ML_project_final

December 24, 2022

[ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import os
sns.set(style='darkgrid', context='notebook', color_codes=True)
```

[ ]:
```
import os
from six.moves import urllib

data_files = ["canberra-data.csv" , "london-data.csv" , "los-angeles-data.csv" ,␣
↪"mumbai-data.csv" , "new-york-data.csv"]

for fn_dst in data_files:
    fn_src = 'https://raw.githubusercontent.com/palakkeni5/ML-project/main/data/
↪%s' % fn_dst

    if os.path.isfile(fn_dst):
        print('File %s is already downloaded' % fn_dst)
    else:
        print('Downloaded %s' % fn_dst)
        urllib.request.urlretrieve(fn_src, fn_dst)
```

```
File canberra-data.csv is already downloaded
File london-data.csv is already downloaded
File los-angeles-data.csv is already downloaded
File mumbai-data.csv is already downloaded
File new-york-data.csv is already downloaded
```

[ ]:
```
names =[ "time","temperature_2m_max (°F)","temperature_2m_min (°F)",
        "apparent_temperature_max (°F)","apparent_temperature_min (°F)",
        "sunrise (iso8601)","sunset (iso8601)","shortwave_radiation_sum (MJ/m²)",
        "precipitation_sum (mm)","rain_sum (mm)","snowfall_sum␣
↪(cm)","precipitation_hours (h)",
        "windspeed_10m_max (m/s)", "windgusts_10m_max (m/
↪s)","winddirection_10m_dominant (°)","et0_fao_evapotranspiration (mm)"
]
```

[ ]:
```
def read_data(file_name):
    df=pd.read_csv( file_name,header=0 ,index_col="time" , parse_dates=["time" ,"sunrise (iso8601)","sunset (iso8601)"])
    df.dropna(inplace = True  )
    df["sunrise (iso8601)"] = df["sunrise (iso8601)"].dt.time.apply(lambda x:int( x.strftime("%H%M%S")) )
    df["sunset (iso8601)"] = df["sunset (iso8601)"].dt.time.apply(lambda x: int(x.strftime("%H%M%S")) )
    return df
```

[ ]:
```
data_all = []

for file_name in data_files:
    df = read_data(file_name)
    print(file_name)
    display(df)
    data_all.append(df)
```

canberra-data.csv

time	temperature_2m_max (°F)	temperature_2m_min (°F)	\
2002-12-02	88.7	58.6	
2002-12-03	89.4	65.5	
2002-12-04	74.1	48.9	
2002-12-05	66.1	41.2	
2002-12-06	70.9	37.4	
...	...	...	
2022-11-22	64.2	48.9	
2022-11-23	70.2	39.6	
2022-11-24	72.6	46.7	
2022-11-25	72.1	49.7	
2022-11-26	72.2	51.2	

time	apparent_temperature_max (°F)	apparent_temperature_min (°F)	\
2002-12-02	89.4	57.6	
2002-12-03	88.3	62.1	
2002-12-04	64.3	40.6	
2002-12-05	56.2	32.0	
2002-12-06	71.6	32.1	
...	...	...	
2022-11-22	57.3	42.6	
2022-11-23	68.9	35.2	
2022-11-24	70.5	45.2	
2022-11-25	71.7	48.1	
2022-11-26	72.1	48.0	

time	sunrise (iso8601)	sunset (iso8601)
2002-12-02	134000	40500
2002-12-03	134000	40600
2002-12-04	134000	40700
2002-12-05	134000	40800
2002-12-06	134000	40900
...	...	...
2022-11-22	134300	35600
2022-11-23	134200	35700
2022-11-24	134200	35800
2022-11-25	134200	35900
2022-11-26	134100	40000

time	shortwave_radiation_sum (MJ/m²)	precipitation_sum (mm)
2002-12-02	32.01	0.0
2002-12-03	31.97	0.0
2002-12-04	25.96	0.0
2002-12-05	23.92	4.1
2002-12-06	30.77	0.0
...	...	...
2022-11-22	15.87	0.4
2022-11-23	25.65	0.0
2022-11-24	26.08	0.0
2022-11-25	27.22	0.0
2022-11-26	26.58	0.7

time	rain_sum (mm)	snowfall_sum (cm)	precipitation_hours (h)
2002-12-02	0.0	0.0	0.0
2002-12-03	0.0	0.0	0.0
2002-12-04	0.0	0.0	0.0
2002-12-05	4.1	0.0	4.0
2002-12-06	0.0	0.0	0.0
...	...	...	...
2022-11-22	0.4	0.0	3.0
2022-11-23	0.0	0.0	0.0
2022-11-24	0.0	0.0	0.0
2022-11-25	0.0	0.0	0.0
2022-11-26	0.7	0.0	2.0

time	windspeed_10m_max (m/s)	windgusts_10m_max (m/s)
2002-12-02	4.61	10.7
2002-12-03	8.95	19.3
2002-12-04	10.49	20.6

```
2002-12-05                    9.18                    22.5
2002-12-06                    5.82                    13.8
...                            ...                     ...
2022-11-22                    7.03                    14.3
2022-11-23                    6.40                    13.6
2022-11-24                    5.92                    12.1
2022-11-25                    6.00                    12.6
2022-11-26                    6.40                    18.9

            winddirection_10m_dominant (°)  et0_fao_evapotranspiration (mm)
time
2002-12-02                           313.0                             7.08
2002-12-03                           298.0                             7.24
2002-12-04                           283.0                             5.96
2002-12-05                           268.0                             4.14
2002-12-06                           261.0                             5.50
...                                    ...                              ...
2022-11-22                           165.0                             3.00
2022-11-23                           159.0                             4.39
2022-11-24                           150.0                             4.64
2022-11-25                           142.0                             4.88
2022-11-26                           131.0                             4.77

[7255 rows x 15 columns]
```

london-data.csv

```
            temperature_2m_max (°F)  temperature_2m_min (°F)  \
time
2002-12-02                    50.1                    45.5
2002-12-03                    50.0                    42.0
2002-12-04                    49.6                    40.0
2002-12-05                    46.3                    39.0
2002-12-06                    45.7                    42.1
...                            ...                     ...
2022-11-22                    49.4                    40.3
2022-11-23                    53.4                    42.1
2022-11-24                    54.7                    43.5
2022-11-25                    53.8                    44.1
2022-11-26                    53.9                    44.8

            apparent_temperature_max (°F)  apparent_temperature_min (°F)  \
time
2002-12-02                           42.2                           39.3
2002-12-03                           47.8                           38.1
2002-12-04                           45.1                           31.8
2002-12-05                           39.7                           31.2
2002-12-06                           39.2                           34.5
...                                   ...                            ...
```

2022-11-22	43.8	34.1
2022-11-23	46.2	36.0
2022-11-24	46.5	37.0
2022-11-25	48.3	38.3
2022-11-26	47.7	40.6

time	sunrise (iso8601)	sunset (iso8601)	\
2002-12-02	24300	105600	
2002-12-03	24400	105600	
2002-12-04	24600	105500	
2002-12-05	24700	105500	
2002-12-06	24800	105400	
...	...	...	
2022-11-22	22800	110400	
2022-11-23	23000	110300	
2022-11-24	23100	110200	
2022-11-25	23300	110100	
2022-11-26	23400	110000	

time	shortwave_radiation_sum (MJ/m²)	precipitation_sum (mm)	\
2002-12-02	3.72	2.1	
2002-12-03	2.48	0.3	
2002-12-04	0.93	4.4	
2002-12-05	4.01	0.0	
2002-12-06	1.41	0.0	
...	...	...	
2022-11-22	3.85	0.0	
2022-11-23	3.17	6.5	
2022-11-24	3.06	4.3	
2022-11-25	4.86	0.0	
2022-11-26	2.90	3.6	

time	rain_sum (mm)	snowfall_sum (cm)	precipitation_hours (h)	\
2002-12-02	2.1	0.0	6.0	
2002-12-03	0.3	0.0	2.0	
2002-12-04	4.4	0.0	8.0	
2002-12-05	0.0	0.0	0.0	
2002-12-06	0.0	0.0	0.0	
...	...	...	...	
2022-11-22	0.0	0.0	0.0	
2022-11-23	6.5	0.0	6.0	
2022-11-24	4.3	0.0	5.0	
2022-11-25	0.0	0.0	0.0	
2022-11-26	3.6	0.0	11.0	

```
            windspeed_10m_max (m/s)  windgusts_10m_max (m/s)  \
time
2002-12-02                     6.96                     14.4
2002-12-03                     2.85                      5.8
2002-12-04                     5.62                     11.2
2002-12-05                     5.19                      9.9
2002-12-06                     5.91                     11.2
...                             ...                      ...
2022-11-22                     5.02                      9.9
2022-11-23                     7.84                     15.0
2022-11-24                     9.68                     19.7
2022-11-25                     5.01                      9.5
2022-11-26                     8.41                     16.2

            winddirection_10m_dominant (°)  et0_fao_evapotranspiration (mm)
time
2002-12-02                           260.0                             1.05
2002-12-03                           230.0                             0.31
2002-12-04                           295.0                             0.36
2002-12-05                           353.0                             0.55
2002-12-06                            29.0                             0.46
...                                    ...                              ...
2022-11-22                           208.0                             0.52
2022-11-23                           229.0                             0.87
2022-11-24                           244.0                             0.71
2022-11-25                           218.0                             0.66
2022-11-26                           263.0                             0.61

[7292 rows x 15 columns]
```

los-angeles-data.csv

```
            temperature_2m_max (°F)  temperature_2m_min (°F)  \
time
2002-12-02                     68.7                     47.8
2002-12-03                     66.4                     48.5
2002-12-04                     69.2                     47.1
2002-12-05                     70.4                     46.4
2002-12-06                     67.4                     47.6
...                             ...                      ...
2022-11-22                     75.3                     47.3
2022-11-23                     73.2                     45.8
2022-11-24                     79.6                     47.5
2022-11-25                     76.4                     48.0
2022-11-26                     71.0                     46.0

            apparent_temperature_max (°F)  apparent_temperature_min (°F)  \
time
2002-12-02                           65.8                           44.5
```

2002-12-03	63.8	45.7
2002-12-04	66.0	44.2
2002-12-05	67.4	42.0
2002-12-06	64.1	43.6
...	...	...
2022-11-22	69.3	40.1
2022-11-23	67.9	39.4
2022-11-24	71.2	41.1
2022-11-25	69.8	40.8
2022-11-26	66.0	39.7

time	sunrise (iso8601)	sunset (iso8601)
2002-12-02	93900	194500
2002-12-03	94000	194500
2002-12-04	94100	194400
2002-12-05	94200	194400
2002-12-06	94300	194400
...	...	...
2022-11-22	93100	194700
2022-11-23	93100	194700
2022-11-24	93200	194600
2022-11-25	93300	194600
2022-11-26	93400	194600

time	shortwave_radiation_sum (MJ/m²)	precipitation_sum (mm)
2002-12-02	12.30	0.0
2002-12-03	11.91	0.0
2002-12-04	12.34	0.0
2002-12-05	10.06	0.0
2002-12-06	11.96	0.0
...	...	...
2022-11-22	13.62	0.0
2022-11-23	13.72	0.0
2022-11-24	13.77	0.0
2022-11-25	13.42	0.0
2022-11-26	13.17	0.0

time	rain_sum (mm)	snowfall_sum (cm)	precipitation_hours (h)
2002-12-02	0.0	0.0	0.0
2002-12-03	0.0	0.0	0.0
2002-12-04	0.0	0.0	0.0
2002-12-05	0.0	0.0	0.0
2002-12-06	0.0	0.0	0.0
...	...	...	...
2022-11-22	0.0	0.0	0.0

```
2022-11-23           0.0               0.0                    0.0
2022-11-24           0.0               0.0                    0.0
2022-11-25           0.0               0.0                    0.0
2022-11-26           0.0               0.0                    0.0

            windspeed_10m_max (m/s)  windgusts_10m_max (m/s)  \
time
2002-12-02                     2.25                      5.0
2002-12-03                     2.55                      6.0
2002-12-04                     2.15                      5.2
2002-12-05                     2.47                      4.8
2002-12-06                     2.82                      6.7
...                             ...                      ...
2022-11-22                     2.48                      5.9
2022-11-23                     2.24                      5.1
2022-11-24                     3.11                      7.4
2022-11-25                     2.97                      5.8
2022-11-26                     2.69                      6.5

            winddirection_10m_dominant (°)  et0_fao_evapotranspiration (mm)
time
2002-12-02                             3.0                             2.05
2002-12-03                           136.0                             1.93
2002-12-04                           352.0                             2.02
2002-12-05                            23.0                             1.82
2002-12-06                            56.0                             1.98
...                                    ...                              ...
2022-11-22                            81.0                             3.12
2022-11-23                            36.0                             2.78
2022-11-24                            56.0                             3.44
2022-11-25                            96.0                             3.37
2022-11-26                           333.0                             2.63

[7170 rows x 15 columns]
```

mumbai-data.csv

```
            temperature_2m_max (°F)  temperature_2m_min (°F)  \
time
2002-12-02                     89.0                     69.3
2002-12-03                     87.2                     69.7
2002-12-04                     87.3                     69.2
2002-12-05                     88.9                     71.3
2002-12-06                     89.2                     71.1
...                             ...                      ...
2022-11-22                     86.6                     69.7
2022-11-23                     87.6                     69.7
2022-11-24                     88.2                     71.3
2022-11-25                     88.7                     75.7
```

```
2022-11-26                    88.0                    74.3

            apparent_temperature_max (°F)  apparent_temperature_min (°F)  \
time
2002-12-02                           92.9                           74.6
2002-12-03                           91.8                           75.5
2002-12-04                           90.3                           73.3
2002-12-05                           94.3                           76.0
2002-12-06                           95.1                           75.9
...                                   ...                            ...
2022-11-22                           90.3                           71.7
2022-11-23                           93.5                           73.1
2022-11-24                           94.7                           75.2
2022-11-25                           96.9                           83.0
2022-11-26                           97.0                           82.5

            sunrise (iso8601)  sunset (iso8601)  \
time
2002-12-02             202500             73000
2002-12-03             202500             73100
2002-12-04             202600             73100
2002-12-05             202600             73100
2002-12-06             202700             73100
...                       ...               ...
2022-11-22             201800             73000
2022-11-23             201900             73000
2022-11-24             202000             73000
2022-11-25             202000             73000
2022-11-26             202100             73000

            shortwave_radiation_sum (MJ/m²)  precipitation_sum (mm)  \
time
2002-12-02                            18.45                     0.0
2002-12-03                            18.21                     0.0
2002-12-04                            17.80                     0.0
2002-12-05                            17.73                     0.0
2002-12-06                            17.84                     0.0
...                                     ...                     ...
2022-11-22                            18.35                     0.0
2022-11-23                            17.89                     0.0
2022-11-24                            17.82                     0.0
2022-11-25                            16.63                     0.0
2022-11-26                            16.38                     0.0

            rain_sum (mm)  snowfall_sum (cm)  precipitation_hours (h)  \
time
2002-12-02            0.0                0.0                      0.0
2002-12-03            0.0                0.0                      0.0
```

```
2002-12-04           0.0               0.0                    0.0
2002-12-05           0.0               0.0                    0.0
2002-12-06           0.0               0.0                    0.0
...                  ...               ...                    ...
2022-11-22           0.0               0.0                    0.0
2022-11-23           0.0               0.0                    0.0
2022-11-24           0.0               0.0                    0.0
2022-11-25           0.0               0.0                    0.0
2022-11-26           0.0               0.0                    0.0

            windspeed_10m_max (m/s)  windgusts_10m_max (m/s)  \
time
2002-12-02                     3.85                      6.9
2002-12-03                     4.25                      7.3
2002-12-04                     3.22                      5.9
2002-12-05                     3.86                      7.0
2002-12-06                     4.02                      7.1
...                             ...                      ...
2022-11-22                     3.89                      5.9
2022-11-23                     4.19                      6.7
2022-11-24                     3.92                      6.1
2022-11-25                     2.61                      4.3
2022-11-26                     3.54                      6.2

            winddirection_10m_dominant (°)  et0_fao_evapotranspiration (mm)
time
2002-12-02                           306.0                             4.01
2002-12-03                           323.0                             3.93
2002-12-04                           345.0                             4.00
2002-12-05                           358.0                             4.02
2002-12-06                            18.0                             4.01
...                                    ...                              ...
2022-11-22                            79.0                             4.49
2022-11-23                            89.0                             4.28
2022-11-24                            88.0                             4.16
2022-11-25                            91.0                             3.65
2022-11-26                           140.0                             3.49

[7265 rows x 15 columns]
```

new-york-data.csv

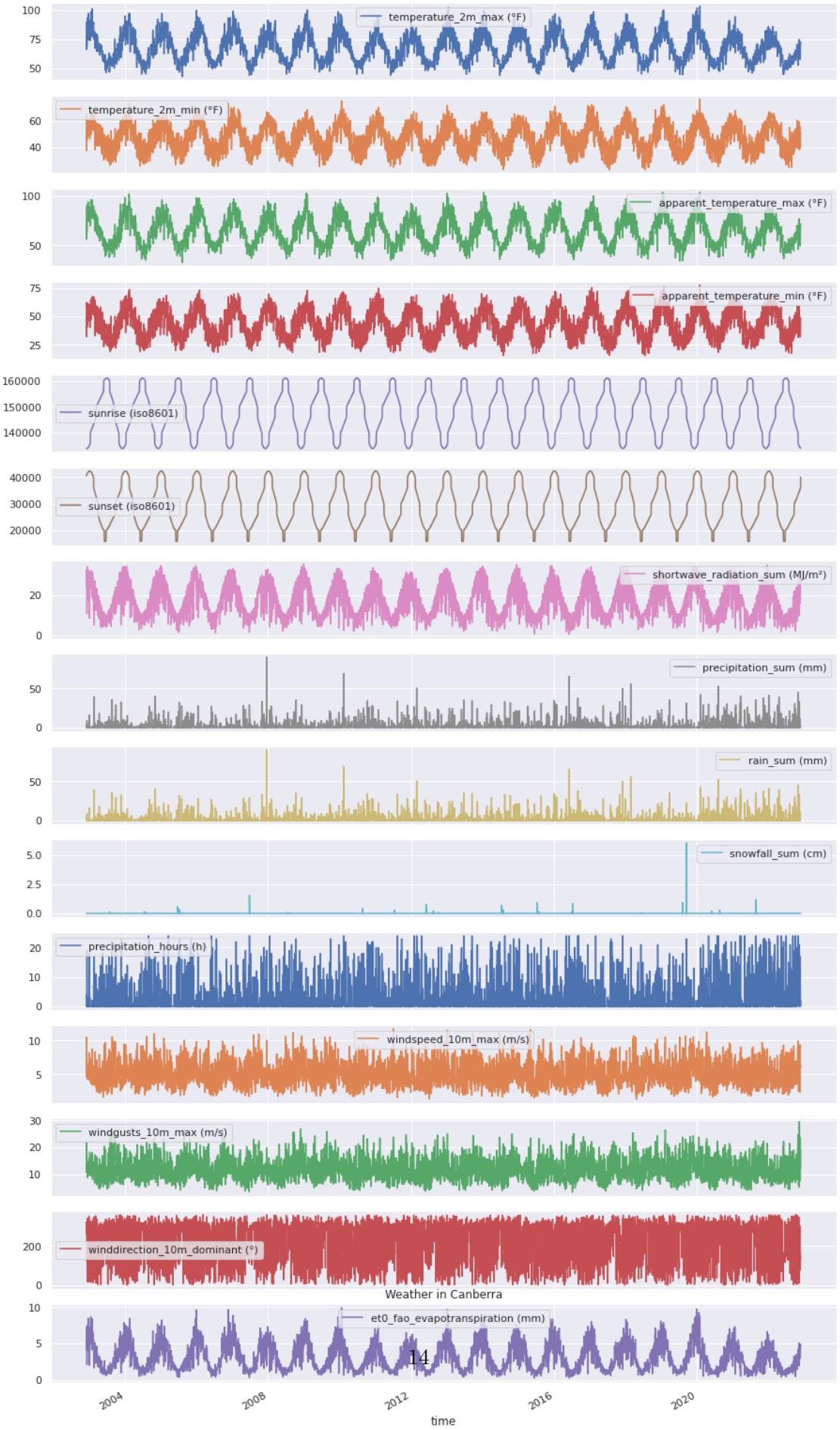
```
            temperature_2m_max (°F)  temperature_2m_min (°F)  \
time
2002-12-02                     39.2                     29.4
2002-12-03                     35.6                     21.0
2002-12-04                     31.6                     18.9
2002-12-05                     30.4                     26.1
2002-12-06                     32.7                     22.9
```

```
...                                 ...                                  ...
2022-11-22                         49.9                                 33.0
2022-11-23                         53.6                                 34.9
2022-11-24                         51.4                                 33.5
2022-11-25                         53.4                                 40.6
2022-11-26                         54.6                                 40.4

            apparent_temperature_max (°F)  apparent_temperature_min (°F)  \
time
2002-12-02                           32.7                           19.8
2002-12-03                           27.4                            8.4
2002-12-04                           23.9                            9.9
2002-12-05                           24.2                           16.6
2002-12-06                           24.7                           15.0
...                                   ...                            ...
2022-11-22                           43.5                           25.8
2022-11-23                           47.3                           28.1
2022-11-24                           46.4                           27.6
2022-11-25                           48.4                           34.9
2022-11-26                           48.5                           33.1

            sunrise (iso8601)  sunset (iso8601)  \
time
2002-12-02              70000            163000
2002-12-03              70100            163000
2002-12-04              70200            163000
2002-12-05              70300            163000
2002-12-06              70400            163000
...                       ...               ...
2022-11-22              64900            163400
2022-11-23              65000            163400
2022-11-24              65100            163300
2022-11-25              65300            163300
2022-11-26              65400            163200

            shortwave_radiation_sum (MJ/m²)  precipitation_sum (mm)  \
time
2002-12-02                             5.98                     0.0
2002-12-03                             9.90                     0.0
2002-12-04                             8.89                     0.0
2002-12-05                             0.93                    15.4
2002-12-06                             8.73                     0.0
...                                     ...                     ...
2022-11-22                            10.27                     0.0
2022-11-23                            10.09                     0.0
2022-11-24                             9.82                     0.0
2022-11-25                             3.51                     2.2
2022-11-26                             9.71                     0.0
```

```
            rain_sum (mm)  snowfall_sum (cm)  precipitation_hours (h)  \
time
2002-12-02            0.0               0.21                      0.0
2002-12-03            0.0               0.07                      0.0
2002-12-04            0.0               0.00                      0.0
2002-12-05            0.0              10.92                     16.0
2002-12-06            0.0               0.00                      0.0
...                   ...                ...                      ...
2022-11-22            0.0               0.00                      0.0
2022-11-23            0.0               0.00                      0.0
2022-11-24            0.0               0.00                      0.0
2022-11-25            2.2               0.00                      6.0
2022-11-26            0.0               0.00                      0.0

            windspeed_10m_max (m/s)  windgusts_10m_max (m/s)  \
time
2002-12-02                     4.56                      8.7
2002-12-03                     6.80                     14.1
2002-12-04                     4.20                      8.5
2002-12-05                     5.75                     10.4
2002-12-06                     3.41                      6.5
...                             ...                      ...
2022-11-22                     4.53                      9.1
2022-11-23                     3.29                      6.6
2022-11-24                     3.13                      5.4
2022-11-25                     5.34                     10.9
2022-11-26                     5.06                      9.4

            winddirection_10m_dominant (°)  et0_fao_evapotranspiration (mm)
time
2002-12-02                           235.0                             1.14
2002-12-03                           321.0                             1.25
2002-12-04                           295.0                             0.87
2002-12-05                            25.0                             0.16
2002-12-06                           299.0                             0.72
...                                    ...                              ...
2022-11-22                           205.0                             1.65
2022-11-23                           154.0                             1.62
2022-11-24                           306.0                             1.24
2022-11-25                           187.0                             1.07
2022-11-26                           172.0                             2.11

[7279 rows x 15 columns]
```

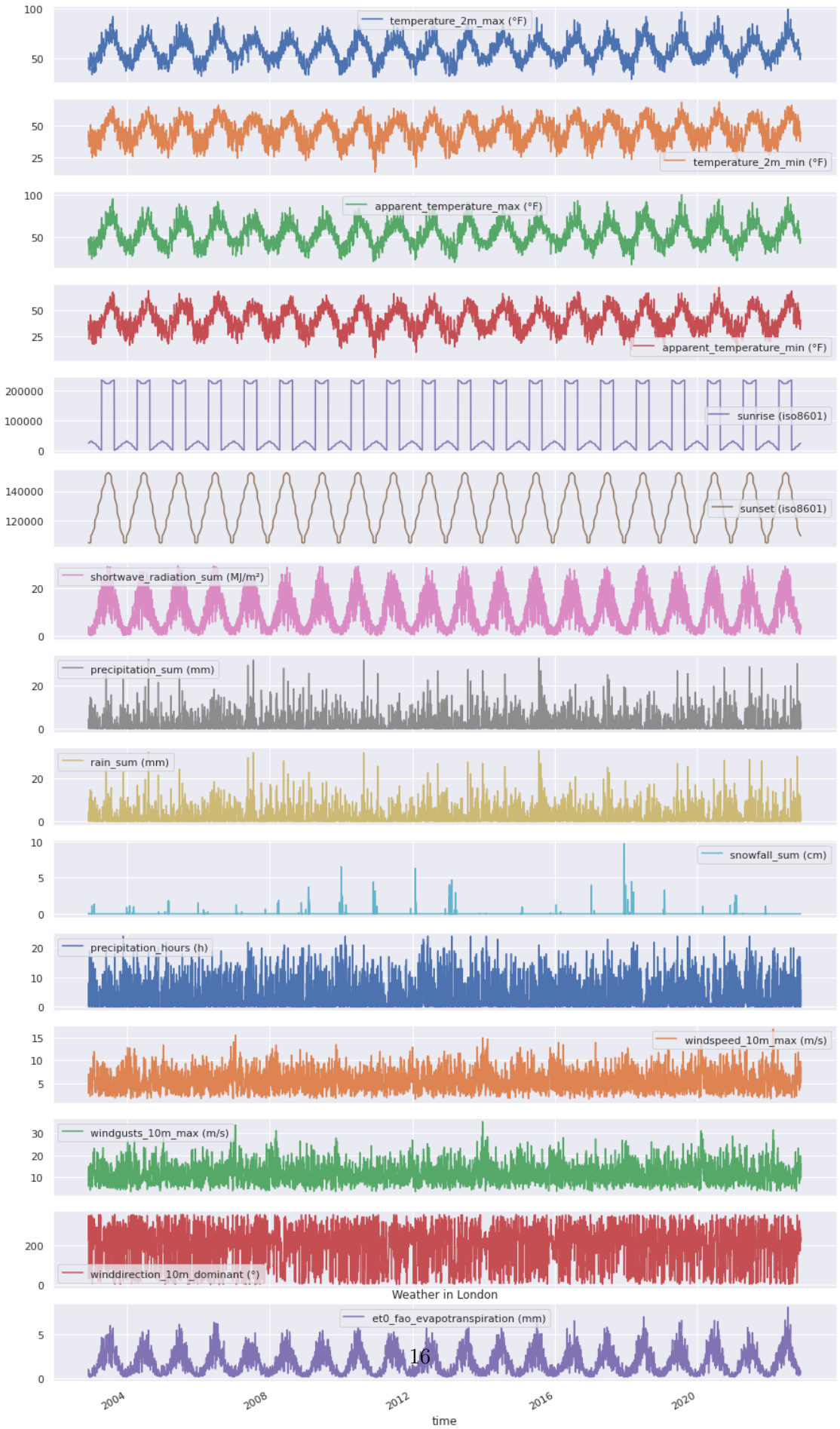
[ ]:
```
data_all[0].plot(subplots=True, figsize=(15,30))
plt.title('Weather in Canberra')
```

```
plt.show()
```



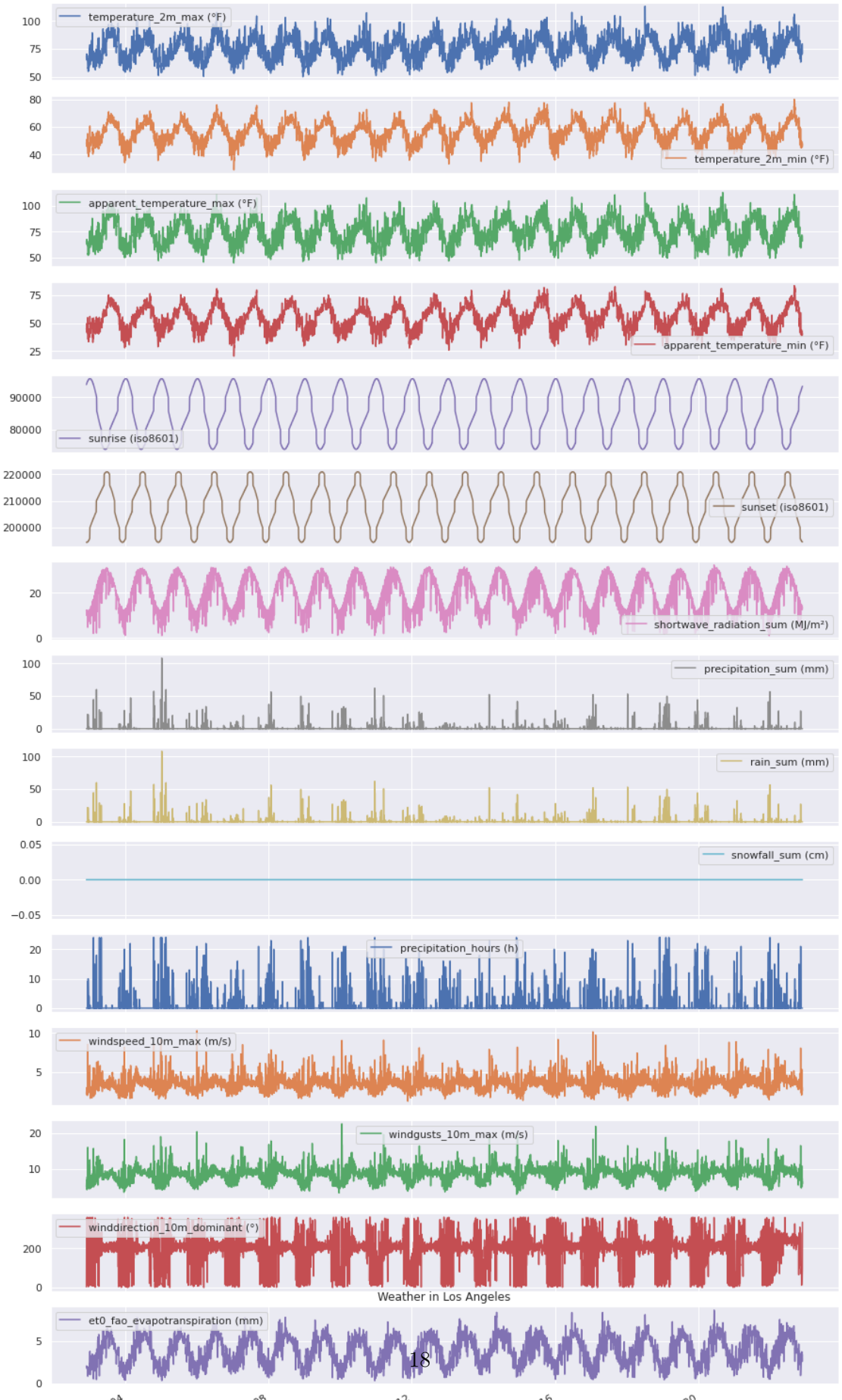
Weather in Canberra

[ ]:
```
data_all[1].plot(subplots=True, figsize=(15,30))
plt.title('Weather in London')
plt.show()
```



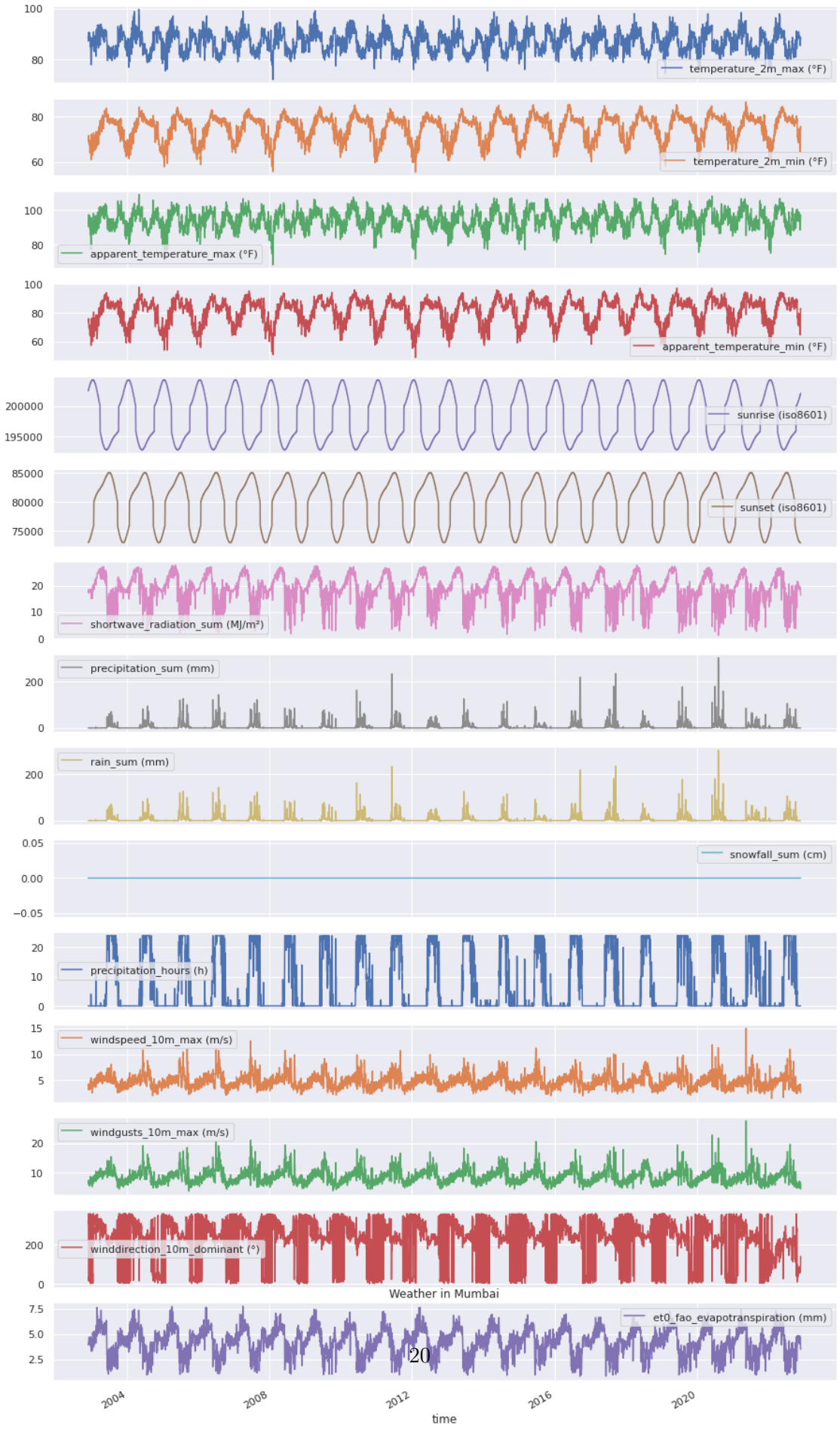
Weather in London

[ ]:
```
data_all[2].plot(subplots=True, figsize=(15,30))
plt.title('Weather in Los Angeles')
plt.show()
```

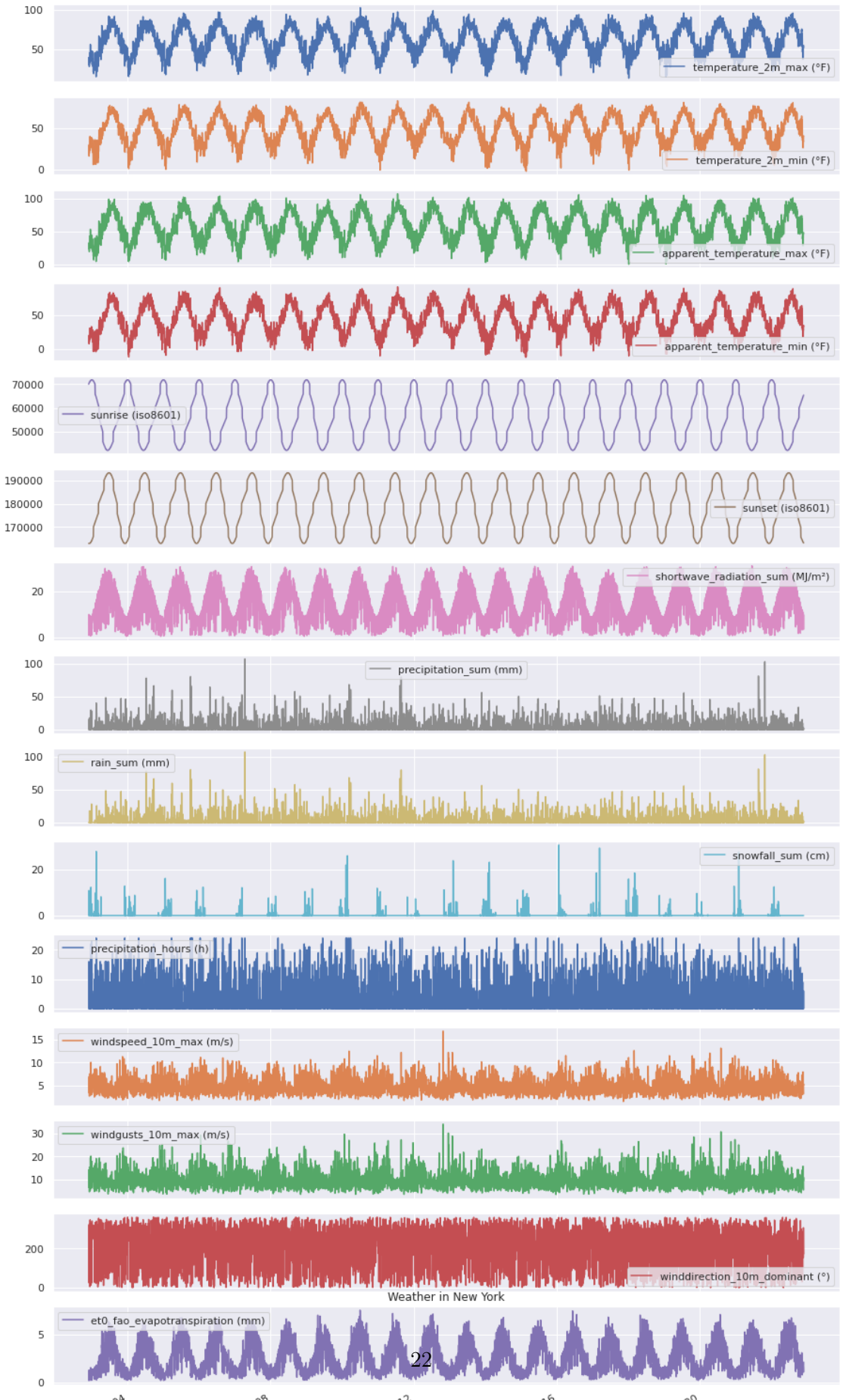


Weather in Los Angeles

[ ]:
```
data_all[3].plot(subplots=True, figsize=(15,30))
plt.title('Weather in Mumbai')
plt.show()
```


temperature_2m_max (°F)
temperature_2m_min (°F)
apparent_temperature_max (°F)
apparent_temperature_min (°F)
sunrise (iso8601)
sunset (iso8601)
shortwave_radiation_sum (MJ/m²)
precipitation_sum (mm)
rain_sum (mm)
snowfall_sum (cm)
precipitation_hours (h)
windspeed_10m_max (m/s)
windgusts_10m_max (m/s)
winddirection_10m_dominant (°)
et0_fao_evapotranspiration (mm)
Weather in Mumbai
2004
2008
2012
2016
2020
time

[ ]:
```
data_all[4].plot(subplots=True, figsize=(15,30))
plt.title('Weather in New York')
plt.show()
```



Weather in New York

##Augmented Dickey-Fuller test

Null hypothesis: Non Stationarity exists in the series.

Alternative Hypothesis: Stationarity exists in the series

[ ]:
```
#time series data testing
from statsmodels.tsa.stattools import adfuller


for index , data in enumerate(data_all):
  tests = pd.DataFrame(columns = ['Column' , 'ADF test statistic' ,'p-value',␣
 ↪'null hypothesis', 'Stationarity'])
  for col in data:
    adfuller_test = adfuller(data[col].tolist() , autolag='AIC')
    hyp = "Reject" if((adfuller_test[1] < 0.05)) else "Fail to reject"
    stationary = "Yes" if((adfuller_test[1] < 0.05)) else "No"
    tests.loc[len(tests.index)] =   [col , adfuller_test[0] , adfuller_test[1] ,␣
 ↪ hyp , stationary ]
  print("Weather data for " + data_files[index])
  display(tests)
```

```
Weather data for canberra-data.csv
```

	Column	ADF test statistic	p-value	\
0	temperature_2m_max (°F)	-5.839659	3.803649e-07	
1	temperature_2m_min (°F)	-5.195630	8.974518e-06	
2	apparent_temperature_max (°F)	-5.782878	5.077674e-07	
3	apparent_temperature_min (°F)	-5.285927	5.855415e-06	
4	sunrise (iso8601)	-8.935483	9.560010e-15	
5	sunset (iso8601)	-8.340064	3.193943e-13	
6	shortwave_radiation_sum (MJ/m²)	-4.997531	2.244418e-05	
7	precipitation_sum (mm)	-67.892419	0.000000e+00	
8	rain_sum (mm)	-67.874690	0.000000e+00	
9	snowfall_sum (cm)	-12.449381	3.603852e-23	
10	precipitation_hours (h)	-12.322070	6.716016e-23	
11	windspeed_10m_max (m/s)	-9.891155	3.555064e-17	
12	windgusts_10m_max (m/s)	-9.489868	3.677300e-16	
13	winddirection_10m_dominant (°)	-8.678142	4.359270e-14	
14	et0_fao_evapotranspiration (mm)	-5.226499	7.760443e-06	

	null hypothesis	Stationarity
0	Reject	Yes
1	Reject	Yes
2	Reject	Yes
3	Reject	Yes
4	Reject	Yes

```
5           Reject          Yes
6           Reject          Yes
7           Reject          Yes
8           Reject          Yes
9           Reject          Yes
10          Reject          Yes
11          Reject          Yes
12          Reject          Yes
13          Reject          Yes
14          Reject          Yes
```

```
Weather data for london-data.csv
```

```
                           Column  ADF test statistic       p-value  \
0           temperature_2m_max (°F)           -5.153924  1.091025e-05
1           temperature_2m_min (°F)           -5.755057  5.845840e-07
2     apparent_temperature_max (°F)           -5.187797  9.310717e-06
3     apparent_temperature_min (°F)           -5.351422  4.280882e-06
4                 sunrise (iso8601)           -7.296681  1.372143e-10
5                  sunset (iso8601)          -13.849524  7.047128e-26
6   shortwave_radiation_sum (MJ/m²)           -5.253814  6.820092e-06
7            precipitation_sum (mm)          -29.601754  0.000000e+00
8                    rain_sum (mm)          -29.477073  0.000000e+00
9                 snowfall_sum (cm)          -15.192134  5.989966e-28
10          precipitation_hours (h)          -27.738896  0.000000e+00
11          windspeed_10m_max (m/s)          -23.987650  0.000000e+00
12          windgusts_10m_max (m/s)          -26.118307  0.000000e+00
13   winddirection_10m_dominant (°)          -30.619206  0.000000e+00
14  et0_fao_evapotranspiration (mm)           -5.201491  8.730679e-06

   null hypothesis Stationarity
0           Reject          Yes
1           Reject          Yes
2           Reject          Yes
3           Reject          Yes
4           Reject          Yes
5           Reject          Yes
6           Reject          Yes
7           Reject          Yes
8           Reject          Yes
9           Reject          Yes
10          Reject          Yes
11          Reject          Yes
12          Reject          Yes
13          Reject          Yes
14          Reject          Yes
```

```
/usr/local/lib/python3.8/dist-
packages/statsmodels/regression/linear_model.py:903: RuntimeWarning: divide by
```

```
zero encountered in log
  llf = -nobs2*np.log(2*np.pi) - nobs2*np.log(ssr / nobs) - nobs2
```

```
Weather data for los-angeles-data.csv
```

	Column	ADF test statistic	p-value	\
0	temperature_2m_max (°F)	-6.558135	8.514004e-09	
1	temperature_2m_min (°F)	-5.542078	1.692781e-06	
2	apparent_temperature_max (°F)	-5.865957	3.325242e-07	
3	apparent_temperature_min (°F)	-5.597673	1.285935e-06	
4	sunrise (iso8601)	-9.321602	9.858460e-16	
5	sunset (iso8601)	-8.660061	4.849685e-14	
6	shortwave_radiation_sum (MJ/m²)	-5.695075	7.908368e-07	
7	precipitation_sum (mm)	-11.580113	2.956925e-21	
8	rain_sum (mm)	-11.580113	2.956925e-21	
9	snowfall_sum (cm)	NaN	NaN	
10	precipitation_hours (h)	-11.246268	1.756514e-20	
11	windspeed_10m_max (m/s)	-7.972888	2.750994e-12	
12	windgusts_10m_max (m/s)	-6.877341	1.463115e-09	
13	winddirection_10m_dominant (°)	-8.026312	2.012972e-12	
14	et0_fao_evapotranspiration (mm)	-5.568551	1.485449e-06	

	null hypothesis	Stationarity
0	Reject	Yes
1	Reject	Yes
2	Reject	Yes
3	Reject	Yes
4	Reject	Yes
5	Reject	Yes
6	Reject	Yes
7	Reject	Yes
8	Reject	Yes
9	Fail to reject	No
10	Reject	Yes
11	Reject	Yes
12	Reject	Yes
13	Reject	Yes
14	Reject	Yes

```
/usr/local/lib/python3.8/dist-
packages/statsmodels/regression/linear_model.py:903: RuntimeWarning: divide by
zero encountered in log
  llf = -nobs2*np.log(2*np.pi) - nobs2*np.log(ssr / nobs) - nobs2
```

```
Weather data for mumbai-data.csv
```

	Column	ADF test statistic	p-value	\
0	temperature_2m_max (°F)	-9.032084	5.410784e-15	
1	temperature_2m_min (°F)	-6.887496	1.382538e-09	
2	apparent_temperature_max (°F)	-9.477970	3.942466e-16	

```
3     apparent_temperature_min (°F)           -6.093609  1.022486e-07
4                 sunrise (iso8601)           -8.647487  5.222892e-14
5                  sunset (iso8601)           -9.293180  1.164864e-15
6   shortwave_radiation_sum (MJ/m²)           -8.163984  8.985797e-13
7            precipitation_sum (mm)           -7.549685  3.213387e-11
8                    rain_sum (mm)           -7.549685  3.213387e-11
9                 snowfall_sum (cm)                 NaN           NaN
10          precipitation_hours (h)           -6.972253  8.604879e-10
11          windspeed_10m_max (m/s)           -7.626191  2.065686e-11
12          windgusts_10m_max (m/s)           -6.790243  2.374828e-09
13   winddirection_10m_dominant (°)           -7.956640  3.024923e-12
14  et0_fao_evapotranspiration (mm)           -7.643523  1.868611e-11

   null hypothesis Stationarity
0           Reject          Yes
1           Reject          Yes
2           Reject          Yes
3           Reject          Yes
4           Reject          Yes
5           Reject          Yes
6           Reject          Yes
7           Reject          Yes
8           Reject          Yes
9   Fail to reject           No
10          Reject          Yes
11          Reject          Yes
12          Reject          Yes
13          Reject          Yes
14          Reject          Yes

Weather data for new-york-data.csv

                            Column  ADF test statistic       p-value  \
0           temperature_2m_max (°F)           -6.078060  1.109220e-07
1           temperature_2m_min (°F)           -5.986613  1.785943e-07
2     apparent_temperature_max (°F)           -6.090552  1.038998e-07
3     apparent_temperature_min (°F)           -6.052694  1.266437e-07
4                 sunrise (iso8601)           -9.518121  3.117150e-16
5                  sunset (iso8601)           -9.648556  1.455465e-16
6   shortwave_radiation_sum (MJ/m²)           -5.118980  1.283796e-05
7            precipitation_sum (mm)          -59.574888  0.000000e+00
8                    rain_sum (mm)          -47.460822  0.000000e+00
9                 snowfall_sum (cm)          -11.184499  2.454137e-20
10          precipitation_hours (h)          -58.672604  0.000000e+00
11          windspeed_10m_max (m/s)           -9.554247  2.523769e-16
12          windgusts_10m_max (m/s)          -11.480157  5.017373e-21
13   winddirection_10m_dominant (°)          -15.105355  7.833532e-28
14  et0_fao_evapotranspiration (mm)           -5.411109  3.209902e-06
```

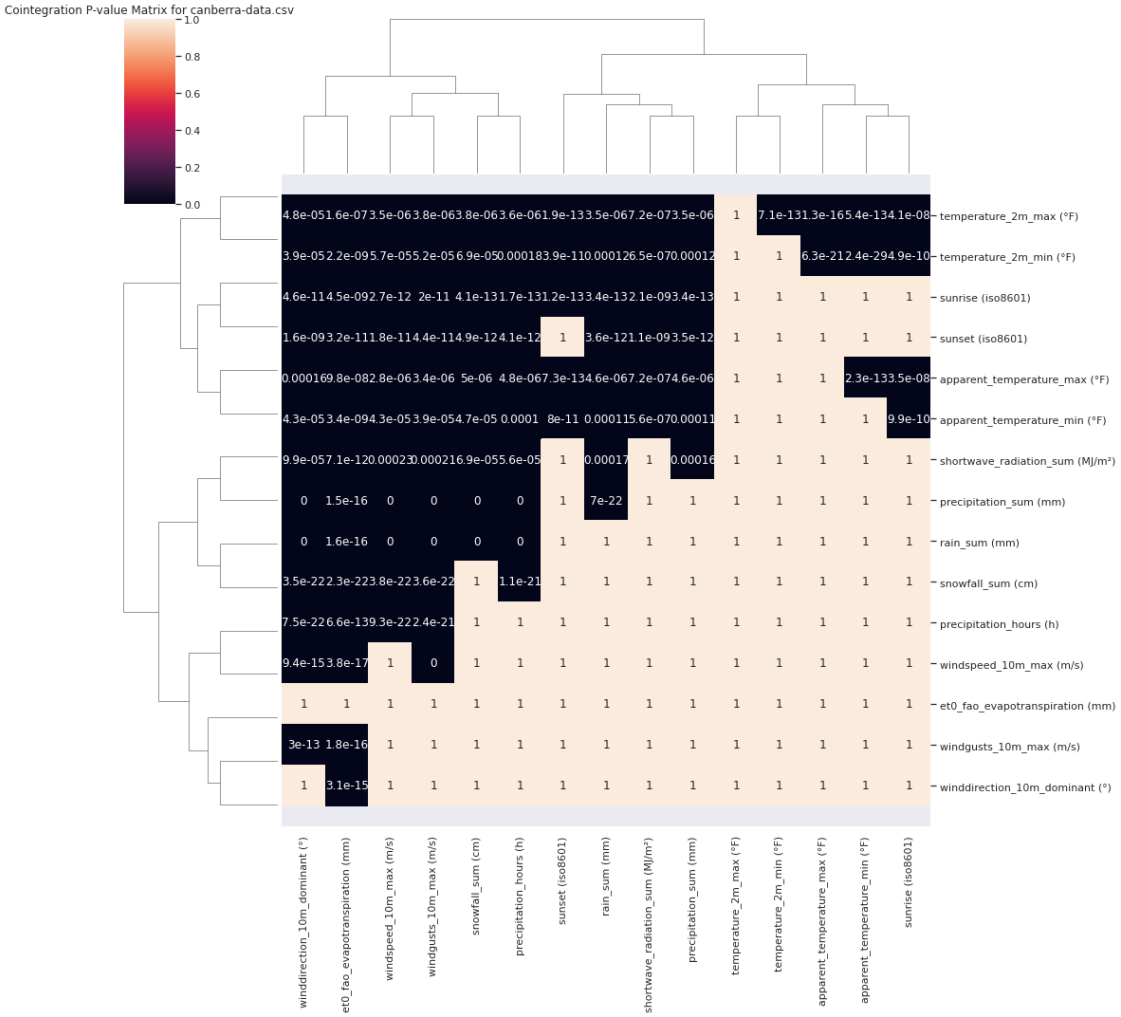
```
   null hypothesis Stationarity
0           Reject          Yes
1           Reject          Yes
2           Reject          Yes
3           Reject          Yes
4           Reject          Yes
5           Reject          Yes
6           Reject          Yes
7           Reject          Yes
8           Reject          Yes
9           Reject          Yes
10          Reject          Yes
11          Reject          Yes
12          Reject          Yes
13          Reject          Yes
14          Reject          Yes
```

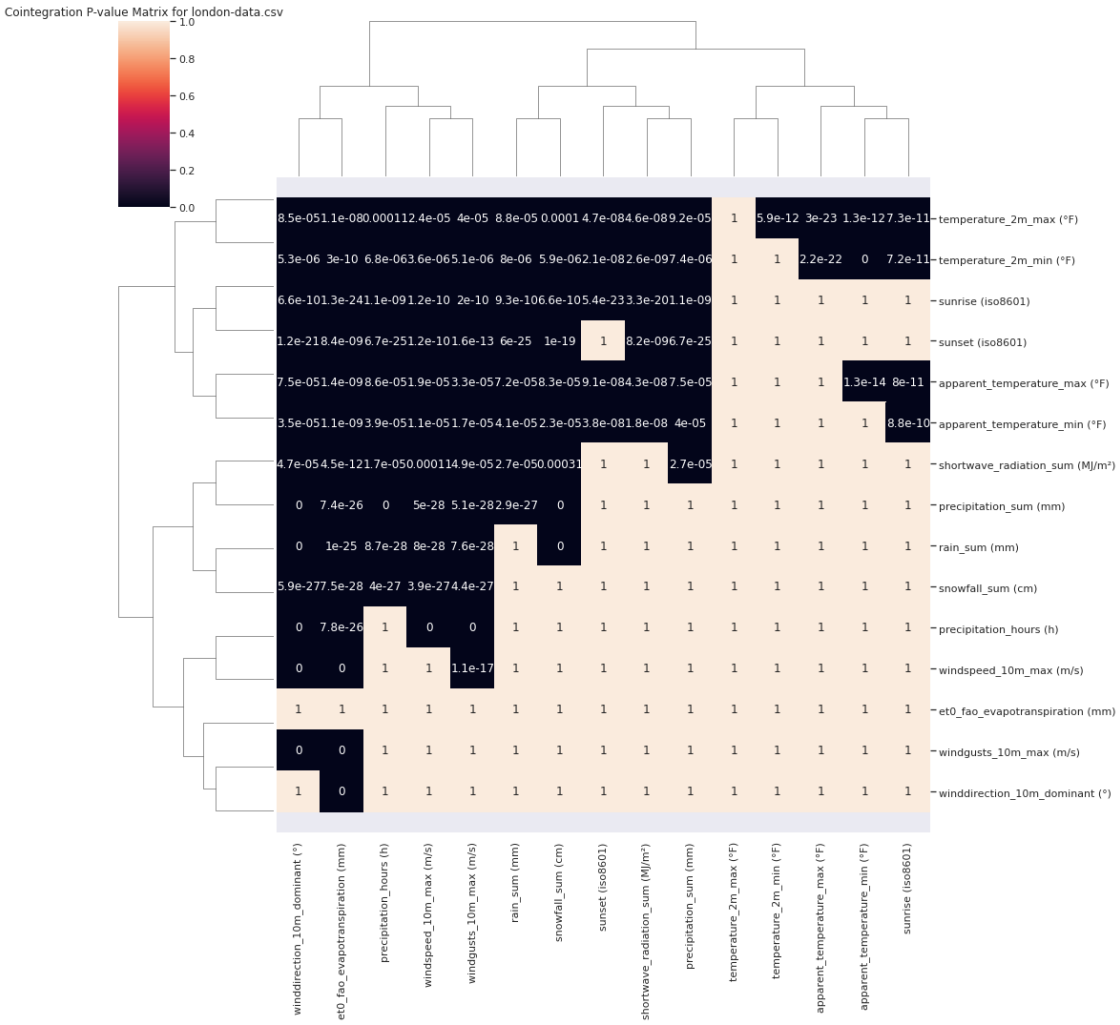
#Cointegration tests

[ ]:
```
from statsmodels.tsa.stattools import coint


def find_cointegrated_pairs(dataframe, critial_level = 0.05):
    n = dataframe.shape[1]
    pvalue_matrix = np.ones((n, n))
    keys = dataframe.columns
    pairs = []
    for i in range(n):
        for j in range(i+1, n):
            series1 = dataframe[keys[i]]
            series2 = dataframe[keys[j]]
            result = coint(series1, series2)
            pvalue = result[1]
            pvalue_matrix[i, j] = pvalue
            if pvalue < critial_level:
                pairs.append((keys[i], keys[j], pvalue))
    return pvalue_matrix, pairs
```

[ ]:
```
for index , data in enumerate(data_all):
    pvalue_matrix, pairs = find_cointegrated_pairs(data)
    coint_pvalue_matrix_df = pd.DataFrame(pvalue_matrix)

    g = sns.clustermap(coint_pvalue_matrix_df, xticklabels=data.
 ↪columns,yticklabels=data.columns, annot=True,
                       figsize=(15, 15))
    plt.title('Cointegration P-value Matrix for '+ data_files[index])
    ax = g.ax_heatmap
    bottom, top = ax.get_ylim()
```

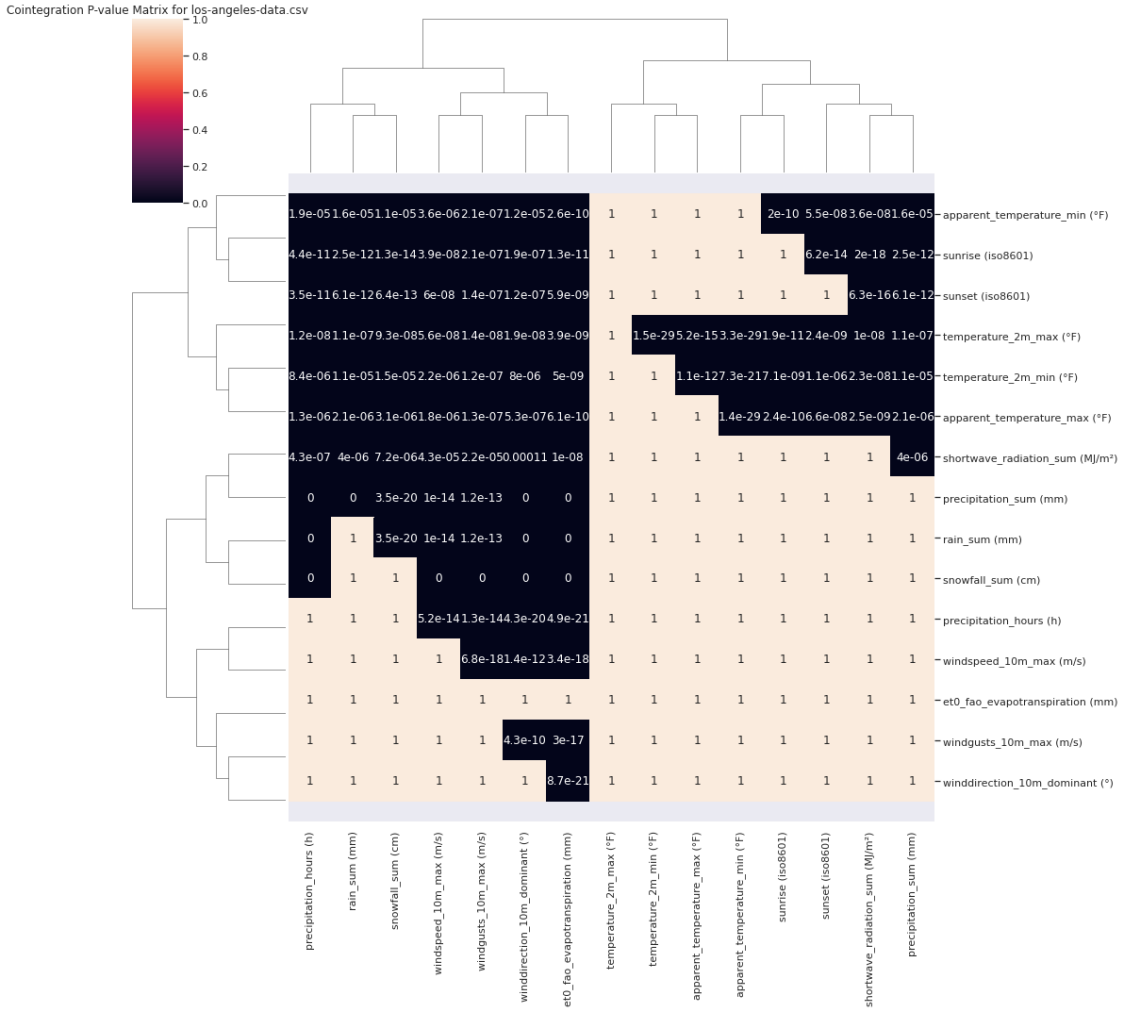
```
    ax.set_ylim(bottom + 0.5, top - 0.5)
    plt.show()
```





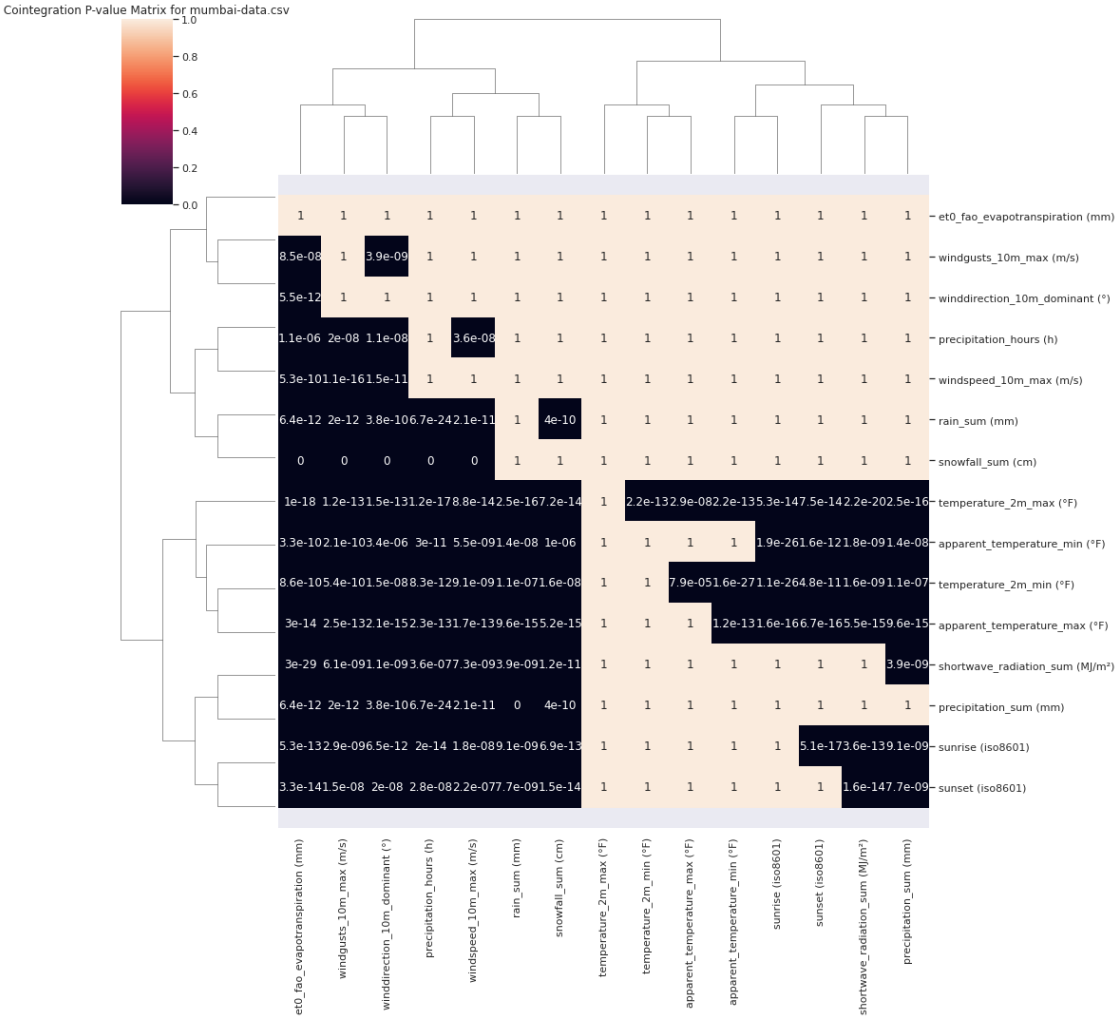

```
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/stattools.py:1605:
CollinearityWarning: y0 and y1 are (almost) perfectly colinear.Cointegration
test is not reliable in this case.
  warnings.warn(
/usr/local/lib/python3.8/dist-
packages/statsmodels/regression/linear_model.py:1715: RuntimeWarning: invalid
value encountered in double_scalars
  return 1 - self.ssr/self.centered_tss
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/stattools.py:1605:
CollinearityWarning: y0 and y1 are (almost) perfectly colinear.Cointegration
test is not reliable in this case.
  warnings.warn(
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/stattools.py:1605:
CollinearityWarning: y0 and y1 are (almost) perfectly colinear.Cointegration
test is not reliable in this case.
  warnings.warn(
```

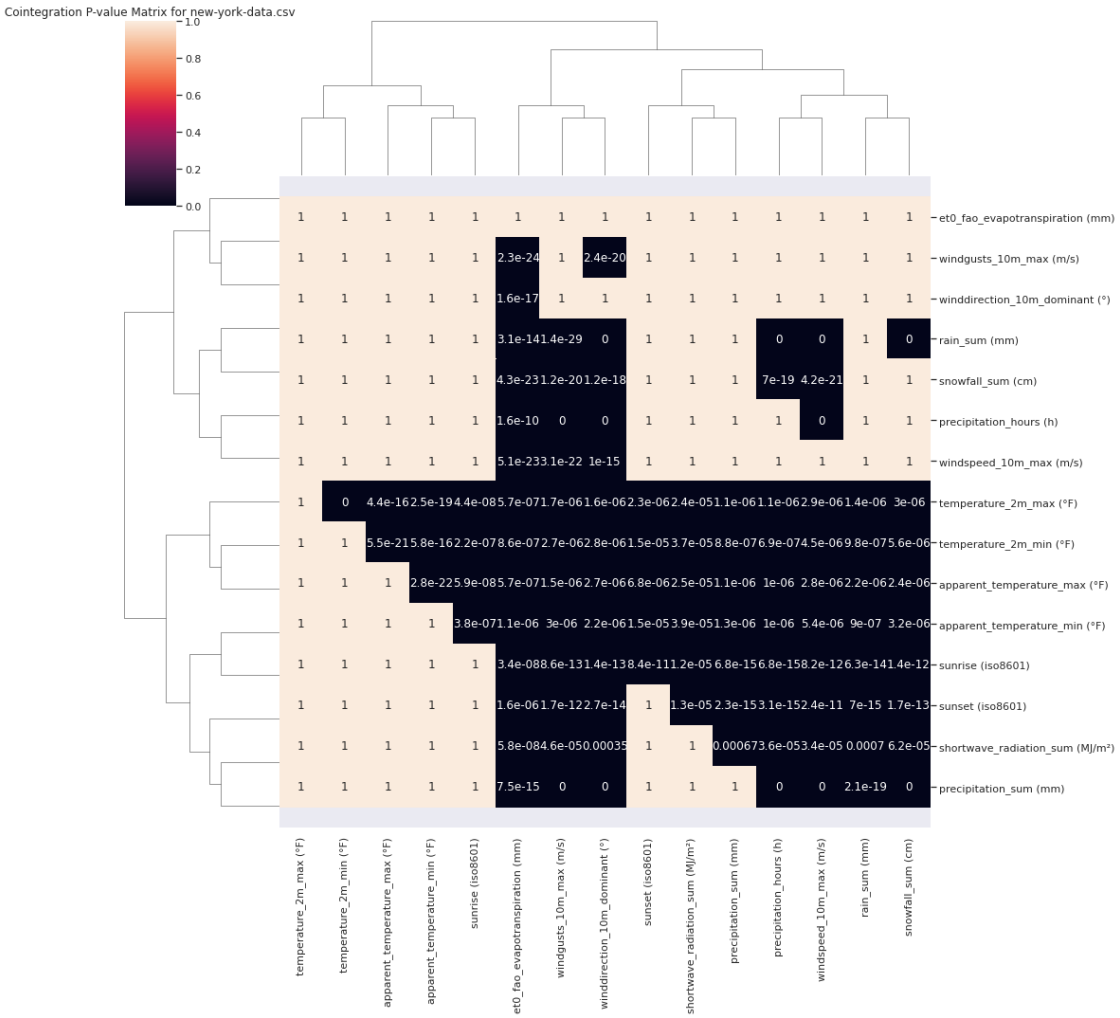
```
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/stattools.py:1605:
CollinearityWarning: y0 and y1 are (almost) perfectly colinear.Cointegration
test is not reliable in this case.
  warnings.warn(
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/stattools.py:1605:
CollinearityWarning: y0 and y1 are (almost) perfectly colinear.Cointegration
test is not reliable in this case.
  warnings.warn(
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/stattools.py:1605:
CollinearityWarning: y0 and y1 are (almost) perfectly colinear.Cointegration
test is not reliable in this case.
  warnings.warn(
```




```
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/stattools.py:1605:
CollinearityWarning: y0 and y1 are (almost) perfectly colinear.Cointegration
test is not reliable in this case.
  warnings.warn(
```

```
/usr/local/lib/python3.8/dist-
packages/statsmodels/regression/linear_model.py:1715: RuntimeWarning: invalid
value encountered in double_scalars
  return 1 - self.ssr/self.centered_tss
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/stattools.py:1605:
CollinearityWarning: y0 and y1 are (almost) perfectly colinear.Cointegration
test is not reliable in this case.
  warnings.warn(
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/stattools.py:1605:
CollinearityWarning: y0 and y1 are (almost) perfectly colinear.Cointegration
test is not reliable in this case.
  warnings.warn(
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/stattools.py:1605:
CollinearityWarning: y0 and y1 are (almost) perfectly colinear.Cointegration
test is not reliable in this case.
  warnings.warn(
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/stattools.py:1605:
CollinearityWarning: y0 and y1 are (almost) perfectly colinear.Cointegration
test is not reliable in this case.
  warnings.warn(
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/stattools.py:1605:
CollinearityWarning: y0 and y1 are (almost) perfectly colinear.Cointegration
test is not reliable in this case.
  warnings.warn(
```


Cointegration P-value Matrix for mumbai-data.csv
1.0
0.8
0.6
0.4
0.2
0.0
et0_fao_evapotranspiration (mm)
windgusts_10m_max (m/s)
winddirection_10m_dominant (°)
precipitation_hours (h)
windspeed_10m_max (m/s)
rain_sum (mm)
snowfall_sum (cm)
temperature_2m_max (°F)
apparent_temperature_min (°F)
temperature_2m_min (°F)
apparent_temperature_max (°F)
shortwave_radiation_sum (MJ/m²)
precipitation_sum (mm)
sunrise (iso8601)
sunset (iso8601)



[ ]:
```
#Data scaling
from sklearn.preprocessing import StandardScaler

def data_scaling(X,y):
    scaler = StandardScaler()
    X_std = scaler.fit_transform(X)
    y_std = scaler.fit_transform(y)
    return X,y
```

[ ]:
```
#VAR
from statsmodels.tsa.api import VAR

nobs = 30
def VAR_regr(X):

    Xtr, Xts = X[0:-nobs], X[-nobs:]
```

```
    model = VAR(Xtr)


    result = model.fit()
    lag_order = result.k_ar
    forecast_input = Xtr.values[-lag_order:]

    Xts_pred = result.forecast(y=forecast_input , steps=nobs)
    rmse = np.sqrt( mean_squared_error( Xts.to_numpy()[:,4]   ,
                                        Xts_pred[:,4] ) )

    return rmse
```

[ ]:
```
#Linear Regression

from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error


def linear_regression(Xtr, Xts , ytr , yts):
    linear_regr = LinearRegression()
    linear_regr.fit(Xtr , ytr)
    yts_pred = linear_regr.predict(Xts)
    score = np.sqrt(mean_squared_error(yts, yts_pred))
    return score
```

[ ]:
```
from sklearn.linear_model import Lasso
from sklearn.linear_model import MultiTaskLassoCV
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import RepeatedKFold
from numpy import absolute
from numpy import mean
from numpy import std

def lasso_regr(Xtr, Xts , ytr , yts):
    lasso_cv_model = MultiTaskLassoCV(cv=5, random_state=0, max_iter=10000)
    lasso_cv_model.fit(Xtr, ytr)
    # lasso_cv_model.alpha_
    lasso_model = Lasso(alpha=lasso_cv_model.alpha_)
    lasso_model.fit(Xtr , ytr)
    yts_pred = lasso_model.predict(Xts)
    score = np.sqrt(mean_squared_error(yts, yts_pred))
    return score
```

[ ]:
```
from sklearn.linear_model import RidgeCV
from sklearn.linear_model import Ridge
```

```
lambda_values = np.random.randint(0,1000,100)

def ridge_regr(Xtr, Xts , ytr , yts) :
    ridgecv = RidgeCV(alphas = lambda_values , scoring =␣
 ↪"neg_mean_squared_error", cv = 10)
    ridgecv.fit(Xtr, ytr)
    # ridgecv.alpha_
    ridge_regr = Ridge(alpha = ridgecv.alpha_)
    ridge_regr.fit(Xtr, ytr)
    yts_pred = ridge_regr.predict(Xts)
    score = np.sqrt(mean_squared_error(yts, yts_pred) )
    return score
```

[ ]:
```
from sklearn import neighbors
from sklearn.neighbors import KNeighborsRegressor
from sklearn.metrics import mean_squared_error
from math import sqrt

def knn_regr(Xtr, Xts , ytr , yts) :
    rmse_val = []

    for K in range(30):
        K = K+1
        model = KNeighborsRegressor(n_neighbors = K)

        model.fit(Xtr, ytr)  #fit the model
        yts_pred = model.predict(Xts) #make prediction on test set
        error = sqrt(mean_squared_error(yts , yts_pred)) #calculate rmse
        rmse_val.append(error) #store rmse values

    neighbors = 1 if rmse_val.index( min(rmse_val) ) < 1 else rmse_val.index(␣
 ↪min(rmse_val) )
    model = KNeighborsRegressor( n_neighbors = neighbors )
    model.fit(Xtr, ytr)  #fit the model
    yts_pred = model.predict(Xts)

    score = np.sqrt(mean_squared_error(yts, yts_pred) )
    return score
```

[ ]:
```
# Decision Tree Regressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error
from sklearn.tree import plot_tree

target_columns = ["temperature_2m_min (°F)","temperature_2m_max␣
 ↪(°F)","apparent_temperature_min (°F)","apparent_temperature_max (°F)"]
```

```
def decision_tree_regr(Xtr, Xts , ytr , yts):

      model = DecisionTreeRegressor()
      model.fit(Xtr, ytr)
      yts_pred = model.predict(Xts)

      dtree_rmse = np.sqrt(mean_squared_error(yts, yts_pred))
      return dtree_rmse
```

[ ]:
```
from sklearn.model_selection import train_test_split

scores_list=[]

for data in data_all:
    y = data[["temperature_2m_max (°F)","temperature_2m_min (°F)",
        "apparent_temperature_max (°F)","apparent_temperature_min (°F)"]]
    X = data[["sunrise (iso8601)","sunset (iso8601)","shortwave_radiation_sum␣
 ↪(MJ/m²)",
        "precipitation_sum (mm)","rain_sum (mm)","snowfall_sum␣
 ↪(cm)","precipitation_hours (h)",
        "windspeed_10m_max (m/s)", "windgusts_10m_max (m/
 ↪s)","winddirection_10m_dominant (°)","et0_fao_evapotranspiration (mm)"]]

    scaled_data = data_scaling(X,y)

    Xtr, Xts, ytr, yts = train_test_split(scaled_data[0] , scaled_data[1] ,␣
 ↪test_size = 0.25 )
    temp_scores_list = []
    # Using Linear regression
    temp_scores_list.append(linear_regression(Xtr, Xts, ytr, yts))
    # Using Lasso regression
    temp_scores_list.append(lasso_regr(Xtr, Xts, ytr, yts))
    # Using Ridge regression
    temp_scores_list.append(ridge_regr(Xtr, Xts, ytr, yts))
    # Using K Neighbours regression
    temp_scores_list.append(knn_regr(Xtr, Xts, ytr, yts))
    # Using Decision Tree regression
    temp_scores_list.append(decision_tree_regr(Xtr, Xts, ytr, yts))
    #Using VAR
    temp_scores_list.append(VAR_regr(data))

    scores_list.append(temp_scores_list)
    # print(X.shape)
```

```
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/base/tsa_model.py:581:
```

```
ValueWarning: A date index has been provided, but it has no associated frequency
information and so will be ignored when e.g. forecasting.
  warnings.warn('A date index has been provided, but it has no'
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/base/tsa_model.py:581:
ValueWarning: A date index has been provided, but it has no associated frequency
information and so will be ignored when e.g. forecasting.
  warnings.warn('A date index has been provided, but it has no'
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/base/tsa_model.py:581:
ValueWarning: A date index has been provided, but it has no associated frequency
information and so will be ignored when e.g. forecasting.
  warnings.warn('A date index has been provided, but it has no'
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/base/tsa_model.py:581:
ValueWarning: A date index has been provided, but it has no associated frequency
information and so will be ignored when e.g. forecasting.
  warnings.warn('A date index has been provided, but it has no'
/usr/local/lib/python3.8/dist-packages/statsmodels/tsa/base/tsa_model.py:581:
ValueWarning: A date index has been provided, but it has no associated frequency
information and so will be ignored when e.g. forecasting.
  warnings.warn('A date index has been provided, but it has no'
```

[ ]:
```
combined_data = (pd.concat(data_all) )
```

[ ]:
```
y = combined_data[["temperature_2m_max (°F)","temperature_2m_min (°F)",
    "apparent_temperature_max (°F)","apparent_temperature_min (°F)"]]
X = combined_data[["sunrise (iso8601)","sunset␣
↪(iso8601)","shortwave_radiation_sum (MJ/m²)",
    "precipitation_sum (mm)","rain_sum (mm)","snowfall_sum␣
↪(cm)","precipitation_hours (h)",
    "windspeed_10m_max (m/s)", "windgusts_10m_max (m/
↪s)","winddirection_10m_dominant (°)","et0_fao_evapotranspiration (mm)"]]

scaled_data = data_scaling(X,y)
# print(scaled_data[0].shape)
# print(scaled_data[1].shape)
Xtr, Xts, ytr, yts = train_test_split(scaled_data[0] , scaled_data[1] ,␣
↪test_size = 0.25 )
temp_scores_list = []
# Using Linear regression
temp_scores_list.append(linear_regression(Xtr, Xts, ytr, yts))
# Using Lasso regression
temp_scores_list.append(lasso_regr(Xtr, Xts, ytr, yts))
# Using Ridge regression
temp_scores_list.append(ridge_regr(Xtr, Xts, ytr, yts))
# Using K Neighbours regression
temp_scores_list.append(knn_regr(Xtr, Xts, ytr, yts))
# Using Decision Tree regression
temp_scores_list.append(decision_tree_regr(Xtr, Xts, ytr, yts))
```

```
# We cannot use VAR for combined data
temp_scores_list.append(None)
scores_list.append(temp_scores_list)
# print(X.shape)
```

[ ]:
```
for i in range(len(data_files)):
  data_files[i] = data_files[i].replace("-data.csv","")
```

[ ]:
```
score_cols = ["Linear Regression" , "Lasso Regression" , "Ridge Regression" , "K␣
↪Neighbours Regression", "Decision Tree Regression" , "VAR"]
data_files.append("combined_locations")
scores_list_df = pd.DataFrame(scores_list, columns=score_cols , index =␣
↪data_files)

scores_list_df
```

[ ]:

	Linear Regression	Lasso Regression	Ridge Regression	\
canberra	5.236979	7.565535	5.246868	
london	6.007466	8.255969	6.015647	
los-angeles	5.160602	8.309621	5.160769	
mumbai	3.114037	4.079621	3.114070	
new-york	7.595215	12.162461	7.609076	
combined_locations	8.294295	15.918454	8.299724	

	K Neighbours Regression	Decision Tree Regression	\
canberra	6.363952	6.217231	
london	6.185384	6.688432	
los-angeles	6.038354	5.034846	
mumbai	3.002185	3.014948	
new-york	8.003590	8.499780	
combined_locations	6.207913	6.496159	

	VAR
canberra	4336.884746
london	20169.191912
los-angeles	1128.403680
mumbai	516.199766
new-york	574.082491
combined_locations	NaN

[ ]:
```
from matplotlib.pyplot import figure
```

[ ]:
```
figure(figsize=(10, 6), dpi=200)

scores = scores_list_df.drop("VAR",axis=1)
plt.plot(scores,'o')
```

[ ]: [<matplotlib.lines.Line2D at 0x7f59d4488310>,
 <matplotlib.lines.Line2D at 0x7f59d4488cd0>,
 <matplotlib.lines.Line2D at 0x7f59d4488f70>,
 <matplotlib.lines.Line2D at 0x7f59d44887c0>,
 <matplotlib.lines.Line2D at 0x7f59d4488880>]