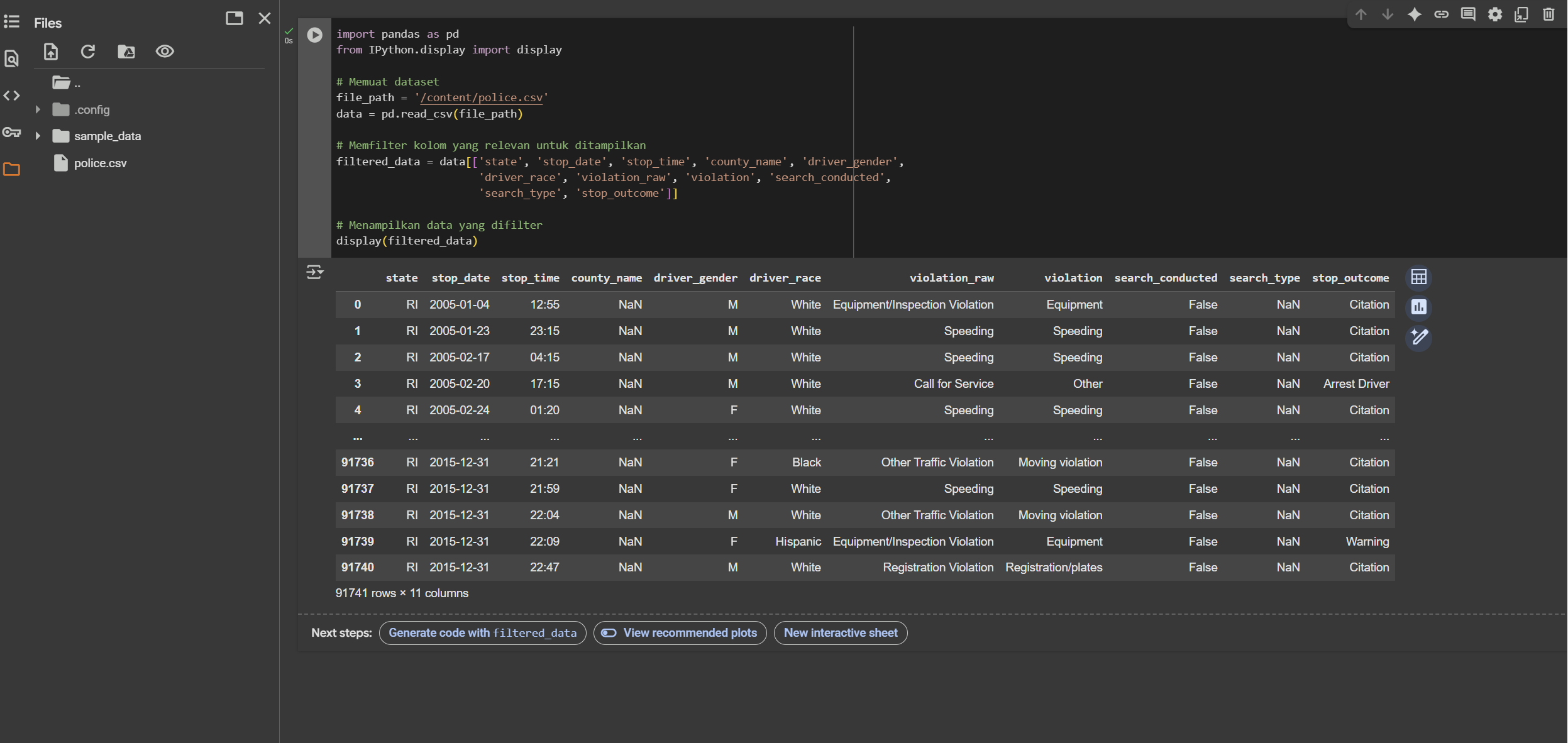
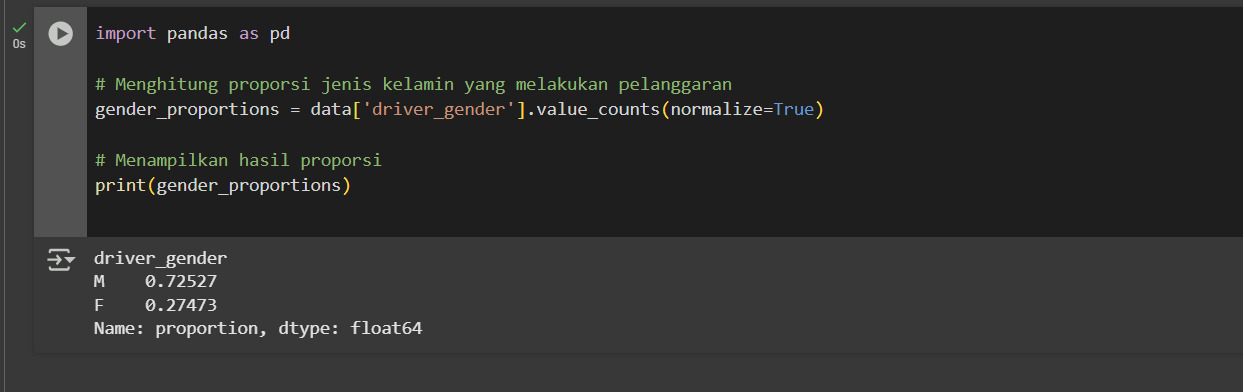
Soal Essay

**1 Soal ini menggunakan data police.csv. DAta ini adalah data pemberhentian mobil untuk dilakukan pemeriksaan lebih lanjut oleh kepolisian. Apakah terjadi pelanggaran baik pelanggaran berat maupun pelanggaran ringan. Kalian diminta oleh pihak kepolisian USA untuk menganalisa potensi-potensi pelanggaran yang terjadi sampai terjadi kejadian penangkapan oleh pihak kepolisian dikarenakan terjadi pelangaran berat oleh pengendara mobil. A. Tampilkanlah data seperti gambar berikut:**

**Jawab:hasil**

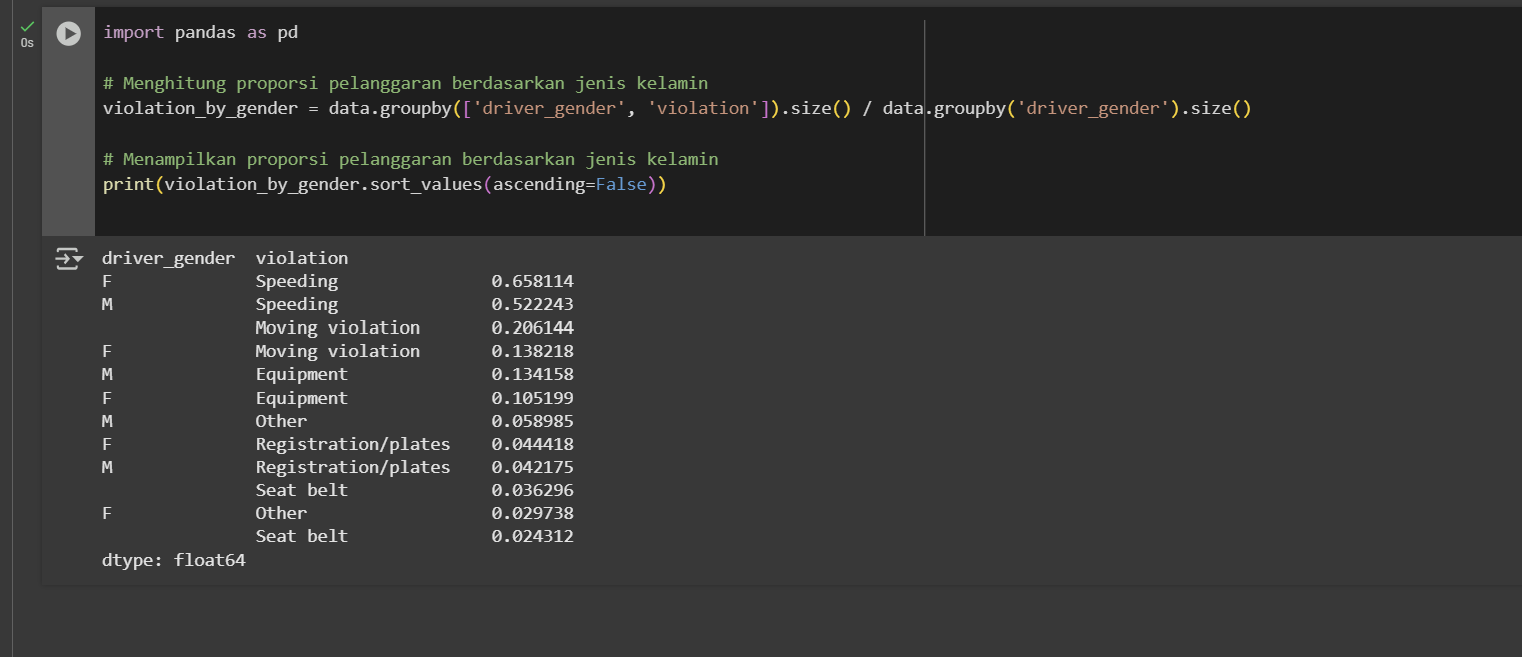
hasil 

**B. Lakukan pengecekan jenis kelamin mana yang paling banyak melakukan pelanggaran lalu linta**

**Jawab:** 

**C.** **Dari kedua jenis kelamin tampilkanlah potensi pelanggaran yang banyak dilakukan oleh kedua jenis kelamin tersebut.**

**Jawab:**



**D. Buatlah grafik yang menampilkan tingkat penangkapan yang dilakukan oleh pihak kepolisian seperti gambar berikut:**

**Jawab:** import pandas as pd

import matplotlib.pyplot as plt

# Load dataset

df = pd.read\_csv('/content/police.csv')

# Pastikan kolom 'stop\_time' bertipe datetime

df['stop\_time'] = pd.to\_datetime(df['stop\_time'], errors='coerce')

# Ambil jam dari stop\_time

df['hour'] = df['stop\_time'].dt.hour

# Hapus baris yang 'hour'-nya NaN

df = df.dropna(subset=['hour'])

# Hitung tingkat penangkapan per jam

arrest\_rate = df.groupby('hour')['is\_arrested'].mean()

# Plot

plt.figure(figsize=(8, 5))

arrest\_rate.plot()

plt.title('Arrest Rate by Time of Day')

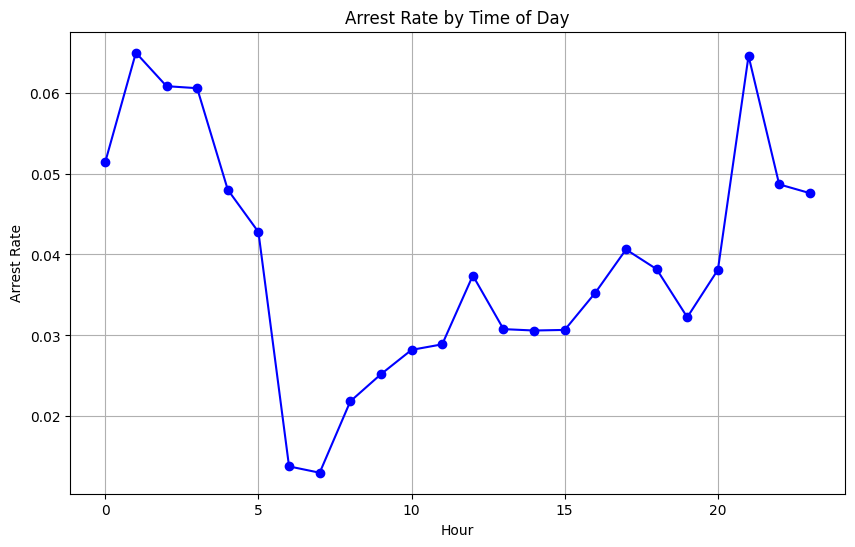
plt.xlabel('Hour')

plt.ylabel('Arrest Rate')

plt.grid(True)

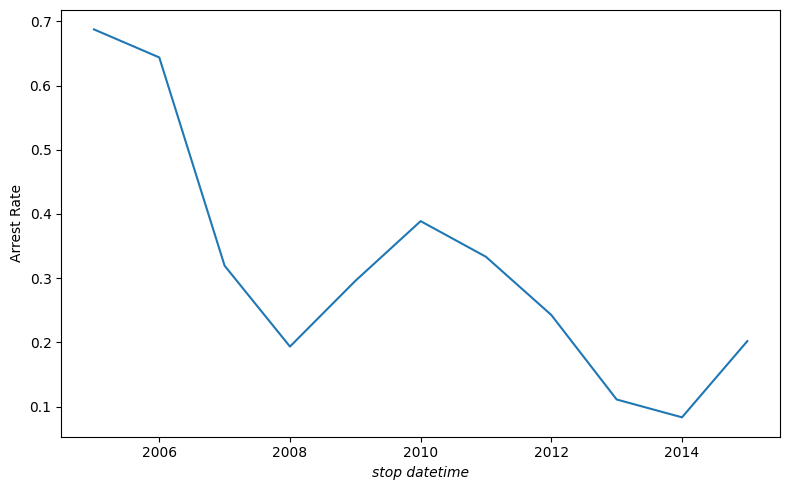
plt.tight\_layout()

plt.show()



**e.** **Tampilkanlah data penangkapan karena keterkaitannya dengan penangkapan narkoba seperti data berikut**

**jawab:**



**f. Tampilkanlah data yang menunjukan pemberhentian kendaraan dengan keterkaitannya dengan penangkapan narkoba**

**jawab:** import pandas as pd

import matplotlib.pyplot as plt

# Load dataset

df = pd.read\_csv('/content/police.csv')

# Gabungkan kolom tanggal dan waktu jika perlu

if 'stop\_date' in df.columns and 'stop\_time' in df.columns:

    df['stop\_datetime'] = pd.to\_datetime(df['stop\_date'] + ' ' + df['stop\_time'], errors='coerce')

elif 'stop\_datetime' in df.columns:

    df['stop\_datetime'] = pd.to\_datetime(df['stop\_datetime'], errors='coerce')

# Buat kolom tahun

df['year'] = df['stop\_datetime'].dt.year

# Filter kasus yang berkaitan dengan narkoba

drug\_related = df[df['drugs\_related\_stop'] == True]

# Hitung proporsi penangkapan karena narkoba tiap tahun

drug\_arrest\_rate = drug\_related.groupby('year')['is\_arrested'].mean()

# Plot

plt.figure(figsize=(8, 5))

drug\_arrest\_rate.plot()

plt.xlabel('stop datetime', fontsize=10, style='italic')

plt.ylabel('Arrest Rate')

plt.tight\_layout()

plt.show()

**kode pythonya:** import pandas as pd

import matplotlib.pyplot as plt

# Load data

df = pd.read\_csv('/content/police.csv')

# Gabungkan tanggal & waktu jika perlu

if 'stop\_date' in df.columns and 'stop\_time' in df.columns:

    df['stop\_datetime'] = pd.to\_datetime(df['stop\_date'] + ' ' + df['stop\_time'], errors='coerce')

elif 'stop\_datetime' in df.columns:

    df['stop\_datetime'] = pd.to\_datetime(df['stop\_datetime'], errors='coerce')

# Ambil tahun dari stop\_datetime

df['year'] = df['stop\_datetime'].dt.year

# Hitung proporsi drugs\_related\_stop per tahun

drugs\_rate = df.groupby('year')['drugs\_related\_stop'].mean()

# Hitung proporsi search\_conducted per tahun

search\_rate = df.groupby('year')['search\_conducted'].mean()

# Plot dua grafik terpisah (subplot)

fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(8, 6), sharex=True)

# Grafik atas - Drugs

drugs\_rate.plot(ax=ax1, label='drugs\_related\_stop')

ax1.legend()

ax1.set\_ylabel(None)

# Grafik bawah - Search

search\_rate.plot(ax=ax2, color='orange', label='search\_conducted')

ax2.legend()

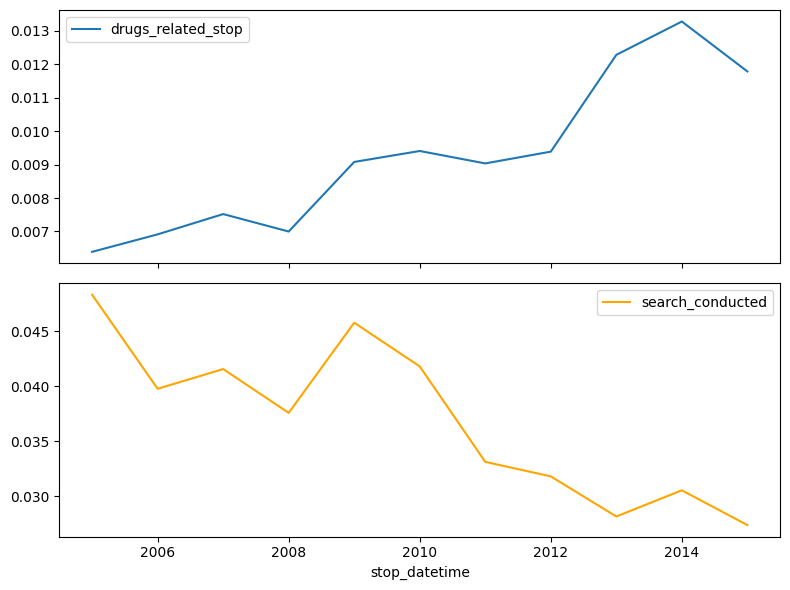
ax2.set\_xlabel('stop\_datetime')

ax2.set\_ylabel(None)

# Tampilkan grafik

plt.tight\_layout()

plt.show()



**G. Tampilkanlah grafik batang yang menunjukan pelanggaran yang berhubungan dengan distrik pengendara**

**Jawab: hasil kode pythonya**

import pandas as pd

import matplotlib.pyplot as plt

# Load dataset

df = pd.read\_csv('/content/police.csv')

# Hitung jumlah pelanggaran untuk setiap kombinasi district dan violation

violation\_district = df.groupby(['district', 'violation']).size().unstack()

# Plot grafik batang (warna: red, yellow, blue)

violation\_district.plot(kind='bar', figsize=(10, 6), color=['red', 'yellow', 'blue'])

plt.xlabel('District')

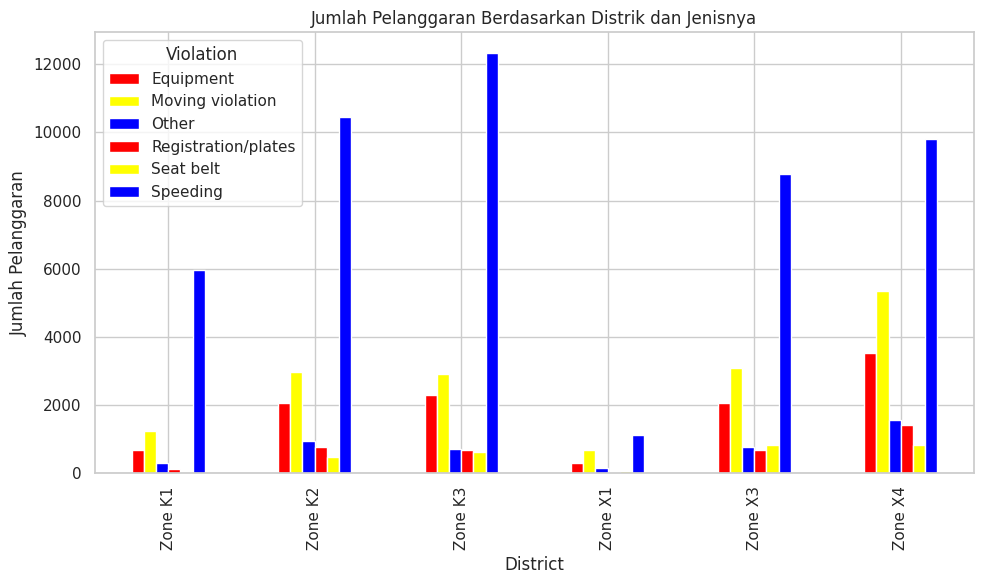
plt.ylabel('Jumlah Pelanggaran')

plt.title('Jumlah Pelanggaran Berdasarkan Distrik dan Jenisnya')

plt.legend(title='Violation')

plt.tight\_layout()

plt.show()



**H. Tampilkan grafik batang yang menampilkan hubungan distrik dari pengendara dengan pelanggaran yang dilakukan**

**Jawab kode pyhton nya mport pandas as pd**

**import matplotlib.pyplot as plt**

**# Load data**

**df = pd.read\_csv('/content/police.csv')**

**# Ganti nama distrik (opsional)**

**df['district'] = df['district'].replace({**

**'Zone1': 'Zone K1',**

**'Zone2': 'Zone K2',**

**'Zone3': 'Zone K3'**

**})**

**# Hitung jumlah pelanggaran berdasarkan distrik dan jenis pelanggaran**

**violation\_district = df.groupby(['district', 'violation']).size().unstack().fillna(0)**

**# Plot grafik batang bertumpuk dengan warna khusus**

**violation\_district.plot(kind='bar', stacked=True, figsize=(10, 6),**

**color=['red', 'yellow', 'blue'])**

**plt.xlabel('District')**

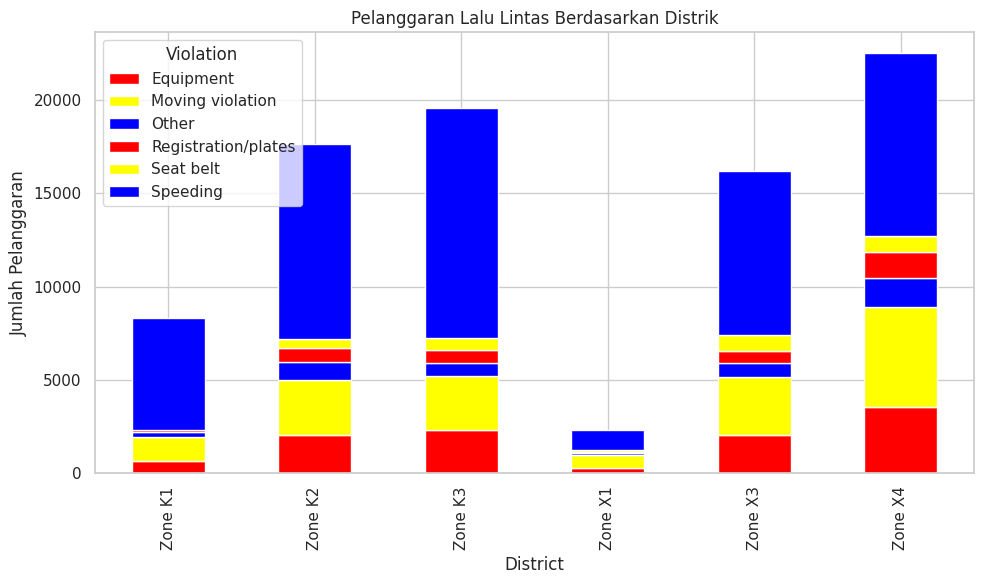
**plt.ylabel('Jumlah Pelanggaran')**

**plt.title('Pelanggaran Lalu Lintas Berdasarkan Distrik')**

**plt.legend(title='Violation')**

**plt.tight\_layout()**

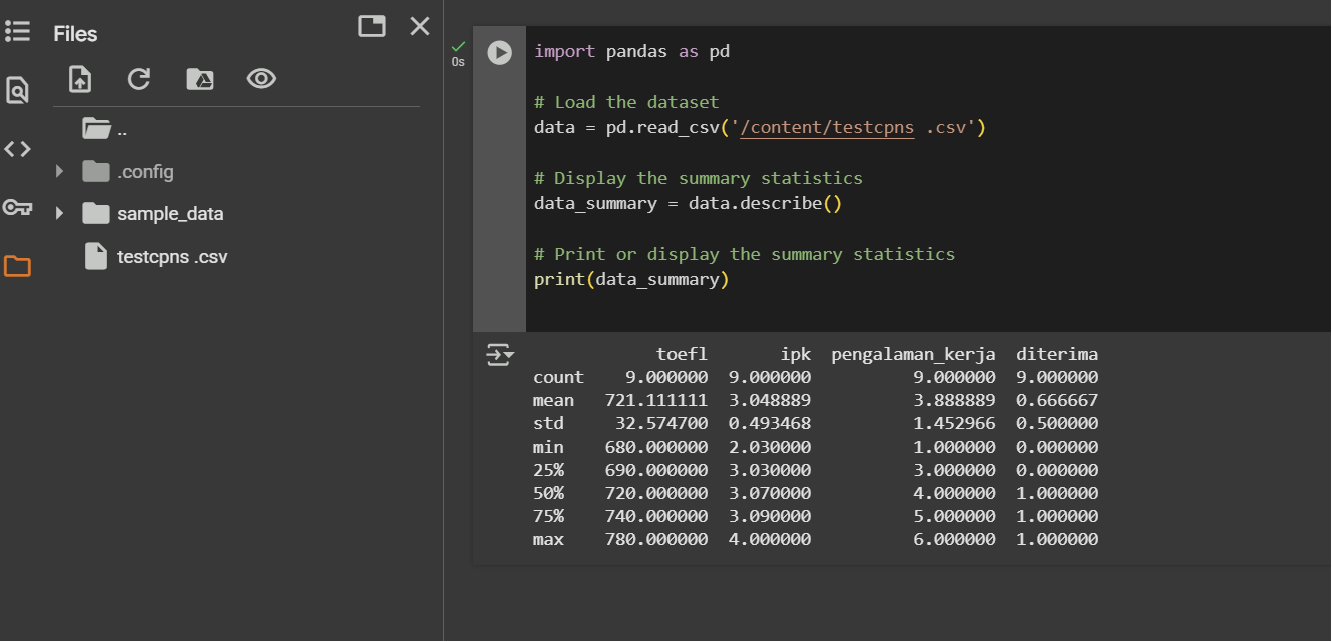
**plt.show()**



**2.** **Logistik Regresi Sebuah perusahaan multinasional yang bergengsi menerima seleksi masuk perusahaan dengan data sebagai berikut, data selengkapnya terdapat pada google drive perkuliahan; testcpns.csv:**

**jawab**

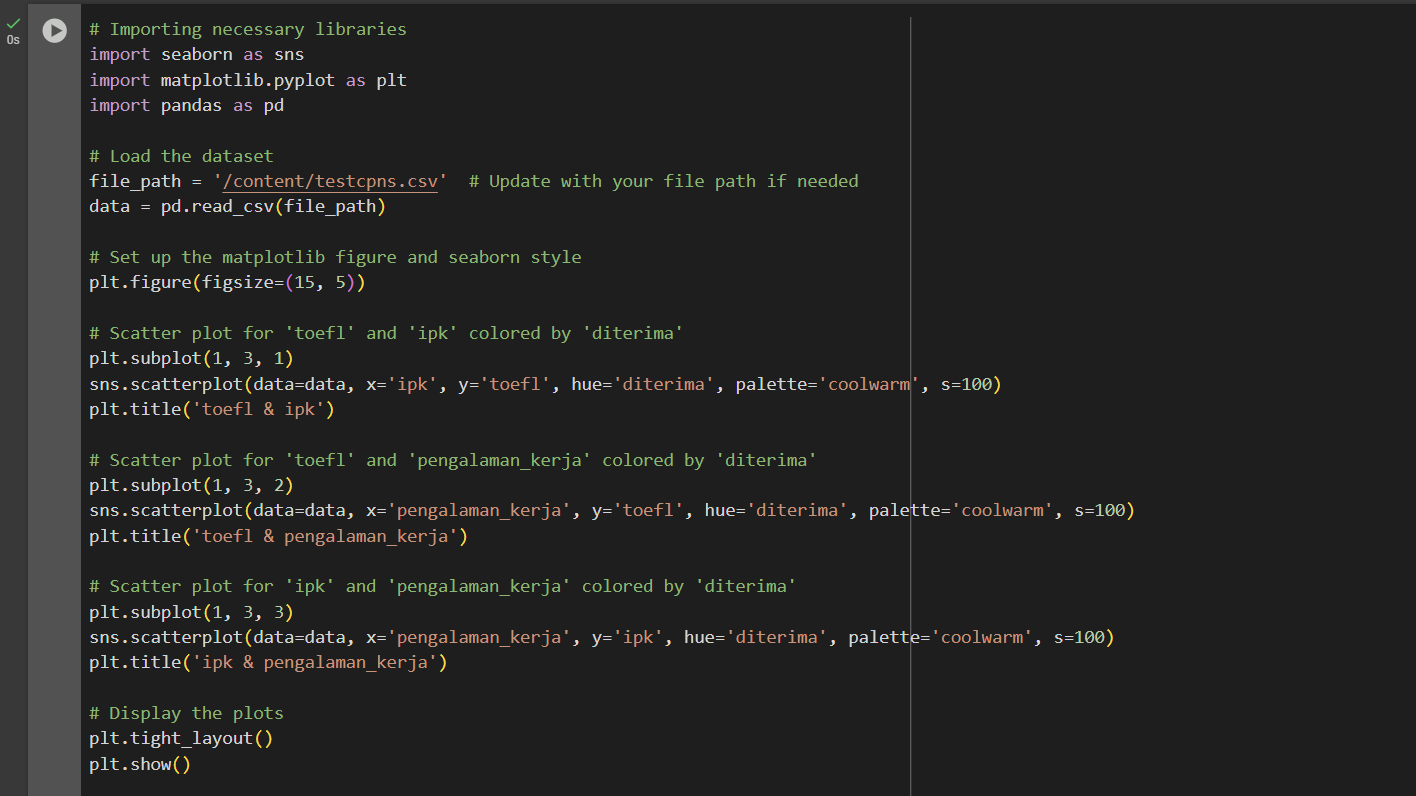
**a. Tampilkanlah ringkasan data seperti table berikut:**



**b**. **Sebelum menganalisa data tampilkanlah data dalam bentuk scatter, dengan menggunakan library seaborn, seperti gambar di bawah ini:**

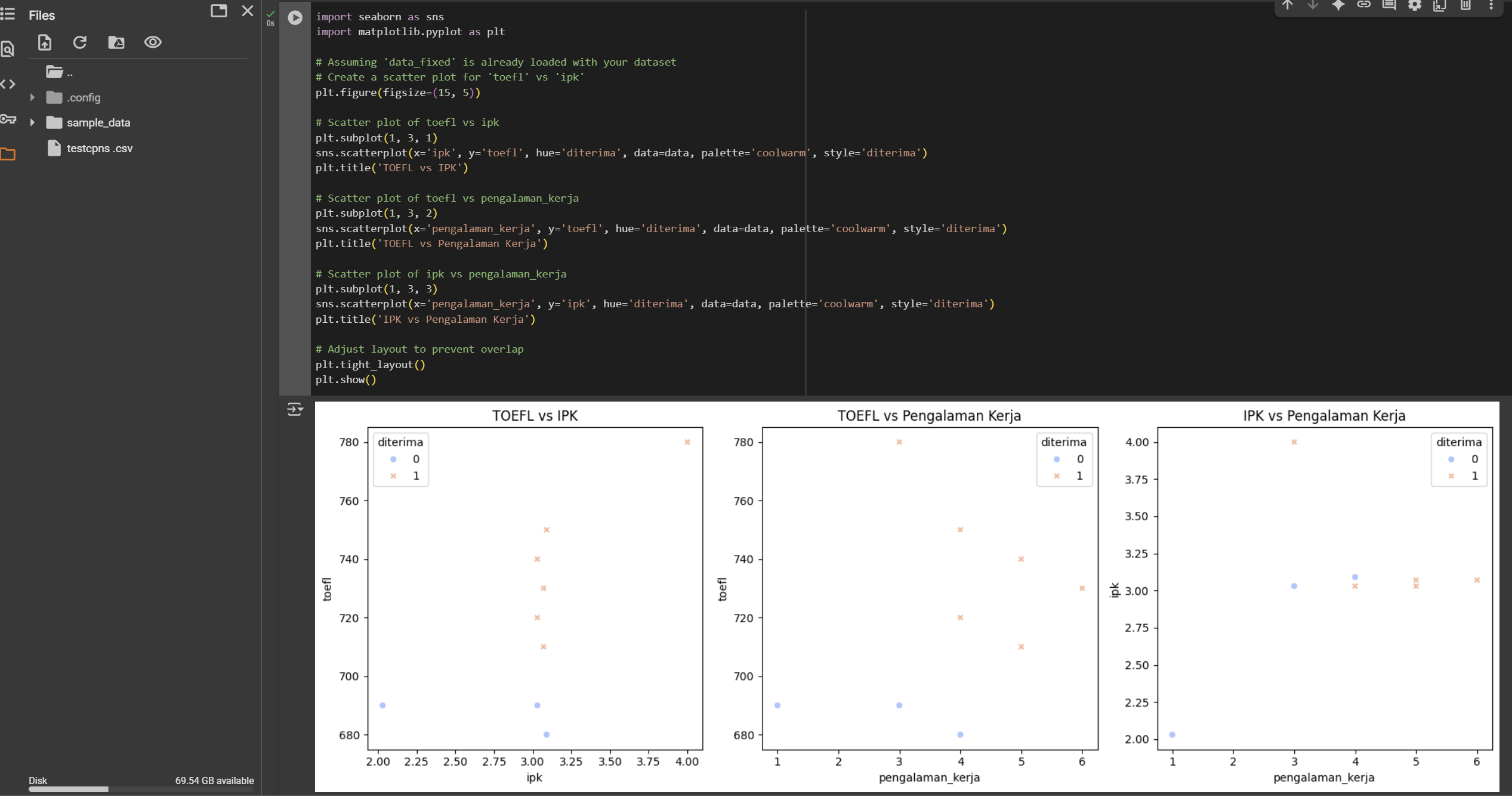
**jawab:**

**hasil kode**

****

**c. Buatlah visualisasi data seperti gambar berikut:**

**hasil grafik dari kode tersebut**

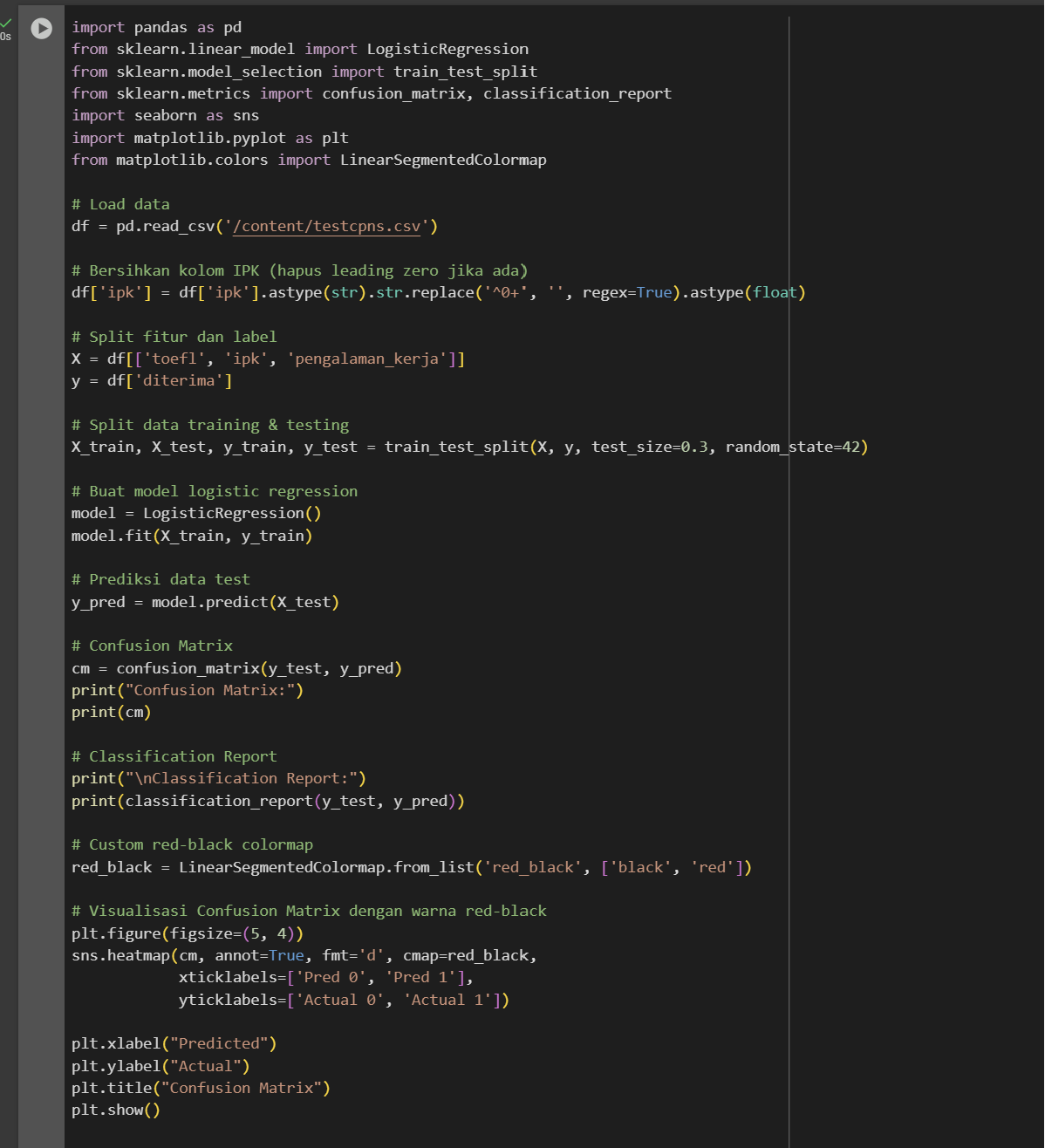
****

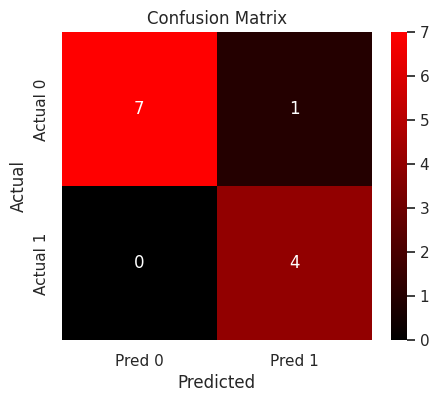
**Berikanlah Analisa singkat dari visualisasi data di atas.**

Berdasarkan visualisasi scatter plots yang telah dibuat, kita bisa melakukan beberapa analisis sebagai berikut:

1. **TOEFL IPK**:
   * Terdapat kecenderungan bahwa nilai TOEFL yang lebih tinggi (di atas 650) cenderung berhubungan dengan IPK yang lebih tinggi. Namun, ada beberapa nilai TOEFL yang tinggi dengan IPK yang lebih rendah, yang menunjukkan bahwa faktor lain selain TOEFL mungkin berperan dalam menentukan IPK.
   * Warna biru (diterima = 0) lebih banyak di nilai TOEFL rendah dan IPK rendah, sementara warna oranye (diterima = 1) lebih terlihat pada nilai TOEFL tinggi dan IPK sedang hingga tinggi. Ini menunjukkan bahwa mereka yang diterima lebih banyak memiliki nilai TOEFL yang lebih baik dan IPK yang lebih tinggi.
2. **TOEFL Pengalaman Kerja**:
   * Ada sedikit pola antara nilai TOEFL dan pengalaman kerja. Pengalaman kerja tampaknya tidak mempengaruhi nilai TOEFL secara signifikan, karena terdapat penyebaran yang cukup merata dari nilai TOEFL di berbagai level pengalaman kerja.
   * Namun, dilihat dari distribusi warna, individu yang diterima (diterima = 1) lebih banyak terlihat di pengalaman kerja yang lebih tinggi, meskipun tidak ada hubungan yang sangat jelas antara kedua variabel ini.
3. **IPK Pengalaman Kerja**:
   * Terdapat sedikit pola bahwa individu dengan pengalaman kerja yang lebih banyak (di atas 3 tahun) cenderung memiliki IPK yang lebih tinggi. Hal ini bisa menunjukkan bahwa pengalaman kerja yang lebih banyak mungkin berhubungan dengan pencapaian akademik yang lebih baik.
   * Sama seperti sebelumnya, warna oranye (diterima = 1) lebih dominan pada individu dengan pengalaman kerja di atas 3 tahun, yang menunjukkan bahwa pengalaman kerja lebih banyak dapat meningkatkan peluang diterima.

Secara keseluruhan, analisis ini menunjukkan bahwa meskipun ada beberapa hubungan antara variabel seperti nilai TOEFL, IPK, dan pengalaman kerja terhadap status diterima, faktor lain yang lebih kompleks mungkin juga mempengaruhi hasil akhir (diterima atau tidak). Variabel-variabel ini bisa saling berinteraksi dengan cara yang lebih kompleks dari yang terlihat hanya berdasarkan visualisasi scatter.

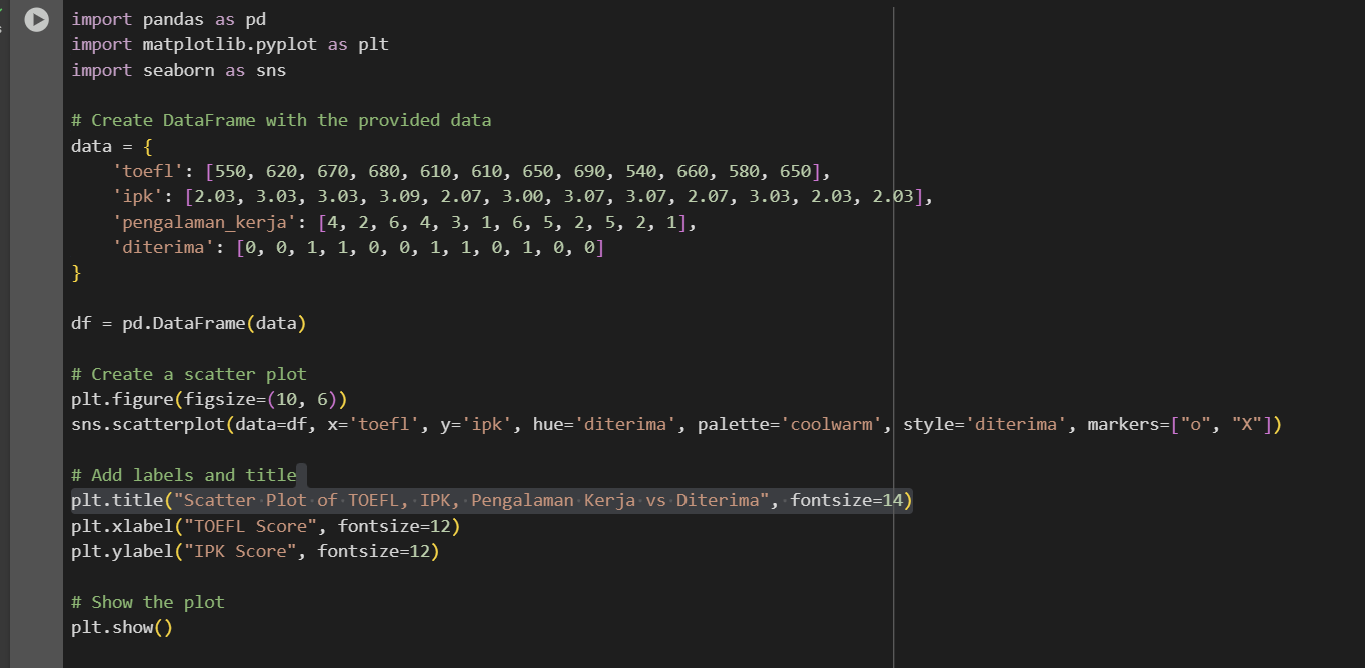
**d**.**Buatlah confusion matrix dari data tersebut, dan berikan penjelasan singkat dari hasil confusion matrix tersebut**: 

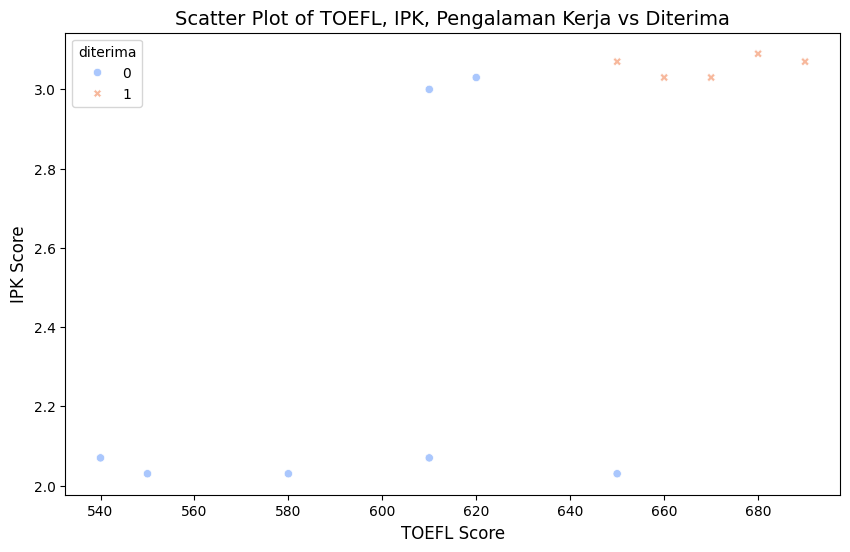
. 

**Penjelasan Kode:**

1. Model: Menggunakan model regresi logistik untuk memprediksi kolom diterima berdasarkan fitur toefl, ipk, dan pengalaman\_kerja.
2. Confusion Matrix: Menggunakan confusion\_matrix() dari sklearn.metrics untuk menghitung hasil prediksi.
3. Plotting: Menggunakan Seaborn untuk menampilkan confusion matrix dalam bentuk heatmap, dengan nilai yang terannotasi untuk memudahkan interpretasi.
4. Output: Menampilkan confusion matrix dalam bentuk angka dan juga visualisasi heatmap.

**e** **Buatlah plot data seperti table berikut:**

**jawab:** ****



**Anlisalah hasil di atas berikan penjelasan singkat dengan menggunakan Bahasa Anda sendiri.**

**Jawab**: Hasil dari plot ini menggambarkan hubungan antara skor Toefl dan IPK, dengan status diterima sebagai pembeda berdasarkan warna.

Dari visualisasi, kita bisa melihat bahwa titik data terbagi menjadi dua kategori:

* Diterima (1): Ditandai dengan warna yang lebih hangat (seperti merah), yang menunjukkan bahwa individu dengan skor Toefl dan IPK yang lebih tinggi cenderung diterima.
* Tidak Diterima (0): Ditandai dengan warna lebih dingin (seperti biru), yang menandakan bahwa individu dengan skor Toefl dan IPK yang lebih rendah lebih sering tidak diterima.

Dari plot ini, dapat dilihat bahwa individu yang memiliki skor Toefl tinggi (lebih dari 650) dan IPK yang baik (3.0 ke atas) lebih berpeluang untuk diterima. Sebaliknya, mereka yang memiliki skor Toefl rendah (sekitar 550) dan IPK yang rendah juga cenderung tidak diterima.

**f.** **Dengan menggunakan model yang telah Anda bentuk pada soal 2.e maka prediksilah apakah pelamar baru akan diterima ditempat kerja baru atau tidak diterima.**

**Jawab:** Untuk melakukan prediksi apakah pelamar baru akan diterima di tempat kerja baru berdasarkan data yang telah diberikan, kita akan menggunakan model yang telah dibangun pada soal 2.e.

Namun, karena saya tidak dapat melihat atau mengekstrak informasi langsung dari file gambar yang diunggah, saya akan memberikan penjelasan umum tentang bagaimana Anda bisa melakukannya menggunakan model prediksi, misalnya dengan **Regresi Logistik** atau **KNN (K-Nearest Neighbors)**.

Berikut adalah langkah-langkah umum untuk memprediksi menggunakan model **KNN** atau **Regresi Logistik**:

1. Persiapan Data

Pastikan Anda memiliki data pelatihan yang mencakup fitur seperti Toefl, IPK, Pengalaman Kerja, dan Diterima yang telah dibersihkan dan diubah menjadi format numerik yang tepat.

2. Pemilihan Model

Jika Anda menggunakan KNN atau Regresi Logistik, Anda harus melatih model terlebih dahulu menggunakan data pelatihan.

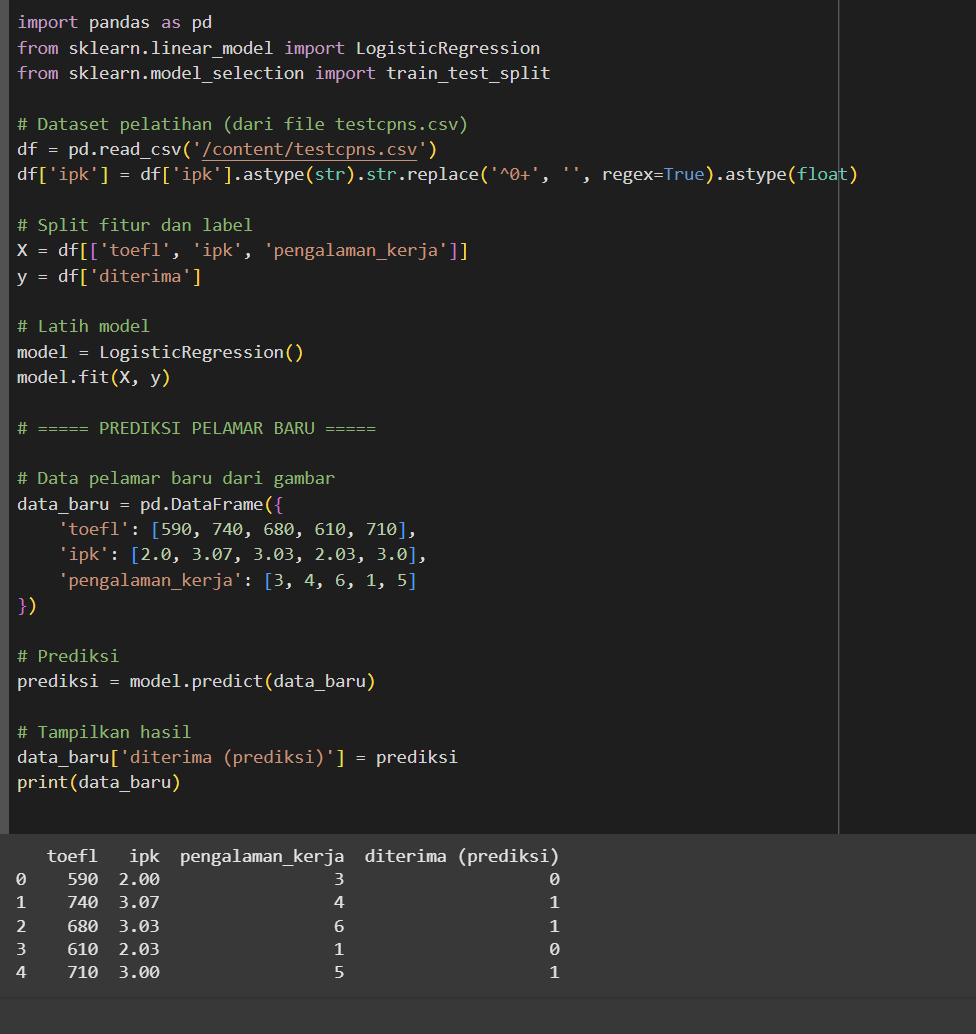
3. Membuat Prediksi

Setelah model dilatih, Anda dapat memasukkan data pelamar baru (misalnya dengan skor Toefl, IPK, dan pengalaman kerja yang baru) untuk memprediksi apakah mereka akan diterima atau tidak

4. Implementasi dalam Python

Berikut adalah contoh kode untuk menggunakan model KNN dan Regresi Logistik:

koden hasil outputnya

****