IMDB Movie Analysis – Final Report

1. **Introduction**

* The IMDB 5000 Movie Dataset contains details of over 5,000 movies, including attributes such as title, director, genre, budget, gross revenue, duration, IMDb rating, and number of user votes.
* The objective of this project is to perform a complete data analytics workflow by cleaning the dataset, analyzing it through multiple tools, and deriving meaningful insights.
* This project demonstrates the use of Python, Excel, SQL, and Power BI to explore trends, identify top-performing movies/directors, and visualize patterns in the film industry.

**2.Methodology**

***🔹 Python (Data Cleaning & EDA)***

* + Imported dataset using Pandas.
  + Cleaned missing values and duplicates.
  + Performed Univariate Analysis (IMDb scores, duration, budgets).
  + Performed Bivariate Analysis (Budget vs Gross, Genre vs IMDb score).
  + Created plots: histograms, scatter plots, boxplots, heatmaps.

***🔹 Excel (Pivot & Dashboard)***

* Built PivotTables to analyze:
* Top 10 highest-grossing movies.
* Average IMDb rating per genre.
* Movie production trend per year.
* Applied Conditional Formatting:
* Highlighted Top 5 Directors (by movie count).
* Highlighted movies with IMDb score > 8.
* Designed an interactive dashboard with slicers for genre, director, and year.

***🔹 SQL (Structured Queries)***

* + Imported dataset into MySQL.
  + Executed queries:
* Find top 5 highest-rated movies.
* Get average IMDb score per director.
* List movies with gross > 100M.
* Count movies per genre.
* Find director with most movies.

***🔹 Power BI (Interactive Dashboard)***

* + Imported dataset into Power BI.
  + Built dashboard with visuals:
    - Total gross revenue and average IMDb score (KPIs).
    - Genre-wise movie distribution.
    - Yearly production trend.
    - Scatter plot (Budget vs Gross).
    - Added slicers for filtering by genre, director, year.

**3.Key Findings & Insights Correlation:**

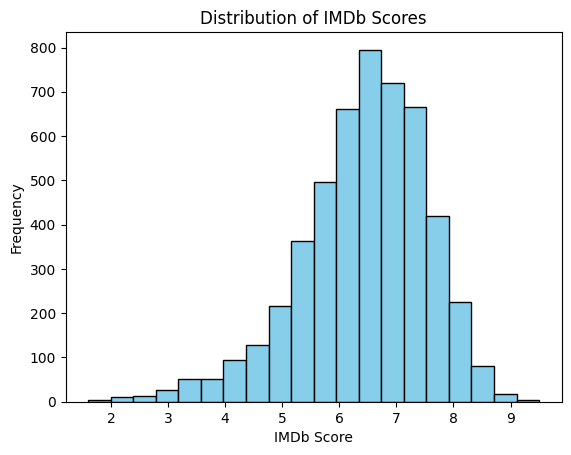
* Higher budgets generally lead to higher gross revenue.
* Genre Popularity: Drama, Comedy, and Action dominate the dataset.
* Directors: Steven Spielberg directed the most movies.
* IMDb Ratings: Most movies fall between 5.5 – 7.5.
* Trends: After 2000, movies show larger budgets and revenues.

**4.GitHub Repository :**

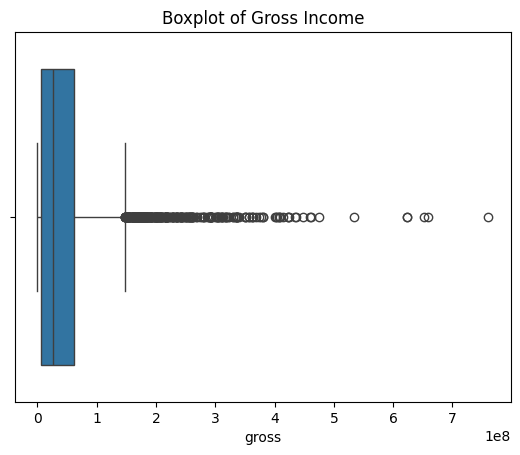
**5. Visuals :**

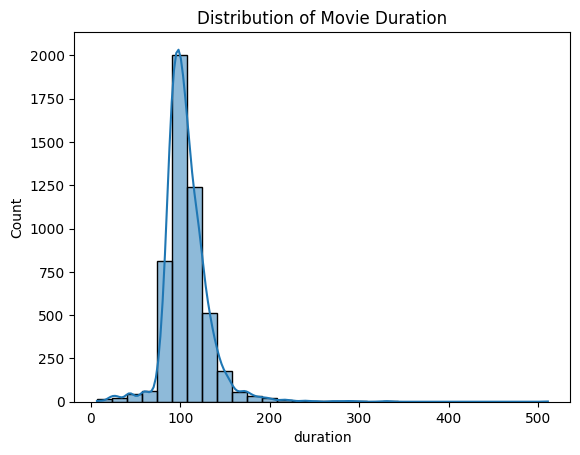
Charts &Plots:

Univariate analysis:

****

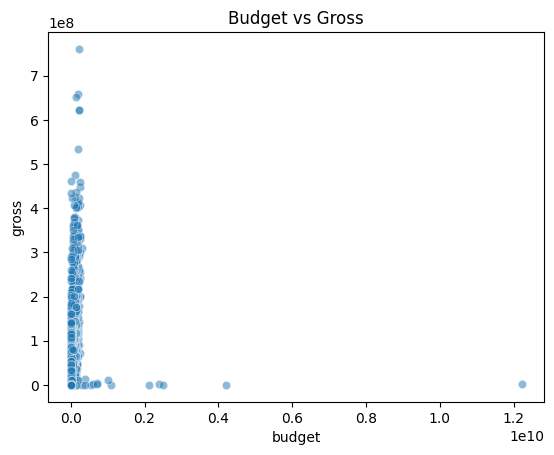
IMDb score distribution

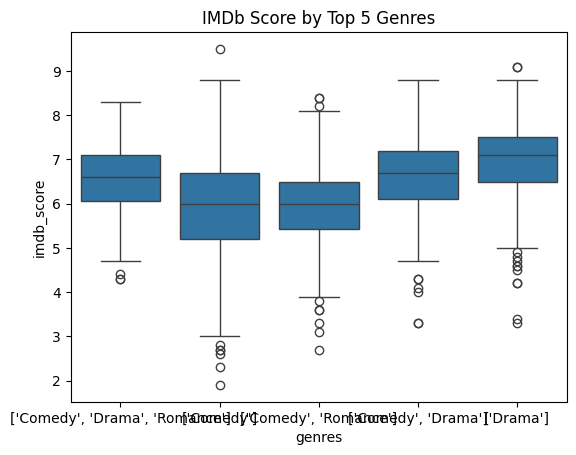
****

**** Boxplot of Gross

Duration distribution

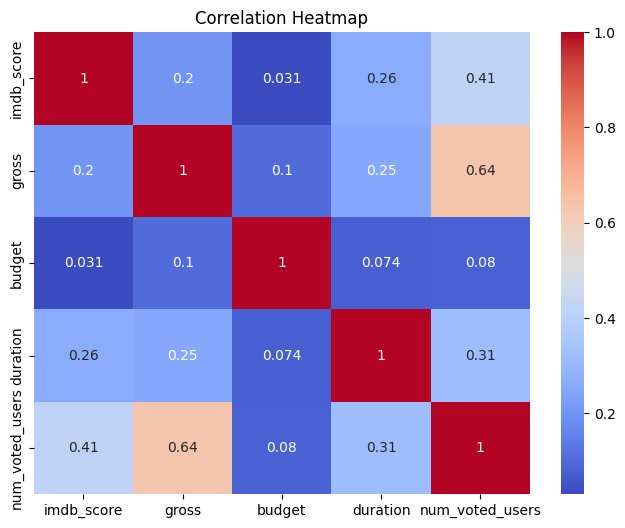
Bivariate analysis:

****

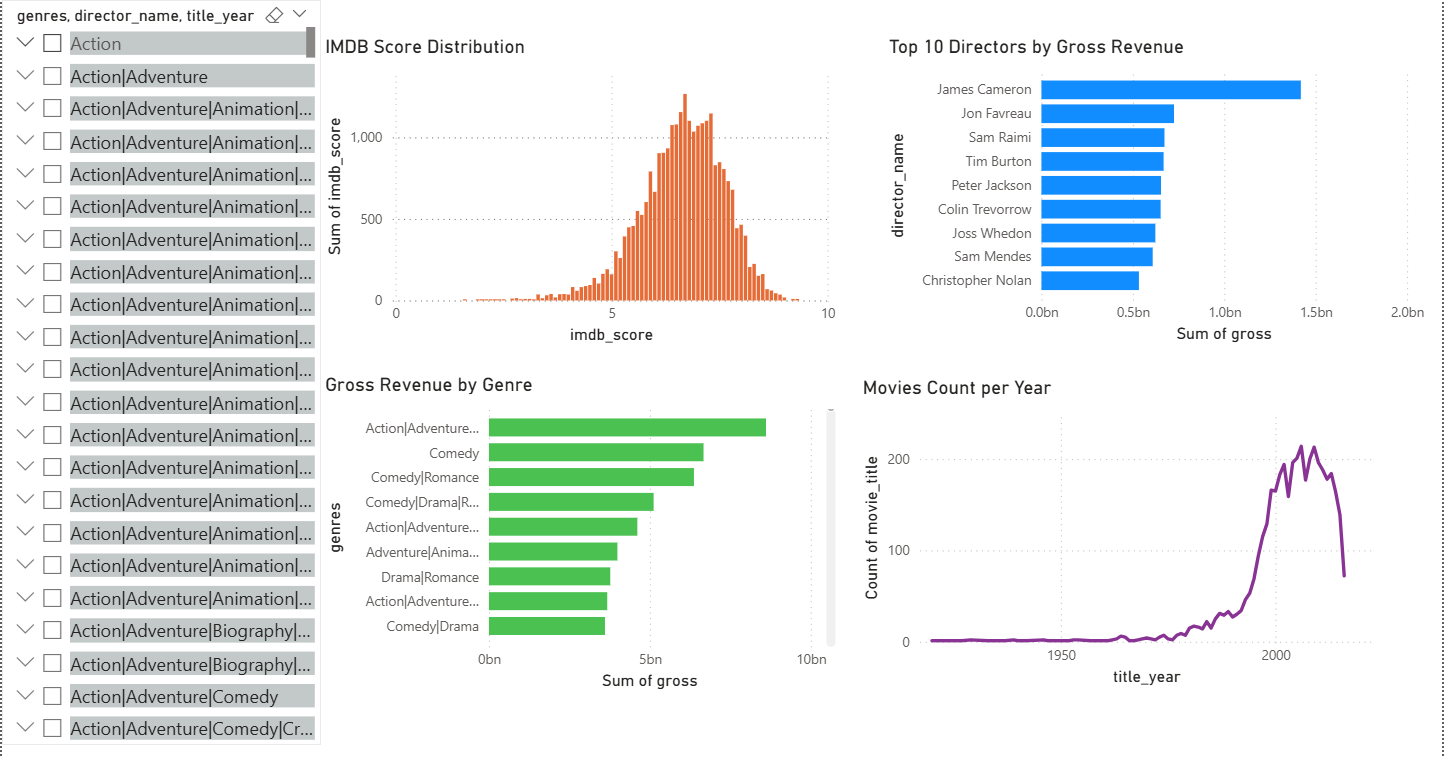
**** Budget vs Gross

Genre vs IMDb Score

Visualizations using Matplotlib & Seaborn:

****

Heatmap

****Dashboard:

**5. Conclusion**

This project highlights how multiple tools complement each other in analytics:

* Python – for data cleaning and exploratory analysis.
* Excel – for pivot analysis and dashboards.
* SQL – for structured querying.
* Power BI – for advanced interactive visualizations.

The analysis provided meaningful insights into movie trends, revenue patterns, genre performance, and director contributions.