In [181]:
```python
import pandas as pd
```

1.

In [182]:
```python
df = pd.read_csv("Forecast_Data_Set.csv")
```

In [183]:
```python
df
```

Out[183]:

	station	Date	Present_Tmax	Present_Tmin	LDAPS_RHmin	LDAPS_RHmax	LDAPS_Tmax_lapse	LDAPS_Tmin_lapse	LDAPS_WS	LDAPS_LH	...	LDAPS_PPT2	LDAPS_PPT3	LDAPS_PPT4	lat	lon	DEM	Slope	Solar radiation	Next_Tmax	Next_Tmin
0	1.0	2013-06-30	28.7	21.4	58.255688	91.116364	28.074101	23.006936	6.818887	69.451805	...	0.000000	0.000000	0.000000	37.6046	126.991	212.3350	2.785000	5992.895996	29.1	21.2
1	2.0	2013-06-30	31.9	21.6	52.263397	90.604721	29.850689	24.035009	5.691890	51.937448	...	0.000000	0.000000	0.000000	37.6046	127.032	44.7624	0.514100	5869.312500	30.5	22.5
2	3.0	2013-06-30	31.6	23.3	48.690479	83.973587	30.091292	24.565633	6.138224	20.573050	...	0.000000	0.000000	0.000000	37.5776	127.058	33.3068	0.266100	5863.555664	31.1	23.9
3	4.0	2013-06-30	32.0	23.4	58.239788	96.483688	29.704629	23.326177	5.650050	65.727144	...	0.000000	0.000000	0.000000	37.6450	127.022	45.7160	2.534800	5856.964844	31.7	24.3
4	5.0	2013-06-30	31.4	21.9	56.174095	90.155128	29.113934	23.486480	5.735004	107.965535	...	0.000000	0.000000	0.000000	37.5507	127.135	35.0380	0.505500	5859.552246	31.2	22.5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7747	23.0	2017-08-30	23.3	17.1	26.741310	78.869858	26.352081	18.775678	6.148918	72.058294	...	0.000000	0.000000	0.000000	37.5372	126.891	15.5876	0.155400	4443.313965	28.3	18.1
7748	24.0	2017-08-30	23.3	17.7	24.040634	77.294975	27.010193	18.733519	6.542819	47.241457	...	0.000000	0.000000	0.000000	37.5237	126.909	17.2956	0.222300	4438.373535	28.6	18.8
7749	25.0	2017-08-30	23.2	17.4	22.933014	77.243744	27.939516	18.522965	7.289264	9.090034	...	0.000000	0.000000	0.000000	37.5237	126.970	19.5844	0.271300	4451.345215	27.8	17.4
7750	NaN	NaN	20.0	11.3	19.794666	58.936283	17.624954	14.272646	2.882580	-13.603212	...	0.000000	0.000000	0.000000	37.4562	126.826	12.3700	0.098475	4329.520508	17.4	11.3
7751	NaN	NaN	37.6	29.9	98.524734	100.000153	38.542255	29.619342	21.857621	213.414006	...	21.621661	15.841235	16.655469	37.6450	127.135	212.3350	5.178230	5992.895996	38.9	29.8

7752 rows × 25 columns

2.

In [184]:
```python
count = len(df)
print("Number of data points:", count)
```

```
Number of data points: 7752
```

3.

In [185]:
```python
print(df.head(10))
```

```
   station        Date  Present_Tmax  Present_Tmin  LDAPS_RHmin  LDAPS_RHmax  \
0      1.0  2013-06-30          28.7          21.4    58.255688    91.116364
1      2.0  2013-06-30          31.9          21.6    52.263397    90.604721
2      3.0  2013-06-30          31.6          23.3    48.690479    83.973587
3      4.0  2013-06-30          32.0          23.4    58.239788    96.483688
4      5.0  2013-06-30          31.4          21.9    56.174095    90.155128
5      6.0  2013-06-30          31.9          23.5    52.437126    85.307251
6      7.0  2013-06-30          31.4          24.4    56.287189    81.019760
7      8.0  2013-06-30          32.1          23.6    52.326218    78.004539
8      9.0  2013-06-30          31.4          22.0    55.338791    80.784607
9     10.0  2013-06-30          31.6          20.5    56.651203    86.849632

   LDAPS_Tmax_lapse  LDAPS_Tmin_lapse  LDAPS_WS    LDAPS_LH  ...  LDAPS_PPT2  \
0         28.074101         23.006936  6.818887   69.451805  ...         0.0
1         29.850689         24.035009  5.691890   51.937448  ...         0.0
2         30.091292         24.565633  6.138224   20.573050  ...         0.0
3         29.704629         23.326177  5.650050   65.727144  ...         0.0
4         29.113934         23.486480  5.735004  107.965535  ...         0.0
5         29.219342         23.822613  6.182295   50.231389  ...         0.0
6         28.551859         24.238467  5.587135  125.110007  ...         0.0
7         28.851982         23.819054  6.104417   42.011547  ...         0.0
8         28.426975         23.332373  6.017135   85.110971  ...         0.0
9         27.576705         22.527018  6.518841   63.006075  ...         0.0

   LDAPS_PPT3  LDAPS_PPT4      lat      lon       DEM   Slope  \
0         0.0         0.0  37.6046  126.991  212.3350  2.7850
1         0.0         0.0  37.6046  127.032   44.7624  0.5141
2         0.0         0.0  37.5776  127.058   33.3068  0.2661
3         0.0         0.0  37.6450  127.022   45.7160  2.5348
4         0.0         0.0  37.5507  127.135   35.0380  0.5055
5         0.0         0.0  37.5102  127.042   54.6384  0.1457
6         0.0         0.0  37.5776  126.838   12.3700  0.0985
7         0.0         0.0  37.4697  126.910   52.5180  1.5629
8         0.0         0.0  37.4967  126.826   50.9312  0.4125
9         0.0         0.0  37.4562  126.955  208.5070  5.1782

   Solar radiation  Next_Tmax  Next_Tmin
0      5992.895996       29.1       21.2
1      5869.312500       30.5       22.5
2      5863.555664       31.1       23.9
3      5856.964844       31.7       24.3
4      5859.552246       31.2       22.5
5      5873.780762       31.5       24.0
6      5849.233398       30.9       23.4
7      5863.992188       31.1       22.9
8      5876.901367       31.3       21.6
9      5893.608398       30.5       21.0

[10 rows x 25 columns]
```

4.

In [186]:
```python
print(df.tail(10))
```

```
      station        Date  Present_Tmax  Present_Tmin  LDAPS_RHmin  \
7742     18.0  2017-08-30          23.3          18.2    30.259682
7743     19.0  2017-08-30          22.6          15.0    25.536045
7744     20.0  2017-08-30          22.7          15.9    38.216373
7745     21.0  2017-08-30          23.1          17.8    24.688997
7746     22.0  2017-08-30          22.5          17.4    30.094858
7747     23.0  2017-08-30          23.3          17.1    26.741310
7748     24.0  2017-08-30          23.3          17.7    24.040634
7749     25.0  2017-08-30          23.2          17.4    22.933014
7750      NaN         NaN          20.0          11.3    19.794666
7751      NaN         NaN          37.6          29.9    98.524734

      LDAPS_RHmax  LDAPS_Tmax_lapse  LDAPS_Tmin_lapse   LDAPS_WS    LDAPS_LH  \
7742    86.565193         26.474403         17.289157   6.522731   94.203783
7743    78.779045         26.596280         17.856040   6.905963   54.030075
7744    94.428116         24.037222         15.777677   6.478060  112.580310
7745    78.261383         27.812697         18.303014   6.603253    9.614074
7746    83.690018         26.704905         17.814038   5.768083   82.146707
7747    78.869858         26.352081         18.775678   6.148918   72.058294
7748    77.294975         27.010193         18.733519   6.542819   47.241457
7749    77.243744         27.939516         18.522965   7.289264    9.090034
7750    58.936283         17.624954         14.272646   2.882580  -13.603212
7751   100.000153         38.542255         29.619342  21.857621  213.414006

      ...  LDAPS_PPT2  LDAPS_PPT3  LDAPS_PPT4      lat      lon       DEM  \
7742  ...    0.000000    0.000000    0.000000  37.4832  127.024   56.4448
7743  ...    0.000000    0.000000    0.000000  37.5776  126.938   75.0924
7744  ...    0.000000    0.000000    0.000000  37.6181  127.004  146.5540
7745  ...    0.000000    0.000000    0.000000  37.5507  127.040   26.2980
7746  ...    0.000000    0.000000    0.000000  37.5102  127.086   21.9668
7747  ...    0.000000    0.000000    0.000000  37.5372  126.891   15.5876
7748  ...    0.000000    0.000000    0.000000  37.5237  126.909   17.2956
7749  ...    0.000000    0.000000    0.000000  37.5237  126.970   19.5844
7750  ...    0.000000    0.000000    0.000000  37.4562  126.826   12.3700
7751  ...   21.621661   15.841235   16.655469  37.6450  127.135  212.3350

         Slope  Solar radiation  Next_Tmax  Next_Tmin
7742  1.231300      4435.611816       27.8       16.7
7743  1.767800      4495.473633       27.1       16.6
7744  4.729600      4554.965332       26.7       16.3
7745  0.572100      4456.024414       27.6       17.7
7746  0.133200      4441.803711       28.0       17.1
7747  0.155400      4443.313965       28.3       18.1
7748  0.222300      4438.373535       28.6       18.8
7749  0.271300      4451.345215       27.8       17.4
7750  0.098475      4329.520508       17.4       11.3
7751  5.178230      5992.895996       38.9       29.8

[10 rows x 25 columns]
```

5.

In [187]:
```python
df.rename(columns = {"Date": "Year_Month_Day"}, inplace = True)
```

In [188]:
```python
df.head()
```

Out[188]:

	station	Year_Month_Day	Present_Tmax	Present_Tmin	LDAPS_RHmin	LDAPS_RHmax	LDAPS_Tmax_lapse	LDAPS_Tmin_lapse	LDAPS_WS	LDAPS_LH	...	LDAPS_PPT2	LDAPS_PPT3	LDAPS_PPT4	lat	lon	DEM	Slope	Solar radiation	Next_Tmax	Next_Tmin
0	1.0	2013-06-30	28.7	21.4	58.255688	91.116364	28.074101	23.006936	6.818887	69.451805	...	0.0	0.0	0.0	37.6046	126.991	212.3350	2.7850	5992.895996	29.1	21.2
1	2.0	2013-06-30	31.9	21.6	52.263397	90.604721	29.850689	24.035009	5.691890	51.937448	...	0.0	0.0	0.0	37.6046	127.032	44.7624	0.5141	5869.312500	30.5	22.5
2	3.0	2013-06-30	31.6	23.3	48.690479	83.973587	30.091292	24.565633	6.138224	20.573050	...	0.0	0.0	0.0	37.5776	127.058	33.3068	0.2661	5863.555664	31.1	23.9
3	4.0	2013-06-30	32.0	23.4	58.239788	96.483688	29.704629	23.326177	5.650050	65.727144	...	0.0	0.0	0.0	37.6450	127.022	45.7160	2.5348	5856.964844	31.7	24.3
4	5.0	2013-06-30	31.4	21.9	56.174095	90.155128	29.113934	23.486480	5.735004	107.965535	...	0.0	0.0	0.0	37.5507	127.135	35.0380	0.5055	5859.552246	31.2	22.5

5 rows × 25 columns

6.

In [189]:
```python
df.drop("LDAPS_PPT4", axis = 1, inplace = True)
```

In [190]:
```python
print(df.head())
```

```
   station Year_Month_Day  Present_Tmax  Present_Tmin  LDAPS_RHmin  \
0      1.0     2013-06-30          28.7          21.4    58.255688
1      2.0     2013-06-30          31.9          21.6    52.263397
2      3.0     2013-06-30          31.6          23.3    48.690479
3      4.0     2013-06-30          32.0          23.4    58.239788
4      5.0     2013-06-30          31.4          21.9    56.174095

   LDAPS_RHmax  LDAPS_Tmax_lapse  LDAPS_Tmin_lapse  LDAPS_WS    LDAPS_LH  ...  \
0    91.116364         28.074101         23.006936  6.818887   69.451805  ...
1    90.604721         29.850689         24.035009  5.691890   51.937448  ...
2    83.973587         30.091292         24.565633  6.138224   20.573050  ...
3    96.483688         29.704629         23.326177  5.650050   65.727144  ...
4    90.155128         29.113934         23.486480  5.735004  107.965535  ...

   LDAPS_PPT1  LDAPS_PPT2  LDAPS_PPT3      lat      lon       DEM   Slope  \
0         0.0         0.0         0.0  37.6046  126.991  212.3350  2.7850
1         0.0         0.0         0.0  37.6046  127.032   44.7624  0.5141
2         0.0         0.0         0.0  37.5776  127.058   33.3068  0.2661
3         0.0         0.0         0.0  37.6450  127.022   45.7160  2.5348
4         0.0         0.0         0.0  37.5507  127.135   35.0380  0.5055

   Solar radiation  Next_Tmax  Next_Tmin
0      5992.895996       29.1       21.2
1      5869.312500       30.5       22.5
2      5863.555664       31.1       23.9
3      5856.964844       31.7       24.3
4      5859.552246       31.2       22.5

[5 rows x 24 columns]
```

7.

In [191]:
```python
df["Next_Taverage"] = (df["Next_Tmin"] + df["Next_Tmax"]) / 2
```

In [192]:
```python
print(df.head())
```

```
   station Year_Month_Day  Present_Tmax  Present_Tmin  LDAPS_RHmin  \
0      1.0     2013-06-30          28.7          21.4    58.255688
1      2.0     2013-06-30          31.9          21.6    52.263397
2      3.0     2013-06-30          31.6          23.3    48.690479
3      4.0     2013-06-30          32.0          23.4    58.239788
4      5.0     2013-06-30          31.4          21.9    56.174095

   LDAPS_RHmax  LDAPS_Tmax_lapse  LDAPS_Tmin_lapse  LDAPS_WS    LDAPS_LH  ...  \
0    91.116364         28.074101         23.006936  6.818887   69.451805  ...
1    90.604721         29.850689         24.035009  5.691890   51.937448  ...
2    83.973587         30.091292         24.565633  6.138224   20.573050  ...
3    96.483688         29.704629         23.326177  5.650050   65.727144  ...
4    90.155128         29.113934         23.486480  5.735004  107.965535  ...

   LDAPS_PPT2  LDAPS_PPT3      lat      lon       DEM   Slope  \
0         0.0         0.0  37.6046  126.991  212.3350  2.7850
1         0.0         0.0  37.6046  127.032   44.7624  0.5141
2         0.0         0.0  37.5776  127.058   33.3068  0.2661
3         0.0         0.0  37.6450  127.022   45.7160  2.5348
4         0.0         0.0  37.5507  127.135   35.0380  0.5055

   Solar radiation  Next_Tmax  Next_Tmin  Next_Taverage
0      5992.895996       29.1       21.2          25.15
1      5869.312500       30.5       22.5          26.50
2      5863.555664       31.1       23.9          27.50
3      5856.964844       31.7       24.3          28.00
4      5859.552246       31.2       22.5          26.85

[5 rows x 25 columns]
```

8.

In [193]:
```python
sr_min = df["Solar radiation"].min()
sr_max = df["Solar radiation"].max()
```

In [194]:
```python
df["Solar radiation"] = ((df["Solar radiation"] - sr_min) / (sr_max - sr_min))
```

In [195]:
```python
print(df.head())
```

```
   station Year_Month_Day  Present_Tmax  Present_Tmin  LDAPS_RHmin  \
0      1.0     2013-06-30          28.7          21.4    58.255688
1      2.0     2013-06-30          31.9          21.6    52.263397
2      3.0     2013-06-30          31.6          23.3    48.690479
3      4.0     2013-06-30          32.0          23.4    58.239788
4      5.0     2013-06-30          31.4          21.9    56.174095

   LDAPS_RHmax  LDAPS_Tmax_lapse  LDAPS_Tmin_lapse  LDAPS_WS    LDAPS_LH  ...  \
0    91.116364         28.074101         23.006936  6.818887   69.451805  ...
1    90.604721         29.850689         24.035009  5.691890   51.937448  ...
2    83.973587         30.091292         24.565633  6.138224   20.573050  ...
3    96.483688         29.704629         23.326177  5.650050   65.727144  ...
4    90.155128         29.113934         23.486480  5.735004  107.965535  ...

   LDAPS_PPT2  LDAPS_PPT3      lat      lon       DEM   Slope  \
0         0.0         0.0  37.6046  126.991  212.3350  2.7850
1         0.0         0.0  37.6046  127.032   44.7624  0.5141
2         0.0         0.0  37.5776  127.058   33.3068  0.2661
3         0.0         0.0  37.6450  127.022   45.7160  2.5348
4         0.0         0.0  37.5507  127.135   35.0380  0.5055

   Solar radiation  Next_Tmax  Next_Tmin  Next_Taverage
0         1.000000       29.1       21.2          25.15
1         0.925703       30.5       22.5          26.50
2         0.922242       31.1       23.9          27.50
3         0.918280       31.7       24.3          28.00
4         0.919835       31.2       22.5          26.85

[5 rows x 25 columns]
```

9.

In [196]:
```python
check = df.isnull().sum()
print("MISSING DATA")
print(check)
```

```
MISSING DATA
station              2
Year_Month_Day       2
Present_Tmax        70
Present_Tmin        70
LDAPS_RHmin         75
LDAPS_RHmax         75
LDAPS_Tmax_lapse    75
LDAPS_Tmin_lapse    75
LDAPS_WS            75
LDAPS_LH            75
LDAPS_CC1           75
LDAPS_CC2           75
LDAPS_CC3           75
LDAPS_CC4           75
LDAPS_PPT1          75
LDAPS_PPT2          75
LDAPS_PPT3          75
lat                  0
lon                  0
DEM                  0
Slope                0
Solar radiation      0
Next_Tmax           27
Next_Tmin           27
Next_Taverage       27
dtype: int64
```

In [197]:
```python
# filling in missing values with zeros
df.fillna(value = 0, inplace = True)
```

In [198]:
```python
print(df.tail())
```

```
      station Year_Month_Day  Present_Tmax  Present_Tmin  LDAPS_RHmin  \
7747     23.0     2017-08-30          23.3          17.1    26.741310
7748     24.0     2017-08-30          23.3          17.7    24.040634
7749     25.0     2017-08-30          23.2          17.4    22.933014
7750      0.0              0          20.0          11.3    19.794666
7751      0.0              0          37.6          29.9    98.524734

      LDAPS_RHmax  LDAPS_Tmax_lapse  LDAPS_Tmin_lapse   LDAPS_WS    LDAPS_LH  \
7747    78.869858         26.352081         18.775678   6.148918   72.058294
7748    77.294975         27.010193         18.733519   6.542819   47.241457
7749    77.243744         27.939516         18.522965   7.289264    9.090034
7750    58.936283         17.624954         14.272646   2.882580  -13.603212
7751   100.000153         38.542255         29.619342  21.857621  213.414006

      ...  LDAPS_PPT2  LDAPS_PPT3      lat      lon       DEM     Slope  \
7747  ...    0.000000    0.000000  37.5372  126.891   15.5876  0.155400
7748  ...    0.000000    0.000000  37.5237  126.909   17.2956  0.222300
7749  ...    0.000000    0.000000  37.5237  126.970   19.5844  0.271300
7750  ...    0.000000    0.000000  37.4562  126.826   12.3700  0.098475
7751  ...   21.621661   15.841235  37.6450  127.135  212.3350  5.178230

      Solar radiation  Next_Tmax  Next_Tmin  Next_Taverage
7747         0.068411       28.3       18.1          23.20
7748         0.065441       28.6       18.8          23.70
7749         0.073239       27.8       17.4          22.60
7750         0.000000       17.4       11.3          14.35
7751         1.000000       38.9       29.8          34.35

[5 rows x 25 columns]
```

10.

In [199]:
```python
# resetting dataframe
df = pd.read_csv("Forecast_Data_Set.csv")
df
```

Out[199]:

	station	Date	Present_Tmax	Present_Tmin	LDAPS_RHmin	LDAPS_RHmax	LDAPS_Tmax_lapse	LDAPS_Tmin_lapse	LDAPS_WS	LDAPS_LH	...	LDAPS_PPT2	LDAPS_PPT3	LDAPS_PPT4	lat	lon	DEM	Slope	Solar radiation	Next_Tmax	Next_Tmin
0	1.0	2013-06-30	28.7	21.4	58.255688	91.116364	28.074101	23.006936	6.818887	69.451805	...	0.000000	0.000000	0.000000	37.6046	126.991	212.3350	2.785000	5992.895996	29.1	21.2
1	2.0	2013-06-30	31.9	21.6	52.263397	90.604721	29.850689	24.035009	5.691890	51.937448	...	0.000000	0.000000	0.000000	37.6046	127.032	44.7624	0.514100	5869.312500	30.5	22.5
2	3.0	2013-06-30	31.6	23.3	48.690479	83.973587	30.091292	24.565633	6.138224	20.573050	...	0.000000	0.000000	0.000000	37.5776	127.058	33.3068	0.266100	5863.555664	31.1	23.9
3	4.0	2013-06-30	32.0	23.4	58.239788	96.483688	29.704629	23.326177	5.650050	65.727144	...	0.000000	0.000000	0.000000	37.6450	127.022	45.7160	2.534800	5856.964844	31.7	24.3
4	5.0	2013-06-30	31.4	21.9	56.174095	90.155128	29.113934	23.486480	5.735004	107.965535	...	0.000000	0.000000	0.000000	37.5507	127.135	35.0380	0.505500	5859.552246	31.2	22.5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7747	23.0	2017-08-30	23.3	17.1	26.741310	78.869858	26.352081	18.775678	6.148918	72.058294	...	0.000000	0.000000	0.000000	37.5372	126.891	15.5876	0.155400	4443.313965	28.3	18.1
7748	24.0	2017-08-30	23.3	17.7	24.040634	77.294975	27.010193	18.733519	6.542819	47.241457	...	0.000000	0.000000	0.000000	37.5237	126.909	17.2956	0.222300	4438.373535	28.6	18.8
7749	25.0	2017-08-30	23.2	17.4	22.933014	77.243744	27.939516	18.522965	7.289264	9.090034	...	0.000000	0.000000	0.000000	37.5237	126.970	19.5844	0.271300	4451.345215	27.8	17.4
7750	NaN	NaN	20.0	11.3	19.794666	58.936283	17.624954	14.272646	2.882580	-13.603212	...	0.000000	0.000000	0.000000	37.4562	126.826	12.3700	0.098475	4329.520508	17.4	11.3
7751	NaN	NaN	37.6	29.9	98.524734	100.000153	38.542255	29.619342	21.857621	213.414006	...	21.621661	15.841235	16.655469	37.6450	127.135	212.3350	5.178230	5992.895996	38.9	29.8

7752 rows × 25 columns

In [200]:
```python
# using dropna() to get rid of rows with missing values
drop_df = df.dropna(axis = 0)
```

In [201]:
```python
drop_df
```

Out[201]:

	station	Date	Present_Tmax	Present_Tmin	LDAPS_RHmin	LDAPS_RHmax	LDAPS_Tmax_lapse	LDAPS_Tmin_lapse	LDAPS_WS	LDAPS_LH	...	LDAPS_PPT2	LDAPS_PPT3	LDAPS_PPT4	lat	lon	DEM	Slope	Solar radiation	Next_Tmax	Next_Tmin
0	1.0	2013-06-30	28.7	21.4	58.255688	91.116364	28.074101	23.006936	6.818887	69.451805	...	0.0	0.0	0.0	37.6046	126.991	212.3350	2.7850	5992.895996	29.1	21.2
1	2.0	2013-06-30	31.9	21.6	52.263397	90.604721	29.850689	24.035009	5.691890	51.937448	...	0.0	0.0	0.0	37.6046	127.032	44.7624	0.5141	5869.312500	30.5	22.5
2	3.0	2013-06-30	31.6	23.3	48.690479	83.973587	30.091292	24.565633	6.138224	20.573050	...	0.0	0.0	0.0	37.5776	127.058	33.3068	0.2661	5863.555664	31.1	23.9
3	4.0	2013-06-30	32.0	23.4	58.239788	96.483688	29.704629	23.326177	5.650050	65.727144	...	0.0	0.0	0.0	37.6450	127.022	45.7160	2.5348	5856.964844	31.7	24.3
4	5.0	2013-06-30	31.4	21.9	56.174095	90.155128	29.113934	23.486480	5.735004	107.965535	...	0.0	0.0	0.0	37.5507	127.135	35.0380	0.5055	5859.552246	31.2	22.5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7745	21.0	2017-08-30	23.1	17.8	24.688997	78.261383	27.812697	18.303014	6.603253	9.614074	...	0.0	0.0	0.0	37.5507	127.040	26.2980	0.5721	4456.024414	27.6	17.7
7746	22.0	2017-08-30	22.5	17.4	30.094858	83.690018	26.704905	17.814038	5.768083	82.146707	...	0.0	0.0	0.0	37.5102	127.086	21.9668	0.1332	4441.803711	28.0	17.1
7747	23.0	2017-08-30	23.3	17.1	26.741310	78.869858	26.352081	18.775678	6.148918	72.058294	...	0.0	0.0	0.0	37.5372	126.891	15.5876	0.1554	4443.313965	28.3	18.1
7748	24.0	2017-08-30	23.3	17.7	24.040634	77.294975	27.010193	18.733519	6.542819	47.241457	...	0.0	0.0	0.0	37.5237	126.909	17.2956	0.2223	4438.373535	28.6	18.8
7749	25.0	2017-08-30	23.2	17.4	22.933014	77.243744	27.939516	18.522965	7.289264	9.090034	...	0.0	0.0	0.0	37.5237	126.970	19.5844	0.2713	4451.345215	27.8	17.4

7588 rows × 25 columns

In [202]:
```python
# resetting dataframe
df = pd.read_csv("Forecast_Data_Set.csv")
df
```

Out[202]:

	station	Date	Present_Tmax	Present_Tmin	LDAPS_RHmin	LDAPS_RHmax	LDAPS_Tmax_lapse	LDAPS_Tmin_lapse	LDAPS_WS	LDAPS_LH	...	LDAPS_PPT2	LDAPS_PPT3	LDAPS_PPT4	lat	lon	DEM	Slope	Solar radiation	Next_Tmax	Next_Tmin
0	1.0	2013-06-30	28.7	21.4	58.255688	91.116364	28.074101	23.006936	6.818887	69.451805	...	0.000000	0.000000	0.000000	37.6046	126.991	212.3350	2.785000	5992.895996	29.1	21.2
1	2.0	2013-06-30	31.9	21.6	52.263397	90.604721	29.850689	24.035009	5.691890	51.937448	...	0.000000	0.000000	0.000000	37.6046	127.032	44.7624	0.514100	5869.312500	30.5	22.5
2	3.0	2013-06-30	31.6	23.3	48.690479	83.973587	30.091292	24.565633	6.138224	20.573050	...	0.000000	0.000000	0.000000	37.5776	127.058	33.3068	0.266100	5863.555664	31.1	23.9
3	4.0	2013-06-30	32.0	23.4	58.239788	96.483688	29.704629	23.326177	5.650050	65.727144	...	0.000000	0.000000	0.000000	37.6450	127.022	45.7160	2.534800	5856.964844	31.7	24.3
4	5.0	2013-06-30	31.4	21.9	56.174095	90.155128	29.113934	23.486480	5.735004	107.965535	...	0.000000	0.000000	0.000000	37.5507	127.135	35.0380	0.505500	5859.552246	31.2	22.5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7747	23.0	2017-08-30	23.3	17.1	26.741310	78.869858	26.352081	18.775678	6.148918	72.058294	...	0.000000	0.000000	0.000000	37.5372	126.891	15.5876	0.155400	4443.313965	28.3	18.1
7748	24.0	2017-08-30	23.3	17.7	24.040634	77.294975	27.010193	18.733519	6.542819	47.241457	...	0.000000	0.000000	0.000000	37.5237	126.909	17.2956	0.222300	4438.373535	28.6	18.8
7749	25.0	2017-08-30	23.2	17.4	22.933014	77.243744	27.939516	18.522965	7.289264	9.090034	...	0.000000	0.000000	0.000000	37.5237	126.970	19.5844	0.271300	4451.345215	27.8	17.4
7750	NaN	NaN	20.0	11.3	19.794666	58.936283	17.624954	14.272646	2.882580	-13.603212	...	0.000000	0.000000	0.000000	37.4562	126.826	12.3700	0.098475	4329.520508	17.4	11.3
7751	NaN	NaN	37.6	29.9	98.524734	100.000153	38.542255	29.619342	21.857621	213.414006	...	21.621661	15.841235	16.655469	37.6450	127.135	212.3350	5.178230	5992.895996	38.9	29.8

7752 rows × 25 columns

In [203]:
```python
columns = ['station', 'Present_Tmax', 'Present_Tmin', 'LDAPS_RHmin', 'LDAPS_RHmax', 'LDAPS_Tmax_lapse', 'LDAPS_Tmin_lapse', 'LDAPS_WS', 'LDAPS_LH', 'LDAPS_PPT2', 'LDAPS_PPT3', 'lat', 'lon', 'DEM', 'Slope', 'Solar radiation', 'Next
```

In [204]:
```python
# using mean() to fill columns in with the mean of said column. Applies to all columns except date
df = df.fillna(df[columns].mean())
```

In [205]:
```python
df
```

Out[205]:

	station	Date	Present_Tmax	Present_Tmin	LDAPS_RHmin	LDAPS_RHmax	LDAPS_Tmax_lapse	LDAPS_Tmin_lapse	LDAPS_WS	LDAPS_LH	...	LDAPS_PPT2	LDAPS_PPT3	LDAPS_PPT4	lat	lon	DEM	Slope	Solar radiation	Next_Tmax	Next_Tmin
0	1.0	2013-06-30	28.7	21.4	58.255688	91.116364	28.074101	23.006936	6.818887	69.451805	...	0.000000	0.000000	0.000000	37.6046	126.991	212.3350	2.785000	5992.895996	29.1	21.2
1	2.0	2013-06-30	31.9	21.6	52.263397	90.604721	29.850689	24.035009	5.691890	51.937448	...	0.000000	0.000000	0.000000	37.6046	127.032	44.7624	0.514100	5869.312500	30.5	22.5
2	3.0	2013-06-30	31.6	23.3	48.690479	83.973587	30.091292	24.565633	6.138224	20.573050	...	0.000000	0.000000	0.000000	37.5776	127.058	33.3068	0.266100	5863.555664	31.1	23.9
3	4.0	2013-06-30	32.0	23.4	58.239788	96.483688	29.704629	23.326177	5.650050	65.727144	...	0.000000	0.000000	0.000000	37.6450	127.022	45.7160	2.534800	5856.964844	31.7	24.3
4	5.0	2013-06-30	31.4	21.9	56.174095	90.155128	29.113934	23.486480	5.735004	107.965535	...	0.000000	0.000000	0.000000	37.5507	127.135	35.0380	0.505500	5859.552246	31.2	22.5
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7747	23.0	2017-08-30	23.3	17.1	26.741310	78.869858	26.352081	18.775678	6.148918	72.058294	...	0.000000	0.000000	0.000000	37.5372	126.891	15.5876	0.155400	4443.313965	28.3	18.1
7748	24.0	2017-08-30	23.3	17.7	24.040634	77.294975	27.010193	18.733519	6.542819	47.241457	...	0.000000	0.000000	0.000000	37.5237	126.909	17.2956	0.222300	4438.373535	28.6	18.8
7749	25.0	2017-08-30	23.2	17.4	22.933014	77.243744	27.939516	18.522965	7.289264	9.090034	...	0.000000	0.000000	0.000000	37.5237	126.970	19.5844	0.271300	4451.345215	27.8	17.4
7750	13.0	NaN	20.0	11.3	19.794666	58.936283	17.624954	14.272646	2.882580	-13.603212	...	0.000000	0.000000	0.000000	37.4562	126.826	12.3700	0.098475	4329.520508	17.4	11.3
7751	13.0	NaN	37.6	29.9	98.524734	100.000153	38.542255	29.619342	21.857621	213.414006	...	21.621661	15.841235	16.655469	37.6450	127.135	212.3350	5.178230	5992.895996	38.9	29.8

7752 rows × 25 columns

In [206]:
```python
df["station"].mean()
```

Out[206]:
```
13.0
```

11.

a)

In [207]:
```python
correlations = df.corr()

#Resource used: https://datatofish.com/correlation-matrix-pandas/
```

In [208]:
```python
correlations
```

Out[208]:

	station	Present_Tmax	Present_Tmin	LDAPS_RHmin	LDAPS_RHmax	LDAPS_Tmax_lapse	LDAPS_Tmin_lapse
station	1.000000	0.112655	0.131646	-0.066733	-0.166984	0.069107	0.104758
Present_Tmax	0.112655	1.000000	0.618760	-0.204571	-0.301787	0.570129	0.623496
Present_Tmin	0.131646	0.618760	1.000000	0.123143	-0.015140	0.464151	0.762993
LDAPS_RHmin	-0.066733	-0.204571	0.123143	1.000000	0.579141	-0.564580	0.089476
LDAPS_RHmax	-0.166984	-0.301787	-0.015140	0.579141	1.000000	-0.373404	-0.114143
LDAPS_Tmax_lapse	0.069107	0.570129	0.464151	-0.564580	-0.373404	1.000000	0.654021
LDAPS_Tmin_lapse	0.104758	0.623496	0.762993	0.089476	-0.114143	0.654021	1.000000
LDAPS_WS	0.005277	-0.121820	-0.034675	0.294361	0.135333	-0.311996	-0.130035
LDAPS_LH	-0.133357	0.135463	-0.009258	-0.070858	0.238579	0.048010	-0.134761
LDAPS_CC1	0.006953	-0.313370	0.085190	0.613818	0.436652	-0.438439	0.010901
LDAPS_CC2	0.003413	-0.214864	0.091057	0.745443	0.391330	-0.523619	0.047727
LDAPS_CC3	0.000697	-0.144741	-0.002797	0.689679	0.226957	-0.541327	-0.044018
LDAPS_CC4	0.006156	-0.141435	-0.044725	0.514075	0.130619	-0.429539	-0.080335
LDAPS_PPT1	-0.002723	-0.109553	0.113837	0.262665	0.268449	-0.111065	0.038056
LDAPS_PPT2	-0.008250	-0.099449	0.068176	0.390967	0.229050	-0.242122	0.000870
LDAPS_PPT3	-0.012461	-0.120367	-0.046126	0.240642	0.134607	-0.188115	-0.088575
LDAPS_PPT4	-0.010452	-0.100107	-0.063518	0.168595	0.117853	-0.160273	-0.094655
lat	-0.237502	-0.052615	-0.078475	0.087099	0.195798	-0.042093	-0.096257
lon	-0.118704	0.009020	-0.043461	-0.076237	0.025667	0.090666	-0.026701
DEM	-0.255830	-0.187121	-0.250275	0.102115	0.177168	-0.178895	-0.195455
Slope	-0.090061	-0.105809	-0.146339	0.123744	0.219599	-0.162332	-0.185849
Solar radiation	-0.019001	-0.025439	0.053578	0.244080	0.149261	0.047971	0.159975
Next_Tmax	0.107983	0.610348	0.471806	-0.437898	-0.283738	0.827189	0.586983
Next_Tmin	0.128293	0.622741	0.796670	0.095019	-0.071997	0.586868	0.880176

	LDAPS_WS	LDAPS_LH	LDAPS_CC1	...	LDAPS_PPT2	LDAPS_PPT3	LDAPS_PPT4	lat	lon	DEM	Slope	Solar radiation	Next_Tmax	Nex
station	0.005277	-0.133357	0.006953	...	-0.008250	-0.012461	-0.010452	-0.237502	-0.118704	-0.255830	-0.090061	-0.019001	0.107983	0.
Present_Tmax	-0.121820	0.135463	-0.313370	...	-0.099449	-0.120367	-0.100107	-0.052615	0.009020	-0.187121	-0.105809	-0.025439	0.610348	0.
Present_Tmin	-0.034675	-0.009258	0.085190	...	0.068176	-0.046126	-0.063518	-0.078475	-0.043461	-0.250275	-0.146339	0.053578	0.471806	0.
LDAPS_RHmin	0.294361	-0.070858	0.613818	...	0.390967	0.240642	0.168595	0.087099	-0.076237	0.102115	0.123744	0.244080	-0.437898	0.
LDAPS_RHmax	0.135333	0.238579	0.436652	...	0.229050	0.134607	0.117853	0.195798	0.025667	0.177168	0.219599	0.149261	-0.283738	-0.
LDAPS_Tmax_lapse	-0.311996	0.048010	-0.438439	...	-0.242122	-0.188115	-0.160273	-0.042093	0.090666	-0.178895	-0.162332	0.047971	0.827189	0.
LDAPS_Tmin_lapse	-0.130035	-0.134761	0.010901	...	0.000870	-0.088575	-0.094655	-0.096257	-0.026701	-0.195455	-0.185849	0.159975	0.586983	0.
LDAPS_WS	1.000000	0.006711	0.289445	...	0.191886	0.161684	0.144085	0.036657	-0.059466	0.191053	0.171629	0.121759	-0.342729	-0.
LDAPS_LH	0.006711	1.000000	-0.147296	...	-0.077545	0.016102	0.019979	0.133221	0.025822	0.055133	0.086781	-0.047384	0.157616	-0.
LDAPS_CC1	0.289445	-0.147296	1.000000	...	0.389453	0.168742	0.091722	-0.008170	-0.006876	-0.014159	-0.020853	0.218949	-0.456539	-0.
LDAPS_CC2	0.261090	-0.264280	0.779683	...	0.478472	0.240134	0.159548	-0.001505	-0.003599	-0.012886	-0.018161	0.170903	-0.495092	0.
LDAPS_CC3	0.242991	-0.245583	0.517493	...	0.337569	0.339800	0.302357	0.004128	0.016119	0.000701	-0.003355	0.119642	-0.519227	-0.
LDAPS_CC4	0.220533	-0.172766	0.361046	...	0.288195	0.267494	0.376049	-0.009151	-0.002489	-0.007370	-0.014530	0.120335	-0.450262	-0.
LDAPS_PPT1	0.152587	-0.012542	0.446328	...	0.383831	0.033004	0.024626	0.000568	-0.001231	0.011345	0.016389	0.088273	-0.119867	0.
LDAPS_PPT2	0.191886	-0.077545	0.389453	...	1.000000	0.221172	0.139555	0.019564	0.018600	0.008357	0.010214	0.110182	-0.178383	-0.
LDAPS_PPT3	0.161684	0.016102	0.168742	...	0.221172	1.000000	0.309076	0.035110	0.021662	0.017186	0.025333	-0.054071	-0.194143	-0.
LDAPS_PPT4	0.144085	0.019979	0.091722	...	0.139555	0.309076	1.000000	0.013275	0.044963	-0.003640	0.000118	0.035060	-0.180297	-0.
lat	0.036657	0.133221	-0.008170	...	0.019564	0.035110	0.013275	1.000000	0.289029	0.033165	0.074763	0.038513	-0.053510	-0.
lon	-0.059466	0.025822	-0.006876	...	0.018600	0.021662	0.044963	0.289029	1.000000	0.007587	0.038070	0.002182	0.003052	-0.
DEM	0.191053	0.055133	-0.014159	...	0.008357	0.017186	-0.003640	0.033165	0.007587	1.000000	0.783576	0.056224	-0.172538	-0.
Slope	0.171629	0.086781	-0.020853	...	0.010214	0.025333	0.000118	0.074763	0.038070	0.783576	1.000000	0.028199	-0.102725	-0.
Solar radiation	0.121759	-0.047384	0.218949	...	0.110182	-0.054071	0.035060	0.038513	0.002182	0.056224	0.028199	1.000000	0.008795	0.
Next_Tmax	-0.342729	0.157616	-0.456539	...	-0.178383	-0.194143	-0.180297	-0.053510	0.003052	-0.172538	-0.102725	0.008795	1.000000	0.
Next_Tmin	-0.096576	-0.055440	-0.011094	...	-0.008058	-0.062622	-0.104300	-0.083533	-0.046603	-0.246906	-0.146882	0.120212	0.621427	1.

24 rows × 24 columns

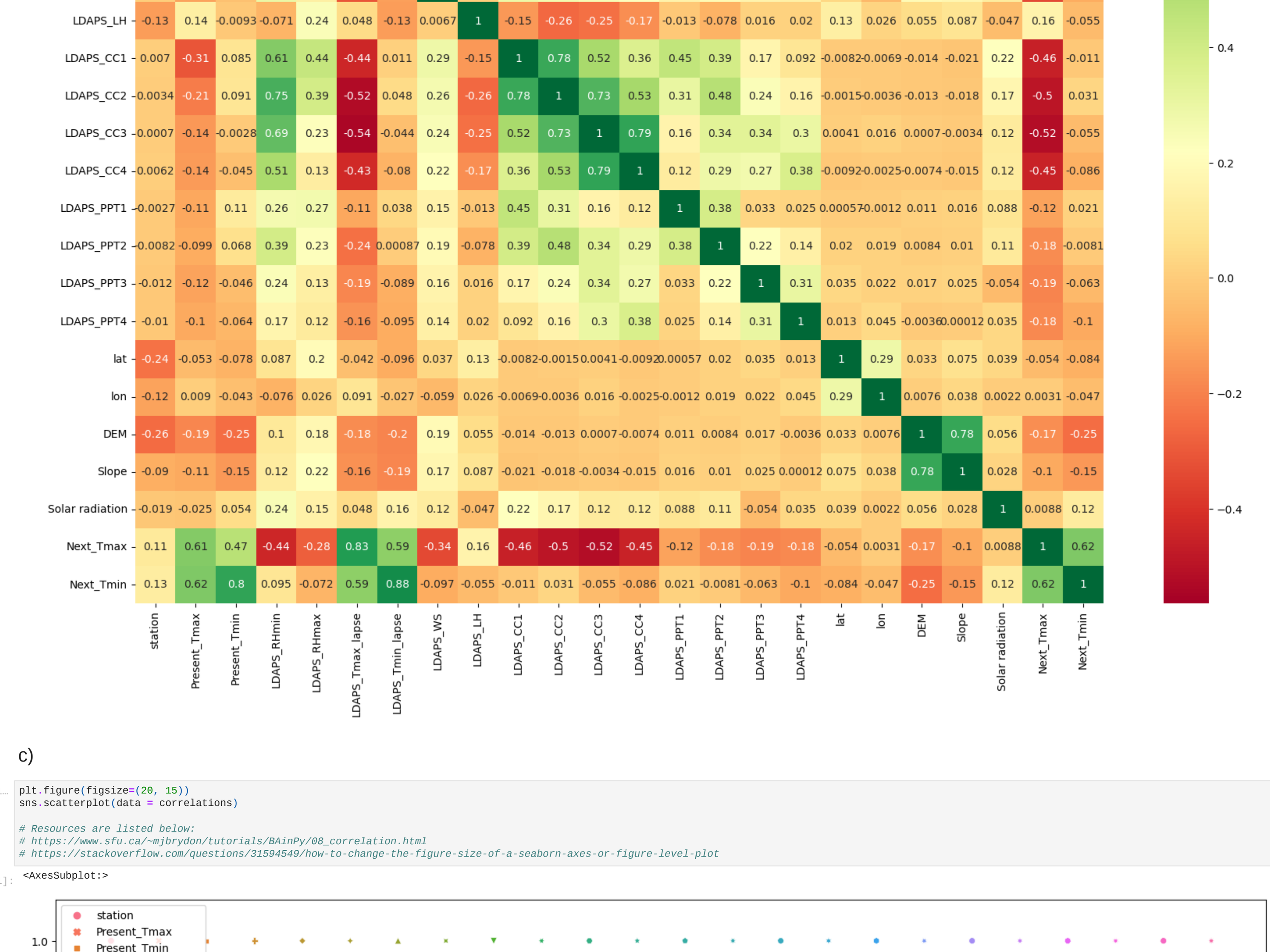
b)

In [209]:
```python
import matplotlib.pyplot as plt
import seaborn as sns
```

In [210]:
```python
plt.figure(figsize=(20, 15))
sns.heatmap(correlations, annot=True, cmap="RdYlGn")

# Resource used: https://blog.quantinsti.com/creating-heatmap-using-python-seaborn/
```

Out[210]:
```
<AxesSubplot:>
```



c)

In [211]:
```python
plt.figure(figsize=(20, 15))
sns.scatterplot(data = correlations)

# Resources are listed below:
# https://www.sfu.ca/~mjbrydon/tutorials/BAinPy/08_correlation.html
# https://stackoverflow.com/questions/31594549/how-to-change-the-figure-size-of-a-seaborn-axes-or-figure-level-plot
```

Out[211]:
```
<AxesSubplot:>
```

12.

In [212]:
```python
import numpy as np

upper = correlations.where(np.triu(np.ones(correlations.shape), k = 1).astype(bool))

removed_corr = [column for column in upper.columns if any((upper[column] > 0.7) | (upper[column] < -0.7))]

correlations.drop(removed_corr, axis = 1, inplace = True)

# Resources used:
# https://stackoverflow.com/questions/29294983/how-to-calculate-correlation-between-all-columns-and-remove-highly-correlated-on
# https://www.andrews.edu/~calkins/math/edrm611/edrm05.htm
```

In [213]:
```python
removed_corr
```

Out[213]:
```
['LDAPS_Tmin_lapse',
 'LDAPS_CC2',
 'LDAPS_CC3',
 'LDAPS_CC4',
 'Slope',
 'Next_Tmax',
 'Next_Tmin']
```

The variables that had a magnitude of absolute(0.7) and over were removed. This was because +/- 0.7 and upward (to +/- 1.0) incidates a very high correlation. This means that the 7 columns listed above were dropped for this reason. If we look above at the correlations chart in 11a, we can see that these columns had strong correlations with others.