Wrangle and Analyze Data

By Fahad Alsaeed

Introduction:

This project focuses on wrangling and visualizing a dataset that has been gathered from different sources. In order to explore a twitter account called @WeRateDogs where the account post a tweet of a dog and letting the audience rate.

Project Details:

The tasks of this project are as follows:

• Gathering Data

• Assessing Data

• Cleaning Data

Gathering Data:

The data used in this project was collected from three different sources:

- **twitter_archive.csv** given from Udacity.
- **Image_predictions.tsv** extracted from Udacity sever.
- **tweet.json** scraped from Twitter api.

Assessing Data:

Once the three data files were generated, I programmatically assessed the data using python code and jupyter notebooks. I also manually inspected the .csv files using Excel. In addition, I listed the quality and tidiness issues and addressed them step-by-step.

Cleaning Data:

This part of the data wrangling was divided into three parts:

- Define.
- Code.
- Test.

These three steps were executed on each of the issues described in the assess section (this included 'melting' the dog stages into one column instead of four columns as originally

presented in the twitter archive file and changing data types for some columns and remove another).

Conclusion:

Using the data analysis skills, I learned at udacity, I was able to wrangle the WeRateDogs Twitter data to create an interesting and trustworthy analysis and visualization.

- I have used Python programming language and some of its packages.
- For gathering data there are several packages that help scraping data off the web for example I used Tweepy for gather data from twitter.
- I can Handling, assessing, cleaning and visualizing of data is possible programmatically using code.