

**Understanding the factors that influence the popularity of
mobile applications - A comparative study between Android
apps and IOS apps**

Poornima Joshi, M.S.

Nima Zahadat, Ph.D.

The George Washington University

ABSTRACT

Technological advancements especially in the mobile field have started to demand the need for mobile applications to not just survive but thrive in the business world of today. There is an inherent need to be fast, cheap and efficient to conquer the mobile market. In order to be durable, there is a need to understand the factors that affect the popularity or likeability of these mobile apps among the users. Rating in the app distribution platform is a significant performance metric for measuring how well an app is doing in the app market. It is important to explore the factors which affect the popularity in order to develop competitive applications, understand where to invest time or money, and succeed in the business world. This research presents a case study, involving quantitative studies performed on the two most leading mobile markets - Google Inc.'s Play store and Apple Inc.'s App store. Thus, our approach is to perform quantitative data analysis on these stores with a desire to effectively investigate and communicate the findings of this case study.

Keywords: Mobile applications, Mobile apps marketing, Mobile software engineering, Software popularity.

INTRODUCTION

The ubiquity of mobile apps in everyday life make them an attractive source for mining digital knowledge (Liu, et al., 2018). Mobile apps are software applications that are designed for handheld smart phones, tablets and other smart devices. Present day mobile phones are almost an extension of a human body and numerous apps form the backbone of the mobile device to enable users to perform their day to day tasks. These apps are developed by software developers and distributed to the users using an application distribution platform. Google Inc. and Apple Inc. with their Play Store and App Store respectively are the biggest players in the app markets (Annapurna, Teja, & Murthy, 2016). Mobile apps are being developed in diverse areas such as education, business, health, finance and entertainment and are distributed as either free or paid services.

App development is usually market driven (Lim, Bentley, Kanakam, Ishikawa, & Honiden, 2015). Although the developers and product experts put extensive efforts into market research to understand the end user needs for their apps, it is often challenging for them to guarantee user

satisfaction. During the development of the app, the developers have limited exposure to the potential users. Success of an app is often perceived by the number of downloads, daily active users and average rating of the application (Lim, Bentley, Kanakam, Ishikawa, & Honiden, 2015). However, accurately measuring the success of a software system is difficult as there is neither a universal metric nor a ranking scheme (German, 2007). Therefore, this study is aimed at trying to understand different factors that influence the popularity of mobile applications and also draw some comparisons between user behavior of Android and IOS apps.

LITERATURE REVIEW

The benefit of mobile apps is evident. However, there is a need to understand which factors specifically contribute to the success of mobile applications. This section presents the related literature of mobile applications, their popularity with respect to Android and IOS devices.

Understanding which characteristics influence the high rating of mobile applications is crucial (Tian, Nagappan, Lo, & Hassan, 2015). Tian et al. discuss in their study, that 28 factors with 8 dimensions could be potentially associated to app ratings - a measure of popularity in their study. According to them, three of the 28 factors have a major influence on the app rating - install size of an app, number of promotional images, and target SDK version. Tian et al. considered only free apps for their research purposes which may not support in generalizing the results to all apps of Android platform.

Lee et al. performed a study on the top 300 apps in the Apple's App store and examined how factors at different levels affect the app's survival in the top 300 charts. Furthermore, they found that providing free applications, investment in relatively less popular categories and constantly providing updates on app features increase sales performance (Lee & T. S, 2014). However, the study only focused on the longevity of the app and not on the factors that influence the success of the app.

Fu et. Al collected and studied over 13 million user reviews from Google Play. They proposed a system called WisCom, which is an integrated system to analyze user reviews (Fu, et al., 2013). It

supports the mobile app markets, benefiting end-users, app developers, market operators and other relevant stakeholders in mobile app ecosystem. However, Fu et.al in their study, they did not account other factors of the app itself like size, number of install or date of install in their analysis, instead focused on the user reviews only.

Frank et al. crawled a corpus of 188,389 Android apps from several Android app stores including the official Google Play Store. Their primary objective was to discover the patterns in the Android permission requests by applying Boolean matrix factorization (Frank, Dong, Felt, & Song, 2012). They used probabilistic model to mine permission request patterns from Android and Facebook applications based on reputation only and did not look into it the other way around, i.e. to look if permissions impact the reputation of an application.

In addition, there have been several researches on the app user behavior for mobile applications based on country. App market is an extremely competitive market where the sale margin is low (Lim, Bentley, Kanakam, Ishikawa, & Honiden, 2015). App user's behavior differs significantly across countries and have different expectations in various geographical locations (Lim, Bentley, Kanakam, Ishikawa, & Honiden, 2015). Lim et al. only focus their study on country as a factor for popularity and do not discuss the other factors.

To encapsulate all of these findings of several researchers and understand the factors that contribute to the popularity of mobile applications at a high level, we have designed this study.

RESEARCH

Purpose

The purpose of this study is to understand what factors influence the popularity of mobile applications with respect to the two most popular mobile ecosystems - Android's Play store apps and Apple's App store apps. Following are the research questions that we have attempted to answer in this study.

RQ #1: How does popularity vary per genre, application size, and price? The goal here is to provide an initial view about the popularity of the apps, by comparing the average user ratings according to genre, application size, and price of the application.

RQ #2: Does popularity correlate with the characteristics of an app like number of installations and number of ratings? This investigation is essential to check whether these factors can be improved to increase an app's popularity.

RQ #3: What is the impact that new versions have on popularity of an application? This study can demonstrate if relevant gains in popularity happen due to new versions (implemented in new releases).

Methodology

This study is a case study on mobile applications and application distribution systems. Case study is mainly chosen as the research methodology, as it is the most widely used research methods in information systems research (Darke, Shanks, & Broadbent, 1998). It is well suited to understand the interactions between information technology-related to innovations and organizational contexts (Darke, Shanks, & Broadbent, 1998) (Myers, 1997). Thus, we study the existing popular application distributed system, scrape data from it and examined it in a way to identify the causality of popularity amongst them.

This study is conducted as a two-part simultaneous case study where the Android's Play store and Apple's App store are examined simultaneously to observe the behavior of users in each of these systems. Consequently, we have chosen the mobile app rating in the respective stores as the unit of measure for popularity in our analysis.

Moreover, the purpose of most researches is to generate a theory and prove it (Norman, 1991). It can be achieved by either qualitative data collection which is concerned with words and meanings or quantitative methods which are concerned with numbers and measurement (Yin, 2009); or sometimes combination of both. Since our study involves human behavior, verification cannot rest

on intuition, argument, or opinion (Norman, 1991); therefore, it was decided to base our verification on quantitative data analysis with a desire to effectively investigate and communicate the findings of this case study with a stimulating and valuable report.

DATASET

This study aims to understand which factors influence the popularity or high rating of the mobile applications in both Android Play Store and Apple App Store. The rating of each app was a primary requirement in the data set, since rating is directly associated with popularity (Liu, Au, & Choi, 2014). Thus, two data sets were obtained – one of Play Store Kaggle (Gupta, 2018) and another of App store (Ramanathan, 2018) from Kaggle. The Play store dataset in Kaggle was generated by scraping the Play Store web page using Selenium. It consists of information such as name of app, category, number of installs, rating, price and much more. The App Store data set was extracted from the iTunes Search API at the Apple Inc website. R and Linux web scraping tools were used. The App store data contains information like name, size, price, rating, genre and much more.

DATA ANALYSIS

Data Pre-processing

The primary motivation in this analysis was to understand and explore the factors which affect the popularity in order to develop competitive applications, understand where to invest time or money, and succeed in the business world. In order to accomplish it, the two datasets from Kaggle were cleaned in Python using Pandas and NumPy libraries, to make it easy to compare. In the Play store data set, the size column consisting of different apps were converted to megabytes to maintain consistency. The ‘+’ at the end of ‘Number of Installs’ column was removed. Finally, all the duplicates were dropped, and NaN values were removed. Similarly, in the App store data set the size of the apps were converted to megabytes, duplicates were dropped, and NaN values were removed. Lastly, the Apple description data in a separate data frame was merged into the apple data set.

Data Mining

Python and several python packages were utilized to perform the data mining tasks. Pandas and NumPy were used for data restriction and manipulation. Matplotlib, Seaborn and Plotly were used to generate graphs. The Seaborn color palette was predominantly used throughout.

Before attempting to answer the research questions, it was important to understand the data, since it will add a new dimension to the way we would approach our data (Velleman & Hoaglin, 1981). A pie chart (Appendix, Figure 1 and Appendix, Figure 2) was drawn to see what categories or genres of apps are trending in the app markets. The observation was that family, game, entertainment and education form the top 50% of all apps. Further a bar charts of counts were plotted to analyze the free and paid apps in the app markets. Android has very high number of free apps compared to paid apps (Appendix, Figure 3). And, Apple has a mix of free and paid apps (Appendix, Figure 4).

In addition, it was discovered that the average size of an app in Play Store is about 20.3 Mb and average size of an app in Apple Store 190 Mb. Finally, an investigation on our primary interest - the rating distribution of apps was conducted. It was observed that a majority of ratings were slightly greater than 4.5 in Play Store (Figure 1). When the same was repeated for App Store, it was observed that a majority of ratings were around than 4.5 and a significant number of apps even have 0 rating (Figure 2). Proceeding with these findings as a basis, we go ahead to answer our research questions which will be discussed in the next section.

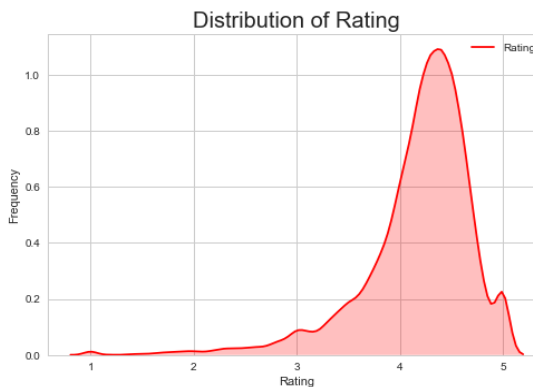


Figure 1: Distribution of Ratings in Play Store

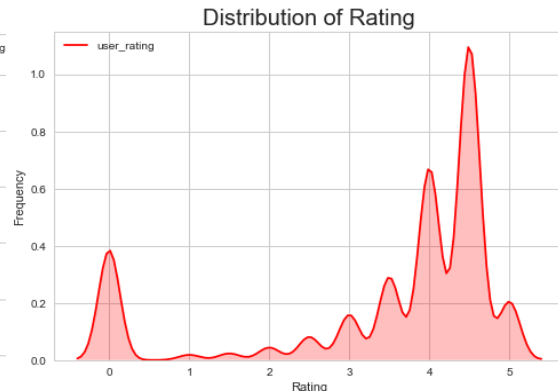


Figure 2: Distribution of Rating in App Store

KEY FINDINGS

In this section, we use the described datasets to answer the three research questions listed in the paper's research section.

RQ #1: How does popularity vary per genre, application-size, and price?

To begin answering *RQ #1*, we compare the popularity that varies per genre. Figure 2 illustrates that Family followed by Games and Tools consist of the max number of apps in the Google Play Store. Further, from Figure 1 it can be observed that most app categories are performing reasonably well with medians lying around the overall average. 'Health_and_Fitness' and 'Books_and_Reference' are outstanding with 50% apps having greater than 4.5 rating. On the other end, 50% of apps in the Dating category have a rating lesser than the average rating. This would make us speculate that these categories are saturated markets.

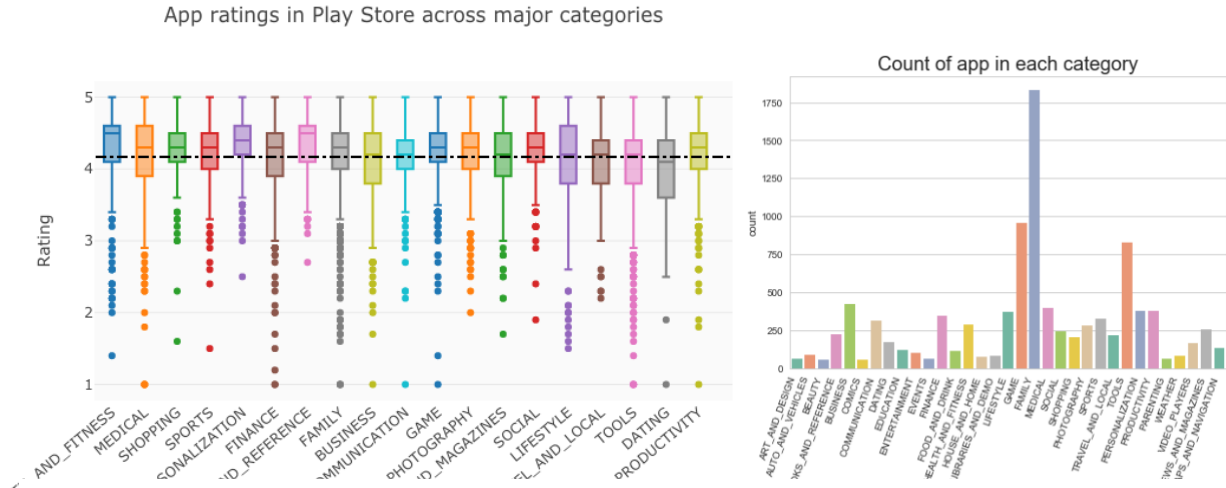


Figure 1: Box plot of app ratings in Play Store across major categories Figure 2: Bar count plot of app in each category

Further, comparing the popularity of apps in Apple's App store with respect to genre, Figure 4 shows that Games constitutes the maximum number of apps. In Figure 3, we can observe that most app categories performing reasonably ok with medians lying around the overall average, however not as well as android apps on comparing with the averages. Productivity, Shopping, Health_and_Fitness and Education are outstanding with 50% apps having greater than 4.5 rating. On the other end, 50% of apps in the Social Networking have a rating lesser than the average rating. This would make us believe that these categories are saturated markets.

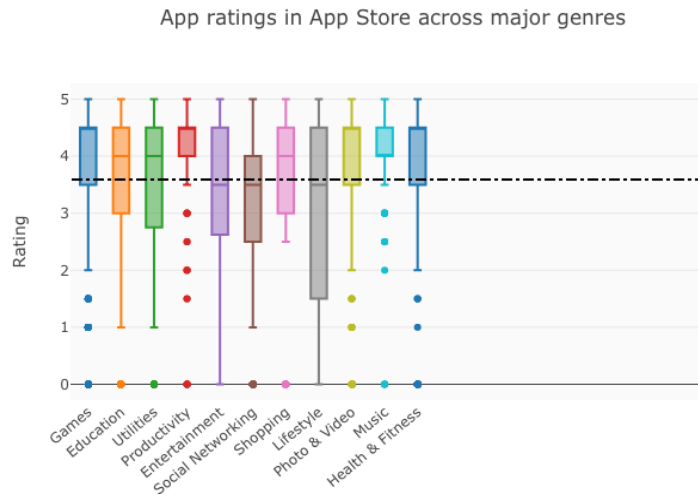


Figure 3: Box plot of app ratings in App store across major genres in App Store

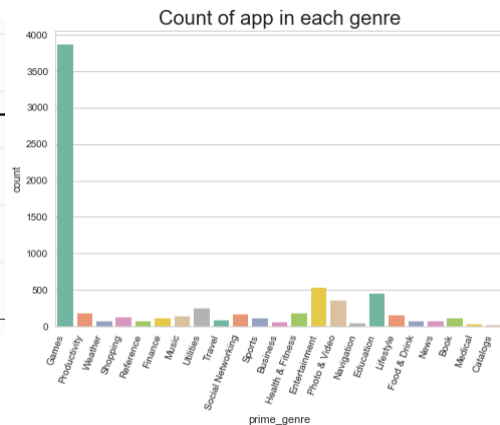


Figure 4: Bar count plot of app in each genre in App Store

A scatter regression plot of Rating v/s Price in Google Play Store showed a slight negative correlation. Therefore, increasing the price of app does not help in making it popular was our first conclusion. However, majority of apps are within the \$50 range. If we would dive deep and observe, we can see a positive correlation in Figure 5. If users are paying a price for an app means, there is a higher quality or value involved.

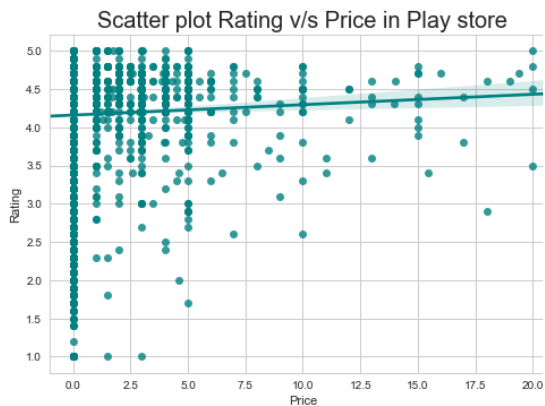


Figure 5: A scatter plot of Rating v/s Price in Play store

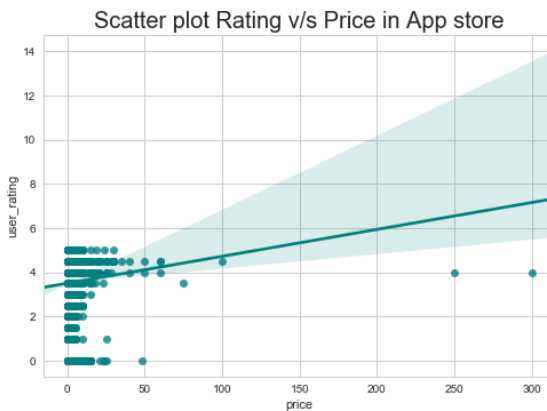


Figure 6: A scatter plot of Rating v/s Price in App

Similarly repeating the same in Apple App store, a positive correlation for price and ratings is observed. The main take away would be that, users don't mind paying a price for a better-quality product. This is reassured when we observe the apps between \$0 – \$20 (Figure 6). There is a clear positive correlation between price and rating. It indicates that users are willing to pay a price for

high quality apps. Another interesting finding was that the highest priced app costing \$400 in Android store was “I’m Rich - Trump Edition” in the Lifestyle category with 3.6 average rating and highest priced app costing \$299.99 in App store was “LAMP Words For Life” app in Education category with average of 4.0 rating. This was an interesting finding in the rating v/s price comparison (Figure 7).

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
4367	I'm Rich - Trump Edition	LIFESTYLE	3.6	275	7.3	9.0	Paid	400.0	Everyone	Lifestyle	May 3, 2018	1.0	4.1 and up

	id	track_name	size_bytes	currency	price	rating_count_tot	rating_count_ver	user_rating	user_rating_ver	ver	cont_rating	prime_genre	su
1479	551215116	LAMP Words For Life	556.243164	USD	299.99	41	0	4.0	0.0	1.5.5	4+	Education	

Figure 7: Highest priced apps in Play Store and App Store

Lastly, comparing popularity with application size (in Mb) in both Play Store and App store, we noticed no significant correlation between the two in Android Play store (Figure 8). The following joint plot (Figure 9) demonstrates that in App store as the size increases better is the rating of an app.

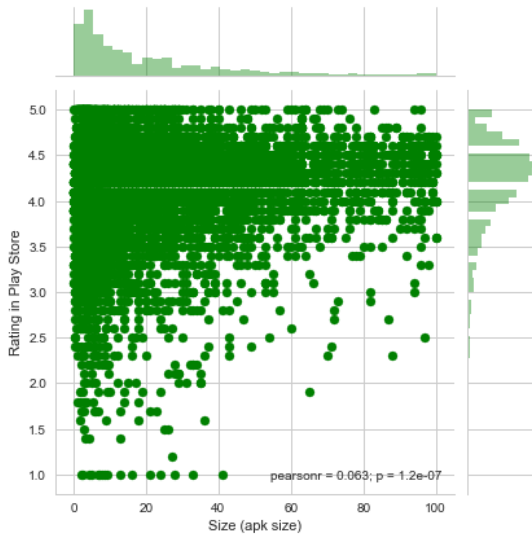


Figure 8: Joint plot of Rating and Size of app (in Mb)

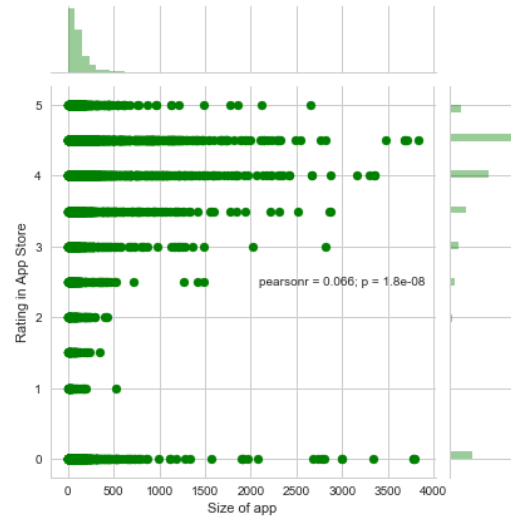


Figure 9: Joint plot of Rating and Size of app (in Mb)

RQ #2: Does popularity correlate with the characteristics of an app like number of installations and number of ratings?

Figure 10 and Figure 11 illustrate the rating v/s installs in Play Store and App Store respectively. We can notice a positive correlation between the rating of an app and number of installs. A higher

rating is observed as the number of installs increase in both Play Store and App store. We could interpret that higher installs results in an increase in app's popularity.

Rating v/s Installs in Play Store

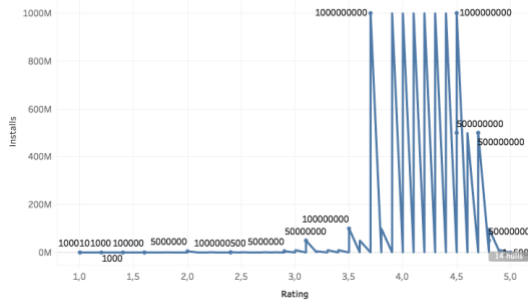


Figure 10: A plot of Rating v/s Installs in Play store

Rating v/s Number of Installs in App Store

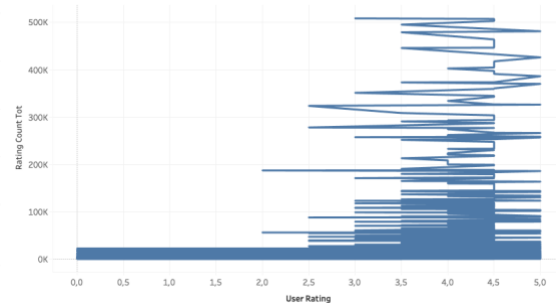


Figure 11: A plot of Rating v/s Installs in App store

Q #3: What is the impact that new versions have on popularity of an application?

Figure 12 and Figure 13 is a scatter plot of popularity v/s version. This is conducted to see if relevant gains in popularity happen due to new versions of the application. On illustrating rating v/s version in Play store for all data points, we observe a slight positive correlation (Appendix 1) between version and rating. However, significant number of apps have versions between 0 and 20 which led to focusing only on these apps. For apps having versions between 0 and 20 there is nearly no correlation between releasing versions and being more popular in Play Store. To maintain consistency, a comparison of rating and version was performed on the App Store app having version number between 0 and 20. A positive correlation between user rating and releasing versions was observed. It indicates that in App store, versions are significantly improving products.

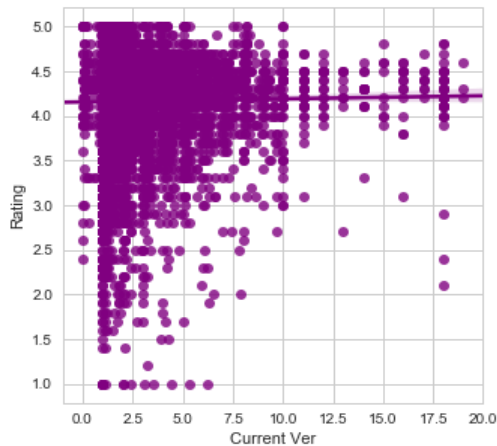


Figure 12: A scatter plot of Rating v/s Version in

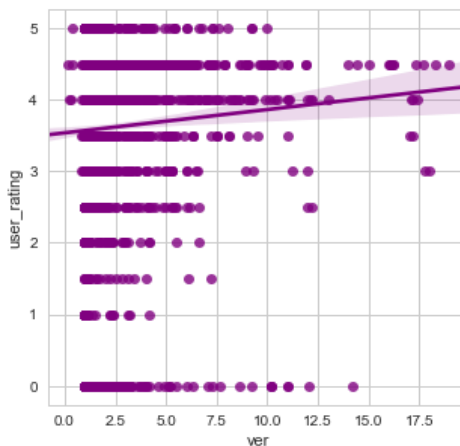


Figure 13: A scatter plot of Rating v/s Version in

CONCLUSIONS

In this paper, we first studied the popularity of mobile apps aiming to answer three research questions. (RQ #1 – Popularity variation per genre) We concluded that Games is a category, that is performing well in both in Android as well as IOS apps. Health & fitness and education/ books are generally high rated. Dating apps have low average rating in both Android and IOS apps. (RQ #1 – Popularity variation with price) Interestingly, between 0 - 20 USD, there is a positive correlation between price and rating. It indicates that users are willing to pay a price for high quality apps. (RQ #1 – Popularity variation by application size) In android apps, there is no correlation between size of an app and rating. However, in IOS apps, the larger the apps are, higher ratings are observed. (RQ #2) We found that there is positive correlation between number of installations and rating of an app. We conclude that, higher the rating the more likely it is that user will install the app. (RQ #3) New versions have no impact on popularity in Android apps, however in IOS apps, as the number of versions increase popularity has gradually increased.

FUTURE WORK

As future work, it will be interesting to investigate app markets other than the App Store and Play Store like TV apps and compare them with these popular ones. It will also be a very interesting study to correlate the independent variables themselves. For example, we could compare size of an app with the price or number of installations and category. Moreover, it will be very useful to investigate models for predicting app popularity, which can be used to warn developers when signs of stagnation are detected in their apps and steer the development of app in the right direction.

BIOGRAPHY

Poornima Joshi is a graduate student in the Data Science Program at The George Washington University.

Dr. Nima Zahadat is a professor of information systems and computer science. He has also held positions as Chief Security Officer, Chief Information Officer, Director of security, Director of Training Solutions, Dean of Computer Science, Program Chair of Information Systems, and Director of Operations. Dr. Zahadat has worked extensively with public and private sectors through- out the years. Dr. Zahadat has taught at George Mason University and the George Washington University in the fields of information systems, engineering, data science, web development, and security. He has developed and taught over 100 different Information Systems, Security, and Project Management curricula throughout his career.

He has an undergraduate degree in Mathematics from George Mason, a graduate degree in Information Systems from George Washington, and a Ph.D. in Systems Engineering and Engineering Management from George Washington. Dr. Zahadat's research interests are mobile security, information security, digital forensic, risk management, data mining, and information visualization. Dr. Zahadat enjoys biking, photography, travel, skiing, and writing.

APPENDIX

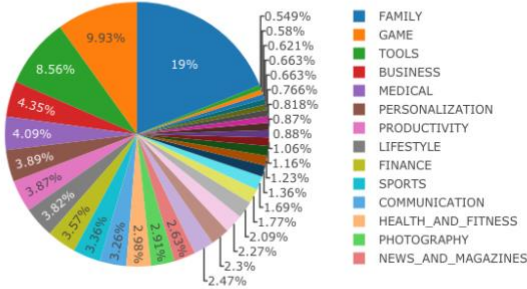


Figure 1: Pie chart of Number of Apps in different categories in Play Store

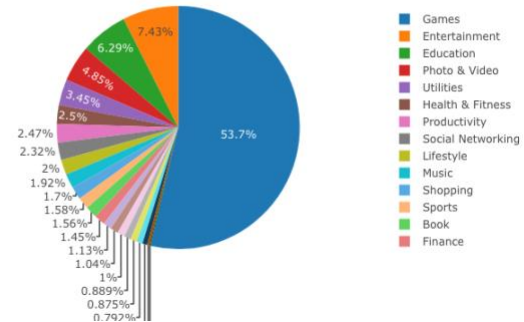


Figure 2: Pie chart of Number of Apps in different categories in App Store

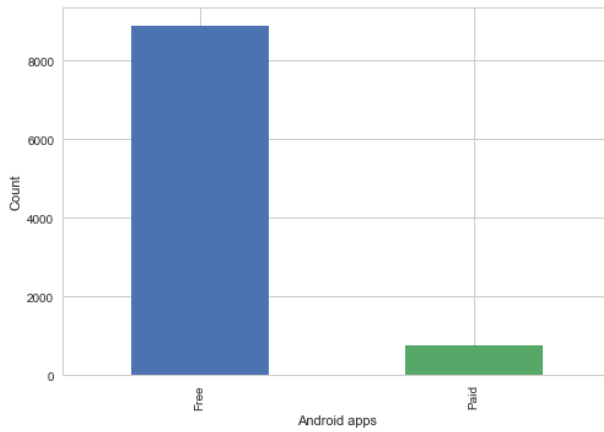


Figure 3: Bar count chart of Number of Apps that are free and paid in Play Store

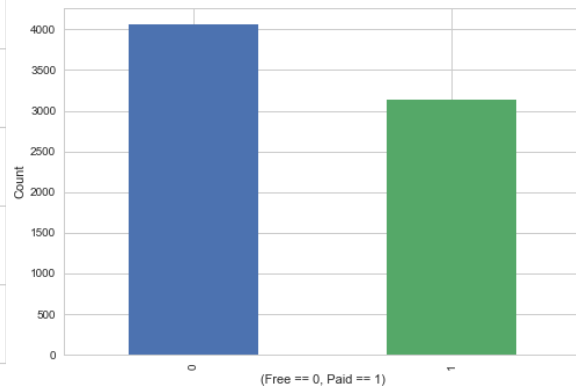


Figure 4: Bar count chart of Number of Apps that are free and paid in Play Store

REFERENCES

- Annapurna, Teja, P. K., & Murthy, S. Y. (2016). A Comparative Study on Mobile Platforms (Android vs. IOS). *International Journal of Advanced Research in Computer Engineering & Technology*, 547-553.
- Darke, P., Shanks, G., & Broadbent, M. (1998). Successfully completing case study research: combining rigour, relevance and pragmatism. *Information systems journal*, 273-289.
- Frank, M., Dong, B., Felt, A. P., & Song, D. (2012). Mining Permission Request Patterns from Android and Facebook Applications. *12th International Conference on Data Mining*. IEEE.
- Fu, B., Lin, J., Li, L., Faloutsos, C., Hong, J., & Sadeh, N. (2013). Why People Hate Your App — Making Sense of User Feedback in a Mobile App Store. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- German, D. M. (2007). Using Software Distributions to Understand the Relationship among Free and Open Source Software Projects. *Fourth International Workshop on Mining Software Repositories (MSR '07)* (p. 24). Washington, DC, USA: IEEE Computer Society.
- Gupta, L. (2018). *Kaggle*. Retrieved from Kaggle: <https://www.kaggle.com/lava18/google-play-store-apps>
- Lee, G., & T. S, R. (2014). Determinants of Mobile Apps Success: Evidence from App Store Market. *Journal of Management Information Systems*, 33-170.
- Lim, S. L., Bentley, P. J., Kanakam, N., Ishikawa, F., & Honiden, S. (2015). Investigating Country Differences in Mobile App User Behavior and Challenges for Software Engineering. *IEEE Transactions on Software Engineering*, 40-64.
- Liu, C. Z., Au, Y. A., & Choi, H. S. (2014). Effects of Freemium Strategy in the Mobile App Market: An Empirical Study of Google Play. *Journal of Management Information Systems*, 326-354.
- Liu, T. F., Craft, M., Situ, J., Yumer, E., Mech, R., & Kumar, R. (2018). Learning Design Semantics for Mobile Apps. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)* (pp. 569-579). New York: ACM.

- Myers, M. D. (1997). Qualitative Research in Information Systems. *Management Information Systems Quarterly*, 241-242 .
- Norman, K. L. (1991). *The psychology of menu selection: Designing cognitive control at the human/computer interface*. Intellect Books.
- Ramanathan. (2018). *Kaggle*. Retrieved from Kaggle: <https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps>
- Tian, Y., Nagappan, M., Lo, D., & Hassan, A. E. (2015). What Are the Characteristics of High-Rated Apps? A Case Study on Free Android Applications. *IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 301-310.
- Velleman, P. F., & Hoaglin, D. C. (1981). *Applications, Basics, and Computing of Exploratory Data Analysis*. Duxbury Press.
- Yin, R. (2009). *Case study research: Design and methods*. Sage publications.