

# Act Report

Fatema Buhuligah

After storing the master clean dataframe as twitter\_archive\_master.csv, I will analyze and visualize the data with at least three (3) insights and one (1) visualization.

First, I used the describe function to get a brief analysis of the numeric columns.

```
In [296]: df1_clean.describe()
```

Out[296]:

	rating_numerator	rating_denominator	retweet_count	favorite_count	img_num	p1_conf	p2_conf	p3_conf
count	1694.000000	1694.0	1687.000000	1687.000000	1579.000000	1579.000000	1.579000e+03	1.579000e+03
mean	13.496458	10.0	2813.545347	9493.161826	1.245725	0.607810	1.345590e-01	5.968674e-02
std	44.927358	0.0	4645.249729	12701.745607	0.610771	0.268446	1.014477e-01	5.162794e-02
min	10.000000	10.0	11.000000	69.000000	1.000000	0.055379	1.011300e-08	1.740170e-10
25%	10.000000	10.0	661.500000	2286.500000	1.000000	0.373729	5.216185e-02	1.532015e-02
50%	11.000000	10.0	1509.000000	4973.000000	1.000000	0.608537	1.186220e-01	4.836810e-02
75%	12.000000	10.0	3233.000000	12143.000000	1.000000	0.855207	1.950575e-01	9.153815e-02
max	1776.000000	10.0	75351.000000	152326.000000	4.000000	1.000000	4.676780e-01	2.734190e-01

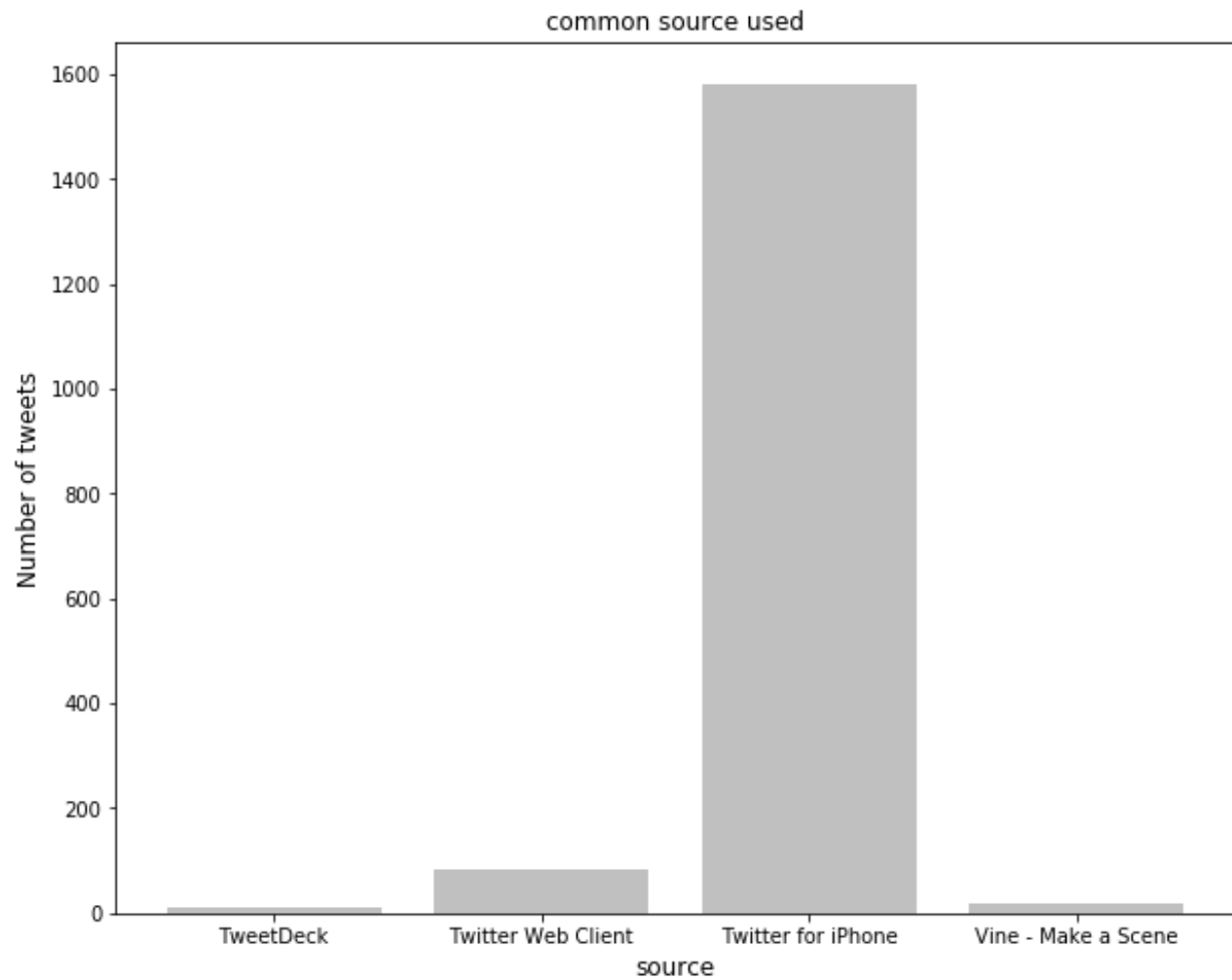
from the above output I found:

- The denominator has the same value (10) for all entries.
- The average numerator score is 13.49.
- The maximum number of retweets is 75351 whereas the maximum number of favorites is 152326.

I will answer these questions in my analysis process:

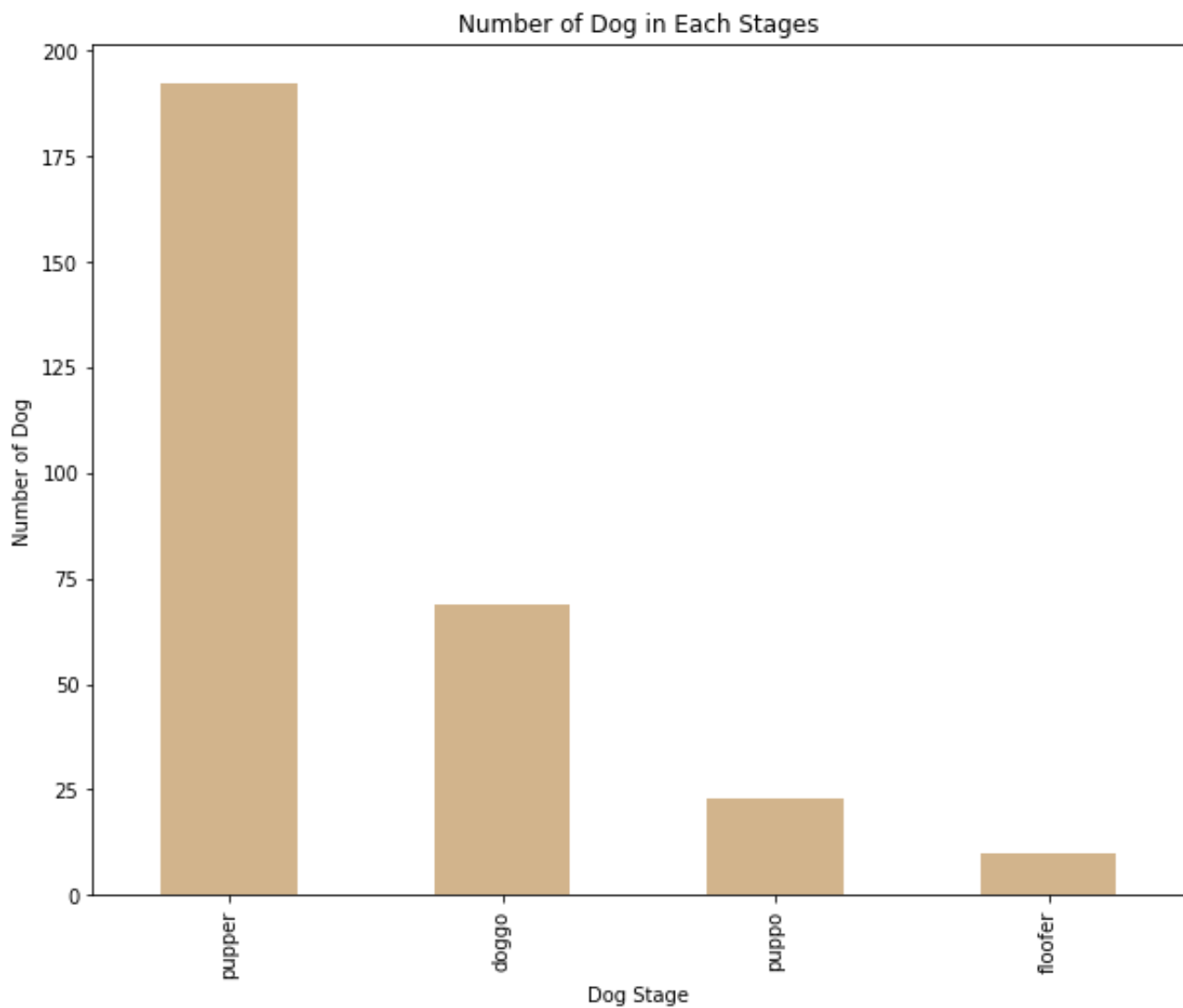
- What is the most common source used for tweet in this dataset?
- What is the most common dog stage?
- What is the most 10 common dog names?
- What is the stage that receive higher favorite count, and what is the stage that receive higher retweet count?

What is the most common source used for tweet in this dataset?



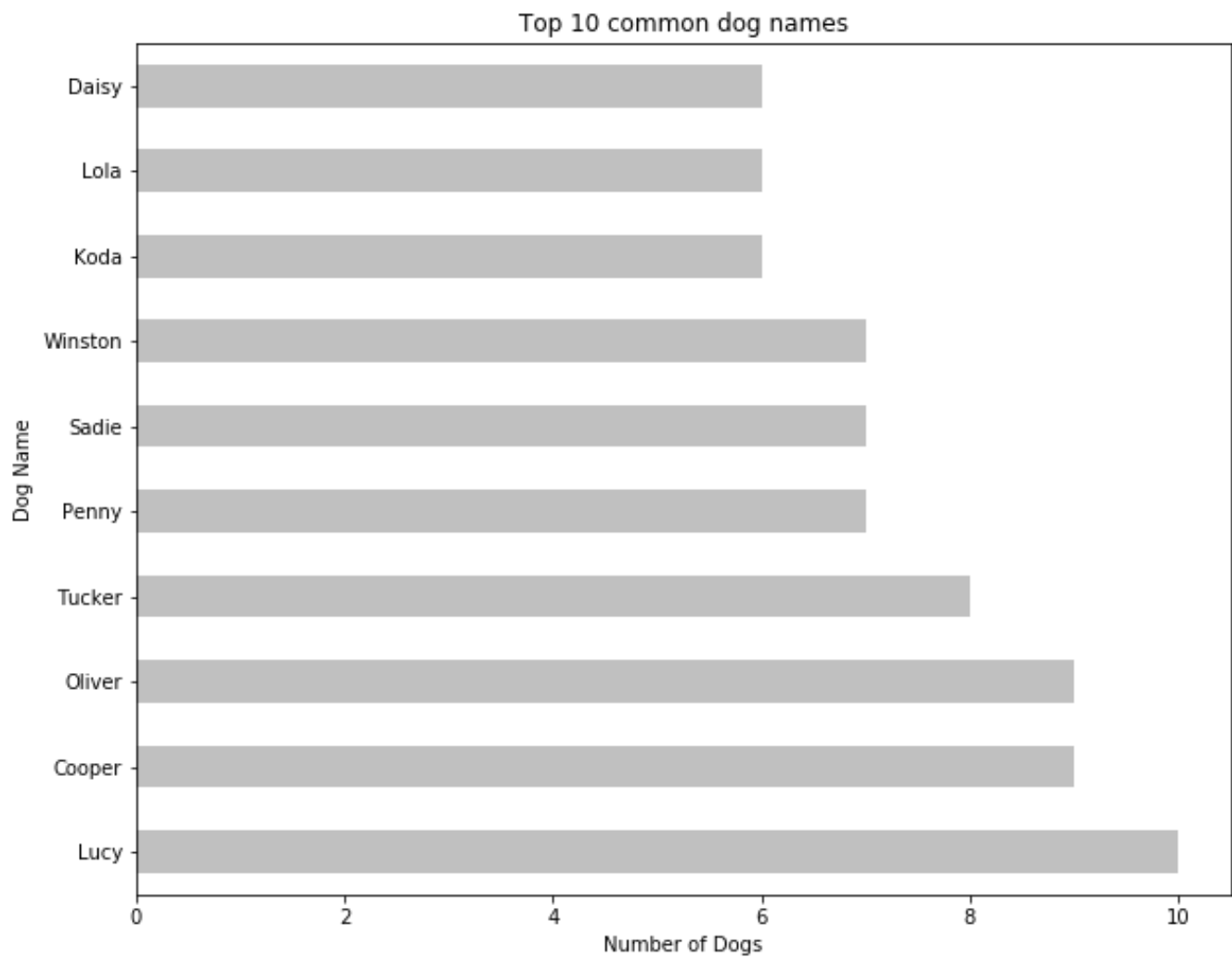
From the above bar chart, I found that the **Twitter for iPhone** is the most common source used.

### What is the most common dog stage?



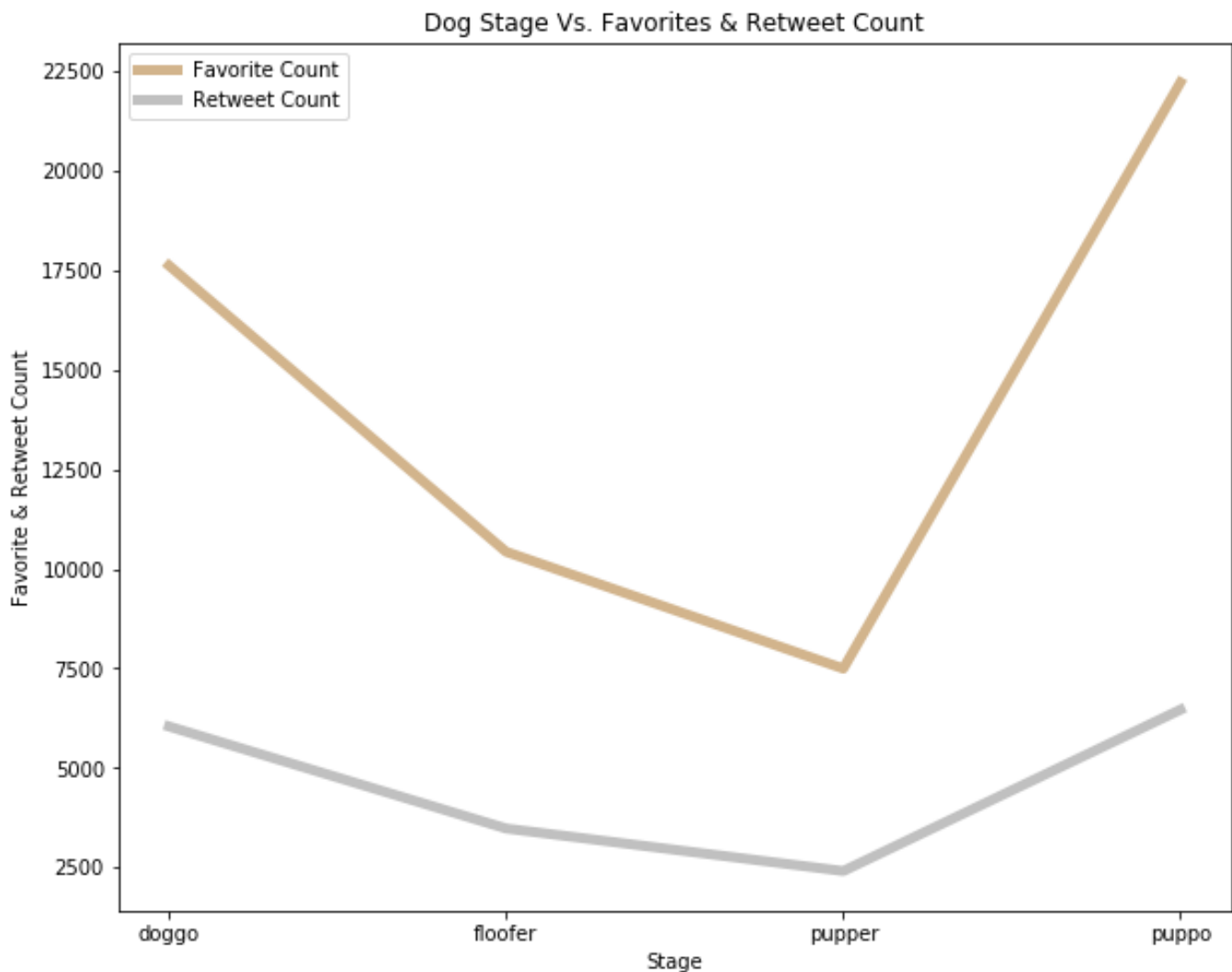
From the above bar chart, I found that the **Pupper** is the most common stage of dog in the dataframe.

What is the most 10 common dog names?



From the above bar chart, I found that the **Lucy** is the most common dog name, then **Cooper & Oliver**.

What is the stage that receive higher favorite count, and what is the stage that receive higher retweet count?



As we can see in the above chart, we found that **Puppo** is the stage that received the higher favorite and retweet then the stage **doggo**. On the other hand, **Pupper** received the least favorite and retweet.