



Microsoft®

Studios

# MICROSOFT STUDIOS PROJECT

By Daniel Muruthi

Data findings and insight.

# Introduction

Based on our movie data, the main process I undertook was Exploratory data analysis (EDA). This step was crucial to the process since it involved analyzing and summarizing the main characteristics of the dataset. It allows me to gain insights into the data, detect patterns, and identify potential outliers or errors. The process of EDA involves several techniques such as visualizing data, computing descriptive statistics, and detecting missing or incomplete data.

This report aims to highlight my findings from the IMDB and Box OfficeMojo data sources and the recommendations made from the findings.

# Steps involved

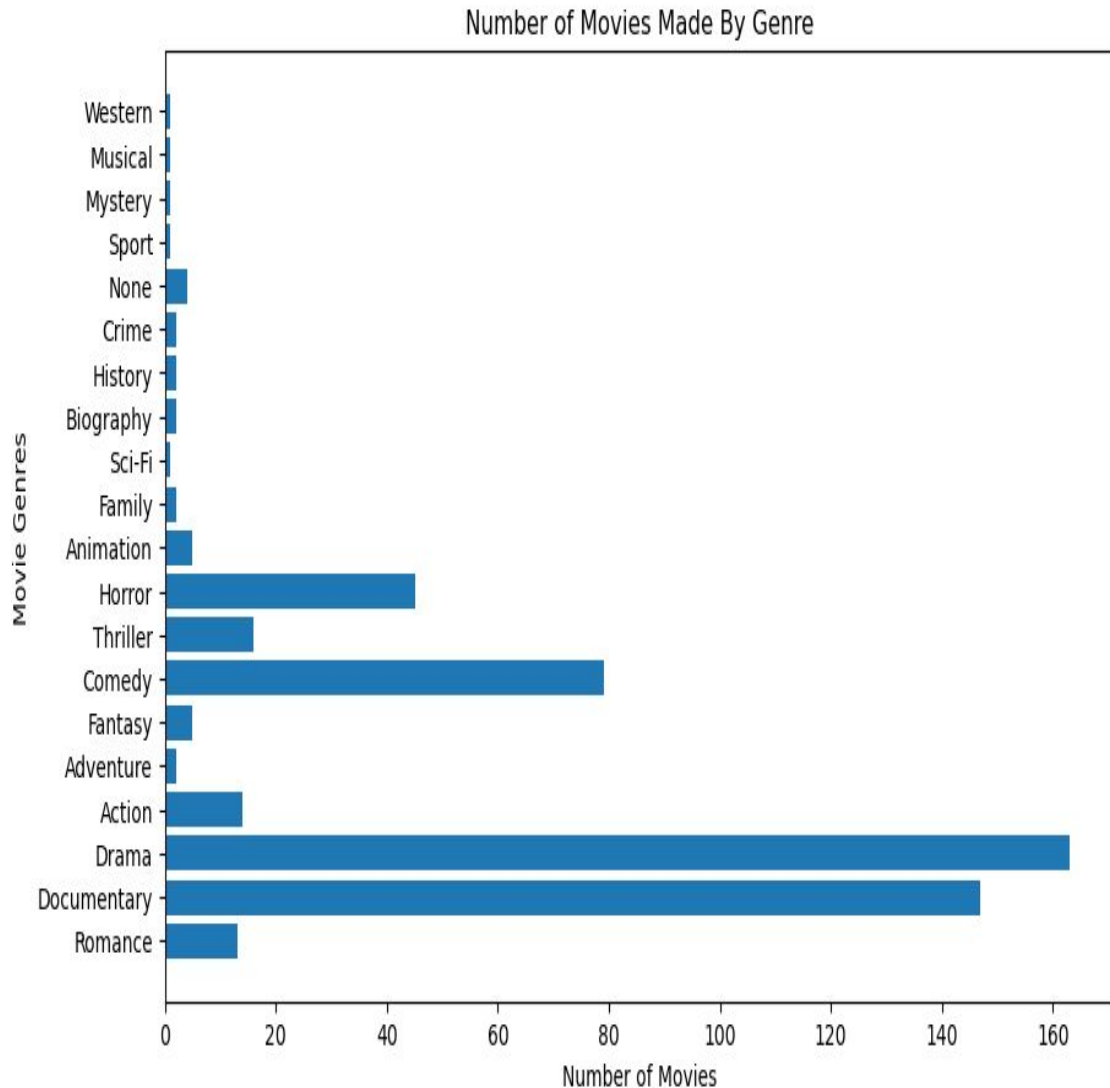
Methods used to achieve this are:

1. Data Loading: This was accessing the different data sources to gain our data
2. Data inspection: This was to get information on the data sources eg tables, columns in tables to find our relevant data.
3. Exploratory data analysis: Exploring the data to find data errors, identify patterns, missing values, uncover relationships between variables and inconsistencies in the data.
4. Visualizations: graphical presentations of the results for a better understanding of the results.

# Findings

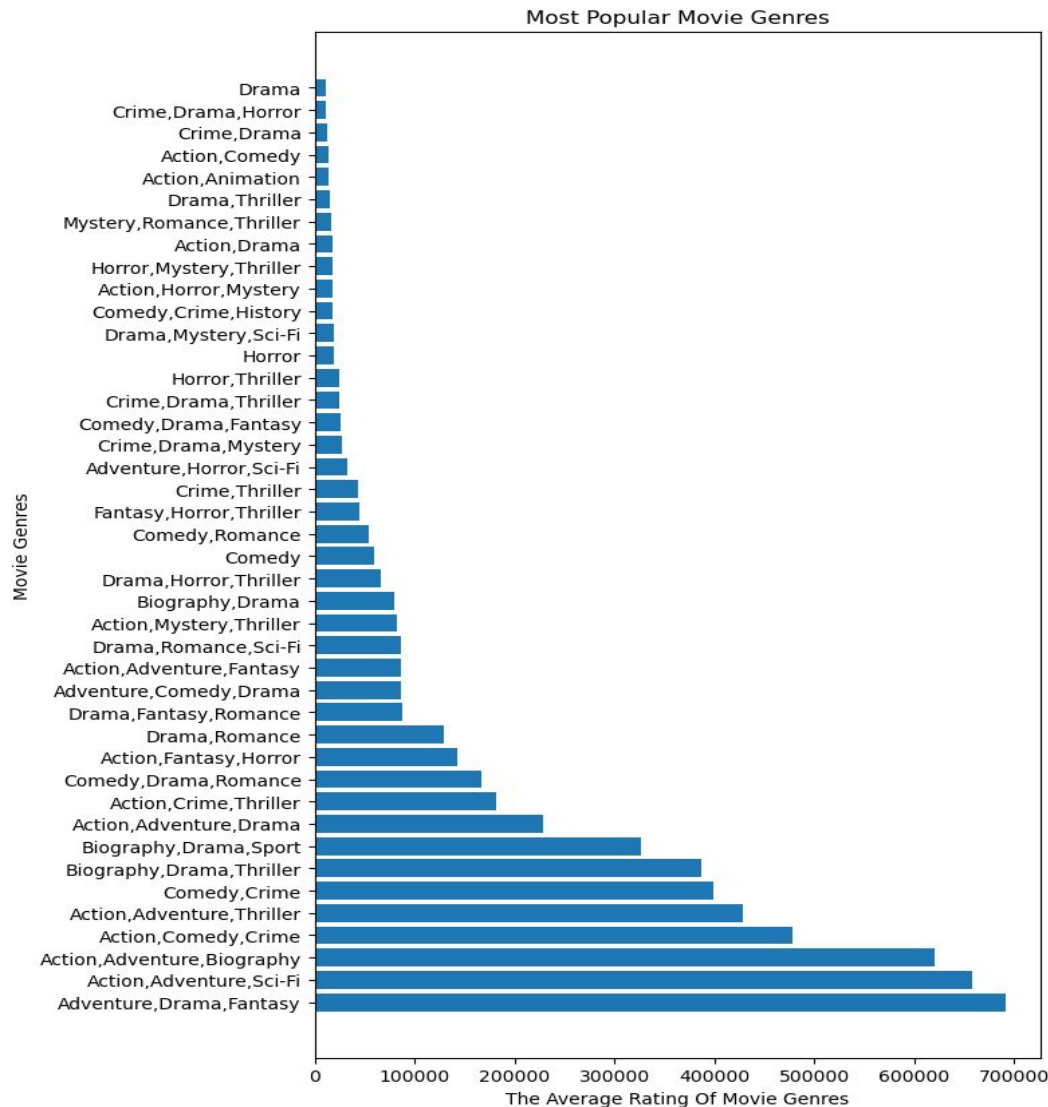
- As the dataset was huge i sampled the first 1000 rows
- I found out that Drama, Documentary, Comedy, Horror and thriller in that order we the most popular individually.
- However, movies that combined eg (adventure, drama,fantasy), (action, adventure, scifi) got the highest ratings
- The ratings showed that people prefer movies with a longer runtime to shorter runtime
- Movies That made the most money domestically and foreign had high views

# Findings On Our Most Popular genres



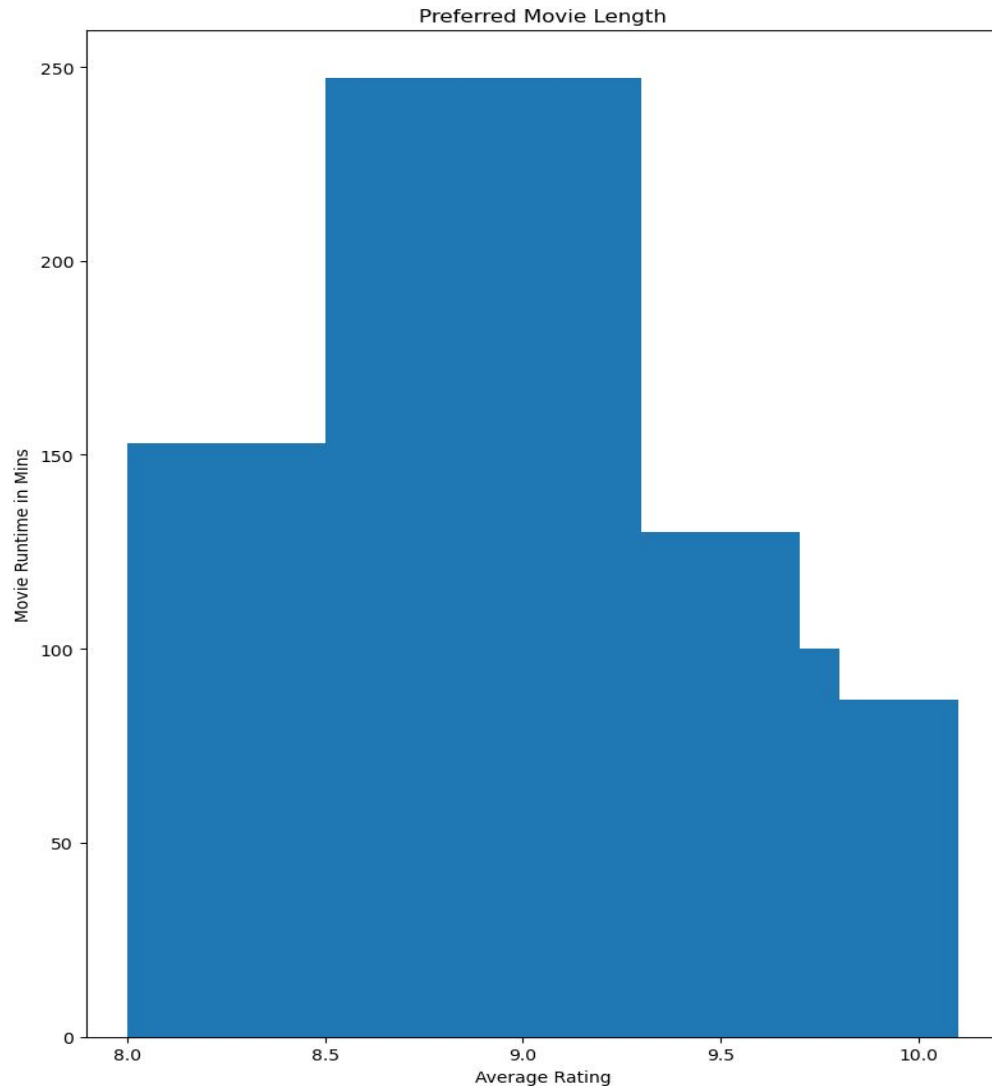
- I plotted a bar graph containing the frequency each movie genre and got the top five movie genres being: Drama, Documentary, Comedy, Horror and Thriller respectively.
- To be noted though, movies usually cover more than one genre, which is evidenced by our data.
-

# Movie Genres Against Movie Genres



- I then plotted the themes of the movies genres against average rating to find out if there was a preference from viewers to various movie themes. To be noted some movies also featured more than one genre but i treated them as a single unit.
- Movies with a combination of (adventure,drama,fantasy) turned out to be the most popular according to average findings.

# Findings on Runtime

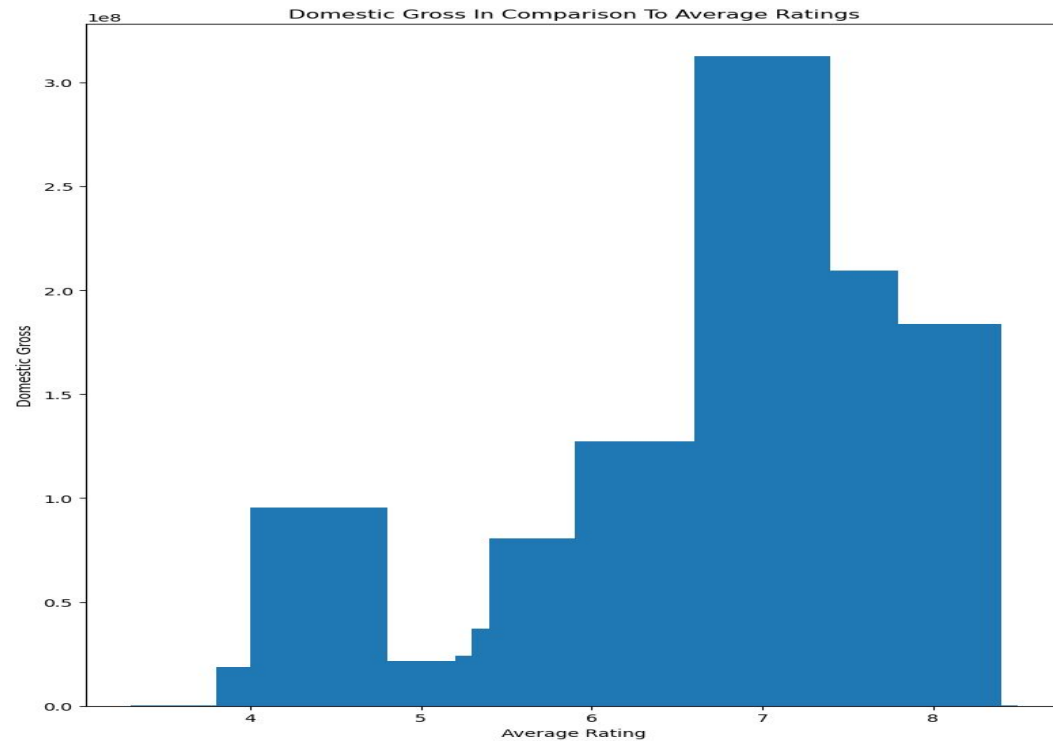


- I investigated a movies runtime against average rating to find consumer preference.
- There was a higher preference for movies to be longer than there was for shorter movies.

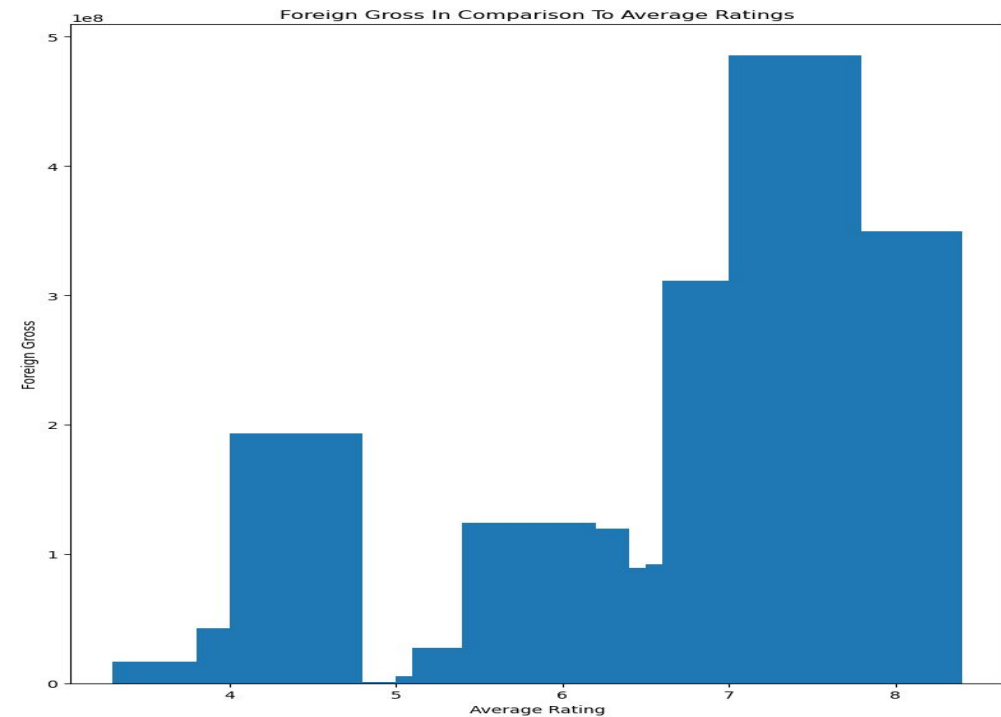


# Findings on The effect of ratings on movie income

## Domestic\_Gross



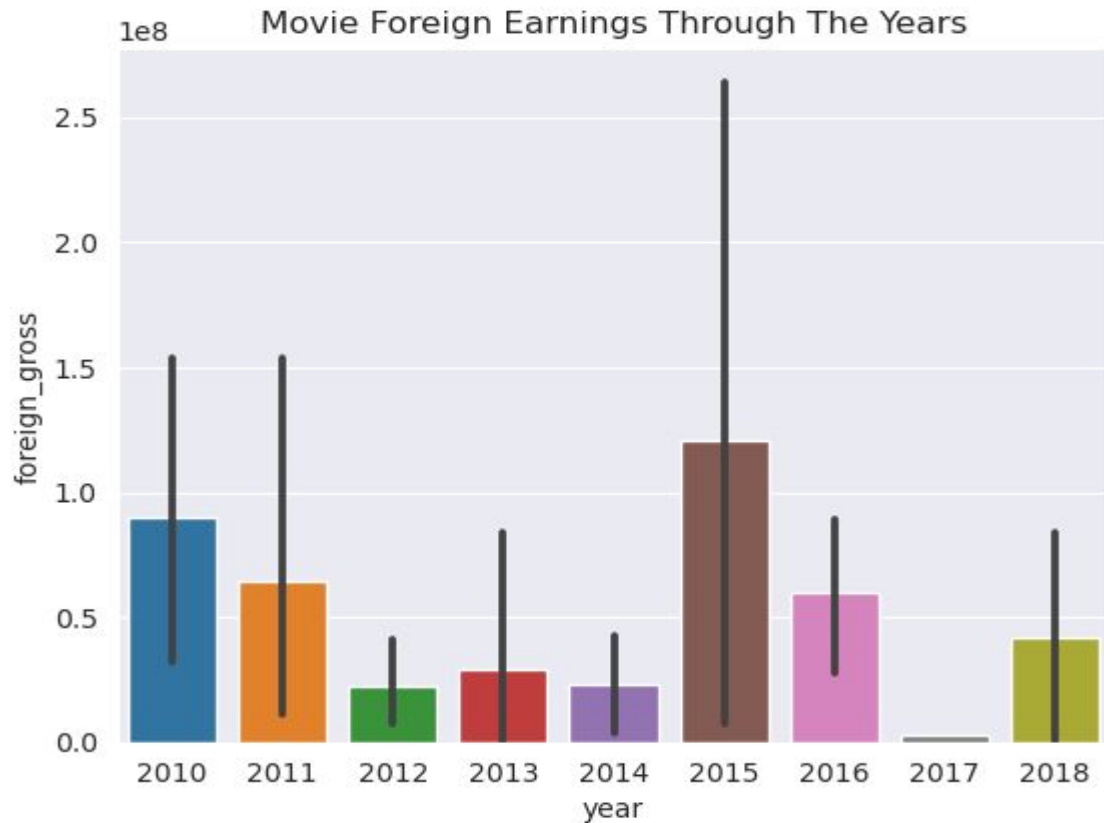
## Foreign\_Gross



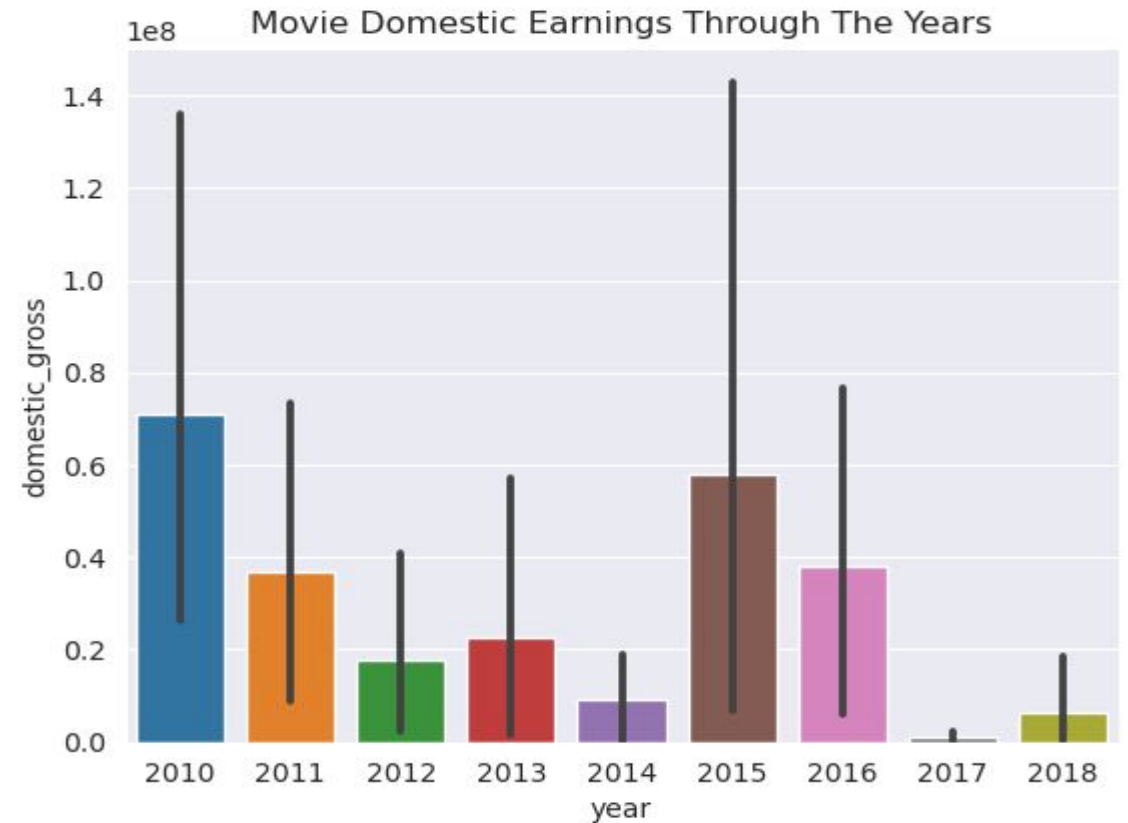
In both income increased proportionally with average ratings

# Movie Earnings Through The Years

## Foreign\_Gross



## Domestic\_Gross



Movie earning slowed down in the latter years and more investigation is required to understand why?

# Recommendations.

Based on the findings. Below are the things I would recommend:

- Microsoft Studio should be customer oriented and make movies from the most popular genres eg Drama Movies and a combination of top genres
- Microsoft should make longer movies in comparison to shorter movies
- Microsoft Studios should do a deeper study on why through the years movie earnings slowed down