Google Play Store: Health & Fitness Report and Analysis

1. Introduction

This report will include data and information regarding the Google Play Store and will be used to answer a user story. The data utilized for answering the user story consists of 10,841 rows and 13 columns all related to Google Play Store statistics and metrics. These ranged from the category of the app, the overall rating, installs, current version, price, and other similar factors. Meaning there are 10,841 apps and relevant information contained within the data.

The question we are tasked with is, “User would like to know the mean price and mean number of installs for Health and Fitness ”. In this report I will be discussing how I cleaned the data to present a more clear picture, as well as discuss my hypothesis that the cheaper the app the more downloads it will have. Free having the most downloads.

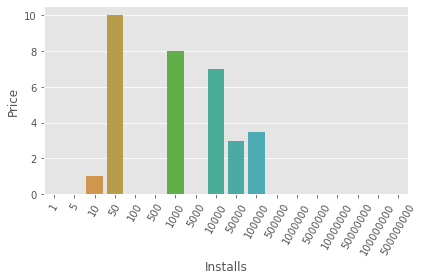
1. Data

Initially I look at the data and realized that there are too many columns for the scope of our analysis. So my first step was to clean the data by removing all of the columns except for: ‘Name’, ‘Category’, and ‘Price’. This allowed me take out some of the white noise in the data. Next I removed all of the categories except ‘Health & Fitness’ because that is the category we are focused on. Now with my dataframe of just the installs and price of all of the ‘Health & Fitness’ apps I started to look at the numbers. Both the price and the installs were object strings, and are therefore unable to perform mathematical operations. I first removed the ‘$’ sign as well as the ‘,’. I then converted the ‘Price’ column from a string to a float. This allowed me to find the mean price for all of the Health & Fitness apps. I then noticed the Installs were large numbers as well as some had a ‘+’. I removed the ‘+’ and ‘,’ and converted the column to float as well as find the mean installs. The numbers were so large that I converted them into millions allowing for better readabliliy.

1. Analysis/Results/Conclusion

The mean price of all the apps is $.20 and the mean installs is 4,642,441 installs. From the data we can show that: 1. People like their health apps, and 2. The means can be a bit misleading because of how top heavy the graph is.

From the results, a staggering 325 out of 341 total apps, or 95.3% of the Health & Fitness apps were priced as free. Of those 325 free apps, 77% have more than 1,000 installs, 71% have more than 10,000 installs, and 57% have more than 100,000 installs. There are only 9 out of the total 341 apps, or 2.6%, that have 10,000 or more installs and cost more than $1. And there are only 4 apps, or 1.1%, that cost more than $1 and have 100,000 installs or more.



The chart above shows that the error bars are too large for the app’s price to show any statistical significance enough to have any weight about how many installs your app will get based on the price. There is however a pretty strong case to suggest if you want an app for the masses it needs to be free.