Wrangle act

Eya Sahli

Udacity Data Analyst Nanodegree

Overview

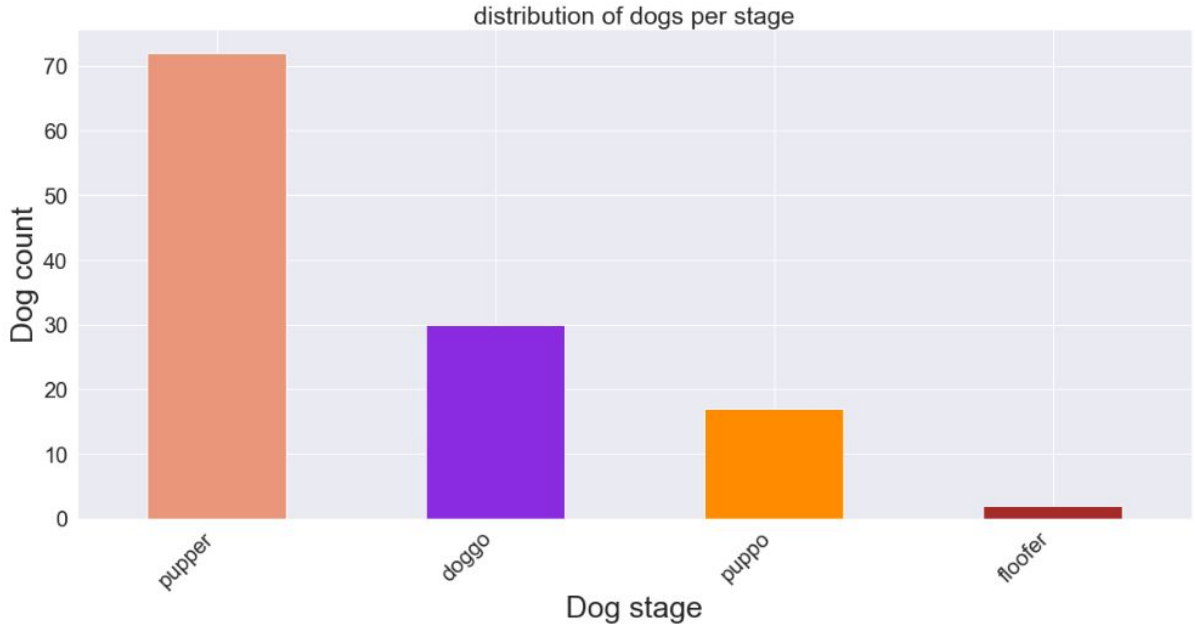
In this document I'm concisely describing my analysis efforts of my wrangle and analyze data project .

Questions :

- ## Question 1 : what is the distribution of dogs per stage ?

  In this first question , I tried to identify the dog stage distribution of the dogs in my dataset using value counts and then i made a bar chart with seaborn to visualize the results . turns out pupper is the most common dog stage , and floofer is the least common !

  Barchart :



Question 2: what are the top 5 dog breeds with the most favorites ?

In this question i used groupby , sum and sort_values to get and sum the number of favorites for each dog breed than sort them to get the top5 dog breeds with the

most favorites , then i plotted my results into a barchart , the top5 dog breeds when favorites are concerned are :

golden retriever with 1193996 favorites.

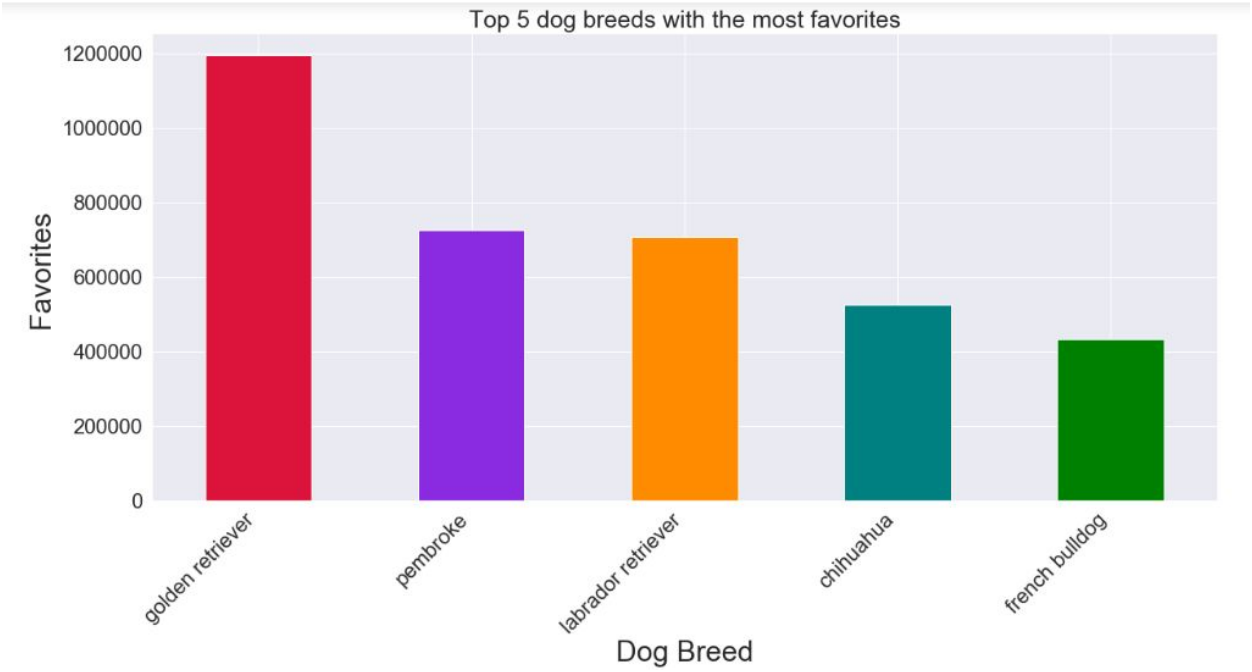
pembroke with 723837 favorites.

labrador retriever with 706936 favorites.

Chihuahua with 524573 favorites.

french bulldog with 433549 favorites.

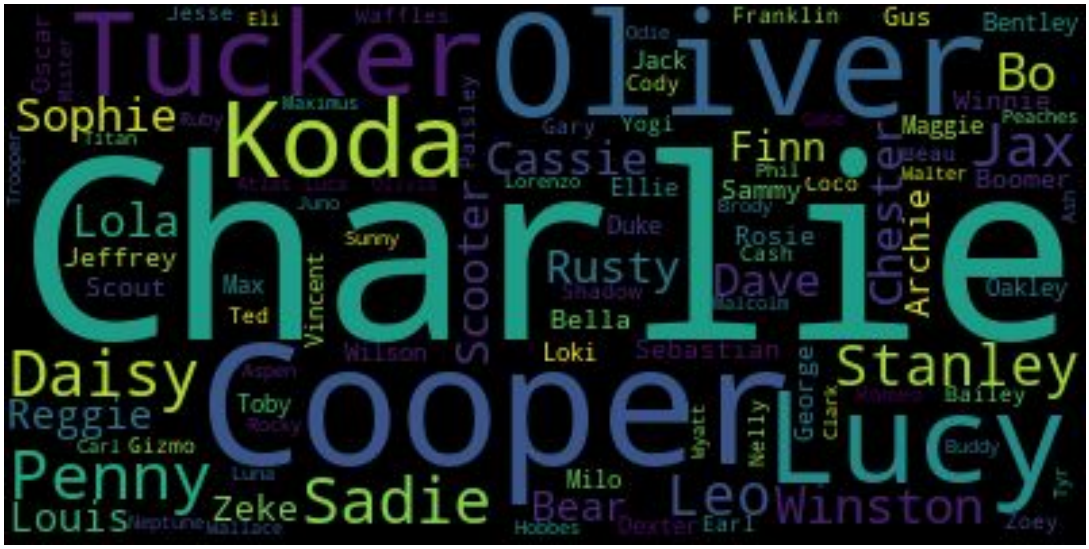
Barchart :



- ## Question 3 : what are the 100 most common Dog names ?

To answer this question , i created a wordcloud to display the most common 100 names in our dataset , the instructions in this resource helped me achieve the following wordcloud :

https://www.datacamp.com/community/tutorials/wordcloud-python



As shown in the wordcloud , we can spot some very common names like Charlie , which i assume is the most common , Cooper , Tucker and Oliver , Lesser commun names are Penny , Daisy and Stanely and Less common names are Zoey , Wallace and Yogi .

- ## Question 4 : what is the distribution of source for this dataset ?

  For the distribution of source , a simple value_counts() was used , the results were later plotted into a barchart to show us the following :

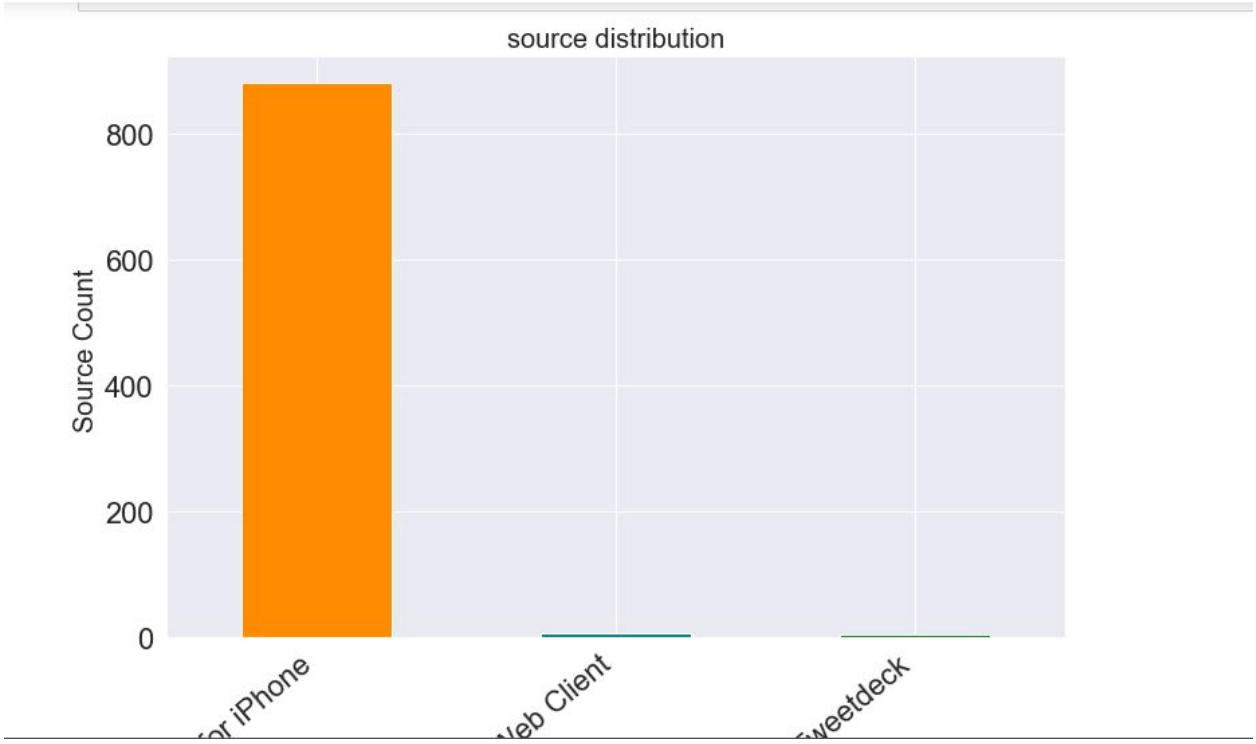
  Twitter for iPhone : 881

  Twitter Web Client : 8

  Tweetdeck : 6

So Twitter for Iphone is the most used medium in WeRateDogs while TweetDeck is the least used

Here is my barchart :


source distribution
Source Count
800
600
400
200
0
For iPhone
Web Client
Tweetdeck