Microsoft's new movie studio

BY

ZAINABU CHEROTICH RONO

Instructors name; Rose Delilah, Asha Deen, Noah Kandie, Anthony Muiko

NOVEMBER, 2022

Outline

- Introduction and Business Understanding
- Familiarization with Data
- Summary of Data Cleaning
- Summary of Aggregated and Joined Data
- Summary of Data Visuals and Interpretation
- Business Conclusion and Recommendations

Introduction and Business Understanding

- **Business Problem**

- How can Microsoft determine the nature of Studio that it can set up for it to create movies that are widely consumed and makes financial sense to the business?

- The presentation will highlight Data exploration from various sources to determine the movies that Microsoft can invest in for their new studio. This will be based on three key points;
  - *The most consumed movies across the globe.*
  - *The most rewarding movies, businesswise.*
  - *The best kind of movies to invest in.*

Familiarization with Data

- The data utilized for these analysis are extracted from box office. The three major sources are;
    - *bom.movie_gross-csv file*
    - *imdb.title.basics-csv file*
    - *imdb.title.ratings-csv file*
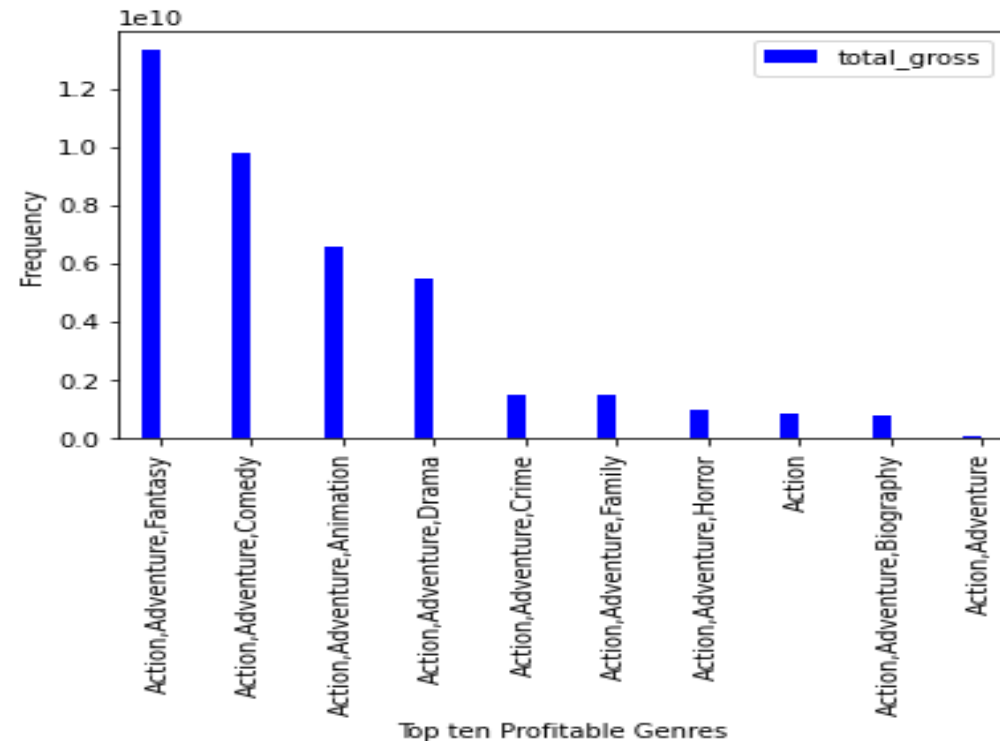
Summary of Data Cleaning

- ***bom.movie_gross***
  - Analyzed using Pandas. Missing values were present in studios, domestic_gross and foreign_gross columns. They were dropped apart from foreign_gross, whose missing values were replaced with respective median. There were no duplicates nor extraneous values.
- ***imdb.title.basics***
  - Analyzed using Pandas. Missing values were present in runtime_minutes and genres columns but were all dropped. There were no duplicates nor extraneous values.
- ***imdb.title.ratings***
  - Analyzed using Pandas. There were no missing values, nor duplicates, nor extraneous values.

Summary of Aggregated and Joined Data

- The three data sets were joined using index column, tconst, from imdb.title.basics and imdb.title.ratings. The resulting dataframe was joined to bom.movie_gross using primary_title column.

- Runtime_minutes and genres had missing values, which were dropped.

- There were no duplicate values.

- Column names were reset to a standard format for uniformity.

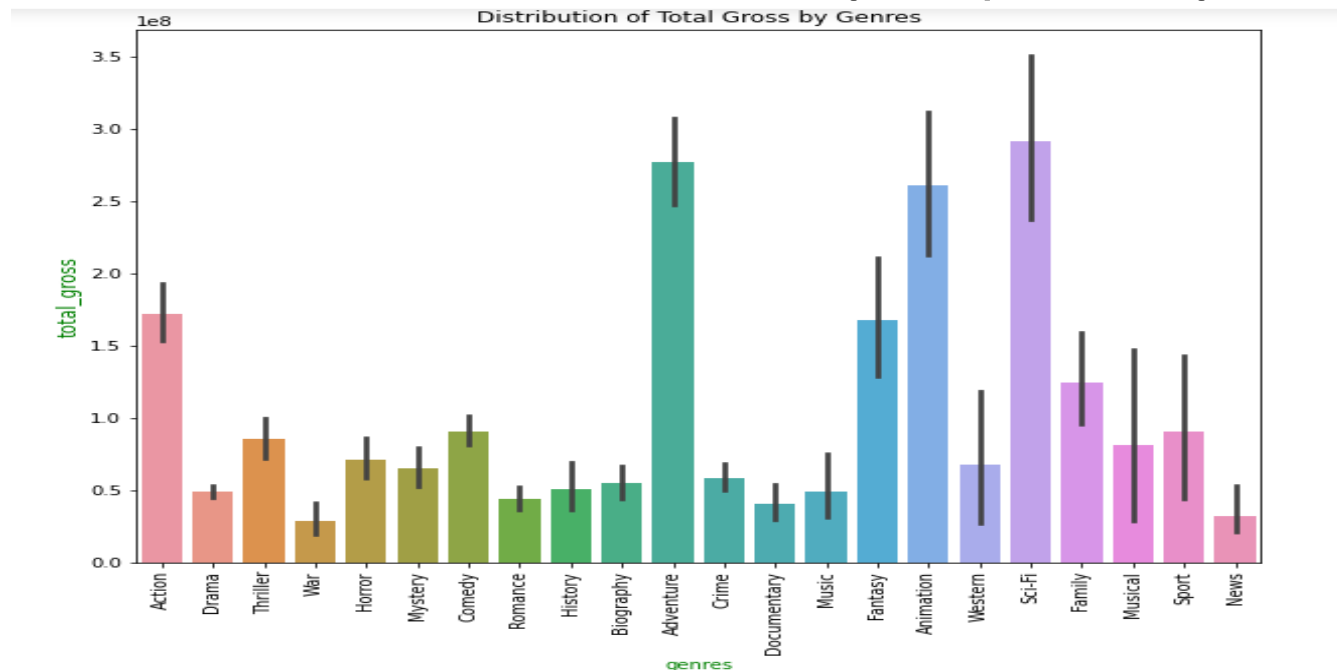- Foreign_gross column was converted from an object data type to a float data type.

Summary of Data Visuals and Interpretation

- **Histogram of genres that have higher gross**

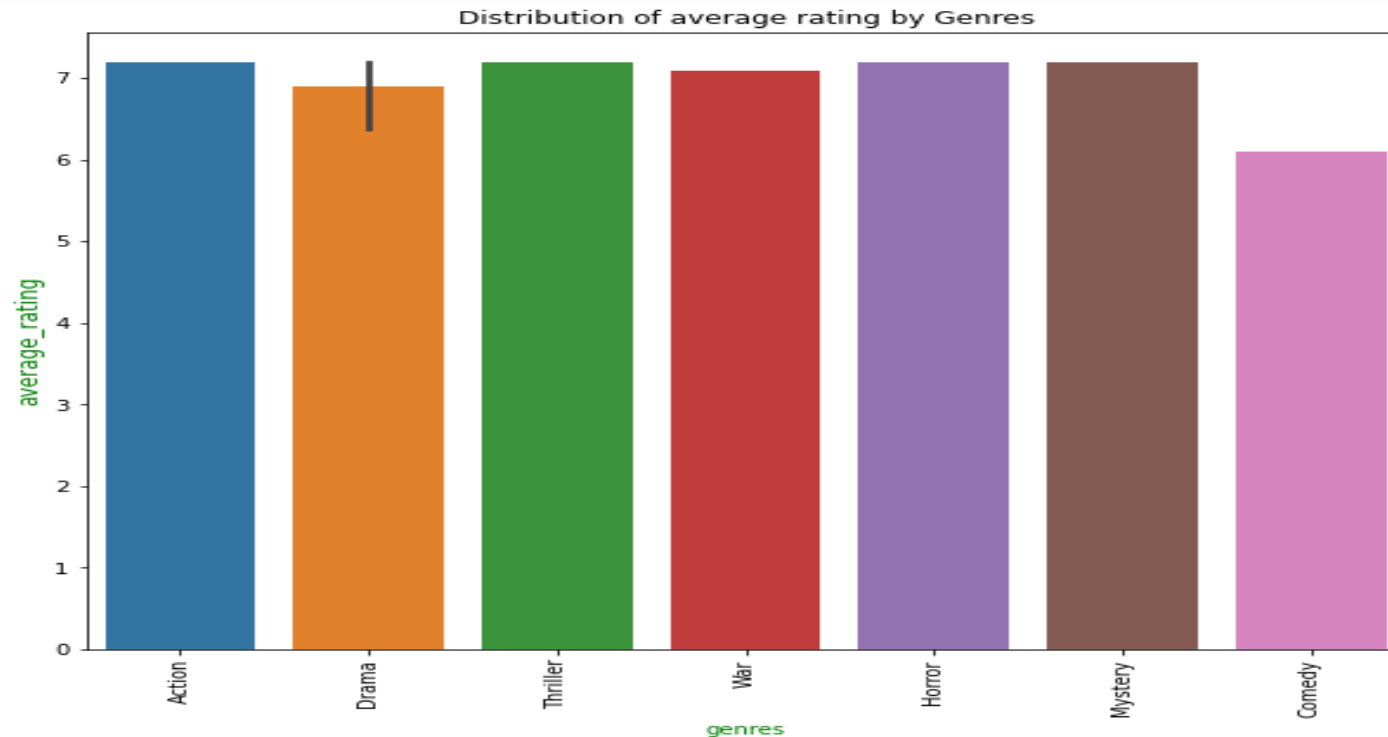- A summary of the top ten performing movies in box office

Summary of Data Visuals and Interpretation

- **Relationship between genres and total_gross**

- The top five most watched genres, with higher pay are; Sci-Fi, Adventure, Animation, Action and Fantasy respectively.
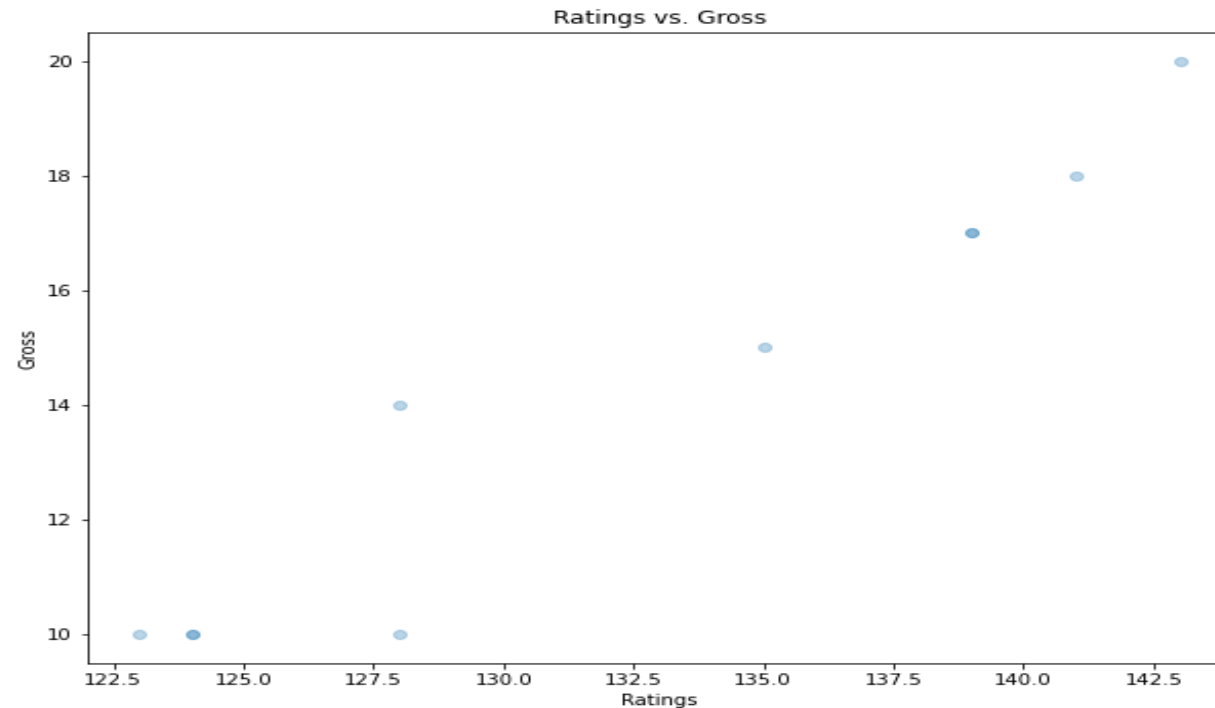
Summary of Data Visuals and Interpretation

- **Relationship between genres and ratings**

- From ratings vs genres, Action, Thriller, War, Horror and Mystery have the highest ratings.

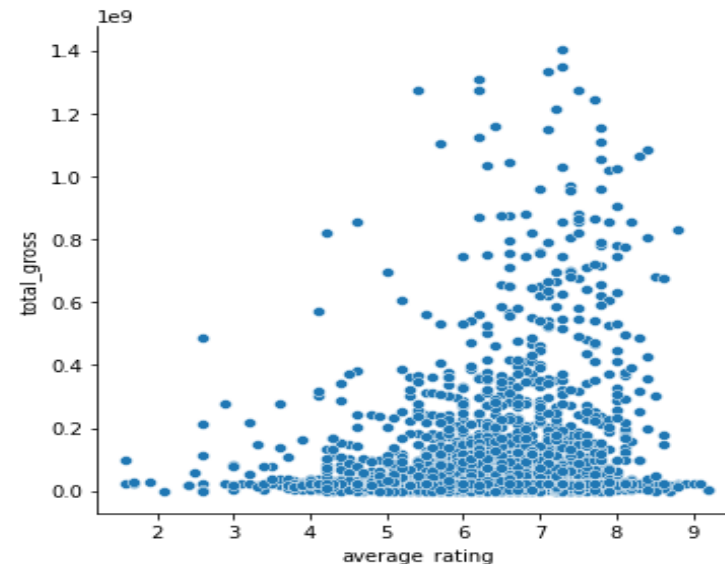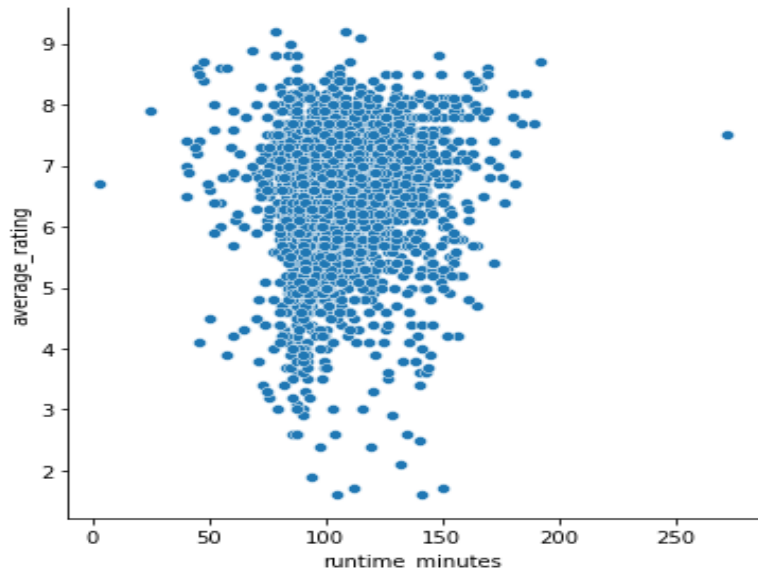Distribution of average rating by Genres

Summary of Data Visuals and Interpretation

- **Relationship between total gross and ratings**

- From the total_gross vs Ratings scatter plot diagram, there exist a positive correlation between the two.
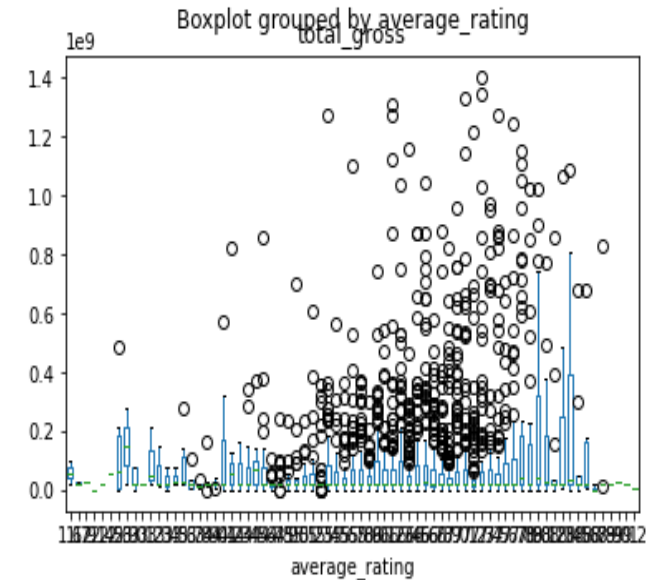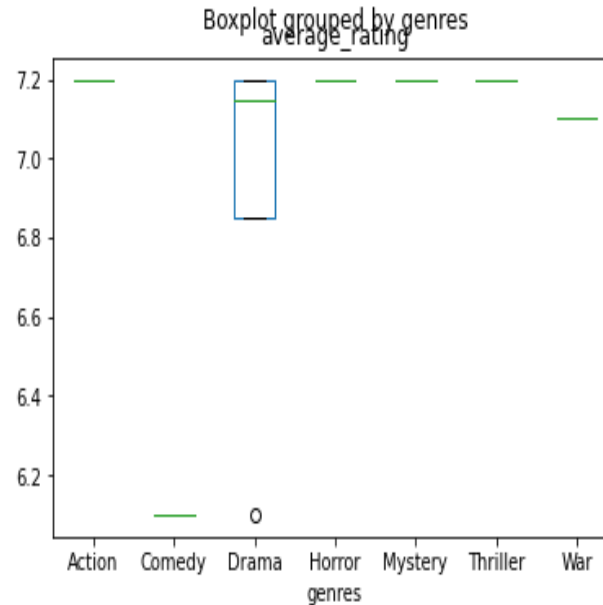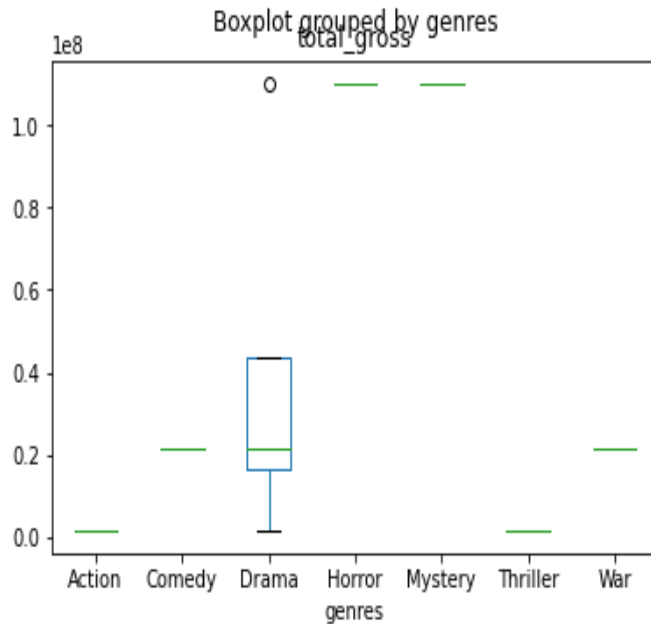
Summary of Data Visuals and Interpretation

- **Relationship between Ratings and Runtime minutes and total gross**

- Most movies with runtime_minutes of between 100-150 minutes are more common and tend to have a good rating of between 6.5-7.

- Ratings vs gross have imperfect positive correlation. Higher ratings tend to have higher gross.

Summary of Data Visuals and Interpretation

- **Relationship between top 10 most watched genres, highest income and highest ratings**

- From the box plot, the genre that has higher rating and higher gross is drama.

Business Conclusion and Recommendations

- From the analysis, the genres that have higher gross are Drama, Horror, Mystery, Thriller and Action.

- This analysis, therefore, recommends the following three business solutions for Microsoft on intent of creating Music studios;
  - 1. The top-five movies doing good in box office are Drama, Horror, Mystery, Thriller and Action. They have high ratings, which translates to higher gross.
  - 2. While producing studio, Microsoft should consider combining 2-3 kinds of genres from the top five, since a combination of two or more have higher ratings and higher gross as compared to individual gender.
  - 3. The movies should be between 100-150 minutes, which have better ratings.