Nama : Harsena Argretya

Nim : 22.11.4958

**Collection Private Dataset**

HUBUNGAN KEPADATAN PENDUDUK DAN TINGKAT POLUSI UDARA DI DKI  
JAKARTA TAHUN 2024

1. Dari Tugas week 1-2 , Lakukan proses Collection Privat Dataset,

Lakukan prosesnya secara steap by steap dan dikumpulkan dalam

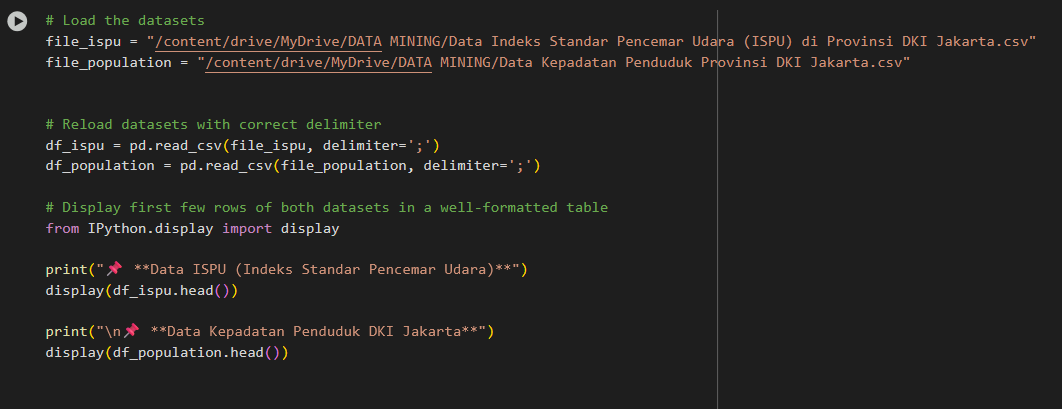
bentuk pdf serta upload hasil datasetnya.

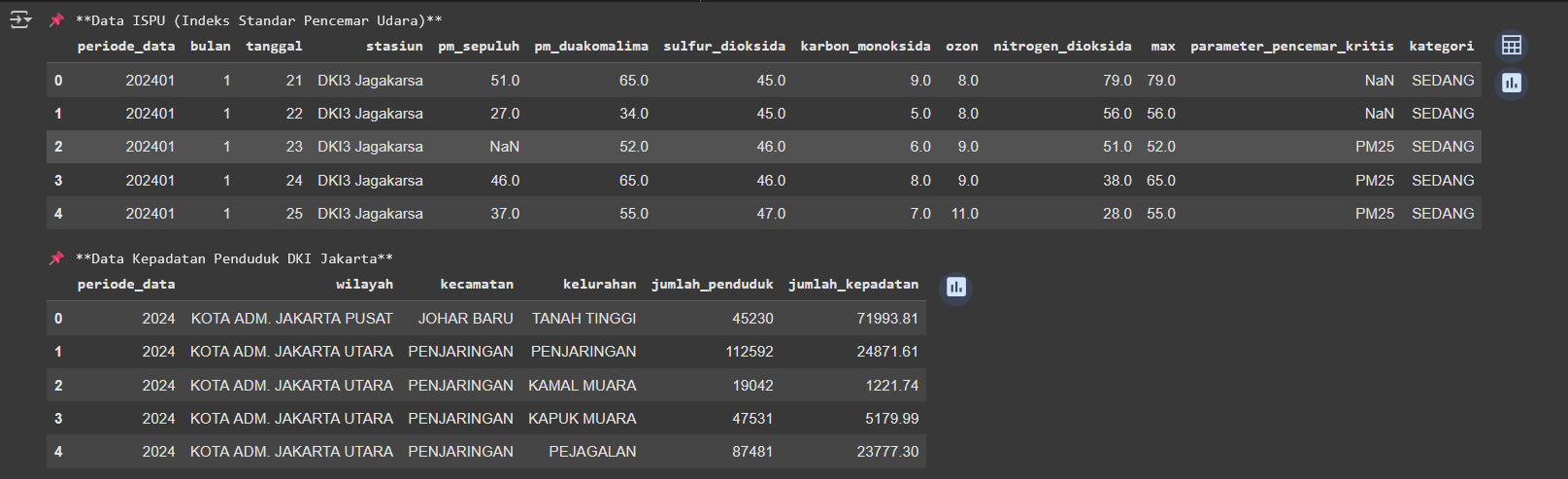
1. **Data Selection (Seleksi Data)**

Jawab

Dua dataset yang saya pilih untuk analisis:

* Data ISPU (Indeks Standar Pencemaran Udara) di DKI Jakarta 2024
* Data Kepadatan Penduduk di DKI Jakarta 2024



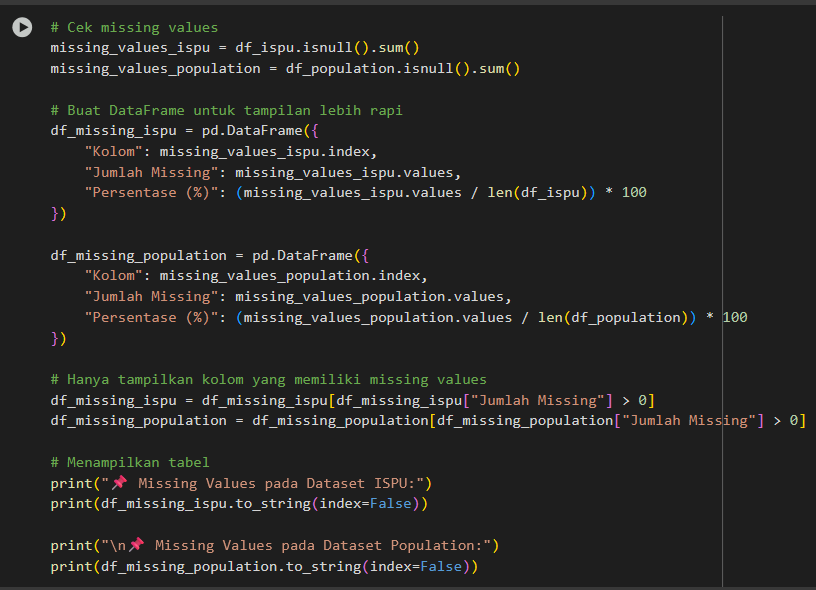


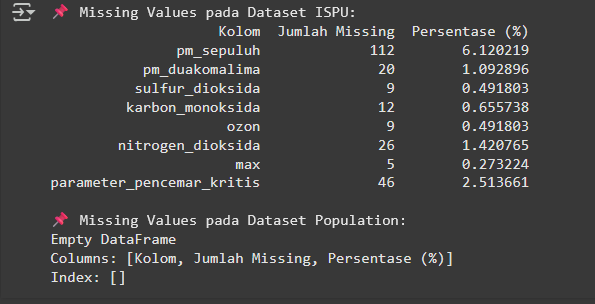
1. **Data Cleaning (Pembersihan Data)**

Jawab

Beberapa langkah data cleaning :

* Pengecekan missing value

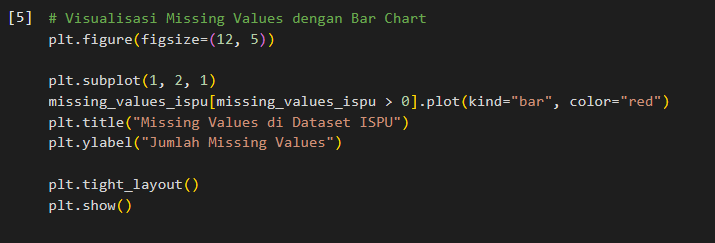


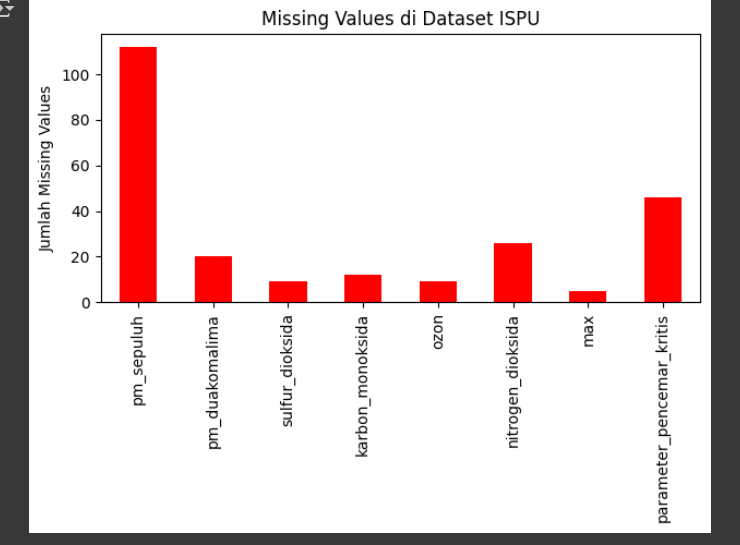


Penjelasan :

Pada Dataset ISPU terjadi beberapa missing value dan pada data Populasi tidak ada missing value

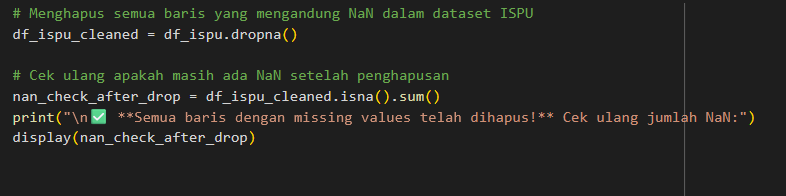
* Visualisasi missing value menggunakan bar chart.

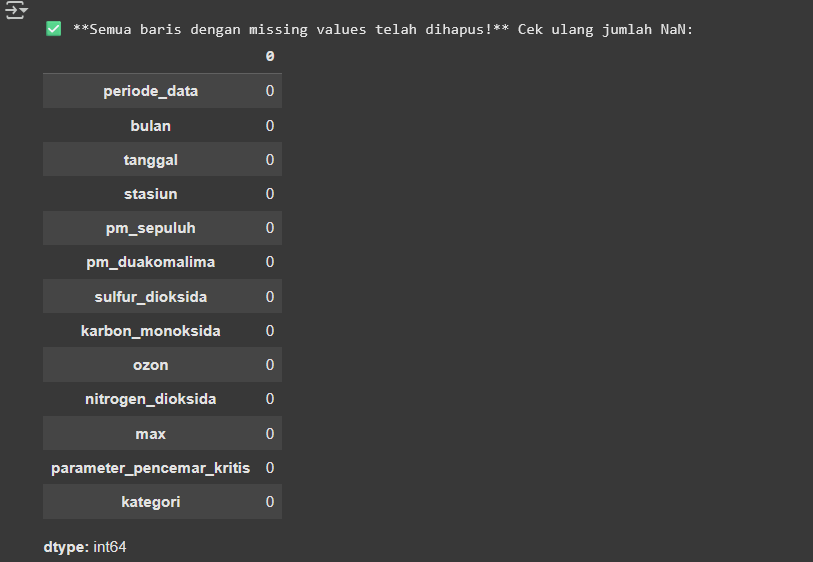




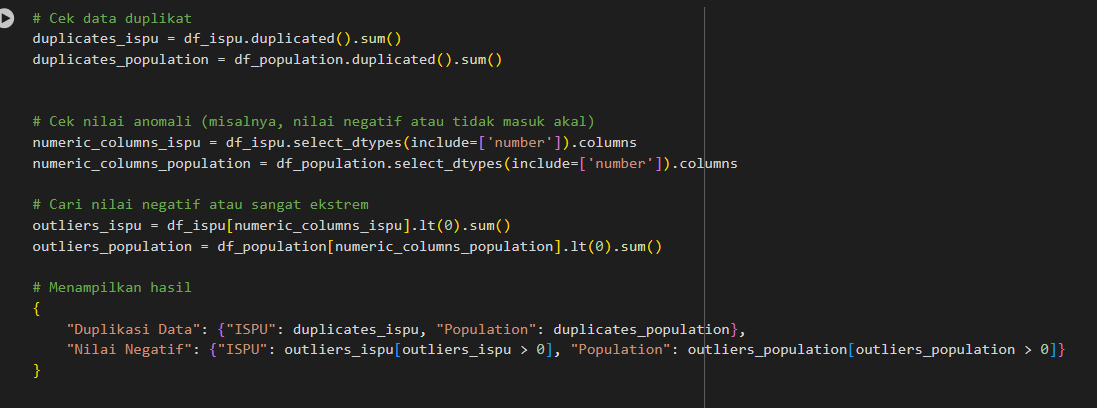
Penjelasan : Terlihat jelas dari bar chart missing value paling banyak pada kolom pm\_sepuluh

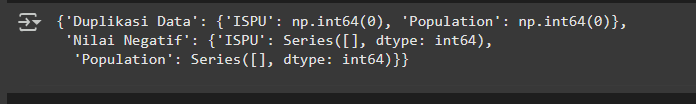
* Missing values di-dropna langsung (dropna())





* Cek duplikat data

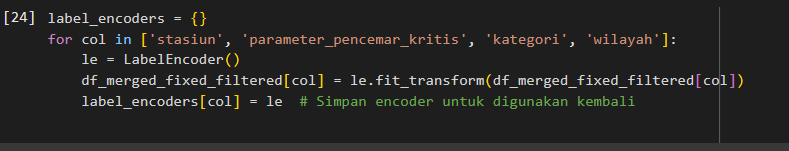


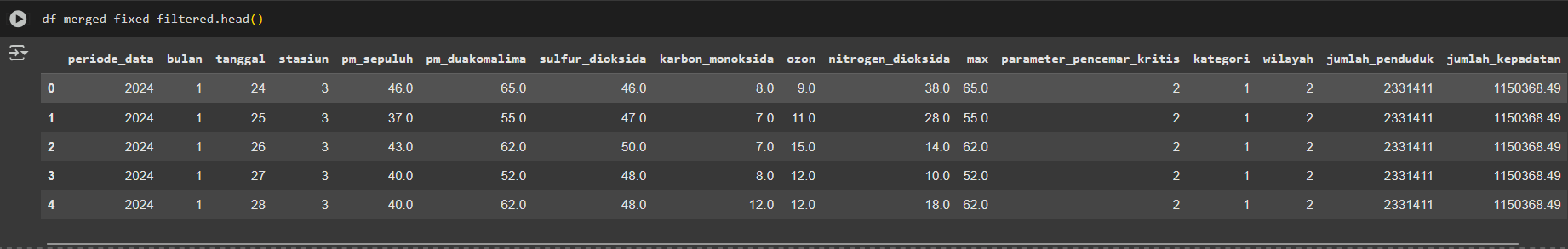


1. **Data Transformation (Transformasi Data)**

Jawab

* Label Encoding digunakan untuk mengubah data kategori menjadi numerik menggunakan LabelEncoder().



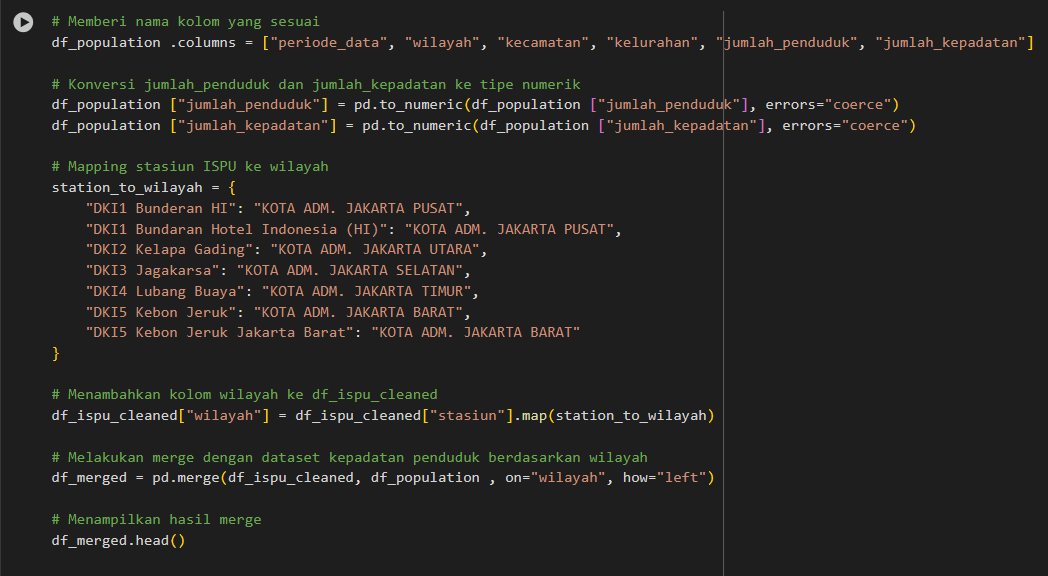


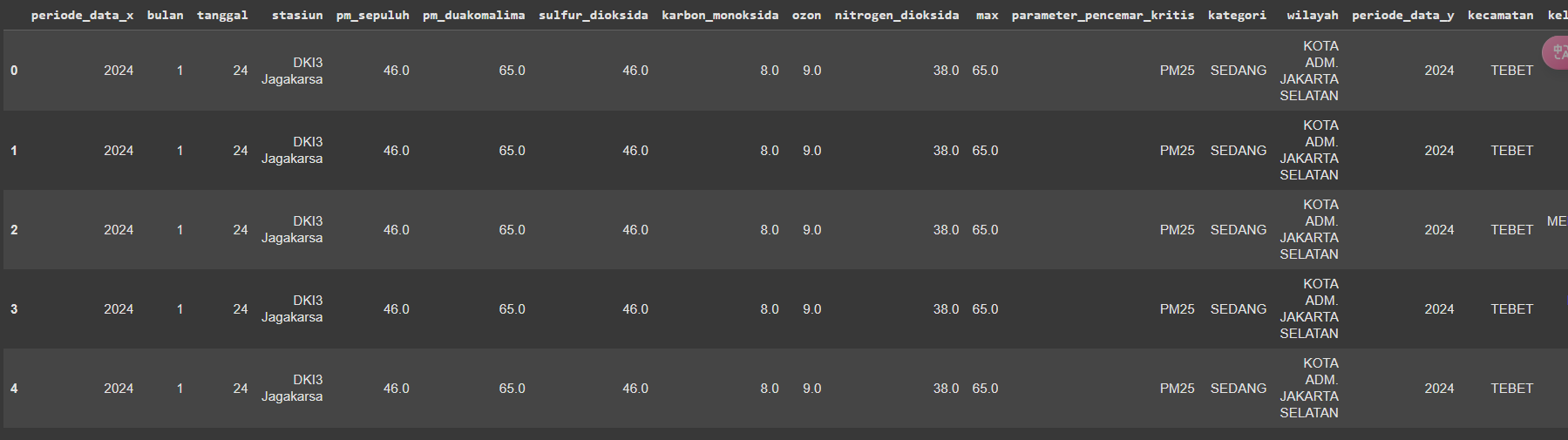
* Mengganti semua nilai dalam kolom ‘periode\_data’ menjadi 2024



1. **Data Integration (Integrasi Data)**

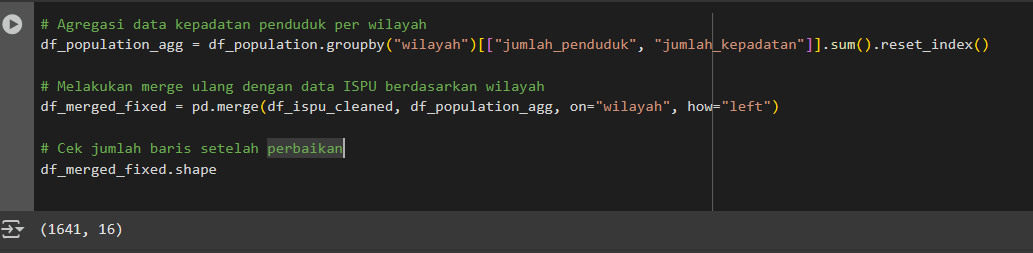
* Mapping wilayah dari stasiun ISPU: Stasiun ISPU dipetak ke wilayah administrasi DKI Jakarta

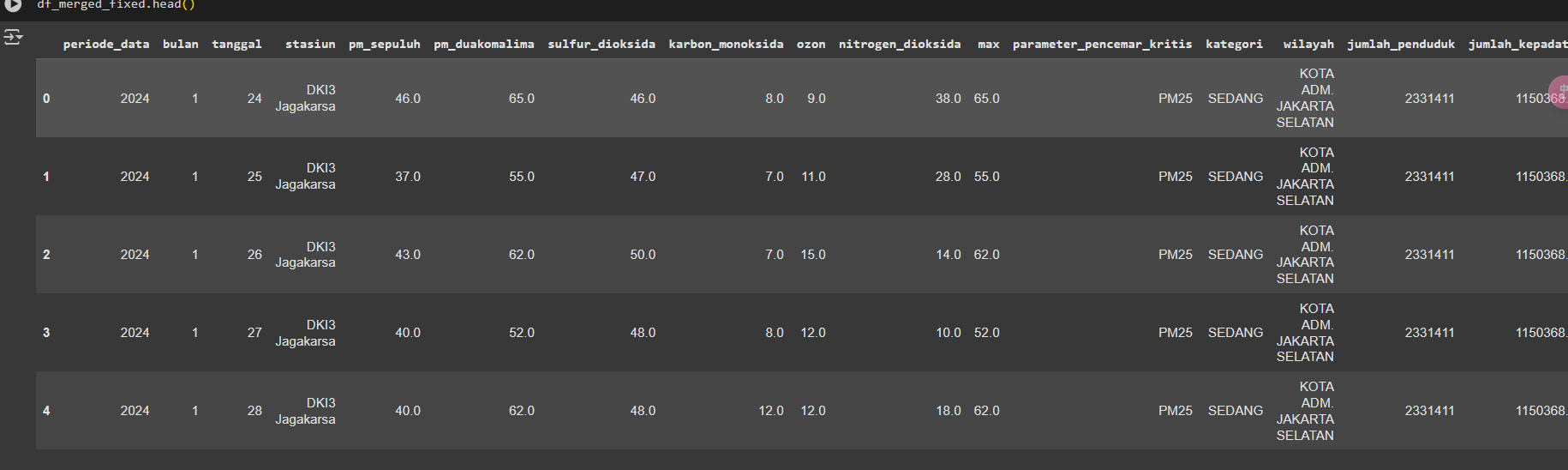
****

****

* Merge antara Data ISPU dan Data Kepadatan Penduduk:

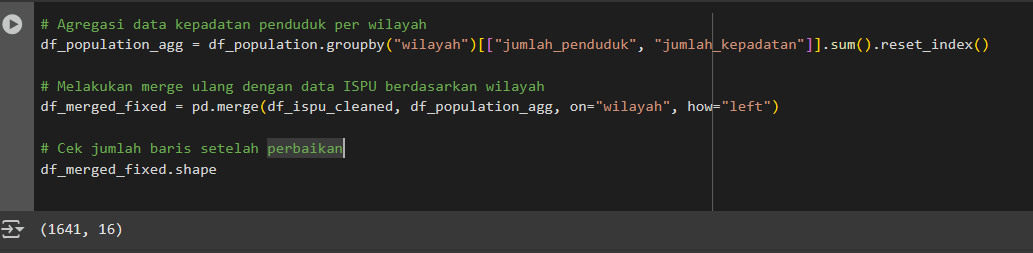
1. Data ISPU digabungkan dengan kepadatan penduduk berdasarkan kolom “wilayah”

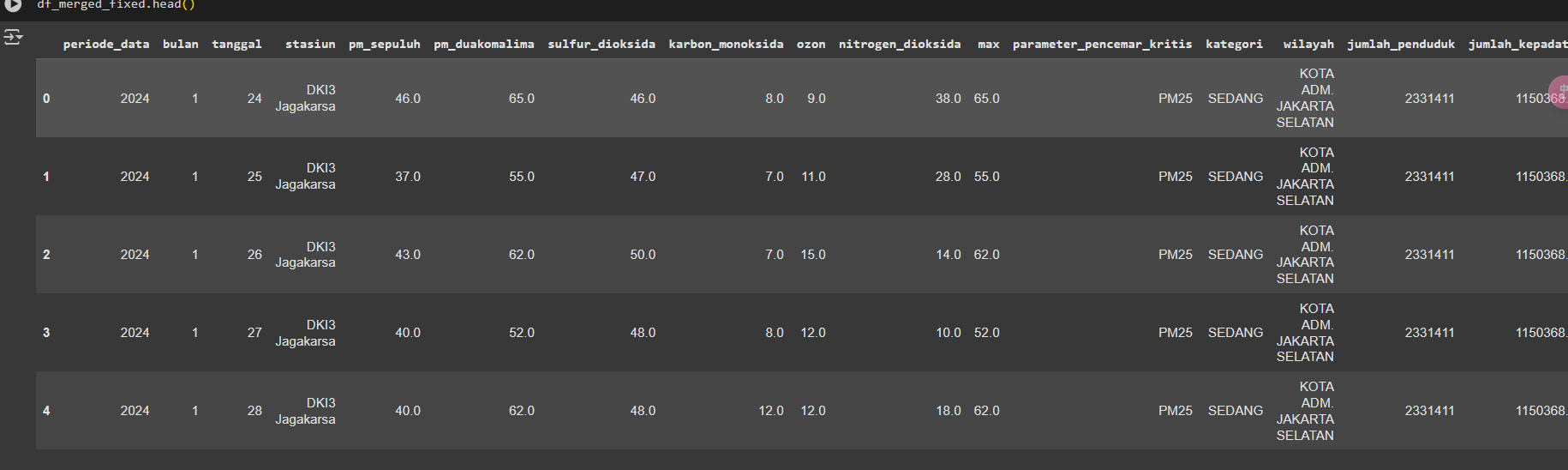




1. **Data Reduction (Reduksi Data)**

* Agregasi kepadatan penduduk per wilayah untuk menyederhanakan data.





1. Berikan pendekatan sesuai dengan collection dataset yang didapatkan

pada soal 1 terkait Missing Value Imputation dan Class Imbalance.

1. Missing Value