DSFT-07 PHASE 1 PROJECT PRESENTATION

By Mainye Gesanda Peter

PROJECT OVERVIEW

The project aims to perform a comprehensive Exploratory Data Analysis (EDA) on a movies dataset to extract insights .

The dataset contains information about various movies, including details such as budget, revenue, genre, release date, and audience ratings.

PROJECT OBJECTIVES

TO find out;

1. Which is the most common genre combination from the Dataset?
2. which is the top popular genre combination from the Dataset?
3. Which is the most voted movie?
4. Which movies made the highest gross profits and losses?
5. Whats is the duration of movies?
6. Which are the movies with the lowest and highest budget?
7. Which movies produced highest and lowest revenue?

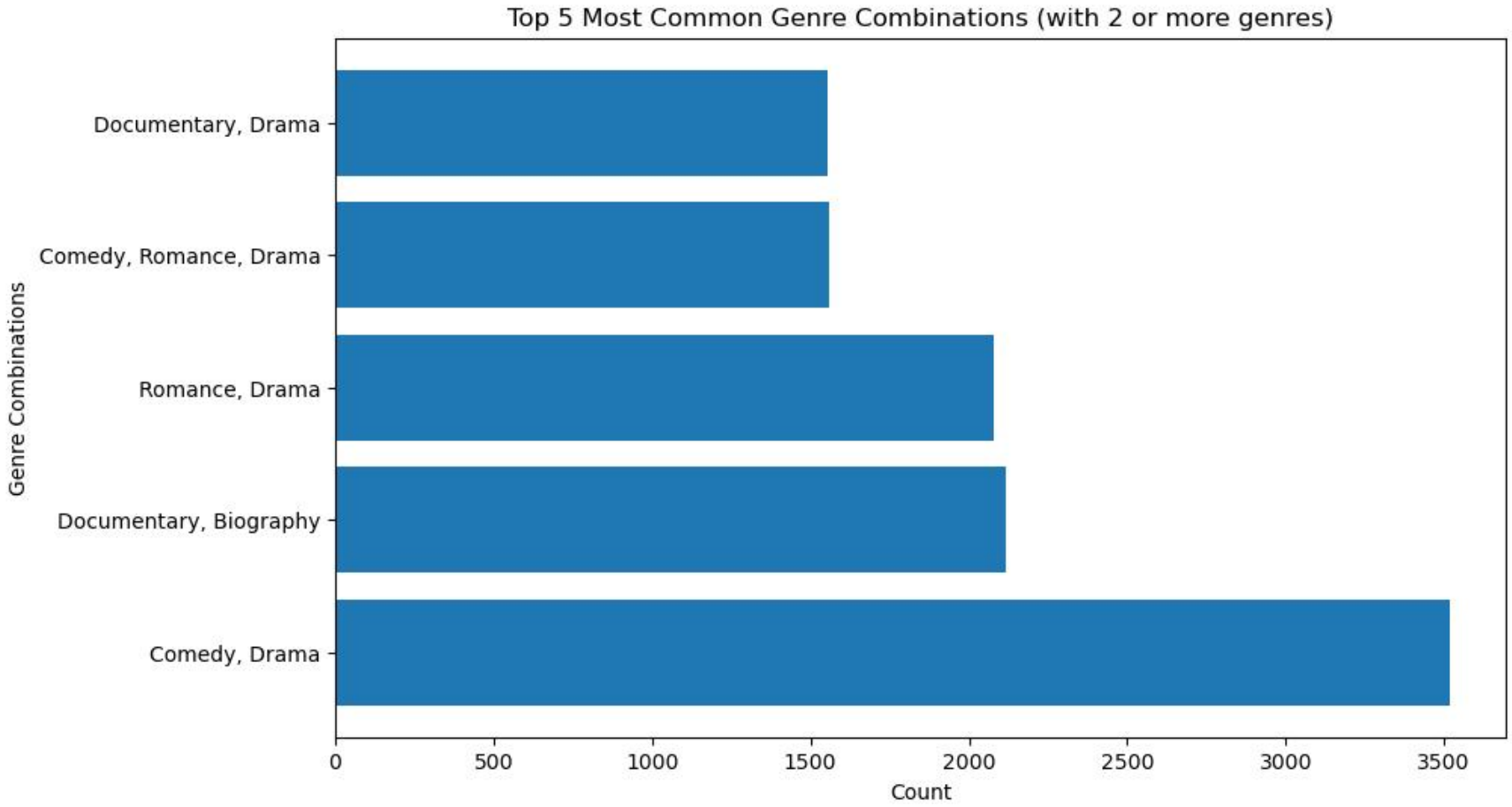
DATA

- 1.SQL Database: There is a im.db relational database.
- 2.TSV Files: Two TSV (Tab-Separated Values) files were used, presumably containing supplementary data for the analysis. One has reviews
- 3.CSV Files: Two CSV (Comma-Separated Values) files were included

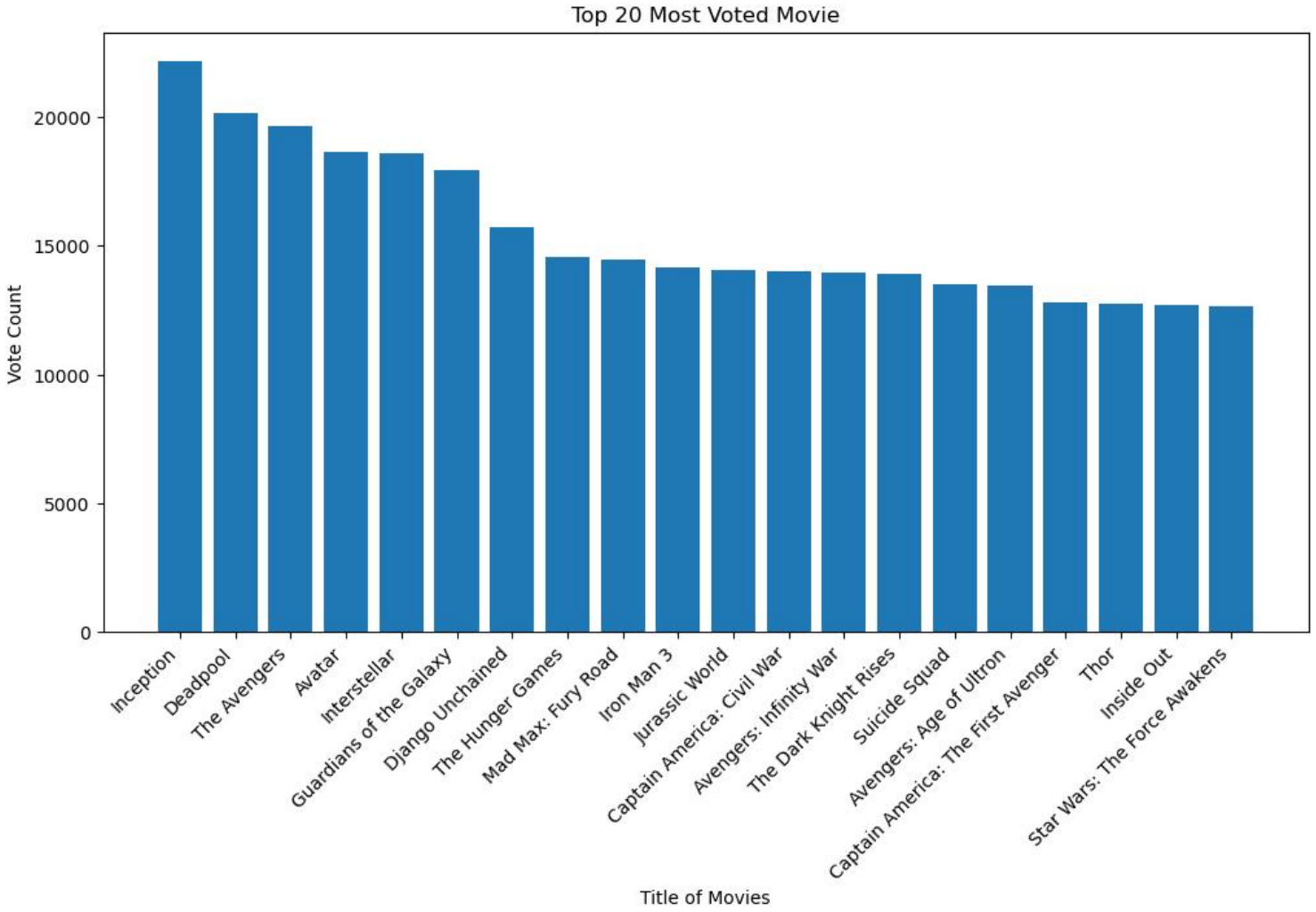
PROCEDURES

- Importing all necessary packages i.e numpy, pandas, sqlite3, matplotlib.plotly & pyplot
- Loading the datasets and creating dataframes
- Cleaning the data
- Data preparation and analysis
- Data visualization

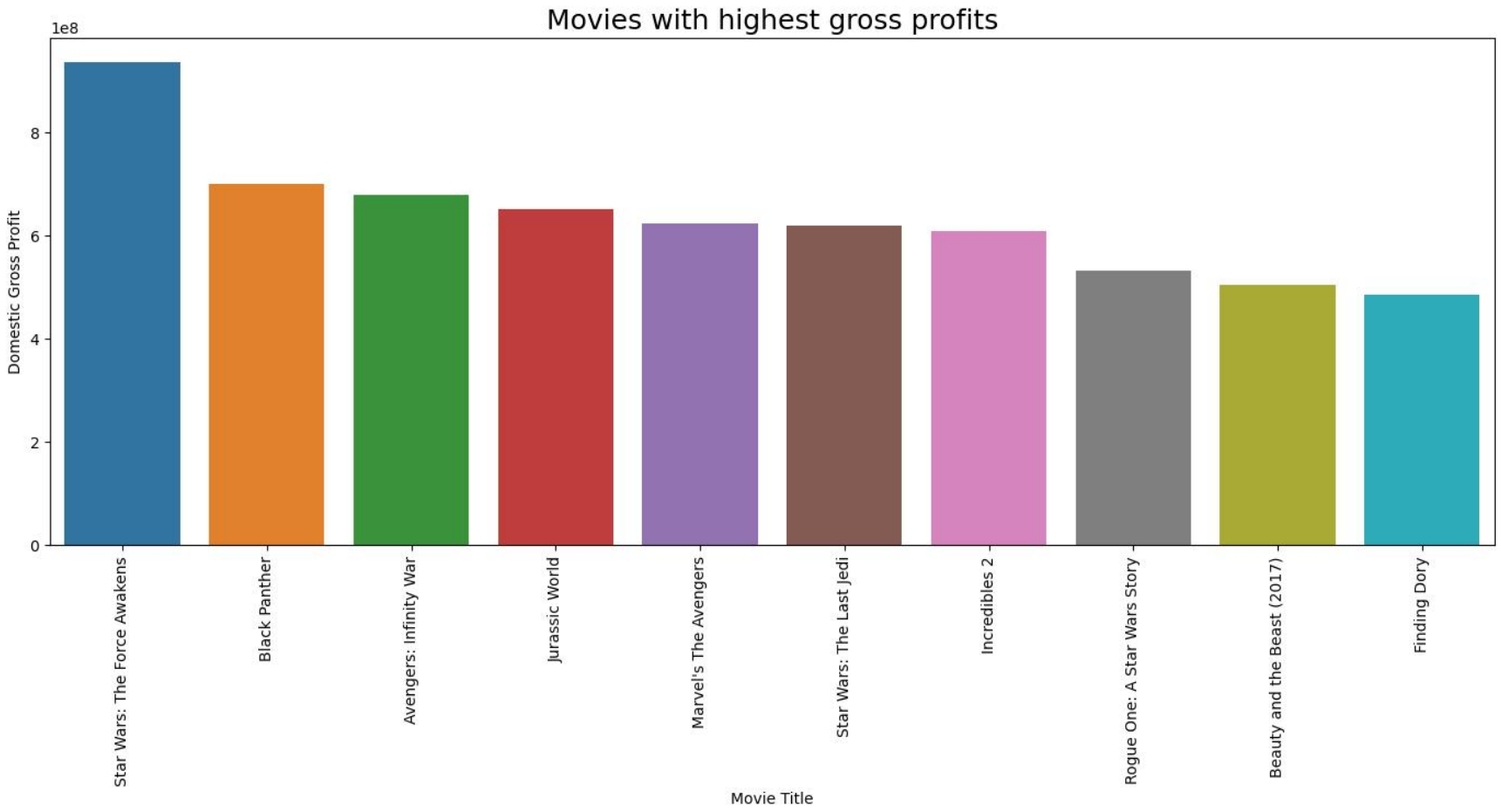
RESULTS


Top 5 Most Common Genre Combinations (with 2 or more genres)
Genre Combinations
Documentary, Drama
Comedy, Romance, Drama
Romance, Drama
Documentary, Biography
Comedy, Drama
0
500
1000
1500
2000
2500
3000
3500
Count

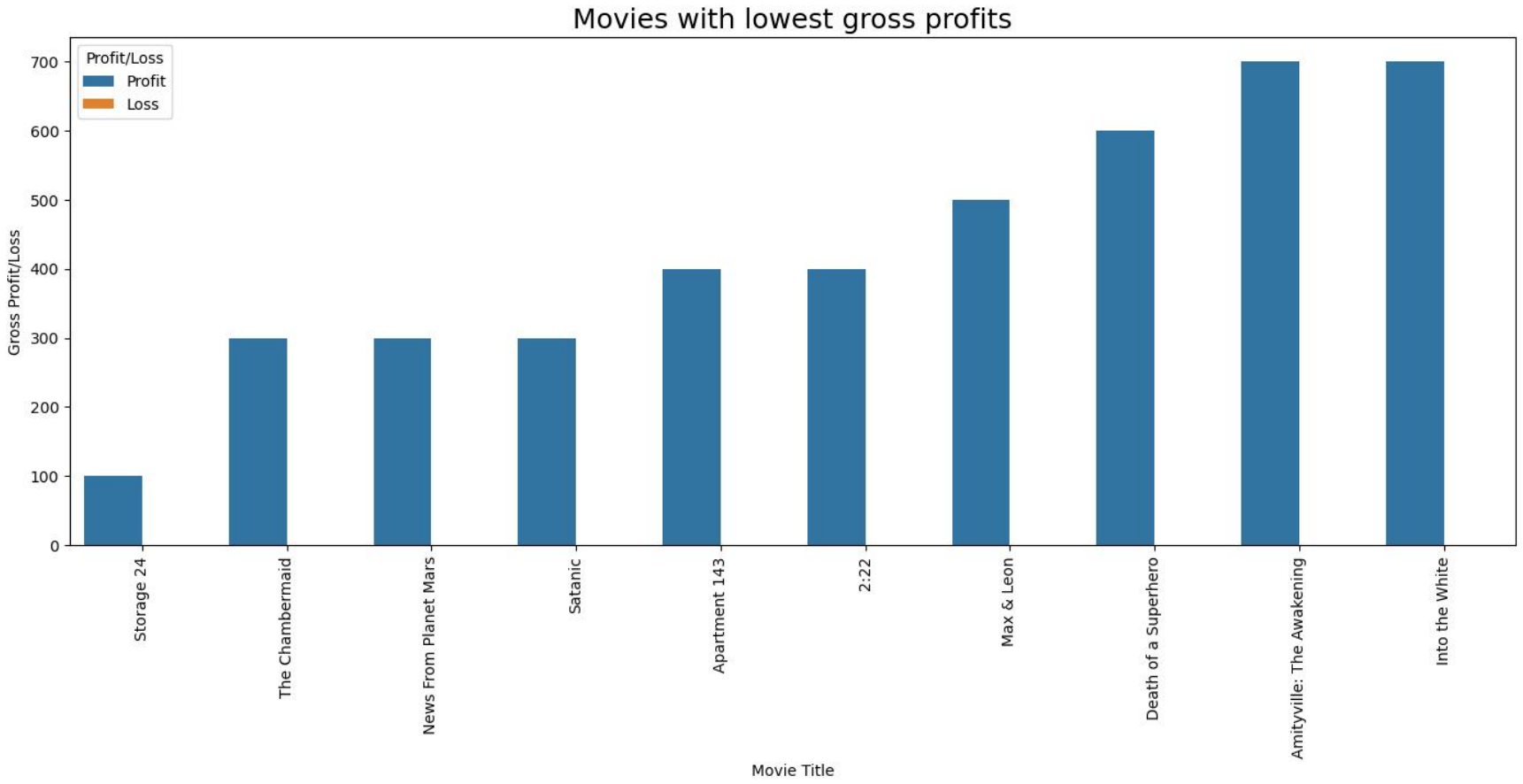
The most voted movie



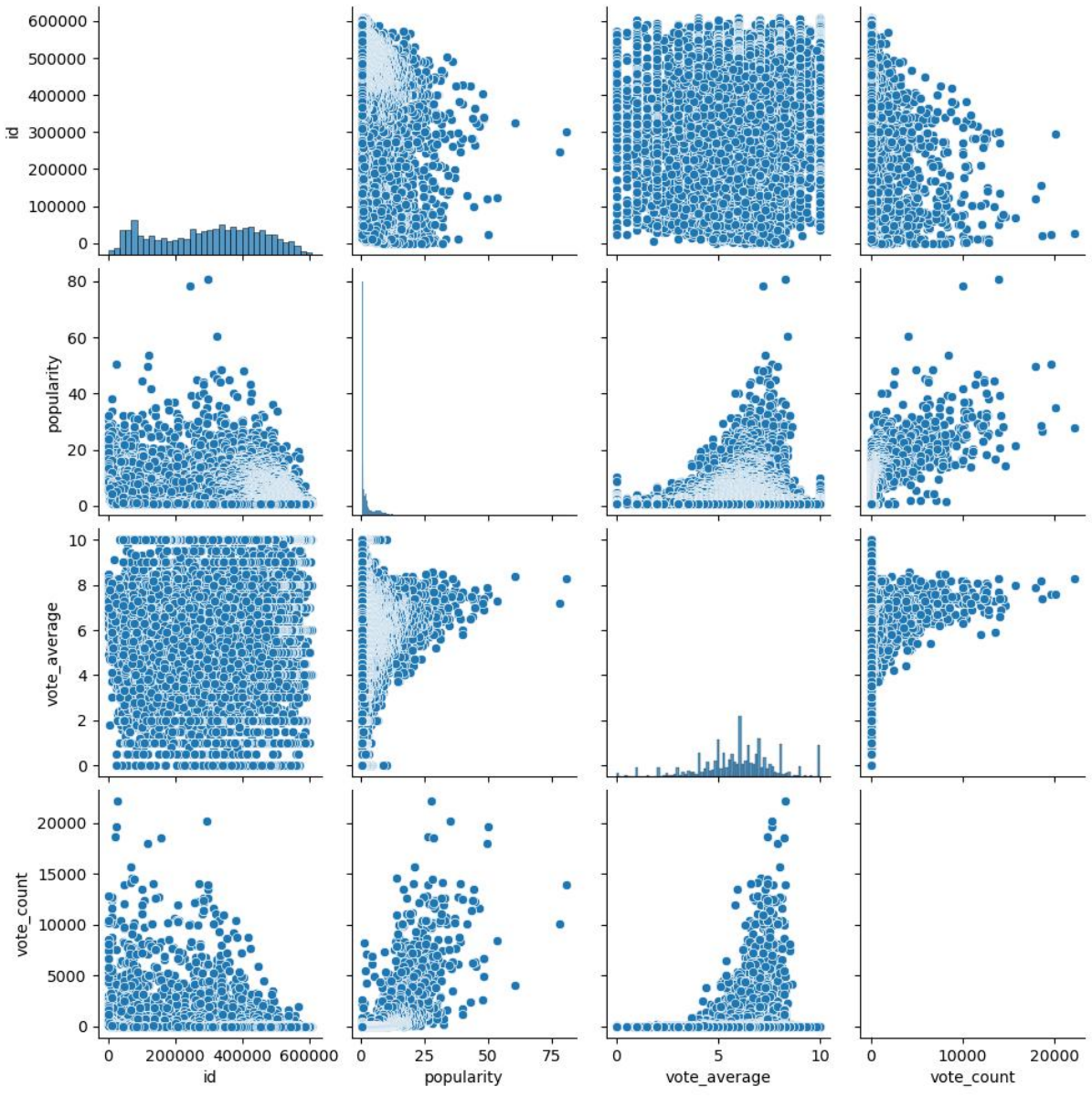
Movies with highest gross profit



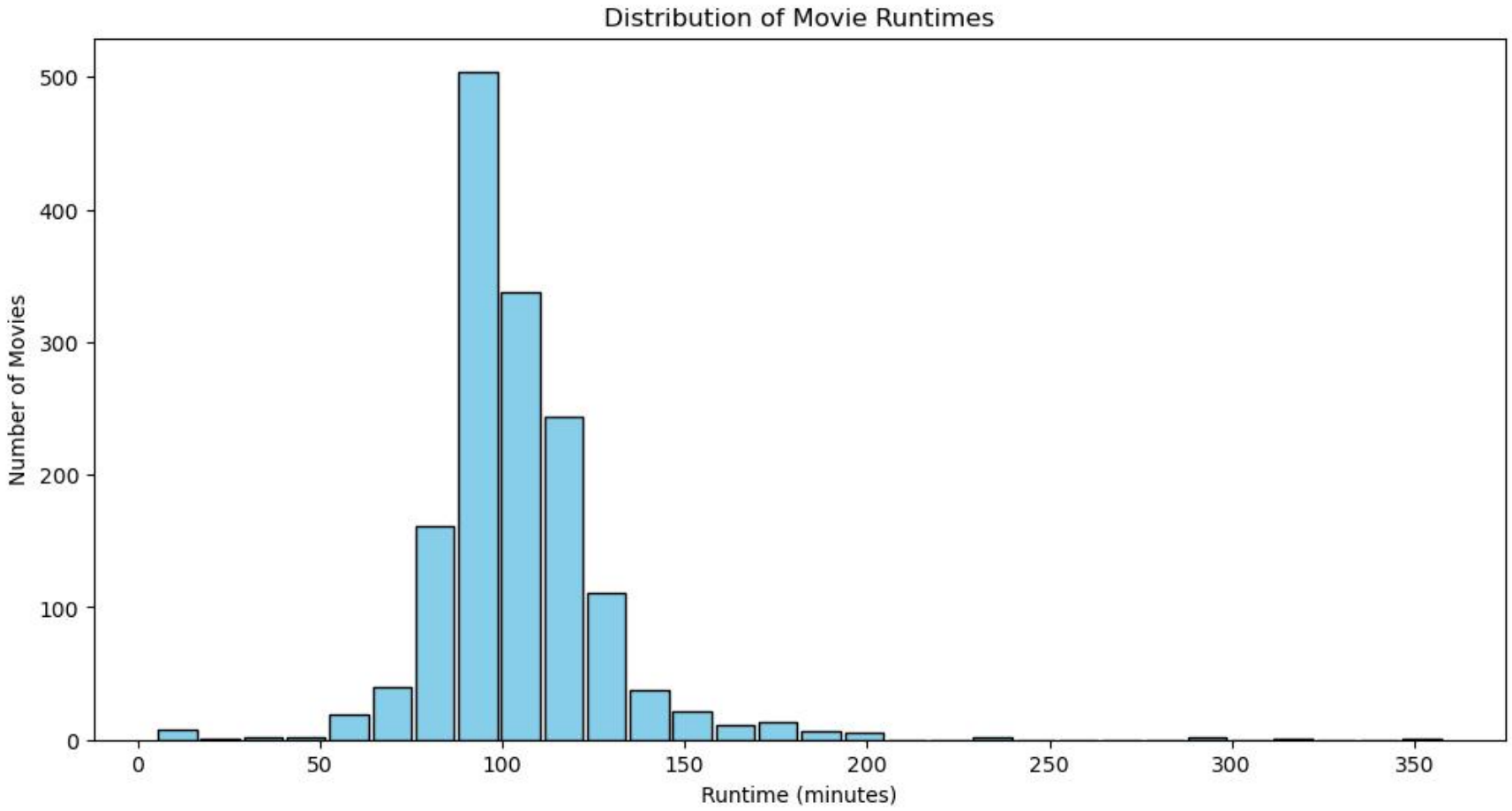
Movies with lowest gross profit



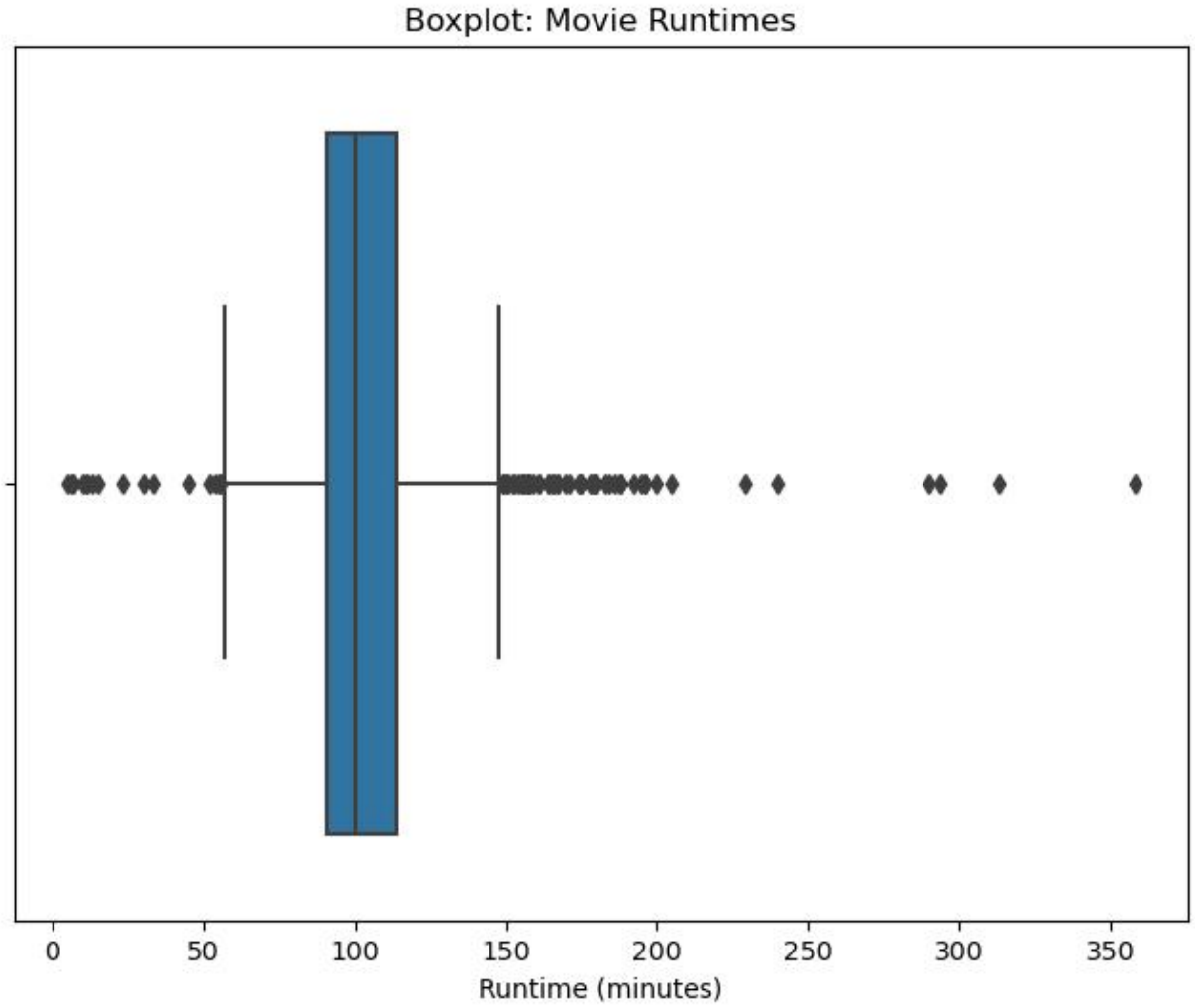
Pairplot


id
popularity
vote_average
vote_count
0
100000
200000
300000
400000
500000
600000
0
20
40
60
80
0
2
4
6
8
10
0
5000
10000
15000
20000
0
200000
400000
600000
0
25
50
75
0
5
10
0
10000
20000

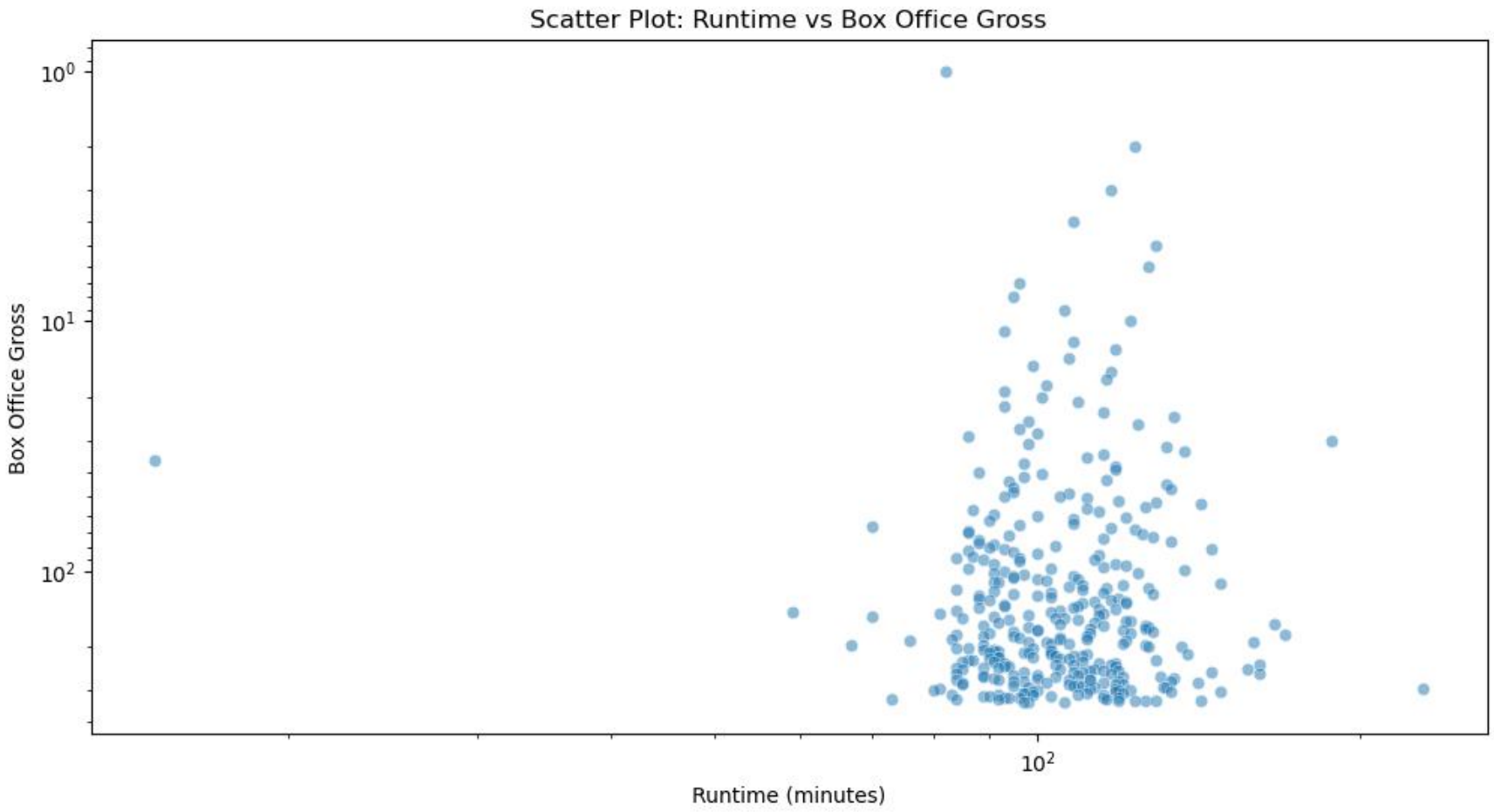
Duration of the movies(runtime)



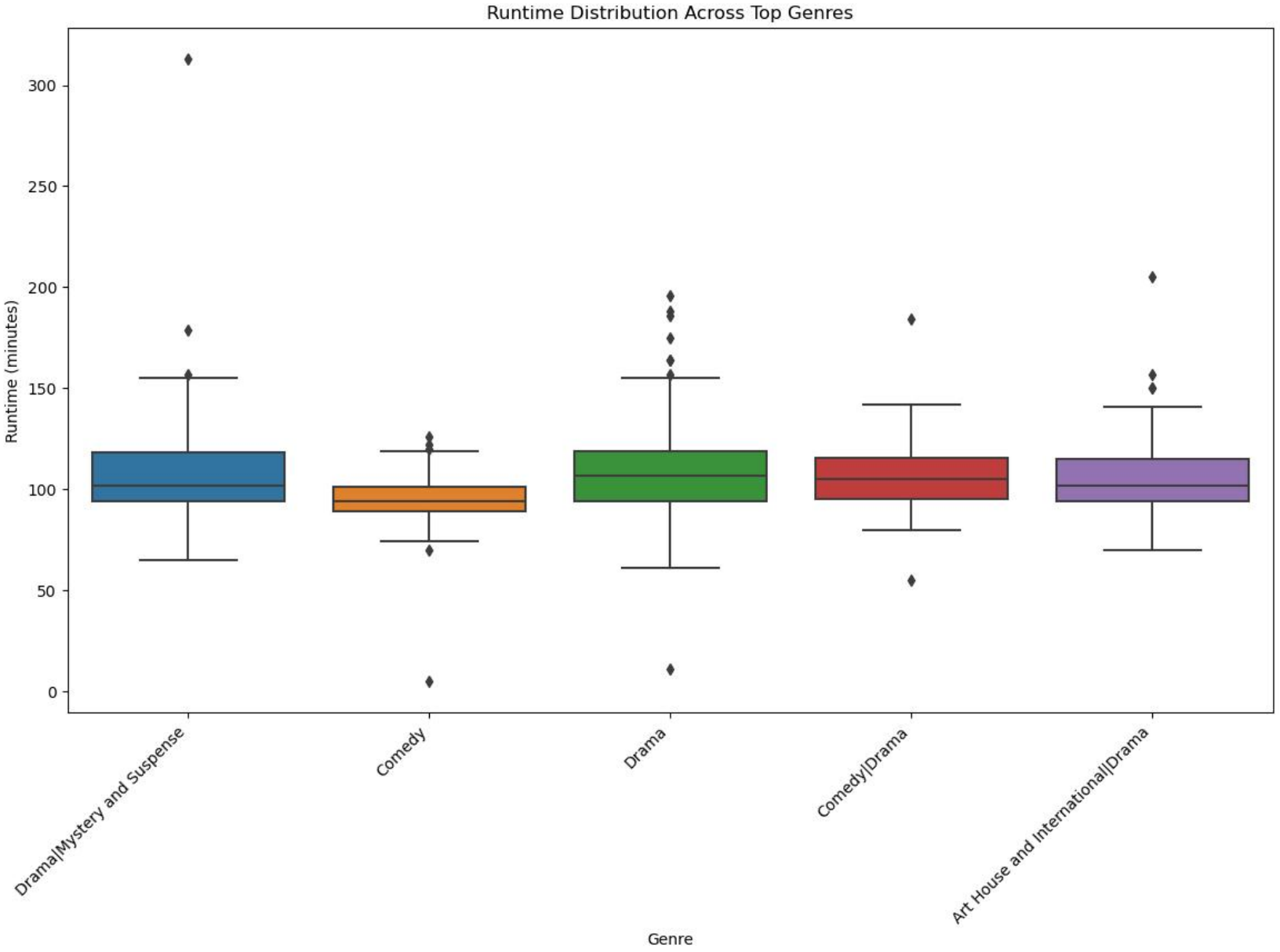
checking out outliers using a boxplot



scatter plot



runtime distribution



CONCLUSION

- The Movie 'Inception' received most numbers of votes from people
- 'Comedy, Drama' combination is the Top most genre combination from the Dataset
- The top popular genre is the 'Action,Crime,Drama’
- Star Wars:The Force Awakens made the highest profit
- Storage 24 made the lowest gross profit
- The runtime for most movies is 100
- Movies are mostly made in drama, comedy, thriller, action, romance, adventure but apart from adventure & action, animation, family, fantasy, science Fiction are the genres making gross profit

Recommendation

- I recommend the creation comedy and drama movies and also creation of movies that yield more profit.

THANK
YOU !