

Wrangle and Analyze Data Project

Act report

By: Abdullellah Al-hudaithy

Introduction: After finishing the wrangling process before I started analyzing and visualizing the twitter-archive-master table I removed unnecessary columns that I will not use in the analysis and visualization steps. And its was shown like this:

- **tweet_id** the Tweet ID
- **timestamp** when the author has published the tweet
- **source** the source of the tweet whether it has been tweeted from iPhone, web etc..
- **text** the content of the tweet
- **retweeted_status_timestamp** when the tweet has been retweeted
- **expanded_urls** tweet URL
- **rating_numerator** that rating that the account will give.
- **rating_denominator** these ratings almost always have a denominator of 10.
- **name** the name of the dog
- **Dog_type** weather its doggo, floofer, pupper, or puppo
- **favourites** How many people click the favourite button for the tweet
- **retweets** How many people click the retweet button for the tweet
- **jpg_url** Image's url for the dog in each tweet
- **img_num** Number of images in each tweet
- **p1** the algorithm's #1 prediction for the image in the tweet
- **p1_conf** how confident the algorithm is in its #1 prediction
- **p1_dog** whether the #1 prediction is a breed of dog or not
- **p2** the algorithm's second most likely prediction
- **p2_conf** how confident the algorithm is in its #2 prediction
- **p2_dog** whether the #2 prediction is a breed of dog or not

- **p3** the algorithm's third most likely prediction
- **p3_conf** how confident the algorithm is in its #3 prediction
- **p3_dog** whether the #3 prediction is a breed of dog or not

Analysis and Visualization

Question 1: Who the most dog who take like (favorites)?

name

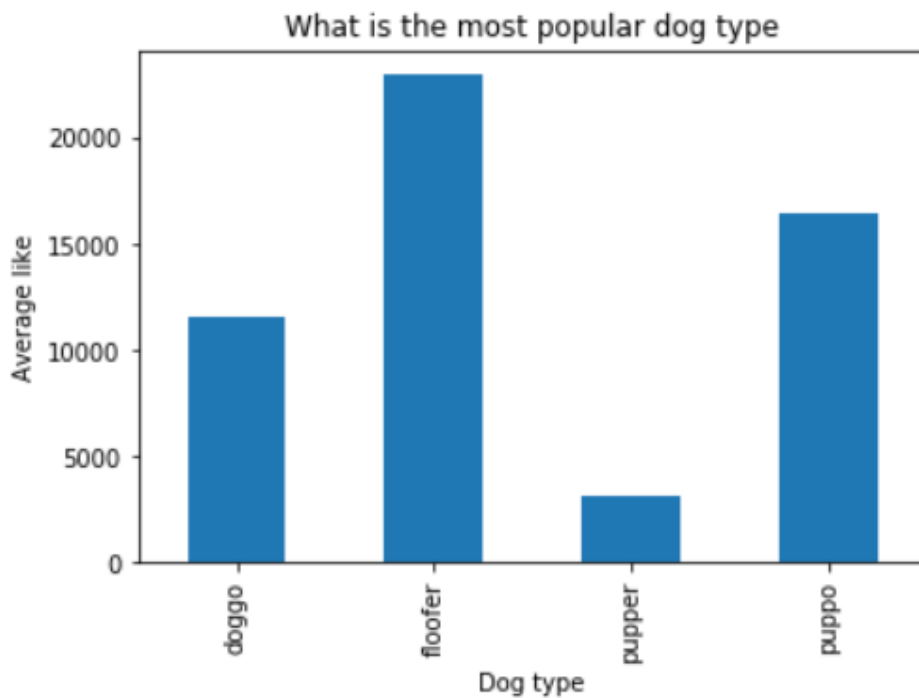
Bo 185922.0

Charlie 115468.0

Zoey 113581.0

As we can notice that Bo receive 185922 like and he/she is so far from the second one Charlie.

Question 2: What is the most popular dog type based on like (favorites) median?



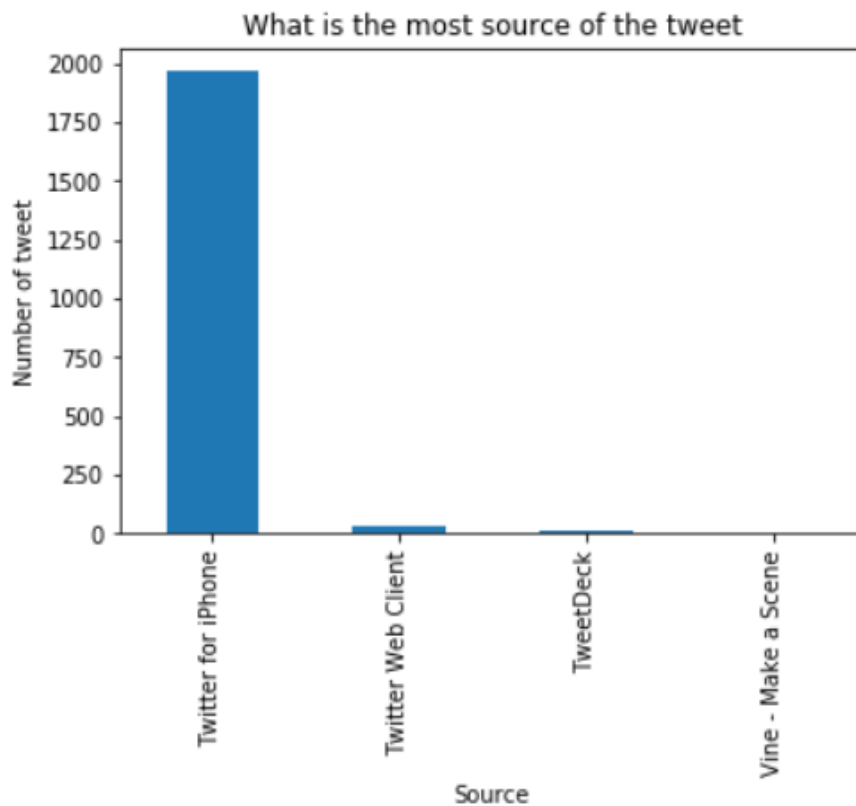
We can see both floofer is the popular in the account.

Note: I take the median to avoid the large number e.g. Bo.

Question 3: What is the rating numerator average?

Answer: 12.30

Question 4: What is the most source that the tweets comes from?



We can see a lot of people or most of the people tweeting from iPhone.