# Analysis and insight

Visualizing Data

This is a tweets posted trend and the decreased over the time, plot between time and number of tweets.

```
In [86]: data = df.tweet_id.groupby([df.year, df.month]).count()
         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')
```


WeRateDogs tweeting trend over time
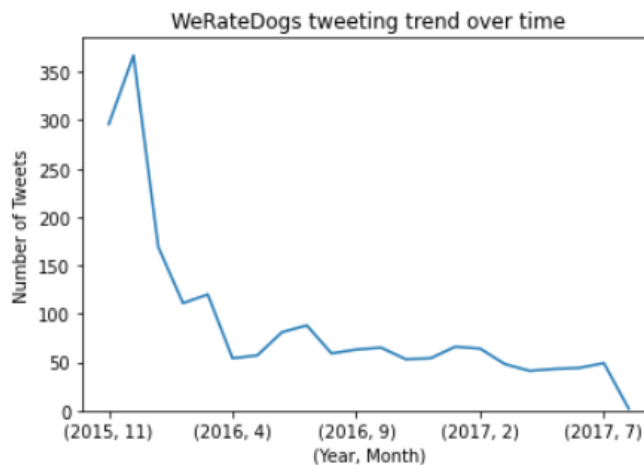
The Twitter account started in 2015, tweeted a lot in frist year. For example, it posted around 300 tweets in the month of November 2015, which increased to around 350 tweets in exactly the next month but its tweeting activity reduced drastically in the subsequent months, averaging between 50 to 100 since April, 2016

```
In [87]: df['source'].value_counts().plot(kind='barh', figsize=(11,5), title='Most used Twitter source').set_xla
         plt.savefig('twitter_source')
```
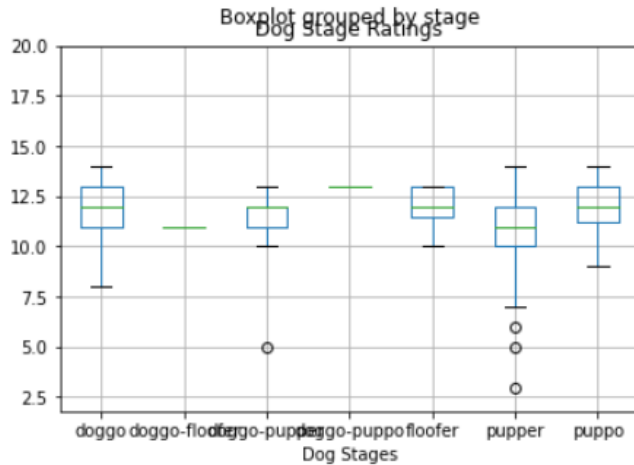
Most used Twitter source
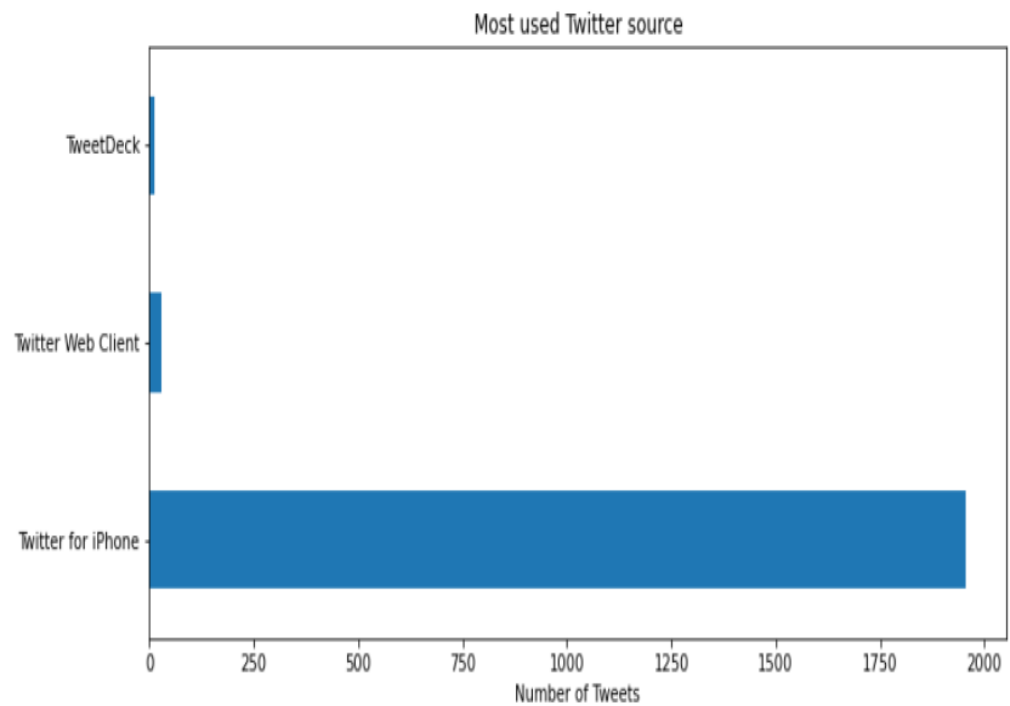
the most common source petween tweetdeck , twitter web cllient and iphone user is iphine
user

Ratings by Dog Stages on WeRateDogs Twitter

```
[85]: df[df['stage'] != 'None'].boxplot(column = ['rating_numerator'], by = ['stage'])

plt.title ('Dog Stage Ratings')
plt.xlabel('Dog Stages')
plt.ylim(ymax=20)
plt.show();
```


Boxplot grouped by stage
Dog Stage Ratings

WeRateDogs appears to rate puppers lower than doggos, floofers and puppos.

Most of the dogs are rated 12/10 , the highest rating received by any dog is 14/10 we can say that WeRateDogs is probably very conservative when it comes to rating a dog the highest

```
In [80]: print('Doggo')
         print('%s\t%s' % ('Mean Retweet Count',
                           round(df.retweet_count[df.stage == 'doggo'].mean())))
         print('%s\t%s' % ('Mean Favorite Count',
                           round(df.favorite_count[df.stage == 'doggo'].mean())))

         print('Floofer')
         print('%s\t%s' % ('Mean Retweet Count',
                           round(df.retweet_count[df.stage == 'floofer'].mean())))
         print('%s\t%s' % ('Mean Favorite Count',
                           round(df.favorite_count[df.stage == 'floofer'].mean())))

         print('Pupper')
         print('%s\t%s' % ('Mean Retweet Count',
                           round(df.retweet_count[df.stage == 'pupper'].mean())))
         print('%s\t%s' % ('Mean Favorite Count',
                           round(df.favorite_count[df.stage == 'pupper'].mean())))

         print('Puppo')
         print('%s\t%s' % ('Mean Retweet Count',
                           round(df.retweet_count[df.stage == 'puppo'].mean())))
         print('%s\t%s' % ('Mean Favorite Count',
                           round(df.favorite_count[df.stage == 'puppo'].mean())))
```
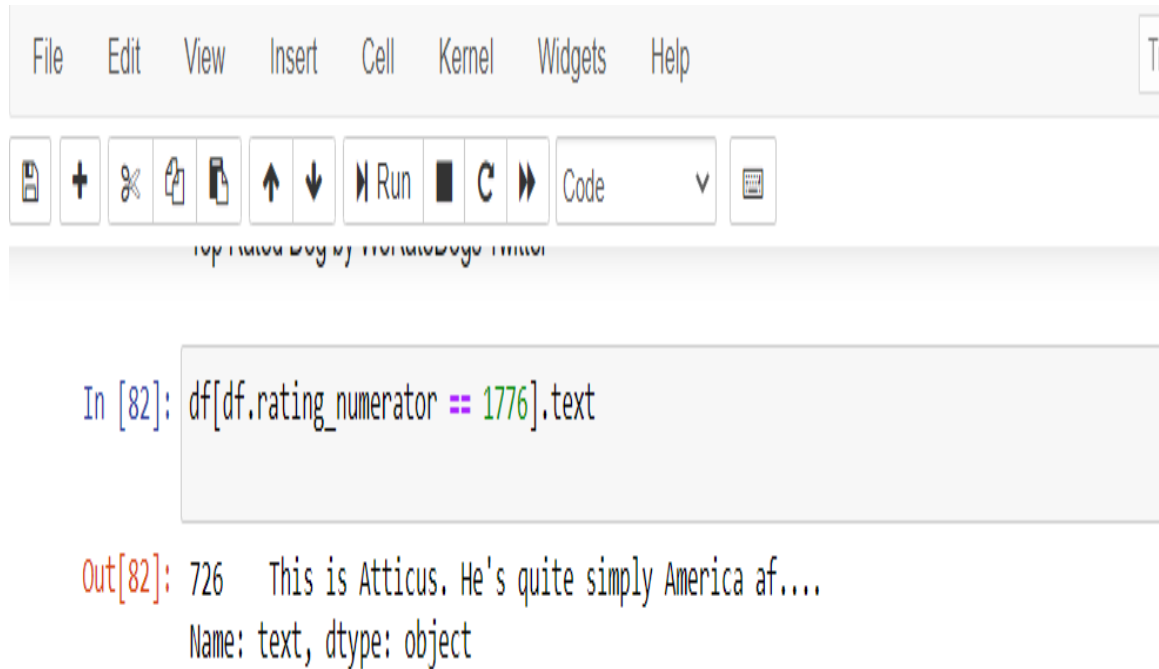
```
Doggo
Mean Retweet Count      7126
Mean Favorite Count     19356
Floofer
Mean Retweet Count      4969
Mean Favorite Count     13206
Pupper
Mean Retweet Count      2364
Mean Favorite Count     7198
Puppo
Mean Retweet Count      6474
Mean Favorite Count     21582
```

From the above table , it is evident that tweets of dogs that have increase in relweeting and favoriting of tweets having a mention of the * dog stage e Doggo , Floofer or Puppo) , the

exception being Pupper Pupper being a small doggo usually younger is probably less popular scompared to its older counterparts Having said that I can see a strong correlation between tweets mentioning the dog stage and its chances of being retweeted and favorited more

File   Edit   View   Insert   Cell   Kernel   Widgets   Help
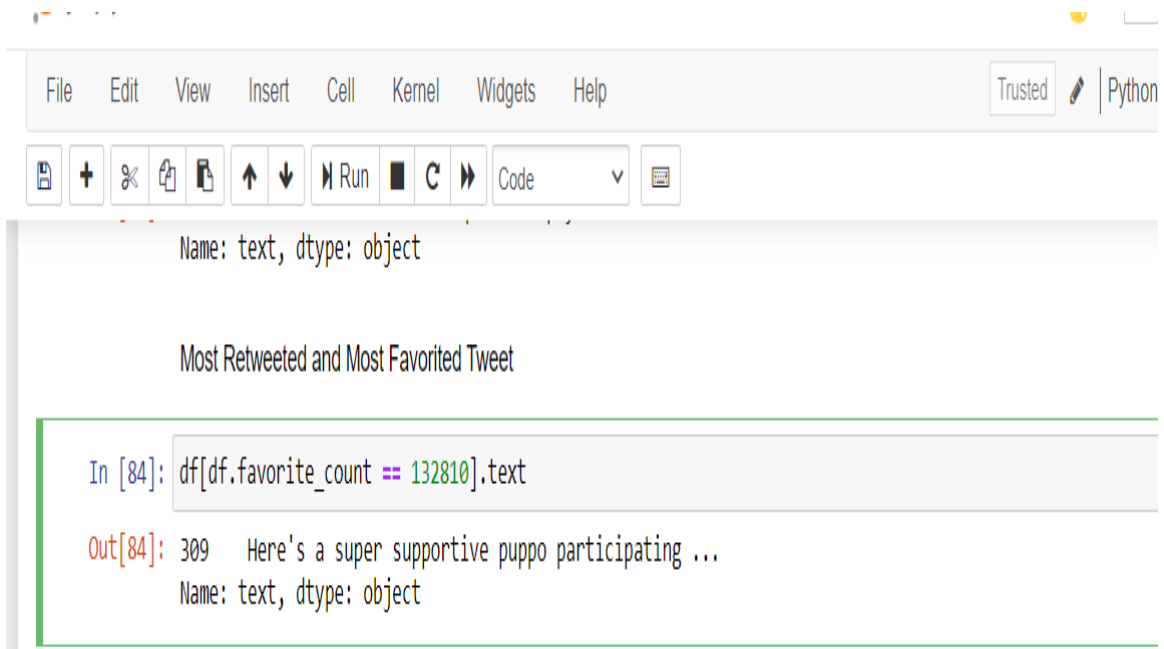
```
In [82]: df[df.rating_numerator == 1776].text
```

```
Out[82]: 726    This is Atticus. He's quite simply America af....
         Name: text, dtype: object
```

The top rated dog by WeRateDogs Twitter is a dog wrapped in an American flag. The rating 1776 is a reference to Events from the year 1776 in the United States. This year is celebrated in the United States as the official beginning of its nationhood, with the Declaration of Independence issued on July 4.

Most Retweeted and Most Favorited Tweet

Name: text, dtype: object

Most Retweeted and Most Favorited Tweet

```
In [84]: df[df.favorite_count == 132810].text
```

```
Out[84]: 309    Here's a super supportive puppo participating ...
         Name: text, dtype: object
```

The most retweeted and most favorited tweet is a video of a puppo