

Report on data analysis as part of deliveries for 2nd project of Udacity data science Nanodegree

By : Mohamed Hashem

egyptfwd
initiative

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. The account was started in 2015 by college student Matt Nelson, and has received international media attention for its popularity the Twitter account has nearly 7.6 million followers, and Nelson sees 30,000 likes on a post as being viral. His most popular post was of a dog marching in the 2017 Women's March, which was retweeted more than 50,000 times and favorited 134,000 times.

It mainly 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.

WeRateDogs downloaded their Twitter archive and sent it to Udacity via email exclusively for you to use in this project. This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017. More on this soon.

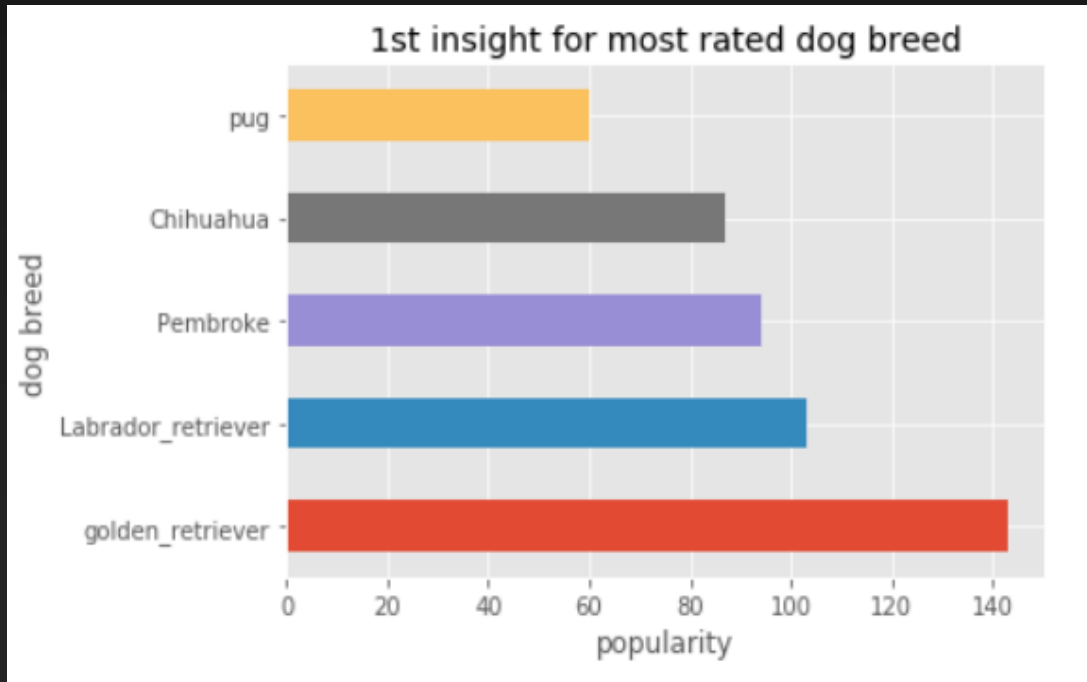
Using data analysis techniques I have fetched some data from this account and applied some processing on it to get some useful insights with the help of python programming language, actually I have got 3 insights and visualized all of them as following



Most Rated Dog Breed

>140

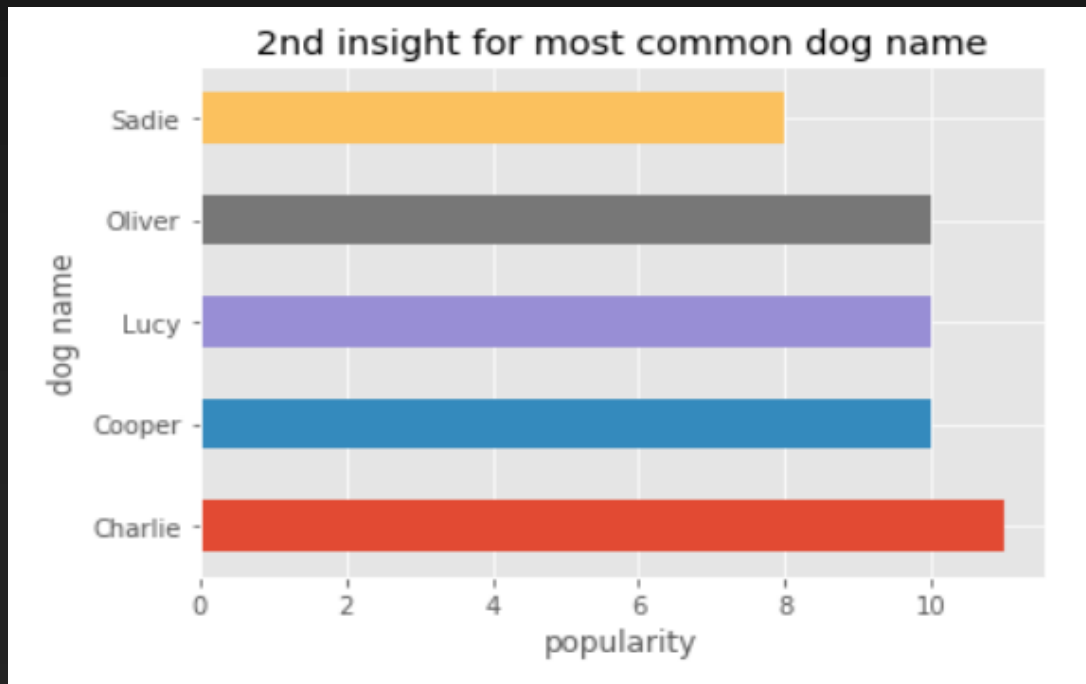
After analyzing the available tweeter data, I have checked how much every breed is present and after that I have visualized the highest 5 values in a horizontal bar chart as shown which indicates that the Golden Retriever is the most common dog breed on this account tweets for the mentioned time period .



Most Common Dog name

>10

The 2nd insight I have fetched through my journey on this dataset is the most common dog name through this account, which found to be Charlie, then a visualization process had been made for the top 5 names as horizontal bar chart as shown



Relation between retweets and favorites

+ve

correlation

And finally to analyze a different type of data and visualize it on a different shape, a scatter plot had been produced to declare that there is a positive relation between retweets and favorites, i.e. as retweets increase favorites also increase

The following plot shows that

