

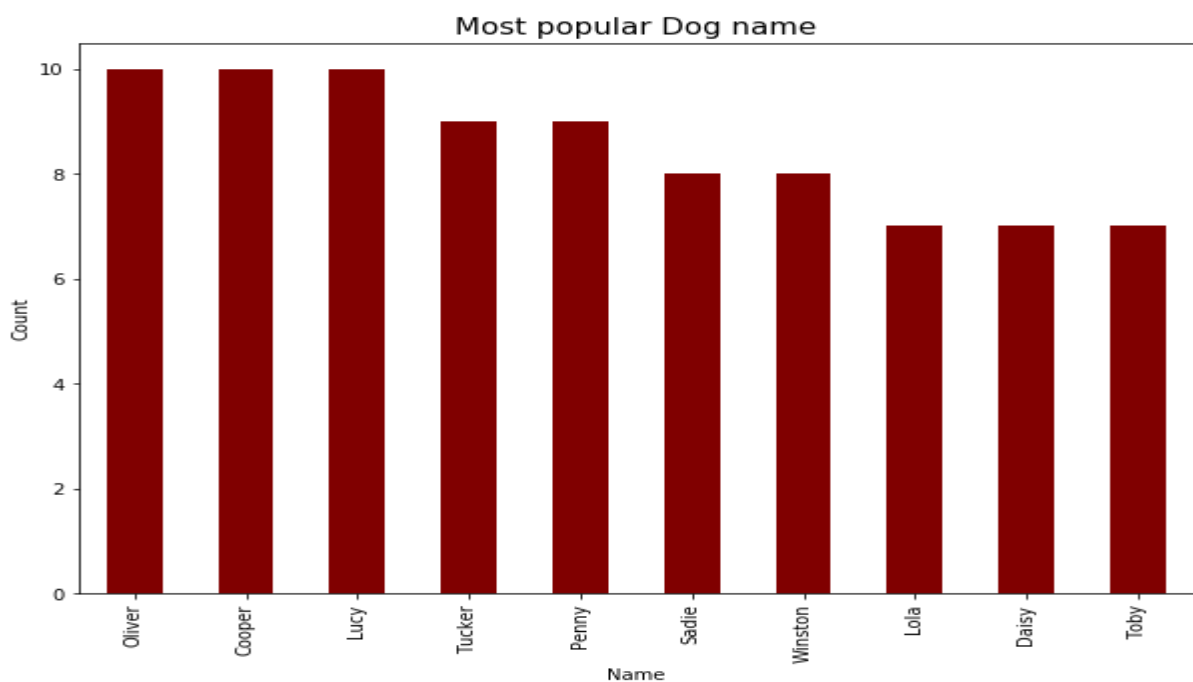
Data Analysis and Visualization

After cleaning our dataset we proceed to visualising our data and to gain useful insights from it. We had four research questions which we wanted to answer.

Research Question 1: What are the most common dog names?

To answer this question, we chose the top 10 most common dog names through `value_counts()` since there are a lot of different dog names to simplify our visualisation. We plotted a bar chart for our analysis.

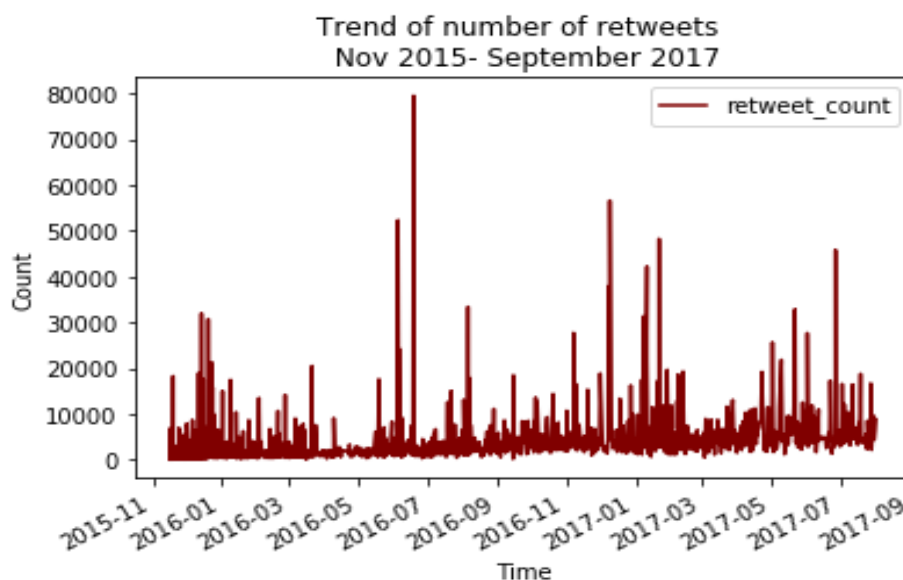
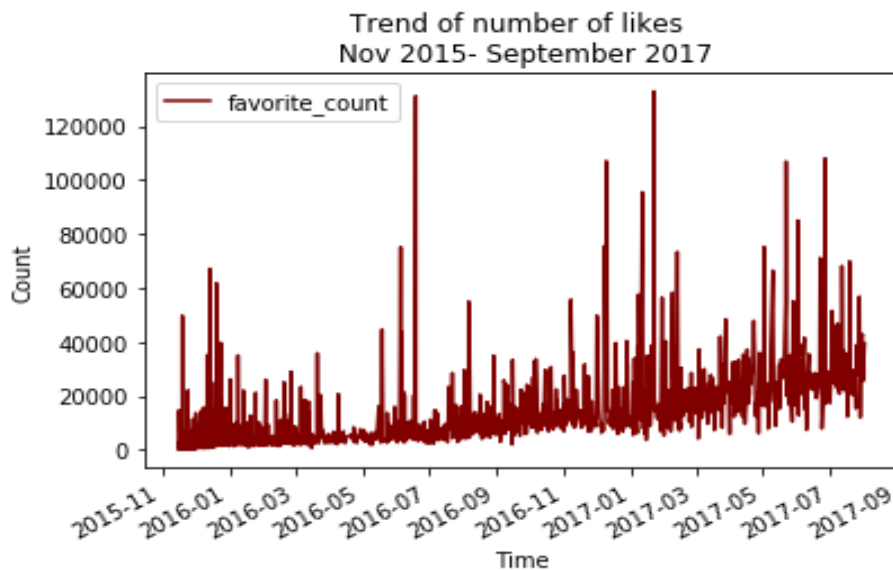
Insight: We saw that Lucy, Oliver and Cooper are the most common dog names. Each was named 10 times in this dataset.



Research Question 2: What is the trend of number of likes and retweets in a tweet?

For this we wanted to see the trend of likes and retweets in a tweet by **Weratedogs** and if it has increased or decreased over the last 2 years. We plotted a Line chart to show this trend using `plot.line()`.

Insight: We could see that number of likes and retweets have increased on an average over the last 2 years



Research Question 3: What is the average number of likes and retweets for a tweet?

We wanted to see that if the page tweets a post, on an average how many likes and retweets is it expected to get. We used `describe()` and `mean()` on our dataset to give us the possible results.

Insight: Average number of likes and retweets in a tweet by Weratedogs were 8896 and 2767 respectively.

Research Question 4: What is the most viral tweet?

Here we wanted to identify the most viral tweet this page has seen over this 2-year period from Nov 2015 – Sep 2017. For this we again make use of `describe()` and further queried the post with most likes.

Insight: The most viral tweet was the one that got 132,810 likes and was a post about a supportive puppo getting a 13/10 rating.

