Case Study:

Analysing Movie Dataset for Revenue and Popularity Insights

Name: Niketa Sengar

Objective

The goal of this study is to explore and derive insight from the dataset of movies and directors spanning a 40- year period (1976-2016). By analysing the relationships between budget, revenue, popularity and other attributes, this study aims to identify key factors contributing to a movie's financial success and audience engagement.

Data set Overview

The merged dataset (directors.csv & movies.csv) has 1465 rows × 13 columns including :

Features	Description
Budget	Production budget of the movie.
Revenue	Revenue generated by the movie
Popularity	A metric indicating audience interest.
Vote Average	Average rating given by users.
Vote Count:	Number of votes received.
Year, Month, Day	Release date details.
director_name	Directors name
id	Unique ID corresponding to each director
gender	Directors gender

Link to dataset (click here)

Link to kaggle notebook

The movie dataset shows no missing values and spans a wide range of budgets, revenues, and popularity scores.

The director data set had 26% of missing entries in gender column, which was imputed using .fillna().mode()

```
In [5]: df.isnull().mean()

Out[5]:
Unnamed: 0      0.0
id              0.0
budget          0.0
popularity      0.0
revenue         0.0
title           0.0
vote_average    0.0
vote_count      0.0
director_id     0.0
year            0.0
month           0.0
day             0.0
dtype: float64
```

```
df_dir['gender']=df_dir['gender'].fillna(df_dir['gender'].mode()[0])
```

Exploratory Data Analysis

Descriptive Statistics

Budget: Ranges from $0 to $380 million, highlighting varying production scales.

Revenue: Peaks at $2.79 billion, with some movies earning no measurable revenue.

Popularity: Scores range from 0 to 724, indicating significant disparity in audience engagement.

Vote Average: Average rating is 6.4, with a minimum of 3.0 and a maximum of 8.3.

Key Observations

Movies with higher budgets tend to generate higher revenues.

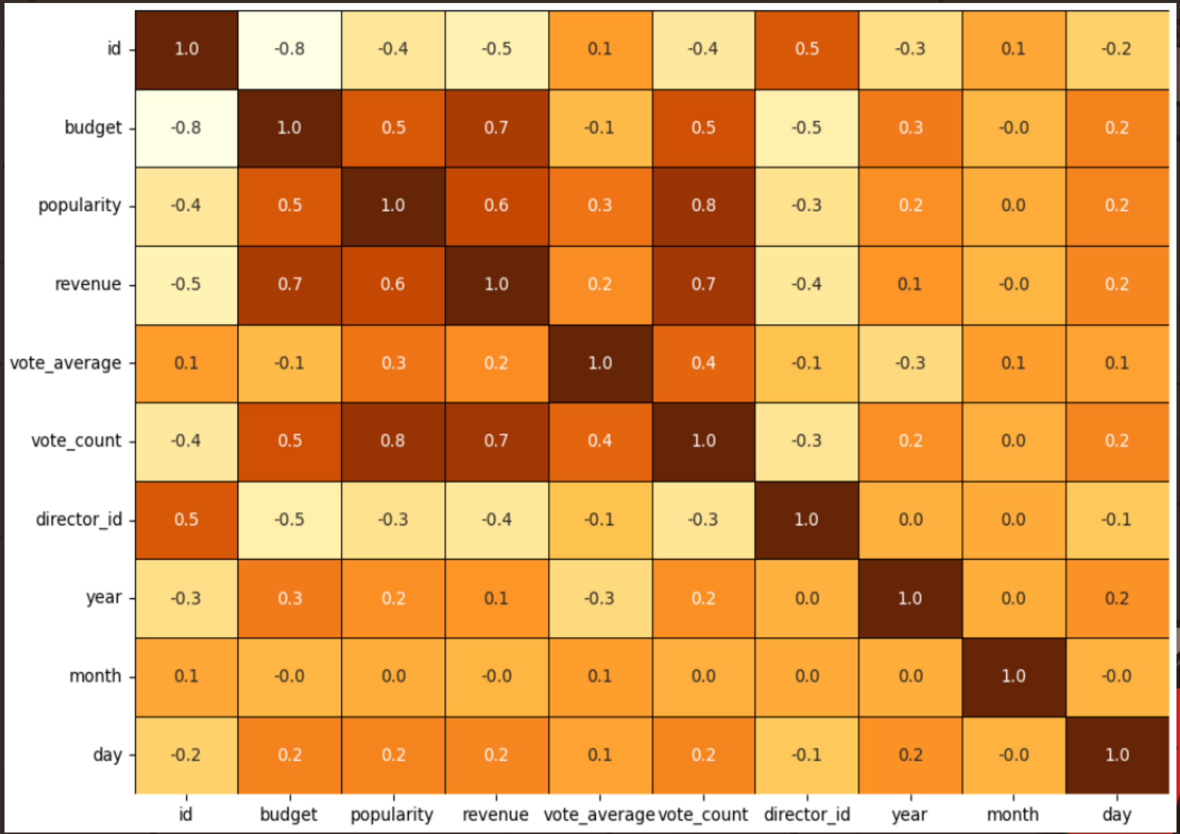
Popularity is strongly correlated with vote count, suggesting that well-liked movies attract more audience feedback

Correlation Analysis

Revenue and Popularity
Positive correlation (0.6), showing that popular movies tend to generate higher revenues.

Popularity and Vote Count
Very strong correlation (0.8), suggesting that popular movies receive more audience engagement.

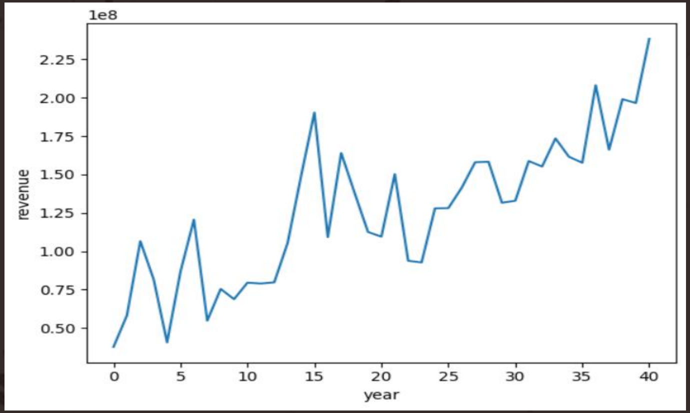
Budget and Revenue
Strong positive correlation (0.7), indicating that higher investments often lead to higher earnings



Trends over Time

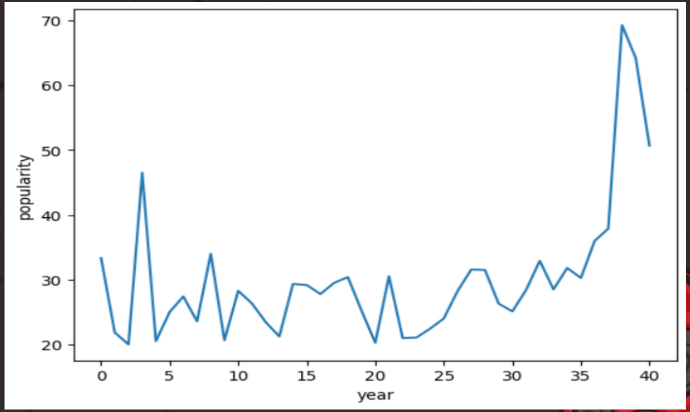
Revenue over Time

Revenue exhibits an upward trend over the years, with notable peaks likely driven by blockbuster releases.
Periodic fluctuations indicate the volatile nature of the movie industry.



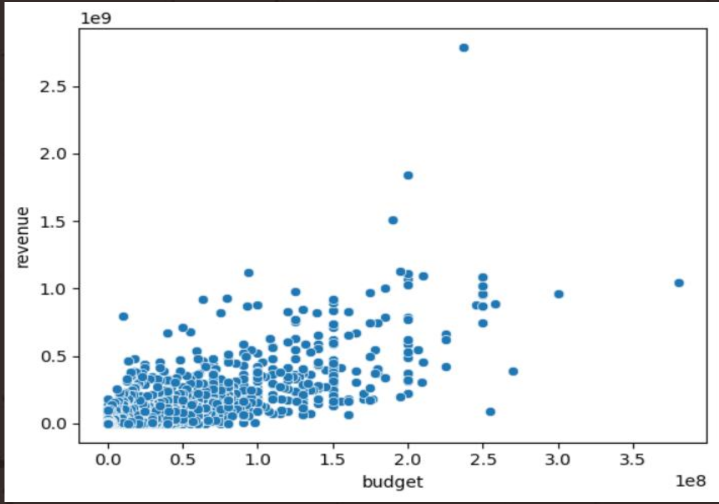
Popularity Over Time

Popularity fluctuated in earlier years, possibly reflecting changing audience preferences.
A sharp rise in recent years (post-2000) suggests increased global movie consumption, potentially driven by advancements in distribution channels and marketing strategies.



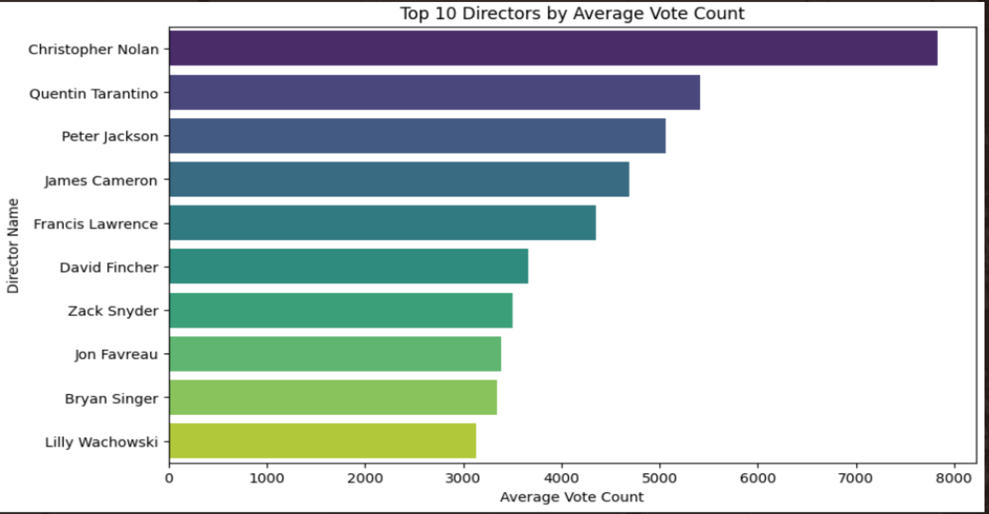
Budget Vs Revenue

Movies with larger budgets often yield higher revenues, though some exceptions exist.
A few low-budget movies performed exceptionally well, indicating the role of content quality and audience appeal.



Top 10 Directors

- **Christopher Nolan Leads**:
Christopher Nolan stands out significantly with the highest average vote count (close to 8000 votes). This indicates his movies consistently attract a large and engaged audience.
- **Close Competitors**:
Quentin Tarantino and Peter Jackson follow in second and third places, respectively, with substantial but noticeably lower average vote counts compared to Nolan.



Gender-Wise Success

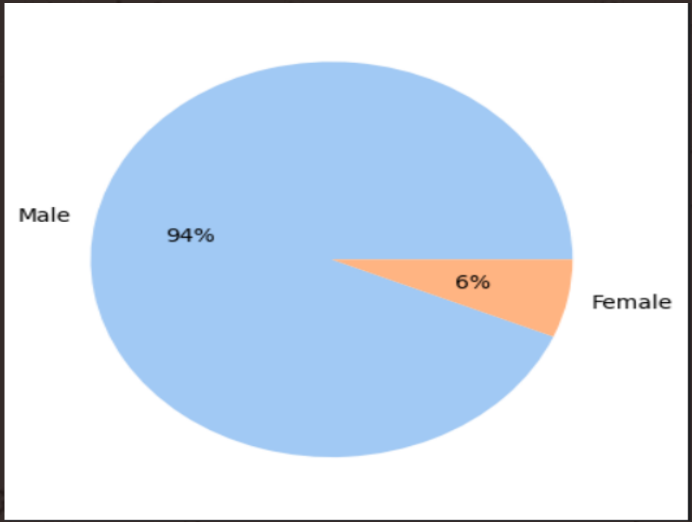
There are small differences in vote averages, revenue, and popularity between films directed by men and women, with male-directed films performing slightly better on all metrics.

The differences are not vast, suggesting that, when provided similar resources and opportunities, both male and female directors deliver comparable results in terms of ratings, revenue, and popularity.

	vote_average	revenue	popularity
gender			
Female	6.262500	1.384982e+08	29.000000
Male	6.370551	1.433601e+08	30.897418

Gender Distribution

The data set has more male directors than Female directors suggesting less females working in this field.



Conclusion

This study highlights the critical factors driving movie success. By understanding the interplay between budget, popularity, and revenue, filmmakers and producers can better strategize their investments and marketing efforts. Future analyses could focus on deeper insights, such as the impact of genre, director reputation, and marketing campaigns on movie success.