Advances in Economics, Business and Management Research, volume 160
Proceedings of the International Conference on Business and Management Research (ICBMR 2020)

Big Data Analysis of Paid and Free Applications in Google Playstore and Apple App Store to Know Application Characteristics and Monetization Opportunities for New Startup in Indonesia

Joshua Aritonang[1] Rofikoh Rokhim[2,*]

[1] *Faculty of Economics and Business Universitas Indonesia*
[*] *Corresponding author. Email: rofikoh.rokhim@ui.ac.id*

ABSTRACT
In this paper, we research the performance of Apple App Store and Google Play Store in Indonesia. In this paper, we compare the distribution of each application category in their respective markets. Researchers are interested in application performance because of the increasing number of internet users and applications, especially Indonesians. The researcher examines the mobile app's total download per day based on its rank in the respective market. The first ranked app on the Apple App Store Indonesia has a total daily download of 289 times. Simultaneously, the first ranked application on Google Playstore Indonesia has a total download of 17,106 times.

Keywords: *Internet, Smartphone, Mobile App, Google Play, Apple App Store.*

1. INTRODUCTION

Indonesia is one of the countries with the largest internet users in the world. According to a survey from APJII (2018), internet user penetration in Indonesia has reached 171.17 million. If seen from the total population of Indonesia, this amount is at a percentage of 64.8%. In 2017, internet user penetration in Indonesia totaled 143.26 million [1]. Data shows an increase in the number of internet users in Indonesia by more than 27 million. The magnitude of the increase in 1 year proves that Indonesia's business potential in using services via the internet.

At first, mobile applications were developed for entertainment purposes. However nowadays, they have even reached essential industries such as health, finance, and other industries [2]. A report linked by [3] provides information that Google Play is the most popular software market for the Android operating system. Google Play offers more than 2.8 million applications for various domains. The iOS operating system, through the App Store market, offers more than 2.2 million applications [3]. Google Play and the App Store offer applications that are numerous and varied. In fact, some applications are exclusive only on each operating system.

The researcher collected several studies that form the basis of the start of this study. The first study was a study from [4], discussing the measurement and forecasting of game applications' engagement. In that study, it examines the effect of diversifying product developers' application lines on their performance. It explains the type of platform ecosystem (open versus closed) on which they commercialize applications [5]. The study provides a basic idea in the use of descriptive statistics as an initial description. It uses the OLS regression method on the calculation of daily income with the shape and scale parameters.

The second study is [6]. This second study discusses the income implications of mobile visits, using an online travel agent example. The researcher refers to the article because it compares mobile vs. desktop with income as an indicator. The study results in the report that mobile traffic provides a significant increase in revenue compared to desktops. The second result of the study states that mobile channel traffic has a more significant short-term effect than desktop [6]. That study provides a strong basis that smartphones have better income projections than on the desktop as a whole.

The third study [7], discusses monetization by measurement and forecasting in mobile game

applications. This research found that there is good potential for game-based in-app purchases [7].

The fourth study is [8], discussing the factors that influence mobile apps' success. This study generates the opinion that applications with high ratings have a low probability of being in the top 50 apps on the App Store. An example is Netflix, which is a top-ranking application, but only has a user rating of 3.5 out of 5 points maximum. In the latest research, the basis for the researchers in using the OLS method is the research conducted by [9]. The gap taken from these studies is whether this calculation is applicable in the mobile app market in Indonesia. Can you know the total downloads per day? How is the situation inside those markets?

2. LITERATURE REVIEW

This section will explain the theoretical basis for conducting this research. This research builds upon several theories that will be used, carried out on several references.

2.1. Mobile App on Smartphone

An application on a smartphone is a vehicle or an essential tool in meeting diverse customer satisfaction of their needs and interests [10]. Applications on smartphones can make users addicted because they are presented very interestingly and can engage many users [10]. According to survey results, most people see and check smartphones about 85 times a day and touch the smartphones 2617 times [11]. According to [11], there is a blurred line between smartphones when we carry out personal and professional activities. In the same book, [11] revealed that mobile platforms have many capabilities that desktops do not have, which dramatically changes consumers' experience and utility. In this explanation, a brief conclusion can be drawn that the potential for application development on smartphones is still vast, with business capabilities that can yet be explored, even combined. Furthermore, some relevant works and articles that the reader might find useful such as [12], [13], [14], [15] and [16].

2.2. Big Data

Big Data has a keyword that is “big”. The word "big" in the term refers to at least 3 different directions, namely [17]:

- Velocity

  At this time, of course, we can create large amounts of data (structured, unstructured, and semi-structured data) that require considerable resources to carry out the process if it still uses conventional methodology. Therefore, it is necessary to build new architectures and tools to process large and unstructured data so that the analysis process can be faster

- Complexity

  The term complexity in big data usually comes from the observation and analysis of the unstructured data because of the natural conditions taken from smartphones, sensors, social media, internet search, e-mail, GPS, and others. This problem has become an event experienced by researchers every day. Two methods make it possible to reduce the data dimensions, namely data projection and variable selection. These methodologies have proven effective for broad issues and their relevance to big data

- Big Samples

  Big data certainly has variations from large populations. This volume of the data is inseparable from the activities that make the information very varied.

The ‘Big Data’ term has been very popular nowadays. Readers might find this an interesting topic in [18], [19], [20] and [21].

2.3. Pareto Distribution (Pareto Law)

This research will allude to Pareto's distribution by the assumption of sales related to ranking on the Apple App Store Indonesia and Google Play Store Indonesia. According to [9], sales and ranking in each application have an assumption related to the Pareto or power-law distribution method, which explains that the small number of application products available in the application market represents a large portion of the population in the application market. According to the Pareto distribution, the researchers applied the formula used in predicting the sale of applications in the Apple App Store Indonesia and Google Playstore Indonesia. Following is the procedure described as follows:

$$sales = b \times (rank) - a + \varepsilon \quad (1)$$

In this formula, b represents the scale parameter, and a represents the shape parameter. Both of these values will be obtained through OLS regression results. In principle, Pareto is famous for the term 80%: 20%, in which 20% of the population has the highest activity or role.

3. RESEARCH METHODOLOGY

In the process, data which has been obtained from the source and then divided into several columns, namely Name, Category, Rating, Review, Size, Installation, Type, Price, Content, and Genre. These columns’ use is useful for determining the characteristics of the application based on the column that has been prepared. The data used is the population of applications in Google Play and Apple App Store.

3.1. Dependent Variable

The dependent variable is a variable that is measured to find out the influence of other variables. According to [22], it is a variable affected and becomes a result due to independent variables' impact. In this study, researchers choose the dependent variable of the ranking of applications based on the top paid application in each Apple App Store Indonesia and Google Play Store Indonesia (see Table 1). This decision is based on a journal reference from [9], and recent research using the same method, namely research on the diversification of application developers on the App Store and Google Playstore [5]. Researchers also analysed the formula in [9] using Apple App Store Indonesia and Google Playstore Indonesia data, in which the dependent variable was ranking applications on top-grossing applications and top paid applications.

3.2. Independent Variable

Independent variables are variables that can influence and cause changes in a dependent variable. According to [22], the independent variable is the variable that causes the change or the occurrence of the dependent variable. In this research, the researcher decides the variables that refer to the study conducted by [9], namely the application ranking on the top-grossing application and price on each Apple App Store Indonesia and Google Play Store Indonesia (see Table 2). This variable is adjusted to the scraping results conducted on the two application markets, namely Apple App Store Indonesia and Google Playstore Indonesia. The researchers also analyzed the formula in [9] research using Apple App Store Indonesia and Google Playstore Indonesia data

3.3. Method

This study analyses activities carried out on the object carefully using multiple linear regression tools with the Ordinary Least Square (OLS) regression. The researcher will adjust the formula used in this study by including two calculation schemes based on [5] and [9].

Table 1. Dependent Variables

Dependent Variables	Variables
Researcher's Scenario	Top Paid App
Garg & Telang's Scenario	Top Grossing App

Table 2. Independent Variables

Independent Variables	Variables
Researcher's Scenario	Price, Top Grossing App
Garg & Telang's Scenario	Price, Top Paid App

There are two regression formulas which form the basis of this research, as follows:

$$Log(r_g) = \beta 0 + \beta 1 \times Log(r_p) + \beta 2 \times Log(P) \quad (2)$$

$$Log(r_p) = \beta 0 + \beta 1 \times Log(r_g) + \beta 2 \times Log(P) \quad (3)$$

Next will be an analysis of the proposed model with adjustments to the Apple Store Indonesia and Google Playstore Indonesia by analyzing the scale parameters and shape parameters. This calculation aims to determine the prediction of the download with the assumption that the Pareto distribution is following the reference research conducted by [9].

$$a_g = -1 \times (1/\beta_2) \quad (4)$$

$$a_p = -1 \times (\beta_1/\beta_2) \quad (5)$$

$$b_p = \left(\sum_{r_{p=1}}^{N} d_{r_p}\right) \div \left(\sum_{r_{p=1}}^{N} r_p^{-a_p}\right) \quad (6)$$

$$b_g = exp\ exp\left(-1 \times \left(\frac{\beta 1}{\beta 2}\right)\right) \times \left(\sum_{r_{p=1}}^{N} d_{r_p}\right) \div \left(\sum_{r_{p=1}}^{N} r_p^{-a_p}\right) \quad (7)$$

The four formulas above will be used to calculate predictions from total daily income using a formula from research by [9] adjusted to Apple Store Indonesia and Google Playstore Indonesia. After getting the results on these four values, we put the values into the distribution formula, as follows:

$$d_{r_p} = b_p \times r_p^{-a_p} \quad (8)$$

Table 3. Hypothesis Formulation F-stat.

Hypothesis	F-stat
Garg & Telang	There is a significant influence between Price and Rank Paid together on the Grossing Rank.
Researcher's method	There is a significant effect between Price and Rank Grossing together on Rank Paid

Table 4. Hypothesis Formulation t-stat.

Hypothesis	t-stat
Garg & Telang	Partially there is a significant influence between the Rank Paid variable on the Grossing Rank
	Partially there is a significant influence between the Price variable on the Grossing Rank
Researcher's method	Partially there is a significant influence between the Grossing Rank and the Paid Rank
	Partially there is a significant effect between Price and Rank Paid

In this study, the results that will be evaluated are the values of f-statistics, and t-statistics.

4. RESULTS

This chapter will explain the results of exploration and calculation. Table 5 and Table 6 show the Exploratory Data Analysis and list the best category based on the apps sold in the respective market. Table 1 shows the Apple App Store results with parameters that have been determined by researchers — showing the results that the application category with the largest population is Lifestyle and followed by Photo and Video. Table 2 shows the effects on the Google Play Store with parameters that have been decided by researchers. Displaying the products that the application category with the largest population is games and followed by education.

Table 7 and Table 8 below show the regression results, after which the results of the shape and scale parameters are entered into the daily download distribution formula. It illustrates that based on the calculation using the regression formula from [9], at the

Table 5. Exploratory Data in Apple App Store.

EDA	Category
Total of Applications per Categories	Lifestyle
	Photo & Video
	Entertainment
	Games
Total of Applications per Price Categories	Free
Average Size per Application Categories	Games
	Education
Total of Reviews of Application Categories	Photo & Video
	Music
Average Rating of Application Categories	Games
	Photo & Video
	Productivity

Table 6. Exploratory Data in Google Play Store.

EDA	Kategori
Total of Applications per Categories	Games
	Education
	Tools
	Entertainment
Total of Applications per Price Categories	Free
Average Size per Application Categories	Games
	Parenting
Total of Reviews of Application Categories	Tools
	Game Action
Average Rating of Application Categories	Books & References
	Music and Audio
	Personalization

Table 7. Apple App Store Download Formula.

Apple App Store	Daily Download Formula
Garg & Telang (2013)	$289{,}187 \times r_p^{-01279}$
Researcher's Method	$289{,}187 \times r_p^{-0{,}53}$

Table 8. Google Play Store Download Formula.

Google Play Store	Daily Download Formula
Garg & Telang (2013)	$17.106{,}97 \times r_p^{-0{,}89}$
Researcher's Method	$17.106{,}97 \times r_p^{-0{,}79}$

Apple App Store Indonesia, the total daily downloads for the app at the first rank are 289 times more than the app at the ranking of 100. The calculation uses a regression formula from [9] to illustrate that on Google Playstore Indonesia, the total daily downloads for the app at the first rank are 60 times more than the app at the ranking of 100. The calculation used the regression formula from the researchers' proposal illustrates that at Apple App Store Indonesia, the total daily downloads for the app at the first rank are 12 times more than the app at the ranking of 100. The calculation using the regression formula from the researchers' proposal illustrates that on Google Play Store Indonesia, the total daily downloads for the app at the first rank are 38 times more than the app ranking at the ranking of 100.

From these results, it can be illustrated that the visitation from Google Playstore Indonesia is more than the Apple App Store Indonesia. This visitation is illustrated from the total downloads generated using the revenue formula proposed by [9]. However, the number of verified applications in the Google Playstore Indonesia market is more significant than its competitors. Thus, it can be concluded that there is a higher competition between application developers on Google Play Store than Apple App Store Indonesia.

Table 9 and Table 10 show the results of the F and T statistical tests. In the table, it can be seen that the results of the test of the data using the hypothesis that has been determined by the researcher give an illustration that the data being tested are officially registered applications on the Apple App Store and Google Play Store Indonesia. The data will be a representation of the information that has been tested using OLS regression.

5. DISCUSSION

The first regression proposed by [9], at the Apple App Store Indonesia and Google Playstore Indonesia, obtained the result that the price and ranking variables paid did not significantly influence the grossing rank variables simultaneously or per variable. The result is evident from both the Statistical F and t statistical tests, producing insignificant values. The second regression

Table 9. Hypothesis Testing Results for F-statistic.

F-Statistic	Score	Results
Apple App Store		
Garg & Telang (2013)	0,1919	H0 Accepted
Researcher's Method	0,6572	H0 Accepted
Google Playstore		
Garg & Telang (2013)	4,65E+05	H0 Accepted
Researcher's Method	4,65E+05	H0 Accepted

Table 10. Hypothesis Testing Results for t-statistic.

T-Statistic	Score	Results
Apple App Store		
Garg & Telang (2013)		
Rank Paid - Rank Grossing	-0,522	H0 Accepted
Price - Rank Grossing	-0,438	H0 Accepted
Researcher's Method		
Rank Grossing - Rank Paid	-0,522	H0 Accepted
Price - Rank Paid	-1,056	H0 Accepted
Google Playstore		
Garg & Telang (2013)		
Rank Paid - Rank Grossing	964,652	H0 Rejected
Price - Rank Grossing	0,307	H0 Accepted
Researcher's Method		
Rank Grossing - Rank Paid	964,652	H0 Rejected
Price - Rank Paid	-0,259	H0 Accepted

proposed by researchers at the Apple App Store Indonesia and Google Playstore Indonesia obtained the result that the price variable and grossing ranking did not significantly affect the variable rank paid simultaneously or per variable. The result is evident from the F Statistical test, producing insignificant nominal values. However, in the t-statistical test, there is a significant influence between store rank grossing and rank paid.

From this finding, new companies, or often called startups, can choose the right strategy to enter the Apple App Store Indonesia or Google Play Store Indonesia market. Besides, the new company that will make the application to be more concerned with the value aspects offered in the application than the price consideration because the price of an application does not significantly affect store paid rankings and grossing ranks. Estimates from the researchers that most applications that are on top-grossing are applications that have a free price or free. The implementation can be seen and compared from the results of the hypothesis test on the F-statistic test and t-statistic.

6. CONCLUSION

This study derived several findings from a population taken from the Apple App Store and Google Play Store. The results obtained can give an idea to the application development companies that will or have been involved in the application market. This study has several limitations. First, the data is only taken on January 10, 2020. Second, the data taken is in the form of a population on the Apple App Store Indonesia and Google Playstore. Third, the manual process for scrap coding and cleaning activities.

For further research, an area can still be explored, which is a picture of revenue per category in the Apple App Store Indonesia and Google Play Store Indonesia. This can be formed using a regression formula that researchers do. Businesses in the application and technology industries can reconsider the revenue strategy that must be determined, due to the insignificant effect of the sales price stated on the Apple App Store Indonesia and Google Play Store Indonesia with application revenue.

REFERENCES

[1] APJII, "Penetrasi & Profil Perilaku Pengguna Internet Indonesia Tahun 2018," Jakarta, 2019. [Online]. Available: www.apjii.or.id.

[2] H. Muccini, A. Di Francesco, and P. Esposito, "Software testing of mobile applications: Challenges and future research directions," in 2012 7th International Workshop on Automation of Software Test (AST), 2012, pp. 29–35, doi: 10.1109/IWAST.2012.6228987.

[3] Statista, "Number of apps available in leading app stores as of 1st quarter 2020.," 2020. https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ (accessed Mar. 20, 2020).

[4] P. Roma and D. Ragaglia, "Revenue models, in-app purchase, and the app performance: Evidence from Apple's App Store and Google Play," Electron. Commer. Res. Appl., vol. 17, pp. 173–190, 2016, doi: https://doi.org/10.1016/j.elerap.2016.04.007.

[5] P. Roma and M. Vasi, "Diversification and performance in the mobile app market: The role of the platform ecosystem," Technol. Forecast. Soc. Change, vol. 147, pp. 123–139, 2019, doi: https://doi.org/10.1016/j.techfore.2019.07.003.

[6] Y. Zhao, P. Song, and F. Feng, "What are the revenue implications of mobile channel visits? Evidence from the online travel agency industry," Electron. Commer. Res. Appl., vol. 36, p. 100865, 2019, doi: https://doi.org/10.1016/j.elerap.2019.100865.

[7] O. Rutz, A. Aravindakshan, and O. Rubel, "Measuring and forecasting mobile game app engagement," Int. J. Res. Mark., vol. 36, no. 2, pp. 185–199, 2019, doi: https://doi.org/10.1016/j.ijresmar.2019.01.002.

[8] W. N. Picoto, R. Duarte, and I. Pinto, "Uncovering top-ranking factors for mobile apps through a multimethod approach," J. Bus. Res., vol. 101, pp. 668–674, 2019, doi: https://doi.org/10.1016/j.jbusres.2019.01.038.

[9] R. Garg, R., & Telang, "Inferring App Demand from Publicly Available Data. MIS Quarterly," MIS Q., vol. 37, no. 4, pp. 1253–1264, [Online]. Available: https://misq.org/inferring-app-demand-from-publicly-available-data.html.

[10] W.-K. Tan, Y.-J. Hsiao, S.-F. Tseng, and C.-L. Chan, "Smartphone application personality and its relationship to personalities of smartphone users and social capital accrued through use of smartphone social applications," Telemat. Informatics, vol. 35, no. 1, pp. 255–266, 2018, doi: https://doi.org/10.1016/j.tele.2017.11.007.

[11] S. Van Till, "Chapter 1 - From Packages to People," S. B. T.-T. F. T. F. D. S. Van Till, Ed. Butterworth-Heinemann, 2018, pp. 1–13.

[12] S. Barnett, I. Avazpour, R. Vasa, and J. Grundy, "Supporting multi-view development for mobile applications," J. Comput. Lang., vol. 51, pp. 88–96, 2019, doi: 10.1016/j.cola.2019.02.001.

[13] F. Keating, "Got a smartphone? You probably check Facebook fourteen times a day," 2013. http://www.dailymail.co.uk/sciencetech/article-2300466/Smartphone-users-check-Facebook-14-times-day-admit-looking-movies.html (accessed Mar. 20, 2020).

[14] A. A. Lashitew, R. van Tulder, and Y. Liasse, "Mobile phones for financial inclusion: What explains the diffusion of mobile money innovations?," Res. Policy, vol. 48, no. 5, pp. 1201–1215, 2019, doi: https://doi.org/10.1016/j.respol.2018.12.010.

[15] A. Mendoza, "Why a mobile app does not make sense.," 2017. https://www.mobilemarketer.com/ex/mobilemarketer/cms/opinion/columns/8605.html (accessed Mar. 20, 2020).

[16] A. Mendoza, Mobile User Experience. Elsevier, 2013.

[17] J. L. Torrecilla and J. Romo, "Data learning from big data," Stat. Probab. Lett., vol. 136, pp. 15–19, 2018, doi: https://doi.org/10.1016/j.spl.2018.02.038.

[18] W. H. Inmon, D. Linstedt, and M. Levins, Data Architecture: A Primer for the Data Scientist (2nd Edition). 2019.

[19] B. K. Williams and S. C. Sawyer, Using Information Technology: A Practical Introduction to Computers & Communications : Complete Version, 6th ed. McGraw-Hill Technology Education, 2009.

[20] D. Talia, P. Trunfio, and F. Marozzo, "Chapter 1 - Introduction to Data Mining," in Computer Science Reviews and Trends, D. Talia, P. Trunfio, and F. B. T.-D. A. in the C. Marozzo, Eds. Boston: Elsevier, 2016, pp. 1–25.

[21] S. Sagiroglu and D. Sinanc, "Big data: A review," in 2013 International Conference on Collaboration Technologies and Systems (CTS), 2013, pp. 42–47, doi: 10.1109/CTS.2013.6567202.

[22] Sugiyono, Metode Penelitian Kuantitatif Kualitatif dan R & D. Gramedia, 2011.