SQL Data Analysis Project Documentation

Project Overview

Project Title: Apple App Store Data Analysis

Project Description: This project involves analyzing data from the Apple App Store to gain insights into app ratings, genres, pricing, and user preferences. The goal is to provide valuable recommendations based on the analysis to inform business decisions and potential opportunities in the app market.

Data Sources

CSV files: appleStore_description1, appleStore_description2, appleStore_description3 , appleStore_description4 & AppleStore
These tables contain information about app descriptions. Table 'AppleStore' contains information about apps, including user ratings, genres, pricing, and more.

Data Integration

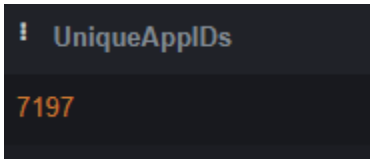
```
CREATE TABLE appleStore_description_combined AS
SELECT * FROM appleStore_description1
UNION ALL
SELECT * FROM appleStore_description2
UNION ALL
SELECT * FROM appleStore_description2
UNION ALL
SELECT * FROM appleStore_description3
UNION ALL
SELECT * FROM appleStore_description4;
```

To perform a comprehensive analysis, data from multiple sources was integrated. A new table, `appleStore_description_combined`, was created by combining data from `appleStore_description1`, `appleStore_description2`, `appleStore_description3`, and `appleStore_description4`. This consolidated dataset was used for further analysis.

Exploratory Data Analysis (EDA)

Unique App IDs

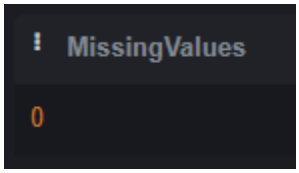
```
SELECT COUNT(DISTINCT id) AS UniqueAppIDs
FROM AppleStore;

SELECT COUNT(DISTINCT id) AS UniqueAppIDs
FROM appleStore_description_combined
```



*The total number of unique app IDs in both `AppleStore` and `appleStore_description_combined` was found to be **7,197**.*

Missing Values

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



There were no missing values in critical fields such as `track_name`, `user_rating`, or `prime_genre` in the `AppleStore` table.

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

MissingValues
0

Similarly, no missing values were found in the `app_desc` field of `appleStore_description_combined`.

Genre Analysis

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

prime_genre	NumberOfApps
Games	3862
Entertainment	535
Education	453
Photo & Video	349
Utilities	248
Health & Fitness	180

Most popular genres were Games, Entertainment and Education in 1st, 2nd & 3rd respectively.

Ratings Overview

```
SELECT MIN(user_rating) AS MinimumRating,
      MAX(user_rating) AS MaximumRating,
      AVG(user_rating) AS AverageRating
FROM AppleStore;
```

MinimumRating	MaximumRating	AverageRating
0	5	3.526955675976101

An overview of app ratings, including minimum, maximum, and average ratings, was found to be 0,5 & 3.5 respectively.

Data Analysis

Paid vs. Free Apps

```
SELECT CASE
        WHEN price > 0 THEN 'Paid'
        ELSE 'Free'
    END AS App_Type,
    AVG(user_rating) AS Average_Rating
FROM AppleStore
GROUP BY App_Type;
```

App_Type	Average_Rating
Free	3.3767258382642997
Paid	3.720948742438714

The analysis compared the average user ratings of paid and free apps. Paid apps were found to have slightly higher average ratings.

Language Support and Ratings

```
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) AS AverageRating
FROM AppleStore
GROUP BY Language_Bucket
ORDER BY AverageRating DESC;
```

language_bucket	AverageRating
10-30 languages	4.1305120910384066
>30 languages	3.7777777777777777
<10 languages	3.368327402135231

Apps were categorized based on the number of supported languages, and their average ratings were compared. Apps with language support between 10 and 30 languages had higher average ratings than others.

Genres with Lower Ratings

```
SELECT prime_genre,
    AVG(user_rating) AS AverageRating
FROM AppleStore
GROUP BY prime_genre
ORDER BY AverageRating
LIMIT 3;
```

prime_genre	AverageRating
Catalogs	2.1
Finance	2.4326923076923075
Book	2.4776785714285716

Top 3 Genres with lower average ratings (2.5 or lower) were Catalogs, Finance and Book, indicating potential opportunities for improvement or new app development.

Description Length vs. Ratings

```
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(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;
```

Description_Length_bucket	Average_rating
Long	3.88292469352014
Medium	3.2596747055524395
Short	2.6159600997506236

The length of app descriptions was categorized, and its correlation with user ratings was analyzed. Longer descriptions were associated with higher average ratings.

Top-Rated Apps by Genre

```
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;
```

prime_genre	track_name	user_rating
Book	Color Therapy Adult Coloring Book for Adults	5
Business	TurboScan™ Pro - document & receipt scanner: scan multiple p…	5
Catalogs	CPlus for Craigslist app - mobile classifieds	5
Education	Elevate - Brain Training and Games	5
Entertainment	Bruh-Button	5
Finance	Credit Karma: Free Credit Scores, Reports & Alerts	5
Food & Drink	Domino's Pizza USA	5
Games	Head Soccer	5
Health & Fitness	Yoga Studio	5

Recommendations

Based on the analysis, the following recommendations can be made:

1. Prioritize Paid Apps: Invest in the development of paid apps, as they tend to have better user ratings.

2. Focus on Language Support: Consider providing support for 10 to 30 languages, as this correlates with higher average ratings.

3. Explore Opportunities in Low-Rated Genres: Genres such as Catalogs, Finance, and Books have lower ratings, suggesting opportunities for app development or improvement in these categories.

4. Optimize App Descriptions: Longer app descriptions are associated with higher average ratings, so ensure that app descriptions are comprehensive and engaging.

Conclusion

This data analysis project provides valuable insights into the Apple App Store ecosystem. By understanding user preferences, genre popularity, and other factors influencing app ratings, businesses and developers can make informed decisions to enhance their apps' success in the market.