

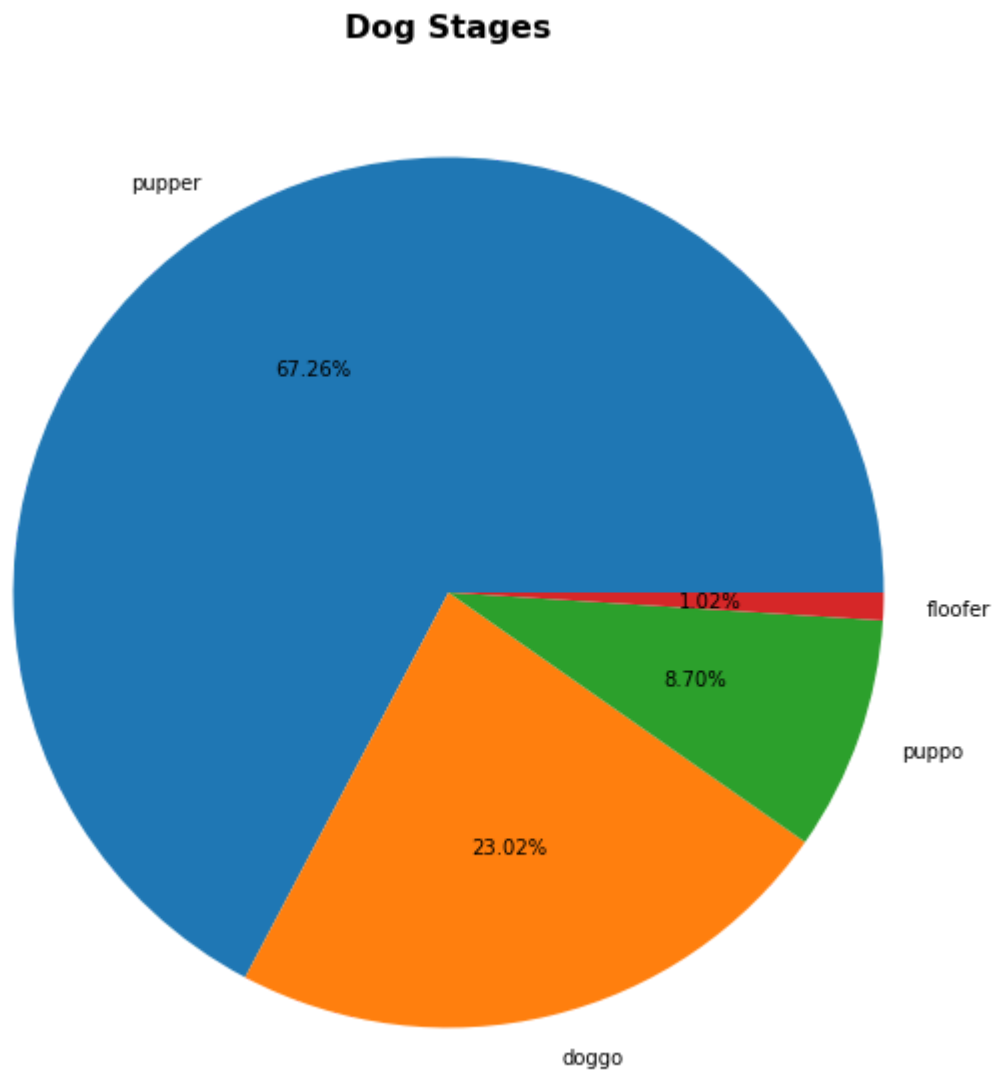
WeRateDogs analysis and insights

After cleaning the data set I made some visualizations functions in order to make more information about the data

Pie chart for doggy meter

```
In [101... #showing a Pie chart for the doggy meter
tweets['doggy_meter'].value_counts().plot(kind='pie', autopct='%.2f%%', figsize=(10,10)

plt.ylabel('')
plt.title('Dog Stages', weight='bold', fontsize=16);
```



we can notice that Pupper has the largest percentage with a 67.3 and the least was floofer

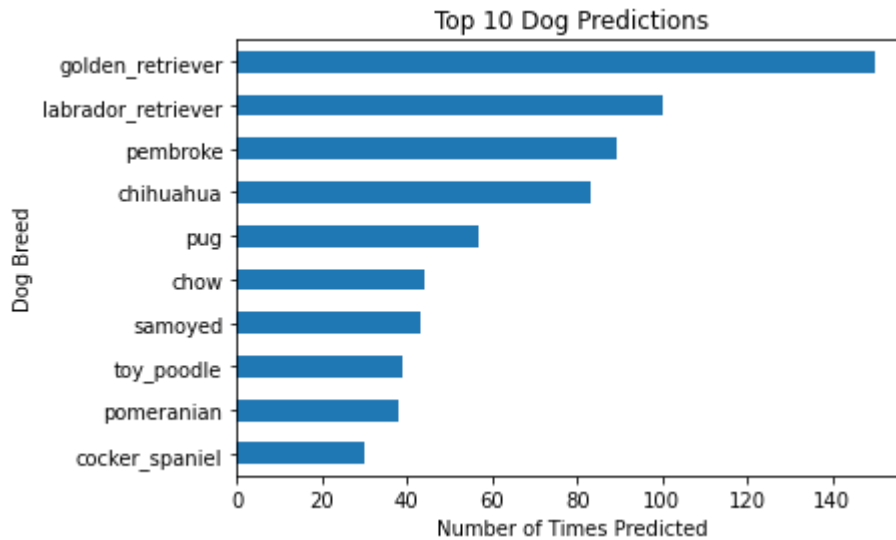
Its not a strange result as the pupper is the medium dogs witch in tern would be

the most common dogs

In []:

Top 10 predictions

```
In [103... #Make a bar plot for top 10 dog breeds
tweets['p1'].value_counts()[9::-1].plot(kind='barh')
plt.title('Top 10 Dog Predictions')
plt.xlabel('Number of Times Predicted')
plt.ylabel('Dog Breed')
plt.fontsize = 12
```



In the plot above you will notice that the golden retriever is in about 140 photos, that may give more than conclusion, it may because the most number of photos, or its the most common dog in certain era of time, it depends ofcourse there's a lot of variables may affect that

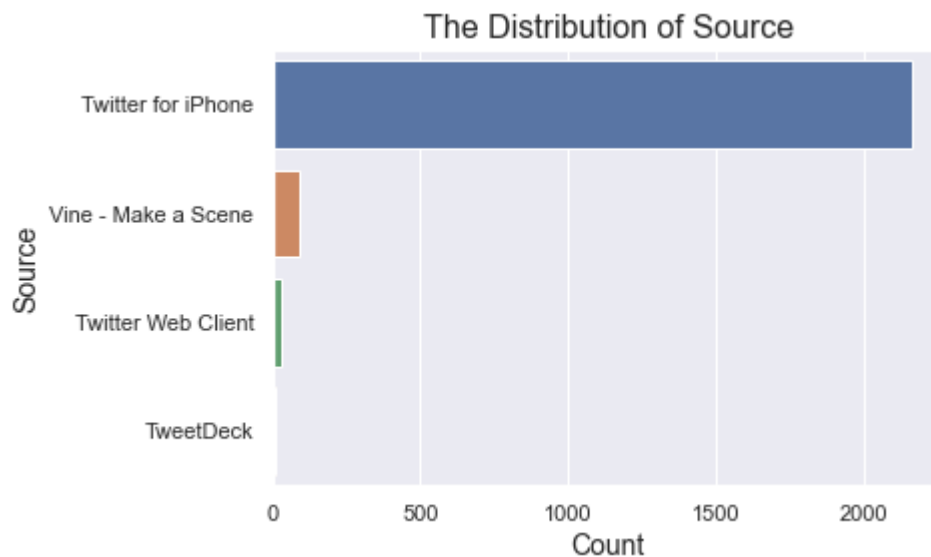
then you will pass the other types that the detection mentioned, to reach the least one witch is cocker spaniel

Source effects

```
In [109... # how the tweets counts affected by the source
sorted_source = tweets['source'].value_counts().index
print(tweets['source'].value_counts())
sns.set(style="darkgrid")
sns.countplot(data = tweets, y = 'source', order = sorted_source)
plt.xticks(rotation = 360)
plt.xlabel('Count', fontsize=14)
plt.ylabel('Source', fontsize=14)
plt.title('The Distribution of Source', fontsize=16)
```

```
Twitter for iPhone      2163
Vine - Make a Scene      91
Twitter Web Client      32
TweetDeck               11
Name: source, dtype: int64
```

Out[109... Text(0.5, 1.0, 'The Distribution of Source')



Twitter for Iphone is the most number of sources, that may give you the sense that apple users are more interested in dogs, that may be translated to someways of richness, the very least source is tweetdeck

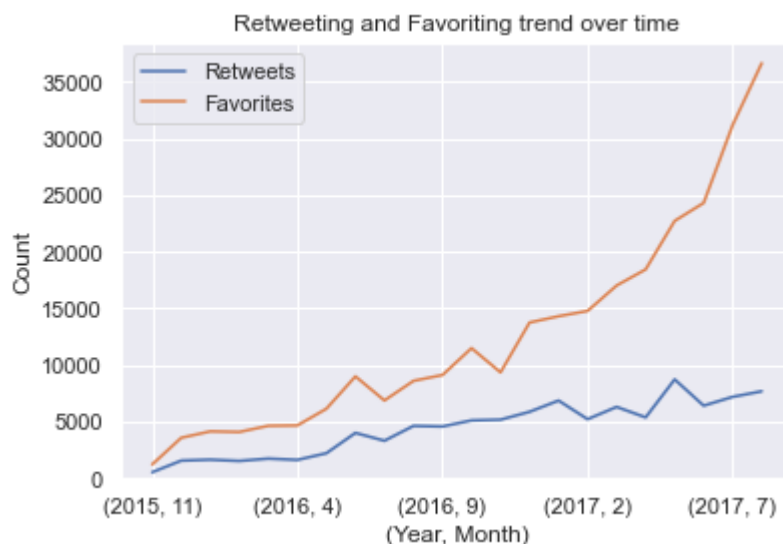
In []:

In []:

Retweets vs favourites

In [140]...

```
# the relation between retweeting and favourite
plt.subplots(1, sharex=True)
tweets.retweet_count.groupby([tweets["timestamp"].dt.year, tweets["timestamp"].dt.month
tweets.favorite_count.groupby([tweets["timestamp"].dt.year, tweets["timestamp"].dt.month
plt.title('Retweeting and Favoriting trend over time')
plt.ylabel('Count')
plt.xlabel('(Year, Month)')
plt.legend(('Retweets', 'Favorites'))
plt.savefig('ret_fav')
```



by the time and noticing the relation between retweets and favourites, you will

find that favourites is increasingly dramatically more than retweets, and the gap is continuously increasing, in 2016 for example the gap was about 3000 but in 2017 the gap was about 30000, that may have a lot of meaning like decreasing in the social sense, or may be an increasing usage of favourite as a way to save tweets to return for sometime later

In []:

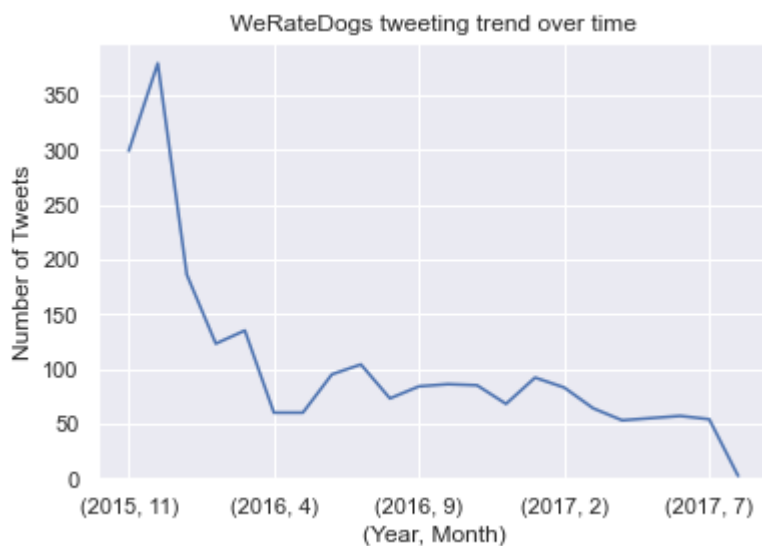
In []:

In []:

Tweets decreasing over the time

In [143...]

```
#tweeting through the time period
data = tweets tweet_id.groupby([tweets["timestamp"].dt.year, tweets["timestamp"].dt.mon
ax = data.plot(kind='line', title='WeRateDogs tweeting trend over time')
ax.set_xlabel("(Year, Month)")
ax.set_ylabel("Number of Tweets")
ax.set_ylim(0, )
plt.savefig('tweet_trend')
```



as tweets was increasing in 2015 but then it face a continuous loose till 2017, the most inclination was at 2016 then decreasing tends to take a fair rate, there was some slightly increasing in april 2016 jan 2017