Conclusions of the Analysis

Brandelyn Nie

What are the most popular categories?

After exploring the data with SQL, the most popular category was Games with 3,327 more apps than the next category Entertainment. For a stakeholder, creating a new game to put in the App Store could be good as there is demand for games, but on the other hand it would be hard to create a game that stands out in such a saturated genre.

```
42 -- See number of apps in each genre.
43
44 SELECT prime_genre, COUNT(*) AS NumApps
45 FROM AppleStore
46 GROUP BY prime_genre
47 ORDER BY NumApps DESC -- the genre with most number of apps displayed at the topAppleStore
48
```

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

What price would be good?

Paid apps on average have a higher user rating than free ones. This could be an indication that app users who paid for apps would have higher engagement with them and see a higher quality in such apps since they had to pay for them.

```
66 /*** DATA ANALYSIS ***/
67
68 -- Let's see if paid apps have higher ratings than free ones.
69
70 SELECT CASE
71         WHEN price > 0 THEN 'Paid'
72         ELSE 'Free'
73     END AS App_Type,
74     avg(user_rating) AS Avg_Rating
75 FROM AppleStore
76 GROUP BY App_Type   -- we see that paid apps have a higher average rating than paid ones, but only by a .4 difference
77
```

App_Type	Avg_Rating
Free	3.3767258382642997
Paid	3.720948742438714

By binning the prices of the apps, the majority of apps are under $10 dollars, and at a glance the majority of paid apps are in the $0.99 to $4.99 range. So if the app developer does choose to put a price on their app these market trends are something to take into consideration.

```sql
56 -- A overview of the distribution of the prices.
57
58 SELECT
59     (price/2) * 2 AS PriceBinStart,
60     ((price/2) * 2) + 2 AS PriceBinEnd,
61     COUNT(*) AS NumApps
62 FROM AppleStore
63 GROUP BY PriceBinStart
64 ORDER BY PriceBinStart
65
```

PriceBinStart	PriceBinEnd	NumApps
0	2	4056
0.99	2.99	728
1.99	3.99	621
2.99	4.99	683
3.99	5.99	277
4.99	6.99	394
5.99	7.99	52
6.99	8.99	166
7.99	9.99	33
11.99	13.99	6
12.99	14.99	5
13.99	15.99	6
14.99	16.990000000000002	21
15.99	17.990000000000002	4
16.99	18.99	2
17.99	19.99	3
18.99	20.99	1
19.99	21.99	13
20.99	22.99	2
21.99	23.99	1
22.99	24.99	2
23.99	25.99	2
24.99	26.99	8
27.99	29.99	2
29.99	31.99	6
34.99	36.99	1
39.99	41.99	2
47.99	49.99	1
49.99	51.99	2
59.99	61.99	3

On the other hand, there is a high volume of free apps, but for the stakeholder there are again concerns about user retention if there are so many other apps offered for free.

```sql
66 -- Apps that are free
67 SELECT price = 0,
68     COUNT(*) AS NumApps
69 FROM AppleStore
70
```

price = 0	NumApps
0	7197

How to maximize user ratings?

The average user rating for all apps is about 3.5 stars, so the app developer will want to consider features that can ensure their app will rate higher than 3.5 stars. Some possible features will be analyzed later in these conclusions.

```sql
49 -- An overview of the ratings.
50
51 SELECT min(user_rating) AS MinRating,
52        max(user_rating) AS MaxRating,
53        avg(user_rating) AS AvgRating
54 FROM AppleStore
55
```

MinRating	MaxRating	AvgRating
0	5	3.526955675976101

An aspect that can be considered is how many languages to support. This is demographic information that could be explored more to pinpoint specific languages to implement, but this dataset only contains the number of languages, and not the specific languages supported.

From the query, having 10 to 30 supported languages has the highest user ratings on average, so the stakeholder quality over quantity in this aspect is more beneficial to their app getting high ratings.

```sql
83 -- Let's see if apps that support more languages have higher ratings.AppleStore
84
85 SELECT CASE
86        WHEN lang_num <10 THEN '<10 Languages'
87        WHEN lang_num BETWEEN 10 AND 30 THEN '10-30 Languages'
88        ELSE '>30 Languages'
89     END AS language_buckets,
90     avg(user_rating) AS Avg_Rating
91 FROM AppleStore
92 GROUP BY language_buckets
93 ORDER BY Avg_Rating DESC -- the 10-30 range has the highest ratings, so more languages isn't a big focus to improve rating
94
```

language_buckets	Avg_Rating
10-30 Languages	4.1305120910384066
>30 Languages	3.777777777777777
<10 Languages	3.368327402135231

Finding the genres with the lowest average ratings could give the stakeholder a great starting point for what type of app to create. The top ten lowest average rated genres are listed in the result below. Creating an app that has the qualities current offerings in those categories do not address would have a high chance of getting higher user ratings than other apps in those categories.

```sql
95  -- Let's check the genres with low ratings.
96
97  SELECT prime_genre,
98         avg(user_rating) AS Avg_Rating
99  FROM AppleStore
100 GROUP BY prime_genre
101 ORDER BY Avg_Rating ASC -- to see the lowest
102 LIMIT 10 -- only ten results
```

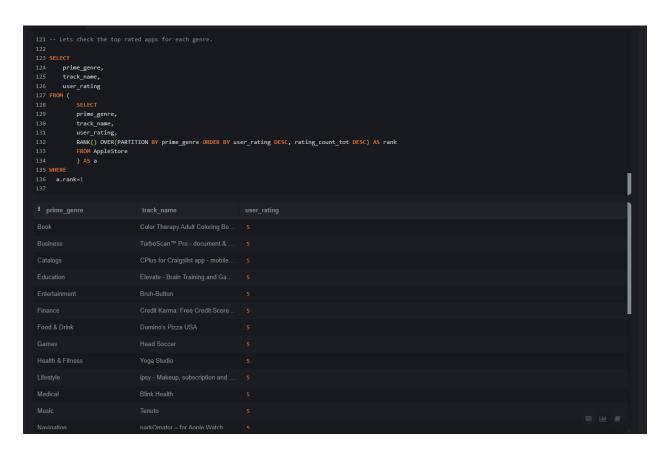| prime_genre	Avg_Rating
Catalogs	2.1
Finance	2.4326923076923075
Book	2.4776785714285716
Navigation	2.6847826086956523
Lifestyle	2.8055555555555554
News	2.98
Sports	2.982456140350877
Social Networking	2.9850299401197606
Food & Drink	3.1825396825396823
Entertainment	3.2467289719626167

Apps with longer descriptions have the highest average user ratings. This positive correlation indicates that it would be a good idea to have a longer description so that those looking to download can fully understand the capabilities of the app, and know what they are downloading. This clarity would make app users happy and leave a positive review for the app.

```sql
104
105 -- Let's check if there is correlation between the length of the app description and user rating.
106     -- we will need to join our two tables for this query
107
108 SELECT CASE
109         WHEN length(B.app_desc) < 500 THEN 'Short'
110         WHEN length(B.app_desc) BETWEEN 500 AND 1000 THEN 'Medium'
111         ELSE 'Long'
112     END AS descrip_length_buckets,
113     avg(user_rating) AS Avg_Rating
114 FROM AppleStore AS A
115 JOIN appleStore_description_combined AS B
116     ON A.id = B.id
117 GROUP BY descrip_length_buckets
118 ORDER BY user_rating ASC
119
120
```

descrip_length_buckets	Avg_Rating
Long	3.855946944988041
Medium	3.232809430255403
Short	2.533613445378151

A new app creator can also take some inspiration from apps that have already done well. Below is a query that produces the highest user rated app in each genre. For example, if the new app developer was considering building a new Health & Fitness app, they could look to see what made Yoga Studio so successful. Some possible ways to go about this would be to read through the reviews with high and low ratings, to find what users like and did not like. For the features that users did not like about Yoga Studio, it would be beneficial to see what those app developers did to remedy those complaints and try to avoid them at the launch of a new app.

```
121 -- Lets check the top rated apps for each genre.
122
123 SELECT
124     prime_genre,
125     track_name,
126     user_rating
127 FROM (
128         SELECT
129         prime_genre,
130         track_name,
131         user_rating,
132         RANK() OVER(PARTITION BY prime_genre ORDER BY user_rating DESC, rating_count_tot DESC) AS rank
133         FROM AppleStore
134         ) AS a
135 WHERE
136     a.rank=1
137
```

prime_genre	track_name	user_rating
Book	Color Therapy Adult Coloring Bo...	5
Business	TurboScan™ Pro - document & ...	5
Catalogs	CPlus for Craigslist app - mobile...	5
Education	Elevate - Brain Training and Ga...	5
Entertainment	Bruh-Button	5
Finance	Credit Karma: Free Credit Score...	5
Food & Drink	Domino's Pizza USA	5
Games	Head Soccer	5
Health & Fitness	Yoga Studio	5
Lifestyle	ipsy - Makeup, subscription and ...	5
Medical	Blink Health	5
Music	Tenuto	5
Navigation	parkOmator – for Apple Watch	5

A note:

These are only some of the insights that can be drawn from this dataset, and there are many ways to further explore some of these considerations. When actually working with a stakeholder, they would likely have some expectations in mind that the analysis can later be catered towards. The analysis done so far is a good start but the ideas can be further expanded depending on the needs of the app developer.