Seminario de Big Data

Dataset: Top Películas más taquilleras de todos los tiempos.

1. Elegimos el dataset “Top 1000 películas más taquilleras de todos los tiempos” desde la web Kaggle.
2. La representación elegida es en su mayoría de gráfico de barras. Ya que, considerando los datos dentro del dataset, es posible representar valores para comparar costos, taquilla, año de lanzamiento, etc. Puntos a fines del tema del dataset. Es posible representarlos de manera escalada, ascendente o descendentemente, mostrar evolución de presupuesto de las películas taquilleras, entre otros.
3. Los pasos realizados en el proceso de preparación y selección de datos, fueron en su mayor parte, de limpieza de datos nulos.  
   El dataset elegido tenía los valores de sus filas y columnas bastante limpios, pero presentaban algunos que tenían caracteres no coincidentes a los esperados. Por ejemplo, se encontró un campo de valor “\*\*\*\*\*” dentro de la columna “Gross” (presupuesto), la cual era string y para convertirlo a numérico, los “\*” daban problemas.
4. Una vez preparados los datos, elegimos las librerías de Pandas y Plotly para la realización de dataframes y graficar los datos de forma clara y precisa, para el dashboard.
5. Realizados dentro del dashboard.
6. Realizados dentro del dashboard.
7. Tanto la alfabetización visual como la metáfora visual son abordados en los gráficos. La comprensión, interpretación y comunicación de elementos complejos través de elementos visuales. Entre los procesos destacamos:
   1. Selección de gráficos ideales
   2. Uso de colores.
   3. Títulos y etiquetas claras.
   4. Simplificación de datos complejos.