

Apple App Store Project

In this scenario, the stakeholder is an app developer who wishes to maximize their efforts towards a marketable successful app. That said, in order to achieve their intent this professional seeks to find out what the most popular app categories are, what price to set and and a way to maximize platform user ratings.

In this scenario, the stakeholder is an app developer who wishes to maximize their efforts towards a marketable, successful app. That said, to achieve their intent, this professional seeks to find out what the most popular app categories are, what price to set, and how to maximize platform user ratings.

First, we need to find some datasets suitable for the task. In [kaggle.com](https://www.kaggle.com), there are many files up to the task; however, two CSV files stand out from the others. The first one is called appleStore_description, approximately 12.5 MB; the second one is named AppleStore, with 819 KB.

Once these are found, we can start an Exploratory Data Analysis (EDA). Throughout this process, we can find any possible issues with the data, including some missing or inconsistent data errors, and outliers, among others, which can save us a lot of time.

Let's look deep into it.

SQL

query_2

Query results

7197 rows

| column00 | id | track_name | size_bytes | currency | price | rating_count_to |
|----------|-----------|--|------------|----------|-------|-----------------|
| 1 | 281656475 | PAC-MAN Premium | 100788224 | USD | 3.99 | 21292 |
| 2 | 281796108 | Evernote - stay organized | 158578688 | USD | 0 | 161065 |
| 3 | 281940292 | WeatherBug - Local Weather, Radar, Maps, Alerts | 100524032 | USD | 0 | 188583 |
| 4 | 282614216 | eBay: Best App to Buy, Sell, Save! Online Shopping | 128512000 | USD | 0 | 262241 |
| 5 | 282935706 | Bible | 92774400 | USD | 0 | 985920 |
| 6 | 283619399 | Shanghai Mahjong | 10485713 | USD | 0.99 | 8253 |
| 7 | 283646709 | PayPal - Send and request money safely | 227795968 | USD | 0 | 119487 |
| 8 | 284035177 | Pandora - Music & Radio | 130242560 | USD | 0 | 1126879 |
| 9 | 284666222 | PCalc - The Best Calculator | 49250304 | USD | 9.99 | 1117 |

< 1 2 144 >

Download

Some good stuff in the dataset above that we can use in our analysis: user_ratings, user_ratings_tot, prime_genre, track_name, price and lang_num.

| > SQL | | | | AppleStoreDescription |
|---------------|--|------------|--|-----------------------|
| Query results | | | | 7197 rows |
| id | track_name | size_bytes | app_desc | |
| 283646709 | PayPal - Send and request money safely | 227795968 | Description TAP INTO YOUR MONEY Send money | |
| 284035177 | Pandora - Music & Radio | 130242560 | Find the music you love and let the music y | |
| 284666222 | PCalc - The Best Calculator | 49250304 | PCalc is the powerful choice for scientists | |
| 284736660 | Ms. PAC-MAN | 70023168 | Now with MFi controller support! Ms. PAC-MAI | |
| 284791396 | Solitaire by MobilityWare | 49618944 | Solitaire by MobilityWare is the ORIGINAL m | |
| 284815117 | SCRABBLE Premium | 227547136 | CONNECT WITH FRIENDS. PLAY WITH WORDS. Get | |
| 284815942 | Google - Search made just for mobile | 179979264 | The Google app keeps you in the know about | |
| 284847138 | Bank of America - Mobile Banking | 160925696 | Bank conveniently and securely with the Ban | |
| 284862767 | FreeCell | 55153664 | Play the BEST FreeCell Solitaire game on yo | |
| < 1 2 144 > | | | | |

Here, we have track_name and app_desc, that are worth it to analyze.

2 - EXPLORATORY DATA ANALYSIS

2.1 - Data Cleaning

Beforehand we might need to check if both datasets have the exact apps list.

| > SQL | | track_name_check |
|---------------|------------|------------------|
| Query results | | 0 rows |
| source | track_name | |
| | | |

Since we did not find any discrepancies between the two tables, we can now proceed to checking unique app numbers, as well as identifying any null or missing values.

| | |
|------------------------------------|------------------------------------|
| > SQL | UniqueAppIDs_AppleStoreDescription |
| ▼ Query results | 1 row |
| UniqueAppIDs_AppleStoreDescription | |
| 7197 | |

| | |
|---------------------------------|---------------------------------|
| > SQL | UniqueAppIDs_AppleStore_Dataset |
| ▼ Query results | 1 row |
| UniqueAppIDs_AppleStore_Dataset | |
| 7197 | |

| | |
|----------------------------------|----------------------------------|
| > SQL | MissingValues_AppleStore_DataSet |
| ▼ Query results | 1 row |
| MissingValues_AppleStore_DataSet | |
| 0 | |

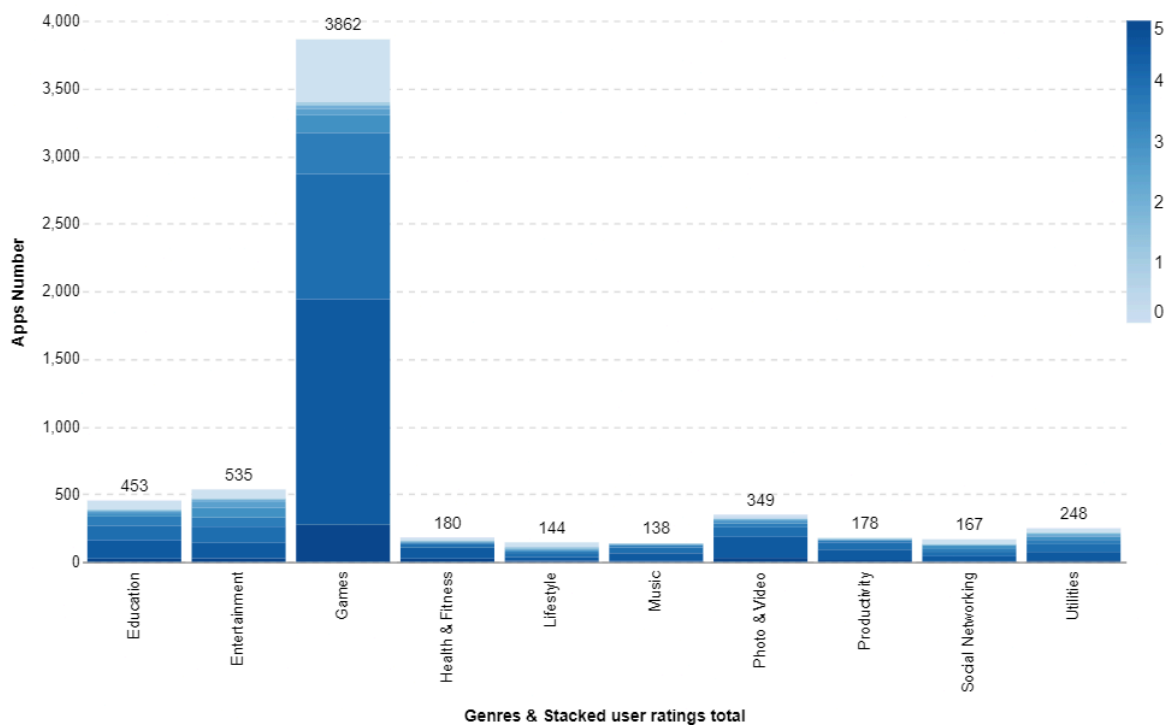
| | |
|-------------------------------------|-------------------------------------|
| > SQL | MissingValues_AppleStoreDescription |
| ▼ Query results | 1 row |
| MissingValues_AppleStoreDescription | |
| 0 | |

After completing the data cleaning process, we can proceed to the next step.

2.2 - Pre-processing and Visualizations

So what are the *most numerous genres*?

| prime_genre | NumApps |
|-------------------|---------|
| Games | 3862 |
| Entertainment | 535 |
| Education | 453 |
| Photo & Video | 349 |
| Utilities | 248 |
| Health & Fitness | 180 |
| Productivity | 178 |
| Social Networking | 167 |
| Lifestyle | 144 |



By hovering over the columns, we can observe the ratings ratio for each gender, providing valuable insights.

An outlier has been identified: the majority of store apps fall under the Games category, with 3,862 apps. This suggests that the Games category might be more competitive and therefore harder to succeed in.

Overall, what is the **average user rating**?

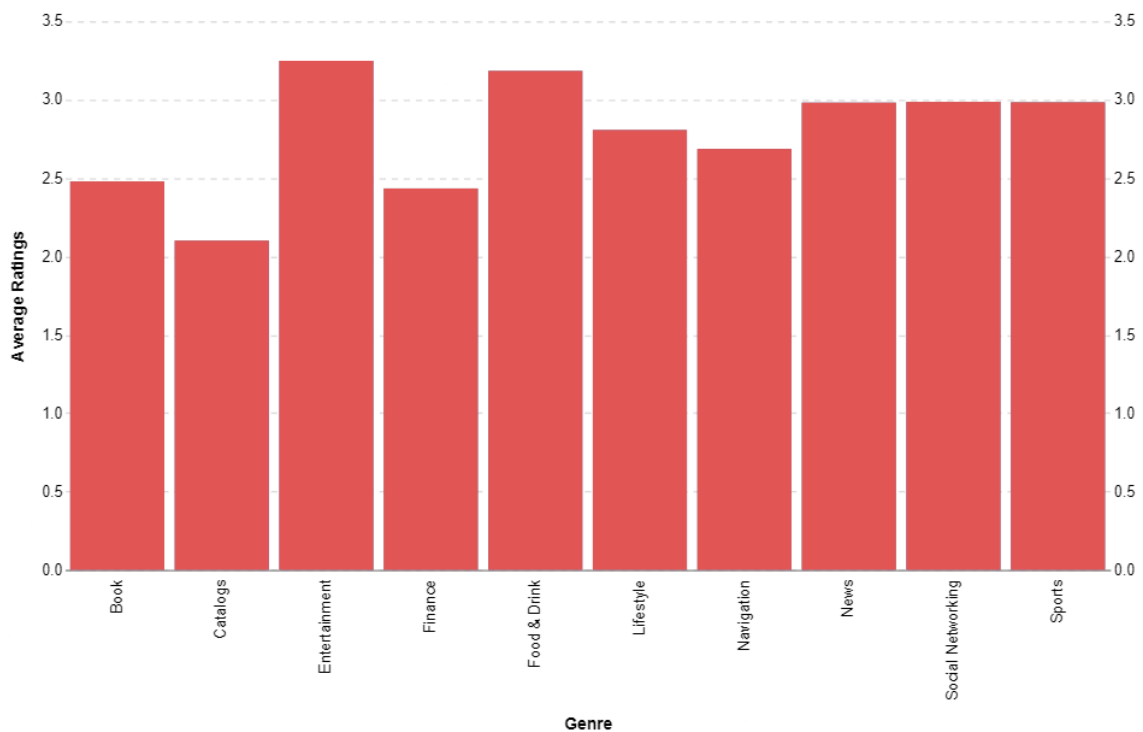
| > SQL | | | user_ratings |
|---------------|-----------|-------------|--------------|
| Query results | | | 1 row |
| MinRating | MaxRating | AvgRating | |
| 0 | 5 | 3.526955676 | |

From this, we can conclude that an app with a 3.5 rating would be around the average for the Apple App Store.

Referring back to the total number of apps and user ratings by genre in the visualization above, we can see that approximately 75% of games are rated 4 out of 5 or higher, highlighting that the Games category is even more competitive than we initially thought.

However, this analysis does not provide much additional insight yet. Examining the lowest-rated genres could potentially reveal opportunities in areas that

| > SQL | | AvgRatings_Genre |
|-------------------|--------------|------------------|
| Query results | | 10 rows |
| prime_genre | Avg_Rating | |
| Catalogs | 2.1 | |
| Finance | 2.4326923077 | |
| Book | 2.4776785714 | |
| Navigation | 2.6847826087 | |
| Lifestyle | 2.8055555556 | |
| News | 2.98 | |
| Sports | 2.9824561404 | |
| Social Networking | 2.9850299401 | |
| Food & Drink | 3.1825396825 | |



This looks promising, as there are **ten categories (genres)** that are rated in **average 3.25** or below.

Next, let's identify which apps have **higher ratings within these lower-rated group**. To **focus on a more relevant subset**, we'll **limit our search to apps that have been rated at least 5,000 times**.

| > SQL | | | query_5 |
|-------------------|--|-------------|---------|
| Query results | | | 10 rows |
| prime_genre | track_name | user_rating | |
| Entertainment | Bruh-Button | 5 | |
| Catalogs | CPlus for Craigslist app - mobile classifieds | 5 | |
| Book | Color Therapy Adult Coloring Book for Adults | 5 | |
| Social Networking | We Heart It - Fashion, wallpapers, quotes, tattoos | 5 | |
| Lifestyle | insy - Makeup, subscription and beauty tips | 5 | |

On a side note, some apps have longer descriptions than others. **Could the length of the description influence the ratings?** Let's define short descriptions as those with fewer than 500 characters, medium descriptions as those between 500 and 1,000 characters, and long descriptions as those exceeding 1,000 characters.

| > SQL | | query_9 |
|---------------------------|----------------|---------|
| Query results | | 3 rows |
| description_length_bucket | average_rating | |
| Long | 3.855946945 | |
| Medium | 3.2328094303 | |
| Short | 2.5336134454 | |

It appears that users tend to prefer apps with detailed descriptions that explain the app's purpose, features, and other key details, over shorter descriptions. **Therefore, apps with descriptions under 500 characters seem less favoured.**

Next, we could **explore whether highly rated apps support a significant number of languages.**

While some apps are only available in a single language - often English - others offer a broad selection, with some supporting over 30 languages. Given this variety, **it would be interesting to examine whether the number of supported languages has an impact on app ratings.**

| > SQL | | languages_access |
|-----------------|--------------|------------------|
| Query results | | 3 rows |
| language_bucket | Avg_Rating | |
| 10-30 LANGUAGES | 4.130512091 | |
| > 30 LANGUAGES | 3.777777778 | |
| <10 LANGUAGES | 3.3683274021 | |

That's an interesting finding: something between 10-30 languages could do very well.

Continuing with our task of gaining insights for launching an app, **let's explore which genres are the least common and lowest-rated.**

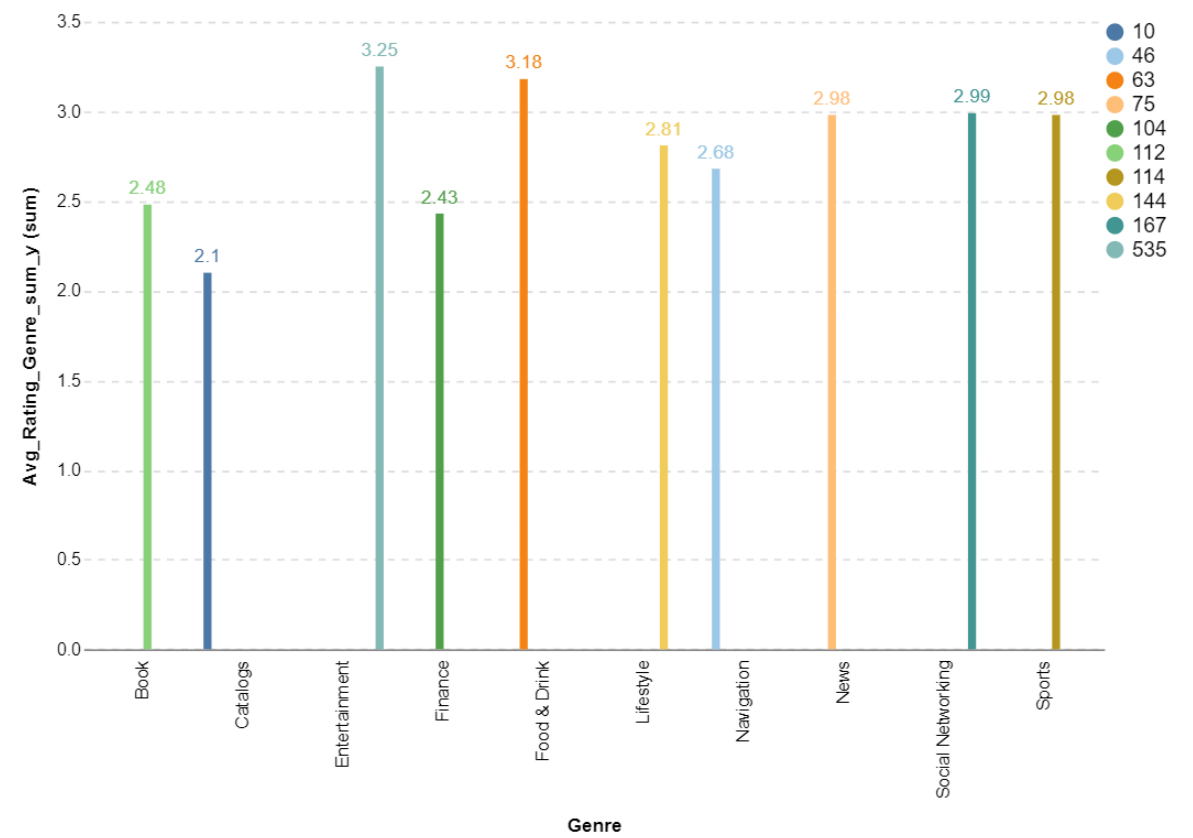
> SQL

Low_Rated_Count_Avg

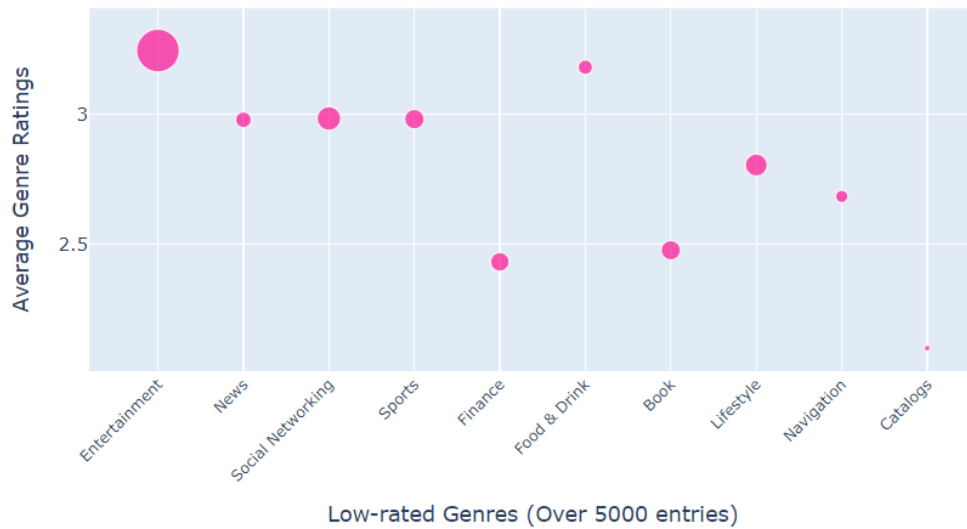
▼ Query results

10 rows

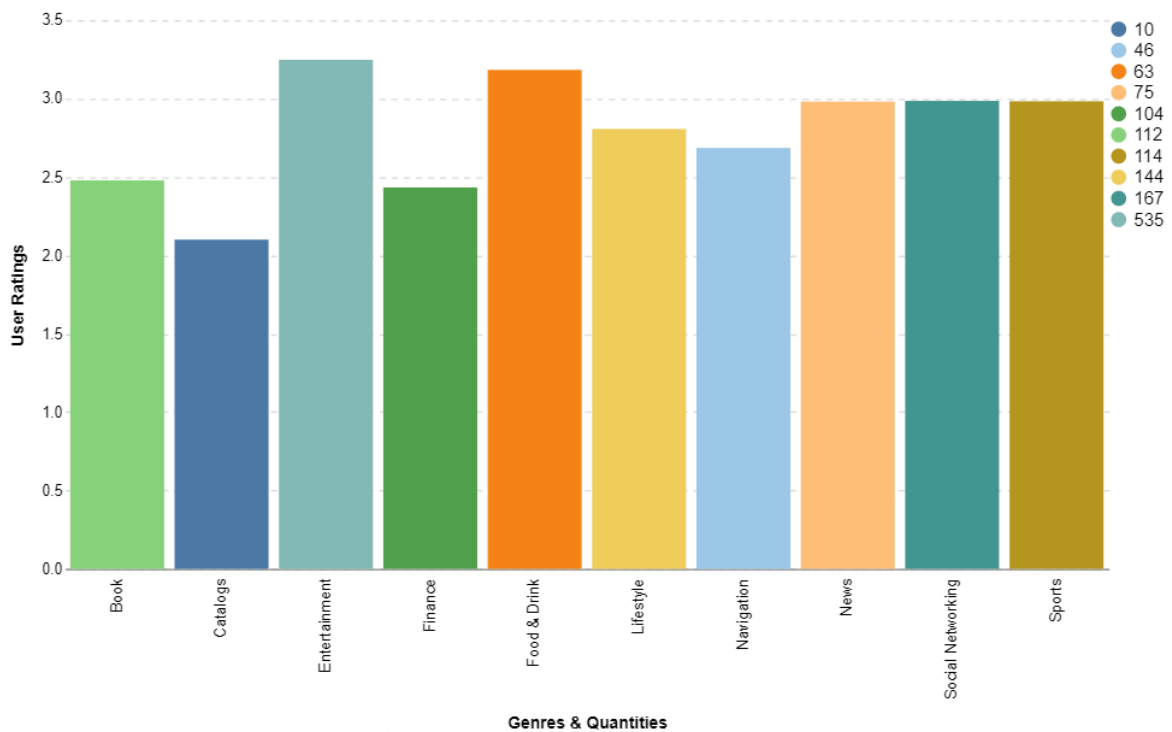
| prime_genre | track_name | lowrated_app_user_rating | price | app_count | Av |
|-------------------|--|--------------------------|-------|-----------|----|
| News | WIRED Magazine | 2.5 | 0 | 75 | 2. |
| Social Networking | Grindr - Gay and same sex guys chat, meet and date | 2.5 | 0 | 167 | 2. |
| Sports | NBC Sports | 2.5 | 0 | 114 | 2. |
| Finance | Wells Fargo Mobile | 3 | 0 | 104 | 2. |
| Food & Drink | Allrecipes Dinner Spinner | 3.5 | 0 | 63 | 3. |
| Book | Kindle - Read eBooks, Magazines & Textbooks | 3.5 | 0 | 112 | 2. |
| Lifestyle | Tinder | 3.5 | 0 | 144 | 2. |
| Navigation | MotionX GPS | 3.5 | 1.99 | 46 | 2. |
| Catalogs | CPlus for Craigslist app - mobile classifieds | 5 | 0 | 10 | 2. |



Average Genre Ratings vs Low-rated Genres



The ***bubble visualization*** above highlights the ***10 lowest-rated genres***, each one with over ***5,000 user ratings***, along with their respective quantities



From the tables above, we can observe the apps with the ***ten lowest ratings, along with their respective prices and the total number of apps in these low-rated categories, all of which have at least 5,000 user ratings.***

This data confirms that almost all of these apps are free, which raises an important question: What about the paid apps? What are their user ratings, and what is their average price?

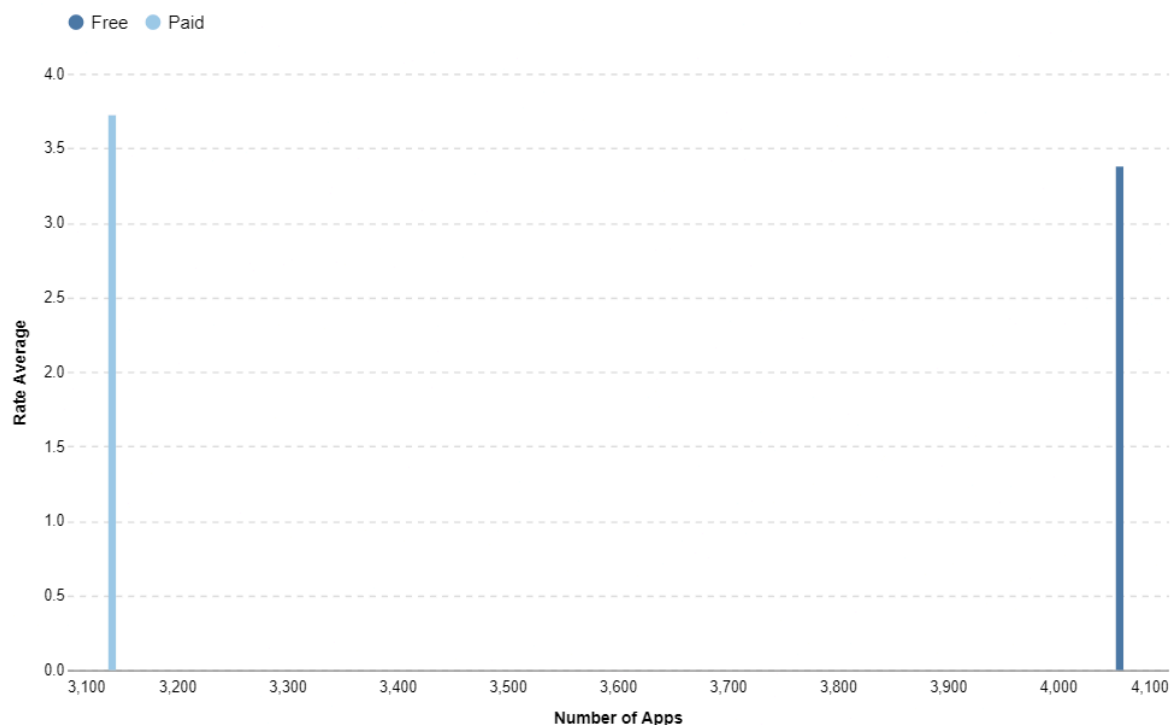
Regarding the tables above, we can see the apps with the ten lowest rates, also their respective prices, and the total of apps that are in these low-rated categories. Also, important to mention, at least 5,000 user ratings.

With this information, we can confirm that almost all of them are free, which brings a question: ***what about the paid ones: what are their user ratings like and what is the average price?***

Let's find out if there is some correlation.

First, what about ***free apps***? Are they ***better rated than paid apps***? Which kind are ***more numerous***?

| Query results | | | 2 rows |
|---------------|--------------|----------------|--------|
| App_Type | Avg_Rating | Number_of_Apps | |
| Paid | 3.7209487424 | 3141 | |
| Free | 3.3767258383 | 4056 | |



We can observe that the ***average ratings of both categories are roughly the same***, which certainly ***affects our analysis***. As for the number of apps in each category, while ***free apps are more numerous***, we do not have the data to determine whether they are freemium-based.

Since we are already analyzing whether apps are free or paid, as well as their distribution, **let's also examine the price range.**

| PriceRange | NumApps |
|-----------------|---------|
| 0 | 4056 |
| 0.01 - 1 | 728 |
| 1.00 - 2.99 | 1304 |
| 3.00 - 4.99 | 671 |
| 5.00 - 6.99 | 218 |
| 7.00 - 8.99 | 42 |
| 9.00 - 10.99 | 81 |
| 11.00 - 12.99 | 11 |
| 13.00 - 14.99 | 27 |
| 15.00 - 16.99 | 6 |
| 17.00 - 18.99 | 4 |
| 19.00 - 20.99 | 15 |
| 21.00 - 22.99 | 3 |
| 23.00 - 24.99 | 10 |
| 27.00 - 28.99 | 2 |
| 29.00 - 30.99 | 6 |
| 33.00 - 34.99 | 1 |
| 39.00 - 40.99 | 2 |
| 47.00 - 48.99 | 1 |
| 49.00 - 50.99 | 2 |
| 59.00 - 60.99 | 3 |
| 73.00 - 74.99 | 1 |
| 99.00 - 100.99 | 1 |
| 249.00 - 250.99 | 1 |
| 299.00 - 300.99 | 1 |

Regarding checking the price, it's interesting to note that **in the market some apps are free to download and sign up**, but it's **likely that they require payments for extra features**. Tinder, for instance, is one of the apps that uses this **market strategy**, which is called **Freemium**.

The table above also reveals a **trend** where most **paid apps are priced between \$1 and \$3**.

At this stage, **we can say that there is plenty of room for bringing to market a new app** in **News**, **Entertainment**, and **Navigation**. We'll bring the details over the recommendations below.

3 - RECOMMENDATIONS

In conclusion, based on the analysis, we recommend developing *an app in underrepresented, low-rated categories like Catalogs, Finance, and Navigation*. These categories *offer potential* due to their *smaller number of apps*, which presents *less competition*. *A well-designed app with a superior user experience could address existing gaps and capitalize on the lack of strong contenders in these spaces.*

For the Books category, while there is uncertainty around whether low ratings apply to the content or e-reader functionality, further investigation could help clarify if this area presents a viable opportunity.

Additionally, we *suggest* launching a *paid* app priced *between USD 1.00 and USD 4.00* or adopting a *Freemium* model, *offering basic features for free while charging for advanced options*. This model allows you to attract a larger user base initially while monetizing