A Movie Analysis for Microsoft

By Benjamin Dean

Introduction

This project is intended to display an understanding of the material covered in Phase 1 of the Flatiron School's Part Time Data Science program. Students were given a scenario in which Microsoft Corporation has decided to start their own movie studio and has reached out for guidance. This project covers the processes of data exploration, cleaning, and modeling. It then provides recommendations for Microsoft based on this analysis.

Outline

- Business Problem
- Data Exploration
- Methods
- Results
- Conclusions

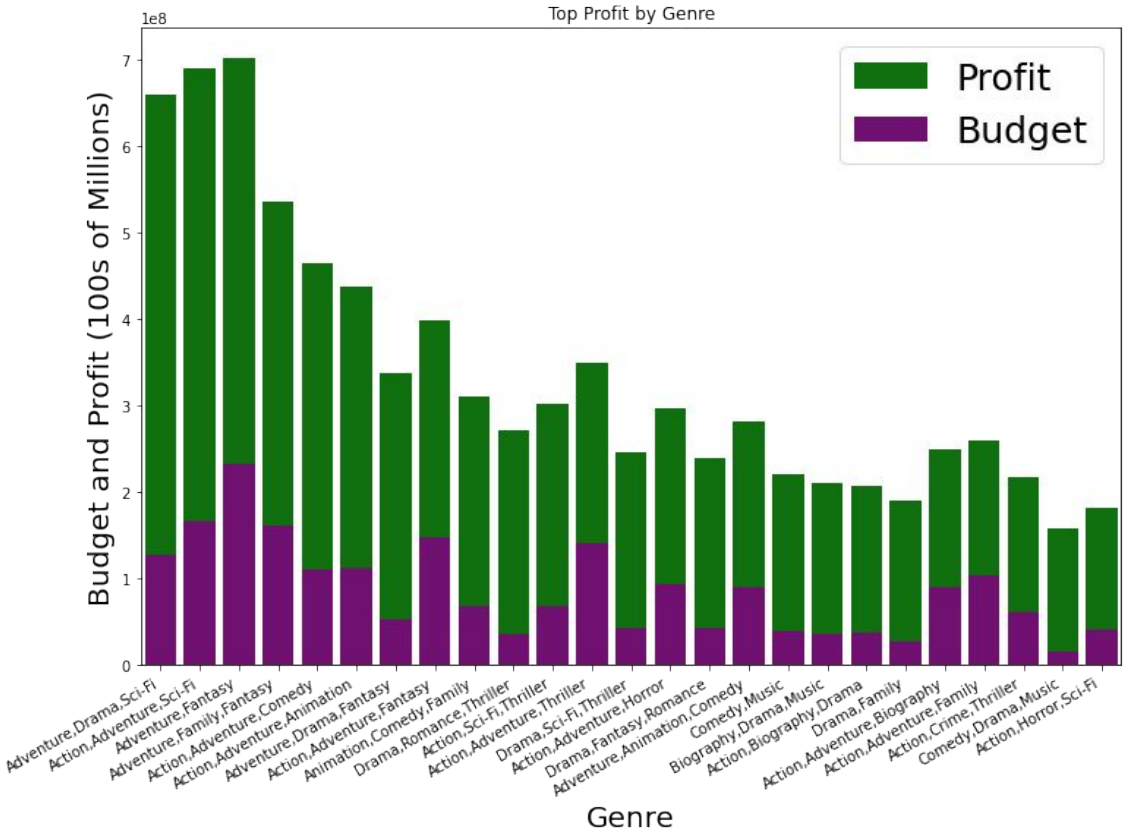
Data

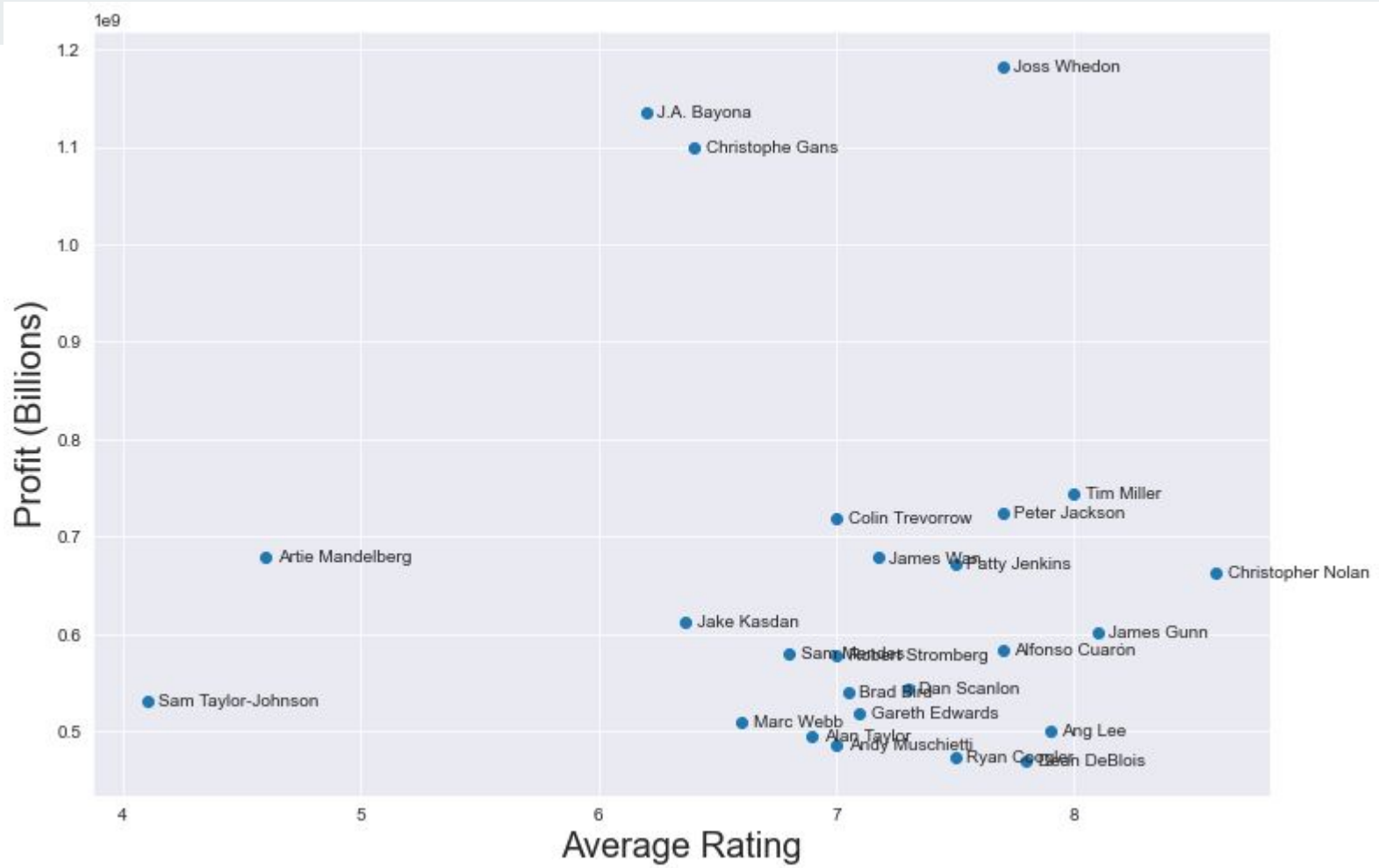
There were eleven datasets initially provided and for this project we focus on six of them from IMDB.com, TheMovieDb.org, and the-numbers.com. After determining the profit and ratings as the variables we wanted to target we selected these datasets because they contained data on production budgets, worldwide grosses, director names, movie titles, etc. We then created a profit column by subtracting the production budgets series from the worldwide gross series.

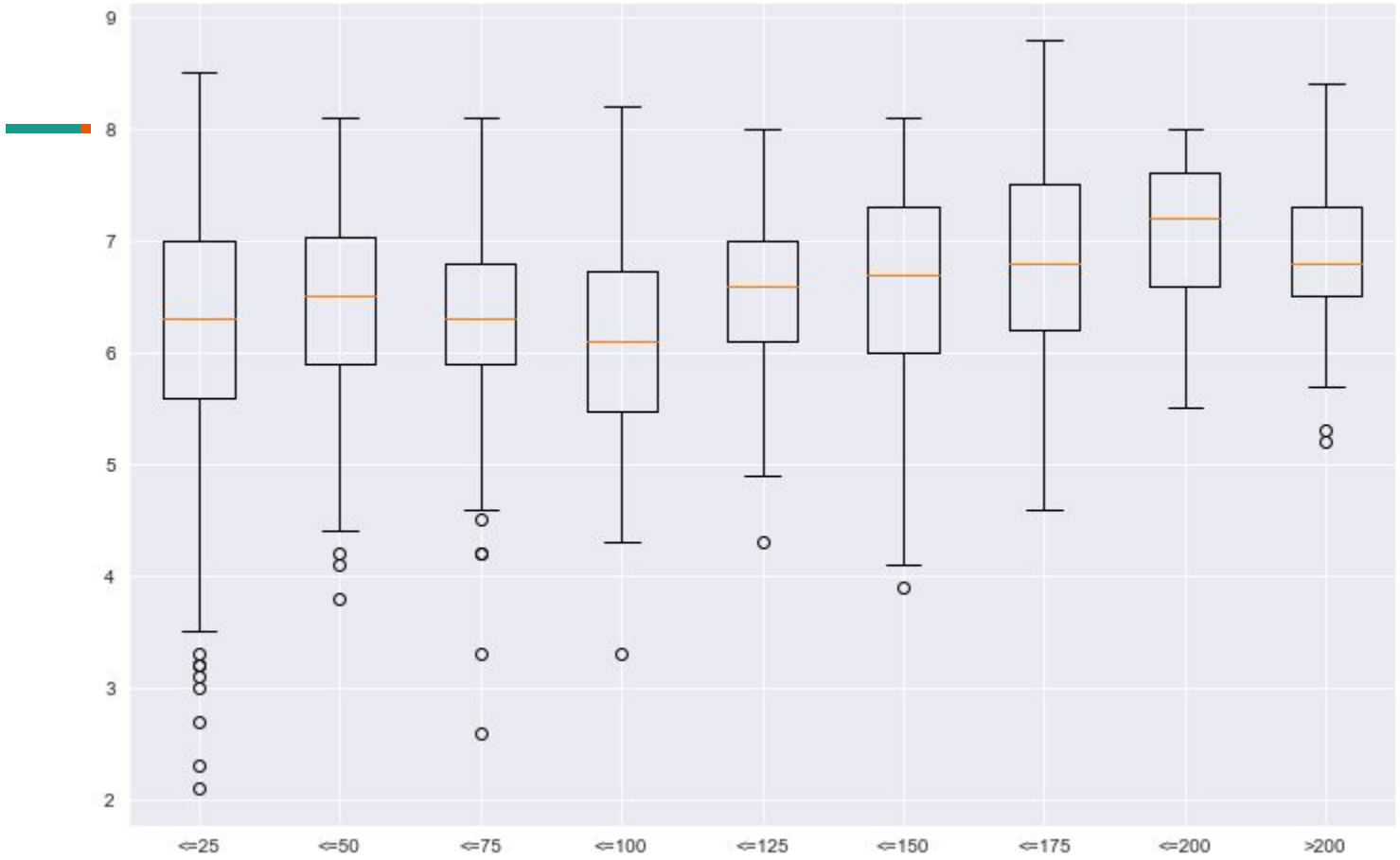
Methods

Once we had selected appropriate data it was time to prepare and model our graphs. We first cleaned the data by getting rid of columns and rows that we did not want. We also removed movie genres in which the count was only one. We then renamed a number of columns so that we could easily merge datasets. The first three datasets were merged on the tconst column, a column that associated movies with a code. The following three datasets were merged on the columns ‘nconst’, ‘primary_title’, and ‘primary_title’ and ‘release_date’.

Results


Top Profit by Genre
1e8
Budget and Profit (100s of Millions)
Profit
Budget
0
1
2
3
4
5
6
7
Adventure,Drama,Sci-Fi
Action,Adventure,Sci-Fi
Adventure,Fantasy
Adventure,Family,Fantasy
Action,Adventure,Comedy
Action,Adventure,Animation
Adventure,Drama,Fantasy
Action,Adventure,Fantasy
Animation,Comedy,Family
Drama,Romance,Thriller
Action,Sci-Fi,Thriller
Action,Adventure,Thriller
Drama,Sci-Fi,Thriller
Action,Adventure,Horror
Drama,Fantasy,Romance
Adventure,Animation,Comedy
Comedy,Music
Biography,Drama,Music
Action,Biography,Drama
Drama,Family
Action,Adventure,Biography
Action,Adventure,Family
Action,Crime,Thriller
Comedy,Drama,Music
Action,Horror,Sci-Fi
Genre


1e9
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
Profit (Billions)
4
5
6
7
8
Average Rating
Joss Whedon
J.A. Bayona
Christophe Gans
Tim Miller
Colin Trevorrow
Peter Jackson
Artie Mandelberg
James Wan
Patty Jenkins
Christopher Nolan
Jake Kasdan
James Gunn
Alfonso Cuarón
Dan Scanlon
Sam Taylor-Johnson
Gareth Edwards
Marc Webb
Ang Lee
Alan Taylor
Andy Muschietti
Ryan Coogler
Dean DeBlois


9
8
7
6
5
4
3
2
<=25
<=50
<=75
<=100
<=125
<=150
<=175
<=200
>200

Conclusion

Our analysis provides a number of suggestions for Microsoft regarding multiple aspects of film production. We conclude that Microsoft should begin picking up scripts within the genres action, adventure, fantasy, and sci-fi. We also advise they considering working with directors Joss Whedon and Christopher Nolan.