

PROJECT UAS DATA MINING & BI

**EKSPLORASI DATA
KESEHATAN DENGAN
PROSES MINING**

ANGGOTA KELompok

Chatlea Shakira Haq

2106725116

Fernaldy

2106706464

Jihan Sandrina Halim

2106708160

Niken Salsabila Helmelia

2106724933

Zahrah Mahfuzah

2106704004

DATA UNDERSTANDING

Data yang digunakan terdiri dari **289.800 baris** dan **79 kolom**, menyajikan informasi tentang sejumlah besar individu.

Setiap baris mencakup data tentang satu individu, sementara setiap kolom mewakili berbagai atribut atau variabel yang mencakup berbagai aspek seperti **demografi** (usia, jenis kelamin, wilayah, dsb.), **kesehatan** (riwayat penyakit, tinggi badan, berat badan, dsb.), **kebiasaan hidup** (merokok, konsumsi alkohol, pola makan, dsb.), dan **lingkungan sekitar** (fasilitas umum, sanitasi rumah tangga, dsb.).

TUJUAN ANALISIS

Dari data yang diberikan, kami akan fokus menganalisis beberapa hal sebagai berikut:

- Penyakit apa yang paling umum diderita masyarakat
- Korelasi penyakit kronis
- Distribusi umur untuk setiap penyakit
- Distribusi demografis individu dengan kesehatan mental yang membutuhkan perhatian khusus berdasarkan wilayah, jenis kelamin, status perkawinan, dan kelompok umur
- Distribusi demografis individu penderita obesitas berdasarkan wilayah, jenis kelamin, dan kelompok umur
- Pola konsumsi dan aktivitas fisik individu penderita obesitas
- Hubungan antara kesehatan mental dengan status perkawinan dan kondisi obesitas

PREPROCESSING

1. Memeriksa Tipe Variabel

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 289801 entries, 0 to 289800
Data columns (total 79 columns):
 #   Column           Non-Null Count  Dtype  
 ---  --  
 0   ID               289801 non-null   int64  
 1   WILAYAH          289801 non-null   object  
 2   JENIS KELAMIN    289801 non-null   object  
 3   USIA             289801 non-null   float64 
 4   LAMA BERDINAS   289801 non-null   float64 
 5   PENDIDIKAN TERAKHIR  289801 non-null   object  
 6   TINGGI BADAN     289801 non-null   object  
 7   BERAT BADAN      289801 non-null   float64 
 8   LINGKAR PINGGANG 289801 non-null   float64 
 9   Status Perkawinan 289801 non-null   object  
 10  PENGELOUARAN/BULAN 289801 non-null   object
```

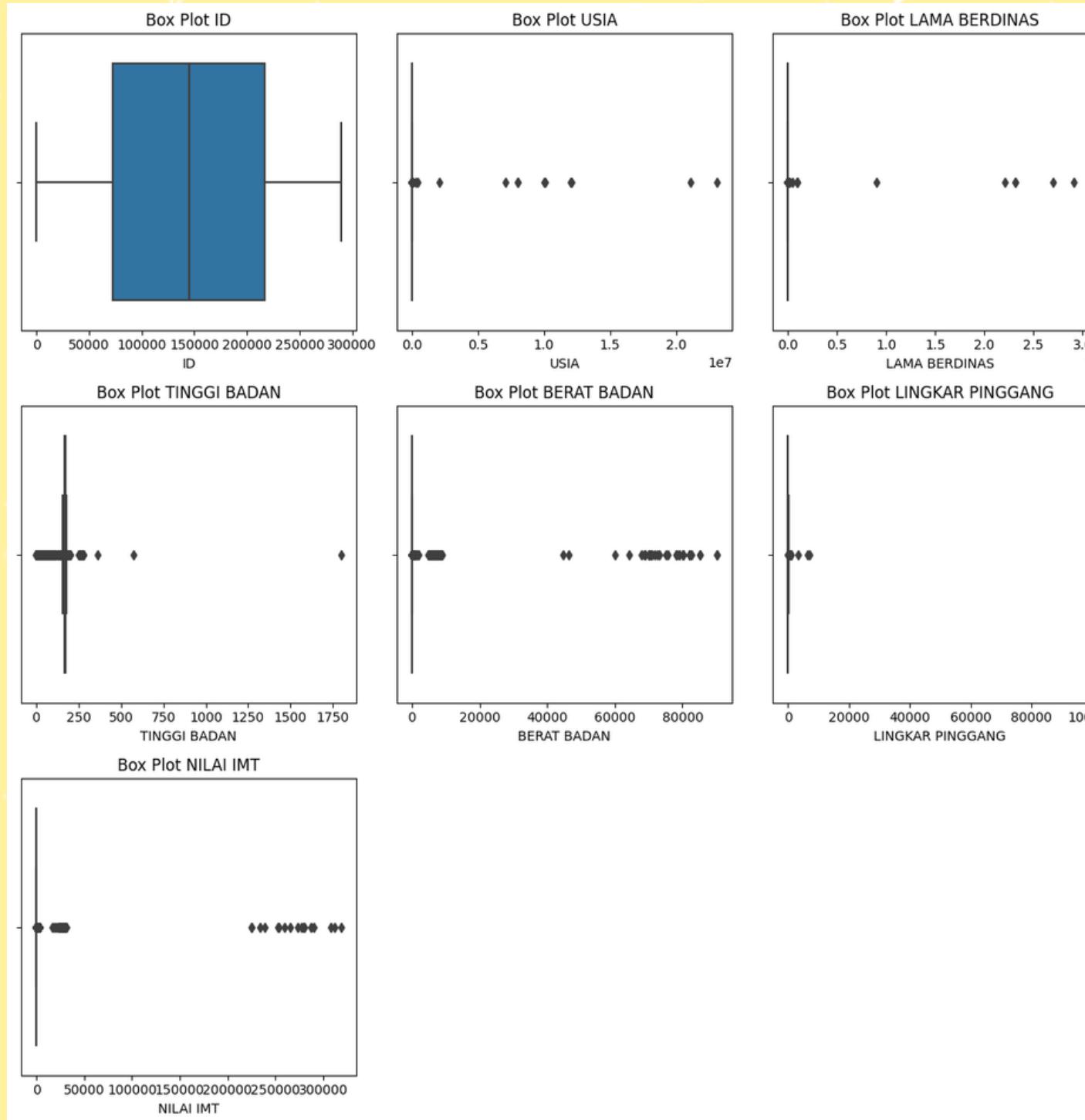
Diperlihatkan tipe data dari 10 variabel teratas dan diperoleh bahwa terdapat ketidaksesuaian tipe data untuk variabel TINGGI BADAN, yang mana seharusnya adalah *float*. Sehingga, terlebih dahulu harus disesuaikan.

Dengan menggunakan kode berikut akan diubah tipe data untuk TINGGI badan menjadi *float*

```
1 df["TINGGI BADAN"] = df["TINGGI BADAN"].replace(",",".", regex = True)
2 df["TINGGI BADAN"] = df["TINGGI BADAN"].astype(float)
```

PREPROCESSING

2. Memeriksa Noise dan Outlier



Dapat dilihat bahwa sangat banyak outlier untuk setiap variabel numberik dalam data. Namun, berdasarkan tujuan analisis, maka **kami hanya mempertimbangkan untuk menangani outlier pada variabel USIA, TINGGI BADAN, dan BERAT BADAN**. Sedangkan, NILAI IMT nanti akan menyesuaikan dengan nilai tinggi dan berat badan.

PREPROCESSING

2.1 Menangani Noise dan Outlier TINGGI BADAN

Diketahui bahwa **range nilai tinggi badan dalam data adalah 1,64 - 1799**. Kami mengasumsikan bahwa satuan tinggi badan dalam data adalah dalam cm, sehingga beberapa pertimbangan yang kami lakukan untuk menangani noise dan outlier dalam tinggi badan sebagai berikut.

```
1 df["TINGGI BADAN"] = df["TINGGI BADAN"].replace(".", "", regex = True)
2 df.loc[df["TINGGI BADAN"] <= 2, "TINGGI BADAN"] = df["TINGGI BADAN"]*100
3 df.loc[df["TINGGI BADAN"] == 5, "TINGGI BADAN"] = df["TINGGI BADAN"]*30
4 df.loc[(df["TINGGI BADAN"] > 10) & (df["TINGGI BADAN"] < 20), "TINGGI BADAN"] = df["TINGGI BADAN"]*10
5 df.loc[(df["TINGGI BADAN"] >= 20) & (df["TINGGI BADAN"] < 30), "TINGGI BADAN"] = df["TINGGI BADAN"]*6
6 df.loc[(df["TINGGI BADAN"] >= 30) & (df["TINGGI BADAN"] < 40), "TINGGI BADAN"] = df["TINGGI BADAN"]*5
7 df.loc[(df["TINGGI BADAN"] >= 40) & (df["TINGGI BADAN"] < 50), "TINGGI BADAN"] = df["TINGGI BADAN"]*4
8 df.loc[(df["TINGGI BADAN"] >= 50) & (df["TINGGI BADAN"] < 100), "TINGGI BADAN"] = df["TINGGI BADAN"]+100
9 df.loc[(df["TINGGI BADAN"] >= 100) & (df["TINGGI BADAN"] < 140), "TINGGI BADAN"] = df["TINGGI BADAN"]+50
10 df.loc[(df["TINGGI BADAN"] >= 200) & (df["TINGGI BADAN"] < 250), "TINGGI BADAN"] = df["TINGGI BADAN"]-50
11 df.loc[(df["TINGGI BADAN"] >= 250) & (df["TINGGI BADAN"] < 300), "TINGGI BADAN"] = df["TINGGI BADAN"]-100
12 df.loc[(df["TINGGI BADAN"] >= 300) & (df["TINGGI BADAN"] < 400), "TINGGI BADAN"] = df["TINGGI BADAN"]/2
13 df.loc[(df["TINGGI BADAN"] >= 400) & (df["TINGGI BADAN"] < 1000), "TINGGI BADAN"] = df["TINGGI BADAN"]/3
14 df.loc[df["TINGGI BADAN"] > 1000, "TINGGI BADAN"] = df["TINGGI BADAN"]/10
```

Kemudian, setelah dilakukan penyesuaian, diperoleh **range nilai tinggi badan yang baru adalah 140 - 199 cm**.

PREPROCESSING

2.1 Menangani Noise dan Outlier TINGGI BADAN

Diketahui bahwa **range nilai tinggi badan dalam data adalah 1,64 - 1799**. Kami mengasumsikan bahwa satuan tinggi badan dalam data adalah dalam cm, sehingga beberapa pertimbangan yang kami lakukan untuk menangani noise dan outlier dalam tinggi badan sebagai berikut.

```
1 df["TINGGI BADAN"] = df["TINGGI BADAN"].replace(".", "", regex = True)
2 df.loc[df["TINGGI BADAN"] <= 2, "TINGGI BADAN"] = df["TINGGI BADAN"]*100
3 df.loc[df["TINGGI BADAN"] == 5, "TINGGI BADAN"] = df["TINGGI BADAN"]*30
4 df.loc[(df["TINGGI BADAN"] > 10) & (df["TINGGI BADAN"] < 20), "TINGGI BADAN"] = df["TINGGI BADAN"]*10
5 df.loc[(df["TINGGI BADAN"] >= 20) & (df["TINGGI BADAN"] < 30), "TINGGI BADAN"] = df["TINGGI BADAN"]*6
6 df.loc[(df["TINGGI BADAN"] >= 30) & (df["TINGGI BADAN"] < 40), "TINGGI BADAN"] = df["TINGGI BADAN"]*5
7 df.loc[(df["TINGGI BADAN"] >= 40) & (df["TINGGI BADAN"] < 50), "TINGGI BADAN"] = df["TINGGI BADAN"]*4
8 df.loc[(df["TINGGI BADAN"] >= 50) & (df["TINGGI BADAN"] < 100), "TINGGI BADAN"] = df["TINGGI BADAN"]+100
9 df.loc[(df["TINGGI BADAN"] >= 100) & (df["TINGGI BADAN"] < 140), "TINGGI BADAN"] = df["TINGGI BADAN"]+50
10 df.loc[(df["TINGGI BADAN"] >= 200) & (df["TINGGI BADAN"] < 250), "TINGGI BADAN"] = df["TINGGI BADAN"]-50
11 df.loc[(df["TINGGI BADAN"] >= 250) & (df["TINGGI BADAN"] < 300), "TINGGI BADAN"] = df["TINGGI BADAN"]-100
12 df.loc[(df["TINGGI BADAN"] >= 300) & (df["TINGGI BADAN"] < 400), "TINGGI BADAN"] = df["TINGGI BADAN"]/2
13 df.loc[(df["TINGGI BADAN"] >= 400) & (df["TINGGI BADAN"] < 1000), "TINGGI BADAN"] = df["TINGGI BADAN"]/3
14 df.loc[df["TINGGI BADAN"] > 1000, "TINGGI BADAN"] = df["TINGGI BADAN"]/10
```

Kemudian, setelah dilakukan penyesuaian, diperoleh **range nilai tinggi badan yang baru adalah 140 - 199 cm**.

PREPROCESSING

2.2 Menangani Noise dan Outlier BERAT BADAN

Diketahui bahwa **range nilai berat badan dalam data adalah 1 - 90210**. Kami mengasumsikan bahwa satuan berat badan dalam data adalah dalam kg, sehingga beberapa pertimbangan yang kami lakukan untuk menangani noise dan outlier dalam berat badan sebagai berikut.

```
1 df["BERAT BADAN"] = df["BERAT BADAN"].replace(".", "", regex = True)
2 df.loc[df["BERAT BADAN"] == 1, "BERAT BADAN"] = df["BERAT BADAN"]*50
3 df.loc[df["BERAT BADAN"] == 6.77, "BERAT BADAN"] = df["BERAT BADAN"]*10
4 df.loc[(df["BERAT BADAN"] > 10) & (df["BERAT BADAN"] < 30), "BERAT BADAN"] = df["BERAT BADAN"]+50
5 df.loc[(df["BERAT BADAN"] > 150) & (df["BERAT BADAN"] <= 250), "BERAT BADAN"] = df["BERAT BADAN"]-100
6 df.loc[(df["BERAT BADAN"] > 250) & (df["BERAT BADAN"] <= 2500), "BERAT BADAN"] = df["BERAT BADAN"]/10
7 df.loc[(df["BERAT BADAN"] > 4000) & (df["BERAT BADAN"] <= 10000), "BERAT BADAN"] = df["BERAT BADAN"]/100
8 df.loc[(df["BERAT BADAN"] > 10000) & (df["BERAT BADAN"] <= 100000), "BERAT BADAN"] = df["BERAT BADAN"]/1000
```

Kemudian, setelah dilakukan penyesuaian, diperoleh **range nilai berat badan yang baru adalah 30 - 202.2 kg**.

PREPROCESSING

2.2 Menangani Noise dan Outlier BERAT BADAN

Diketahui bahwa **range nilai berat badan dalam data adalah 1 - 90210**. Kami mengasumsikan bahwa satuan berat badan dalam data adalah dalam kg, sehingga beberapa pertimbangan yang kami lakukan untuk menangani noise dan outlier dalam berat badan sebagai berikut.

```
1 df["BERAT BADAN"] = df["BERAT BADAN"].replace(".", "", regex = True)
2 df.loc[df["BERAT BADAN"] == 1, "BERAT BADAN"] = df["BERAT BADAN"]*50
3 df.loc[df["BERAT BADAN"] == 6.77, "BERAT BADAN"] = df["BERAT BADAN"]*10
4 df.loc[(df["BERAT BADAN"] > 10) & (df["BERAT BADAN"] < 30), "BERAT BADAN"] = df["BERAT BADAN"]+50
5 df.loc[(df["BERAT BADAN"] > 150) & (df["BERAT BADAN"] <= 250), "BERAT BADAN"] = df["BERAT BADAN"]-100
6 df.loc[(df["BERAT BADAN"] > 250) & (df["BERAT BADAN"] <= 2500), "BERAT BADAN"] = df["BERAT BADAN"]/10
7 df.loc[(df["BERAT BADAN"] > 4000) & (df["BERAT BADAN"] <= 10000), "BERAT BADAN"] = df["BERAT BADAN"]/100
8 df.loc[(df["BERAT BADAN"] > 10000) & (df["BERAT BADAN"] <= 100000), "BERAT BADAN"] = df["BERAT BADAN"]/1000
```

Kemudian, setelah dilakukan penyesuaian, diperoleh **range nilai berat badan yang baru adalah 30 - 202.2 kg**.

PREPROCESSING

2.3 Menangani Noise dan Outlier USIA

Diketahui bahwa **range nilai usia dalam data adalah 15 - 23091982**. Kami mengasumsikan bahwa satuan usia dalam data adalah dalam tahun, sehingga beberapa pertimbangan yang kami lakukan untuk menangani noise dan outlier dalam usia sebagai berikut.

```
1 df["USIA"] = df["USIA"].replace(".", "", regex = True)
2 df["USIA"] = df["USIA"].replace(".", "", regex = True)
3 df.loc[(df["USIA"] > 150) & (df["USIA"] <= 560), "USIA"] = df.loc[(df["USIA"] > 150) & (df["USIA"] <= 560), "USIA"].apply(lambda x: math.floor((x - x % 10) / 10))
4 df.loc[(df["USIA"] > 2000) & (df["USIA"] <= 6000), "USIA"] = df.loc[(df["USIA"] > 2000) & (df["USIA"] <= 6000), "USIA"].apply(lambda x: math.floor(x / 100))
5 df['USIA'] = df['USIA'].replace({280701:28,
6                         462525:46,
7                         10032003:20,
8                         10051996:27,
9                         12041979:44,
10                        12051987:36,
11                        21071984:39,
12                        23091982:41,
13                        8011996:27,
14                        2071967:56,
15                        10032023:32,
16                        7091992:31,
17                        8021981:42})
```

Kemudian, setelah dilakukan penyesuaian, diperoleh **range usia yang baru adalah 15 - 100 tahun**.

PREPROCESSING

3. Memeriksa Variabel Kelompok Umur, NILAI IMT, dan IMT

Berdasarkan data yang diberikan, diketahui bahwa **variabel Kelompok Umur, NILAI IMT, dan IMT diperoleh dengan menyesuaikan pada variabel USIA, TINGGI BADAN, dan BERAT BADAN**. Sebelumnya, kami telah menangani noise dan outlier dari variabel-variabel tersebut, oleh karena itu akan dilakukan penyesuaian ulang untuk variabel Kelompok Umur, NILAI IMT, dan IMT sebagai berikut.

```
1 df['Kelompok Umur'] = df['USIA'].apply(lambda usia: '15 - 24' if usia <= 24 else  
2                                         '25 - 34' if usia <= 34 else  
3                                         '35 - 44' if usia <= 44 else  
4                                         '45 - 54' if usia <= 54 else  
5                                         '55 - 64' if usia <= 64 else  
6                                         '> 65')  
  
1 df['NILAI IMT'] = df['BERAT BADAN'] / ((df['TINGGI BADAN'] / 100)**2)  
2 sorted(df["NILAI IMT"].unique().tolist())  
  
1 df['IMT'] = df['NILAI IMT'].apply(lambda value: 'Sangat Kurus' if value < 17 else  
2                                         'Kurus' if value < 18.5 else  
3                                         'Normal' if value < 25 else  
4                                         'Gemuk' if value <= 27 else  
5                                         'Obesitas')
```

PREPROCESSING

4. Memeriksa Missing Value

Diperoleh total missing value sebagai berikut:

```
1 # Memeriksa Total Missing Value
2 df.isnull().sum()[df.isnull().sum() > 0]
```

```
Jika Ya berapa jumlah gigi saudara yang Karies           223744
Jika Ya, Berapa Batang Saudara merokok dalam sehari?      152543
Berapa lama saudara menggosok gigi ?                      1779
Berapa jarak tempat sampah dengan rumah saudara ?        41649
dtype: int64
```

```
1 # Memeriksa Persentase Missing Value
2 (df.isnull().sum()[df.isnull().sum() > 0] / len(df)) * 100
```

```
Jika Ya berapa jumlah gigi saudara yang Karies           77.206083
Jika Ya, Berapa Batang Saudara merokok dalam sehari?      52.637154
Berapa lama saudara menggosok gigi ?                      0.613870
Berapa jarak tempat sampah dengan rumah saudara ?        14.371586
dtype: float64
```

Perhatikan bahwa missing value pada variabel “Jika Ya berapa jumlah gigi saudara yang Karies” dan “Jika Ya, Berapa Batang Saudara merokok dalam sehari?” artinya **individu tersebut tidak memiliki gigi yang karies dan/atau tidak merokok.** Sehingga, akan diimputasi dengan nol.

PREPROCESSING

4. Memeriksa Missing Value

Diperoleh total missing value sebagai berikut:

```
1 # Memeriksa Total Missing Value
2 df.isnull().sum()[df.isnull().sum() > 0]
```

```
Jika Ya berapa jumlah gigi saudara yang Karies           223744
Jika Ya, Berapa Batang Saudara merokok dalam sehari?      152543
Berapa lama saudara menggosok gigi ?                      1779
Berapa jarak tempat sampah dengan rumah saudara ?        41649
dtype: int64
```

```
1 # Memeriksa Persentase Missing Value
2 (df.isnull().sum()[df.isnull().sum() > 0] / len(df)) * 100
```

```
Jika Ya berapa jumlah gigi saudara yang Karies           77.206083
Jika Ya, Berapa Batang Saudara merokok dalam sehari?      52.637154
Berapa lama saudara menggosok gigi ?                      0.613870
Berapa jarak tempat sampah dengan rumah saudara ?        14.371586
dtype: float64
```

Perhatikan bahwa missing value pada variabel “Jika Ya berapa jumlah gigi saudara yang Karies” dan “Jika Ya, Berapa Batang Saudara merokok dalam sehari?” artinya **individu tersebut tidak memiliki gigi yang karies dan/atau tidak merokok.** Sehingga, akan diimputasi dengan nol.

PREPROCESSING

4. Memeriksa Missing Value

Sedangkan, untuk menangani missing value pada variabel “Berapa lama saudara menggosok gigi ?” dan “Berapa jarak tempat sampah dengan rumah saudara ?” akan diperiksa terlebih dahulu baris-barisnya. Setelah diperiksa, ternyata **terdapat missing value dikarenakan memang individu tersebut tidak membersihkan/menggosok gigi dan/atau tidak terdapat tempat sampah di sekitar rumah**. Oleh karena itu, untuk menangani missing value pada dua variabel tersebut juga akan diimputasi dengan nol.

5. Memeriksa Duplikasi Data

```
2 df.duplicated().sum()  
0
```

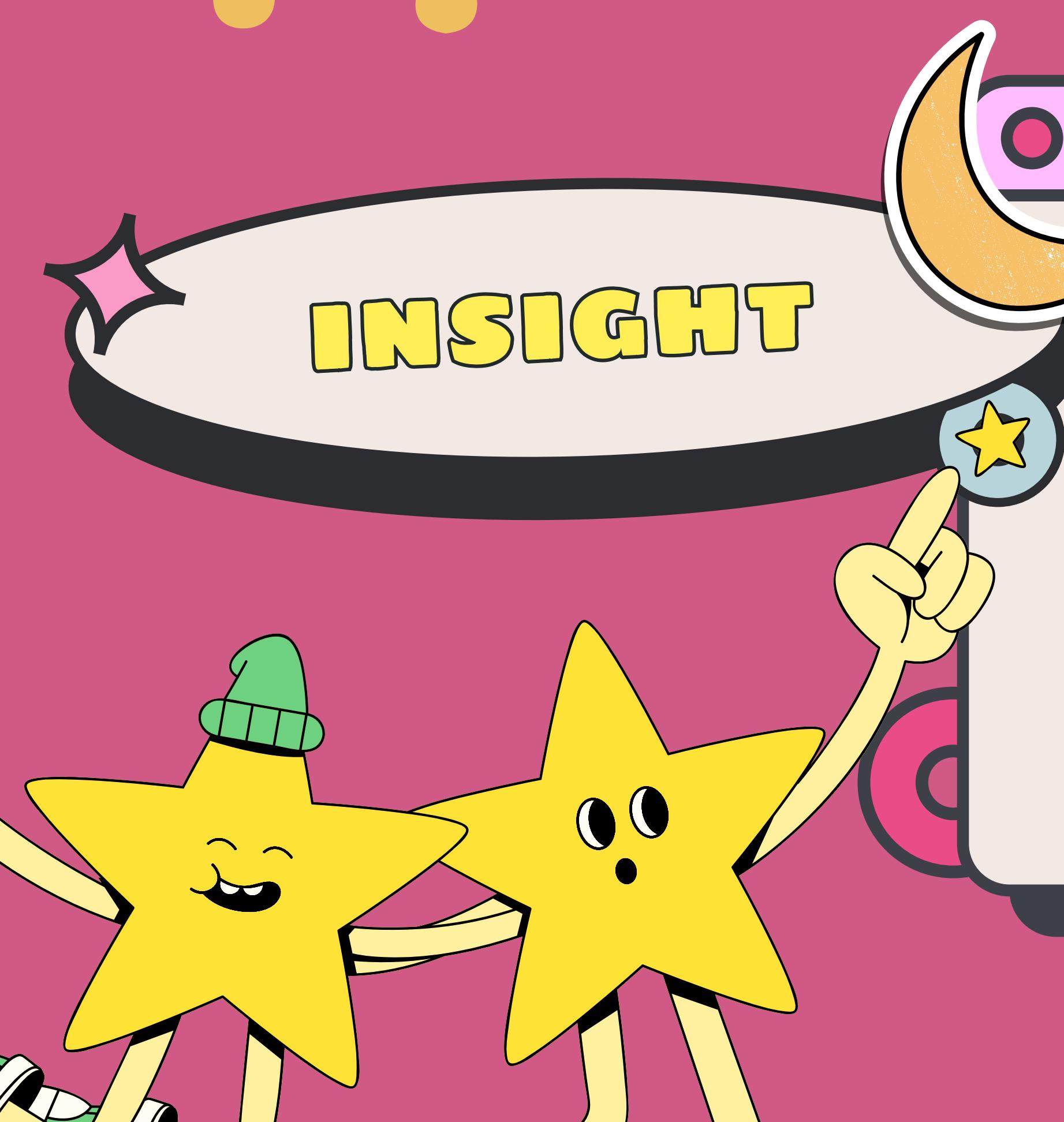
Tidak ditemukan duplikasi dalam data.

PREPROCESSING

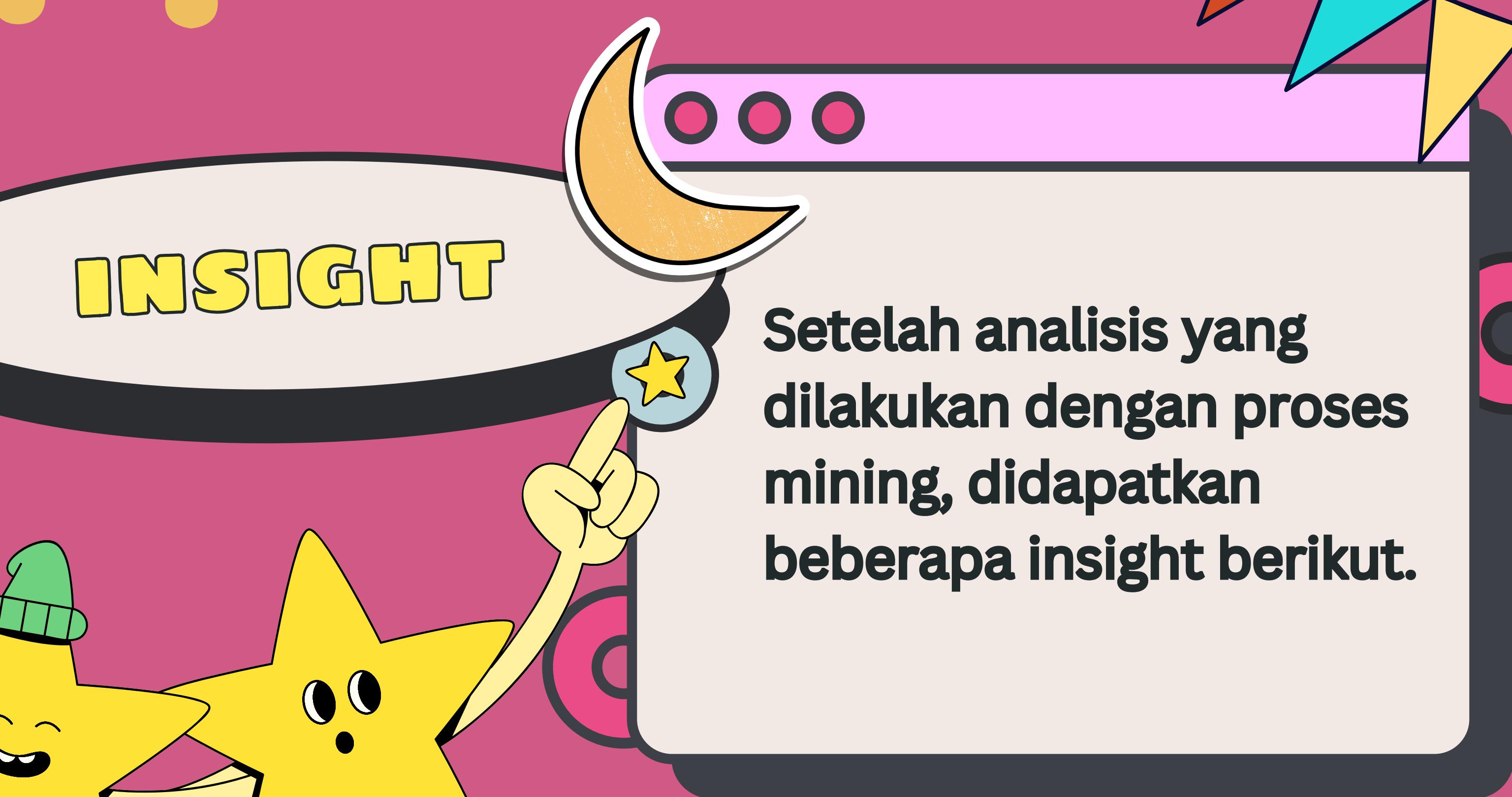
6. Korelasi



Berdasarkan *heatmap* di samping, dapat dilihat bahwa **terdapat korelasi positif antara NILAI IMT dengan USIA, TINGGI BADAN, BERAT BADAN, dan LINGKAR PINGGANG**. Artinya, seiring bertambahnya usia, tinggi badan, berat badan, dan lingkar pinggang, maka nilai IMT juga cenderung meningkat.



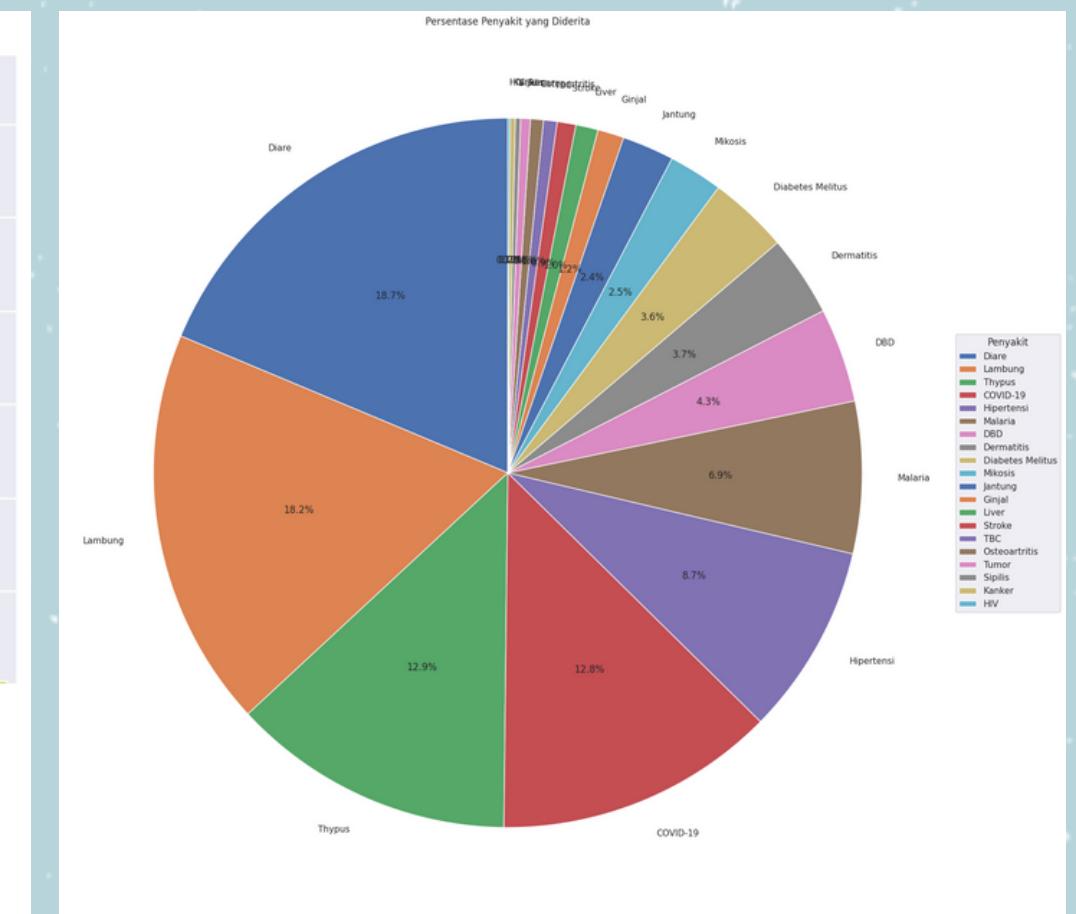
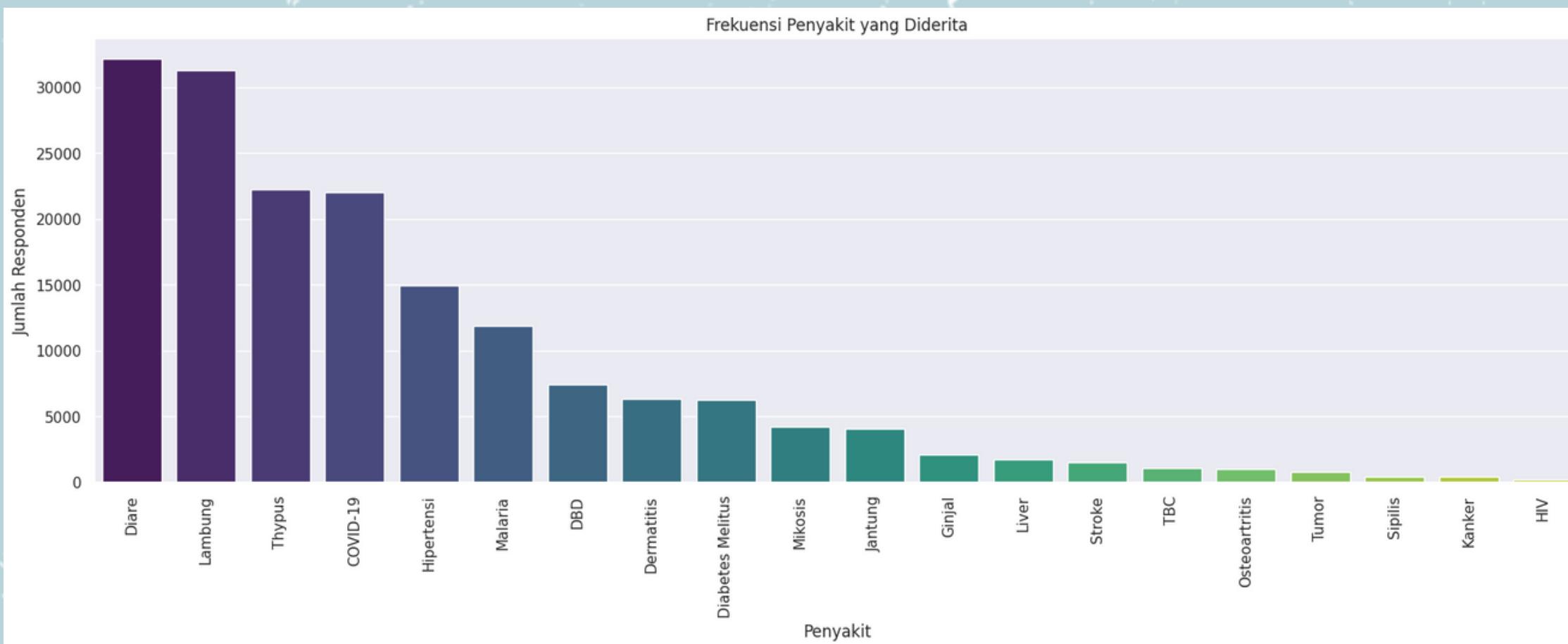
INSIGHT



Setelah analisis yang dilakukan dengan proses mining, didapatkan beberapa insight berikut.

Penyakit Umum

Penyakit apa saja yang paling banyak diderita?

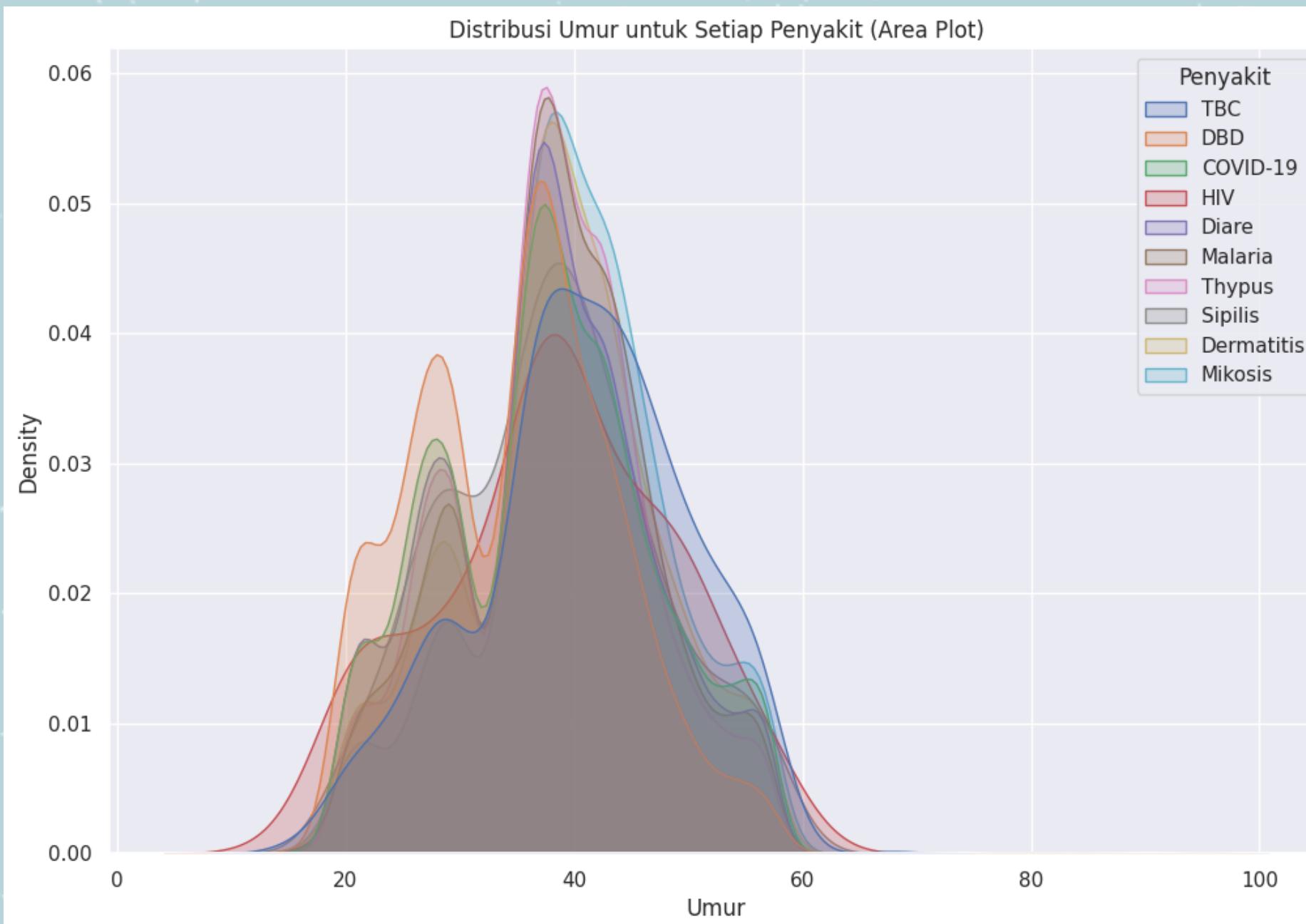


Berdasarkan hasil analisis kami dari data penyakit, didapatkan penyakit-penyakit yang paling banyak diderita, berikut sepuluh di antaranya:

1. Diare
2. Lambung
3. Thypus
4. COVID-19
5. Hipertensi
6. Malaria
7. DBD
8. Dermatitis
9. Mikosis
10. Jantung

Distribusi Umur

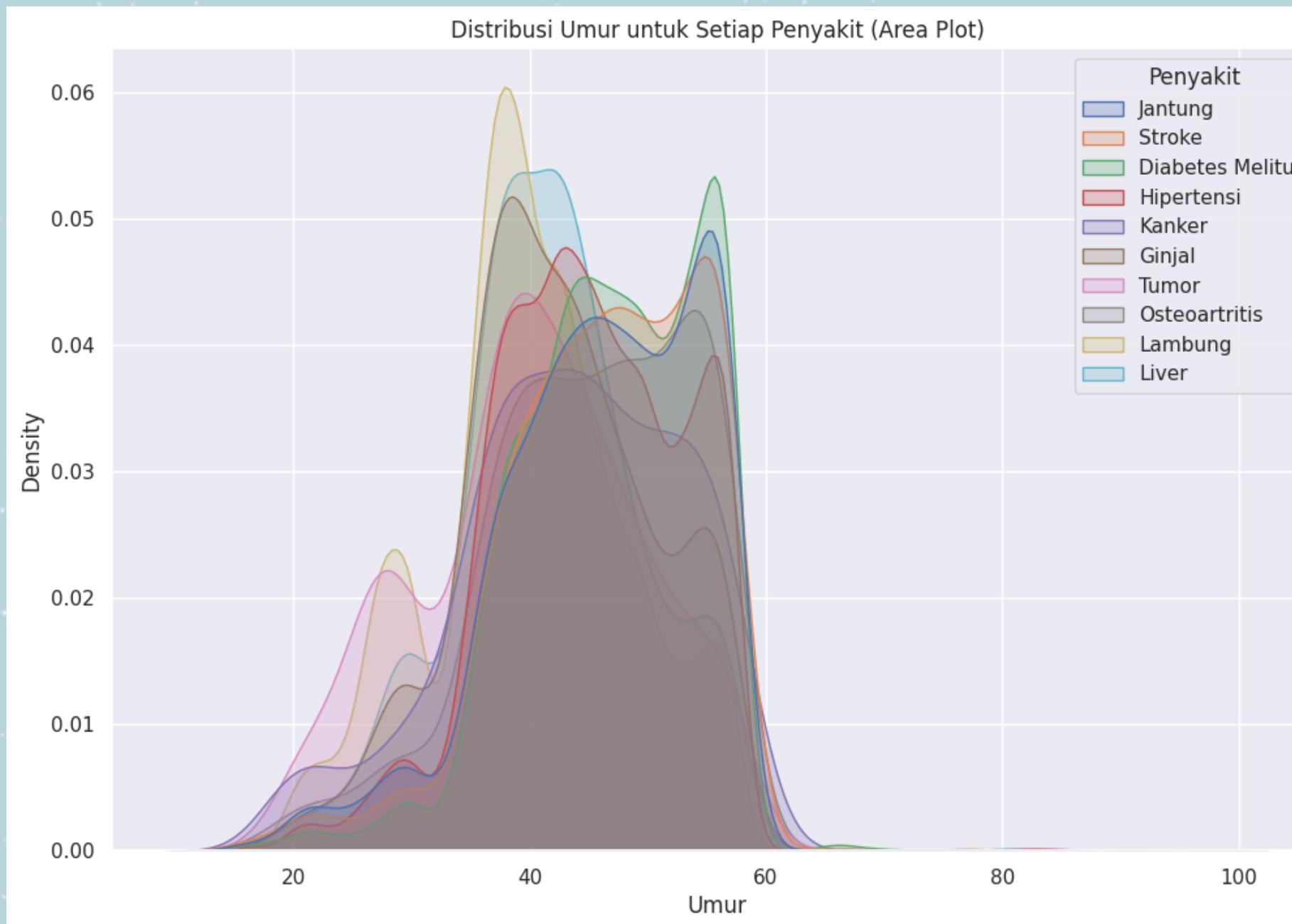
Penyakit Infeksi: TBC, DBD, COVID-19, HIV, Diare, Malaria, Thypus, Sipilis, Dermatitis, Mikosis



Berdasarkan visualisasi data, terlihat bahwa penyakit yang umumnya disebabkan oleh infeksi memiliki **puncak distribusi umur responden di sekitar rentang usia 35-55 tahun**. Tidak banyak lansia yang berumur lebih dari 50 tahun mengidap penyakit kelompok ini. Hal ini mungkin terjadi karena biasanya lansia tidak banyak berinteraksi dengan banyak orang sehingga minim resiko terkena infeksi.

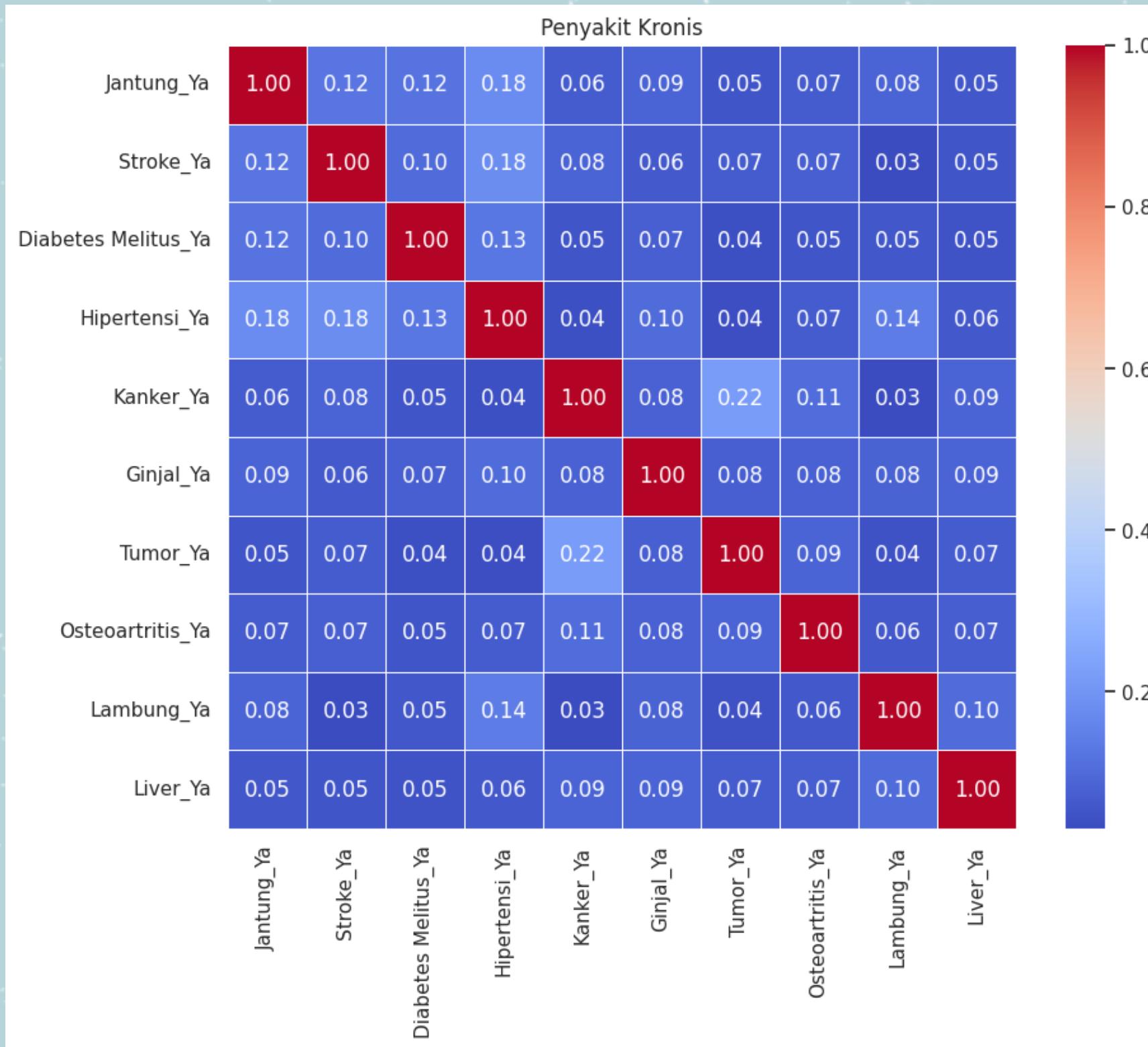
Distribusi Umur

Penyakit Kronis: Jantung, Stroke, Diabetes Melitus, Hipertensi, Kanker, Ginjal, Tumor, Osteoarthritis, Lambung, Liver.



Berbeda dengan kelompok penyakit sebelumnya, penyakit yang umumnya bersifat kronis memiliki **puncak distribusi umur responden di sekitar rentang usia 35 - 45 dan 55 - 65 tahun**. Hal ini mengindikasikan bahwa baik individu yang muda maupun tua penting untuk pemantauan kesehatan dan pencegahan penyakit yang termasuk dalam kelompok.

Korelasi Penyakit Kronis



Penyakit **Tumor** dan **Kanker** memiliki korelasi yang cukup kuat dibandingkan penyakit kronis lainnya, **Jantung** dan **Hipertensi** memiliki korelasi yang cukup kuat sama dengan **Stroke** dan **Hipertensi**. Maka, akan dicek frekuensi responden yang terkena penyakit yang memiliki korelasi kuat.

Korelasi Penyakit Kronis

		Hipertensi	Tidak	Ya
		Jantung	Stroke	
Jantung	Tidak	2718	34	12755
	Ya	500	649	
Ya	Tidak	2441	1306	
	Ya	74	242	

		Kanker	Tidak	Ya
		Tumor		
Tumor	Tidak	2887	69	273
	Ya	638		121

- Terdapat **242** orang yang terdiagnosa **3 penyakit kronis sekaligus**, yaitu **Jantung, Stroke, dan Hipertensi**. Orang-orang ini butuh perhatian penuh agar kondisinya tidak memburuk.
- Terdapat **1306** orang yang terdiagnosa penyakit **Jantung dan Hipertensi**. Orang-orang ini butuh perhatian khusus karena **beresiko terkena Stroke**.

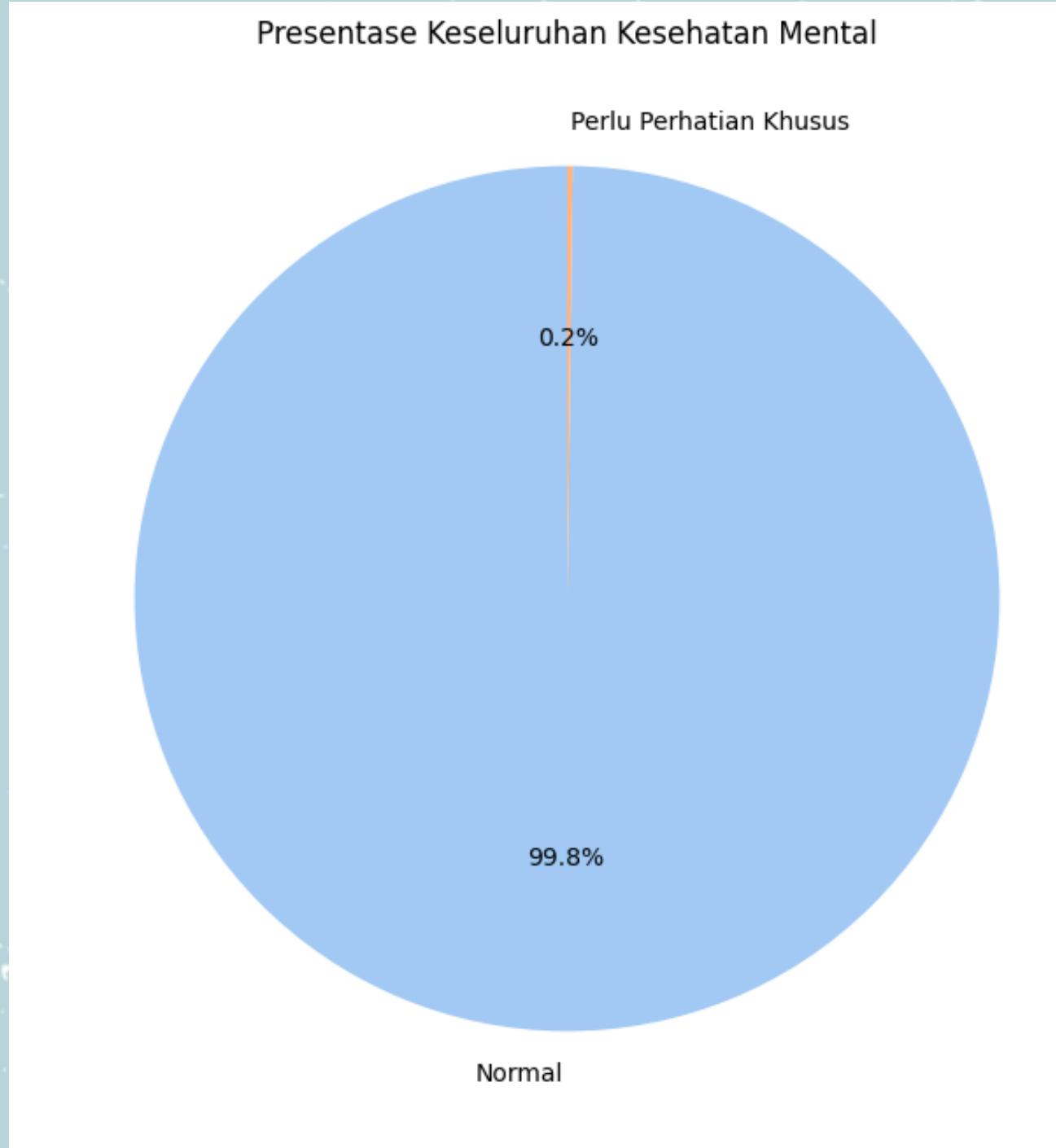
- Terdapat **121** orang yang terdiagnosa **Tumor dan Kanker sekaligus**
- Terdapat **638** orang yang terdiagnosa **Tumor**, tetapi **tidak terdiagnosa Kanker**. Orang-orang ini harus memeriksakan dirinya secara berkala jika belum mendapatkan tindakan sebelumnya karena terdapat resiko tumor berkembang menjadi sel kanker.

Korelasi Penyakit Kronis

		Konsumsi air mineral		
		2 Liter	Kurang dari 2 Liter	Lebih dari 2 liter
		Ginjal?		
Diabetes?	Tidak	Tidak	81223	26900
		Ya	493	410
Ginjal?	Ya	Tidak	1660	730
		Ya	79	79
				145

- Terdapat **1660** orang yang konsumsi **air mineral 2 liter per hari** terdiagnosa Diabetes Melitus, tidak terdiagnosa penyakit Ginjal.
- Terdapat **3554** orang yang terdiagnosa Diabetes Melitus, tidak terdiagnosa penyakit Ginjal dan konsumsi air mineral **lebih dari 2 liter per hari**. Orang-orang ini perlu memeriksakan fungsi Ginjalnya, jika ada **penurunan fungsi ginjal maka konsumsi air mineral perlu dibatasi**.
- Terdapat **730** orang yang terdiagnosa Diabetes Melitus, tidak terdiagnosa penyakit Ginjal, dan mengkonsumsi **kurang dari 2 liter air mineral per hari**. Orang-orang ini perlu meningkatkan jumlah konsumsi air mineral agar tidak dehidrasi karena dehidrasi dapat berpengaruh terhadap kadar gula dalam darah.

KESEHATAN MENTAL



Presentase Kesehatan Mental

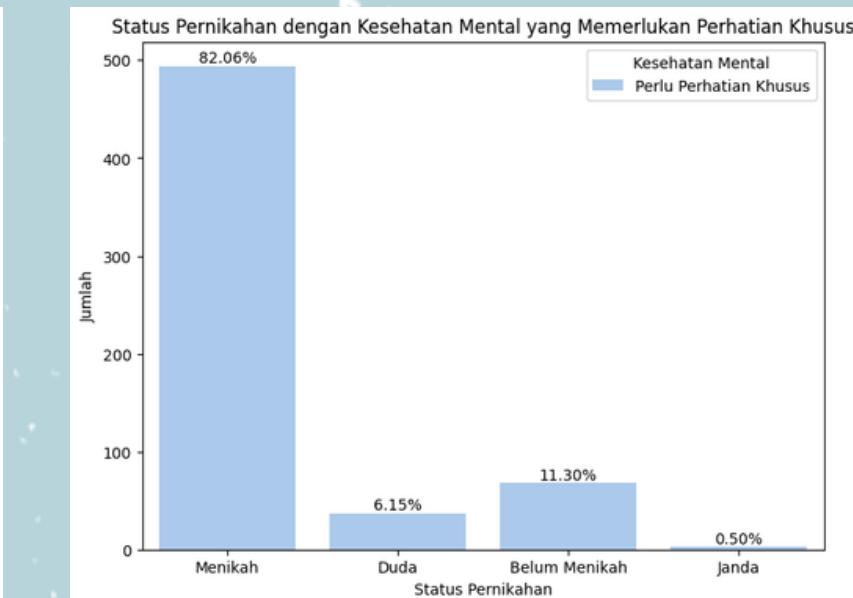
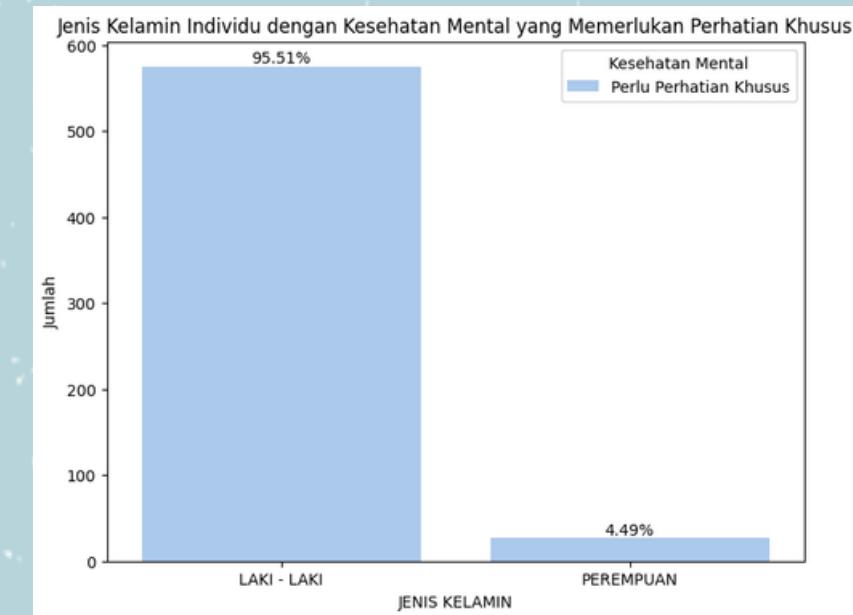
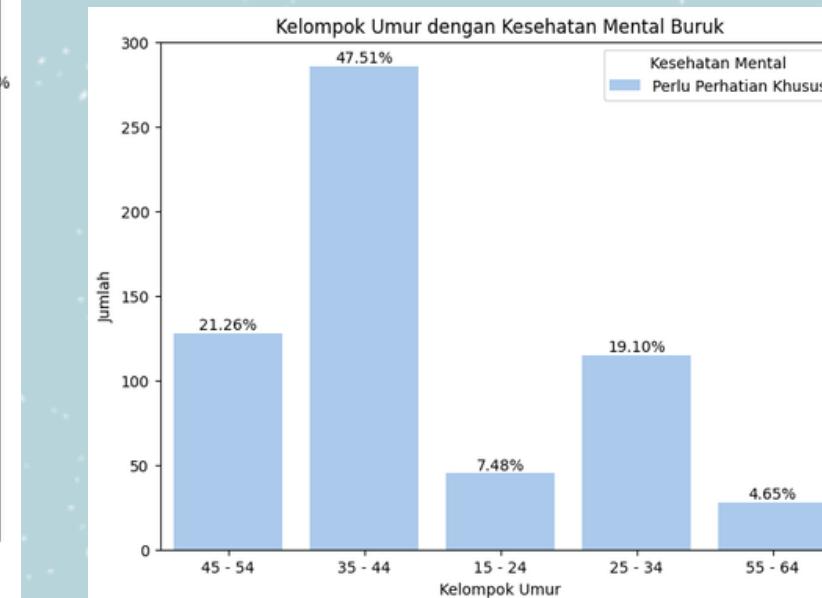
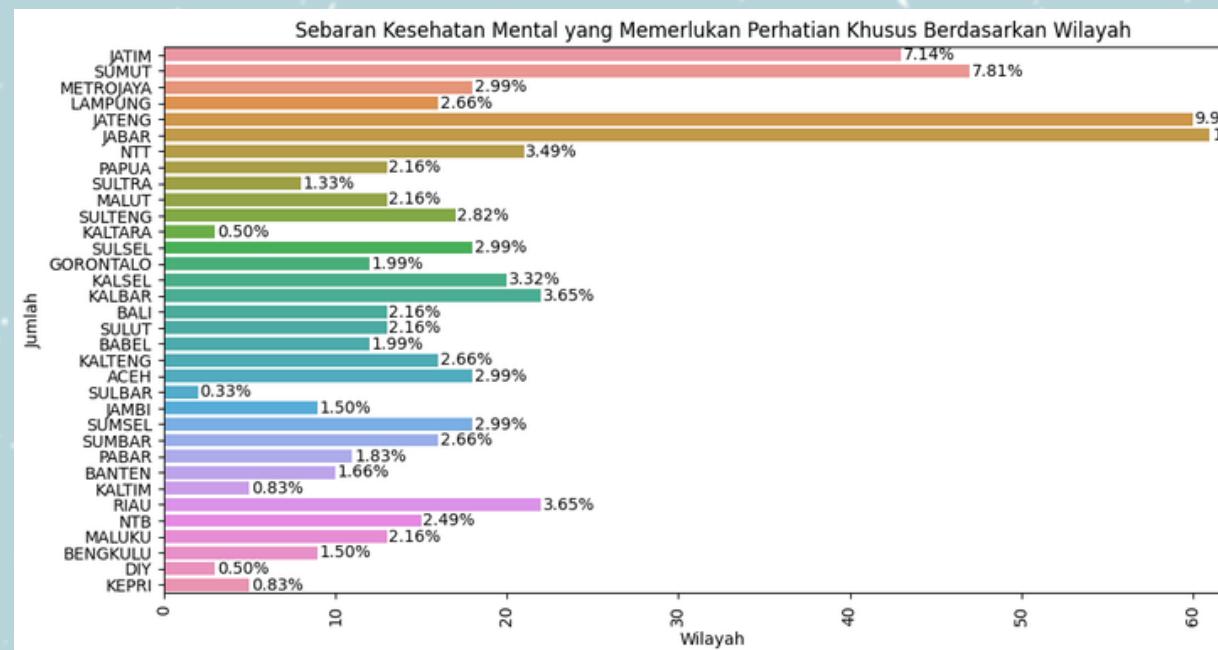
Dengan proses mining yang dilakukan terhadap variabel-variabel berisi pertanyaan tentang kesehatan mental, dapat dilihat bahwa dari 289.801 individu, diklasifikasikan **99.8% individu mempunyai mental normal dan 0.2% lainnya butuh perhatian khusus terhadap kesehatan mentalnya.** Jika dihitung, terdapat 602 individu yang memerlukan perhatian khusus terhadap mentalnya.

Kesehatan Mental	Jumlah
Normal	289.199
Perlu Perhatian Khusus	602

KESEHATAN MENTAL

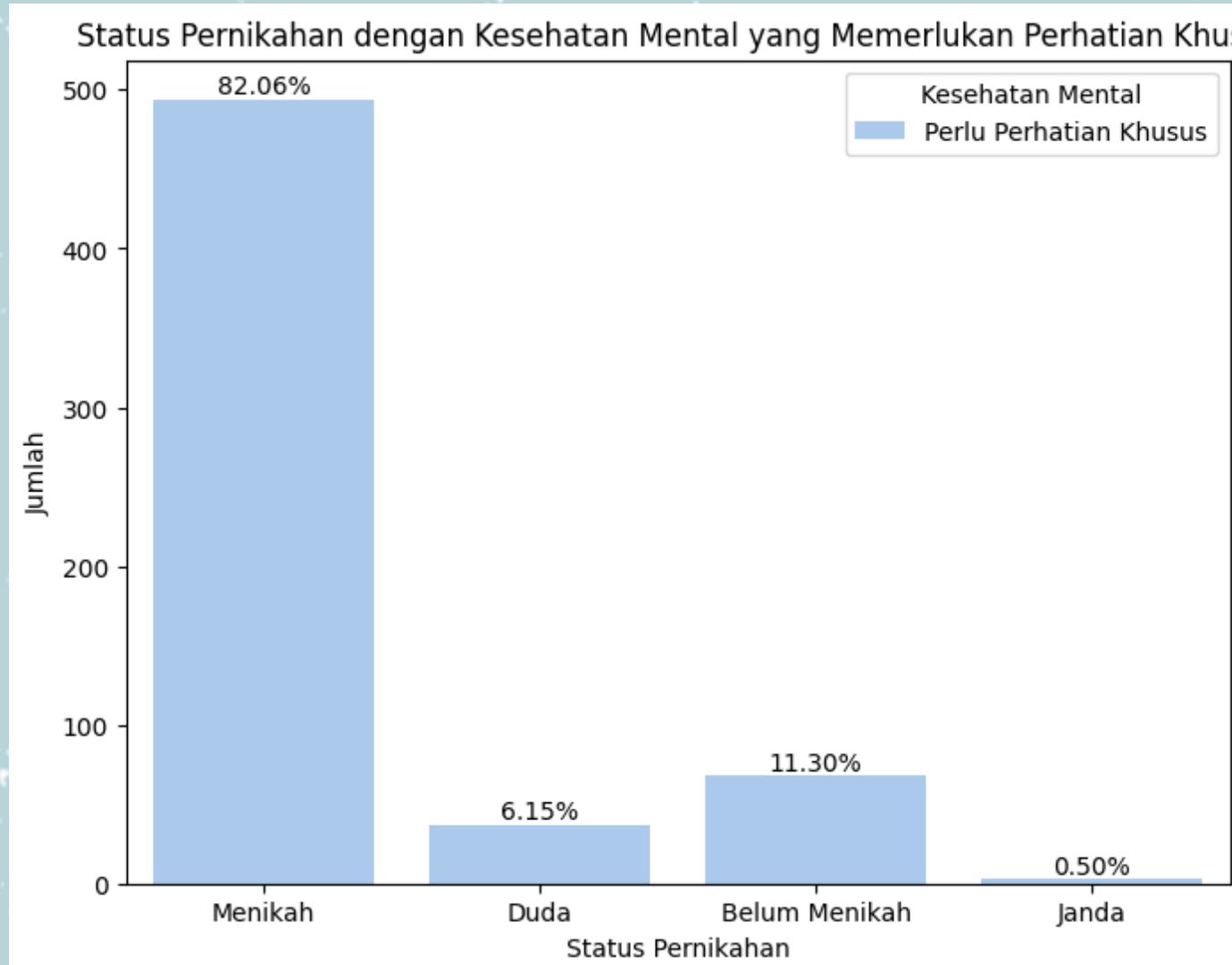
Demografi Individu dengan Kesehatan Mental Kurang Baik

Dari 602 individu dengan kesehatan mental yang butuh perhatian khusus, **10.13%nya berada di wilayah Jawa Barat, 95.51% berjenis kelamin laki-laki, 82.06% berstatus menikah, dan 47.51% berada di kelompok umur 35-44 tahun.**



KESEHATAN MENTAL

Apakah Status Perkawinan Berhubungan dengan Kesehatan Mental yang Kurang Baik?



Dari 602 individu dengan kesehatan mental yang butuh perhatian khusus, **82.06%** di antaranya adalah individu yang sudah menikah. Angka ini sangat banyak sehingga menarik perhatian kami untuk melihat apakah status pernikahan berhubungan dengan kesehatan mental yang perlu perhatian khusus. Oleh karena itu, hubungan ini akan dianalisis dengan regresi logistik.

KESEHATAN MENTAL

Apakah Status Perkawinan Berhubungan dengan Kesehatan Mental yang Kurang Baik?

Logit Regression Results						
Dep. Variable:	Kesehatan Mental_Perlu Perhatian Khusus	No. Observations:	289801			
Model:	Logit	Df Residuals:	289797			
Method:	MLE	Df Model:	3			
Date:	Thu, 21 Dec 2023	Pseudo R-squ.:	inf			
Time:	13:04:38	Log-Likelihood:	-inf			
converged:	True	LL-Null:	0.0000			
Covariance Type:	nonrobust	LLR p-value:	1.000			
coef	std err	z	P> z	[0.025	0.975]	
const	-6.5953	0.121	-54.349	0.000	-6.833	-6.357
Status Perkawinan_Menikah	0.4390	0.129	3.392	0.001	0.185	0.693
Status Perkawinan_Janda	0.3695	0.591	0.626	0.532	-0.788	1.527
Status Perkawinan_Duda	1.7058	0.205	8.328	0.000	1.304	2.107

y = 1, jika kesehatan mental perlu diperhatikan; 0, lainnya

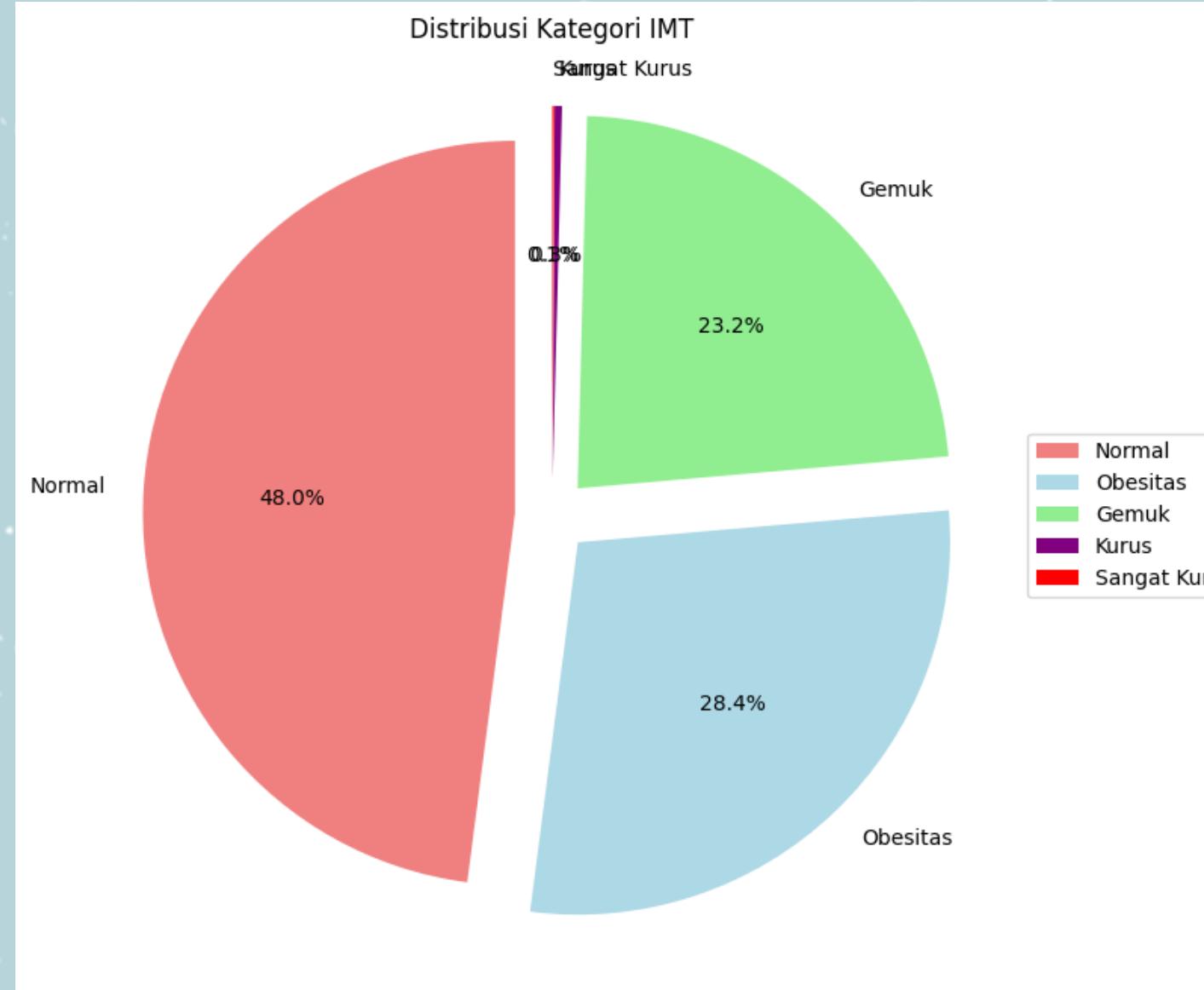
x1 = 1, jika Sudah Menikah; 0, lainnya

x2 = 1, jika Janda; 0 lainnya

x3 = 1, jika Duda; 0 lainnya

Dari hasil analisis regresi logistik didapatkan nilai koefisien variabel-variabel yang telah menikah semuanya positif. Sehingga, jika **individu sudah/pernah menikah, maka peluang individu tersebut mengalami gangguan kesehatan mental lebih tinggi daripada individu yang belum menikah.** contoh interpretasi: $\exp(0.439) = 1.55$, individu yang sudah menikah memiliki peluang mengalami gangguan mental 1.55 kali lipat daripada yang belum menikah.

OBESITAS



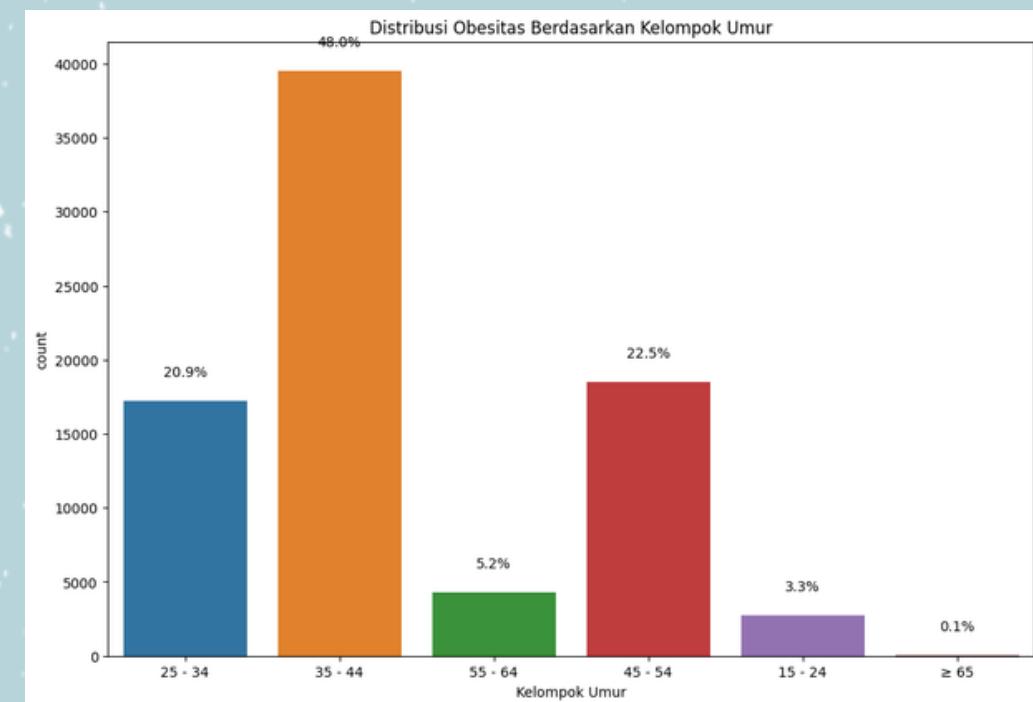
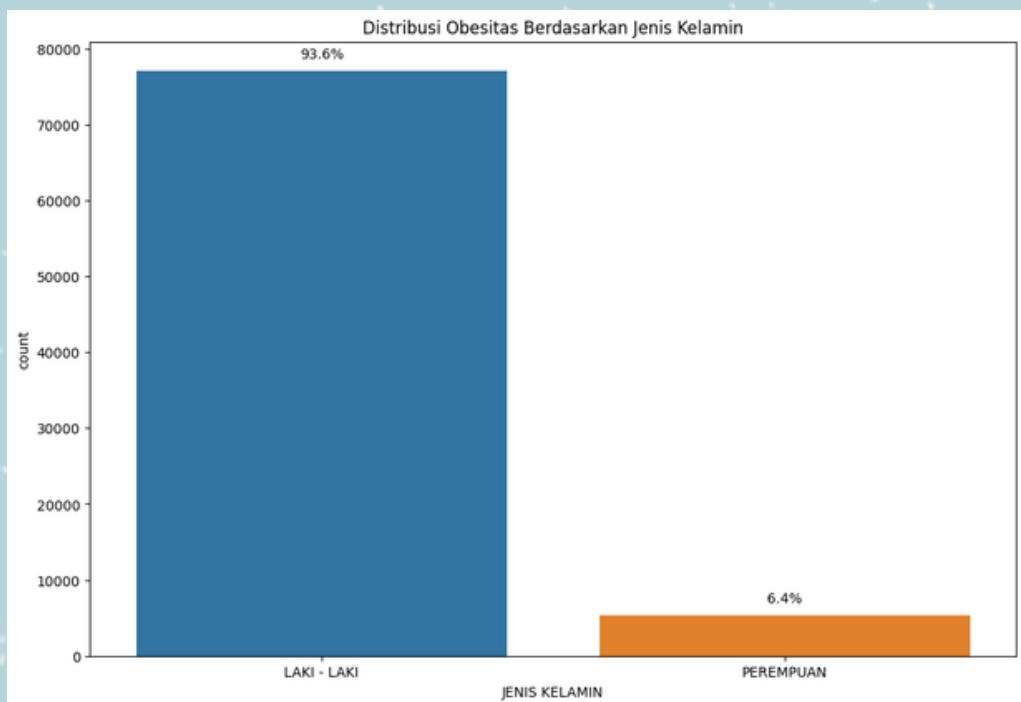
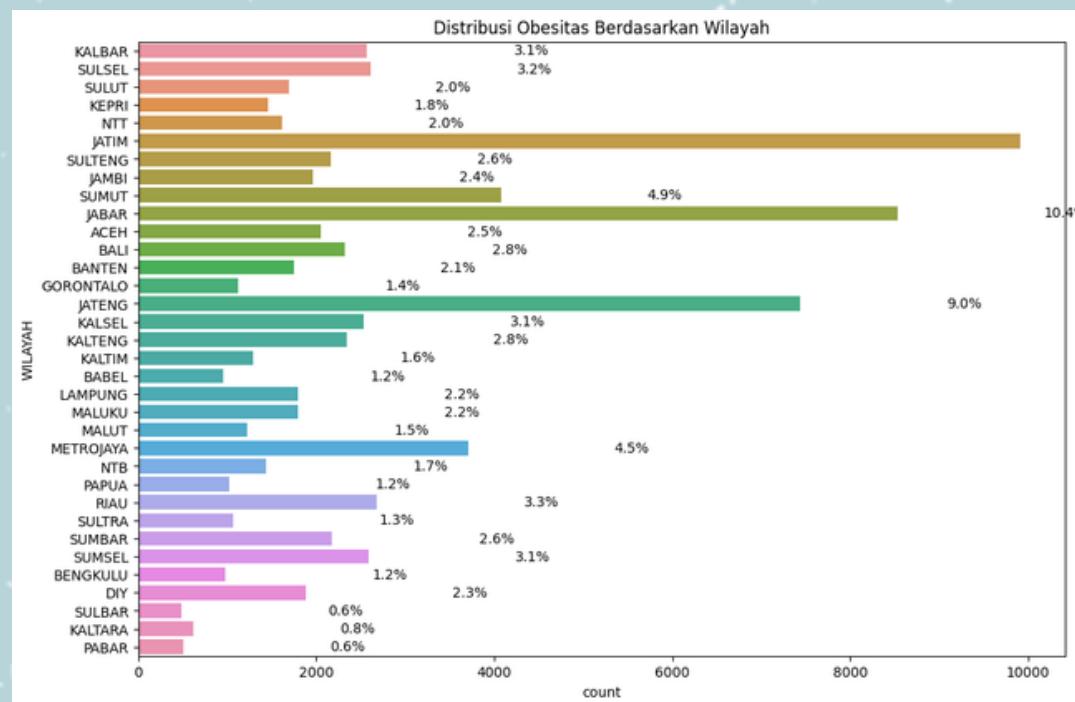
Presentase Katgeori IMT

Kategori IMT	Jumlah	Persentase
Normal	138.961	48,0%
Obesitas	82.360	28,4%
Gemuk	67.319	23,2%
Kurus	919	0,3%
Sangat Kurus	242	0,1%

Dapat dilihat bahwa dari 289.800 individu, **28,4%** atau **sekitar 82.300 orang adalah penderita obesitas** dan **23,2%** atau **sekitar 67.300 orang berbadan gemuk**. Untuk lebih detailnya dapat dilihat pada tabel di atas.

OBESITAS

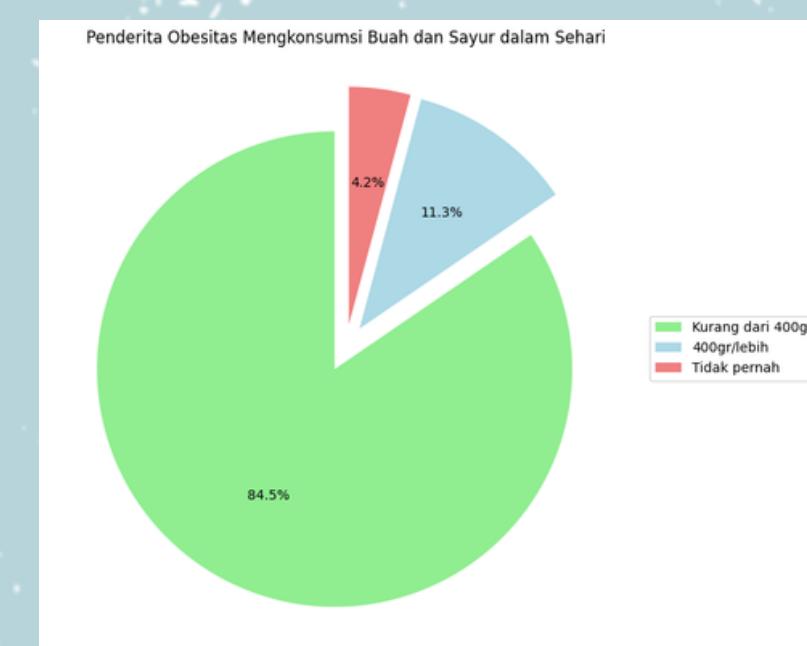
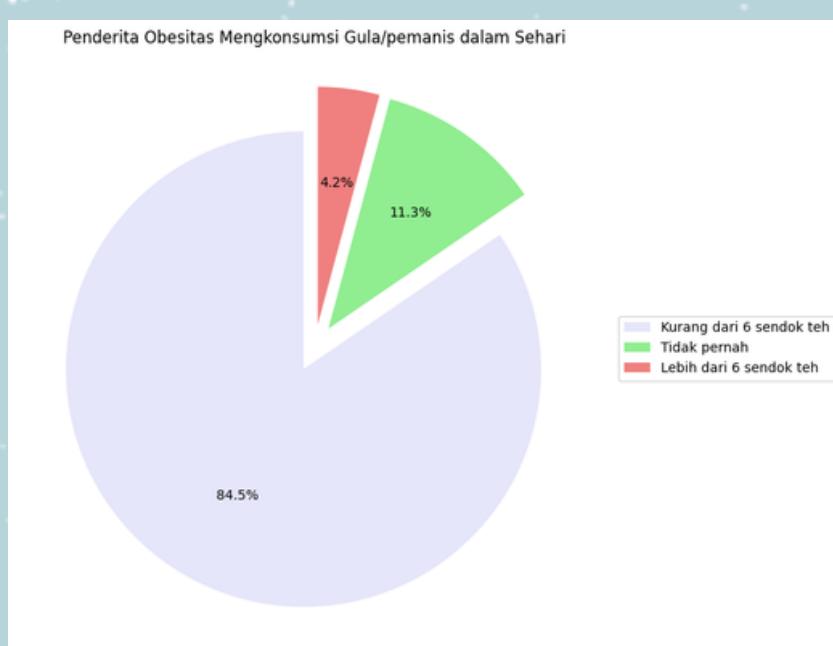
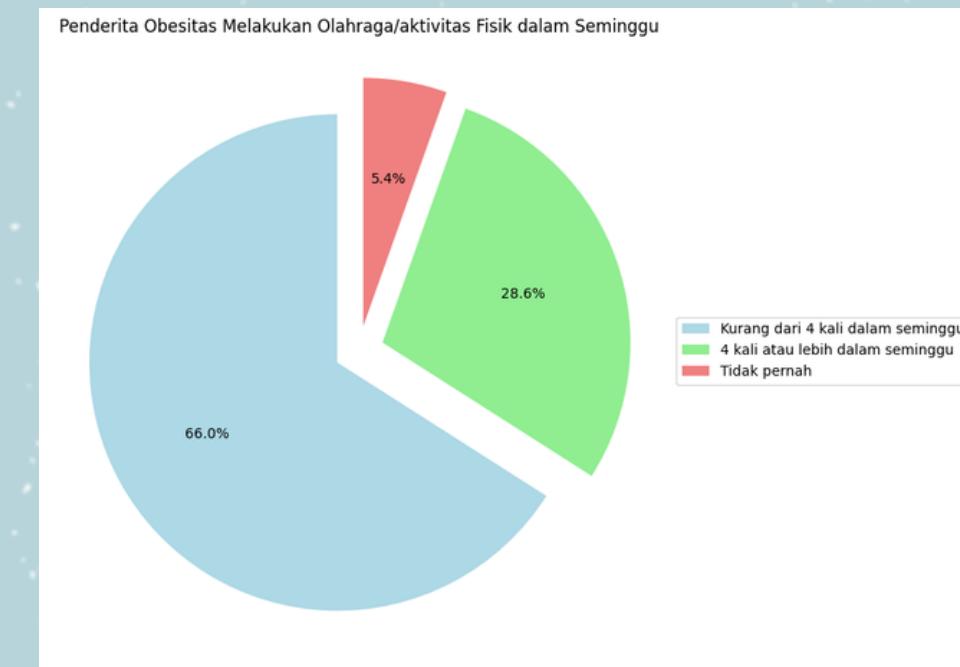
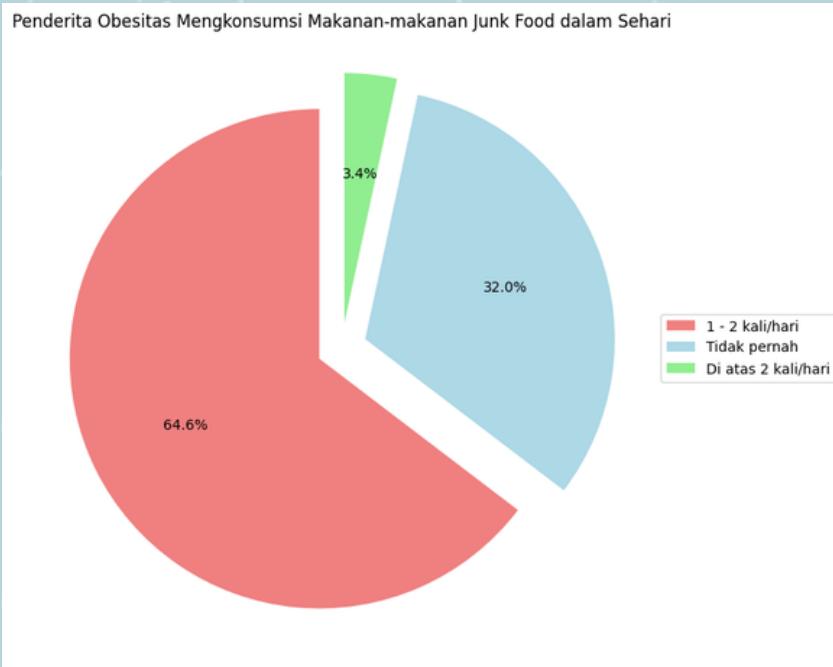
Demografi Penderita Obesitas



Dapat dilihat bahwa dari 82.360 individu penderita obesitas, **12% berada di wilayah Jawa Timur, 93,6% berjenis kelamin laki-laki, dan 48,0% berada di kelompok umur 35 - 44 tahun.** Dapat diketahui juga bahwa **tiga wilayah dengan jumlah penderita obesitas terbanyak berada di pulau Jawa**, yaitu Jawa Timur, Jawa Barat, dan Jawa Tengah.

OBESITAS

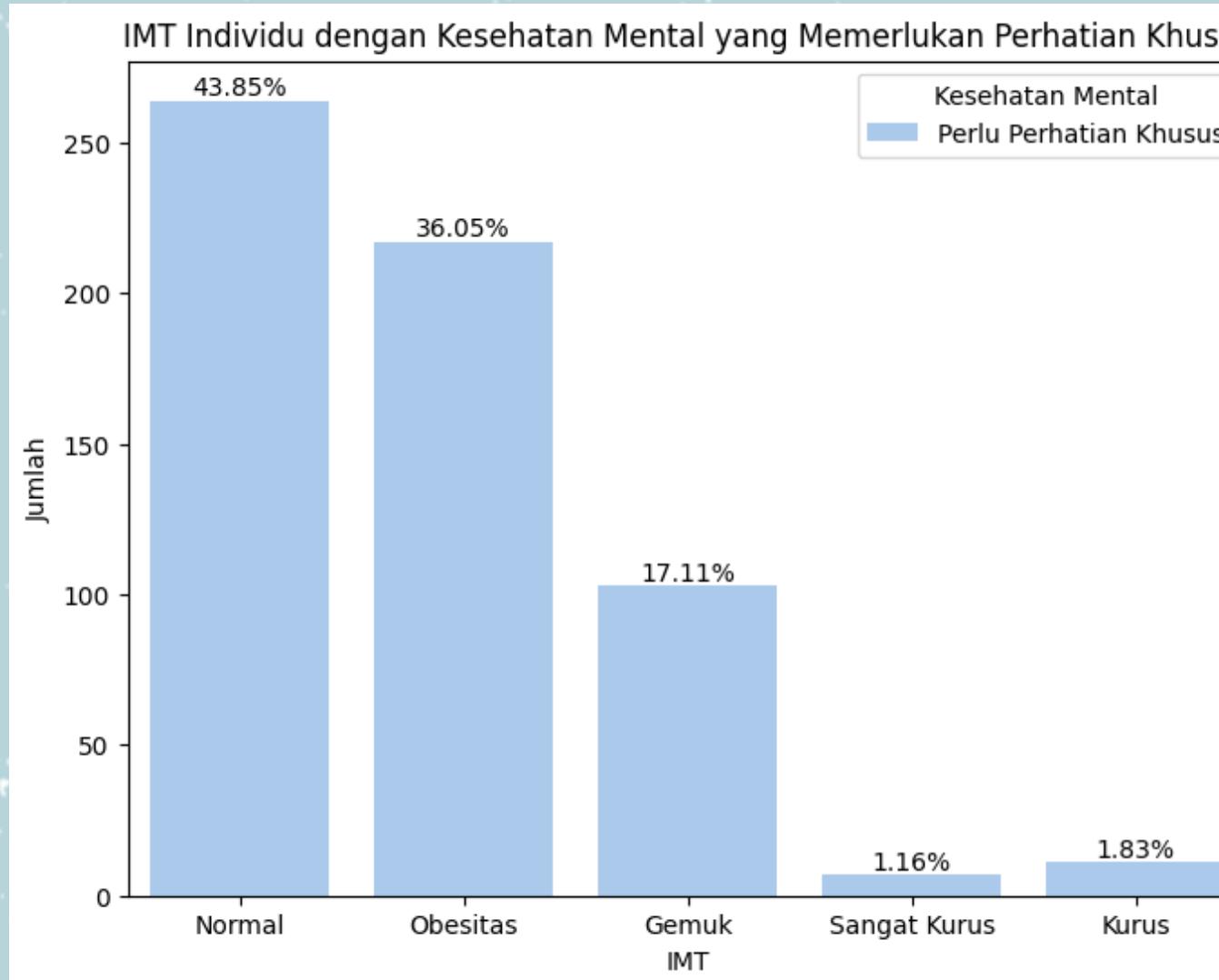
Pola Konsumsi dan Aktivitas Fisik



Dapat dilihat bahwa walaupun penderita obesitas kebanyakan mengonsumsi gula dalam batas normal, kebanyakan penderita obesitas tsb mempunyai pola konsumsi yang **kurang sehat**, yaitu mengonsumsi junk food sebanyak 1-2 kali tetapi mengonsumsi sayuran dan buah kurang dari 400 gram dalam sehari. Ajuran pola konsumsi yang sehat adalah mengonsumsi junk food maksimal satu kali dalam seminggu dan 400 gr dalam sehari. Kebanyakan penderita obesitas juga melakukan aktivitas fisik/olahraga kurang dari 4 kali dalam seminggu.

KESEHATAN MENTAL

Apakah Kategori IMT Obesitas Berhubungan dengan Kesehatan Mental yang Kurang Baik?



Dari 602 individu dengan kesehatan mental yang butuh perhatian khusus, **36.05%** di antaranya adalah individu dengan IMT Obesitas. Angka ini merupakan sepertiga dari jumlah individu yang butuh perhatian khusus. Karena jumlahnya cukup banyak, akan dilihat apakah benar terdapat hubungan antara kedua keadaan tersebut dengan regresi logistik.

KESEHATAN MENTAL

Apakah Kategori IMT Obesitas Berhubungan dengan Kesehatan Mental yang Kurang Baik?

Logit Regression Results						
Dep. Variable:	Kesehatan Mental_Perlu Perhatian Khusus	No. Observations:	289801			
Model:		Logit	Df Residuals:	289799		
Method:		MLE	Df Model:	1		
Date:	Thu, 21 Dec 2023	Pseudo R-squ.:	inf			
Time:	13:04:04	Log-Likelihood:	-inf			
converged:		True LL-Null:	0.0000			
Covariance Type:		nonrobust LLR p-value:	1.000			
coef	std err	z	P> z	[0.025	0.975]	
const	-6.2875	0.051	-123.255	0.000	-6.387	-6.188
IMT_Obesitas	0.3512	0.085	4.132	0.000	0.185	0.518

y = 1, jika kesehatan mental perlu diperhatikan; 0, lainnya
x = 1, jika IMT Obesitas; 0, lainnya

Dari hasil analisis regresi logistik didapatkan nilai koefisien variabel IMT jika individu obesitas adalah 0.3512 yang signifikan karena p-value < 0.05. Sehingga dapat diartikan individu dengan IMT Obesitas akan mengalami gangguan kesehatan mental sebesar $\exp(0.3512) = 1.42$ kali lipat daripada individu dengan IMT bukan obesitas. Oleh karena itu, dapat disimpulkan keadaan **IMT Obesitas ini dapat memicu seseorang untuk memiliki gangguan kesehatan mental**.

LAMPIRAN

Link Google Colab: [Preprocessing & Analisis Data](#)

Dataset dalam excel: [Data Project UAS DatMin & BI](#)



Terima Kasih