

DOCUMENTACION

Planificación:

- ¿Cómo se dividieron las tareas entre los miembros del equipo?

Danny Bryan Tejaxun Pichiyá	Susan Pamela Herrera Monzon
-Conexión de base de datos -Normalización y carga de datos -Limpieza de datos -Preparación de columnas -Obtención de datos de la base de datos. - Documentación	- Script de base de datos - Análisis de datos - Generación de gráficos - Validaciones y conclusiones - Documentación

- ¿Qué herramientas y tecnologías decidieron utilizar y por qué?
 1. **MySQL**
para gestionar la base de datos y las relaciones entre tablas.
 2. **Python + Pandas**
para la manipulación y análisis de datos.
 3. **Matplotlib**
para visualizaciones personalizadas de las métricas clave.
 4. **Jupyter Notebook**
para documentar el proceso paso a paso y generar un reporte reproducible.
 5. **Seaborn**
Para documentar los datos mediante generación de graficas.
 6. **SQLalchemy**
Para establecer la conexión con la base de datos.

- ¿Cómo establecieron los plazos para cada fase del proyecto?

Sprint 1	Sprint 2	Sprint 3
<ul style="list-style-type: none"> - Planeación y repartición de tareas. - Creación de servidor para base de datos - Normalización 	<ul style="list-style-type: none"> - Lectura de datos con pandas. - Limpieza de datos con pandas. - Conexión de base de datos - Carga de datos con pandas. 	<ul style="list-style-type: none"> - Análisis de datos y reporte en jupyter - Realización de graficas basadas en los resultados. - Documentación y presentación.

Proceso de análisis:

- Describa el enfoque paso a paso que siguieron para limpiar y preparar los datos.

- Conversión de columnas de fechas con **``pd.to_datetime``**.
- Manejo de valores nulos (imputación o eliminación según el caso).
- Normalización de categorías (por ejemplo, estandarizar nombres de categorías de productos).
- Creación de columnas derivadas como **``año_mes` y `grupo_edad``**.

- Explique las decisiones tomadas durante el análisis exploratorio de datos.
- Agrupamos por mes para encontrar tendencias de ventas estacionales.
- Calculamos totales por cliente, producto y categoría para descubrir patrones de compra.
- Segmentamos clientes en grupos de edad para observar diferencias de comportamiento.

- Detalle los desafíos encontrados durante el análisis y cómo los superaron.

La columna `fecha_compra` no estaba en formato datetime.

La función `mode()` en `agg()` causaba error.

Metodología

- Barras para comparar categorías (productos más vendidos, regiones de envío, métodos de pago).
- Barras apilada y barras simples normalizadas para mostrar proporciones de métodos de pago por categoría.
- Dispersión (scatter plot) para analizar correlaciones (edad vs total de compra).