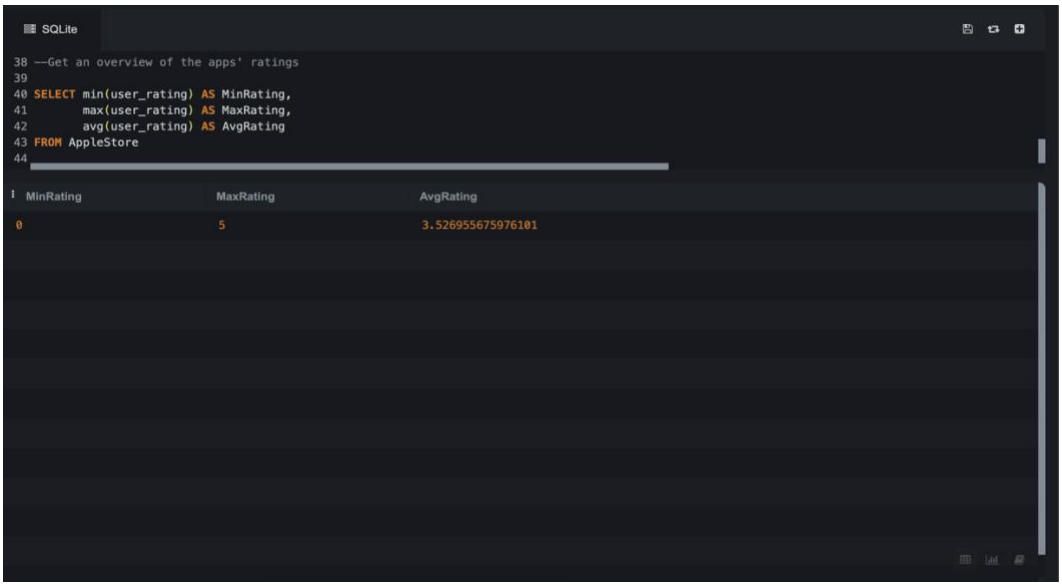
Exploratory Data Analysis

Once I imported the dataset into SQLite Online Studio, I began my EDA process. My goal with this process was to overview the data so I know what I should analyze in the data analysis process.

- I started off by checking the number of unique apps in both tables by using COUNT(DISTINCT).
    - The unique count for each table was 7,197 apps; I don't have to worry about sampling bias since the distribution will be approximately Normal.

- Next, I checked for any NULL values.
    - I made sure to only retrieve NULL values by creating WHERE statements.

- Then, I found the number of apps per genre by using COUNT(*).
    - I created a GROUP BY statement to group all the genres into its distinct values.
- I noticed that the Games and Entertainment genres have the most apps.
    - This means that these genres are extremely competitive and saturated.



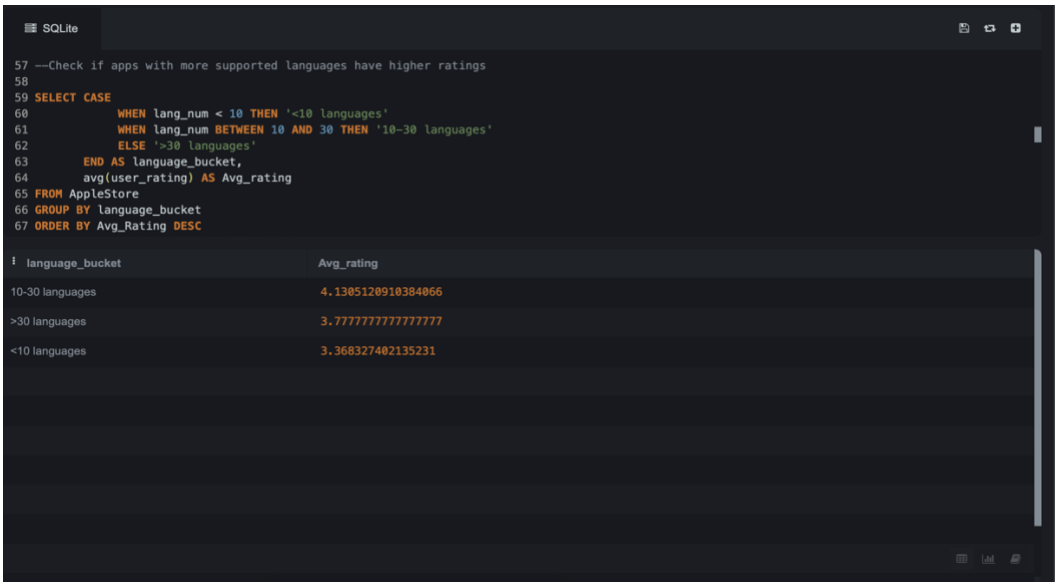
- Finally, I queried for the minimum, maximum, and average app ratings.
    - This tells me that the top 50% of apps had a rating above 3.5.

Data Analysis

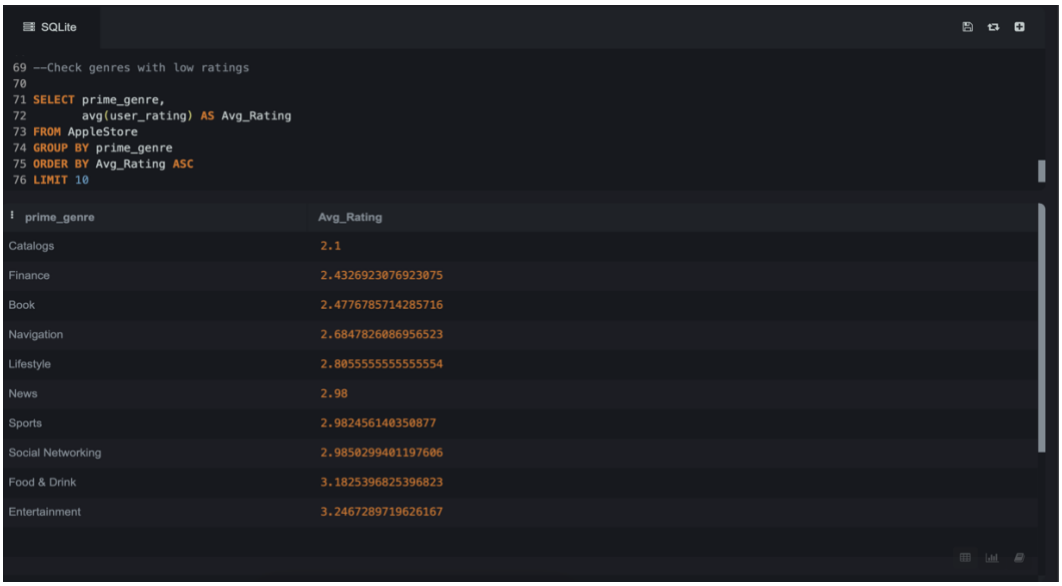
After going through the EDA process, the most important information I obtained was that the top 50% of apps had a rating above 3.5. And since the app developer wants to maximize user ratings, I came up with a few questions:

1. How do paid-app ratings compare with free-app ratings?
2. Which genres have the lowest app ratings?
3. What relationship do additional features like language support and app description have with app ratings?

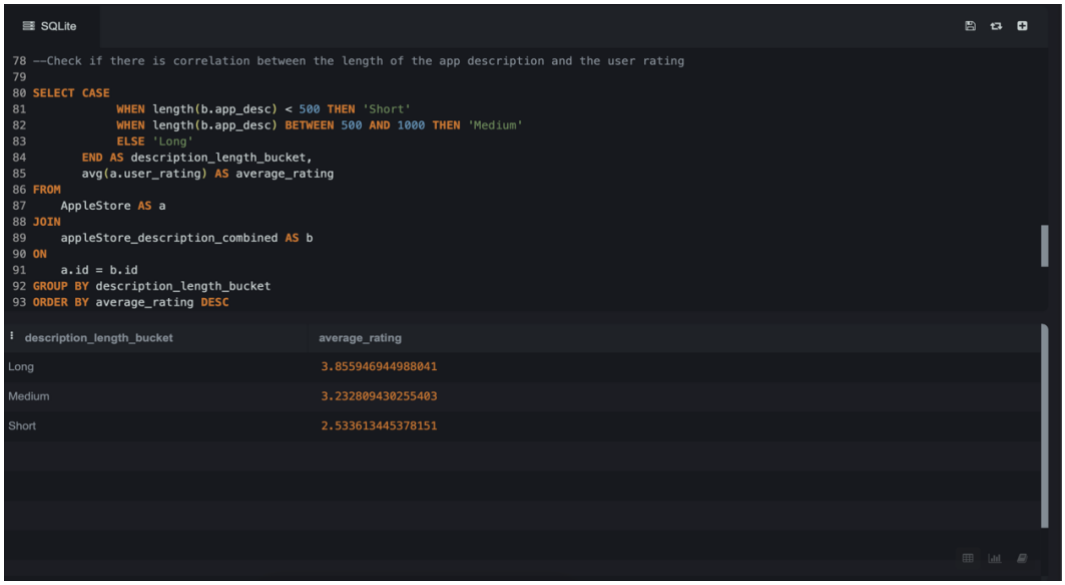
- I compared the average ratings of free and paid apps by creating a CASE statement where I assigned variables (Paid/Free) under a column called App_Type. Then, I group the Paid/Free apps into one field with their average ratings.
- I noticed that the average rating for paid apps is higher than the average rating for free apps.
    - This may be because paying for an app leads to more engagement, thus higher user satisfaction.



- This was a similar query from the “Free App rating vs. Paid App rating” query. The only thing different was that I created language support amount brackets.
- I noticed that apps that support 10-30 languages have the highest ratings.
    - This shows that it’s not about how many languages an app supports. It’s about supporting certain languages to reach a wider audience.



- I retrieved genres with the lowest average ratings by grouping genres into distinct fields and sorting the average rating by ascending order.
- This query showed me that the lowest-rated genres like Catalogs and Finance had customers that weren’t satisfied by the services and features within a genre.
    - Thus, creating Catalogs or Finance apps are a great opportunity to maximize user ratings because there aren’t many apps within those genres (less competition), and it’ll be a chance to fulfill missing user satisfaction.



- Again, I used a CASE statement to assign data into specific fields (Long/Short/Medium) under a column. Then, I used an INNER JOIN between both tables to join data that have matching IDs.

- Here, I noticed that a long app description has the highest average rating compared to medium and short app descriptions.
    - This could be the fact that a long app description gives a clearer understanding of an app's feature and services, leading to better user satisfaction.

Insights

After exploring and analyzing the App Store dataset, here are the insights and recommendations:

1. Paid apps are more likely to have higher ratings than free apps.
    a. People who pay for an app are more likely to have engagement with the app = more satisfied.
2. Apps between 10-30 languages have the highest ratings.
    a. It's not about how many languages that an app can support; it's about supporting languages that can reach a wider audience.
3. Finance and Book apps have the lowest ratings.
    a. This is a great opportunity to create quality Finance/Book apps to fulfill missing customer satisfaction.
4. Apps with longer descriptions have better ratings.
    a. People are more satisfied with an app if they clearly understand its services and features.
5. New apps should aim for a rating above 3.5.
    a. Having a rating above 3.5 will help the app stand out (top 50%).
6. Games and Entertainment genres have the highest competition.
    a. It's extremely tough to stand out due to saturation.
    b. However, these genres have a high demand = high risk, high reward.