**SUMMARY FOR THE RSVP MOVIES**

**SEGMENT 1**

For the first segment, we looked into the dataset and its shape, by finding the number of rows and number of null values in the columns in each table. Then we looked at the frequency of the movies releases, compartmentalizing it both my monthly and yearly. We also looked at the number of movies produced in countries such as India and United States. We also looked at different genres present in the data set which includes: Drama, Fantasy, Thriller, Comedy, Horror, Family, Romance, Adventure, Action, Sci-Fi, Crime, Mystery and Others. Then we looked for the genre which comprises the greatest number of movies, which was Drama. Which in turn means that RSVP should focus more on “Drama” genre. Then we looked into filtering out the movies that belong to only one genre. We also found the average duration of movie in each genre. We also used rank function to find the rank of the genre based on the number of movies produces in each genre.

**SEGMENT 2**

Here we looked at the data more discretely. We devised a query to find the maximum and minimum of average rating, total votes, and median ratings to calculate the outliers in the table. On the basis of average ratings, we found the top 10 movies. We summarized the ratings table based on the movie count by median ratings. After that we looked at the top production houses based on the number of hit movies they’ve produced which was solely defined by the condition of average ratings.

**SEGMENT 3**

Here we analyzed the names table by checking the null values in each column. In this segment we found the top three directors of all time corresponding to the top three genres we found in the previous segment. We also looked at the top actors and actresses by filtering them out based on their movie’s median ratings. Previously we looked at the production houses based on the number of hit movies, but in this segment we sorted them out based on the number of votes received. Then we targeted actors and actresses of India based on their hits. Then we used case statements to rate the thriller movies based on certain conditions.

**SEGMENT 4**

In this segment we looked at the broader picture of the dataset. Supposedly by focusing on the average duration, running total of the average duration and the moving average duration. Then we looked at the top 5 movies of all time based on the top 3 genre which we have already devised in the previous segment. Then we tried something different where we were required to use the POSITION clause in order to find the multilingual movies. At last, we concluded with the details of the top 9 directors in the whole movie dataset.