UDACITY

DAND Wrangle and Analyses Data

Omar Abdulaziz Alashikh

Introduction:

In this project we will gather data from various sources and various formats to assess its quality and tidiness, after that we will start cleaning the data

Gathering:

1-From (CSV) file that contains archive tweets for WeRateDog account, which was given by Udacity.
2-From (TSV) file that contains detailed data about the images in each tweet.
3-From (JSON) file that contains detailed data about the count of the favorite and retweets.

Assessing Data:

I searched through the datasets to find any quality or tidiness issues, either visually or through programming

Quality Issus:

1-Missing Values
2-Incorrect dog names
3-Dupliated tweets

Tidiness Issues:

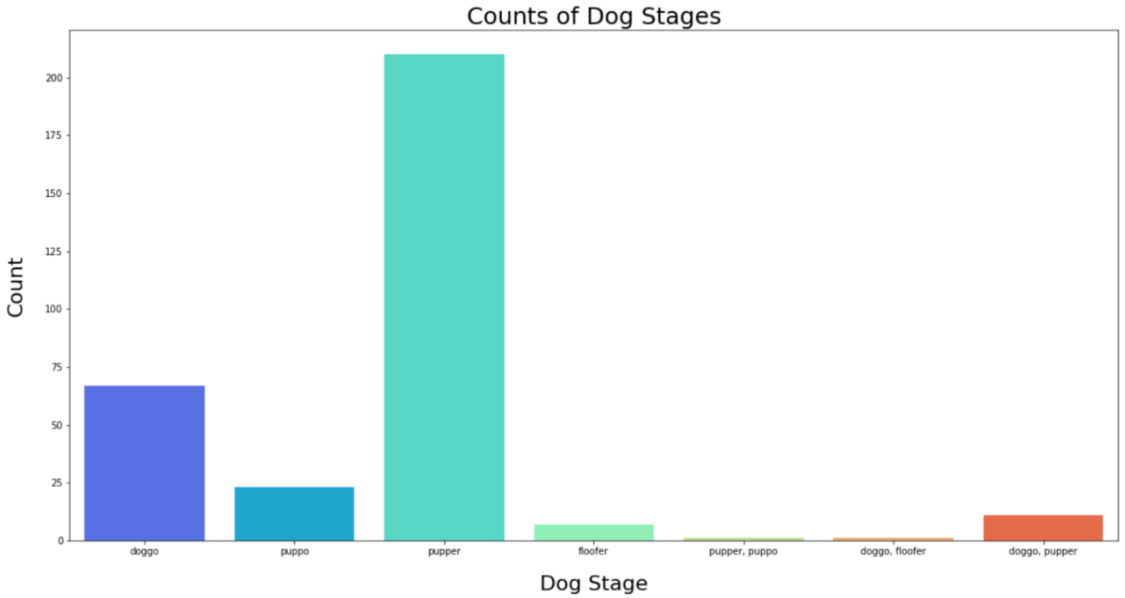
1-Convert the nominator rating and denominator rating to actual rating.
2-Create single column that express the type of dog instead of four columns for each type.
3-Merge the data frames together

Cleaning Data:

I tidied up the Twitter archive data frame, then the image prediction and the Twitter API data frame.

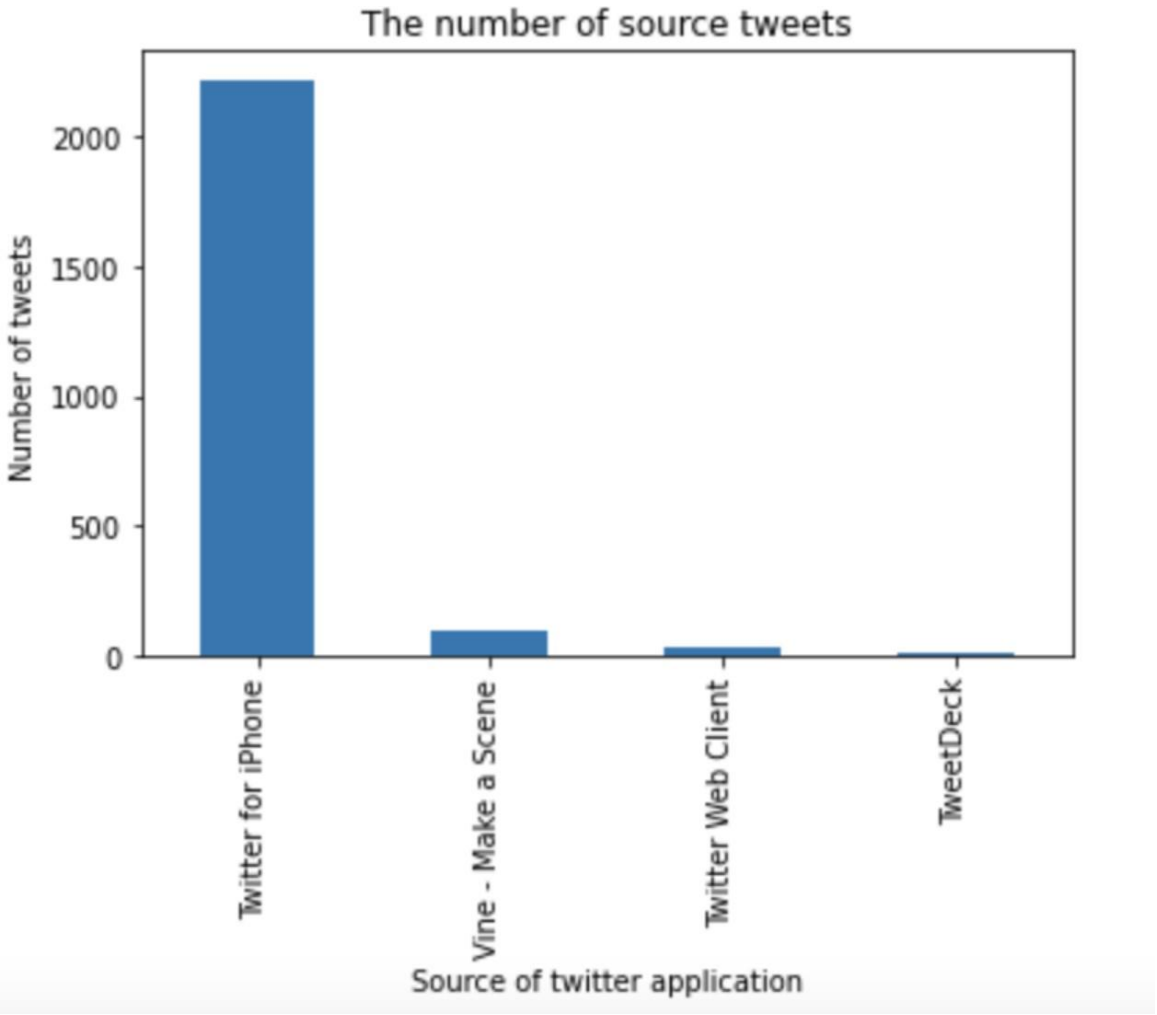
Analyses

Q1: What are the most Dog stages



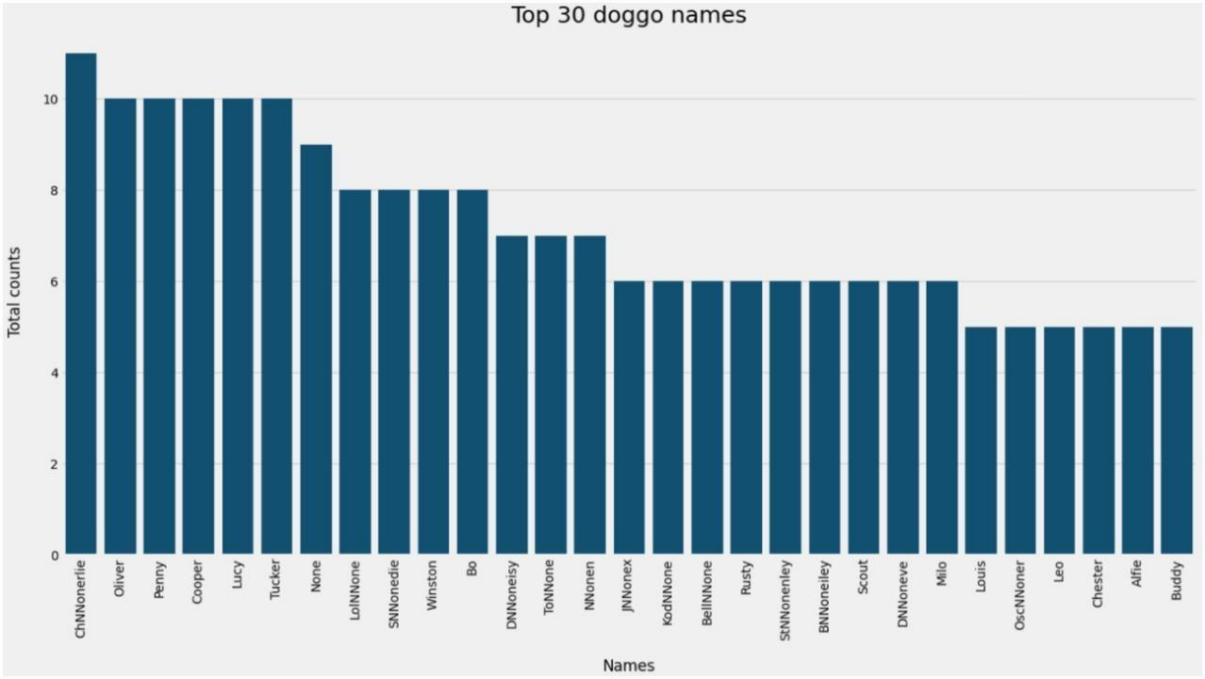
It’s the pupper breed

Q2: The Statics of source tweets



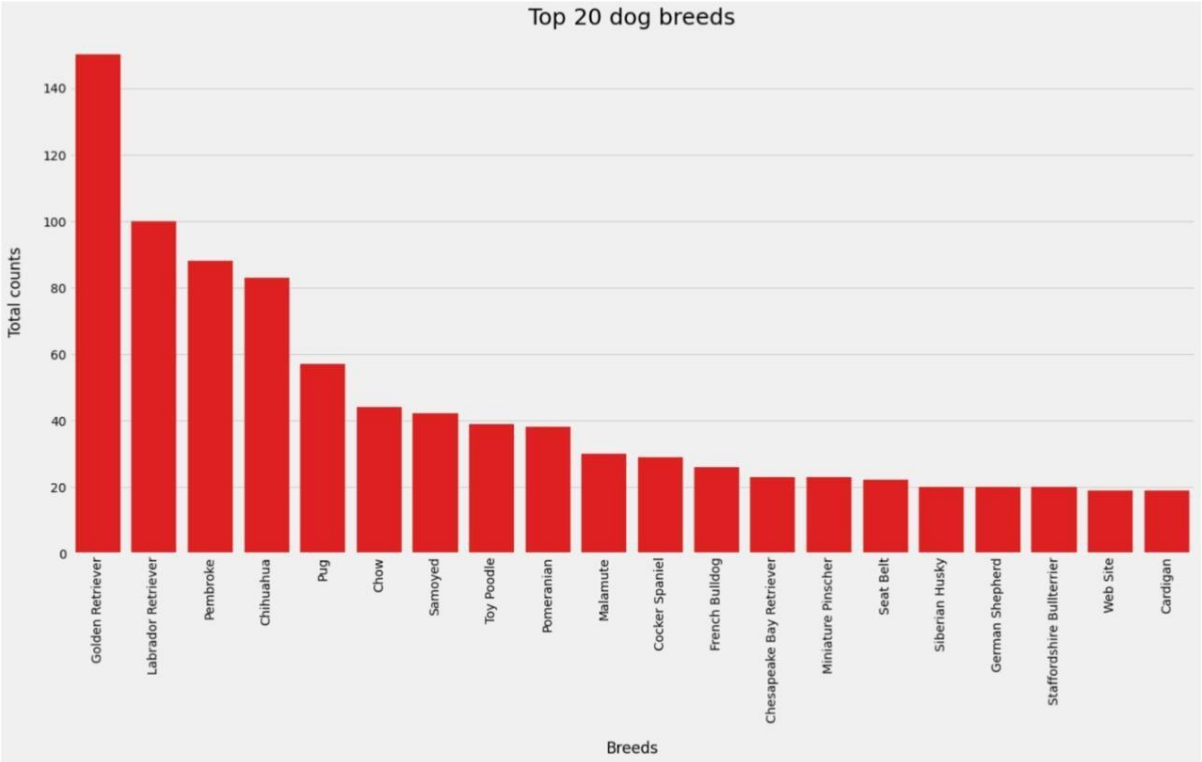
Twitter for iPhone users generates the most tweets.

Q3: What is the Top 30 doggo names



It’s not a static thing you can name your dog whatever you want

Q4: What are the most popular dog breeds



the most popular breed is
the Golden Retriever