

**Escola Tècnica Superior d'Enginyeria  
Electrònica i Informàtica La Salle**

Treball Final de Màster

Màster Universitari en Enginyeria de Telecomunicació

**Data Science en Auditoria de Sistemas**

**Alumne**

Daniel Gómez García

**Professor Ponent**

Rosa Maria Alsina Pagès

**Professor Coponent**

Joan Lluís Pijoan Vidal



---

# ACTA DE L'EXAMEN DEL TREBALL FI DE CARRERA

---

Reunit el Tribunal qualificador en el dia de la data, l'alumne

D.

va exposar el seu Treball de Fi de Carrera, el qual va tractar sobre el tema següent:

Acabada l'exposició i contestades per part de l'alumne les objeccions formulades pels Srs. membres del tribunal, aquest valorà l'esmentat Treball amb la qualificació de

Barcelona,

VOCAL DEL TRIBUNAL

VOCAL DEL TRIBUNAL

PRESIDENT DEL TRIBUNAL





# Resum

Aquest Projecte Fi de Màster té com a objectiu acostar l'àmbit i el coneixement de l'Auditoria financera dins de l'àrea pertanyent a les Tecnologies de la Informació pel fet que els sistemes de Tecnologies de la Informació intervenen cada vegada més i de manera significativa en els processos de negoci de cada entitat.

Aquesta presència dels sistemes TU, porta a l'existència d'Auditories en les empreses, a fi de garantir que aquestes entitats reflecteixin una imatge fidel dels seus estats financers. Sorgeix així, la necessitat de cobrir una sèrie de requeriments i procediments que compleixin amb les Normes Internacionals d'Auditoria.

Seguint la metodologia de treball establerta, s'han realitzat diversos casos d'estudi amb la finalitat de poder realitzar un seguiment de tots els moviments econòmics realitzats per una entitat al llarg del període fiscal auditat davant un possible risc de frau i l'aportació de valor en la informació econòmica de la qual ja disposen les empreses però que en determinades ocasions només contemplen el seu emmagatzematge.

# Resumen

Este Proyecto Fin de Máster tiene como objetivo acercar el ámbito y el conocimiento de la Auditoría financiera dentro del área perteneciente a las Tecnologías de la Información debido a que los sistemas de Tecnologías de la Información intervienen cada vez más y de manera significativa en los procesos de negocio de cada entidad.

Esta presencia de los sistemas TI, lleva a la existencia de Auditorías en las empresas, con objeto de garantizar que dichas entidades reflejen una imagen fiel de sus estados financieros. Surge así, la necesidad de cubrir una serie de requerimientos y procedimientos que cumplan con las Normas Internacionales de Auditoría.

Siguiendo la metodología de trabajo establecida, se han realizado diversos casos de estudio con el fin de poder realizar un seguimiento de todos los movimientos económicos realizados por una entidad a lo largo del período fiscal auditado ante un posible riesgo de fraude y la aportación de valor en la información económica de la que ya disponen las empresas pero que en determinadas ocasiones sólo contemplan su almacenamiento.

# Abstract

This Master's Final Project aims to bring the scope and knowledge of the Financial Audit closer together within the area belonging to Information Technology because Information Technology systems are increasingly and significantly involved in each entity's business processes.

This presence of IT systems leads to the existence of Audits in companies, in order to ensure that these entities reflect a faithful image of your financial statements. Thus arises, the need to cover a number of requirements and procedures that comply with the International Auditing Standards.

Following the established working methodology, a number of case studies have been carried out in order to be able to track all economic movements made by an entity over the fiscal period audited in the face of a potential risk of fraud and the contribution of value in the economic information already available to companies but sometimes only contemplate its storage.







# Contenido

1	Introducción .....	3
1.1	Motivación.....	3
1.2	Objetivos .....	3
1.3	Organización de la memoria .....	3
2	Estado del arte.....	5
2.1	Auditoría de Sistemas .....	5
2.1.1	Objetivos de la auditoría de sistemas .....	6
2.2	Análisis de datos .....	6
2.2.1	Técnicas ADA.....	7
2.2.2	Pruebas de auditoría .....	8
2.2.3	Tipos de datos en auditoría .....	9
3	Diseño.....	11
3.1	Metodología en la auditoría.....	11
3.2	Planificación.....	12
3.2.1	Entendimiento de TI .....	12
3.2.2	Risk Assesment .....	12
3.2.3	Walkthrough de los proceso de negocio .....	12
3.2.4	Herramientas utilizadas .....	13
3.3	Contabilidad en auditoría IT.....	17
3.3.1	Conceptos generales .....	17
3.4	Data Science.....	18
4	Desarrollo y análisis de resultados .....	25
4.1	Caso de estudio 1.....	25
4.2	Caso de estudio 2.....	37
5	Presupuesto.....	51
6	Conclusiones y trabajo futuro .....	53
6.1	Conclusiones .....	53
6.2	Trabajo futuro .....	54
7	Referencias bibliográficas.....	55

# **Acrónimos**

*ADA: Advanced Display Automanagement.*

*API: Application Program Interface.*

*CAAT: Computer Assisted Audit Technique.*

*CRM: Customer Relationship Management.*

*DAX: Data Analysis Expressions.*

*ERP: Enterprise Resource Planning.*

*IAASB: International Auditing and Assurance Standards Board.*

*IoT: Internet of Things.*

*ISA: International Standard on Auditing.*

*IT: Information Technology.*

*NIA: Normas Internacionales de Auditoría.*

*RDBMS: Relational database management system.*

*RGPD: Reglamento General de Protección de Datos.*

*SAP: Systems, Applications and Products.*

*SQL: Structured Query Language.*

*TFM: Trabajo Final de Máster.*

*TI: Tecnologías de la Información.*

# 1 Introducción

Hoy en día, la cantidad de información y datos que se genera en las empresas es inmensa. Actualmente, la información que se es capaz de procesar y tratar dentro de una empresa es casi incalculable. Este suceso hace que las empresas, cada vez en mayor medida, necesiten de herramientas tecnológicas con las que poder trabajar, a través de softwares y equipos informatizados para poder desarrollar sus actividades de manera óptima y eficiente.

Esta presencia de softwares y tecnología, provoca la necesidad de la auditoría de sistemas.

## 1.1 Motivación

Este proyecto surge de la idea de poder abarcar y acercar otros ámbitos técnicos presentes en la actividad laboral con el estudio y análisis de los diferentes datos masivos existentes hoy en día mediante la aplicación de herramientas y metodologías a través de *Data Science*, de modo que se pueda aportar valor a la información.

## 1.2 Objetivos

El objetivo de este Trabajo Final de Máster, es el acercamiento a los usuarios de los conocimientos y actividades relacionadas con la Auditoría de Sistemas, que se encuentra presente dentro de las Auditorías Financieras.

## 1.3 Organización de la memoria

La memoria consta de los siguientes capítulos:

- Capítulo 1: Introducción
- Capítulo 2: Estado del arte
- Capítulo 3: Diseño
- Capítulo 4: Desarrollo y análisis de resultados
- Capítulo 5: Presupuesto
- Capítulo 6: Conclusiones y trabajo futuro
- Capítulo 7: Referencias bibliográficas



## 2 Estado del arte

### 2.1 Auditoría de Sistemas

¿En qué consiste una auditoría de sistemas?

El objetivo principal que tiene la auditoría es la capacidad de validar la integridad de la información y datos almacenados en las bases de datos de los sistemas de información de cada empresa, así como su procesamiento. Es un tipo de auditoría que va más allá del factor económico.

Dicha auditoría consta de la revisión y evaluación de los controles y sistemas de información de las entidades, incluyendo su utilización, eficiencia y seguridad en la empresa que realiza el proceso de la información tratada. Uno de los beneficios de la auditoría de sistemas como alternativa de control, la revisión y su seguimiento permite que el proceso informático y las tecnologías se realicen de un modo más eficiente y seguro, mejorando una adecuada toma de decisiones.

En definitiva, la auditoría de sistemas consiste en:

- Identificar controles que procesen la información e instalación de sistemas, con el objetivo de poder evaluar su efectividad y presentar también alguna recomendación y o posibilidad de mejora.
- Identificar y juzgar de manera objetiva la información tratada.
- Examinar y evaluar los procesos en cuanto a l tratamiento de los datos se refiere. También evaluar la cantidad de recursos invertidos, la rentabilidad de cada proceso, así como su eficacia y eficiencia.

El análisis y la evaluación realizada a través de la auditoría de sistemas debe ser objetiva, crítica, sistemática e imparcial.

El informe de auditoría final deberá ser un claro ejemplo de la realidad de la empresa en cuanto a los procesos y la informatización se refiere, ya que nos proporciona una imagen fiel de la situación actual de la empresa de cara a los accionistas o los estados financieros.

## 2.1.1 Objetivos de la auditoría de sistemas

En el ámbito empresarial, cada vez se encuentra una mayor presencia tecnológica que hace necesario un sistema de control que implique un seguimiento y posterior análisis, siendo necesaria la auditoría de sistemas.

Uno de los primeros factores más importantes, es procurar garantizar la seguridad durante el tratamiento de los datos de información, proporcionando privacidad y confidencialidad, así como su buen uso evitando la manipulación de la misma. La utilización del sistema informático es necesario para obtener un proceso más eficiente y rentable que nos permita detectar posibles errores para una mejor toma de decisiones de modo inmediato.

Algunos de los objetivos de la auditoría de sistemas son:

- Agilización en la toma de decisiones
- Aumentar la seguridad de la información de los usuarios de los sistemas
- Procurar la confidencialidad e integridad de los datos a través de los sistemas de seguridad y el entorno de control
- Minimizar la existencia de riesgos en la manipulación del tratamiento de la información
- Educar a los usuarios sobre el control de los sistemas de información, al tratarse de un sector muy dinámico que requiere de un conocimiento actualizado para poder cumplir con las legislaciones vigentes

A modo de resumen, la auditoría de sistemas es una manera de controlar y evaluar los sistemas informáticos de cada entidad, incluyendo todo tipo de información que afecte a los estados financieros de la empresa.

Actualmente la auditoría de sistemas es clave y necesaria para poder garantizar la seguridad y el tratamiento de la información de los sistemas informáticos de una empresa.

## 2.2 Análisis de datos

Los auditores, con el objetivo de poder tratar y analizar la información de los sistemas informáticos para obtener las evidencias necesarias que soporten las conclusiones de auditoría, necesitan de herramientas y técnicas de auditoría asistida por ordenador (CAAT) y su análisis de datos de auditoría (ADA).

Las técnicas ADA actualmente se consideran una evolución de las CAATs al permitir al auditor abordar los objetivos de auditoría con mayor eficacia en la visualización de la información pudiendo ser utilizadas a lo largo de todo el proceso de auditoría desde el inicio hasta la emisión del informe de fiscalización.



El Análisis de Datos en la Auditoría lo define el *IAASB*, mediante el *Data Analytics Working Group*, de la siguiente manera:

“Cuando es usado para obtención de evidencia de auditoría, es la ciencia y el arte de descubrir y analizar patrones, desviaciones e inconsistencias y la extracción de otro tipo de información útil de los datos subyacentes relacionados con el asunto en cuestión a auditar a través del análisis, el uso de modelos y la visualización con el propósito de planear o ejecutar el trabajo de auditoría de dicho asunto.”

Desde que las empresas actualizaron sus sistemas de información y sistemas financieros, los auditores necesitan de sistemas informáticos con los que poder realizar pruebas de auditoría de la información extraída de la empresa para poder dar a respuesta a los posibles riesgos como fraude entre otros, que puedan existir en la información contable.

Finalmente, el enfoque de *Data Science* se aloja en la inferencia, es decir, es el proceso de obtención de conclusiones basadas únicamente en los conocimientos del propio auditor.

### **2.2.1 Técnicas ADA**

Existen diferentes técnicas *ADA*, que permiten al auditor abordar los principales objetivos de la auditoría siendo las más utilizadas las siguientes:

- Descriptivas:  
Son las más frecuentes al ser un tipo de análisis que permite describir el pasado utilizando información detallada, para poder responder a preguntas como, ¿qué ha ocurrido?
- De evaluación:  
Esta técnica se basa fundamentalmente en procurar analizar diferentes perspectivas en la correlación de los datos para intentar entender el por qué han sucedido los hechos.
- Predictivas:  
Se trata de modelos estadísticos para poder responder sobre posibles tendencias futuras.
- Prescriptivas:  
Se basan en proponer acciones que puedan mejorar algoritmos ya implementados que puedan ayudar a los usuarios a la automatización de sistemas o la mejora en la toma de decisiones.

Algunas de las ventajas que podemos proporcionar con la utilización de estas técnicas serían:

- Análisis de grandes volúmenes de datos en un breve período de tiempo
- Eficiencia en la implementación de los recursos de auditoría
- Análisis de la totalidad de la población de los datos
- Facilidad en la repetición de las pruebas

La utilización de estas técnicas permite al auditor un mayor alcance en la extensión de las pruebas a realizar.

El objetivo de la utilización de dichas técnicas es facilitar al auditor una orientación a la hora de realizar las pruebas de auditoría que nos ayuden a establecer unos criterios homogéneos siguiendo una metodología común, disponer de una adecuada planificación de las pruebas, poder estandarizar la documentación realizada, así como poder automatizar en la medida de lo posible cualquier ejecución para mejorar la eficiencia llevada a cabo.

## 2.2.2 Pruebas de auditoría

Hay diversas pruebas de auditoría, aunque hoy en día, algunas de las más importantes llevadas a cabo son:

- Integridad de la información:  
Esta prueba, se encuentra formada por un conjunto de pruebas sustantivas que nos permitan evaluar la completitud y la exactitud de los datos de información, así como su integridad y consistencia.  
Estas pruebas resultarán como un indicador sobre la fortaleza o debilidad en el diseño de los controles así como su implementación en el procesado y tratamiento de la información durante las rutinas de validación.
- Muestreo:  
Se realizan seleccionando una muestra de manera aleatoria o mediante estadística aplicada para poder llevar a cabo diferentes planteamientos posibles y ejecutar la comprobación sobre una cantidad limitada de la totalidad de la población de datos, de modo, que permita obtener la seguridad necesaria en la opinión emitida y sustentada por el auditor.

El auditor puede solicitar la extracción de una muestra aleatoria sobre la totalidad de las transacciones almacenadas en la base de datos del sistema para poder realizar las validaciones necesarias sobre las pruebas realizadas.

## 2.2.3 Tipos de datos en auditoría

### 2.2.3.1 Datos estructurados

Los datos estructurados suelen encontrarse en las bases de datos relacionales (*RDBMS*), donde se almacena información como registros de identificación de reservas, albaranes, facturas, información relacionada con el cliente, datos personales, nombres, direcciones, cadenas de texto de longitud variable...etc. Toda la información permanece contenida en los registros, siendo una manera fácil de búsqueda y almacenamiento de la misma.

La información puede ser generada tanto de forma humana como de manera mecánica, mientras sean creados siguiendo una estructura de base de datos relacionales. Este formato nos ayuda a poder realizar consultas de búsqueda de la información, sea de manera manual o mediante algoritmos en función del tipo de datos y los campos correspondientes.

Algunas de las aplicaciones comunes basadas en bases de datos relaciones con datos estructurados pueden incluir sistemas de reservas de alojamientos en hoteles o aerolíneas, transacciones bancarias, control de inventario sobre los almacenes de una entidad, todo quedaría registrado y almacenado en la base datos.

Para poder realizar búsquedas o consultas de los datos estructurados dentro de las bases de datos relacionales, se utiliza el lenguaje de consulta estructurado (*SQL*).

Aunque la mayoría de los datos de *CRM* se encuentran estructurados, existen bases de datos relacionales que almacenan o apuntan a datos no estructurados, como por ejemplo, las aplicaciones relacionadas con la gestión de clientes, siendo más complicada las consultas de búsqueda en comparación con las bases de datos tradicionales.

Como hemos mencionado anteriormente, las técnicas *CAAT* son técnicas que resultan muy útiles al tratar la información de cada empresa siempre y cuando los datos se encuentren disponibles en bases de datos estructuradas, ya que permite al auditor realizar pruebas sobre los controles o pruebas sustantivas con mayor facilidad para el cumplimiento de las auditorías financieras.

Sin embargo, debido a la gran expansión y generación de grandes conjuntos o volúmenes de información y datos no estructurados como puede ser el *Big Data* o el *IoT*, provoca una limitación para trabajar con las técnicas *CAAT*.

En este punto, y con vistas a un futuro no muy lejano los auditores ya empezaríamos a contemplar las tecnologías basadas en la inteligencia artificial.

### **2.2.3.2 Datos no estructurados**

El resto de información o datos no estructurados, son aquellos que aunque tienen una estructura interna asociada, pero no se siguen un modelo de datos o esquema de manera predefinida pudiendo ser almacenados en bases de datos no relacionales como *NoSQL*, y siendo generados tanto de manera manual por el usuario o de manera automática por el sistema.

Las auditorías financieras o de sistemas, actualmente no se encuentran trabajando con este tipo de datos ya que, los principales sistemas de información de las entidades trabajan con datos estructurados dentro de bases de datos relacionales como por ejemplo el *ERP SAP* o *AS400* entre otros. Aunque sí se contempla la puesta en marcha del trabajo de datos no estructurados en un futuro próximo.

## 3 Diseño

### 3.1 Metodología en la auditoría

Para poder llevar a cabo una metodología adecuada es necesario entender algunos conceptos fundamentales como son las Normas Internacionales de Auditoría conocidas como NIA o /ISA.

Estas normas son estándares o reglas profesionales que tratan con las responsabilidades del auditor al realizar la inspección o auditoría financiera de la información facilitada por la entidad económica auditada.

El principal objetivo de la aplicación de estas reglas en una auditoría no es otro que el poder aumentar la confianza de los usuarios o accionistas en los estados financieros de la entidad. Este hecho se lleva a cabo mediante la opinión o expresión que el auditor indica sobre dichos estados financieros para comprobar si cumplen o no en todos los aspectos materiales y siguiendo el marco de información financiera aplicable.

El fin de general de dicha opinión sobre los marcos de información viene a referirse a si los estados financieros de la entidad se encuentran reflejando una imagen fiel a los aspectos materiales legislados. El auditor permite formar dicha opinión a través de dicha auditoría basada y conforme a las normativas y requerimientos de las NIA.

El auditor debe lograr una seguridad razonable a través de los requerimientos y objetivos recogidos en las NIA, sirviendo como una guía de aplicación o anotación explicativa.

Uno de los requisitos recogidos en las NIA, es que el propio auditor debe aplicar su juicio y escepticismo profesional durante la planificación y ejecución de la auditoría, además de los siguientes apartados:

- Identificar y valorar los riesgos de incorrección material, debido a fraude o error, basándose en el conocimiento de la entidad y de su entorno, incluido el control interno de la entidad.
- Obtener la evidencia de auditoría suficiente y adecuada sobre si existen incorrecciones materiales, mediante el diseño y la implementación de respuestas adecuadas a los riesgos valorados.
- Formar una opinión sobre los estados financieros basada en las conclusiones alcanzadas a partir de la evidencia de auditoría obtenida.

## 3.2 Planificación

### 3.2.1 Entendimiento de TI

Como primer paso en la auditoría, el auditor tiene que ser capaz de identificar y valorar el conocimiento de la entidad y de su entorno.

Debe conocer las políticas y los procedimientos implementados por la empresa, así como la infraestructura propia de las TI (hardware, sistemas operativos, etc.) y las aplicaciones de software implicadas en la operativa de negocio con el fin de poder obtener una base sobre el diseño y la implementación de los riesgos valorados de incorrección material.

### 3.2.2 Risk Assessment

La evaluación de los riesgos es clave para la realización de todas las auditorías de los estados financieros. La idea de un enfoque "basado en el riesgo" para la auditoría se refiere al enfoque del proceso de auditoría en aquellas áreas que están en mayor riesgo de error material.

Para ello, hay que hacer un énfasis especial en las siguientes características:

- *Completeness:*  
Toda la información que debería haber sido registrada en el período contable, ha sido registrada.
- *Existence:*  
Los activos, pasivos, y el saldo de capital que aparecen en los estados financieros de una empresa, realmente existen.
- *Accuracy:*  
Las transacciones se registraron por las cantidades apropiadas.
- *Valuation:*  
Las cifras presentadas en un estado financiero son precisas y se basan en una valoración adecuada de los activos, pasivos y saldos de patrimonio.
- *Ocurrence:*  
Las transacciones realmente han ocurrido.

### 3.2.3 Walkthrough de los proceso de negocio

Para establecer una gestión basada en procesos en la organización se tienen que identificar todos los procesos relevantes que participan dentro del alcance de los estados financieros de la entidad.

El mapa de procesos es la representación gráfica, un diagrama que nos proporcionará un entendimiento del funcionamiento de los procesos y subprocesos de la entidad y nos dará la visibilidad necesaria sobre los riesgos y controles de los que la empresa dispone.

### 3.2.4 Herramientas utilizadas

Para poder trabajar con los datos estructurados de las bases de datos relacionales, utilizaremos la herramienta o software llamada IDEA.

IDEA es un software de la firma CASEWARE diseñada por y para ayudar los auditores y profesionales dedicados al ámbito financiero, para poder realizar análisis de datos de una manera rápida y eficaz, permitiendo analizar el 100% de la población de datos, y garantizando la integridad de los mismos a través de las tareas correspondientes de auditoría.

Las principales funcionalidades que nos permite este software es poder importar ficheros de datos sin importar el formato en el que éstos se encuentren, el único inconveniente que conllevan, es que, en función del formato de origen hay que aplicar una serie de criterios y reglas para poder adaptarlos a un modo legible y visual con el que podamos realizar el análisis correspondiente.

A estas extracciones de datos se les podrá aplicar un filtrado para poder quedarnos con la información relevante, podremos comprobar las distintas posibilidades existentes en un campo, podremos detectar celdas vacías en facturas, duplicidades dentro del mismo fichero de información, etc.

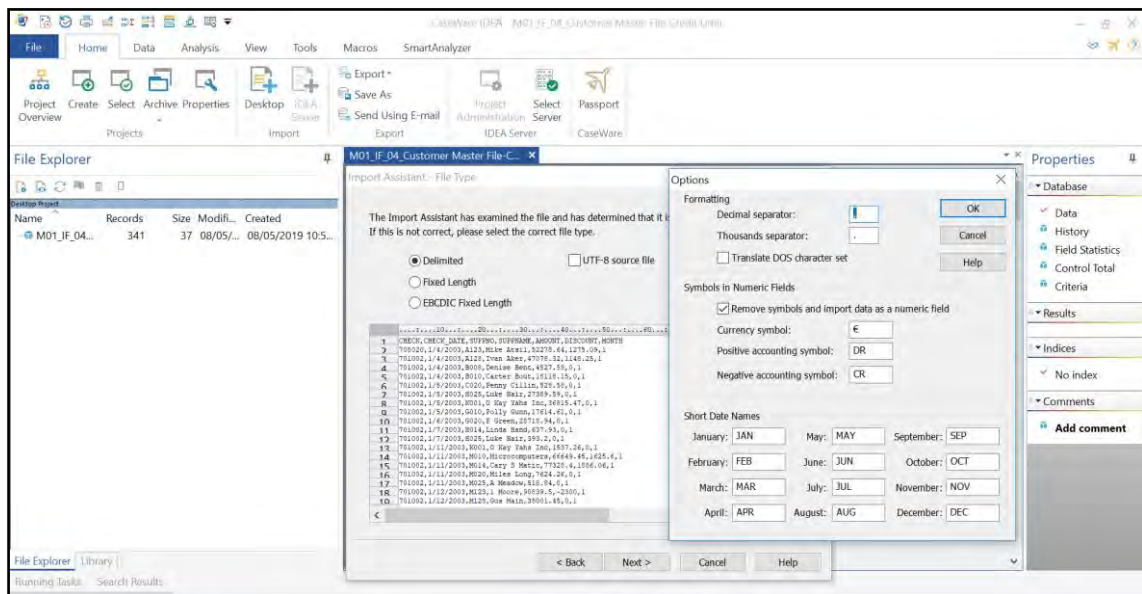


Figura 1. Importación de un fichero txt.

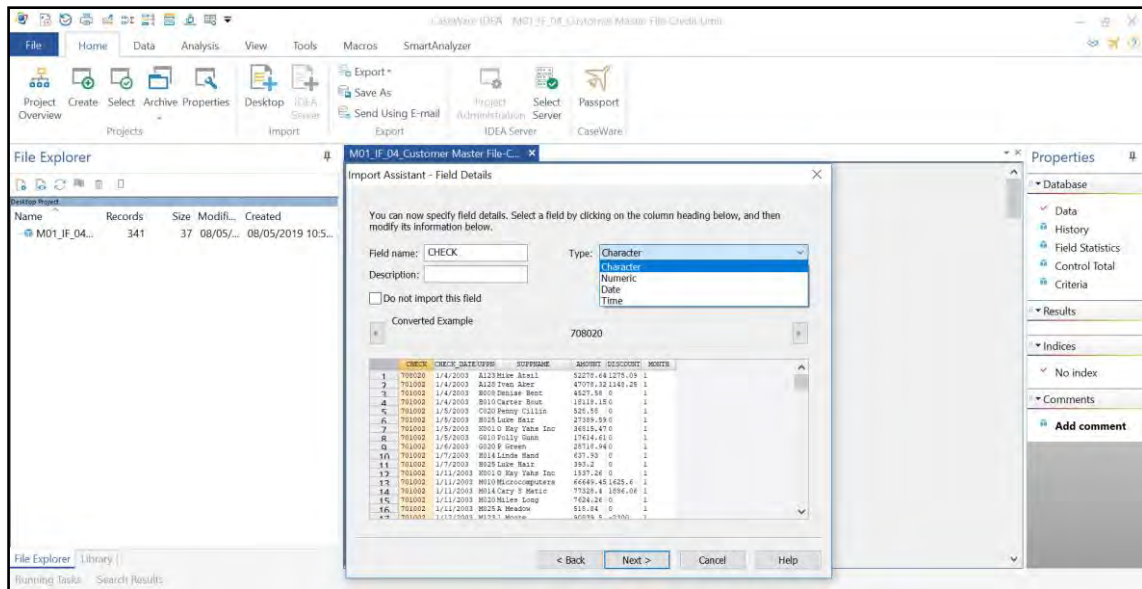


Figura 2. Definición del tipo de campos.

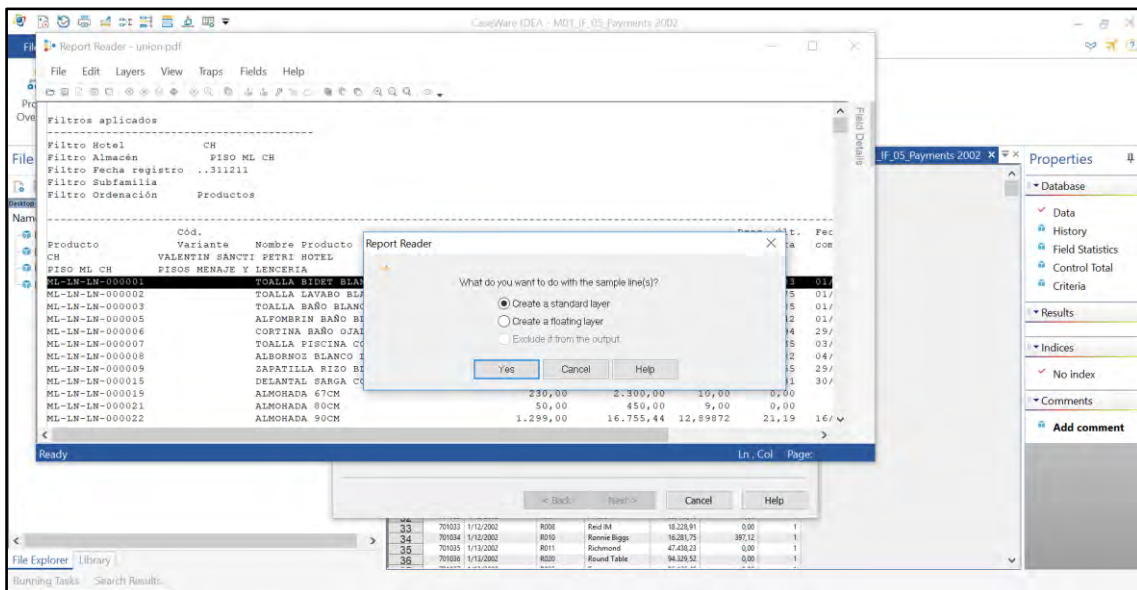


Figura 3. Importación de un archivo pdf.



Cód.	Variante	Nombre Producto	Stock Actual	Importe Coste	PMP	Prec. últ.	Fec. últ.	Prec. últ.	Fec. últ.
		TOALLA BIOTER BLANCO LOGO 28X4	2.132,00	1.589,89	0,74573	0,83	01/09/11	1,00	01/03/07
		TOALLA LAVABO BLANCO LOGO 50X	5.940,00	14.823,38	2,49552	2,75	01/09/11	3,00	01/03/07
		TOALLA BAÑO BLANCO LOGO 100X1	5.416,00	35.512,14	6,55689	8,85	01/09/11	6,00	01/03/07
		ALFOMBRIN BAÑO BLANCO 45X60	2.212,00	4.072,59	1,84113	2,42	01/09/11	1,80	01/03/07
		CORTINA BAÑO GUALES Y LOGO 18	1.412,00	17.546,53	12,42672	18,94	29/03/10	0,00	26/11/07
		TOALLA PISCINA COLIE CELESTE 1	1.699,00	14.376,96	8,44699	10,35	03/11/11	10,00	01/03/07
		ALBORNOS BLANCO LOGO	699,00	7.740,63	11,07386	10,82	04/10/10	0,00	01/11/07
		ZAPATILLA RIZO BLANCO	446,00	468,30	1,05	1,55	29/09/11	0,00	
		DELANTAL SARGA CON PETO BLAN	159,00	522,98	3,31	3,31	30/05/08	3,31	01/03/07
		ALMOHADA 67CM	230,00	2.300,00	10,00	0,00		10,00	01/03/07
		ALMOHADA 80CM	50,00	450,00	9,00	0,00		9,00	01/03/07
		ALMOHADA 90CM	1.299,00	16.755,44	12,89072	21,19	16/09/11	13,00	01/03/07
		FUNDA ALMOHADA BLANCO 1 BOCA	599,00	1.225,96	2,0501	2,75	02/09/11	1,89	01/03/07
		FUNDA ALMOHADA BLANCO 2 BOCA	3.759,00	10.909,43	2,90299	3,75	11/08/11	2,82	01/03/07
		FUNDA COJIN CUADRANTE BLANCO	60,00	174,00	2,90	0,00		2,90	01/03/07
		FUNDA ALMOHADA CUNA BLANCO 2	360,00	403,20	1,12	0,00		1,12	01/03/07
		SABANA ENCIMERA BLANCO 190X30	3.031,00	25.825,37	8,52041	10,27	02/09/11	8,30	01/03/07
		SABANA ENCIMERA BLANCO 250X30	271,00	3.078,18	11,3586	13,08	11/09/11	10,00	01/03/07
		SABANA ENCIMERA BLANCO 300X30	312,00	4.045,70	12,96689	17,40	11/09/11	12,00	01/03/07
		SABANA BAJERA BLANCO 190X300	2.829,00	17.152,71	6,06317	7,85	11/09/11	6,00	01/03/07
		SABANA BAJERA BLANCO 250X300	316,00	2.543,92	8,05038	9,70	11/09/11	7,80	01/03/07
		SABANA BAJERA BLANCO 300X300	294,00	2.808,39	9,55235	7,66	21/07/10	9,50	01/03/07
		SABANA CUNA BLANCO 105X170	448,00	1.066,16	2,37992	2,26	30/06/09	2,40	01/03/07
		CUBRECOCHEON BLANCO AJUSTAB	1.514,00	22.719,00	15,00	0,00		0,00	16/11/10

Figura 4. Adaptación de un archivo pdf.

Una vez, tenemos la información correctamente importada, se procederá al análisis de la misma.

Para poder llevarlo a cabo, el software tiene un lenguaje propio del editor, que nos ofrece funciones de gran utilidad para el tratado y posterior análisis de la información, siendo algunas de las más relevantes las siguientes:

- **Joins:**  
Función que sirve al igual que en el lenguaje SQL para poder combinar y analizar diferentes bases de datos, existiendo las siguientes opciones:
  - **Matches only:**  
El output contendrá aquellos registros que se encuentren en ambas bases de datos.
  - **Records with no secondary match:**  
El output contendrá aquellos registros que se encuentren en la primera base de datos pero no en la segunda.
  - **Records with no primary match:**  
El output contendrá aquellos registros que se encuentren en la segunda base de datos pero no en la primera.
  - **All records in primary file:**  
El output contendrá todos los registros que se encuentren en ambas bases de datos y los de la primera base de datos.
  - **All records in both files:**  
El output contendrá todos los registros de ambas bases de datos.

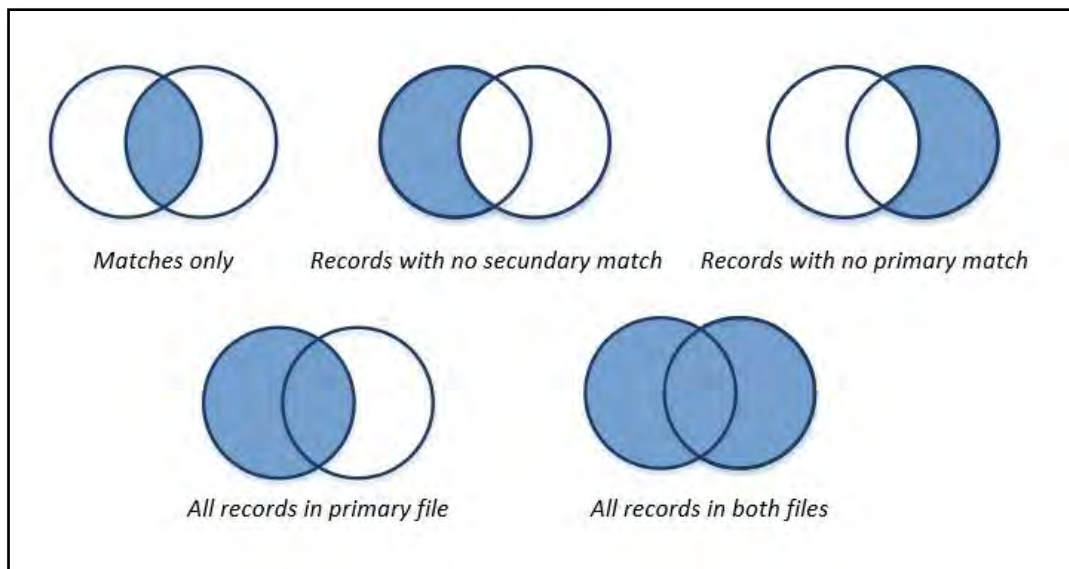


Figura 5. Tipos de *Join*.

- *Append:*  
Esta función se utiliza para unir bases de datos o insertar campos en una de ellas como puede ser la *query INSERT* de *SQL*.
- *Summarization:*  
La función de sumariación es muy para agrupar y crear registros en cualquier base de datos, siendo similar a la *query SUM* de *SQL*.
- *Stratification:*  
Esta función permite analizar valores de grupo dentro de unos límites establecidos.
- *Duplicate key:*  
Hay dos variantes en función de si queremos encontrar registros que tengan campos idénticos o especificar cuáles de esos campos deben ser iguales y cuáles distintos.

También podemos aplicar criterios o especificar filtros en función del análisis que queramos llevar a cabo, crear campos virtuales editables, o incluso utilizar el lenguaje propio de la herramienta para análisis concretos.

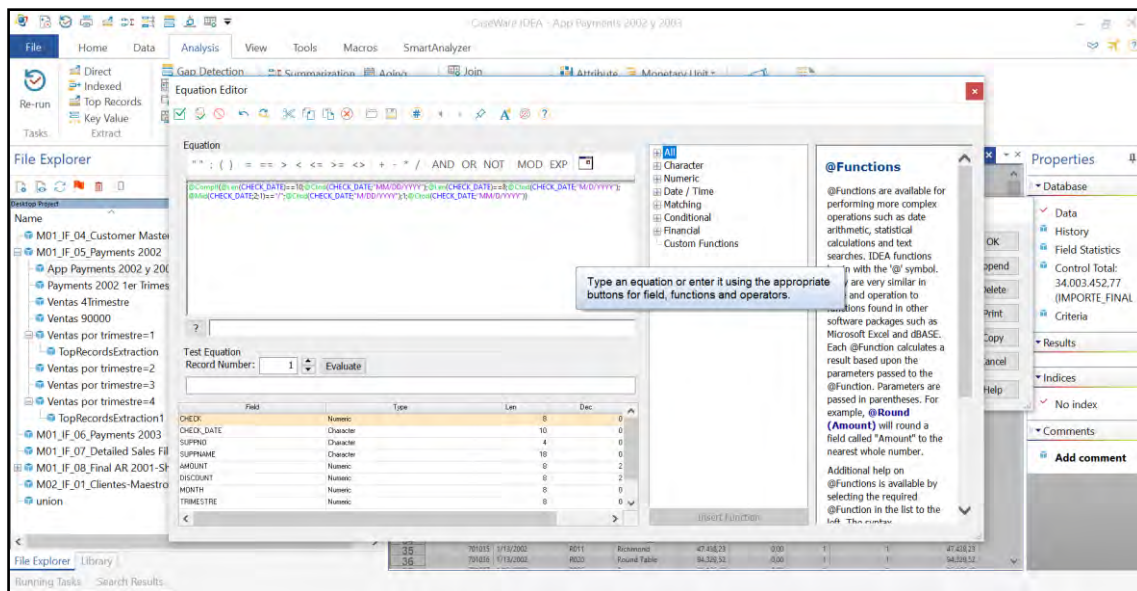


Figura 6. Análisis mediante lenguaje propio.

## 3.3 Contabilidad en auditoría IT

### 3.3.1 Conceptos generales

Antes de poder comenzar con el primer caso de estudio, es necesario mencionar y especificar los principales conceptos contables llevados a cabo dentro de una auditoría financiera.

Las transacciones es todo acontecimiento que afecta económicamente a la empresa como pueden ser operaciones de compra, operaciones de venta, entre otras, siendo reflejadas en asientos contables.

El Libro Diario es todo aquello que genere una transacción u operación económica para la empresa. Todas estas operaciones quedan reflejadas en diferentes cuentas contables que son agrupadas en el Libro Mayor dónde se especifica el detalle de cada movimiento de cada cuenta contable.

Existen diferentes tipos de asientos contables entre los que destacan:

- Inmovilizado
- Proveedores
- Acreedores
- Aprovisionamientos
- Ingresos

Finalmente, cuando nos referimos a los estados financieros de la entidad nos estamos refiriendo a las cuentas anuales de la empresa que se componen del balance y la cuenta de resultados.

El balance es una cuenta oscilante al estar compuesta por los activos, pasivos y el patrimonio neto. Mientras que la cuenta de resultados está compuesta por el registro de los ingresos y los gastos.

A modo de resumen, los asientos contables son el conjunto de anotaciones o apuntes que se hacen en el Libro Diario con el fin de poder registrar un hecho económico, que provoca una modificación cuantitativa o cualitativa en la composición del balance y/o resultado de una empresa, y por tanto, un movimiento en las cuentas contables.

Las cuentas anuales están formadas por el balance, la cuenta de pérdidas y ganancias, cambios de estado en el patrimonio neto, estado de flujo de efectivo y una memoria.

Los principios contables básicos, son los siguientes:

- Empresa en funcionamiento
- Devengo
- Uniformidad
- No compensación
- Importancia relativa en términos cuantitativos o cualitativos
- Prudencia

Otro concepto importante es la materialidad. La materialidad es el margen de error permitido en las correcciones materiales de la empresa que pueden provocar una variación en la opinión emitida por el auditor según su juicio profesional.

## **3.4 Data Science**

### **3.4.1.1 Definición**

Cuando hablamos de *Data Science* o Ciencia de los Datos, nos referimos al estudio de grandes cantidades de datos combinando la estadística, las matemáticas y la informática para su interpretación. Ya que, el objetivo de la misma es la toma de decisiones.

Por ejemplo, la principal diferenciación entre *Big Data* y *Data Science* es que cuando hablamos de *Big Data*, nos referimos a resolver problemas de gestión y almacenamiento de los datos al tener un gran volumen de los mismos, mientras de *Data Science* se refiere a las herramientas que transforman esos datos en valor.

### 3.4.1.2 Campos en Data Science

#### Data mining:

La minería de datos, abarca una gran parte dentro de *Data Science*, y consiste en la búsqueda y obtención de valor en la información donde en un comienzo no parece existir.

Generalmente, suele tratarse con datos no estructurados como fuente de origen y se termina como fuente de destino en información de valor tras aplicar diferentes algoritmos y técnicas de análisis.

La minería de datos procura responder a diferentes tipos de preguntas en sus análisis como pueden ser:

- ¿Cómo afectaría a mis ventas si restaurase el hotel y ofreciese descuentos a los usuarios en redes sociales?
- ¿De qué manera podría clasificar a mis clientes para hacerles llegar información personalizada?
- ¿Con qué porcentaje de seguridad puedo predecir si este año se superarán el número de ventas fijadas como objetivo?

Las principales fases de las que se compone la minería de datos son:

- Obtención de los datos de información  
La recogida de información es parte del proceso y puede realizarse a través de formularios, sensores, redes sociales, APIs, etc. Ya que no siempre se puede disponer de la fuente de información en una base de datos.
- Tratamiento y pre-procesado de la información recibida  
En esta fase, se debe preparar y analizar bien la información sin perder valor en ella, teniendo como objetivo optimizar los algoritmos y reducir costes de ejecución de la siguiente fase.

Los principales métodos llevados a cabo en esta fase serían:

- Normalización
- Cuantificación
- Reducción de la dimensionalidad
- Discreción de variables

- *Machine Learning*  
Estos algoritmos son capaces de predecir y clasificar información tras aprender de datos anteriores ya entrenados. Entrenar el modelo significa alimentar los algoritmos de aprendizaje automático con nuestra información.

Existen algoritmos supervisados que son aquellos que mediante la observación encuentran las relaciones existentes entre variables de entrada y variables de salida ya conocidas o mediante ejemplos.

Mientras que los algoritmos no supervisados son aquellos que pueden proporcionar conocimiento únicamente con los datos que se proporcionan como entrada sin necesidad de conocer el resultado a obtener. Es decir, analiza los datos para identificar patrones mediante la interpretación de los mismos.

Y por último, existen los algoritmos por refuerzo. Estos algoritmos se proporcionan con un conjunto de acciones, parámetros y valores finales. De este modo, el algoritmo intenta determinar cuál resultado es el óptimo entre las diferentes opciones y posibilidades. Por consecuente, aprende mediante el ensayo y error.

Los algoritmos de aprendizaje automático más populares son los siguientes:

- De regresión  
Este algoritmo debe comprender las relaciones existentes entre las variables ya que existe una variable dependiente y el resto cambiantes.
- Bayesianos  
Este algoritmo se basa en el algoritmo de *Bayes*, donde cada valor es independiente de cualquier otro.
- De agrupación  
Este algoritmo se utiliza para la búsqueda de grupos dentro de la información en el aprendizaje no supervisado.
- Árboles de decisión  
Este algoritmo resulta una estructura en forma de árbol donde cada nodo del árbol representa una prueba con una variable específica, y cada rama son los resultados de las pruebas.
- Redes neuronales  
Este algoritmo se distribuye en capas con un gran número de elementos conectados entre sí, que aprenden con el ejemplo y la experiencia y son muy útiles en modelar relaciones no lineales de alta dimensión.
- De reducción de dimensión  
Este algoritmo se basa en reducir variables hasta encontrar la información exacta y requerida.
- De aprendizaje profundo  
Este algoritmo ejecuta datos a través de capas de algoritmos de redes neuronales, las cuáles, facilitan una representación simplificada de la información a la siguiente capa, aprendiendo de este modo y de manera progresiva con cada transferencia entre capas.

- Pruebas y desarrollo  
En términos generales, no existen algoritmos mejores o peores, sino, que depende de cada caso a examinar. Por consiguiente, la obtención de una mejora de resultados conllevará probar varios algoritmos hasta encontrar la mejor solución.
- Interpretación y visualización de los resultados obtenidos

Finalmente, y tras aplicarse los algoritmos correspondientes, se obtiene una representación del conocimiento obtenido buscando la mejor forma de poder visualizar dicha información aportando así el valor de la misma.

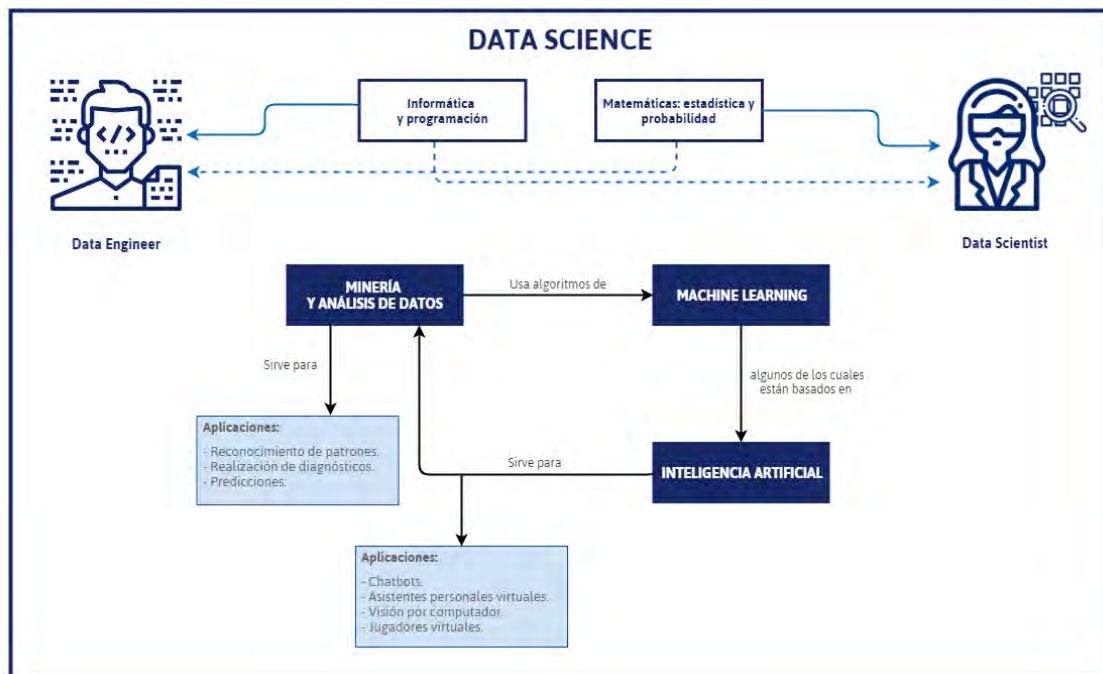


Figura 7. Disciplinas en *Data Science*.

### 3.4.1.3 Microsoft Power BI

Microsoft ofrece con *Power BI* una colección de servicios de software, aplicaciones y conectores que funcionan en conjunto para transformar diversos orígenes de datos en información relevante de una manera visualmente interactiva.

*Power BI* consta de 3 elementos fundamentales en la dinámica de trabajo:

- *Power BI Desktop*: Software de procesamiento de la información importada
- *Power BI Service*: Visualización de los reportes generados
- Aplicaciones móviles: Posibilidad de compartir de la información con otros usuarios

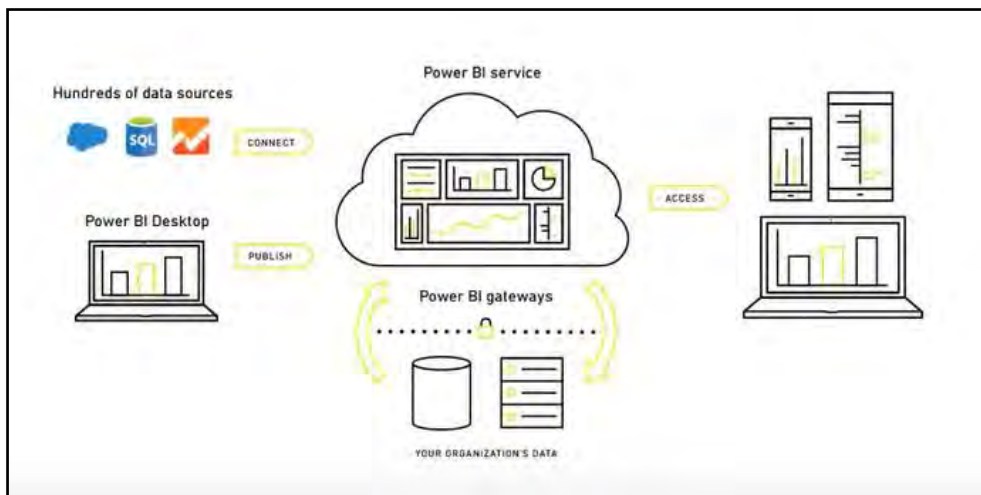


Figura 8. Organización de *Power BI*.

### 3.4.1.4 Bloques de creación

- Visualizaciones:

Los gráficos, mapas u objetos visuales, como su propio nombre indica son todos aquellos elementos para poder representar la información de manera visual.

- Conjunto de datos:

Son todas las colecciones de datos que *Power BI* utiliza para crear las visualizaciones. Los conjuntos de datos pueden ser una combinación de múltiples orígenes de datos con lo que se puedan crear consultas, campos calculados, etc.

- Informes:

Grupo de visualizaciones relacionadas entre sí.



- **Paneles:**  
Colección de objetos visuales de una sola página, a diferencia de los informes, se especifica una perspectiva rápida y general de la información.
- **Iconos:**  
Visualización única dentro de un informe o un panel.

#### **3.4.1.5 Relaciones de datos**

Cuando tenemos un grupo de tablas y queremos un funcionamiento lógico entre ellas, necesitamos una relación entre los datos.

Existen los siguientes tipos de relaciones de modelado:

- **Varios a uno (\*: 1) :**  
La mayoría común, el tipo predeterminado, lo que significa que la columna de una tabla puede tener más de una instancia de un valor y la otra tabla relacionada, a menudo conocida como tabla de búsqueda, solo tiene una instancia de un valor.
- **Uno a uno (1:1) :**  
La columna de una tabla solo tiene una instancia de un valor determinado, y la otra tabla relacionada solo tiene una instancia de un valor determinado.
- **Relaciones de varios a varios:**  
Con modelos compuestos, puede establecer relaciones de varios a varios entre tablas, lo que elimina los requisitos para valores únicos en tablas. También permite descartar las soluciones alternativas anteriores, como el hecho de presentar nuevas tablas solo para establecer relaciones.

#### **3.4.1.6 Columnas calculadas**

Cuando hablamos de columnas calculadas se define como cálculos que combinan o transforman datos de dos o más elementos existentes en una tabla.

Para poder llevar a cabo algunas de las transformaciones, se necesitan expresiones del tipo *DAX*.



## 4 Desarrollo y análisis de resultados

### 4.1 Caso de estudio 1

Nuestro primer caso de estudio y prueba realizada, estará enfocado al control de los asientos contables realizados por una importante entidad perteneciente al ámbito de la cosmética y productos de belleza. Esta realización de la prueba, tiene una importancia significativa y trascendental, ya que se deben reflejar todos los movimientos económicos realizados por la empresa a lo largo del período fiscal auditado, existiendo un posible riesgo relacionado con el fraude.

Para ello, siguiendo nuestra metodología aplicada, se ha solicitado a la entidad la extracción de los *Journal Entries* del sistema de la entidad. El personal responsable del área de IT, nos ha informado, que disponen de transacciones personalizadas para su propio sistema, también conocidas como transacciones *custom*, por lo que, éste hecho nos hace cambiar nuestro enfoque y deberemos comprobar si sus transacciones personalizadas reflejan todos los movimientos contables que quedan grabados de manera estándar en el sistema, incluyendo las justificaciones necesarias en caso de existir discrepancias.

Esta entidad, dispone de un sistema *ERP SAP*, por lo que, una vez han sido extraídos los asientos contables del sistema, conviene conocer el significado de los principales campos que cada asiento lleva incluido para su posterior análisis:

Campo	Descripción	Carácter
BUKRS	Sociedad	Obligatorio
BELNR	Número asiento contable	
GJAHR	Año del ejercicio fiscal	
BUZEI	Ubicación asiento	
SHKZG	Indicador Debe o Haber	
DMBTR	Importe en moneda local	
HKONT	Cuenta contable	
SGTXT	Detalle descriptivo	
AUGDT	Fecha de compensación	
AUGBL	Documento compensatorio	Adicional
KUNNR	Cliente	
LIFNR	Acreedor	

Figura 9. Tabla BSEG para detalle de los asientos contables.

Soc.	Nº doc.	Año	Pos	Compensación	Doc.comp.	D/H	Importe ML	Texto	Libro mayor	Cliente	Acreedor
0001	100000612	2017	2			S	42.013,40	12000009 - Reset per 20170531	113199		
0001	100000613	2017	1			S	42.013,40	12000009 - Reset per 20170531 Reversal	230010		
0001	100000613	2017	2			H	42.013,40	12000009 - Reset per 20170531 Reversal	113199		
0001	100000614	2017	1			S	42.013,40	12000009 - Valuation on 20170531	230010		
0001	100000614	2017	2			H	42.013,40	12000009 - Valuation on 20170531	113199		
0001	100000615	2017	1			H	42.013,40	12000009 - Valuation on 20170531 Reversal	230010		
0001	100000615	2017	2			S	42.013,40	12000009 - Valuation on 20170531 Reversal	113199		
0001	100000616	2002	1			S	1.334,92		12000008		
0001	100000616	2002	2			H	1.334,92		12000009		
0001	100000617	2002	1			S	23.746,70		12000008		
0001	100000617	2002	2			H	23.746,70		12000009		
0001	100000618	2017	1			S	16.842,54		12000008		
0001	100000618	2017	2			H	16.842,54		12000009		
0001	100000619	2004	1			S	14.463,79		12000020		
0001	100000619	2004	2			H	14.463,79		12000021		
0001	100000620	2004	1			S	26.239,50		12000020		
0001	100000620	2004	2			H	26.239,50		12000021		
0001	100000621	2004	1			S	3.131,46	12000021 - Valuation on 20040730	230010		
0001	100000621	2004	2			H	3.131,46	12000021 - Valuation on 20040730	113199		
0001	100000622	2004	1			H	3.131,46	12000021 - Valuation on 20040730 Reversal	230010		
0001	100000622	2004	2			S	3.131,46	12000021 - Valuation on 20040730 Reversal	113199		
0001	100000623	2004	1			S	3.131,46	12000021 - Valuation on 20040731	230010		
0001	100000623	2004	2			H	3.131,46	12000021 - Valuation on 20040731	113199		
0001	100000624	2004	1			H	3.131,46	12000021 - Valuation on 20040731 Reversal	230010		
0001	100000624	2004	2			S	3.131,46	12000021 - Valuation on 20040731 Reversal	113199		
0001	100000625	2017	1	26.07.2017	100000627	S	100,00		140000	21	
0001	100000625	2017	2			S	200,00		140000	1050	
0001	100000625	2017	3			S	281,00		140000	1050	
0001	100000625	2017	4			S	281,00		140000	1050	
0001	100000625	2017	5			H	941,00		140000	1050	
0001	100000625	2017	6			H	21,00		140000	1050	
0001	100000626	2017	1	26.07.2017	100000627	H	100,00		140000	21	
0001	100000626	2017	2			H	200,00		140000	1050	
0001	100000626	2017	3			H	281,00		140000	1050	
0001	100000626	2017	4			H	281,00		140000	1050	
0001	100000626	2017	5			S	941,00		140000	1050	
0001	100000626	2017	6			S	21,00		140000	1050	

Figura 10. Asientos contables en el sistema.

La entidad dispone de numerosas sociedades internacionales, de las cuáles, nos centraremos en la sociedad existente en España, al estar dentro de nuestro enfoque o alcance auditado.

Al intentar realizar la extracción de la información, nos encontramos con un primer inconveniente, ya que, el propio sistema tiene una limitación de tamaño para la exportación de archivos. Como se está analizando el período fiscal anual, se ha tomado la iniciativa de poder facilitar la información de manera mensual.

El primer paso para ejecutar el análisis, será poder adaptar para su correcta importación los ficheros BSEG que contienen la información de los asientos contables. Para ello, seleccionaremos los campos a tratar, especificando a su vez el tipo de cada uno de los datos, así como posibles delimitadores, etc.

Para optimizar este proceso de importación de la información, realizaremos un *Script* para poder importar a IDEA todas las bases de datos disponibles:

```

New Script.iss
Function ImportBBDD1
  dbName = "abril10.IMD"
  Client.ImportDelimFile "C:\Users\danielgomez\Desktop\TFM\Script\abril10.xls",
  dbName, FALSE, "", "C:\Users\danielgomez\Documents\My IDEA Documents\IDEA Projects\Samples\BSEG_CUSTOM\Custom.rdf", FALSE
Client.OpenDatabase (dbName)
End Function

Function ImportBBDD2
  dbName = "agosto10.IMD"
  Client.ImportDelimFile "C:\Users\danielgomez\Desktop\TFM\Script\agosto10.xls",
  dbName, FALSE, "", "C:\Users\danielgomez\Documents\My IDEA Documents\IDEA Projects\Samples\BSEG_CUSTOM\Custom.rdf", FALSE
Client.OpenDatabase (dbName)
End Function

Function ImportBBDD3
  dbName = "diciembre10.IMD"
  Client.ImportDelimFile "C:\Users\danielgomez\Desktop\TFM\Script\diciembre10.xls",
  dbName, FALSE, "", "C:\Users\danielgomez\Documents\My IDEA Documents\IDEA Projects\Samples\BSEG_CUSTOM\Custom.rdf", FALSE
Client.OpenDatabase (dbName)
End Function

Function ImportBBDD4
  dbName = "enero10.IMD"
  Client.ImportDelimFile "C:\Users\danielgomez\Desktop\TFM\Script\enero10.xls",
  dbName, FALSE, "", "C:\Users\danielgomez\Documents\My IDEA Documents\IDEA Projects\Samples\BSEG_CUSTOM\Custom.rdf", FALSE
Client.OpenDatabase (dbName)
End Function
  
```

Figura 10. Script Importación base de datos.

Una vez tenemos la información correctamente importada en IDEA, procuraremos adaptar las bases de datos en función de si la información ha sido tratada de forma estándar o de manera *custom* para proceder a realizar el análisis con nuestra herramienta de auditoría IDEA.

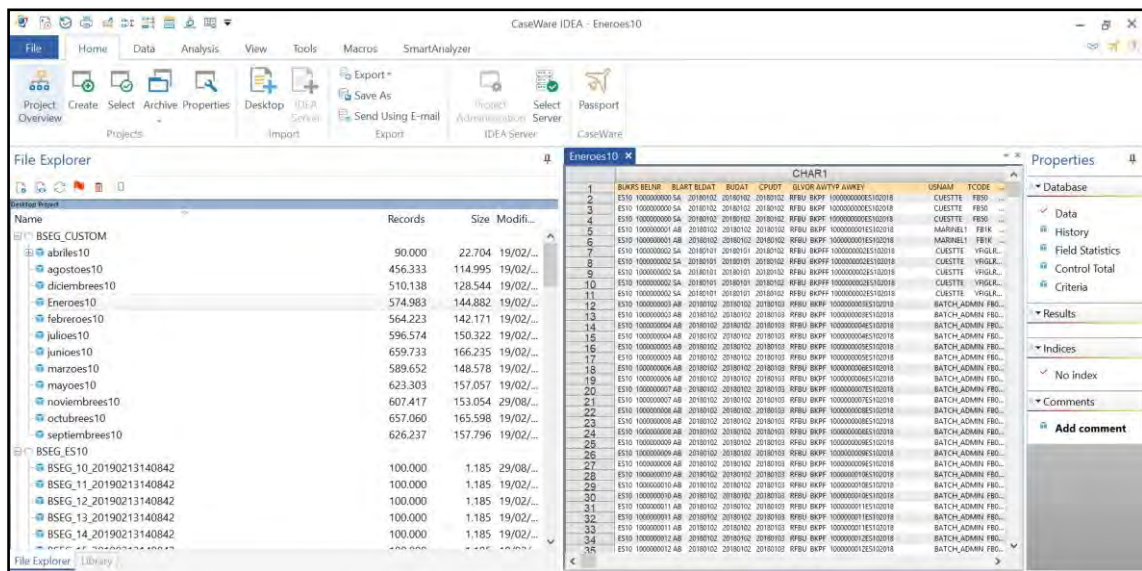


Figura 12. Contenido BSEG.

Para poder disponer de las 2 bases de datos, deberemos realizar la unión de los múltiples ficheros mensuales que las componen para formar el período fiscal anual, para ello, utilizaremos la función *Append* tanto en los asientos estándares como *custom*:

```
New Script.iss
Function AppendBBDD1|
  Set db = Client.OpenDatabase("abril10.IMD")
  Set task = db.AppendDatabase
  task.AddDatabase "agosto10.IMD"
  task.AddDatabase "diciembre10.IMD"
  task.AddDatabase "Enero10.IMD"
  task.AddDatabase "febrero10.IMD"
  task.AddDatabase "julio10.IMD"
  task.AddDatabase "junio10.IMD"
  task.AddDatabase "marzo10.IMD"
  task.AddDatabase "mayo10.IMD"
  task.AddDatabase "noviembre10.IMD"
  task.AddDatabase "octubre10.IMD"
  task.AddDatabase "septiembre10.IMD"
  dbName = "App All BSEG_Custom.IMD"
  task.PerformTask dbName, ""
  Set task = Nothing
  Set db = Nothing
  Client.OpenDatabase (dbName)
End Function
```

Figura 11. *Script Append* de bases de datos *custom*.

Nuestra primera base de datos la nombraremos BSEG\_CUSTOM para hacer referencia a la información relacionada con los asientos contables tratados de manera personalizada, y nuestra segunda base de datos la nombraremos BSEG\_ES10, siendo ésta última la que contendrá la información estándar del sistema.

Una vez disponemos de ambas bases de datos, podemos visualizar el número total de registros existentes en cada una de ellas. En BSEG\_CUSTOM existen 6.555.653 asientos mientras que en BSEG\_ES10 existen 7.065.565 de asientos contables.

Como primer análisis comprobamos que existe una diferencia de aproximadamente unos 500.000 asientos contables extraídos del propio sistema. Luego, debemos continuar con las pruebas para poder emitir las conclusiones oportunas.

El siguiente paso que realizaremos será eliminar las cabeceras para obtener un listado más "limpio" de asientos. El campo *key* existente en ambas bases de datos y que utilizaremos para seguir realizando pruebas será BELNR, es decir, el identificador único de cada asiento contable.

A continuación, realizaremos un *Join* del tipo *matches only* entre ambas bases de datos para comprobar si se encuentran o no todos los asientos contables. Y comprobamos que existen diferencias al resultar 5.991.419 de asientos coincidentes.

```

New Script.iss
Function JoinDBMATCH
    Set db = Client.OpenDatabase("BSEG_CUSTOM\App All BSEG_Custom.IMD")
    Set task = db.JoinDatabase
    task.FileToJoin "BSEG_ES10\Append Databases.IMD"
    task.IncludeAllPFields
    task.IncludeAllSFields
    task.AddMatchKey "BELNR", "BELNR", "A"
    task.CreateVirtualDatabase = False
    dbName = "BSEG_CUSTOM\Join Databases.IMD"
    task.PerformTask dbName, "", WI_JOIN_MATCH_ONLY
    Set task = Nothing
    Set db = Nothing
    Client.OpenDatabase (dbName)
End Function
    
```

Figura 14. Script Match entre bases de datos.

Name	Records	Size
BSEG_CUSTOM		
abril10	90.000	22.704
App All BSEG_Custom	6.555.653	3.514.765
BSEG_CUSTOM_Detalle	6.555.641	2.106.290
Join Databases	5.991.419	1.819.700
agosto10	456.333	114.995
diciembre10	510.138	128.544
Enero10	574.983	144.882
febrero10	564.223	142.171
julio10	596.574	150.322
junio10	659.733	166.235
marzo10	589.652	148.578
mayo10	623.303	157.057
noviembre10	607.417	153.054
octubre10	657.060	165.598
septiembre10	626.237	157.796

Figura 15. Tratamiento Custom en IDEA.

Con esta información, continuaremos con el tratamiento de los datos estándares realizando un *Join* del tipo *records with no secondary match* entre las 2 bases de datos y obtenemos un total de 1.074.143 de asientos que se encuentran en la base de datos BSEG\_ES10 pero no aparecen en la base de datos BSEG\_CUSTOM.

```
New Script.iss
Function JoinDBNSMATCH
    Set db = Client.OpenDatabase("BSEG_ES10\Append Databases.IMD")
    Set task = db.JoinDatabase
    task.FileToJoin "BSEG_CUSTOM\App All BSEG_Custom.IMD"
    task.IncludeAllFields
    task.AddMatchKey "BELNR", "BELNR", "A"
    task.CreateVirtualDatabase = False
    dbName = "BSEG_ES10\Join Databases.IMD"
    task.PerformTask dbName, "", WI_JOIN_NOC_SEC_MATCH
    Set task = Nothing
    Set db = Nothing
    Client.OpenDatabase (dbName)
End Function
```

Figura 16. *Script No Secondary* entre bases de datos.

Como hay muchos asientos que pueden encontrarse más de una vez, procederemos a realizar la función *summarization* que nos permite agrupar dichos asientos y obtenemos que existen 1.092.131 de asientos que aparecen más de una vez en la base de datos BSEG\_ES10 y un total de 945.507 asientos en la base de datos BSEG\_CUSTOM.

```
New Script.iss
Function SUMBSEG
    Set db = Client.OpenDatabase("BSEG_ES10\Append Databases.IMD")
    Set task = db.Summarization
    task.AddFieldToSummarize "BELNR"
    task.IncludeAllFields
    dbName = "BSEG_ES10\Sum BSEG_Estándar por BELNR.IMD"
    task.OutputDBName = dbName
    task.CreatePercentField = FALSE
    task.PerformTask
    Set task = Nothing
    Set db = Nothing
    Client.OpenDatabase (dbName)
End Function

Function SUMCUSTOM
    Set db = Client.OpenDatabase("BSEG_CUSTOM\App All BSEG_Custom.IMD")
    Set task = db.Summarization
    task.AddFieldToSummarize "BELNR"
    task.Criteria = "CABECERA == ""Detalle""
    dbName = "BSEG_ES10\Sum BSEG_Custom por BELNR.IMD"
    task.OutputDBName = dbName
    task.CreatePercentField = FALSE
    task.PerformTask
    Set task = Nothing
    Set db = Nothing
    Client.OpenDatabase (dbName)
End Function
```

Figura 127. *Script Summarization* entre bases de datos.



El siguiente paso que llevaremos a cabo será comprobar cuántos de estos asientos coinciden entre ambas bases de datos. Por lo que, realizaremos un *Join* del tipo *matches only* entre ambas bases de datos y obtenemos 942.242 asientos coincidentes, por tanto, estos asientos aparecen más de una vez en cada una de ellas.

Para asegurarnos, realizaremos un *Join* del tipo *records with no primary match*, y efectivamente, aparecen 3.265 asientos que sumados a los anteriores nos resultan los 945.507 asientos que aparecen en la BSEG\_CUSTOM. Aprovecharemos este punto para realizar un *Join* del tipo *records with no secondary match* que utilizaremos más adelante, obteniendo 149.889 asientos repetidos que se encuentran en la base de datos BSEG\_ES10 pero no en la BSEG\_CUSTOM.

```

New Script.iss
Function JoinDBSUMNPMATCH
  Set db = Client.OpenDatabase("BSEG_ES10\Sum BSEG_Estándar por BELNR.IMD")
  Set task = db.JoinDatabase
  task.FileToJoin "BSEG_ES10\Sum BSEG_Custom por BELNR.IMD"
  task.IncludeAllPFields
  task.IncludeAllSFields
  task.AddMatchKey "BELNR", "BELNR", "A"
  task.CreateVirtualDatabase = False
  dbName = "BSEG_ES10\Join NP BSEG_Estándar-Custom.IMD"
  task.PerformTask dbName, "", WI_JOIN_NOC_PRI_MATCH
  Set task = Nothing
  Set db = Nothing
  Client.OpenDatabase (dbName)
End Function

```

Figura 18. Script No Primary entre bases de datos.

Una vez hemos llegado a este punto, realizaremos un *visual conector* siendo el campo único de selección BELNR, entre 3.265 de asientos que se repiten y existen en la base de datos BSEG\_CUSTOM y el total de registros de dicha base de datos formada por 6.555.641 de asientos sin las cabeceras, para obtener los posibles valores para esos asientos.

```

New Script.iss
Function VISUALCONNECTDB
  Set db = Client.OpenDatabase("BSEG_ES10\Join NP BSEG_Estándar-Custom.IMD")
  Set task = db.VisualConnector
  id0 = task.AddDatabase ("BSEG_ES10\Join NP BSEG_Estándar-Custom.IMD")
  id1 = task.AddDatabase ("BSEG_CUSTOM\BSEG_CUSTOM_Detalle.IMD")
  task.MasterDatabase = id0
  task.AppendDatabaseNames = FALSE
  task.IncludeAllPrimaryRecords = FALSE
  task.AddRelation id0, "BELNR1", id1, "BELNR"
  task.IncludeAllFields
  task.CreateVirtualDatabase = False
  dbName = "BSEG_ES10\VisCon para añadir información original.IMD"
  task.OutputDatabaseName = dbName
  task.PerformTask
  Set task = Nothing
  Set db = Nothing
  Client.OpenDatabase (dbName)
End Function

```

Figura 19. Script Visual Conector entre bases de datos.

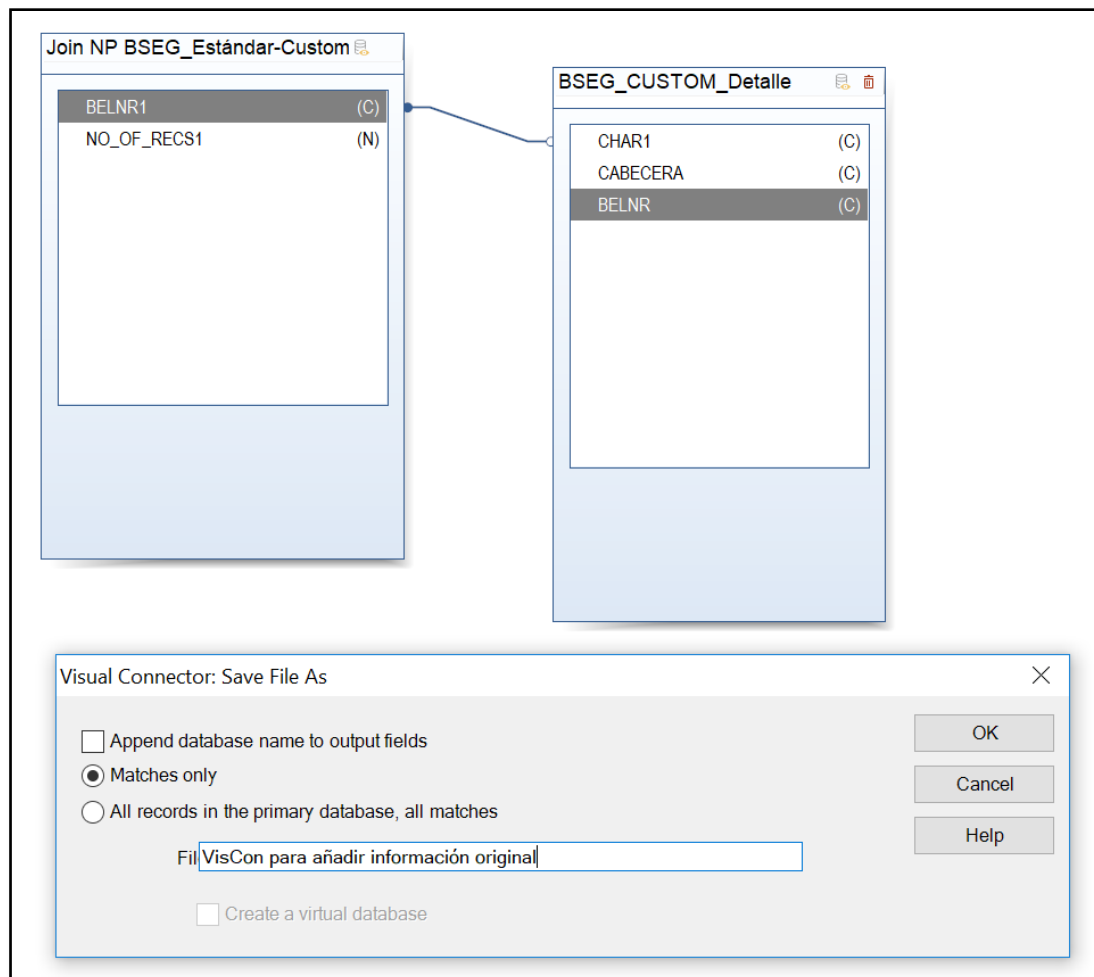


Figura 20. Visual Conector en IDEA.

Y obtenemos 564.222 registros. Por consiguiente, el próximo paso es realizar una *Summarization* para agrupar los registros que se encuentren repetidos, dando como resultado un total de 88.995 registros como posibles valores.

Rescatando los 149.889 asientos repetidos que se encuentran en la base de datos BSEG\_ES10 pero no en la BSEG\_CUSTOM, realizaremos los 3 tipos de *Join* que hemos realizado anteriormente con los posibles 88.995 registros, *matches only* encontrando todos como resultado, *records with no secondary match* no encontrando ningún asiento, y por último, *records with no primary match* donde podemos observar que de los 88.995 posibles asientos, existen 60.894 posibles asientos en la base de datos BSEG\_ES10.

Como último paso, realizaremos la implementación de otro *visual conector*, utilizando como valor de campo único el campo BELNR entre la base de datos primaria Join NP BSEG\_Estándar-Custom y la segunda base de datos BSEG\_CUSTOM\_Detalle para poder obtener todos los posibles valores que coincidan entre ambas bases de datos.

El *visual conector* nos ha proporcionado el número de los asientos contables a los que correspondería esos posibles valores, dando como resultado 509.921 asientos contables, que son los asientos que existen como diferencia entre la base de datos BSEG\_ES10 y la base de datos personalizada BSEG\_CUSTOM.

Name	Records	Size
BSEG_1_20190213140842	100.000	1.185
Append Databases	7.065.565	186.351
Join Databases	1.074.143	12.646
Sum BSEG_Estándar por BELNR	1.092.131	38.448
Join Match BSEG_Estándar-Custom	942.242	52.550
Join NP BSEG_Estándar-Custom	3.265	267
VisCon para añadir información original	564.222	186.392
Sum S	88.995	5.808
Join Match entre Febrero (Custom) - Sólo en ESTAN...	88.995	5.034
Join NP entre Febrero (Custom) - Sólo en ESTANDAR	60.894	4.389
VisCon Detalle encontrados	509.921	1.860.976
ES10	509.921	1.851.712
Summarization	60.894	3.309
Join NS entre Febrero (Custom) - Sólo en ESTANDAR	0	75
Join NS BSEG_Estándar-Custom	149.889	5.363
BSEG_20_20190213140842	100.000	1.185
BSEG_21_20190213140842	100.000	1.185
BSEG_22_20190213140842	100.000	1.185
BSEG_23_20190213140842	100.000	1.185

Figura 21. Tratamiento Estándar en IDEA.

Pero para poder concluir con mayor detalle nuestro análisis, además de haber logrado encontrar los asientos contables, realizaremos la creación de una columna importe, para poder conocer el grado de detalle entre ambas bases de datos, obteniendo como resultado una diferencia de 264.710.623€.

	CHAR19	IMPORTE OK	CHAR87	CHAR20	CHAR21	CHAR22	CHAR23	CHAR24	CHA
1	702454.90	-702.454,90	0004502009	702454.90	0.00	702454.90	EUR	0.00	0.00
2	12048.01	-12.048,01	0001152091	12048.01	0.00	12048.01	EUR	0.00	0.00
3	1130.91	-1.130,91	0001154091	1130.91	0.00	1130.91	EUR	0.00	0.00
4	21589.15	-21.589,15	0006537129	21589.15	0.00	21589.15	EUR	0.00	0.00
5	8787.56	-8.787,56	0006321329	8787.56	0.00	8787.56	EUR	0.00	0.00
6	0.14	-0,14	0003431401	0.14	0.00	0.14	EUR	0.00	0.00
7	147.62	-147,62	0001060003	147.62	0.00	147.62	EUR	0.00	0.00
8	0.72	-0,72	0002101039	0.72	0.00	0.72	EUR	0.00	0.00
9	0.89	-0,89	0001037100	0.89	0.00	0.89	EUR	0.00	0.00
10	5.11	-5,11	0003433201	5.11	0.00	5.11	EUR	0.00	0.00
11	20.24	-20,24	0002321305	20.24	0.00	20.24	EUR	0.00	0.00
12	25.90	-25,90	0002101022	25.90	0.00	25.90	EUR	0.00	0.00
13	56.54	-56,54	0001026001	56.54	0.00	56.54	EUR	0.00	0.00
14	57.52	-57,52	0001033109	57.52	0.00	57.52	EUR	0.00	0.00
15	456.06	-456,06	0003011001	456.06	0.00	456.06	EUR	0.00	0.00
16	223.40	-223,40	0002272002	223.40	0.00	223.40	EUR	0.00	0.00
17	95.77	-95,77	0002241001	95.77	0.00	95.77	EUR	0.00	0.00
18	0.30	-0,30	0002101039	0.30	0.00	0.30	EUR	0.00	0.00
19	0.28	-0,28	0001037100	0.28	0.00	0.28	EUR	0.00	0.00
20	2.81	-2,81	0003433201	2.81	0.00	2.81	EUR	0.00	0.00
21	24.33	-24,33	0001033109	24.33	0.00	24.33	EUR	0.00	0.00
22	230.75	-230,75	0003011001	230.75	0.00	230.75	EUR	0.00	0.00
23	32.46	-32,46	0002241001	32.46	0.00	32.46	EUR	0.00	0.00
24	0.45	-0,45	0003431401	0.45	0.00	0.45	EUR	0.00	0.00
25	441.95	-441,95	0001060003	441.95	0.00	441.95	EUR	0.00	0.00
26	0.37	-0,37	0001037100	0.37	0.00	0.37	EUR	0.00	0.00
27	34.00	-34,00	0001033109	34.00	0.00	34.00	EUR	0.00	0.00
28	0.02	-0,02	0002101039	0.02	0.00	0.02	EUR	0.00	0.00
29	0.05	-0,05	0001037100	0.05	0.00	0.05	EUR	0.00	0.00
30	0.84	-0,84	0003433201	0.84	0.00	0.84	EUR	0.00	0.00
31	4.32	-4,32	0001033109	4.32	0.00	4.32	EUR	0.00	0.00
32	72.60	-72,60	0003011001	72.60	0.00	72.60	EUR	0.00	0.00
33	15.25	-15,25	0002241001	15.25	0.00	15.25	EUR	0.00	0.00
34	0.02	-0,02	0001037100	0.02	0.00	0.02	EUR	0.00	0.00
35	0.41	-0,41	0003433201	0.41	0.00	0.41	EUR	0.00	0.00

Figura 22. Conclusiones obtenidas según importe.

Como hemos detectado los asientos contables que se encuentran incluidos en la base de datos estándar del sistema, procederemos a visualizar los que reflejen un mayor importe, ya que, serán los que incidan en un mayor impacto y relevancia en los estados financieros de la entidad.

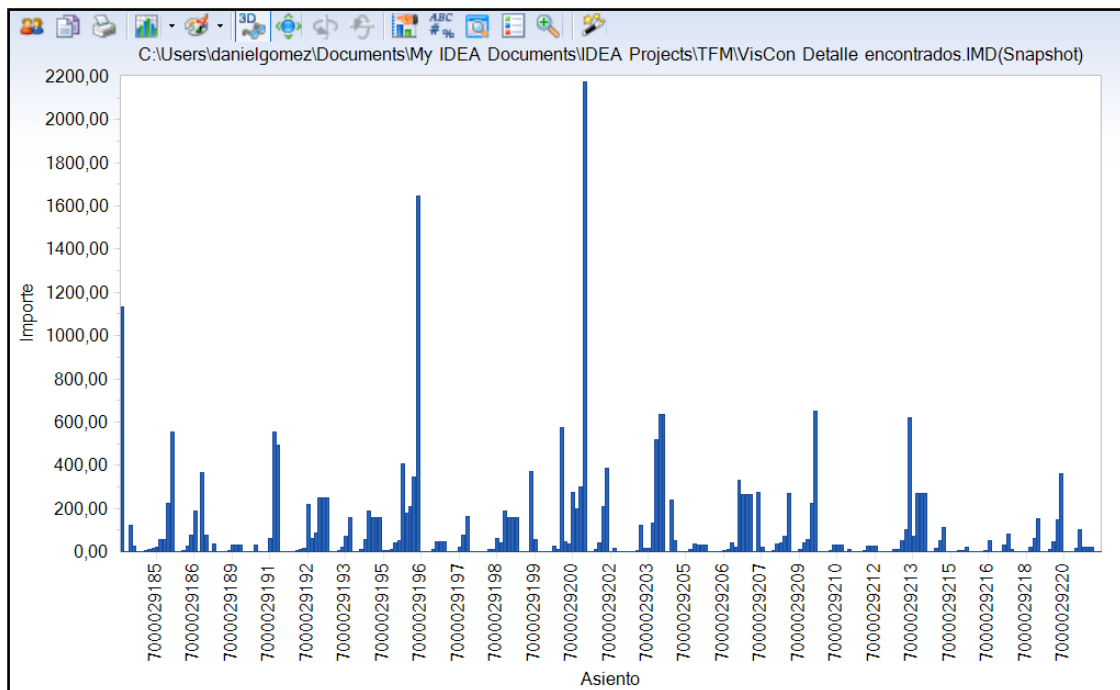


Figura 23. Asientos contables y su impacto.

Ante la existencia de tal diferencia económica entre ambas bases de datos, se solicitó la justificación pertinente a la entidad auditada, quién nos facilitó dicha justificación en persona al poder comprobar una serie de filtros existentes en las transacciones estándares para poder excluir de la contabilidad apuntes como pueden ser las nóminas de los empleados.

Para finalizar, observaremos la visualización de todas las pruebas y pasos realizados entre las distintas bases de datos que nos han permitido llegar a la conclusión estimada.

Ante la existencia de multitud de ramificaciones existentes, resulta difícil poder visualizar todo el conjunto en una única visualización.

A continuación detallaremos los pasos más importantes llevados a cabo:

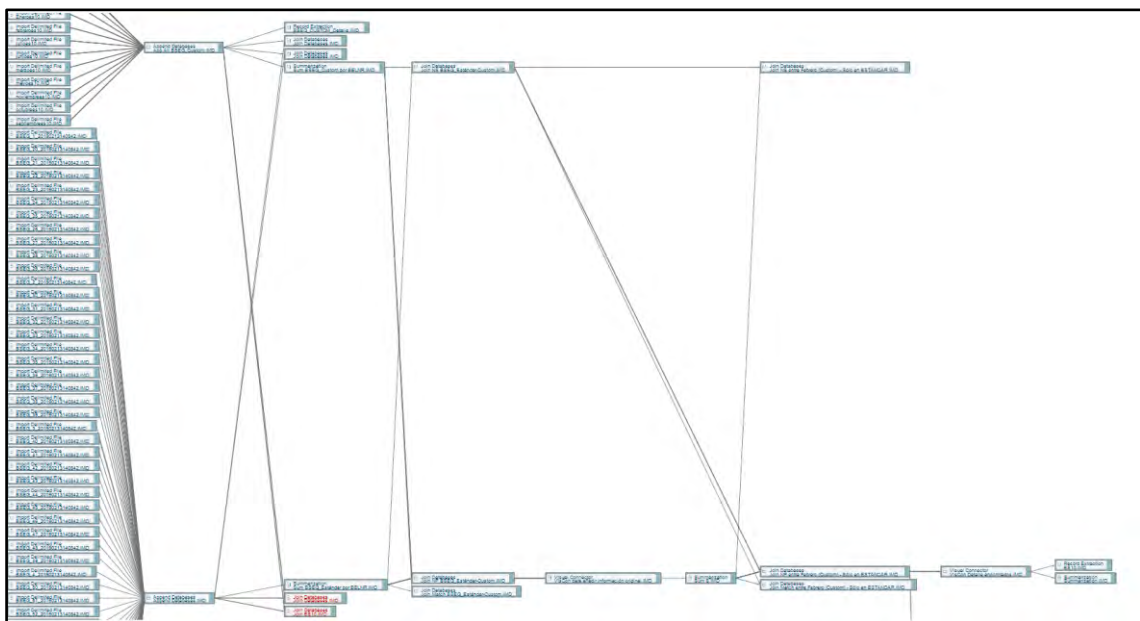


Figura 24. Pruebas realizadas sobre las bases de datos con IDEA.

Además de la realización de cada uno de los pasos realizados en esta prueba, las funciones utilizadas se han implementado en *Scripts*, de manera que, de este modo se pueda re-utilizar el código para futuras pruebas alterando o modificando las bases de datos que vayan a ser analizadas optimizando así el rendimiento y la automatización de este procedimiento.



## 4.2 Caso de estudio 2

Para la realización de nuestro segundo caso de estudio, nos hemos marcado como objetivo la obtención de valor en la información proporcionada por una entidad relacionada con el ámbito de la hostelería, de manera que, de la información almacenada y sin uso aparente por la empresa, se puedan extraer conclusiones que optimicen o re-direccionen futuras estrategias o líneas de negocio.

Una vez, la entidad nos ha facilitado la extracción de la información de los sistemas de los estados financieros de la misma, el primer paso que debemos realizar es una adecuada importación de los datos, para que no puedan verse afectados o podamos perder información de valor para el posterior análisis.

A través de la herramienta de análisis de datos *Power BI*, podemos seleccionar distintas bases de datos locales, servicios en la nube, archivos .xls, .csv, .txt, o directamente incluso a través de datos tabulares existentes en una url.

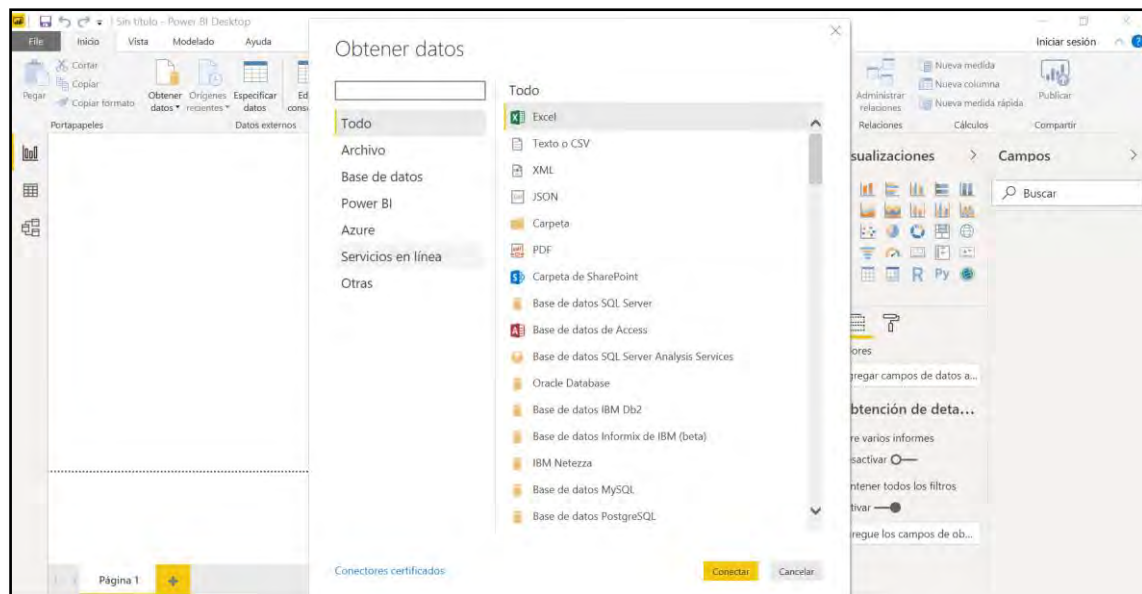


Figura 25. Importación de la información en *Power BI*.

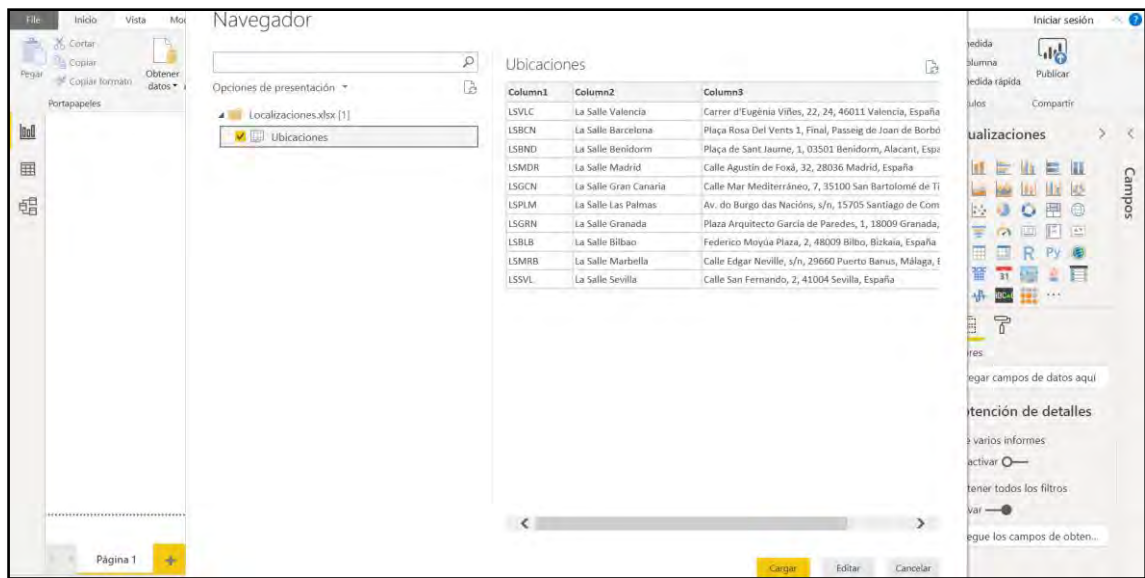


Figura 26. Pre-visualización y adaptación de la información en Power BI.

Cuando disponemos de la información importada en la herramienta, podremos acceder a visualizar todo el contenido de los registros de cada base de datos, dónde a través del editor de *queries*, podremos aplicar directamente la creación de campos y columnas calculadas, medidas, o consultas de lenguaje *DAX* con la que dar forma a la información y podamos visualizarla correctamente para los posteriores informes.

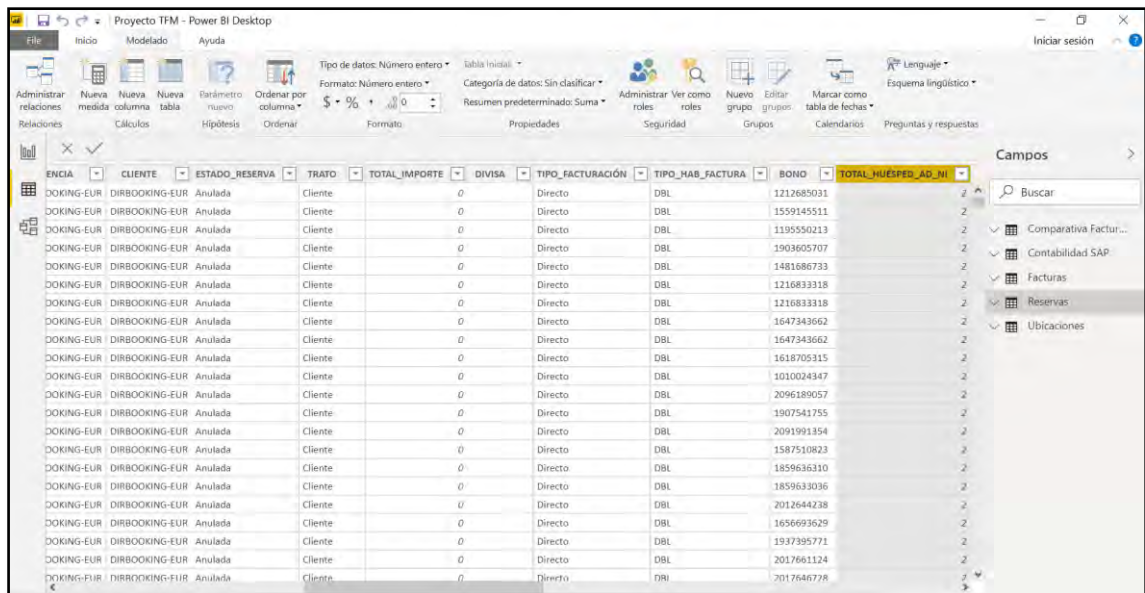


Figura 27. Tratamiento de las bases de datos en Power BI.



Por motivo de la RGPD que tiene por objetivo garantizar la confidencialidad respecto al tratamiento de la información personal o comercial, se han modificado los registros de la información facilitada, para poder conservar y preservar la confidencialidad e integridad de la entidad en la que se ha llevado a cabo el estudio.

Hemos importado la información facilitada por el cliente, que se encuentra formada por varias bases de datos:

- Ubicaciones
- Reservas
- Facturas
- Contabilidad *SAP*

La empresa, nos ha explicado a modo de entendimiento, que su proceso de ventas se inicia con el registro o la reserva de las estancias o pernoctaciones llevadas a cabo por el usuario en las distintas sedes de la cadena, a través tanto de la aplicación propia como por medio web. Cada sede registra las reservas y posteriormente realizan el volcado de las facturas agrupadas por hotel al sistema contable de la entidad, siendo el sistema contable analizado *SAP*.

Para poder obtener conclusiones por medio de las visualizaciones en la elaboración de nuestro informe a través del *dashboard*, hemos comenzado situando las diferentes sedes de la cadena hotelera en forma de mapa para conocer exactamente sus ubicaciones.

El primer paso realizado ha sido acceder a la base de datos que contiene las Ubicaciones de cada uno de los hoteles, y mapear el campo *DIRECCION HOTEL* para la visualización del tipo *Arc GIS Map*. En función de la información proporcionada en este campo, se podrá visualizar correctamente o deberá poder ser adaptada mediante coordenadas de latitud y longitud.



Figura 28. Ubicación de los hoteles.

Aprovechando que tenemos esta visualización activada, podemos observar de un modo rápido, la facturación que ha llevado a cabo a lo largo del período fiscal cada uno de los hoteles.

Para ello, hemos mapeado de la base de datos Ubicaciones el campo DIRECCIÓN HOTEL como campo ubicación, el campo tamaño TOTAL\_IMPORTE perteneciente a la base de datos Facturas, y el color HOTEL de la base de datos Ubicaciones para poder diferenciarlos visualmente según la facturación obtenida por cada sede.

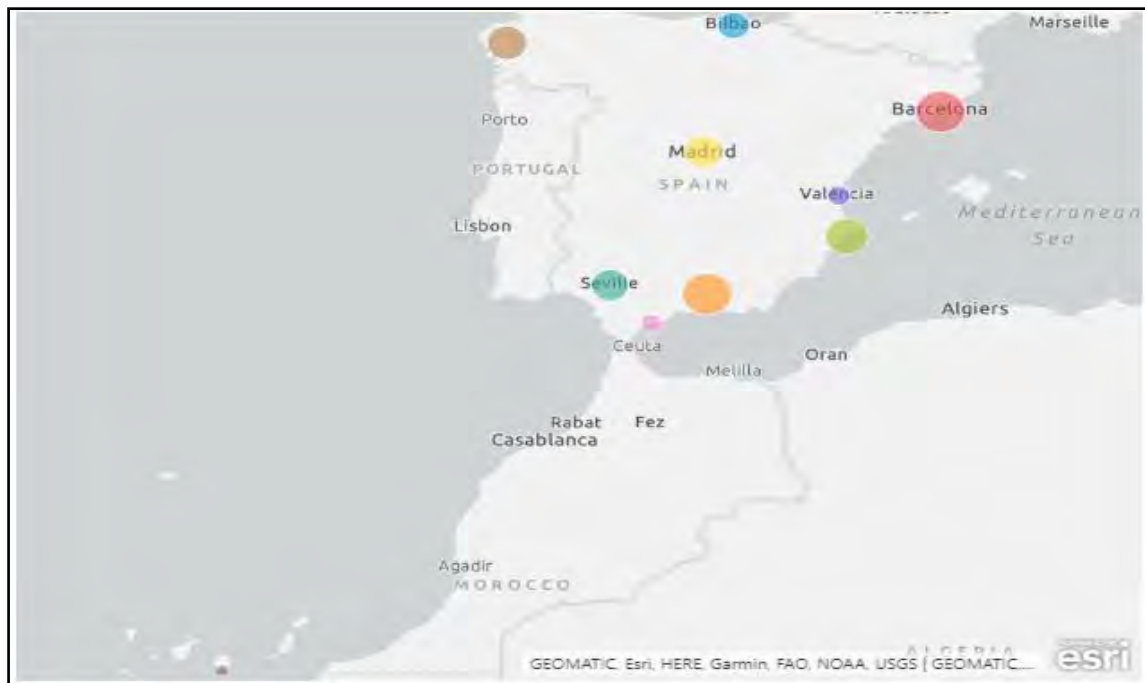


Figura 29. Facturación por hotel.

De este modo, podemos localizar y diferenciar los hoteles que han logrado facturar más y aquellos que han facturado menos, siendo éstos ordenados de manera descendente según las cifras económicas obtenidas seleccionando la visualización tabla para el campo HOTEL de la base de datos Ubicaciones e Importe Reservas de la base de datos Comparativa Facturación.

LSBCN	12.566.305,000 €
LSGRN	10.734.153,000 €
LSBND	9.141.434,000 €
LSPLM	8.675.492,000 €
LSMDR	7.558.758,000 €
LSSVL	7.417.914,000 €
LSBLB	4.969.396,000 €
LSVLC	1.614.525,000 €
LSMRB	1.380.765,000 €
LSGCN	388.774,000 €
<b>Total</b>	<b>64.447.516,000 €</b>

Figura 30. Importes facturados por hotel.

Como hemos logrado obtener la facturación total conjunta de la cadena hotelera en el período fiscal auditado, vamos a realizar la comparativa con la información existente en el sistema contable de la entidad para poder comprobar si existen o no diferencias, y cuáles son los importes existentes en los estados financieros.

Hemos analizado el importe total de la facturación realizada por cada uno de los hoteles en el período anual, y se ha obtenido un importe total de 64.447.516€, mientras que el sistema contable SAP, refleja una facturación total de 64.969.011€, por lo que, existe una diferencia entre las facturas y la contabilidad en los estados financieros de la compañía de 124.048€.

El principal motivo o la causa de esta desviación, resulta de que para poder realizar el devengo se ha partido de un precio medio por noche, y esta situación provoca pequeñas desviaciones respecto al importe en SAP, ya que, el precio por noche es un precio fluctuante a lo largo de todo el período fiscal.

Para poder visualizarlo, hemos seleccionado un gráfico de columnas agrupadas y de líneas, mapeando los campos contenidos en la base de datos Comparativa Facturación, según columnas y ejes, salvo en Importe SAP, proveniente de la base de datos Contabilidad SAP, campo DMBTR\_D\_H\_SUM (Importe en moneda local).

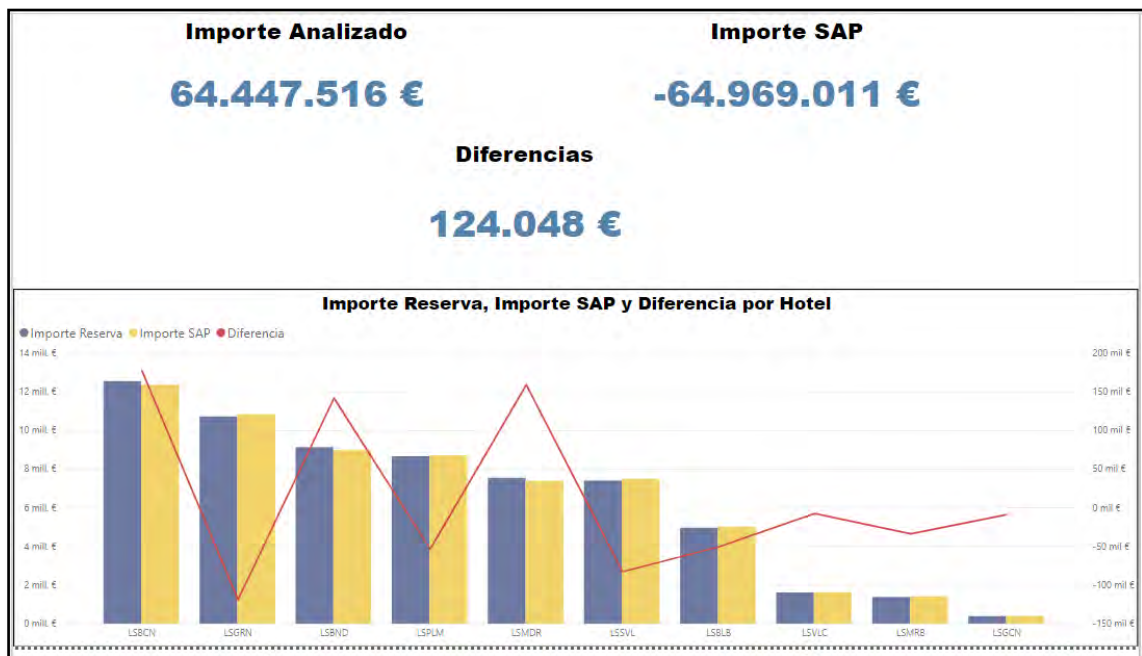


Figura 31. Comparativa del sistema contable de la cadena hotelera.

Continuando con nuestro análisis, hemos realizado una comparativa para cada hotel sobre las variaciones existentes en cada uno de ellos, según el precio por noche, mes, y el tipo de habitación solicitada por el usuario.

Para representar esta información, se han utilizado varias segmentaciones de datos, para poder filtrar cada hotel y tipo de habitación variando la información en tiempo real, una tabla que pueda representar de la base de datos Reservas, los campos que aparecen en la imagen inferior, y finalmente un gráfico de columnas apiladas y líneas, dónde de la base de datos Reservas, hemos mapeado el campo FECHA\_ENTRADA filtrando por Mes, el promedio de PRECIO\_NOCHE y el recuento del número de reservas que pertenecen al campo HARD, filtrando aquellos valores que sean mayores a 0 para evitar posibles errores de representación.



Figura 32. Reservas mensuales según precio medio en Granada.

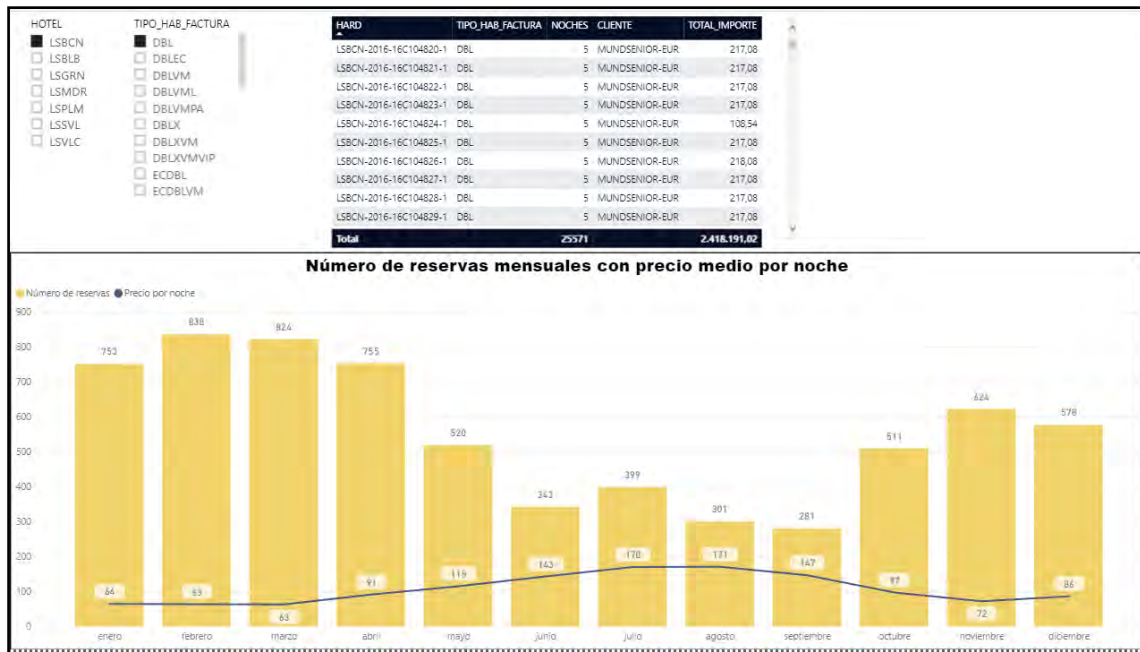


Figura 33. Reservas mensuales según precio medio en Barcelona.

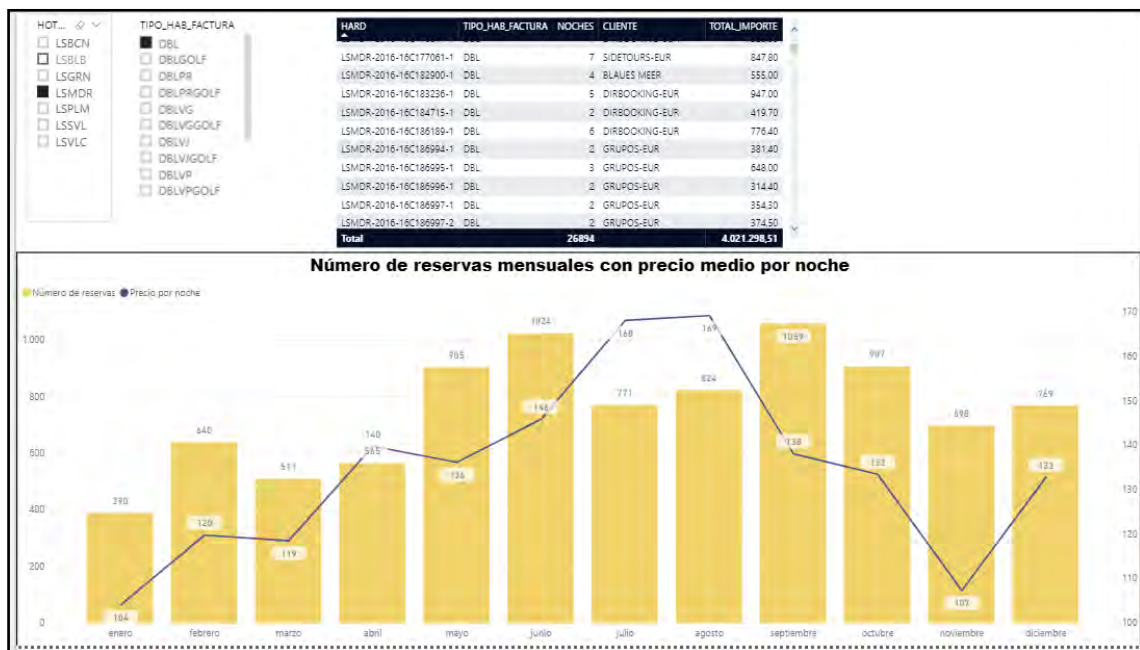


Figura 34. Reservas mensuales según precio medio en Madrid.

Como se puede observar, aunque existen variables y factores externos o internos como puede ser la ubicación del hotel, o en función del mes, según la temporada climática, generalmente se puede concluir, que a mayor subida del precio medio por noche en el hotel, hay un descenso en las reservas de los mismos, fundamentada y directamente relacionada por dicha subida del precio medio por noche, excluyendo de casos o determinadas particularidades excepcionales en algún hotel, como puede ser Valencia, ya que, aun existiendo subidas en el precio medio por noche, prácticamente no descienden las reservas y en algunos meses incluso aumentan.

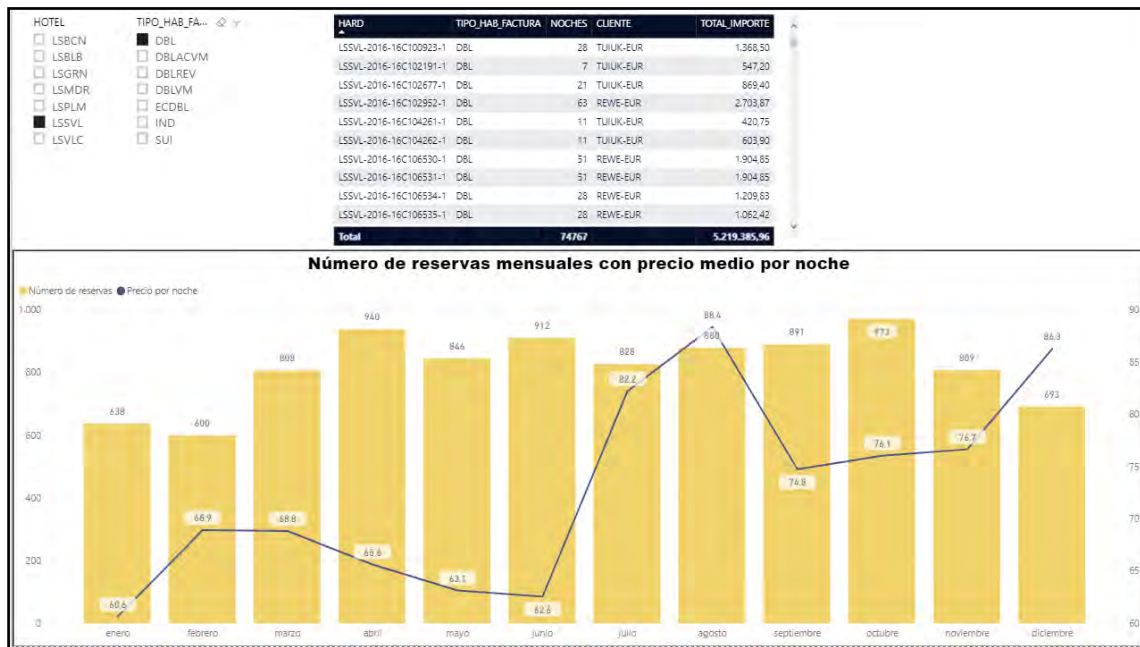


Figura 35. Reservas mensuales según precio medio en Valencia.

Otro punto de vista que hemos querido comprobar es la facturación realizada sin reserva previa por parte del usuario. Este tipo de facturas son directamente volcadas al propio sistema contable SAP.

Para ello, hemos seleccionado varias segmentaciones de datos, para poder representar las distintas variantes posibles en función del hotel y las facturas asociadas a cada uno de ellos, un gráfico circular donde mapearemos los campos HOTEL y TOTAL\_IMPORTE de la base de datos Facturas y realizaremos un filtrado de página para aquellas facturas con ESTADO igual a SIN RESERVA.

Y por último, un *CalendarVisual* seleccionando los campos FECHA\_FACTURACIÓN y PRECIO\_FACTURA\_SIN\_IVA de la base de datos Facturas, para poder representar en el calendario en función de la intensidad del color, qué facturación ha existido cada uno de los días del año.

Las conclusiones que podemos obtener de este análisis, a lo largo de todo el año, la concentración de la facturación sin reserva previa, generalmente, es realizada desde el final de la Primavera hasta el final del Otoño, concentrándose especialmente en la época veraniega.

Si seleccionamos un día de verano como puede ser en Julio, observamos la proporción de esas facturaciones sin reserva, obteniendo los mayores porcentajes la mayoría de hoteles costeros que representan un total del 79,93% (Barcelona 56,06%, Valencia 9,06%, Benidorm 7,59%, Bilbao 7,22%), frente a los hoteles de interior que representan un total del 20,07% (Granada 11,22%, Madrid 4,65%, Sevilla 3,12% y Santiago de Compostela 1,08%).

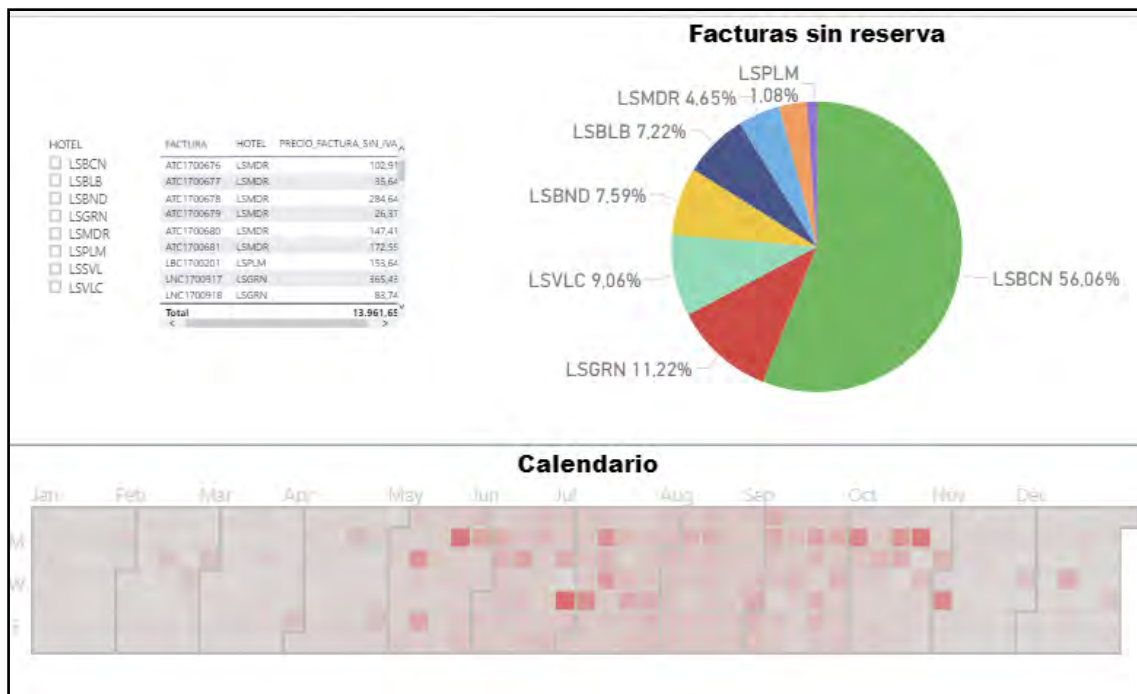


Figura 36. Época de la facturación en verano sin reserva previa.

Si por el contrario, analizamos la proporción de la facturación sin reserva previa llevada a cabo un día de Enero, en la época invernal, observamos que los hoteles de interior representan el 76,4% (Granada 37,87%, Madrid 28,08%, Sevilla 5,57%, Santiago de Compostela 4,89%) frente a los hoteles de costa que representan un total de 23,6% (Bilbao 14,49%, Valencia 9,11%).

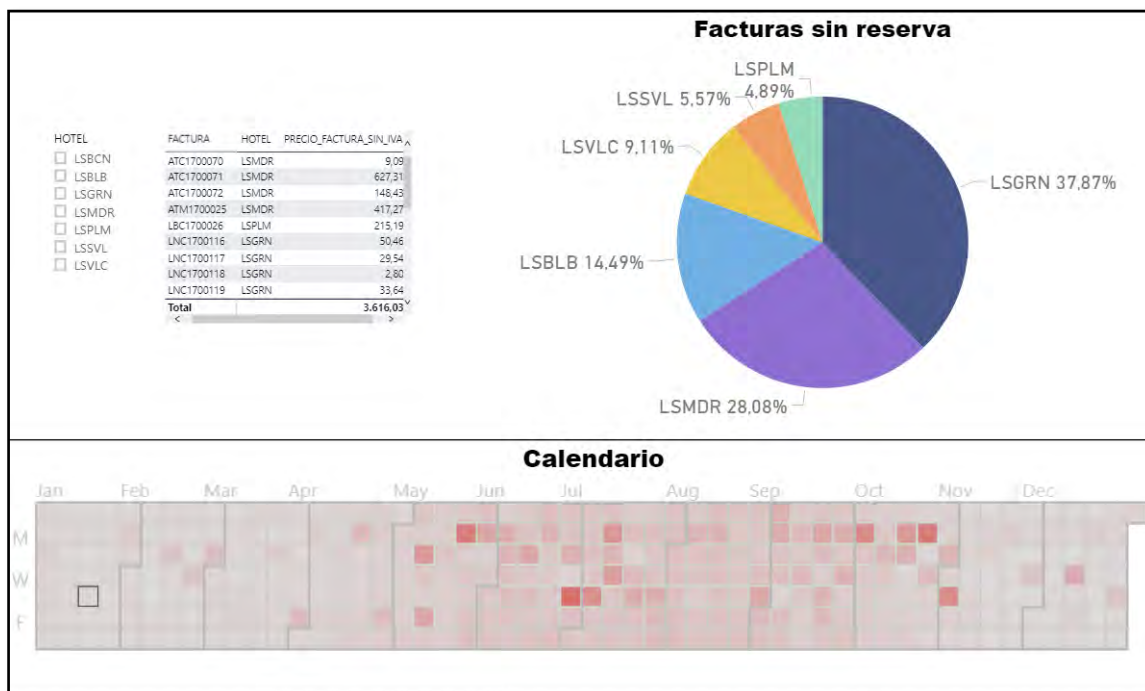


Figura 37. Época de la facturación en invierno sin reserva previa.

El siguiente apartado que nos gustaría poder conocer, es analizar con cuánta antelación el usuario suele reservar su estancia en cada uno de los hoteles.

Para llevar esto a cabo, mediante el lenguaje de programación *DAX* realizaremos en primer lugar la creación de una nueva columna de cálculo, y después obtendremos la diferencia existente en número de días, entre la fecha de venta de la reserva realizada por el usuario, y la fecha de entrada a la estancia en el hotel.

La sentencia que hemos utilizado después de modelar la creación de una nueva columna en la base de datos Reservas, ha sido la siguiente:

```
1 DIAS_ATELACION = DATEDIFF(Reservas[FECHA_VENTA].[Date];Reservas[FECHA_ENTRADA].[Date];DAY)
```

The screenshot shows the Power BI Desktop interface. The formula bar at the top contains the DAX formula: `1 DIAS_ATELACION = DATEDIFF(Reservas[FECHA_VENTA].[Date];Reservas[FECHA_ENTRADA].[Date];DAY)`. Below the formula bar, a data table is displayed with the following columns: HE, PRECIO\_RESERVA\_ESTANCIA, PRECIO\_RESERVA\_ESTANCIA\_SIN\_IVA, IMPORTE\_TOTAL\_SIN\_IVA, COLZO, DIF\_DEVENGADO, HARD, and DIAS\_ATELACION. The DIAS\_ATELACION column contains numerical values representing the number of days between the reservation date and the check-in date for each record. The table also includes a 'Campos' (Fields) pane on the right showing a list of fields from the 'Reservas' table.

HE	PRECIO_RESERVA_ESTANCIA	PRECIO_RESERVA_ESTANCIA_SIN_IVA	IMPORTE_TOTAL_SIN_IVA	COLZO	DIF_DEVENGADO	HARD	DIAS_ATELACION
0	0	0	0	0	0	LSVLC-2017-2189-1	0
0	0	0	0	0	0	LSVLC-2017-17C100262-1	26
0	0	0	0	0	0	LSVLC-2017-17C101354-1	4
0	0	0	0	0	0	LSVLC-2017-17C101950-1	40
0	0	0	0	0	0	LSVLC-2017-17C101955-1	6
0	0	0	0	0	0	LSVLC-2017-17C102503-1	25
0	0	0	0	0	0	LSVLC-2017-17C102504-1	25
0	0	0	0	0	0	LSVLC-2017-17C102570-1	21
0	0	0	0	0	0	LSVLC-2017-17C102571-1	21
0	0	0	0	0	0	LSVLC-2017-17C103420-1	77
0	0	0	0	0	0	LSVLC-2017-17C103446-1	4
0	0	0	0	0	0	LSVLC-2017-17C104127-1	2
0	0	0	0	0	0	LSVLC-2017-17C104787-1	11
0	0	0	0	0	0	LSVLC-2017-17C104992-1	1
0	0	0	0	0	0	LSVLC-2017-17C105471-1	75
0	0	0	0	0	0	LSVLC-2017-17C105682-1	75
0	0	0	0	0	0	LSVLC-2017-17C105835-1	5
0	0	0	0	0	0	LSVLC-2017-17C107069-1	29
0	0	0	0	0	0	LSVLC-2017-17C107772-1	33
0	0	0	0	0	0	LSVLC-2017-17C108200-1	1
0	0	0	0	0	0	LSVLC-2017-17C108805-1	10
0	0	0	0	0	0	LSVLC-2017-17C108819-1	74

Figura 38. Cálculos en lenguaje *DAX*.

Para poder agrupar los datos que nos permitan ver, analizar y explorar más claramente tendencias en los objetos visuales, es necesario realizar una discretización de los mismos. De este modo, crearemos un nuevo grupo en la nueva columna creada *DIAS\_ATELACION*:



The image shows a software dialog box titled "Grupos" with a close button (X) in the top right corner. It contains the following fields and controls:

- Nombre:** A text input field containing "DIAS\_ATELACION (ubicaciones)".
- Campo:** A text input field containing "DIAS\_ATELACION".
- Tipo de grupo:** A dropdown menu with "Discretizar" selected.
- Valor mínimo:** A text input field containing "-1".
- Tipo de discretización:** A dropdown menu with "tamaño de las discretizaciones" selected.
- Valor máximo:** A text input field containing "1812".

Below these fields is a paragraph of text: "La discretización divide datos numéricos o de fecha y hora en grupos de igual tamaño. El tamaño de discretización predeterminado se calcula según sus datos."

Underneath the text is a text input field for "Tamaño de la discretización" with the value "17".

At the bottom center is a button labeled "Restablecer valores predeterminados".

At the bottom right are two buttons: "Aceptar" (highlighted in yellow) and "Cancelar".

**Figura 39. Discretización de variables.**

Una vez hemos realizado los cálculos necesarios, en función de cada hotel, visualizaremos con cuánta antelación los usuarios realizan la reserva de su estancia en función del nuevo grupo creado con los valores discretizados DIAS\_ATELACION.

Seleccionaremos dicha columna mencionada anteriormente, junto al campo HARD (número de reservas) pertenecientes ambos a la base de datos Reservas, seleccionando la visualización embudo. Para el importe total neto seleccionaremos un medidor mapeando IMPORTE\_TOTAL\_SIN\_IVA dentro de la base de datos Reservas.

El número de total de reservas lo representaremos mediante una tarjeta seleccionando el campo HARD dentro de la base de datos Reservas.

El primer ejemplo que hemos seguido, ha sido el hotel situado en Barcelona, del que podemos visualizar que los usuarios suelen realizar las reservas mayoritariamente con un margen inferior al mes de antelación. Existiendo 26.480 reservas a lo largo de todo el año, cuyo importe neto refleja una cantidad de 22,06 millones de euros.

Otro análisis importante será comprobar el número de huéspedes alojados en el hotel según la fecha de entrada, ya que, de este modo, obtendremos los meses de mayor o menor ocupación del hotel a lo largo de todo el año. Para el hotel de Barcelona, la ocupación aunque se encuentra repartida todo el período, dispone de una mayor concentración entre Mayo y Octubre coincidiendo con una mejor época ambiental.



Figura 40. Reservas según antelación y ocupación en Barcelona.

Si comparamos el caso anterior con el hotel situado en Madrid, podemos observar que los usuarios suelen realizar las reservas de este hotel con hasta un mínimo de un mes de antelación, hay un porcentaje importante de reservas en un período comprendido entre 2 y 8 meses de antelación, existiendo un número destacable de reservas con hasta un año de antelación, siendo éstas un total de 442.

Se refleja un importe neto de 12,99 millones de euros repartidas en las 16,700 reservas existentes.

En cuanto al nivel de ocupación del hotel, según el número de huéspedes por fecha de entrada, podemos observar que la ocupación se mantiene prácticamente estable a lo largo de todo el período anual, notando un ligero descenso entre los meses de Julio y Agosto especialmente, coincidiendo con el período vacacional.

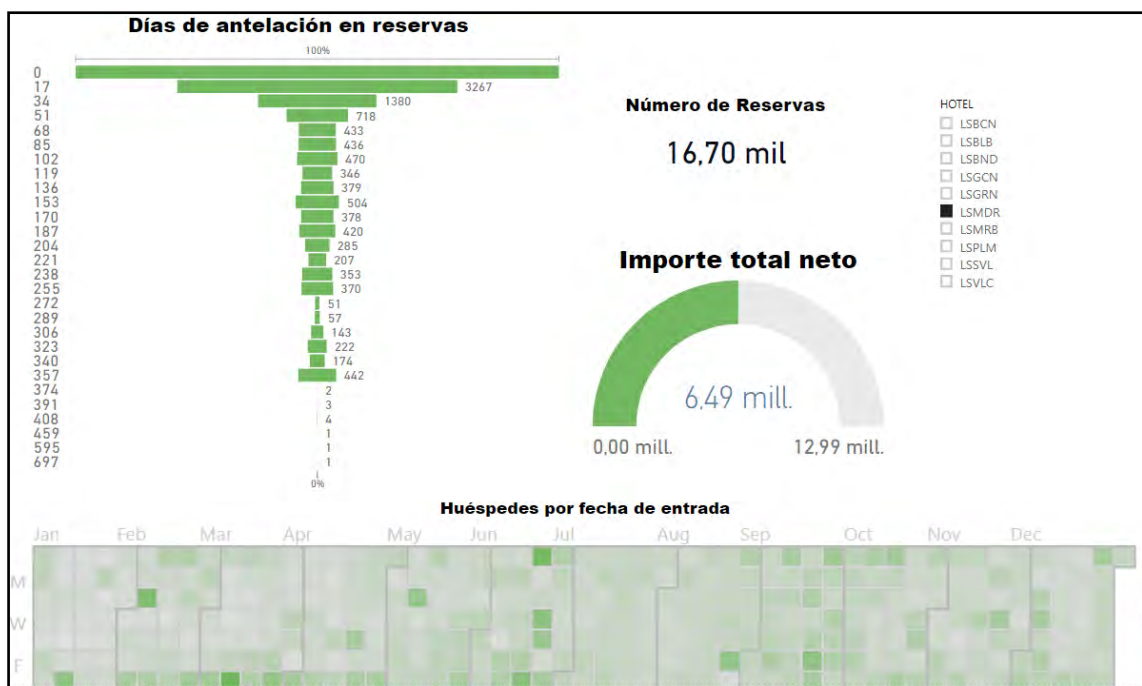


Figura 41. Reservas según antelación y ocupación en Marbella.

Para poder llevar a cabo este conjunto de análisis de la información proporcionada, hemos tenido que administrar las relaciones existentes entre las diferentes bases de datos importadas y creadas.

La clave primaria o *primary key* seleccionada como identificador único para las relaciones de modelado utilizadas ha sido el atributo diferenciador HOTEL.

Este campo nos permite disponer de los tipos de relaciones de modelado Varios a uno (\*: 1) para referirnos desde la base de datos Ubicaciones a la base de datos Reservas y Facturas, y relacionar de este modo los hoteles con los distintos atributos contenidos en las entidades reservas y facturas.

Y la relación Uno a uno (1:1) entre la base de datos o entidad Ubicaciones y la base de datos o entidad Comparación Facturación que hemos creado a través del atributo HOTEL, ya que, el análisis de la facturación tiene que ser única para cada hotel.

Si intentamos relacionar datos que no tienen ninguna relación entre ellos, se indicará que existirá normalidad entre ambos datos relacionados, no existiendo compatibilidad entre los mismos.

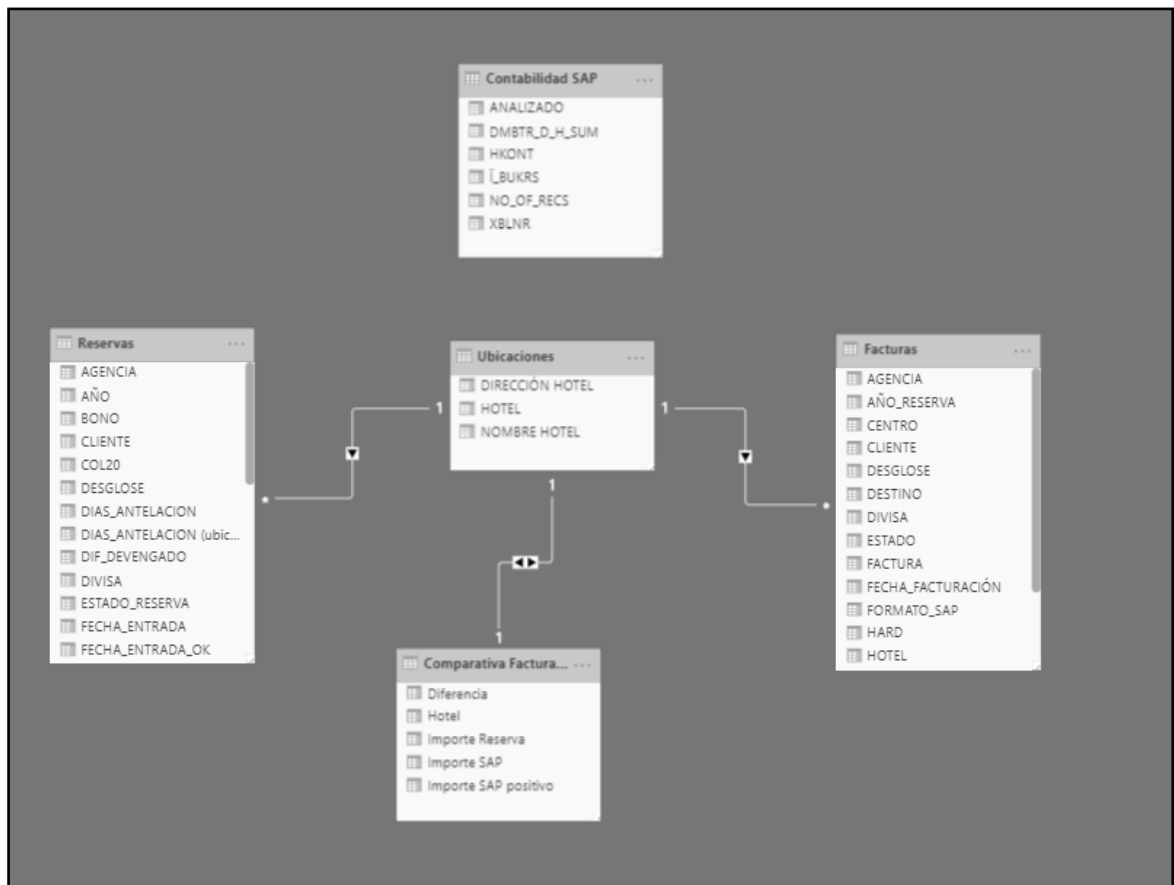


Figura 42. Modelo relacional entre las bases de datos.

## 5 Presupuesto

La construcción y la elaboración del proyecto llevado a cabo, viene determinado por los principales elementos a tener en consideración:

- Ordenador de trabajo
- Licencia de software *CASEWARE IDEA*
- Licencia de software *Microsoft Power BI*

El equipo de trabajo disponible ha sido un ordenador HP con procesador i5 y 8GB de memoria RAM, que actualmente se encuentra en el mercado por un precio alrededor de los 700€.

La licencia del software de la firma *CASEWARE IDEA* se encuentra en torno a los 4000€, a diferencia del software *Microsoft Power BI*, ya que la versión Desktop, es una versión totalmente gratuita y accesible a multitud de herramientas para la elaboración de los informes.

Tomando como precio medio, el precio por hora que la universidad proporciona a los estudiantes según convenios con distintas entidades colaboradoras, estimaremos un precio de 6€/h. Se han necesitado 753 horas de trabajo para la realización de este proyecto, en función del precio medio por hora ponderado, resultaría un total de 4518€.

Por tanto, la suma de todos los importes implicados en la consecución de los objetivos incluidos en el alcance de este proyecto ascenderían a la cantidad de 9218€.



# 6 Conclusiones y trabajo futuro

## 6.1 Conclusiones

El objetivo real del proyecto, era poder analizar y trabajar con casos reales basados en la parte de la Auditoría que desarrolla el área de las Telecomunicaciones y Sistemas de las Tecnologías de la Información.

A lo largo de todo el desarrollo del proyecto, se han manejado grandes volúmenes de datos, que no siempre resultan fáciles de manejar para poder obtener unos resultados concluyentes detallando éstos dentro de los informes generados.

Esta parte de la Auditoría, engloba 2 grandes áreas como son la rama financiera y la tecnológica, siendo una complementaria a la otra, y sin ambas, resulta tedioso y difícil poder realizar un trabajo eficiente y fiel a la realidad.

Es importante el poseer de grandes actitudes y conocimientos en ambas ramas, pues el Auditor debe basar su opinión en la emisión del informe de los estados financieros a través del tratamiento de la información facilitada por la entidad.

Por tanto, una vez hemos finalizado el proyecto, podemos decir que se han cumplido con los objetivos establecidos.

Durante el transcurso del proyecto se han adquirido conocimientos sobre los siguientes temas:

- Conceptos financieros
- Programación en IDEA
- Programación en DAX
- Utilización de las herramientas para el análisis y tratamiento de bases de datos:
  - IDEA
  - *Microsoft Power BI*

## 6.2 Trabajo futuro

La Auditoría Tecnológica se encuentra basada en un proceso de cambio constante. Actualmente, la línea de trabajo está enfocada a través del análisis descriptivo mediante el entendimiento y la recogida de información, pero lo que se está buscando es poder comenzar a implementar el análisis predictivo mediante *Big Data*, Inteligencia Artificial, y algoritmos más complejos en la búsqueda de la eficiencia y el refuerzo del papel del auditor como generador de confianza.

Algunas limitaciones que existen hoy en día, son el almacenamiento de la información por las entidades en servidores físicos, lo cual implica una capacidad y un acceso restringido a la información. Siendo la ubicación de los servidores en la nube, como una posible solución ante estas limitaciones, permitiendo una mayor escalabilidad y acceso en tiempo real a la información.

La introducción del *Big Data* nos permitirá poder trabajar con grandes cantidades de datos no estructurados pudiendo optimizar algunos procesos de manera automatizada.

Por tanto, la implementación de estas posibles soluciones prácticas, serían una buena línea de trabajo futuro a seguir.



## 7 Referencias bibliográficas

- [1] <https://www.emprendepyme.net/auditoria-de-sistemas.html>
- [2] <https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article>
- [3] <http://contaduriapublica.org.mx/2019/04/01/analisis-de-datos-en-la-auditoria/>
- [4] <https://www.auditool.org/blog/auditoria-externa/6432-analisis-de-datos-de-auditoria-ada-en-la-evaluacion-de-riesgos>
- [5] <https://www.analiticanegocios.com/guias/diferencias-datos-estructurados-no-estructurados/>
- [6] <https://normasdeinformacionfinanciera.com/normas-internacionales-auditoria/>
- [7] [https://en.wikipedia.org/wiki/Management\\_assertions](https://en.wikipedia.org/wiki/Management_assertions)
- [8] <https://www.investopedia.com/articles/financial-analysis/063016/what-are-financial-statement-assertions.asp>
- [9] <https://aukera.es/blog/data-science-que-es-y-que-no-es/>
- [10] <https://www.myaccountingcourse.com/accounting-cycle/journal-entries>
- [11] <https://news.microsoft.com/es-xl/power-bi-analitica-de-negocios-para-obtener-informacion-de-valor/>
- [12] <https://www.apd.es/algoritmos-del-machine-learning/>