Alex Yom

Inferential Statistics

After wrangling my dataset and getting it prepared for exploration, I thought it would effective to use inferential statistics to receive a general idea of what the data looks like. There seem to be several factors that I would like to focus more on. For this specific project, I think it would be best to look into the impact categories, content, rating, and type. Other factors such as app size and latest update date do not seem like salient factors so much of the research and exploration will focus on the other features because those factors would project to be potentially very important in being able to find the success of an app.

It seems like there is some correlation between these factors as apps that have a high rating would be downloaded more and certain types of apps such as communication apps are downloaded more often. Apps with low ratings would obviously not be downloaded often, so some positive relationship between ratings and installs would be expected. Preliminary analysis showed that certain categories of apps do get downloaded more often on average. We can dive deeper and explore the relationship between category and installs more thoroughly through statistics and visual representation.

Finding the statistics behind each category of app would illuminate a bigger picture of the totality of data. By reviewing the minimum, mean, quartiles, maximum, of the data, we can see the difference between each category and make better judgements. Using box plots will help visualize this data and if we compare the plots for each category, it would make comparing very easy and useful. Using boxplots could be helpful in assessing different categories because while a certain category may have a greater number of installs in total, another category may have a higher average of downloads which would be useful information for prospective app developers.