**Pauta de Trabajo Práctico - Unidad II**

**Ramo: Fundamentos para Data Science**

**Unidad II: Análisis Exploratorio, Imputación de Datos y Visualización**

**Fecha de entrega: 28 de mayo 11:59pm**

**Enviar: diego.miranda@academico.cftlr.cl**

**Instrucciones:**

* Los análisis exploratorios, imputación de datos faltantes y visualización de datos deben realizarse en un script de Python que contenga el código utilizado para cada tarea. El archivo debe estar debidamente comentado para facilitar su comprensión.
* El informe final debe contener exclusivamente la interpretación de los resultados, conclusiones y gráficos obtenidos, evitando incluir el código en el cuerpo del informe. Si es necesario, el script de Python puede adjuntarse como anexo.

1. **Análisis Exploratorio** 
   * Realizar un análisis exploratorio utilizando al menos 5 codigos. (5 puntos)
   * Calcular las medidas estadísticas (media, mediana, moda, desviación estándar) para variables numéricas. Interpretar cada resultado indicando qué información aporta. (3 puntos)
2. **Imputación de Datos Faltantes** 
   * Identificar las columnas con valores faltantes y calcular el porcentaje de datos perdidos por cada columna. (2 puntos)
   * Proponer un método de imputación para cada variable: media, mediana o moda para numéricas y moda o 'Sin Información' para categóricas. Justificar cada elección. (4 puntos)
   * Realizar la imputación de los datos faltantes utilizando el método seleccionado. (5 puntos)
   * Analizar cómo cambian las distribuciones antes y después de la imputación, interpretando si el método elegido fue el adecuado. (2 puntos)
3. **Visualización de Datos** 
   * Generar gráficos para representar la distribución de cada variable (histogramas para numéricas, barras o tortas para categóricas). Explicar los patrones observados. (4 puntos)
   * Graficar relaciones entre variables relevantes (scatter plot para numéricas, gráficos de barras cruzadas para categóricas) y analizar correlaciones o asociaciones encontradas. (4 puntos)
   * Crear un heatmap de correlación para las variables numéricas y explicar cuáles tienen mayor relación y por qué podría ser relevante. (2 puntos)
4. **Informe Final**

El informe debe estar estructurado de manera clara y ordenada, siguiendo los siguientes apartados:

* + 1. Portada: Título del trabajo, nombre del estudiante, fecha y curso. (0.5 puntos)
    2. Introducción: Breve descripción del objetivo del análisis y la estructura del informe. (1 punto)
    3. Desarrollo: Incluye el análisis exploratorio, imputación de datos faltantes y visualización de datos, con gráficos bien etiquetados y conclusiones parciales. (5 puntos)
    4. Interpretación de Resultados: Análisis crítico de los resultados obtenidos, destacando patrones relevantes, anomalías detectadas y posibles relaciones entre variables. (2 puntos)
    5. Conclusiones: Resumen de los hallazgos más importantes y sugerencias para futuras investigaciones o aplicaciones prácticas. (1.5 puntos)
  + El informe debe tener un enfoque interpretativo, donde cada gráfico y estadística presentada sea analizada en el contexto del problema planteado.
  + Evitar la inclusión de códigos o procedimientos técnicos en el informe, estos deben ir en un anexo si se consideran relevantes.
  + El informe debe estar bien redactado, evitando faltas de ortografía y con un estilo claro y conciso.
  + Elaborar un informe en formato PDF con el análisis completo, gráficos bien etiquetados y conclusiones detalladas.
  + Interpretar los resultados obtenidos de manera clara y proponer posibles acciones a partir del análisis realizado.

**Observaciones:**

* **Utilizar librerías de análisis de datos como Pandas, NumPy, Matplotlib y Seaborn.**
* **La calidad del análisis, la correcta interpretación de los resultados y el uso de visualizaciones adecuadas serán fundamentales para la evaluación.**