Name of the Program	program_09.py
Course Name	ABE651 Environmental Informatics
Assignment	09-Data Quality Checking
Name of Program Creator	Mukhamad Suhermanto

Description

The program imports the text file as a dataframe and the date contained in the file is used as index. Four checks are carried out using the program for the data quality check. In addition, the program provides the detail of the checked/evaluated data, their comparison in the form of text file and PNG visualization.

Input	*DataQualityChecking.txt* It is a time series data containing four variables from 1915-01-01 to 1916-12-31.
Output	4 Comparison Plots quality checked data and failure checks in the form of csv file
Checking	1. Replace the No Data (-999) values with NaN 2. Removes Gross Error in Precipitation, Maximum Temperature, Minimum Temperature and Wind Speed and replaces with with NaN 3. Swapping of Maximum and Minimum Temperature when needed 4. Replacing Daily Temperature > 25 ^{0}C with NaN

Some of the excerpts from the raw data and processed can be seen at the Figure below:

```
Raw data.....
             Precip    Max Temp    Min Temp  Wind Speed
count  731.000000  731.000000  731.000000  731.000000
mean     0.288098   14.167227    0.548413    2.904172
std     53.773216   54.738379   53.477046    1.597814
min   -999.000000 -999.000000 -999.000000   -2.500000
25%      0.000000    6.735000   -4.080000    2.045000
50%      0.000000   18.560000    3.610000    2.910000
75%      2.237500   26.195000   11.875000    3.600000
max    279.000000  194.800000   26.100000   39.900000

Missing values removed.....
             Precip    Max Temp    Min Temp  Wind Speed
count  729.000000  729.000000  729.000000  731.000000
mean     3.029630   16.946835    3.290658    2.904172
std     12.191541   13.293172   10.740052    1.597814
min     -3.475000  -10.080000  -82.600000   -2.500000
25%      0.000000    6.850000   -4.030000    2.045000
50%      0.000000   18.580000    3.680000    2.910000
75%      2.275000   26.200000   11.900000    3.600000
max    279.000000  194.800000   26.100000   39.900000
```

Figure 1 Raw data and -999 to NaN Replacement

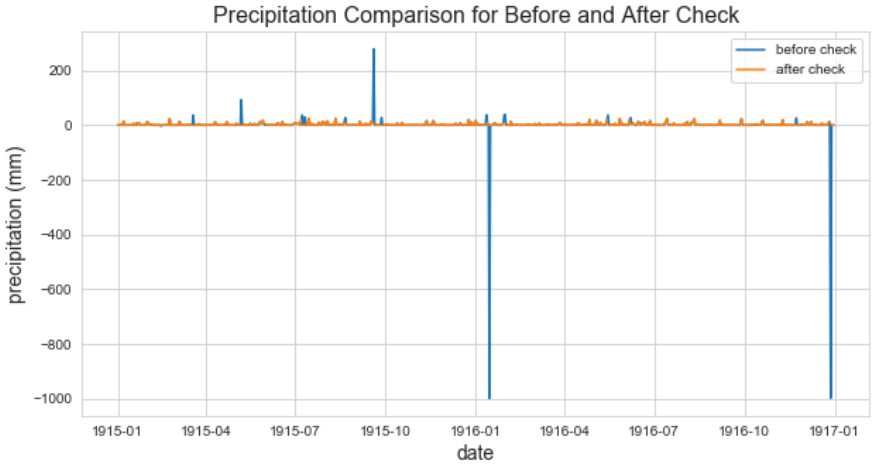
```
Check for gross errors complete.....
             Precip    Max Temp    Min Temp  Wind Speed
count  714.000000  715.000000  727.000000  729.000000
mean     2.070588   16.329263    3.512641    2.860837
std      4.291815   11.311690    9.879578    0.798721
min      0.000000  -10.080000  -18.630000    1.500000
25%      0.000000    6.680000   -3.935000    2.050000
50%      0.000000   18.310000    3.680000    2.910000
75%      1.950000   25.880000   11.910000    3.600000
max     24.050000   34.960000   26.100000    4.280000

Check for swapped temperatures complete.....
             Precip    Max Temp    Min Temp  Wind Speed
count  714.000000  715.000000  727.000000  729.000000
mean     2.070588   16.379790    3.462948    2.860837
std      4.291815   11.278356    9.852212    0.798721
min      0.000000  -10.080000  -18.630000    1.500000
25%      0.000000    6.715000   -3.935000    2.050000
50%      0.000000   18.360000    3.610000    2.910000
75%      1.950000   25.930000   11.875000    3.600000
max     24.050000   34.960000   23.900000    4.280000
```

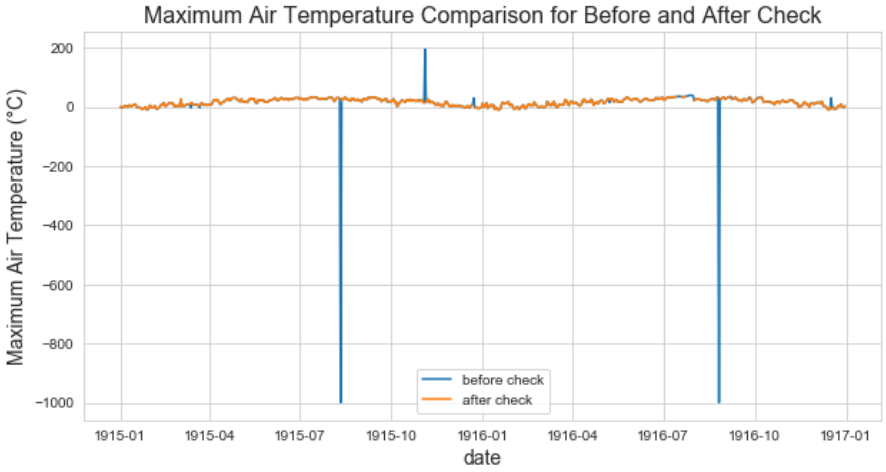
Figure 2 Gross Error and Swapping Temperature

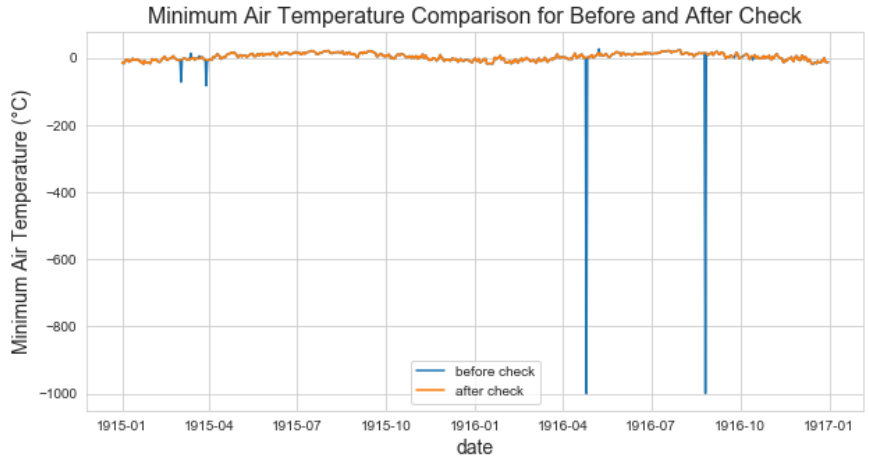
Table 1 Data Checking Summary

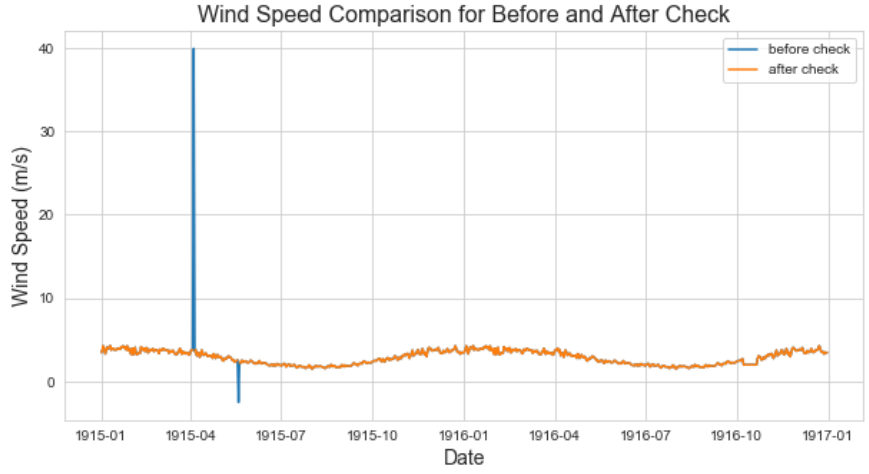
0	NaN	Precip	Max Temp	Min Temp	Wind Speed
1	1. No Data	2.0	2.0	2.0	0.0
2	2. Gross Error	15.0	14.0	2.0	2.0
3	3. Swapped	0.0	4.0	4.0	0.0
4	4. Range Fail	0.0	5.0	5.0	0.0

Table 1 shows the summary of the processed data checking (4 types of checking). The comparisons are shown in the plots of


Precipitation Comparison for Before and After Check
before check
after check
precipitation (mm)
200
0
−200
−400
−600
−800
−1000
1915-01
1915-04
1915-07
1915-10
1916-01
1916-04
1916-07
1916-10
1917-01
date


Maximum Air Temperature Comparison for Before and After Check
Maximum Air Temperature (°C)
200
0
−200
−400
−600
−800
−1000
before check
after check
1915-01
1915-04
1915-07
1915-10
1916-01
1916-04
1916-07
1916-10
1917-01
date


Minimum Air Temperature Comparison for Before and After Check
Minimum Air Temperature (°C)
0
−200
−400
−600
−800
−1000
before check
after check
1915-01
1915-04
1915-07
1915-10
1916-01
1916-04
1916-07
1916-10
1917-01
date


Wind Speed Comparison for Before and After Check
Wind Speed (m/s)
40
30
20
10
0
before check
after check
1915-01
1915-04
1915-07
1915-10
1916-01
1916-04
1916-07
1916-10
1917-01
Date