Wrangle and Analyze Data Report

Name: Gilbert Woasieworm Adjei

This report includes insights gained from analyzing the WeRataDogs twitter page.

The data is gathered from WeRataDogs twitter archive and it was wrangled and cleaned with the help of python and the jupyter notebook.

After gathering the data, there were some quality issues which includes the following,

1. Missing data
2. Time data recognize as object
3. Unwanted data
4. Different names for tweet IDs

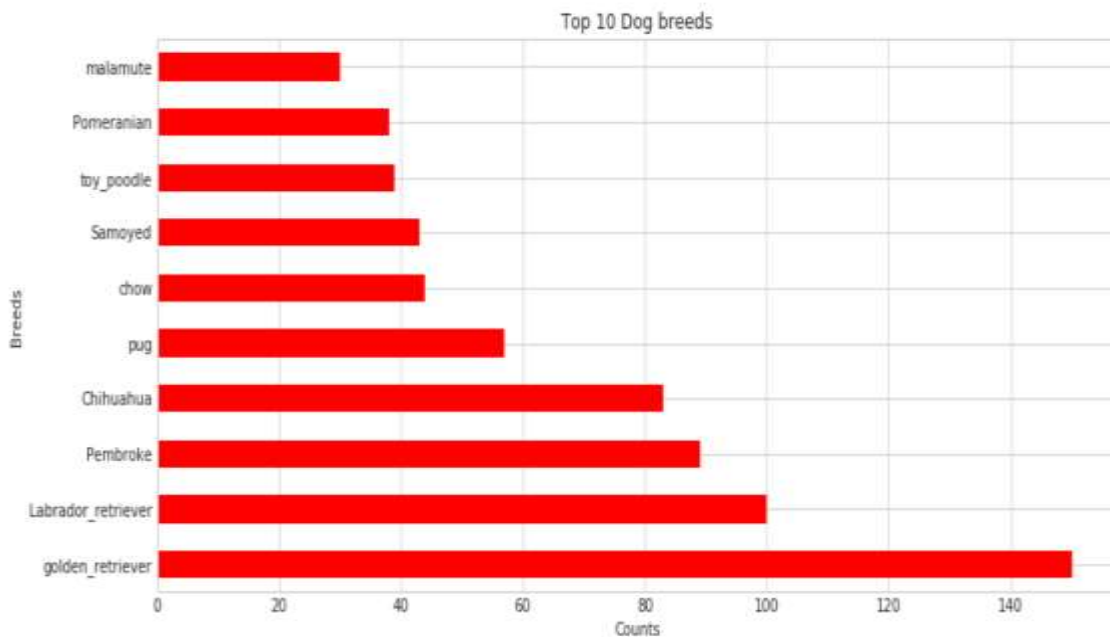Also, there were some tidiness issues such as,

1. Different Data Frames
2. DataFrames has different shapes

These issues were addressed and the cleaned data stored.

The data cleaning stage was followed by the visualization stage where the bar charts and regression plots were used to gain insights from the data and relationships between some of the data.
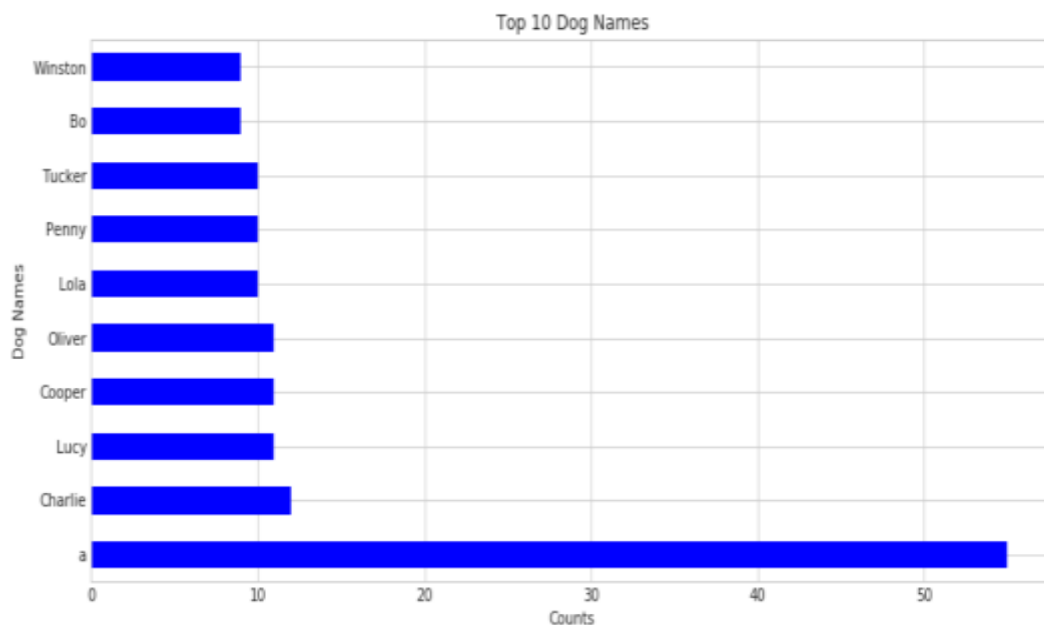
The visualizations display the top 10 dog breeds, top 10 dog names and also, shows that there is a strong positive correlation between the retweet counts and favorite counts from the data.

The horizontal graph below shows the top 10 dog breeds from the wrangled data. This shows that the Golden retriever, the Labrador retriever and the Pembroke are the top three dogs breeds to be rated.

The next horizontal graph below shows the top 10 dog names from the wrangled data.
This shows the counts of the names of the top ten dogs to be rated.

The regression plot below shows the correlation between the favorite counts and retweet counts from the data. It shows that there is a strong positive correlation between the favorite and retweet counts.