

# Movie Analysis For Microsoft

Dhruv Rangunathan

# Introduction

- Microsoft is investing in a new movie studio.
- They want a data-driven approach to determine what factors matter most when producing movies.
- This analysis seeks to find variables that correlate with high-performing movies in the box office.
- Actionable insights stemming from the analysis will be discussed.

# Business Context

- In this analysis, we will measure the effect of ratings, runtime, genres, and role on movie success.
- Return-on-Invest (ROI) is used to measure the success of a movie in the box office.
- Movie studios typically see half of the box office gross.
- Therefore, our ROI calculation is half of the gross divided by the production cost.

# Data

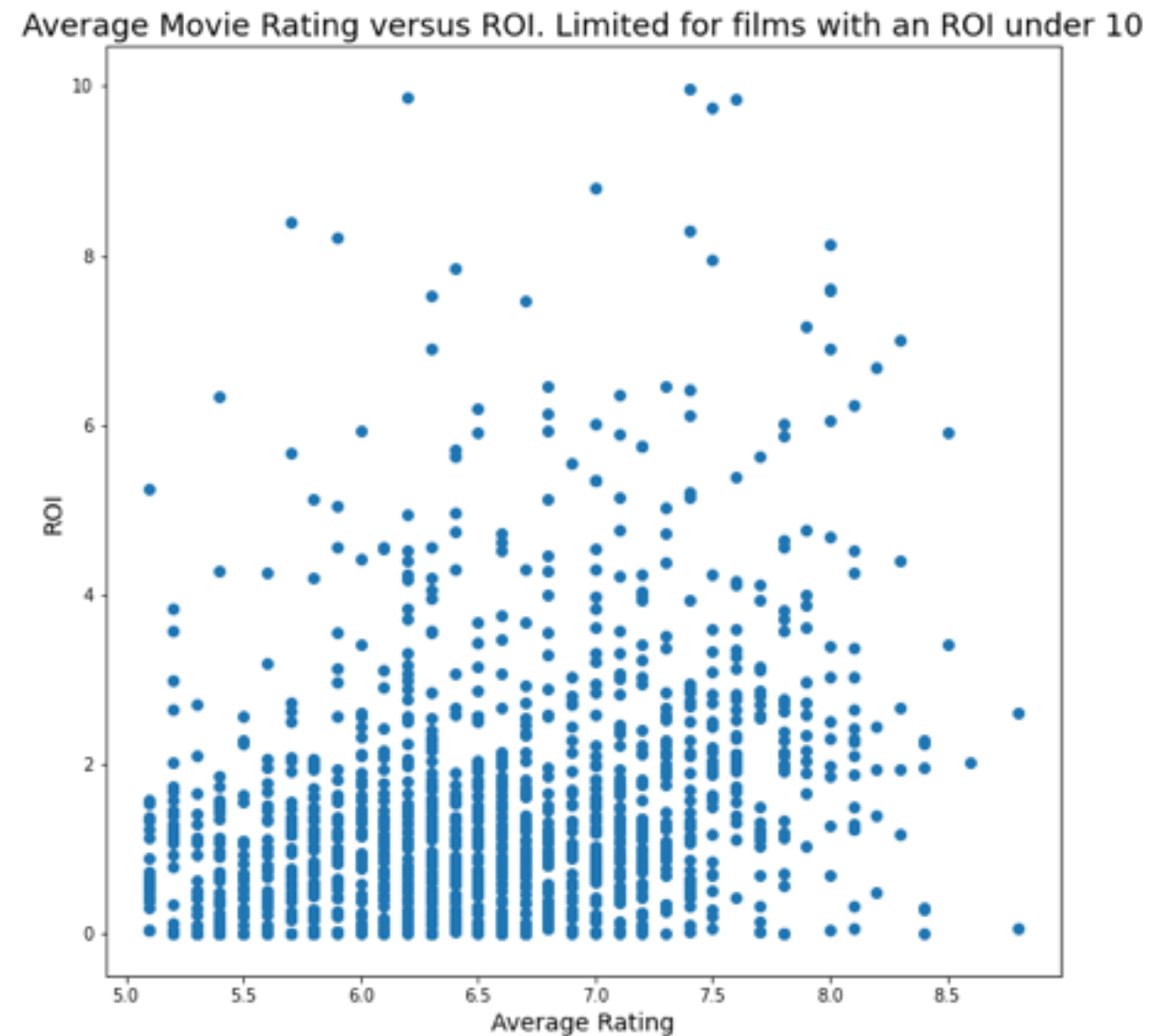
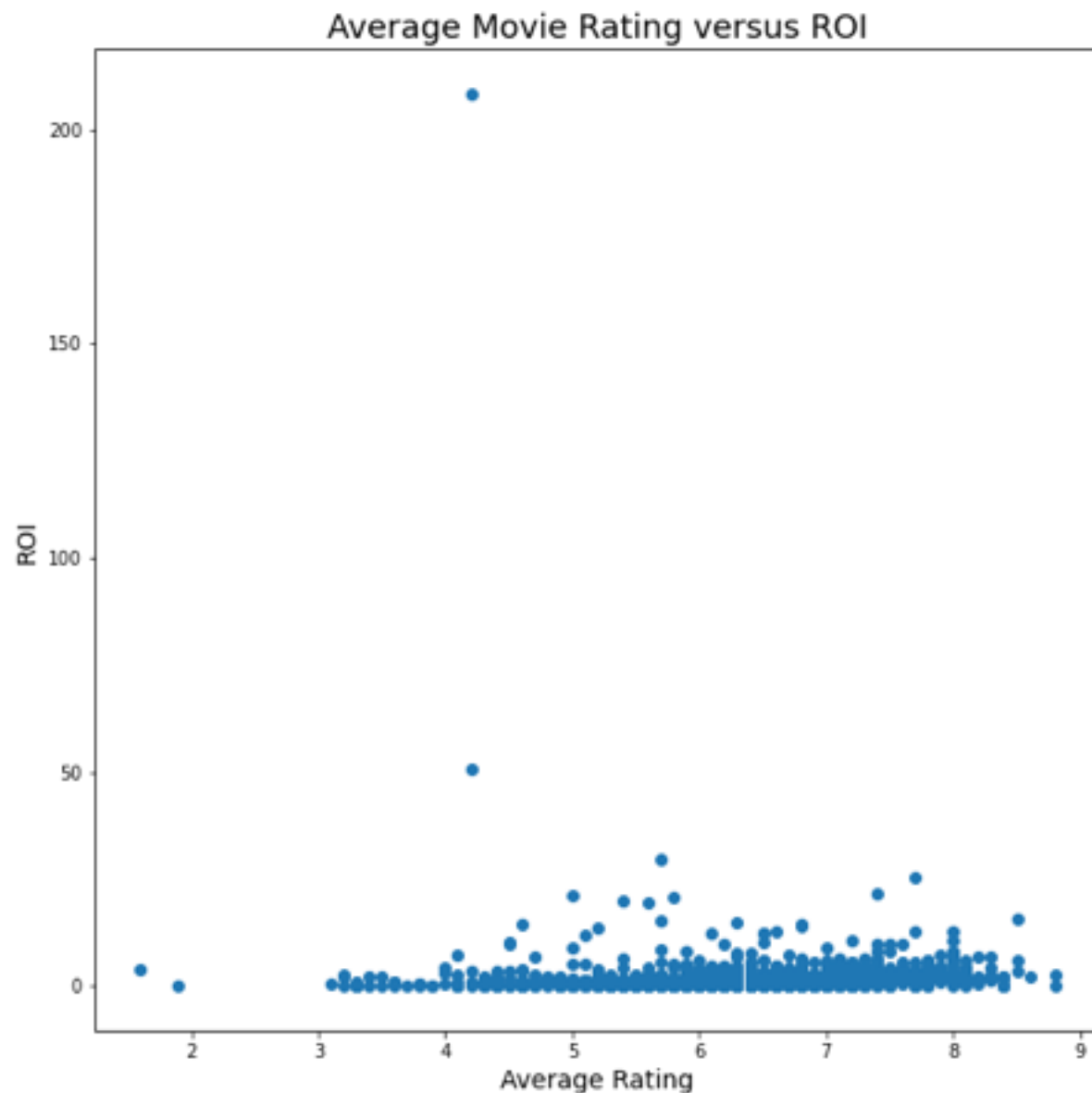
- Data from the following sources were used in the analysis.
  - The Numbers: Data stored as a CSV.
  - IMDb: Data stored in a SQL database.
- The Numbers: Used to get data on the production budgets and worldwide gross.
- IMDb: Database containing information on ratings, runtime, genres, actors and directors associated with movies.

# Process Steps

- Import data from The Numbers into data frame
- Clean data by resolving missing data issues and data type conversion of string data to numerical.
- Calculate the ROI for each movie in the data frame
- Import data from IMDb SQL database.
- Clean data by resolving any missing data issues and store in a data frame for further analysis.
- Join data frames to conduct further data analysis.

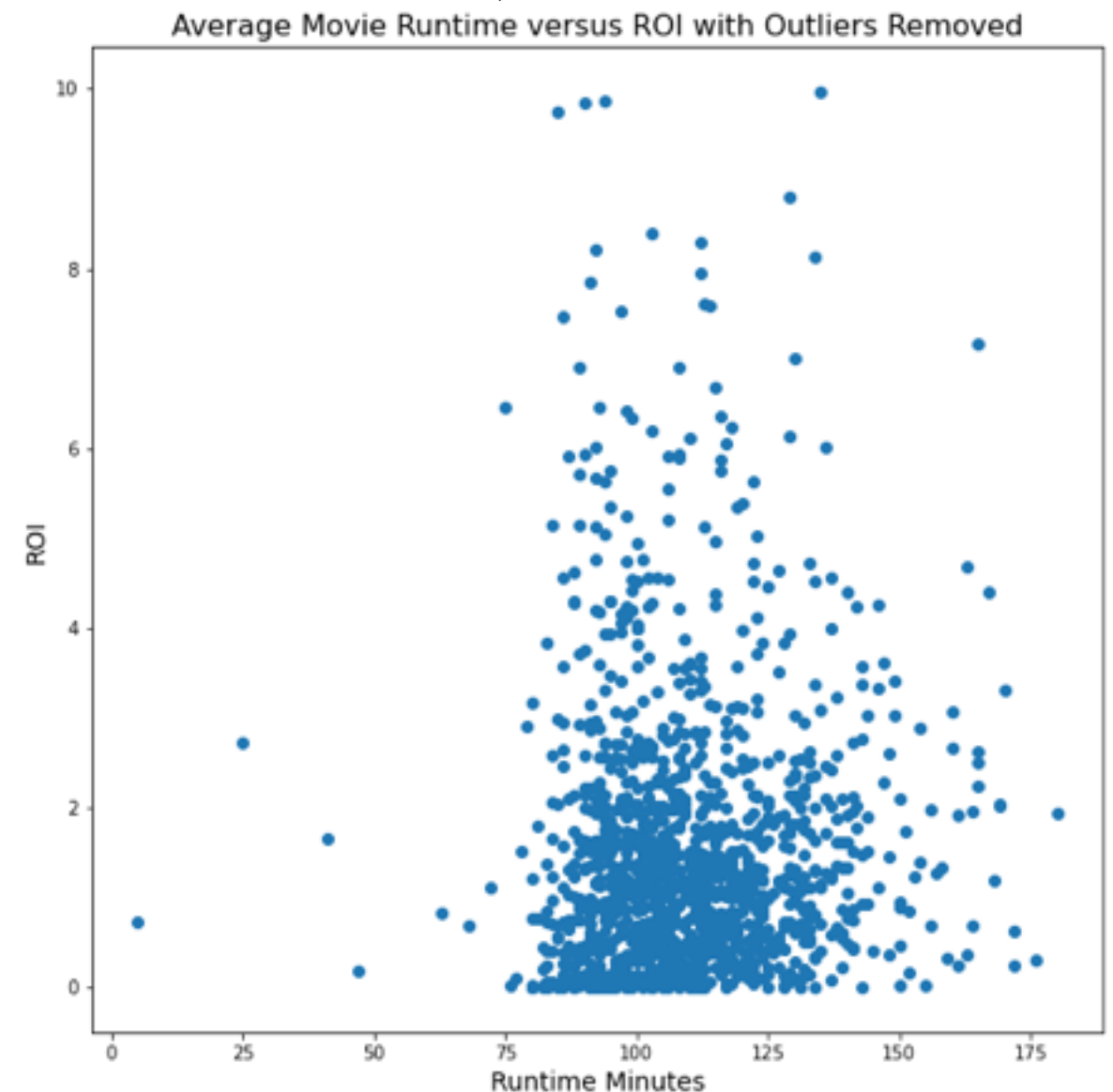
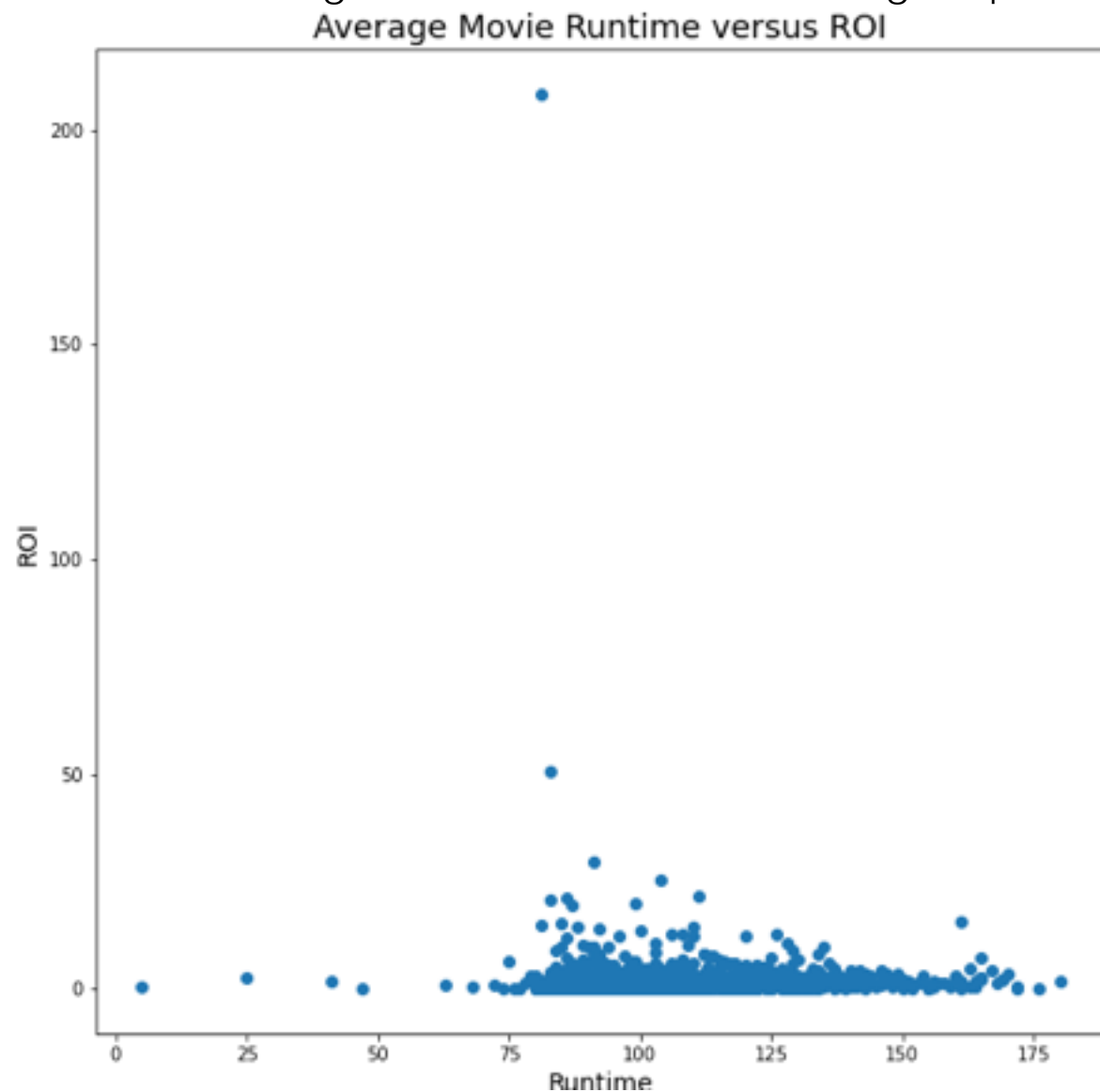
# Results - Effect of Ratings on ROI

- No correlation found between movie rating and ROI.
- Correlation Coefficient is 0.002.
- Removing the outliers in the ROI/ratings improves the correlation coefficient to 0.28, suggesting a weak correlation.



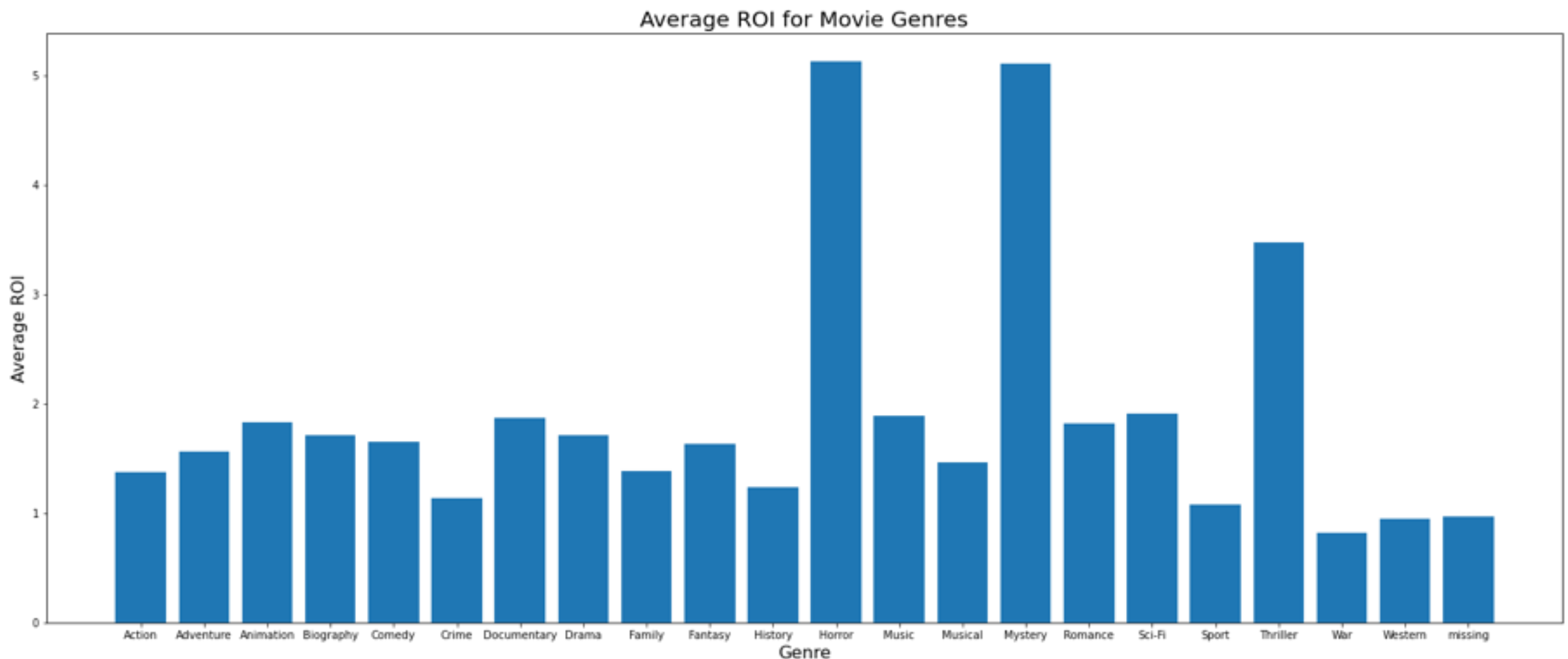
# Results - Effect of runtime on ROI

- No correlation found between runtime and ROI.
- Correlation Coefficient is 0.005.
- Removing the outliers in the ROI/ratings improves the correlation coefficient to 0.05, which is still a weak correlation.



# Results - Genres with the highest ROI

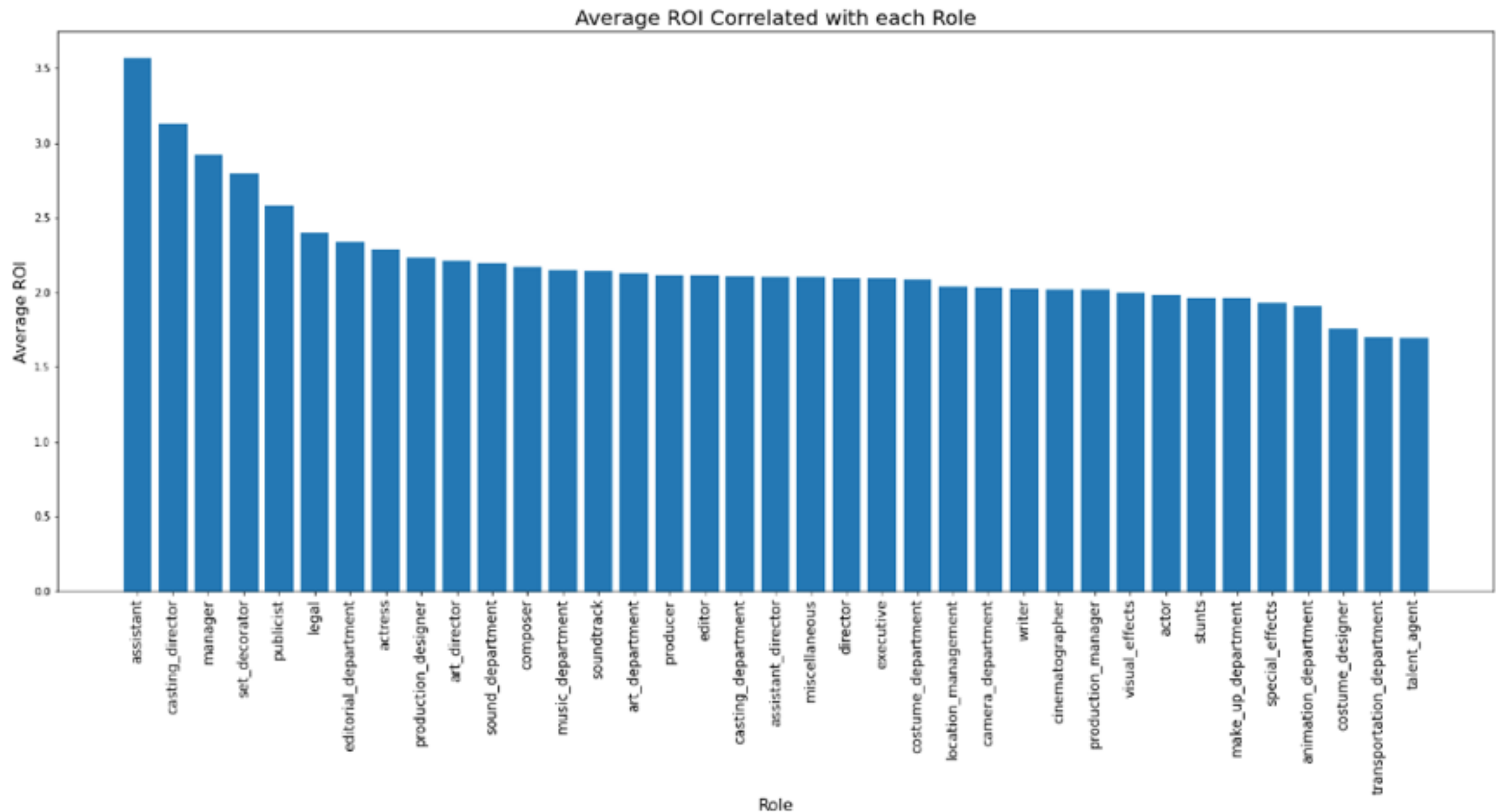
- We grouped each genre by ROI.
- We aggregated the numbers by getting the mean of the ROI across each genre.
- The genres that had the best ROI were horror, mystery, and thriller.





# Results - Roles Correlated with ROI

- We calculated the aggregate average ROI by role.
- The roles that had the best ROI on average were assistant, casting directors, and managers.



# Business Applications

- Initially invest in the horror, thriller, and mystery genres as they have the highest ROI on average.
- Since there is no correlation between movie runtime and ROI, Microsoft may not want to producer longer and more expensive movies.
- Since there is no correlation or a weak correlation between movie ratings and ROI, Microsoft does not need to invest time and energy investing in pleasing the critics.
- Microsoft should focus on hiring quality assistants, casting directors, and managers since these roles seem to be correlated with higher ROIs.

# Future Improvement Ideas

- Limit the movie timeframe: For example, movies that were successful in the 40s were popular with an audience that isn't the majority today.
- Examine any correlations between release months and ROI. This will help Microsoft determine what months work best for release dates.
- Examine if the rise streaming platforms have reduced the profitability of the traditional movie industry.

# CONTACT INFORMATION

- Name: Dhruv Rangunathan
- Business Email: dhruvrangunathan@gmail.com
- LinkedIn Profile: <https://www.linkedin.com/in/dhruv-rangunathan-908993b1/>
- GitHub Profile: <https://github.com/dragunat2016>