Capstone Project-1

Play Store App Review Analysis

Team

Rishika Verma

PROBLEM STATEMENT

During the course of time with the increasing awareness of the people with regard to Google Play Store, it has become certain that now a days it has become the one stop platform for the people to download apps of various categories. Google Play Store dispenses millions of apps to people across the globe and hence can aptly be said as the hub of apps for Android devices. Advancement in Information Technology and increasing awareness of people with regard to digitalization have made the Google Play Store the most solicited digital store. In our project a comprehensive analysis of Play Store App has been done by taking into account and subsequently comparing

thousands of apps and reviews across different categories. The fetched data will help in formulating strategies for the growth and retention

APPROACH

In this project, we analyzed thousands of applications on Google Play Store and their corresponding reviews to understand the Android App market. The project has been divided into 2 parts: Apps analysis and Sentiment Analysis of User reviews. Initially, we imported, cleaned and analyzed various features of the Apps dataset through data exploration and visualizations.

In the next part, we loaded and cleaned the User Reviews dataset. The two datasets are then merged to visualize the composition of the total reviews and the sentiment polarity distribution.

DATA SUMMARY

Apps DataFrame Summary

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
8143	Breweries (CZ/SK)	LIFESTYLE	4.4	1740	4.4M	50,000+	Free	0	Teen	Lifestyle	July 27, 2017	3.1.0	4.1 and up
10279	Underdog FC	SPORTS	4.6	216	40M	10,000+	Free	0	Everyone	Sports	August 2, 2018	23.17	4.1 and up
72	Android Auto - Maps, Media, Messaging & Voice	AUTO_AND_VEHICLES	4.2	271920	16M	10,000,000+	Free	0	Teen	Auto & Vehicles	July 11, 2018	Varies with device	5.0 and up
4206	H-Kakashi - theme Xperia™	PERSONALIZATION	4.6	621	5.4M	50,000+	Free	0	Everyone	Personalization	December 1, 2017	a.2	4.4 and up
5039	AF Hydro	HEALTH_AND_FITNESS	2.0	136	15M	10,000+	Free	0	Everyone	Health & Fitness	May 12, 2018	1.0.5	4.4 and up

After proper Data Cleaning

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9660 entries, 0 to 9659
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   App             9660 non-null   object 
 1   Category        9660 non-null   object 
 2   Rating          8197 non-null   float64
 3   Reviews         9660 non-null   object 
 4   Size            9660 non-null   object 
 5   Installs        9660 non-null   object 
 6   Type            9659 non-null   object 
 7   Price           9660 non-null   object 
 8   Content Rating  9659 non-null   object 
 9   Genres          9660 non-null   object 
 10  Last Updated    9660 non-null   object 
 11  Current Ver     9652 non-null   object 
 12  Android Ver     9657 non-null   object 
dtypes: float64(1), object(12)
memory usage: 981.2+ KB
```

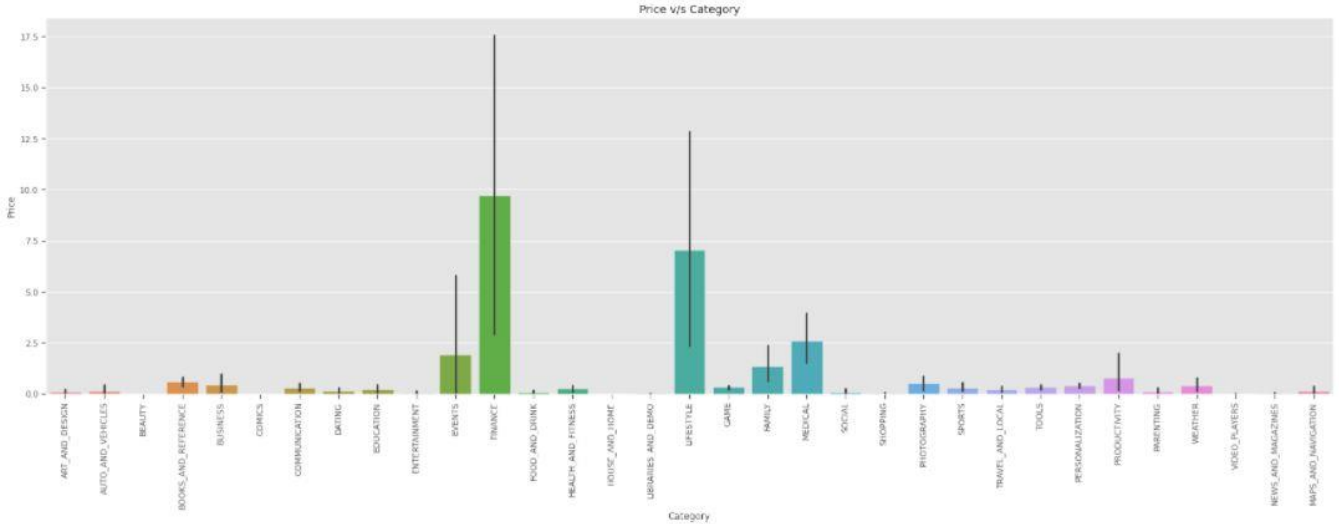
Correcting Data Type

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9660 entries, 0 to 9659
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   App             9660 non-null   object 
 1   Category        9660 non-null   object 
 2   Rating          9660 non-null   float64
 3   Reviews         9660 non-null   object 
 4   Size            9660 non-null   object 
 5   Installs        9660 non-null   object 
 6   Type            9660 non-null   object 
 7   Price           9660 non-null   object 
 8   Content Rating  9660 non-null   object 
 9   Genres          9660 non-null   object 
 10  Last Updated    9660 non-null   object 
 11  Current Ver     9660 non-null   object 
 12  Android Ver     9660 non-null   object 
dtypes: float64(1), object(12)
memory usage: 981.2+ KB
```

Missing Value Procedure Done

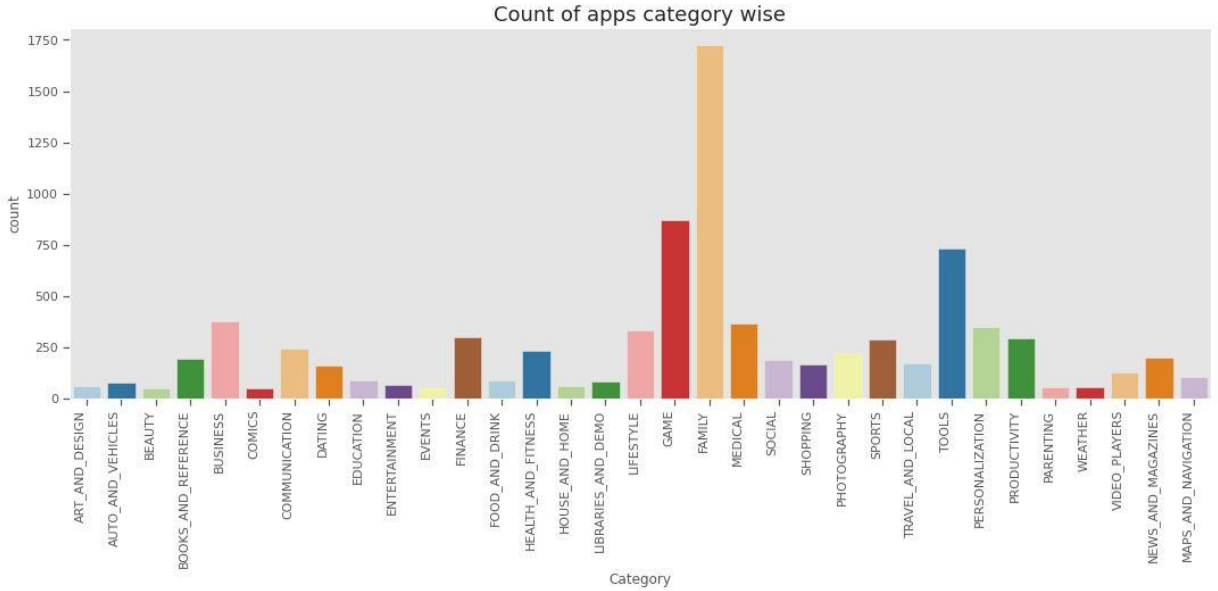
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8422 entries, 0 to 8431
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   App             8422 non-null   object 
 1   Category        8422 non-null   object 
 2   Rating          8422 non-null   float64
 3   Reviews         8422 non-null   int64  
 4   Size            8422 non-null   float64
 5   Installs        8422 non-null   int64  
 6   Type            8422 non-null   object 
 7   Price           8422 non-null   float64
 8   Content Rating  8422 non-null   object 
 9   Genres          8422 non-null   object 
 10  Last Updated    8422 non-null   object 
 11  Current Ver     8422 non-null   object 
 12  Android Ver     8422 non-null   object 
dtypes: float64(3), int64(2), object(8)
memory usage: 921.2+ KB
```

DATA VISUALIZATIONS

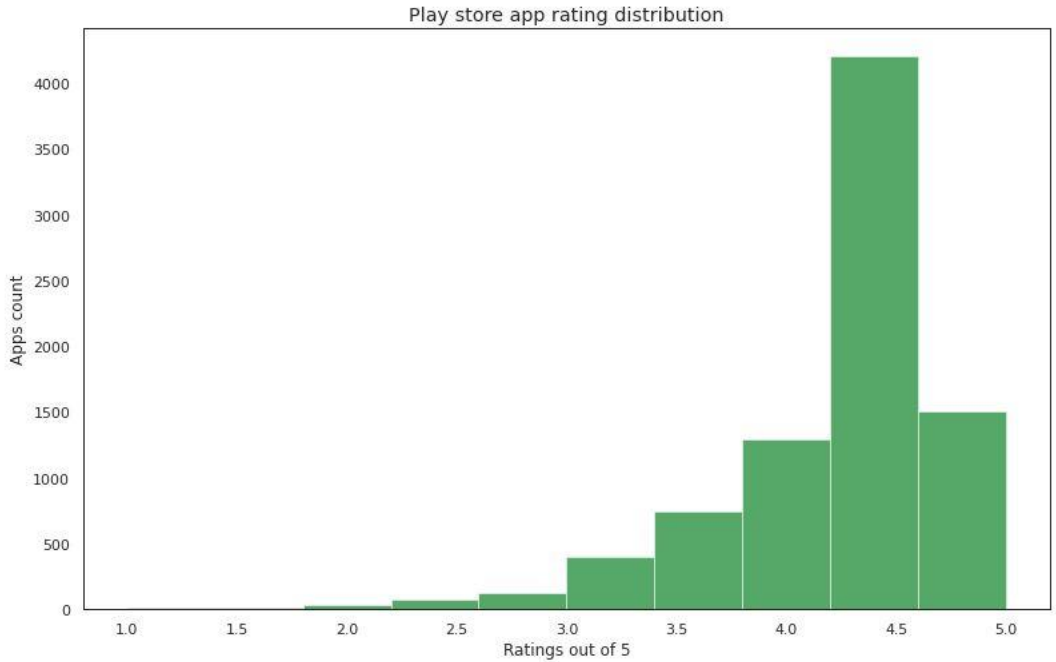


<Figure size 2160x576 with 0 Axes>

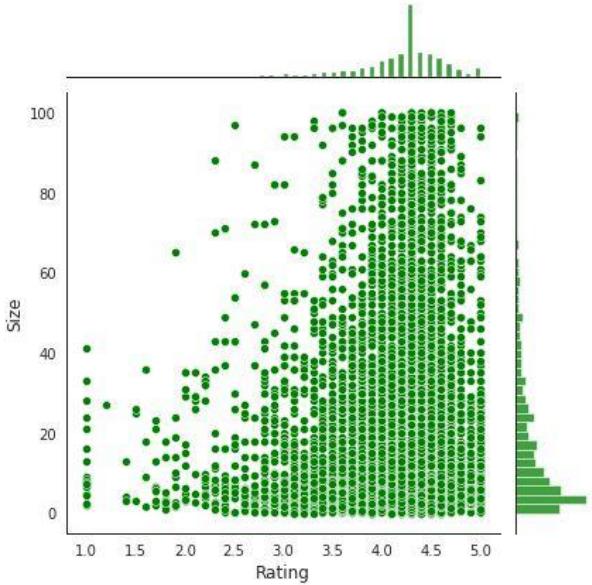
Category Wise count of App



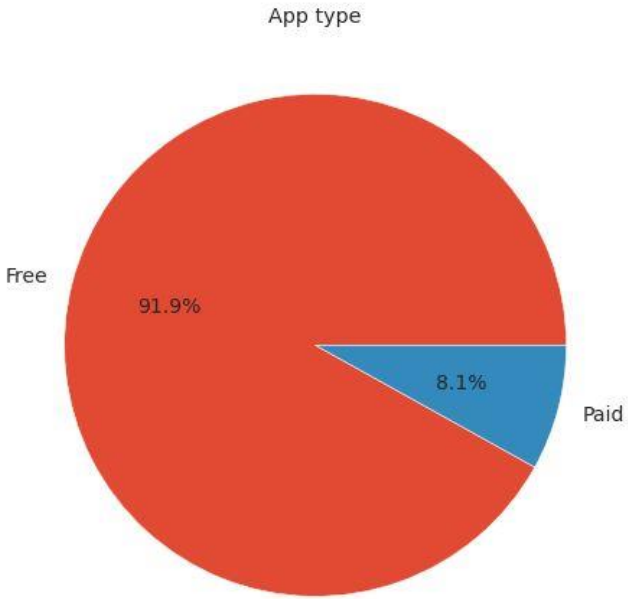
Rating of Apps

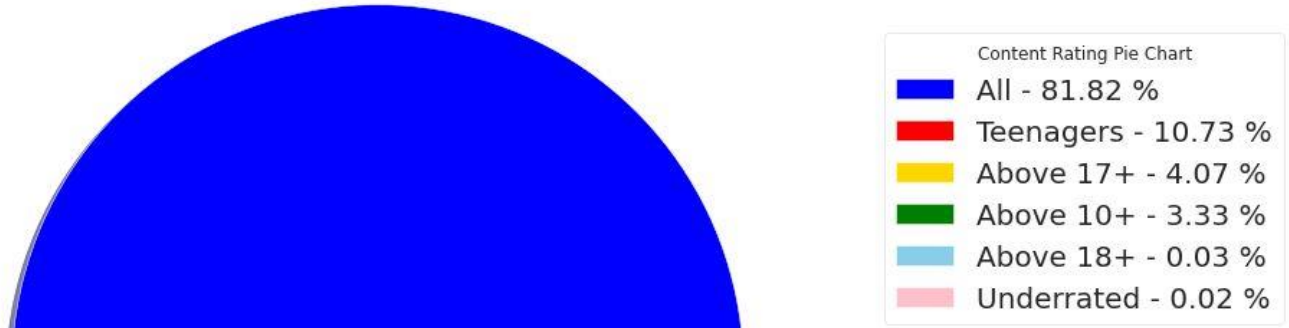


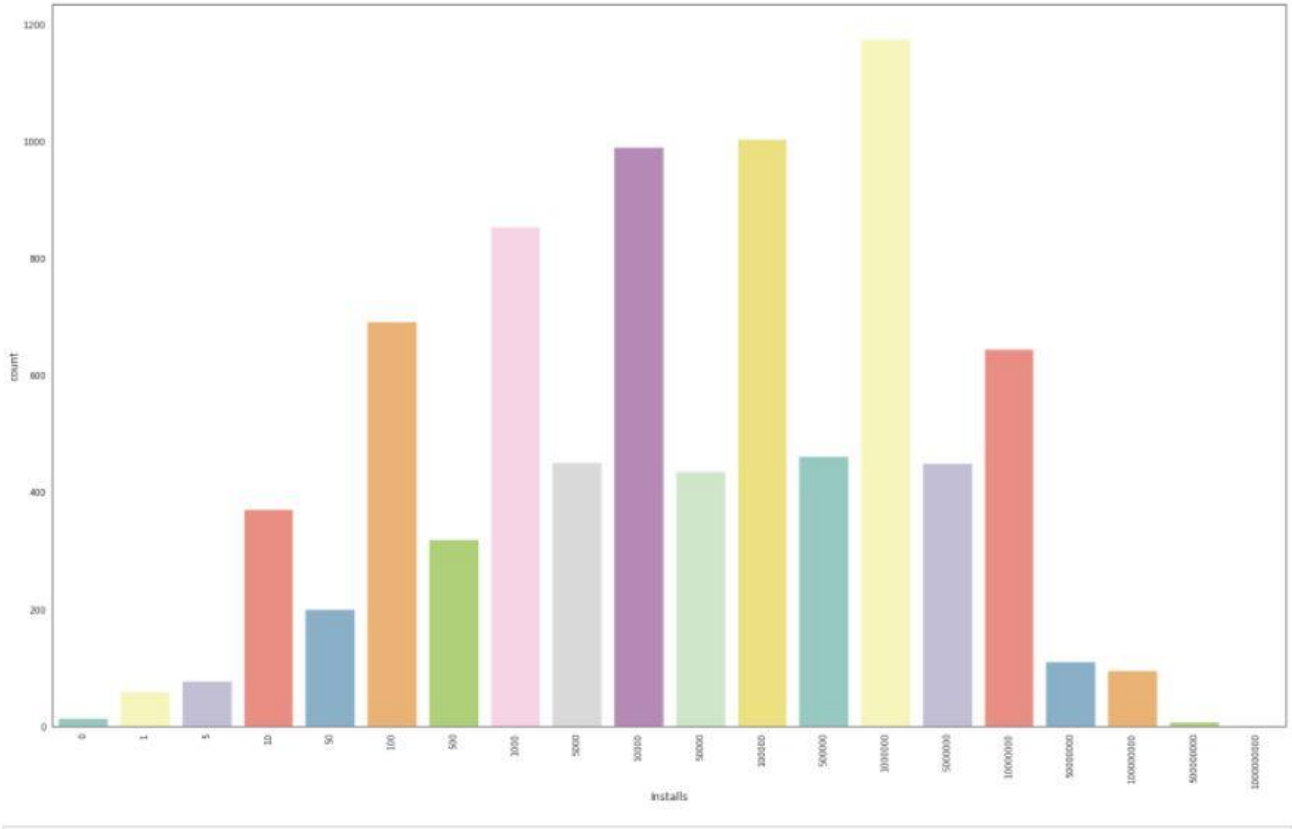
Size of Apps



Price Category

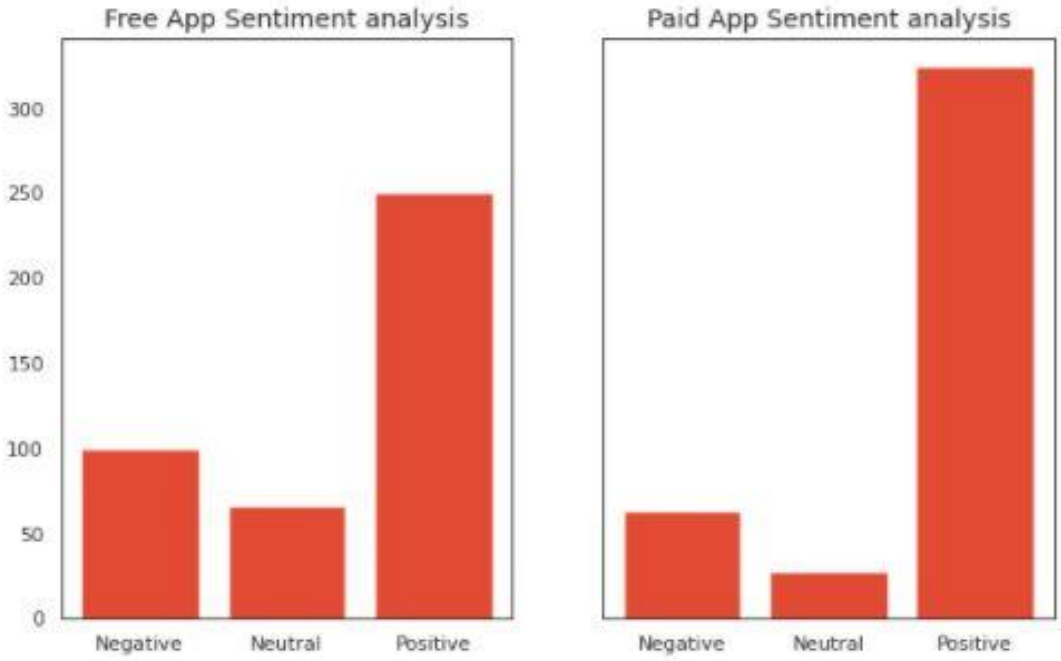



Content Rating Pie Chart
All - 81.82 %
Teenagers - 10.73 %
Above 17+ - 4.07 %
Above 10+ - 3.33 %
Above 18+ - 0.03 %
Underrated - 0.02 %


1200
1000
800
600
400
200
0
count
0
1
5
10
50
100
500
1000
5000
10000
50000
100000
500000
1000000
5000000
10000000
50000000
100000000
500000000
1000000000
installs

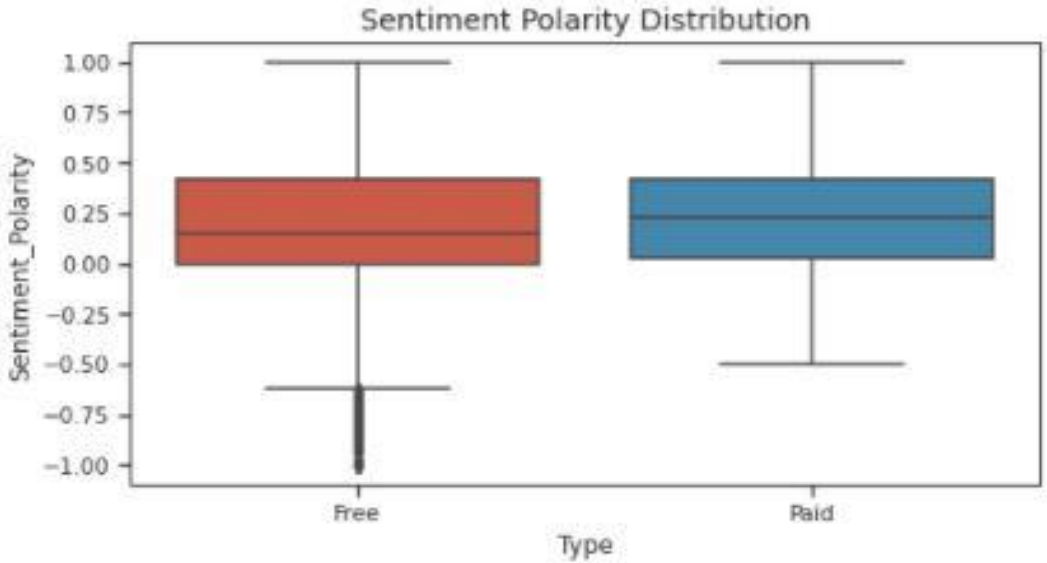
Rating
Reviews
Size
Installs
Price
1
0.062
0.046
0.041
-0.019
0.062
1
0.18
0.6
-0.0087
0.046
0.18
1
0.13
-0.022
0.041
0.6
0.13
1
-0.01
-0.019
-0.0087
-0.022
-0.01
1
Rating
Reviews
Size
Installs
Price
1.0
0.8
0.6
0.4
0.2
0.0

SENTIMENT ANALYSIS OF USER REVIEW


Free App Sentiment analysis
300
250
200
150
100
50
0
Negative
Neutral
Positive
Paid App Sentiment analysis
Negative
Neutral
Positive

Sentiment Analysis
Sentiment Subjectivity
Sentiment Polarity
Sentiment
Positive
Neutral
Negative

Sentiment Polarity



WORDCLOUD plot for positive and negative reviews

still
using
playing
fix
take
ads
way
account
play
worst
know
keep
good
people
day
time
much
back
try
money
level
new
open
fun
phone
update
terrible
want
thing
say
problem
pay
screen
always
start
add
let
lot
see
give
work
make
bad
player
well
suck
never
need
annoying
hate
great
useless
show
tried
stupid
star
issue
getting
frustrating
slow
waste time
got
think
option
load
hard
wrong
card
trying
horrible
hour
change
anything
used
disappointed
help
minute
free
something
little
without
put
second
log
another
buy
stop
enough
nothing
dit
ridiculous
actually
find
working
dificult
first
point
picture
sometime
control
uninstall
different
crashes
passwords
notification
purchase
bug
data
lost
tell
month
full

people
well
star
better
way
want
look
time
lot
best
free
nice
back
thank
take
using
find
need
great
make
update
problem
help
awesome
easy
work
day
feature
phone
add
play
think
level
keep
far
see
new
good
thing
know
fun
say
option
still
much
fix
able
change
sometime
never
right
many
playing
useful

CONCLUSIONS

1. The most popular App Category is "Game".
2. A large number of Apps fall into "Family" Category i.e., this is the category with highest number of subsequent apps.
3. The total average rating of Play Store Apps is 4.18.
4. The App Categories with least and highest average ratings are "Dating" and "Events" respectively.
5. Free apps are highly popular when compared to Paid apps.
6. Most of the apps receive Positive reviews i.e., about 63.5%. Free apps might have a broader range of reviews than Paid apps, but the median of Sentiment Polarity is higher in Paid apps.