



UDACITY

Project 4: Data Wrangling & Analysis (act_report)

By

MOHAMMED IBRAHEEM ALSHONEEFI

Introduction:

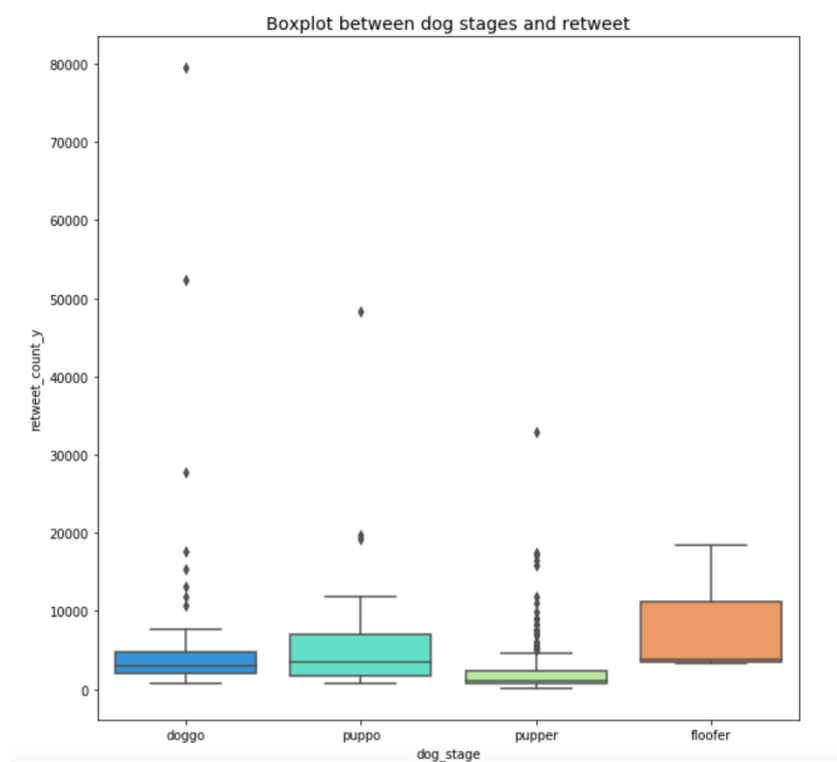
This project (Data Wrangling & Analysis) is a part of Udacity Data Analyst Nano Degree. Moreover, this report involves multiple stages consist of gathering data, assessing data, and clean the data. The first step includes gathering data from various sources – the sources required are from twitter- need to be merged. The next step responsible for accessing the data by take snapshots, build assumptions, and conclude results. Finally, cleaning the data is the last stage that include lots of effort and creative work. Starts with defining the qualities, and tidiness issues. Then start solves each issue separately.

Datasets:

The datasets are collected from twitter API and it were three datasets. Image prediction, twitter archive, and Json file should convert to csv file. The account that collected from is @dog_rate which known as WeRateDogs, this account has 4 million followers that rate the dogs out of 10 in some cases it has more Than 10. Image prediction dataset has the tweet details in addition to the image classifier prediction results. Twitter archive dataset has the tweet details.

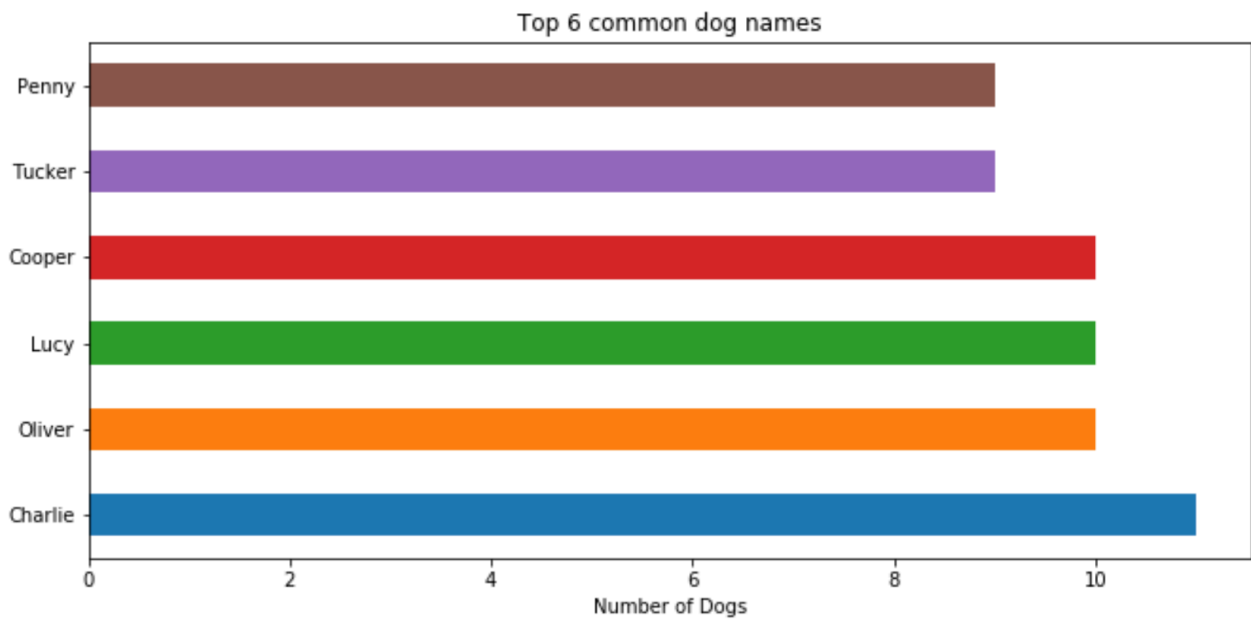
Visualization:

Graph 1:



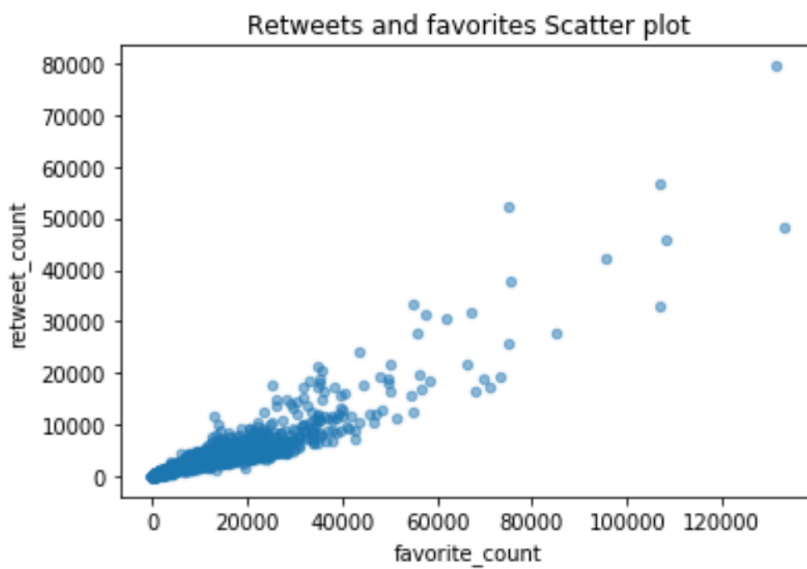
Due to the huge number of the followers for the account it reflects on the activity to the tweets, so we find out that the average of the retweets is 2766, and the maximum number of retweets is 79515.

Graph 2:



Clearly the common name among dogs is Charlie.

Graph 3:



As we see here the strong coloration between the retweet count, and the favorite count the have direct proportion