A picture containing text, font, screenshot, graphics

Description automatically generated

*Report: Film Analysis*

Name: Mahenoor Syed Abdul Haq Sahib Name: Yashwini Dilip Moolchandani

Student ID: 2202970 Student ID: 2203066

Name: Jalila Farida Syed Abdul Haq Sahib Name: Huzaifa Bin Asim

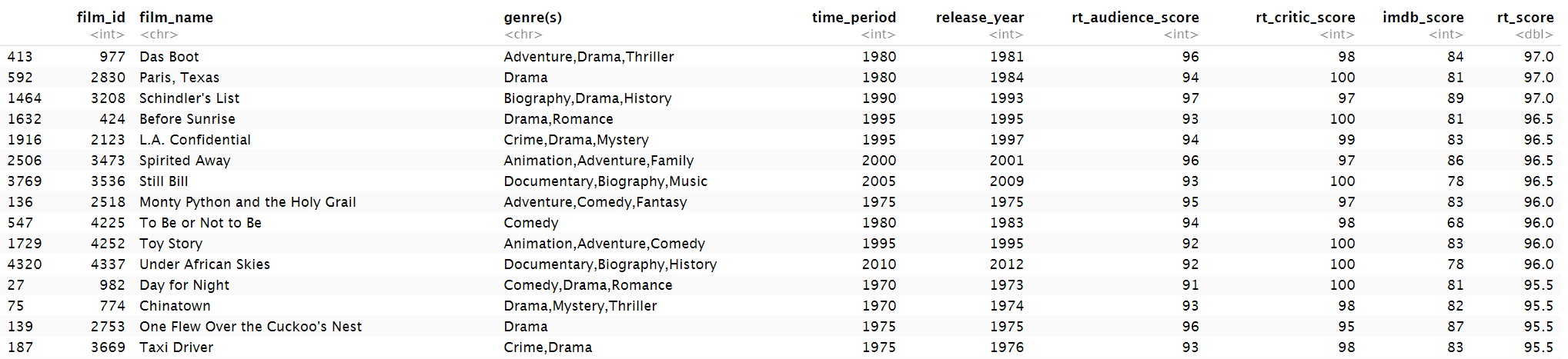
Student ID: 2202971 Student ID: 2202906

Name: Burchin Mutlu

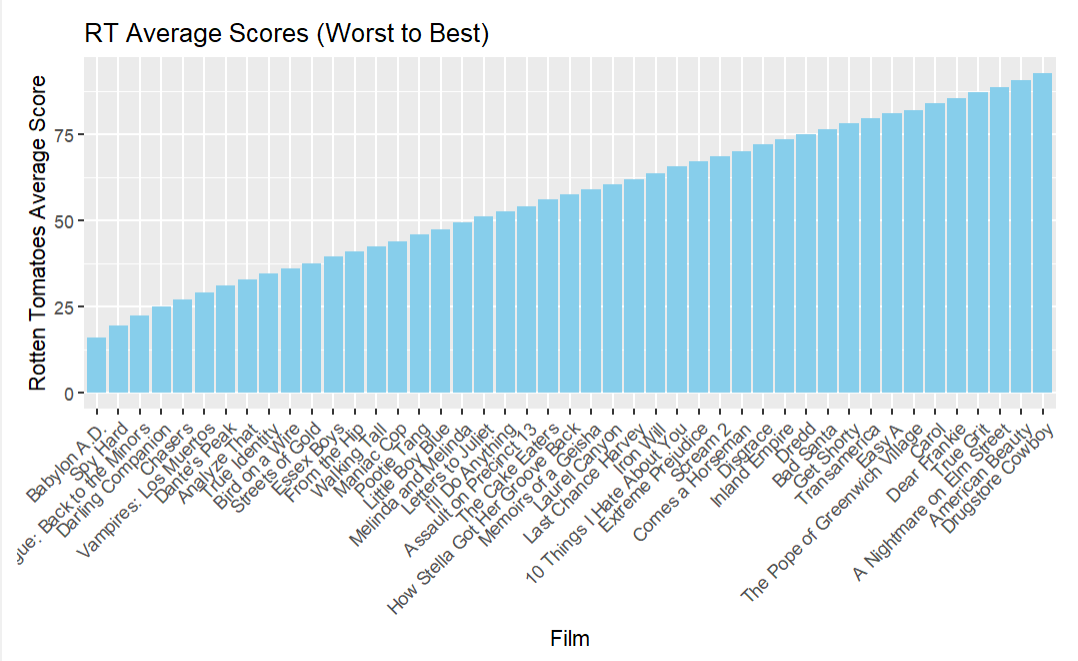
Student ID: 2203064

# REPORT: PROGRAMMING AND DATA VISUALISATION EXAM PROJECT

1. After combining the tables in the given database and embedding the SQL query as R chunk, we came across four tables that needed to be analyzed.
   * rt\_audience\_score (Rotten Tomatoes movie scores given by audience)
   * rt\_critic\_score (Rotten Tomatoes movie scores given by site authors)
   * imdb\_score (Average IMDB scores based on audience ratings)
   * imdb\_likes (Movie likes per movie on IMDB, given by audience as well)
2. There are 4620 movies in total and according to these columns, 2594 movies have null values in imdb\_likes. There are several methods in order to fill those but for this Project, we disregarded as there are so many missing values that causes us not be able to work on. So our analysis consists of three key scores; rt\_audience\_score, rt\_critic\_score and imdb\_score.
3. Our approach to these values will be to evaluate the Rotten Tomatoes ones separately and the IMDB ones separately, because of the fact that each individual site has it’s algorithms in order to finalize and rate viewing results (like anti-spam features on their web applications just to prevent nationalist spam botting etc.) Before that, please check our scatter plot on step 19 ( row 107) about these scores.
4. We’ve found average scores for Rotten Tomatoes (rt\_score= (rt\_audience\_score + rt\_critic\_score)/2)



1. Here is a bar chart about rt\_score, ordered by worst to best movies.



A picture containing text, screenshot

Description automatically generated

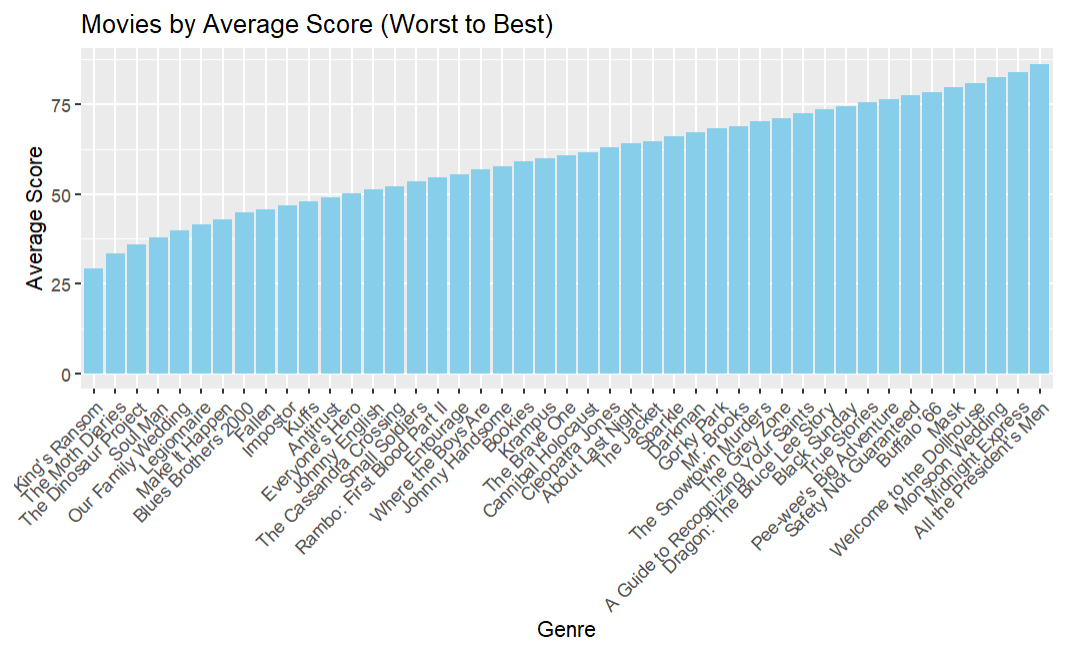
1. A scatter plot between imdb\_scores and rt\_score. Looks more linearly distributed than RT scores.

1. In order to find a final score, we averaged these for each movie. (avg\_score = (rt\_score + imdb\_score)/2) After dropping the unneecessary columns, here are the results of best to worst movies.

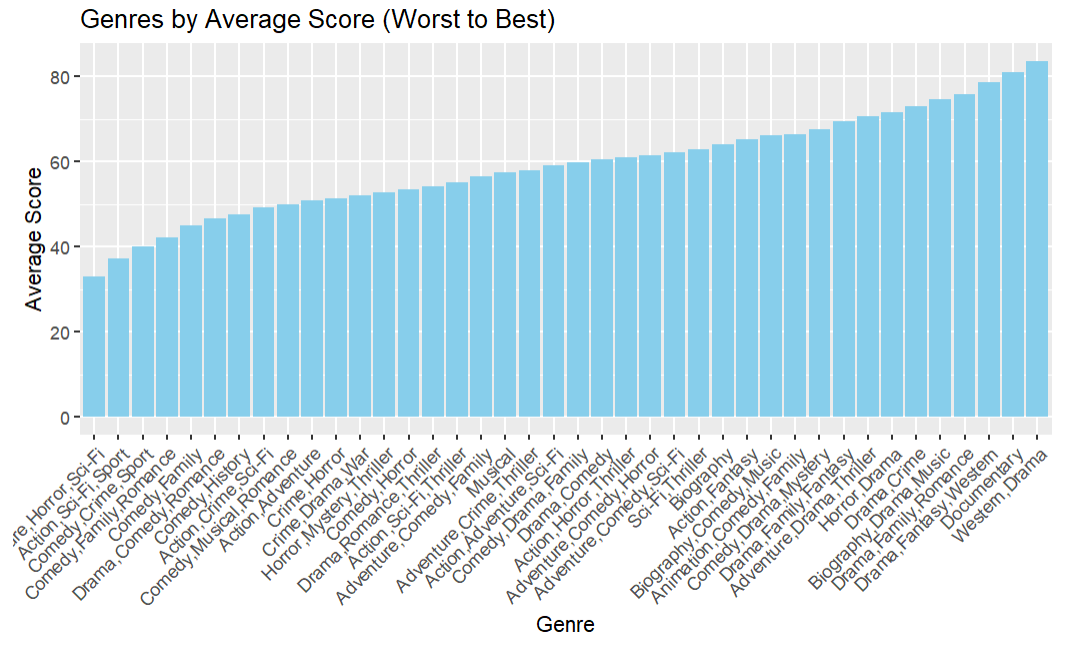


Worst to best movies depending on avg\_score.





1. To analyze the genres, we created another dataset in order to average genre, depending on avg\_score. Here are the results  
   



# RESULTS

* According to these results our best movie is “The Shawshank Redemption” and our worst movie is “Alone in the Dark”
* Also worst genre combined, according to our foundings are “Family, Horror, Mystery” and best genre are “Drama, Fantasy, War”.