SQL Data Analysis Project:

Problem Statement: Analyze Apps in Apple store, figure what type of Apps are popular. Based on the analysis recommend to the client which type of App can be successful.

Questions to be asked: What APP Categories are most popular?
What price should I set?
How can I maximize user ratings?

Pre-requisites: This project is performed on online DB based SQLite, SQLiteonline.com. Due to the website limitation for size of dataset to 4mb, the large dataset is divided into 4 smaller datasets.

Applestore.csv → appleStore_description1
appleStore_description2
appleStore_description3
appleStore_description4

Exploratory Data Analysis (EDA) - Missing/ inconsistent data, errors/ outliers

Break down into small files and use union all to combine them into single dataset.

```
1
2 -- Create a single table by combining all the small datasets
3 CREATE TABLE appleStore_description_combined AS
4
5 SELECT * FROM appleStore_description1
6 UNION ALL
7
8 SELECT * FROM appleStore_description2
9 UNION ALL
10
11 SELECT * FROM appleStore_description3
12 UNION ALL
13
14 SELECT * FROM appleStore_description4
15
```

Check the number of unique apps in both tables AppleStore and appleStore_description_combined.

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

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

UniqueAppIDs
7197

SQLite

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

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

UniqueAppIDs
7197

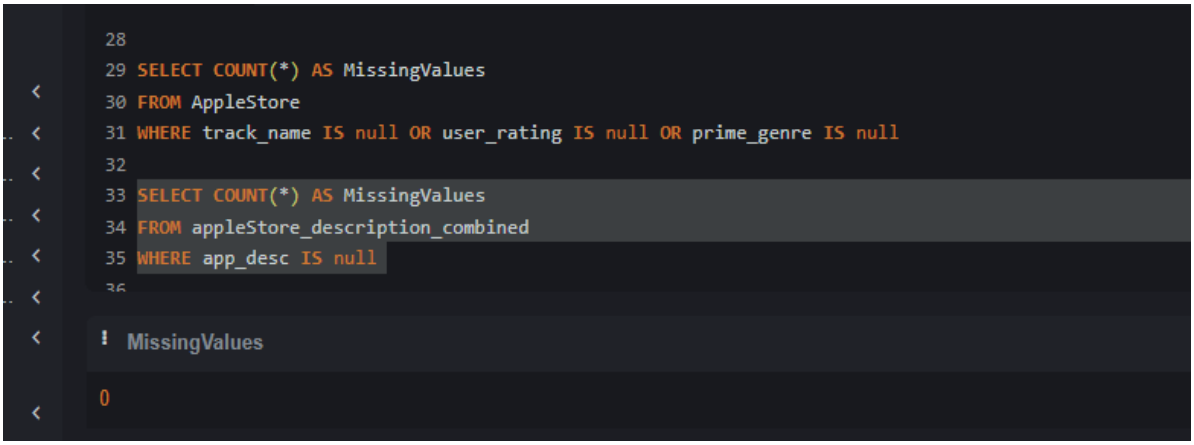
From two tables unique App Ids are 7197 and no missing data between two tables.

Check for any missing values in key fields AppleStore and appleStore_description_combined.

```sql
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
```

MissingValues
0



No missing values in both tables.

Find out the number of apps per genre.

SQLite

```
40
41 SELECT prime_genre, COUNT(*) AS NumApps
42 FROM AppleStore
43 GROUP BY prime_genre
44 ORDER BY NumApps DESC
45
```

prime_genre	NumApps
Games	3862
Entertainment	535
Education	453
Photo & Video	349
Utilities	248
Health & Fitness	180
Productivity	178
Social Networking	167
Lifestyle	144
Music	138
Shopping	122
Sports	114
Book	112
Finance	104
Travel	81
News	75
Weather	72

Game apps are the most popular category followed by Entertainment and education.

Get an overview of the app's ratings.

```
47
48 SELECT min(user_rating) AS MinRating,
49        max(user_rating) AS MaxRating,
50        avg(user_rating) AS AvgRating
51 FROM AppleStore
52
```

MinRating	MaxRating	AvgRating
0	5	3.526955675976101

Minimum rating given is '0' and Maximum rating given is '5' and overall Average rating is '3.5'.

Determine whether paid apps have higher ratings than free apps.

```
54
55 SELECT CASE
56             WHEN price > 0 THEN 'paid'
57             ELSE 'Free'
58         END AS App_Type,
59         avg(user_rating) AS Avg_Rating
60 FROM AppleStore
61 GROUP BY App_Type
62
```

App_Type	Avg_Rating
Free	3.3767258382642997
paid	3.720948742438714

Paid apps have a bit higher rating compared to free apps.

Check if apps with more supported language have higher ratings.

```
65 SELECT CASE
66             WHEN lang_num < 10 THEN '<10 languages'
67             WHEN lang_num BETWEEN 10 AND 30 THEN '10-30 languages'
68             ELSE '>30 languages'
69         END AS language_bucket,
70         avg(user_rating) AS Avg_Rating
71 FROM AppleStore
72 GROUP BY language_bucket
73 ORDER BY Avg_Rating DESC
74
```

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

Apps with languages in between 10-30 have higher ratings.

Check genres with low ratings.

```
76
77 SELECT prime_genre,
78        avg(user_rating) AS Avg_Rating
79 FROM AppleStore
80 GROUP BY prime_genre
81 ORDER BY Avg_Rating ASC
82 LIMIT 10
83
```

prime_genre	Avg_Rating
Catalogs	2.1
Finance	2.4326923076923075
Book	2.4776785714285716
Navigation	2.6847826086956523
Lifestyle	2.8055555555555554
News	2.98
Sports	2.982456140350877
Social Networking	2.9850299401197606
Food & Drink	3.1825396825396823
Entertainment	3.2467289719626167

Apps belonging to Catalogs category have lower ratings.

Check if there is a correlation between the length of the app description and the user rating.

```sql
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
```

description_length_bucket	average_rating
Long	3.855946944988041
Medium	3.232809430255403
short	2.533613445378151

Apps with better descriptions have more ratings.

Finding the Insights - rank over window function that assigns rank to each row assigns within window of rows and then partition by prime_genre creating a separate window for each unique genre and finally order by user rating in descending order**.**

check the top-rated apps for each group.

```
106 SELECT
107       prime_genre,
108       track_name,
109       user_rating
110 FROM (
111       SELECT
112       prime_genre,
113       track_name,
114       user_rating,
115       RANK() OVER(PARTITION BY prime_genre ORDER BY user_rating DESC, rating_count_tot DESC) AS rank
116       FROM
117       appleStore
118      ) AS a
119 WHERE
120 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 pag...	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
Lifestyle	ipsy - Makeup, subscription and beauty tips	5
Medical	Blink Health	5
Music	Tenuto	5

Final Recommendations for the Client:

1. Paid Apps have better ratings.
2. Apps supporting between 10 and 30 languages have better ratings.
3. Finance and Book Apps have low ratings.
4. Apps with a longer description have better ratings.
5. A new App should aim for an average rating above 3.5.
6. Games and entertainment have high competition.