



ACT_REPORT

PROJECT: WRANGLE AND ANALYZE DATA

Abstract

This is a report of wrangling (analyzing and visualizing) dataset which is a tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog.

UMAR FARUQ ZUBAIRU
umarfaruqzubairu@gmail.com

INTRODUCTION

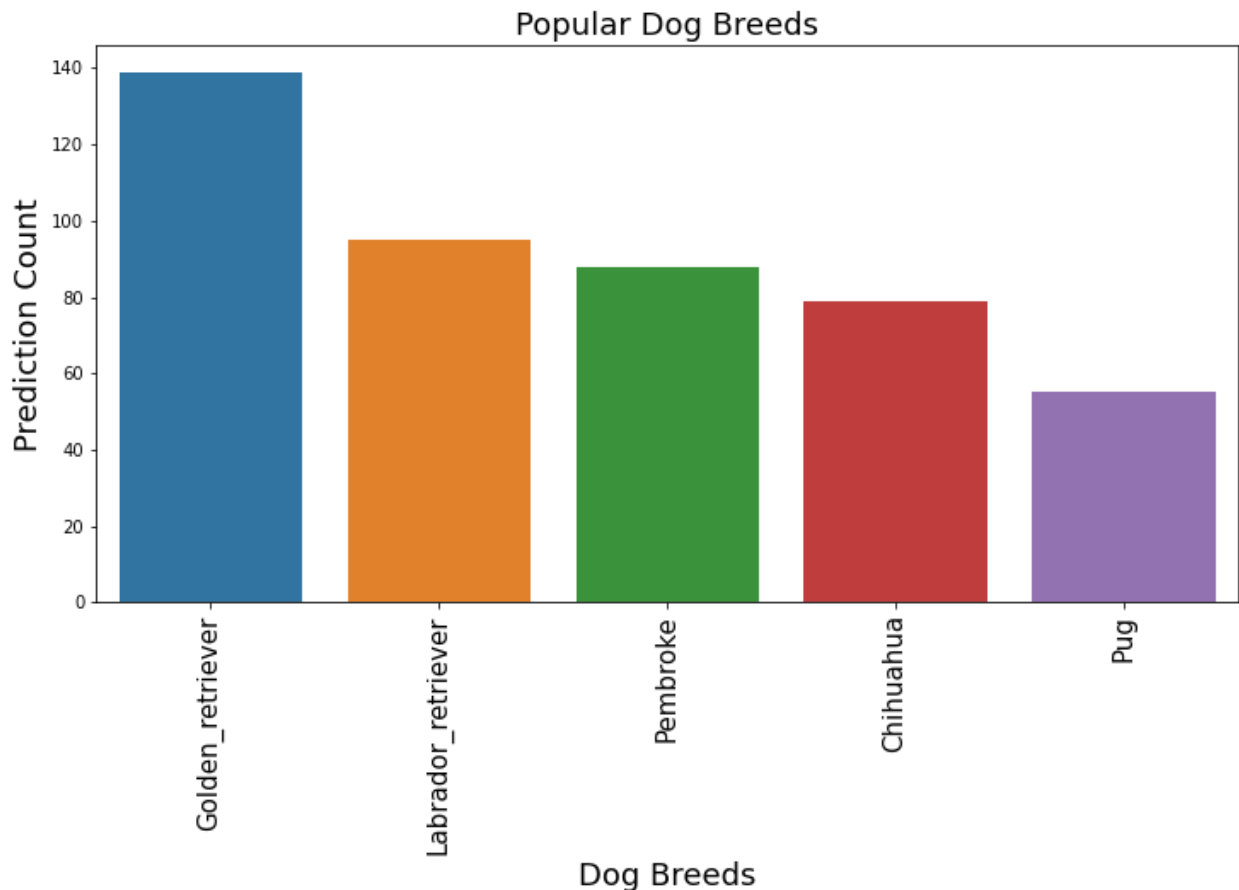
The dataset that I wrangled is the tweet archive of Twitter user [@dog_rates](#), also known as [WeRateDogs](#). [WeRateDogs](#) is a Twitter account that rates people's dogs with a humorous comment about the dog.

But what are these ratings for? Which is the most popular dog breed? What's the relationship between Retweets and Favorites counts? Among others are some questions that disturb my mind.

I have wrangled, analyze the data and found some interesting insights about the data. Of course, the awesome Python libraries help me a lot in the process. Thanks to the large community of internet users that help me get through some challenges by visiting forums such as stackoverflow.

MOST POPULAR DOG BREED

I have analyzed the data and found the top 5 most popular dog breeds based on **first prediction**.



The above chart shows the top 5 most popular dog breeds. Where **Golden_retriever** is the top.

TOP 3 RATED DOGS AMONG THE POPULAR DOG BREED – GOLDEN RETRIEVER

Analyzed the most popular dog breed and found out that these are the top 3 rated Golden retriever dogs:

- I. Sam



II. Ollie



ollievuesomuch

PRANK >



III. One



RATINGS, RETWEET AND FAVORITE COUNTS

From what I found, there is strong positive correlation between Favorite Counts and Retweet Counts i.e a linear relationship between the retweet counts and the favorite counts.

