# Taller 1: Análisis de Game of Thrones

**Objetivo:** Este taller tiene como objetivo que los estudiantes apliquen técnicas de análisis de datos vistas en clase, como la fusión de datasets, la imputación de valores faltantes, y la creación de visualizaciones. Se proporcionan dos archivos con información sobre la serie *Game of Thrones* y los estudiantes deberán realizar un informe detallado siguiendo las instrucciones.

# Instrucciones:

1. **Descripción del dataset (10 puntos)**

Proporcione un párrafo detallado que describa el conjunto de datos y su estructura (títulos de columnas, tipo de datos, etc.).

1. **Fusión de Archivos (10 puntos)**
   * Fusionen ambos datasets utilizando primary\_key como clave primaria.
   * Justifiquen la selección de la unión (inner join) y verifiquen que el primary\_key también representa el orden de emisión de los capítulos.
2. **Estadísticos Descriptivos (10 puntos)**
   * Para cada columna numérica, calcule: media, mediana, moda, varianza, desviación estándar y rango.
   * Incluya una tabla con estos valores.
3. **Imputación de Datos (20 puntos)**
   * Identifiquen las columnas con valores nulos y realicen la imputación usando la media para columnas numéricas y la moda para otras.
   * Indiquen los criterios para realizar la imputación y comparen cómo cambian la media y la varianza antes y después de imputar.

Pregunta: ¿Por qué no sería adecuado reemplazar los valores nulos de la columna "Written by" con la media?

1. **Encoding (15 puntos)**
   * Realicen **One-Hot Encoding** para la columna "Written by".
   * Realicen **Label Encoding** para la columna "Directed by".
   * Justifiquen por qué **One-Hot Encoding** no sería ideal para la columna "Directed by".
2. **Imputación de Fechas (10 puntos)**
   * Traten los valores faltantes en la columna "Original air date" de forma que se mantenga la coherencia temporal de los episodios.
   * Expliquen cómo manejarían esta situación.
3. **Tratamiento de Outliers (20 puntos)**
   * Hagan un gráfico de boxplot para la columna "Imdb rating" e identifiquen los outliers.
   * Elijan una técnica para manejar los outliers (eliminación, reemplazo, etc.) y vuelvan a graficar el boxplot.
   * Expliquen si la media y la varianza cambian después del tratamiento de outliers.
4. **Visualización de Datos (30 puntos)**
   * Realicen un gráfico de torta para la columna "Written by" y explíquenlo.
   * Realicen un gráfico de línea para el promedio de "Imdb rating" por temporada y explíquenlo.
   * Realicen un gráfico de línea para "U.S. viewers (millions)" por temporada y explíquenlo.
5. **Punto Extra (10 puntos)**
   * Agreguen una visualización adicional que consideren relevante, justifiquen su elección y expliquen su interpretación.

# Consideraciones Generales

* Incluyan justificación y explicaciones en cada gráfico.
* Entreguen los gráficos en formato PDF o Word, junto con el código fuente (Python o Excel).
* Los trabajos deben ser en grupos de hasta 3 personas.

### Instrucciones adicionales:

**Nota Importante:** Además de realizar las actividades solicitadas en este taller, **el informe final debe seguir el formato entregable que se adjunta** (véase documento "Formato de Entregable"). Asegúrense de cumplir con todas las secciones especificadas, incluyendo la portada, el resumen ejecutivo, la metodología, y las demás partes descritas en dicho formato.