**Cleaning Data**

NOAA data Station list:

1. NEW PLYMOUTH AWS, NZ
2. INVERCARGILL AIRPOR, NZ
3. RAOUL ISL KERMADEC, NZ
4. GISBORNE AERODROME, NZ
5. CHRISTCHURCH INTERNATIONAL, NZ
6. HOKITIKA AERODROME, NZ
7. KAITAIA, NZ
8. PARAPARAUMU AWS, NZ
9. CHATHAM ISLANDS AWS, NZ
10. TARA HILLS, NZ
11. AUCKLAND AERO AWS, NZ
12. KAIKOURA, NZ

|  |  |  |  |
| --- | --- | --- | --- |
| STATION NAME & ID | Start | End | Coverage |
| **AUCKLAND** AERO AWS, NZ  GHCND:NZM00093110 | 1994-08-02 | 2023-10-04 | 100% |
| CHATHAM ISLANDS AWS, NZ  GHCND:NZ000939870 | 1956-10-01 | 2005-12-31 | 97% |
| CHRISTCHURCH INTERNATIONAL, NZ  GHCND:NZM00093781 | 1954-03-02 | 2023-10-04 | 95% |
| GISBORNE AERODROME, NZ  GHCND:NZ000093292 | 1962-02-01 | 2023-10-04 | 97% |
| HOKITIKA AERODROME, NZ  GHCND:NZ000936150 | 1964-01-01 | 2023-10-04 | 97% |
| INVERCARGILL AIRPOR, NZ  GHCND:NZ000093844 | 1948-05-31 | 2023-10-04 | 98% |
| KAIKOURA, NZ  GHCND:NZM00093678 | 1997-01-01 | 2023-10-04 | 99% |
| KAITAIA, NZ  GHCND:NZ000093012 | 1965-12-31 | 2023-10-04 | 94% |
| NEW PLYMOUTH AWS, NZ  GHCND:NZ000933090 | 1944-01-01 | 2023-10-04 | 98% |
| PARAPARAUMU AWS, NZ  GHCND:NZ000093417 | 1972-01-01 | 2023-10-04 | 97% |
| RAOUL ISL KERMADEC, NZ  GHCND:NZ000093994 | 1940-03-08 | 2023-10-04 | 89% |
| TARA HILLS, NZ  GHCND:NZ000937470 | 1949-11-01 | 2005-12-31 | 96% |

To handle missing value of temperature series data using **Technical note on the treatment of missing data in the 11-station series** from Niwa from historical data around 1930 to 2009 that provide an idea how to fill the missing data. This dataset only provides guidance as annual mean value.

<https://niwa.co.nz/our-science/climate/information-and-resources/nz-temp-record/temperature-trends-from-raw-data/technical-note-on-the-treatment-of-missing-data>

To replace the missing value in more details is necessary for example temperature change in different season. A combined monthly data for mean temperature base on seven station available on **NIWA 'seven-station' temperature series: monthly data for mean temperature**, <https://niwa.co.nz/seven-stations>, this data contains the monthly data for mean temperature from 1909-2018 locations in the "seven station" series, which including:

* Auckland
* Masterton
* Wellington
* Hokitika
* Nelson
* Lincoln
* Dunedin

StatsNZ

A dataset are available from Ministry for the Environment and Statistics New Zealand, which were use to develop ‘Our Atmosphere and Climate’ report which contains: rain, drought, temperature, atmospheric data from 30site on StatsNZ.

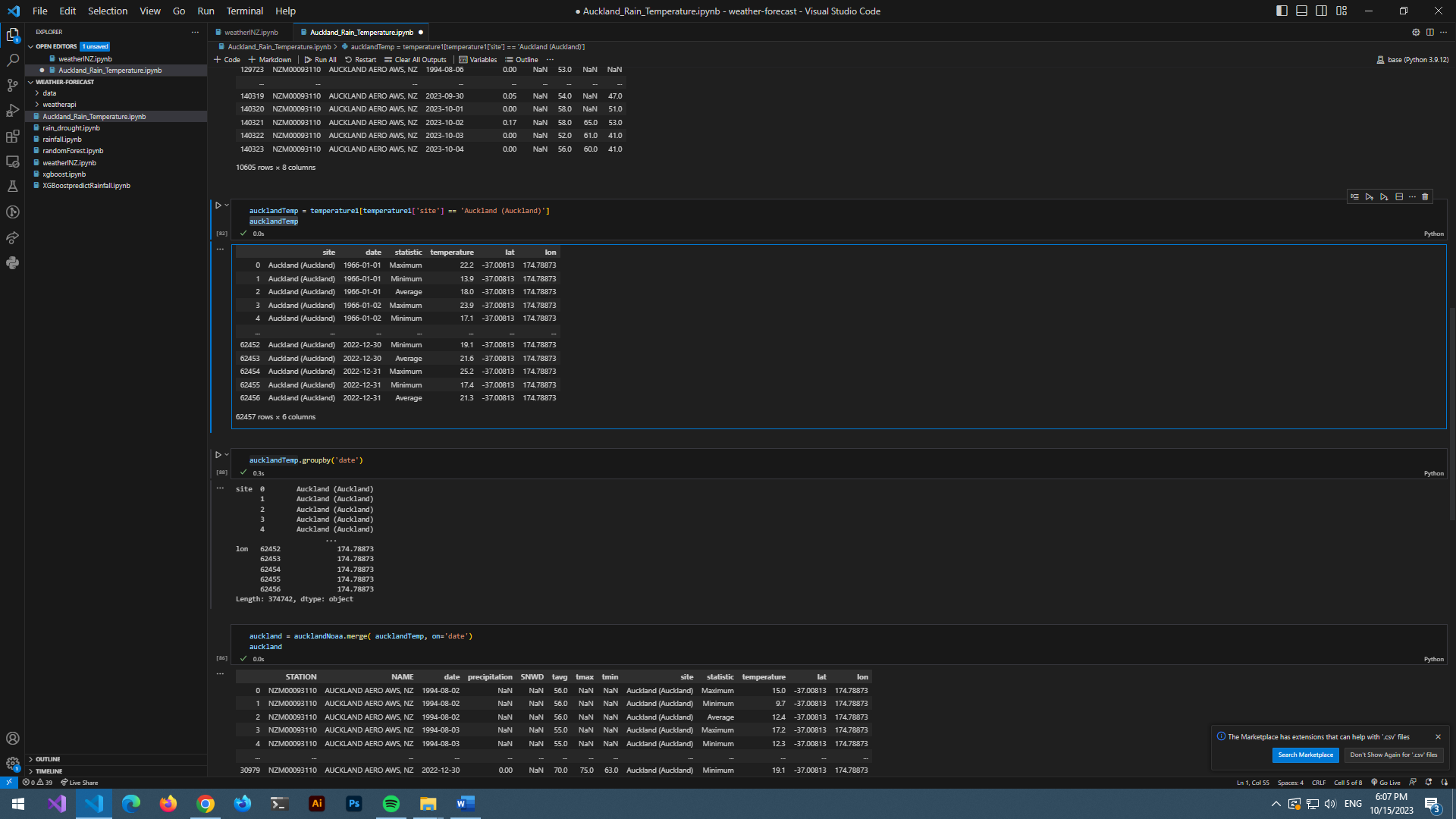
<https://data.mfe.govt.nz/table/105056-daily-temperature-1909-2019/>

these site including:

1. Gisborne (Gisborne)
2. Dunedin (Otago)
3. Invercargill (Southland)
4. Dannevirke (Manawatū-Whanganui)
5. Christchurch (Canterbury)
6. Hokitika (West Coast)
7. Auckland (Auckland)
8. Blenheim (Marlborough)
9. Gore (Southland)
10. Hamilton (Waikato)
11. Lake Tekapo (Canterbury)
12. Milford Sound (Southland)
13. Napier (Hawke's Bay)
14. Nelson (Nelson)
15. New Plymouth (Taranaki)
16. Kerikeri (Northland)
17. Masterton (Wellington)
18. Queenstown (Otago)
19. Reefton (West Coast)
20. Rotorua (Bay of Plenty)
21. Wellington (Wellington)
22. Tauranga (Bay of Plenty)
23. Taumarunui (Manawatū-Whanganui)
24. Tara Hills (Canterbury)
25. Whangārei (Northland)
26. Taupō (Waikato)
27. Timaru (Canterbury)
28. Waiouru (Manawatū-Whanganui)
29. Whanganui (Manawatū-Whanganui)
30. Whangaparāoa (Auckland)

To combine, fill data from both set of data from StatsNZ and NOAA database. The site are different, which in this model we narrow down to just base on Auckland. In StatsNZ, all the value are stacked.:

For example :



To create extra column:

* First grouped by the ‘date’, and apply aggregation function(max, min, mean) and reset as extra column
* Make sure both dataset have lowercase for merging base on ‘date’