DAND Wrangle And Analyze Project

Act Report

By Hamad Sami AlAssafi

Table of Contents

Introduction:

After I finished the wrangling process, the analysis begins by creating master dataframe that contains all the dataframes. In our master dataframe we have multiple columns and I will go through them one by one to understand our dataset and make wise analyses in our process:

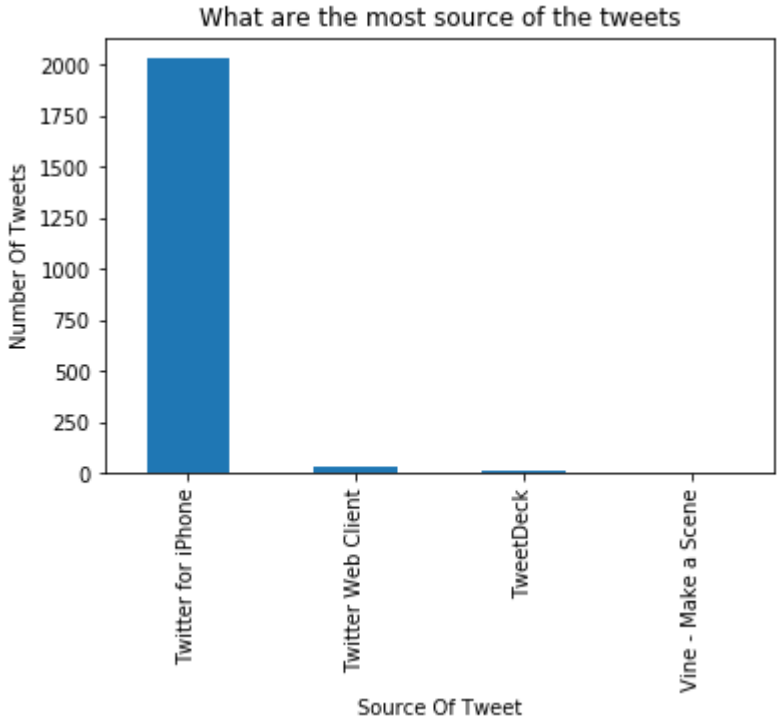
- 1-(tweet_id) the Tweet ID
- 2-(in_reply_to_status_id) means if the tweet is reply tweet then it should have it unique ID
- 3-(in_reply_to_user_id) If the tweet is a reply tweet then the value will be an integer of the original Tweet's author id
- 4-(timestamp) When the author has published the tweet
- 5-(source) The source of the tweet (whether it has been tweeted from Phone Computer etc..)
- 6-(text) The content of the tweet
- 7-(retweeted_status_id) It shows what tweet has been retweeted based on tweet ID
- 8-(retweeted_status_user_id) ID for the person who retweet that certain tweet
- 9-(retweeted_status_timestamp) When the tweet has been retweeted
- 10-(expanded_urls) Tweet URL
- 11-(Rating) The rating of the dog
- 12-(name) The name of the dog
- 13&14&15&16-(doggo, floofer, pupper, puppo) Different criteria for the dogs that determine their age and their type
- 17-(jpg_url) Image(s) url for the dog in each tweet
- 18-(img_num) Number of images in each tweet
- 19- (p1) Is the algorithm's #1 prediction for the image in the tweet
- 20-(p1_conf) Is how confident the algorithm is in its #1 prediction

- 21-(p1_dog) Is whether or not the #1 prediction is a breed of dog
- 22-(p2) Is the algorithm's second most likely prediction
- 23-(p2_conf) Is how confident the algorithm is in its #2 prediction
- 24-(p2_dog) Is whether or not the #2 prediction is a breed of dog
- 25-(p3) Is the algorithm's third most likely prediction
- 26-(p3_conf) Is how confident the algorithm is in its #3 prediction
- 27-(p3_dog) Is whether or not the #3 prediction is a breed of dog
- 28-(favourites) How many people click the favourite button for the tweet
- 29-(retweets) How many people click the retweet button for the tweet

Analysis and visualization:

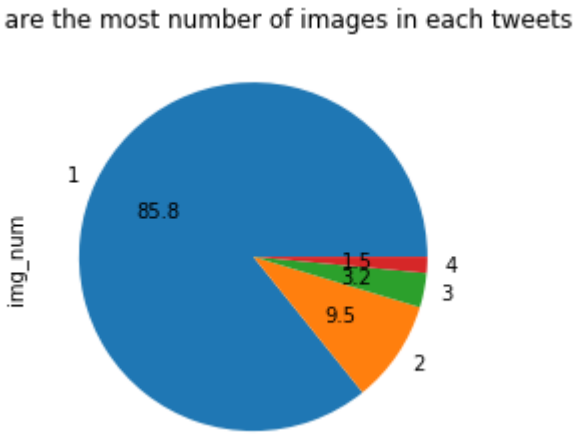
I have explored the dataset and came up with four questions two of them with visualizations using (matplotlib):

Question 1: What are the most source of the tweets in our dataset



The most tweets have been tweeted from (Twitter application in IPhone)

Question 2: What are the most image number in each image



Most of the tweets have only one image with (85.8) percent

Question 3: What is the most popular dog type in our dataset

pupper	229
doggo	75
puppo	29
floofer	3

Is it shown the most popular dog type in our data set is (pupper) with (229) dog

Question 4: What is the average rating of the dogs:

1.7672621921776919

Is it shown the average (mean) is 1.7672621921776919, and in average it considered as good average.

Thank you for reading, I hope it is useful insights and visualizations.