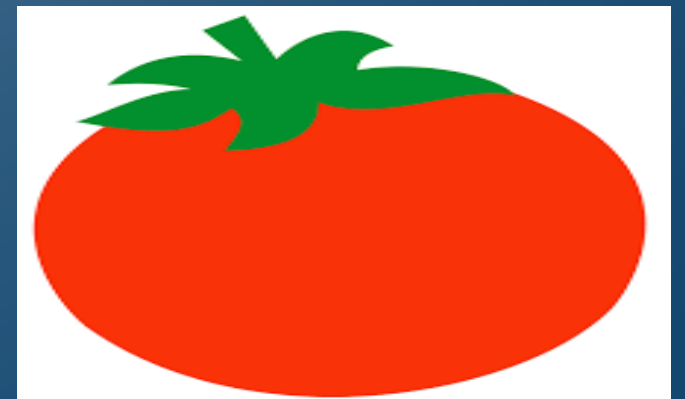# FIDELIS WANALWENGE

MOVIE DATA EXPLORATION & ANALYSIS
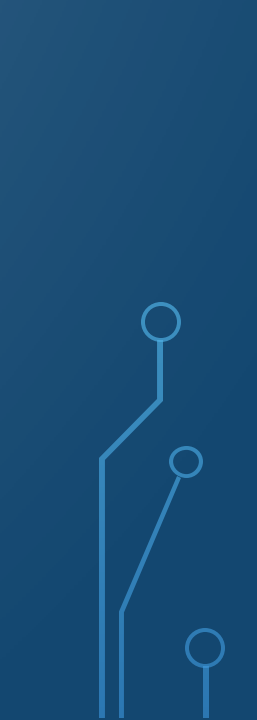# FOR MICROSOFT SRUDIOS

JULY 24, 2023

Summary

We used exploratory data analysis (EDA) to clean, assemble, and interpret movie data to make recommendations for creating successful films based off of profit/loss, genre popularity, and movie runtime

Our results displayed some potential intersections for these metrics to maximize return on investment, and viewer ratings.

Outline

- Business Problem

- Data

- Methods

- Results

- Conclusion

- Business Recommendations

Business Problem

This project uses data from IMDB, and The Numbers and analyses it to come up with three suggestions for helping the head Microsoft's new movie studio department. It will aim to answer the following:

1. What is the best time of the year to release movies?

2. What is the most popular genre?

3. What is the relationship between rating vs profit/loss and rating vs release year?

4. What is the Average movie runtime?

Data

1. [Box Office Mojo](#)

2. [IMDB](#)

3. [Rotten Tomatoes](#)
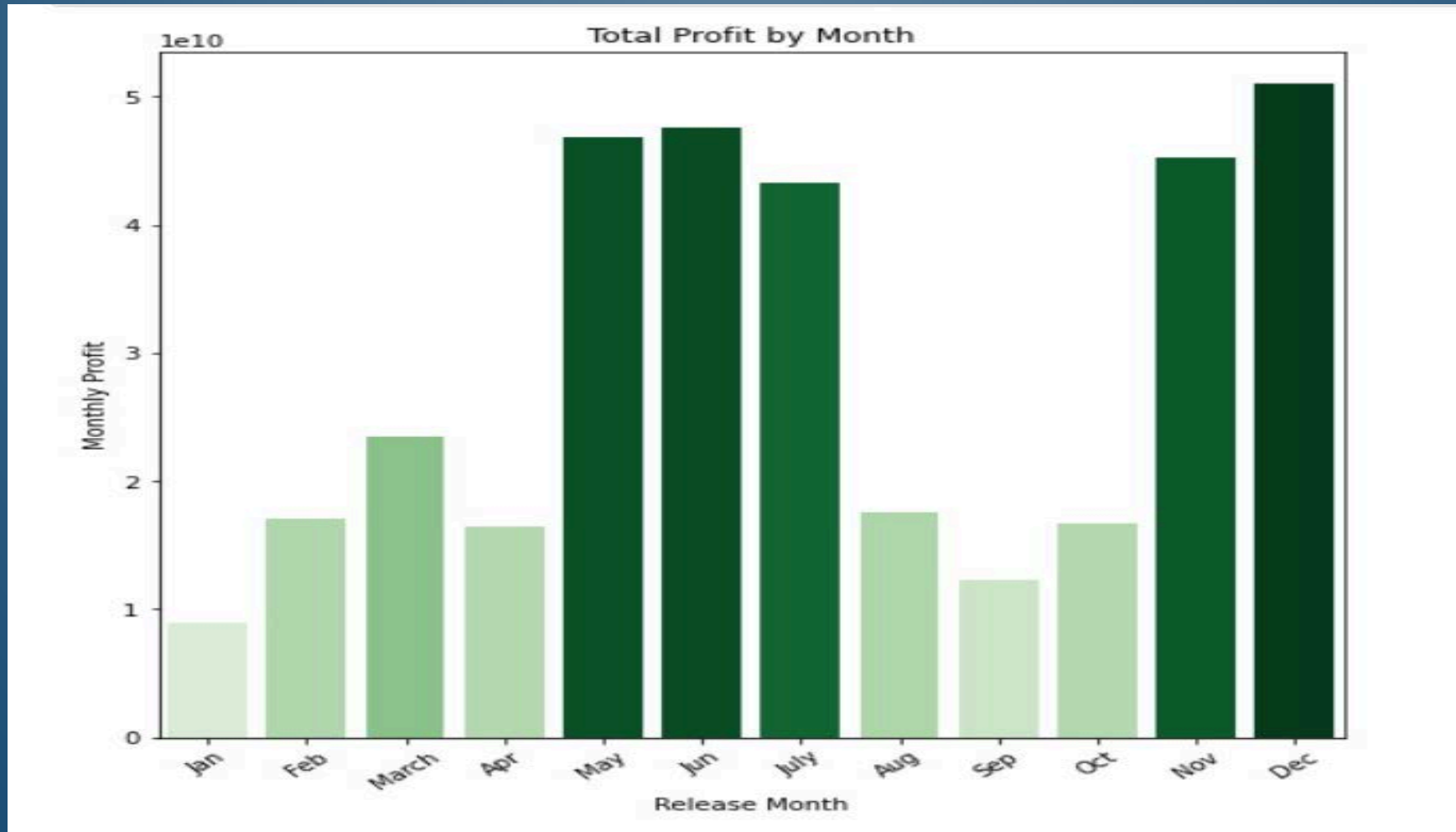
4. [TheMovieDB](#)
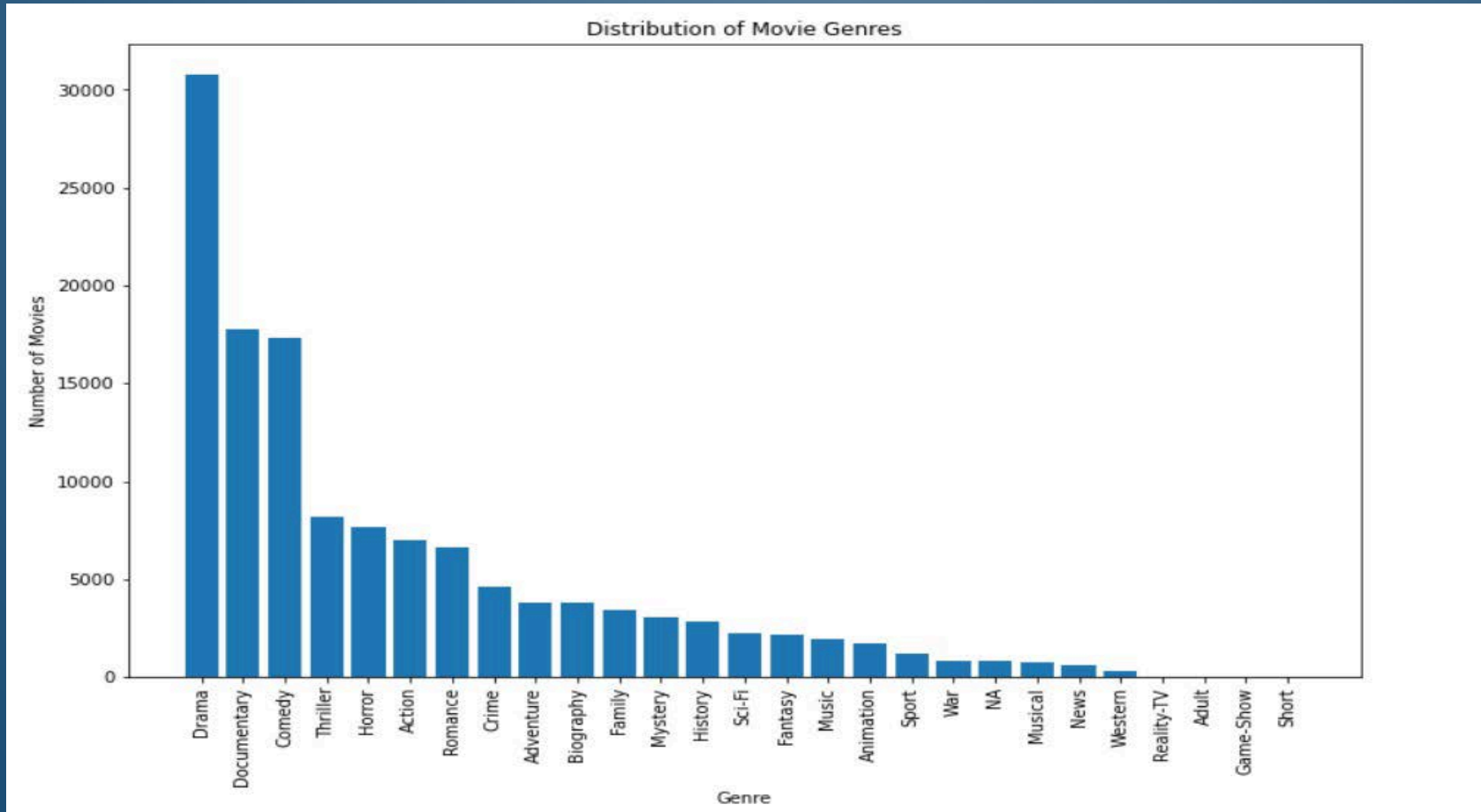
5. [The Numbers](#)

Methods

1. Import libraries

2. Open relevant files

3. Exploratory data analysis and Data cleaning

4. Data analysis

5. Join files for further analysis

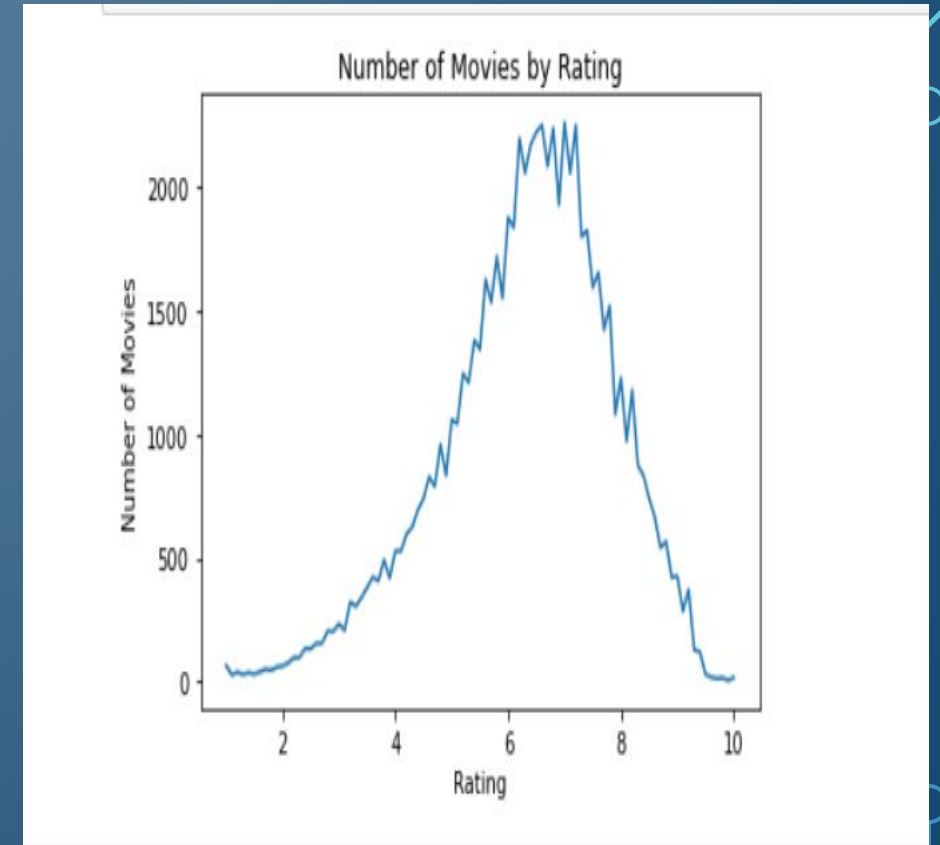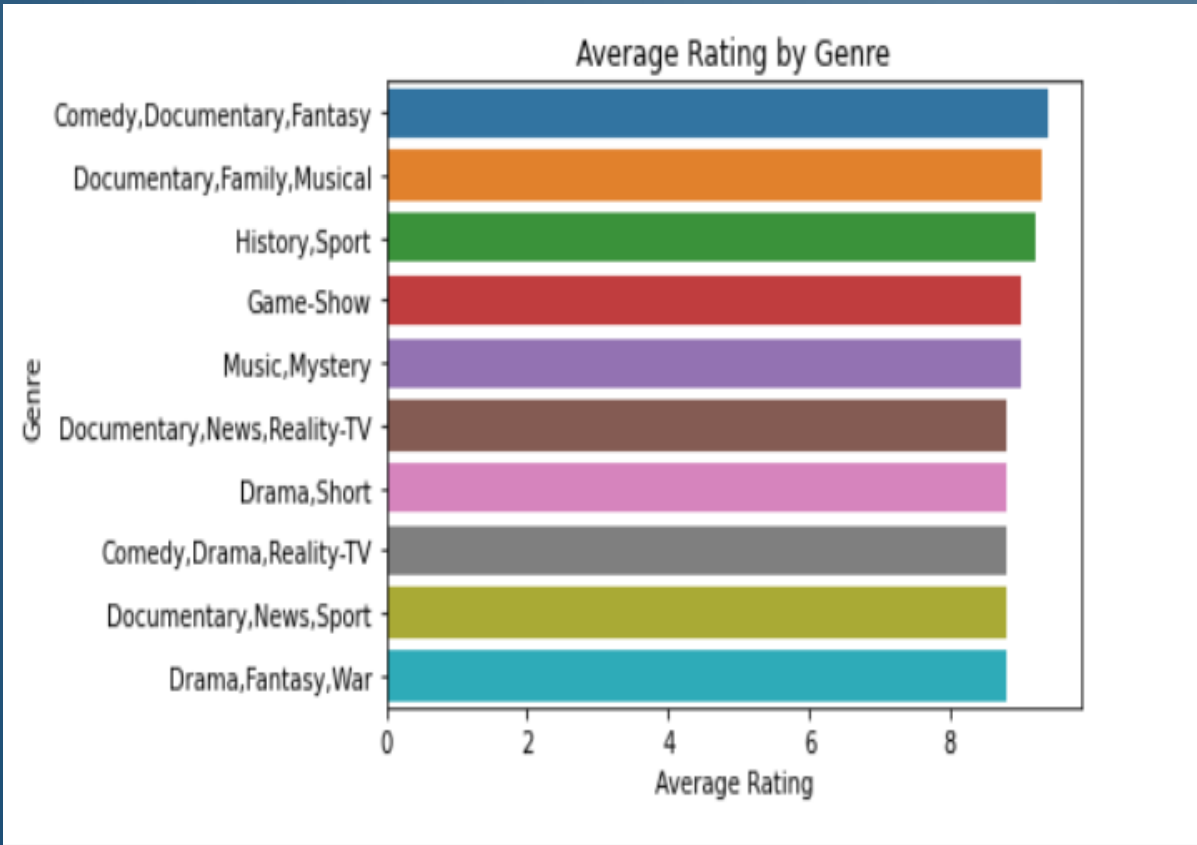6. Data analysis and visualization

Best Time to Release a Movie

- The top 3 months to release the movie are December, June and May for better profits

Most Popular Genre by Distribution
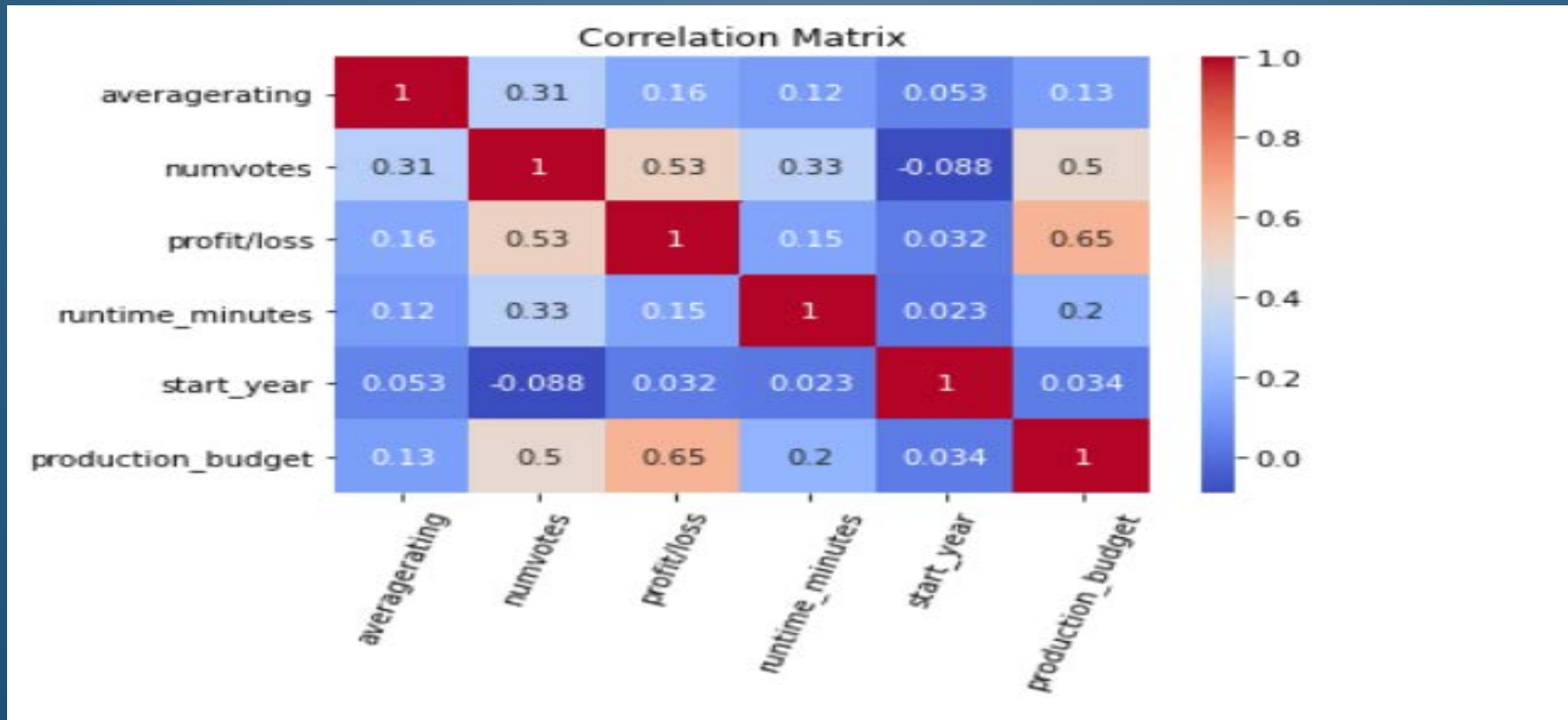

Distribution of Movie Genres

- The most common Genre is Drama
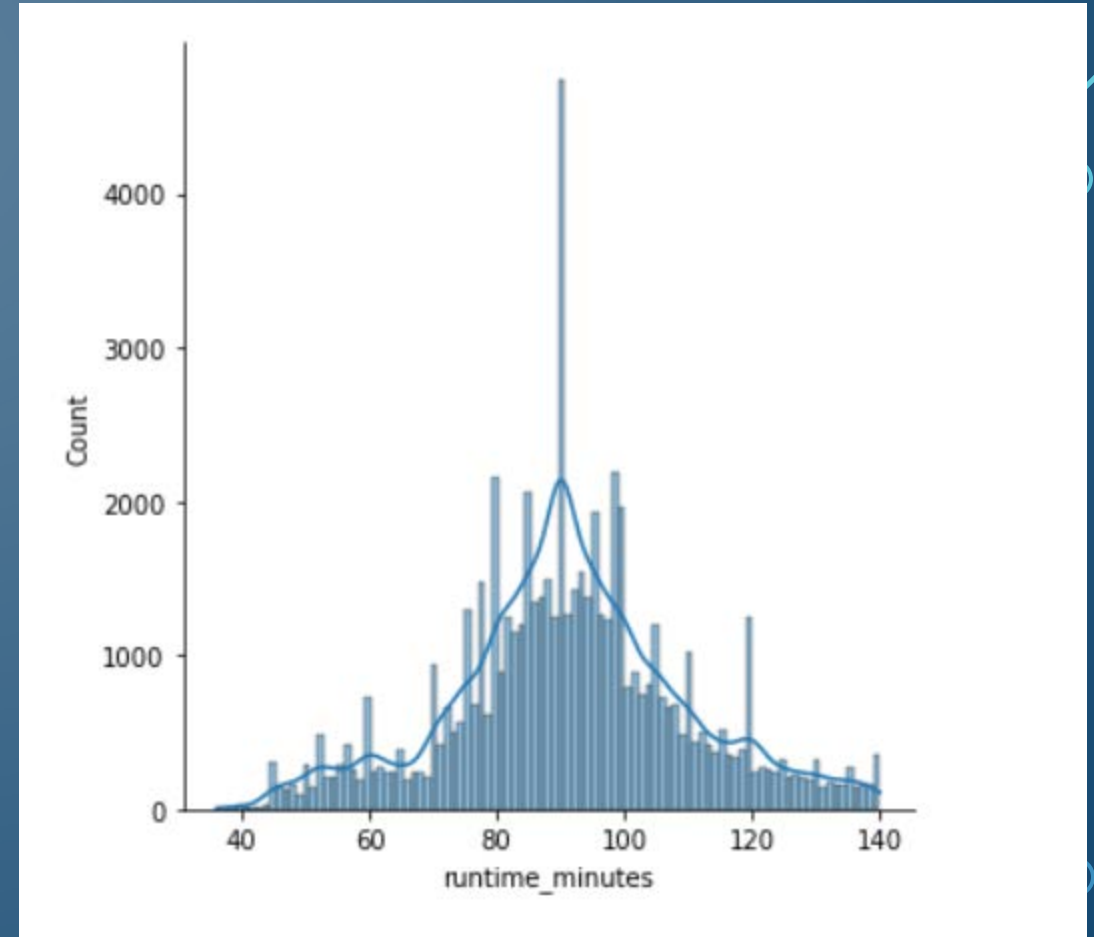
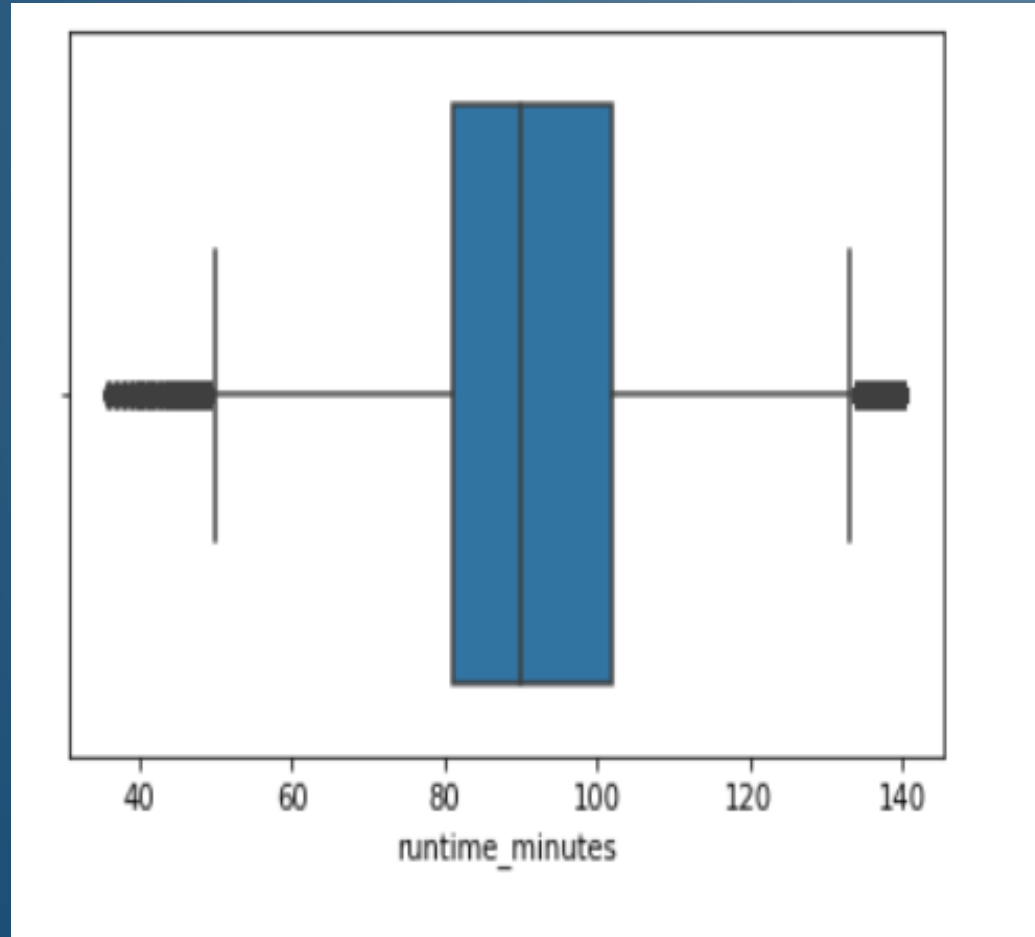# Most Popular Genre by Average Rating

- The Genre Comedy, Documentary, Fantasy is rated highest by average rating

- The average rating of most movies is between 6-7

Relationship between rating vs profit/loss and rating vs release year

- Averagerating Vs profit/loss have 0.16 correlation between them. This means there is a low possibility that movies with higher average ratings have higher profits.

- Production_budget Vs profit both have 0.65 correlation between them. This means there is a good possibility that movies with higher investments result in better revenue.

- Start_year Vs num votes have a negative(-0.088) correlation. This means that movie rating (num votes) do not depend on the release year.

Average movie runtime

- The average movie run_time after eliminating outliers is 91 minutes

Recommendations

1. Plan the production schedule for a release in December

2. Higher production budgets are likely to have a higher profit owing to the correlation of 0.65

3. Comedy, Documentary, Fantasy films are highest rated while Drama is the most common

4. The average length of films is 91 minutes hence this can serve as a benchmark for production length of films

THANK YOU!