Nama : Irja’ Multazamy

NIM : 200411100155

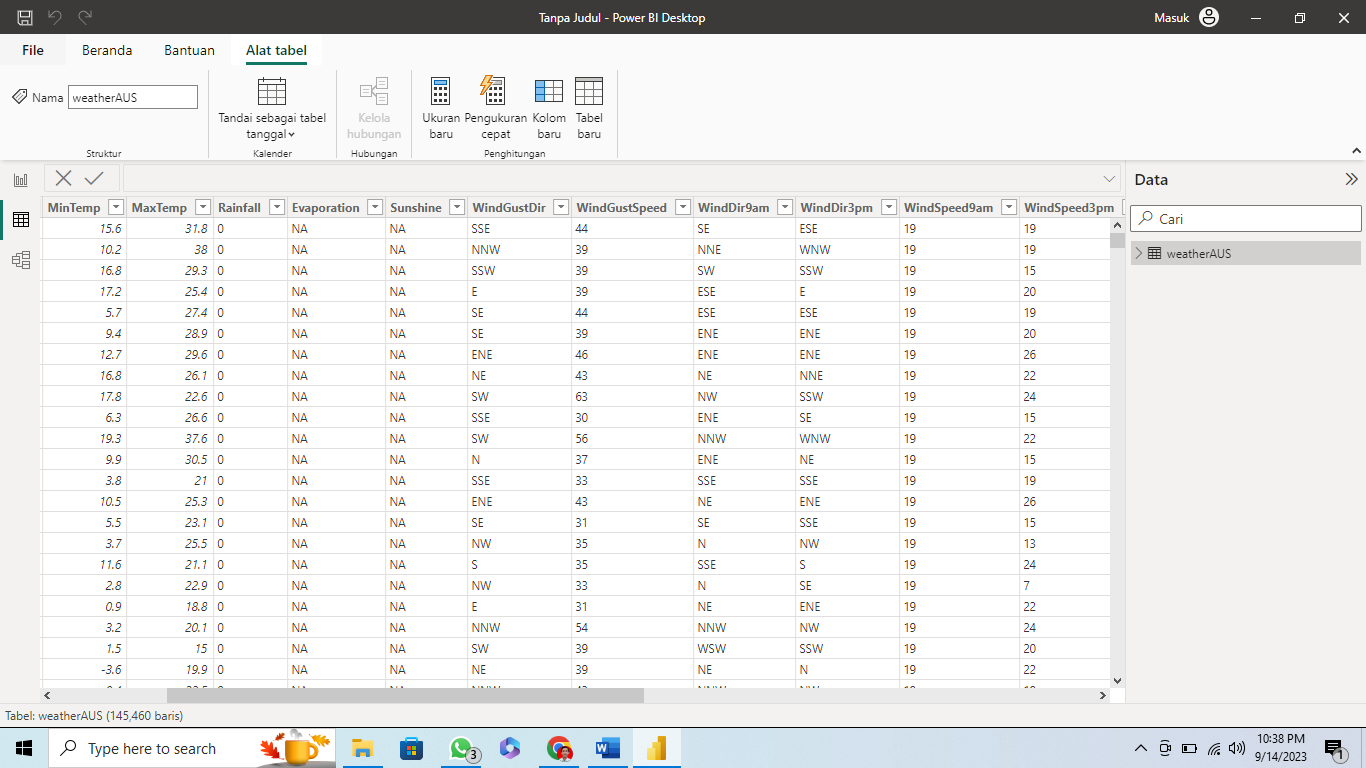
Mata Kuliah : Proyek Sains Data

**Tugas 1 (Analisis Data)**

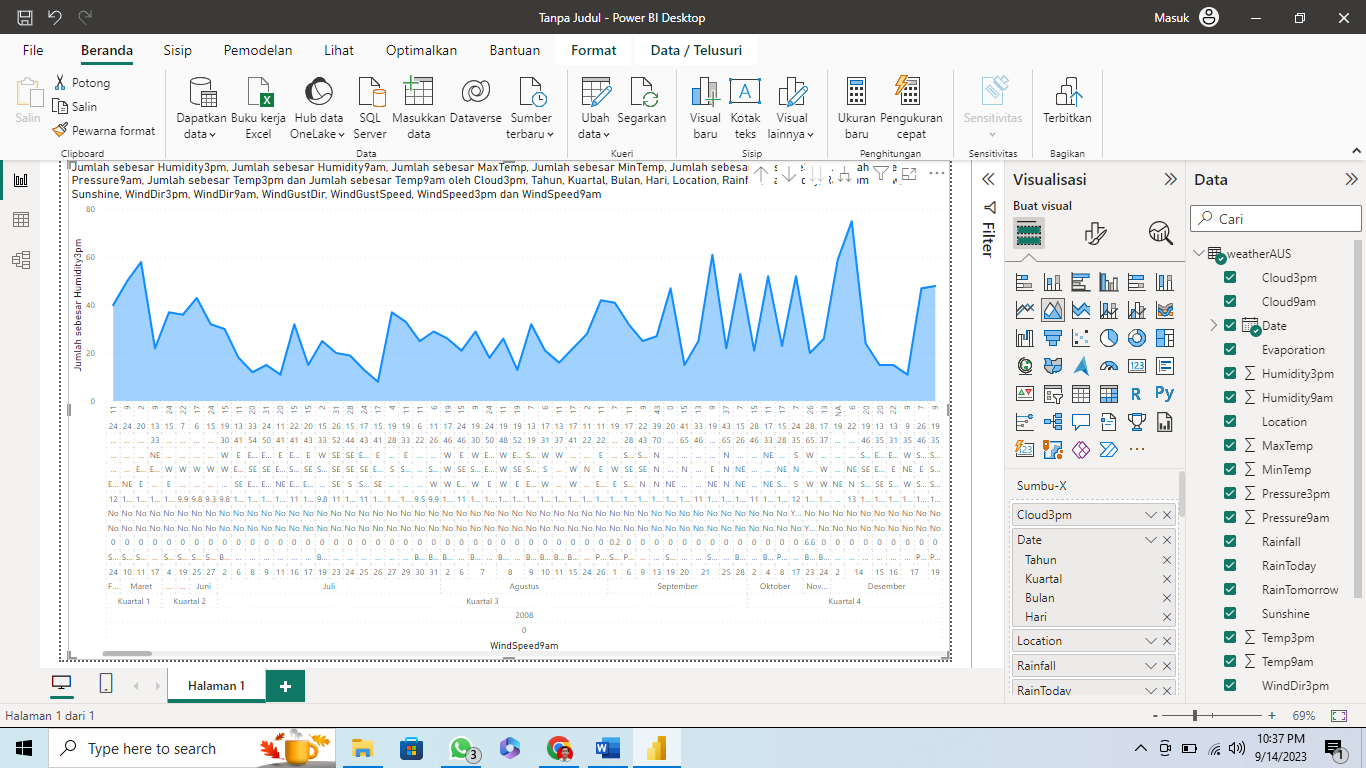
1. Cari data di kaggle dan lakukan analisis data menggunakan Power BI

Rain in Australia, link: <https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package> , Data ini adalah kumpulan data hasil pengamatan cuaca harian dari berbagai stasiun cuaca di Australia selama 10 tahun. Data ini memiliki 23 fitur dan 145460 baris.

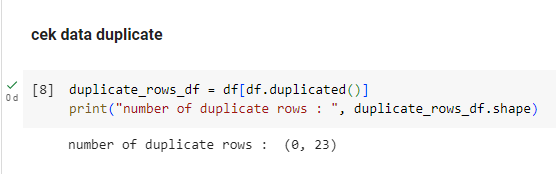
* Import data ke Power BI



* Setelah data dimasukkan, analisis data dapat dilakukan mulai dari transform data, menampilkan diagram garis, visualisasi data dan banyak operasi lainnya.



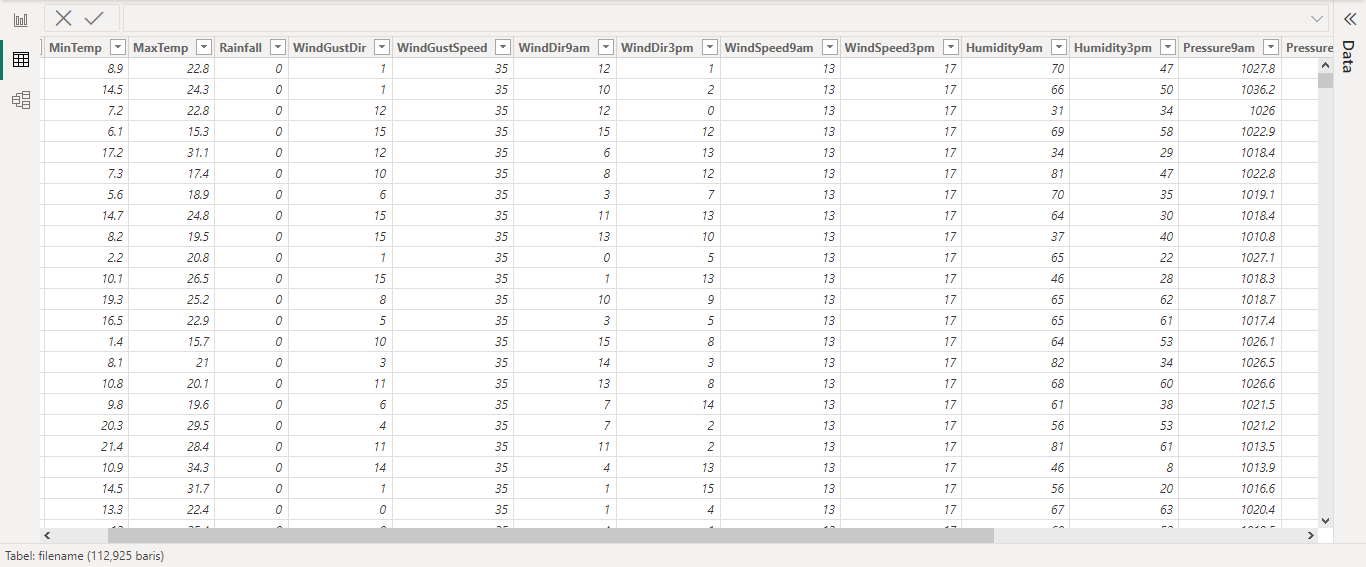
1. Cek data yang duplikat



1. Menentukan fitur yang akan digunakan

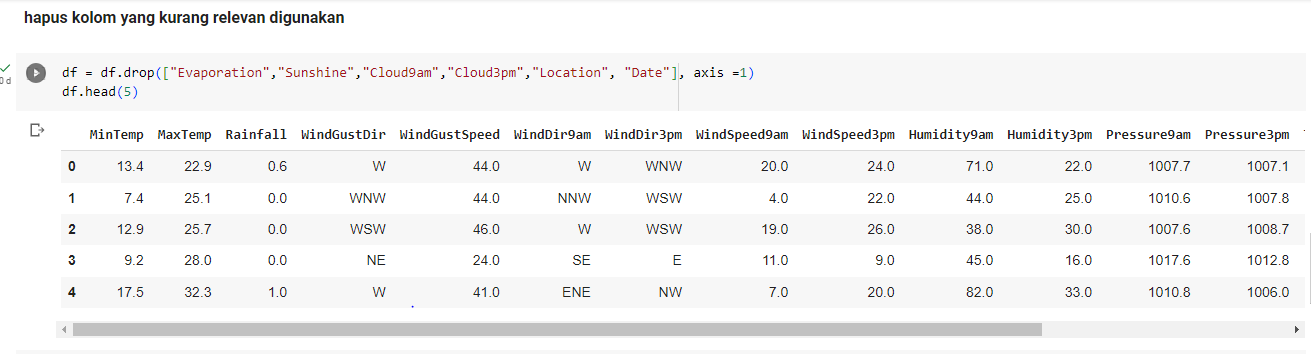
Ada 17 dari 23 fitur yang digunakan yaitu:

* MinTemp : suhu minimum dalam derajat celsius
* MaxTemp : suhu maximal dalam derajat celsius
* Rainfall : Jumlah curah hujan yang tercatat pada hari itu dalam mm
* WindGustDir : Arah hembusan angin terkuat dalam 24 jam
* WindGustSpeed : Kecepatan (km/jam) hembusan angin terkuat dalam 24 jam
* WindDir9am : Arah angin pada jam 9 pagi
* WindDir3pm : Arah angin pada jam 3 sore
* WindSpeed9am : Kecepatan angin pada jam 9 pagi
* WindSpeed3pm : Kecepatan angin pada jam 3 sore
* Humidity9am : kelembapan pada jam 9 pagi
* Humidity3pm : kelembapan pada jam 3 sore
* Pressure9am : Tekanan udara
* Pressure3pm : Tekanan udara
* Temp9am : suhu pada jam 9 pagi
* Temp3pm : suhu pada jam 3 sore
* RainToday : hari ini hujan? Yes atau No
* RainTomorrow : besok hujan? Yes atau No



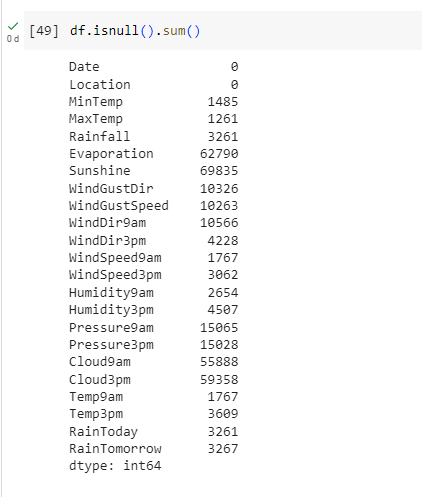
1. Fitur yang dihapus

Ada 6 fitur yang dihapus yaitu: Evaporation (penguapan), Sunshine (berapa jam matahari menyinari dalam sehari), Cloud9am, Cloud3pm, Location, Date.

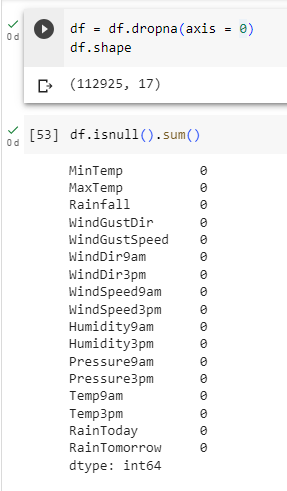


1. Melakukan proses clean data dengan cara mengatasi missing data

Setelah melakukan pengecekan, missing data dapat ditemukan

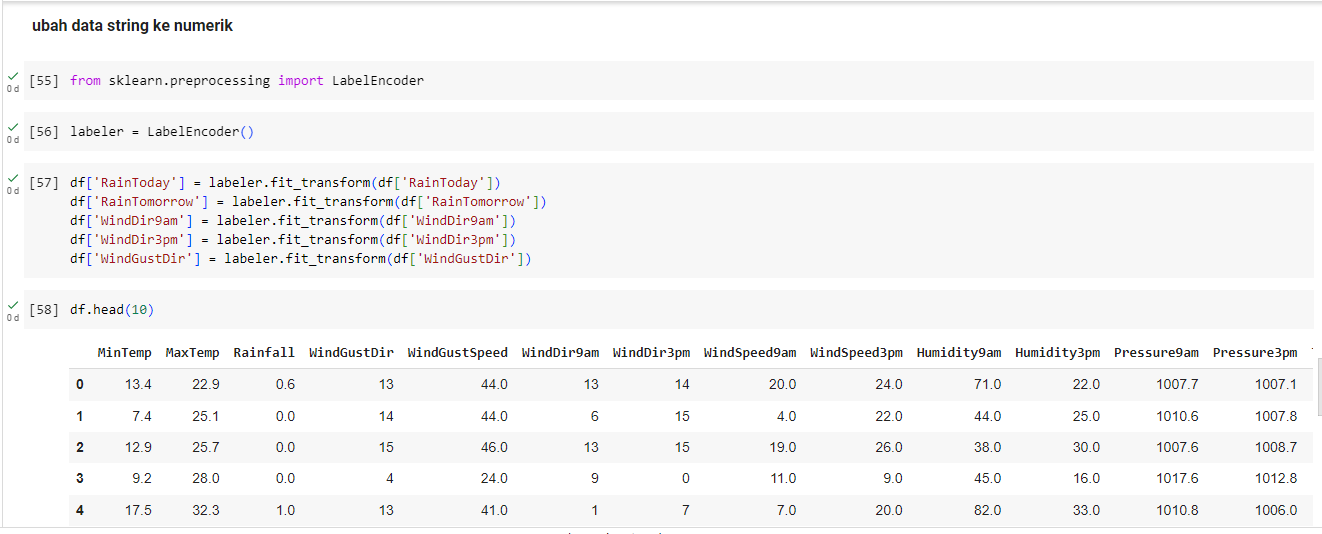


Menggunakan method dropna() untuk menghapus baris atau kolom yang mengandung missing values, seperti pada contoh



1. Merubah data string menjadi numerik

Merubah nilai string menjadi numerik dapat dilakukan dengan menerapkan Label encoder merupakan sebuah library python dari 'SciKit Learn' yang berfungsi mengkonfersi data-data kategorikal dan string yang bersifat kategorikal menjadi numeric yang dapat dengan mudah dipahami model.



1. Splitting dan Balancing data

Melakukan splitting data dan balancing data. Splitting data merupakan metode yang digunakan untuk membagi dataset menjadi data train dan data test.

