



# **GRADO EN COMERCIO**

## **TRABAJO FIN DE GRADO**

### **“APLICACIÓN DE BUSINESS INTELLIGENCE y BIG DATA”**

**Luis Miguel Martín Peña**

**FACULTAD DE COMERCIO  
VALLADOLID, Julio 2016**





# **UNIVERSIDAD DE VALLADOLID**

## **GRADO EN COMERCIO**

CURSO ACADÉMICO 2015/2016.

### **TRABAJO FIN DE GRADO**

**“APLICACIÓN DE BUSINESS  
INTELLIGENCE y BIG DATA”**

**Trabajo presentado por:** Luis Miguel Martín Peña

Firma:

**Tutor:** Juan Carlos de Margarida Sanz

Firma:

**FACULTAD DE COMERCIO**  
**Valladolid, Julio 2016**



### **Agradecimiento:**

A mi amigo, Juan Carlos, tutor en el desarrollo de este TFG que me ayudo desde un primer momento, estando pendiente y disponible. Sin él no hubiera sido posible llegar a tan buen puerto.

## **Índice:**

- 1 Introducción.
- 2 Nacimiento y evolución del Business Intelligence.
  - 2.1 Terminología de B.I.
  - 2.2 Terminología de Big Data.
- 3 Funcionamiento y requisitos de las herramientas.
- 4 Casos prácticos aplicados en la empresa y sus éxitos.
- 5 Finalidad.
- 6 Implantación Sector Alimentación.
  - 6.1 Primeros pasos.
  - 6.2 Recopilación de datos y uso de los mismos.
  - 6.3 Fases.
  - 6.4 Equipo humano y técnico necesario.
  - 6.5 Procedimiento habitual tras analizar los datos.
- 7 Conclusiones.
- 8 Bibliografía.
- 9 Anexos.

## **Índice de Acrónimos:**

- BI: Business Intelligence.
- Big Data: Macrodatos o Datos masivos.
- BBDD: Base de Datos.
- CDM: Cuadro de Mando.
- VC: Venta Cruzada.
- OLTP: OnLine Transaction Processing.
- OLAP: On-Line Analytical Processing.
- ETL: Extract, Transform and Load.
- FTP: File Transfer Protocol.

## **Índice de Ilustraciones:**

|  |                |
|--|----------------|
| Ilustración 1. Tablero de control                  | página 12      |
| Ilustración 2. Cuadro de mando                     | página 13      |
| Ilustración 3. OLTP frente a OLAP                  | página 16      |
| Ilustración 4. Fases de proceso datos              | página 19      |
| Ilustración 5. Pirámide herramientas-rol           | página 19      |
| Ilustración 6. Entrada y resultado decisor         | página 22      |
| Ilustración 7. Herramientas Front-end              | página 23      |
| Ilustración 8. Evolución consultas web términos BI | página 23      |
| Ilustración 9. Problemas baja calidad datos        | página 26      |
| Ilustración 10. Oportunidades de mejora            | página 38      |
| Ilustración 11. Tareas genéricas y específicas     | página 39      |
| Ilustración 12. Equipo humano fase I               | página 40      |
| Ilustración 13. Equipo humano fase II              | página 40      |
| Ilustración 14. SAS GUIDE                          | página 41      |
| Ilustración 15. Fases del proyecto                 | página 43      |
| Ilustración 16. FTP                                | página 56 y 57 |
| Ilustración 17. Ficheros que recibo                | página 57      |
| Ilustración 18. Fichero Clientes                   | página 58      |
| Ilustración 19. Fichero Consumo                    | página 59 y 60 |
| Ilustración 20. Fichero Direcciones                | página 60 y 61 |
| Ilustración 21. Fichero LOPD                       | página 61 y 62 |
| Ilustración 22. Fichero Navegación bruto           | página 62 y 63 |
| Ilustración 23. Fichero Productos                  | página 63 y 64 |
| Ilustración 24. Ficheros que creo                  | página 65      |

|  |                |
|--|----------------|
| Ilustración 25. Fichero Canales              | página 65      |
| Ilustración 26. Fichero Empresa              | página 66      |
| Ilustración 27. Fichero navegación creada    | página 67      |
| Ilustración 28. Fichero Operaciones          | página 68      |
| Ilustración 29. Fichero Persona campaña      | página 68 y 69 |
| Ilustración 30. Fichero Persona empresas     | página 69 y 70 |
| Ilustración 31. Fichero Proveedores          | página 70 y 71 |
| Ilustración 32. Fichero Tipo producto        | página 71      |
| Ilustración 33. Cuadro relaciones            | página 72      |
| Ilustración 34. Ficheros resultantes         | página 73      |
| Ilustración 35. Fichero Hábitos              | página 73 y 74 |
| Ilustración 36. Fichero Hábitos provinciales | página 74 y 75 |
| Ilustración 37. Fichero Campañas             | página 75      |
| Ilustración 38. Cuadro relaciones resultados | página 76      |
| Ilustración 39. Ejemplo SAS Guide Enterprise | página 77      |
| Ilustración 40. Cruce tablas I, II y III     | página 78 y 79 |
| Ilustración 41. Raíz almacenamiento proyecto | página 80      |

## **1 - Introducción:**

El objetivo del presente trabajo de fin de grado será analizar la naturaleza y evolución de la tecnología Business Intelligence basada en el análisis de grandes cantidades de datos, que disponen las empresas de gran volumen a todos los niveles, con el fin de obtener conclusiones que permitan la toma de decisiones alienadas con las directrices marcadas, así como anticiparse a los cambios del mercado.

Además de disponer de un grandísimo volumen de datos necesitamos las herramientas apropiadas para la creación de modelos estadísticos con los que poder tratar esos datos y así llegar al fin deseado.

Para ello se analizará la historia sobre Business Intelligence los programas y métodos de trabajo (SAS y Big Data), que son objeto de este estudio, desgranando cómo funcionan dichas herramientas y que es todo lo necesario para lograr que funcionen correcta y eficientemente. Es necesario conocer en qué consiste y como funciona cada uno de los métodos de trabajo hasta la fecha y la forma en que han evolucionado desde su nacimiento.

Una vez que dispongamos de la información necesaria y conozcamos los entresijos del funcionamiento, nos dispondremos a enumerar algunos casos de éxito en grandes empresas a nivel mundial, comentando las diferentes ventajas que aporta esta joven tecnología sin olvidarnos del hándicap que pueden existir, especialmente para empresas medianas o pequeñas.

Para concluir este trabajo me centraré en analizar donde está posicionada a día de hoy tanto la tecnología enunciada como las finalidades para las que se utilizará en un futuro muy cercano.

## **2 - Nacimiento y evolución del Business Intelligence:**

La primera ocasión en la que escuche hablar de inteligencia empresarial data del año 2011 nada más llegar a Madrid e incorporarme a mi actual puesto de trabajo, ya entonces era una frase de moda que no hizo otra cosa con el paso del tiempo, sino que crecer ya que tiene un significado concreto y además trasmite un valor real para las empresas.

Podemos decir que la inteligencia empresarial o BI [business intelligence] se refiere a la ayuda que esta proporciona a una empresa para que dicha empresa sea capaz de “entenderse” mucho mejor y de manera más amplia sobre sí misma. A esto habría que sumarle todo lo positivo (aplicaciones, mejoras, habilidades) de las que saca fruto la empresa con todo esto. Esta datado en el año 1958 la primera vez que se utilizó el término, hoy por todos conocido, como inteligencia empresarial, “capacidad de aprehender las interrelaciones de los datos presentados de modo tal que se pueda orientar la acción hacia el objetivo deseado” (Hans Peter Luhn, en 1958).

Debemos puntualizar que esta nueva o joven tecnología no es únicamente un gran almacenamiento de datos o únicamente una analítica pues se deben dar ambas y deben ir de la mano para así poder orientarse a determinados retos o beneficios, bien sea aumentar cuotas de clientes, servicios postventas, etc. Dentro del sector TI tratan el business intelligence como la forma de aclarar relaciones y así poder crear las futuras situaciones hipotéticas con el objetivo de perfeccionar la toma de decisiones.

Una de las grandes ventajas y beneficios del business intelligence es aglutinar todo ese montante de datos que recaen en las empresas (a mayor volumen de la empresa mayor será el de sus datos) y así poder correlacionarlos, manera en que podrán ser de utilidad. Por si mismos los datos pueden no ser significativos pero una vez que se ponen a disposición de la inteligencia empresarial adoptan formas que si la tienen. Se suele decir, con toda la razón del mundo, que el BI no crea nuevos datos, sino que en la parte que ayuda es en la relación y análisis de los mismos, ya que aún pareciendo disparatado o de difícil comprensión todo en las empresas está relacionado de alguna manera.

Quizás no seamos conscientes de la cantidad que tiene las empresas registrados de diferentes transacciones pues cada acto propio de la empresa, bien sean los pedidos,

pagos a proveedores, de nominas, etc, que conlleva una acción diferente que registra datos.

Cada movimiento diferente que ello engloba suelen modificarse de forma individual y no suelen estar preparadas para “hablarse” entre ellas, por lo que ahí es donde entra a jugar un papel importante el business intelligence para proporcionar;

- Poder visualizar datos de varias BBDD de una sola vista, de tal manera que podremos relacionar ventas con determinados gastos.
- Ser capaces de ver un resumen de datos mezclando gastos totales con lo desembolsado en las nominas para una temporalidad determinada y acotada.
- Lograr obtener una “foto” de datos respecto al tiempo pasado.
- El ser capaces de realizarnos preguntas a futuro y con datos antiguos poder dar respuestas.

Ese tipo de respuestas podríamos meterlos de forma directa en todos aquellos sistemas que se encarguen de planificar dentro de la empresa con lo que estamos ayudando de forma muy significativa a la definición de las diferentes áreas (presupuestos, objetivos generales, objetivos particulares departamentales, etc). Con ello podemos decir sin duda que el “pasado” está ayudando y decidiendo sobre la toma de decisiones “futuras”.

Existen varias razones por las que el BI está en auge desde hace varios años y una de ellas es el hecho de no necesitar romper con los sistemas que ya estén montados, sino que se basa de ellos para absorber lo que necesita y con ello trabajar, por lo que se dice que se trata de una tecnología de bajo impacto sobre la ya existente dentro de las empresas.

Lo comentado en el párrafo anterior hace que el riesgo a la hora de implantar alguna herramienta de business intelligence en nuestra empresa no conlleve casi riesgo y por lo general si está bien diseñada puede darnos un alto rendimiento de las inversiones.

Para llegar al conocimiento de BI tal y como lo conocemos hoy en día fueron necesarios el paso de los años e hitos en el camino que fueron definiendo lo que llegaría a ser lo que es hoy en día. Por épocas y años un buen resumen sería:

- Años 60: Datos almacenados en sistemas de archivo.
- Finales de los 60: Aparece el concepto de BBDD gracias a Codd.
- Años 70: Desarrollo de las primeras BBDD y aplicaciones empresariales (SAP, JD Edwards, Siebel, PeopleSoft). Con ello se consiguió realizar “data entry” en los sistemas, aumentando la información disponible.
- Años 80: Aparece el concepto Datawarehouse, así como los primeros sistemas de reporting. En estos años seguía siendo complicado y con una funcionalidad no muy extensa. Ciertamente había sistemas de BBDD pero sin aplicaciones que facilitasen su explotación.
- Finales de los 80: Aparece el término Business Intelligence.
- Años 90: Aparecen aplicaciones de business Intelligence, pero resultaba muy caro. Así mismo nacen las primeras BBDD multidimensionales (cubos) y el concepto de Data Mining.
- Año 2000: Se estabilizan las aplicaciones BI, básicamente en las plataformas (Oracle, SAP, IBM, Microsoft).

Conforme pasaban los años y las diferentes tecnologías se iban alimentando de las anteriores y aportaban alguna mejora respecto a la anterior.

## **2.1 - Terminología de B.I:**

Para esta parte del trabajo nos sumergiremos en lo elemental que alberga una solución de business intelligence, describiendo las características más importantes que se dan a alto nivel.

### *a) Data Warehouses o Almacenes de Datos y Data Marts o mercados de datos*

Podemos decir que tanto los almacenes de datos como los mercados de datos pertenecen a la parte importante de todo arranque en los proyectos de BI, aunque si profundizamos en ambos términos llegaremos a la conclusión de que técnicamente son muy parecidos, ya que no se puede decir que son iguales. La diferencia en esencia entre ambas es la amplitud de sus datos, es decir, hasta dónde quiere llegar cada uno de ellos, pues el primero busca satisfacer las necesidades que pudiera presentar toda la empresa y por el contrario el segundo de los términos le basta con agasajar a una pequeña parte de la misma (un departamento concreto por ejemplo).

A la hora de hablar sobre los diferentes enfoques voy a comentar los dos más importantes, bien desde arriba hacia abajo como el contrario desde abajo hacia arriba, el cual suele ser el más sencillo de implantar pues desgranamos las necesidades de partes pequeñas para acabar viendo cuales podrían ser a nivel general de toda la empresa. La razón de por qué es más sencillo es debido a que es mucho más sencillo canalizar los departamentos o áreas e ir poco a poco para llegar a lo general que justo al revés.

Debemos tener en cuenta que poner un sistema de BI en una empresa de grandes dimensiones es complejo puesto que la visión global de esta no es fácil y lleva asociado tiempo y dinero, razón por la que esto puede verse facilitado cuando se trata de empresas de menos envergadura ya que es más fácil el poder tener una “foto” de toda la empresa. Además hay que tener en cuenta que las soluciones prefabricadas de antemano para esas empresas menores suelen tener un resultado mucho más positivo.

En el mundo de la consultoría, en el que llevo inmerso desde hace casi seis años, una frase muy recurrente a la hora de abordar nuevos proyectos, problemas genéricos o asilados dice explícitamente “un paso atrás” con lo que se nos indica la necesidad de coger perspectiva, desgranar los problemas para ir dando soluciones más pequeñas que sean posibles de resolver para después ir encajando las piezas con el problema o proyecto general. Como es comprensible esta metodología consigue rapidez a la hora de

resolver problemas de menos tamaño, lo que lleva implícito ahorro de dinero y enfoque directo al éxito.

b) Minería de datos

Para definir de una manera sencilla que es y para qué sirve la minería de datos, también conocido como data mining, diré que es el proceso por el cual somos capaces de encontrar la aguja del pajar, es decir saca los patrones de un conjunto de datos (siempre y cuando estén) ya que no se inventa aquello que no está.

Una vez que tengamos las conclusiones debemos comprobar que lo que nos muestra tiene sentido común.

c) Informes, tableros de control y cuadros de mando

Cabe la posibilidad que tanto volumen de datos no nos dejen ver con claridad para lo cual necesitaremos algo más simplificado, manera con la que ser capaz de hacernos una idea general y clara sobre cómo están las cosas. Por una parte los informes son algo coherente de un almacén de datos y podemos encontrarlos de diversos estilos aunque dos instrumentos muy utilizados son el tablero de control y el CDM (cuadro de mando).



Tablero de control ([www.axoft.com](http://www.axoft.com))

La imagen que precede corresponde al primero de ellos, el tablero de control, lo que nos permitirá tener una idea general (sin mucha profundidad) de cómo están las cosas en la empresa.

Con ello podemos tomar el pulso y temperatura a diferentes aspectos de la empresa y nos llevará a decidir si debemos profundizar más o no sobre qué aspectos (clientes, económicos, proveedores, etc).

Claramente ese tablero lo que nos va revelar es si las “cosas” de la empresa están significativamente dentro o fuera de lo planteado, estamos mejor o peor y de cómo debiéramos estar, esa es la pregunta que de un vistazo responderá.

Por otra parte la segunda forma de la que hablamos en líneas anteriores es el CDM (cuadro de mando) que difiere del tablero de control en que este además es más completo, nos aporta más datos, y por ende podremos saber cuánto hemos progresado en términos respecto a los objetivos que tiene la compañía.



Cuadro de mando ([www.pgconocimiento.com](http://www.pgconocimiento.com))

La imagen anterior, tal y como apunta el pie de foto, se trata de un completo cuadro de mando en el que poder ver en qué punto estamos respecto a los objetivos,

manera en que podremos ver que problemas tenemos, en qué áreas tenemos problemas, etc.

Es cierto que ambos dos, tableros y cuadros de mando, realizan funciones parecidas, aunque por diferenciarlos un poco más diré que mientras uno revela datos en términos absolutos el otro lo hace respecto a los objetivos.

d) Analítica predictiva

Este concepto, analítica predictiva, engloba varias disciplinas (data mining, búsqueda de patrones y realizar las predicciones a futuro basándose en datos históricos).

Uno de los casos clásicos respecto a este concepto es la VC (venta cruzada) de artículos basándose en aquellos que el cliente ya adquirió en el pasado, de forma y manera que al ofrecerles determinados artículos evitamos que el cliente tenga que ir a buscarlos o se acuerde de ellos si es que en algún momento estuvieron en su cabeza, lo que ayuda a ver aumentados los ingresos.

Es cierto que al leer estas líneas pensemos que muchas empresas vienen utilizando este parámetro sin ni siquiera planificarlo pues con los datos que tienen de años pasados saben que hay determinadas épocas del año en que las ventas superan con creces a las del resto del año, razón por la cual podrán hacer una predicción sobre cuanto personal necesitan, si han de aumentar su stock en almacén para hacer frente a toda la demanda que en principio se les viene encima.

e) Gestión del rendimiento empresarial

Quizás el título de este apartado de a entender que se trata de algo muy genérico y la verdad es que se trata de una mezcla entre informes, la planificación, forecasting, presupuestos y el análisis predictivo. Busca dar un “foto” que organice y de perspectivas nuevas para así poder optimizar el rendimiento.

Si bien es cierto que el concepto es bastante nuevo la realidad nos enseña que las prácticas se vienen desarrollando desde hace mucho tiempo ya que la gran mayoría de los directores y gerentes de una compañía tienen la buena costumbre de revisar las ventas del día anterior cuando empiezan a arrancar el día. Todo esto no es a capricho, la razón es clara, es la manera de saber como fue el día anterior, de qué manera llevan las ventas respecto a objetivos personales, empresariales, etc. De esta manera saben si la dinámica es buena, si hay que apretar la marcha por qué no llegan, etc.

Por otra parte esto también les ayuda para decidir si se debe incrementar la velocidad de los pedidos por que haya una previsión de romper stock ante una demanda creciente. Una gran ventaja que no hemos comentado es la automatización del sistema BPM, el cual se alimenta por sí mismo y actualiza de la misma manera.

Debemos dedicar algunas líneas a la definición y análisis de la tecnología subyacente en las soluciones de BI.

### OLTP frente a OLAP

Las bases de datos de procesamiento de transacciones en línea (OLTP) tienen estas características:

- Información detallada, disponen de información resumida y también contiene información desgranada.
- Procesan transacciones, realiza un dato por vez: un pedido, un producto, un cliente. Podemos realizar informes básicos, aunque la estructura puede admitir el acceso veloz a pequeños grupos de datos razón por la que contiene muchas tablas.
- Esquema rígido y nada variable.
- La información es reciente y puede actualizarse en tiempo real.
- Los volúmenes de datos de entrada suelen ser muy alto.

Por otra parte tenemos almacén de datos es una base de datos con procesamiento analítico en línea (OLAP):

- Pudiera presentar datos resumidos, sin detalles. Trabaja con una agregación de datos.
- Incide en el análisis y las decisiones. Es más importante analizar grandes cantidades de datos y así intentar canalizar las posibles tendencias.
- Como ya hemos comentados trabaja con una agregación de datos por lo que la información es histórica. Se centra en las tendencias con los datos antiguos

- De gran importancia para todos estos procesos es que los datos estén limpios, razón por la que los datos habitualmente se “limpian”, siendo estos así más fiables lo que nos mostrará tendencias más claras.

Es muy habitual que las empresas que disponen de un almacén de datos vayan a disponer también de una o más bases de datos “normales” con las que abastecer al almacén de datos, haciendo así que sean independientes en el funcionamiento.

Quizás la manera más visual y sencilla de comprender las diferencias entre ambas sea mediante una imagen en la que diferenciar donde no coinciden ambas dos, OLTP frente a OLAP;

|                                | OLTP                             | OLAP                               |
|--------------------------------|----------------------------------|------------------------------------|
| <b>Usuario típico</b>          | Operacional                      | Analista                           |
| <b>Función</b>                 | Almacenamiento/consulta de datos | Análisis                           |
| <b>Registros accedidos</b>     | Decenas                          | Millones                           |
| <b>Unidad de procesamiento</b> | Transacción simple               | Agregación                         |
| <b>Redundancia de datos</b>    | Causa ineficiencias              | Clave para procesamiento eficiente |
| <b>Modelo de datos</b>         | Relacional                       | Multidimensional                   |

*OLTP vs OLAP (Elaboración propia)*

### Datos normalizados frente a dimensionales

Cuando hablamos de OLTP estamos haciendo referencia a datos normalizados, es decir que quieren reducir los datos duplicados. Ciertamente es que a la hora de devolvernos consultas en estas bases de datos suele realizarse de forma más lenta lo que incrementa el tiempo de respuesta en una búsqueda de BI. El almacén de los datos utiliza una estructura diferente pues los datos los reúne en una tabla, conocida como tabla de dimensiones. En una segunda tabla de dimensiones carga todos los datos referentes a los productos, por ejemplo, y en una tercera tabla de dimensiones tendrá cargados los

países desde donde se realizan compras, de esta forma se monta una tabla de hecho que de lo que se va a encargar es de relacionar las tres a la vez.

Muchas veces hemos escuchado aquello de que los almacenes prefieren cambiar tamaño por velocidad.

### Modelado de datos

Todos conocemos la historia en que un profesor con un jarrón vacío pregunta a sus alumnos que cosas son las que entran en dicho jarrón de los diferentes artículos que en la mesa tiene, lo interesante de esto es la necesidad del orden pues no es lo mismo meter primero la arena y luego las piedras grandes, que para el modelado de datos hace referencia al esquema. El esquema de cómo se organicen los datos hará que sea mucho más eficiente y eficaz dicha bbdd en las futuras búsquedas, análisis, etc. Lo primero que debe preguntarse es justo que preguntas queremos que se nos respondan, por que de esta manera se estructurara de una u otra manera

### Transformación y carga de datos

Cuando ya tenemos claro que esquema, visto en el punto anterior, vamos a seguir y que preguntas buscamos responder veremos que el siguiente hito a desarrollar es la propia carga de datos. Se comento en líneas anteriores pero merece la pena recordar que el buen funcionamiento del almacén de datos en parte es gracias a que se copia los datos y por tanto no interfiere en los sistemas de producción.

Aunque se le podría echar en cara que los datos están algo desactualizados, pues si bien es cierto, eso no acarrea ningún tipo de problema pues debemos tener en cuenta que el BI funciona con tendencias, de largo plazo, con lo cual no tiene mayor inconveniente. Para el proceso de copiado de datos se le denomina (ETL) extraer, transformar y cargar.

El primero de ellos “extraer” lo que hace es una conexión con la base de datos de donde “bebe” y extrae de ella los datos necesarios. Para el paso “transformar” lo que hace es reordenar los datos en el esquema que use el almacén de datos. Para terminar, el momento de “cargar”, lo que realiza es colocar los datos transformados en el almacén de datos.

### Análisis en memoria

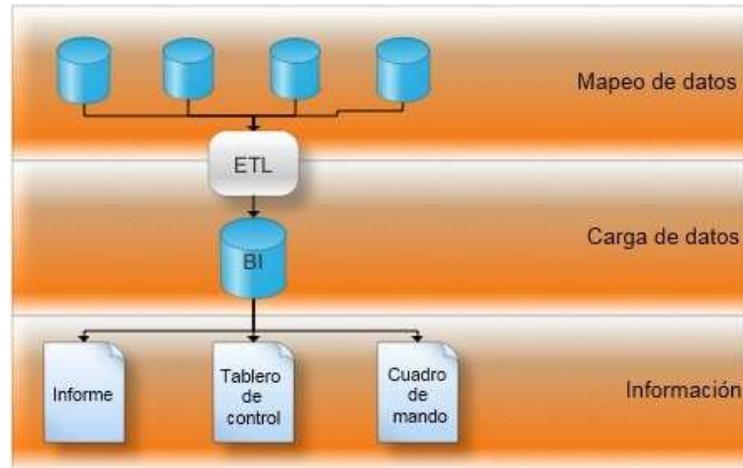
Existe la posibilidad realizar los análisis mientras se leen en la memoria del servidor y aunque esta idea no es nueva debemos reconocer que desde no hace mucho tiempo es viable, debido a la disponibilidad de potentes ordenadores así como una disminución en el precio de estos.

### Realizar informes, análisis y visualización

Para finalizar debemos utilizar los datos del almacén, será donde entren a jugar los sistemas de BI, que es quien dispone de la “inteligencia” para poder pasar de los datos en “bruto” a informes, cuadros de mando, tableros, etc. Con la elaboración de informes, el análisis y la visualización cabe la posibilidad que pudieran llegar a parecer muy similares por lo que vamos a desarrollar un poco más la idea y las diferencias entre ambas;

- Informes, engloban variedad de necesidades empresariales, pueden ser personalizados, podemos decir que son muy flexibles lo que es muy valorado.
- Tableros de control son lo que modifican la compleja información en imágenes de alto impacto. Visto esto somos conscientes que son más intuitivos además de ayudar a los gerentes totalmente informados, sin olvidar que les indican de forma rápida los problemas que pudieran tener, por lo que es ideal como herramienta diaria.
- Sin olvidar que los cuadros de mando, al igual que los tableros de control, ofrecen información de un vistazo hay que sumarle la comparación con objetivos específicos.
- Para concluir con este apartado diré que el análisis nos permite visualizar información de toda la empresa lo que a su vez nos trasmite el por qué de las cosas, pudiendo bucear entre la información más directa hasta la más profunda.

Con la siguiente imagen podemos ver los diferentes niveles y resultados a los que podemos llegar con los mismos, se trata de una ilustración de lo redactado con anterioridad.



*Fases proceso (Elaboración propia)*

Diferentes roles necesitarán funcionalidad diferente de la herramienta pues no todos abarcan lo mismo ni de la misma manera tal y como se refleja en la siguiente imagen mediante una pirámide herramientas-rol;



*Pirámide herramientas-rol (Elaboración propia)*

## **2.2 - Terminología de Big Data:**

Big Data tiene como objetivo resolver problemas antiguos y nuevos de manera más eficiente enfocada a extensos y heterogéneos volúmenes de datos no usados hasta el momento, posiblemente la próxima tecnología revolucionaria en tecnologías de la Información que apareció en la primera década del siglo XXI. Genera valor a un negocio a partir del almacenamiento y procesamiento de cantidades muy grandes de información digital que no puede ser analizada con técnicas tradicionales de computación.

Existen diferentes definiciones sobre Big Data y seguramente muchas serían válidas, en nuestro caso lo definiremos como un conjunto de tecnologías, herramientas (por ejemplo, SAS Guide) y técnicas que hacen posible la recogida, procesamiento y análisis de volúmenes masivos y heterogéneos de datos. Además de todo esto lo que buscamos son unos resultados, a modo de tableros o no, con lo que operar (estadísticas, patrones de comportamiento, análisis de rendimiento, etc).

¿Por qué se habla de la inteligencia de Big Data?, ¿Qué hace inteligente a un sistema de Big Data?

- **Análisis:** Maneja grandes cantidades de datos y de muy diversa naturaleza, no sólo los directamente relacionados con el problema, sino también cualquier otro dato que pueda influir en la decisión.
- **Instrumentación:** Para recoger los datos debe existir un conjunto de instrumentos capaces de medir la información deseada.
- **Interconexión:** Debe existir una infraestructura preparada para recibir y almacenar cualquier tipo de datos recogido y procesarlos de manera eficiente.

Para entender con mayor profundidad debemos hablar de la minería de datos, la cual ya definimos en páginas anteriores como la ciencia que se dedica a encontrar las agujas en el pajar del Big Data, extrayendo lo valioso de entre la ingente cantidad de información. Cada vez más cantidad de datos es almacenada y el desarrollo tecnológico permite computar, almacenar y transmitir datos. Algunos factores que han permitido este avance son;

- Abaratamiento de los sistemas de almacenamiento.
- Incremento de la velocidad de cómputo.

- Mejorar en la confiabilidad.
- Creación de sistemas que valgan de administradores de bbdd.
- Los datos en bruto tienen poco valor, surge la necesidad de estudiar y analizar los datos para obtener información valiosa, es decir el conocimiento.

El objetivo fundamental de la minería de datos es encontrar modelos inteligibles a partir de los datos y para ser efectiva debe de ser un proceso automático o semiautomático.

Los modelos descubiertos deben ayudar a tomar decisiones que aporten algún beneficio al sistema. El data mining dentro de Big Data da un paso más allá ya que no sólo tiende a buscar patrones a partir de los datos sino que además analiza esos patrones.

Podemos resumir lo esencial de Big Data en cuatro V de la siguiente manera;

- Volumen. Hace referencia al volumen (tamaño) de datos que se generan, para hacernos una idea debemos saber que lo producido a nivel mundial en dos días equivale a todos los que se habían generado antes del año 2003.
- Velocidad (y frecuencia): Debido al volumen y la velocidad con que estos se crean es muy fácil que los datos se queden desfasados por lo que dejan de tener interés.
- Variedad: Desde hace no demasiados años vemos como son múltiples los aparatos o las formas en que lo humanos generamos datos, desde un teléfono móvil, un sistema GPS, sensores de calor, etc.
- Veracidad: En puntos anteriores ya comenté la importancia de tener datos, pero que esos datos estuvieran limpios, pues uno de los mayores problemas es que vengan sucio, incompletos, etc. Esto nos lleva tener muchos datos pero incorrectos y necesitamos de “arreglar”, ya que será de la única manera en que nos proporcionaran valor.

La manera de comprender mejor el escenario de Big Data es describir las cuatro fases principales en el proceso, centradas en “Educar, explorar, interactuar y ejecutar”.

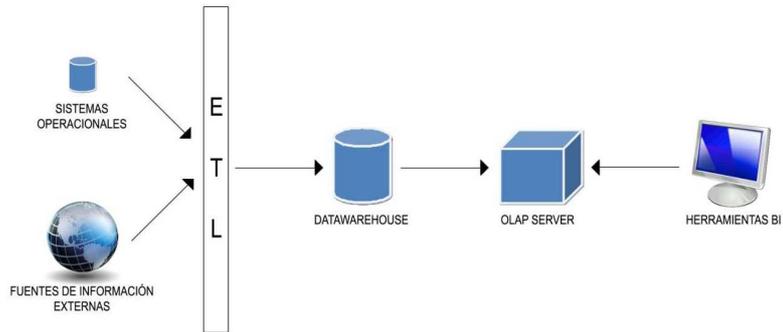
- Educar: Básicamente se trata de realizar la base de conocimiento centrándonos en la concienciación y el desarrollo del conocimiento. Veremos que el resultado en los directivos empresariales es claro, aún no han llegado a comprender del todo el potencial de esta herramienta y por tanto queda recorrido de trabajo por hacer.
- Explorar: En esta parte nos centramos en desarrollar la hoja de ruta de la empresa. Por lo general lo principal se basa en desarrollar un caso de negocio cuantificable y crear un proyecto de Big Data.
- Interactuar: Esto suele suceder cuando las empresas comienzan a comprobar el valor de negocio de Big Data.
- Ejecutar: Se trata de implementar el propio Big Data.

**INFORMACIÓN**  **DECISIÓN**  **ACCIÓN**

*Entrada y resultado decisor (Elaboración propia)*

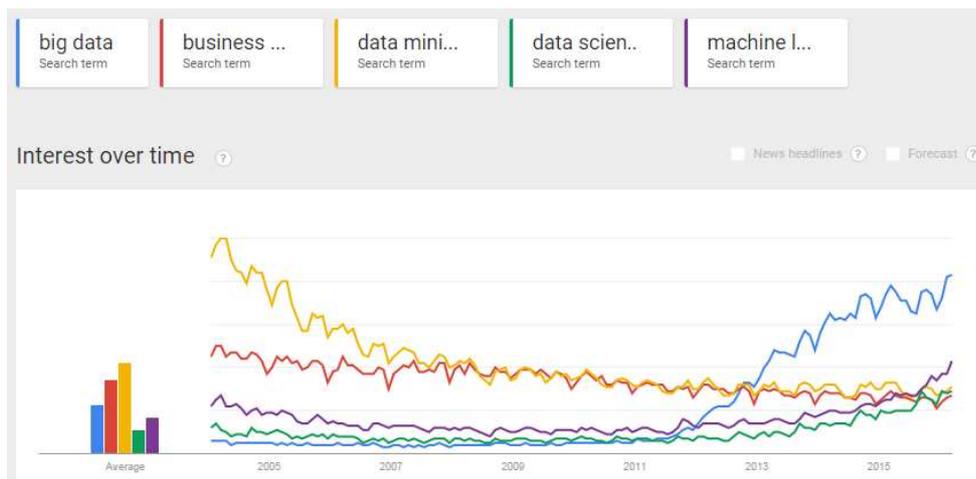
Componentes que conforman un sistema BI:

- Fuentes de información, que alimentan al sistema con los datos.
- El Proceso ETL, que extrae, transforma y carga los datos de las fuentes de información al datawarehouse.
- Datawarehouse que será el encargado de reunir los datos de la organización. Se trata de la colección de datos integrada, es decir, tiene los datos de todos los sistemas operacionales de la organización, por lo que la información no volátil, no se modifica ni se elimina.
- Motor OLAP que permite capacidad de cálculo, consultas o análisis de escenarios.
- Por último las herramientas Front-end, que nos permiten la visualización por los datos para su posterior análisis.



*Proceso ETL (Elaboración propia)*

Si analizamos el ámbito de las consultas web y comparamos la frecuencia con la que estos términos han sido buscados en Google desde 2004 hasta hoy, podemos mediante el siguiente gráfico ver la evolución y hacia donde apuntan cada una de las líneas para este y el próximo año.



*Evolución consultas web (www.mikelnino.com)*

### **3 - Funcionamiento y requisitos de las herramientas:**

Históricamente la manera de entrada en las organizaciones del BI era mediante su instalación en un servidor, razón por la que es tan necesario comprender bien todos los procesos y herramientas que están por detrás. Por ello en este tercer apartado, explicaré los procesos básicos mediante los cuales se introduce el Business Intelligence, además de ver por que tradicionalmente se trataba de algo carísimo de arrancar y mantener.

#### **Obtener los datos empresariales**

Básicamente el objetivo primordial del BI es reunir datos de fuentes diferentes, cuantas más fuentes de las que “beber” mejor ya que la muestra será no sólo más amplia sino que también más rica, manera en que podrá comenzar a hacer conexiones, hallar patrones y detectar tendencias que, de otro modo, no serían evidentes.

En la actualidad lo que suele ocurrir muy a menudo es que los datos dentro de la empresa estén alojados en bdd del tipo transaccional (para los pedidos, la facturación, etc) y además es probable que físicamente se encuentren en bases de datos como SQL server, DB2 de IBM o alguna de las más conocidas de Oracle. Cabe la posibilidad y es muy habitual verlo en las empresas que además de todo esto dispongan de más datos en otro tipo de formatos o soportes cómo hojas de cálculo Excel.

A todo esto se le denomina sistemas fuente, que será precisamente de donde absorberá los datos necesarios, ya que el Business Intelligence no varía eso, de lo único que se encargara es de mirar esos datos y copiarlos en otro lugar diferente.

#### **Mapeo directo de datos**

Con esta técnica lo que consigue el BI es conectarse a los datos de nuestra empresa allá donde estén ubicados, de forma y manera que cada vez que necesite algo, algún dato concreto, lo que hará será conectarse y cogerlo, con lo que de primeras entendemos que la mayor ventaja es que siempre los datos estarán actualizados (siempre leerá el último dato que se haya registrado). Por el contrario hay que mencionar que el inconveniente que esto presenta es claro, funcionando así podemos ver ralentizado el rendimiento de los sistemas.

Por lo general los sistemas de BI se basan en un sistema transaccional que les permiten una rápida consulta de datos, aunque también se da el caso en que esa

consulta no es tan sólo una consulta como tal sino que además lleva incorporado el agregar más datos.

### Carga de datos

El momento en que realizamos la carga de datos viene dado por no querer permitir que un sistema pueda acceder a los que ya están en producción mientras los usuarios están creando, metiendo, modificando datos de la índole que sea. Para la función que se busca, el sistema de BI, lo que hará habitualmente será que mediante un mapa de datos copiara esos datos necesarios de sistemas transaccionales para transformarlos a modo en que se requiere en business intelligence.

No debemos olvidar que la mayoría, por no decir la totalidad de los sistemas de BI, admiten la carga personalizada de los datos provengan de la bbdd que sea. Lo que realmente hace que el BI sea tan útil es la transformación de ETL, un proceso de transformar será capaz de solucionar algunos problemas como por ejemplo;

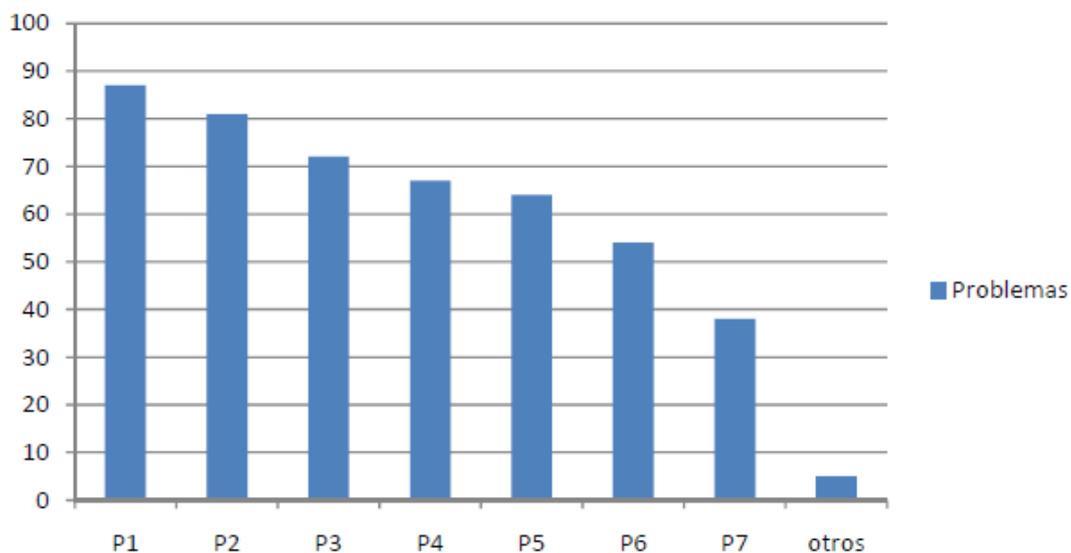
- Completar y solucionar las abreviaturas que muchas veces nos encontramos en los datos, bien de nombres, ciudades, productos, etc.
- Quedarse con los datos que serán relevantes.
- Limpieza, verificación de datos (direcciones, numéricos).
- Calcular los totales de los pedidos de clientes.
- Filtrar la entrada para evitar redundancias.
- Se pueden mezclar de diversas fuentes.
- Corroborar los datos, para así tener la certeza de que los datos usados son correctos y actualizados.
- También podrá correlacionar los datos que vienen de diferentes sistemas.

En todas las empresas en las que trabajé estos años atrás, en todas ellas con diferentes proyectos lo que garantizó el poder ver cómo operan en los diferentes departamentos, vi como la carga de datos se realizaba fuera del horario laboral, razón por la cual no interfería en absoluto para los trabajadores.

En líneas anteriores ya comenté que los datos no estarán actualizados al minuto pero esto no es necesario pues los sistemas de BI trabajan con patrones y por temporalidad.

Es importante desde un primer momento asegurar, si es posible, que la calidad de los datos sea la correcta, a mayor calidad mejor para todo el proceso y los resultados que obtendremos. Nos podemos encontrar con los principales problemas que provoca una baja calidad;

- Problema1. Demasiado tiempo extra para conciliar los datos.
- Problema2. Una baja credibilidad en el sistema.
- Problema3. Encontrarnos con costes extra.
- Problema4. Que los clientes estén insatisfechos.
- Problema5. Demoras para el desarrollo de nuevos sistemas.
- Problema6. Unas pérdidas de ingresos.
- Problema7. Algunos problemas de conformidad.



*Problemas en datos (sasybi.blogspot.com)*

Decimos que para poder decir que los datos que tenemos son de calidad antes debemos comprobar varias variables:

- Precisión: Representen con precisión la realidad.
- Integridad: Estructura entre datos y relaciones serán invariables.
- Totalidad: Necesitamos que estén todo los necesarios.

- Coherencia: Que se encuentren bien definidos y además sean comprensibles pues de otra manera nos servirían para muy poco.
- Validez: Existen unos rangos aceptables según la empresa, cada una tendrá los suyos propios, entre los que deben estar el valor de los datos.
- Accesibilidad: Que sea fácil acceder.
- Disponibilidad: Que podamos disponer de los datos siempre que sea necesario.

### Mercados de datos o Data Marts y almacenes de datos o Data Warehouses

Como ya describí con anterioridad los almacenes es donde BI recopila habitualmente sus datos para después hacerlos llegar a sus sistemas transaccionales, en producción, por lo que debe contener los datos de una forma correcta, limpia y estandarizada para que no se generen problemas a la hora de tratarlos. Por otra parte un mercado puede ayudar a tomar las decisiones amplias en un ámbito global de empresa.

Es cierto que algunos almacenes de datos se crean a partir de un conjunto de mercados de datos (lo relatado en párrafos anteriores sobre los diferentes enfoques desde abajo hacia arriba o viceversa).

### Análisis en memoria

Respecto al análisis en memoria también se desarrollo con anterioridad cuando dije que el enfoque se basa en que en vez de cargar todos sus datos en un almacén de datos, los almacene dentro de la memoria de un ordenador y trabaje allí con ellos, interactuando con una parte muy pequeña. Con todo esto una de las cosas que se pueden lograr y que nos ayuda (ahorrando tiempo) es el hecho de que borremos de la ecuación la necesidad de transformar los datos, con lo que los datos se pueden consultar desde sus sitios (bbdd) de procedencia. Algunos de los motivos que también debemos tener en cuenta son;

- Para crear un almacén de datos necesitamos cierta competencia técnica, tanto para montarlo desde cero como para conseguir que el rendimiento sea el más óptimo sin embargo como los análisis en memoria son del tipo ad hoc nos van a requerir menos destreza técnica (esta más a la mano de todo tipo de públicos).

- Los análisis de memoria son más rápidos que los análisis en disco debido a que en estos segundos se ralentizan más que los primeros (diferencias entre memorias solidas y no).
- Esta comprobado que el análisis en memoria permitir análisis más flexibles ya que los datos pueden a estructurar rápidamente.

Para ser sinceros hay que recalcar que este tipo de análisis no es una solución perfecta ya que no es ilimitada o gratuita.

### Creación de una cultura empresarial impulsada por la información

Históricamente a nivel mundial hemos podido ver como las empresas, especialmente las pequeñas y familiares se administraban por instinto, esto solía venir dado por el hecho de que los gerentes y directivos de dichas empresas llevaban muchos años en ellas. Este factor les hacía pensar en todo momento que controlaban la empresa de principio a fin y se veían con la falsa capacidad de tener todo controlado (lo que se conoce como visualizar la “foto” de forma entera y global), ni que decir tiene que este tipo de modus operandi para nada es fiable ni mucho menos el más eficiente para una empresa, independientemente el sector, tamaño o procedencia.

Conforme vamos dando pasos, el primero será la creación de todo lo que alberga el sistema de BI e ir introduciendo datos, no debemos olvidar que el acto de introducir datos no basta pues es necesario que confiemos en ellos ya que de otra manera acabaran por no utilizarse quedando en el olvido. He conocido (y en algunos casos he trabajado para ellos) a determinados gerentes que se pasaban todo el día mirando hojas de cálculo y nunca matizaban toda esa información con una concienciación práctica. Necesitamos que la información contribuya con el proceso de administración, no dominarlo.

### Elaboración de informes

Los informes pueden ofrecer información muy valiosa y los buenos gerentes aman los informes por esa razón, son capaces de proporcionarles mucha ventaja frente a competidores, proveedores, etc que sin duda quieren aprovechar para beneficio de su empresa.

El sistema de BI nos puede ayudar a ver tendencias por volúmenes en determinados ámbitos de nuestra empresa. Los informes de un sistema de BI pueden ser

extremadamente flexibles, dinámicos ventaja que nos permite usar ese informe para hacer preguntas a futuro sobre situaciones hipotéticas, lo cual sin duda nos ayudara a tomar decisiones de acuerdo con los resultados de las tendencias que seamos capaces de sacar y ver.

### Tableros de control

De los tableros de control diremos que son tipos especializados de informes que utilizan visualizaciones gráficas para resumir información clave. Muchas veces el tablero de control se les compara con el cuadro de instrumento de un coche, en el cual podemos ver de un vistazo como esta todo, si hay problemas (luces rojas) que llamen nuestra atención y sobre las cuales debemos incidir rápidamente, etc.

### Cuadros de mando

Para finalizar con este apartado debo hablar de los cuadros de mando, que al igual que los tableros de control, lo que buscan es brindar una “foto” general de alto nivel. Se dice que la gran diferencia entre ambos es que los cuadros de mando aplican una calificación absoluta en función de umbrales relativos.

Respecto al cuadro de mando en cada uno de los indicadores nos indica si la empresa está yendo “bien”, “aceptable” o “mal”. Por describirlo de modo sencillo diré que si todo tiene una “A” o una “B”, entonces en esencia está yendo bien y por el contrario si los puntos mayoritarios son una "C" o una "D" la marcha de la empresa no es la esperada y requiere más atención y esfuerzo por parte de la gerencia. El último de ellos sería el “F” por cuya virtud necesita una atención inmediata y significativa que corrija esos casos.

## **4 - Casos prácticos aplicados en la empresa y sus éxitos:**

### *a) Sector de los Seguros*

Una importante empresa dentro del sector de los seguros y reaseguros de nuestro país presento un caso de éxito al analizar las pólizas de directo desde punto de vista de siniestralidad. Existen segmentos de clientes que tienen ratios de siniestralidad elevados que se corresponden con perfiles que no nos interesa captar. Para ello situó a los clientes directo en el medio y alrededor de ese eje y se situaron los diferentes grupos de que disponía la compañía en términos y catalogados por la siniestralidad creando así cuatro subconjuntos (Segmentos A, B, C y D). Dentro del segmento A englobamos los clientes con una siniestralidad muy alta, en el B lo de siniestralidad alta, los clientes con siniestralidad dentro de la media se encuentran en el segmento C y para los de baja siniestralidad el D.

Para llegar al objetivo deseado, captar clientes rentables y reducir el coste por póliza de un 20%, focalizo su selección del público objetivo analizado desde el departamento de marketing para clientes con póliza de auto, target al que le restaron la suma de los segmentos A y B, con lo que dejamos fuera del publico objetivo todos aquellos clientes de muy alta y alta siniestralidad.

### *b) Sector de las Telecomunicaciones*

Para el sector de las telcos tan grande e interrelacionado desde hace unos cuantos años, donde la competencia directa es grande y fuerte comentare un caso de éxito de gran relevancia en el que se busco ganancias mediante el freno de dos problemas habituales en toda empresa, siendo estos dos hechos para las grandes compañías una parte de mucho peso en sus balances, que desarrollamos desde mi empresa (proyecto en el participe activamente durante más de un año y medio).

Una parte enfoca la retención de los posibles abandonos de la compañía hacia la competencia y por otra la prevención del impago de facturas por parte de sus clientes. Para ello partimos de un análisis exhaustivo tanto de la empresa, el sector y la competencia más directa. Una vez cumplimentado este hito del proyecto arrancamos con el estudio y disgregación de los datos que disponíamos en diferentes bases de datos dentro de la empresa para así poder determinar qué datos necesitamos utilizar para la implementación de la segmentación de cartera mediante modelos estadísticos en SAS.

El objetivo es poder disponer de modelos de propensión que identifiquen los targets de las diferentes campañas, además de clientes en riesgo:

- Clientes con riesgo de baja voluntaria/portabilidad -> Desconexiones voluntarias
- Clientes con riesgo de desconexión por Impago -> Desconexiones por impago

El desarrollo de los modelos permitió a la empresa aumentar su conocimiento de los perfiles y/o patrones que siguen los clientes incluidos en los dos grupos así como identificar sobre qué clientes concretos se pueden aplicar las acciones de retención y fidelización más óptimas. Basamos nuestro trabajo mediante el método SEMMA (Sample, Explore, Modify, Model, Asses);

- Muestreo y selección -> Muestreo, selección y establecer niveles de confianza.
- Limpieza de datos -> Datos inexistentes, o no clasificados, identificación de los extremos así como la eliminación del ruido.
- Transformación de datos -> Reducir la dimensionalidad, creación de características, normalizar los datos, variables correlacionadas y discretización.
- Modelado -> Clasificación, regresión, agrupamiento, asociación y secuenciación.
- Reportes y visualización -> Entrega de los resultados.

Al tratarse de la elaboración de modelos de comportamiento (alguien hace algo o no), se utilizan modelos de elección binaria porque la variable de salida es discreta, es decir, es SÍ (1) o NO (0). La mejora que produjo el modelo sobre una elección aleatoria fue de un 3,4 para el 10% más propenso. Al contactar al 8% más propenso cada mes sobre el total de cartera (alrededor de 7600 clientes) se acierta en el 30% de las desconexiones voluntarias que ocurrirán a los dos meses de la fecha de ejecución.

Se utilizó un horizonte de predicción 2 meses, esto significa que con los datos de cierre del mes de septiembre (que se reciben en octubre) se predecirán las desconexiones de móvil que van a ocurrir durante el mes de diciembre. El listado de clientes propensos estará listo para llamar a finales de octubre, y las llamadas se



realizarán durante noviembre para prevenir las desconexiones de móvil que ocurrirán en diciembre.

## **5 - Finalidad:**

La finalidad o finalidades como comenté son varias, basándonos en el análisis predictivo y trabajando con grandes volúmenes de datos lo que buscamos es la obtención de patrones de comportamiento que nos permitan decidir la estrategia más afín a nuestros objetivos.

Con ello podemos determinar qué productos para un determinado sector de nuestra cartera son los más recomendables ofertar, realizar ventas cruzadas y hasta calibrar que clientes tienen una mayor propensión a darse de baja o dejar un impago en sus próximas facturas. Una vez que tengamos en nuestras manos todas las herramientas y datos necesarios podemos crear los tableros de scoring para la cartera de clientes.

## **6 - Implantación Sector Alimentación:**

La aportación que se persigue y pretende implantar como mejora para una gran y conocida empresa nacional dentro del sector de la alimentación, con la aplicación de todo lo analizado y explicado en este trabajo, es determinar los diferentes tipos de clientes que frecuentan los establecimientos que dispone en toda España dicha empresa y así analizar que hábitos de consumo tienen, que propensión de compra y hacia qué productos con lo que incrementar las ventas habituales de cada cliente y poder realizar con éxito la venta cruzada de otros productos que no se encuentra en la lista habitual de cada cliente.

Para ello necesitamos dos flujos de datos, de entrada y salida;

- Datos de entrada -> Crear una base de datos con todos los datos personales de cada cliente así como de todos los productos que compra siempre que acude a alguno de nuestros establecimientos. Además crearemos una tabla donde reflejar los hábitos de consumo, otra con todo lo referente a LOPD, una tabla con todos los productos (referencias) y sus características que disponemos en nuestros supermercados y por último una tabla con todos los datos web generados por las cookies.
- Datos de salida -> Para presentar las ofertas y productos a los clientes que podría tener gran propensión de compra (tanto de productos habituales en su cesta de la compra como aquellos que no lo son) mediante pantallas digitales en los carritos de los supermercados, a través de la app en su dispositivo móvil y mediante e-mail.

## **6.1 – Primeros pasos:**

Nuestro primer objeto siempre va a ser crear la base de datos sobre nuestros clientes con sus datos personales para lo cual se solicitarán mediante un folleto para todos aquellos que deseen pertenecer al club de descuentos y ventajas de nuestros centros comerciales. Mediante dicho formulario, se recogerán todos los datos básicos con los que empezar a trabajar. El porcentaje de clientes dispuestos a ceder voluntariamente sus datos es limitado, puesto que no todos están dispuestos de primeras a proporcionarlos sin más.

Para el grupo de clientes que en un primer momento no estén dispuestos a proporcionar sus datos se les ofrecerá un descuento especial para su próxima compra, descuento directo sobre factura, a condición de cumplimentar de forma correcta y completa el folleto de información y hacerse miembro del club debido a que para la aplicación de dicho descuento deberán entregar su tarjeta de socio en alguna de las cajas de cobro antes de pagar.

En la recogida de datos utilizamos la lista de la compra cada vez que el cliente entrega su tarjeta de fidelización en las cajas de cobro, de esta forma tenemos vinculado cada cliente con todos los productos que adquiere. Son estos datos los que necesitamos para empezar a montar nuestra BBDD y los modelos estadísticos con los que después llegar a conclusiones con el suficiente peso como determinar que acciones finales y productos se ofrece a cada uno de los diferentes grupos que creemos.

Todos los clientes con tarjeta de fidelización recibirán las ofertas y propuestas de productos que determinemos mediante el análisis de todos los datos de que disponemos mediante pantallas digitales que montarán nuestros carritos nada más llegar al supermercado y tras identificarse con dicha tarjeta. Dentro de este proyecto está pensado el envío de dichas ofertas sobre productos habituales del cliente como de los que establezcamos para la venta cruzada a través de una campaña de e-mail semanal, haciendo el envío el día más propenso para cada uno de los grupos de clientes.

El objetivo como se indicó es doble, por un lado ofrecer productos en promoción que sean habituales en su cesta de la compra que aun haciéndoles gastar más dinero les pueda compensar y por otra parte realizar ventas cruzadas de productos que aún no siendo habituales en su cesta de la compra por el perfil de cliente que presenta sea altamente potencial a la adquisición de alguno de los productos para esa venta cruzada.

## **6.2 – Recopilación de datos y uso de los mismos:**

Para crear la bbdd con la que posteriormente trabajar y crear los modelos estadísticos que nos indiquen todo lo que ya comentamos líneas atrás necesitamos tener todos los datos en diferentes tablas con las que poder trabajar, las cuales deben hacer referencia a;

- Datos personales -> Id Cliente, Nombre y apellidos, DNI, edad, sexo, provincia y dirección de residencia, estado civil, número de personas con las que convive en el domicilio y el parentesco respecto a ellos, profesión, email, método de pago, teléfono, teléfono móvil, fax.
- Datos de consumo -> Gasto mensual del cliente en nuestros establecimientos, tipos de productos que compra y catalogación de los mismo, en que día de la semana se realiza la compra así como la estacionalidad en que estas compras ocurren, Id Cliente, Id Empresa, precio, unidades, fecha compra.
- Hábitos -> Estacionalidad, rango de gasto medio.
- LOPD -> Todo lo referente a lopd (Id Empresa, Id Cliente, fecha alta, etc).
- Productos -> Referencias que disponemos en los centros comerciales y algunas características más relevantes de los mismos.
- Datos WEB -> Movimientos por la página web, cookies, etc.
- Campañas Ofertadas -> Datos básicos sobre que campañas y a que clientes se les han ofrecido.

Así mismo a cada cliente dentro de nuestra base de datos le daremos un número de cliente (IdCliente = IdPersona) que será único con el que poder trabajar y también se calculará la antigüedad del cliente.

### **6.3 – Fases:**

Para garantizar el éxito del proyecto es importante subdividir en dos fases todo el proceso, de forma que en la primera de ellas (“piloto”), se realizan los análisis y pruebas con un grupo reducido de clientes para así poder conocer que alcance están teniendo los modelos estadísticos, que grado de acierto y mejora estamos logrando y rectificar las posibles desviaciones que se pudieran originar.

1. Seguir la evolución de los resultados:

- Realizando seguimiento a los clientes que aceptaron el descuento.
- Profundizando en los motivos de rechazo (entre 6 meses y 1 año).

2. Convertir el piloto en una campaña periódica, semanal, con las siguientes consideraciones:

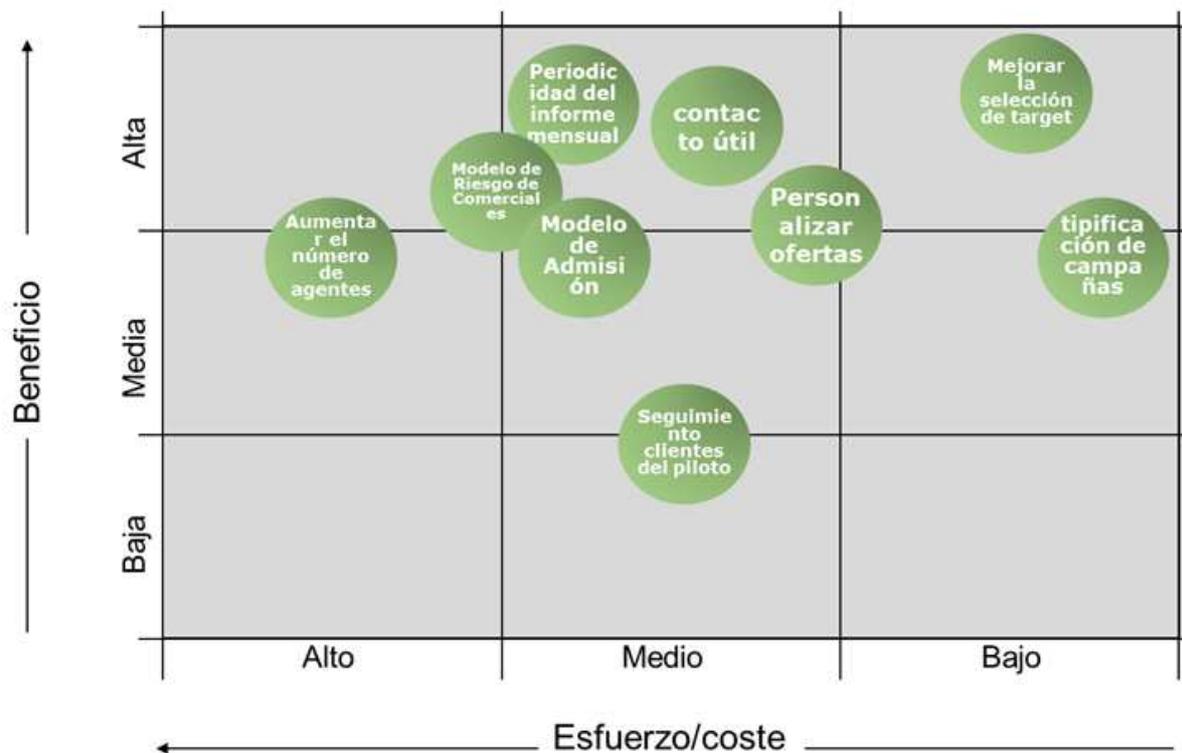
Fase 1 (Con los recursos actuales);

- Seleccionar el target, limitándolo a los 1.500 clientes con score más alto (15% de acierto) añadiendo los clientes que respecto al período anterior hayan aumentado su score.
- Diseñar las ofertas de modo “personalizado” siguiendo la estrategia de “Visión 360º cliente” y plantear, además de los descuentos:
  - Ofertas mediante mail y móvil para los clientes que, además de estar en los puntos más altos de abandono, están en los puntos más altos de propensión de compra para productos de venta cruzada.
  - Upgrades de venta para los clientes que tengan bajas compras (compras del tipo 3x2, descuento al comprar 3 o más unidades de un determinado producto).
- Mejorar el % de contacto útil y la tipificación de motivos de rechazo.

Fase 2 (Ampliando recursos);

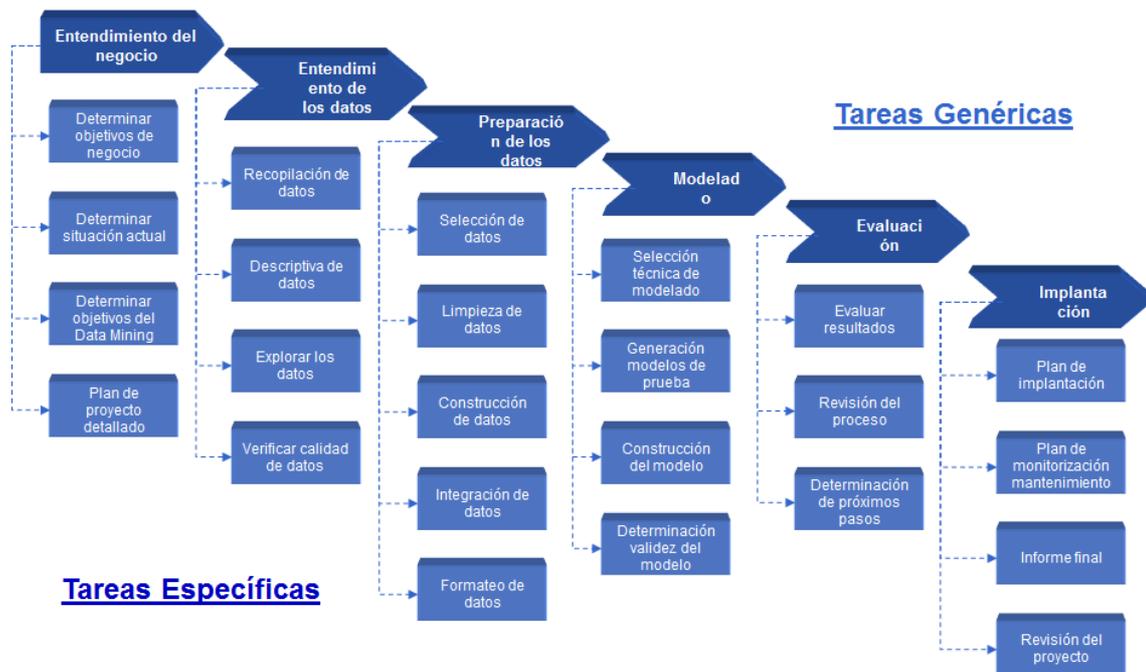
- Aumentando el target de la campaña a los 8.000 clientes.
- Aumentando la periodicidad del informe.

- Automatizando la salida de los Modelos señalando la oferta más adecuada.
- Ampliar la visión del cliente con el desarrollo de nuevos modelos.
- Revisión del Modelo de Admisión: Revisión y mejora de los actuales procesos de admisión de clientes de empresa. Segmentación de clientes por actividad económica (para grandes clientes, empresas).
- Oferta comercial personalizada: Desarrollar una oferta comercial personalizada por cliente.



*Oportunidades de mejora (elaboración propia)*

Para comprender como clasificamos las tareas, genéricas y específicas, echemos un vistazo al siguiente cuadro en el que se describen las que se engloba dentro de cada una de las genéricas y el orden de las específicas dentro de cada paso.

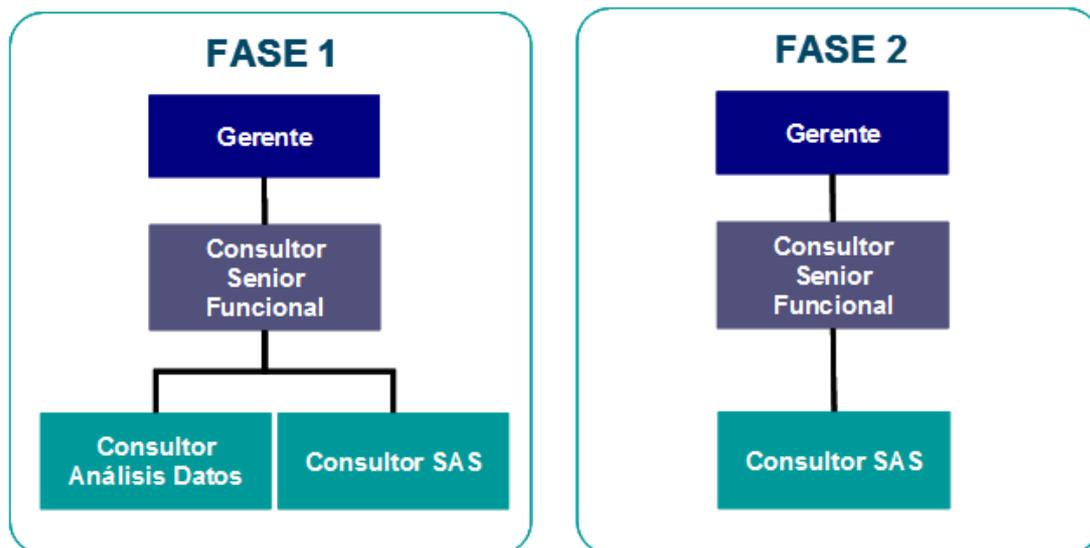


Tareas genéricas y específicas (elaboración propia)

## **6.4 – Equipo humano y técnico necesario:**

Considero necesaria la participación de tres consultores en la fase 1: un consultor senior experimentado en el análisis de datos y gestión de proyectos, un consultor de apoyo para el análisis de datos y un consultor experto en SAS.

Para la fase 2, al aprovechar gran parte del análisis de datos de la fase previa, se prescinde del consultor de análisis de datos.



*Equipo trabajo Fase I y Fase II (elaboración propia)*

Dentro del apartado técnico además de ordenadores y conexión a internet necesitamos varios programas con los que realizar los estudios, procesos y análisis posteriores. Dentro de los programas necesarios se encuentra SAS Enterprise GUIDE 9.4, herramienta sobre la que se montaran los modelos estadísticos para la obtención de resultados con la exportación de tabloneros finales. En dichos tabloneros vienen determinadas las ofertas y todo aquello que comentamos con anterioridad y base del trabajo.

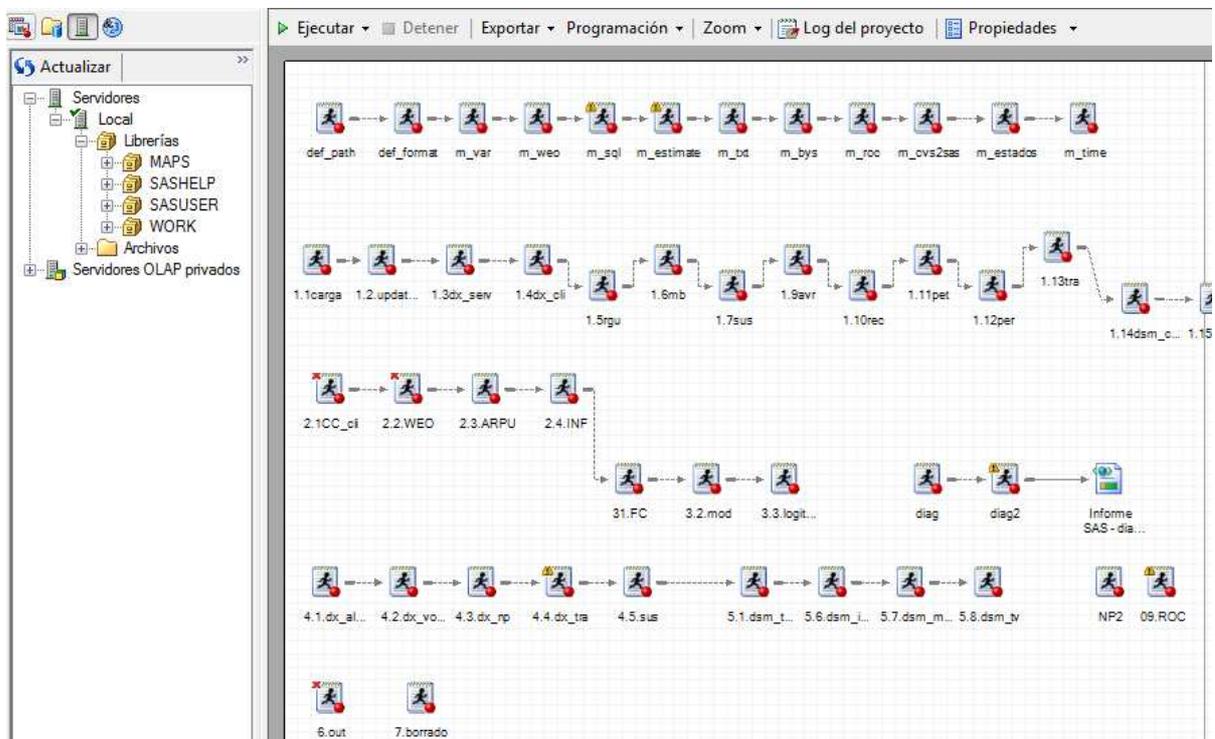
Para ello debemos arrancar el programa y abrir el archivo SAS logit\_v40.egp que se encuentra en el directorio raíz. El proyecto en SAS se compone de una serie de módulos que se deben de ejecutar secuencialmente.

El proyecto está dividido en 6 apartados.

- Cabecera (Es lo primero que debo compilar cada vez que arrancamos SAS para que pueda funcionar el resto de las cargas)

- Carga de datos
- Tablas de caracterización
- Estimación
- Modelos
- Resultados

Cada uno de estos apartados se corresponde aproximadamente con el conjunto de módulos que se sitúan en cada una de las líneas horizontales del proyecto.



*Nodos trabajo SAS Guide (SAS Guide)*

## **6.5 – Procedimiento habitual tras analizar los datos:**

Una vez que dispongamos de los modelos estadísticos montados sobre SAS donde se analizan todos los datos y hábitos de compra para nuestros clientes seremos capaces de determinar dentro del marco nacional (España es el marco de venta y actuación de nuestro cliente) qué grupos de clientes por rangos de edades, sexo y población son más idóneos para ofrecerles unos u otros productos. Con el análisis realizado deberemos preguntarnos qué características tienen cada uno de ellos, es decir, que diferencias encontramos en las diferentes ciudades y comunidades autónomas para así determinar si es posible agrupar por comunidad autónoma o debemos sumergirnos en mayor profundidad al nivel de cada una de las ciudades.

Si tenemos claro que, como hice referencia en varios momentos anteriores del trabajo, la línea de productos ofrecidos serán por una parte los que ya se consumen con cierta regularidad (carro de la compra) y por otra parte aquellos que sin estar de forma habitual en el carro de la compra nos sirven para realizar ventas cruzadas necesitamos tener claro que los grupos con lo que obtener el scoring estarán diferenciados pues tras realizar los análisis cabría la posibilidad de encontrarnos comportamientos muy diferenciadores entre comunidades autónomas o ciudades.

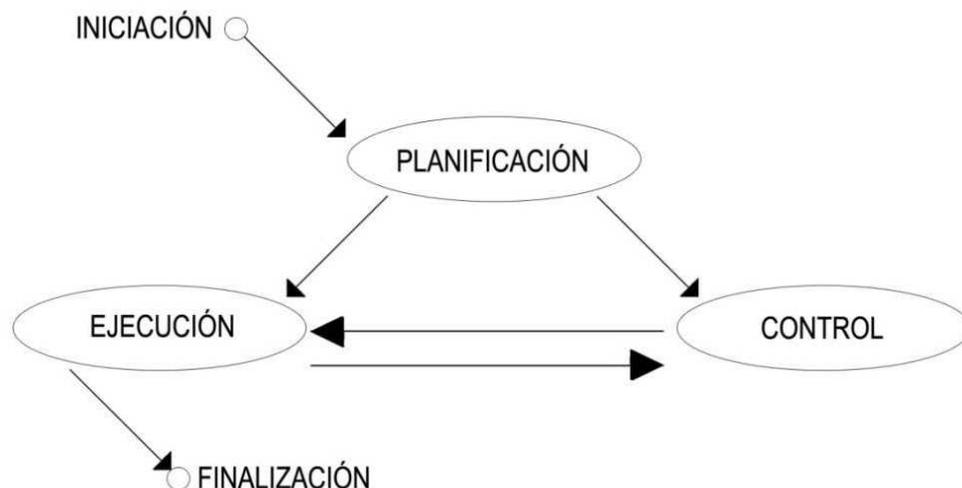
Así mismo cabe la posibilidad que un mismo producto en una zona del país funcione perfectamente con una demanda creciente y sin embargo en otra presente una demanda casi nula (influenciado por hábitos, climatología, costumbres culturales, etc) razón por la que debemos analizar la estructura del mapa nacional a nivel de consumo por productos. Con el mapa dibujado debemos hacer lo propio con el resto de algoritmos que hemos recogido y con ellos determinar por cada sexo que grupos de edades podemos agrupar, de forma y manera que al final tendremos una disgregación por cada sexo (masculino y femenino). Dentro de dichos grupos vamos a realizar subgrupos por rangos de edades los cuales vendrán muy marcados por la característica de la situación familiar y número que lo componen.

Nos encontraremos que dentro de los subgrupos de las edades, tanto para los grupos de ambos sexos, en determinados rangos la situación familiar estará más inclinada hacia una unidad familiar de pareja con hijos. Otro grupo predominante será la de aquellos, también para ambos sexos, en que la situación familiar sea la de la propia unidad, es decir hombres y mujeres que viven solos. De este otro grupo tendremos mucho peso en subgrupos para los rangos de edades menores de 36 años y por encima

de los 67 años, que hacen mención a personas aún sin establecer un nicho familiar en pareja o aquellas mayores en situación de viudedad o separación.

Con todo esto y analizando que productos en que fechas se adquieren y agrupando por la estacionalidad en que ocurren podremos obtener un tablón en el que poder determinar para cierto grupo de personas, con un rango de edad de los que describamos, con una comunidad autónoma determinada y sexo concreto los productos que más consumen y por tanto los validos para ventas cruzadas son unos muy concretos.

Para la gestión del proyecto y para su posterior comprobación debemos acometer las siguientes fases que se pueden ver ilustradas en la siguiente imagen;



*Fases del proyecto (elaboración propia)*

La manera de proceder una vez que disponemos de toda la información para realizar nuestro trabajo será realizando un mailing todos los días, el target será diferente y dependiente de los resultados de nuestro análisis, se mandarían los mails personalizados a los grupos de clientes con las ofertas que entendemos serán de su interés ya que la mayoría componen artículos que se encuentran de forma habitual en su carro de la compra y por otra parte irán incluidos artículos de venta cruzada con esos que ya son habituales.

De esta manera conseguimos incrementar las ventas que ya eran habituales ya que al ofertar una ventaja a modo de ahorro para el cliente le interesará aunque sea incrementando el volumen de la compra y con ello nosotros vemos incrementado el

volumen de negocio, por tanto un mayor beneficio para la empresa (economías de escala).

Por otra parte está el beneficio que nos reportaran todas esas ventas cruzadas, hasta la fecha se realizan sin determinar ningún target y únicamente colocando determinados productos en localizaciones específicas de cada centro comercial, con los análisis y trabajo realizado conseguiremos que las ofertas de ventas cruzadas que se lanzan tenga un mayor éxito puesto que se enfocan a los diferentes targets que obtuvimos del estudio.

Como ya he comentado con anterioridad, una vez que obtengamos los resultados con los que trabajar, se lanzara un mail diario por tipo de target al que vaya enfocado, que está directamente relacionado con el peso específico que dicho target tenga en realizar la compra ese mismo día.

Esa información de forma personalizada recaerá directamente en su cuenta de e-mail pero además debemos plasmarla en la nueva aplicación de la empresa que el cliente puede instalarse en su teléfono móvil, no solo porque en ella estarán reflejadas las ofertas propias de su target sino porque además con ella podrá acceder a novedades de nuestros centros comerciales, identificarse con su tarjeta de cliente de manera virtual pues la tarjeta física no la llevamos siempre con nosotros.

Además de todo lo novedoso respecto al tratamiento, análisis y obtención de resultados con los datos, mediante métodos estadísticos existe otra gran novedad aún sin implantar a nivel nacional a la que hice referencia en párrafos anteriores, se trata de pantallas inteligentes en los carros de la compra de nuestras grandes superficies mediante los cuales el cliente una vez se identifique podrá ir observando todo aquello que está ligado a su cuenta de cliente respecto a ofertas, productos, promociones y artículos que designamos para nuestras ventas cruzadas.

Por otra parte, mediante esas pantallas digitales en los carros de la compra, podremos con un escáner de productos situados en el anverso de dicha pantalla (la parte que apunta directamente al interior del carro de la compra) procesar que artículos se están cargando y así indicar al cliente que otros productos serían buenos o necesarios para combinar con los que está cargando en el carro, o bien porque los pudiera necesitar para cocinar.



Esta función ayudara al cliente a no dejar sin comprar productos que pudiera necesitar, lo que sería una pérdida de ventas para la empresa.

Aprovechando que dispondremos de las pantallas digitales en los carros de los supermercados crearemos dentro del programa informático una lista de las recetas más utilizadas en todo el panorama nacional, y de esa lista lo que haremos será cargar en la aplicación un apartado denominado ¿Qué productos necesito? por cuya virtud el cliente podrá seleccionar una receta que desee preparar para que en la pantalla le indique cuales son todos los productos necesarios.

Esta última sección será muy interesante para cocineros principiantes o para aquellos más avanzados que deseen preparar alguna receta que no hubieran hecho hasta la fecha.

## **7 - Conclusiones:**

Las conclusiones acerca del Business Intelligence, utilizando un gran volumen de datos sobre la cartera de clientes y mediante métodos estadísticos apropiados podremos lograr muestras fiables que serán de gran utilidad para anticiparse a las decisiones del mercado y saber de antemano datos relevantes que nos proporcionarán esa información que necesitamos para la toma de decisiones.

Con el tiempo el análisis predictivo ha llegado a su madurez y se ha convertido en una práctica corporativa esencial y necesaria para mantener la ventaja sobre la competencia. Esta tecnología representa una fase completamente nueva en la evolución empresarial en tanto que lleva a escena el aprendizaje organizativo, que hace que el negocio crezca puesto que implementa una sola forma de gestión de riesgos basada en datos a través de diversos escenarios. Este trabajo presenta siete objetivos estratégicos que solo podrán alcanzarse absolutamente haciendo uso del análisis predictivo, a saber, Competir, Crecer, Consolidar, Mejorar, Satisfacer, Aprender y Actuar.

### *1. Competir*

Cuando los productos que ofrece una empresa llegan a ser de consumo masivo, la ventaja competitiva dependerá de las mejoras que se realicen en los procesos de negocio. Aplicando el análisis predictivo, cuya propuesta de valor consiste en tomar decisiones empresariales con mayor prontitud y en hacer progresar sus funciones con el aprendizaje organizativo aplicado.

Este proceso de aprendizaje proporciona una ventaja competitiva única y revela los puntos débiles de la competencia. Los resultados de las encuestas demuestran que un "entorno competitivo más duro" es con diferencia la razón de mayor peso por la que las organizaciones adoptan el análisis predictivo. El análisis predictivo ofrece una poderosa, singular y cualitativa diferenciación al proporcionar inteligencia empresarial con la que podemos competir en ventas y en preservación de clientes. Un modelo predictivo generado a partir de los datos de su empresa se sirve de la experiencia de la que solo ésta es concedora, por tanto, la perspicacia y la inteligencia del modelo están fuera del alcance del conocimiento común, y las altas probabilidades que marca componen una lista de contactos personalizada. En contraste con las tendencias más extendidas que pueden medirse a gran escala, un modelo predictivo consigue patrones y tendencias de compra más refinados en forma de micro segmentos de clientes.

Puesto que los datos por los que se instruye el modelo incluyen las ventas a clientes que presumiblemente tienen conocimiento no solamente de sus productos y marketing, sino también de los de la competencia, el proceso de modelado aprende a distinguir eficazmente micro segmentos de clientes que eligen su empresa de aquellos que se pasan a la competencia. De este modo, podemos identificar exactamente de qué adolece su competidor, sus puntos débiles, tal como revelan las tendencias de comportamiento que se encuentran codificadas en sus datos. La empresa promueve entonces este conocimiento, actuando en función de estas oportunidades y llevando a cabo actividades de ventas y marketing con un enfoque más preciso.

## 2. Crecer

Las aplicaciones de marketing y ventas de los análisis predictivos constituyen las principales propuestas de valor. Cada cliente recibe una puntuación predictiva en lo que respecta al comportamiento relacionado con las ventas, ya sean compras, respuestas, fugas y clics. Las puntuaciones determinan entonces las operaciones empresariales con respecto al comportamiento en marketing, ventas, atención al cliente y sitio web. De este modo, el análisis predictivo ofrece su ventaja competitiva única a determinadas actividades que incluyen el trato directo con el cliente.

La propuesta de valor es simple, al suprimir aquellos clientes con menor probabilidad de respuesta, los costes se reducen drásticamente y los beneficios crecen.

Si el 40% de los clientes identificados como los de mayor probabilidad de respuesta incluyen el 80% del total de los que responden, se puede eliminar una buena parte del otro 60%, lo que permite ahorrar sobremanera en los costes de las campañas. La creación de modelos de fuga de clientes puede ser la aplicación empresarial más atractiva del análisis predictivo.

Puesto que conservar los clientes es el principal objetivo de muchas organizaciones, los incentivos para lograr conservarlos de forma eficiente, como las ofertas de descuentos, pueden ser bastante costosos. La ganancia se consigue cuando una oferta semejante se dirige solamente a aquellos clientes con mayor probabilidad de fuga.

### 3. Consolidar

Cuando las transacciones crecen considerablemente en número y se automatizan, se multiplican las oportunidades de delinquir. En todos los sectores, las transacciones fraudulentas relacionadas con facturas, compras con tarjeta de crédito, devoluciones de impuestos, reclamaciones de pólizas de seguros, llamadas de teléfonos móviles, clics sobre publicidad online y cheques bancarios de clientes incurren en enormes costes. Puntuar y clasificar las transacciones con un modelo predictivo sirviéndose de la experiencia recopilada por la empresa en relación con el fraude contribuye a aumentar sobremanera la detección del fraude.

Puesto que un equipo de investigadores solamente puede inspeccionar un número fijo de transacciones sospechosas por semana, el hecho de ofrecer una base mejor depurada de transacciones sospechosas –con menos “falsas alarmas” (falsos positivos)- hace que se detecten más fraudes y que se impidan más pérdidas o que se recuperen.

Por ejemplo, la detección de fraudes en las reclamaciones de seguros de automóviles ha multiplicado por 6,5 veces su capacidad de detección si la comparamos con lo que se consigue sin medios para clasificar y puntuar las reclamaciones de seguros. De forma análoga, el análisis predictivo se extiende también a la seguridad de la información con la detección de intrusiones online por parte de hackers y virus, así como a la identificación de criminales exigida por ley.

### 4. Mejorar

Después de aumentar las ventas y consolidar la integridad de las transacciones empresariales, queda mejorar el producto de la empresa y la eficiencia con la que se produce. Ya se trate de fabricar productos o de prestar servicios, el cometido central de la empresa es producir y suministrar cada vez con mayor eficacia. A este respecto, la predicción desempeña un papel clave en el progreso de la actividad principal del negocio. En el sector de los seguros el valor y la posición que ocupen los productos de la empresa -pólizas de seguros- dependen de la puntuación predictiva en función del riesgo.

Si se identifica de forma más precisa a aquellos solicitantes con un riesgo mayor de dar curso a cuantiosas reclamaciones, los proveedores de seguros pueden llevar a cabo una selección más eficaz y tomar decisiones asimismo eficaces en relación con los precios de dichas pólizas.

Las entidades crediticias como los bancos gestionan los riesgos de forma similar, puesto que lo predicen puntuando a los solicitantes de préstamos y de tarjetas de crédito de acuerdo con su probabilidad de que eventualmente no cumplan los pagos. El análisis predictivo mejora la fabricación, inspección y reparación de productos de distintas maneras. Durante la producción los artículos defectuosos son detectados en la cadena de montaje. Una vez puestos los productos en circulación, la creación de modelos de fiabilidad determina qué componentes tienen mayor probabilidad de fallar o, en respuesta a las llamadas de los clientes, cuáles probablemente van a necesitar ser reparados. Se acerca una ola prácticamente sin límites de nuevas “aplicaciones rompedoras”.

Las recientes innovaciones continúan ampliando la gama de aplicaciones de análisis predictivo, en tanto que mejoran su capacidad esencial y hacen progresar la actividad principal de las empresas de distintos tipos de negocios. Otros ejemplos incluyen la optimización de la cadena de suministro con predicción de la demanda según el inventario, procesamiento de solicitudes mediante la predicción de aprobaciones y denegaciones, rendimiento de recursos humanos y creación de modelos de las bajas de personal que servirán de base a la hora de seleccionar nuevas incorporaciones, contrataciones y retención del capital humano, marketing sanitario proactivo mediante la predicción de riesgos relacionados con la salud, comercialización algorítmica (caja negra) mediante la predicción de mercados.

##### 5. Satisfacer

Durante más de una década se ha dicho que la experiencia del consumidor será el “próximo campo de batalla para ganar en competitividad”. Dando por supuesta la agudeza de la competencia, el análisis predictivo será el medio con el que mantenerse de pie en este campo. Las ventajas que proporciona a la empresa –de las que ya hablé en este trabajo- suponen un “reverso” no menos próspero: ventajas para el consumidor que consigue más por menos, de forma más fácil y fiable. Mayor relevancia mediante un marketing que selecciona sus objetivos de forma más precisa. A medida que pasa el tiempo cada vez se tolera menos la publicidad por correo y el SPAM. Las recomendaciones sobre productos son cada vez más visibles, se valoran más e incluso se esperan.

El análisis predictivo constituye un punto de venta explícito para el consumidor final, mejores productos y servicios mejorando la capacidad esencial del negocio. Las mejoras que ofrecen el control de calidad analítico, la creación de modelos de fiabilidad,

los servicios perfeccionados y el rápido procesamiento de las aplicaciones satisfacen esas demandas que cada vez son más exigentes del consumidor. Los consumidores finales, al igual que los clientes corporativos B2B, siempre se inclinarán hacia un mayor conocimiento, agilidad y un gusto refinado. Crecen las opciones para elegir proveedores, productos, servicios y los consumidores se sienten confiados gracias a un conocimiento exhaustivo, como el que se obtiene online.

A medida que aumentan los niveles de fraude, los consumidores permanecen expectantes ante el grado de protección, y además son cada vez más sensibles a los inconvenientes que conllevan las advertencias de fraude que resultan ser "falsas alarmas" y que interrumpen el procesamiento de las transacciones de pago. Precios más baratos mediante una mayor eficiencia. Con una fuerza competitiva asegurada, el éxito genera éxito. El crecimiento empresarial comporta nuevos ahorros, tanto en la propia ejecución de análisis predictivos como en la productividad de la empresa en general.

#### 6. Aprender

Las técnicas de elaboración de informes empresariales, incluidas las tarjetas de resultados, paneles de control, indicadores de rendimiento, procesamiento analítico en línea, consultas ad hoc y segmentación estándar de marketing como el método RFM, ofrecen un análisis retrospectivo. La capacidad del análisis predictivo para aprender de la experiencia es lo que hace que esta tecnología sea predictiva, distinguiéndola de otros sistemas de inteligencia artificial y técnicas analíticas. Los modelos predictivos se han diseñado de forma excepcional para optimizar las predicciones más inmediatas, como la fuga de clientes. Este proceso de optimización constituye un acto de aprendizaje puesto que para tener éxito debe ejecutar generalizaciones a partir de datos en forma de patrones y pautas matemáticas, y estas generalizaciones deben confirmarse cuando se apliquen a próximos casos aún no conocidos.

La obtención de un sólido modelo predictivo a partir de datos de la organización es un verdadero acto de aprendizaje a partir de la experiencia. El arte antecede a la ciencia en esta "mágica" habilidad para generalizar a partir de ejemplos. Aun cuando existan multitud de datos, con millones o billones de registros, el diseño de un método por el cual la máquina puede generalizar, encontrando patrones que se dan no sólo en los datos disponibles sino en general (conocido también como inducción), recae en la intuición y las ideas humanas. Afortunadamente, la ciencia sirve para hacer que estas

ideas sean formales y para medir el grado de eficacia en que generaliza realmente el modelo predictivo resultante.

Los métodos establecidos de análisis predictivo son competentes y sólidos, en tanto que producen modelos predictivos que funcionan. Estos métodos incluyen árboles de decisiones, regresión logística, Naïve Bayes y redes neuronales. El análisis predictivo integra y aprovecha las potentes fuentes de datos como los datos sociales y el texto no estructurado. Una gran empresa de telecomunicaciones duplicó el rendimiento de los modelos de fuga de clientes integrando datos sociales, ya que si los amigos de un suscriptor de llamadas cambian de operador de telefonía móvil, es muy probable que el suscriptor lo haga también. De forma análoga, una gran compañía de telecomunicaciones de Norteamérica ha demostrado que los clientes que ya han realizado una cancelación en su red de telefonía tienen una probabilidad del 600% de volver a hacerlo. Una importante universidad online se centra en los amigos de los actuales estudiantes, puesto que las probabilidades de que éstos también se matriculen son del 100%.

### 7. Actuar

Los profundos conocimientos que se obtienen a partir de los sistemas estándar de inteligencia empresarial y de elaboración de informes no son inmediatamente procesables, primero deben ser llevados a la práctica por medio del juicio humano. Los métodos de medición, los informes, los paneles de control y otros análisis retrospectivos son componentes importantes de la inteligencia empresarial, pero su ejecución será apropiada en tanto que esté claro a priori qué tipo de acciones o decisiones se recomiendan, si es que se recomienda alguna.

En cambio, el análisis predictivo está específicamente diseñado para generar imperativos de acciones concluyentes. Toda puntuación predictiva de un cliente lleva a realizar una acción con respecto a ese cliente. De este modo, el análisis predictivo se convierte, por su diseño, en el modelo de mayor capacidad de procesamiento de la inteligencia empresarial.

El análisis predictivo ofrece una suculenta ganancia global en tanto que impulsa millones de decisiones operativas, tales como si conviene enviar un correo, llamar, ofrecer un descuento, recomendar un producto, mostrar un anuncio publicitario o dedicar recursos de ventas en una iniciativa.

En cuanto a la gestión de fraudes, el modelo predictivo fomenta las decisiones para llevar a cabo auditorías, investigaciones o impedir los fraudes. Y en el caso de aplicaciones empresariales esenciales, las decisiones basadas en modelos analíticos incluyen si conviene inspeccionar un artículo o sistema por si presenta defectos, cargar un componente en un camión de reparaciones, enviar asistencia, realizar un rápido seguimiento de una aplicación o comprar activos. Los modelos predictivos nos hablan sobre los clientes, nos revelan intenciones, preferencias y las causas de su comportamiento. Además de impulsar las decisiones con las puntuaciones resultantes, un modelo predictivo ofrece valor en una segunda manera: inspeccionando las reglas o patrones internos del modelo a menudo se descubren conocimientos estratégicos.

Aunque por naturaleza son apropiados para su fin, tales conocimientos son un bien común y proporcionan un saber selectivo, puesto que todo modelo predictivo se optimiza para un objetivo de predicción concreto. Por ejemplo, un modelo de fuga de clientes reveló que los usuarios con intentos fallidos de inicio de sesión –posiblemente por haber olvidado su contraseña- tenían un alto riesgo de abandono. Saber esto da lugar a la creación de un programa asistente con recordatorios de contraseña, con lo que potencialmente se ahorran los costosos descuentos de permanencia que dichos clientes recibirían si solo se utilizara la puntuación predictiva. Los conocimientos sobre modelos que adquirió una prestigiosa red social en línea hicieron que identificara los factores clave para garantizar la permanencia desde un principio, tales como si el usuario había cargado una fotografía, y que implicaron acciones que doblaron la permanencia de los clientes de una y dos semanas.

## **8 - Bibliografía:**

### **Noticias/Artículos**

- Matthew B. Rice (2004), “*Business Intelligence* tools can help turn out savings in core cost areas”, Managed Healthcare Alliance.
- Mark Beyer y Donald Feinberg (2006)“The Basic of datawarehousing”.
- Wayne W. Eckerson (2002), “Data quality and the bottom line”, TDWI, Report Series.
- Hugh. J. Watson (2006), “Recent Development in datawarehousing” publicado en [http://www.terry.uga.edu/~hwatson/dw\\_tutorial.ppt](http://www.terry.uga.edu/~hwatson/dw_tutorial.ppt).
- D. Vesset (2006),“QlikTech´s Approach to Business Intelligence: Keep it Simple and Flexible”.
- “Next-Gen BI is here”.(2009). Information Week, InformationWeekanalytics.com.
- BARC. Business Application Research Center.(2007).<http://www.bi-verdict.com/>
- Lyndsay Wise. (2007), “Cinco pasos para lograr un proyecto de Business Intelligence exitoso”.
- Turban, E. (1995). Decision support and expert systems: management support systems. Englewood Cliffs, N.J., Prentice Hall.
- W. Eckerson et (2005) “Business Intelligence: Competir con la información”. Josep Lluís Cano.
- “The future of I.T: What´s on tap for 2007 and beyond”.
- Stephen Brobst Chief Technology Officer (2010) - Teradata Corporation.
- "A Business Intelligence System" IBM Journal. Julio 2008
- Data Quality for BI  
[http://www.trilliumsoftware.com/home/solutions/enterprise/datawarehouse\\_bi.as](http://www.trilliumsoftware.com/home/solutions/enterprise/datawarehouse_bi.as)
- “La gestión del conocimiento. Su importancia como recurso estratégico para la organización” Sanguino, R, (2003).
- “Repositorios OLAP”. Carlos Hurtado Larrain. (2005) Presentación en el Congreso de Ciencia y Tecnología Computacional, Tuluá.

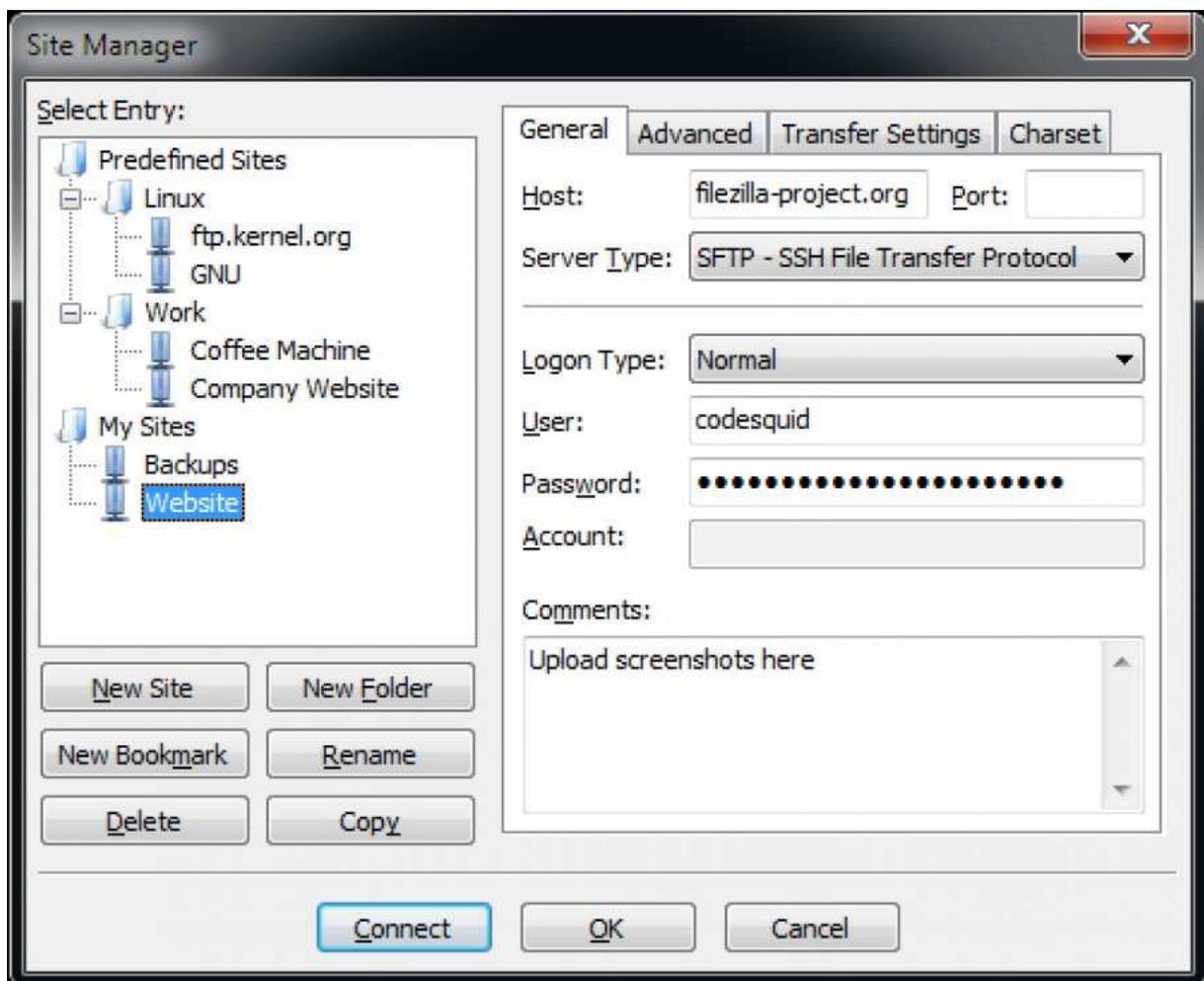
- “OLAP and Datamining: Bridging the Gap” <http://www.datamining.com/bridge.htm>
- Manuel de la Herrán Gascón. (2001) “Cómo diseñar grandes variables en bases de datos multidimensionales”.
- Sergio A. Moriello. (1999) “Datamining: Oráculos de silencio. Buceando en un mar de información”.
- Douglas D. Dankel, Avelino J Gonzales. Prentice-Hall, (1993). “The engineering of knowledge-based systems theory and practice”.
- Wayne Eckerson, (2005) “Enterprise BI: Strategies and technologies for deploying BI on an enterprise scale”.
- Harvard Business School Press, Boston, (1996). “The Balanced Scorecard: Translating Strategy into Action”.
- Ralph Kimball, Laura Reeves. Wiley computer publishing, (1998). “The Datawarehouse lifecycle toolkit”.
- Monjo Carrió y Luis Maldonado Ramos (2001). “Patología y técnicas de intervención en estructuras arquitectónicas”.

Enlaces web

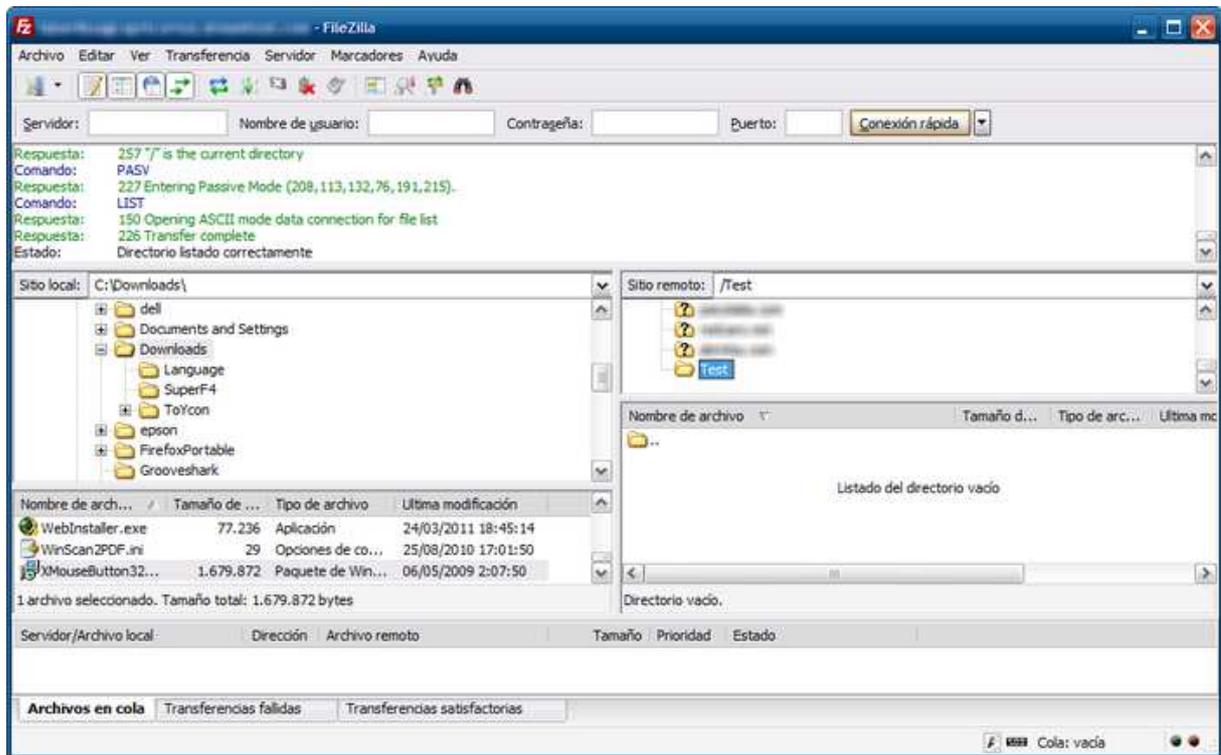
- Microsoft BI: <http://www.microsoft.com/bi/>
- [http://itilv3.osiatis.es/proceso\\_mejora\\_continua\\_servicios\\_TI/DIKW.php](http://itilv3.osiatis.es/proceso_mejora_continua_servicios_TI/DIKW.php)
- [www.gartner.com](http://www.gartner.com).
- <http://www.bi-spain.com>
- SAP BI: [www.sap.com](http://www.sap.com)
- QlikView: <http://www.qlikview.com>
- Oracle BI: <http://www.oracle.com/>
- Targit: <http://www.targit.com>
- [www.nase-it.com](http://www.nase-it.com)
- [www.todobi.blogspot.com](http://www.todobi.blogspot.com)
- [www.BIScorecard.com](http://www.BIScorecard.com)

## 9 - Anexos:

Para concluir con el TFG veremos a través de pantallazos más una breve descripción de cada uno de ellos, cuáles serán los pasos habituales desde que el cliente nos facilita los archivos necesarios para nuestro trabajo y análisis en un servidor FTP, que será la manera de proceder al intercambio de archivos entre la empresa y nosotros, hasta la obtención de resultados con los que poder proceder.



*FTP (elaboración propia)*



#### FTP (elaboración propia)

Una vez logados accedemos al gestor de FTP donde la empresa nos habrá dejado los ficheros correspondientes, los copiaremos en nuestro ordenador para comprobar que todos disponen de las características y parámetros necesarios.

| Nombre                        | Fecha de modifica... | Tipo              | Tamaño |
|-------------------------------|----------------------|-------------------|--------|
| 20160701_Clientes.txt         | 30/06/2016 18:07     | Notepad++ Docu... | 5 KB   |
| 20160701_Consumo.txt          | 30/06/2016 18:50     | Notepad++ Docu... | 2 KB   |
| 20160701_Direcciones.txt      | 30/06/2016 18:52     | Notepad++ Docu... | 3 KB   |
| 20160701_LOPD.txt             | 30/06/2016 18:49     | Notepad++ Docu... | 1 KB   |
| 20160701_Navegacion_Bruto.txt | 30/06/2016 18:56     | Notepad++ Docu... | 2 KB   |
| 20160701_Productos.txt        | 30/06/2016 18:51     | Notepad++ Docu... | 3 KB   |

#### Ficheros recibidos (elaboración propia)

Para los ficheros que nos proporciona la empresa mensualmente necesitamos que dentro contengan la siguiente información y que además presenten cada uno de ellos unos parámetros diferentes y característicos como muestran las siguientes imágenes;

Fichero Clientes:

| IDPERSONA | TIPONIF | NIF           | NOMBRE                   | APELLIDO1 | APELLIDO2 | FECHANACIMIENTO | SEXO | ESTADOCIVIL | TELEFONO_FIJO | TELEFONO_MOVIL | FAK                                 | EMAIL      | SITUACIONLABORAL | SITUACION_FAMILIAR |            |
|-----------|---------|---------------|--------------------------|-----------|-----------|-----------------|------|-------------|---------------|----------------|-------------------------------------|------------|------------------|--------------------|------------|
| 1         | 1       | nif 09089944E | RAMON JESUS GONZALEZ     | FIGUEREDO |           | 02/06/1936      | H    | soltero     | 986090808     |                |                                     |            |                  | Activo             | 14/04/2016 |
| 2         | 2       | nif 52234482X | SEVERIANA TAMAYO         | CAMPON    |           | 11/04/1944      | M    | casado      | 986690783     | 666777888      | Activo                              | 20/02/2016 |                  |                    |            |
| 3         | 3       | nif 23044649Y | JUAN IGNACIO VAZQUEZ     | RIVERO    |           | 19/05/1944      | H    | divorciado  | 654386373     |                | FCARUNCHO@TERRA.ES                  | Desempleo  | 17/03/2016       |                    |            |
| 4         | 4       | nif 11989122G | JUAN ANTONIO FERNANDEZ   | CUENCA    |           | 01/06/1944      | H    | casado      | 655934486     |                |                                     |            | Jubilado         | 14/04/2016         |            |
| 5         | 5       | nif 13292380J | SANTIAGO BENITO          | GARCIA    |           | 01/09/1930      | H    | divorciado  | 686918754     |                | INMIS@GRUPOINTERPATTY.COM           | Jubilado   | 17/03/2016       |                    |            |
| 6         | 6       | nif 50651900Q | ELENA CAMPO              | DEL       |           | 01/11/1937      | M    | casado      | 987411731     | 638129775      | CENTRAL@GRUPOINTERPATTY.COM         | Activo     | 17/03/2016       |                    |            |
| 7         | 7       | nif 72816118G | MARINA COLETO            | CABACO    |           | 01/01/1940      | M    | soltero     | 914571647     | 696994597      | PABLOBLANCO@GLOVALMULTICOVER.COM    | Activo     | 17/03/2016       |                    |            |
| 8         | 8       | nif 07763883D | MANUEL ALBERTO DOMINGUEZ | SANZ      |           | 30/08/1944      | H    | soltero     | 915552942     | 696994590      | ALONSO@RAOPEL.COM                   | Desempleo  | 21/04/2016       |                    |            |
| 9         | 9       | nif 13441771P | JENIFER CASADO           | MARTIN    |           | 30/09/1944      | NULL | soltero     | 961343728     | 663928905      | JARAMIREZ@ASERBROKERS.ES            | Activo     | 14/04/2016       |                    |            |
| 10        | 10      | nif 09706493J | SHEILA QUINTANA          | GARCIA    |           | 01/03/1944      | M    | soltero     | 963911132     | 687723616      | AUTOTAKIMERCEDES@RAMADA@HOTMAIL.COM | Activo     | 20/05/2015       |                    |            |
| 11        | 11      | nif 09706467R | MICHEL ANGEL BLAS        | SIERRA    |           | 01/12/1942      | H    | divorciado  | 961343728     |                | FERMOAUTO@FERMOAUTO.COM             | Desempleo  | 26/02/2014       |                    |            |
| 12        | 12      | nif 09706474P | JOSE ANTONIO ALONSO      | BUENO     |           | 01/12/1944      | H    | casado      | 961340666     |                | ANA.BELEN@FERMOAUTO.COM             | Activo     | 19/02/2014       |                    |            |
| 13        | 13      | nif 09706477B | VERONICA ALBENCA         | TEJERA    |           | 01/07/1944      | M    | viudo       | 608959311     |                | TEST@TEST.COM                       | Activo     | 17/07/2015       |                    |            |
| 14        | 14      | nif 09706531L | VICTOR MUÑOZ             | MORENO    |           | 01/08/1942      | H    | casado      | 961343727     |                | GRUASCHIC@TERRA.COM                 | Desempleo  | 26/02/2014       |                    |            |
| 15        | 15      | nif 09706613D | M JOSE LEON              | LEVYA     |           | 01/01/1950      | NULL | divorciado  | 961343728     | 676080642      | ADMINISTRACION@GRUASDELACHICA.COM   | Activo     | 26/02/2014       |                    |            |
| 16        | 16      | nif 09706640J | JOSE FRANCO              | DELGADO   |           | 01/08/1946      | H    | casado      | 961343728     | 961343728      | ADMINISTRACION@GRUASDELACHICA.ES    | Activo     | 26/02/2014       |                    |            |
| 17        | 17      | nif 09706663J | MARIA ARACELI HERRERO    | GONZALEZ  |           | 01/08/1948      | M    | divorciado  | 961340027     | 699317300      | MARIB@JBAENA.E-TELEFONICA.NET       | Desempleo  | 26/04/2014       |                    |            |
| 18        | 18      | nif 09706703F | FRANCISCO ULECIA         | GARCIA    |           | 01/02/1940      | H    | casado      | 963811132     | 963811132      | TALLERESDELAPUENTE@TELEFONICA.NET   | Activo     | 19/02/2014       |                    |            |
| 19        | 19      | nif 09706846N | ENCARNACION SORIA        | SERRANO   |           | 07/05/1944      | M    | divorciado  | 938880201     | 658926292      | TALLERES_DAVILA@YAHOO.ES            | Activo     | 29/06/2015       |                    |            |
| 20        | 20      | nif 09706861G | LINA MARIA GOYNECHE      | ARETA     |           | 01/09/1945      | NULL | casado      | 934590723     |                | GRUASCAMPOS1@GMAIL.COM              | Desempleo  | 31/07/2015       |                    |            |
| 21        | 21      | nif 09706861G | RICARDO MERCHAN          | RODRIGUEZ |           | 01/03/1940      | NULL | divorciado  | 938852058     | 669440199      | GRUASJUAN@GRUASJUAN.COM             | Activo     | 27/02/2014       |                    |            |
| 22        | 22      | nif 09706881R | PAOLA ARMARIO            | DELGADO   |           | 01/09/1940      | M    | casado      | 938894169     |                | TAKRAPIT@TAKRAPIT.ES                | Activo     | 26/02/2014       |                    |            |
| 23        | 23      | nif 09707104V | JULIAN PAREJO            | RUIZ      |           | 01/12/1944      | H    | divorciado  | 934590723     |                | TAXIONDA@HOTMAIL.COM                | Desempleo  | 27/04/2015       |                    |            |
| 24        | 24      | nif 09707160G | MIGUEL ANGEL SANCHEZ     | MOLINA    |           | 27/07/1944      | H    | casado      | 9869300611    | 676996799      | TALLERESRUFINO@BOSCH-BCS.COM        | Activo     | 26/02/2014       |                    |            |
| 25        | 25      | nif 09707247E | LUISA GONZALEZ           | SANCHEZ   |           | 01/12/1944      | M    | viudo       |               |                | GRUASSEBASTIAN@GONZALEZ@GMAIL.COM   | Activo     | 26/04/2013       |                    |            |
| 26        | 26      | nif 09707394V | MARIA CARMEN GARCIA      | PICON     |           | 14/12/1944      | M    | casado      |               |                | TAKISANGEL@VALENCIA@HOTMAIL.COM     | Desempleo  | 26/02/2014       |                    |            |
| 27        | 27      | nif 09707496H | MANUEL ALVARADO          | GALIA     |           | 30/07/1939      | H    | divorciado  | 986847430     | 699475938      | TALLERES@GRUAS@MARTINEZ@HOTMAIL.COM | Activo     | 26/02/2014       |                    |            |
| 28        | 28      | nif 09707571R | VICENTE SIMON SANCHEZ    |           |           | 10/12/1936      | H    | casado      | 958450389     |                | AUTOSIGUERMAN@AUTOSIGUERMAN.ES      | Activo     | 27/05/2015       |                    |            |
| 29        | 29      | nif 09707618R | AMELIO ANTONIO ALONSO    | PAVON     |           | 01/09/1945      | H    | casado      | 629463084     |                | AUTOSIGUERMAN@AUTOSIGUERMAN.ES      | Desempleo  | 26/02/2014       |                    |            |
| 30        | 30      | nif 09707645V | ANASTASIO CASTILLO       | GONZALEZ  |           | 01/10/1969      | H    | divorciado  | 629859192     |                | ADMN@AUTOSIGUERMAN.ES               | Activo     | 26/02/2014       |                    |            |
| 31        | 31      | nif 09707668Y | JOSE MARIA GUDINO        | TRINIDAD  |           | 01/10/1934      | H    | casado      | 649712242     |                | ADMN@AUTOSIGUERMAN.ES               | Jubilado   | 13/08/2015       |                    |            |
| 32        | 32      | nif 09707684E | ANTONIO MARIA BARQUERO   | PEREZ     |           | 17/01/1943      | H    | divorciado  |               |                |                                     |            | Activo           | 14/03/2014         |            |
| 33        | 33      | nif 09707707E | PABLO FRIAS              | TORRES    |           | 01/08/1944      | NULL | divorciado  | 961343737     | 667509743      | TAXI_RAMOS@HOTMAIL.COM              | Jubilado   | 26/02/2014       |                    |            |
| 34        | 34      | nif 09707766N | GABRIEL FRIAS            | TORRES    |           | 01/09/1943      | H    | casado      | 961343728     | 696994648      | TFVELASCO@HOTMAIL.COM               | Activo     | 26/02/2014       |                    |            |
| 35        | 35      | nif 09707773L | FRANCISCO BORBA          | TALEGO    |           | 01/10/1942      | NULL | soltero     | 961340666     | 696994649      |                                     | Jubilado   | 25/02/2015       |                    |            |

*Fichero Clientes (elaboración propia)*

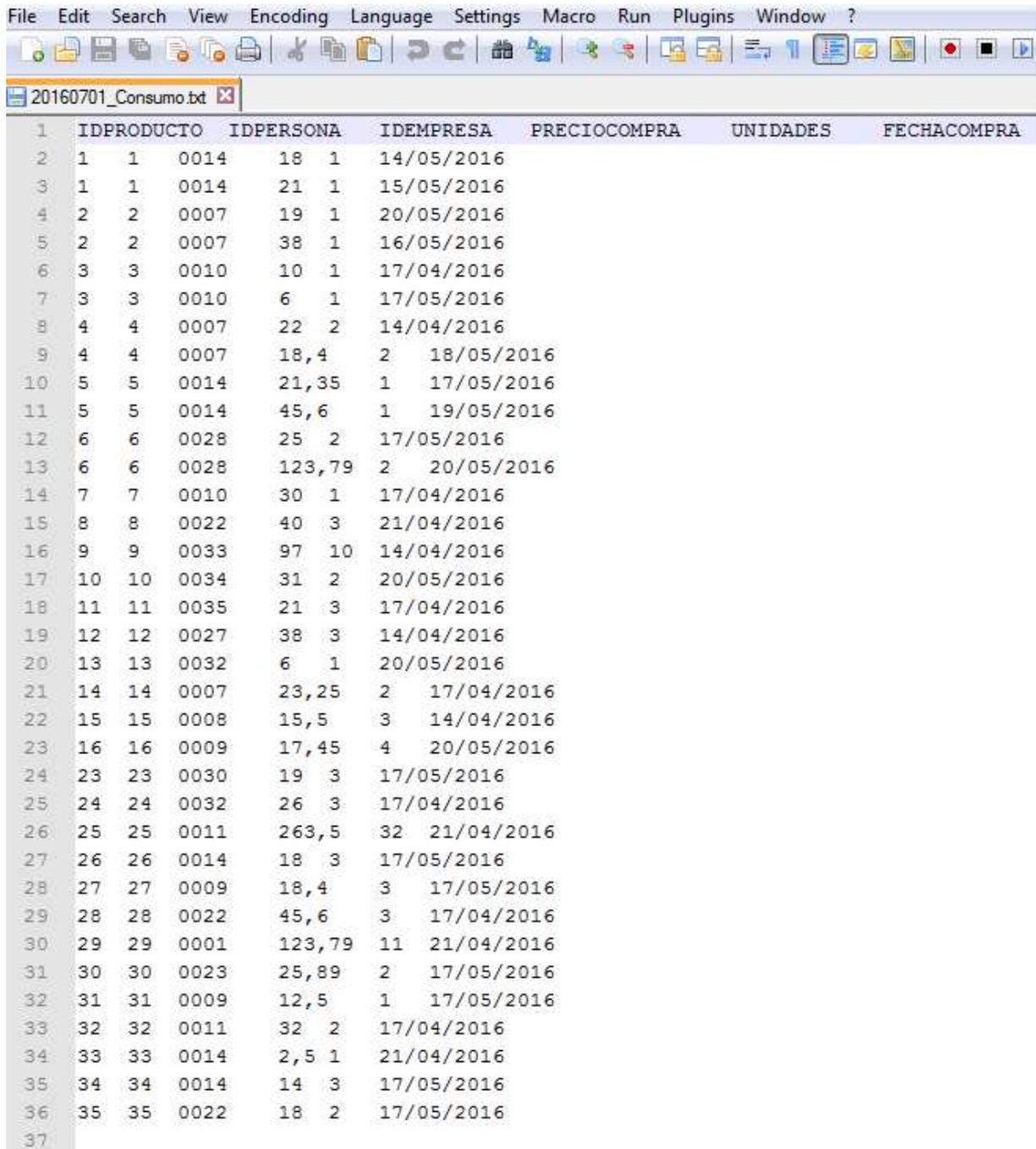
|    | A        | B                  | C                    | D          | E               | F                                 |
|----|----------|--------------------|----------------------|------------|-----------------|-----------------------------------|
| 1  | Nº CAMPO | CAMPO              | FORMATO              | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES                  |
| 2  | 1        | IDPERSONA          | NUMÉRICO             | NO         | 50              |                                   |
| 3  | 2        | TIPONIF            | CADENA DE CARACTERES | NO         | 15              |                                   |
| 4  | 3        | NIF                | ALFANUMERICO         | NO         | 10              |                                   |
| 5  | 4        | NOMBRE             | CADENA DE CARACTERES | NO         | 50              |                                   |
| 6  | 5        | APELLIDO1          | CADENA DE CARACTERES | NO         | 50              |                                   |
| 7  | 6        | APELLIDO2          | CADENA DE CARACTERES | NO         | 50              |                                   |
| 8  | 7        | FECHANACIMIENTO    | FECHA                | SI         | 10              |                                   |
| 9  | 8        | SEXO               | CADENA DE CARACTERES | NO         | 10              |                                   |
| 10 | 9        | ESTADOCIVIL        | CADENA DE CARACTERES | SI         | 10              |                                   |
| 11 | 10       | TELEFONO_FIJO      | NUMÉRICO             | SI         | 9               |                                   |
| 12 | 11       | TELEFONO_MOVIL     | NUMÉRICO             | NO         | 9               |                                   |
| 13 | 12       | FAX                | NUMÉRICO             | SI         | 9               |                                   |
| 14 | 13       | EMAIL              | CADENA DE CARACTERES | SI         | 250             |                                   |
| 15 | 14       | SITUACIONLABORAL   | CADENA DE CARACTERES | NO         | 50              | Activo<br>Desempleado<br>Jubilado |
| 16 | 15       | SITUACION_FAMILIAR | CADENA DE CARACTERES | SI         | 50              | Otros<br>Nucleo Familiar          |
| 17 | 16       | FECHA_ALTA         | FECHA                | NO         | 10              |                                   |
| 18 |          |                    |                      |            |                 |                                   |

*Fichero Clientes (elaboración propia)*

El fichero de clientes debe contener toda la información personal de los clientes, referente a los datos básicos de cada cliente. Además de nosotros vamos a incluir el

IDPERSONA, que será único por cada cliente y que iremos añadiendo de forma correlativa.

Fichero Consumo:



|    | IDPRODUCTO | IDPERSONA | IDEMPRESA | PRECIOCOMPRA | UNIDADES | FECHACOMPRA |
|----|------------|-----------|-----------|--------------|----------|-------------|
| 1  |            |           |           |              |          |             |
| 2  | 1          | 1         | 0014      | 18           | 1        | 14/05/2016  |
| 3  | 1          | 1         | 0014      | 21           | 1        | 15/05/2016  |
| 4  | 2          | 2         | 0007      | 19           | 1        | 20/05/2016  |
| 5  | 2          | 2         | 0007      | 38           | 1        | 16/05/2016  |
| 6  | 3          | 3         | 0010      | 10           | 1        | 17/04/2016  |
| 7  | 3          | 3         | 0010      | 6            | 1        | 17/05/2016  |
| 8  | 4          | 4         | 0007      | 22           | 2        | 14/04/2016  |
| 9  | 4          | 4         | 0007      | 18,4         | 2        | 18/05/2016  |
| 10 | 5          | 5         | 0014      | 21,35        | 1        | 17/05/2016  |
| 11 | 5          | 5         | 0014      | 45,6         | 1        | 19/05/2016  |
| 12 | 6          | 6         | 0028      | 25           | 2        | 17/05/2016  |
| 13 | 6          | 6         | 0028      | 123,79       | 2        | 20/05/2016  |
| 14 | 7          | 7         | 0010      | 30           | 1        | 17/04/2016  |
| 15 | 8          | 8         | 0022      | 40           | 3        | 21/04/2016  |
| 16 | 9          | 9         | 0033      | 97           | 10       | 14/04/2016  |
| 17 | 10         | 10        | 0034      | 31           | 2        | 20/05/2016  |
| 18 | 11         | 11        | 0035      | 21           | 3        | 17/04/2016  |
| 19 | 12         | 12        | 0027      | 38           | 3        | 14/04/2016  |
| 20 | 13         | 13        | 0032      | 6            | 1        | 20/05/2016  |
| 21 | 14         | 14        | 0007      | 23,25        | 2        | 17/04/2016  |
| 22 | 15         | 15        | 0008      | 15,5         | 3        | 14/04/2016  |
| 23 | 16         | 16        | 0009      | 17,45        | 4        | 20/05/2016  |
| 24 | 23         | 23        | 0030      | 19           | 3        | 17/05/2016  |
| 25 | 24         | 24        | 0032      | 26           | 3        | 17/04/2016  |
| 26 | 25         | 25        | 0011      | 263,5        | 32       | 21/04/2016  |
| 27 | 26         | 26        | 0014      | 18           | 3        | 17/05/2016  |
| 28 | 27         | 27        | 0009      | 18,4         | 3        | 17/05/2016  |
| 29 | 28         | 28        | 0022      | 45,6         | 3        | 17/04/2016  |
| 30 | 29         | 29        | 0001      | 123,79       | 11       | 21/04/2016  |
| 31 | 30         | 30        | 0023      | 25,89        | 2        | 17/05/2016  |
| 32 | 31         | 31        | 0009      | 12,5         | 1        | 17/05/2016  |
| 33 | 32         | 32        | 0011      | 32           | 2        | 17/04/2016  |
| 34 | 33         | 33        | 0014      | 2,5          | 1        | 21/04/2016  |
| 35 | 34         | 34        | 0014      | 14           | 3        | 17/05/2016  |
| 36 | 35         | 35        | 0022      | 18           | 2        | 17/05/2016  |
| 37 |            |           |           |              |          |             |

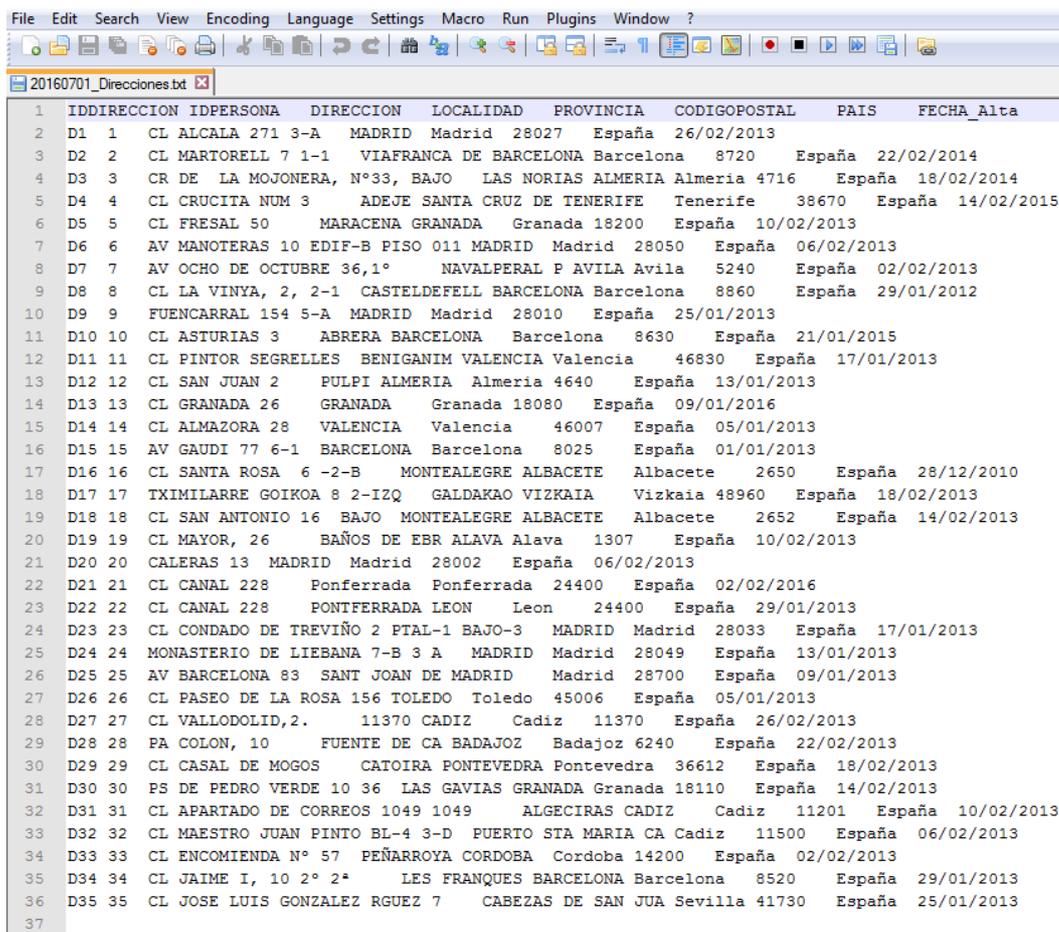
*Fichero Consumo (elaboración propia)*

|   | A        | B            | C        | D          | E               | F                      |
|---|----------|--------------|----------|------------|-----------------|------------------------|
| 1 | Nº CAMPO | CAMPO        | FORMATO  | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES       |
| 2 | 1        | IDPRODUCTO   | NUMÉRICO | NO         | 50              |                        |
| 3 | 2        | IDPERSONA    | NUMÉRICO | NO         | 50              |                        |
| 4 | 3        | IDEMPRESA    | NUMÉRICO | NO         | 50              |                        |
| 5 | 4        | PRECIOCOMPRA | NUMÉRICO | NO         | 4               | Se admitiran decimales |
| 6 | 5        | UNIDADES     | NUMÉRICO | NO         | 3               |                        |
| 7 | 6        | FECHACOMPRA  | FECHA    | NO         | 10              |                        |
| 8 |          |              |          |            |                 |                        |

*Fichero Consumo (elaboración propia)*

El fichero de consumo debe contener la información que hace referencia a los diferentes productos consumidos (carro de la compra) y en este fichero también vamos a incluir el IDPERSONA para tener relacionados los productos y los clientes, según nuestra codificación.

*Fichero Direcciones:*



| ID | IDPERSONA | DIRECCION                         | LOCALIDAD                    | PROVINCIA  | CODIGOPOSTAL | PAIS   | FECHA      | Alta |
|----|-----------|-----------------------------------|------------------------------|------------|--------------|--------|------------|------|
| 1  |           | IDDIRECCION                       |                              |            |              |        |            |      |
| 2  | D1        | 1 CL ALCALA 271 3-A               | MADRID Madrid                | 28027      | España       |        | 26/02/2013 |      |
| 3  | D2        | 2 CL MARTORELL 7 1-1              | VIAFRANCA DE BARCELONA       | Barcelona  | 8720         | España | 22/02/2014 |      |
| 4  | D3        | 3 CR DE LA MOJONERA, N°33, BAJO   | LAS NORIAS ALMERIA           | Almeria    | 4716         | España | 18/02/2014 |      |
| 5  | D4        | 4 CL CRUCITA NUM 3                | ADEJE SANTA CRUZ DE TENERIFE | Tenerife   | 38670        | España | 14/02/2015 |      |
| 6  | D5        | 5 CL FRESAL 50                    | MARACENA GRANADA             | Granada    | 18200        | España | 10/02/2013 |      |
| 7  | D6        | 6 AV MANOTERAS 10 EDIF-B PISO 011 | MADRID Madrid                | 28050      | España       |        | 06/02/2013 |      |
| 8  | D7        | 7 AV OCHO DE OCTUBRE 36,1°        | NAVALPERAL P AVILA           | Avila      | 5240         | España | 02/02/2013 |      |
| 9  | D8        | 8 CL LA VINYA, 2, 2-1             | CASTELDEFELL BARCELONA       | Barcelona  | 8860         | España | 29/01/2012 |      |
| 10 | D9        | 9 FUENCARRAL 154 5-A              | MADRID Madrid                | 28010      | España       |        | 25/01/2013 |      |
| 11 | D10       | 10 CL ASTURIAS 3                  | ABRERA BARCELONA             | Barcelona  | 8630         | España | 21/01/2015 |      |
| 12 | D11       | 11 CL PINTOR SEGRELLES            | BENIGANIM VALENCIA           | Valencia   | 46830        | España | 17/01/2013 |      |
| 13 | D12       | 12 CL SAN JUAN 2                  | PULPI ALMERIA                | Almeria    | 4640         | España | 13/01/2013 |      |
| 14 | D13       | 13 CL GRANADA 26                  | GRANADA                      | Granada    | 18080        | España | 09/01/2016 |      |
| 15 | D14       | 14 CL ALMAZORA 28                 | VALENCIA                     | Valencia   | 46007        | España | 05/01/2013 |      |
| 16 | D15       | 15 AV GAUDI 77 6-1                | BARCELONA                    | Barcelona  | 8025         | España | 01/01/2013 |      |
| 17 | D16       | 16 CL SANTA ROSA 6 -2-B           | MONTEALEGRE ALBACETE         | Albacete   | 2650         | España | 28/12/2010 |      |
| 18 | D17       | 17 TXIMILARRE GOIKOA 8 2-IZQ      | GALDAKAO VIZKAIA             | Vizkaia    | 48960        | España | 18/02/2013 |      |
| 19 | D18       | 18 CL SAN ANTONIO 16              | BAJO MONTEALEGRE ALBACETE    | Albacete   | 2652         | España | 14/02/2013 |      |
| 20 | D19       | 19 CL MAYOR, 26                   | BAÑOS DE EBR ALAVA           | Alava      | 1307         | España | 10/02/2013 |      |
| 21 | D20       | 20 CALERAS 13                     | MADRID Madrid                | 28002      | España       |        | 06/02/2013 |      |
| 22 | D21       | 21 CL CANAL 228                   | Ponferrada                   | Ponferrada | 24400        | España | 02/02/2016 |      |
| 23 | D22       | 22 CL CANAL 228                   | PONTFERRADA LEON             | Leon       | 24400        | España | 29/01/2013 |      |
| 24 | D23       | 23 CL CONDADO DE TREVIÑO 2 PTAL-1 | BAJO-3 MADRID                | Madrid     | 28033        | España | 17/01/2013 |      |
| 25 | D24       | 24 MONASTERIO DE LIEBANA 7-B 3 A  | MADRID Madrid                | 28049      | España       |        | 13/01/2013 |      |
| 26 | D25       | 25 AV BARCELONA 83                | SANT JOAN DE MADRID          | Madrid     | 28700        | España | 09/01/2013 |      |
| 27 | D26       | 26 CL PASEO DE LA ROSA 156        | TOLEDO                       | Toledo     | 45006        | España | 05/01/2013 |      |
| 28 | D27       | 27 CL VALLODOLID,2.               | 11370 CADIZ                  | Cadiz      | 11370        | España | 26/02/2013 |      |
| 29 | D28       | 28 PA COLON, 10                   | FUENTE DE CA BADAJOZ         | Badajoz    | 6240         | España | 22/02/2013 |      |
| 30 | D29       | 29 CL CASAL DE MOGOS              | CATOIRA PONTEVEDRA           | Pontevedra | 36612        | España | 18/02/2013 |      |
| 31 | D30       | 30 PS DE PEDRO VERDE 10 36        | LAS GAVIAS GRANADA           | Granada    | 18110        | España | 14/02/2013 |      |
| 32 | D31       | 31 CL APARTADO DE CORREOS 1049    | 1049 ALGECIRAS CADIZ         | Cadiz      | 11201        | España | 10/02/2013 |      |
| 33 | D32       | 32 CL MAESTRO JUAN PINTO BL-4 3-D | PUERTO STA MARIA CA          | Cadiz      | 11500        | España | 06/02/2013 |      |
| 34 | D33       | 33 CL ENCOMIENDA N° 57            | PEÑARROYA CORDOBA            | Cordoba    | 14200        | España | 02/02/2013 |      |
| 35 | D34       | 34 CL JAIME I, 10 2° 2ª           | LES FRANQUES BARCELONA       | Barcelona  | 8520         | España | 29/01/2013 |      |
| 36 | D35       | 35 CL JOSE LUIS GONZALEZ RGUEZ 7  | CABEZAS DE SAN JUA           | Sevilla    | 41730        | España | 25/01/2013 |      |
| 37 |           |                                   |                              |            |              |        |            |      |

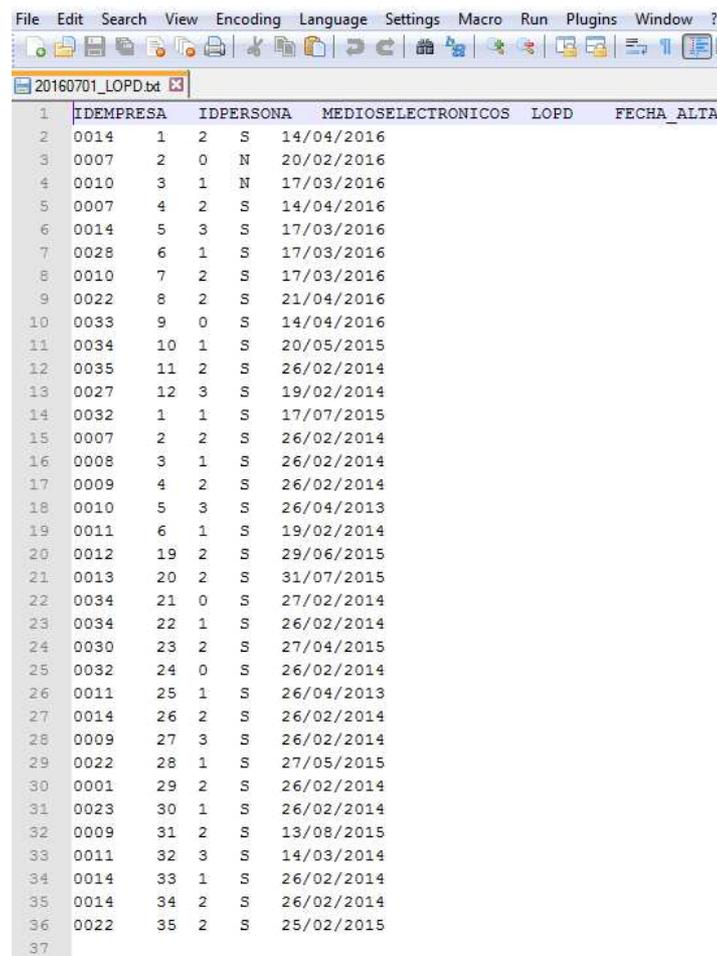
*Fichero Direcciones (elaboración propia)*

|   | A        | B            | C                    | D          | E               | F                |
|---|----------|--------------|----------------------|------------|-----------------|------------------|
| 1 | Nº CAMPO | CAMPO        | FORMATO              | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES |
| 2 | 1        | IDDIRECCION  | NUMÉRICO             | NO         | 50              |                  |
| 3 | 2        | IDPERSONA    | NUMÉRICO             | NO         | 50              |                  |
| 4 | 3        | DIRECCION    | CADENA DE CARACTERES | NO         | 150             |                  |
| 5 | 4        | LOCALIDAD    | CADENA DE CARACTERES | NO         | 50              |                  |
| 6 | 5        | PROVINCIA    | CADENA DE CARACTERES | NO         | 50              |                  |
| 7 | 6        | CODIGOPOSTAL | CADENA DE CARACTERES | NO         | 5               |                  |
| 8 | 7        | PAIS         | CADENA DE CARACTERES | NO         | 50              |                  |
| 9 | 8        | FECHA_Alta   | FECHA                | NO         | 10              |                  |

*Fichero Direcciones (elaboración propia)*

El fichero de direcciones debe contener toda la información que hace referencia a la dirección de cada cliente, también en este fichero vamos a incluir el IDPERSONA para tener la relación entre las direcciones completas y los clientes.

*Fichero LOPD:*



|    | IDEMPRESA | IDPERSONA | MEDIOSELECTRONICOS | LOPD | FECHA_ALTA |
|----|-----------|-----------|--------------------|------|------------|
| 1  |           |           |                    |      |            |
| 2  | 0014      | 1         | 2                  | S    | 14/04/2016 |
| 3  | 0007      | 2         | 0                  | N    | 20/02/2016 |
| 4  | 0010      | 3         | 1                  | N    | 17/03/2016 |
| 5  | 0007      | 4         | 2                  | S    | 14/04/2016 |
| 6  | 0014      | 5         | 3                  | S    | 17/03/2016 |
| 7  | 0028      | 6         | 1                  | S    | 17/03/2016 |
| 8  | 0010      | 7         | 2                  | S    | 17/03/2016 |
| 9  | 0022      | 8         | 2                  | S    | 21/04/2016 |
| 10 | 0033      | 9         | 0                  | S    | 14/04/2016 |
| 11 | 0034      | 10        | 1                  | S    | 20/05/2015 |
| 12 | 0035      | 11        | 2                  | S    | 26/02/2014 |
| 13 | 0027      | 12        | 3                  | S    | 19/02/2014 |
| 14 | 0032      | 1         | 1                  | S    | 17/07/2015 |
| 15 | 0007      | 2         | 2                  | S    | 26/02/2014 |
| 16 | 0008      | 3         | 1                  | S    | 26/02/2014 |
| 17 | 0009      | 4         | 2                  | S    | 26/02/2014 |
| 18 | 0010      | 5         | 3                  | S    | 26/04/2013 |
| 19 | 0011      | 6         | 1                  | S    | 19/02/2014 |
| 20 | 0012      | 19        | 2                  | S    | 29/06/2015 |
| 21 | 0013      | 20        | 2                  | S    | 31/07/2015 |
| 22 | 0034      | 21        | 0                  | S    | 27/02/2014 |
| 23 | 0034      | 22        | 1                  | S    | 26/02/2014 |
| 24 | 0030      | 23        | 2                  | S    | 27/04/2015 |
| 25 | 0032      | 24        | 0                  | S    | 26/02/2014 |
| 26 | 0011      | 25        | 1                  | S    | 26/04/2013 |
| 27 | 0014      | 26        | 2                  | S    | 26/02/2014 |
| 28 | 0009      | 27        | 3                  | S    | 26/02/2014 |
| 29 | 0022      | 28        | 1                  | S    | 27/05/2015 |
| 30 | 0001      | 29        | 2                  | S    | 26/02/2014 |
| 31 | 0023      | 30        | 1                  | S    | 26/02/2014 |
| 32 | 0009      | 31        | 2                  | S    | 13/08/2015 |
| 33 | 0011      | 32        | 3                  | S    | 14/03/2014 |
| 34 | 0014      | 33        | 1                  | S    | 26/02/2014 |
| 35 | 0014      | 34        | 2                  | S    | 26/02/2014 |
| 36 | 0022      | 35        | 2                  | S    | 25/02/2015 |
| 37 |           |           |                    |      |            |

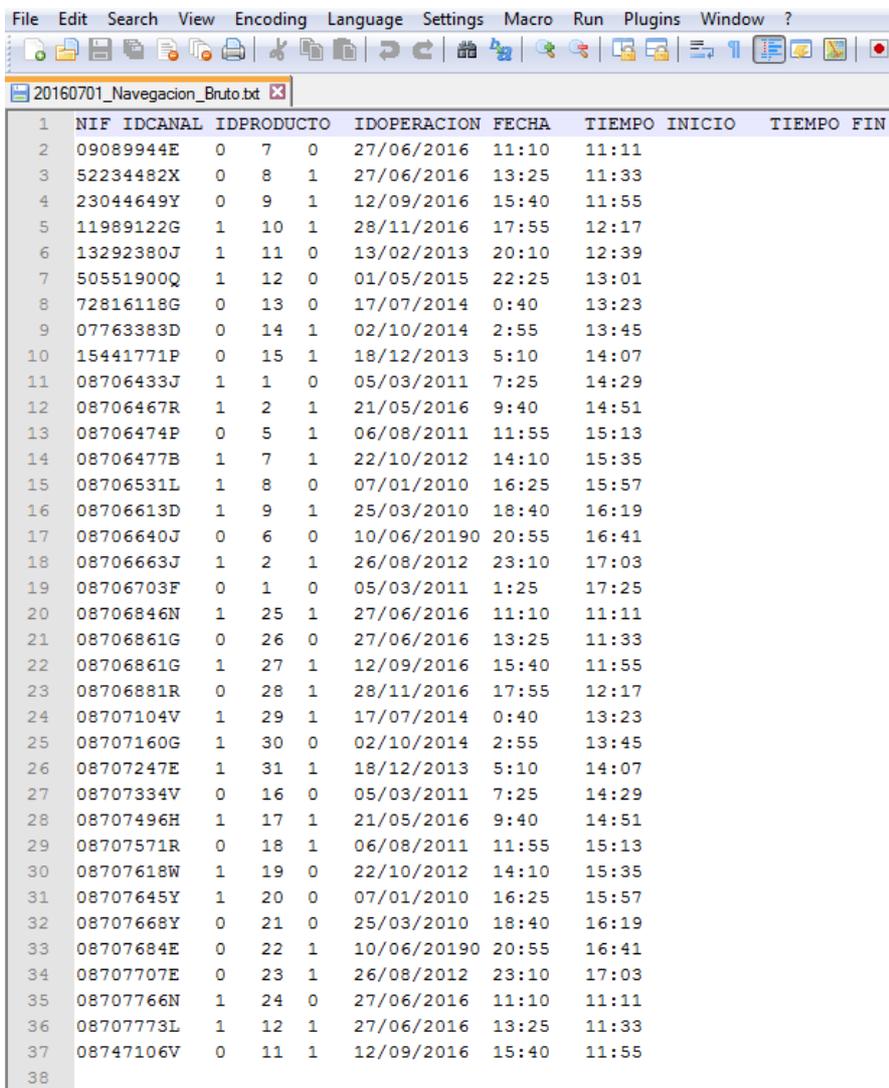
*Fichero LOPD (elaboración propia)*

|   | A        | B                  | C                    | D          | E               | F                              |
|---|----------|--------------------|----------------------|------------|-----------------|--------------------------------|
| 1 | Nº CAMPO | CAMPO              | FORMATO              | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES               |
| 2 | 1        | IDEMPRESA          | NUMÉRICO             | NO         | 50              |                                |
| 3 | 2        | IDPERSONA          | NUMÉRICO             | NO         | 50              |                                |
| 4 | 3        | MEDIOSELECTRONICOS | NUMÉRICO             | NO         |                 | 0-->Si Email<br>1 1-->No Email |
| 5 | 4        | LOPD               | CADENA DE CARACTERES | NO         |                 | S<br>1 N                       |
| 6 | 5        | FECHA_ALTA         | FECHA                | NO         | 10              |                                |

*Fichero LOPD (elaboración propia)*

Para fichero LOPD necesitamos relacionar los clientes con la firma y concesión de datos por lo que incluimos el IDPERSONA.

*Fichero Navegación bruto:*



| 1  | NIF       | IDCANAL | IDPRODUCTO | IDOPERACION | FECHA       | TIEMPO INICIO | TIEMPO FIN |
|----|-----------|---------|------------|-------------|-------------|---------------|------------|
| 2  | 09089944E | 0       | 7          | 0           | 27/06/2016  | 11:10         | 11:11      |
| 3  | 52234482X | 0       | 8          | 1           | 27/06/2016  | 13:25         | 11:33      |
| 4  | 23044649Y | 0       | 9          | 1           | 12/09/2016  | 15:40         | 11:55      |
| 5  | 11989122G | 1       | 10         | 1           | 28/11/2016  | 17:55         | 12:17      |
| 6  | 13292380J | 1       | 11         | 0           | 13/02/2013  | 20:10         | 12:39      |
| 7  | 50551900Q | 1       | 12         | 0           | 01/05/2015  | 22:25         | 13:01      |
| 8  | 72816118G | 0       | 13         | 0           | 17/07/2014  | 0:40          | 13:23      |
| 9  | 07763383D | 0       | 14         | 1           | 02/10/2014  | 2:55          | 13:45      |
| 10 | 15441771P | 0       | 15         | 1           | 18/12/2013  | 5:10          | 14:07      |
| 11 | 08706433J | 1       | 1          | 0           | 05/03/2011  | 7:25          | 14:29      |
| 12 | 08706467R | 1       | 2          | 1           | 21/05/2016  | 9:40          | 14:51      |
| 13 | 08706474P | 0       | 5          | 1           | 06/08/2011  | 11:55         | 15:13      |
| 14 | 08706477B | 1       | 7          | 1           | 22/10/2012  | 14:10         | 15:35      |
| 15 | 08706531L | 1       | 8          | 0           | 07/01/2010  | 16:25         | 15:57      |
| 16 | 08706613D | 1       | 9          | 1           | 25/03/2010  | 18:40         | 16:19      |
| 17 | 08706640J | 0       | 6          | 0           | 10/06/20190 | 20:55         | 16:41      |
| 18 | 08706663J | 1       | 2          | 1           | 26/08/2012  | 23:10         | 17:03      |
| 19 | 08706703F | 0       | 1          | 0           | 05/03/2011  | 1:25          | 17:25      |
| 20 | 08706846N | 1       | 25         | 1           | 27/06/2016  | 11:10         | 11:11      |
| 21 | 08706861G | 0       | 26         | 0           | 27/06/2016  | 13:25         | 11:33      |
| 22 | 08706861G | 1       | 27         | 1           | 12/09/2016  | 15:40         | 11:55      |
| 23 | 08706881R | 0       | 28         | 1           | 28/11/2016  | 17:55         | 12:17      |
| 24 | 08707104V | 1       | 29         | 1           | 17/07/2014  | 0:40          | 13:23      |
| 25 | 08707160G | 1       | 30         | 0           | 02/10/2014  | 2:55          | 13:45      |
| 26 | 08707247E | 1       | 31         | 1           | 18/12/2013  | 5:10          | 14:07      |
| 27 | 08707334V | 0       | 16         | 0           | 05/03/2011  | 7:25          | 14:29      |
| 28 | 08707496H | 1       | 17         | 1           | 21/05/2016  | 9:40          | 14:51      |
| 29 | 08707571R | 0       | 18         | 1           | 06/08/2011  | 11:55         | 15:13      |
| 30 | 08707618W | 1       | 19         | 0           | 22/10/2012  | 14:10         | 15:35      |
| 31 | 08707645Y | 1       | 20         | 0           | 07/01/2010  | 16:25         | 15:57      |
| 32 | 08707668Y | 0       | 21         | 0           | 25/03/2010  | 18:40         | 16:19      |
| 33 | 08707684E | 0       | 22         | 1           | 10/06/20190 | 20:55         | 16:41      |
| 34 | 08707707E | 0       | 23         | 1           | 26/08/2012  | 23:10         | 17:03      |
| 35 | 08707766N | 1       | 24         | 0           | 27/06/2016  | 11:10         | 11:11      |
| 36 | 08707773L | 1       | 12         | 1           | 27/06/2016  | 13:25         | 11:33      |
| 37 | 08747106V | 0       | 11         | 1           | 12/09/2016  | 15:40         | 11:55      |
| 38 |           |         |            |             |             |               |            |

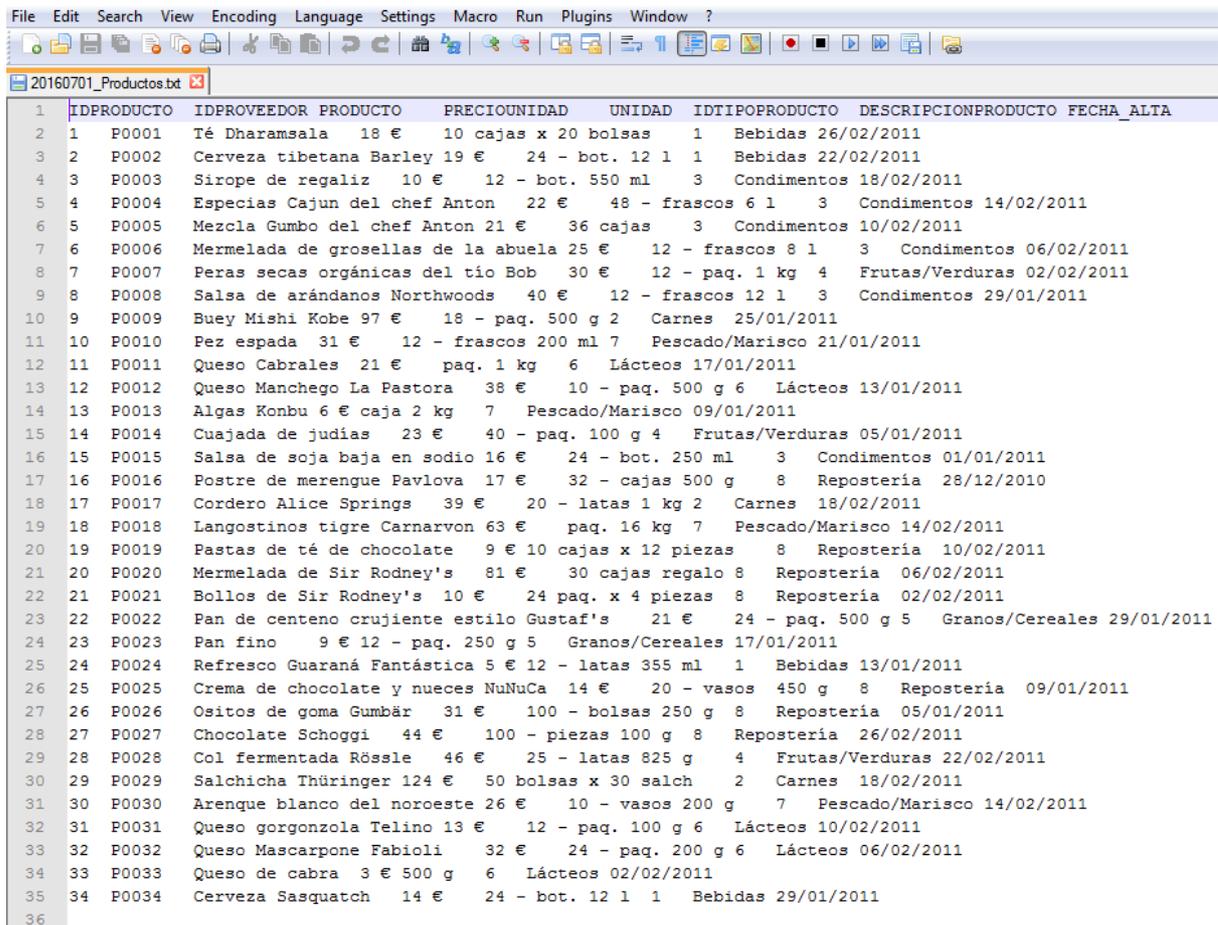
*Fichero Navegación bruto (elaboración propia)*

|   | A        | B             | C            | D          | E               | F                |
|---|----------|---------------|--------------|------------|-----------------|------------------|
| 1 | Nº CAMPO | CAMPO         | FORMATO      | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES |
| 2 | 1        | NIF           | ALFANUMERICO | NO         | 10              |                  |
| 3 | 2        | IDCANAL       | NUMÉRICO     | NO         | 50              |                  |
| 4 | 3        | IDPRODUCTO    | NUMÉRICO     | NO         | 50              |                  |
| 5 | 4        | IDOPERACION   | NUMÉRICO     | NO         | 50              |                  |
| 6 | 5        | FECHA         | FECHA        | NO         | 10              |                  |
| 7 | 6        | TIEMPO INICIO | HORA         | NO         | 50              |                  |
| 8 | 7        | TIEMPO FIN    | HORA         | NO         | 50              |                  |

*Fichero Navegación bruto (elaboración propia)*

La navegación en bruto, es la que nos proporciona la empresa, y necesitamos que contenga determinados datos recogidos de las cookies de navegación, con lo que posteriormente poder relacionar aquellos datos de navegación por la página web con el listado de clientes.

*Fichero Productos:*



|    | IDPRODUCTO | IDPROVEEDOR                              | PRODUCTO | PRECIOUNIDAD         | UNIDAD | IDTIPOPRODUCTO  | DESCRIPCIONPRODUCTO | FECHA_ALTA |
|----|------------|--|----------|----------------------|--------|-----------------|---------------------|------------|
| 1  | P0001      | Té Dharamsala                            | 18 €     | 10 cajas x 20 bolsas | 1      | Bebidas         | 26/02/2011          |            |
| 2  | P0002      | Cerveza tibetana Barley                  | 19 €     | 24 - bot. 12 l       | 1      | Bebidas         | 22/02/2011          |            |
| 3  | P0003      | Sirope de regaliz                        | 10 €     | 12 - bot. 550 ml     | 3      | Condimentos     | 18/02/2011          |            |
| 4  | P0004      | Espicias Cajun del chef Anton            | 22 €     | 48 - frascos 6 l     | 3      | Condimentos     | 14/02/2011          |            |
| 5  | P0005      | Mezcla Gumbo del chef Anton              | 21 €     | 36 cajas             | 3      | Condimentos     | 10/02/2011          |            |
| 6  | P0006      | Mermelada de grosellas de la abuela      | 25 €     | 12 - frascos 8 l     | 3      | Condimentos     | 06/02/2011          |            |
| 7  | P0007      | Peras secas orgánicas del tío Bob        | 30 €     | 12 - paq. 1 kg       | 4      | Frutas/Verduras | 02/02/2011          |            |
| 8  | P0008      | Salsa de arándanos Northwoods            | 40 €     | 12 - frascos 12 l    | 3      | Condimentos     | 29/01/2011          |            |
| 9  | P0009      | Buey Mishi Kobe                          | 97 €     | 18 - paq. 500 g      | 2      | Carnes          | 25/01/2011          |            |
| 10 | P0010      | Pez espada                               | 31 €     | 12 - frascos 200 ml  | 7      | Pescado/Marisco | 21/01/2011          |            |
| 11 | P0011      | Queso Cabrales                           | 21 €     | paq. 1 kg            | 6      | Lácteos         | 17/01/2011          |            |
| 12 | P0012      | Queso Manchego La Pastora                | 38 €     | 10 - paq. 500 g      | 6      | Lácteos         | 13/01/2011          |            |
| 13 | P0013      | Algas Konbu                              | 6 €      | caja 2 kg            | 7      | Pescado/Marisco | 09/01/2011          |            |
| 14 | P0014      | Cuajada de judías                        | 23 €     | 40 - paq. 100 g      | 4      | Frutas/Verduras | 05/01/2011          |            |
| 15 | P0015      | Salsa de soja baja en sodio              | 16 €     | 24 - bot. 250 ml     | 3      | Condimentos     | 01/01/2011          |            |
| 16 | P0016      | Postre de merengue Pavlova               | 17 €     | 32 - cajas 500 g     | 8      | Repostería      | 28/12/2010          |            |
| 17 | P0017      | Cordero Alice Springs                    | 39 €     | 20 - latas 1 kg      | 2      | Carnes          | 18/02/2011          |            |
| 18 | P0018      | Langostinos tigre Carnarvon              | 63 €     | paq. 16 kg           | 7      | Pescado/Marisco | 14/02/2011          |            |
| 19 | P0019      | Pastas de té de chocolate                | 9 €      | 10 cajas x 12 piezas | 8      | Repostería      | 10/02/2011          |            |
| 20 | P0020      | Mermelada de Sir Rodney's                | 81 €     | 30 cajas regalo      | 8      | Repostería      | 06/02/2011          |            |
| 21 | P0021      | Bollos de Sir Rodney's                   | 10 €     | 24 paq. x 4 piezas   | 8      | Repostería      | 02/02/2011          |            |
| 22 | P0022      | Pan de centeno crujiente estilo Gustaf's | 21 €     | 24 - paq. 500 g      | 5      | Granos/Cereales | 29/01/2011          |            |
| 23 | P0023      | Pan fino                                 | 9 €      | 12 - paq. 250 g      | 5      | Granos/Cereales | 17/01/2011          |            |
| 24 | P0024      | Refresco Guaraná Fantástica              | 5 €      | 12 - latas 355 ml    | 1      | Bebidas         | 13/01/2011          |            |
| 25 | P0025      | Crema de chocolate y nueces NuNuCa       | 14 €     | 20 - vasos 450 g     | 8      | Repostería      | 09/01/2011          |            |
| 26 | P0026      | Ositos de goma Gumbär                    | 31 €     | 100 - bolsas 250 g   | 8      | Repostería      | 05/01/2011          |            |
| 27 | P0027      | Chocolate Schoggi                        | 44 €     | 100 - piezas 100 g   | 8      | Repostería      | 26/02/2011          |            |
| 28 | P0028      | Col fermentada Rössle                    | 46 €     | 25 - latas 825 g     | 4      | Frutas/Verduras | 22/02/2011          |            |
| 29 | P0029      | Salchicha Thüringer                      | 124 €    | 50 bolsas x 30 salch | 2      | Carnes          | 18/02/2011          |            |
| 30 | P0030      | Arenque blanco del noroeste              | 26 €     | 10 - vasos 200 g     | 7      | Pescado/Marisco | 14/02/2011          |            |
| 31 | P0031      | Queso gorgonzola Telino                  | 13 €     | 12 - paq. 100 g      | 6      | Lácteos         | 10/02/2011          |            |
| 32 | P0032      | Queso Mascarpone Fabioli                 | 32 €     | 24 - paq. 200 g      | 6      | Lácteos         | 06/02/2011          |            |
| 33 | P0033      | Queso de cabra                           | 3 €      | 500 g                | 6      | Lácteos         | 02/02/2011          |            |
| 34 | P0034      | Cerveza Sasquatch                        | 14 €     | 24 - bot. 12 l       | 1      | Bebidas         | 29/01/2011          |            |

*Fichero Productos (elaboración propia)*

|   | A        | B                   | C            | D          | E               | F                      |
|---|----------|---------------------|--------------|------------|-----------------|------------------------|
| 1 | Nº CAMPO | CAMPO               | FORMATO      | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES       |
| 2 | 1        | IDPRODUCTO          | NUMÉRICO     | NO         | 50              |                        |
| 3 | 2        | IDPROVEEDOR         | NUMÉRICO     | NO         | 50              |                        |
| 4 | 3        | PRODUCTO            | ALFANUMERICO | NO         | 70              |                        |
| 5 | 4        | PRECIOUNIDAD        | NUMÉRICO     | NO         | 4               | Se admitiran decimales |
| 6 | 5        | UNIDAD              | NUMÉRICO     | NO         | 3               |                        |
| 7 | 6        | IDTIPOPRODUCTO      | NUMÉRICO     | NO         | 50              |                        |
| 8 | 7        | DESCRIPCIONPRODUCTO | ALFANUMERICO | NO         | 70              |                        |
| 9 | 8        | FECHA_ALTA          | FECHA        | NO         | 10              |                        |

*Fichero Productos (elaboración propia)*

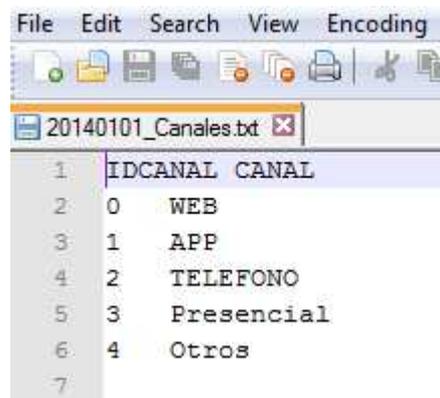
El fichero con el listado de productos debe contener todo lo relacionado a todas las referencias de productos de que se dispongan en los establecimientos, así como una breve descripción de los mismos.

Además de los ficheros que nos proporciona la empresa necesitamos crearnos nuestros propios ficheros, la mayoría sólo se cargarán una vez en el sistema y serán fijos a lo largo del tiempo y otros además de cargarlos desde un principio también se irán alimentando de datos históricos.

| Nombre                         | Fecha de modifica... | Tipo              | Tamaño |
|--------------------------------|----------------------|-------------------|--------|
| 20160701_Canales.txt           | 30/06/2016 18:58     | Notepad++ Docu... | 1 KB   |
| 20160701_Empresas.txt          | 30/06/2016 18:49     | Notepad++ Docu... | 2 KB   |
| 20160701_Navegacion_Creada.txt | 30/06/2016 18:57     | Notepad++ Docu... | 2 KB   |
| 20160701_Operaciones.txt       | 30/06/2016 18:58     | Notepad++ Docu... | 1 KB   |
| 20160701_Persona_Campañas.txt  | 30/06/2016 18:54     | Notepad++ Docu... | 1 KB   |
| 20160701_Persona_Empresa.txt   | 30/06/2016 18:50     | Notepad++ Docu... | 1 KB   |
| 20160701_Proveedores.txt       | 30/06/2016 18:53     | Notepad++ Docu... | 1 KB   |
| 20160701_Tipo_Producto.txt     | 30/06/2016 18:52     | Notepad++ Docu... | 1 KB   |

*Ficheros Creados (elaboración propia)*

**Fichero Canales:**



```

1 IDCANAL CANAL
2 0 WEB
3 1 APP
4 2 TELEFONO
5 3 Presencial
6 4 Otros
7

```

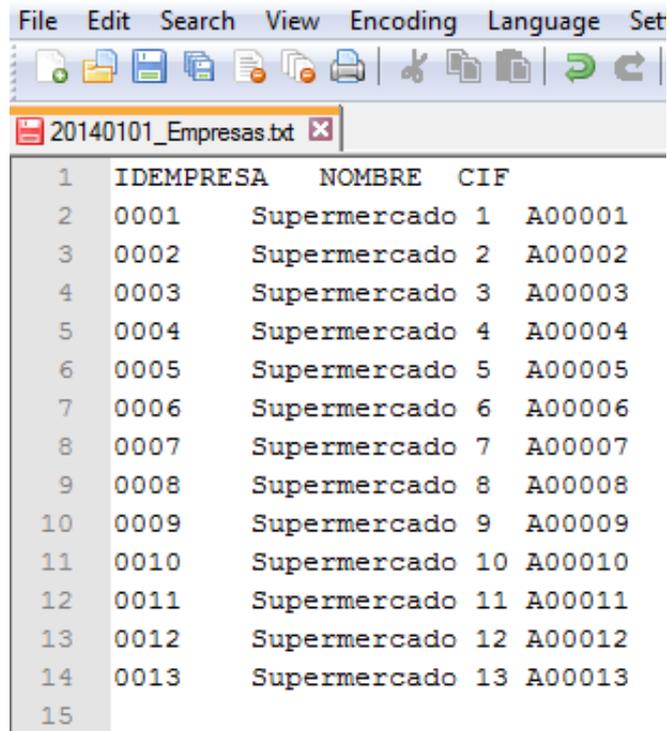
*Fichero Canales (elaboración propia)*

|   | A        | B       | C                    | D          | E               | F   |
|---|----------|---------|----------------------|------------|-----------------|---|
| 1 | Nº CAMPO | CAMPO   | FORMATO              | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES  |
| 2 | 1        | IDCANAL | NUMÉRICO             | NO         | 50              | 0-->WEB<br>1-->APP<br>2-->TELEFONO<br>3-->Presencial<br>4-->Otros |
| 3 | 2        | CANAL   | CADENA DE CARACTERES | NO         | 150             |   |

*Fichero Canales (elaboración propia)*

El fichero que agrupa los canales se carga al principio y salvo que se agregará algún otro canal este será de esos ficheros fijos que no cambian todos los meses.

Fichero Empresas:



|    | IDEMPRESA | NOMBRE          | CIF    |
|----|-----------|-----------------|--------|
| 1  |           |                 |        |
| 2  | 0001      | Supermercado 1  | A00001 |
| 3  | 0002      | Supermercado 2  | A00002 |
| 4  | 0003      | Supermercado 3  | A00003 |
| 5  | 0004      | Supermercado 4  | A00004 |
| 6  | 0005      | Supermercado 5  | A00005 |
| 7  | 0006      | Supermercado 6  | A00006 |
| 8  | 0007      | Supermercado 7  | A00007 |
| 9  | 0008      | Supermercado 8  | A00008 |
| 10 | 0009      | Supermercado 9  | A00009 |
| 11 | 0010      | Supermercado 10 | A00010 |
| 12 | 0011      | Supermercado 11 | A00011 |
| 13 | 0012      | Supermercado 12 | A00012 |
| 14 | 0013      | Supermercado 13 | A00013 |
| 15 |           |                 |        |

*Fichero Empresas (elaboración propia)*

|   | A        | B         | C                    | D          | E               | F                |
|---|----------|-----------|----------------------|------------|-----------------|------------------|
| 1 | Nº CAMPO | CAMPO     | FORMATO              | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES |
| 2 | 1        | IDEMPRESA | NUMÉRICO             | NO         | 50              |                  |
| 3 | 2        | NOMBRE    | CADENA DE CARACTERES | NO         | 150             |                  |
| 4 | 3        | CIF       | ALFANUMERICO         | NO         | 10              |                  |

*Fichero Empresas (elaboración propia)*

El fichero de las empresas reflejara todos los supermercados que están incluidas en la matriz, con los indicadores para diferenciarlos.

Fichero Navegación Creada:

|    | IDNAVEGACION | IDPERSONA | NIF | IDCANAL | IDPRODUCTO | IDOPERACION | FECHA | TIEMPO INICIO | TIEMPO FIN |
|----|--------------|-----------|-----|---------|------------|-------------|-------|---------------|------------|
| 1  | N0001        | 09089944E | 0   | 7       | 0          | 27/06/2016  | 11:10 | 11:11         |            |
| 2  | N0002        | 52234482X | 0   | 8       | 1          | 27/06/2016  | 13:25 | 11:33         |            |
| 3  | N0003        | 23044649Y | 0   | 9       | 1          | 12/09/2016  | 15:40 | 11:55         |            |
| 4  | N0004        | 11989122G | 1   | 10      | 1          | 28/11/2016  | 17:55 | 12:17         |            |
| 5  | N0005        | 13292380J | 1   | 11      | 0          | 13/02/2013  | 20:10 | 12:39         |            |
| 6  | N0006        | 50551900Q | 1   | 12      | 0          | 01/05/2015  | 22:25 | 13:01         |            |
| 7  | N0007        | 72816118G | 0   | 13      | 0          | 17/07/2014  | 0:40  | 13:23         |            |
| 8  | N0008        | 07763383D | 0   | 14      | 1          | 02/10/2014  | 2:55  | 13:45         |            |
| 9  | N0009        | 15441771P | 0   | 15      | 1          | 18/12/2013  | 5:10  | 14:07         |            |
| 10 | N0010        | 08706433J | 1   | 1       | 0          | 05/03/2011  | 7:25  | 14:29         |            |
| 11 | N0011        | 08706467R | 1   | 2       | 1          | 21/05/2016  | 9:40  | 14:51         |            |
| 12 | N0012        | 08706474P | 0   | 5       | 1          | 06/08/2011  | 11:55 | 15:13         |            |
| 13 | N0013        | 08706477B | 1   | 7       | 1          | 22/10/2012  | 14:10 | 15:35         |            |
| 14 | N0014        | 08706531L | 1   | 8       | 0          | 07/01/2010  | 16:25 | 15:57         |            |
| 15 | N0015        | 08706613D | 1   | 9       | 1          | 25/03/2010  | 18:40 | 16:19         |            |
| 16 | N0016        | 08706640J | 0   | 6       | 0          | 10/06/2019  | 20:55 | 16:41         |            |
| 17 | N0017        | 08706663J | 1   | 2       | 1          | 26/08/2012  | 23:10 | 17:03         |            |
| 18 | N0018        | 08706703F | 0   | 1       | 0          | 05/03/2011  | 1:25  | 17:25         |            |
| 19 | N0019        | 08706846N | 1   | 25      | 1          | 27/06/2016  | 11:10 | 11:11         |            |
| 20 | N0020        | 08706861G | 0   | 26      | 0          | 27/06/2016  | 13:25 | 11:33         |            |
| 21 | N0021        | 08706861G | 1   | 27      | 1          | 12/09/2016  | 15:40 | 11:55         |            |
| 22 | N0022        | 08706881R | 0   | 28      | 1          | 28/11/2016  | 17:55 | 12:17         |            |
| 23 | N0023        | 08707104V | 1   | 29      | 1          | 17/07/2014  | 0:40  | 13:23         |            |
| 24 | N0024        | 08707160G | 1   | 30      | 0          | 02/10/2014  | 2:55  | 13:45         |            |
| 25 | N0025        | 08707247E | 1   | 31      | 1          | 18/12/2013  | 5:10  | 14:07         |            |
| 26 | N0026        | 08707334V | 0   | 16      | 0          | 05/03/2011  | 7:25  | 14:29         |            |
| 27 | N0027        | 08707496H | 1   | 17      | 1          | 21/05/2016  | 9:40  | 14:51         |            |
| 28 | N0028        | 08707571R | 0   | 18      | 1          | 06/08/2011  | 11:55 | 15:13         |            |
| 29 | N0029        | 08707618W | 1   | 19      | 0          | 22/10/2012  | 14:10 | 15:35         |            |
| 30 | N0030        | 08707645Y | 1   | 20      | 0          | 07/01/2010  | 16:25 | 15:57         |            |
| 31 | N0031        | 08707668Y | 0   | 21      | 0          | 25/03/2010  | 18:40 | 16:19         |            |
| 32 | N0032        | 08707684E | 0   | 22      | 1          | 10/06/2019  | 20:55 | 16:41         |            |
| 33 | N0033        | 08707707E | 0   | 23      | 1          | 26/08/2012  | 23:10 | 17:03         |            |
| 34 | N0034        | 08707766N | 1   | 24      | 0          | 27/06/2016  | 11:10 | 11:11         |            |
| 35 | N0035        | 08707773L | 1   | 12      | 1          | 27/06/2016  | 13:25 | 11:33         |            |
| 36 |              |           |     |         |            |             |       |               |            |
| 37 |              |           |     |         |            |             |       |               |            |

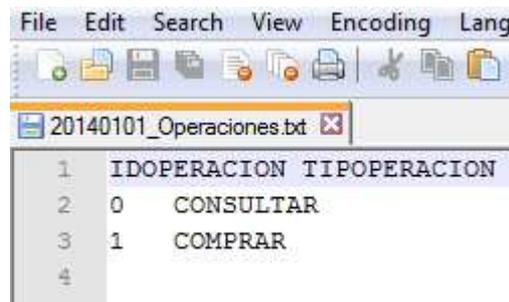
Fichero Navegación creada (elaboración propia)

|    | A        | B             | C            | D          | E               | F                |
|----|----------|---------------|--------------|------------|-----------------|------------------|
| 1  | Nº CAMPO | CAMPO         | FORMATO      | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES |
| 2  | 1        | IDNAVEGACION  | ALFANUMERICO | NO         | 10              |                  |
| 3  | 2        | IDPERSONA     | NUMÉRICO     | NO         | 50              |                  |
| 4  | 3        | NIF           | ALFANUMERICO | NO         | 10              |                  |
| 5  | 4        | IDCANAL       | NUMÉRICO     | NO         | 50              |                  |
| 6  | 5        | IDPRODUCTO    | NUMÉRICO     | NO         | 50              |                  |
| 7  | 6        | IDOPERACION   | NUMÉRICO     | NO         | 50              |                  |
| 8  | 7        | FECHA         | FECHA        | NO         | 10              |                  |
| 9  | 8        | TIEMPO INICIO | HORA         | NO         | 50              |                  |
| 10 | 9        | TIEMPO FIN    | HORA         | NO         | 50              |                  |

Fichero Navegación creada (elaboración propia)

La navegación creada contiene todos los datos del fichero navegación en bruto más dos campos creados, IDNAVEGACION y el IDPERSONA (el cual hemos utilizado en varios ficheros anteriores).

Fichero Operaciones:



```

1 IDOPERACION TIPOOPERACION
2 0 CONSULTAR
3 1 COMPRAR
4

```

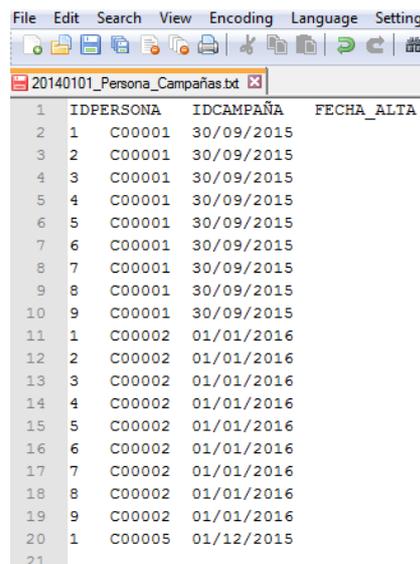
*Fichero Operaciones (elaboración propia)*

|   | A        | B               | C                    | D          | E               | F                |
|---|----------|-----------------|----------------------|------------|-----------------|------------------|
| 1 | Nº CAMPO | CAMPO           | FORMATO              | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES |
| 2 |          | 1 IDOPERACION   | NUMÉRICO             | NO         | 50              | 0 --> Consultar  |
| 3 |          | 2 TIPOOPERACION | CADENA DE CARACTERES | NO         | 150             | 1 --> Comprar    |

*Fichero Operaciones (elaboración propia)*

Otro de los ficheros que serán fijos desde el principio y que no se modificaran, salvo que hubiera que añadir algún tipo de operación, es el fichero de las operaciones.

Fichero Persona Campañas:



```

1 IDPERSONA IDCAMPAÑA FECHA_ALTA
2 1 C00001 30/09/2015
3 2 C00001 30/09/2015
4 3 C00001 30/09/2015
5 4 C00001 30/09/2015
6 5 C00001 30/09/2015
7 6 C00001 30/09/2015
8 7 C00001 30/09/2015
9 8 C00001 30/09/2015
10 9 C00001 30/09/2015
11 1 C00002 01/01/2016
12 2 C00002 01/01/2016
13 3 C00002 01/01/2016
14 4 C00002 01/01/2016
15 5 C00002 01/01/2016
16 6 C00002 01/01/2016
17 7 C00002 01/01/2016
18 8 C00002 01/01/2016
19 9 C00002 01/01/2016
20 1 C00005 01/12/2015
21

```

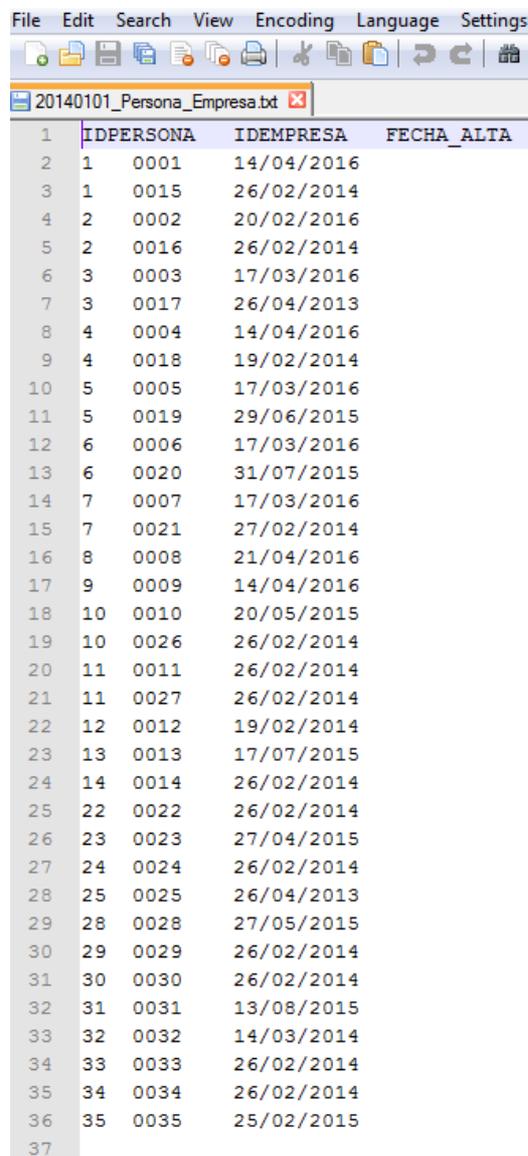
*Fichero Persona campañas (elaboración propia)*

|   | A        | B          | C        | D          | E               | F                |
|---|----------|------------|----------|------------|-----------------|------------------|
| 1 | Nº CAMPO | CAMPO      | FORMATO  | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES |
| 2 | 1        | IDPERSONA  | NUMÉRICO | NO         | 50              |                  |
| 3 | 2        | IDCAMPAÑA  | NUMÉRICO | NO         | 50              |                  |
| 4 | 3        | FECHA_ALTA | FECHA    | NO         | 10              |                  |

*Fichero Persona campañas (elaboración propia)*

El fichero de persona campañas contendrá todos los clientes y sobre que campañas han sido objeto, como la fecha en que se dio de alta dicha campaña.

*Fichero Persona Empresa:*



|    | IDPERSONA | IDEMPRESA | FECHA_ALTA |
|----|-----------|-----------|------------|
| 1  |           |           |            |
| 2  | 1         | 0001      | 14/04/2016 |
| 3  | 1         | 0015      | 26/02/2014 |
| 4  | 2         | 0002      | 20/02/2016 |
| 5  | 2         | 0016      | 26/02/2014 |
| 6  | 3         | 0003      | 17/03/2016 |
| 7  | 3         | 0017      | 26/04/2013 |
| 8  | 4         | 0004      | 14/04/2016 |
| 9  | 4         | 0018      | 19/02/2014 |
| 10 | 5         | 0005      | 17/03/2016 |
| 11 | 5         | 0019      | 29/06/2015 |
| 12 | 6         | 0006      | 17/03/2016 |
| 13 | 6         | 0020      | 31/07/2015 |
| 14 | 7         | 0007      | 17/03/2016 |
| 15 | 7         | 0021      | 27/02/2014 |
| 16 | 8         | 0008      | 21/04/2016 |
| 17 | 9         | 0009      | 14/04/2016 |
| 18 | 10        | 0010      | 20/05/2015 |
| 19 | 10        | 0026      | 26/02/2014 |
| 20 | 11        | 0011      | 26/02/2014 |
| 21 | 11        | 0027      | 26/02/2014 |
| 22 | 12        | 0012      | 19/02/2014 |
| 23 | 13        | 0013      | 17/07/2015 |
| 24 | 14        | 0014      | 26/02/2014 |
| 25 | 22        | 0022      | 26/02/2014 |
| 26 | 23        | 0023      | 27/04/2015 |
| 27 | 24        | 0024      | 26/02/2014 |
| 28 | 25        | 0025      | 26/04/2013 |
| 29 | 28        | 0028      | 27/05/2015 |
| 30 | 29        | 0029      | 26/02/2014 |
| 31 | 30        | 0030      | 26/02/2014 |
| 32 | 31        | 0031      | 13/08/2015 |
| 33 | 32        | 0032      | 14/03/2014 |
| 34 | 33        | 0033      | 26/02/2014 |
| 35 | 34        | 0034      | 26/02/2014 |
| 36 | 35        | 0035      | 25/02/2015 |
| 37 |           |           |            |

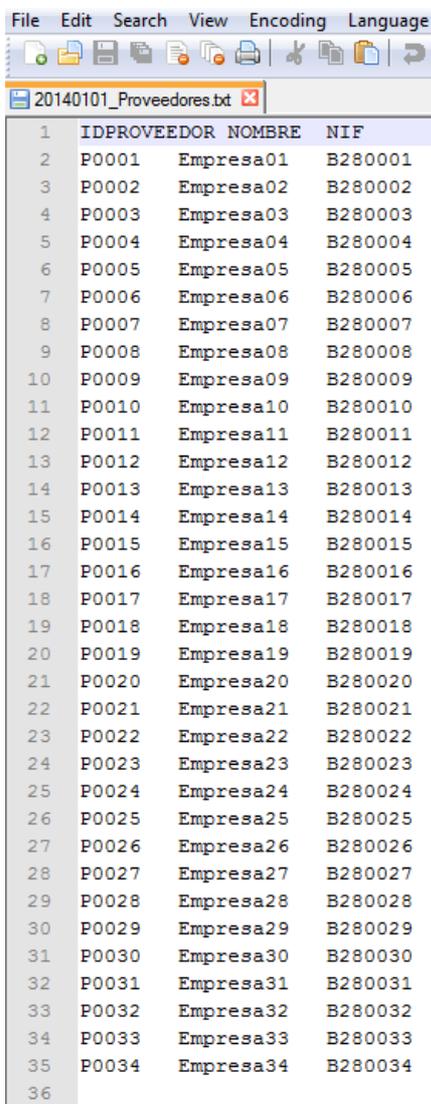
*Fichero Persona empresa (elaboración propia)*

|   | A        | B          | C        | D          | E               | F                |
|---|----------|------------|----------|------------|-----------------|------------------|
| 1 | Nº CAMPO | CAMPO      | FORMATO  | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES |
| 2 | 1        | IDPERSONA  | NUMÉRICO | NO         | 50              |                  |
| 3 | 2        | IDEMPRESA  | NUMÉRICO | NO         | 50              |                  |
| 4 | 3        | FECHA_ALTA | FECHA    | NO         | 10              |                  |

*Fichero Persona empresa (elaboración propia)*

El fichero de persona empresa contendrá todos los clientes y sobre que empresas (supermercados) se registró el alta.

*Fichero Proveedores:*



|    | IDPROVEEDOR | NOMBRE    | NIF     |
|----|-------------|-----------|---------|
| 2  | P0001       | Empresa01 | B280001 |
| 3  | P0002       | Empresa02 | B280002 |
| 4  | P0003       | Empresa03 | B280003 |
| 5  | P0004       | Empresa04 | B280004 |
| 6  | P0005       | Empresa05 | B280005 |
| 7  | P0006       | Empresa06 | B280006 |
| 8  | P0007       | Empresa07 | B280007 |
| 9  | P0008       | Empresa08 | B280008 |
| 10 | P0009       | Empresa09 | B280009 |
| 11 | P0010       | Empresa10 | B280010 |
| 12 | P0011       | Empresa11 | B280011 |
| 13 | P0012       | Empresa12 | B280012 |
| 14 | P0013       | Empresa13 | B280013 |
| 15 | P0014       | Empresa14 | B280014 |
| 16 | P0015       | Empresa15 | B280015 |
| 17 | P0016       | Empresa16 | B280016 |
| 18 | P0017       | Empresa17 | B280017 |
| 19 | P0018       | Empresa18 | B280018 |
| 20 | P0019       | Empresa19 | B280019 |
| 21 | P0020       | Empresa20 | B280020 |
| 22 | P0021       | Empresa21 | B280021 |
| 23 | P0022       | Empresa22 | B280022 |
| 24 | P0023       | Empresa23 | B280023 |
| 25 | P0024       | Empresa24 | B280024 |
| 26 | P0025       | Empresa25 | B280025 |
| 27 | P0026       | Empresa26 | B280026 |
| 28 | P0027       | Empresa27 | B280027 |
| 29 | P0028       | Empresa28 | B280028 |
| 30 | P0029       | Empresa29 | B280029 |
| 31 | P0030       | Empresa30 | B280030 |
| 32 | P0031       | Empresa31 | B280031 |
| 33 | P0032       | Empresa32 | B280032 |
| 34 | P0033       | Empresa33 | B280033 |
| 35 | P0034       | Empresa34 | B280034 |
| 36 |             |           |         |

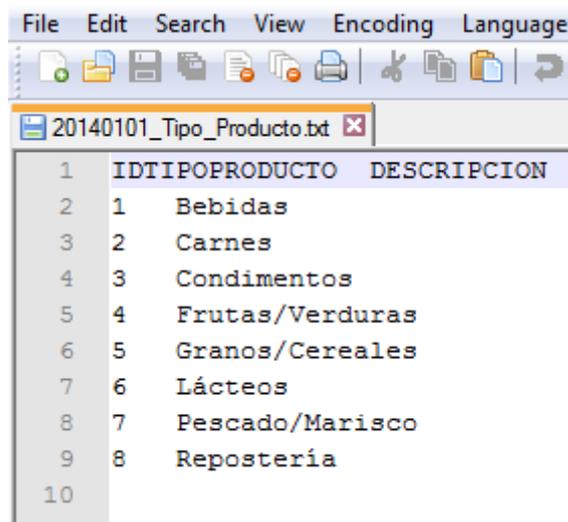
*Fichero Proveedores (elaboración propia)*

|   | A        | B           | C                    | D          | E               | F                |
|---|----------|-------------|----------------------|------------|-----------------|------------------|
| 1 | Nº CAMPO | CAMPO       | FORMATO              | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES |
| 2 | 1        | IDPROVEEDOR | NUMÉRICO             | NO         | 50              |                  |
| 3 | 2        | NOMBRE      | CADENA DE CARACTERES | NO         | 150             |                  |
| 4 | 3        | NIF         | ALFANUMERICO         | NO         | 10              |                  |

*Fichero Proveedores (elaboración propia)*

El fichero con los proveedores contendrá un IDPROVEEDOR junto con su nombre comercial y su NIF para poder relacionarlos.

*Fichero Tipo Producto:*



|    | A              | B               | C | D | E | F |
|----|----------------|-----------------|---|---|---|---|
| 1  | IDTIPOPRODUCTO | DESCRIPCION     |   |   |   |   |
| 2  | 1              | Bebidas         |   |   |   |   |
| 3  | 2              | Carnes          |   |   |   |   |
| 4  | 3              | Condimentos     |   |   |   |   |
| 5  | 4              | Frutas/Verduras |   |   |   |   |
| 6  | 5              | Granos/Cereales |   |   |   |   |
| 7  | 6              | Lácteos         |   |   |   |   |
| 8  | 7              | Pescado/Marisco |   |   |   |   |
| 9  | 8              | Repostería      |   |   |   |   |
| 10 |                |                 |   |   |   |   |

*Fichero Tipo producto (elaboración propia)*

|   | A        | B              | C                    | D          | E               | F                |
|---|----------|----------------|----------------------|------------|-----------------|------------------|
| 1 | Nº CAMPO | CAMPO          | FORMATO              | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES |
| 2 | 1        | IDTIPOPRODUCTO | NUMÉRICO             | NO         | 50              |                  |
| 3 | 2        | DESCRIPCION    | CADENA DE CARACTERES | NO         | 150             |                  |

*Fichero Tipo producto (elaboración propia)*

El último de los ficheros creados contiene la información acerca de los productos respecto a su descripción para poder catalogarlos, este fichero no se modificará todos los meses salvo que se incluyan nuevos productos con categorías que no estén dentro de dicho fichero.

Todos ellos presentan, al menos mediante un campo, una relación para así poder trabajar con todos ellos, dicha relación entre los ficheros como el campo clave por el que se unen se muestra en la siguiente imagen;

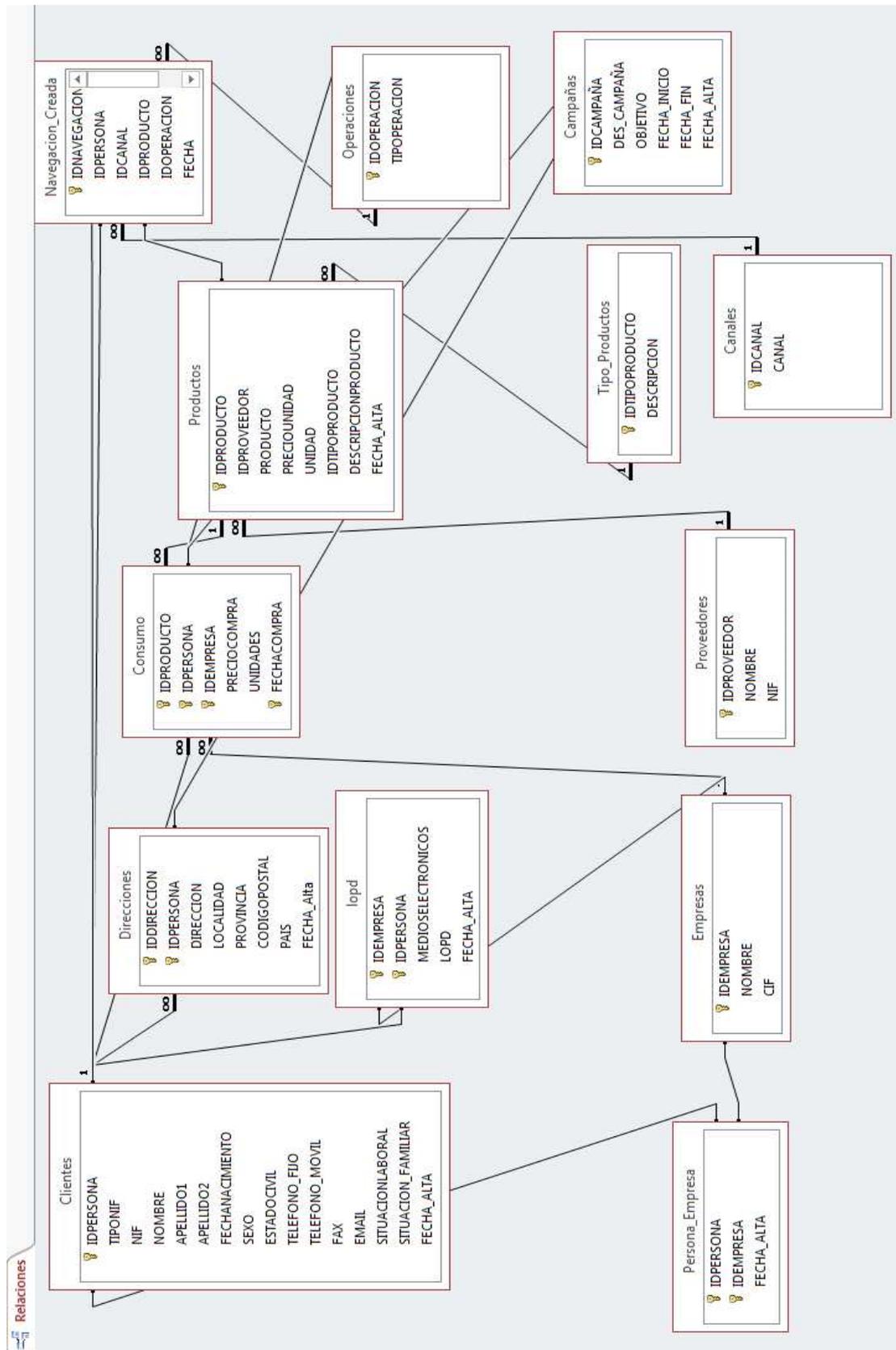


Imagen Relaciones tablas (elaboración propia)

Además de los ficheros que nos proporciona la empresa y los que creamos nosotros, tendremos finalmente tres ficheros creados a partir de los cruces entre ambos tipos de ficheros (recibidos y creados).

| Nombre                            | Fecha de modifica... | Tipo              | Tamaño |
|-----------------------------------|----------------------|-------------------|--------|
| 20160701_Campañas.txt             | 30/06/2016 18:54     | Notepad++ Docu... | 2 KB   |
| 20160701_Habitos.txt              | 30/06/2016 18:55     | Notepad++ Docu... | 2 KB   |
| 20160701_Habitos_Provinciales.txt | 30/06/2016 18:56     | Notepad++ Docu... | 1 KB   |

*Ficheros creados (elaboración propia)*

Fichero Hábitos:

| File                 | Edit      | Search          | View        | Encoding | Language | Settings | Macr |
|----------------------|-----------|-----------------|-------------|----------|----------|----------|------|
| 20140101_Habitos.txt |           |                 |             |          |          |          |      |
| 1                    | IDPERSONA | Promedio_Compra | Gasto_Medio |          |          |          |      |
| 2                    | 1042228   | 0,017713746     | 18 €        |          |          |          |      |
| 3                    | 915343682 | 0,018697843     | 19 €        |          |          |          |      |
| 4                    | 7620507   | 0,00984097      | 10 €        |          |          |          |      |
| 5                    | 7653048   | 0,021650134     | 22 €        |          |          |          |      |
| 6                    | 7586770   | 0,021010471     | 21 €        |          |          |          |      |
| 7                    | 288287    | 0,024602425     | 25 €        |          |          |          |      |
| 8                    | 7620541   | 0,02952291      | 30 €        |          |          |          |      |
| 9                    | 7652250   | 0,03936388      | 39 €        |          |          |          |      |
| 10                   | 5988362   | 0,095457408     | 95 €        |          |          |          |      |
| 11                   | 6499025   | 0,030507007     | 31 €        |          |          |          |      |
| 12                   | 6498958   | 0,020666037     | 21 €        |          |          |          |      |
| 13                   | 736448    | 0,037395686     | 37 €        |          |          |          |      |
| 14                   | 1536818   | 0,005904582     | 6 €         |          |          |          |      |
| 15                   | 6067083   | 0,022880255     | 23 €        |          |          |          |      |
| 16                   | 6499186   | 0,015253503     | 15 €        |          |          |          |      |
| 17                   | 5063706   | 0,017172493     | 17 €        |          |          |          |      |
| 18                   | 402070    | 0,038379783     | 38 €        |          |          |          |      |
| 19                   | 55197     | 0,061506062     | 62 €        |          |          |          |      |
| 20                   | 6499527   | 0,009053692     | 9 €         |          |          |          |      |
| 21                   | 5832505   | 0,079711856     | 80 €        |          |          |          |      |
| 22                   | 1470543   | 0,00984097      | 10 €        |          |          |          |      |
| 23                   | 3414478   | 0,020666037     | 21 €        |          |          |          |      |
| 24                   | 929189643 | 0,008856873     | 9 €         |          |          |          |      |
| 25                   | 6499111   | 0,004428436     | 4 €         |          |          |          |      |
| 26                   | 918224546 | 0,013777358     | 14 €        |          |          |          |      |
| 27                   |           |                 |             |          |          |          |      |

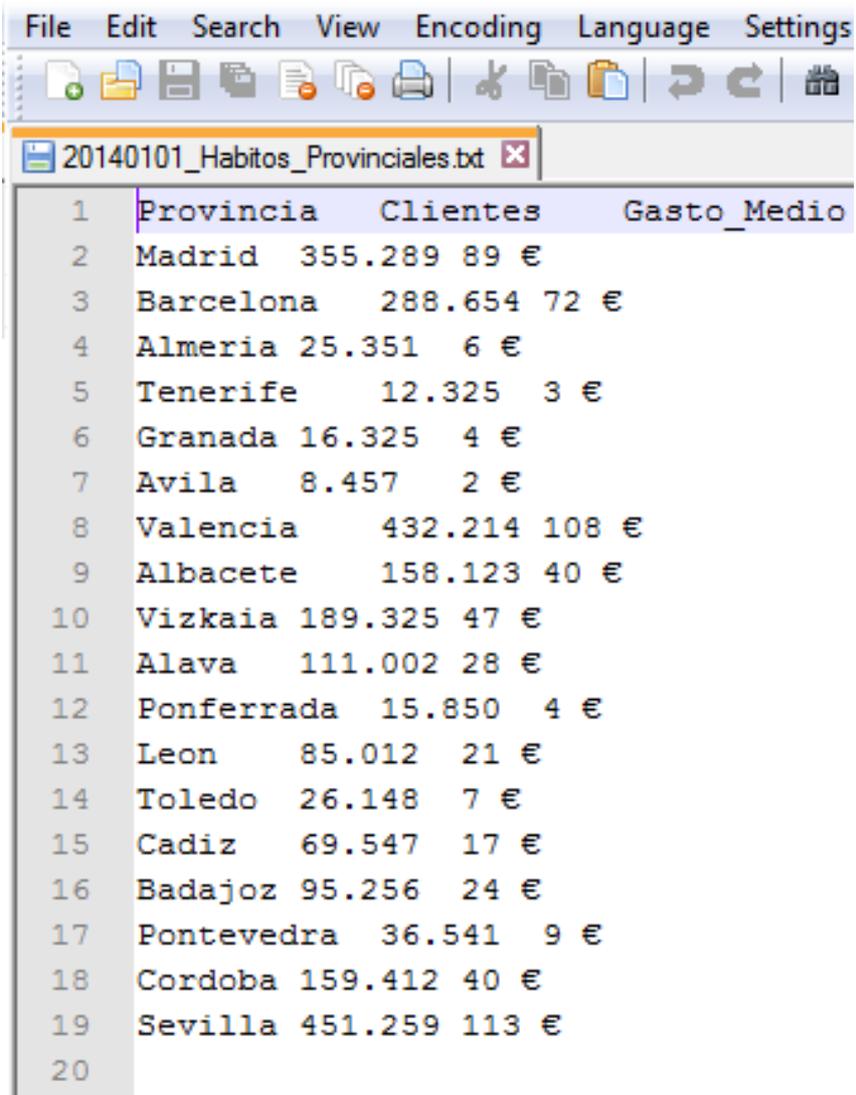
*Fichero Hábitos (elaboración propia)*

|   | A        | B               | C        | D          | E               | F                |
|---|----------|-----------------|----------|------------|-----------------|------------------|
| 1 | Nº CAMPO | CAMPO           | FORMATO  | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES |
| 2 | 1        | IDPERSONA       | NUMÉRICO | NO         | 50              |                  |
| 3 | 2        | PROMEDIO_COMPRA | NUMÉRICO | NO         | 150             |                  |
| 4 | 3        | GASTO_MEDIO     | NUMÉRICO | NO         | 10              |                  |

*Fichero Tipo Hábitos (elaboración propia)*

Dentro del fichero de hábitos tendremos todos nuestros clientes con su promedio en las compras como el gasto medio de las mismas.

*Fichero Hábitos Provinciales:*



| 1  | Provincia  | Cientes       | Gasto_Medio |
|----|------------|---------------|-------------|
| 2  | Madrid     | 355.289 89 €  |             |
| 3  | Barcelona  | 288.654 72 €  |             |
| 4  | Almeria    | 25.351 6 €    |             |
| 5  | Tenerife   | 12.325 3 €    |             |
| 6  | Granada    | 16.325 4 €    |             |
| 7  | Avila      | 8.457 2 €     |             |
| 8  | Valencia   | 432.214 108 € |             |
| 9  | Albacete   | 158.123 40 €  |             |
| 10 | Vizkaia    | 189.325 47 €  |             |
| 11 | Alava      | 111.002 28 €  |             |
| 12 | Ponferrada | 15.850 4 €    |             |
| 13 | Leon       | 85.012 21 €   |             |
| 14 | Toledo     | 26.148 7 €    |             |
| 15 | Cadiz      | 69.547 17 €   |             |
| 16 | Badajoz    | 95.256 24 €   |             |
| 17 | Pontevedra | 36.541 9 €    |             |
| 18 | Cordoba    | 159.412 40 €  |             |
| 19 | Sevilla    | 451.259 113 € |             |
| 20 |            |               |             |

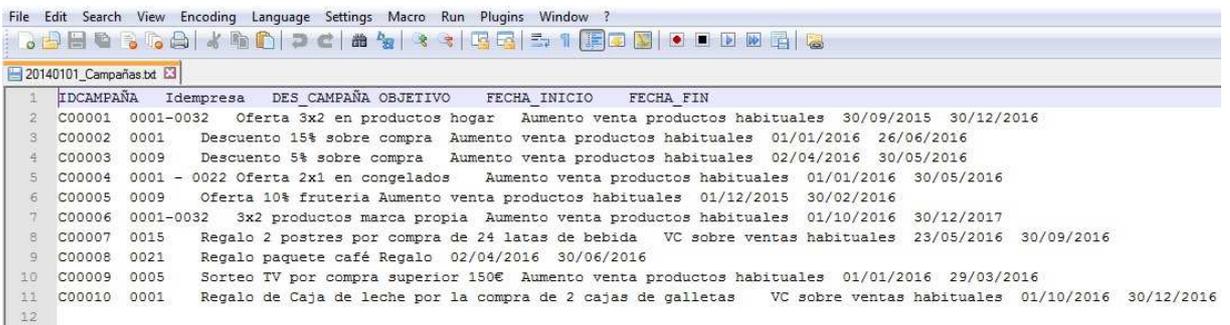
*Fichero Tipo Hábitos provinciales (elaboración propia)*

|   | A        | B           | C                    | D          | E               | F                |
|---|----------|-------------|----------------------|------------|-----------------|------------------|
| 1 | Nº CAMPO | CAMPO       | FORMATO              | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES |
| 2 | 1        | PROVINCIA   | CADENA DE CARACTERES | NO         | 150             |                  |
| 3 | 2        | CLIENTES    | NUMÉRICO             | NO         | 150             |                  |
| 4 | 3        | GASTO_MEDIO | NUMÉRICO             | NO         | 10              |                  |

*Fichero Tipo Hábitos provinciales (elaboración propia)*

Con el fichero de hábitos provinciales agruparemos por las diferentes provincias el volumen de los clientes que allí tenemos y el gasto medio de ellos.

### Fichero Campañas:



| 1  | IDCAMPAÑA | Idempresa   | DES CAMPAÑA  | OBJETIVO                           | FECHA_INICIO | FECHA_FIN  |
|----|-----------|-------------|--|------------------------------------|--------------|------------|
| 2  | C00001    | 0001-0032   | Oferta 3x2 en productos hogar                                | Aumento venta productos habituales | 30/09/2015   | 30/12/2016 |
| 3  | C00002    | 0001        | Descuento 15% sobre compra                                   | Aumento venta productos habituales | 01/01/2016   | 26/06/2016 |
| 4  | C00003    | 0009        | Descuento 5% sobre compra                                    | Aumento venta productos habituales | 02/04/2016   | 30/05/2016 |
| 5  | C00004    | 0001 - 0022 | Oferta 2x1 en congelados                                     | Aumento venta productos habituales | 01/01/2016   | 30/05/2016 |
| 6  | C00005    | 0009        | Oferta 10% frutería  | Aumento venta productos habituales | 01/12/2015   | 30/02/2016 |
| 7  | C00006    | 0001-0032   | 3x2 productos marca propia                                   | Aumento venta productos habituales | 01/10/2016   | 30/12/2017 |
| 8  | C00007    | 0015        | Regalo 2 postres por compra de 24 latas de bebida            | VC sobre ventas habituales         | 23/05/2016   | 30/09/2016 |
| 9  | C00008    | 0021        | Regalo paquete café Regalo                                   | 02/04/2016                         | 30/06/2016   |            |
| 10 | C00009    | 0005        | Sorteo TV por compra superior 150€                           | Aumento venta productos habituales | 01/01/2016   | 29/03/2016 |
| 11 | C00010    | 0001        | Regalo de Caja de leche por la compra de 2 cajas de galletas | VC sobre ventas habituales         | 01/10/2016   | 30/12/2016 |
| 12 |           |             |  |                                    |              |            |

*Fichero Campañas (elaboración propia)*

|   | A        | B            | C                    | D          | E               | F                |
|---|----------|--------------|----------------------|------------|-----------------|------------------|
| 1 | Nº CAMPO | CAMPO        | FORMATO              | VALOR NULO | LONGITUD MÁXIMA | POSIBLES VALORES |
| 2 | 1        | IDCAMPAÑA    | NUMÉRICO             | NO         | 50              |                  |
| 3 | 2        | Idempresa    | NUMÉRICO             | NO         | 50              |                  |
| 4 | 3        | DES_CAMPAÑA  | CADENA DE CARACTERES | NO         | 250             |                  |
| 5 | 4        | OBJETIVO     | CADENA DE CARACTERES | NO         | 250             |                  |
| 6 | 5        | FECHA_INICIO | FECHA                | NO         | 10              |                  |
| 7 | 6        | FECHA_FIN    | FECHA                | NO         | 10              |                  |

*Fichero Campañas (elaboración propia)*

El fichero de las campañas contendrá todas aquellas campañas que ya hemos realizado, sea para el target que sea, así como las que se van a lanzar. Cada vez que realice una nueva se cargara en el sistema para posteriormente poder tenerla en cuenta cada vez que se utilicen los ficheros.

El mapa de relaciones, tanto con ficheros recibidos, creados y resultantes de los cruces con el que trabajar queda de la forma que presenta la siguiente imagen;

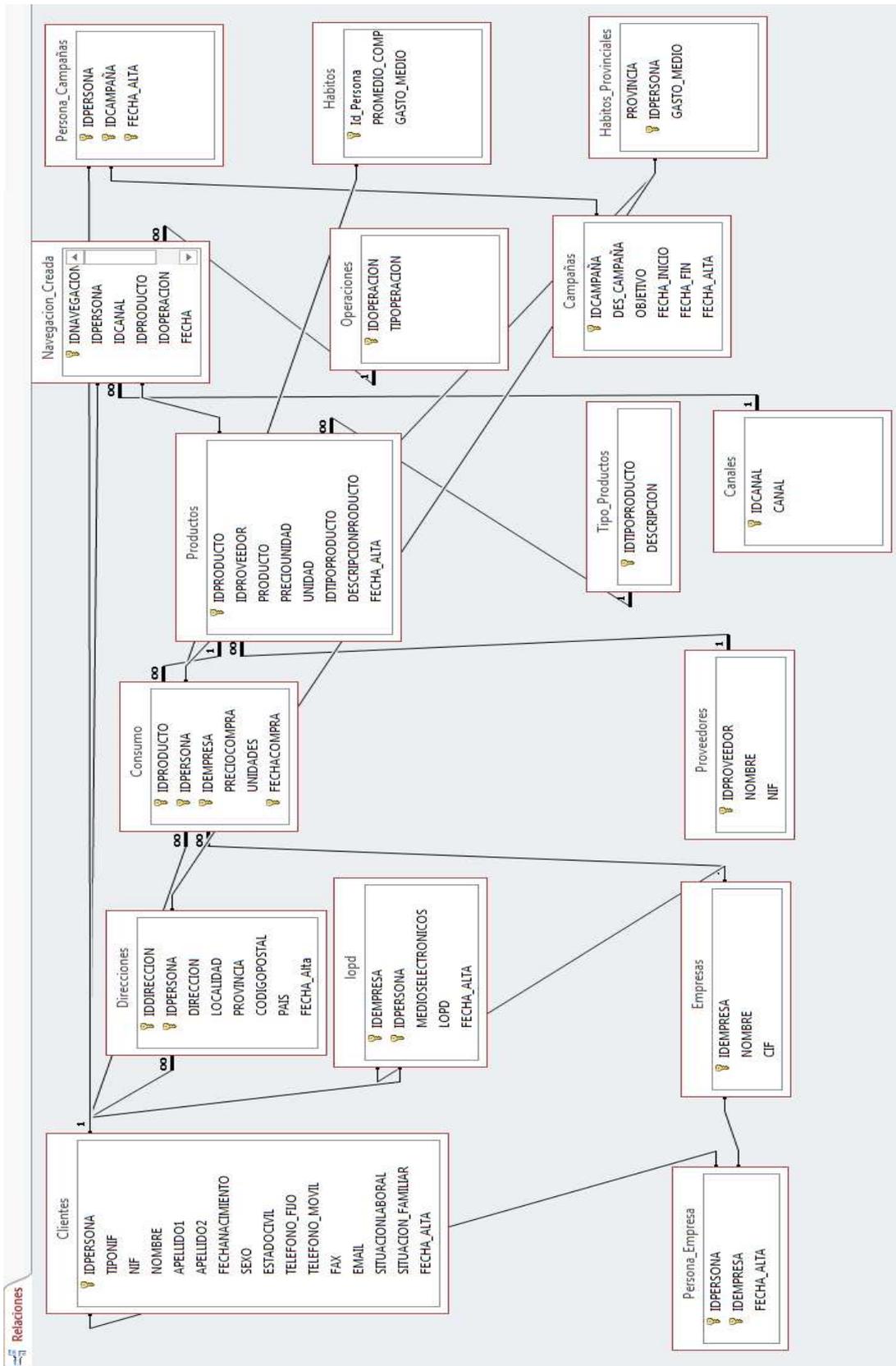
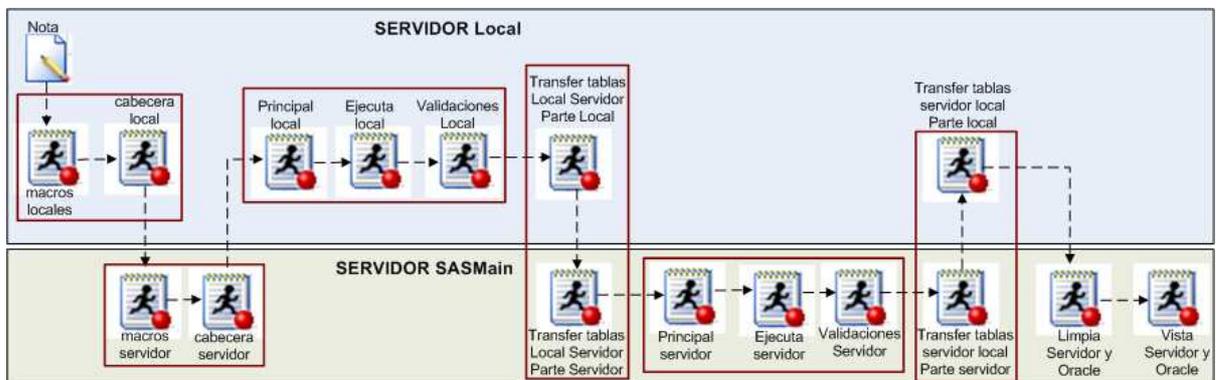
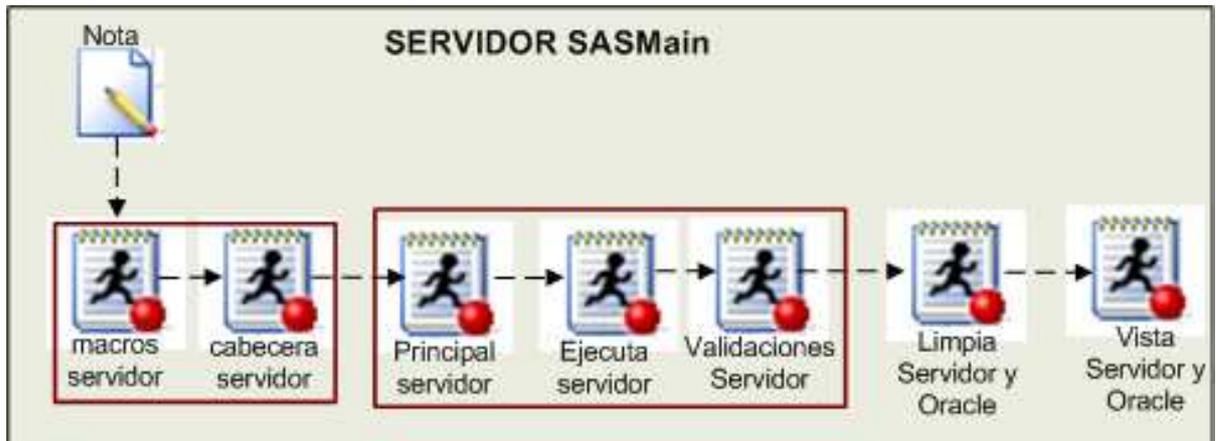
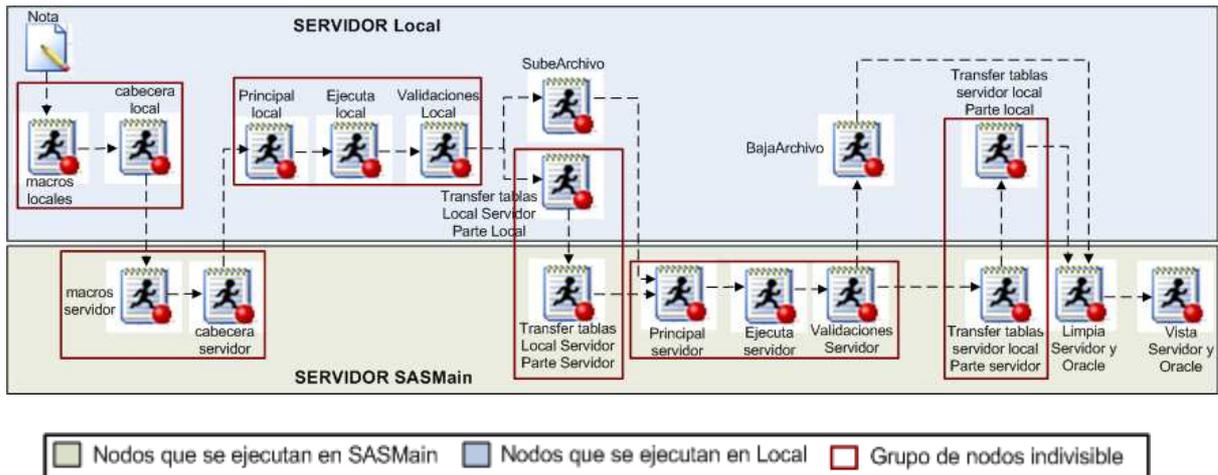


Imagen Relaciones tablas (elaboración propia)

Para finalizar mostraré algunas imágenes de la estructura y jerarquía de SAS Enterprise Guide, los diferentes nodos, relaciones, etc.

**Estructura de un proyecto de SAS Enterprise Guide**

Un proyecto SAS estará compuesto por los siguientes nodos:



*Ilustraciones SAS GUIDE (elaboración propia)*

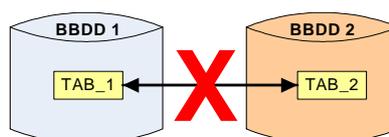
### **Cruce de información**

A la hora de cruzar información dentro del entorno SAS hay que tener presentes dos premisas básicas para la utilización correcta;

- SAS sólo se debe cruzar con SAS.
- Oracle sólo se debe cruzar con Oracle y DENTRO DE LA MISMA INSTANCIA.

Dado la limitación de recursos del servidor, y que este es compartido por todas las áreas a la hora de realizar cruces de información deberemos tener muy claro el tamaño de las tablas a cruzar y la ubicación de las mismas. Deberemos evitar;

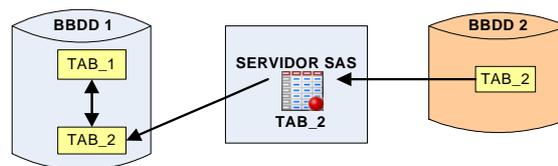
1. Para el cruce de tablas en el servidor es recomendable subir a Oracle las tablas SAS y cruzarlas en dicho entorno. En caso de ver más factible el cruce en SAS por las dimensiones reducidas de las tablas, nunca crear el resultado en la librería WORK sino en una librería definida por nosotros.
2. Cruce de una tabla SAS con una tabla de Oracle. La recomendación en este caso es subir a Oracle la tabla SAS siempre y cuando el tamaño de esta sea considerablemente inferior al tamaño de la tabla Oracle.
3. Cruce de dos tablas en la misma base de datos Oracle. Para el cruce de dos tablas que se encuentran en la misma BBDD usaremos el procedimiento de SAS “Pass through” mediante el uso de las macros `crea_tabla_or_or_s`, `inserta_datos_or_s`, ...
4. Cruce de dos tablas de distintas bases de datos Oracle. No realizaremos nunca el cruce directo de dos tablas Oracle que se encuentren en distintas BBDD ya que esto pone en peligro la estabilidad de los sistemas implicados.



Cruce tablas I (elaboración propia)

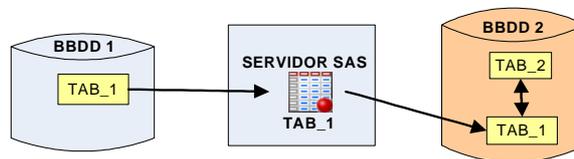
Para realizar este cruce tendremos que usar el servidor SAS como intermediario analizando previamente el tamaño de cada una de las tablas a cruzar;

- Creamos en el servidor una tabla SAS a partir de la tabla Oracle de la BBDD 2 dado que esta es la de menor tamaño. Posteriormente subimos a Oracle a la BBDD 1 la tabla SAS y así podremos realizar el cruce.



Cruce tablas II (elaboración propia)

- Creamos en el servidor una tabla SAS a partir de la tabla Oracle de la BBDD 1 dado que esta es la de menor tamaño. Posteriormente subimos a Oracle a la BBDD 2 la tabla SAS y así podremos realizar el cruce.

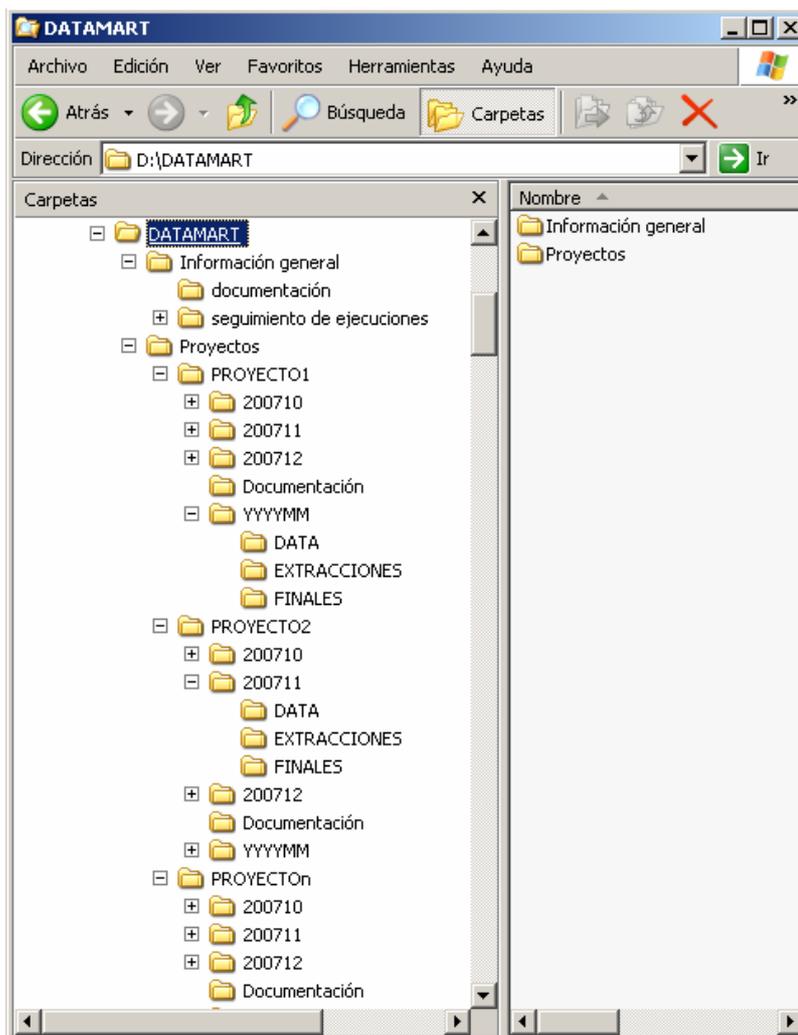


Cruce tablas III (elaboración propia)

5. Filter&Query. Únicamente se debería utilizar la utilidad para consultas de detalle de control de proceso o depurado de datos, eliminándose del proyecto una vez se haya resuelto el problema. No utilizar la utilidad 'Filter&Query' para cruzar dos tablas ya que se pierde el control de la transacción al no poder utilizar los mecanismos de control de errores, evita la automatización de los proyectos, induce a errores al permitir el cruce de dos tablas Oracle de distinta instancia y hace más costosa la lectura y depuración del proyecto.

### **Almacenamiento de proyectos**

Tanto los proyectos SAS para la carga de datos en la BBDD de nuestro DataMart, como los proyectos SAS generados para los informes y análisis, deberán ser almacenados en una estructura similar a la siguiente dentro de una unidad de red en la que se realicen backups periódicos de la información:



*Raíz almacenamiento proyecto (elaboración propia)*

### **Nomenclatura de librerías SAS en el servidor**

Debido a que las librerías dentro del servidor SAS sólo pueden contener como máximo 8 caracteres nos vemos obligados a seguir una nomenclatura que evite sobrescribir las librerías de los demás componentes del grupo de trabajo. La nomenclatura a seguir debe ser la siguiente:



Mcccccc

[M].- Inicial del usuario que ejecuta el proceso

[cccccc].-Estos 7 caracteres son de libre disposición pero tendremos que tener cuidado de no sobrescribir información.