Capstone Project 1

PLAY STORE APP REVIEW ANALYSIS

BY – GAURAV BISHT

ANALYSIS PLAY STORE APP

1. Defining complication statements.
2. Inspecting the dataset.
3. Assembling the dataset, getting it ready.
4. Importing different libraries & attributes (Google play store & user reviews).
5. Exploratory data analysis [EDA].
    - % of each application app.
    - Categories having highest & lowest no. of apps.
    - Reviewing top 5 categories.
    - Paid vs Free
    - Review on the basis of category wise dataset
    - Average rating of apps in google play store.
    - Categories & their sizes.
    - Highest number of installed apps.
    - Heatmap for numerical columns
    - Inferences & conclusion
    - Q & A

Google play also branded as the google play store is a digital app which is run by google and is widely used by billions of people, downloading there favourite app according to their need, it's a app used for certified devices running over Andriod OS as well as Chrome OS, over 3+ million apps are there in play store, The total revenue of the platform is over $11.2 billion (82,000 crores) in 2019.

COMPLICATION STATEMENT

Play store is a digital marketplace app which is used to download a variety of apps for andriod smartphones, as smartphons are increasing day by day and peoples need and requirement over apps is also increasing accordingly, also it's a very profitable & money making market for app developers.

As peoples demand changes with time and trend according to their need and comfort so developers need to know this thing too, like social, entertainment and games categories also are in hype.

So it's always important to find out what type of app is required in market and what's trending as what people are downloading and liking the most before developing the app,

Also it's important to keep in mind that how size, rating, type, price & review affect the sentiment of apps and user

INSPECTING THE DATASET

A) Play Store Data.csv - This data.csv contains all the required information about the applications present in google play store app, as it contains 13 attributes or bascially subsets that describes the given app. It has 10841 apps out of which 9659 are unique.

B) User Review.csv – This user.csv files contains 4 attributes i.e. App, Review, Sentiment Category & Sentiment Score.

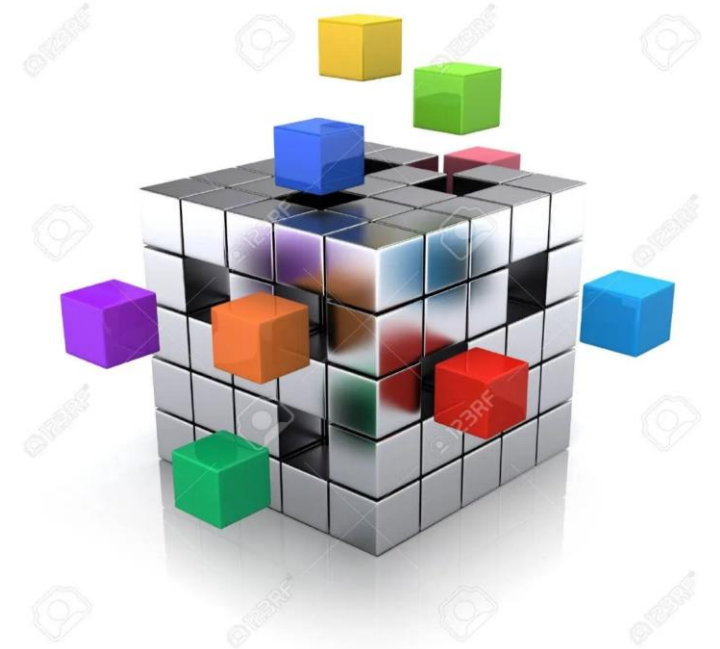
ASSEMBLING THE DATASET

Loading the Dataset – We are given two different dataset play store dataset & user review of apps, so here we'll import various python inbuild libraries as - numpy & pandas.

Importing libraries – Using Matplotlib, seaborn & plotly to work on the dataset.

Data cleaning – Here we'll remove null values, find and remove identical or duplicate data, remove duplicate rows also filling the missing values with mode & numerical values with median, fixing prize, size and other attributes.

Exploratory Data Analysis – Analyzing the data and using different statical graphs and also data visualizations methods to make it more presentable.



ATTRIBUTES (Google play store)

App : Name of the app.

Rating : Gives current average rating (out of 5) of the app on Google Play.

Reviews : Number of user reviews given on the app.

Category : Category of the app. for examples are: FAMILY, SOCIAL, TOOLS, COMMUNICATION etc.

Size : Size of the app in MB (megabytes), KB (kilobytes), GB (gigabytes).

Type : paid or free.

Price : Price of the app in US$ on Google Play Store.

Last Updated : Date on which the app was last updated on Google Play.

Installs : NO. of times any given app is or was downloaded/installed from Google Play.

Content Rating : Sutable age group for different Applications.

Genre : App can belong to any version.

Android Version : It tells you on which version the app will run or will work.

Current Version : It will the current version of the app.

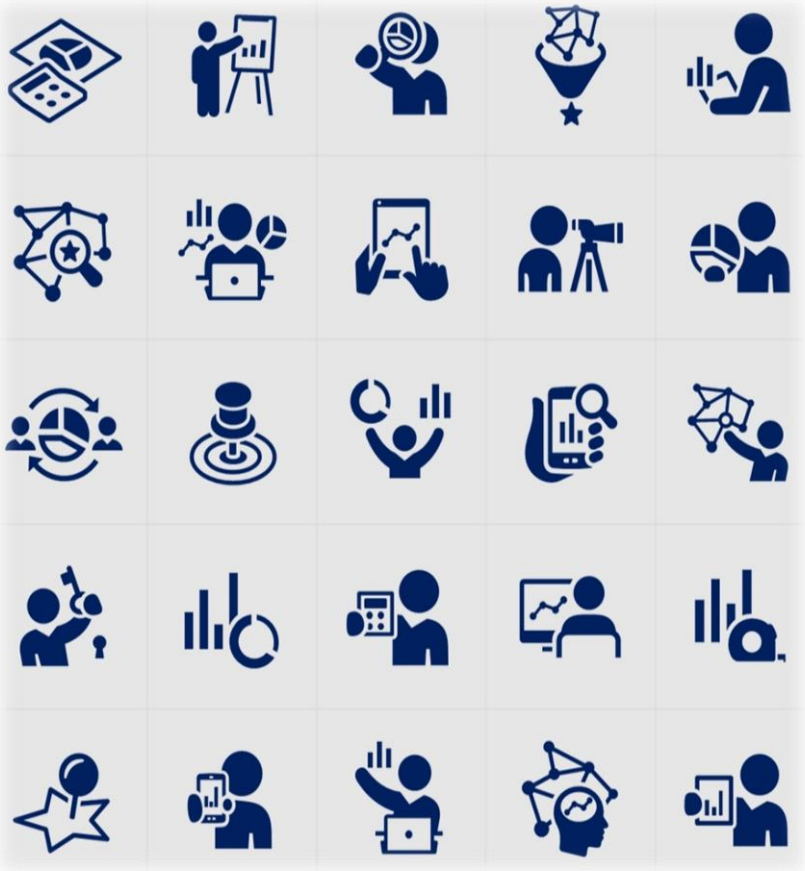
ATTRIBUTES (User Review)

Review : Preliminary processed user review text.

App : Name of the app on which the user review was provided. Matches the App column of the apps.csv file.

Sentiment Category : User review - Positive, Negative or Neutral.

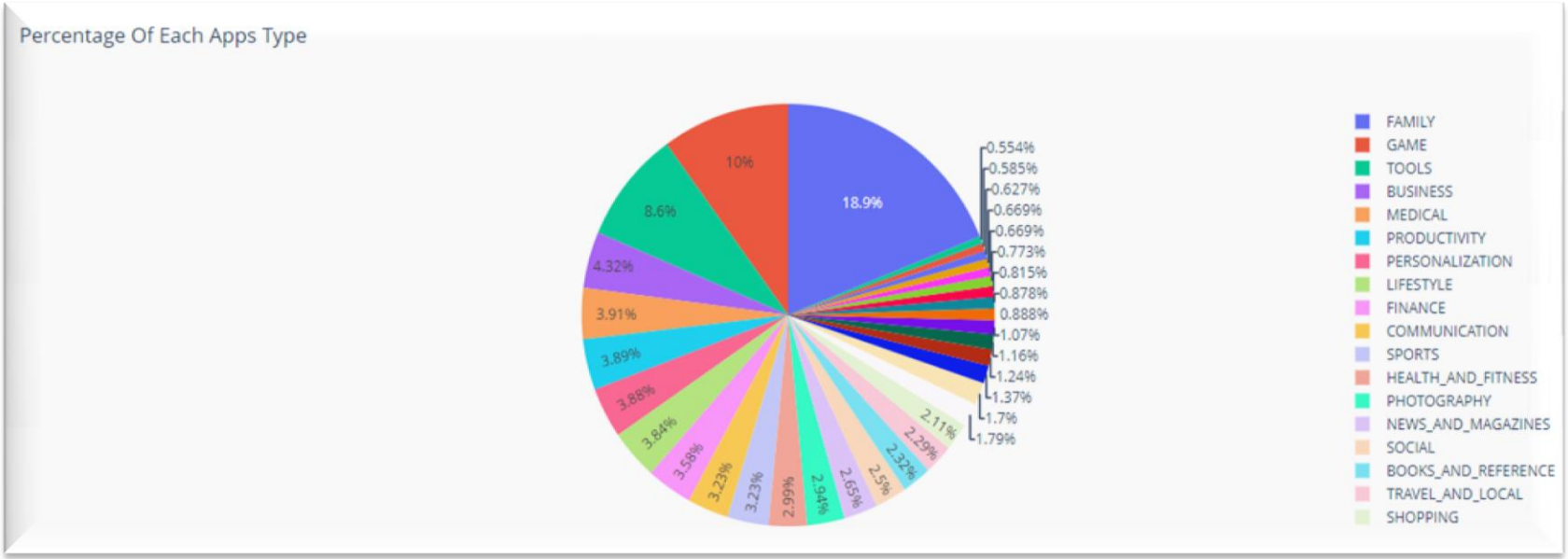
Sentiment Score : Sentiment score of the user review.



EXPLORATORY DATA ANALYSIS [EDA]

1. Percentage of each application type in google play store -

Here clearly we can se that categories like FAMILY, GAME & TOOLS have the highest % share in play store



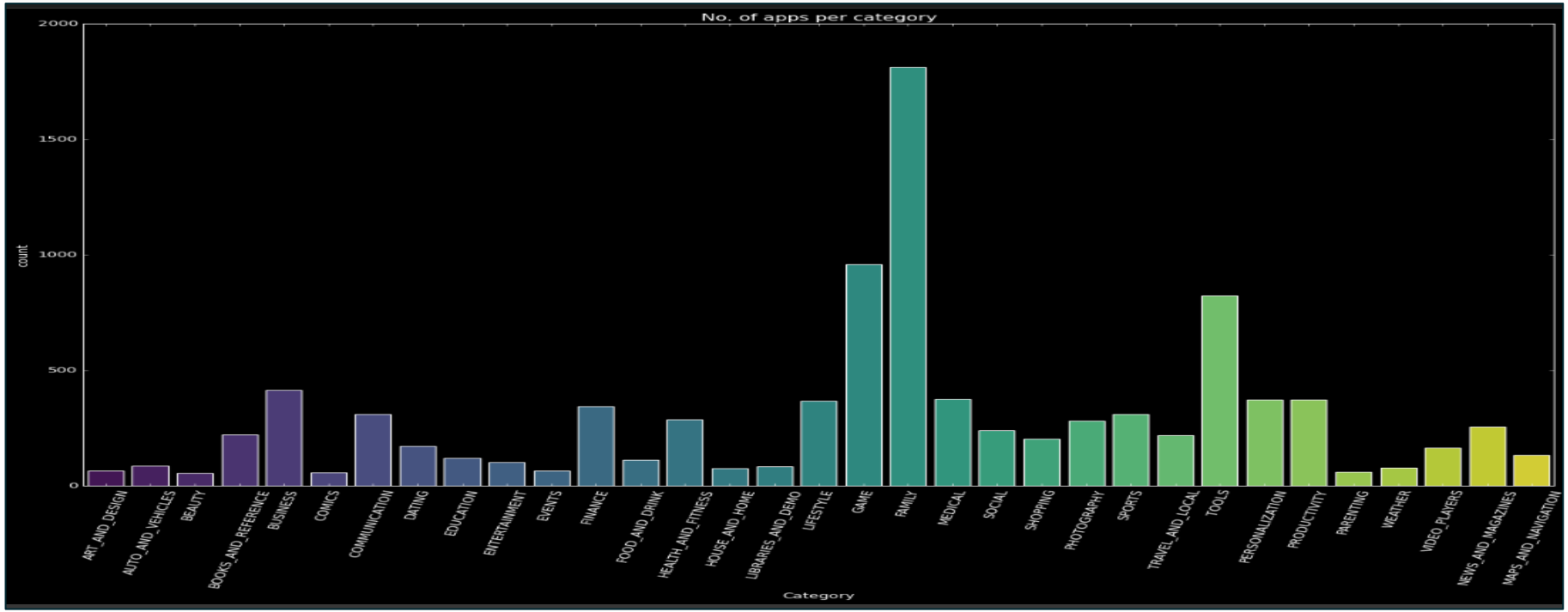
EXPLORATORY DATA ANALYSIS [EDA]

2. Categories having highest & lowest number of applications in play store.

ANALYSIS –

Highest no. Of apps – **Family**

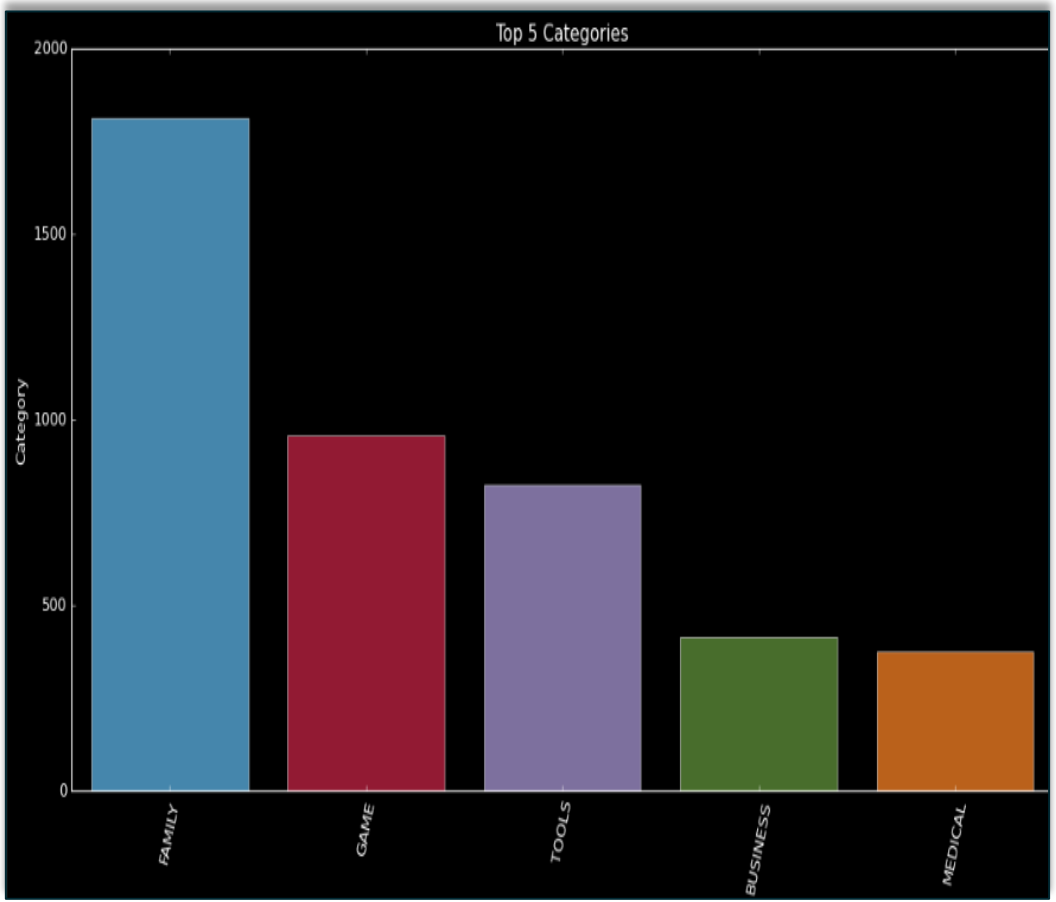
Lowest no. Of apps – **Beauty**



Reviewing top 5 categories.

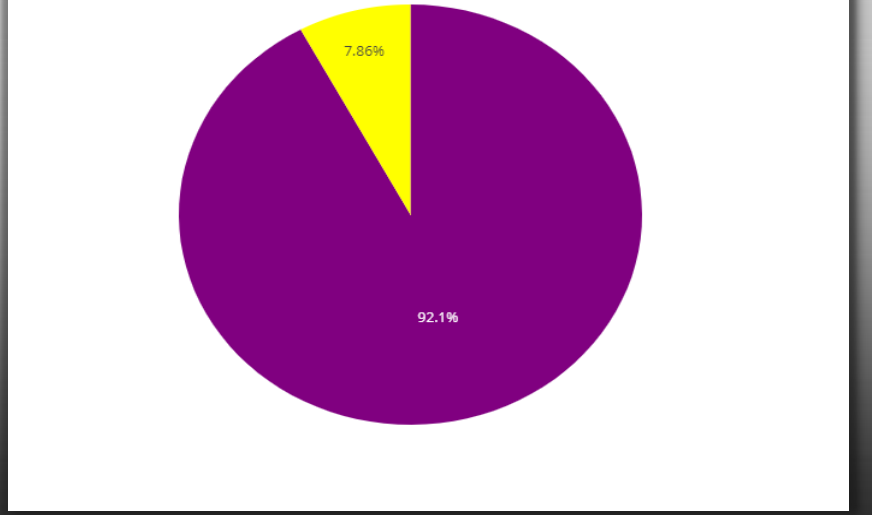
The top 5 categories having max number of apps are as follows –

1. Family
2. Game
3. Tools
4. Business
5. Medical



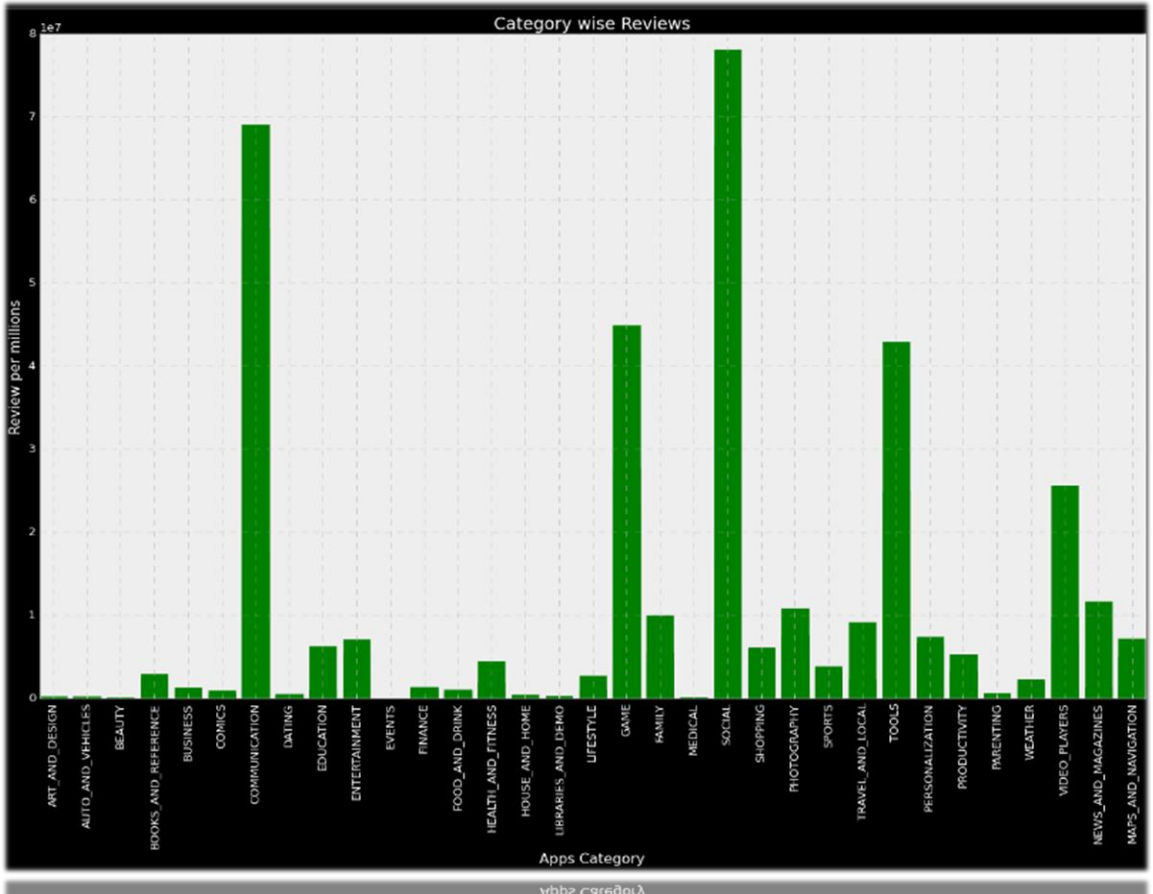
3. Percentage of paid & free apps & also the amount of paid & free apps in different categories in google play store.

Over 92.1% are free & rest 7.86% are paid in play store, Another pie chart denotes the amount of paid & free apps in diff. Categories in play store.

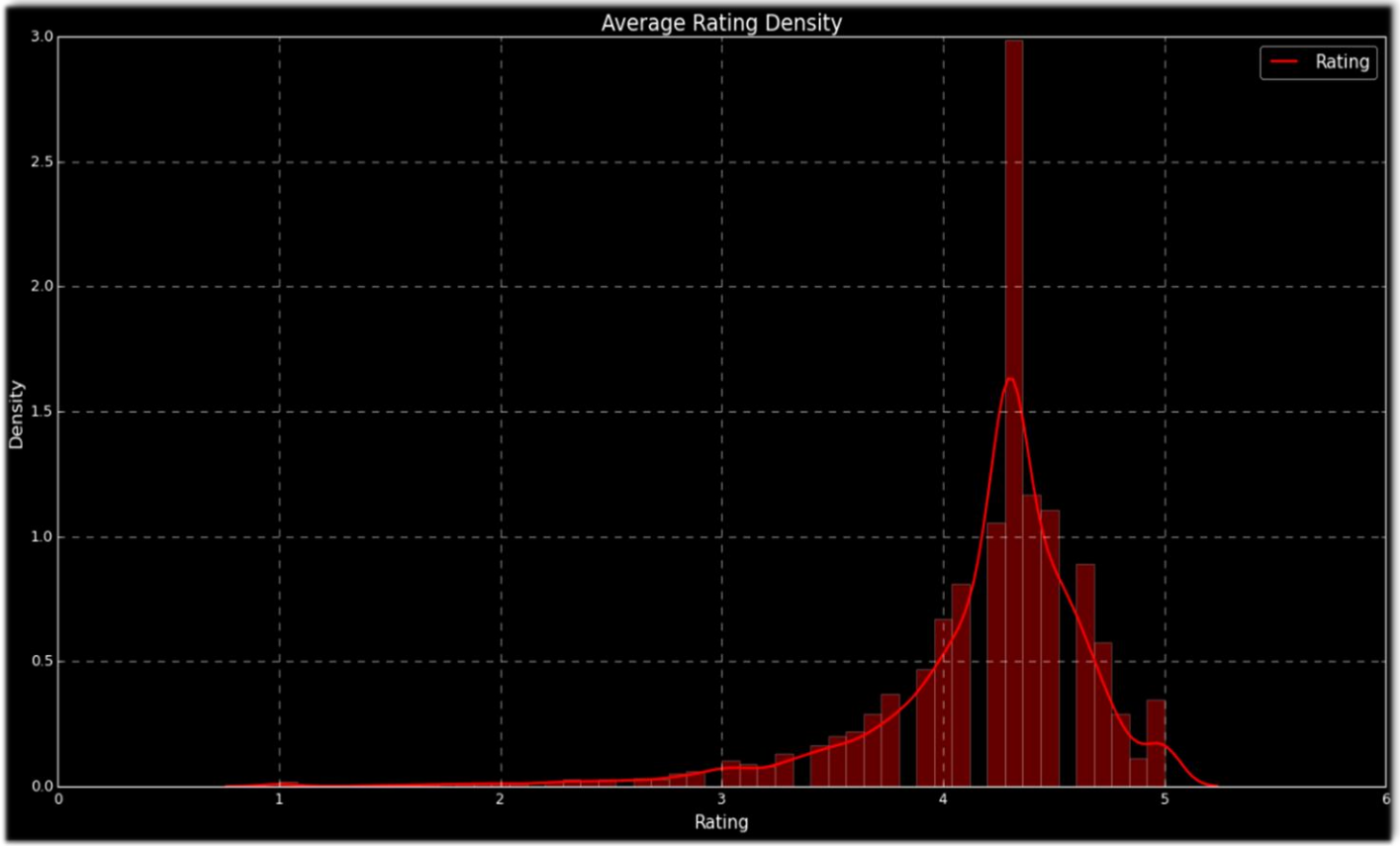


4. Review on the basis of category wise dataset.

Here by looking at this bar graph we conclude that most of apps which are built as social media or communication apps have highest reviews follwed by 'game' & 'family'.

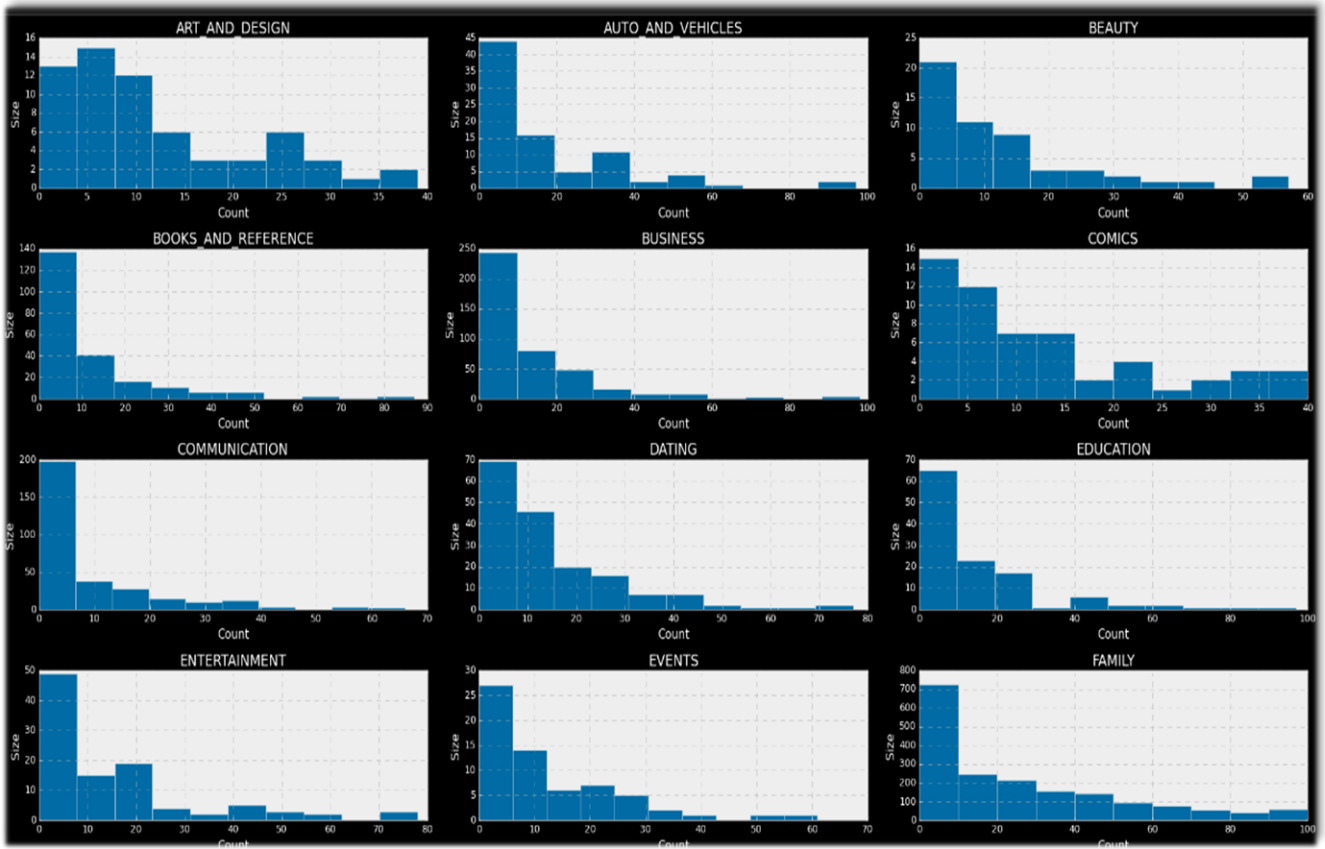


5. AVERAGE RATING OF APPS IN GOOGLE PLAY STORE



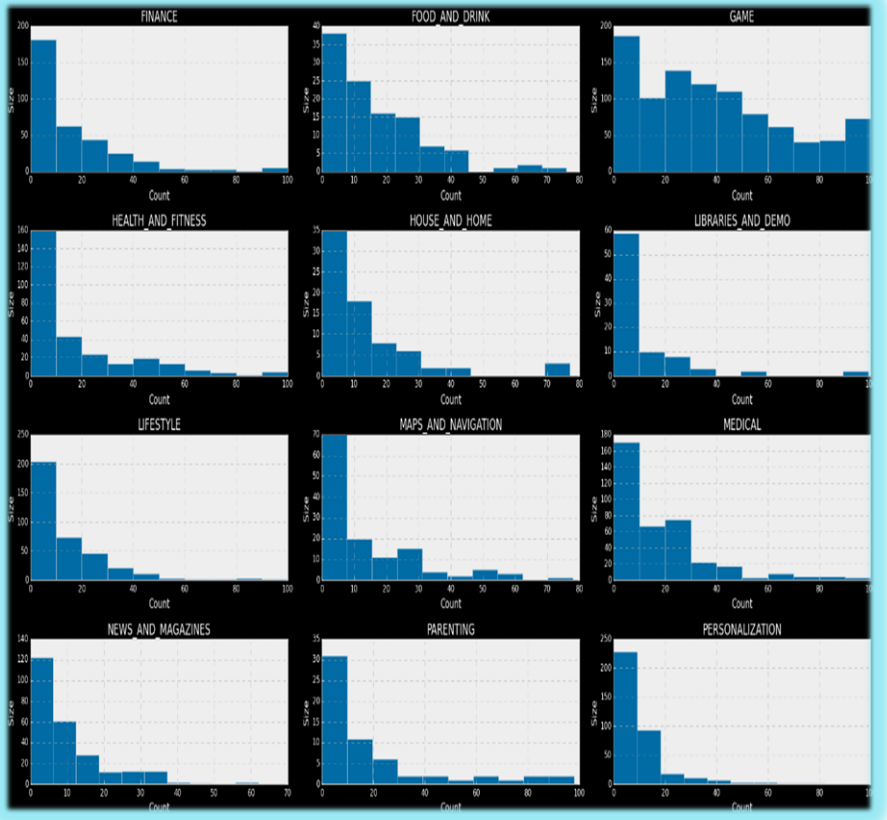
The average rating of apps in google play store is 4.25

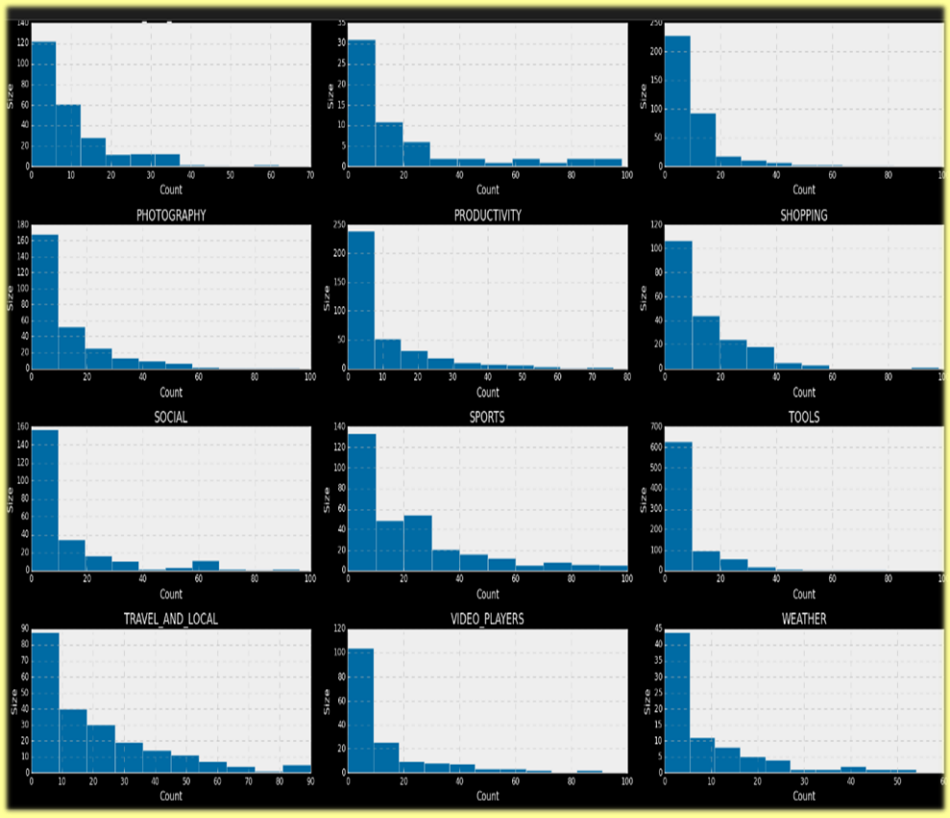
6. Different categories & their sizes.



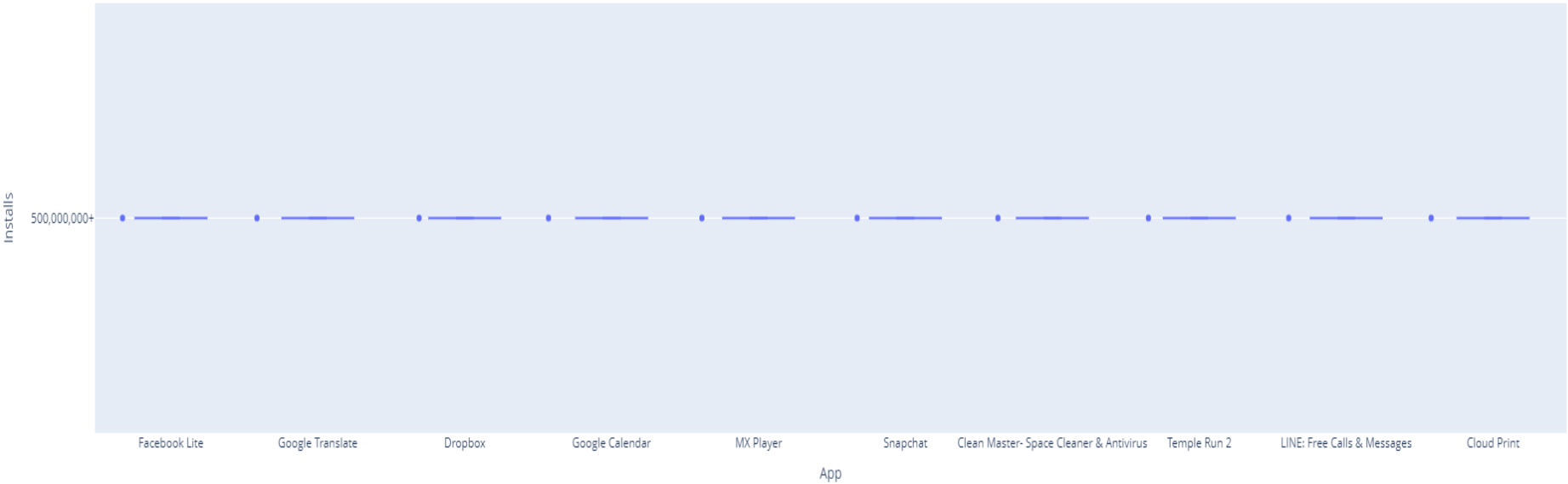
These visualizations tells that most of the apps are low to mid range in size as compared to gmes where the apps are slightly higher in size.

Remaining 2 visualizations based on the app size dataset.





7. Top ten apps installed highest no. of times.

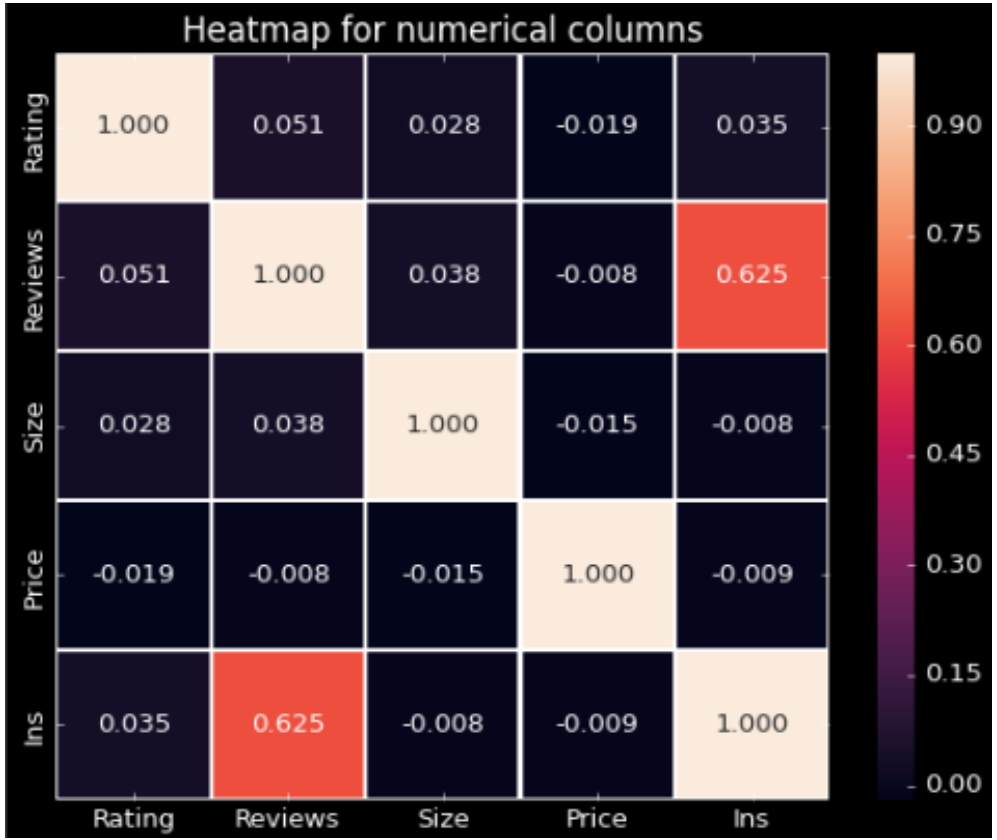


Alls the apps present in above voilin chart by plotly are downloaded more than 500 million times from google store.

8. HeatMap for numerical columns.

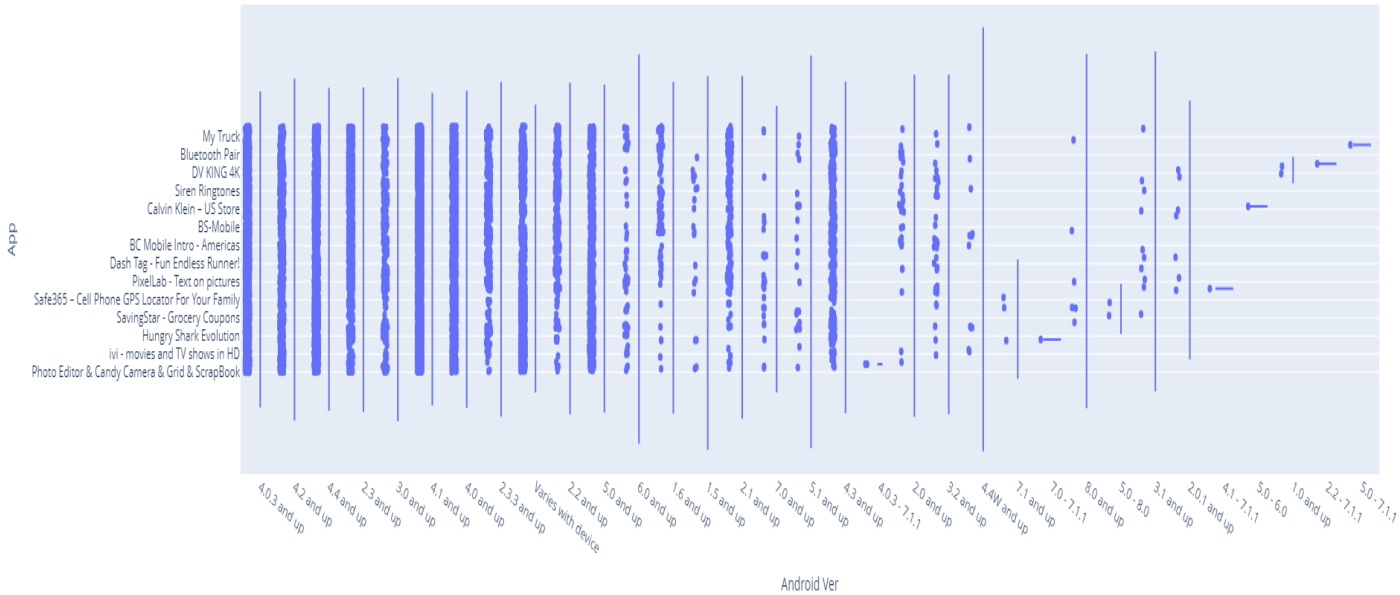
Above map on inspecting clearly shows the positive corelationship between reviews and install which is 0.625 i.e. as the the number of reviews increases, the number of installations also increase.

Similarly we can see corelations between other numeric columns too.



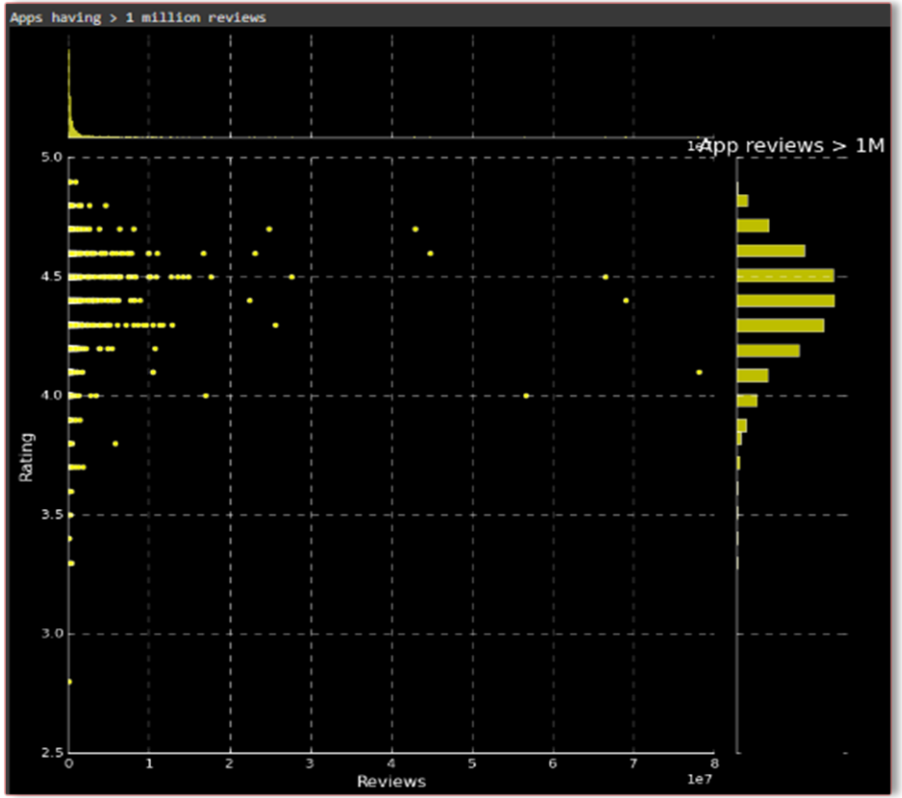
9. App vs Android Version

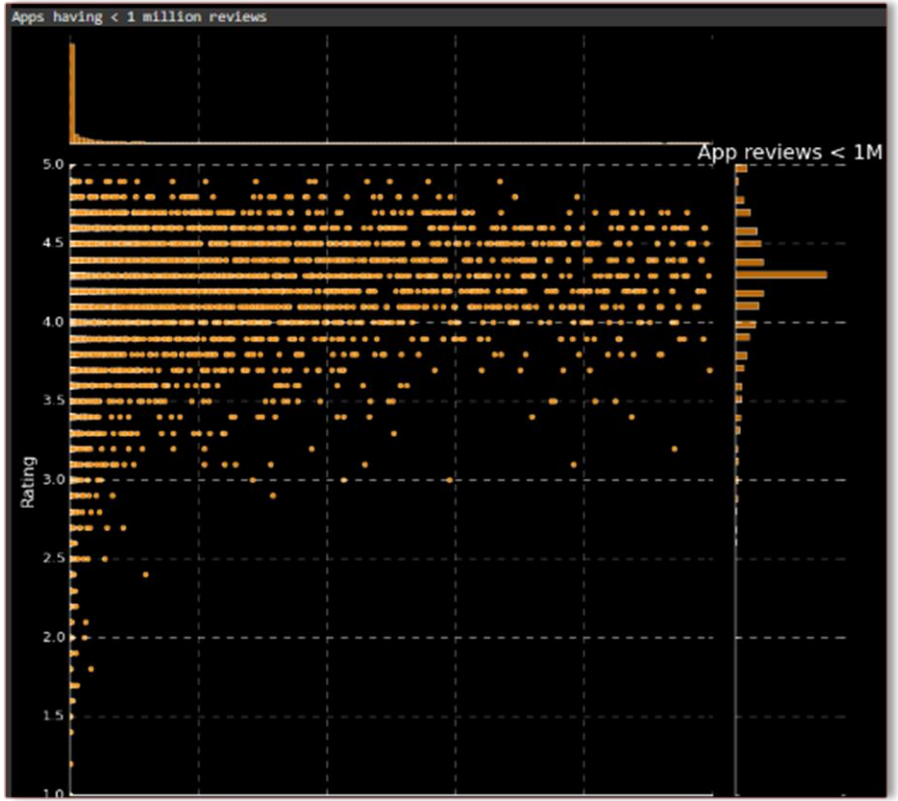
```
4.1 and up                2165
4.0.3 and up              1376
4.0 and up                1274
Varies with device         986
4.4 and up                 817
2.3 and up                 614
5.0 and up                 506
4.2 and up                 369
2.3.3 and up               271
2.2 and up                 237
3.0 and up                 231
4.3 and up                 218
2.1 and up                 133
1.6 and up                 116
6.0 and up                  53
7.0 and up                  42
3.2 and up                  36
2.0 and up                  32
5.1 and up                  22
1.5 and up                  20
4.4W and up                 11
3.1 and up                  10
2.0.1 and up                 7
8.0 and up                   6
7.1 and up                   3
4.0.3 - 7.1.1                2
5.0 - 8.0                    2
1.0 and up                   2
7.0 - 7.1.1                  1
4.1 - 7.1.1                  1
5.0 - 6.0                    1
2.2 - 7.1.1                  1
5.0 - 7.1.1                  1
Name: Android Ver, dtype: int64
```



The above visualization shows us that which app from google play store is running on which version of android OS.

10. Relationship between Rating & Review





The above observation shows that the the most reviewed apps are likely to have better rating.

INFERENCES & CONCLUSION -

After analyzing and exploring the data we got to know that this dataset gives us so many intresting insights and useful info about the play store app & thus it will deliver useful info to the customers as well as will direct the developers to get new apps in market and to popularize the product.
Also while using the visualization libraries i got to know about them alot as each library has it's own strength & weaknesses which are as follows. -

MATPLOTLIB - It's great for distribution analysis but low - level interface.
SEABORN - It has simple & short code but it dosen't have wide varities as matplotlib.
PLOTLY - It's very interactive, has versatile graphics & high - level interface. Bascially much better and revised version of matplotlib & seaborn.

* Family & games app have the highest share ratio in play store app.
* From the dataset we got to know that over 92.1% app in play store are free & 7.86% are paid.
* Most of apps which are built as social media or communication apps have highest reviews follwed by 'game' & 'family'
* Top 5 categories in Google play store are - family, games, tools, business, medical.
* The average rating in playstore is 4.25 & there are over 265 apps in the play store whose rating is 5.0
* Most of the apps are low to mid range in size as compared to games where apps are slightly higher in size.
* There is high positive corelationship between reviews and install which is 0.065.
* The most reviewed apps are likely to be better rated in play store app.
* Over 2165 apps are working/running over android version 4.1 and up in play store.

THANK YOU