Metadata

Description of Program:

The program is called program\_09.py and it was created by Avnika Manaktala as part of a submission for Assignment 9 of ABE651: Environmental Informatics class. The program imports a dataset containing precipitation, minimum temperature, maximum temperature and wind speed and performs data quality checking on the data frame. The program outputs the processed data table along with a table containing the type of error fixed/found for each variable. Plots are then generated showing a comparison between the raw data and processed data.

Summary of Data Quality Checks:

1. Removing NoData values: The program removes all values of -999 in the dataframe. This removes 2 values each from the columns Precip, Max Temp and Min Temp but does not remove any values from Wind Speed (Table 1).
2. Checking for gross errors: Ensures precipitation is between 0 and 25 mm, temperature values are between -25C and 35C and wind speed is between 0 and 10m/s. This removes 15 values from Precip, 14 from Max Temp, 2 from Min Temp and 2 from Wind Speed (Table 1).
3. Check that Max Temp is always greater than Min Temp: Swapping values where Max Temp is Greater than Min Temp. This swap temperature values of 4 recordings. (Table 1).
4. Check temperature range. Checking both Min and Max Temp if they exceed 25C and if they do, they are replaced with NaN. 5 records are discarded due to this check (Table 1).

Correction Count Table:

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

*Table 1: Replaced Value Information*

Processed data Plots (variable correction):

![A screenshot of a cell phone

Description automatically generated]()

*Figure 1: Variable correction plot for precipitation variable. 2 no data values and 15 gross errors are removed.*

*![A screenshot of a map

Description automatically generated]()*

*Figure 2: Variable correction plot for Max Temp variable. Several values removed, mostly due to gross errors (15) and range fail (5).*

*![A screenshot of a cell phone

Description automatically generated]()*

*Figure 3: Variable correction plot for Min Temp variable. Maximum values removed due to range fail (5) and some were swapped with Max Temp values (4).*

*![A screenshot of a cell phone

Description automatically generated]()*

*Figure 4: Variable correction plot for Wind Speed variable. Only 2 values removed due to gross errors.*