# Wrangle and Analyze WeRateDogs Data

By Rawan Alaufi

## Introduction

Every second, on average, around 6,000 tweets are tweeted on Twitter, which corresponds to over 350,000 tweets sent per minute, 500 million tweets per day and around 200 billion tweets per year.[1]

For this project, the goal is to wrangle and analyze the twitter user @dog_rates, also known as WeRateDogs.

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. It was started in 2015 by college student Matt Nelson, and has received international media coverage both for its popularity and for the attention drawn to social media copyright law when it was suspended by Twitter. [2]

## Data wrangling process:

### 1. Gathering Data

I gathered data of this project from three sources.

1. Twitter archive csv file is given by Udacity.

2. Image predictions file is hosted on Udacity's servers I downloaded it programmatically.

3. Querying the twitter API by using tweetpy python library to extract retweets and favorites counts.

### 2. Assessing Data

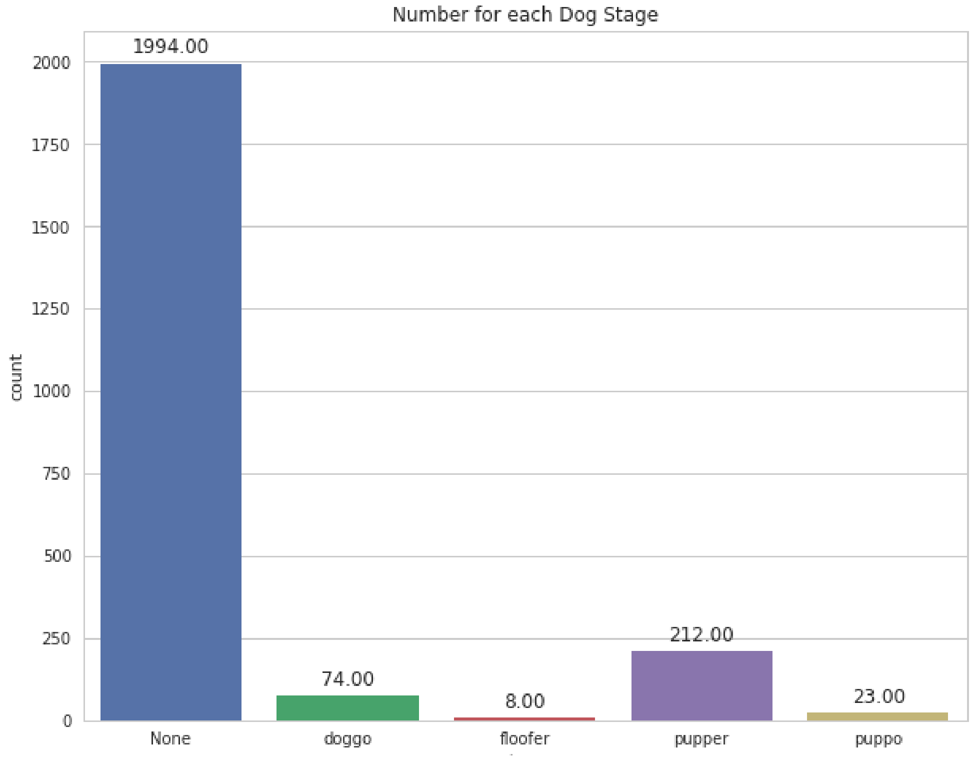
Assessing data is the second step in data wrangling. I detected eight nine quality issues and three tidiness issues while assessing the data.  Data quality issues are content issues and of tidiness, issues are structural issues.

### 3. Cleaning Data

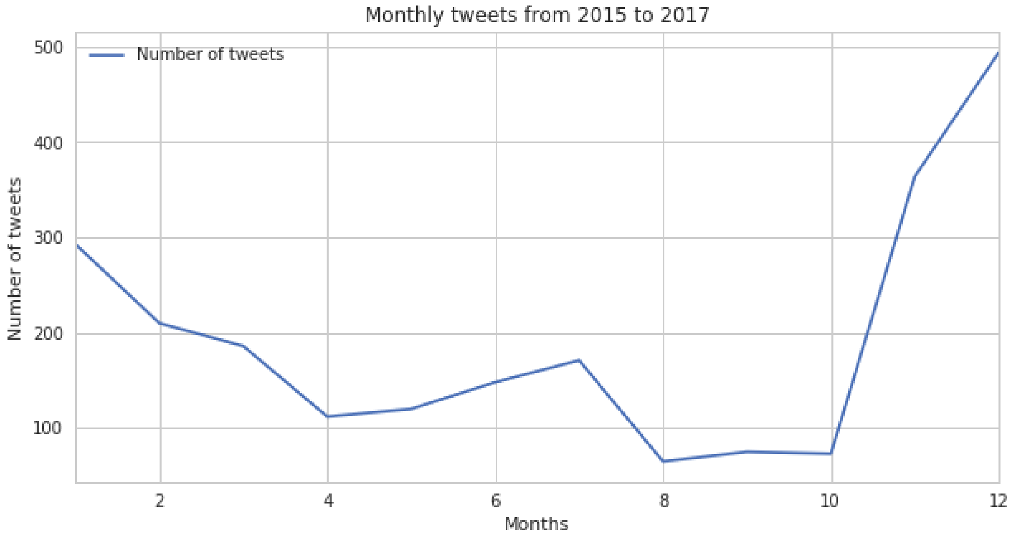
Is the third step in data wrangling. in these phase I fixed all quality and tidiness issues I identified in the assess step.

### Visualizing:

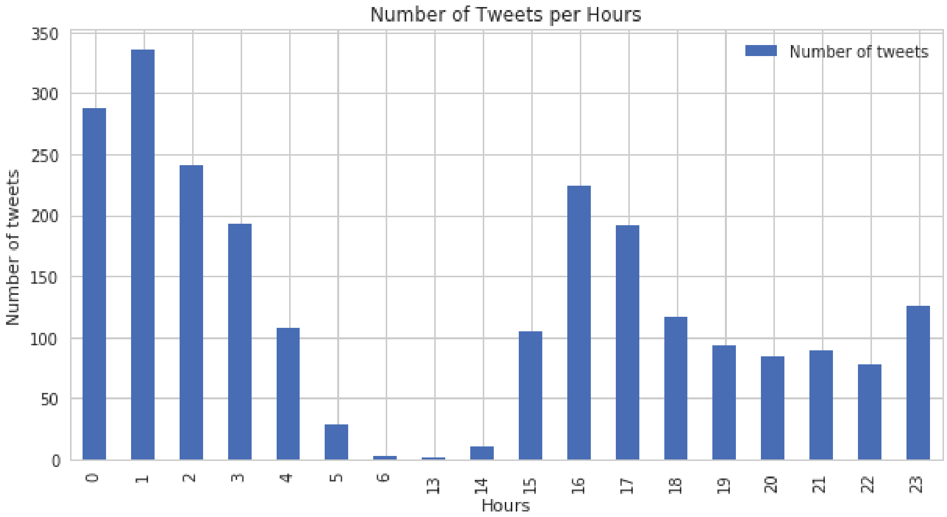
I made some plots from the final data file after applying the whole process of data wrangling.



The above plot shows the number for each dog stage, the most frequent stage of dogs is pupper stage.



The above plot shows the the number of tweets over months, most tweets were in December while fewer tweets were in October.



The above plot shows the number of tweets per hours, most tweets at 1 am.

## References:

1. http://www.internetlivestats.com/twitter-statistics/
2. https://en.wikipedia.org/wiki/WeRateDogs