

Assignment 09
Data Quality Checking

Program name: Program_09.py

Program Creator: Asmita Gautam

Description:

This program completes the data quality checking of the meteorological dataset which has data on date, precipitation(mm), maximum temperature (Celsius), minimum temperature (Celsius), and wind speed(m/s).

Data Quality Checking analysis:

1. Removing the no data values/Check 01: The no data values (-999) on the data are replaced by nan. The number of data values removed through check 01 were 2 from precipitation, 2 from maximum temperature, 2 from minimum temperature. The count, minimum, range, mean are altered by removing the no data points on the variables like precipitation, minimum and maximum temperature.

2. Removing the data out of range/Check 02 Gross Error : The data values which are out of range are removed. The out of range of the variables considered are less than 0 and more than 25 for precipitation, less than -25 and greater than 35 for temperature and less than 0 and more than 10 for wind speed. The number of data values removed through Check 02 as gross error were 15 from precipitation, 14 from maximum temperature, 2 from minimum temperature and 2 from wind speed. Count, maximum, minimum, range, mean and other parameters are altered by check02.

3. Check 03 Swap/ Removing the maximum and minimum temp when max temperature is greater than minimum temp of the daily observation : The number of data values removed from Check 03 were 4 for both maximum and minimum temperature.

It helps to make the mean temperature realistic.

4. Check 04 Range fail : If the range of the minimum and maximum temperature exceeds 25 then both maximum and minimum temperature value of the daily observation is replaced by nan. The number of data values removed from check 04 were 5 for both maximum and minimum temperature.

It helps to make the mean temperature realistic

Table 1: No of the removed data values from each check for the given variables

	Precipitation	Max temp	Minimum temp	Wind speed
Check01: No data	2	2	2	0
Check02: Gross Error	15	14	2	2
Check03: Swap	0	4	4	0
Check04: Range fail	0	5	5	0

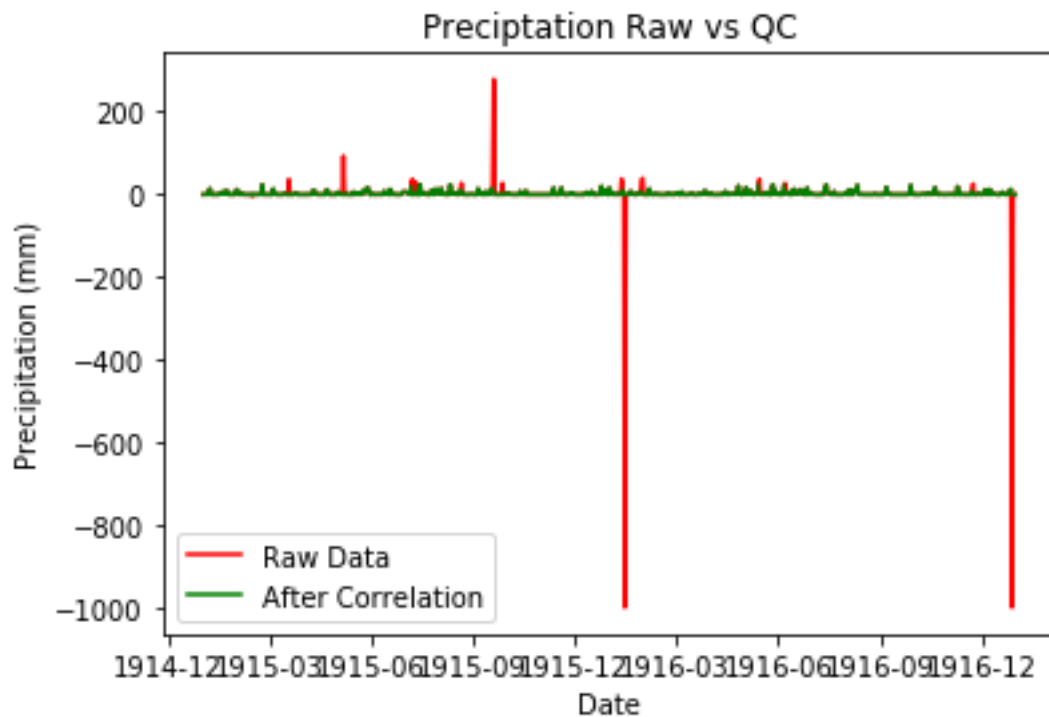


Fig 1: Precipitation before and after data quality checking

Figure 1 indicates the removal of the errors as displayed in the red and the final data on green

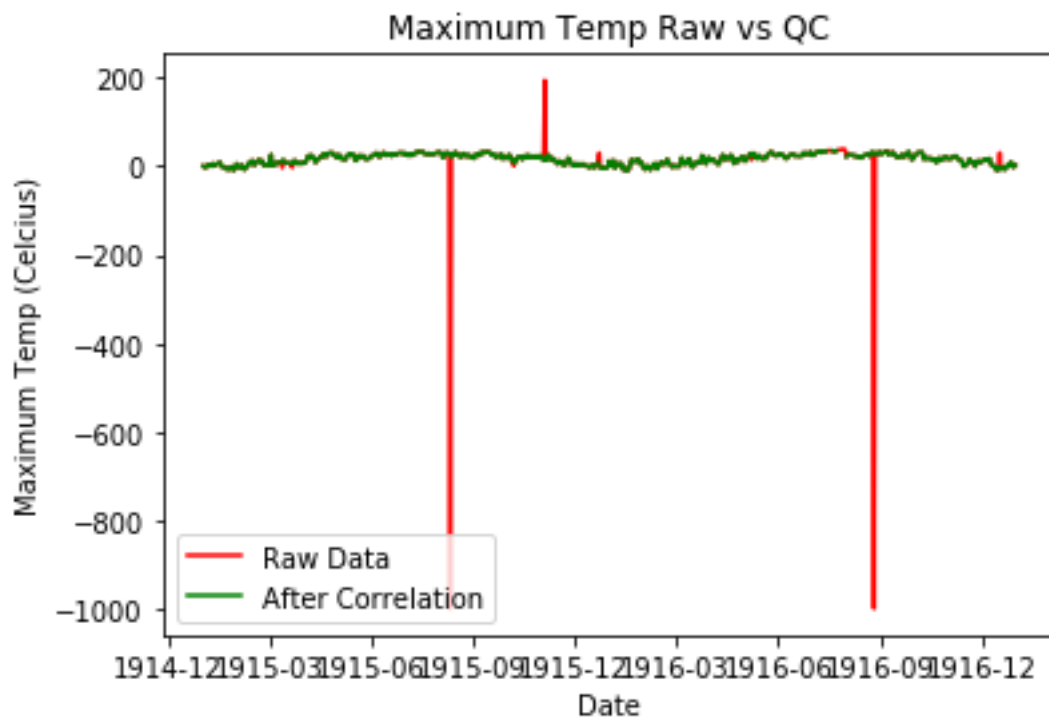


Fig 2: Maximum temperature comparison before and after data quality checking

The figure indicates the removal of the data values indicated by red lines and the final data sets indicated by green lines

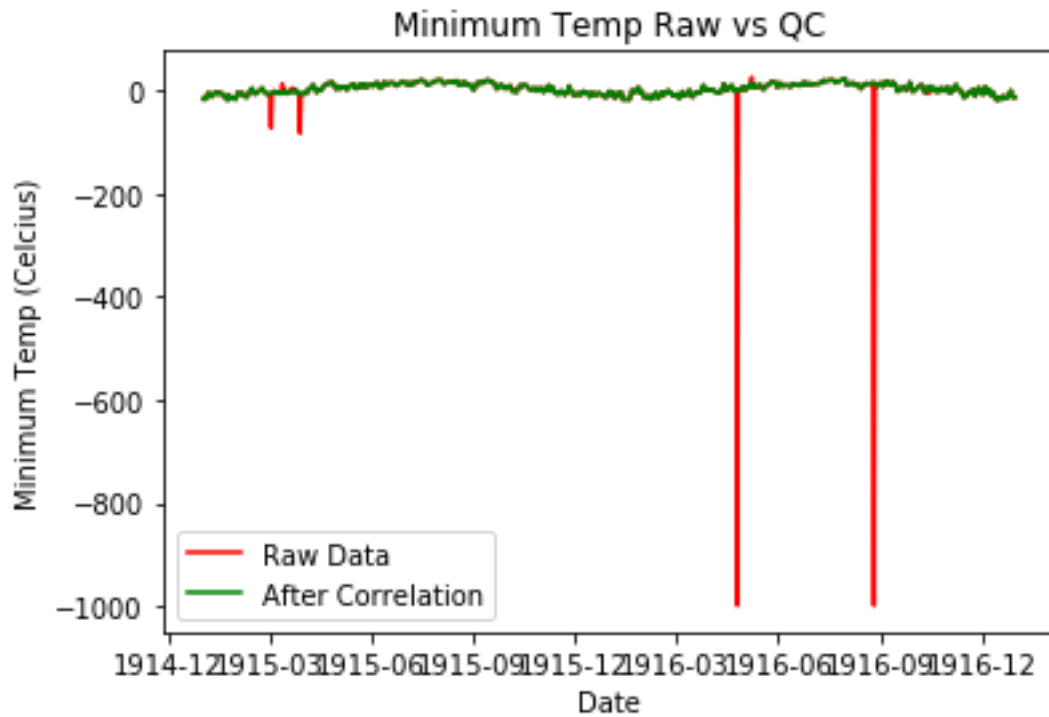


Fig 3: Minimum temperature comparison before and after data quality checking

The given fig3 indicates the few red data values which are removed after data quality checking indicated in green lines.

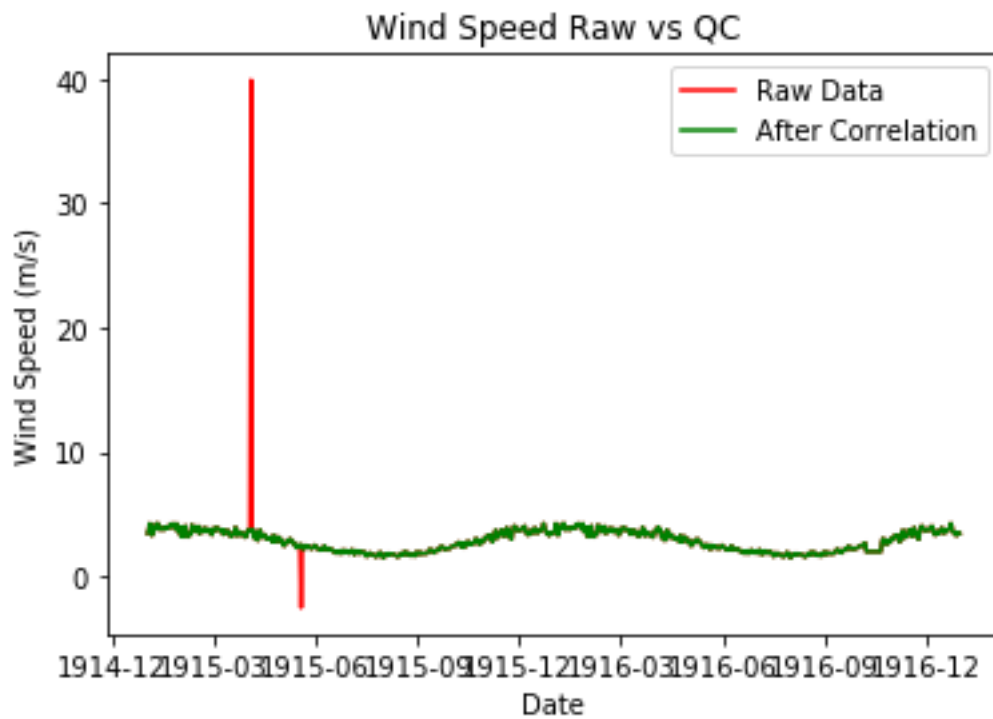


Fig 4: Wind Speed comparison before and after data quality checking.

The given figure clearly indicates the 2 data points (in red color) which are removed from data quality checking.