|  |
| --- |
| Photo displaying partial image of two pie charts on a canvas-textured page |
| Play Store App Review  Deriving valuable insights using Python |
| |  |  |  | | --- | --- | --- | | Md. Nawab Ali |  | Almabetter- EDA Capstone -1 | |

Play Store App Review

Mr. Md Nawab Ali, Ms. Jahnavi Jaolekar and Mr. Kaustubh M.Amare

Data science trainees, Alma better, Bangalore

# Abstract:

Play store is one of the largest and widely accepted App libraries. User can search for and download the most suitable app according to need for free, as arroun0d 81% of apps are free to install. We have used 2 raw data sets for exploratory data analysis, google play store dataset & User reviews data sets. This study was done to produce and observe meaningful insights and derive helpful conclusions. This study will help to observe the data with various visuals aiding to make the data crisp and easy to understand.

# Introduction:

Playstore is an app market for Android operating systems developed and regulated by Google, offering apps of wide variety and purposes, currently there are 3.48 million apps on playstore and the number is still growing by an impressive rate. Its home to app kinds of app and provide them majorly free of cost. It is trusted by billions around the world due it its security policies and surveys. Thus studying this large amount of data makes it important to understand trends and get other valuable Insights from a developers point of view.

Logos representing playstore are:

# Problem statement:

The Play Store apps data has enormous potential to drive app-making businesses to success. Actionable insights can be drawn for developers to work on and capture the Android market.

Each app (row) has values for catergory, rating, size, and more. Another dataset contains customer reviews of the android apps.

Explore and analyze the data to discover key factors responsible for app engagement and success.

# Methodology:

We collected the data of play store and user reviews from website of alma better. Our basic approach was to make a copy of original clean the data and make it ready for data analysis and data visualization. Steps will be explained in details below.

## Step 1:- Data obtaining

Google play store apps data frame and user review data frame were obtained from official alma better website. Data was then imported into running environment i.e., google collab in this case.

## Step 2:- Data overview

Necessary copies of data were made to save the parent data from any permanent modifications. General layout of data was studied using .info (), .describe () and .head() methods.

## Step 3:- Data cleaning

* Function “complete\_info()” was made to review the dataset again and again after processing to determine the null values ,duplicates and data types.
* Duplicates were determined and dropped from the data, using “.drop\_duplicates()” method.
* Some columns were processed and null values were replaced by mean or mode if applicable or else dropped in order to make the data complete.

## Step 4:- Data preparation

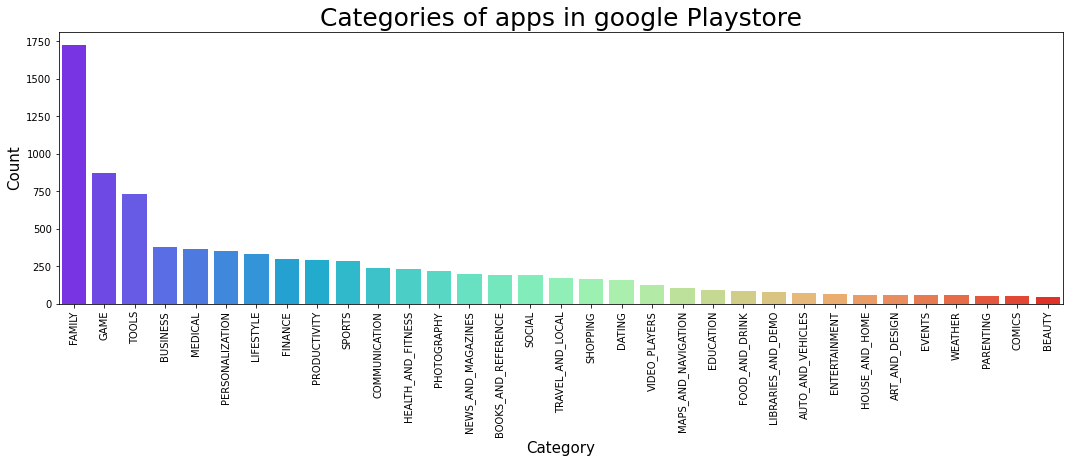
Converting some columns datatype in more suitable datatypes, appropriate for data analysis i.e.

* Size: - It was in from of object (M, K, varies with device, etc.) with this units thus function was made to convert them all into simple numerical string and then column was converted to float using “.to\_numeric()” method. KB is converted to MB using formula mb= kb/1024 in function.
* Installs: - It was in the form of object e.g., 500,000,000+, which was converted to number after replacing ‘+’ and ‘comma’ with suitable characters, to get 500000000 in integer format.
* Price: - It was in format of object from with unit “$” thus converted to basic numerical value by removing “$” and converting it to float.
* Reviews: - This column was clean thus directly converted it to integer using “.dtype(int)” function.

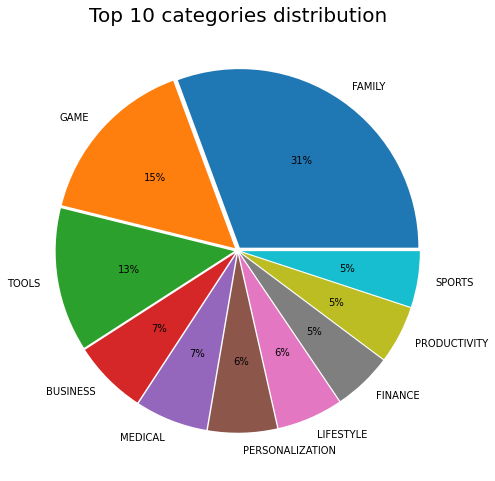
## Step 5:- Data wrangling & data visualization

Data was explored based on categories, in total 33 unique categories were found in the data.

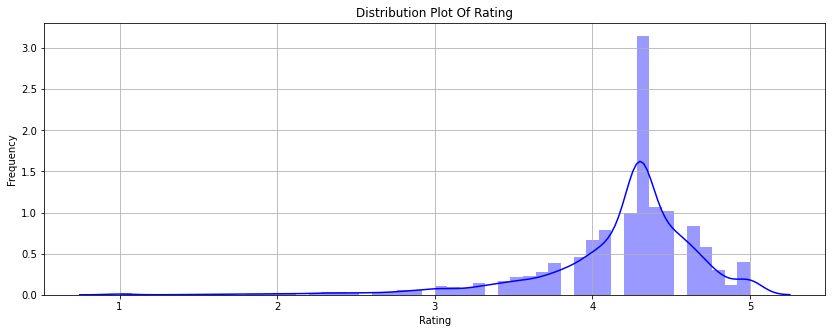
1. Graph plot for No. of Apps in each category.



1. From these 33 categories top 10 categories were shortlisted depending on their population distribution and a pie chart is plotted.



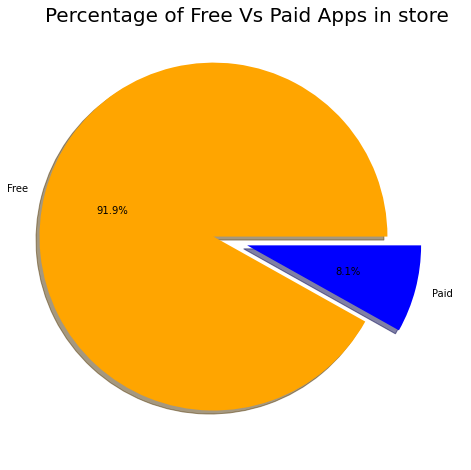
1. Cumulative average rating of all apps in play store is 4.18/5.0 determining that the maximum apps are useful and trustworthy.



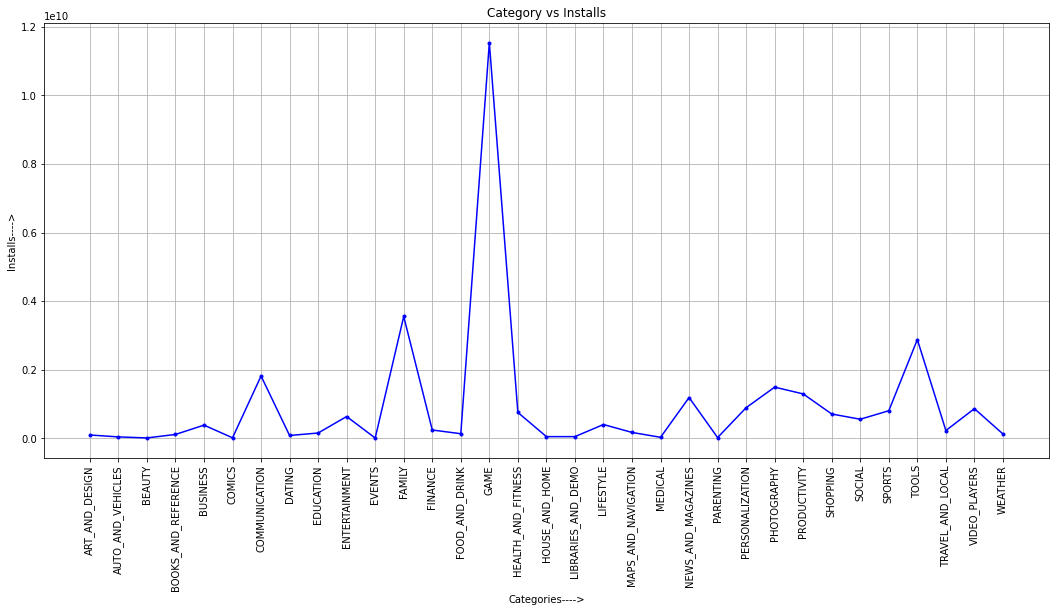
1. Apps were also categorized by content rating i.e., mainly Everyone, mature 17+ and Teen. After studying this parameter we found out that majority if the apps available in play store falls under ‘Everyone’ category thus can have vast audience from all age group.



1. Paid Apps vs free Apps

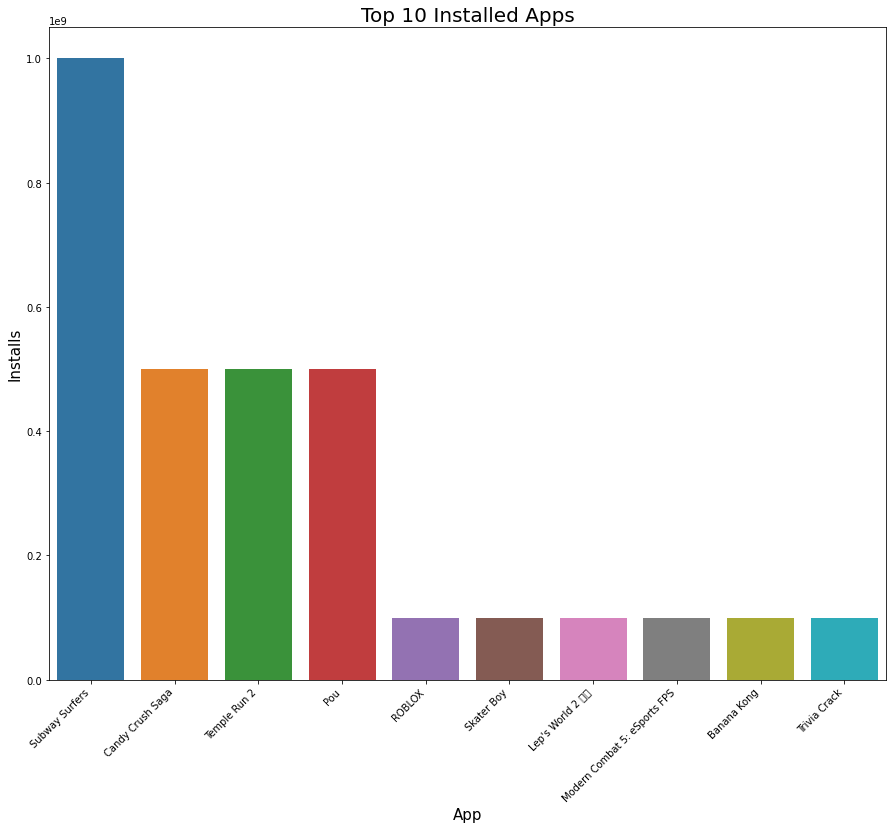
Majority of apps on play store is free showing the why it is preferred on other stores, like windows store and app store.

1. Categories of app installed highly by user population.



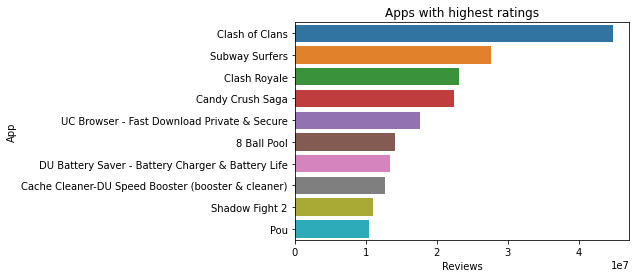
As we can observe in above graph, apps from category of ‘Game’ is installed very highly by the masses.

1. Top 10 installs named “top10incategory\_installs”
   1. A function is made to give top 10 based on their category. It was used for the games

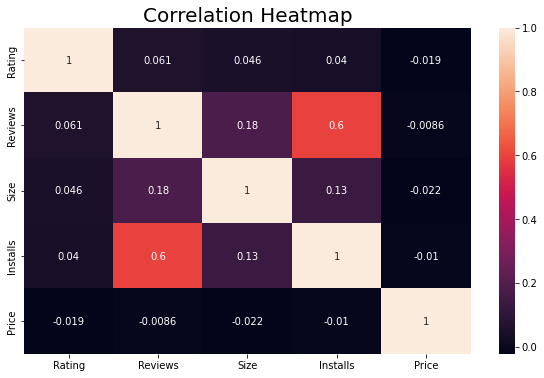


Thus, it can be observed that subway surfers candy crush and temple run 2 are on trending in play store and are installed by mass population.

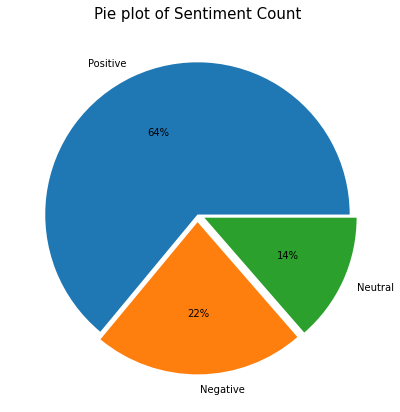
1. Highly reviewed apps on play store are mostly from “Game” category as observed

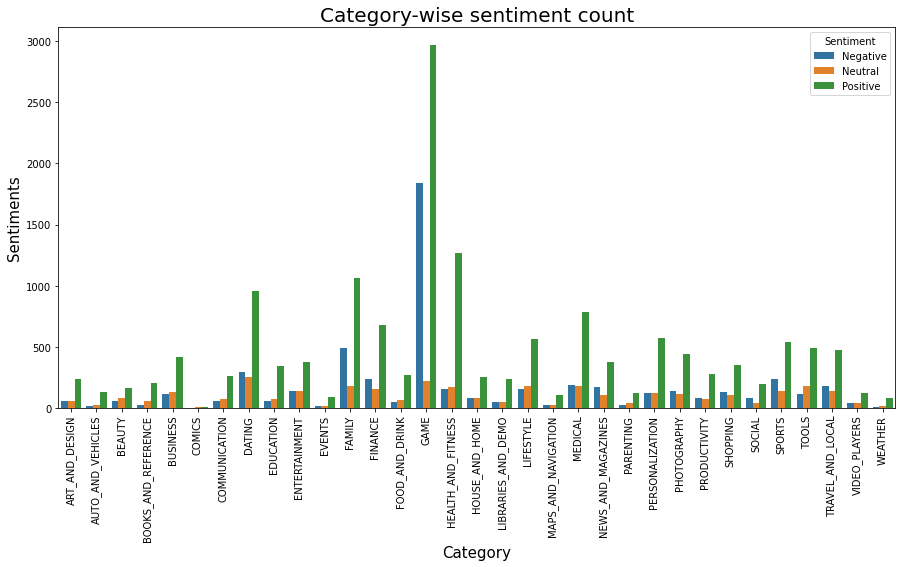


1. Correlation of features of play store data set.
   1. Installs and Reviews are positively correlated.
   2. Size might have some effect on Installation capacity of user.
   3. Rest of the parameters are not affecting each other directly



1. Sentiment analysis.
   1. Two data frames i.e. playstore apps data frame and User\_review data frame is merged to get the sentiments of users regarding the app. Sentiments are distinguished as ‘Positive’, ‘Negative’ and ‘Neutral’
   2. Sentiments of overall app database of play store is POSITIVE.



* 1. Sentiments of user for each category:- The graph shows that people are highly sentimental about ‘Gaming’ category, followed by ‘house and home’ and ‘Dating’ category showing majority of positive sentiments.

# Conclusion

* Most of the app's belongs to Family category. They are approximately 19% among all categories.
* Ratings are negatively skewed with mean rating of 4.12
* Sizes are positively skewed with median size of 12 MB.
* Prices are positively skewed with mean price of 1,median price of 0 and max price of 400.
* Most of The Applications Are Created For Everyone
* Approximately 92% apps are free.
* Tools, Entertainment and Education are the top 3 genres.
* Approximately 64% sentiments are positive, 22% are negative and 14% are neutral.
* There are 266 apps which have 5-star rating. In which 67 apps belong to the Family category.
* Among Top 50 install apps 22% belongs to the communication category and 16% belongs to the Game category.
* Events category have the maximum mean rating of 4.395313
* Genres of Communication, Tools, Productivity and Social have the maximum no. of installs.
* Apps belonging to Finance category have the highest mean price (8.408203) among all categories.
* Family category has the highest no. of paid apps.
* Game And Family Category Have The Highest No. Of Installs Of Paid Apps.
* Installs are positively correlated with reviews with correlation value of 0.63.

# References

1. [stackoverflow](https://stackoverflow.com/) - [https://stackoverflow.com/](%20https:/stackoverflow.com/)
2. [grepper](https://www.codegrepper.com/)- <https://www.codegrepper.com/>
3. [Geekforgeeks](http://www.geeksforgeeks.org/) - <http://www.geeksforgeeks.org/>