Universidades de Burgos, León y Valladolid

Máster universitario

Inteligencia de Negocio y Big Data en Entornos Seguros

ESTUDIO DEL IMPACTO EN LA PRIVACIDAD DE LAS APP'S Y SU CONTEXTUALIZACIÓN EN EL MARCO LEGISLATIVO

Presentado por María Elena Brezmes Alonso
en Universidad de Valladolid
2 de Setiembre de 2025

Tutor: D. Quiliano Isaac Moro Sancho

Resumen

Desde la entrada en vigor del Reglamento General de Protección de Datos en mayo de 2018, en la Unión Europea se establecen una serie de medidas por ley para el cuidado de la privacidad y seguridad de los ciudadanos europeos.

Al aplicar la metodología OWASP se pueden detectar riesgos de seguridad en la auditoría de las *app's*, especialmente donde hacen alusión a aspectos como controles de privacidad inadecuados. Es decir, la falta de medidas correctas para proteger la privacidad de los datos personales de los usuarios en las *app's*, una mala configuración de seguridad o un almacenamiento inadecuado de los datos, se relaciona con la falta de medidas adecuadas para proteger los datos que se almacenan en un dispositivo móvil y por último insuficiente criptografía, es decir, un uso inadecuado o débil de técnicas de cifrado en una *app* para proteger datos sensibles.

La pregunta clave es si han tomado conciencia los programadores de *app`s* de Android de las implicaciones en la privacidad y seguridad de los usuarios el mal uso de los permisos en los cuales se basa la seguridad de las *app`s* de Android, así como si los usuarios son conscientes de lo que implica el consentimiento de ciertos permisos que se solicitan en la instalación y ejecución de las *app's* de acuerdo con el RGPD.

En este TFM vamos a crear cuadros de mando a partir de los datos de las *app's* de Android alojadas en Google Play Store con los que informar acerca del cuidado de la privacidad por parte de los programadores y acerca del cuidado de la privacidad por parte de los usuarios que hacen uso de las *app's*.

También se valora el impacto a la privacidad del uso de los diferentes grupos de permisos peligrosos que se declaran en las *app's* y que el usuario consiente al ser informado de forma clara y transparente para el funcionamiento de las *app's* en su vida diaria.

Palabras clave: *app´s,* programadores, permiso, privacidad, responsabilidad proactiva, riesgo, evaluación de impacto, RGPD, protección de datos desde el diseño.

Abstract

Since the entry into force of the General Data Protection Regulation in May 2018, the European Union has established a series of measures by law to protect the privacy and security of European citizens.

By applying the OWASP methodology, security risks can be detected in app audits, especially those related to aspects such as inadequate privacy controls. This includes a lack of adequate measures to protect the privacy of users' personal data in apps, poor security configuration, or inadequate data storage. This is related to a lack of adequate measures to protect data stored on a mobile device. Finally, insufficient cryptography, that is, inadequate or weak use of encryption techniques in an app to protect sensitive data.

The key question is whether Android app developers have realized the implications for user privacy and security of misuse of permissions, which underpin Android app security, and whether users are aware of the implications of consenting to certain permissions requested when installing and running apps under the GDPR.

In this Master, we will create dashboards based on data from Android apps hosted on the Google Play Store to report on the privacy practices of developers and the privacy practices of users who use the apps.

The privacy impact of using the various groups of dangerous permissions declared in the apps, which the user consents to by being clearly and transparently informed, is also assessed. This information is used to assess the app's functionality in their daily lives.

Keywords: app´s, programmers, permission, privacy, proactive responsibility, risk, impact assessment, GDPR, data protection by design.

Agradezco a mis padres y hermanos el inculcarme el seguir aprendiendo ante nuevos retos.

Agradezco la guía y acompañamiento del tutor en el desarrollo del TFM.

Índice general

Índice de figuras

Índice de tablas

MEMORIA

1. INTRODUCCIÓN

1.1. Contexto

En la última década prolifera el uso de los teléfonos móviles y la disponibilidad de aplicaciones móviles gratuitas en su mayoría en espacios como Google Play Store en el caso del Sistema Operativo Android.

La privacidad se entiende como el ámbito de la vida que se tiene derecho a proteger de cualquier intromisión. Ciertamente, se impacta de forma negativa al acceder a datos privados de los usuarios por las *app´s*[1].

La Unión Europea mediante el RGPD[2] **y en España** con la LOPDGDD[3] (menciona en sus inicios "*La protección de las personas físicas en relación con el tratamiento de datos personales es un derecho fundamental protegido por el artículo 18.4 de la Constitución española*"), establecen medidas legales para la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos.

Así el RGPD [67] establece la protección de las personas físicas en relación con el tratamiento de datos personales, lo considera un derecho fundamental.

La Carta de los Derechos Fundamentales de la Unión Europea («la Carta») en el artículo 8, apartado 1, y el TFUE[4] en el artículo 16, apartado 1, establecen que toda persona tiene derecho a la protección de los datos de carácter personal que le conciernan.

Por su parte, el Sistema Operativo Android considera la parte de seguridad, es un sistema operativo de código abierto, no limita a los pequeños programadores y valida las aplicaciones antes de instalarlas en un dispositivo electrónico. Propone un esquema de seguridad, para proteger a los usuarios sin necesidad de centralizar y controlar el sistema por una única empresa como realiza Apple en las aplicaciones que se destinan a su uso en iOS.

La seguridad de Android se fundamenta en tres pilares [51]:

[1] App (Application): programas informáticos se conciben para un cometido concreto y se dirigen a un determinado conjunto de dispositivos electrónicos como teléfonos inteligentes o tabletas que se conectan a internet
[2] RGPD: Reglamento General de Protección de Datos
[3] LOPDGDD: Ley Orgánica 3/2018 de Protección de Datos y Garantías de Protección de Datos
[4] TFUE: Tratado de Funcionamiento de la Unión Europea

- Android, se basa en Linux, incorpora toda la parte de seguridad de dicho sistema operativo y de esta forma impide que las aplicaciones tengan acceso directo al hardware o interfiera con recursos de otras aplicaciones.
- Toda aplicación se firma con un certificado digital, identifica a su autor. La firma digital garantiza que el fichero de la aplicación no se modifica.
- Si la aplicación accede a partes del sistema que pueden comprometer la seguridad del sistema o la privacidad del usuario, se utiliza un modelo de permisos, de forma que el usuario conozca los riesgos antes de instalar la aplicación.

Google, por su parte, contribuye a la seguridad de las aplicaciones incorpora un formulario de **Seguridad de los datos** [23] en la página de contenido de la *app* a rellenar por el programador, se informa acerca de aspectos como los permisos que se utilizan por la aplicación móvil, al ser uno de los mecanismos de seguridad que se utilizan para la protección de la privacidad en las *app´s* y a partir de los cuales interactúa con el usuario que descarga e instala la *app* en su dispositivo.

Decir a todo esto, que la cantidad de aplicaciones alojadas en Google Play Store a inicios del año 2025 estaba en torno a 3.3 millones y se actualizan constantemente ver el gráfico en la Figura 1:

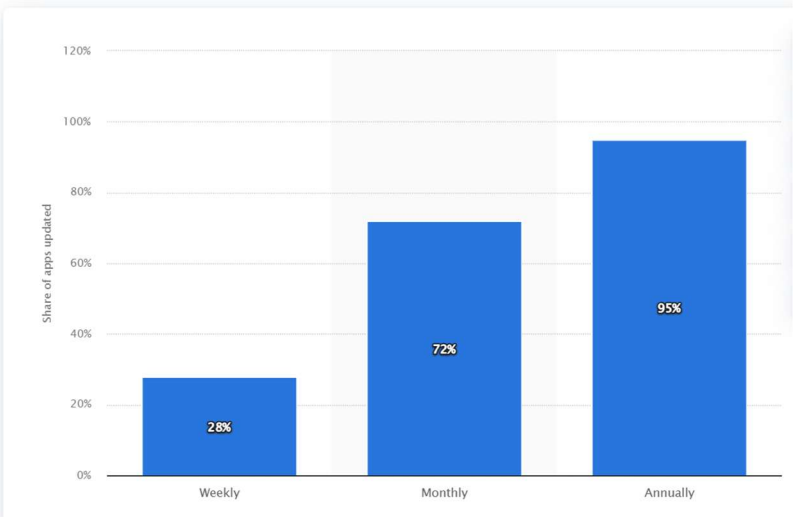

Figura 1. Frecuencia con la que se actualizan las *app´s* en Google Play Store [73]

1.2. Estado del Arte

Desde que los dispositivos móviles contienen información privada de los usuarios, Android explota el sistema de permisos para regular como se accede a ciertos recursos por parte de las *app´s*.

Sin embargo, en muchos casos se declaran permisos peligrosos sin ser necesarios en realidad para su funcionalidad en la *app*, utilizan de forma excesiva permisos por la ausencia de estrictas especificaciones de desarrollo y una efectiva supervisión de su uso.

Además, los usuarios apenas tienen conocimiento acerca de los permisos de Android y no tienen herramientas para detectar el excesivo uso de permisos por parte de las *app´s*. La mayoría de los usuarios consienten los permisos que solicitan las *app´s*, esto conlleva un alto riesgo en cuanto a la privacidad. [55]

Los permisos son uno de los aspectos más importantes para proteger la privacidad y seguridad de los usuarios de Android. Los programadores muchas veces no los utilizan correctamente, solicitan más o menos permisos de los que se necesitan de acuerdo a la funcionalidad de la *app*.

Una de las buenas prácticas para proporcionar seguridad y privacidad al usuario es usar el menor número de permisos posible. [59]

Usar los teléfonos móviles se convierte en algo habitual para las personas, las *app´s* proporcionan un amplio abanico de funcionalidades como la navegación, entretenimiento, fitness, etc, para ello las *app´s* necesitan acceder a datos sensibles del usuario.

El RGPD y la LOPDGDD demandan mejores prácticas a la privacidad, proteger a los usuarios frente a la invasión a la privacidad y no convertir al ecosistema vulnerable.

Los programadores deben diseñar sus *app´s* de forma que sean compatibles con el marco legislativo. Sin embargo, no es fácil comprender los términos legales y cómo afectan al desarrollar las *app´s*.

No comprender la legislación se materializa en diferencias entre lo que los programadores mencionan en la política de seguridad y lo que realmente implementan en sus *app´s*. [30]

Se relaciona el contenido del RGPD con los requisitos al recolectar y procesar datos personales. Se mencionan los siguientes principios: principio de finalidad (Artículo 5-1(b)), principio de minimización de datos (Artículo 5-1(c)), principio de transparencia

(Artículo 5-1 (a)), principio de protección de datos desde el diseño, configuración y procesamiento mínimo de datos (Artículo 25-2).

El consentimiento se presenta de tal forma que se distinga claramente de los demás asuntos, de fácil acceso, con un lenguaje claro y sencillo. Esto implica un cambio en las *app´s*, el consentir el uso de los permisos debe incluir la opción de sólo en este momento y así libremente y consciente consentir el riesgo de procesar datos personales, consentimiento parcial. (Artículo 7-2)

A partir de la versión 9.0 de Android el usuario tiene más control sobre el consentimiento de los permisos. Sin embargo, al consentir, no tiene forma de retirar dicho consentimiento.

Con la entrada en vigor del RGPD el usuario debe poder retirar el consentimiento en cualquier momento, aunque consienta el uso de un tipo de permiso peligroso por parte de la *app*, poder retirar dicho consentimiento en cualquier momento, supone cambios en la forma de programar. [68]

Los programadores de *app´s* deben conocer los permisos necesarios para cada API y de esta forma elegir los permisos correctos con mínimos privilegios.

Los permisos se declaran en el fichero *AndroidManifest.xml* y los permisos peligrosos se consienten por el usuario en tiempo de ejecución. [72]

En Europa la protección a la privacidad se considera un derecho humano, sin embargo, auditar y testear las *app´s* contra los requisitos legales de protección de datos no resulta trivial.

Hay distancia entre los requisitos legales del RGPD y cómo se trasladan dichos requisitos a soluciones prácticas, se necesitan herramientas con las cuales testear, verificar y auditar *app´s.*

En el estudio "*GDPR Compliance Assessment for Cross Border Personal Data Transfers in Android App´s*" [33] establecen una metodología para comparar la política de privacidad con el RGPD, en las *app´s* de Android. La ubicación, contactos, identificador único del dispositivo del usuario incluyendo el IMEI[5], IMSI[6], UDID[7], número del teléfono móvil, datos biométricos, identidad del teléfono, registros de

[5] IMEI: International Mobile Equipment Identify
[6] IMSI: International Mobile Subscriber Identity
[7] UDID: Unique Device Identifier

llamadas, SMS, historial de navegación, correos electrónicos, fotos y videos se denominan datos personales en el RGPD.

En el estudio "*Share First, Ask Later (or Never?) Studying Violations of GDPR's Explicit Consent in Android App´s*"[71] utilizan una muestra de 24.838 app´s y comprueban que el 34,3% de la muestra envían datos personales sin un consentimiento previo por parte del usuario de acuerdo al RGPD. Comunican a los programadores el resultado y constatan la necesidad de una mejor información y documentación, así como herramientas para evitar el incumplimiento de los requisitos legales del RGPD.

Al analizar *app´s* de categoría social, se parte del hecho de que están pensadas para socializar en internet, por lo cual, algunos permisos como acceder a la red son vitales en este tipo de *app´s*. La posibilidad de compartir ficheros multimedia, permitir el acceso a la cámara, ubicación o grabar audio se añaden a la funcionalidad de la *app*, se establecen como uso muy común. Esto significa que, por el simple hecho de utilizar *app´s* de categoría social, el usuario asume riesgos a la privacidad y protección de datos ya que deben consentir ciertos permisos para el correcto funcionamiento de la *app*. [63]

En la versión 12 de Android se añade un panel de privacidad disponible para la consulta por parte del usuario, proporciona una vista rápida de los permisos consentidos. [14]

1.3. Preguntas de investigación

Ante los hechos que se mencionan en el contexto y estado del arte, (ver apartados 1.1 y 1.2) se plantean dos preguntas:

RQ1: ¿Tienen conciencia los programadores de aplicaciones móviles de Android de lo que implica el RGPD en el uso de los permisos por parte de las *app´s*?

RQ2: ¿Tienen conciencia los usuarios de aplicaciones móviles de las consecuencias en su privacidad el consentimiento de los permisos que solicitan las *app´s*?

1.3.1 Información para los Programadores

Una de las principales amenazas de la privacidad en los dispositivos móviles proviene de instalar aplicaciones móviles por los permisos que utilizan. Los programadores deben realizar cambios desde el diseño, informar de forma adecuada al usuario y así entienda lo que acepta.

Android en sus políticas establece solicitar el consentimiento del usuario en el momento de la instalación de la *app* de aquellos permisos que utilizan datos sensibles, preguntar al usuario de forma explícita.

El RGPD demanda mejores prácticas por parte de los programadores de *app´s*, proteger la privacidad de los usuarios lo marca la ley, sin embargo, para los programadores no es fácil comprender los principios legales que aplican a su desempeño.

En el ecosistema de los dispositivos móviles se promete una cosa en la política de privacidad y se implementa otra con los diferentes permisos, se observa una brecha entre los principios legales de privacidad, seguridad y lo que se implementa.

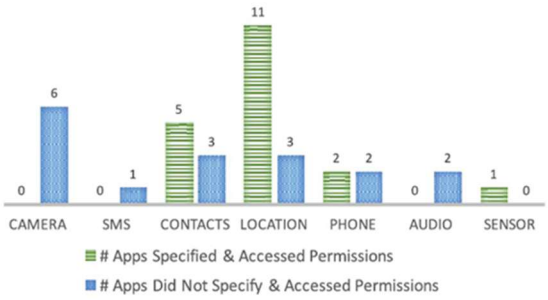

Figura 2. Diferencias entre comunicar e implementar [16]

Las *app´s* muchas veces solicitan más permisos de los que realmente necesitan para su funcionamiento. Muchos permisos son de tipo *signature* o *signatureOrSystem* los cuales no son fáciles de desinstalar por parte del usuario, constituyen una amenaza para su privacidad.

El RGPD intenta revertir la falta de conciencia por parte de los programadores del cuidado de los aspectos de privacidad desde el diseño de las *app´s*. Los programadores deben tomar conciencia y asumir su responsabilidad en la protección de la privacidad de los usuarios.

Las graves amenazas al acceder a datos sensibles de los usuarios desde las *app´s*, convierte en urgencia la necesidad de **desarrollar desde el diseño de las *app´s* la privacidad y seguridad de los usuarios por parte de los programadores de aplicaciones móviles**.

Al evaluar 86.163 *app´s* de Android se encuentran 24.838 *app´s* que envían datos personales del usuario hacia proveedores de publicidad sin un consentimiento explícito por parte del usuario (se realiza el análisis en Alemania a finales del año 2020 [71]).

Los problemas de un mal uso de los principios legales del RGPD se deben a los proveedores de publicidad, al distribuidor de las aplicaciones, Google en este caso y a los programadores de aplicaciones.

Muchos programadores no analizan todas las posibles situaciones cuando un usuario no acepta el consentimiento a utilizar ciertos permisos, no observan los cambios del sistema operativo, ni el funcionamiento por parte de librerías de terceros sobre el uso de los permisos en tiempo de ejecución.

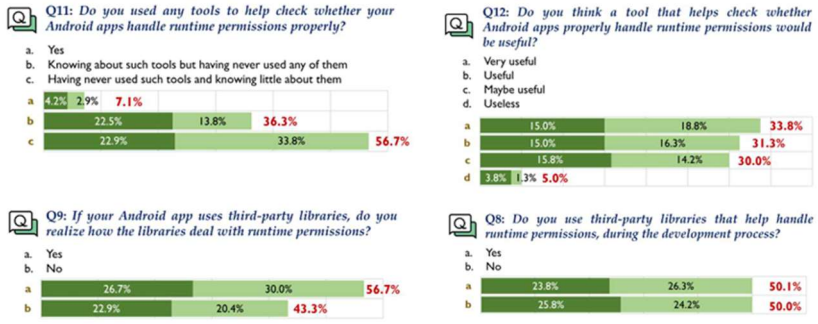

Figura 3. Preguntas del uso de los permisos por parte del programador [69]

Valoran el añadir una opción más en el momento del consentimiento de los permisos por el usuario, no sólo *sí* o *no,* sino "*puede ser*" se lo plantean desde la perspectiva del cumplimiento legal del RGPD, punto de vista desde un **consentimiento parcial**. [1] [2] [4] [5] [28] [30] [33] [42] [59] [68] [69]

1.3.2 Sobre los usuarios

Los usuarios no se detienen ante las ventanas emergentes de las aplicaciones móviles al informar y solicitar el consentimiento del uso de los permisos a utilizar por éstas; esto convierte en amenazas los datos que se alojan en sus dispositivos.

La mayoría de los usuarios no comprenden el funcionamiento de los permisos de las aplicaciones móviles, los aceptan sin más y esto conlleva un riesgo a la privacidad, permiten el uso de permisos que en realidad no necesita la *app* en su funcionamiento. El usuario al aceptar ciertos permisos respecto a una aplicación no suele cambiar su decisión.

Esto constituye, en definitiva, la falta de conciencia del usuario en cuanto al cuidado de su privacidad. El RGPD intenta revertir esta situación: el usuario debe otorgar el consentimiento de forma explícita antes de instalar cierto tipo de permisos las aplicaciones móviles.

Reason for *Granting* Permissions	
I wanted to enable a feature of the app.	227
The app asked for it.	117
I didn't think about it.	106
Other	18

Figura 4. Motivos por los cuales consentir permisos de forma correcta [1]

Reason for *Granting* Permissions	
I didn't need the feature.	357
I was concerned about my privacy.	212
I didn't think about it.	135
I was concerned about the security of my device.	88
Other	22

Figura 5. Motivos por los cuales consentir permisos de forma incorrecta [1]

Permission Name	#req.	Denied (%)	Granted (%)
read phone numbers	731	83.99	16.01
get accounts	4139	64.89	35.11
access background location	1771	63.13	36.87
Bluetooth scan	1490	62.55	37.45
camera	6342	58.33	41.67
read calendar	1534	57.24	42.76
record audio	4557	55.63	44.37
read phone state	4909	54.94	45.06
access fine location	5905	51.96	48.04
write calendar	1228	51.95	48.05
access coarse location	6445	51.34	48.66
read contacts	5120	51.07	48.93
write external storage	8575	50.24	49.76
read external storage	9792	48.72	51.28
access media location	1384	47.47	52.53
write contacts	2077	33.37	66.63
activity recognition	657	32.27	67.73
query all packages	2891	31.75	68.25
read SMS	1237	21.18	78.82
read call log	1327	19.67	80.33
body sensors	176	18.75	81.25

Figura 6. Porcentaje de permisos que deniegan y consienten los usuarios [1]

El estudio [1] demuestra que muchos usuarios no tienen conciencia o conocimiento acerca de los permisos que aceptan cuando se los solicitan las aplicaciones.

Solicitar el consentimiento explícito al usuario de ciertos permisos, permite construir un modelo al usuario acerca de lo que acepta y los permisos que considera necesarios en la aplicación.

La mayoría de los usuarios no revoca los permisos de sensores sensibles como el micrófono y la cámara, sólo sino afecta a la funcionalidad principal de la aplicación y ésta no se utiliza con frecuencia.

Se puede observar en el gráfico de la figura 6 como los permisos de localización y almacenamiento se solicitan con una mayor frecuencia por las *app´s* y no están entre los que más se deniegan por los usuarios (observar figura 7) a pesar de ser datos sensibles y por lo tanto poner en peligro su privacidad.

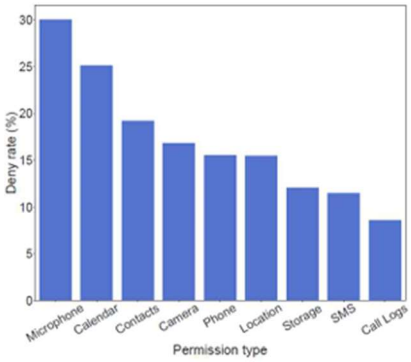

Figura 7. Permisos que se deniegan en la muestra del estudio en 10 países [3]

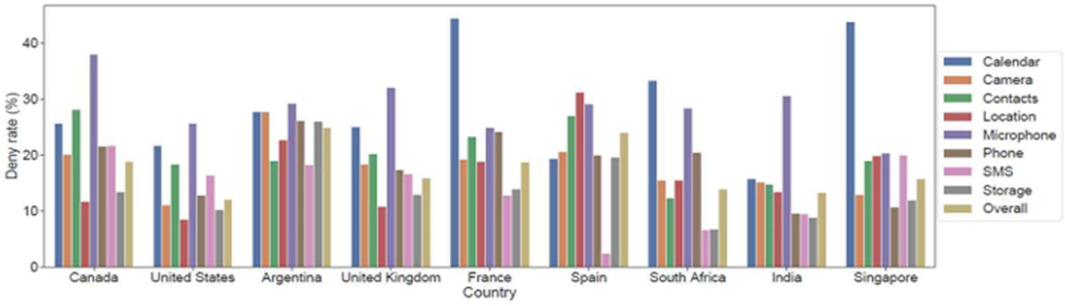

Figura 8. Ratio grupo de permisos peligrosos que se deniegan en cada país [3]

Observar en la figura 9 la discrepancia entre los altos valores en la actitud del usuario ante la privacidad y los bajos valores al denegar permisos que la invaden, lo denominan la paradoja de la privacidad.

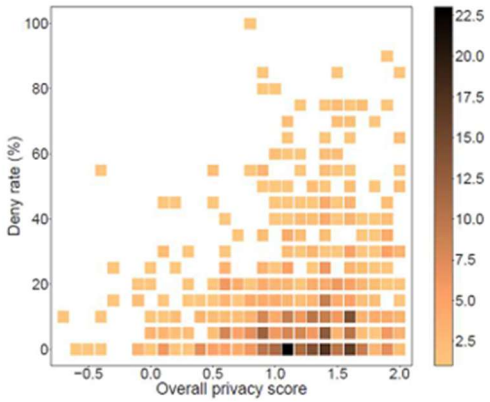

Figura 9. Correlación de la privacidad y porcentaje de permisos que se deniegan [3]

Los usuarios aceptan los permisos al pensar en el buen hacer del programador. [1] [2] [3] [5] [33] [71]

1.4. Motivación

El RGPD plantea un nuevo contexto en relación al cuidado de la privacidad, las *app´s* impactan en la privacidad, cuestionar si los programadores de aplicaciones móviles toman conciencia de los riesgos a la privacidad de un mal uso de los permisos al diseñar e implementar *app´s* motiva la realización del proyecto BI.

Al realizar la auditoría de aplicaciones móviles de acuerdo a la metodología OWASP[8] [56] en el año 2024, se observan entre los riesgos más significativos algunos que se relacionan con la privacidad de los usuarios como se menciona en el resumen al iniciar el documento.

Esto presenta un escenario en el cual las *app´s* no cumplen con lo que se estable en el marco legislativo o bien no se hace buen uso de ellas.

Figura 10. OWASP Mobile Top 10 2024 (M6, M9, M10) [57]

El programador dispone del esquema de permisos de Android para usar recursos y características especiales del hardware, si bien toda *app* al acceder a estos recursos está obligada a comunicar de forma concisa al usuario, utilizar un lenguaje claro y sencillo, con fácil acceso y fácil de entender.

Facilitar esta información en forma electrónica cumple con el principio de transparencia que se contempla en el RGPD.

[8] OWASP: Open Web Application Security Project

El ecosistema de las *app´s*, lo componen diferentes actores a tener en cuenta a la hora de valorar su buen funcionamiento. Uno de ellos lo forman los programadores, si bien los usuarios también forman parte al usar las *app´s* en su vida diaria.

Concienciar del impacto en la privacidad de los usuarios el mal uso de los permisos y observar si los usuarios toman conciencia del impacto en su privacidad al instalar las *app´s* supone una motivación más, ante la necesidad de reducir los riesgos que se detectan en la auditoría de OWASP.

Otro punto importante es el disponer de un volumen de datos de *app´s* con los diferentes permisos que se declaran en cada una de ellas, analizar y construir la aplicación BI y así conocer si la forma de utilizar los permisos por parte de los programadores es la correcta.

A esto añadir el tiempo que transcurre desde la entrada en vigor del RGPD en mayo de 2018, parece interesante el realizar el estudio acerca del impacto en la privacidad por parte de las *app´s* de acuerdo al marco legislativo.

1.5. Organización de la memoria

1. **Introducción**: situar el contexto donde se desenvuelve lo que va a ser el proyecto de BI realizar un recorrido por los documentos, conferencias, publicaciones y aporten luz a los objetivos del proyecto de BI y constatar resultados y la motivación para su realización.
2. **Objetivos del proyecto**: establecer los objetivos a alcanzar mediante el desarrollo del proyecto BI.
3. **Conceptos teóricos:** definir los conceptos en los cuales se basará el proyecto, la metodología a seguir en el proyecto BI a partir de los objetivos planteados, hasta la obtención de los resultados.
4. **Técnicas y Herramientas:** definir las diferentes técnicas y herramientas a utilizar a lo largo del proyecto desde la obtención de los datos hasta los resultados
5. **Impacto en la Privacidad de las app´s de acuerdo con la AEPD**: aterrizar en los aspectos relevantes en el desarrollo del proyecto BI para la obtención de los objetivos planteados de acuerdo a la guía de la gestión del riesgo y evaluación de impacto en tratamientos de datos personales de la AEPD.

6. **Buenas Prácticas:** aspectos a tener en cuenta desde el diseño de las *app´s* para minimizar los riesgos de la privacidad y con ello minimizar el impacto de la privacidad de las *app´s* de acuerdo al marco legislativo.
7. **Conclusiones:** comentar los resultados obtenidos en el proyecto de BI y posibles trabajos futuros que amplíen el conocimiento del estudio.

2. OBJETIVOS

El objetivo principal del TFM es ayudar a la toma de decisiones por parte de los programadores a la hora de construir una *app* desde su diseño, descubrir aquello que les puede aportar la recopilación de los datos, conocer la intromisión en la privacidad y anonimizar de acuerdo con el marco legislativo vigente en la Unión Europea, Reglamento General de Protección de Datos y su transposición en la Ley Orgánica de Protección de Datos Personales y Garantía de los Derechos Digitales en España.

Este objetivo general se desglosa en los siguientes componentes:

- Analizar los permisos que utilizan las *app´s* en Android y obtener los más influyentes en cuando a la privacidad.
- Analizar la intromisión en datos personales por parte de la *app*
- Conocer el impacto del acceso a información de registro de actividades vitales por la *app*
- Conocer el impacto del acceso a aspectos personales de los usuarios por la *app*
- Conocer el riesgo del acceso a preferencias de consumo, hábitos, gustos, necesidades del usuario por la *app*
- Conocer el riesgo del acceso a la ubicación del dispositivo donde se aloja la *app*
- Analizar el riesgo de acceso a metadatos por parte de las *app´s*
- Descubrir el riesgo del acceso a identificadores únicos de los dispositivos electrónicos
- Conocer el riesgo del acceso a datos y metadatos de las comunicaciones electrónicas
- Descubrir el riesgo de acceso a datos de navegación
- Conocer el comportamiento del usuario ante el cuidado de su privacidad
- Disponer de diferentes filtros o segmentos como son la categoría de la *app*, tipo de público, fecha de lanzamiento y programador en el análisis.
- Obtener cuadros de mando a partir de los datos y conocer a partir de la aplicación BI el impacto a la privacidad de las *app´s* de acuerdo al marco legislativo.
- Utilizar la evaluación de riesgo de la AEPD en la valoración del impacto.

Con todo ello concienciar a los programadores de la parte importante en su desempeño para la mejora de la privacidad de los usuarios en el uso de las *app´s* en su vida diaria.

3. CONCEPTOS TEÓRICOS

3.1. Big Data

Al tratar los Big Data o datos masivos, del qué no del porqué de los datos, podemos dejar que los datos hablen por sí mismos, sin necesitar conocer la causa de un fenómeno.

Antes de los datos masivos, los análisis se limitan habitualmente a someter a prueba un reducido número de hipótesis que definen con precisión muchas veces antes de recopilar los datos.

Ahora, con el Big Data, se establecen conexiones entre los datos que no se sospechan. Una gran cantidad de cosas se pueden medir, almacenar, analizar y compartir se convierten hoy en día en datos: datos de los sensores, GPS de los teléfonos móviles, clics en la red y en Twitter.

Disponer de un volumen de datos ofrece mucha más libertad para explorar, para estudiar datos desde diferentes perspectivas, o examinar más de cerca determinados aspectos.

Los datos masivos son un recurso y una herramienta. Informan antes que explican. Las predicciones se basan en correlaciones son el corazón de los datos masivos. En lugar del enfoque que se sustenta en hipótesis se emplea uno que se sustenta por datos.

El uso del Big Data probablemente permite que nuestros resultados sean menos sesgados y más precisos, obtener información mucho más deprisa.

La palabra data significa “dato”, en el sentido de “hecho”. Así “datificar” un fenómeno es plasmar en un formato cuantificado y se pueda tabular y analizar.

Por ejemplo, al instalar módulos inalámbricos en los vehículos, al datificar la localización transforman el concepto de los seguros. Los datos ofrecen una vista pormenorizada de los tiempos, localizaciones y distancias de conducción real permiten un precio mejor en función del riesgo. Transforman la naturaleza del seguro: se basan en la actuación individual y no grupal y así incentivan el buen comportamiento.

La capacidad de recopilar datos de geolocalización de los usuarios los convierte en algo muy valioso. A escala individual permiten la publicidad personalizada allí donde esté o se prevean vaya a estar una persona por medio de los teléfonos iPhone y los teléfonos Android al recopilar datos sobre la ubicación y WiFi, y remitirla a Apple y Google respectivamente sin que los usuarios sean conscientes de ello. Es más, la información se agrega para revelar tendencia.

El estudio de este TFM es una datificación. Las colecciones de datos se fusionan y se obtiene información acerca del impacto a la privacidad de las *app´s* al utilizar diferentes filtros o segmentaciones de los datos. Se informa al programador cómo implementa los grupos de permisos en las *app´s*.

Se informa de las *app´s* de forma individual, permite concienciar al programador y reducir el impacto a la privacidad de las *app´s* pero también se informa de forma conjunta del nivel de conocimiento de los programadores acerca del uso de los grupos de permisos peligrosos y el impacto a la privacidad del usuario.

Los usos no comerciales de la geolocalización terminan siendo los más importantes, se conoce como *reality mining*, "minería de la realidad". Procesan enormes cantidades de datos procedentes de teléfonos móviles, extraen inferencias y predicciones sobre el comportamiento humano.

En uno de sus estudios, analizan los movimientos y los patrones de llamadas les permite identificar a personas que contraen la gripe antes de que ellas mismas sepan que están enfermas. Ahora bien, en manos irresponsables, el poder del *reality mining* puede tener consecuencias indeseables.

El entusiasmo por el "internet de las cosas" tiene que ver con el networking o las redes de contactos, casi tanto o más con datificar todo cuanto nos rodea. Al datificar el mundo, los usos potenciales de la información no tienen más límite que el ingenio personal.

La tecnología llega a un punto en el que es capaz de capturar y almacenar unas cantidades inmensas de información por poco dinero. Resulta frecuente recopilar datos de forma pasiva, sin mucho esfuerzo o sin que sean conscientes siquiera aquellos a los que registran.

Al caer tanto el coste del almacenamiento se justifica la conservación de los datos, no se descarta la información con el uso, el valor de los datos es lo que se gana de todas las formas posibles en que los empleen.

Existen tres importantes vías para desencadenar el valor de opción de los datos: reutilizarlos de forma básica, fusión de conjuntos de datos y hallazgos de combinaciones "dos por uno" (la fusión de conjuntos de datos la suma es más valiosa que sus partes) [18].

3.2. Privacidad de las *App´s*

Actualmente la mayoría de las *app´s* se desarrollan con el Sistema Operativo de iOS o con el Sistema Operativo Android. Aquí se desarrollan conceptos teóricos del Sistema Operativo Android al pertenecer el volumen de datos del proyecto BI a *app´s* que se instalan en dispositivos electrónicos con dichos sistemas operativos.

Android se crea por Google y la *Open Handset Alliance*, paquete de software de código abierto, se usa y configura el sistema sin pagar un canon. Es una arquitectura con un alto nivel de seguridad y se basa en Linux. Se aíslan unos programas de otros al ejecutarse dentro de una caja (*sandbox*) que incorpora la máquina virtual *Dalvik.*

Cada *app* dispone una serie de permisos accede a recursos y servicios del sistema. Por medio de los permisos ofrece seguridad a los usuarios.

El programador declara los permisos a utilizar en el fichero *AndroidManifest.xml*, describe qué permisos solicita la *app*, qué servicios se ejecutan en un segundo plano sin necesidad de disponer de interfaz de usuario y qué clases son las encargadas de recibir y manejar eventos del sistema. [26]

Desde el punto de vista de la privacidad y seguridad, resulta importante verificar que la aplicación sigue el **principio de mínimos privilegios** y no se registran más permisos de los que se necesitan para su funcionamiento. Otorgar excesivos permisos y privilegios conforma una de las vulnerabilidades más comunes, crea la mayor parte de problemas de privacidad en las aplicaciones móviles.

Las aplicaciones se instalan en el dispositivo móvil y tienen privilegios de acceso: acceso a la lista de contactos, recepción y envío de mensajes, acceso a micrófono, acceso a la cámara, etc.

El modelo de permisos en Android contempla diferentes niveles de protección de permisos. Actualmente hay 4 tipos de permisos:

- **Normal**: en el momento de la instalación de la *app* se permite estos permisos por su bajo riesgo sin necesidad del consentimiento previo por parte del usuario, es decir, el sistema otorga automáticamente en el momento de la instalación este tipo de permisos por su menor riesgo y el usuario no puede revocarlos.
- ***Signature***: se concede en el momento de la instalación, pero solo cuando la aplicación se firma por el mismo certificado que la aplicación que define el permiso, es decir, sólo otorga permisos para las aplicaciones que firma el mismo programador.

- ***SignatureOrSystem***: se diferencia del tipo *signature* en que además se puede usar por el sistema.
- ***Dangerous***: por su alto riesgo, sólo se accede a este tipo de permisos si el usuario da su consentimiento en el momento de ejecución de la *app*. Son permisos con **acceso a datos privados del usuario**.

Los grupos de permisos se detallan en la Tabla 1, agrupan permisos peligrosos o *dangerous*, al otorgar el consentimiento a uno de los permisos peligrosos, se otorga el consentimiento a todos los permisos del mismo grupo de permisos.

Grupo de permisos	Descripción
Calendar	Acceder a la agenda del dispositivo
Contacts	Acceder a los contactos del dispositivo
Location	Acceder a la ubicación del dispositivo
Microphone	Acceder al audio del dispositivo
Phone	Acceder al teléfono del dispositivo
Device ID & call information	Acceder a los datos del dispositivo y al registro de llamadas
Wearable sensors/Activity data	Acceder a los datos de los sensores
SMS	Acceder a los datos de los mensajes en el dispositivo
Storage	Acceder a los datos almacenados en el dispositivo

Tabla 1. Grupo de permisos peligrosos

Consultar el apéndice A. "Permisos en el Sistema Operativo de Android" para conocer los permisos que se pueden declarar en cada uno de los grupos peligrosos en una *app*.

3.3. Marco legislativo

La protección de datos experimenta una evolución técnica y legislativa sin precedentes en ningún campo del Derecho, tanto a nivel nacional, como europeo. La entrada en vigor en mayo de 2018 del Reglamento General de Protección de Datos de la Unión Europea (679/2016) supone una amplia revisión de la normativa sobre la materia.

La Constitución Española reconoce en su artículo 18.3 y 18.4 el derecho a la intimidad, se garantiza el secreto de las comunicaciones telefónicas y limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos al tratarse de derechos fundamentales.

Estos derechos también se sustentan a nivel supranacional, en la Carta de los Derechos Fundamentales de la Unión Europea y en la Declaración Universal de los Derechos Humanos.

El revolucionario desarrollo de las tecnologías digitales provoca un rotundo cambio de paradigma tanto en el funcionamiento de nuestra sociedad, como en nuestro propio modo de comprender la realidad y de interactuar en el mundo.

El uso indebido de las herramientas digitales como los dispositivos móviles propicia consecuencias con un gran impacto, tanto en su perspectiva social como en la individual, al potenciar exponencialmente los riesgos ya existentes y generar otros nuevos de gran calado que afectan directamente a la seguridad y privacidad de los usuarios.

En este escenario, los poderes públicos establecen un marco normativo acorde a la cambiante realidad y proporcionan medios suficientes para llevar a cabo el proceso de implantación de las medidas de seguridad y hacen partícipe a la sociedad del cuidado de su privacidad. [25]

El RGPD contempla el uso de los datos recogidos por parte de las *app´s* desde su diseño se mencionan en los siguientes artículos:

- **Artículo 4** – Definiciones – "Datos Personales"[9]
- **Artículo 5** – Principios relativos al tratamiento – donde se establece los datos personales
- **Artículo 7** – Condiciones para el consentimiento
- **Artículo 25** – Protección de datos desde el diseño y por defecto
- **Artículo 32** – Seguridad de los datos personales

Consultar el contenido de estos artículos en el apéndice B. "Artículos del RGPD".

3.3.1 Privacidad de datos desde el diseño

La AEPD[10] traslada a la práctica el *principio de protección de datos desde el diseño*, reconoce la importancia de incorporar los principios de privacidad dentro de los procesos de diseño.

El Reglamento General de Protección de Datos 2016/679 (UE), en su artículo 25 y bajo el epígrafe '*Protección de datos desde el diseño y por defecto*', incorpora a la normativa de protección de datos la práctica de considerar los requisitos de privacidad desde las primeras etapas del diseño de productos y servicios.

[9] «datos personales»: toda información sobre una persona física identificada o identificable («el interesado»); se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona;
[10] AEPD: Agencia Española de Protección de Datos

La privacidad desde el diseño implica utilizar un enfoque orientado a la gestión del riesgo y de responsabilidad proactiva, establecer estrategias que incorporen la protección de la privacidad a lo largo de todo el ciclo de vida del objeto (ya sea este un sistema, un producto hardware o software, un servicio o un proceso).

El objetivo es la protección de datos desde el diseño, en las primeras fases de desarrollo y no sea una capa añadida a un producto o sistema. La privacidad debe formar parte integral de la naturaleza de dicho producto o servicio.

La privacidad desde el diseño se plantea para proporcionar al usuario el máximo nivel de privacidad, los datos personales estén automáticamente protegidos en cualquier sistema, aplicación, producto o servicio.

Figura 11. Privacidad desde el diseño como suma integral del enfoque al riesgo y la responsabilidad proactiva [41]

Se ha de establecer la configuración por defecto desde el diseño a aquel nivel que resulte lo más respetuoso posible en términos de privacidad. En caso de que el sujeto no tome ninguna acción de configuración, garantizar su privacidad y mantenerse intacta, pues se integra en el sistema y se configura por defecto.

La transparencia en el tratamiento de datos es el pilar demuestra la diligencia y la responsabilidad proactiva ante la Autoridad de Control y como medida de confianza ante los sujetos cuyos datos se tratan.

Tal y como establece el considerando 39 del RGPD, para las personas físicas debe quedar totalmente claro que se recogen, utilizan, consultan o tratan de otra manera datos personales que les conciernen, así como la medida en que dichos datos son tratados.

El artículo 83 considera sancionable no atender a la obligación de la protección de datos desde el diseño, al igual que su correcta aplicación constituye uno de los criterios para baremar la gravedad de una infracción.

Si bien el cumplimiento de esta obligación aplica específicamente al responsable del tratamiento, a la luz del considerando 78 y lo que se establece en el artículo 28 del RGPD, la protección de datos desde el diseño se proyecta sobre otros actores participantes en el tratamiento de datos personales como son los proveedores y prestadores de servicios, programadores de productos y aplicaciones o fabricantes de dispositivos.

Comprender cómo un tratamiento de los datos personales puede llegar a afectar a la privacidad de los individuos es la clave para diseñar y desarrollar sistemas confiables desde un punto de vista de protección de datos.

El RGPD consagra en su artículo 5 los principios básicos que se tienen en cuenta a la hora de realizar los tratamientos, de modo que estos seis principios (licitud, lealtad y transparencia, limitación de la finalidad, minimización de datos, exactitud, limitación del plazo de conservación, integridad y confidencialidad) unidos al de responsabilidad proactiva se convierten en el núcleo de la norma y en el objetivo que todo sistema, aplicación, servicio o proceso debe garantizar en su diseño, además de los requisitos o requerimientos funcionales a satisfacer propios del sistema.

Tradicionalmente, el diseño de sistemas seguros y confiables se centra en analizar los riesgos y dar respuesta a las amenazas que afectan a los objetivos de la seguridad que están más orientados a la privacidad:

- **Confidencialidad**, evitar los accesos no autorizados a los sistemas,
- **Integridad**, proteger de modificaciones no autorizados a los sistemas,
- **Disponibilidad**, garantizar que los datos y los sistemas están disponibles cuando es necesario.

A los objetivos de seguridad se les suman tres nuevos objetivos de protección: **control, transparencia y desvinculación**, configurar un marco global de protección en el tratamiento de los datos personales.

En las fases iniciales de concepción del objeto y del análisis de sus requisitos, hay que trabajar con **estrategias de privacidad**, enfoques genéricos a alto nivel dirigidos a identificar aquellas tácticas a seguir durante las diferentes etapas del procesamiento de los datos para garantizar los objetivos de privacidad y el cumplimiento de los principios de tratamiento. Las estrategias proporcionan un modelo accesible a través del cual los ingenieros que diseñan el objeto concretan los requisitos de privacidad identificados durante las fases de análisis y requisitos. Las estrategias de privacidad sirven de puente

entre los principios de tratamiento impuestos por la norma y la implementación de la privacidad en soluciones concretas.

Se identifican ocho estrategias de diseño de la privacidad que se conocen como 'minimizar', 'ocultar', 'separar', 'abstraer', 'informar', 'controlar', 'cumplir' y 'demostrar'.

Las primeras, al incluir las estrategias de 'minimizar', 'ocultar', 'separar' y 'abstraer' son estrategias que se orientan al tratamiento de los datos, mientras las estrategias de 'informar', 'controlar', 'cumplir' y 'demostrar' se orientan a la definición de procesos, a implementar una gestión responsable de los datos personales.

Aunque, depende del contexto, determinadas estrategias pueden ser más aplicables que otras en el marco de desarrollo de un sistema, estas ocho estrategias, se consideran desde las etapas iniciales de su concepción y análisis y aplicadas conjuntamente, permiten incorporar salvaguardas y medidas de protección en las operaciones y procedimientos de tratamiento de los datos personales y conseguir que los resultados finales tengan en cuenta los requisitos de privacidad y garantizar los derechos y libertades de los sujetos de datos. [41]

3.3.2 Evaluación de Riesgos

El RGPD no establece un criterio práctico-metodológico para la gestión de los riesgos. La gestión de riesgos en el RGPD es la protección de la persona, en su dimensión individual y social, como sujeto de los datos o afectado por el tratamiento.

En la evaluación del riesgo, hay que evaluar qué impacto puede tener para el individuo y la sociedad, ya que hay brechas cuyo impacto social hace más difícil minimizar los riesgos.

En el RGPD se utiliza el término "riesgo" como sinónimo de "amenaza" es un término bien definido en la literatura de gestión de riesgos, y en dicha literatura el "riesgo" se deriva de la probable materialización con un determinado impacto de una amenaza.

El impacto depende del daño que se pueda ocasionar a los sujetos en particular y a la sociedad en su conjunto, en el ámbito de sus derechos y libertades, a corto, medio y a largo plazo.

La AEPD determina el nivel de un riesgo específico en función de su impacto y probabilidad en el siguiente mapa de calor:

Probabilidad	Muy alta	Medio	Alto	Muy alto	Muy alto
	Alta	Bajo	Alto	Muy alto	Muy alto
	Baja	Bajo	Medio	Alto	Muy alto
	Improbable	Bajo	Bajo	Medio	Muy alto
		Muy limitado	Limitado	Significativo	Muy significativo
		Impacto			

Figura 12. Matriz Probabilidad x Impacto para determinar el nivel del riesgo [34]

En la evaluación de riesgos, se establecen las siguientes categorías de factores de riesgo:

- Operaciones relacionadas con los fines de tratamiento
- Tipos de datos utilizados
- Extensión y alcance del tratamiento
- Categorías de interesados
- Factores técnicos del tratamiento
- Recogida y generación de datos
- Efectos colaterales del tratamiento
- Categoría del responsable/encargado
- Comunicaciones de datos

En cada una de las categorías de factores de riesgo se establece el nivel de riesgo asociado a cada uno de los diferentes factores, se evalúan aspectos que se relacionan con los datos que recogen, procesan e infieren las *app´s* y se utilizan en el tratamiento por parte del responsable.

En la tabla 2 se detallan los diferentes factores de riesgo de la categoría “Tipos de datos utilizados” y el nivel de riesgo asociado a cada uno de ellos.

Por otro lado, en la categoría “Factores técnicos en el tratamiento”, se menciona como factor de riesgo las aplicaciones móviles y se establece un nivel de riesgo medio en este factor.

Factor de Riesgo	**Nivel de Riesgo**
Documentos personales, p.ej.: • Correos electrónicos • Notas de lectores de libros electrónicos	Medio
Información de aplicaciones de registro de actividades vitales	Alto
Aspectos personales, p.ej.: • Personas o grupos con los que se relaciona • Roles sociales • Gustos/preferencias de contenidos audiovisuales (televisión interactiva, plataformas de contenidos, redes sociales, …) • Cuidado de salud	Medio
Preferencias de consumo, hábitos, gustos, necesidades, etc. que no permitan inferir informaciones relacionadas con categorías especiales de datos, p.ej.: • Hábitos de consumo (tarjetas de fidelización de clientes, actividad web, …) • Preferencias de contenidos audiovisuales en diferentes medios (televisión interactiva, plataformas de contenidos, redes sociales, …)	Medio
Datos de localización, p.ej: • Registro de desplazamientos • Registro de lugares habituales	Medio
Categorías especiales de datos o que permitan inferirlos	Muy Alto
Categorías especiales de datos seudonimizados	Alto
Metadatos, p.ej.: • Datos de tráfico de las comunicaciones electrónicas • Identificación de emisor y/o receptor en las Comunicaciones • Datos en conexiones a internet: localización; características software y hardware del dispositivo con el que se conecta; redes sociales o páginas en general en las que se ha logado, conexión (IP, proveedor de servicios, velocidad de descarga)	Medio
Identificadores únicos, p.ej.: • Dirección IP • Dirección MAC • Identificadores únicos derivados de las características del dispositivo (p. e. acceso a la información de la batería de un dispositivo, id publicitario del dispositivo) • Identificadores únicos añadidos a archivos (p.e. metadatos de fotografías subidas a redes sociales)	Medio
Datos y metadatos de las comunicaciones electrónicas y datos inferidos de las comunicaciones electrónicas, p.ej.: • Correos electrónicos • Mensajes instantáneos • Llamadas de teléfonos • Video llamadas	Medio
Datos de navegación web, p.ej.: • Registro de páginas visitadas (p. e. historial de navegación, logs de servidores web) • Navegador utilizado	Medio

Tabla 2. Factores de riesgo que relacionan el ámbito del tratamiento con los datos recogidos [34]

4. TÉCNICAS Y HERRAMIENTAS

Un proceso de Ciencia de Datos convierte datos en resultados prácticos, convierte un problema en una solución y trata el problema de negocio como un proyecto BI (*Business Intelligence*), lo cual facilita extraer, perfilar, transformar, almacenar, administrar y visualizar los datos de modo eficaz.

4.1. Analítica de Big Data

La analítica de datos, "implica los procesos y las actividades que se diseñan para obtener y evaluar datos y así extraer información útil" (ISACA[11]).

Las técnicas que más se emplean en analítica de datos son: consultas e informes, visualización de datos, Minería de datos, análisis predictivo de datos, lógica difusa, optimización, *streaming* de audio, vídeo o fotografía.

Se examinan datos en bruto con el propósito de obtener conclusiones, responder a objetivos, con la información que contienen. La relevancia de los datos en bruto no es aceptable hasta que no se contextualizan y procesan en información útil. **La analítica es el proceso de extracción y creación de información a partir de datos en bruto mediante la filtración, procesamiento, categorización, contextualización, análisis y visualización de datos**.

La información se organiza y estructura para inferir conocimiento relativo al sistema y/o usuarios, su entorno y sus operaciones, esto constituye a los sistemas más inteligentes y eficaces.

Hay diferentes tipos de analítica de datos: descriptiva, predictiva, de diagnóstico, de descubrimiento, aumentada y prescriptiva.

4.2. Analítica Descriptiva

La analítica descriptiva es la que más se utiliza y su objetivo es realizar una instantánea de la situación para tomar decisiones con un alto grado de éxito. Utilizar datos históricos, identificar comportamientos y dibujar cómo se hacen las cosas. Se consultan diferentes *indicadores de negocio* para obtener una visión de lo que ha pasado y está pasando.

[11] ISACA: Asociación de Auditoría y Control de Sistemas de Información

Preparar y analizar datos históricos e *identificar patrones y tendencias*. Lograr un profundo conocimiento a partir de cuadros de mandos, agrupaciones, informes. Mediante un análisis descriptivo **responder a la pregunta ¿qué pasó?**

La analítica descriptiva utiliza técnicas como: *modelos de regresión, modelados de datos y visualización de datos*.

4.3. Analítica Predictiva

La analítica predictiva realiza la **creación de modelos** que permiten vaticinar lo que va a ocurrir con antelación. Busca obtener conocimiento en forma de patrones, modelos o tendencias que ayuden a acertar con situaciones futuras. Responder a la pregunta **¿qué pasará?, ¿qué va a ocurrir?**

En la analítica predictiva se utilizan modelos predictivos que se entrenan con datos existentes. **Estos modelos aprenden patrones y tendencias de los datos existentes y predicen la probabilidad de un suceso o el resultado probable de un evento** (modelo de clasificación) o números previsibles (*forecast*, modelos de regresión).

Se basa en métodos matemáticos avanzados de *estadística o el Aprendizaje Automático* y así predecir los datos que faltan y describir lo que va a suceder.

4.4. Metodología a seguir en un proyecto BI

Existen numerosas metodologías para realizar un proyecto BI, así como el ciclo de vida correspondiente. Normalmente todas ellas se componen de diferentes fases y coinciden en gran medida, en el número y nombre de las fases, así como en sus contenidos.

Las metodologías ofrecen diferentes etapas de forma según la fuente, los títulos de las etapas pueden variar, así como los contenidos internos de cada etapa. De cinco a seis etapas son el número que más se acepta en las diferentes metodologías:

- Etapa 1. Comprender el negocio
- Etapa 2. Comprender los datos
- Etapa 3. Preparar los datos
- Etapa 4. Modelar
- Etapa 5. Evaluar
- Etapa 6. Desplegar

4.4.1 Comprender el ¿negocio?

Corresponde a la fase inicial del proyecto BI, se concentra en comprender los objetivos del proyecto y en definir las necesidades del cliente.

Busca convertir el conocimiento de los datos en la definición de un problema de Minería de Datos y en un plan preliminar que se diseña para alcanzar los objetivos, y entender los objetivos del estudio y requisitos del proyecto, desde una perspectiva del estudio y no técnica.

Cada dominio y estudio funciona con un conjunto de reglas y objetivos. Para adquirir los datos correctos, se debe entender el estudio a realizar. Hacer preguntas sobre el conjunto de datos ayuda a adquirir datos de forma correcta.

De manera similar al conocimiento previo del área del estudio del proyecto, se recopila el conocimiento previo en los datos. Comprender cómo recopilar, almacenar, transformar, informar y utilizar los datos para el proceso.

Se examinan todos los datos disponibles para responder a la pregunta u objetivo principal del estudio del proyecto y conseguir así los nuevos datos a obtener. Se tienen que valorar aspectos como la calidad de los datos, volumen de datos, disponibilidad de los datos, datos perdidos, o si la falta de datos obliga a cambiar alguno de los objetivos.

Como resultado de este paso se obtiene un conjunto de datos para responder a los objetivos.

El modelo inferido es tan bueno como los datos que se utilizan para crearlo.

4.4.2 Comprender los datos

En la fase de compresión y estudio de los datos recopilar y familiarizarse con los datos, descubrir conocimiento preliminar sobre esos datos, identificar los problemas de calidad y analizar las primeras potencialidades, y/o descubrir subconjuntos interesantes para dar respuesta a los objetivos que se plantean como pueden ser las diferentes segmentaciones.

En la fase de recolección se ha de responder a la procedencia. Los datos provienen de numerosas fuentes: registros de servidores web, datos de repositorios en línea, datos de bases de datos, datos de redes sociales, datos de sensores, datos de la Internet de las cosas. Elegir las fuentes con los objetivos presentes de la fase de comprensión del estudio de proyecto.

En general, esta fase tiene cuatro actividades:

1. Recolectar datos iniciales. Adquirir los datos de diferentes fuentes y cargar en la herramienta de análisis.
2. Descubrir los datos. Examinar los datos y documentar sus propiedades fundamentales como: formato de datos, número de registros o identidades de los campos.
3. Explorar los datos. Profundizar en los datos, consultar, visualizar e identificar las relaciones entre ellos.
4. Verificar la calidad de los datos. Los datos son limpios o "sucios", documentar todas sus características.

4.4.3. Preparar los datos

La fase de preparación de datos se conoce como *preprocesamiento*, consta de las actividades para construir el conjunto final de datos a partir de los datos en bruto iniciales. Incluir la selección de tablas, registros y atributos, así como la transformación y la limpieza para las herramientas con las que modelar los datos. Preparar el conjunto de datos y se adapte al estudio del proyecto, es la parte del proceso que consume más tiempo, suele llegar al 80% del tiempo del proyecto completo.

4.4.4. Modelar los datos

En esta fase de modelación de los datos nos dedicamos a seleccionar y aplicar las técnicas de modelado de los datos, calibrar sus parámetros a valores óptimos, definir métricas o KPI's, obtener requisitos específicos sobre la forma de los datos. El objetivo de esta fase es obtener el *data mart* del proyecto (datos que se relacionan con un aspecto del negocio o del estudio que se relacionan con los objetivos y sirven de base para implementar la aplicación BI).

4.4.5. Evaluar

Hasta aquí se construyen uno o varios modelos con una calidad suficiente desde la perspectiva del análisis de datos. Antes de proceder al despliegue final del modelo, es importante la evaluación y comparación del modelo obtenido con los objetivos del negocio o estudio. La obtención de resultados es la fase final.

4.4.6. Desplegar

El objetivo del modelo normalmente es aumentar el conocimiento de los datos. Este conocimiento se organiza y presenta para que el cliente pueda usarlo.

Al depender de los requisitos, en esta fase se generan uno o más cuadros de mando. El objetivo de esta fase es la distribución y puesta en producción, explotar la potencialidad de los modelos e integrar en los procesos de toma de decisiones de la organización.

La metodología a utilizar en el proyecto de BI es la del *ciclo de vida de Kimball* se representa en la Figura 13.

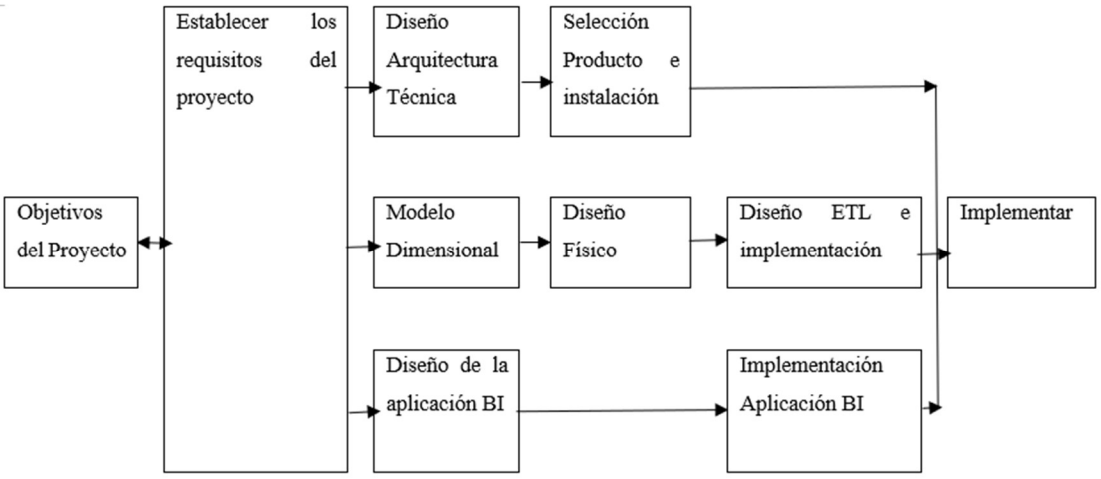

Figura 13. Esquema metodología del ciclo de vida de *Kimball* de un proyecto BI [20]

4.5. Visualización de datos

La analítica Big Data solo es útil al garantizar que las personas a quienes se va a destinar obtienen la información adecuada, en el formato adecuado y el momento en que se necesita y así tomar las decisiones correctas de la forma más eficiente posible.

Al proliferar las aplicaciones de Big Data en la Inteligencia de Negocios facilitan la creación de una alta gama de herramientas de visualización, permiten presentar los resultados de los análisis de los datos con un aspecto más atractivo y comprensible.

Se consolida la tendencia de integrar los datos en formatos sencillos dentro de una misma presentación y mezclar todo tipo de datos. Hoy en día, estas soluciones muy eficaces se obtienen mediante herramientas de visualización se conocen como **cuadros de mando** (*dashboard*), es decir, visualizar en una única pantalla gráficos y diagramas o tablas, métricas o indicadores clave de rendimiento, ayudar a la gestión de una empresa.

En síntesis, un cuadro de mando ayuda a conseguir los objetivos de la inteligencia y de la analítica de negocios: transformar los datos en información, la información en conocimiento, y el conocimiento en mejores decisiones de los empleados.

4.5.1. Características de un Cuadro de Mando

El cuadro de mando para tomar decisiones consta de diferentes elementos: métricas y KPI's, gráficos, tablas, diagramas, mapas, infografías, alertas visuales y todo se combina de manera precisa y uniforme. Disponer de un menú de navegación y en su caso una infografía. Esta información se basa~~,~~ en los KPI del negocio y/o estudio, así como en las tendencias de negocio:

- **Indicadores clave de rendimiento KPI adecuados**. Realizar una búsqueda exhaustiva de los indicadores, añadir valor al negocio y/o estudio.
- **Presentación visual**. Los cuadros, gráficos, tablas y restante información sean ilustrativos, claros y estén bien estructurados.
- **Datos comprensibles y accionables**. Contextualizar los datos para compararlos e interpretarlos, permitir establecer valores útiles.
- **Personalizado**. Ajustar el cuadro de mando a los objetivos de cada negocio y/o estudio, es decir, no es estándar para todas las empresas y estrategias funcionales. Elaborar una presentación a medida, en función de los objetivos.

Los cuadros de mando son herramientas de administración del rendimiento empresarial, presentan ante los usuarios una visualización de los indicadores empresariales.

Los usuarios *exploran y visualizan datos masivos* mediante *gráficos interactivos*, Cuadros de Mando Integral, paneles de Control o cuadros de mando y visualizan informes de resultados en tiempo real [20].

4.6. Herramientas de Big Data

La analítica de Big Data y sus herramientas permiten a los usuarios analizar los datos masivos de un modo rápido y económico.

4.6.1. Fuentes de datos

Una de las fuentes de datos masivos para la analítica de datos es el portal ***Kaggle*** [46]. Se trata de un sitio web para compartir conocimiento, disponer de *Datasets*, es decir, conjuntos de datos públicos de alta calidad para explorar y analizar datos.

4.6.2. Almacenamiento de los datos

En el contexto de los datos masivos aparecen nuevas necesidades, las formas de procesar y almacenar la información cambian. Lo más frecuente es tratar con datos no

estructurados o semiestructurados, en grandes volúmenes y la consistencia de los datos deja de ser un requisito rígido y puede ser flexible, prevalece la veracidad de los datos por encima de un esquema fijo para el modelado de datos.

En este nuevo contexto nace la necesidad de una nueva clase de base de datos, denominada bases de datos *NoSQL*, convertir el procesamiento distribuido en una característica de primer orden y surgen conceptos como los clústeres, la réplica de datos y el escalado horizontal.

4.6.2.1. Base de datos *NoSQL: MongoDB*

Una de las bases de datos *NoSQL* que se utilizan es *MongoDB*, se crea por la compañía *10gen* en el año 2007. Se caracteriza por almacenar los datos en documentos de tipo *JSON* con un esquema dinámico que se denomina "*BSON*".

- **Documentos**

  Los documentos son la unidad básica de organización de la información en *MongoDB*, y desempeñan un papel equivalente a una fila en las bases de datos relacionales.

- **Tipos de datos**

  Los documentos en *MongoDB* soportan tipos de datos como: Nulo (representar el valor nulo o bien un campo que no existe. Por ejemplo {"x":*null*}), Booleanos (valores true o false), Números (distinguir entre números reales y números enteros), Cadenas (cualquier cadena de caracteres), Fechas (almacenar la fecha en milisegundos), Arrays (representar un conjunto o lista de valores), Documentos embebidos (pueden contener documentos embebidos como valores de un documento padre) e Indicadores de objetos (es un identificador de 12 bytes para un documento) entre otros.

  Cada documento tiene una clave se denomina "_id", es de tipo *ObjectId* y valor único. Al insertar un documento, si no tiene un valor para la clave "_id", se genera automáticamente una por *MongoDB*.

- **Colecciones**

  Una colección es un grupo de documentos, desempeña el papel análogo a las tablas en las bases de datos relacionales. Dentro de una colección puede haber cualquier número de documentos con diferentes estructuras [45].

Las aplicaciones de BI se realizan con un gran número de herramientas disponibles en código abierto (*open source*), en plataformas en la nube o software de pago.

De acuerdo con la categoría de aplicación BI se establece la siguiente clasificación:

- **Minería de Datos y análisis de datos**: *Knime, Google Analytics, RapidMiner*
- **Visualización de datos**: *Tableau, Qlik Sense[65], Power BI y Excell*
- **Lenguaje de programación**: *Python, Jupyter [6] (Notebook), JavaScript, SQL, MATLAB, Databriks* [23]
- **Bibliotecas de *Python***: *Matplotlib, NumPy, Pandas, Scipy, NTLK, Scikit-Learn, Seaborn*
- **Marcos de trabajo** (*frameworks*): *Apache Hadoop y Apache Spark*

Figura 14. Cuadrante mágico de *Gartner* de plataformas de Analítica y de Inteligencia de Negocio. (Mayo 2025) [22]

Se puede observar en la Figura 14 herramientas líderes como *Microsoft Power BI*, *Tableau(Salesforce)* y *Qlik (View y Sense)* entre otras.

5. IMPACTO EN LA PRIVACIDAD DE LAS APP´S DE ACUERDO CON LA AEPD

Vamos a construir el proyecto BI a partir de la metodología del ciclo de vida de *Kimball*, cuyas principales fases se comentan en el apartado 4 con técnicas y herramientas de la presente Memoria.

5.1. Objetivos del proyecto

Al inicio del proyecto se establece el objetivo del proyecto BI: ayudar a la toma de decisiones a los programadores a la hora de construir una *app* desde el diseño, conocer aquello que les puede aportar la recopilación de los datos, conocer la intromisión en la privacidad y su anonimización de acuerdo con el marco legislativo en la Unión Europea: Reglamento General de Protección de Datos y, a nivel nacional, la ley orgánica 3/2018 de Protección de Datos Personales y Garantía de los Derechos Digitales.

5.2. Requisitos del Proyecto

Una de las partes importantes de la seguridad de las aplicaciones móviles para asegurar la privacidad de los datos personales son los permisos declarados en la aplicación, en concreto en el fichero *AndroidManifest.xml*.

Ha de considerarse aspectos como, la toma de decisiones, establecer buenas prácticas en el uso de los permisos por los programadores, concienciar de la importancia de incluir la privacidad desde la fase de diseño de las *app´s*, tener presente el marco legislativo y contribuir con ello a minorar los riesgos a la seguridad que se detectan en la auditoría realizada en el año 2024 por OWASP en las aplicaciones móviles de Android, todos estos aspectos constituyen objetivos del proyecto BI.

Partiendo del marco que se declara para el desarrollo del Proyecto BI, se contemplan los objetivos y se identifican requisitos.

Requisitos
Conocer el grado de intromisión a la privacidad de las aplicaciones móviles.
Conocer si los usuarios cuando valoran una *app*, consideran parte importante o no la privacidad.
Conocer si las aplicaciones más intrusivas a la privacidad son de las más descargas por los usuarios.
Comprobar si al entrar en vigor el RGPD cambia el uso de los permisos en las *app´s* por parte de los programadores.
Saber si utilizan las *app´s* los permisos necesarios de acuerdo a su funcionalidad.
Conocer el comportamiento de los programadores y de los usuarios que hacen uso de las *app´s* ante la privacidad

Conocer si la intromisión en la privacidad por parte de las *app´s* es la misma en todo tipo de público o no.
Conocer si las *app´s* de pago son menos intrusivas a la privacidad respecto a las *app´s* gratuitas o no.
Ver la tendencia a lo largo del tiempo, conocer las *app´s* más intrusivas en cuanto a la privacidad.
Conocer la intromisión en datos personales por parte de la *app*
Conocer el riesgo de acceso a información de registro de actividades vitales por la *app*
Descubrir el acceso a preferencias de consumo, hábitos, gustos, necesidades del usuario y el riesgo asociado a ello
Conocer el riesgo del acceso a aspectos personales de los usuarios por la *app*
Descubrir el riesgo de acceso a la ubicación del dispositivo donde está instalada la *app*
Conocer el riesgo de acceso a metadatos por parte de las *app´s*
Analizar el riesgo del acceso a identificadores únicos de los dispositivos electrónicos
Analizar el riesgo del acceso a datos y metadatos de las comunicaciones electrónicas
Conocer el riesgo de acceso a datos de navegación

Tabla 3. Requisitos del proyecto de BI

5.3 Diseño de la arquitectura

En el diseño de la arquitectura se establecen los requisitos del proyecto y se responde a qué es lo que se necesita hacer, respondiendo con la arquitectura a la pregunta de cómo se va a realizar.

En su diseño, se establece las fuentes de los datos, su almacenamiento y que datos se necesitan en el proyecto.

5.3.1. Fuentes de Datos

A continuación, se consultan posibles fuentes de datos con un volumen suficiente de datos y veraces a utilizar en el proyecto.

Realizando búsquedas en internet mediante los términos: *permission*, permiso, android, *app´s*, *ios*, aplicaciones móviles, se obtienen un total de 12 posibles fuentes de datos para el proyecto, no se incluyen en el *Data Lake* algunas de las fuentes de datos [15] [27] [50] [52] se considera no apropiado el volumen de datos, muestran conjuntos de datos repetidos o no disponen de atributos que se consideran importantes en el proyecto.

Se alojan las fuentes de datos [9] [10][14] [36] [37] [38][51] [54] en el *Data Lake* y se analizan de forma más amplia su contenido, determinando si se utilizan o no en el proyecto BI:

- ***AndroidPermission*** [9]: conjunto de datos, con un volumen de 54.34 MB, en un fichero "*Extracted_permssion_updated.csv*", Aplicaciones y permiso para

Android. Enlace de datos https://www.kaggle.com/code/saurabhshahane/android no disponible.

- ***AppAndroid*** [14]: conjunto de datos, con un volumen de 38.97 MB, en dos ficheros "*Android app´s Excel.xlsx*" (12.14 MB) y "*Android app´s csv.csv*" (26.83 MB), registros de teléfonos con Sistema Operativo Android (a partir de 14.000 usuarios).
- ***AppPermission*** [51]: conjunto de datos, con un volumen de 145.95 KB, en un fichero "*extracted_permission_updated.csv*", lista completa de aplicaciones móviles junto con sus respectivas categorías, recuentos de descargas y uso de permisos.
- ***AppPermissionGoogleplay*** [10]: conjunto de datos, con un volumen de 1.6 GB, en un fichero "*googleplay-app-permission.json*", los datos se respaldan en el repositorio https://github.com/gauthamp10/android-permissions-dataset. Corresponden a *app´s* alojadas en *Google Play Store*.
- ***GooglePlayStore*** [37]: conjunto de datos, con un volumen de 676.46 MB, en un fichero "*Google-Playstore.csv*", los datos se respaldan en el repositorio https://github.com/gauthamp10/Google-Playstore-Dataset. Disponen de datos de *app´s* alojadas en *Google Play Store*. Los datos se recopilan en el año 2021.
- ***MobileAppPermission*** [36]: conjunto de datos, con un volumen de 2.42 GB, se dispone de dos ficheros: "*Permissions_all.csv*" (1.49 GB) y "*with_perm_cols.csv*" (933 MB). No proporcionan información de la procedencia de los datos que se alojan en los diferentes ficheros.
- ***NaticusDroidPermission*** [54]: conjunto de datos, con un volumen de 4.9 MB, se dispone de un fichero "*data.csv*", contiene datos para la detección de malware mediante los permisos que se declaran en las *app´s*.
- ***PlayStoreGoogle*** [38]: conjunto de datos, con un volumen de 9.03 MB, se dispone de los ficheros "*googleplaystore.csv*" (1.36 MB) y *googleplaystore_user_reviews.csv* (7.67 MB), contienen datos de *app´s* que se alojan en *Google Play Store*.

Se determina alojar los datos de las diferentes fuentes en una base de datos *NoSQL* y se realiza el análisis observando los atributos que las componen.

Se opta por dicho alojamiento al obtener los datos en ficheros con formato *csv* y *json* y se permite importar tanto datos estructurados, como no estructurados.

5.3.2 Almacenamiento de los datos

Se almacenan los datos en el *Data Lake*, en una base de datos *NoSQL MongoDB*, creando una colección para cada una de las fuentes de datos seleccionadas y se alojan los datos de los diferentes ficheros en cada una de ellas.

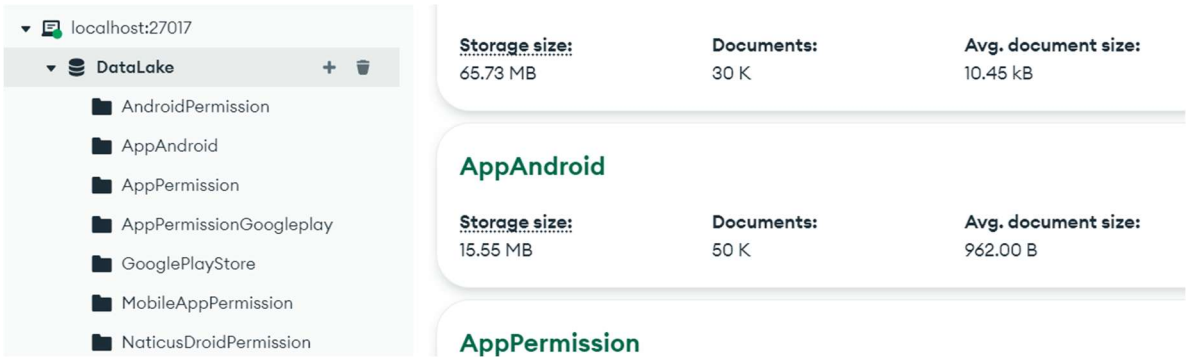

Figura 15. Diferentes fuentes de datos se alojan en *MongoDB (NoSQL)*

Las fuentes de datos que se incluyen en el *Data Lake* son posibles candidatas para el proyecto BI, se realiza un ligero perfilado de los datos y se determina incluir las fuentes de datos ***AppPermissionGoogleplay*** y ***GooglePlayStore*** en el *data warehouse*.

Los dos *dataset* contienen un volumen de datos por encima de 1 millón de registros y contienen datos de calidad, se descargan de *Kaggle.*

Contienen datos de las mismas *app´s,* es decir, la misma *app* aparece en los registros de ambos *dataset* pero con información diferente acerca de ella, se puede realizar una fusión de ambos *dataset* y así se dispone de los datos para el desarrollo del proyecto BI.

Se pueden consultar los datos de las fuentes de datos que se utilizan en el proyecto BI, *AppPermissionGoogleplay* y *GooglePlayStore* en el apéndice C.

5.3.3 Analítica de datos del proyecto BI

Con los datos de las fuentes que se eligen para realizar el proyecto BI, se crea una nueva base de datos "*Data Warehouse*" en *MongoDB* y se alojan los datos de los *dataset* en ambas colecciones: *AppPermissionGooglePlay* y *GooglePlayStore*.

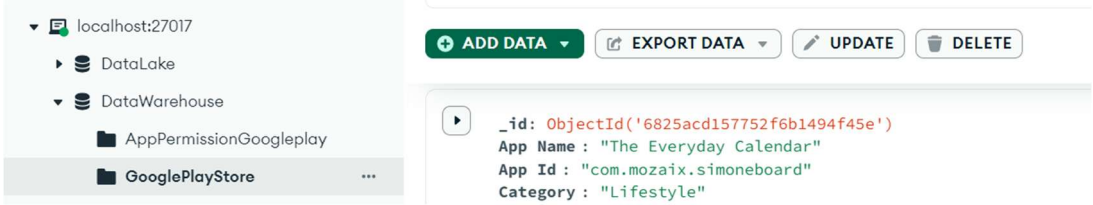

Figura 16. Colecciones de datos del proyecto BI en *MongoDB (NoSQL)*

Se piensa en el atributo del "Nombre" de la *app* como dato único en ambas colecciones, pero se comprueba la existencia de *apps* con el mismo nombre y diferente Id.

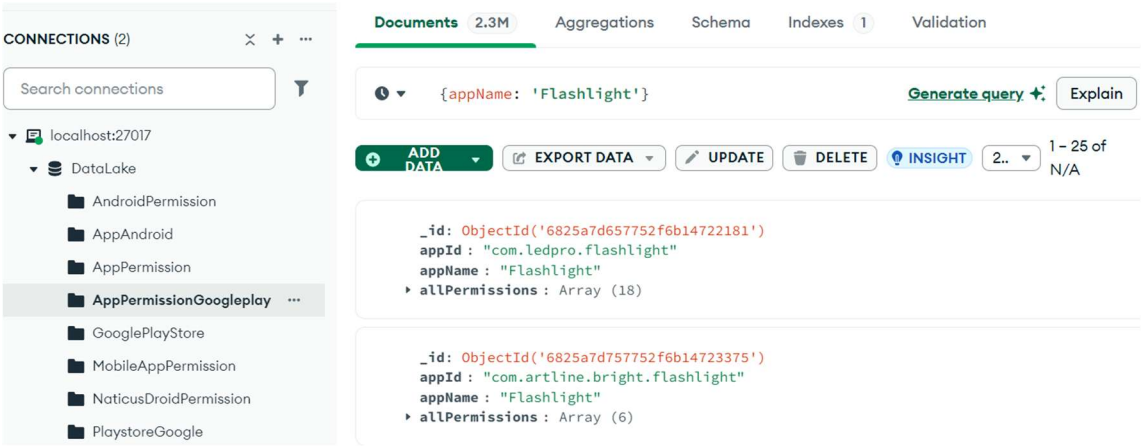

Figura 17. Se constata la duplicidad en los nombres de las aplicaciones

Se identifica el dato "App Id" y "appId" como el dato que se utiliza en la construcción del *join*[12] de ambas colecciones, se obtiene información relevante acerca de los requisitos del proyecto BI.

En la colección ***AppPermissionGoogleplay*** se tiene más de 2.3 millones de instancias y en la colección ***GooglePlayStore*** se dispone de 2.2 millones de instancias al realizar el *join* de ambas colecciones, se dispone de un volumen de datos de al menos 2.1 millones de instancias.

5.4 Modelo dimensional

El modelado dimensional es una técnica de diseño lógico permite estructurar los datos de forma que puedan dar respuesta a los objetivos del proyecto.

El modelo dimensional divide el mundo en mediciones y contexto. Utiliza lo que se conoce como hechos, referencian a qué se va a medir, y las dimensiones indican cómo se va a medir.

Los hechos son los valores con los que se puede calcular y permiten obtener las métricas o indicadores a partir de los requisitos del proyecto.

En cuanto a las dimensiones, son atributos de valor discreto y se utilizan para filtrar o segmentar los datos del proyecto BI y proporcionan información relacionada con los requisitos del proyecto, se utilizan para filtrar la información que se muestra en la aplicación BI, un cuadro de mando.

[12] Join: es una operación fundamental en SQL que permite combinar datos de diferentes tablas o colecciones

Así, en el proyecto BI se dispone de los siguientes hechos (se encuentran descritos en el apéndice C):

- appId
- Rating
- Installs
- Minimum Android
- Free
- Developer Website
- Developer Email
- Privacy Policy
- Type

Type (consta de 16 distintos valores en el conjunto de datos, ver tabla 27)

Y las dimensiones:

- Category
- Developer Id
- Released
- Content Rating

La dimensión Fecha constituye una jerarquía con los datos del año, mes y día.

5.4.1 Métrica y KPI

En este punto del proyecto, se establecen métricas y KPI's partiendo de la tabla de requisitos del proyecto BI (ver tabla 3), se define el indicador, forma de calcularlo, se indica si hay alguna excepción, se mencionan las fuentes de datos que se utilizan y los artículos del RGPD que se relacionan con la medición.

Se establecen los siguientes KPI's:

- Nº de grupo de permisos de datos sensibles (VPDS)
- Nº de grupo de permisos de alto riesgo o permisos peligrosos (VPP)
- Nº de grupo de permisos de red (VPR)
- Nº de grupo de permisos de acceso al almacenamiento (VPA)
- Nº de grupo de permisos de acceso a la ubicación (VPU)

Por último, pero no por ello menos importante, se establece una métrica a partir de los grupos de permisos peligrosos que se declaran en las *app´s* según su funcionalidad.

Utilizando el estudio *Overprivileged Permission Detection for Android Applications* [55] se establecen los grupos de permisos peligrosos que utilizan las *apps* de acuerdo con la funcionalidad o categoría que proporcionan al usuario. Clasifican 17 categorías de las 48 disponibles en los datos del proyecto BI.

Se define la métrica:

- Nº de grupos de permisos innecesarios por categoría (VPNC)

En el apéndice D se muestra el detalle de cada uno de los KPI's.

Todos los indicadores se declaran para cada una de las segmentaciones, es decir, categoría, tipo de público, programador y fecha de lanzamiento. Se utiliza la misma fórmula de cálculo, pero se agrupa por cada segmentación o dimensión.

5.4.1.1 Evaluación de Riesgo

La AEPD establece y detalla factores de riesgo en función de los tipos de datos a utilizarse (ver tabla 2).

Partiendo de los grupos de permisos (ver tabla 1) se declaran diferentes KPI's permite medir los diferentes factores de riesgo que se relaciona con los datos, permite medir el cumplimiento normativo y permite una aproximación inicial a la evaluación del riesgo y su impacto (ver en la figura 12).

Factor de Riesgo	Nivel de Riesgo	KPI
Documentos personales	Medio	VPA
Información de aplicaciones de registro de actividades vitales	Alto	VPA, VPP
Aspectos personales (Perfilado)	Medio	VPA, VPU
Preferencias de consumo, hábitos, gustos, necesidades, etc. que no permitan inferir informaciones relacionadas con categorías especiales de datos	Medio	VPDS, VPA, VPU
Datos de localización	Medio	VPU
Datos muy personales no recogidos en clasificaciones anteriores	Alto	VPP
Categorías especiales de datos seudonimizados	Alto	VPDS
Metadatos	Medio	VPP, VPA
Identificadores únicos	Medio	VPP
Datos y metadatos de las comunicaciones electrónicas y datos inferidos de las comunicaciones electrónicas	Medio	VPR
Datos de navegación web	Medio	VPR, VPA

Tabla 4. Factores de riesgo que se asocian a los tipos de datos y KPI's para su medición

Asociando la tabla de factores de riesgo de los tipos de datos con los KPI's se conoce el nivel de riesgo al utilizar las *app´s* los grupos de permisos y con ello el impacto en el contexto para el usuario y la sociedad en general.

Observando los diferentes cuadros de mando donde se visualizan los KPI's, y la tabla 4 se conoce el nivel de riesgo que se asocia a los diferentes KPI's y con ello el impacto de las *app´s* en la contextualización del marco legislativo a partir de los diferentes grupos de permisos que se declaran en su desarrollo. (ver figura 12)

5.4.2. *Profiling* o perfilado de los datos

Siguiendo la metodología del proyecto BI, a continuación se realiza el perfilado de los datos, para determinar si se dispone de los datos que se utilizan en la obtención de los indicadores de rendimiento o KPI's.

Hay que realizar un perfilado de los datos en profundidad con la intención de detectar tantos problemas como sean posibles respecto a los datos elegidos.

Se utiliza la herramienta de *Jupyter Notebooks* y se realizan consultas de los datos mediante el lenguaje de programación *Python*.

Se construye el *dataframe* a partir de la colección ***GooglePlayStore*** alojada en el *Data Warehouse* y se analizan principalmente los datos de los atributos correspondientes a los hechos y dimensiones del proyecto y así poder dar respuesta a los requisitos.

Se obtienen de todos ellos información relevante en cuanto a su contenido, se dispone de un volumen de datos de cada una de las segmentaciones o filtros por lo cual se decide su inclusión en el diseño del modelo conceptual dimensional.

Se construye un segundo *dataframe* a partir de la colección ***AppPermissionGoogleplay***, se constata la calidad y volumen de los datos respecto a los grupos de permisos de las *app´s*. Se observan las nomenclaturas de los grupos de permisos y si aparecen grupos de permisos repetidos en las *app's*.

En el apéndice E. *Analítica de datos* se amplía la información, se detalla en profundidad el contenido de los datos de ambas colecciones.

5.4.3 Modelo en estrella

Habiendo perfilado los datos con los cuales se obtienen las diferentes métricas definidas a partir de los requisitos del proyecto BI, se construye el **modelo dimensional** con los **hechos y dimensiones**.

Los hechos corresponden a los datos que se utilizan en la obtención de las diferentes métricas tanto a nivel individual como a nivel en conjunto de las *app's*. Los hechos a incluir se mencionan en el apartado 5.4

A continuación, se obtienen las dimensiones a partir de los diferentes filtros o segmentaciones de los datos. En el proyecto BI se definen las dimensiones: categoría (*category*), tipo de público (*Content Rating*), programador (*developer Id*) y fecha de lanzamiento (*released*).

El resultado del modelo dimensional del proyecto BI, constituye la base del almacén de datos, se muestra en el apéndice F.

5.4.4 Diseño físico

Mediante el modelo dimensional de datos, con la tabla de hechos y las tablas de dimensiones, se realiza el diseño lógico de datos. Se define cada una de las tablas a alojar en la base de datos a partir de la tabla de hechos y las tablas de dimensiones obteniendo el modelo físico de los datos.

Con las fuentes de datos del proyecto BI (*AppPermissionGoogleplay* y *GooglePlayStore*) se diseña el modelo físico con las tablas: **FactPrivacidad, DimCategoria, DimTipoPublico, DimFecha y DimProgramador**. Con todo esto, se declara la naturaleza/Tipo e indicador de cada dato en origen: esto conlleva el detalle de cada uno de los atributos que se utiliza en el proyecto BI en las mencionadas tablas, se establecen las claves foráneas en referencia a tablas que se utilizan en el proceso de ETL. Ver en el apéndice G el detalle de las tablas.

En este paso se obtiene el **modelo físico** y se utiliza como entrada en el proceso de ETL.

5.4.5 ETL

El proceso ETL (Extract, Transformation and Load) o proceso de extracción, transformación y carga de los datos, consume alrededor del 80% del total del tiempo de la construcción del proyecto BI.

Se parte de los datos originales de las fuentes de datos, se realizan transformaciones y finalmente se cargan los datos en tablas permitiendo realizar consultas al usuario. Una vez realizado el proceso ETL se obtienen los datos que se utilizan en la implementación de los cuadros de mando, es decir, permiten construir los cuadros de mando donde se visualizan los KPI's que se calculan a partir de dichos datos.

Con los resultados del perfilado de los datos y el análisis con las herramientas de *Jupyter Notebooks* y *Databriks*, (ver apéndice C y apéndice E) se establecen las

transformaciones necesarias en los datos antes de que se alojen en el *Data Mart* y así se obtienen resultados mediante la aplicación BI.

Se construye el *Data Mart* de Privacidad con 5 ficheros CSV: *Privacidad, Categoría, TipoPúblico, Programador y Fecha* constituyen el destino de los datos. El detalle con el contenido de cada uno de los ficheros se muestra en el apéndice H: identificador de cada dato, la naturaleza/tipo y si las tablas son hechos o dimensiones.

Se disponen los datos de las diferentes tablas y ficheros del *Data Mart*, y se procede a detallar la relación del dato origen – dato destino e indicar la tabla de origen de la cual provienen para cada uno de los ficheros *csv* (ver apéndice H).

Se obtienen los diferentes ficheros CSV, constituyen el *Data Mart* de Privacidad a utilizar en la aplicación BI: se visualizan los diferentes KPI's y se obtiene información acerca de los objetivos del proyecto de BI.

Calidad de los datos

En el siguiente paso se contrasta la calidad de las Fuentes de datos del proyecto BI. (Ver apartado 5.3.1)

Limpieza de los datos

Se realiza el análisis de los datos mediante un perfilado. Están limpios los atributos que se utilizan en el proyecto BI, si bien hay información a desechar al no ser utilizada en la aplicación BI como son el atributo "permission" en la colección *AppPermissionGoogleplay* y los datos duplicados correspondientes a los mismos grupos de permisos en las *app`s* se eliminan. En la colección *GooglePlayStore* se desechan aquellos atributos que no se incluyen en los hechos y dimensiones del proyecto BI (ver apéndice C y apéndice E.)

5.5 Diseño de la aplicación BI

En el diseño de la aplicación BI se constituye un ciclo analítico, se tiene en mente cómo va a acceder el usuario a la información del *Data Mart* y pueda tomar decisiones a partir de los datos.

Al inicio de la aplicación BI se dispone de una página con un menú de navegación para el acceso a los cuadros de mando y una infografía donde se muestra información relevante acerca de la evaluación del impacto en el contexto por parte de las *app's* de acuerdo al marco legislativo y los correspondientes KPI's que se utilizan en su medición y que se visualizan en los cuadros de mando que constituyen la aplicación BI. (Ver figura 18 y 19)

AEPD. Nivel de riesgo e Impacto en la privacidad por las *apps*

Probabilidad					
	Muy alta	Medio	Alto	Muy alto	Muy alto
	Alta	Bajo	Alto	Muy alto	Muy alto
	Baja	Bajo	Medio	Alto	Muy alto
	Improbable	Bajo	Bajo	Medio	Muy alto
		Muy limitado	Limitado	Significativo	Muy significativo
		Impacto			

Factor de Riesgo	Nivel de Riesgo	KPI
Documentos personales	Medio	VPA
Información de aplicaciones de registro de actividades vitales	Alto	VPA, VPP
Aspectos personales (Perfilado)	Medio	VPA, VPU
Preferencias de consumo, hábitos, gustos, necesidades, etc. que no permitan inferir informaciones relacionadas con categorías especiales de datos	Medio	VPDS, VPA, VPU
Datos de localización	Medio	VPU
Datos muy personales no recogidos en clasificaciones anteriores	Alto	VPP
Categorías especiales de datos seudonimizados	Alto	VPDS
Metadatos	Medio	VPP
Identificadores únicos	Medio	VPP
Datos y metadatos de las comunicaciones electrónicas y datos inferidos de las comunicaciones electrónicas	Medio	VPR
Datos de navegación web	Medio	VPR

Figura 18: Infografía: AEPD. Nivel de riesgo e Impacto en la privacidad por las *app´s*

5.5.1 Diseño Cuadros de Mando

Los cuadros de mando se construyen a partir del patrón elegido en forma de Z, es decir, se distribuye la información más relevante de acuerdo a la forma del patrón, se consideran zonas más visibles y por lo tanto el usuario se detiene más en dichas partes del cuadro de mando.

Se sitúan los diferentes elementos: KPI's, gráficos, tablas y filtros correspondientes a las diferentes dimensiones del almacén de datos en un marco, constituye lo que se conoce como *mockup* o prototipo de diseño del cuadro de mando.

Los cuadros de mando se diseñan con un fondo en tono gris y se perfila cada uno de los diferentes componentes con el tono anaranjado que utiliza la AEPD en sus documentos, se consideran representativos del proyecto BI, de la privacidad.

Se detalla el diseño del cuadro de mando, se modifica el contenido y se obtiene información acerca de los diferentes requisitos del proyecto BI.

Se diseñan 4 cuadros de mando:

- Comportamiento del usuario ante la Privacidad
- Comportamiento del Programador ante la Privacidad
- Cuidado de la Privacidad en el tiempo por las *app´s*
- Cuidado de la Privacidad por las *app´s* según su funcionalidad

Se muestra a continuación el diseño de cada uno de los cuadros de mando:

Cuadro de Mando: Comportamiento del usuario ante la Privacidad

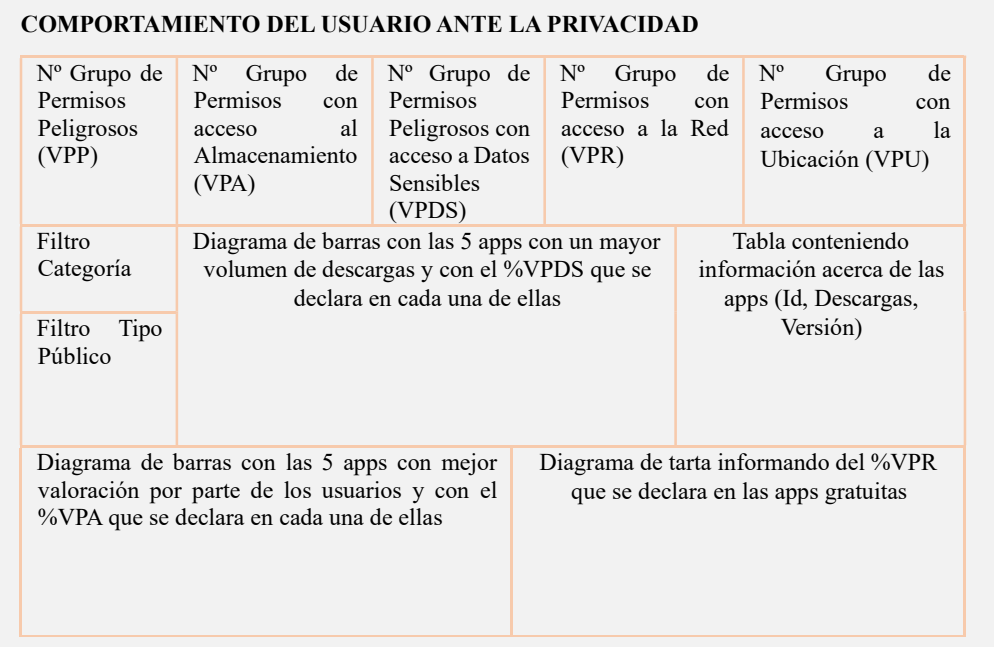

Figura19. Prototipo o *Mockup* cuadro de mando: Comportamiento del usuario ante la Privacidad

La visualización informa acerca del comportamiento del usuario ante la privacidad, se comprueba si tiene en cuenta el cuidado de su privacidad ante el uso de las *app´s* o no.

Con la información que se muestra en el cuadro de mando se conoce si las aplicaciones más intrusivas a la privacidad son de las más descargas por los usuarios, se conoce si los usuarios cuando valoran una *app*, consideran parte importante o no la privacidad, se conoce si las *app´s* de pago son menos intrusivas a la privacidad respecto a las *app´s* gratuitas o no, se conoce si la intromisión en la privacidad por parte de las *app´s* es la misma en todo tipo de público o no y se descubre el acceso a preferencias de consumo, hábitos, gustos, necesidades del usuario y el riesgo asociado a ello. Todo ello son parte de los requisitos del proyecto BI.

Se informa acerca de la concienciación o no por parte del usuario del cuidado de la privacidad al hacer uso de las *app´s*.

Cuadro de Mando: Comportamiento del Programador ante la Privacidad

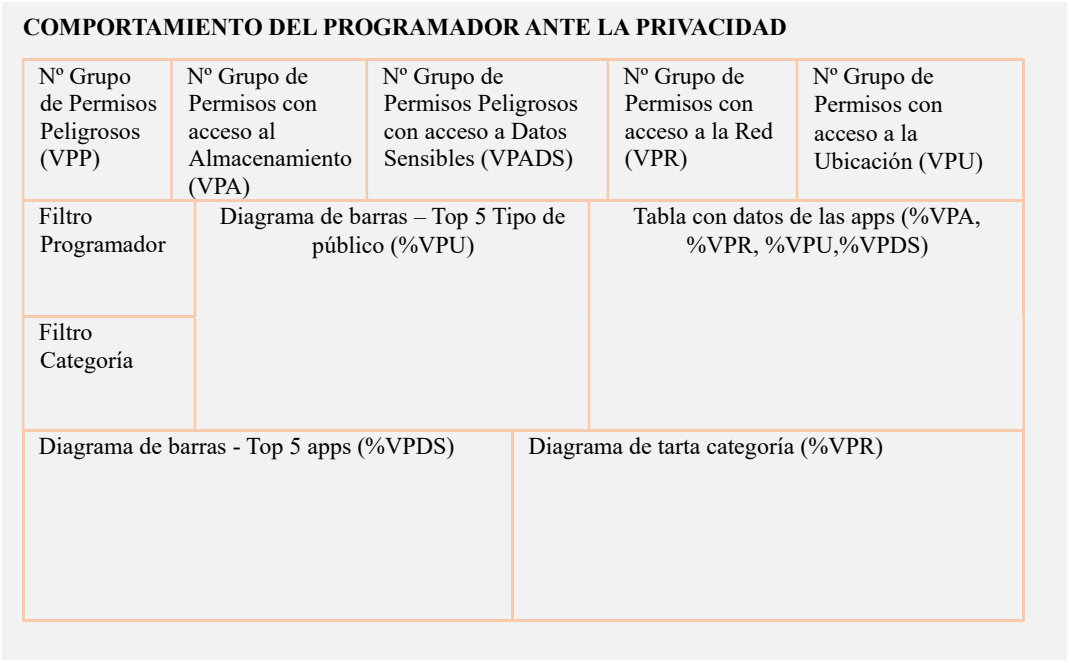

Figura 20. Prototipo o *Mockup* cuadro de mando: Comportamiento del Programador ante la Privacidad

El cuadro de mando permite conocer el comportamiento del programador ante la Privacidad, es decir, si la tiene presente o no a la hora de diseñar y programar las *app´s. Se c*onoce el uso que hace de los diferentes grupos de permisos y con ello los riesgos y el impacto en el contexto del usuario.

La información que se muestra se conoce el riesgo de acceso a datos de navegación, se conoce el grado de intromisión a la privacidad de las *app's*, se conoce el comportamiento de los programadores, se conoce la intromisión en datos personales por parte de la *app, se c*onoce el riesgo de acceso a información de registro de actividades vitales por la *app, se c*onoce el riesgo del acceso a aspectos personales de los usuarios por la *app, se* descubre el riesgo de acceso a la ubicación del dispositivo donde está instalada la *app, se* conoce el riesgo de acceso a metadatos por parte de las *app's,* se analiza el riesgo del acceso a identificadores únicos de los dispositivos electrónicos, se analiza el riesgo del acceso a datos y metadatos de las comunicaciones electrónicas, todo ello se contempla en los requisitos del proyecto BI y mediante la tabla 12 se conoce el impacto a la privacidad de las *app's* en el contexto del marco legislativo.

Cuadro de Mando: Cuidado de la Privacidad en el tiempo por las *app´s*

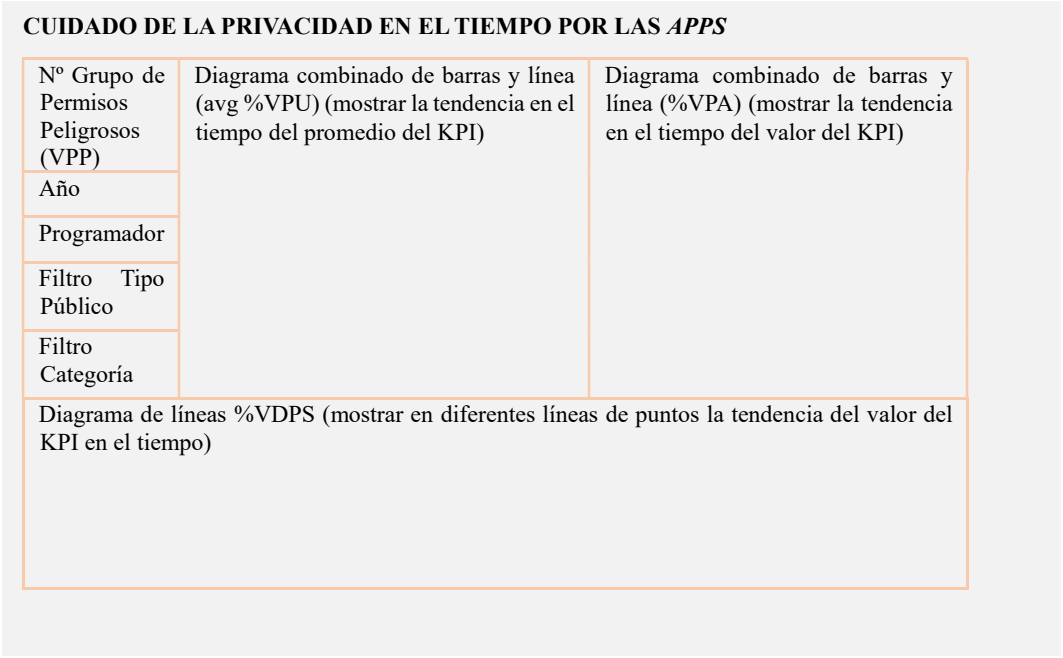

Figura 21. Prototipo o *Mockup* cuadro de mando: Cuidado de la Privacidad en el tiempo por las *app´s*

El cuadro de mando permite conocer información en el tiempo, acerca del uso de los grupos de permisos peligrosos por parte de las *app´s,* se informa si ha habido un cambio de tendencia o no en el tiempo a partir de la entrada en vigor del RGPD en mayo de 2018.

Cuadro de Mando: Cuidado de la Privacidad por las *app´s* según su funcionalidad

Un cuarto cuadro de mando informa acerca del impacto en la privacidad de las *app´s* de acuerdo con los grupos de permisos necesarios según su funcionalidad (Ver las 17 categorías y los grupos de permisos necesarios en la Tabla 17).

Se conoce el grado de formación de los desarrolladores en cuanto a los grupos de permisos que deben utilizar en cada una de las diferentes categorías y con ello se minimizan los riesgos a la privacidad del usuario al utilizar sólo aquellos que se necesitan para la funcionalidad que ofrece la *app* al usuario.

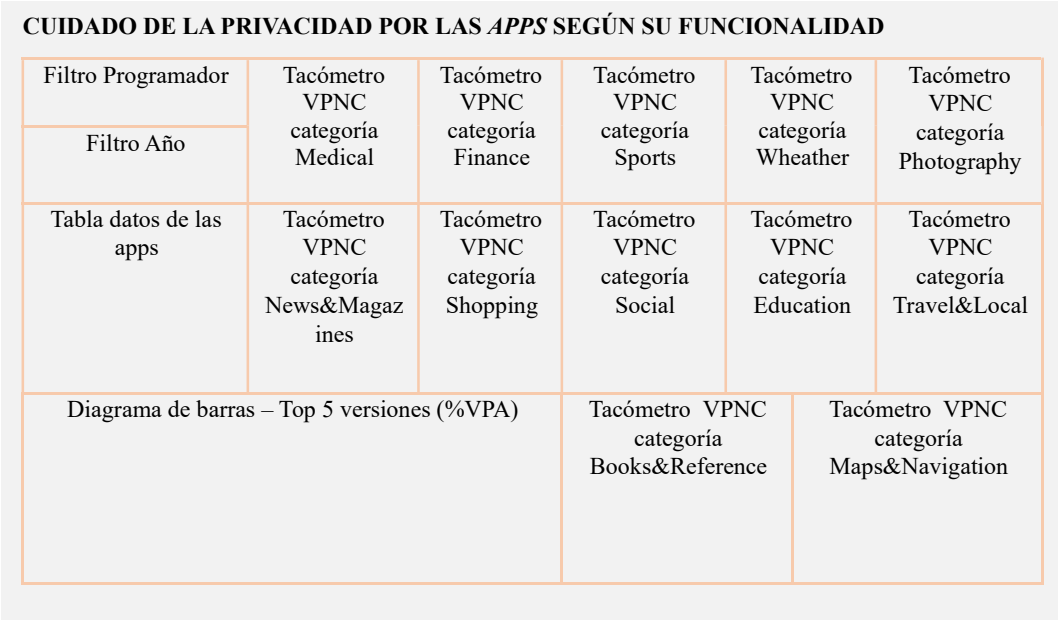

Figura 22. Prototipo o *Mockup* cuadro de mando: Cuidado de la Privacidad por las *app´s* según su funcionalidad

5.5.2 Cuadros de Mando

Una vez se diseñan los diferentes cuadros de mando que conforman la aplicación BI, se decide utilizar la herramienta *Qlik Sense* para su implementación.

Se procede a la carga de los 5 ficheros del *Data Mart* de Privacidad (ver modelo lógico en el apéndice E) y se establecen las relaciones entre las diferentes tablas de acuerdo al modelo dimensional (ver apéndice E).

Se obtiene el modelo de datos a utilizar por parte de la herramienta del proyecto BI, es decir, los diferentes cuadros de mando.

Los KPI's se definen en la propia herramienta *Qlik Sense* y se asocian a los diagramas donde se visualizan los datos.

Habiendo alojado los datos del proyecto BI en las diferentes tablas, se implementan los cuadros de mando de acuerdo al diseño que se muestra en el apartado 5.5.1 y con ello se informa acerca del impacto a la privacidad de las *app´s* de acuerdo al marco legislativo y así pueda tomar conciencia el programador de los riesgos del uso de los grupos de permisos peligrosos que se declaran en las *app´s* y con ello el impacto a la privacidad del usuario al hacer uso de la *app*.

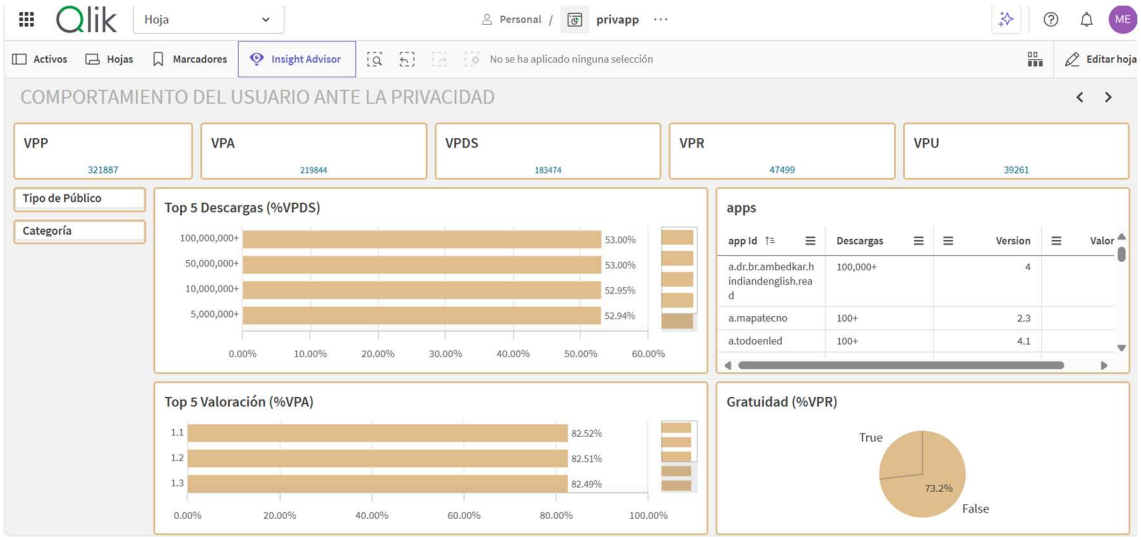

Figura 23. Cuadro de Mando: Comportamiento del usuario ante la Privacidad

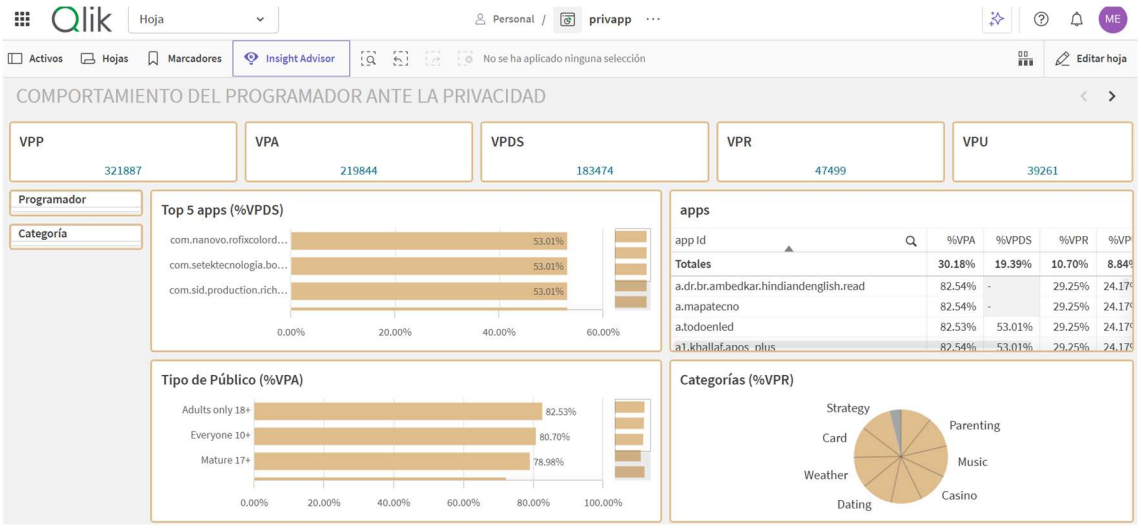

Figura 24. Cuadro de Mando: Comportamiento del Programador ante la Privacidad

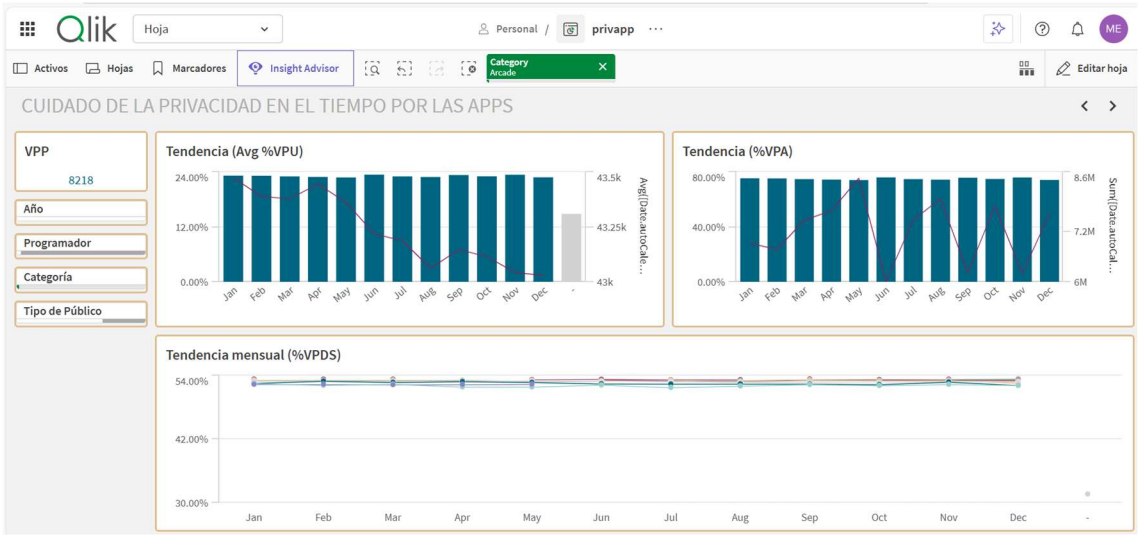

Figura 25. Cuadro de Mando: Cuidado de la Privacidad en el tiempo por las *app´s*

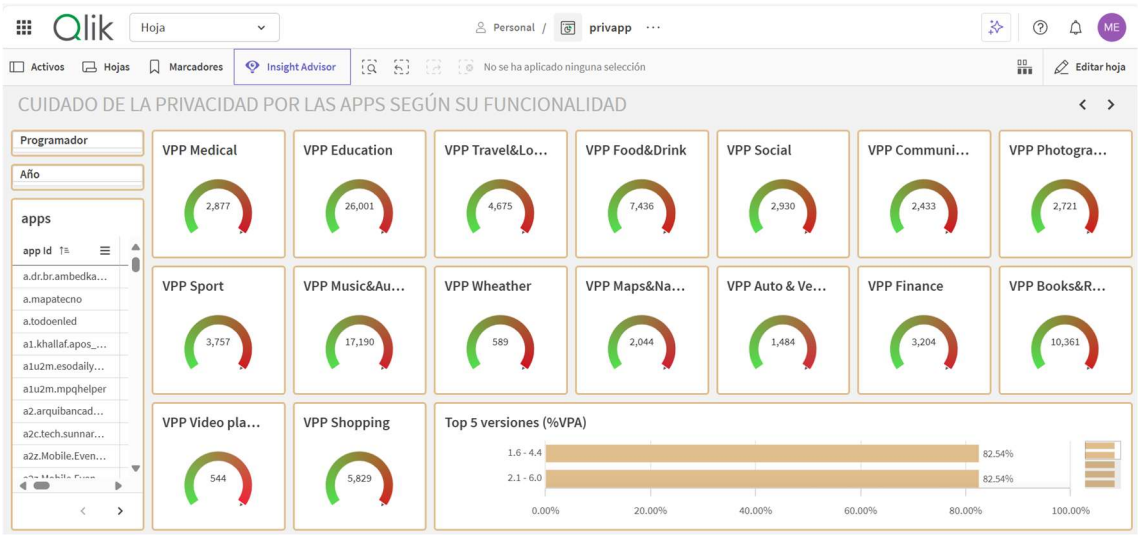

Figura 26. Cuadro de Mando: Cuidado de la Privacidad por las *app´s* según su funcionalidad

6. BUENAS PRÁCTICAS

Aquí se presentan una colección de buenas prácticas a aplicar desde el diseño de una *app* que se encaminan a reducir el impacto en la privacidad de las *app´s* de acuerdo al marco legislativo.

Desde el diseño de la *app* se ha de tener presente la recogida, el almacenamiento y el tratamiento de datos personales en el dispositivo electrónico.

- Incluir el principio del mínimo privilegio por defecto, según el cual las *app´s* acceden sólo a los datos que realmente necesitan para poder realizar la función para la cual se diseñan.
- Utilizar el mínimo número de permisos, no más de los que necesite la aplicación móvil de acuerdo con su funcionalidad.
- Utilizar el consentimiento previo por parte del usuario si se va a utilizar un permiso peligroso.
- Informar de forma transparente al usuario del uso del permiso peligroso, la forma de informar al usuario se realizará conforme al artículo 7 – 2 del RGPD.
- Ofrecer las opciones de denegar o aceptar de una forma parcial, es decir, mientras se utiliza la *app* o sólo para esa vez, de la forma menos invasiva posible.
- Diseñar la *app* con medidas para evitar el acceso no autorizado a datos personales, garantizar la protección de datos tanto en tránsito y como de datos que se almacenan en el dispositivo electrónico.
- Incluir mecanismos que permitan al usuario ver qué datos tratan en la *app*, permitir activar o desactivar de manera selectiva los permisos.
- Usar librerías de terceros sin auditar previamente, puede conllevar el uso indebido de los datos por la *app*.

7. CONCLUSIONES

El estudio realizado permite conocer el impacto en la privacidad de las *app´s* de acuerdo al marco legislativo y en concreto según la tabla de riesgos de la AEPD, se combina la probabilidad y el impacto de factores de riesgo (ver tabla 2).

Al seguir la metodología del ciclo de vida de *Kimball* se obtiene la aplicación BI con unos cuadros de mando donde se informa acerca de los objetivos del proyecto BI, se obtienen respuesta a los diferentes requisitos en los que se desglosan los objetivos (ver tabla 3) con lo que se busca ayudar en la toma de decisiones al programador desde la fase de diseño de la *app* al tener presente el cuidado de la privacidad desde el inicio y con ello mejorar los aspectos de privacidad que se mencionan en la auditoría realizada por OWASP.

Esto reporta una mejora en la privacidad de las *app´s* contribuye a mejorar el ecosistema.

Al aplicar el principio de minimización de datos, se obtiene el valor mejor posible de los diferentes KPI's que se utilizan en el proyecto y si se aplican medidas de seudonimización, los datos que se obtienen a partir de las *app´s* permiten datificar la realidad sin ser tan invasivas en la privacidad del usuario y causar un menor impacto en la privacidad de la sociedad.

Trabajos futuros

Entre otros, se proponen los siguientes:

* Una de las partes importantes de la seguridad de las aplicaciones móviles la forman los permisos que se declaran para su funcionamiento, a partir de los mismos, conocer el grado de intromisión a la privacidad de las aplicaciones móviles, saber si se utilizan los permisos necesarios de acuerdo con la funcionalidad o categoría de la aplicación móvil o, por el contrario, utilizan más permisos de los necesarios.

* Etiquetar las *app´s* de acuerdo con su impacto en la privacidad, partir de los permisos necesarios para su funcionamiento por categoría que establecen en el estudio "*Overprivileged Permission Detection for Android Applications*"[55].

* Construir un clasificador de *app´s* de acuerdo con las etiquetas de bajo, medio y alto impacto en cuanto a la privacidad.

REFERENCIAS BIBLIOGRAFICAS

[1] “I do (not) need that Feature!” - Understanding Users' Awareness and Control of Privacy Permissions on Android Smartphones https://dl.acm.org/doi/10.5555/3696899.3696923 (2024) Ultimo acceso Mayo 2025
[2] A Comprehensive Analysis of the Android Permissions System https://ieeexplore.ieee.org/abstract/document/9272963 (2020) Ultimo acceso Mayo 2025
[3] A large scale study of user behavior, expectations and engagement with android permissions https://www.usenix.org/conference/usenixsecurity21/presentation/cao-weicheng (2021) Ultimo acceso Mayo 2025
[4] Accept - Maybe - decline: Introducing partial consent for the permission-based access control model of Android https://dl.acm.org/doi/10.1145/3381991.3395603 (2020) Ultimo acceso Mayo 2025
[5] An Android Application Risk Evaluation Framework Based on Minimum Permission Set Identification https://www.sciencedirect.com/science/article/abs/pii/S0164121220300169, (2020) Ultimo acceso 2025
[6] Anaconda (JupyterNotebooks) https://nb.anaconda.com/jupyterhub/user Ultimo acceso Junio 2025
[7] Anaconda Cloud https://nb.anaconda.cloud/jupyterhub/ Ultimo acceso Mayo 2025
[8] Android 12 https://www.android.com/android-12/ Ultimo acceso Mayo 2025
[9] Android Permission https://www.kaggle.com/code/yashmehta648/android-permission Ultimo acceso 2025
[10] Android Permissions Dataset https://www.kaggle.com/datasets/gauthamp10/app-permissions-android, Ultimo acceso Marzo 2025
[11] Android Permissions: Evolution, Attacks, and Best Practices, https://ieeexplore.ieee.org/document/10747397, Ultimo acceso Mayo 2025
[12] Android programación. Autor: Ed Burnette Editorial: Anaya
[13] Android social applications permission overview from a privacy perspective https://ieeexplore.ieee.org/document/9484128 (2021) Ultimo acceso Mayo 2025
[14] Androids App Metadata (50.000 app´s) https://www.kaggle.com/datasets/kboghe/android-app´s-metadata Ultimo acceso Mayo 2025
[15] Apple App´store & Google Play Store data https://datarade.ai/data-products/apple-app´store-google-play-store-data-cleardata Ultimo acceso Abril 2025
[16] Assisting Developers in Preventing Permissions Related Security Issues in Android Application https://link.springer.com/book/10.1007/978-3-030-86507-8 Ultimo acceso Mayo 2025
[17] Big data La revolución de los datos masivos. Autores: Viktor Mayer-Schönberger y Kenneth Cukier. Editorial: Turner Noema
[18] Brightdata https://brightdata.com/cp/datasets/browse/gd_lsk382l8xei8vzm4u?id=hl_4b01fc93 Ultimo acceso Mayo 2025
[19] Ciencias de Datos un enfoque práctico de tecnologías, herramientas y aplicaciones. Autor: Luis Joyanes Aguilar. Editorial: Marcombo
[20] Cómo elaborar trabajos académicos y científicos (TFG, TFM, tesis y artículos) Autor: Ángel Cervera Rodríguez Editorial: Alianza editorial

[21] Cuadrante mágico de Gartner https://www.qlik.com/es-es/gartner-magic-quadrant-business-intelligence Ultimo acceso Junio 2025

[22] Databricks https://community.cloud.databricks.com/ Ultimo acceso Junio 2025
[23] Declara el uso de datos de tu app https://developer.android.com/privacy-and-security/declare-data-use?hl=es-419 Ultimo acceso Mayo 2025
[24] Derechos digitales. Autores: Lucrecio Rebollo Delgado y Pilar Zapatero Martín. Editorial: Dykinson, s.l.
[25] Desarrollo seguro en ingeniería del software. Aplicaciones seguras con Android, Nodejs, Python y C++ Autor: José Manuel Ortega Candel. Editorial Marcombo
[26] Deteccion Android Malware from App Permissions https://www.kaggle.com/code/quackaddict7/detecting-android-malware-from-app-permissions Ultimo acceso Mayo 2025
[27] Do Android App Developers Accurately Report Collection of Privacy-Related Data? https://dl.acm.org/doi/10.1145/3691621.3694949 (2022) Ultimo acceso Mayo 2025
[28] El gran libro de Android. Autor: Jesús Tomás Girones. Editorial: Marcombo ediciones técnicas
[29] Engineering Privacy in Smartphone App´s: A Technical Guideline Catalog for App Developers https://ieeexplore.ieee.org/document/9001128 (2020) Ultimo acceso Mayo 2025
[30] Enhancing Fidelity of Description in Android App´s With Category-Based Common Permissions https://ieeexplore.ieee.org/document/9496656 Ultimo acceso Mayo 2025
[31] Exodus https://reports.exodus-privacy.eu.org/es/ Ultimo acceso Mayo 2025
[32] Freely Given Consent?: Studying Consent Notice of Third-Party Tracking and Its Violations of GDPR in Android App´s https://dl.acm.org/doi/abs/10.1145/3548606.3560564 (2022) Ultimo acceso Mayo 2025
[33] GDPR Compliance Assessment for Cross Border Personal Data Transfers in Android App´s https://ieeexplore.ieee.org/document/9328756 (2021) Ultimo acceso Mayo 2025 https://www.sciencedirect.com/science/article/pii/S0167404823001724 Ultimo acceso Mayo 2025
[34] Gestión del riesgo y evaluación de impacto en tratamiento de datos personales https://www.aepd.es/guias/gestion-riesgo-y-evaluacion-impacto-en-tratamientos-datos-personales.pdf Ultimo acceso Junio 2025
[35] Gestionar permisos desde el panel de privacidad https://support.google.com/android/answer/13530434?hl=es#:~:text=En%20el%20%EE%80%80panel%20de%20privacidad,%EE%80%81 Ultimo acceso Mayo 2025
[36] Google Play Permissions Data https://www.kaggle.com/datasets/sdkruse/google-play-permissions-data/data Ultimo acceso Mayo 2025
[37] Google Play Store App´s https://www.kaggle.com/datasets/gauthamp10/google-playstore-app´s Ultimo acceso Abril 2025
[38] Google Play Store App´s https://www.kaggle.com/datasets/lava18/google-play-store-app´s Ultimo acceso Mayo 2025
[39] Grupo de trabajo “Artículo 29 sobre protección de datos” https://www.aepd.es/sites/default/files/2019-12/wp202_es.pdf Ultimo acceso Julio 2025
[40] Grupos de permisos https://developer.android.com/guide/topics/manifest/permission-group-element.html?hl=es-419 Ultimo acceso Mayo 2025
[41] Guía de privacidad desde el diseño https://www.aepd.es/guias/guia-privacidad-desde-diseno.pdf Ultimo acceso Julio 2025
[42] Identification of Possibly Intemperate Permission Demands in Android App´s https://ieeexplore.ieee.org/document/9753830 (2022) Ultimo acceso Mayo 2025

[43] Ingeniería de datos. Diseño, implementación y optimización de flujos de datos en Python. Autor: José Manuel Ortega Candel Editorial: Ra-Ma
[44] Ingeniería de la protección de datos https://www.aepd.es/documento/enisa ingenieria-de-la-proteccion-de-datos.pdf Ultimo acceso Mayo 2025
[45] Introducción a las bases de datos NoSQL usando MongoDB. Autor: Antonio Sarasa. Editorial: UOC
[46] Kaggle https://www.kaggle.com/ Ultimo acceso Junio 2025
[47] Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673
[48] Manifest.permission https://developer.android.com/reference/android/Manifest.permission Ultimo acceso Mayo 2025
[49] Manifest.permission https://developer.android.com/reference/android/Manifest.permission Ultimo acceso Mayo 2025
[50] Mobile App Development Trends To Follow in 2023 https://www.kaggle.com/datasets/infowindtechnologies/mobile-app-development-trends-to-follow-in-2023/discussion?sort=hotness Ultimo acceso Mayo 2025
[51] Mobile App Permissions for Ciber-Security Analysis https://www.kaggle.com/datasets/farukalam/mobile-app-permissions-for-cyber-security-analysis Ultimo acceso Mayo 2025
[52] Mobile Permissions https://mobileevolution.github.io/ Ultimo acceso Abril 2025
[53] MongoDB https://www.mongodb.com/ Ultimo acceso Junio 2025
[54] NATICUSdroid (Android Permissions) https://www.archive.ics.uci.edu/dataset/722/naticusdroid%2Bandroid%2Bpermissions%2Bdataset Ultimo acceso Abril 2025
[55] Overprivileged Permission Detection for Android Applications https://ieeexplore.ieee.org/document/8761572 (2019) Ultimo acceso Abril 2025
[56] OWASP https://owasp.org/www-project-mobile-top-10/ Ultimo acceso Mayo 2025
[57] OWASP Mobile Top 10 2024 https://owasp.org/www-project-mobile-top-10/ Ultimo acceso 2025
[58] Permisos en Android https://developer.android.com/guide/topics/permissions/overview?hl=es-419 Ultimo acceso Mayo 2025
[59] Permission Issues in Open-Source Android App´s: An Exploratory Study https://ieeexplore.ieee.org/document/8930838 (2019) Ultimo acceso Mayo 2025
[60] Permission-Educator: App for Educating Users About Android Permissions https://link.springer.com/book/10.1007/978-3-030-98404-5 Ultimo acceso Mayo 2025
[61] Preventing Permissions Security Issues in Android: a Developer's Perspective https://gtsslr21-reseau.sciencesconf.org/354299/gt2021.pdf Ultimo acceso Mayo 2025
[62] Privacidad y Anonimización de Datos. Autores: Jordi Casas Roma y Cristina Romero Tris. Editorial: UOC
[63] Privacy issues of android application permissions: A literature review https://onlinelibrary.wiley.com/doi/full/10.1002/ett.3773 Ultimo acceso Mayo 2025
[64] Privacypatterns.eu - collecting patterns for better privacy https://privacypatterns.eu/#/?limit=6&offset=12 Ultimo acceso Julio 2025
[65] Qlik https://www.qlik.com/es-es ultimo acceso Junio 2025

[66] Qlik Sense https://www.qlik.com/us/trial/qlik-talend-cloud Ultimo acceso Agosto 2025
[67] Reglamento General de Protección de Datos https://www.boe.es/doue/2016/119/L00001-00088.pdf
[68] Runtime and Design Time Completeness Checking of Dangerous Android App Permissions Against GDPR, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10373786 (2024), Ultimo acceso Mayo 2025
[69] Runtime Permission Issues in Android App´s: Taxonomy, Practices, and Ways Forward https://ieeexplore.ieee.org/document/9705152 (2023) Ultimo acceso Abril 2025
[70] Semantic-aware Comment Analysis Approach for API Permission Mapping on Android https://dl.acm.org/doi/10.1145/3443279.3443312 (2021) Ultimo acceso Abril 2025
[71] Share First, Ask Later (or Never?) Studying Violations of GDPR's Explicit Consent in Android App´s https://www.usenix.org/conference/usenixsecurity21/presentation/nguyen (2021) Ultimo acceso Mayo 2025
[72] Sin privacidad no hay ciberseguridad https://www.aepd.es/prensa-y-comunicacion/blog/sin-privacidad-no-hay-ciberseguridad Ultimo acceso Julio 2025
[73] Statista https://www.statista.com/study/11677/google-statista-dossier/ Ultimo acceso Abril 2025
[74] The Data Wartehouse Lifecycle Toolkit. Autores: Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy y Bob Becker. Editorial: Kimball Group
[75] Understanding the Bad Development Practices of Android Custom Permissions in the Wild https://www.computer.org/csdl/journal/tq/5555/01/10826575/23i0s5AJdra Ultimo acceso Mayo 2025

APÉNDICES

A. Permisos en el Sistema Operativo de Android

Normal Permissions	
PERSISTENT_ACTIVITY	GET_PACKAGE_SIZE
RESTART_PACKAGES	INTERNET
GET_TASKS	MODIFY_AUDIO_SETTINGS
ACCESS_LOCATION_EXTRA_COMMANDS	READ_SYNC_SETTINGS
ACCESS_NETWORK_STATE	READ_SYNC_STATS
ACCESS_WIFI_STATE	RECEIVE_BOOT_COMPLETED
BLUETOOTH	REORDER_TASKS
BLUETOOTH_ADMIN	SET_WALLPAPER
BROADCAST_STICKY	SET_WALLPAPER_HINTS
CHANGE_NETWORK_STATE	VIBRATE
EXPAND_STATUS_BAR	WAKE_LOCK
WRITE_SYNC_SETTINGS	CHANGE_WIFI_STATE

Figura 27. Permisos de tipo Normal [2]

Signature Permissions	
SET_PREFERRED_APPLICATIONS	DELETE_CACHE_FILES
WRITE_SETTINGS	BATTERY_STATS
SYSTEM_ALERT_WINDOW	CHANGE_CONFIGURATION
CLEAR_APP_CACHE	

Figura 28. Permisos de tipo *Signature* [2]

SignatureorSystem Permissions	
ACCESS_CHECKIN_PROPERTIES	MODIFY_PHONE_STATE
BROADCAST_PACKAGE_REMOVED	MOUNT_UNMOUNT_FILESYSTEMS
CALL_PRIVILEGED	READ_INPUT_STATE
CHANGE_COMPONENT_ENABLED_STATE	READ_LOGS
CONTROL_LOCATION_UPDATES	REBOOT
DELETE_PACKAGES	SET_ALWAYS_FINISH
DIAGNOSTIC	SET_ANIMATION_SCALE
DUMP	SET_DEBUG_APP
FACTORY_TEST	SET_PROCESS_LIMIT
INSTALL_PACKAGES	SET_TIME_ZONE
MASTER_CLEAR	SIGNAL_PERSISTENT_PROCESSES
WRITE_APN_SETTINGS	STATUS_BAR
WRITE_GSERVICES	

Figura 29. Permisos de tipo *SignatureorSystem* [2]

	Permisos peligrosos	**Grupo de permisos**
1	Read_Calendar	Calendar
2	Write_Calendar	Calendar
3	Camera	Camera
4	Read_Contacts	Contacts
5	Write_Contacts	Contacts
6	Get_Accounts	Contacts
7	Access_Fine_Location	Location
8	Access_Coarse_Location	Location
9	Record_Audio	Microphone
10	Read_Phone_State	Phone
11	Read_Phone_Numbers	Phone
12	Call_Phone	Device ID & call information
13	Answer_Phone_Calls	Device ID & call information
14	Read_Call_Log	Device ID & call information
15	Write_Call_Log	Device ID & call information
16	Add_VoiceMail	Phone
17	Use_sip	Phone
18	Process_outgoing_calls	Device ID & call information
19	Body_Sensors	Wearable sensors/Activity data
20	Send_Sms	SMS
21	Receive_Sms	SMS
22	Read_sms	SMS
23	Receive_Wap_Push	SMS
24	Receive_Mms	SMS
25	Read_External_Storage	Storage
26	Write_External_Storage	Storage

Tabla 5. Permisos peligrosos y el grupo de permisos *dangerous*

En la web para programadores de Android se consultan los diferentes permisos y el grupo de permisos al que pertenecen [24].

B. Artículos del RGPD

Artículo 4 - Definiciones
1) «datos personales»: toda información sobre una persona física identificada o identificable («el interesado»); se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona;

Tabla 6. Artículo 4. Definiciones

Artículo 5 - Principios relativos al tratamiento - donde se establece los datos personales
1.(a) tratados de manera lícita, leal y transparente en relación con el interesado («licitud, lealtad y transparencia»);
1.(b) recogidos con fines determinados, explícitos y legítimos, y no serán tratados ulteriormente de manera incompatible con dichos fines; de acuerdo con el artículo 89, apartado 1, el tratamiento ulterior de los datos personales con fines de archivo en interés público, fines de investigación científica e histórica o fines estadísticos no se considerará incompatible con los fines iniciales («limitación de la finalidad»);
1.(c) adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados («minimización de datos»);
1.(e) mantenidos de forma que se permita la identificación de los interesados durante no más tiempo del necesario para los fines del tratamiento de los datos personales; los datos personales podrán conservarse durante períodos más largos siempre que se traten exclusivamente con fines de archivo en interés público, fines de investigación científica o histórica o fines estadísticos, de conformidad con el artículo 89, apartado 1, sin perjuicio de la aplicación de las medidas técnicas y organizativas apropiadas que impone el presente Reglamento a fin de proteger los derechos y libertades del interesado («limitación del plazo de conservación»)
1.(f) tratados de tal manera que se garantice una seguridad adecuada de los datos personales, incluida la protección contra el tratamiento no autorizado o ilícito y contra su pérdida, destrucción o daño accidental, mediante la aplicación de medidas técnicas u organizativas apropiadas («integridad y confidencialidad»).

Tabla 7. Artículo 5 – Principios relativos al tratamiento – donde se establece los datos personales

Artículo 7 - Condiciones para el consentimiento
1.Cuando el tratamiento se base en el consentimiento del interesado, el responsable deberá ser capaz de demostrar que aquel consintió el tratamiento de sus datos personales.
2. Si el consentimiento del interesado se da en el contexto de una declaración escrita que también se refiera a otros asuntos, la solicitud de consentimiento se presentará de tal forma que se distinga claramente de los demás asuntos, de forma inteligible y de fácil acceso y utilizando un lenguaje claro y sencillo. No será vinculante ninguna parte de la declaración que constituya infracción del presente Reglamento.
3. El interesado tendrá derecho a retirar su consentimiento en cualquier momento. La retirada del consentimiento no afectará a la licitud del tratamiento basada en el consentimiento previo a su retirada. Antes de dar su consentimiento, el interesado será informado de ello. Será tan fácil retirar el consentimiento como darlo.
4. Al evaluar si el consentimiento se ha dado libremente, se tendrá en cuenta en la mayor medida posible el hecho de si, entre otras cosas, la ejecución de un contrato, incluida la prestación de un servicio, se supedita al consentimiento al tratamiento de datos personales que no son necesarios para la ejecución de dicho contrato

Tabla 8. Artículo 7. Condiciones para el consentimiento

Artículo 25 - Protección de datos desde el diseño y por defecto
1. Teniendo en cuenta el estado de la técnica, el coste de la aplicación y la naturaleza, ámbito, contexto y fines del tratamiento, así como los riesgos de diversa probabilidad y gravedad que entraña el tratamiento para los derechos y libertades de las personas físicas, el responsable del tratamiento aplicará, tanto en el momento de determinar los medios de tratamiento como en el momento del propio tratamiento, medidas técnicas y organizativas apropiadas, como la seudonimización, concebidas para aplicar de forma efectiva los principios de protección de datos, como la minimización de datos, e integrar las garantías necesarias en el tratamiento, a fin de cumplir los requisitos del presente Reglamento y proteger los derechos de los interesados.

2. El responsable del tratamiento aplicará las medidas técnicas y organizativas apropiadas con miras a garantizar que, por defecto, solo sean objeto de tratamiento los datos personales que sean necesarios para cada uno de los fines específicos del tratamiento. Esta obligación se aplicará a la cantidad de datos personales recogidos, a la extensión de su tratamiento, a su plazo de conservación y a su accesibilidad. Tales medidas garantizarán en particular que, por defecto, los datos personales no sean accesibles, sin la intervención de la persona, a un número indeterminado de personas físicas.

Tabla 9. Artículo 25. Protección de datos desde el diseño y por defecto

Artículo 32 - Seguridad de los datos personales
a) La seudonimización y el cifrado de datos personales

Tabla 10. Artículo 32. Seguridad de los datos personales

C. Fuentes de Datos

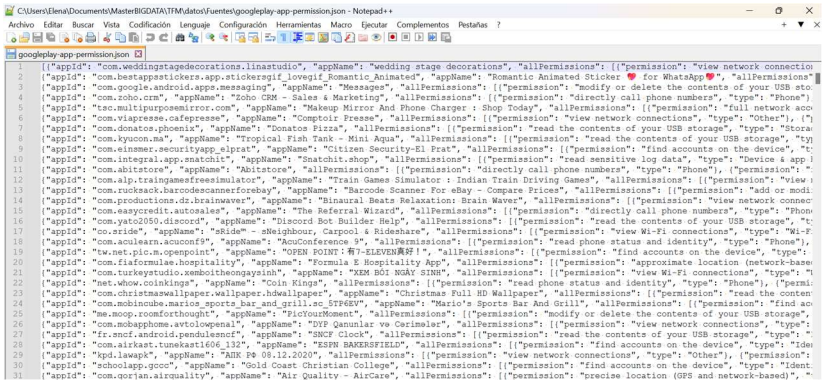

Figura 30. Contenido del fichero *googleplay-app-permission.json*

Atributo	Descripción	Tipo
appId	Id de la aplicación	String
appName	Nombre de la aplicación	String
allPermissions	Array formado por tuplas <Permission, Type >	String
Permission	Permiso	String
Type	Grupo de permisos	String

Tabla 11. Datos correspondientes a la colección *AppPermissionGoogleplay*

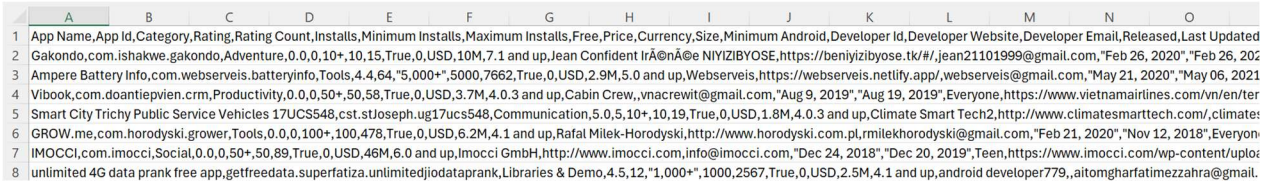

Figura 31. Contenido del fichero *Google-Playstore.csv*

Atributo	Descripción	Tipo
App Name	Nombre de la aplicación	String
App Id	Identificador de la aplicación	String
Category	Categoría de la aplicación	String
Rating	Valoración de la aplicación por el usuario	Float
Rating Count	Recuento de valoraciones que realizan los usuarios	Integer
Installs	Descargas de las *app´s* por los usuarios	String
Minimum Installs	Descargas mínimas	Integer
Maximum Installs	Descargas máximas	Integer
Free	Pagar o no por el uso de la *app*	Boolean
Price	Precio a pagar por descargar la *app*	Integer
Currency	Divisa del pago de la *app*	String
Size	Tamaño de la *app*	String
Minimum Android	Versión mínima de Android para poder instalar la *app*	String
Developer Id	Identificador del programador	String
Developer Website	Web del programador	String
Developer Email	Correo electrónico del programador	String
Released	Fecha de lanzamiento de la *app*	String
Privacy Policy	Web de la política de privacidad de la *app*	String
Last Updated	Fecha última actualización de la *app*	String
Content Rating	Tipo de público al que se dirige la *app*	String
Ad Supported	Si tiene soporte la *app*	Boolean
In app purchases	Compras en aplicaciones	Boolean
Editor Choice	Elección del Editor	Boolean

Tabla 12. Datos correspondientes a la colección *GooglePlayStore*

D. KPIs

- **Nº de grupo de permisos peligrosos con acceso a datos sensibles** (categoría/Tipo de Publico /Programador/Fecha de lanzamiento):

Volumen de permisos de datos sensibles (VPDS)	
Definición	Grupos de Permisos peligrosos con acceso a datos sensibles que se declaran en la *app*. Se consideran datos sensibles del usuario los contactos, SMS, registros de llamadas. VPDS Se obtiene el número de grupos de permisos con acceso a datos sensibles que se declaran en todas las *app´s*. El indicador obtiene la intromisión a la privacidad por parte de las *app´s*. Grupos de Permisos que acceden a datos privados del usuario.
Cálculo	Sumar todos los grupos de permisos de acceso a datos sensibles declarados en la *app*. VPDS = ∑ Grupos de Permisos que acceden a datos sensibles del usuario por *app*
Excepciones	
Fuentes de datos	*AppPermissionGoogleplay, GooglePlayStore*
RGPD	La forma de proteger dichos datos personales se establece en al Artículo 25 – 2, en el artículo 5 – 1(b) y artículo 5 – 1(c) mencionan el principio de minimización de datos.

- **Nº de grupo de permisos peligrosos de alto riesgo** (categoría/Tipo de Publico/Programador/Fecha de lanzamiento):

Volumen de permisos de alto riesgo o permisos peligrosos (VPP)	
Definición	Grupos de Permisos peligrosos o de alto riesgo que se declaran en la *app*. Se consideran grupos de permisos peligrosos los que acceden a recursos como la cámara, el micrófono, la ubicación, contactos, SMS y registros de llamadas. VPP Se obtiene el número de grupos de permisos peligrosos que se declaran en todas las *app´s*. El indicador obtiene la intromisión a la privacidad por parte de las *app´s* en un nivel de riesgo alto. Grupos de Permisos que acceden a datos privados del usuario.
Cálculo	Sumar todos los grupos de permisos peligrosos declarados en la *app*. VPP = ∑ Grupo de Permisos peligrosos por *app*
Excepciones	
Fuentes de datos	*AppPermissionGoogleplay, GooglePlayStore*
RGPD	Usar los Grupos de permisos peligrosos implica el consentimiento previo por parte del usuario, Artículo 4 (11). Las condiciones del consentimiento se contemplan en el artículo 7. La forma de proteger dichos datos se establece en al Artículo 25 – 2, en el artículo 5 – 1(b) y artículo 5 – 1(c) mencionan el principio de minimización de datos.

- **Nº de grupos de permisos peligrosos de acceso a la red** (categoría/Tipo de Publico/Programador/Fecha de lanzamiento):

Volumen de permisos de acceso a la red (VPR)	
Definición	Grupo de Permisos de acceso a la Red que se declaran en la *app*. Se consideran grupos de permisos de acceso a la red, grupos de permisos que acceden a la Red y a los datos del dispositivo. VPR Se obtiene el número de grupos de permisos con acceso a la red que se declaran en las *app´s*. El indicador obtiene la intromisión a la privacidad por parte de las *app´s*. Grupos de Permisos que transmiten información del dispositivo a servidores remotos.
Cálculo	Sumar todos los grupos de permisos peligrosos con acceso a la red que se declaran en la *app*. VPR = ∑ Grupos de Permisos de acceso a la red por *app*
Excepciones	
Fuentes de datos	*AppPermissionGoogleplay, GooglePlayStore*
RGPD	Usar los grupos de permisos peligrosos con acceso a la red implica el consentimiento previo por parte del usuario, Artículo 4 (11). La forma de proteger dichos datos se establece en el Artículo 25 – 2, en el artículo 5 – 1(b) y artículo 5 – 1(c) mencionan el principio de minimización de datos.

- **Nº de grupos de permisos peligrosos con acceso al almacenamiento**

(categoría/Tipo de Publico/ Programador/Fecha de lanzamiento):

Volumen de permisos de acceso al almacenamiento (VPA)	
Definición	Grupo de Permisos peligrosos con acceso al almacenamiento que se declaran en la *app*. Se consideran grupos de permisos peligrosos con acceso al almacenamiento, grupos de permisos que acceden a los datos alojados en el dispositivo móvil. VPA Se obtiene el número de grupos de permisos con acceso al almacenamiento del dispositivo móvil de todas las *app´s*. El indicador obtiene la intromisión a la privacidad por parte de las *app´s*. Grupos de Permisos que acceden a los datos almacenados en el dispositivo móvil por parte de la aplicación.
Cálculo	Sumar todos los grupos de permisos con acceso al almacenamiento que se declaran en las *app´s*. VPA = ∑ Grupos de Permisos con acceso al almacenamiento por *app*
Excepciones	
Fuentes de datos	*AppPermissionGoogleplay, GooglePlayStore*
RGPD	Usar grupos de permisos con acceso a los datos que se alojan en el dispositivo electrónico implica el consentimiento previo por parte del usuario, Artículo 4 (11). La forma de proteger dichos datos se establece en el Artículo 25 – 2 del RGPD, el artículo 5 – 1(b) y artículo 5 – 1(c) mencionan el principio de minimización de datos.

- **Nº de grupos de permisos peligrosos con acceso a la ubicación** (categoría/Tipo de Publico/ Programador/Fecha de lanzamiento):

Volumen de permisos de acceso a la ubicación (VPU)	
Definición	Grupo de Permisos peligrosos con acceso a la ubicación que se declaran en la *app*. Se consideran grupos de permisos con acceso a la ubicación, grupos de permisos que acceden a recursos como el GPS del dispositivo móvil. VPU Se obtiene el número de grupos de permisos con acceso a la ubicación del dispositivo móvil que se declaran en todas las *app´s*. El indicador obtiene la intromisión a la privacidad por parte de las *app´s*. Permisos que acceden a datos de ubicación del usuario.
Cálculo	Sumar todos los grupos de permisos con acceso a la ubicación que se declaran en la *app*. VPU = ∑ Grupos de Permisos con acceso a la ubicación por *app*
Excepciones	
Fuentes de datos	*AppPermissionGoogleplay, GooglePlayStore*
RGPD	Usar grupos de permisos con acceso a la ubicación del dispositivo electrónico implica el consentimiento previo por parte del usuario, Artículo 4 (11). La forma de proteger dichos datos se establece en al Artículo 25 – 2 del RGPD, artículo 5 – 1(b) y artículo 5 – 1(c) mencionan el principio de minimización de datos.

- **Nº de grupo de permisos peligrosos innecesarios por categoría**

Volumen de grupo de permisos innecesarios por categoría (VPNC)	
Definición	Grupo de Permisos peligrosos innecesarios que se declaran en las *app´s* de acuerdo con la funcionalidad o categoría Se consideran innecesarios grupos de permisos que no se incluyen en la clasificación que realizan en el estudio “Overprivileged Permission Detection for Android Applications" [55]. VPNC Se obtiene el número de grupos de permisos no necesarios en la *app* de acuerdo con su funcionalidad o categoría. El indicador obtiene la intromisión a la privacidad por parte de las *app´s*. Declarar más grupos de permisos de los necesarios en su categoría puede ser un indicador de que la *app* recopila datos innecesarios, aumenta el riesgo al filtrar información sensible.
Cálculo	Sumar todos los grupos de permisos que se declaran en la *app* y restar los grupos de permisos necesarios de acuerdo con su categoría. VPNC = ∑ grupos de permisos que se declaran en la app - ∑ grupos de permisos necesarios de acuerdo con la funcionalidad de la app
Excepciones	Incluir sólo las categorías de la tabla 19
Fuentes de datos	*AppPermissionGoogleplay, GooglePlayStore*
RGPD	Se excede en el número de grupos de permisos por parte de las *app´s*, solicita permisos innecesarios para su funcionamiento. De acuerdo con el Art. 5 (1c): “los datos personales serán adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados («minimización de datos»)”.

Definir los diferentes grupos de permisos a incluir en cada una de las métricas o KPIs:

Métrica	Perfilado de dato	Colección
VPDS	Grupo de permisos	*AppPermissionGoogleplay*
	Filtros (categoría/Tipo de Publico/ Programador/Fecha de lanzamiento)	*GooglePlayStore*
VPP	Grupo de permisos	*AppPermissionGoogleplay*
	Filtros (categoría/Tipo de Publico/ Programador/Fecha de lanzamiento)	*GooglePlayStore*
VPR	Grupo de permisos	*AppPermissionGoogleplay*
	Filtros (categoría/Tipo de Publico/ Programador/Fecha de lanzamiento)	*GooglePlayStore*
VPA	Grupo de permisos	*AppPermissionGoogleplay*
	Filtros (categoría/Tipo de Publico/ Programador/Fecha de lanzamiento)	*GooglePlayStore*
VPU	Grupo de permisos	*AppPermissionGoogleplay*
	Filtros (categoría/Tipo de Publico/ Programador/Fecha de lanzamiento)	*GooglePlayStore*
VPNC	Grupo de Permisos	*AppPermissionGoogleplay*
	Filtros : Categoría	*GooglePlayStore*

Tabla 13. Métricas y datos que utilizar.

KPI	Grupo de permisos
VPDS	Contacts, Location, SMS, Wearable sensors/Activity data, Device ID & call information, Phone
VPP	Microphone, Location, SMS, Contacts, Phone, Device ID & call information, Calendar, Wearable sensors/Activity data, Identity, Storage
VPR	Wi-Fi Connection Information
VPA	Storage, SMS, Contacts, Calendar, Device ID & call information
VPU	Location

Tabla 14. Métricas y grupos de permisos (Ver tabla1)

Utilizar los siguientes datos en la métrica **Nº de Grupo de permisos innecesarios por categoría (VPNC):**

	Permisos peligrosos	Grupo de permisos
1	Read_Calendar	Calendar
2	Write_Calendar	Calendar
3	Camera	Camera
4	Read_Contacts	Contacts
5	Write_Contacts	Contacts
6	Get_Accounts	Contacts
7	Access_Fine_Location	Location
8	Access_Coarse_Loation	Location
9	Record_Audio	Microphone
10	Read_Phone_State	Phone
11	Read_Phone_Numbers	Phone
12	Call_Phone	Device ID & call information
13	Answer_Phone_Calls	Device ID & call information
14	Read_Call_Log	Device ID & call information
15	Write_Call_Log	Device ID & call information
16	Add_VoiceMail	Phone
17	Use_sip	Phone
18	Process_outgoing_calls	Device ID & call information
19	Body_Sensors	Wearable sensors/Activity data

20	Send_Sms	SMS
21	Receive_Sms	SMS
22	Read_Sms	SMS
23	Receive_Wap_Push	SMS
24	Receive_Mms	SMS
25	Read_External_Storage	Storage
26	Write_External_Storage	Storage

Tabla 15. Permisos peligrosos declarados en las *app´s*

	Categoría	Permisos Peligrosos	Grupos de permisos Peligrosos
1	Sports	6, 7, 8, 10, 25, 26	Contacts, Location, Phone, Storage
2	Travel&Local	3, 6, 7, 8, 10, 25, 26	Camera, Contacts, Location, Phone, Storage
3	Medical	7, 8, 25, 26	Location, Storage
4	Books&Reference	3, 6, 10, 25, 26	Camera, Contacts, Phone, Storage
5	Maps&Navigation	3, 7, 8, 10, 25, 26	Camera, Location, Phone, Storage
6	Wheather	7, 8, 10, 25, 26	Location, Phone, Storage
7	Photography	3, 7, 8, ,9 ,25, 26	Camera, Location, Microphone, Storage
8	Education	3, 6, 9, 10, 25,26	Camera, Contacts, Microphone, Phone, Storage
9	News&Magazines	6, 7, 8, 10, 25, 26	Contacts, Location, Phone, Storage
10	Social	3, 4, 6, 7, 8, 9, 10, 25, 26	Camera, Contacts, Location, Microphone, Phone, Storage
11	Video players &Editors	3, 9, 10, 25, 26	Camera, Microphone, Phone, Storage
12	Finance	3, 4, 7, 8, 10, 21, 22, 25, 26	Camera, Contacts, Location, Phone, SMS, Storage
13	Shopping	3, 7, 8, 10, 25, 26	Camera, Location, Phone, Storage
14	Auto & Vehicles	7, 8, 10, 12, 25, 26	Location, Phone, Device ID & call information, Storage
15	Communication	3, 4, 6, 7, 8, 9, 10, 21, 25, 26	Camera, Contacts, Location, Microphone, Phone, SMS, Storage
16	Music & Audio	9, 10, 25, 26	Microphone, Phone, Storage
17	Food & Drink	7, 8, 25, 26	Location, Storage

Tabla 16. Categorías de app´s y permisos peligrosos a utilizar de acuerdo a su funcionalidad

E. Analítica de datos

Se analizan los datos que se alojan en las colecciones de la base de datos *NoSQL* que se seleccionan para el proyecto BI.

E.1 Colección de datos *GooglePlayStore*

Se construye el fichero *Google-Playstore-csv.csv* a partir de los datos que se alojan en la colección *GooglePlayStore,* se carga en la herramienta *Jupyter (notebooks*) y se analiza su contenido a partir del *dataset.*

```
#Cargamos el fichero con la muestra de los datos a analizar y construimos
#el dataframe apps
apps = pd.read_csv('./Google-Playstore-csv.csv',
                   #index_col=0
                  )
```

Figura 32. Carga del fichero *Google-Playstore-csv* en la herrami enta *Jupyter* y análisis

```
#Obtenemos información de las variables de tipo object:
# count: número de instancias contenidas en el dataframe con datos en la variabl
# unique: valores diferentes que tendría la variable dentro del dataframe
# top:  vaalor de la variable con una mayor frecuencia en el dataframe
# freq: la mayor frecuencia de un valor dentro de la variable del dataframe
apps.describe(include = object)
```

	App Name	App Id	Category	Installs	Currency	Size	Mini An
count	2312939	2312944	2312944	2312837	2312809	2312748	23(
unique	2177943	2312944	48	22	15	1657	
top	Tic Tac Toe	com.yyazilim.biliyormusun	Education	100+	USD	Varies with device	4.
freq	382	1	241090	443368	2311548	74777	6(

Figura 33. Consulta del contenido del *dataset app´s*

- **Categoría *(Category)***

En los datos se constatan 48 categorías diferentes en el atributo "*category*". Se valora la agrupación de algunas por sus diferentes fines: social, financiera, deportes, etc.

```
Category
Education                241090
Music & Audio            154906
Tools                    143988
Business                 143771
Entertainment            138276
Lifestyle                118331
Books & Reference        116728
Personalization           89210
Health & Fitness          83510
Productivity              79698
Shopping                  75256
Food & Drink              73927
Travel & Local            67288
Finance                   65466
Arcade                    53792
Puzzle                    51168
Casual                    50813
Communication             48167
Sports                    47483
Social                    44734
News & Magazines          42807
Photography               35552
Medical                   32065
Action                    27555
Maps & Navigation         26722
Simulation                23282
Adventure                 23203
Educational               21308
Art & Design              18539
Auto & Vehicles           18280
House & Home              14369
Video Players & Editors   14015
Events                    12841
Trivia                    11795
Beauty                    11772
Board                     10588
```

Figura 34. Valores del atributo "*Category*" en el *dataset app´s*

- **Tipo de Público *(Content Rating)***

En los datos se muestran 6 tipos de público diferentes en el atributo "*content rating*". En los correspondientes a "Unrated" y "Adults only 18+" el número de instancias es mínimo si se compara con el resto y con el volumen de instancias del *dataset,* más de 2.3 millones.

```
#Obtenemos los datos de los tipos de público ordenados de más a menos por el núm
apps["Content Rating"].value_counts()

Content Rating
Everyone           2022089
Teen                196375
Mature 17+           60289
Everyone 10+         33901
Unrated                154
Adults only 18+        136
Name: count, dtype: int64
```

Figura 35. Valores del atributo "*Content Rating*" en el *dataset app´s*

- **Descargas (*Install*)**

Se analiza el atributo *Install* (descargas) y se observa el formato "100+"

	count	unique	top	freq
Installs	2312837	22	100+	443368

Figura 36. Valor más alto del atributo "*Installs*" en el *dataset app´s*

- **Gratuidad *(Free)***

Se analizan los datos del atributo "*Free*" y se observa una frecuencia absoluta mayor en las *app´s* gratuitas frente a las *app´s* de pago.

```
#Mostramos el recuento de cada uno de los desarrolladores
apps.groupby(['Free']).count()
```

Free	App Name	App Id	Category	Rating	Rating Count	Installs	Minimum Installs	Maximum Installs
False	45068	45068	45068	44841	44841	44961	44961	45068
True	2267871	2267876	2267876	2245220	2245220	2267876	2267876	2267876

Figura 37. Valores del atributo "*Free*" en el *dataset app´s*

- **Programador**

Se analizan los datos, se observa un elevado número de valores diferentes en el atributo "*Developer Id"*, muchos tienen frecuencia absoluta 1.

```
#Obtenemos los datos de los desarrolladore ordenados de más a menos por el númer
apps["Developer Id"].value_counts()
```

```
Developer Id
Subsplash Inc                          5422
TRAINERIZE                             5153
ChowNow                                4865
OrderYOYO                              2884
Phorest                                2821
                                       ...
karel.srt                                 1
Isabel Technologies, Inc.                 1
Evergreen Publications (India) Ltd.       1
DevDoma                                   1
UWash Mobile                              1
Name: count, Length: 758371, dtype: int64
```

Figura 38: Valores del atributo "*Developer Id*" en el *dataset app´s*

- **Fecha de lanzamiento de la aplicación *(*released*)***

Se analizan los datos, se observan en el atributo "*released*" datos desde el año 2010 hasta el año 2021. La franja de tiempo permite descubrir tendencias antes y después de la entrada en vigor del RGPD en mayo de 2018.

Formato de los datos alojados en el atributo *released*: "May 19, 2021"

```
#Obtenemos los datos de los lanzamientos de las apps ordenados de mayor a menor
apps["Released"].value_counts()
```

```
Released
Jun 16, 2020    2051
Feb 27, 2020    2034
Jun 15, 2020    2025
Jun 24, 2020    2022
Feb 19, 2020    2015
                ...
Feb 28, 2010       2
Apr 5, 2010        1
Jun 16, 2010       1
Apr 10, 2010       1
Jun 16, 2021       1
```

Figura 39. Valores del atributo "*Released*" en el *dataset app´s*

E.2 Colección de datos *AppPermissionGoogleplay*

Partiendo de los datos de la colección *AppPermissionGoogleplay* que se aloja en el *DataWarehouse* se construye el *dataframe*.

Cada uno de los grupos de permisos, se etiqueta como *Type*, es decir, cada una de las *app´s* contiene tuplas con los datos {permission,type} corresponden a cada uno de los permisos que se declara en la *app* el conjunto de tuplas se etiqueta como *allPermissions*.

{"appId": "com.zoho.crm", "appName": "Zoho CRM - Sales & Marketing", "allPermissions": [{"permission": "directly call phone numbers", "type": "Phone"}, {"permission": "read phone status and identity", "type": "Phone"}, {"permission": "read call log", "type": "Phone"}, {"permission": "view Wi-Fi connections", "type": "Wi-Fi connection information"}, {"permission": "find accounts on the device", "type": "Contacts"}, {"permission": "modify your contacts", "type": "Contacts"}, {"permission": "read your contacts", "type": "Contacts"}, {"permission": "precise location (GPS and network-based)", "type": "Location"}, {"permission": "approximate location (network-based)", "type": "Location"}, {"permission": "record audio", "type": "Microphone"}, {"permission": "take pictures and videos", "type": "Camera"}, {"permission": "read the contents of your USB storage", "type": "Storage"}, {"permission": "modify or delete the contents of your USB storage", "type": "Storage"}, {"permission": "find accounts on the device", "type": "Identity"}, {"permission": "read phone status and identity", "type": "Device ID & call information"}, {"permission": "read the contents of your USB storage", "type": "Photos/Media/Files"}, {"permission": "modify or delete the contents of your USB storage", "type": "Photos/Media/Files"}, {"permission": "disable your screen lock", "type": "Other"}, {"permission": "run at startup", "type": "Other"}, {"permission": "create accounts and set passwords", "type": "Other"}, {"permission": "full network access", "type": "Other"}, {"permission": "draw over other app´s", "type": "Other"}, {"permission": "control vibration", "type": "Other"}, {"permission": "view network connections", "type": "Other"}, {"permission": "change your audio settings", "type": "Other"}, {"permission": "use accounts on the device", "type": "Other"}, {"permission": "prevent device from sleeping", "type": "Other"}, {"permission": "read Google service configuration", "type": "Other"}, {"permission": "modify system settings", "type": "Other"}]},

Figura 40. Contenido de la *app* (*com.zoho.crm*) en el fichero *json* que se descarga de *Kaggle* [10]

Se analizan los datos del *dataframe* partiendo de la información sin las etiquetas en las tuplas, es decir, se convierte en el siguiente formato:

{"appId": "com.zoho.crm", "appName": "Zoho CRM - Sales & Marketing", {directly call phone numbers, Phone}, {read phone status and identity, Phone},{read call log, Phone},{view Wi-Fi connections, Wi-Fi connection information},{find accounts on the device, Contacts},{modify your contacts, Contacts},{read your contacts, Contacts}, {precise location (GPS and network-based), Location}, {approximate location (network-based), Location},{ record audio, Microphone}, {take pictures and videos, Camera}, {read the contents of your USB storage, Storage}, {modify or delete the contents of your USB storage, Storage}, {find accounts on the device, Identity}, {read phone status and identity, Device ID & call information}, {read the contents of your USB storage, Photos/Media/Files}, {modify or delete the contents of your USB storage, Photos/Media/Files}, {disable your screen lock, Other}, {run at startup, Other}, {create accounts and set passwords, Other}, {full network access, Other}, {draw over other app´s, Other}, {control vibration, Other}, {view network connections, Other}, {change your audio settings, Other}, {use accounts on the device, Other}, { prevent device from sleeping, Other}, {read Google service configuration, Other}, {modify system settings, Other}}

Figura 41: se transforma y se analizan datos en la *app* "*com.zoho.crm*"

En la *app* "*com.zoho.crm*" se observan los datos:

Permiso	Grupo de permisos
directly call phone numbers	Phone
read phone status and identity	Phone
read call log	Phone
view Wi-Fi connections	Wi-Fi connection information
find accounts on the device	Contacts
modify your contacts	Contacts

read your contacts	Contacts
precise location (GPS and network-based)	Location
approximate location (network-based)	Location
record audio	Microphone
take pictures and videos	Camera
read the contents of your USB storage	Storage
modify or delete the contents of your USB storage	Storage
find accounts on the device	Identity
read phone status and identity	Device ID & call information
read the contents of your USB storage	Photos/Media/Files
modify or delete the contents of your USB storage	Photos/Media/Files
disable your screen lock	Other
run at startup	Other
create accounts and set passwords	Other
full network Access	Other
draw over other app´s	Other
control vibration	Other
view network connections	Other
change your audio settings	Other
use accounts on the device	Other
prevent device from sleeping	Other
read Google service configuration	Other
modify system settings	Other

Tabla 17: análisis de los datos del *dataframe* a partir de la colección *AppPermissionGoogleplay*

La *app* muestra datos que se repiten, es decir, el mismo grupo de permisos pertenece a más de un permiso.

En la *app* "*com.zoho.crm*" el grupo de permisos se eliminan los repetidos que contienen: *Phone, Wi-Fi connection information*, *Contacts, Location, Microphone, Camera, Storage, Identity, Device ID & call information, Photos/Media/Files y Other*.

	Grupo de permisos
1	Calendar
2	Camera
3	Cellular data settings
4	Contacts
5	Device & app history
6	Device ID & call information
7	Identity
8	Location
9	Microphone
10	Other
11	Phone
12	Photos/Media/Files
13	SMS
14	Storage
15	Wearable sensors/Activity data
16	Wi-Fi connection information

Tabla 18. Grupos de permisos (“type”) en la colección *AppPermissionGooglePlay*

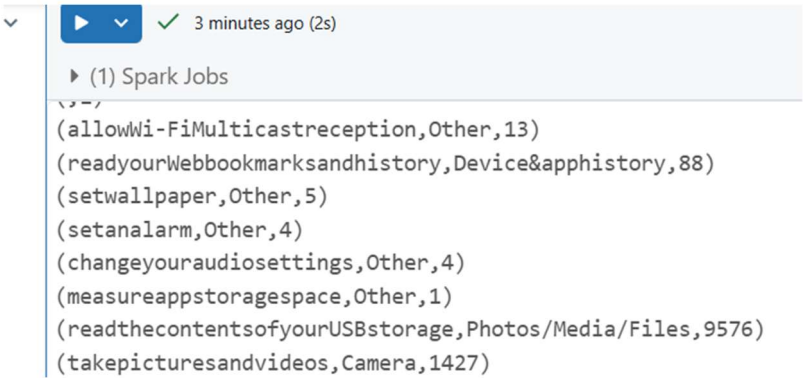

Figura 42. Análisis de las tuplas (*permission,type*) en *Databricks*

Se realiza la limpieza de los datos, se obtiene en cada registro el id de la app y los diferentes grupos de permisos que se declaran, se eliminan los repetidos.

E.3 Modelo dimensional

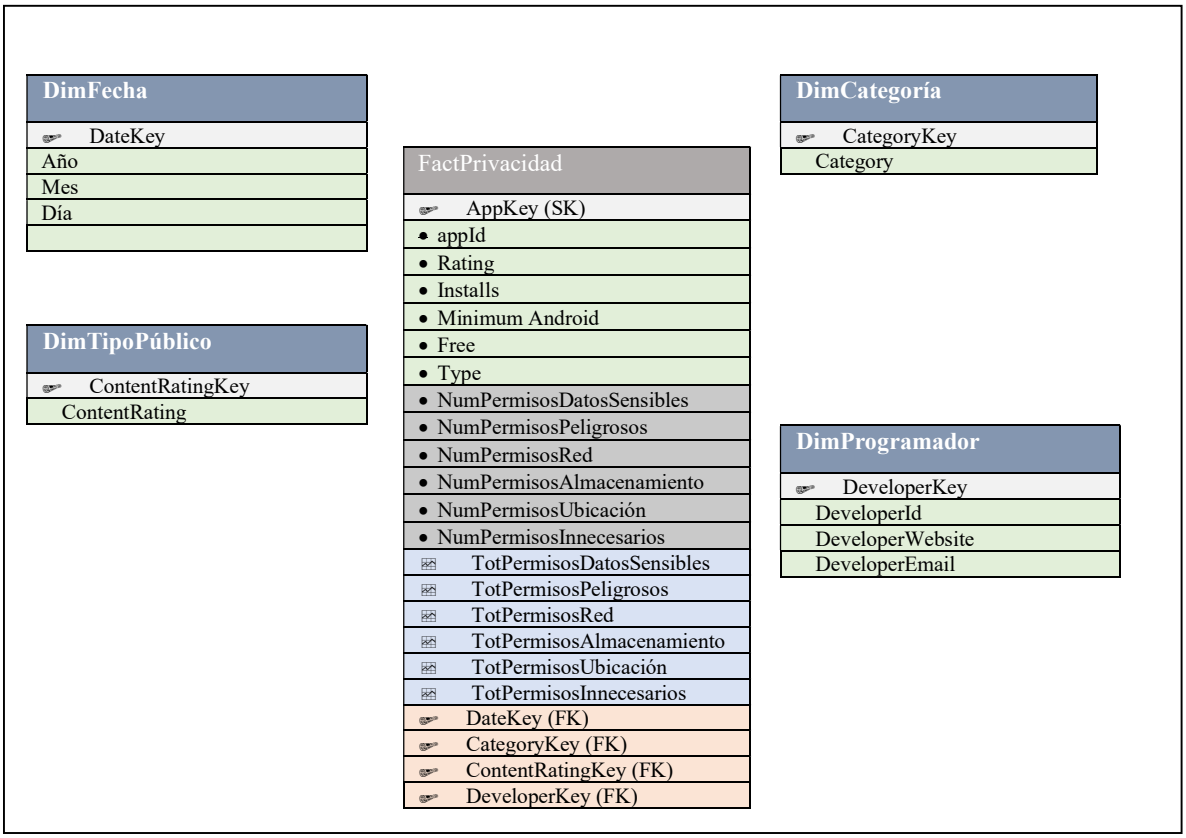

Figura 43. Modelo dimensional en estrella

E.4 Modelo físico

Tabla FactPrivacidad		
Nombre de Campo	**Tipo**	**Comentario**
AppKey	Cadena de caracteres.	Clave primaria
CategoryKey	Entero.	Clave foránea a DimCategory
ContentRatingKey	Entero.	Clave foránea a DimContentRating
DeveloperId	Entero.	Clave foránea a DimDeveloper

DateKey	Fecha.	Clave foránea a DimDate
appId	Cadena de caracteres.	
Rating	Entero.	
Installs	Cadena de caracteres.	
Free	Cadena de caracteres.	
Minimum Android	Cadena de caracteres.	
Policy	Cadena de caracteres.	
Type	Cadena de caracteres.	16 valores diferentes ver tabla 26

Tabla 19. Tipo e identificadores de la tabla FactPrivacidad

Tabla DimCategoría		
Nombre de Campo	**Tipo**	**Comentario**
CategoryKey	Cadena de caracteres.	Clave primaria
Category	Cadena de caracteres.	

Tabla 20. Tipo e identificadores de la tabla DimCategoría

Tabla DimTipoPúblico		
Nombre de Campo	**Tipo**	**Comentario**
ContentRatingKey	Cadena de caracteres.	Clave primaria
ContentRating	Cadena de caracteres.	

Tabla 21. Tipo de Identificadores de la tabla DimTipoPúblico

Tabla DimProgramador		
Nombre de Campo	**Tipo**	**Comentario**
DeveloperKey	Cadena de caracteres.	Clave primaria
DeveloperWebsite	Cadena de caracteres.	
DeveloperEmail	Cadena de caracteres.	

Tabla 22. Tipo de Identificadores de la tabla DimProgramador

Tabla DimFecha		
Nombre de Campo	**Tipo**	**Comentario**
DateKey	Cadena de caracteres.	Clave primaria
Año	Entero.	Año
Mes	Cadena de Caracteres.	Mes
Día	Entero.	Día

Tabla 23. Tipo de Identificadores de la tabla DimFecha

E.5 Data Mart de Privacidad

Fichero Privacidad.csv		**Tabla: Hechos**
Nombre de Campo	**Tipo**	**Comentario**
AppKey	Entero.	Identificador
CategoryKey	Entero.	Clave foránea al fichero Categoría.csv
ContentRatingKey	Entero.	Clave foránea al fichero TipoPúblico.csv
DeveloperKey	Cadena de caracteres.	Clave foránea al fichero Programador.csv

DateKey	Date.	Clave foránea al fichero Fecha.csv
appId	Cadena de Caracteres	
Rating	Entero.	
Installs	Entero.	
Free	Boolean.	
Version	Cadena de caracteres.	
Polícy	Cadena de caracteres.	
		Una columna por cada uno de los grupos de permisos.
Type	Entero.	Valor=1 si en la app utiliza el grupo de permisos

Tabla 24. Características del fichero "Privacidad.csv" del DataMart de Privacidad

Fichero Categoría.csv		**Tabla: Dimensión**
Nombre de Campo	**Tipo**	**Comentario**
CategoryKey	Entero.	Identificador
Nombre	Cadena de caracteres.	Nombre de la Categoría de la app

Tabla 25. Características del fichero "Categoría.csv" del DataMart de Privacidad

Fichero TipoPúblico.csv		**Tabla: Dimensión**
Nombre de Campo	**Tipo**	**Comentario**
ContentRatingKey	Entero.	Identificador
Nombre	Cadena de caracteres.	Nombre del tipo de público

Tabla 26. Características del fichero "TipoPúblico.csv" del DataMart de Privacidad

Fichero Programador.csv		**Tabla: Dimensión**
Nombre de Campo	**Tipo**	**Comentario**
DeveloperKey	Cadena de caracteres.	Identificador
PaginaWeb	Cadena de caracteres.	Página Web del programador
CorreoElectronico	Cadena de caracteres.	Correo Electrónico del programador

Tabla 27. Características del fichero "Programador.csv" del DataMart de Privacidad

Fichero Fecha.csv		**Tabla: Dimensión**
Nombre de Campo	**Tipo**	**Comentario**
DateKey	Date.	Identificador
Año	Date.	
Mes	Date.	
Día	Date.	

Tabla 28. Características del fichero "Fecha.csv" del DataMart de Privacidad

Relacionar dato origen – dato destino de los diferentes ficheros: Privacidad.csv, Categoría.csv, TipoPúblico.csv, Programador.csv y Fecha.csv:

Fichero Privacidad.csv		
Nombre Campo Destino	**Tabla de Origen**	**Campo en Tabla Origen**
AppKey	← FactPrivacidad	AppKey
CategoryKey	← FactPrivacidad	CategoryKey

ContentRatingKey	← FactPrivacidad	ContentRatingKey
DeveloperKey	← FactPrivacidad	DeveloperKey
DateKey	← FactPrivacidad	DateKey
appId	← FactPrivacidad	App Id
Valoración	← FactPrivacidad	Ratings
Descargas	← FactPrivacidad	Installs
Gratuita	← FactPrivacidad	Free
Version	← FactPrivacidad	Minimum Android
Política	← FactPrivacidad	Policy
GrupoPermisos	← FactPrivacidad	Type

Tabla 29. Relación dato origen – dato destino del fichero "Privacidad.csv" del Datamart de Privacidad.

Fichero Categoría.csv		
Nombre Campo Destino	**Tabla de Origen**	**Campo en Tabla Origen**
CategoryKey	← DimCategoría	CategoryKey
Nombre	← DimCategoría	Category

Tabla 30. Relación dato origen – dato destino del fichero "Categoría.csv" del Datamart de Privacidad.

Fichero TipoPúblico.csv		
Nombre Campo Destino	**Tabla de Origen**	**Campo en Tabla Origen**
ContentRatingKey	← DimTipoPúblico	ContentRatingKey
Nombre	← DimTipoPúblico	

Tabla 31. Relación dato origen – dato destino del fichero "TipoPúblico.csv" del Datamart de Privacidad.

Fichero Programador.csv		
Nombre Campo Destino	**Tabla de Origen**	**Campo en Tabla Origen**
DeveloperKey	← DimProgramador	DeveloperKey
PaginaWeb	← DimProgramador	DeveloperWebsite
CorreoElectronico	← DimProgramador	DeveloperEmail

Tabla 32. Relación dato origen – dato destino del fichero "Programador.csv" del Datamart de Privacidad.

Fichero Fecha.csv		
Nombre Campo Destino	**Tabla de Origen**	**Campo en Tabla Origen**
DateKey	← DimFecha	DateKey
Año	← DimFecha	
Mes	← DimFecha	
Día	← DimFecha	

Tabla 33. Relación dato origen – dato destino del fichero "Fecha.csv" del Datamart de Privacidad.

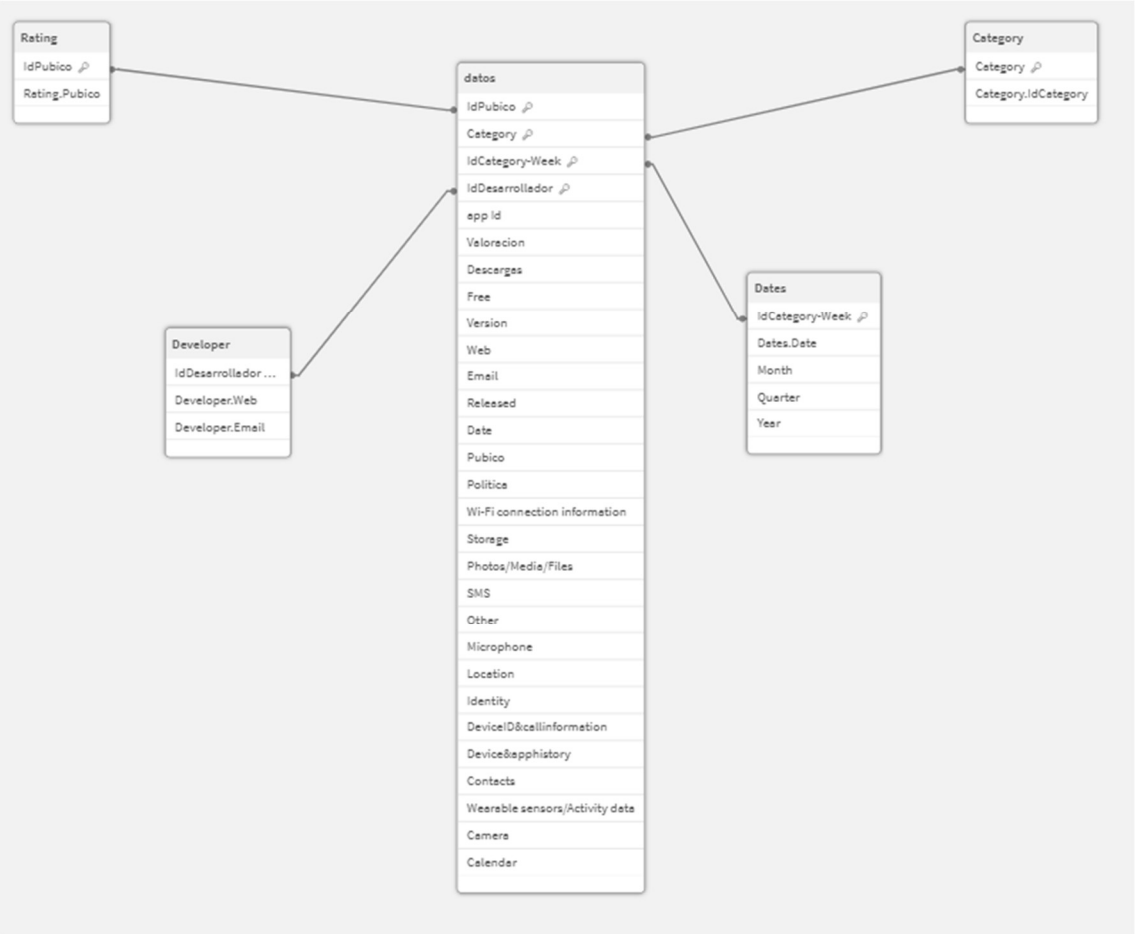

Figura 44. Modelo físico de los datos