

Project: Wrangle and Analyze Data

Student: Sherif Shawkat



Act Report

Introduction:

This report is part of the requirements needed to complete the project “Wrangle and Analyze Data” for Udacity Data Analyst nanodegree. The project aims at analyzing dog tweets on the social media giant Twitter. We are collecting tweets, cleaning the received data then analyzing and visualizing them.

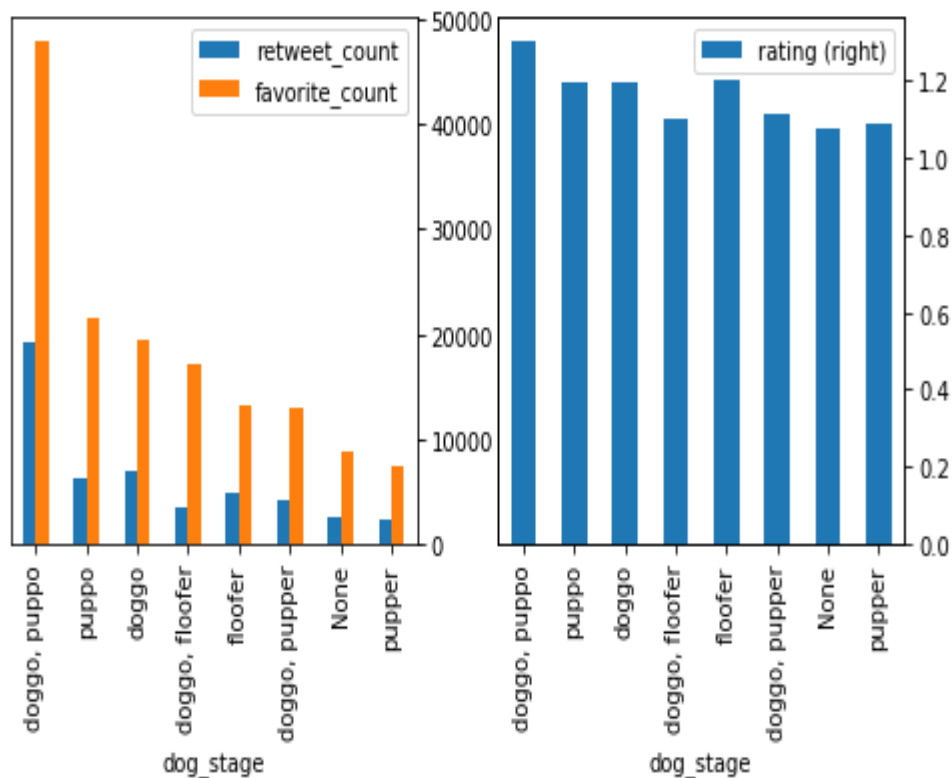
Analysis & Visualizations:

1- Favorite dog stage:

We are analyzing the 4 stages of dogs (doggo, floofer, pupper and puppo) against the ratings, retweets and favorites given for each dog stage.

Results show that the most favorite dog stage is “doggo + puppo”, while the least favorite one is “pupper”.

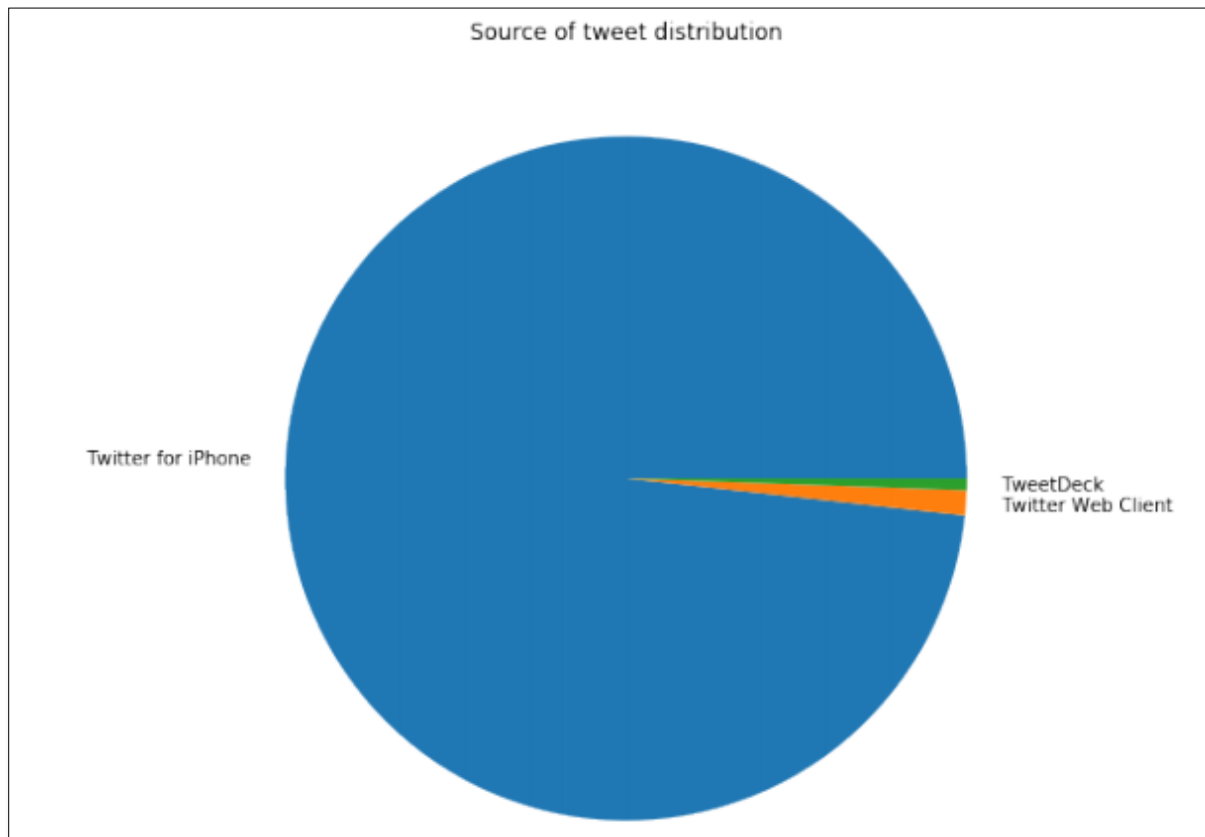
There are many tweets that didn't include the dog type though.



2- Most used source of tweets:

We are analyzing the “source” column, which indicates from which platform was this tweet tweeted.

Results show that over 90% of tweets come from “iPhone”.



3- Most used names:

We are having a small flavors on the most common names given for dogs. Results show that the most commonly used names for dogs are (Charlie, Cooper, Lucy and Oliver).

Note: There are many tweets where the tweeters didn't mention their dog name.

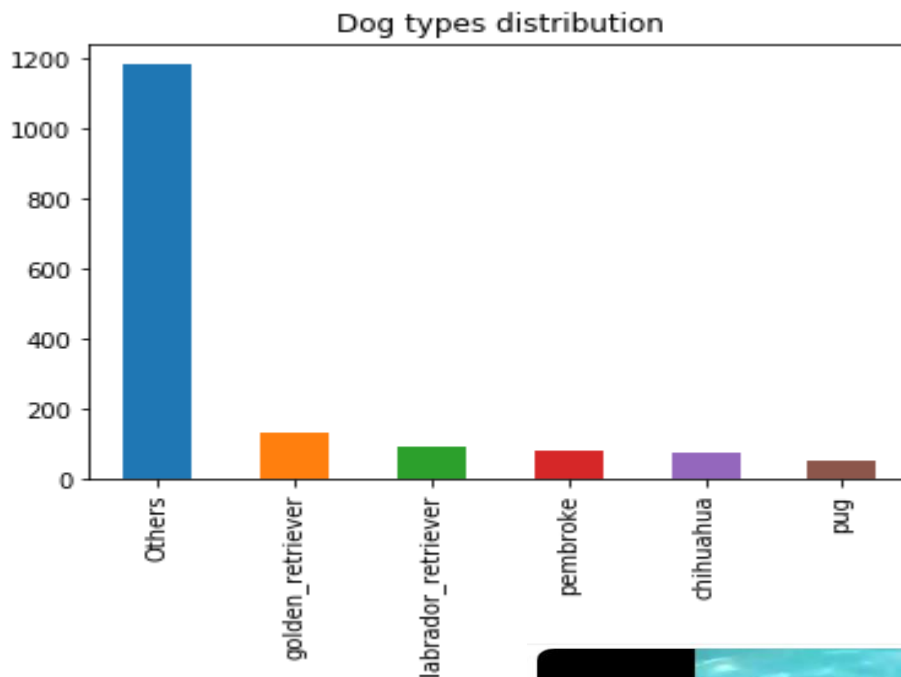
Charlie	10
Cooper	10
Lucy	10
Oliver	9

4- Most common dog type:

We are analyzing the most common dog type using only the 1st prediction of the pictures “p1”.

Since there are very large no. of types, I only plotted the most 5 common types, and other types referred as “Others”.

The most common types are the retrievers “Golden or Labrador”



5- Most favorite/retweeted tweet:

This is a snap of the video from the most favorite/retweeted tweet showing a labrador retriever doggo trying to stand up in a pool. The video has over 12.8M views right now.



Conclusion:

Analyzing & Visualizing data is like the icing on the cake, and my most favorite part of the data analysis process. The amount of findings conclusions you can extract from data is enormous. Visualization is very important to help understand the findings and easily reach our conclusions.