Langkah untuk melakukan data cleaning pada dataset sebagai berikut:

1. Memeriksa Missing Values:

* Mengidentifikasi kolom yang memiliki nilai-nilai yang hilang.
* Mengambil keputusan bagaimana menangani nilai-nilai yang hilang tersebut (misalnya, mengganti dengan nilai rata-rata, median, atau menghapus baris).

2. Memeriksa Duplikasi:

* Mengidentifikasi baris-baris yang duplikat dan menghapusnya.

3. Memeriksa Konsistensi Tipe Data:

* Memastikan setiap kolom memiliki tipe data yang benar (misalnya, `age` adalah numerik, `gender` adalah kategori).

4. Menangani Outliers:

* Mengidentifikasi nilai-nilai outlier dalam kolom-kolom numerik seperti `age`, `avg\_glucose\_level`, dan `bmi`.
* Mengambil keputusan apakah akan menghapus atau menyesuaikan outliers tersebut.

5. Memeriksa Data Inconsistencies:

* Memeriksa apakah ada nilai yang tidak konsisten, seperti kesalahan ejaan dalam kolom kategori (`gender`, `work\_type`, `smoking\_status`).

Link spreadsheet :