

## Proyecto de Evaluación:

### "Análisis de Datos para Optimización de Ventas en E-commerce":

---

**Descripción del Proyecto:** Como analista de datos junior, has sido contratado por una empresa de consultoría para servir a un cliente especializado en comercio electrónico para analizar sus datos de ventas y proponer estrategias de mejora. Este proyecto te permitirá aplicar los conocimientos adquiridos a lo largo del bootcamp. Completa las siguientes fases de manera ordenada, siguiendo las indicaciones detalladas a continuación:

#### 1. Diseño de Base de Datos

##### 1.1. Diagrama Entidad-Relación (ERD)

- Identifica las entidades principales: ventas, clientes, productos y transacciones.
- Define los atributos clave de cada entidad, como ID de cliente, nombre del producto, fecha de la transacción, etc.
- Establece las relaciones entre las entidades, asegurándote de que estén correctamente normalizadas para evitar redundancias.
- Utiliza una herramienta de diagramación como Draw.io, Lucidchart, o cualquier otra de tu preferencia para crear el ERD.
- Revisa el diagrama para asegurarte de que refleja correctamente todos los aspectos relevantes del negocio.

##### 1.2. Implementación en el DBMS

- Selecciona un DBMS con el que te sientas cómodo (PostgreSQL).
- Transforma el ERD en una serie de sentencias SQL CREATE TABLE, asegurándote de incluir claves primarias, claves foráneas y restricciones necesarias.
- Implementa las tablas en el DBMS y verifica que estén correctamente estructuradas.

- Realiza pruebas para garantizar que las relaciones entre tablas funcionen según lo esperado, utilizando comandos JOIN para verificar integridad referencial.

## 2. Extracción y Manipulación de Datos

### 2.1. Importación de Datos

- Inspecciona los archivos CSV (ventas.csv, clientes.csv, productos.csv) para entender su estructura.
- Utiliza comandos como LOAD DATA INFILE o herramientas de importación en el DBMS para cargar los datos en las tablas correspondientes.
- Asegúrate de que los datos se importen correctamente y que los tipos de datos sean consistentes con el diseño de la base de datos.
- Realiza validaciones para detectar y corregir posibles errores en la importación (por ejemplo, campos nulos inesperados o datos inconsistentes).

### 2.2. Consultas SQL para Extracción de Información

- Escribe consultas SQL para calcular:
  - **Ventas totales por categoría de producto:** Usa GROUP BY en la categoría del producto y calcula la suma de las ventas.
  - **Clientes con mayor valor de compra:** Usa funciones de agregación como SUM para calcular el total gastado por cada cliente, ordenando los resultados de mayor a menor.
  - **Productos más vendidos por región:** Combina GROUP BY y ORDER BY para determinar los productos más populares en cada región, basándote en la cantidad total vendida.
- Guarda estas consultas y sus resultados para su inclusión en el informe final.

## 3. Análisis Exploratorio de Datos

### 3.1. Análisis Estadístico Descriptivo

- **Objetivo:** Obtener un panorama general de los datos de ventas mediante estadísticas descriptivas.

- **Instrucciones:**

- Utiliza Python (con librerías como Pandas y NumPy) o R para realizar cálculos de estadísticas básicas, como medias, medianas, desviaciones estándar, y distribuciones.
- Identifica variables importantes, como el número de transacciones, ventas promedio por cliente, etc.
- Analiza la distribución de las ventas para detectar patrones, como ventas estacionales o concentraciones en determinados productos.

### **3.2. Identificación de Patrones, Tendencias y Anomalías**

- Analiza las series temporales de ventas para identificar tendencias ascendentes o descendentes.
- Usa técnicas de agrupamiento para segmentar a los clientes según comportamiento de compra, detectando grupos clave (por ejemplo, clientes frecuentes vs. esporádicos).
- Identifica anomalías, como picos de ventas inusuales, productos que no se venden como se esperaba o clientes con comportamientos fuera de lo común.

### **3.3. Generación de Visualizaciones**

- Usa herramientas de visualización como Matplotlib, Seaborn (Python) o ggplot2 (R) para crear gráficos de barras, líneas, histogramas, y diagramas de dispersión.
- Asegúrate de que cada visualización esté claramente etiquetada y acompañada de un breve análisis de lo que representa.
- Considera la creación de mapas de calor o gráficos de series temporales para resaltar tendencias y patrones.

## **4. Creación de Dashboard**

### **4.1. Desarrollo del Dashboard en Power BI**

- Importa tus datos procesados a Power BI.

- Desarrolla un resumen de KPIs clave (ventas totales, margen de beneficio, número de transacciones, etc.).
- Implementa gráficos que muestren las tendencias de ventas a lo largo del tiempo, segmentadas por diferentes criterios (como producto o región).
- Crea un mapa interactivo que muestre la distribución geográfica de las ventas, permitiendo filtrado por región o producto.
- Desarrolla una sección de análisis de la cesta de compra, mostrando qué productos suelen comprarse juntos y cómo esto puede influir en las ventas futuras.

## **5. Implementación de Modelo Predictivo (Opcional)**

### **5.1. Creación del Modelo de Machine Learning**

- Selecciona un subconjunto de datos relevante para el modelado, asegurándote de que esté bien preparado (limpieza de datos, normalización, etc.).
- Prueba diferentes algoritmos de machine learning (como regresión lineal, árboles de decisión, random forest) para determinar cuál es más adecuado.
- Entrena el modelo con un conjunto de datos de entrenamiento y valida su rendimiento con un conjunto de datos de prueba.
- Evalúa el modelo utilizando métricas como precisión, recall, RMSE (Root Mean Squared Error) u otras que consideres apropiadas.

### **5.2. Documentación del Proceso de Modelado**

- Documenta tu razonamiento detrás de la selección del modelo, incluyendo las características que consideraste y cómo realizaste la validación cruzada.
- Incluye gráficos o tablas que ilustren el rendimiento del modelo y justifiquen por qué es el más adecuado.
- Describe cualquier ajuste de hiperparámetros que hayas realizado y cómo esto mejoró el rendimiento del modelo.

## **6. Reporte y Presentación**

## 6.1. Preparación del Informe Escrito

- **Resumen Ejecutivo:** Escribe un resumen de 1-2 páginas que destaque los hallazgos más importantes y las recomendaciones clave.
- **Metodología:** Describe los métodos y herramientas utilizadas en cada fase del proyecto, explicando por qué fueron elegidas.
- **Insights Principales:** Resalta los descubrimientos más relevantes del análisis de datos, utilizando gráficos y estadísticas para respaldar tus conclusiones.
- **Recomendaciones:** Propón estrategias basadas en los datos analizados para mejorar las ventas, como la optimización de inventario, ajustes en la estrategia de precios, o segmentación de clientes.

## 6.2. Creación de la Presentación

- Prepara una presentación de aproximadamente 10 minutos, con un enfoque en los puntos más importantes del informe.
- Usa diapositivas claras y visuales, evitando sobrecargar con texto; apóyate en gráficos y tablas para ilustrar los puntos clave.
- Practica la presentación para asegurarte de que el contenido se ajusta al tiempo asignado y que es comprensible para una audiencia técnica y no técnica.

## 8. Consideraciones Adicionales

### 8.1. Aplicación de Buenas Prácticas de Codificación

- Sigue las convenciones de codificación recomendadas para los lenguajes utilizados (PEP 8 para Python, por ejemplo).
- Implementa un manejo de errores robusto para evitar que fallos inesperados interrumpan la ejecución del código.
- Usa control de versiones para gestionar el desarrollo del proyecto, realizando commits frecuentes y descriptivos.

### 8.2. Optimización del Código y Consultas SQL

- Revisa tus consultas SQL para asegurarte de que estén optimizadas, evitando operaciones costosas como subconsultas innecesarias.
- Utiliza índices en las bases de datos donde sea relevante para mejorar la velocidad de las consultas.
- Optimiza el código de análisis, evitando cálculos redundantes y utilizando técnicas como el vectorizado en Python.

### **8.3. Provisión del Código con Comentarios**

- Asegúrate de que cada parte del código esté bien comentada, explicando la lógica detrás de bloques de código clave.
- Organiza el código en archivos y carpetas de manera estructurada, facilitando su navegación.
- Incluye instrucciones sobre cualquier dependencia o configuración necesaria para ejecutar el código.

### **8.4. Inclusión de un README**

- El README debe incluir:
  - Descripción del proyecto.
  - Instrucciones para la configuración del entorno.
  - Detalles sobre cómo correr el código y reproducir los análisis y visualizaciones.
  - Información sobre el autor y cualquier otra referencia relevante.

## **Proceso de Entrega**

### **Subida de Archivos a GitHub**

- Crea un repositorio público o privado en GitHub y sube todos los archivos relevantes: código fuente, informes, presentaciones, y archivos de Power BI.

- Asegúrate de que el repositorio esté bien organizado, con carpetas claramente etiquetadas y un README principal.
- Proporciona instrucciones en el README sobre cómo acceder y utilizar los archivos.

### **Preparación para la Sesión de Q&A**

- **Objetivo:** Demostrar un profundo conocimiento del proyecto y defender tus conclusiones.
- **Instrucciones:**
  - Prepárate para responder preguntas sobre la metodología utilizada, decisiones tomadas durante el análisis y recomendaciones propuestas.
  - Revisa posibles preguntas que podrían surgir sobre el rendimiento del modelo predictivo, la integridad de los datos, o la aplicabilidad de tus recomendaciones.
  - Practica explicaciones claras y concisas para temas complejos, asegurándote de que puedas comunicarte de manera efectiva tanto con audiencias técnicas como no técnicas.

**Plazo Final de Entrega:** 9 de Diciembre 11:59PM.

**Defensa:** 12 de Diciembre 6:00PM