

DPS

Integrated Data Analytics Competition:
Entertainment, Hospitality, and Health Insights

ANGGOTA KELOMPOK:



MUHAMMAD PANDU



DELYANA IKA



ASYIFA IZAYANI

LATAR BELAKANG

Pencegahan penyakit kronis menjadi fokus utama karena meskipun persentase pasien sehat cukup tinggi, kasus hipertensi, diabetes, dan obesitas masih menonjol. Hal ini menuntut adanya kampanye gaya hidup sehat melalui olahraga, pola nutrisi seimbang, serta edukasi berhenti merokok. Integrasi layanan diabetes dan hipertensi juga penting karena banyak pasien mengalami kedua kondisi tersebut secara bersamaan, sehingga diperlukan screening terpadu untuk deteksi risiko ganda yang lebih efisien. Selain itu, jumlah pasien dengan obesitas dan kebiasaan merokok yang lebih tinggi pada kelompok berpenyakit menegaskan perlunya program manajemen berat badan dan perilaku. Untuk mendukung keterlibatan pasien, peningkatan akses aplikasi digital kesehatan dengan fitur visualisasi hasil pemeriksaan, notifikasi rutin, serta rekomendasi personal menjadi solusi yang relevan.

Dari sisi bisnis dan pemasaran, fokus diarahkan pada segmen dominan yaitu kelompok muda dan dewasa melalui personalisasi penawaran, serta memperkuat pasar keluarga (small family) sebagai segmen tamu terbesar dengan pengembangan paket dan fasilitas yang sesuai. Profitabilitas juga dapat ditingkatkan dengan mengalihkan pemesanan dari travel agent ke kanal langsung seperti website dan walk-in yang memiliki biaya komisi lebih rendah.

Dalam industri hiburan, data penjualan tiket menunjukkan bahwa kategori festival merupakan yang paling diminati dengan persentase 30%. Artis populer seperti Bernadya menjadi magnet utama untuk menarik penonton, sementara pemilihan venue strategis dan penyelenggaraan konser di peak season dapat memaksimalkan pendapatan. Selain itu, strategi penjualan perlu disesuaikan dengan preferensi tiap kota agar konser lebih tepat sasaran dan memberikan hasil optimal.

KATEGORI KESEHATAN

DATA PREPARATION

1. Import Data & Data Understanding



```
DPS.py

import pandas as pd

# import data
df = pd.read_csv("/content/drive/MyDrive/Developer Battle/Kesehatan/kesehatan.csv")

# data info
df.info()
```

CodelImage

Pada bagian awal ini, kami melakukan import pandas sebuah library yang digunakan untuk pengolahan data tabular. Sehingga kami mampu melakukan import dataset yang akan diolah. Setelah itu kami menggunakan fungsi .info() untuk mendapatkan info mengenai tipe data

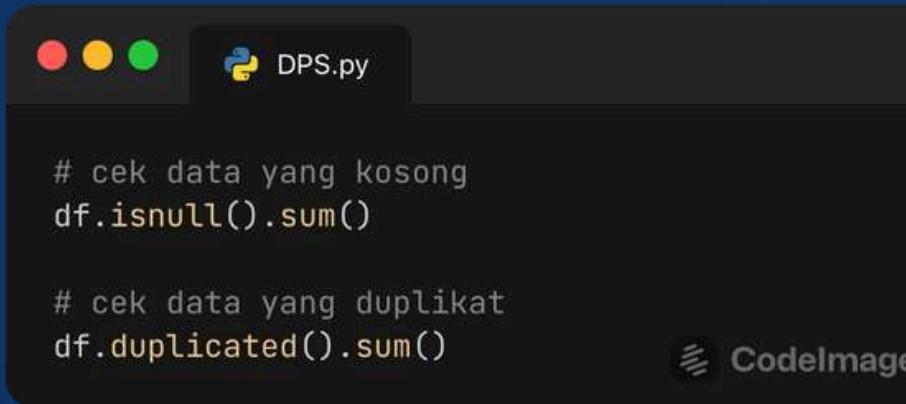
| # | Column | Non-Null Count | Dtype |
|----|---------------------------|----------------|---------|
| 0 | ID | 1545 non-null | int64 |
| 1 | Nama | 1545 non-null | object |
| 2 | Jenis_Kelamin | 1545 non-null | object |
| 3 | Usia | 1545 non-null | int64 |
| 4 | Tinggi_cm | 1545 non-null | float64 |
| 5 | Berat_kg | 1545 non-null | int64 |
| 6 | Tekanan_Darah | 1545 non-null | object |
| 7 | Diabetes | 1545 non-null | object |
| 8 | Cholesterol | 1545 non-null | object |
| 9 | Aktifitas_Fisik | 1468 non-null | object |
| 10 | Merokok | 1468 non-null | object |
| 11 | Riwayat_Penyakit_Keluarga | 1545 non-null | object |
| 12 | Diagnosa_Penyakit | 1545 non-null | object |
| 13 | Tanggal_Pemeriksaan | 1545 non-null | object |

dtypes: float64(1), int64(3), object(10)
memory usage: 169.1+ KB

DATA PREPARATION

2.

Cek Duplikat & Data kosong



```
# cek data yang kosong  
df.isnull().sum()  
  
# cek data yang duplikat  
df.duplicated().sum()
```

CodelImage

Bagian ini dilakukan untuk mengecek data yang kosong serta duplikat. Sehingga nanti ketika proses analysis dilakukan, data yang kosong dan duplikat ini tidak mengganggu hasil analysisnya

output :

| | |
|---------------------------|----|
| ID | 0 |
| Nama | 0 |
| Jenis_Kelamin | 0 |
| Usia | 0 |
| Tinggi_cm | 0 |
| Berat_kg | 0 |
| Tekanan_Darah | 0 |
| Diabetes | 0 |
| Cholesterol | 0 |
| Aktifitas_Fisik | 77 |
| Merokok | 77 |
| Riwayat_Penyakit_Keluarga | 0 |
| Diagnosa_Penyakit | 0 |
| Tanggal_Pemeriksaan | 0 |

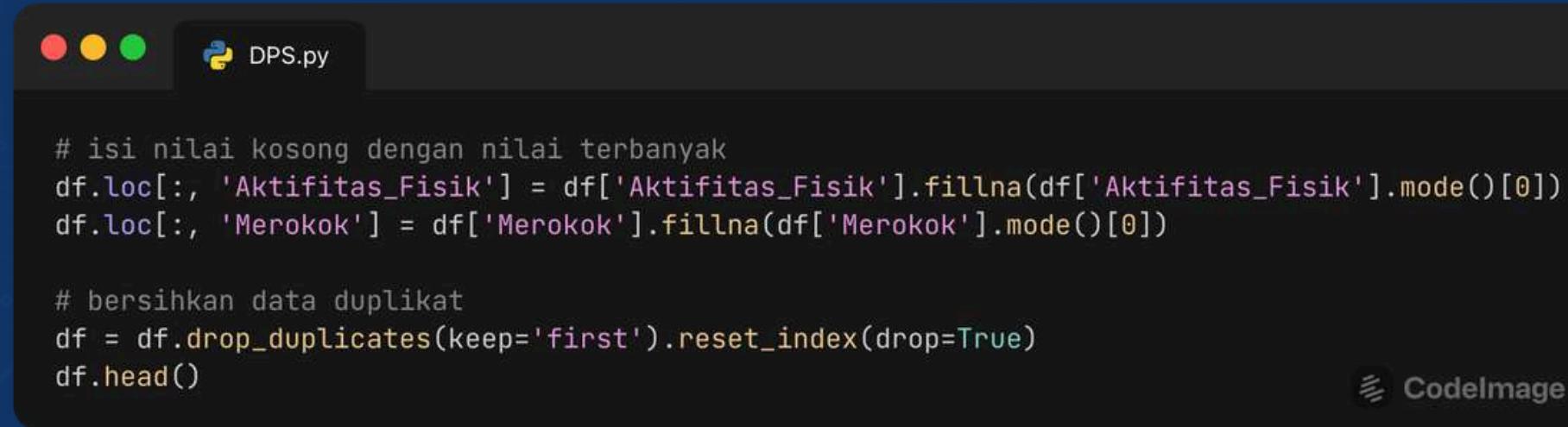


np.int64(29)

DATA PREPARATION

3.

Bersihkan data duplikat & kosong



```
# isi nilai kosong dengan nilai terbanyak
df.loc[:, 'Aktifitas_Fisik'] = df['Aktifitas_Fisik'].fillna(df['Aktifitas_Fisik'].mode()[0])
df.loc[:, 'Merokok'] = df['Merokok'].fillna(df['Merokok'].mode()[0])

# bersihkan data duplikat
df = df.drop_duplicates(keep='first').reset_index(drop=True)
df.head()
```

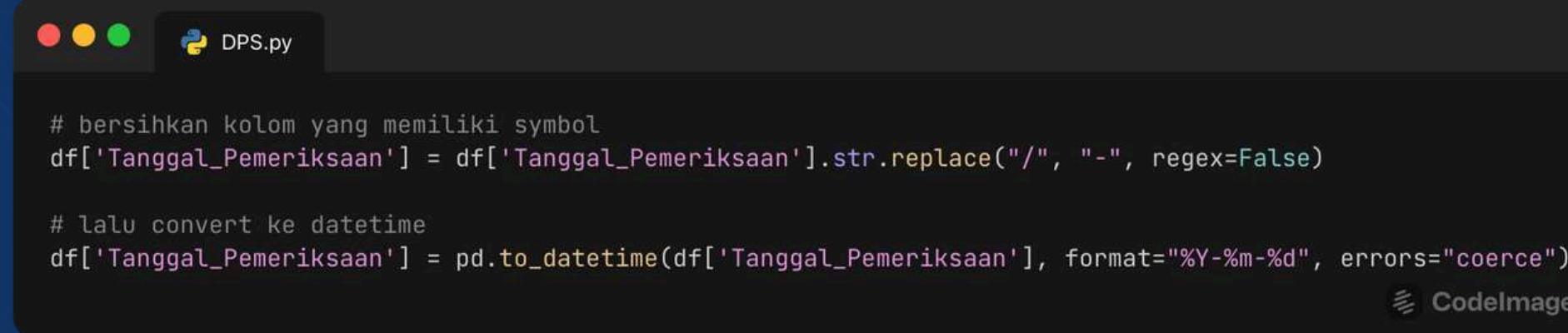
CodelImage

Setelah melakukan pengecekan data duplikat dan kosong di bagian sebelumnya. Bagian ini kami melakukan pembersihan, pada data yang duplikat kami memilih untuk menghapusnya karena dirasa tidak cocok bagi data berkategori kesehatan memiliki fitur yang duplikat. Sementara di bagian data yang kosong kami memilih untuk mengisinya berdasarkan data yang memiliki jumlah terbanyak atau modus

DATA PREPARATION

4.

Ubah tipe data kolom tanggal



```
# bersihkan kolom yang memiliki symbol
df['Tanggal_Pemeriksaan'] = df['Tanggal_Pemeriksaan'].str.replace("/", "-", regex=False)

# lalu convert ke datetime
df['Tanggal_Pemeriksaan'] = pd.to_datetime(df['Tanggal_Pemeriksaan'], format="%Y-%m-%d", errors="coerce")
```

CodelImage

Output:

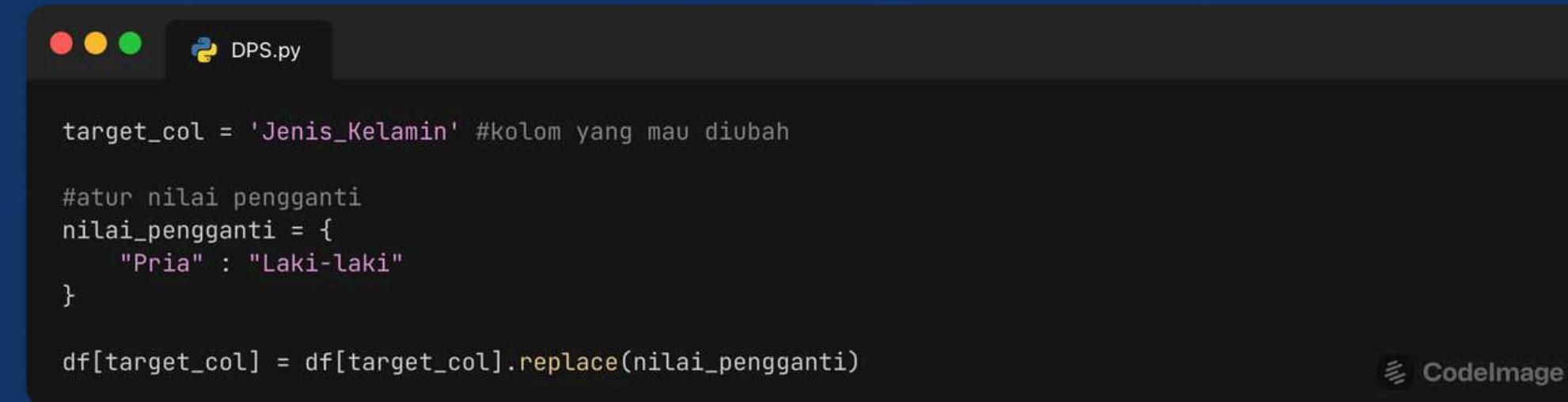
| Tanggal_Pemeriksaan | |
|---------------------|------------|
| 0 | 2024-08-13 |
| 1 | 2024-11-11 |
| 2 | 2024-10-26 |
| 3 | 2025-03-28 |
| 4 | 2023-08-10 |

Karena kolom tanggal memiliki type data yang tidak sesuai (object), kami kemudian mengubah type data dari kolom tanggal. Dimulai dari membersihkan symbol-symbol yang tidak sesuai pada kolom tanggal, kemudian diconvert dalam type data datetime

DATA PREPARATION

5.

Perbaiki kolom Jenis Kelamin



```
target_col = 'Jenis_Kelamin' #kolom yang mau diubah

#atur nilai pengganti
nilai_pengganti = {
    "Pria" : "Laki-laki"
}

df[target_col] = df[target_col].replace(nilai_pengganti)
```

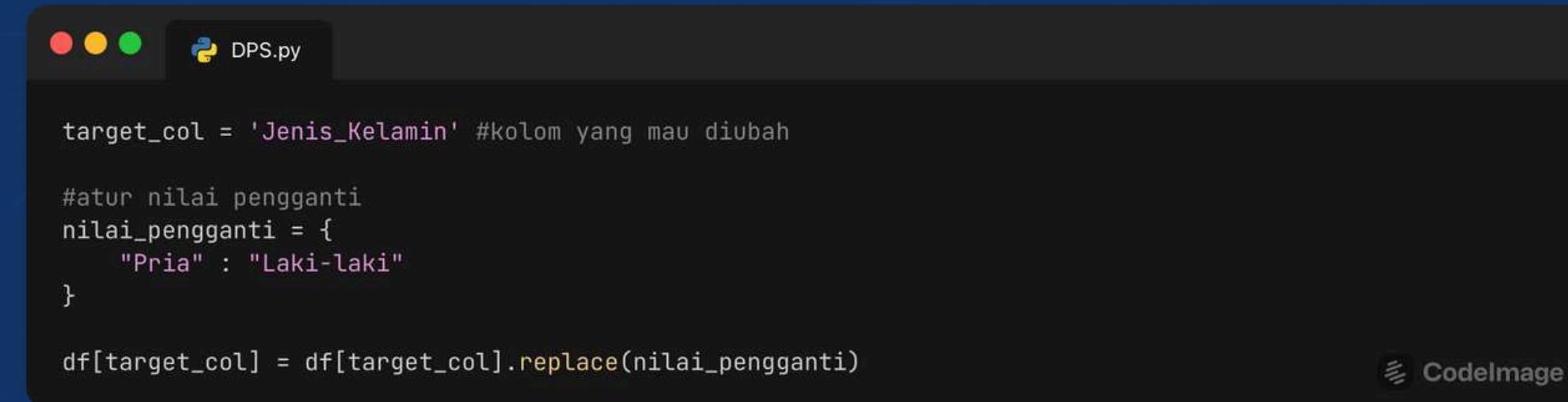
CodelImage

Dikarenakan kolom “Jenis Kelamin” terdapat 2 attribut yang memiliki arti sama, yaitu Pria dan Laki-laki. Hal tersebut dapat menyebabkan hasil analisis data menjadi kacau, maka dari itu kami memperbaiki kolom jenis kelamin. Dengan menyeragamkan perbedaan attribut “Laki-laki” dan “Pria” menjadi satu attribut saja, menjadi “Laki-laki”.

DATA PREPARATION

5.

Perbaiki kolom Jenis Kelamin



```
target_col = 'Jenis_Kelamin' #kolom yang mau diubah

#atur nilai pengganti
nilai_pengganti = {
    "Pria" : "Laki-laki"
}

df[target_col] = df[target_col].replace(nilai_pengganti)
```

CodelImage

Dikarenakan kolom “Jenis Kelamin” terdapat 2 attribut yang memiliki arti sama, yaitu Pria dan Laki-laki. Hal tersebut dapat menyebabkan hasil analisis data menjadi kacau, maka dari itu kami memperbaiki kolom jenis kelamin. Dengan menyeragamkan perbedaan attribut “Laki-laki” dan “Pria” menjadi satu attribut saja, menjadi “Laki-laki”.

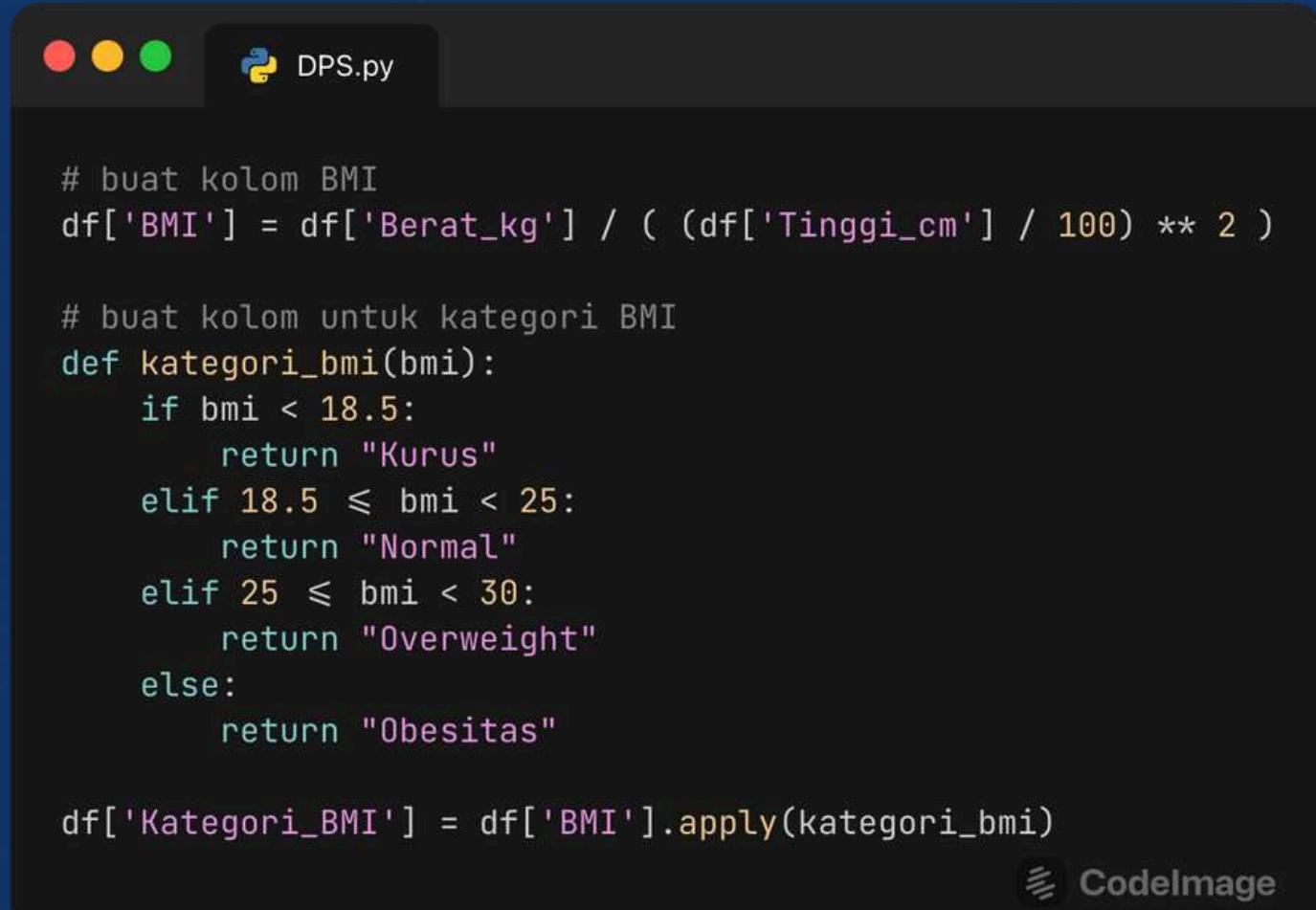
PREPROCESSING

Feature Engineering

1.

Buat kolom BMI dan Kategori BMI

Di bagian ini kami melakukan feature engineering, yaitu membuat kolom BMI serta kolom Kategori BMI. Hal ini dilakukan agar kami bisa mengekstrak fitur serta pola tersembunyi yang ada pada data. Sehingga bisa membantu analisis data



```
# buat kolom BMI
df['BMI'] = df['Berat_kg'] / ((df['Tinggi_cm'] / 100) ** 2)

# buat kolom untuk kategori BMI
def kategori_bmi(bmi):
    if bmi < 18.5:
        return "Kurus"
    elif 18.5 ≤ bmi < 25:
        return "Normal"
    elif 25 ≤ bmi < 30:
        return "Overweight"
    else:
        return "Obesitas"

df['Kategori_BMI'] = df['BMI'].apply(kategori_bmi)
```

CodelImage

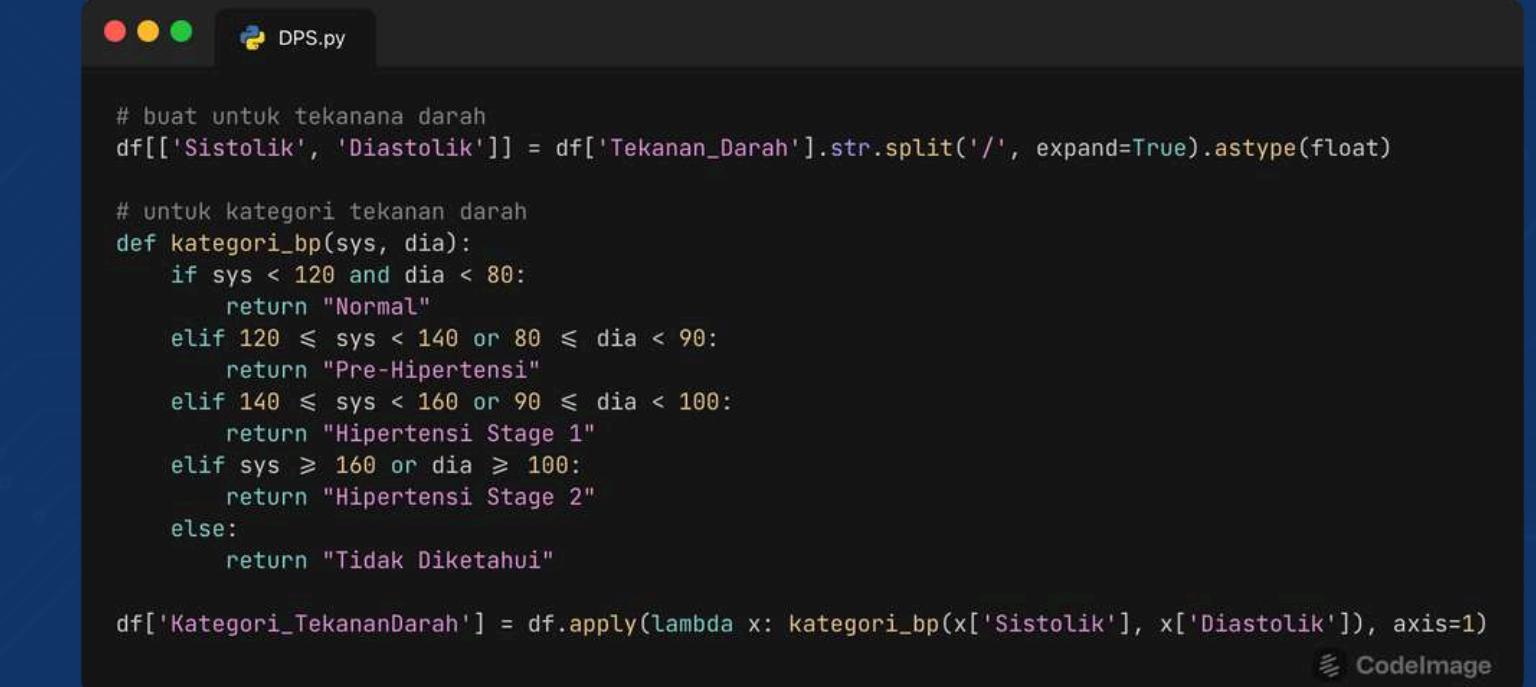
PREPROCESSING

Feature Engineering

2.

Buat kolom Sistolik, Diastolik, serta Kategori

Pada kolom Tekanan_Darah awalnya terdapat tekanan darah sistolik dan diastolik yang digabung. Disini kami memutuskan untuk memisahkannya agar bisa membantu dalam analisis. Kami juga membuat kolom Kategori tekanan darah, yang didasarkan pada rentang tekanan darah sistolik dan diastolik.



```
# buat untuk tekanan darah
df[['Sistolik', 'Diastolik']] = df['Tekanan_Darah'].str.split('/', expand=True).astype(float)

# untuk kategori tekanan darah
def kategori_bp(sys, dia):
    if sys < 120 and dia < 80:
        return "Normal"
    elif 120 ≤ sys < 140 or 80 ≤ dia < 90:
        return "Pre-Hipertensi"
    elif 140 ≤ sys < 160 or 90 ≤ dia < 100:
        return "Hipertensi Stage 1"
    elif sys ≥ 160 or dia ≥ 100:
        return "Hipertensi Stage 2"
    else:
        return "Tidak Diketahui"

df['Kategori_TekananDarah'] = df.apply(lambda x: kategori_bp(x['Sistolik'], x['Diastolik']), axis=1)
```

Codelimage

EXPLORATORY DATA ANALYSIS (EDA)

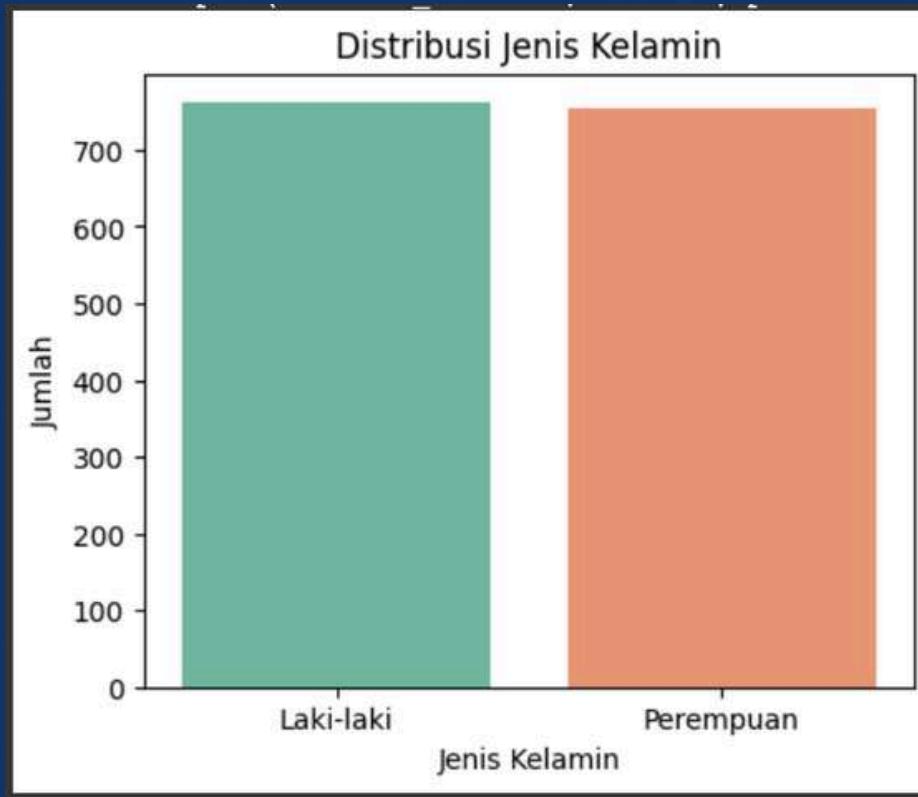
1.

Statistik Deskriptif

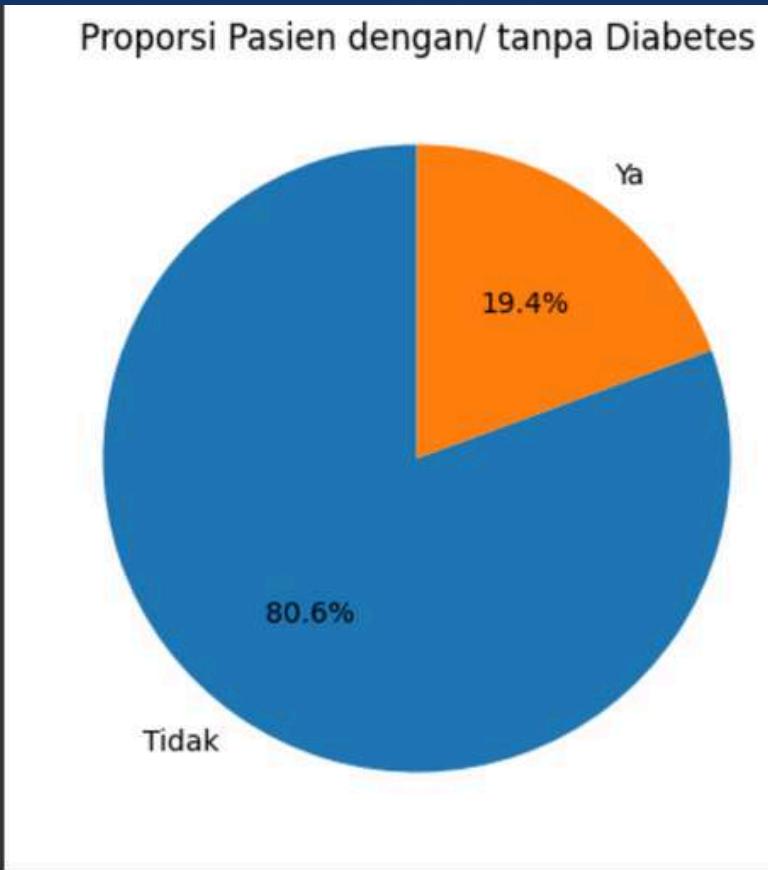
```
# statistik deskriptif fitur numerik
df[['Usia', 'Tinggi_cm', 'Berat_kg', 'BMI', 'Sistolik', 'Diastolik']].describe()
```

| | Usia | Tinggi_cm | Berat_kg | BMI | Sistolik | Diastolik |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|
| count | 1516.000000 | 1516.000000 | 1516.000000 | 1516.000000 | 1516.000000 | 1516.000000 |
| mean | 53.173483 | 165.008575 | 91.552111 | 34.344777 | 133.573219 | 85.112797 |
| std | 22.194927 | 14.731365 | 104.377624 | 39.648882 | 26.088335 | 14.749922 |
| min | 15.000000 | 140.000000 | 45.000000 | 12.742382 | 90.000000 | 60.000000 |
| 25% | 33.000000 | 152.000000 | 61.000000 | 22.597585 | 111.750000 | 72.000000 |
| 50% | 55.000000 | 165.000000 | 78.000000 | 28.085581 | 132.000000 | 85.000000 |
| 75% | 72.000000 | 178.000000 | 94.000000 | 35.224514 | 155.000000 | 98.000000 |
| max | 90.000000 | 190.000000 | 1070.000000 | 431.457986 | 180.000000 | 110.000000 |

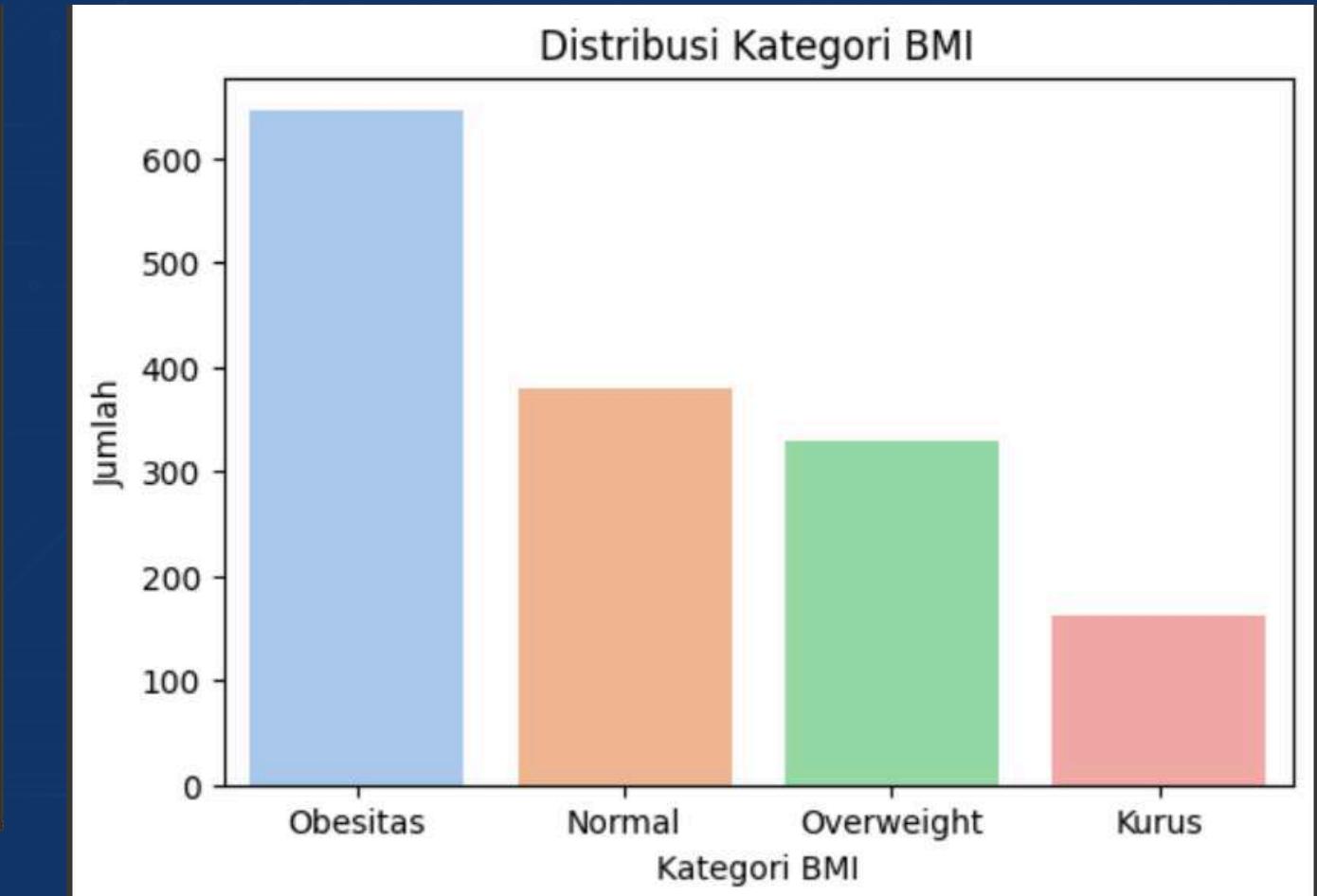
VISUALISASI DATA



Distribusi jenis kelamin laki-laki dan perempuan bisa dilihat seimbang

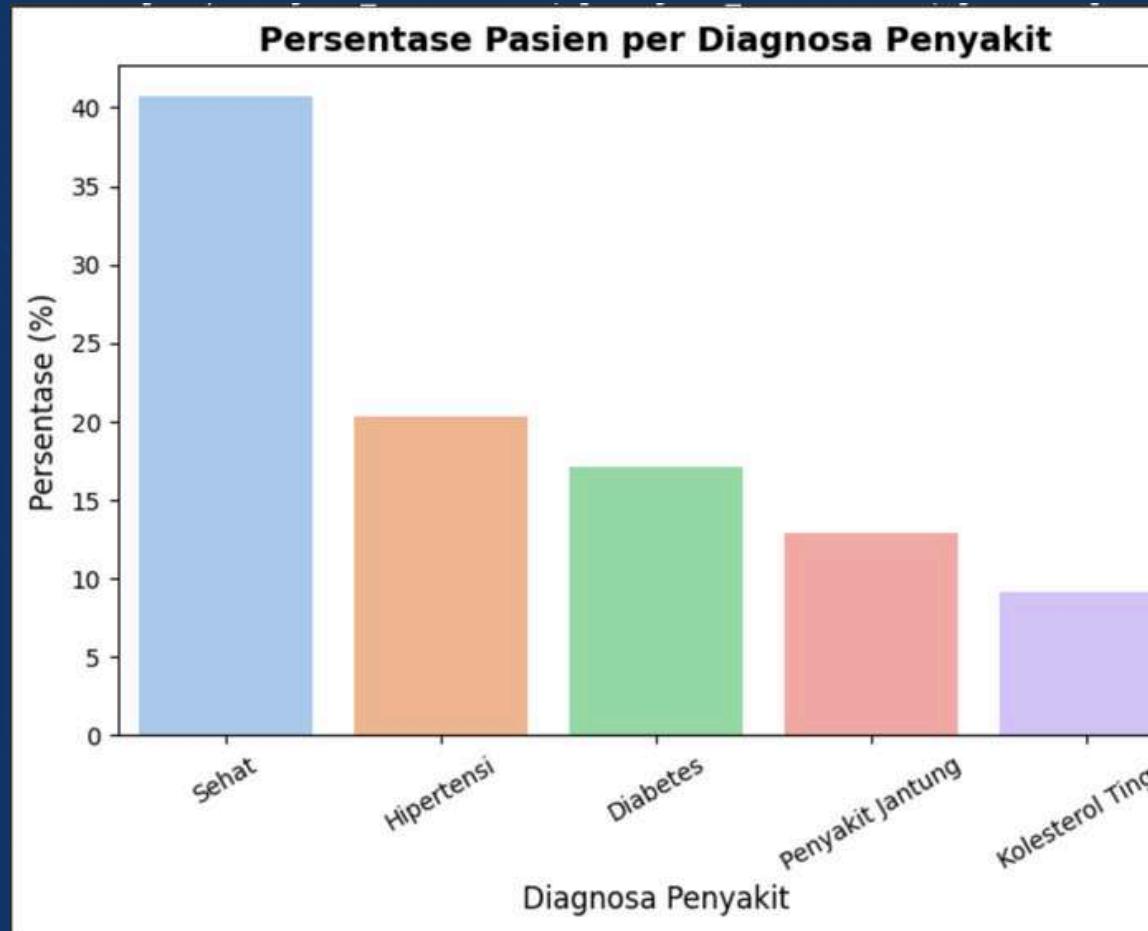


Ternyata ada lebih banyak pasien yang tidak mengidap diabetes

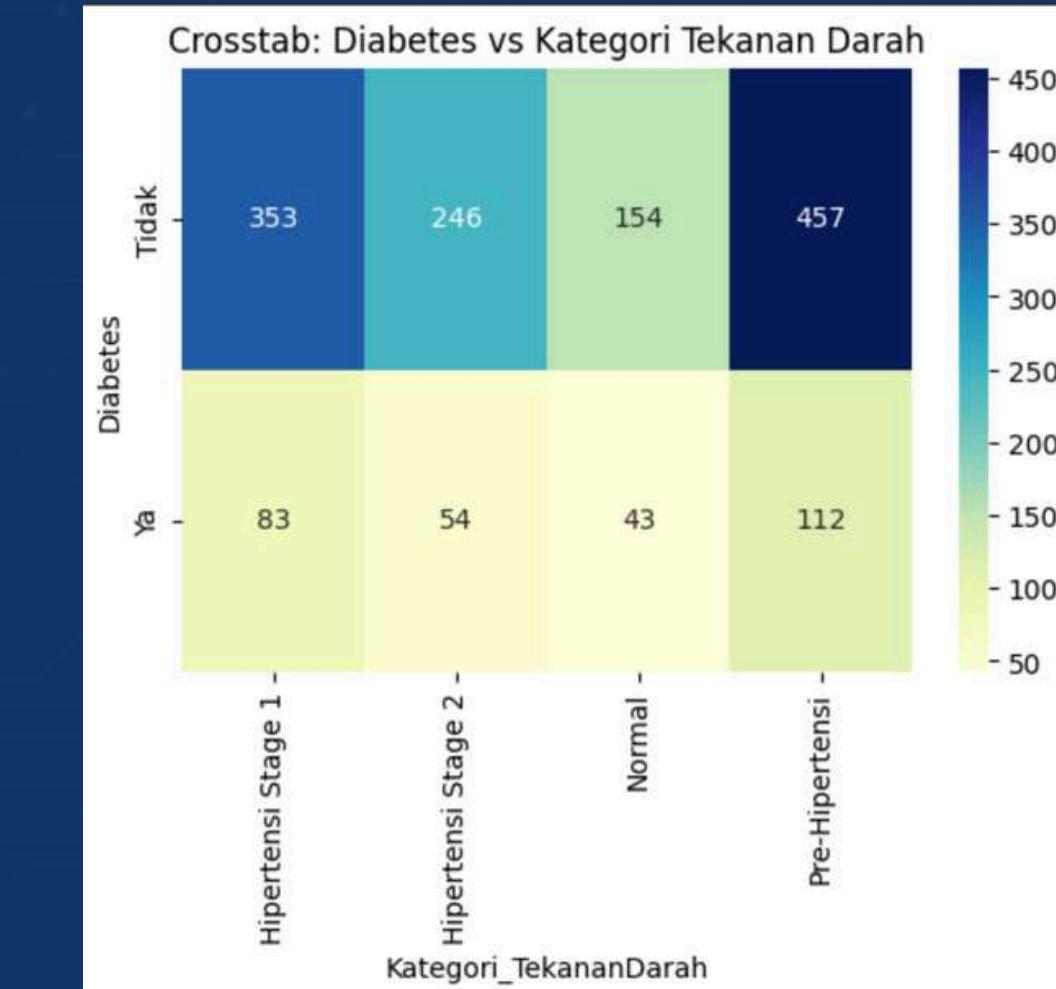


Ternyata ada lebih banyak pasien yang obesitas, dan perbedaan jumlahnya cukup jauh dibandingkan kategori normal

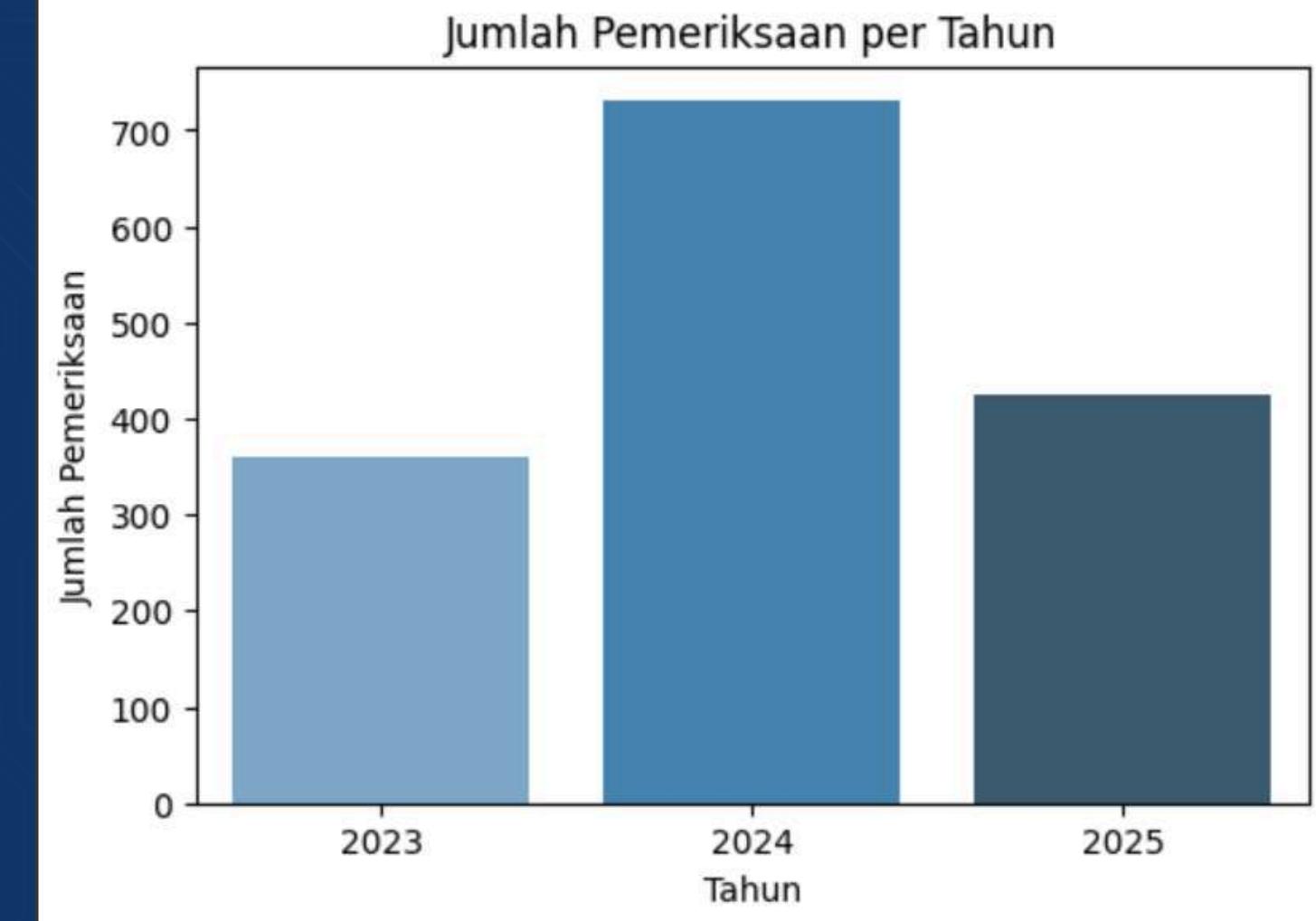
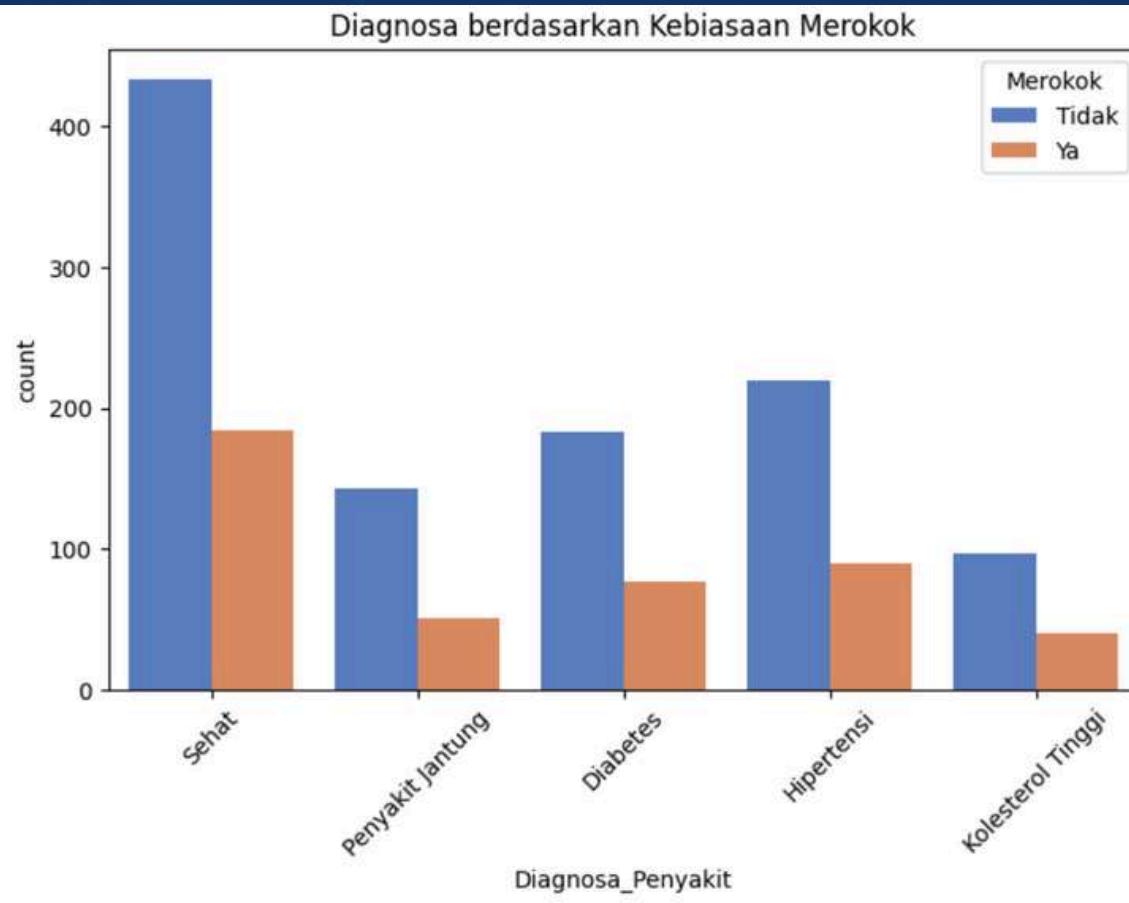
VISUALISASI DATA



Pasien sehat persentasenya lebih tinggi dibandingkan
pasien yang mengidap penyakit tertentu



VISUALISASI DATA



HASIL REKOMENDASI

- **Fokus Pencegahan Penyakit Kronis**

Persentase pasien sehat cukup tinggi, namun kasus hipertensi, diabetes, dan obesitas juga menonjol. Perlu kampanye gaya hidup sehat (olahraga, nutrisi, berhenti merokok) untuk mencegah lonjakan penyakit kronis.

- **Integrasi Layanan Diabetes & Hipertensi**

Banyak pasien diabetes juga memiliki tekanan darah tidak normal. Disarankan adanya screening terpadu agar deteksi risiko ganda lebih cepat dan efisien.

- **Program Manajemen Berat Badan & Perilaku**

Jumlah pasien dengan obesitas dan kebiasaan merokok lebih tinggi pada kelompok berpenyakit. Perlu program intervensi khusus (nutrisi, aktivitas fisik, dan edukasi berhenti merokok).

- **Peningkatan Akses Aplikasi Digital Kesehatan**

Tren pemeriksaan menurun di 2025 bisa disebabkan pengecekan kesehatan pasien yang kurang. Solusi: bangun platform digital kesehatan dengan visualisasi hasil pemeriksaan sederhana, notifikasi rutin, dan rekomendasi personal agar pasien lebih terlibat.

KATEGORI RESERVASI

Colab atau Link Analisis Data

DATA PREPARATION

1. Bagaimana distribusi tamu hotel berdasarkan kelompok usia dan tipe keluarga?
2. Segmen tamu mana yang menjadi target utama dan segmen mana yang memiliki potensi dikembangkan?
3. Bagaimana cara meningkatkan pemesanan langsung (Website, Walk-In, Mobile App) untuk mengurangi ketergantungan pada Travel Agent?

PREPROCESSING

1. Menganalisis karakteristik tamu hotel berdasarkan usia, tipe keluarga, dan cara pemesanan.
2. Mengidentifikasi peluang bisnis jangka pendek dan jangka panjang berdasarkan segmen tamu.
3. Menyusun rekomendasi strategi pemasaran dan layanan untuk meningkatkan profitabilitas hotel.

ANALYSIS



Penelitian : Analisis data tamu hotel dari laporan internal dan sistem reservasi.



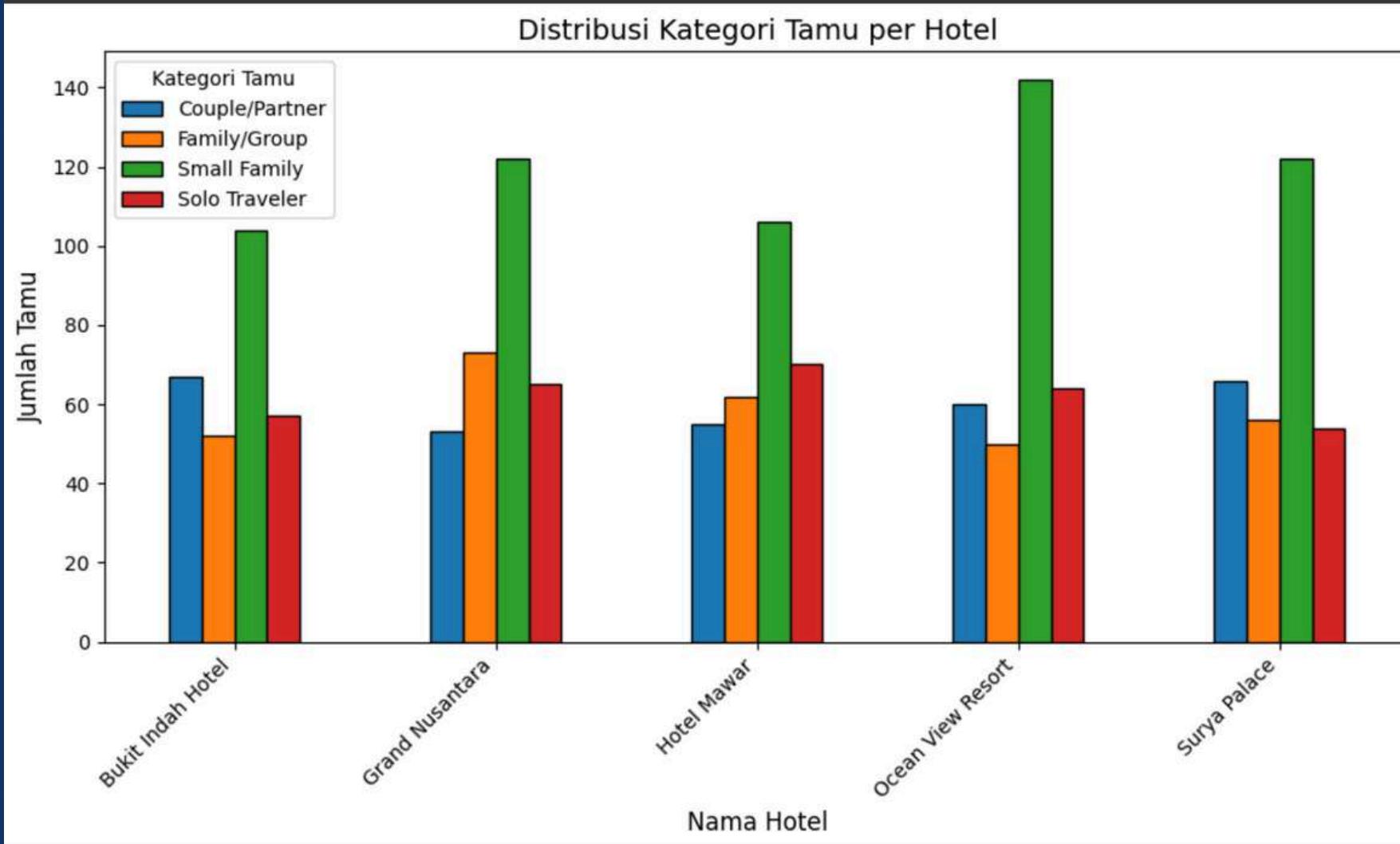
Sumber Data: Dataset reservasi hotel mencakup usia, tipe keluarga, jenis pemesanan, dan informasi hotel terkait.



Analisis Data: Mengkaji strategi inovasi dan strategi promosi yang tepat berdasarkan usia, pola pemesanan dan segmentasi tamu



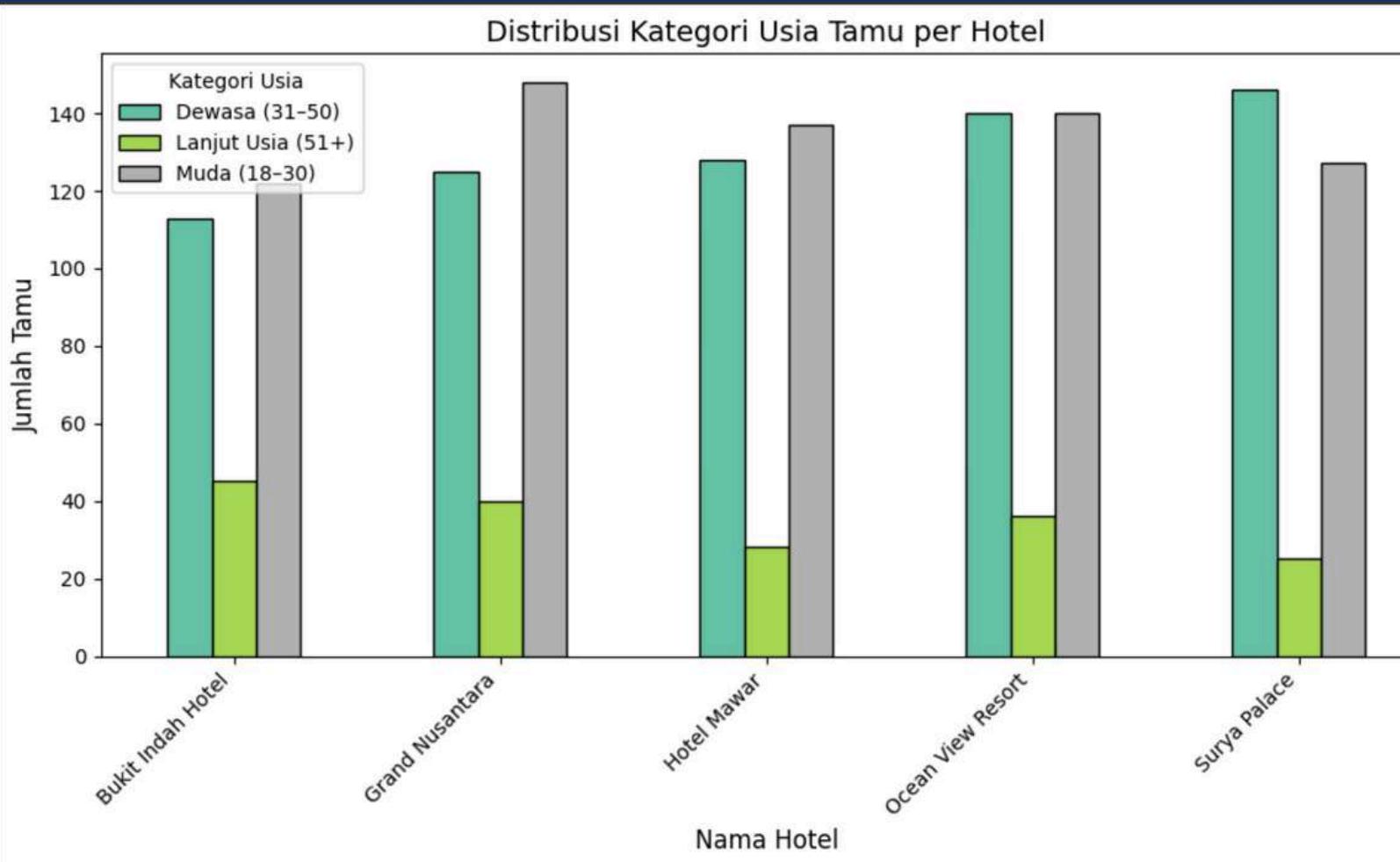
DATA PENELITIAN



Peluang Pengembangan Produk: Memetakan Permintaan dari Segmen Solo, Pasangan, dan Keluarga

Grafik ini membantu memahami "purpose of travel" (tujuan perjalanan) tamu, yang sangat penting untuk pengembangan paket promosi, penyesuaian fasilitas, dan strategi penetapan harga.

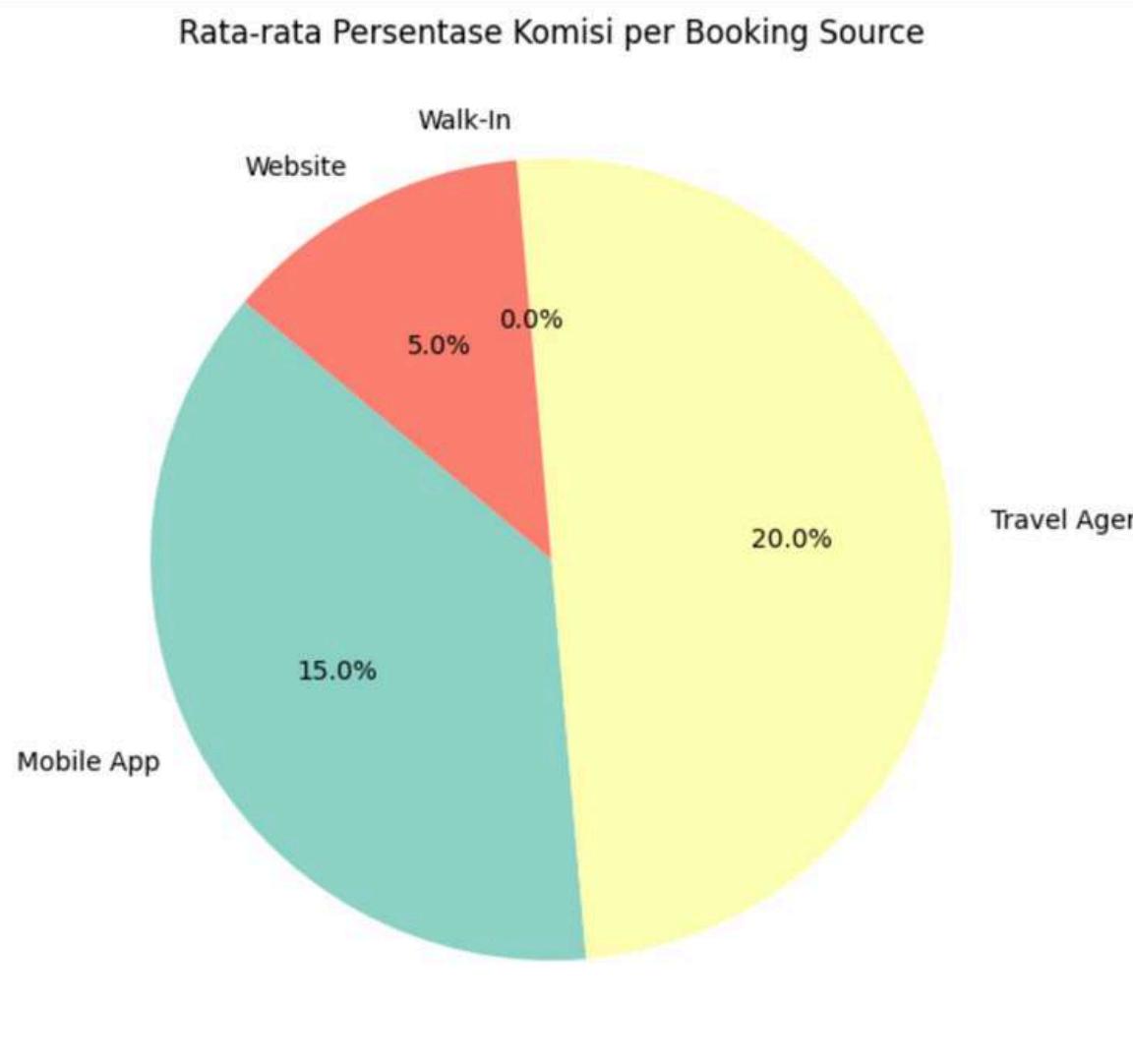
DATA PENELITIAN



Dasar Penyesuaian Strategi Pemasaran: Memahami Siapa Profil Tamu Utama Kita

Grafik ini krusial untuk menentukan target pasar utama, sehingga strategi pemasaran, penawaran produk, dan komunikasi bisa disesuaikan agar lebih efektif dan tepat sasaran.

DATA PENELITIAN



Peluang Peningkatan Margin Melalui Kanal Pemesanan Langsung (Direct Booking)

Grafik ini adalah kunci untuk memahami profitabilitas per saluran penjualan dan mengidentifikasi peluang untuk meningkatkan margin keuntungan dengan mendorong kanal yang lebih murah.

HASIL REKOMENDASI



- **Fokuskan pemasaran pada segmen dominan (Muda & Dewasa), dengan personalisasi penawaran per hotel (contoh: Grand Nusantara untuk tamu Muda, Surya Palace untuk Dewasa/Keluarga).**
- **Menangkan pasar Keluarga (Small Family) dengan memprioritaskan pengembangan paket, fasilitas, dan layanan yang relevan karena ini adalah segmen tamu terbesar.**
- **Tingkatkan profitabilitas dengan mengalihkan pemesanan dari Travel Agent (komisi tinggi) ke kanal langsung seperti Website dan Walk-In (komisi rendah/nol).**

KATEGORI PENJUALAN

Collab Data Penjualan Konser

DATA PREPARATION

```
[29]
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1575 entries, 0 to 1574
Data columns (total 10 columns):
 #   Column      Non-Null Count  Dtype  
0   Order_ID    1501 non-null    object 
1   Artist       1505 non-null    object 
2   City         1526 non-null    object 
3   Venue        1512 non-null    object 
4   Date         1506 non-null    object 
5   Ticket_Catogory 1512 non-null  object 
6   Ticket_Price 1516 non-null    float64
7   Buyer_Name   1516 non-null    object 
8   Quantity     1524 non-null    float64
9   Total_Price  1502 non-null    float64
dtypes: float64(3), object(7)
memory usage: 123.2+ KB

[38]
df.isnull().sum()

Order_ID    74
Artist      70
City        49
Venue       63
Date        69
Ticket_Catogory 63
Ticket_Price 59
Buyer_Name   59
Quantity     51
Total_Price  73

dtype: int64
```

```
[110]
# 1. Jika Total_Price kosong -> isi dengan Quantity * Ticket_Price
df['Total_Price'] = df['Total_Price'].fillna(df['Quantity'] * df['Ticket_Price'])

# 2. Jika Quantity kosong -> isi dengan Total_Price / Ticket_Price
df['Quantity'] = df['Quantity'].fillna(df['Total_Price'] / df['Ticket_Price'])

# 3. Jika Ticket_Price kosong -> isi dengan Total_Price / Quantity
df['Ticket_Price'] = df['Ticket_Price'].fillna(df['Total_Price'] / df['Quantity'])

# Untuk kolom teks (object), bisa pakai modus (nilai terbanyak)
object_cols = df.select_dtypes(include='object').columns

for col in object_cols:
    if col != 'Date': # skip kolom Date
        df[col] = df[col].fillna(df[col].mode()[0])
```

```
[111]
df.isnull().sum()

Order_ID    0
Artist      0
City        0
Venue       0
Date        0
Ticket_Catogory 0
Ticket_Price 0
Buyer_Name   0
Quantity     0
Total_Price  0

dtype: int64
```

```
[112]
# Normalisasi venue -> jadi huruf besar semua
df["venue_norm"] = df["Venue"].str.upper()

# Normalisasi artis -> kapitalisasi sesuai format judul
df["artis_norm"] = df["Artist"].str.title()
```

```
[114]
# Ubah kolom Date ke format datetime (abaikan error biar NaN kalau format salah)
df["Date"] = pd.to_datetime(df["Date"], errors="coerce")

# Tambahkan kolom bulan
df['Month'] = df['Date'].dt.to_period('M')

# Hitung revenue per bulan
monthly_revenue = df.groupby('Month')['Total_Price'].sum()

plt.figure(figsize=(10,5))
monthly_revenue.plot(kind='line', marker='o')
plt.title('Tren Total Revenue per Bulan')
plt.ylabel('Total Revenue (Rp)')
plt.xlabel('Month')
plt.grid(True, linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

PREPROCESSING

1.

Penanganan Nilai yang Hilang (Missing Values)

Kami memeriksa nilai kosong menggunakan `df.isnull().sum()`, lalu mengisi nilai yang hilang pada kolom numerik (`Total_Price`, `Quantity`, `Ticket_Price`) berdasarkan perhitungan antar kolom, dan pada kolom teks (kecuali `Date`) menggunakan modus (nilai terbanyak).

2.

Konversi Tipe Data

Konversi tipe data dilakukan agar data sesuai dengan tujuan analisis, perhitungan, maupun visualisasi.

3.

Normalisasi Data Teks

Nama venue dinormalisasi menjadi huruf kapital semua di kolom `venue_norm`, dan nama artis dinormalisasi menjadi format judul di kolom `artis_norm` untuk memastikan konsistensi data.

ANALYSIS

- Distribusi Kategori Tiket

Data menunjukkan bahwa penjualan tiket terbagi dalam empat kategori dengan proporsi cukup seimbang. Kategori Festival mendominasi dengan 30,1%, diikuti oleh Gold (24,7%), Silver (22,2%), dan VIP (23,1%). Hal ini mengindikasikan bahwa peminat tiket festival lebih tinggi dibanding kategori lainnya.

- Top 5 Artis Berdasarkan Total Revenue

Dari sisi pendapatan, artis Bernadya memberikan kontribusi terbesar yang sangat signifikan dibandingkan artis lain. Sementara itu, artis seperti Efek Rumah Kaca, Isyana Sarasvati, Nadin Amizah, dan Pamungkas memberikan kontribusi lebih kecil, menunjukkan adanya ketimpangan distribusi pendapatan per artis.

- Tren Total Revenue per Bulan

Grafik menunjukkan adanya lonjakan pendapatan yang cukup tajam pada bulan tertentu, sementara bulan lainnya relatif stabil dengan nilai rendah. Hal ini bisa disebabkan oleh adanya konser besar atau event tertentu pada periode tersebut.

- Distribusi Revenue Berdasarkan Venue

Analisis per venue memperlihatkan bahwa ada satu venue dengan kontribusi terbesar terhadap revenue, sedangkan venue lainnya relatif kecil. Hal ini bisa menjadi dasar evaluasi untuk menentukan venue potensial di masa depan.

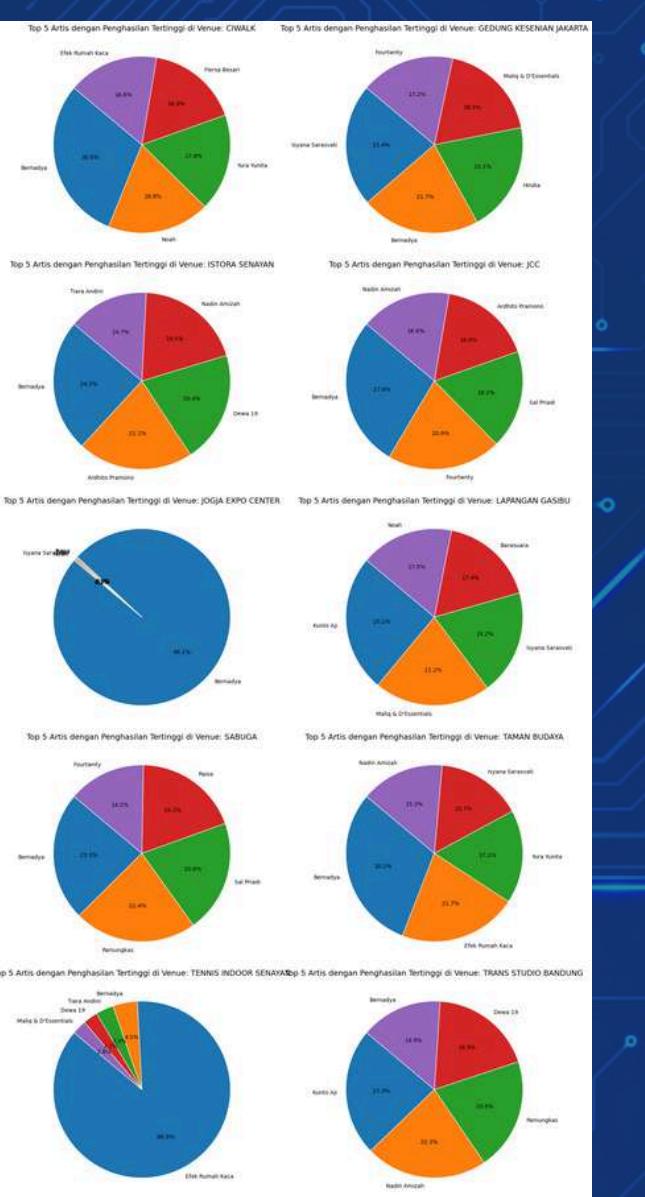
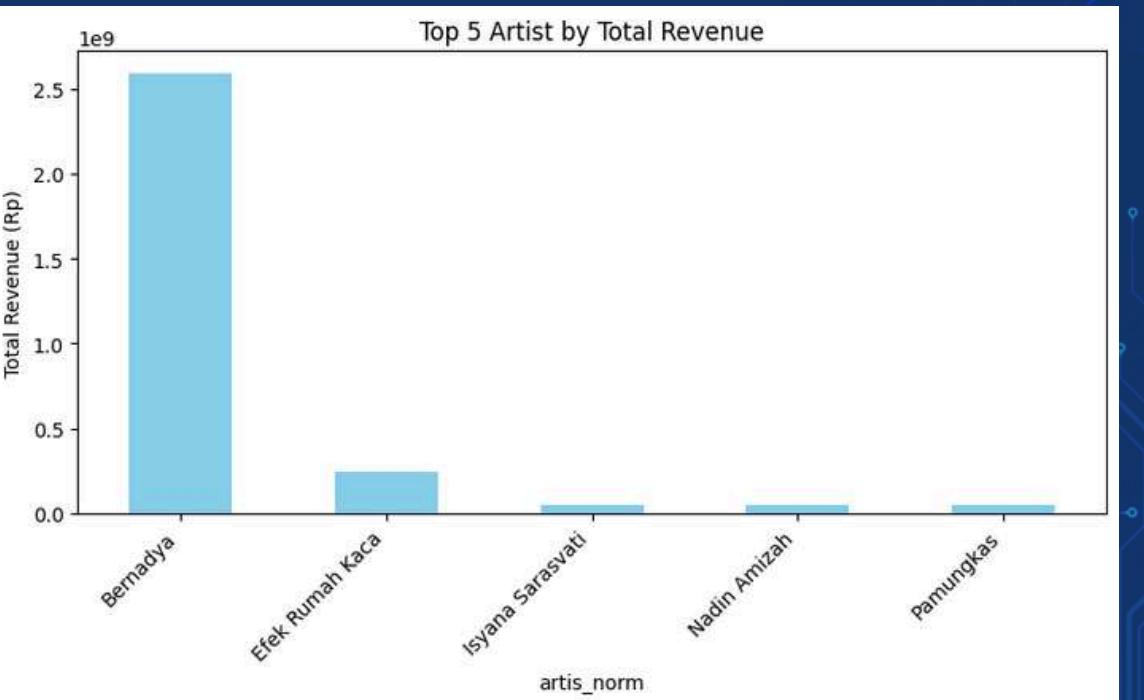
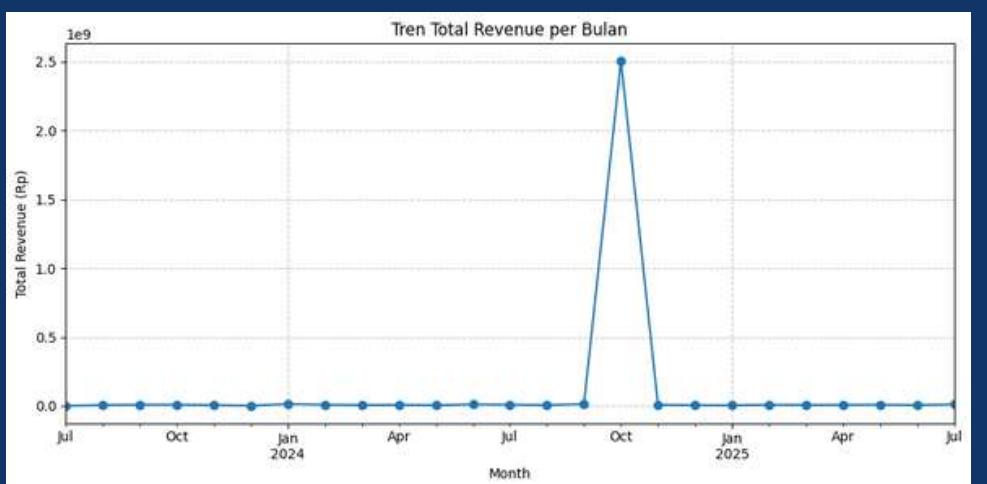
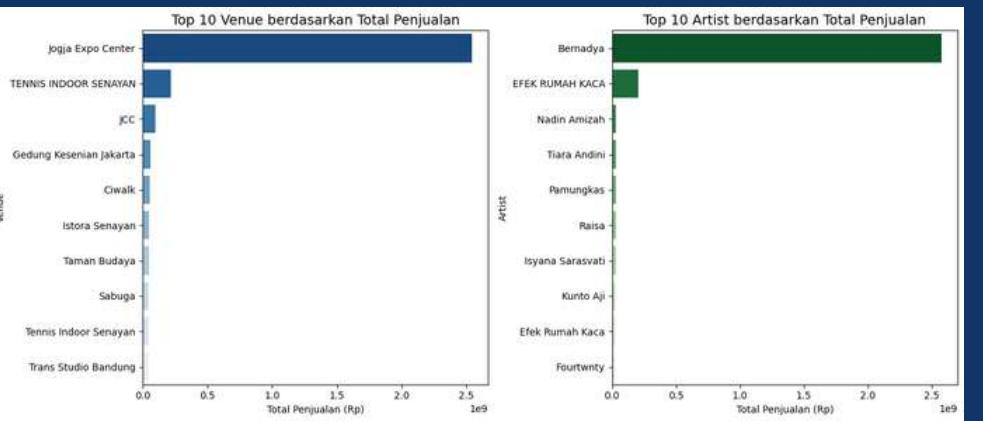
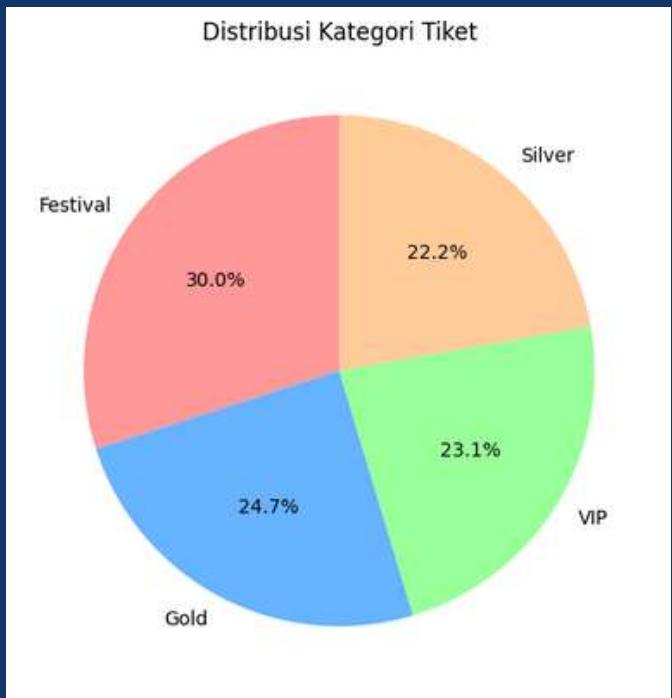
- Distribusi Revenue Berdasarkan Kota

Kota tertentu terlihat lebih dominan dalam menghasilkan revenue, menunjukkan bahwa lokasi memiliki pengaruh besar terhadap tingkat penjualan tiket. Kota besar dengan populasi tinggi cenderung memberikan kontribusi lebih signifikan.

- Perbandingan Visualisasi Pie Chart Lain

Beberapa diagram lingkaran yang ditampilkan menunjukkan distribusi data tambahan, misalnya perbandingan kategori tiket di setiap kota atau venue. Dari visualisasi ini terlihat bahwa preferensi pembelian tiket bervariasi tergantung lokasi, meskipun tren dominasi kategori Festival tetap konsisten.

VISUALISASI DATA



HASIL REKOMENDASI



- Perbanyak tiket Festival dapat dilihat kategori paling diminati sebanyak 30%.
- Prioritaskan artis populer untuk konser utama, seperti: Bernadya, dll.
- Fokus pada venue strategis dengan revenue tertinggi.
- Gelaran konser besar di peak season (lonjakan penjualan terlihat).
- Sesuaikan strategi per kota sesuai preferensi tiket.

KESIMPULAN

Berdasarkan hasil analisis, dapat disimpulkan bahwa strategi pengembangan perlu difokuskan pada tiga aspek utama. Pada bidang kesehatan, penting untuk memperkuat pencegahan penyakit kronis, integrasi layanan diabetes-hipertensi, serta program manajemen obesitas dan perilaku, yang didukung dengan platform digital agar pasien lebih terlibat. Pada sektor reservasi, optimalisasi dapat dilakukan dengan memaksimalkan segmen dominan (Muda, Dewasa, dan Keluarga) melalui personalisasi penawaran, sekaligus mendorong pemesanan langsung untuk meningkatkan profitabilitas. Sementara pada penjualan, strategi terbaik adalah memperbanyak tiket festival dan konser populer, menghadirkan artis utama, memilih venue strategis, serta mengatur jadwal pada peak season untuk mendorong pendapatan maksimal.

TERIMA KASIH