

Universidad Tecnica Nacional

Minería de datos - ISW 911

Laboratorio # 3

Asignación Grupal - 2 personas.

- Utilice uno de los documentos que están disponibles en la carpeta del DRIVE con nombre **“Archivos”**
 - **Rentadebicis.xlsx / csv**
 - **Plan_de_compras_2022.xlsx**
 - **Estadisticaspoliciales2023.xlsx**
 - **Estadísticas_vinos.xlsx / csv**
- Cree un proyecto GIT para guardar el laboratorio, este será el enlace que deberá cargar al aula en línea para la revisión, asegúrese de asignar los derechos al usuario **utn-frocha**
- Use R Studio para realizar las siguientes tareas EDA:
 1. Cargue el set de datos en su ambiente de trabajo.
 - a. Exploración Inicial de los Datos:
 - i. Obtener información sobre las columnas y tipos de datos (5 puntos)
 - ii. Resumen estadístico de las variables numéricas (5 puntos)
 - b. Limpieza y Tratamiento de Datos:
 - i. Determine valores nulos (10 puntos)
 - ii. Determine valores atípicos (10 puntos)
 - iii. Determine concentraciones y tendencia de datos (10 puntos)
 - iv. Determine acciones correctivas (imputación) (15 puntos)
 - c. Visualización de datos.
 - i. Realice las gráficas de datos y agrupaciones para las variables categóricas en el dataset (20 puntos)
 - d. Análisis de Relaciones de Datos:
 - i. Análisis Univariable y Bivariable (10 puntos)
 2. Documente apropiadamente las líneas de código que genera.
 3. Adjunte el link de GIT al repositorio correspondiente del campus virtual, ***archivos con el formato diferente al solicitado no se revisarán y se calificaran con 0 automáticamente***

Métrica evaluadora

Indicadores a Evaluar	Excelente	Muy Bueno	Bueno	Insuficiente	Obtenido
Cargue el set de datos en su ambiente de trabajo	Cargó los datos correctamente y sin errores (5 puntos)	Cargó los datos con mínimos errores (4 puntos)	Cargó los datos, pero con algunos errores que no afectan significativamente el análisis (3 puntos)	No pudo cargar los datos correctamente (0 puntos)	5
Exploración Inicial de los Datos	Proporciona información detallada sobre todas las columnas y tipos de datos (5 puntos)	Proporciona información sobre la mayoría de las columnas y tipos de datos (4 puntos)	Proporciona información sobre algunas columnas y tipos de datos (3 puntos)	No proporciona información sobre las columnas y tipos de datos (0 puntos)	5
Exploración Inicial de los Datos (Resumen estadístico de las variables numéricas)	Proporciona un resumen detallado y comprensivo de todas las variables numéricas (5 puntos)	Proporciona un resumen detallado de la mayoría de las variables numéricas (4 puntos)	Proporciona un resumen básico de algunas variables numéricas (3 puntos)	No proporciona un resumen estadístico de las variables numéricas (0 puntos)	5
Limpieza y Tratamiento de Datos (Determine valores nulos)	Identifica y maneja eficientemente todos los valores nulos en el conjunto de datos (10 puntos)	Identifica y maneja la mayoría de los valores nulos en el conjunto de datos (8 puntos)	Identifica y maneja algunos valores nulos en el conjunto de datos (6 puntos)	No identifica ni maneja los valores nulos en el conjunto de datos (0 puntos)	10
Limpieza y Tratamiento de Datos (Determine valores atípicos)	Detecta y trata todos los valores atípicos de manera adecuada (10 puntos)	Detecta y trata la mayoría de los valores atípicos de manera adecuada (8 puntos)	Detecta y trata algunos valores atípicos de manera adecuada (6 puntos)	No detecta ni trata los valores atípicos de manera adecuada (0 puntos)	10
Limpieza y Tratamiento de Datos (Determine concentraciones y tendencia de datos)	Realiza un análisis profundo de las concentraciones y tendencias en los datos (10 puntos)	Realiza un análisis sólido de las concentraciones y tendencias en los datos (8 puntos)	Realiza un análisis básico de las concentraciones y tendencias en los datos (6 puntos)	No realiza un análisis de las concentraciones y tendencias en los datos (0 puntos)	10
Limpieza y Tratamiento de Datos (Determine acciones correctivas - imputación)	Implementa estrategias avanzadas y precisas para la imputación de datos faltantes (15 puntos)	Implementa estrategias efectivas para la imputación de datos faltantes (12 puntos)	Implementa estrategias básicas para la imputación de datos faltantes (9 puntos)	No implementa estrategias para la imputación de datos faltantes (0 puntos)	15
Visualización de datos	Realiza las gráficas de datos y agrupaciones para las variables categóricas en el dataset (20 puntos)	Realiza las gráficas de datos y agrupaciones para la mayoría de las variables categóricas en el dataset (16 puntos)	Realiza algunas gráficas de datos y agrupaciones para algunas variables categóricas en el dataset (12 puntos)	No realiza gráficas de datos y agrupaciones para las variables categóricas en el dataset (0 puntos)	20
Análisis de Relaciones de Datos	Realiza un análisis univariable y bivariado de manera adecuada (10 puntos)	Realiza un análisis univariable y bivariado de manera adecuada con algunos errores menores (8 puntos)	Realiza un análisis univariable y bivariado con errores significativos (6 puntos)	No realiza un análisis univariable y bivariado de manera adecuada (0 puntos)	10

----- Ultima Linea -----