Miriam Wanjohi

Wrangle_report

This project utilized the 'WeRateDogs' data downloaded from the 'WeRateDogs' Twitter account. The account rates people's dogs on a scale of 1-10 and the ratings are saved as the value over the maximum (10) such that if a dog is rated 3, the rating numerator is saved as 13 with a denominator of 10.

For the data wrangling process, first I had to gather data first by downloading and reading the twitter archive csv data already provided for the assignment. The initial Twitter Archive data had over 5000 tweets but had been filtered to 2356 tweets. I then downloaded the images data programmatically using request library from the Udacity library.

Additional data was needed for favorite and retweet count and for this; I had to query Twitter API using the tweepy library and stored the JSON data into a JSON file which I used to create a dataframe for easy analysis.

The second step of wrangling was assessment, first visually and then programmatically (using code). I recorded most of the areas within the dataset that would be classified as sources of messy and untidy data. I then used this as a guide in the cleaning process which is the last step of wrangling. The clean individual datasets were then merged to form one master dataset which I used for analysis and visualization.

Project Advantages:

This was a fun and challenging project to work on given the reality of it given that the data was from a real life twitter account as opposed to data created for use in a classroom project. It was a good learning experience on learning how to gather data from different API's.

Limitations of the Project:

One major limitation I came across while completing this project is trying to figure out how to query Twitter API for the additional data. It took me nearly a week of writing/editing/researching codes that worked in order to download and read the data. The biggest issue I had was in writing and reading the JSON file line by line.

I am also not sure I am exploiting the dataset into its maximum potential and would like to dig deeper into more exploratory analysis.