**KELOMPOK:**

Moh. Naufal Faqih 10222044

Firman Firdaus 10222033

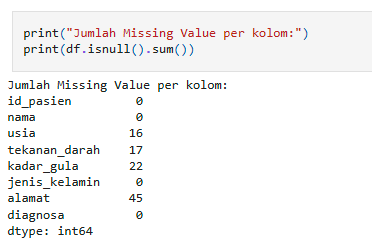
Ryan Azis S. 10222041

**Studi Kasus: Analisis Data Pasien Klinik Kesehatan**

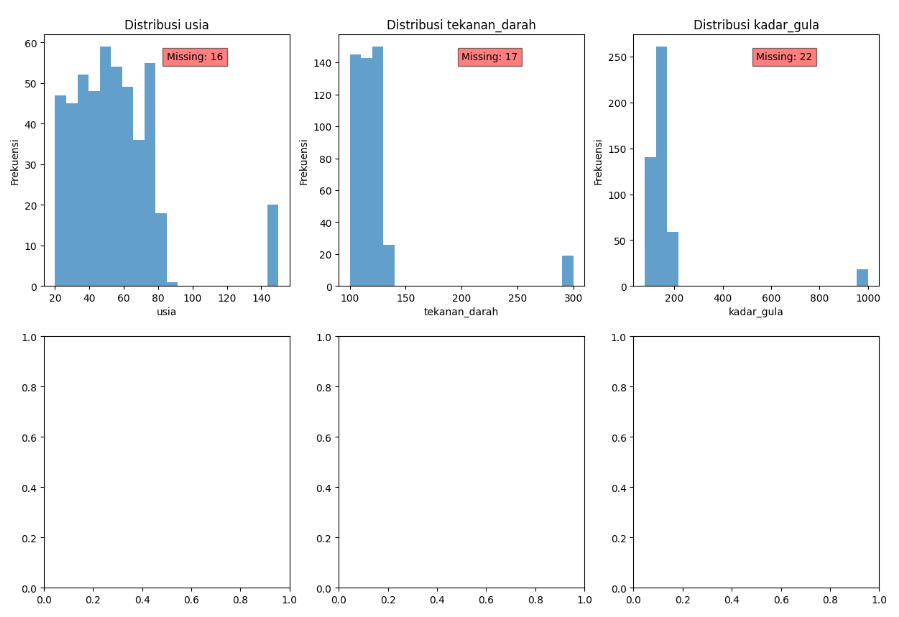
1. Memuat Dataset Pasien



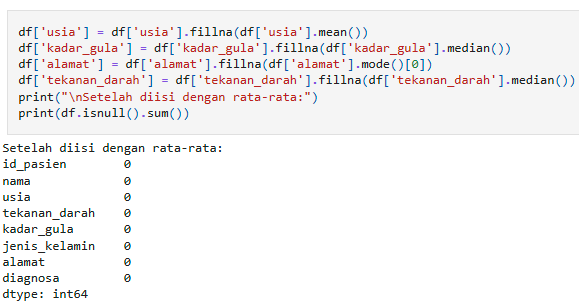
1. Mendeteksi Missing Value



1. Visualisasi Sebelum Data Dibersihkan

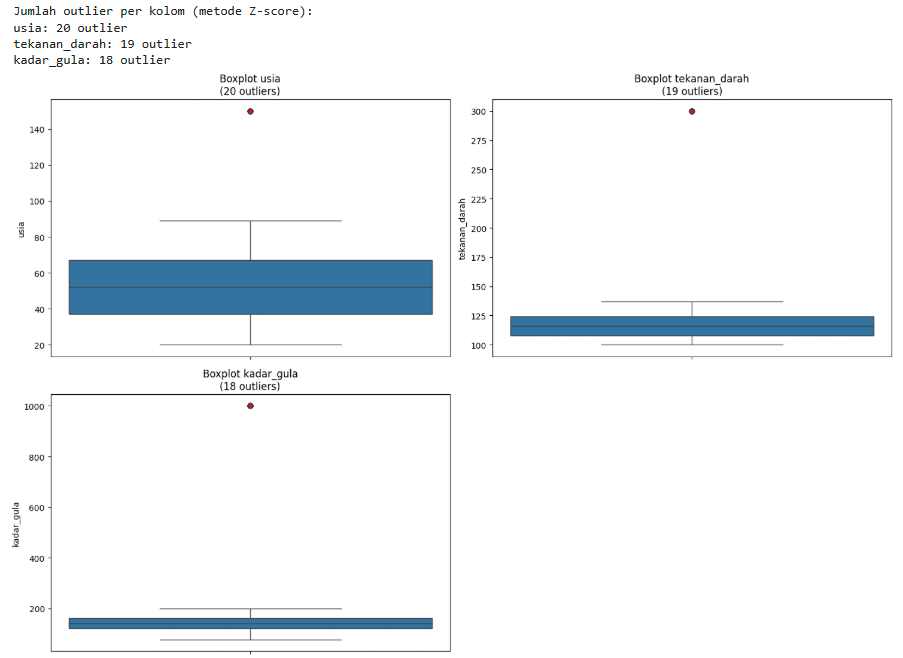


1. Menangani beberapa Missing Value

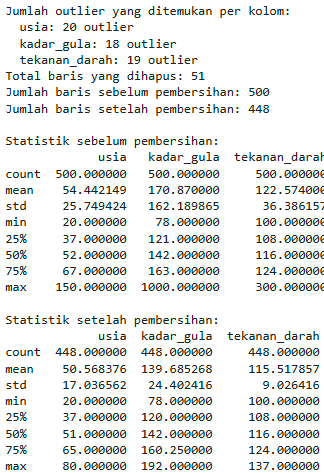


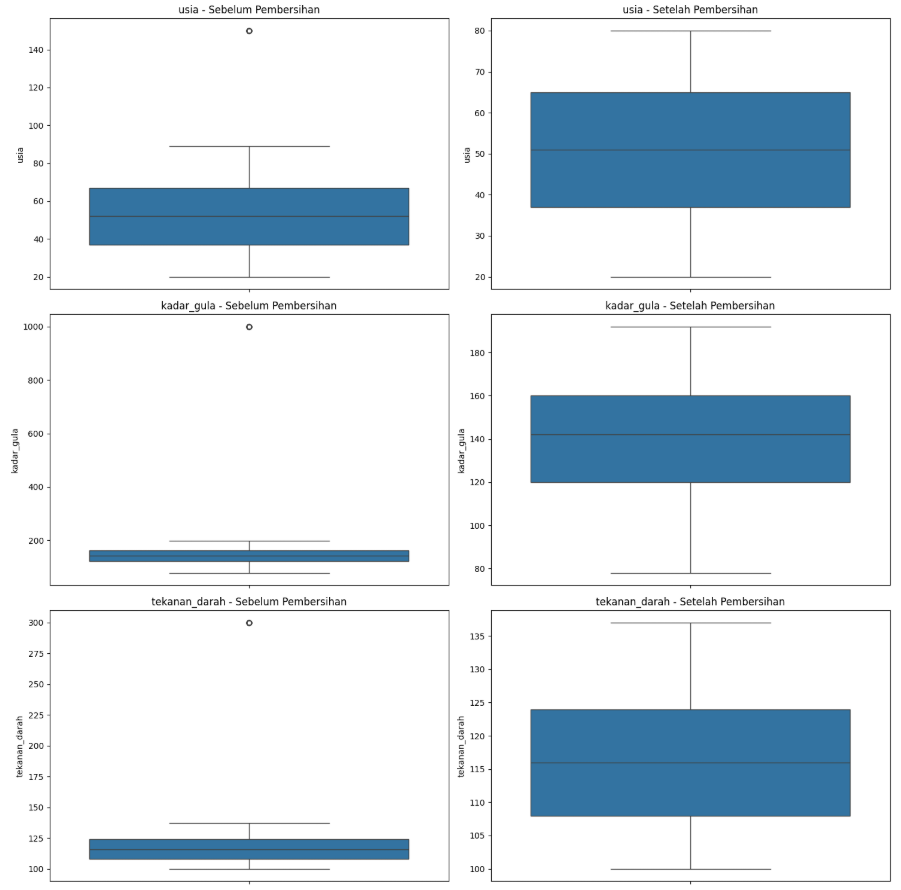


1. Mendeteksi Outlier



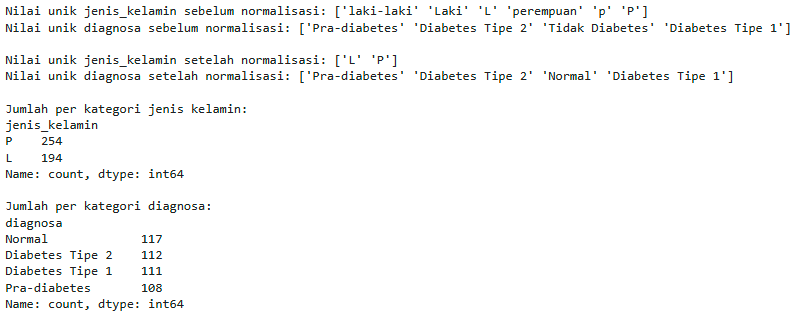
1. Tangani Outlier dengan Teknik Remove



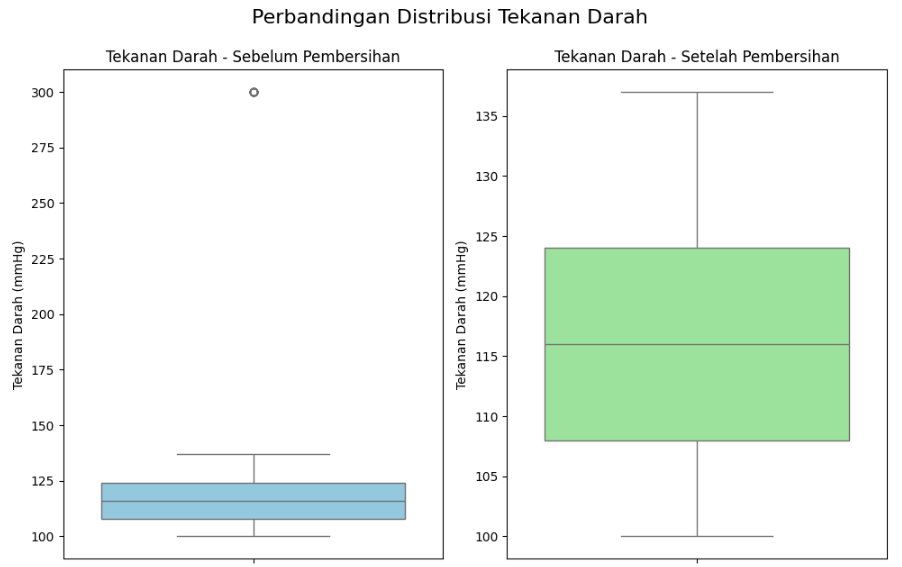


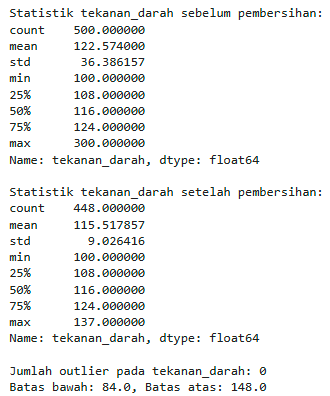


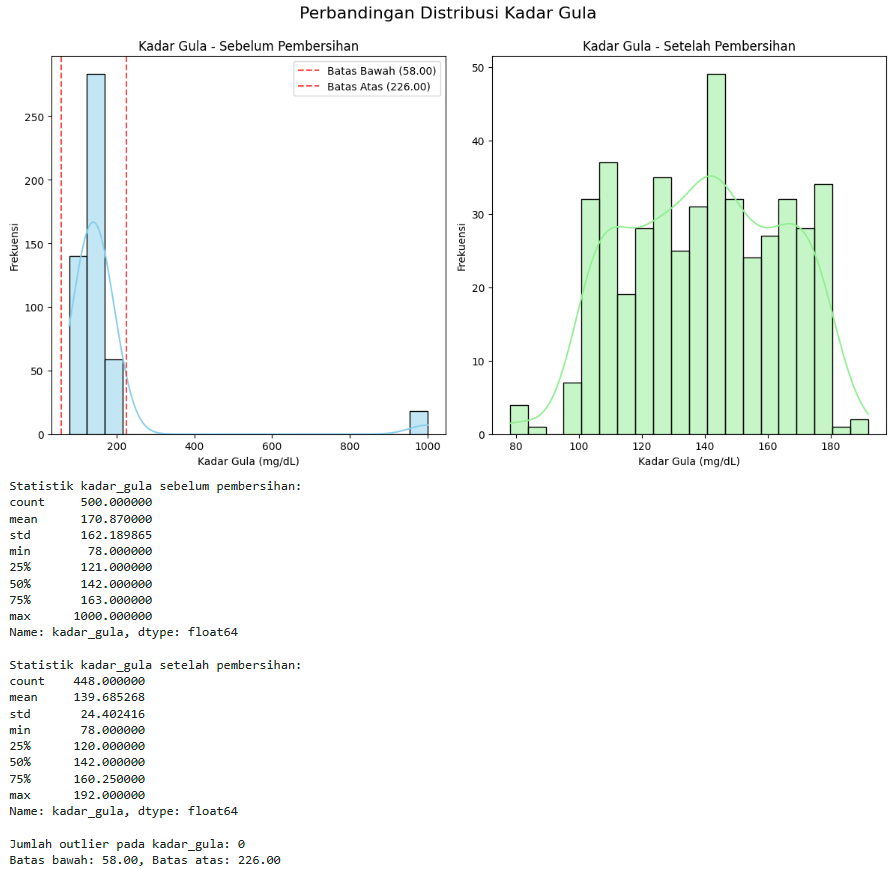
1. Normalisasi Kolom Gender dan Diagnosa

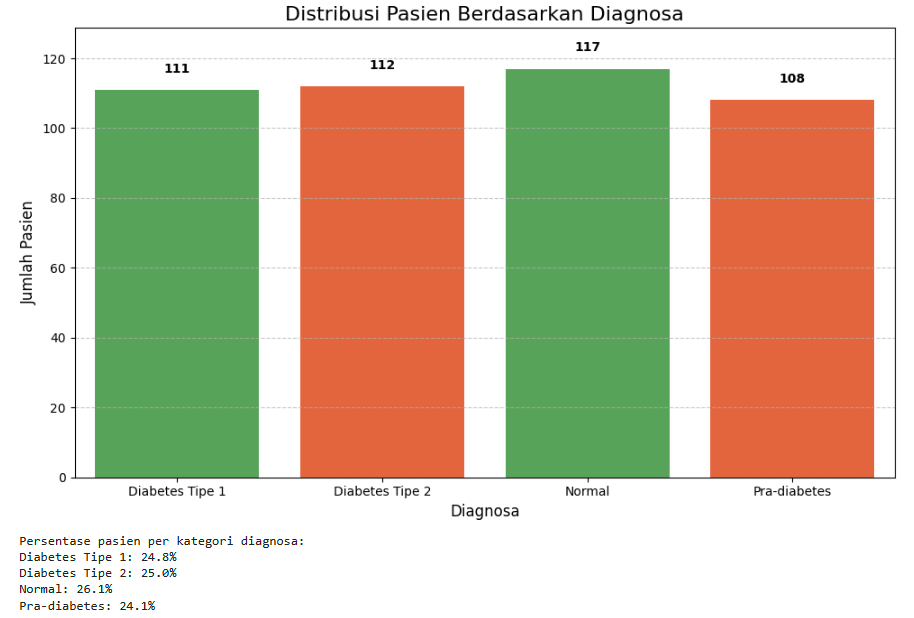


1. Visualisasi









1. Kesimpulan
   1. Penanganan Missing Value

Pengisian missing values dengan nilai statistik yang tepat (mean, median, mode) memastikan:

* Dataset lengkap tanpa nilai kosong, sehingga analisis dapat dilakukan tanpa gangguan.
* Lebih representatif karena menggunakan nilai-nilai yang mendekati distribusi data asli.
* Jumlah data tetap terjaga mempertahankan kekuatan statistik dataset.
  1. Penannganan Outlier

Penghapusan outlier memberikan dampak positif pada kualitas analisis:

* **Distribusi Data Lebih Normal**: Terlihat dari visualiasi boxplot dan histogram yang menunjukkan distribusi lebih seimbang setelah pembersihan.
* **Statistik Deskriptif Lebih Representatif**: Nilai mean dan standar deviasi tidak lagi terpengaruh oleh data yang ekstrem nilainya.
* **Peningkatan Keandalan Analisis**: Mengurangi risiko bias pada proses analisis selanjutnya.
  1. Perubahan Pada Visualiasi

Perbandingan visualisasi menunjukkan perubahan signifikan

* **Boxplot Tekanan Darah**: Hilangnya titik-titik ekstrem pada visualisasi setelah pembersihan, menunjukkan distribusi yang lebih bagus.
* **Histogram Kadar Gula**: Kurva distribusi menjadi lebih mendekati distribusi normal setelah penghapusan outlier.
* **Bar Chart Diagnosa**: Menunjukkan perbandingan proporsi pasien dengan diabetes dan non-diabetes yang lebih akurat.
  1. Normalisasi Data Kategorikal

Standarisasi nilai pada kolom jenis\_kelamin dan diagnosa menghasilkan:

* **Konsistensi Format Data**: Semua nilai gender menjadi format standar ‘L’ dan ‘P’.
* **Kemudahan Interpretasi**: Nilai diagnosa yang konsisten memudahkan analisis dan visualisasi.
* **Peningkatan Keakuratan Perhitungan**: Pengelompokkan yang tepat saat analisis kategorikal.

Secara keseluruhan, proses pembersihan data telah mengubah dataset yang semula memiliki missing values dan outlier menjadi dataset yang lebih berkualitas dan representative. Visualisasi setelah pembersihan menunjukkan gambaran yang lebih akurat tentang distribusi dan karakteristik pasien diabetes yang sangat penting untuk analisis medis dan pengambilan keputusan klinis.