Intelligence Competency Center)**.

El BICC, cuya traducción al español suena forzosa, es un equipo transversal a la organización compuesto por un equipo de variados perfiles cuya función principal es la de realizar un uso efectivo y eficiente del BI en toda la organización. Un BICC ha de garantizar tanto el uso efectivo del BI como la entrega adecuada y a tiempo de la información que requieren los usuarios. Entre otras funciones el BICC debe (Páez, 2010):

- ✓ Definir en gran medida, la estrategia del BI.
- ✓ Conseguir apoyos y fondos necesarios en la directiva.
- ✓ Establecer métodos y estándares.
- ✓ Desarrollar habilidades analíticas y formar a usuarios.

El equipo ha de estar formado por personas con tres habilidades: de negocio, analíticas e informáticas (especialistas IT, Information Technology), tal y como se muestra en la Ilustración 11. Es decir, debe aunar a personas del nivel directivo y personas involucradas en el aspecto más técnico del proceso de BI.

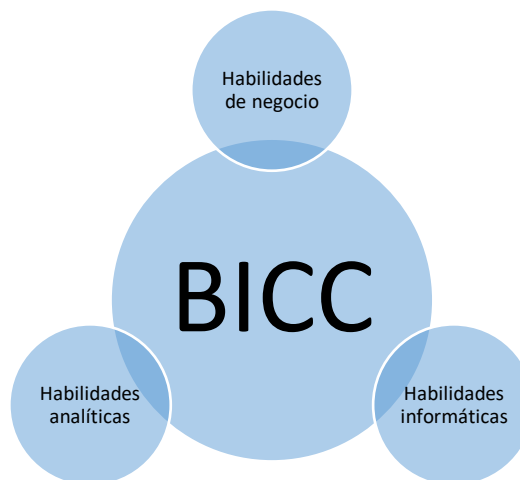


Ilustración 11. Habilidades del BICC. Fuente: elaboración propia

- **Habilidades de Negocio**
Debe contener integrantes que conozcan plenamente la organización, sus procesos internos y el mercado exterior. Deben ser capaces de definir métricas, transmitir la estrategia de la organización al BI, desarrollar planes de acción y tener conocimiento de los recursos y procesos.
- **Habilidades analíticas**
Se necesita de personal con amplios conocimientos en análisis de datos, estadística y algoritmos. Es común que este personal sean



estadísticos o con experiencia en modelación, análisis y optimización de procesos.

- **Habilidades informáticas**

Todos estos conocimientos avanzados en análisis han poder ser puestos en valor mediante un sistema o herramienta. Es donde se hace necesario al personal del departamento de informática (IT), los cuales han de ocuparse del mantenimiento, gestión de los datos y servir de apoyo a la implementación de los conocimientos de los estadísticos.

En caso de no poder crear un BICC, debemos alinear los objetivos del departamento de informática con los de la organización en la mayor medida de lo posible (Conesa et al, 2010).

- Identificar procesos/departamentos que necesitan aplicaciones específicas para extracción y gestión de datos para evitar su aislamiento.
- Incluir absolutamente todos los resultados de procesos analíticos en los procesos de negocios para ayudar en toma de decisiones.
- Revisar las herramientas o proceso de BI actuales en un contexto coste-riesgo/beneficios.
- Considerar inversiones en herramientas o procesos (novedosos o con potencial de ser muy útiles) específicos cuyo ROI esté en el plazo de un año.
- Revisar y analizar casos de éxito (o fracaso) en la industria a la que se pertenece para intentar “imitar” los pasos a seguir (o evitar).
- Analizar quién necesita de BI y poner especial atención en aquellos procesos/departamentos que realmente necesiten de BI en su actividad normal (y no aquellos que más lo pidan).

Los proyectos de BI suponen un proyecto de transformación digital, y por ello ha de suponer también un cambio en la cultura corporativa, lo cual en términos y monetarios y de recursos supone una **inversión muy alta**. Por tanto, incluso antes de tratar de realizar los anteriores puntos para tener una estrategia general, **debemos saber responder con certeza y seguridad a las siguientes cuestiones:**

1. Qué problemas o necesidades se desean resolver con BI.
2. En dichos problemas, qué preguntas se han de resolver y cuáles serán las medidas a tomar.
3. Cuál es el motivo de no poder responder a las preguntas en la actualidad.
4. Cuál es el impacto de no tener esta información (no saber responder esas preguntas).
5. Dónde se obtendrán los datos necesarios.
6. Cuál es el grado de alineación entre las fuentes de datos/información (producto, cliente, departamentos...).
7. Cuál es la diferencia entre los datos existentes actualmente y los necesarios para responder a las preguntas.
8. Cómo es la calidad de los datos.
9. Cuál es la cantidad de datos actual e histórica necesaria, además de los períodos de actualización.
10. Cuál es el período de actualización de las respuestas.



2.6 Modelos de madurez³

Durante el anterior apartado no se han establecido pautas fijas y específicas de cómo llevar a cabo una estrategia, sino más bien, se han dado pinceladas de cómo ha de desarrollarse ésta. Bien es cierto, que **el mejor punto de partida es calificar o establecer nuestro grado de madurez** en lo que se refiere a BI. Por ello, existen varios **modelos de madurez**: BIMM (Business Intelligence Maturity Model), Forrester BI Maturity Model, Business Information Maturity Model, Gartner's Maturity Model for BI and Performance Management y BI Maturity Hierarchy, entre otros.

2.6.1 Business Intelligence Maturity Model (BIMM)

Este modelo de madurez clasifica nuestra organización en función del grado de madurez de implantación de sistemas de BI. Es una manera simple y directa de identificar y clasificar. Se define por fases:

- **Fase 1. No existe BI.** Caracterizado por la utilización de sistemas transaccionales (OLTP) en ciertos sectores de la organización. Decisiones basadas en intuición y experiencia. No existen herramientas y los datos están dispersos e inaccesibles.
- **Fase 2. No existe BI pero los datos son accesibles.** No existe un proceso estandarizado para el uso de datos en toma de decisiones pero es posible el acceso a datos de calidad (y algunos usuarios pueden usarlos para tomar decisiones). Normalmente, soluciones basadas en Excel y reporting simple. Se intuye la existencia de mejores soluciones pero se desconoce el concepto de BI.
- **Fase 3. Aparición de procesos formales de toma de decisiones basada en datos.** Se caracteriza por la creación de un equipo que gestione los datos corporativos y genere información para fundamentar las decisiones. Utilización de datos "vírgenes", sin depuración ni modelación, directamente extraídos de OLTP. No existe data warehouse.
- **Fase 4. Aparición del DW.** El uso directo excesivo de datos desde OLTP hace necesario la existencia de un repositorio de datos, un data warehouse. Reporting sigue siendo personal.
- **Fase 5. Crecimiento del DW y formalización del reporting.** El acceso al DW permite que el reporting se expanda a toda la organización. Se comienza a entender el concepto de OLAP, pero pocos saben aprovecharlo.
- **Fase 6. Despliegue de OLAP.** Los usuarios requieren cada vez de respuestas a preguntas más sofisticadas, y es cuando aparece el OLAP. Las decisiones repercuten cada vez más en el negocio.
- **Fase 7. Business Intelligence se formaliza.** Necesidad creciente de nuevos procesos como la minería de datos, los CMI, CRM, Supply Chain Management, etc.

³ Adaptado del curso: "Introducción al Big Data para PYMES", impartido por la entidad BPMSAT s.l.



2.6.2 Forrester BI Maturity Model

Existen 4 niveles en dicho modelo (Maler et al, 2013).

- **Nivel 1. Static Intelligence.** Las fuentes de datos están predefinidas, son escasas y están obsoletas, al igual que los datos que contiene. El acceso a los datos está restringido a la propia organización.
- **Nivel 2. Reactive Intelligence.** Actualización de las fuentes de datos de forma periódica. Movimiento y compartición de datos unidireccional.
- **Nivel 3. Interactive Intelligence.** Multitud y variedad de fuentes de datos, algunas incluso en tiempo real. Los datos pueden ser compartidos a terceros, externos a la corporación. Se evoluciona a medida que se tienen más datos.
- **Nivel 4. Adaptive Intelligence.** Existencia de fuentes de datos colaborativas y actualizadas. El conocimiento adquirido proviene de respuestas a preguntas realizadas “just-in-time” mediante datos provenientes de diferentes socios estratégicos.

2.6.3 Business Information Maturity Model

El modelo tiene como objetivo aumentar la importancia del BI en la organización. Propuesto por N. Williams en 2003, define tres factores fundamentales para el éxito del BI: alineación y gestión, influencia y distribución. Está compuesto por tres niveles (Williams et al, 2003):

- **Fase 1.** Este primer escenario se puede entender como la anteposición del uso de los datos antes que su almacenamiento (el DW no es conocido ni usado). Los usuarios proporcionan al departamento de informática un “listado” de información que necesitan, por lo que el departamento centra sus esfuerzos en **qué** les gustaría recibir o qué es lo que quieren.
- **Fase 2.** La fase media supone una gran mejora respecto a la primera. La organización entiende que debe reinventar el rol de la información corporativa para obtener beneficio de sus inversiones. Se analiza la información que se necesita para cumplir objetivos o para desarrollar procesos. Es decir, en este caso el departamento de informática se centra en **por qué** la información es necesaria para el usuario.
- **Fase 3.** La etapa final pretende determinar qué procesos están en mejor posición de usar información e investigar qué procesos pueden ser mejorados cuando se tiene la información necesaria. Se pretende implantar procesos a nivel corporativo de toma de decisiones (en pos de procesos individuales) para optimizar el uso de la información.



2.6.4 Gartner's Maturity Model for BI and Performance Management

Gartner, una de las consultoras tecnológicas más relevantes en el ámbito del BI, propuso el siguiente modelo en 2008 para ayudar a las empresas a mejorar su BI y trazar una hora de ruta para estas mejoras. El modelo tiene 5 niveles (Ilustración 12):

- **Nivel 1. Unaware.** A modo de resumen, se llevan a cabo multitud de malas prácticas. Los datos, en términos generales, son inconsistentes y mal interpretados. Las demandas individuales de información dictan los cambios. Destaca por el uso de Excel y falta de herramientas de reporting, además de la inexistencia de métricas y la importancia de éstas. El departamento de informática es el que tiene en sus manos el presupuesto y la capacidad de gestionar la información y generar informes.
- **Nivel 2. Tactical.** El departamento de informática comienza a invertir en BI. Comienza el uso de métricas de rendimiento (KPI) a nivel de departamentos (algunos). Se usan programas comerciales, no específicos ni personalizados al negocio, su acceso tanto a ellas como a los datos es difícil. Los directivos se muestran reacios a la bondad de los resultados obtenidos.
- **Nivel 3. Focused.** El BI ha sido aplicado a parte de la organización y comienzan a vislumbrarse los primeros beneficios. Ciertos departamentos e incluso directivos apoyan plenamente el BI y aumentan los recursos disponibles, siendo éstos de diferentes fuentes. El BI evoluciona y se disponen de paneles de indicadores a nivel departamental, sin embargo, suele ocurrir la inconsistencia entre las métricas disponibles y los objetivos planteados (no son de utilidad o no concuerdan). **Aparece el BICC.**
- **Nivel 4. Strategic.** El BI es plenamente apoyado por el equipo directivo. El BI se extrapola incluso a socios y comienza a abarcar la gestión del rendimiento de los procesos clave del negocio. La información ya es accesible a nivel corporativo y el BICC tiene recursos suficientes para alcanzar objetivos. Se desarrolla un portfolio de estándares de TIC y una política de gobierno de la información. Los usuarios del BI son formados en procesamiento de datos y en el uso de la información de forma efectiva para transformaciones digitales estratégicas y tácticas.
- **Nivel 5. Pervasive.** El BI ya abarca la gestión del rendimiento a nivel corporativo, y tienen la confianza de toda la organización. Además, tanto BI como la monitorización del rendimiento son flexibles para adaptarse a los cambios en las demandas de información. El BI aparece en los procesos de negocio y se extiende a clientes, proveedores y socios. Los usuarios son formados y evaluados sobre sus conocimientos en gestión de la calidad de los datos, además de comenzar a poseer acceso a la información para creación de conocimiento para crear valor. El BICC toma un carácter proactivo y dinámico.

2.6.5 Business Intelligence Maturity Hierarchy

El modelo consiste en cuatro etapas, propuesto por R. Deng en 2007. Las etapas son:

- **Fase 1. Data.** Esta fase sirve como punto de partida para la introducción del DW y el BI. Se tiene capacidad de recoger, depurar, estandarizar y almacenar datos de diversas fuentes.
- **Fase 2. Information.** Los datos se comienzan a analizar y la organización es capaz de determinar indicadores de rendimientos (KPI) y emplearlos para CMI. Esta información de rendimiento está claramente definida y fácil de entender.
- **Fase 3. Knowledge.** Se usa el BI para identificar causas, buscar tendencias y relaciones y realizar análisis de escenarios. La forma avanzada de esta fase es crear un sistema experto que produzca nuevos conocimientos en función de experiencia pasada.
- **Fase 4. Wisdom.** La productividad de la organización en esta fase es mucho más alta, además proporcionar ventajas competitivas en periodos de entrega, objetivos empresariales, desarrollo del producto y calidad del servicio.

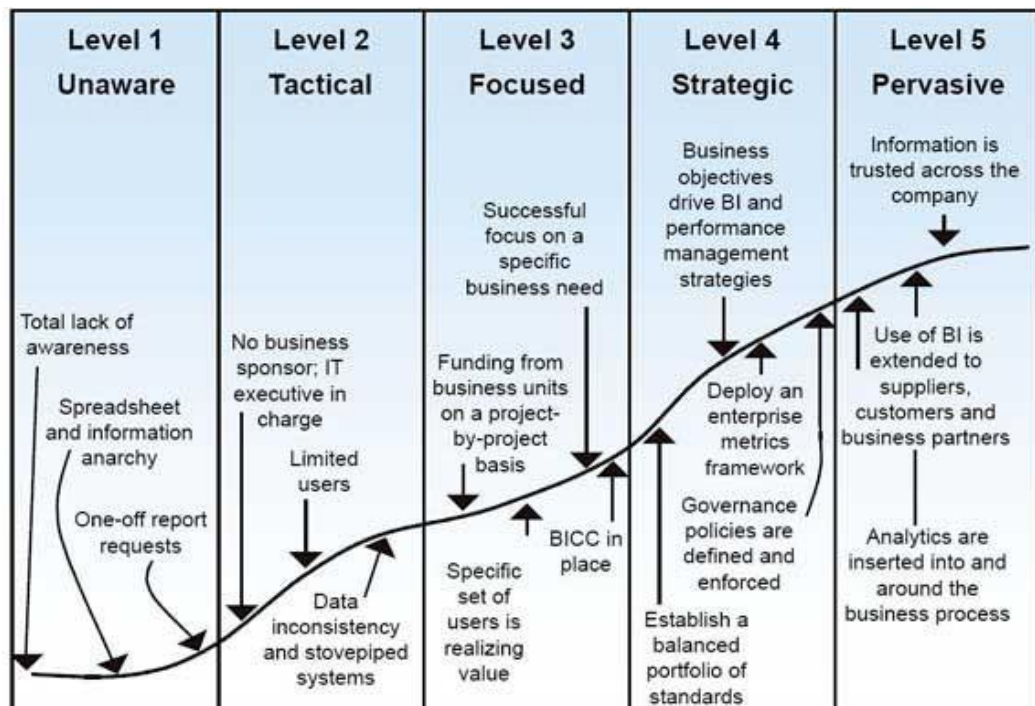


Ilustración 12. Gartner's Maturity Model for BI and Performance Management (recuperado de: <https://co.pinterest.com/pin/372039619207360824/>)



2.6.6 Errores comunes al implantar BI

Algunos de los errores comunes citados a continuación ya han sido identificados durante el capítulo como malas prácticas en los procesos de BI. Sin embargo, es muy probable encontrarse algunas de las siguientes situaciones:

1. Usar hojas de Excel como método de análisis (caos de Excel).
2. Considerar que un DW ya es poseer BI, cuando en realidad, un DW simplemente es un almacén de datos.
3. Carecer de una estrategia clara y definida de BI.

Por otra parte, se consideran otros errores más específicos:

4. No garantizar la calidad de los datos. Los datos son la fuente de información para cualquier informe o gráfico, por lo que se debe asegurar su calidad e integridad.
5. No evaluar las herramientas de BI disponibles en el mercado, cuando se debe elegir aquella que mejor se adapte o permita más flexibilidad de personalización al negocio.
6. No considerar el carácter dinámico del BI. El BI debe evolucionar en la misma medida que evolucionan los objetivos empresariales, tanto los datos que se proporcionan, como los resultados en los informes.
7. Subcontratar BI. Normalmente supondrá un ahorro en inversiones y coste, sin embargo, el BI ha de llevarse a cabo de manera interna.
8. Exigir informes, CMI y otros resultados sin tiempo ni presupuesto.
9. Diferentes sistemas de BI en función del departamento. Esto provoca un claro desajuste la orientación de la estrategia corporativa y la departamental. Se debe conseguir una visión única y conjunta.

2.7 Breve Introducción al Big Data

En la actualidad, la comunidad científica y los medios de comunicación nos bombardean con el término Big Data (al cual comenzaremos a denominar BD por comodidad a partir de ahora). Pretenden inculcarnos su importancia y nos describen a éste como el “petróleo del siglo XXI”, una fuente de increíbles beneficios. Este bombardeo es fácilmente demostrable mediante datos de la fuente a la que acude en primera instancia cualquier persona que desconoce o quiere profundizar en un tema: Google. Mediante su herramienta Google Trends, es posible ver la popularidad de búsquedas de términos en una escala temporal. Por curiosidad, compararemos en la Ilustración 13 la popularidad de la búsqueda de BD con BI.

Se observa la clara tendencia creciente del término BD y la progresiva decadencia en la popularidad de BI. Esto, se puede explicar fundamentándonos en que poco a poco, el término BD se confunde con BI como una generalización de éste.

● big data
Término de búsqueda● business intelligence
Término de búsqueda

+ Añadir comparación

Todo el mundo ▼ 2004 - hoy ▼ Todas las categorías ▼ Búsqueda web ▼

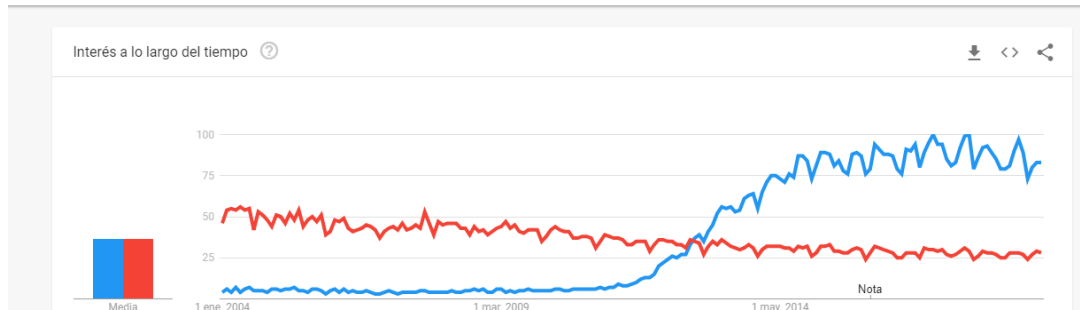


Ilustración 13. Popularidad de Big Data y Business Intelligence (recuperado de: <https://trends.google.es/trends/explore?date=all&geo=ES&q=big%20data,business%20intelligence>, visualizado el 23/05/19)

La alusión más temprana conocida al BD, aunque no se usara dicho término explícitamente, es de 2001. En ella, se describían los retos y oportunidades que se presentaban debido al aumento de los datos mediante el **modelo de las 3 V**: aumento del **Volumen**, aumento de la **Velocidad** y aumento de la **Variedad** de los datos (Laney D., 2001).

Aunque en esta publicación no se hiciera referencia directa al término de BD, el modelo de las 3 V se ha extendido a definiciones formales de BD posteriores. Entonces, podemos definir BD como activos de información caracterizados por un volumen, velocidad y variedad tan altos que requieren una tecnología específica y métodos analíticos para su transformación en valor (De Mauro, Greco et al, 2016).

Recientemente, ciertos autores reconocidos en el ámbito del BD han promulgado la adición de 4 V a este modelo que complementan a las 3 anteriores, convirtiéndose en el **modelo de las 7 V**. Se consideran:

- **Veracidad**: incertidumbre y grado de fiabilidad de los datos.
- **Viabilidad**: capacidad de ser usados para crear valor para un individuo u organización.
- **Visualización**: modo en que los datos son presentados.
- **Valor**: capacidad de los datos que son convertidos en información para proporcionar ayuda en las tomas de decisiones.

Debemos considerar que en el BD, los activos de información usados pueden ser de tres tipos.

1. **Estructurados**. Es el tipo de datos al que las personas y empresas están acostumbrados. Es información ordenada (normalmente por filas y columnas con títulos) bajo un formato definido en la que conocemos su significado. Es la contenida en bases de datos comunes. Además, son fáciles de procesar por herramientas específicas.
 - a. **Creados** (Rayo, 2016). Son datos que son obtenidos activamente, es decir, se ha de requerir a una organización o usuario que los



- proporcione de manera consciente a través de mecanismos de captura y análisis. También pueden ser semi-estructurados. Ejemplos: encuestas, registros en páginas, programas de fidelización...
- b. **Provocados.** Son datos obtenidos de forma pasiva, es decir, no se le ha exigido dicha información aunque ha sido proporcionada. Pueden ser semi-estructurados. Ejemplos: valoraciones de lugares, reviews de productos, etc.
 - c. **Transaccionales.** Datos obtenidos de un cliente o vendedor al realizarse una transacción. Método muy popular para obtener información por parte de empresas.
 - d. **Compilados.** Datos recopilados previamente por empresas o instituciones que pueden ser comercializados a terceros. Ejemplos: censos, volúmenes de compra, datos estatales...
 - e. **Experimentales.** Caso híbrido entre datos creados y transaccionales. Son datos obtenidos al recopilar la respuesta de usuarios ante estímulos (experimentación) o datos obtenidos mediante simulaciones. Pueden ser semiestructurados.
2. **No estructurados.** Es información que no puede ser almacenada bajo un modelo relacional de datos predefinido ni tiene ninguna estructura interna identificable. Generalmente son datos en forma de texto, binarios o de comportamientos de usuarios que son difícilmente analizables o clasificables (sin metadatos). Es tal su importancia que se considera que más de la mitad de los datos de una organización son datos no estructurados y más del 80% de la información relevante para un negocio proviene de fuentes no estructuradas.
- a. **Capturados.** Datos recogidos de forma pasiva sobre la conducta de usuarios y máquinas en la red, sin ser ellos conscientes de la recogida de dicha información.
 - b. **Generados por el usuario.** Datos que han sido generados de manera consciente, ya sean textos, como publicaciones de fotos y vídeos, como páginas web...
3. **Multi-estructurados o híbridos** (Calvo, 2017). A mitad de camino entre estructurados y no estructurados, son aquellos que no tienen un significado y sentido fácilmente comprendido por el usuario, necesitando de complejas reglas para ello. No tienen una estructura rígida ni pueden ser ordenables en filas y columnas.

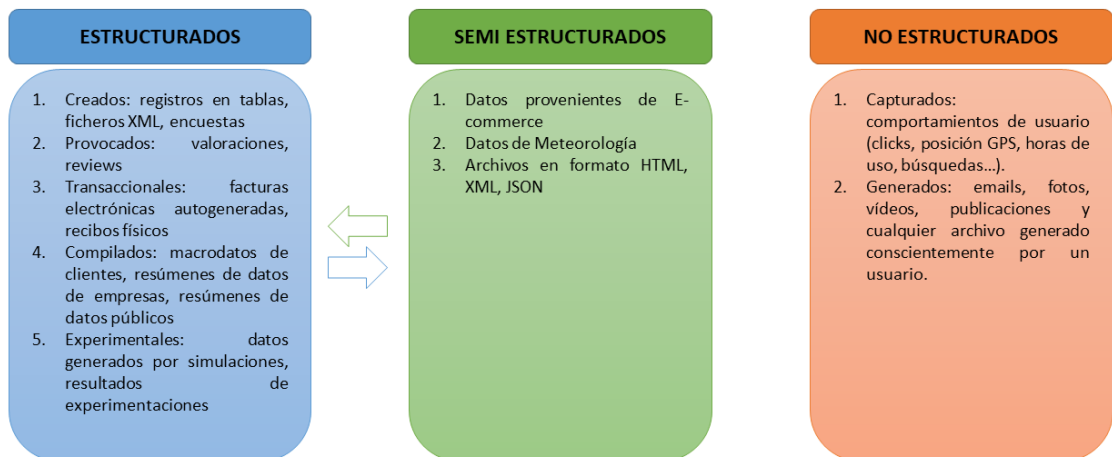


Ilustración 14 Cuadro resumen de tipos de datos. Fuente: elaboración propia

La principal y más rompedora característica del BD es permitir a individuos y organizaciones la capacidad de responder a preguntas que ni siquiera se habían planteado o sabían de la existencia de una respuesta. El uso adecuado de una gran cantidad de datos permite identificar, solucionar y prevenir problemas. Es decir, proporciona un nuevo punto de referencia para la visión global.

Para empresas, el BD es un gran avance en aspectos como la eficiencia, la relación con clientes, reducción de riesgos o búsqueda de nuevas oportunidades. Convierte a la empresa tradicional en una más eficiente, con más ingresos y con clientes satisfechos.

Veamos estas y otras ventajas que proporciona el BD en profundidad, comentando además aplicaciones reales de empresas nacionales.

- **Mejorar la producción y reducir costes.** Existe un número creciente de empresas que comienzan a dar uso al BD para aumentar su competitividad. Tanto en el proceso de producción como en el de almacenamiento de datos se pueden reducir costes. Por una parte, empresas como Confectionary Holdings o Mesbook, que instalan sensores en sus plantas para optimizar sus procesos y mejorar la toma de decisiones. Por otra parte, el almacenamiento de datos y su posterior análisis es cada vez más asequible gracias a almacenamientos en nube.
- **Mejorar la captación, relación y satisfacción del cliente.** Conocer a nuestros clientes es esencial para cualquier empresa. El BD permite a la organización conocer más detalladamente las necesidades de los consumidores, en especial el qué y el cuándo, lo que facilita la captación de nuevos clientes. Por otra parte, también ayuda en el entendimiento de los hábitos de compra permitiendo generar experiencias y ofertas personalizadas para su fidelización y satisfacción.

Por ejemplo, la cadena de supermercados Kruger ha usado BD para personalizar las campañas de marketing obteniendo una tasa de retorno del 70%.



- **Nuevos productos y servicios.** La analítica de los datos permite a los negocios dar al consumidor lo que quiere o considera que debe haber. Zara y Desigual, unas de las grandes del sector textil, utilizan el análisis de datos para el diseño de los productos para los clientes que consideran potenciales o ya fidelizados mediante observadores y datos de venta, visualizando tendencias globales y locales.
- **Agilizar y mejorar la toma de decisiones.** La aparición de herramientas cada vez más potentes con tecnología de análisis en memoria permite analizar datos en tiempo real, obtener información rápidamente y facilitar la toma de decisiones, basadas en datos y no en intuición o experiencia.
- **Mejorar la competitividad de industrias tradicionales.** Sectores como el agrícola, aunque se consideran poco digitalizados, también pueden hacer uso del BD. Empresas como Bodega Pago de Carraovejas usa analítica de datos para optimizar factores como el agua regada o la cantidad de fertilizantes, además de darle uso preventivo, tal y como usan los datos meteorológicos para evitar la pérdida de cosechas.
- **Mejorar la competitividad de las PYMES.** Cualquier empresa, independientemente de su tamaño puede incluir algún uso del BD en sus procesos. Existen multitud de herramientas gratuitas para análisis de datos comunes como transacciones entre cliente y empresa, interacciones con la página web o redes sociales...

2.7.1 Diferencias entre Business Intelligence y Big Data

Habiendo introducido brevemente los conceptos de BI y BD, parece que no está clara la línea divisoria de las soluciones propuestas de ambos, pudiendo incluso confundirse, malinterpretarse o usarse de manera incorrecta como términos. Entonces, ¿dónde reside su diferencia?

Su **principal diferencia** radica en la **fuentes de datos**. Mientras que las soluciones de BD utilizan cualquiera de los tres tipos de datos, las de BI se centran exclusivamente en los datos meramente estructurados. Si bien es cierto que, las últimas y más actualizadas herramientas de BI ya incluyen funcionalidades para el tratamiento de datos no estructurados en una clara búsqueda de ventaja competitiva frente a su competencia.

Evidentemente, aunque esta sea la diferencia más visible, existen otras más específicas.

- ✓ Mientras que las soluciones de BD utilizan un sistema de ficheros distribuido (SFD), las de BI se aprovechan de un servidor central (topología en estrella).
- ✓ Por otra parte, con BD podemos trabajar con datos históricos o datos en tiempo real, mientras que con BI, en la mayoría de casos (algunas herramientas tienen posibilidad de trabajar con datos en tiempo real), son datos históricos.
- ✓ En BD se usa la tecnología de MPP (procesamiento paralelo masivo), dividiendo las tareas en varias sub tareas que se ejecutan en paralelo para



mejorar la velocidad de análisis, aunque ciertas herramientas de BI comienzan a aprovecharse de ella también (véase más adelante ThoughtSpot).

En resumen, el BI pretende dar respuesta a preguntas habituales en empresas, proporcionar información sobre puntos fuertes o débiles en el negocio o mejorar relación con los clientes. Es decir, otorgar la posibilidad de hacer decisiones basadas en los datos. En cuanto a BD, nos ofrece las mismas posibilidades pero con la ventaja de poder utilizar diferentes fuentes de datos y mayor número de datos (evidentemente a un coste monetario y de recursos mayor).

Además, también podemos entender que **BI y BD son conceptos complementarios** el uno al otro. El BD nos permite estructurar y darle un sentido a esa información 'desestructurada' para que el BI la transforme en información con valor para la toma de decisiones empresariales.

2.8 Otros enfoques

Hasta el presente capítulo, hemos mencionado ciertos conceptos que rondan alrededor del mundo del BI. OLAP, OLTP, Data Mining, Big Data... Ahora, profundizaremos en otros dos conceptos que cobran fuerza en los últimos años: **Data Science, Analytics y Business Analytics**.

2.8.1 Data Science

El término Data Science surge en los años 90 en ambiente en el que se observó la necesidad de dar rigor matemático al análisis computacional de grandes volúmenes de datos. El término apareció en 1997 en la publicación de C.F. Jeff Wu llamada "Statistic=Data Science?" en el que destacaba algunas novedades en estadística.

Data Science reúne una serie de principios, definiciones de problemas, algoritmos y procesos para la extracción de información, en principio, desconocida a partir de grandes volúmenes de datos (Kelleher et al, 2018). Los elementos principales del Data Science provienen, a su vez, de campos relacionados tales como Machine Learning (ML) y Data Mining. A veces incluso utilizados indistintamente, ya que sus similitudes son evidentes, el Data Science abarca un espectro mayor de técnicas y posibilidades que el de los demás.

Machine Learning se ocupa en términos generales, del diseño y evaluación de algoritmos para la extracción automática de patrones y tendencias en los datos. Por otra parte, Data Mining se ocupa del análisis de datos estructurados en aplicaciones comerciales. Sin embargo, Data Science también se encarga de estas tareas y además, también toma parte en los procesos de extracción, depuración y transformación de datos no estructurados. Data Science se aprovecha al máximo en casos de grandes volúmenes de datos donde los patrones a buscar son demasiado complicados para ser realizado por humanos manualmente.

Por otra parte, Data Science se caracteriza por una serie de tareas fundamentales que puede realizar en un set de datos:

1. Clustering (segmentación) (**será visto en capítulos posteriores**)



2. Reglas de asociación (**será visto en capítulos posteriores**)
3. **Detección de anomalías** (outliers)

La detección de anomalías se basa en la búsqueda e identificación de elementos que no concuerdan con los elementos típicos en un set de datos. La aproximación más común y básica para esta actividad es crear reglas en la propia base de datos para separar las anomalías. Esta creación de reglas ya ha sido agregada en lenguaje SQL a través del comando MATCH_RECOGNIZE. Sin embargo, esta aproximación solo nos permite identificar de manera paulatina cada vez más anomalías (cuantas más anomalías diferentes se den, más cambiarán las reglas). Esto supone una desventaja, ya que solo identificaremos la anomalía cuando el sistema reconozca que es una anomalía. Sin embargo, este proceso puede ser realizado mediante la agrupación (clustering) a la inversa. Podemos agrupar los datos “normales” en un clúster y los anómalos en otros, o también podemos clasificarlos por “distancias” al centro de los datos (visto en Data Mining también). También podemos utilizar un modelo de árbol de decisiones, lo cual es más efectivo pero a la vez más complicado de modelar.

Una aplicación muy común, por ejemplo, es la búsqueda en los mercados financieros de actividades potencialmente fraudulentas. O incluso, en transacciones bancarias.

4. **Predicción** (clasificación y regresión)

La clasificación es la actividad que pretende medir la probabilidad de que un cliente tome una decisión o realice una acción. Normalmente, en la clasificación se hace referencia a “modelos de propensión” para determinar estos hechos.

El otro caso es la regresión, técnica por la que podemos predecir la medida de cualquier elemento mediante un modelo. Lo que podemos medir, por ejemplo, es el precio de cualquier cosa, desde un producto material hasta un servicio.

Las aplicaciones empresariales del Data Science son infinitas, desde las tradicionales empresas de manufacturación, pasando por sanidad hasta en los últimos años, tomar importancia en el deporte. Sin embargo, donde más utilidad tiene es en ventas y marketing. La necesidad de conocer lo que quiere el cliente es el proceso más complicado para cualquier negocio, y el método de prueba y error ya no tiene cabida en la industria actual. Data Science es el que ayuda a las personas a tomar decisiones sobre campañas de marketing y productos en función de los datos de los clientes.

La popularidad de Data Science es debida a varios factores. Por una parte, el emergente Big Data y la necesidad de abordar de alguna manera esa gran cantidad de datos que las organizaciones generan y almacenan. Por otra parte, también es debido al abaratamiento del almacenamiento tanto en la nube como en servidores físicos. Las economías de escala permiten que la unidad de almacenamiento sea más barata que nunca. Y por último, la aparición de tarjetas gráficas y sus Graphical Processing Units (GPUs) que han sufrido una mejora exponencial, pudiendo ser utilizadas para ML, procesamiento de datos y modelación.

La creciente popularidad de herramientas de BI ha provocado que el Data Science se haya empezado a incluir en ellas como una característica más, permitiendo a dichos softwares convertirse en herramientas muy potentes de análisis de datos y predicción.

2.8.2 Analytics y Business Analytics

En el otro lado tenemos a Analytics y Business Analytics, cuyos sendos términos suelen estar relacionados y acompañados también de Business Intelligence.

Analytics puede ser definido como el uso de técnicas estadísticas (tendencias, gráficas, etc), software de sistemas de información y metodologías de optimización de procesos (programación lineal) para explorar, visualizar y descubrir patrones. Analytics es una denominación más antigua que refiere a todas las disciplinas, no solo a los negocios. Hay diferentes tipos de Analytics (análisis), y tal y como se comentó en el capítulo 2.1, puede ser descriptivo, predictivo o anticipativo

Es aquí donde se encuentran las diferencias entre Analytics y Business Analytics. Mientras que Analytics pretende generar información de provecho a partir de las fuentes de datos, Business Analytics va un paso más allá y busca una mejora medible en el rendimiento corporativo. Por tanto, podemos definir Business Analytics como un proceso que comienza con la extracción de datos corporativos y que consiste en la aplicación secuencial de componentes de análisis descriptivo, predictivo y anticipativo para ayudar en el rendimiento del negocio y su toma de decisiones. El Business Analytics debe ser más que el mero Analytics, obteniendo resultados que son factibles de implementar y que realmente supondrán una mejora en el negocio.

Características	Analytics	BA	BI
Objetivo	¿Qué está pasando y qué pasará?	¿Qué está pasando ahora, qué pasará y cuál es la mejor estrategia para abordarlo?	¿Qué está pasando y qué hemos hecho anteriormente para abordarlo?
Usa análisis descriptivo	Sí	Sí	Sí
Usa análisis predictivo	Sí	Sí	No en principio, pero últimas herramientas sí
Usa análisis prescriptivo	Sí	Sí	No en principio, pero últimas herramientas sí
Usa los tres análisis conjuntamente	No	Sí	Podría con últimas herramientas
Centrado en negocios	Puede	Sí	Sí
Centrado en el almacenamiento y gestión de datos	No	No	Sí

Tabla 5. Comparación Analytics, BA, BI. Fuente: Schniederjans et al (2014)



La relación de dichos conceptos con BI, es que según la mayoría de autores, **BI engloba a estos dos, incluyendo a Data Science y los Sistemas de Información**. A mayores, BI también engloba el proceso de almacenamiento de la información, su arquitectura y su gestión además de los procesos de consulta, realización de informes y otras características visuales en las que se centra el BI. Es decir, BI busca conocer qué está pasando y dónde, y cuáles son las acciones a tomar basado en experiencia pasada, lo cual, es diferente en BA, el cual busca el por qué y cuál es el posible resultado futuro.

A modo de resumen, se expone en la Tabla 5 una comparativa de los tres términos: Analytics, BA y BI (Schniederjans et al, 2014).



Universidad de Valladolid



ESCUELA DE INGENIERÍAS
INDUSTRIALES



3. Herramientas de BI

Habiendo conocido brevemente el significado y concepto de BI, debemos considerar qué herramientas existen. Las herramientas son utilizadas por empresas de manera generalizada por sus empleados. Cada uno de ellos puede tener acceso pleno o no a las herramientas y sus datos, dependiendo del rol que tengan.

Los diferentes tipos de herramientas coinciden y vienen dados por las diferentes etapas del BI. En resumidas cuentas, este está formado por:

- I. Recopilación de datos.
- II. Procesamiento de datos.
- III. Estudio y análisis de datos.
- IV. Visualización e interpretación.
- V. Distribución.

Así pues, podemos sintetizar dicho proceso mediante 3 tipos de herramientas de BI.

1. **Herramientas para la gestión de los datos (Data Management Tools).** Pueden englobar los pasos de I y II del proceso. Permiten la manipulación de los datos, desde su estado inicial al ser obtenidos, desorganizados y no depurados, hasta su estado final de almacenamiento en sistemas tras haber sido filtrados, organizados y estandarizados para su análisis.
2. **Herramientas para el descubrimiento de datos (Data Discovery Tools).** Engloban específicamente la etapa III del proceso. Permiten la obtención de nueva información, no vista anteriormente a través de los datos (minería de datos, data mining), mediante el análisis de los mismos. Algunas, además cuentan con técnicas de análisis predictivo para pronósticos y evoluciones posibles.
3. **Herramientas de reporting.** Engloban las etapas IV y V. Sirven para la visualización gráfica e interpretación de la nueva información obtenida. Generalmente, consisten en herramientas que crean gráficos interactivos, cuadros de mando personalizables y facilitan el proceso de distribución al resto de trabajadores de la organización, además de poseer normalmente un sistema de restricción al acceso de datos en función del usuario.

Debemos destacar que la rápida y constante evolución de la tecnología ha permitido que, en términos generales, las **herramientas más vendidas** y conocidas disponibles en el mercado **reúnan en una sola herramienta los tres tipos**.

Durante el apartado 2.2 Arquitectura del BI se hizo referencia a que en secciones siguientes se haría hincapié en el 'stage' de **Reporting**, y es aquí donde lo haremos considerando las **características básicas** y funcionalidades primarias de los **softwares de BI** actuales.

1. **Cuadros de Mando Integrales (CMI).** Surgieron en 1992 de la mano de Kaplan y Norton como un tablero que permite convertir objetivos estratégicos del negocio en unos operativos del proceso de producción para conseguirlos. En

definitiva, es una herramienta que permite medir y monitorizar nuestros objetivos de forma periódica y de forma visual. Permite obtener una visión genérica de la empresa en las cuatro perspectivas básicas de ésta: finanzas, clientes, procesos internos y aprendizaje y conocimiento.

Para cada una de las perspectivas existen cuatro elementos (Tabla 6).

Perspectiva	Aspecto clave para la formulación de la estrategia
Objetivo estratégico	Aquello que queremos lograr a través de la estrategia.
Mapa estratégico	Interconexión de los objetivos estratégicos a través de un grafismo
Indicador estratégico	Mide el nivel o porcentaje del logro que hemos conseguido
Meta	Valor del indicador en un momento dado

Tabla 6. Características de las perspectivas. Fuente: elaboración propia

- a. **Perspectiva financiera.** Esta perspectiva es la considerada de hechos y acontecimientos pasados, ya que la contabilidad de una empresa no puede ser realizada en tiempo real (debido a plazos de pago por parte de proveedores, por ejemplo) . En ella se han de incluir objetivos relacionados con la contabilidad y otros datos económicos de la empresa. Ejemplos: crecimiento de beneficios, optimización de gastos...
- b. **Perspectiva clientes.** El funcionamiento óptimo de la empresa y de la perspectiva financiera pasa por la satisfacción de los clientes. Como es obvio, ha de buscarse una buena satisfacción del cliente y a la vez, mantener su fidelidad, ya que sin los consumidores, el negocio no sería posible. Ejemplos: creación de valor, imagen de marca, mejorar relación y comunicación...
- c. **Perspectiva procesos internos.** Analiza los proceso propios para averiguar si son adecuados para mejorar la perspectiva cliente y la financiera. Existen cuatro tipos de procesos:
 - i. Procesos de operaciones
 - ii. Procesos de gestión de clientes
 - iii. Procesos de innovación
 - iv. Procesos relacionados con el medioambiente y sociedad.Ejemplos: calidad del producto, rendimiento de productos...
- d. **Perspectiva aprendizaje y conocimiento.** Considera los conocimientos, herramientas y posibilidades que tiene nuestro equipo en su mano para conseguir los objetivos definidos. Ejemplos: capacitación de empleados, motivación de empleados...



Cada uno de los ejemplos de objetivos propuestos de cada perspectiva puede tener uno o varios indicadores. Además, los **objetivos estratégicos** han de cumplir el acrónimo **S.M.A.R.T.**

- **Specific (Específico).** Los objetivos deben ser concretos y detallados, de manera que cualquier persona, incluso personas ajenas al negocio, comprendan el significado del objetivo.
- **Measurable (Medible).** Los objetivos deben poder ser cuantificados, especialmente a la hora de crear indicadores para ellos y medir la consecución de estos.
- **Attainable (Alcanzable).** Los objetivos deben suponer un reto y una mejora, pero debe ser realizable y posible.
- **Realistic (Realista).** Cualquier objetivo debe tener en cuenta nuestras capacidades y recursos disponibles, y por ello, han de ser realistas.
- **Time-bound (Acotado en el tiempo).** Independientemente del objetivo y su ámbito, éste ha de tener una línea temporal y estar definido en el tiempo para poder marcar los pasos y etapas.

En la Ilustración 15 se expone un ejemplo en el que se puede observar un CMI en su totalidad, desde el mapa estratégico donde se interrelacionan cada uno de los objetivos estratégicos, pasando por los indicadores de cada uno de los objetivos, hasta el estado actual de los indicadores.

1. KPIs (Key Performance Indicator)

En el apartado anterior, se habló de indicadores. Estos indicadores, no exclusivamente los de un Cuadro de Mando Integral, se conocen como KPIs. Sirve de medida del nivel de desempeño de un proceso, habiendo fijado un objetivo previamente. Las funciones principales de un KPI son informar, controlar, evaluar y ayudar en la toma de decisiones. Es por ello, que no existe una lista prediseñada de KPI, si no que éstos varían en cada empresa, incluso entre las de un mismo sector.

Al igual que los objetivos, **suele decirse que los KPI**, a su vez, han de ser **también SMART**.

A continuación se muestran algunos ejemplos por sector/departamento comúnmente utilizados (Rodríguez, 2016).

- a. Finanzas.
 - i. Beneficio Neto
 - ii. TSR (Rentabilidad total para el accionista)
 - iii. Tasa de crecimiento de ingresos
 - iv. ROI (Rentabilidad de la inversión)
 - v. ROCE (Rentabilidad sobre el capital empleado)...
- b. Marketing.
 - i. Cuota de mercado
 - ii. Valor de la marca
 - iii. Tasa de conversión a clientes
 - iv. Cuota de menciones en línea (OSOV)
 - v. Índice de Klout...

Mapa Estratégico		Objetivos Estratégicos		Indicadores		(selección de datos)		Metas (d. Actual)		Semáforo / éxito	
Finanzas		Aumentar Rentabilidad	Reducción de gasto	Grado de rentabilidad		Margenes		0,0485	-0,1178	●	fracaso
				Disminuir gastos totales trimestrales		Margen operativo TTM		3000	2951,33	●	mejorar
				Disminuir gasto en instrumentos trimestrales		31 Dic. 16		500	400,18	●	mejorar
				Disminuir deuda financiera anual		2015		800	2696,3	●	fracaso
				Fortalecer Financiera		Ayuda a largo plazo/ Total fondos propios M		1,2369	2,0343	●	éxito
				Ventas totales anuales		2015		60000	58000	●	mejorar
				Número países con presencia		Teila		12	14	●	éxito
				Ingreso total trimestrales		2 Semestre 2014		3000	2294,63	●	mejorar
				Índice satisfacción del cliente respecto competidores		31 Dic. 16		7,8	9,1	●	éxito
				Mejorar experiencia del cliente		2013		25000	23000	●	mejorar
Clientes		Mejorar Experiencia Cliente	Mejorar problema de cuellos de botell	Porcentaje clientes volverían a comprar un Teila		Teila 2017		60	51	●	éxito
				Número de vehículos del nuevo Model 3 producidos		21 2017		1000	40	●	fracaso
				Número de trabajadores		2015		15000	13058	●	mejorar
				Producción anual		4 T. 2016		25000	22200	●	mejorar
Proceso I.		Mejorar Problemas	Aplicación de tecnología desarrollad	Cantidad de características nuevas en un modelo de automóvil		2013		10	14	●	éxito
				Porcentaje de coches con fallos por trimestre		2 T. 2017		5	17	●	fracaso
				Porcentaje de empleados identificados con la empresa		2015		85	80	●	mejorar
Aprendizaje		Efectividad Organizacional	Productividad personal	Cursos de formación recibidos desarrollo anuales		2013		12	8	●	mejorar
				Porcentaje de aprobados tests examen de cursos		2014		80	70	●	mejorar
				Número de sugerencias innovadoras		2013		40	50	●	éxito
Aprendizaje		Mejorar Línea Producción	Efectividad Organizacional	Porcentaje productividad por persona y mes		2013		100	90	●	mejorar
				Porcentaje eficiencia		2013		40	29	●	mejorar

Ilustración 15. Ejemplo de CMI. Fuente: elaboración propia



- c. Marketing.
 - i. Cuota de mercado
 - ii. Valor de la marca
 - iii. Tasa de conversión a clientes
 - iv. Cuota de menciones en línea (OSOV)
 - v. Índice de Klout...
- d. Operaciones
 - i. Eficiencia global de equipos (OEE)
 - ii. Medida del valor ganado (EV)
 - iii. Tasa de contracción de inventario (ISR)
 - iv. Nivel de residuos de los proceso
 - v. Tasa de entrega completa y a tiempo (DIFOT)...
- e. Recursos humanos
 - i. Valor agregado del capital humano (HCVA)
 - ii. Ingresos por empleado
 - iii. Índice de satisfacción de los empleados
 - iv. Promedio de permanencia del personal
 - v. Factor de Bradford (absentismo)...
- f. Clientes
 - i. Tasa de retención de clientes
 - ii. Índice de satisfacción de clientes
 - iii. Rentabilidad del cliente
 - iv. Quejas de clientes...

2. Minería de datos (Datamining)

La minería de datos es un campo de la estadística y las ciencias de la computación dedicado a extraer conocimiento útil, comprensible y novedoso de grandes volúmenes de datos, siendo su principal objetivo encontrar información oculta o implícita que no es posible obtener mediante métodos estadísticos convencionales (Moine et al, 2011). La minería de datos suele confundirse o usarse como sinónimo de KDD (Knowledge Discovery in Databases). Sin embargo, la realidad es que el KDD, acuñado en 1989, es la extracción de conocimiento a partir de bases de datos mientras que la minería de datos es la mera etapa de análisis del KDD (como la preparación de datos, selección y limpieza de éstos, etc).

En cuanto a la minería de datos se refiere, también normalmente es confundida su función real. Es usada para referirse simplemente a grandes volúmenes de datos, procesamiento de la propia información (recolección, extracción, almacenamiento y análisis) o hasta cualquier tipo de aplicación para la ayuda de toma de decisiones. Lejos de la realidad, la minería de datos es el análisis automático o semiautomático de grandes cantidades de datos para la búsqueda de patrones o correlaciones (podemos entenderlo como resúmenes de los datos de entrada).

La minería de datos tiene un fundamento muy marcado en la inteligencia artificial, aprendizaje automático y técnicas estadísticas. Las técnicas, en general, más utilizadas son (Maimon et al, 2010) (UIAF, 2014):

- a. **Redes neuronales.** Modelo de computación similar al sistema neuronal biológico basado en modelos matemáticos y utilizado para aprendizaje



y procesamiento automático. Los resultados obtenidos mediante redes neuronales provienen de tres funciones:

- i. Propagación (función de excitación)
- ii. Activación (función de modificación de la función anterior)
- iii. Transferencia (aplicación del valor devuelto por ii.)

Existen diferentes tipos de redes neuronales y diferentes clasificaciones de éstas (por topología, tipo de aprendizaje, tipo de entrada).

Sus ventajas principales son: aprendizaje adaptativo, autorganización de la información, procesamiento en tiempo real, tolerancia a fallos y fácil integración.

- b. **Algoritmos de agrupación.** Son conocidos a su vez como procedimientos de “*clustering*”. Basa su funcionamiento en disponer los datos analizados en conjuntos de vectores que sean lo más similares posibles, y a la vez, que los diferentes conjuntos presenten más diferencias entre sí. Se establecen distancias entre elementos y distancia al centro de los conjuntos. La complejidad radica más en establecer los criterios para similitudes que en el propio proceso. Ejemplos:
 - i. Algoritmo K-means
 - ii. Algoritmo K-medoids
- c. **Reglas de asociación.** Considerando un conjunto de datos, recorren todas las combinaciones de características (dimensiones) de la base de datos para descubrir los hechos que mayor probabilidad tienen de ocurrir o son comunes en el conjunto.
- d. **Árboles de decisión.** Modelo de predicción que divide un conjunto de análisis, buscando el mayor grado de pureza entre los grupos resultantes mediante un diagrama de construcciones lógicas. Se dispone de un nodo inicial denominado raíz que incluye todos los individuos y se va dividiendo en nodos de acuerdo a una función discreta sobre las variables. Ejemplos:
 - i. Algoritmo ID3
 - ii. Algoritmo C4.5 (Ilustración 16)
 - iii. CART
 - iv. CHAID Random Tree
- e. **Regresión lineal.** Método estadístico más utilizado para correlacionar los datos a través de un modelo matemático. Ineficaz en espacios multidimensionales.



Ilustración 16. Ejemplo de Algoritmo C4.5. Fuente: UIAF (2014)

- f. **Máquinas de soporte vectorial (MSV).** Utilizados en problemas de clasificación y regresión, la idea fundamental consiste en que, teniendo un conjunto de puntos (que a su vez es subconjunto de un conjunto mayor denominado espacio) donde cada uno de los puntos puede pertenecer solamente a dos categorías, el algoritmo de MSV pretende predecir el posicionamiento de un nuevo punto que se añade de entre las dos categorías. El MSV buscará la separación óptima de las dos categorías mediante un hiperplano, buscando el máximo margen o distancia con los puntos más cercanos a éste (Ilustración 17).

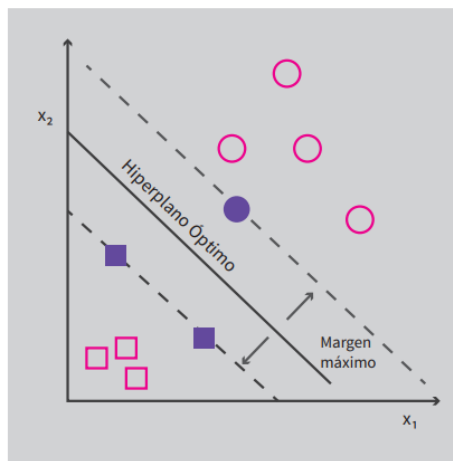


Ilustración 17. Ejemplo MSV. Fuente: UIAF (2014)

Aunque existen más técnicas, se han nombrado las más destacadas. Por otra parte, es fácil determinar que la **minería de datos es la parte fundamental en cualquier herramienta de BI**, ya que es donde radica la capacidad de análisis y funcionalidades más importantes a la hora de tratar los datos.



3. **Dashboards (“tableros”).** Son la principal funcionalidad de las herramientas de BI. Los dashboards son elementos de las herramientas de BI que se utilizan a modo de panel de control en el que se expresan, generalmente de manera visual, todo tipo de información que se considera relevante. A la vez, se puede usar para representar los principales KPI definidos de la organización. La capacidad de personalización de estos dashboards son prácticamente ilimitadas: desde colores, hasta disposición de las gráficas, generación de varias hojas (al estilo Excel) de dashboards, diseño, modificación de tamaños, etc.

También suelen tener capacidad de convertirse en interactivos, ya sea mediante características drag-and-drop o funcionalidades de agrupación o desagregación de los datos en cada uno de los elementos visuales.

La Ilustración 18 muestra un ejemplo de dashboard de una empresa.

4. **Informes (Reports).** La prestación última de cualquier herramienta de BI es la capacidad de distribuir la información generada. Cualquiera de las herramientas disponibles permiten la generación de informes automáticos, distribuidos y periódicos de la información generada a través de su análisis. Los informes pueden ser presentados mediante plantillas y distribuidos por departamentos y restringiendo su acceso por rango o rol. Permiten multitud de personalizaciones al igual que los dashboards, y sirven de ancla a los métodos más tradicionales de comunicación corporativa. Su gestión, creación y seguridad también puede ser restringida y atribuida solo a ciertos usuarios.

El propio proceso de esta distribución de información requiere de plantearse quién requiere de información, cuándo la requiere y qué exactamente requiere. Se debe generar el exceso o escasez de información por parte de los usuarios que lo requieran, ya que puede producir la pasividad del usuario por el exceso o aparición de problemas por su escasez.

La mayoría de las soluciones comerciales de BI ofrecen los servicios de informes por un precio adicional en función del número de usuarios que se beneficiarán de éste.

El mercado de BI está en un constante crecimiento tal y como se prevé que crezca desde los 15,64 mil millones de dólares en 2016 hasta los 29,48 mil millones de 2022, un crecimiento de más del 88% en tan solo 6 años (Statistics MRC, 2017). Un pronóstico aún más esperanzador intuye que crezca hasta los 34.3 mil millones de dólares en 2022 (Market Research Future, BI Market Research Report - Global Forecast to 2022, 2019).

Esto se traduce en una amplia gama de vendedores y productos que ofrecen soluciones, generalmente, parecidas y con las mismas prestaciones.

Sin embargo, en esta comparativa se escogen los tres productos por excelencia como herramientas de BI, ampliamente reconocidas y usadas, tildadas como líderes, con una buena valoración y en constante mejora: Power BI, Tableau y Qlik. Por otra parte, incluiremos como cuarta opción aquella que está surgiendo en los últimos años como una de las opciones más innovadoras y con mayor carácter diferenciador del mercado, además de estar cosechando muy buenas valoraciones por los usuarios y por estudios en herramientas de BI: ThoughtSpot.

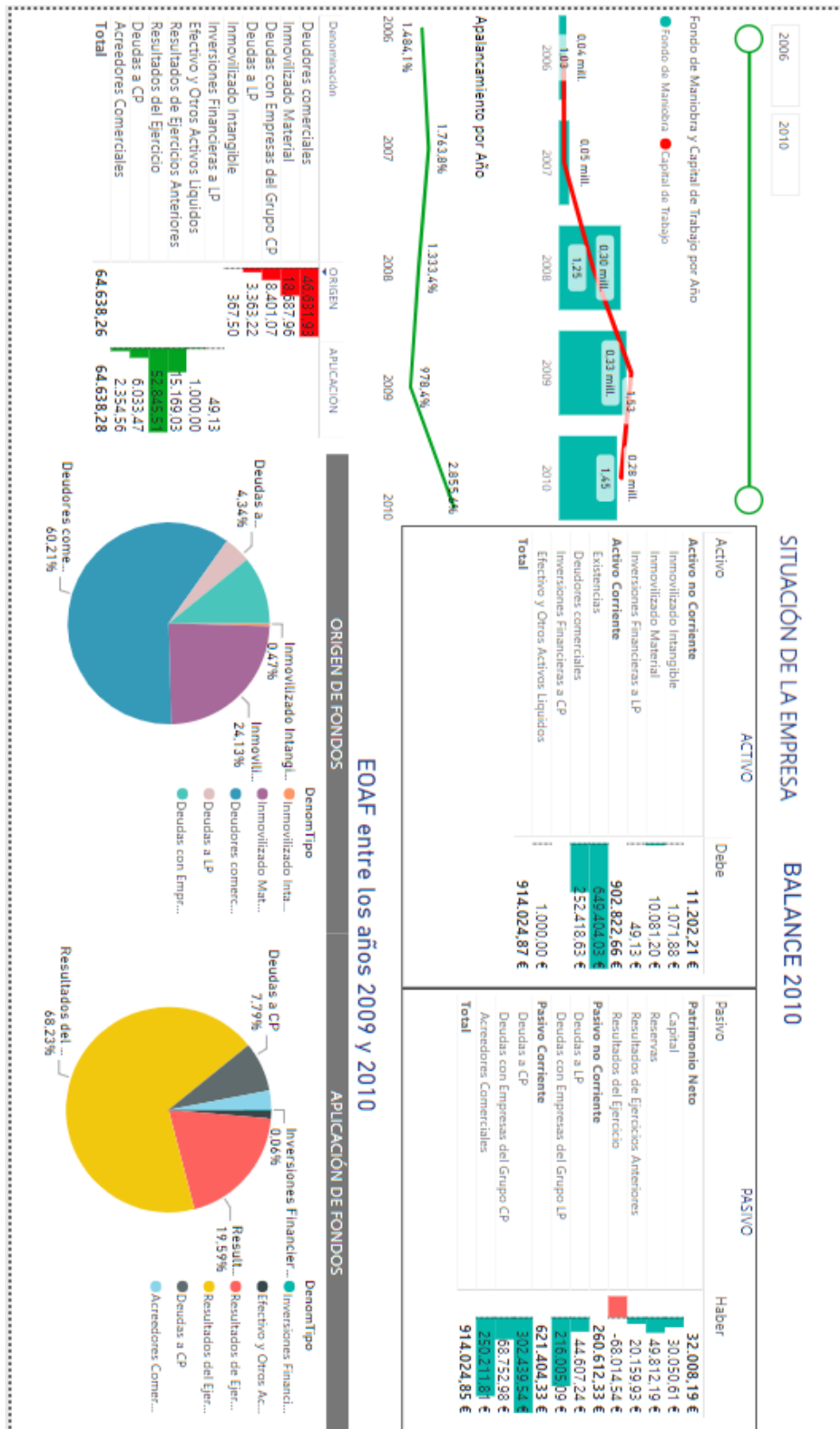


Ilustración 18. Ejemplo de dashboard en Power BI. Fuente: cedido por entidad BPMSat s.l

3.1 Microsoft Power BI

La apuesta de Microsoft para el terreno del descubrimiento de datos y análisis es Power BI. Es una herramienta de análisis empresarial basada en la visualización (mediante gráficos y cuadros de mando interactivos) y distribución de la información generada con la organización.

Los variantes de su producto que ofrece Microsoft son progresivas, es decir, desde la versión básica gratuita de creador, pasando a las de pago que permiten compartir informes y datos.

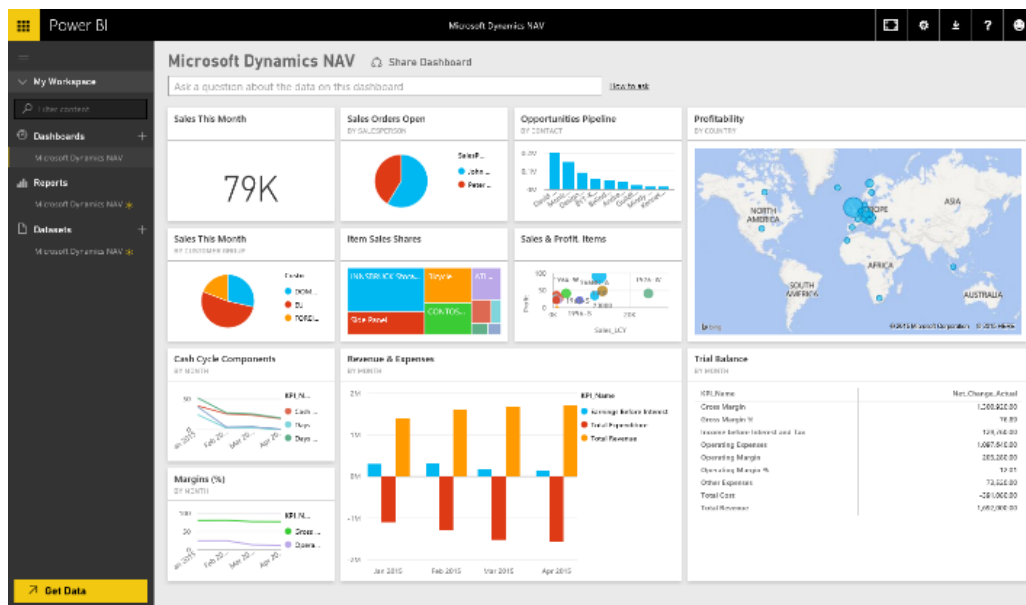


Ilustración 19. Interfaz Power BI (recuperado de: <https://powerbi.microsoft.com/es-es/blog/exploring-your-microsoft-dynamics-nav-data-with-power-bi/>)

Por una parte, la versión gratuita Power BI Desktop es el cliente para el análisis y creación de cuadros de mandos interactivos y visuales, además de gráficos y tablas. Es decir, el software solo permite la preparación y descubrimiento de datos, pero no la distribución. Tiene características de analítica aumentada, por la que busca a través de algoritmos sofisticados patrones de tendencias y correlaciones en conjuntos de datos no estructurados. Microsoft, además, pone a disposición del usuario una librería online de recursos visuales personalizados y personalizables, creados por Microsoft y sus partners, para sacar el máximo provecho del análisis visual.

Power BI posee una gran accesibilidad a cientos de fuentes de datos, desde archivos locales o en la nube, como Dynamics 365, Salesforce, Azure SQL DB, Excel y SharePoint (Power BI, 2019).

La preparación de datos ha sido facilitada mediante la funcionalidad “Power Query”, ya conocida por los usuarios de Excel, la cual ingiere, transforma, integra y enriquece los datos. El lenguaje de preparación de datos para funciones avanzadas es DAX, haciendo su uso difícil para usuarios poco experimentados. Además, todos los pasos y etapas seguidos en la preparación de los datos pueden ser rastreados y editados, lo



que además de hacerlo flexible, lo convierte en seguro. También posee búsquedas NLQ (Natural Language Query), a través del cual permite la búsqueda de información con lenguaje no programado ni específico.

Una característica importante para comenzar a estudiar y analizar datos que resultan poco conocidos, extraños o poco estandarizados es “Quick Insights”, la cual se puede definir como un asistente (comúnmente conocido como “wizard”) que analiza los datos de manera automática buscando anomalías y tendencias, proporcionando sugerencias para representaciones o interpretaciones.

Las versiones de pago son las Power BI PRO y Power BI Premium. Son aquellas que permiten la distribución de informes.

Por una parte Power BI Pro permite la distribución interna de datos. También permite la regulación del acceso a esos datos mediante la creación de roles en el sistema. Su modelo de precios es por usuario/licencia.

Por otra, Power BI Premium permite la distribución de datos, además de creación y distribución de informes, con usuarios externos a la organización sin la necesidad de un pago de licencia por usuario. A diferencia de Power BI Pro, Power BI Premium no se rige por precios por usuario/licencia, permitiendo tener un número ilimitado de usuarios. Sin embargo, su precio varía según la memoria RAM (25, 50, 100 GB) deseada en servidores dedicados. Además, tiene características de escalabilidad para aumentos de usuarios posteriores.

Otros productos complementarios son Power BI Mobile (plataforma de consulta para dispositivos móviles), Power BI Embedded (dependiente de Azure y utilizado para consulta de datos en webs y aplicaciones externas) y Servidor de Informes Power BI (generación de informes).

Servidor de Informes Power BI está incluido en Power BI Premium.

Según el informe anual de Gartner sobre las principales herramientas de BI, Microsoft Power BI se encuentra en el cuadrante de los “Líderes”.

Ventajas

- ✓ Capacidad de rastreo de pasos en preparación de datos y capacidad de realizar cambios en ellos.
- ✓ Facilidad y asistencia en la preparación de datos mediante Power Query con posibilidad de añadir reglas o funciones.
- ✓ Buena relación calidad-precio. En comparación a sus competidores, los precios ofrecidos por Microsoft son bajos.
- ✓ Alta satisfacción de los clientes que usan el software. La facilidad de uso y el precio siguen siendo dos de los factores clave a la hora de elegir Power BI como herramienta de BI.
- ✓ Actualizaciones constantes y hoja de ruta basada en tecnología innovadora y disruptiva. Microsoft está invirtiendo en tecnología puntera que desea integrar en Power BI. Entre ellas, analítica aumentada, nuevas características de AutoML (Auto Machine Learning, disponible en Azure ML), Cognitive Services de Azure (análisis de textos, sentimientos e imágenes) o Inteligencia Artificial.



Desventajas

- ✓ Tamaño limitado para importación de datos y restricción a consulta en vivo de éstos a ciertos tipos (BARC Score Data Discovery, 2018).
- ✓ Falta de unificación de productos. Power BI tiene una gama de productos extensa y segmentada, sin poseer una opción que posea todas y cada una de las herramientas disponibles integradas e incluidas en el precio.
- ✓ Amplio abanico de análisis y visualización, bajo en generación de informes. Servidor de Informes Power BI, centrado en generación de informes (tipo 3), tiene carencias en sus prestaciones en comparación con los demás (más centrados en el análisis y distribución de elementos visuales, tipo 1 y 2), tales como la autorización de contenidos y distribución selectiva.
- ✓ Funciones de análisis automático solo disponible en la nube y deficientes.

3.2 Tableau

Tableau nace del proyecto científico Polaris llevado a cabo por la universidad de Stanford (Computer Graphics at Stanford University, 2000-2002). Tableau fue fundada en 2003 y pretende desarrollar un software que requiere poca formación y conocimiento. Es por ello que ofrece una herramienta basada en interfaz de búsqueda y exploración visual e interactiva que permite a los usuarios acceder, preparar, analizar y presentar y distribuir los resultados de los datos sin necesidad de conocimientos o habilidades en programación.

Tableau ofrece cuatro productos: Tableau Prep, Tableau Desktop, Tableau Server y Tableau Online.

Aunque Tableau Desktop reúne ciertas funcionalidades de preparación de datos, el lanzamiento de Tableau Prep ha mejorado considerablemente la experiencia en el proceso. Tableau Prep tiene compatibilidad con cientos de fuentes de datos (Tableau, 2019), ya sean datos en instalaciones físicas o en la nube o ya sean hojas de cálculo o bases de datos. Ha conseguido acercar el proceso a una metodología visual, con sugerencias de modelado, perfilado y trazabilidad de los datos. Permite al usuario editar en tiempo real los datos y deshacer pasos si es necesario. Además, es capaz de transformar tareas repetitivas como las agrupaciones, en tareas de un solo clic. Por último, permite la supervisión del flujo de trabajo mediante Tableau Prep Conductor en el entorno del servidor.

Tableau Desktop es el producto central de la empresa. Se centra en la capacidad humana de detectar patrones mediante la vista. Es un cliente de escritorio a partir del cual se pueden realizar multitud de análisis de los datos. Pretende ser un software de fácil manejo, poseyendo una interfaz casi completamente basada en drag-and-drop. Esto ha convertido al software en una elección muy común en el sector empresarial, donde seguramente no todo el mundo tenga conocimientos profundos en análisis de datos. Tableau Desktop permite también la manipulación de datos en tiempo real mientras se trabaja con ello, convirtiéndolo en una herramienta muy flexible en descubrimiento de datos. Una gran característica de Tableau son sus "Mapas", en las que tiene incorporado todos los códigos postales de más de 50 países, permitiendo geolocalizar cualquier dato y crear mapas interactivos y dinámicos. Por último, también permite el trabajo sin conexión, con la capacidad de extraer grandes cantidades de datos a la memoria en segundos. Esto es en gran parte posible gracias a su motor de

datos in-memory de alto rendimiento Hyper, siendo una tecnología puntera en el sector.

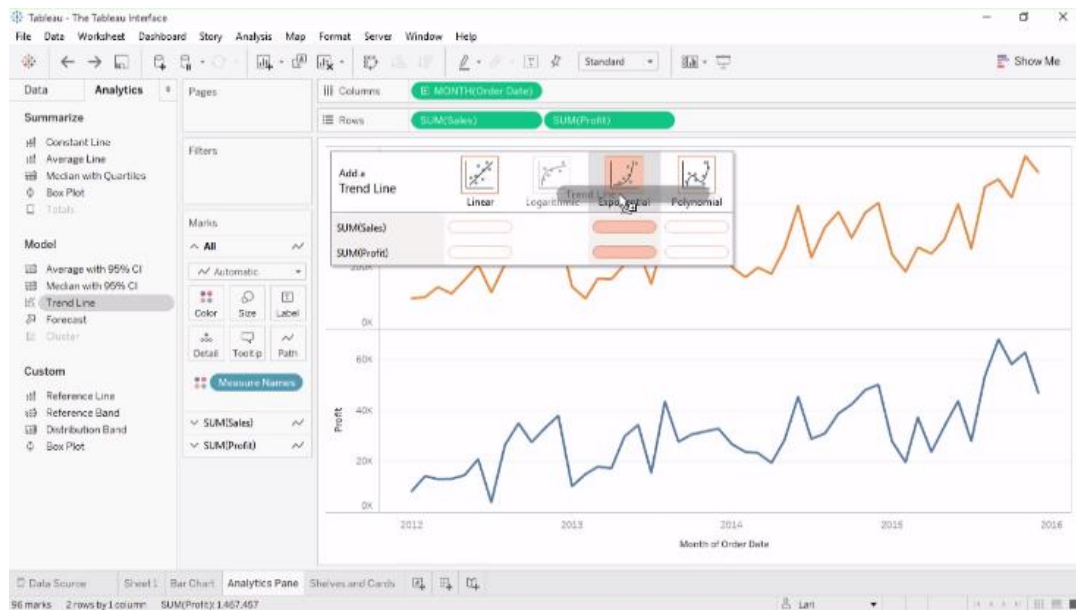


Ilustración 20. Interfaz Tableau (1) (recuperado de: <https://www.tableau.com/es-es/learn/tutorials/on-demand/tableau-interface>)

Por último, las opciones para análisis self-service en organizaciones que ofrece Tableau son Tableau Server y Tableau Online. La principal diferencia es que Tableau Online se basa en una 'nube', para aquellas organizaciones que no deseen poseer su propio servidor, siendo su gran ventaja el ahorro en costes de infraestructura e instalación, además del ahorro en tiempo.

Ambos permiten la compartición de datos y gráficos, además de capacitar al usuario a explorar, crear cuadros de mando y realizar consultas con lenguaje natural (NLQ). Por otra parte, Tableau Online permite también las mismas funcionalidades desde la web.

Por último, Tableau ofrece la aplicación para dispositivos móviles Tableau Mobile, la cual actúa como ventana hacia Tableau Server o Tableau Online, y funcionalidades "embedded" mediante creación de aplicaciones y análisis con la API de JavaScript.

El modelo de precios de Tableau es por suscripción por usuario/licencia. Ofrece diferentes precios para organizaciones dependiendo si es utilizado Tableau Server u Online (siendo Online más caro). También ofrece diferentes roles a diferentes precios en función de las funcionalidades (Creator, Explorer, Viewer), intentando abaratar su instalación.

Tableau, según el informe de Gartner es también un "Líder".

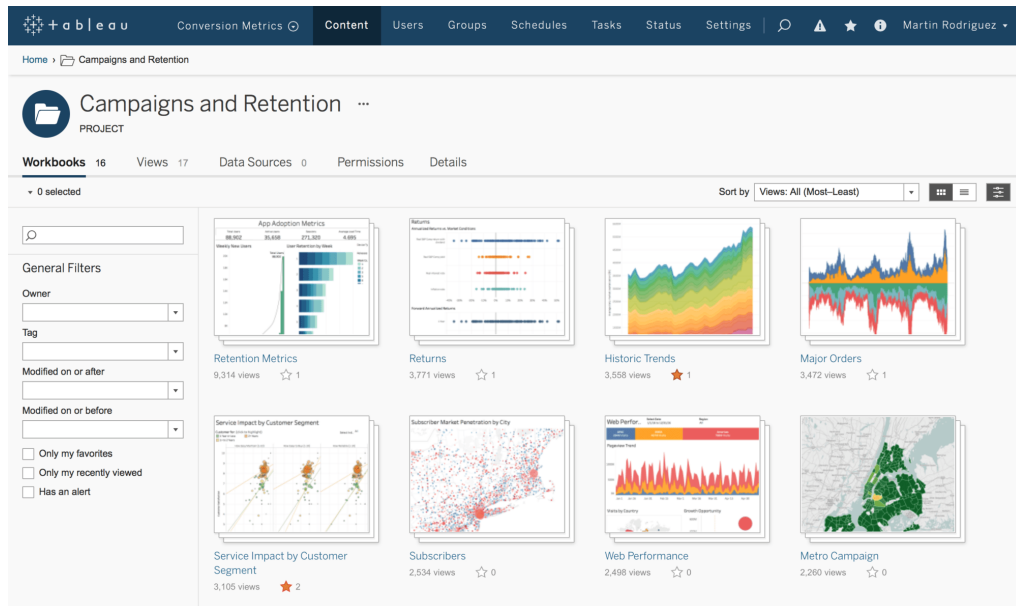


Ilustración 21. Interfaz Tableau (2) (recuperado de: <https://www.tableau.com/es-es/products/server>)

Ventajas

- ✓ Interfaz de fácil manejo para análisis y preparación de datos con drag-and-drop.
- ✓ Análisis visual inteligente con alta interactividad.
- ✓ Tableau Desktop incluye funcionalidades (limitadas) de Tableau Prep.
- ✓ Posibilidad de interacción con información e informes distribuidos. Posibilidad de anotaciones, debates y conversaciones entre usuarios (BARC Score Data Discovery, 2018).

Desventajas

- ✓ Carencias en productos. Ausencia de opciones distribución parcial de informes (bursted reports) en los que se incluye solo la información necesaria para un individuo concreto.
- ✓ La preparación de datos requiere de una herramienta independiente (Tableau Prep), lo que provoca la necesidad de cambio de software durante el trabajo.
- ✓ Escasa información y asistencia al usuario para análisis avanzado.
- ✓ Cambios en suscripciones. La suscripción 'Viewer' ha sido añadida recientemente, necesitando cambios en suscripciones y usuarios para aquellos que ya tenían contratado Tableau (Gartner Inc., 2019)
- ✓ Limitaciones en las cantidades de cada tipo de usuario.

3.3 Qlik

Hasta la aparición en 2014 de Qlik Sense, Qlik solo tenía un único producto, Qlik View. Ahora, la empresa proporciona un amplio catálogo de herramientas para análisis de datos: Qlik Sense Enterprise, Qlik Sense Cloud, Qlik Analytics Platform, Qlik View, Qlik NPrinting y otros (Qlik GeoAnalytics, Qlik Associative Big Data Index, Qlik DataMarket, Qlik Data Analyst) En 2017 adquirió Idevio para mejorar sus prestaciones en análisis geográfico con su herramienta Qlik GeoAnalytics.

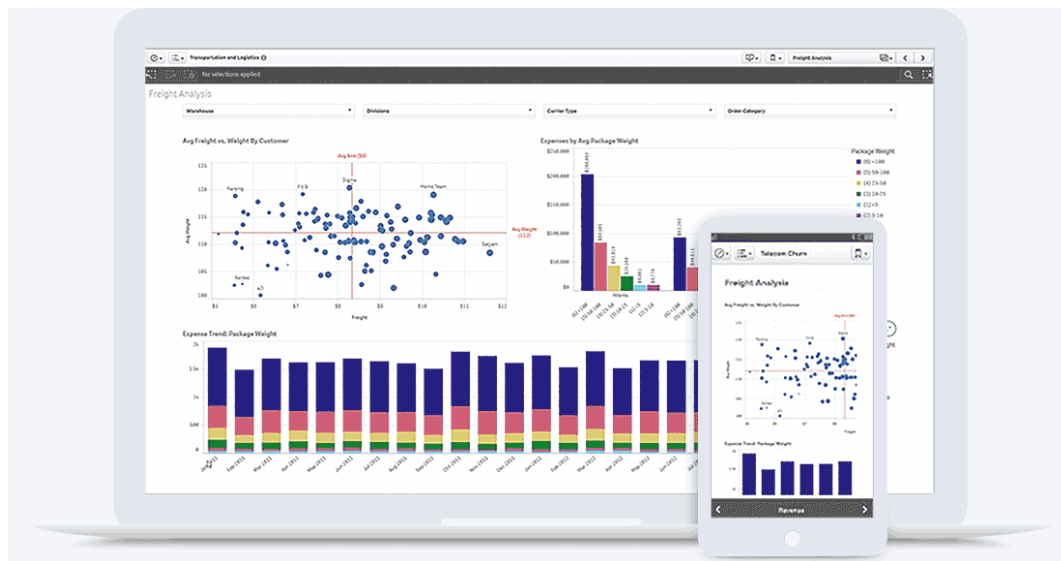


Ilustración 22. Interfaz QlikSense (recuperado de: <https://www.qlik.com/es-es/products/qlik-sense>)

Qlik proporciona un servicio y herramienta de exploración y descubrimiento de datos self-service arrojando resultados inmediatos. Su producto central es Qlik Sense, con sus dos variantes Qlik Sense Enterprise (empresas de gran tamaño) y Qlik Sense Cloud Business (individuos y equipos, basado en nube). El desarrollo del software basado en su motor asociativo QIX permite combinar fuentes de datos en una única visualización, además de indexar cada relación posible entre los datos para proporcionar sugerencias (Qlik, 2019). Por otra parte Qlik Sense, a diferencia de otras herramientas, no está limitada a consultas programadas y predefinidas, por lo que hace mucho más fácil el descubrimiento de datos, sin la necesidad de recurrir a programadores ni desarrollar nuevos tipos de consulta.

La preparación de datos en Qlik Sense también se ha visto mejorada en las últimas versiones con ayudas visuales a la preparación y funcionalidad profiling, es decir, sugerencias y análisis personalizado en función de lo realizado por el usuario, sin necesidad de scripts.

La herramienta ofrece atractivas representaciones visuales, además de realizar sugerencias en tiempo real de visualización de datos que estén siendo seleccionados o utilizados.

Qlik Sense también ofrece compatibilidad completa con dispositivos móviles.

La oferta “embedded” de Qlik es Plataforma analítica Qlik, con capacidad de incrustar en aplicaciones externas, portales o en cualquier otro sitio los servicios de análisis visual de datos. Qlik cuenta con una librería de APIs a disposición del cliente para crear rápidamente extensiones.

Por otra parte, QlikView es la forma ‘primitiva’ de QlikSense, reuniendo parte de sus características y funcionalidades, además de compartir el famoso motor asociativo. Sin embargo, tiene capacidades más limitadas y menos carácter destinado al descubrimiento de datos.

Qlik Core es la plataforma de desarrollo de aplicaciones para instalaciones en nube, El potente motor in-memory permite desarrollar aplicaciones robustas, agradables a la vista y fiables a la hora de analizar datos. Puede soportar multitud de fuentes de datos, modelos de datos complejos y cálculos avanzados.



Ilustración 23. Interfaz Qlik View (recuperado de: <https://www.qlik.com/es-es/products/qlikview>)

Por último, Qlik NPrinting está dedicada a la generación de informes a partir de datos y análisis de Qlik Sense y Qlik View.

Gartner sitúa a Qlik en el cuadrante de “Líderes”.

Ventajas

- ✓ Características del producto y escalabilidad. Fuertes funcionalidades de analítica aumentada con Cognitive Engine, Insight Advisor y otros, permiten al usuario encontrar información y resultados más rápido con creación de gráficos automáticos (Gartner Inc., 2019)
- ✓ Navegación intuitiva, rápida y flexible dentro de los datos, con capacidad de aplicar filtros dinámicos.
- ✓ Comunidad de usuarios muy activa. Gracias a la participación de la comunidad, es muy fácil formarse a partir de ella y obtener respuestas a dudas frecuentes.
- ✓ Actualizaciones con tecnologías recientes. Basa su hoja de ruta en el big data, analítica aumentada y preparación de datos mejorada.
- ✓ Integración de R y Python disponible para preparación de datos y análisis visual (BARC Score Data Discovery, 2018).

Desventajas

- ✓ Segmentación de productos. Falta de herramienta completa que reúna el flujo de trabajo normal en ella. La existencia de multitud de productos complementarios entre sí hace necesario el cambio de herramienta para hacer diferentes tareas y completar el flujo de trabajo.



- ✓ Carencias en la visualización de datos. En comparación con sus competidores, Qlik ofrece gráficos, tableros y cuadros de mando menos flexibles y fáciles de comprender.
- ✓ Siguen siendo necesarios ciertos scripts para la preparación de datos a pesar de mejorar la facilidad de uso.
- ✓ Amplias diferencias en funcionalidades entre Qlik Sense y QlikView (destinados al mismo mercado) y mala migración de QlikView a QlikSense.

3.4 ThoughtSpot

A pesar de que la compañía fue fundada en 2012 por antiguos ejecutivos de Google, no ha sido hasta hace un par de años cuando sus productos empezaron a destacar por su innovación y tecnología.

ThoughtSpot es una herramienta que puede ser instalada en la nube o en dispositivo, con los datos cargados in-memory en sistemas de procesamiento MPP (“*Massively Parallel Processing*”) e indexado para consultas rápidas. Es una herramienta de análisis de datos basado en búsquedas con el primer motor de búsqueda relacional. Esto permite analizar datos a través de preguntas simples en su campo de búsqueda (NLQ, “*Natural Language Queries*”), realizando cálculos y operaciones necesarias en el momento indicado.

Destaca en ella la herramienta SpotIQ, la cual a través de IA y el motor de búsqueda in-memory creado desde cero para ThoughtSpot realiza análisis automáticos de los datos. Es capaz de autodetectar anomalías y outliers a través de complejos algoritmos, encontrar correlaciones entre datos y mostrar tendencias que no se habían detectado en amplios conjuntos de datos no purgados (Thoughtspot, 2019). Además, el propio SpotIQ, con el algoritmo de aprendizaje automático DataRank es capaz de ‘aprender’ lo que es importante para los usuarios, detectando consultas frecuentes o ámbitos importantes de la empresa. Con ello, es capaz de realizar sugerencias de búsquedas frecuentes, errores de escritura y sinónimos y tendencias personales del usuario. Así pues, ofrece un servicio dinámico, personalizado y predictivo.

Destaca en ella la herramienta SpotIQ, la cual a través de IA y el motor de búsqueda in-memory creado desde cero para ThoughtSpot realiza análisis automáticos de los datos. Es capaz de autodetectar anomalías y outliers a través de complejos algoritmos, encontrar correlaciones entre datos y mostrar tendencias que no se habían detectado en amplios conjuntos de datos no purgados (Thoughtspot, 2019). Además, el propio SpotIQ, con el algoritmo de aprendizaje automático DataRank es capaz de ‘aprender’ lo que es importante para los usuarios, detectando consultas frecuentes o ámbitos importantes de la empresa. Con ello, es capaz de realizar sugerencias de búsquedas frecuentes, errores de escritura y sinónimos y tendencias personales del usuario. Así pues, ofrece un servicio dinámico, personalizado y predictivo.

Por otra parte, en la última versión de ThoughtSpot 5, ha sido lanzado el novedoso módulo SearchIQ. Gracias a él se pueden realizar búsquedas y análisis de datos por voz con lenguaje natural. Fue especialmente pensado para dispositivos móviles y cuando no se tiene acceso a un ordenador.

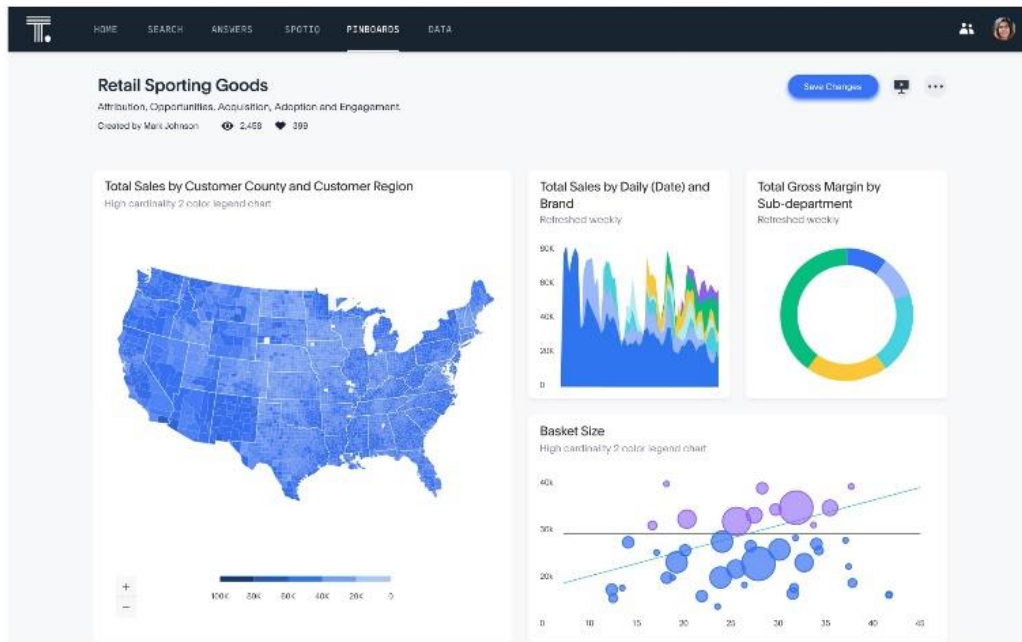


Ilustración 24. Interfaz ThoughtSpot (1) (recuperado de: <https://www.thoughtspot.com/product>)

A diferencia de otras herramientas, se puede estar seguro de la bondad de los resultados y análisis que deseamos. Search Inspector nos ofrece la opción de observar el proceso de cómo ha conseguido los datos, a través de las operaciones y relaciones que ha establecido entre tablas y conjuntos de datos contenidos en la memoria.

Por último, ThoughtSpot también dispone de análisis integrado (incrustado) en otras aplicaciones, webs o procesos de trabajo. La plataforma ThoughtSpot Embedded Analytics permite proporcionar y mostrar datos y análisis a terceros (teniendo el control sobre qué puede ser visto) con amplio abanico de posibilidades y personalizaciones sin la necesidad de inversiones adicionales ni pagos de licencias por usuarios. Tanto las búsquedas con el motor relacional, como gráficos y cuadros de mando pueden ser integrados y ofrecidos a individuos externos. Además, también es posible crear aplicaciones y portales personales a través de ThoughtSpot Data APIs.

La gran marca de diferenciación de ThoughtSpot es su modelo de precios. A diferencia de todos y cada uno de sus competidores, ThoughtSpot no basa su modelo en licencias ni usuarios, si no en dos versiones Enterprise y Enterprise Extended sin límite de usuarios variando su precio en función de la capacidad de la memoria (250GB mínimo).

Gartner, por primera vez hasta ahora, sitúa a ThoughtSpot como un “Líder”, debido particularmente por la satisfacción de sus clientes, rápida innovación y buena ejecución.

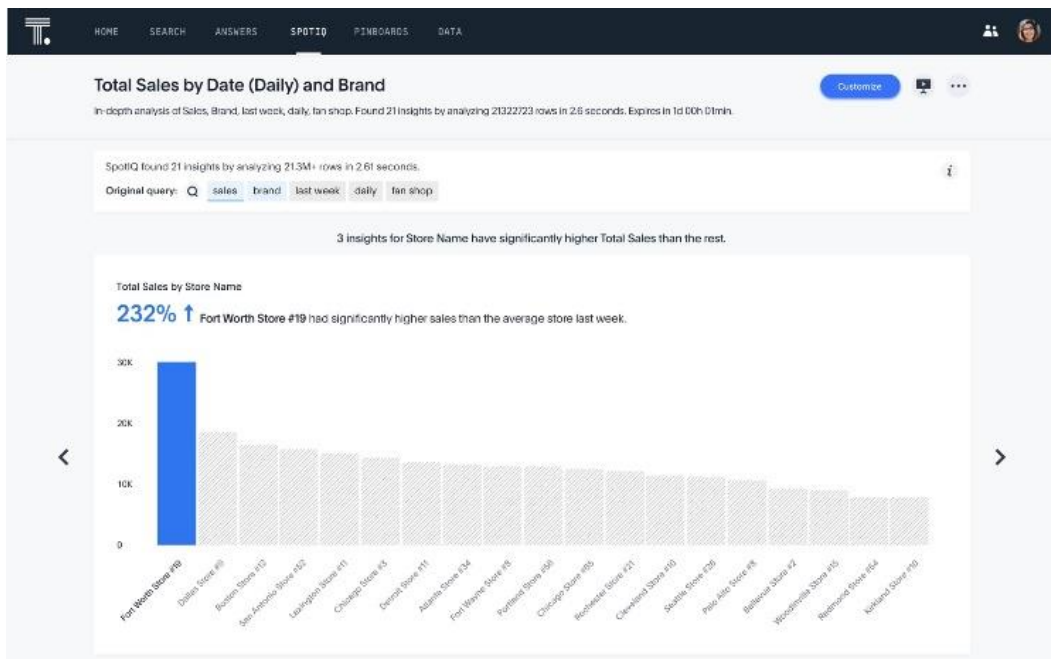


Ilustración 25. Interfaz ThoughtSpot (2) (recuperado de: <https://www.thoughtspot.com/product>)

Ventajas

- ✓ Facilidad de uso para gente poco experimentada. La gran ventaja de ThoughtSpot es su sistema de búsqueda basado en NLQ, lo que permite que hasta personas que jamás han usado BI puedan utilizarlo (Gartner Inc., 2019).
- ✓ Modelo de precios. Beneficioso para empresas grandes.
- ✓ Innovación y tecnología. Puede considerarse que los productos que ofrece ThoughtSpot son tecnológicamente de los más avanzados.

Desventajas

- ✓ Atraso respecto de sus competidores en preparación de datos, exploración visual y cuadros de mando. Preparación de datos se ha de realizar a través de herramientas externas. Carece de ciertos gráficos y opciones visuales comunes. No permite manipulación de datos en tiempo real. Los datos han de ser cargados en la memoria (MPP).
- ✓ Ausencia de contenido específico prediseñado. Las aplicaciones para casos específicos (por sector, industria) han de ser creados por el propio cliente.
- ✓ Escasa disponibilidad geográfica. Tiene pocos 'partners' y contratos de colaboración de empresas externas. Está directamente presente en pocos países.

3.5 Comparativa

El proceso de elección de la herramienta de BI a utilizar por una organización es crítico. Se ha de estudiar el propósito de ella, las funcionalidades que se exigen y el presupuesto disponible para su implementación. Cualquier proyecto de Transformación Digital incurrirá en un gran esfuerzo monetario y humano para conseguir su objetivo. Dicho esto, la elección del sistema es una de las primeras fases



del proyecto. Además, se debe considerar la capacidad de adaptación y flexibilidad de las herramientas para personalizaciones a nuestro modelo de negocio.

En un intento de mejora de la toma de esta decisión, en el presente Trabajo de Fin de Grado se realiza en las Tablas 7 y 8 una comparativa de las principales herramientas citadas. A continuación detallaremos los criterios de comparación que hemos elegido como más importantes:

1. **Versiones.** Se exponen las versiones existentes de cada uno de los proveedores.
2. **Motor de datos.** Es importante conocer qué tipo de motor se utiliza (es decir, qué tipo de bases de datos) para saber cuál se adecúa mejor a nuestros propósitos.
3. **Implementación.** Tal y como se ha comentado a lo largo del capítulo existen las implementaciones on-premise (en servidores de la empresa) o en Cloud (en la nube), teniendo ambas sus ventajas y desventajas.
4. **Solución all-in-one.** Determina si mediante la adquisición del producto, obtendremos una herramienta de BI lo suficientemente completa como para realizar análisis generales de la organización, sin necesidad de software complementario del proveedor.
5. **NLQ (Natural Language Query).** La capacidad de hacer preguntas al sistema en lenguaje natural es muy importante para los empleados con escasa familiarización con estas herramientas. Debemos determinar qué tipos de usuarios la manejarán.
6. **Búsqueda por voz en lenguaje natural.** Complementario a lo anterior, es una de las características más novedosas de este software.
7. **Precio.** Aspecto crítico y que se ha de conocer. Bien es cierto, que algunos de los proveedores no ofrecen los precios públicamente.
8. **Versión móvil.** En la actualidad, cualquier persona tiene a mano un Smartphone o dispositivo electrónico con capacidad de instalar aplicaciones. Es un bonus para todos los usuarios que han de viajar o no siempre tienen un ordenador a mano.
9. **Fuentes de datos compatibles.** Debemos conocer si nuestros datos actuales son compatibles con el software, ya que en caso contrario, necesitaremos de un proyecto de conversión (esfuerzo y dinero).
10. **Actualizaciones.** Vivimos en un mundo dinámico y cambiante, por lo que es necesario conocer si nuestro software ofrece actualizaciones periódicas incluyendo nuevas funcionalidades

Características	POWER BI	TABLEAU	QLIK	THOUGHTSPOT
Versiones	Desktop, PRO, Premium, Mobile y Embedded	Tableau Prep, Tableau Desktop, Tableau Server y Tableau Online	Qlik Sense Enterprise, Qlik Sense Cloud, Qlik Analytics Platform, Qlik View, Qlik NPrinting, Qlik GeoAnalytics, Qlik Associative Big Data Index, Qlik DataMarket, Qlik Data	ThoughtSpot Enterprise, ThoughtSpot Enterprise Extended
Motor de datos	Power Query (Heredado de Excel)	Hyper (In-memory de alta velocidad)	QIX Indexación asociativa (In-memory)	Relational Search Engine in columns (MPP In-memory)
Implementación	On-premise, Cloud (Azure)	On-premise, Cloud	On-premise, Cloud	On-premise, Cloud
Solución all-in-one	<p>✗ (Versiones básicas)</p> <p>✓ (Power BI Premium)</p>	<p>✓</p> <p>(Tableau Prep no es indispensable)</p>	✗	✓
NLQ (Natural Language Query)	✓	✗	✗	✓
Búsqueda por voz con lenguaje natural	✓	✗	✗	✓
Precio (Mayo 2019)	<p>Desktop: gratuito</p> <p>Pro: 8,40€ lic/mes</p> <p>Premium: por capacidad (nodo/mes)</p>	<p>Tableau Server (TS), Tableau Online (TO) por lic/mes:</p> <p>Creator: 70 \$ TO/TS</p> <p>Explorer: 35 \$ TS o 42 \$ TO</p>	A CONSULTAR PARA CADA HERRAMIENTA	Exclusivamente por capacidad (consultar): Enterprise Extended Enterprise

Tabla 7. Comparativa herramientas BI (1). Fuente: elaboración propia

Características	POWER BI	TABLEAU	QLIK	THOUGHTSPOT
Versión móvil (Android, iOS)	✓	✓	✓	✗
Fuentes de datos compatibles	+100 fuentes (compatibilidad completa con software Microsoft)	84 fuentes (algunas solo disponibles en Windows)	91 fuentes (nativos o con conectores)	12 fuentes nativas (posibilidad de más mediante APIs)
Actualizaciones	Mensuales	Casi mensuales	Casi mensuales (en productos importantes)	Cada varios meses (actualizaciones importantes)

Tabla 8. Comparativa herramientas BI (2). Fuente: elaboración propia

3.6 Conclusiones

A modo de resumen, podemos utilizar el método de clasificación gráfica utilizado por Gartner, denominado como “*Magic Quadrant*”, en el que se representan en el eje X la adecuación de su visión (la capacidad de anticiparse y adaptarse a la demanda del mercado) y en el eje Y la capacidad de ejecución (las prestaciones y la capacidad de comercializar su producto). La Ilustración 26 sirve a efectos ilustrativos de la evolución de un año a otro de las herramientas analizadas en los informes de Gartner.

Observamos cómo ThoughtSpot ha conseguido colocarse en el cuadrante de los “Líderes” desde el de los “Visionarios” de un año para otro. Las otras 3 herramientas han mantenido su puesto. Sin embargo, Tableau, la cual en el año 2018 parecía estar a la altura de Microsoft Power BI en Capacidad de Ejecución, sin embargo, se ha alejado de él y se ha quedado atrás.

Resumiendo:

- **La opción más robusta y completa es MS Power BI.** La compatibilidad, además de sus amplias funcionalidades (a pesar de no ser siempre las más innovadoras) junto con su precio, hacen que se convierta en la herramienta más adecuada para cualquier escenario.
- La opción más innovadora y con mayor proyección de futuro es ThoughtSpot. Sin embargo, la falta mucho camino por recorrer para alcanzar a los tres grandes, sobre todo en aspectos básicos como preparación de datos y exploración visual.
- Tableau es en términos generales una muy buena opción, y destaca por su facilidad de uso y el poco conocimiento que se necesita para usarla. Además, reúne todas las funcionalidades necesarias para el flujo de trabajo en una (se puede usar Tableau Prep para preparación de datos asistida y mejorada, pero no es estrictamente necesaria).
- Qlik es también una opción muy interesante, sobre todo a nivel tecnológico. Sin embargo, la multitud y diversidad de aplicaciones segmentadas que ofrece, además de la poca información ‘oficial’ que ofrece sobre la funcionalidad de estas, junto a ninguna información de sus precios, hacen que no sea tan atractiva como otras.

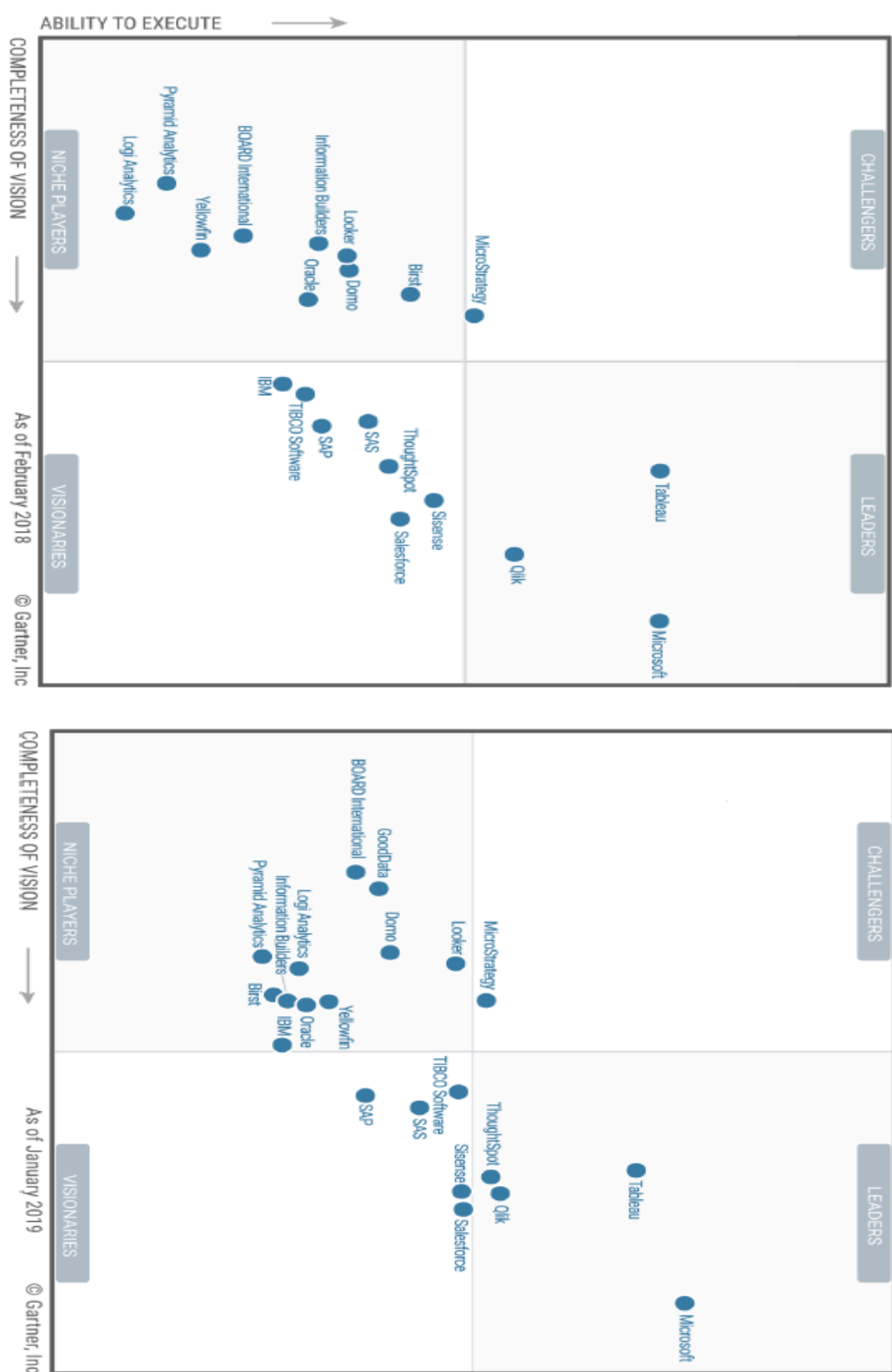


Ilustración 26. Magic Quadrant 2018 y 2019. Fuente: Gartner Inc. (2018), Gartner Inc. (2019)



4. Descripción del caso de estudio

El marco teórico, extenso, propuesto hasta este momento sirve como base para el caso práctico propuesto a continuación. A efecto ejemplificador de todo ello, durante las siguientes páginas se propone una implementación de Business Intelligence en una empresa de carácter industrial (manufactura o producción).

Por una parte, se pretende la creación de una herramienta de soporte a la actividad diaria rutinaria industrial, de manera que sirva como refuerzo para aprovecharse de las fortalezas de la organización y solventar sus debilidades. Dicha herramienta, analizará los datos estructurados provenientes de la actividad de los diferentes departamentos a partir de una base de datos propia.

Por otra, se propone un análisis más novedoso e innovador. En los últimos años, con la expansión del concepto Big Data ha surgido la necesidad de analizar datos no estructurados. Con ello, en la siguiente exposición del caso práctico se realizará un análisis de datos no estructurados, los cuales, en este caso, serán mensajes de redes sociales destinados a la compañía y mensajes internos de trabajadores. Más conocido como análisis de sentimientos, a través de él conoceremos la valoración externa de los clientes y el grado de satisfacción de los empleados.

Para describir el caso, en primer lugar, hemos de delimitar las características de la empresa en particular. En cualquier proyecto de adopción de BI por parte de una empresa es fundamental comprender el entorno interno de la empresa para ofrecer una solución que se adopte a las características y particularidades de ésta. Es decir, las soluciones han de ser lo más únicas posibles, adaptadas a los propósitos de la organización, de sus objetivos y de sus capacidades. Los resultados dependerán de comprender el funcionamiento y procesos internos.

Consideramos la empresa EJEMPLO S.A, una PYME pequeña (entre 10 y 50 empleados) con sede en Valladolid, y un periodo de actividad delimitado entre el 1 de febrero de 2019 y el 17 de mayo de 2019 (aproximadamente, 4 meses), dedicada a la fabricación y manufactura de tres productos en sus instalaciones. Los productos en cuestión son los productos A, B y C.

Estos productos, son fabricados en dos líneas existentes de manera que los productos A y B son producidos en la Línea 1 y el producto C en la Línea 2, sin posibilidad de fabricación de A y B en la Línea 2 ni C en la Línea 1. Por otra parte, también se cuenta con cinco operarios (Operario 1, Operario 2, etc.), de los cuales, los operarios 1 a 3 se dedican exclusivamente a la Línea 1 y los Operarios 4 y 5 a la Línea 2. Estos **operarios se dedican de manera única y exclusiva a un pedido** cuando se le es asignado, de manera que un pedido u orden es realizado por un operario de principio a fin. Se ha de destacar las paradas (planificadas o no) en las líneas de producción y que son trascendentes en los planes de producción y los plazos. Tanto para los Operarios como para las Líneas se consideran unos salarios por hora y unos gastos por hora, respectivamente.

Los productos A, B y C fabricados en sendas líneas, a su vez, están conformados de tres materias primas (MP) X, Y y Z. Estas MP son adquiridas externamente a cuatro



proveedores (Proveedor 1, Proveedor 2, etc.), de manera que es posible la adquisición de las tres a cualquiera de los cuatro proveedores con diferencias en los precios y tiempos de suministro.

Las MP poseen un almacén en el que son inventariadas hasta la emisión de una orden de producción, momento en el cual quedan comprometidas y desaparecen del inventario. El almacén cuenta con un Inventario Inicial, Stock de Seguridad (SS) y una Capacidad de Almacenamiento para cada una de las MP. Estos datos, y otros relevantes en cuanto a la producción serán desvelados más adelante.

Los Productos, cuentan con un almacén las mismas características donde son inventariadas antes de su venta.

La venta de los productos cuenta con cuatro canales posibles de distribución: Comercial de ventas, pedidos de establecimientos de Grandes Superficies, Minoristas o Venta Directa en las instalaciones de la compañía.

4.1 Proceso detallado

En las anteriores líneas se ha descrito cualitativamente la actividad de la empresa EJEMPLO S.A. Sin embargo, debemos concretar los datos específicos de cada ámbito para realizar un mejor análisis.

En primer lugar, comenzando en el principio del proceso productivo de la empresa, se dispone de cuatro proveedores. Los pedidos (órdenes de compra) son exclusivamente de una MP, considerando un identificador para cada una de las compras. La Tabla 9 muestra los datos de cada uno de los proveedores.

Nombre	LeadTime (horas)	CosteX/ud	CosteY/ud	CosteZ/ud
prov1	48	4	3	4
prov2	60	2	5	4
prov3	24	3	5	3
prov4	72	6	2	4

Tabla 9. Datos de proveedores

En la fabricación de los Productos, se ha citado la necesidad de MP para la fabricación de éstos (Tabla 10).

Producto	mat_prima_X	mat_prima_Y	mat_prima_Z
A	2	0	1
B	0	2	0
C	1	1	1

Tabla 10. Lista de materiales de productos

A partir de la tabla se infiere que para cada producto A son necesarias 2 unidades de X, 0 de Y y 1 de Z. De igual manera se definen los demás.

En cuanto a los Operarios de la compañía y las Líneas, la Tabla 11 y 12 muestran los datos relativos a éstos.



Nombre	Línea	Salario por hora
OP1	L1	10
OP2	L1	10
OP3	L1	10
OP4	L2	12
OP5	L2	12

Tabla 11. Datos de operarios

Nombre línea	Coste por hora
L1	60
L2	40

Tabla 12. Datos de líneas

Por otra parte, las Paradas de las líneas se nombran mediante un identificador conocido por los trabajadores y tienen un tiempo estimado de parada (Tabla 13).

Código de parada	Descripción	Tiempo estimado en horas
par01	Mant. Programado	0,5
par02	Preparación de la línea	1
par03	Accidente leve	0,25
par04	Accidente grave	3
par05	Modernización y cambios	18

Tabla 13. Datos de paradas de línea

Por último, los datos sobre Productos (añadidos a los anteriores) y MP restantes y necesarios para el análisis se muestran en la Ilustración 27.

Acompañaremos a estos datos de un esquema simplificado de los procesos internos considerados (Ilustración 28).

4.2 Estado de la compañía

En el anterior apartado se ha definido el flujo de trabajo interno para la obtención de los productos que fabrica la compañía. Aunque este proceso es elemental a la hora de presentar una solución de BI, también es necesario considerar los recursos actuales de la compañía en materia de BI, es decir, su estado de digitalización o las capacidades reales y actuales de ésta.

Se considera, en este aspecto, a la compañía EJEMPLO S.A como una compañía completamente aislada del mundo de BI, en la que existen métodos rudimentarios de control (OLTP) y en la que no existe ningún tipo de solución de BI. Basándonos en los modelos de madurez anteriormente citados podemos situarla en las fases iniciales.



Nombre producto	mat_prima_X	mat_prima_Y	mat_prima_Z	Precio de venta	Tiempo de fabricación (min)	Línea	Capacidad almacén	SS	
A	2	0	1	55	25	L1		400	40
B	0	2	0	59	30	L1		400	40
C	1	1	1	65	50	L2		400	40

Nombre MP	Invinicial	SS	Capacidad Almacén
X	50	80	1000
Y	30	80	1000
Z	0	80	1000

Ilustración 27. Datos de Productos y MP.

1 Febrero 2019 – 17 Mayo 2019

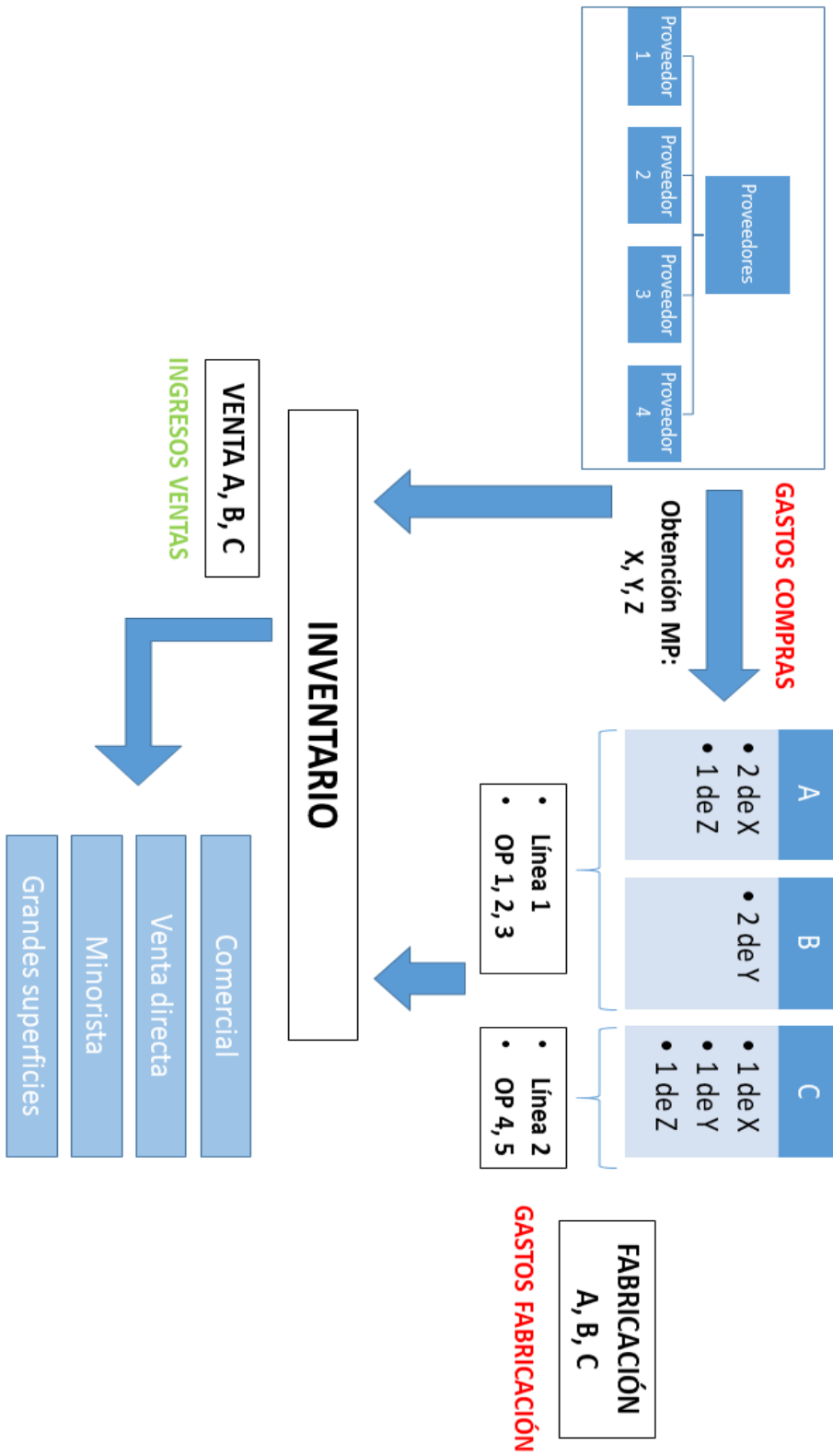


Ilustración 28. Esquema simplificado del flujo de trabajo. Elaboración propia



BUSINESS INTELLIGENCE MATURITY MODEL (BIMM)

Fase 1. No existe BI. Caracterizado por la utilización de sistemas transaccionales (OLTP) en ciertos sectores de la organización. Decisiones basadas en intuición y experiencia. No existen herramientas y los datos están dispersos e inaccesibles.

FORRESTER BI MATURITY MODEL

Nivel 1. Static Intelligence. Las fuentes de datos están predefinidas, son escasas y están obsoletas, al igual que los datos que contiene. El acceso a los datos está restringido a la propia organización.

BUSINESS INFORMATION MATURITY MODEL

*Fase 1. Este primer escenario se puede entender como la anteposición del uso de los datos antes que su almacenamiento (el DW no es conocido ni usado). Los usuarios proporcionan al departamento de informática un "listado" de información que necesitan, por lo que el departamento centra sus esfuerzos en **qué** les gustaría recibir o qué es lo que quieren.*

GARTNER'S MATURITY MODEL FOR BI AND PERFORMANCE MANAGEMENT

Nivel 1. Unaware. A modo de resumen, se llevan a cabo multitud de malas prácticas. Los datos, en términos generales, son inconsistentes y mal interpretados. Las demandas individuales de información dictan los cambios. Destaca por el uso de Excel y falta de herramientas de reporting, además de la inexistencia de métricas y la importancia de éstas.

El departamento de informática es el que tiene en sus manos el presupuesto y la capacidad de gestionar la información y generar informes.

BUSINESS INTELLIGENCE MATURITY HIERARCHY

Fase 1. Data. Esta fase sirve como punto de partida para la introducción del DW y el BI. Se tiene capacidad de recoger, depurar, estandarizar y almacenar datos de diversas fuentes.



Cualquiera de las anteriores definiciones de etapas, e incluso, combinaciones de ellas explican de manera fiel la situación de la empresa.

- ✓ Uso de sistemas transaccionales (OLTP).
- ✓ Decisiones basadas en intuición.
- ✓ Datos poco actualizados y con poca frecuencia actualizados, con orígenes escasos.
- ✓ La escasa información distribuida es bajo encargo del departamento de informática y bajo demanda de lo que se les exige.
- ✓ Sin DataWarehouse.
- ✓ Demandas de información por parte de directiva, escasas métricas, uso de Excel como silos de información.
- ✓ Sin procesos de depurado o estandarización de datos.

Es por ello que se encarga la generación de un sistema de BI que permita a la compañía a alcanzar etapas o fases posteriores de cada uno de los modelos de madurez expuestos.

4.3 Objetivos y requisitos del sistema de Business Intelligence

Un proyecto dedicado a la implementación de BI en el seno de una organización requiere también de actividades conjuntas. La bilateralidad de las comunicaciones es un aspecto clave en cualquier proyecto, y especialmente en este tipo. En una sesión conjunta de primera toma de contacto se ha de escuchar la idea y concepto inicial de la compañía de qué esperan que reúna la herramienta y su visión de ella. Han de establecerse ciertos vínculos entre ambos equipos para permitir que el proyecto se desarrolle fluida y satisfactoriamente.

En lo relativo a EJEMPLO S.A, se han determinado los siguientes requisitos y objetivos en relación a la solución implementada:

- ✓ Creación y puesta en marcha de un Data Warehouse para la compañía, a través de la cual se gestionarán los datos corporativos.
- ✓ Integración del Data Warehouse con la herramienta seleccionada para el análisis de los datos.
- ✓ La elección de la herramienta será nuestra decisión y ha de estar justificada y englobada en el presupuesto disponible para el proyecto.
- ✓ La interfaz ha de mostrar aspectos clave de la empresa como: inventarios, compras, producción y ventas.
- ✓ En cuanto a los inventarios, el requisito mínimo es que se ha de poder visualizar la evolución del estado de inventarios de cada uno de los productos y de MP.
- ✓ Para las compras, se exige poder ver el resumen mensual de gastos por proveedor y por producto. Además, es de sobra conocido que algunos de los pedidos contienen piezas defectuosas, por lo que se considerará indispensable poder conocer qué proveedores ofrecen mayor número de piezas defectuosas y cuántos de los pedidos los han contenido. A su vez, queremos comprobar la calidad del tiempo de suministro ofrecido por contrato con los proveedores, por lo que se deberá de ofrecer una manera de comparar el Lead Time ofrecido por contrato y el real.



- ✓ Para la producción, por una parte tenemos el proceso específicamente productivo. La directiva desea tener a mano un resumen de la cantidad de productos fabricados y su coste, las no-calidades por cada uno de los productos e información referente a los tiempos de producción. También, por experiencia, los trabajadores han mostrado su preocupación por la utilización en mayor medida que la calculada de las MP para la fabricación de los Productos. Este desajuste deberá poder ser observado mediante algún grafismo para determinar sus posibles causas.
Por otro lado, se desea conocer la productividad de los trabajadores, el funcionamiento de las líneas o información relevante en cuanto a las paradas en las líneas. Deben permitir encontrar correlaciones o indicios sobre las causas raíz de problemáticas.
- ✓ Por último, para las ventas es esencial poder mostrar el EBIDTA mensual de la compañía, además de algún análisis de los canales de venta y su rentabilidad.

Al tratarse de una empresa sencilla dedicada a un proceso relativamente sencillo y además, poco introducida en el BI, los objetivos y requisitos son bastante genéricos y poco específicos. Sin embargo, son útiles como guía para la realización del proyecto.

4.4 Modelo del sistema

Habiendo conocido la organización, sus procesos y sus requisitos del sistema, procederemos a detallar durante los siguientes capítulos la solución concebida.

4.4.1 Software y arquitectura

En primer lugar, debemos citar la herramienta de BI elegida para la solución. En nuestro caso, será Power BI. La flexibilidad de la herramienta, su alta compatibilidad y la fiabilidad que ofrece un producto de Microsoft, además de ser un producto líder absoluto provoca que su elección sea la óptima para nuestro proyecto. Por otra parte, su bajo coste por usuario es una ventaja muy importante para una empresa local/regional que no está dispuesta ni en condiciones de pagar precios desorbitados por herramientas sobrecapacitadas para sus propósitos. Por ello, elegimos la versión Power BI PRO, con un precio de 8,40€/mes para cada usuario, la cual permite la creación y distribución de informes entre los empleados, además del acceso a la aplicación Power BI Mobile.

En cuanto a la arquitectura del sistema, la envergadura y los propósitos de la empresa juegan un papel muy importante al decidirla. Al tratarse de una PYME pequeña con un proceso de manufactura relativamente sencillo y una digitalización casi inexistente (salvo utilización de ficheros Excel), se propone una arquitectura sencilla que sirva como primer paso hacia una Transformación Digital aceptable, eficaz y escalable. Para ello, utilizaremos como modelo la arquitectura propuesta en el apartado 2.2 (Ilustración 28).

En ella, consideramos como los ficheros Excel ya existentes y los datos operacionales (sensores, planes de producción, etc) como las únicas fuentes de datos. Estas fuentes

serán redirigidas al DW corporativo donde serán depurados y estructurados en la mayor medida posible.

A partir de él, Power BI extraerá los datos, los preparará en la manera oportuna y se servirá de ellos para la creación de gráficos e informes que se necesiten.

Debemos destacar que el presente **Trabajo de Fin de Grado se atañe única y exclusivamente a la parte del análisis de datos**. Es decir, durante el presente texto, no se realizará ninguna explicación ni referencia al modo de creación e implantación del DW ni las conexiones a las fuentes de datos, ya que su complejidad queda fuera de las posibilidades y conocimientos adquiridos durante el Grado.

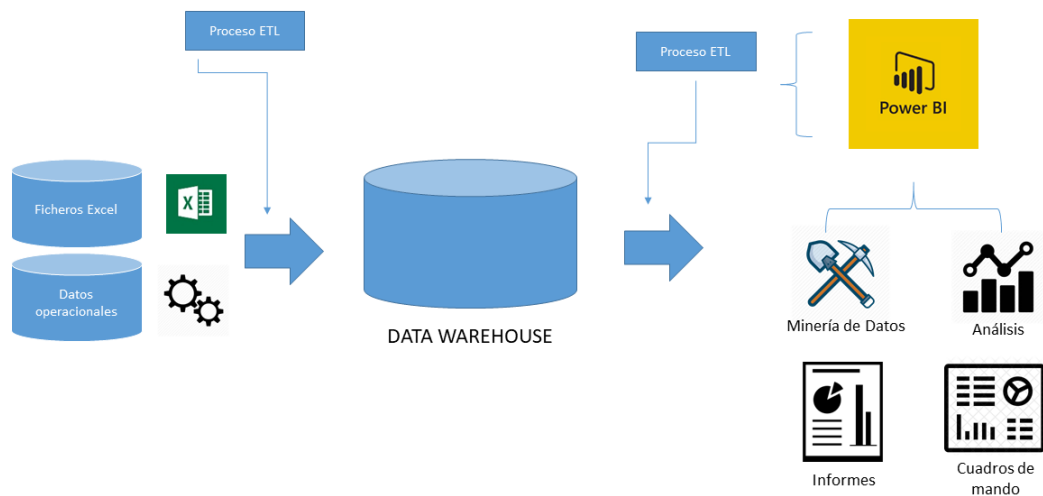


Ilustración 29. Arquitectura de la solución para el caso práctico. Elaboración propia

4.4.2 Obtención de los datos

La empresa EJEMPLO S.A basa su producción y compras en un Plan Maestro de Producción (PMP) (a partir de un Plan Agregado de Producción) y en un Plan de Requerimientos de Material (MRP, “Material Requirements Planning”), respectivamente. En ellos, se planifican las órdenes de producción y las órdenes de compra de MP teniendo en cuenta el máximo número de factores que pueden influir sobre la producción. A tener en cuenta:

- ✓ Plan Maestro de Producción
 - Previsiones a Corto y Medio plazo.
 - Pedidos comprometidos.
 - Pedidos pendientes.
 - Número de trabajadores (mínimo y máximo), posibilidad de subcontratación y horas extras o retrasos en pedidos.
 - Costes de contratación, despidos, salarios, horas extra, horas subcontratadas, etc.
 - Existencias, stock de seguridad.
 - Etcétera.
- ✓ Plan de Requerimientos de Material
 - Plan Maestro de Producción.
 - Lista de materiales.



- Estado de inventarios.
- Demandas independientes.
- Método y tipos de lotes y sus parámetros asociados.
- Aprovechamiento (coeficiente para compensar defectos).
- Etcétera.

Cabe destacar que aunque se puede obtener un Plan Maestro de Producción Propuesto (con el que se cubren las necesidades), las características y parámetros (esencialmente, los tipos de lotes) del MRP son los que determinan los momentos y cantidades de cada uno de los productos y MP.

Considerando que las fechas consideradas para este análisis son desde el 1 de Febrero de 2019 al 17 de Mayo de 2019, la creación de sendos planes se extenderá hasta el 31 de Mayo para redondearlos a cuatro meses (16 semanas). Así pues, ambos planes para 16 semanas se muestran en la Ilustración 31.

Debe notarse que el PMP muestra cómo se satisfacen las necesidades de manera óptima. Sin embargo, los parámetros del MRP (tamaño y método de lote, tiempos de suministros, consideración de otros parámetros para las MP, etc) provocan una disposición diferente pero que también satisface las necesidades.

Teniendo en cuenta esto, nuestros datos estarán basados en el MRP generado. Obtenemos 31 órdenes de compra y 34 órdenes de producción. Se considerarán 70 órdenes de ventas que se han sucedido durante este período. Además, considerando que la fecha última es el 17 de mayo, existen dos órdenes de producción en curso.

5. Análisis de datos estructurados

5.1 Datos

Los planes de producción considerados son puestos en marcha y se recopilan los datos necesarios para la creación del panel de mando. Estos datos son introducidos en el DW y posteriormente son extraídos a Power BI. Tras la preparación de los datos pertinentes para la obtención de los paneles deseados, se tienen finalmente cargadas 16 tablas, tal y como muestra la Ilustración 30

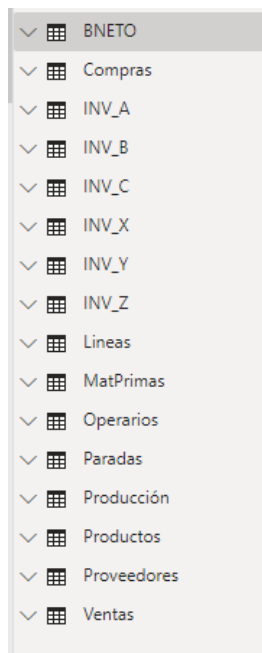


Ilustración 30. Tablas para el análisis de datos en Power BI

Las tablas Líneas, MatPrimas, Operarios, Paradas, Productos y Proveedores contienen información relativa a cada uno de los términos. Es decir, contienen la información mostrada en el capítulo anterior.

En cuanto a las demás, éstas contienen la información necesaria para el análisis. Las tablas Compras, Producción y Ventas ordenan la información por cada una de las órdenes, de manera que reúnen la siguiente información:

1. **Compras.** Muestra para cada “orden” el “proveedor” al que se realizó el pedido, la “materia prima”, la “cantidad” de materia prima, el “coste” del pedido, las “fechas de emisión y entrega”, el “lead time real”, los “defectos” en el pedido y **otras columnas auxiliares** que serán utilizadas como por ejemplo una de carácter binario para conocer si el pedido ha llegado a tiempo, el *lead time* teórico, etc. (Ilustración 32).



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
PMP A	60	60	60	0	60	60	60	60	0	60	0	60	0	60	60	60
PMP B	50	50	50	0	50	50	0	50	50	0	50	50	50	0	50	50
PMP C	40	40	0	0	40	40	0	40	0	40	0	40	40	40	0	40

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	50	75	50	0	75	50	50	75	0	50	0	75	0	75	75	50
B	70	35	35	0	70	35	0	70	35	0	70	35	70	0	70	50
C	60	30	0	0	45	45	0	30	0	45	0	45	45	45	0	45
X	202	58	0	172	172	128	172	0	172	0	172	0	172	0	172	0
Y	80	110	0	90	110	0	100	110	0	110	110	0	0	90	110	0
Z	133	64	0	108	108	64	108	0	108	0	108	108	108	0	0	108

Ilustración 31. PMP y MRP propuestos



2. **Producción.** Como en Compras, cada “orden” tiene el “producto” y “cantidad” asociados, además de las “fechas de emisión y entrega” y su “coste”. Además, se han incluido “materias primas reales” utilizadas para la fabricación de cada orden para mostrar los desajustes ampliamente conocidos. Por otra parte, se ha incluido la información relativa al “operario” encargado de la realización de la orden, la “línea” en la que se ha realizado, la “duración real” de la producción, las “no calidades” en el pedido y si ha habido alguna “parada” durante el pedido y de qué tipo ha sido. También se han incluido una gran variedad de columnas auxiliares (creadas en la preparación de los datos) como por ejemplo columnas que muestran las materias primas en teoría necesarias, la duración teórica, el desajuste producido, el tiempo de fabricación unitario, las unidades por hora, si ha habido retraso o no, etc. (Ilustración 33)
3. **Ventas.** Al igual que en las anteriores, las “órdenes” tienen el “producto” vendido asociado, la “cantidad” vendida, el “canal de venta” y el “ingreso” que supone. Además, se incluye la “fecha de emisión o salida del almacén” (Ilustración 34).

Estas tres tablas son de importancia primordial, ya que **el resto de tablas no citadas hasta ahora son derivadas de ellas tres.**

La tabla BNETO reúne las “órdenes de compra”, “órdenes de producción” y las “órdenes de ventas”, con sus “fechas de emisión”, las “cantidades” y el “coste o ingreso” que supone para facilitar la visualización del EBITDA.

Por otra parte, las de inventario de productos A, B y C, reúne las “órdenes de producción” y las “órdenes de venta” con sus “productos”, “cantidades” y “fechas de entrega” y “fecha de emisión”, respectivamente, para considerar cada una de ellas como entradas y salidas.

Lo mismo ocurre con las tablas de inventarios de X, Y y Z, reuniendo las “órdenes de compra” y las “órdenes de producción”, con sus “materias primas”, “cantidades” y “fechas de entrega” y “fecha de emisión”, respectivamente, para considerar las entradas y salidas de inventario.

5.2 Consideraciones generales del análisis

Los representantes de la empresa EJEMPLO S.A exigieron poder visualizar los datos de 4 ámbitos esenciales para ellos: compras, producción, inventarios y ventas. En ello es lo que basaremos nuestra solución en Power BI. Se plantea la creación de seis páginas en Power BI: **Estado General, Inventarios MP, Inventarios Productos, Compras de MP, Producción y Operarios y Líneas de Producción.**

La división en dos páginas de inventarios es por mera cuestión de espacio, ya que es difícil reunir la información de seis ítems en una sola página. Por otra parte, se ha separado Operarios y Líneas de Producción de la propia Producción porque es un ámbito del que se puede obtener gran cantidad de información relativa a correlación entre líneas y operarios y factores de producción críticos. Por último, es evidente la ausencia de una página dedicada a Ventas, sin embargo, se ha incluido dicha información en la **página Estado General**, ubicada en el primer espacio, **dedicada especialmente a directivos** que su máxima preocupación son los datos monetarios y es lo que desean ver en primer lugar, independientemente del resto de información.



id_compra	id_MP	mat_primaria	id_prov	Proveedor	Cantidad	Coste	f_emision_compra	f_entrega_compra	lead_time_teo	lead_time_real	defecto	estado_X
C1	1	X	3	prov3	202	606	viernes, 1 de febrero de 2019	lunes, 4 de febrero de 2019	24	48	1	201
C2	2	Y	3	prov3	80	400	viernes, 1 de febrero de 2019	lunes, 4 de febrero de 2019	24	74	3	
C3	3	Z	3	prov3	133	399	viernes, 1 de febrero de 2019	lunes, 4 de febrero de 2019	24	62	0	
C4	1	X	1	prov1	58	232	lunes, 4 de febrero de 2019	jueves, 7 de febrero de 2019	48	46	3	55
C5	2	Y	2	prov2	110	550	lunes, 4 de febrero de 2019	jueves, 7 de febrero de 2019	60	63	1	

estado_Y	estado_Z	Presupuesto	Atiempo
77		6000	1
	133	6000	1
		6000	0
109		6000	1

Ilustración 32. Tabla Compras

id_orden	id_producto	nombre_producto	cantidad	f_emision_orden	f_entrega_orden	mat_prima_X_teo	mat_prima_X_real	mat_prima_Y_teo
P1	1 A		50	miércoles, 5 de febrero de 2019	jueves, 7 de febrero de 2019	100	100	0
P2	2 B		70	jueves, 7 de febrero de 2019	lunes, 11 de febrero de 2019	0	0	140
P3	3 C		60	miércoles, 5 de febrero de 2019	lunes, 11 de febrero de 2019	60	60	60
P4	1 A		75	lunes, 11 de febrero de 2019	jueves, 14 de febrero de 2019	150	152	0
P5	2 B		35	jueves, 14 de febrero de 2019	lunes, 18 de febrero de 2019	0	0	70

mat_prima_Y_real	mat_prima_Z_teo	mat_prima_Z_real	linea	operario	duracion_teorica_h	duracion_real_h	id_parada	nombre_parada	tiempo_real_par	no_calidad
0	50	51 L1	L1	OP1	20,8333333333333	21	No parada			0
142	0	0 L1	L1	OP2		35	par02	Preparación de la línea	0,75	0
61	60	60 L2	L2	OP4		49	No parada			3
0	75	76 L1	L1	OP2		31,25	24 par01	Mant. Programado	0,5	0
70	0	0 L1	L1	OP1		17,5	No parada			1

coste_orden	estado_A	estado_B	estado_C	TabUnitario	X_desajuste	Y_desajuste	Z_desajuste	UdPorHora	Retraso
1458,33333333333	50			25,2	0	0	1	0,0396825396825397	1
2450		70		30	0	2	0	0,0333333333333333	0
2600			57	49	0	1	0	0,0204081632653061	0
2187,5	75			19,2	2	0	1	0,0520833333333333	0
1225		34		34,2857142857143	0	0	0	0,0291666666666667	1

Ilustración 33. Tabla Producción



id_venta	prod_venta	cantidad	f_expedicion	canal_venta	ingreso	A	B	C
V1	A	55	miércoles, 6 de febrero de 2019	Comercial	3025	55		
V2	C	30	jueves, 7 de febrero de 2019	Gr. Superficies	1950			30
V3	B	25	sábado, 16 de febrero de 2019	Comercial	1475		25	
V4	B	20	domingo, 17 de febrero de 2019	Minorista	1180		20	
V5	A	60	lunes, 18 de febrero de 2019	Gr. Superficies	3300	60		

Ilustración 34. Tabla Ventas

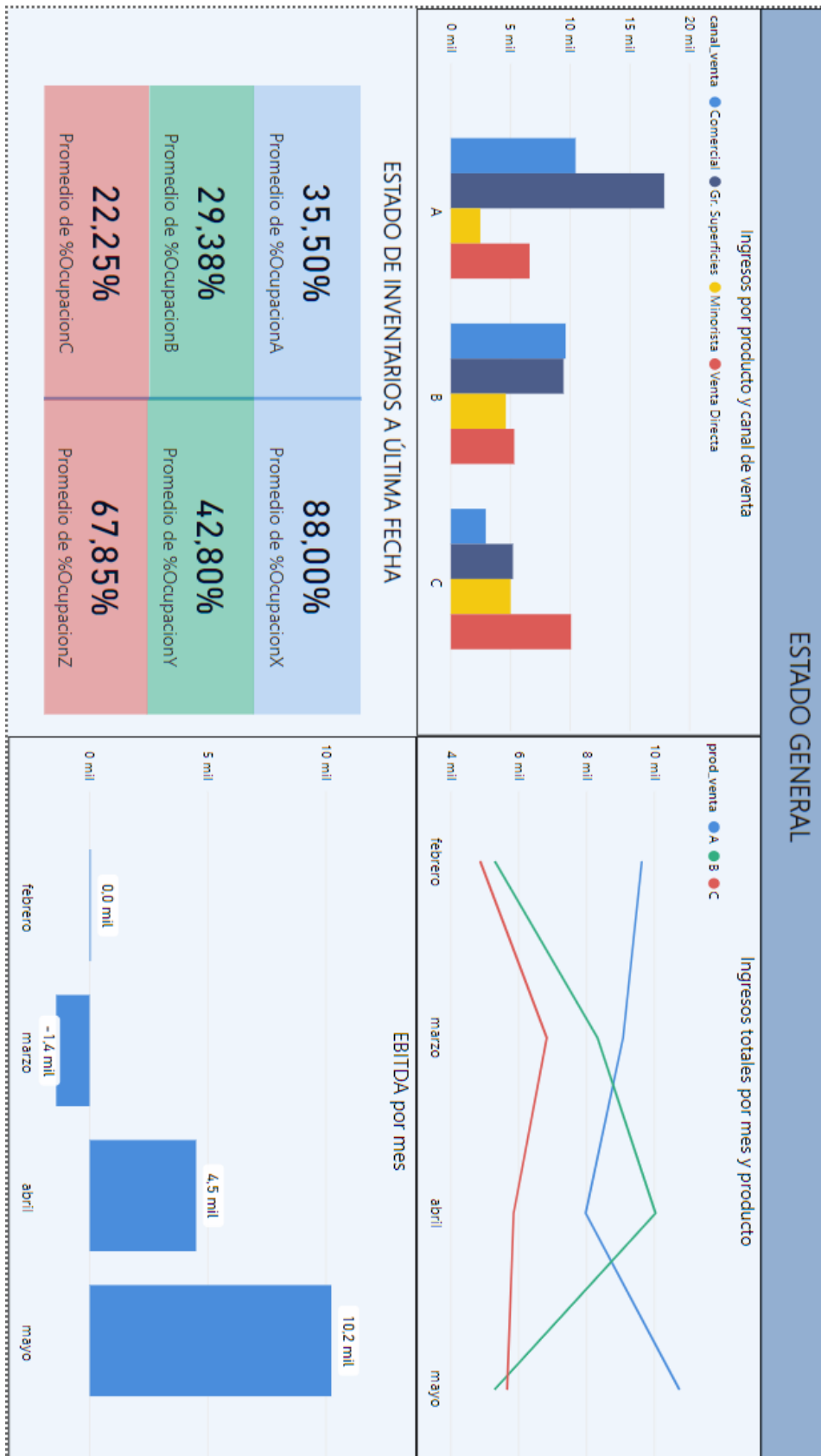


Ilustración 35. Página 1: Estado General



En cada uno de los siguientes epígrafes se procede a realizar una justificación de los gráficos utilizados con un posterior análisis de la información más significativa en ellos.

5.3 Cuadro de mando: ESTADO GENERAL

Como su nombre indica, debe proporcionar una visión del estado de la compañía a la fecha actual. Está especialmente pensada para los responsables y la directiva, para informarse sobre el estado, especialmente financiero, de la empresa. Se ha propuesto la inclusión de:

1. Ingresos por producto y canal de venta. La directiva quiere y pretende conocer cuáles son sus productos y sus canales de ventas más rentables. Por ello, se distribuyen los productos a través de un gráfico de barras y una leyenda por colores para el canal de venta.
2. Evolución de ingresos totales por producto y mes. Al igual que es importante conocer nuestros canales más rentables, es igual de importante conocer nuestros meses más rentables para encontrar las causas de bajadas o subidas, compensar con campañas de marketing en futuras ocasiones, o ajustar la producción en función de las causas para el futuro.
3. Estado de inventarios a última de fecha. Reunir en una sola tabla el estado de los inventarios a la última fecha que hubo un movimiento para informar a los responsables si se necesita un aumento de producción o al contrario, se pueden parar los procesos debido a un sobre inventariado.
4. EBIDTA mensual. Tal vez el gráfico más relevante para un directivo. Muestra el resultado financiero mes a mes de la actividad productiva (ingresos de ventas menos los gastos de producción y compras).

El resultado obtenido con esta estructura propuesta se muestra en la Ilustración 35.

Lo más llamativo, en primer lugar, son los ingresos que genera el producto A en las Grandes Superficies (Ilustración 36). Los ingresos que genera están muy lejos de cualquier otra fuente para cualquier otro producto. Si además, seleccionamos dicha barra, podemos ver la evolución temporal en la interacción en el gráfico adyacente, viendo como de Abril a Mayo duplica sus números (Ilustración 37).

Este descubrimiento, permite a la empresa invertir fuertemente en la promoción de A para mantener o incluso aumentar las ventas de dicho producto en Grandes Superficies, o incluso Comercial, el cual es la segunda fuente de ingresos absoluta.

También es curiosa la popularidad del producto C en Venta Directa en las propias instalaciones de la compañía (Ilustración 36), por lo que podemos intuir que tiene gran demanda por clientes cerca de las instalaciones. Se debe incentivar la compra del producto por Venta Directa ya que es un claro nicho de mercado.

Lo siguiente que llama la atención es el desplome de los ingresos del producto B durante el mes de Mayo y el despunte de los ingresos de A (muy influenciados por las Grandes Superficies, con unos 5250 euros de ingresos de los más de 10000 totales). Si seleccionamos el mes de Mayo vemos la distribución de las ventas según los canales, según se ve en la Ilustración 38.

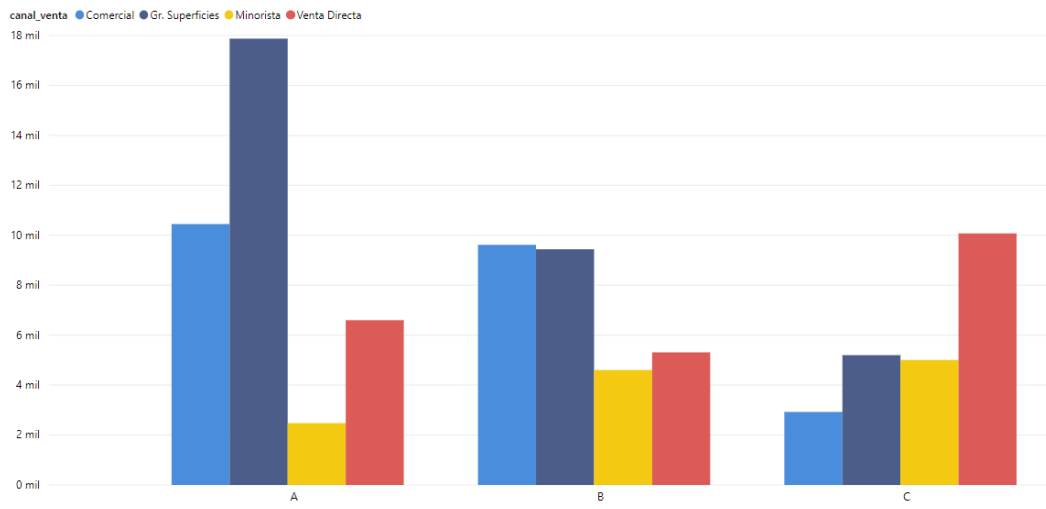


Ilustración 36. Ingresos por canal de venta y producto

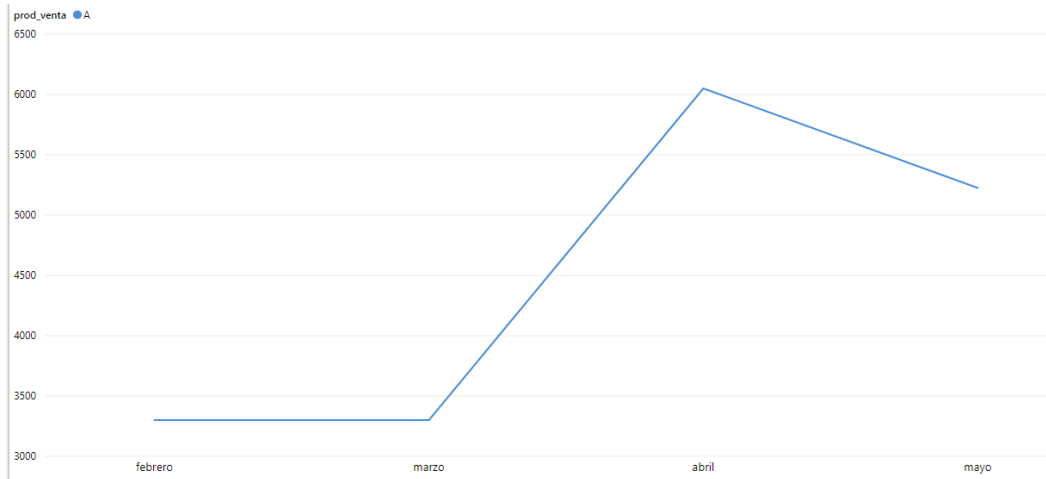


Ilustración 37. Ingresos de A mensuales

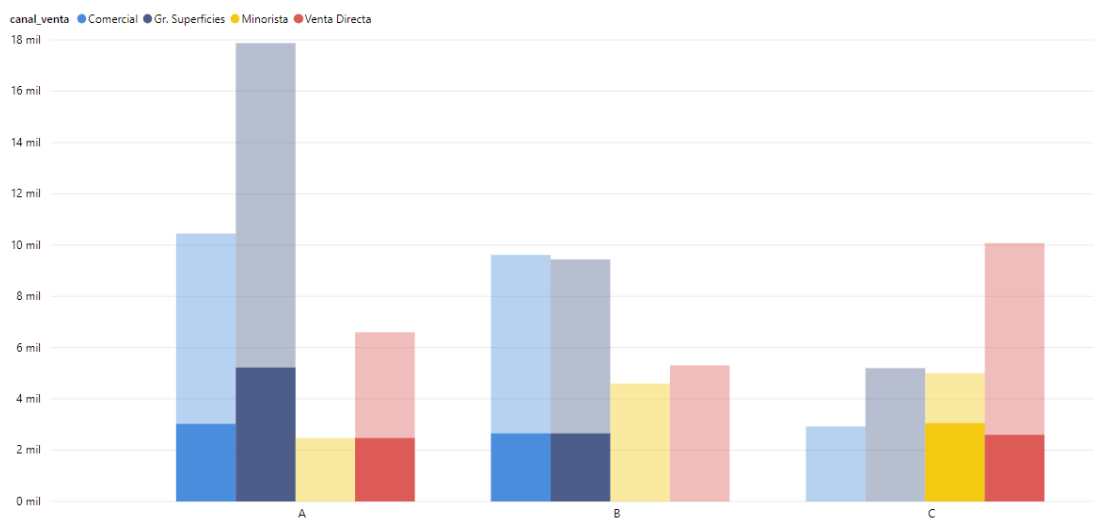


Ilustración 38. Ingresos por canal de venta y producto en Mayo

Y si lo comparamos con Abril en la Ilustración 39, es fácil observar cómo se han perdido por completo las ventas de Venta Directa de un mes a otro (Abril supone un 75% de las ventas totales por Venta Directa) y la mitad de las ventas en Grandes Superficies.

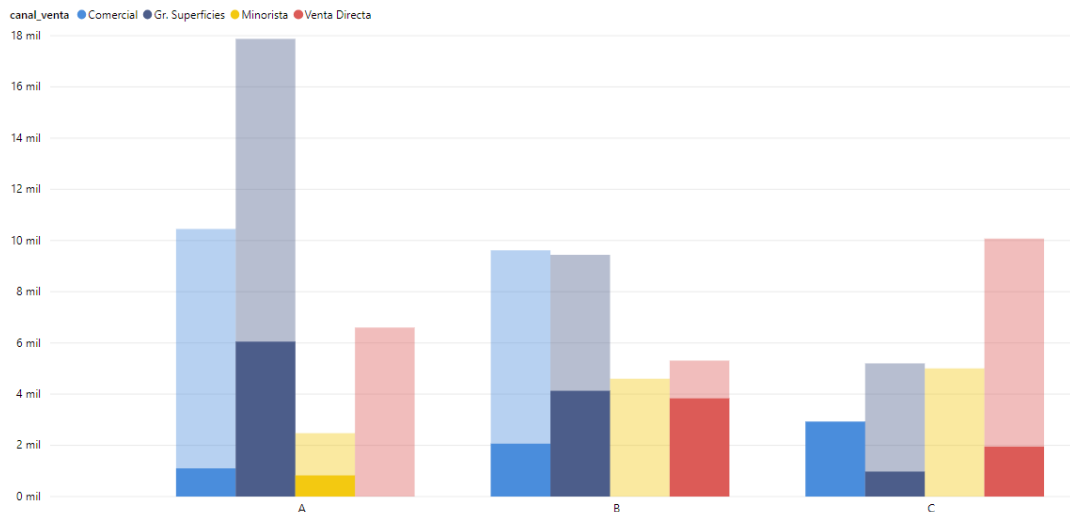


Ilustración 39. Ingresos por canal de venta y producto en Abril

Ya de por sí, es un problema que el 75% de las ventas en un canal sean exclusivamente en un mes, pero también es un problema perderlos en un mes por completo. Deben analizarse las causas y mejorarse las estrategias de marketing para estabilizar las demandas, o incluso reanalizar las estimaciones de la demanda para evitar un desajuste de tal magnitud.

En cuanto a los inventarios a última fecha (Ilustración 40), el dato más llamativo es que el almacén de X está ocupado al 88%, un número excesivamente alto. Deben revisarse los planes de producción y de materiales para revisar si se ha cometido un error ya que implica una gran cantidad de MP sin utilizar.



Ilustración 40. Inventarios a última fecha de movimiento

Por último, en cuanto al EBIDTA mensual, el cual se puede observar en la Ilustración 41, resulta curioso el hecho de que en Marzo el resultado es negativo a pesar de que las ventas fueron menores que en Febrero. Podemos suponer que hubo una gran cantidad de órdenes de producción en dicho mes, por lo que se recomienda la revisión del plan de producción para conseguir uno más homogéneo en futuras ocasiones.

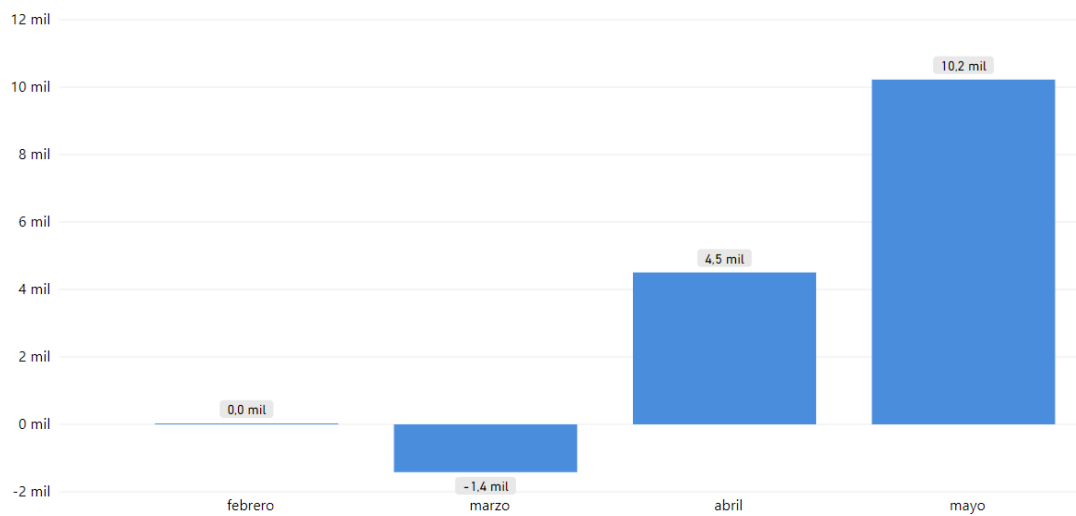


Ilustración 41. EBITDA mensual

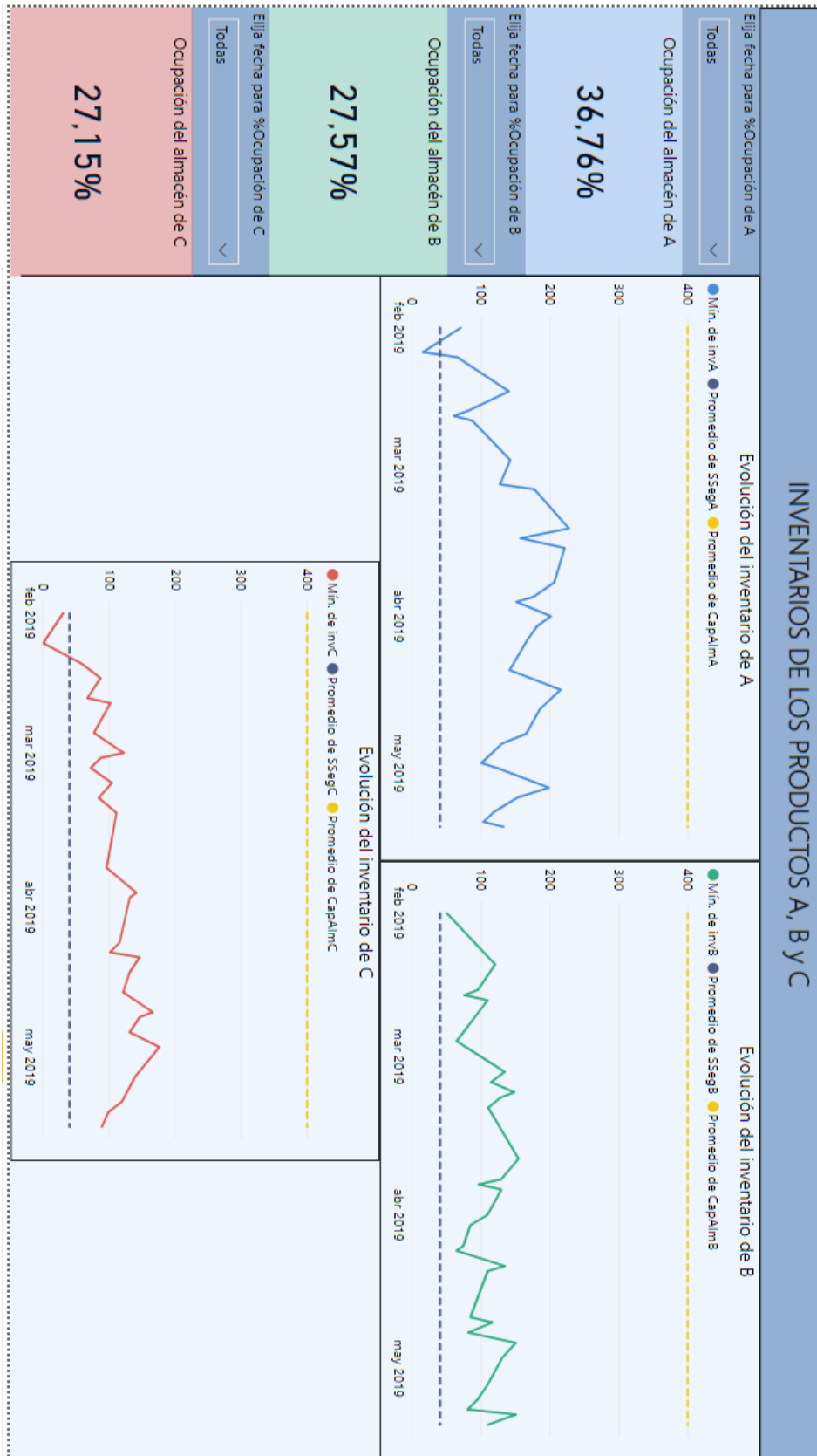
5.4 Cuadro de mando: INVENTARIOS DE PRODUCTOS Y MP

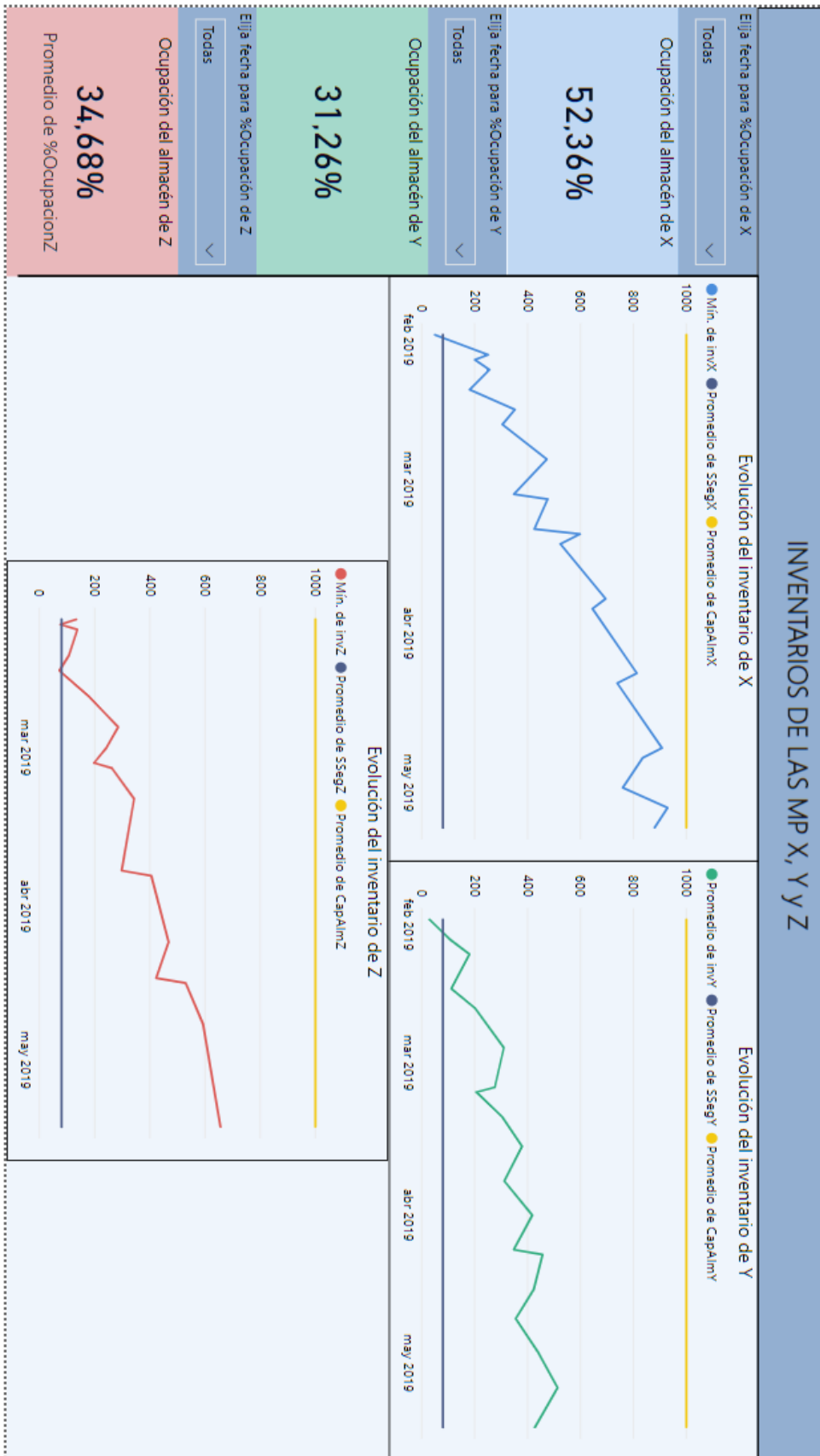
Resumiremos en una sola sección ambos, ya que se utilizarán los mismos recursos aunque dedicados por separado a Productos y MP. El estado de inventarios es fundamental e imprescindible en cualquier análisis o informe de BI.

1. **Evolución de los inventarios por mes.** Se representa gráficamente en función del tiempo el inventario acumulado. Además, se acompaña de las referencias de la capacidad del almacén y el Stock de Seguridad para cada ítem. Es un gráfico básico en cualquier organización.
2. **Ocupación del inventario por día.** Tres ventanas con un desplegable para la elección del día a considerar permiten visualizar el porcentaje de ocupación del almacén. Muy útil para conocer el estado en un día concreto en casos de problemas o hechos que hayan perturbado la producción. Permite esclarecer causas o indagar en conflictos ocurridos o cómo influyó solventarlos.

El resultado de sendos cuadros de mando propuestos para Productos y MP se muestra en las Ilustraciones 42 y 43.

Si comenzamos con el análisis de los resultados, tenemos que en los inventarios de los Productos no existe ningún hecho significativo en las gráficas. Obtenemos evoluciones, dentro de lo que cabe, uniformes, sin tendencias. Lo más destacable es el uso del Stock de Seguridad tanto de A como de C en las primeras semanas (Ilustración 44 y 45, respectivamente), seguramente, debido a una mala estimación de la demanda a corto plazo. En principio, no es algo que debe ser preocupante, ya que son hechos aislados en los que en las primeras semanas hubo un aumento de la demanda de dichos productos.





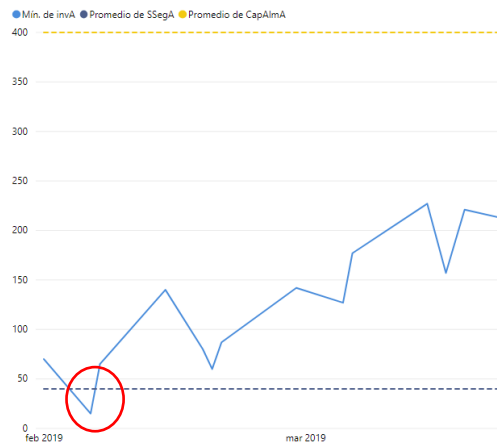


Ilustración 44. Inventario de A. Reposición de Stock de Seguridad

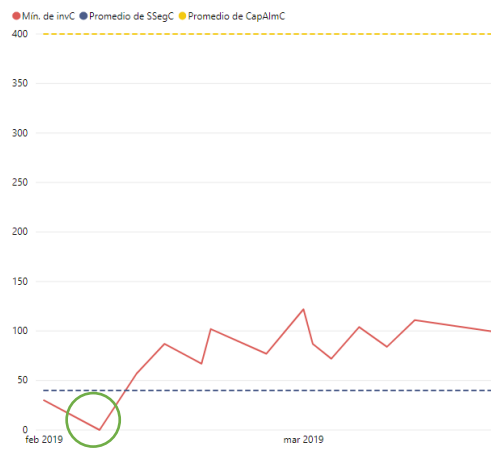


Ilustración 45. Inventario de C. Reposición del Stock de Seguridad

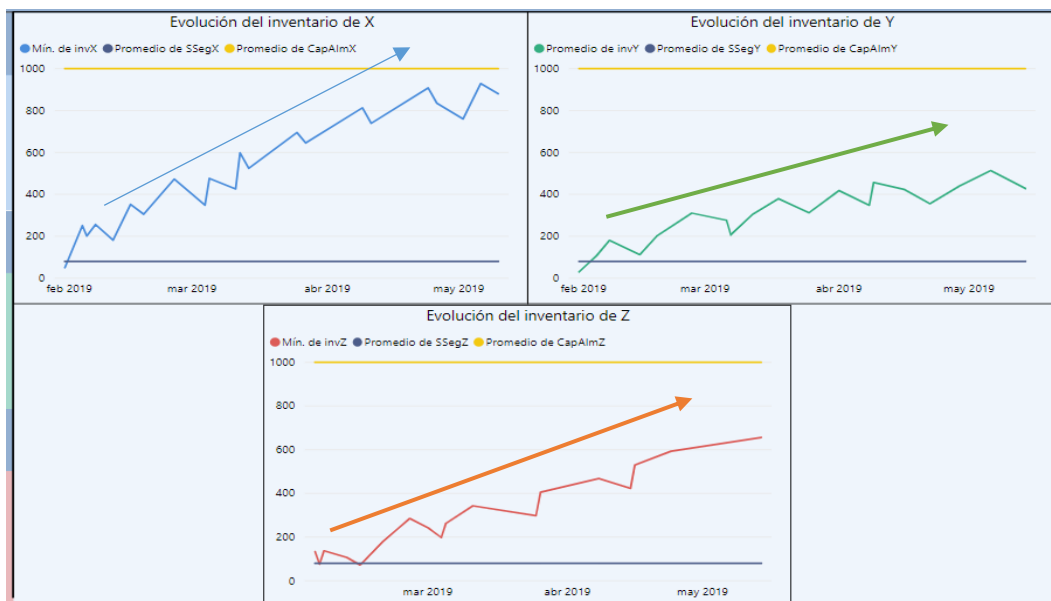


Ilustración 46. Aumento en los inventarios de MP

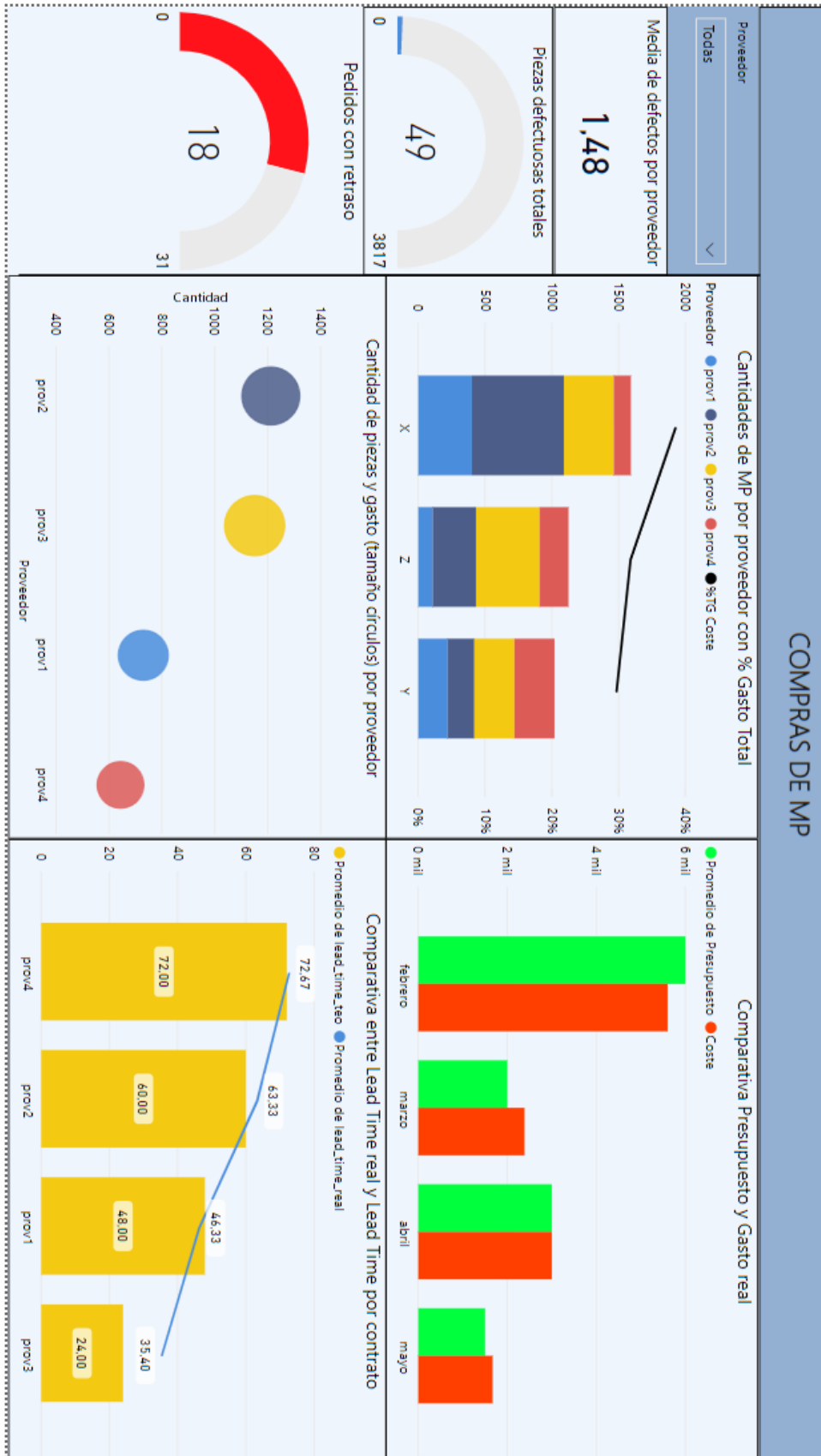


Ilustración 47. Página 4: Compras de MP



5.5 Cuadro de mando: COMPRAS DE MP

La complejidad de las páginas va en aumento, según vamos avanzando en su explicación. La página de Compras de MP muestra mediante cuatro gráficos y tres ventanas la información más importante en cuanto a los pedidos y los proveedores, tal y como se muestra en la Ilustración 47. Pone especial interés en los gastos y en la calidad del servicio presado por los proveedores.

1. **Cantidades de MP por proveedor con gasto total en cada MP.** En dicho gráfico de barras resumimos gran cantidad de información. Por una parte, se muestra para cada MP la cantidad total comprada de cada uno, donde en cada barra se representa la parte proporcional adquirida a cada proveedor en función del color. Se incluye una línea que representa el porcentaje del gasto total en compras que supone cada MP, lo cual se refiere al típico análisis ABC.
2. **Comparativa entre el presupuesto mensual y el gasto mensual.** Es sabido que el departamento de compras tiene un presupuesto mensual asignado. Como es evidente, respetar los presupuesto o incluso, ahorrar, es de gran importancia para el estado financiero de la compañía.
3. **Gráfico de burbujas de la cantidad de piezas adquiridas y el gasto en cada proveedor.** La propensión a la compra a ciertos proveedores puede indicar multitud de información encubierta: mejores condiciones, no objetividad a la hora de asignar pedidos, deficiencia en el servicio, etc. Para cualquier empleado alejado del departamento propio es difícil conocer esta información. Por ello, un método fácil que sintetiza cualquier tipo de correlación es el gráfico de burbujas (donde el tamaño indica el gasto en dicho proveedor).
4. **Comparativa del Lead Time real y por contrato.** En dicho gráfico representamos el Lead Time definido por contrato por barras y representamos la media del Lead Time real en una línea. La calidad del servicio de un proveedor es en gran parte debida a su tiempo de suministro. Es por ello que se debe tener conciencia de cuál de ellos lo respeta o incluso mejora, y aquel que se suele retrasar.

Los siguientes tres gráficos se acompañan de un desplegable que permite seleccionar el proveedor a que se referirán (o también una media de todos).

5. **Media de piezas defectuosas.** Otra gran parte de la calidad del servicio proporcionada por el proveedor depende de la cantidad de piezas defectuosas que nos entrega. Aunque es evidente que cualquier control de calidad no es infalible, algunos pueden presentar mayores medias y nos permitan identificar a aquellos que sus controles no son lo suficientemente rigurosos.
6. **Piezas defectuosas totales.** En la misma temática que el anterior, se ofrece un medidor en el que se representan las piezas defectuosas respecto del total de piezas adquiridas.
7. **Pedidos con retraso.** El gráfico comparativo del Lead Time es útil pero nos ofrece una información parcial, ya que una media cercana a la establecida por contrata puede suponer, por ejemplo, que la gran mayoría de pedidos fueron entregados con retraso de algunas horas (la diferencia es poca, pero son retrasos). Por ello, se ofrece un medidor en el que se representan respecto del total de pedidos, cuántos fueron entregados con retraso.

En esta página comenzamos a tener información realmente relevante sobre nuestros proveedores. Sin embargo, comencemos con las materias primas. En el primer gráfico, en la Ilustración 48, el cual se ha creado con intenciones de realizar un análisis ABC, observamos que en cierta parte se cumple que el alrededor del 70% de nuestros gastos se reparten dos de nuestros productos (66% de los productos). Esto supone que el 70% de nuestros gastos están en unas 2700 piezas de un total de 3700, lo cual es un gasto muy significativo.

También es fácil diferenciar que los proveedores a los que más órdenes se les asignan son los proveedores 2 y 3. Además, el proveedor 2 es un proveedor clave para la MP X. Continuando el análisis de los proveedores, vemos en el gráfico de burbujas (Ilustración 49) que los proveedores 2 y 3 son, con diferencia, los proveedores que mayor cantidad de piezas suplen y además, lógicamente, los que mayor gasto incurren.

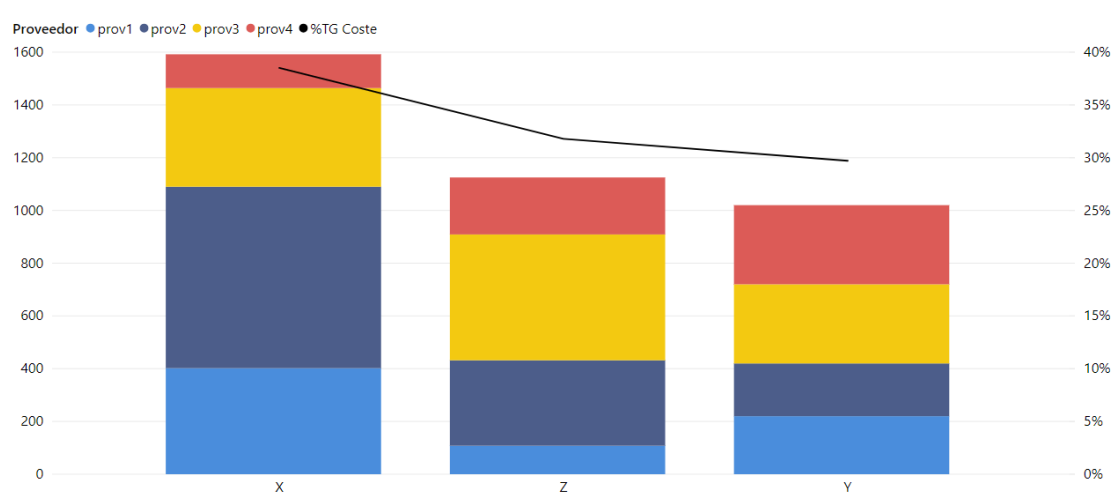


Ilustración 48. Cantidades de MP compradas por proveedor con % del Gasto Total

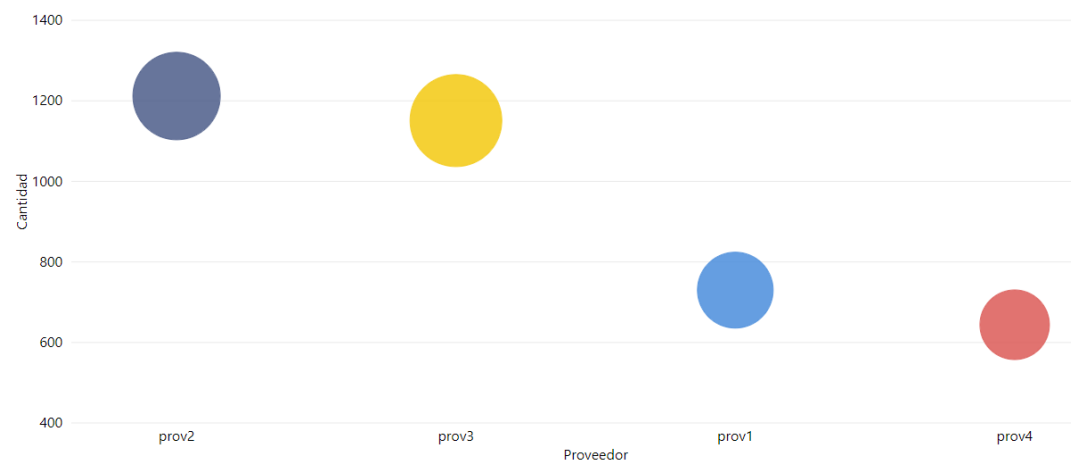


Ilustración 49. Gráfico de burbujas de cantidades adquiridas y gasto en cada proveedor

Si tenemos en mente dicho factor y echamos un ojo a la Ilustración 50, es inevitable ver como los proveedores 2 y 3, irónicamente, son los que más se desvían de su Lead

Time teórico, con diferencias de 3 horas por parte del proveedor 2 y hasta 11 horas en el caso del proveedor 3. A su vez, es curioso como el proveedor 1 incluso tiene una desviación de adelanto frente a su Lead Time teórico.

Continuando la línea de los proveedores, también podemos ver claras diferencias entre los defectos y los retrasos de cada uno de ellos en las Ilustraciones 51 y 52.

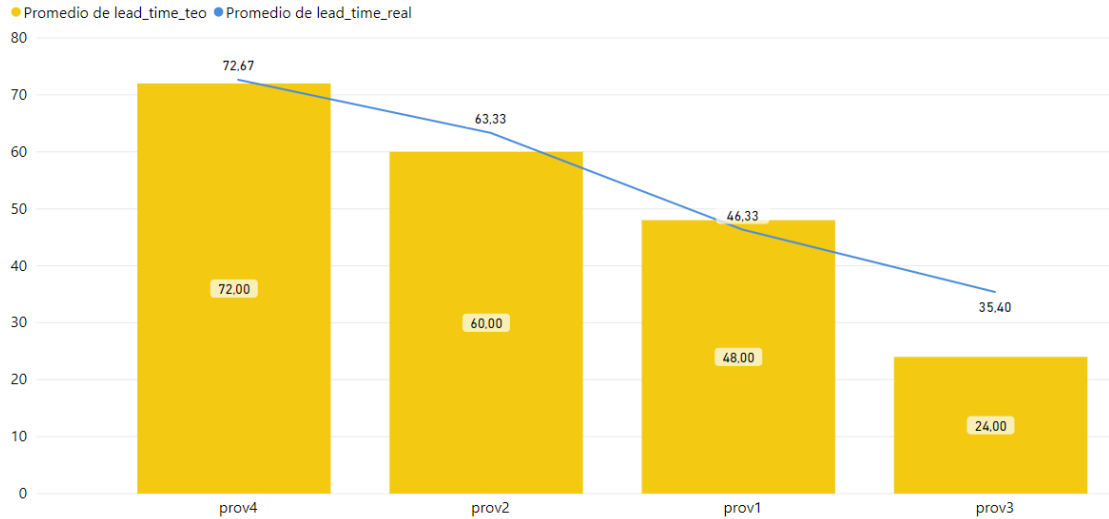


Ilustración 50. Comparativa del Lead Time teórico y real de proveedores

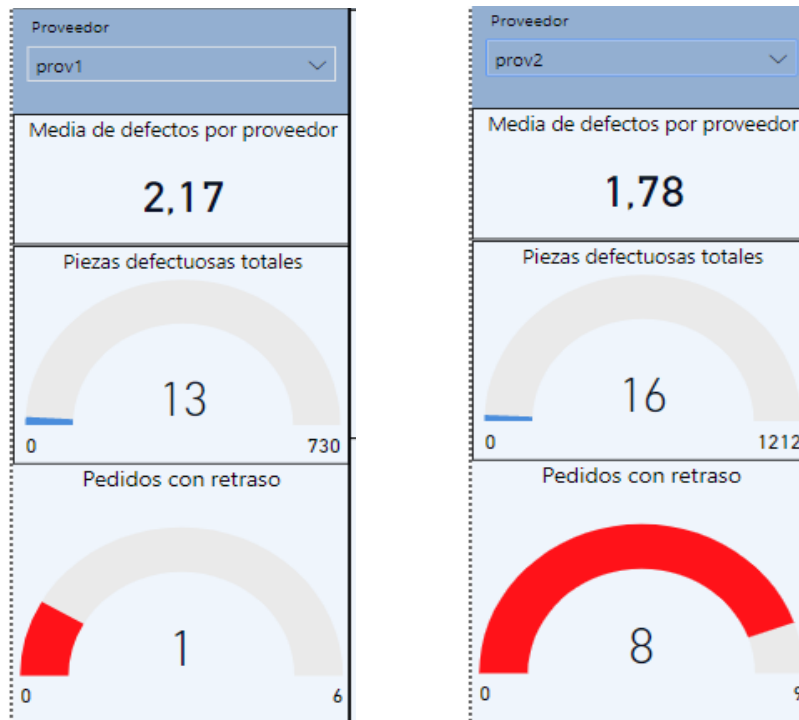


Ilustración 51. Defectos y retrasos de Proveedores 1 y 2

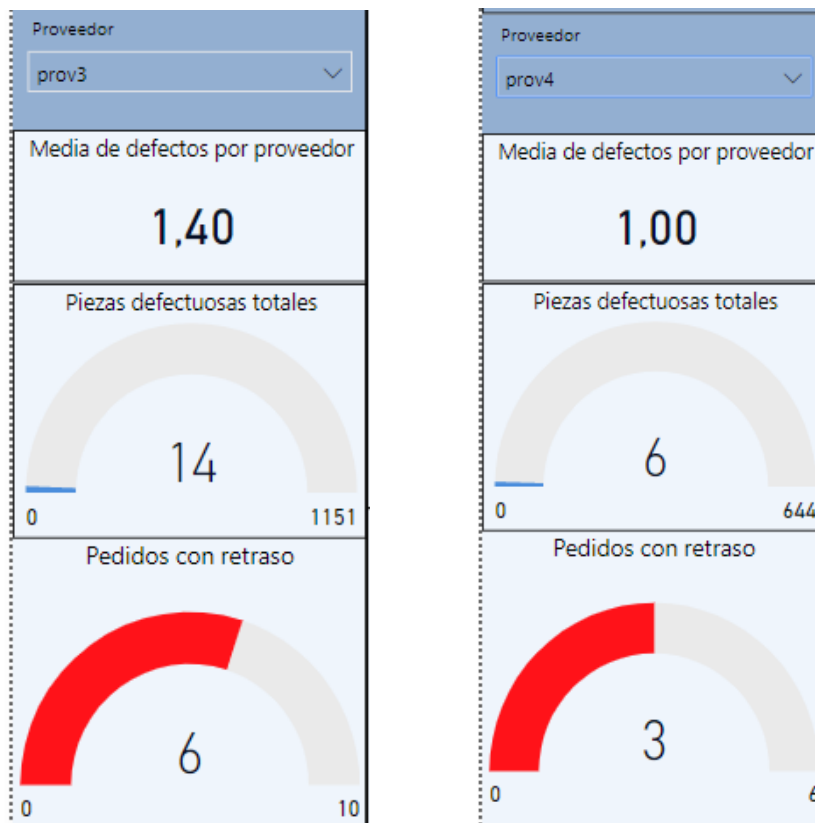


Ilustración 52. Defectos y retrasos de Proveedores 3 y 4

Por otra parte, observamos que se confirman nuestras sospechas. Los proveedores 2 y 3 son aquellos que más pedidos con retrasos han servido (más del 90% y un 60% respectivamente). Sin embargo, también es cierto que el proveedor 1 es el que más productos defectuosos por pedido de media entrega.

Recordando los precios por MP de cada uno de los proveedores en la Tabla 14, y viendo que no existe una gran diferencia entre el proveedor 1 y el proveedor 3 en cuanto a los precios, debería considerarse la incorporación del proveedor 1 como fuente principal de suministros a pesar de una mayor media de defectos por pedido (aunque no es exageradamente alta).

Nombre	LeadTime (horas)	CosteX/ud	CosteY/ud	CosteZ/ud
prov1	48	4	3	4
prov2	60	2	5	4
prov3	24	3	5	3
prov4	72	6	2	4

Tabla 14. Datos de proveedores (2)

Por último, si se echa un vistazo al gráfico comparativo de gastos mensuales frente el presupuesto en la Ilustración 53, es preocupante que en dos de ellos se superen los presupuestos. Además, Marzo, coincide con el mes en el que el EBITDA fue negativo, por lo que los gastos en MP son la principal causa de las pérdidas. Se deben tomar medidas más restrictivas para controlar los gastos y ajustarlos a los presupuestos.

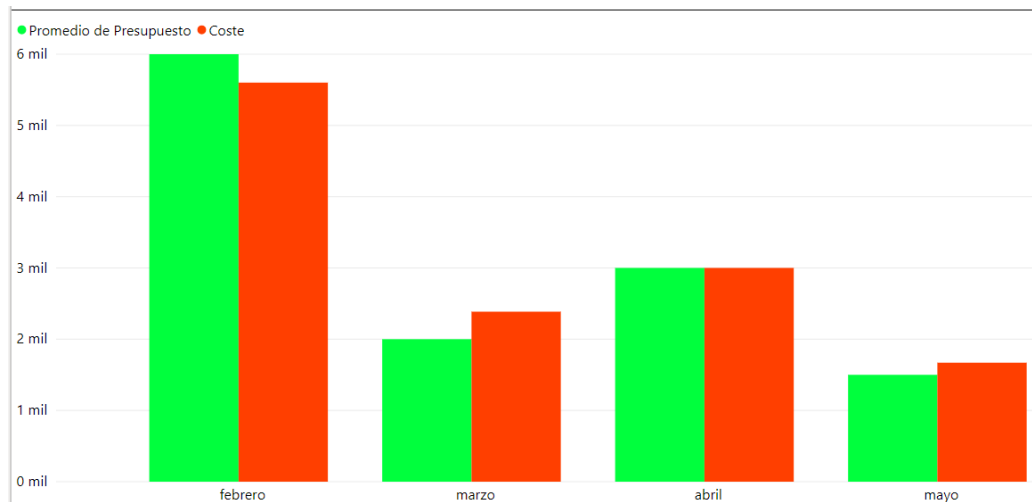


Ilustración 53. Comparativa de gasto y presupuesto mensual

5.6 Cuadro de mando: PRODUCCIÓN

El tipo de empresa que es EJEMPLO S.A hace que la página de Producción sea el corazón del análisis. De esta página dependen muchas de las decisiones que serán tomadas en cuanto a su proceso productivo. Por ello, se ha de asegurar que la información contenida en ella sea de máxima utilidad y genere un punto de vista que de otra manera no podría ser posible. En esta página nos centramos especialmente en los productos (cantidades, costes, tiempos), y en menor medida, en las paradas en la línea (Ilustración 54).

- 1. Cantidades de producto fabricado con gasto total en cada producto.** Análogamente al primer gráfico de Compras de MP, representamos las cantidades de cada producto fabricado con el porcentaje del gasto total que representa para así, al igual que antes, asemejarlo a un análisis ABC.
- 2. Diferencia respecto de las MP planificadas.** Los tan conocidos desajustes debido a una mala utilización de las MP pueden ahora ser controlados a través de un gráfico que muestra la cantidad (diferencia) de desajuste total para cada producto respecto de las materias primas. Evidentemente, por ejemplo, no se producen desajustes para el producto B en las MP X y Z, ya que no son necesarias para su fabricación. Es importante conocer de dónde proceden las fuentes principales de desajuste.
- 3. Comparativa de TFabricación teórica y real.** La productividad de los procesos (en términos generales, no individuales) puede ser medida de diferentes formas. Una de ellas es comparar los tiempos de fabricación unitarios reales (a través de la duración de las operaciones) respecto a los tiempos de fabricación calculados y establecidos. La comparación puede sernos útil para detectar si existen deficiencias en los procesos productivos, y además, descubrir a qué producto afecta en mayor medida (puede ser por diferentes causas, ya sea por malas estimaciones, problemas en la línea, etc).

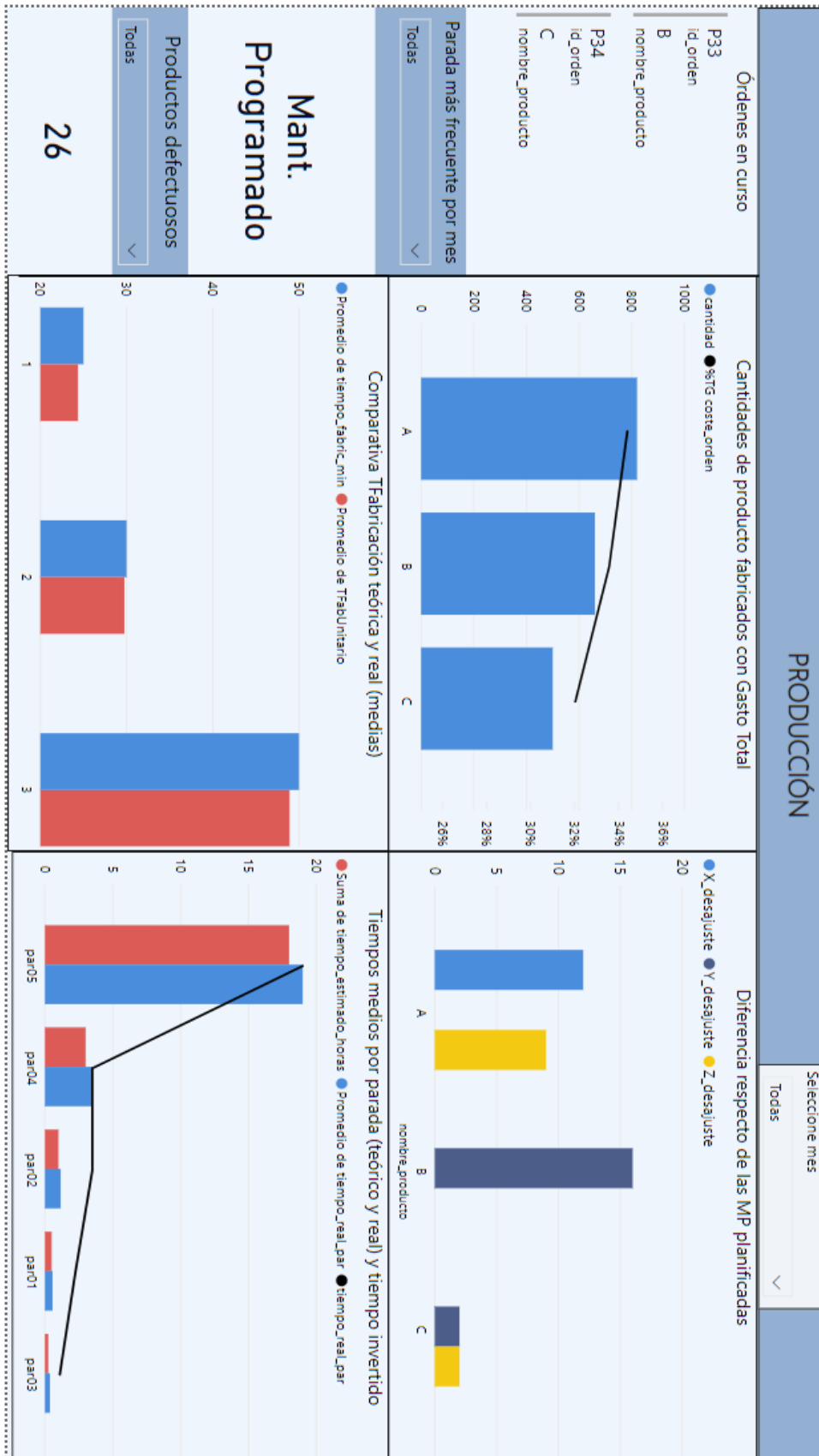


Ilustración 54. Página 5: Producción

4. **Tiempos medios por parada (teórico y real) y tiempo invertido.** Las paradas son otra fuente de pérdidas (si las máquinas no trabajan, las máquinas no producen pero sigue habiendo gastos fijos). Se deben esclarecer las paradas más frecuentes, las causas y las desviaciones respecto de los tiempos estimados de parada para conseguir un mayor aprovechamiento del capital humano. Es por ello, que en esta gráfica se comparan los tiempos estimados por cada una de las paradas respecto de las medias reales, y por otra parte, se muestra una línea que representa el tiempo total invertido en ellos (mayor desviación respecto de las barras, mayor frecuencia).

Los **gráficos anteriores** se acompañan de un **desplegable** para **escoger el mes** que se quiere analizar.

5. **Órdenes de producción en curso.** A mero propósito informativo, aunque no por ello irrelevante, se muestra una ventana con los pedidos en curso que no han sido finalizados, junto con el producto que se está fabricando.
6. **Parada más frecuente por mes.** Aunque en el gráfico de comparación de tiempos de paradas pueda intuirse cuáles son las paradas más frecuentes, en esta ventana, acompañada de un desplegable para elegir el mes, se puede encontrar la parada más frecuente. Esto es especialmente útil para detectar problemas recurrentes durante periodos de tiempo con un solo clic, sin necesidad de comunicarse con los encargados, ahorrando tiempo y recursos.
7. **Productos defectuosos.** Es una ventana indispensable en un proceso productivo. La ventana, la cual funciona junto con el desplegable de los meses de las paradas, también se acompaña con otra para elegir el producto. Es una manera rápida y eficaz de observar el buen o mal hacer del proceso productivo.

Analizando los datos, análogamente a Compras de MP, nos fijaremos en el gráfico que pretende basarse en un análisis ABC en la Ilustración 55. En este caso, alrededor de un 67% de los costes están provocados por 2 de los 3 productos (66%), los cuales suponen unas 1500 piezas de unas 2000 totales. En este caso, no se acerca tanto a la regla común, pero también es un buen indicador.

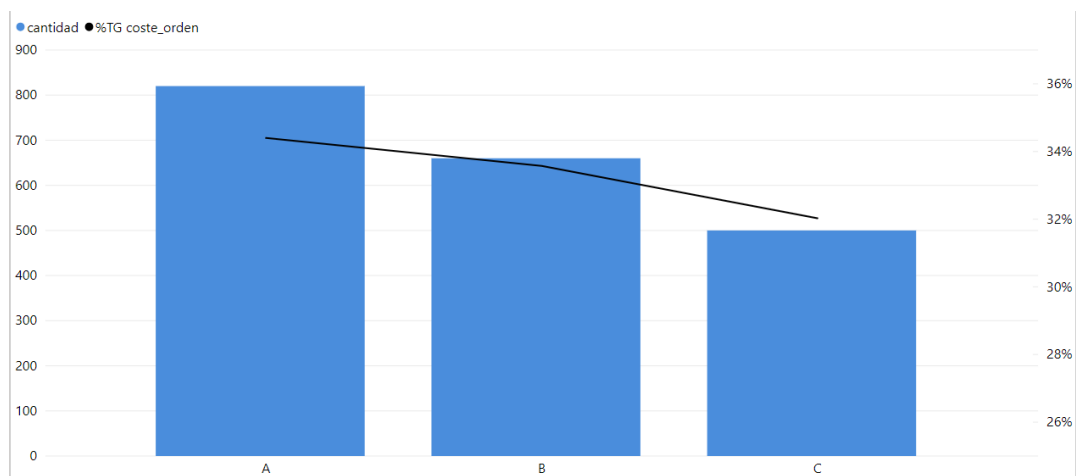


Ilustración 55. Cantidades fabricadas de producto con % del gasto total

Por otra parte, en los desajustes de MP en las fabricaciones de Productos (Ilustración 56), se evidencia que el mayor desajuste de Y se produce en la fabricación de B, y una parte mínima en C. Lo mismo ocurre con X y Z en la fabricación de A, habiendo una mínima desviación en Z para C llegando incluso a no haber para X en C. Esto nos da a entender que el problema de los desajustes está relacionado con las líneas y los trabajadores, ya que los productos A y B, que tienen unos grandes desajustes en su MP, solo se fabrican en la Línea 1, a diferencia de C. Se procederá a una investigación más rigurosa de las causas, ya que podría tratarse de algo más preocupante que defectos, como por ejemplo, robos.

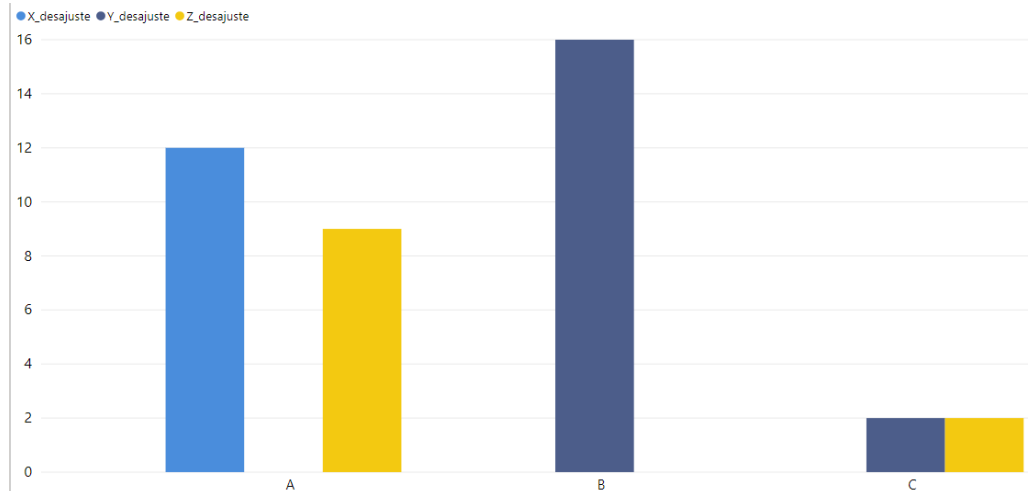


Ilustración 56. Desajuste de MP por productos

Pasando al gráfico de los tiempos de fabricación unitarios en la Ilustración 57, nos llevamos una grata sorpresa viendo cómo todos los productos son realizados de media, en un tiempo menor que lo establecido. Es por ello, que se recomienda revisar el proceso productivo, estandarizar y crear una cultura de lo que se hace actualmente para poder incluso tratar de mejorarlo.

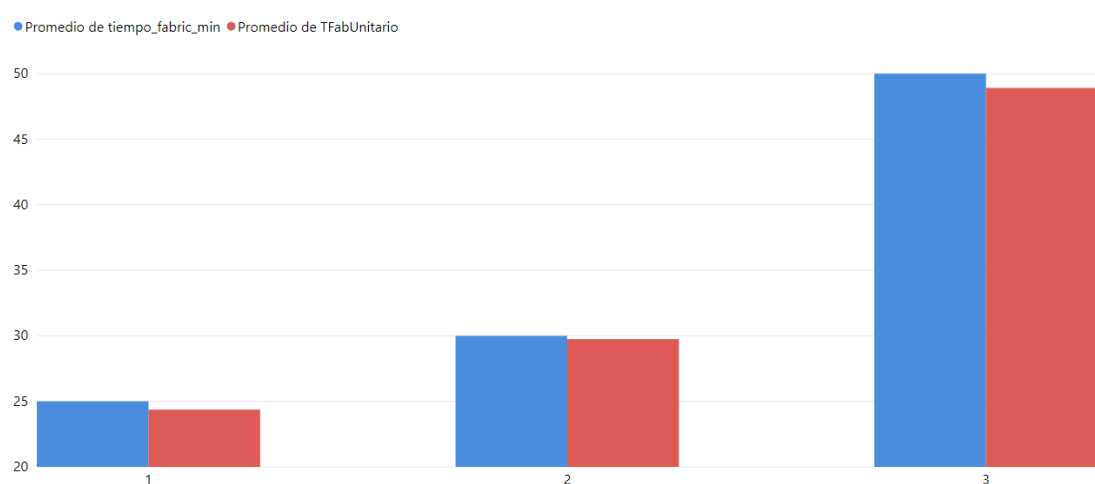


Ilustración 57. Comparativa de tiempos de fabricación teóricos y reales

Por otra parte, en cuanto a las paradas, las medias de cada una de las paradas andan muy parejas en cuanto a las teóricas, tal y como es visible en la Ilustración 58 (hay que

tener en cuenta la escala en horas del gráfico). Lo que sí resulta curioso es como el tiempo invertido (línea negra) es mayor para aquellas que tienen una menor duración, existiendo una mayor distancia entre las barras y la línea. Así ocurre con par02 (preparación de la línea), pa01 (mantenimiento programado) y par03 (accidente leve).

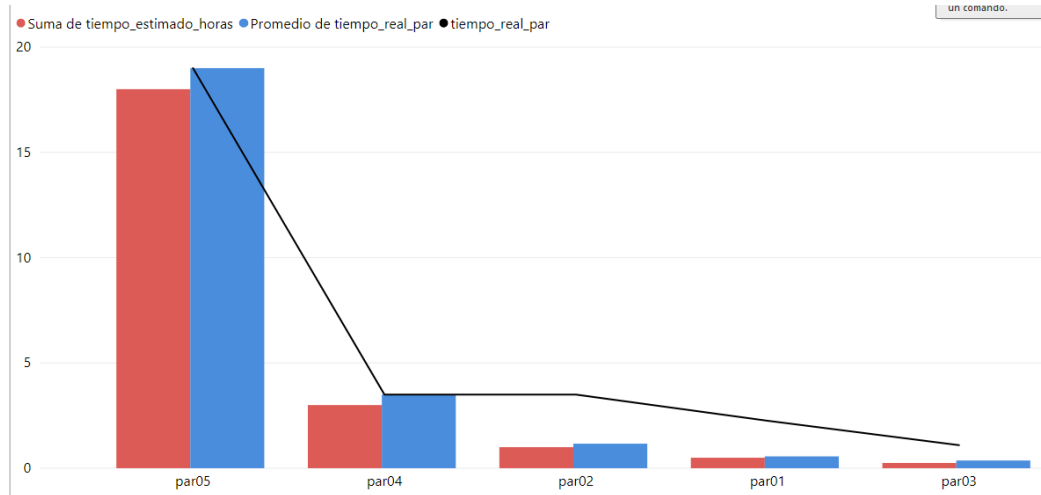


Ilustración 58. Tiempos teóricos y reales de paradas y tiempo invertido

En la Ilustración 59 se visualizan las paradas más frecuentes por mes. Podemos deducir que los accidentes leves son imprevisibles, aunque sí evitables con unas medidas de prevención más estrictas para evitar tantas paradas. Sin embargo, los accidentes graves son un asunto muy importante a tratar y minimizar su aparición. Por otra parte, los mantenimientos programados y las preparaciones de línea sí son eventos planificados. Los mantenimientos programados son típicos y se realizan, de manera general, con periodos y frecuencias establecidas y son necesarios. Sin embargo, las preparaciones de línea sí pueden ser evitados en mayor medida si se establece un plan de producción que requiera de menos preparaciones (línea cambia de fabricación de producto) con el máximo de pedidos del mismo producto consecutivamente.



Ilustración 59. Paradas más frecuentes por mes

Por último, analizaremos las no-calidades en los productos. Por producto, no detectamos ninguna desviación, obteniendo en total, 8 defectos para A, 9 para B y 9 para C (Ilustración 60). Una distribución prácticamente equitativa.

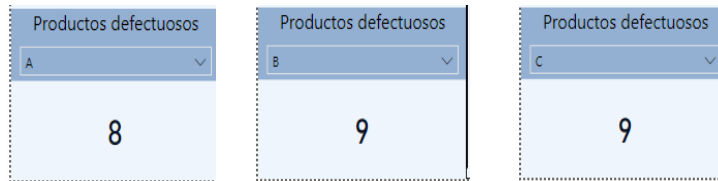


Ilustración 60. Productos defectuosos

Sin embargo, si visualizamos esto por meses, es curioso como el mes de marzo reúne 11 de los 27 defectos posibles. Además, filtrando por Productos, el principal motivo es el producto C con 6 piezas (Ilustración 61), frente a 1 de A y 4 de B.

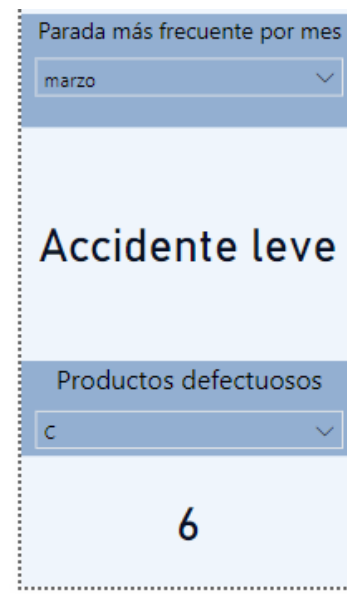


Ilustración 61. Productos defectuosos de C en Marzo

5.7 Cuadro de mando: OPERARIOS Y LÍNEAS DE FABRICACIÓN

La página de Producción ya es bastante útil para informar sobre los procesos de producción y encontrar ciertas causas. Sin embargo, las correlaciones entre líneas y operarios son las que realmente pueden evidenciar causas o problemas en nuestra actividad. Por ello, en esta pestaña nos centraremos en la información relativa a éstos y las no-calidades, retrasos y paradas. La característica más importante de esta página es la capacidad de complementar los análisis de cada uno de los gráficos con el resto. Todos entre ellos son complementarios y aportan más información a cada descubrimiento. El resultado de este cuadro de mando se visualiza en la Ilustración 62.

- 1. Horas trabajadas por operario.** La mera disposición de las horas realizadas por operario puede evidenciar una mala planificación, o más tarde, permitirá entender las no-calidades, retrasos, etc. debido a una excesiva (o al contrario, escasa) actividad en los procesos.





2. **NO-Calidad por Producto y Operario.** Correlación fundamental que no podía faltar. Se analizan la cantidad de no-calidades de cada uno de los operarios con una leyenda por colores en las barras para cada producto. Cualquier correlación o desviación puede ser acompañada por la interpretación en otro de los gráficos.
3. **Ud/hora por operario.** Al igual que se midió la productividad de los procesos anteriormente mediante los tiempos de producción, en este caso mediremos las productividades de cada uno de los operarios mediante las unidades fabricadas por hora de media (comprendiendo que cada producto tiene un tiempo de fabricación diferente, y no todos los operarios trabajan en todos los productos). Es un gráfico fundamental para descubrir las carencias de nuestros empleados o para premiar a los mejores.
4. **NO-Calidad por Producto y Línea.** Al igual que puede existir una correlación entre defectos y operarios, también puede existir por la línea, por ello es importante descubrir si la línea juega un papel importante en los defectos en los productos, o actúa de especial manera sobre algunos.
5. **Horas trabajadas por línea.** Cualquier usuario que conoce el proceso productivo conoce las horas disponibles de cada línea semanal o mensualmente. Por ello, se acompaña al gráfico una ventana que muestra las horas trabajadas durante el periodo analizado.
6. **Mayor razón de paradas por línea.** Acompañado de un desplegable para elegir línea, al igual que en Producción podemos diferenciar parada más frecuente por mes, ahora podemos detectar la parada más frecuente en cada una de las líneas. Como es evidente, conocer cuál es la más frecuente puede evidenciar problemas recurrentes que a la larga, puede derivar en algo peor. Por ello, debemos controlar las paradas en función de la línea.
7. **Retrasos por operario.** Ninguno de los operarios puede trabajar en las dos líneas, por lo que se ha descartado la posibilidad de seleccionar la línea para buscar dificultades de cierto operario en alguna línea. En cambio, mediante un menú desplegable, se puede elegir el operario y se mostrarán la cantidad de retrasos debidos a ese operario en función del total de pedidos en los que ha estado involucrado. Permite diferenciar a los trabajadores más eficaces de los que tal vez, necesiten una mejor formación o un cambio de línea.

Sin duda alguna, tal y como se ha repetido hasta ahora, probablemente es la página más interesante para ser analizada. Lo primero que nos llama la atención en el primer gráfico (Ilustración 63) es el reparto desigual de horas trabajadas en la Línea 1, en la que el OP1 ha trabajado 200 horas, OP2 más de 300 y OP3 poco más de 30. Sorprende la escasa participación del OP3.

Si seguimos al OP3, vemos como en el gráfico que relaciona Operarios y NO-Calidades en la Ilustración 64, se obtiene de manera lógica, que el OP3 no ha estado presente en ninguna NO-Calidad. También es curioso como el OP5 trabaja menos horas que OP1 pero sin embargo, obtiene un menor número de NO-Calidades en sus productos (OP5 solo fabrica un producto).

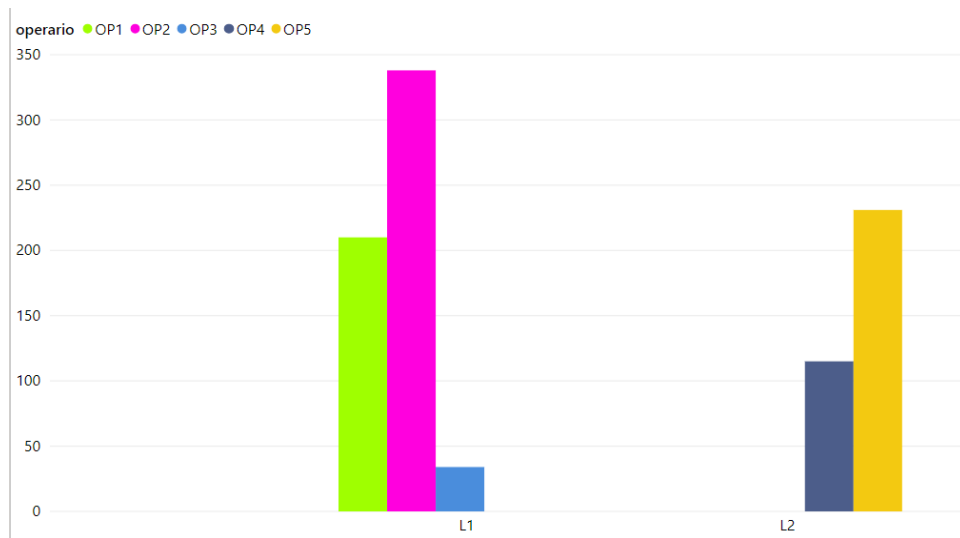


Ilustración 63. Horas de trabajo por operario

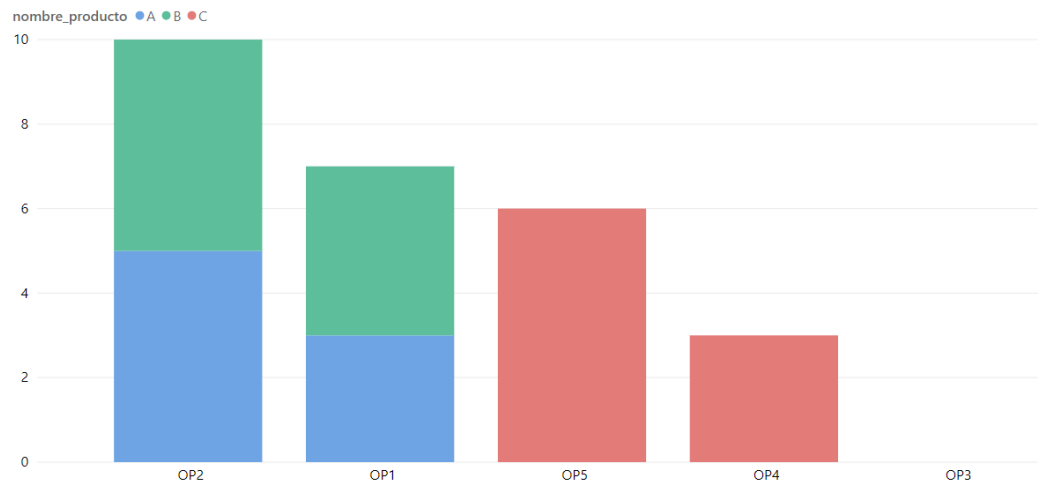


Ilustración 64. NO-calidades por operario

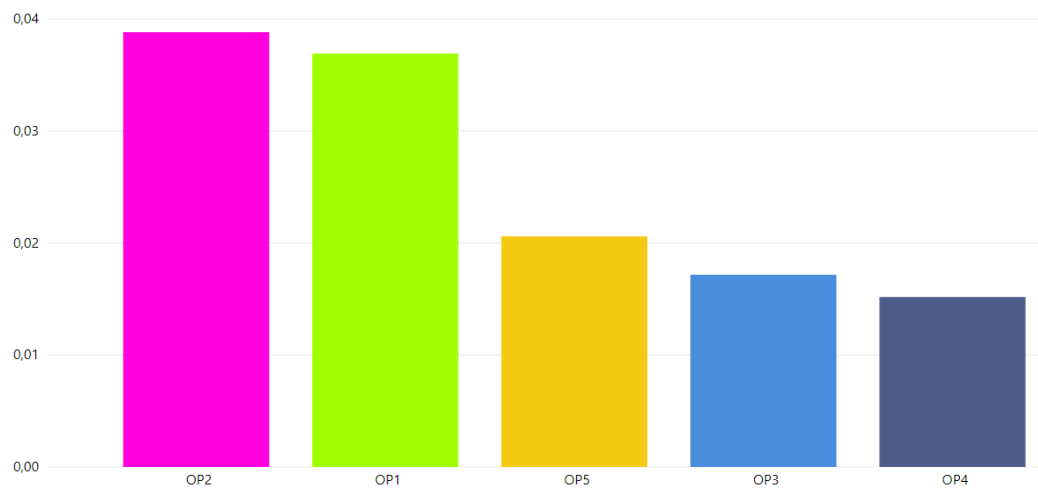


Ilustración 65. Productividad (unidades por hora) por operario

Sin embargo, lo que es más sorprendente aún es en el gráfico de las productividades, visible en la Ilustración 65, como el OP3 a pesar de ser con diferencia el que menos horas ha trabajado, es el que tiene la segunda peor marca de fabricación de unidades por hora.

También debemos destacar la mala labor del OP5, habiendo trabajado las mismas horas que OP1 y obteniendo una productividad que es casi la mitad de OP1. A su vez, se debe felicitar y premiar al OP2, siendo el operario más trabajador y productivo de los compañeros con diferencia.

Siguiendo con los Operarios, si consultamos los retrasos en las órdenes de fabricación (Ilustración 67), de un total de 34 órdenes, solo se produjeron 12 retrasos. Lo que llama la atención es que de los 12, 7 de los retrasos fueron debidos a OP1, el cual además participó en 8 órdenes (más del 90% de sus pedidos fueron retrasos). De nuevo, se debe de felicitar al OP2 el cual solo produjo 2 retrasos en 13 órdenes que fabricó.

En definitiva, el OP2 es un excelente trabajador, el cual ha trabajado con diferencia más horas que los demás, obteniendo la mejor productividad y muy pocos retrasos. A diferencia de él, el OP3 ha trabajado poco más de 30 horas y ha presentado la segunda peor productividad. OP5, a su vez, ha sido el segundo con más horas invertidas pero obteniendo una productividad que supone casi la mitad que sus semejantes (OP1, OP2). Por último, OP1, a pesar de presentar buena productividad y trabajar bastantes horas, su índice de retrasos es más del 90%.

Se propone incentivar a OP2 para motivar este ritmo de trabajo, además de nivelar las cargas de trabajo de una manera más homogénea (OP3, en particular) y tratar de disminuir las horas de trabajo de OP1, el cual es productivo pero entrega con retraso demasiados pedidos para tener un ritmo de trabajo más holgado. Por último, se debe incentivar a trabajadores como OP5 para mejorar su productividad, o en su defecto, disminuir sus horas.

Por último, en cuanto a las NO-Calidades en función de las Líneas, no se observa nada destacable. Las Líneas 1 y 2 trabajan 582 y 346 horas respectivamente (Ilustración 66) y en términos generales, se encuentran casi el doble de NO-Calidades en la 1 respecto de la 2 (Ilustración 68). Es una proporción bastante cercana a las horas trabajadas y no se puede sacar ninguna conclusión medianamente razonable.



Ilustración 66. Horas trabajadas por línea

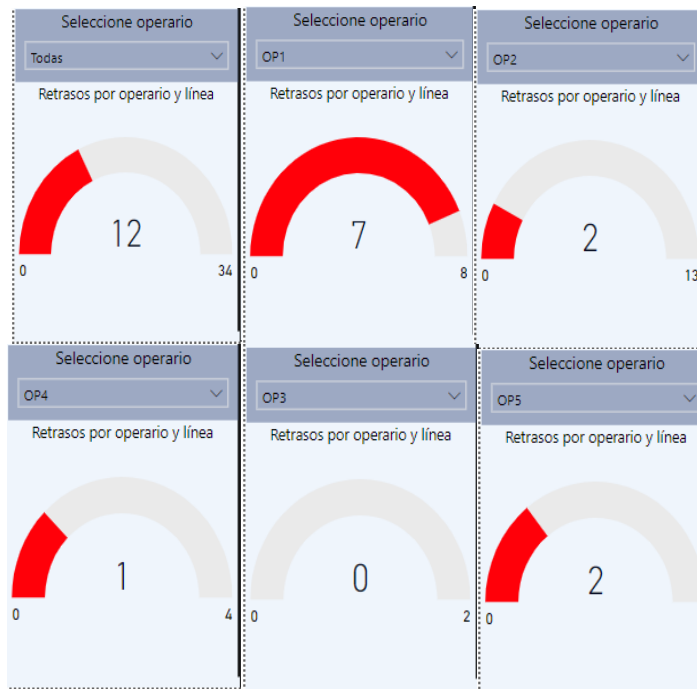


Ilustración 67. Retrasos en pedidos por operario

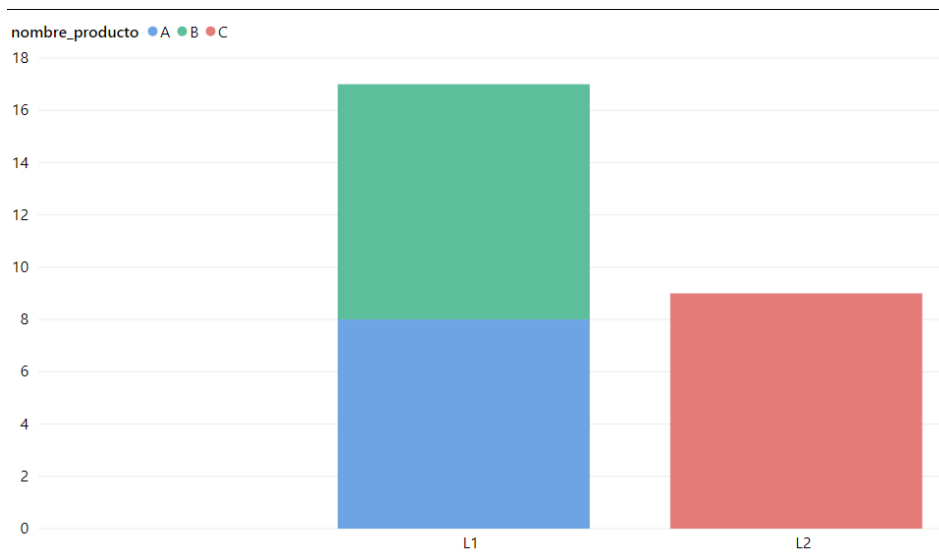


Ilustración 68. NO-calidades por línea



6. Análisis de datos no estructurados

6.1 Introducción y propósito del análisis

Marco teórico

La manipulación de los datos corporativos internos de los procesos y departamentos de una organización son, evidentemente, esenciales en los procesos de BI. La necesidad de conocimiento dentro de nuestros muros es clave para obtener ventajas competitivas sobre el resto del mercado. Sin embargo, durante la última década esto pasado solo a ser una parte. La opinión del consumidor y de la opinión pública ha pasado a ser transcendental en las ventas.

Han surgido nuevos métodos de comunicación y transmisión de información que aumentan la complejidad de las comunicaciones interpersonales, las redes sociales. Los últimos datos indican que Facebook y Twitter reúnen en su plataforma a más de 2200 millones y 320 millones de usuarios activos en el mundo, respectivamente (WeAreSocial, 2019). La creación masiva de contenido en estas páginas por parte de (potenciales) clientes y usuarios hace de éstas un nicho de información que presenta ventajas y oportunidades y que cualquier empresa debería tener en cuenta. El auge de las redes sociales como métodos de comunicación e influencia sobre consumidores ha provocado el nacimiento en los últimos años de una nueva profesión denominada “*influencer*”. Es la máxima demostración de este hecho. Personas que acumulan un considerable número de seguidores o fans y son contratados por marcas y empresas para hacer publicidad o promoción (a veces incluso, encubierta) de sus productos en sus cuentas en las redes sociales.

Esto demuestra la fragilidad de la empresa frente a las opiniones públicas. Se debe analizar y comprender a los usuarios para aumentar su satisfacción. Las empresas están tomando una postura proactiva con el tema, creando y mejorando su imagen de marca a través de la creación de perfiles oficiales en las redes sociales, en los que se publicitan sus actividades públicas, noticias, ofertas de empleo y otros propósitos de marketing.

Esta exposición crítica de los negocios frente a la opinión pública tiene dos principales causas (Stieglitz et al, 2014). Por una parte, los avances tecnológicos permiten un seguimiento, monitorización y análisis automático de los datos generado en las redes sociales. Y por otra, la alta complejidad de las comunicaciones actuales debido a las redes sociales. Han propiciado comunicaciones entre individuos inconexos y desconocidos entre sí, aumentado el rango y escala de las difusiones de información, a la vez que han disminuido los tiempos de propagación masiva de información (superando ampliamente a medios de comunicación oficiales).

Más allá de la complejidad de las propias comunicaciones actuales, se debe tener en cuenta la complejidad de analizar correctamente las redes sociales. Cada una de ellas ofrece diferentes medios (fotos, vídeos, “likes”, emoticonos, puntuaciones, etc.) para expresar opiniones, puntos de vista, emociones o actitudes. Es donde surge el “*Social*



Media Analytics” (SMA) como el ámbito de investigación interdisciplinar científico que pretende cubrir todos estos factores. Su principal objetivo es la creación y evaluación de un método científico, a la vez que se generan técnicas y softwares para el seguimiento, modelaje y análisis de datos de redes sociales para diferentes propósitos. Generalmente, en la industria el **SMA se considera un subconjunto del Business Intelligence** en el que se transforman los datos vírgenes en información útil.

El SMA abarca multitud de fuentes, ya no solo Facebook o Twitter, si no también otras redes sociales, RSS, blogs, wikis, canales de noticias o cualquier fuente de textos y datos no estructurados. El **acceso a estos datos** puede ser puede realizarse de tres maneras: ⁴

1. **Repositorios gratuitos de datos.** Pueden ser descargados gratuitamente, como por ejemplo Wikipedia, la cual ofrece copias gratuitas de todo su contenido, incluidos los proyectos de Wikimedia Foundation. Otro ejemplo es World Bank data, el cual ofrece innumerables conjuntos de datos de multitud de ámbitos diferentes (estadísticas sobre género, clima, finanzas, energía, etc.) e incluso bases de datos completas.
2. **Acceso mediante herramientas.** Algunas fuentes permiten un acceso controlado a sus datos para facilitar la obtención de respuestas a la vez que condicionar el acceso completo a datos. Pueden considerarse a su vez:
 - a. **Fuentes gratuitas.** Sus datos son accesibles gratuitamente, pero sus herramientas limitan y protegen el acceso a los datos “vírgenes”. Un ejemplo es Google con su amplio abanico de herramientas como Trends o InSights.
 - b. **Fuentes de pago o comerciales.** Distribuidores de datos que cobran por el acceso a los datos de las redes sociales. Algunos cobran por el acceso a sus propios datos, pero otros son vendedores con licencia de ofrecer datos de terceros (véase Gnip y DataSift con Twitter o Thomson Reuters con noticias). En este contexto, Gnip es una de las herramientas más importantes del mercado, ofreciendo acceso a Twitter, Tumblr, Disqus y Wordpress, entre otras. Ofrece datos en tiempo real mediante PowerTrack que permite monitorizar datos basados en palabras clave, lugares, frases o tipo de contenido.
 - c. **Acceso mediante APIs.** Los repositorios de datos ofrecen acceso basado en HTTP programable a los datos mediante APIs. Para investigadores y analistas, es la forma más útil de acceso a datos. Destacaremos, por la relevancia en el presente Trabajo de Fin de Grado, los accesos a Twitter y Facebook (posible con Wikipedia, RSS, noticias, geolocalización, etc.)
 - i. **Twitter.** La plataforma Twitter dispone de la opción de restringir el acceso público a cualquier publicación aunque dicha opción está solo presente en menos del 10% de las cuentas. Considerando estas cuentas públicas, sus mensajes están disponibles en formato JSON (*JavaScript Object Notation*, archivo estándar basado en texto derivado de JavaScript) desde Twitter Search API. Por otra parte, permite también acceso a datos **casí** en tiempo real mediante Streaming API

⁴ Adaptado de Batrinca et al, 2015.

(este servicio funciona mediante niveles de permisos y ha lanzado un programa denominado Data Grants para facilitar el acceso a subconjuntos de datos para investigación).

Los filtros que se pueden aplicar para obtener los mensajes que se desean pueden ser por usuario, palabras clave o localización. El archivo resultante es un JSON matricial cuyos objetos son los Tweets filtrados y cada uno está claramente estructurado (usuario o fecha de creación, por ejemplo). En la Ilustración 68 se muestra un ejemplo de archivo de salida para la búsqueda “*financial times*”.

```
{
  // Results page-specific nodes:
  "completed_in":0.019, // Seconds taken to generate the results page
  "max_id":270492897391034368, // Tweets maximum ID to be displayed up to
  "max_id_str":"270492897391034368", // String version of the max ID
  "next_page":"?page=2&max_id=270492897391034368&q=financial%20times&rpp=1&include_entities=1&result_type=mixed", // Next results page parameters
  "page":1, // Current results page
  "query":"financial+times", // Search query
  "refresh_url":"?since_id=270492897391034368&q=financial%20times&result_type=mixed&include_entities=1", // Current results page parameters
  // Results node consisting of a list of objects, i.e. Tweets:
  "results":[
    {
      // Tweet-specific nodes:
      "created_at":"Sun, 18 Nov 2012 16:51:58 +0000", // Timestamp Tweet was created at
      "entities":{"hashtags":[],"urls":[],"user_mentions":[]}, // Tweet metadata node
      "from_user":"zerohedge", // Tweet author username
      "from_user_id":18856867, // Tweet author user ID
      "from_user_id_str":"18856867", // String representation of the user ID
      "from_user_name":"zerohedge", // Tweet author username
      "geo":null, // Geotags (optional)
      "id":270207733444263936, // Tweet ID
      "id_str":"270207733444263936", // String representation of the Tweet ID
      "iso_language_code":"en", // Tweet language (English)
      "metadata":{"recent_retweets":6,"result_type":"popular"}, // Tweet metadata
      // Tweet author profile image URL (secure and non-secure HTTP):
      "profile_image_url":"http://a0.twimg.com/profile_images/72647502/vtyler_normal.jpg",
      "profile_image_url_https":"https://si0.twimg.com/profile_images/72647502/vtyler_normal.jpg",
      // Tweet source (whether it was posted from Twitter Web or another interface):
      "source":"&lt;a href=&quot;http://www.tweetdeck.com&quot;&gt;TweetDeck&lt;/a&gt;",
      "text":"Investment Banks to Cut 40,000 More Jobs, Financial Times Says", // Tweet content
      // Recipient details (if any):
      "to_user":null,
      "to_user_id":0,
      "to_user_id_str":"0",
      "to_user_name":null
    }
  ],
  // Other results page-specific nodes:
  "results_per_page":1, // Number of Tweets displayed per results page
  "since_id":0, // Minimum Tweet ID
  "since_id_str":"0" // String representation of the 'since_id' value
}
```

Ilustración 69. Obtención de datos de Twitter en formato JSON con REST API. Fuente: Batrinca et al (2015)

- ii. **Facebook.** Las opciones de privacidad de usuarios de Facebook son bastante más complejas que las de Twitter, por lo que algunos mensajes son más difíciles de obtener. Facebook dispone de API Graph y API de Marketing. A diferencia de Twitter, Facebook permite el acceso a más elementos que los propios “*post*” de cada usuario. Las consultas permiten acceder a la búsqueda de términos en función de páginas, eventos, grupos, lugares, localizaciones o páginas cuyos temas sean esa consulta. Sin embargo, se

necesitan permisos para algunas consultas: Graph API necesita un permiso incluido en la consulta, búsqueda de páginas y lugares necesita un acceso de la aplicación, mientras que las demás búsquedas necesitan un permiso de usuario. El archivo devuelto por Facebook también es JSON, aunque los campos son completamente diferentes que los de Twitter, debido a la variedad. En la Ilustración 69 se muestra el resultado de la consulta “Centrica” para páginas.

```
{
  "id": "96184651725",
  "name": "Centrica",
  "picture": "http://profile.ak.fbcdn.net/hprofile-ak-snc4v71177_96184651725_7616434_s.jpg",
  "link": "http://www.facebook.com/centricapl",
  "likes": 427,
  "category": "Energy/Utility",
  "website": "http://www.centrica.com",
  "username": "centricapl",
  "about": "We're Centrica, meeting our customers' energy needs now...and in the future. As a leading integrated energy company, we're investing more now than ever in new sources of gas and power. http://www.centrica.com",
  "location": {
    "street": "Millstream, Maidenhead Road",
    "city": "Windsor",
    "country": "United Kingdom",
    "zip": "SL4 5GD",
    "latitude": 51.485694848812,
    "longitude": -0.63927860415725
  },
  "phone": "+44 (0)1753 494000",
  "checkins": 228,
  "talking_about_count": 5
}
```

Ilustración 70. Ejemplo de búsqueda en Facebook Graph API Search. Fuente: Batrinca et al, 2015

Hasta ahora hemos comentado la importancia del SMA y las técnicas de extracción de mensajes de redes sociales. Sin embargo, no hemos ahondado en la principal pregunta, ¿qué es un análisis de sentimientos y cómo se realiza?

El SMA es un ámbito que abarca las tres principales técnicas para obtención de información de redes sociales: “Social Network Analysis” (SNA), análisis de tendencias y análisis de textos/sentimientos (adaptado de (Stieglitz et al, 2014)).

1. **Social Network Analysis.** El SNA estudia las relaciones entre personas, organizaciones, grupos de interés, estados, etc. analizando la estructura de sus conexiones. En el contexto del SMA, el SNA es utilizado para identificar usuarios influyentes y líderes de masas, además de grupos relevantes de usuarios. En detección de diferentes grupos, SNA también es útil con diferentes métodos y algoritmos de detección de grupos tales como el algoritmo Girvan-Newman u otros métodos de clustering como el agrupamiento difuso o el de K-medias. El SNA continúa teniendo problemas para descubrir grupos dinámicos y cambiantes en grandes conjuntos de datos.
2. **Análisis de tendencias.** Es la técnica que hace uso de los más novedosos avances en informática y estadística para predecir temas “candentes” y emergentes en las redes sociales. La mayoría de los modelos están basados en el modelo oculto de Markov donde se asume que el modelo es un proceso de Markov donde se han de determinar una serie de parámetros desconocidos a partir de parámetros observables (Modelo oculto de Markov, 2019).



3. Análisis de textos/sentimientos. Es una técnica de investigación dentro del análisis de contenidos en el que se trata de extraer inferencias válidas a partir de textos dentro del contexto en el que han sido usados. Los métodos cuantitativos de análisis de textos son necesarios debido al aumento de la cantidad y longitud de los textos en las redes sociales. La minería de textos basa su objetivo en tratar de obtener de forma automática modelos, tendencias, patrones o reglas de datos no estructurados. A su vez, también pretende la obtención automática de información y conocimiento a partir de ellos (He et al, 2015). Las técnicas principales de la minería de textos son: análisis clúster, categorización, extracción de información (resumen), etc. Algunos de ellos ya han sido comentados en capítulos anteriores.

Por otra parte, el análisis de sentimientos es la detección y estudio de opiniones, sentimientos, emociones y subjetividades en textos mediante métodos computacionales. La aplicación más importante es la extracción de comentarios positivos o negativos de los textos. Esta aplicación del análisis de sentimientos basa su funcionamiento en técnicas de Machine Learning como SVM (Support Vector Machine), algoritmo clasificador bayesiano ingenuo, máxima entropía, etc. Existe otra vertiente más tradicional que basa su funcionamiento en el de un diccionario en el que se polarizan las palabras para identificar su negatividad o positividad y su fuerza. En términos generales, el análisis de sentimientos sigue encontrando dificultades en las redes sociales debido a la aparición de emoticonos, lenguaje vulgar, sarcasmo... que, en principio, las máquinas no son capaces de distinguir. Por lo tanto, es necesario la supervisión humana en estos procesos para mejorar los algoritmos o detectar errores.

Propósito

Conocido el ámbito, es necesario explicar el planteamiento del análisis a realizar. La empresa, en un esfuerzo de máxima adopción de Business Intelligence, se ha propuesto su introducción en el SMA. Desea conocer la opinión del cliente y de sus empleados a través de una de las herramientas principales del SMA: **el análisis de sentimientos**. En la misma temática, espera encontrar temas recurrentes o problemas que ocurren frecuentemente que son reportados por los clientes y los empleados para corregir u ofrecer medidas en un intento de mejorar su satisfacción.

Se considera la existencia de perfiles oficiales de la organización en las dos redes sociales más importantes: Twitter y Facebook. A su vez, se considera también la existencia de una "Intranet" en la que los empleados son capaces de realizar ruegos, sugerencias, quejas o cualquier mensaje que deseen para ser leídos por la directiva o los responsables que se consideren. Estas serán las fuentes de mensajes para el análisis. En Twitter se considerarán las menciones al perfil de la empresa (mensajes que contengan @+"nombre"). En cambio, en Facebook, se considerarán los post en la propia página de la empresa. En la Intranet, los empleado han de rellenar un formulario con los siguientes campos obligatorios: un código identificador único (generado automáticamente), la fecha, el origen del mensaje, el departamento al que va dirigido y el propio mensaje.

En el presente Trabajo Fin de Grado, **se presupone la extracción de los mensajes de Twitter, Facebook (a través de los códigos comentados) y la Intranet a un fichero Excel** que será importado a Power BI. Este fichero además ha de ser manipulado



manualmente para introducir el departamento de destino del mensaje, ya que las redes sociales no ofrecen un campo para su especificación.

Por otra parte, el período de tiempo considerado para el análisis ha sido desde el 01/02/2019 al 31/05/2019, contando con una muestra de cincuenta (50) mensajes en total. El mismo análisis se recomienda realizarse en frecuencias iguales o mayores con períodos menores de tiempo.

En este caso, se decide realizar un análisis de sentimientos mediante extracción de palabras clave y valoración del mensaje en una escala de cero a uno [0-1], donde cero hace referencia a un comentario muy negativo y uno a uno muy positivo. Estos procedimientos se realizarán gracias a la herramienta de análisis de textos de **Microsoft Azure Text Analytics**. Esta herramienta está incluida en el bloque de **Cognitive Services** de Microsoft: análisis de textos, análisis de multimedia, reconocimiento por voz, reconocimiento de idioma y consultas en lenguaje natural, etc. Cognitive Services, en resumen, reúne todo tipo de herramientas de Inteligencia Artificial y Machine Learning. Pero, ¿cómo funciona Text Analytics?

Text Analytics es un servicio basado en la nube que proporciona un procesamiento avanzado de idioma natural sobre texto sin formato e incluye cuatro funciones principales: análisis de sentimiento, extracción de palabras clave, detección de lenguaje y reconocimiento de entidades (Microsoft Docs, 2019A).

Para poder aprovecharnos de las capacidades de ésta, debemos poseer una cuenta de Microsoft Azure, ya que es un servicio de pago y no está disponible gratuitamente. La cuenta es necesaria debido a que el análisis de texto requiere de una clave propia y única para el acceso a la API. Text Analytics utiliza un algoritmo de aprendizaje automático (Machine Learning) para generar una puntuación en el que el modelo se entrena previamente con un cuerpo de texto amplio con las asociaciones de opinión (aunque este texto actualmente no es accesible al público). Las técnicas usadas por el modelo son: análisis de categorías gramaticales, ubicación de las palabras y las asociaciones de palabras (Microsoft Docs, 2019B). El modelo posee una fase de evaluación de objetividad, para detectar si es objetivo o contiene opiniones. Un texto que es considerado objetivo, no continúa siendo analizado en la fase de opiniones, obteniendo una puntuación de 0,5.

Más adelante serán explicados en profundidad el esquema de la interfaz de Power BI resultante y el análisis completo de ésta, además de las conclusiones obtenidas gracias a él.

6.2 Preparación de los datos

Los datos del archivo Excel son cargados en Power BI de la forma mostrada en la Ilustración 71. Antes de proceder a su análisis debemos configurar y crear las funciones de análisis de sentimientos en Power Query.

id	fecha	AB departamento	AB origen	AB C text
1	01/02/2019	Producción	Interno	Tenemos malos horarios y no podemos compaginar familia y trabajo. ...
2	03/02/2019	Logística	Facebook	Envío recibido en solo un día. ¡Incredible!
3	04/02/2019	Calidad	Facebook	El producto comprado tiene defectos a simple vista. No volveré a com...
4	04/02/2019	Producción	Interno	El trato recibido por el jefe de producción es nefasto. Se necesitan ca...
5	05/02/2019	Calidad	Interno	Utilizamos las mejores y más novedosas técnicas del mercado. Es un pl...
6	07/02/2019	Atención al cliente	Twitter	Me encanta que mis dudas se resuelvan tan rápido. ¡Gracias!
7	08/02/2019	Atención al cliente	Facebook	Llamé para preguntar sobre el estado de mi pedido y la persona que m...
8	10/02/2019	Calidad	Twitter	Los acabados del productos son de primera calidad. Mi enhorabuena a...
9	12/02/2019	Logística	Interno	Completamente inseguro trabajar en la zona de salidas y expediciones...
10	13/02/2019	Atención al cliente	Interno	Las jornadas de trabajo son muy largas y más de 8 horas. Los descanso...
11	15/02/2019	Calidad	Twitter	Cada día pienso más que ha sido una excelente compra, me facilita mu...
12	15/02/2019	Producción	Interno	No he podido ver a mi familia en varios días. Hemos pedido a nuestro]...
13	15/02/2019	Atención al cliente	Twitter	He tardado mucho tiempo en contactar con un agente. Péssimo servici...
14	15/02/2019	Producción	Interno	Nuestro jefe de producción no tolera ninguna sugerencia. De los peore...
15	15/02/2019	Atención al cliente	Facebook	Jamás recibí respuesta al correo electrónico que envié. No compraré n...
16	18/02/2019	Producción	Interno	Ya era hora de que se implantaran los turnos rotativos. Ahora podré di...
17	20/02/2019	Logística	Twitter	Me aseguraron tener mi producto en un día en casa y así fue. ¡Genial!
18	24/02/2019	Calidad	Interno	Se deben realizar controles más estrictos de calidad. ¡Es inadmisible q...
19	25/02/2019	Calidad	Facebook	Estoy contento con el producto comprado, pero he visto que tiene alg...
20	28/02/2019	Logística	Interno	La mayoría de estanterías del almacén estan casi rotas. Necesitamos C...
21	02/03/2019	Producción	Interno	Hemos superado nuestro record de producción en una semana. ¡Enho...
22	04/03/2019	Atención al cliente	Facebook	Es decepcionante encontrarse con personas maleducadas en vuestro s...
23	07/03/2019	Producción	Interno	Me encanta trabajar con mis compañeros de turno, pero al jefe de pro...

Ilustración 71. Análisis de sentimientos. Carga previa

En primera instancia, aunque no es el caso, debemos depurar los datos. En primer lugar, debemos comprobar que los “id” (identificadores únicos, “Primary Key”) de los mensajes no están duplicados ni presentan errores. Para ello, debemos seleccionar la columna y con clic derecho, “Quitar duplicados” y “Quitar errores” (Ilustración 72). Lo mismo debemos hacer con la columna de “text”.

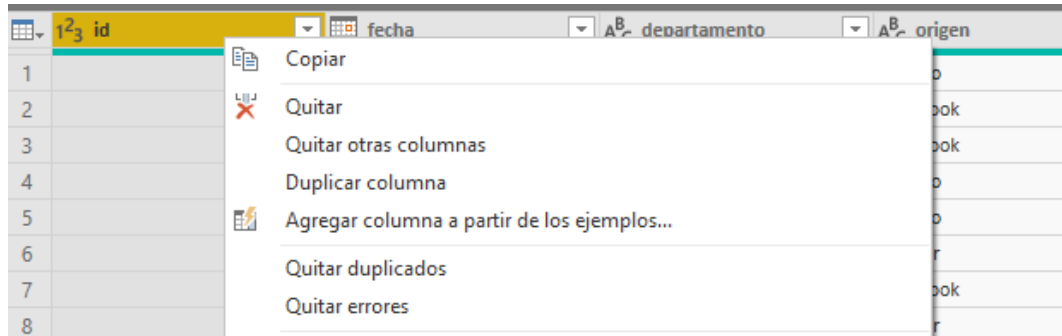


Ilustración 72. Análisis de sentimientos. Purgado de datos

Para realizar sendos análisis, necesitamos tres funciones de “Consulta en blanco”. Éstas, realizarán de forma automática lo que nosotros mismos podemos hacer manualmente a través de la web de Cognitive Services: introducir el texto a analizar y obtener los resultados que deseamos. Es la gran ventaja de disponer de APIs en estos servicios.

En primer lugar, requerimos de una función que detecte el lenguaje del mensaje, la cual llamaremos “IdentificaLenguaje”. Esta función devuelve en formato ISO el lenguaje del mensaje (“es” para español, “en” para inglés, etc.) La creamos y accedemos al “Editor Avanzado” para introducir el código mostrado en la Ilustración 73.

IdentificaLenguaje

Opciones de presentación ▾

```
(text) => let
  apikey = "08833ca19dbd4b03b547526f4d480a4c",
  endpoint = "https://westcentralus.api.cognitive.microsoft.com/text/analytics/v2.0/languages",
  jsontext = Text.FromBinary(Json.FromValue(Text.Start(Text.Trim(text), 500))),
  jsonbody = "{ documents: [ { language: \"en\", id: \"0\", text: \" & jsontext & \" } ] }",
  bytesbody = Text.ToBinary(jsonbody),
  headers = [#"Ocp-Apim-Subscription-Key" = apikey],
  bytesresp = Web.Contents(endpoint, [Headers=headers, Content=bytesbody]),
  jsonresp = Json.Document(bytesresp),
  language = jsonresp[documents][0][detectedLanguages][0][iso6391Name]
in language
```

Ilustración 73. Análisis de sentimientos. Función IdentificaLenguaje

La siguiente función que creamos será “PalabrasClave” (Ilustración 74), la cual nos devolverá en una nueva columna las palabras clave de los mensajes.

A mayores, crearemos otra función que genere las palabras clave de cada mensaje, en la que a diferencia de lo anterior, lo almacenará en una “Lista” (Ilustración 75). Esto nos será útil para un análisis específico por palabras. De la anterior forma obtendríamos un conjunto de palabras por mensaje que no podemos analizar individualmente, si no solo como conjunto. Es decir, si en un mensaje tenemos tres palabras, solo podremos analizar, por ejemplo, si ese conjunto de palabras aparece más veces, lo cual es inútil para nuestros propósitos.

PalabrasClave

Opciones de presentación ▾

```
(text) => let
  apikey = "08833ca19dbd4b03b547526f4d480a4c",
  idioma = IdentificaLenguaje(text),
  endpoint = "https://westcentralus.api.cognitive.microsoft.com/text/analytics/v2.1/keyPhrases",
  jsontext = Text.FromBinary(Json.FromValue(Text.Start(Text.Trim(text), 5000))),
  jsonbody = "{ documents: [ { language: "" & idioma & "", id: ""0"", text: " & jsontext & " } ] }",
  bytesbody = Text.ToBinary(jsonbody),
  headers = [{"Ocp-Apim-Subscription-Key" = apikey}],
  bytesresp = Web.Contents(endpoint, [Headers=headers, Content=bytesbody]),
  jsonresp = Json.Document(bytesresp),
  keyphrases = Text.Lower(Text.Combine(jsonresp[documents]{0}[keyPhrases], ", "))
in keyphrases
```

Ilustración 74. Análisis de sentimientos. Función PalabrasClave

PalabrasClaveLista

Opciones de presentación ▾

```
// Returns key phrases from the text as a list object
(text) => let
  apikey = "08833ca19dbd4b03b547526f4d480a4c",
  idioma = IdentificaLenguaje(text),
  endpoint = "https://westcentralus.api.cognitive.microsoft.com/text/analytics/v2.1/keyPhrases",
  jsontext = Text.FromBinary(Json.FromValue(Text.Start(Text.Trim(text), 5000))),
  jsonbody = "{ documents: [ { language: "" & idioma & "", id: ""0"", text: " & jsontext & " } ] }",
  bytesbody = Text.ToBinary(jsonbody),
  headers = [{"Ocp-Apim-Subscription-Key" = apikey}],
  bytesresp = Web.Contents(endpoint, [Headers=headers, Content=bytesbody]),
  jsonresp = Json.Document(bytesresp),
  keyphrases = jsonresp[documents]{0}[keyPhrases]
in keyphrases
```

Ilustración 75. Análisis de sentimientos. Función PalabrasClaveLista

Observamos que “idioma”, en sendas funciones, está definido por la función “IdentificaLenguaje” que previamente creamos y su objeto es la columna “text”.

Por último, creamos la función “Resultados” (Ilustración 76) para la valoración numérica del mensaje.

Resultados

Opciones de presentación ▾

```
(text) => let
  apikey = "08833ca19dbd4b03b547526f4d480a4c",
  idioma = IdentificaLenguaje(text),
  endpoint = "https://westcentralus.api.cognitive.microsoft.com/text/analytics/v2.0/sentiment",
  jsontext = Text.FromBinary(Json.FromValue(Text.Start(Text.Trim(text), 5000))),
  jsonbody = "{ documents: [ { language: "" & idioma & "", id: ""0"", text: " & jsontext & " } ] }",
  bytesbody = Text.ToBinary(jsonbody),
  headers = [{"Ocp-Apim-Subscription-Key" = apikey}],
  bytesresp = Web.Contents(endpoint, [Headers=headers, Content=bytesbody]),
  jsonresp = Json.Document(bytesresp),
  sentiment = jsonresp[documents]{0}[score]
in sentiment
```

Ilustración 76. Análisis de sentimientos. Función Resultados.

Ahora, crearemos las columnas con los resultados de las funciones (Ilustración 77). Agregamos la columna nueva “Palabras Clave” mediante “Invocar función personalizada”. Lo mismo hacemos con “Resultados análisis”:

Invocar función personalizada

Invoque una función personalizada definida en este archivo para cada fila.

Nuevo nombre de columna
Palabras Clave

Consulta de función
PalabrasClave

text (opcional)
text

Aceptar Cancelar

Ilustración 77. Análisis de sentimientos. Invocar funciones.

Por otra parte, “PalabrasClaveLista” ha de ser invocada en otra tabla idéntica a la inicial que duplicaremos mediante la herramienta “**Referenciar**”, ya que al invocarla y crear la lista con las palabras individuales, duplicara cada uno de los “id” y todos los datos asociados para cada una de las palabras clave.

Por último, debemos establecer que la columna de “**Resultados análisis**” es de tipo “**Número decimal**”, ya que por defecto, es de tipo texto. Teniendo ambas filas nuevas generadas, debemos “**Cerrar, y Aplicar**” para entrar a la interfaz de análisis y creación de tableros de Power BI.

6.3 Planteamiento y análisis

Al finalizar el anterior proceso obtendremos las tablas mostradas en las Ilustraciones 78 y 79.

En este momento, ya podemos iniciar el análisis de los datos que tenemos. En primer lugar, instalaremos el objeto visual “**WordCloud**” para hacer un análisis preliminar de las palabras y la frecuencia con las que aparecen. Además, se filtrado en las opciones las palabras no relevantes como la preposición “de” para obtener el elemento de la Ilustración 80.

Una visualización general nos permite ver que ciertas palabras aparecen tanto en singular y plural (e incluso con errores de escritura, véase ‘profuctos’ o ‘atención’). Esto, es un defecto de este tipo de análisis que requiere del razonamiento humano para determinar que hacen referencia a lo mismo, y pueden ser considerados iguales.

En la fase de preparación de datos, podría haberse considerado una transformación general de todas las palabras a singular. Sin embargo, los recursos y la dificultad del problema (complejidad de la lengua española para formar plurales y un código que tenga en cuenta este factor) hacen más fácil el proceso de análisis visual para determinar qué palabras requieren de transformación.

id	fecha	departamento	origen	text	PalabrasClave	Resultados Análisis
1	viernes, 1 de febrero de 2019	Producción	Interno	Tenemos malos horarios y no podemos compaginar familia y trabajo	familia, malos horarios, turnos rotativos, trabajo	0,229432821273804
2	domingo, 3 de febrero de 2019	Logística	Facebook	Envío recibido en solo un día. Increíble	envío recibido, día	0,527485013008118
3	lunes, 4 de febrero de 2019	Calidad	Facebook	El producto comprado tiene defectos a simple vista. No volveré a comprar	simple vista, producto comprado, defectos	0,574985563755095
4	lunes, 4 de febrero de 2019	Producción	Interno	El trato recibido por el jefe de producción es nefasto. Se necesitan cambios	jefe de producción, trato recibido, cambios	0,339923590421677
5	martes, 5 de febrero de 2019	Calidad	Interno	Utilizamos las mejores y más novedosas técnicas del mercado. Es un placer	novedosas técnicas, mercado, placer	1
6	jueves, 7 de febrero de 2019	Atención al cliente	Twitter	Me encanta que mis dudas se resuelvan tan rápido. Gracias	dudas, gracias	0,80926966671753
7	viernes, 8 de febrero de 2019	Atención al cliente	Facebook	Llamé para preguntar sobre el estado de mi pedido y la persona que me atendió me resolvió todas las dudas. Gracias	persona, estado, pedido, grosera	0,0454545468091965
8	domingo, 10 de febrero de 2019	Calidad	Twitter	Los acabados del producto son de primera calidad. Mi enhorabuena	primera calidad, departamento de calidad, acabados, productos, enhorabuena	1
9	martes, 12 de febrero de 2019	Logística	Interno	Completamente inseguro trabajar en la zona de salidas y expediciones	zona de salidas, expediciones, carretilleras	0,310216873884201
10	miércoles, 13 de febrero de 2019	Atención al cliente	Interno	Las jornadas de trabajo son muy largas y más de 8 horas. Los descansos son cortos. Jornadas de trabajo, presión, horas	encargado, descansos cortos, jornadas de trabajo, presión, horas	0,52143305400848
11	viernes, 15 de febrero de 2019	Calidad	Twitter	Cada día pienso más que ha sido una excelente compra, me facilitó la vida, excelente compra, día	vida, excelente compra, día	1
12	viernes, 15 de febrero de 2019	Producción	Interno	No he podido ver a mi familia en varios días. Hemos pedido a un jefe de producción, turnos rotativos, familia, días	jefe de producción, turnos rotativos, familia, días	0,245338633656502
13	viernes, 15 de febrero de 2019	Atención al cliente	Twitter	He tardado mucho tiempo en contactar con un agente. Péximo servicio, parte, tiempo, agente	péximo servicio, parte, tiempo, agente	0,189710736274719
14	viernes, 15 de febrero de 2019	Producción	Interno	Nuestro jefe de producción no tolera ninguna sugerencia. De los jefes de producción, sugerencia, peores encargados	jefe de producción, sugerencia, peores encargados	0,35219466662488
15	viernes, 15 de febrero de 2019	Atención al cliente	Facebook	Jamás recibí respuesta al correo electrónico que envié. No compra respuesta, correo electrónico, empresa	respuesta, correo electrónico, empresa	0,39635303783417
16	lunes, 18 de febrero de 2019	Producción	Interno	Ya era hora de que se implantaran los turnos rotativos. Ahora podemos trabajar en un día en casa y así fue. Genial	días, semana, familia, turnos rotativos, hora, gracias	0,88676744694781
17	miércoles, 20 de febrero de 2019	Logística	Twitter	Me aseguraron tener mi producto en un día en casa y así fue. Genial	producto, día, casa	0,890904724597931
18	domingo, 24 de febrero de 2019	Calidad	Interno	Se deben realizar controles más estrictos de calidad. Es inadmisible	productos, mercado, defectos detectados, controles, calidad	0,689599692821503
19	lunes, 25 de febrero de 2019	Calidad	Facebook	Estoy contento con el producto comprado, pero he visto que tiene defectos superficiales	producto comprado, defectos superficiales	0,888901715278625
20	jueves, 28 de febrero de 2019	Logística	Interno	La mayoría de estanterías del almacén están casi rotas. Necesitamos un almacén, mayoría de estanterías	almacén, mayoría de estanterías	0,59366762638092
21	sábado, 2 de marzo de 2019	Producción	Interno	Hemos superado nuestro record de producción en una semana. El record de producción, semana, enhorabuena	record de producción, semana, enhorabuena	1
22	lunes, 4 de marzo de 2019	Atención al cliente	Facebook	Es decepcionante encontrarse con personas maleducadas en vuestras oficinas	personas maleducadas, servicio de atención, cliente, queja	0,10301873087883
23	jueves, 7 de marzo de 2019	Producción	Interno	Me encanta trabajar con mis compañeros de turno, pero el jefe de producción es un problema	jefe de producción, compañeros de turno, cambio	0,75711661577247
24	domingo, 10 de marzo de 2019	Producción	Interno	La máquina número tres tiene una alta tasa de defectos y es la más mala	máquina número, alta tasa de defectos, acciones, productos	0,670984387397756
25	miércoles, 13 de marzo de 2019	Atención al cliente	Interno	Las quejas por una compañera han sido numerosas y el encargado no ha respondido	encargado, compañera, quejas, directiva, problema	0,38877063986393
26	viernes, 15 de marzo de 2019	Logística	Twitter	Hice mi pedido por la tarde y tuve mi producto en un día en casa. Excelente servicio	producto, tarde, día, casa, pedido	1
27	sábado, 16 de marzo de 2019	Logística	Facebook	El pedido debía llegar en un día pero llegó en dos días. Al menos si el pedido llegara en un día sería un gran éxito	pedido, día, casa, casa, casa, casa	0,861543416978929
28	sábado, 16 de marzo de 2019	Producción	Interno	Seguimos con un ritmo excepcional de producción pero se debería mejorar	producción, problemas, ritmo excepcional, máquina	0,633909250831604

Ilustración 78. Análisis de Sentimientos. Tabla con palabras clave en cadena



id	fecha	departamento	origen	text	PalabrasClave	Resultados Análisis	Palabras en Lista
45	Jueves, 16 de mayo de 2019	Atención al cliente	Facebook	Es la segunda vez que contacto y ya han arreglado sus problemas	problemas, correos electrónicos, eficaz, emhorabuena	0,543596923351288	1 emhorabuena
45	Jueves, 16 de mayo de 2019	Atención al cliente	Facebook	Es la segunda vez que contacto y ya han arreglado sus problemas	problemas, correos electrónicos, eficaz, emhorabuena	0,543596923351288	1 eficaz
45	Jueves, 16 de mayo de 2019	Atención al cliente	Facebook	Es la segunda vez que contacto y ya han arreglado sus problemas	problemas, correos electrónicos, eficaz, emhorabuena	0,543596923351288	1 correos elect
42	miércoles, 1 de mayo de 2019	Atención al cliente	Twitter	Servicio rápido y de calidad. La atención al cliente resuelve tus dudas	alternativas, servicio rápido, calidad, minutos, atención, ci	0,543596923351288	1 problemas
42	miércoles, 1 de mayo de 2019	Atención al cliente	Twitter	Servicio rápido y de calidad. La atención al cliente resuelve tus dudas	alternativas, servicio rápido, calidad, minutos, atención, ci	0,543596923351288	1 problema
42	miércoles, 1 de mayo de 2019	Atención al cliente	Twitter	Servicio rápido y de calidad. La atención al cliente resuelve tus dudas	alternativas, servicio rápido, calidad, minutos, atención, ci	0,543596923351288	1 cliente
42	miércoles, 1 de mayo de 2019	Atención al cliente	Twitter	Servicio rápido y de calidad. La atención al cliente resuelve tus dudas	alternativas, servicio rápido, calidad, minutos, atención, ci	0,543596923351288	1 atención
42	miércoles, 1 de mayo de 2019	Atención al cliente	Twitter	Servicio rápido y de calidad. La atención al cliente resuelve tus dudas	alternativas, servicio rápido, calidad, minutos, atención, ci	0,543596923351288	1 minutos
42	miércoles, 1 de mayo de 2019	Atención al cliente	Twitter	Servicio rápido y de calidad. La atención al cliente resuelve tus dudas	alternativas, servicio rápido, calidad, minutos, atención, ci	0,543596923351288	1 calidad
42	miércoles, 1 de mayo de 2019	Atención al cliente	Twitter	Servicio rápido y de calidad. La atención al cliente resuelve tus dudas	alternativas, servicio rápido, calidad, minutos, atención, ci	0,543596923351288	1 Servicio rápid
42	miércoles, 1 de mayo de 2019	Atención al cliente	Twitter	Servicio rápido y de calidad. La atención al cliente resuelve tus dudas	alternativas, servicio rápido, calidad, minutos, atención, ci	0,543596923351288	1 alternativas
34	miércoles, 3 de abril de 2019	Atención al cliente	Facebook	Contactar con vosotros por correo ha sido un infierno. He tenido q	infierno, correos electrónicos, días	0,184377357363701	0 días
34	miércoles, 3 de abril de 2019	Atención al cliente	Facebook	Contactar con vosotros por correo ha sido un infierno. He tenido q	infierno, correos electrónicos, días	0,184377357363701	0 correos elect
33	sábado, 30 de marzo de 2019	Atención al cliente	Twitter	Es de muy mal gusto esperar media hora para hablar con un ageni	media hora, atención, mal gusto, empresa, cliente, agente, salu	0,232155039906502	1 saludo
33	sábado, 30 de marzo de 2019	Atención al cliente	Twitter	Es de muy mal gusto esperar media hora para hablar con un ageni	media hora, atención, mal gusto, empresa, cliente, agente, salu	0,232155039906502	1 agente
33	sábado, 30 de marzo de 2019	Atención al cliente	Twitter	Es de muy mal gusto esperar media hora para hablar con un ageni	media hora, atención, mal gusto, empresa, cliente, agente, salu	0,232155039906502	1 cliente
33	sábado, 30 de marzo de 2019	Atención al cliente	Twitter	Es de muy mal gusto esperar media hora para hablar con un ageni	media hora, atención, mal gusto, empresa, cliente, agente, salu	0,232155039906502	1 empresa
33	sábado, 30 de marzo de 2019	Atención al cliente	Twitter	Es de muy mal gusto esperar media hora para hablar con un ageni	media hora, atención, mal gusto, empresa, cliente, agente, salu	0,232155039906502	1 mal gusto
33	sábado, 30 de marzo de 2019	Atención al cliente	Twitter	Es de muy mal gusto esperar media hora para hablar con un ageni	media hora, atención, mal gusto, empresa, cliente, agente, salu	0,232155039906502	1 atención
22	lunes, 4 de marzo de 2019	Atención al cliente	Facebook	Es decepcionante encontrarse con personas maleducadas en vuesi	personas maleducadas, servicio de atención, cliente, queja	0,10301873087883	0 queja
22	lunes, 4 de marzo de 2019	Atención al cliente	Facebook	Es decepcionante encontrarse con personas maleducadas en vuesi	personas maleducadas, servicio de atención, cliente, queja	0,10301873087883	0 cliente
22	lunes, 4 de marzo de 2019	Atención al cliente	Facebook	Es decepcionante encontrarse con personas maleducadas en vuesi	personas maleducadas, servicio de atención, cliente, queja	0,10301873087883	0 servicio de at
15	viernes, 15 de febrero de 2019	Atención al cliente	Facebook	Jamás recibí respuesta al correo electrónico que envié. No compra	personas maleducadas, servicio de atención, cliente, queja	0,396335303783417	0 personas mal
15	viernes, 15 de febrero de 2019	Atención al cliente	Facebook	Jamás recibí respuesta al correo electrónico que envié. No compra	personas maleducadas, servicio de atención, cliente, queja	0,396335303783417	0 empresa
15	viernes, 15 de febrero de 2019	Atención al cliente	Facebook	Jamás recibí respuesta al correo electrónico que envié. No compra	personas maleducadas, servicio de atención, cliente, queja	0,396335303783417	0 correo electr

Ilustración 79. Análisis de sentimientos. Tabla con una palabra por fila (id repetido)



- conocer sobre el rendimiento de los clientes y opiniones de trabajadores y clientes.
2. **Departamentos.** Continuando hacia un contexto más específico, debemos conocer el estado puntual de cada departamento. Debe proporcionarnos, para cada departamento información sobre su estado, su desempeño y la opinión o información proporcionada por clientes y trabajadores.
 3. **Orígenes.** De la misma manera que se puede hacer un estudio en cuanto departamentos, es lógico realizarlo recíprocamente respecto de orígenes, para tratar de encontrar patrones sobre los usuarios de cada plataforma.
 4. **Palabras.** La parte fundamental de este análisis está en las palabras. Por ello, es vital conocer su origen, su valoración y su frecuencia. A través de esta página debemos ser capaces de obtener información sobre sucesos en cada departamento y su impacto de cara al trabajador y los sucesos.
 5. **Evolución.** En nuestra última página, se visualizará la evolución de cada una de estas palabras, de manera que nos permita conocer el estado de los problemas que han denotado las palabras de la anterior página. Debemos extraer conclusiones sobre la resolución o estado de lo que trabajadores o clientes han detectado o exigido.

GENERAL

Para cumplir el propósito de esta página se ha deducido que para cualquier persona que dispone de escaso tiempo para analizar en profundidad los datos (directiva, encargados, etc.) la importancia reside en la concentración de la máxima información en pocos elementos.

Por ello, se plantean únicamente tres elementos visuales. En primer lugar, la **nube de palabras**. Al igual que la nube de palabras preliminar que creamos para tener una referencia general de aquello que más se repite, ahora establecemos el mismo con los datos depurados.

En segundo lugar, la **evaluación del desempeño de cada departamento**. Para ello, como ya mencionamos, cada mensaje es valorado en una escala de cero a uno (0-1) donde 0 denota un mensaje muy negativo y uno, muy positivo. Para ello, creamos un gráfico con la media de valoraciones de los mensajes pertenecientes a cada departamento. Proporciona un método eficaz de conocer dónde se necesitan medidas correctoras, o en caso contrario, de reforzamiento o recompensa.

Por último, el **desempeño de la organización a lo largo del periodo** de tiempo considerado. De manera general, también se puede desear conocer si el estado de la empresa desde la opinión externa o interna está mejorando o empeorando a lo largo del tiempo. Por ello, se establece un gráfico temporal por meses en el que se muestra la valoración media de los mensajes de cada mes, independientemente del origen o departamento. A la vez que puede detectar tendencias, puede detectar anomalías, las cuales en las siguientes páginas, si se desea, podrán ser analizadas en profundidad.

El resultado de esta propuesta se muestra en la Ilustración 82.



DEPARTAMENTOS

El análisis departamental se centra en aquella información que permite descubrir (además de cuantitativamente) el rendimiento departamental. Para ello se han establecido cuatro elementos visuales, tal y como muestra la Ilustración 83.

En primer lugar, es indispensable una manera de valorar numéricamente el desempeño de éstos. Se ha creado un **gráfico de barras con el promedio de las valoraciones** de los mensajes para cada departamento. Por ende, se puede diferenciar de manera clara y visual aquellos departamentos que están siendo bien gestionados según las opiniones.

Por otra parte, también resulta interesante para el responsable de analizar los datos, el origen de los mensajes y por tanto, también se ha creado un **gráfico de sectores de los orígenes de los mensajes** de cada departamento. Así, es posible conocer el grado de implicación interna por parte de los empleados, además de los departamentos que son más reclamados por entes externos (incluyendo la diferenciación por redes).

Siendo complementario al anterior gráfico, se dispone de un gráfico temporal para conocer la **distribución de llegada de mensajes por mes y por departamento**, lo cual es útil para comprobar si existe alguna relación entre el momento y las valoraciones.

La parte fundamental de esta página reside en el último elemento visual. Se ha dispuesto de un **gráfico de barras para conocer qué palabras son las más repetidas y su reparto por departamento**. Es decir, se han escogido aquellas más repetidas de manera global y se distribuye su número por los departamentos. Esto resulta esencial para encontrar patrones entre cambios en valoraciones, cantidad de mensajes e incluso, entender las fuentes de origen.

ORÍGENES

A pesar de que la intuición haga pensar que los orígenes de los mensajes no son influyentes, cabe la posibilidad de que posean correlaciones con aspectos como departamento de destino o valoración. Por ello se presenta en esta página tres elementos visuales.

En primer lugar, un **gráfico de sectores para mostrar en función del origen, el departamento de destino**, es decir, el gráfico recíproco de la anterior página de departamentos. La distribución nos permitirá conocer la actividad de los usuarios con nuestras páginas.

En segundo lugar, se presenta un **gráfico de valoraciones medias según los tres orígenes**. Conocer quiénes son los más críticos o favorables con nuestra organización, facilita centrar los esfuerzos para mejorar la opinión de ese nicho (ya sea mediante promociones, estrategias de marketing, etc.).

Por último, se presenta un **gráfico de columnas y línea** en el que las columnas representan la frecuencia por origen y la línea la valoración media de los mensajes en los que aparece la palabra. Permite conocer qué tipo de usuarios han hecho referencia a dichos términos en mayor medida, a la vez que su impacto.

El resultado se puede visualizar en la Ilustración 84.



Ilustración 83. Análisis de sentimientos. Página 2: Departamentos

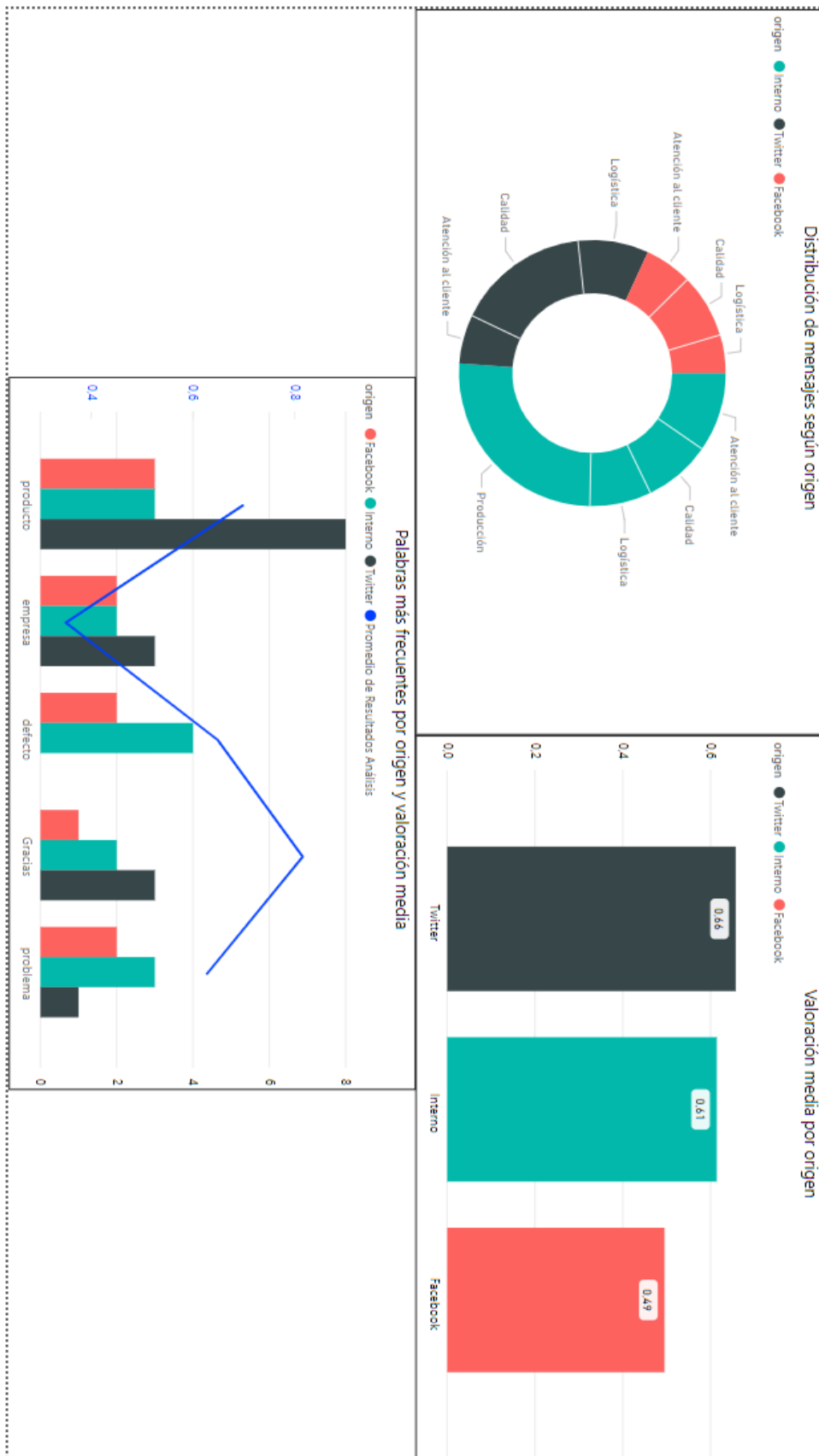


Ilustración 84. Análisis de sentimientos. Página 3: Orígenes

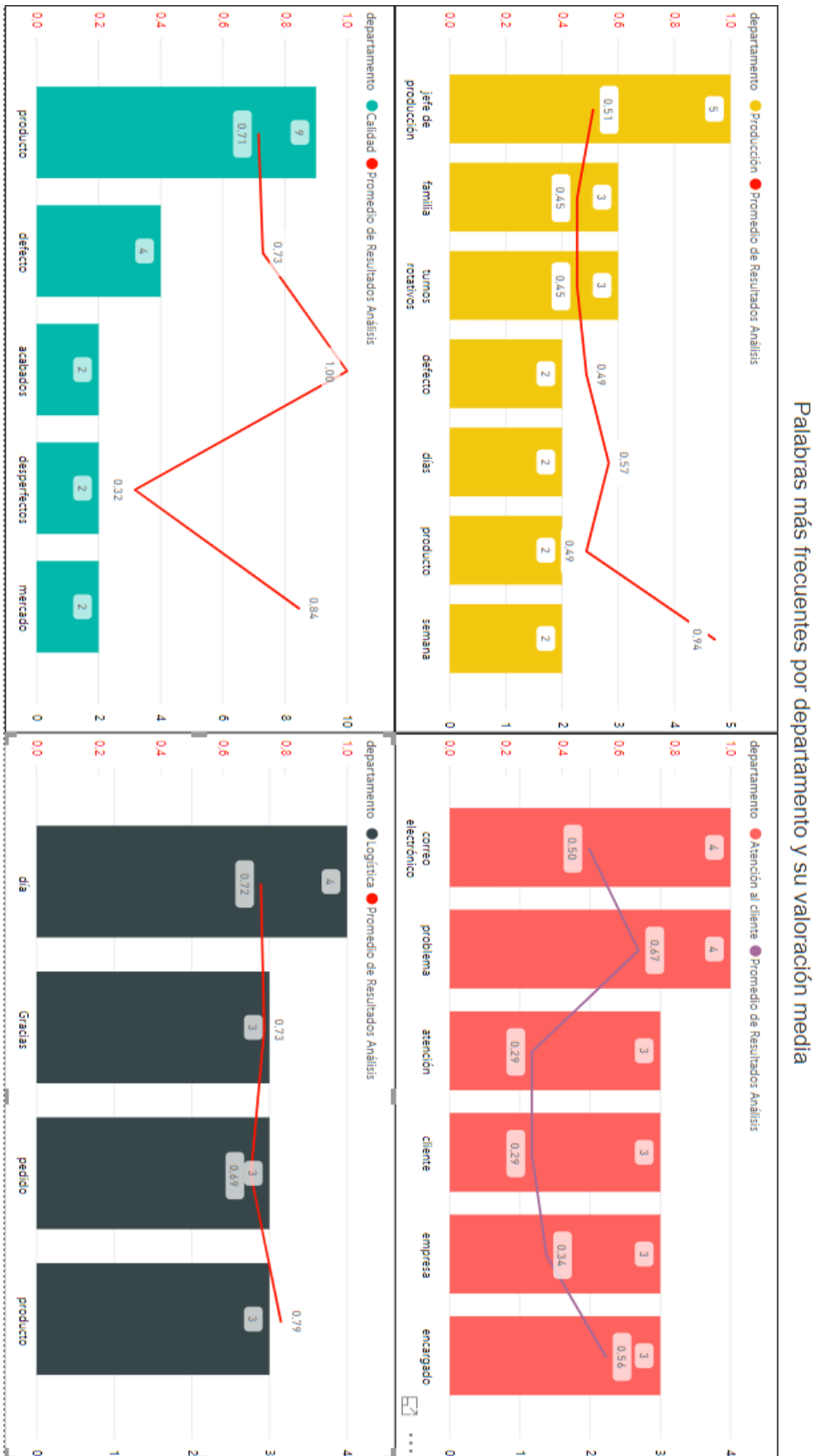


Ilustración 85. Análisis de sentimientos. Página 4: Palabras

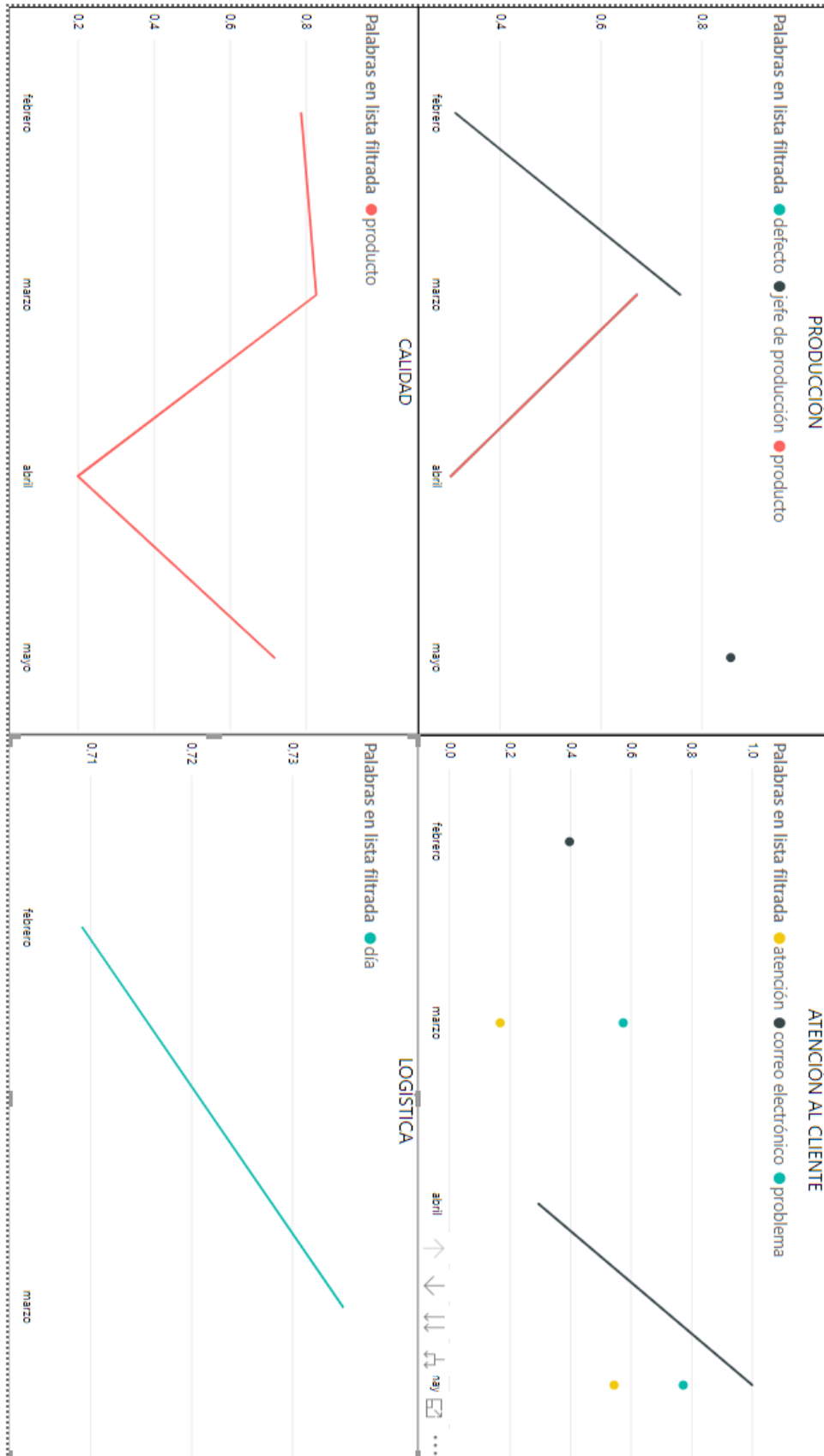


Ilustración 86. Análisis de sentimientos. Página 5: Evolución



PALABRAS

Es importante entender que las páginas se han dispuesto de manera que la información es cada vez más específica, y supone mayor necesidad de análisis humano de los datos en cada una de ellas.

Por ello, en esta página (Ilustración 85) se ha expandido la información del último gráfico de la anterior página. Se ha dispuesto departamentalmente, las palabras más repetidas ilustrando mediante las barras su frecuencia, y en las líneas su valoración media.

Es vital la necesidad de un humano para entender la correlación entre la palabra, el promedio y el departamento en el que aparece. Además, también se necesita de la visión crítica de un humano para detectar si el análisis ofrecido por Cognitive Services es correcto. Es decir, comprobar si es coherente una valoración promedio alta para una palabra evidentemente negativa. A la vez que profundizamos en la información y somos capaces de entender más información, también aparece la posibilidad de detectar errores o incoherencias que nos permiten depurar los datos aún más, o encontrar anomalías.

EVOLUCIÓN

A través del buen criterio del analista se presentan para cada departamento la palabras más relevantes (ya sea por frecuencia o por especial interés en ellas) y la evolución de su nota media a lo largo de los meses.

Esta última página, mostrada en la Ilustración 85, es la que ha de servir en una última instancia para detectar problemáticas que han sido o no resueltas o han pasado desapercibidas para la directiva. Además de un carácter causativo, esta página puede usarse con carácter preventivo para adoptar las medidas convenientes (las puestas en marchas para solventar la problemática que surgió, por ejemplo) para que no se vuelva a repetir. Como es evidente, la mano humana en la elección de las palabras relevantes es vital para hacer el análisis óptimo.

6.4 Resultados del análisis

Comencemos a analizar cada una de las páginas en búsqueda de información relevante para nuestro negocio. En primer lugar, como es lógico, acudimos a la página General. En ella, las palabras que más nos llaman la atención en la nube de palabras (Ilustración 80) son “producto”, “producción”, “defecto”, “empresa”, “gracias” y “problema”. A primera vista, los términos “defecto” y “problema” son alarmantes debido a sus connotaciones negativas. Por otra parte, “gracias” tiene unas claras connotaciones positivas que puede hacernos pensar que cualquier “defecto” o “problema” ha sido solventado.

En cuanto al rendimiento de los departamentos en la Ilustración 87, es esperanzador que el departamento de Calidad tenga una nota promedio tan alta (0,74) y es preocupante la baja nota de Atención al cliente (0,46).

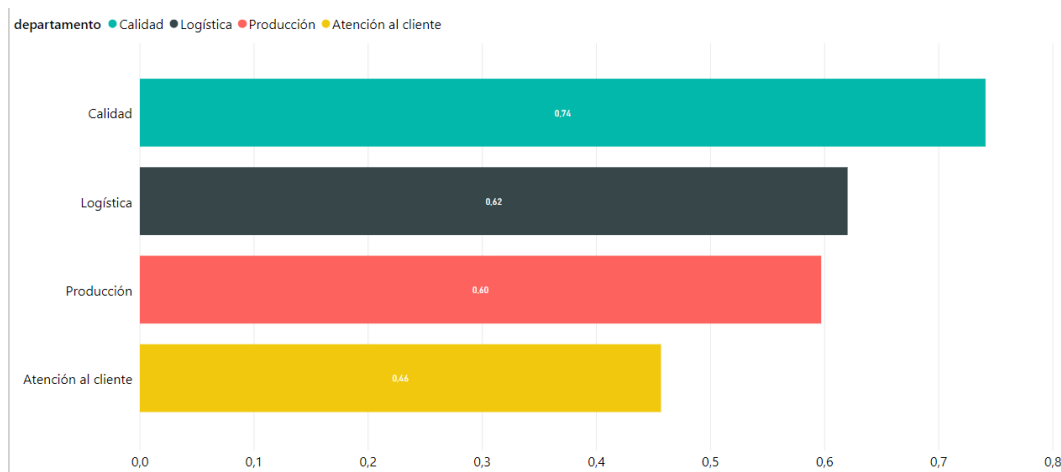


Ilustración 87. Análisis de sentimientos. Valoración por departamento

Por otra parte, en el gráfico de desempeño general (visible en la Ilustración 82) hay una **evidente bajada en el mes de abril** que ha de ser considerado y analizado en profundidad.

Pasando a la página de Departamentos, lo primero que llama la atención es la subida en valoración media de los cuatro departamentos en el último mes de mayo (Ilustración 88). Por tanto, podemos concluir que **cualquier problemática existente ha sido solventada** (a falta de indicios que exista alguna en el propio mes de mayo) y nuestro **análisis** se convierte en uno plenamente de **detección de hechos pasados**. Es destacable el gran desempeño del departamento de Atención al cliente durante el mes de mayo, llegando a rozar la nota de 0,9. Es evidente que ha solventado cualquier cuestión ha sido solucionada.

Por otra parte, destacar una mala nota inicial en el mes de febrero en el departamento de Producción con una brusca subida en Marzo y otra bajada en Abril, que ha de ser estudiada.

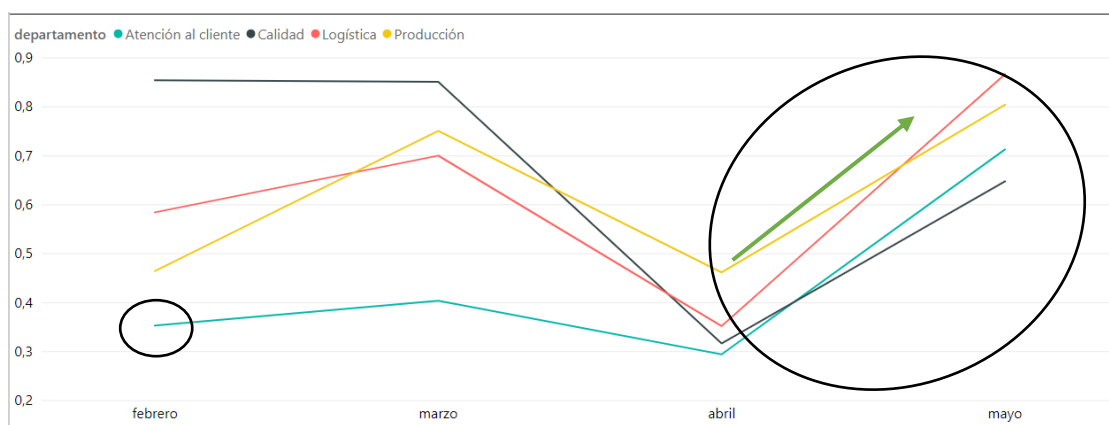


Ilustración 88. Análisis de sentimientos. Valoración departamento por mes

La cantidad de mensajes es más alta durante Febrero debido a, probablemente, los periodos navideños en los que más ventas se ha tenido y, después de un tiempo de prueba, los clientes han decidido poner una opinión sobre el producto o servicio prestado.

Por otra parte, en cuanto a las palabras más frecuentes en la Ilustración 89, lo **más sorprendente que se puede mencionar es:**

- ✓ “Producto”, como es lógico, aparece de forma mayoritaria nueve (9) veces en el departamento de Calidad frente a las tres (2) y dos (2) en Logística y Producción, respectivamente. La buena nota en promedio del departamento Calidad y la cantidad de veces que aparece en él, hace que pensemos que nuestro producto genera unas muy buenas sensaciones.
- ✓ Curiosa la aparición de “jefe de producción” en la tabla y únicamente en el departamento de Producción. Se debe analizar.
- ✓ “Defecto” aparece en departamentos de Calidad y Producción, lo cual también ha de ser analizado. Sin embargo, es curioso como existe una nota alta global de Calidad a pesar de las veces que aparece “defecto”. Puede que sea la razón de la nota baja en Abril.
- ✓ “Día” aparece en Logística bastantes veces. Considerando que la empresa ofrece envíos en un día en sus productos, podemos intuir que se refieren a ello.

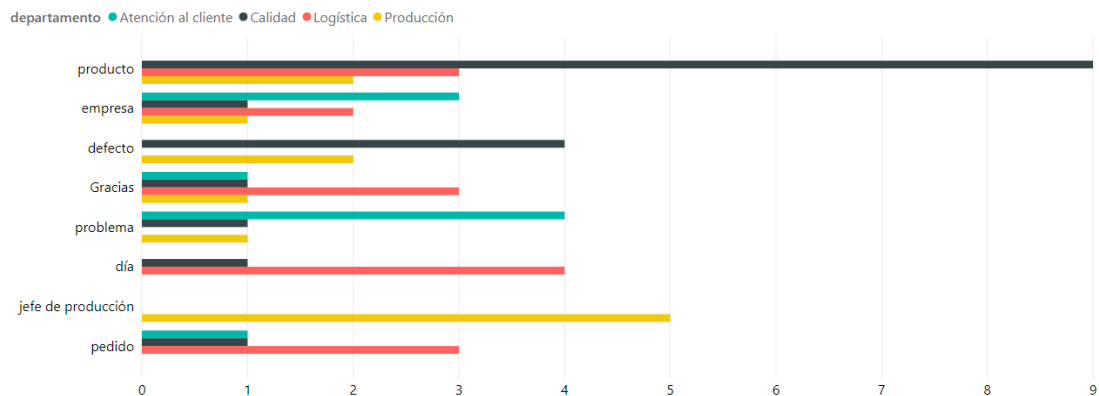


Ilustración 89. Análisis de sentimientos. Palabras más frecuentes

Respecto a los Orígenes (visible en la Ilustración 84), en primer lugar, el origen Interno supone más del 50% de los mensajes, lo que nos demuestra alta participación por parte de los empleados en estas iniciativas. Lo más atractivo, es la nota baja promedio que ofrecen los usuarios de Facebook. Esto, tal vez es debido a que, al menos en España, los usuarios activos de dicha plataforma son en términos generales mayores que los usuarios de Twitter y sus exigencias y críticas son más acusadas.

En la página de Palabras hemos de ir por partes. Comenzamos con el departamento de Producción. “Jefe de producción” tiene una nota media de 0,51 (Ilustración 90). Esto evidencia un problema con él. Si nos movemos a la página de Evolución, donde se ha escogido dicha palabra, vemos como en Febrero tuvo una nota de 0,31, pasando a 0,76 en Marzo y 0,86 en Mayo (Ilustración 91). Por supuesto, esto se refiere a un antiguo jefe de producción que debido a las quejas de los trabajadores fue despedido y los últimos mensajes evidencian el buen hacer del nuevo jefe. Esto está directamente correlado con “familia” y “turnos”, ya que la principal exigencia de los trabajadores eran los turnos rotativos para conciliar vida familiar y profesional.

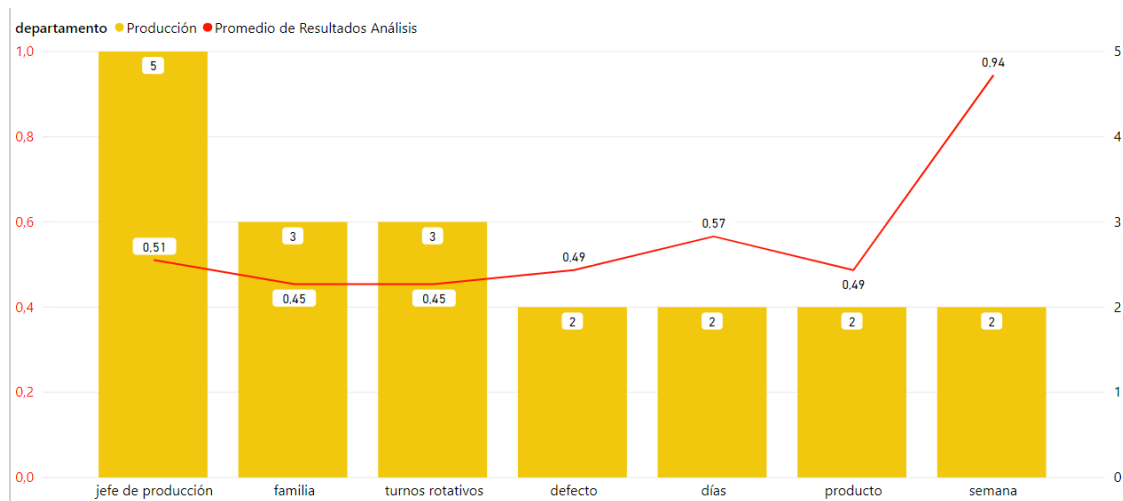


Ilustración 90. Análisis de sentimientos. Palabras de Producción

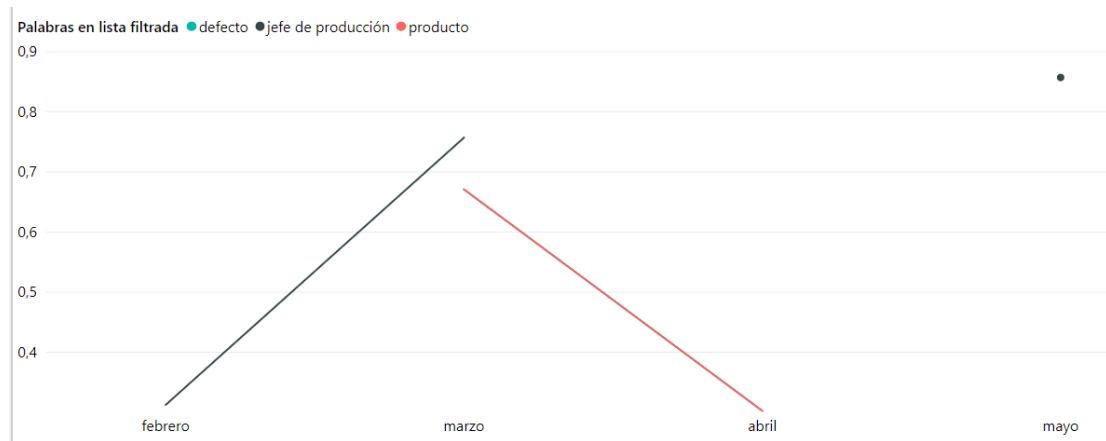


Ilustración 91. Análisis de sentimientos. Evolución palabras Producción

La palabra “Producto” es la palabra principal de Calidad con una nota media alta de 0,71 (visible en la Ilustración 84). Por otra parte, resulta curioso cómo en Producción tiene una nota media relativamente baja (Ilustración 90). Si escogemos dicha palabra para ver su evolución tanto en Producción como en Calidad, observamos una bajada idéntica en el mes de Abril para ambos departamentos (Ilustraciones 91 y 92). Esto pone en evidencia lo que ocurrió durante Marzo en la que se detectaron altas tasas de defectos en ciertas máquinas por parte de trabajadores y posteriormente salieron al mercado a pesar de ser defectuosos (comentarios negativos en Abril por parte de clientes). **Se observa como “defecto” y “producto” comparten pendiente en Producción.**

En Calidad, además, la palabra “desperfectos”, si la incluimos en la gráfica de Evolución aparece únicamente durante el mes de Abril, confirmando nuestras sospechas.

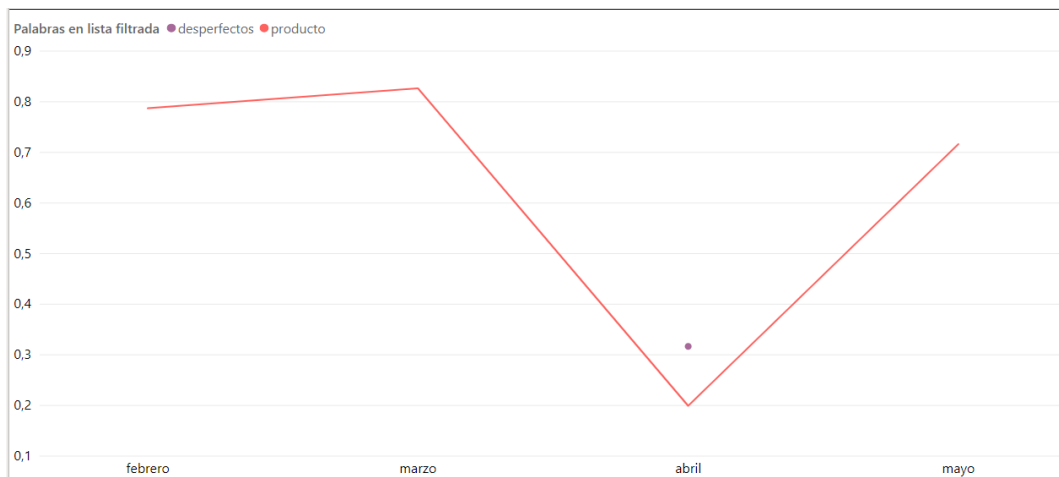


Ilustración 92. Análisis de sentimientos. Producto y desperfectos en Calidad

En términos generales, en Calidad, la palabra “producto” con una nota alta nos hace sentir seguros sobre el diseño y prestaciones de lo que producimos y por otra, la nota alta en la palabra “día”, “pedido” “producto” y “Gracias” confirma también que nuestros clientes están muy satisfechos por nuestro servicio de entrega en un solo día.

Por último, el departamento de Atención al cliente (Ilustración 93), las palabras más repetidas son “problema”, “correo electrónico” y “atención” y “cliente”. Consideramos que las dos últimas son palabras clave de “Atención al cliente”, ya que tienen exactamente la misma nota media. Es preocupante que el nombre del departamento tenga una nota tan baja.

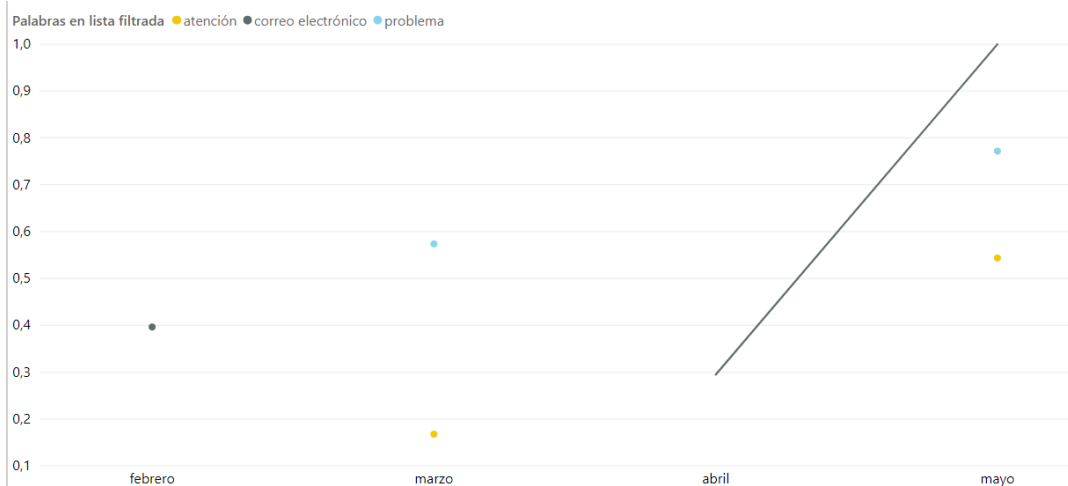


Ilustración 93. Análisis de sentimientos. Evolución palabras de Atención al cliente.



Si profundizamos en este hecho, descubrimos que hubo varias quejas por parte de clientes sobre una agente del servicio durante Marzo, tanto por parte de clientes como por parte de empleados. Las notas tan bajas del departamento, en términos generales son debido a los malos comentarios hacia dicha empleada. En el mes de Abril fue despedida y es evidente la mejora de la nota media de “atención al cliente”.

Por otra parte, “correo electrónico” tiene notas bajas durante los primeros meses, para finalmente tener una subida drástica en Mayo. Esto se entiende como un error en las páginas de contacto de la empresa donde los correos electrónicos no funcionaban. Las quejas fueron reportadas por los clientes en Febrero, pero debido a la inexistencia de un sistema como el actual, fue imposible descubrirlo hasta Mayo.



7. Conclusiones y extensiones

7.1 Conclusiones

La realización de este Trabajo de Fin de Grado nos ha permitido, tanto al autor como a cualquier lector, conocer de forma genérica el estado actual del Business Intelligence y sus posibilidades.

En una primera instancia, mediante un enfoque completamente teórico que se necesitaba para poseer una base de conocimientos en el ámbito, se ha puesto en contexto el porqué de la existencia y la necesidad de un sistema capaz de habilitar a los responsables de cualquier organización de tomar las decisiones basándose en la infinidad de datos poseídos en sus sistemas de almacenamiento. Posteriormente, se han descrito de manera científica, apoyándose en los principales autores de cada campo, la historia, componentes fundamentales, estrategias, ventajas o modelos de madurez, entre otros, del Business Intelligence.

Finalmente, se ha puesto en valor y en práctica toda esta información mediante la aplicación de un sistema de Business Intelligence en una empresa industrial de ámbito regional (PYME pequeña). A su vez, además de un análisis de datos estructurados, se ha implementado un novedoso análisis de mensajes de empleados y clientes (análisis de datos no estructurados conocido como Análisis de Sentimientos) para calificar, evaluar y descubrir información a través de ellos, con una metodología de estructuración previa. Además, dentro del apartado del análisis de sentimientos se ha incluido un apartado de enfoque teórico de las técnicas usadas en la obtención de datos a partir de redes sociales.

La implementación del sistema de BI en la compañía nos ha permitido averiguar información crítica que de otra manera habría sido conocer. Se destaca:

- ✓ La dominancia de los ingresos debidos al producto A en Grandes Superficies. Además, esta dominancia ha aparecido en los últimos dos meses. Por otra parte, se debe tratar de homogeneizar la venta de productos en otros canales, habiendo grandes diferencias entre canales por producto e incluso meses.
- ✓ Aumento progresivo de los inventarios de las MP. La tendencia positiva continua ha provocado que la MP X esté rozando el 90% de ocupación en el mes de Mayo, algo inadmisibile en cuanto a la planificación MRP. Se ha de revisar la planificación en busca de fallos o negligencias.
- ✓ El hecho más destacable de los proveedores son las desviaciones del proveedor 2 y proveedor 3 en el Lead Time (con retrasos de media de 3 y 12 horas, respectivamente) además de la entrega de los pedidos con retraso en más del 90% y 60% de los casos, respectivamente. Además, ambos proveedores son proveedores clave debido a las grandes cantidades que proveen. Por tanto, observando además que el proveedor 1 consigue adelantar incluso las fechas estimadas de media, y que sus precios son bastante parecidos a los competidores, resulta lógico pensar en convertirlo en un proveedor principal a costa de los proveedores 2 y 3.



- ✓ En cuanto a los gastos, el presupuesto ha sido superado en 2 de los 4 meses analizados. Además, se observó que el EBITDA es negativo en el mes de Marzo, el cual coincide con ser el mes en el que hubo un gasto considerablemente superior al presupuesto. Se han de controlar los gastos de una manera más estricta para evitar pérdidas.
- ✓ Los desajustes de MP en los procesos de producción son prácticamente por completo debido a los desajustes en los productos A y B, los cuales se fabrican en la Línea 1. Se necesitan controles más exhaustivos de las causas, sin descartar la posibilidad de hurtos o malas prácticas por parte de los operarios.
- ✓ La aparición de Accidente Leve y Accidente Grave como las causas más frecuentes de parada durante los meses de Marzo y Mayo, hacen que se deban realizar campañas de concienciación de prevención de riesgos laborales y se traten de aumentar las medidas de seguridad.
- ✓ El OP2 destaca por ser el trabajador con mayor número de horas trabajadas, mayor productividad y menor tasa de entrega de pedidos con retraso. En el lado opuesto, el OP3 es el trabajador con menos horas trabajadas de lejos, además de tener baja productividad. Algo parecido sucede con el OP5, el cual tiene una productividad casi la mitad que OP1, el cual trabaja casi las mismas horas que él. En definitiva, se han de homogeneizar las horas de trabajo de los operarios, especialmente en la Línea 1, incentivando al OP2 para que siga con ese ritmo de trabajo, y su vez, tomar las medidas necesarias con los operarios que están por debajo del rendimiento esperado.

La inclusión del análisis de sentimientos, por otra parte, ha provocado un aumento drástico en las notas medias de todos los departamentos. La posibilidad de análisis y descubrimiento de problemas y contratiempos ha permitido solucionar algunos que estaban presentes desde hace varios meses atrás. Las capacidades que ofrece un análisis así mes a mes son más potentes que cualquier equipo de personas que se encargue de supervisar todo.

Si se hubiera contado con un sistema de análisis de sentimientos se podrían haber **evitado**: los problemas con el antiguo jefe de producción, los correos electrónicos inhabilitados, la agente de atención al cliente y la salida al mercado de productos defectuosos, entre otros.

En definitiva, el Trabajo de Fin de Grado provee de una magnífica base de introducción al mundo del Business Intelligence para todo aquel interesado tanto en su marco teórico y en un ejemplo del potencial del Business Intelligence en cualquier tipo de empresa, independientemente del sector o tamaño. Permite descubrir la importancia de la cultura de los datos en las organizaciones, y los beneficios que supone tener un control total sobre ellos.

7.2 Extensiones futuras

En cuanto a las líneas futuras de trabajo, podemos sintetizarlas en un aumento de la complejidad del sistema de BI.

En un primer lugar, se podrían tomar una mayor gama de fuentes de datos para su análisis. Entre ellas, podemos destacar la fuente de datos del ERP y CRM, sistemas muy comunes y extendidos en las compañías industriales y que, actualmente, constituyen para muchas de las empresas la fuente principal de información.



En un segundo lugar, aunque es trivial, también podemos considerar la implementación en una empresa de mayor tamaño. La elección de una empresa así supone retos como la integración de datos y buen funcionamiento del sistema para decenas e incluso cientos de usuarios. A su vez, se debe crear un excelente flujo de la información, además de la creación por roles para el acceso restringido a información, por ejemplo, confidencial o sensible.

También podemos considerar una mayor complejidad de proceso productivo en la que existan más líneas funcionando, más productos e incluso productos intermedios. A su vez, también cabría la posibilidad de permitir la intercambiabilidad de la fabricación de los productos en diferentes líneas y el uso de estas por todos los operarios.

Por otra parte, el sistema descrito es uno meramente descriptivo. En una evolución de este sistema se puede considerar añadir elementos de predicción que faciliten evitar futuros problemas.

Y por último, cabe la posibilidad de introducir en los sistemas de BI instalados ciertas funcionalidades de predicción mediante inteligencia artificial o "*Machine Learning*" basadas en soluciones ya existentes de Microsoft Cognitive Services.





8. BIBLIOGRAFÍA

BARC Score Data Discovery (2018). Data Preparation, Visual Analysis and Guided Advanced Analytics for Business Analysts

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *Ai & Society*.

BBVA (2015). Reinventar la empresa en la era digital. Editorial TURNER.

Bernabéu, Darío R. (2010). DATA WAREHOUSING: Investigación y Sistematización de Conceptos. HEFESTO: Metodología para la Construcción de un Data Warehouse.

Cano, Lluís J. (2007). Business Intelligence: Competir con Información. Fundación Banesto.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 36(4).

Codd E.F, Codd S.B & Salley C.T. Technical Report. San Jose, Calif: Codd EF & Associates; 1993. *Providing OLAP (Online Analytical Processing) to User-Analysts: An IT Mandate*.

Conesa, J. & Curto, J. (2010). Introducción al BI. 1ª Edición. Editorial UOC.

De Mauro A., Greco M., Grimaldi M., (2016) "A formal definition of Big Data based on its essential features", Library Review.

Gartner Inc. (2018). Magic Quadrant for Analytics and BI Platforms.

Gartner Inc. (2019). Magic Quadrant for Analytics and BI Platforms.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*.

H. Inmon, William (1992). *Building the Data Warehouse*. 1st Edition. Wiley and Sons.

IDC (2018). The Digitization of the World From Edge to Core.

Kelleher, John D. & Tierney, B. (2018). Data science. Cambridge, MA : The MIT Press.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*.



Luhn, H. P. (1958). A business intelligence system. *IBM Journal of Research and Development*, 314-319.

Maimon, O. & Rokach L. (2010). *Data Mining and Knowledge Discovery Handbook*. Second Edition. Springer.

Market Research Future (2019). *BI Market Research Report - Global Forecast to 2022*.

Moine, J. M., Haedo, A. S., & Gordillo, S. E. (2011). Estudio comparativo de metodologías para minería de datos. In *XIII Workshop de Investigadores en Ciencias de la Computación*.

O'Brien, J. A., & Marakas, G. M. (2009). *Management information systems* (9th ed.). Boston, MA: McGraw-Hill/Irwin.

Piattini, Mario G. (2006). *Tecnología y Diseño de Bases de Datos*. 1ª Edición. Editorial Ra-Ma S.A.

Schniederjans, M. J., Schniederjans, D. G., & Starkey, C. M. (2014). *Business analytics principles, concepts, and applications: what, why, and how*. Pearson Education.

Stieglitz, S., Dang-Xuan, L., Bruns, A., & Neuberger, C. (2014). Social media analytics. *Business & Information Systems Engineering*.

Statistics MRC (2017). *BI (BI) - Global Market Outlook (2016-2022)*.

Tamayo, M. & Moreno, F. J. (2006). Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP. *Ingeniería e investigación*, 26(3).

UIAF (Unidad de Información y Análisis Financiero de la República Colombiana) (2014). *TÉCNICAS DE MINERÍA DE DATOS PARA LA DETECCIÓN Y PREVENCIÓN DEL LAVADO DE ACTIVOS Y LA FINANCIACIÓN DEL TERRORISMO (LA/FT)*. Documentos UIAF.

Varga, M. (2001). On the Differences of Relational and Dimensional Data Model. En *The 12th International Conference on Information and Intelligent Systems IIS 2001*.

Williams, N., & Thomann, J. (2003). BI Maturity and ROI: How Does Your Organization Measure Up?. *Decision Path*.



9. Páginas web consultadas

Calvo, D. (2017). Tipos de datos: estructurados, semiestructurados y no estructurados. Recuperado de: <http://www.diegocalvo.es/tipos-de-datos-estructurados-semiestructurados-y-no-estructurados/>. Accedido el: 03/03/19.

Computer Graphics at Stanford University (2000-2002). Polaris: Database and Data Cube Visualization. Recuperado de: <http://www.graphics.stanford.edu/projects/polaris/>. Accedido el 07/03/19.

Maler E. & Khatibloo F. (2013). Where BI Falls Short: Taking A Singular Point Of View. Recuperado de: https://go.forrester.com/blogs/13-08-26-where_bi_falls_short_taking_a_singular_point_of_view/. Accedido el: 14/05/19.

Microsoft Docs (2019A). ¿Qué es Text Analytics API? Funcionalidades Azure Cognitive Services Microsoft Docs. Recuperado de: <https://docs.microsoft.com/es-es/azure/cognitive-services/text-analytics/overview>. Accedido el 02/06/2019.

Microsoft Docs (2019B). Análisis de sentimiento mediante Text Analytics de Azure Cognitive Services | Microsoft Docs. Recuperado de: <https://docs.microsoft.com/es-es/azure/cognitive-services/text-analytics/how-tos/text-analytics-how-to-sentiment-analysis>. Accedido el: 02/06/2019.

Modelo oculto de Markov (2019). En Wikipedia. Recuperado de: https://es.wikipedia.org/wiki/Modelo_oculto_de_M%C3%A1rkov. Accedido el: 02/06/2019.

Páez, F. (2012). BUSINESS INTELLIGENCE COMPETENCY CENTER. Recuperado de: <http://cmigestion.es/2012/11/13/business-intelligence-competency-center/> . Accedido el 13/05/19.

Power BI (2019). Power BI Desktop: Informes interactivos | Microsoft Power BI. Recuperado de: <https://powerbi.microsoft.com/es-es/desktop/>. Accedido el 20/02/19.

Qlik (2019). La plataforma para la próxima generación de BI | Qlik. Recuperado de: <https://www.qlik.com/es-es/products/why-qlik-is-different>. Accedido el 22/02/19.

Rayo, Ángel M. (2016). Tipos de datos en Big Data: clasificación por categoría y por origen. Recuperado de: <https://www.bit.es/knowledge-center/tipos-de-datos-en-big-data/>. Accedido el 04/03/19.



Rodríguez, M. (2016). Los 75 KPI que todo gerente debe conocer. Recuperado de: <https://www.euroresidentes.com/empresa/exito-empresarial/los-75-kpi-que-todo-gerente-debe-conocer>. Accedido el 22/05/19.

Tableau (2019). Tableau Prep: La nueva solución de preparación de datos. Recuperado de: <https://www.tableau.com/es-es/products/prep>. Accedido el 22/05/19.

Thoughtspot (2019). SpotIQ - AI-Driven Analytics for Humans | ThoughtSpot. Recuperado de: <https://www.thoughtspot.com/spotiq>. Accedido el 25/02/19.

WeAreSocial (2019). Digital 2019: Global Internet Use Accelerates - We Are Social. Recuperado de: <https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates>. Accedido el 03/06/2019.