

# **Box Office Trends and Insights: Microsoft New Movie Studio**

AUTHOR:KHADIJA ALI

July 2023

# Project Outline

- ▶ Introduction
- ▶ Your Task
- ▶ Business Understanding
- ▶ Data Understanding
- ▶ Methods
- ▶ Answering the business questions
- ▶ Summary
- ▶ Recommendations

# Introduction

In this analysis for Microsoft, we will make use of everything we have learned about pandas, data cleaning, and exploratory data analysis.

In order to complete this analysis, you will have to import, clean, combine, reshape and visualize data to answer the questions provided. Even as a junior data scientist, you can produce interesting analyses by combining multiple datasets . You will get a chance to practice all of these skills with the movie datasets which contains information about movies.

# Your Task : Analyzing movies datasets to provide insights for Microsoft New Movie Studio



# Business Understanding

You are charged with exploring what types of films are currently doing the best at the box office. You must then translate those findings into actionable insights that the head of Microsoft's new movie studio can use to help decide what type of films to create.

The business questions to be answered are:

- ▶ Analyzing the distribution of genres in movies dataset and the mean runtime minutes by genre to find out how long most movies run.

- ▶ Analyzing the distribution of genres in movies dataset and the mean runtime minutes by genre to find out how long most movies run.
- ▶ Analyzing the distribution of genres in movies dataset and the mean runtime minutes by genre to find out how long most movies run.
- ▶ Analyzing the distribution of genres in movies dataset and the mean runtime minutes by genre to find out how long most movies run.

# Data Understanding

The data sets offered contain a wide range of data essential to understanding our business concerns. There are three selected datasets.

The data is contained in four separate CSV files:

- ▶ `imdb.title.basics.csv`: which provides information about movies that emphasize various genres,titles,start years and their running times.
- ▶ `imdb.title.ratings.csv`: which provides information about average rating of movies and their number of votes

- ▶ bom.moviegross.csv: that highlights the box office earnings of movies over a particular period of time.
- ▶ tn.moviebudgets.csv: that highlights the production budgets and domestic, worldwide profits.

# Methods

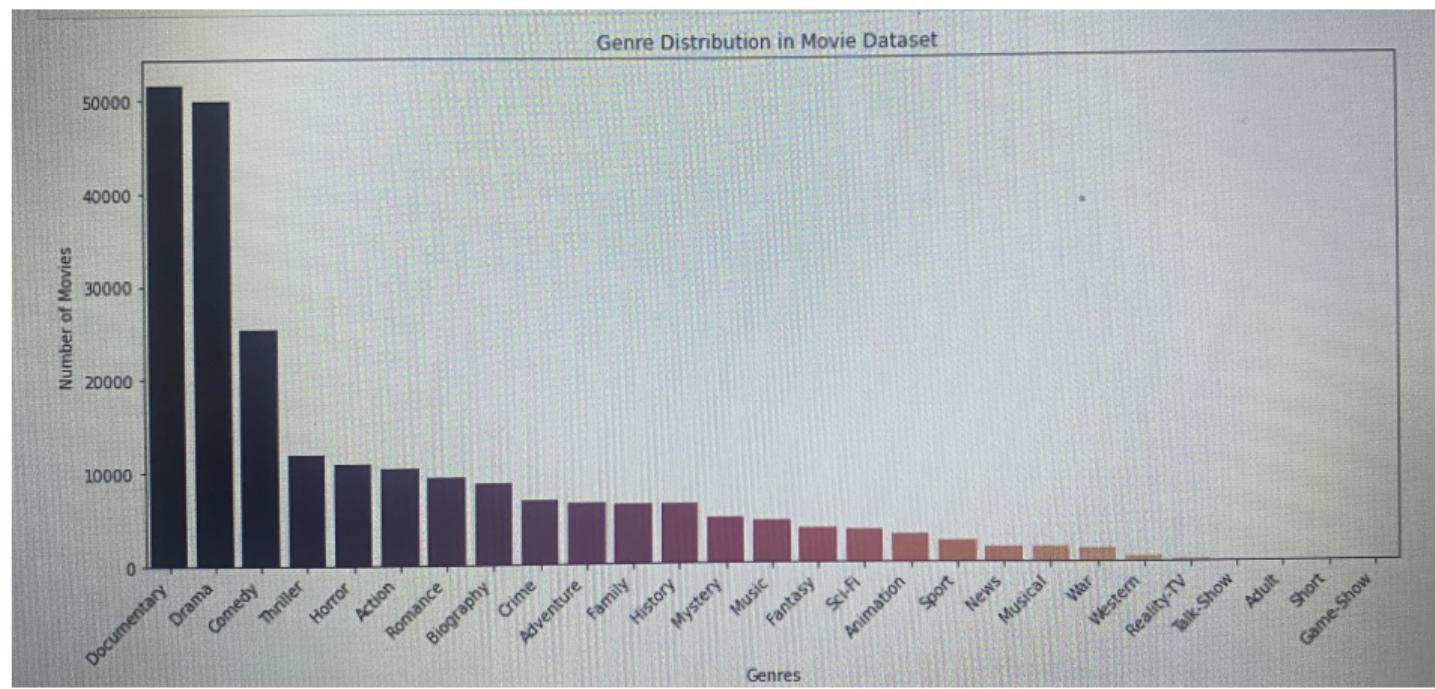
In this analysis the required libraries were imported the content on the four csv files was loaded separately.

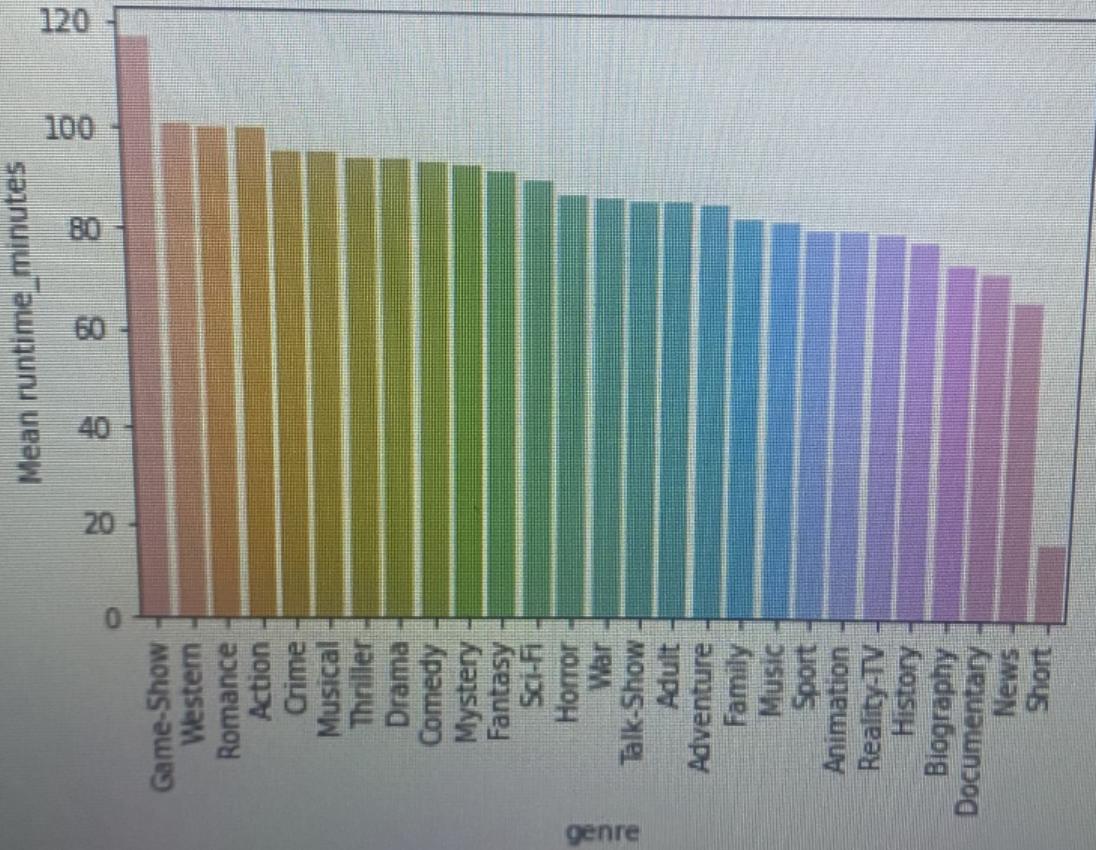
Data cleaning was done which involved dealing with missing values,outliers and duplicates.The datatypes were also changed to the correct form.

Merging of dataframes ,grouping and creation of new columns was done where there it was necessary.

The data was visualized using visualization tools such as bar graphs and scatterplots

Analyzing the distribution of genres in movies dataset and the mean runtime minutes by genre to find out how long most movies run.

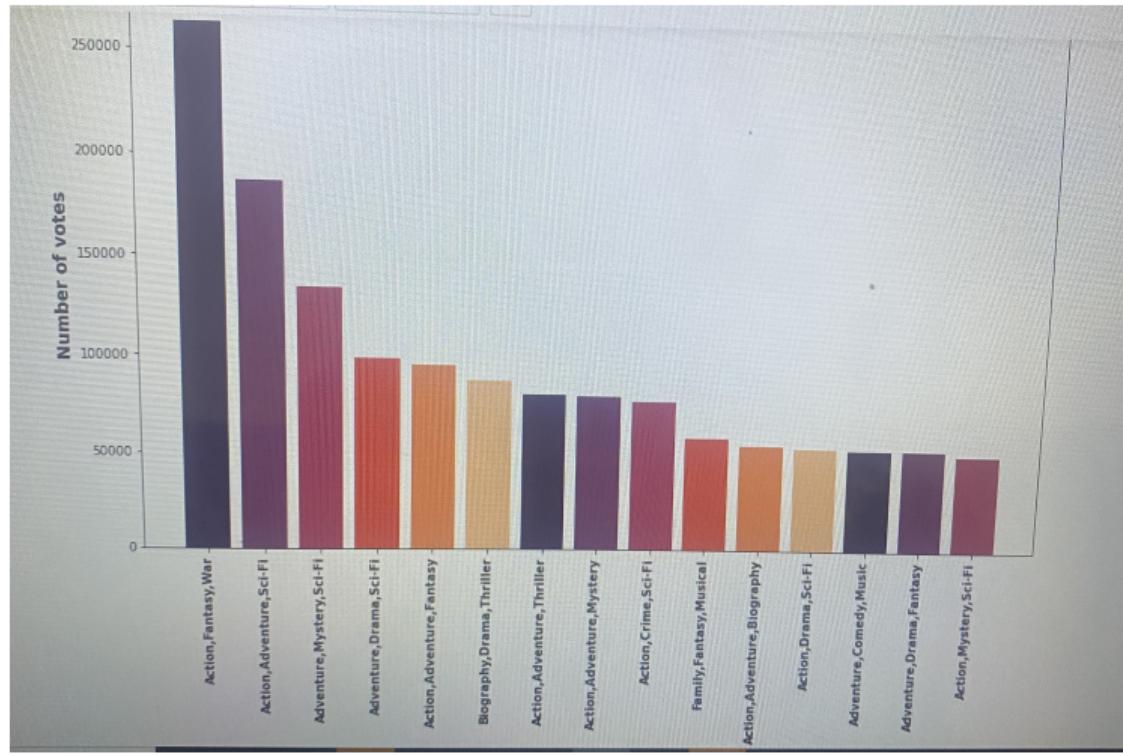




Based on the analysis of genre distribution in the movie dataset we see that Documentary,Drama,Comedy,Thriller and horror have the highest number of movies,Microsoft should consider investing in creating films in these genres.

Based on the analysis of the mean runtimes by genres we see that most of these genres have a runtime of 80 to 100 minutes,Microsoft should consider investing in films that have a runtime of 80-100 minutes.

# Analyzing the most popular and profitable genres to invest in.



The top 3 most popular genres combination according to the number of votes are:

- ▶ (Action,Fantasy,War),
- ▶ (Action,Adventure,SciFi)
- ▶ (Adventure,Mystery, SciFi).

The top three genre combinations with the highest ratings are:

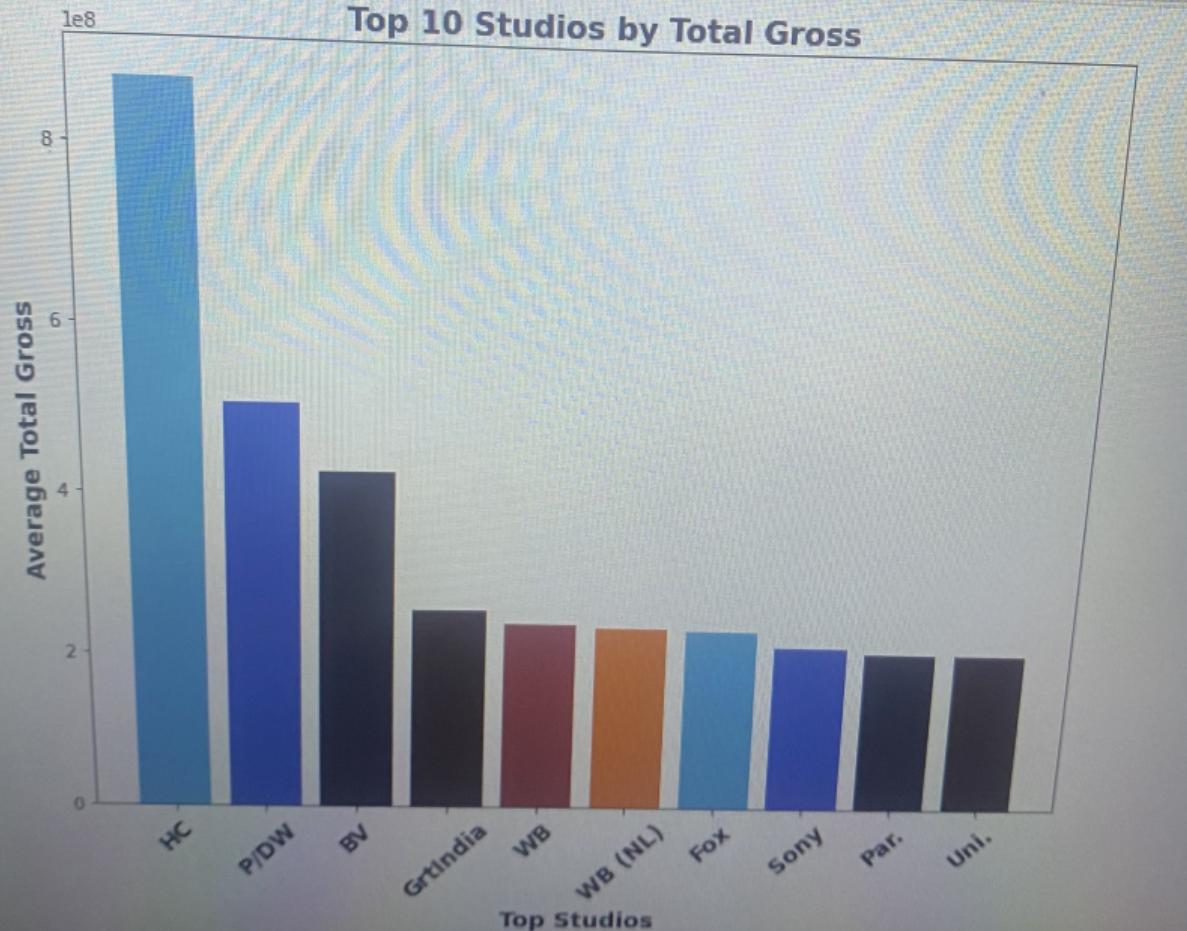
- ▶ (Comedy,Documentary,Fantasy)
- ▶ (Documentary,Family,Medical)
- ▶ (History,Sport).

Microsoft should consider creating films that combine elements from these genres. By creating films that combine elements from popular genres , Microsoft can appeal to a wide audience and increase their chances of success at the Box office.

# Analyzing studios with the highest average total gross.

- ▶ From the graph below, HC and P/DW are the top 2 studios with the highest average total gross.
- ▶ Also, Par. and Uni. are the bottom 2 studios with the lowest average total gross hence the least market share.

Microsoft should consider partnering with HC and P/DW to create high grossing films. By partnering with them, Microsoft can leverage their expertise in creating successful films that appeal to a wide audience increasing their success at the box office.



# Analyzing the relationship between total gross profit and the production budget.

Microsoft should consider investing in high budget films. There is a moderate positive correlation between production budget and total gross profit which suggests that in general as the production budget increases ,gross profit also tends to increase.

However, it is important to note that there is still a variability in the data and other factors might influence the relationship.



# Summary

From the above analysis of the four business problems, Microsoft may create a film production company that successfully competes with current studios and creates engaging material for audiences by investing in high budget films.

Partnering with HC and P/DW to create high grossing films, consider creating films that combine elements from the most popular genres, consider investing in films that have a runtime of 80-100 minutes.

# Recommendations

Aside from the analysis other recommendations i would give to Microsoft new movie studio include creating movies with strong character development with high quality content.

Microsoft should consider partnering with well known directors and actors to help bring their movies to life and create a distinct brand and strategy.