

Untitled

December 12, 2020

1 Analysis

```
In [11]: import pandas as pd
import matplotlib.pyplot as plt
```

```
In [12]: df_master = pd.read_csv('twitter_archive_master.csv')
```

```
In [13]: df_master.head()
```

```
Out[13]:
```

	Unnamed: 0	tweet_id	in_reply_to_status_id	in_reply_to_user_id \
0	0	892420643555336193	NaN	NaN
1	1	892177421306343426	NaN	NaN
2	2	891815181378084864	NaN	NaN
3	3	891689557279858688	NaN	NaN
4	4	891327558926688256	NaN	NaN

	tweet_timestamp \
0	2017-08-01 16:23:56 +0000
1	2017-08-01 00:17:27 +0000
2	2017-07-31 00:18:03 +0000
3	2017-07-30 15:58:51 +0000
4	2017-07-29 16:00:24 +0000

	source \
0	<a href="http://twitter.com/download/iphone" r...
1	<a href="http://twitter.com/download/iphone" r...
2	<a href="http://twitter.com/download/iphone" r...
3	<a href="http://twitter.com/download/iphone" r...
4	<a href="http://twitter.com/download/iphone" r...

	tweet_text	retweeted_status_id \
0	This is Phineas. He's a mystical boy. Only eve...	NaN
1	This is Tilly. She's just checking pup on you...	NaN
2	This is Archie. He is a rare Norwegian Pouncin...	NaN
3	This is Darla. She commenced a snooze mid meal...	NaN
4	This is Franklin. He would like you to stop ca...	NaN

	retweeted_status_user_id	retweeted_status_timestamp	...	\
--	--------------------------	----------------------------	-----	---

0	NaN	NaN	...
1	NaN	NaN	...
2	NaN	NaN	...
3	NaN	NaN	...
4	NaN	NaN	...

	dog_rating_out_of_ten	rating_denominator	dog_name	dog_category	img_num	\
0	13	10	Phineas	None	NaN	
1	13	10	Tilly	None	1.0	
2	12	10	Archie	None	1.0	
3	13	10	Darla	None	NaN	
4	12	10	Franklin	None	2.0	

	dog_breed_prediction	prediction_confidence	timestamp				\
0	NaN	NaN	Tue	Aug	01	16:23:56 +0000	2017
1	Chihuahua	0.323581	Tue	Aug	01	00:17:27 +0000	2017
2	Chihuahua	0.716012	Mon	Jul	31	00:18:03 +0000	2017
3	NaN	NaN	Sun	Jul	30	15:58:51 +0000	2017
4	basset	0.555712	Sat	Jul	29	16:00:24 +0000	2017

	favorites	retweets
0	39467	8853
1	33819	6514
2	25461	4328
3	42908	8964
4	41048	9774

[5 rows x 21 columns]

In [14]: df_master.tail()

```
Out[14]:
```

	Unnamed: 0	tweet_id	in_reply_to_status_id	\
2075	2075	666049248165822465	NaN	
2076	2076	666044226329800704	NaN	
2077	2077	666033412701032449	NaN	
2078	2078	666029285002620928	NaN	
2079	2079	666020888022790149	NaN	

	in_reply_to_user_id	tweet_timestamp	\
2075	NaN	2015-11-16 00:24:50 +0000	
2076	NaN	2015-11-16 00:04:52 +0000	
2077	NaN	2015-11-15 23:21:54 +0000	
2078	NaN	2015-11-15 23:05:30 +0000	
2079	NaN	2015-11-15 22:32:08 +0000	

	source	\
2075	<a href="http://twitter.com/download/iphone" r...	
2076	<a href="http://twitter.com/download/iphone" r...	

```

2077 <a href="http://twitter.com/download/iphone" r...
2078 <a href="http://twitter.com/download/iphone" r...
2079 <a href="http://twitter.com/download/iphone" r...

```

	tweet_text	retweeted_status_id	\
2075	Here we have a 1949 1st generation vulpix. Enj...	NaN	
2076	This is a purebred Piers Morgan. Loves to Netf...	NaN	
2077	Here is a very happy pup. Big fan of well-main...	NaN	
2078	This is a western brown Mitsubishi terrier. Up...	NaN	
2079	Here we have a Japanese Irish Setter. Lost eye...	NaN	

	retweeted_status_user_id	retweeted_status_timestamp	...	\
2075	NaN	NaN	...	
2076	NaN	NaN	...	
2077	NaN	NaN	...	
2078	NaN	NaN	...	
2079	NaN	NaN	...	

	dog_rating_out_of_ten	rating_denominator	dog_name	dog_category	img_num	\
2075	5	10	None	None	1.0	
2076	6	10	a	None	1.0	
2077	9	10	a	None	1.0	
2078	7	10	a	None	1.0	
2079	8	10	None	None	1.0	

	dog_breed_prediction	prediction_confidence	\
2075	miniature_pinscher	0.560311	
2076	Rhodesian_ridgeback	0.408143	
2077	German_shepherd	0.596461	
2078	redbone	0.506826	
2079	Welsh_springer_spaniel	0.465074	

	timestamp	favorites	retweets
2075	Mon Nov 16 00:24:50 +0000 2015	111	41
2076	Mon Nov 16 00:04:52 +0000 2015	311	147
2077	Sun Nov 15 23:21:54 +0000 2015	128	47
2078	Sun Nov 15 23:05:30 +0000 2015	132	48
2079	Sun Nov 15 22:32:08 +0000 2015	2535	532

[5 rows x 21 columns]

```
In [17]: df_master.dog_category.value_counts()
```

```

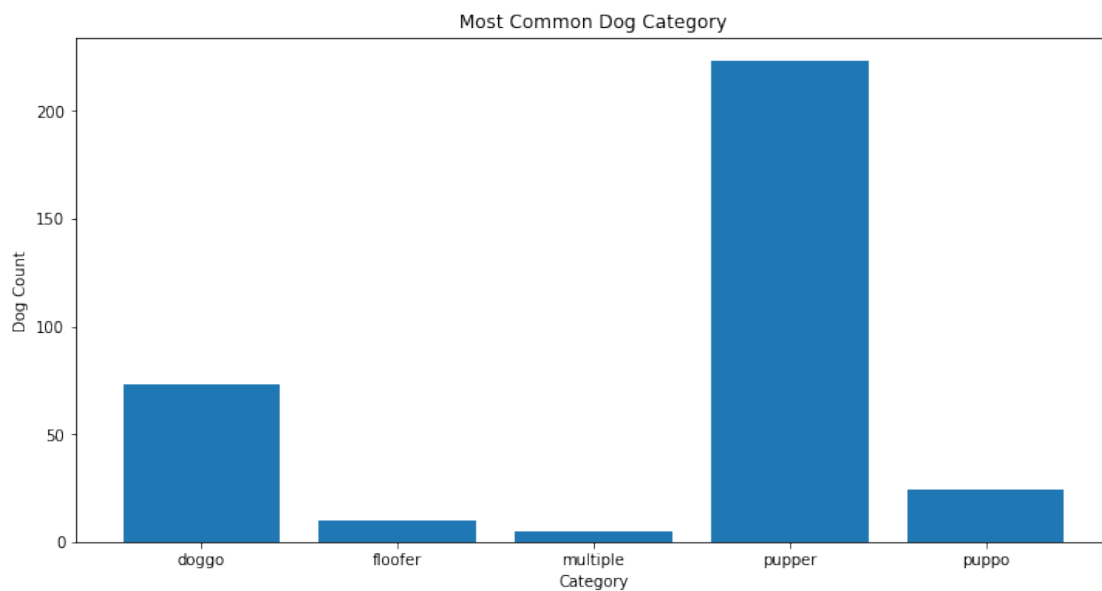
Out[17]: None      1744
         pupper     223
         doggo      73
         puppo      24
         floofer     10

```

```
multiple      5
none          1
Name: dog_category, dtype: int64
```

```
In [18]: dog_type = ['pupper', 'doggo', 'puppo', 'floofer', 'multiple']
dog_counts = [223, 73, 24, 10, 5]
```

```
fig,ax = plt.subplots(figsize = (12,6))
ax.bar(dog_type, dog_counts, width = 0.8)
ax.set_ylabel('Dog Count')
ax.set_xlabel('Category')
plt.title("Most Common Dog Category")
plt.show()
```



```
In [20]: df_master.dog_rating_out_of_ten.describe()
```

```
Out[20]: count      2080.000000
mean         11.683173
std          39.814228
min           0.000000
25%          10.000000
50%          11.000000
75%          12.000000
max          1776.000000
Name: dog_rating_out_of_ten, dtype: float64
```

```
In [21]: df_master.dog_breed_prediction.value_counts()
```

```

Out[21]: golden_retriever      134
         Labrador_retriever    91
         Pembroke              88
         Chihuahua             78
         pug                   54
         chow                   41
         Samoyed               40
         Pomeranian           38
         toy_poodle            37
         malamute              29
         cocker_spaniel        27
         French_bulldog        26
         Chesapeake_Bay_retriever 23
         miniature_pinscher    21
         German_shepherd       20
         Siberian_husky        19
         Staffordshire_bullterrier 19
         Shetland_sheepdog     18
         Eskimo_dog            18
         Maltese_dog           17
         beagle                17
         Shih-Tzu              17
         Cardigan              17
         Italian_greyhound     16
         Rottweiler            16
         kuvasz                15
         Lakeland_terrier      15
         vizsla                13
         basset                13
         Pekinese              13
         ...
         Norwich_terrier       4
         bluetick              4
         Mexican_hairless      4
         Weimaraner            4
         Brabancon_griffon     3
         Scottish_deerhound    3
         Welsh_springer_spaniel 3
         Irish_water_spaniel   3
         curly-coated_retriever 3
         Ibizan_hound          3
         komondor              3
         Leonberg              3
         Afghan_hound          3
         cairn                 3
         briard                3
         Greater_Swiss_Mountain_dog 3
         Appenzeller           2

```

Sussex_spaniel	2
giant_schnauzer	2
Australian_terrier	2
toy_terrier	2
wire-haired_fox_terrier	2
black-and-tan_coonhound	2
groenendael	1
standard_schnauzer	1
Scotch_terrier	1
Japanese_spaniel	1
clumber	1
silky_terrier	1
EntleBucher	1

Name: dog_breed_prediction, Length: 111, dtype: int64

2 Insights

The most popular dog type is a "pupper"

The most popular dog, based on image predictions, is a Golden Retriever
 he median and mean ratings out of ten are 11 and 11.7 respectively

In []: