act_report

November 11, 2019

** Data Vitualization and Insights**

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

In [3]:
```
df = pd.read_csv('twitter_archive_master.csv')
```

In [4]:
```
df.head(5)
```

Out[4]:

	tweet_id	jpg_url	timestamp	text	expanded_urls	rating_numerator	rating_denominator	name	favorites	retweets	user_followers	dog_stage
0	872820683541237760	https://pbs.twimg.com/media/DBzhxOPWAAEhl0E.jpg	2017-06-08 14:20:41	Here are my favorite #dogsatpollingstations \n...	https://twitter.com/dog_rates/status/872820683...	13.0	10.0	None	14142	3490	8490262	None
1	872620804844003328	https://pbs.twimg.com/media/DBwr_hzXkAEnZBW.jpg	2017-06-08 01:06:27	This is Monkey. She's supporting owners everyw...	https://twitter.com/dog_rates/status/872620804...	13.0	10.0	Monkey	19939	3443	8490262	None
2	872486979161796608	https://pbs.twimg.com/media/DBuyRlTUwAAYhG9.jpg	2017-06-07 16:14:40	We. Only. Rate. Dogs. Do not send in other thi...	https://twitter.com/dog_rates/status/872486979...	12.0	10.0	None	39138	8497	8490262	None
3	872122724285648897	https://pbs.twimg.com/media/DBpm-5UXcAUeCru.jpg	2017-06-06 16:07:15	This is Kody. He's a baller. Wishes he was a l...	https://twitter.com/dog_rates/status/872122724...	12.0	10.0	Kody	33067	7647	8490262	None
4	871879754684805121	https://pbs.twimg.com/media/DBmKAmBXUAE-pQ-.jpg	2017-06-06 00:01:46	Say hello to Lassie. She's celebrating #PrideM...	https://twitter.com/dog_rates/status/871879754...	13.0	10.0	Lassie	36316	10422	8490262	None

```
     prediction_algo  conf_level dog_gender
0                pug    0.999120        NaN
1     cocker_spaniel    0.513191     female
2           Pembroke    0.931861        NaN
3                pug    0.066736       male
4  Shetland_sheepdog    0.969171     female
```

In [6]:
```
# The most favorites
df.favorites.max(), df.favorites.min()
```

Out[6]:
```
(159705, 73)
```

In [7]:
```
# The most favorite is Labrador_retriever in the Dog stage.
df[df['favorites'] ==159705]
```

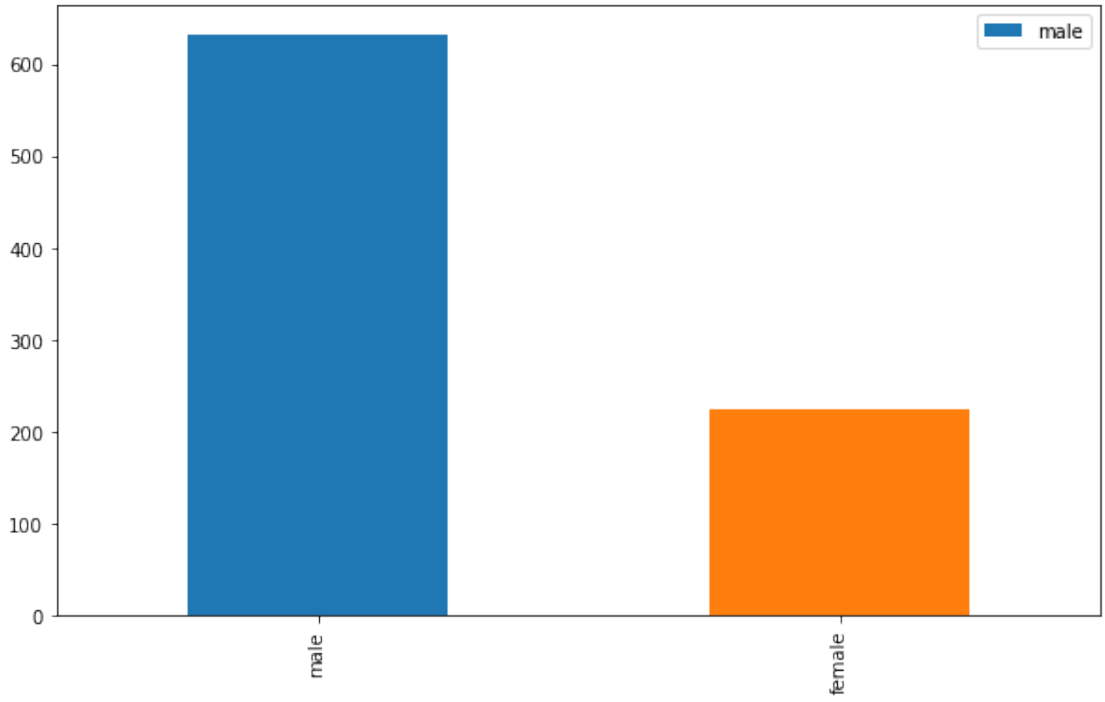
Out[7]:
```
                tweet_id                                           jpg_url  \
1691  744234799360020481  https://pbs.twimg.com/ext_tw_video_thumb/74423...

                timestamp                                               text  \
1691  2016-06-18 18:26:18  Here's a doggo realizing you can stand in a po...

                                        expanded_urls  rating_numerator  \
1691  https://twitter.com/dog_rates/status/744234799...              13.0

      rating_denominator  name  favorites  retweets  user_followers dog_stage  \
1691                10.0  None     159705     79705         8490278     doggo

         prediction_algo  conf_level dog_gender
1691  Labrador_retriever    0.825333        NaN
```

In [8]:
```
# Less favorite is English_setter in the Dog stage.
df[df['favorites'] ==73]
```

Out[8]:
```
                tweet_id                                           jpg_url  \
130  666102155909144576  https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg

                timestamp                                               text  \
130  2015-11-16 03:55:04  Oh my. Here you are seeing an Adobe Setter giv...

                                        expanded_urls  rating_numerator  \
130  https://twitter.com/dog_rates/status/666102155...              11.0

     rating_denominator  name  favorites  retweets  user_followers dog_stage  \
130                10.0  None         73        11         8490299      None

    prediction_algo  conf_level dog_gender
130  English_setter    0.298617        NaN
```

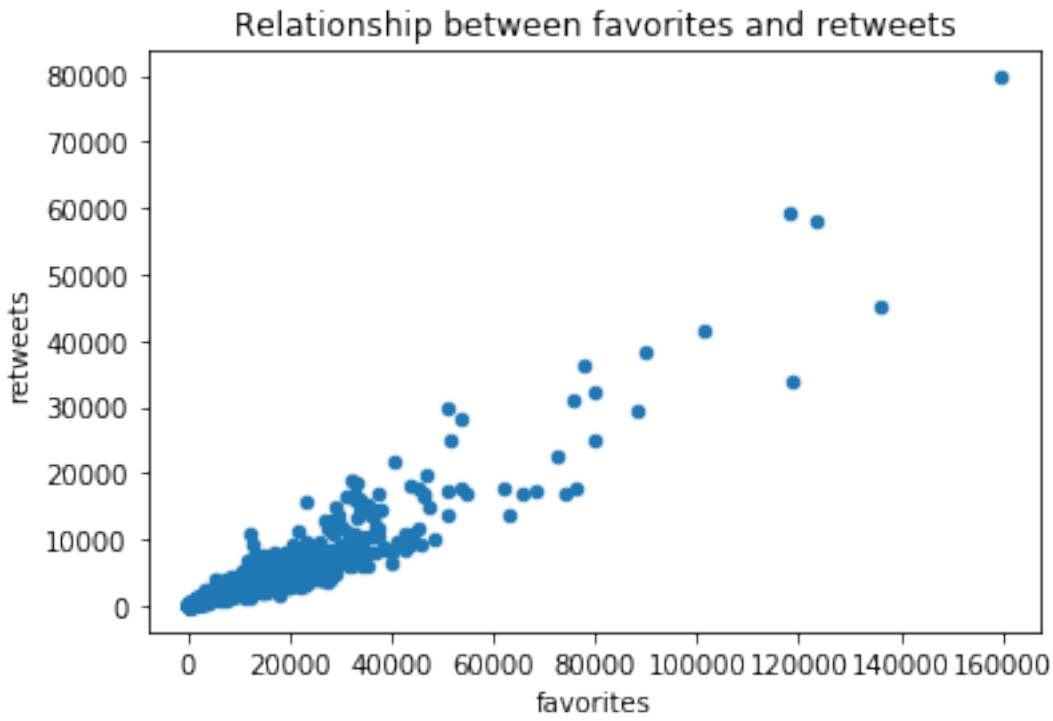
In [9]:
```
# Most favorite Dog gender
from matplotlib.pyplot import *
```

```
fig, ax = subplots()
df['dog_gender'].value_counts().plot(kind = 'bar', figsize = (10, 6), ax=ax)
ax.legend(['male', 'female']);
```



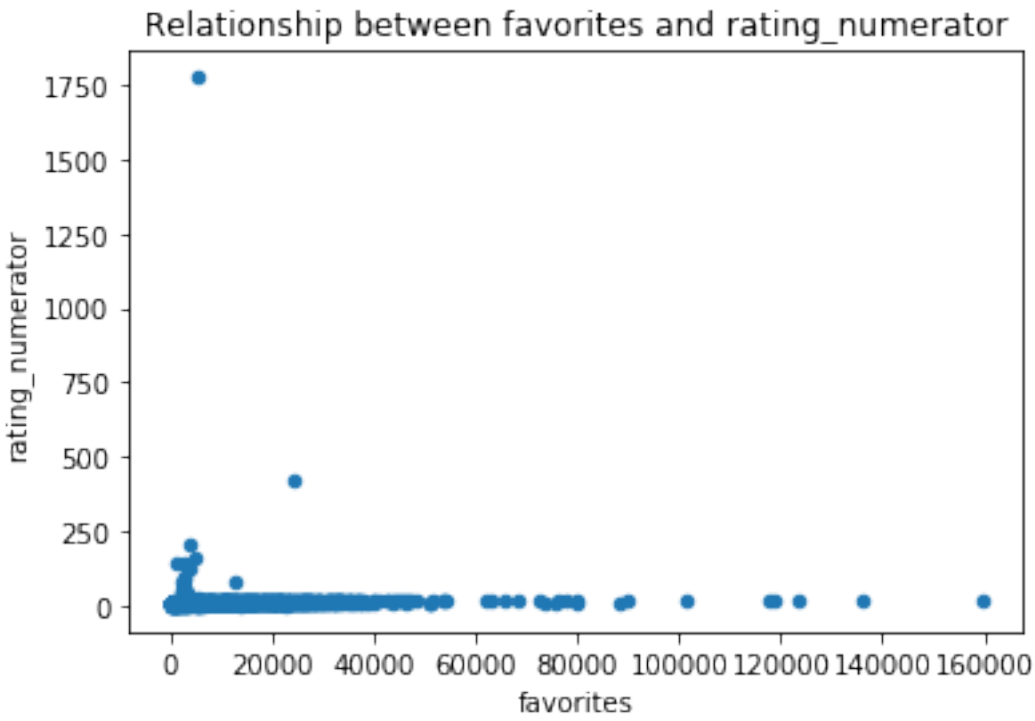
The male dogs are the most common dog gender

```
In [11]: # Relationship between favorites and retweets
         df.plot.scatter('favorites', 'retweets')
         plt.title('Relationship between favorites and retweets');
```



There's a close relationship between favorites and retweets. Labrador_retriever is the most favorites and also has the most retweets while English_setter is the least favorites and also has the least retweets.

```
In [12]: # Relationship between favorites and ratings
         df.plot.scatter('favorites', 'rating_numerator')
         plt.title('Relationship between favorites and rating_numerator');
```




```
In [14]: # The most retweets
         df.retweets.max(), df.retweets.min()
```

```
Out[14]: (79705, 11)
```

```
In [15]: # The most retweet is Labrador_retriever in the Dog stage.
         df[df['retweets'] ==79705]
```

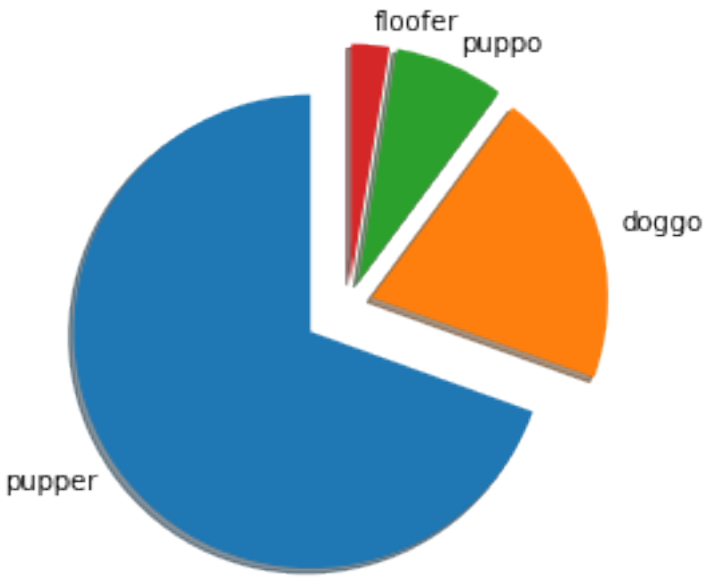
```
Out[15]:                tweet_id                                          jpg_url  \
         1691  744234799360020481  https://pbs.twimg.com/ext_tw_video_thumb/74423...

                        timestamp                                               text  \
         1691  2016-06-18 18:26:18  Here's a doggo realizing you can stand in a po...

                                              expanded_urls  rating_numerator  \
         1691  https://twitter.com/dog_rates/status/744234799...              13.0

               rating_denominator  name  favorites  retweets  user_followers dog_stage  \
         1691                10.0  None     159705     79705         8490278     doggo

                  prediction_algo  conf_level dog_gender
         1691  Labrador_retriever    0.825333        NaN
```

```
In [16]: # The least retweets is English_setter in the Dog stage.
         df[df['retweets'] ==11]
```

Out[16]:
```
               tweet_id                                          jpg_url  \
130  666102155909144576  https://pbs.twimg.com/media/CT54YGiWUAEZnoK.jpg

               timestamp                                               text  \
130  2015-11-16 03:55:04  Oh my. Here you are seeing an Adobe Setter giv...

                                        expanded_urls  rating_numerator  \
130  https://twitter.com/dog_rates/status/666102155...              11.0

     rating_denominator  name  favorites  retweets  user_followers dog_stage  \
130                10.0  None         73        11         8490299      None

    prediction_algo  conf_level dog_gender
130  English_setter    0.298617        NaN
```

In [23]:
```
# Plot the data partitioned by dog stage
dog_stage_none = list(df[df['dog_stage'] != 'None']['dog_stage'].value_counts())[0:4]
dog_stages = df[df['dog_stage'] != 'None']['dog_stage'].value_counts().index.tolist()[0
explode = (0.2, 0.1, 0.1, 0.1)

fig1, ax1 = plt.subplots()
ax1.pie(dog_stage_none, explode = explode, labels = dog_stages, shadow = True, startang
ax1.axis('equal');

df[df['dog_stage'] != 'None'].groupby('dog_stage')['rating_numerator'].mean()
```

Out[23]:
```
dog_stage
doggo      11.870968
floofer    11.875000
pupper     10.652217
puppo      12.043478
Name: rating_numerator, dtype: float64
```



Most people has pupper. This means most people likes, favorites and retweet pupper dogs. Pupper dogs are the most adorable.

Conclusion

Most people who rated dogs on we rate dog twitter handles favorites and retweets the males dogs. However, from our analysis, the most favorites and retweets dog in the dog stage is doggo and does not neccessarily mean its the best dog_stage ever. Further analysis shows pupper dog_stage has more collective ratings than doggo.

In [ ]: