

P R E S E N T A T I O N

Analisis Risiko Stroke



BY : Melzha Amanda



github.com/amandamelzha/Analisis-Risiko-Stroke/



My Previous Project

**Transaction Effectiveness &
Revenue Analysis – Amazon Sales**

Evaluasi Efektivitas Transaksi dan Dampak
terhadap Revenue

PRESENT BY: MELZHA AMANDA

Create with Canva

amazon Canva

1 / 22

DESKRIPSI

- Menganalisis data transaksi dan membuat KPI (Cancel Rate & Pending Rate).
- Membangun dashboard interaktif untuk memantau performa pendapatan dan status transaksi.
- Mengidentifikasi potensi kebocoran pendapatan yang disebabkan oleh tingginya transaksi batal dan tertunda.
- Memberikan rekomendasi berbasis data untuk meningkatkan completion rate dan efisiensi operasional.

Tools: Tableau, Excel

Link: [Klik Disini](#)

My Previous Project



DESKRIPSI

- Membersihkan dan mentransformasi dataset asuransi menggunakan Python (Pandas).
- Melakukan Exploratory Data Analysis (EDA) untuk menganalisis pola biaya berdasarkan usia, BMI, dan status merokok.
- Mengembangkan fitur segmentasi serta menghitung metrik untuk perbandingan biaya.
- Membangun dashboard interaktif menggunakan Tableau untuk memvisualisasikan faktor pendorong biaya kesehatan.
- Menghasilkan insight yang mengidentifikasi kebiasaan merokok dan BMI tinggi sebagai kontributor utama peningkatan biaya.

Tools : Python, Tableau

Link: [Klik Disini](#)

My Previous Project

Campaign Performance Evaluation, Optimization & A/B Testing Analysis

Klik disini: [Dashboard](#)



DESKRIPSI

- Membersihkan dan mentransformasi dataset performa campaign digital marketing Facebook dan Instagram menggunakan Python (Pandas).
- Melakukan Exploratory Data Analysis (EDA) untuk menganalisis performa funnel pemasaran.
- Mengembangkan metrik turunan seperti CTR, conversion rate, dan cost per purchase untuk mengevaluasi efektivitas campaign.
- Menganalisis segmentasi audience berdasarkan usia dan gender untuk menemukan segmen dengan performa pembelian tertinggi.
- Mengevaluasi variasi campaign sebagai bagian dari A/B testing untuk mengidentifikasi strategi dengan conversion rate terbaik.
- Membangun dashboard interaktif menggunakan Tableau untuk memvisualisasikan funnel performance, efisiensi biaya, dan segmentasi audience.

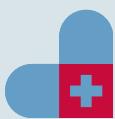
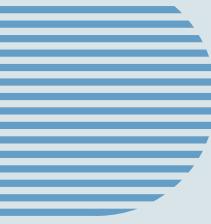
Tools : Python, Tableau

Link: [Klik Disini](#)



Problem Statement

Stroke merupakan salah satu penyebab utama kematian dan disabilitas secara global. Institusi kesehatan membutuhkan pendekatan berbasis data untuk mengidentifikasi faktor risiko yang paling berpengaruh terhadap kejadian stroke, guna mendukung strategi pencegahan yang lebih efektif dan tepat sasaran.



Borcalle Medical



Tujuan Analisa:

Insight yang dihasilkan dari analisis ini diharapkan dapat membantu tenaga medis, rumah sakit, dan pemerintah dalam:

- Deteksi dan pencegahan dini stroke
- Penyusunan program kesehatan berbasis data
- Peningkatan kesadaran masyarakat terhadap faktor risiko stroke

Bussines Question:

- Faktor apa yang paling berkontribusi terhadap kejadian stroke?
- Pada kelompok usia berapa risiko stroke meningkat signifikan?
- Apakah hipertensi, penyakit jantung, dan kadar glukosa tinggi meningkatkan risiko secara signifikan?
- Apakah kombinasi komorbid (misal: hipertensi + penyakit jantung) memperbesar risiko dibanding faktor tunggal?
- Kelompok demografis mana yang paling rentan terhadap stroke?



Data Understanding

HEALTCARE DATASET STROKE

Variabel	Keterangan
id	Nomor identitas unik untuk setiap pasien (tidak digunakan dalam analisis).
gender	Jenis kelamin pasien (Male, Female, Other).
age	Usia pasien (dalam tahun).
hypertension	Status hipertensi pasien: 1 = memiliki hipertensi, 0 = tidak memiliki hipertensi.
heart_disease	Status penyakit jantung: 1 = memiliki penyakit jantung, 0 = tidak memiliki penyakit jantung.
ever_married	Status pernikahan: "Yes" = pernah/sedang menikah, "No" = belum pernah menikah.
work_type	Jenis pekerjaan pasien: • private = karyawan swasta • self-employed = wiraswasta • govt_job = pegawai pemerintah • never_worked = tidak pernah bekerja • children = anak-anak (belum bekerja)
Residence_type	Jenis tempat tinggal: • Urban = daerah perkotaan • Rural = daerah pedesaan
avg_glucose_level	Rata-rata kadar glukosa darah pasien.
bmi	Indeks Massa Tubuh (Body Mass Index).
smoking_status	Status merokok: • never smoked = tidak pernah merokok • formerly smoked = pernah merokok (sudah berhenti) • smokes = masih merokok • unknown = tidak diketahui
stroke	Status stroke: 1 = pernah mengalami stroke, 0 = tidak pernah mengalami stroke.

Dataset ini berisi informasi pasien terkait faktor risiko kesehatan dan kejadian stroke. Setiap baris merepresentasikan 1 pasien.

Total data: 4.860 pasien

PERMASALAHAN KUALITAS DATA

1

Memeriksa struktur data, nama kolom, dan tipe data

#	Column	Non-Null Count	Dtype
0	id	4860 non-null	int64
1	gender	4860 non-null	object
2	age	4860 non-null	float64
3	hypertension	4860 non-null	int64
4	heart_disease	4860 non-null	int64
5	ever_married	4860 non-null	object
6	work_type	4860 non-null	object
7	Residence_type	4860 non-null	object
8	avg_glucose_level	4860 non-null	float64
9	bmi	4458 non-null	float64
10	smoking_status	4860 non-null	object
11	stroke	4860 non-null	int64

Seluruh kolom telah sesuai tipe datanya

2

Cek konsistensi variabel categorical

	column	dtype	n_unique	unique_values
0	id	int64	4831	[21202, 113623, 64489, 70429, 163939, 52234, 56179, 67465, 31091, 110303, 165438, 138843, 12336, 52220, 71192, 61338, 31806, 35276, 173124, 11244, 135493, 19805, 65907, 167502, 20655, 128325, 65321, 25287, 10390, 170311, 39661, 107690, 8960, 68438, 51958, 39017, 137269, 140506, 10351, 134900, 47803, 56799, 59663, 53478, 156030, 102815, 59275, 136736, 46093, 61238, 2244, 15964, 141287, 174624, 60602, 44583, 63467, 21521, 49785, 59649, 11605, 32766, 169086, 35123, 60675, 152102, 58820, 45788, 143953, 137386, 23946, 166541, 64174, 44978, 24163, 175687, 68094, 67620, 56734, 41615, 72011, 137026, 27626, 32826, 19364, 145853, 147597, 37993, 9143, 70428, 2633, 123880, 115569, 138284, 10243, 68249, 42082, 3746, 10995, 24245, ...]
1	gender	object	2	[Female, Male]
2	age	float64	1176	[27.0, 63.82, 56.0, 33.0, 78.95, 72.0, 29.0, 20.0, 34.0, 79.91, 56.04, 74.8, 73.0, 26.0, 11.0, 40.0, 70.0, 6.0, 57.88, 77.0, 72.94, 60.0, 49.0, 61.01, 16.0, 64.87, 54.0, 8.0, 80.03, 18.0, 42.0, 51.0, 62.0, 38.78, 31.94, 50.0, 79.08, 37.0, 76.0, 28.0, 73.15, 79.06, 10.0, 59.95, 44.0, 64.0, 57.24, 80.05, 9.0, 25.0, 59.81, 1.24, 48.0, 78.04, 53.0, 77.84, 77.06, 3.0, 56.09, 59.0, 39.0, 12.0, 57.1, 46.0, 30.0, 1.4, 79.73, 7.0, 50.04, 55.91, 36.0, 17.0, 32.0, 81.92, 79.13, 64.91, 13.0, 66.0, 55.0, 80.76, 65.0, 35.0, 2.0, 60.02, 75.0, 5.0, 54.91, 60.96, 80.12, 73.75, 81.0, 22.0, 24.0, 57.95, 79.96, 70.96, 45.0, 80.0, 66.18, 52.0, ...]
3	hypertension	int64	2	[0, 1]
4	heart_disease	int64	2	[0, 1]
5	ever_married	object	2	[Yes, No]
6	work_type	object	5	[Private, Govt_job, Self-employed, children, Never_worked]
7	Residence_type	object	2	[Urban, Rural]
8	avg_glucose_level	float64	3976	[80.57, 74.19, 73.02, 84.48, 126.6, 104.05, 207.58, 117.59, 106.23, 72.98, 161.83, 109.58, 87.56, 154.08, 56.33, 65.47, 91.25, 84.1, 71.02, 83.06, 220.51, 84.14, 206.53, 209.94, 94.96, 206.38, 64.55, 92.95, 67.33, 175.66, 140.52, 71.6, 73.41, 90.78, 199.78, 118.22, 83.82, 75.56, 67.02, 92.55, 173.97, 82.35, 107.74, 89.61, 189.89, 174.62, 58.03, 91.73, 56.47, 122.23, 80.75, 99.4, 215.67, 73.72, 68.68, 70.02, 150.0, 103.28, 128.97, 119.96, 108.2, 106.41, 96.43, 84.2, 221.08, 107.49, 86.36, 197.79, 60.88, 161.5, 97.31, 161.46, 204.86, 72.49, 116.04, 67.14, 124.92, 66.01, 82.83, 126.18, 87.74, 76.15, 266.59, 87.74, 74.96, 102.13, 249.53, 66.47, 67.87, 76.98, 71.5, 103.34, 168.74, 101.01, 73.04, 85.6, 99.71, 76.83, 267.6, 90.97, ...]
9	bmi	float64	1309	[39.8, 28.79, 31.1, 44.5, 27.73, 33.5, 22.8, 17.1, nan, 28.88, 27.3, 24.1, 20.2, 18.1, 36.0, 19.8, 27.0, 28.56, 32.3, 21.5, 46.1, 17.4, 41.0, 16.7, 31.53, 27.4, 28.93, 56.0, 45.2, 21.9, 26.23, 29.88, 24.24, 26.3, 38.9, 38.5, 41.2, 36.51, 23.95, 35.92, 22.7, 30.5, 30.9, 29.1, 34.14, 25.5, 17.7, 18.8, 31.34, 35.85, 26.9, 24.9, 34.7, 31.8, 25.09, 28.15, 25.3, 21.4, 24.2, 26.27, 30.1, 24.5, 18.4, 26.45, 27.2, 15.0, 28.6, 27.6, 19.0, 27.9, 20.4, 25.62, 35.36, 26.1, 34.8, ...]
10	smoking_status	object	4	[smokes, Unknown, never smoked, formerly smoked]
11	stroke	int64	2	[0, 1]

Ditemukan inkonsistensi penulisan pada variabel work_type dan smoking_status,
di mana terdapat perbedaan penggunaan huruf kapital dan huruf kecil.
sehingga perlu dilakukan standarisasi penulisan dengan menggunakan
format kapital di awal kata (Title Case).

PERMASALAHAN KUALITAS DATA

3

Memeriksa Data Kosong pada Dataset

```
id                      0  
gender                  0  
age                     0  
hypertension             0  
heart_disease            0  
ever_married              0  
work_type                 0  
Residence_type            0  
avg_glucose_level          0  
bmi                     402  
smoking_status              0  
stroke                   0  
dtype: int64
```

Terdapat data kosong pada BMI dan akan di Handling dengan cara inputasi nilai median

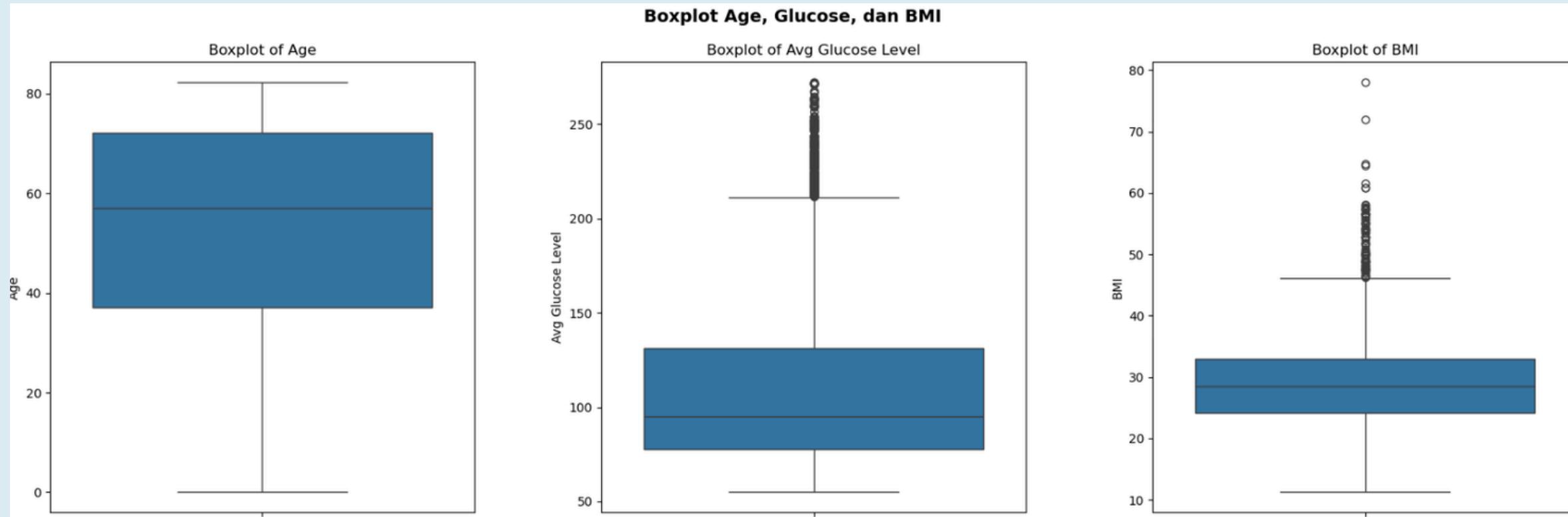
4

Memeriksa Data Duplikat

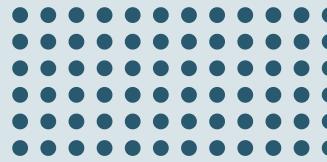
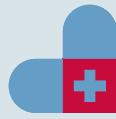
Jumlah data duplikat: 0

Tidak ditemukan data duplikat pada data ini.

PERMASALAHAN KUALITAS DATA



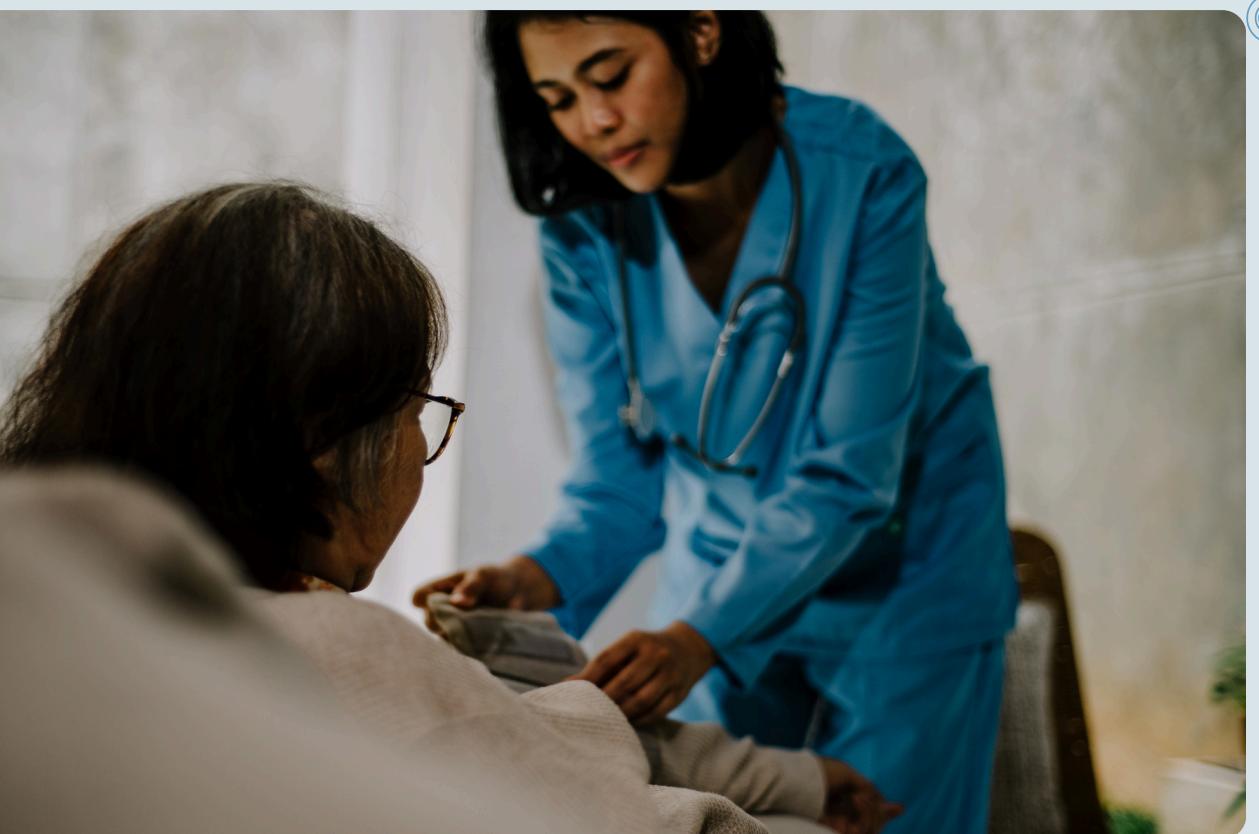
- Terdapat outlier pada variabel Glucose dan BMI berdasarkan metode IQR.
- Namun setelah dilakukan validasi terhadap rentang nilai medis, data tersebut masih masuk kategori realistik secara klinis.
- Oleh karena itu, tidak dilakukan penghapusan atau transformasi terhadap outlier tersebut.

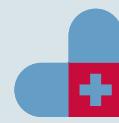


Transformasi dan Feature Engineering

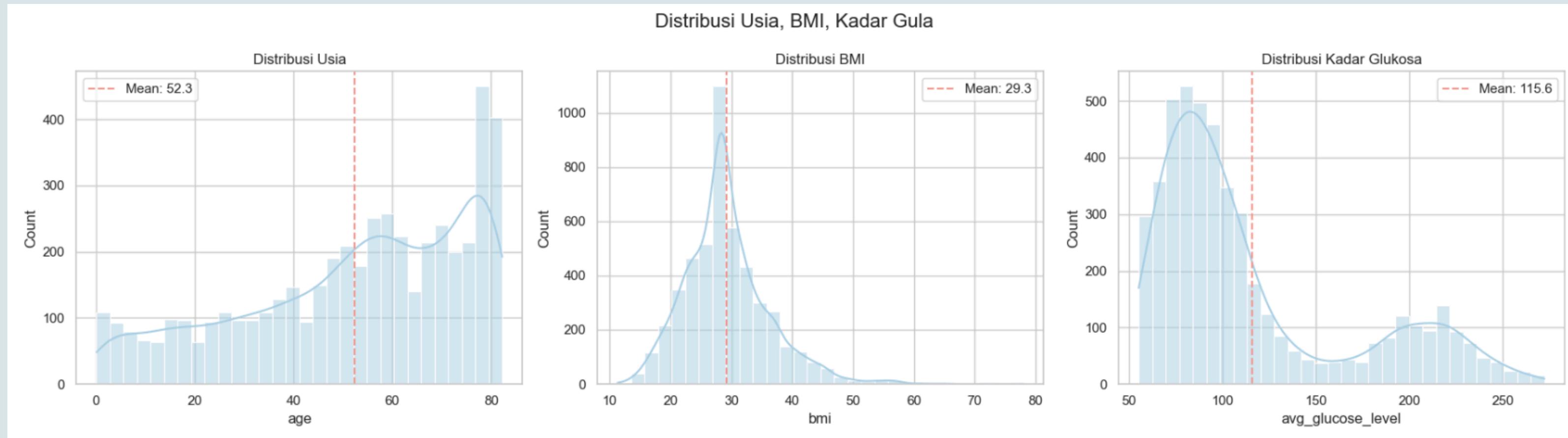
- Age → Age Category (Children, Young Adults, Middle Aged, Older Adults)
- BMI → BMI Category (Underweight, Normal, Overweight, Obese)
- Avg Glukosa → Glukosa Category (Low, Normal, Prediabetes, Diabetes)
- Membuat Stroke Label (Stroke / Tidak Stroke)
- Membuat variabel Kombinasi Risiko (Hipertensi + Jantung)
- Menghitung Stroke % sebagai metrik utama analisis

Tujuan: mempermudah segmentasi & interpretasi visualisasi.



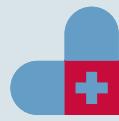


Exploratory Data Analysis (EDA)

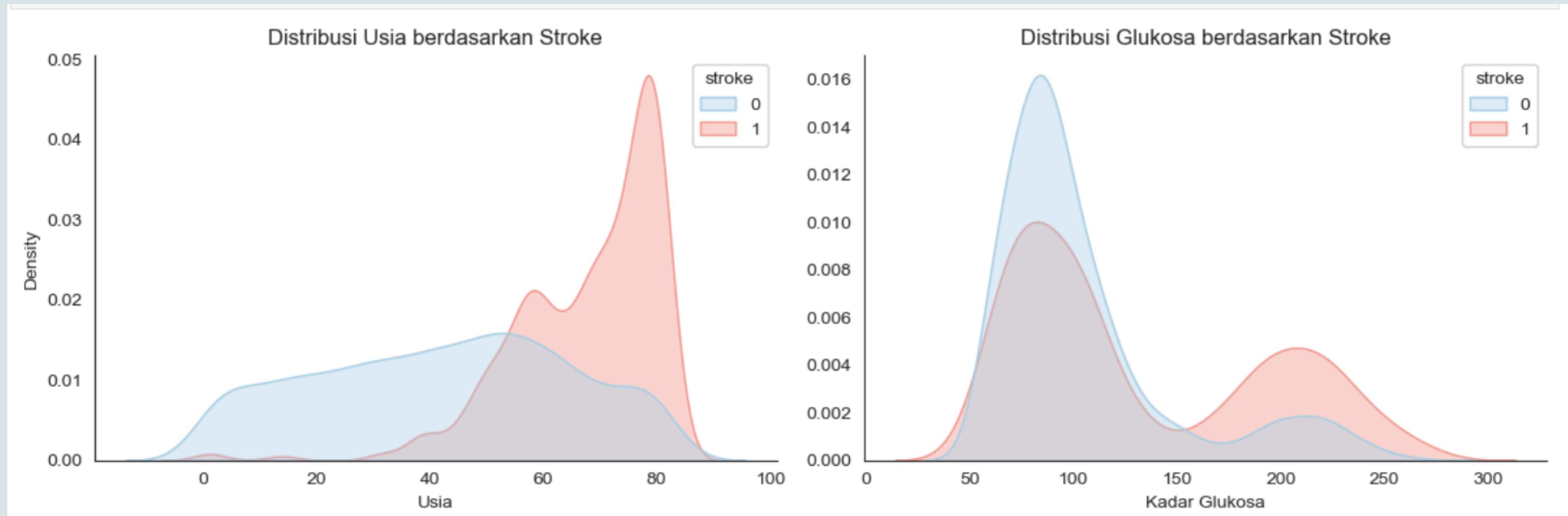


Berdasarkan analisis distribusi:

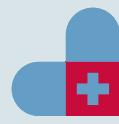
- Mayoritas pasien berada pada usia dewasa hingga lansia.
- Sebagian besar pasien memiliki BMI dalam kategori overweight.
- Terdapat kelompok pasien dengan kadar glukosa tinggi yang berpotensi meningkatkan risiko stroke.
- Temuan ini menjadi dasar untuk analisis hubungan antara faktor-faktor tersebut dengan kejadian stroke pada tahap selanjutnya.



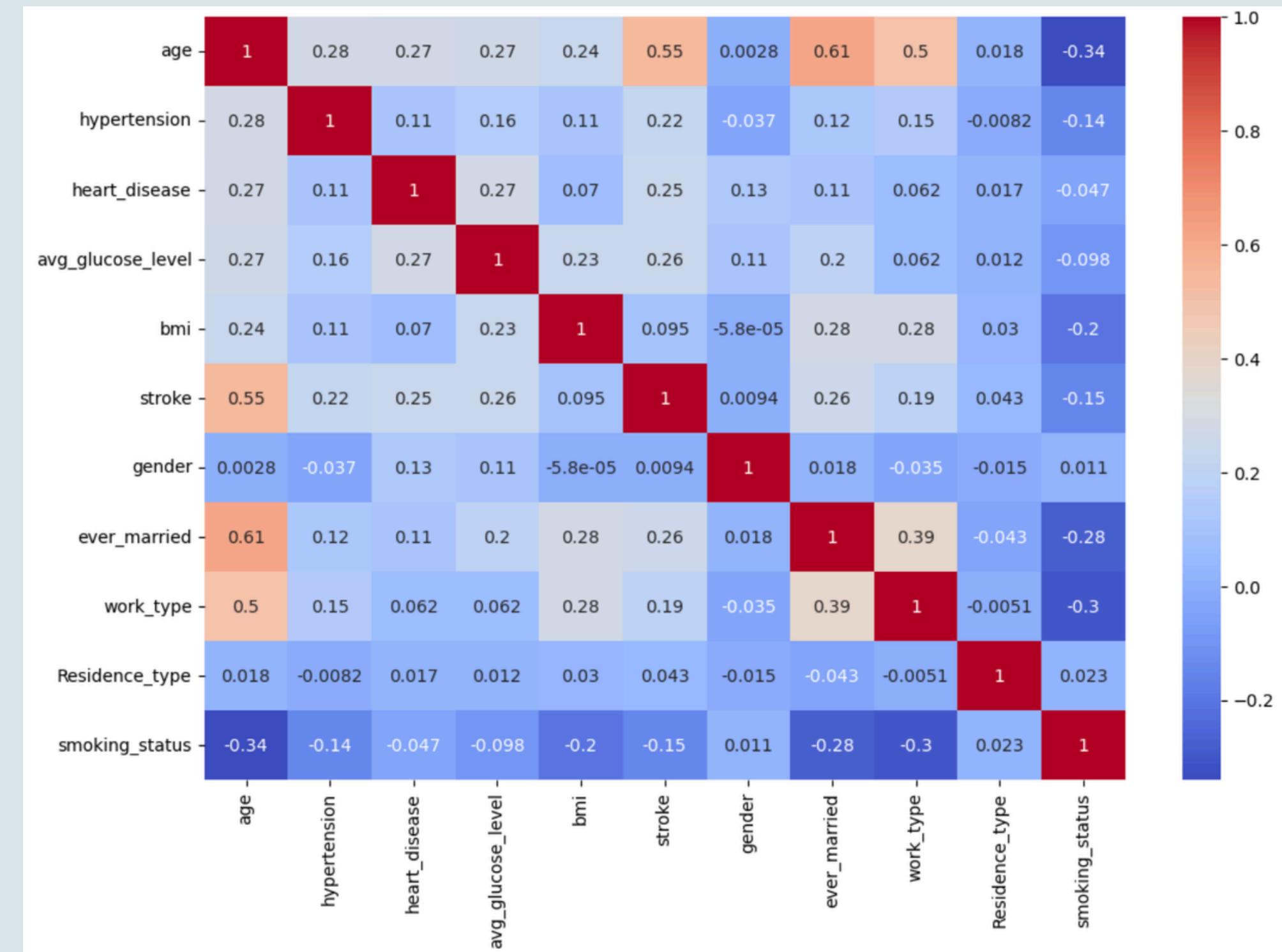
Exploratory Data Analysis (EDA)



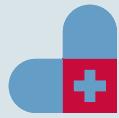
- Usia dan kadar glukosa menunjukkan perbedaan distribusi yang jelas antara pasien stroke dan non-stroke, mengindikasikan potensi faktor risiko penting.



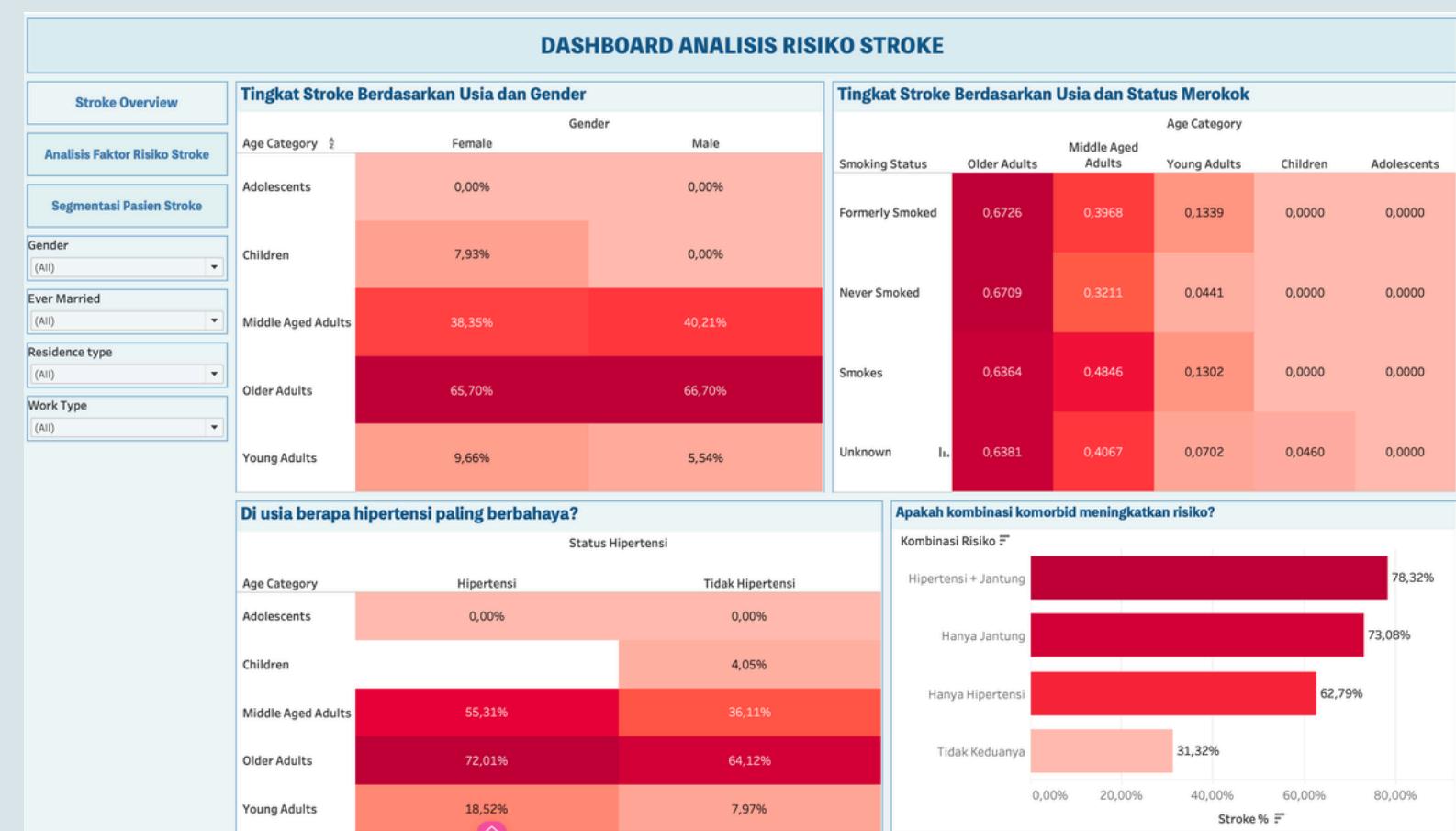
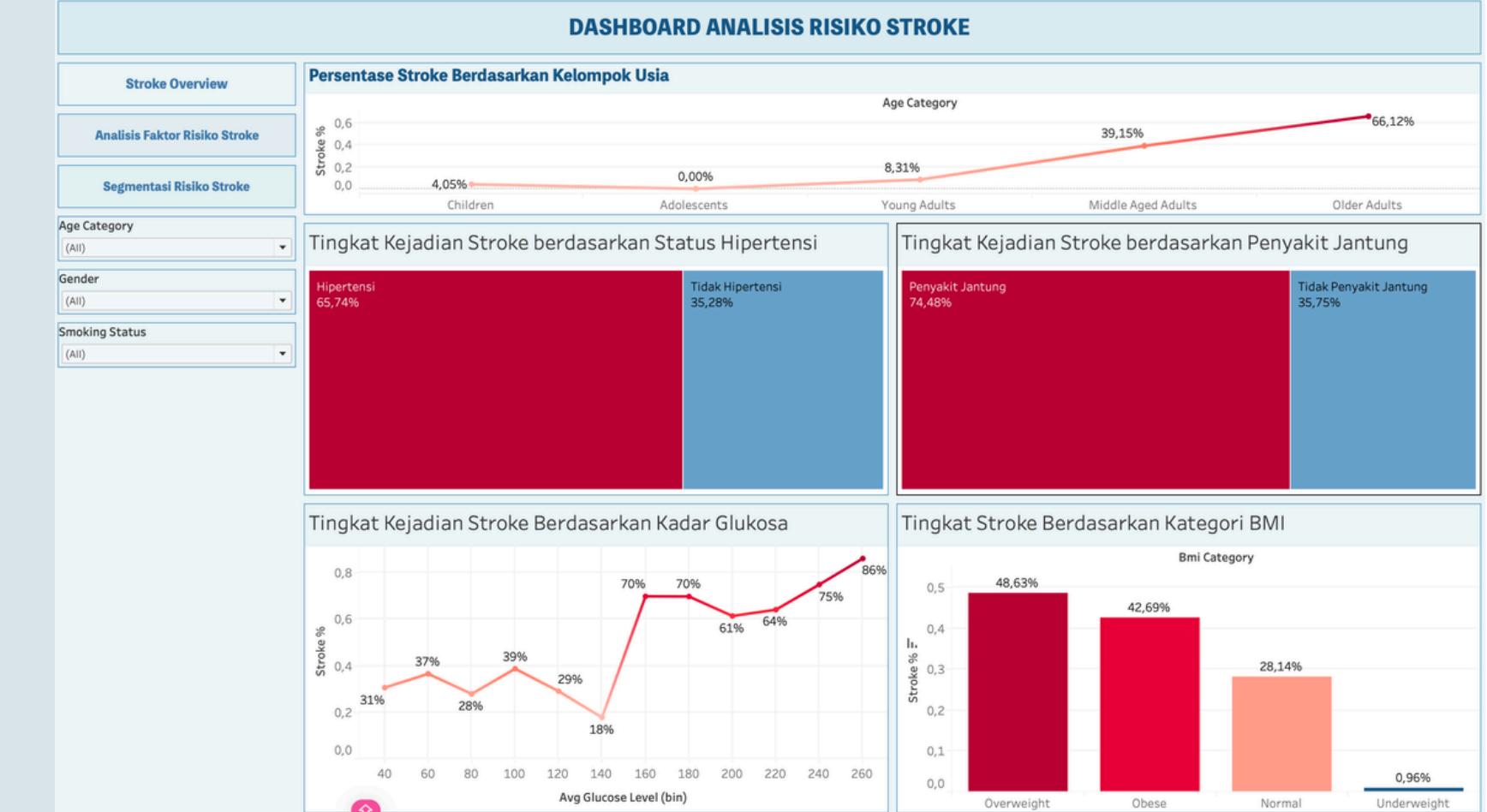
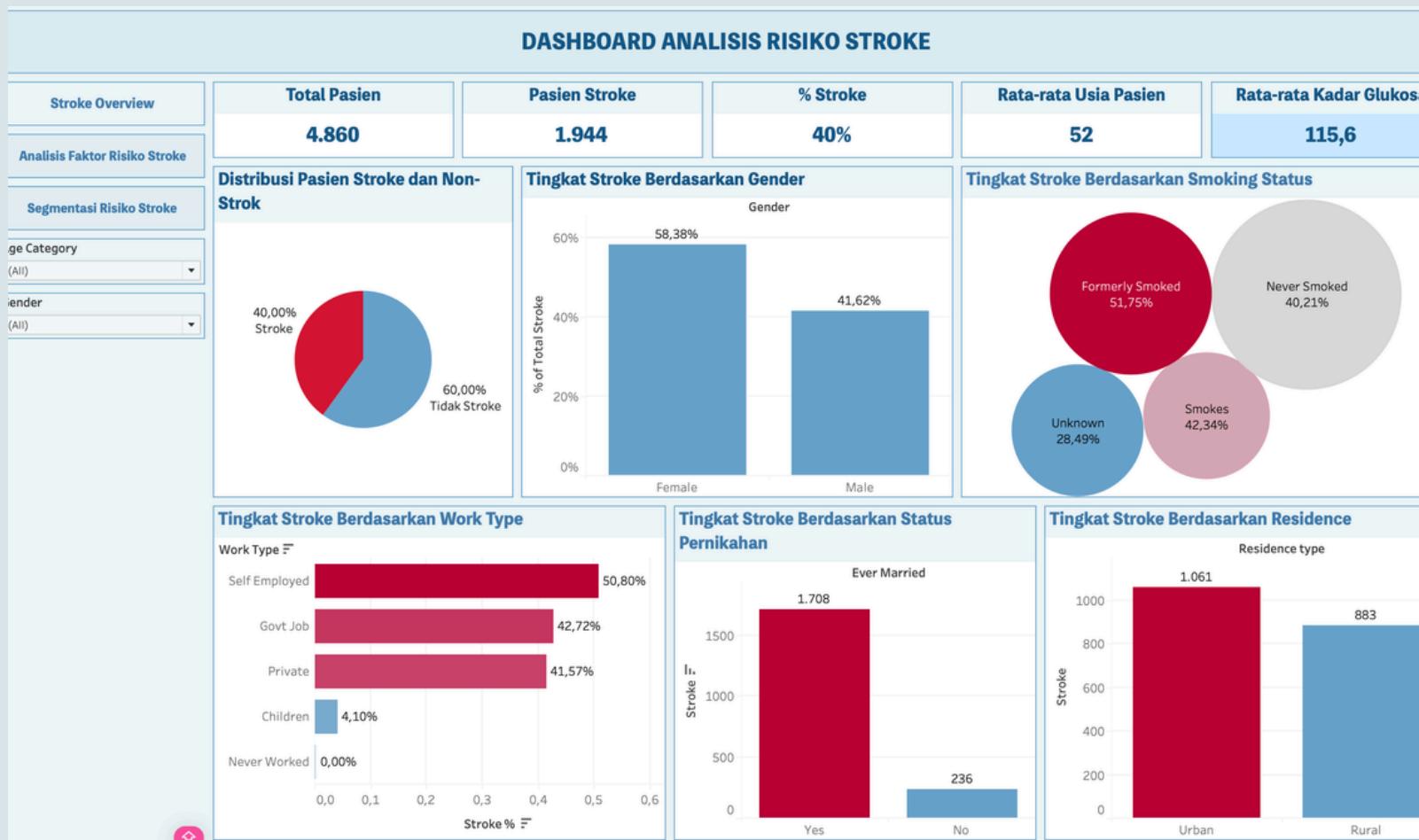
Exploratory Data Analysis (EDA)

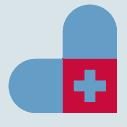


- Faktor usia, hipertensi, penyakit jantung, dan kadar glukosa merupakan variabel yang paling relevan untuk dianalisis lebih lanjut dalam risiko stroke.

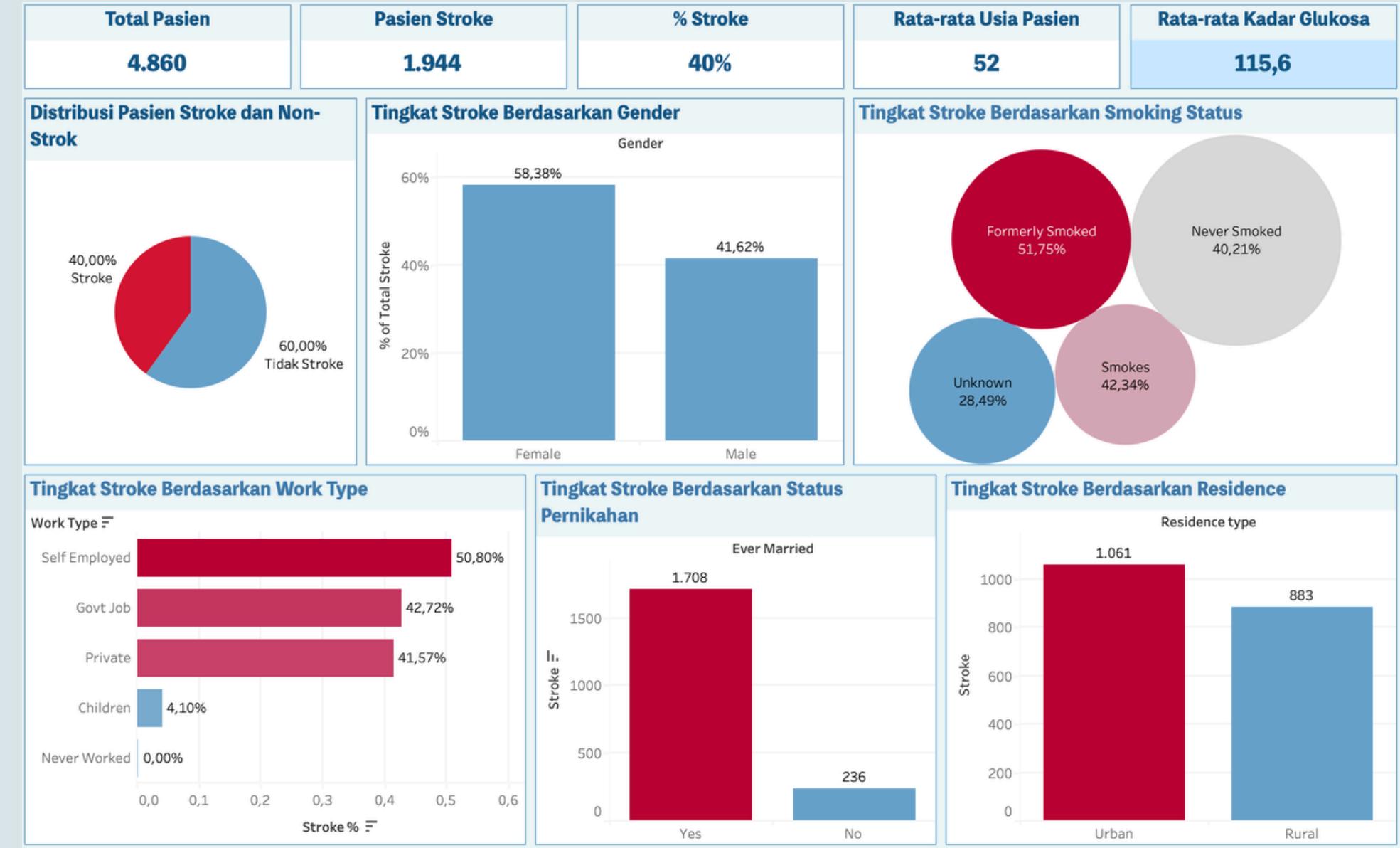


DASHBOARD ANALISIS RISIKO STROKE



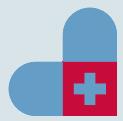


DASHBOARD - Stroke Overview

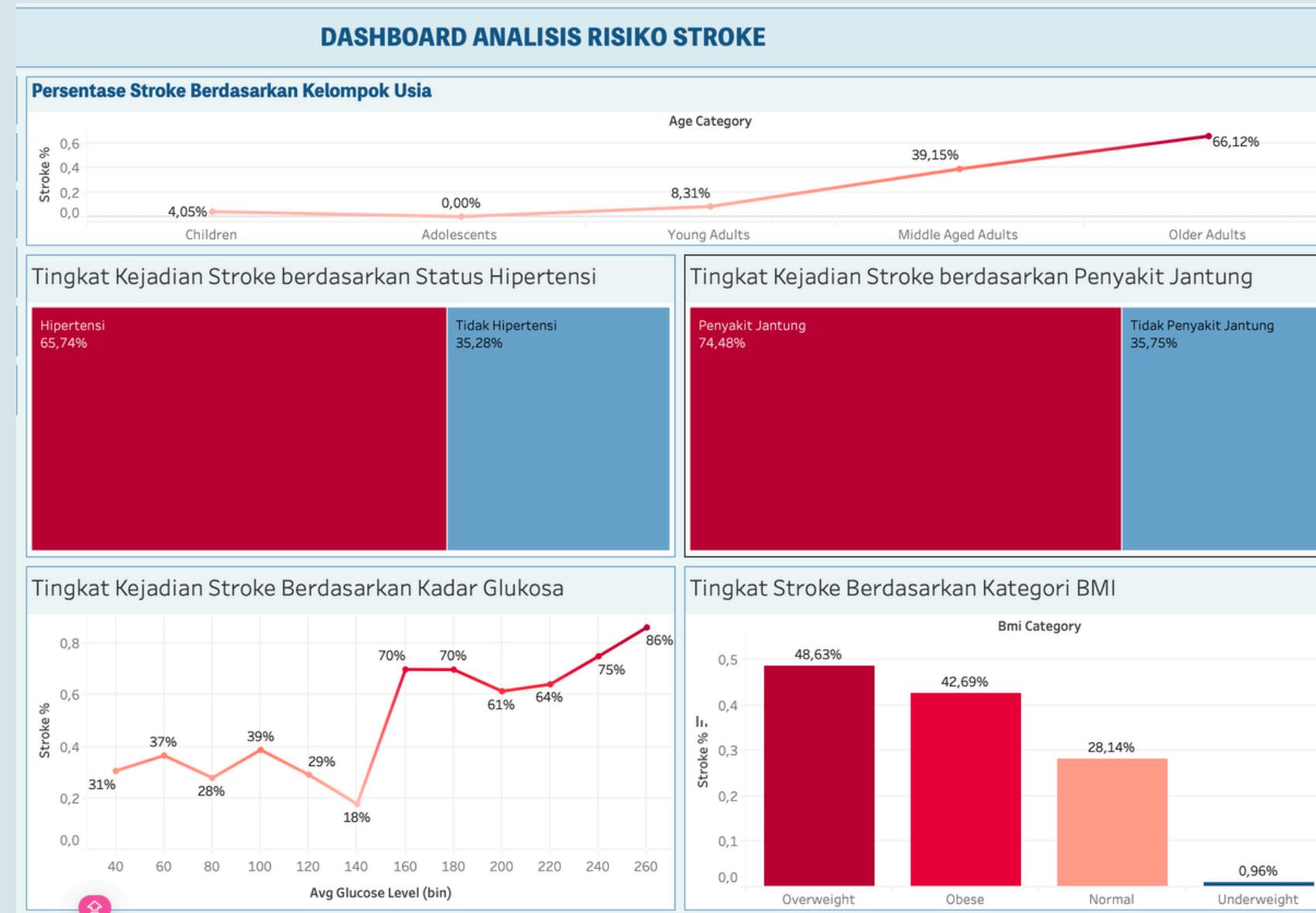


Dari total 4.860 pasien, 40% memiliki riwayat stroke.

- Mayoritas pasien berada pada usia paruh baya ke atas, dengan rata-rata usia 52 tahun
- Stroke lebih banyak terjadi pada kelompok usia lanjut
- Persentase stroke meningkat signifikan pada kelompok Middle Aged dan Older Adults
- Perempuan sedikit lebih dominan dibanding laki-laki
- Pasien yang pernah menikah menunjukkan jumlah stroke lebih tinggi
- Wilayah urban memiliki jumlah kasus lebih banyak dibanding rural

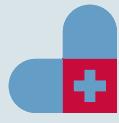


DASHBOARD - Analisis Faktor Risiko Stroke

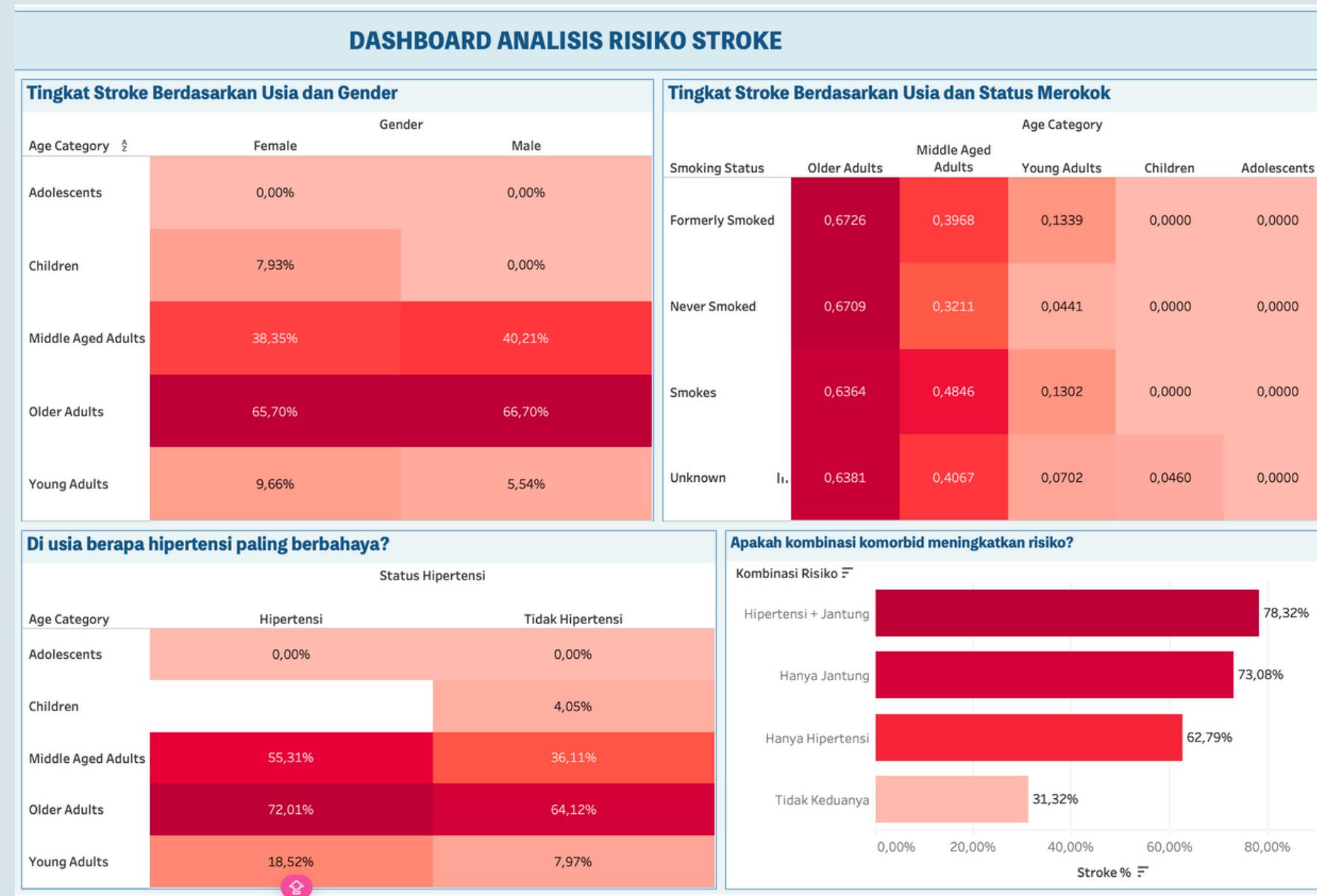


- Pasien dengan hipertensi memiliki tingkat stroke jauh lebih tinggi dibanding yang tidak.
- Penyakit Jantung merupakan faktor paling dominan dengan persentase tertinggi.
- Risiko stroke meningkat seiring meningkatnya kadar glukosa.
- BMI: Kelompok overweight dan obese menunjukkan risiko lebih tinggi dibanding normal.

Hasil analisis menunjukkan bahwa hipertensi, penyakit jantung, kadar glukosa tinggi, dan obesitas merupakan faktor medis utama yang berkontribusi signifikan terhadap peningkatan risiko stroke.

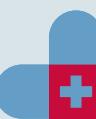


DASHBOARD – SEGMENTASI RISIKO STROKE



- Risiko tertinggi ditemukan pada usia lanjut dengan hipertensi
- Kombinasi hipertensi + penyakit jantung menunjukkan risiko sangat tinggi ($\approx 80\%$)
- Usia memiliki korelasi tertinggi terhadap kejadian stroke dibanding variabel lain
- Risiko meningkat signifikan pada pasien dengan komorbiditas ganda





Insight

1. Risiko stroke meningkat signifikan pada usia menengah hingga lanjut.
2. Penyakit jantung dan hipertensi merupakan faktor klinis utama yang meningkatkan risiko stroke.
3. Kombinasi hipertensi dan penyakit jantung menunjukkan risiko tertinggi.
4. Kadar glukosa tinggi berkorelasi dengan peningkatan probabilitas stroke.
5. BMI overweight/obese dan riwayat merokok memperkuat risiko, meskipun tidak sekuat faktor usia dan komorbid.
6. Kasus stroke lebih banyak ditemukan pada populasi urban dibanding rural.

Rekomendasi

1. Fokus pemeriksaan rutin untuk pasien usia di atas 50 tahun
2. Pantau lebih ketat pasien yang memiliki lebih dari satu penyakit (misalnya hipertensi dan penyakit jantung)
3. Dorong kontrol gula darah secara rutin untuk mencegah risiko stroke
4. Tingkatkan edukasi tentang pola hidup sehat (berhenti merokok, jaga berat badan, aktif bergerak)
5. Perkuat program pencegahan stroke di wilayah perkotaan



Thank You



amandamerzha@gmail.com



Jakarta, Indonesia

