Kristina Frazier
Project Submission: Investigate a Relational Database

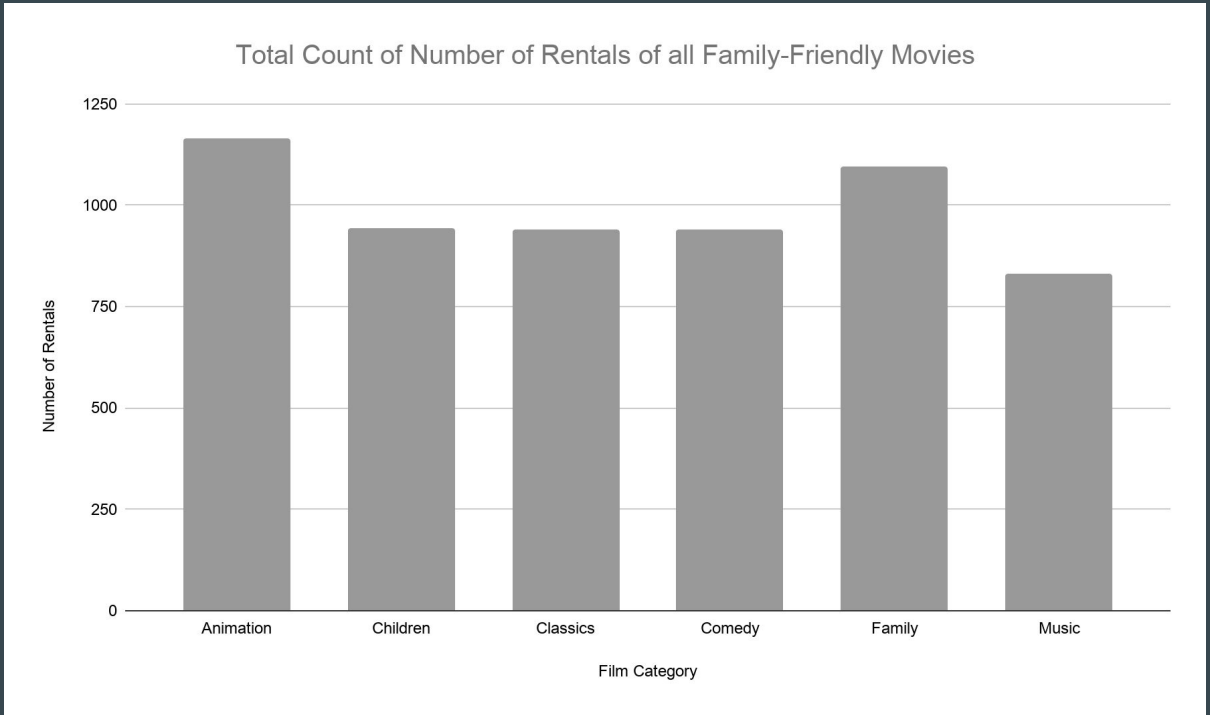
•••

Programming for Data Science with Python Nanodegree Program

Question 1

Question: We want to understand more about the movies that families are watching. The following categories are considered family movies: Animation, Children, Classics, Comedy, Family and Music. Create a query that lists each movie, the film category it is classified in, and the number of times it has been rented out.

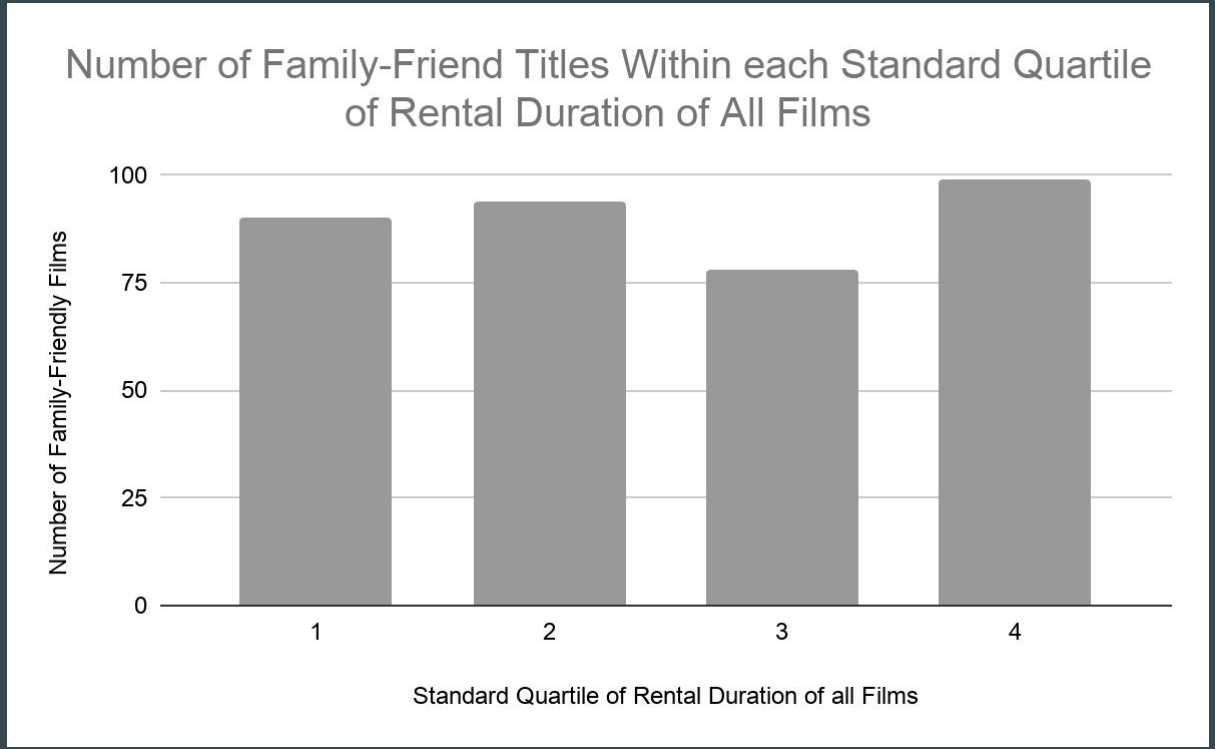
Solution: See “Query 1” in submitted text file. The query result lists each family movie film and the number of rentals for each film. The visualization at right shows an aggregation of total rentals of all films within a certain family category. As we can see, renters do not dramatically favor one type of film category or another. However, animation films have the most rentals than all other categories.; Although from this analysis we cannot be sure if this is because there are more films in the animation category than any other film.



Question 2

Question: Now we need to know how the length of rental duration of these family-friendly movies compares to the duration that all movies are rented for. Can you provide a table with the movie titles and divide them into 4 levels (first_quarter, second_quarter, third_quarter, and final_quarter) based on the quartiles (25%, 50%, 75%) of the rental duration for movies across all categories?

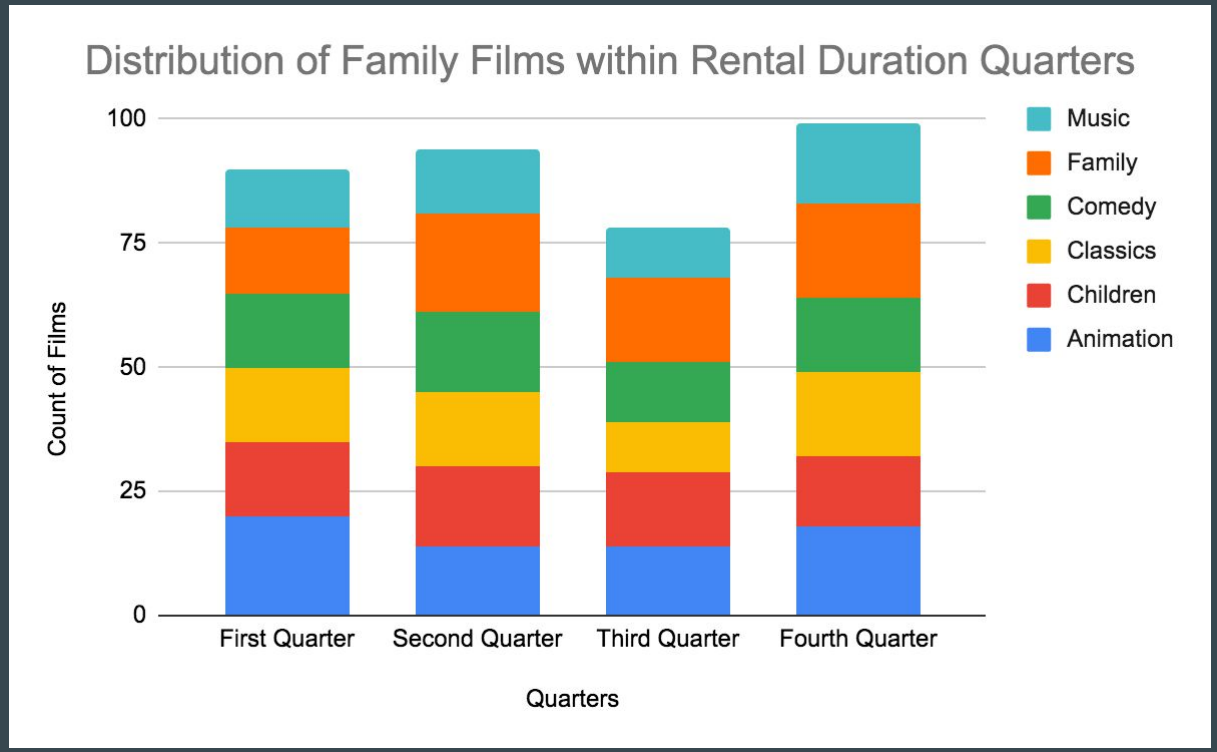
Solution: See “Query 2” in submitted text file. The resulting table lists each family movie, their category, and the quartile of their duration as compared to the rental duration of all films in any category. The bar chart at right aggregates the number of films within each rental duration quartile - the results are almost evenly split between each quartile, meaning that family movies are not dramatically rented for either longer or shorter durations when compared to all other films.



Question 3

Question: Finally, provide a table with the family-friendly film category, each of the quartiles, and the corresponding count of movies within each combination of film category for each corresponding rental duration category.

Solution: See “Query 3” in submitted text file. The resulting table lists the number of family films by category that fall within each quarter of rental duration across the entire film dataset. The clustered bar chart at right visualizes this distribution, showing that the the count of films within each family category is fairly evenly split among the quarters of rental duration, i.e. no category of family films is overrepresented in longer or shorter rental durations.



Question 4

Question: We want to find out how the two stores compare in their count of rental orders during every month for all the years we have data for. Write a query that returns the store ID for the store, the year and month and the number of rental orders each store has fulfilled for that month.

Solution: See “Query 4” in submitted text file. The query result shows the count of rental orders aggregated by month for each store. The visualisation at right shows that rental orders were similar for each store in each month represented in the dataset, but most of the time Store 2 had more rental orders than Store 1. Store 2 outperformed Store 1 in May 2005, July 2005, and February 2006.