

KUIS : Visualisasi Data

Reza Chairul Manam [120140086]

26 Maret 2025

1 Soal

1.1 Perhitungan Tipe Dataset Kuis Visdat

Diketahui tipe data soal dengan rumus:

$$\text{Tipe Dataset Kuis Visdat} = \frac{4!}{3! \times 2!} - (M \bmod 2)^3 \times (0!)$$

Dengan $M = 6$ (digit terakhir NIM), maka:

Hitung Faktorial: $4! = 24$, $3! = 6$, $2! = 2$, $0! = 1$.

Hitung Bagian Kiri: $\frac{4!}{3! \times 2!} = \frac{24}{6 \times 2} = \frac{24}{12} = 2$.

Hitung Bagian Kanan: $M \bmod 2 = 6 \bmod 2 = 0$, $(0)^3 = 0$, $(0)^3 \times 0! = 0 \times 1 = 0$.

Hasil Akhir: $2 - 0 = 2$.

Jadi, hasilnya adalah 2.

Link Spreadsheet Kuis: *Dataset 2*

Link Github Kuis: *Kuis VisDat*

2 Pre-processing / Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk membersihkan dan menyiapkan data sebelum dianalisis. Teknik yang digunakan adalah **Variable Transformation**.

Teknik **Variable Transformation** dipilih sebagai metode pra-pemrosesan yang paling efisien untuk *Dataset 2 - Consumer Behavior* karena dapat meningkatkan kualitas data tanpa kehilangan informasi. Transformasi ini diperlukan untuk memastikan bahwa data memiliki format yang sesuai agar siap digunakan dalam analisis. Misalnya, kolom *Purchase_Amount* yang awalnya berupa string dengan simbol \$ dikonversi menjadi numerik agar dapat dihitung dan divisualisasikan dengan benar. Selain itu, *Time_of_Purchase* diubah ke format *datetime*, memungkinkan analisis tren pembelian berdasarkan waktu.

Kolom biner seperti *Discount_Used* dan *Customer_Loyalty_Program_Member* juga dikonversi menjadi angka 0 dan 1, sehingga lebih mudah diproses dalam analisis statistik. Dengan transformasi ini, dataset menjadi lebih bersih, konsisten, dan siap digunakan untuk eksplorasi lebih lanjut, termasuk pembuatan visualisasi dan pengambilan keputusan berbasis data.

2.1 Implementasi dalam Python

Berikut adalah skrip Python untuk melakukan pra-pemrosesan data dengan teknik **Variable Transformation**:

Listing 1: Pra-pemrosesan Data dengan Variable Transformation

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4 # Load dataset
5 file_path = "Dataset_2_-_consumer_behavior.csv"
6 df = pd.read_csv(file_path)
7 # Variable Transformation
8 # 1. Menghapus simbol '$' dan mengubah ke numerik
9 df['Purchase_Amount'] = df['Purchase_Amount'].replace('[\$,]', '', regex=True).astype(float)
10 # 2. Mengonversi Time_of_Purchase ke format datetime
11 df['Time_of_Purchase'] = pd.to_datetime(df['Time_of_Purchase'])
12 # 3. Mengubah Discount_Used dan Customer_Loyalty_Program_Member ke format numerik (0 dan 1)
13 df['Discount_Used'] = df['Discount_Used'].map({'Yes': 1, 'No': 0})
14 df['Customer_Loyalty_Program_Member'] = df['Customer_Loyalty_Program_Member'].map({'Yes': 1, 'No': 0})
15 # Menampilkan hasil transformasi
16 print(df.head())
17 # Simpan hasil ke file baru jika diperlukan
18 df.to_csv("processed_dataset.csv", index=False)
```

3 Visualisasi Data

3.1 Implementasi dalam Python

Berikut adalah skrip Python untuk melakukan visualisasi data dalam bentuk Bar Chart dan Histogram:

Listing 2: Visualisasi data dalam bentuk Barchart dan Histogram

```
1 # Visualisasi 1: Bar Chart - Jumlah pelanggan berdasarkan kategori usia
2 plt.figure(figsize=(10, 5))
3 sns.countplot(x='Age', data=df, palette='viridis')
4 plt.title('Jumlah_Pelanggan_Berdasarkan_Kategori_Usia')
5 plt.xlabel('Usia')
6 plt.ylabel('Jumlah_Pelanggan')
7 plt.xticks(rotation=45)
8 plt.show()
9 # Visualisasi 2: Histogram - Distribusi Jumlah Pembelian
10 plt.figure(figsize=(10, 5))
11 sns.histplot(df['Purchase_Amount'], bins=20, kde=True, color='blue')
12 plt.title('Distribusi_Jumlah_Pembelian_(Purchase_Amount)')
13 plt.xlabel('Jumlah_Pembelian_($)')
14 plt.ylabel('Frekuensi')
15 plt.show()
```

Berikut adalah hasil visualisasi data:

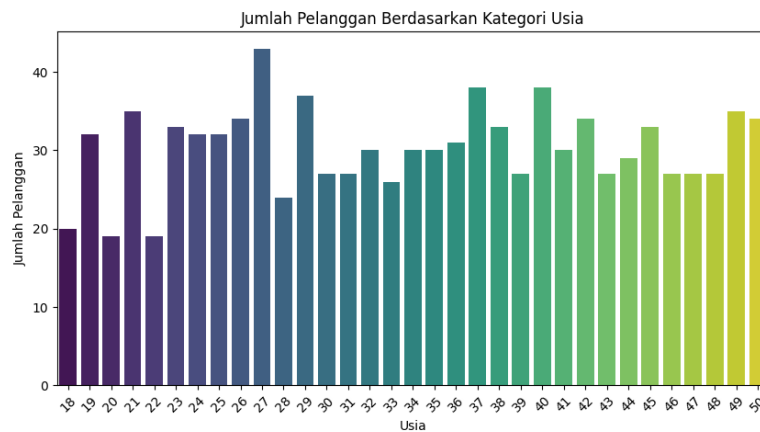


Figure 1: Visualisasi Data Pertama dengan Bar Chart

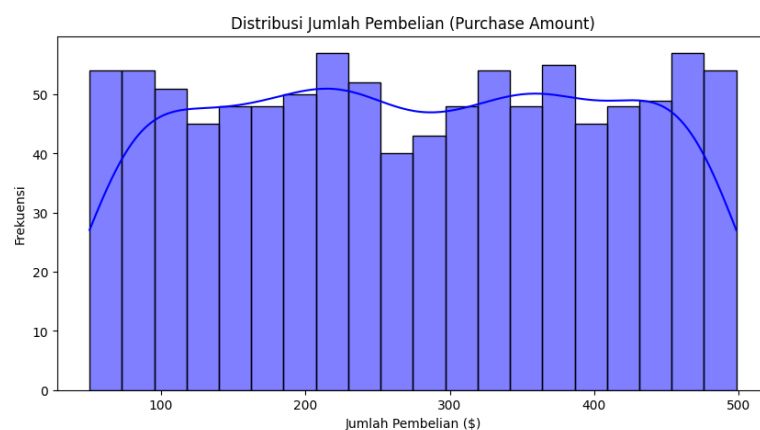


Figure 2: Visualisasi Data Kedua dengan histogram

4 Analisis Hasil Visualisasi

4.1 Analisis Bar Chart (Visualisasi Pertama)

Alasan Pemilihan Chart:

- Bar chart dipilih karena data yang divisualisasikan bersifat kategorikal, yaitu kategori usia pelanggan.
- Bar chart memungkinkan perbandingan jumlah pelanggan dalam setiap kategori usia dengan jelas.

Pemilihan Warna:

- Menggunakan gradasi warna dari ungu ke hijau untuk memberikan efek visual yang menarik dan memudahkan pemahaman distribusi data.
- Warna yang lebih gelap menunjukkan jumlah pelanggan yang lebih sedikit, sementara warna lebih terang menunjukkan jumlah pelanggan lebih banyak.

4.2 Analisis Histogram (Visualisasi Kedua)

Alasan Pemilihan Chart:

- Histogram dipilih untuk menampilkan distribusi jumlah pembelian (*Purchase Amount*).
- Dengan histogram, kita bisa melihat apakah data memiliki pola distribusi tertentu, misalnya apakah data normal, miring ke kiri/kanan, atau memiliki *outlier*.

Pemilihan Warna:

- Histogram menggunakan warna biru dengan transparansi untuk memberikan tampilan yang lebih *soft*.
- Ditambahkan garis distribusi (*kde plot*) untuk membantu memahami pola penyebaran data.