

## Act Report :

In this report I'll illustrate the insights and observations about the data with one visualization. This steps comes after prepare the data very well and eliminate all the quality and tidiness issues.

After creating the the master dataframe , this is the time to write insights

see the describtion of numeric data :

	rating	retweet_count	favorite_count	pre_conf
count	2356.0	1420.000000	1420.000000	1751.000000
mean	inf	2603.376761	8287.144366	0.548992
std	NaN	4049.137824	11141.534237	0.298804
min	0.0	2.000000	52.000000	0.000010
25%	1.0	578.500000	1786.750000	0.303098
50%	1.1	1311.000000	3903.000000	0.548907
75%	1.2	3103.750000	10280.000000	0.820481
max	inf	56625.000000	107015.000000	0.999956

- we can see that the largest number of retweets of one tweet is 56625
- and for favorite it is equal to 107015
- and the big accurate number of oredition is very good, it is nearly to be 100% , also we can see the prediction is good if we look at the quartiles.

Quantites variables :

	tweet_id	timestamp	text	name	stage_of_dog	jpg_url	prediction
count	2356	2356	2356	2356	2356	1751	1751
unique	2356	2356	2356	957	5	1691	113
top	687826841265172480	2015-11-18 19:43:11 +0000	This is Lizzie. She's about to fist bump the l...	None	None	<a href="https://pbs.twimg.com/ext_tw_video_thumb/81596...">https://pbs.twimg.com/ext_tw_video_thumb/81596...</a>	golden_retriever
freq	1	1	1	745	1976	2	173

- we can clearly see the none names frequent 754 times and this is big number.Unfortunately many entries have none in names.
- The same problem happend with stage\_of\_dog.
- we can see that two same picture tweeted in different times. May we caount this as duplicate.

## Answering the questions:

- What is the most frequent name:

```
None      745
a          55
Charlie    12
Cooper     11
Lucy       11
...
Aqua       1
Odin       1
Huck       1
Brooks     1
Champ      1
Name: name, Length: 957, dtype: int64
```

- a is the most frequent name

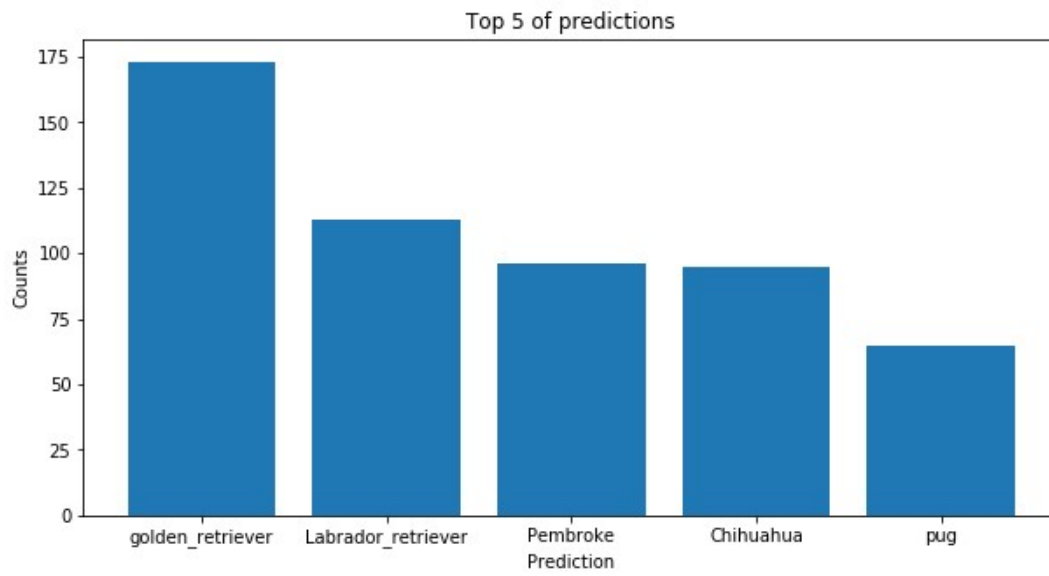
- What is the most favoriated and retweeted tweet:

			rating	retweet_count	favorite_count	pre_conf
prediction	stage_of_dog	name				
Chihuahua	None	Stephan	1.3	56625.0	107015.0	0.505370
standard_poodle	doggo	Bo	1.4	42228.0	95450.0	0.351308
golden_retriever	None	Zoey	1.3	27680.0	85011.0	0.841001
malamute	None	quite	1.2	25661.0	75193.0	0.216163
Labrador_retriever	None	Lilly	1.3	19297.0	73397.0	0.908703
...	...	...	...	...	...	...
vizsla	None	Beau	1.3	NaN	NaN	0.960513
		Bodie	0.9	NaN	NaN	0.876413
		Kreg	1.1	NaN	NaN	0.657275
		Philbert	1.1	NaN	NaN	0.307794
whippet	doggo	Miguel	1.2	NaN	NaN	0.601712

1364 rows × 4 columns

- Most favoraited and retweeted dog is called Stephan and the predection of it's stage is Chihuahua

- Visualization of the Top 5 predictions repeated:



- We can see that the top 5 prediction was to these types