**Play Store App reviews analysis**

**Kritisha Panda, Data science trainee**

**Alma Better, Bangalore**

**Abstract**

App reviews and ratings are a crucial part of market analysis campaign of any app making business. It lets the business target the right audience and scale the sales. As per a report published by Apptentive, app reviews can influence 70% of people in installing the app on Apple devices. On the other hand, 75% read app reviews and downloaded them. Hence the importance of app reviews and ratings can’t be emphasized enough to help app-makers sky-rocket their business. For a app developer, the crucial point of interest must be the app revenue. App reviews directly affect customer retention and app revenue.

Throughout literature many analysis have been performed to study the effects of reviews and ratings on app installs. In this project, a humble attempt has been made to categorise the apps and study their relation to ratings, reviews and price. Also, the relation of app ratings and installs has also been briefly covered

**Introduction**

Reviews are crucial for a business as they help potential customers decide what to buy. For an app developer, reviews tell how people feel about the app and help them make decisions on installing the app. Not everyone amongst the user will be a tech-savvy. Research suggests that 47% of customers complaint about the app online.

Reviews and ratings on a category of apps, for example, financial apps, can give a developer idea on the current market trend and competitions before entering the market. Hence, he can customize and

personalize the app to be ahead in the competition. Data visualization plays an important role in understanding the market trends and make informed decision. It is a graphical representation of numerical data which helps to understand trends and patterns easily. Data visualization helps in understanding data better as it gives an easy way to represent collected data in form of graphs and charts and other visual elements so that people involved in decision-making process can make decisions easily. Whether these visualizations are dynamic or static, it is not of much importance to them. However, they can easily take the necessary steps to improve their performance when they know what is going on in their business or organization through these data visualizations.

These days consumers are getting more skeptical and aware of digital app marketing than ever. Other people suggestions in the form of app store ratings and reviews stand alone as the best and greatest driver of discovery and conversion. GDPR and other privacy regulations have moved to the forefront of marketing and product development, so trusted recommendation in the form of a rating or review is more important than ever.

**Exploratory Data Analysis**

Exploratory Data Analysis (EDA) is a method to analyse datasets and find preliminary insights, information, or uncover underlying patterns and trends in the data. Instead of making assumptions, data can be processed and analysed in a systematic way to gain insights and make informed decisions. Some advantages of EDA include:

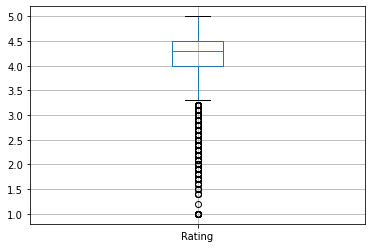
1. Improve understanding of variables by extracting averages, mean, maximum and minimum, values, etc.
2. Discover outliers, errors, and missing values in the data.
3. Identify patterns by visualizing data in graphs such as box plots histograms and scatter plots.

Hence, the main goal is to understand the data better and use tools so as to gain valuable insights or draw conclusions. Exploratory data analysis (EDA) is used by data scientists to scrutinize and investigate data sets and brief their main characteristics, often employing data visualization methods. It helps determine and decide how best to manipulate data sources to get the answers you need, making it easier than ever before for data scientists to discover patterns, anomalies, ,spot, test a hypothesis, or check assumptions. EDA is primarily used to see what data can convey beyond the formal modelling or hypothesis testing a task and provides a better understanding of data set variables and the relationships between them. It can also help decide if the statistical techniques one are considering for data analysis are accurate or not. Originally developed by American mathematician John Tukey in the 1970s, EDA techniques continue to be a widely used and accepted method in the data discovery process today.

**Data wrangling**

Data wrangling is the process of taking data in its rawest form and then manipulating it so that it can be used for further analysis. This process can be incredibly important for a number of reasons. Firstly, it ensures that the data is in a usable format. This is important because if the data is not in a usable format, then it cannot be properly analysed. Secondly, data wrangling can also help to improve the quality of the data. This is important because the better the quality of the data, the more accurate the results of the analysis will be. Thirdly, data wrangling can help to speed up the process of analysing data. This is because if the data is in a usable format and of good quality, then the analyst will be able to work with it more quickly and efficiently. Finally, data wrangling can also help to reduce the costs associated with data analysis as less time and money will be needed to cleanse and prepare the data for analysis.

Before beginning the EDA process, it is essential to have a overview of the data and check if our data set contains null values. After importing the necessary libraries, i.e, pandas, numpy, seaborn and matplotlib.pyplot we use the .head(), .info() and .describe() methods to gain insight into our data. Using the .isnull() method returns the number of null values in our dataset. We can either replace values by the mean, median or mode of the column or if 90% of the data in a row or column is null the particular row or column can be dropped.

The columns of interest to us in our dataset is mainly ‘Ratings’, ‘Reviews’ and ‘Price’. Number of installs has also been considered. Figure 1 shows the variations of ratings across our dataset

*Figure 1*

**Analysing the trend**

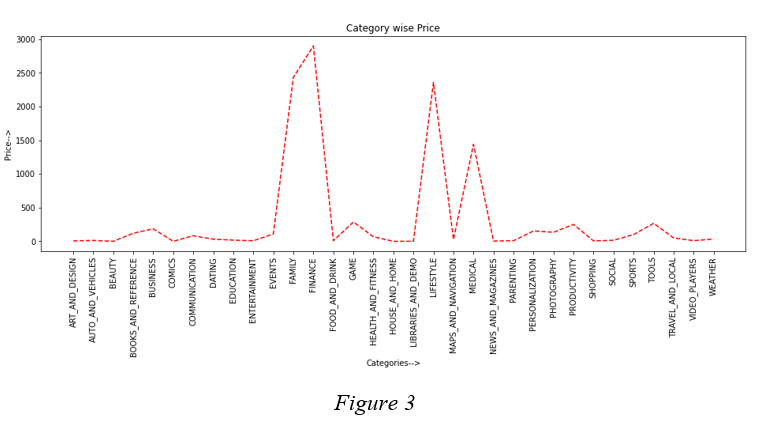
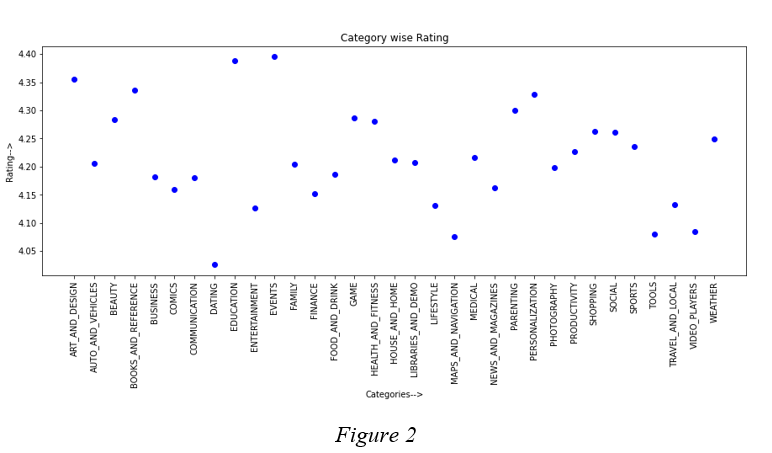
****The data is then grouped by category and compared individually with ‘Ratings’, ‘Reviews’ and ‘Price’ to see the trend. It gives us a visual representation of how the ratings, reviews and price vary. The groupby() method is used to achieve the task. Three variables a,b and c are assigned the values of columns ‘Ratings’, ‘Reviews’ and ‘Price’ from the dataset and graphs are plotted to see the trend with categories.

Figure 2 shows the variation of ‘Rating’ and with categories. It is evident from the plot that education has the highest number of ratings followed by events. The figure thus gives a clear insight on the ratings by people. Similarly, Figure 3 depicts the variation of ‘Price’ with categories. We can see that finance apps are most expensive followed by family.

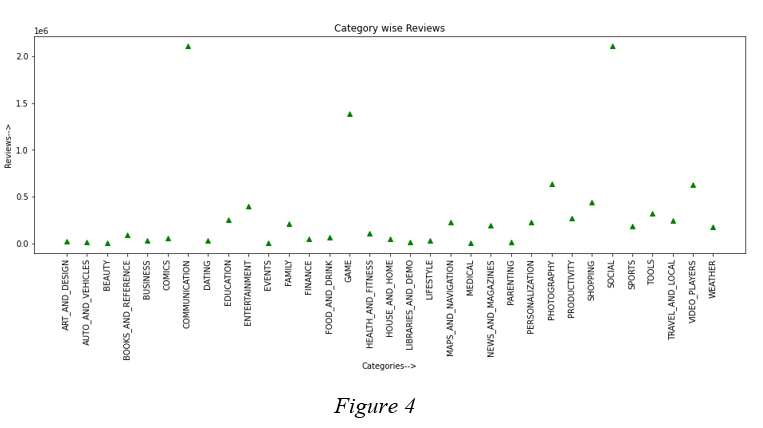
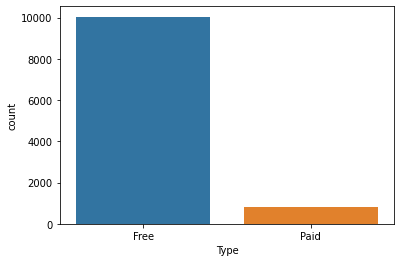
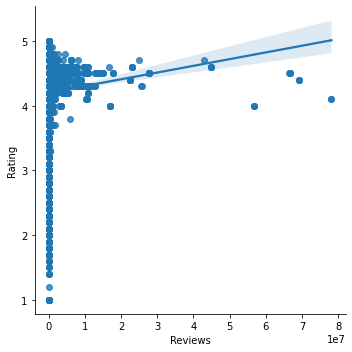


Figure 4 similarly depicts the variation of reviews with categories. The visual data thus represents that communication and social media apps have the highest number of reviews compared with other categories. Also, from figure 5 we note that paid apps are less preferred than free apps. There is a clear distinction between the number of installs of free and paid apps.



*Figure 5*

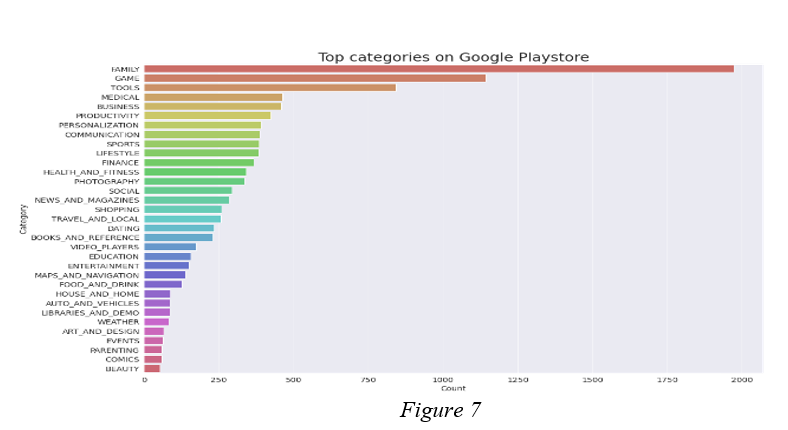
As per stats, free apps are faster to download than their paid counterparts. Also, not having a paywall makes it easier for the customers to try the app.

Users feel more comfortable to test the free version before trying the paid version. Once they start to trust the app, they can change to paid version. Thus, many app makers have both the free and paid version.

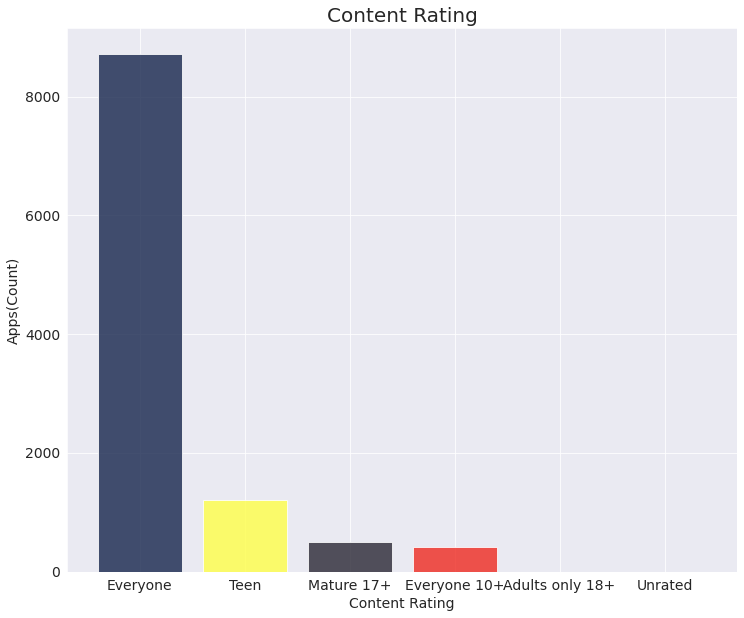
**Rating and Reviews**

Contrary to popular belief that more ratings means more installs, we see that the average rating for all number of installs is between 4 and 5. Also, there is no probable relation between ratings and reviews as can be noted from figure 6.

*Figure 6*

****

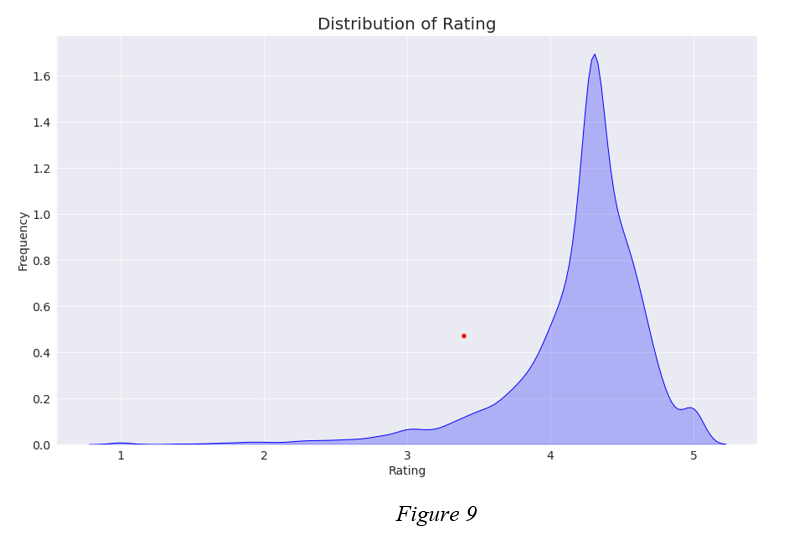
**Analysis of the obtained results**

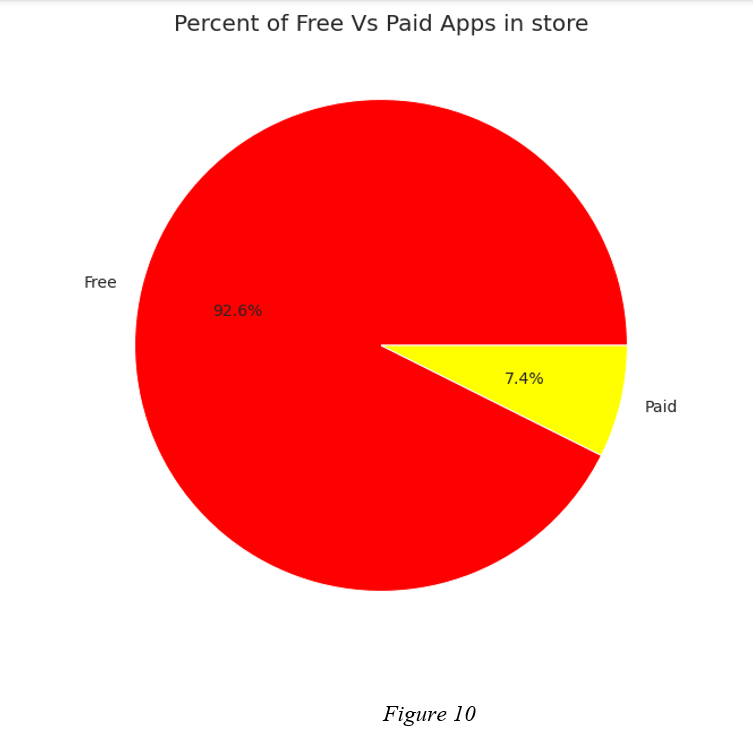
As per figure 7 we can se that Family and gaming apps are most popular in the category list. Searches for parenting/family-related apps, such as "allowance and chores app for kids" and "brushing teeth app for kids free," have grown 65% YoY. The first question in our analysis: Which is the top category to be installed is thus answered here. It is evident from the figure 7 that family and gaming apps are ahead of others. The parenting years can be a hectic time in life. And the challenges COVID-19 brought us made it especially difficult to have meaningful face-to-face interactions with many of our loved ones. While we can always make a phone call or see people via Zoom video chats, there’s something especially comforting about knowing there are family apps designed to help make our life easier. The next analysis was done for the apps which was ahead as per content rating. The results show that apps meant for all age groups grouped as ‘Everyone’ has the highest number of installs. General apps like gaming, health and fitness, nutrition, die etc which can be used by everyone has highest percentage of installs compared to

apps meant for specific ages. This is evident from figure 8. Next, we analyzed the distribution of apps as per ratings. It was seen that most of the apps had a 3.5 to 4.8 star ratings in playstore as evident from Figure 9. Because the analysis assumed categories across all age groups, such outcome is quite expected. Then we move on to the next important hurdle faced by app-making businesses: What percentage of apps are paid and free. It was noticed that free apps were the most installed ones, as

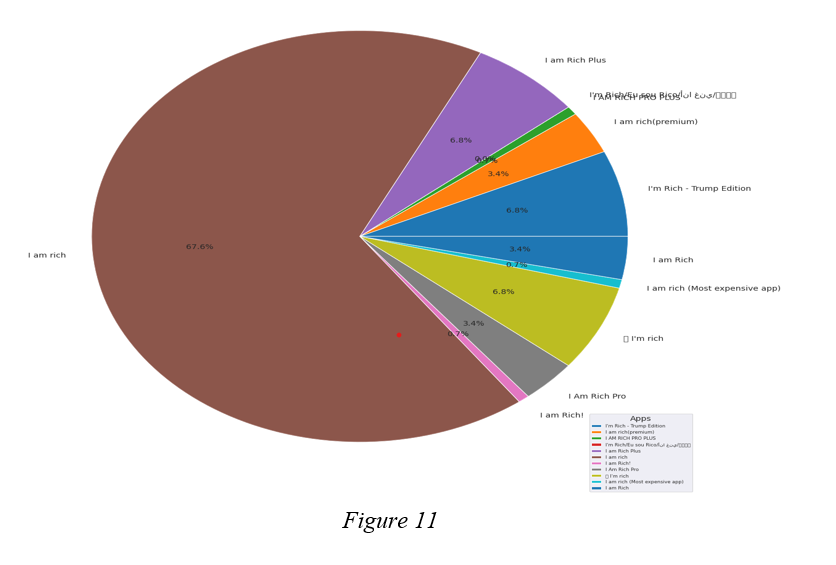
*Figure 8*

Expected [Figure 10]. Customers take time to trust the app hence want to try the free version first. Once they gain trust they may shift to the paid version.

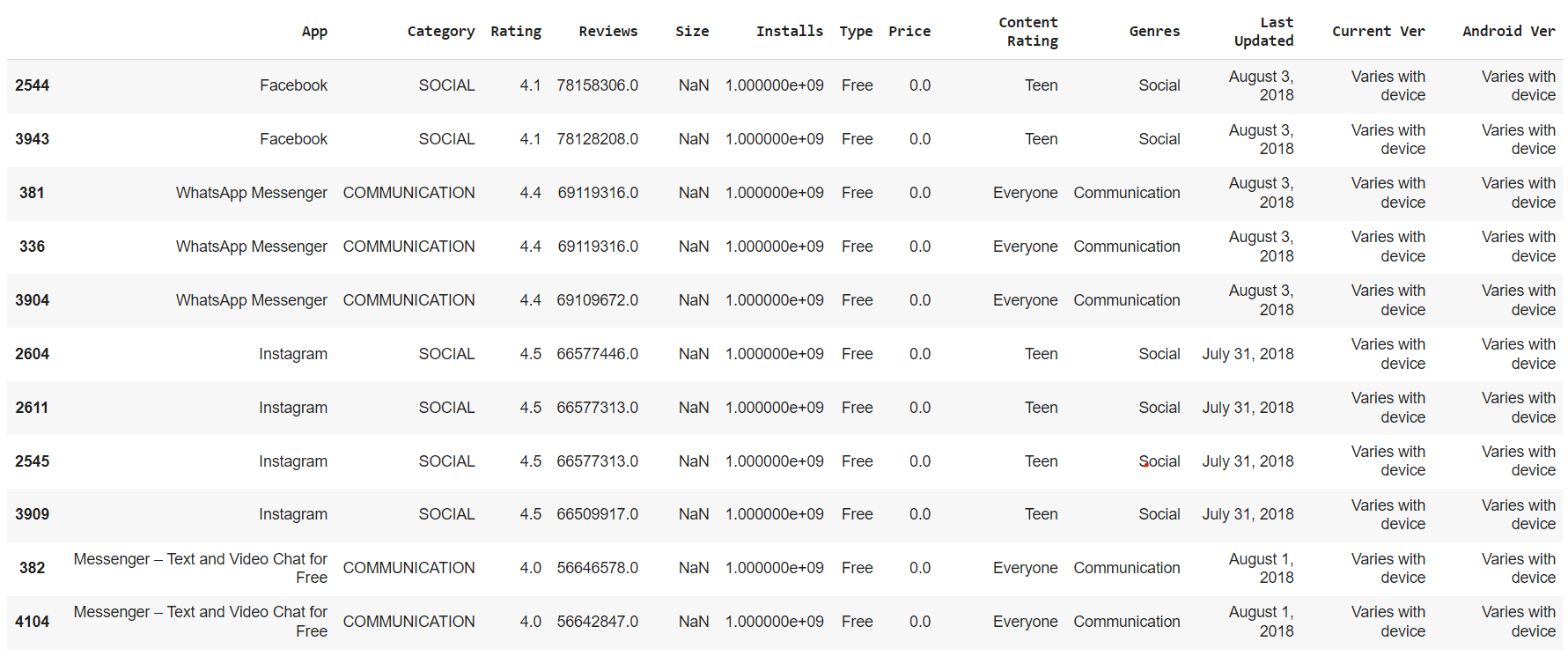




Next we take a look at categories of apps that were installed more. Depicted in Figure 11, it is clear that gaming apps were more installed followed by communication apps. Productivity apps lie far behind in the race. This may pertain to the fact that games belonging to ‘Everyone’ category can be easily installed which provides a change of mood for people and an entertaining rejuvenation. It'd be unfortunate if we don’t take a look at the top 10 paid apps in play store. This analysis provides an overview of the competitors for an app-making business. As shown in Figure 11, the top 10 paid apps are mostly financial applications. Its quite logical to have paid versions as people must be curious to know the tricks to multiply their finances at a small cost to be paid. Hence, paid versions of finance category apps can sustain in the market.



Last but not the least, we take a look at what are the apps with highest number of reviews. It is an essential parameter for an app-developer to know which category of apps get the highest reviews. It gives an idea of overall competition and quality of the market in that category. As shown in Figure 12, we can infer that communication-based apps are the ones having highest number of reviews. Today the entire world lives in internet as netizens. Hence, its not uncommon to have such an outcome where communication-based apps like the social media apps are having the highest reviews. Given a raw data set, its highly impossible to infer conclusions about different market scenarios. However, with data visualization life becomes easier. With a pictorial or graphical representation which answers our questions, inferring conclusions become a piece of cake walk. Also, it saves a lot of time and effort for the data scientist to infer conclusions from the graphs rather than the raw unprocessed data.

Hence, we see that a large number of conclusions can be inferred from the processed data just in a few moments. We represent the simplicity of data visualization and effectiveness of graphical plots over analysing Raw crude data.

**Conclusion**

Thus, we perform an in-depth analysis and see that family, finance and lifestyle apps are more expensive than others. Communication and social media apps have highest number of reviews. Education and events apps have highest number of ratings followed by art and design. Difference between installation of paid and free apps is huge. Free apps were installed in much higher number than the paid apps.Family, Games followed by Tolls are the top categories as per the analysis. Apps with content for age group ‘Everyone’ are ahead of others in the race. 92.6% apps are free whereas only 7.4 percent are paid. ‘I am rich’ app is the most installed app. This is in accordance with the observation that finance apps are one of the most installed ones.

*Figure 12*

**References**

1. Analytical Vidhya
2. Google
3. Coursera
4. Wikipedia