realDonaldTrumpTweets

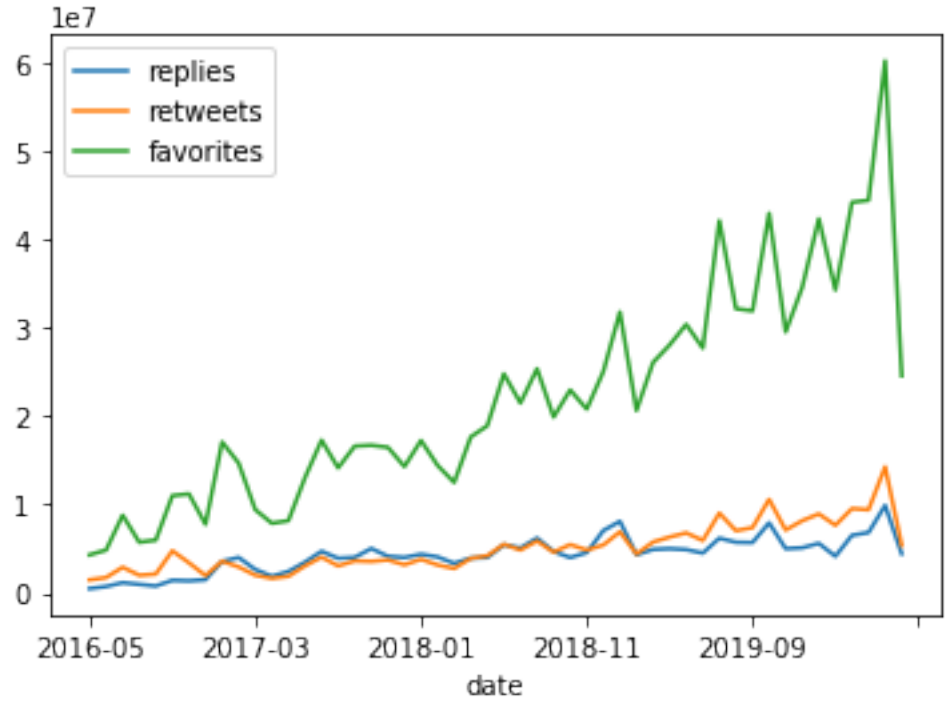
June 10, 2020

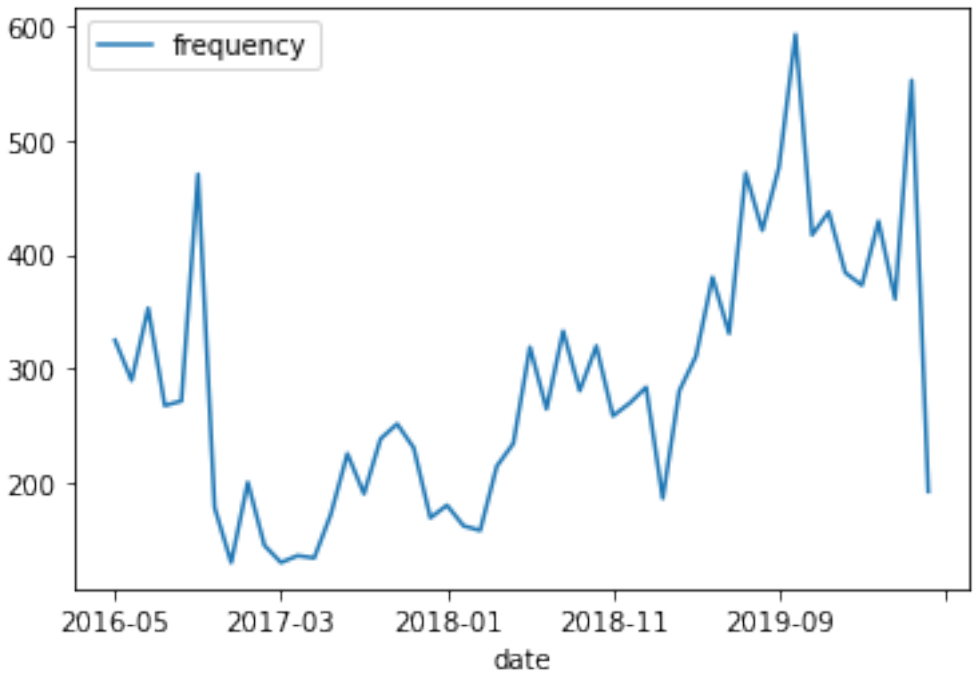
[1]:
```
import pandas

url = '/Users/min/OneDrive - The University of Texas at Dallas/UTD/Courses/EPPS␣
 ↪7V81 Advanced Data Programming/Exercise 1/realDonaldTrump.csv'
df = pandas.read_csv(url)

df['date'] = pandas.to_datetime(df['date'])

dfm1 = df.groupby(df['date'].dt.strftime('%Y-%m')).sum()

dfm2 = df.groupby(df['date'].dt.strftime('%Y-%m')).count()

dfm1 = dfm1.drop(['geo','id'], axis=1)

dfm2 = dfm2.
 ↪drop(['username','to','replies','retweets','favorites','text','geo','mentions','hashtags','
 ↪axis=1)

dfm2.columns = ['frequency']

dfm3 = pandas.concat([dfm1, dfm2], axis=1)

#dfm1['counts'] = pandas.Series(data=dfm1.groupby(dfm1['date']).count())

#counts = df.groupby(df['date'].dt.strftime('%B')).['date'].count()

print(dfm3)

dfm1.plot()
dfm2.plot()
dfm3.plot()
```

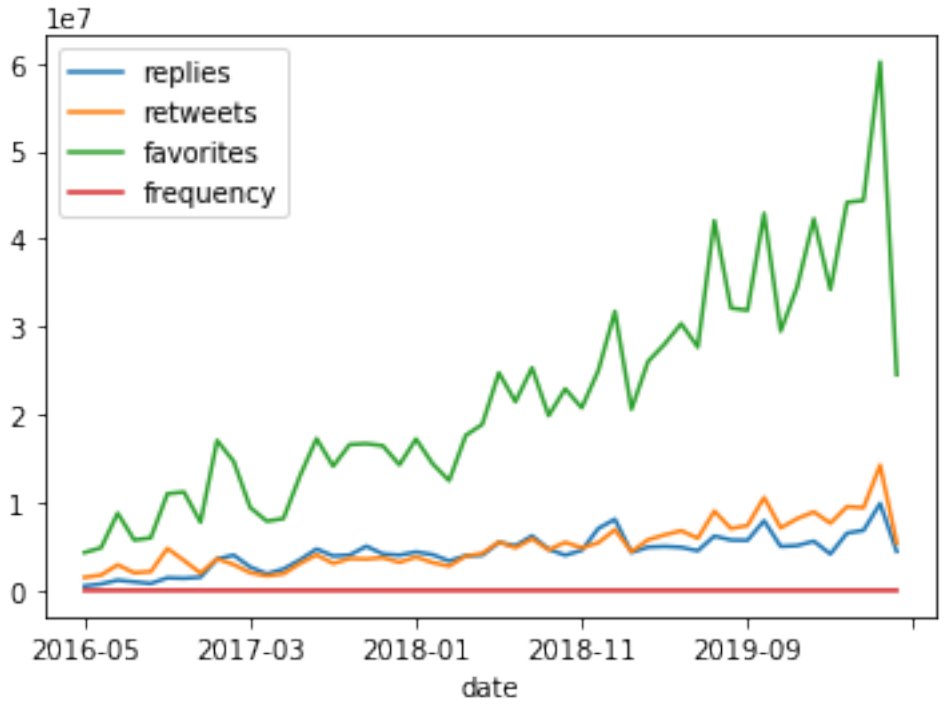
```
          replies    retweets   favorites  frequency
date
2016-05   497639.0   1481575.0   4309566.0        325
2016-06   732431.0   1746171.0   4852132.0        290
2016-07  1166429.0   2912240.0   8792222.0        353
2016-08   974934.0   2003823.0   5739703.0        268
```

2016-09	790090.0	2142912.0	5982613.0	272
2016-10	1433707.0	4768195.0	11005189.0	470
2016-11	1377526.0	3413486.0	11196803.0	179
2016-12	1500045.0	1966408.0	7764445.0	131
2017-01	3573588.0	3599981.0	17065721.0	201
2017-02	4017672.0	2944686.0	14675472.0	146
2017-03	2657039.0	2009576.0	9422961.0	131
2017-04	1888395.0	1660982.0	7881802.0	137
2017-05	2427946.0	1884934.0	8170410.0	135
2017-06	3478841.0	3065709.0	13036666.0	173
2017-07	4715252.0	4048937.0	17269000.0	226
2017-08	3925454.0	3065041.0	14142139.0	191
2017-09	4024275.0	3626714.0	16604062.0	239
2017-10	5030746.0	3570307.0	16698224.0	252
2017-11	4145891.0	3732339.0	16469030.0	231
2017-12	4014768.0	3192018.0	14294271.0	170
2018-01	4377597.0	3804248.0	17234442.0	181
2018-02	4078667.0	3168443.0	14423691.0	163
2018-03	3349013.0	2767793.0	12492671.0	159
2018-04	3944364.0	3907695.0	17640195.0	215
2018-05	4007668.0	4208369.0	18905480.0	235
2018-06	5458511.0	5505441.0	24770741.0	319
2018-07	5089196.0	4858991.0	21476050.0	265
2018-08	6207122.0	5877316.0	25341978.0	333
2018-09	4693975.0	4614792.0	19895797.0	281
2018-10	4004275.0	5471543.0	22944502.0	320
2018-11	4569435.0	4852052.0	20784554.0	259
2018-12	7055468.0	5426277.0	25031426.0	270
2019-01	8099190.0	6915655.0	31758974.0	284
2019-02	4324207.0	4405362.0	20629331.0	187
2019-03	4927358.0	5760091.0	26064702.0	281
2019-04	5008934.0	6324528.0	28037058.0	311
2019-05	4909342.0	6789916.0	30362076.0	380
2019-06	4511780.0	5942285.0	27687912.0	331
2019-07	6193904.0	9022321.0	42089476.0	471
2019-08	5736332.0	7044162.0	32135583.0	421
2019-09	5705325.0	7390191.0	31891325.0	476
2019-10	7928670.0	10582272.0	42917891.0	592
2019-11	5008768.0	7118812.0	29513066.0	417
2019-12	5107822.0	8173483.0	34559862.0	437
2020-01	5609144.0	8942520.0	42303334.0	384
2020-02	4147625.0	7655803.0	34245644.0	373
2020-03	6521245.0	9539524.0	44164501.0	429
2020-04	6836206.0	9409940.0	44394067.0	361
2020-05	9894789.0	14258930.0	60161546.0	552
2020-06	4432992.0	5388251.0	24557817.0	193

[1]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc03fd29ed0>






1e7
6
5
4
3
2
1
0
replies
retweets
favorites
frequency
2016-05
2017-03
2018-01
2018-11
2019-09
date