

---

# DATA ACT REPORT

---

Displays the visualization(s) produced from wrangled data.

According with our dataset clean called all\_data and saved as twitter\_archive\_master.csv this document will provide the graphs obtained and its respective interpretation.

1. First, what is the 10 most popular dog breed?

Using the code: `sort_values()` and `value_counts()`, which sort by the values along either axis from highest to lowest value. Mainly column was image\_predic.

```
popular_dogs=all_data.image_predic.sort_values().value_counts().head(11)
```

```
NaN                268
golden_retriever   145
Labrador_retriever  97
Pembroke           82
Chihuahua          82
pug                60
toy_poodle         49
chow               40
Pomeranian         38
Samoyed            38
Chesapeake_Bay_retriever  30
Name: image_predic, dtype: int64
```

As we can observe, highest number is NaN which means that we do not have that data. Nevertheless, the most popular dog was golden\_retriever with 145 follows by Labrador\_retriever with 97.

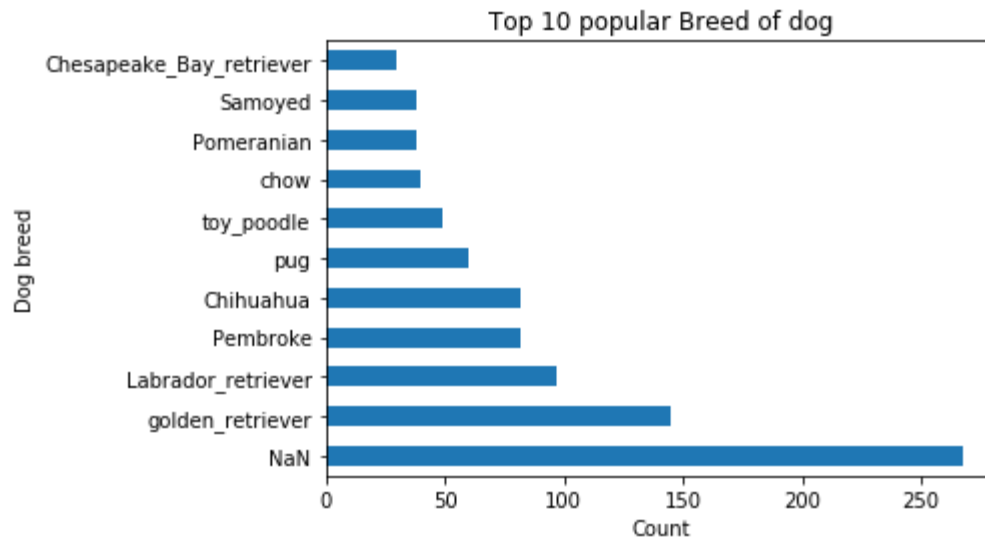
Importing important libraries such as `import matplotlib.pyplot as plt` I plotted the 10 most popular dogs and the code is:

```
popular_dogs.plot(kind = 'barh');
```

```
plt.title('Top 10 popular Breed of dog')
```

```
plt.xlabel('Count')
```

```
plt.ylabel('Dog breed');
```



2. What is the 10 least popular dog breed?

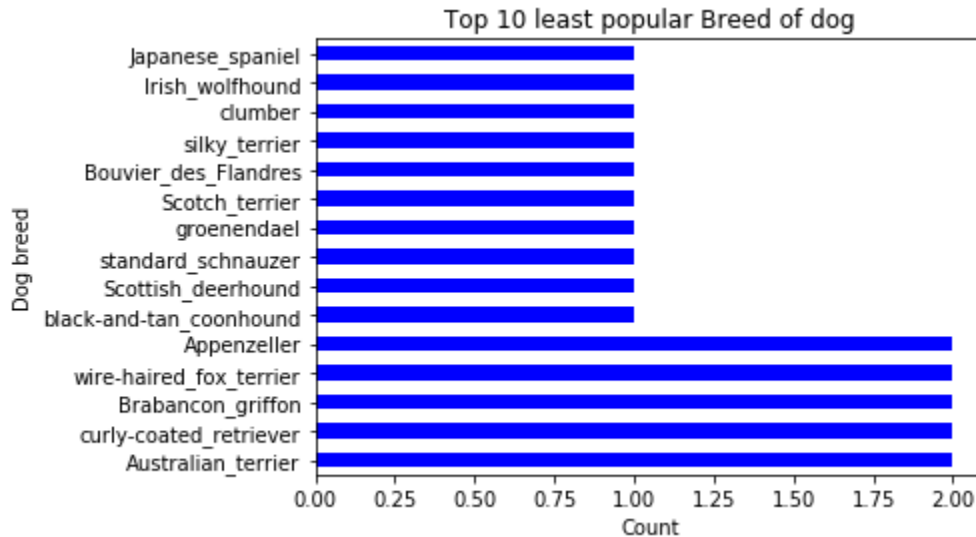
I did the same procedure from first point and the code was:

```
least_popular=all_data.image_predic.sort_values( ascending=True).value_counts().tail(10)
```

least\_popular

```
Australian_terrier      2
curly-coated_retriever  2
Brabancon_griffon       2
wire-haired_fox_terrier  2
Appenzeller             2
black-and-tan_coonhound  1
Scottish_deerhound      1
standard_schnauzer       1
groenendael             1
Scotch_terrier           1
Bouvier_des_Flandres     1
silky_terrier            1
clumber                 1
Irish_wolfhound          1
Japanese_spaniel         1
Name: image_predic, dtype: int64
```

Plotting those data, we obtain:



If I compare with the first graph, I could say that the top 10 least popular dogs and I included 5 dogs more it could be insignificant since the number is smaller.

3. What are the favorite dogs? It means which has the most like.

Using the code: `sort_value` and getting the first 10 value. The main columns were favorites and image\_predic.

```
favorite_dogs= all_data[['favorites','image_predic']]
```

```
favorite_dogs_popular=favorite_dogs.sort_values(by=['favorites'],ascending=False).head(10)
```

```
favorite_dogs_popular
```

	favorites	image_predic
1014	150231.0	chow
517	116050.0	Maltese_dog
129	112595.0	golden_retriever
1054	111130.0	soft-coated_wheaten_terrier
65	96026.0	toy_poodle
433	85385.0	NaN
162	83515.0	cocker_spaniel
111	76363.0	Norwegian_elkhound
1740	76263.0	Samoyed
520	73313.0	chow

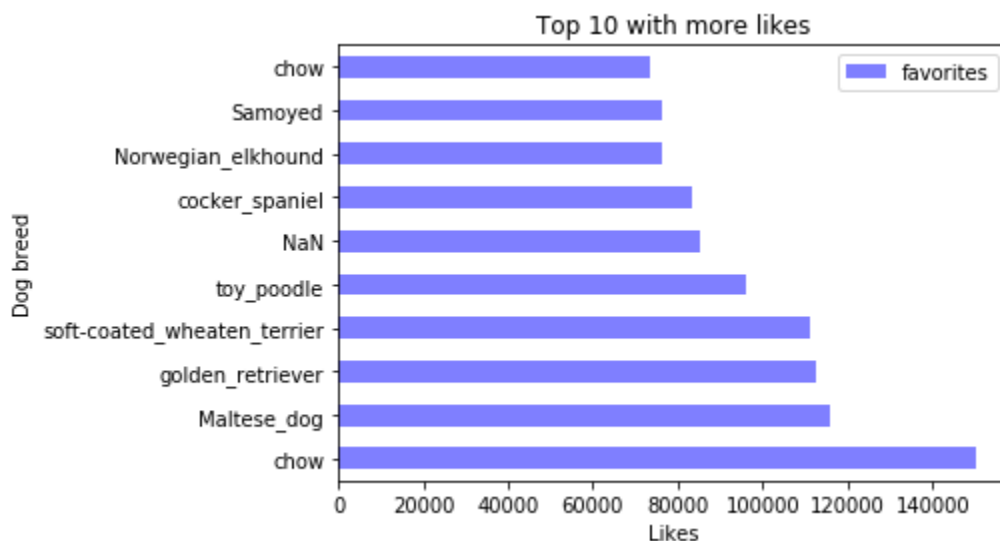
Plotting the last table I got:

```
favorite_dogs_popular.plot(kind='barh',x='image_predic',y='favorites', alpha = 0.5, color='Blue');
```

```
plt.title('Top 10 with more likes')
```

```
plt.xlabel('Likes')
```

```
plt.ylabel('Dog breed');
```



Chow is the favorite follow by Maltese\_dog. Effectively, golden\_retriever has more like which match with the most popular dogs.

#### 4. What does dog obtain more retweets?

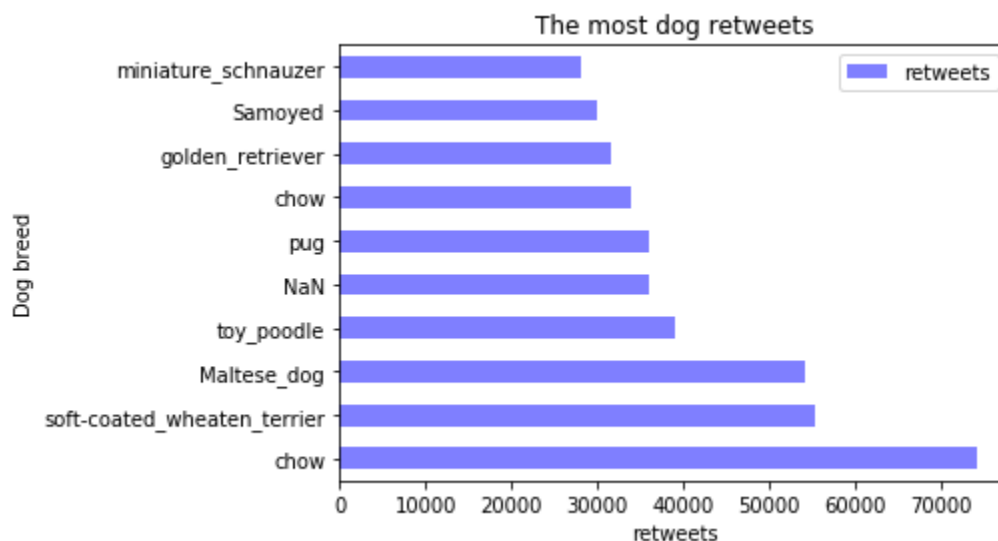
Focus in the columns retweets and image\_predic I could filter which dog has more retweets.

```
retweets_dogs= all_data[['retweets','image_predic']]
```

```
retweets_dogs_most=retweets_dogs.sort_values(by=['retweets'],ascending=False).head(10)
```

```
retweets_dogs_most
```

	retweets	image_predic
1014	74122.0	chow
1054	55233.0	soft-coated_wheaten_terrier
517	54040.0	Maltese_dog
65	38991.0	toy_poodle
433	36007.0	NaN
430	36007.0	pug
520	33815.0	chow
129	31553.0	golden_retriever
1740	29965.0	Samoyed
158	28018.0	miniature_schnauzer



Chow dog obtained 74122 retweets and the second place was soft\_coated\_wheaten\_terrier with 55233. Last graph showed that chow is the favorite, therefore make sense that it had more retweets.

5. What does dog obtain least retweets?

To Obtain this graph I had to focus on retweets and image\_predic columns.

```
retweets_dogs= all_data[['retweets','image_predic']]
least_retweets_dogs=retweets_dogs.sort_values(by=['retweets'],ascending=True).head(10)
least_retweets_dogs
```

	retweets	image_predic
279	1.0	basenji
328	2.0	pug
1270	2.0	NaN
262	2.0	beagle
29	3.0	NaN
1055	6.0	collie
54	8.0	golden_retriever
63	10.0	miniature_pinscher
178	15.0	golden_retriever
1498	16.0	chow

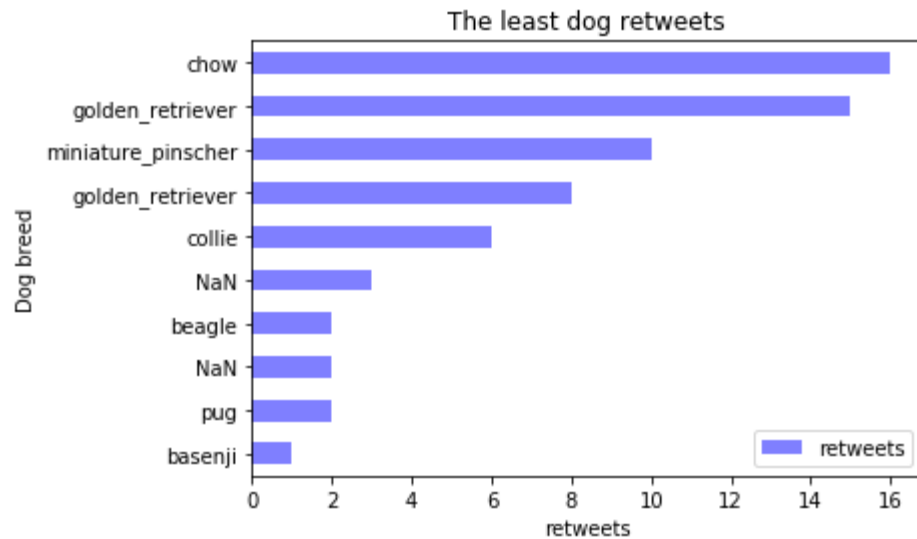
Its graph is:

```
least_retweets_dogs.plot(kind='barh',x='image_predic',y='retweets', alpha = 0.5,
color='blue');
```

```
plt.title('The least dog retweets')
```

```
plt.xlabel('retweets')
```

```
plt.ylabel('Dog breed');
```



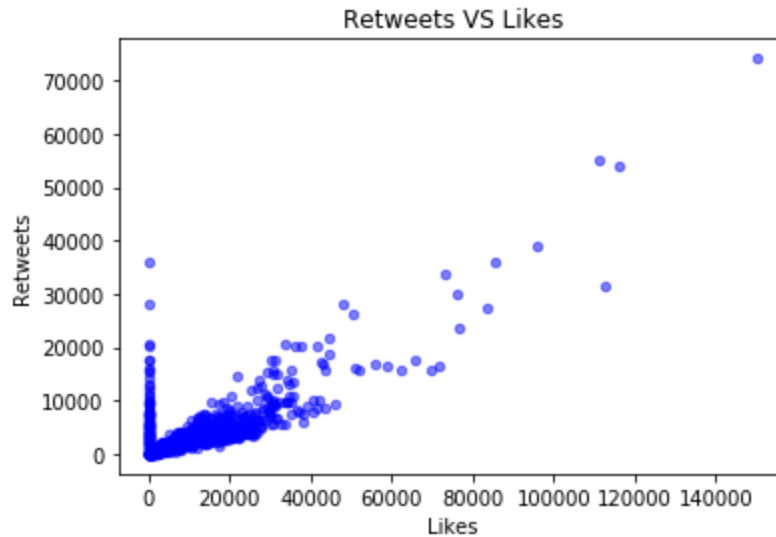
As it was expected, the number the retweets is lower as the least favorites dogs. This make sense because There are few numbers of favorites and as consequences affect the reweets.

#### 6. Plot favorites (like) VS retweets

With whole graph we can plot favorite vs retweets. According with my graph I expect a graph quite linear or exponential which indicate that the number of favorites should be match the number of retweets.

Code was:

```
all_data.plot(kind='scatter',x='favorites',y='retweets', alpha = 0.5, color='b');  
plt.xlabel('Likes')  
plt.ylabel('Retweets')  
plt.title('Retweets VS Likes');
```



As we saw the last graphs (The most dog retweets and the top 10 with more likes), the number of likes and retweets are bigger with some breeds. So, this last graph confirmed a positively correlated with Likes and retweets. Also, exist some breeds that are not favorite as consequence it has least retweets.

## Conclusion

During whole analysis and visualization. I plot the most and least popular dog breeds. In addition, which dog breeds have more 'likes' and retweets. This allowed to verify which retweets vs likes are growing according with the favorites and popular breeds. Also, the least popular favorites make sense with the last graph where there are some points with 1 in like but some of them were retweets.