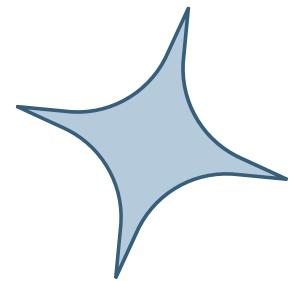


# Project UAS Data Mining

Anggota:

- 
- 
1. Aulia Shinta Madani
  2. Christi Novelina Panggabean
  3. Nadia Amelia Kirana
  4. Nadia Sukesi Sianipar
  5. Zahrah Aulia Putri

# INSIGHT 1

OLS Regression Results

<b>Dep. Variable:</b>	Gangguan_Mental	<b>R-squared:</b>	0.035
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.035
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	755.7
<b>Date:</b>	Thu, 21 Dec 2023	<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	12:30:05	<b>Log-Likelihood:</b>	-7.2202e+05
<b>No. Observations:</b>	289801	<b>AIC:</b>	1.444e+06
<b>Df Residuals:</b>	289786	<b>BIC:</b>	1.444e+06
<b>Df Model:</b>	14		
<b>Covariance Type:</b>	nonrobust		

Dibentuk model regresi dengan Gangguan Mental sebagai variabel respon, lalu Jenis\_Kelamin, Status\_Perkawinan, PengeluaranperBulan, serta Pendidikan\_Terakhir, USIA, dan Lama-Berdinas sebagai variabel prediktor.

	coef	std err	t	P> t	[0.025 0.975]
<b>Intercept</b>	1.6983	0.120	14.187	0.000	1.464 1.933
<b>C(Jenis_Kelamin)[T.PEREMPUAN]</b>	0.2234	0.022	10.343	0.000	0.181 0.266
<b>C(Status_Perkawinan)[T.Duda]</b>	1.3796	0.045	30.498	0.000	1.291 1.468
<b>C(Status_Perkawinan)[T.Janda]</b>	1.0666	0.082	13.020	0.000	0.906 1.227
<b>C(Status_Perkawinan)[T.Menikah]</b>	0.5675	0.019	30.538	0.000	0.531 0.604
<b>C(PengeluaranperBulan)[T.Rp. 11 Juta s.d. 14 Juta]</b>	-0.8214	0.120	-6.864	0.000	-1.056 -0.587
<b>C(PengeluaranperBulan)[T.Rp. 2 Juta s.d. 6 Juta]</b>	-1.5761	0.105	-15.044	0.000	-1.781 -1.371
<b>C(PengeluaranperBulan)[T.Rp. 7 Juta s.d. 10 Juta]</b>	-1.2417	0.106	-11.761	0.000	-1.449 -1.035
<b>C(Pendidikan_Terakhir)[T.S1]</b>	-0.3497	0.049	-7.180	0.000	-0.445 -0.254
<b>C(Pendidikan_Terakhir)[T.S2]</b>	-0.7862	0.056	-14.029	0.000	-0.896 -0.676
<b>C(Pendidikan_Terakhir)[T.S3]</b>	-0.5890	0.204	-2.884	0.004	-0.989 -0.189
<b>C(Pendidikan_Terakhir)[T.SMA]</b>	0.3483	0.048	7.212	0.000	0.254 0.443
<b>C(Pendidikan_Terakhir)[T.SMP]</b>	0.8328	0.090	9.301	0.000	0.657 1.008
<b>USIA</b>	0.0175	0.001	13.656	0.000	0.015 0.020
<b>Lama_Berdinas</b>	0.0093	0.001	7.225	0.000	0.007 0.012
<b>Omnibus:</b>	217534.491	<b>Durbin-Watson:</b>	1.955		
<b>Prob(Omnibus):</b>	0.000	<b>Jarque-Bera (JB):</b>	4638421.882		
<b>Skew:</b>	3.481	<b>Prob(JB):</b>	0.00		
<b>Kurtosis:</b>	21.321	<b>Cond. No.</b>	1.71e+03		

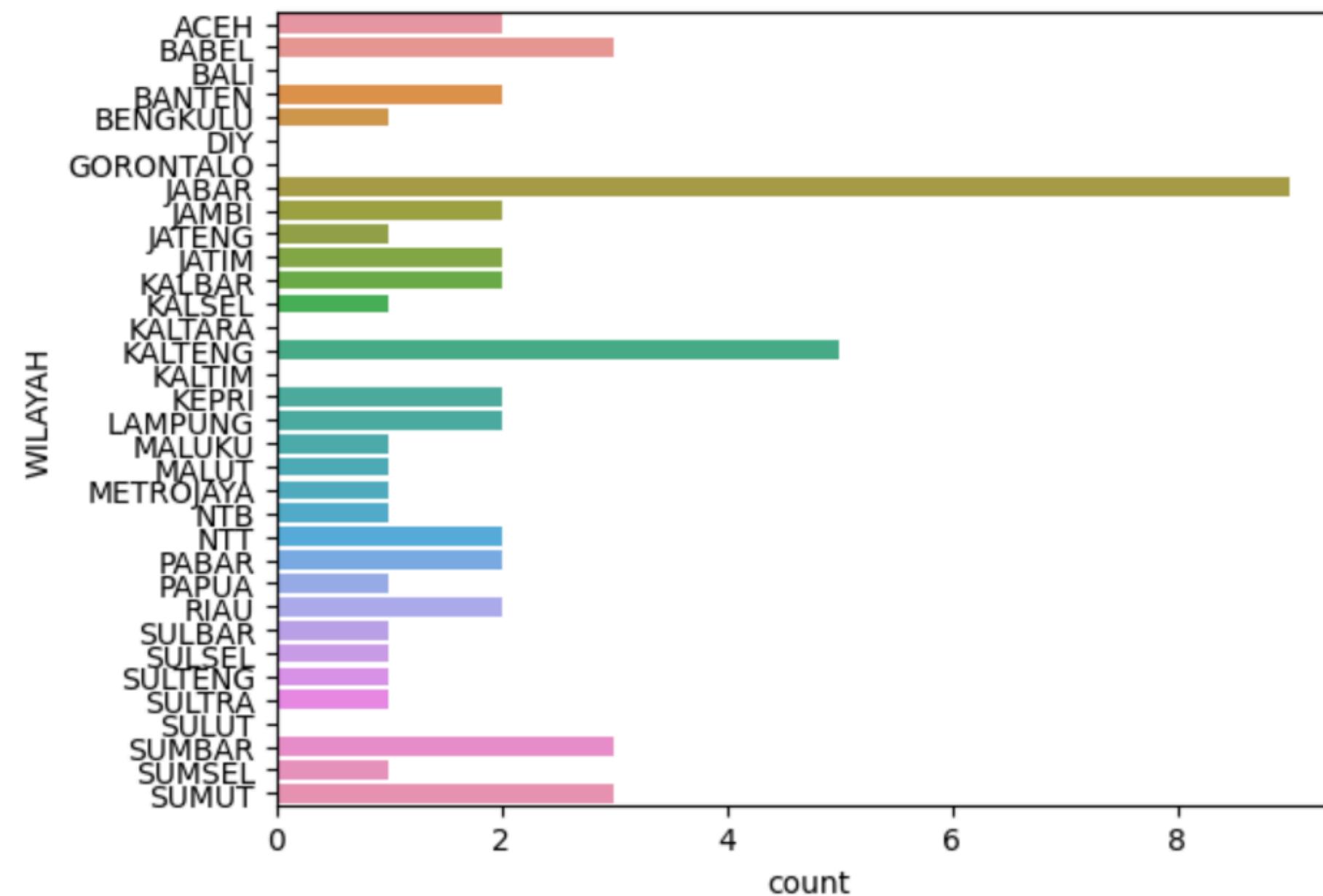
## INSIGHT 1

Dari hasil regresi, terlihat bahwa semua parameter signifikan dengan tingkat signifikansi 5% sehingga beberapa informasi ini signifikan, yaitu:

- Terdapat kecenderungan bahwa perempuan memiliki tingkat gangguan mental lebih tinggi daripada laki-laki.
- Terdapat kecenderungan bahwa duda memiliki tingkat gangguan mental lebih tinggi.
- Individu dengan pengeluaran lebih tinggi cenderung memiliki tingkat gangguan mental yang lebih tinggi juga.
- Individu dengan tingkat pendidikan terakhir SMP cenderung memiliki tingkat gangguan mental lebih tinggi, sedangkan individu dengan tingkat pendidikan terakhir S2 cenderung memiliki tingkat gangguan mental lebih rendah.
- Semakin bertambahnya usia/lama berdinasti cenderung meningkatkan tingkat gangguan mental.

# INSIGHT 1

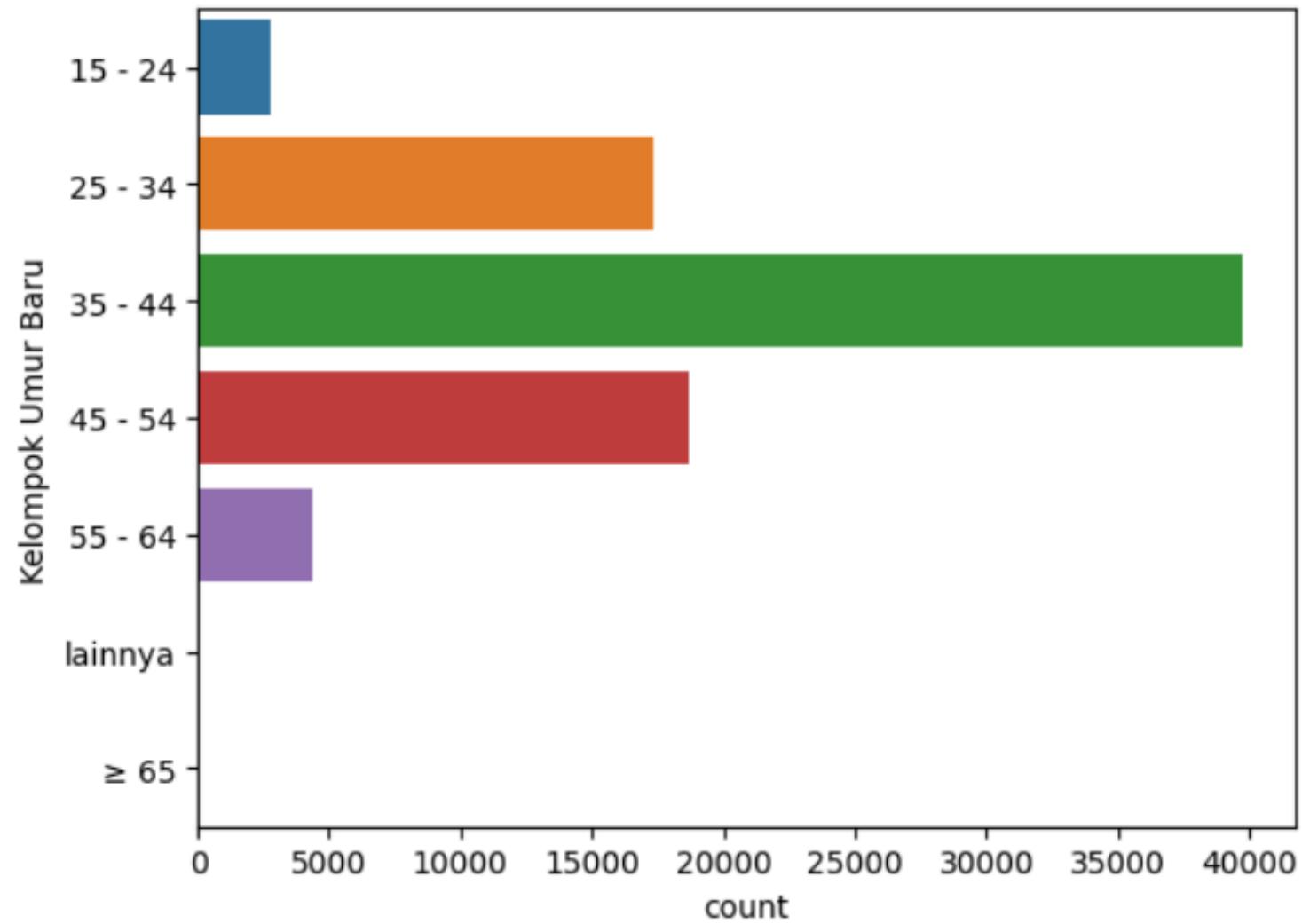
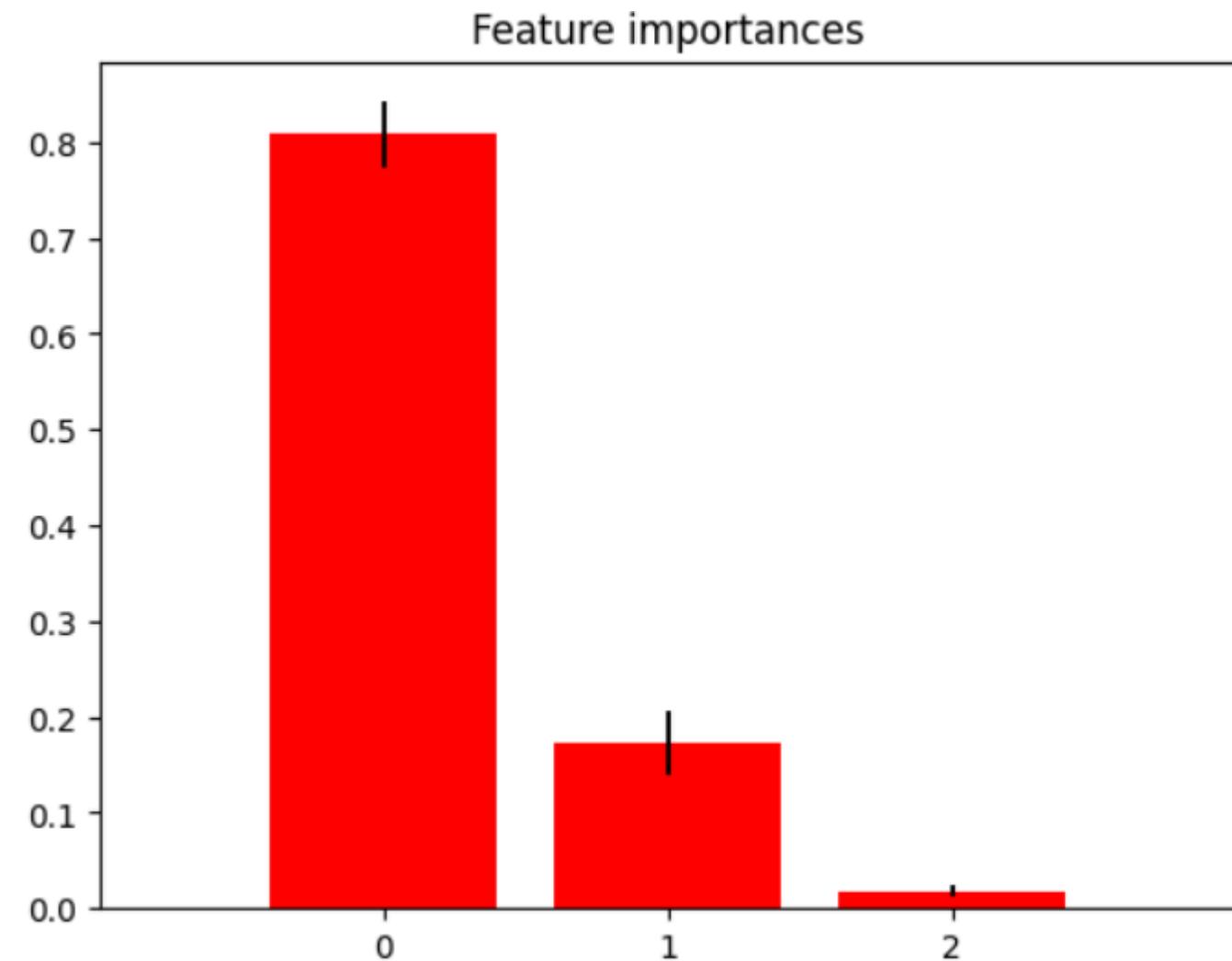
Dari model regresi diatas, juga dapat diketahui bahwa **individu dengan pengeluaran di atas Rp 15 juta per bulan cenderung memiliki tingkat gangguan mental lebih tinggi.**



Pada grafik, terlihat bahwa individu yang memiliki **pengeluaran di atas Rp 15 juta dengan tingkat gangguan mental yang tinggi mayoritas berada di daerah Jawa Barat**. Pada tahun 2023, UMR tertinggi di Jawa Barat sebesar Rp 5.1 juta, dimana nilai ini sangat jauh dari pengeluaran sebesar Rp 15 juta. Terdapat kemungkinan bahwa tingkat gangguan mental di Jawa Barat disebabkan oleh tidak cukupnya pendapatan seseorang untuk menutupi pengeluaran per bulannya.

## INSIGHT 2

Stres eating merupakan kecenderungan seseorang untuk mengalihkan rasa cemas, sedih, atau marah dengan makan berlebih. Dibuat hipotesis bahwa tingkat gangguan mental merupakan penyebab individu mengalami obesitas.



Variabel USIA merupakan variabel yang paling penting dalam model klasifikasi yang diperoleh, yaitu sebesar 80%. Sedangkan, variabel Gangguan Mental berperan sebesar 17%. Hal ini mengindikasikan bahwa variabel USIA berperan lebih penting dari variabel Gangguan Mental dalam pengklasifikasian 'Obesitas'. Dari grafik dibawah, terlihat bahwa individu yang mengalami obesitas paling banyak berusia di rentang 35 - 44 tahun.

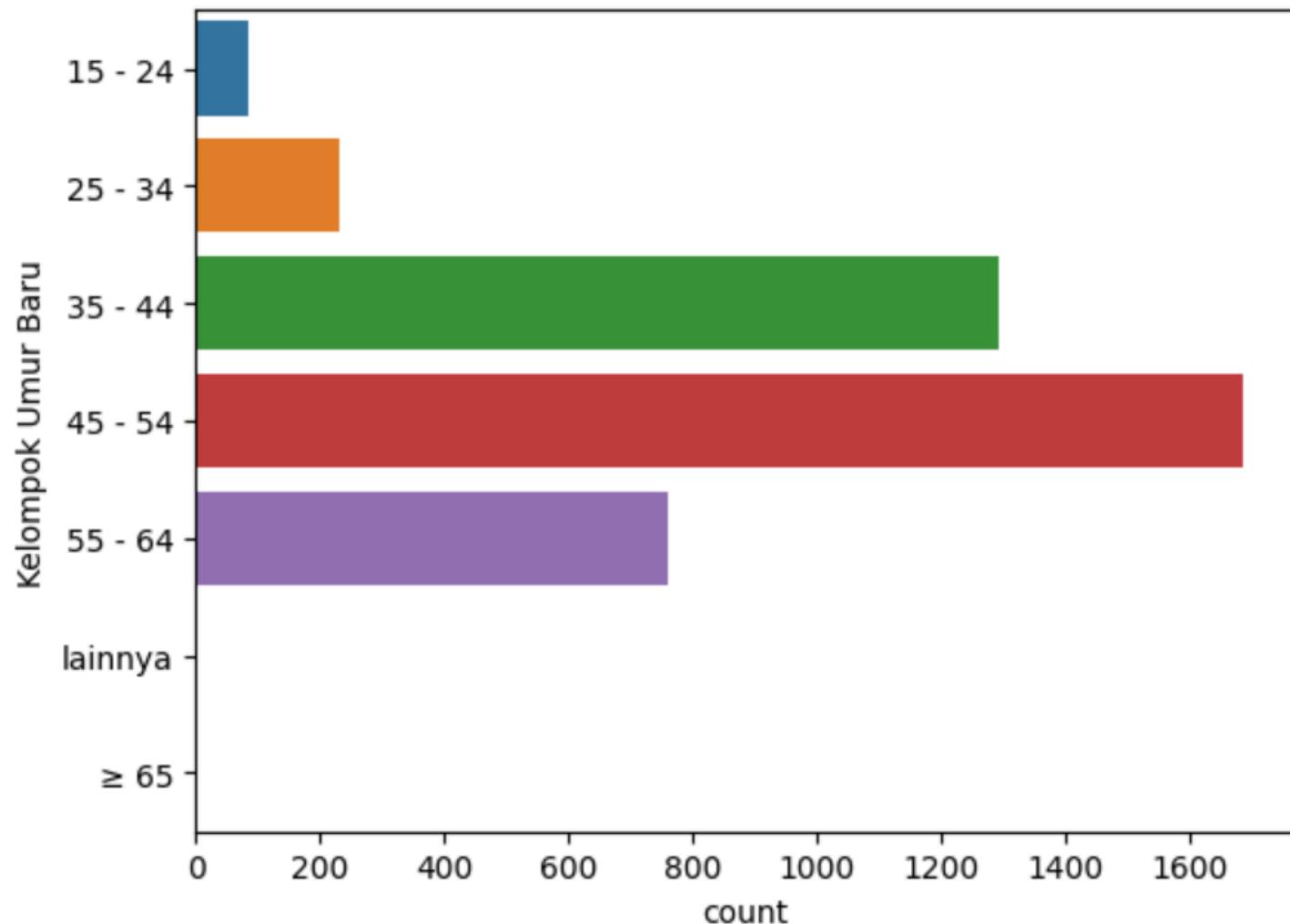
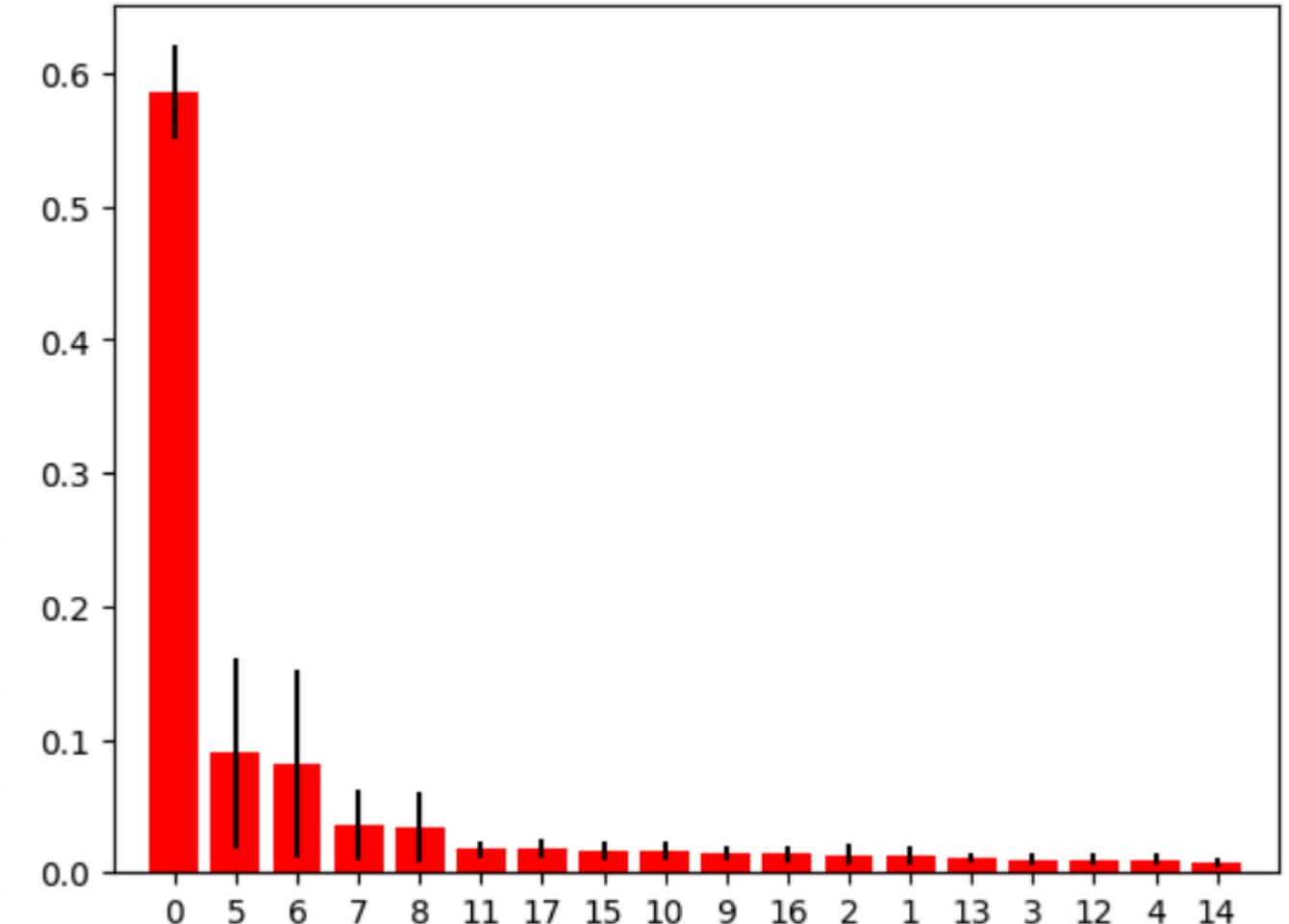
# INSIGHT 3

## Model klasifikasi 'Penyakit Jantung'

Feature ranking:

1. feature 0 (0.586178)
2. feature 5 (0.089953)
3. feature 6 (0.082064)
4. feature 7 (0.035791)
5. feature 8 (0.034380)
6. feature 11 (0.017912)
7. feature 17 (0.017363)
8. feature 15 (0.016993)
9. feature 10 (0.016547)
10. feature 9 (0.014532)
11. feature 16 (0.013797)
12. feature 2 (0.013301)
13. feature 1 (0.012275)
14. feature 13 (0.011046)
15. feature 3 (0.010119)
16. feature 12 (0.010086)
17. feature 4 (0.009686)
18. feature 14 (0.007979)

Feature importances



Variabel **USIA** merupakan variabel yang paling penting dalam model klasifikasi penyakit jantung, yaitu sebesar 58%, dimana pada grafik kedua terlihat bahwa **majoritas individu yang memiliki penyakit jantung berusia di rentang 45 - 54 tahun**. Di sisi lain, penyakit hipertensi dan diabetes melitus juga berperan dalam model sebesar masing-masing 8% dan 3%.

## INSIGHT 3

Mengenai kegiatan individu, berikut peran beberapa variabel dalam pengklasifikasian penyakit jantung:

- Variabel tidak pernah melakukan aktivitas fisik berperan sekitar 1.7%
- Variabel junkfood 1-2 kali/sehari berperan sekitar 1.6%

Dari informasi tersebut, mayoritas individu yang terkena penyakit jantung adalah individu yang tidak pernah melakukan aktivitas fisik, mengonsumsi junkfood 1-2 kali/sehari, memiliki riwayat hipertensi dan diabetes melitus.

Oleh karena itu, untuk mengurangi risiko terkena penyakit jantung, seorang individu perlu untuk mengurangi konsumsi junkfood, lebih banyak melakukan aktivitas fisik/berolahraga, dan menjaga konsumsi gula dengan tujuan mengurangi risiko penyakit diabetes melitus.

# INSIGHT 4

		OLS Regression Results			coef	std err	t	P> t	[0.025 0.975]
<b>Dep. Variable:</b>	Diare	<b>R-squared:</b>	0.006	Intercept	0.1403	0.003	46.160	0.000	0.134 0.146
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.006	C(Air_bersih)[T.Sumur/Kali]	-0.0070	0.001	-5.913	0.000	-0.009 -0.005
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	164.6	C(Air_bersih)[T.Tadah hujan]	0.0303	0.007	4.562	0.000	0.017 0.043
<b>Date:</b>	Thu, 21 Dec 2023	<b>Prob (F-statistic):</b>	0.00	C(Air_bersih)[T.Tidak ada]	-0.0262	0.010	-2.701	0.007	-0.045 -0.007
<b>Time:</b>	18:48:24	<b>Log-Likelihood:</b>	-74643.	C(Cuci_tangan)[T.Selalu]	-0.0416	0.002	-21.429	0.000	-0.045 -0.038
<b>No. Observations:</b>	289801	<b>AIC:</b>	1.493e+05	C(Cuci_tangan)[T.Tidak/Jarang sekali]	-0.0288	0.010	-2.893	0.004	-0.048 -0.009
<b>Df Residuals:</b>	289790	<b>BIC:</b>	1.494e+05	C(Jarak_tempat_sampah)[T.Lebih dari 10 meter]	-0.0042	0.001	-3.078	0.002	-0.007 -0.002
<b>Df Model:</b>	10			C(Jarak_tempat_sampah)[T.Tidak ada]	-0.0147	0.002	-7.782	0.000	-0.018 -0.011
<b>Covariance Type:</b>	nonrobust			C(junkfood)[T.Di atas 2 kali/hari]	0.0237	0.004	6.751	0.000	0.017 0.031
				C(junkfood)[T.Tidak pernah]	-0.0341	0.001	-27.176	0.000	-0.037 -0.032
				<b>USIA</b>	0.0007	5.89e-05	11.738	0.000	0.001 0.001
				<b>Omnibus:</b>	133900.690	<b>Durbin-Watson:</b>	1.678		
				<b>Prob(Omnibus):</b>	0.000	<b>Jarque-Bera (JB):</b>	494840.699		
				<b>Skew:</b>	2.458	<b>Prob(JB):</b>	0.00		
				<b>Kurtosis:</b>	7.100	<b>Cond. No.</b>	664.		

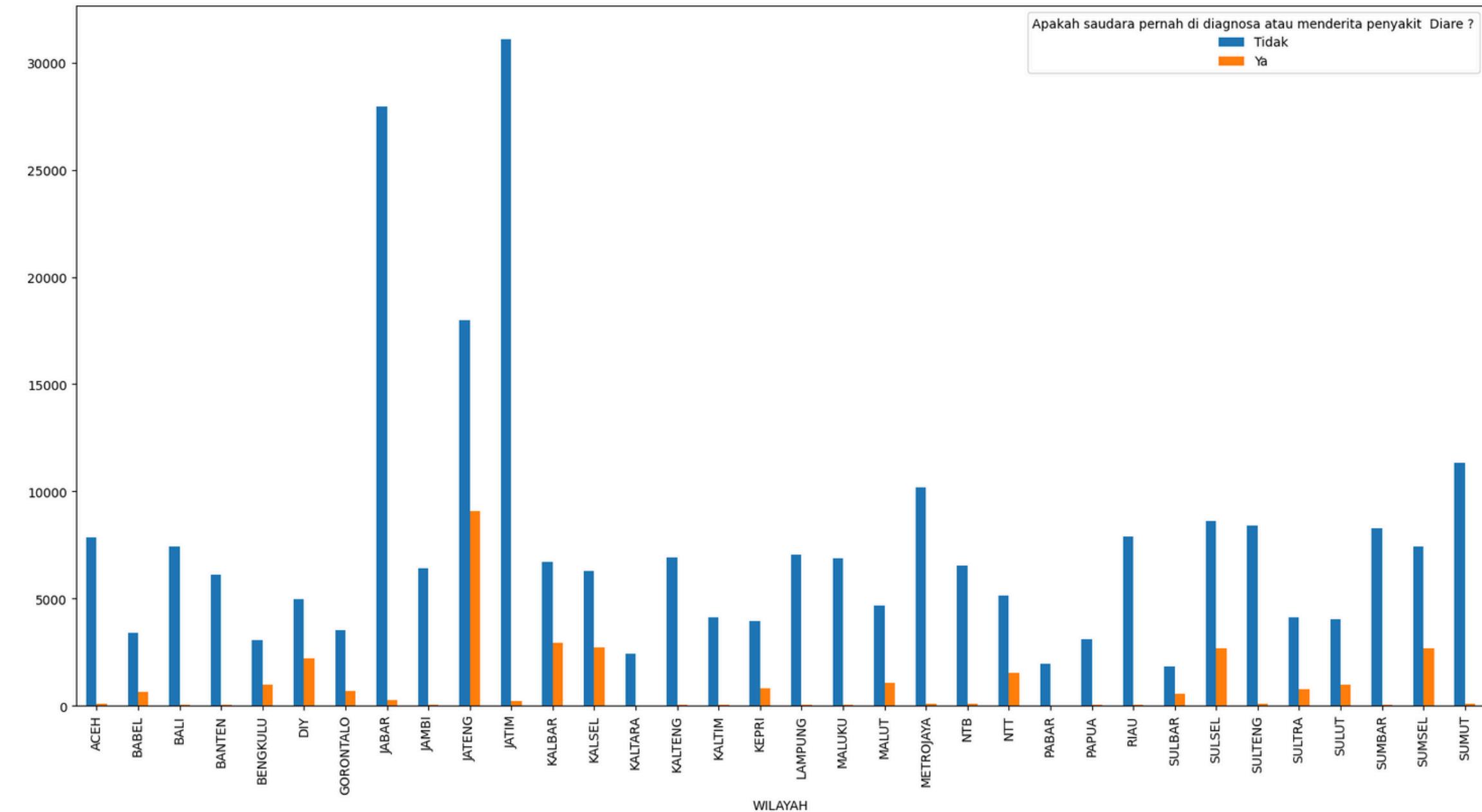
Dibentuk model regresi dengan Diare sebagai variabel respon, lalu variabel Air\_bersih, Cuci\_tangan, jarak\_tempat\_sampah, junkfood, dan USIA sebagai variabel prediktor.

## INSIGHT 4

Dari hasil regresi, terlihat bahwa semua parameter signifikan dengan tingkat signifikansi 5% sehingga beberapa informasi ini signifikan, yaitu:

- Sumber air bersih yang berasal dari tadah hujan cenderung lebih dapat menyebabkan terjadinya diare daripada sumber air bersih lainnya.
- Individu yang jarang atau bahkan sama sekali tidak mencuci tangan dengan sabun sebelum makan cenderung lebih mungkin terkena diare.
- Rumah yang di sekitarnya tidak memiliki tempat pembuangan sampah memiliki kecenderungan yang lebih rendah untuk terkena diare.
- Individu yang mengonsumsi junkfood lebih dari 2 kali per hari memiliki kecenderungan yang lebih tinggi untuk terkena diare.
- Seiring bertambahnya usia, maka dapat meningkatkan kemungkinan terkena diare.

# INSIGHT 4



Berdasarkan grafik di atas, dapat diketahui bahwa kasus terjadinya individu terdiagnosis diare paling banyak terdapat di Provinsi Jawa Tengah.

# INSIGHT 5

OLS Regression Results	
<b>Dep. Variable:</b>	KARIES
<b>R-squared:</b>	0.072
<b>Model:</b>	OLS
<b>Adj. R-squared:</b>	0.072
<b>Method:</b>	Least Squares
<b>F-statistic:</b>	1862.
<b>Date:</b>	Fri, 22 Dec 2023
<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	02:20:24
<b>Log-Likelihood:</b>	-1.4870e+05
<b>No. Observations:</b>	289801
<b>AIC:</b>	2.974e+05
<b>Df Residuals:</b>	289788
<b>BIC:</b>	2.976e+05
<b>Df Model:</b>	12
<b>Covariance Type:</b>	nonrobust

Dibentuk model regresi dengan KARIES sebagai variabel respon, lalu variabel JK, merokok, konsumsi\_gula, jlh\_alkohol, jlh\_gosok\_gigi, lama\_gosok\_gigi, dan USIA sebagai variabel prediktor.

	coef	std err	t	P> t  [0.025 0.975]
<b>Intercept</b>	0.0184	0.005	3.618	0.000 0.008 0.028
<b>JK[T.PEREMPUAN]</b>	0.0307	0.003	10.449	0.000 0.025 0.036
<b>C(merokok)[T.Ya]</b>	0.1014	0.002	64.040	0.000 0.098 0.105
<b>C(konsumsi_gula)[T.Lebih dari 6 sendok teh]</b>	0.1050	0.004	26.670	0.000 0.097 0.113
<b>C(konsumsi_gula)[T.Tidak pernah]</b>	-0.0435	0.002	-17.782	0.000 -0.048 -0.039
<b>C(jlh_alkohol)[T.Tidak pernah]</b>	-0.0674	0.004	-18.068	0.000 -0.075 -0.060
<b>C(jlh_alkohol)[T.lebih dari 25ml]</b>	-0.0358	0.014	-2.587	0.010 -0.063 -0.009
<b>C(jlh_gosok_gigi)[T.2 Kali]</b>	-0.1251	0.005	-25.211	0.000 -0.135 -0.115
<b>C(jlh_gosok_gigi)[T.&gt; 2 Kali]</b>	-0.1690	0.005	-33.320	0.000 -0.179 -0.159
<b>C(jlh_gosok_gigi)[T.Tidak pernah]</b>	-0.1742	0.008	-20.996	0.000 -0.190 -0.158
<b>C(lama_gosok_gigi)[T.1 - 2 menit]</b>	0.0622	0.003	20.400	0.000 0.056 0.068
<b>C(lama_gosok_gigi)[T.Di atas 3 menit]</b>	0.0367	0.003	11.227	0.000 0.030 0.043
<b>C(lama_gosok_gigi)[T.Di bawah 1 menit]</b>	0.0938	0.003	27.078	0.000 0.087 0.101
<b>USIA</b>	0.0081	7.7e-05	105.688	0.000 0.008 0.008
<b>Omnibus:</b>	43278.644		Durbin-Watson:	1.983
<b>Prob(Omnibus):</b>	0.000		Jarque-Bera (JB):	63927.887
<b>Skew:</b>	1.139		Prob(JB):	0.00
<b>Kurtosis:</b>	2.680		Cond. No.	1.92e+16

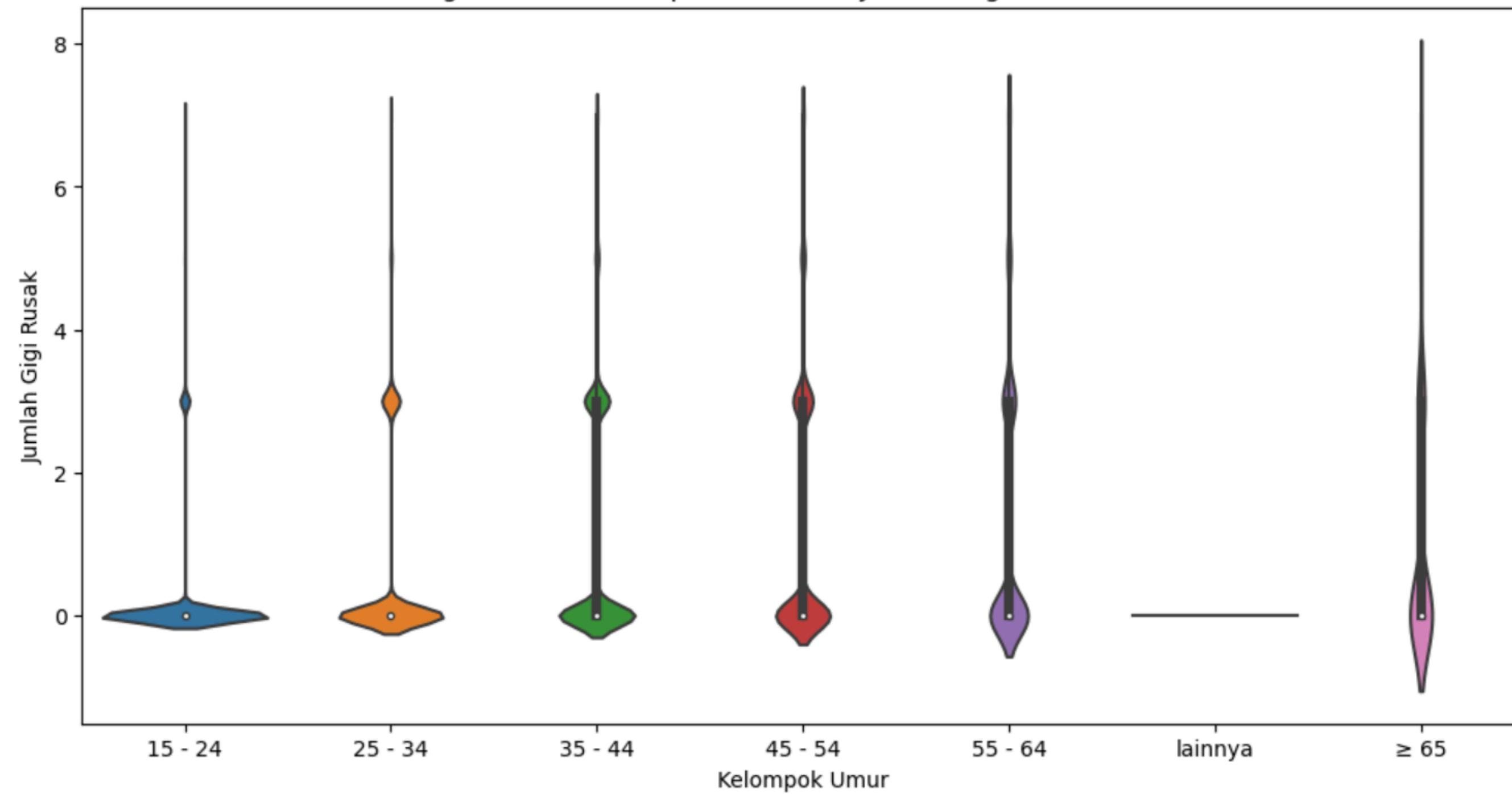
## INSIGHT 5

Dari hasil regresi, terlihat bahwa semua parameter signifikan dengan tingkat signifikansi 10% sehingga beberapa informasi ini signifikan, yaitu:

- Terdapat kecenderungan bahwa perempuan memiliki tingkat penyakit karies lebih tinggi daripada laki-laki.
- Individu yang perokok memiliki tingkat penyakit karies yang cenderung lebih tinggi dibandingkan tidak merokok.
- Konsumsi gula berpengaruh terhadap risiko penyakit karies. Pada data diperoleh bahwa tingkat penyakit karies lebih tinggi pada individu yang mengonsumsi gula  $> 6$  sendok teh.
- Konsumsi alkohol juga berpengaruh terhadap risiko penyakit karies. Pada data diperoleh bahwa tingkat penyakit karies lebih tinggi pada individu yang mengonsumsi alkohol  $> 25$  ml.
- Jumlah gosok gigi tidak terlalu berpengaruh secara signifikan terhadap terjadinya penyakit karies.
- Semakin lama menggosok gigi cenderung menurunkan nya risiko terkena karies. Pada data diperoleh bahwa individu yang menggosok gigi diatas 3 menit memiliki tingkat penyakit karies yang lebih rendah.
- Semakin bertambahnya usia cenderung meningkatkan tingkat terjadinya penyakit karies.

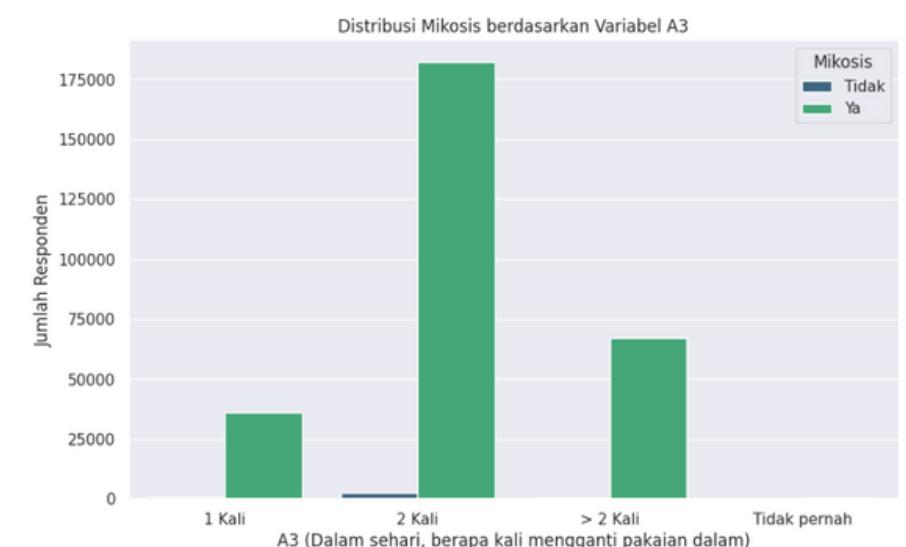
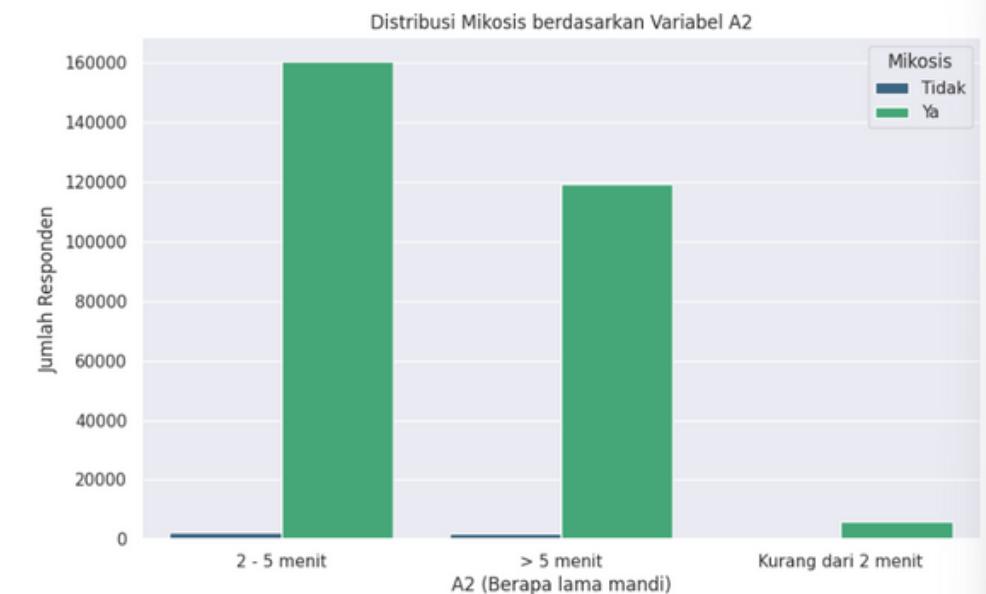
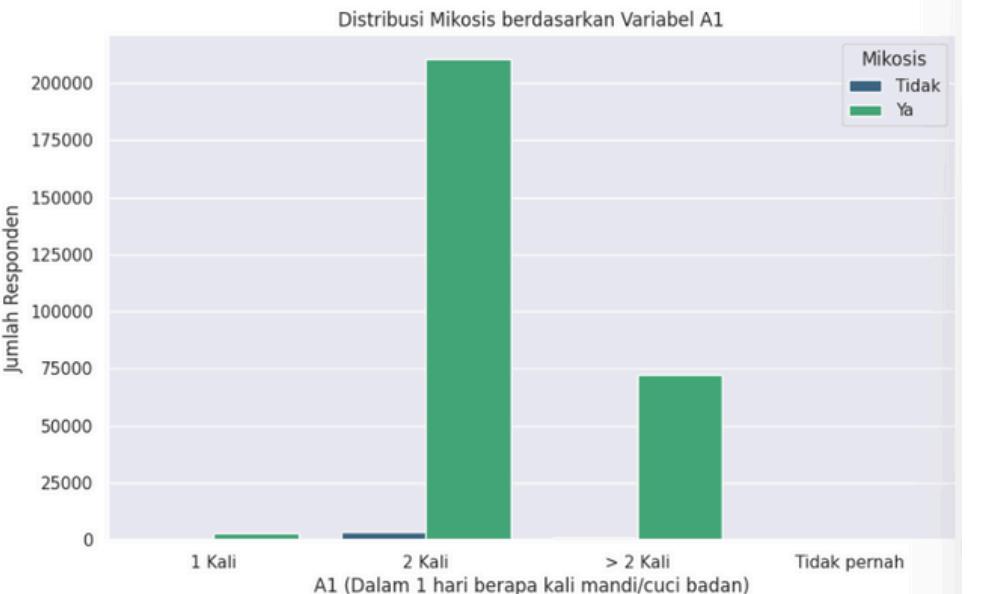
# INSIGHT 5

Hubungan antara Kelompok Umur dan Jumlah Gigi Rusak (Violin Plot)



# INSIGHT 6

OLS Regression Results						
<b>Dep. Variable:</b>	Mikosis	<b>R-squared:</b>	0.001			
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.001			
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	44.19			
<b>Date:</b>	Fri, 22 Dec 2023	<b>Prob (F-statistic):</b>	1.75e-71			
<b>Time:</b>	09:59:37	<b>Log-Likelihood:</b>	2.0352e+05			
<b>No. Observations:</b>	289801	<b>AIC:</b>	-4.070e+05			
<b>Df Residuals:</b>	289792	<b>BIC:</b>	-4.069e+05			
<b>Df Model:</b>	8					
<b>Covariance Type:</b>	nonrobust					
		coef	std err	t	P> t	[0.025 0.975]
Intercept		0.9581	0.002	422.788	0.000	0.954 0.963
C(A1)[T.2 Kali]		0.0210	0.002	9.241	0.000	0.017 0.025
C(A1)[T.> 2 Kali]		0.0227	0.002	9.770	0.000	0.018 0.027
C(A1)[T.Tidak pernah]		0.0133	0.008	1.768	0.077	-0.001 0.028
C(A2)[T.> 5 menit]		-0.0010	0.000	-2.175	0.030	-0.002 -9.92e-05
C(A2)[T.Kurang dari 2 menit]		-0.0009	0.002	-0.570	0.569	-0.004 0.002
C(A3)[T.2 Kali]		0.0065	0.001	9.469	0.000	0.005 0.008
C(A3)[T.> 2 Kali]		0.0100	0.001	12.091	0.000	0.008 0.012
C(A3)[T.Tidak pernah]		-0.0191	0.005	-3.604	0.000	-0.029 -0.009
Omnibus:	389243.157	Durbin-Watson:	1.982			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	51603316.803			
Skew:	-8.078	Prob(JB):	0.00			
Kurtosis:	66.344	Cond. No.	52.7			

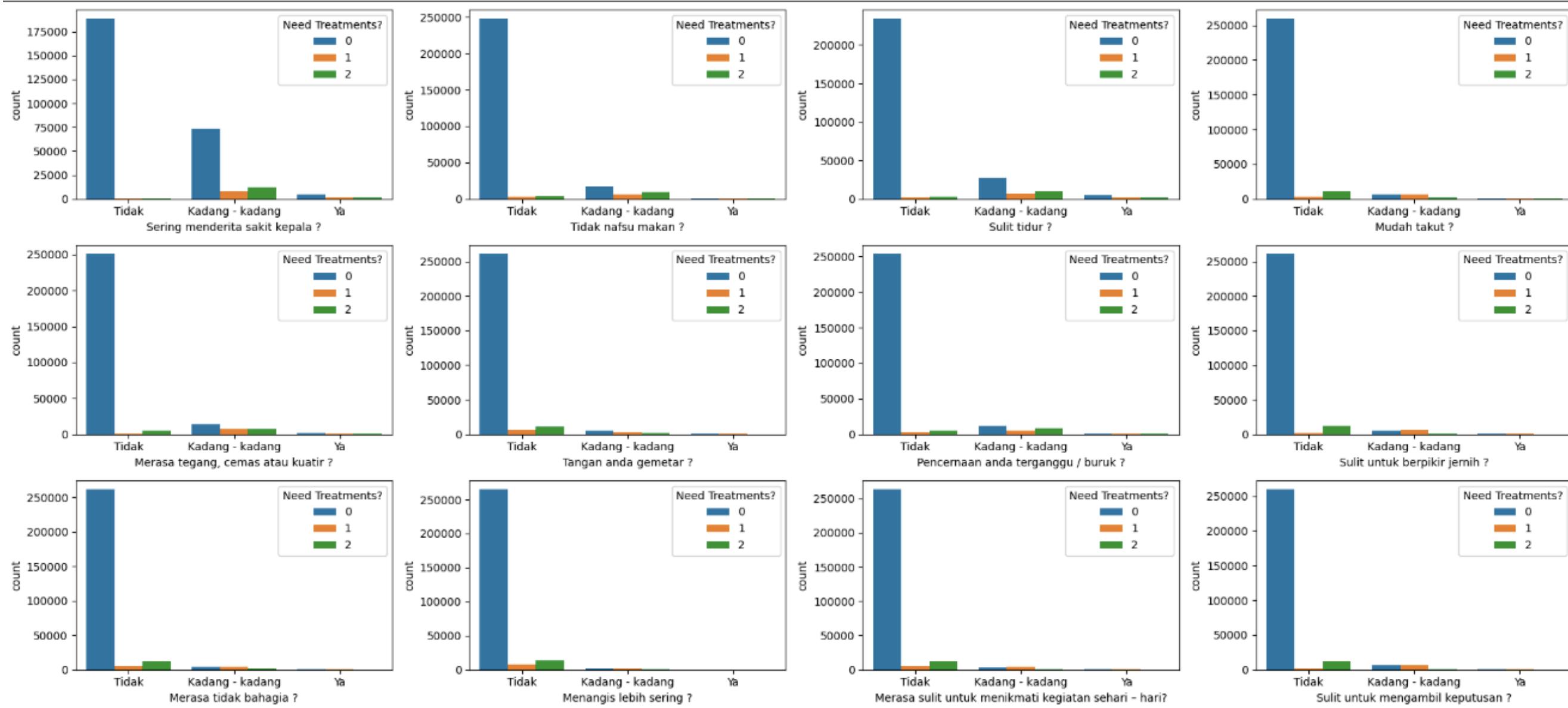


Dari hasil ols diatas, dengan taraf signifikansi 0,5 terlihat bahwa seluruh parameter signifikan. Didapatkan informasi sebagai berikut :

1. Orang yang mandi kurang dari 2 - 5 menit lebih banyak yang terkena mikosis dibandingkan dengan yang lain
2. Orang yang mengganti celana dalam 2 kali dalam sehari lebih banyak yang terkena mikosis dibandingkan dengan yang lain.
3. Orang yang mandi 2 kali dalam sehari lebih banyak terkena penyakit mikosis daripada yang lainnya.

Dari informasi diatas, anehnya orang-orang yang terkena mikosis adalah orang-orang yang rajin menjaga kebersihan. Hal ini menandakan ada penyebab lain yang menyebabkan seseorang terkena penyakit tersebut

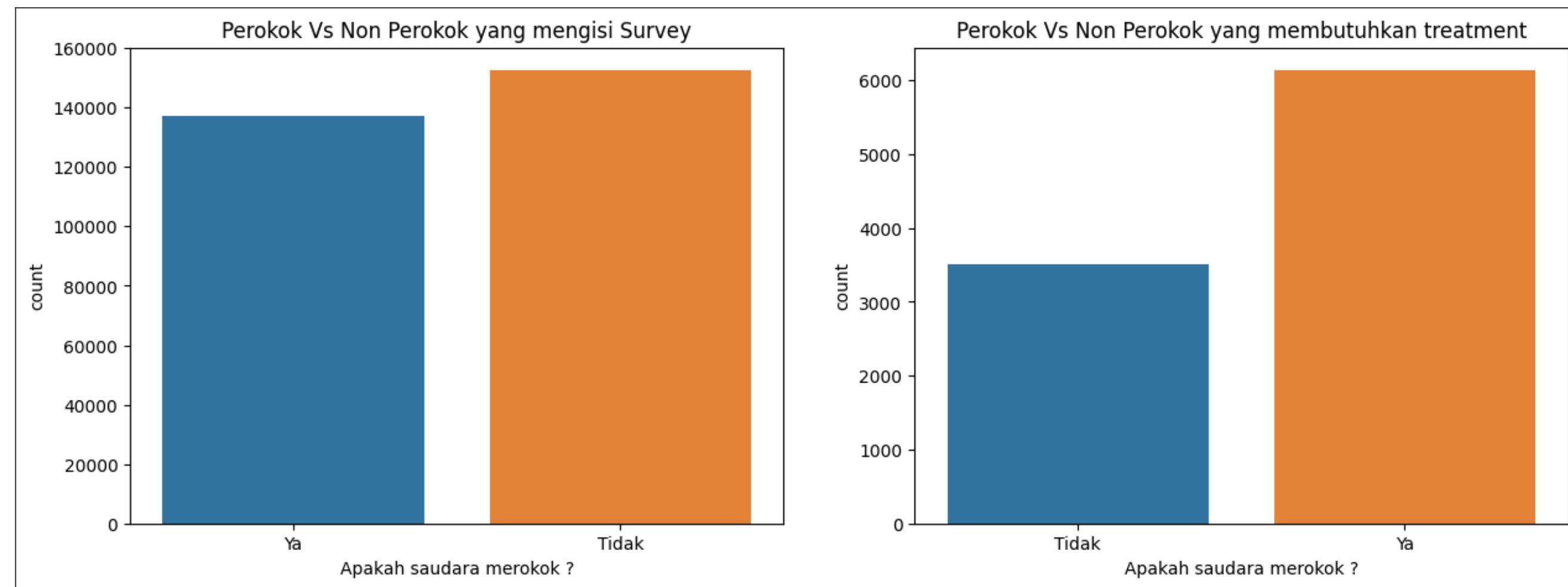
## Treatment Mental Health Menggunakan Clustering K-Modes



Berdasarkan hasil visualisasi clustering di samping dapat membantu untuk menentukan jenis cluster. Diambil kesimpulan :

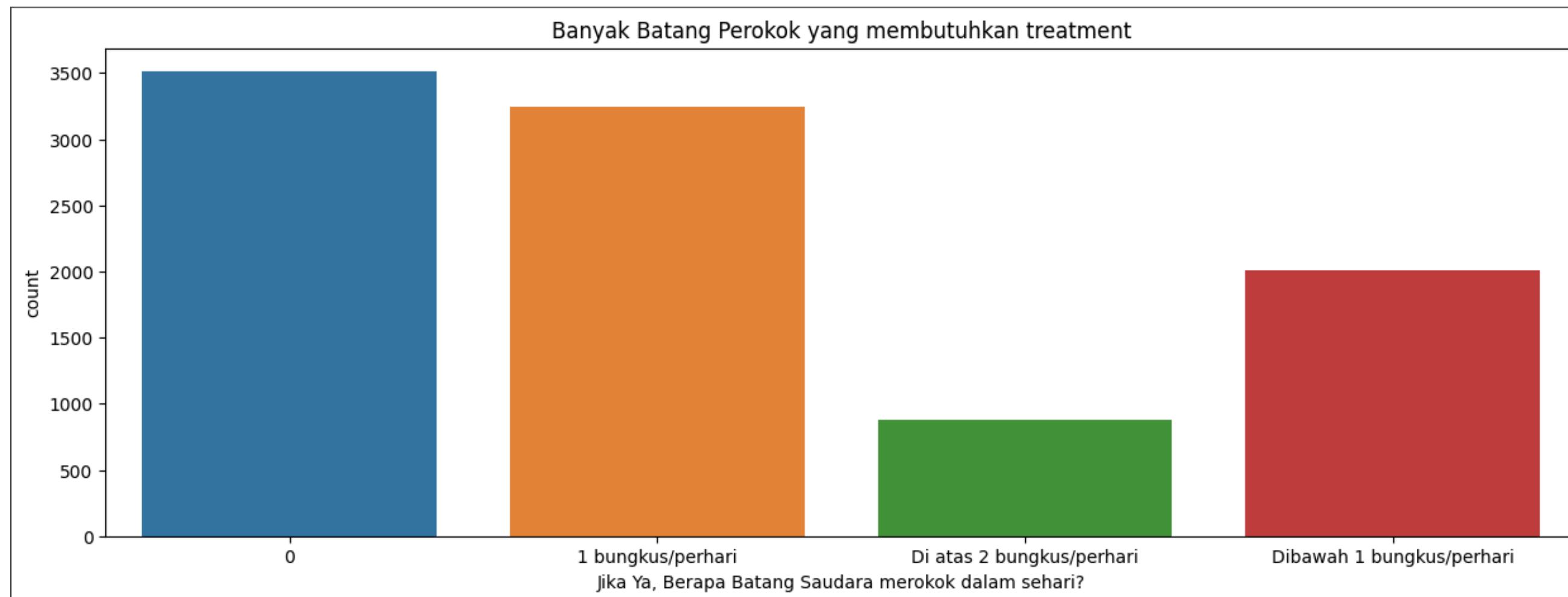
- Cluster 0 : Tidak membutuhkan treatment mental health**
- Cluster 1 : Sangat membutuhkan atau harus melakukan mental helath treatment**
- Cluster 2 : Mungkin membutuhkan atau dianjurkan untuk melakukan treatment mental health**

# Perokok Membutuhkan Treatment Mental Health



Berdasarkan hasil visualisasi di atas **orang-orang yang membutuhkan treatment mental health cenderung merokok**. Hal ini dapat disebabkan karena **nikotin pada rokok dapat memberikan rasa tenang atau relaksasi sementara**. Terapi dan dukungan yang tepat, termasuk berhenti merokok, dapat menjadi bagian penting dari perawatan untuk meningkatkan kesehatan mental.

## Banyak Batang Perokok yang Membutuhkan Treatment



Dari hasil visualisasi di atas, **sebagian besar perokok yang membutuhkan treatment merupakan perokok sedang**. Hal ini terjadi diindikasi bahwa **rokok hanya digunakan untuk memberikan rasa tenang atau relaksasi** sehingga mereka tidak mengonsumsi jumlah rokok yang tinggi setiap hari.

## Score Mental Healt yang Membutuhkan Treatment

```
count      9649.000000
mean       12.428749
std        4.895776
min        6.000000
25%        9.000000
50%        11.000000
75%        15.000000
max        39.000000
Name: Score, dtype: float64
```

Jika dilihat dari hasil di samping, **orang yang harus melakukan treatment** untuk kesehatan mental adalah **orang yang memiliki score di antara 6 sampai 39 dengan rata-rata score 12.43**

# Terima Kasih

