Tugas Analisis Visualisasi Data Praktikum

Tugas 2 : Eksplorasi Data



Oleh :

Ferry Triwantono – 082111633094

**FAKULTAS SAINS DAN TEKNOLOGI**

**PROGRAM STUDI SISTEM INFORMASI**

**UNIVERSITAS AIRLANGGA**

**2023**

**Boxplot, Histogram, dan Heapmap**

1. Import package dan dataset

*# Import dataset dan package*

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

dataset = pd.read\_csv(

    'D:\Coolyeah\Mata Kuliah\SMT 5\Analisis Visualisasi Data Praktikum\Boxplot, Histogram, Scatter Plot, dan Heat am\heart.csv')

df = pd.DataFrame(dataset)

Syntax tersebut digunakan untuk mengimpor Pustaka atau modul dan juga dataset yang telah disesuaikan dengan perintah pd.DataFrame yang diperlukan untuk memproses, menganalisis, dan menggambar visualisasi data.

*# 3. Cek apakah terdapat Outlier, analisa, dan visualisasikan dengan boxplot*

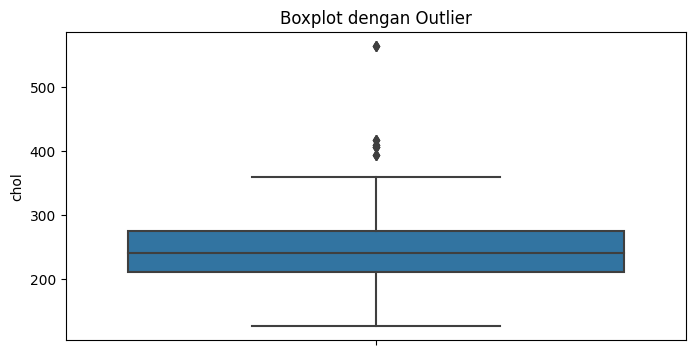
plt.subplots(*figsize*=(8, 4), *dpi*=100)

sns.boxplot(*y*=df['chol'])

plt.title("Boxplot dengan Outlier")

plt.show

Kemudian setelah dataset dan modul telah diimport kita akan memerlukan beberapa syntax seperti pada contoh diatas untuk membuat boxplot dengan seaborn menggunakan matplotlib sebagai control pengaturan tampilan.

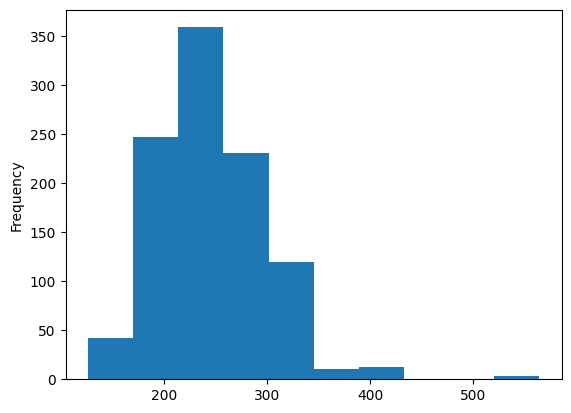
****

**Gambar 1.1** Visualisasi Boxplot dengan Outlier

*# 4. Tentukan distribusi dari dataset tersebut, analisa, dan visualisasikan dengan Histogram*

df['chol'].plot.hist()

Digunakan untuk membuat histogram dari data yang terdapat pada kolom ‘chol’ dari Dataframe yang telah dibuat yaitu ‘df’ dan berasal dari dataset utama. Hasilnya akan menjadi histogram yang menggambarkan distribusi data dalam kolom ‘chol’.



**Gambar 1.2** Visualisasi Histogram

*# 5. Carilah korelasi antar variabel dengan variabel output, analisa, dan visualisasikan dengan Scatterplot dan Heatmap*

y = df['age']

x = df['chol']

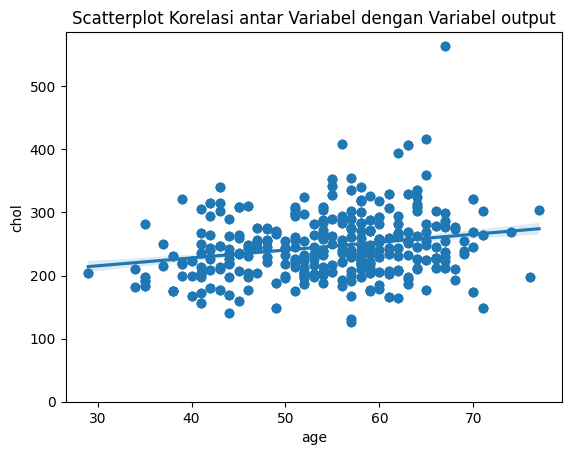
*# a. Scatterplot*

plt.title("Scatterplot Korelasi antar Variabel dengan Variabel output")

plt.scatter(x, y)

plt.ylim(0,)

Hasilnya adalah scatterplot yang menampilkan hubungan antara dua variabel dengan judul Scatterplot Korelasi antar Variabel dengan Variabel output dengan sumbu y pada plot dibatasi mulai dari 0 ke atas tanpa batas atas yang ditentukan



**Gambar 1.3** Visualisasi Scatterplot Korelasi antar Variabel dengan Variabel output

*# b. Heapmap*

select\_column = ['age', 'trestbps', 'chol']

dfbaru = df[select\_column]

correlation\_matrix = dfbaru.corr()

plt.figure(*figsize*=(8, 6))

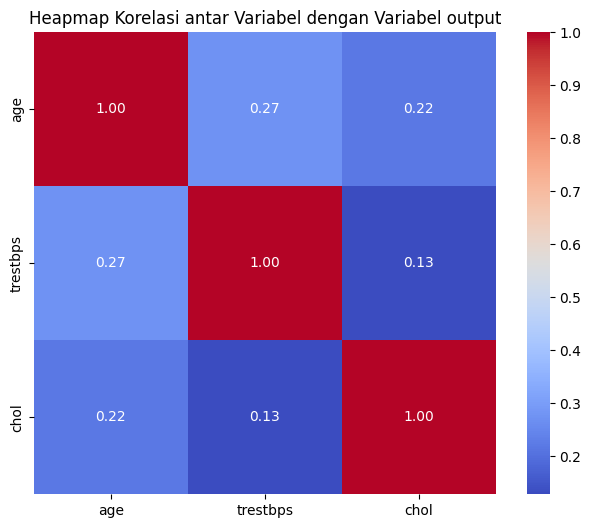
sns.heatmap(correlation\_matrix, *annot*=True,

*cmap*='coolwarm', *fmt*=".2f", *square*=True)

plt.title("Heatmap Korelasi antar Variabel dengan Variabel output")

plt.show()

Hasilnya akan berupa heapmap yang menunjukan korelasi antara kolom ‘age’, ‘trestbps’, dan ‘chol’ dalam dataframe ‘dfbaru’. Heapmap ini memberikan gambaran visual tentang seberapa kuat atau lemah hubungan korelasi antara kolom tersebut



**Gambar 1.4** Visualisasi Heapmap Korelasi antar Variabel dengan Variabel Output