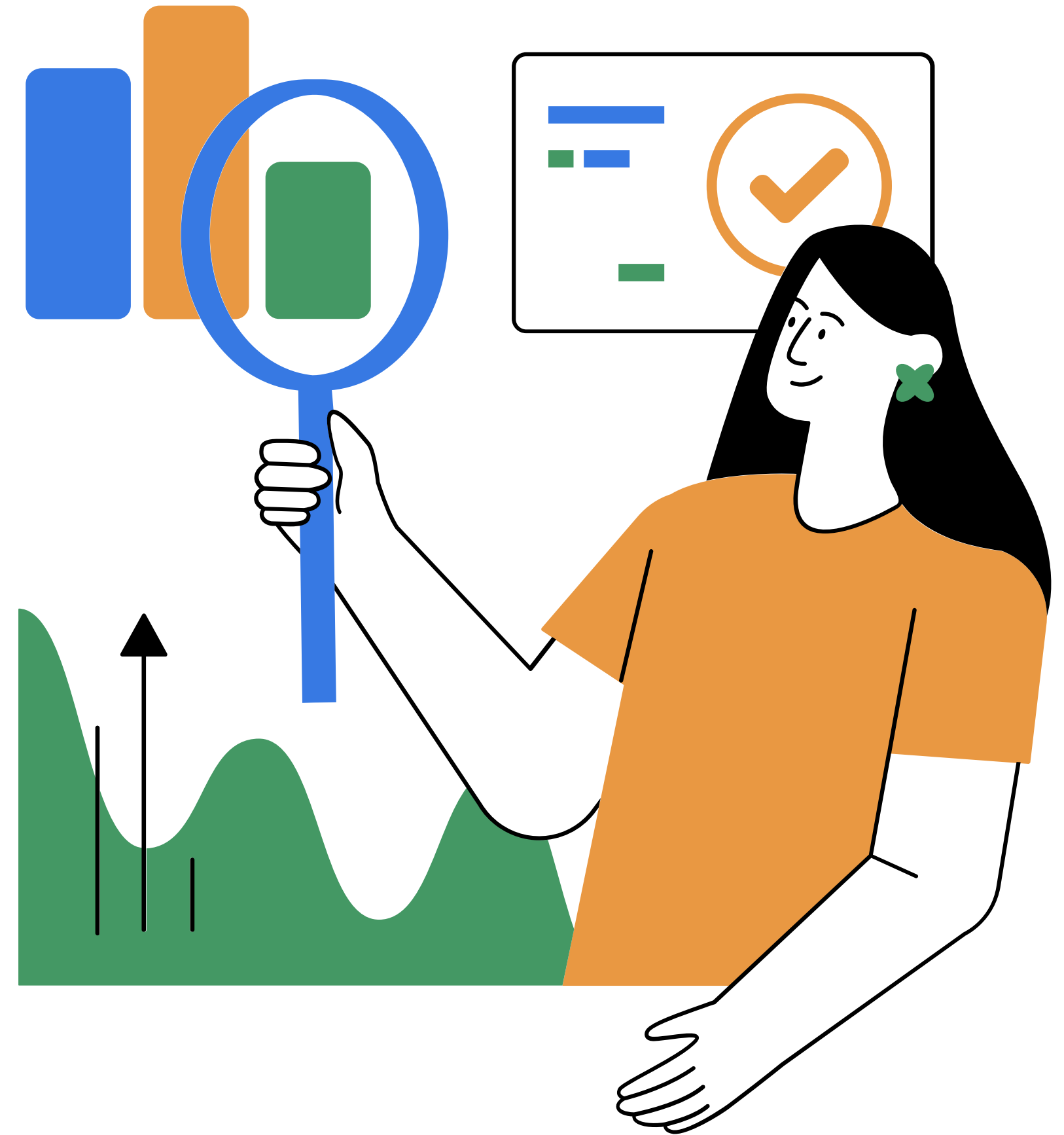


Digital Skill Fair 41.0

Muzhaffar Rafi
Faculty Of Data: Data Science Analyst



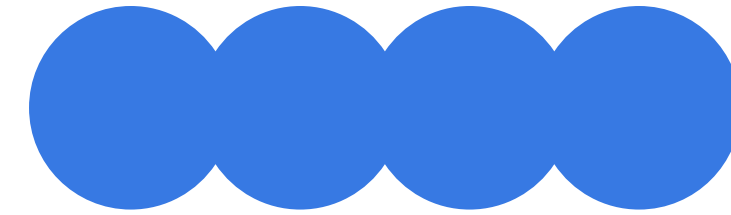


ABOUT ME

"Saya adalah fresh graduate dari jurusan Sistem Informasi yang memiliki semangat tinggi untuk terus belajar dan mengembangkan pengetahuan di bidang data. Saya telah mengikuti event Digital Skill Fair 41.0 sebagai upaya untuk memperkuat pemahaman saya terhadap dasar-dasar Data Science serta mulai membangun portofolio di bidang tersebut."



DIBIMBING.ID DIGITAL SKILL FAIR 41.0: FACULTY OF DATA



Bootcamp singkat 5 hari untuk memperkuat keahlian dalam analisis dan pengolahan data.



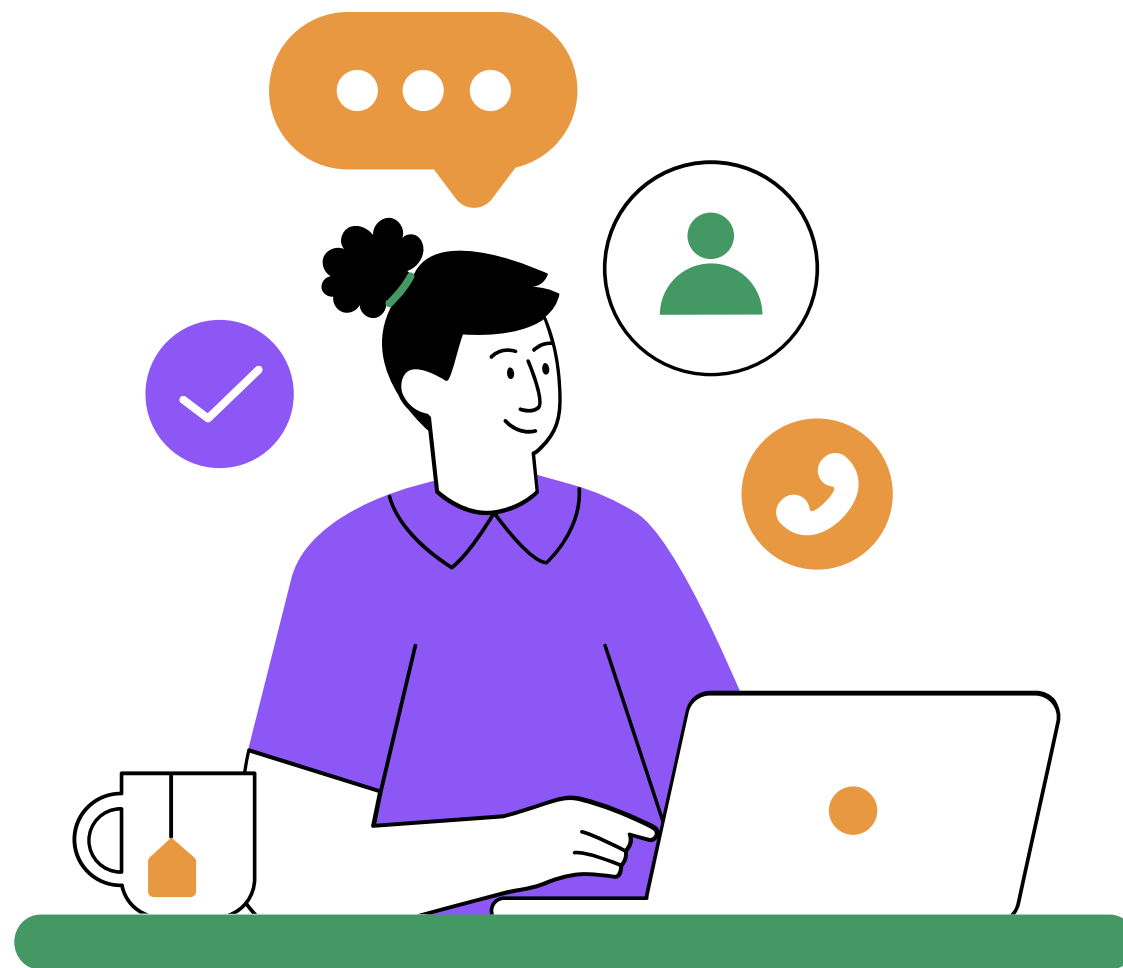
Difokuskan untuk mengasah keterampilan sebagai Data Science Analyst



Mempelajari dasar-dasar Python, praktik exploratory data analysis (EDA), serta mendapatkan pembekalan karier melalui sesi career preparation



MY 5-DAY LEARNING JOURNEY



Day 1-2: Building the foundation

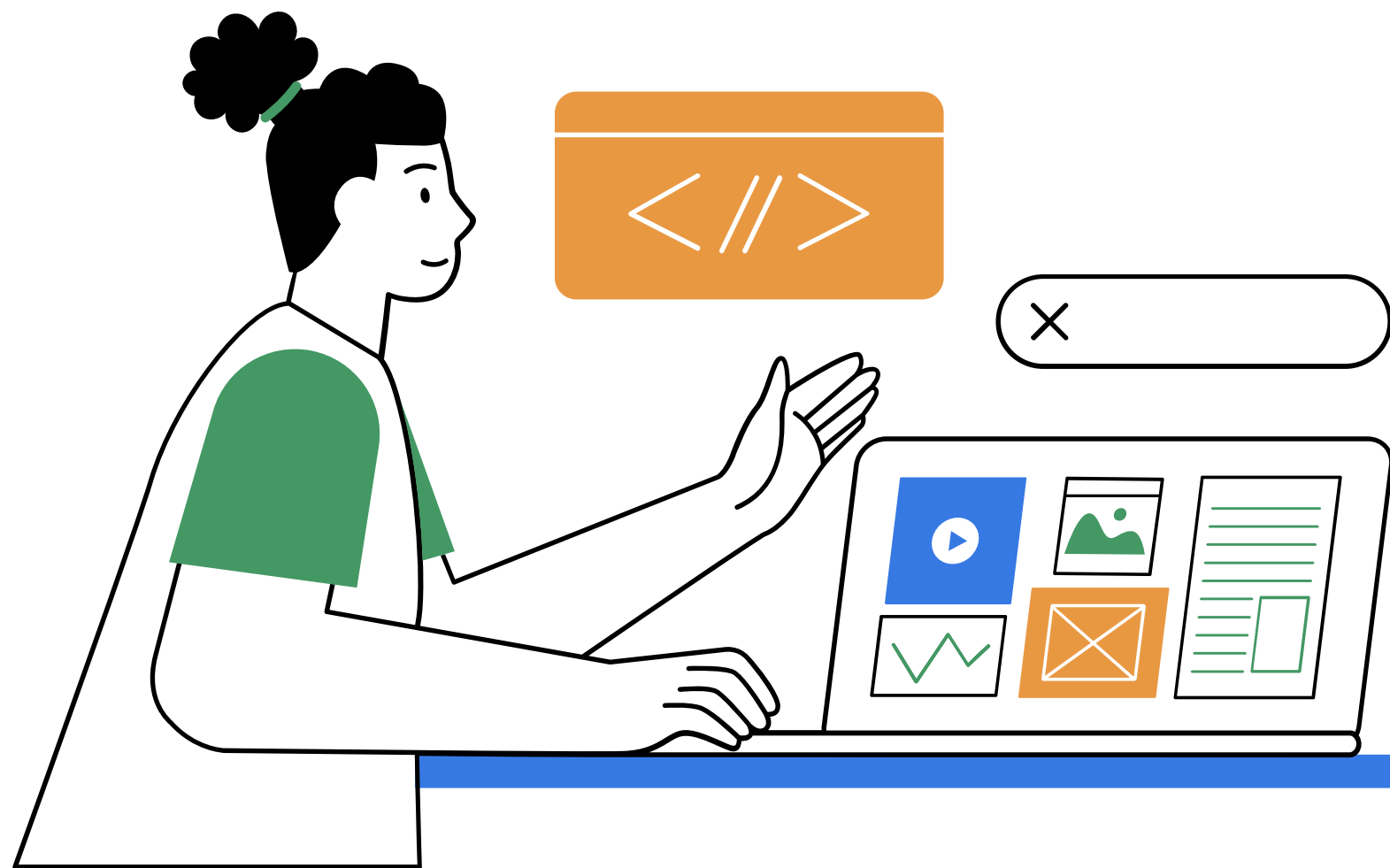
- Mendapatkan pemahaman umum tentang industri data serta pengenalan dasar bahasa pemrograman Python.
- Tugas pertama yang dikerjakan adalah membuat program sederhana untuk input data mahasiswa menggunakan logika percabangan if-elif-else.

Day 3: Diving Into The Core Skill - EDA

- Mendalami tahapan Exploratory Data Analysis (EDA), dimulai dari memahami data mentah hingga menghasilkan insight.
- Fokus pada data cleansing: menghapus duplikasi dan menangani missing values

Day 4-5: Career Preparation

- Memperoleh pemahaman tentang cara mengoptimalkan profil LinkedIn dan GitHub agar lebih profesional dan menarik bagi recruiter.
- Belajar menyusun CV yang efektif, serta mendapatkan kesempatan untuk mereview portofolio bersama mentor.

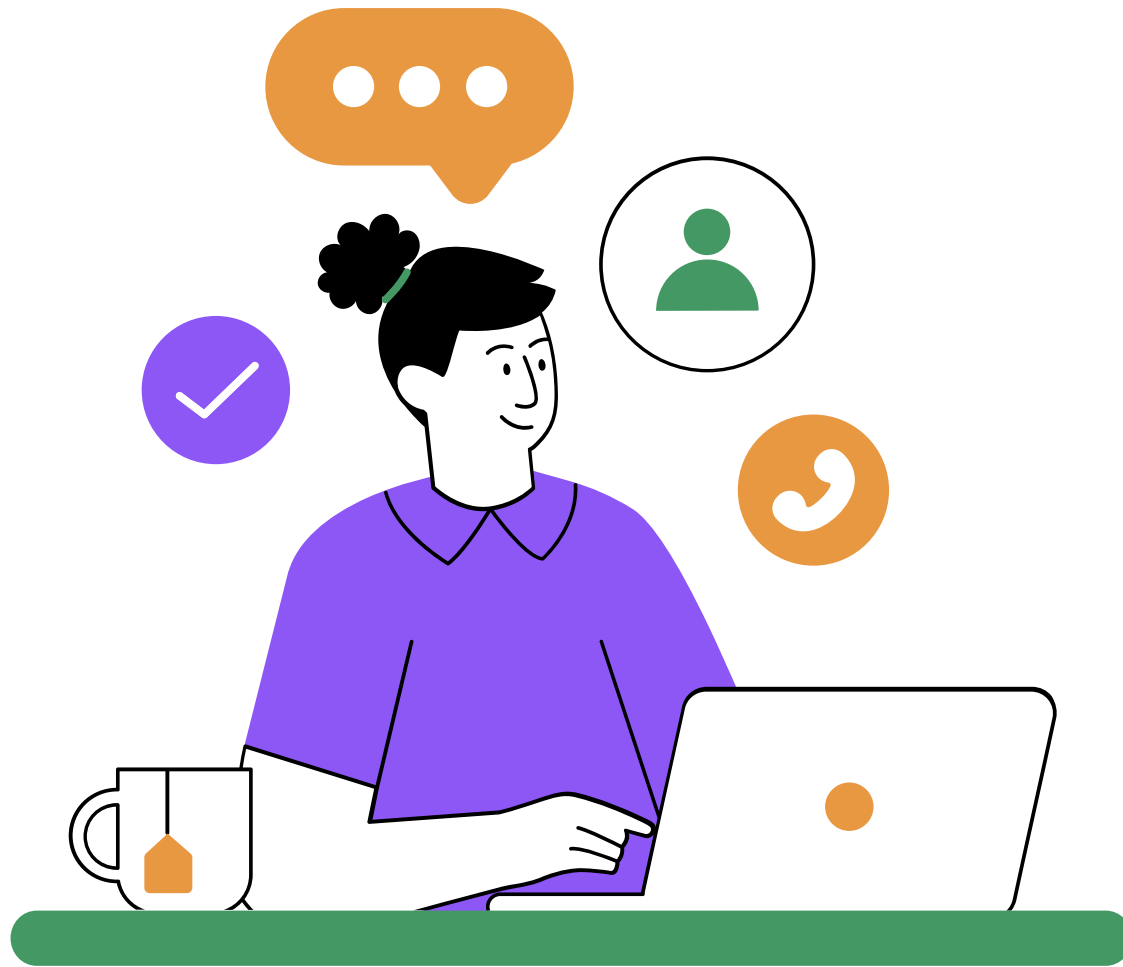


Project Overview:
Exploratory Data Analysis
to Hypertension Risk
Prediction

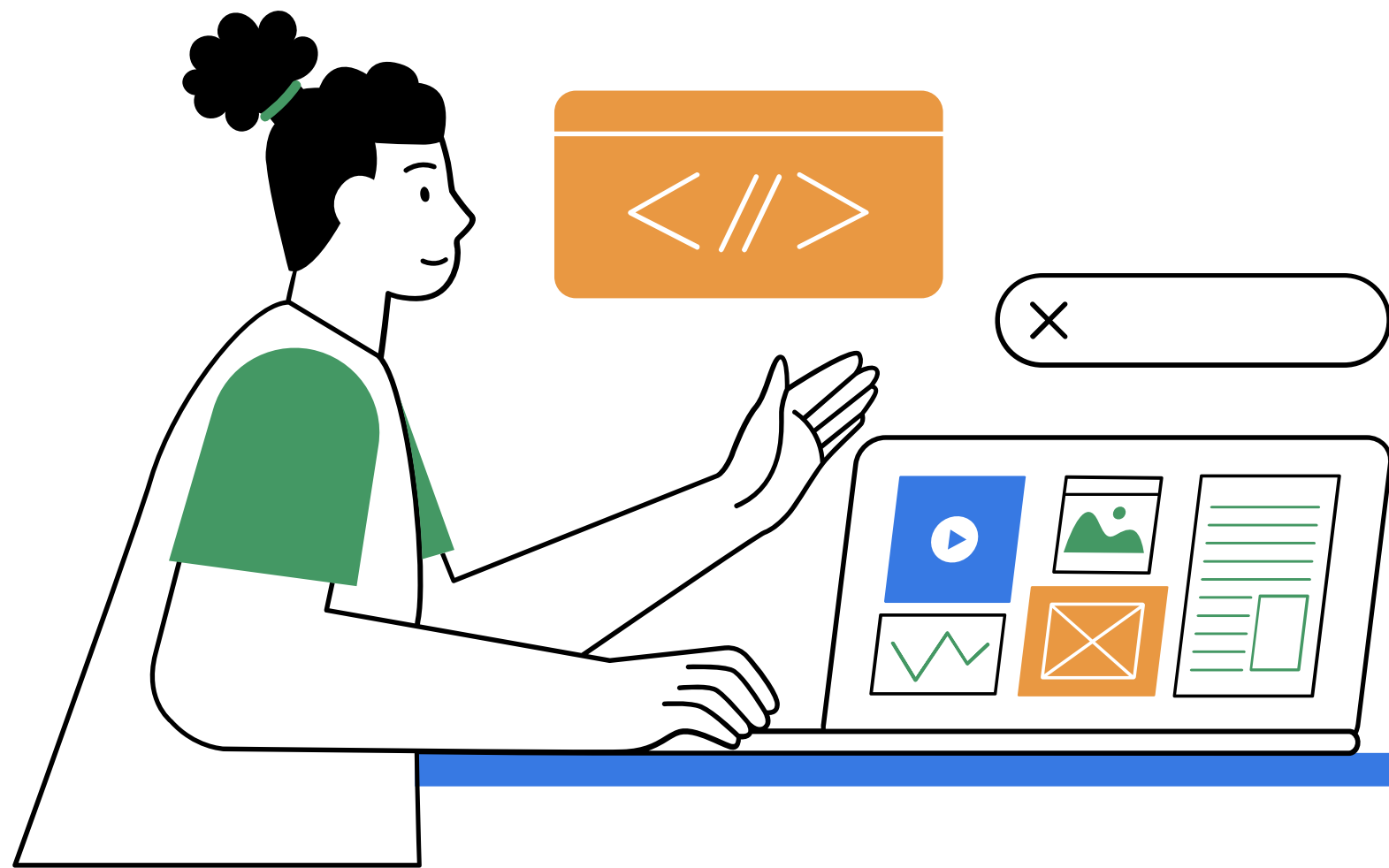
Tools &
Technologies

BUSINESS PROBLEM

Bagaimana mengidentifikasi secara akurat faktor-faktor risiko utama yang memengaruhi hipertensi? Dataset yang digunakan yaitu dataset Hypertension Risk Prediction Dataset dari Kaggle.



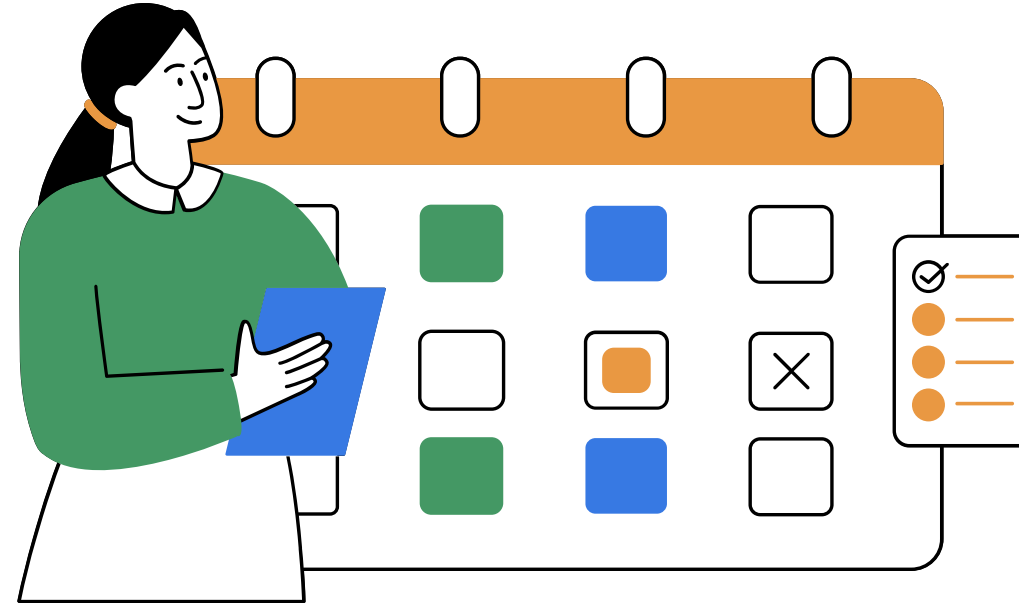
1. Membersihkan dan mempersiapkan data pasien hipertensi untuk analisis prediksi
2. Menggunakan pendekatan deskriptif untuk mengeksplorasi pola dan temuan yang relevan dari data
3. Menyampaikan hasil analisis utama melalui representasi data yang visual



THE PROCESS: DATA CLEANSING & DATA PREPROCESSING

Tools & Technologies

Handling Missing Values



Masalah

Ditemukan sejumlah data yang hilang pada kolom Medication

Solusi

Baris data yang kosong di isi menggunakan nilai modus untuk kategorikal sedangkan di isi menggunakan nilai median untuk numerik

Gambar

```
[ ] # Cek missing values
print("Missing values sebelum pembersihan:")
print(df.isnull().sum())

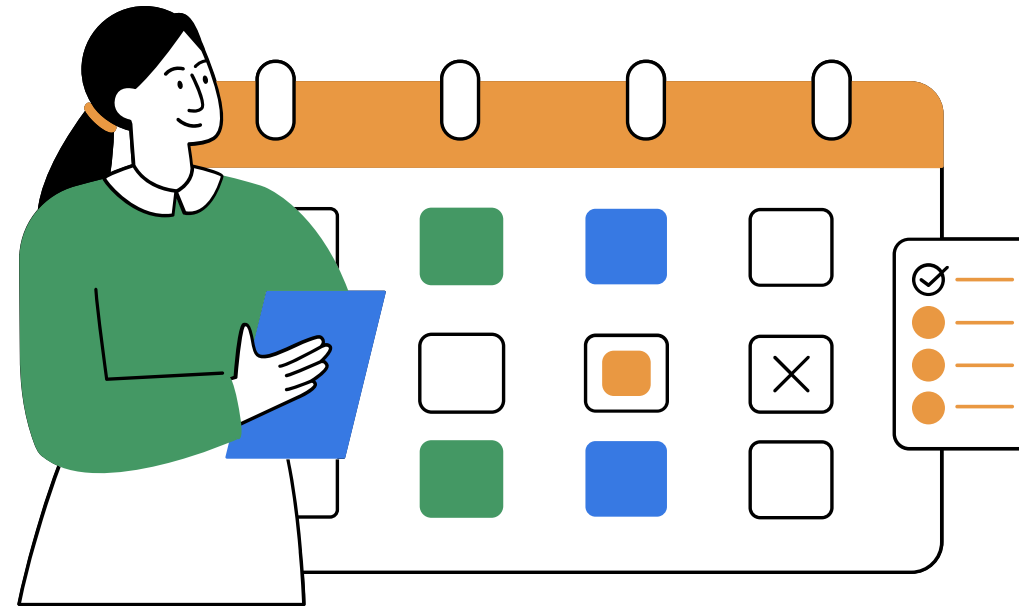
# Menangani missing values
for column in df.columns:
    if df[column].dtype == 'object':
        # Untuk kolom kategorikal, isi dengan modus
        df[column].fillna(df[column].mode()[0], inplace=True)
    else:
        # Untuk kolom numerik, isi dengan median
        df[column].fillna(df[column].median(), inplace=True)

print("\nMissing values setelah pembersihan:")
print(df.isnull().sum())
```

```
Missing values sebelum pembersihan:
Age          0
Salt_Intake  0
Stress_Score 0
BP_History   0
Sleep_Duration 0
BMI          0
Medication   799
Family_History 0
Exercise_Level 0
Smoking_Status 0
Has_Hypertension 0
dtype: int64
```

```
Missing values setelah pembersihan:
Age          0
Salt_Intake  0
Stress_Score 0
BP_History   0
Sleep_Duration 0
BMI          0
Medication   0
Family_History 0
Exercise_Level 0
Smoking_Status 0
Has_Hypertension 0
dtype: int64
```


Handling Duplicates



Masalah

Pada dataset ini tidak memiliki kolom yang memiliki duplicate

Gambar

```
[ ] # Cek duplikat
duplicate_count = df.duplicated().sum()
print(f"Jumlah baris duplikat: {duplicate_count}")

# Hapus duplikat
df.drop_duplicates(inplace=True)

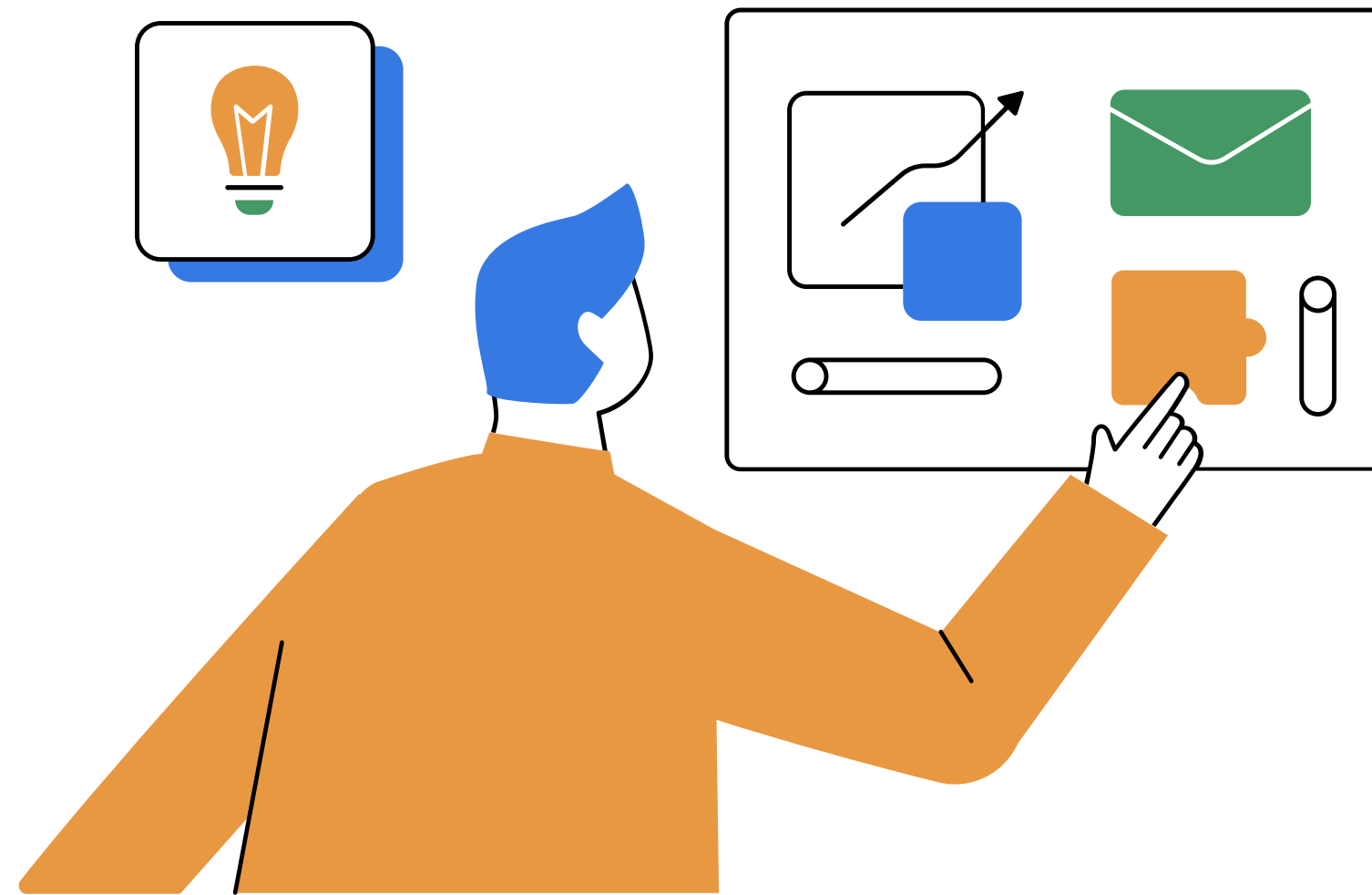
print(f"Jumlah baris setelah menghapus duplikat: {len(df)}")
```



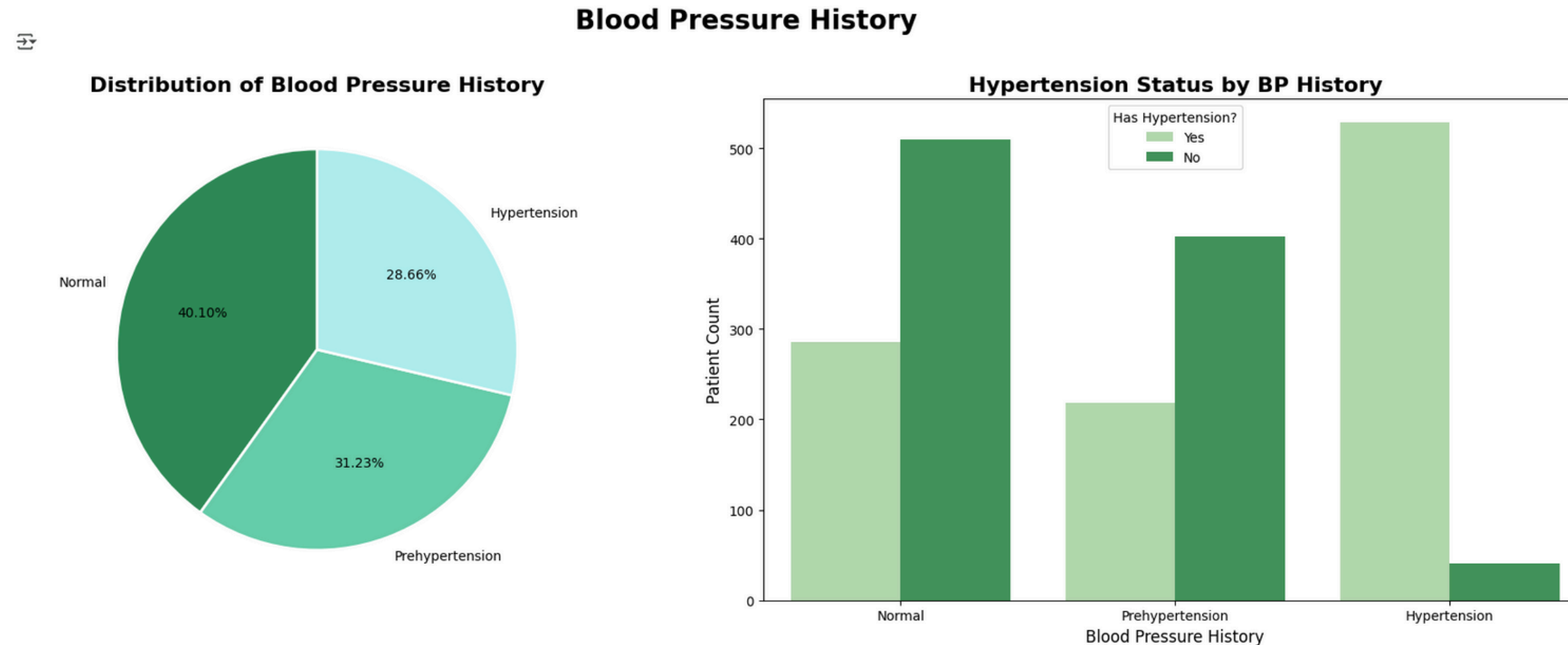
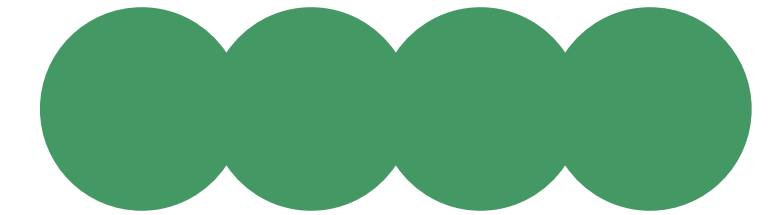
Jumlah baris duplikat: 0

Jumlah baris setelah menghapus duplikat: 1985

Exploratory of Data Analysis

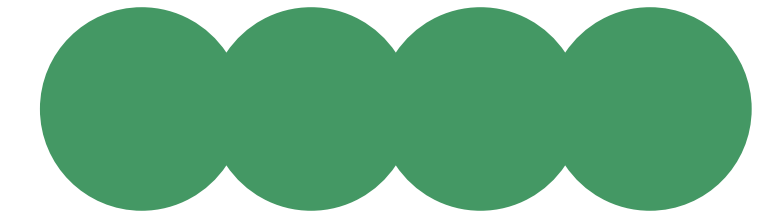


Exploratory of Data Analysis

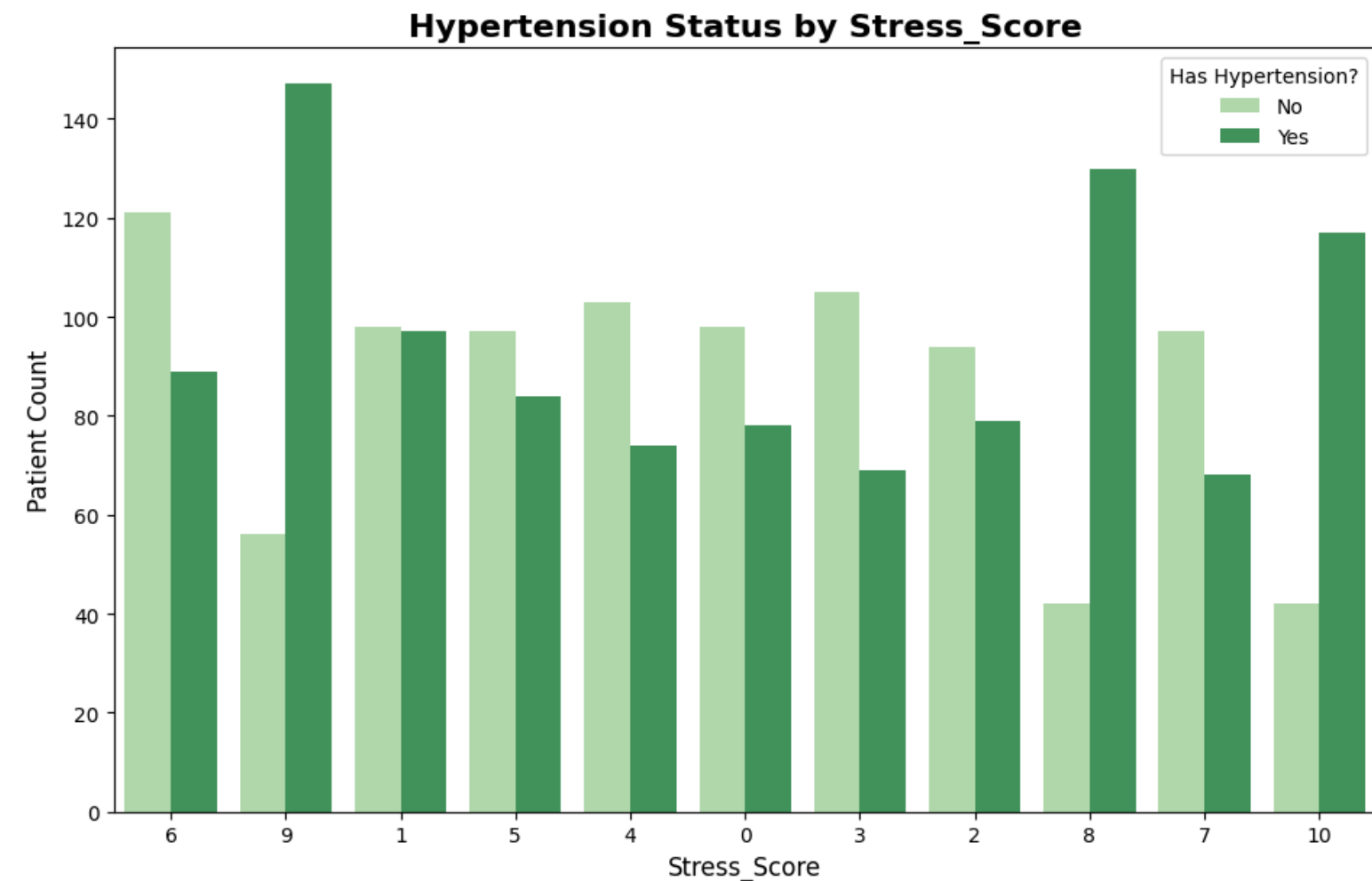
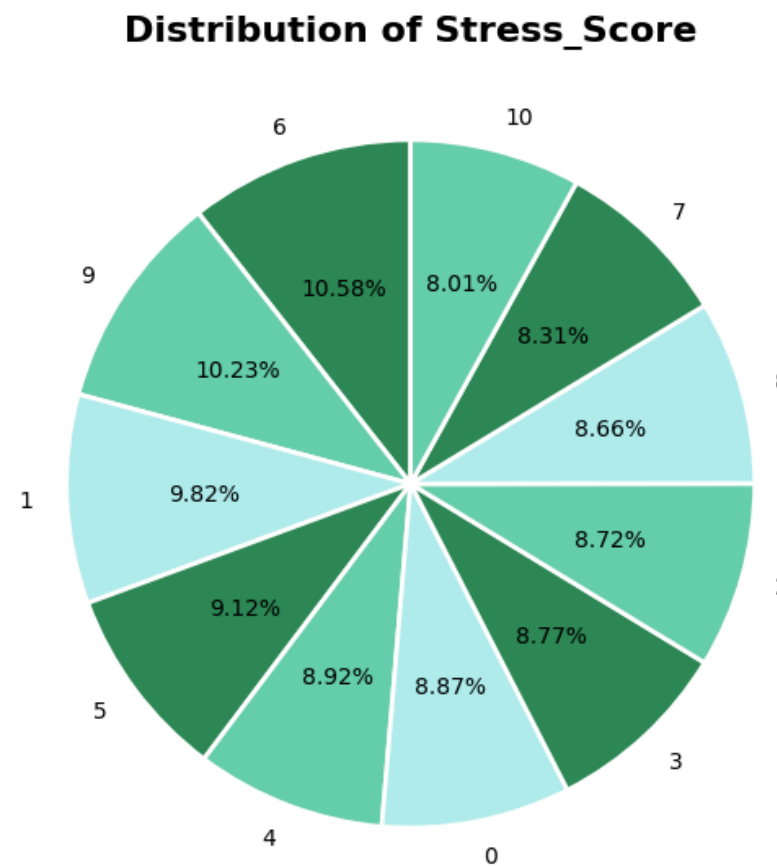


Dari diagram pie, terlihat bahwa sebagian besar pasien memiliki tekanan darah normal (40,10%), disusul oleh kategori pra-hipertensi (31,23%) dan hipertensi (28,66%). Namun, diagram batang menunjukkan bahwa sebagian besar pasien dengan riwayat hipertensi juga terdiagnosis hipertensi (sekitar 530 dari total 570 pasien), sedangkan mayoritas pasien dengan tekanan darah normal dan pra-hipertensi tidak mengalami hipertensi.

Exploratory of Data Analysis

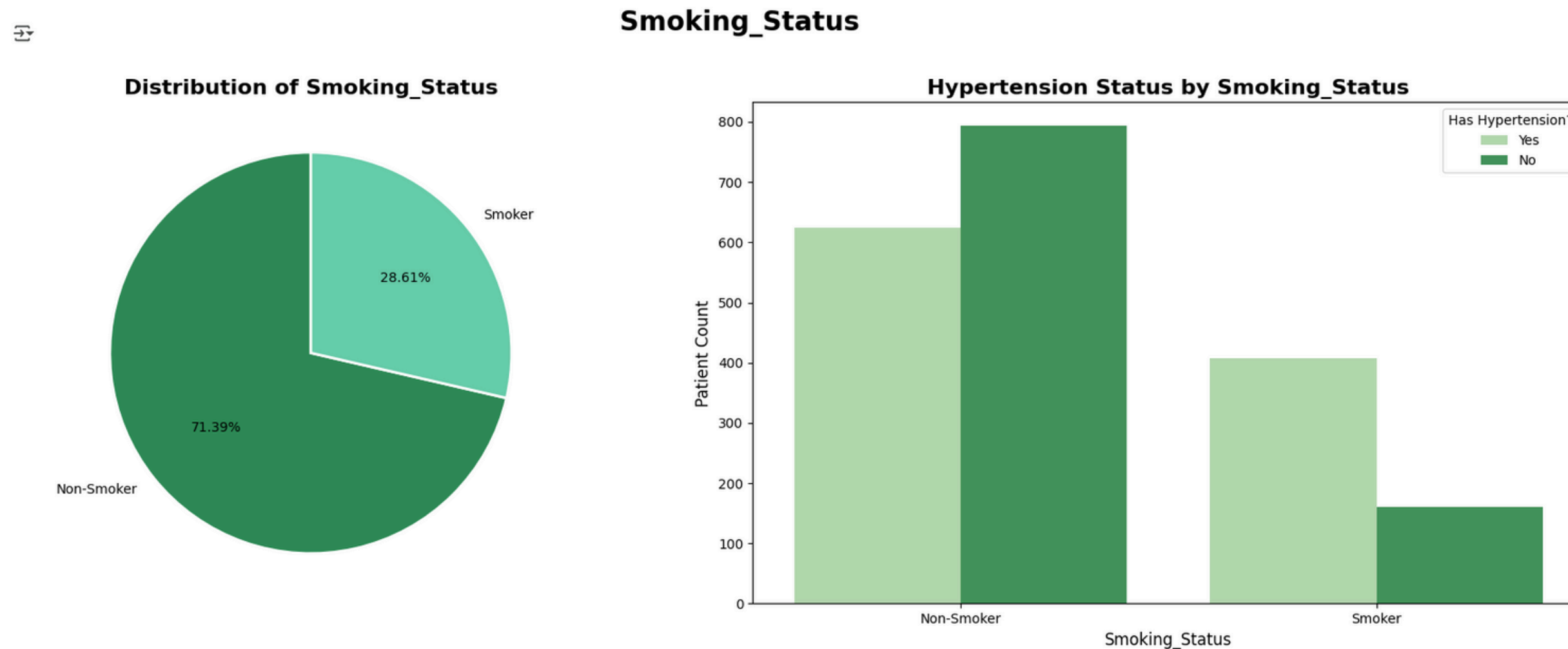
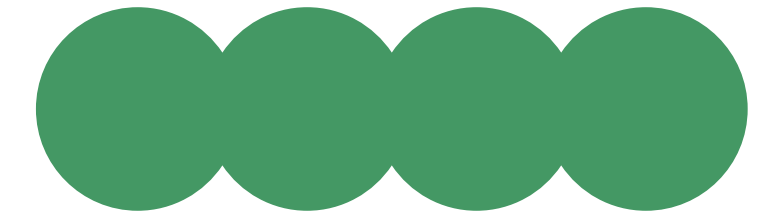


Stress_Score



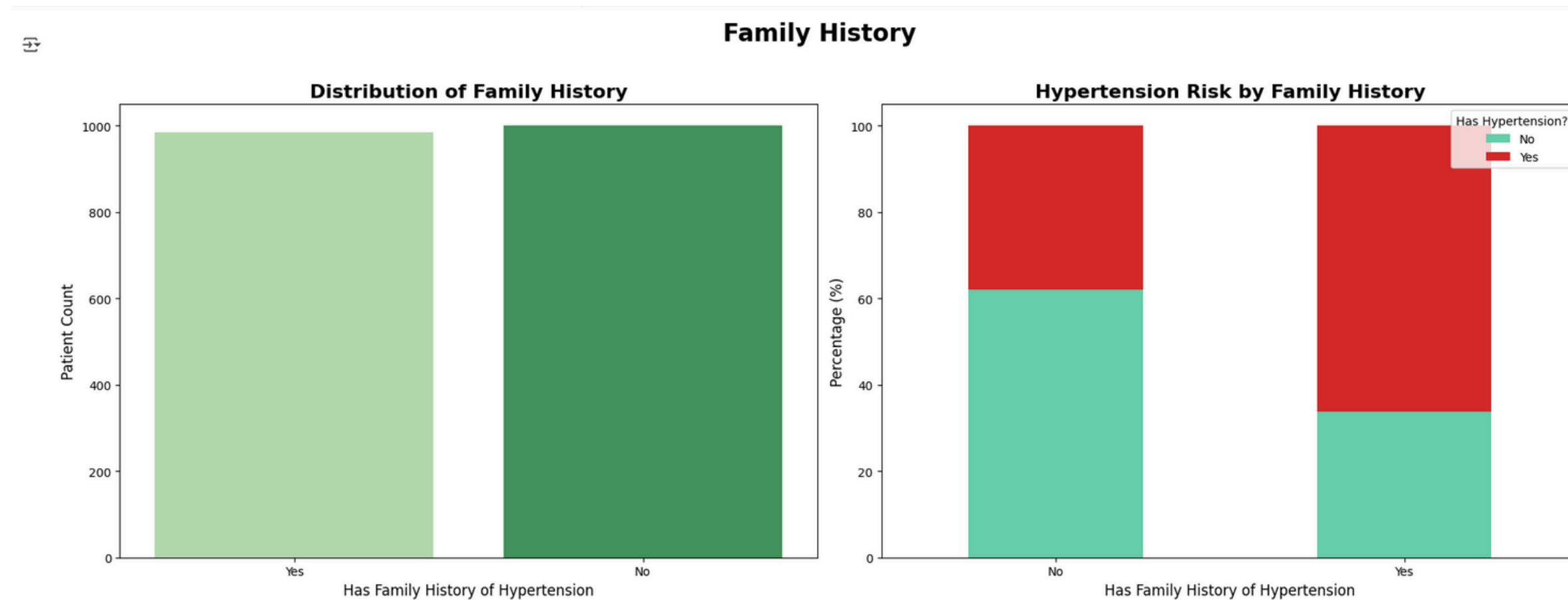
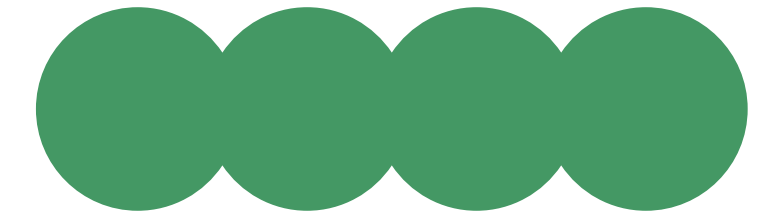
Gambar menunjukkan bahwa skor stres tersebar merata, namun pasien dengan skor stres tinggi (8–10) lebih banyak yang menderita hipertensi. Sebaliknya, pada skor rendah hingga sedang, lebih banyak pasien yang tidak hipertensi. Hal ini menunjukkan bahwa semakin tinggi tingkat stres, semakin besar risiko hipertensi.

Exploratory of Data Analysis



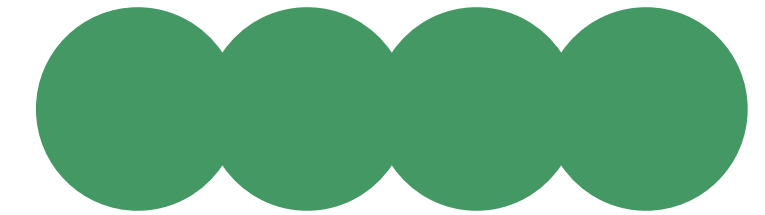
Gambar menunjukkan bahwa sebagian besar pasien adalah non-perokok (71,39%), namun proporsi hipertensi lebih tinggi pada perokok. Di antara perokok, mayoritas menderita hipertensi, sedangkan non-perokok lebih banyak yang tidak hipertensi. Hal ini menunjukkan adanya hubungan kuat antara merokok dan risiko hipertensi.

Exploratory of Data Analysis

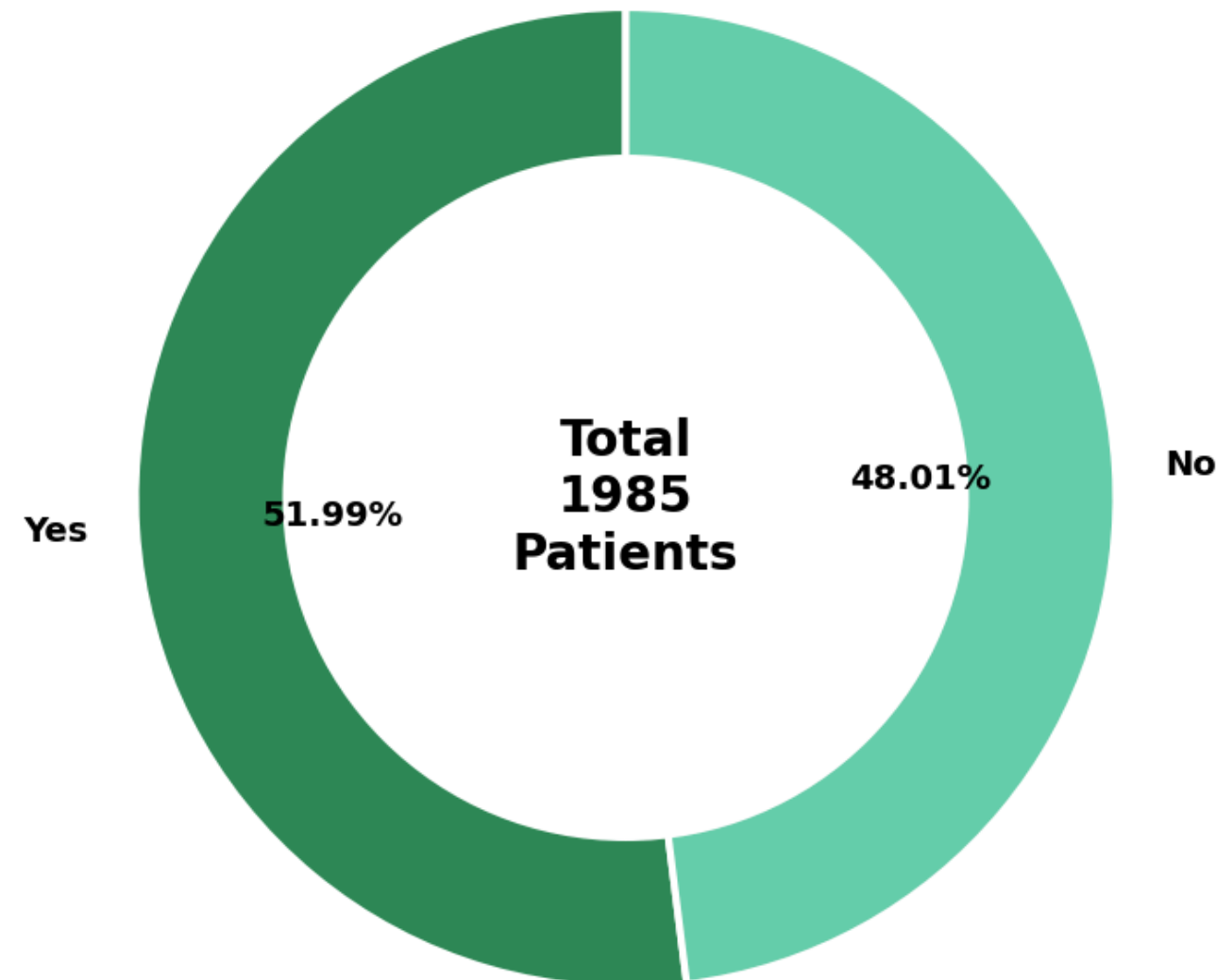


Gambar menunjukkan bahwa pasien dengan riwayat hipertensi keluarga memiliki risiko hipertensi lebih tinggi dibandingkan yang tidak. Artinya, riwayat keluarga menjadi faktor yang signifikan dalam meningkatkan kemungkinan seseorang terkena hipertensi.

Exploratory of Data Analysis



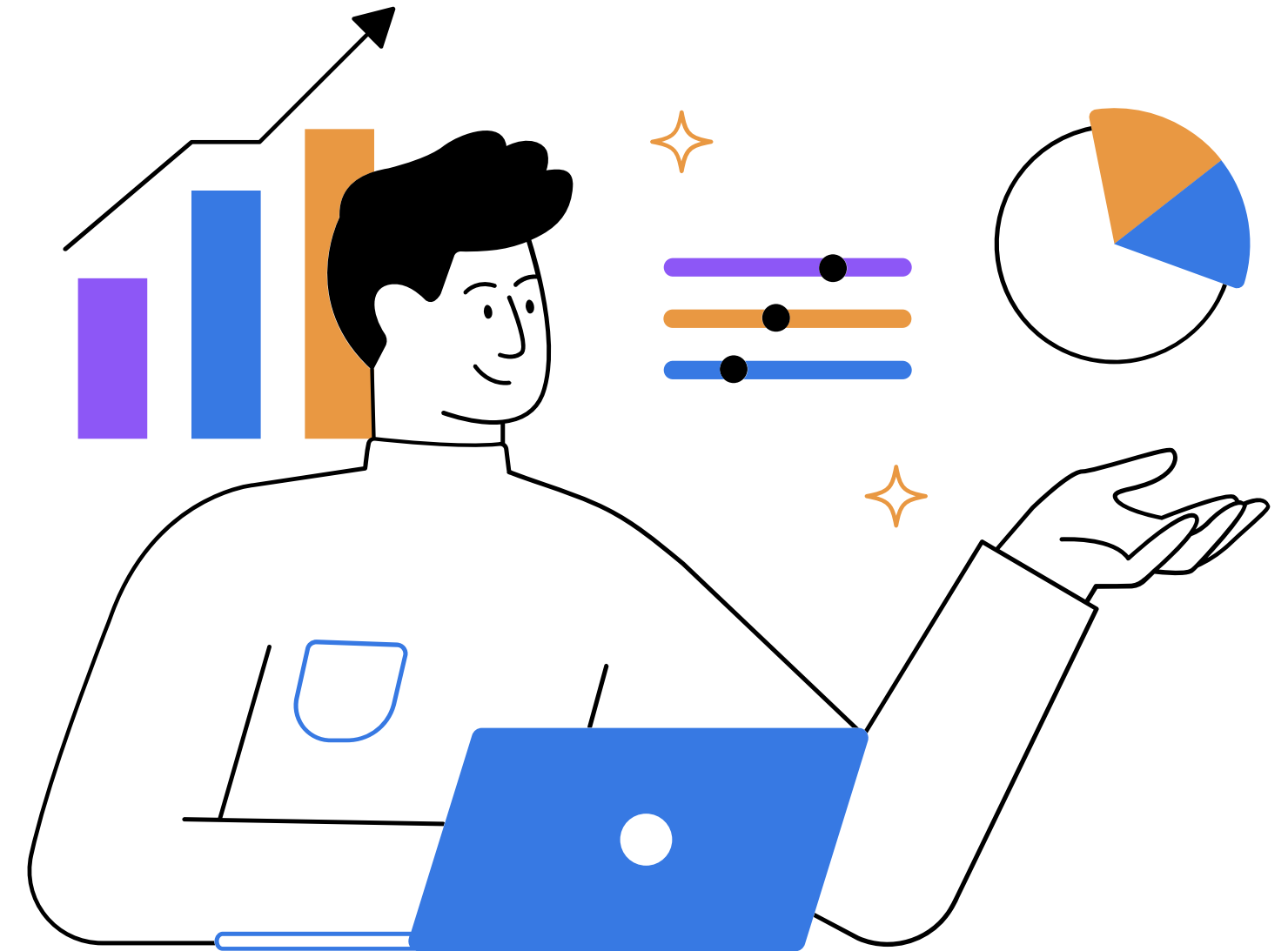
 **Target Variable Distribution: Has Hypertension**



Gambar menunjukkan distribusi pasien berdasarkan status hipertensi, di mana 51,99% pasien menderita hipertensi dan 48,01% tidak. Ini menunjukkan bahwa mayoritas pasien dalam data memiliki hipertensi.

Kesimpulan

Kesimpulan dari analisis menunjukkan bahwa mayoritas pasien (51,99%) menderita hipertensi. Risiko hipertensi meningkat pada pasien dengan riwayat tekanan darah tinggi, perokok, tingkat stres yang tinggi, dan riwayat hipertensi dalam keluarga. Faktor gaya hidup, kondisi klinis, dan genetik memiliki peran penting dalam perkembangan hipertensi, sehingga pencegahan dapat dilakukan melalui deteksi dini dan perubahan gaya hidup sehat.



Thank You



● muzhaffar.rafi05@gmail.com

● <https://www.linkedin.com/in/muzhaffarrafi/>