Group 3 – Project 1

The analysis seeks to answer the following key questions:

* Is a movie's revenue dependent on the ratings it earns?
* Is the count of votes a key factor towards a movie's ratings?
* Are there some genres that earn higher revenue despite receiving lesser votes or ratings?
* Is there a correlation between a movie's ratings and revenue and count of votes and revenue?

**Movie’s revenue vs OMDB rating and count of votes**

**Data Analysis:**

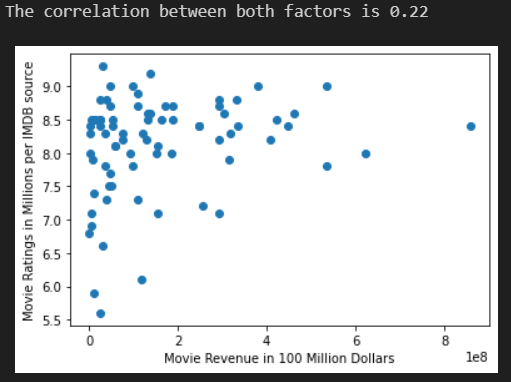
For the purpose of performing the analysis, a sample dataset of 76 movies released during the period 1990 to 2023 were reviewed.

Following observations were made:

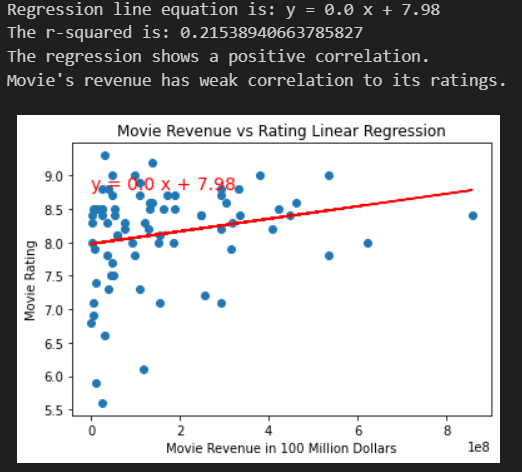
* Average count of votes a movie received was 1M.
* Average revenue per movie was $158M.
* Average rating was 8.0.
* In terms of revenue:
* the genre ‘Action, Drama, Thriller’ was the top earning revenue, followed by ‘Animation, Adventure, Drama’,
* while the categories ‘Crime, Thriller’ and ‘Comedy, Drama’ did not do very well at the Box Office.
* In terms of count of votes, it was observed that:
* the genre ‘Drama, Romance’ earned the highest number of votes but this category stood at 6th rank (out of 44) for the revenue it earned.
* the genre ‘Drama’ earned second highest count of votes but it stood at 25th position in the ranking.
* In terms of rating, it was observed that:
* the categories that earned the highest revenue were rated lower compared to other categories i.e., rated 8.0, while the highest rating observed in this dataset was 9.3.
* From the statistical analysis we observed there is a positive but weak correlation between movie’s revenue and its ratings or count of votes.

Following statistical correlation analysis were performed:

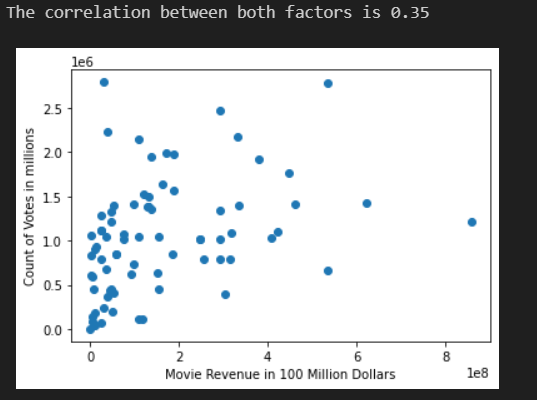
* St Pearson Correlation Analysis: Analysis was performed between movie’s revenue and rating; the result was 0.22 which indicated weak positive correlation between the two.

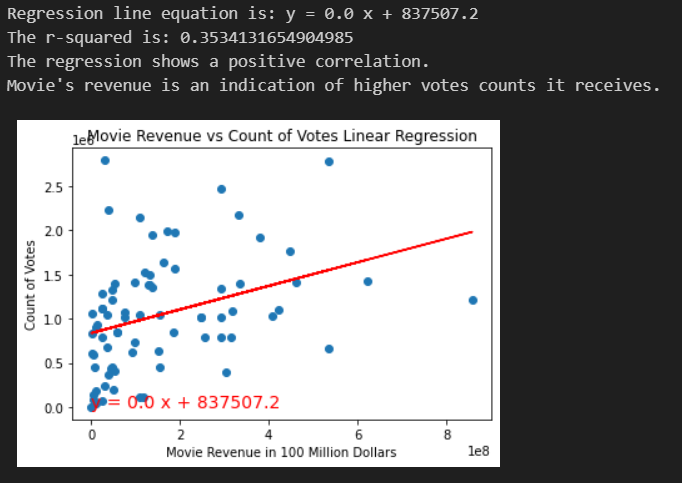


* Similarly, the r-squared value was 0.21 which returned the same observation as above:

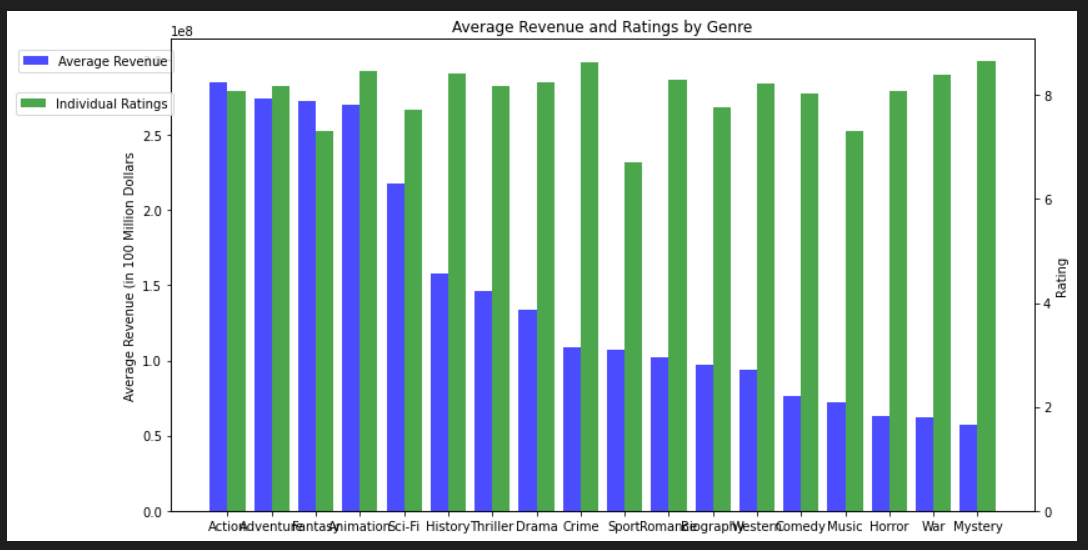


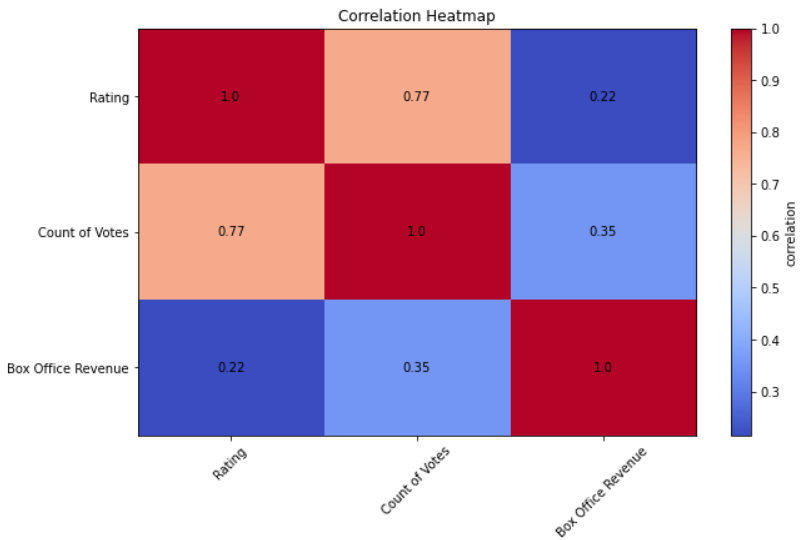
* St Pearson Correlation Analysis: Analysis was performed between movie’s revenue and count of votes; the result was 0.35 which indicated weak positive correlation between the two.



* Similarly, the r-squared value was 0.35 which returned the same observation as above:
* 
* We observed that whilst high revenue earning movies could receive votes and ratings but these are not the deciding factors towards a movie’s performance or earnings. The bar chart below explains this position.

Example: Categories such as Comedy, Music, Horror, War, Mystery earned ratings equivalent to top revenue earning categories but yet they were in the lowest tier of revenue earnings.





**Hypothesis:**

**Null Hypothesis:** There is no correlation between a movie's revenue and its ratings or count of votes.

**Alternate Hypothesis:** Count of votes and movie's ratings contribute to movie's revenue and success

A one t-test was performed to assess if the alternate hypothesis holds true. The t-test result showed pvalue of 0.007 (which is less than 0.05) which means the null hypothesis is rejected. In other words, the alternate hypothesis that the count of votes and movies ratings are correlated to movie’s revenue and performance.

**Conclusion:**

From the data analysis and various statistical outcomes, we can derive the following conclusions:

* There is positive correlation between count of votes and movie’s revenue and ratings and movie’s revenue.
* The correlation is however weak i.e., though there is a correlation, the count of votes or ratings are not ultimately strong factors leading to a movie’s success. In other words, viewers do not necessarily submit votes or ratings even when they made the spend to go watch the movie.
* However, there are some viewers who based on ratings or count of votes, decide to go watch the movie or recommend others to watch the movie.
* There t-test value of less than 0.05 also rejects the null hypothesis, which confirms the conclusions made above.