Capstone Project - I

EDA – Play Store App Review

Prepared by – **Sandeep K**

Cohort - Montreal

Abstract - Since most Play Store applications are free, the income model is very obscure and inaccessible regarding how the in-application buys, adverts and memberships add to the achievement of an application. In this way, an application's prosperity is normally dictated by the quantity of installation of the application and the client appraisals that it has gotten over its lifetime instead of the income is created. Application (App) ratings are feedback provided voluntarily by users and function important evaluation criteria for apps. However, these ratings can often be biased due to insufficient or missing votes. Additionally, significant differences are observed between numeric ratings and user reviews. This Study aims to predict the ratings of Google Play Store apps using machine learning Algorithms. I have tried to perform Data Analysis and prediction into the Google Play store application dataset that I have collected from Kaggle. Using Machine Learning Algorithms, I have tried to discover the relationships among various attributes present in my dataset such as which application is free or paid, about the user reviews, rating of the application.

1. PROBLEM STATEMENT

I had received a mandate from ABC Co. Ltd. That, company wants to build a successful Mobile App, which should be accepted by a huge customer base of Play Store and create a positive Brand Image among customers. It should be one of the Successful App on Play Store.

DELIVERABLES :

- Analyze and identify key factors responsible for app engagement and success.
- Submit recommendations to select and plan development of a Successful App
- Suggest Actionable Insights to Developers to build a competitive App in best possible manner

MY APPROACH :

- Import data & important Libraries, Clean the data, perform Data wrangling and prepare copy datasets (Free App data, paid App data, Merged data etc)
- Understand data features and explore important data variables
- Analyze and understand Business Models for App making Business
- Identify Top 10 Performing : Genres | Categories | Apps
- Identify Top 3 Content rating categories and their contribution in App Success
- Identify most preferred factors like Size, Android Version, Categories
- Analyze important factors responsible for App engagement and success
- Correlation Analysis of all Quantitative Variables & their impact (Rating, Reviews, Updates, Size, Price, Installs, Android Version, Current Version, Sentiment Polarity, Sentiment, Sentiment Subjectivity)
- Used Word Cloud to demonstrate, how it can be used to manage sentiments of people
- Conclusion – Qualitative Recommendations to Management Team and actionable technical insights to App Development Team.

2. INTRODUCTION

- Mobile App Market is set to **GROW** 20% by 2023
- Android Apps comprise 90% of the Mobile App Market
- Android is the most popular operating system in the world, with over 2.5 billion active users spanning over 190 countries.
- Google Play was launched on March 6, 2012, bringing together Android Market marking a shift in Google's digital distribution strategy .
- Android is the dominant mobile operating system today more than 85% of all mobile devices running Google's OS. The Google Play Store is the largest and most popular Android app store.
- There are more than 3.04 million apps found on Google Play Store
- The Play Store apps data has enormous potential to drive app-making businesses to success.

2.1 PLAY STORE DATA & REVIEW ANALYSIS

In today's scenario we can see that mobile apps playing an important role in any individual's life. It has been seen that the development of the mobile application advertise has an incredible effect on advanced innovation. Having said that, with the consistently developing versatile application showcase there is additionally an eminent ascent of portable application designers inevitably bringing about high as can be income by the worldwide portable application industry. With enormous challenge from everywhere throughout the globe, it is basic for a designer to realize that he is continuing in the right heading. To hold this income and their place in the market the application designers may need to figure out

how to stick into their present position. The Google Play Store is observed to be the biggest application platform. It has been seen that although it creates more than two-fold the downloads than the Apple App Store yet makes just a large portion of the cash contrasted with the App Store. In this way, I scratched information from the Play Store to direct our examination on it.

With the fast development of advanced cells, portable applications (Mobile Apps) have turned out to be basic pieces of our lives. Be that as it may, it is troublesome for us to follow along the fact and to understand everything about the apps as new applications are entering market each day. It is accounted for that Android market achieved a large portion of a million applications in September 2011. Starting at now, 0.675 million Android applications are accessible on Google Play App Store. Such a lot of applications are by all accounts an extraordinary open door for clients to purchase from a wide determination extend. We trust versatile application clients consider online application surveys as a noteworthy impact for paid applications. It is trying for a potential client to peruse all the literary remarks and rating to settle on a choice. Additionally, application engineers experience issues in discovering how to improve the application execution dependent on generally speaking evaluations alone and would profit by understanding a huge number of printed remarks.

We develop Android apps & release on Play Store. As an Developer or say Business Perspective it's very important to know whether users are enjoying the app or facing any issues. To know this Play Store has a Ratings & reviews section for each app released on play store. Users can submit the ratings and has a freedom to write a review for a particular app. This approach is quite a lengthy to rate & review app i.e. navigate to Play store to submit feedback or redirect leaving a current app workflow to open Play Store App link using URI. We never wanted our customers to leave our application, but with this flow, we are forced to redirect the control to Play store app.

Google Play store Dataset

The dataset consists of Google play store application and is taken from Almabetter, which is the world's largest community for data scientists to explore, analyze and share data.

This dataset is for Web scratched information of 10k Play Store applications to analyze the market of android. Here it is a downloaded dataset which a user can use to examine the Android market of different use of classifications music, camera etc. With the assistance of this, client can predict see whether any given application will get lower or higher rating level. This dataset can be moreover used for future references for the proposal of any application. Additionally, the disconnected dataset is picked so as to choose the estimate exactly as online data gets revived all around a great part of the time. With the assistance of this dataset, I will examine various qualities like rating, free or paid and so forth utilizing Hive and after that I will likewise do forecast of various traits like client surveys, rating etc.

The data set contains the following columns:

App: This Column contains the name of the app

Category: This contains the category to which the app belongs. The category column contains 33 unique values.

Rating: This column contains the average value of the individual rating the app has received on the play store. Individual rating values can vary between 0 to 5.

Reviews: This column contains the number of people that have given their feedback for the app.

Size: This column contains the size of the app i.e. The memory space that the app occupies on the device after installation.

Installs: This column indicates the number of time that the app has been downloaded from the play store, these are approximate values and not absolute values.

Type: This column contains only two values- free and paid. They indicate whether the user must pay money to install the app on their device or not.

Price: For paid apps this column contains the price of the app, for free apps it contains the value 0.

Content Rating: It indicates the targeted audience of the app and their age group.

Genre: This column contains to which genre the app belongs to, genre can be considered as a sub division of Category.

Last updated: This column contains the info about the date on which the last update for the app was launched.

Current version: Contains information about the current version of the app available on the play store.

Android version: Contains information about the version of the android OS on which the app can be installed.

2.2 USER REVIEW DATA

- User reviews data frame has 64295 rows and 5 columns. The 5 columns are identified as follows:
- **App:** Contains the name of the app with a short description (optional).
- **Translated Review:** It contains the English translation of the review dropped by the user of the app.
- **Sentiment:** It gives the attitude/emotion of the writer. It can be 'Positive', 'Negative', or 'Neutral'.
- **Sentiment Polarity:** It gives the polarity of the review. Its range is [-1,1], where 1 means 'Positive statement' and -1 means a 'Negative statement'.
- **Sentiment Subjectivity:** This value gives how close a reviewer's opinion is to the opinion of the general public. Its range is [0,1]. Higher the subjectivity,

closer is the reviewer's opinion to the opinion of the general public, and lower subjectivity indicates the review is more of a factual information.

2.3 PYTHON

Most of the info scientist use python due to the good built-in library functions and therefore the decent community. Python now has 70,000 libraries. Python is simplest programing language to select up compared to other language. That is the most reason data scientists use python more often, for machine learning and data processing data analyst want to use some language which is straightforward to use. That is one among the most reasons to use python. Specifically, for data scientist the foremost popular data inbuilt open-source library is named panda. As we have seen earlier in our previous assignment once we got to plot scatterplot, heat maps, graphs, 3-dimensional data python built-in library comes very helpful.

2.4 DATA CLEANING & PREPARATION

Preprocessing is important into transitioning raw data into a more desirable format. Undergoing the preprocessing process can help with completeness and compellability. For instance, you'll see if certain values were recorded or not. Also, you'll see how trustable the info is. It could also help with finding how consistent the values are. We need preprocessing because most real-world data are dirty. Data can be noisy i.e. the data can contain outliers or simply errors generally. Data can also be incomplete i.e. there can be some missing values.

The available data is raw and unusable for Exploratory data analysis, so before we do anything with the data we will have to explore and clean it to prepare it for data analysis.

- **Step1:** I write a function play store info (), that will display 5 attributes about all the columns: Data type, Count of non-null values, Count of null values, number of unique values in that column and percentage of null value in that columns in the play store dataset.

- **Step2:** I start off with the column 'Type' we can see that it has one null value. We checked this row and found out from the play store that it is a free app. We use filllna() function of the pandas library to fill this value.

- **Step 3:** I have not dropped any column, I tried to substitute the missing values or inconsistencies with the help of Mode and Median values to avoid data loss and attain precise insights.

- **Step 4:** The 'Rating' column has 1474 null values. Due to low variations in the rating values and a lot of repeated values the 'median' would be a suitable statistical indicator to replace the null values with. We calculate the median of the column using the median () aggregate method, and fill this value in place of null values using the fillna() function.

- **Step 5:** The 'Reviews' column despite being a numerical indicator is of the 'object' data type, we will convert this to 'int' data type using the as type(int) function.

- **Step 6:** The size column, which should be numeric, is of the data type 'object', it also has characters 'k' and 'M' in the values which stand for kilobytes and Megabytes, I will replace the 'k' with 1000 and 'M' with 1000000. Some values also have '+' sign in them, which will be removed. Next, we will convert this column into 'int' datatype.

- **Step 7:** The 'Installs' column values contain the characters '+' and ',' which are going to prevent us from converting this column into a numeric datatype. We will get rid of these using the strip() and replace() functions.

- **Step 8:** The values in the column 'Price' is having the '$' sign in some values and the column is of the datatype 'object'. I will first remove the '$' sign using the strip() function and then convert the column into 'int' datatype.

- **Step 9:** Handling the duplicates in the App column we drop the no of duplicate rows that are present in the App columns. Final App count was 9660.

- **Step 10:** We write a function Ur info(), that will display 5 attributes about all the columns: Data type, Count of non-null values, Count of null values ,number of unique values in that column and percentage of null value in that columns in the User review dataset.

- **Step11:** In the User review dataset the columns are App, Translated Review, Sentiment, Sentiment Polarity, Sentiment Subjectivity in this total 26863 NaN value are present so we drop them using dropna() function.

3. EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset.

EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better. In this article, we will understand EDA with the help of an example dataset. We will use Python language (Pandas library) for this purpose.

3.1 Free vs Paid – Market Share

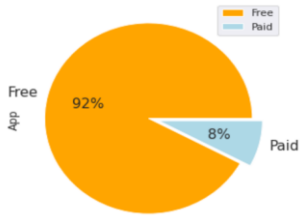


Fig -1: Free vs Paid

INFERENCE:

Here we can see that 92% apps are free, and 8% apps are paid on Google Play Store, so we can say that Most of the apps are free on Google Play Store.

3.2 Pair wise plot – 4 variables

In the below plot, we plotted the pair wise graphs for Rating, Size, Installs and Reviews to understand their relationship and impact on each other. Also to understand their distribution differences in Free and Paid Apps

Fig -2: Pair wise plot 1

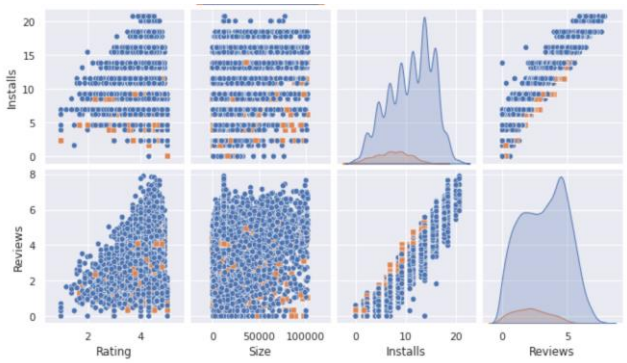
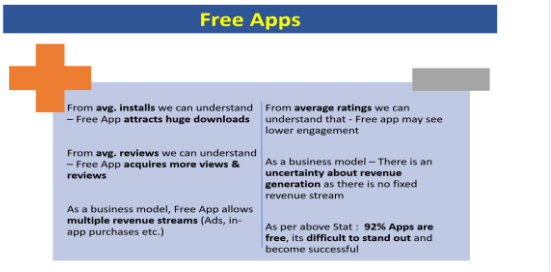
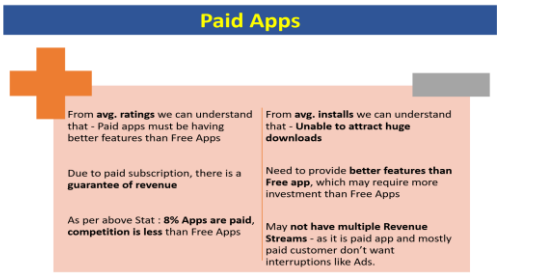


Fig -3: Pair wise plot 2

INFERENCE:

- The Market share of Free Apps and Paid Apps is drastically different
- Contribution of Customer segments are also different in both
- Sentiment Polarity & subjectivity is not same for the both
- Pros & Cons of Free App & Paid App Business Models





3.3 Install Vs Reviews

The below plots presents relation of Installs and reviews of the Free Apps and Paid Apps,

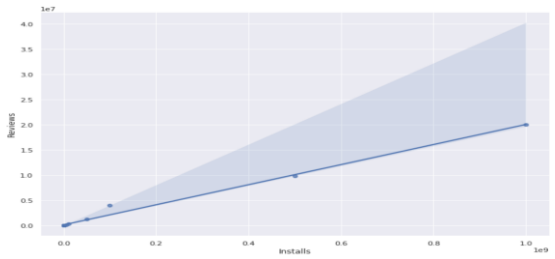


Fig -4 : Free App – Install Vs Reviews

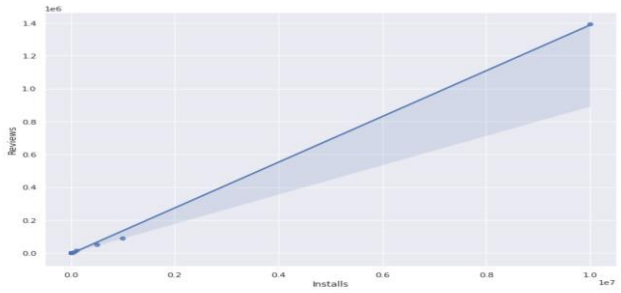


Fig -5 : Paid App – Install Vs Reviews

INFERENCE -

1. For free apps - reviews are directly proportional to installs
2. For Paid Apps also, reviews are directly proportional to installs

3.4 Install Vs Ratings

The below plots presents relation of Installs and Rating of the Free Apps and Paid Apps,

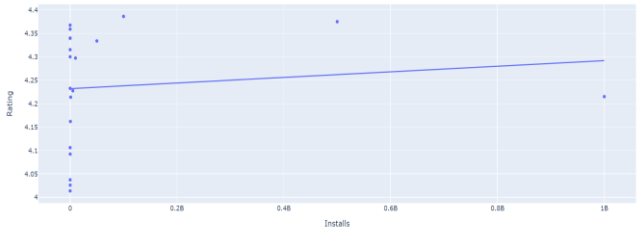


Fig -6 : Free App – Install Vs Rating

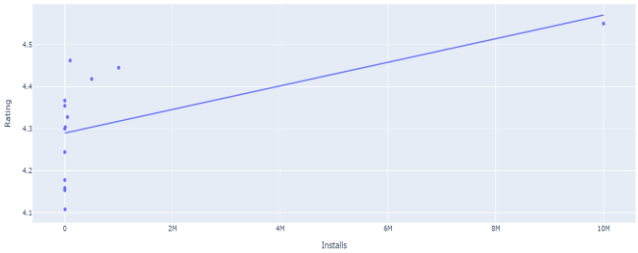


Fig -7 : Paid App – Install Vs Rating

INFERENCE -

1. Here we can see, as per graph, there is positive increment in Ratings with respect to Installs for Free Apps
2. Rate of increase is slow but it's crucial as it affects performance of the App significantly.
3. As per Paid App graph, there is a good positive increment in Ratings with respect to Installs for Paid Apps
4. Rate of increase of rating is higher than the increment rate of ratings in Free Apps

3.5 Content rating wise analysis

The below plots presents content rating category wise apps % presence in Free App & Paid App category,

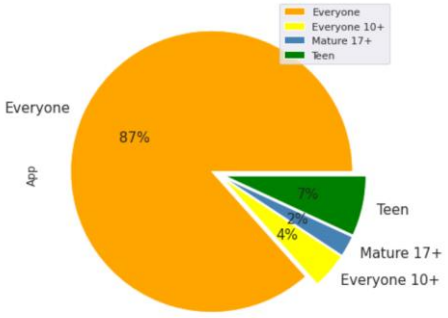


Fig -8 : Paid App – Rating wise App %

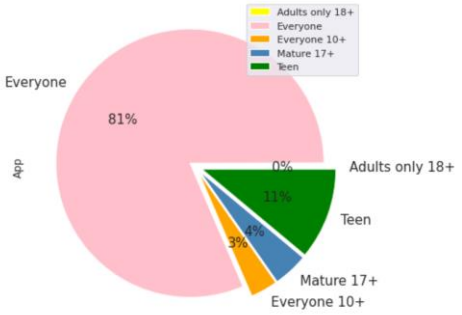


Fig -9 : Free App – Rating wise App %

INFERENCE –

- In Paid Apps segment 'Everyone' category apps contributing to 87% of the App Market. Which is bigger than Free Apps. Hence we can expect very high competition in this segment.
- In Free Apps segment 'Everyone' category apps contributing to 81% of the App Market. Hence we can expect very high competition in this segment.

3.6 Top 3 Content Rating – Contributing to best ratings

The below table presents 3 top content rating groups which contributing to higher ratings in the Free Apps and Paid Apps,

Content_Rating	Rating
Everyone	4.443094
Everyone 10+	4.444000
Teen	4.427742
Mature 17+	4.369444

Fig -10 : Free App – Top 3 Content rating providing best Ratings

Content_Rating	Rating
Everyone	4.565306
Teen	4.521429
Everyone 10+	4.500000
Mature 17+	4.500000

Fig -11 : Paid App – Top 3 Content rating providing best Ratings

3.7 Top 3 Best Content Rating – Free & paid Apps

The below table presents 3 top content rating groups which contributing to higher ratings in the Free Apps and Paid Apps,

Content_Rating	Rating
Everyone	4.443094
Everyone 10+	4.444000
Teen	4.427742
Mature 17+	4.369444

Fig -12 : Paid App – Top 3 Content rating Free Apps

Content_Rating
Mature 17+
Everyone 10+
Everyone
Teen

Fig -13 : Paid App – Top 3 Content rating paid apps

3.8 Top 3 Content Rating – Maximum contribution in successful - Free & Paid Apps

The below table presents 3 top content rating groups which contributing to higher ratings in the Free Apps and Paid Apps,

Fig -14 : Free App – Top 3 Content rating contributors to success

Fig -15 : Paid App – Top 3 Content rating contributors to success

3.9 Identify most preferred app size - Free & Paid Apps

The below table presents stats of Free Apps and Paid Apps Size,

```
count       784.000000
mean      29818.216837
std       26343.315824
min        1228.000000
25%       12288.000000
50%       12288.000000
75%       46080.000000
max      102400.000000
Name: Size, dtype: float64
```

Fig -16 : Free App – Most preferred App Size

```
count        73.000000
mean      30489.452055
std       27352.354761
min          50.000000
25%       12288.000000
50%       21504.000000
75%       44032.000000
max      102400.000000
Name: Size, dtype: float64
```

Fig -17 : Paid App – Most preferred App Size

3.10 Identify content rating wise preferred app size - Free & Paid Apps

The below table presents stats of content rating wise most preferred Free Apps and Paid Apps Size,

Content_Rating	Size
Everyone 10+	51507.200000
Teen	38573.741935
Mature 17+	28407.444444
Everyone	25415.329650

Fig -18 : Paid App – Content rating wise preferred App Size

Content_Rating	Size
Teen	44609.714286
Everyone 10+	43904.000000
Mature 17+	28160.000000
Everyone	24360.040816

Fig -19 : Free App – Content rating wise preferred App Size

3.11 Identify most preferred app size - Free & Paid Apps

The below table presents stats of the top 5 heavy apps in terms of size for Free Apps and Paid Apps Size,

Category	Size
PARENTING	63488.000000
FAMILY	47610.154472
GAME	46954.540541
FINANCE	42170.181818
SPORTS	34443.636364

Fig -20 : Free App – Category wise preferred App Size

Category	Size
SPORTS	74752.000000
GAME	48038.375000
FAMILY	30760.900000
BUSINESS	23449.500000
PHOTOGRAPHY	20684.333333

Fig -21 : Paid App – Category wise preferred App Size

3.12 Correlation analysis - Free & Paid Apps

The below plot presents statistical correlation of various variables of play store data for Free Apps and Paid Apps Size,

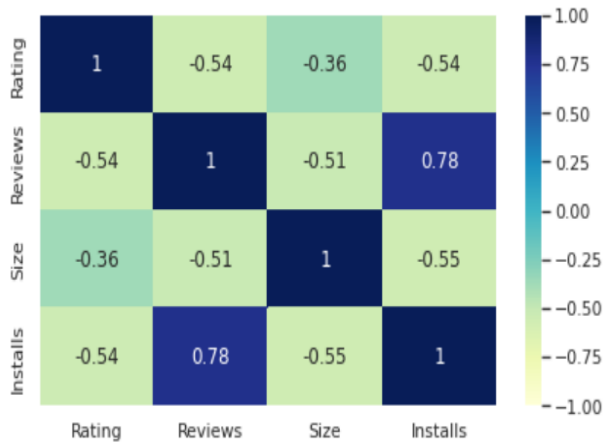


Fig -22 : Correlation heatmap – Free Apps

INFERENCE -

1. For Free App category, heatmap given beside shows strong positive correlation between Installs & Revi ews i.e. 0.78
2. Size & Installs and Rating & Installs shows significant negative correlation
3. Lowest negative correlation is between rating and Size but it's significant i.e. -0.36
4. Need to further investigate these relations to understand exact trend of these variables and its long term impact on App performance

Fig -23 : Correlation heatmap – paid Apps

INFERENCE -

1. For Paid App category, heatmap given beside shows strong positive correlation between Installs & Revi ews i.e. 0.96
2. Price, Size & rating shows negative correlation with respect to all variables
3. Lowest negative correlation is between Size & rating and Size and installs, they must be stable variables for paid apps
4. Let's further investigate these relations to understand exact trend of these variables and its long term impact on App performance

3.12 IMPACT OF UPDATES ON SIZE OF - FREE & PAID APPS

The below plot presents impact of updates on the size of the Free Apps and Paid Apps,

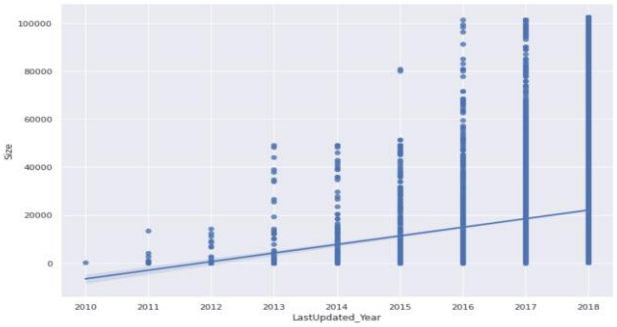


Fig -24 : Update vs Size – Free Apps

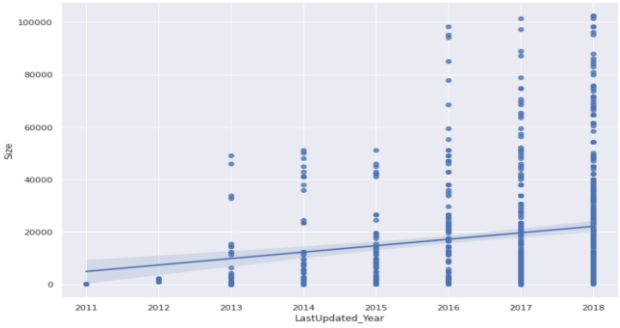


Fig -25 : Update vs Size – Paid Apps

INFERENCE –

1. In Free apps, as graph shows the Size of the App is increasing with the new updates in Apps
2. New updates helps to enhance new features in app, quality of the visuals and processing speed of the data, due to these features Apps must be getting heavy

3. In Paid apps also, as graph shows the Size of the App is increasing with the new updates in Apps
4. New updates helps to enhance new features in app, quality of the visuals and processing speed of the data, due to these features Apps must be getting heavy

3.13 PRICE VS SIZE - PAID APPS

The below plot presents impact of size on the price of the Paid Apps,

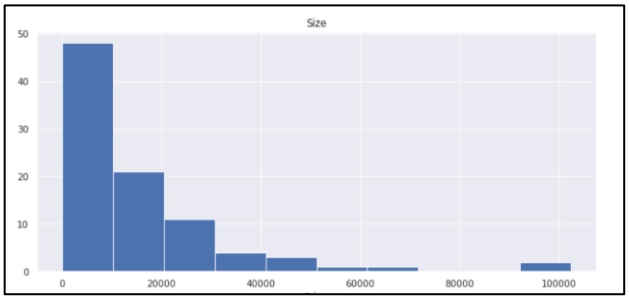


Fig -26 : Price vs Size

INFERENCE -

Here we can see most of the Apps are having app size up to 30 MB and they all are saturated around price range of Rs 10 - Rs 20/-

3.14 PRICE TREND ANALYSIS

Below plot shows the trend relation of price and various categories of paid apps,

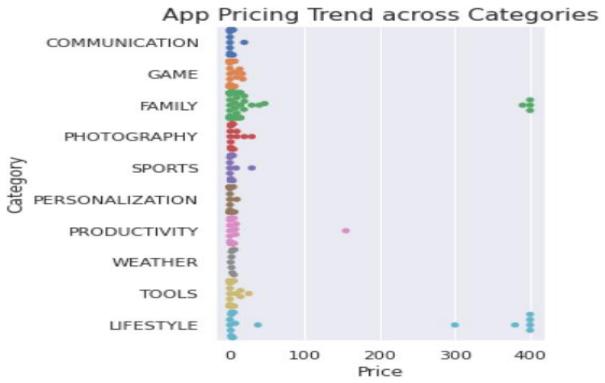


Fig -27 : Price trend analysis

INFERNCE –

1. As per price trend reflecting in chart given beside, most of the Apps in all the categories are trying to maintaining their price in the range of 10 - 20 Rs.
2. And as we have seen earlier, it is the most acceptable price among customers for Paid Apps

3.15 TOP 5 ANDROID VERSIONS – FREE & PAID APPS

Below plot shows the trend relation of android versions and no of users for free and paid apps,

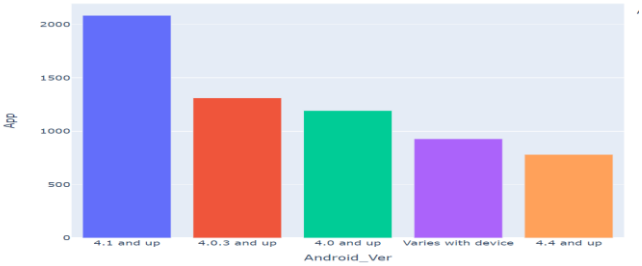


Fig -28: Top 5 Android Versions - Free Apps

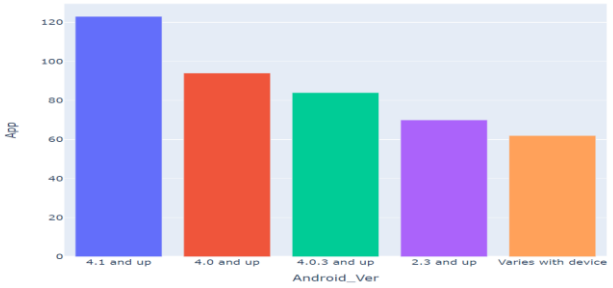


Fig -29: Top 5 Android Versions - Paid Apps

INFERNCE –

As per plot most accepted top 5 android versions are of series 4.0 and above for both Free and Paid Apps.

3.16 TOP 5 CURRENT VERSIONS – FREE & PAID APPS

Below plot shows the trend relation of current versions and no of users for free and paid apps,

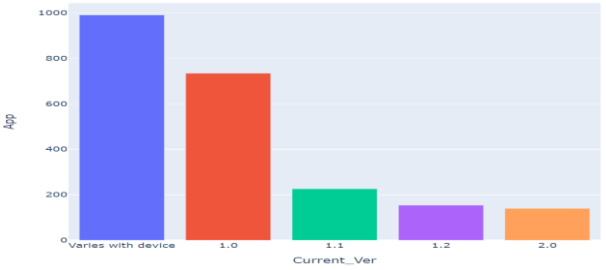


Fig -30: Top 5 Current Versions - Free Apps

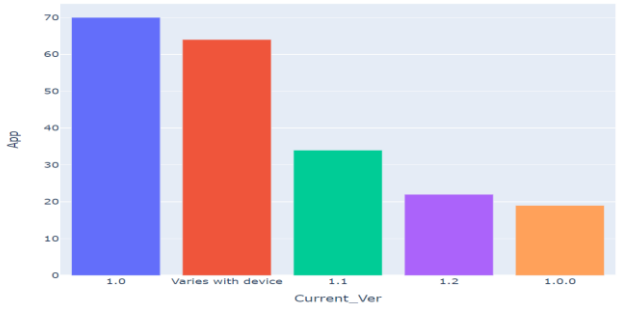


Fig -31: Top 5 Current Versions - Paid Apps

INFERNCE –

As per plot most accepted top 5 current versions are of series 1.0 and above for both Free and Paid Apps.

3.17 CURRENT VERSION VS SIZE – FREE & PAID APPS

Below plot shows the relation of current versions and size of the free and paid apps,

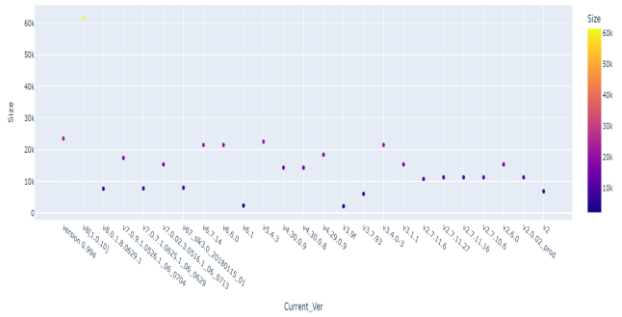


Fig -32: Current Versions Vs Size - Free Apps

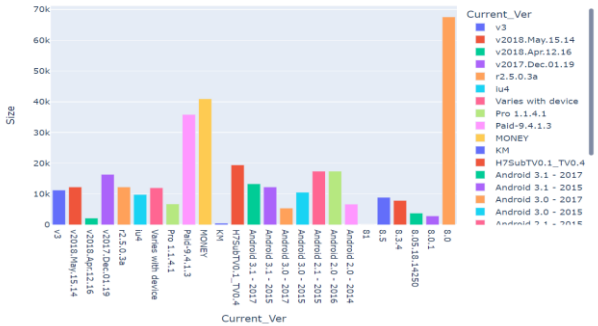


Fig -33: Current Versions Vs Size - Paid Apps

INFERENCE –

1. In Free Apps also, as per chart above, we can say that, No relation can be established between Current version and Size of the App.
2. In Paid Apps, as per chart below, we can say that, No relation can be established between Current version and Size of the App

3.18 ANDROID VERSION VS RATING – FREE & PAID APPS

Below plot shows the relation of android versions and rating of the free and paid apps,

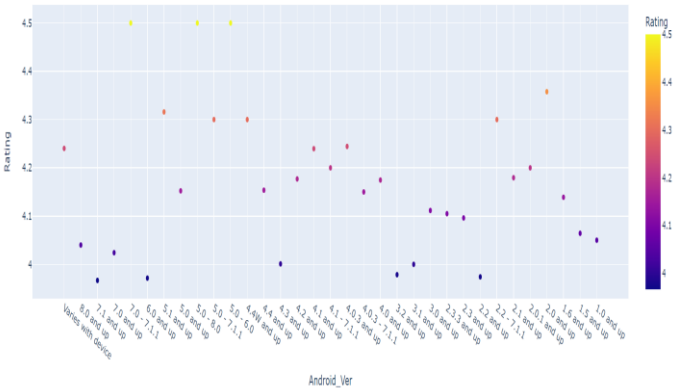


Fig -34: Android Versions Vs Rating - Free Apps

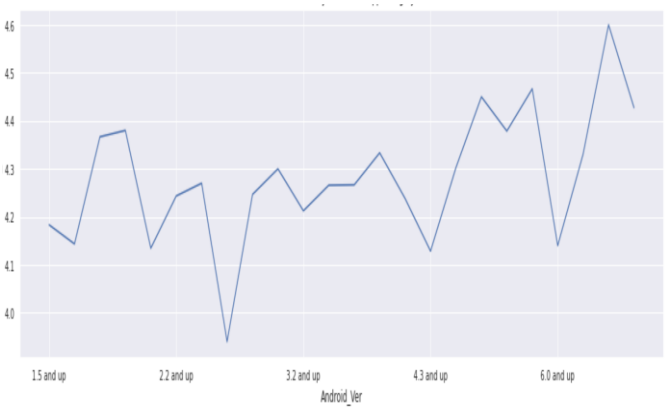


Fig -35: Android Versions Vs Rating - Paid Apps

INFERENCE –

1. In Free Apps, as per chart above, Android versions of Series 4.0 and above are showing good ratings and customers positive acceptance
2. In Paid Apps, as per chart below, Android versions of Series 4.0 to Series 4.4 are showing good ratings and customers positive acceptance

3.19 ANDROID VERSION VS REVIEWS – FREE & PAID APPS

Below plot shows the relation of android versions and reviews of the free and paid apps,

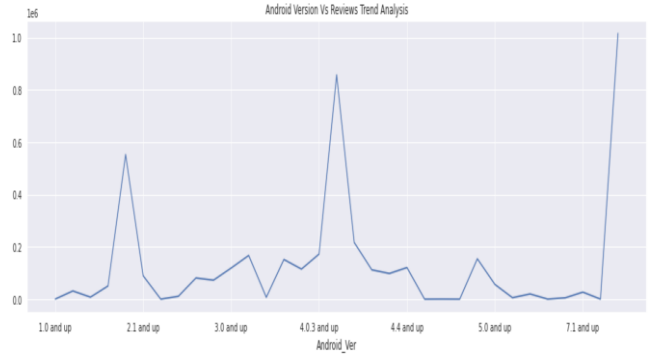


Fig -36: Android Versions Vs Reviews - Free Apps

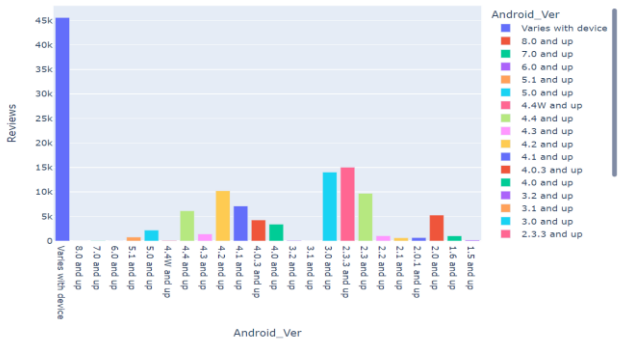


Fig -37: Android Versions Vs Reviews - Paid Apps

INFERENCE –

1. In Free Apps, Android versions of Series 4.0 to Series 4.4 are showing highest mean reviews.
2. In Paid Apps also, Android versions of Series 4.0 to Series 4.4 and Series 2.3 to Series 3.0 are showing highest mean reviews.

3.20 ANDROID VERSION VS SIZE – FREE & PAID APPS

Below plot shows the relation of android versions and size of the free and paid apps,

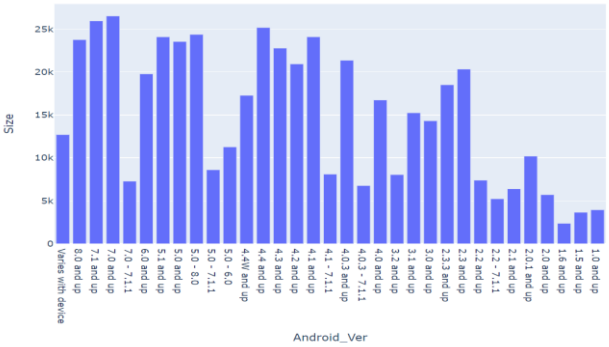


Fig -38: Android Versions Vs Size - Free Apps

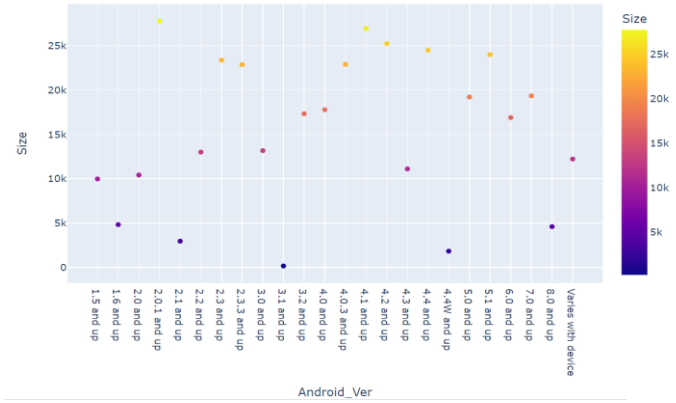


Fig -39: Android Versions Vs Size - Paid Apps

INFERENCE –

1. In Free Apps, as per above graph, we can easily make out the trend of increase in size with latest Android versions.
2. In Paid Apps also, we can say Size increment trend with latest Android versions.

3.21 ANDROID VERSION VS PRICE – FREE & PAID APPS

Below plot shows the relation of android versions and price of the paid apps,

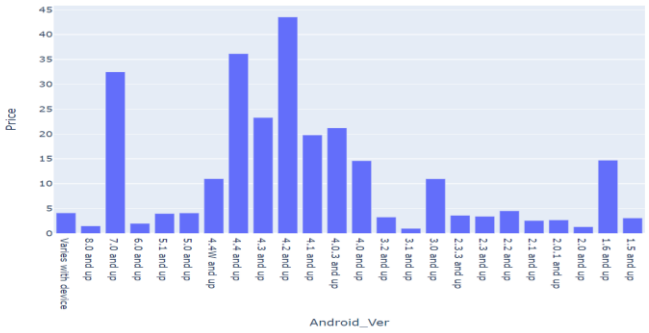


Fig -40: Android Versions Vs Price - Paid Apps

INFERENCE –

1. In Paid Apps, Android Version Series of 4.0 to 4.4 are mostly accepted version
2. As we can check in graph most of the prices are concentrated around Android version 4.0 to 4.4
3. And these Android versions are having average subscription cost of around Rs 10 - 20/-

3.22 REVIEWS VS RATING – FREE & PAID APPS

Below plot shows the relation of reviews and rating of the paid apps,

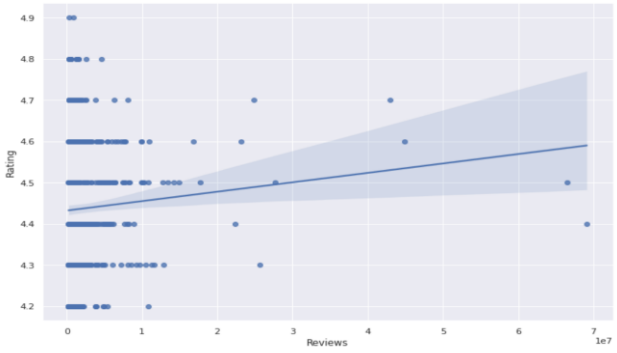


Fig -41: Reviews Vs Rating - Free Apps

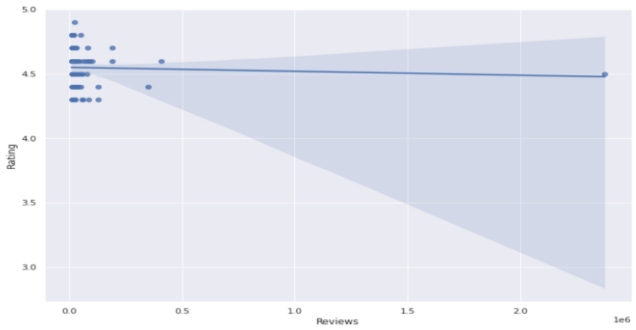


Fig -42: Reviews Vs Rating - Paid Apps

INFERENCE –

1. As per below chart, in successful Free Apps, Ratings are showing positive incremental trends with increasing reviews.
2. In Successful Paid Apps, Ratings are showing slightly negative incremental trends with increasing reviews.

3.23 REVIEWS VS SIZE – FREE & PAID APPS

Below plot shows the relation of reviews and size of the paid apps,

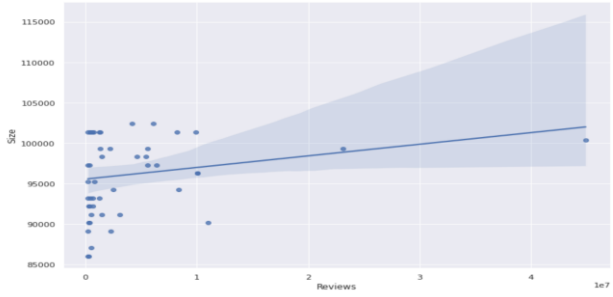


Fig -43: Reviews Vs Size - Free Apps

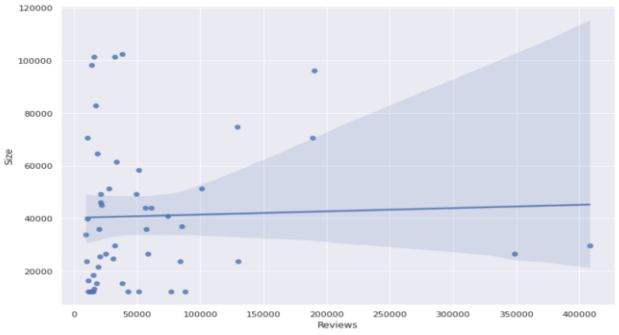


Fig -44: Reviews Vs Size - Paid Apps

INFERENCE –

1. In Free Apps, with increase in Size reviews are showing positive incremental trend.
2. In Paid Apps, with increase in Size of the app, reviews are not getting changed, they are stable.

3.24 PRICE VS RATING– PAID APPS

Below plot shows the relation of price and rating of the paid apps,

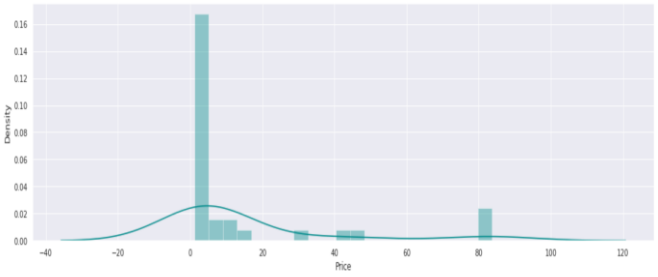


Fig -45: Price Vs Rating

INFERENCE – People have highest acceptance of price, in the range of Rs 10 – 20/-

3.25 UPDATE VS RATING– PAID APPS

Below plot shows the relation of price and rating of the paid apps,

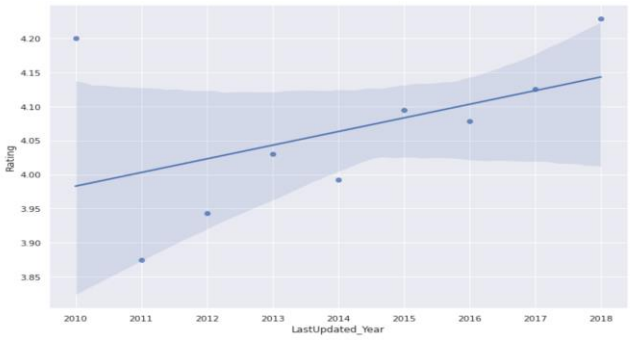


Fig -46: Update Vs Rating – Free Apps

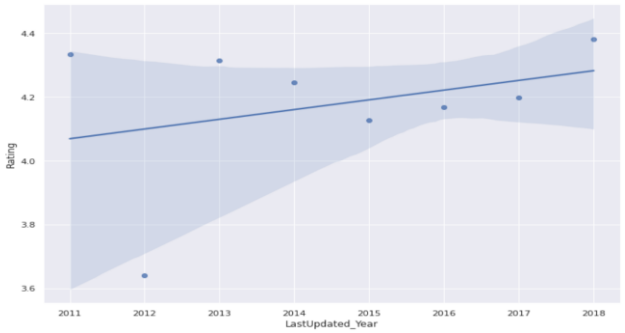


Fig -47: Update Vs Rating – Paid Apps

INFERENCE –

1. In Free Apps, with Latest updates of Apps their rating is also increasing.
2. In Paid Apps, with Latest updates of Apps their rating is also increasing

3.27 SENTIMENT CORRELATION ANALYSIS

Below plot shows the correlation of the various variables of sentiment data for the free and paid apps,

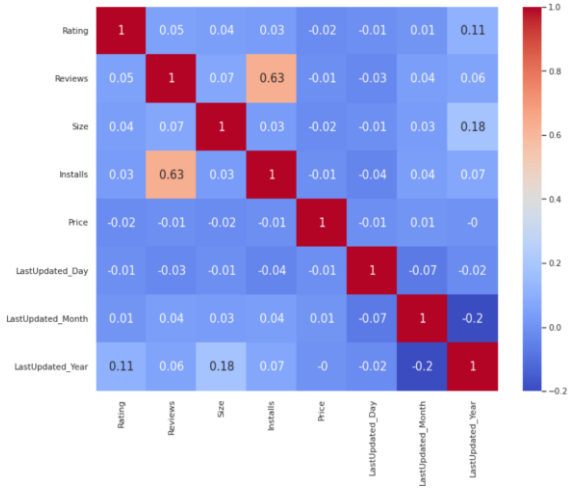


Fig -48: Correlation Heatmap Sentiments data

INFERENCE :

1. Here we can see strong correlation of Installs & Rating.
2. Other parameters are not showing any significant correlation with each other.
3. As per some normal incremental correlation coefficient no. like 0.11 (Last Update & Rating), 0.18 (Last Update & Size) etc. may show us some slight trends over a period.

3.28 SENTIMENT SUBJECTIVITY DISTRIBUTION

Below plot shows the distribution of sentiment subjectivity between Free and Paid Apps ,

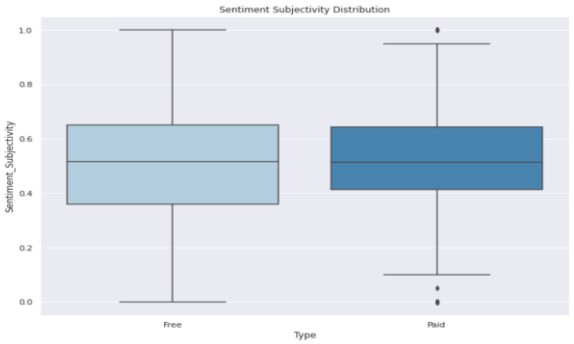


Fig -49: Sent. Subjectivity Distribution

INFERENCE –

1. Here this Sentiment Subjectivity Box Plot also giving us same sentiment trend like polarity
2. Here also, we can confirm that customers sentiment subjectivity i.e. intensity of emotions while expressing their views is stable and better in terms of Paid Apps. That might be due to Paid Apps are providing good services than the Free Apps.
3. We can also see that Free Apps sentiment subjectivity is slightly negatively skewed as compared to Paid Apps

3.29 SENTIMENT POLARITY DISTRIBUTION

Below plot shows the distribution of sentiment polarity between Free and Paid Apps ,

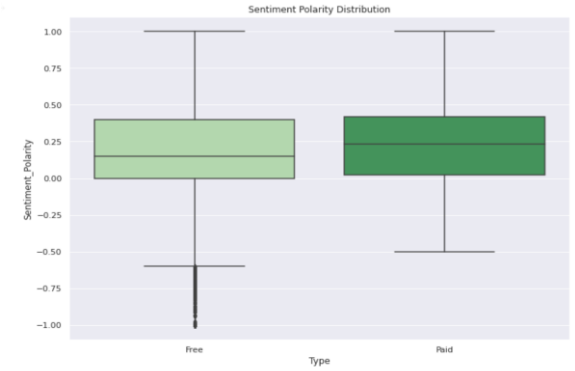


Fig -50: Sent. Polarity Distribution

INFERENCE –

1. Here this Box Plot gives us five important numbers related to relation of Sentiment Polarity in Free & Paid Apps
2. Free apps have higher sentiment polarity parameters with respect to Paid Apps (Median, Max, Inter Quartile Ranges etc.)
3. Here we understand that customers sentiment are stable and better in terms of Paid Apps. That might be due to Paid Apps are providing good services than the Free Apps

4. Here we can see that Free Apps sentiment polarity is slightly negatively skewed as compared to Paid Apps

3.30 OVERALL POSITIVITY AND SUBJECTIVITY DISTRIBUTION

Below plot shows the distribution of overall sentiment positivity and subjectivity between Free and Paid Apps ,

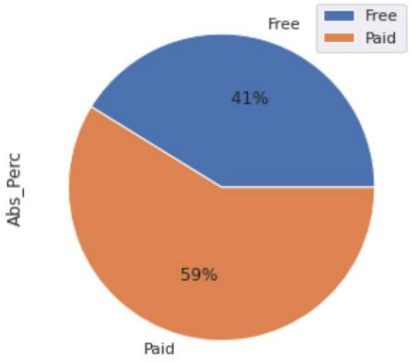


Fig -51: Overall Positivity Distribution

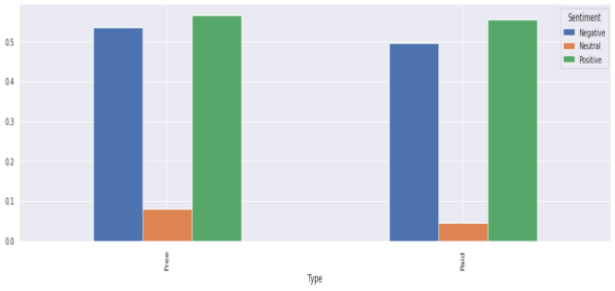


Fig -52: Overall Subjectivity Distribution

INFERENCE -

1. In this pie chart we can see that as a whole people are positive about Free & Paid both the apps
2. People are slightly conservative while giving their reviews about Apps
3. We can also see that overall positivity of Paid Apps is more that Free Apps. This means people are finding Paid App service better than Free Apps
4. In above bar chart we can see positive bar is higher than negative bar, people are overall positive about Free & Paid Apps
5. Very few people are neutral, most of the people are vocal about their experiences
6. Overall Positive subjectivity is more

3.31 CONTENT RATING WISE SENTIMENT POLARITY

Below plot shows the distribution of sentiment polarity content rating wise between Free and Paid Apps ,

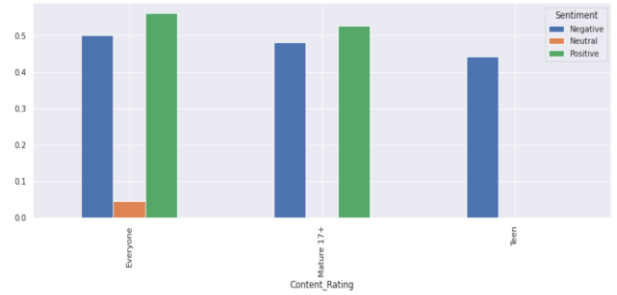


Fig -53: Content rating wise Sent. Polarity – Paid Apps

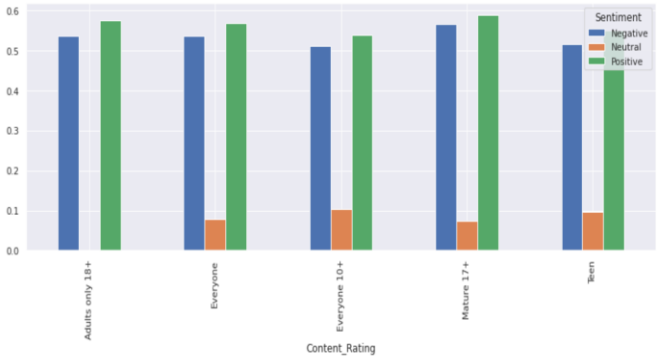


Fig -54: Content rating wise Sent. Polarity – Free Apps

INFERENCE -

1. In both the bar charts its prominent that overall sentiments are positive about the free as well paid apps
2. People looks conservative while giving negative feedback

3.32 IMPACT OF SIZE ON SENTIMENTS – FREE AND PAID APPS

Below plot shows the impact of size on the sentiment polarity and subjectivity among users for Free and Paid Apps ,

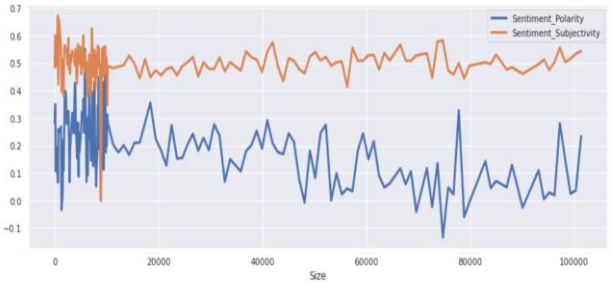


Fig -55: Size Vs Sentiment Polarity & Sub.– Free Apps

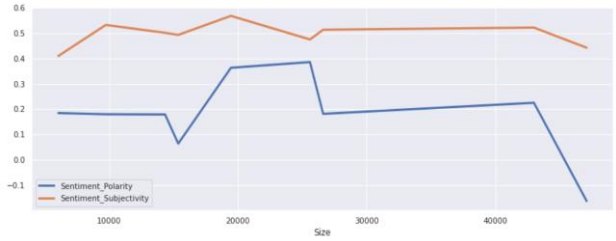


Fig -56: Size Vs Sentiment Polarity & Sub. – Paid Apps

INFERENCE –

1. For Free Apps, increase in size does not have significant impact on subjectivity
2. For Paid Apps also, increase in size does not have significant impact on subjectivity

3.33 IMPACT OF INSTALLS ON SENTIMENTS – FREE AND PAID APPS

Below plot shows the impact of installs on the sentiment polarity and subjectivity among users for Free and Paid Apps ,

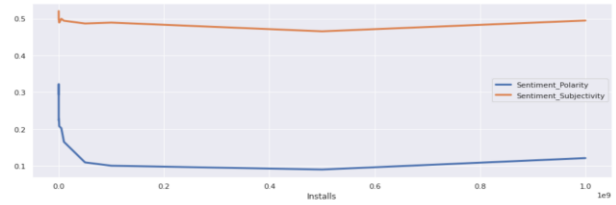


Fig -57: Installs Vs Sentiment Polarity & Sub. – Free Apps

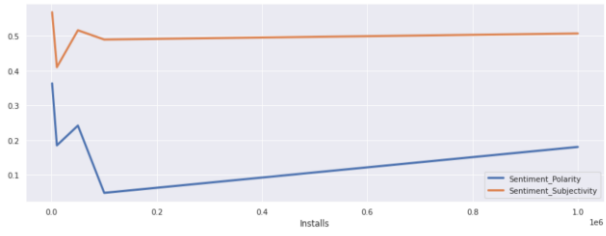


Fig -58: Installs Vs Sentiment Polarity & Sub. – Paid Apps

INFERENCE –

1. For Free Apps, increasing installs does not show any significant positive / negative trend for both sentiment polarity as well as sentiment subjectivity.
2. For Paid Apps, increasing installs are showing slight positive trend for sentiment polarity.
3. Subjectivity is stable while increasing installs

3.34 IMPACT OF RATING ON SENTIMENTS – FREE AND PAID APPS

Below plot shows the impact of rating on the sentiment polarity and subjectivity among users for Free and Paid Apps ,

Fig -59: Rating Vs Sentiment Polarity & Sub. – Free Apps

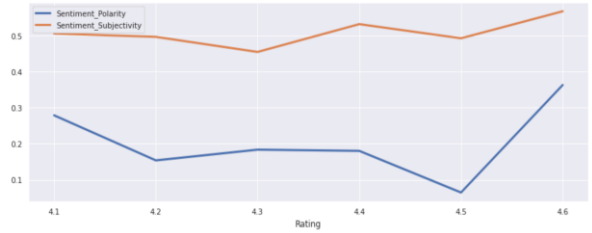


Fig -60: Rating Vs Sentiment Polarity & Sub. – Paid Apps

INFERENCE –

1. For Free Apps, in chart we can see rating is highly correlated with rating.
2. For Paid Apps also, increasing rating are showing positive trend for sentiment polarity & subjectivity both.
3. Rate of polarity increase is higher than subjectivity

3.35 IMPACT OF UPDATE ON SENTIMENTS – FREE AND PAID APPS

Below plot shows the impact of update on the sentiment polarity and subjectivity among users for Free and Paid Apps ,

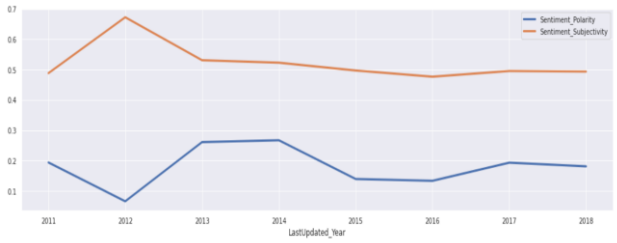


Fig -61: Rating Vs Sentiment Polarity & Sub. – Free Apps

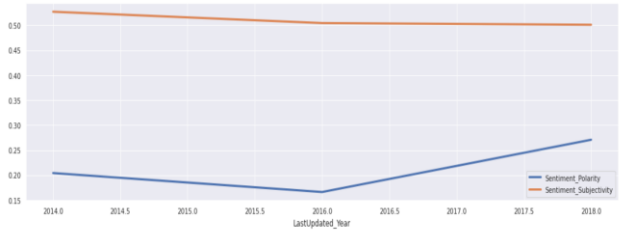


Fig -62: Rating Vs Sentiment Polarity & Sub. – Paid Apps

INFERENCE –

1. For Paid Apps, recent last updates are showing positive trend for sentiment polarity.
2. For Free Apps, recent last updates are not showing significant positive trend for both sentiment polarity as well as Sentiment Subjectivity
3. Polarity & Subjectivity both looks stable

3.36 RELATION OF SENTIMENT POLARITY AND SUBJECTIVITY

Below plot shows the relation of sentiment polarity and subjectivity for Free and Paid Apps ,

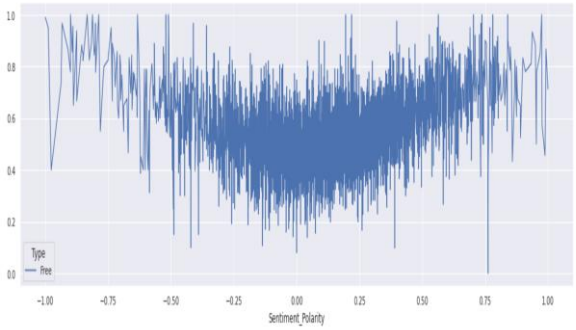


Fig -63: Sent. Polarity Vs Sentiment Subjectivity

INFERENCE –

1. Here we can see that, as polarity is going towards two extremes sentiment subjectivity is increasing.
2. People having their views very intense and emotional for highly negative or highly positive reviews
3. Polarity is moving towards high subjectivity to positive as well as negative end.

3.37 WORD CLOUD ANALYSIS OF REVIEWS

Below plot shows the word cloud representation of reviews text data for Free and Paid Apps ,

Fig -64: Word Cloud Free App

INFERENCE –

4. Game – It's an prominent word here, it means many people who are Using Gaming Apps are writing reviews and expressing their views frequently. (some similar words like level, play etc gives same indication)
5. Great / good / Love / best / nice / thank – All these words are showing positive vibes about reviews and app features.
6. Update / account / work / still – Such words indicating concerns of people. So need to focus on such feedbacks.
7. Start slow / another / note / bad device – These words also represent serious concerns or feedback from customers. Need to look after these reviews as well.

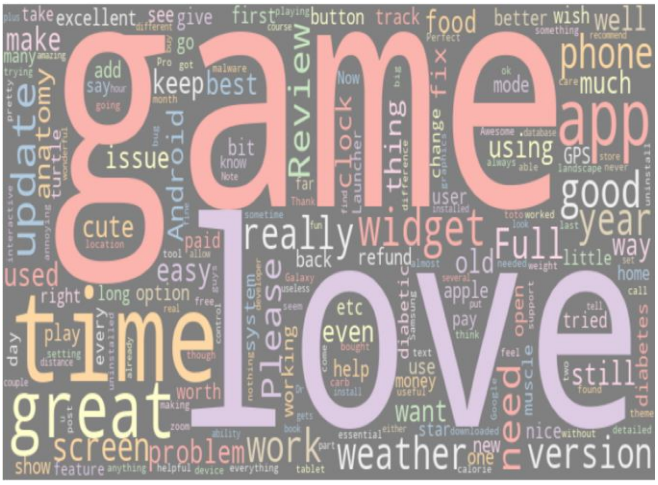


Fig -65: Word Cloud Paid App

INFERENCE –

1. Great / good – These kind of prominent words are showing positive sentiments about paid apps.

2. Game – many paid apps are Gaming apps, hence this word is also prominent here
3. Love – These kinds of word indicates, that there are many dating apps in paid apps type and people are frequently using this word.
4. Diabetes / tablet / calorie – These words indicates that there are many health related apps in paid apps type.
5. Update / Please / Problem / Version / Change / User / Issue - Such words indicating concerns of people. They might be wanted App makers to update their version, fix some bugs etc. So need to focus on such feedbacks.

3.38 CONCLUSION :

Conclusion cum Recommendations (App makers Management Team)

1. In order to start planning or strategizing launch of any New App, you need to finalise a Business Model first. Either Free App model or Paid App Model
2. Every model has their own Pros & Cons (refer slide 17 & 18), hence you need to check the same, that which model suits to your budget, resources and skill sets.
3. If you want to explore both the Business models and want to pursue their cost benefit analysis, then you have to choose an App category which works best as per your capabilities.
4. Here we would like to suggest some important genres and categories which may help you to build and launch a profitable App
5. In Free App category – Most successful categories are as below Annexure – 1 of our report will help you with all these details
6. You can select desired category and app to start your Free App planning accordingly.
7. If you want to initiate a Paid Type of App then you can start with below top Paid App categories
    - *Gaming Apps | Communication | Lifestyle | Sports | Family*
    - *Annexure – 2 of our report will help you with all these details*
8. You can select desired category and app to start your Free App planning
9. Also you can it by your Target Audience category, there are five different customer segments in Free & paid App category,
    - Everyone | Everyone 10+ | Teen | Mature 17+ | Adult 18+
10. In these categories Everyone & Everyone 10+ categories are in top three customer bases of both Free and Paid Apps
11. You can target these categories because,
    - These 2 categories contributing at the extent of around 66% and 50% respectively to the success of Free and paid apps
    - Mature 17+ category alone contributing @ 50% to the success of paid apps
    - In Annexure 3 to 8, we have given most preferred Apps list by each of these categories separately. You can refer the same for finalizing an app type.
    - We also need to consider a huge competition in theses three categories, more than 90% apps on play store are serving these three categories. Hence competition will be tough. Our App needs to have an uniqueness in it's services to become successful.
12. In Free app category people are preferring an app size of 12 – 30 MB and in paid App category people are preferring an avg. app size between 23 to 45 MB
13. You can also use varies with devise app versions as people have positive acceptance towards that as well.
14. App needs to get updated frequently with latest technology
15. Periodical updates affects user experience, reviews and rating of app.
16. Due to additional features / quality of user interface / storage these things are making paid apps slightly heavier than free apps
17. If we go in details content rating wise also we can select appropriate size for our app.
18. In terms of technology people are preferring Android Version apps having version series 4.0 and above. These updated versions are mostly appreciated and liked by people.
19. Certain categories like Sport / family / Game includes video streaming and storage hence these apps are most heavier apps in terms of size on the play store
20. As we investigated correlation of various important factors which affects performance or success of apps, we found below observations

21. Size and Ratings in Free App category are negatively correlated with each other.
22. Installs and Reviews are strongly correlated with each other. Reviews are mostly positive.
23. As a whole and on an average People are showing overall positivity in reviews.
24. Installs and rating are slightly positively correlated. Slow rate of rating growth with increasing installs.
25. These positive reviews are pushing installations of app
26. But as installations are increasing, as people have already gone through positive reviews they expect higher than those reviews.
27. And Free apps are not satisfying their expectation in terms of features and hence people are getting slightly negative and showing a low rate of growth in rating with increasing installs.
28. After an installation threshold of 0.5 Billion rating rate slows down, Need to review customer sentiment at this at 0.5 B installs of app.
29. In paid apps also we saw a positive correlation of Installs and reviews.
30. Hence for success of an app, reviews are very important, they are driving success of the app. You need to develop a mechanism in app which will frequently ask for reviews and gain maximum reviews.
31. In paid apps Size and Price are adversely affecting apps success.
32. Make sure to have App pricing between Rs 10 -20/- Maximum people are tend to pay this amount on an average. Because of this you can gain maximum customers and to generate more revenue you can explore another methods like in app purchases, temp paid services, temp subscription etc
33. While planning operations of the App, you need to have mechanism to address people queries or negative reviews.
34. Sentiment Subjectivity is high within Mature 17+ , Teen and Everyone 10+ categories.
35. So while planning of these category apps need to take this feedback redressal mechanism
36. TECHNICAL RECOMMENDATIONS

Conclusion – Technical Recommendations

To App Development Team

Technical Recommendations	Details
Technology	Android Version - Series 4.0 and above or Android version - Varies with Devise versions
Size	Free App - between 12 - 30 MB | Paid App - 23 - 45 MB | Heavy App - Max. 60 MB for paid Apps and 40 MB for Free Apps
Review Mechanism	A pop-up mechanism for review after every long activity on App Periodical review mechanism for people who gives frequent review Negative feedback tracing and tracking of redressal - Mechanism / Feature
User Interface	Free App - Can be of moderate technology | Paid App - Should be HD technology for better user experience
App Updates	App updation should be done periodically with latest technology
App Experience	Our Annexures consist of some top play store apps, study their technical specification to get an edge over other competitors
Annexures	Annexure 1 - Top 10 Genres | Catgories | Apps (Free Apps)
	Annexure 2 - Top 10 Genres | Catgories | Apps (Paid Apps)
	Annexure 3 - Top 10 Apps (Everyone Segment - Free apps)
	Annexure 4 - Top 10 Apps (Everyone 10+ Segment - Free apps)
	Annexure 5 - Top 10 Apps (Teen Segment - Free apps)
	Annexure 6 - Top 10 Apps (Mature 17+ Segment - Paid apps)
	Annexure 7 - Top 10 Apps (Everyone 10+ Segment - Paid apps)
	Annexure 8 - Top 10 Apps (Everyone Segment - Paid apps)
	Annexure 9 - Top 10 Apps (Most Positive Sentiments - Free apps)
	Annexure 10 - Top 10 Apps (Most Positive Sentiments - Paid apps)

References~

- GeeksforGeeks
- Analytics Vidhya
- Stackoverflow
- Towards data science
- Python libraries documentation
- Data camp

Thank You