

Proyecto de Evaluación:

"Análisis de datos para optimización de ventas en e-commerce"

Presentado por:

Barrientos, Sandra k00002692

Herrera, Elizabeth k00002712

Martel, Rodrigo k00002721

Rivera, Edwin k00002715

ÍNDICE

[**Resumen ejecutivo 1**](#_62l5shd43tot)

[**Introducción 2**](#_rbbwrf7uf4bv)

[Objetivos del proyecto 2](#_h6iwx2d5efjm)

[Alcance del proyecto 2](#_4gezvwq1syl2)

[**Metodología 3**](#_1ynxn7shc7pj)

[Diseño de la base de datos 3](#_7mz2238seekd)

[Modelo Snowflake 3](#_6lilr2s3rcrx)

[Utilidad de este modelo en el proceso 3](#_guv4sfdh0sfb)

[Lucidchart 3](#_wmgnrsgckf6g)

[PostgreSQL como DBMS 4](#_gh72dyw5bjwg)

[Puntos considerados para trabajar con PostgreSQL 4](#_mo1urs3yw4ku)

[Recopilación, preparación y manipulación de datos 5](#_h4jp1zn5kfss)

[Importación de datos desde archivos CSV 5](#_jc0s134x5xc)

[Validación de datos 5](#_uf36mj3s53ji)

[Extracción y manipulación de datos 5](#_q526lquudkzh)

[Análisis exploratorio de datos 6](#_z1bttgwyr6t)

[Jupyter Notebook 6](#_isxn0x5hjbag)

[PowerBI 6](#_whfbdmun9z47)

[Visual Studio 7](#_p4x1qyfjxkj2)

[Librerías como herramientas utilizadas 7](#_o4hl6j19dnsi)

[Resultados principales 8](#_kzljsslgp2ya)

[KPIs clave 8](#_885spmmt4jkk)

[Tendencias de ventas 8](#_3zt4fqqnb4nr)

[Análisis de clientes 8](#_kl0r2u2nt7uw)

[Análisis de productos 8](#_wxbo4cnftk73)

[**Conclusiones y recomendaciones 8**](#_37q49ic6ue19)

[Anexos 10](#_xdj9hov6wty9)

# Resumen ejecutivo

Este informe presenta un análisis de los datos de las ventas de un cliente especializado en comercio electrónico, con el objetivo de optimizar sus estrategias de ventas.

Como analistas de datos junior, hemos aplicado los conocimientos adquiridos a lo largo del bootcamp para completar las diferentes fases del proyecto. Estas fases incluyen el diseño e implementación de una base de datos, la extracción y manipulación de datos, análisis exploratorio y la creación de visualizaciones y dashboard.

Nuestro análisis permitió identificar patrones y tendencias clave que han sido fundamentales para proponer recomendaciones basadas en datos, con el fin de mejorar las ventas del cliente. Entre los principales hallazgos se encuentran un aumento significativo en la venta de ciertos productos en regiones específicas y un incremento en las ventas durante determinados períodos del año.

Basándonos en estos hallazgos, proponemos estrategias específicas para optimizar el inventario, ajustar la estrategia de precios y segmentar a los clientes de manera más efectiva, con el objetivo de maximizar las ventas. Estas recomendaciones están diseñadas para ayudar al cliente a tomar decisiones informadas y a mejorar su rendimiento comercial de manera sostenible.

# Introducción

## Objetivos del proyecto

El objetivo principal de este proyecto es analizar los datos de ventas de un cliente especializado en comercio electrónico para optimizar sus estrategias de ventas. Utilizando diversas técnicas de análisis de datos, buscamos identificar patrones y tendencias clave que permitan proponer recomendaciones basadas en datos para mejorar las ventas del cliente.

Además, el proyecto tiene como objetivo aplicar los conocimientos adquiridos durante el bootcamp en un entorno real, abordando todas las fases del análisis de datos, desde el diseño de la base de datos hasta la creación de un dashboard interactivo.

## Alcance del proyecto

El alcance del proyecto abarca las siguientes fases:

1. **Diseño de base de datos**: creación de un diagrama entidad-relación y su implementación en un DBMS.
2. **Extracción y manipulación de datos:** importación de datos desde archivos CSV y ejecución de consultas SQL para la extracción de información relevante.
3. **Análisis exploratorio de datos:** realización de análisis estadístico descriptivo, identificación de patrones y tendencias y generación de visualizaciones.
4. **Creación de dashboard:** desarrollo de un dashboard interactivo en PowerBI para presentar los hallazgos y las métricas clave.
5. **Reporte y presentación:** preparación de un informe escrito detallado y una presentación clara y visual de los hallazgos y recomendaciones.
6. **Consideraciones adicionales:** aplicación de buenas prácticas de codificación, optimización del código y consultas SQL y provisión de código comentado.

# Metodología

## Diseño de la base de datos

### Modelo Snowflake

Se seleccionó el modelo Snowflake para trabajar con la base de datos en estudio. Este es un modelo variante del modelo estrella, que permite ser herramienta visual para diseñar, entender y optimizar las bases de datos.

En el modelo Snowflake las tablas de dimensiones (como venta y producto en este caso) se dividen en subtablas adicionales para normalizarlas. La estructura resultante tiene un aspecto de “copo de nieve” debido a las múltiples ramas que surgen las tablas normalizadas.

#### Utilidad de este modelo en el proceso

* Representar relaciones complejas: muestra las conexiones entre las tablas de hechos y las múltiples capas de tablas de dimensión normalizadas.
* Optimización del almacenamiento: ayuda a visualizar cómo la normalización reduce la redundancia al descomponer las dimensiones.
* Facilitar la consulta de datos: permite identificar cómo se deben unir las tablas para consultas analíticas en el almacén de datos.
* Diseño eficiente del esquema: ayuda a identificar áreas donde el modelo puede ser optimizado o simplificado sin perder integridad de los datos.
* Mejor comprensión y documentación: sirve como documentación visual para desarrolladores, annalisa y administradores de bases de datos.

### Lucidchart

Es una herramienta en línea ampliamente versátil y utilizada para crear diagramas, flujos de trabajo y representaciones visuales de procesos y sistemas.

Se seleccionó esta herramienta por varias razones:

* Interfaz intuitiva y fácil de usar: facilita la creación de diagramas, es amigable y accesible para usuarios tanto principiantes como avanzados.
* Variedad de plantillas y formas: ofrece una amplia biblioteca de plantillas prediseñadas y elementos personalizables para diferentes tipos de diagramas, que se necesitan para modelar bases de datos.
* Colaboración en tiempo real: permite que varios usuarios trabajen en el mismo proyecto de forma simultánea, logrando aportaciones de los integrantes del equipo.
* Integración con otras herramientas: es posible compartir el proyecto mediante otras plataformas y aplicaciones que son utilizadas en las empresas. En este caso, se utilizaron las aplicaciones colaborativas de Google (Drive) y Office 365.
* Accesibilidad multiplataforma: al ser una herramienta basada en la nube se puede acceder al proyecto desde cualquier dispositivo con conexión a internet.
* Exportación y compatibilidad: permite exportar diagramas en múltiples formatos, como PDF, PNG, SVG y otros. Esto facilita la integración de los diagramas en presentaciones, informes y documentación técnica.

### PostgreSQL como DBMS

PostgreSQL es un Sistema de Gestión de Bases de Datos relacional (DBMS, por sus siglas en inglés) de código abierto, reconocido por su potencia, flexibilidad y características avanzadas. Adicionalmente, se seleccionó por ser un DBMS con el que se trabajó a lo largo del bootcamp.

#### Puntos considerados para trabajar con PostgreSQL

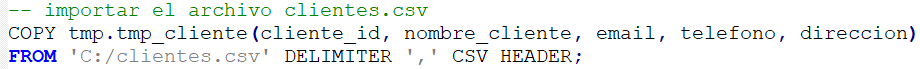
* Es completamente gratuito y no requiere licencias costosas, lo que lo hace ideal para proyectos de cualquier escala.
* Cumple con el estándar SQL y admite características avanzadas, como subconsultas, funciones de ventana, joins complejos, etc.
* Maneja desde pequeñas bases de datos, como en este caso, hasta grandes volúmenes de datos de manera eficiente, permitiendo el fácil manejo si la información en la base de datos crece.
* Es compatible con arquitecturas de nube (como AWS, Google Cloud, Azure) y entornos locales.
* Es altamente extensible, lo que permite personalizarlo y adaptarlo a necesidades específicas, como es el caso de este proyecto.
* Permite crear funciones personalizadas en lenguajes como Python, Java, C, o PL/SQL.
* Ofrece seguridad robusta.

## Recopilación, preparación y manipulación de datos

### Importación de datos desde archivos CSV

La importacion de los datos se realizó desde los archivos CSV provistos por el ciente, utilizando consultas SQL específicas para cargar la información en las tablas temporales creadas en PostgreSQL.

Un ejemplo de las consultas utilizadas es la siguiente:



En esta consulta, se está utilizando el comando COPY para importar los datos desde el archivo clientes.csv ubicado en la ruta especificada. El delimitador de los campos es una coma (,) y se incluye la opción CSV HEADER para indicar que la primera fila del archivo contiene los nombres de las columnas.

#### Validación de datos

Como parte del proceso de importación, se realizaron diversas validaciones para asegurar la integridad y la calidad de los datos. Algunas de estas validaciones incluyeron:

* Verificación de la existencia de valores nulos en campos obligatorios
* Confirmación de que los valores numéricos estén dentro de rangos esperados
* Revisión de la consistencia y unicidad de los identificadores clave

### Extracción y manipulación de datos

Una vez importados y validados los datos, se procedió a la ejecución de consultas SQL para extraer y manipular la información relevante. Algunas de las operaciones realizadas incluyeron:

* Filtrado de datos: selección de registros que cumplieran con ciertos criterios específicos.
* Agregación de datos: cálculo de métricas como sumas, promedios y conteos.
* Uniones de tablas: combinación de datos provenientes de diferentes tablas para obtener una visión integrada.

Estos pasos aseguraron que los datos estuvieran listos para el análisis exploratorio y la generación de visualizaciones en la siguiente fase del proyecto.

## Análisis exploratorio de datos

### Jupyter Notebook

Es una herramienta interactiva basada en la web que permite combinar código, texto, visualizaciones y resultados en un mismo documento. Para los integrantes del proyecto es de fácil acceso e ideal para experimentación, análisis y comunicación de datos.

Otras consideraciones para usar Jupyter Notebook:

* Interactividad: ejecuta y muestra resultados inmediatamente.
* Documentación integrada: incluye texto enriquecido (markdown) junto con el código.
* Visualizaciones: fácil integración con gráficos y bibliotecas como Matplotlib y Seaborn.
* Uso compartido: exporta notebooks como HTML o PDF para compartir ideas.
* Versatilidad: compatible con Python y otros lenguales mediante kernels.

### PowerBI

Es una herramienta de inteligencia de negocios desarrollada por Microsoft que permite analizar datos y crear informes interactivos y visualizaciones. Está diseñada para ayudar a empresas y usuarios a tomar decisiones informadas basadas en datos, es una solución poderosa, flexible y accesible para transformar datos en decisiones estratégicas.

Consideraciones para utilizar PowerBI:

* Amplia biblioteca de gráficos y visualizaciones interactivas.
* Personalización sencilla para adaptar informes a tus necesidades.
* Compatible con cientos de fuentes como Excel, SQL Server, SharePoint, Google Analytics, APIs, etc.
* Uso de DAX (Data Analysis eXpressions) para cálculos personalizados.
* Soporte para modelado avanzado de datos.
* Informes accesibles en la nube, escritorio y dispositivos móviles.
* Publicación y uso compartido de informes con equipos a través del servicio PowerBI o Microsoft Teams.
* Planes gratuitos y opciones de pago asequibles según las necesidades de los usuarios.

### Visual Studio

Visual Studio es un entorno de desarrollo integrado (IDE, por sus siglas en inglés) de Microsoft, diseñado para crear, depurar y desplegar aplicaciones. Es ideal para desarrollar proyectos de software en varios lenguajes como C#, C++, Python, JavaScript, entre otros.

Puntos adicionales por los cuales se utilizó Visual Studio:

* Soporte para múltiples lenguajes y plataformas.
* Herramientas avanzadas para desarrollo y depuración.
* Integración con servicios en la nube y control de versiones.
* Amplia comunidad y recursos de aprendizaje.

### Librerías como herramientas utilizadas

Para el proyecto, fue fundamental utilizar librerías para aprovechar al máximo las capacidades del lenguaje y mejorar la eficiencia del desarrollo.

Los puntos que ayudan al trabajar con librerías son:

✓ Productividad

✓ Ahorro de tiempo

✓ Código reutilizable

✓ Calidad y confiabilidad

✓ Facilidad de uso

Librerías estándar: incluidas en el lenguaje de programación (por ejemplo, math o datetime en Python, os ).

Librerías externas: Pandas, Numpy o Matplotlib en Python, dotenv.

Librerías para áreas concretas como:

* Análisis de datos: Pandas
* Visualización: Matplotlib, Seaborn
* Bases de datos: SQLAlchemy

## Resultados principales

### KPIs clave

* Ventas totales: las ventas totales representan la suma de todas las ventas realizada durante el período disponible y analizado. Este KPI proporciona una visión general del volumen de ventas y es fundamenta para evaluar el rendimiento general.
* Número de transacciones: refleja la cantidad todal de ventas realizadas, lo que ayuda a entender la actividad comercial y la frecuencia de las compras.

### Tendencias de ventas

* Se analizaron las ventas a lo largo del tiempo, segmentadas por trimestre, utilizando la tabla calendario creada. Esto permitió identificar patrones estacionales y tendencias a lo largo del año.
* Las ventas segmentadas por región mostraron que la región Este lidera en términos de volumen de ventas.
* Las ventas segmentadas por categoría mostraron que Hogar lidera en términos de volumen de ventas.

### Análisis de clientes

* Clientes con mayor número de compras: se identificaron los 8 clientes que más compras realizaron, tanto en términos de cantidad como de valor, proporcionando una visión de los clientes más frecuentes y valiosos en términos de ingresos, respectivamente.
* Clientes con menor número de compras: se identificaron también los clientes con menor actividad para identificar posibles oportunidades de mejora.

### Análisis de productos

* Productos más vendidos: se identificaron los productos más vendidos por categoría y región, identificando los productos más populares entre los cientes.
* Productos menos vendidos: se analizaron los productos con menor rotación para ajustar estrategias de inventario.

# Conclusiones y recomendaciones

El análisis permitió identificar patrones clave en las ventas y el comportamiento de los clientes. Las ventas totales proporcionan una visión clara del desempeño general, mientras que el análisis geográfico y de productos ofrece información detallada para la toma de decisiones.

Entre las recomendaciones que podemos ofrecer están:

**Optimización de inventario**

Basado en el análisis de las ventas por producto y región, se recomienda ajustar el inventario para asegurar la disponibilidad de productos de alta demanda y reducir el exceso de stock en productos con baja rotación.

**Estrategias regionales**

Las regiones con menores ventas pueden beneficiarse de campañas de marketing específicas.

**Fidelización de clientes**

Se deben implementar programas de fidelización para los clientes más frecuentes y estrategias para reactivar a los clientes con menos compras.

**Tendencias de ventas**

Las ventas mostraron un patrón estacional con picos en la segunda mitad del año. Esto sugiere que las estrategias de marketing y promoción deben ajustarse para aprovechar estos picos de demanda.

**Distribución geográfica**

La región Este lidera en términos de volumen de ventas, seguida por la región Norte. Esto indica una fuerte presencia comercial en esa región, lo que puede ser aprovechado para expandir operaciones o mejorar la distribución.

# 

# Anexos