

Lanjutan analisis data

1. Apa fungsi library pandas dalam analisis data?

Untuk mengolah, membersihkan, dan menganalisis data dalam bentuk tabel (DataFrame) dengan mudah.

2. Apa perbedaan antara matplotlib dan seaborn?

- Matplotlib: library visualisasi dasar, lebih low-level, harus mengatur banyak hal manual.
- Seaborn: dibangun di atas Matplotlib, tampilannya lebih menarik dan mudah digunakan untuk visualisasi statistik.

3. Mengapa perlu melakukan data cleaning sebelum analisis?

Agar data bersih, akurat, dan konsisten, sehingga hasil analisis tidak salah atau menyesatkan.

4. Fungsi apa yang digunakan untuk menampilkan 5 data pertama dari DataFrame?

Menggunakan fungsi:

df.head()

atau

df.head(5)

5. Bagaimana cara menyimpan data hasil analisis ke file CSV?

Menggunakan:

df.to_csv('nama_file.csv', index=False)

codingan

```
import pandas as pd
```

```
import numpy as np
```

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

```
import seaborn as sns
```

```
df = pd.read_csv("Data_Penjualan_Toko_Online.csv")
```

```
print(df.head())
```

```
print(df.info())
```

```
print(df.isnull().sum())
```

```
# Jika ada nilai kosong
```

```

df = df.dropna()

print("Statistik Deskriptif:")
print(df.describe())

print("\nProduk Terlaris:")
print(df.groupby("Produk")["Jumlah"].sum().sort_values(ascending=False))

plt.figure(figsize=(8,5))
sns.barplot(data=df, x="Produk", y="Pendapatan", estimator=sum,
ci=None)
plt.title("Total Pendapatan per Produk")
plt.show()

plt.figure(figsize=(10,5))
sns.lineplot(data=df, x="Tanggal", y="Pendapatan", marker="o")
plt.title("Tren Pendapatan Mingguan")
plt.show()

df.to_csv("hasil_analisis_penjualan.csv", index=False)

print("Data hasil analisis telah disimpan sebagai hasil_analisis_penjualan.csv")

```

Hasil terminal

	Tanggal	Produk	Kategori	Harga	Jumlah	Pendapatan
0	2024-01-07	Laptop	Elektronik	8000000	1	8000000
1	2024-01-14	Mouse	Aksesoris	150000	2	300000
2	2024-01-21	Keyboard	Aksesoris	250000	4	1000000
3	2024-01-28	Monitor	Elektronik	2000000	9	18000000
4	2024-02-04	Headset	Aksesoris	300000	1	300000

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 25 entries, 0 to 24

Data columns (total 6 columns):

```
# Column    Non-Null Count Dtype
---  --  ---  --
0  Tanggal    25 non-null   object
1  Produk     25 non-null   object
2  Kategori   25 non-null   object
3  Harga      25 non-null   int64
4  Jumlah     25 non-null   int64
5  Pendapatan 25 non-null   int64
dtypes: int64(3), object(3)
memory usage: 1.3+ KB
None
Tanggal    0
Produk     0
Kategori   0
Harga      0
Jumlah     0
Pendapatan 0
dtype: int64
Statistik Deskriptif:
          Harga   Jumlah   Pendapatan
count  2.500000e+01 25.000000 2.500000e+01
mean   2.642000e+06 4.520000 1.011000e+07
std    3.459639e+06 2.725191 1.653172e+07
min    1.500000e+05 1.000000 3.000000e+05
25%   2.500000e+05 2.000000 7.500000e+05
50%   3.000000e+05 5.000000 2.700000e+06
75%   2.000000e+06 6.000000 1.000000e+07
max   1.000000e+07 9.000000 7.200000e+07
```

Produk Terlaris:

Produk

Monitor 33

Keyboard 25

Laptop 21

Mouse 18

Headset 16

Name: Jumlah, dtype: int64

