Nama : Yoga Tirta Permana

NIM : 200411100142

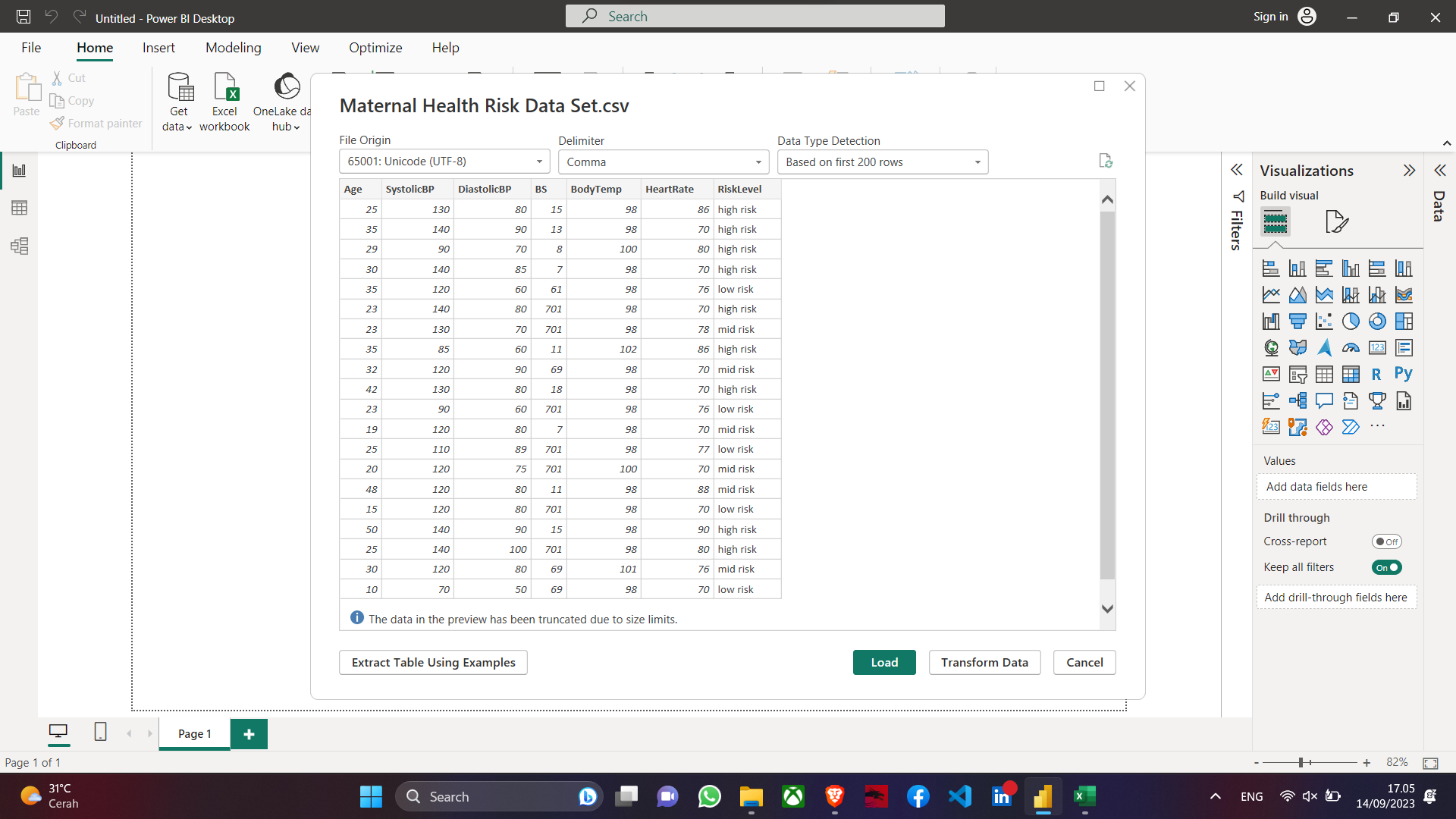
Kelas : Proyek Sains Data E

**Tugas 1 Analisis Data**

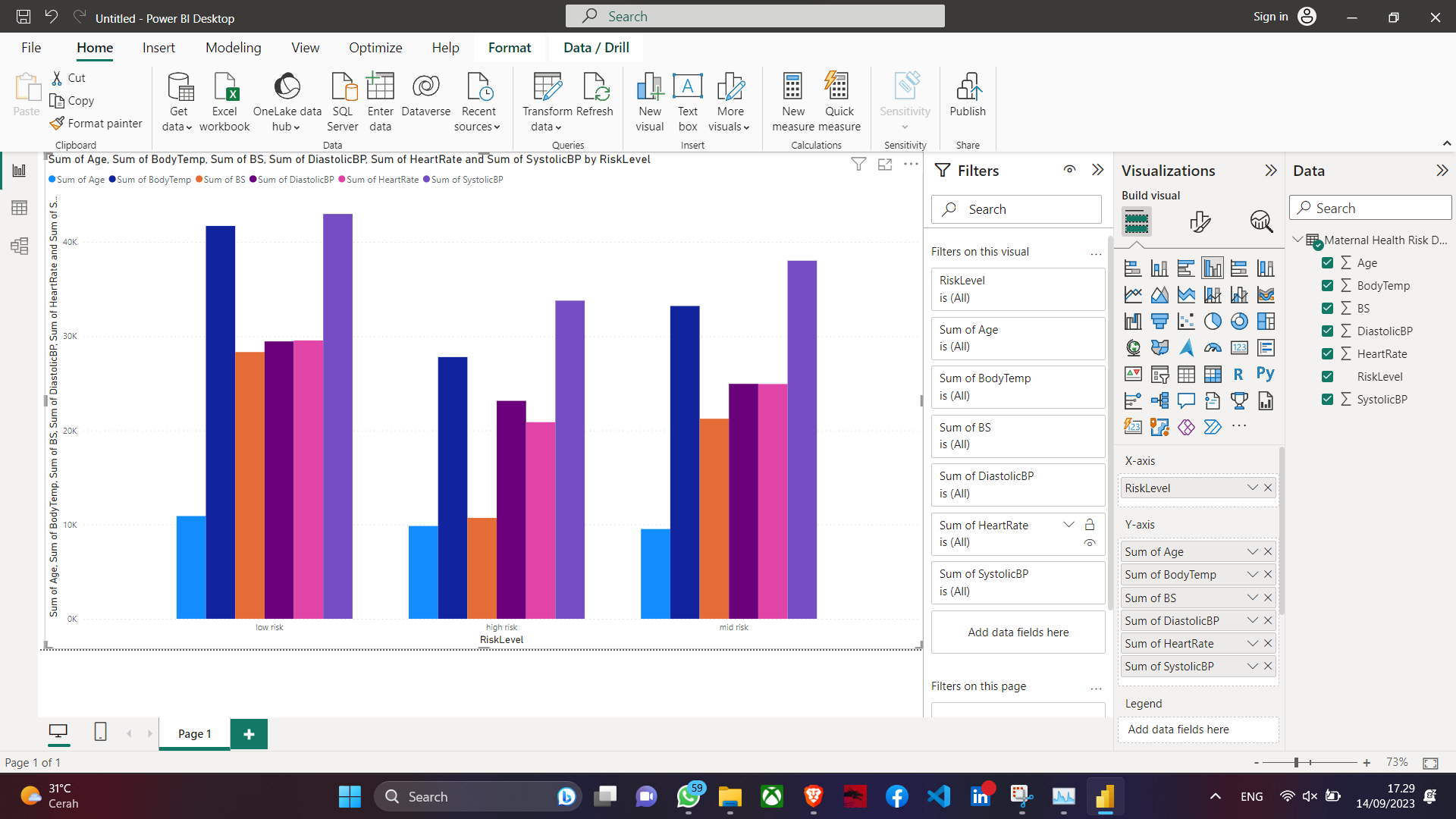
1. Cari data di Kaggle dan lakukan analisis data menggunakan Power BI

Data yang penulis gunakan adalah [Maternal Health Risk Data](https://www.kaggle.com/datasets/csafrit2/maternal-health-risk-data). Dataset tersebut berisi prediksi resiko kesehatan bagi pasien hamil yang telah dikumpulkan dari berbagai rumah sakit, klinik komunitas, dan layanan kesehatan ibu melalui sistem pemantauan risiko berbasis IoT. Dataset tersebut terdiri dari 7 fitur dan 1014 kolom.

* Untuk mulai menganalisis data, kita import data kita kedalam Power BI



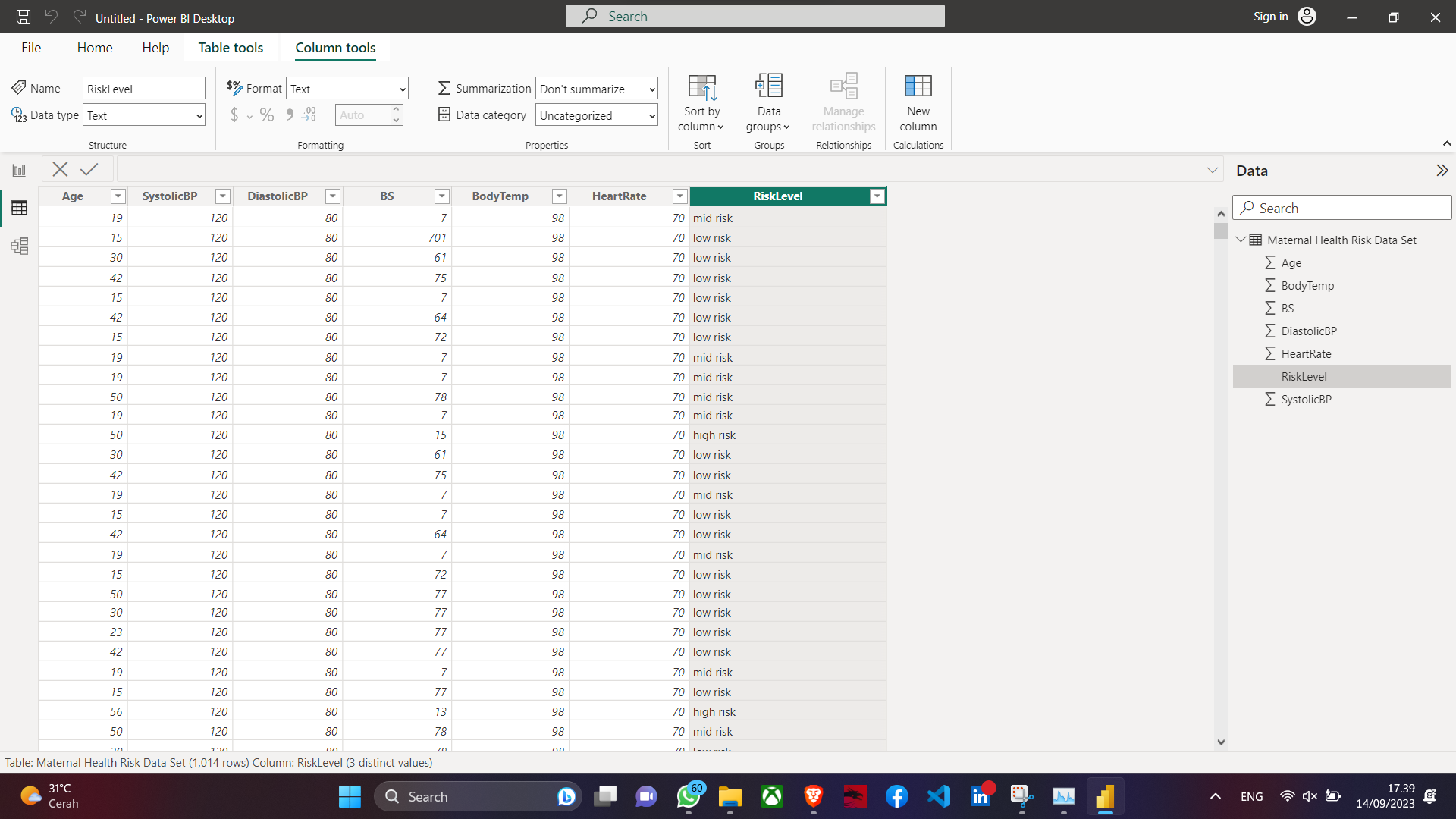
* Setelah data dimasukkan, kita bisa menganalisis data mulai dari transform data, menampilkan diagram, dan visualisasi data.



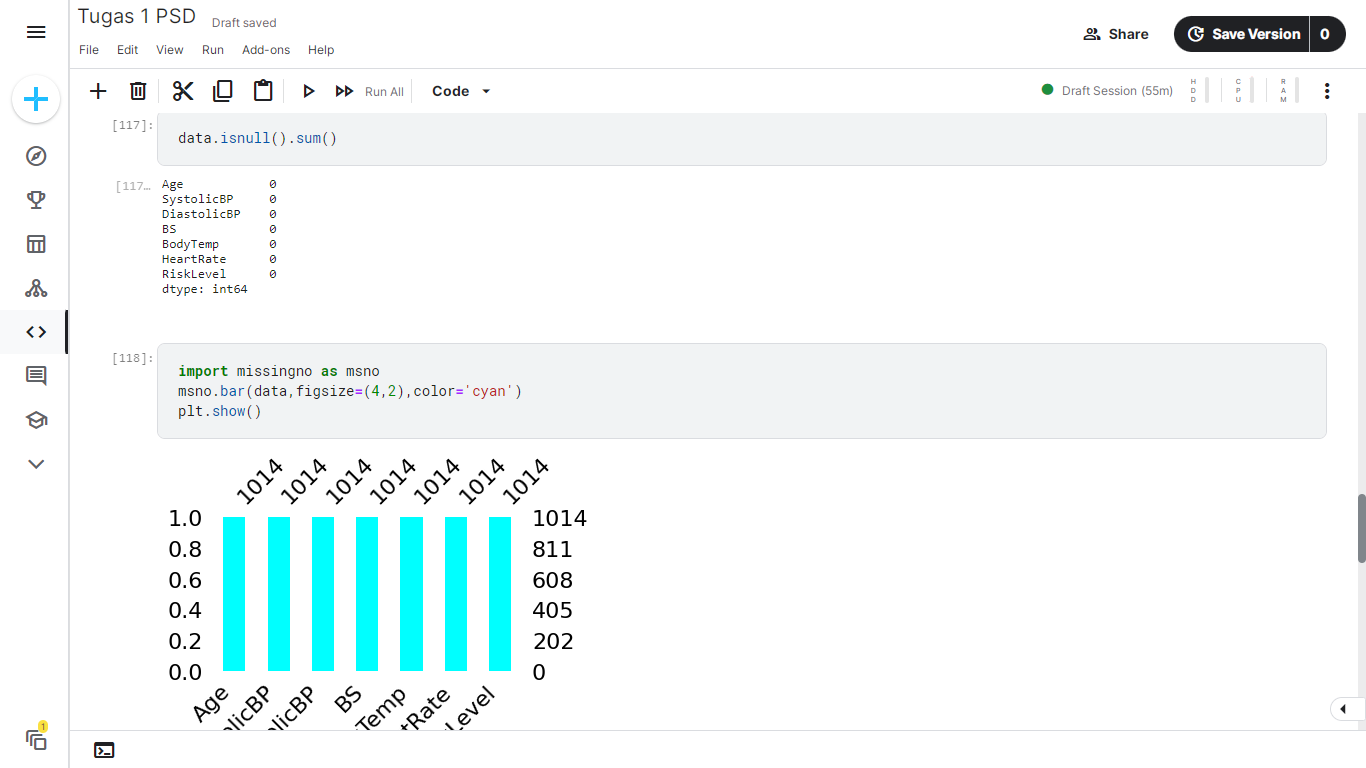
1. Analisis hubungan antar parameter yang ada

Pada dataset [Maternal Health Risk Data](https://www.kaggle.com/datasets/csafrit2/maternal-health-risk-data) terdapat 7 parameter atau fitur yang terdiri dari:

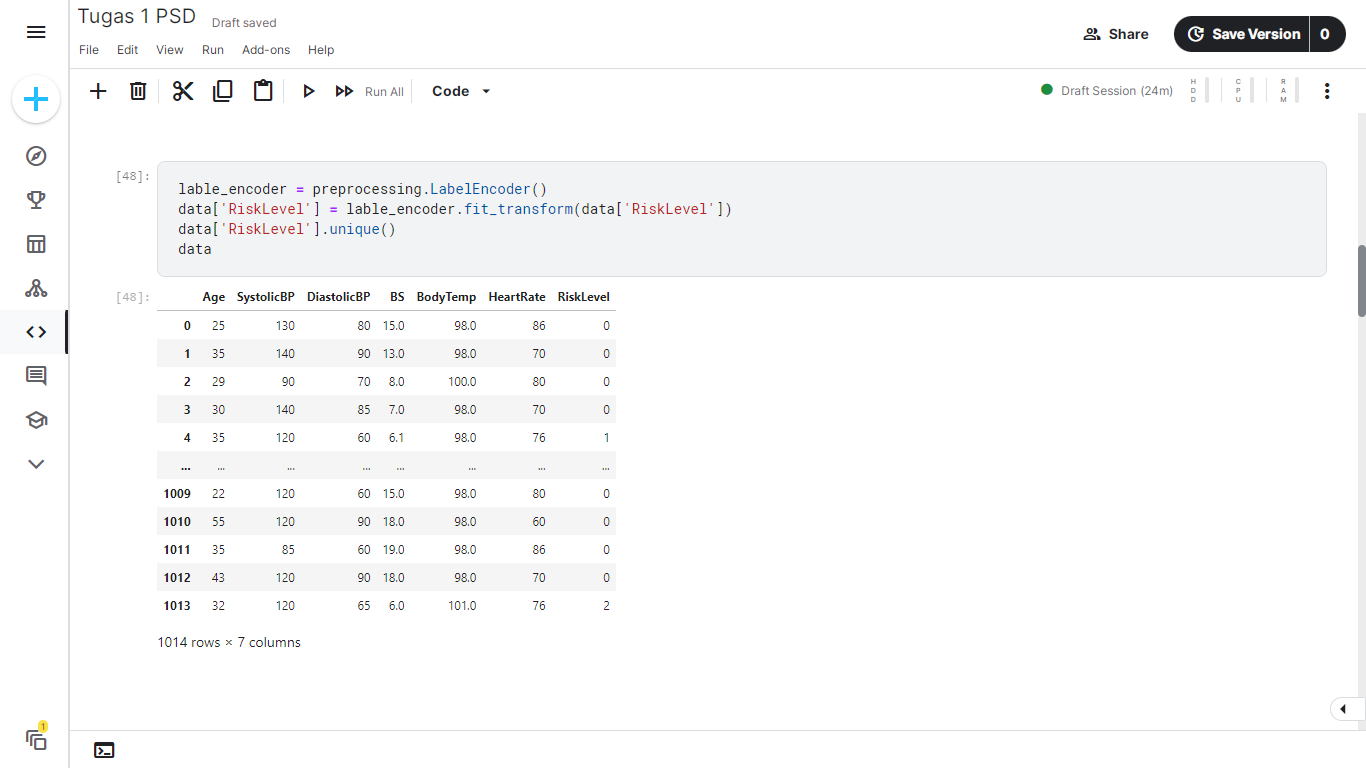
* **Age**: Usia dalam tahun saat seorang wanita hamil.
* **SystolicBP**: Nilai Tekanan Darah tertinggi dalam mmHg, atribut penting lainnya selama kehamilan.
* **DiastolicBP**: Nilai Tekanan Darah yang lebih rendah dalam mmHg, atribut penting lainnya selama kehamilan.
* **BS**: Kadar glukosa darah dinyatakan dalam konsentrasi molar, mmol/L.
* **BodyTemp**: Suhu tubuh dalam farenheit
* **HeartRate**: Denyut jantung istirahat normal dalam detak per menit.
* **RiskLevel**: Prediksi Tingkat Intensitas Risiko selama kehamilan berdasarkan atribut sebelumnya. Semua parameter saling terhubung dengan target untuk mendapatkan hasil akurasi yang maksimal



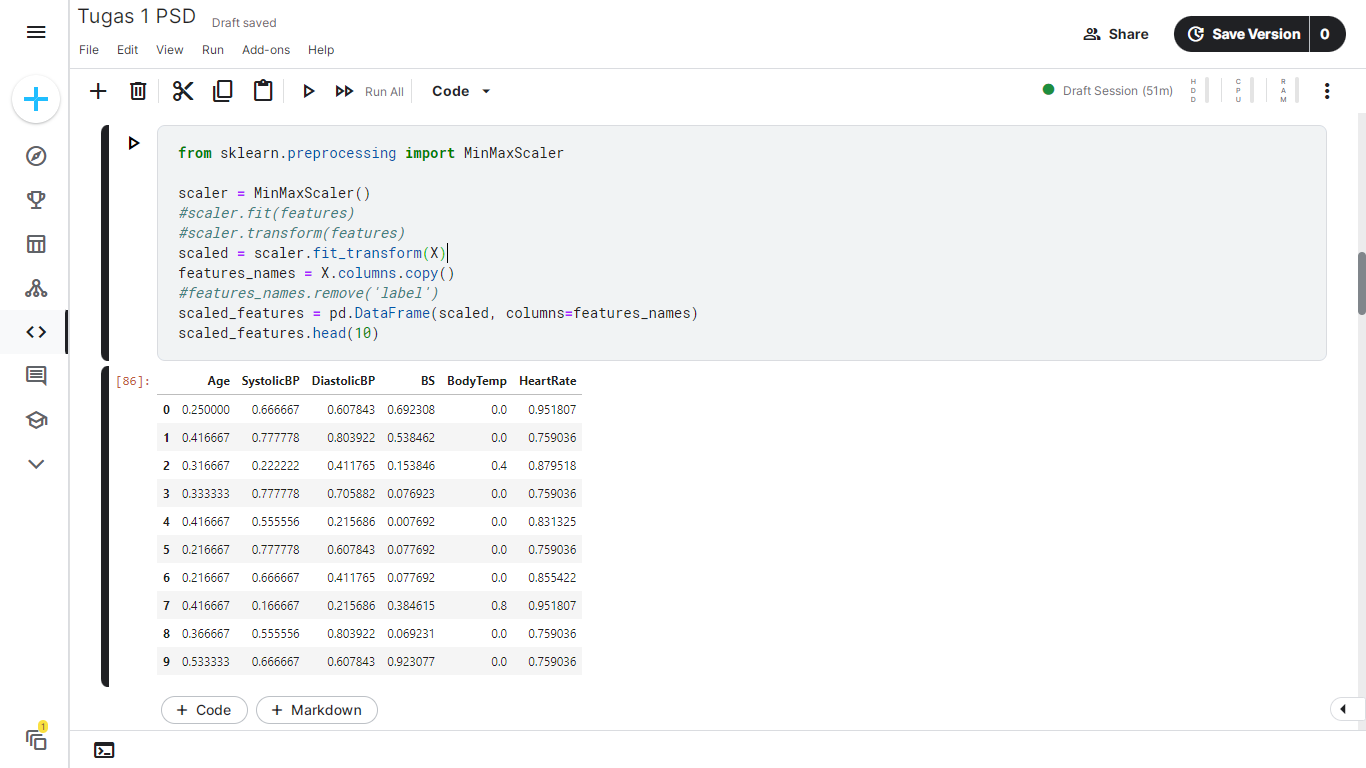
1. Melakukan proses Cleaning Data dengan cara :
2. Mengatasi missing data
3. Transformasi Data
4. Normalisasi Data
5. Dataset [Maternal Health Risk Data](https://www.kaggle.com/datasets/csafrit2/maternal-health-risk-data) sudah berbentuk data clean, karena tidak ada missing value yang terdeteksi, hal tersebut bisa dibuktikan ketika penulis mencoba untuk melakukan operasi mencari missing value di kaggle langsung menggunakan python seperti contoh dibawah ini:



1. Proses transformasi data adalah ketika kita ingin merubah data kedalam bentuk numerik agar bisa diolah lebih lanjut



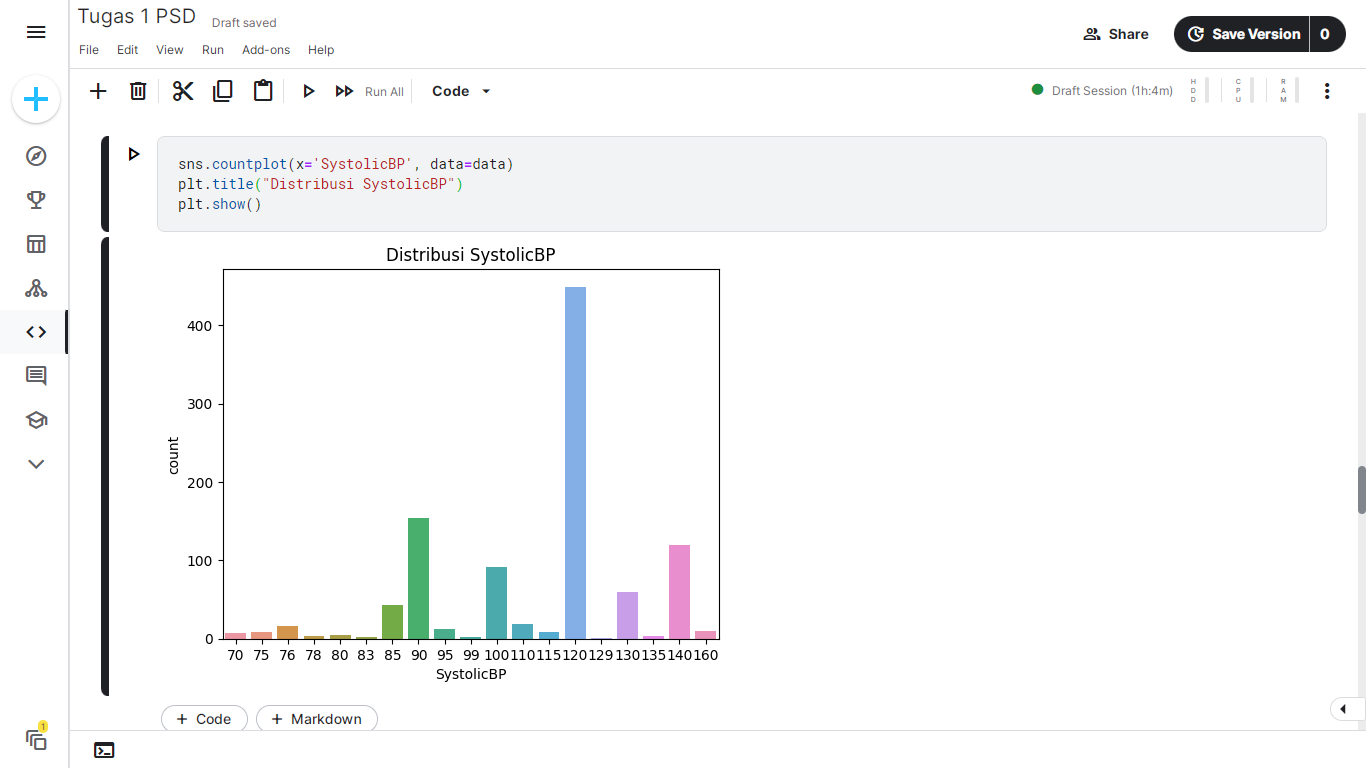
1. Proses normalisasi adalah ketika kita merubah range data numerik yang sudah ditransformasi, menjadi range antara 0 sampai 1 biasanya menggunakan min-max scaler.

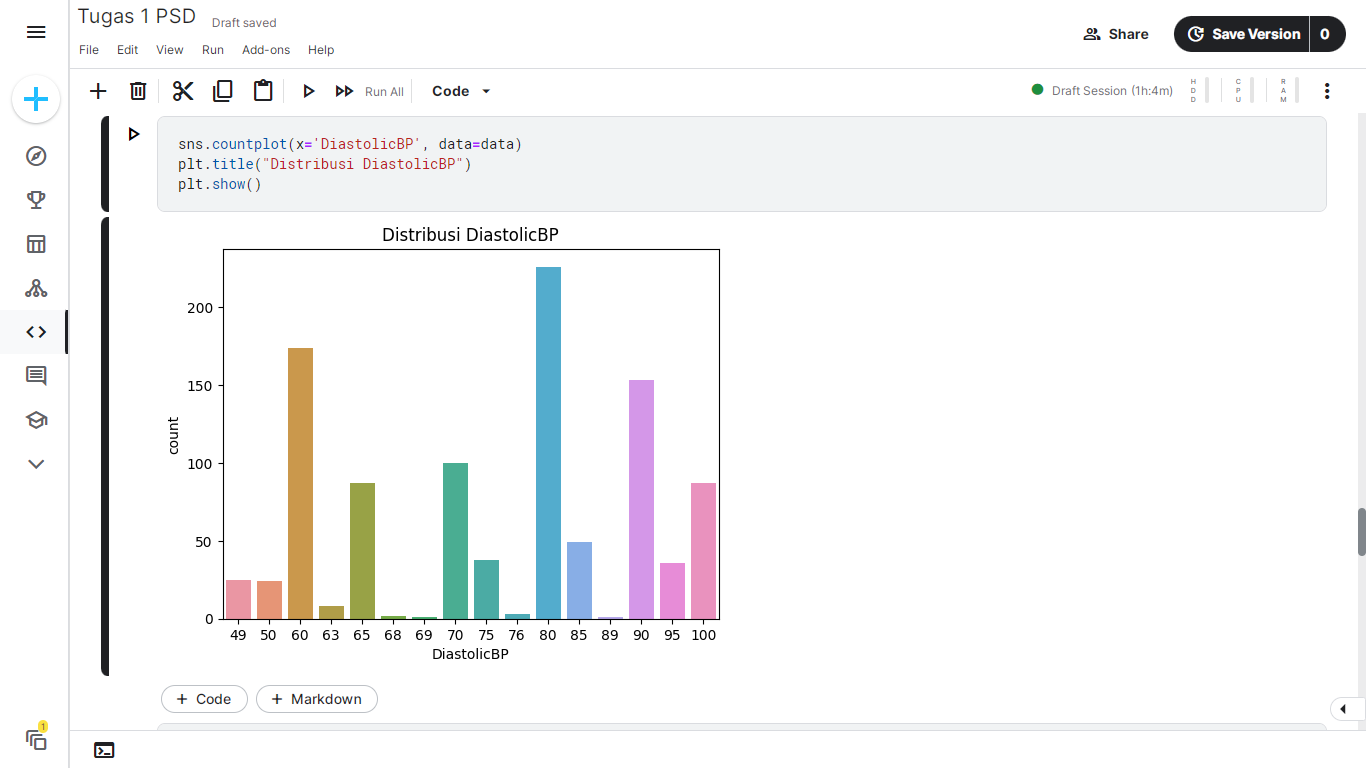


1. Balancing data

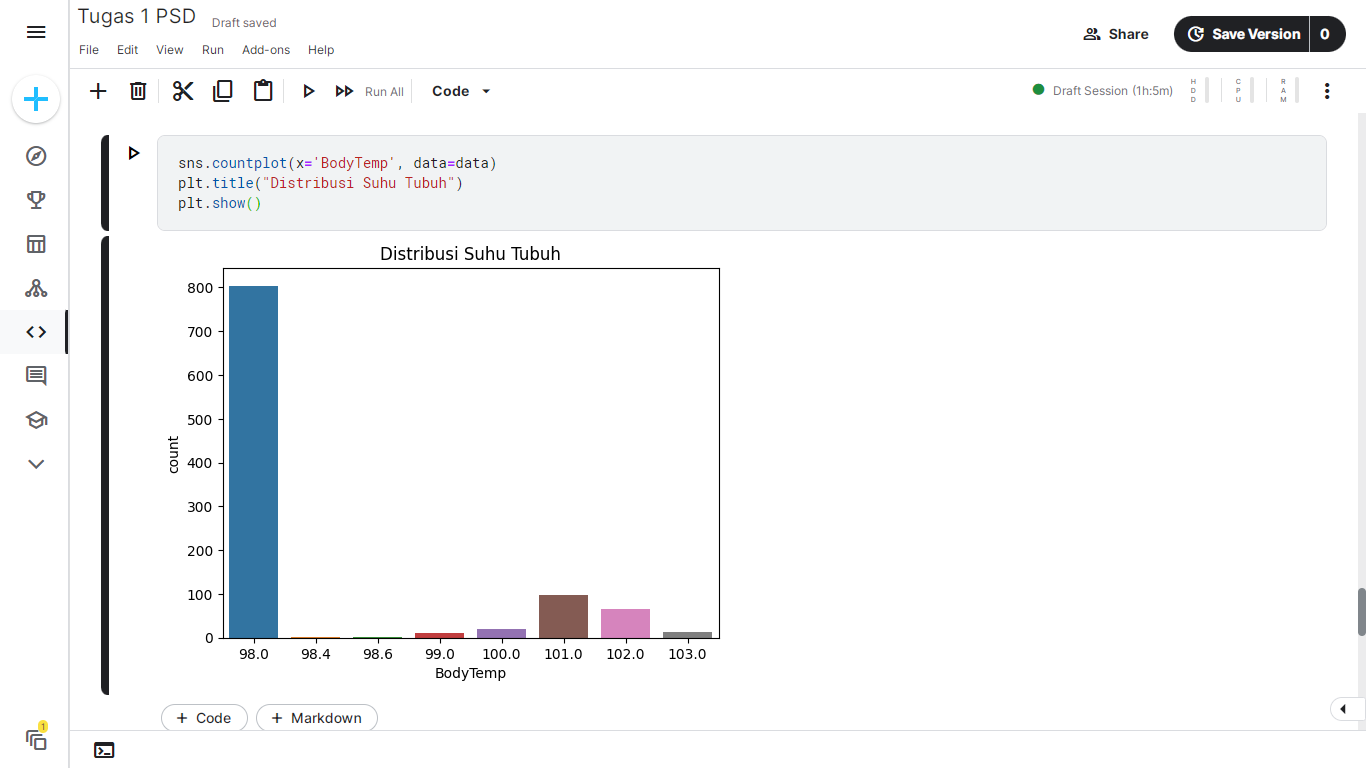
Proses Balancing data dilakukan dalam kondisi ketika dimana salah satu atau lebih dari class yang ada dalam suatu himpunan memiliki jumlah yang cukup timpang atau memliki perbedaan yang cukup jauh diantara class yang ada. Dalam kasus ini kita harus mengetahui sebaran data dari masing masing parameter seperti berikut:

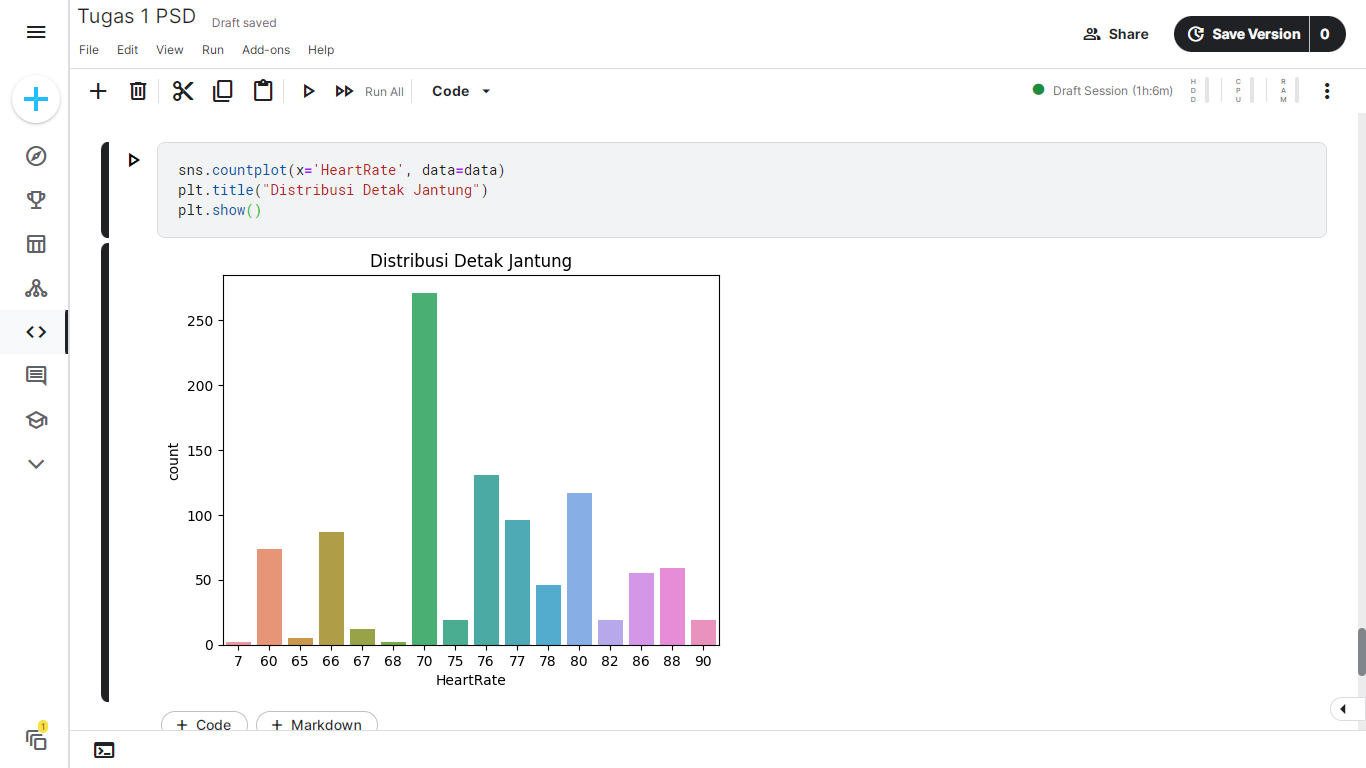


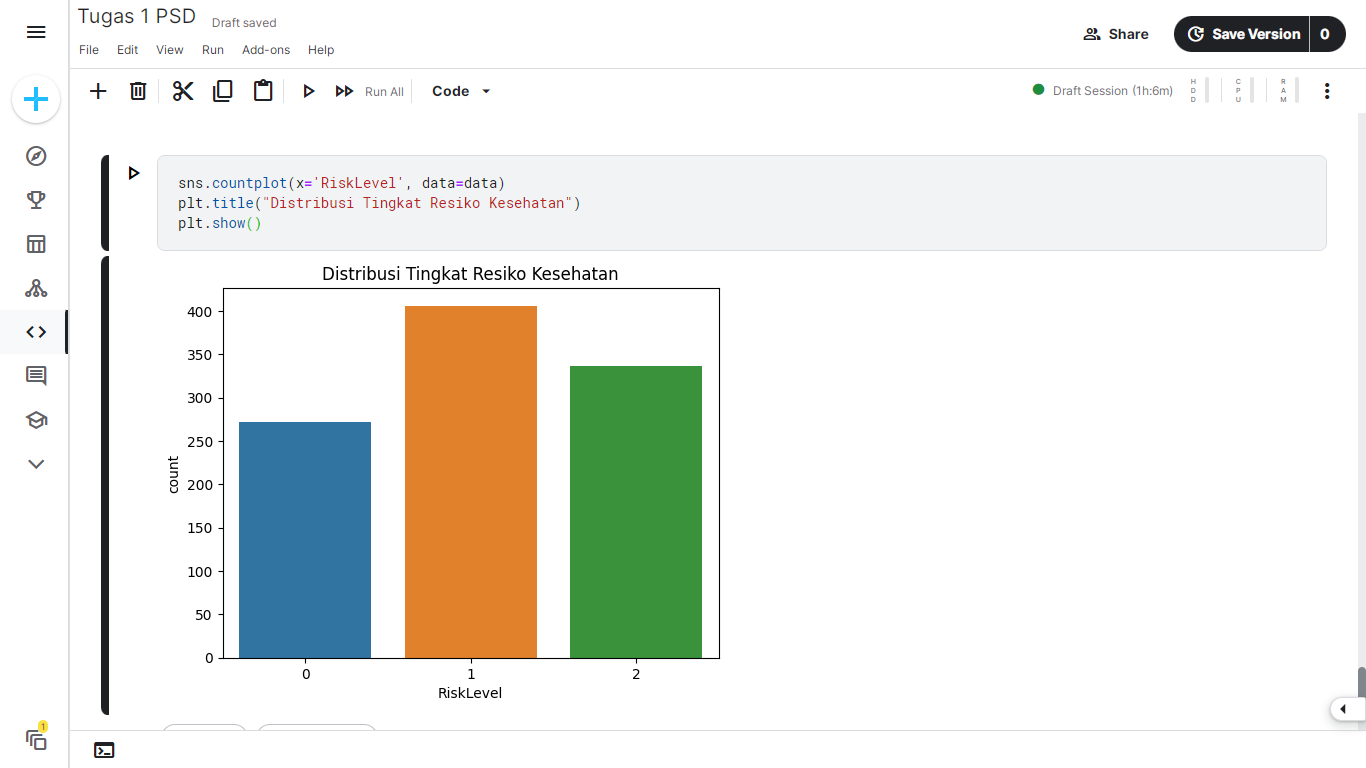












Dari sebaran data diatas, tidak terjadi Imbalance, karena presentase selisih dari label RiskLevel tidak lebih dari 10%.

Kaggle : https://www.kaggle.com/code/yogatirtapermana/tugas-1-psd