

Twitter – WeRateDogs

Wrangle Report

Satyam Chauhan



Introduction

This project is a part of Udacity Data Analyst Nano Degree Program. One of the requirements of the nano degree is to complete the data wrangling and analysis of data obtained via various

methods and generate meaningful insights and visualization. For this project we are working on the data sets from WeRateDogs Twitter account. This account tweets entertaining and funny dog posts. They are famous for their quirky rating system where they score dogs in an unconventional way.

Insights

We obtained data from various methods, and we wrangled them using various techniques to generate a master data set to arrive at some cool insights

- Followership
 - Over the years, WeRateDogs followership hasn't seen significant variation, but it had the highest followership in 2015

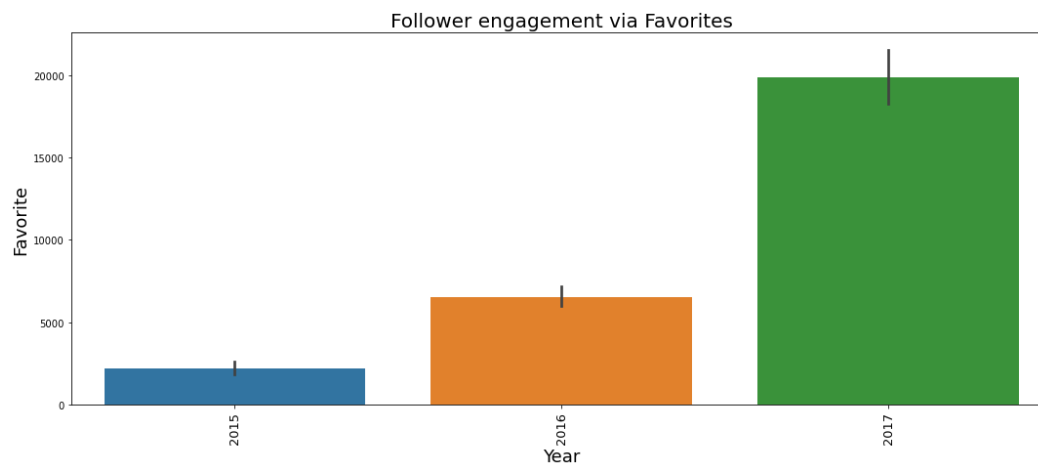
followers	
year	
2015	9100223
2016	9100221
2017	9100219

- Popular dogs via tweets from WeRateDogs
 - We analyzed the image predictions data provided by Udacity.
 - All three predictions were compared and the one with highest avg confidence was considered for this analysis
 - first_prediction turned out to be the one with highest average confidence
 - Thereafter, we utilized value_counts function to recognize the breed with highest prediction observation
 - Conclusion –
 - Golden Retriever turned out to be the most popular breed when it comes to the tweets from WeRateDogs
- Popular breed via follower engagement
 - In the previous analysis we realized popular breed from WeRateDog's perspective
 - Now we would like to check out which breeds are popular among followers
 - Retweet and Favorite counts are a good measure of follower engagement with the tweets by the accounts
 - We created a matrix to identify maximum retweet and favorite count value and the queried that number to identify the breeds for each variable. We used first_prediction for dog breed
 -

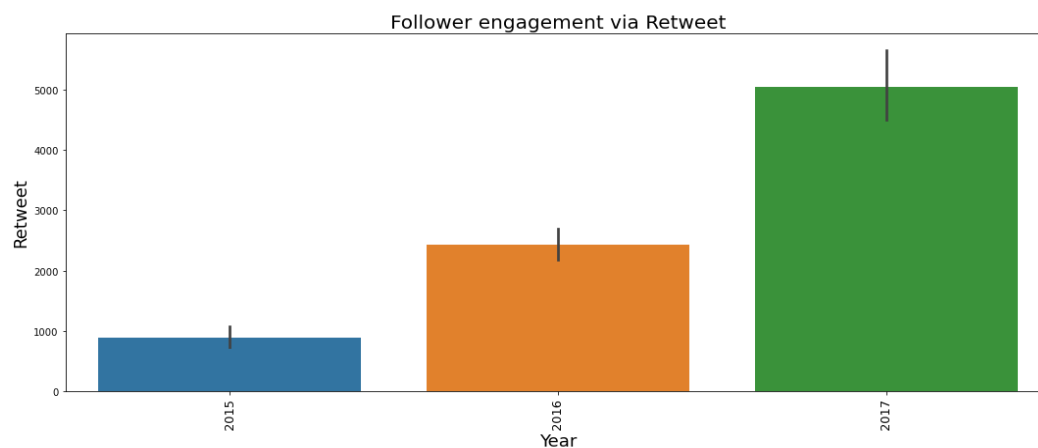
- So, we utilized the prediction column to generate another column named as 'breed' where dog breeds were stored as it is and anything else was stored as 'object'
- Conclusion
 - Standard poodle turned to be the popular breed based on Retweet
 - Saluki turned out to be the popular breed based on Favorites

Visualization

- Follower engagement via favorites
 - This visualization shows the follower engagement via favorite over the years of 2015, 2016 and 2017
 - The follower engagement via favorite sees an upward trend



- Follower engagement via retweets
 - This visualization depicts the follower engagement via retweet over the years of 2015, 2016 and 2017
 - The follower engagement via retweet sees an upward trend



- Popular breed each year
 - For this visualization we created a dataframe for each year and concatenated them to create one dataframe

- We used the value_function to get the top breed for each year
- The below visualization shows that Golden Retriever and Labrador retrievers have been popular consistently whereas we see chihuahua being the top dog in 2015
- This visualization gives us a unique perspective as to how golden retriever became WeRateDogs favorite dog over the years.

