Report: act_report

- Create a **250-word-minimum written report** called "act_report.pdf" or "act_report.html" that communicates the insights and displays the visualization(s) produced from your wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

Introduction

After the data is gathered, assessed and cleaned and stored as a csv file, it is then loaded into python to be analyzed to come up with actionable insights. It is the job of a data analyst to come up with this analysis and to back it up with visualizations. The insights that were obtained from analyzing the data are listed below.

Insights

1. Tweets were grouped by the year that they were posted and calculations of the average number of retweets for each tweet in that year were done. It was observed that on average the number of retweets in 2017 were more than twice as much as the retweets in 2016 and more than 6 times the retweets in 2015. This shows how the page became more popular over time.
2. Favorites were grouped by the month that they were posted and calculations of the average number of favorites for each tweet in that month were done.The month with the most average number of favorites is the month of July with the least average number of favorites November. Since the twitter account used is based in the US, it can be assumed that most of the people who interact with the account are from the US. By following this assumption it can be seen that on average Summer months which are June, July and August in the US have more average number of favorites as compared to the winter months in the US which are December, January and February.
3. Analysis was done on the source column of the dataset and it was discovered that more than 98% of the tweets that were posted on the Twitter account were posted using an iPhone and the other ways of posting the tweets were using a Web Client or a Tweet Deck.
4. Analysis was done on the dog stages column of the dataset and it was discovered that most of the dogs in the final dataset do not have a dog stage listed and the dog stage with the least numbers is the floofer with only six dogs recorded as being in this dog stage. If the dogs that do not have a dog stage listed are not counted, the dog stage that is most popular is pupper stage with 118 dogs recorded being in this stage.

Visualizations

1. A bar graph was plotted to see the distribution of the dog ratings in the final dataset. From the bar graph below it can be seen that the 3 most common numerator ratings are 11, 10 and 12 in ascending order. The numerator ratings follow a normal distribution. Figure 1 below is a bar graph showing the distribution of the dog ratings.

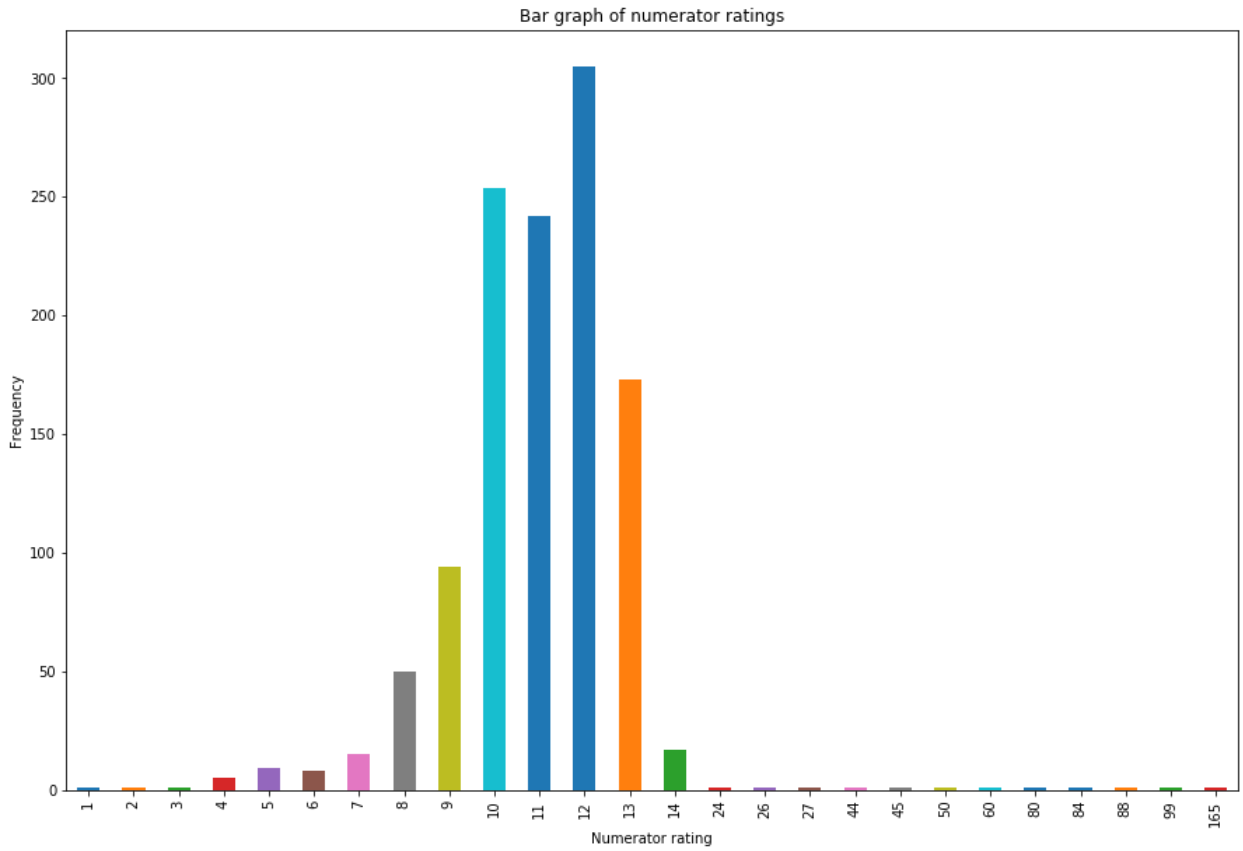

Figure 1: Bar graph of dog ratings

2. A time series was plotted to see the trend of the number of favorites for each tweet with time. It can be seen in the time series below that the general trend is that as time progresses the number of favorites also increases. This could be as a result of the Twitter account being more popular over time. Figure 2 below is a time series plot of the number of favorites for each tweet.

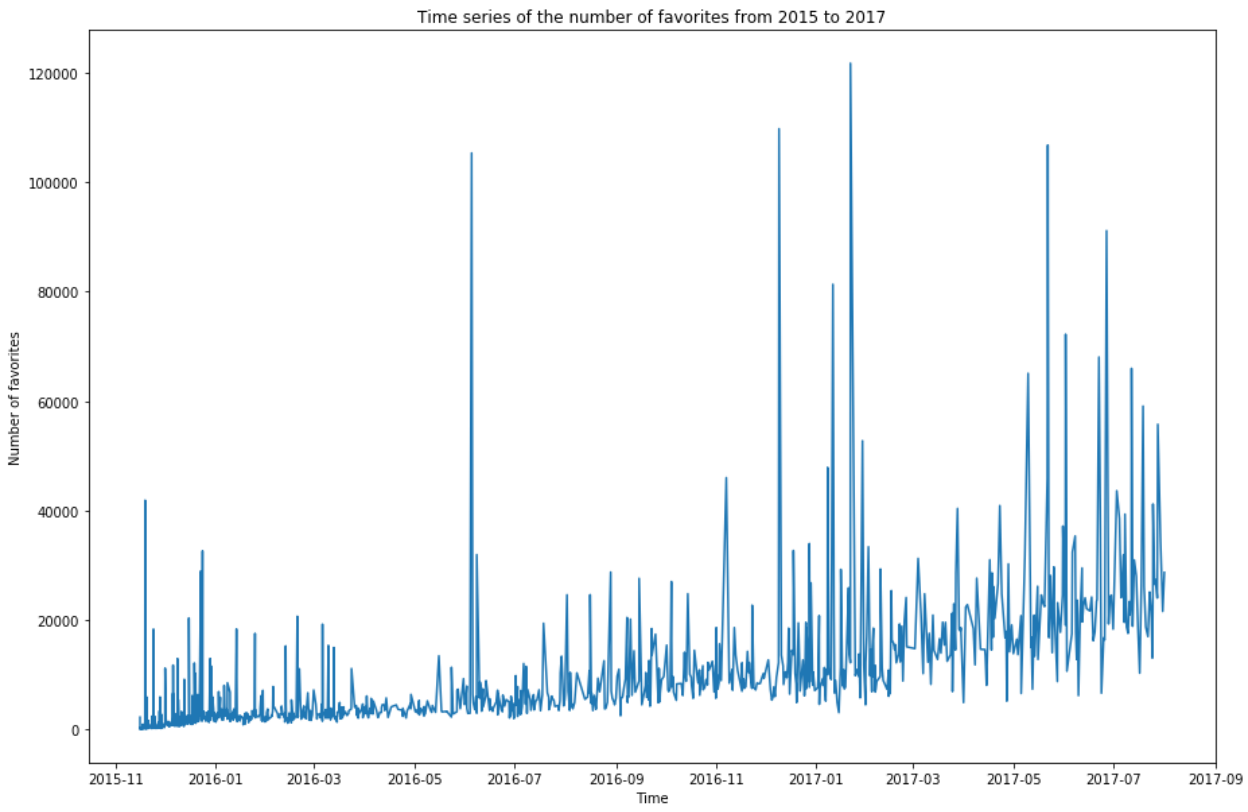

Figure 2: Time series of the number of favorites for each tweet