

SEPTEMBER 07, 2022

Data Wrangling Report

Gathering, Assessing, Cleaning and
Visualizations

PRESENTED TO

Udacity

PRESENTED BY

Dania Alaa
Adimi






Table of Contents

1 Data Visualization

2 Correlation Heatmap

3 Tweets over time

4 Dogs stages

5 Dogs stages with
Ratings

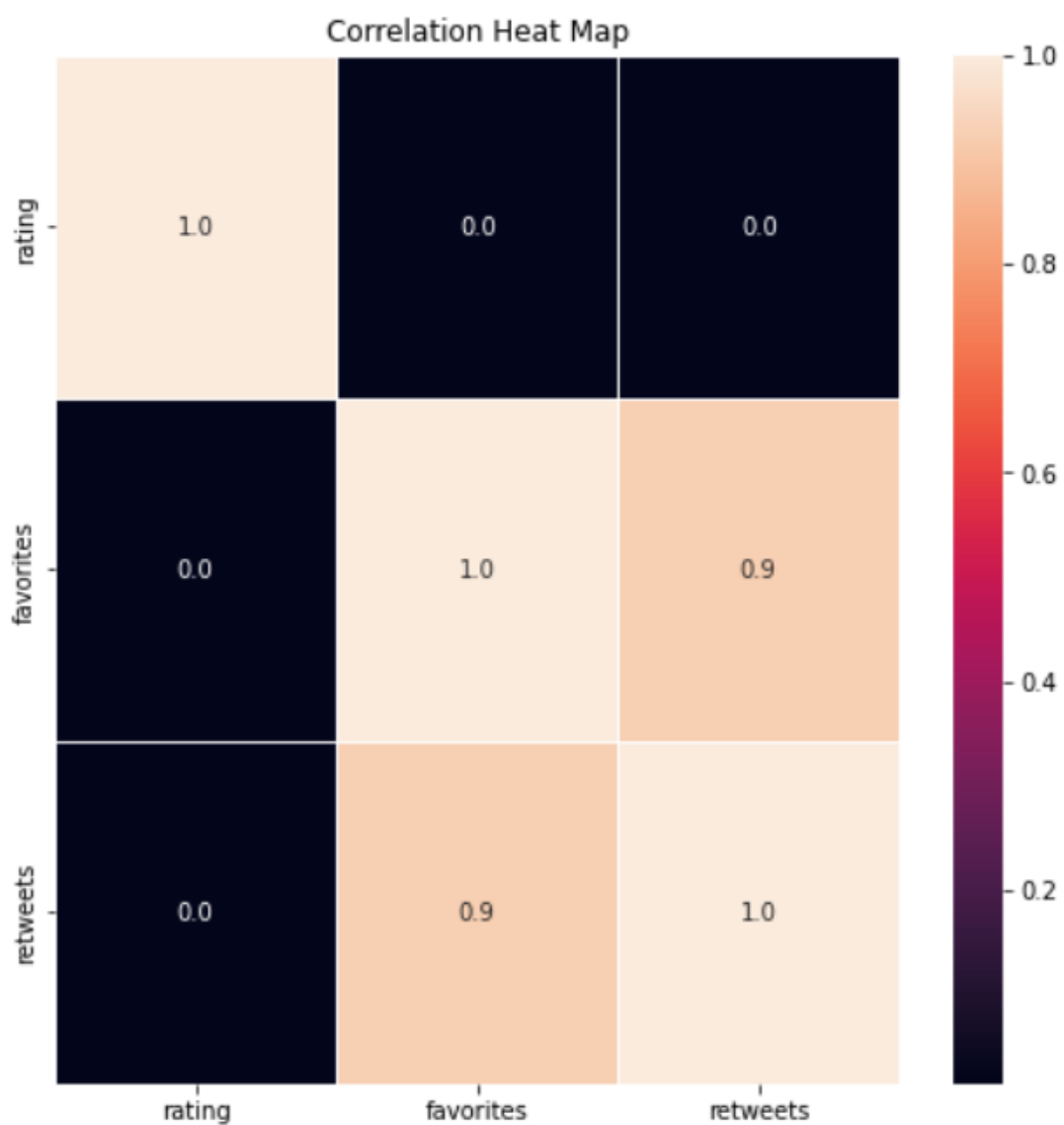
Data Visualization

Data visualization gives us a clear idea of what the information means by giving it visual context through maps or graphs. This makes the data more natural for the human mind to comprehend and therefore makes it easier to identify trends, patterns, and outliers within large data sets.

Since our dataset is now cleaned and ready. I can do some exciting visuals and extract meaning from them. In this report, I will be presenting the 04 visuals that I chose to work on.

Correlation Heatmap

In order to see the correlation between our variables, we can plot a heatmap.

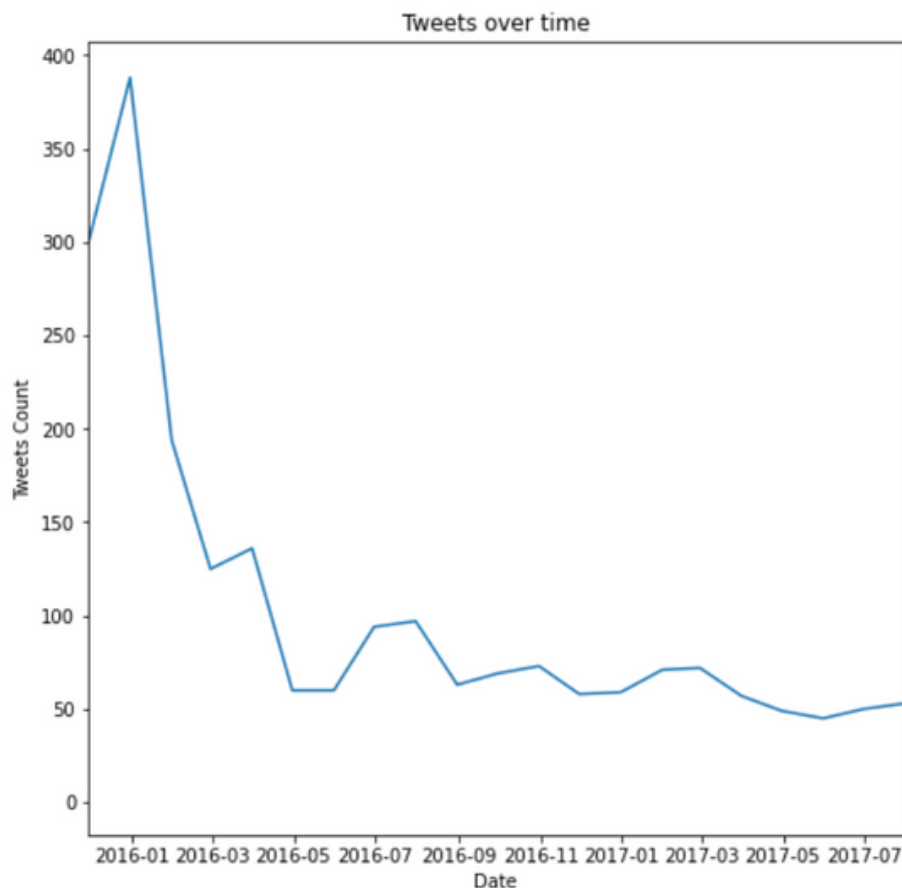


Insights

- We can clearly see that there's a strong correlation between favorites and retweets, which kind of expected.
- There's no correlation between retweets and rating, favorites and rating.

Tweets over time

In order to see the evolution of tweets over time, we can use a graph for better insights.

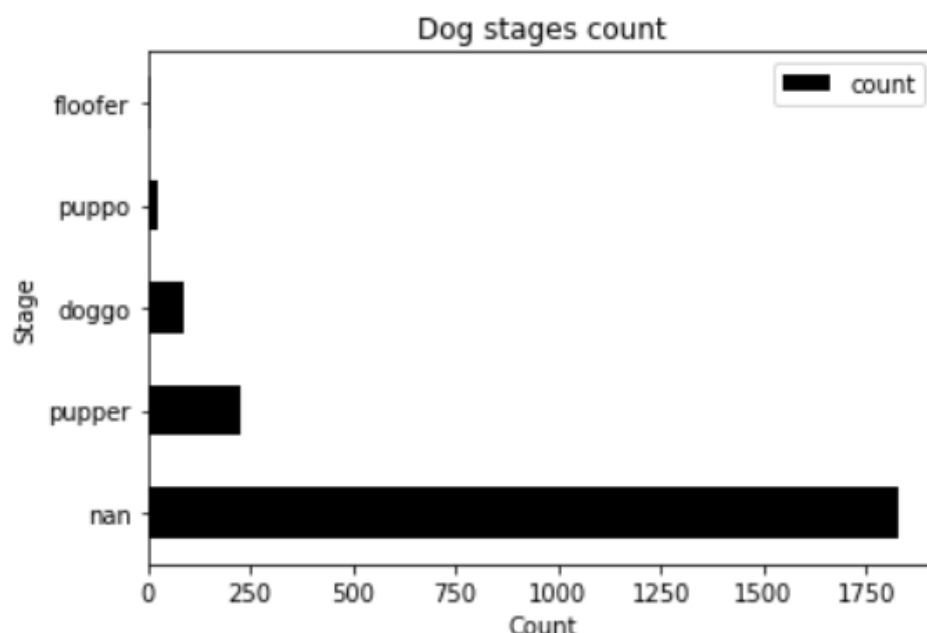


Insights

- We can clearly see that the tweets count was decreasing tremendously.

Dogs stages

Another interesting thing that we can do is to check which are the dogs stages that are more frequent in our dataset. We can do that using a bar plot.

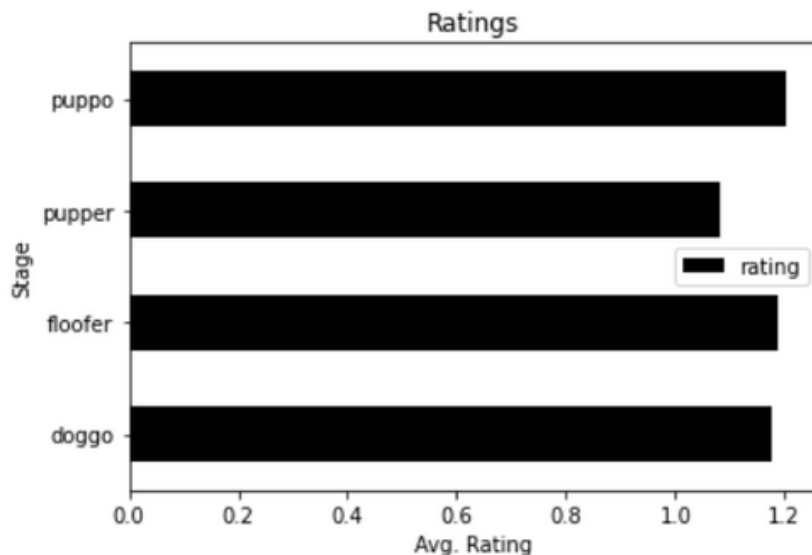


Insights

- Dogs in Pupper stage of dog life cycle get most tweets, which is expected.
- However, most dogs in our dataset are missing the stage.

Dogs stages with ratings

Even more interesting is to check which are the dogs stages that are have the highest avg. rating.



Insights

- We can clearly see that puppos are on the top of the list.