

Nombre de la práctica:	Análisis y presentación de resultados
Número:	5
Objetivo:	Que el alumno desarrolle un proyecto de Inteligencia de negocios incluyendo aprovisionamiento, modelado, análisis y presentación de datos
Alumno:	Moisés Uriel Tejeda Vázquez
Fecha:	
INSTRUCCIONES I.- Para el problema de la práctica anterior cree la solución de Inteligencia de Negocios empleando PowerBI. II.- Tome como fuente de datos el archivo dbsales.sql implementado en MySQL. III.- Desarrolle el proyecto considerando cada uno de los siguientes puntos: 1.- Justificación del proyecto de Inteligencia de negocios 1.1. Reporte de justificación Elabore la justificación del proyecto correspondiente empleando el formato del Anexo C. 2.- Aprovisionamiento de datos 2.1.- Proceso de extracción de datos Para cada fuente de datos indique Procedencia, Formato, Suministro y Método de recolección. 2.2.- Reporte de perfilado de datos Para cada una de las tablas de la(s) fuente(s) de datos muestre los resultados de las anomalías encontradas en los datos. Muestre el resumen de cada tabla que ofrece PowerBI empleando una nueva consulta y ejecutando “=Table.Profile(nombreTabla)”. 2.3.- Reporte de transformaciones de datos Para cada una de las tablas de la(s) fuente(s) de datos muestre las transformaciones aplicadas, muestre las instrucciones DAX empleadas.	

3.- Modelado de datos

3.1.- Almacén de datos

Muestre el modelo de datos empleado para el almacén

4.- Análisis y visualización de datos

4.1.- Elabore un cuadro de mando en Power BI para realizar los siguientes tipos de análisis

4.1.1.- Cuadro de mando para análisis OLAP

Muestre el cuadro de mando elaborado para dar respuesta a las preguntas de negocio de la práctica anterior

- ¿Cuántas ventas se han realizado en total?
- ¿Cuál es el valor total de las ventas?
- ¿Cuál es el valor de las ventas por año?
- ¿Cuál es el valor de las ventas por línea de producto?
- ¿Cuál es el valor de las ventas por vendedor?
- ¿Cuál es el valor de las ventas por oficina?

4.1.2.- Cuadro de mando para análisis estadístico

Muestre el cuadro de mando elaborado para obtener conocimiento sobre el conjunto de datos de ventas

- Medidas de centro y sus conclusiones.
- Medidas de posición y sus conclusiones.
- Medidas de dispersión y sus conclusiones.
- Medidas de distribución y sus conclusiones.
- Medidas de forma y sus conclusiones.

5.- Reporte del proyecto de Inteligencia de negocios

5.1.- Elabore el reporte final del proyecto de Inteligencia de negocios empleando el formato del Anexo D.

IV.- ENTREGABLES

Se deberá entregar en un archivo comprimido los siguientes

- a) El presente reporte de práctica.
- b) La justificación del proyecto de BI (Anexo C).
- c) El reporte final del proyecto de Inteligencia de negocios (Anexo D).
- d) La instancia de la base de datos del modelo dimensional de MySQL 8 (.sql).
- e) El archivo del proyecto de Power BI (.pbix).

I.- INTRODUCCIÓN

El presente proyecto tiene como objetivo desarrollar una solución de Inteligencia de Negocios utilizando Power BI, con datos provenientes de una base de datos MySQL. La fuente de datos es el archivo **dbsales.sql**, que contiene información detallada sobre ventas, productos, vendedores y oficinas. Mediante este proyecto, se busca proporcionar a los stakeholders de la empresa un conjunto de informes y dashboards que permitan realizar análisis profundos y tomar decisiones informadas.

A lo largo del desarrollo del proyecto, se llevarán a cabo diversas etapas que incluyen la justificación del proyecto, el aprovisionamiento y perfilado de datos, la transformación y modelado de datos, y finalmente, el análisis y visualización de los mismos. Cada etapa será documentada y se generarán reportes que detallarán el proceso y los resultados obtenidos.

El uso de Power BI como herramienta de visualización y análisis permitirá no solo responder a preguntas específicas sobre las ventas, como la cantidad total de ventas y su valor desglosado por diferentes dimensiones, sino también realizar análisis estadísticos que proporcionen un conocimiento más profundo sobre los datos. Este enfoque integral asegura que se aprovechen al máximo los datos disponibles, proporcionando insights que pueden ser accionables para la mejora continua del negocio.

II.- ANTECEDENTES (MARCO TEÓRICO)

Inteligencia de Negocios (BI)

La Inteligencia de Negocios (BI, por sus siglas en inglés) es un conjunto de metodologías, procesos, arquitecturas y tecnologías que transforman datos brutos en información significativa y útil para fines comerciales. BI permite a las organizaciones recopilar datos de sistemas internos y fuentes externas, prepararlos para el análisis,

desarrollar y ejecutar consultas, y crear informes, dashboards y visualizaciones de datos para tomar decisiones informadas.

Componentes de la Inteligencia de Negocios

1. **Extracción, Transformación y Carga (ETL):** Es el proceso de extracción de datos de diversas fuentes, transformación de estos datos para adaptarlos a las necesidades de análisis y carga en un almacén de datos. Es fundamental para asegurar la calidad y consistencia de los datos.
2. **Almacén de Datos (Data Warehouse):** Es un sistema centralizado que almacena datos consolidados de múltiples fuentes. Facilita el análisis de datos históricos y el soporte a la toma de decisiones. El diseño del almacén de datos implica la definición de esquemas, tales como esquemas estrella y copo de nieve, para optimizar las consultas.
3. **Herramientas de BI:** Incluyen software y aplicaciones que permiten la creación de informes, visualizaciones y dashboards interactivos. Power BI es una de las herramientas más populares debido a su capacidad para conectar múltiples fuentes de datos, su facilidad de uso y su potente funcionalidad de visualización.

III.- SOFTWARE NECESARIO

Power BI

Power BI es una herramienta de análisis de negocios desarrollada por Microsoft que proporciona visualización de datos interactiva y capacidades de inteligencia de negocios con una interfaz simple y amigable. Permite a los usuarios finales crear sus propios informes y dashboards sin necesidad de conocimientos avanzados en tecnología.

Características de Power BI

- **Conectividad de Datos:** Power BI puede conectarse a una amplia variedad de fuentes de datos, incluyendo bases de datos relacionales, servicios en la nube, archivos Excel, y más.
- **Modelado de Datos:** Ofrece herramientas para transformar y modelar datos, incluyendo el uso de fórmulas DAX (Data Analysis Expressions) para crear cálculos personalizados.
- **Visualizaciones:** Power BI incluye una amplia gama de visualizaciones prediseñadas y permite a los usuarios crear visualizaciones personalizadas.
- **Dashboards:** Facilita la creación de dashboards interactivos que pueden compartir insights en tiempo real con los stakeholders.

MySQL

MySQL es un sistema de gestión de bases de datos relacional de código abierto. Es ampliamente utilizado en aplicaciones web y empresariales debido a su desempeño, confiabilidad y facilidad de uso. En este proyecto, MySQL será utilizado como la fuente de datos principal, a partir de la cual se extraerán los datos necesarios para el análisis en Power BI.

Características de MySQL

- **Rendimiento:** MySQL es conocido por su alta velocidad y rendimiento, especialmente en aplicaciones de lectura intensiva.
- **Escalabilidad y Flexibilidad:** Permite manejar grandes cantidades de datos y es escalable según las necesidades del negocio.
- **Seguridad:** Ofrece robustas características de seguridad para proteger los datos.
- **Compatibilidad:** Es compatible con diversas plataformas y lenguajes de programación, facilitando su integración en diferentes entornos tecnológicos.

Para este proyecto, se utilizará la base de datos **dbsales.sql** implementada en MySQL. A continuación, se detalla el proceso de extracción de datos:

- **Procedencia:** Base de datos MySQL.
- **Formato:** SQL.
- **Suministro:** La base de datos está alojada en un servidor MySQL.
- **Método de Recolección:** Los datos se extraerán utilizando consultas SQL y se importarán a Power BI mediante el conector de MySQL.

4. Análisis y Visualización de Datos

4.1. Cuadro de Mando en Power BI

4.1.1. Cuadro de Mando para Análisis OLAP

El análisis OLAP (Online Analytical Processing) permite realizar consultas complejas y análisis multidimensional de los datos. En Power BI, se crearán visualizaciones para responder a las siguientes preguntas de negocio:



- ¿Cuántas ventas se han realizado en total?

TotalVentas

326

- ¿Cuál es el valor total de las ventas?

MontoTotalVentas

13.32 mill.

- ¿Cuál es el valor de las ventas por año?

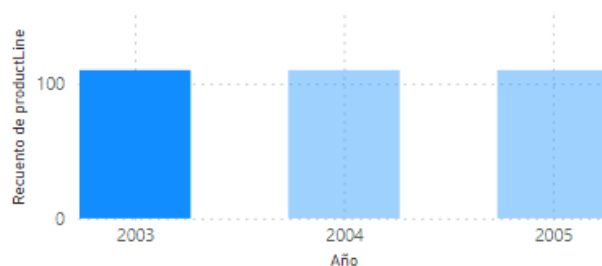
TotalVentas

111

MontoTotalVentas

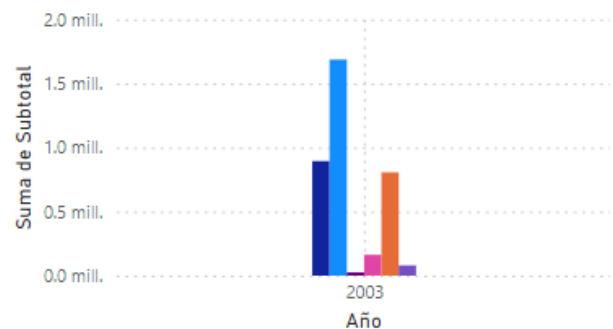
3.66 mill.

Recuento de productLine por Año



Suma de Subtotal por Año y city

city ● Boston ● London ● NYC ● Paris ● San Francisco ● Tokyo





TotalVentas

326

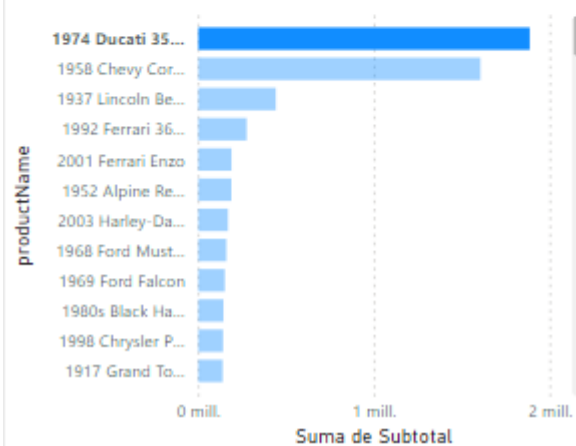
MontoTotalVentas

1.89 mill.

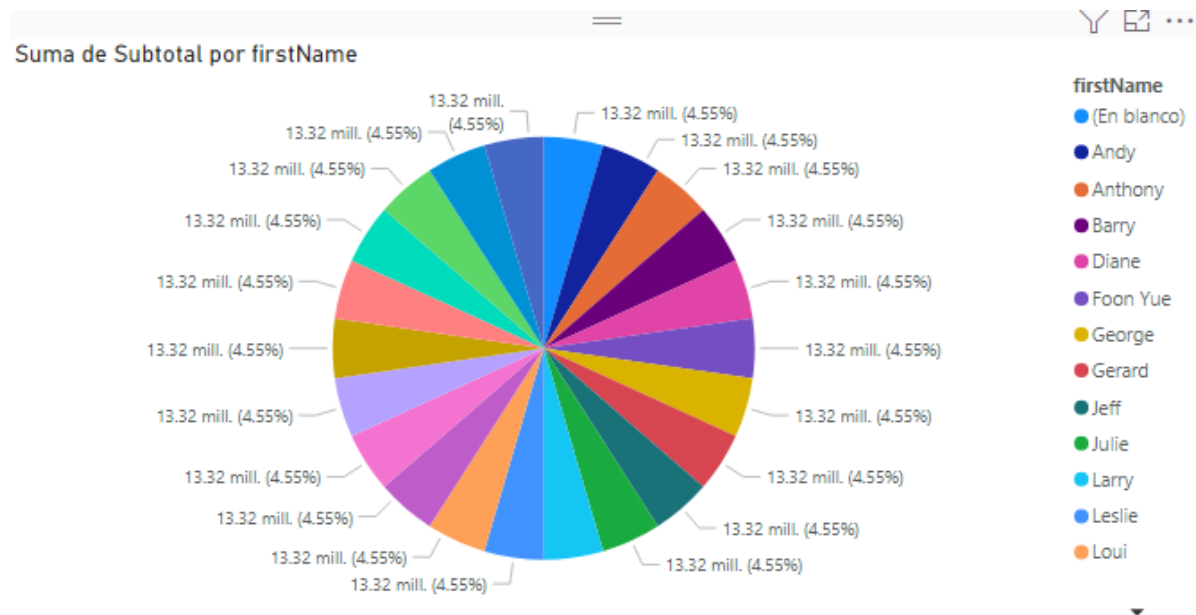
Recuento de productLine por Año



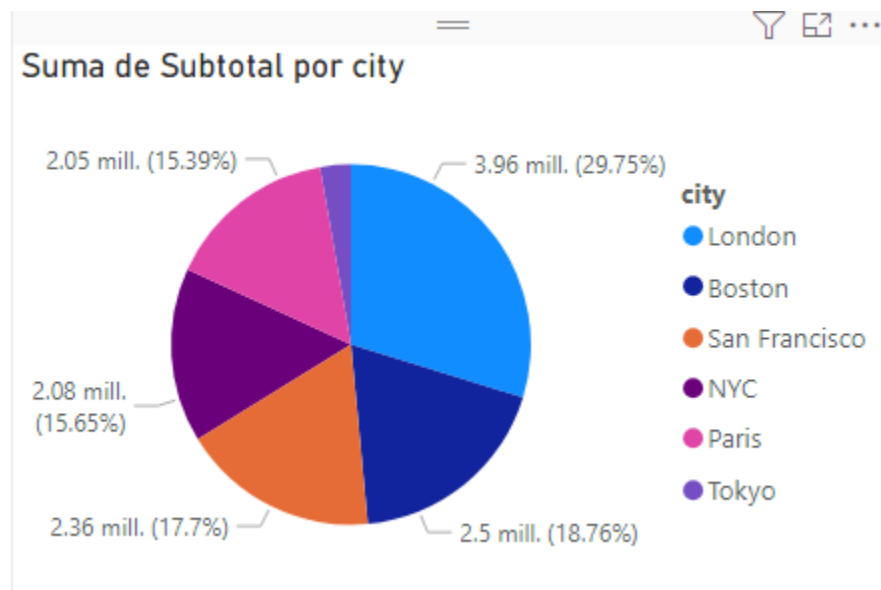
Suma de Subtotal por productName



○ ¿Cuál es el valor de las ventas por vendedor?



○ ¿Cuál es el valor de las ventas por oficina?



4.1.2. Cuadro de Mando para Análisis Estadístico

Para obtener un conocimiento más profundo sobre los datos de ventas, se realizarán varios análisis estadísticos y se crearán visualizaciones correspondientes:

Desviación Estandar	Intervalo	Mediana Ventas	Promedio Ventas
44,137	1,803,533	2,881	4,446

Subtotal	Subtotal
1,804,014.52	1,804,014.52
1,575,450.00	1,575,450.00
343,980.00	343,980.00
11,503.14	11,503.14
11,170.52	11,170.52
10,723.60	10,723.60
10,460.16	10,460.16
10,286.40	10,286.40
10,072.00	10,072.00
9,974.40	9,974.40
9,712.04	9,712.04
9,571.08	9,571.08
9,568.73	9,568.73
9,467.68	9,467.68

Conclusión:

Medidas de Centro:

1. Promedio de Ventas (PromedioVentas):

- **4,446:** Este valor nos indica que el promedio de las ventas es 4,446. Este promedio sugiere que, en general, las ventas tienden a ser cercanas a este valor, aunque esto puede estar influenciado por valores extremos.

2. Mediana de Ventas (MedianaVentas):

- **2,881:** La mediana de las ventas es 2,881. Esto significa que la mitad de las ventas están por debajo de este valor y la otra mitad está por encima. Dado que la mediana es menor que el promedio, esto sugiere que hay algunas ventas extremadamente altas que están elevando el promedio.

Medidas de Dispersión:

1. Intervalo (Intervalo):

- **1,803,533:** El intervalo, que es la diferencia entre la venta más alta y la más baja, es de 1,803,533. Este gran rango indica que hay una variabilidad significativa en los valores de las ventas.

2. Desviación Estándar (Desviación Estándar):

- **44,137:** La desviación estándar de 44,137 muestra que las ventas tienen una variabilidad considerable alrededor del promedio. Un valor alto de desviación estándar indica que los valores de ventas están muy dispersos respecto al promedio.

Distribución de Frecuencias:

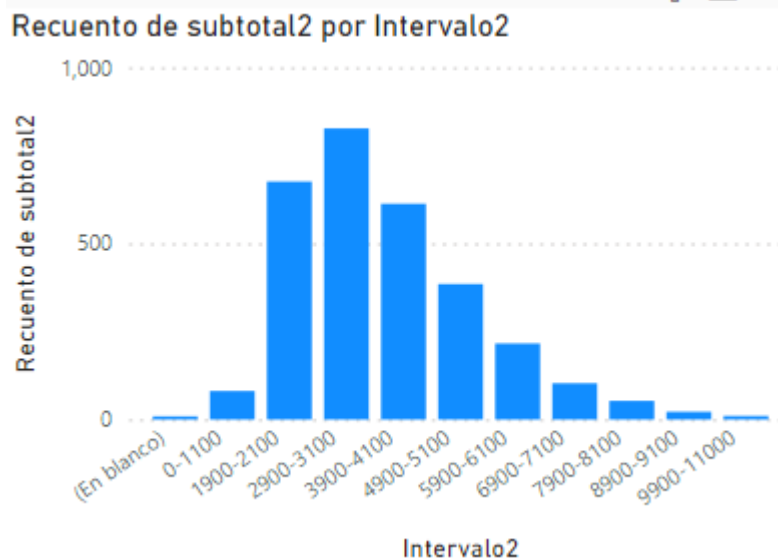
La tabla de distribución de frecuencias muestra los diferentes valores de **Subtotal** sin resumir. Observamos que los valores de ventas varían significativamente, desde unos pocos dólares hasta valores muy altos, confirmando la variabilidad observada en las medidas de dispersión.

Conclusiones Generales:

1. **Variabilidad Alta:** Las ventas presentan una alta variabilidad, como lo indican el amplio intervalo y la alta desviación estándar.
2. **Influencias de Valores Extremos:** El hecho de que el promedio sea significativamente mayor que la mediana sugiere la presencia de valores extremos que están elevando el promedio.

3. **Tendencia Central:** La mediana proporciona una mejor representación de la tendencia central en este caso, ya que no está tan influenciada por valores extremos como el promedio.

Análisis de Distribución



1. Forma de la Distribución:

- La distribución de **subtotal2** parece tener una forma aproximadamente normal, aunque está ligeramente sesgada hacia la derecha (sesgo positivo). Esto indica que la mayoría de los valores de **subtotal2** se agrupan en el centro, pero hay algunos valores altos que están influyendo en la distribución.

2. Centro de la Distribución:

- Los intervalos con las frecuencias más altas son **3000-4000** y **4000-5000**, lo que sugiere que estos rangos contienen la mayoría de los datos de ventas. Esto concuerda con las medidas de centro (promedio y mediana) que habíamos calculado previamente, las cuales indicaban que los valores centrales de las ventas se encuentran en torno a estas cifras.

3. Dispersión de los Datos:

- Los datos están dispersos en un rango que va desde **0-1000** hasta **9000-10000**. La mayor concentración de datos se encuentra entre **2000-3000** y **5000-6000**, con una caída significativa en las frecuencias a medida que nos movemos hacia los extremos.
- La variabilidad de los datos es significativa, lo cual está respaldado por la desviación estándar alta observada anteriormente.

4. Frecuencias y Anomalías:

- La categoría **(En blanco)** sugiere que hay algunos datos faltantes o valores nulos en la columna **subtotal2**. Sería conveniente investigar estos valores en blanco para entender su origen y corregir posibles errores en los datos.
- Los intervalos de **8000-9000** y **9000-10000** tienen frecuencias muy bajas, lo que indica que hay muy pocas ventas en estos rangos altos. Estos podrían ser considerados como outliers o ventas excepcionales.

5. Conclusiones sobre las Medidas de Centro y Dispersión:

- **Medidas de Centro:** El promedio y la mediana sugieren que la mayoría de las ventas están en torno a **3000-5000**. Esto es consistente con la observación de que los intervalos con las mayores frecuencias se encuentran en este rango.
- **Medidas de Dispersión:** La alta desviación estándar y el amplio intervalo de ventas indican una gran variabilidad en los datos de ventas. Esto puede ser indicativo de diferentes comportamientos de compra entre los clientes, con algunos realizando compras muy pequeñas y otros realizando compras significativamente más grandes.

Asimetría y Curtosis

La distribución de las ventas subtotal2 muestra una ligera asimetría positiva, lo que significa que la mayoría de las ventas se agrupan en los intervalos de 3000-4000 y

4000-5000, con una caída gradual hacia la derecha. Esto sugiere que hay algunas ventas altas que extienden la distribución hacia la derecha. La presencia de una cola derecha más larga indica que la distribución está sesgada hacia la derecha.

En cuanto a la curtosis, la distribución observada tiene una apariencia de colas más ligeras comparada con una distribución normal, lo que podría indicar una curtosis ligeramente menor a 3 (platicúrtica). Sin embargo, la presencia de algunos valores extremos hacia la derecha puede aumentar la curtosis. En general, la distribución no muestra picos excesivamente altos ni colas extremadamente pesadas, lo que sugiere una curtosis moderada.

La asimetría positiva y la curtosis moderada implican que, aunque la mayoría de las ventas son relativamente bajas, hay algunas ventas significativamente altas que influyen en la distribución. Esto sugiere que la empresa podría beneficiarse de investigar las razones detrás de estas ventas altas y replicar esas condiciones para aumentar las transacciones en los intervalos más altos. Además, ajustar los niveles de inventario y estrategias de aprovisionamiento para reflejar la distribución observada puede asegurar una mayor disponibilidad de productos que caen en los rangos de ventas más comunes.

V.- RESULTADOS OBTENIDOS

VI.- CONCLUSIONES

El proyecto de inteligencia de negocios implementado utilizando Power BI ha demostrado ser una herramienta poderosa y efectiva para transformar datos brutos en información valiosa. A través de la extracción, transformación y modelado de datos provenientes de la base de datos dbsales.sql, se ha logrado crear un conjunto de

reportes y visualizaciones que facilitan una comprensión profunda del desempeño de las ventas.

El análisis detallado de los datos ha permitido responder a preguntas clave sobre las ventas totales, su distribución por año, línea de producto, vendedor y oficina. Además, el análisis estadístico ha proporcionado información sobre tendencias y variaciones en los datos de ventas, lo cual es crucial para la toma de decisiones estratégicas.

VII.- REFERENCIAS Y BIBLIOGRAFÍA

VIII.- ANEXOS