act_report

January 30, 2019

1 Analyzing, and Visualizing Data

```
In [30]: import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         %matplotlib inline
         import seaborn as sns

In [31]: df=pd.read_csv('twitter_archive_master.csv')

In [32]: df.head()
```

Out[32]:

	Unnamed: 0	tweet_id	in_reply_to_status_id	in_reply_to_user_id	timestamp	source	text	expanded_urls	rating_numerator
0	0	892420643555336193	NaN	NaN	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://twitter.com/dog_rates/status/892420643...	13.0
1	1	892177421306343426	NaN	NaN	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://twitter.com/dog_rates/status/892177421...	13.0
2	2	891815181378084864	NaN	NaN	2017-07-31 00:18:03	<a href="http://twitter.com/download/iphone" r...	This is Archie. He is a rare Norwegian Pouncin...	https://twitter.com/dog_rates/status/891815181...	12.0
3	3	891689557279858688	NaN	NaN	2017-07-30 15:58:51	<a href="http://twitter.com/download/iphone" r...	This is Darla. She commenced a snooze mid meal...		
4	4	891327558926688256	NaN	NaN	2017-07-29 16:00:24	<a href="http://twitter.com/download/iphone" r...	This is Franklin. He would like you to stop ca...		

```
3  https://twitter.com/dog_rates/status/891689557...              13.0
4  https://twitter.com/dog_rates/status/891327558...              12.0

   rating_denominator      name  favorite_count  retweet_count  \
0                10.0   Phineas           37986           8302
1                10.0     Tilly           32617           6132
2                10.0    Archie           24554           4058
3                10.0     Darla           41327           8442
4                10.0  Franklin           39507           9147

                                          jpg_url  img_num dog_type  \
0  https://pbs.twimg.com/media/DGKD1-bXoAAIAUK.jpg        1      NaN
1  https://pbs.twimg.com/media/DGGmoV4XsAAUL6n.jpg        1      NaN
2  https://pbs.twimg.com/media/DGBdLU1WsAANxJ9.jpg        1      NaN
3  https://pbs.twimg.com/media/DF_q7IAWsAEuuN8.jpg        1      NaN
4  https://pbs.twimg.com/media/DF6hr6BUMAAzZgT.jpg        2      NaN

                                          txt_splt Extracted_name
0  ['This', 'is', 'Phineas.', "He's", 'a', 'mysti...        Phineas
1  ['This', 'is', 'Tilly.', "She's", 'just', 'che...          Tilly
2  ['This', 'is', 'Archie.', 'He', 'is', 'a', 'ra...         Archie
3  ['This', 'is', 'Darla.', 'She', 'commenced', '...          Darla
4  ['This', 'is', 'Franklin.', 'He', 'would', 'li...       Franklin
```

In [34]:
```
df['timestamp'] = pd.to_datetime(df['timestamp'])
```

In [35]:
```
df.dtypes
```

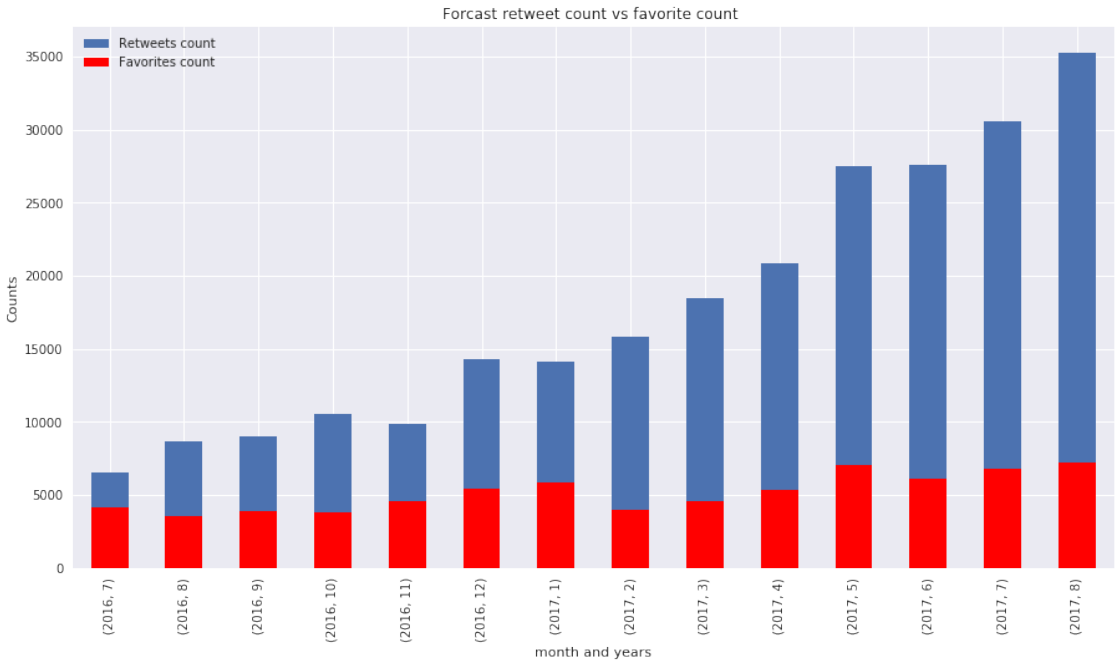
Out[35]:
```
Unnamed: 0                       int64
tweet_id                         int64
in_reply_to_status_id          float64
in_reply_to_user_id            float64
timestamp                datetime64[ns]
source                           object
text                             object
expanded_urls                    object
rating_numerator               float64
rating_denominator             float64
name                             object
favorite_count                   int64
retweet_count                    int64
jpg_url                          object
img_num                          int64
dog_type                         object
txt_splt                         object
Extracted_name                   object
dtype: object
```

In [27]:
```
sns.set(color_codes=True)
sns.lmplot(x="retweet_count",y="favorite_count", data=df, size = 9, line_kws={'color':
```

```
plt.title(' Retweet Count vs. Favorite count ')
plt.xlabel('Retweet Count')
plt.ylabel('Favorite Count');
```

In [71]:
```
fig = plt.figure(figsize=(15,8))
df['favorite_count'].groupby([df["timestamp"].dt.year, df["timestamp"].dt.month]).mean(
df['retweet_count'].groupby([df["timestamp"].dt.year, df["timestamp"].dt.month]).mean()
plt.legend(('Retweets count', 'Favorites count'))
plt.title('Forcast retweet count vs favorite count ')
plt.ylabel('Counts')
plt.xlabel('month and years')
```

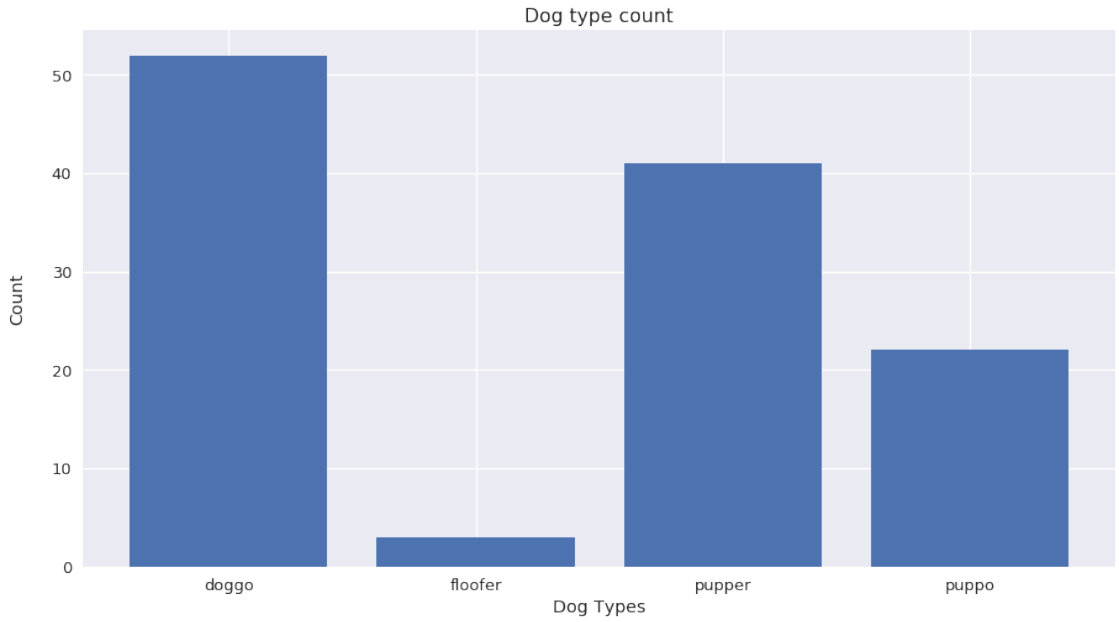
Out[71]:
```
Text(0.5,0,'month and years')
```




```
In [77]: df['favorite_count'].describe()

Out[77]: count       692.000000
         mean      16284.377168
         std       15155.876710
         min           0.000000
         25%        7806.750000
         50%       12425.000000
         75%       21483.000000
         max      140174.000000
         Name: favorite_count, dtype: float64

In [78]: df['retweet_count'].describe()

Out[78]: count      692.000000
         mean      4980.871387
         std       5587.125889
         min         23.000000
         25%       2189.750000
         50%       3418.000000
         75%       5490.500000
         max      60866.000000
         Name: retweet_count, dtype: float64

In [102]: #top 5 popular Name
          df['Extracted_name'].value_counts().head(5)
```

```
Out[102]: Charlie    6
          Tucker     5
          Penny      5
          Daisy      4
          Bo         4
          Name: Extracted_name, dtype: int64
```

```
In [100]:  plt.subplots(figsize=(15, 8))
           plt.bar([1, 2, 3, 4], df.groupby('dog_type')['tweet_id'].count(), tick_label=['doggo',
           plt.title('Dog type count')
           plt.xlabel('Dog Types')
           plt.ylabel('Count\n');
```




```
In [95]: df.groupby('dog_type')['tweet_id'].count()
```

```
Out[95]: dog_type
         doggo      52
         floofer     3
         pupper     41
         puppo      22
         Name: tweet_id, dtype: int64
```

```
In [80]: df.dtypes
```

```
Out[80]: Unnamed: 0                    int64
         tweet_id                      int64
         in_reply_to_status_id       float64
```

```
in_reply_to_user_id            float64
timestamp                datetime64[ns]
source                          object
text                            object
expanded_urls                   object
rating_numerator               float64
rating_denominator             float64
name                            object
favorite_count                   int64
retweet_count                    int64
jpg_url                         object
img_num                          int64
dog_type                        object
txt_splt                        object
Extracted_name                  object
dtype: object
```

```
In [123]: pd.set_option('max_colwidth', 800)
          index=df['favorite_count'].nlargest(3).index
          heigh_faviort_dog = df[['tweet_id','name','timestamp','favorite_count','dog_type','jpg
          heigh_faviort_dog
```

Out[123]:

	tweet_id	name	timestamp	favorite_count	dog_type	
322	822872901745569793	NaN	2017-01-21 18:26:02	140174	puppo	
419	807106840509214720	Stephan	2016-12-09 06:17:20	126816	NaN	https:/
111	866450705531457537	Jamesy	2017-05-22 00:28:40	121772	pupper	

Stephan
Jamesy