act_report

July 4, 2021

1 Wrangle and Analyze Data

2 1. Import libraries

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import json
import datetime
import requests
import tweepy
from tweepy import OAuthHandler
from timeit import default_timer as timer
from subprocess import call
from tqdm import trange,tqdm
```

2.1 5.3 Visualizing Data

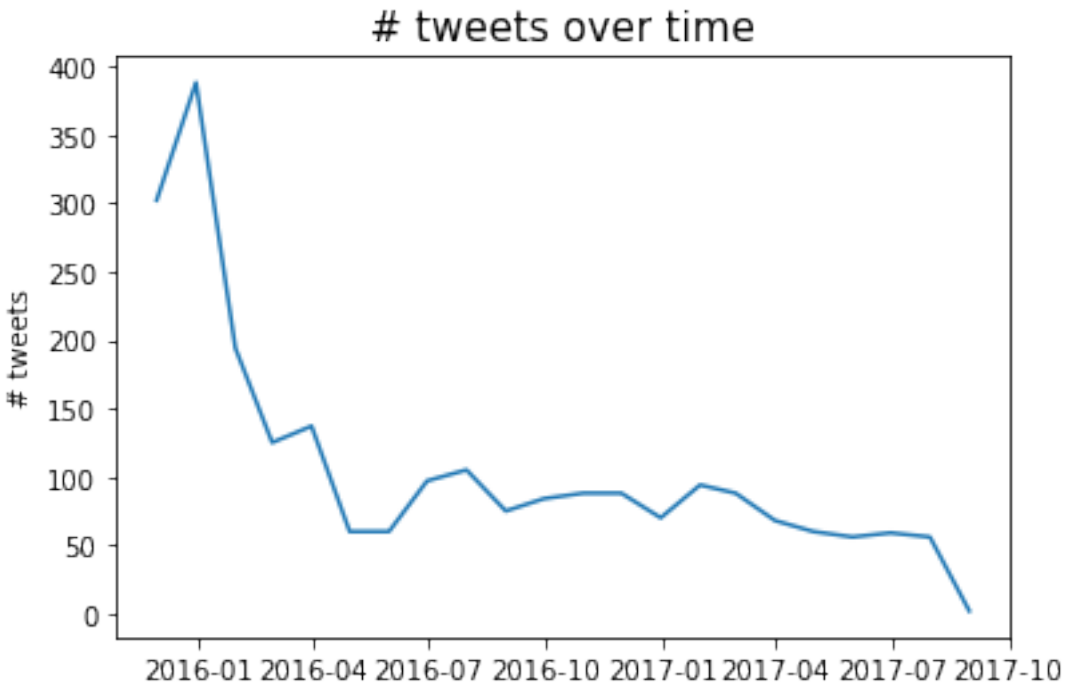
[93]:
```
df_merge.head(2)
```

[93]:
```
             tweet_id                  timestamp  \
0  892420643555336193 2017-08-01 16:23:56+00:00
1  892177421306343426 2017-08-01 00:17:27+00:00

                                              source  \
0  <a href="http://twitter.com/download/iphone" r…
1  <a href="http://twitter.com/download/iphone" r…

                                                text  rating_numerator  \
0  This is Phineas. He's a mystical boy. Only eve…                 13
1  This is Tilly. She's just checking pup on you…                 13

   rating_denominator     name  \
0                  10  Phineas
1                  10    Tilly
```

```
                                          jpg_url  img_num  \
0  https://pbs.twimg.com/media/CT4udn0WwAA0aMy.jpg      1.0
1  https://pbs.twimg.com/media/CT42GRgUYAA5iDo.jpg      1.0

                     p1  …   p2_conf p2_dog                   p3  \
0  Welsh_springer_spaniel  …  0.156665   True    Shetland_sheepdog
1                redbone  …  0.074192   True  Rhodesian_ridgeback

    p3_conf p3_dog retweet_count  favorite_count followers_count  \
0  0.061428   True        7237.0         34675.0       9001806.0
1  0.072010   True        5420.0         30046.0       9001806.0

   retweeted_status                      url
0    Original tweet  https://t.co/MgUWQ76dJU
1    Original tweet  https://t.co/aQFSeaCu9L

[2 rows x 23 columns]
```

2.1.1 5.3.1 tweets over time

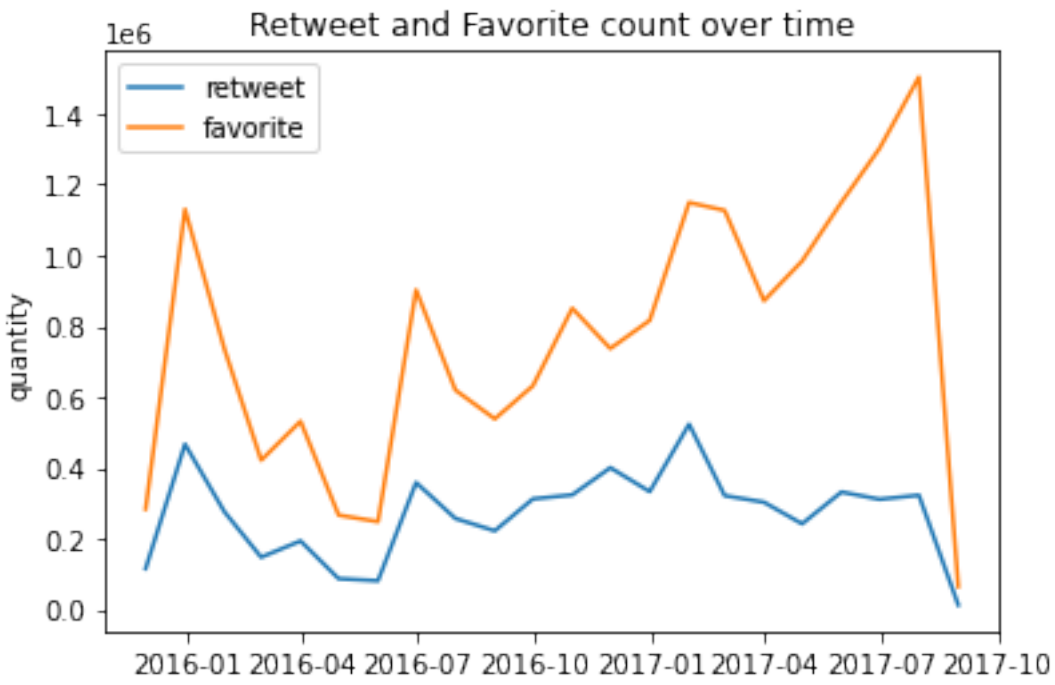
[113]:
```
data_group = df_merge.groupby(pd.Grouper(key = 'timestamp', freq = "M")).
 ↪count().reset_index()
data_group = data_group[['timestamp', 'tweet_id']]
plt.plot(data_group['timestamp'], data_group['tweet_id'])
plt.title('# tweets over time')
plt.ylabel('# tweets')
plt.savefig('tweets_over_time')
plt.show()
```



- **Insight**: Through time, the number of tweets has decreased, in 2016 there were almost 400 tweets per month, however in 2017 this quantity decreased near to zero.

2.1.2 5.3.2 retweet and favorite count over time

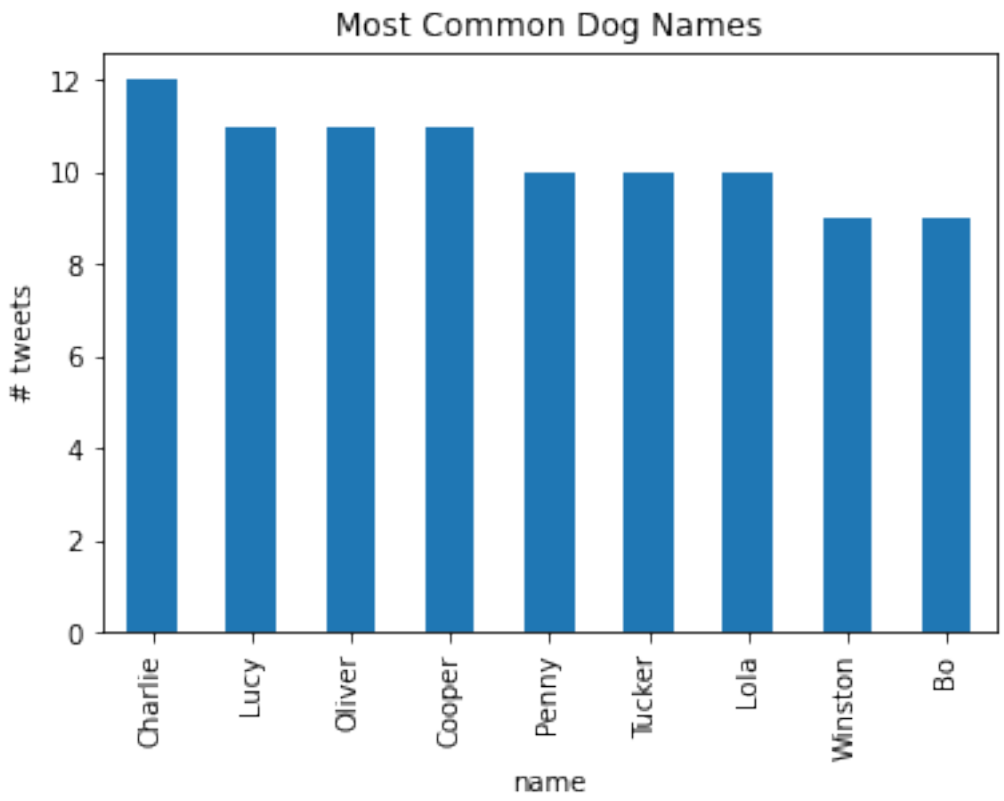
[131]:
```
data_group = df_merge.groupby(pd.Grouper(key = 'timestamp', freq = "M")).
 ↪agg({'retweet_count':'sum', 'favorite_count':'sum'}).reset_index()
plt.plot(data_group['timestamp'], data_group['retweet_count'])
plt.plot(data_group['timestamp'], data_group['favorite_count'])
plt.title('Retweet and Favorite count over time')
plt.ylabel('quantity')
plt.legend(['retweet', 'favorite'], loc='upper left')
plt.savefig('retweet_favorite_count_over_time')
plt.show()
```



- **Insight**: Through time, the number of retweets and favorites have followed a similar trend, the only difference is that the number of favorites has been greater than retweets.

2.1.3 5.3.3 Most common dog names

[147]:
```
data_group = df_merge.groupby('name').agg({'tweet_id':'count'}).reset_index().
↪sort_values(by='tweet_id', ascending=False)
data_group[0:9].plot(kind='bar',x='name', legend=False)
plt.title('Most Common Dog Names')
plt.ylabel('# tweets')
plt.savefig('dog_names_over_time')
plt.show()
```



- **Insight**: Through time, Charlie, Lucy and Oliver are the three most common names.

[ ]: