

1. Proyecto

1.2. Entrega 2 - Exploración y Análisis de Datos (EDA)

Objetivo:

Realizar una exploración y análisis exhaustivo del conjunto de datos proporcionado, utilizando técnicas de EDA para comprender la estructura de los datos, identificar patrones y preparar los datos para el modelado.

Descripción de la segunda actividad:

1. Exploración Inicial de Datos:

- **Carga del Dataset:** Cargar el conjunto de datos en el entorno de trabajo. Asegurarse de que los datos se hayan cargado correctamente y se encuentren en el formato adecuado.
- **Revisión General:** Realizar una revisión general del conjunto de datos para obtener una visión inicial de su estructura. Incluyan el número de filas y columnas, nombres de columnas, y tipos de datos de cada columna.

2. Análisis Univariado:

- **Distribución de Variables:** Realizar un análisis de las variables individuales para entender su distribución. Utilizar histogramas, boxplots y estadísticas descriptivas (media, mediana, desviación estándar, etc.) para las variables numéricas. Para las variables categóricas, calcular frecuencias y proporciones.
- **Identificación de Outliers:** Detectar posibles outliers en las variables numéricas utilizando técnicas como el análisis de boxplots o el cálculo del z-score.

3. Análisis Bivariado:

- **Relaciones entre Variables:** Analizar la relación entre pares de variables para identificar patrones o correlaciones. Utilizar gráficos de dispersión (scatter plots) y mapas de calor de correlación (heatmaps) para evaluar la relación entre variables numéricas.
- **Comparaciones Categóricas:** Explorar cómo las variables categóricas afectan a las variables numéricas, utilizando gráficos de barras y boxplots.

4. Limpieza de Datos:

- **Tratamiento de Valores Faltantes:** Identificar valores faltantes en el conjunto de datos. Aplicar métodos adecuados para el tratamiento de estos valores, ya sea mediante imputación, eliminación de filas/columnas o sustitución con valores medianos/modales.
- **Normalización y Transformación:** Realizar transformaciones necesarias para preparar los datos para el análisis posterior. Esto puede incluir normalización y estandarización.

5. Preparación del Informe:

- **Documentación del Proceso:** Documentar detalladamente el proceso de EDA, incluyendo los hallazgos importantes, las decisiones tomadas durante la limpieza y las interpretaciones de las visualizaciones realizadas.
- **Informe Final:** Preparar un informe que resuma los análisis realizados, las visualizaciones generadas y las conclusiones obtenidas. Este informe debe ser claro y proporcionar una comprensión completa del conjunto de datos.

Entregables:

• Informe de EDA:

- **Contenido:** El informe debe incluir la descripción del conjunto de datos, el análisis univariado y bivariado realizado, la limpieza de datos y cualquier transformación aplicada. Incluyan gráficos, tablas y visualizaciones relevantes.
- **Formato:** El informe debe tener entre 3-7 páginas y puede ser presentado en un documento PDF o similar.

• Presentación de Resultados:

- **Contenido:** Preparar una presentación de 5 minutos para exponer los principales hallazgos del EDA al supervisor. La presentación debe destacar las observaciones clave y las decisiones de limpieza y transformación de datos.
- **Formato:** La presentación puede incluir diapositivas con gráficos y visualizaciones para apoyar los puntos clave.

Rubrica

Las evaluaciones se basarán en:

- **Análisis y Comprensión:** La profundidad del análisis univariado y bivariado y la capacidad para identificar patrones y relaciones significativas.
- **Calidad de la Limpieza de Datos:** La efectividad en el tratamiento de valores faltantes y la calidad de las transformaciones aplicadas.
- **Documentación y Comunicación:** La claridad y exhaustividad del informe, y la capacidad de comunicar los resultados de manera efectiva en la presentación.

◀ Entrega Final

Ir a...

Sala de Coloquio - Promoción ▶