Matthew Cornelius

02/23/2020

DSC 530

Bellevue University

**Final Summary Paper**

**Statistical/Hypothetical Question**

The data that I have decided to do my project on is a dataset that has the details of over 15,000 entries concerning apps for your mobile devices. The data set has many different aspects to it such as user rating count, average user rating, size of the app, etc.. The question I decided to ask based on the data set is, what are or if any components of an app influences the average user rating?

**Outcome of your EDA**

From the efforts of my analysis of my data set I have not made a complete conclusive decision of what exactly determines the average user rating. However, I have noticed some trends that I do believe are indicators that might help in answering my question. From my findings I think that price and age rating seem to have the strongest influence on the average user rating.

**What do you feel was missed during the analysis?**

I feel the main thing that I might have missed from my analysis, is choosing the correct variables to try and find relationships. I say this due to the fact that multiple times when I was trying to perform for example the linear regression part of my analysis. I kept having a hard time performing this because of a multitude of problems, most of which were errors concerning that something wasn’t matching. During this part I actually received quite a few different types of errors, most of which I just don’t understand at this level in my programing career.

**Were there any variables you felt could have helped in the analysis?**

While I was going over the data trying to figure out what had the strongest influence on the average user rating, an idea to this question did come to mind. A few categories that I felt could have been added to my data set is, the time of duration for the usage of the app. As in how long did they use the app before trying to give it a review. And if and how often the app would ask its user to give the app a rating.

**Were there any assumptions made you felt were incorrect?**

My strongest initial assumption was that price would have an overwhelming large influence on what the average user rating was for an app. However, I do not believe this to be so now. I think the biggest influence on the average user rating was what age was the app intended for.

**What challenges did you face, what did you not fully understand?**

My biggest challenge was in all honestly trying to code everything. This is my second semester in the program, so I have only been coding for around the duration of the start of the program. Most of my struggles were just trying to understand the code I was trying to write. I looked toward the thinkstats book for guidance, but it usually proved fruitless and I ended up having to do a lot of out of class research. This took away the time I would have put into trying to get results from the data.