



FINAL PROJECT DATA MINING

Andri Naufal Al'afif - 2106725085

Kalisha Rahma Firza - 2106725053

Ricky - 2106724952

Selvira Salsabila - 2106656232

Sabrina Zahra Salsabila - 2106726895

IMPORT PACKAGE DAN DATA

```
import warnings; warnings.simplefilter('ignore')
import pandas as pd, seaborn as sns, matplotlib.pyplot as plt, numpy as np
import gdown
import random
import plotly.express as px
from scipy import stats
from scipy.stats import shapiro, anderson
from scipy.stats import mannwhitneyu
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error
import statsmodels.api as sm
from statsmodels.formula.api import ols
import statsmodels.formula.api as smf
```

```
[ ] #Import Data Set
!gdown 1RSkEOn6UldxS1VEC7-2lPLhvZ_MtfgHc
data = pd.read_csv("/content/data_kesehatan-Mental-SRQ-2023-Filtered.csv", sep = ";")
```

```
Downloading...
From: https://drive.google.com/uc?id=1RSkEOn6UldxS1VEC7-2lPLhvZ\_MtfgHc
To: /content/data_kesehatan-Mental-SRQ-2023-Filtered.csv
100% 194M/194M [00:02<00:00, 92.8MB/s]
```

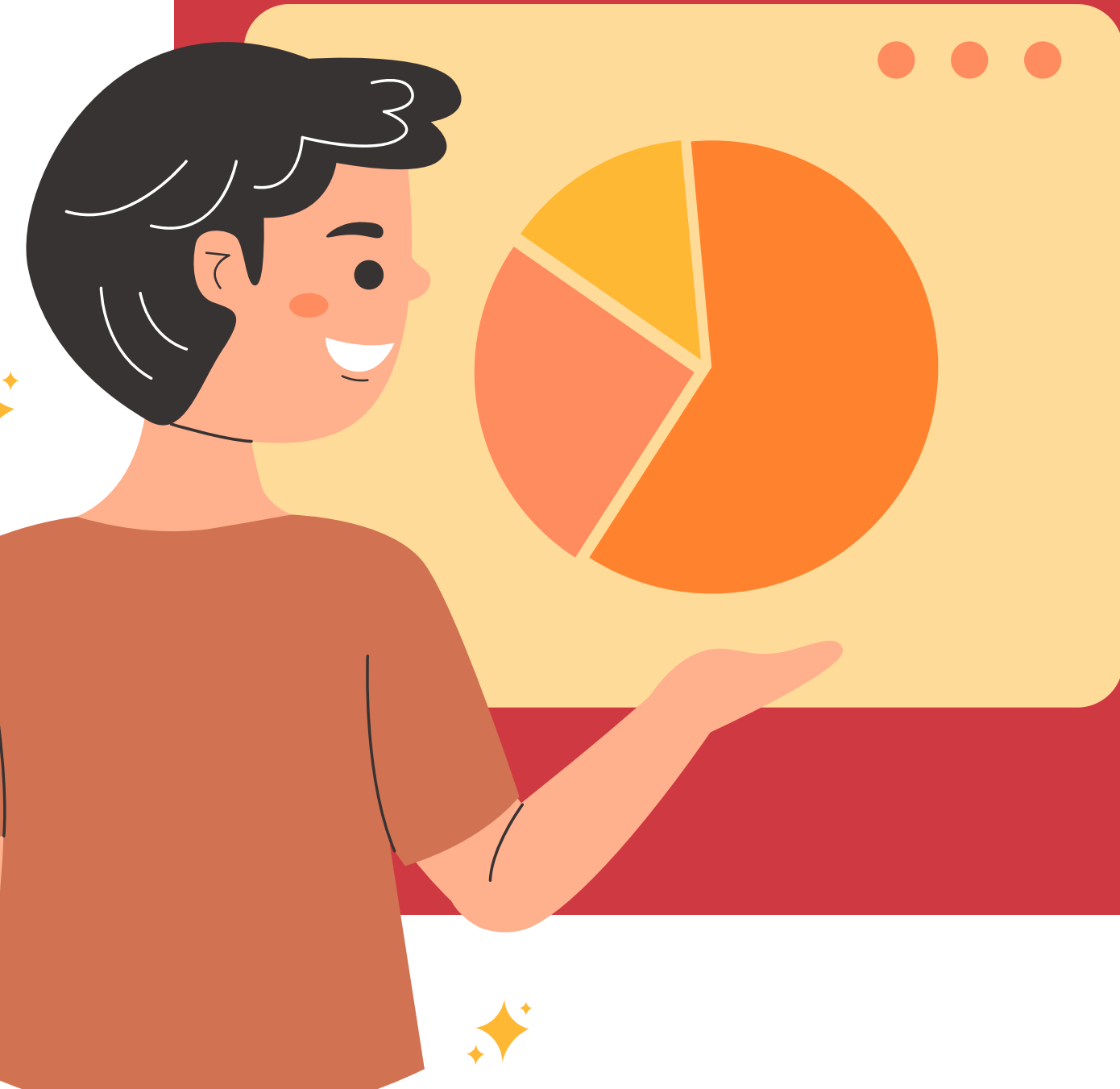


DATA INFO

```
data.info()

<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   PENGELUARAN/BULAN                        289801 non-null  object
11   Apakah saudara pernah di diagnosa atau menderita penyakit Jantung ?  289801 non-null  object
12   Apakah saudara pernah di diagnosa atau menderita penyakit Stroke ?    289801 non-null  object
13   Apakah saudara pernah di diagnosa atau menderita penyakit Diabetes Melitus ? 289801 non-null  object
14   Apakah saudara pernah di diagnosa atau menderita penyakit Hipertensi ?    289801 non-null  object
15   Apakah saudara pernah di diagnosa atau menderita penyakit Kanker ?       289801 non-null  object
16   Apakah saudara pernah di diagnosa atau menderita penyakit Ginjal ?        289801 non-null  object
17   Apakah saudara pernah di diagnosa atau menderita penyakit Tumor ?         289801 non-null  object
18   Apakah saudara pernah di diagnosa atau menderita penyakit Osteoarthritis ? 289801 non-null  object
19   Apakah saudara pernah di diagnosa atau menderita penyakit Lambung ?        289801 non-null  object
20   Apakah saudara pernah di diagnosa atau menderita penyakit Liver ?          289801 non-null  object
21   Apakah saudara pernah di diagnosa atau menderita penyakit Tuberculosis (TBC) ? 289801 non-null  object
22   Apakah saudara pernah di diagnosa atau menderita penyakit Demam Berdarah Dengue (DBD) ? 289801 non-null  object
23   Apakah saudara pernah di diagnosa atau menderita penyakit COVID-19 ?       289801 non-null  object
24   Apakah saudara pernah di diagnosa atau menderita penyakit HIV/AIDS ?       289801 non-null  object
25   Apakah saudara pernah di diagnosa atau menderita penyakit Diare ?           289801 non-null  object
26   Apakah saudara pernah di diagnosa atau menderita penyakit Malaria ?        289801 non-null  object
27   Apakah saudara pernah di diagnosa atau menderita penyakit Gonore atau Sipilis ? 289801 non-null  object
28   Apakah saudara pernah di diagnosa atau menderita penyakit Thypus ?         289801 non-null  object
29   Apakah saudara pernah di diagnosa atau menderita penyakit Dermatitis (Skabies)/ peradangan atau iritasi kulit ? 289801 non-null  object
30   Apakah saudara pernah di diagnosa atau menderita penyakit Mikosis (infeksi jamur) ? 289801 non-null  object
31   Sering menderita sakit kepala ?       289801 non-null  object
32   Tidak nafsu makan ?                   289801 non-null  object
33   Sulit tidur ?                         289801 non-null  object
34   Mudah takut ?                        289801 non-null  object
35   Merasa tegang, cemas atau kuatir ?    289801 non-null  object
36   Tangan anda gemetar ?                 289801 non-null  object
37   Pencernaan anda terganggu / buruk ?   289801 non-null  object
38   Sulit untuk berpikir jernih ?         289801 non-null  object
39   Merasa tidak bahagia ?               289801 non-null  object
40   Menangis lebih sering ?             289801 non-null  object
41   Merasa sulit untuk menikmati kegiatan sehari - hari? 289801 non-null  object
42   Sulit untuk mengambil keputusan ?     289801 non-null  object
43   Pekerjaan anda sehari - hari terganggu ? 289801 non-null  object
44   Tidak mampu melakukan hal - hal yang bermanfaat dalam hidup? 289801 non-null  object
45   Kehilangan minat pada berbagai hal ?  289801 non-null  object
46   Tidak merasa berharga?               289801 non-null  object
47   Mempunyai pikiran untuk mengakhiri hidup? 289801 non-null  object
48   Merasa lelah sepanjang waktu?         289801 non-null  object

48 Merasa lelah sepanjang waktu? 289801 non-null object
49 Mengalami rasa tidak enak di perut? 289801 non-null object
50 Mudah lelah ? 289801 non-null object
51 Apakah saudara menderita penyakit gigi dan mulut (karies/gigi berlubang) 289801 non-null object
52 Jika Ya berapa jumlah gigi saudara yang Karies 66057 non-null object
53 Berapa kali saudara olahraga/aktifitas fisik selama seminggu ? 289801 non-null object
54 Seberapa banyak Saudara mengkonsumsi air mineral dalam satu hari? 289801 non-null object
55 Seberapa sering Saudara mengkonsumsi buah dan sayur dalam sehari? 289801 non-null object
56 Dalam 1 hari berapa kali Saudara mandi/ mencuci badan? 289801 non-null object
57 Berapa lama saudara mandi ? 289801 non-null object
58 Apakah saudara merokok ? 289801 non-null object
59 Jika Ya, Berapa Batang Saudara merokok dalam sehari? 137258 non-null object
60 Seberapa banyak saudara mengkonsumsi gula/pemanis setiap harinya ? 289801 non-null object
61 Seberapa banyak Saudara mengkonsumsi Alkohol setiap hari ? 289801 non-null object
62 Seberapa banyak Saudara mengkonsumsi makanan-makanan junk food dan berminyak dalam sehari 289801 non-null object
63 Dalam satu hari berapa kali Saudara membersihkan/ menggosok gigi ? 289801 non-null object
64 Berapa lama saudara menggosok gigi ? 288022 non-null object
65 Dalam sehari, berapa kali Saudara mengganti pakaian dalam? 289801 non-null object
66 Apakah Saudara selalu mencuci tangan dengan air bersih dan sabun sebelum makan? 289801 non-null object
67 Apakah tempat tinggal Saudara mempunyai penghawaan/sirkulasi udara yang cukup? 289801 non-null object
68 Apakah lingkungan tempat tinggal Saudara saat ini terkontaminasi/ terganggu akibat polusi udara yang tidak sehat? 289801 non-null object
69 Sumber air bersih di tempat tinggal saudara ? 289801 non-null object
70 Penggunaan air bersih untuk makan & minum bersumber dari : 289801 non-null object
71 Berapa jarak jamban dengan sumur dirumah saudara ? 289801 non-null object
72 Bagaimana saluran pembuangan air (sanitasi) rumah tangga? 289801 non-null object
73 Apakah terdapat tempat sampah di sekitar rumah? 289801 non-null object
74 Berapa jarak tempat sampah dengan rumah saudara ? 248152 non-null object
75 Apakah di sekitar tempat tinggal Saudara tersedia fasilitas umum (fasum) untuk kegiatan/aktivitas olah raga? 289801 non-null object
76 Kelompok Umur 289801 non-null object
77 NILAI IMT 289801 non-null object
78 IMT 289801 non-null object
dtypes: float64(4), int64(1), object(74)
memory usage: 174.7+ MB
```



PRE PROCESSING

DROP DATA & CEK TIPE DATA

```
data = data.drop(['ID',"Berapa lama saudara menggosok gigi ?","Dalam sehari, berapa kali Saudara mengganti pakaian dalam?","Apakah Saudara selalu mencuci tangan dengan air bersih dan sabun sebelum makan?","Apakah Saudara selalu mencuci tangan dengan air bersih dan sabun sebelum makan?","Apakah tempat tinggal Saudara mempunyai penghawaan/sirkulasi udara yang cukup?","Apakah lingkungan tempat tinggal Saudara saat ini terkontaminasi/ terganggu akibat polusi udara yang tidak sehat?","Sumber air bersih di tempat tinggal saudara ?","Penggunaan air bersih untuk makan & minum bersumber dari : ","Berapa jarak jamban dengan sumur dirumah saudara ?","Bagaimana saluran pembuangan air (sanitasi) rumah tangga?","Apakah terdapat tempat sampah di sekitar rumah? ","Berapa jarak tempat sampah dengan rumah saudara ?","Berapa lama saudara mandi ?","Dalam 1 hari berapa kali saudara mandi/ mencuci badan?","Jika Ya berapa jumlah gigi saudara yang Karies","Apakah saudara menderita penyakit gigi dan mulut (karies/gigi berlubang)","Apakah saudara pernah di diagnosa atau menderita penyakit Mikosis (infeksi jamur) ?","Jika Ya, Berapa Batang Saudara merokok dalam sehari?","Kelompok Umur","Pencernaan anda terganggu / buruk ?","Apakah saudara pernah di diagnosa atau menderita penyakit Dermatitis (Skabies)/ peradangan atau iritasi kulit ?","Apakah saudara pernah di diagnosa atau menderita penyakit Thypus ?","Apakah saudara pernah di diagnosa atau menderita penyakit Gonore atau Sipilis ?","Apakah saudara pernah di diagnosa atau menderita penyakit Malaria ?","Apakah saudara pernah di diagnosa atau menderita penyakit Diare ?","Apakah saudara pernah di diagnosa atau menderita penyakit HIV/AIDS ?","Apakah saudara pernah di diagnosa atau menderita penyakit COVID-19 ?","Apakah saudara pernah di diagnosa atau menderita penyakit Demam Berdarah Dengue (DBD) ?","Apakah saudara pernah di diagnosa atau menderita penyakit Liver ?","Apakah saudara pernah di diagnosa atau menderita penyakit Lambung ?","Apakah saudara pernah di diagnosa atau menderita penyakit Osteoarthritis ?","Apakah saudara pernah di diagnosa atau menderita penyakit Tumor ?","Apakah saudara pernah di diagnosa atau menderita penyakit Ginjal ?","Apakah saudara pernah di diagnosa atau menderita penyakit Kanker ?","Apakah saudara pernah di diagnosa atau menderita penyakit Tuberculosis (TBC) ?","Dalam satu hari berapa kali Saudara membersihkan/ menggosok gigi ?","Apakah di sekitar tempat tinggal Saudara tersedia fasilitas umum (fasum) untuk kegiatan/aktivitas olah r.
```

```
data["IMT"] = data["IMT"].astype("category")
data["Merasa lelah sepanjang waktu?"] = data["Merasa lelah sepanjang waktu?"].astype("category")
data["WILAYAH"]=data["WILAYAH"].astype("category")
data["JENIS KELAMIN"]=data["JENIS KELAMIN"].astype("category")
data["PENDIDIKAN TERAKHIR"]=data["PENDIDIKAN TERAKHIR"].astype("category")
data["Status Perkawinan"]=data["Status Perkawinan"].astype("category")
data["PENGELUARAN/BULAN"]=data["PENGELUARAN/BULAN"].astype("category")
data["Apakah saudara pernah di diagnosa atau menderita penyakit Jantung ?"]=data["Apakah saudara pernah di diagnosa atau menderita penyakit Jantung ?"].astype("category")
data["Apakah saudara pernah di diagnosa atau menderita penyakit Stroke ?"]=data["Apakah saudara pernah di diagnosa atau menderita penyakit Stroke ?"].astype("category")
data["Apakah saudara pernah di diagnosa atau menderita penyakit Diabetes Melitus ?"]=data["Apakah saudara pernah di diagnosa atau menderita penyakit Diabetes Melitus ?"].astype("category")
data["Apakah saudara pernah di diagnosa atau menderita penyakit Hipertensi ?"]=data["Apakah saudara pernah di diagnosa atau menderita penyakit Hipertensi ?"].astype("category")
data["Sering menderita sakit kepala ?"]=data["Sering menderita sakit kepala ?"].astype("category")
data["Tidak nafsu makan ?"]=data["Tidak nafsu makan ?"].astype("category")
data["Sulit tidur ?"]=data["Sulit tidur ?"].astype("category")
data["Mudah takut ?"]=data["Mudah takut ?"].astype("category")
data["Merasa tegang, cemas atau kuatir ?"]=data["Merasa tegang, cemas atau kuatir ?"].astype("category")
data["Tangan anda gemetar ?"]=data["Tangan anda gemetar ?"].astype("category")
data["Sulit untuk berpikir jernih ?"]=data["Sulit untuk berpikir jernih ?"].astype("category")
data["Pekerjaan anda sehari - hari terganggu ?"]=data["Pekerjaan anda sehari - hari terganggu ?"].astype("category")
data["Merasa tidak bahagia ?"]=data["Merasa tidak bahagia ?"].astype("category")
data["Menangis lebih sering ?"]=data["Menangis lebih sering ?"].astype("category")
data["Merasa sulit untuk menikmati kegiatan sehari - hari?"]=data["Merasa sulit untuk menikmati kegiatan sehari - hari?"].astype("category")
data["Sulit untuk mengambil keputusan ?"]=data["Sulit untuk mengambil keputusan ?"].astype("category")
data["Tidak mampu melakukan hal - hal yang bermanfaat dalam hidup?"]=data["Tidak mampu melakukan hal - hal yang bermanfaat dalam hidup?"].astype("category")
data["Kehilangan minat pada berbagai hal ?"]=data["Kehilangan minat pada berbagai hal ?"].astype("category")
data["Tidak merasa berharga?"]=data["Tidak merasa berharga?"].astype("category")
data["Mempunyai pikiran untuk mengakhiri hidup?"]=data["Mempunyai pikiran untuk mengakhiri hidup?"].astype("category")
data["Merasa lelah sepanjang waktu?"]=data["Merasa lelah sepanjang waktu?"].astype("category")
data["Mengalami rasa tidak enak di perut?"]=data["Mengalami rasa tidak enak di perut?"].astype("category")
data["Mudah lelah ?"]=data["Mudah lelah ?"].astype("category")
data["Berapa kali saudara olahraga/aktifitas fisik selama seminggu ?"]=data["Berapa kali saudara olahraga/aktifitas fisik selama seminggu ?"].astype("category")
data["Seberapa banyak Saudara mengkonsumsi air mineral dalam satu hari?"]=data["Seberapa banyak Saudara mengkonsumsi air mineral dalam satu hari?"].astype("category")
data["Seberapa sering Saudara mengkonsumsi buah dan sayur dalam sehari?"]=data["Seberapa sering Saudara mengkonsumsi buah dan sayur dalam sehari?"].astype("category")
data["Apakah saudara merokok ?"]=data["Apakah saudara merokok ?"].astype("category")
data["Seberapa banyak saudara mengkonsumsi gula/pemanis setiap harinya ?"]=data["Seberapa banyak saudara mengkonsumsi gula/pemanis setiap harinya ?"].astype("category")
data["Seberapa banyak Saudara mengkonsumsi Alkohol setiap hari ?"]=data["Seberapa banyak Saudara mengkonsumsi Alkohol setiap hari ?"].astype("category")
data["Seberapa banyak Saudara mengkonsumsi makanan-makanan junk food dan berminyak dalam sehari"]=data["Seberapa banyak Saudara mengkonsumsi makanan-makanan junk food dan berminyak dalam sehari"].astype("category")
```


DROP DATA & CEK TIPE DATA

```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 289801 entries, 0 to 289800
Data columns (total 42 columns):
#   Column                                                                 Non-Null Count  Dtype
---  -
0   WILAYAH                                                                289801 non-null  category
1   JENIS KELAMIN                                                         289801 non-null  category
2   USIA                                                                  289801 non-null  float64
3   LAMA BERDINAS                                                         289801 non-null  float64
4   PENDIDIKAN TERAKHIR                                                  289801 non-null  category
5   TINGGI BADAN                                                         289801 non-null  object
6   BERAT BADAN                                                          289801 non-null  float64
7   LINGKAR PINGGANG                                                     289801 non-null  float64
8   Status Perkawinan                                                    289801 non-null  category
9   PENGELUARAN/BULAN                                                   289801 non-null  category
10  Apakah saudara pernah di diagnosa atau menderita penyakit Jantung ? 289801 non-null  category
11  Apakah saudara pernah di diagnosa atau menderita penyakit Stroke ?   289801 non-null  category
12  Apakah saudara pernah di diagnosa atau menderita penyakit Diabetes Melitus ? 289801 non-null  category
13  Apakah saudara pernah di diagnosa atau menderita penyakit Hipertensi ? 289801 non-null  category
14  Sering menderita sakit kepala ?                                       289801 non-null  category
15  Tidak nafsu makan ?                                                  289801 non-null  category
16  Sulit tidur ?                                                         289801 non-null  category
17  Mudah takut ?                                                         289801 non-null  category
18  Merasa tegang, cemas atau khawatir ?                                  289801 non-null  category
19  Tangan anda gemetar ?                                                 289801 non-null  category
20  Sulit untuk berpikir jernih ?                                         289801 non-null  category
21  Merasa tidak bahagia ?                                                289801 non-null  category
22  Menangis lebih sering ?                                              289801 non-null  category
23  Merasa sulit untuk menikmati kegiatan sehari - hari?                289801 non-null  category
24  Sulit untuk mengambil keputusan ?                                     289801 non-null  category
25  Pekerjaan anda sehari - hari terganggu ?                             289801 non-null  category
26  Tidak mampu melakukan hal - hal yang bermanfaat dalam hidup?      289801 non-null  category
27  Kehilangan minat pada berbagai hal ?                                  289801 non-null  category
28  Tidak merasa berharga?                                                289801 non-null  category
29  Mempunyai pikiran untuk mengakhiri hidup?                            289801 non-null  category
30  Merasa lelah sepanjang waktu?                                         289801 non-null  category
31  Mengalami rasa tidak enak di perut?                                   289801 non-null  category
32  Mudah lelah ?                                                         289801 non-null  category
33  Berapa kali saudara olahraga/aktifitas fisik selama seminggu ?     289801 non-null  category
34  Seberapa banyak Saudara mengkonsumsi air mineral dalam satu hari?   289801 non-null  category
35  Seberapa sering Saudara mengkonsumsi buah dan sayur dalam sehari?   289801 non-null  category
36  Apakah saudara merokok ?                                              289801 non-null  category
37  Seberapa banyak saudara mengkonsumsi gula/pemanis setiap harinya ? 289801 non-null  category
38  Seberapa banyak Saudara mengkonsumsi Alkohol setiap hari ?          289801 non-null  category
39  Seberapa banyak Saudara mengkonsumsi makanan-makanan junk food dan berminyak dalam sehari 289801 non-null  category
40  NILAI IMT                                                            289801 non-null  object
41  IMT                                                                  289801 non-null  category
dtypes: category(36), float64(4), object(2)
memory usage: 23.2+ MB
```

MISSING VALUES & DUPLICATES

```
data.isna().sum()
WILAYAH 0
JENIS KELAMIN 0
USIA 0
LAMA BERDINAS 0
PENDIDIKAN TERAKHIR 0
TINGGI BADAN 0
BERAT BADAN 0
LINGKAR PINGGANG 0
Status Perkawinan 0
PENGELUARAN/BULAN 0
Apakah saudara pernah di diagnosa atau menderita penyakit Jantung ? 0
Apakah saudara pernah di diagnosa atau menderita penyakit Stroke ? 0
Apakah saudara pernah di diagnosa atau menderita penyakit Diabetes Melitus ? 0
Apakah saudara pernah di diagnosa atau menderita penyakit Hipertensi ? 0
Sering menderita sakit kepala ? 0
Tidak nafsu makan ? 0
Sulit tidur ? 0
Mudah takut ? 0
Merasa tegang, cemas atau kuatir ? 0
Tangan anda gemetar ? 0
Sulit untuk berpikir jernih ? 0
Merasa tidak bahagia ? 0
Menangis lebih sering ? 0
Merasa sulit untuk menikmati kegiatan sehari - hari? 0
Sulit untuk mengambil keputusan ? 0
Pekerjaan anda sehari - hari terganggu ? 0
Tidak mampu melakukan hal - hal yang bermanfaat dalam hidup? 0
Kehilangan minat pada berbagai hal ? 0
Tidak merasa berharga? 0
Mempunyai pikiran untuk mengakhiri hidup? 0
Merasa lelah sepanjang waktu? 0
Mengalami rasa tidak enak di perut? 0
Mudah lelah ? 0
Berapa kali saudara olahraga/aktifitas fisik selama seminggu ? 0
Seberapa banyak Saudara mengkonsumsi air mineral dalam satu hari? 0
Seberapa sering Saudara mengkonsumsi buah dan sayur dalam sehari? 0
Apakah saudara merokok ? 0
Seberapa banyak saudara mengkonsumsi gula/pemanis setiap harinya ? 0
Seberapa banyak Saudara mengkonsumsi Alkohol setiap hari ? 0
Seberapa banyak Saudara mengkonsumsi makanan-makanan junk food dan berminyak dalam sehari 0
NILAI IMT 0
IMT 0
dtype: int64

[ ] data = data.dropna(how='all', axis=0)
```

```
[10] data = data.dropna(how='all', axis=0)

[11] data.duplicated().sum()
89

[12] data = data.drop_duplicates()

data.shape
(289712, 42)
```

CEK VARIABEL

1. TINGGI BADAN

```
data["TINGGI BADAN"].unique()

array(['168.9', '169.5', '166.8', '164.5', '165.5', '168.5', '175',
      '172.5', '166.5', '166', '172', '167', '175.8', '163.5', '167.8',
      '72.5', '168.8', '174.5', '161', '172.6', '160.5', '175.5',
      '171.5', '17', '165', '170', '150', '165.9', '156', '168', '167.5',
      '166.3', '159.5', '169', '173.5', '171', '1.65', '166.9', '176.5',
      '1.71', '166.6', '1.76', '171.3', '156.5', '157.5', '17.1', '160',
      '1.75', '1.66', '1.72', '1.68', '162.5', '17.5', '177.5', '169.2',
      '178', '176', '180.5', '16.6', '1.7', '178.5', '167.2', '171.8',
      '170.3', '165.4', '1.69', '173.4', '170.2', '15.5', '182', '158.5',
      '185.5', '158', '164.4', '165.7', '167.6', '1.64', '161.5',
      '179.5', '165.58', '169.7', '168.2', '1.77', '173', '170.5',
      '17.6', '163', '16.7', '164', '171.6', '176.6', '16.39', '188.5',
      '173.8', '172.8', '164.7', '155', '174.8', '168.58', '5', '174',
      '153', '180', '181', '16', '177', '162', '159', '187', '15', '179',
      '185', '151', '157', '18', '152', '26', '20', '148', '154', '19',
      '72', '23', '30', '29', '32', '33', '35', '39', '38', '65', '63',
      '67', '68', '69', '71', '70', '45', '46', '76', '75', '53', '183',
      '78', '80', '55', '57', '60', '84', '62', '66', '59', '56', '145',
      '64', '54', '74', '73', '61', '79', '86', '82', '87', '89', '90',
      '88', '95', '85', '97', '108', '101', '110', '120', '100', '195',
      '140', '196', '188', '136', '134', '137', '143', '144', '146',
      '147', '184', '190', '130', '189', '186', '138', '149', '135', '156',
      '170', '147', '158', '183', '159', '171', '178', '160', '161', '172', '179', '162', '163',
      '173', '164', '180', '174', '165', '166', '175', '167', '176', '182', '168', '169', '177',
      '184', '155', '153', '152', '157', '150', '181', '187', '189', '185', '154', '148', '136',
      '145', '190', '186', '143', '149', '191', '193', '151', '188', '146', '135', '192', '195',
      '144', '140', '1799', '141', '197', '194', '200', '142', '270', '271', '138', '198', '196',
      '199', '273', '267', '268', '269', '265', '275', '264', '279', '266', '254', '276', '274',
      '253', '277', '272', '256', '257', '262', '263', '260', '255', '365', '575'],
      dtype=object)

[15] data["TINGGI BADAN"] = data["TINGGI BADAN"].replace("", "", regex=True)

[16] data["TINGGI BADAN"] = data["TINGGI BADAN"].astype(float)

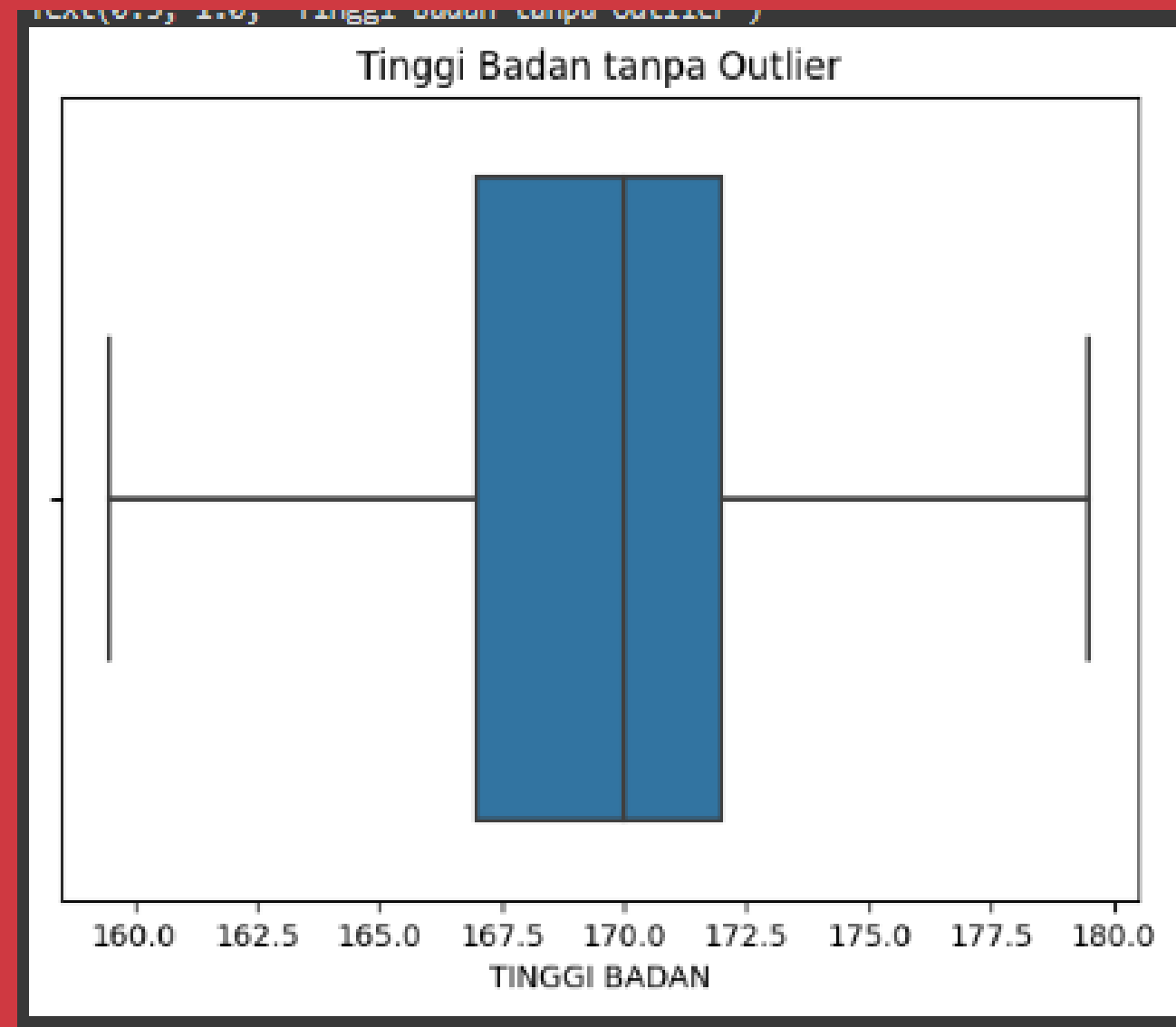
[17] data.loc[data["TINGGI BADAN"] < 2, "TINGGI BADAN"] = data["TINGGI BADAN"]*100

[45] Q1 = data["TINGGI BADAN"].quantile(0.25)
      Q3 = data["TINGGI BADAN"].quantile(0.75)
      IQR = Q3 - Q1

      lower_bound = Q1 - 1.5 * IQR
      upper_bound = Q3 + 1.5 * IQR

      data_cleaned_tb = data[(data["TINGGI BADAN"] >= lower_bound) & (data["TINGGI BADAN"] <= upper_bound)]

sns.boxplot(x="TINGGI BADAN", data=data_cleaned_tb)
plt.title('Tinggi Badan tanpa Outlier')
```



CEK VARIABEL

2. BERAT BADAN

```
[22] data.loc[data["BERAT BADAN"] < 10, "BERAT BADAN"] = data["BERAT BADAN"]*10

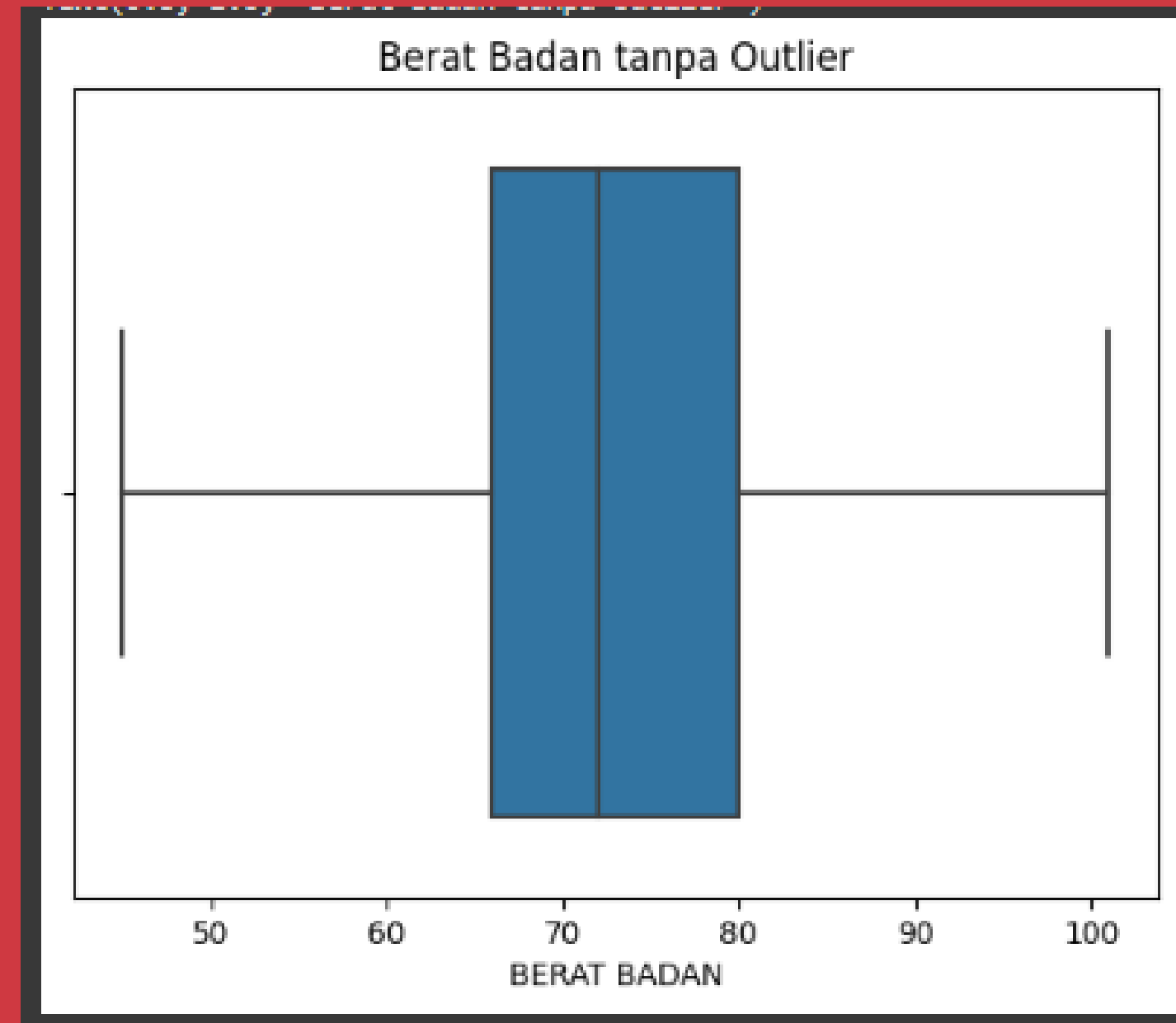
[23] data["BERAT BADAN"] = data["BERAT BADAN"].astype(float)

[44] Q1 = data['BERAT BADAN'].quantile(0.25)
      Q3 = data['BERAT BADAN'].quantile(0.75)
      IQR = Q3 - Q1

      lower_bound = Q1 - 1.5 * IQR
      upper_bound = Q3 + 1.5 * IQR

      data_cleaned_bb = data[(data['BERAT BADAN'] >= lower_bound) & (data['BERAT BADAN'] <= upper_bound)]

sns.boxplot(x='BERAT BADAN', data=data_cleaned_bb)
plt.title('Berat Badan tanpa Outlier')
```



CEK VARIABEL

3. USIA

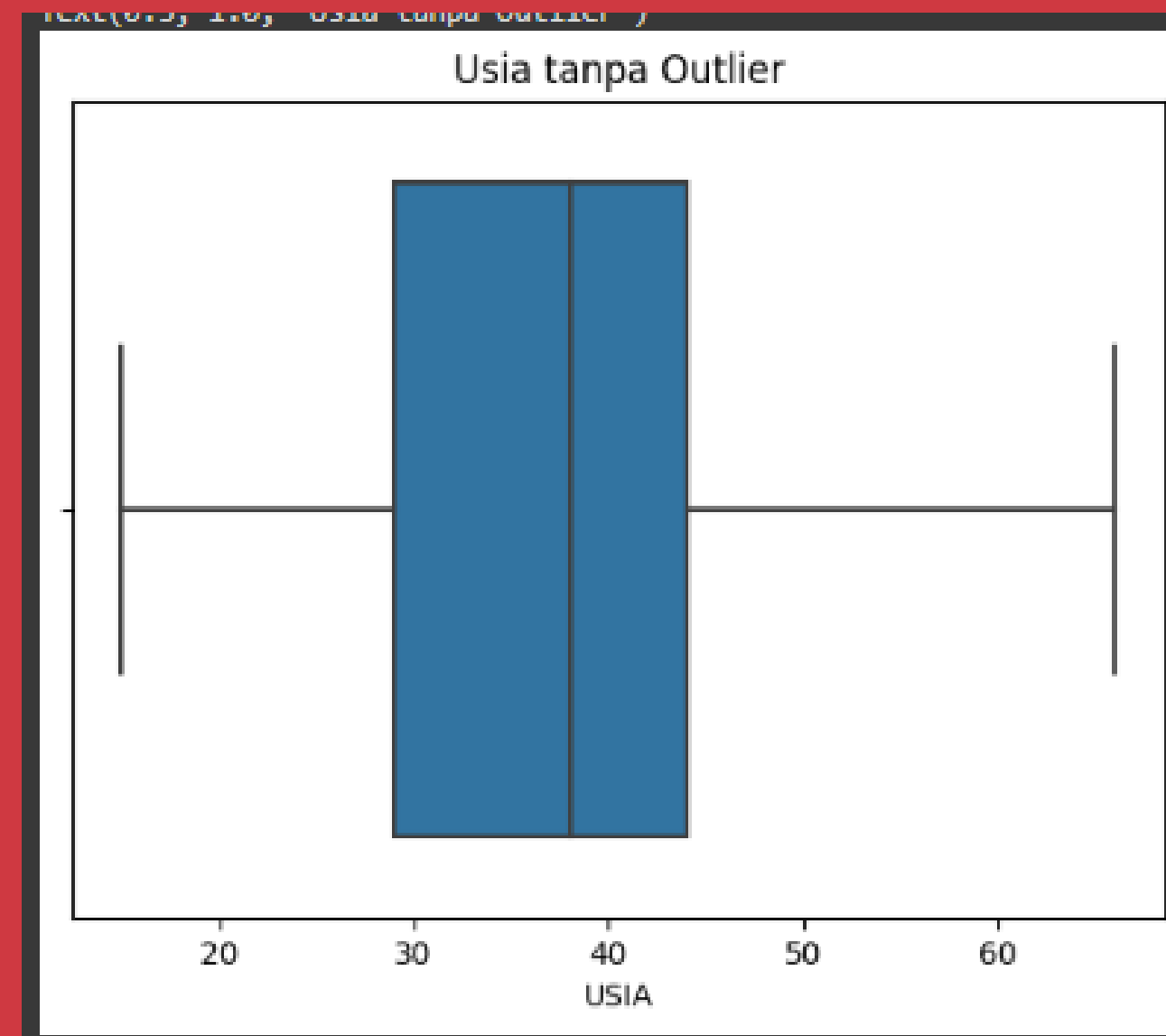
```
[28] data["USIA"] = data["USIA"].astype(int)

[43] Q1 = data['USIA'].quantile(0.25)
      Q3 = data['USIA'].quantile(0.75)
      IQR = Q3 - Q1

      lower_bound = Q1 - 1.5 * IQR
      upper_bound = Q3 + 1.5 * IQR

      data_cleaned_usia = data[(data['USIA'] >= lower_bound) & (data['USIA'] <= upper_bound)]

sns.boxplot(x='USIA', data=data_cleaned_usia)
plt.title('Usia tanpa Outlier')
```



CEK VARIABEL

4. LINGKAR PINGGANG

```
[31] data["LINGKAR PINGGANG"] = data["LINGKAR PINGGANG"].replace("-", "", regex = True)

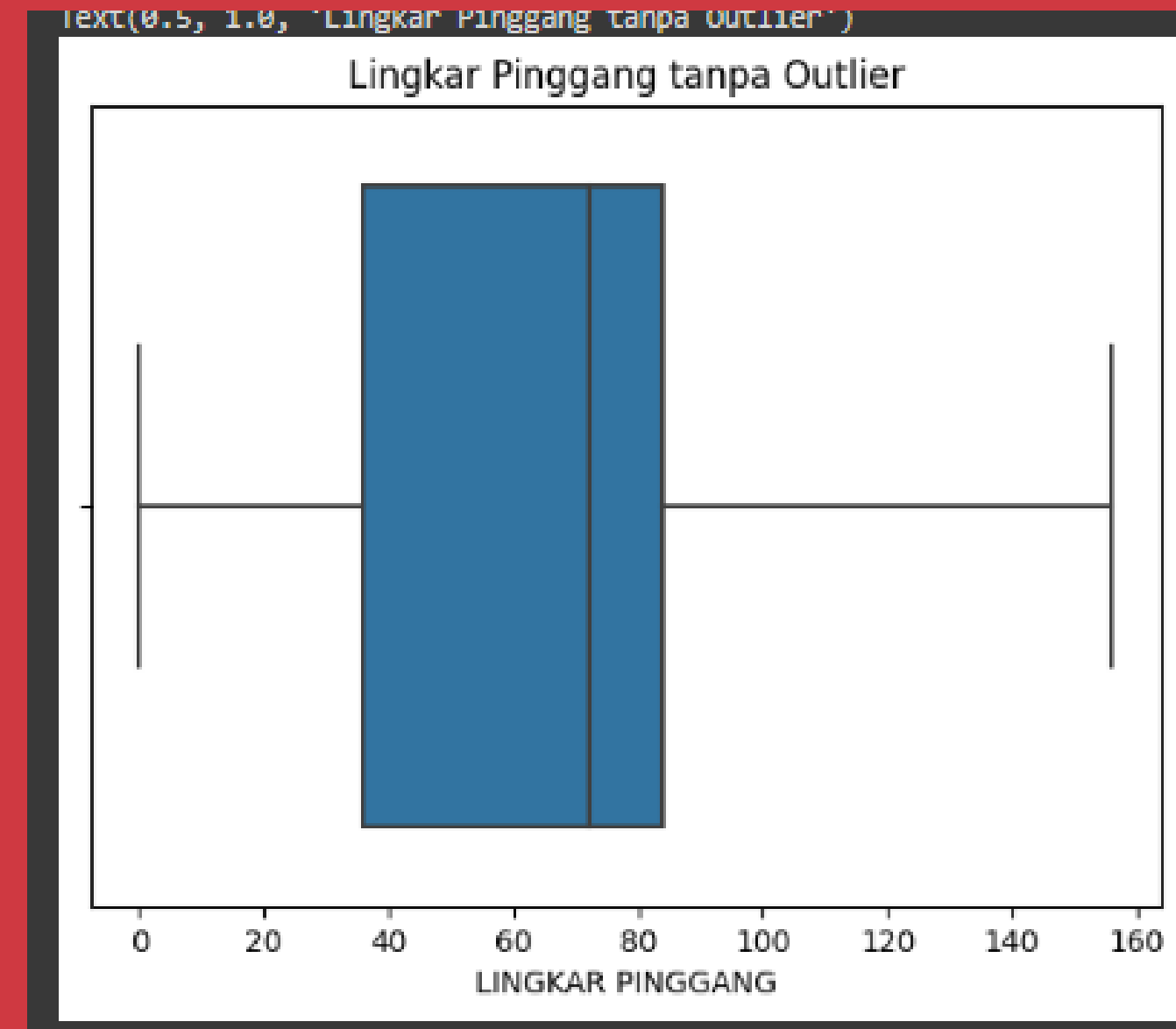
[32] data["LINGKAR PINGGANG"] = data["LINGKAR PINGGANG"].astype(float)

[40] Q1 = data['LINGKAR PINGGANG'].quantile(0.25)
     Q3 = data['LINGKAR PINGGANG'].quantile(0.75)
     IQR = Q3 - Q1

     lower_bound = Q1 - 1.5 * IQR
     upper_bound = Q3 + 1.5 * IQR

     data_cleaned_lp = data[(data['LINGKAR PINGGANG'] >= lower_bound) & (data['LINGKAR PINGGANG'] <= upper_bound)]

sns.boxplot(x='LINGKAR PINGGANG', data=data_cleaned_lp)
plt.title('Lingkar Pinggang tanpa Outlier')
```



CEK VARIABEL

5. LAMA BERDINAS

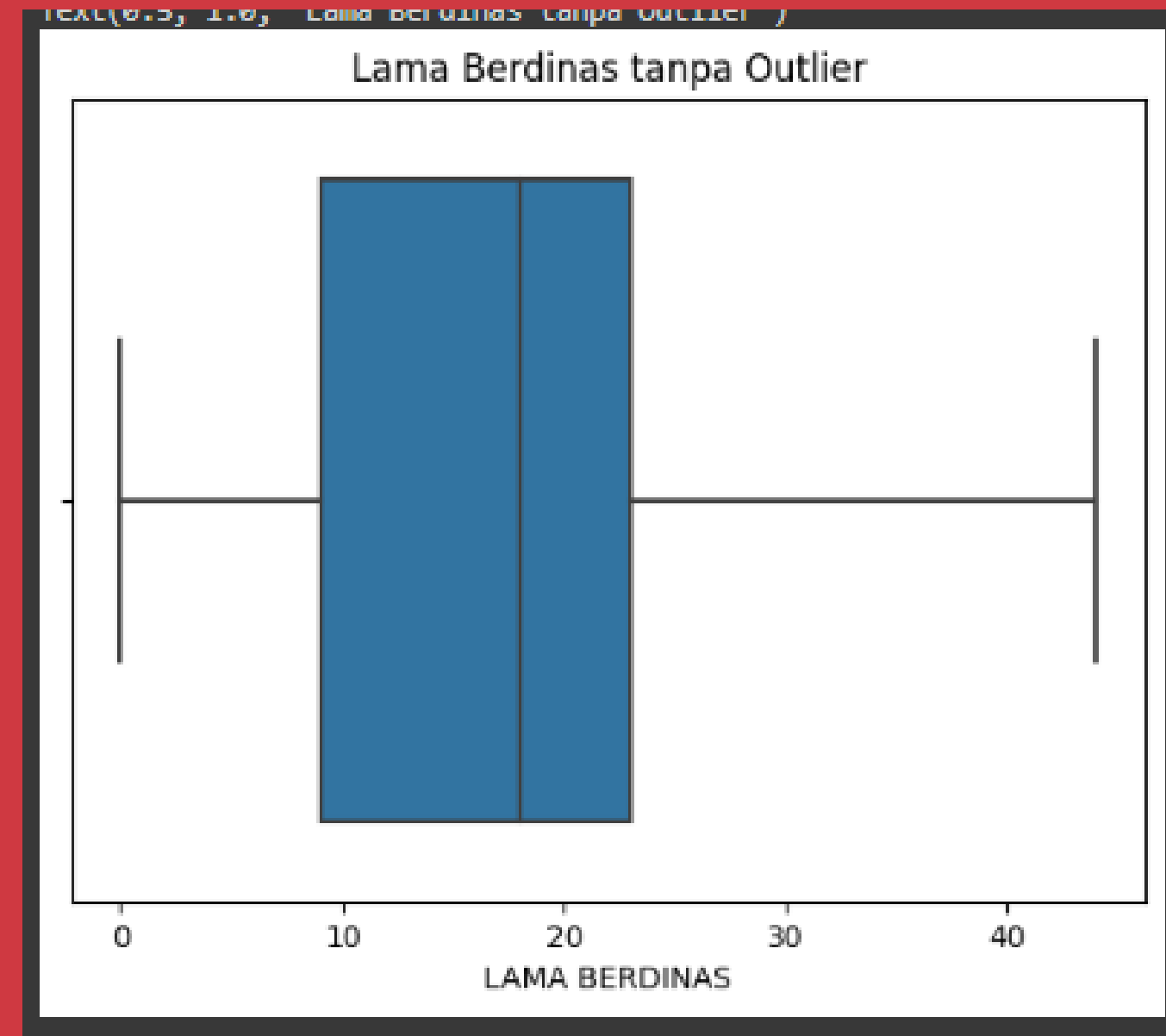
```
[36] data["LAMA BERDINAS"] = data["LAMA BERDINAS"].astype(float)

Q1 = data['LAMA BERDINAS'].quantile(0.25)
Q3 = data['LAMA BERDINAS'].quantile(0.75)
IQR = Q3 - Q1

lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

data_cleaned_dinas = data[(data['LAMA BERDINAS'] >= lower_bound) & (data['LAMA BERDINAS'] <= upper_bound)]

[38] sns.boxplot(x='LAMA BERDINAS', data=data_cleaned_dinas)
plt.title('Lama Berdinas tanpa Outlier')
```



KELOMPOK LAMA BERDINAS

```
[47] bins = [0, 10, 20, 30, 40, 50, float('inf')]
      labels = ['1-10', '11-20', '21-30', '31-40', '41-50', '>50']

      data['Kelompok Lama Berdinas'] = pd.cut(data['LAMA BERDINAS'], bins=bins, labels=labels, right=False)
```

```
[48] data['Kelompok Lama Berdinas']
```

0	11-20
1	21-30
2	1-10
3	21-30
4	1-10
	...
289796	1-10
289797	21-30
289798	>50
289799	11-20
289800	1-10

Name: Kelompok Lama Berdinas, Length: 289712, dtype: category
Categories (6, object): ['1-10' < '11-20' < '21-30' < '31-40' < '41-50' < '>50']

PENGGABUNGAN DATA

```
[51] data_combined = data_combined.dropna(how="any", axis = 0)

[52] data_combined.loc[data_combined['USIA'] <= 24, 'KELOMPOK USIA'] = '15-24'
data_combined.loc[(data_combined['USIA'] > 24) & (data_combined["USIA"] <=34), 'KELOMPOK USIA'] = '25-34'
data_combined.loc[(data_combined['USIA'] > 34) & (data_combined["USIA"] <=44), 'KELOMPOK USIA'] = '35-44'
data_combined.loc[(data_combined['USIA'] >44) & (data_combined["USIA"] <=54), 'KELOMPOK USIA'] = '45-54'
data_combined.loc[(data_combined['USIA'] >54) & (data_combined["USIA"] <=64), 'KELOMPOK USIA'] = '54-64'
data_combined.loc[data_combined['USIA'] > 64, 'KELOMPOK USIA'] = '>=65'

[53] data_combined["KELOMPOK USIA"] = data_combined["KELOMPOK USIA"].astype("category")

[54] data_combined["NILAI IMT"] = data_cleaned_bb["BERAT BADAN"]/((data_cleaned_tb["TINGGI BADAN"]/100)**2)

[55] data_combined["NILAI IMT"] = data_combined["NILAI IMT"].astype(float)

data_combined.loc[data_combined['NILAI IMT'] < 17, 'IMT'] = 'Sangat Kurus'
data_combined.loc[(data_combined['NILAI IMT'] >= 17) & (data_combined["NILAI IMT"] <18.5), 'IMT'] = 'Kurus'
data_combined.loc[(data_combined['NILAI IMT'] >= 18.5) & (data_combined["NILAI IMT"] <25), 'IMT'] = 'Normal'
data_combined.loc[(data_combined['NILAI IMT'] >= 25) & (data_combined["NILAI IMT"] <=27), 'IMT'] = 'Gemuk'
data_combined.loc[data_combined['NILAI IMT'] > 27, 'IMT'] = 'Obesitas'

[57] data_combined["IMT"] = data_combined["IMT"].astype("category")
```

data_combined

TINGGI BADAN	BERAT BADAN	USIA	LINGKAR PINGGANG	LAMA BERDINAS	MELAYAH	JENIS KELAMIN	PENDIDIKAN TERAKHIR	Status Perkawinan	PENGELUARAN/BULAN	...	Seberapa banyak saudara mengonsumsi air mineral dalam satu hari?	Seberapa sering saudara mengonsumsi buah dan sayur dalam sehari?	Apakah saudara merokok ?	Seberapa banyak saudara mengonsumsi gula/pemais setiap harinya ?	Seberapa banyak Saudara mengonsumsi Alkohol setiap hari ?	Seberapa banyak Saudara mengonsumsi makanan-makanan junk food dan bersmyang dalam sehari	Kelompok Lama Berdinass	KELOMPOK USIA	NILAI ZMT	ZMT	
0	168.9	85.0	34	100.0	16.0	KALBAR	LAKI-LAKI	SMA	Menikah	Rp. 7 Juta s.d. 10 Juta	---	Lebih dari 2 liter	Kurang dari 400gr	Ya	Kurang dari 6 sendok teh	Tidak pernah	1-2 kali/hari	11-20	25-34	29.796114	Obesitas
1	169.5	70.0	43	80.0	23.0	KALBAR	LAKI-LAKI	SMA	Menikah	Rp. 2 Juta s.d. 6 Juta	---	Lebih dari 2 liter	Kurang dari 400gr	Tidak	Tidak pernah	Tidak pernah	Tidak pernah	21-30	35-44	24.364563	Normal
2	169.5	60.0	21	78.0	1.0	KALBAR	LAKI-LAKI	SMA	Belum Menikah	Rp. 2 Juta s.d. 6 Juta	---	2 Liter	400gr/lebih	Tidak	Kurang dari 6 sendok teh	Tidak pernah	1-2 kali/hari	1-10	15-24	20.883912	Normal
3	166.8	80.0	40	34.0	23.0	KALBAR	LAKI-LAKI	S2	Menikah	Rp. 7 Juta s.d. 10 Juta	---	2 Liter	Kurang dari 400gr	Tidak	Lebih dari 6 sendok teh	Tidak pernah	1-2 kali/hari	21-30	35-44	28.753975	Obesitas
4	164.5	62.0	28	32.0	9.0	KALBAR	LAKI-LAKI	SMA	Menikah	Rp. 2 Juta s.d. 6 Juta	---	2 Liter	Kurang dari 400gr	Ya	Kurang dari 6 sendok teh	Tidak pernah	1-2 kali/hari	1-10	25-34	22.911836	Normal
...
269794	169.0	60.0	46	32.0	26.0	DIY	LAKI-LAKI	SMA	Menikah	Rp. 2 Juta s.d. 6 Juta	---	2 Liter	400gr/lebih	Ya	Kurang dari 6 sendok teh	Tidak pernah	Tidak pernah	21-30	45-54	21.007668	Normal
269795	169.0	60.0	53	70.0	33.0	JATENG	LAKI-LAKI	SMA	Menikah	Rp. 2 Juta s.d. 6 Juta	---	Lebih dari 2 liter	400gr/lebih	Ya	Kurang dari 6 sendok teh	Tidak pernah	Tidak pernah	31-40	45-54	21.007668	Normal
269796	169.0	60.0	23	78.0	4.0	KALSEL	LAKI-LAKI	SMA	Menikah	Rp. 2 Juta s.d. 6 Juta	---	Lebih dari 2 liter	400gr/lebih	Tidak	Kurang dari 6 sendok teh	Kurang dari 25ml	1-2 kali/hari	1-10	15-24	21.007668	Normal
269799	179.0	55.0	35	68.0	17.0	JABAR	LAKI-LAKI	SMA	Menikah	Rp. 2 Juta s.d. 6 Juta	---	Lebih dari 2 liter	Kurang dari 400gr	Ya	Kurang dari 6 sendok teh	Tidak pernah	1-2 kali/hari	11-20	35-44	17.165507	Kurus
269800	171.0	100.0	19	33.0	1.0	JABAR	LAKI-LAKI	SMA	Belum Menikah	Rp. 2 Juta s.d. 6 Juta	---	Lebih dari 2 liter	400gr/lebih	Ya	Kurang dari 6 sendok teh	Tidak pernah	1-2 kali/hari	1-10	15-24	34.198557	Obesitas

269795 rows x 43 columns

data_combined.to_csv("data_combined.csv")

SAVE PRE-PROCESSING

```
[1] !gdown 1f_r9VmYAIKTzm7zuQV1bzkojtVPO4i-d

Downloading...
From: https://drive.google.com/uc?id=1f_r9VmYAIKTzm7zuQV1bzkojtVPO4i-d
To: /content/data_combined.csv
100% 104M/104M [00:00<00:00, 126MB/s]

data_combined = pd.read_csv("https://drive.google.com/uc?id=1f_r9VmYAIKTzm7zuQV1bzkojtVPO4i-d",index_col=0)
data_combined
```

	TINGGI BADAN	BERAT BADAN	USIA	LINGKAR PINGGANG	LAMA BERDINAS	WILAYAH	JENIS KELAMIN	PENDIDIKAN TERAKHIR	Status Perkawinan	PENGELUARAN/BULAN	...	Seberapa banyak Saudara mengonsumsi air mineral dalam satu hari?	Seberapa sering Saudara mengonsumsi buah dan sayur dalam sehari?	Apakah saudara merokok ?	Seberapa banyak saudara mengonsumsi gula/pemanis setiap harinya ?	Seberapa banyak Saudara mengonsumsi Alkohol setiap hari ?	Seberapa banyak Saudara mengonsumsi makanan-makanan junk food dan berminyak dalam sehari	KELOMPOK USIA	NILAI IMT	IMT	KELOMPOK LAMA BERDINAS
0	168.9	85.0	34	100.0	16.0	KALBAR	LAKI - LAKI	SMA	Menikah	Rp. 7 Juta s.d. 10 Juta	...	Lebih dari 2 liter	Kurang dari 400gr	Ya	Kurang dari 6 sendok teh	Tidak pernah	1 - 2 kali/hari	25-34	29.796114	Obesitas	11-20
1	169.5	70.0	43	80.0	23.0	KALBAR	LAKI - LAKI	SMA	Menikah	Rp. 2 Juta s.d. 6 Juta	...	Lebih dari 2 liter	Kurang dari 400gr	Tidak	Tidak pernah	Tidak pernah	Tidak pernah	35-44	24.364563	Normal	21-30
2	169.5	60.0	21	78.0	1.0	KALBAR	LAKI - LAKI	SMA	Belum Menikah	Rp. 2 Juta s.d. 6 Juta	...	2 Liter	400gr/lebih	Tidak	Kurang dari 6 sendok teh	Tidak pernah	1 - 2 kali/hari	15-24	20.883912	Normal	1-10
3	166.8	80.0	40	34.0	23.0	KALBAR	LAKI - LAKI	S2	Menikah	Rp. 7 Juta s.d. 10 Juta	...	2 Liter	Kurang dari 400gr	Tidak	Lebih dari 6 sendok teh	Tidak pernah	1 - 2 kali/hari	35-44	28.753975	Obesitas	21-30
4	164.5	62.0	28	32.0	9.0	KALBAR	LAKI - LAKI	SMA	Menikah	Rp. 2 Juta s.d. 6 Juta	...	2 Liter	Kurang dari 400gr	Ya	Kurang dari 6 sendok teh	Tidak pernah	1 - 2 kali/hari	25-34	22.911836	Normal	1-10
...
289794	169.0	60.0	46	32.0	26.0	DIY	LAKI - LAKI	SMA	Menikah	Rp. 2 Juta s.d. 6 Juta	...	2 Liter	400gr/lebih	Ya	Kurang dari 6 sendok teh	Tidak pernah	Tidak pernah	45-54	21.007668	Normal	21-30
289795	169.0	60.0	53	70.0	33.0	JATENG	LAKI - LAKI	SMA	Menikah	Rp. 2 Juta s.d. 6 Juta	...	Lebih dari 2 liter	400gr/lebih	Ya	Kurang dari 6 sendok teh	Tidak pernah	Tidak pernah	45-54	21.007668	Normal	31-40
289796	169.0	60.0	23	78.0	4.0	KALSEL	LAKI - LAKI	SMA	Menikah	Rp. 2 Juta s.d. 6 Juta	...	Lebih dari 2 liter	400gr/lebih	Tidak	Kurang dari 6 sendok teh	Kurang dari 25ml	1 - 2 kali/hari	15-24	21.007668	Normal	1-10
289799	179.0	55.0	35	68.0	17.0	JABAR	LAKI - LAKI	SMA	Menikah	Rp. 2 Juta s.d. 6 Juta	...	Lebih dari 2 liter	Kurang dari 400gr	Ya	Kurang dari 6 sendok teh	Tidak pernah	1 - 2 kali/hari	35-44	17.165507	Kurus	11-20
289800	171.0	100.0	19	33.0	1.0	JABAR	LAKI - LAKI	SMA	Belum Menikah	Rp. 2 Juta s.d. 6 Juta	...	Lebih dari 2 liter	400gr/lebih	Ya	Kurang dari 6 sendok teh	Tidak pernah	1 - 2 kali/hari	15-24	34.198557	Obesitas	1-10

269795 rows × 43 columns

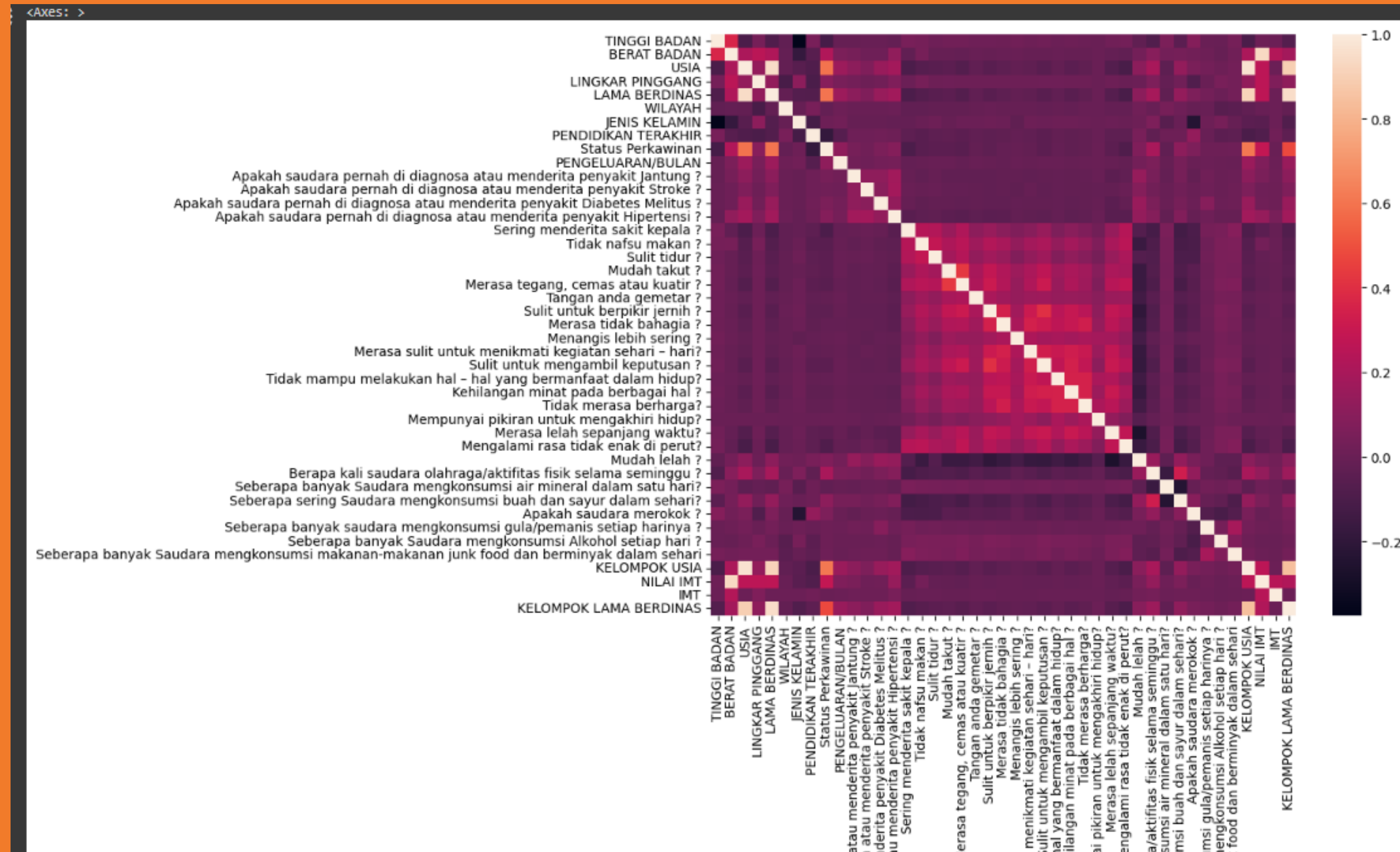
KLASIFIKASI



TAHAPAN KLASIFIKASI

1. Menentukan vektor X dan variabel Y
2. Melakukan Transformasi data dengan label encoder untuk variabel kategorik
3. Melakukan feature selection menggunakan plot korelasi
4. Membuat data train dan data test

PLOT KORELASI



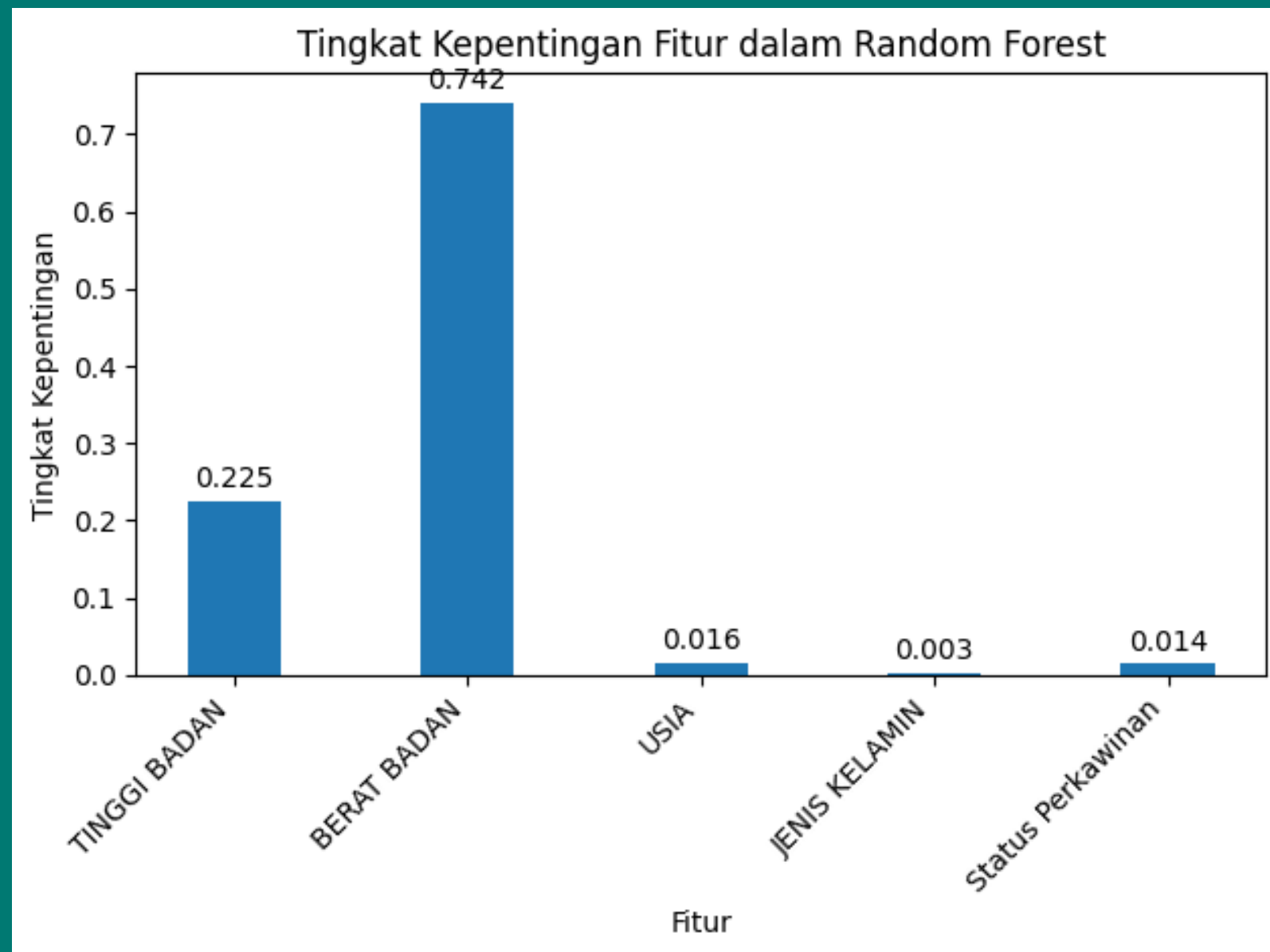
TABEL AKURASI MODEL

Model	Akurasi (%)
Regresi Logistik	98
KNN	99
Decision Tree (Gini)	89
Decision Tree (Entropy)	87
Naive Bayes	79
Random Forest	100
SVM	100
Gradient Boost	100
XGBoost	100
LightGBM	97
CatBoost	37
AdaBoost	43



INSIGHT

FAKTOR MEMENGARUHI OBESITAS



- Berat Badan **74.2%**
- Tinggi Badan **22.5%**
- Usia **1.6%**
- Status Perkawinan **1.4%**
- Jenis Kelamin **0.03%**



HUBUNGAN OBESITAS DENGAN PENYAKIT STROKE, JANTUNG, DIABETES MELITUS, DAN HIPERTENSI (REGRESI LOGISTIK)

Accuracy: 0.718267573528049

Confusion Matrix:

[[37483 1296]

[13906 1274]]

Classification Report:

	precision	recall	f1-score	support
0	0.73	0.97	0.83	38779
1	0.50	0.08	0.14	15180
accuracy			0.72	53959
macro avg	0.61	0.53	0.49	53959
weighted avg	0.66	0.72	0.64	53959

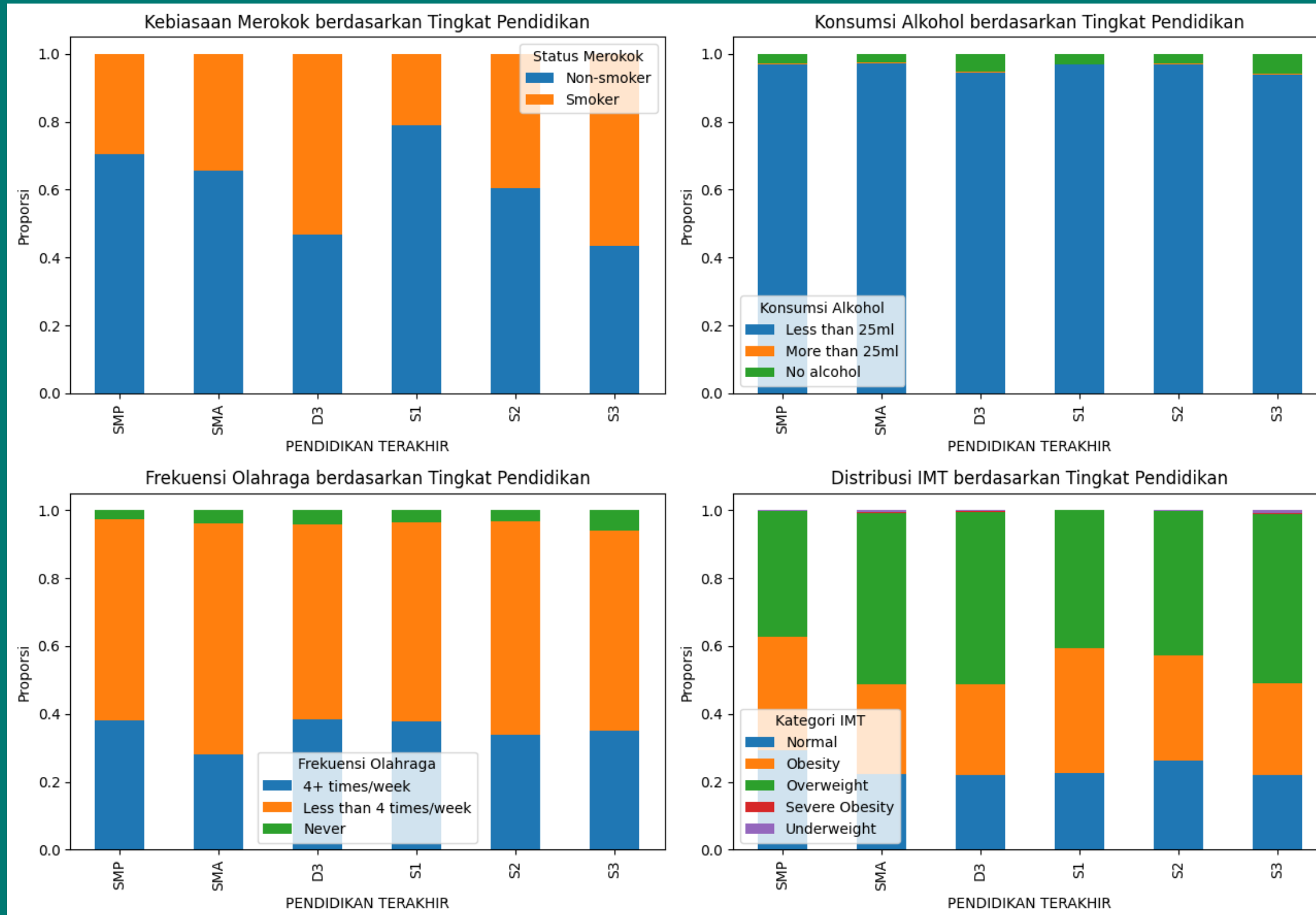
Intercept Model: [-1.00608111]

Koefisien Model: [[0.39767038 -0.25913835 -0.00123403 1.01827954]]

Odds Ratio: [[1.48835336 0.77171625 0.99876673 2.76842769]]

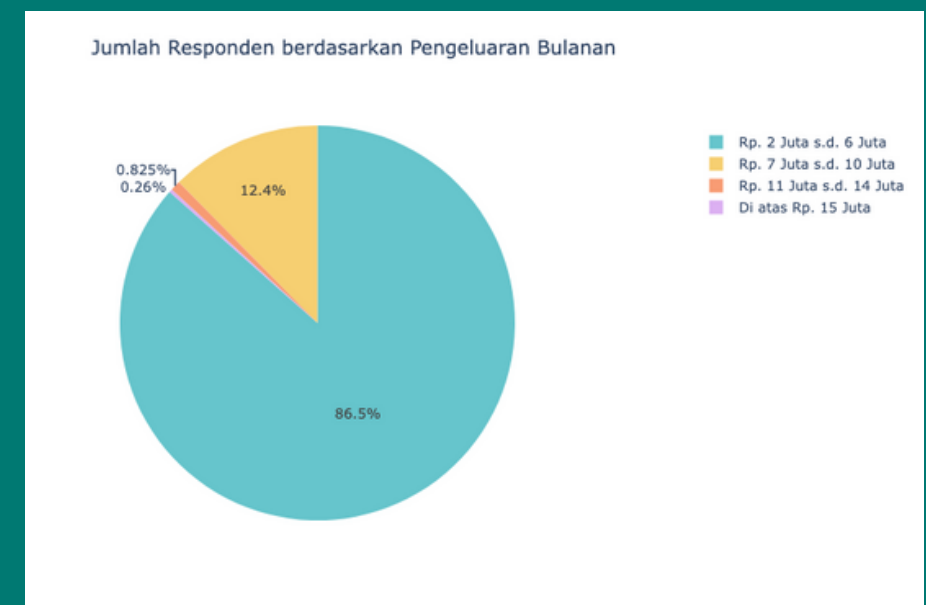
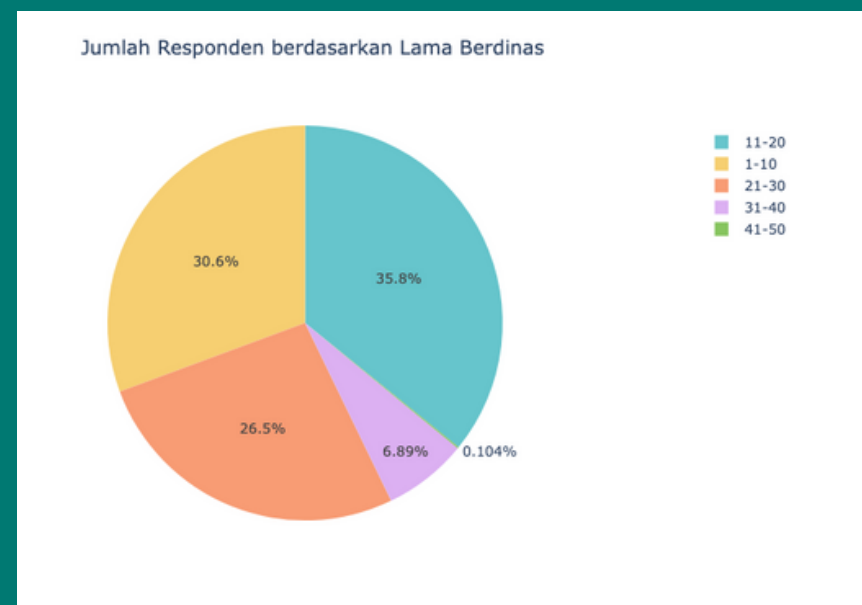
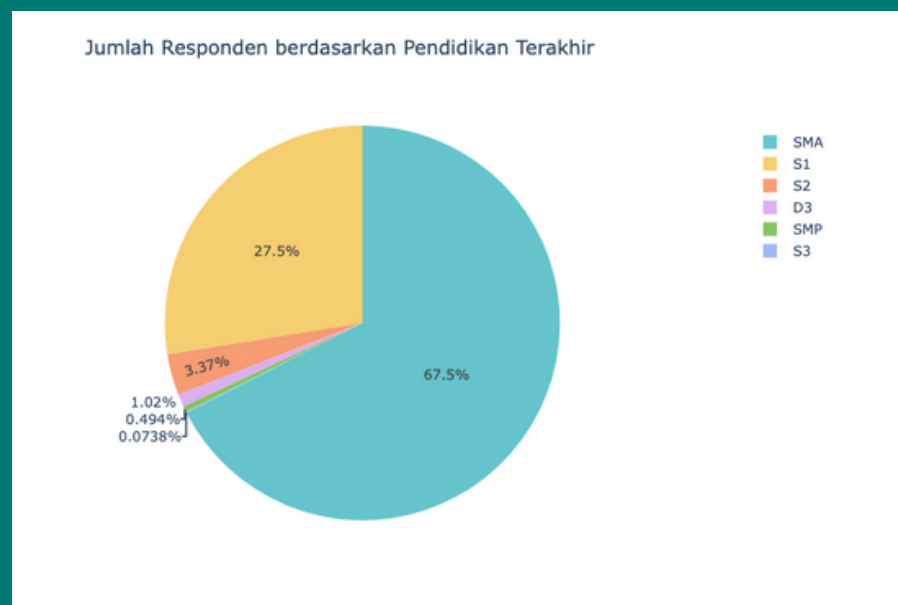
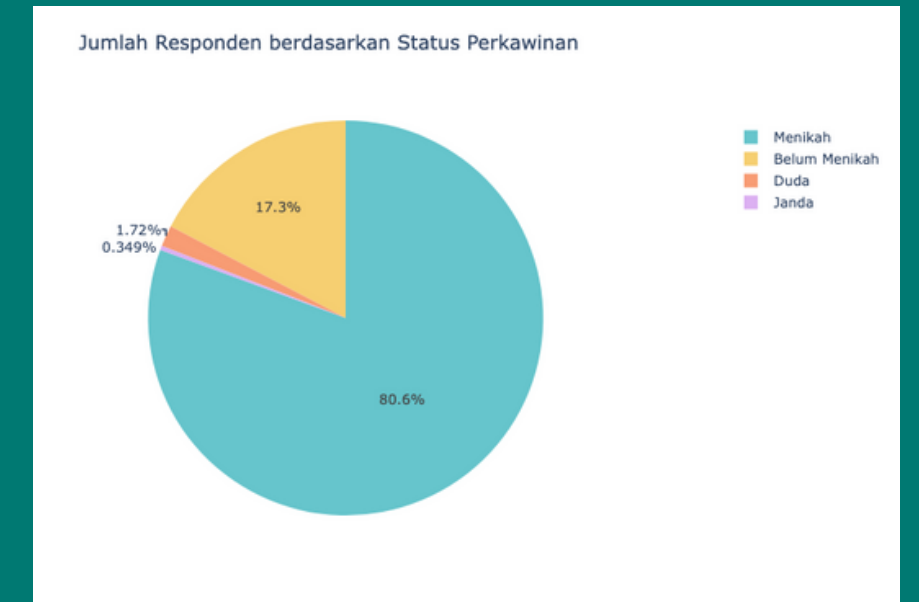
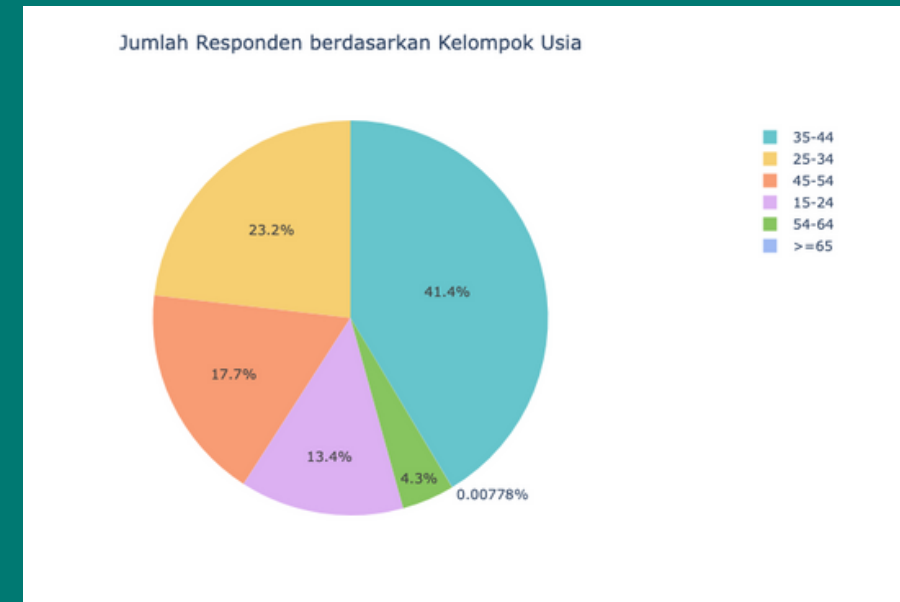
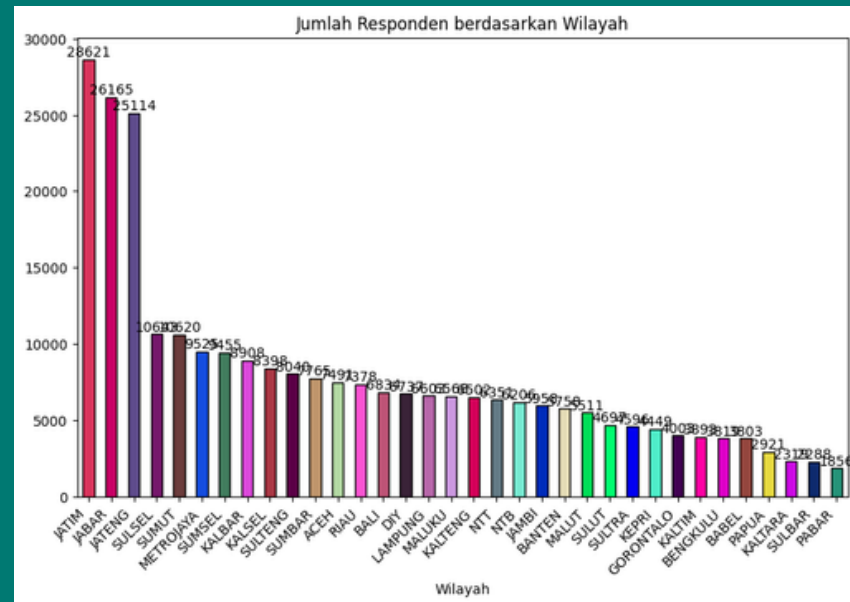


HUBUNGAN TINGKAT PENDIDIKAN TERHADAP PILIHAN GAYA HIDUP DAN KESEHATAN



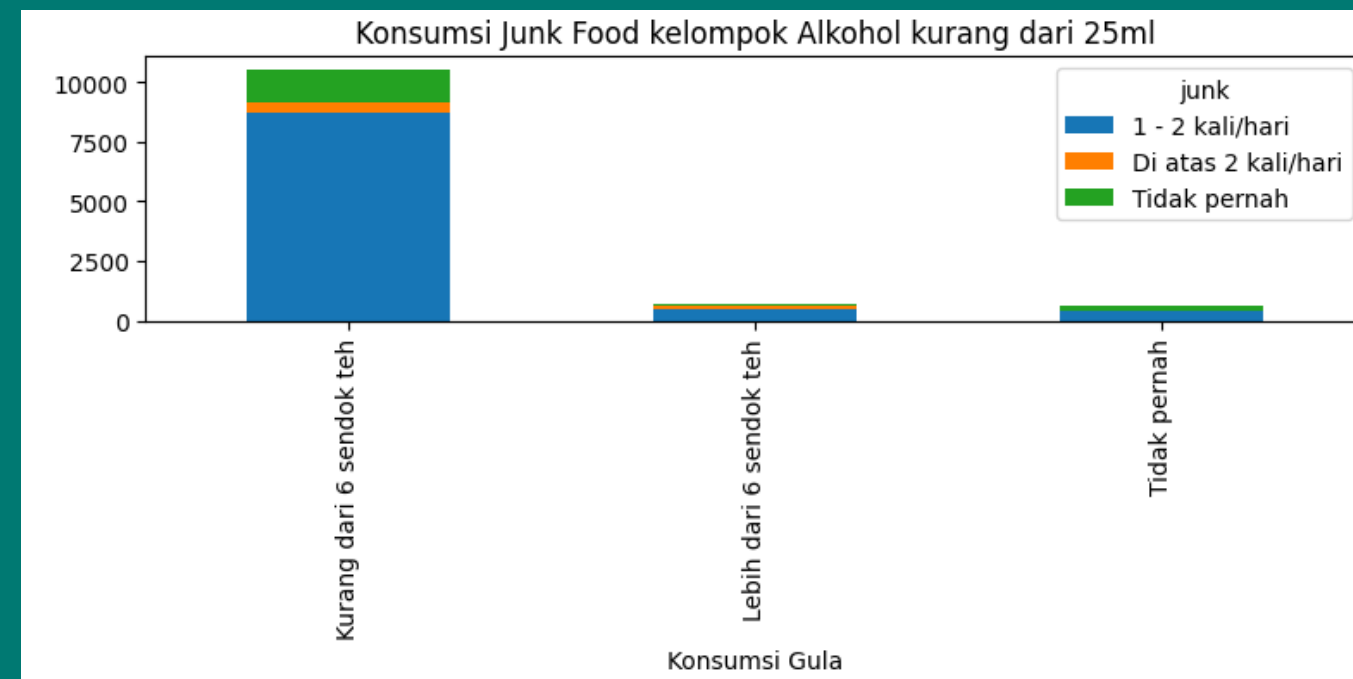
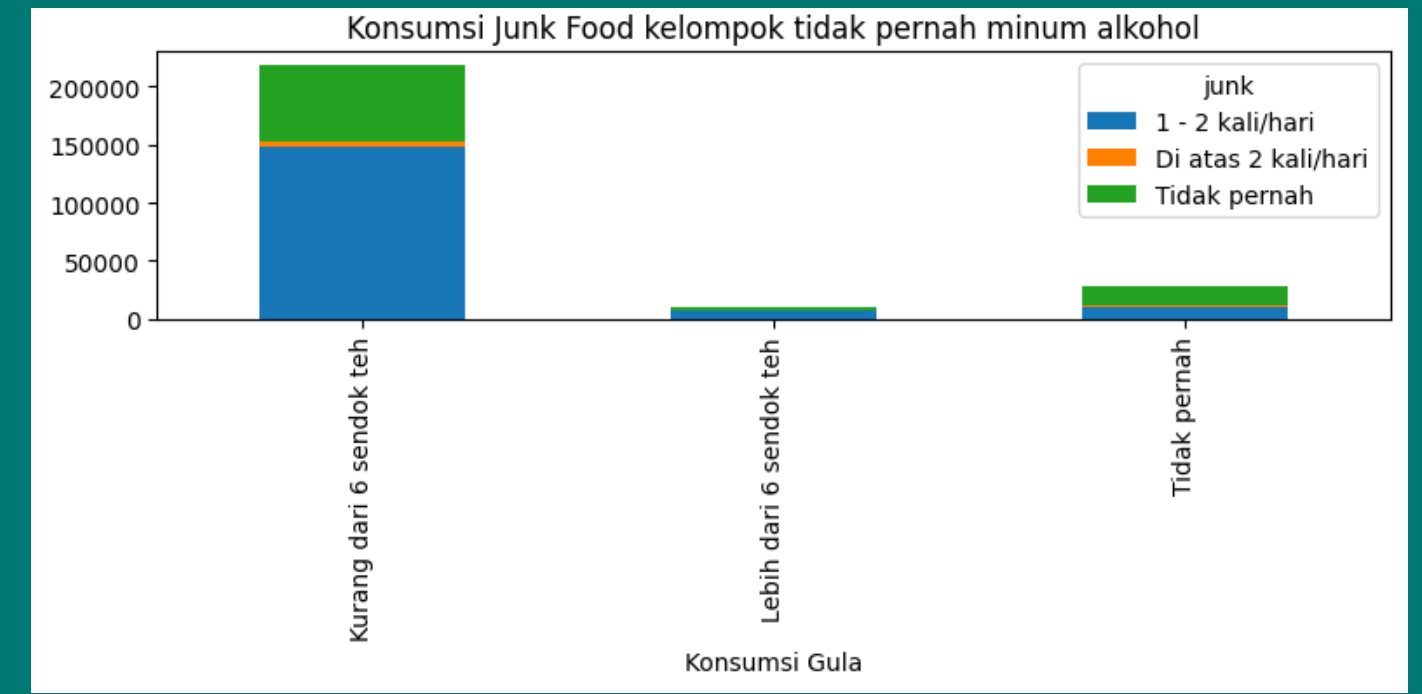
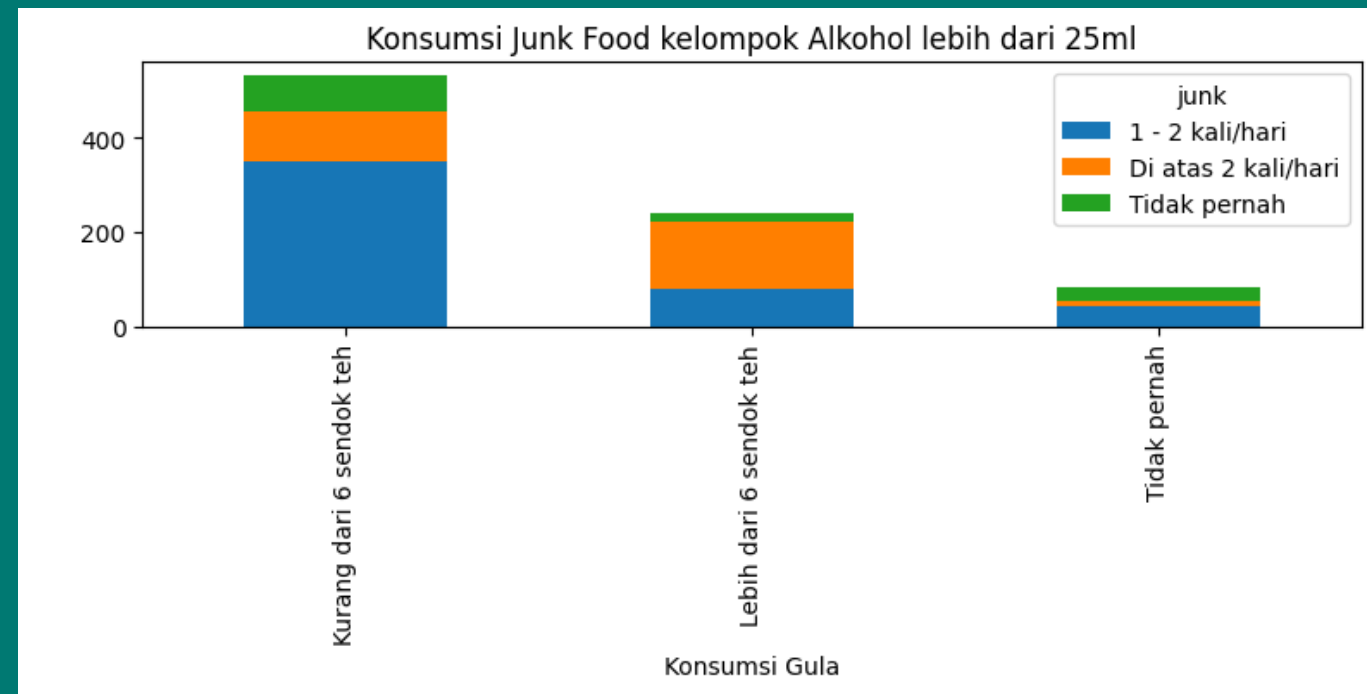


DATA DEMOGRAFIS



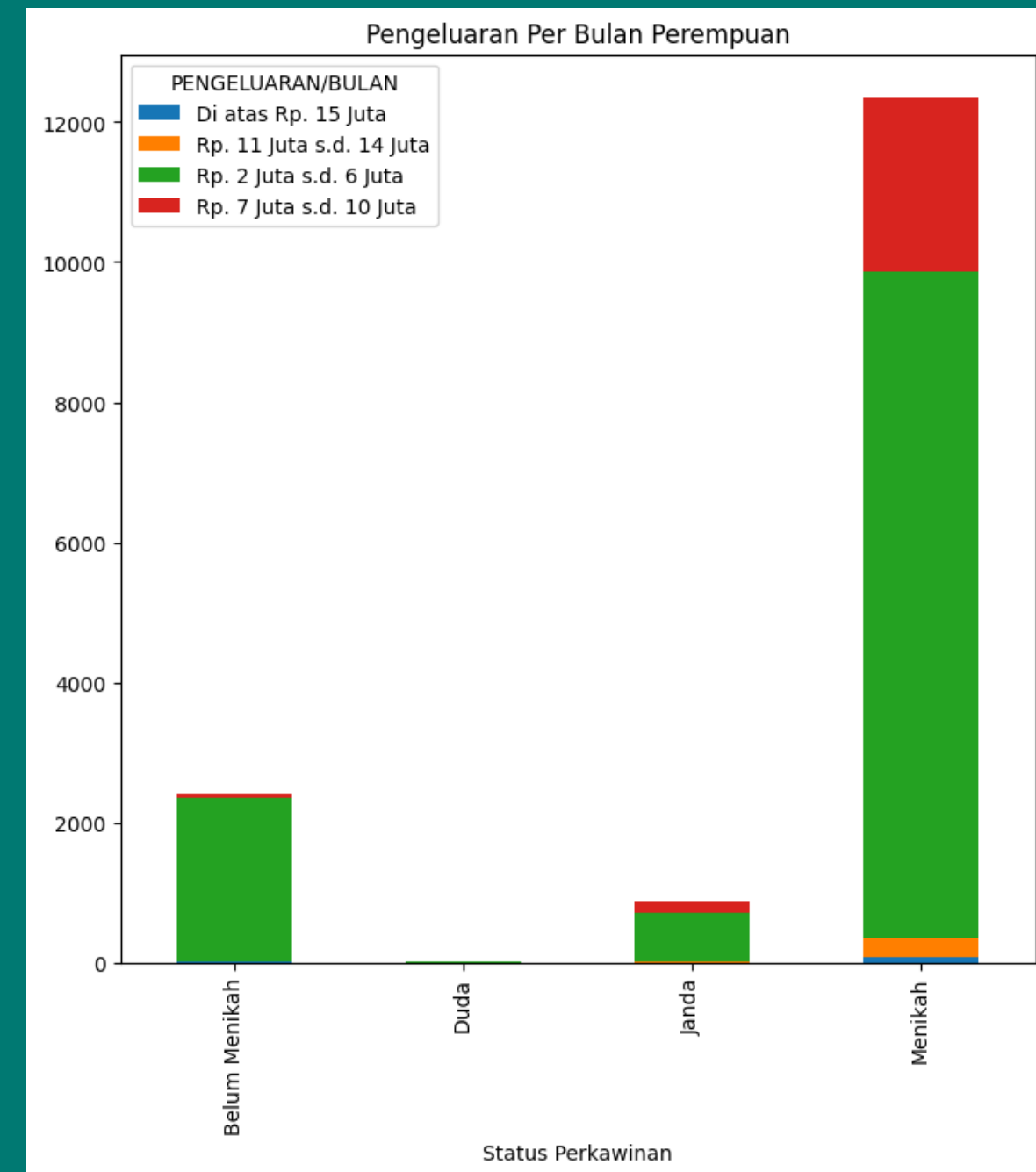
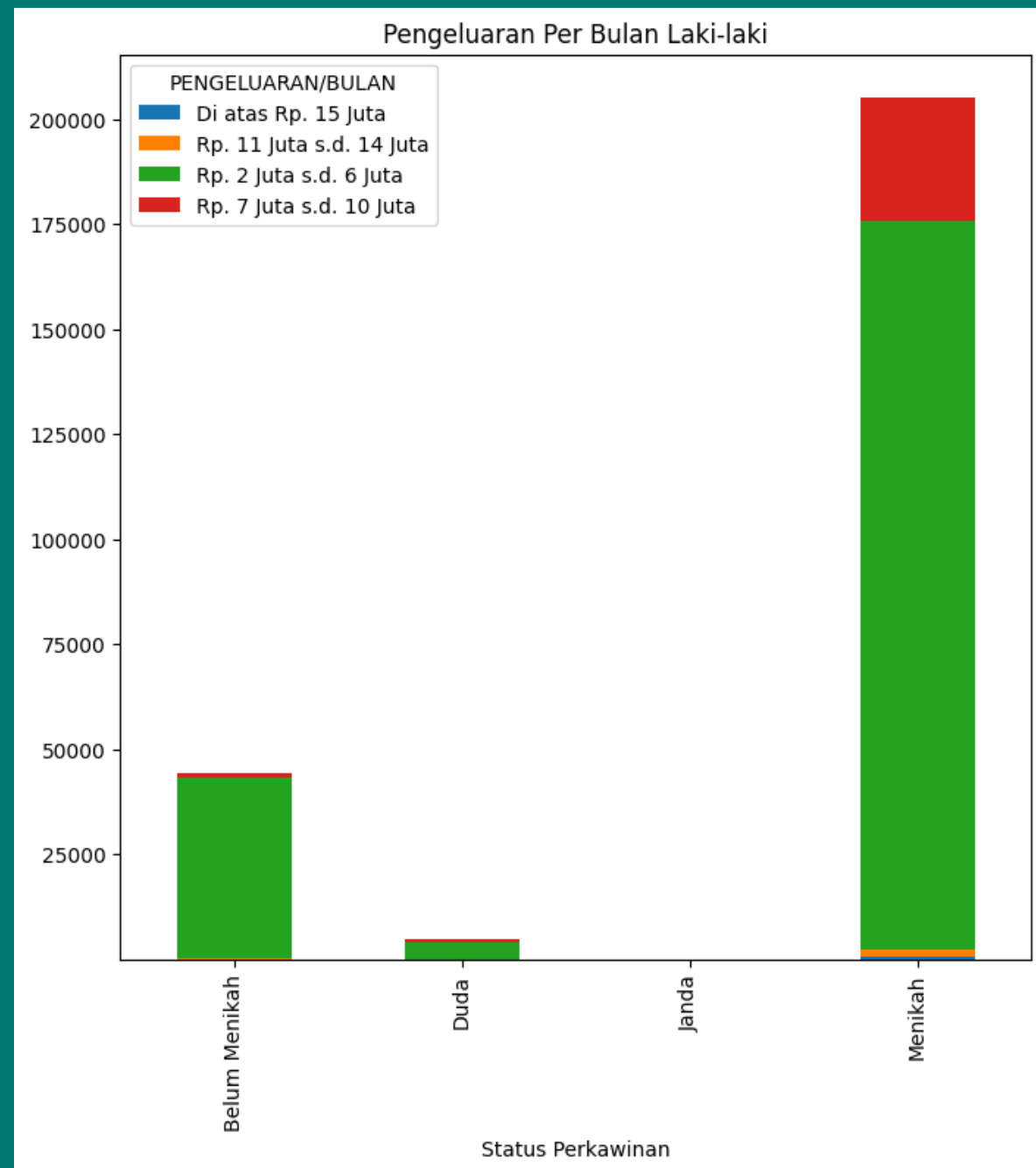


KECENDURUNGAN KONSUMSI MAKANAN DAN MINUMAN TIDAK SEHAT



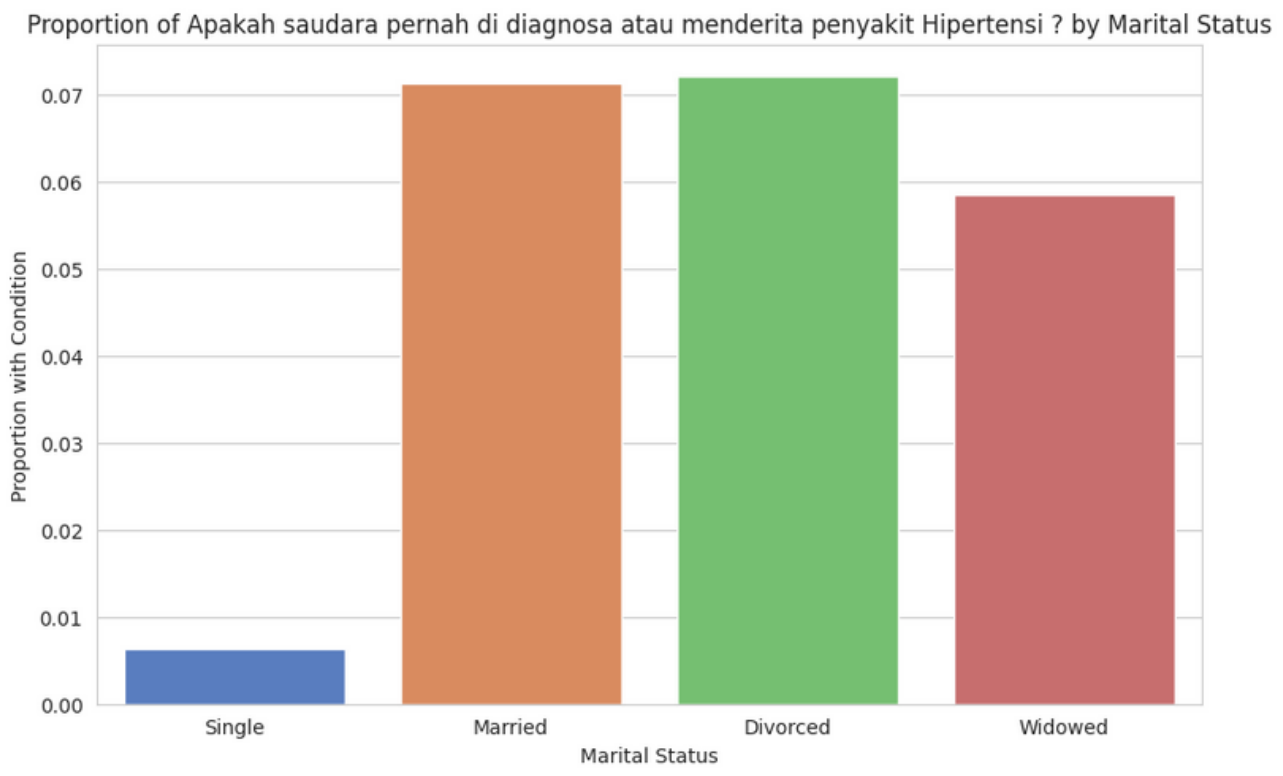
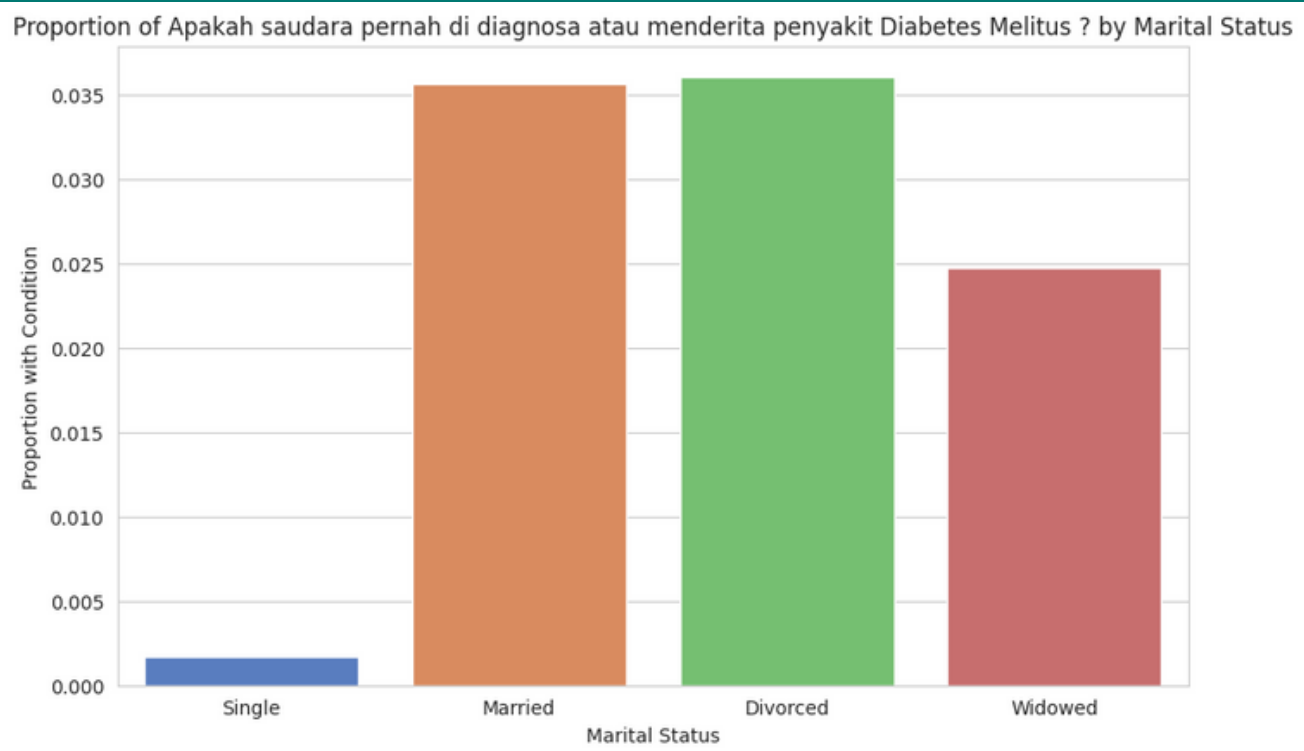
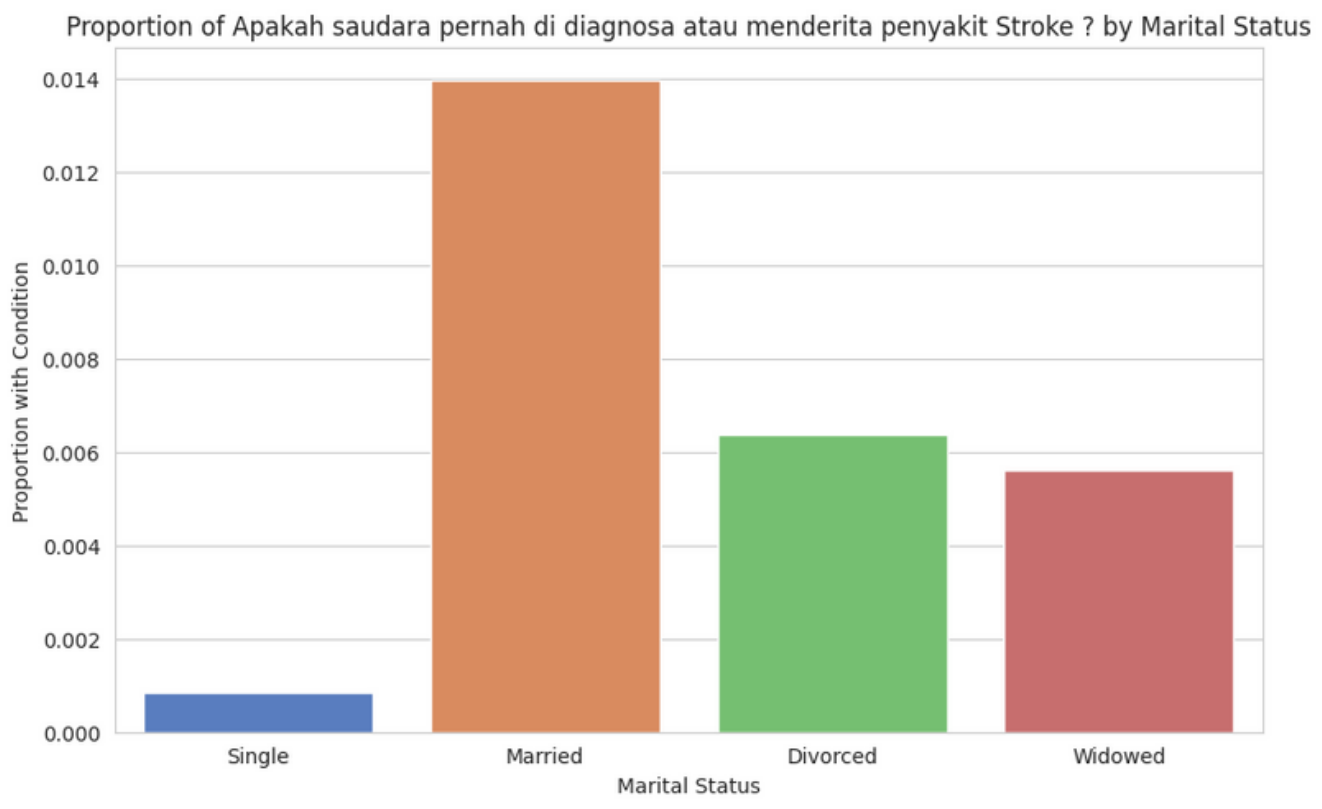
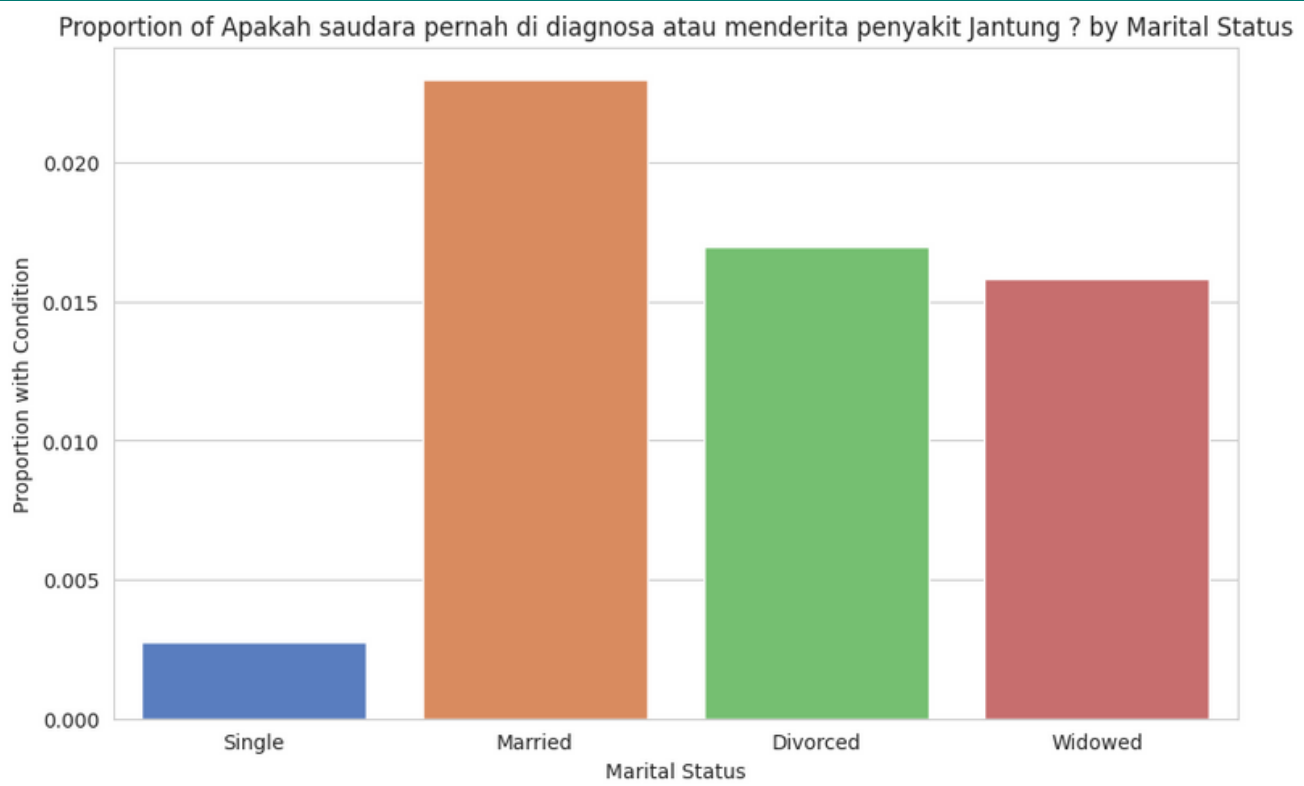


PENGELUARAN PER BULAN BERDASARKAN JENIS KELAMIN DAN STATUS PERKAWINAN

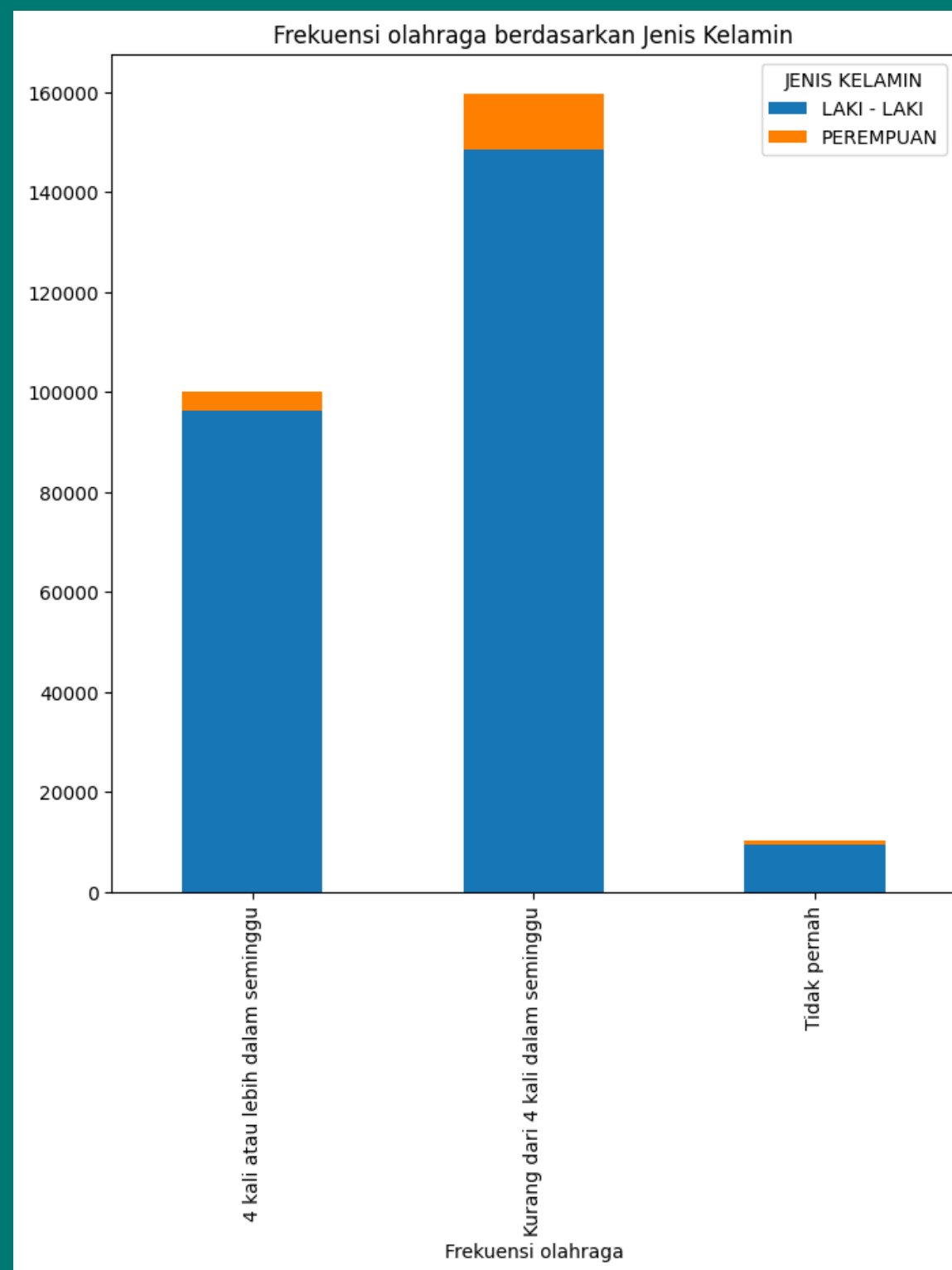




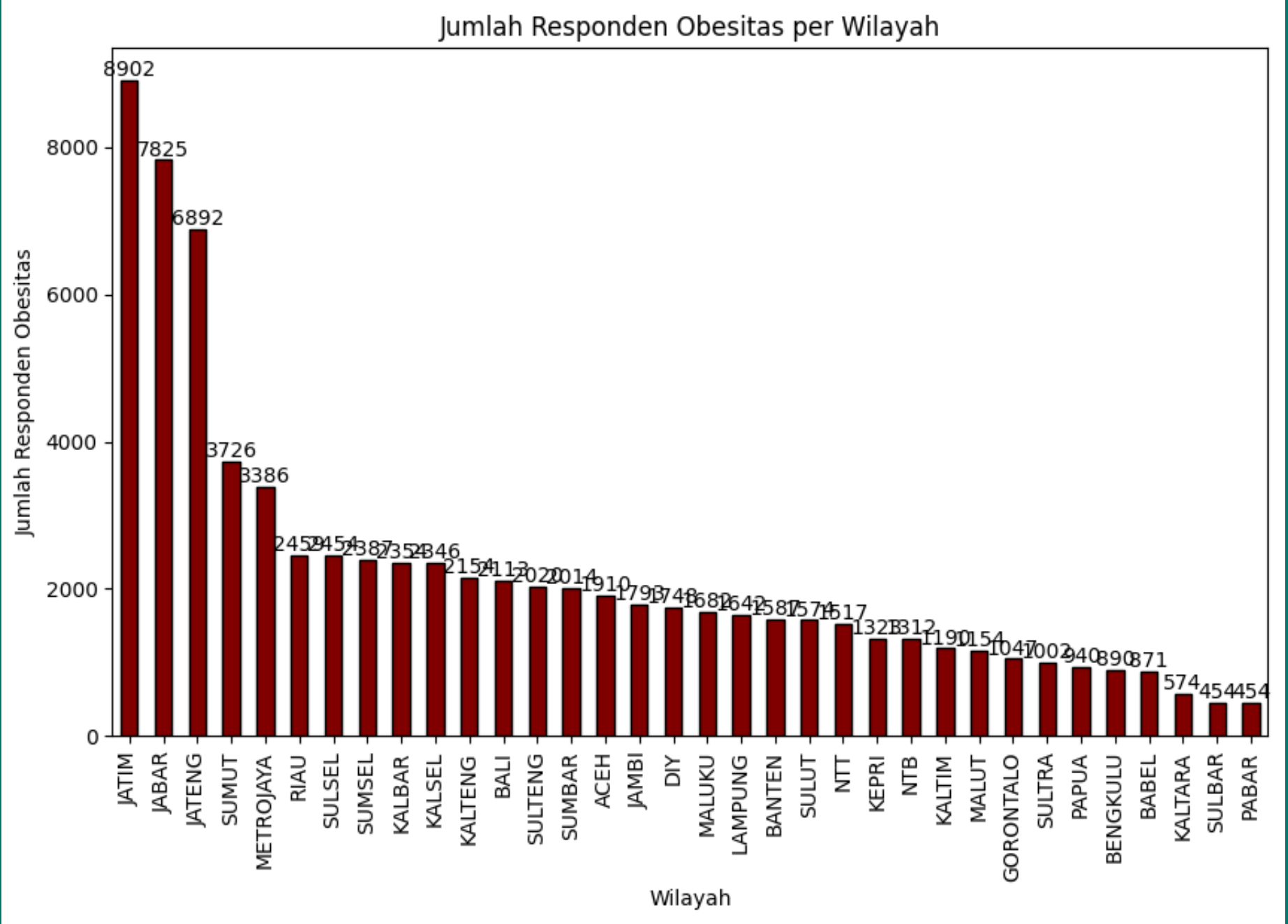
PERBEDAAN KONDISI KESEHATAN BERDASARKAN STATUS PERNIKAHAN



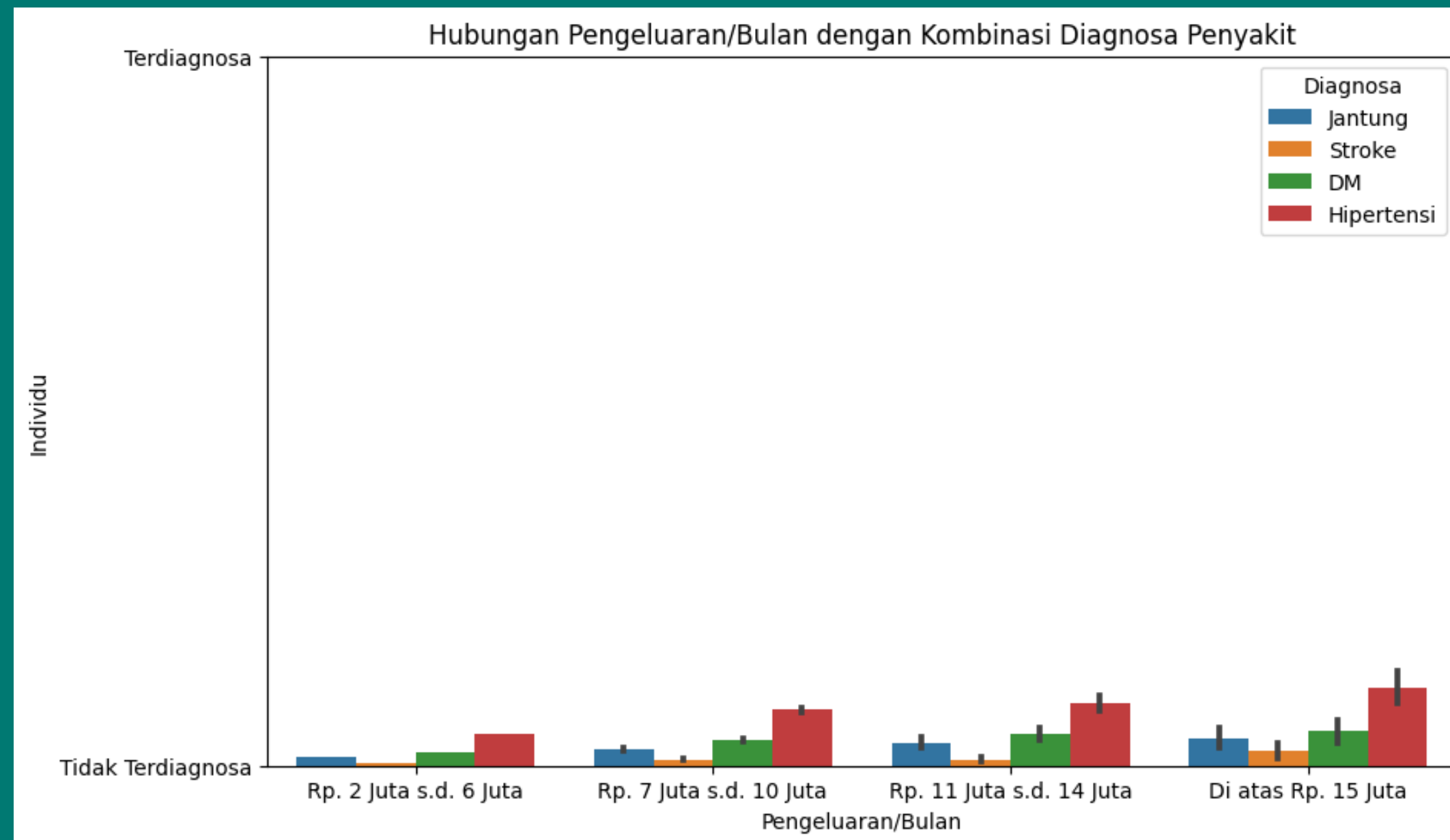
FREKUENSI OLAHRAGA BERDASARKAN JENIS KELAMIN



JUMLAH OBESITAS DI TIAP DAERAH

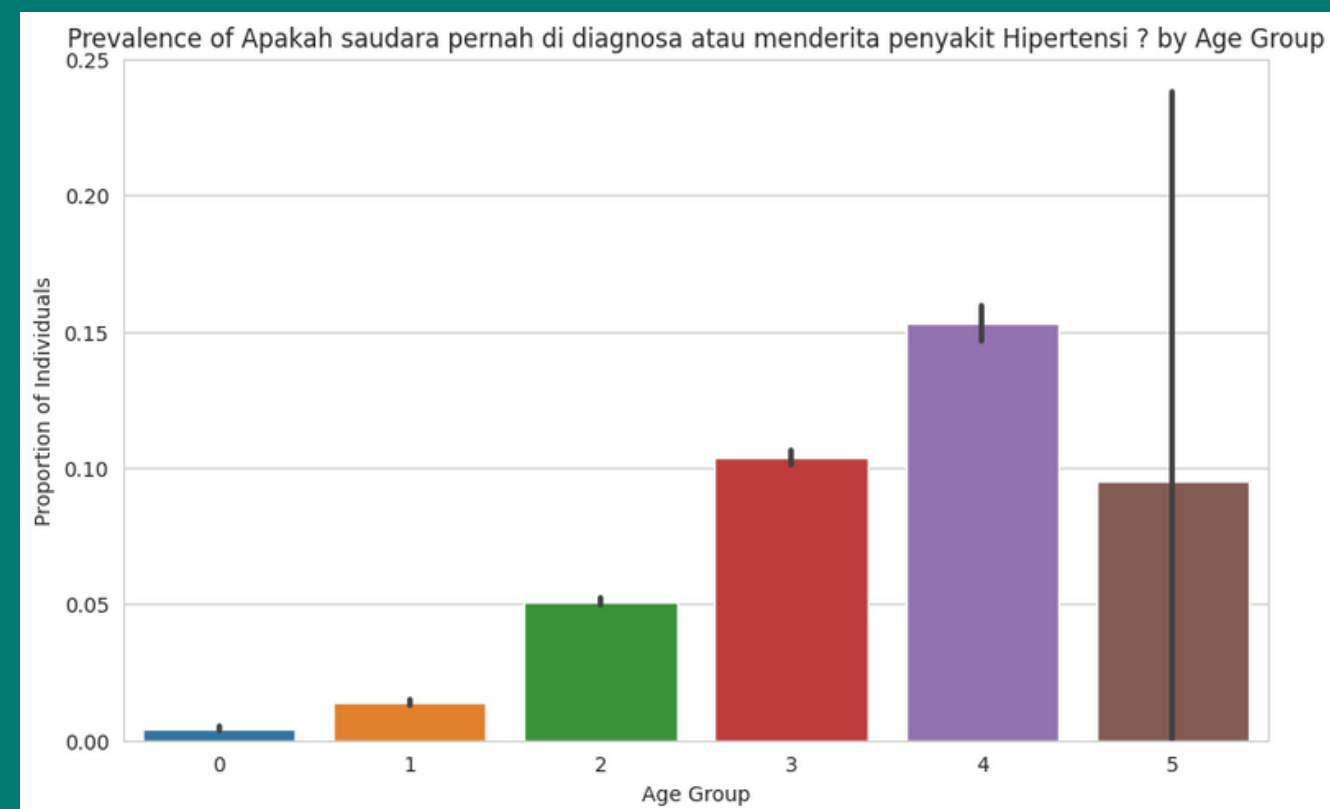
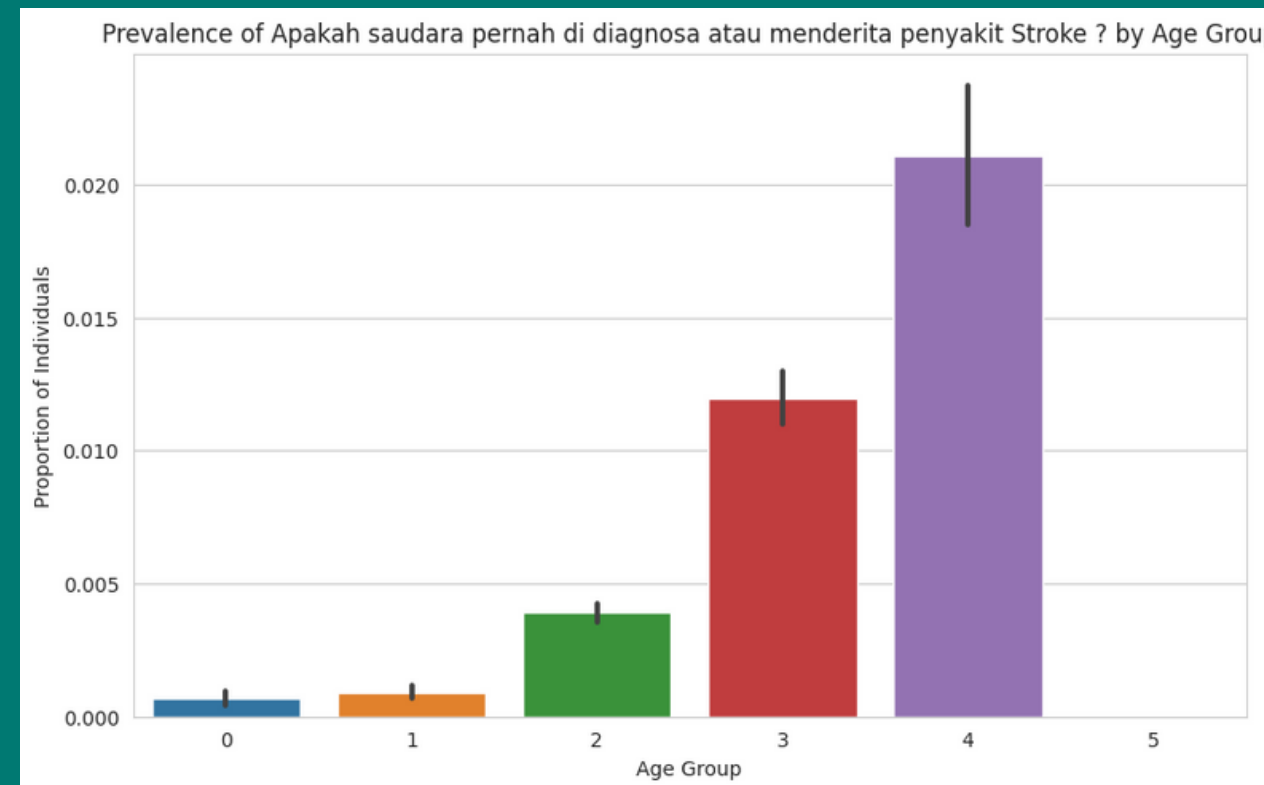
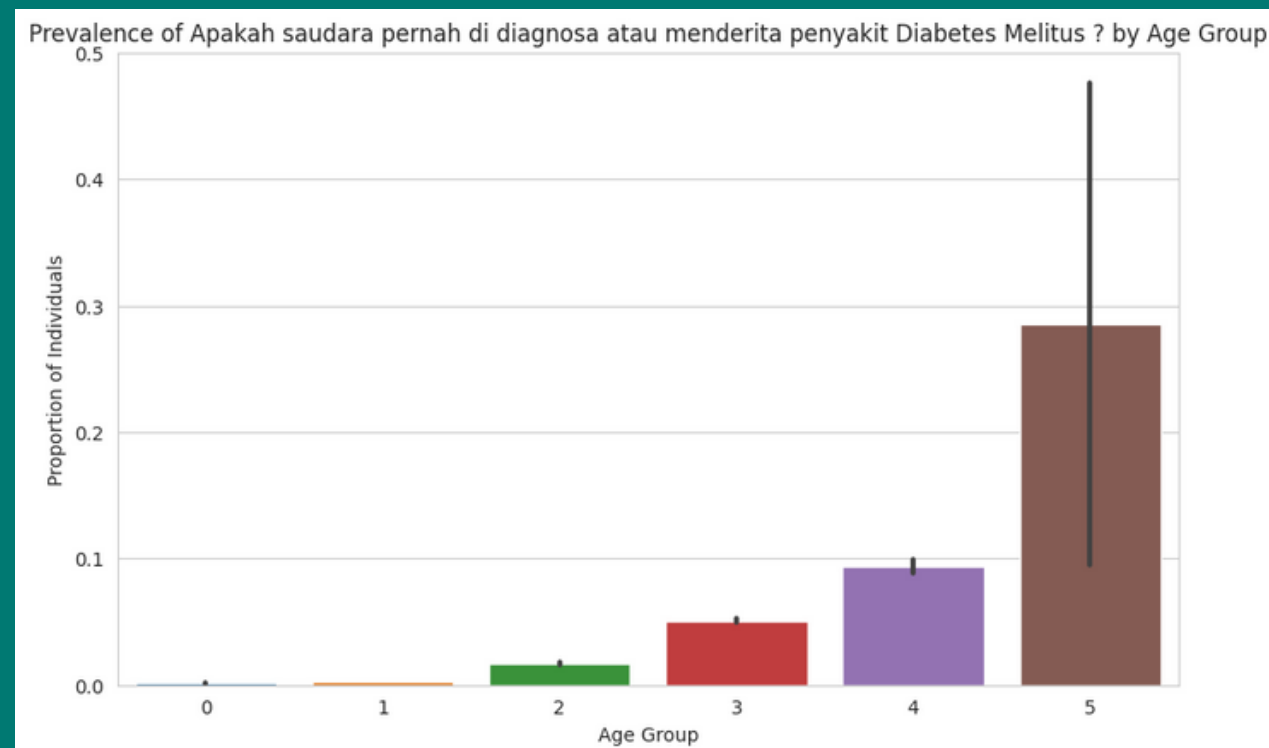
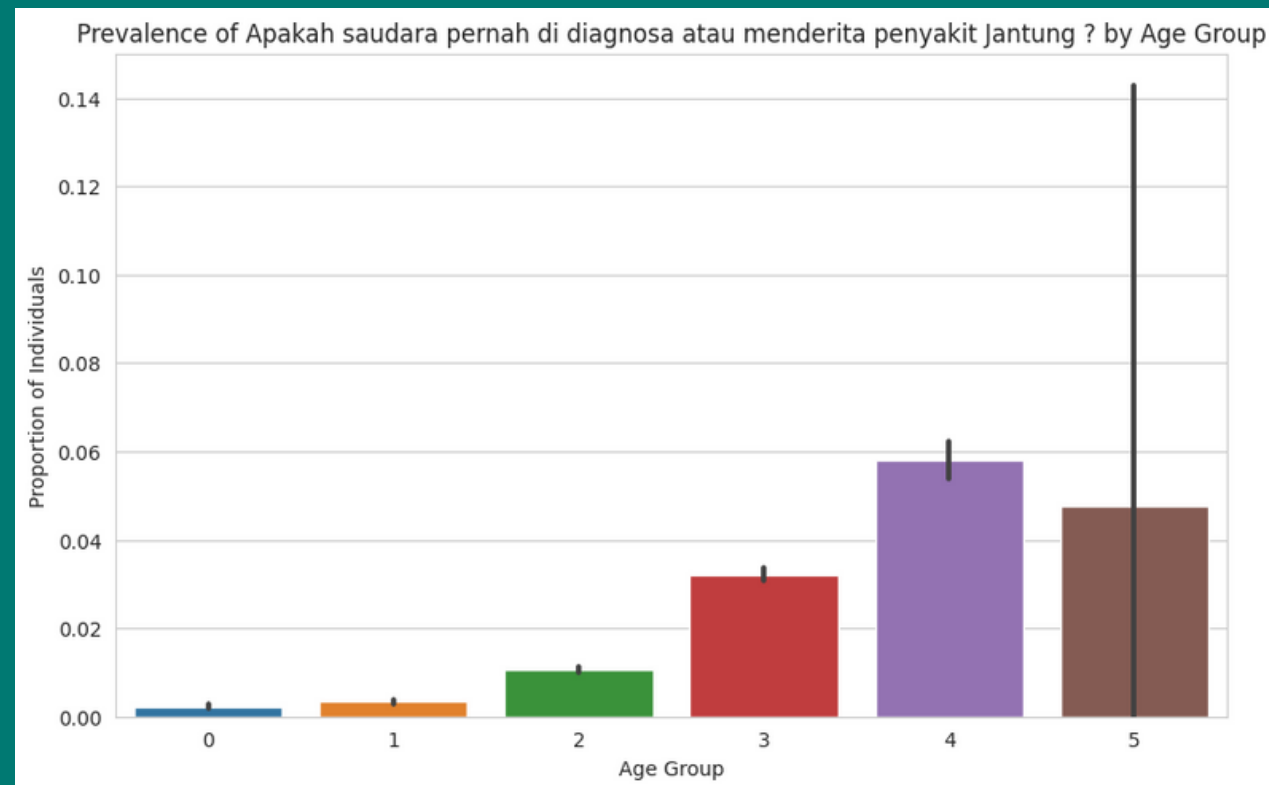


KECENDERUNGAN DIAGNOSA PENYAKIT (JANTUNG, STROKE, DIABETES MELITUS, DAN HIPERTENSI) BERDASARKAN PENGELUARAN/BULAN

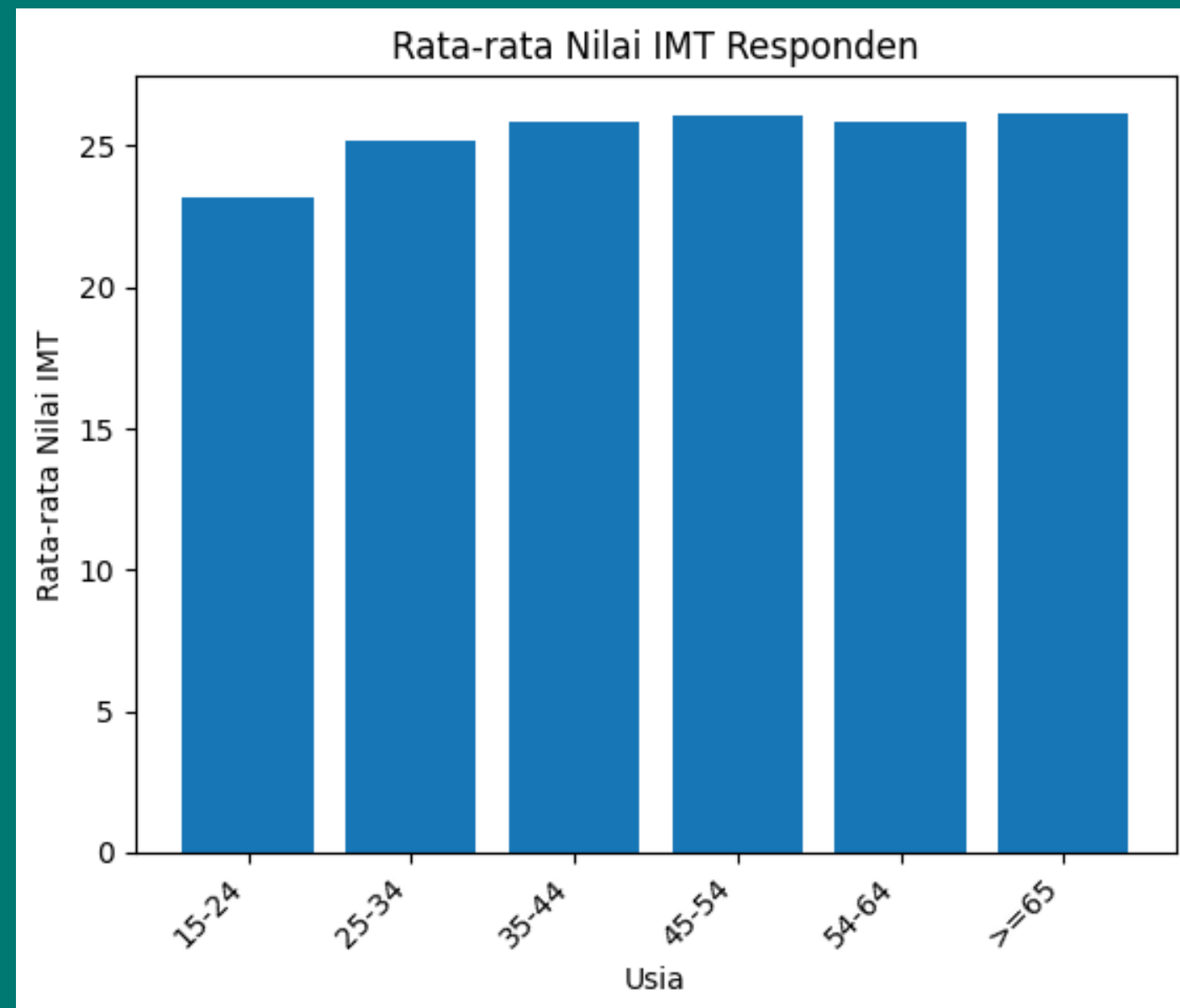




HUBUNGAN USIA DENGAN PENYAKIT TERKAIT GAYA HIDUP



ASOSIASI NILAI RATA-RATA IMT BERDASARKAN KELOMPOK USIA





UJI NORMALITAS

UJI	P value	
	NILAI IMT	SKOR KESEHATAN MENTAL
Shapiro-Wilk	0.0	0.0
Kolmogorov-Smirnov	0.0	0.0





MANN-WHITNEY TEST

	Nilai IMT		Skor Kesehatan Mental	
	Laki-laki	Perempuan	Laki-laki	Perempuan
Mean	25.42	24.73	19.49	19.48
p-value	2.12e-211		9.87e-05	





MODEL REGRESI

OLS Regression Results

Dep. Variable:	Total Skor	R-squared:	0.002			
Model:	OLS	Adj. R-squared:	0.002			
Method:	Least Squares	F-statistic:	226.9			
Date:	Thu, 21 Dec 2023	Prob (F-statistic):	3.80e-99			
Time:	03:44:58	Log-Likelihood:	-5.2894e+05			
No. Observations:	215836	AIC:	1.058e+06			
Df Residuals:	215833	BIC:	1.058e+06			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	12.2843	0.035	355.353	0.000	12.217	12.352
NILAI IMT	0.0427	0.002	21.299	0.000	0.039	0.047
JENIS KELAMIN_LAKI - LAKI	6.1306	0.020	309.880	0.000	6.092	6.169
JENIS KELAMIN_PEREMPUAN	6.1536	0.023	264.712	0.000	6.108	6.199
=====						
Omnibus:	160319.739	Durbin-Watson:	2.006			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3100168.100			
Skew:	3.468	Prob(JB):	0.00			
Kurtosis:	20.223	Cond. No.	2.94e+15			
=====						





THANK
YOU!



LINK PRESENTASI

