Analysis and Visualizations

Notebook:	Udacity		
Created:	28-Sep-19 11:12 AM	**Updated:**	28-Sep-19 11:21 AM
Author:	prashant kumar		
URL:	https://www.google.com/search?q=analyse+and+visualizations+data+analys&rlz=1C1...		

Wrangle and Analyze Data
Udacity Nano Degree Program

Data visualization for this project describes all my efforts to help people understand the significance of WeRateDogs twitter handleHere,I analyse different aspects of the data available and place all the relevant information into a visual context.

Different kinds of charts and plots such bar grapg,scatter plot etc. has been used to make the visualizion and answer certain important questions:-

For example:-

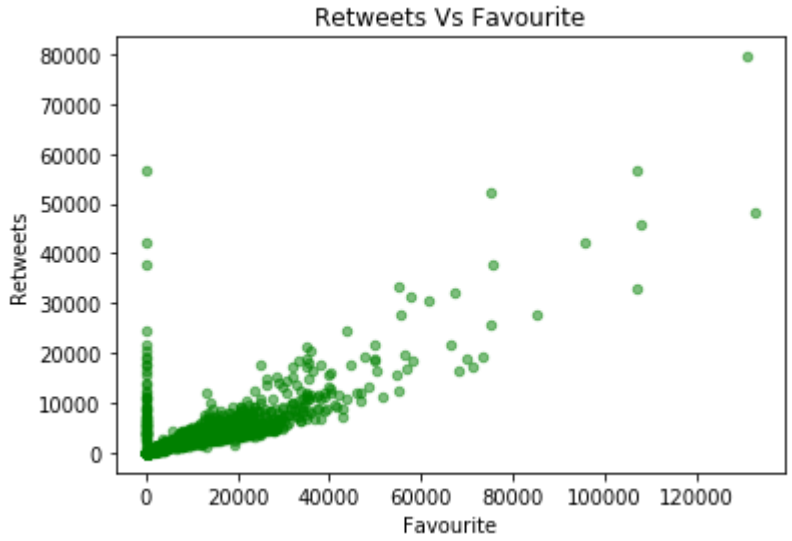
What is the relation between Retweets and likes(favourite count)?

In [256]:
```
df=pd.read_csv('twitter_master.csv')
df.head()
```

Out[256]:

	tweet_id	timestamp	source	text	
0	892420643555336193	2017-08-01 16:23:56	<a href="http://twitter.com/download/iphone" r...	This is Phineas. He's a mystical boy. Only eve...	https://twitte
1	892177421306343426	2017-08-01 00:17:27	<a href="http://twitter.com/download/iphone" r...	This is Tilly. She's just checking pup on you....	https://twitte
2	891815181378084864	2017-07-31 00:18:03	<a href="http://twitter.com/download/iphone" r...	This is Archie. He is a rare Norwegian Pouncin...	https://twitte
3	891689557279858688	2017-07-30 15:58:51	<a href="http://twitter.com/download/iphone" r...	This is Darla. She commenced a snooze mid meal...	https://twitte
4	891327558926688256	2017-07-29 16:00:24	<a href="http://twitter.com/download/iphone" r...	This is Franklin. He would like you to stop ca...	https://twitte

In [413]:
```
#Retweets vs. Favourites
df.plot(kind='scatter',x='favorite_count',y='retweet_count', alpha = 0.5, colo
r = 'green');
plt.xlabel('Favourite');
plt.ylabel('Retweets');
plt.title('Retweets Vs Favourite');
```



The scatter plot above shows that the Retweet_count and Favourite count are directly proportional.

Most common names for the dogs?

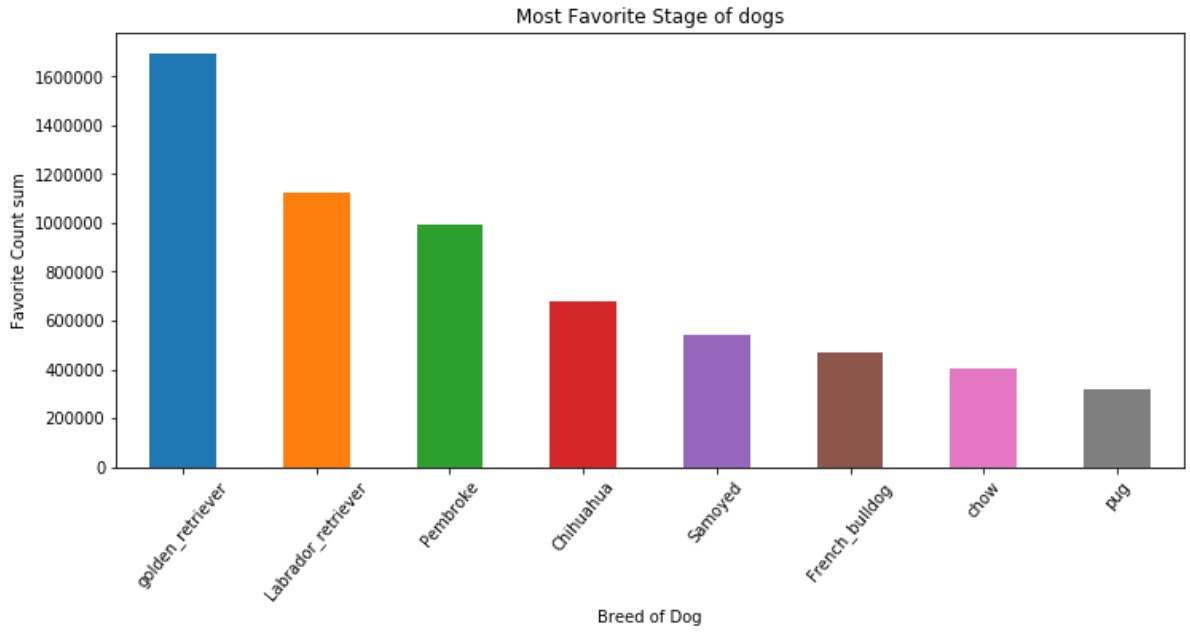
In [265]:
```
df.name.value_counts().head(50)
```

Out[265]:
```
None        677
Charlie      11
Cooper       10
Penny        10
Lucy         10
Tucker       10
Oliver       10
Sadie         8
Bo            8
Lola          8
Winston       8
Daisy         7
Toby          7
Dave          6
Bella         6
Rusty         6
Bailey        6
Milo          6
Koda          6
Jax           6
Scout         6
Stanley       6
Alfie         5
Chester       5
Buddy         5
Louis         5
Oscar         5
Leo           5
Larry         5
Clarence      4
Scooter       4
Sophie        4
Finn          4
Oakley        4
Bentley       4
Ruby          4
Brody         4
Jerry         4
Bruce         4
Walter        4
Archie        4
George        4
Maggie        4
Chip          4
Dexter        4
Loki          4
Winnie        4
Reggie        4
Sunny         4
Phil          4
Name: name, dtype: int64
```

Most common name for the dogs include Charlie,cooper,panie etc

Most Favourite breeds?

In [409]:
```
fig= plt.figure(figsize=(12,5))
df.groupby('p1')['favorite_count'].sum().sort_values(ascending=False).head(8).
plot(kind='bar');
plt.xlabel('Breed of Dog')
plt.xticks(rotation=50)
plt.ylabel('Favorite Count sum')
plt.title('Most Favorite Breed of dogs');
```



Golden Retriever is found to be the most favourite breed.Then next come the Labrador Retriever

Most rated dog?

In [343]:
```
dog_rated = df['rating_numerator'].sort_values(ascending=False).head(1)
print dog_rated
df.query('rating_numerator=="1776"')
```

```
802    1776
Name: rating_numerator, dtype: int64
```

Out[343]:

	tweet_id	timestamp	source	text	
802	749981277374128128	2016-07-04 15:00:45	<a href="https://about.twitter.com/products/tw...	This is Atticus. He's quite simply America af....	https://twitt

The most rated dog is Atticus with rating 1776/10

Image of the most rated dog

In [390]:
```
most_rated=df.query('rating_numerator=="1776"')['jpg_url']
most_rated=list(most_rated)
Image(url= t[0], width=200, height=200)
```

Out[390]:

Most favourite dog

In [348]:
```
dog_favorite = df['favorite_count'].sort_values(ascending=False).head(1)
print dog_favorite
df.query('favorite_count=="132810"')
```

```
329    132810
Name: favorite_count, dtype: int64
```

Out[348]:

	tweet_id	timestamp	source	text	
329	822872901745569793	2017-01-21 18:26:02	<a href="http://twitter.com/download/iphone" r...	Here's a super supportive puppo participating ...	https://twit

**Most favourite dog has 132810 favourite count.

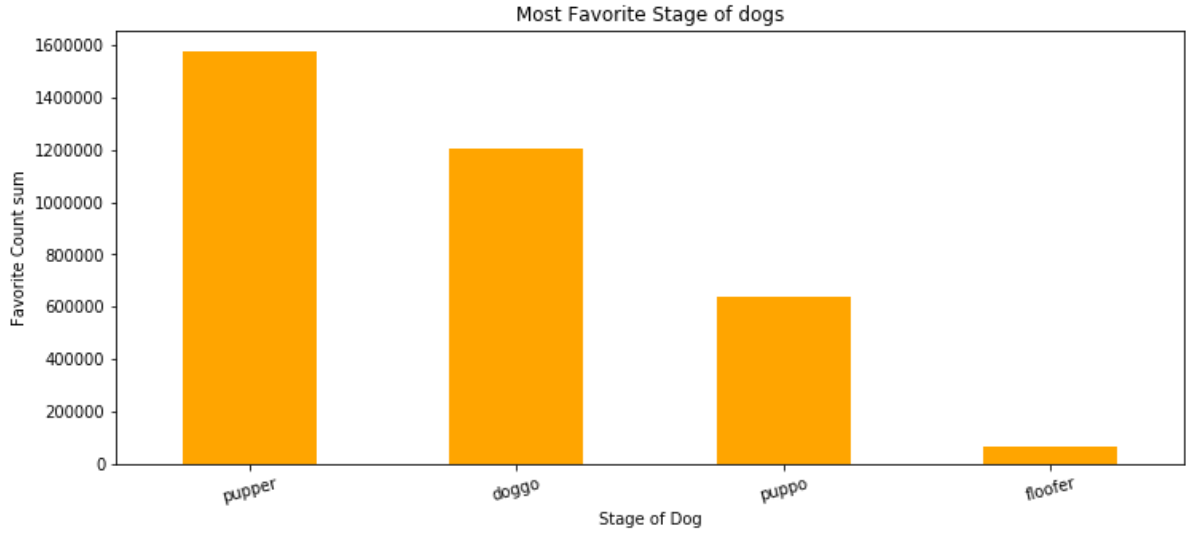
Image of the most favourite dog

In [354]:
```
most_favourite=df.query('favorite_count=="132810"')['jpg_url']
most_favourite=list(most_favourite)
Image(url= t[0], width=170, height=150)
```

Out[354]:

Most favourite Stage of dogs

In [408]:
```
fig= plt.figure(figsize=(12,5))
labels = ['Pupper', 'Doggo', 'Puppo', 'Floofer']
df.groupby(['dogs_types']).favorite_count.sum().sort_values(ascending=False).p
lot(kind='bar',color='orange')
plt.xlabel('Stage of Dog')
ax.set_xticklabels(labels)
plt.xticks(rotation=15)
plt.ylabel('Favorite Count sum')
plt.title('Most Favorite Stage of dogs');
```



Pupper is the most favourite stage of dog.Then next comes the doggo

Favorite Counts have increased over the years ?

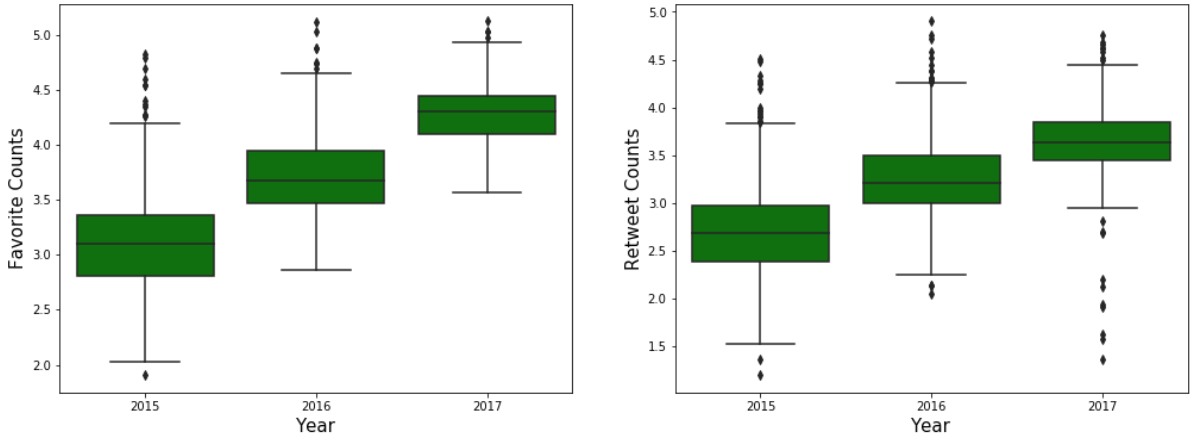
In [377]:
```
df['date'] = pd.DatetimeIndex(df.timestamp).normalize()
df['year'] = df['date'].dt.year
df.groupby('year').count()[['favorite_count','retweet_count']]
```

Out[377]:

year	favorite_count	retweet_count
2015	665	665
2016	1022	1022
2017	386	386

In [393]:
```
fig,(ax1,ax2) = plt.subplots(1,2,figsize = (17,6))
sns.boxplot(x = df['year'],y = np.log10(df['favorite_count']),ax = ax1,color=
'green');
sns.boxplot(x = df['year'], y = np.log10(df['retweet_count']),ax = ax2,color=
'green');
ax1.set_xlabel('Year',fontsize = 15);
ax1.set_ylabel('Favorite Counts',fontsize = 15);
ax2.set_xlabel('Year',fontsize = 15);
ax2.set_ylabel('Retweet Counts',fontsize = 15);
```

```
C:\Users\ptkr\AppData\Local\Continuum\anaconda_2\lib\site-packages\ipykernel_
launcher.py:2: RuntimeWarning: divide by zero encountered in log10
```



From the box plot above we can see that the favourite counts have increased over the years.