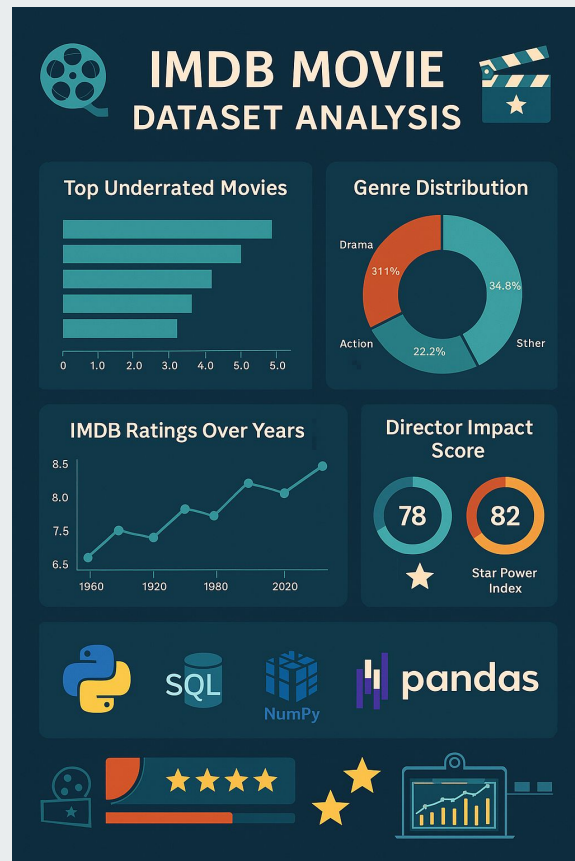# IMDB Movie Dataset Analysis

Project By
Gary Felix A

ABSTRACT

The IMDb Movie Dataset Analysis project aims to explore and extract valuable insights from a curated dataset of top 1000 movies listed on IMDb. Using Python, Pandas, NumPy, File Handling and MySQL, the project involves data cleaning, transformation, and visualization to identify patterns and trends in the movie industry

The dataset is preprocessed to remove duplicates, handle missing values, and standardize formats. Key analytical tasks include identifying top directors, genre popularity, top underrated movies, decade-wise votings and best film. Visualization tools such as Plotly are used to present findings in an interactive manner.

The outcome of this project benefits the film producers, critics, data enthusiast, and recommendation systems.

Project Description

The IMDb Movie Dataset Analysis is a data analytics project designed to explore and gain insights from a dataset containing detailed information about the top 1000 movies on IMDb. In this project, Data is preprocessed for analysis and visualization using tools such as Python (Pandas, NumPy, Plotly) and MySQL for storing cleaned data.

Key Features:

- **Data Cleaning:** Removal of duplicates, handling of missing values, format standardization.
- **Analysis:** Top 10 directors, Top 10 Genres, Decade wise best films, Decade wise total number of votes, Ratings distribution over years, Top underrated movies, Gross vs Imdb ratings.

Project Description

Key Features:

- **Custom Visualization:** Plotly is used to present insights (e.g., bar charts, line graphs, histogram, treemap.
- **Database Integration:** Cleaned data is stored in MySQL database and queried through Python.

Tools & Technologies:

- **Programming Language:** Python
- **Libraries:** Pandas, NumPy, Plotly
- **Database:** MySQL
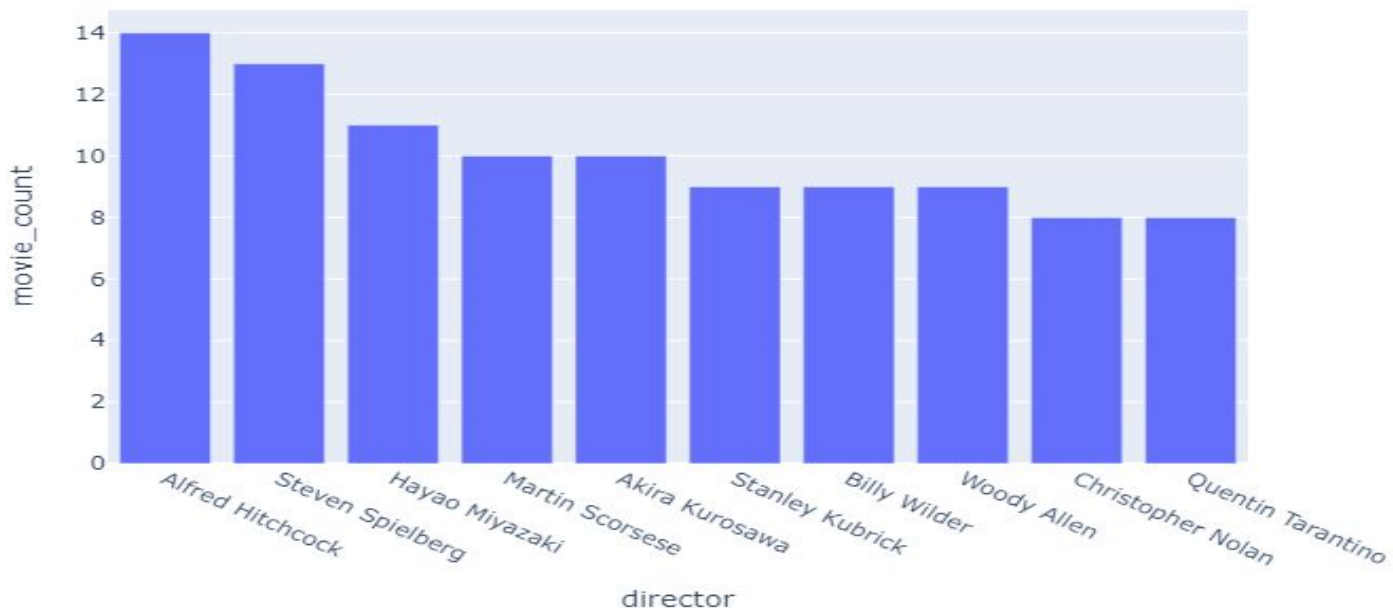- **Data Handling:** CSV
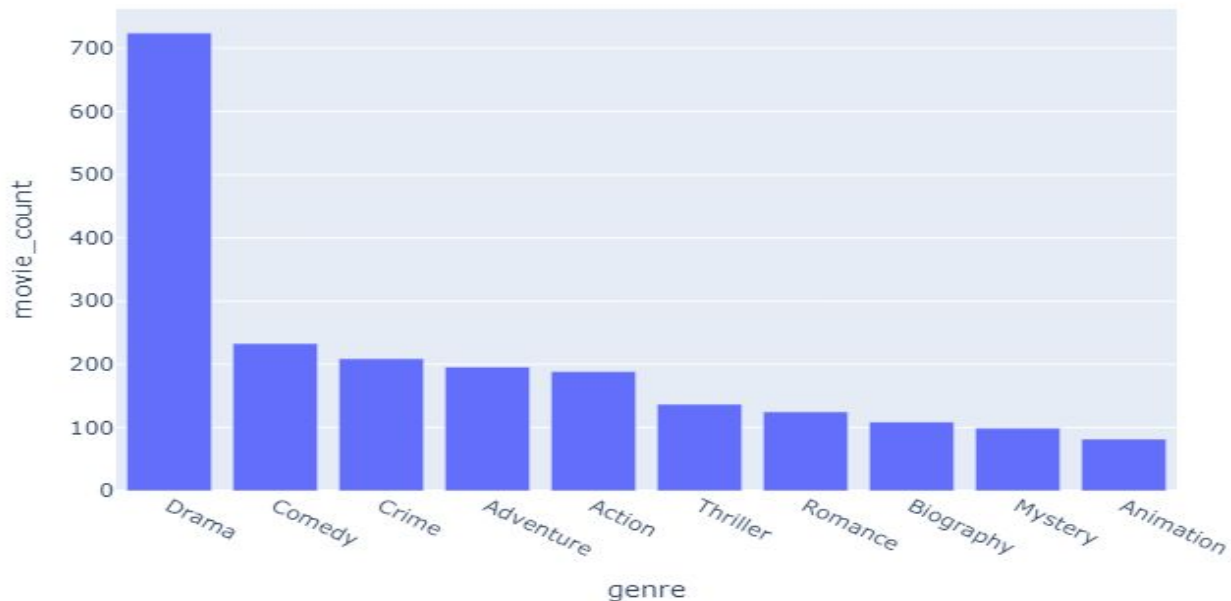- **Environment:** Jupyter Notebook

Tech Stack

Insights & Results

Top 10 Directors with Most Movies With High Ratings

Insights & Results
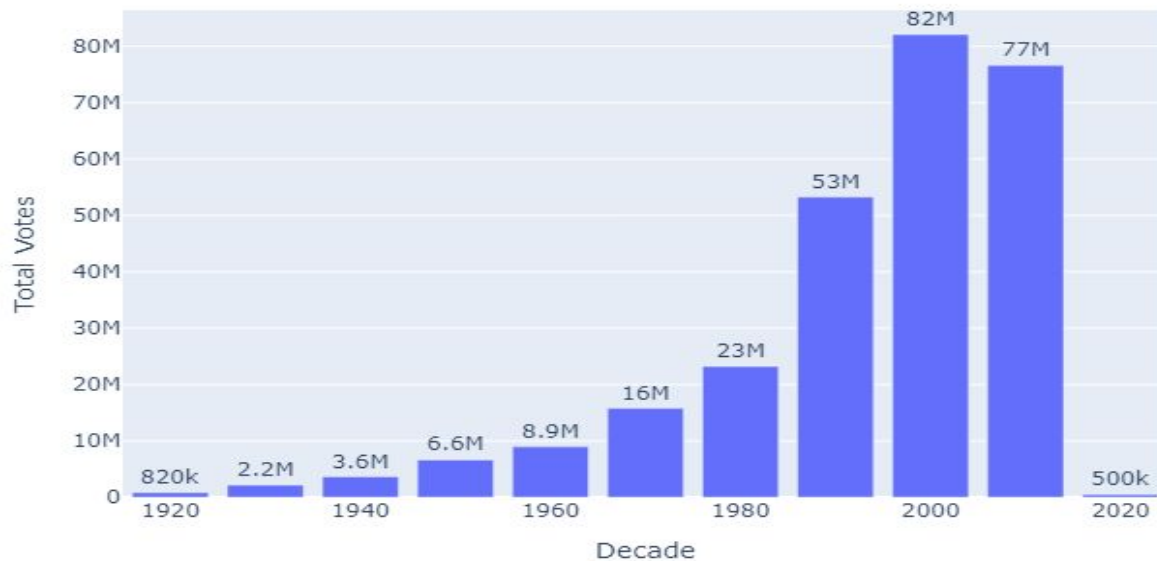
Top 10 Genres with Most Movies

Insights & Results

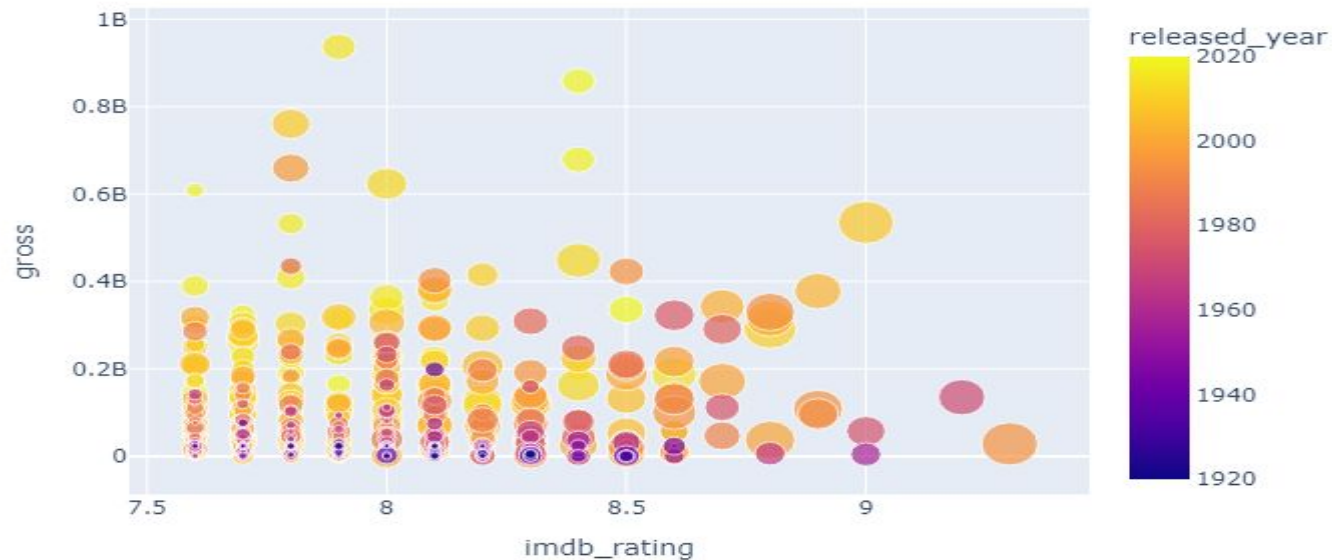Best Movie of Each Decade (Treemap)

Insights & Results

Decade-wise Total Number of Votes

Insights & Results

Thank You

References

Raw Dataset - IMDB Movie Dataset - Top 1000 movies