Phase 1 Project Presentation

NAME: KIMEU FAITH MWENDE

GROUP: DSFP COHORT 15

~Moringa School

MICROSOFT' S NEW MOVIE STUDIO STRATEGY

Business Problem

- Microsoft sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but they don’t know anything about creating movies. You are charged with exploring what types of films are currently doing the best at the box office. You must then translate those findings into actionable insights that the head of Microsoft's new movie studio can use to help decide what type of films to create.

OBJECTIVES

- Import the required libraries
- Load the given data
- Inspect the data
- Perform data cleaning
- Begin data analysis
- Ask relevant questions that need to be answered in the form of visualizations like bar plots to the head of Microsoft.
- Derive conclusions from these visualizations

Import the required libraries

- The following are the libraries that I used:
  - Pandas
  - Matplotlib
  - Seaborn

Load the given data

- The data is in csv format.
- All the required data was loaded with the aid of a function into my notebook.

Inspect the data.

- Look at the shape of the data frame, the information contained in each table and check for missing (Nan) values as well as duplicates.
- Get a feel of the data and have a general idea of what each table can contribute to in the strategic plan of Microsoft geared towards a new movie studio.

Data Cleaning

- Check the percentage of missing values
- Drop or replace null values in the affected columns or rows.
- The replace technique used was filling the null values using the median.
- Why median?

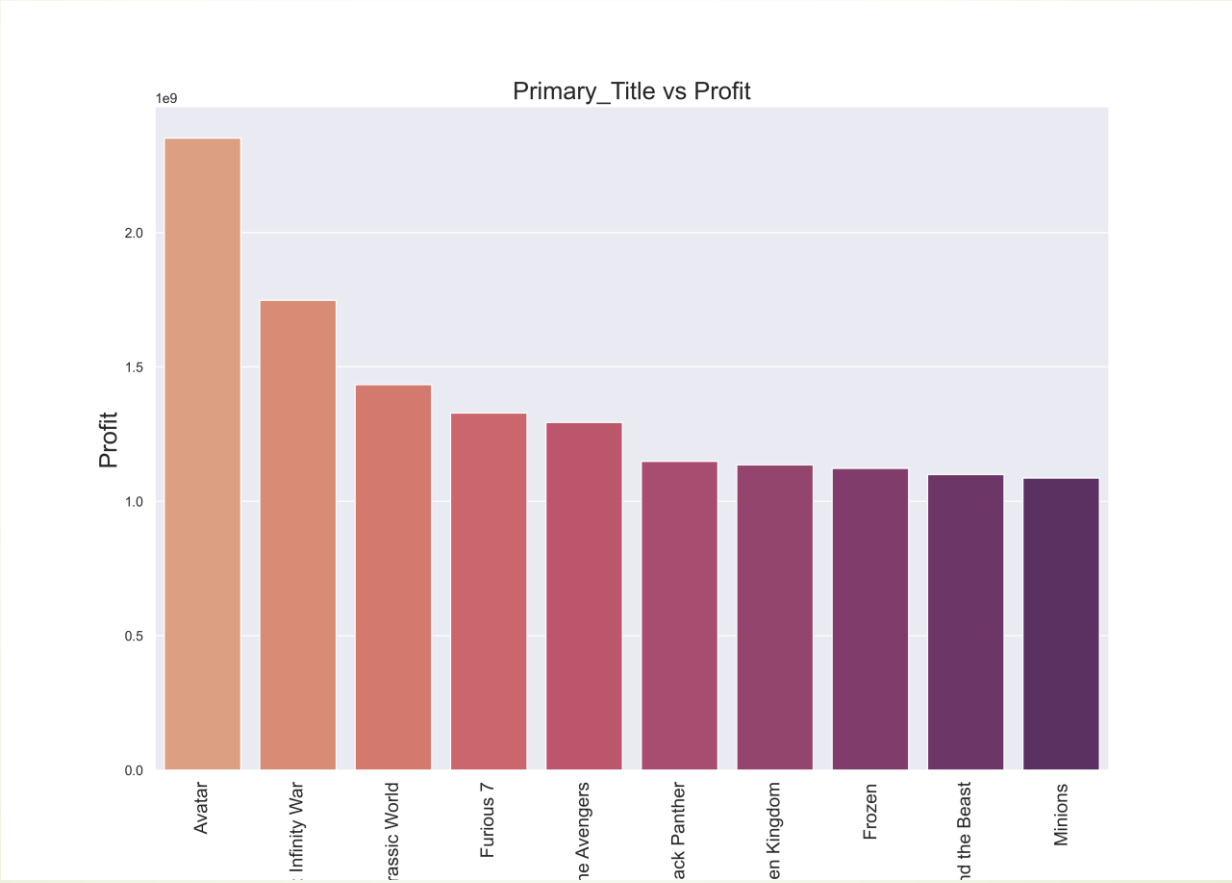
  The choice was informed by the fact that the mean could be affected by outliers and thus could offset our values or rather may end up giving a false impression.
- Drop duplicates
- Confirm the data frames are all clean to proceed.

Data Analysis

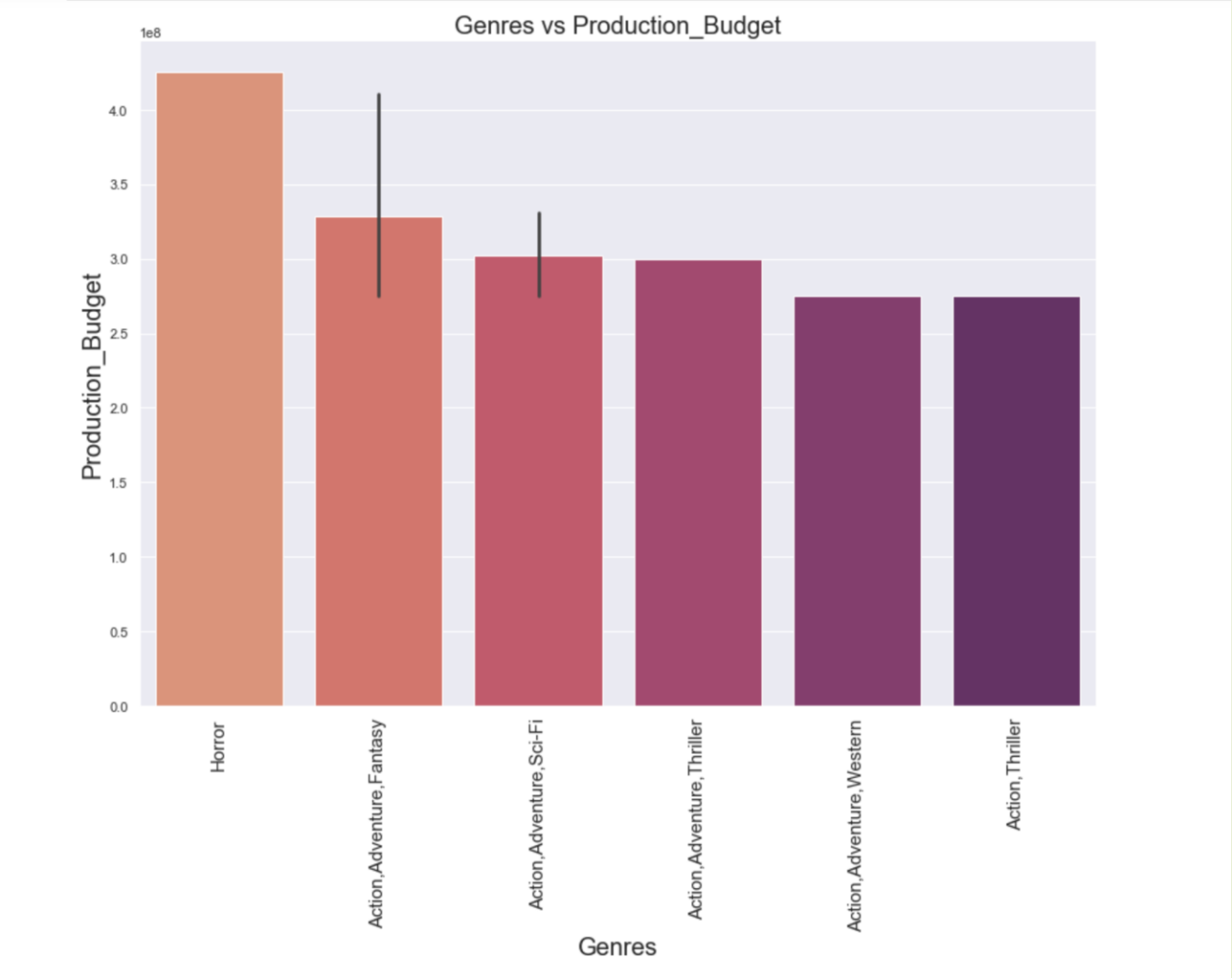
- Merge the relevant tables on a unique identifier contained in both.
- Remove symbols from integers so as to perform mathematical operations seamlessly
- Add important columns that can help derive more useful information from the tables.
- Ensure there are no duplicates in the final merged data frame.

Questions with their answers in form of plots

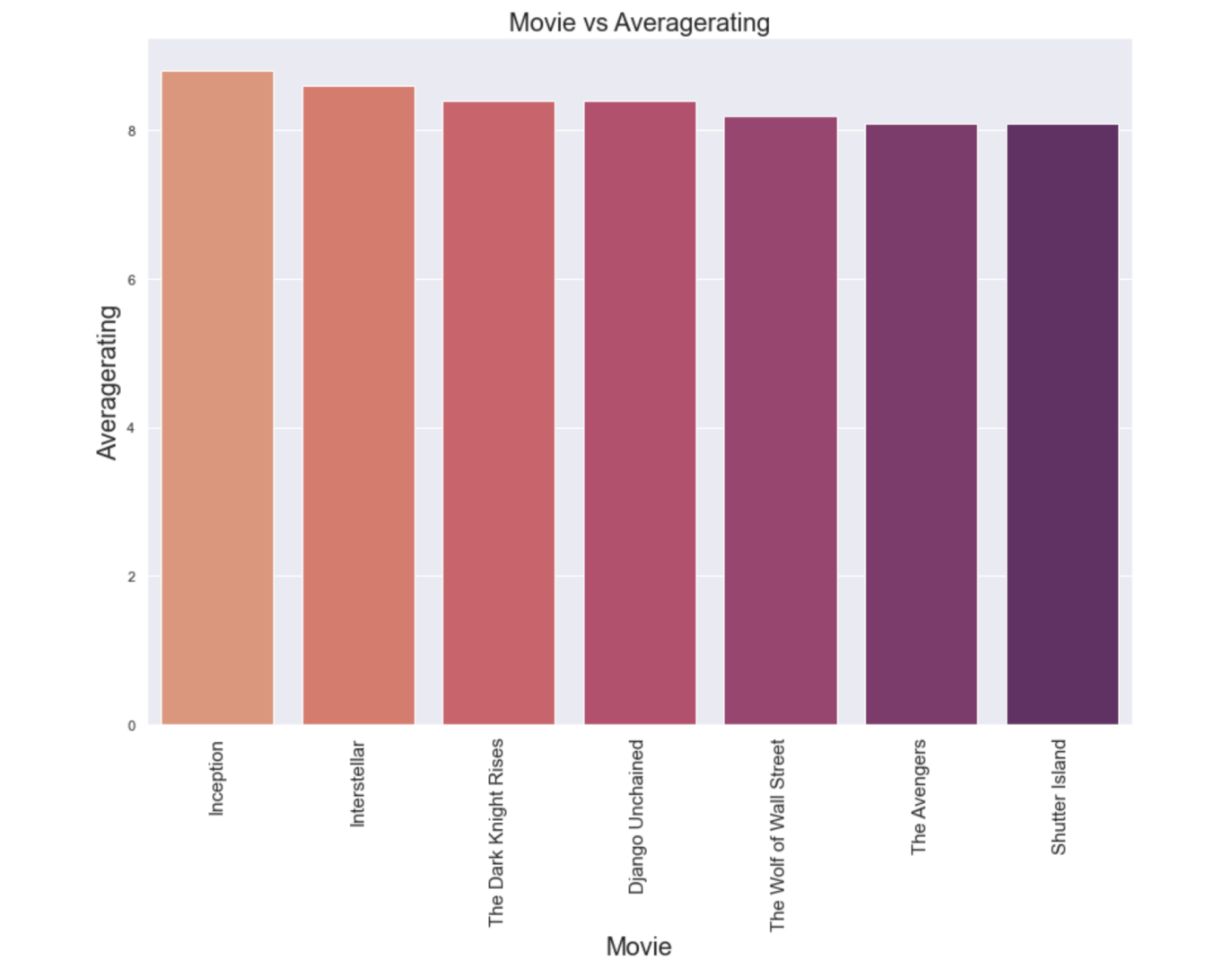
Q1: Which are the top 10 profitable movies?



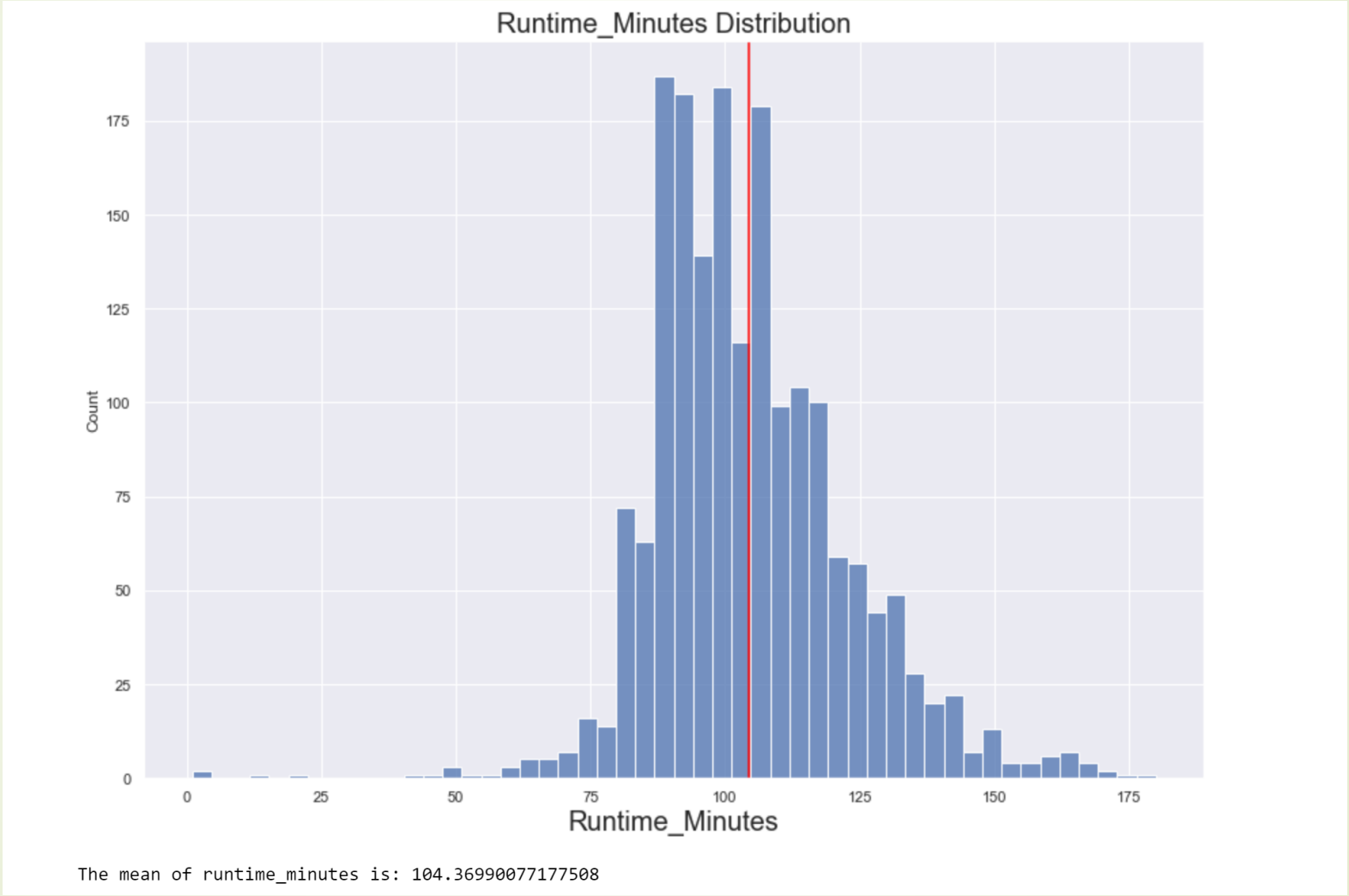
Q2: What genre of movies have high production budgets?



Q3: What are the top-rated movies?

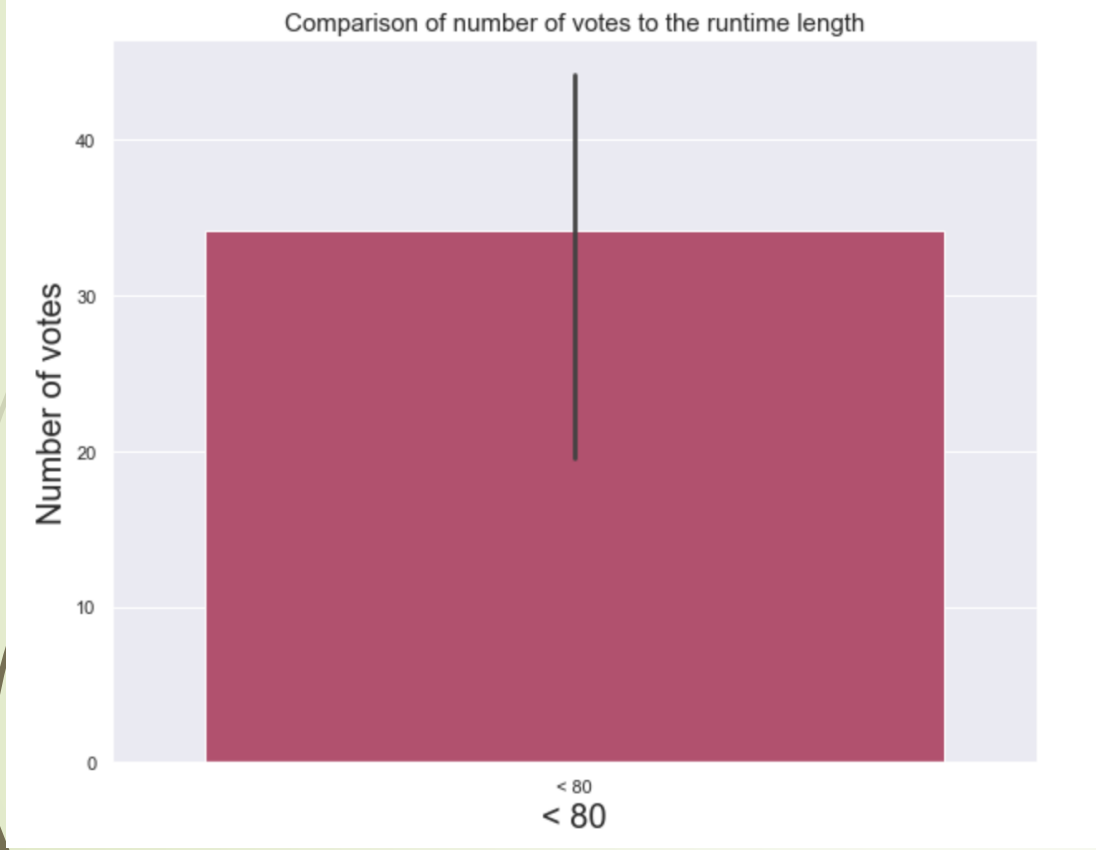


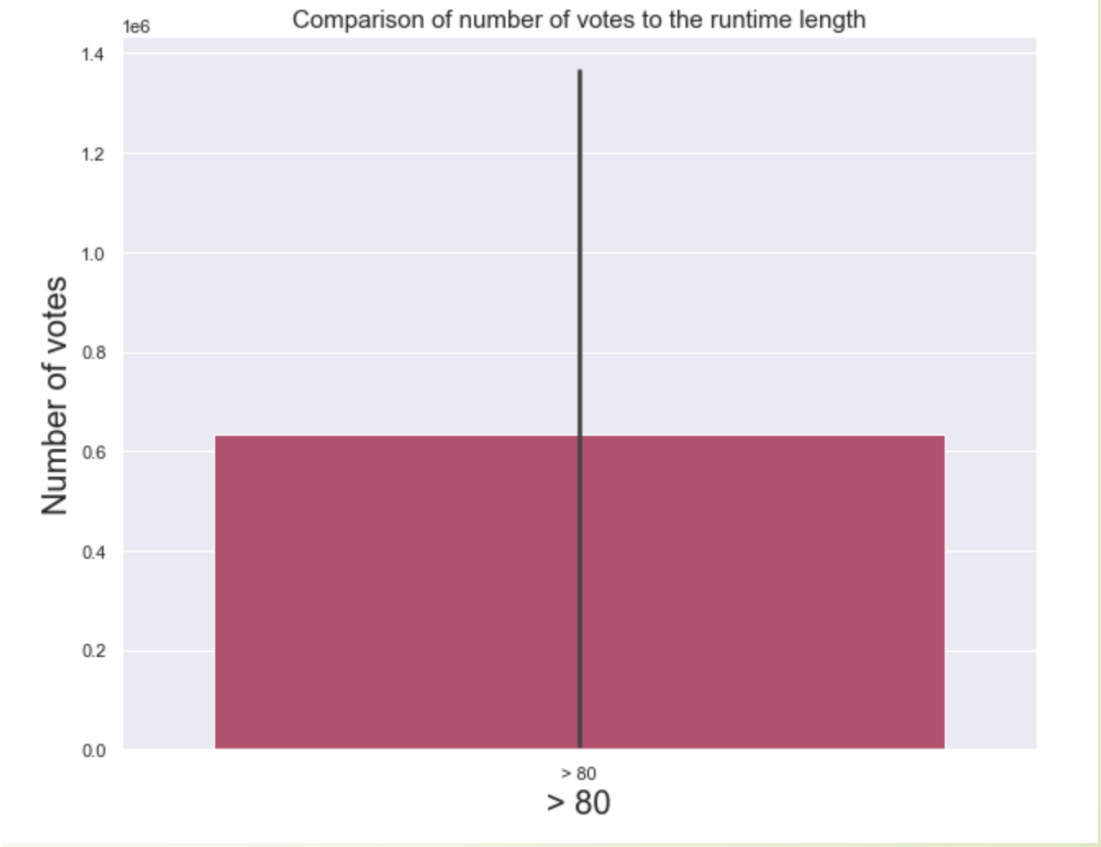
Q4: What is the duration of majority of the movies?



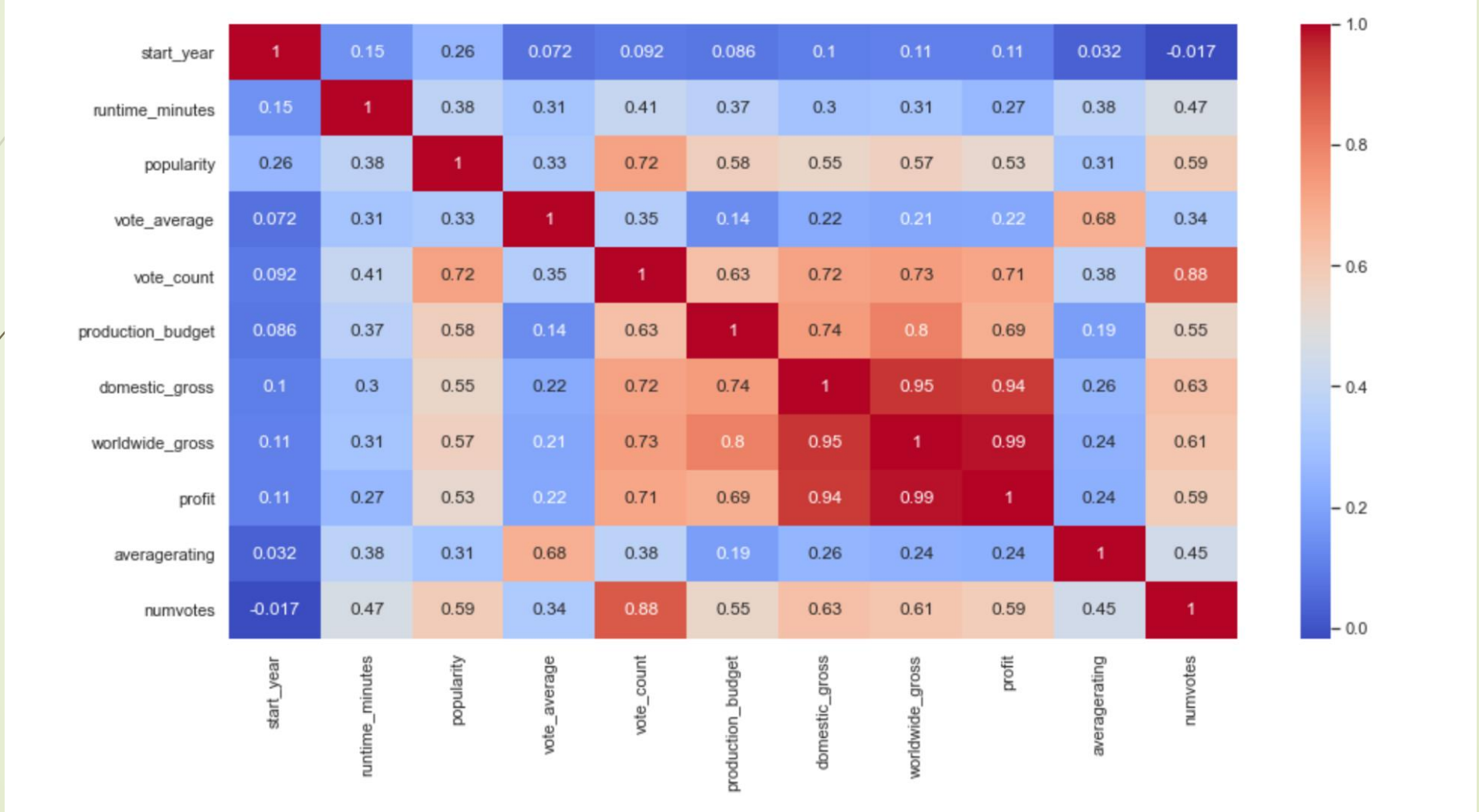
The mean of runtime_minutes is: 104.36990077177508

Comparison of number of votes to runtime length

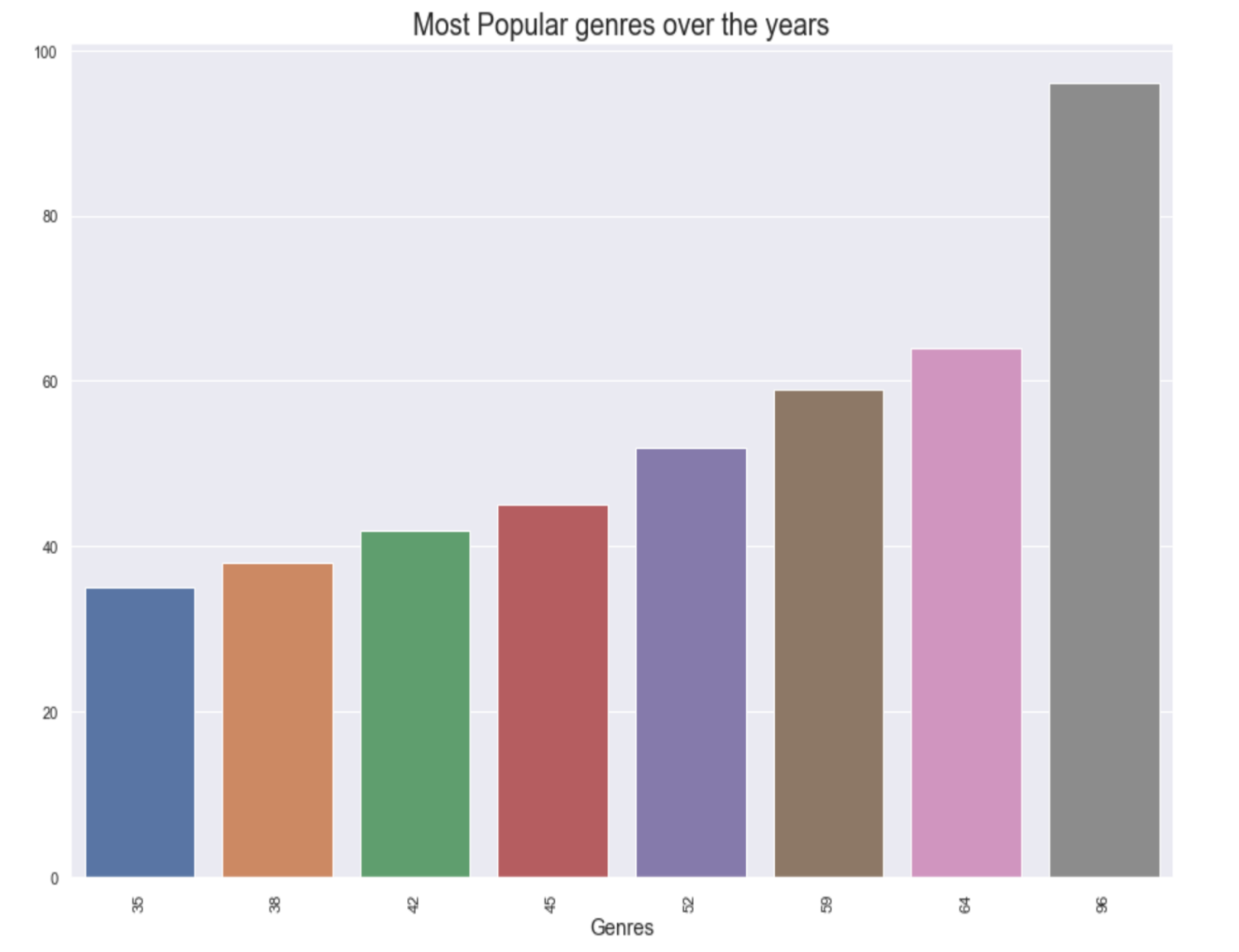




Q5: What is the correlation between the attributes of the given dataset?



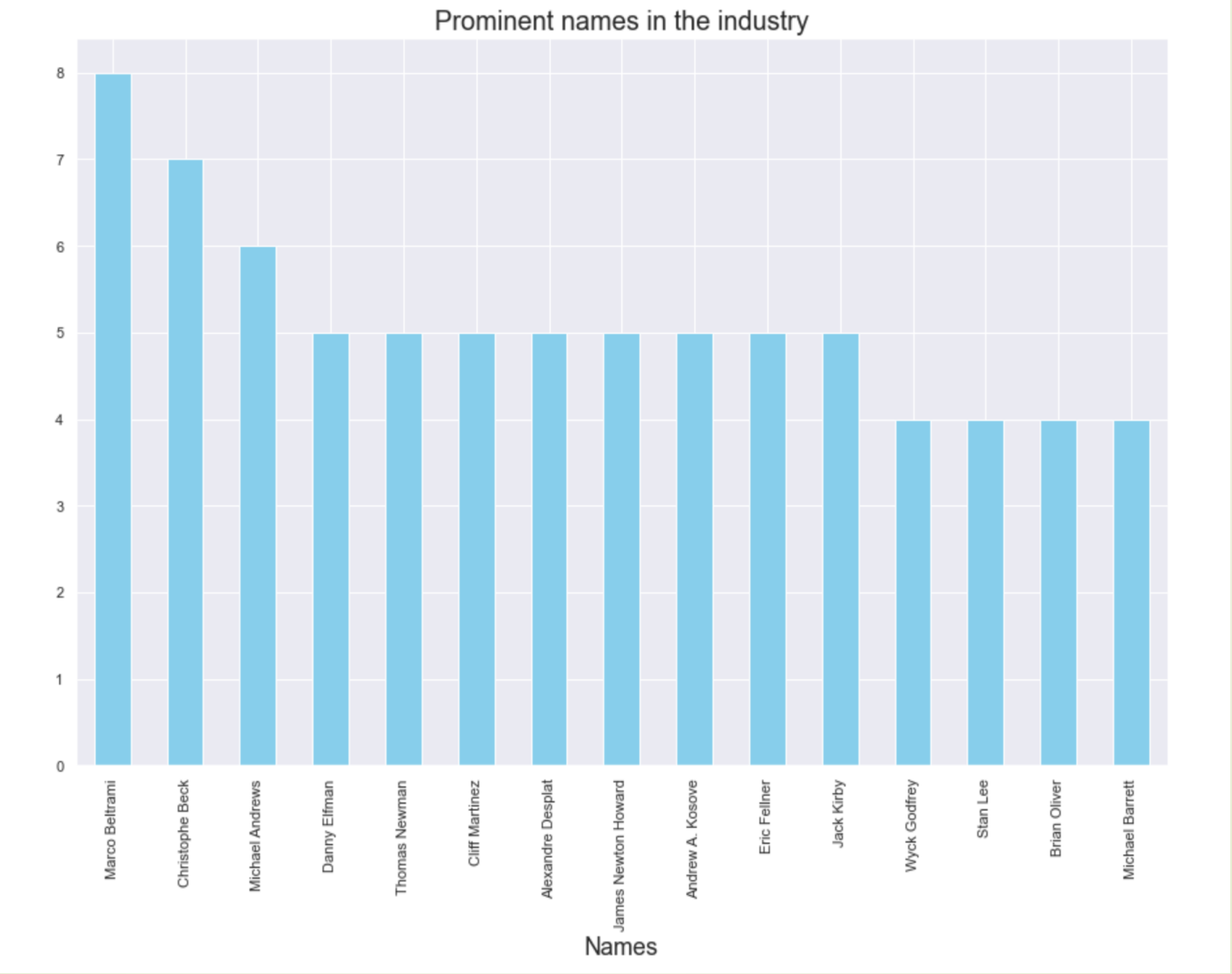
Q6: Which is the most popular genre over the years?



LEGEND

Drama……96
Comedy, Drama……64
Comedy, Drama, Romance……64
Adventure, Anime, Comedy……64
Comedy……59
Action, Adventure, Sci-Fi……52
Drama, Romance……45
Comedy, Romance……42
Action, Crime, Drama……38
Documentary……35

Q7: Who are the most valued names (people) in the movie industry?



Conclusions

Based on Qn 1:

- The graph displays indicators of the revenue that the company may attain from the production of such kind of movies.
- It forms a guide that could lead the company down the right path if it releases similar kind of movies and even provides a benchmark to build on and work with moving forward.

Qn 2:

- The bar plot just gives a general idea of where to allocate most funds of the production budget based on the genres being produced.
- It looks like horror films take up a major chunk of the money. This is important so as to allocate funds appropriately and plan accordingly in terms of the release date.
- For example, horror films are majorly saturated during the Halloween season and it would be ideal to release them at such times.

Qn 3:

- The bar plot displayed mainly provides insight on what the audience feels about the films. Higher ratings are obviously associated with good movies and vice versa. This is a plus because it is feedback from the audience. Both the fan base and critic base are useful gauges of the kind of movies you would want to create.

Qn 4;

- A majority of the movies have a runtime centered between 80 and 125 minutes with the average value being 104 minutes.
- The length of the movie influences the audience's attention.

- Very long movies might end up boring a person and conversely, short movies may just hasten the storyline hence not fun to watch.
- This might not always necessarily be the case depending on the target audience's preference or taste in movies.
- While this is the case, we may arrive at a general conclusion that there is no ideal length of movie runtimes.
- But, it would be great to base and kick off with the most frequent runtime followed by most movies as depicted above.

- From the comparison of number of votes to runtime length bar plots, more number of votes have been cast for movies with a runtime of above 80 minutes.
- However, very few people showed interest in voting for movies with a lower than 80 minutes runtime length.
- This is a huge contrast and how the film audience reacts is fundamental to keep track of the productions.
- If a movie receives a high number of votes from people, it increases its popularity and thus results in the movie making more profit.

Qn 5;

- If the coefficient value lies between ± 0.50 and ± 1, then it is said to be a strong correlation. Therefore;
  - Popularity has a high positive correlation with vote count, production budget, domestic gross, worldwide gross, profit and number of votes.
  - This indicates that the more popular a movie is, the more likely it is to get votes from fans or critics. Also, the more popular a movie is, the more money it makes.

- Essentially, these tables have a high correlation with one another. It is advisable to capture the film audience as we can derive a lot of information from these statistics.

- The number of votes are equally important to keep in mind as well as its popularity because these will generate more revenue for the company from the fan or critic base.

Qn 6;

- Drama is the most popular genre over the years as well as its combination with other genres.
- For example, the second most popular combination of genres is Comedy, Drama.
- In totality, drama, comedy and romance appear to have the most instances as the popular genres.
- This data may help in prioritizing the genre of movies to film and release in comparison to poor performing genres.

Qn 7:

- Marco Beltrami is a composer in the music department and has worked on movies like 'The Wolverine', 'Carrie', 'Afraid of the Dark', 'A Good day to Die Hard' etc.

- Michael Andrews' primary title is in the music department also, as a track composer and has worked on the sound track of movies like 'Neighbors', 'Bad Teacher', 'She's out of my league' just to mention but a few.

- Andrew A. Kosove falls in the category of a producer and has worked on movies like 'Blade Runner 2049' and 'prisoners'.

- Stan Lee is a renown writer especially known for 'X-Men: Days of Future Past', 'The Amazing Spider-Man 2', 'Ant-man and the Wasp'.

- These are key people that may elevate the nature of the movie in their various sectors of expertise providing next level, top-notch, highly profitable movies for the benefit of the company.

Thank You for Your Attention!

Questions?