Alvaro Hu

October 22, 2018

SpringBoard

Google Play Store Analysis Proposal

Almost 20 years into the 21st century, almost everything can be done in a virtual space. And many of these virtual spaces can be downloaded from the internet in a matter of seconds in the form of an app. It's hard to believe that only 25 years ago, the internet was not accessible by even the richest people in society. You want to take notes or make a grocery list? There's an app for that. You want to watch Football on your phone while waiting for your haircut? There's an app for that. You want to keep track of your stock holdings to make sure you don't lose money on an investment? There's an app for that. There seems to be an app for everything already, yet all over the world, especially in regions like Silicon Valley or metropolitan areas like Dallas/ Fort Worth, there are innovators creating new and improved versions of apps every day. With my analysis, I will be able to steer app developers towards making apps that will yield in the highest returns, both in amount of installs and in revenue earned.

I will use the two CSV files available on Kaggle.com. The first one provides me with over 10,000 different apps, and the second provides me with over 60,000 different reviews for the various apps. These are not every review for each app, rather it has provided me with up to 100 of the "most relevant" reviews for each app. Some of the more important features in the first dataset are the name of the app, number of reviews, rating out of 5, the category the app is in, the price of the app, and the number of installs. The important features in the second CSV file are the review and the sentiment of the review, which is either positive, negative, or neutral (which has been preprocessed, though I may want to do my own sentiment analysis to practice my own skills). I can use these features to figure out the answers to my questions.

My client is every single developer out there looking for what their next project should be focused on, depending on if they want recognition (downloads) or money (revenue). In an ideal world, all developers and mobile developers would be working on some sort of “passion project”, but the truth of the matter is that many of them are also willing to adapt what they work on to more suit the needs and desires of the general public. So by providing an in-depth analysis of what genres and types of apps give the most downloads in general, then it would alleviate some of the decision making process for these developers when they are deciding what to work on, or what project they want to help fund or hop in on.

The way I will be approaching this problem is by checking either the mean or the median installs for each of the categories of apps (which is around 30). I will also be basing the revenue aspect of this project on the proportions when compared to the entire category. For example, I will check to see how the average or mean amount of installs for apps in the Productivity category changes when the app is no longer free, so that I can adequately recommend to the developer whether or not it would be safe to charge for the service or app. As for whether or not negative reviews impact sales and installs, I can join the two datasets on the name of the app, and figure out if, overall, a majority negative reviews has the outcome of having less downloads, and if that depends on if the app is paid or if it is free.