Data Science Phase 1 Project

Analysis of Movie Data to Determine the best Production Genre for Microsoft Movie Production

Author : Vivian Maiyo
Moringa School

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

Microsoft Movie Studio is looking to create movies and wants an analysis of movie data to help them decide the types of movies to produce

Data

We are using the below data sets from IMDB and The Numbers movie database sites

- IMDb. title.basics
- IMDb. title.ratings
- tn.movie_budgets

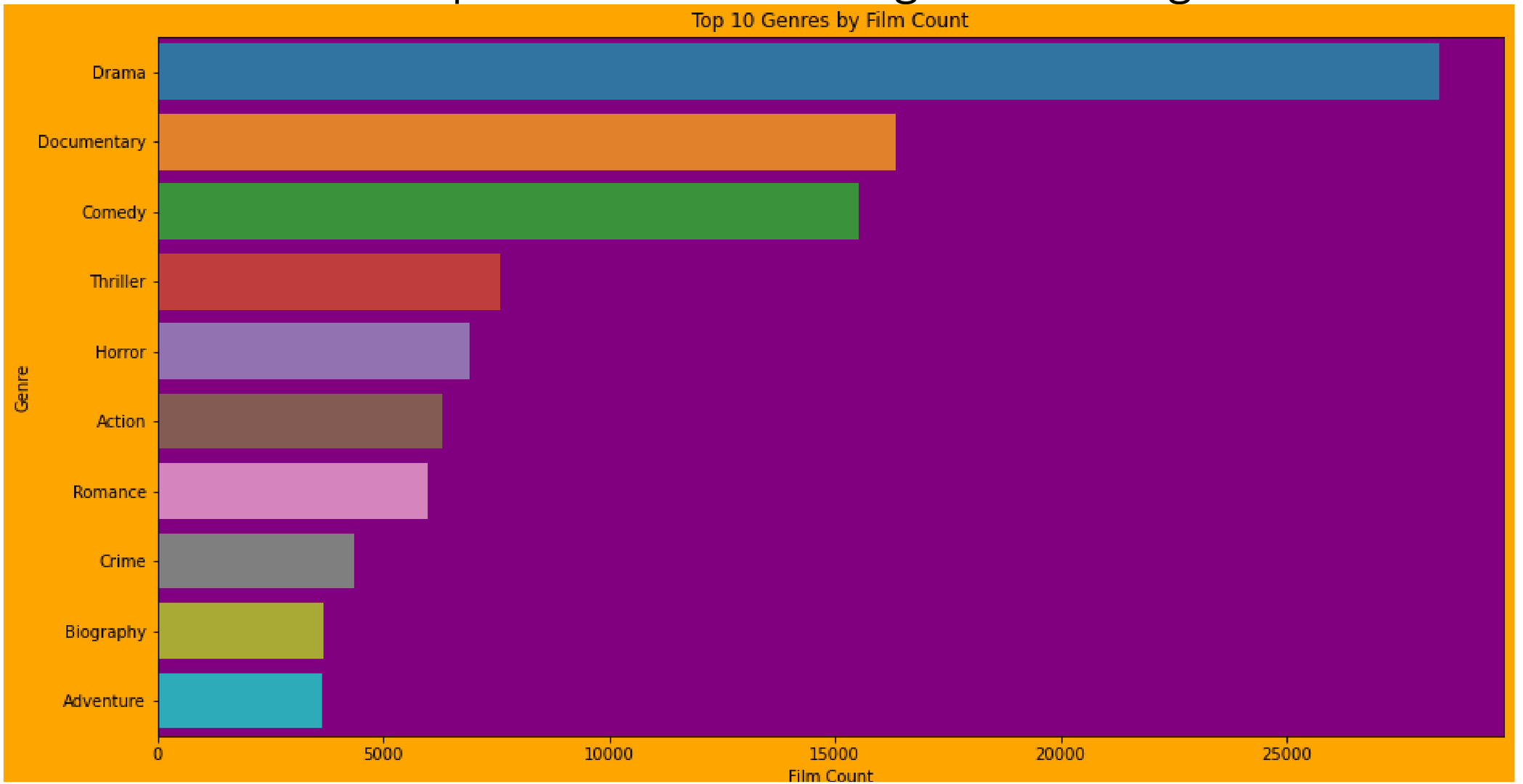
The data sets contain data on movie ratings, genre, production cost, domestic and worldwide gross revenue, movie release date, number of votes and movie run time in minutes.

Methods

In this analysis we used descriptive statistics to describe relationships between variables and further went ahead to visualize the results using different vizualization tools such as heatmaps, line graph, bar graphs and scatter plots.

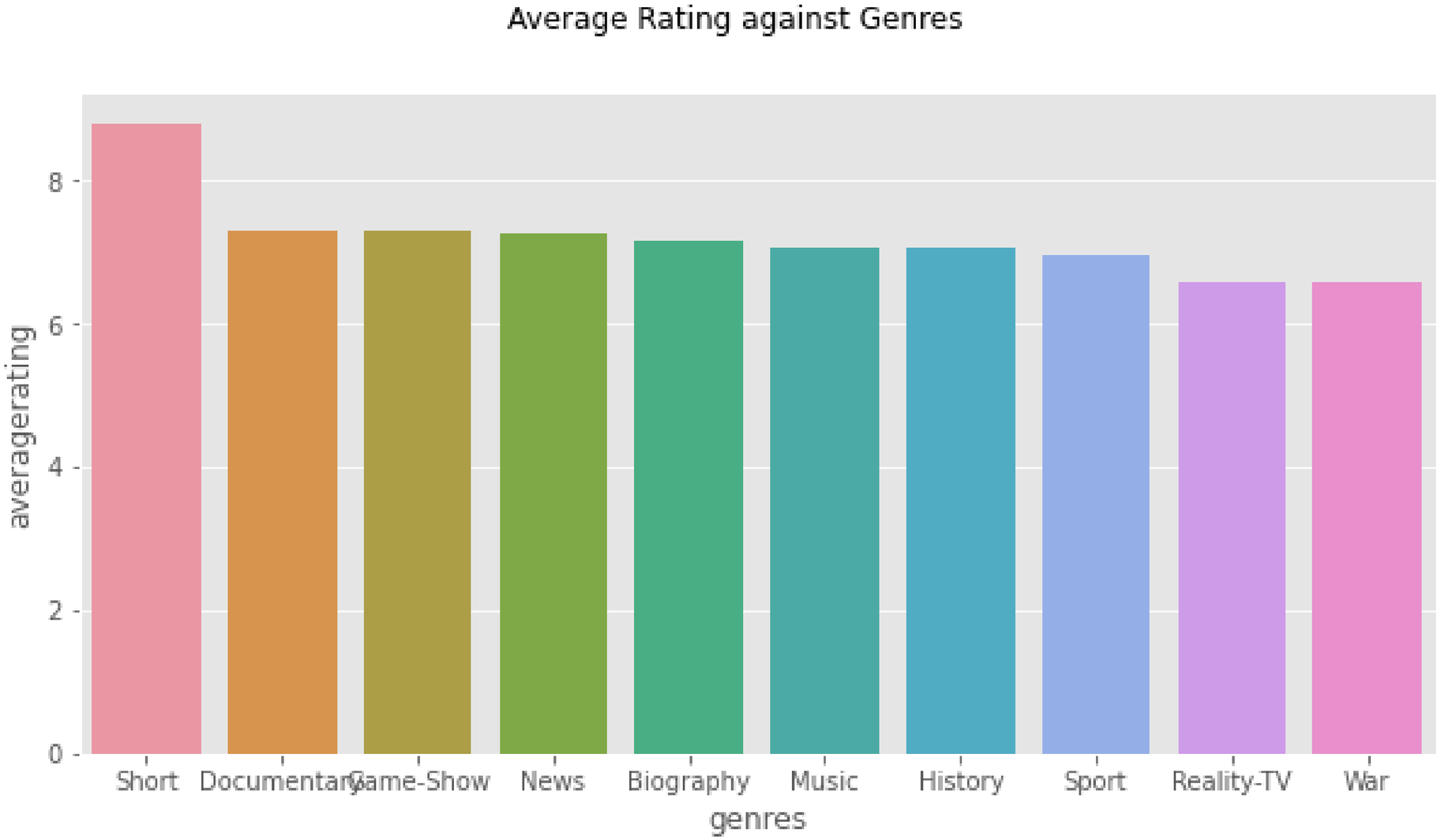
Results

The Horizontal Bar Graph below shows the genre with highest film count



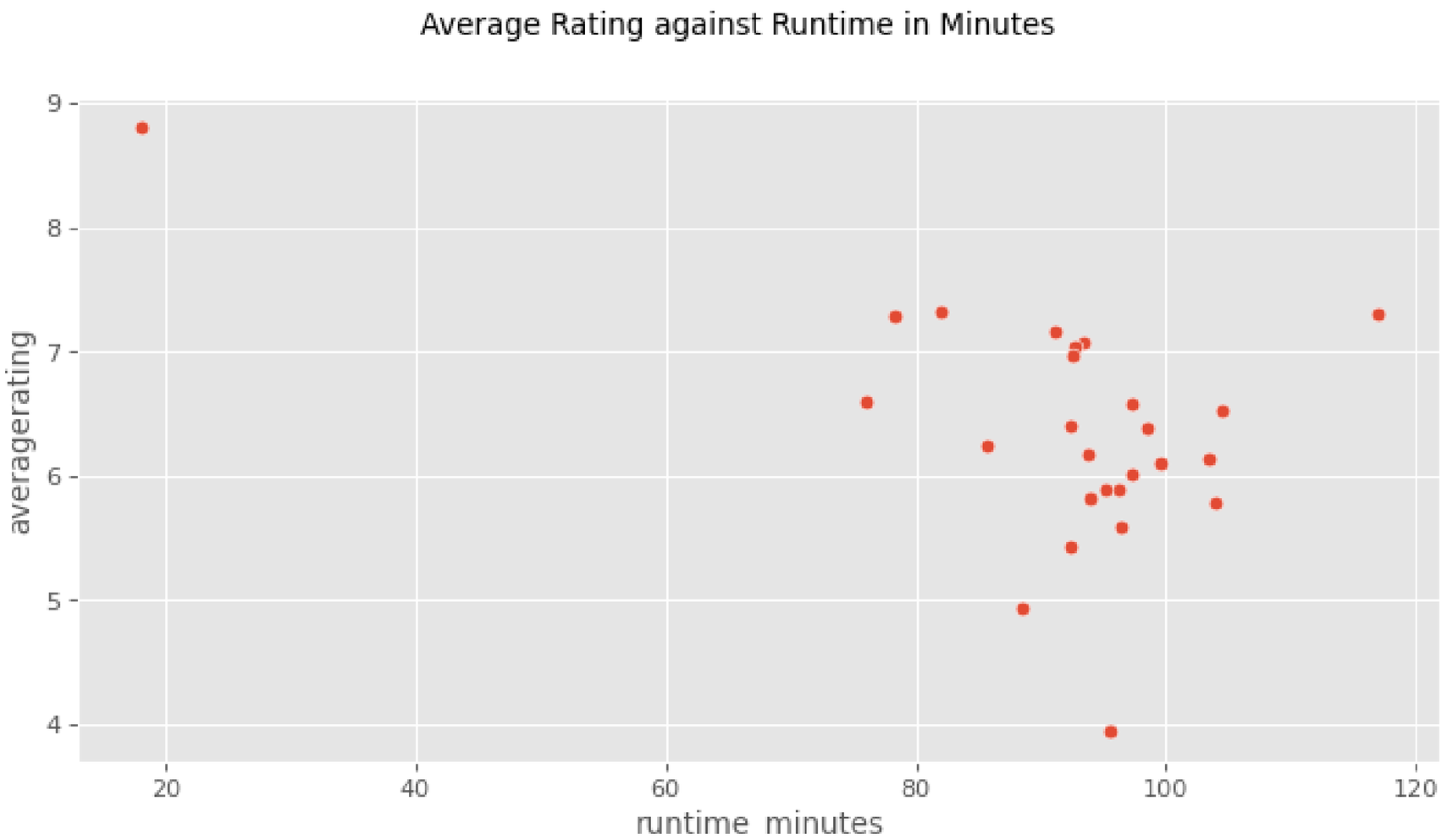
The most popular genre by film count is drama, followed by documentary then comedy. This implies that three genres are the most popular genre amongst film producers.

Bar chart showing average rating by Genre



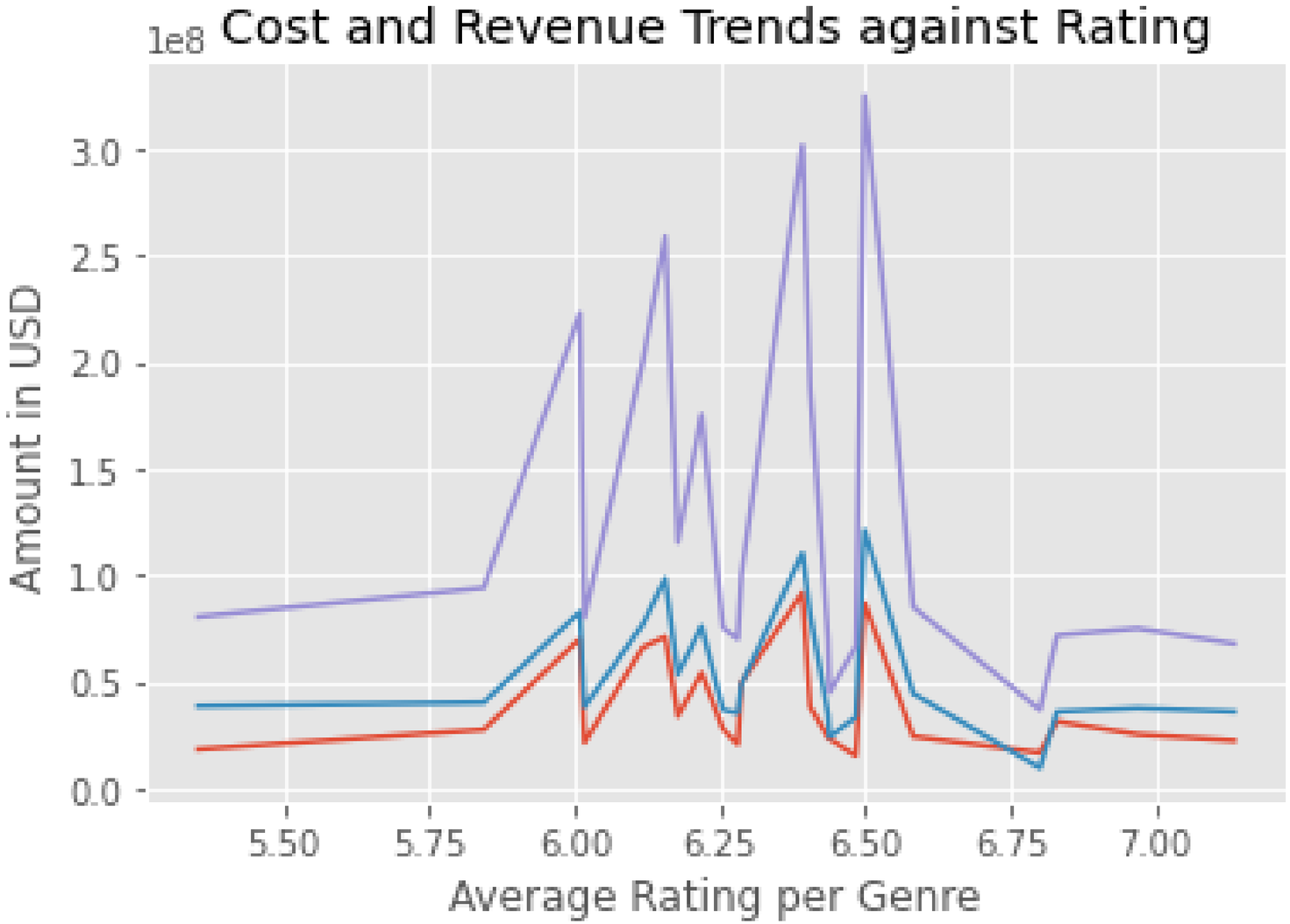
The bar chart outputs average film rating by genre. Short films are rated best followed by documentaries. Interestingly, both drama and comedy genres are not among the top 10 rated

Scatter Plot showing the correlation between runtime and average rating



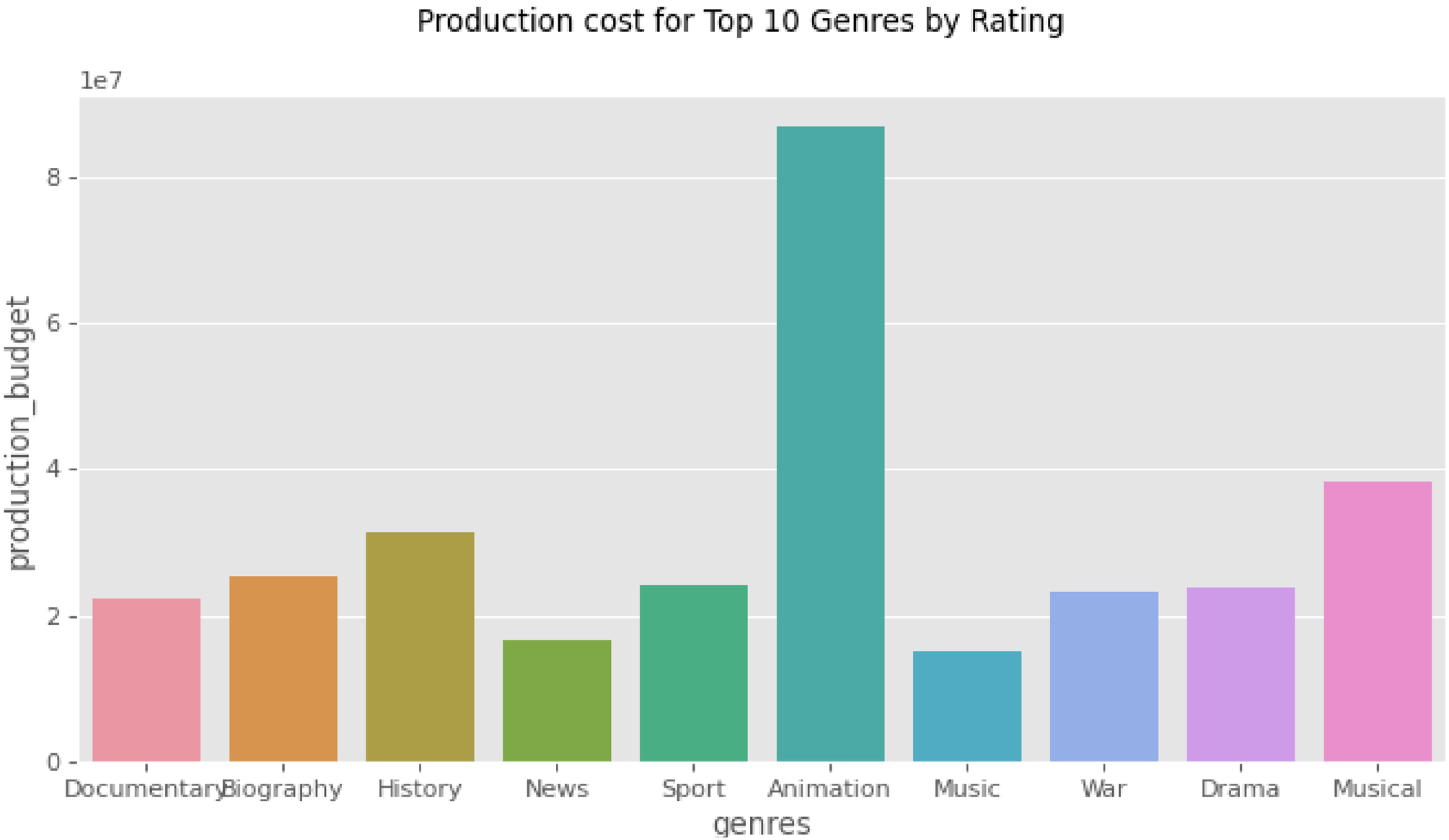
The scatter plot shows weak positive correlation between run time and average rating.This implies that film run time does not affect its rating.

Line Graph showing trends in cost of production and revenue



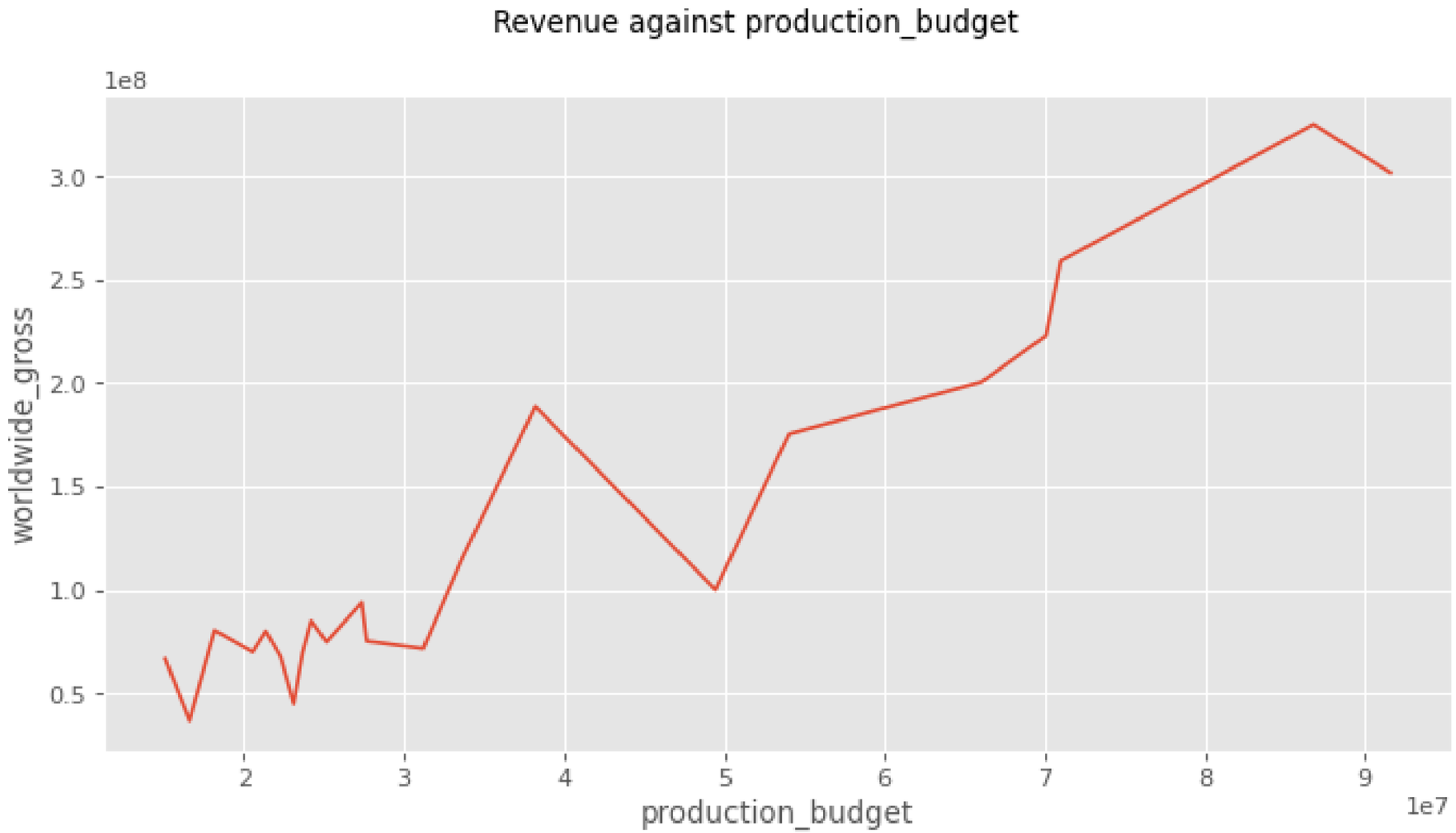
Even though the cost of production, domestic gross and worldwide gross shows similar trend against average rating by genre, it is worth noting that the cost of production is higher among films rated good(rating between as compared to those rated excellent (rating above 7). This implies that high productions cost does not translate to highly rated movies.

Bar Graph showinhg cost of production for Top 10 Genres by Rating



Generally, highly rated films have lower production costs, which implies that high production costs does is not an indicator of a good film.

A line graph showing how revenue changes with cost of production



Worldwide gross revenue generally increases with the cost of production. This implies that is no significant change in Return on Investment with an increase in cost of production

Conclusion

- Film producers generally prefer dramas and documentaries
- Short films and documentaries are rated the highest
- There is no correlation between film run time and its rating
- The cost of production does not determine the film's rating
- Worldwide gross revenue generally increases with cost of production

The End

Thank you!