

MICROSOFTS NEW MOVIE STUDIO ANALYTICS



1.BUSINESS PROBLEM

Microsoft have decided to create a new movie studio, but they don't know anything about creating movies. Taking into consideration that the film industry is a highly competitive industry. We are to explore what types of films are currently doing the best at the box office and will also maximize the profits

2. DATA UNDERSTANDING

The data set are csv files giving the

1.Title ratings,

2.Title basics,

3.Movie studios,

4.Movie titles the domestic and foreign gross sales of movie titles.

5.Movie Release year

The data source IMDb website which is a worldwide online database for information realting to movies,Tv programmes and online content streaming.

3.DATA PREPARATION

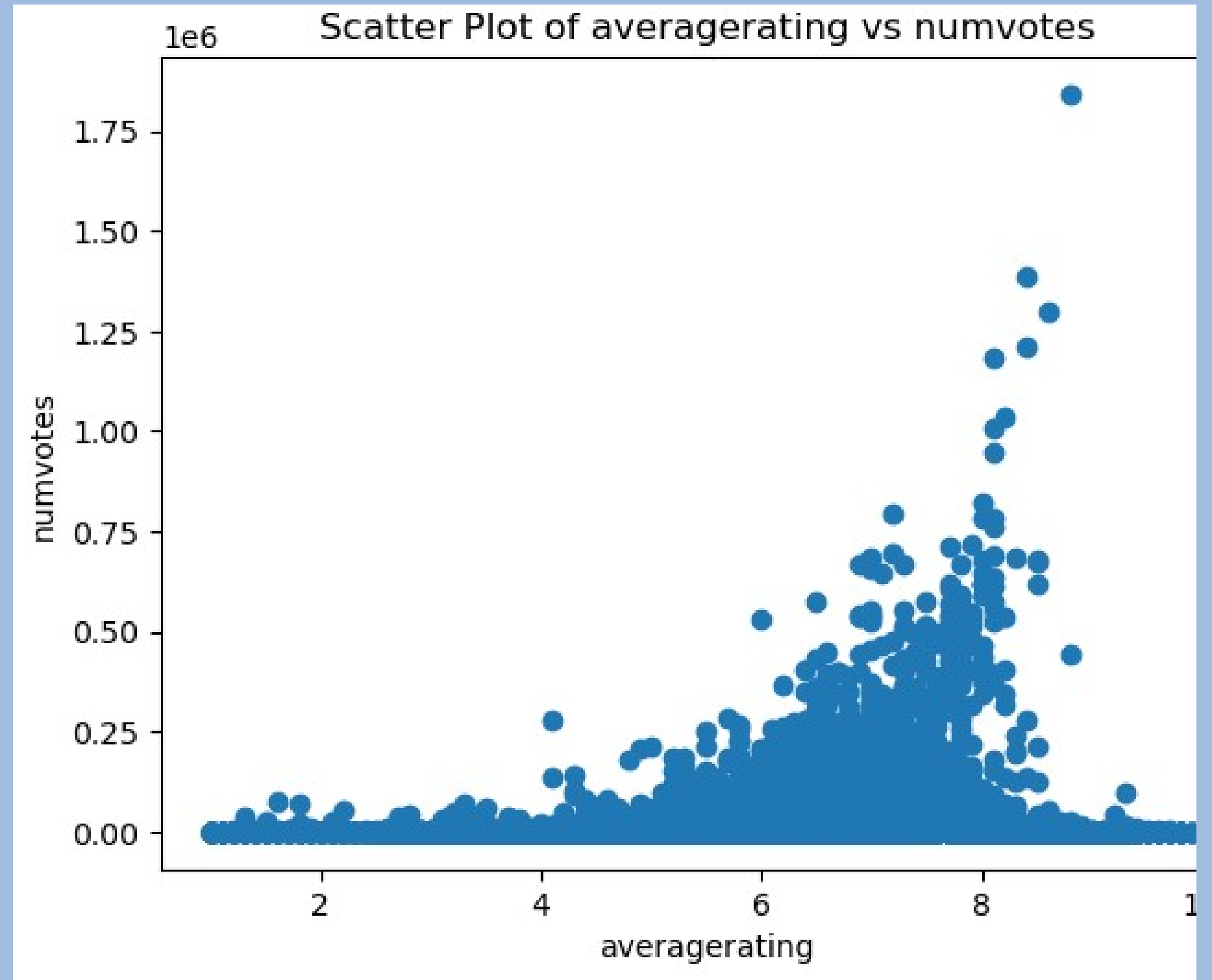
All the three files were used in this analysis for Microsoft to gain better insights relating to the new movie studio.I analysed the column types and description of the data sets, followed by data cleaning on the data sets which include

- checking for duplicates,filling the NaN values with the mean of the relevant columns,populated new columns where needed.**

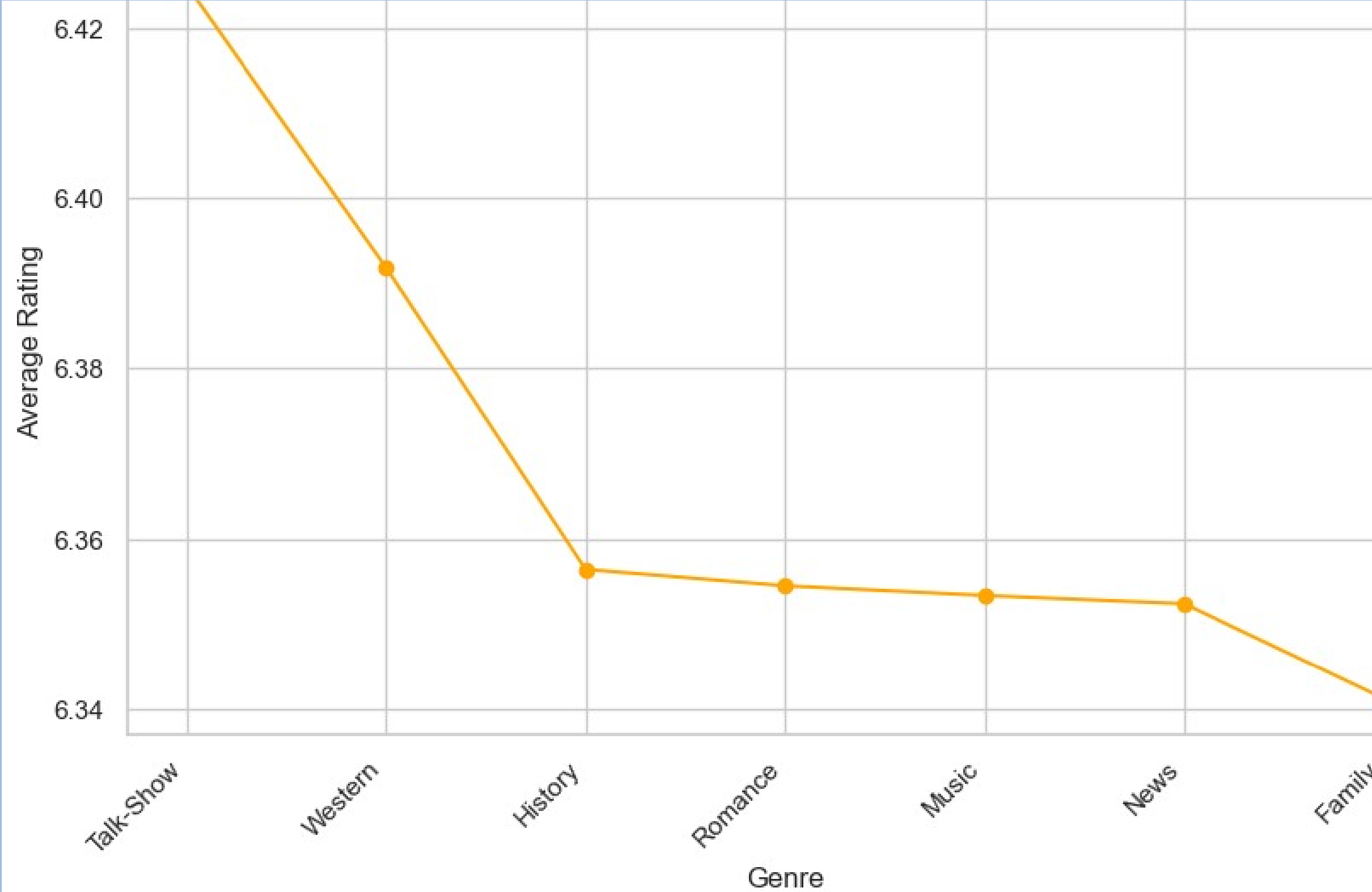
Then combined the title basics data with the title ratings data using the unique identifier tconst to create a combined_title_df.

4. DATA ANALYSIS

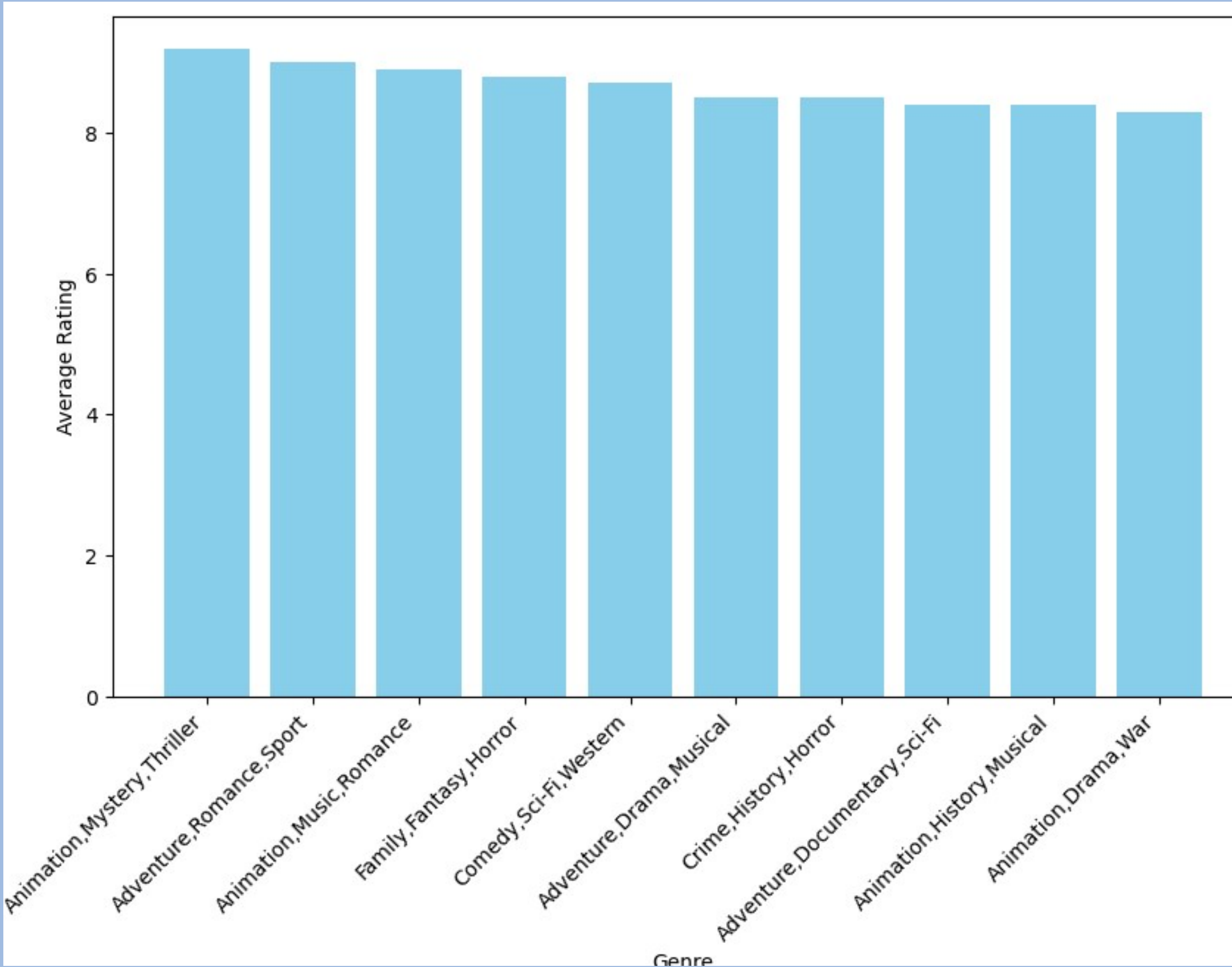
Correlation was checked between average ratings and no of votes and there was no significant relationship between the two.



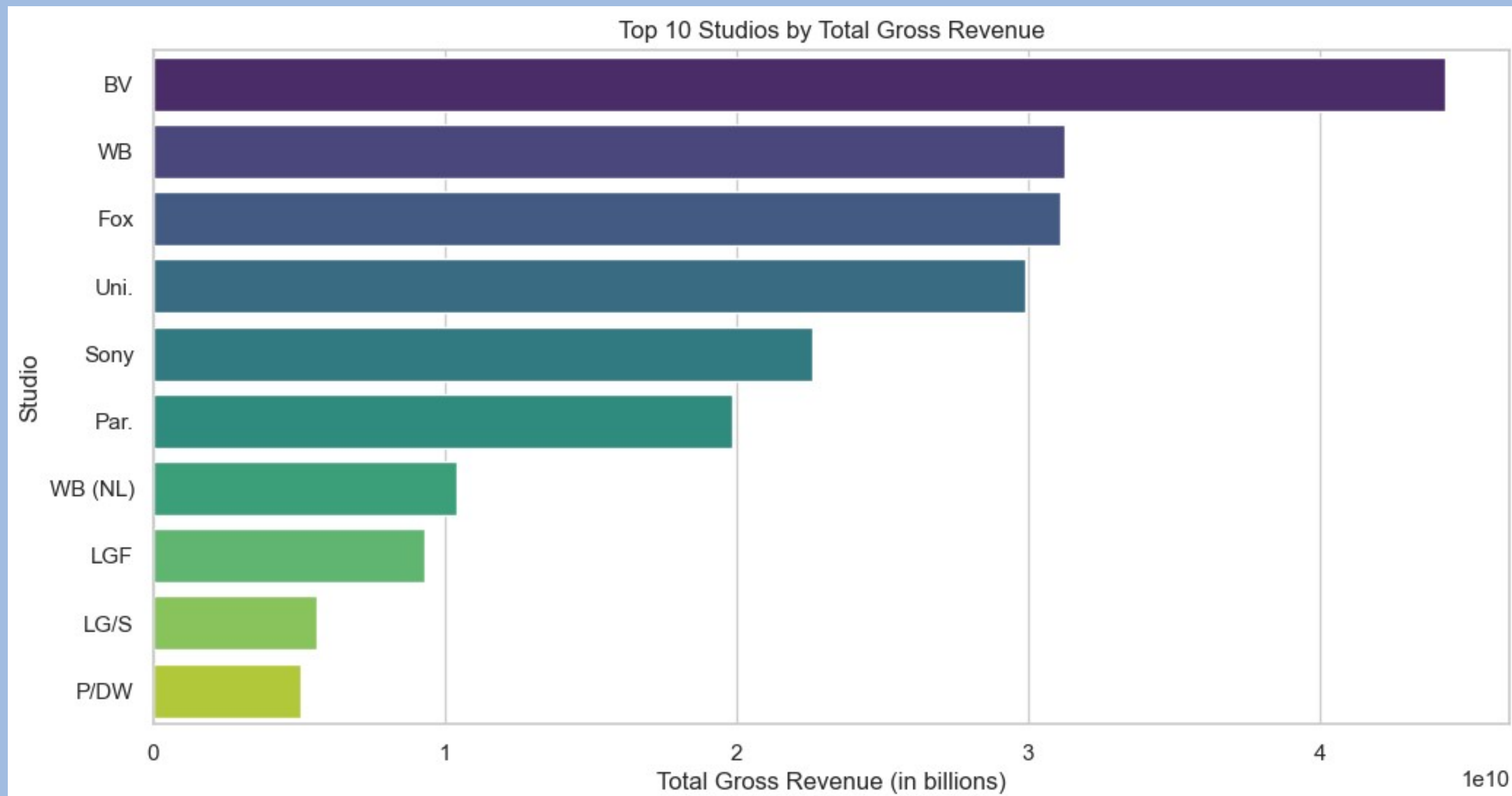
**Top 7 Genre
combination
by average
rating.**



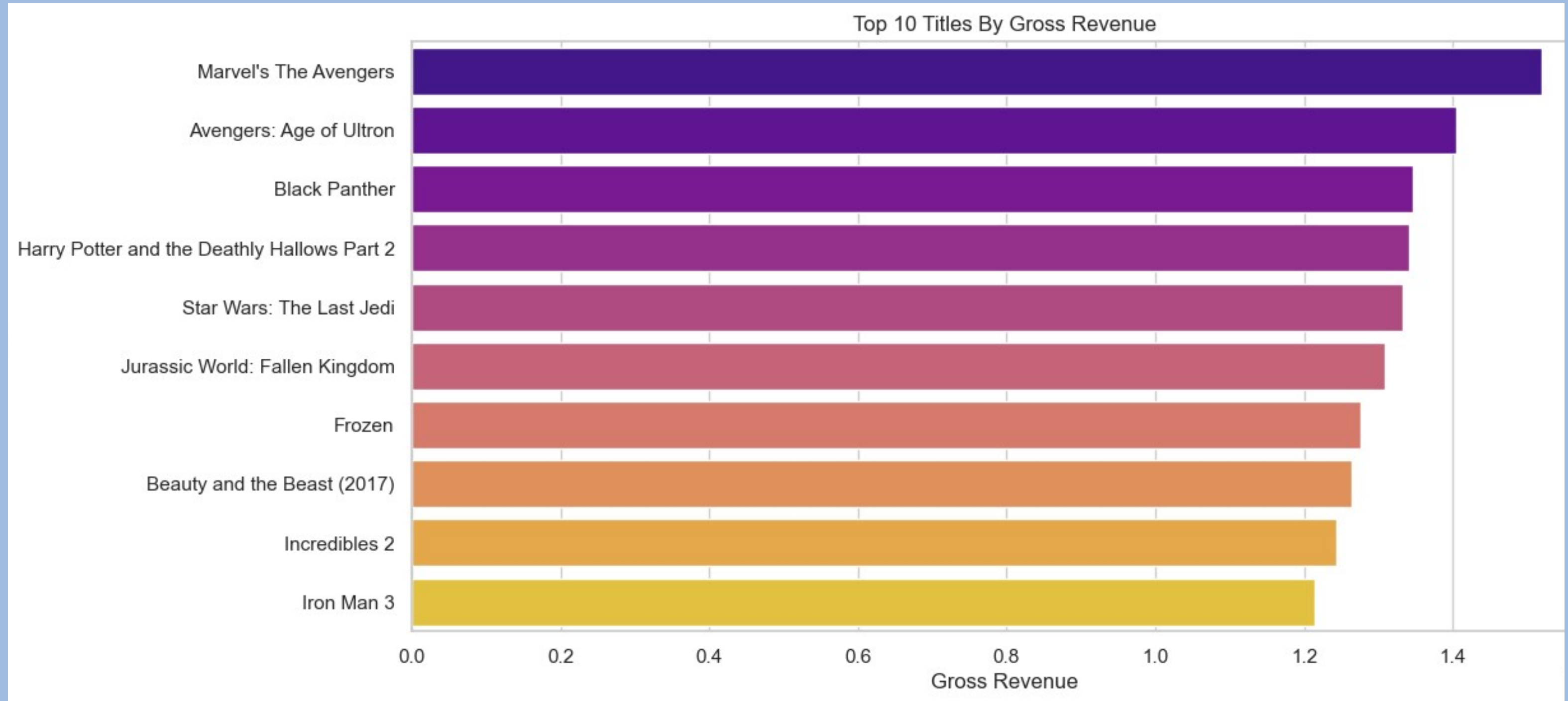
Top 10 Genres by average rating



Top 10 studios by Total Gross Revenues



Top 10 Titles by Gross Revenues



Microsoft can stay informed and upto date about the industry,to be creative in the production of movies and aligning the genre combination with the target audience.

CODE QUALITY

Codes were coded as per the PEP8