

Analyzing App Releasing and the Updating Behavior of Android Apps Developers

Iqra Saher, Akmal Khan, M. Zeeshan Jhandir, Rafaqat Kazmi, and Imran Sarwar Bajwa

Department of Computer Science, The Islamia University of Bahawalpur, Pakistan

Corresponding author: Akmal Khan (e-mail: Akmal.Shahbaz@iub.edu.pk).

Abstract- Google Play Store is the emerging platform in which millions of applications are releasing every day. Users install various applications on their Android phones. We perform detailed analysis on dataset for finding a relationship among various app attributes. Therefore, we use web scraping and collected the metadata of various categories of applications for one month from 25 March to 24 April 2020 from Google Play Store. We identify the developers' involvement and activeness in app market and try to find the number of apps updated by the developers. We found the top 10 categories among 48 categories, and the education category was at the top when we treated all subcategories as a separate category. The majority of applications are requiring android version 4.1 and up because application versions vary. We also found the involvement of developers after app releasing and our result shows that 65% of applications have been updated by developers while some applications have not been updated but still have the highest rating. Further, the recently updated applications have also increased total installs. Our detailed analysis will help new developers to decide app releasing and updating time and will provide insights to developers about the most competitive categories of applications in Google Play Store, and the importance of different app attributes.

Index Terms— Android, App Analysis, Mobile Application, Play Store, Web Scraping.

I. INTRODUCTION

Android phones are well-known among people in this era due to its low cost and variety of apps and Google Play Store is the central place of all applications (apps) created by developers because it contains millions of apps for Android users. It is one of the vast Play Store for searching any category of apps in a very short time. Large numbers of apps are released every day by developers so it is the biggest challenge to create apps with popular features used by the app users. As the number of apps is rising, the failure ratios of apps are also increasing. Developers are releasing their apps quickly in various Android markets and it is not possible to maintain their app ranking in Play Store statistics because different apps are already available and almost providing similar features of other apps. Some developers are extremely active and they hastily release their apps in various Android markets. Quality is not important for such developers. Different Android markets and Google Play Store offers both free and paid apps to their Android users. Many free apps are also available that offer a similar functionality, which is offered by the paid app. Due to this reason, Android users prefer to install free apps and the rating for paid apps are also getting low due to this reason. Many developers are releasing large amounts of apps and some new developers have also come in the emerging market for publishing several apps, but new developers face difficulty during app updating and releasing processes because they have no proper knowledge about android applications.

The developers choose Play Store for developing and releasing their apps by purchasing an account for \$25 and then can release number of apps. Due to this reason, Play Store contains double apps than the apple app store. There, we choose Google Play Store for analysis of apps updating and releasing behavior, because it is one of the biggest app markets in the world and it is very important to guide new developers because they are focusing on app quantity, not quality. Our main contributions are as follows:

- We performed a detailed analysis of Google Play Store Applications for finding a relationship between different attributes available in app description.
- We tried to guide developers about different top categories of apps, varying android versions of apps, rating, installs etc.
- Our main aim of this research is examine the involvement of developers by using the latest update attribute with other attribute.

In this paper, we perform a detailed analysis and try to find the answers to the following research questions.

1. Which Categories of apps are more popular among users?
2. How many apps are not updating in our dataset for many years?
3. Do app versions support the same Android version?
4. Do developers are active and update their App timely?
5. Are Rating and Install app attributes are dependent on each other?

Many Researchers performed a longitudinal study of Google Play Store apps, studied developer behaviors in various aspects and performed an analysis of various app attributes but researchers used only specific attributes for analysis i.e., Rating, Installs, etc. We tried to help developers at app releasing and updating time by performing detailed analysis for a better understanding of the relationship among different attributes. Due to this reason, we have chosen the Google Play Store for extracting data of various categories of apps and for providing some guidance to new developers about different collected attributes.

II. RELATED WORK

Researchers also performed analysis on Google Play Store and extract data by using web scraping[1] but focused on only some attributes i.e., App type (free, paid), Rating, Update, etc. Many researchers have worked on understanding the different behaviors of Android app developers [5] with detailed analysis and provide statistics about apps. Previous studies also performed a detailed analysis [6] for a better understanding of the Google Play Store and designed several research questions and tried to find the answers to these designed research questions. All research questions are related to the app market, app update duration, top apps, etc. Researchers also characterized developers [10] into four types according to the number of apps released in different Android market. Some researchers surveyed for understanding the App releasing [8] and App updating [9] criteria and the researcher also performed a detailed analysis [2]-[4], [6]-[7], [10]-[11] of Google Play Store apps. Figure 1 shows that researchers focused only some attributes i.e. rating, install, etc. but we have collected more number of attributed and tried to find the effect of one attribute to another.

Proposed Methodology	Year	# of Apps	Survey	Attributes	Crawler	Developers Behavior
N. Viennot et al.	2014	Over 1.1 Million	✗	✓	✓	✗
B. Carbanar et al.	2015	More than 1,60,000	✓	✗	✗	✗
M. Nayebei et al.	2016	-----	✓	✓	✗	✗
H. Wang et al.	2017	Over 1 Million	✗	✓	✓	✗
R. Potharaju et al.	2017	More than 1,60,000	✗	✓	✓	✗
H. Wang et al.	2018	Over 3.6 Million	✗	✗	✗	✓
H. Wang et al.	-----	Above 6 million	✗	✓	✓	✓
H. Wang et al.	2019	Over 6.2 Million	✗	✗	✗	✓
H. Wang et al.	2019	5.3 million	✗	✓	✓	✓
R. M. A. Latif et al.	2019	Free 10k Paid 3600	✗	✗	✓	✗
Our Work	2020	Over 3500	✗	✓	✓	✓

FIGURE 1. COMPARISON TABLE

III. BACKGROUND

Google Play Store is one of the popular Play Store in the world because it provides a variety of apps to their Android users in one place. The Apps are classifying into different classes called app categories. Each App belongs to one specific category according to the nature of the app. Getting any app from the Play Store into your Android phone is called app installs. The App Installs increase when there is a positive keyword relevance given by developer and high rating and positive reviews from user side. App rating is the number, which range from one to five and rated by the Android app users. One is the lowest rating while five is the highest number, which is assign by the user when they satisfy with the provided app features. Several reasons for the low rating of any apps are bug, Negative Reviews, Poor Performance, Fake Apps and Poor Customer Support. When the developers add, new features with their already released apps are called app update. The updating any app increases the size of the app, but apps maintenance is also necessary for getting better rating and installs. After creating an app, the developers choose different Android markets for publishing their apps. Uploading apps in different Android Markets are called app release. Either free or paid apps can be offered to Android users by developers are called app type. Some developers offer free apps to their Android users, but some charge dollars for using apps. Moreover, a user can give review about app after installing in their Android phones. A review is an opinion and it may be negative or positive depends on the satisfaction of users with apps.

IV. METHODOLOGY AND RESULTS

We built our web scraper and extract metadata of Google Play Store Applications for one month. Our scraper takes 6 to 7 hours daily to collect data from Google Play Store. Figure 2 shows our complete methodology comprising of three steps: 1) Data Collection by Web Scraping, 2) Data Cleaning, 3) Data Analysis and Visualization.

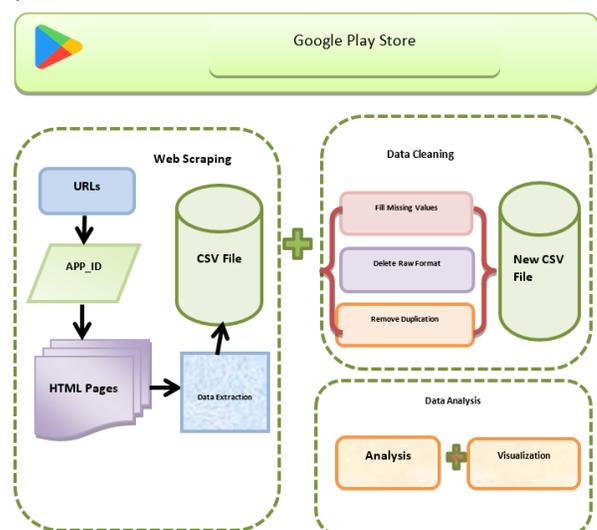


FIGURE 2. METHODOLOGY DIAGRAM

A. DATA COLLECTION BY WEB SCRAPING

First, we collect Uniform Resource Locator (URL) from different categories of Apps. Using these URLs; we collect packages names of various categories of apps and then extract data from Google Play Store. Using this technique, we collect 3,509 unique app data for one month and the attributes we collect are 1) App Title 2) Rating 3) Category 4) Developers email 5) Update 6) Size 7) Installs 8) Version 9) Required Android 10) Offer by and 11) No. of User involved Rating.

B. DATA CLEANING

We delete all duplicated rows and find total null values in each column using a heat map and filled missing values by taking the mean of columns. We converted size k into e^3 and M into e^6 , removed + sign from Installs & User Rating and split Update Date 28-12-2018 into the year, month, and day into three columns. For the current version of the app, we keep the version begin in the dataset but removed version end. Moreover, we collect developer email during web scraping but we did not use this attribute and store this attribute in another CSV file, which can be accessed easily when needed.

C. DATA ANALYSIS AND VISUALIZATION

After data cleaning, our data is ready for analysis and visualization. We do not try to collect data of any specific category and we collect all categories of data to find more about each category.

a) APPS CATEGORY VS. APP COUNT AND TOP 10

Different researchers combine subcategories of apps into a single category due to space limitations but we consider all apps in our dataset as separate categories. Moreover, we do not create a separate column for app genres and try to find the total number of apps of all categories available in our dataset.

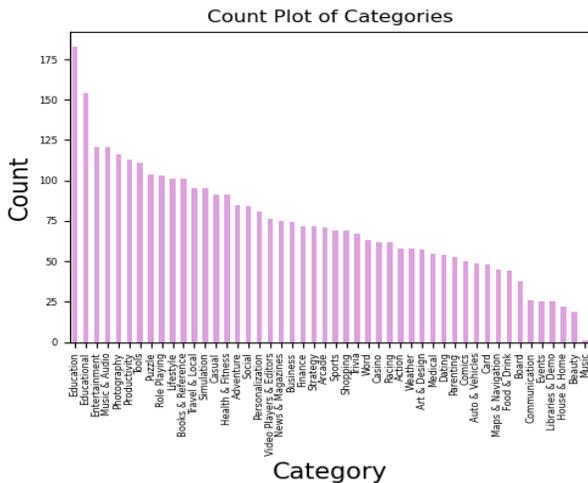


FIGURE 3. BAR CHART OF CATEGORIES BY NUMBER OF APPS

Figure 3 shows that the Education category contains a huge amount of apps while the music category has the lowest number of apps. The education category has above 175 apps in our dataset while only one app belongs to the music category. We further try to find the top 10 categories that contain more numbers of apps in our data sets.

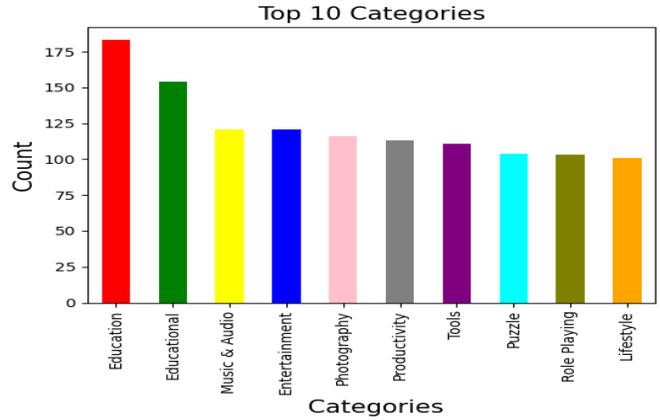


FIGURE 4. TOP 10 CATEGORIES VS. APP COUNTS

Figure 4 shows the top 10 categories among 48 different categories. The education category is at the top while the educational category was on the second.

b) ANDROID VERSIONS VS. APP VERSIONS

Different app versions are supporting different Android phones. We try a different version of apps to find which types of Android phones are required for these versions of apps.

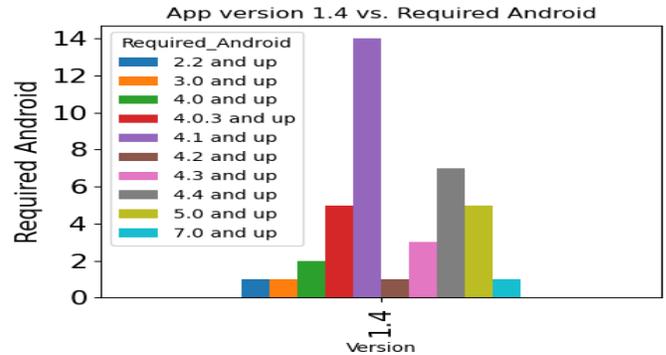


FIGURE 5. APP VERSIONS 1.4 AGAINST REQUIRED ANDROID

Figure 5 shows that Android version 4.1, and up is more suitable for app version 1.4. Version 4.0.3 and 5.0 are equally supporting version 1.4 but It does not much support the Android versions 2.2, 3.0, 4.2, and 7.0 versions. There are a total of 10 Android versions which are supporting this app version 1.4.

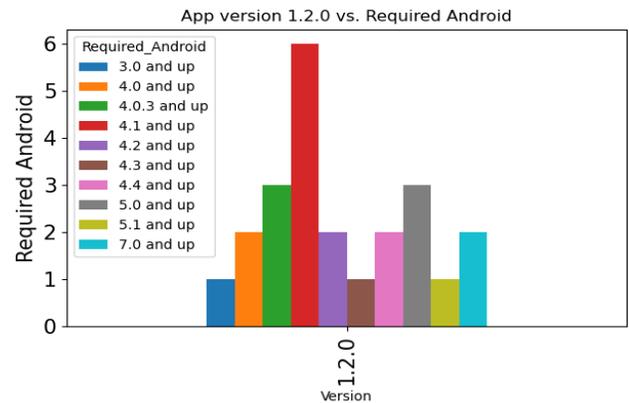


FIGURE 6. APP VERSIONS 1.2.0 BY REQUIRED ANDROID

Figure 6 shows app version 1.2.0 is supporting the android version 4.1 and up and it is also supporting many other android versions. The app versions 1.2.0 are equally supporting the 4.0.3 and 5.0 and up android version and 3.0, 4.3, and 5.1 are up are also supporting the app version 1.2.0.

c) APP CATEGORY VS. RATING

We calculated the mean average of rating attributes to find out which categories of apps has the highest rating and is popular among users.

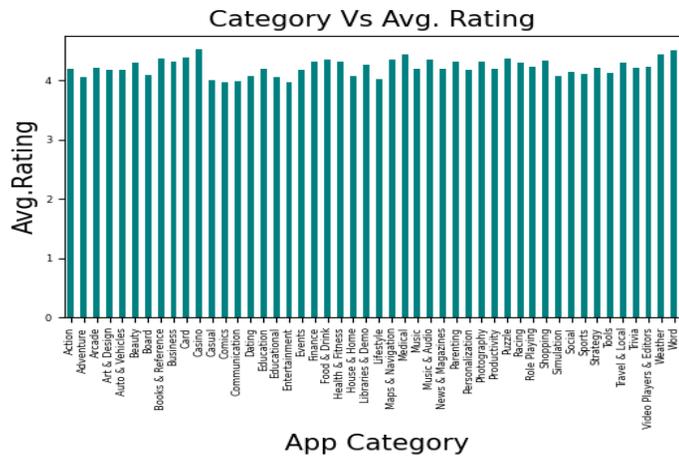


FIGURE 7. BAR CHART OF APPS CATEGORIES BY AVERAGE RATING

Figure 7 shows that all categories of apps contain the mean rating above 3.5 but the casino and word categories contain the same highest rating i.e., 4.5, and also the majority of categories contain ratings above 4.

d) APP UPDATE VS. COUNT, RATING AND INSTALLS

The Year of Update attribute is also an important attribute. We have above 3500 unique apps in our dataset and we find the total number of apps updated by year.

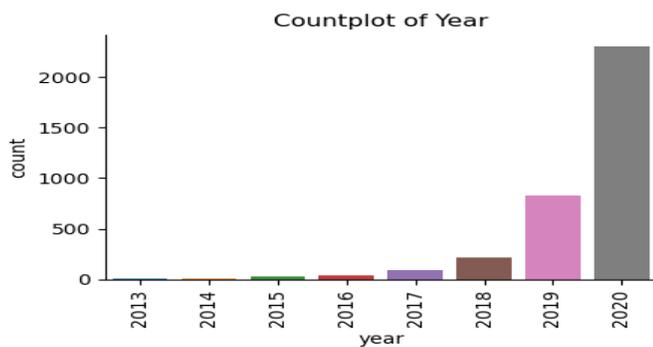


FIGURE 8. COUNT PLOT OF UPDATE YEAR BY TOTAL NUMBER OF APPS

Figure 8 shows, more than 2000 apps in our dataset have updated in the year 2020. It means developers are very active and they are updating their apps timely, but none of the apps have updated in the year 2013. Next, we find the rating of apps in different updating year.

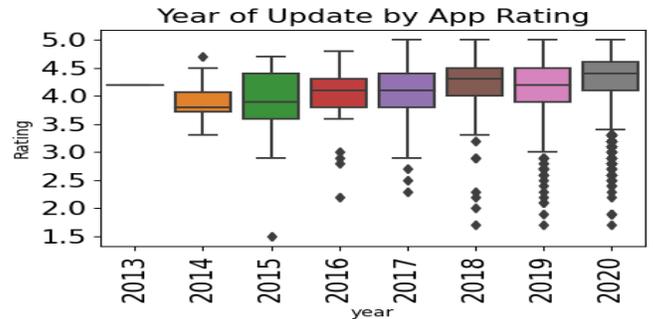


FIGURE 9. BOXPLOT OF UPDATE YEAR BY TOTAL RATING OF APPS

Figure 9 shows that the app updated in the year 2020 has a minimum rating above 3.2 while the maximum rating value is 5.0. The apps, which are updated in different years, all have minimum values above 2.5 it means none of the apps has a rating below 2.5. Next, we relate the total app installs with the year of update and tried to find the number of installs in different updated years.

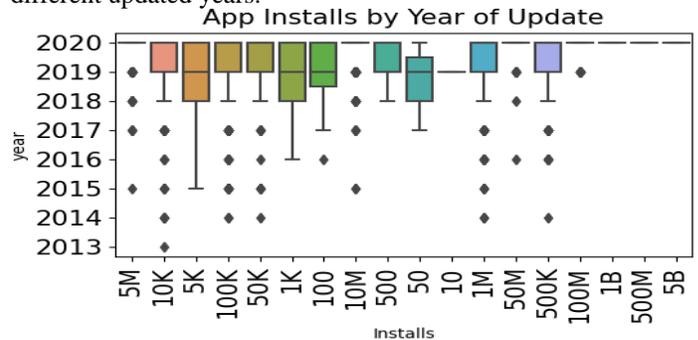


FIGURE 10. BOXPLOT OF TOTAL APP INSTALLS BY YEAR OF UPDATE

Figure 10 shows that the apps updated in the year 2020 have a maximum of 5B installs while a minimum of only 50 installs. The app updated from 2018 to 2020 has a minimum of 500 installs. The apps that are updating from 2015 to 2020 have not received higher installs.

e) APP INSTALLS VS. RATING

We try to find a relationship by using rating and install attributes. We considered both attributes are important for success of any Android Market. We tried to find a relationship by using both attributes for analysis.

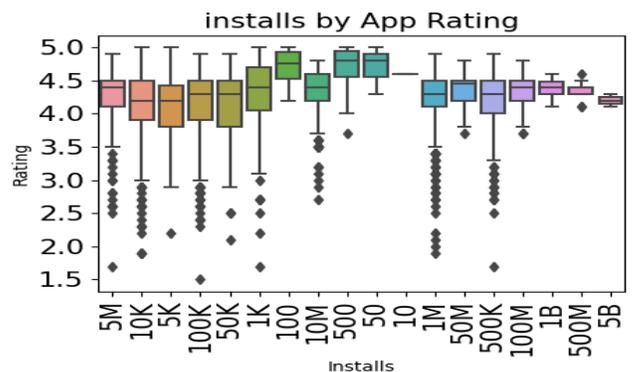


FIGURE 11. BOXPLOT OF TOTAL APP INSTALLS BY RATING

Figure 11 shows that the apps which have the highest installs 5B have a rating below 4.5 while the apps which have only 50 installs have a rating above 4.5. It means the rating does not depend on the high installs. All app installs have minimum rating value are above 2.7 while has maximum rating value 5. The apps which have only 10 installs also have rating 4.5.

f) APP INSTALLS VS. TOTAL APPS

Next, we find the total number of app installs in our dataset.

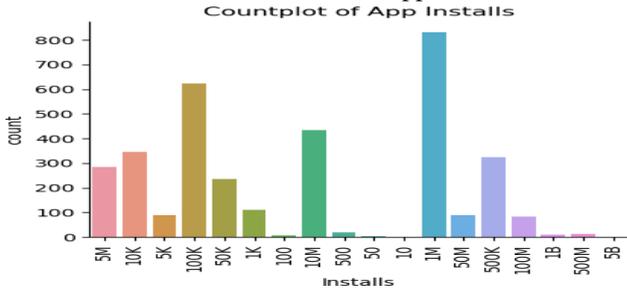


FIGURE 12. COUNT PLOT OF TOTAL APPS INSTALLS

Figure 12 shows that more than 800 apps have 1M installs while only one app is available in our dataset has only 10 installs. There are few apps in our dataset, which has the highest number of installs i.e., 5 billion. There are above 50 apps in our dataset that has the good installs 5K, 50M, 100M, etc.

g) APP CATEGORIES AND UPDATE YEAR

Next, we find different categories of the app have updated in the year 2019 and 2020 because we want to find the categories, which are ignored by the developer during the updating process.

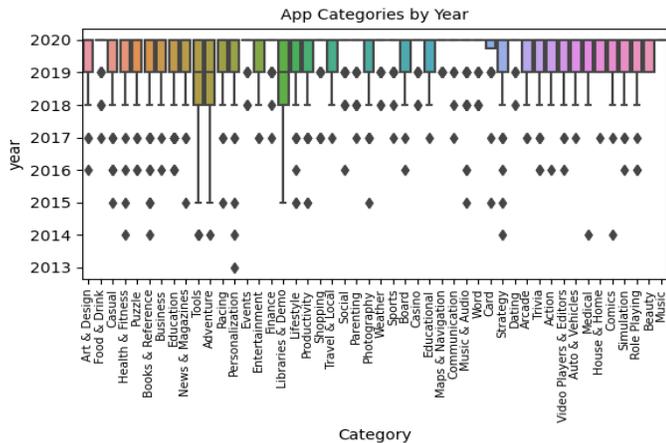


FIGURE 13. PLAY STORE CATEGORIES BY YEAR OF UPDATE

Figure 13 shows that all categories of apps updated in the year 2020. Mostly categories of apps are updating since 2018 e.g., Tools, Adventures, Libraries, and Demo and still the developers are updating these categories because some apps require proper maintenance with time

a) REQUIRED ANDROID AND UPDATE YEAR

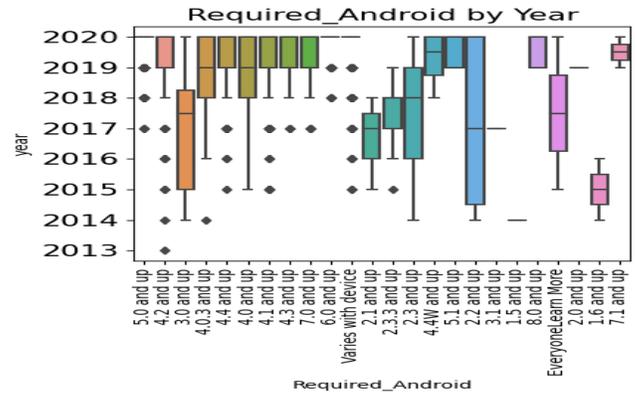


FIGURE 14. BOXPLOT OF REQUIRED ANDROID BY YEAR OF UPDATE

Figure 14 shows that the apps updating in the year 2013 are not supporting any Android version, but the apps updated in the year 2014 have a minimum value of Android are 3.0, 2.3, 2.2, etc. The android version 1.6 is not using now because it has been used in previous years and now latest versions are used i.e., 4.2, 4.0.3, 7.0, etc.

h) APP CATEGORY VS. RATING

Finally, we tried to find the ratings of different app with various available categories.

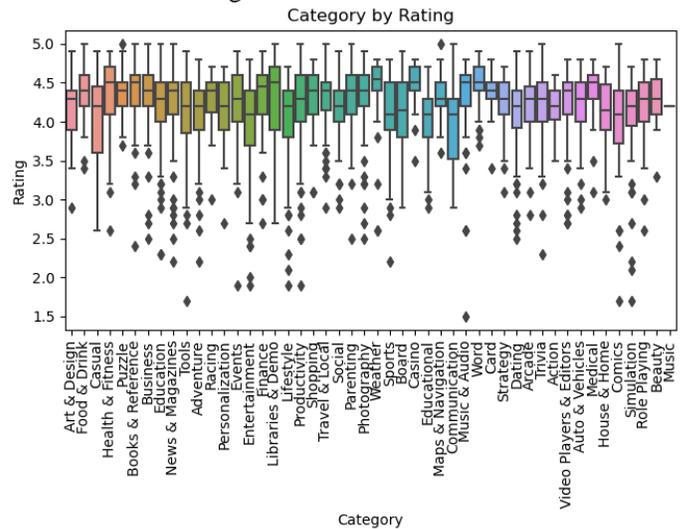


FIGURE 15 BOXPLOT OF CATEGORIES BY RATING OF AN APP

Figure 15 shows that all categories of apps have a minimum rating above 2.5 while there are also some other categories of apps are available whose maximum rating is 5.0. We observe that we have 3509 apps in our dataset and more than 2000 apps are recently updated in the year 2020. It means that developers are very active and they are maintaining their apps timely. We find that 65% of apps have been updated and most apps that have updated in the year 2020 have a higher rating than the previous rating. It means ratings of apps have improved after updating. The education category contains a large number of apps than other categories when we treated all subcategories separately. These results will be helpful and provides additional

knowledge to developers at app releasing and updating time and also guide to new developers about the app popularity and the effect of one attribute to another.

V. CONCLUSION

We collected the metadata of application from the Google Play Store for one month by using web scraping and then we performed an analysis of our dataset for a better understanding of the various categories of apps and we observe that Education category has a large number of apps i.e., 183 while Music category has only one app in our dataset. All top 10 categories of apps in our dataset has above 100 apps in which lifestyle category was on number 10th and has 101 apps .When we have taken the App version 1.4 then total 10 Android Versions were supporting. Android Version 4.0.3 and 5.0 are equally supporting the app version 1.4 .Above 1000 Apps are supporting the Android version 4.1. When we have taken the mean average of rating attribute then mostly categories of apps has received above 3.8 rating .Word and Casino categories has the highest mean rating ,i.e., 4.5. When we found the number of apps in different updating years, then above 2000 apps have been updated recently. When we relate year and rating attribute then recently updated apps has minimum rating i.e., above 3.2 while some apps are also receiving maximum rating i.e., 5.0. And also recently updated apps have also received better installs. Some recently updated apps in our dataset have maximum installs .i.e., 5 Billion. The apps which have the highest installs are receiving rating range below 4.5 but there are also some apps which have only 50 installs but have rating 4.5. It means rating and installs are not dependent to each other. We performed our detailed analysis but releasing date are not present in the app description, the researcher can build any scraper or use app for collecting the app releasing date and also same dataset can be used in future for rating prediction.

REFERENCES

- [1] R. M. A. Latif and F. Ijaz, "Data Scraping from Google Play Store and Visualization of its Content for Analytics," *2019 2nd Int. Conf. Comput. Math. Eng. Technol.*, pp. 1–8, 2019.
- [2] H. Wang, H. Li, and Y. Guo, "Understanding the evolution of mobile app ecosystems: A longitudinal measurement study of Google play," *Web Conf. 2019 - Proc. World Wide Web Conf. WWW 2019*, pp. 1988–1999, 2019, doi: 10.1145/3308558.3313611.
- [3] N. Viennot, E. Garcia, and J. Nieh, "A measurement study of Google Play," *Perform. Eval. Rev.*, vol. 42, no. 1, pp. 221–234, 2014, doi: 10.1145/2591971.2592003.
- [4] H. Wang *et al.*, "An explorative study of the mobile app ecosystem from app developers' perspective," *26th Int. World Wide Web Conf. WWW 2017*, pp. 163–172, 2017, doi: 10.1145/3038912.3052712.
- [5] H. Wang, X. Wang, and Y. Guo, "Characterizing the global mobile app developers: A large-scale empirical study," *Proc. - 2019 IEEE/ACM 6th Int. Conf. Mob. Softw. Eng. Syst. MOBILESoft 2019*, pp. 150–161, 2019, doi: 10.1109/MOBILESoft.2019.00031.
- [6] B. Carbunar and R. Potharaju, "A longitudinal study of the Google app market," *Proc. 2015 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2015*, pp. 242–249, 2015, doi: 10.1145/2808797.2808823.
- [7] R. Potharaju, M. Rahman, and B. Carbunar, "A Longitudinal Study of Google Play," *IEEE Trans. Comput. Soc. Syst.*, vol. 4, no. 3, pp. 135–149, 2017, doi: 10.1109/TCSS.2017.2732167.
- [8] M. Nayebi, B. Adams, and G. Ruhe, "Release Practices for Mobile Apps -- What do Users and Developers Think?," pp. 552–562, 2016, doi: 10.1109/saner.2016.116.
- [9] Y. Tian, B. Liu, W. Dai, B. Ur, P. Tague, and L. F. Cranor, "Supporting privacy-conscious app update decisions with user reviews," *SPSM 2015 - Proc. 5th Annu. ACM CCS Work. Secur. Priv. Smartphones Mob. Devices, co-located with CCS 2015*, pp. 51–61, 2015, doi: 10.1145/2808117.2808124.
- [10] H. Wang, H. Li, L. Li, Y. Guo, and G. Xu, "Why are Android apps removed from Google Play?: A large-scale empirical study," *Proc. - Int. Conf. Softw. Eng.*, pp. 231–242, 2018, doi: 10.1145/3196398.3196412.
- [11] H. Wang *et al.*, "Beyond Google Play : A Large-Scale Comparative Study of Chinese Android App Markets," no. Section 5.