

Google Play Store Analysis

CMSC6950

Computer Based Research Tools and Applications

Khosh Raftar Nouri, Aida (akhoshraftar)
Salcedo, Carlos (cdsalcedo) Sarker, Sourav (souravs)
Yazdanpanah, Fatemeh (fyazdanpanah)

June 2019

1. Introduction to data

While many public datasets (on Kaggle and the like) provide Apple App Store data, there are not many counterpart datasets available for Google Play Store apps anywhere on the web. On digging deeper, it can be found out that iTunes App Store page deploys a nicely indexed appendix-like structure to allow for simple and easy web scraping. On the other hand, Google Play Store uses sophisticated modern-day techniques (like dynamic page load) using JQuery making scraping more challenging. 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 category, rating, reviews, size, installs, price, rated, last updated, and version.

2. Objectives of Google-Play Store Analysis

This project focuses on the analysis of the Play Store data set in Kaggle. The aim of this project is:

- Using the data to analyze consumer trends and determine which type of apps are the most popular and profitable.
- Classifying applications based on their categories.
- Presenting the growth of applications from 2016 to 2018.

- Comparing different categories of applications based on the Android version.
- Comparing the rates in different kinds of applications.
- Assessing supported Android version with numbers of reviews based on different categories.

3. Top 20 Applications categories Based on the Application Numbers

In our dataset we have more than 9000 applications. The pie chart (Figure 1) represents top 20 application categories based on application number. The top number of application is belonged to family. The following high number of applications are related to Game, Tools, and Business respectively. The result indicates that more users are interested in the four top categories and most of the developers put more emphasis on improving them.

- Family is included of 1832 numbers of applications.
- Game is included of 959 numbers of applications.
- Tools is included of 827 numbers of applications.
- Business is included of 420 numbers of applications.
- Medical is included of 395 numbers of applications.

This result shows that the major numbers of applications is related to Family.

4. Growth of Application Numbers from 2016-2018

According to bar chart, (Figure 2), we can see the growth of application numbers which are categorized in different years. Most of the applications have grown in all three years from 2016-2018. Family reached to the top of growth around 200, 400, and 800 in 2016, 2017, and 2018 respectively. The following highest application number in all three years are belonged to Tools and Business.

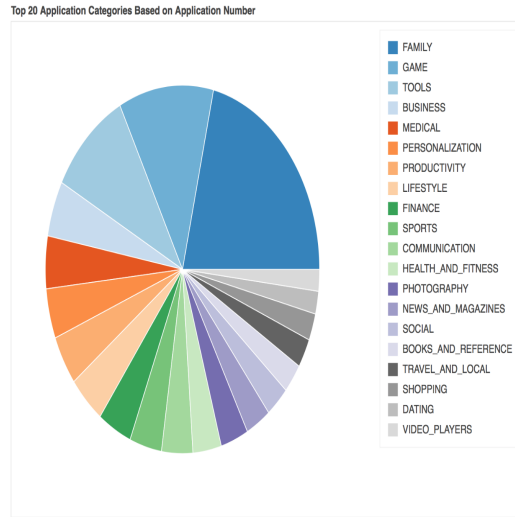


Figure 1: Top 20 Application Categories Based on Application Number

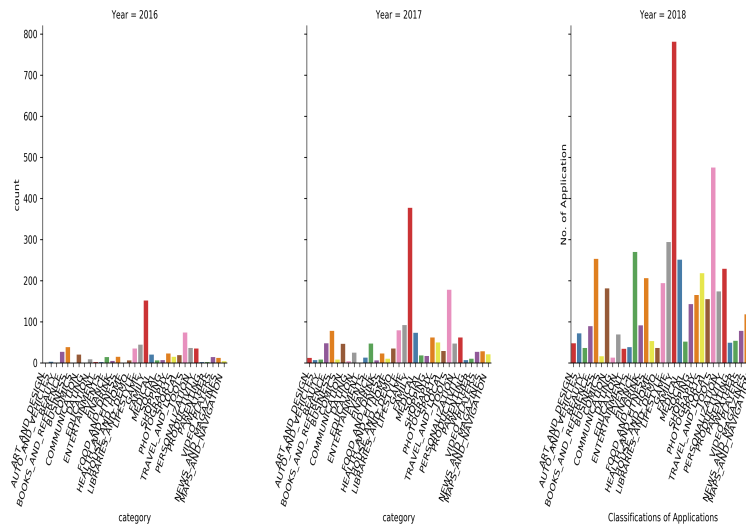


Figure 2: Growth of Application Numbers from 2016-2018

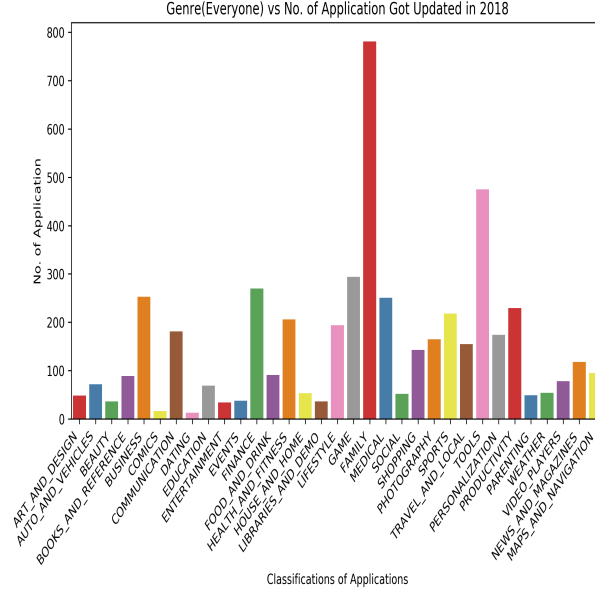


Figure 3: Genre (Everyone) Vs. Number of Application Got Updated in 2018

5. Application Update

5.1 Number of Application Update

From the survey carried out, Family, Tools, and Business have the most updated numbers of applications which reaches around 800 in 2018. While, Dating and Comics have less updated compared to others. (Figure 3)

5.2 Different Months of Application Update

This figure (Figure 4) represents the update of application based on the differences between months and years. The highest rate of update is belonged to month of July in 2018. As shown in figure 4, in 2018, there is a great trend of update compared to other years.

6. Android Version

6.1 Supported Android Version with No. Reviews Based on Game

Figure 5 describes the supported Android version with number of reviews based on game.

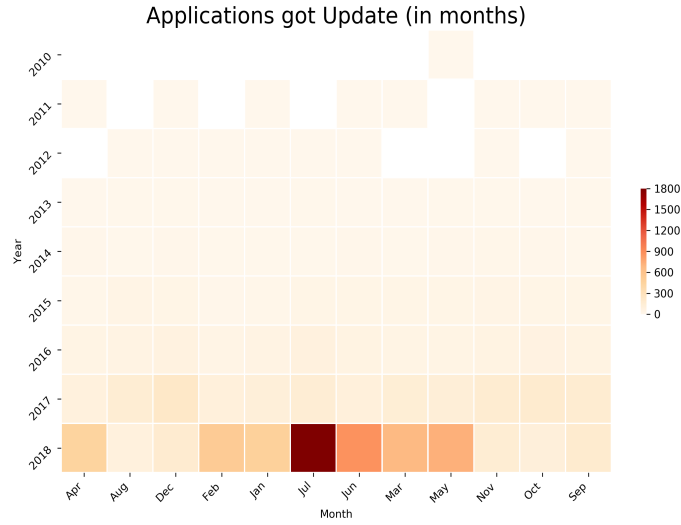


Figure 4: Application Got Update (in Month)

- The majority number of reviews are belonged to version of 4.1, 4.0.3, 2.3, 4.0, 4.4 and up.
- Among Everyone (orange dot) and Teen (red dot), 4.1,4.0.3, and 2.3 are seen significantly.
- Some versions which are less than two have no review as shown in the figure.

6.2 Category Vs. Number of Applications Based on Android Versions

As we can see in figure 6, there is a relationship between application classification and version of Android. The version of 4.1 and up has the highest usage in Family applications which is reached to above 400. Additionally, it can be seen that the Family category has the most diversity of version compare with other categories. Version 5.0-6.0, 5.0-7.1.1, and 5.0-8.0 has their lowest usage among all applications.

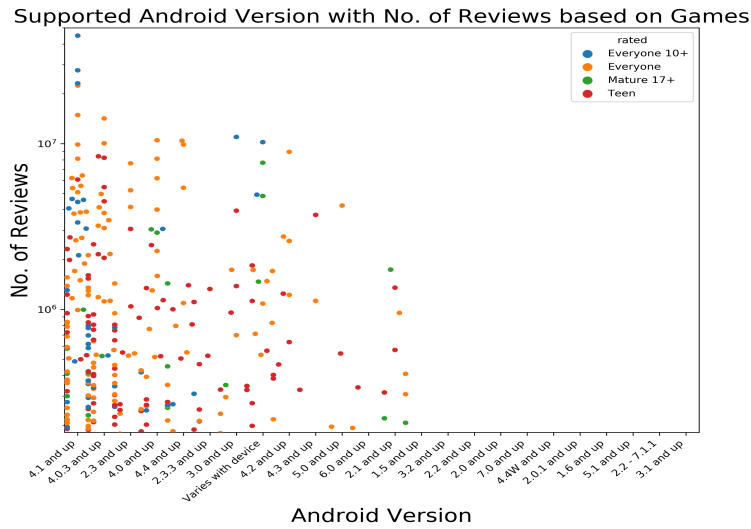


Figure 5: Supported Android Version with No. reviews based on Game

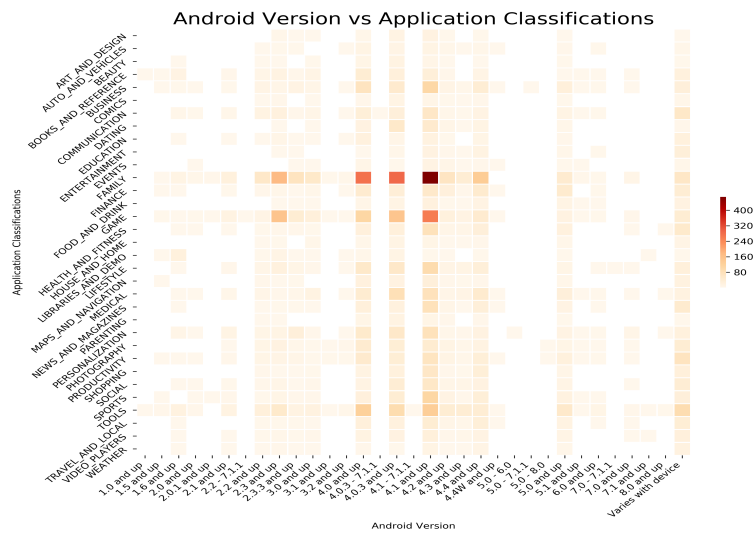


Figure 6: Category Vs. No. of Apps Based on Android Versions

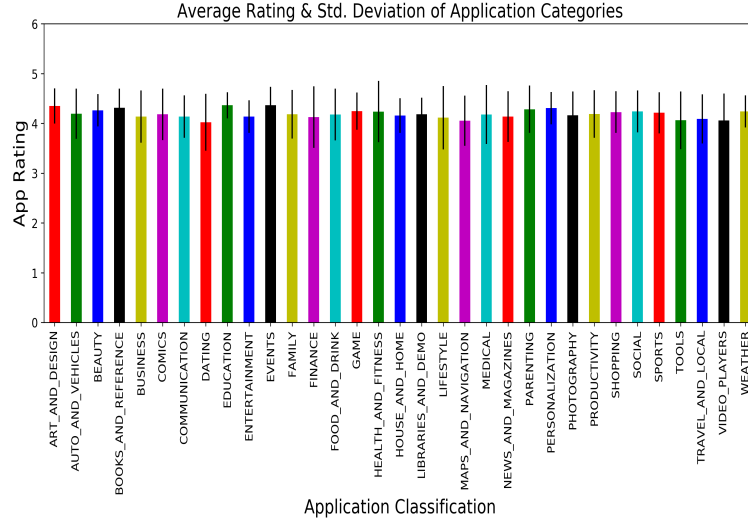


Figure 7: Average Rating and Std. Devision of Application Categories

7. Application Rating

7.1 Average Rating and Std. Devision of Application Categories

Figure 7 indicates the average rating of application categories. This figure is trying to compare rating of different applications. In fact all the applications have the same range of rating above 4. The rating of all application are so close to each other in all categories.

7.2 Rating in each year

From 2016 to 2018, the rate of applications is around 4 to 5. Also, it is axiomatic that it has increased slightly in all these three years and there is a subtle difference among them. In 2018, it started from 4.075 and it reached to nearly 4.22. (Figure 8)

7.3 Application Rating Vs. No. of Applications

According to the Figure 9, there is a relationship between number of installation and rating:

- The more rating, the more number of application installed.
- The most number of application ratings are ranged from 3.3 to 5 .

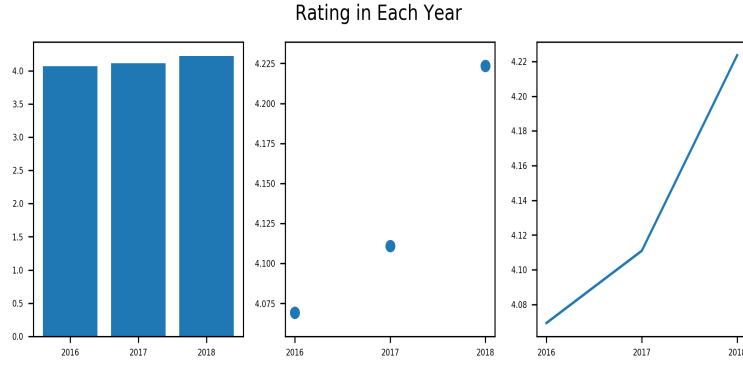


Figure 8: Rating in each year

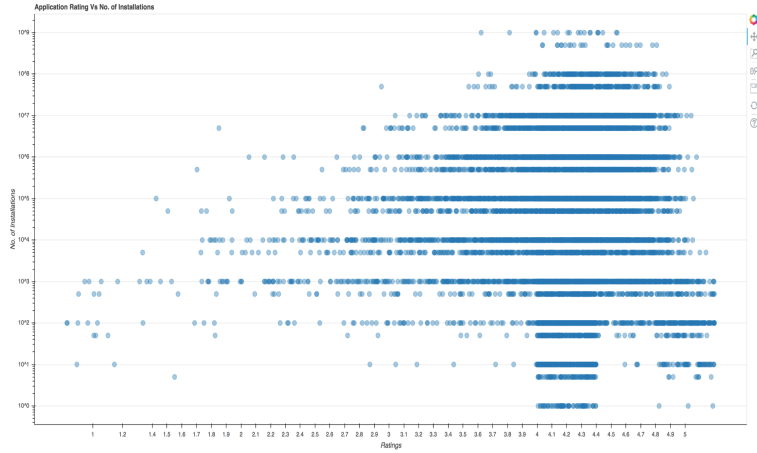


Figure 9: Application Rating Vs. No. of Applications

- The most number of application installations are ranged between 10^3 to 10^5 .

8. Average Size of applications

The bar chart (Figure 10) represents the average size of applications. Through all the applications, Game is considered as a huge average size. The other applications such as Travel, Sports, Family, Education, and entertainment have occupied the following big size.

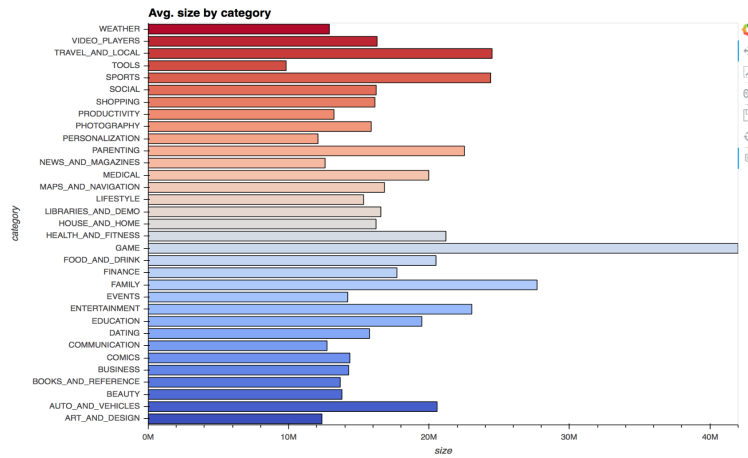


Figure 10: Avg. Size by Categories

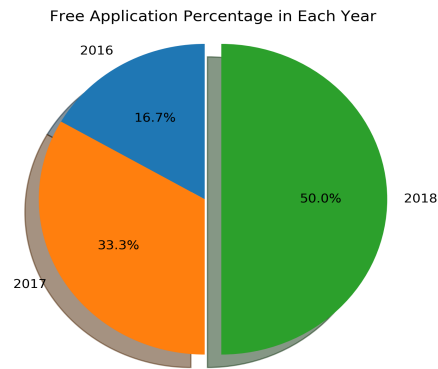


Figure 11: Free Applications Percentage in Each Year

9. Free Applications

9.1 Free Applications Percentage in Each Year

The Pie chart (Figure 11) presents the growth trend of free applications from 2016 to 2018. In 2016, it started by 16% and in following years the percentage of free applications has increased and reached to the 50%. It shows that the developers invested more on free applications in 2018.

9.2 Application Rating Vs. Number of Price

This graph (Figure 12) shows the relationship between installation and price. It presents the more installation, the less price is valued. Free application has

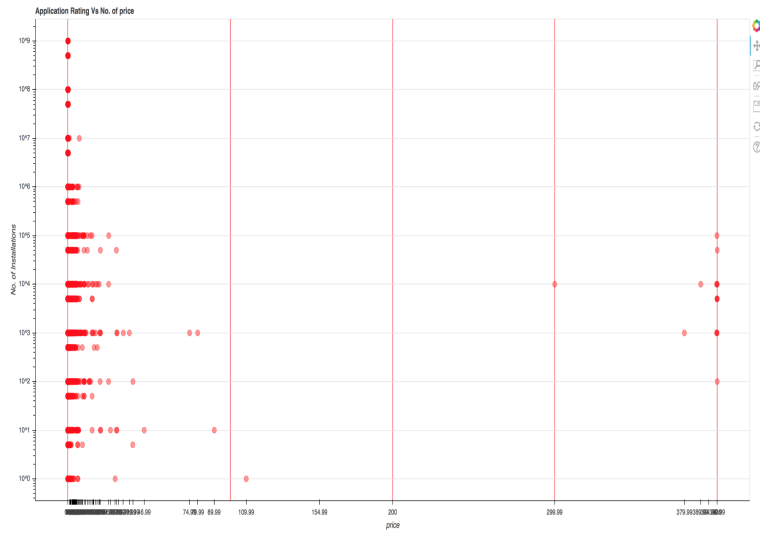


Figure 12: Application Rating Vs. No. of Price

the highest installation through all applications. Moreover, the graph clearly displays the increase of price leads users less download the application.