

In [158]:

```
!pip install tweepy
```

```
Requirement already satisfied: tweepy in /Users/stevenhansen/opt/anaconda3/lib/python3.7/site-packages (3.9.0)
Requirement already satisfied: requests-oauthlib>=0.7.0 in /Users/stevenhansen/opt/anaconda3/lib/python3.7/site-packages (from tweepy) (1.3.0)
Requirement already satisfied: requests[socks]>=2.11.1 in /Users/stevenhansen/opt/anaconda3/lib/python3.7/site-packages (from tweepy) (2.22.0)
Requirement already satisfied: six>=1.10.0 in /Users/stevenhansen/opt/anaconda3/lib/python3.7/site-packages (from tweepy) (1.12.0)
Requirement already satisfied: oauthlib>=3.0.0 in /Users/stevenhansen/opt/anaconda3/lib/python3.7/site-packages (from requests-oauthlib>=0.7.0->tweepy) (3.1.0)
Requirement already satisfied: urllib3!=1.25.0,!<1.25.1,<1.26,>=1.21.1 in /Users/stevenhansen/opt/anaconda3/lib/python3.7/site-packages (from requests[socks]>=2.11.1->tweepy) (1.24.2)
Requirement already satisfied: certifi>=2017.4.17 in /Users/stevenhansen/opt/anaconda3/lib/python3.7/site-packages (from requests[socks]>=2.11.1->tweepy) (2019.9.11)
Requirement already satisfied: idna<2.9,>=2.5 in /Users/stevenhansen/opt/anaconda3/lib/python3.7/site-packages (from requests[socks]>=2.11.1->tweepy) (2.8)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /Users/stevenhansen/opt/anaconda3/lib/python3.7/site-packages (from requests[socks]>=2.11.1->tweepy) (3.0.4)
Requirement already satisfied: PySocks!=1.5.7,>=1.5.6; extra == "socks" in /Users/stevenhansen/opt/anaconda3/lib/python3.7/site-packages (from requests[socks]>=2.11.1->tweepy) (1.7.1)
```

In [193]:

```
import pickle
import os
import tweepy
from tweepy.streaming import StreamListener
from tweepy import OAuthHandler
import csv
import pandas as pd
```

In [210]:

```
#https://developer.twitter.com/en/apps/18356662
if not os.path.exists('./secret_twitter_credentials3.pkl'):
    Twitter={}
    Twitter['Consumer Key'] = ''
    Twitter['Consumer Secret'] = ''
    Twitter['Access Token'] = ''
    Twitter['Access Token Secret'] = ''
    with open('secret_twitter_credentials3.pkl','wb') as f:
        pickle.dump(Twitter, f)
else:
    Twitter=pickle.load(open('secret_twitter_credentials2.pkl','rb'))
```

In [211]:

```
auth = tweepy.OAuthHandler(Twitter['Consumer Key'], Twitter['Consumer Secret'] )
auth.set_access_token(Twitter['Access Token'], Twitter['Access Token Secret'])
api = tweepy.API(auth)
```

In [212]:

```
api
```

Out[212]:

```
<tweepy.api.API at 0x11bdeef10>
```

In [196]:

```
# https://towardsdatascience.com/twitter-data-collection-tutorial-using-python-3267d7cfa93e
def get_all_tweets(screen_name,path):
    alltweets = []
    new_tweets = api.user_timeline(screen_name = screen_name, count=200)
    alltweets.extend(new_tweets)
    oldest = alltweets[-1].id -1
    while len(new_tweets)>0:
        #print("getting tweets before %s" %(oldest))
        new_tweets = api.user_timeline(screen_name = screen_name,count=200,max_id=oldest)
        alltweets.extend(new_tweets)
        oldest = alltweets[-1].id -1
        #print("...%s tweets downloaded so far" % (len(alltweets)))
    outtweets = [[tweet.id_str, tweet.created_at, tweet.text, tweet.favorite_count,
tweet.in_reply_to_screen_name, tweet.retweeted] for tweet in alltweets]
    with open(path + '%s_tweets.csv' % screen_name, 'w') as f:
        writer = csv.writer(f)
        writer.writerow(["id","created_at","text","likes","in reply to","retweeted"])
        writer.writerows(outtweets)
    pass
```

In [5]:

```
def simple_analysis(tweets):
    return(tweets.shape,tweets.created_at.min(),tweets.created_at.max())
```

In [7]:

```
top_ten_twitters = pd.read_csv("./Top_ten_twitters.csv")
## from wikipedia
```

In [8]:

```
top_ten_twitters.iloc[9,2]
```

Out[8]:

```
'Ariana Grande'
```

In [12]:

```
congress_twitter_handles = pd.read_csv("./116Congress.csv")
```

In [137]:

```
congress_twitter_handles.shape
```

Out[137]:

```
(534, 8)
```

In [22]:

```
null_cspan = congress_twitter_handles[congress_twitter_handles.CSPAN.isnull()]
```

In [23]:

```
null_cspan
```

Out[23]:

	Wikipedia Names	Position	Joined Congress	ODU-WSDL	CSPAN	TweetCongress	Github	
186	Collin Peterson	Rep	3-Jan-91	collinpeterson	NaN	NaN	NaN	colli
264	Greg Gianforte	Rep	21-Jun- 17	GregForMontana	NaN	NaN	NaN	GregFo
269	Gregorio Sablan	Rep	3-Jan-09	Kilili_Sablan	NaN	Kilili_Sablan	Kilili_Sablan	
404	Michael San Nicolas	Rep	3-Jan-19	NaN	NaN	NaN	NaN	NO

In [24]:

```
congress_twitter_handles=congress_twitter_handles[congress_twitter_handles.CSPAN.is  
null()]==False]
```

In [58]:

```
congress_twitter_handles.shape
```

Out[58]:

```
(534, 8)
```

In [181]:

```
sen_search = congress_twitter_handles[congress_twitter_handles.CSPAN == "RepCasten"]
sen_search
```

Out[181]:

	Wikipedia Names	Position	Joined Congress	ODU- WSDL	CSPAN	TweetCongress	Github	REMARK
475	Sean Casten	Rep	3-Jan-19	RepCasten	RepCasten	RepCasten	RepCasten	NaN

In [183]:

```
congress_twitter_handles.CSPAN[476]
```

Out[183]:

'RepSeanDuffy'

In [178]:

```
sen_index_search = congress_twitter_handles.iloc[455]
sen_index_search
```

Out[178]:

Wikipedia	Names	Roger Marshall
Position		Rep
Joined Congress		3-Jan-17
ODU-WSDL		RepMarshall
CSPAN		RepMarshall
TweetCongress		RepMarshall
Github		RepMarshall
REMARK		NaN

Name: 459, dtype: object

In [157]:

```
doesnotexist = [177,360,364,404]
```

In [155]:

```
error_34_index = [286,329,441]
```

In [166]:

```
notenoughtweets_index = [186,231,264,269,449,460,455]
```

In [184]:

```
suspended_name = ['RepSeanDuffy']
```

In [185]:

```
i=477
while i<534:
    handle = "@"+congress_twitter_handles.CSPAN[i]
    get_all_tweets(handle,"./Congress_tweets/")
    i+=1
```

In [213]:

```
get_all_tweets("@realDonaldTrump","./")
```

In [15]:

Out[15]:

```
'@sendougjones'
```

In [19]:

```
get_all_tweets('@sendougjones','./Congress_tweets/')
```

In [167]:

```
trump_tweets = pd.read_csv('./realDonaldTrump_tweets.csv')
```

In [139]:

```
shap, early, last = simple_analysis(trump_tweets)
```

In [140]:

```
shap
```

Out[140]:

```
(3237, 6)
```

In [157]:

```
early
```

Out[157]:

```
'2020-04-11 16:01:37'
```

Note the time is 5 hrs behind

In [89]:

```
obama_tweets = pd.read_csv("./@BarackObama_tweets.csv")
```

In [90]:

```
simple_analysis(obama_tweets)
```

```
(3230, 6)  
2014-11-15 14:09:41  
2020-07-18 05:25:25
```

In [169]:

```
get_all_tweets("@justinbieber", "./")
```

In [170]:

```
get_all_tweets("@NateSilver538", "./")
```

In [176]:

```
get_all_tweets("@ananavarro", "./")
```

In [179]:

```
get_all_tweets("@donaldJtrumpjr", "./")
```

In [180]:

```
get_all_tweets("@HillaryClinton", "./")
```

In [93]:

```
bieber_tweets = pd.read_csv('./@justinbieber_tweets.csv')
```

In [94]:

```
simple_analysis(bieber_tweets)
```

```
(3148, 6)  
2015-09-10 19:20:10  
2020-07-03 16:20:20
```

In [188]:

```
top_ten_twitters = pd.read_csv("./Top_ten_twitters.csv")
```

In [189]:

```
top_ten_twitters.head()
```

Out[189]:

	Rank	Account name	Owner	Followers	Activity
0	1	@BarackObama	Barack Obama	120	Former U.S. president
1	2	@justinbieber	Justin Bieber	112	Musician
2	3	@katyperry	Katy Perry	108	Musician
3	4	@rihanna	Rihanna	97	Musician and businesswoman
4	5	@taylorswift13	Taylor Swift	87	Musician

In [101]:

```
get_all_tweets(top_ten_twitters['Account name'][0], "./test")
```

In [154]:

```
top_ten_twitters['test']=1
```

In [216]:

```
for top_twitters in top_ten_twitters["Account name"]:  
    twitts = get_all_tweets(top_twitters, './TopTenAccounts/')
```

In [156]:

top\_ten\_twitterers

Out[156]:

	Rank	Change	Account name	Owner	Followers	Activity	test
0	1	Steady	@BarackObama	Barack Obama	120	Former U.S. president	test
1	2	Steady	@justinbieber	Justin Bieber	112	Musician	1
2	3	Steady	@katyperry	Katy Perry	108	Musician	1
3	4	Steady	@rihanna	Rihanna	97	Musician and businesswoman	1
4	5	Steady	@taylorswift13	Taylor Swift	87	Musician	1
5	6	Steady	@Cristiano	Cristiano Ronaldo	86	Footballer	1
6	7	Increase	@realDonaldTrump	Donald Trump	83	Current U.S. president	1
7	8	Decrease	@ladygaga	Lady Gaga	82	Musician	1
8	9	Steady	@TheEllenShow	Ellen DeGeneres	80	Comedian and television personality	1
9	10	Steady	@ArianaGrande	Ariana Grande	75	Musician and actress	1

In [ ]: