

Modul Praktikum

Pertemuan 4 – Exploratory Data Analysis (EDA)

Mata Kuliah Data Science

September 7, 2025

Tujuan

- Mahasiswa mampu melakukan eksplorasi data (EDA).
- Mahasiswa dapat menghitung statistik deskriptif menggunakan Pandas.
- Mahasiswa dapat membuat visualisasi sederhana dengan Matplotlib/Seaborn.
- Mahasiswa dapat menarik insight awal dari dataset Titanic.

Persiapan

1. Pastikan Python 3 terinstal dengan library:
 - pandas
 - matplotlib
 - seaborn
2. Unduh dataset `titanic.csv` dari Kaggle: <https://www.kaggle.com/c/titanic>
3. Simpan file `titanic.csv` di folder kerja Anda.

Langkah Praktikum

1. Membaca Dataset

```
import pandas as pd

# membaca file Titanic
df = pd.read_csv("titanic.csv")

# tampilkan 5 baris pertama
print(df.head())

# informasi dataset
print(df.info())
```

2. Statistik Deskriptif Numerik

```
# ringkasan umur
print(df["Age"].describe())

# ringkasan harga tiket
print(df["Fare"].describe())
```

3. Statistik Deskriptif Kategorikal

```
# distribusi jenis kelamin
print(df["Sex"].value_counts())

# distribusi kelas tiket
print(df["Pclass"].value_counts())
```

4. Visualisasi Dasar

```
import matplotlib.pyplot as plt
import seaborn as sns

# Histogram Age
df["Age"].hist(bins=20)
plt.title("Distribusi Umur Penumpang")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.show()

# Countplot Sex
sns.countplot(x="Sex", data=df)
plt.title("Distribusi Jenis Kelamin Penumpang")
plt.show()

# Boxplot Fare vs Pclass
sns.boxplot(x="Pclass", y="Fare", data=df)
plt.title("Distribusi Fare berdasarkan Kelas")
plt.show()

# Scatterplot Age vs Fare dengan hue Survived
sns.scatterplot(x="Age", y="Fare", hue="Survived", data=df)
plt.title("Hubungan Age dan Fare terhadap Survival")
plt.show()
```

Soal Latihan

1. Hitung rata-rata umur penumpang (Age) dan median harga tiket (Fare).
2. Buat bar chart distribusi penumpang berdasarkan Pclass.

3. Buat boxplot harga tiket (`Fare`) berdasarkan jenis kelamin (`Sex`).
4. Buat scatterplot antara `Age` dan `Fare` dengan pewarnaan berdasarkan `Pclass`.

Tugas Pertemuan 4

Kerjakan eksplorasi data (EDA) pada dataset Titanic. Laporan (2–3 halaman, PDF) berisi:

- Deskripsi dataset (jumlah baris/kolom, tipe data).
- Statistik deskriptif untuk variabel numerik dan kategorikal.
- Minimal 3 visualisasi (histogram, bar chart, boxplot, scatter plot).
- Insight yang diperoleh dari hasil eksplorasi.