Microsoft Movie Studio Analysis

By Ashley Eakland

Summary

So you want to open a movie studio? GREAT!

My completed analysis will provide insight into rating importance, which genres you should focus on and just how long your movie runtimes should be.

Outline

- Business Problem
  - *My approach to your task*
- Data
  - *Introduction to the data*
- Methods
  - *How I prepared/analyzed data*
- Results
  - *The findings*
- Conclusions
  - *Recommendations*

Business Problem

So you want a piece of the movie-making pie, but you don't know where to start... Who could blame you!

Here are the things you should be considering:

- Genre
- Movie Length
- Ratings

Data

IMDB - well known source for film information

- Title Basics dataset of 146,144 titles
- Title Ratings dataset of 73,856 title numbers

Box Office Mojo - IMDB company, dedicated to tracking box office revenue in an algorithmic manner

- Gross Profit Data for 3387 titles both domestic and foreign markets

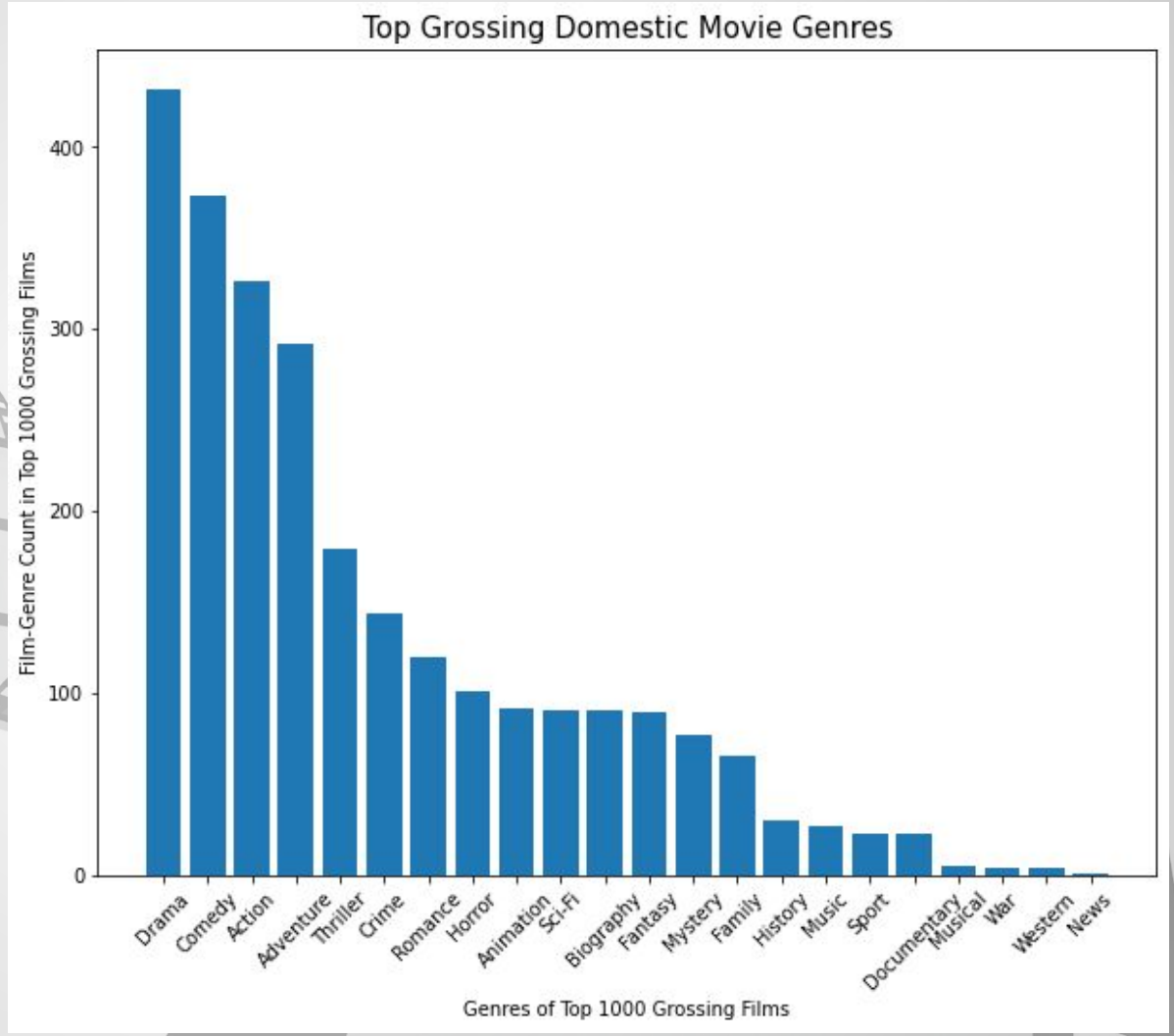
Methods

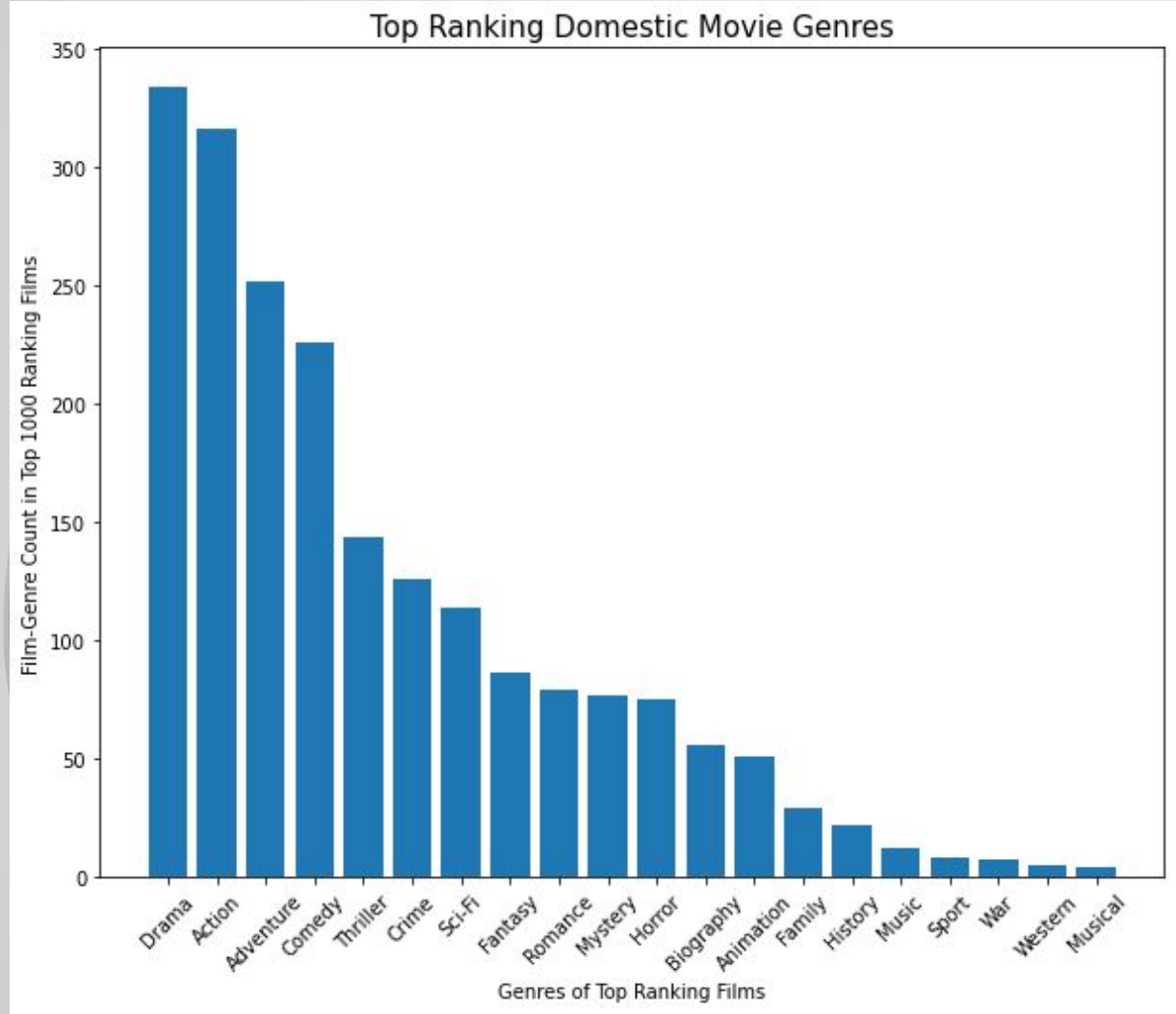
Using exploratory data analysis, statistics, and data visualization, I identified what I believe to be important factors in a films success rate.

Results

Genres

As shown here, and on the following slide, top performing genres include Adventure, Action, Comedy, Sci-Fi, Animation, Drama, Thriller, Fantasy, and Biography.





Results

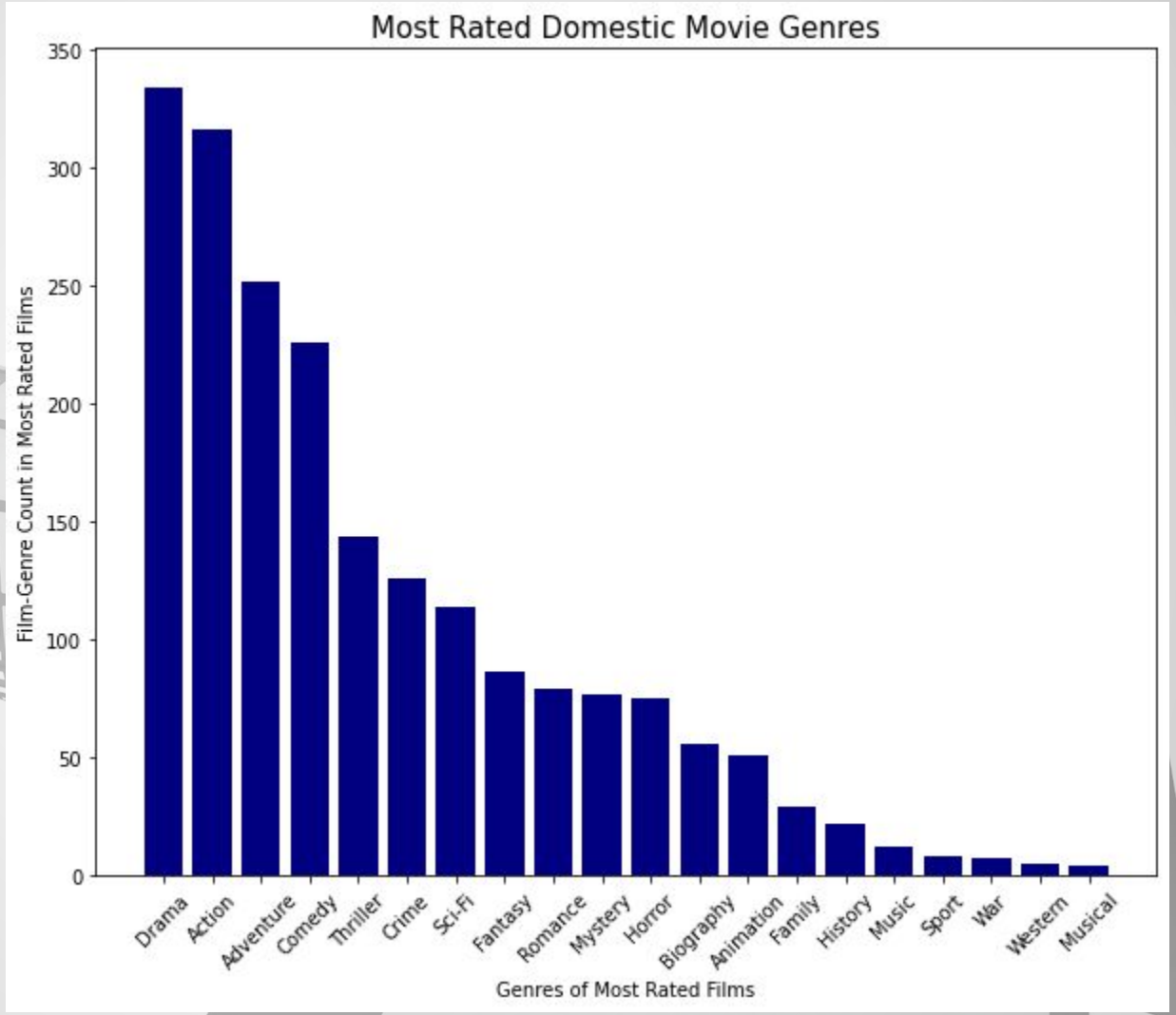
Genres cont.

Top Ranking genres closely match top grossing genres.

Results

Genres cont.

Most Rated Genres follow this trend with Drama, Comedy, Adventure, Action and Thriller rounding out the top 5 most ranked film genres.

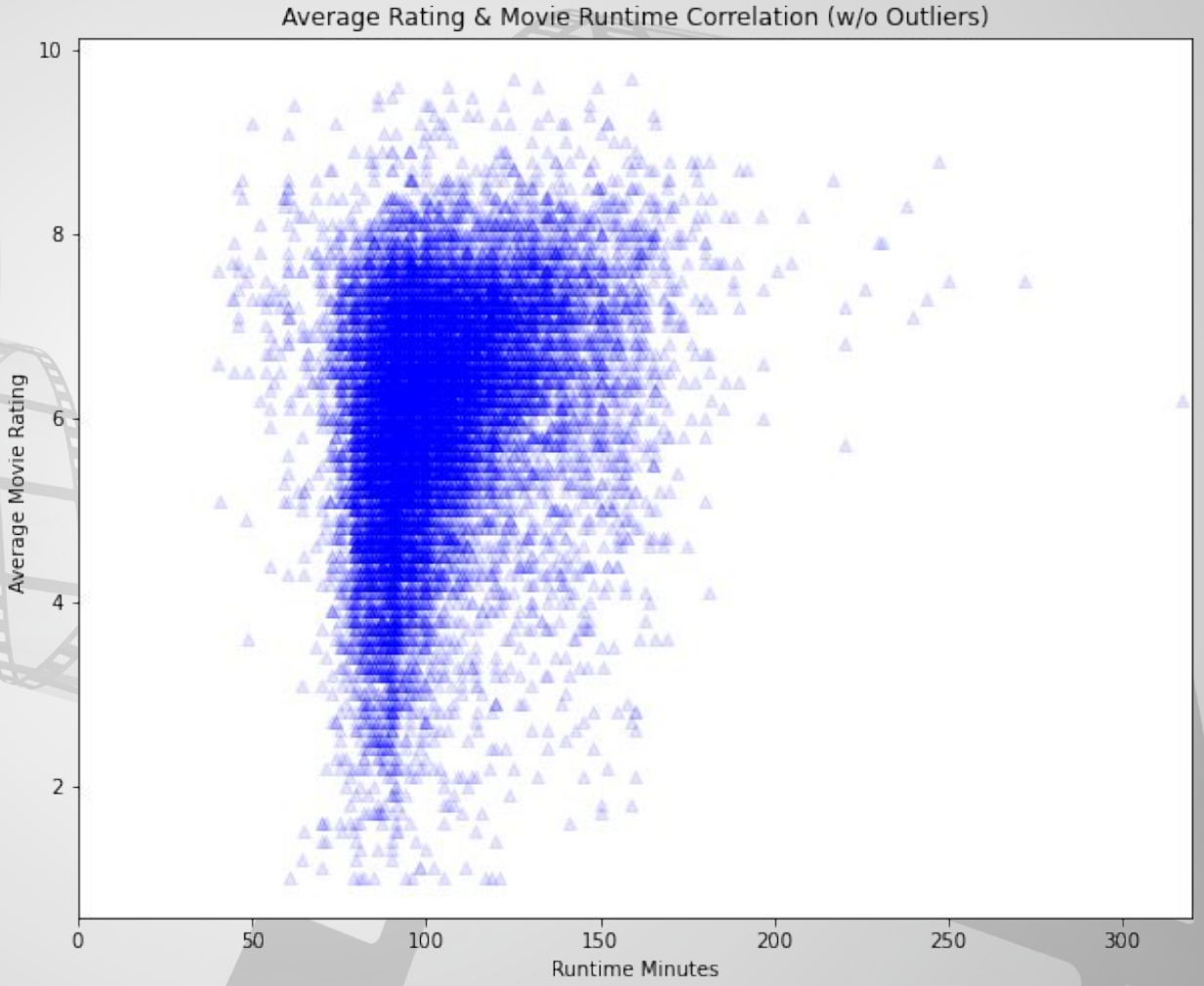


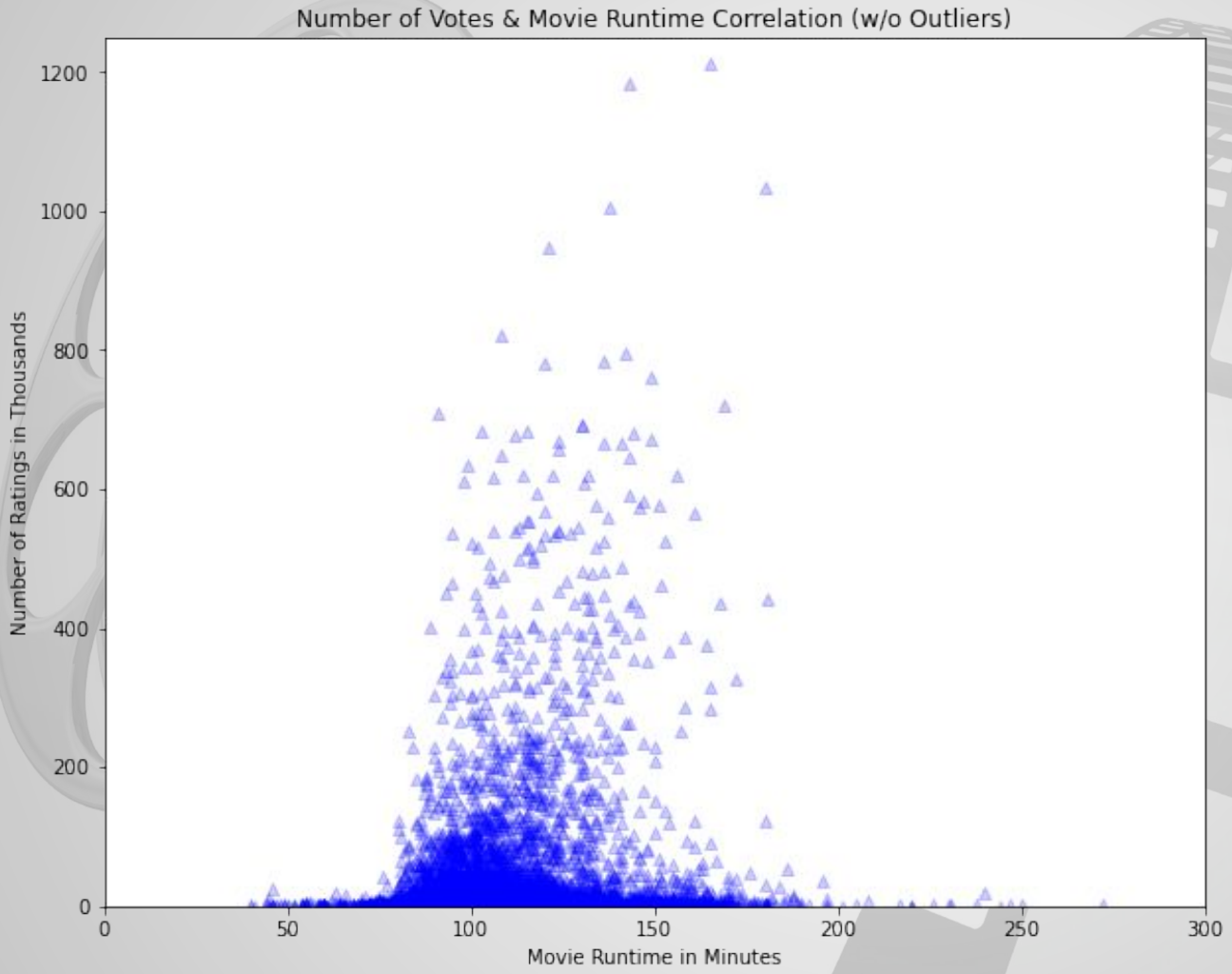
Results

Movie Length

How long should the movie be?

The scatter plot here shows a runtime range for optimal average rating.





Results

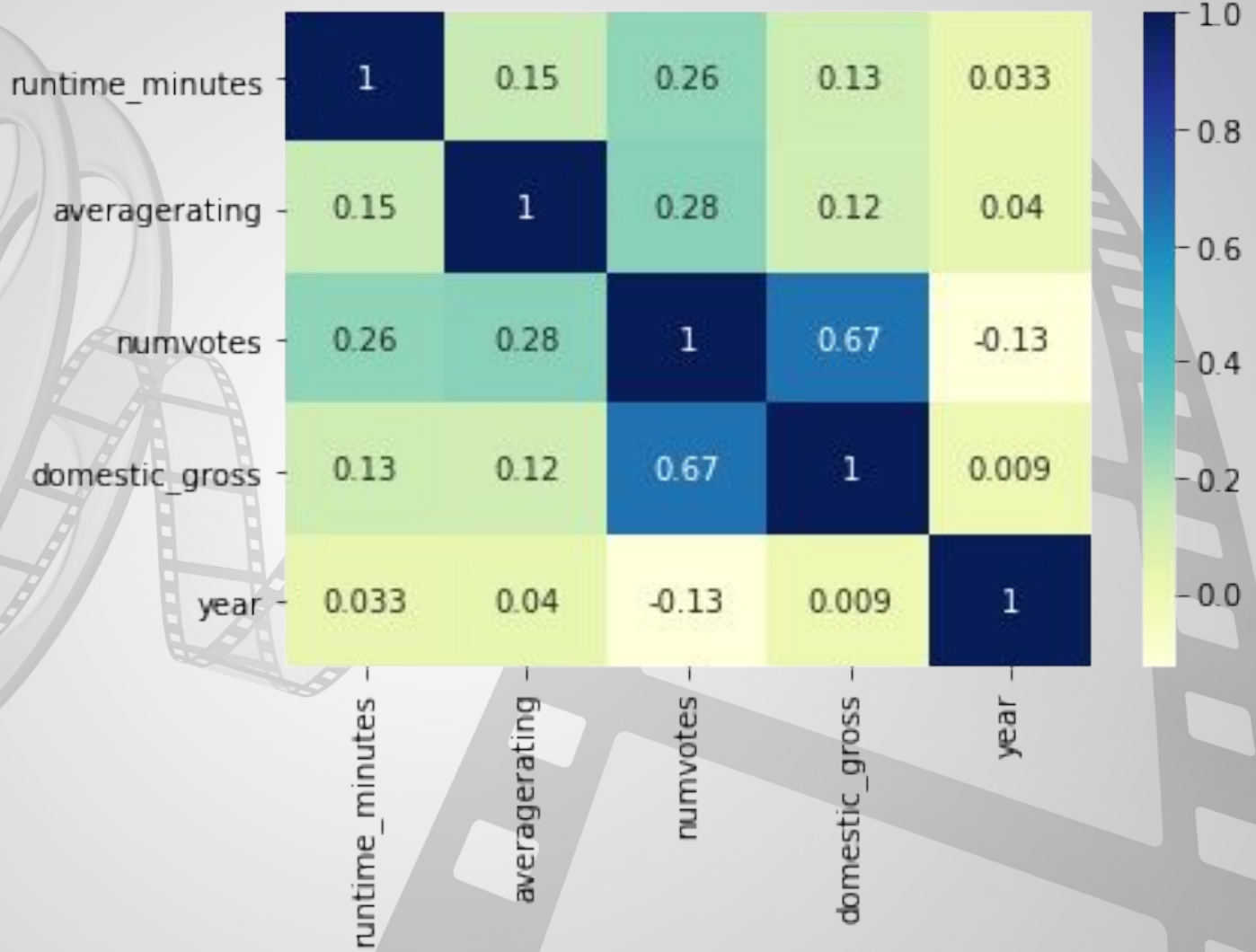
Movie Length cont.

Inline with the average rating correlation, we see clustering around the 100-120 minute mark for overall number of ratings as well.

Results

Rating importance

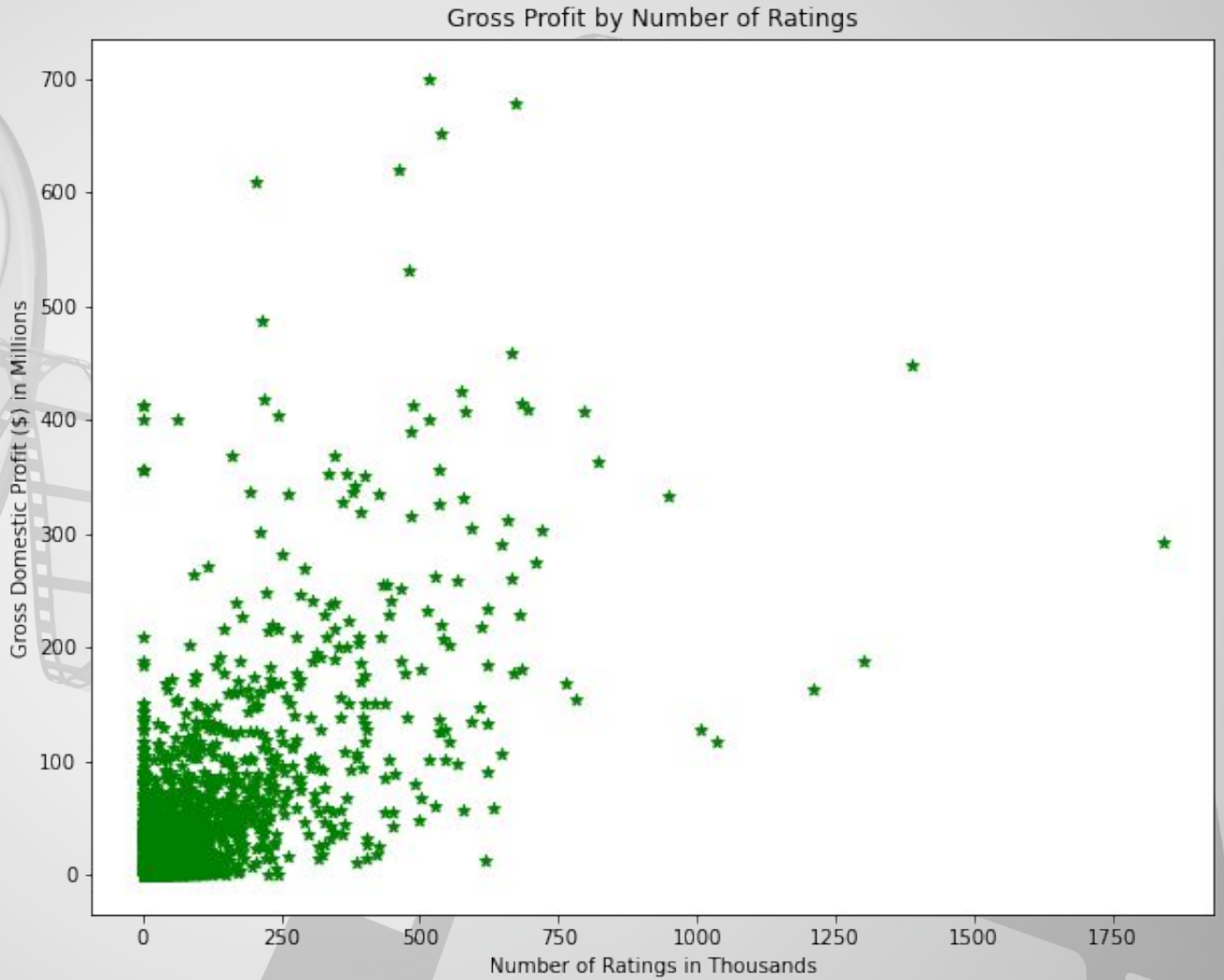
With a correlation of .67 in the heatmap here, we see a pretty strong positive correlation between the number of ratings and the domestic gross profit.



Results

Rating importance cont.

Scatter plot representation of the heatmap previously shown.



In Conclusion...

- Target Adventure, Action, Comedy and Drama genres.
- To maximize ratings, stay around 120 minutes.
- Go for volume.

Limitations of this analysis include the lack of in depth insight into genres and their relationships with gross profit due to lack of time. Coming soon to a data science student near you...

Thank You!

Email: ashley@eakland.net
GitHub: @smashley-eakland
LinkedIn: https://www.linkedin.com/in/ashleyeakland/