# EPPS7V81 Exercise 2

June 15, 2020

[1]:
```
import pandas as pd
df= pd.read_csv('cdc.csv')
print (df)
df['date']=pd.to_datetime(df['date'])
```

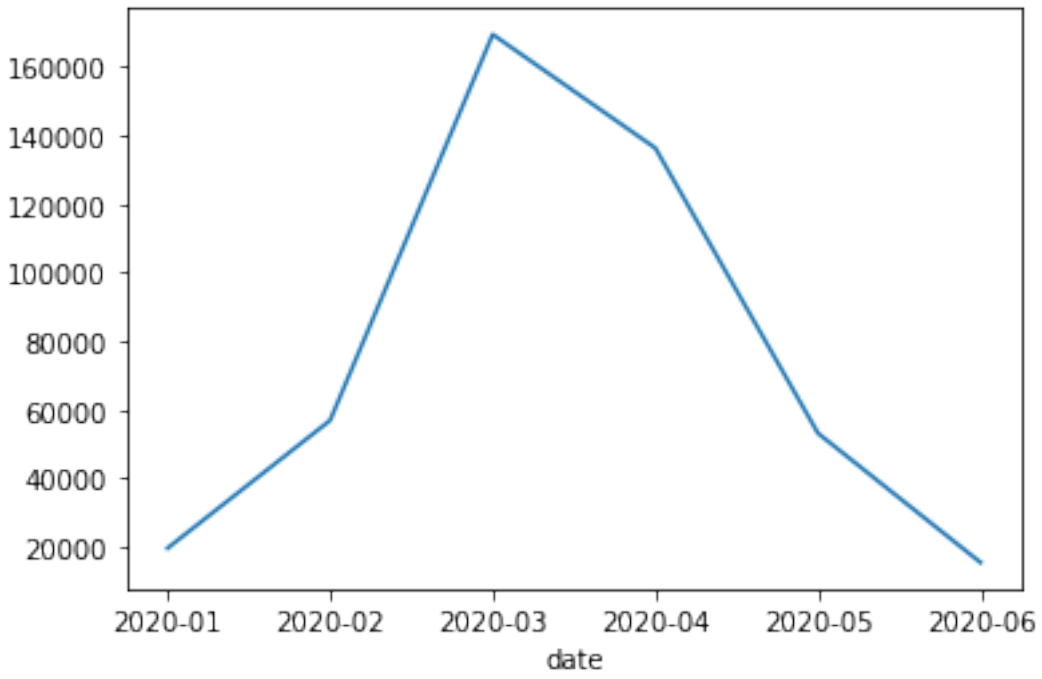
```
                   date username   to  replies  retweets  favorites  \
0    2020-06-12 21:45:03   CDCgov  NaN       79       170        300
1    2020-06-12 21:00:00   CDCgov  NaN      151       321        485
2    2020-06-12 20:03:02   CDCgov  NaN      218      1309       1992
3    2020-06-12 18:35:02   CDCgov  NaN       94       262        437
4    2020-06-12 17:45:04   CDCgov  NaN      121       610        674
..                   …      …  …      …        …         …
933  2020-01-02 20:08:56   CDCgov  NaN       87       414        400
934  2020-01-02 18:07:13   CDCgov  NaN        0        32         27
935  2020-01-02 16:34:01   CDCgov  NaN        0        42         52
936  2020-01-02 15:31:01   CDCgov  NaN        5        95        111
937  2020-01-01 15:33:01   CDCgov  NaN        6        32        248

                                                  text  geo  mentions  \
0    The latest CDC #COVIDView report shows that, a…  NaN       NaN
1    The coming weeks could see a possible increase…  NaN       NaN
2    #Employers: Antibody tests for #COVID19 should…  NaN       NaN
3    It’s time to slow the spread of #COVID19. Heal…  NaN       NaN
4    As of June 8, forecasts suggest the total numb…  NaN       NaN
..                                                 …  …       …
933  OUTBREAK UPDATE: As of 12/27, 2,561 hospitaliz…  NaN       NaN
934  NEW Report: Recommendations for Providing Qual…  NaN  @CDCMMWR
935  #DYK? At least 1.7 million adults in the U.S. …  NaN       NaN
936  Vaccinate your son or daughter against HPV whe…  NaN       NaN
937  CDC wishes you and your family a healthy and #…  NaN       NaN

                  hashtags                   id  \
0      #COVIDView #COVID19  1271559161583714305
1                 #COVID19  1271547824400990211
2      #Employers #COVID19  1271533487665881088
3                 #COVID19  1271511344785821697
4                 #COVID19  1271498768274382848
..                      …                   …
```

```
933                    NaN  1212828138734010368
934  #STD #STDcare #STDQCS  1212797510923800576
935  #DYK #sepsis #NewYear  1212774053783724033
936                    NaN  1212758200338006016
937          #HappyNewYear  1212396317168881664

                                              permalink
0    https://twitter.com/CDCgov/status/127155916158…
1    https://twitter.com/CDCgov/status/127154782440…
2    https://twitter.com/CDCgov/status/127153348766…
3    https://twitter.com/CDCgov/status/127151134478…
4    https://twitter.com/CDCgov/status/127149876827…
..                                                 …
933  https://twitter.com/CDCgov/status/121282813873…
934  https://twitter.com/CDCgov/status/121279751092…
935  https://twitter.com/CDCgov/status/121277405378…
936  https://twitter.com/CDCgov/status/121275820033…
937  https://twitter.com/CDCgov/status/121239631716…

[938 rows x 12 columns]
```

[2]:
```
df['date']=pd.to_datetime(df['date'])
df_retweets=df.groupby(df['date'].dt.strftime('%Y-%m'))['retweets'].sum()
order=['Jan''Feb''Mar''Apr''May''June']
df2=df.reindex(order, axis=1)
print(df_retweets)
df_retweets.plot()
```

```
date
2020-01     19635
2020-02     56910
2020-03    169495
2020-04    136362
2020-05     53185
2020-06     15473
Name: retweets, dtype: int64
```

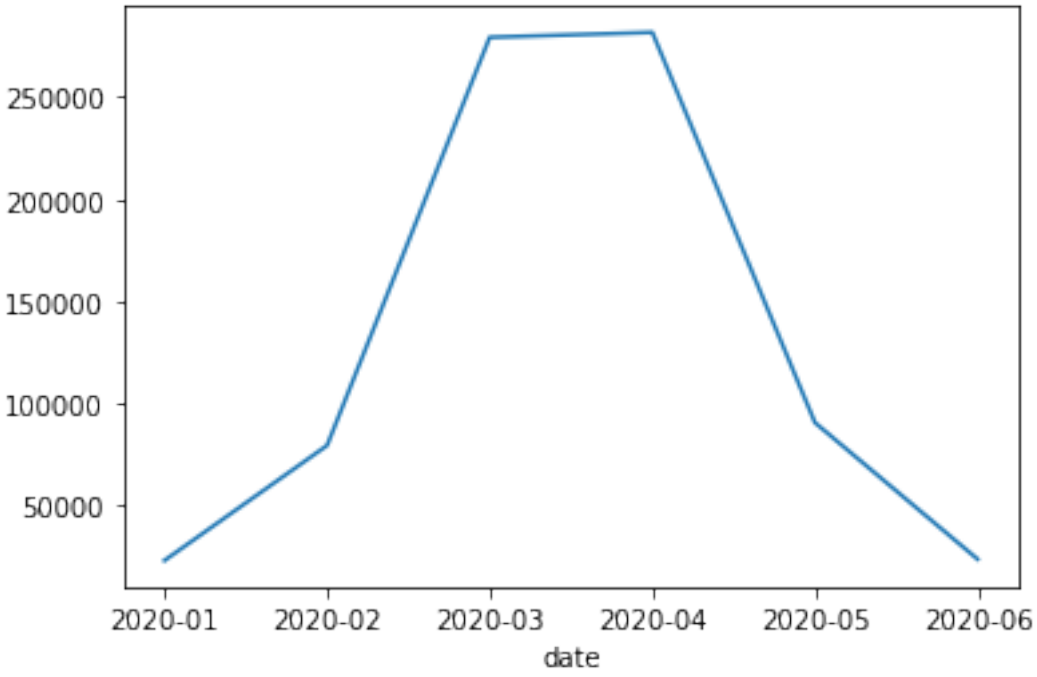
[2]: <matplotlib.axes._subplots.AxesSubplot at 0x1147e9490>




[3]:
```
df['date']=pd.to_datetime(df['date'])
df_fav=df.groupby(df['date'].dt.strftime('%Y-%m'))['favorites'].sum()
order=['Jan''Feb''Mar''Apr''May''June']
df2=df.reindex(order, axis=1)
print(df_fav)
df_fav.plot()
```

```
date
2020-01     23260
2020-02     79540
2020-03    279190
2020-04    281582
2020-05     90700
2020-06     23935
Name: favorites, dtype: int64
```

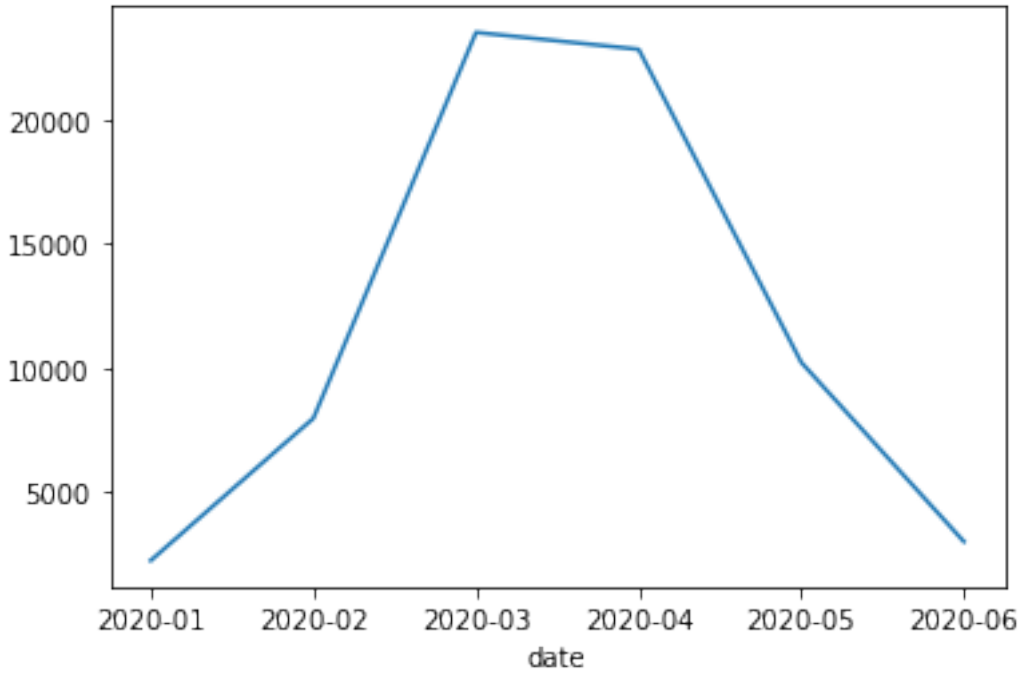
[3]: `<matplotlib.axes._subplots.AxesSubplot at 0x1208e65d0>`




[4]:
```
df['date']=pd.to_datetime(df['date'])
df_rep=df.groupby(df['date'].dt.strftime('%Y-%m'))['replies'].sum()
order=['Jan''Feb''Mar''Apr''May''June']
df3=df.reindex(order, axis=1)
print(df_rep)
df_rep.plot()
```

```
date
2020-01     2201
2020-02     7956
2020-03    23565
2020-04    22881
2020-05    10216
2020-06     2965
Name: replies, dtype: int64
```

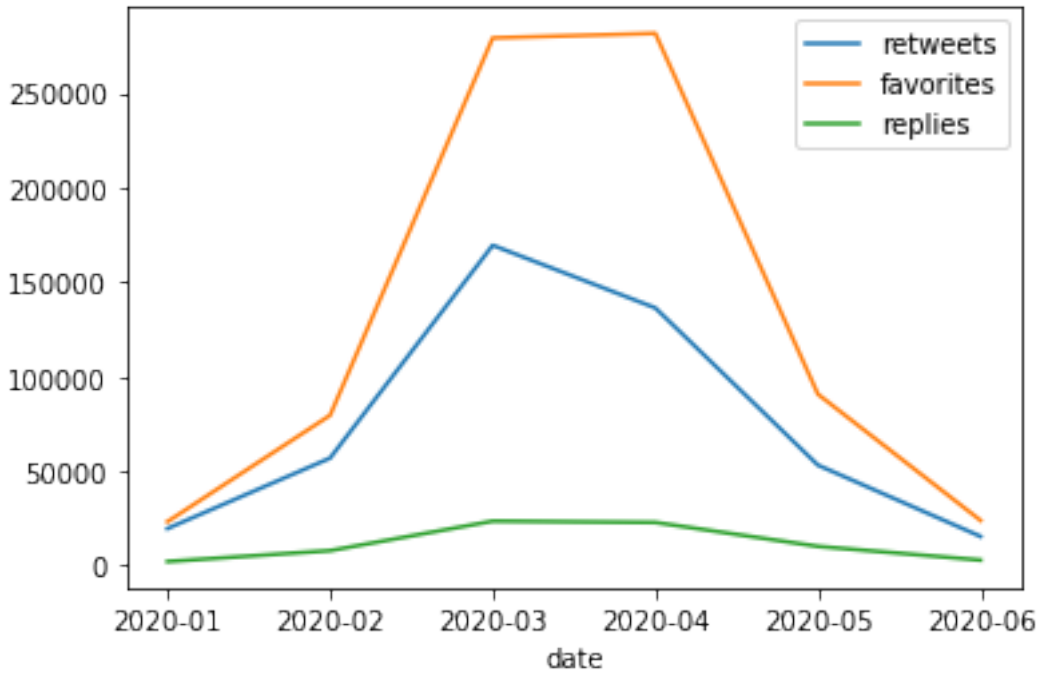
[4]: `<matplotlib.axes._subplots.AxesSubplot at 0x112d33390>`




[5]:
```
df['date']=pd.to_datetime(df['date'])
df_retweets=df.groupby(df['date'].dt.strftime('%Y-%m'))['retweets'].sum()
df_fav=df.groupby(df['date'].dt.strftime('%Y-%m'))['favorites'].sum()
df_rep=df.groupby(df['date'].dt.strftime('%Y-%m'))['replies'].sum()
order=['Jan''Feb''Mar''Apr''May''June']
df1=df.reindex(order, axis=0)
print(df_retweets,df_fav)
df4 = pd.concat([df_retweets, df_fav, df_rep], axis=1)
df4.plot()
```

```
date
2020-01     19635
2020-02     56910
2020-03    169495
2020-04    136362
2020-05     53185
2020-06     15473
Name: retweets, dtype: int64 date
2020-01     23260
2020-02     79540
2020-03    279190
2020-04    281582
2020-05     90700
2020-06     23935
Name: favorites, dtype: int64
```

[5]: <matplotlib.axes._subplots.AxesSubplot at 0x112d9ac10>




[6]:

```
      ␣
↪---------------------------------------------------------------------------

       TypeError                                 Traceback (most recent call␣
↪last)

       <ipython-input-6-57c60942be02> in <module>
         1 df['date']=pd.to_datetime(df['date'])
   ----> 2 df_fav=df.groupby(df['date'].dt.strftime('%Y-%m'))['mentions'].sum()
         3 order=['Jan''Feb''Mar''Apr''May''June']
         4 df2=df.reindex(order, axis=1)
         5 print(df_fav)


       ~/opt/anaconda3/lib/python3.7/site-packages/pandas/core/groupby/groupby.
↪py in f(self, **kwargs)
      1390
      1391                 # apply a non-cython aggregation
   -> 1392                 result = self.aggregate(lambda x: npfunc(x,␣
↪axis=self.axis))
```

```
      1393                 return result
      1394
```

```
      ~/opt/anaconda3/lib/python3.7/site-packages/pandas/core/groupby/generic.
↪py in aggregate(self, func, *args, **kwargs)
       263
       264             try:
   --> 265                 return self._python_agg_general(func, *args,␣
↪**kwargs)
       266             except (ValueError, KeyError):
       267                 # TODO: KeyError is raised in _python_agg_general,
```

```
      ~/opt/anaconda3/lib/python3.7/site-packages/pandas/core/groupby/groupby.
↪py in _python_agg_general(self, func, *args, **kwargs)
       934                 pass
       935
   --> 936             result, counts = self.grouper.agg_series(obj, f)
       937             assert result is not None
       938             key = base.OutputKey(label=name, position=idx)
```

```
      ~/opt/anaconda3/lib/python3.7/site-packages/pandas/core/groupby/ops.py␣
↪in agg_series(self, obj, func)
       639
       640         try:
   --> 641             return self._aggregate_series_fast(obj, func)
       642         except ValueError as err:
       643             if "Function does not reduce" in str(err):
```

```
      ~/opt/anaconda3/lib/python3.7/site-packages/pandas/core/groupby/ops.py␣
↪in _aggregate_series_fast(self, obj, func)
       664         group_index = algorithms.take_nd(group_index, indexer,␣
↪allow_fill=False)
       665         grouper = libreduction.SeriesGrouper(obj, func, group_index,␣
↪ngroups, dummy)
   --> 666         result, counts = grouper.get_result()
       667         return result, counts
       668
```

```
      pandas/_libs/reduction.pyx in pandas._libs.reduction.SeriesGrouper.
↪get_result()
```

```
      pandas/_libs/reduction.pyx in pandas._libs.reduction._BaseGrouper.
↪_apply_to_group()


      ~/opt/anaconda3/lib/python3.7/site-packages/pandas/core/groupby/groupby.
↪py in <lambda>(x)
      911     def _python_agg_general(self, func, *args, **kwargs):
      912         func = self._is_builtin_func(func)
  --> 913         f = lambda x: func(x, *args, **kwargs)
      914
      915         # iterate through "columns" ex exclusions to populate output␣
↪dict


      ~/opt/anaconda3/lib/python3.7/site-packages/pandas/core/groupby/groupby.
↪py in <lambda>(x)
     1390
     1391                 # apply a non-cython aggregation
  -> 1392                 result = self.aggregate(lambda x: npfunc(x,␣
↪axis=self.axis))
     1393                 return result
     1394


      <__array_function__ internals> in sum(*args, **kwargs)


      ~/opt/anaconda3/lib/python3.7/site-packages/numpy/core/fromnumeric.py in␣
↪sum(a, axis, dtype, out, keepdims, initial, where)
     2227
     2228     return _wrapreduction(a, np.add, 'sum', axis, dtype, out,␣
↪keepdims=keepdims,
  -> 2229                           initial=initial, where=where)
     2230
     2231


      ~/opt/anaconda3/lib/python3.7/site-packages/numpy/core/fromnumeric.py in␣
↪_wrapreduction(obj, ufunc, method, axis, dtype, out, **kwargs)
       86                 return reduction(axis=axis, dtype=dtype, out=out,␣
↪**passkwargs)
       87             else:
  ---> 88                 return reduction(axis=axis, out=out, **passkwargs)
       89
       90     return ufunc.reduce(obj, axis, dtype, out, **passkwargs)
```

```
      ~/opt/anaconda3/lib/python3.7/site-packages/pandas/core/generic.py in␣
↪stat_func(self, axis, skipna, level, numeric_only, min_count, **kwargs)
    11182             skipna=skipna,
    11183             numeric_only=numeric_only,
  > 11184             min_count=min_count,
    11185         )
    11186


      ~/opt/anaconda3/lib/python3.7/site-packages/pandas/core/series.py in␣
↪_reduce(self, op, name, axis, skipna, numeric_only, filter_type, **kwds)
     3889                 )
     3890             with np.errstate(all="ignore"):
  -> 3891                 return op(delegate, skipna=skipna, **kwds)
     3892
     3893         # TODO(EA) dispatch to Index


      ~/opt/anaconda3/lib/python3.7/site-packages/pandas/core/nanops.py in␣
↪_f(*args, **kwargs)
       67             try:
       68                 with np.errstate(invalid="ignore"):
  ---> 69                     return f(*args, **kwargs)
       70             except ValueError as e:
       71                 # we want to transform an object array


      ~/opt/anaconda3/lib/python3.7/site-packages/pandas/core/nanops.py in␣
↪nansum(values, axis, skipna, min_count, mask)
      491     elif is_timedelta64_dtype(dtype):
      492         dtype_sum = np.float64
  --> 493     the_sum = values.sum(axis, dtype=dtype_sum)
      494     the_sum = _maybe_null_out(the_sum, axis, mask, values.shape,␣
↪min_count=min_count)
      495


      ~/opt/anaconda3/lib/python3.7/site-packages/numpy/core/_methods.py in␣
↪_sum(a, axis, dtype, out, keepdims, initial, where)
       36 def _sum(a, axis=None, dtype=None, out=None, keepdims=False,
       37          initial=_NoValue, where=True):
  ---> 38     return umr_sum(a, axis, dtype, out, keepdims, initial, where)
       39
       40 def _prod(a, axis=None, dtype=None, out=None, keepdims=False,


      TypeError: unsupported operand type(s) for +: 'int' and 'str'
```

[ ]: