Insights and visualizations from twitter account WeRateDogs

By Johan Tibeus

Introduction:

The dataset that I 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. WeRateDogs has over 4 million followers and has received international media coverage. I was given a dataset of ~5000 tweets that WeRateDogs downloaded and sent to Udacity. After wrangling the data was able to perform insights and visualizations on 1994 tweets. There were 3 insights with a matching visualization I wanted to find answers to.

Insights and visualizations:

First, I wanted to know what the most common breed of dog is. Having added the Breed column in my wrangling steps, I simply did a value count of them and it sorted results from highest to lowest count. The answer to my question is that Golden Retriever is the most common dog.

Note. NaN with the count of 308 represents data fields that where left blank (no dog breed entered)

Insight 1: What is the most common dog breed?

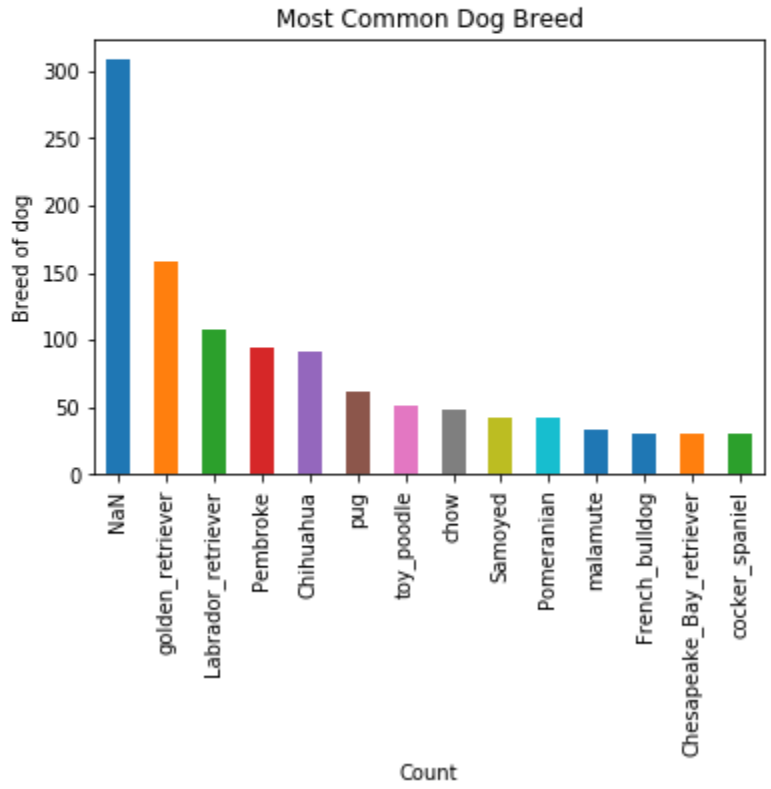
Answer: Golden Retriever

```
twitter_archive_clean.breed.value_counts()
```

```
NaN                 308
golden_retriever    158
Labrador_retriever  108
Pembroke             95
Chihuahua            91
pug                  62
toy_poodle           51
chow                 48
```

I also choose to do a visual bar chart to display this information. I only included the dog breeds that had 30 or more counts.

```
# Bar chart to visualize most common dog breeds
breed = twitter_archive_clean.groupby('breed').filter(lambda x: len(x) >= 30)
breed['breed'].value_counts().plot(kind = 'bar')
plt.title('Most Common Dog Breed')
plt.xlabel('Count')
plt.ylabel('Breed of dog');
```



Secondly, I wanted to know what the most common dog stage is. In my data wrangling I combined the four different stages into one column, so I did a value count of them and got my answer. Pupper is the most common dog stage.

Insight 2: What is the most common dog stage?

Answer: Pupper ¶

```
twitter_archive_clean.stage.value_counts()
```

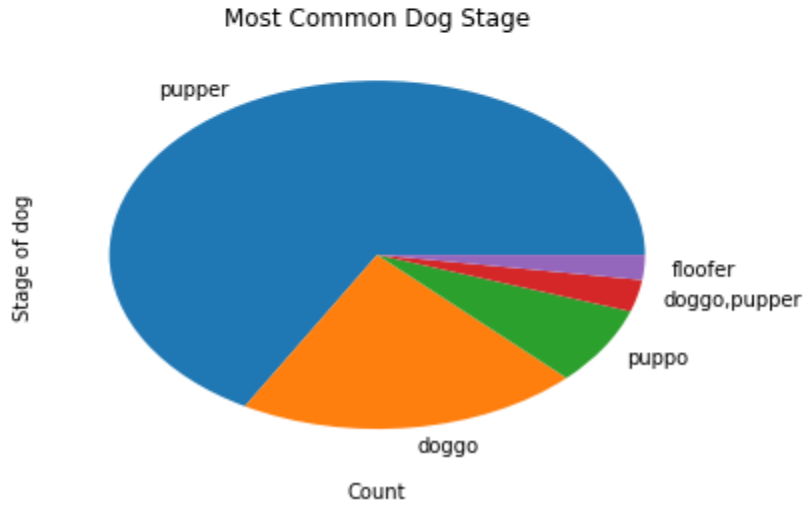
```
pupper            203
doggo              63
puppo              22
doggo,pupper        9
floofer             7
doggo, floofer      1
doggo,puppo         1
Name: stage, dtype: int64
```

Here is the visualization, this time in a pie chart. I only included dog stages with a count of 2 or greater to avoid cluttering the graph.

Visualization 2: What is the most common dog stage?

Answer: Pupper

```
# Pie chart to visualize most common dog stage
stage = twitter_archive_clean.groupby('stage').filter(lambda x: len(x) >= 2)
stage['stage'].value_counts().plot(kind = 'pie')
plt.title('Most Common Dog Stage')
plt.xlabel('Count')
plt.ylabel('Stage of dog');
```



In my third insight and visualization, I wanted to know what the most common dog name is. Again I did a value count in the name column and it listed names in descending order. The most popular dog name is Charlie.

Note. None with the count of 644 represents data fields that where left blank (no dog name entered)

Insight 3: What is the most common dog name?

Answer: Charlie

```
twitter_archive_clean.name.value_counts()
```

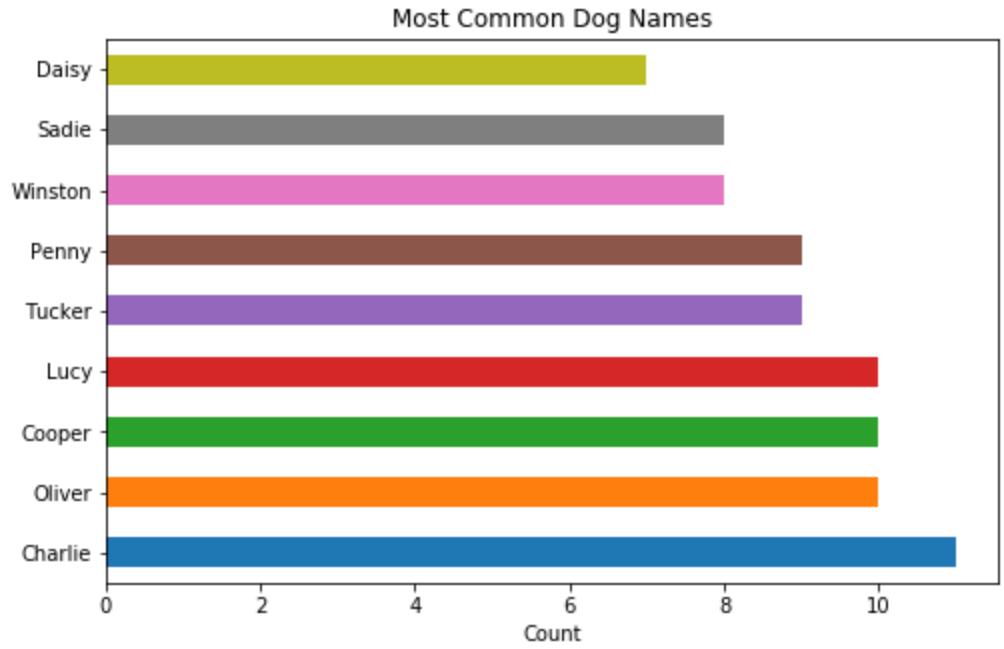
```
None        644
Charlie      11
Oliver       10
Cooper       10
Lucy         10
Tucker        9
Penny         9
Winston       8
Sadie         8
Daisy         7
Lola          7
Toby          7
Bella         6
Bo            6
```

When doing my visualization I wanted to use a line chart, but it was not very helpful when the name “None” was a huge outlier. The graph also looked cluttered if I included all 914 names. To resolve this I used a range of 1-10, where “None” was 0. That way only 9 names displayed and it’s clear to see Charlie is the most common name

Visualization 3: What is the most common dog name?

Answer: Charlie

```
# Line chart to visualize most common dog name
twitter_archive_clean.name.value_counts()[1:10].plot('barh', figsize=(8,5), tit
```



Conclusion:

Doing the data wrangling was the hardest part of this project, once the data is cleaned (quality and tidiness), making sense of it and displaying it in graphs was relatively straight forward.