Wrangle Report

Introduction

Our goal for the project was getting data from different sources regarding a Twitter account @dog_rates, it's a twitter account that rates other people dogs in humorous way, and wrangle all the data, then store it, then analyze the finding and visualize it and finally write a report for the stakeholders.

After cleaning the data and merging all dataframes into one master dataframe that contains all data from the 3 tables we came to the next step of out project

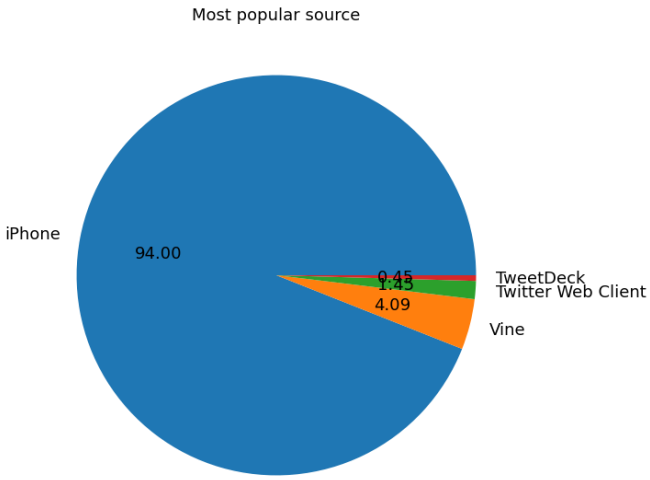
- Analyzing Data
- Visualizing Data

We have three questions to answer

1. What is the most common source of the tweets?
2. What are the most popular dog breeds in the twitter account?
3. Is there a correlation between number of retweets and number of tweets favorite

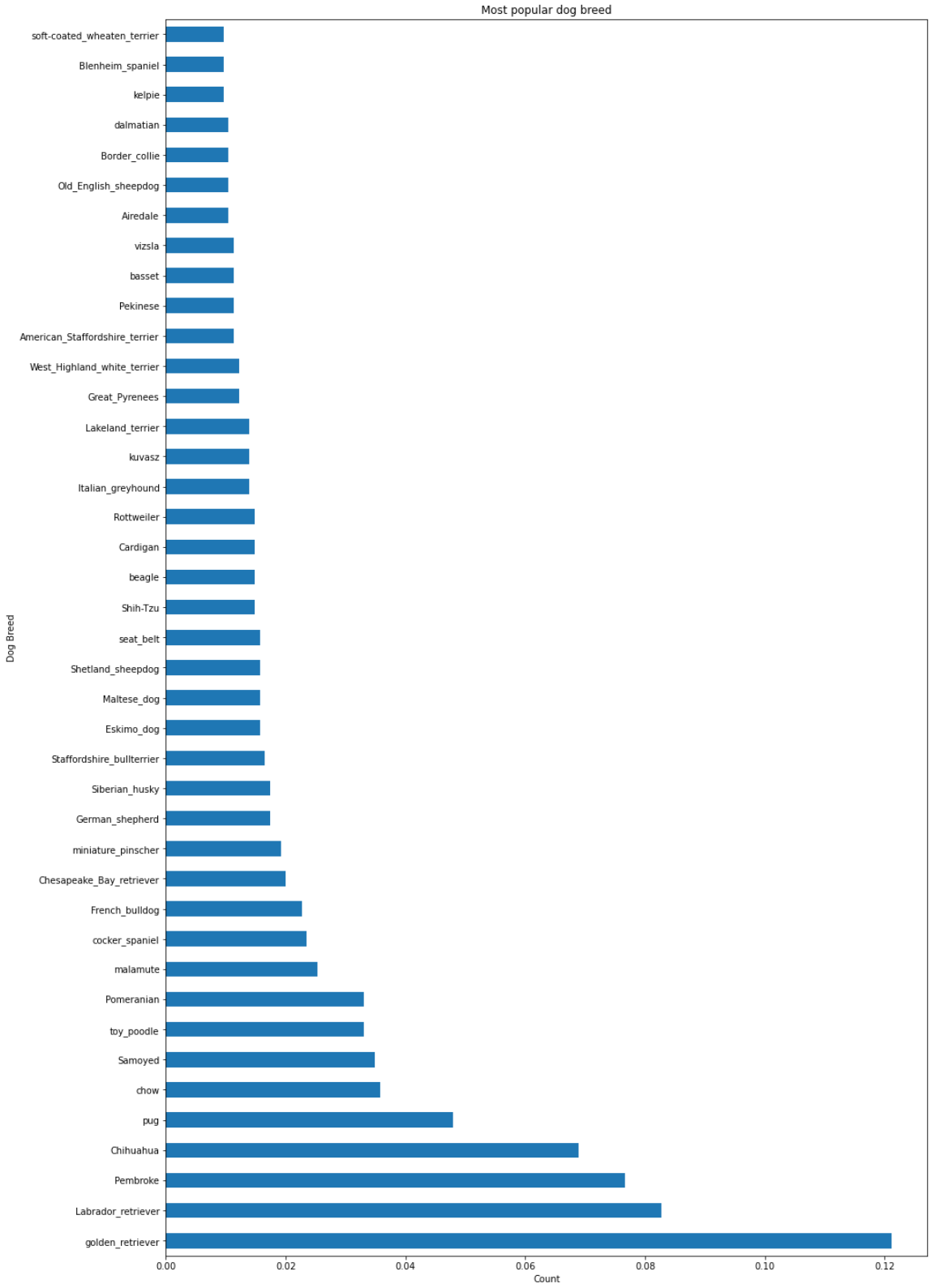
What is the most common source of the tweets?

I Used a pie chart with annotation on the pie chart with the percentage of each tweet source in comparison with total tweets published online by the weratedogs account to find the most common source of tweets, and based on the data that we have 94% of the tweets came from iPhone



What are the most popular dog breeds in the twitter account?

Our second question was to find the most popular dog breeds published online, so I used a horizontal bar chart to represent how many times each dog breed were published online but at least should be 10 times to be include in the bar chart for better readability of the chart, and golden retriever and Labrador retriever were the top 2 dog breed that published online


Most popular dog breed
soft-coated_wheaten_terrier
Blenheim_spaniel
kelpie
dalmatian
Border_collie
Old_English_sheepdog
Airedale
vizsla
basset
Pekinese
American_Staffordshire_terrier
West_Highland_white_terrier
Great_Pyrenees
Lakeland_terrier
kuvasz
Italian_greyhound
Rottweiler
Cardigan
beagle
Shih-Tzu
seat_belt
Shetland_sheepdog
Maltese_dog
Eskimo_dog
Staffordshire_bullterrier
Siberian_husky
German_shepherd
miniature_pinscher
Chesapeake_Bay_retriever
French_bulldog
cocker_spaniel
malamute
Pomeranian
toy_poodle
Samoyed
chow
pug
Chihuahua
Pembroke
Labrador_retriever
golden_retriever
Dog Breed
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Count

Is there a correlation between number of retweets and number of tweets favorite

Our final question if to find if there are a correlation between number of tweets and number of tweets favorite and the best way to find correlation between 2 items is using scatter plot

And based on the graph below there are strong positive correlation between number of tweets and number of tweets favorite.