Author: Varun Aggarwal
Github: https://github.com/Environmental-Informatics/09-data-quality-checking-aggarw82\
Script: Program_09.py
Last Updated: 04/25/2020

The script performs data quality checks on meteorological data. If a data field fails a check, the value is deleted from the analysis. Before and after values are graphed using python matplotlib library. A table highlights the overall count of changes performed on each data field. This table is saved in Stats_for_QDC.txt. In the end, the clean data is saved in the orignal format to file: Clean_Data.txt.

Data Quality Checks:

Check 1: Removes No Data values
Replace all values of -999 in this file with the NumPy NaN values

Check 2: Check for gross errors
Apply the following error thresholds: $0 \leq P \leq 25$; $-25 \leq T \leq 35$, $0 \leq WS \leq 10$ and replace them with NaN

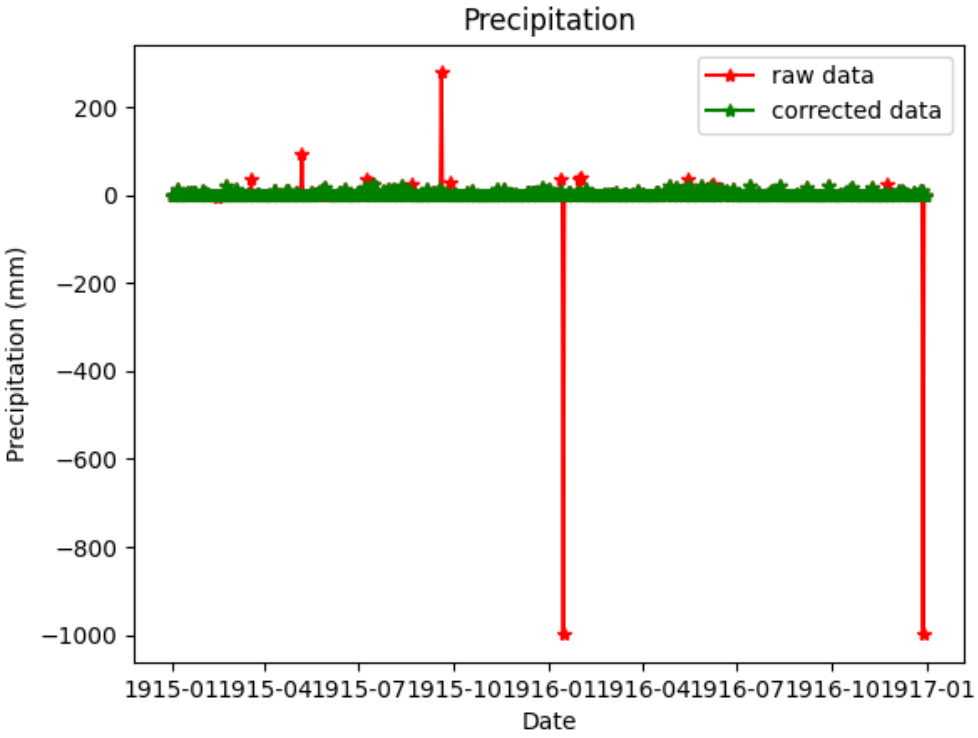
Check 3: Swap Max Temp and Min Temp when Max Temp is less than Min Temp

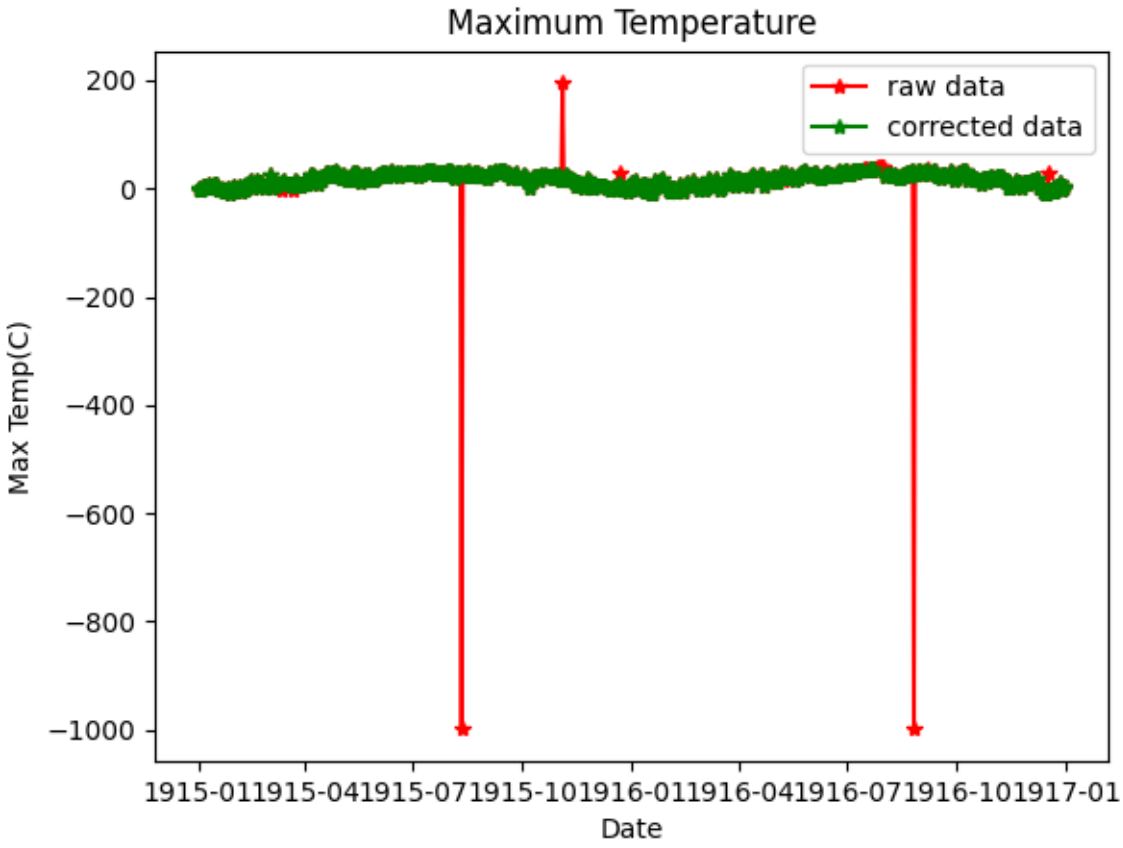
Check 4: Check for daily temperature range exceedence
Identify days with temperature range (Max Temp minus Min Temp) greater than 25°C. When range is exceeded replace both Tmax and Tmin with NaN

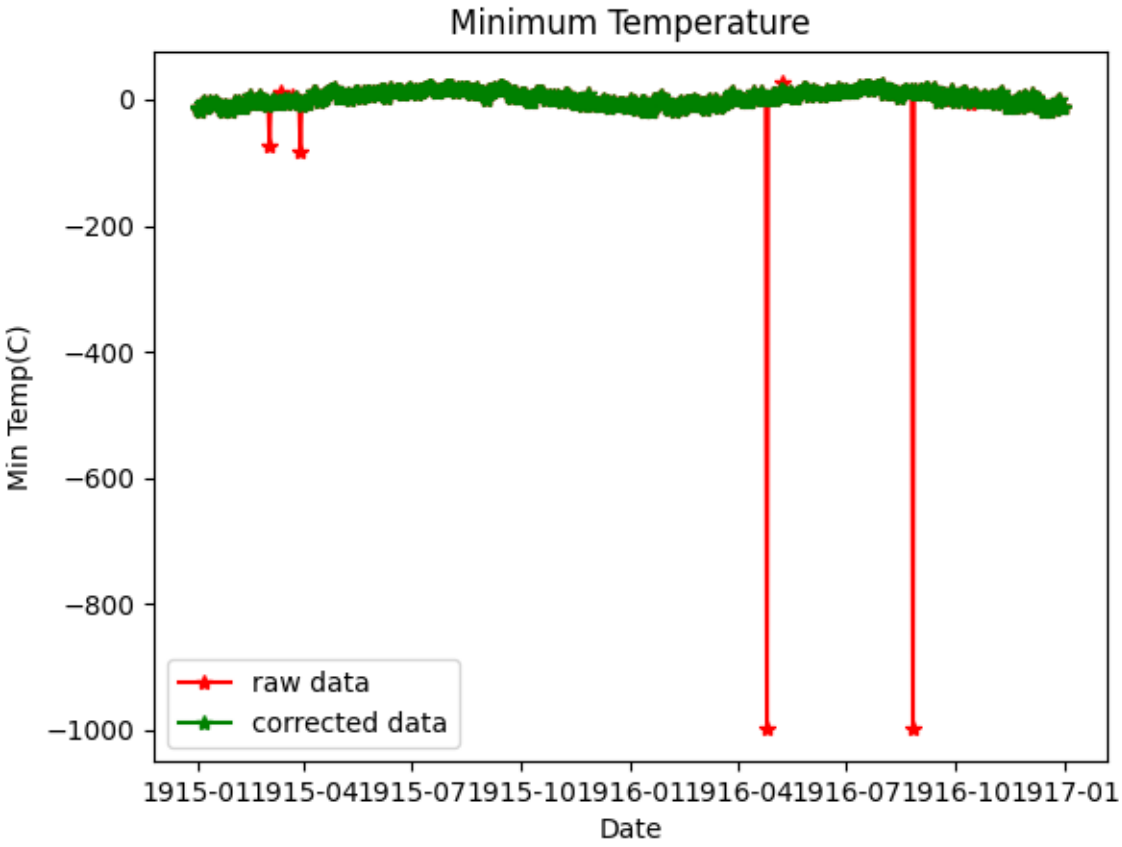
Statistics of Changes

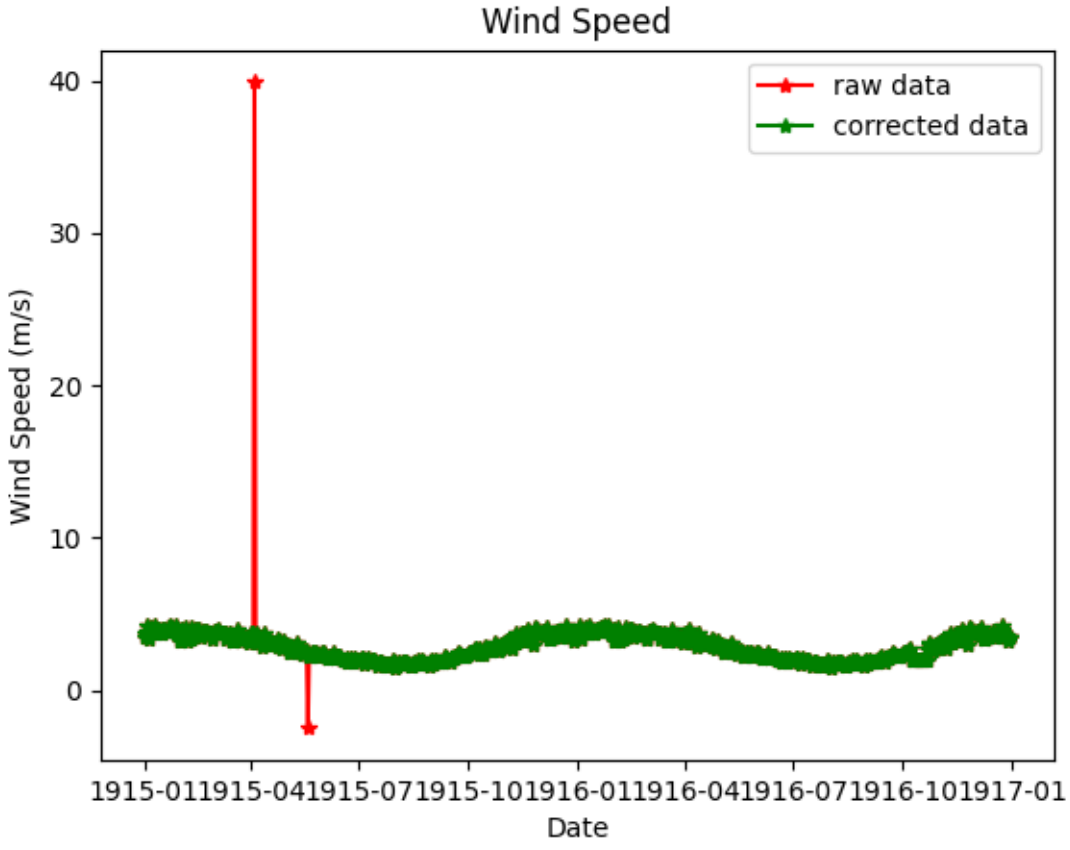
	Precip	Max Temp	Min Temp	Wind Speed
1. No Data	2	2	2	0
2. Gross Error	15	14	2	2
3. Swapped	0	4	4	0
4. Range Fail	0	5	5	0

Graphs Highlighting Changes






Minimum Temperature
0
−200
−400
−600
−800
−1000
Min Temp(C)
raw data
corrected data
1915-011915-041915-071915-101916-011916-041916-071916-101917-01
Date


Wind Speed
40
30
20
10
0
Wind Speed (m/s)
raw data
corrected data
1915-011915-041915-071915-101916-011916-041916-071916-101917-01
Date