Udacity Project

Wrangle and Analyze Data

By Hanan Alshehri

"Act Report"

Introduction

Real-world data rarely comes clean. Using Python and its libraries, I will gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling. I will document my wrangling efforts in a Jupyter Notebook, plus showcase them through analyses and visualizations using Python (and its libraries) and/or SQL.

The dataset that I will be wrangling (and analyzing and visualizing) 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. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

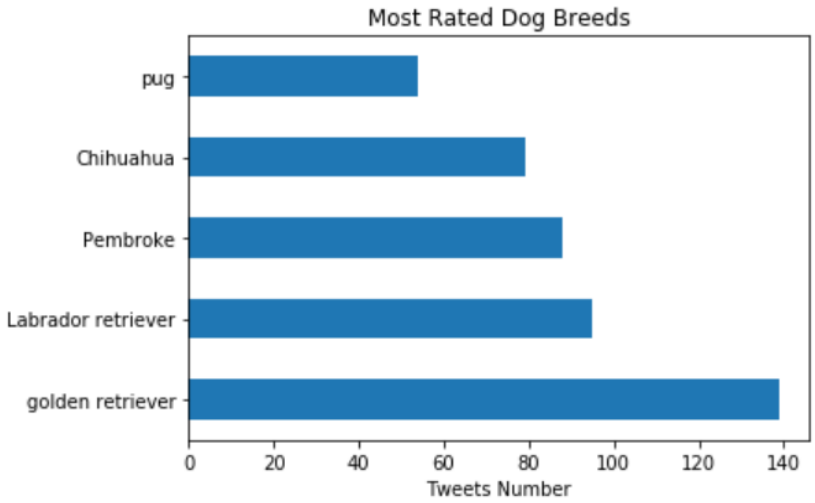
- My goal: wrangle WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. The Twitter archive is great, but it only contains very basic tweet information. Additional gathering, then assessing and cleaning is required for "Wow!"-worthy analyses and visualizations.

In this Project, I will analyze and visualize my wrangled data. At least three (3) insights and one (1) visualization must be produced.

1- Insights

1- Most Popular Dog Breeds (by Num of Tweets).

From the result, the most popular dogs on the WeRateDogs Twitter are Golden Retrievers, then Labrador Retrievers, Pembrokes, Chihuahuas and Pugs.



2- Most Favorite Tweet.

From the result, the most favorited tweet is a picture with #WomensMarch trends: "Here's a super supportive puppo participating in the Toronto #WomensMarch today. 13/10."

3- Prediction Algorithm.

From the result, prediction 2 is the most accurate prediction algorithm because 1495 prediction is true.

2- Visualization

1- The most used Twitter source

As we see, the most used source for Twitter is iPhone: