

What Are the Characteristics of Most-Popular Apps?

A Case Study on Free Android Applications

Darshan Pandhi
Computer Science
University of Manitoba
Winnipeg, Manitoba, Canada
pandhid@myumanitoba.ca

1 Introduction

In today's modern reality, there is a mobile app for everything, literally everything. There are apps that help doctors in their profession and at the same time, there are games like S.M.T.H (Send Me To Heaven)¹ that inspires players to throw their phones as high as they can up in the air. However, the sad and honest reality is that majority of the developers are struggling to make their application a success.

This paper is made in contrast to an existing paper "What are the Characteristics of High-Rated Apps? A Case Study on Free Android Applications" [1]. While this original paper focusses on app ratings as a metric for success, my curiosity inspired me to explore a different success metric - number of downloads.

After all, you can have the best application in the world but if no one is downloading, you will not make a single cent. Besides, there are plenty of examples out there of successful apps that have low ratings. For instance, Google Classroom has a rating of 2.5 with 100M+ downloads, TikTok has a rating of 1.6 with over 1 billion downloads, etc. Hence, I decided to go with number of downloads over user ratings for the context of this paper.

Like the existing paper, I will be looking at 28 factors along 8 dimensions to understand how heavily downloaded apps are different from the less downloaded apps. The eight dimensions include: 1) size of app, 2) complexity of code, 3) dependence on libraries, 4) quality of library code, 5) complexity of user interface, 6) requirements on users, 7) marketing effort, and 8) category of app. Using a random forest classifier, I will also be investigating the most influential factors of a popular app.

It turns out that the size of an app, the number of promotional images that the app displays on its web store page and the target SDK version of an app are the deciding factors for a high-rated app. So, I will be comparing these to the results I obtain from this paper i.e. when the metric is the number of downloads instead.

This research will help future developers to focus on what matters the most. In the words of Yuan et al. [1], knowing the "DNA"

(i.e., characteristics) of most-popular apps is the first step towards successful development and evolution of apps.

2 Related Work

Without doubt, the first piece of related work is obviously the original paper, "What are the Characteristics of High-Rated Apps? A Case Study on Free Android Applications" [1], that inspired me to investigate this space further.

Harman et al. examine the correlations between application's description, download count, and average user rating among thousands of BlackBerry apps [2].

Another short paper titled "The Analysis of Factors Affecting Android App Download Numbers" limits itself to just a theoretical standpoint whilst discussing about 4 different factors affecting download numbers [3].

In the same vein, "Investigating the android apps' success: An empirical study" is a study concerning 474 free Android Apps that finds out that factors such as the number of users' reviews provided for an app, app's category and size appear to have an impact on the app's success [4].

My study is similar to the above studies in that I too will examine the relationships between different factors of an app and the number of downloads. However, I will consider a slew of other factors that have not been investigated before. Also, none of them compare these factors and rank them according to the strength of their impact on the number of downloads of an app like I propose to do.

In regard to the Apple App Store, through their journal article, Rajiv and Rahul develop various strategies that allow them to use public data to infer the rank-demand(number of downloads) relationship for paid iOS apps [5].

3 Data Collection

The paper "Same App, Different App Stores: A Comparative Study" is about a large-scale empirical study on 80,169 iOS and Android mobile app-pairs [6]. Its open-source GitHub repository

¹ <https://play.google.com/store/apps/details?id=com.carrotpop.www.smth&hl=en>

allows one to directly import all the data (available in the form of JSON Objects) in a mongoDB instance.²

Our 28 factors, on the other hand, are mined from information recovered from an app's binary, i.e., its Android Application package (APK), and its Google Play store page.

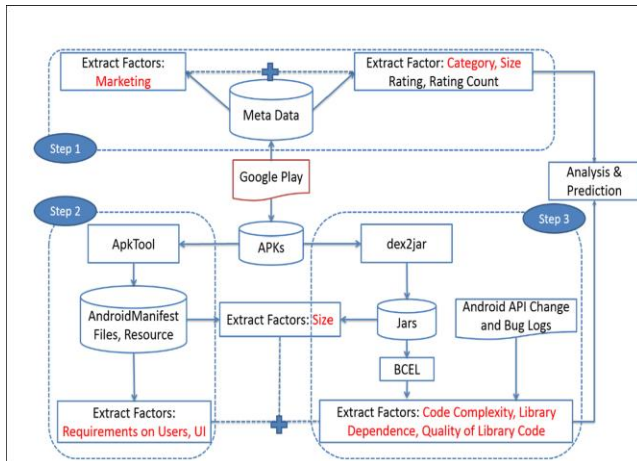


Figure 1: An overview of the planned data mining process

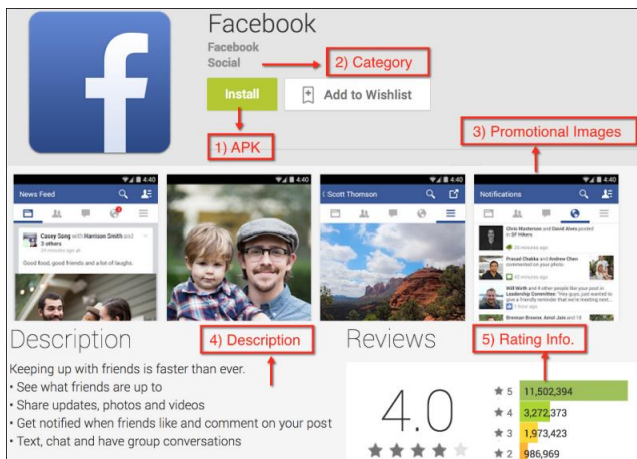


Figure 2: A sample Google Play Store's app page

As shown in Figure 1, factor extraction involves three major steps:

- **Step 1: Process app's store page**
We can extract all information from the app page in the store as shown in Figure 2.
- **Step 2: Process app's AndroidManifest.xml file and resources**

From the app store, we can obtain its APK file. This can be mined using "apktool"³

- **Step 3: Process app's class files and libraries**

We can use any Java Decompiler to view the source code and extract further information like code complexity and information about libraries.

4 Research Questions

The following are 3 research questions that this paper aims to answer:

- **RQ1: Is there a relationship between each factor and the number of downloads?**

The existing paper concluded that generally, high-rated apps have larger sizes, more complex code, more requirements on users, more marketing efforts, more dependence on libraries, and adopt higher quality Android APIs.

Likewise, to understand the contribution of each factor to the number of downloads, I will create two groups of apps from the Google Play store, one consisting of popular apps (more than 1M downloads) and the other of not so popular apps (less than 5K downloads). For each of the 28 factors, I will apply the Mann-Whitney U test to examine whether heavily downloaded apps are statistically significantly different from less downloaded ones. I will also compute the effect size of the differences.

- **RQ2: What are the important factors that could indicate, with high probability, that an app will have a high number of downloads?**

The answer to this question will help us in predicting that given a set of factors, whether the app will be popular or not. And this is akin to reality, wherein number of downloads is impacted by multiple factors rather than just one. In other words, it will allow us to compare and contrast the importance of the different factors.

These can be achieved by inputting these factors to a random-forest classifier. I will be applying correlation and redundancy analyses to better model the combined impact of the factors on the number of downloads.

- **RQ3: How do the characteristics of Most-Popular Apps differ with regards to those of High-Rated Apps?**

This is the question that popped up in my mind when I first read the original paper. What would happen if they

² GitHub Repo: <https://github.com/saltlab/Minning-App-Stores>

³ <https://code.google.com/p/android-apktool/>

had chosen number of downloads instead of app ratings? Will the results still be the same?

The overall approach contains 4 main steps (as mentioned in RQ2):

Stage 1: Correlation analysis

Stage 2: Redundancy analysis

Stage 3: Building a random-forest classifier and testing it

Stage 4: Calculating factor importance through multiple runs

These will allow us to denote a variable importance value for each factor. This juxtaposition of not just the most impactful factors but also the least influential ones will help us gain the best of both worlds. As a matter of fact, it will set the stage up for further analysis of other app success metrics like (analyzing) user reviews/comments for instance.

5 References

- [1] Y. Tian, M. Nagappan, D. Lo and A. E. Hassan, "What are the characteristics of high-rated apps? A case study on free Android Applications," 2015 IEEE International Conference on Software Maintenance and Evolution (ICSME), Bremen, 2015, pp. 301-310, doi: 10.1109/ICSM.2015.7332476.
- [2] M. Harman, Y. Jia, and Y. Zhang. App store mining and analysis: Msr for app stores. In Mining Software Repositories (MSR), 2012 9th IEEE Working Conference on, pages 108–111. IEEE, 2012.
- [3] Wang, Yin Yin. "The Analysis of Factors Affecting Android App Download Numbers." Advanced Materials Research, vol. 912–914, Trans Tech Publications, Ltd., Apr. 2014, pp. 1428–1431. Crossref, doi:10.4028/www.scientific.net/amr.912-914.1428.
- [4] L. Guerrouj and O. Baysal, "Investigating the android apps' success: An empirical study," 2016 IEEE 24th International Conference on Program Comprehension (ICPC), Austin, TX, 2016, pp. 1-4, doi: 10.1109/ICPC.2016.7503724.
- [5] Garg, Rajiv and Telang, Rahul, Inferring App Demand from Publicly Available Data (May 1, 2012). MIS Quarterly, Forthcoming, Available at SSRN: <https://ssrn.com/abstract=1924044> or <http://dx.doi.org/10.2139/ssrn.1924044>
- [6] M. Ali, M. E. Joorabchi and A. Mesbah, "Same App, Different App Stores: A Comparative Study," 2017 IEEE/ACM 4th International Conference on Mobile Software Engineering and Systems (MOBILESoft), Buenos Aires, 2017, pp. 79-90, doi: 10.1109/MOBILESoft.2017.3.