Analyzing and Visualizing We Rate Dogs from Twitter

By Ali Hasan

Introduction

Real-world data rarely comes clean. Using Python and its libraries, you will gather data from a variety of sources regarding one of Twitter accounts (We Rate Dogs) and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling. That's the first part. Second part would be to merge all datasets in one master dataset then conduct different types of analyses and visualization. Third part will be consisted of two PDFs. One for the Data Wrangling efforts and the second one for Analyses and Visualizations

Work Done

I spent more than 70% of my time in the first part about data wrangling. Gathering the data from 3 different data sources was the biggest challenge for me as I had a problem setting up tweepy library. Despite of reinstalling environments, python and using all possible conda and pip work arounds to fix compatibility issue, I failed so fix so that I used classroom workstation and it worked just fine. I shouldn't have waited more than 4 days to do so.

Moving forward to assessing data to come up endless quality and tidiness issues (I chose to work on some). By cleaning, creating one unified dataset from the three initial datasets. We started with 2175 observations from the twitter_archive , 2342 from twitter_api, and 2075 from image, we ended down to 1056 observations in the twitter_archive_master.

Findings

The second part about data analyzing and visualizing was relatively easier. I chose to conduct 4 insight analyses and chose to add visualization to everyone. It is a lot easier to see tabular data next to its visualization.

The first couple of analyses were more about the counts of dog's ratings and their averages. I just wanted to find out favorite dogs then wanted to see their names. Golden Retriever is the winner in both cases.

Third insight was about the correlation between favorite_counts and retweet_counts. As I was anticipating the correlation is positive.

Fourth insight was about the counts and percentages of dog_stages. Pupper stage was the largest