

# Data science

Fauzan Akmal Rabbani

Read more





```
▶ !pip install pandas #untuk menginstal library
```

!pip install pandas

- Tanda ! digunakan dalam Jupyter Notebook untuk menjalankan perintah sistem.
- pip install pandas adalah perintah untuk menginstal library pandas, yang digunakan untuk manipulasi dan analisis data dalam Python.

```
▶ import pandas as pd  
as_ = pd.read_csv('/content/drive/MyDrive/Diabetes/diabetes.csv') #membaca data dari file CSV
```

import pandas as pd

- Mengimpor library pandas dan memberi alias pd agar lebih mudah digunakan.
- as\_ = pd.read\_csv('/content/drive/MyDrive/Diabetes/diabetes.csv')
- pd.read\_csv(...) digunakan untuk membaca file CSV dan mengubahnya menjadi DataFrame pandas.
- '/content/drive/MyDrive/Diabetes/diabetes.csv' adalah path file CSV yang disimpan di Google Drive. Path ini menunjukkan bahwa kode kemungkinan dijalankan di Google Colab.
- Variabel as\_ digunakan untuk menyimpan DataFrame yang dihasilkan.

```
▶ import matplotlib.pyplot as plt

fig, ax = plt.subplots()
for Outcome, d in ds.groupby('Outcome'):
    ax.scatter(d['Glucose'], d['BloodPressure'], label=Outcome)

plt.legend()
plt.title('Diabetes')
plt.xlabel('Glucose')## untuk membuat scatter plot yang memvisualisasikan hubungan antara 'Glucose' dan 'BloodPressure' berdasarkan 'Outcome' dari dataset diabetes.
plt.ylabel('BloodPressure')
plt.grid(True)
plt.show()
```

## Import Matplotlib

```
import matplotlib.pyplot as plt
```

- pustaka matplotlib.pyplot diimpor dan diberi alias plt, yang akan digunakan untuk membuat grafik.

## Membuat Subplot:

```
fig, ax = plt.subplots()
```

- plt.subplots() digunakan untuk membuat figure (fig) dan axes (ax). Ini adalah dasar untuk membuat plot.

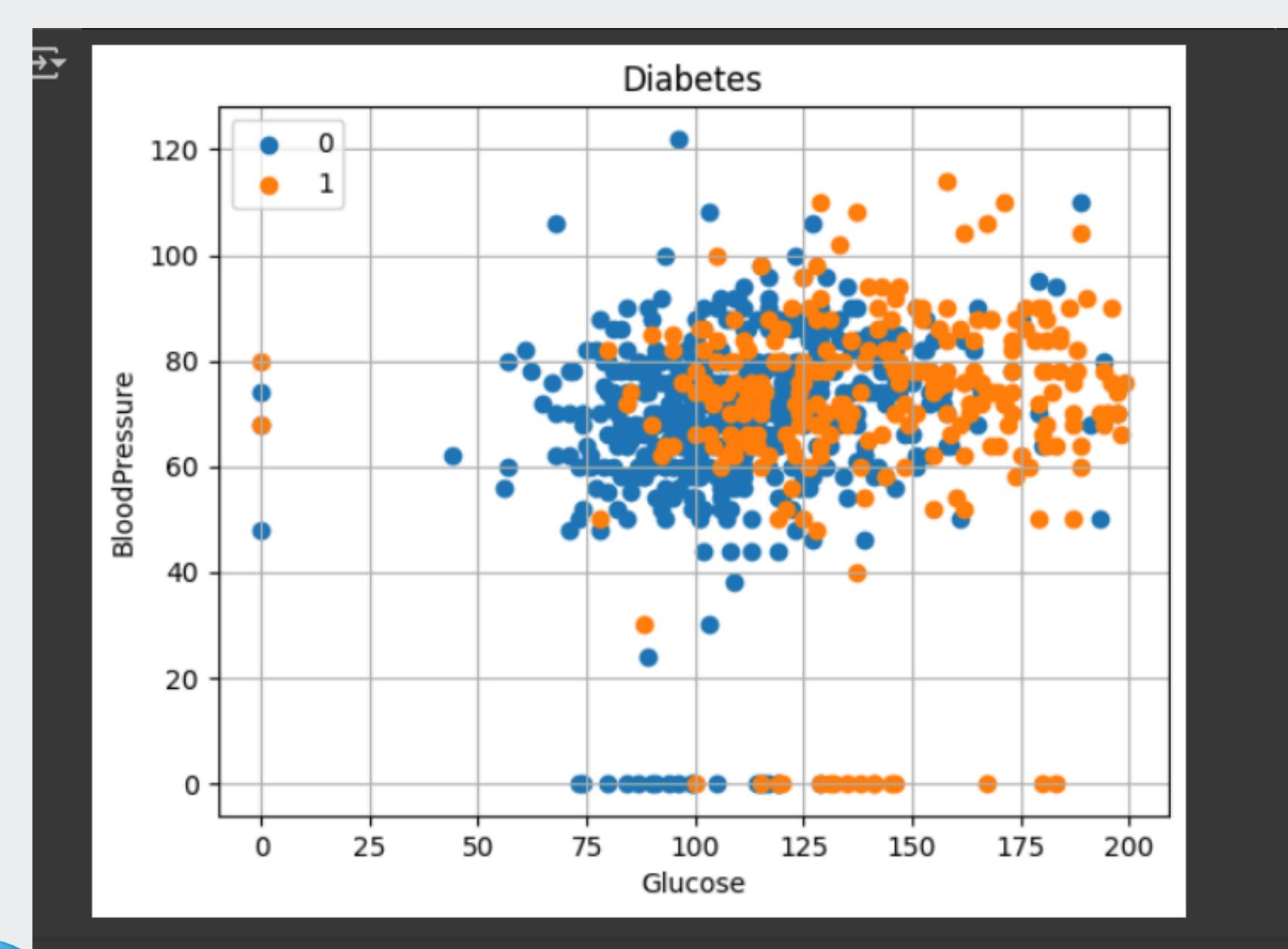
## Looping untuk Setiap Outcome:

```
for Outcome in ds.groupby('Outcome'):
```

- Kode ini melakukan iterasi melalui dataset ds, yang dikelompokkan berdasarkan kolom Outcome. Setiap iterasi akan memberikan nilai Outcome dan subset data yang terkait.
- ax.scatter() digunakan untuk menggambar scatter plot. Di sini, kolom Glucose dan BloodPressure dari dataset ds digunakan sebagai sumbu X dan Y, dan label diatur berdasarkan nilai Outcome.
- ax.legend() menampilkan legenda pada plot, yang membantu mengidentifikasi data berdasarkan label yang diberikan.
- plt.show() digunakan untuk menampilkan plot yang telah dibuat.

Secara keseluruhan, kode ini bertujuan untuk memvisualisasikan hubungan antara dua variabel, Glucose dan BloodPressure, berdasarkan kategori Outcome dalam dataset diabetes.

# Hasil Dari sebelah



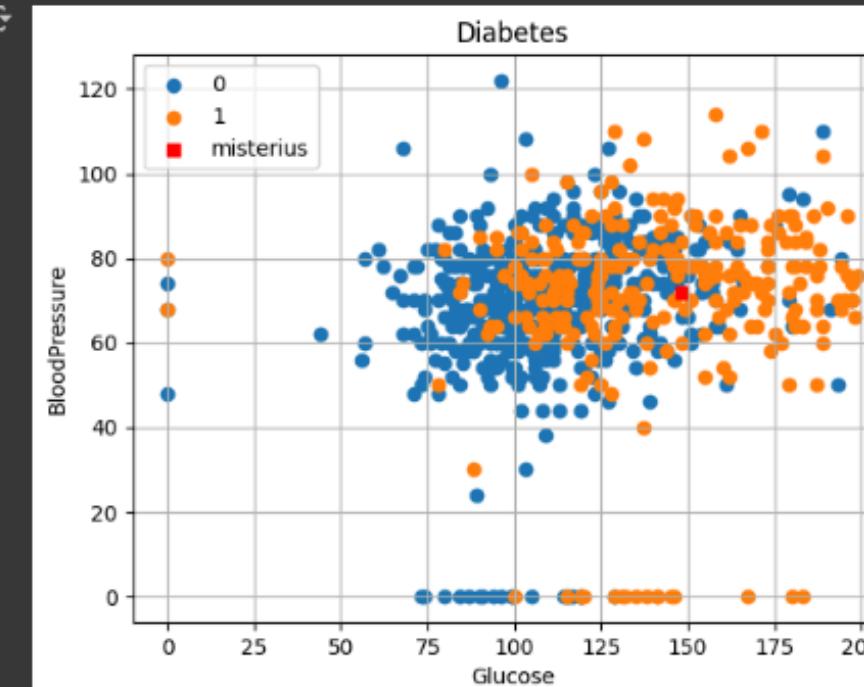
```

fig, ax = plt.subplots()
for Outcome, d in ds.groupby('Outcome'):
    ax.scatter(d['Glucose'], d['BloodPressure'], label=Outcome)

plt.scatter(Glucose,
            BloodPressure,
            color='red',
            marker='s',
            label = 'misterius')

plt.legend(loc='upper left')
plt.title('Diabetes')
plt.xlabel('Glucose')
plt.ylabel('BloodPressure')
plt.grid(True)
plt.show()
plt.show() ## kode ini bertujuan untuk memvisualisasikan data diabetes berdasarkan 'Glucose' dan 'BloodPressure', ##mengelompokkannya berdasarkan 'Outcome', dan menandai data 'misterius' pada plot. Ini membantu dalam memahami distribusi data dan hubungan antar variabel.

```



## Import Matplotlib:

- import matplotlib.pyplot as plt

- Pustaka matplotlib.pyplot diimpor dengan alias plt.

## Membuat Subplot:

```
fig, ax = plt.subplots()
```

- Figure dan axes dibuat untuk menyiapkan area plot.

## Looping untuk Setiap Outcome:

```
for Outcome in ds.groupby('Outcome'):
```

- Kode ini melakukan iterasi melalui dataset yang dikelompokkan berdasarkan kolom Outcome.

## Menggambar Scatter Plot:

```
ax.scatter(ds['Glucose'],
           ds['BloodPressure'], label=Outcome)
```

- Scatter plot digambar dengan Glucose sebagai sumbu X dan BloodPressure sebagai sumbu Y. Label plot ditentukan berdasarkan nilai Outcome.

# Penjelasan Lanjutan



Menambahkan Legenda:  
`ax.legend()`

- Menampilkan legenda pada plot untuk mengidentifikasi data berdasarkan kategori.

Menampilkan Plot:  
`plt.show()`

- Digunakan untuk menampilkan plot yang telah dibuat.
- Secara keseluruhan, kode ini bertujuan untuk memvisualisasikan distribusi data berdasarkan Glucose dan BloodPressure, serta menunjukkan hubungan antara variabel-variabel tersebut dengan membedakan kategori Outcome (misalnya, Misterius) dalam plot.

# Thank You Everyone

