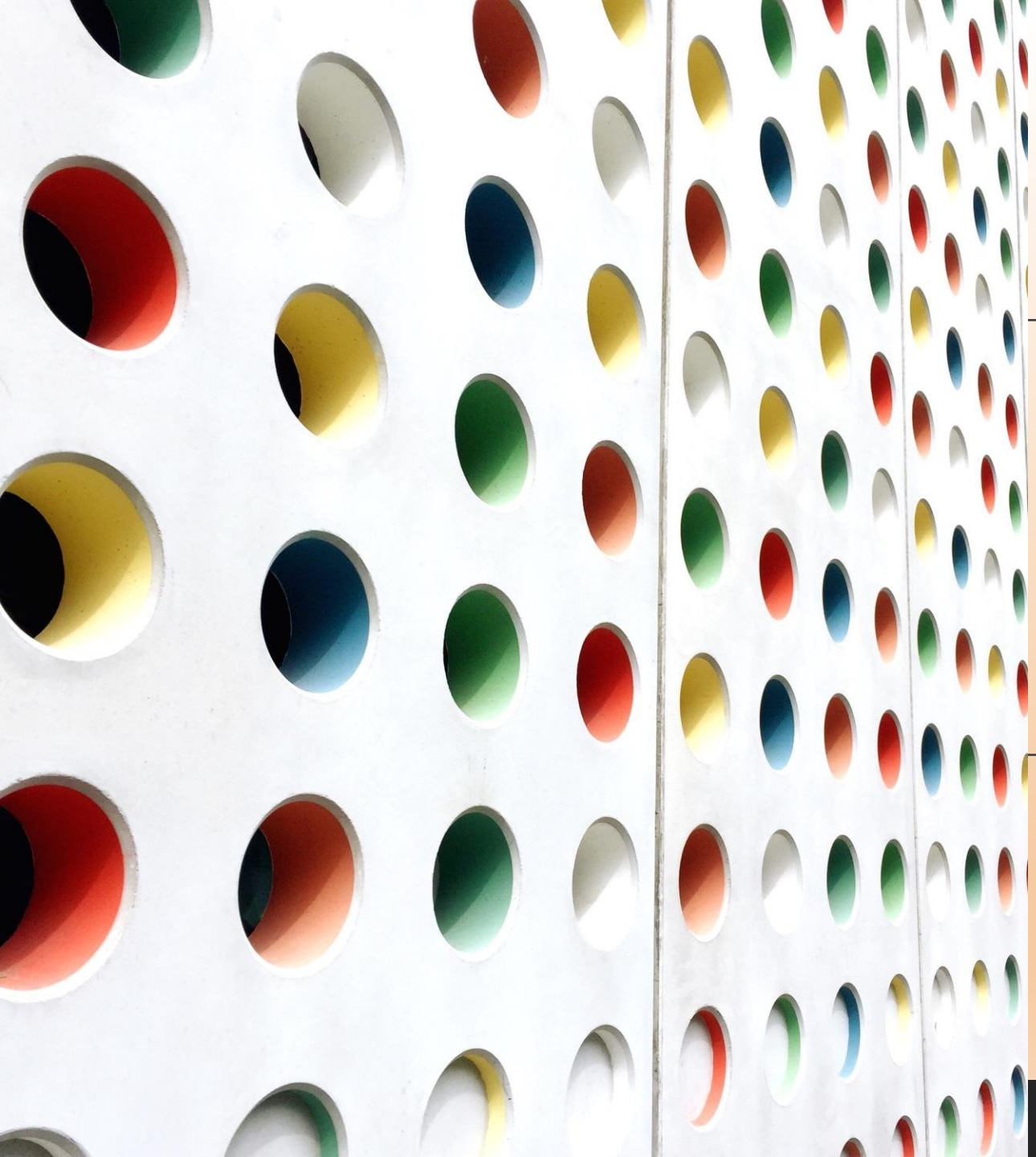


# Microsoft Movie Analysis

FRANK MANDELE

# Introduction.

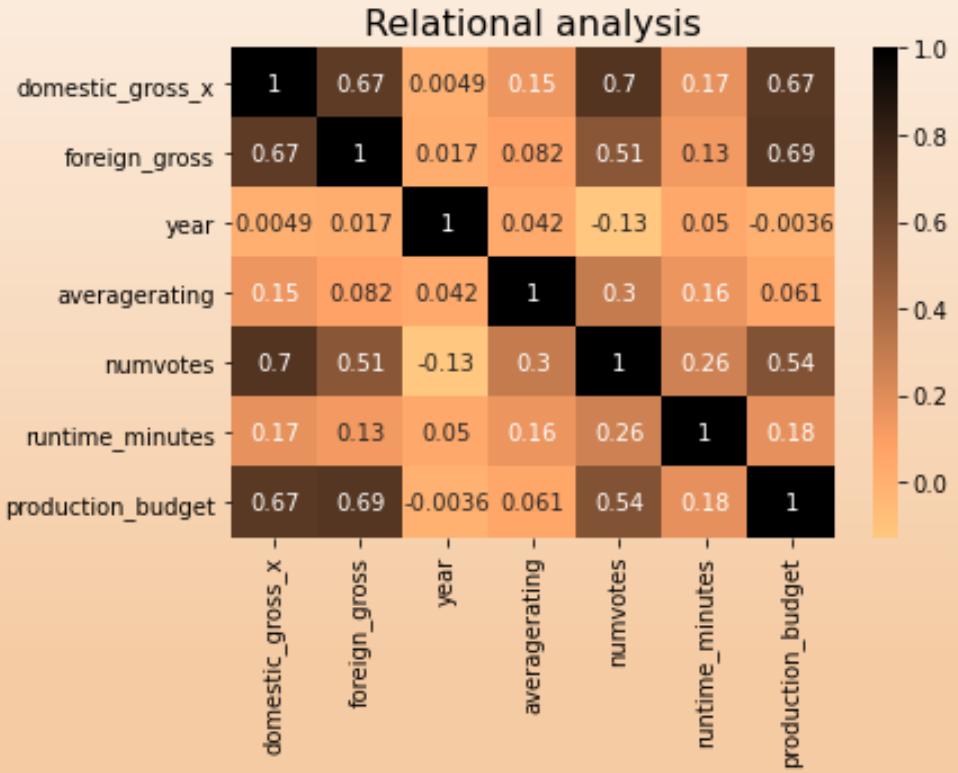
Microsoft has taken a keen interest in film production and wishes to create its own film studio but has limited knowledge of the film industry, this presentation hopes to serve as a guide for Microsoft  and help make an informed decision as they hope to venture into film production.

This presentation features visualizations of analysis of data sourced from: IMDB, The Numbers, and Box Office Mojo and captures data on genres, production budgets, returns from films, viewer' ratings, film durations

# Target Variables:

- Domestic Gross
- Foreign Gross
- Average Rating
- Runtime Minutes
- Production Budget
- Number of Votes

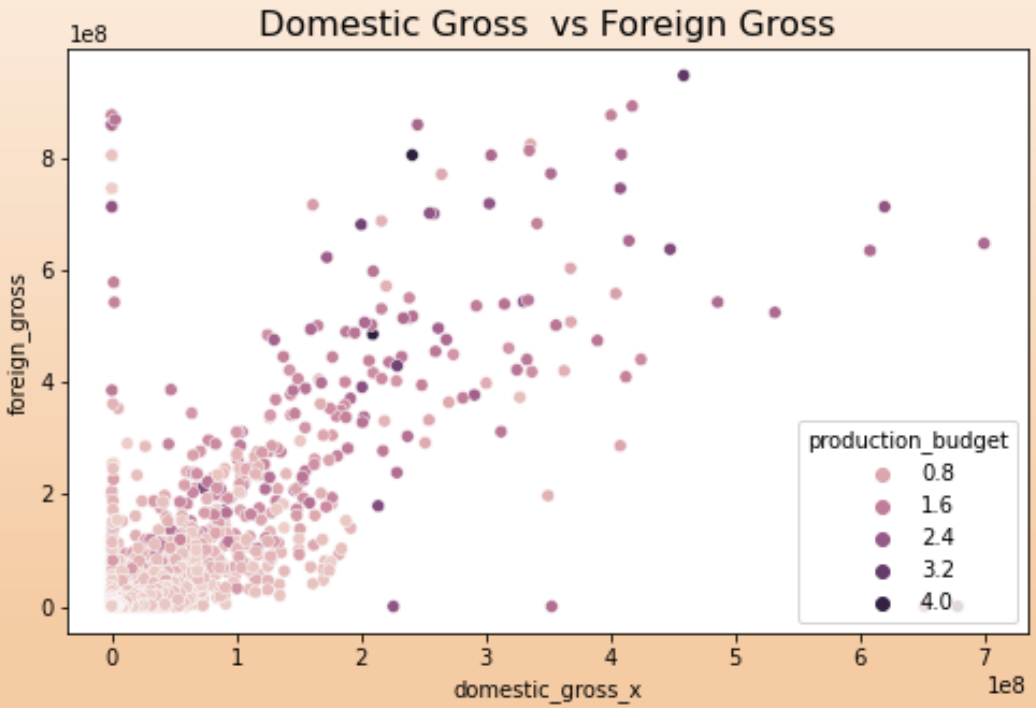
# Relational Analysis of Variables




Overall, the variables in the dataset are positively correlated(except for instances of `year` that has **–ve annotations** with some correlations being fairly stronger than others. This can be observed through the annotations in the plots. Positive values show positive correlation whereas the magnitude of the annotations indicate strength of correlation with `1.0` being the highest value.

In the case of `production_budget`, `foreign_gross` and `domestic_gross_x` for example, the correlation is moderately strong.

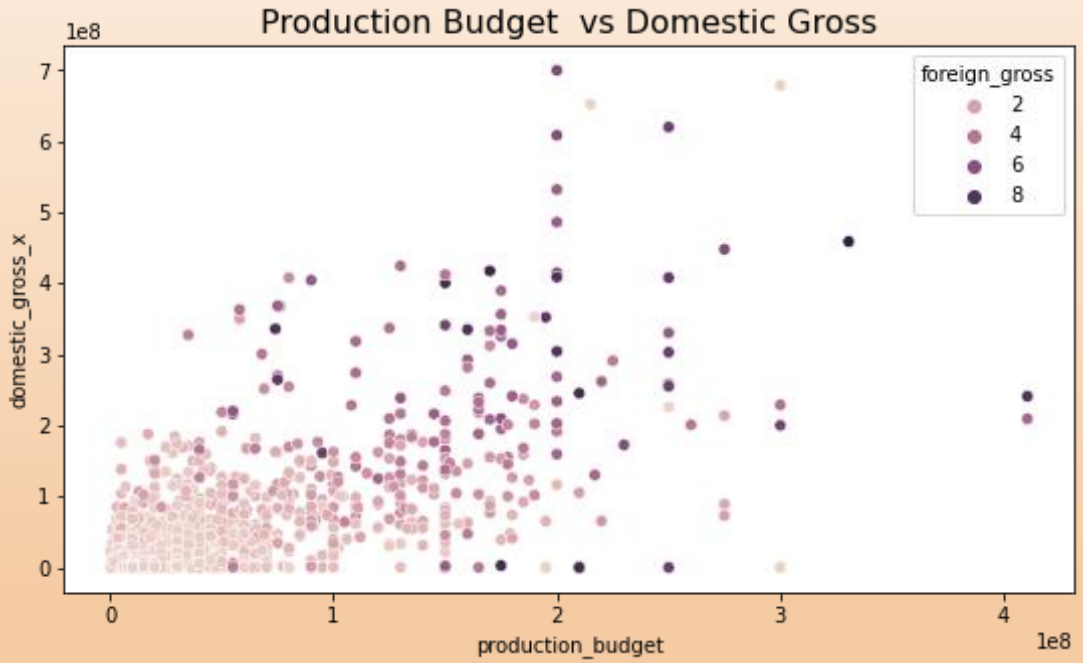
# Domestic Gross vs Foreign Gross




The points in the plot appear to be grouped in a manner that shows an upward trend from the origin (0,0). This is an indication that `domestic_gross_x` and `foreign_gross` are positively correlated.

Whenever `domestic_gross_x` increased, there appear to be a corresponding increment in `foreign_gross` as well. It is also worth noting the increasing number of dark spots representing `production_budget` as the spots spread away from the origin.

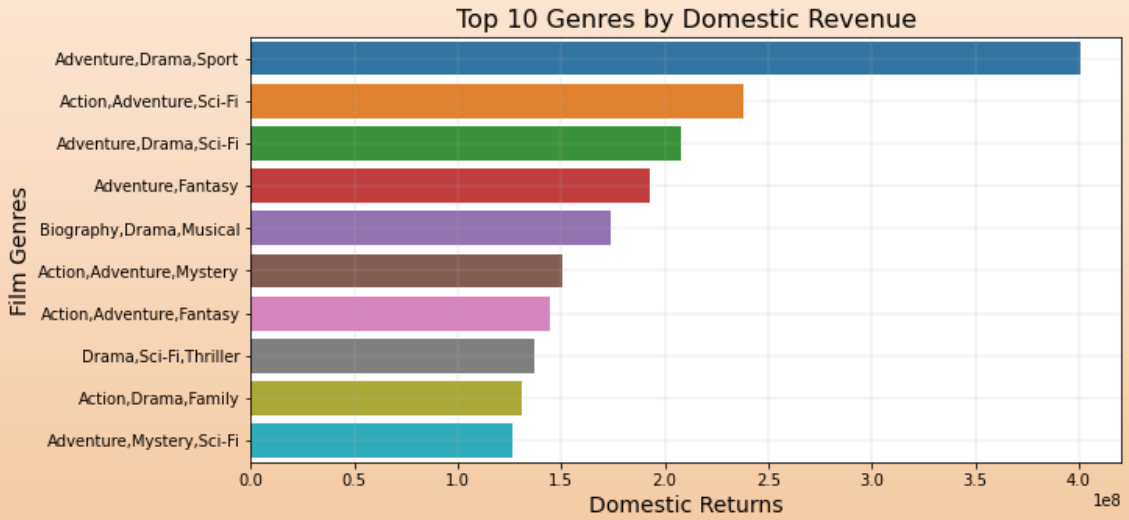
# Production Budget vs Foreign Gross




From the visualization, `production_budget` and `domestic_gross_x` portray an overall positive correlation that appears to weak.

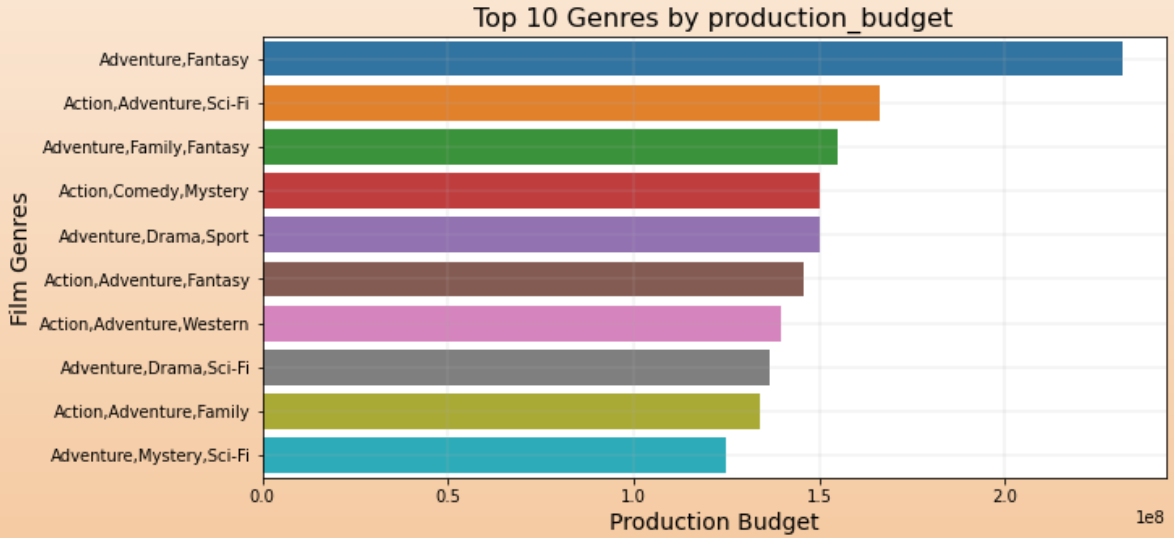
Increased production budget has an corresponding general increment in Domestic gross as well.

# Top Genres by Revenue



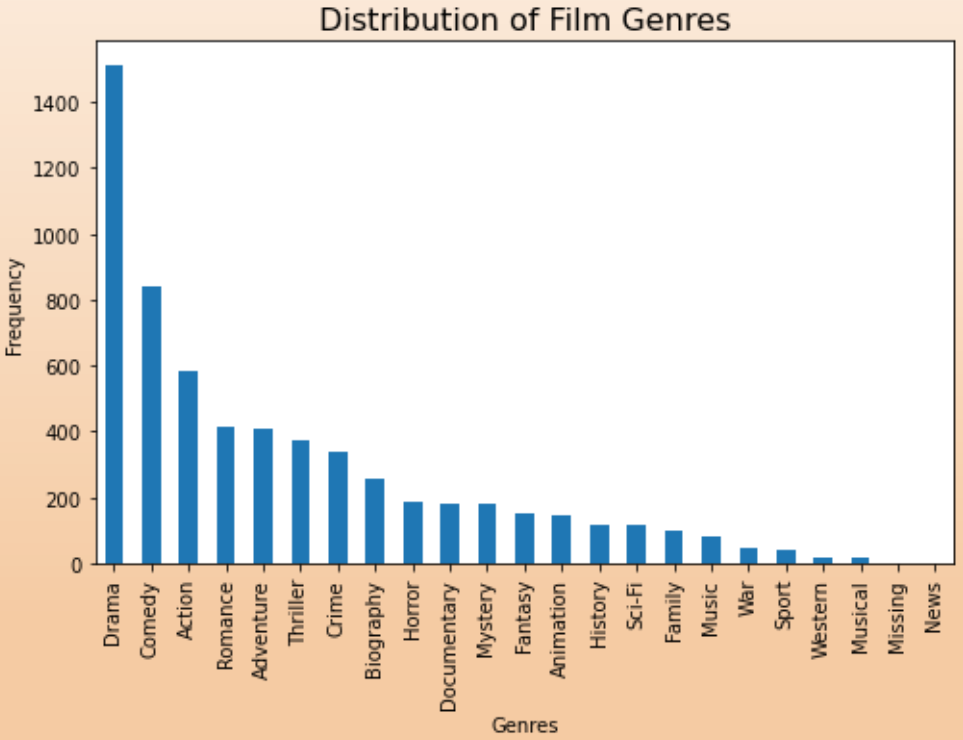

From the visualization, (Adventure, Drama, Sport) genres show the highest sales observed domestically by a significant margin, followed by (Action, Adventure, Sci-Fi) whereas (Adventure, Mystery, Sci-Fi) recorded the lowest sales domestically

# Top Genres by Production Budget



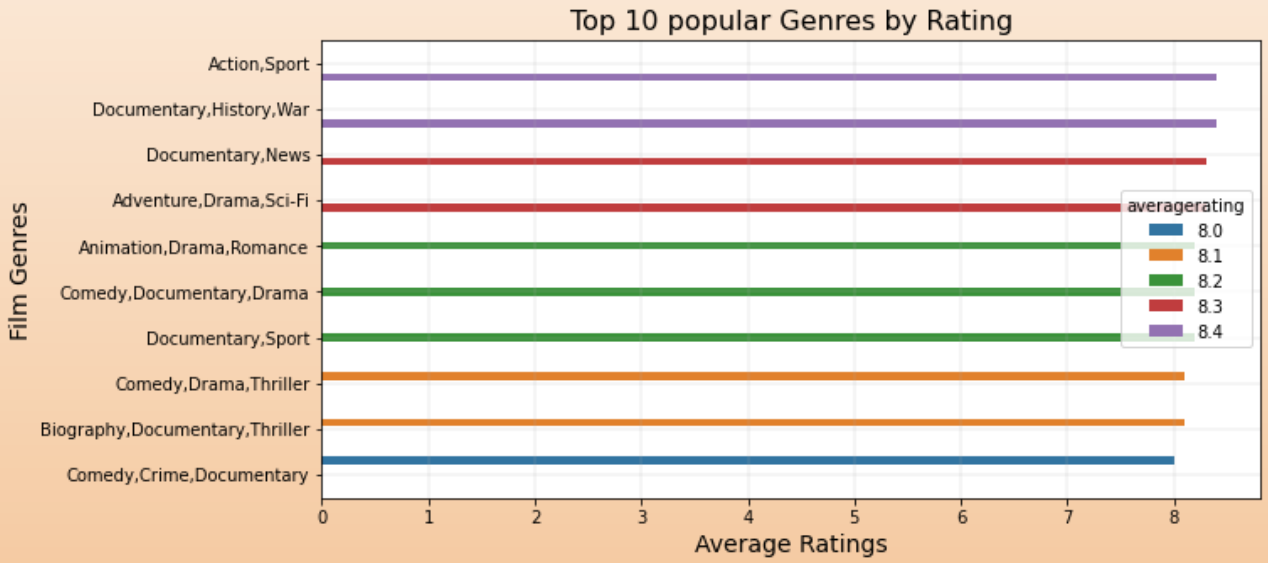

In regards to Film Production Budgets, Adventure, Fantasy recorded the highest Budgets whereas Adventure, Mystery, Sci-fi recorded the lowest production budget.

# Genres Distribution




A great majority of films made are seen to have the genre 'drama' by a big margin, followed by 'comedy' and 'action' films. 'Sport' and 'News' closes the list

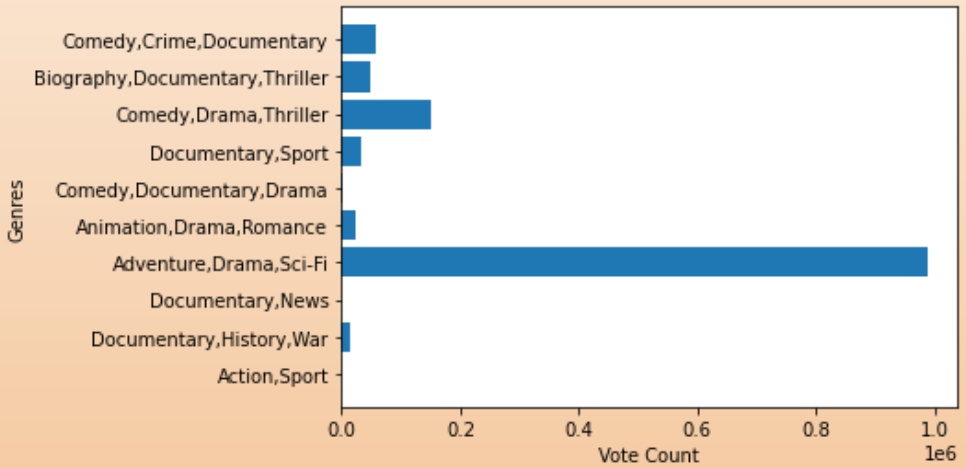
# Popular Genres by Rating




Action, sport and Documentary, History, War genres are rated to be the most popular in terms of rating. This ratings however seeks further clarity since the data might be from a smaller number of votes and therefore might easily create the element of bias in the analysis and therefore not give a true depiction of what is popular to the viewers.

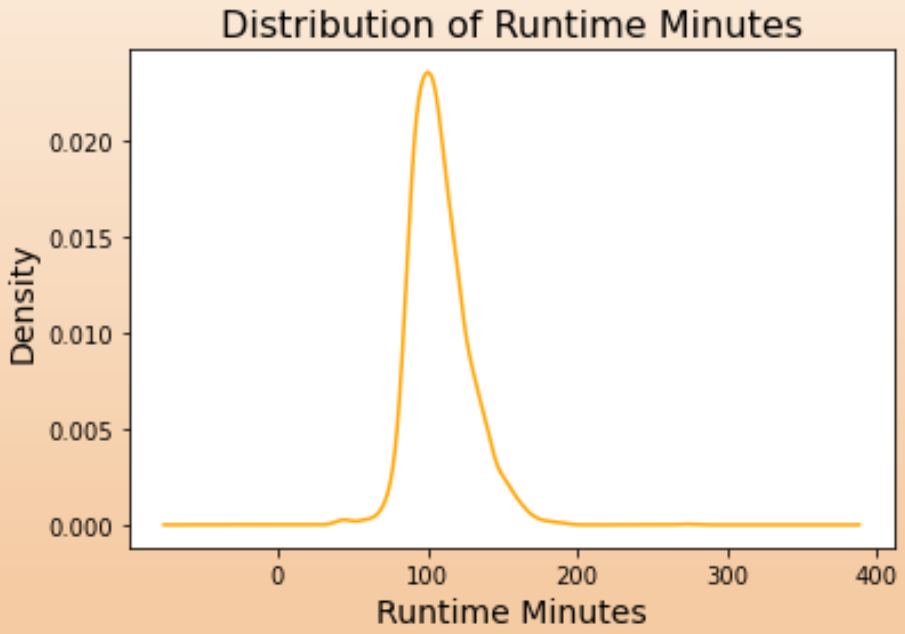
As a result, there's further analysis on the number of votes in the dataset

# Distribution of Votes on Genre Data




As observed from the plot above, there is a huge variation of number of counts with the respective genres. The `ratings` therefore might not be a proper presentation of how popular a film genre is since the rating, as seen from the visualization might be from a smaller group of viewers.

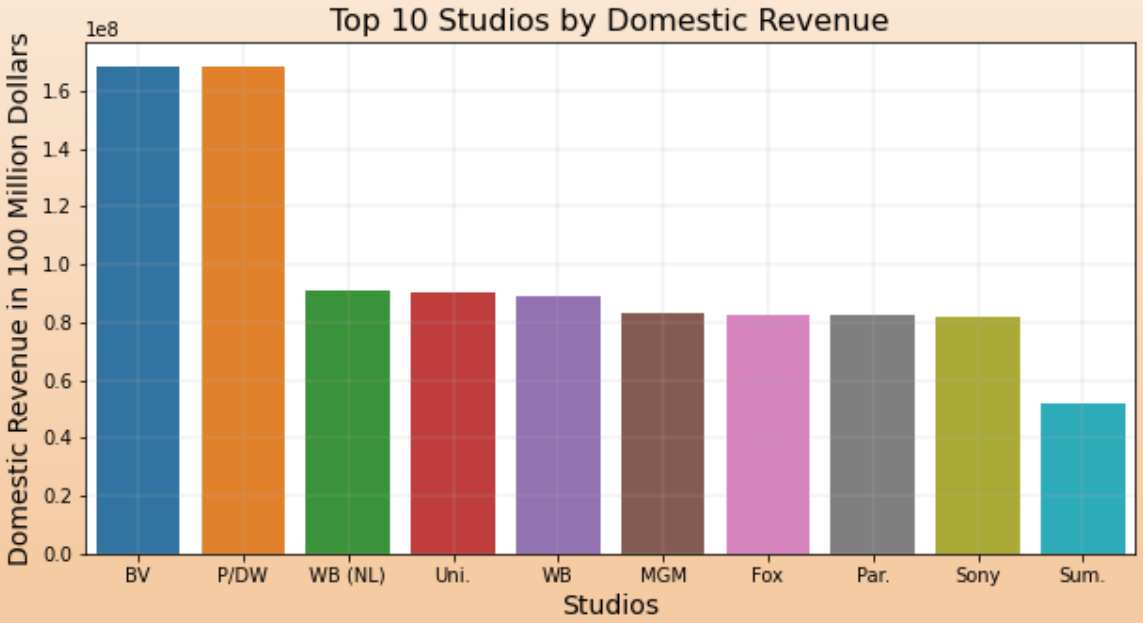
# Distribution of Runtime Minutes




Using a density plot, we can figure out films with durations of approximately 90 minutes were the most common for films because of it being high peak region.

Majority of the films are roughly an hour and a half long

# Additional Analysis: Best Performing Studios




Studios `BV` and `PD` tops the list of Top 10 studios by Domestic Revenue. The films produced under the two studios yielded the highest returns whereas `Sum` closes the top performers list

# Conclusions and Recommendations

- Based on the results of the analysis it appears to be more profitable for Microsoft to engage in films genres that featured in the top 10 categories using film samples produced by some of the top studios to learn what makes them standout in the film industry.
- Production budgets is also estimated to be around 50,000 dollars to around 410 million dollars(min-max generated from descriptive statistics).
- Something worth noting also is the fact that Adventure, Drama, Sci-Fi ranked 3rd in terms of rating but had the highest number of votes count but a staggering majority, we can conclusively say it's easily the most popular genres by numbers despite not appearing as the first, in addition, the difference in rating in quite small.
- It might also be important to consider data on `budget` and `domestic gross` might not be the actual representation of the real values given the fact that the data had quite a number of missing values and therefore this analysis is more of an estimate of the real figures. During the analysis process, I observed cases where similar movies had varying data across the different sources such as different runtime minutes, production budgets, genres and foreign gross.
- This analysis alone might not be sufficient to film production since it only offers insight based on general descriptive data and not technical bit of production.

# The End