

# Microsoft Movies Project

Nicholas Muchiri Kinyua

# Introduction

Three data sets were used to help the head of microsoft start out his movie ambitions

Box office mojo

The Numbers movie budget

Rotten Tomatoes movie info

# Problem Statement

Microsoft wants to get involved in the movie industry. Unfortunately, it has never ventured in this area before. So the head needs to have a clear understanding of how the movie industry is doing, which areas are performing the best, which are not. He needs to have insights and visualizations on how to go about it hence coming up with an appropriate decision which area to focus on and which one not to.

# Business understanding

The head of microsoft will have to answer some questions for him to decide which movies/genres he will keep a close eye on:

Which objectives do I want to achieve?

Which is my target audience?

How do I align my strategies to my objectives in order to achieve my goals?

By doing so, he will have an understanding of his or rather microsoft involvement in the industry, the strategy he will have to employ, the outcome to expect, and of course the challenges he may encounter

# Main Objective

**This project aims to help Microsoft's new movie studio make informed decisions about the types of films to create. By analyzing data from the film industry, insights and trends that will assist in identifying successful strategies for the studio can be uncovered.**

# Specific Objectives

1. **Identify the most successful movie genres that microsoft head will use to make a prompt decision on which type of movies to put more focus on**
2. **To visualize movie budgets (especially using the tn movie budgets file).**
3. **Understanding how different branches of the entire movie field affect the company reputation and sequentials for thaat matter. E.g., rating, genre, runtime.**

# Notebook Structure

**Introduction¶**  
**Problem Statement**  
**Main Objective**  
**Specific Objectives**  
**Notebook Structure**  
**Data cleaning**  
**Data visualizations**  
**Conclusions**  
**Recommendations**

# Data Understanding

First, the data was loaded using the `pd.read_csv()` method

For tsv files, `delimiter = "\t"`

A few ways to understand the data employed were:

- `.describe()`

- `.info()`

- `.head()`

- `.tail()`



# Data Cleaning

The following is the strategy used to clean the data:

Removing any duplicates

Filling missing data with other columns data, mean and median

Dropping off columns

Dropping off rows

Arranging messy columns

# Methodolgy

## Data transformation

Further cleaning and data processing was done:  
bom.movie\_gross was merged with rt\_movie info

This led to data understanding:

.describe()

.info()

Which led to more data cleaning steps:

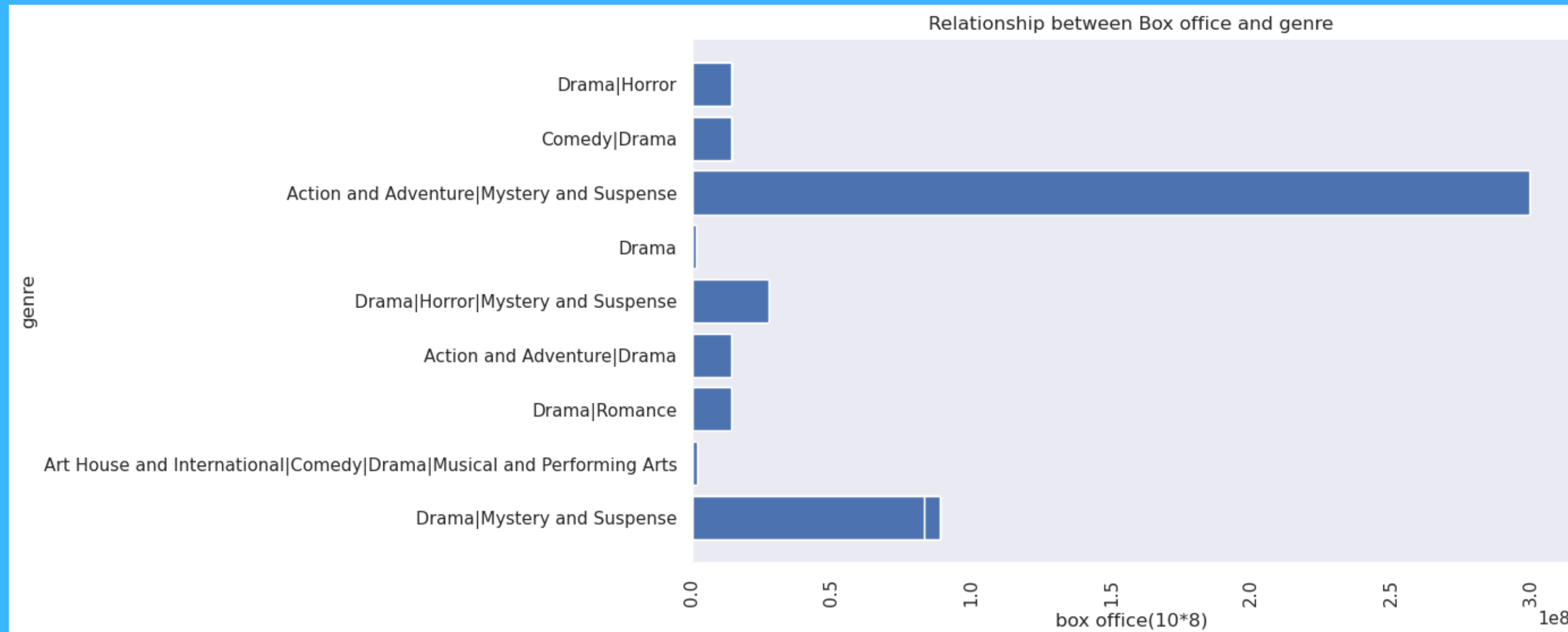
Removing any duplicates

domestic gross was combined with foreign gross to become total gross

domestic gross and foreign gross were subsequently dropped

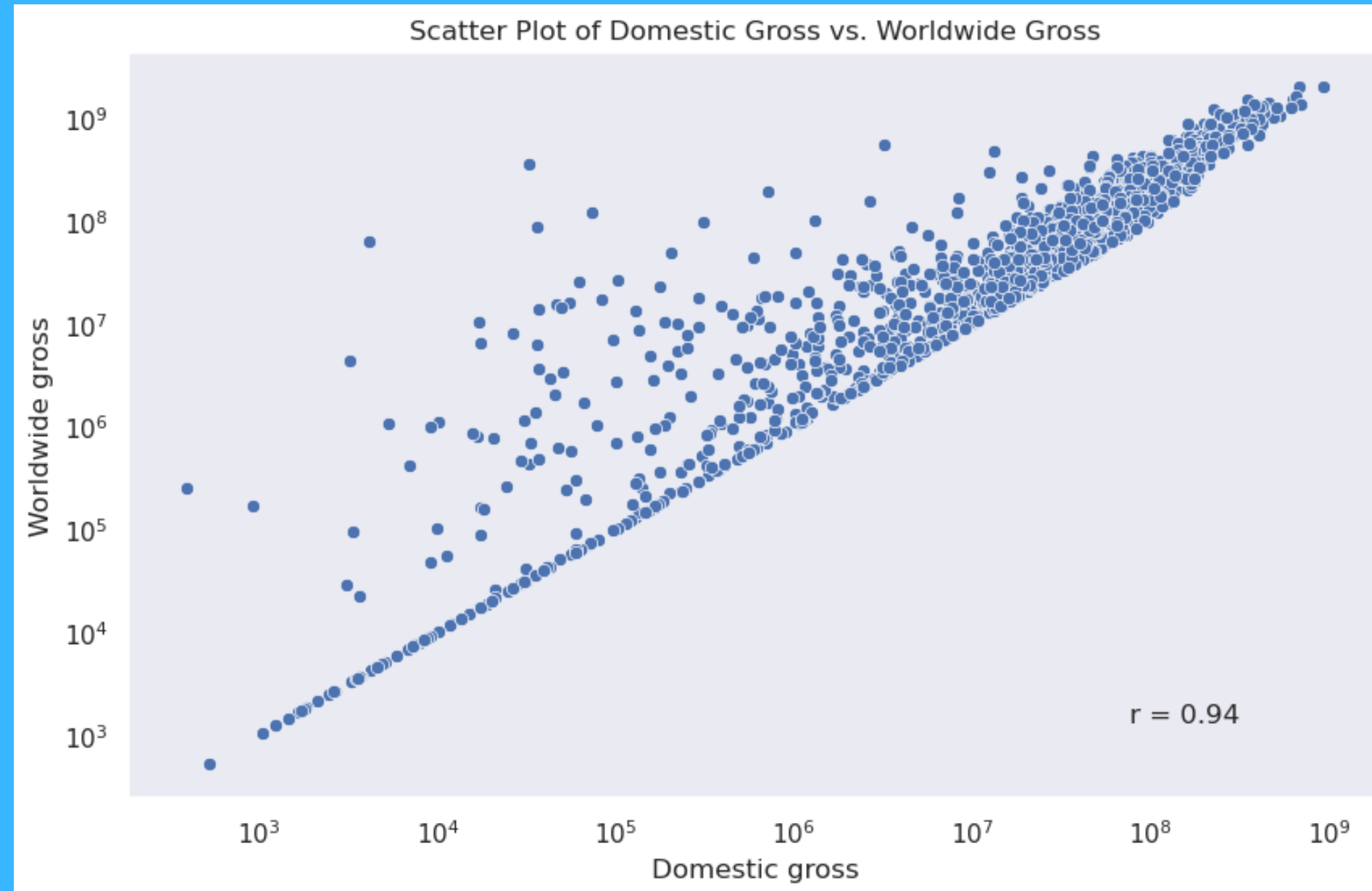
# Data Visualizations

A horizontal bar graph to show the relationship between box office and genre



# Data Visualizations

A scatter plot to depict how domestic gross is comparable to worldwide gross



# Conclusions

**The head of microsoft should consider checking into the data visualizations carefully in order to know where to put more effort and focus on while discarding the other areas.**

**He should only invest on the movies which have the highest number of views. That's why Drama views barely make it in the least (having a total domestic and worldwide gross of less than 5 million dollars).**

**Certainly, there seems to have no significance difference from the domestic market to the worldwide market. That shouldn't be a factor of determining the types of movies to focus on .**

# Recommendations

- 1 a). The head of microsoft should consider venturing into Drama|Mystery and suspense since they have the highest gross income.  
b). The head of microsoft should combine that genre with Action and adventure|Mystery and suspense since they have the highest box office.
2. He should consider investing heavily in production to get a high gross profit both domestic and worldwide.
3. Hes hould never consider to produce PG movies since they have the lowest rating and the lowest total gross too.

# References

bom\_movie gross file

rt.movie\_info file

tn.movie\_budgets file