**Apple store review analysis**

**Project context;** Data set that covers information on apps available in apple app store

**Stakeholder:** Web developer, an aspiring app developer who needs data driven insights to decide which app to develop; Questions, what app category are most popular, what pricing should I set for the app, how can I maximize user ratings

**Business objective:** Find the best app to develop

**Source of data:** Kaggle (Apple store for all the apps gives information on --- and each application description)

**Analysis tool: SQLITE**

**Process of analysis:** Downloaded the data in excel format, Divided the app description files into smaller chunks because SQL has a limited amount of 4MBs it can load

**SQL ANALYSIS;**

1. **Joined the description data into one file**

**CREATE TABLE Combined\_appleStore\_descriptions AS**

**SELECT \* FROM appleStore\_description1**

**UNION ALL**

**SELECT \* FROM appleStore\_description2**

**UNION ALL**

**SELECT \* FROM appleStore\_description3**

**UNION ALL**

**SELECT \* FROM appleStore\_description4**

1. **Checked if the data matches all the entries after joining**

**--Check the number of unique appsAppleStore to see if there is any descrepancies**

**SELECT COUNT(DISTINCT id) As uniq\_apps**

**FROM AppleStore**

**"7197"**

**SELECT COUNT(DISTINCT id) As uniq\_apps**

**FROM Combined\_appleStore\_descriptions**

1. **--Check for any missing values in key fields**

**SELECT COUNT(\*) As missing\_values**

**FROM AppleStore**

**WHERE track\_name is NULL or user\_rating is null or prime\_genre is null**

**"0"**

**SELECT COUNT(\*) As missing\_values**

**FROM Combined\_appleStore\_descriptions**

**WHERE app\_desc is NULL**

**"0"**

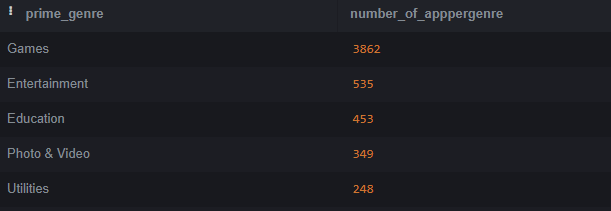
**--Find out the number of apps per genre**

**SELECT prime\_genre, COUNT(\*) As number\_of\_apppergenre**

**FROM AppleStore**

**GROUP BY prime\_genre**

**ORDER BY number\_of\_apppergenre DESC**

****

**--Get an overview of apps rating**

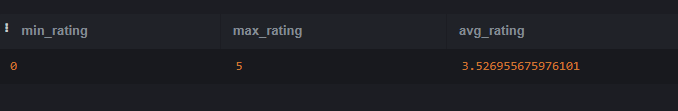
**SELECT**

**MIN(user\_rating) As min\_rating,**

**MAX(user\_rating) As max\_rating,**

**AVG(user\_rating) As avg\_rating**

**from AppleStore**

****

**\*\*DATA ANALYSIS\*\***

**--Determine whether paid apps have higher ratings from app**

**SELECT CASE**

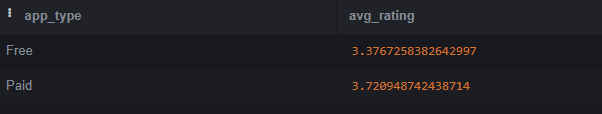
**when price >0 THEN 'Paid'**

**ELSE 'Free'**

**END As app\_type, avg(user\_rating) As avg\_rating**

**FROM AppleStore**

**GROUP BY app\_type**

****

**--check if apps with more languages have higher rating**

**SELECT CASE**

**WHEN lang\_num < 10 THEN '< 10 languages'**

**WHEN lang\_num BETWEEN 10 AND 30 THEN '10-30 languages'**

**ELSE '> 30 languages'**

**end As language\_bucket, avg(user\_rating)**

**FROM AppleStore**

**GROUP by language\_bucket**

****

**--Check which genre has the highest ratingAppleStore**

**SELECT prime\_genre, avg(user\_rating) as avg\_rating**

**FROM AppleStore**

**GROUP BY prime\_genre**

**ORDER BY avg\_rating Asc**

****

**--Check if there is correlation between length of app description and user ratingAppleStore**

**SELECT CASE**

**WHEN length(b.app\_desc)< 500 THEN 'short'**

**WHEN length(b.app\_desc) BETWEEN 500 AND 1000 THEN 'Medium'**

**ELSE 'Long'**

**end As description\_bucket, avg(a.user\_rating)As avg\_rating**

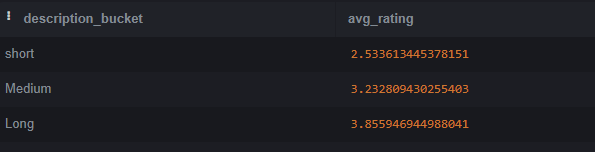
**FROM AppleStore as a**

**JOIN Combined\_appleStore\_descriptions as b**

**on a.id = b.id**

**GROUP BY description\_bucket**

**ORDER BY avg\_rating**

****

**--Check the top rated app per genreAppleStore**

**SELECT**

**prime\_genre,**

**user\_rating,**

**track\_name**

**FROM (**

**SELECT**

**prime\_genre,**

**user\_rating,**

**track\_name,**

**rank() over(partition by prime\_genre ORDER BY user\_rating DESC, rating\_count\_tot DESC) as rank**

**FROM AppleStore) As a**

**WHERE a.rank=1**

****

**RECOMMENDATIONS**

Paid apps have higher ratings than free apps which maybe because users of paid apps have higher engagement rates due to high perceived values and so we can tell our client if the quality of app is high they should consider charging.

Apps supporting multiple languages 10 to 30 have higher rating it’s not number of languages your app support but the focus should be on the right language

Finance and book apps have low rating; suggesting user needs are not being fully met in these genres. So developing a relevant and meaningful app may bring more users

The length of app description has a positive correlation with user rating which implies user appreciate a clear understanding of app features and capabilities before deciding to use the app. So, a well-crafted app description can set clear expectation and eventually increase satisfaction of users

Target rating’; On average all the apps have an average rating of 3.5 so in order to stand out from the crowd we should aim for user rating higher than 3.5

The games and entertainment genres have high volumes of apps suggesting the market maybe saturated. So, entering this space maybe challenging due to high competition, however it also suggest the high user demand in this domain