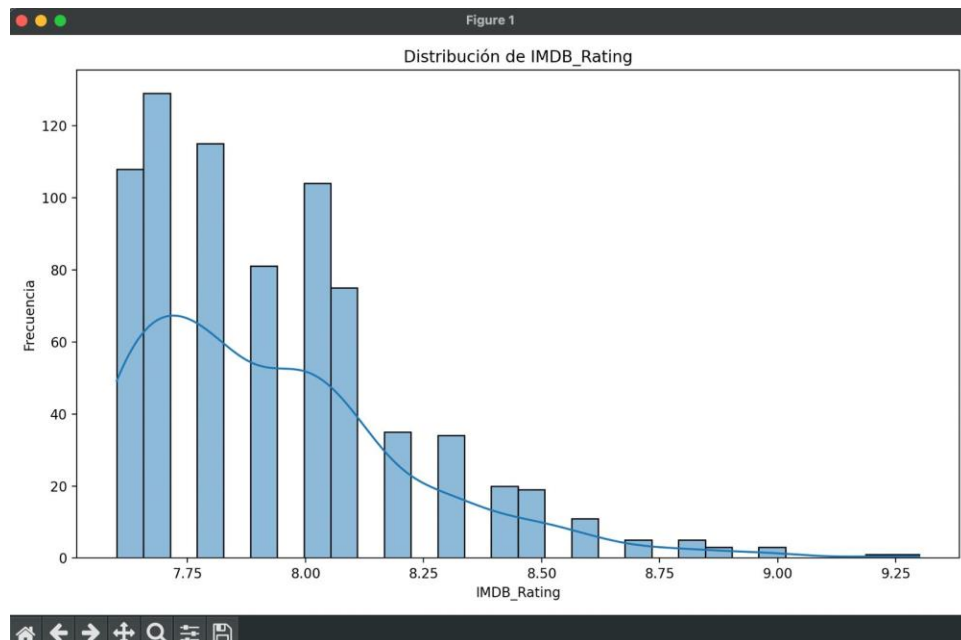


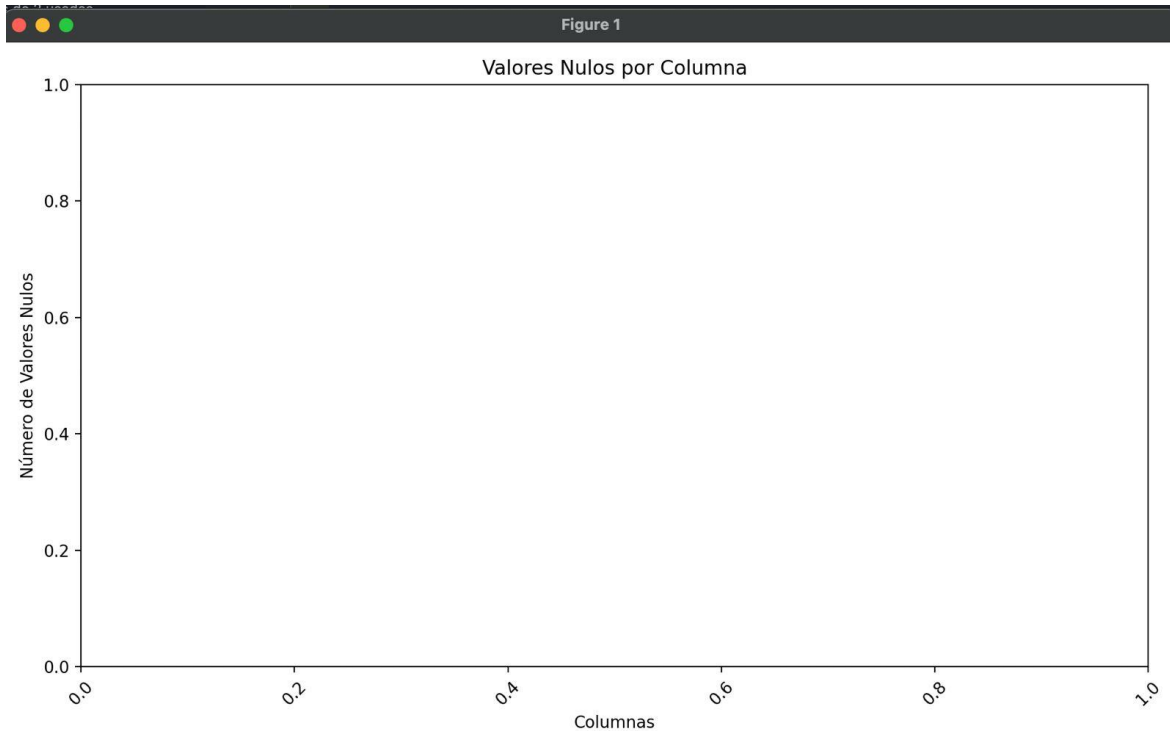
Durante el proceso de preparación del conjunto de datos para su uso en un modelo de machine learning, se realizó una limpieza que consistió en la eliminación de columnas irrelevantes, inconsistentes o difíciles de procesar. Se conservaron únicamente aquellas variables numéricas que aportan información clara, estructurada y directamente útil para tareas predictivas, como **Runtime**, **Released\_Year**, **IMDB\_Rating**, **Meta\_score**, **No\_of\_Votes** y **Gross**. Estas columnas fueron seleccionadas porque presentan un alto potencial explicativo respecto al desempeño o popularidad de una película, además de ser fácilmente manipulables por algoritmos de aprendizaje automático sin requerir transformaciones complejas. Al eliminar campos como nombres de actores, géneros, descripciones textuales o enlaces web, se evitó introducir ruido, ambigüedad o la necesidad de técnicas avanzadas como procesamiento de lenguaje natural o codificación de alta cardinalidad. A continuación, se explicará y mostrará cada gráfica.

- **Distribución de IMDB\_Rating**

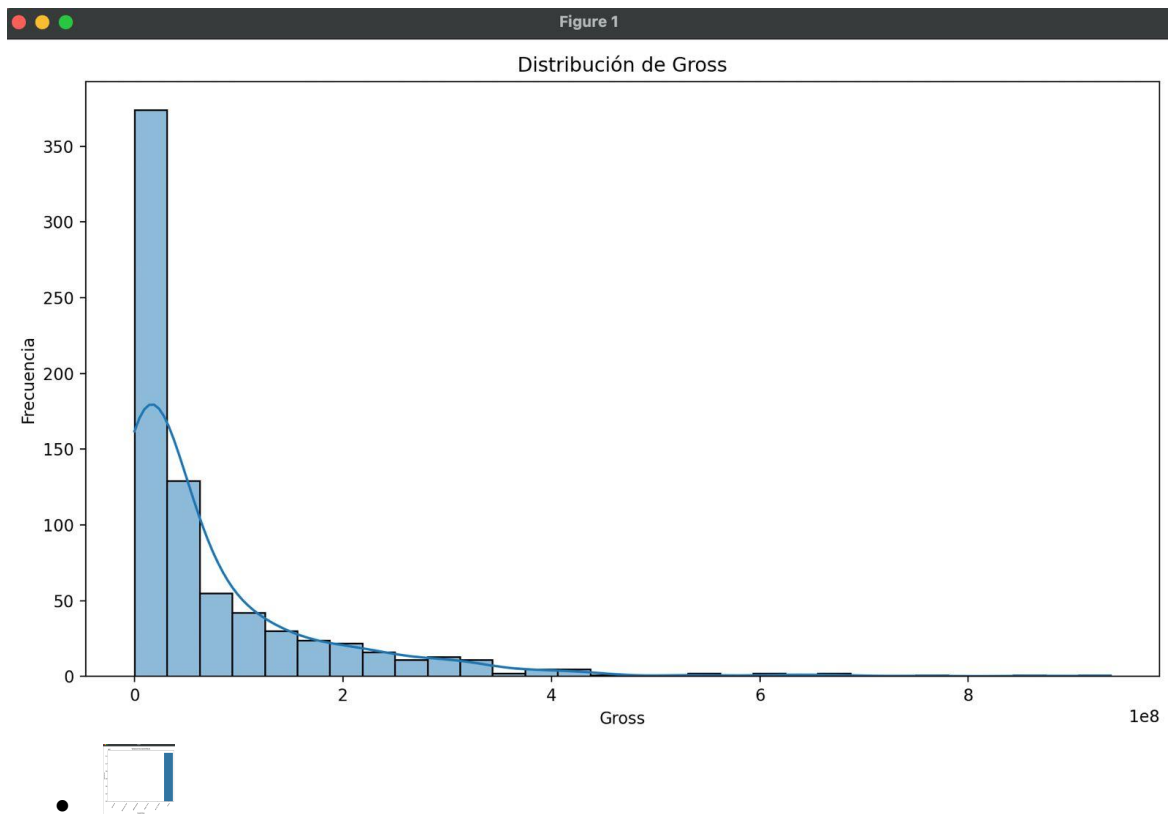


El gráfico muestra la distribución de las calificaciones promedio (IMDB\_Rating) de las películas y series del conjunto de datos. Se observa que la mayoría de los títulos se concentran en un rango de puntuaciones entre 7.6 y 8.1, lo que indica una tendencia general hacia valoraciones positivas. La forma de la distribución es asimétrica hacia la derecha, lo que significa que hay menos títulos con puntuaciones más altas. A medida que aumenta la calificación, la frecuencia disminuye notablemente, siendo escasos los títulos que alcanzan valores por encima de 8.5. Esto sugiere que si bien la mayoría del contenido es bien valorado por los usuarios, las producciones con calificaciones excepcionalmente altas son poco comunes dentro del catálogo analizado.

- Valores Nulos por Columna

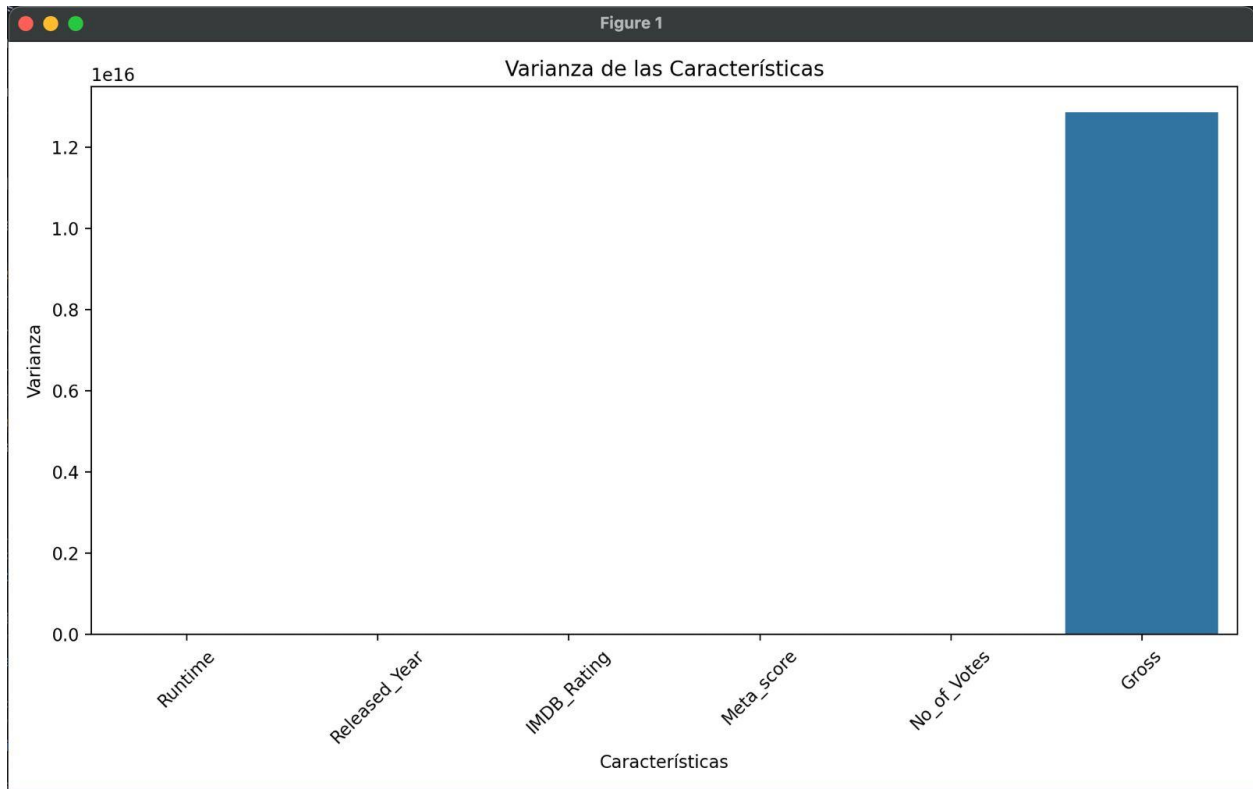


El gráfico muestra el análisis de valores nulos por columna en el conjunto de datos. Como se puede observar, no hay barras visibles en el gráfico, lo que indica que **no existen valores nulos en ninguna de las columnas**. Esto es una buena señal para el preprocesamiento, ya que no se requiere imputación de datos faltantes ni eliminación de filas o columnas por esta causa. La ausencia de nulos permite trabajar directamente con el conjunto de datos sin necesidad de aplicar técnicas de limpieza específicas relacionadas con valores perdidos.

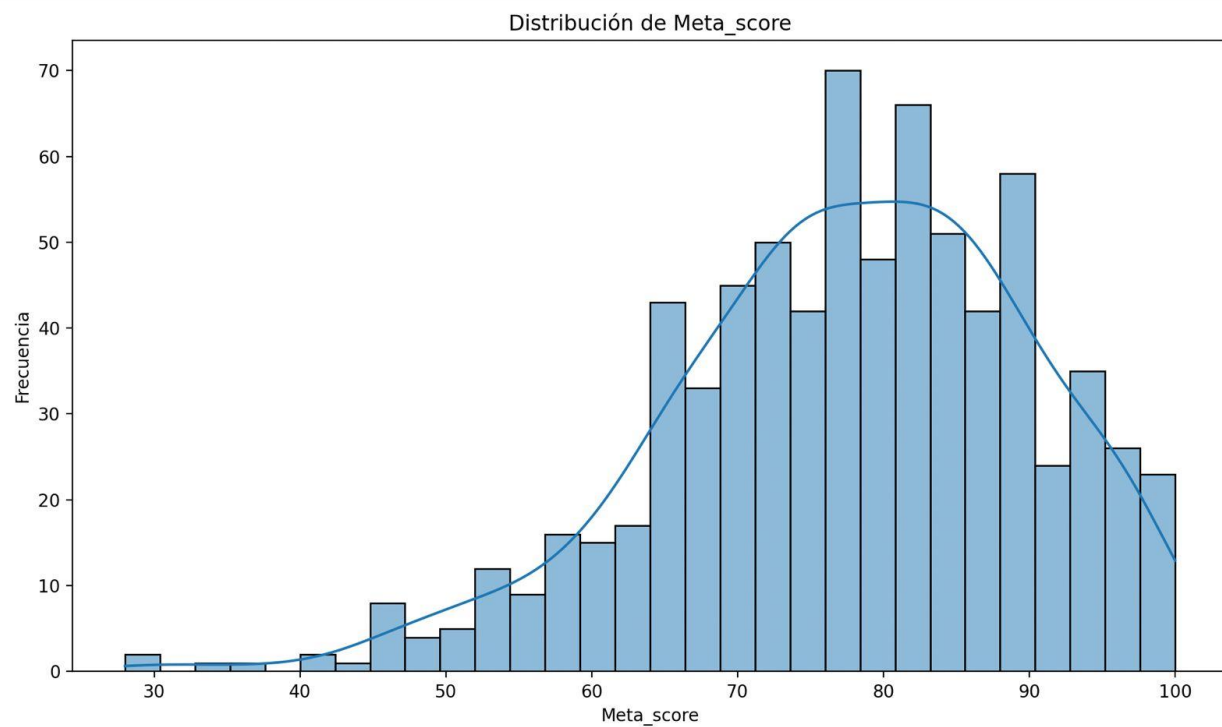


El gráfico muestra la distribución de la variable **Gross**, la cual representa los ingresos brutos generados por las películas. La distribución está claramente sesgada a la derecha, lo que indica que la mayoría de las películas obtienen ingresos relativamente bajos, mientras que un pequeño número de películas alcanza ingresos extremadamente altos. La gran concentración de frecuencias en los valores más bajos sugiere que es común que las películas generen ingresos modestos, mientras que las que superan los 200 millones son menos frecuentes, pero elevan significativamente la media de la variable.

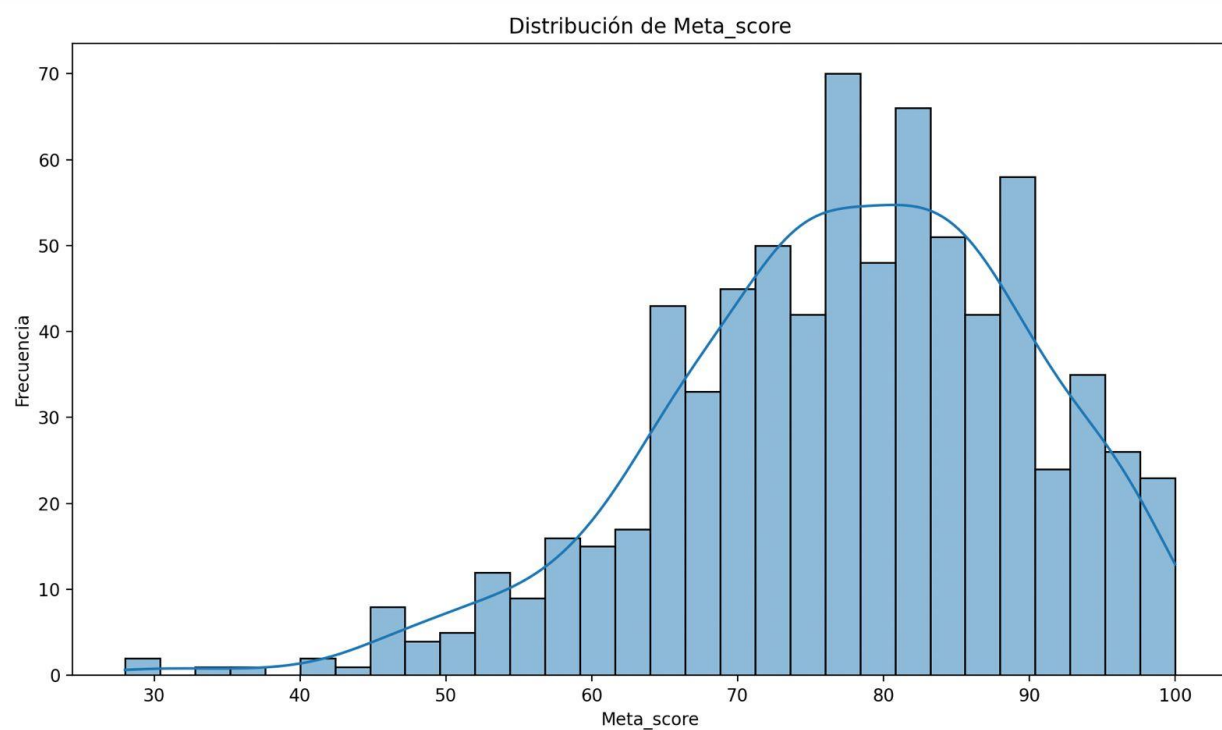
- Varianza de las Características



El gráfico muestra la varianza de diversas características numéricas presentes en el conjunto de datos. Se observa claramente que la variable **Gross** presenta una varianza extremadamente alta en comparación con las demás variables como **Runtime**, **Released\_Year**, **IMDB\_Rating**, **Meta\_score** y **No\_of\_votes**, cuyas varianzas son tan pequeñas en comparación que prácticamente no se visualizan en la gráfica. Esta alta varianza de **Gross** puede atribuirse a su escala y a la presencia de valores atípicos, como se observó anteriormente en su distribución.



- Distribución de Meta\_score



El histograma muestra la distribución de la variable Meta\_score. La forma de la distribución es

asimétrica negativa, aunque se aproxima a una curva normal, con la mayoría de los valores concentrados entre 70 y 90 puntos. Se observa un pico de frecuencia cerca del rango de 75 a 80, lo que indica que muchas películas tienen puntuaciones altas en esta métrica.

Además, hay pocos valores en los extremos bajos, por debajo de 50, lo que sugiere que la mayoría de las películas evaluadas reciben buenas críticas. En comparación con otras variables del conjunto, `Meta_score` presenta una distribución más simétrica y controlada, lo que la hace más adecuada para ser utilizada directamente en análisis estadísticos o modelos de machine learning sin necesidad de transformaciones previas.