******

Unamuno Acha, Isabel

2023 - 2024

Máster Universitario en

Análisis de Datos para la Inteligencia de Negocio / Business Analytics

**Análisis de inventario**

TRABAJO DE FIN DE MÁSTER

**ANÁLISIS DE INVENTARIO**

TRABAJO DE FIN DE MÁSTER PRESENTADO EN: Mondragon Unibertsitatea

PARA LA OBTENCIÓN DEL TÍTULO DE: Máster Universitario en Análisis de Datos para la Inteligencia de Negocio / Business Analytics

AUTOR/A: Isabel Unamuno Acha

DIRECTOR/A: XXX

TUTOR/A: Jon Perez Visaires

ORGANIZACIÓN EN LA QUE HA REALIZADO EL PROYECTO: Datua

FECHA DE DEFENSA: Bilbao, a DD de MM de AAAA

El autor/la autora del Trabajo de Fin de Máster, autoriza a la Facultad de Empresariales de Mondragon Unibertsitatea, con carácter gratuito y con fines exclusivamente de investigación y docencia, los derechos de reproducción y comunicación pública de este documento siempre que: se cite el autor/la autora original, y el uso que se haga de la obra no sea comercial.

****

**Reconocimiento – NoComercial – CompartirIgual (by-nc-sa):** No se permite un uso comercial de la obra original ni de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula la obra original.

DECLARACIÓN DE ORIGINALIDAD

Yo Isabel Unamuno Acha

Declaro que este Trabajo de Fin de Máster es original, fruto de mi trabajo personal, y que no ha sido previamente presentado para obtener otro título o calificación profesional.

Las ideas, formulaciones, imágenes, ilustraciones tomadas de fuentes ajenas han sido debidamente citadas y referenciadas.

**RESUMEN**

**LABURPENA**

**ABSTRACT**

**ÍNDICE**

**ÍNDICE DE FIGURAS**

**ÍNDICE DE TABLAS**

# INTRODUCCIÓN

## Presentación de la empresa

## Problematica

## Marco teórico

## Objetivos

El objetivo de este proyecto es mejorar las prácticas de gestión de inventarios de Bibitor, LLC, una empresa minorista de vinos y licores que opera en múltiples ubicaciones. WBibitor, LLC gestiona ventas y costes de bienes vendidos que alcanzan los cientos de millones, lo que hace que las hojas de cálculo tradicionales resulten insuficientes. El vasto volumen de datos, que incluye millones de registros de ventas, compras e inventario, requiere un enfoque sofisticado para un análisis eficaz. el objetivo de este proyecto es aprovechar el análisis exhaustivo de datos para optimizar el control del inventario y extraer información valiosa del funcionamiento de la empresa, en particular de las ventas y las compras.las tareas del proyecto se dividen en dos objetivos principales:

Analizar el proceso de gestión de inventarios y ofrecer recomendaciones para su optimización.

Extraer información empresarial significativa de los datos y ofrecer recomendaciones prácticas.

## Planificación

xxxxxxxx

# DESARROLLO

Xxxxxxxx

<https://www.kaggle.com/datasets/bhanupratapbiswas/inventory-analysis-case-study>

## Análisis de los datos

Teoría bezela:

El análisis de datos constituye un proceso fundamental en la interpretación y utilización efectiva de conjuntos de datos para obtener insights valiosos y respaldar la toma de decisiones informadas. Este proceso implica una serie de pasos interrelacionados que van desde la entrada de datos hasta la optimización de los resultados. A continuación, se han seguido los siguientes pasos para llevar a cabo el análisis.

* Entrada de Datos: Este paso implica la determinación de los requisitos de datos y la recopilación de información relevante. Se realiza mediante la interacción con stakeholders y la obtención de acceso a los datos necesarios.
* Preparación de Datos: La preparación de datos implica la limpieza y consolidación de datos brutos para transformarlos en una forma adecuada para el análisis. Se verifica continuamente para garantizar la calidad y relevancia de los datos.
* Exploración de Datos: En esta fase, se estudia el conjunto de datos a través de técnicas como el muestreo, análisis estadístico y visualizaciones para identificar patrones y tendencias que puedan conducir a insights valiosos.
* Enriquecimiento de Datos: Los datos se enriquecen con información adicional para proporcionar una perspectiva más completa y revelar insights que de otra manera podrían pasar desapercibidos.
* Ciencia de Datos: Se aplican métodos avanzados de análisis de datos, como algoritmos, aprendizaje automático e inteligencia artificial, para obtener insights más profundos y complejos que pueden ser difíciles de extraer con enfoques convencionales.
* Inteligencia Empresarial: La combinación de datos, software y procesos comerciales junto con la intuición humana permite obtener insights accionables que respaldan la toma de decisiones empresariales.
* Generación de Informes: Los resultados del análisis se organizan y presentan de manera efectiva para compartir los conocimientos adquiridos de manera comprensible y útil.
* Optimización: Se ajustan y mejoran continuamente los modelos y procesos de análisis para mantener su relevancia y eficacia en un entorno en constante cambio.

ESTA PODRIA SER LA RUTA A SEGUIR PARA EL APARTADO DE DESARROLLO.

<https://www.alteryx.com/es/glossary/data-analytics>

Para llevar a cabo la lectura de datos y el procesamiento, se ha optado por emplear el entorno de desarrollo integrado Visual Studio Code, aprovechando su funcionalidad y facilidad de uso en la programación en Python, un lenguaje ampliamente utilizado en el ámbito del análisis de datos y la ciencia de datos.

En la fase inicial del proceso, se ha procedido a importar las bibliotecas y paquetes necesarios para la manipulación, análisis y visualización de datos. Entre estas herramientas esenciales, se incluyen Pandas, una biblioteca de manipulación y análisis de datos que proporciona estructuras de datos flexibles y eficientes, así como funciones para la lectura y escritura de datos en diversos formatos; NumPy, una biblioteca fundamental para la computación numérica en Python, que ofrece soporte para matrices y funciones matemáticas de alto nivel; Matplotlib, una biblioteca ampliamente utilizada para la visualización de datos en dos dimensiones, que permite la creación de gráficos estáticos, interactivos y animados; y Seaborn, una biblioteca basada en Matplotlib que proporciona una interfaz de alto nivel para la creación de gráficos estadísticos atractivos y informativos.

La elección de estas herramientas se sustenta en su capacidad para facilitar las tareas de lectura, procesamiento y análisis de datos, así como en su versatilidad y potencia para generar visualizaciones claras y efectivas que permitan comprender y comunicar adecuadamente los resultados obtenidos durante el análisis de datos.

1. Exploración de datos

A continuación, se han cargado los 6archivos Excel con los que se dispone para llevar adelante el trabajo. Estos corresponden a operaciones comerciales del ejercicio cerrado en 2016. El conjunto de datos incluye las siguientes bases de datos: Inventario inicial del 2016, inventario final del 2016, facturas de compra correspondientes a 2016, precios de compra, datos de compras y datos de ventas. Una vez cargados los datos se han impreso las primeras 5lineas de todos los datos para obtener una visión general de los datos. Una vez leídos las bases de datos, se ha comenzado con el análisis exploratorio de ellos, obteniendo una tabla donde se han recogido detalles clave sobre los datos. Entre ellos, se han recogido los nombres de las columnas, el número de filas, el tipo de los datos, el número de valores únicos de cada columna, los valores no informados y el porcentaje de valores no informados frente al total. Así con la foto clave de cada tabla se han sacado varias conclusiones para después proceder a la limpieza y preprocesamiento de los datos.

Se ha visto que la columna *InventoryId* tiene los mismos registros únicos que los totales, lo que sugiere que podrían servir como claves primaria y foránea para unir las tablas. También se ha visto que varias tablas contienen valores no informados que es necesario identificar y tratar adecuadamente. Además, las columnas de tamaño muestran incoherencias en todas las tablas debido a la variación de unidades de medida. Por ello, es necesario convertirlas a una unidad única y coherente. Además de la columna de volumen, en alguna de las tablas también existe la columna tamaño, la cual contiene el mismo dato, por lo que se puede eliminar una de las dos columnas. Hablando del formato de las fechas se ha observado que utilizan diferentes formatos en cada una de las tablas por lo que es conveniente unificar y ponerlos de la misma manera. Y por último, en la columna *VendorName* se pueden ver espacios o signos tras la palabra que se pueden eliminar.

1. Limpieza y procesamiento

En el apartado anterior se han mencionado cuales son los puntos que se deben abordar para llevar a cabo la limpieza y el procesamiento de los datos, por lo que a continuación se han producido dichos cambios.

* 1. Tratamiento de Missings

La integridad y la calidad de los datos son aspectos fundamentales en cualquier análisis de datos o proyecto de investigación. Los valores no informados, comúnmente conocidos como "missing values" en inglés, representan una preocupación significativa en el contexto del análisis de datos, ya que pueden distorsionar los resultados y conducir a interpretaciones erróneas si no se manejan adecuadamente. Por lo tanto, es crucial detectar y abordar de manera efectiva estos valores faltantes.

Los valores no informados pueden surgir debido a una variedad de razones, que van desde errores en la entrada de datos hasta fallos en la recopilación o transferencia de información. Independientemente de su origen, la presencia de valores no informados puede comprometer la validez y la fiabilidad de cualquier análisis realizado sobre los datos afectados.

Durante el análisis del conjunto de datos que comprende varias tablas, se han detectado valores no informados en cuatro de las tablas analizadas. A continuación, se describen las acciones tomadas para abordar estos valores faltantes:

* Tabla “end\_inventory” : En la columna “City se ha identificado la ausencia de valores. Tras un análisis detallado, se ha observado que todos los nombres de ciudades están presentes excepto para una ubicación de tienda específica. Dado este hallazgo, se ha decidido asignar el nombre "TYWARDREATH" a esta ciudad, basándose en el número de tienda correspondiente.
* Tabla “purchases”: Se han encontrado solo tres valores faltantes en la columna "size", lo que representa un porcentaje mínimo del total de registros. Por lo tanto, se ha optado por por eliminar estas filas para preservar la integridad de los datos restantes.
* Tabla “sales”: En la columna "Approval" se muestra un alto porcentaje de valores faltantes, lo que representa el 93% del total de registros. Dada la magnitud de esta ausencia de información, se ha tomado la decisión de eliminar completamente esta columna para evitar cualquier sesgo o distorsión en el análisis posterior.
* Tabla purchase\_prices”: Se han identificado valores faltantes en las columnas "Description", "Size" y "Volume". Dado que estos valores no informados representan registros individuales, se ha decidido eliminar las filas correspondientes para mantener la coherencia y la integridad de los datos restantes.
  1. Tratamiento de duplicados

Mediante el método duplicated en Python, se ha realizado una verificación para determinar la presencia de duplicados dentro del conjunto de datos representado. Se ha guardado la información en una variable llamada duplicados, la cual captura el resultado booleano, proporcionando una indicación clara sobre la presencia o ausencia de duplicados en los datos. Este enfoque es esencial para garantizar la integridad y la calidad de los datos, ya que la presencia de duplicados puede distorsionar los resultados del análisis y conducir a conclusiones erróneas si no se manejan adecuadamente.

* 1. Irregularidades en los datos

Se han observado irregularidades significativas en las entradas de datos, especialmente en la columna "Size", que exhibe variaciones en todas las tablas de datos, con la excepción del conjunto de datos "purchase\_price". Estas discrepancias se refieren a la representación de información de volumen, expresada en diferentes unidades y formatos, como litros, mililitros, onzas, paquetes (pk) y combinaciones de estas unidades. Este panorama heterogéneo implica la necesidad de estandarizar estas medidas en una unidad de volumen única y coherente.

Para abordar esta cuestión, se ha llevado a cabo un análisis exhaustivo de los recuentos únicos en la columna "Size". Aquellos elementos que exhiben un solo recuento y no presentan un patrón discernible han sido asignados a una tasa equivalente estándar. Por otro lado, aquellos elementos con múltiples recuentos y un patrón discernible han sido transformados de acuerdo con estos patrones identificados.

Este enfoque metodológico busca garantizar la consistencia y la comparabilidad de los datos de volumen a lo largo de las diferentes tablas, lo que es esencial para facilitar un análisis coherente y significativo en el contexto de la investigación o análisis de datos en cuestión. La estandarización de las unidades de volumen contribuye a mitigar posibles sesgos o distorsiones en los resultados del análisis, permitiendo así una interpretación precisa y fiable de los datos.

Además, para garantizar la coherencia y la fiabilidad en el manejo de datos temporales, se ha llevado a cabo una normalización del formato de fecha y hora en un DataFrame específico. Es importante destacar que las columnas de fecha en los conjuntos de datos de "Inventario inicial" e "Inventario final" ya se encuentran en un formato adecuado, por lo que no requieren ninguna manipulación adicional. En consecuencia, nos enfocaremos en formatear las columnas de fecha en los demás conjuntos de datos.

Este proceso de normalización del formato de fecha y hora tiene como objetivo principal garantizar la cohesión y la consistencia en el tratamiento de datos temporales, lo que facilita su posterior análisis y visualización. Al estandarizar el formato de fecha y hora en todas las columnas pertinentes del DataFrame, se promueve la precisión y la interpretabilidad de los datos, lo que es esencial para obtener conclusiones sólidas y confiables en cualquier análisis posterior.

Por último, se ha identificado una discrepancia en los nombres de las ciudades presentes en las tablas de "Inventario inicial" y "Inventario final", donde los nombres no corresponden a ubicaciones geográficas reales. Para abordar esta irregularidad, se ha realizado un proceso de corrección mediante el reemplazo de los nombres de las ciudades no reales por aquellos que sí lo son.

Para llevar a cabo esta corrección, se han extraído todos los nombres de ciudades presentes en las tablas y mediante la función replace se han remplazado con una lista de ciudades geográficamente válidas.

Además, se ha observado que el identificador de inventario (InventoryId) estaba influenciado por el nombre de la ciudad. Por lo tanto, se ha procedido a recrear este identificador combinando el número de tienda, el nombre de la ciudad y el número de marca, separados por un guion. Este proceso de recreación del identificador se ha aplicado en las tablas pertinentes donde se encontraba esta relación entre la ciudad y el InventoryId.

Este enfoque de corrección y recreación de registros asegura la coherencia y la precisión de los datos, lo que facilita su interpretación y análisis posteriores. Además, garantiza que los identificadores de inventario reflejen de manera precisa la relación entre la tienda, la ciudad y la bodega, proporcionando así una representación más fiel de la realidad en el conjunto de datos.

Para terminar, se han impreso otra vez todas las tablas, y se ha asegurado que se ha completado la limpieza y procesamiento de los datos, para así empezar con la siguiente fase.

## Análisis y optimización del inventario

Realizar un Análisis de Control de Inventario basado en parámetros como Tiempo de Entrega (LT), Stock de Seguridad, Punto de Reorden y Análisis ABC.

Extraiga información empresarial significativa relacionada con el inventario respondiendo a las siguientes preguntas:

* ¿Qué productos tienen recuentos de inventario elevados al final del período de inventario y pertenecen a la clase A en el análisis de inventario ABC?
* ¿Qué productos tienen un recuento de existencias bajo al final del periodo de inventario, registraron ventas el año anterior y pertenecen a la clase C en el análisis de inventario ABC?
* Identifique los productos con grandes inventarios pero bajas ventas anuales y, a la inversa, utilice la relación entre las ventas anuales y el inventario final para este análisis.
* ¿Los productos con mayores existencias de seguridad registran mayores volúmenes de ventas, y a qué categoría del análisis ABC pertenecen?
* Enumere los 10 principales productos que requieren un reaprovisionamiento inmediato junto con sus cantidades de reaprovisionamiento.
* Determine la contribución de cada categoría a los ingresos totales basándose en el análisis de inventario ABC.
* Identifique los 10 productos principales de cada categoría que más contribuyen a los ingresos totales anuales.
* Identificar los 10 productos de cada categoría que menos contribuyen a los ingresos totales anuales.

Parte III: extracción de información útil sobre compras y ventas

Extraiga información empresarial significativa relacionada con las compras y las ventas respondiendo a las siguientes preguntas:

* ¿Quiénes son los vendedores o proveedores importantes? Evalúelos en función de la cantidad de artículos suministrados, el valor de los mismos y su contribución a los ingresos totales de la empresa. vendedores más importantes en términos de cantidad y valor de los suministros.
* ¿Quiénes son los vendedores o proveedores menos importantes? Evalúelos en función de la cantidad de artículos suministrados, el valor de los artículos y su contribución a los ingresos totales de la empresa. proveedores más importantes en términos de cantidad y valor de los suministros
* ¿Cuál es el plazo medio de ejecución de los pagos de cada proveedor y cómo se correlaciona con la cantidad o el valor de la compra?
* ¿Qué productos se venden más en términos de cantidad e importe de ventas en dólares y, a la inversa, cuáles se venden menos?
* ¿Cómo se comparan los resultados de ventas de las distintas tiendas y cuáles son las 10 tiendas con mayor rendimiento de ventas?
* ¿Qué tamaño de producto se vende más?
* ¿Podemos proyectar e identificar las tendencias de venta a lo largo del año?

POWERBI:

<https://learn.microsoft.com/es-es/power-bi/create-reports/service-dashboards-design-tips>

**Modelo semántico en powerbi:**

La creación de un modelo semántico excelente es una de las tareas más importantes ya que al hacer este trabajo de forma correcta, ayudará a los usuarios a comprender mejor los datos, lo que facilitará la creación de informes de Power BI útiles para ellos y para el creador.

Un buen modelo semántico ofrece las siguientes ventajas: una exploración de datos más rápida, agregaciones más fáciles de crear, informes más precisos, dedicar menos tiempo al crear el informe y mejor seguimiento a futuro.

<https://learn.microsoft.com/es-es/training/modules/design-model-power-bi/1-introduction>

A la hora de crear el modelo, se ha optado por el modelo estrella.

Esquemas de estrella

Puede diseñar un esquema de estrella para simplificar los datos. No es la única manera de simplificarlos, pero es un método popular; por tanto, todos los analista de datos de Power BI deben comprenderlo. En un esquema de estrella, cada tabla del modelo semántico se define como una tabla de hechos o de dimensiones, como se muestra en el siguiente objeto visual.

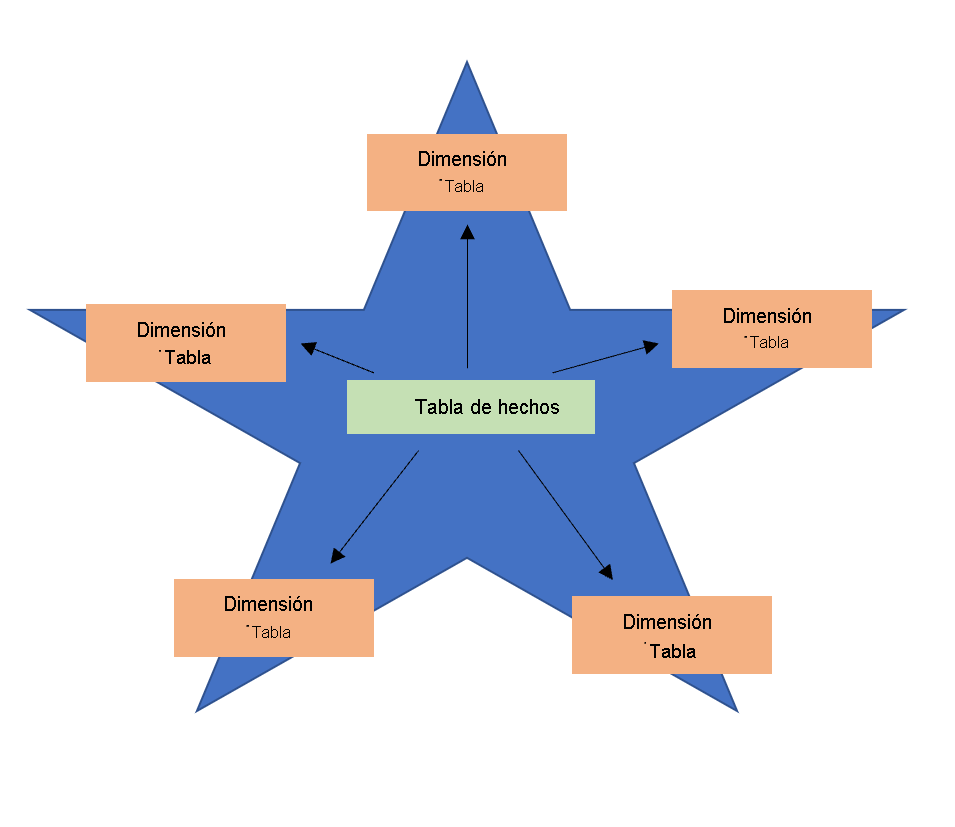


Ilustración del esquema de estrella con una tabla de hechos en el centro y tablas de dimensiones en cada una de las cinco puntas.

Las tablas de hechos contienen valores de datos de eventos o de observación: pedidos de ventas, recuentos de productos, precios, fechas y horas de transacciones, y cantidades. Las tablas de hechos pueden contener varios valores repetidos. Por ejemplo, un producto puede aparecer varias veces en varias filas para diferentes clientes en fechas distintas. Estos valores se pueden sumar para crear objetos visuales. Por ejemplo, un objeto visual del total de pedidos de ventas es una suma de todos los pedidos de ventas en la tabla de hechos. Con las tablas de hechos, es habitual ver columnas rellenadas con números y fechas. Los números pueden ser unidades de medida, como el importe de venta, o pueden ser claves, como un identificador de cliente. Las fechas representan el tiempo que se registra, como la fecha del pedido o la del envío.

Las tablas de dimensiones contienen los detalles sobre los datos de las tablas de hechos: productos, ubicaciones, empleados y tipos de pedido. Estas tablas están conectadas a la tabla de hechos a través de columnas de clave. Las tablas de dimensiones se usan para filtrar y agrupar los datos de las tablas de hechos. Las tablas de hechos, por otro lado, contienen los datos medibles, como ventas e ingresos, y cada fila representa una combinación única de valores de las tablas de dimensiones.

Las tablas de hechos son mucho más grandes que las de dimensiones, porque en ellas se producen numerosos eventos, como ventas individuales. Las tablas de dimensiones suelen ser más pequeñas porque está limitado al número de elementos que puede filtrar y agrupar. Por ejemplo, un año contiene solamente un número fijo de meses y Estados Unidos se compone de un número concreto de estados.

Por ello es muy importante la creación correcta de las tablas, escogiendo únicamente la información necesaria y mostrando de una manera ordenada. Así, como se ha mencionado anterior mente se han creado las siguientes tablas:

M\_CLIENTES

M\_PROVEEDORES

M\_PRODUCTO

M\_TIENDAS

M\_CALENDARIO (<https://learn.microsoft.com/es-es/training/modules/design-model-power-bi/3-date-table>) \* Explicación beyan

H\_VENTAS

H\_COMPRAS

H\_INVENTARIO\_INICIAL

H\_INVENTARIO FINAL

(Explicar la información que se recoge en cada tabla)

Creación de una tabla de fechas común

Se puede crear una tabla de fechas común de estas formas:Datos de origen,DAX,Power Query.

En este caso se ha decidido crear la tabla de calendario mediante DAX.

Puede usar las funciones de expresión de análisis de datos (DAX) CALENDARAUTO() o CALENDAR() para crear la tabla de fechas común. CALENDAR() devuelve un intervalo de fechas contiguo en función de fechas de inicio y de finalización que se especifican como argumentos en la función. Como alternativa, la función CALENDARAUTO() devuelve un intervalo de fechas completo y contiguo que se determina de forma automática a partir del modelo semántico. La fecha de inicio se elige como la más temprana del modelo semántico y la fecha de finalización es la última, además de los datos que se han rellenado para el mes fiscal, que puede optar por incluir como argumento en la función CALENDARAUTO(). Para los fines de este ejemplo, se usa la función CALENDAR() porque solo quiere ver los datos desde el 31 de mayo de 2011 (el primer día en el que las ventas empezaron a realizar el seguimiento de estos datos) en adelante, durante los próximos 10 años. En Power BI Desktop, seleccione **Nueva tabla** y escriba la siguiente fórmula DAX:

Ahora tiene una columna de fechas que puede usar. Pero esta columna es ligeramente dispersa. También quiere ver columnas solo del año, el número del mes, la semana del año y el día de la semana. Puede realizar esta tarea si selecciona **Nueva columna** en la cinta y escribe la ecuación DAX siguiente, que recuperará el año de la tabla de fechas.

## Configuración del modelo semántico y creación de relaciones entre tablas

Aunque se puede realizar el procesos de limpieza de datos en PowerQuery, en este proyecto se ha decido hacerlo mediante visual studio. Por ello, las tablas que se han ingestado en PowerBI cuentan con datos limpios. Así, en el apartado modelo se ha comenzado a crear las relaciones necesarias. (EXPLICACION DE LA CREACION DE RELACIONES)

# RESULTADOS

Xxxx

## Informe en PowerBI

# CONCLUSIONES Y DISCUSIÓN

xxxxxxxx

**BIBLIOGRAFÍA**

|  |
| --- |
| Anexos / Eranskinak |

