

Laporan Praktikum 5

Visualisasi Data dengan Matplotlib, Seaborn, dan Plotly

Peno
221220095
September 30, 2025

Identitas Praktikum

Mata Kuliah : Data Science
Pertemuan : 5
Judul : Visualiasi Data (Matplotlib, Seaborn, Plotly)
Nama : Peno
NIM : 221220095
Tanggal : 30 Oktober 2025

Tujuan

- Mahasiswa memahami peran visualisasi dalam analisis data.
- Mahasiswa dapat membuat visualisasi dasar dengan Matplotlib.
- Mahasiswa dapat membuat visualisasi statistik dengan Seaborn.
- Mahasiswa mengenal visualisasi interaktif dengan Plotly.

Dasar Teori

- Visualisasi data membantu memahami pola, tren, distribusi, dan outlier dalam data.
- **Matplotlib** adalah library visualisasi dasar di Python. Ini digunakan untuk membuat plot standar seperti line plot, bar chart, histogram, dan scatter plot.
- **Seaborn** adalah *library* yang dibangun di atas Matplotlib untuk membuat visualisasi statistik yang lebih menarik dan informatif, seperti countplot, boxplot, dan heatmap.
- **Plotly** adalah *library* yang digunakan untuk membuat visualisasi interaktif yang modern. Pengguna dapat melakukan *zoom*, *hover* (menampilkan info saat kursor di atas data), dan memfilter data langsung pada grafik.

Alat dan Bahan

1. Python 3
2. Library: pandas, matplotlib, seaborn, plotly
3. Dataset: titanic.csv (Kaggle)

Langkah Kerja

1. Membaca dataset Titanic dengan Pandas.
2. Membuat visualisasi dasar dengan Matplotlib (seperti histogram).
3. Membuat visualisasi statistik dengan Seaborn (seperti countplot dan boxplot).
4. Membuat visualisasi interaktif dengan Plotly (seperti scatter plot).
5. Menarik insight awal dari hasil visualisasi yang dibuat.

Kode Program

```
# 1. Import library yang diperlukan
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

# 2. Membaca dataset
df = pd.read_csv("titanic.csv")

# 3. Persiapan Data: Mengisi nilai 'Age' yang hilang (NaN)
# dengan rata-rata umur
df['Age'] = df['Age'].fillna(df['Age'].mean())

# Visualisasi 1: Countplot Status 'Survived' (Seaborn -
# Kategorikal)
plt.figure(figsize=(7, 5))
sns.countplot(x='Survived', data=df, palette='pastel')
plt.title('Distribusi Penumpang Selamat (1) vs Tidak Selamat
(0)')
plt.xlabel('Status Keselamatan')
plt.ylabel('Jumlah Penumpang')
plt.xticks([0, 1], ['Tidak Selamat', 'Selamat'])
plt.show()

# Visualisasi 2: Histogram Distribusi 'Age' (Matplotlib -
# Numerik)
plt.figure(figsize=(8, 5))
plt.hist(df['Age'], bins=30, color='skyblue', edgecolor='black')
plt.title('Distribusi Umur Penumpang')
plt.xlabel('Umur')
```

```

plt.ylabel('Frekuensi')
plt.show()

# Visualisasi 3: Boxplot 'Fare' berdasarkan 'Pclass' (Seaborn - Numerik vs Kategorikal)
plt.figure(figsize=(8, 6))
sns.boxplot(x='Pclass', y='Fare', data=df, palette='Set2')
plt.title('Distribusi Harga Tiket (Fare) berdasarkan Kelas (Pclass)')
plt.xlabel('Kelas Penumpang')
plt.ylabel('Harga Tiket')
plt.ylim(0, 300)
plt.show()

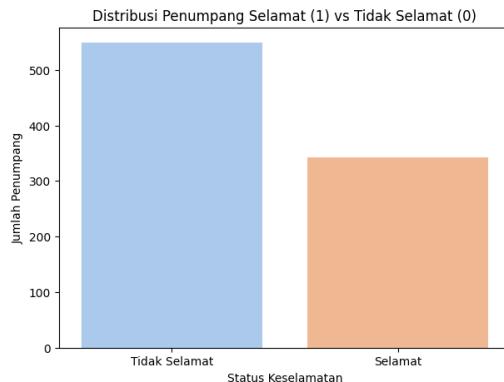
# Visualisasi 4: Countplot 'Embarked' berdasarkan 'Survived' (Seaborn - Kategorikal vs Kategorikal)
plt.figure(figsize=(8, 5))
sns.countplot(x='Embarked', hue='Survived', data=df)
plt.title('Tingkat Keselamatan berdasarkan Pelabuhan Embarkasi')
plt.xlabel('Pelabuhan (C=Cherbourg, Q=Queenstown, S=Southampton)')
plt.ylabel('Jumlah Penumpang')
plt.legend(title='Selamat', labels=['Tidak', 'Ya'])
plt.show()

# Visualisasi 5: Scatter Plot Interaktif 'Age' vs 'Fare' (Plotly - Interaktif)
fig = px.scatter(df,
                  x="Age",
                  y="Fare",
                  color="Pclass",
                  hover_data=['Sex', 'Survived'],
                  title="Scatter Plot Interaktif: Umur vs Harga Tiket (diwarnai Pclass)")
fig.show()

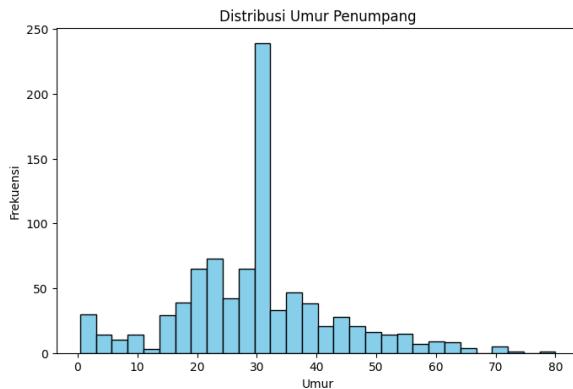
```

Hasil dan Output

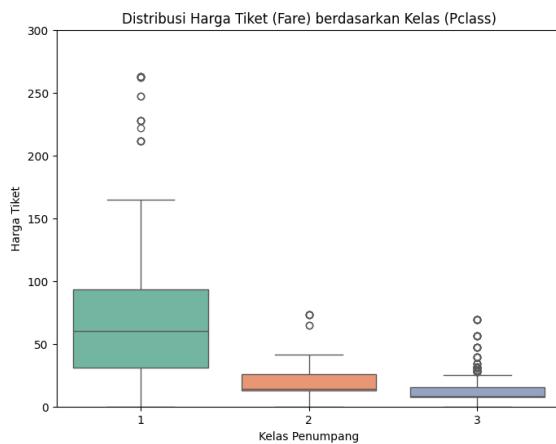
- Visualisasi 1: Countplot Status 'Survived' (Seaborn - Kategorikal):



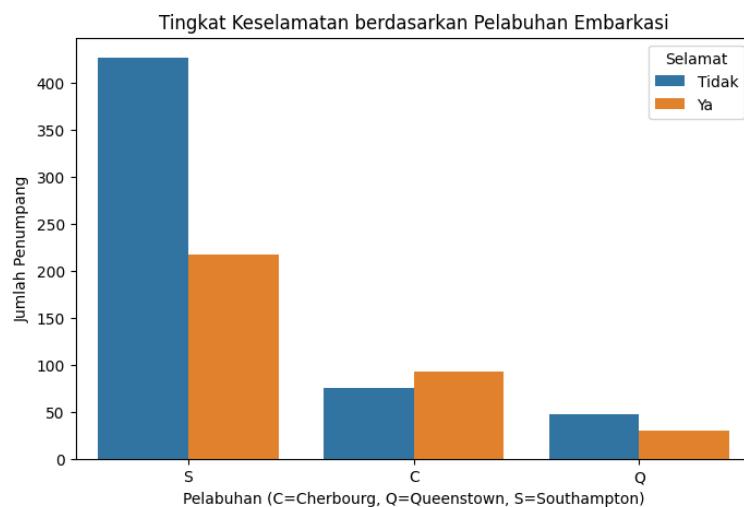
- Visualisasi 2: Histogram Distribusi 'Age' (Matplotlib - Numerik):



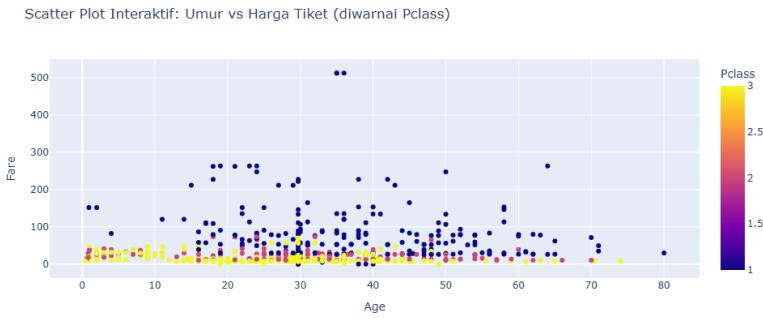
- Visualisasi 3: Boxplot 'Fare' berdasarkan 'Pclass' (Seaborn - Numerik vs Kategorikal):



- Visualisasi 4: Countplot 'Embarked' berdasarkan 'Survived' (Seaborn - Kategorikal vs Kategorikal):



- Visualisasi 5: Scatter Plot Interaktif 'Age' vs 'Fare' (Plotly - Interaktif).



Pembahasan

Berdasarkan 5 visualisasi yang telah dibuat, berikut adalah interpretasi singkatnya:

- **Distribusi Penumpang Selamat:** Dari *countplot* pertama, terlihat jelas bahwa jumlah penumpang yang **tidak selamat (0)** secara signifikan **lebih banyak** daripada jumlah penumpang yang **selamat (1)**.
- **Distribusi Umur Penumpang:** *Histogram* umur menunjukkan bahwa sebagian besar penumpang berusia antara **20 hingga 30 tahun** (setelah nilai *missing* diisi). Distribusinya sedikit miring ke kanan (*right-skewed*).
- **Harga Tiket per Kelas:** *Boxplot* ini dengan jelas menunjukkan perbedaan ekonomi antar kelas. **Kelas 1** memiliki harga tiket (Fare) yang jauh lebih mahal dan dengan rentang harga yang lebih lebar. **Kelas 3** memiliki harga tiket yang paling murah dan terkonsentrasi di bawah 50.
- **Keselamatan berdasarkan Pelabuhan:** *Countplot* ini (sesuai Latihan 3) menunjukkan bahwa mayoritas penumpang naik dari pelabuhan '**S**' (**Southampton**), dan grup ini juga memiliki angka kematian tertinggi. Menariknya, penumpang yang naik dari '**C**' (**Cherbourg**) memiliki proporsi selamat yang lebih tinggi.
- **Scatter Plot (Umur, Tiket, Kelas):** *Scatter plot* interaktif (sesuai Latihan 5) menunjukkan bahwa tidak ada korelasi yang kuat antara Umur dan Harga Tiket. Namun, *insight* utamanya adalah tiket dengan harga **sangat mahal** (misal, Fare > 100) hampir secara eksklusif dimiliki oleh penumpang **Kelas 1**.

Kesimpulan

1. Visualisasi data (menggunakan Matplotlib, Seaborn, dan Plotly) sangat mempermudah dalam mengidentifikasi pola, distribusi, dan *outlier* pada dataset Titanic.

2. Matplotlib efektif untuk plot dasar (histogram), Seaborn sangat baik untuk visualisasi statistik (boxplot, countplot) yang membandingkan variabel, dan Plotly unggul dalam visualisasi interaktif (scatter plot).
3. Dari analisis visual awal, terlihat jelas bahwa faktor-faktor seperti kelas tiket (`Pclass`), pelabuhan (`Embarked`), dan harga tiket (`Fare`) memiliki hubungan yang terlihat jelas dengan data penumpang dan tingkat keselamatan mereka.