

Act report

Raghad Altamrah

From udcity they provide for me three data base and I install it to my computer

The first data is about twitter-archive-enhanced.csv that contine float64(4), int64(3), object(10) and 2356 entries

And the second file image-predictions.tsv that contine bool(3), float64(3), int64(2), object(4) and 2075 entries

And the last file is about tweet-json.txt that conties int64(3) and 2354 entries

I have three data frame is clean so I Marge the data frames and in one CSV file named "twitter_archive_master.csv

Then I write what the information that take it from the data

I will talk about the insights visualization in here

This is the Insights:

the mean of rating_numerator is 13.725702.1

the maxumiam number of image is 4 and the most of the tweets have one image

the maxiumam number of retweets is 56625 and the minumam is 16

i have p1_dog,p2_dog,p3_dog which is type is bool

there is 1740 entries from favorites and retweets

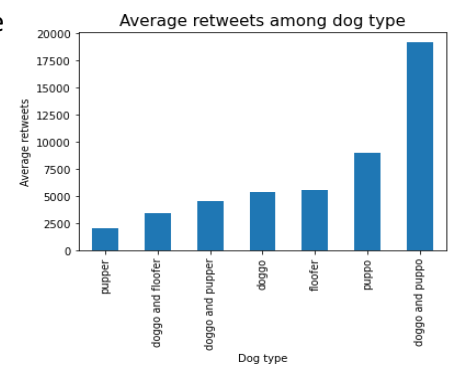
Visualization:

Here I draw a graph between the dog type and the average retweets to see the which dog type is have the biggest retweets

And we see from the graph that the doggo and puppo is the most type is frequently reputed

The single word puppo is reputed more than the other word like pupper and doggo

Also the compined word doggo and puppo is the most reapedt from the other compined word doggo and floofer and doggo and pupper



And I draw a scatter plot between favorites and retweet

So the favorites and retweet have a positive correlation

Between theses two

