In [3]:
```
import pandas as pd
import numpy as np
```

In [4]:
```
from urllib.request import urlretrieve
```

In [5]:
```
italy_covid_url = 'https://gist.githubusercontent.com/aakashns/f6a004fa20c84fec53

urlretrieve(italy_covid_url, './glo/mtn/covid_daily.csv')
```

Out[5]:
```
('./glo/mtn/covid_daily.csv', <http.client.HTTPMessage at 0x105635c8>)
```

In [6]:
```
urlretrieve('https://gist.githubusercontent.com/aakashns/8684589ef4f266116cdce023
            './glo/mtn/locations.csv')
```

Out[6]:
```
('./glo/mtn/locations.csv', <http.client.HTTPMessage at 0x1056eac8>)
```

In [7]:
```
locations_df = pd.read_csv('./glo/mtn/locations.csv')
```

In [8]:
```
files = np.genfromtxt("./glo/covid.txt", delimiter=",", skip_header=1) #opening f
```

In [9]:
```
fili = pd.read_csv("./glo/mtn/covid_daily.txt") #opening file using pandas
```

In [10]:
```
file1 = open("./glo/covid.txt", "r") #opening file using plain python
```

In [11]:
```
file1.readlines()
```

Out[11]:
```
['date,new_cases,new_deaths,new_tests\n',
 '2020-04-21,2256.0,454.0,28045.0\n',
 '2020-03-11,2236.0,451.0,28095.0\n',
 '2020-05-21,2250.0,354.0,28195.0\n',
 '2020-04-21,2256.0,404.0,28090.0\n',
 '2020-05-23,1256.0,804.0,18095.0\n',
 '2020-04-20,2258.0,451.0,21095.0']
```

In [12]:
```
files
```

Out[12]:
```
array([[   nan,  2256.,   454., 28045.],
       [   nan,  2236.,   451., 28095.],
       [   nan,  2250.,   354., 28195.],
       [   nan,  2256.,   404., 28090.],
       [   nan,  1256.,   804., 18095.],
       [   nan,  2258.,   451., 21095.]])
```

Pandas format is simliar to this

covid_data_dict = { 'date': ['2020-08-30', '2020-08-31', '2020-09-01', '2020-09-02', '2020-09-03'], 'new_cases': [1444, 1365, 996, 975, 1326], 'new_deaths': [1, 4, 6, 8, 6], 'new_tests': [53541, 42583, 54395, None, None] }

In [ ]:

In [13]:
```
# Pandas format is simliar to this
covid_data_dict = {
    'date':       ['2020-08-30', '2020-08-31', '2020-09-01', '2020-09-02', '2020-
    'new_cases':  [1444, 1365, 996, 975, 1326],
    'new_deaths': [1, 4, 6, 8, 6],
    'new_tests': [53541, 42583, 54395, None, None]
}
```

In [14]:
```
fili.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 248 entries, 0 to 247
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   date        248 non-null    object 
 1   new_cases   248 non-null    float64
 2   new_deaths  248 non-null    float64
 3   new_tests   135 non-null    float64
dtypes: float64(3), object(1)
memory usage: 7.9+ KB
```

In [15]:
```
fili.describe()
```

Out[15]:

	new_cases	new_deaths	new_tests
count	248.000000	248.000000	135.000000
mean	1094.818548	143.133065	31699.674074
std	1554.508002	227.105538	11622.209757
min	-148.000000	-31.000000	7841.000000
25%	123.000000	3.000000	25259.000000
50%	342.000000	17.000000	29545.000000
75%	1371.750000	175.250000	37711.000000
max	6557.000000	971.000000	95273.000000

In [16]:
```
fili.columns
```

Out[16]:
```
Index(['date', 'new_cases', 'new_deaths', 'new_tests'], dtype='object')
```

In [17]:
```
fili["new_cases"][243]
```

Out[17]:
```
1444.0
```

In [18]:
```
fili.new_cases
```

Out[18]:
```
0         0.0
1         0.0
2         0.0
3         0.0
4         0.0
        ...
243    1444.0
244    1365.0
245     996.0
246     975.0
247    1326.0
Name: new_cases, Length: 248, dtype: float64
```

In [19]:
```
fili.at[243, "new_cases"]
```

Out[19]:
```
1444.0
```

In [20]:
```
fili["date"]
```

Out[20]:
```
0      2019-12-31
1      2020-01-01
2      2020-01-02
3      2020-01-03
4      2020-01-04
          ...
243    2020-08-30
244    2020-08-31
245    2020-09-01
246    2020-09-02
247    2020-09-03
Name: date, Length: 248, dtype: object
```

In [21]:
```
fili.at[2, "date"]
```

Out[21]:
```
'2020-01-02'
```

In [22]:
```
fili[["new_deaths", "new_tests"]]
```

Out[22]:

	new_deaths	new_tests
0	0.0	NaN
1	0.0	NaN
2	0.0	NaN
3	0.0	NaN
4	0.0	NaN
...	...	...
243	1.0	53541.0
244	4.0	42583.0
245	6.0	54395.0
246	8.0	NaN
247	6.0	NaN

248 rows × 2 columns

In [23]:
```
fili = fili.copy()
```

In [24]:
```
fili.loc[243:247]
```

Out[24]:

	date	new_cases	new_deaths	new_tests
243	2020-08-30	1444.0	1.0	53541.0
244	2020-08-31	1365.0	4.0	42583.0
245	2020-09-01	996.0	6.0	54395.0
246	2020-09-02	975.0	8.0	NaN
247	2020-09-03	1326.0	6.0	NaN

In [25]:
```
fili.sample(2)
```

Out[25]:

	date	new_cases	new_deaths	new_tests
15	2020-01-15	0.0	0.0	NaN
174	2020-06-22	224.0	24.0	16152.0

In [26]:
```
fili.head(2)
```

Out[26]:

	date	new_cases	new_deaths	new_tests
0	2019-12-31	0.0	0.0	NaN
1	2020-01-01	0.0	0.0	NaN

In [27]:
```
file3 = fili[["date", "new_deaths"]]
```

In [28]:
```
feel = fili["new_deaths"]
```

In [29]:
```
feel.sum()
```

Out[29]: 35497.0

In [30]:
```
fili.new_deaths.sum()
```

Out[30]: 35497.0

In [31]:
```
fili.new_tests.first_valid_index()
```

Out[31]: 111

In [32]:
```
fili.loc[109:112]
```

Out[32]:

	date	new_cases	new_deaths	new_tests
109	2020-04-18	3493.0	575.0	NaN
110	2020-04-19	3491.0	480.0	NaN
111	2020-04-20	3047.0	433.0	7841.0
112	2020-04-21	2256.0	454.0	28095.0

In [33]:
```
fili
```

Out[33]:

	date	new_cases	new_deaths	new_tests
0	2019-12-31	0.0	0.0	NaN
1	2020-01-01	0.0	0.0	NaN
2	2020-01-02	0.0	0.0	NaN
3	2020-01-03	0.0	0.0	NaN
4	2020-01-04	0.0	0.0	NaN
...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0
244	2020-08-31	1365.0	4.0	42583.0
245	2020-09-01	996.0	6.0	54395.0
246	2020-09-02	975.0	8.0	NaN
247	2020-09-03	1326.0	6.0	NaN

248 rows × 4 columns

In [34]:
```
the_file = fili["new_cases"] > 1000
```

In [35]:
```
the_file
```

Out[35]:
```
0      False
1      False
2      False
3      False
4      False
       ...
243     True
244     True
245    False
246    False
247     True
Name: new_cases, Length: 248, dtype: bool
```

In [36]:
```
fili[the_file]
```

Out[36]:

	date	new_cases	new_deaths	new_tests
68	2020-03-08	1247.0	36.0	NaN
69	2020-03-09	1492.0	133.0	NaN
70	2020-03-10	1797.0	98.0	NaN
72	2020-03-12	2313.0	196.0	NaN
73	2020-03-13	2651.0	189.0	NaN
...	...	...	...	...
241	2020-08-28	1409.0	5.0	65135.0
242	2020-08-29	1460.0	9.0	64294.0
243	2020-08-30	1444.0	1.0	53541.0
244	2020-08-31	1365.0	4.0	42583.0
247	2020-09-03	1326.0	6.0	NaN

72 rows × 4 columns

In [37]:
```
fili["positive_rate"] = fili.new_cases / fili.new_tests
```

In [38]:
```
fili.drop(columns=["positive_rate"], inplace=True)
```

In [39]:
```
thy_file = fili[the_file]
```

In [40]:
```
fili.sort_values("new_cases", ascending=False).tail(10)
```

Out[40]:

	date	new_cases	new_deaths	new_tests
23	2020-01-23	0.0	0.0	NaN
24	2020-01-24	0.0	0.0	NaN
25	2020-01-25	0.0	0.0	NaN
26	2020-01-26	0.0	0.0	NaN
27	2020-01-27	0.0	0.0	NaN
28	2020-01-28	0.0	0.0	NaN
30	2020-01-30	0.0	0.0	NaN
32	2020-02-01	0.0	0.0	NaN
33	2020-02-02	0.0	0.0	NaN
172	2020-06-20	-148.0	47.0	29875.0

In [41]:
```
fili.at[172, "new_cases"] = (fili.at[170, "new_cases"] + fili.at[171, "new_cases"
                             + fili.at[174, "new_cases"]) / 4
```

In [42]:
```
fili.at[172, "new_cases"]
```

Out[42]:
```
286.75
```

In [43]:
```
thy_file.sum()
```

Out[43]:
```
date          2020-03-082020-03-092020-03-102020-03-122020-0...
new_cases                                                224707.0
new_deaths                                                30251.0
new_tests                                               1142213.0
dtype: object
```

In [44]:
```
fili.sort_values("new_deaths", ascending=False).head(2)
```

Out[44]:

	date	new_cases	new_deaths	new_tests
88	2020-03-28	5959.0	971.0	NaN
89	2020-03-29	5974.0	887.0	NaN

In [45]:
```
fili["date"] = pd.to_datetime(fili.date)
```

In [46]:
```
fili["date"]
```

Out[46]:
```
0     2019-12-31
1     2020-01-01
2     2020-01-02
3     2020-01-03
4     2020-01-04
         ...
243   2020-08-30
244   2020-08-31
245   2020-09-01
246   2020-09-02
247   2020-09-03
Name: date, Length: 248, dtype: datetime64[ns]
```

In [47]:
```
file["year"] = pd.DatetimeIndex(file.date).year
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2404\2030919351.py in <module>
----> 1 file["year"] = pd.DatetimeIndex(file.date).year

NameError: name 'file' is not defined
```

In [ ]:
```
file
```

In [ ]:
```
fili["year"] = pd.DatetimeIndex(fili.date).year
fili["month"] = pd.DatetimeIndex(fili.date).month
fili["day"] = pd.DatetimeIndex(fili.date).day
fili["weekday"] = pd.DatetimeIndex(fili.date).weekday
fili["date"] = pd.to_datetime(fili.date)
```

In [ ]:
```
file
```

In [ ]:
```
may_file = fili[fili.month == 5]
```

In [ ]:
```
may_file_total = may_file[["new_cases", "new_deaths", "new_tests"]]
```

In [ ]:
```
may_file_total.sum()
```

In [ ]:
```
fili
```

In [ ]:
```
sunday_cases = fili[fili.weekday == 6]  #sunday is 6
```

In [ ]:
```
total_sunday_cases = sunday_cases.new_cases.mean()
```

In [ ]:
```
total_sunday_cases
```

In [62]:
```
this = fili.groupby("weekday")["new_cases"].mean()
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2404\3809036290.py in <module>
----> 1 this = fili.groupby("weekday")["new_cases"].mean()

C:\New\envs\snakes\lib\site-packages\pandas\core\frame.py in groupby(self, by,
 axis, level, as_index, sort, group_keys, squeeze, observed, dropna)
   7639             squeeze=squeeze,  # type: ignore[arg-type]
   7640             observed=observed,
-> 7641             dropna=dropna,
   7642         )
   7643

C:\New\envs\snakes\lib\site-packages\pandas\core\groupby\groupby.py in __init__
(self, obj, keys, axis, level, grouper, exclusions, selection, as_index, sort,
 group_keys, squeeze, observed, mutated, dropna)
    895                 observed=observed,
    896                 mutated=self.mutated,
--> 897                 dropna=self.dropna,
    898             )
    899

C:\New\envs\snakes\lib\site-packages\pandas\core\groupby\grouper.py in get_grou
per(obj, key, axis, level, sort, observed, mutated, validate, dropna)
    860                 in_axis, level, gpr = False, gpr, None
    861             else:
--> 862                 raise KeyError(gpr)
    863         elif isinstance(gpr, Grouper) and gpr.key is not None:
    864             # Add key to exclusions

KeyError: 'weekday'
```

In [ ]:
```
fili.new_cases.mean()
```

In [ ]:
```
locations_df
```

In [ ]:
```
fili["location"] = "Italy"
```

In [48]:
```
fili
```

Out[48]:

	date	new_cases	new_deaths	new_tests
0	2019-12-31	0.0	0.0	NaN
1	2020-01-01	0.0	0.0	NaN
2	2020-01-02	0.0	0.0	NaN
3	2020-01-03	0.0	0.0	NaN
4	2020-01-04	0.0	0.0	NaN
...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0
244	2020-08-31	1365.0	4.0	42583.0
245	2020-09-01	996.0	6.0	54395.0
246	2020-09-02	975.0	8.0	NaN
247	2020-09-03	1326.0	6.0	NaN

248 rows × 4 columns

In [49]:
```
locations_df
```

Out[49]:

	location	continent	population	life_expectancy	hospital_beds_per_thousand	gdp_per_ca
0	Afghanistan	Asia	3.892834e+07	64.83	0.500	1803
1	Albania	Europe	2.877800e+06	78.57	2.890	11803
2	Algeria	Africa	4.385104e+07	76.88	1.900	13913
3	Andorra	Europe	7.726500e+04	83.73	NaN	
4	Angola	Africa	3.286627e+07	61.15	NaN	5819
...	...	...	...	...	...	
207	Yemen	Asia	2.982597e+07	66.12	0.700	1479
208	Zambia	Africa	1.838396e+07	63.89	2.000	3689
209	Zimbabwe	Africa	1.486293e+07	61.49	1.700	1899
210	World	NaN	7.794799e+09	72.58	2.705	15469
211	International	NaN	NaN	NaN	NaN	

212 rows × 6 columns

In [50]:
```
check = locations_df[locations_df.location == "Italy"]
```

In [51]:
```
check
```

Out[51]:

	location	continent	population	life_expectancy	hospital_beds_per_thousand	gdp_per_capita
97	Italy	Europe	60461828.0	83.51	3.18	35220.084

In [52]:
```
merged_df = fili.merge(locations_df, on="location")
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2404\877527800.py in <module>
----> 1 merged_df = fili.merge(locations_df, on="location")

C:\New\envs\snakes\lib\site-packages\pandas\core\frame.py in merge(self, right,
how, on, left_on, right_on, left_index, right_index, sort, suffixes, copy, indi
cator, validate)
   9201             copy=copy,
   9202             indicator=indicator,
-> 9203             validate=validate,
   9204         )
   9205

C:\New\envs\snakes\lib\site-packages\pandas\core\reshape\merge.py in merge(lef
t, right, how, on, left_on, right_on, left_index, right_index, sort, suffixes,
 copy, indicator, validate)
    117         copy=copy,
    118         indicator=indicator,
--> 119         validate=validate,
    120     )
    121     return op.get_result()

C:\New\envs\snakes\lib\site-packages\pandas\core\reshape\merge.py in __init__(s
elf, left, right, how, on, left_on, right_on, axis, left_index, right_index, so
rt, suffixes, copy, indicator, validate)
    697             self.right_join_keys,
    698             self.join_names,
--> 699         ) = self._get_merge_keys()
    700
    701         # validate the merge keys dtypes. We may need to coerce

C:\New\envs\snakes\lib\site-packages\pandas\core\reshape\merge.py in _get_merge
_keys(self)
   1107                         right_keys.append(rk)
   1108                     if lk is not None:
-> 1109                         left_keys.append(left._get_label_or_level_value
s(lk))
   1110                         join_names.append(lk)
   1111                     else:

C:\New\envs\snakes\lib\site-packages\pandas\core\generic.py in _get_label_or_le
vel_values(self, key, axis)
   1777             values = self.axes[axis].get_level_values(key)._values
   1778         else:
-> 1779             raise KeyError(key)
   1780
   1781         # Check for duplicates

KeyError: 'location'
```

In [54]:
```
merged_df
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2404\1877286824.py in <module>
----> 1 merged_df

NameError: name 'merged_df' is not defined
```

In [55]:
```
merged_df["cases_per_million"] = (merged_df.new_cases * 1000000) / merged_df.popu
merged_df["deaths_per_million"] = (merged_df.new_deaths * 1000000) / merged_df.po
merged_df["tests_per_million"] = (merged_df.new_tests * 1000000) / merged_df.popu
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2404\4066439773.py in <module>
----> 1 merged_df["cases_per_million"] = (merged_df.new_cases * 1000000) / merg
ed_df.population
      2 merged_df["deaths_per_million"] = (merged_df.new_deaths * 1000000) / me
rged_df.population
      3 merged_df["tests_per_million"] = (merged_df.new_tests * 1000000) / merg
ed_df.population

NameError: name 'merged_df' is not defined
```

In [56]:
```
merged_df
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2404\1877286824.py in <module>
----> 1 merged_df

NameError: name 'merged_df' is not defined
```

In [57]:
```
keep = merged_df.to_csv("./glo/mtn/solved_data", index=None)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2404\3564153378.py in <module>
----> 1 keep = merged_df.to_csv("./glo/mtn/solved_data", index=None)

NameError: name 'merged_df' is not defined
```

In [84]:
```
monthly = fili.groupby("month")
```

In [85]:
```
monthly_group = monthly[["new_cases", "new_tests", "new_deaths"]]
```

In [86]:
```
monthly_group.sum()
```

Out[86]:

month	new_cases	new_tests	new_deaths
1	3.00	0.0	0.0
2	885.00	0.0	21.0
3	100851.00	0.0	11570.0
4	101852.00	419591.0	16091.0
5	29073.00	1078720.0	5658.0
6	8206.75	830354.0	1404.0
7	6722.00	797692.0	388.0
8	21060.00	1098704.0	345.0
9	3297.00	54395.0	20.0
12	0.00	0.0	0.0

In [45]:
```
weekly_group = file.groupby("weekday")[["new_cases", "new_tests", "new_deaths"]]
```

In [46]:
```
weekly_group
```

Out[46]:

weekday	new_cases	new_tests	new_deaths
0	2258.0	21095.0	451.0
1	2256.0	28067.5	429.0
2	2236.0	28095.0	451.0
3	2250.0	28195.0	354.0
5	1256.0	18095.0	804.0

In [206]:
```
fili
```

Out[206]:

	date	new_cases	new_deaths	new_tests	year	month	day	weekday	location
0	2019-12-31	0.0	0.0	NaN	2019	12	31	1	italy
1	2020-01-01	0.0	0.0	NaN	2020	1	1	2	italy
2	2020-01-02	0.0	0.0	NaN	2020	1	2	3	italy
3	2020-01-03	0.0	0.0	NaN	2020	1	3	4	italy
4	2020-01-04	0.0	0.0	NaN	2020	1	4	5	italy
...	...	...	...	...	...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0	2020	8	30	6	italy
244	2020-08-31	1365.0	4.0	42583.0	2020	8	31	0	italy
245	2020-09-01	996.0	6.0	54395.0	2020	9	1	1	italy
246	2020-09-02	975.0	8.0	NaN	2020	9	2	2	italy
247	2020-09-03	1326.0	6.0	NaN	2020	9	3	3	italy

248 rows × 9 columns

In [50]:
```
result = file[["date", "new_cases", "new_deaths",]]
```

In [51]:
```
result
```

Out[51]:

	date	new_cases	new_deaths
0	2020-04-21	2256.0	454.0
1	2020-03-11	2236.0	451.0
2	2020-05-21	2250.0	354.0
3	2020-04-21	2256.0	404.0
4	2020-05-23	1256.0	804.0
5	2020-04-20	2258.0	451.0

In [52]:
```
final_resultdf = result.to_csv("./glo/finals.csv", index=False)
```

In [ ]:
```

```

In [53]:
```
fiile = pd.read_csv("./glo/covid.txt")
```

In [54]:
```
fiile
```

Out[54]:

	date	new_cases	new_deaths	new_tests
0	2020-04-21	2256.0	454.0	28045.0
1	2020-03-11	2236.0	451.0	28095.0
2	2020-05-21	2250.0	354.0	28195.0
3	2020-04-21	2256.0	404.0	28090.0
4	2020-05-23	1256.0	804.0	18095.0
5	2020-04-20	2258.0	451.0	21095.0

In [55]:
```
fiile.set_index("date", inplace=True)
```

In [56]:
```
fiile
```

Out[56]:

date	new_cases	new_deaths	new_tests
2020-04-21	2256.0	454.0	28045.0
2020-03-11	2236.0	451.0	28095.0
2020-05-21	2250.0	354.0	28195.0
2020-04-21	2256.0	404.0	28090.0
2020-05-23	1256.0	804.0	18095.0
2020-04-20	2258.0	451.0	21095.0

In [58]:
```
plt.plot(fiile.new_cases);
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2404\956193119.py in <module>
----> 1 plt.plot(fiile.new_cases);

NameError: name 'plt' is not defined
```

In [59]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [60]:
```
fiile.new_cases.plot();
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2404\1699815248.py in <module>
----> 1 fiile.new_cases.plot();

NameError: name 'fiile' is not defined
```

In [57]:
```
pip install matplotlib
```

```
Collecting matplotlibNote: you may need to restart the kernel to use updated pa
ckages.

  Downloading matplotlib-3.5.2-cp37-cp37m-win_amd64.whl (7.2 MB)
Requirement already satisfied: python-dateutil>=2.7 in c:\new\envs\snakes\lib\s
ite-packages (from matplotlib) (2.8.2)
Requirement already satisfied: pyparsing>=2.2.1 in c:\new\envs\snakes\lib\site-
packages (from matplotlib) (3.0.4)
Requirement already satisfied: numpy>=1.17 in c:\new\envs\snakes\lib\site-packa
ges (from matplotlib) (1.21.6)
Collecting cycler>=0.10
  Downloading cycler-0.11.0-py3-none-any.whl (6.4 kB)
Requirement already satisfied: packaging>=20.0 in c:\new\envs\snakes\lib\site-p
ackages (from matplotlib) (21.3)
Collecting kiwisolver>=1.0.1
  Downloading kiwisolver-1.4.2-cp37-cp37m-win_amd64.whl (54 kB)
Collecting fonttools>=4.22.0
  Downloading fonttools-4.33.3-py3-none-any.whl (930 kB)
Collecting pillow>=6.2.0
  Downloading Pillow-9.1.1-cp37-cp37m-win_amd64.whl (3.3 MB)
Requirement already satisfied: typing-extensions in c:\new\envs\snakes\lib\site
-packages (from kiwisolver>=1.0.1->matplotlib) (4.1.1)
Requirement already satisfied: six>=1.5 in c:\new\envs\snakes\lib\site-packages
(from python-dateutil>=2.7->matplotlib) (1.16.0)
Installing collected packages: pillow, kiwisolver, fonttools, cycler, matplotli
b
Successfully installed cycler-0.11.0 fonttools-4.33.3 kiwisolver-1.4.2 matplotl
ib-3.5.2 pillow-9.1.1
```

In [2]:
```
merged_df.new_cases.plot()
merged_df.new_deaths.plot();
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2404\1157960002.py in <module>
----> 1 merged_df.new_cases.plot()
      2 merged_df.new_deaths.plot();

NameError: name 'merged_df' is not defined
```

In [1]:
```
variable = merged_df.new_deaths
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2404\3761457147.py in <module>
----> 1 variable = merged_df.new_deaths

NameError: name 'merged_df' is not defined
```

In [221]:
```
merged_df.set_index("date", inplace=True)
```

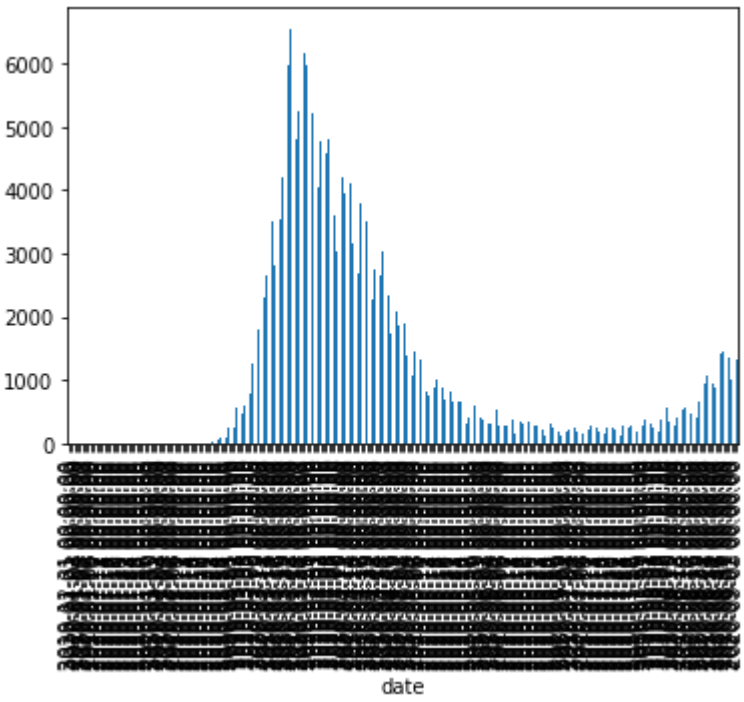
In [61]:
```
merged_df.loc["2020-01-03"]
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_2404\1600618223.py in <module>
----> 1 merged_df.loc["2020-01-03"]

NameError: name 'merged_df' is not defined
```

In [232]:
```
merged_df.new_cases.plot(kind="bar")
```

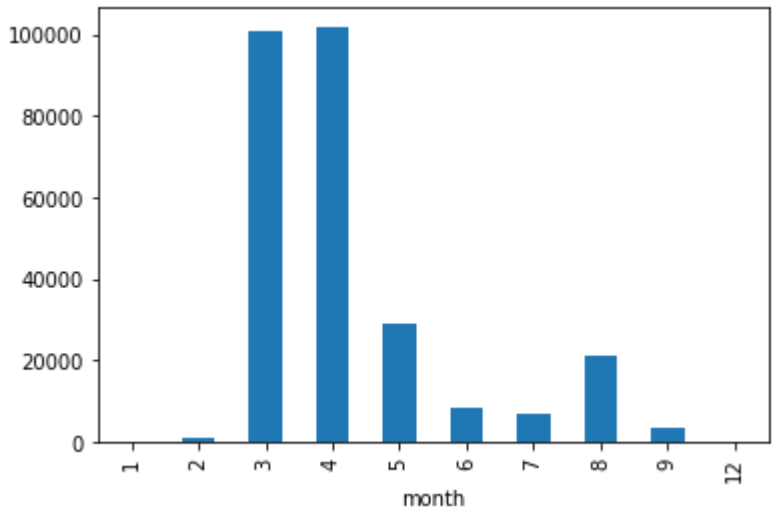
Out[232]: <AxesSubplot:xlabel='date'>



In [240]:
```
monthlyy_group = merged_df.groupby("month")["new_cases"].sum()
```

In [245]:
```
monthlyy_group.plot(kind="bar")
```

Out[245]: <AxesSubplot:xlabel='month'>



In [246]:
```
monthlyy_group
```

Out[246]:
```
month
1          3.00
2        885.00
3     100851.00
4     101852.00
5      29073.00
6       8206.75
7       6722.00
8      21060.00
9       3297.00
12         0.00
Name: new_cases, dtype: float64
```

In [247]:
```
merged_df
```

Out[247]:

date	new_cases	new_deaths	new_tests	year	month	day	weekday	location	continent	popul
2019-12-31	0.0	0.0	NaN	2019	12	31	1	Italy	Europe	60461
2020-01-01	0.0	0.0	NaN	2020	1	1	2	Italy	Europe	60461
2020-01-02	0.0	0.0	NaN	2020	1	2	3	Italy	Europe	60461
2020-01-03	0.0	0.0	NaN	2020	1	3	4	Italy	Europe	60461
2020-01-04	0.0	0.0	NaN	2020	1	4	5	Italy	Europe	60461
...	...	...	...	...	...	...	...	...	...	
2020-08-30	1444.0	1.0	53541.0	2020	8	30	6	Italy	Europe	60461
2020-08-31	1365.0	4.0	42583.0	2020	8	31	0	Italy	Europe	60461
2020-09-01	996.0	6.0	54395.0	2020	9	1	1	Italy	Europe	60461
2020-09-02	975.0	8.0	NaN	2020	9	2	2	Italy	Europe	60461
2020-09-03	1326.0	6.0	NaN	2020	9	3	3	Italy	Europe	60461

248 rows × 16 columns

In [248]:
```
locations_df
```

Out[248]:

	location	continent	population	life_expectancy	hospital_beds_per_thousand	gdp_per_ca
0	Afghanistan	Asia	3.892834e+07	64.83	0.500	1803
1	Albania	Europe	2.877800e+06	78.57	2.890	11803
2	Algeria	Africa	4.385104e+07	76.88	1.900	13913
3	Andorra	Europe	7.726500e+04	83.73	NaN	
4	Angola	Africa	3.286627e+07	61.15	NaN	5819
...	...	...	...	...	...	
207	Yemen	Asia	2.982597e+07	66.12	0.700	1479
208	Zambia	Africa	1.838396e+07	63.89	2.000	3689
209	Zimbabwe	Africa	1.486293e+07	61.49	1.700	1899
210	World	NaN	7.794799e+09	72.58	2.705	15469
211	International	NaN	NaN	NaN	NaN	

212 rows × 6 columns

In [260]:
```
locations_df["population"].sum()
```

Out[260]: 15552778824.0

In [271]:
```
locations_df.life_expect_product = locations_df.population * locations_df.life_ex
```

In [272]:
```
sum_life_exp = locations_df.life_expect_product.sum()
```

In [273]:
```
sum_life_exp
```

Out[273]: 1129919619931.06

In [276]:
```
sum_popu = locations_df.population.sum()
```

In [277]:
```
sum_popu
```

Out[277]: 15552778824.0

In [278]:
```
ove_lif_exp = sum_life_exp / sum_popu
```

In [279]:
```
ove_lif_exp
```

Out[279]: 72.65065829827428

In [280]:
```
locations_df.sort_values("population", ascending=False).head(10)
```

Out[280]:

	location	continent	population	life_expectancy	hospital_beds_per_thousand	gdp_per_ca
210	World	NaN	7.794799e+09	72.58	2.705	15469.
41	China	Asia	1.439324e+09	76.91	4.340	15308.
90	India	Asia	1.380004e+09	69.66	0.530	6426.
199	United States	North America	3.310026e+08	78.86	2.770	54225.
91	Indonesia	Asia	2.735236e+08	71.72	1.040	11188.
145	Pakistan	Asia	2.208923e+08	67.27	0.600	5034.
27	Brazil	South America	2.125594e+08	75.88	2.200	14103.
141	Nigeria	Africa	2.061396e+08	54.69	NaN	5338.
15	Bangladesh	Asia	1.646894e+08	72.59	0.800	3523.
157	Russia	Europe	1.459345e+08	72.58	8.050	24765.

In [286]:
```
locations_df.continent.isna
```

Out[286]:
```
<bound method Series.isna of 0        Asia
1      Europe
2      Africa
3      Europe
4      Africa
        ...
207      Asia
208    Africa
209    Africa
210       NaN
211       NaN
Name: continent, Length: 212, dtype: object>
```

In [289]:
```
locations_df.sort_values("gdp_per_capita", ascending=True).tail(20)
```

Out[289]:

	location	continent	population	life_expectancy	hospital_beds_per_thousand	gdp_per_cap
69	French Polynesia	Oceania	280904.0	77.66	NaN	Na
75	Gibraltar	Europe	33691.0	79.93	NaN	Na
77	Greenland	North America	56772.0	71.70	NaN	Na
79	Guam	Oceania	168783.0	80.07	NaN	Na
81	Guernsey	Europe	67052.0	NaN	NaN	Na
95	Isle of Man	Europe	85032.0	81.40	NaN	Na
100	Jersey	Europe	101073.0	NaN	NaN	Na
113	Liechtenstein	Europe	38137.0	82.49	2.397	Na
127	Monaco	Europe	39244.0	86.75	13.800	Na
130	Montserrat	North America	4999.0	74.16	NaN	Na
137	New Caledonia	Oceania	285491.0	77.55	NaN	Na
142	Northern Mariana Islands	Oceania	57557.0	76.74	NaN	Na
173	Somalia	Africa	15893219.0	57.40	0.900	Na
184	Syria	Asia	17500657.0	72.70	1.500	Na
185	Taiwan	Asia	23816775.0	80.46	NaN	Na
194	Turks and Caicos Islands	North America	38718.0	80.22	NaN	Na
200	United States Virgin Islands	North America	104423.0	80.58	NaN	Na
203	Vatican	Europe	809.0	75.12	NaN	Na
206	Western Sahara	Africa	597330.0	70.26	NaN	Na
211	International	NaN	NaN	NaN	NaN	Na

In [ ]:
```
```