

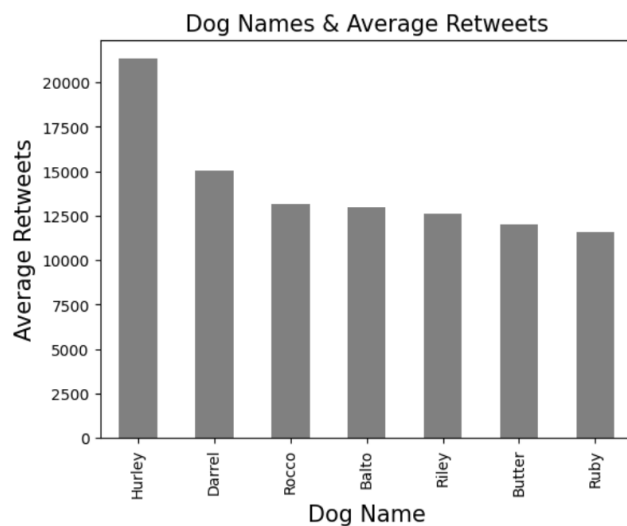
Analyzing and visualizing data

After cleaning data there are many analyses, I found from dataset and those are in following:

Insights:

1. What the dog names with highest average retweeted in 2016?

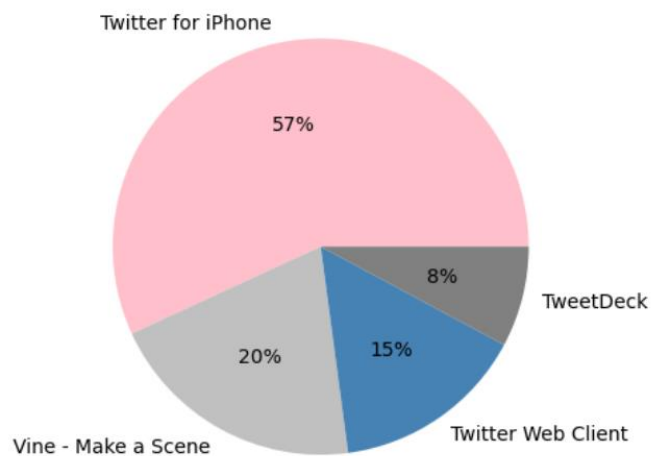
```
Out[78]: name
Hurley    21324.000000
Darrel    15043.000000
Rocco     13140.000000
Balto     12953.000000
Riley     12643.000000
Butter    12014.000000
Ruby      11595.333333
Name: retweet_count, dtype: float64
```



From above chart, the highest average retweeted dogs in 2016 is Hurley with 21324 retweets and next comes Darrel with 15043 retweets and last one Ruby with 11595 retweets.

2. What is the max likes in each source?

```
Out[81]: source
Twitter for iPhone    132810.0
Vine - Make a Scene   47281.0
Twitter Web Client    35178.0
TweetDeck             18278.0
Name: favorite_count, dtype: float64
```



From our Insight, max likes come from Twitter for iPhone with 132810 likes and then the max likes in the Vine - Make a Scene with 47281 likes and next Twitter Web Client with 35178 and last one comes TweetDeck with 18278.

3. What is the most popular dog names by average likes?

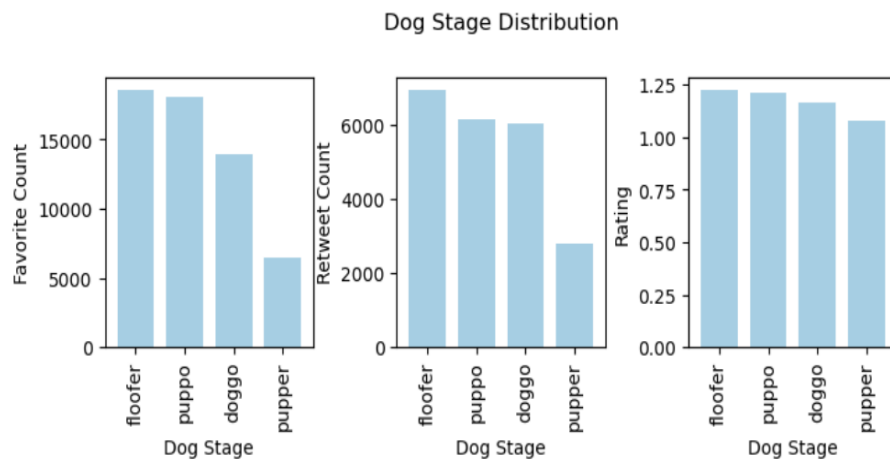
```
Out[87]: name
Duddles      107956.000000
Aja          71144.000000
Canela      69871.000000
Stephan      53507.500000
Jamesy       53413.500000
Barney       48410.000000
Alfy         45880.000000
Dewey        43710.000000
Darla        42908.000000
Cupid        40325.000000
Lassie       39090.000000
Zoey         37860.333333
Clifford     37771.000000
Lilly        36698.500000
Maisey       36689.000000
Thor         36021.000000
Maya         35859.000000
Wiggles      35325.000000
Kody         35324.000000
Harry       35085.000000
Name: favorite_count, dtype: float64
```



As shown above most liked dog names is Duddles with 107956 likes and next comes Aja with 71144 likes and lowest one is Harry with 35085 likes .

Visualization:

4.What the highest dogs stage by average likes and retweets and the rating?



According to our visualization, Floofer is highest dog stage with 18576 likes and 6922 retweeted and 1.225 rating in the average.

Conclusion:

The data wrangling is very important step in the data analysis and the validity of the project depends largely on it.