Wrangling & Analyze Data Project

Analysis & Insights Report

06.10.2021
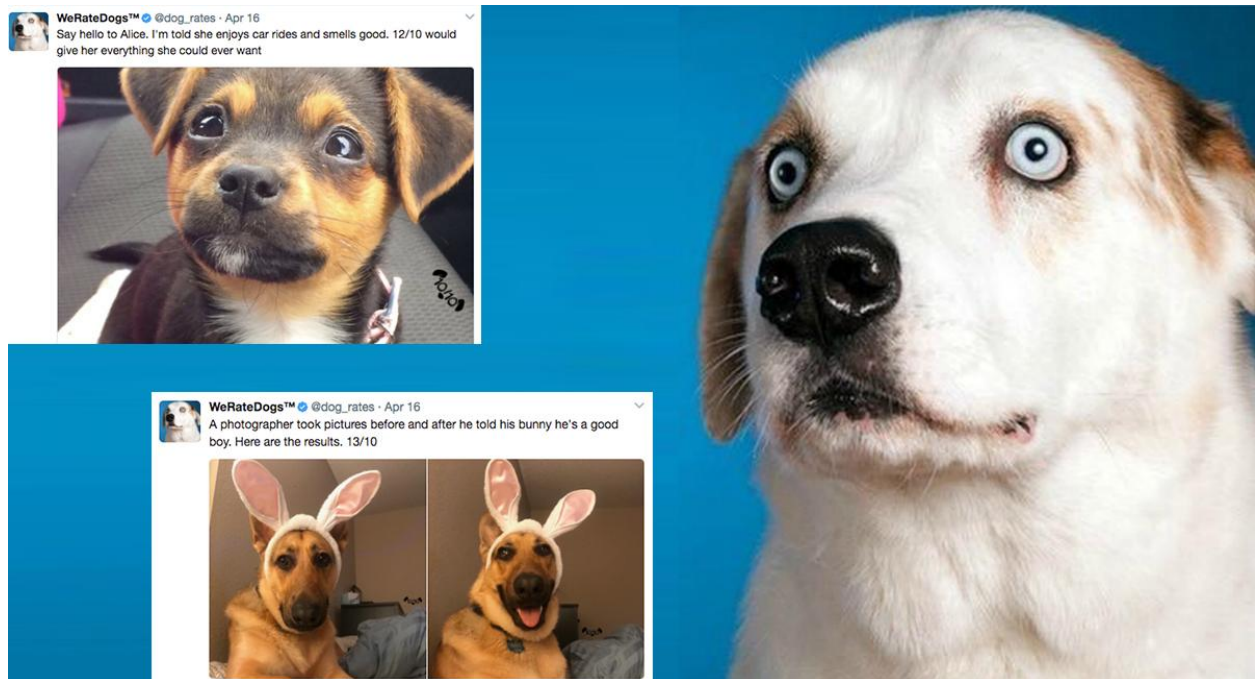—

Mohamed BOUSETTA MAHJOUB
Udacity Data Analyst Nanodegree

Overview

In this report, I will describe some insights found in the wrangled dataset WeRateDogs .

The dataset that is 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. These ratings almost always have a denominator of 10.
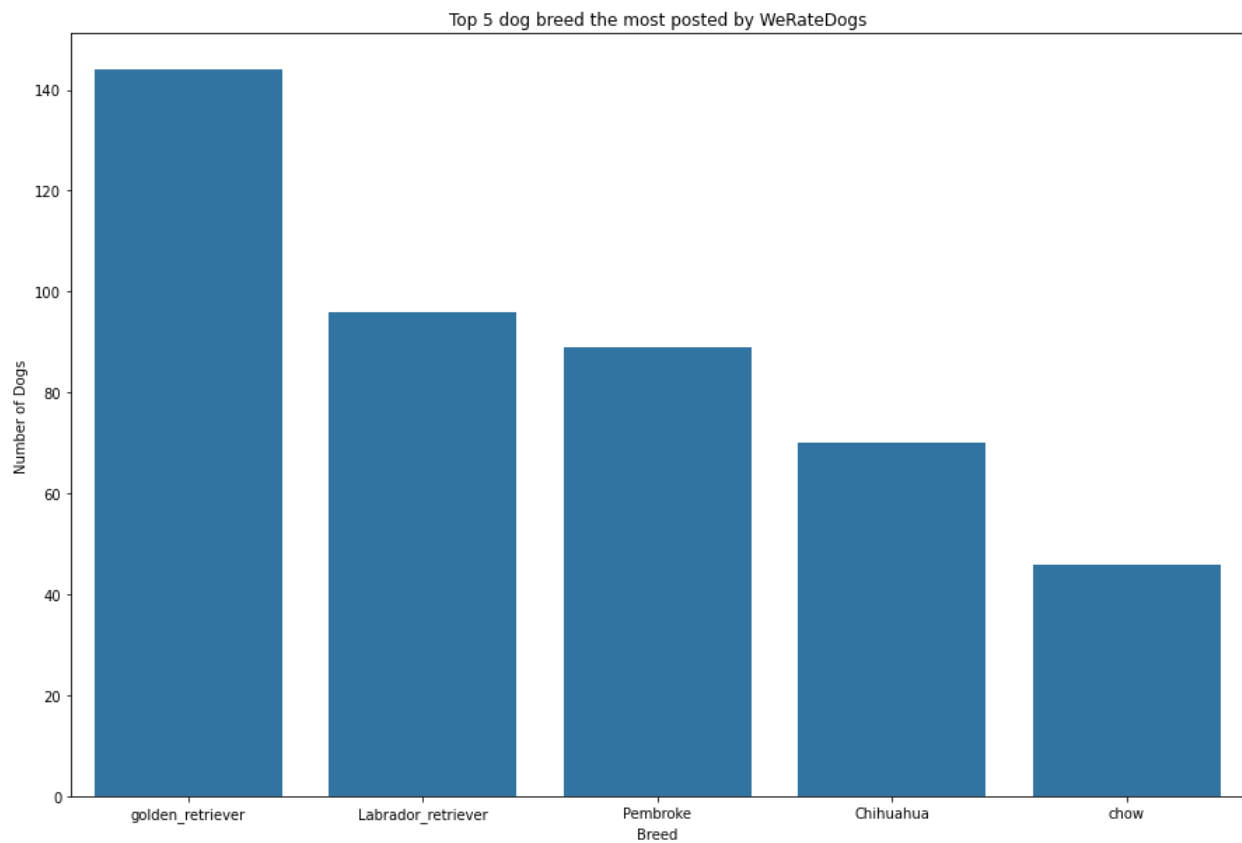

Once the Data is wrangled, I tried to answer some questions about WerateDogs tweets:

- Question 1 : Which dog breed is the most posted ? The Top 5 ?
- Question 2 : For this Top 5, which one is the best rated ? the most liked ? the most retweeted ?
- Question 3 : Is the number of retweets correlated with the number of likes ?
- Question 4 : Is Rating correlated with Number of Retweets and Number of Likes ?

Analysis & Visualisations

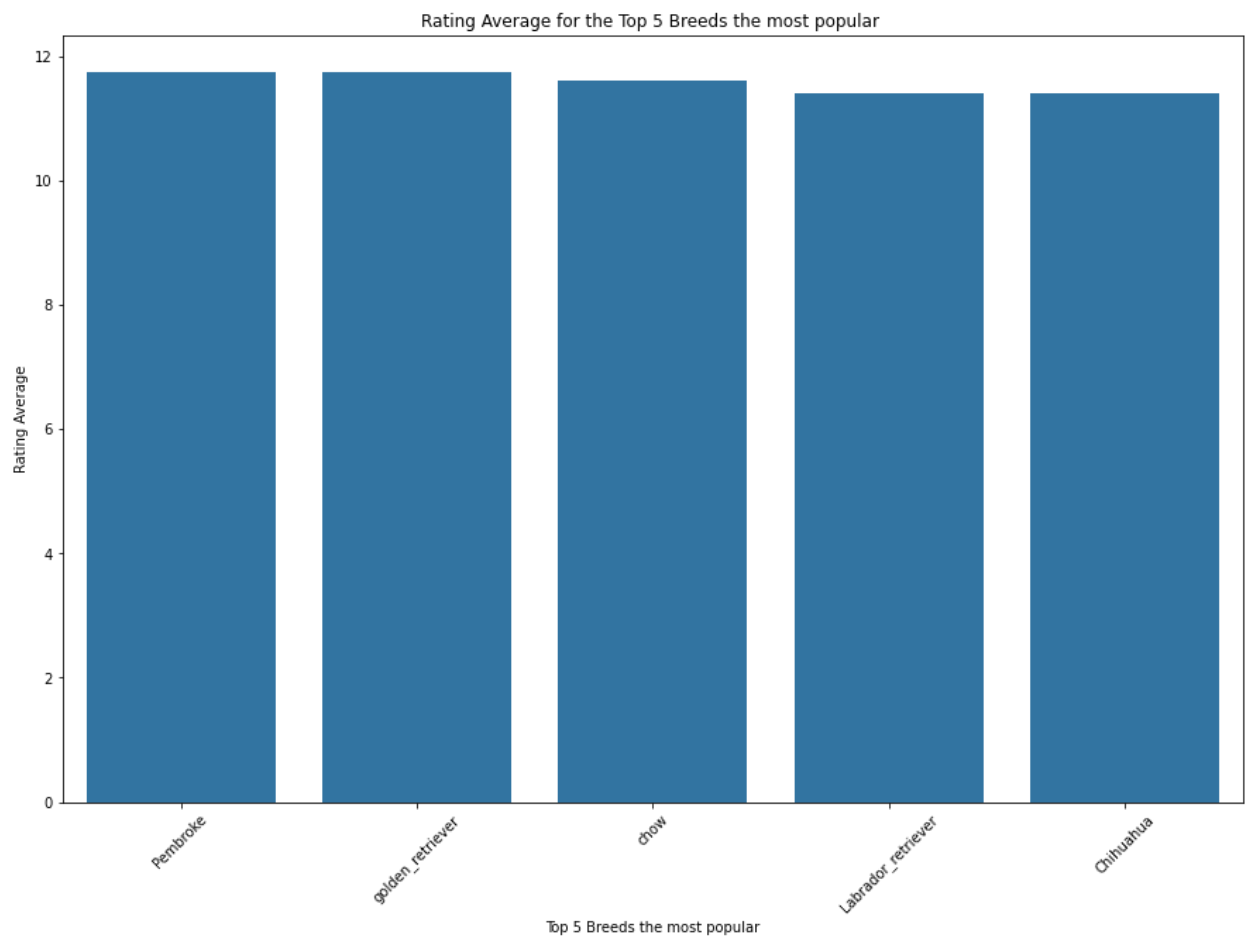## I. Question 1 : Which dog breed is the most posted ? The Top 5 ?

The most posted dog breed by WeRateDogs is a **Golden Retriever**, with a **Labrador Retriever** coming in as the second most popular breed, then **Pembroke**, **Chihuahua** and **Chow** takes the fifth place :

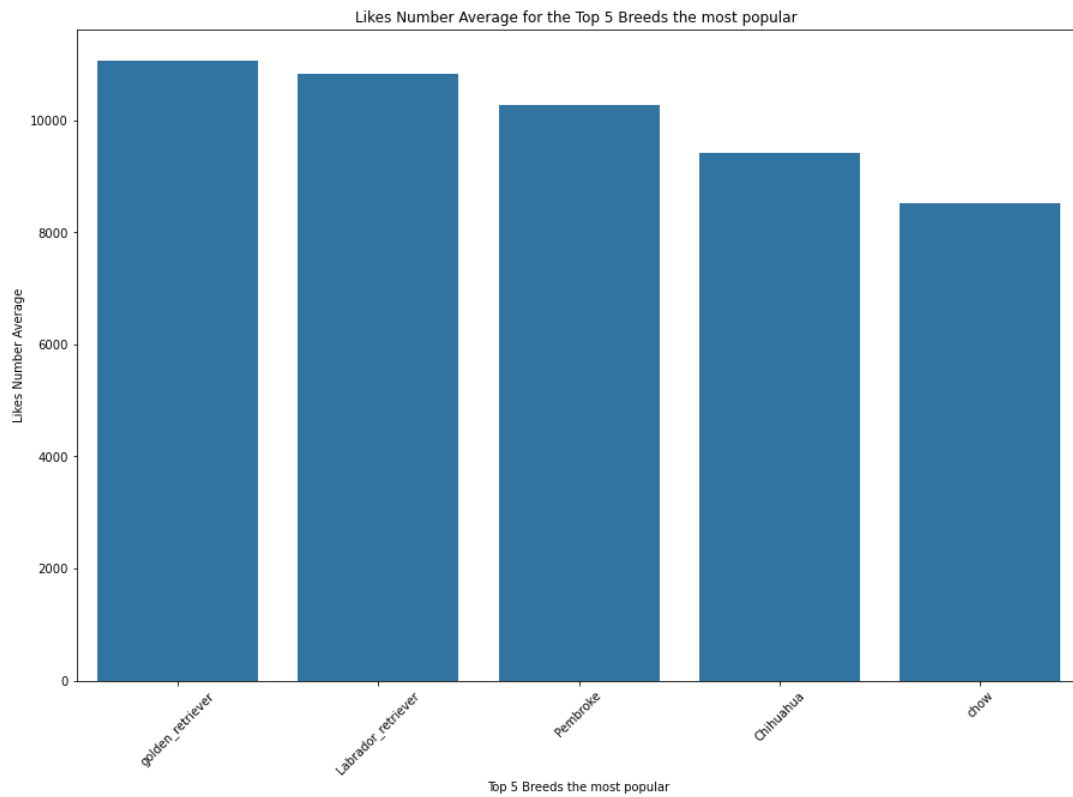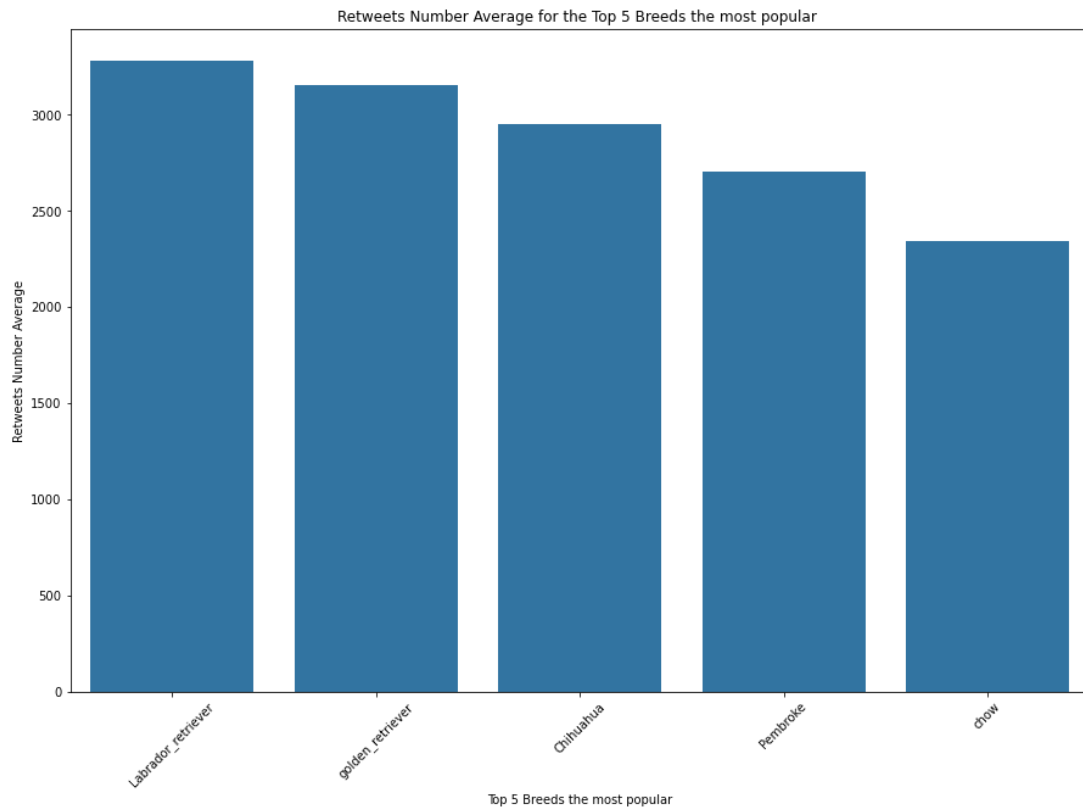Top 5 dog breed the most posted by WeRateDogs

II.   Question 2 : For this Top 5, which one is the best rated ? the most liked ? the most retweeted ?

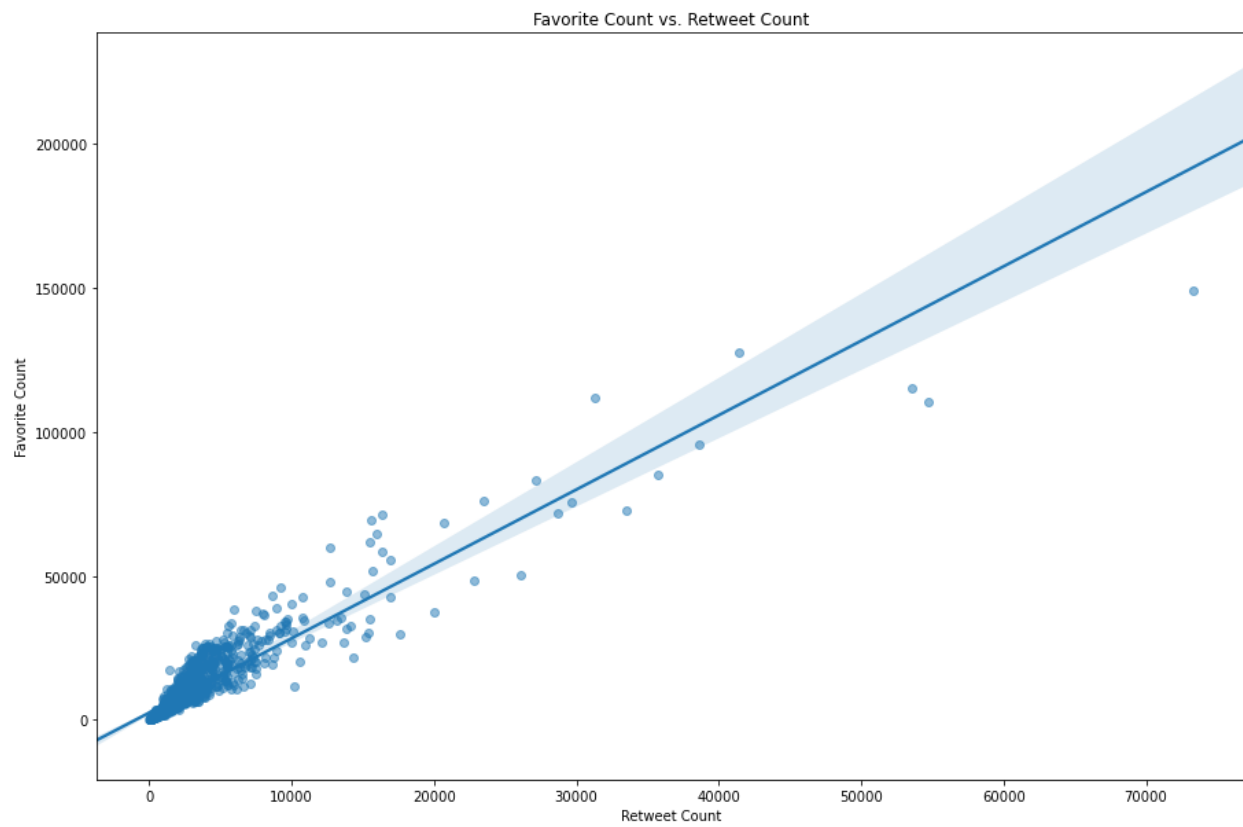For the Top 5 breeds the most popular , we find that :

- **Pembroke** is the best rated
- **Labrador Retriever** is the most retweeted
- **Golden Retriever** is the most liked

Rating Average for the Top 5 Breeds the most popular

Retweets Number Average for the Top 5 Breeds the most popular

Likes Number Average for the Top 5 Breeds the most popular

III.   Question 3 : Is the number of retweets correlated with the number of likes ?

Favorite counts are correlated with **retweet counts** - this is a strong positive correlation :

Favorite Count vs. Retweet Count

IV. Question 4 : Is Rating correlated with Number of Retweets and Number of Likes ?

Ratings are moderately correlated with **Number of Retweets** and **Number of Likes** - it is a positive correlation :

Rating vs. Retweet Count

Rating vs. Favorite Count