Vincent Freger and Matt Daley

Professor Carter

December 9, 2024

MA 705 – 1

Project 3 – Analysis of Movie Ratings by Genre (Movie Ratings Merge)

Report Focus

How do viewer-given movie ratings compare for the three most popular genres (Action, Comedy, and Drama)? Is there a statistical difference among them? Is there a practical difference?

Executive Summary

Full year 2016 movie rating data was analyzed for the three most prevalent genres of Action, Comedy, and Drama (full file was analyzed to determine 3 most prevalent based on counts, and only those three were selected for the merge with ratings due to memory constraints). It was concluded that the Average Rating for each genre is very similar to each other (Please see Table 1). The very minor difference in average ratings is negligible from a practical point of view. When running a statistical test to see if there is a significant difference between any of the three means (average ratings), it showed that there is significant difference (Kruskal Wallis Test was chosen because the ANOVA test assumptions were not met).

Background and Methodology

Two files were selected on www.kaggle.com that were created and maintained by GroupLens research group at the University of Minnesota. The first one contained information about movies including the title, genre, release year, revenue generated, and more, all data at the movie level of approximately 45,000 movies going from the 1970s to July 2017. For the same period, another file collected viewer-generated ratings for movies (at the rating level, as a typical movie would have multiple ratings from many viewers, thus creating one-to-many relationship in the merge). Only 2016 movie ratings were selected to obtain a current state of ratings for a full year and to manage the size of files. At the same time, counts-by-genre was produced and only the top 3 genres by count were selected to manage the size and go around full-memory error. These genres were the subject of interest anyway and are very popular genres. The merged analysis file was produced by joining the two above files on movie_id. The average ratings were calculated using a group by genre function (shown on Table 1 and Chart 1 below). Statistical tests were performed including running the assumptions for one-way ANOVA, which were not met, and Kruskal Wallis was selected for significance testing. The work was done in Deepnote, and the link is provided.

Table 1: Numeric Summary of Movie Ratings by Three Most Popular Genres

	mean float64	count int64
Action	3.7	87634
Comedy	3.64	115528
Drama	3.63	203449

Chart 1: Graphical Summary of Movie Ratings

Average Movie Rating by Genre
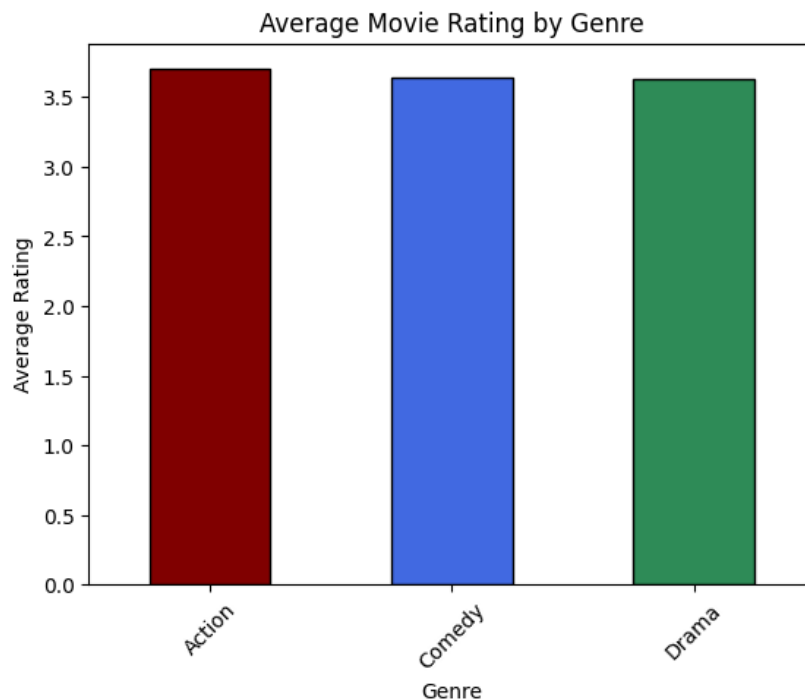
Statistical Testing
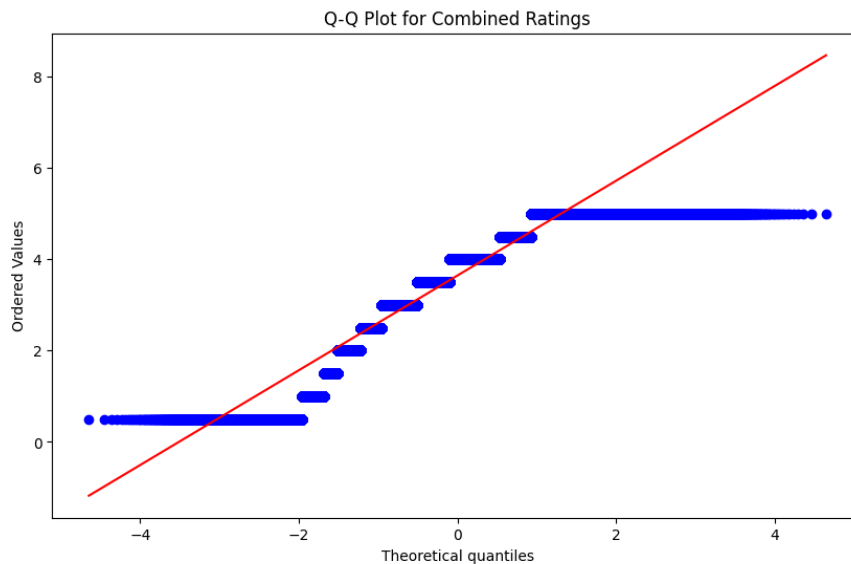
One-Way ANOVA assumption check

1. Independence – This assumption is reasonable because the movies being analyzed do not influence each other, and there is only one genre per movie (mutually exclusive)
2. Homogeneity of Variances – This assumption is not met because the variances are not equal based on Levene's Test. The p-value is less than 0.05, indicating unequal variances, shown in the summary below.
   Homogeneity of Variances:
   ```
   LeveneResult(statistic=15.853692427545809, pvalue=1.3034587764149548e-07)
   ```
3. Normality – This assumption is not reasonable because the Q-Q Plot below shows deviations from the red line.

Chart 2: Q-Q Plot for Normality

Because the assumption check has failed, the ANOVA could not be used to determine statistical significance. Given that, a Kruskal Wallis Test was performed, and the result showed a p-value of less than 0.05 (a standard significance level chosen). As a result, the Null Hypothesis that the means (Average Ratings) are equal, was REJECTED, indicating that at least one Average Rating is different. Below are the formal hypotheses and result of the Kruskal Wallis Test.

NULL Hypothesis: median(Action) = median(Comedy) = median(Drama)

Alternative Hypothesis: At least one median rating of genre is different

```
KruskalResult(statistic=370.50197629083846, pvalue=3.5198007647980167e-81)
```

Conclusions and Recommendations

While the statistical test shows Action to have a significantly higher Average Rating than Comedy and Drama, the visual test shows that the difference is so small that it has no practical implications (3.70 vs 3.64 and 3.70 vs 3.63). The reason for statistical significance is because of sample size, where each genre has close to 100,000 ratings and often more). It is recommended to gather 2017 through present period data to repeat the analysis, and to also look at trend of ratings by year.

Link to Full Data Processing and Analysis in Deepnote below

[Movie Rating Analysis](#)