Investigate a Relational Database

This is a project from the Progamming for Data Science with Python program, provided by Udacity

Question Set #1

In this question set, the customer was always trying to understand the data about the family-friendly movies from his movie rental

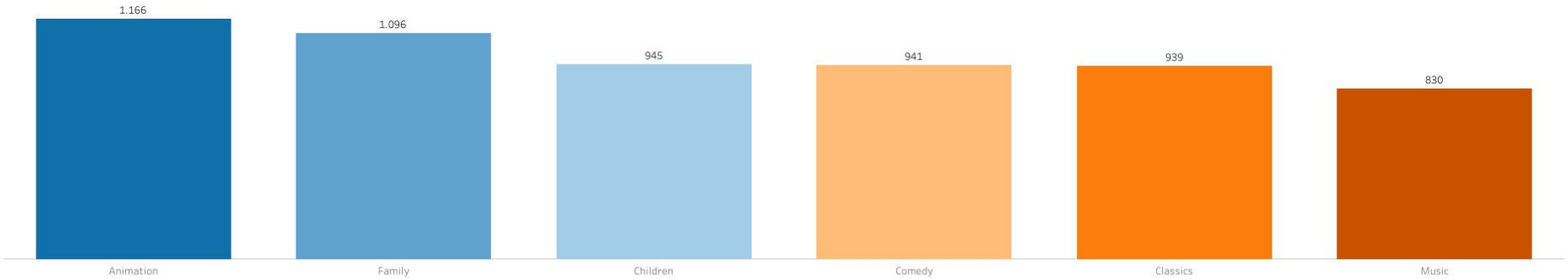
Question 1)

To answer this question, I used a subquery because I wanted to be able to count how many times a given movie was rented and then group the movies by its category.

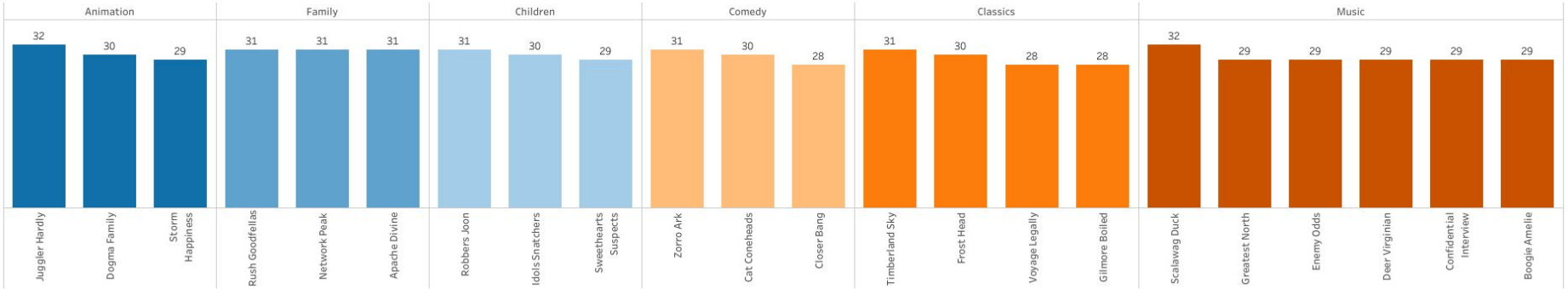
The result showed me the most rented categories and also the most rented movies by category. To better visualize the data, I have tried to print only the three most rented movies for each category.

The visualization can be seen in the next slide. I don’t know why, but when I download the image from the Tableau page, it was not downloaded on high quality and this was reflected in this presentation.

Rental count per category



The top rented movies by category



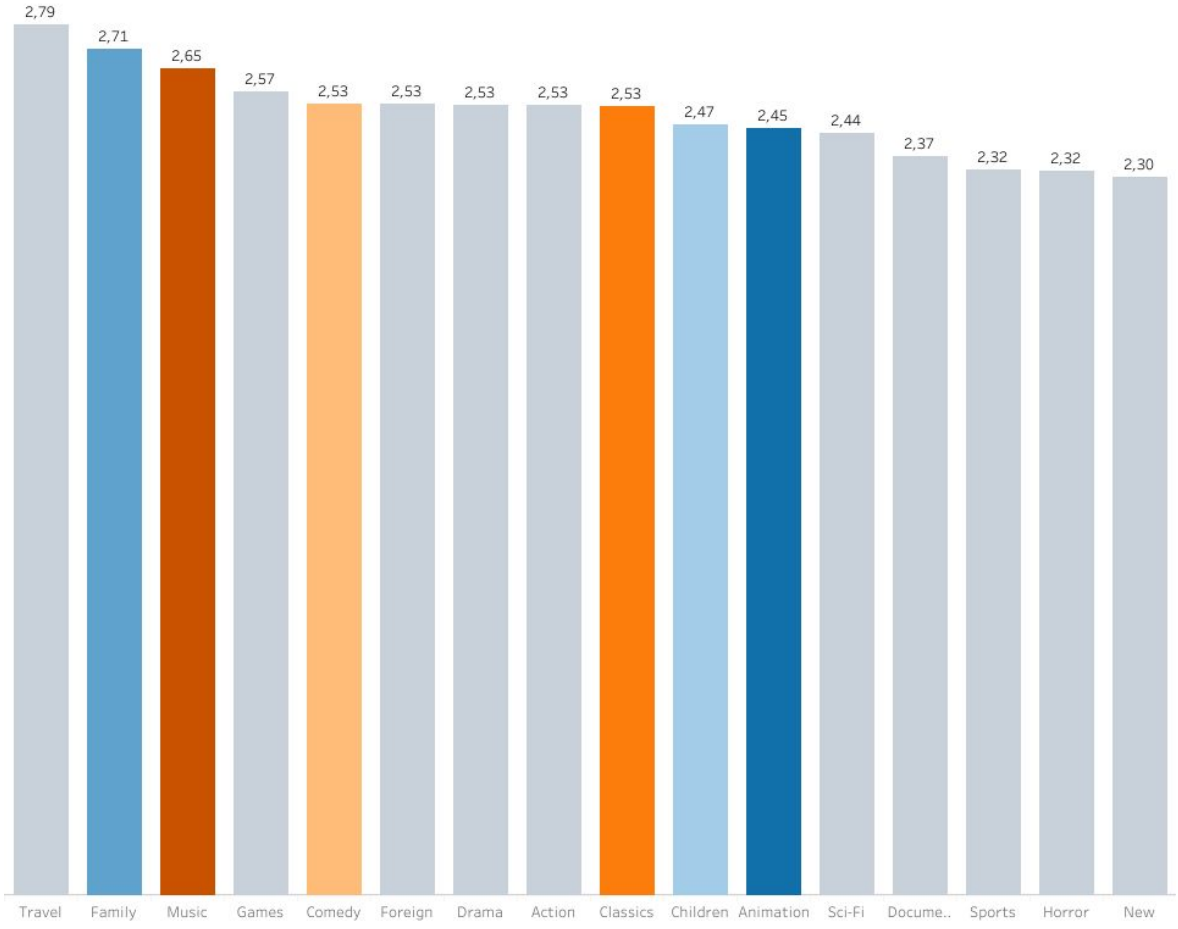
Question 2)

Here the code was pretty simple, using just a NTILE function. In the Tableau app, I choose to show the average standard quartile and color the family-friendly movies with the same color palette I used earlier.

The result showed that the rental duration of the family-friendly movies are, on average, high. In the 10 highest averages, 5 are of these type of movie.

Again, the visualization can be seen in the next slide, and this pattern will continue for the rest of the presentation.

Average rental duration quartile per movie category

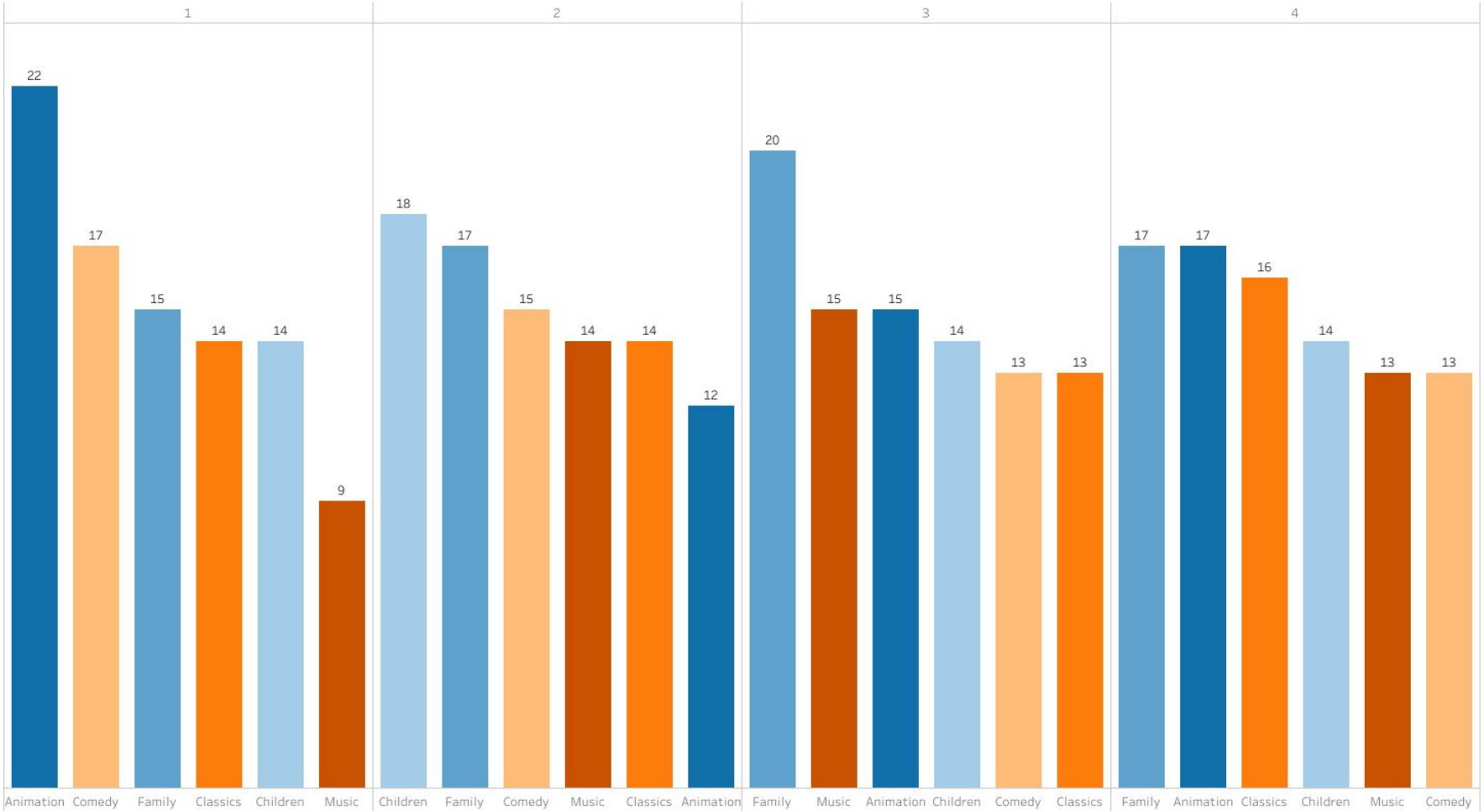


Question 3)

To answer this question, I used a subquery because I wanted to be able to count how many times a given movie was rented and then group the movies by its category and then by its quartile.

In the visualization we can see the most rented movie categories by rental duration standard quartile. In conclusion to this question set, we can see that the most relevant category is the Family category. It is one of the most rented, the average rental duration is one of the highest and, even analyzing in a most granular level of rental time, this category always appears on the top 3.

Movie category rentals per quartile



Question Set #2

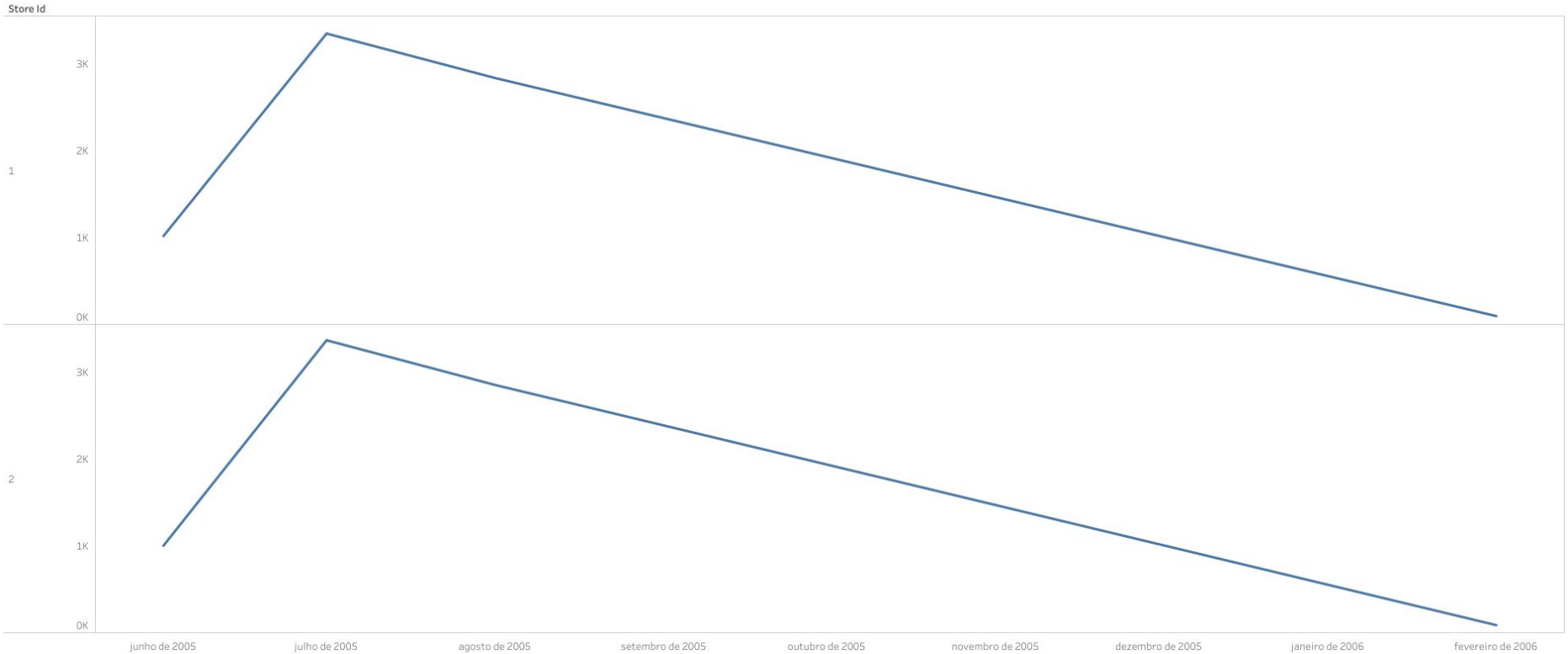
In this set of questions, the customer was trying to understand the financial part of his rental company

Question 1)

For this question, I used a simple COUNT function and grouped the results by date and then by store.

I wasn't able to Dual Axis the visualization, because to do it I needed two measure rows. However, it is possible to see that, for the same period, the rentals quantity was pretty much the same in both stores.

Rentals by month for each store

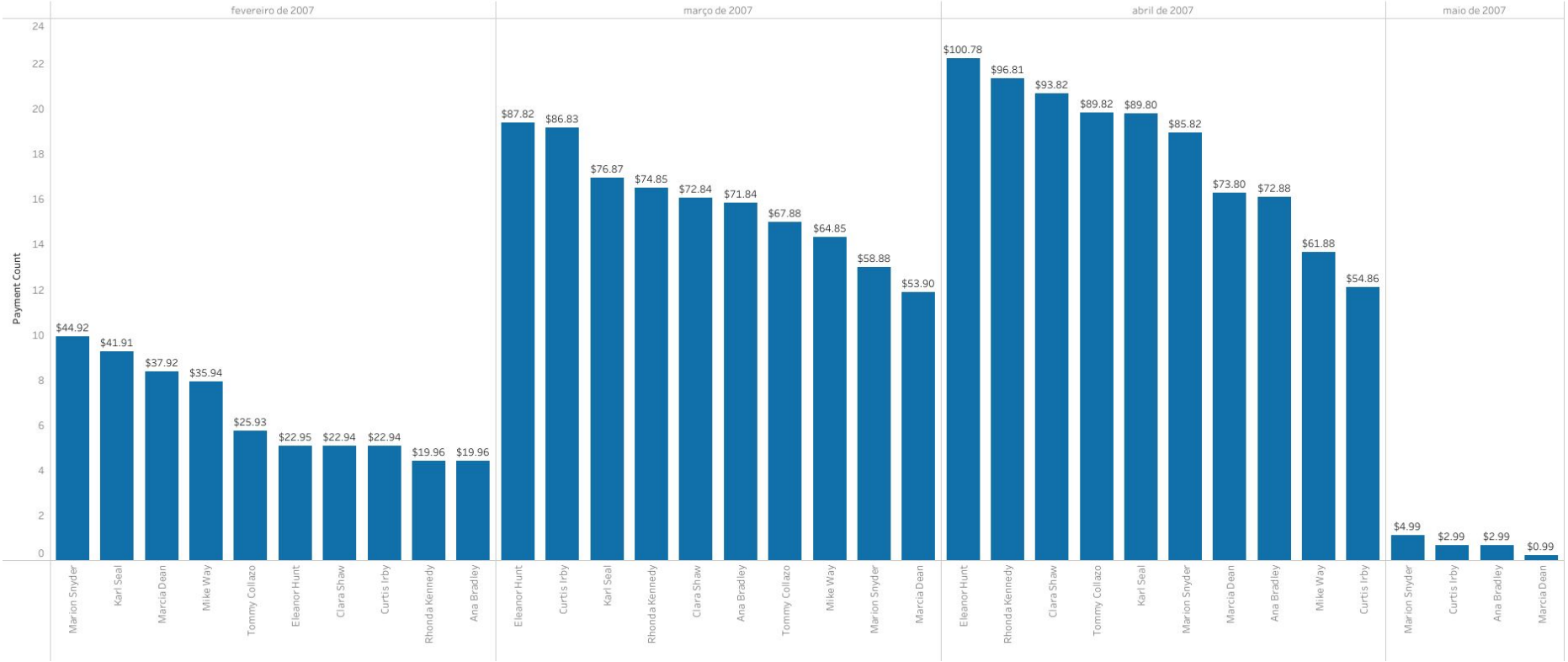


Question 2)

To answer this question, I used a subquery containing the full name of the renter and the sum of the total payment amount. In the final query I have counted the number of payments to discover the rental quantity and grouped it by month and then by customer.

It is possible to see that our customers were spending more and more until April, but something happened in May and the rentals plummeted.

How many times our top renters rented a movie and how much they spent per month



Question 3)

Here, I used two subqueries to find the full name, to sum the payment amount, to group by date, and in the final query I used a LAG function to see the difference between monthly payment.

The visualization will start in March, because February is our first record, so we can't compare it with no month before. By this result, we can see that there was a BOOM in March, but it starts to drop down and in May we have huge negative difference of payment amounts.

Difference of payment amounts by month