Capstone Project:

EDA ON PLAYSTORE APP REVIEW ANALYSIS

- By Mohammad Shahzeb Khan
Mohammad Ammaar
Nada Nasser
Gayatri Gupta
Mansur Shikalgar

DOCKET:

- Introduction
- Data Summary
- Defining Problem Statement
- Data Cleaning
- Exploration and Visualization
- Inferences and Conclusion

INTRODUCTION:

- **➢ Importance of apps**
- • An app for every utility
- • Their importance can’t be overstated
- **➢ Benefits to a business**
- • Greater reach due to smartphones
- • Loyalty and increased customer engagement
- **➢ Challenges and Opportunities**
- • Huge supply: 3.48 Mn apps, increased competition
- • Learn and leverage from existing apps; increase customer satisfaction and
- success
- **➢ Google Play Store dataset**
- • Apps and features
- • User Reviews

DATA SUMMARY

App- The app name
Category- Categorical label, which describes which broad category the app belongs to.
Rating- Continuous variable with a range from 0.0 to 5.0, which describes the average rating the app has received from the users.
Reviews- Continuous variable describing the number of reviews that the app received.
Size- The size of the app. The suffix M is used for megabytes, while the suffix K is used for kilobytes.
Installs- Categorical label that describes the number of installs.
Type- Label that indicates whether the app is free or paid.

shutterstock.com · 235251619

- **Price-** The price value for the paid apps.
- **Content-** Rating Categorical rating that indicates the age group for which the app is suitable.
- **Genre-** Semicolon separated list of genres to which the app belongs.
- **Last Update-** The date the app was last updated.
- **Current Version-** The current version of the app as specified by the developers.
- **Android Version-** The Android operating system the app is compatible with

Defining Problem Statement:

As we know that, there is no shortage when it comes to availability of apps. But one wishes to have the best app for some utility. The challenge is to create such an app, in the current competitive environment, and leverage from existing data. Our aim is to discover the factors/features on which the success (Installs) o f an app depends.

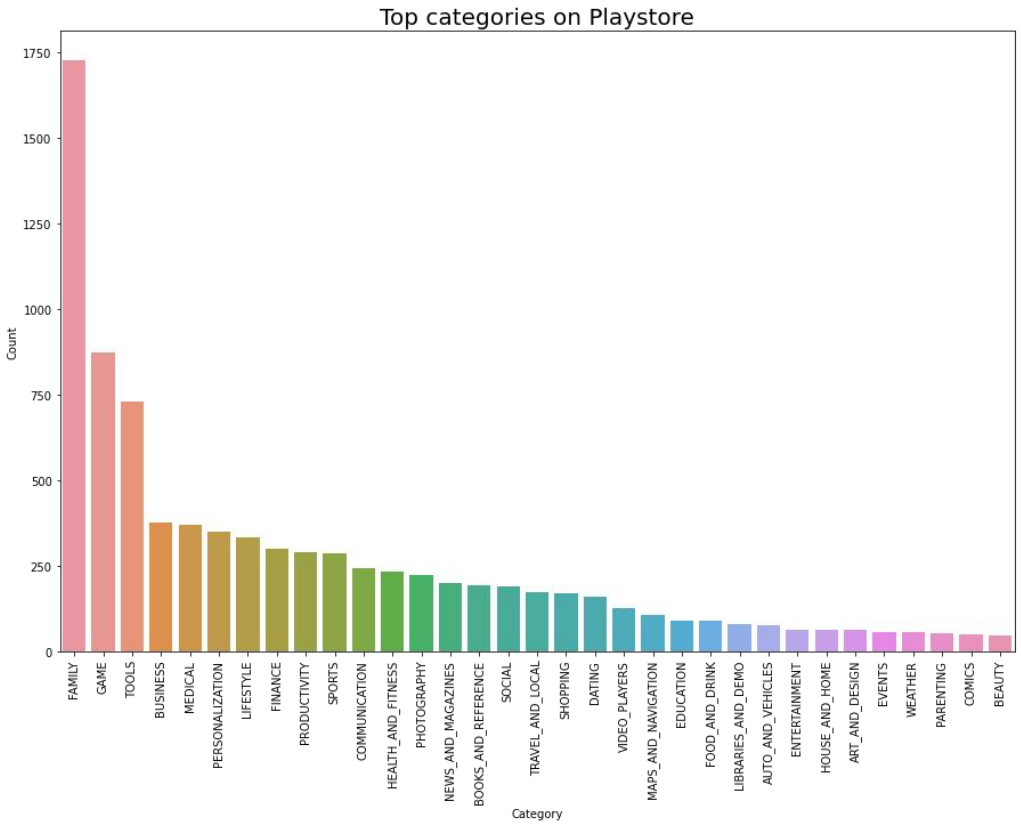
DATA CLEANING:

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled.

Checking for outliers in important column with respect to analysis

- ➢ **Dropping the error information in data (if any)**
- ➢ **Taking care of Null Values**
- ➢ **Converting data type of columns for further operation**

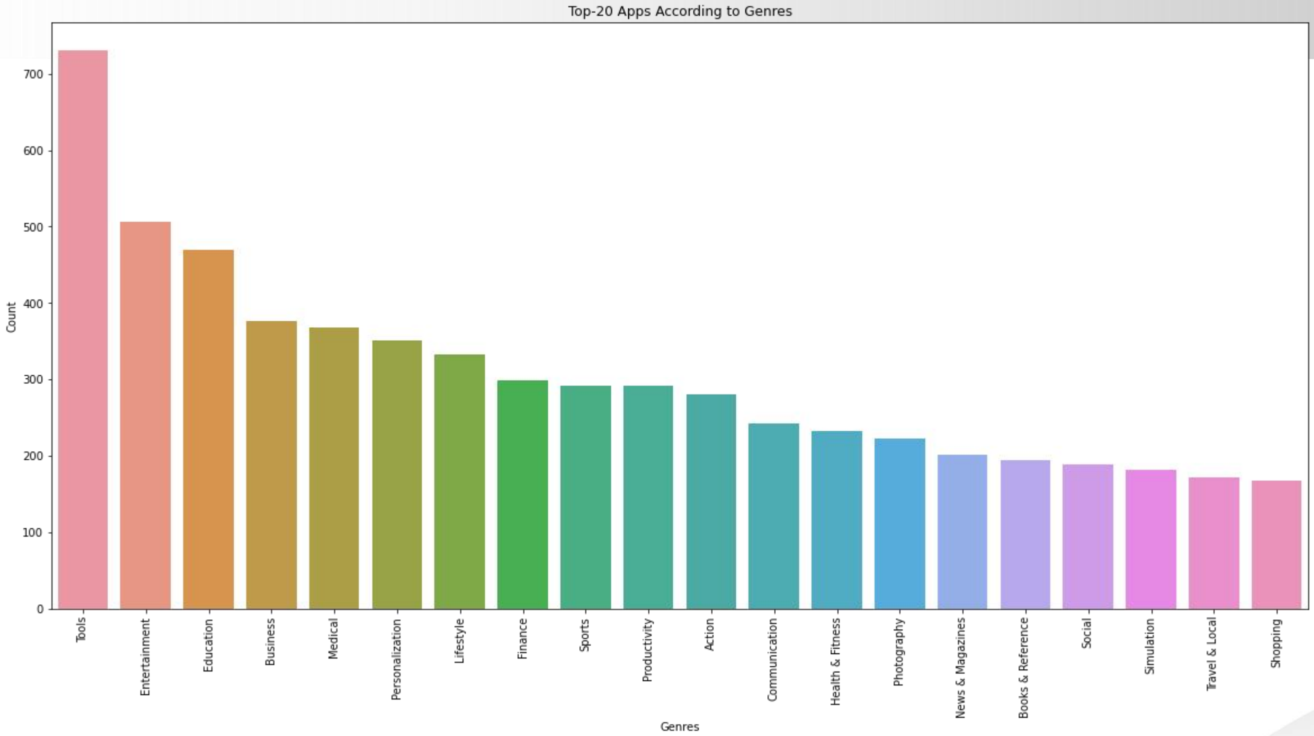
Exploration and Visualization



Question-1:(a) Finding Top apps in playstore as per Category

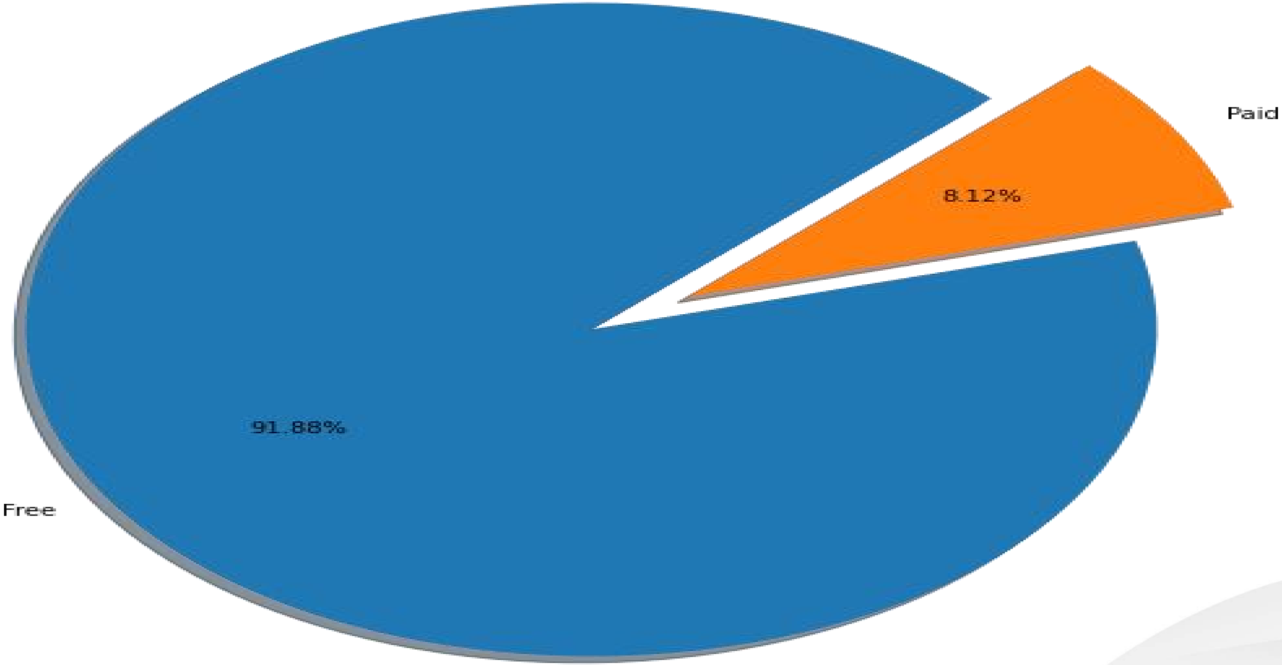
Obervation 1- **As we can see that in Category section 'Family', 'Games','Tools' winning the race. So this Data will give us brief the daily requirements of users.**

Question-1:(b) Finding Top 20 Apps in playstore as per Genres



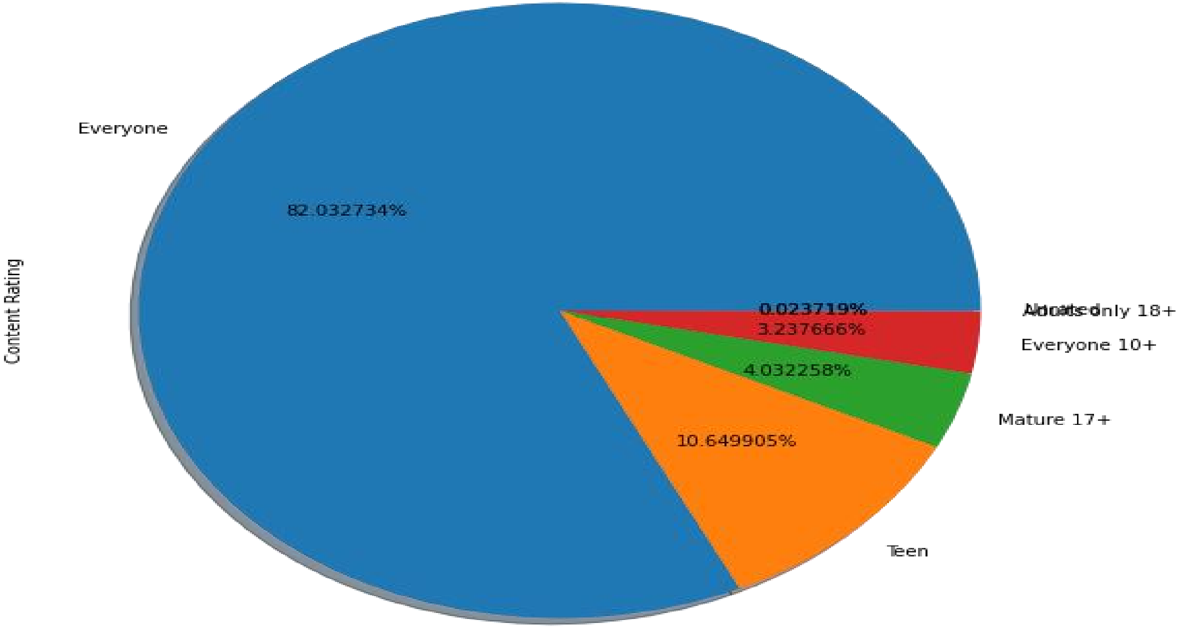
<u>Obsevation 1(b)</u>-By plotting the graph of Top Genres it is clear that the 'Tools','Entertainment','Action' topping the chart.

Question-2(a): Checking the proportion of Free and Paid Apps



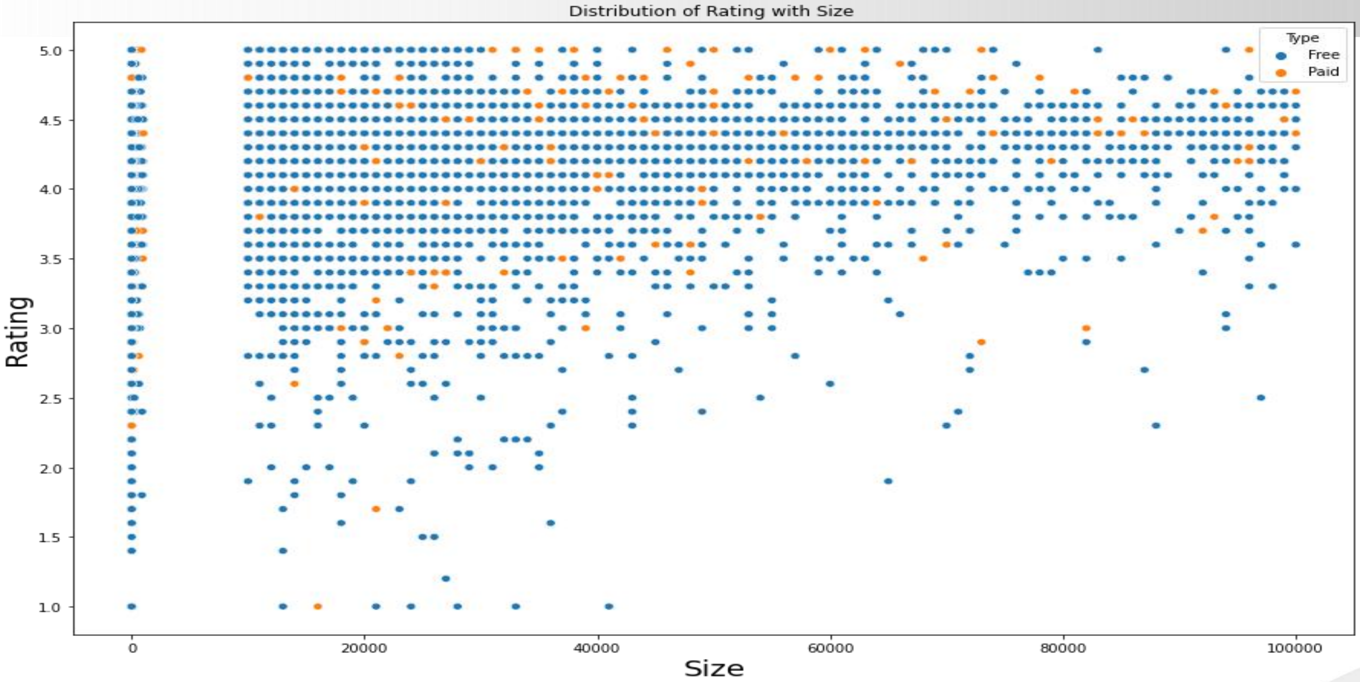
Observation 2(a)- From above Pie chart we can clearly see that majority of apps are free

Question-2(b)- Content Rating Ratios from all apps



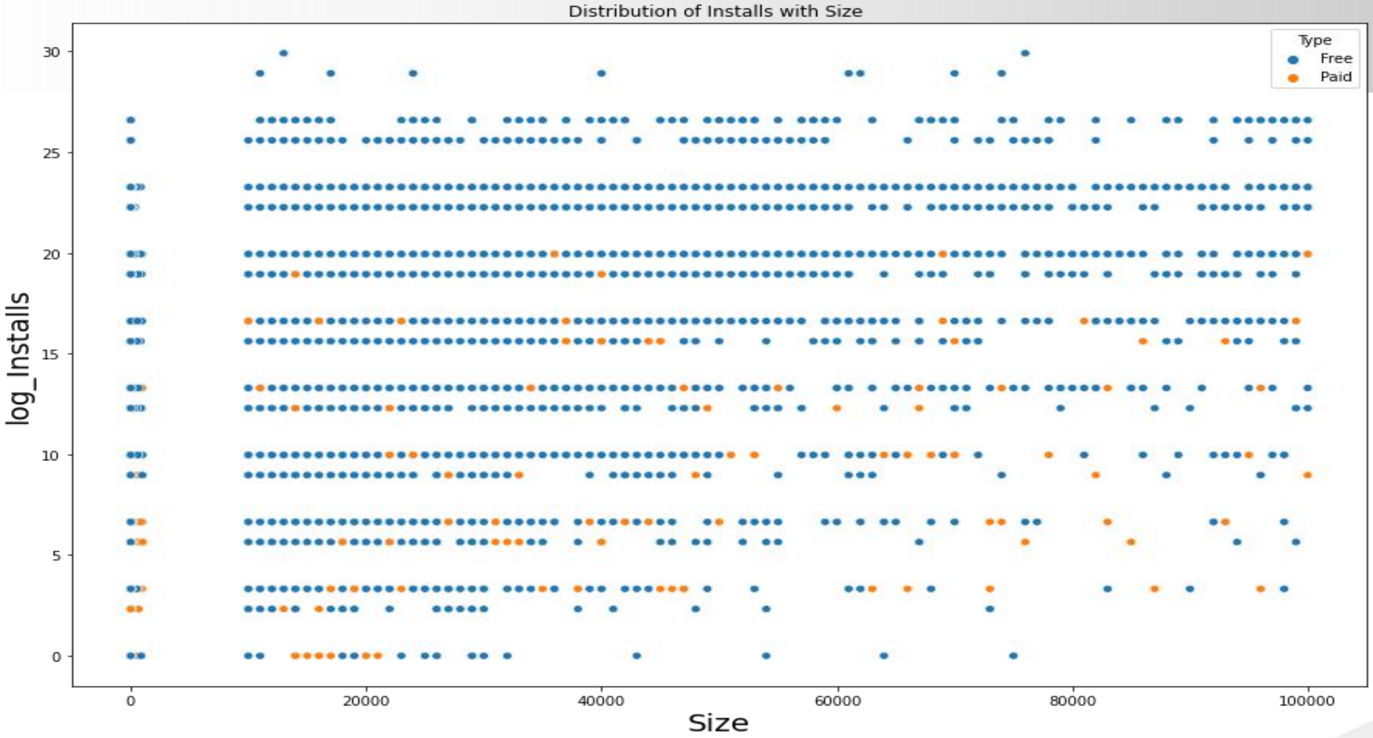
<u>Observation 2(b)-</u> The Majority Content of Apps in Playstore are everyone thus, installing a user-friendly environment

Question-3: Now checking the distribution of apps in terms of Size, Rating & Type



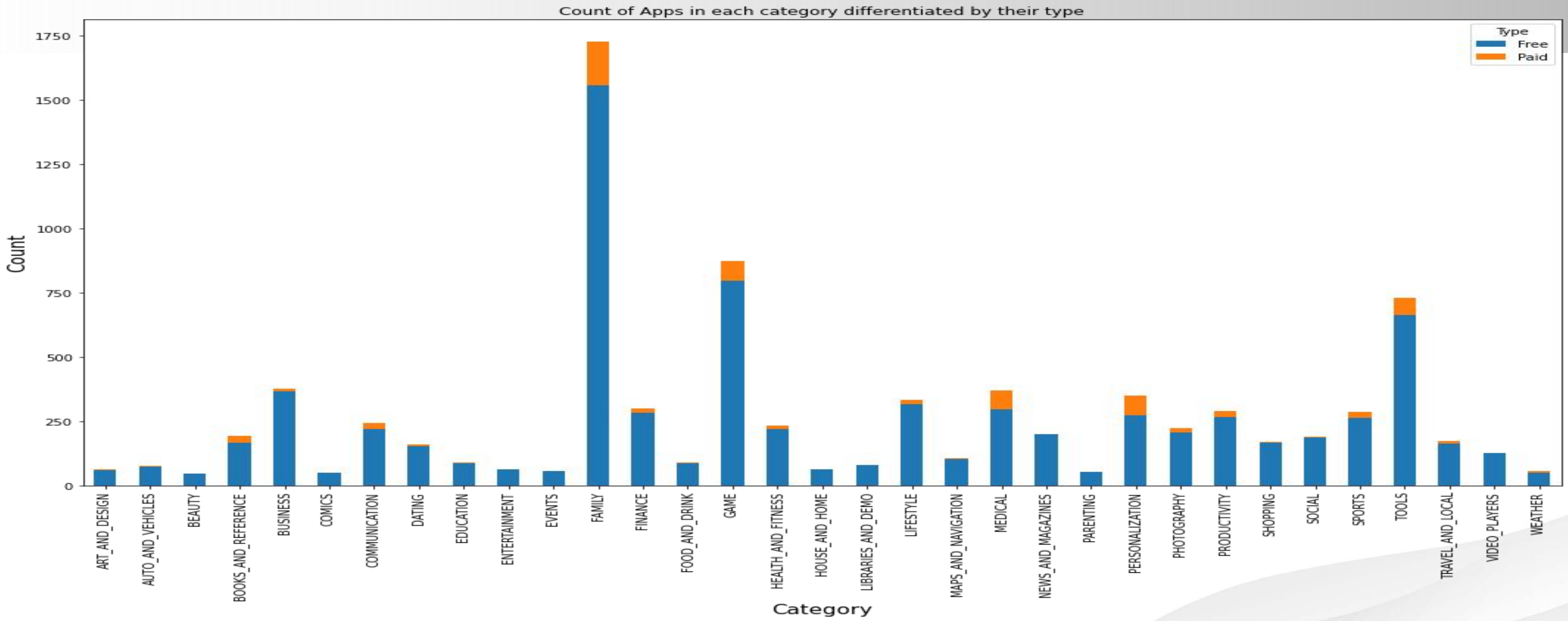
Observation 3- From above scatter plot, we can imply that majority of free apps are small in size and having high rating. While paid apps have quite equal distribution in terms of size and rating.

Question 3(b): Now checking the effect of size on thr Number of Installs

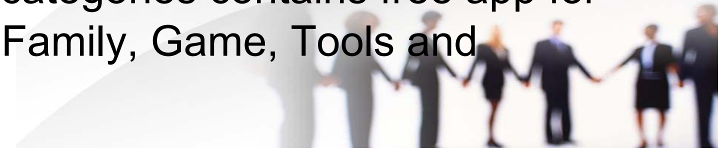


Observation 3(b)- Also, we can say that the bulky apps are less downloaded by user. Size does effect the Rating of the Apps

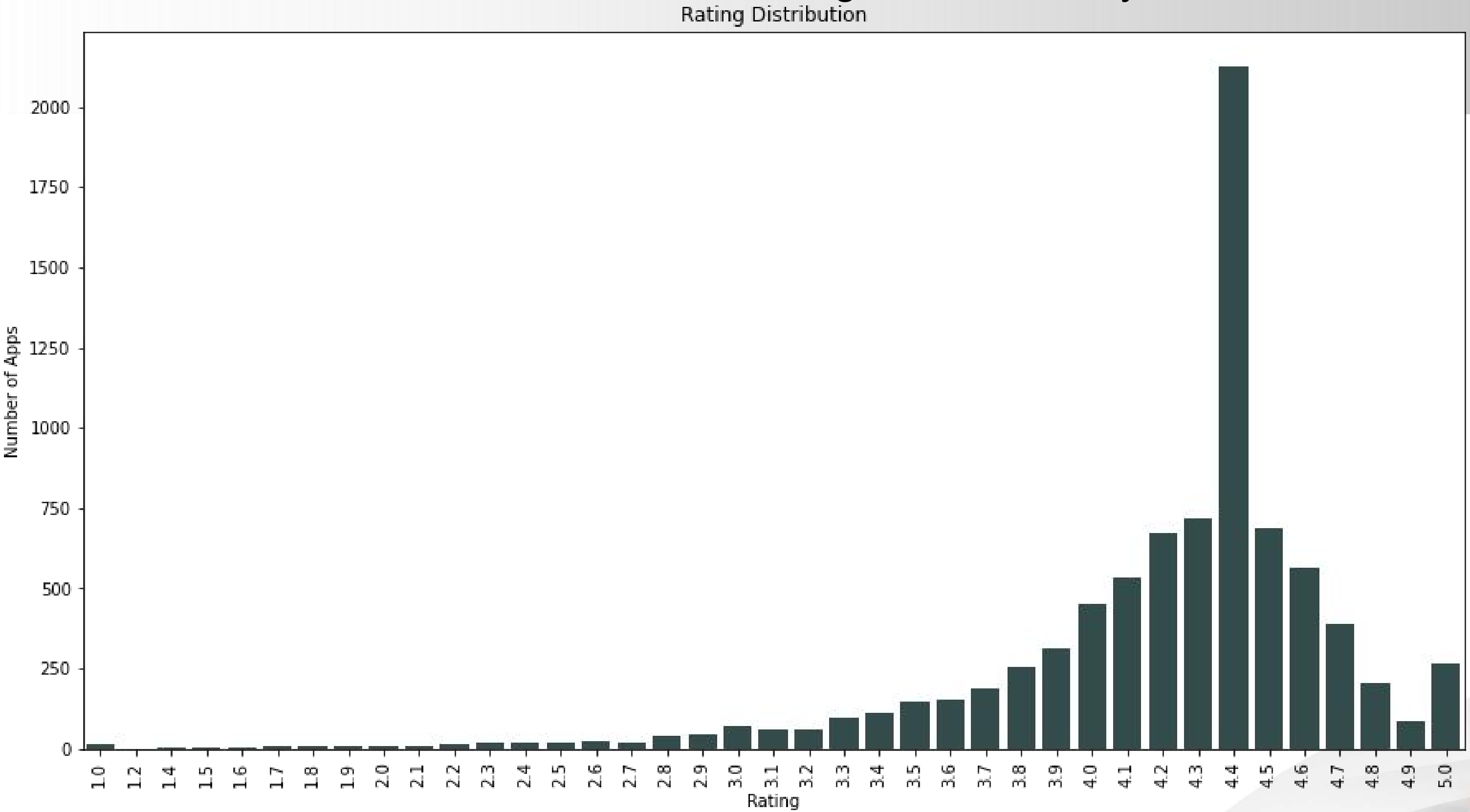
Question-4: Let us examine the Free and Paid Apps available according to Category



Observation 4- The bar plot shows clearly that majority of categories contains free app for download. The most paid apps availbale for download are in Family, Game, Tools and Medical category

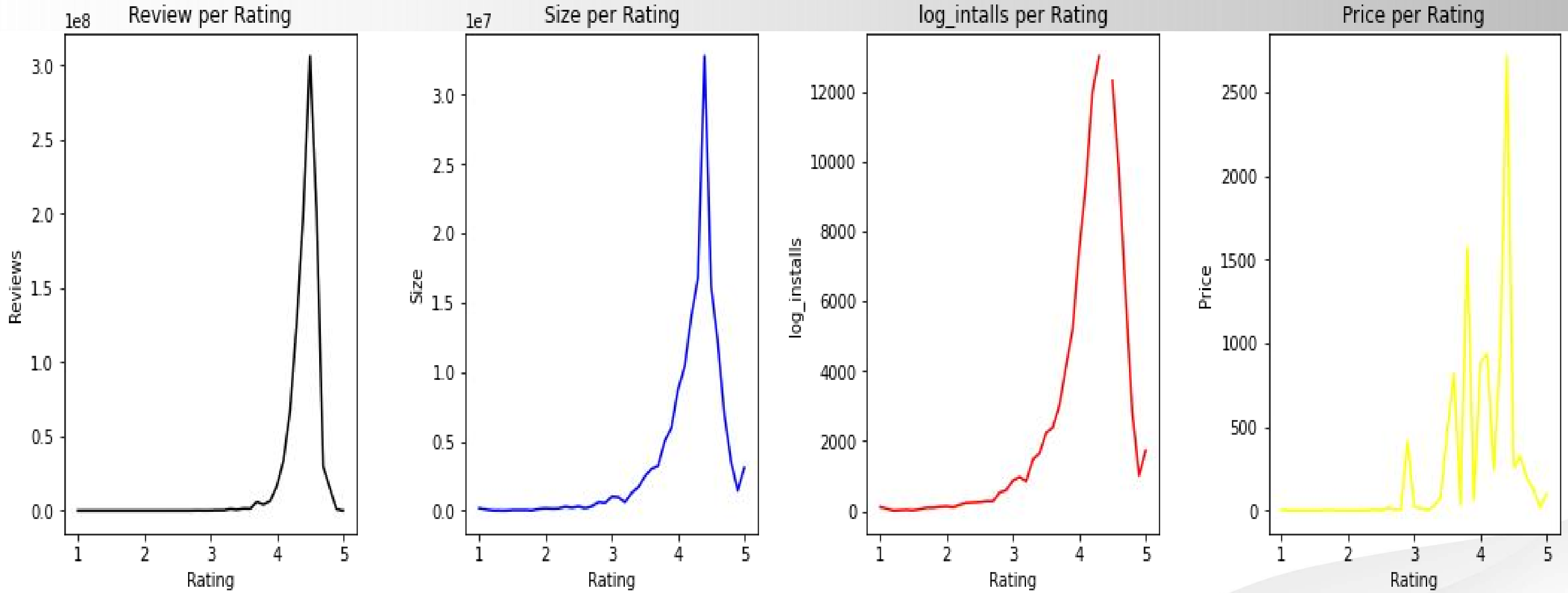


Question 5- Let's now dive into the Rating section and try to establish some meaningful insights.



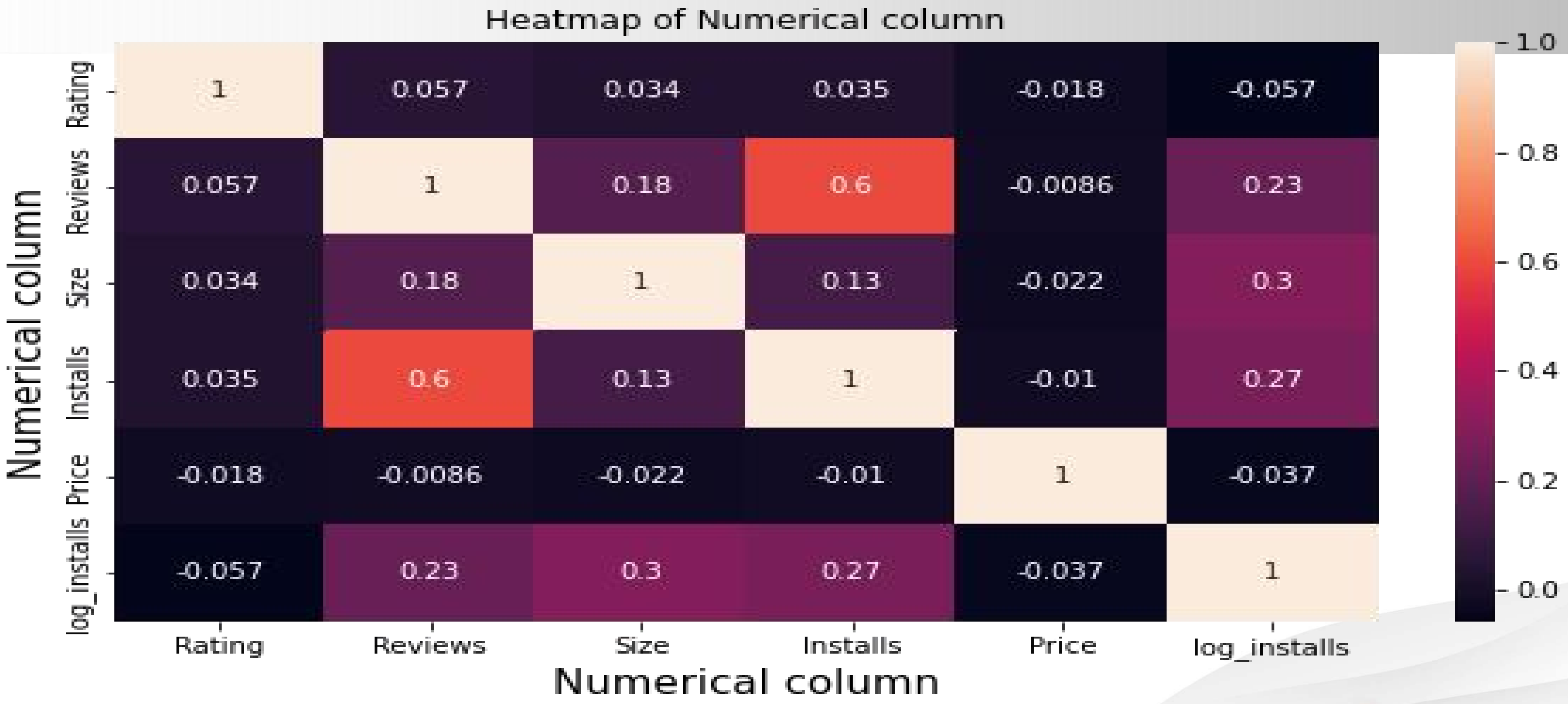
<u>Observation 5- </u>Well well as an honest user must review seems people had taken that fact too seriously as most of the apps are rated above 4 and in the range of 4-4.6

Question 5(b)- Checking the relation of Review, Size, Installs, Price with Rating section and try to establish some meaningful insights



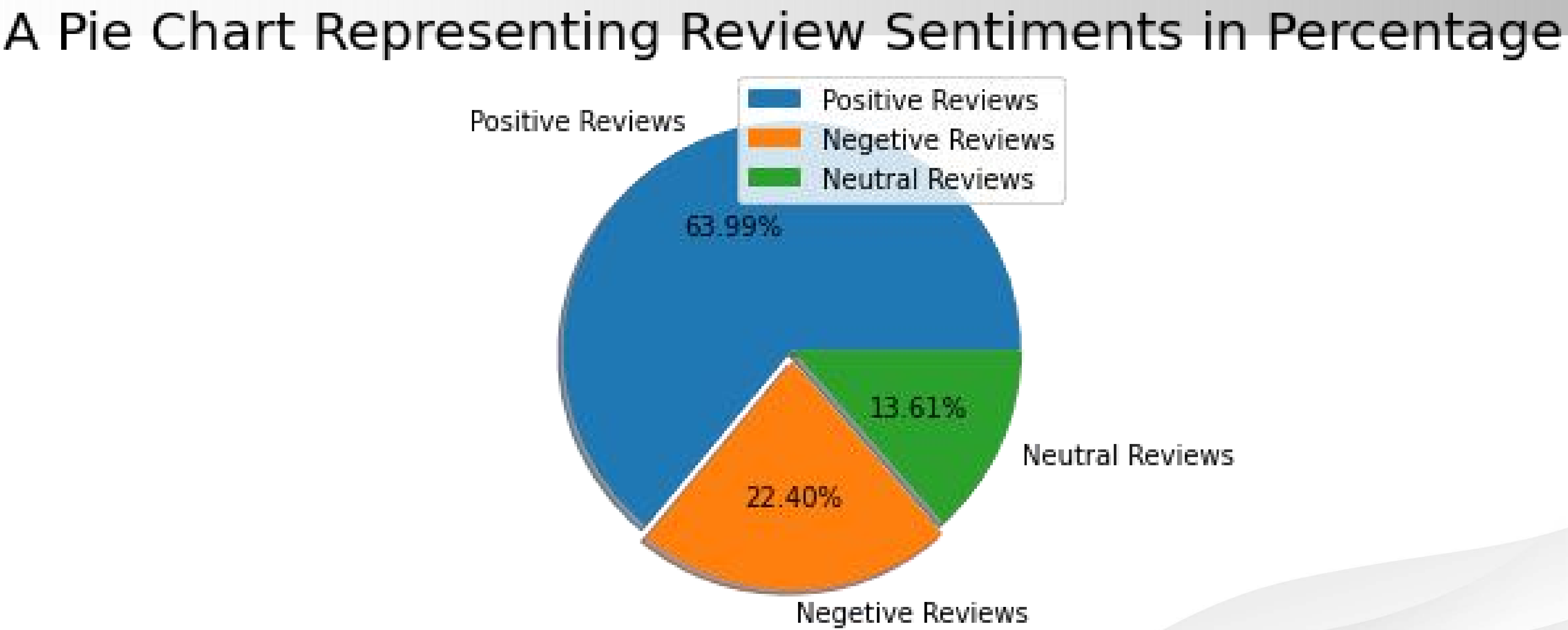
Observation 5- From the above plotting, we can say that most the apps with higher rating range of 4-4.7 are having high amount of reviews, size, and installs. In terms of price, it doesn't reflect a direct relationship with rating, as could see a fluctuation even at the range of high rating

Correlation- The corr() method calculates the relationship between each column in your data set.



Observation 6- From above heat map we can clearly say that Install and Review are correlated.

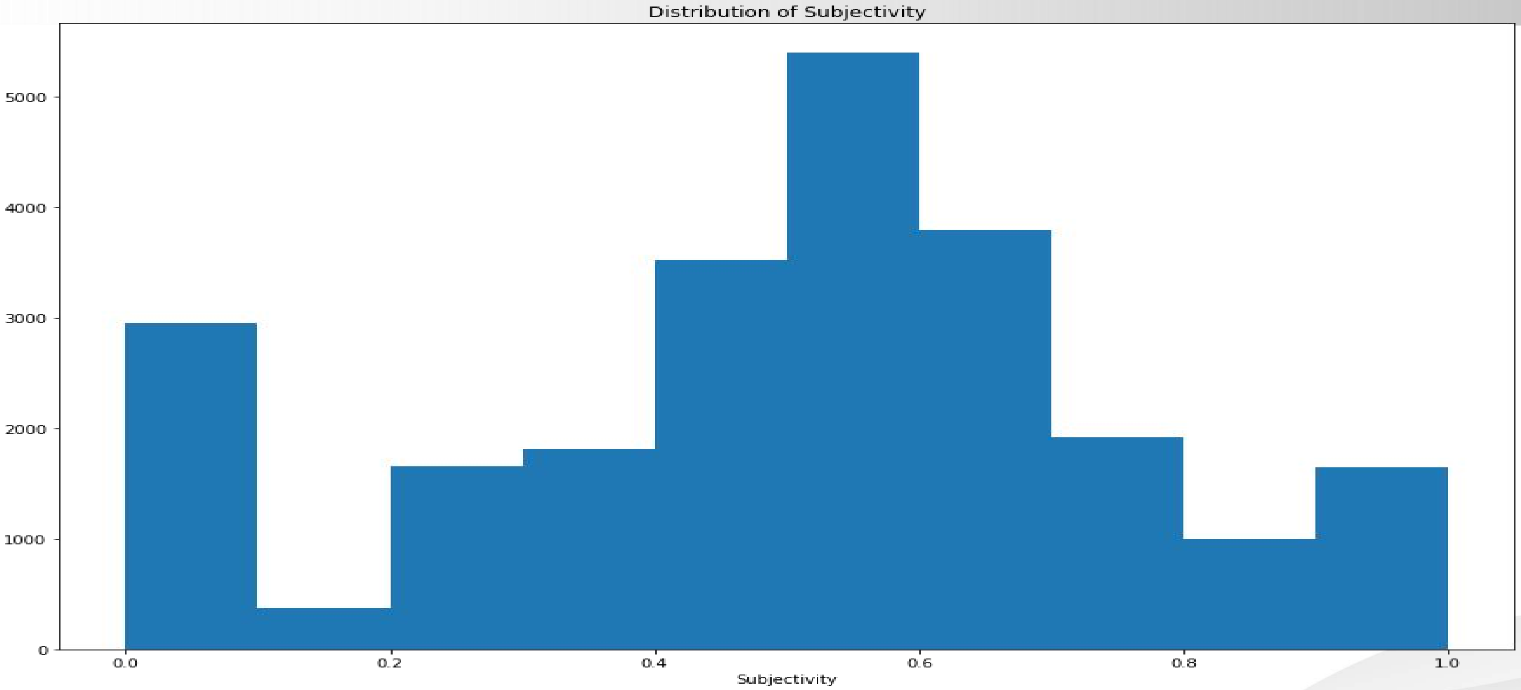
Question 7- Checking the Sentiments Review of different users in Pie Chart



A Pie Chart Representing Review Sentiments in Percentage

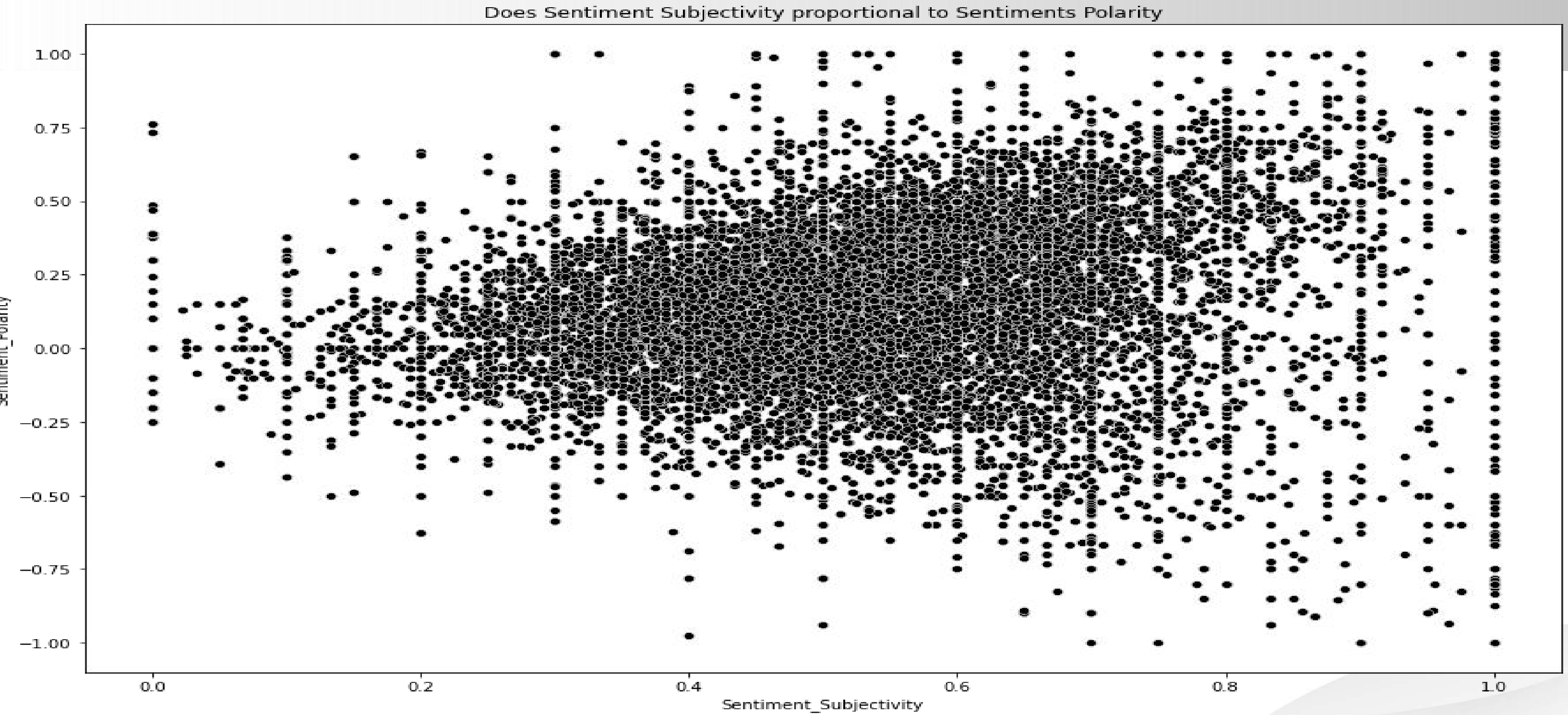
Observation 7- Alas, Positive Reviews from the positive mentality of Users.

Question 7(b)- Checking the Sentiments Subjectivity of different users on Histogram.



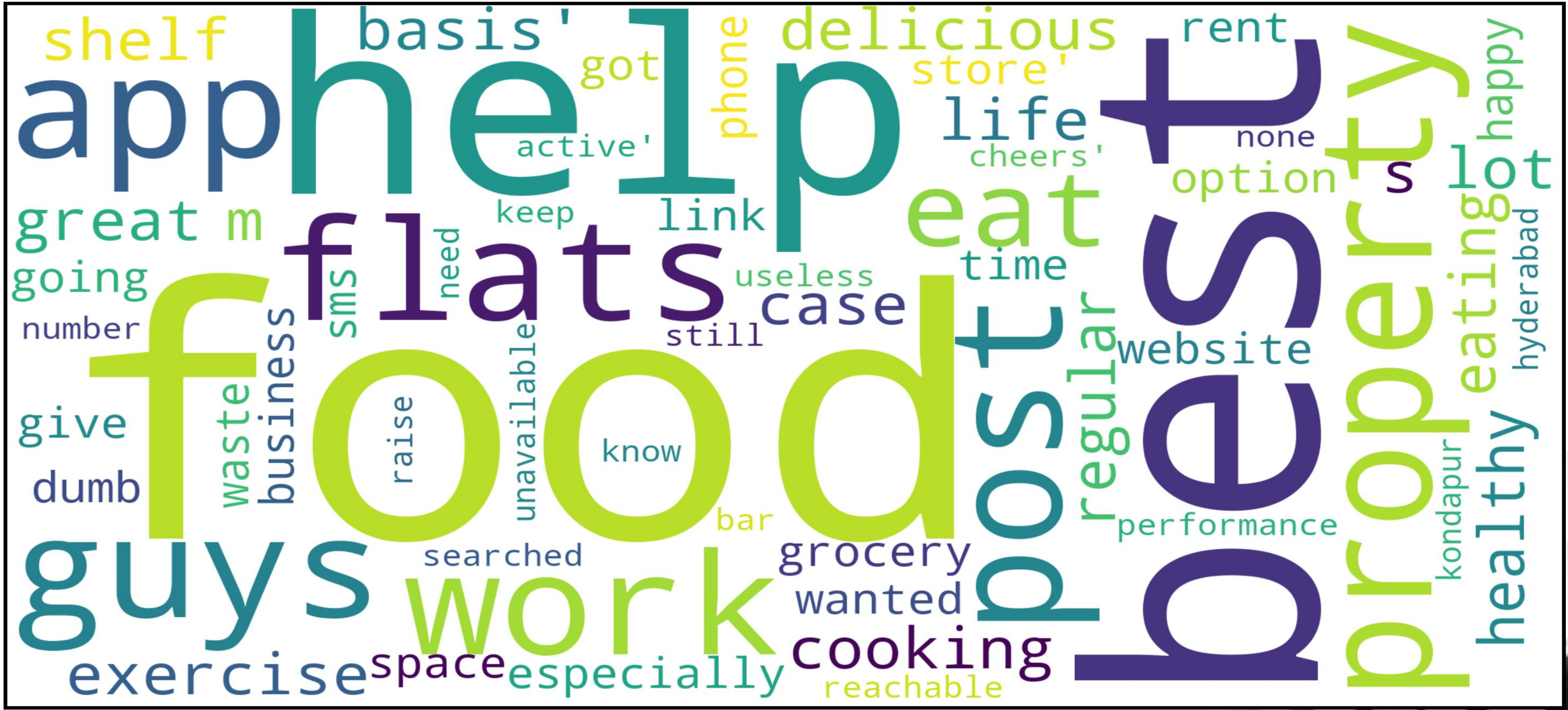
Observation 7(b)- It can be seen that maximum number of sentiment subjectivity lies between 0.4 to 0.7. From this we can conclude that maximum number of users give reviews to the applications, according to their experience

Question 8- Does Sentiment Subjectivity proportional to Sentiments Polarity?



Observation 7- From the above scatter plot it can be concluded that sentiment subjectivity is not always proportional to sentiment polarity but in maximum number of case, shows a proportional behavior, when variance is too high or low.

POSTER BOY!!!



shelf
basis'
delicious
rent
app
help
got
phone
store'
property
happy
life
none
active'
cheers'
best
s
option
lot
great
m
flats
keep
link
eat
eating
hyderabad
going
need
useless
time
sms
case
number
food
still
post
regular
website
business
unavailable
give
waste
raise
know
healthy
kondapur
dumb
bar
performance
guys
searched
grocery
work
wanted
cooking
exercise
space
especially
reachable

Inferences and Conclusion

The Google Play Store Apps report provides some useful insights regarding the trending of the apps in the play store.

As per the graphs visualizations shown above, most of the trending apps (in terms of users' installs) are from the categories like GAME, COMMUNICATION, and TOOL even though the amount of available apps from these categories are twice as much lesser than the category FAMILY.

The trending of these apps are most probably due to their nature of being able to entertain or assist the user. Besides, it also shows a good trend where we can see that developers from these categories are focusing on the quality instead of the quantity of the apps.

Other than that, the charts shown above actually implies that most of the apps having good ratings of above 4.0 are mostly confirmed to have high amount of reviews and user installs.

There are some spikes in term of size and price but it shouldn't reflect that apps with high rating are mostly big in size and pricy as by looking at the graphs they are most probably are due to some minority.

Eventhough apps from the categories like GAME, SOCIAL, COMMUNICATION and TOOL of having the highest amount of installs, rating and reviews are reflecting the current trend of Android users, they are not even appearing as category in the top 5 most expensive apps in the store (which are mostly from FINANCE and LIFESTYLE).

As a conclsuion, we learnt that the current trend in the Android market are mostly from these categories which either assisting, communicating or entertaining apps.

THANK YOU!!!