

```
** App Store Analysis **
```

```
CREATE TABLE appleStore_description_combined AS
```

```
SELECT * FROM appleStore_description1
```

```
UNION ALL
```

```
SELECT * FROM appleStore_description2
```

```
UNION ALL
```

```
SELECT * FROM appleStore_description3
```

```
UNION ALL
```

```
SELECT * FROM appleStore_description4
```

```
** EXPLORATORY DATA ANALYSIS **
```

```
-- Checking the number of unique apps in both tables
```

```
SELECT COUNT( DISTINCT id) AS UniqueAPPIDs  
FROM AppleStore
```

```
SELECT COUNT(DISTINCT id) as UniqueAPPIDs  
FROM appleStore_description_combined
```

```
-- Results (Both the UniqueAppIDs are same so there is no missing data.)
```

```
-- Checking for any missing values in key fields
```

```
SELECT COUNT(*) AS MissingValues  
FROM AppleStore  
WHERE track_name IS NULL OR user_rating IS NULL OR prime_genre IS NULL
```

```
SELECT COUNT(*) AS MissingValues  
FROM appleStore_description_combined  
WHERE app_desc IS NULL
```

```
-- Results ( There are no missing values in both the tables )
```

```
-- Finding out the number of apps per genre
```

```
SELECT prime_genre, COUNT(*) AS NumApps  
FROM AppleStore  
GROUP BY prime_genre  
ORDER BY NumApps DESC
```

```
-- Results ( From the analysis we got to know that Games and Entertainment
has the highest number of apps per genre which is 3862 and 535
respectively.)
```

```
-- Getting an overview of the apps ratingAppleStore
```

```
SELECT min(user_rating) AS minRating,
        max(user_rating) AS maxRating,
        avg(user_rating) AS avgRating
FROM AppleStore
```

```
-- Results ( From the analysis we get to know that minimum rating is of 0,
maximum rating is of 5 and average rating is of 3.52 in AppleStore.)
```

```
** DATA ANALYSIS **
```

```
-- Determining whether paid apps have higher ratings than the free apps
```

```
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
```

```
-- Results ( From the analysis we got know that Paid Apps had an better
avg rating i.e 3.72 than free apps which is around 3.37.)
```

```
-- Check if apps with more supported languages have higher ratings
```

```
SELECT CASE
            WHEN lang_num < 10 THEN '<10 langauages'
            WHEN lang_num BETWEEN 10 AND 30 THEN '10-30 langauges'
            ELSE '>30 langauges'
        END AS language_bucket,
        avg(user_rating) AS Avg_Rating
FROM AppleStore
GROUP BY language_bucket
ORDER BY Avg_Rating DESC
```

```
-- Results ( From here we can see that middel bucket has higher average
user rating. So it is not necessary to work on many languages and we can
focus on other aspects of the app. )
```

```
-- Checking genres with low ratings
```

```
SELECT prime_genre, avg(user_rating) AS AVG_rating
FROM AppleStore
```

```
GROUP BY prime_genre
ORDER BY Avg_Rating ASC
LIMIT 10
```

```
-- Results ( From the analysis we get to know that Catalogs (2.1), Finance
(2.43), Book(2.47) are the genres with the lowest average rating given by
the users. This might be the opportunity to create an app in this genres )
```

```
-- Checking if there is correlation between the length of the app
description and the 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_length_bucket,
        avg(a.user_rating) AS average_rating
```

```
from AppleStore as A
JOIN
appleStore_description_combined AS B
ON A.id = B.id
```

```
GROUP BY description_length_bucket
ORDER BY average_rating DESC
```

```
-- Results ( From the analysis we get to learn that, longer the app
description the better is the average rating of the app.)
```

```
-- Checking the top-rated apps for each genreAppleStore
```

```
SELECT prime_genre, track_name, user_rating
FROM ( SELECT prime_genre, track_name, user_rating, 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
```

```
-- Results ( From the analysis we get a result of all the apps with the
highest total user rating count for each genre.)
```

```
** Final Recommendations **
```

```
-- 1) Paid Apps have better ratings [ This could be because of number of
reasons, one being that users who pay for an app may have higher
engagement and perceived more value which might have lead to better
ratings for the paid apps.]
```

-- 2) Apps supporting between 10 AND 30 Languages have better ratings. [So it is not necessary to work on many languages and we can focus on other aspects of the app.]

-- 3) Finance and Book Apps have low ratings. [From this we can see that user needs are not met in finance and books apps, so this can be an market opportunity to build an app in one of these categories.]

-- 4) Apps with a longer description have better ratings. [Users likely appreciate having an clear understanding of the app's features and capabilities before they download, leading to better ratings.]

-- 5) A new App should aim for an average rating above 3.5. [In app store average rating of all the apps is around 3.5, so for the new app to stand out the rating of the new app should be more than 3.5.]

-- 6) Games and Entertainment genres have high competition. [These categories have a very high volume of apps, so entering these spaces might be a little challenging because of high competition, but also suggests an high user

--

demand in these sectors.]