

Noel Sagara

12030122140198

Pengkodean dan Pemrograman, Kelas D

```
import pandas as pd
```

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

```
import seaborn as sns
```

```
# 1. Memuat data
```

```
# Gantilah 'your_data.csv' dengan nama file data Anda
```

```
data = pd.read_csv('your_data.csv')
```

```
# 2. Memahami struktur data
```

```
print("Beberapa baris pertama dari data:")
```

```
print(data.head())
```

```
print("\nTipe data dari setiap kolom:")
```

```
print(data.dtypes)
```

```
# 3. Membersihkan data
```

```
# Mengecek nilai yang hilang
```

```
print("\nJumlah nilai yang hilang di setiap kolom:")
```

```
print(data.isnull().sum())
```

```
# Mengisi nilai yang hilang atau menghapus baris yang memiliki nilai hilang
```

```
# Misalnya, mengisi nilai hilang dengan median dari kolom tersebut
```

```
data.fillna(data.median(), inplace=True)
```

```
# 4. Eksplorasi data awal
```

```
# Menghitung statistik deskriptif
```

```
print("\nStatistik deskriptif:")
```

```
print(data.describe())
```

```
# Visualisasi dasar
```

```
# Histogram untuk distribusi kolom numerik
```

```
data.hist(bins=30, figsize=(10, 10))

plt.suptitle('Distribusi Kolom Numerik')

plt.show()
```

```
# Scatter plot untuk melihat hubungan antara dua variabel numerik

# Gantilah 'col1' dan 'col2' dengan nama kolom yang ingin Anda lihat

sns.scatterplot(x='col1', y='col2', data=data)

plt.title('Scatter Plot antara col1 dan col2')

plt.show()
```

```
# Heatmap dari matriks korelasi

plt.figure(figsize=(10, 8))

sns.heatmap(data.corr(), annot=True, cmap='coolwarm')

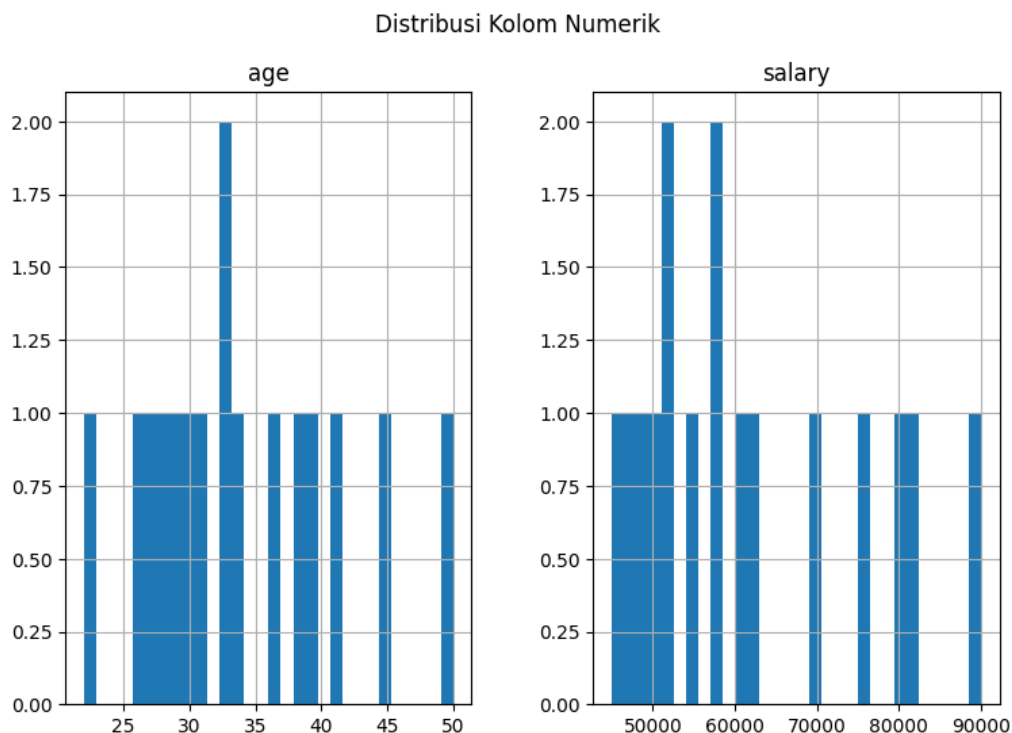
plt.title('Heatmap dari Matriks Korelasi')

plt.show()
```

id	name	age	salary	department	join date
1	John Doe	28	50000	Sales	15/01/2020
2	Jane Smith	34	60000	Marketing	22/03/2018
3	Emily Davis	22	45000	HR	01/06/2021
4	Michael Brown	45	80000	Finance	30/09/2015
5	Jessica Taylor	29	52000	IT	11/11/2019
6	Daniel Anderson	36	70000	Sales	14/02/2017
7	Sarah Lee	30	48000	HR	21/07/2020
8	James Wilson	38	62000	Marketing	05/12/2016
9	Laura Martin	27	51000	IT	12/01/2021
10	Robert Moore	50	90000	Finance	18/05/2010
11	Mary Clark	33		HR	30/08/2019
12	Patricia Lewis	41	75000	Sales	17/03/2016
13	Charles Walker	31	54000	Marketing	23/10/2018
14	Elizabeth Hall	26	47000	Marketing	19/04/2021
15	David Allen	39	82000	Finance	27/06/2015

Hasil :

## Gambar 1



### Interpretasi Grafik: Distribusi Kolom Numerik (Age dan Salary)

Grafik yang ditampilkan adalah histogram untuk kolom 'age' dan 'salary'. Berikut adalah interpretasi dari grafik tersebut:

#### Grafik Distribusi Age

- **Sumbu X (Horizontal):** Menunjukkan rentang usia (age) dari 20 hingga 50 tahun.
- **Sumbu Y (Vertikal):** Menunjukkan frekuensi atau jumlah kejadian untuk setiap rentang usia, dalam skala yang tampaknya normalisasi.

#### Pengamatan dari Grafik:

1. **Rentang Usia 25-30 Tahun:**
  - Frekuensi: Bervariasi antara 1 hingga 2.
  - Pengamatan: Rentang usia ini menunjukkan adanya beberapa individu di setiap interval usia, dengan puncak yang jelas pada usia 30 tahun dengan frekuensi 2.
2. **Rentang Usia 30-35 Tahun:**
  - Frekuensi: Puncak pada usia 30 tahun dengan frekuensi tertinggi (2).
  - Pengamatan: Usia 30 tahun memiliki jumlah kejadian tertinggi, sedangkan usia lain dalam rentang ini memiliki frekuensi yang lebih rendah dan lebih bervariasi.

### 3. Rentang Usia 35-50 Tahun:

- Frekuensi: Bervariasi antara 0 hingga 1.
- Pengamatan: Rentang usia ini menunjukkan distribusi yang lebih tersebar dengan frekuensi yang lebih rendah dan lebih bervariasi, dengan tidak ada usia tertentu yang dominan.

## Grafik Distribusi Salary

- **Sumbu X (Horizontal):** Menunjukkan rentang gaji (salary) dari 40.000 hingga 90.000.
- **Sumbu Y (Vertikal):** Menunjukkan frekuensi atau jumlah kejadian untuk setiap rentang gaji, dalam skala yang tampaknya normalisasi.

### Pengamatan dari Grafik:

#### 1. Rentang Gaji 50.000-60.000:

- Frekuensi: Bervariasi antara 1 hingga 2.
- Pengamatan: Rentang gaji ini menunjukkan adanya beberapa individu di setiap interval gaji, dengan puncak pada gaji sekitar 60.000 dengan frekuensi 2.

#### 2. Rentang Gaji 60.000-90.000:

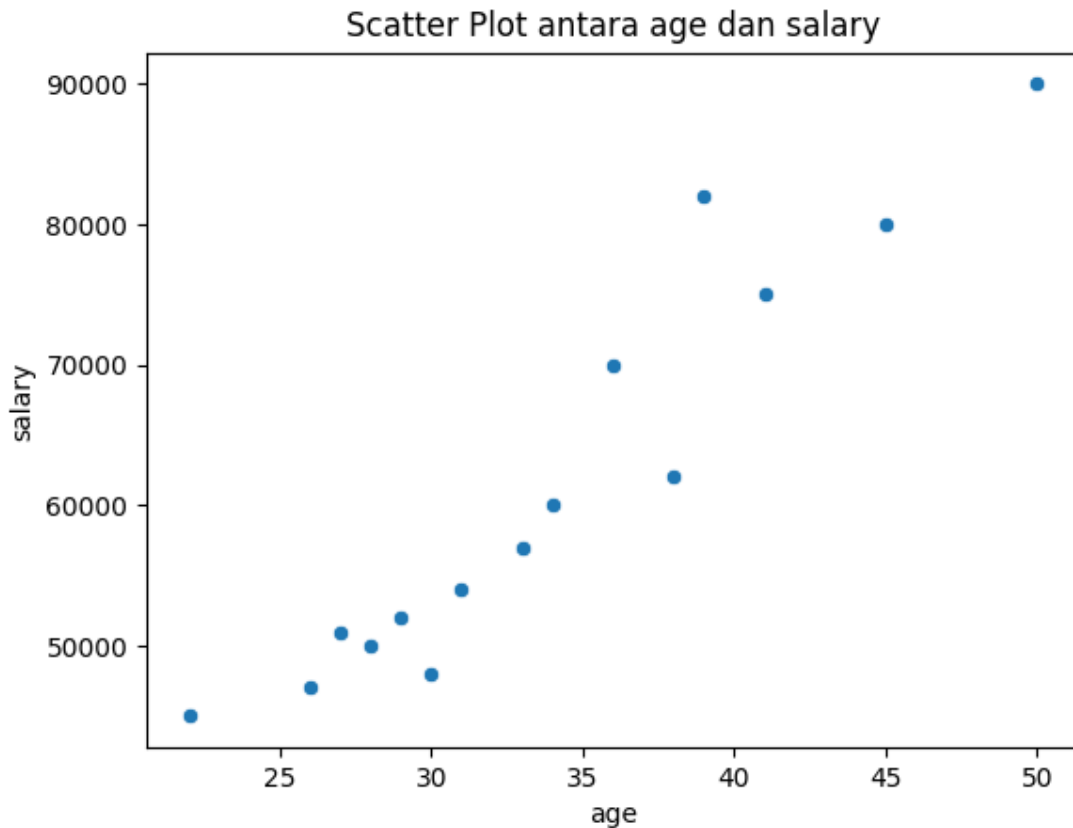
- Frekuensi: Bervariasi antara 0 hingga 1.
- Pengamatan: Rentang gaji ini menunjukkan distribusi yang lebih tersebar dengan frekuensi yang lebih rendah dan lebih bervariasi, dengan tidak ada gaji tertentu yang dominan.

## Kesimpulan

- **Distribusi Usia:** Distribusi usia menunjukkan bahwa sebagian besar individu berada di usia sekitar 30 tahun, dengan distribusi yang lebih tersebar pada usia yang lebih tinggi.
- **Distribusi Gaji:** Distribusi gaji menunjukkan bahwa sebagian besar individu memiliki gaji sekitar 50.000 hingga 60.000, dengan distribusi yang lebih tersebar pada gaji yang lebih tinggi.

Secara keseluruhan, grafik ini memberikan gambaran bahwa distribusi usia dan gaji dalam dataset cenderung berkonsentrasi pada nilai-nilai tertentu dengan penyebaran yang lebih luas pada rentang yang lebih tinggi.

**Gambar 2**



#### Interpretasi Grafik: Scatter Plot antara Age dan Salary

Grafik yang ditampilkan adalah scatter plot yang menunjukkan hubungan antara usia (age) dan gaji (salary). Berikut adalah interpretasi dari grafik tersebut:

- **Sumbu X (Horizontal):** Menunjukkan usia (age) dari 20 hingga 50 tahun.
- **Sumbu Y (Vertikal):** Menunjukkan gaji (salary) dari 40.000 hingga 90.000.

#### Pengamatan dari Grafik:

1. **Pola Hubungan:** Terlihat adanya hubungan positif antara usia dan gaji. Semakin tinggi usia, semakin tinggi gaji yang diterima.
2. **Rentang Usia 20-25 Tahun:**
  - Gaji sekitar: 40.000 hingga 50.000.
  - Pengamatan: Individu yang berada pada rentang usia ini cenderung memiliki gaji yang lebih rendah.
3. **Rentang Usia 25-30 Tahun:**
  - Gaji sekitar: 45.000 hingga 60.000.
  - Pengamatan: Terjadi kenaikan gaji yang signifikan seiring bertambahnya usia dalam rentang ini.
4. **Rentang Usia 30-35 Tahun:**
  - Gaji sekitar: 55.000 hingga 70.000.

- Pengamatan: Kenaikan gaji terus berlanjut dengan peningkatan yang konsisten.
- 5. **Rentang Usia 35-40 Tahun:**
  - Gaji sekitar: 65.000 hingga 80.000.
  - Pengamatan: Peningkatan gaji lebih terlihat pada rentang usia ini.
- 6. **Rentang Usia 40-50 Tahun:**
  - Gaji sekitar: 75.000 hingga 90.000.
  - Pengamatan: Individu dalam rentang usia ini cenderung memiliki gaji tertinggi.

### **Kesimpulan:**

- **Hubungan Positif:** Scatter plot menunjukkan adanya hubungan positif yang kuat antara usia dan gaji. Hal ini berarti seiring dengan bertambahnya usia, gaji cenderung meningkat.
- **Pola Distribusi:** Titik-titik data tersebar membentuk pola linear naik, yang menunjukkan peningkatan gaji yang stabil seiring bertambahnya usia.
- **Variabilitas Gaji:** Variabilitas atau penyebaran gaji lebih terlihat pada rentang usia yang lebih tinggi (40-50 tahun) dibandingkan rentang usia yang lebih muda.

Secara keseluruhan, grafik ini memberikan gambaran bahwa usia mempengaruhi gaji, di mana individu yang lebih tua cenderung memiliki gaji yang lebih tinggi dibandingkan dengan individu yang lebih muda.