

Decoding Film Success:

A Comprehensive Analysis of Directors, Genres, Budgets, and IMDb Scores for Strategic Decision-Making in Film Production.

PRESENTED

BY

MAYUKH DAS

CANDIDATE NUMBER:

277173

Department :

**MS Financial Data
Analytics**

About my report

This report aims to provide a comprehensive analysis of the factors influencing the success of films, with a focus on directors, genres, budgets, and IMDb scores, through a combination of statistical analyses, visualizations, and interpretative summaries.

This dataset has already been provided by the authorities to analyze the IMDb data provided in the movie metadata.csv file.

Exploratory data analyses

In this dataset, we study a total of 28 variables namely 'director_name', 'genre', 'budget', 'imdb_score' etc. Our focus shall remain on the effect of the rest 27 variables on 'gross' variable.

Taking a glimpse at the data:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross
0	Color	James Cameron	723.0	178.0	0.0	855.0	Joel David Moore	1000.0	760505847.0
1	Color	Gore Verbinski	302.0	169.0	563.0	1000.0	Orlando Bloom	40000.0	309404152.0
2	Color	Sam Mendes	602.0	148.0	0.0	161.0	Rory Kinnear	11000.0	200074175.0
3	Color	Christopher Nolan	813.0	164.0	22000.0	23000.0	Christian Bale	27000.0	448130642.0
5	Color	Andrew Stanton	462.0	132.0	475.0	530.0	Samantha Morton	640.0	73058679.0

The minimum and maximum values of our gross variables are 7.6×10^8 and 162, which indicates the presence of huge variations and possible outliers.

```
1 df2[df2.gross == df2.gross.max()]["gross"]
```

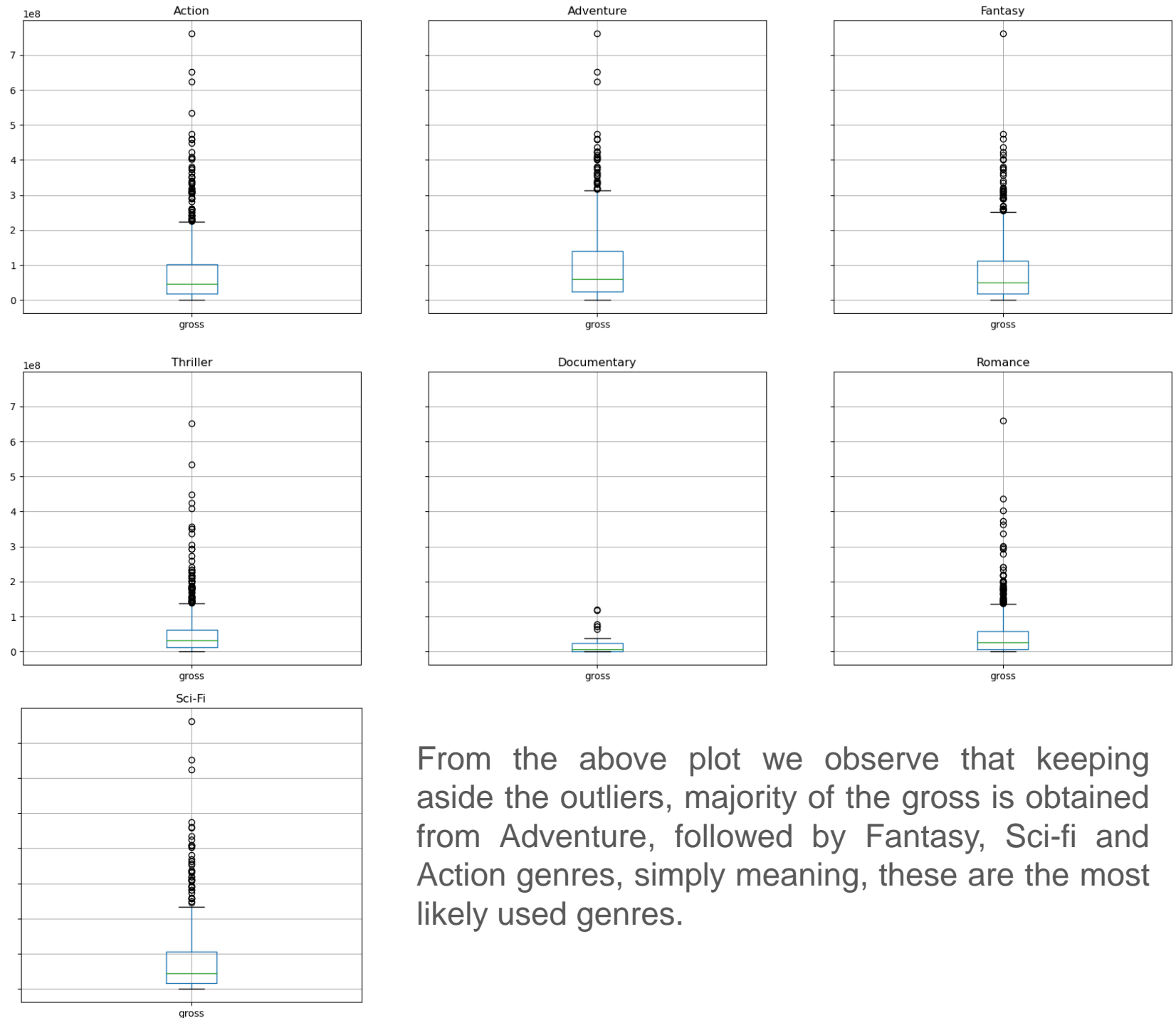
```
0    760505847.0
Name: gross, dtype: float64
```

```
1 df2[df2.gross == df2.gross.min()]["gross"]
```

```
3330    162.0
Name: gross, dtype: float64
```

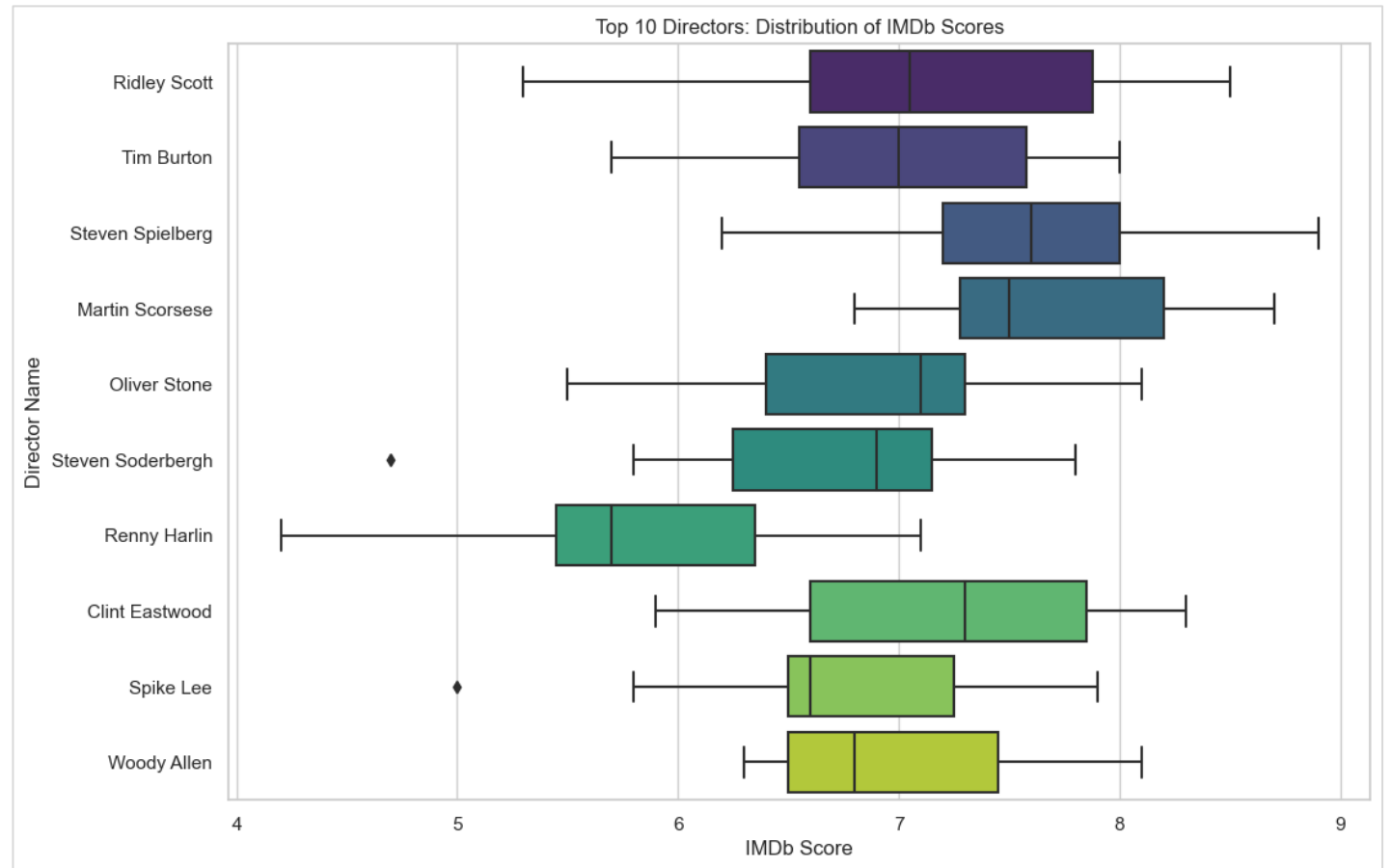
Data Visualisations:

Dependence of gross on genres



From the above plot we observe that keeping aside the outliers, majority of the gross is obtained from Adventure, followed by Fantasy, Sci-fi and Action genres, simply meaning, these are the most likely used genres.

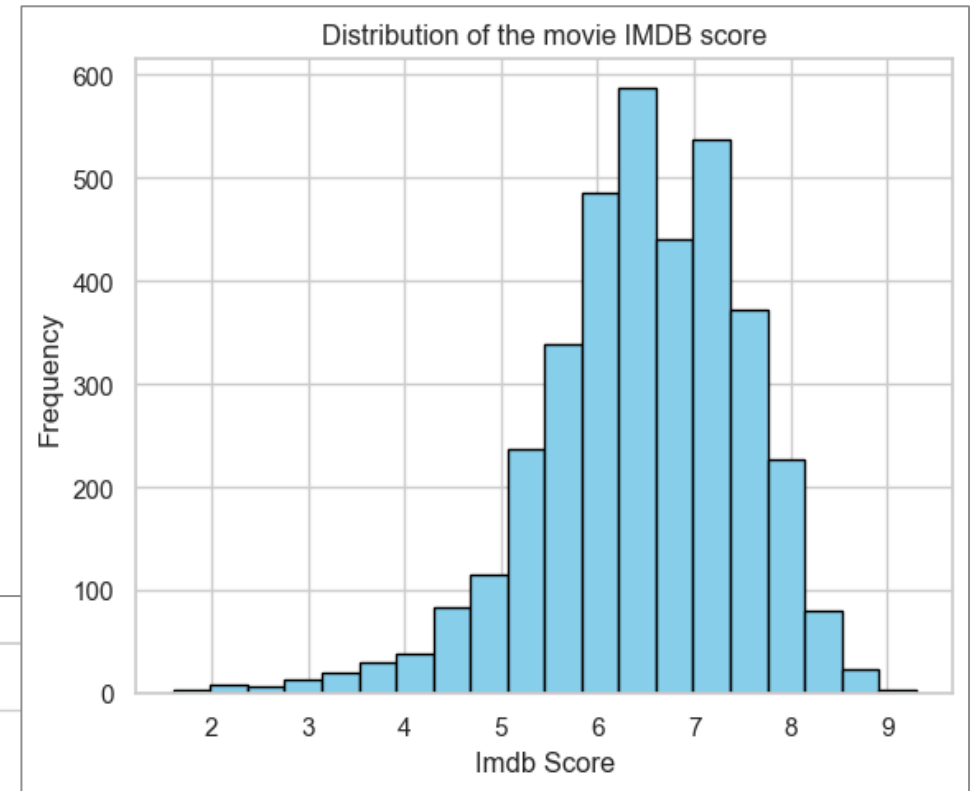
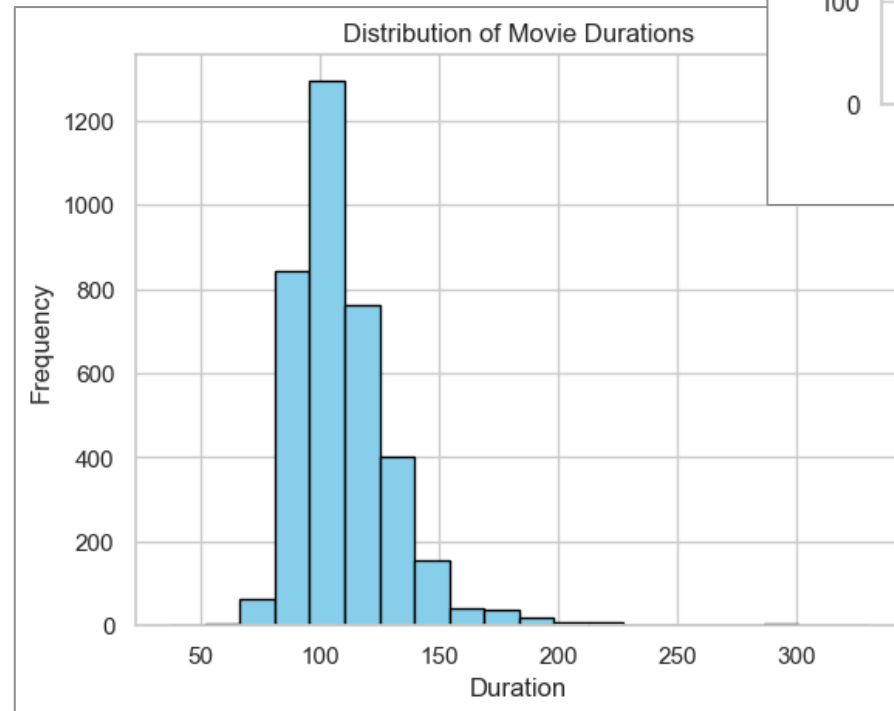
Data Visualisations: Influence of various factors



Directors play a significant role in influencing IMDb scores. The boxplot analysis for top directors reveal varying distributions of IMDb scores. Notably, directors such as Clint Eastwood and Steven Spielberg have a higher median IMDb score compared to others.

Data Visualisations: Distribution plots

The maximum number of movies (more than 1200) have duration around 90-110 minutes, followed by around 850 movies with duration ranging from 80 to < 100 and almost 750 movies having duration of 120-140.



We observe that the majority of movies (nearly 600) have Imdb score approximately 6.2-6.7 or 7-7.3

Exploring variations in movie durations, budgets, IMDB scores and genres

We observe that on average, movies have a duration of around 110 minutes, a gross of approximately 51 million dollars, a budget of around 45 million dollars, and an IMDB score of 6.46. Some genres, such as Film-Noir, Biography, and History, tend to have higher average IMDB scores, while Horror movies have a slightly lower average score.

genres	imdb_score	budget
Adventure Animation Drama Family Musical	8.5	4.500000e+07
Crime Drama Fantasy Mystery	8.5	6.000000e+07
Adventure Animation Fantasy	8.4	2.400000e+09
Adventure Drama Thriller War	8.4	1.400000e+07
Biography Drama History Music	8.3	1.800000e+07

This information gives us insights into the top-rated genres based on IMDB scores, considering a minimum IMDB score of 7.5

Key Hypotheses

- **Hypothesis 1:** Higher movie budget is associated with higher IMDB scores

Correlation between budget and IMDB score: 0.029733230713190027.

The weak correlation in Hypothesis 1 implies that, within this dataset, the budget of a movie has a limited impact on its IMDB score.

- **Hypothesis 2:** Action movies have higher IMDB scores compared to other genres

Mean IMDB score for action movies: 6.285064935064935

Mean IMDB score for non-action movies: 6.523141706334676

Hypothesis 2 indicates that, on average, non-action movies have slightly higher IMDB scores than action movies in this dataset.

- **Hypothesis 3:** There is a positive relationship between director Facebook likes and IMDB scores

Correlation between director Facebook likes and IMDB score:
0.19606683823937404

Hypothesis 3 suggests a weak positive association between the number of director Facebook likes and IMDB scores.

Summary

Based on the analysis of genres with high IMDb scores within the proposed budget range, it is recommended that the studio consider exploring genres such as "Action|Adventure|Drama," "Adventure|Animation|Drama|Family|Musical," and "Adventure|Drama|Thriller|War."

Higher budgets do not necessarily guarantee higher IMDb scores. Other factors, such as genre and director influence, play crucial roles.

Considerations for director selection, genre preferences, and market-specific strategies should be integral to the decision-making process.