Capstone Project-1

Team 1 : Play Store App review analysis

Team Members

Abdul Rahman Talha

Mohd Danish

What is App engagement?

App engagement is defined by a set of metrics that measure user interaction with an app. Common engagement metrics can include how many times a user opens the app, the duration of a usage session, and what actions a user takes within an app.

Basically App engagement and success rate of any app is based on different factors/variables such as No. of downloads, Ratings, Size, Ease of availability (free/paid), reviews and category. The idea of this project is to analyze these factors and take out those obligatory variables that makes an app successful.

Data pipeline

- **Data processing-1:** In this part, we've removed unnecessary features that would not be used in our analysis.
- **Data processing-2:** In this part, most of the variables are object data type which is difficult to analyze so, we converted important features into float.
- **Data processing-3:** In this part, we deleted/replaced null values with randomly chosen values from the same distribution of the data.
- **EDA:**
  - **Univariate Analysis:** In this part, we analyzed individual variable .
  - **Bivariate Analysis:** In this part, we analyzed two or more variable at a time and their relation between them.

Data Summary

rating: Numbers between 1 to 5, which is given by the user who usually give his experience or feedback that current app is fitted in his requirements if number is close to 5, counts in good rating and number close to 1 counts under bad rating.

reviews: How many reviews are coming on that particular app and which type of reviews are coming. Is it a positive review or negative review? This factor matters most from the point of view of success of any app. Positivity and negativity will be consider in user reviews data here, we only consider the numbers.

Data summary

installs: Number of installations of applications on play store data, these numbers may be in thousands, lacs and millions. Number of installations gives us a good lead and insights of the success of any app.

size: size of the applications of different categories available on play store data, these sizes may be in KB(kilobyte), MB(megabyte) and GB(gigabyte). In order to do our analysis size is also a huge factor to consider.

category: this column have the information of categories of different apps as which app belongs to which category like game, fashion, education, communication etc.

Data summary

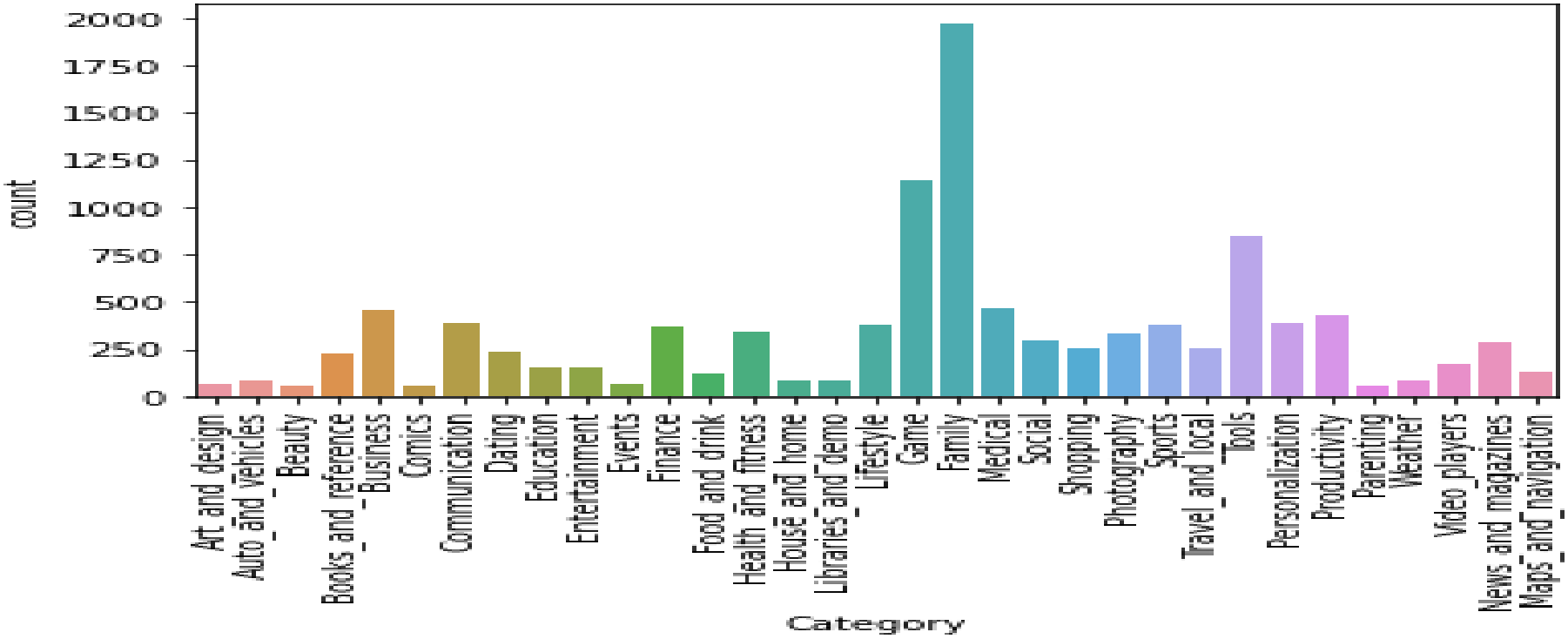
Price: In this column, we get the information of apps if they are free or paid on play store. Play store put money on apps because they think that that app worth that much of money if someone is using it.

Sentiment: A view of or attitude towards a scenario, event or anything this attitude may be positive, negative and neutral. So, this column completely tells the same about different apps means that this column have the information of users' reviews(emotion, opinion) concerning different apps on Play Store.

Data Summary

Sentiment_Polarity: It is the expression that determines the sentimental aspect of an opinion. The Sentiment polarity can be determined as float which lies in the range [-1,1] where 1 means positive statement and -1 means negative statement .
Sentiment_Subjectivity: The subjectivity is a measure of sentiment being objective to subjective and goes from 0 to 1. *subjectivity* is also a float that lies in the range [0,1]. When it is close to 0, it is more about facts and when it increases, it comes close to be an opinion. Subjectivity refer opinion because opinion is important key to client for business concern.

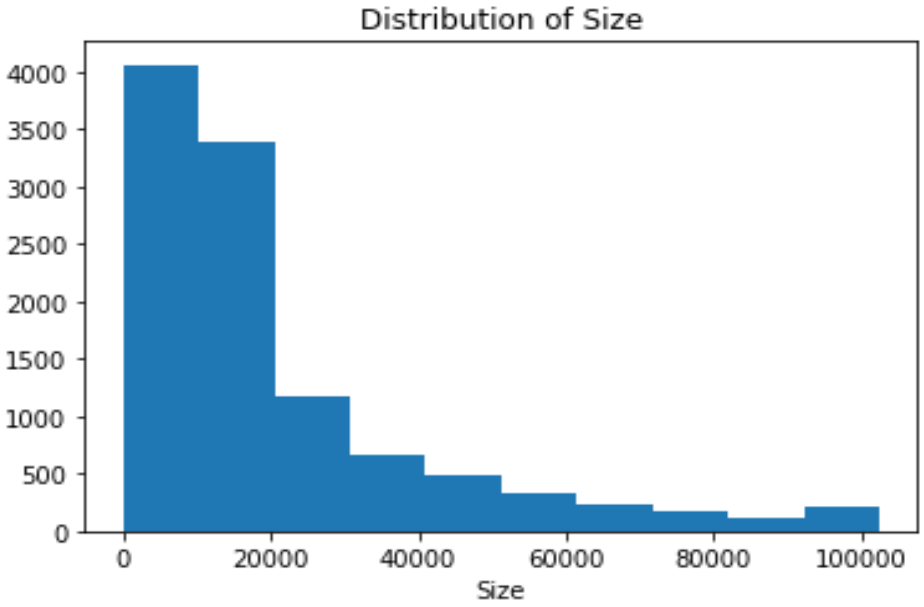
EDA: (Univariate Analysis)



- There are more apps of Category - 'Family’ than 'Game’ & 'Tools'. These are the three category apps which are more in the Play Store data and occupy more space in Play Store.

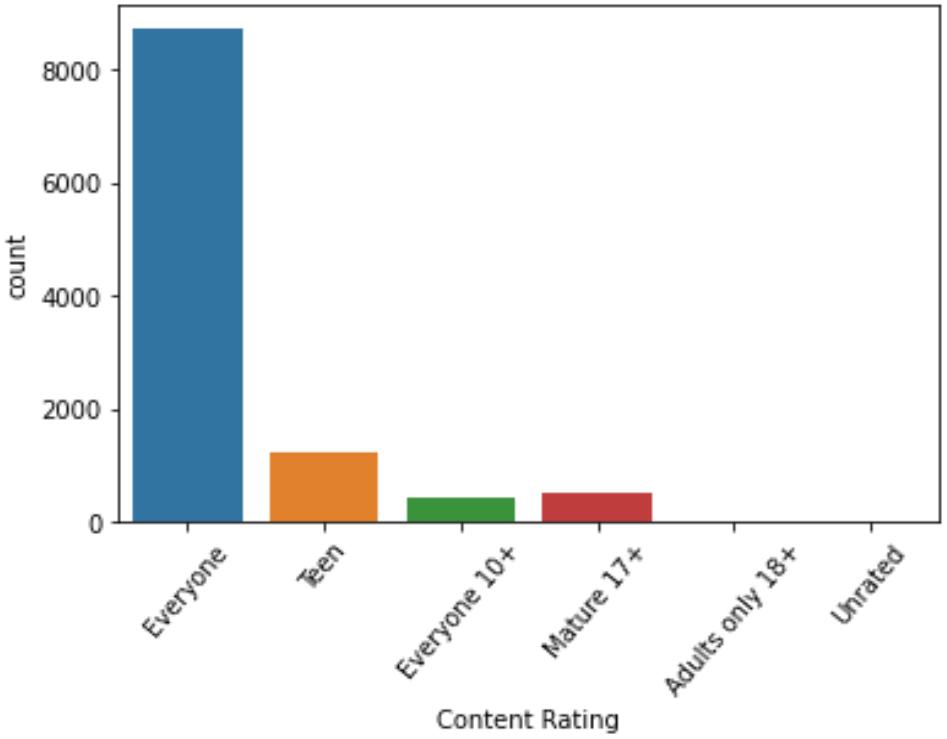
- Family and Game are two categories where population of children more persuaded than any other category. Therefore, clients are making more apps in these categories

Univariate Analysis(Continued)



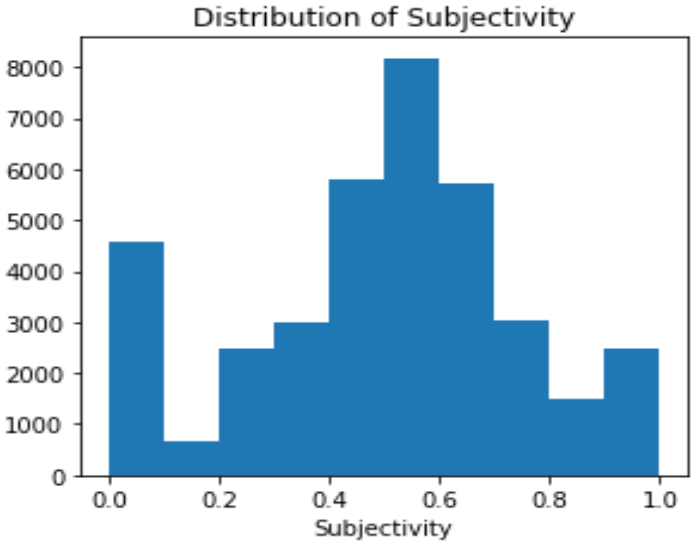
- **Most of the app size is in between 0-20000 KB(200MB).**
- **There are even apps with greater size than normal. They require up to 1,00,000 KB which takes more space of a system.**

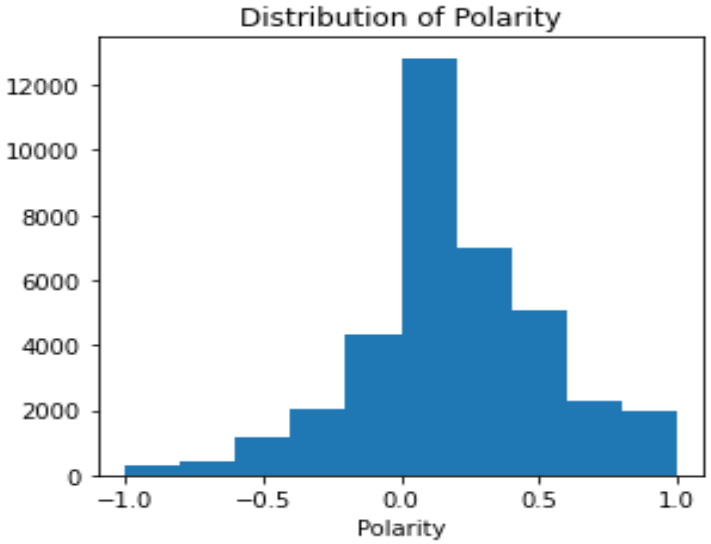
Univariate Analysis(Continued)



- **There are 6 categories in content rating from which most of the apps come under the apps which can be used or rated as used by 'Everyone'.**
- **Next category of content rating apps are used by only 'Teen'.**

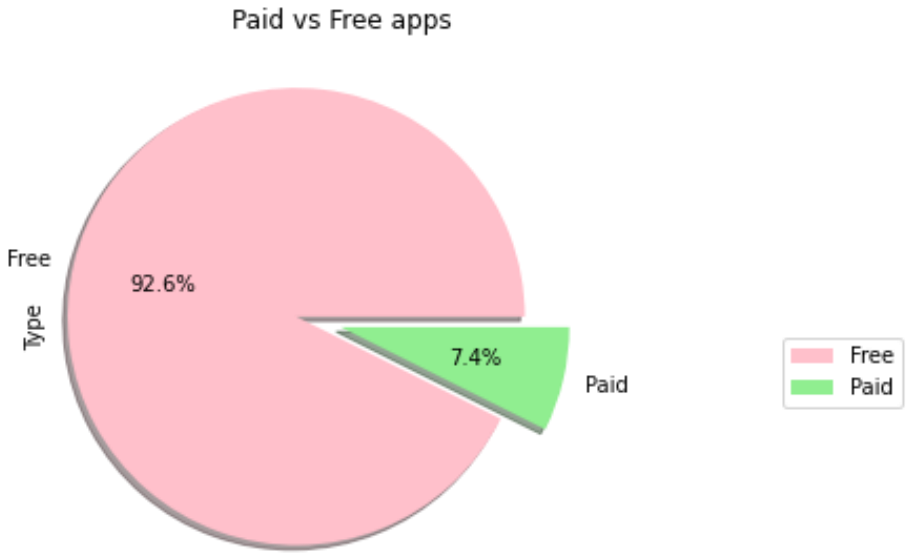
Univariate Analysis(Continued)





- **It can also be seen that there is peakedness in the middle of the data for both subjectivity and polarity.**
- **Most of the reviews are densified in between 0.3-0.7. So, the Subjectivity is increasing towards opinion.**
- **You can see a huge building nearby 0 that means most of the users are neutral they don't have a particular attitude towards app.**

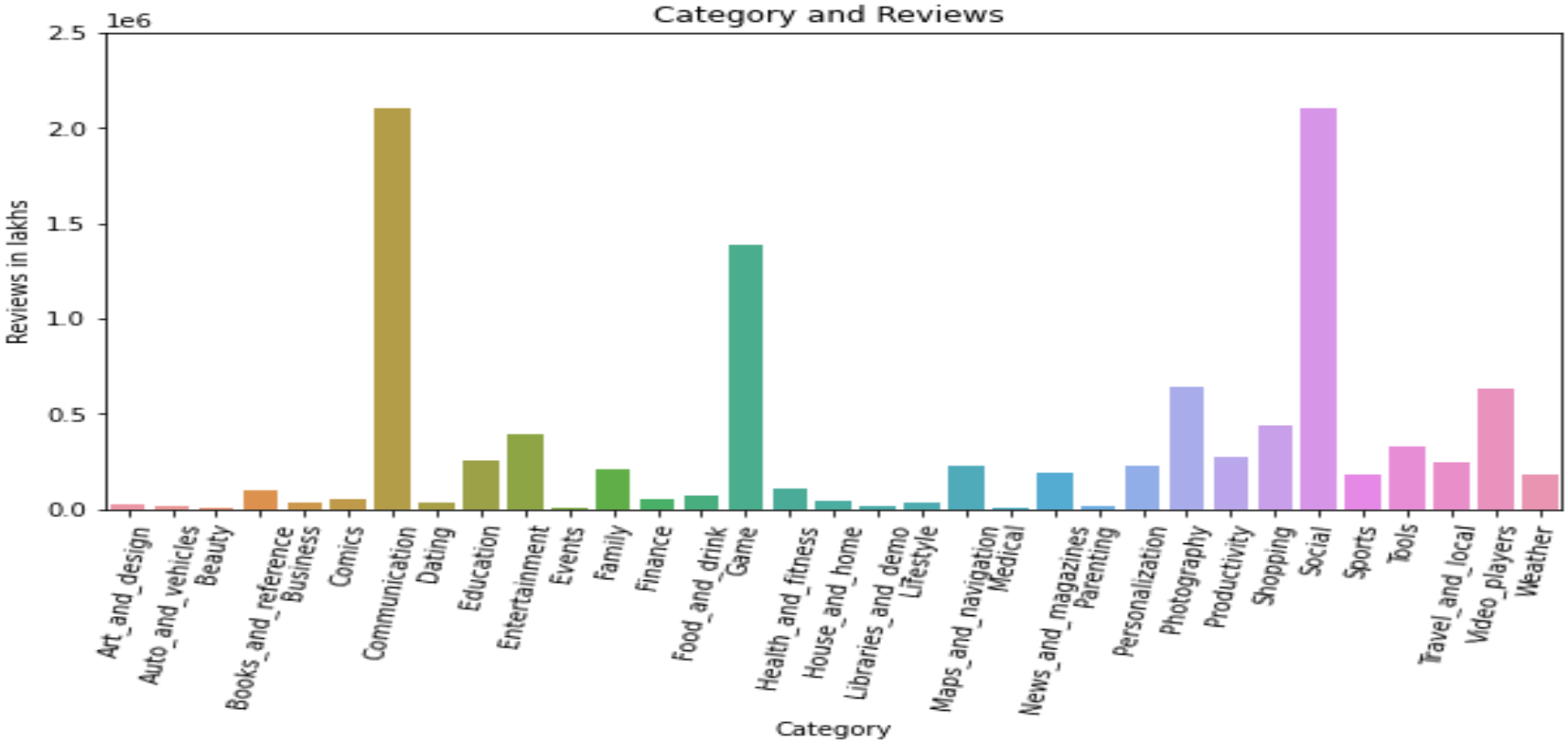
Univariate Analysis(Continued)



There are 92.6% of free apps in t he play store vs 7.4% of paid ap ps.

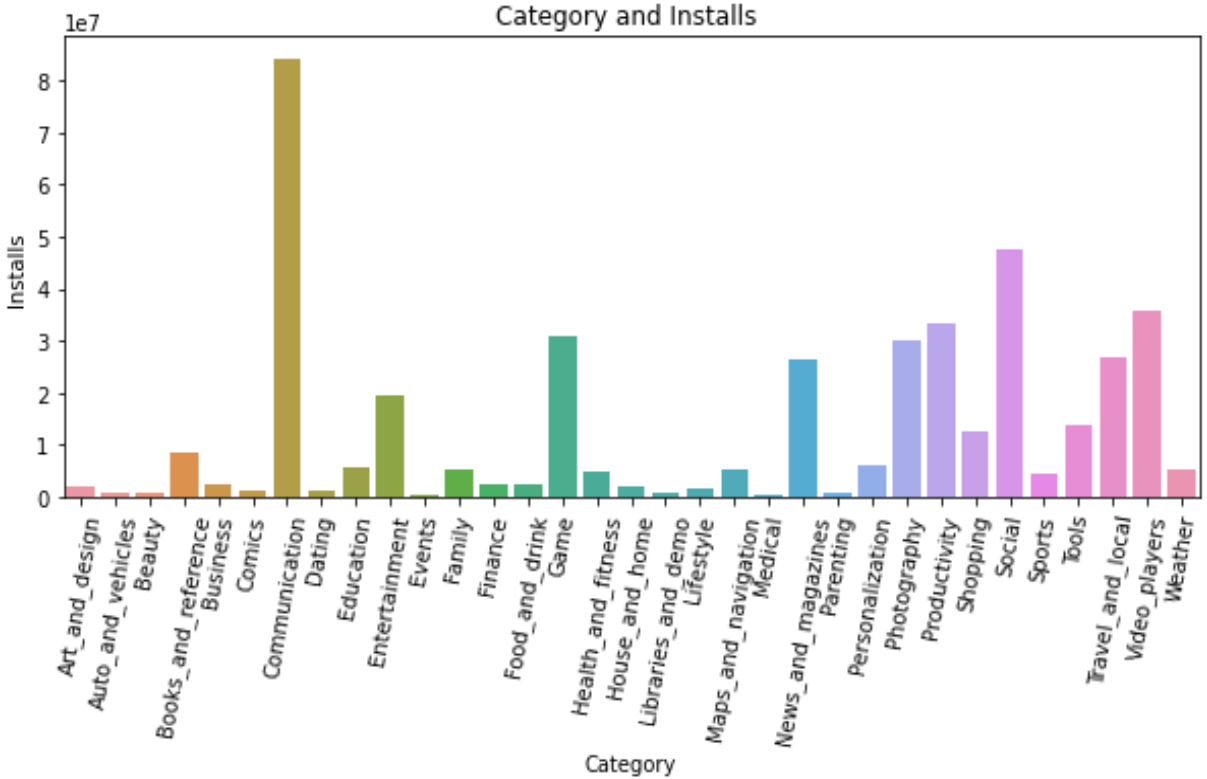
There are more free apps than p aid. So, there are more chances for the common people and ma ny people to install the apps eas ily and use it for their daily need s.

Bivariate Analysis



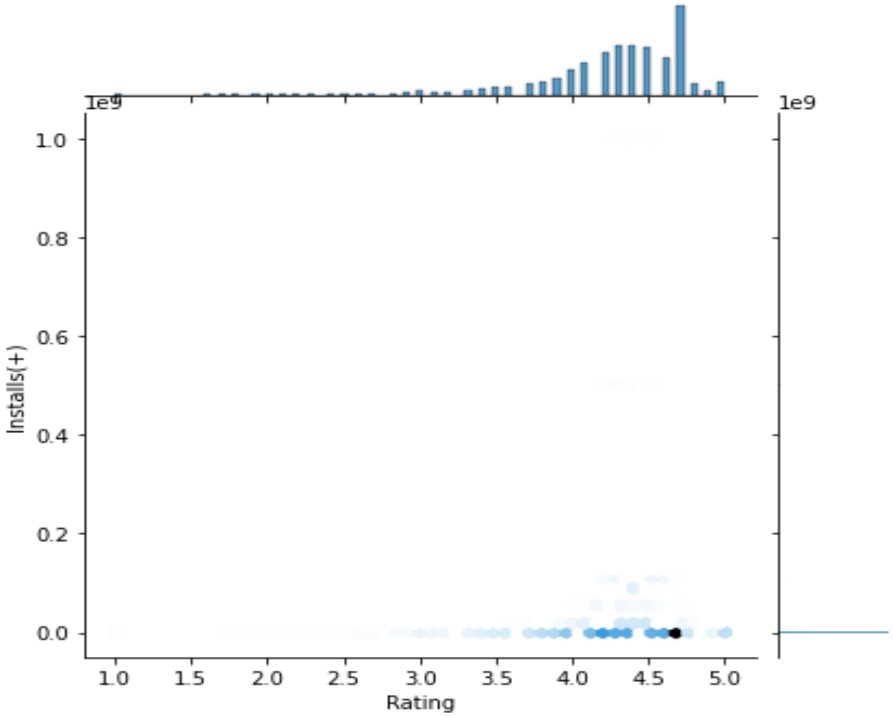
- It can be seen from the above plot, the average number of reviews in each category and also it can be seen that the reviews are highest for Communication, Social and Game categories.
- Yes the category and reviews seem a good match with the existing app use too. Definitely those are priority apps even now. All the users of android mobile do install the communication(chat apps), Social(google, fb, snap..), Games first then go for Photography, Entertainment, Education and so on as present in this data and same for reviews to the respective categories. What are your category apps? Already here for analysis.

Bivariate Analysis(Continued)



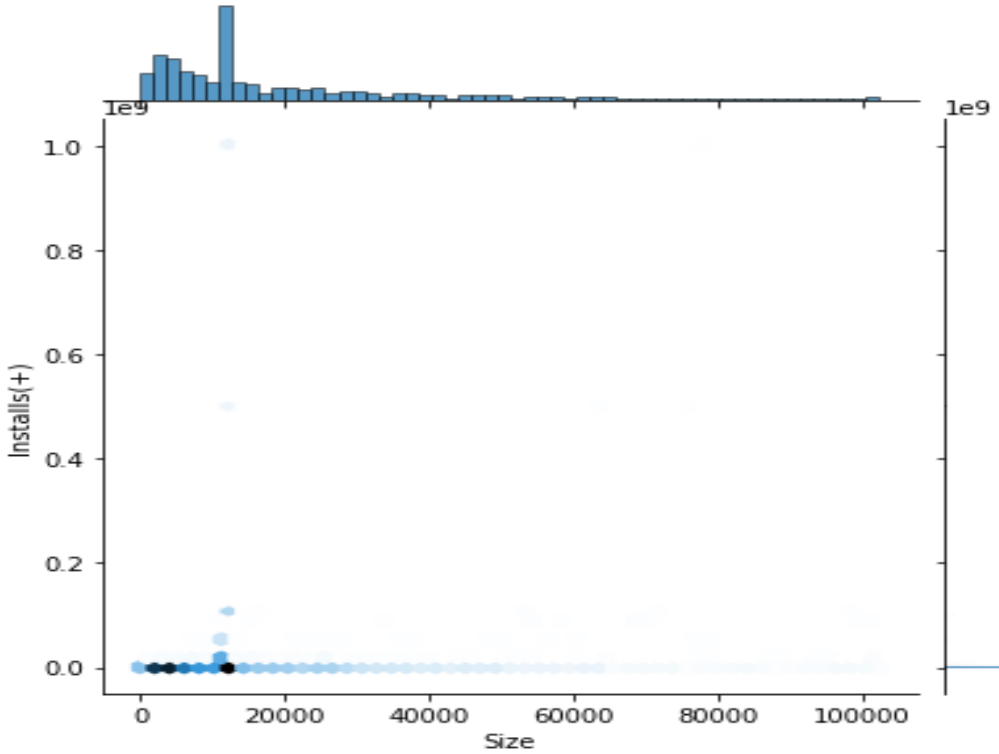
It can be seen that average number of installations are high for Communication apps then Social followed by Game, Video players, photography, News and magazines and, Travel etc.

Bivariate Analysis(Continued)



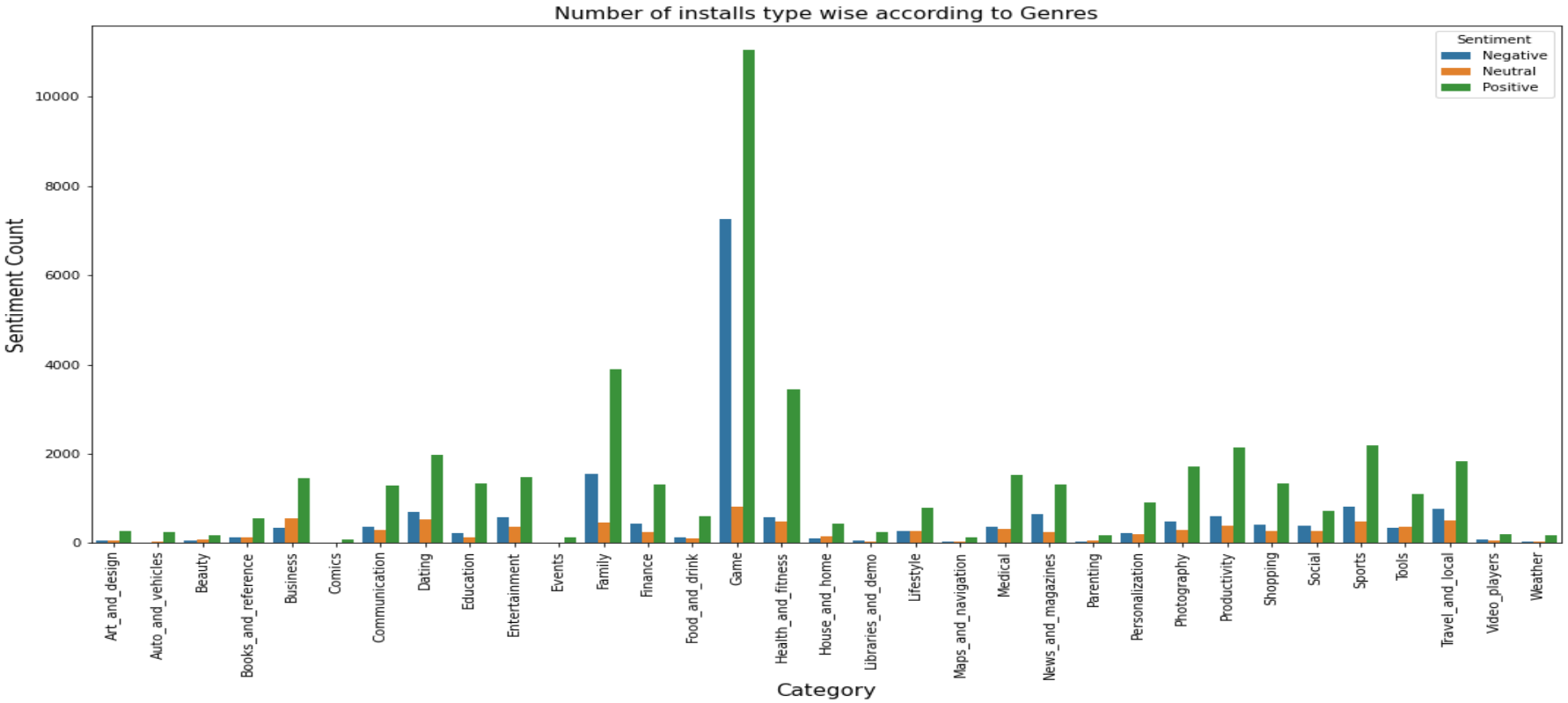
From the above it can be seen that **high rated apps have high number of installations increasing the success rate of an app**

Bivariate Analysis(continued)

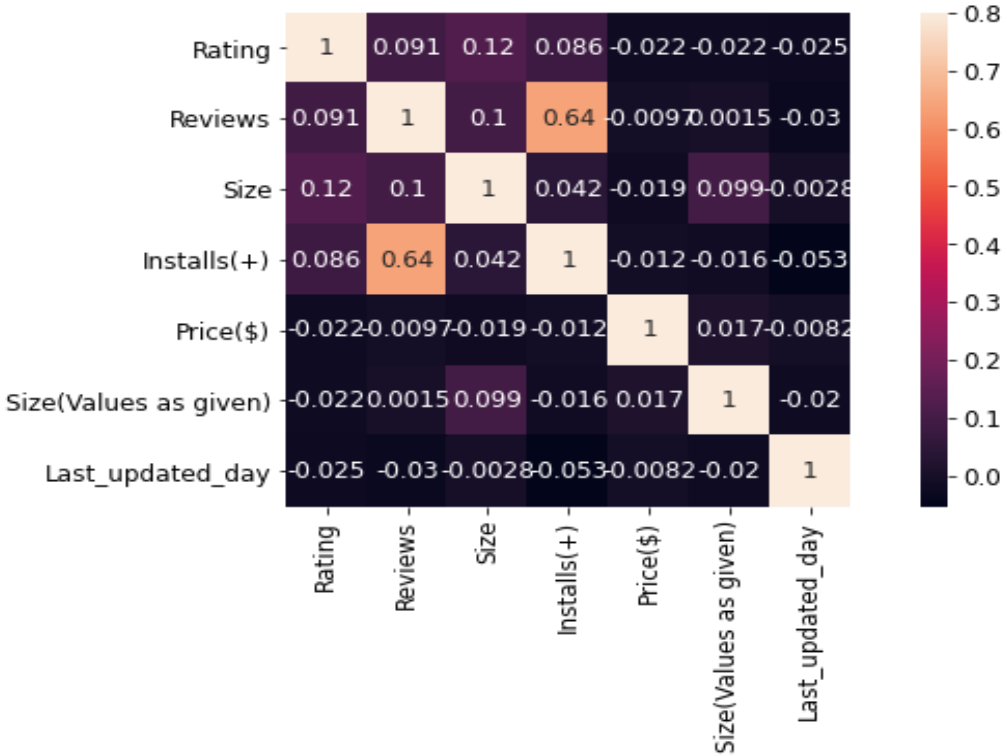


When it comes to size the apps with less size have more installations with some exceptions to it. Like there are more installations with size in between 10000-20000 KB than 0-10000 KB. **So this is one factor which has to be noted and taken care of for increasing the app engagement and success rate of the play store apps.**

Analysis based on sentiment



Correlation Matrix



The correlation between Reviews and Installs is higher and they both are strongly correlated variables. Hence multicollinearity exists between these two features.

Size and Reviews are next correlated variables with correlation of 0.1.

Conclusion and Results

- The features in Play store data that mostly helps in **predicting the success rate of an app are the Rating, Reviews, Installs and Type of an app.**
- From the analysis it can be seen that there are more number of apps with positive reviews than negative and neutral reviews.
- There are many categories of apps present in the play store **that are high in particular category is 'Communication' & 'Social apps'. It can be concluded that these apps are more successful and have high app engagement.**
- **On the basis of these visualization we've concluded that highest rating(4.0-4.5) and large number of reviews increases the success of any app.**

P.S. - Yes, it is obvious it can be seen very easily around us people are more engaged with Social Media Platforms.

Thank You