

Proyecto #2: EDA – Exploratory Data Analysis

Para completar el EDA el grupo deberá desarrollar las siguientes pasos.

1. Crear una carpeta para desarrollar el proyecto con las siguientes consideraciones:
 - a. Descargar la plantilla de cookiecutter para definir la estructura de proyecto.
 - b. Crear el notebook 01_exploratory_data_analysis.ipynb.
 - c. Crear un repositorio local con git.
 - d. Agregar la configuración que considere necesaria en el archivo .gitignore.
 - e. Publicar el repositorio en github.
 - f. Cada miembro del equipo deberá mostrar un screenshot donde hace un pull del repositorio para trabajar de forma local en su computadora.
2. Comprensión inicial del dataset
 - a. Cargar el dataset y visualizar las primeras filas.
 - b. Mostrar el tamaño del dataset (número de filas y columnas).
 - c. Describir brevemente qué representa cada columna (sucursal, producto, fecha, ventas u otras variables relevantes).
 - d. Identificar el rango temporal de los datos (fecha mínima y máxima).
3. Validación de la estructura de los datos
 - a. Verificar los tipos de datos de cada variable (fecha, numérica, categórica).
 - b. Confirmar que la columna de fecha esté correctamente formateada como tipo datetime.
 - c. Validar que la frecuencia temporal sea consistente (diaria, semanal, mensual).
4. Análisis de valores faltantes y duplicados
 - a. Identificar columnas con valores nulos y el porcentaje de datos faltantes.
 - b. Definir la estrategia para tratarlos (eliminación, imputación, etc.)
 - c. Verificar y justificar cómo se manejarán registros duplicados.
5. Análisis descriptivo general
 - a. Calcular métricas estadísticas básicas: media, mediana, mínimo, máximo, desviación estándar.
 - b. Analizar distribución de ventas por:
 - i. Producto
 - ii. Sucursal
 - iii. Periodo de tiempo (día, semana, mes, año)
6. Exploración temporal
 - a. Graficar la evolución de ventas a lo largo del tiempo.
 - b. Identificar patrones de:

- i. Tendencia
 - ii. Estacionalidad
 - iii. Picos o caídas abruptas
 - iv. Analizar si hay ciclos por días de la semana o meses del año.
- 7. Análisis comparativo entre sucursales y productos
 - a. Identificar productos más vendidos y menos vendidos.
 - b. Comparar desempeño entre sucursales.
 - c. Detectar outliers o comportamientos anómalos por sucursal o producto.
- 8. Visualizaciones clave (obligatorias)
 - a. Series de tiempo de ventas totales.
 - b. Gráfico de barras comparando ventas por sucursal.
 - c. Gráfico de barras comparando ventas por producto.
 - d. Heatmap o matriz para visualizar comportamiento de ventas en el tiempo y por sucursal/producto.
- 9. Para cada serie temporal específica (ventas totales, ventas por sucursal, ventas por producto y ventas por sucursal-producto).
 - a. Aplicar:
 - i. Gráfico de Autocorrelación (ACF): muestra la dependencia entre la serie y sus rezagos.
 - ii. Gráfico de Autocorrelación Parcial (PACF): ayuda a identificar la cantidad de términos AR (autoregresivos) que deben incluirse.
 - b. Interpretar:
 - i. ¿Existen correlaciones significativas a diferentes rezagos?
 - ii. ¿Se observa estacionalidad reflejada en patrones repetitivos?
- 10. Insights y conclusiones del EDA
 - a. Identificar variables que afectan directamente las ventas.
 - b. Detectar patrones útiles para el modelo de forecasting.
 - c. Concluir sobre la calidad y preparación del dataset para las etapas siguientes del proyecto.