Trabajo Final de curso - Python - Ciencia de Datos

Cada alumno deberá escoger uno de los siguientes datasets de Kaggle:

1. [Amazon Prime Movies and TV Shows](https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows)
2. [Billboard Hot 100 2000-2023 Data with Features](https://www.kaggle.com/datasets/suparnabiswas/billboard-hot-1002000-2023-data-with-features)
3. [Data Engineering Salary 2024](https://www.kaggle.com/datasets/zeesolver/data-eng-salary-2024)
4. [Netflix Shows](https://www.kaggle.com/datasets/shivamb/netflix-shows)

Python Hackerrank: <https://www.hackerrank.com/python-dia-1-1> (Opcional)

# Requisitos

## Preprocesamiento de Datos

1. **Tratamiento de Valores Nulos**:
   * Identificación y tratamiento de valores nulos.
   * Justificación de la técnica utilizada para manejar los valores nulos.
2. **Tratamiento de Valores Duplicados**:
   * Identificación y eliminación de duplicados.
   * Justificación de la eliminación de duplicados.
3. **Normalización de Datos**:
   * Normalización de las variables pertinentes.
   * Justificación de la elección de la técnica de normalización.
4. **Variables Redundantes**:
   * Identificación y eliminación de variables redundantes.
   * Justificación de la eliminación de estas variables.
5. **Justificación de Decisiones**:
   * Todas las decisiones de preprocesamiento deben ser debidamente justificadas con explicaciones claras.

## Visualización de Datos

1. **Cantidad de Visualizaciones**:
   * Generar al menos 3 visualizaciones.
2. **Complejidad de Visualizaciones de columnas o consultas**:
   * Las visualizaciones deben tener un grado de complejidad adecuado. No se aceptan solo scatter plots o gráficos de barras simples.
   * Usar las variables visuales para enriquecer las visualizaciones.
3. **Descripción e Interpretación**:
   * Cada visualización debe estar acompañada de una descripción detallada y una interpretación de los insights o hallazgos importantes.
4. **Etiquetas y Leyendas**:
   * Todas las visualizaciones deben tener etiquetas en los ejes y leyendas claras.
5. **Herramientas**:
   * Pueden usarse cualquier librería de Python para la generación de las visualizaciones (Matplotlib, Seaborn, Plotly, etc.).

# Entrega y Evaluación

**Fecha de Entrega**: 7 de agosto de 2024 23:59

**Plataforma de Entrega**: Google Classroom.

**Formato de Entrega**:

* Un notebook en Colab donde se realizó el análisis de datos.
* El notebook debe tener la justificación de las decisiones tomadas.

**Criterios de Evaluación**:

* La nota va de 0 a 20 puntos.
* La claridad y justificación de las decisiones de preprocesamiento.
* La calidad y complejidad de las visualizaciones.
* La claridad y profundidad de las interpretaciones y conclusiones.
* La organización y presentación del notebook

### **Rúbrica de Calificación - Trabajo Final de Python - Ciencia de Datos**

| **Criterio** | **Puntos Máximos** | **Descripción** |
| --- | --- | --- |
| **Preprocesamiento de Datos (10)** |  |  |
| **Tratamiento de Valores Nulos** | 3 | Identificación clara y correcta de valores nulos (1.5 puntos). Justificación adecuada de la técnica utilizada para manejar los valores nulos (1.5 puntos). |
| **Tratamiento de Valores Duplicados** | 3 | Identificación correcta de duplicados (1.5 puntos). Justificación adecuada de la eliminación de duplicados (1.5 puntos). |
| **Normalización de Datos** | 2 | Normalización adecuada de las variables pertinentes (1 punto). Justificación clara de la técnica de normalización elegida (1 punto). |
| **Variables Redundantes** | 2 | Identificación y eliminación adecuada de variables redundantes (1 punto). Justificación clara de la eliminación de estas variables (1 punto). |
| **Visualización de Datos (8)** |  |  |
| **Cantidad de Visualizaciones** | 2 | Generación de al menos 5 visualizaciones relevantes (1 punto). Uso correcto de variables visuales para enriquecer las visualizaciones (1 punto). |
| **Complejidad de Visualizaciones** | 2 | Uso de visualizaciones con un grado de complejidad adecuado. |
| **Descripción e Interpretación** | 2 | Descripción detallada y clara de cada visualización (1 punto). Interpretación profunda y relevante de los insights obtenidos (1 punto). |
| **Etiquetas y Leyendas** | 2 | Todas las visualizaciones contienen etiquetas en los ejes y leyendas claras (1 punto). |
| **Organización y Presentación (2)** |  |  |
| **Formato del Notebook** | 2 | El notebook está bien organizado, con todas las justificaciones y decisiones explicadas claramente en el texto. |
| **Total** | **20 puntos** |  |