Wrangle and Analyze Data Project
Act report

By: Mishal AlKhalifa

29th of August 2020

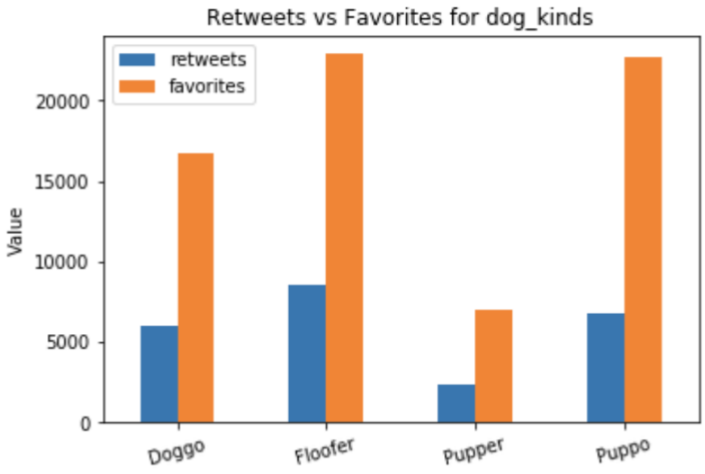
Introduction

The wrangle and analyze project is part of Udacity Data Analysis Nanodegree program. This project is about wrangling of data from multiple sources associated with tweets from twitter application of users rating dogs' pictures. So, in order to wrangle and analyze the data there are main steps to follow as gathering, assessing and cleaning.

Analysis and visualization

This step comes after cleaning the data by involving question to get visualization from the data and come up with some conclusions.

- Q1: What is the most popular dog type based on retweets & favorites using the mean?



So from the chart we can see

- The floofer is the most popular as we see floofer kind got most favorit & retweet.

- Q2: Who the most dog take a 'Retweet'?

Q2: Who the most dog take a 'Retweet'?

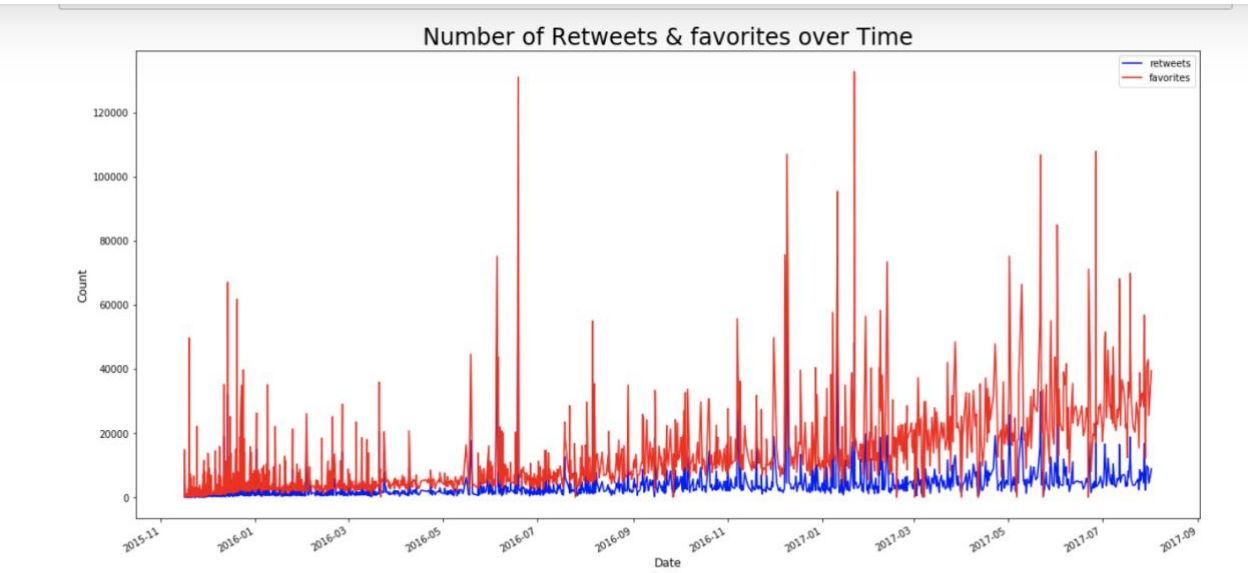
```
In [60]: number_of_retweet_by_dog = twt_archive_main.groupby('name').retweets.sum()
         ## sort it as Descending
         number_of_retweet_by_dog.sort_values(ascending = False)

Out[60]: name
         Bo            73800.0
         Stephan       56625.0
         Duddles       45849.0
         Buddy         39599.0
```

So, The dog called 'Bo' got around 73800 retweets. The dog 'Bo' is so far from the rest of the dogs, compare it with the second one 'Stephan', we will see the difference between the number retweet is far around 17000 retweets

- Q3: What is the rating denominator average ?

```
10.534130543099153
```

- Q4: How the retweets & favorites correlated over time?



As we see

- The number of the favorites is more than number of retweets.
- The relation between the 'favorites' & 'retweet' is suggest there's a direct correlation. When the number of favorites increase the number of retweets will increase.