|  |  |
| --- | --- |
| **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:

|  |  |
| --- | --- |
| A close up of text on a black background  Description automatically generated  Figure 1 Raw data and -999 to NaN Replacement | A close up of text on a black background  Description automatically generated  Figure 2 Gross Error and Swapping Temperature |

Table 1 Data Checking Summary

A picture containing clock, black, room, street

Description automatically generated

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

|  |
| --- |
| A screenshot of a cell phone  Description automatically generated |
| A close up of text on a white background  Description automatically generated |
| A close up of text on a white background  Description automatically generated |
| A picture containing table, man  Description automatically generated |