Metadata

Created April 4th 2020
Author : Ankit Ghanghas

Script Name: program_09.py
Input File : “DataQualityChecking.txt”
Ouput Files : “corrected_data.csv”, “failed_checks_summar.csv”, “precip_plot.png”, “wind_plot.png”, “tmax_plot.png”, “tmin_plot.png”.

Input File Description.

“DataQualityChecking.txt” is a daily time series data set containing Date, Precipitation, Max Temperature, Min Temperature and Wind Speed measurements of a location(location not known). The data is a record from 1915-01-01 to 1916-12-31.

Description of Script.

This script performs a basic data quality test on the input file and keeps track of how many changes are made at each step of the program.

The data quality is done in 4 main steps.
1. Check for no data values and change all no data values (represented by -999.00 initially) by NaN. This ensure that -999.00 does not interfere with other statistics like mean that may be calculated for the data.

2. Check for gross error in the data. The values of precipitation>25 and <0 , values of Temperature>35 and Temperature<-25, values of wind speed>10 and <0 are all set to NaN as these conditions are not feasible in the location where the data has been collected.

3. Check for values where minimum temperature is greater than max temperature. In such cases the values are interchanged. This is done because in no case Minimum temperature can be more than Maximum temperature.

4. Check for if the temperature range is greater than 25. In such cases the values are set to NaN. This condition is kept in accordance with acceptable temperature trends for the location.

The corrected data after all the checks is stored as corrected_data.csv in the present working directory
The Track of how many changes have been made is saved as failed_checks_summar.csv (tab seperated file) in present working directory
Finally a plots of Before Correction and After all correction of all the parameters are saved sepetately as precip_plot.png, wind_plot.png, tmax_plot.png and tmin_plot.png in the working directory.

	Precip	Max Temp	Min Temp	Wind Speed
1. No Data	2	2	2	0
2. Gross Error	15	14	2	2
3. Swapped	0	4	4	0
4. Range	0	5	5	0

Table 1: This table summarizes the number of changes made by each data quality check

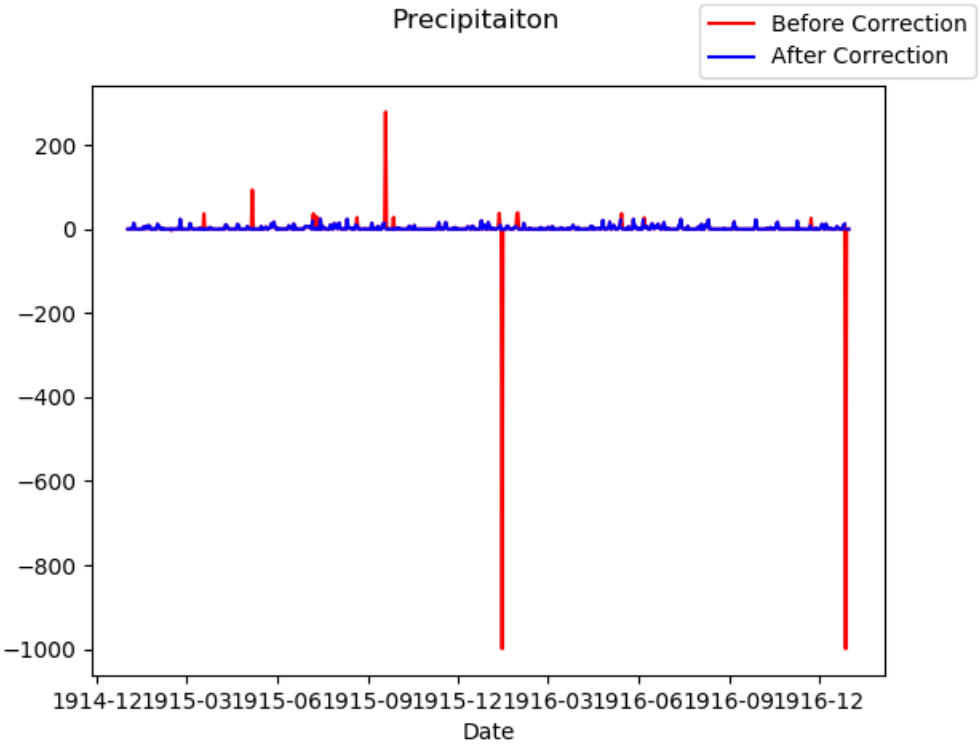

Figure1. This Figure shows the Precipitation data before and after all the checks have been applied. The x axis represents the date and y-axis indicates the value of precipitation.

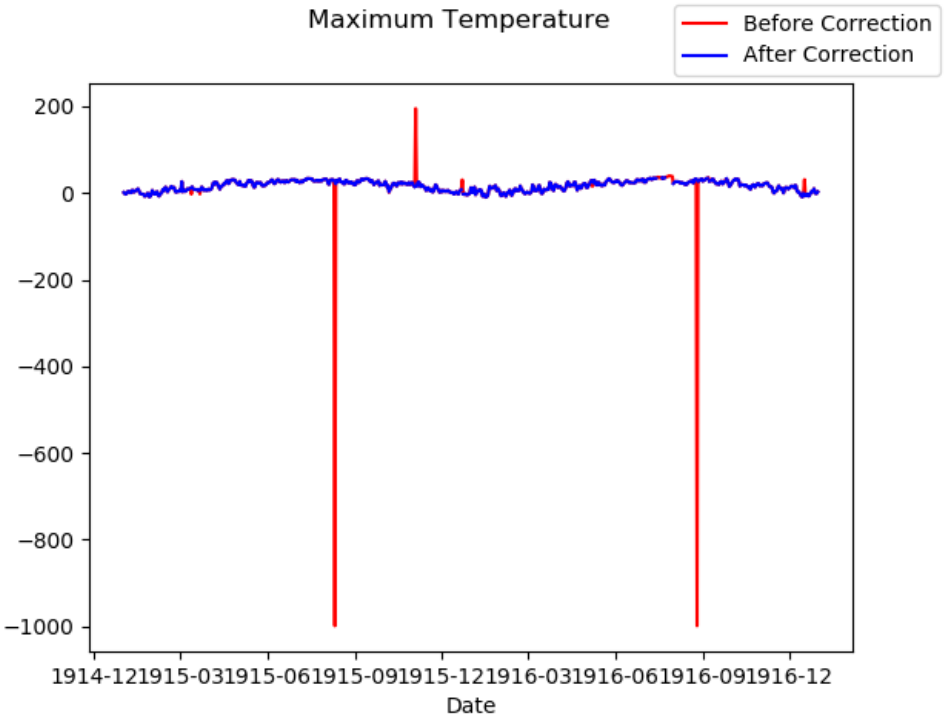

Figure 2. This figure compares the Maximum Temperature before and after all the data checks have been applied.

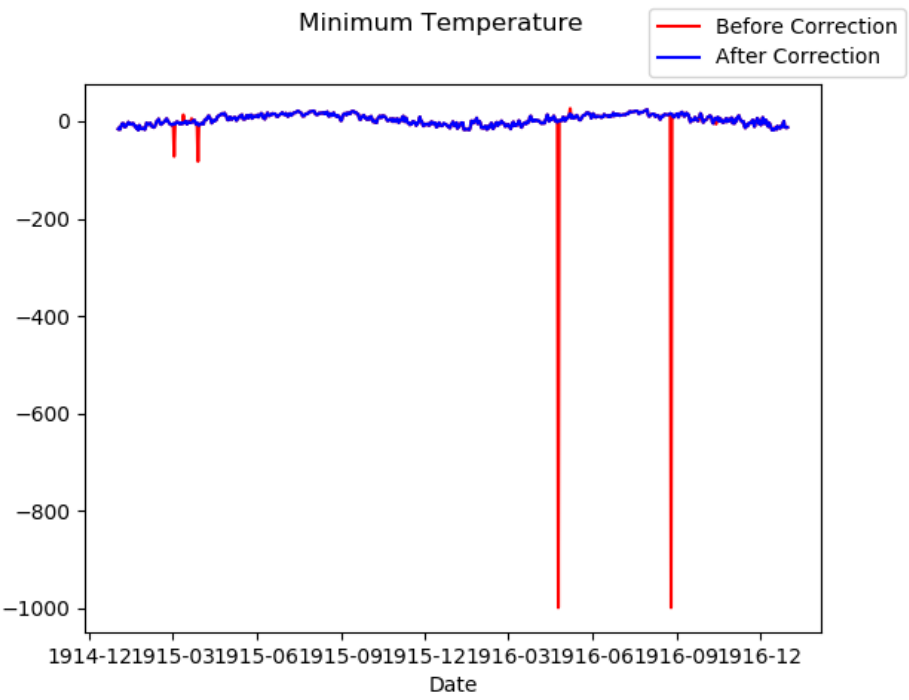

Figure 3. This figure compares the Minimum Temperature before and after all the data checks are applied

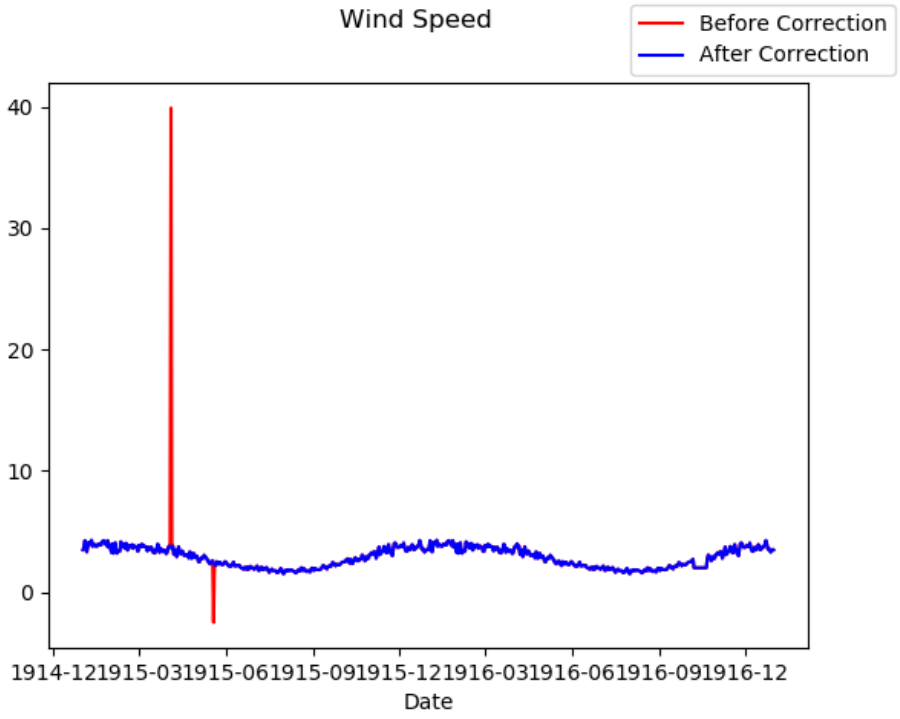

Figure 4. This figure compares the Wind Speed before and after all the data checks are applied