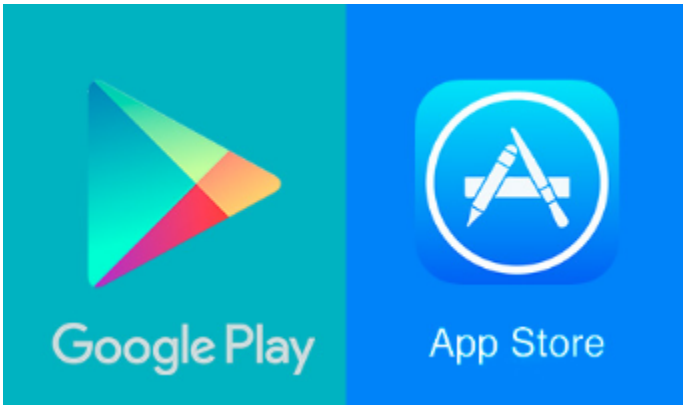
Name - Pratik Bhujade

# Play Store vs. Apps Store

## Comparative Study of Mobile and Tablet Applications on Google Play Store and Apple Apps Store



## Introduction

With the advent of mobile technology and support to innovation the two major ecosystems namely Google Play Store and Apple Apps Store which serve as the backbone of current era of usage of mobile computing technology seems to be the biggest rivals. This ecosystem of mobile application gives a chance to new and emerging mobile apps developers to showcase their talent and create businesses. But with more usage of these technologies it is important to keep a regulatory check on the content being served to the audience. Thus, the aim of this analysis is to analyse the type of third party applications on these ecosystems. The two operating systems were primarily designed for mobile computing devices but now expanded to various other utility and luxury devices. External applications can be created by third party which can be easily downloaded on the device. Android's repository for applications is Play Store while for iOS it is Apps Store. The study aims at understanding and summarizing the data obtained from Kaggle on Android applications and iOS applications in order to compare and predict the future of third party mobile applications.

## Problem Statement

The aim of this analysis will going to answer which type of mobile applications are successful on Play Store and Apps store. With millions of mobile applications available it is difficult for any new mobile app development company to analyse the market and determine the niche. Mobile applications will be analysed on the basis of Genre, Category, and number of reviews, average ratings and user sentiment of applications.

**Data Source**

Data set is taken from the Kaggle data source using the following link

Google Play Store Data: https://www.kaggle.com/lava18/google-play-store-apps

Apple Apps Store Data: https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps/version/2

**Variable Description**

*For Google Play Store Data*

- Application name
- Category of the app
- Overall user rating of the app
- Number of user reviews for the app
- Size of the app
- Number of user installs the app
- Type of App - Paid or Free
- Price of the app
- Content Rating of app - Children / Mature 21+ / Adult
- Genres of An app can belong to multiple genres

- Last Updated date on Play Store
- Current version of the app
- Minimum required Android version
- Name of application
- Translated User Reviews
- SentimentPositive/Negative/Neutral
- Sentiment_PolaritySentiment polarity score
- Sentiment_SubjectivitySentiment subjectivity score

*For Apple App Play Store*

- track_name – Name of the Application
- size_bytes – Size of the application
- currency – Currency in which price of App is mentioned
- price – Price of the Application
- rating_count_tot – Number of ratings given
- rating_count_ver – Ratings on the latest version of app
- user_rating – Average rating
- user_rating_ver – Average rating on latest version of app

- ver – Current version of App
- prime_genre – Primary Genre of App

**Analysis Expectations**

Android and iOS are the two most prominent and widely accepted operating systems for smart phones in the present era. Currently 2.5 billion Android and 800 million iOS users across the world. Thus, the type and category of applications that are being downloaded from these two platforms will allow us to understand the taste and mood of the people across the world.

**Dashboard 1 – Category Details – Google Play Store**

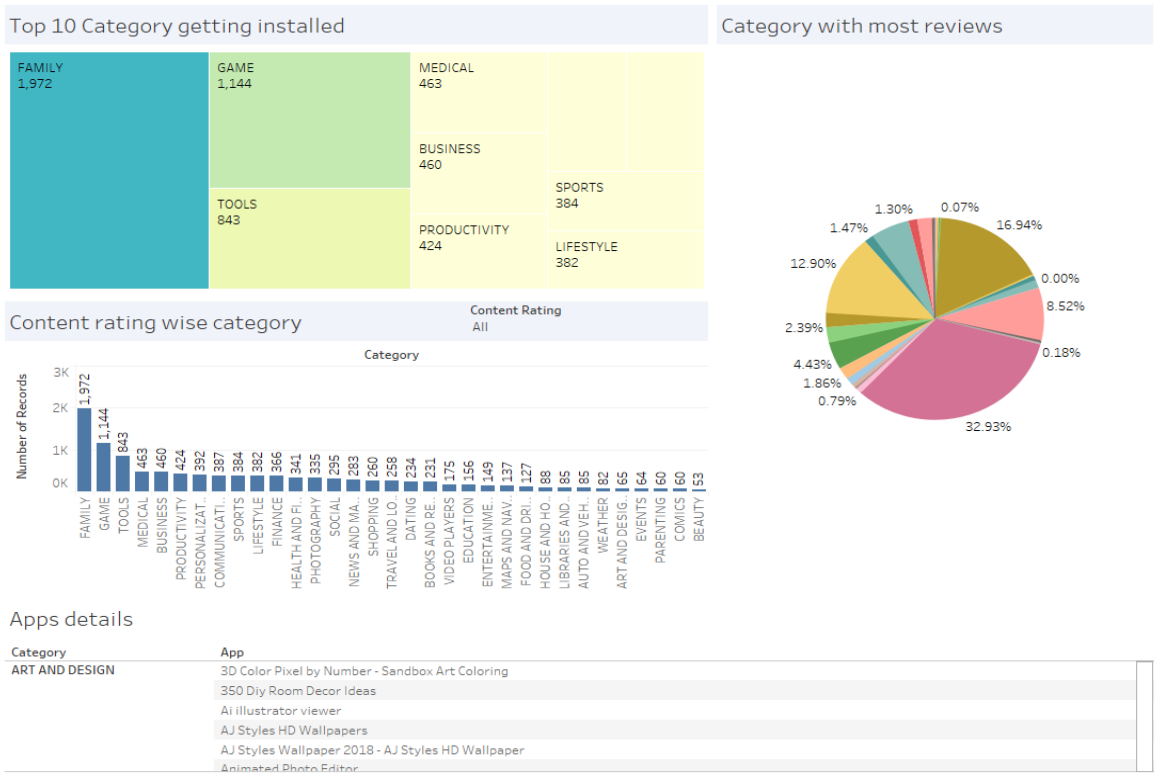



*Figure 1*

First dashboard gives the category details of the android apps that were downloaded from the play store. The top 10 categories downloaded are Family, Game, Tools, Medical, Business, Productivity, Personalization, Communication, Sports and Lifestyle in the decreasing order. Some of the least downloaded categories are comics, beauty and events to name. When it comes to category wise reviews then 33% of the reviews on Play Store are on Game category apps, 17% on the communication apps followed by social media apps and family apps. Some of the categories with least reviews are Health & Fitness and Navigation. As it can be seen that the top category and the second category has difference of nearly 40% compared to first category. This will be useful for a new app development company in order to determine the market share of each category. It can be used by the company to decide which new application they can target.

**Dashboard 2 – Genre Analysis – Google Play Store**

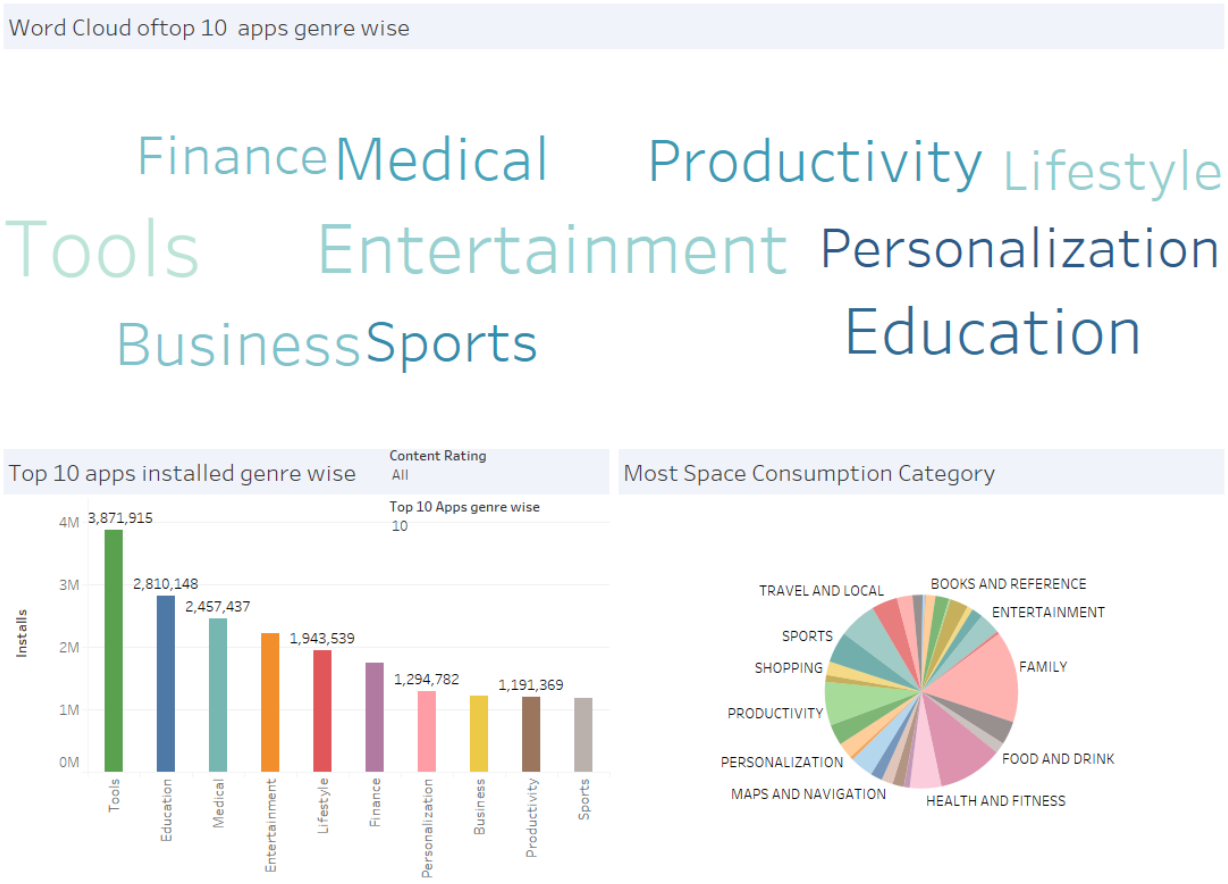



*Figure 2*

Second dashboard considers the genre of the android applications on Google Play Store. The word cloud shows that Entertainment, Education, Productivity, Tools and Lifestyles are some of the most common Genre which are being downloaded from the Play Store. Pie chart shows that most space consuming genre is family apps (15.31%) followed by health & fitness apps (5.29%) and productivity apps. The least space consuming apps are House & Home and Food & Drink. Since the tool based genre is one of the prominent segment, a software application development

company can decide to target this segment and create an app with some innovation to improve on the previously made tool based applications.

**Dashboard 3 – Apps Dashboard – Google Play Store**

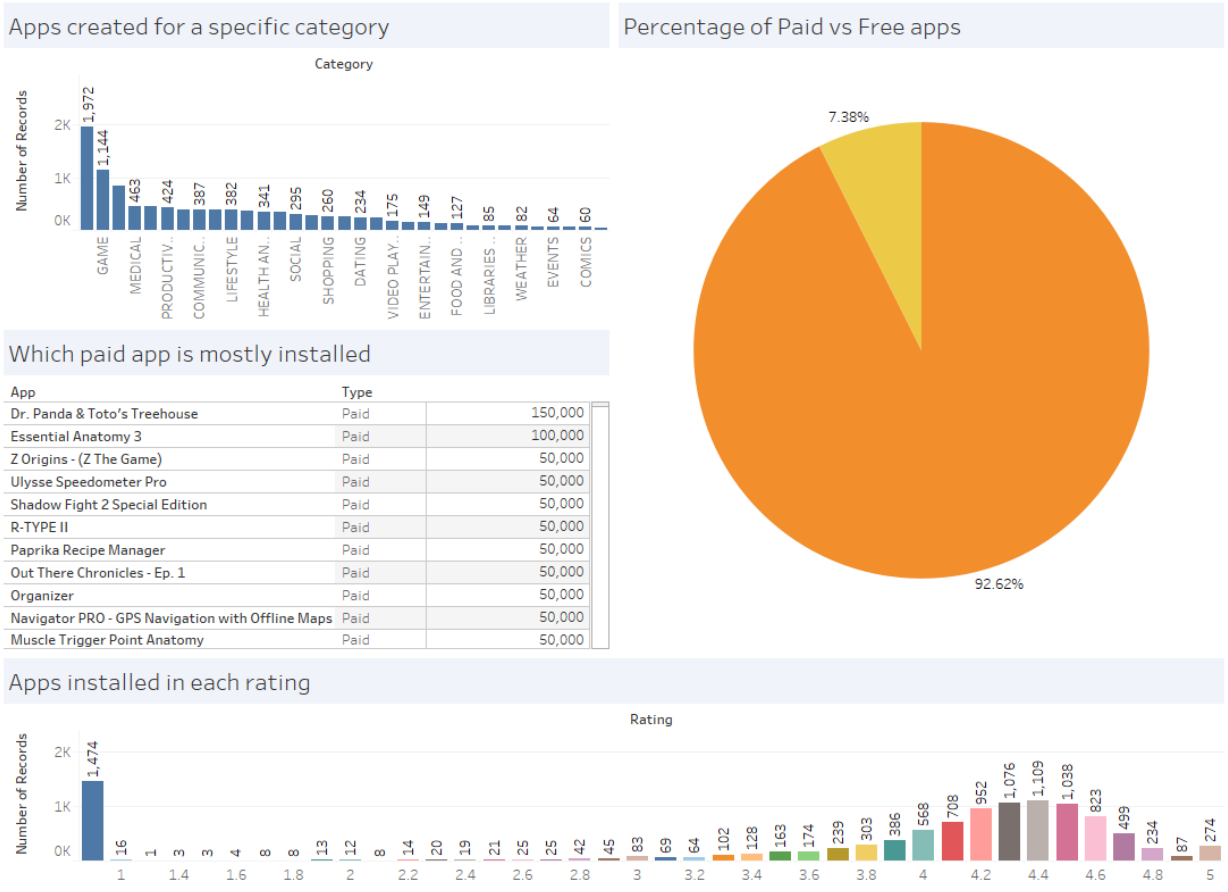



*Figure 3*

This dashboard shows that there are only 7.38% of paid apps while 92.62% are free apps available on Google Play Store. Out of the paid apps some of the most installed apps are Dr. Panda & Toto's Treehouse, Essential Anatomy 3 and Z Origins. As the distribution of rating is skewed to the right there is a clear indication that Play Store is well monitered and quality of

applications is kept high on the platform. This visualization can be used for the moderator of Play Store in order to identify the overall performance of the platform.

**Dashboard 4 - Top n Bottom apps – Google Play Store**

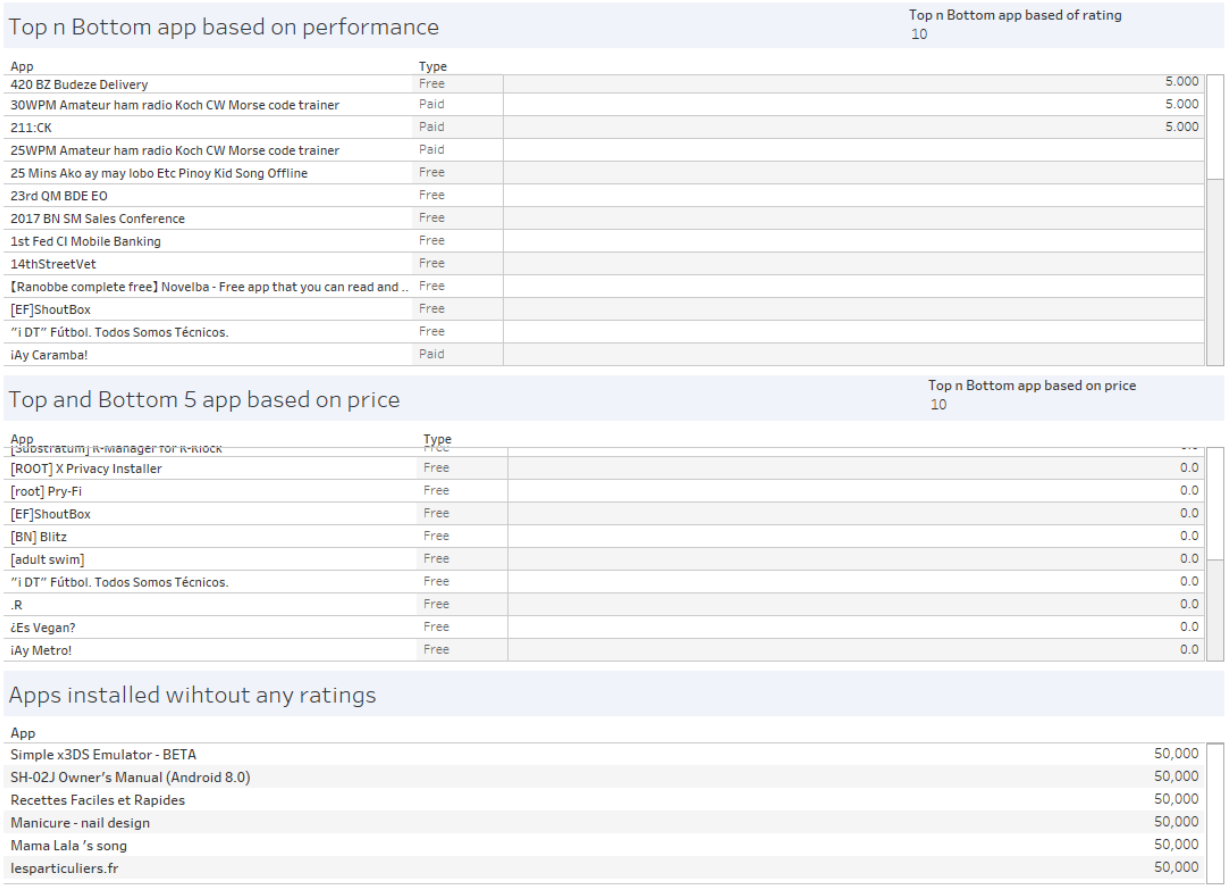



*Figure 4*

On the basis of performance, AJ blue Icon Pack, AI Today, CS & IT Interview Questions, Cric Quick as some of the prominent free applications while I'm rich - Trump Edition is some of the prominent paid applications. Nearly 7% of the applications are paid while others are free. This does not count for in app purchases because there are some applications mostly the gaming apps which are free to download but would have purchases inside the application.

**Dashboard - Rating Dashboard – Google Play Store**

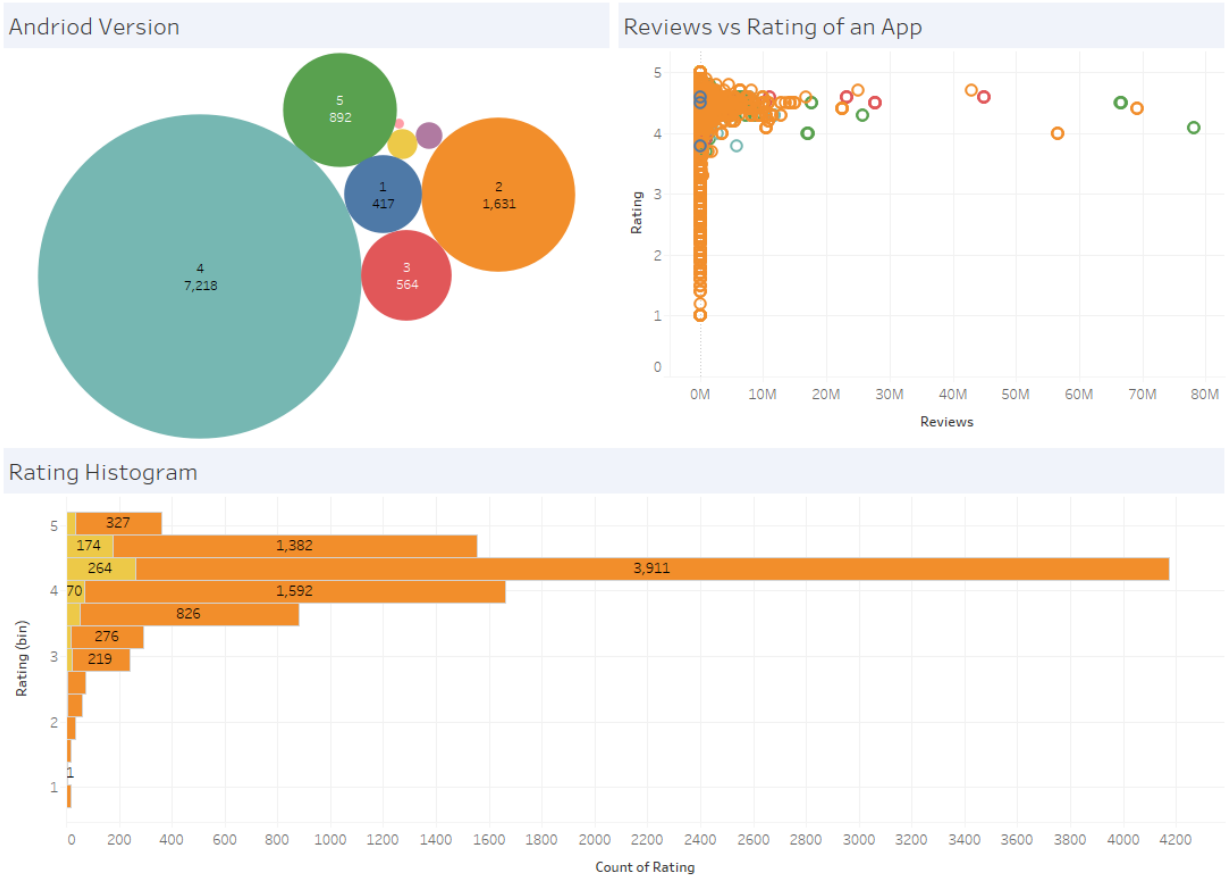



*Figure 5*

For the Android applications, most number of ratings were given after the launch of android version 4 compared to other versions of Android. This shows that Android version 4 is the most interactive version. From the scatterplot of rating vs Reviews, no clear trend is visible. Most of the people like to rate app but do not give any reviews. The distribution of rating is highly skewed to the left. Median rating of applications is close to 4. This can be used by monitoring group in order to determine the overall performance.

**Dashboard - Survey Dashboard – Google Play Store**


*Figure 6*

Analysing the sentiment of users in a survey shows that Bowmasters, Angry Birds, 8 Ball Pool and Calorie Counter are some applications with most positive sentiment while MyFitnessPal, Dead Target and Helix Jump are with most negative sentiment. It can also be seen that some applications are top with positive sentiment as well as the negative sentiment at the same time. This shows the level of popularity of the application and how different groups are formed.

**Dashboard – Category wise iOS Applications**

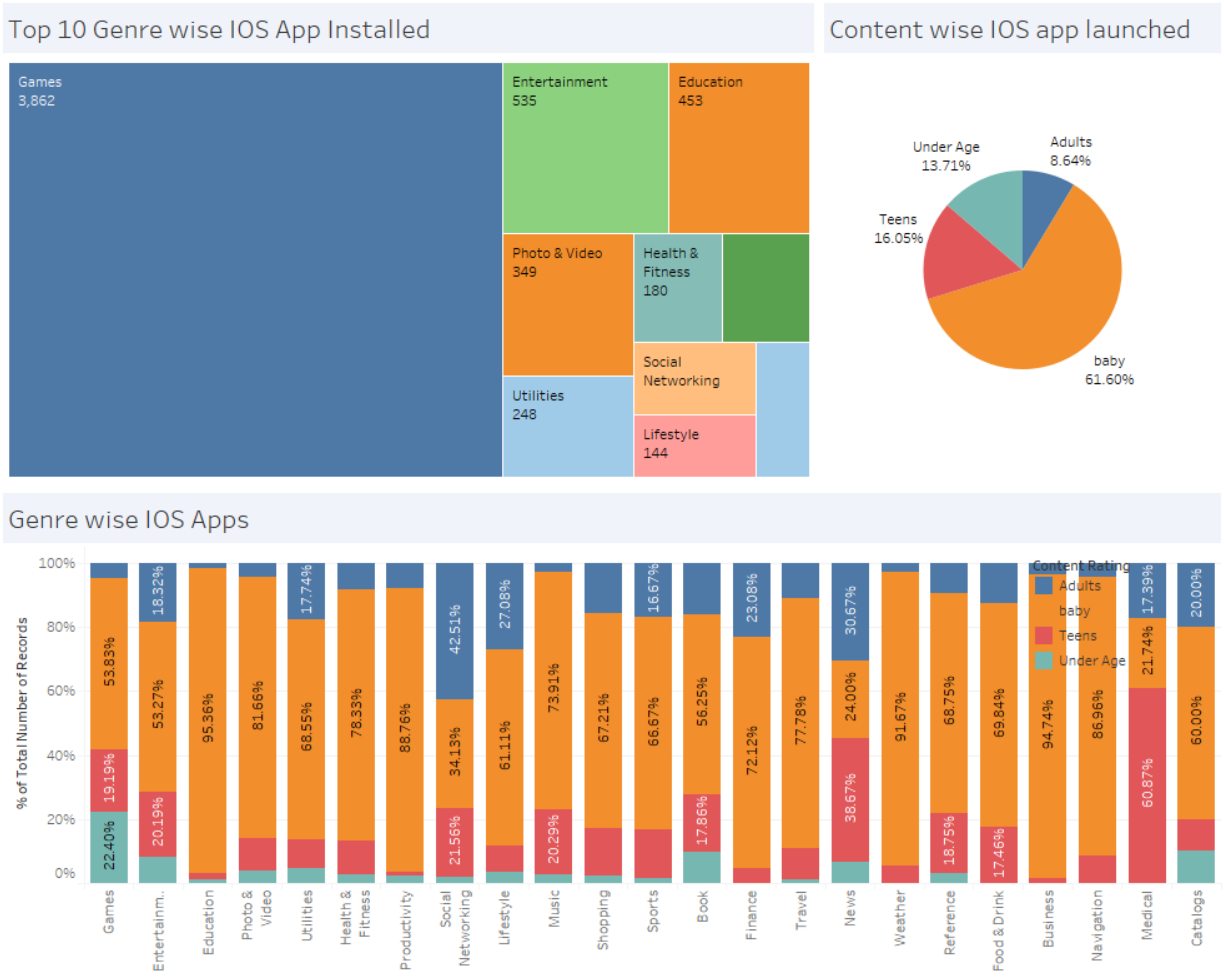



*Figure 7*

Category scenario for iOS applications is completely different from that of the android applications. The highest number of installations is for games followed by entertainment, photo & video, education, health & fitness, social networking, lifestyle and utilities. Content wise the app distribution in iOS shows that 61.6% of the applications were made for baby, 16% for teens while only 9% for adults. Considering the age and genre wise distribution, the game category has

nearly 4.58% of games for adults while 20% for teens and under age users. In case of education, nearly 95% of applications are meant for children. For news category around 76% of applications are for adults or teens.

**Dashboard - Iphone Apps Price and Rating**

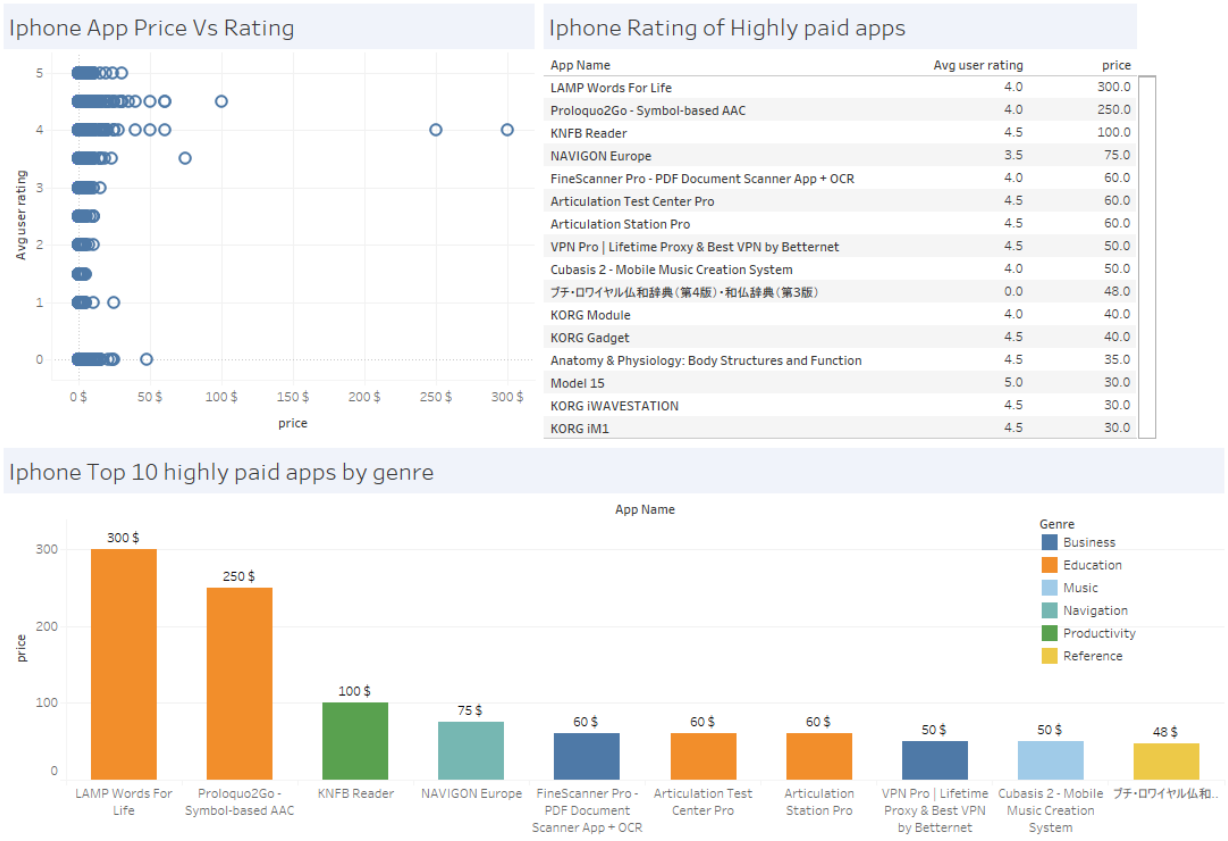



*Figure 8*

Comparison of iphone application price and rating shows that high paying apps generally have high ratings (i.e. above 4). As such no clear relation can be seen between price and rating. Some of the highly paid applications on App Store are LAMP Words for Life and Proloquo2Go-Symbol which belongs to education category. Average user rating for some of the highest paid

applications is between 4 and 4.5. This visual data is useful for the companies which create apps for apps store. It will help the company to decide the pricing of the application.

**Dashboard – I-phone app based on Genre and size**

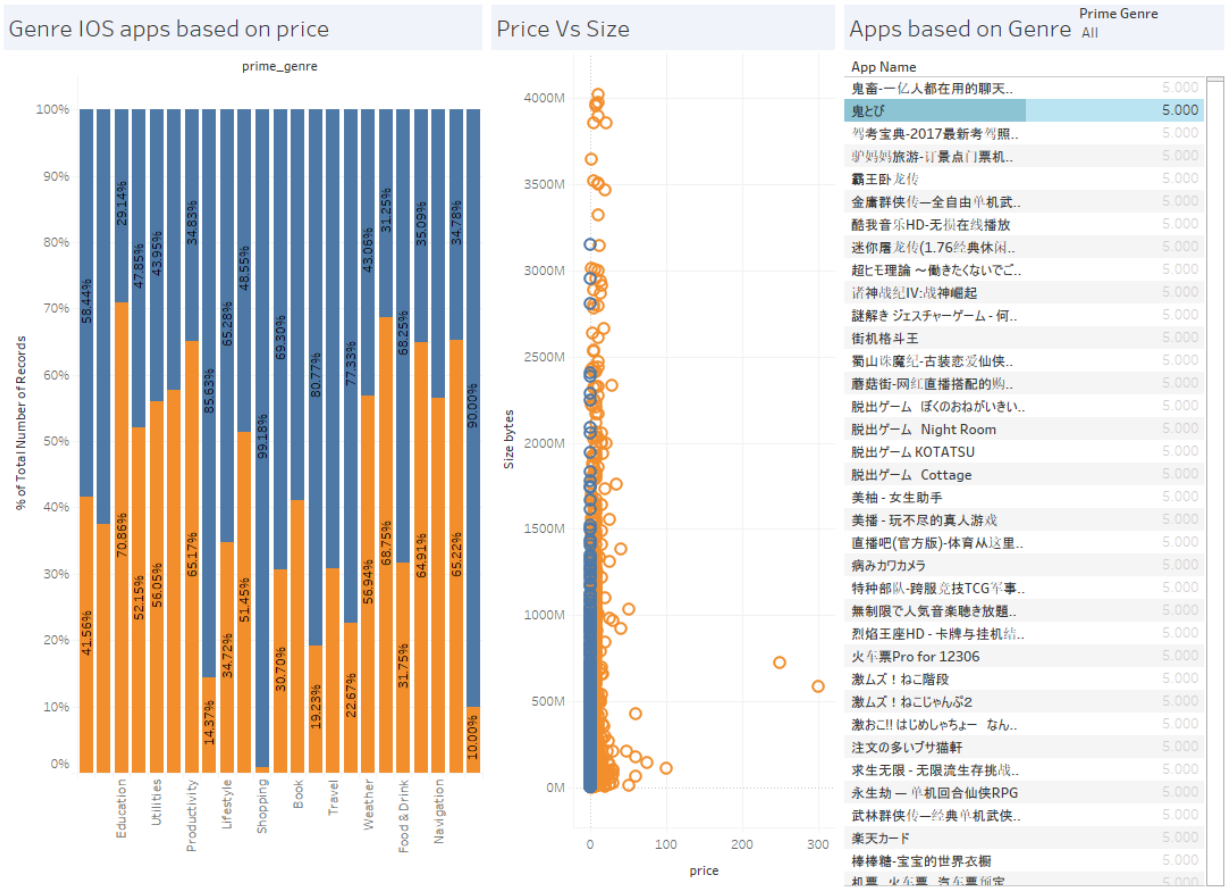



*Figure 9*

From the comparison of paid vs. free applications on App store, the game category has nearly equal share of free and paid applications, education has 70% of the paid applications, while entertainment has 62% of free applications. The category with maximum percentage of paid

applications is education and references. Comparing the price and size of the application there is no clear relationship exist. Applications with highest user ratings on App store are mostly the applications of Chinese origin.

**Dashboard - Relation between Supporting an app and user rating**

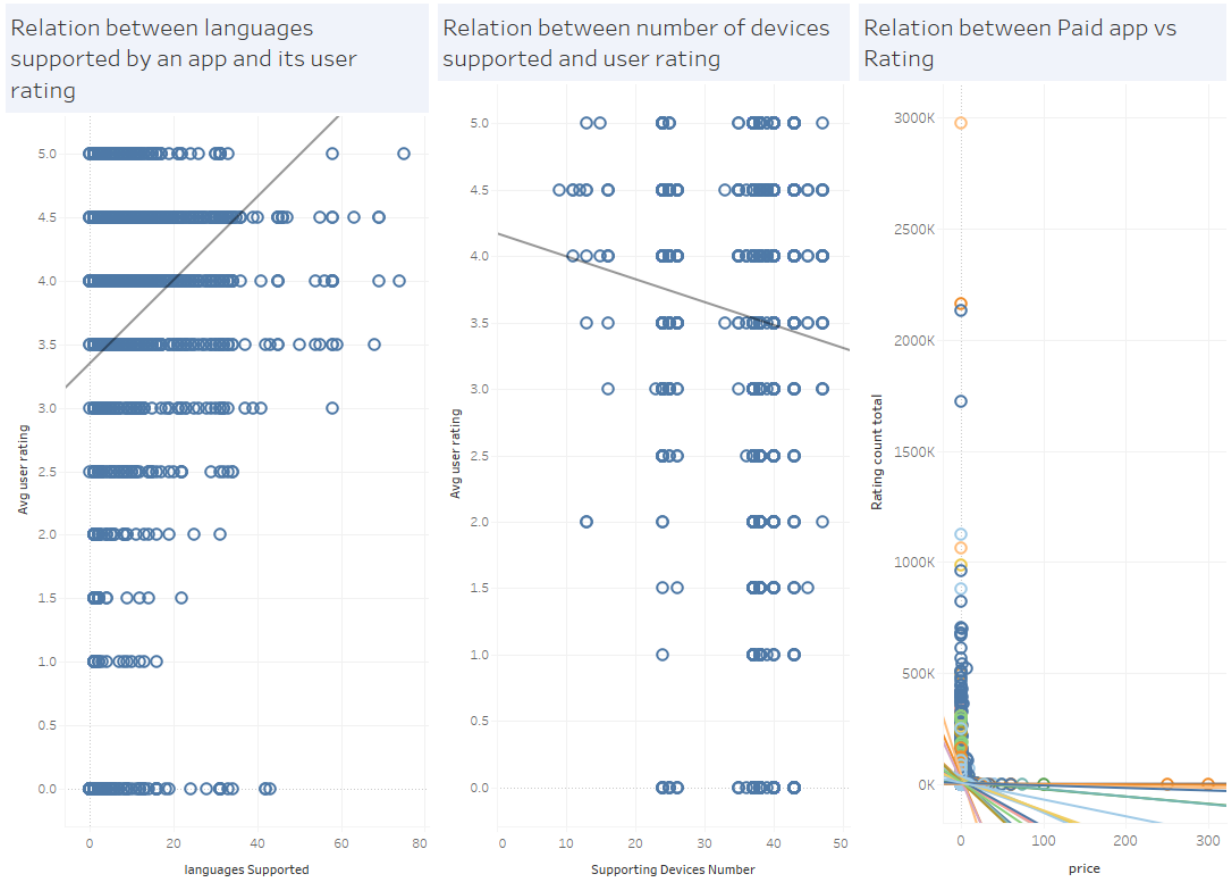



*Figure 10*

It can be clearly seen that relation between number of languages supported by an application and average user rating is passivity related. Relation between number of devices supported and average user rating is negatively related while relation between cost of app and average rating is negatively related on the Apps store. This shows that when an application is made for many

types of devices the performance of application reduces and user rating goes down. This visual data is useful for the technical team of app Development Company while making application so that the architecture of app is designed for specific devices. Also, the language play an important role in the rating as the number of languages are more, ratings will increase. This factor contributes to the psychological behaviour.

**Dashboard - Trending IOS Apps and Games**

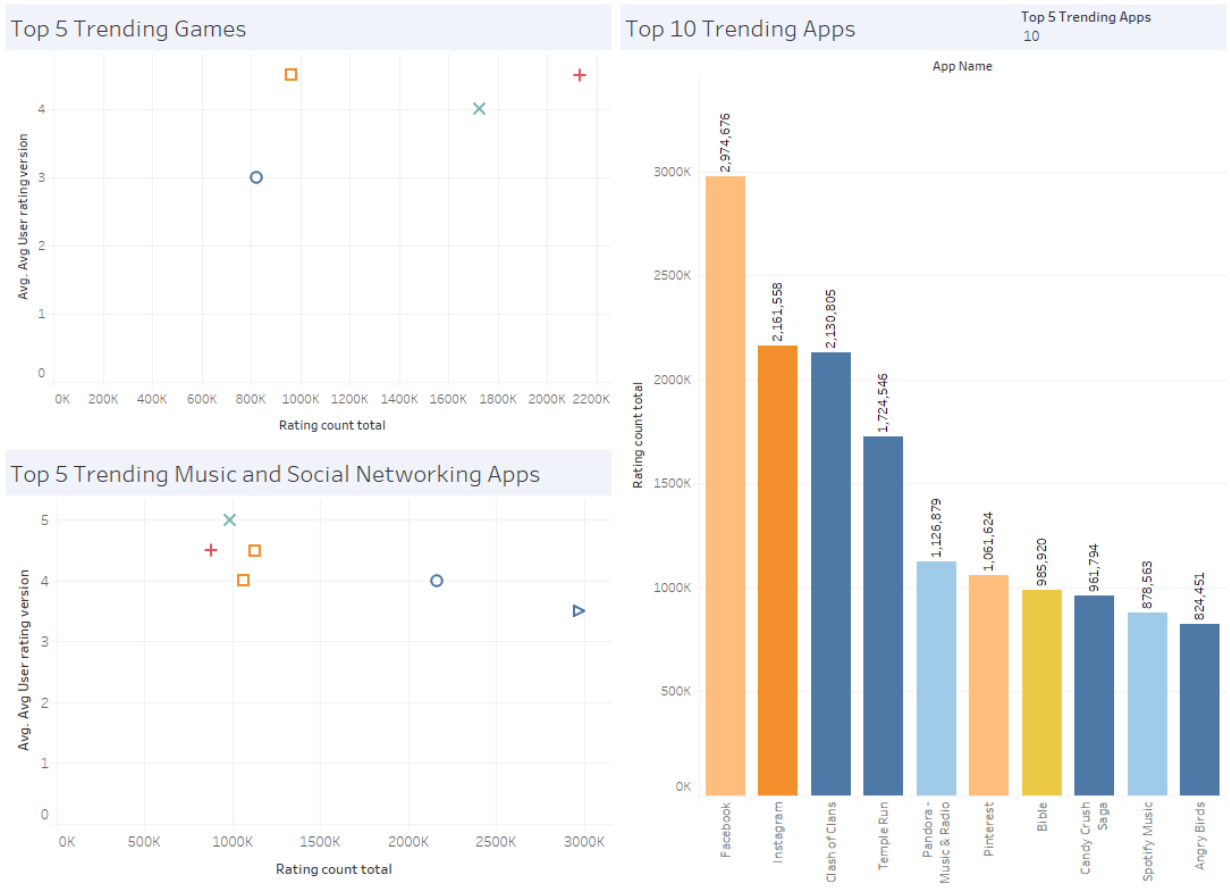



*Figure 11*

The data shows that Candy crunch, temple run, clash of clans and angry birds are some of the most trending applications on Apps store, followed by Pandora - Music and Radio. There is a

negative trend visible between the number of ratings and average rating of an application. It can be seen that the Facebook apps is the most trending and used app on Apps Store followed by Instagram. This means the Apps store users are more focused on downloading social media applications compared to other types of applications.

**Dashboard - Trending app in term of Size and Price for both IOS and Android**

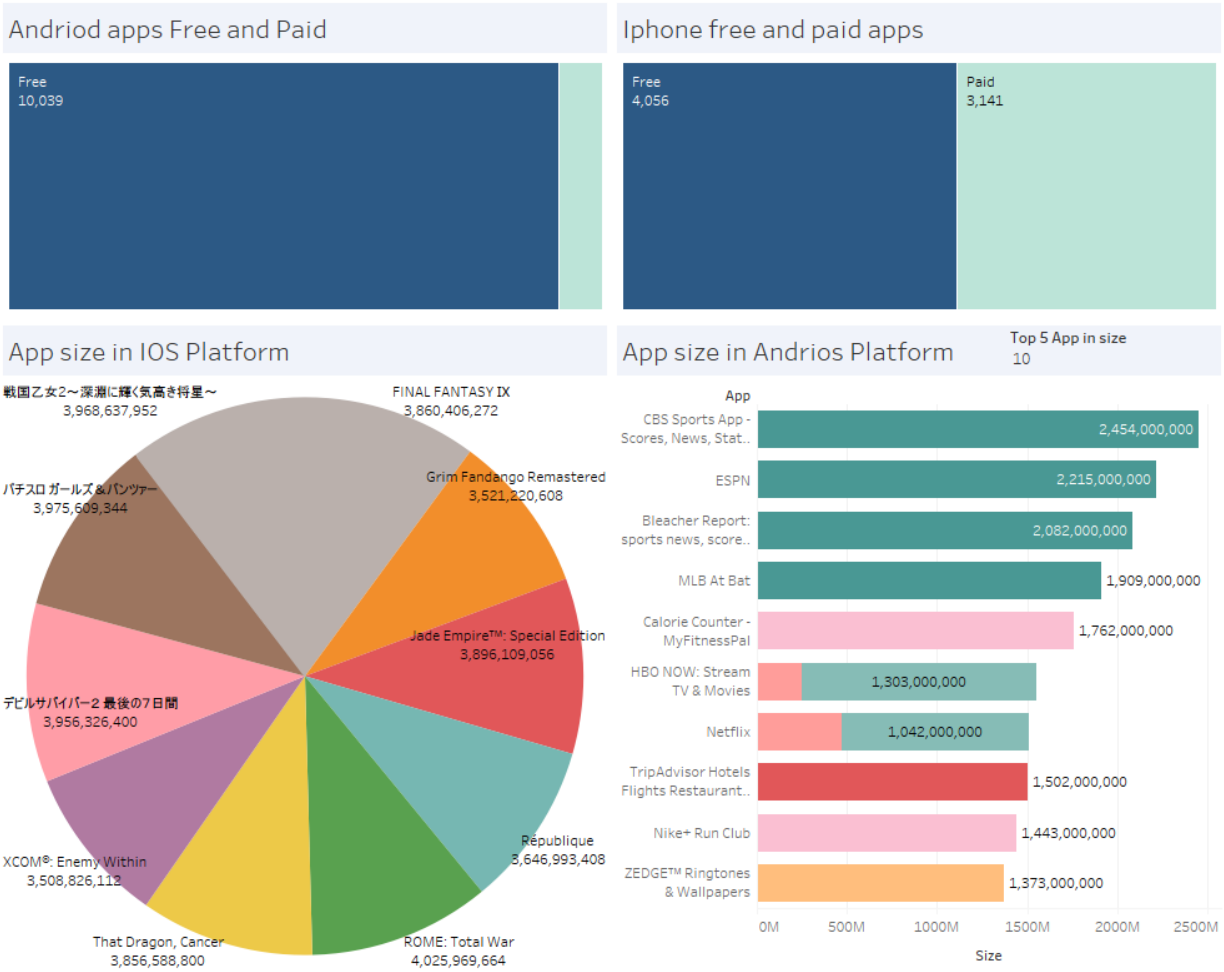



*Figure 12*

Comparative analysis shows that 92.62% of the apps on Android are free while 56.36% of apps on apps store are free. This means there is a huge difference in the ratio of free and paid apps on Google play store and apple apps store respectively. Comparing the size of applications the iOS apps are heavier than android applications. More number of free applications on Play Store suggest the class of people that Play Store App creator focus on. This visual will be useful for strategic decisions while deciding a new application.

**Dashboard - App in different Content for both Market**

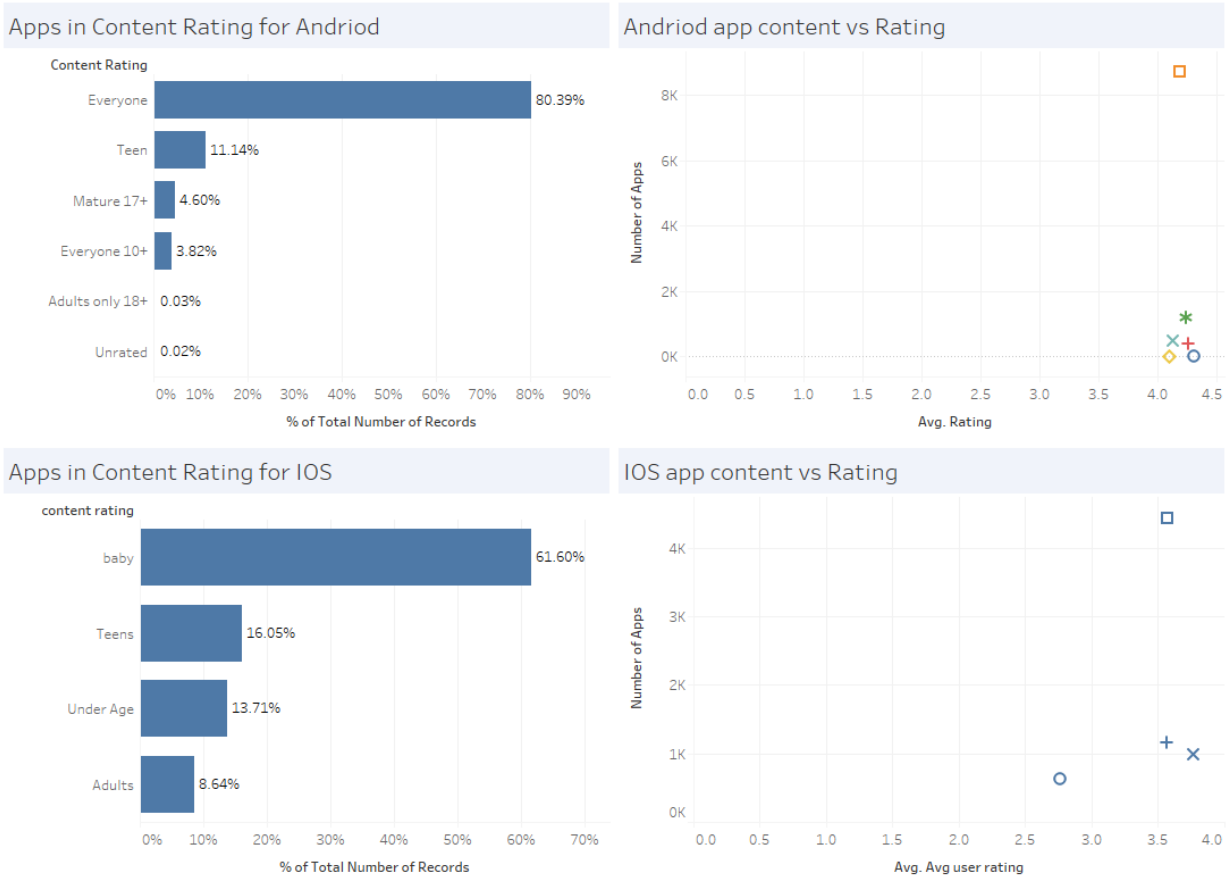



*Figure 13*

The average rating of content on android apps shows that nearly 80% of the content is such that everyone can consume, 11% is for teens while only 0.03% of the content is for Adults. On the contrary, iOS apps have nearly 8.64% of the adult contents. A clear difference in the rating can be seen in iOS and Android applications. For all categories of Android apps the average rating is above 4 while for iOS applications for under age applications is between 2.5 and 3. This means that the iOS app creators are not much focused on creating under age apps. Considering the scatterplot it can be seen that iOS user age group is of higher age levels compared to that of Android users.

**Dashboard - Trend of frequently updating of apps**

*Figure 14*

Considering the time of regular updating by app creators, since 2011 there has been an exponential increase in the number of applications being updated regularly. In 2011 there were only 15 apps which were regularly updated by the creators which has now risen to 7,349. It is predicted that by 2021 nearly 10,105 applications will be on a regular updating. Considering the previous trends in technology it can be said that technology grows in an exponential trend it is more likely that the trend will be exponential this time as well.

**Dashboard – Category wise Variations**

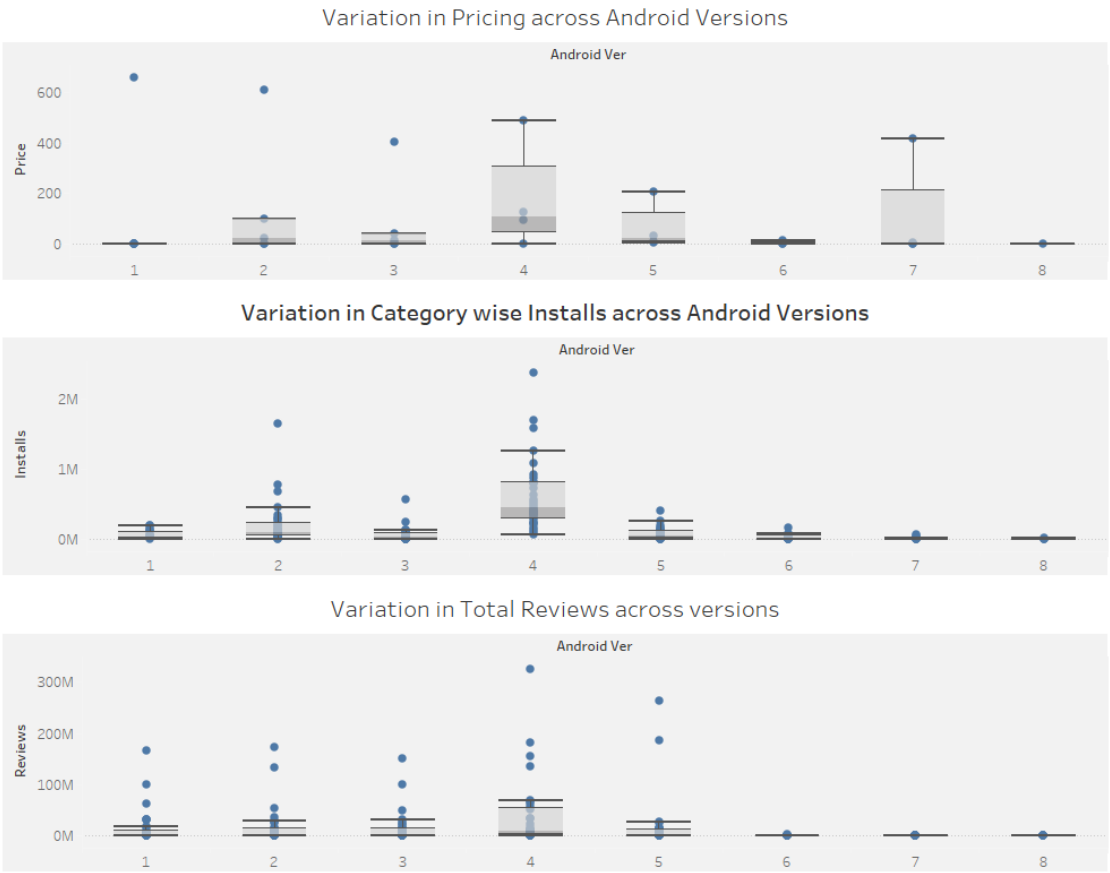



*Figure 15*

There is a huge difference in the pricing of android applications across versions of android. Version 4 and version 7 shows highest amount of variations. Similarly, when it comes to number of installations there is a large variations in categories for version 4 following the same trend the variations seems to be highest in version 4 for total reviews across categories. This shows that version 4 of Android has some major changes in the technology.

**Predictive Analytics**

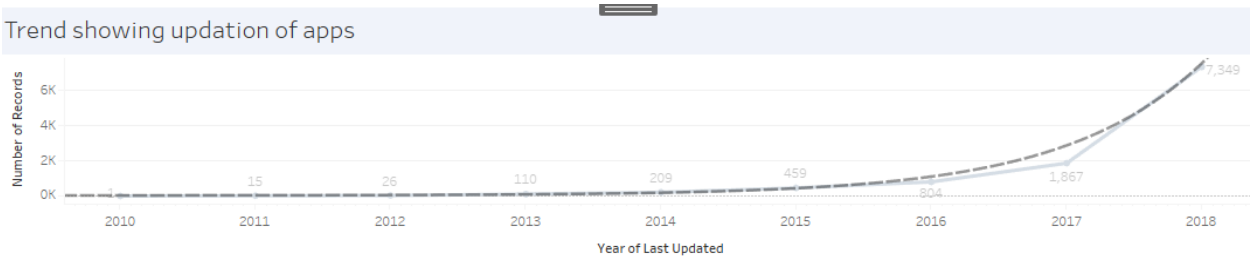



*Figure 16*

The time series in figure 16 shows the trend of updation made by app creators on Google Play Store. The trend clearly shows an exponential increase in the active app creators since 2016. Before 2016 the growth of active creators was low. But with increasing competition the trend becomes exponential and large number of creators started updating the apps on regular basis.

Determining the predictive model by applying curve fitting.

1) Considering the linear trend line

| | |
|---|---|
| **P-value:** | 0.0333592 |
| **Equation :** | Number of Records = 1.68163*Year of Last Updated + -68818.3 |

**Coefficients**

| Term | Value | StdErr | t-value | p-value |
|---|---|---|---|---|
| Year of Last Updated | 1.68163 | 0.63666 | 2.64132 | 0.0333592 |
| intercept | -68818.3 | 26517.2 | -2.59523 | 0.0356722 |

Considering the linear trend model shows that a linear trend model would be a good fit as p-value is less than 5%.

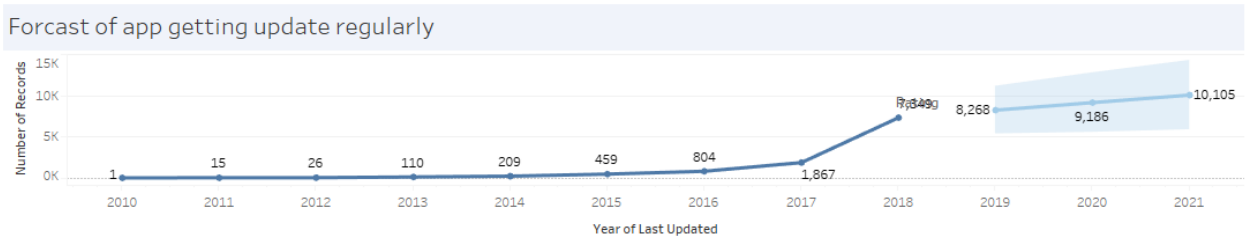



*Figure 17: Forecasting using the linear trend model*

Forecasting using the linear model the active app creators out of the sample considered in this analysis would be 9,186 in 2020 and 10,105 in 2021.

## Forecasting Statistics

### Options Used to Create Forecasts

| | |
|---|---|
| **Time series:** | Year of Last Updated |
| **Measures:** | Sum of Number of Records |
| **Forecast forward:** | 3 years (2019 – 2021) |
| **Forecast based on:** | 2010 – 2018 |
| **Ignore last:** | No periods ignored |
| **Seasonal pattern:** | 1 year cycle |

### Sum of Number of Records

| Initial | Change From Initial | Seasonal Effect | | Contribution | | |
|---|---|---|---|---|---|---|
| 2019 | 2019 – 2021 | High | Low | Trend | Season | Quality |
| 8,268 ± 2,902 | 1,837 | 2021 0 | 2021 0 | 100.0% | 0.0% | Poor |

**Forecasting Model**

All forecasts were computed using exponential smoothing.

**Sum of Number of Records**

| Model | | | Quality Metrics | | | | | Smoothing Coefficients | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Level | Trend | Season | RMSE | MAE | MASE | MAPE | AIC | Alpha | Beta | Gamma |
| Additive | Additive | Additive | 1,764 | 1,250 | 1.36 | 21,628.7% | 147 | 0.500 | 0.000 | 0.500 |

2) Considering the exponential trend line

**P-value:** < 0.0001

**Equation:** ln(Number of Records) = 0.00266348*Year of Last Updated + -105.878

**Coefficients**

| Term | Value | StdErr | t-value | p-value |
|---|---|---|---|---|
| Year of Last Updated | 0.0026635 | 0.0002098 | 12.6968 | < 0.0001 |
| intercept | -105.878 | 8.73727 | -12.1179 | < 0.0001 |

The exponential model shows that the exponential trend is a good fit since the p-value is below 5% level of significance. Thus, the exponential model can forecast the number of active app creators for year above 2018.

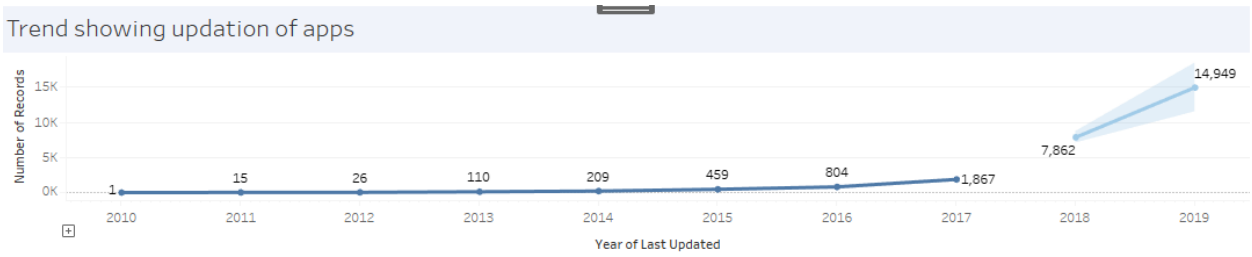



*Figure 18: Exponential Forecast*

Exponential Forecast model shows that in 2018 the active creators would be 7,862 and in 2019 it will be 14,949.

**Prediction of rating for various Content type of App on Play Store based on the number of reviews**

**Trend Lines Model**

A linear trend model is computed for Rating given Reviews. The model may be significant at p <= 0.05. The factor Content Rating may be significant at p <= 0.05.

| | |
|---|---|
| **Model formula:** | Content Rating*( Reviews + intercept ) |
| **Number of modeled observations:** | 9366 |
| **Number of filtered observations:** | 0 |
| **Model degrees of freedom:** | 11 |
| **Residual degrees of freedom (DF):** | 9355 |
| **SSE (sum squared error):** | 2466.34 |
| **MSE (mean squared error):** | 0.263639 |
| **R-Squared:** | 0.007886 |
| **Standard error:** | 0.513458 |
| **p-value (significance):** | < 0.0001 |

**Analysis of Variance:**

| Field | DF | SSE | MSE | F | p-value |
|---|---|---|---|---|---|
| **Content Rating** | 9 | 8.0613272 | 0.895703 | 3.39746 | 0.000355 1 |

**Individual trend lines:**

| Panes | | Color | Line | | Coefficients | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Row** | **Column** | **Content Rating** | **p-value** | **DF** | **Term** | **Value** | **StdErr** | **t-value** | **p-value** |
| Rating | Reviews | Unrated | N/A | 0 | Reviews | 0 | Because the trend line model response variable is constant, there is no information to estimate model statistics. | | |
| | | | | | intercept | 4.1 | | | |
| Rating | Reviews | Teen | 0.0279751 | 1082 | Reviews | 4.648e-09 | 2.112e-09 | 2.2006 | 0.0279751 |
| | | | | | intercept | 4.22864 | 0.0120758 | 350.174 | < 0.0001 |
| Rating | Reviews | Mature 17+ | 0.002956 | 459 | Reviews | 4.88e-08 | 1.633e-08 | 2.98828 | 0.002956 |
| | | | | | intercept | 4.10255 | 0.0243774 | 168.293 | < 0.0001 |
| Rating | Reviews | Everyone 10+ | < 0.0001 | 395 | Reviews | 1.224e-08 | 2.949e-09 | 4.14906 | < 0.0001 |
| | | | | | intercept | 4.2361 | 0.0187669 | 225.722 | < 0.0001 |
| Rating | Reviews | Everyone | < 0.0001 | 7418 | Reviews | 1.406e-08 | 2.634e-09 | 5.33614 | < 0.0001 |
| | | | | | intercept | 4.18107 | 0.0063059 | 663.041 | < 0.0001 |
| Rating | Reviews | Adults only 18+ | 0.48645 | 1 | Reviews | 1.463e-05 | 1.402e-05 | 1.0435 | 0.48645 |
| | | | | | intercept | 3.90336 | 0.452901 | 8.61856 | 0.0735373 |

Clearly from the regression models the application rating depends on the number of reviews as the regression model is significant at 5%. The sign of estimated coefficient shows that as the number of reviews increases for any category the rating will also increase.

**Prediction of Average rating of Apps on Apple Apps Store based on the price of App**

**P-value:** < 0.0001

**Equation:** Avg user rating = 0.0121272*price + 3.50602

**Coefficients**

| Term | Value | StdErr | t-value | p-value |
|---|---|---|---|---|
| price | 0.0121272 | 0.0030646 | 3.95715 | < 0.0001 |
| intercept | 3.50602 | 0.0186411 | 188.08 | < 0.0001 |

The regression model developed shows

$$Rating = 3.506 + 0.012*(price)$$

The model developed is significant and shows that as the price increases the rating also increases.

**Prediction of Size based on the Price of Application on Apple Apps store**

| | |
|---|---|
| **P-value:** | < 0.0001 |
| **Equation:** | Size bytes = 8.88188e+06*price + 2.30117e+08 |

**Coefficients**

| Term | Value | StdErr | t-value | p-value |
|---|---|---|---|---|
| price | 8.88188e+06 | 1.02064e+06 | 8.7022 | < 0.0001 |
| intercept | 2.30117e+08 | 9.39749e+06 | 24.487 | < 0.0001 |

$$Size = 2.30*10^8 + 8.882*10^6*(Price)$$

Clearly the size of application increases as the price increases.

**Cluster Analysis**

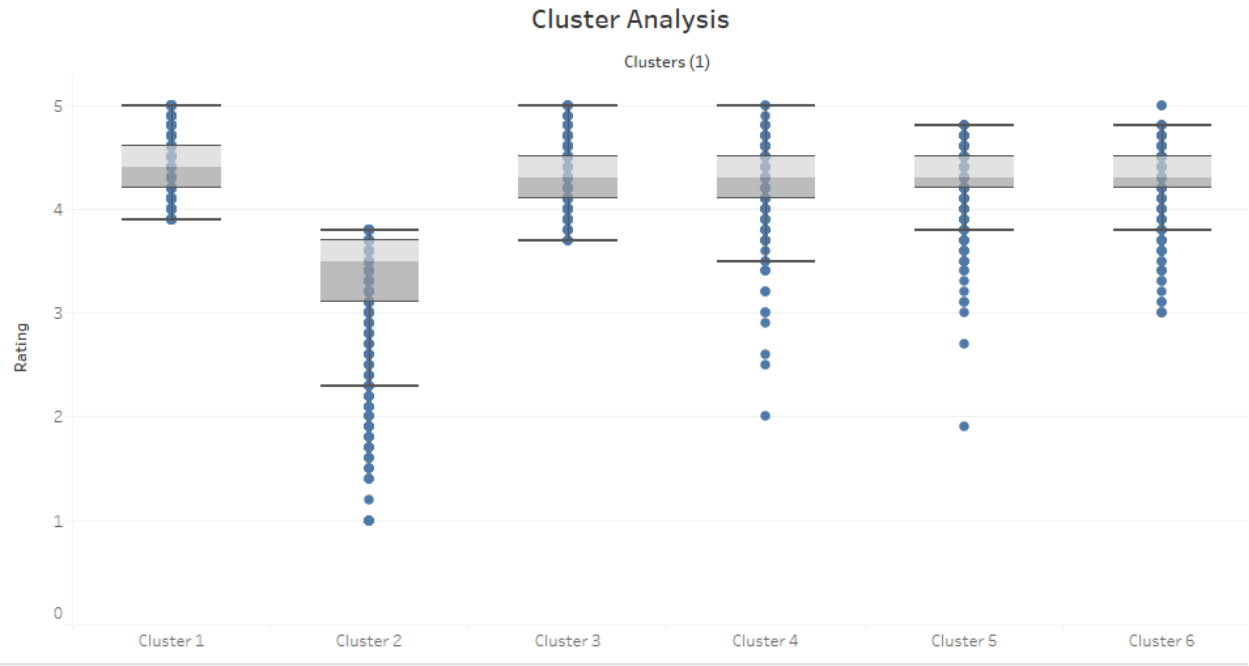



*Figure 19*

## Inputs for Clustering

| | |
|---|---|
| **Variables:** | Avg. Rating<br>Avg. Size<br>Sum of Reviews<br>Category<br>Android Ver<br>Genres |
| **Level of Detail:** | Not Aggregated |
| **Scaling:** | Normalized |

## Summary Diagnostics

| | |
|---|---|
| **Number of Clusters:** | 6 |
| **Number of Points:** | 9366 |

| | |
|---|---|
| **Between-group Sum of Squares:** | 467.49 |
| **Within-group Sum of Squares:** | 136.28 |
| **Total Sum of Squares:** | 603.76 |

| | | Centers | | | Most Common | | |
|---|---|---|---|---|---|---|---|
| **Clusters** | **Number of Items** | **Avg. Rating** | **Avg. Size** | **Sum of Reviews** | **Category** | **Android Ver** | **Genres** |
| **Cluster 1** | 5306 | 4.3881 | 1.161e+07 | 2.7504e+05 | FAMILY | 4 | Tools |
| **Cluster 2** | 1454 | 3.2909 | 1.7235e+07 | 11000.0 | FAMILY | 4 | Tools |
| **Cluster 3** | 1513 | 4.3157 | 6.4223e+07 | 1.449e+06 | GAME | 4 | Action |
| **Cluster 4** | 372 | 4.2618 | 3.0923e+08 | 8.7462e+05 | FAMILY | 4 | Productivity |
| **Cluster 5** | 362 | 4.2644 | 4.3357e+08 | 1.2993e+06 | PRODUCTIVITY | 4 | Productivity |
| **Cluster 6** | 359 | 4.2708 | 1.8827e+08 | 9.7817e+05 | TOOLS | 4 | Tools |
| **Not Clustered** | 1474 | | | | | | |

Clusters have been created based on the average rating, average size, and sum of reviews, category, android version and gender. The clusters are statistically significant and shows that cluster 2 has the highest variation within the cluster.

**Conclusion**

From the above analysis it is difficult to say which ecosystem is better than the other because both have some advantage and some disadvantage. Play Store is advantageous for the users while Apps store is beneficial for App creators in short. In terms of speed the approval of applications is fast on Play Store which may lead to some low quality apps on the platform but in past few years Google has applied strict measures to keep a check on this issue.