**LAPORAN TUGAS**

**ANALISIS DATA WATER POTABILITY DATASET**



**DISUSUN OLEH :**

Oki Bagus Rahmat Prakoso (210411100183)

Dosen Pengampu : EKA MALA SARI ROCHMAN, S.Kom., M.Kom.

**PRODI TEKNIK INFORMATIKA**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS TRUNOJOYO MADURA**

**2023**

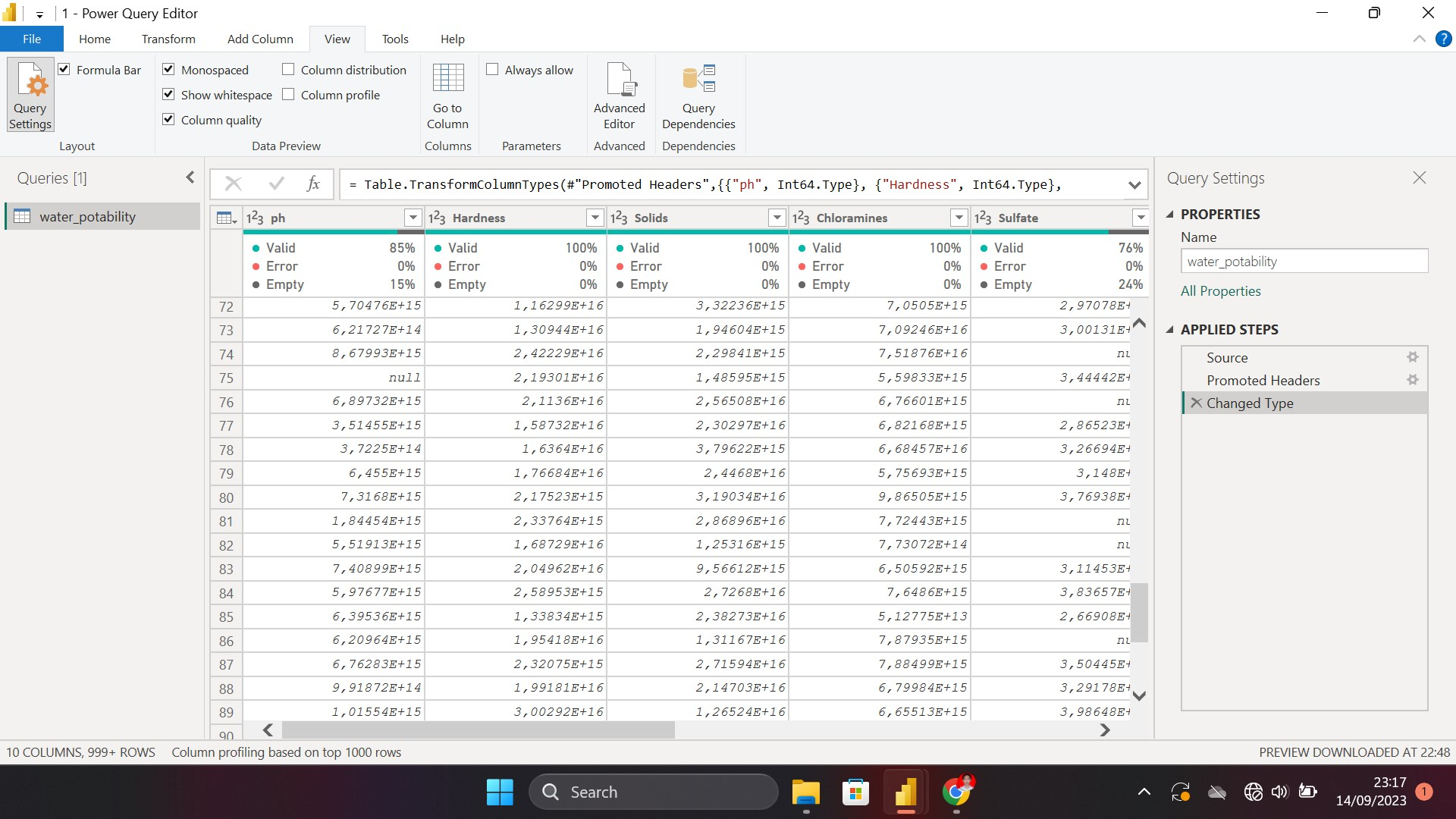
Analisis dataset dari repositori Kaggle yaitu **water\_potability**. Terdapat 3276 baris data dan 9 fitur atau parameter dan 1 atribut labelling bernama Potability yang tertampil pada Tabel 1

Parameter yang berhubungan dengan target apakah air dapat diminum atau tidak.

**Tabel 1**. Parameter Dataset Kualitas Air

|  |  |
| --- | --- |
| Parameter | Deskripsi |
| PH | PH merupakan indikator kondisi asam atau basa status air. PH optimum yang dibutuhkan berkisar 6,5-8 |
| Hardness | Hardness atau kesadahan air merupakan tingkat kemampuan air untuk mengendapkan sabun disebabkan karena adanya ion-ion logam bervalensi seperti kalsium dan magnesium |
| Solids | Solid menggambarkan garam anorganik dan jumlah bahan organik yang berada di larutan dalam air |
| Chloramines | Klorin dan kloramin adalah disinfektan utama yang digunakan dalam sistem air umum. Kloramin terbentuk ketika amonia ditambahkan ke klorin untuk mengolah air minum |
| Sulfate | Sulfate ditemukan di mineral, tanah, batuan. Sulfate dapat menyebabkan rasa yang pekat dengan kadar yang tinggi serta dapat menyebabkan efek pencahar pada konsumen yang tidak terbiasa |
| Conductivity | Konsentrasi jumlah kation atau ion bermuatan positif dan anion atau ion bermuatan negatif di dalam air. Peningkatan konsentrasi ion meningkatkan konduktivitas listrik di air. |
| Organic Carbon | Organic carbon berasal dari bahan atau senyawa organik yang terurai di air |
| Trihalomethanes | Trihalomethanes (THMs) terbentuk dalam air minum terutama sebagai akibat dari klorinasi bahan organik |
| Turbidity | Tingkat kekeruhan yang ada di air |
| Potability | Indikator air layak diminum dan tidak layak diminum |

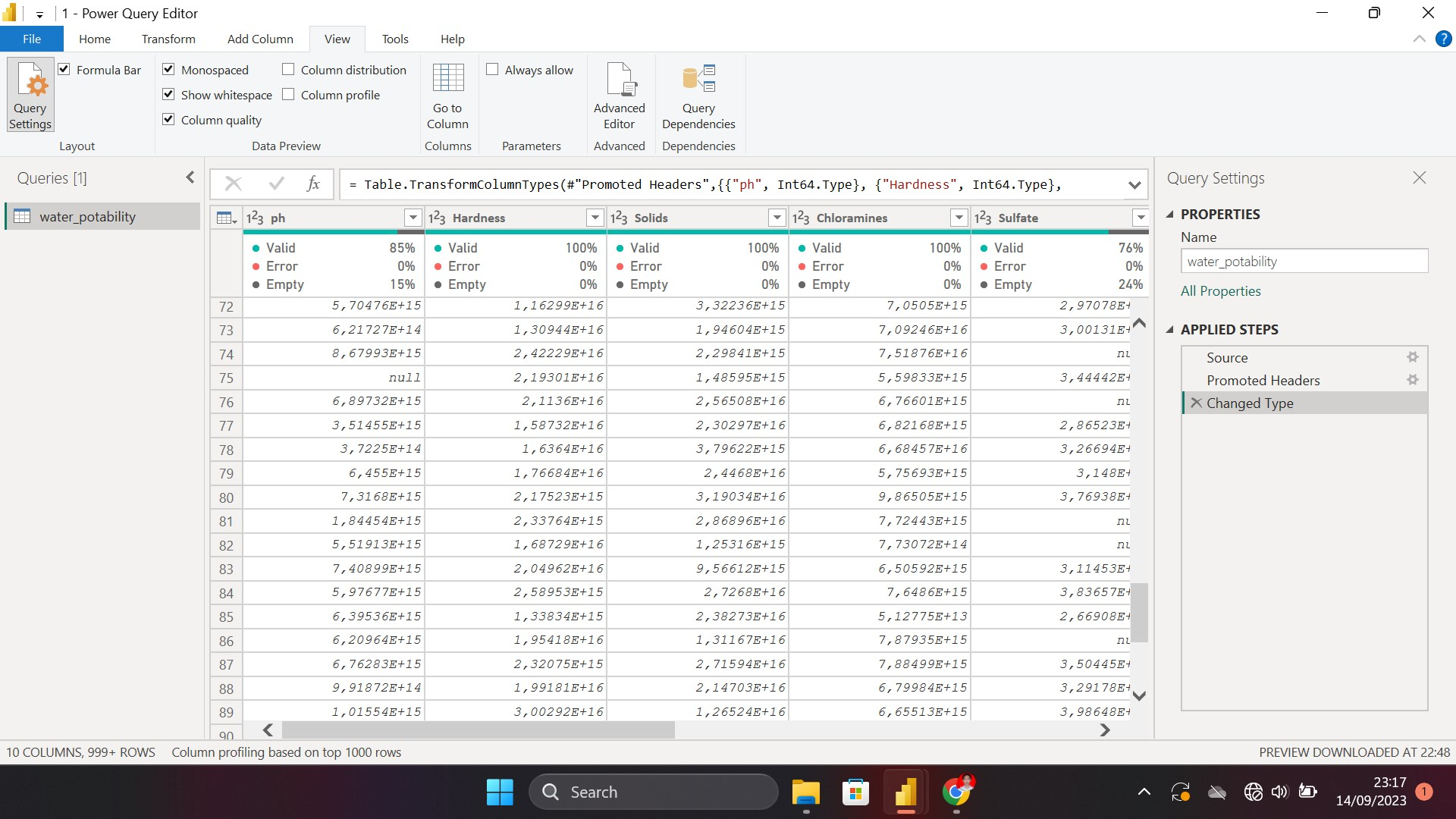
Parameter ini saya gunakan berdasarkan : **Pritalia, G. L. (2022). Analisis Komparatif Algoritme Machine Learning dan Penanganan Imbalanced Data pada Klasifikasi Kualitas Air Layak Minum. *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, *2*(1).**

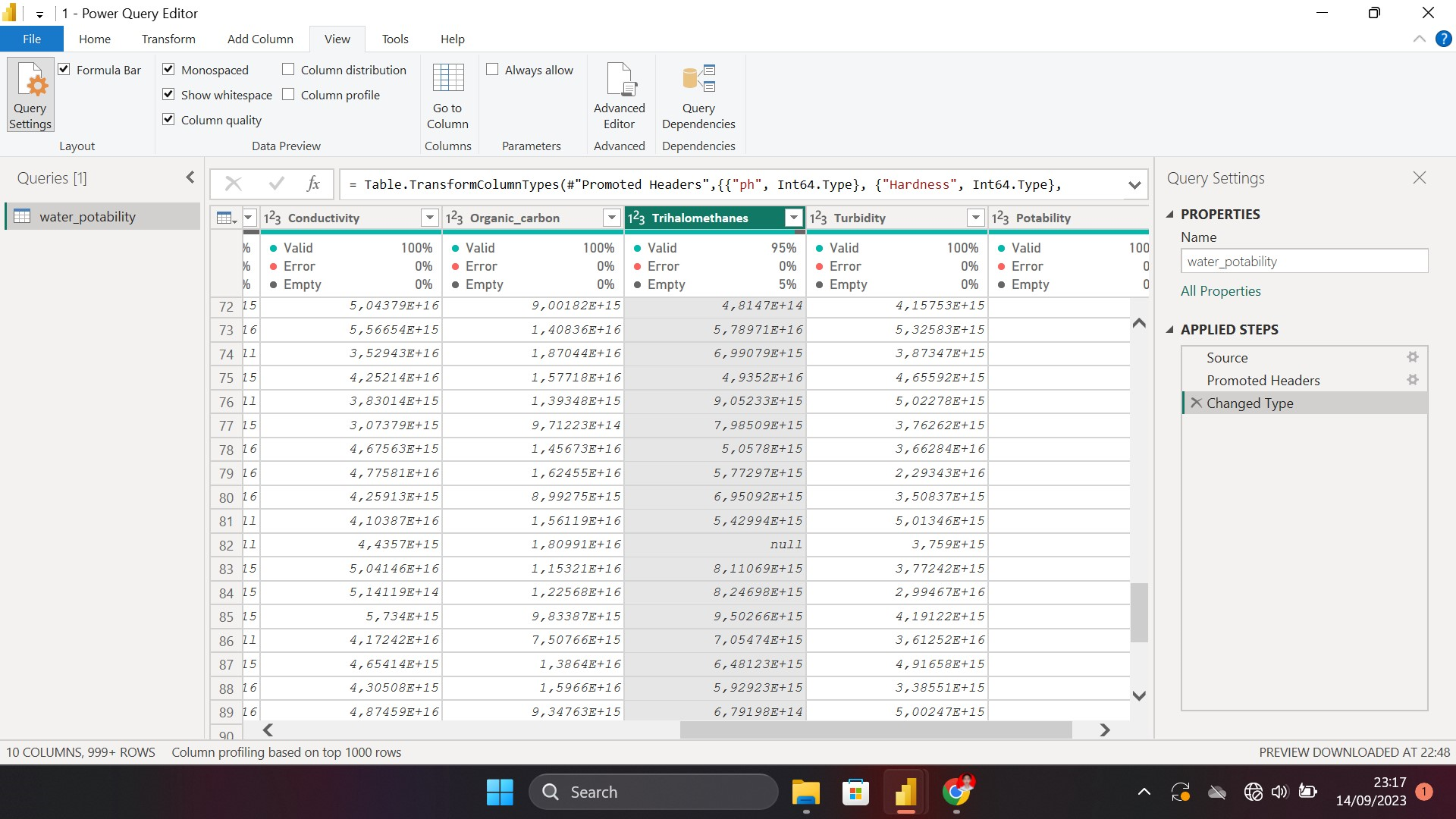


Tahap preprosesing setelah menentukan parameter

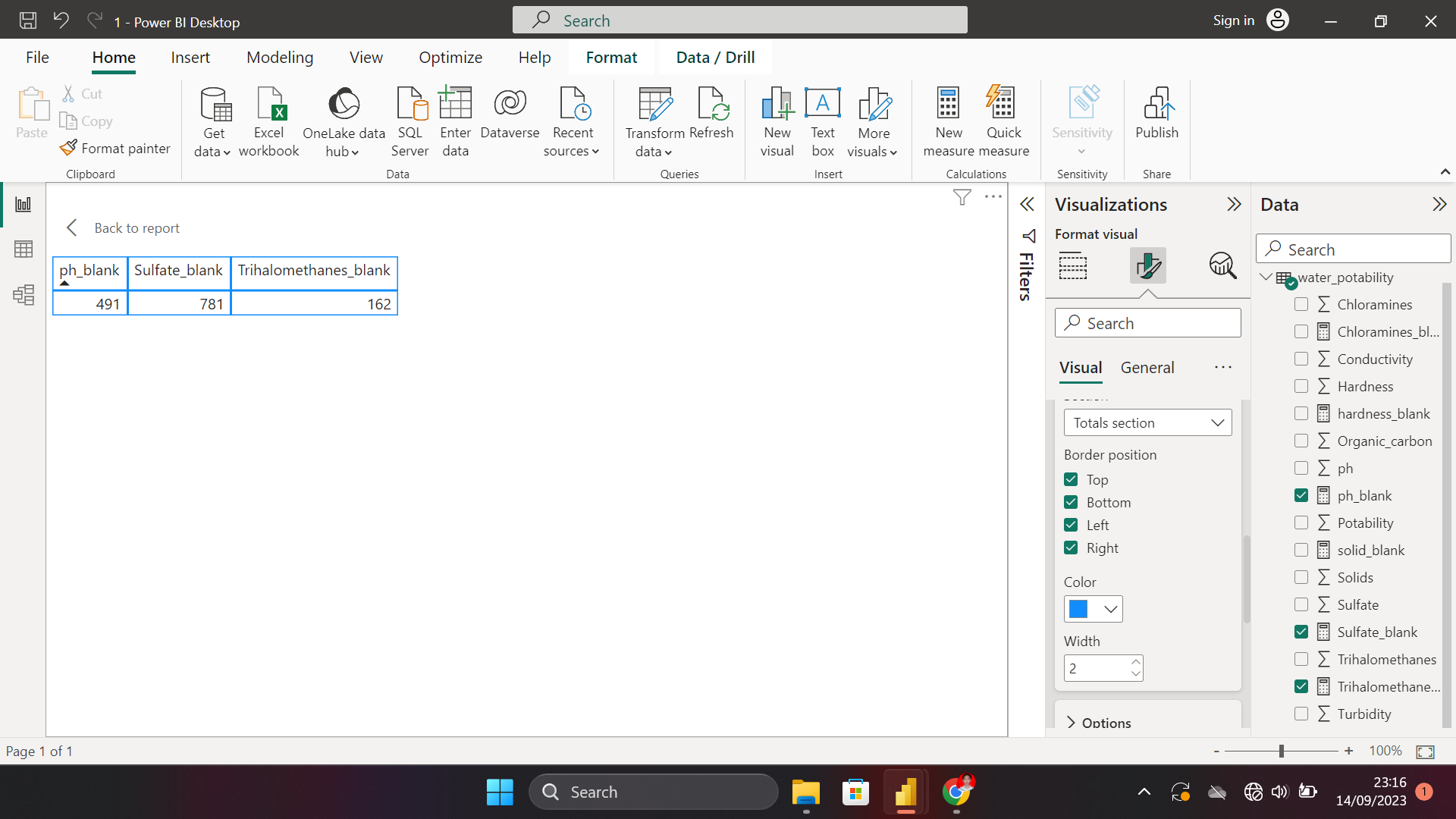
1. Mengecek missing value

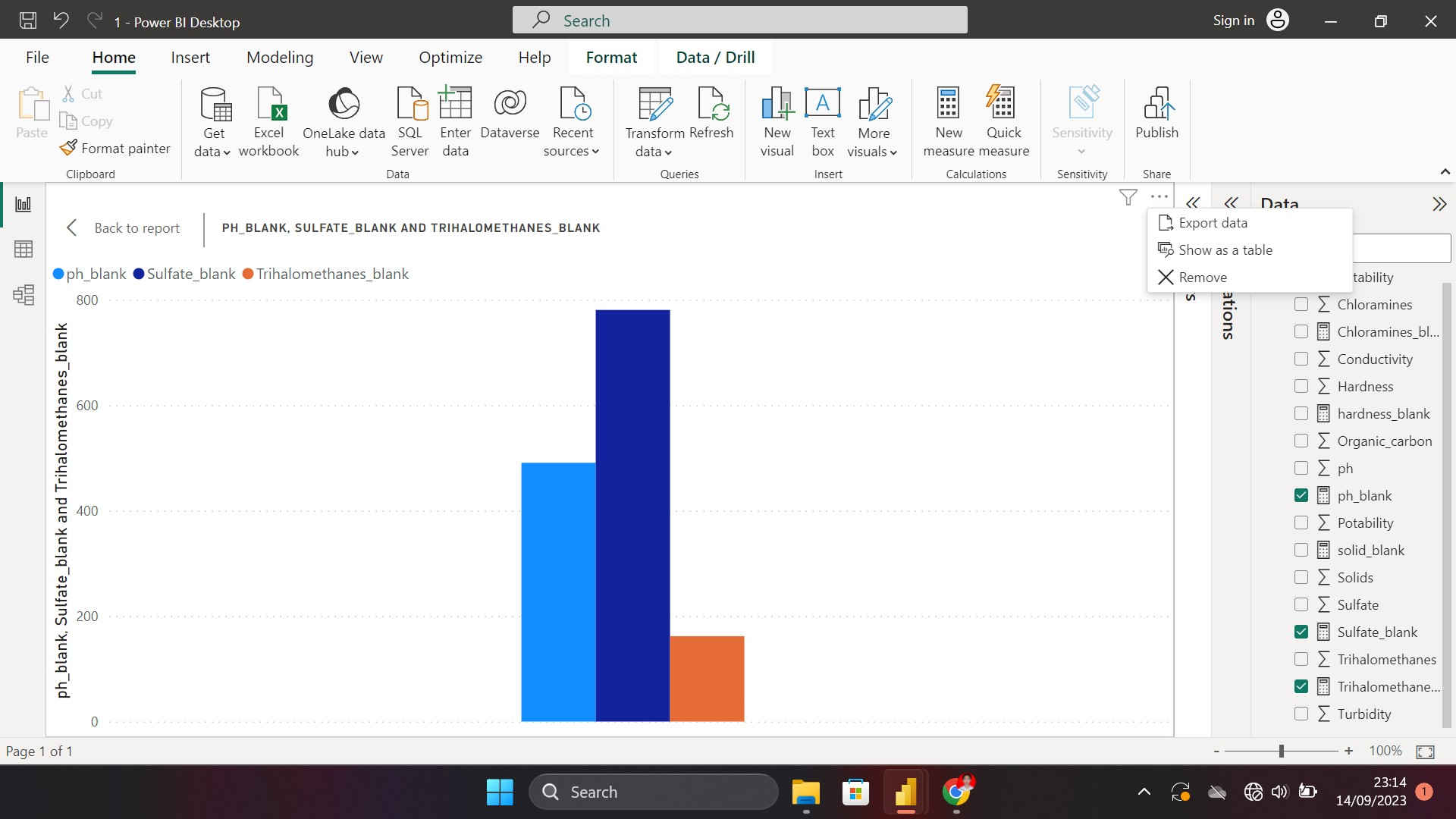
terdata yang NULL atau kosong pada 1434 baris data dengan detail sebagai berikut :





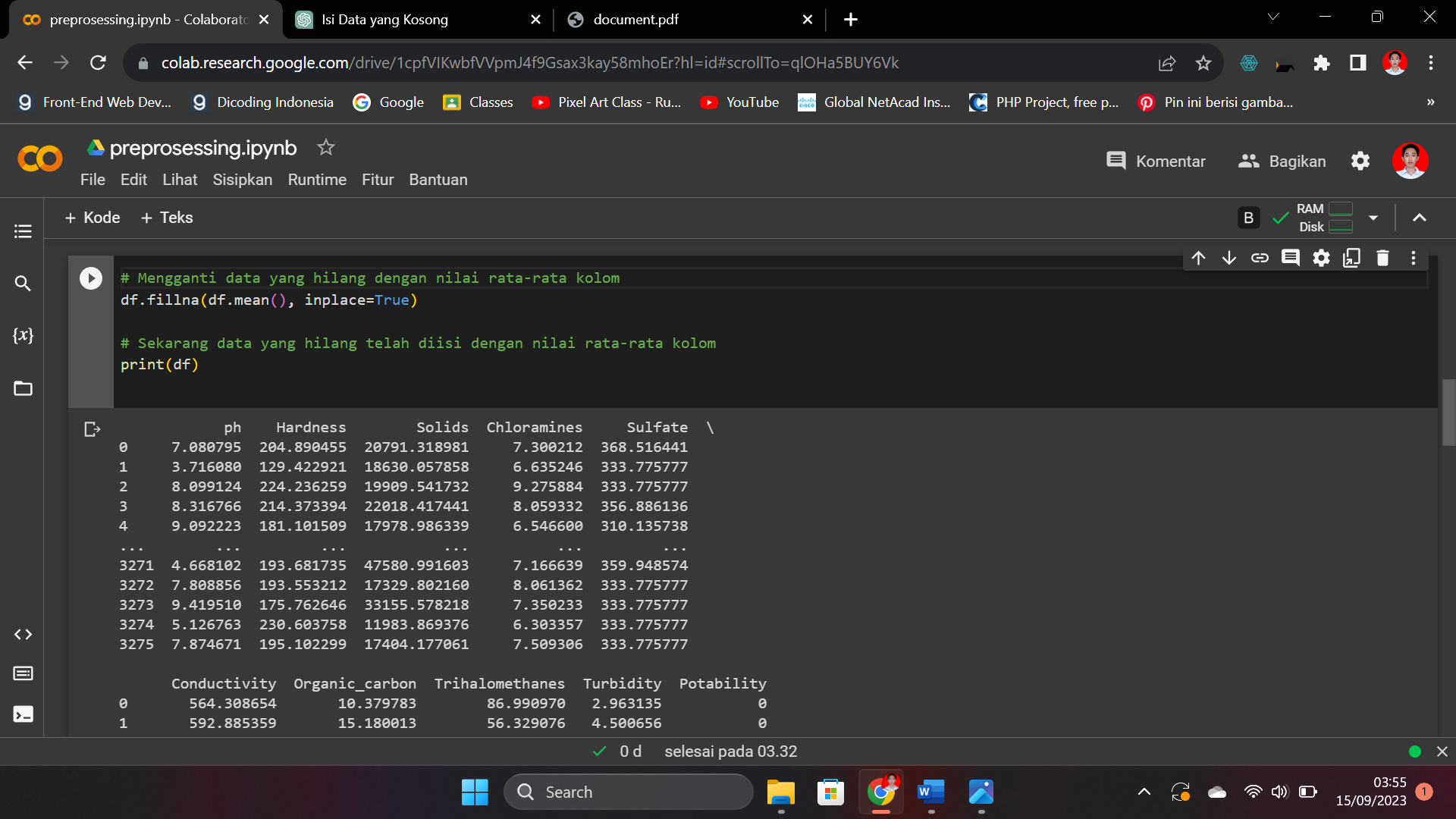
Dari gambar diatas dapat dilihat ada missing value di kolom ph 15% ,sulfate 24% dan Trihalomethanes 5%. Dengan jumlah detail sebagai berikut:



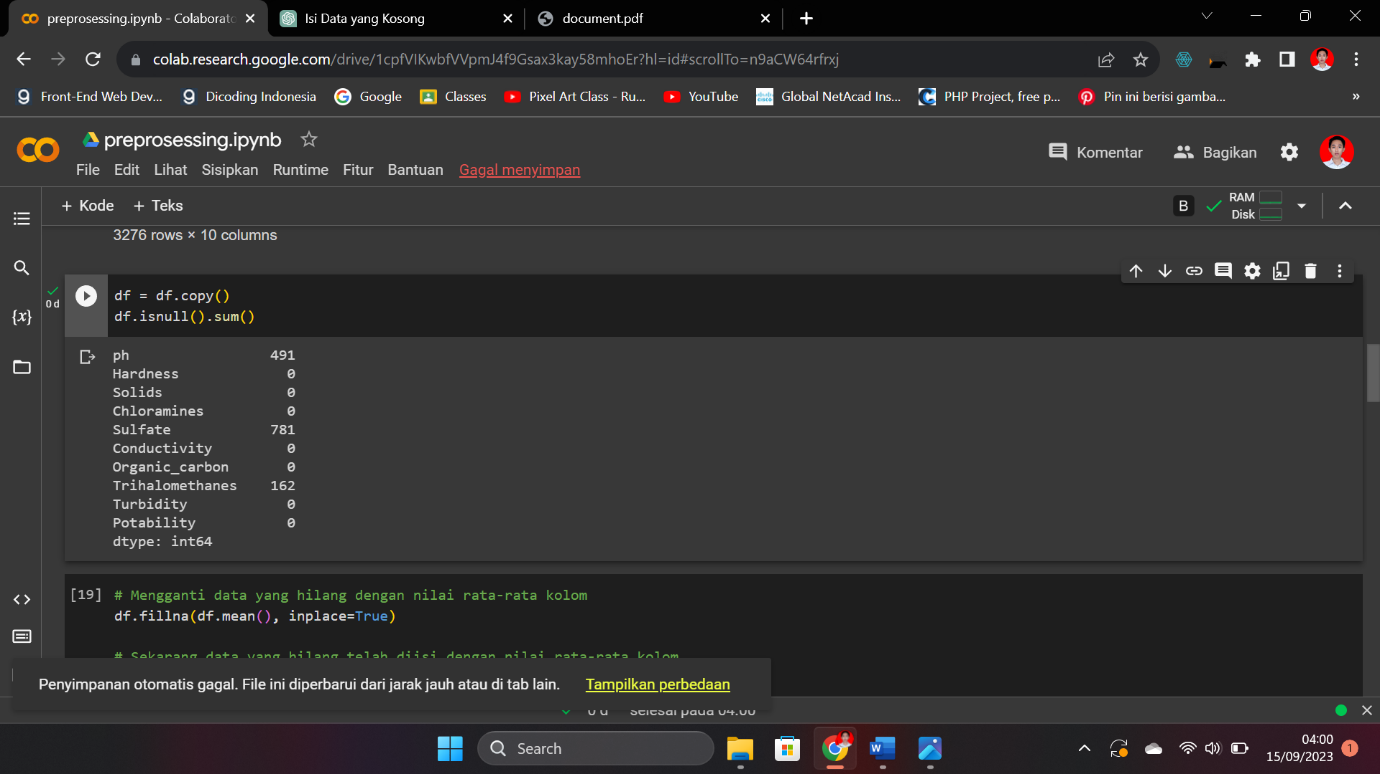
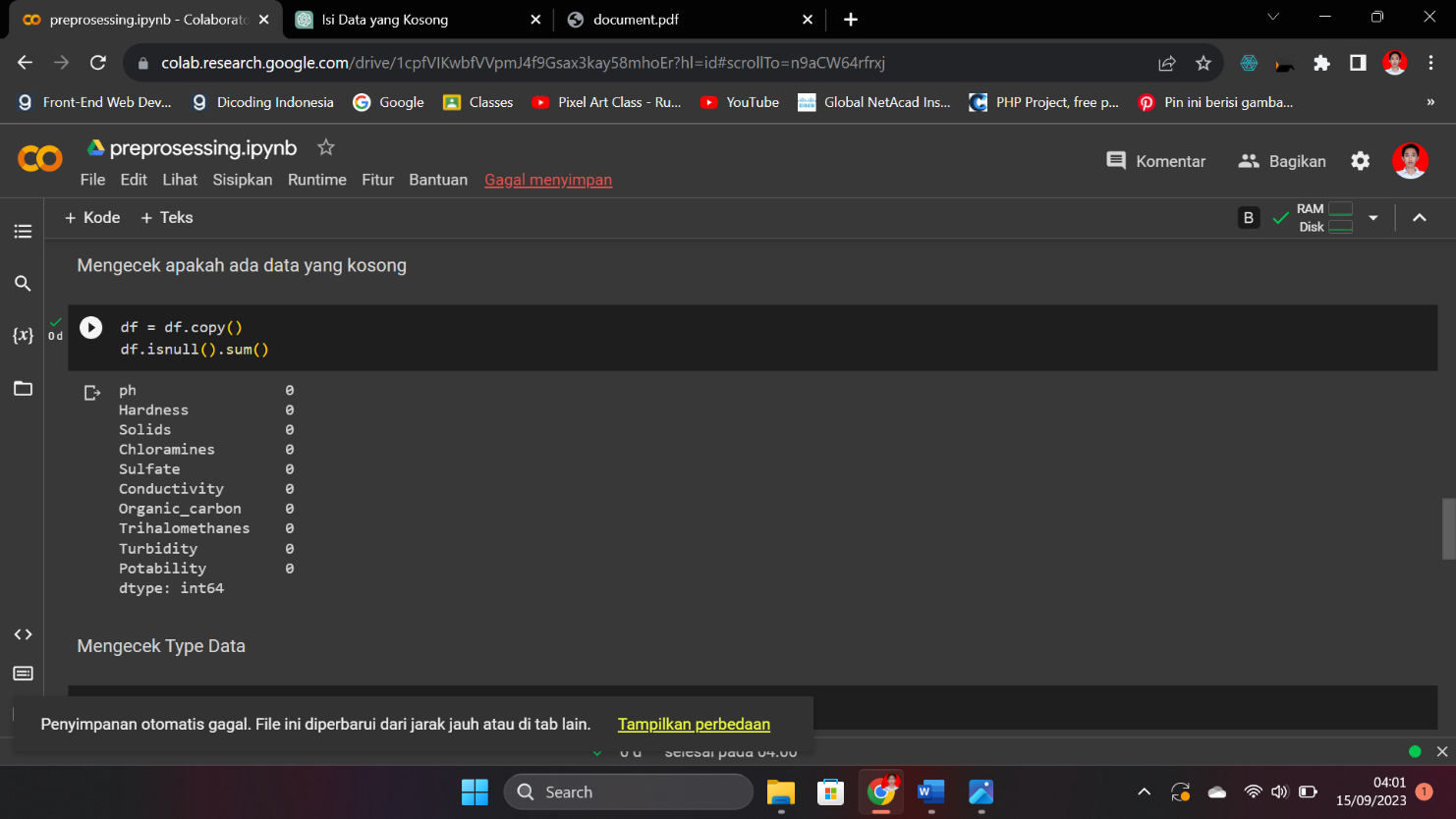


1. Mengisi missing value

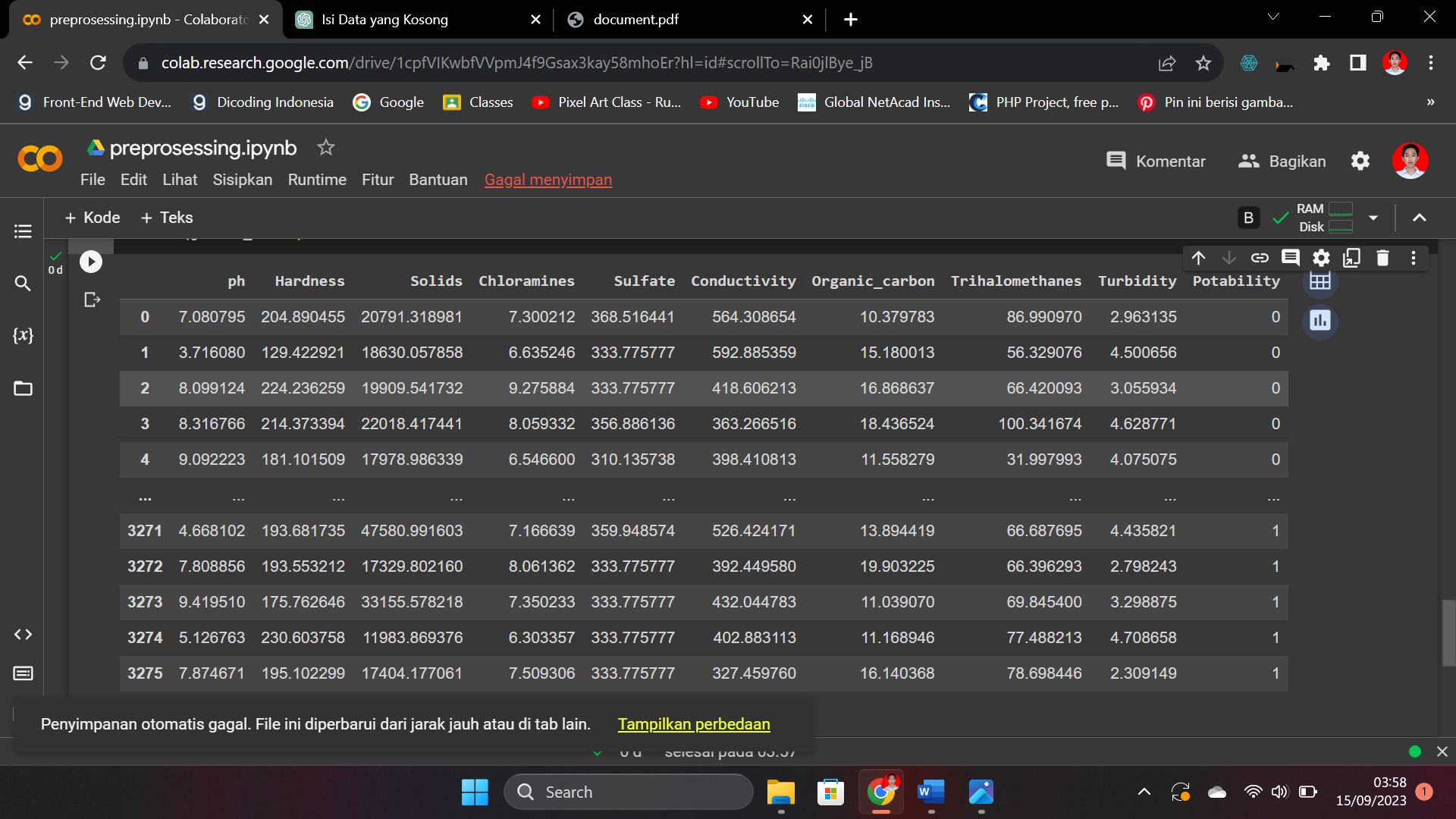
Untuk mengisi missing value menggunakan bantuan library python dengan menggunakan hasil hitung rata rata kolom



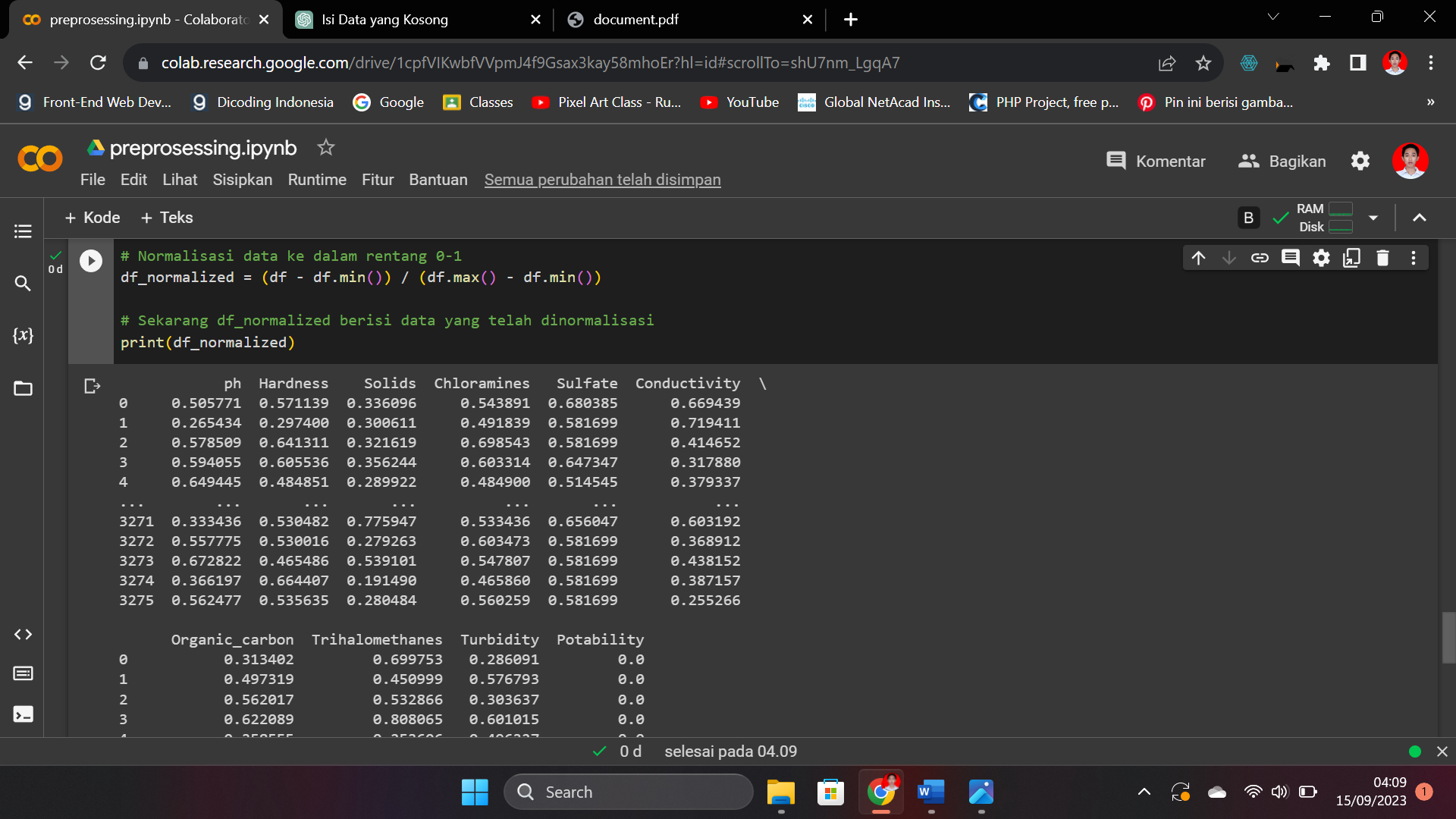
Perbandingan sebelum diisi dengan sesudah di isi



Tampilan data setelah di isi missing valuenya



1. Melakukan normalisasi dengan rentang 0-1 menggunakan rumus mixmax



1. Melakukan balancing yang berfungsi untuk mengatasi masalah ketidakseimbangan kelas dalam dataset saat Anda bekerja pada tugas klasifikasi. Ketidakseimbangan kelas terjadi ketika jumlah sampel atau instansi dalam satu kelas jauh lebih banyak atau jauh lebih sedikit daripada kelas lain dalam dataset.

