| **Nama:**  **(Adrian Halim)**  **NIM: (064002200043)** | C:\Users\RPL-SI 02\Pictures\288px-Trisakti_Logo.svg.png | **MODUL 8**  **Nama Dosen:**  **Dedy Sugiarto** |
| --- | --- | --- |
| **Hari/Tanggal:**  **Senin, 7 AGUSTUS 2023** | **Praktikum Statistika** | **Nama Asisten Labratorium**  **1. Elen Fadilla Estri**  **064002000008**  **2. Rukhy Zaifa Aduhalim**  **064002000041** |

**Eksplorasi Data Menggunakan Python**

1. **Teori Singkat**

histogram berguna untuk memberikan gambaran ukuran tendensi sentral dan kesimetrisan data pengamatan. Penyajian grafis lainnya yang bisa merangkum informasi lebih detail mengenai distribusi nilai-nilai data pengamatan adalah Box and Whisker Plots atau lebih sering disebut dengan BoxPlot atau Box-Plot (kotak-plot) saja. Seperti namanya, Box and Whisker, bentuknya terdiri dari Box (kotak) dan whisker.

Box-plot atau boxplot (juga dikenal sebagai diagram box-and-whisker) merupakan suatu box (kotak berbentuk bujur sangkar). Boxplot adalah salah satu cara dalam statistik deskriptif untuk menggambarkan secara grafik dari data numeris melalui lima ukuran sebagai berikut:

* Nilai observasi terkecil,
* Kuartil terendah atau kuartil pertama (Q1), yang memotong 25% dari data terendah
* Median (Q2) atau nilai pertengahan,
* Kuartil tertinggi atau kuartil ketiga (Q3), yang memotong 25% dari data terbesar
* Nilai observasi terbesar.

Dalam boxplot juga ditunjukkan, jika ada, nilai outlier dari observasi. Boxplot dapat digunakan untuk menunjukkan perbedaan antara populasi tanpa menggunakan asumsi distribusi statistik yang mendasarinya. Karenanya, boxplot tergolong dalam statistik non-parametrik. Jarak antara bagian-bagian dari box menunjukkan derajat dispersi (penyebaran) dan skewness (kecondongan) dalam data. Dalam penggambarannya, boxplot dapat digambarkan secara horizontal maupun vertikal.

1. **Alat dan Bahan**

Hardware : Laptop/PC

Software : R Studio

1. **Elemen Kompetensi**
   1. Latihan pertama – Praktikum

1. Buka Jupyter Notebook atau Google Colab di Browser

|  |
| --- |

2. Lalu jalankan script berikut dan berikan output (gunakan nama variable data dengan nama masing-masing)

| import pandas as pd  from pandas.tools import plotting  import matplotlib.pyplot as plt  import numpy as np  from sklearn.model\_selection import train\_test\_split, cross\_val\_score, KFold, GridSearchCV  from sklearn.linear\_model import LogisticRegression  from sklearn.tree import DecisionTreeClassifier,export\_graphviz  from sklearn.metrics import confusion\_matrix, accuracy\_score  from sklearn.ensemble import GradientBoostingClassifier, RandomForestClassifier |
| --- |

Output:

|  |
| --- |

3. Script

| data\_nama = pd.read\_csv('C:/prakstatik/houseprice.csv')  data\_nama.head(5) |
| --- |

Output:

|  |
| --- |

4. Boxplot dari harga rumah/Price

| col\_list=['Price']  numhouse = data\_nama[data\_nama.columns[data\_nama.columns.isin(col\_list)]]  plt.figure(figsize=(10,5))  numhouse.boxplot(sym='r\*', grid=False)  plt.show() |
| --- |

Output:

|  |
| --- |

5. Histogram dari Price

| plt.figure(figsize=(15,5))  plt.subplot(121)  data\_nama['Price'].plot.hist(bins=10, title='Price')  plt.show() |
| --- |

Output:

|  |
| --- |

6. Scatter dari Price

| col\_list=['Price', 'SqFt']  numhouse = data\_nama[data\_nama.columns[data\_nama.columns.isin(col\_list)]]  numhouse.plot.scatter(x='SqFt', y='Price') |
| --- |

Output:

|  |
| --- |

7. Group Boxplot Berdasarkan Bedrooms

| col\_list=['Price', 'Bedrooms']  numhouse = data\_nama[data\_nama.columns[data\_nama.columns.isin(col\_list)]]  plt.figure(figsize=(15,5))  numhouse.boxplot(by='Bedrooms')  plt.show() |
| --- |

Output:

|  |
| --- |

* 1. Latihan Kedua – Tugas

1. Buatlah Scatter Plot Harga Rumah Berdasarkan Bedrooms!

Scritpt:

| col\_list=['Bedrooms', 'SqFt']  numhouse = data\_Adrian[data\_Adrian.columns[data\_Adrian.columns.isin(col\_list)]]  numhouse.plot.scatter(x='SqFt', y='Bedrooms') |
| --- |

Output:

|  |
| --- |

Penjelasan: Kode yang Anda berikan memiliki tujuan untuk membuat scatter plot (diagram titik) yang menunjukkan hubungan antara dua variabel dalam dataframe data\_Adrian, yaitu 'SqFt' (luas tanah) dan 'Bedrooms' (jumlah kamar tidur).

2. Buatlah Scatter Plot Harga Rumah berdasarkan Bathrooms!

Script:

| col\_list=['Bathrooms', 'SqFt']  numhouse = data\_Adrian[data\_Adrian.columns[data\_Adrian.columns.isin(col\_list)]]  numhouse.plot.scatter(x='SqFt', y='Bathrooms') |
| --- |

Output:

|  |
| --- |

Penjelasan: Kode yang Anda berikan memiliki tujuan untuk membuat sebuah scatter plot (diagram titik) yang menunjukkan hubungan antara dua variabel dalam dataframe data\_Adrian, yaitu 'SqFt' (luas tanah) dan 'Bathrooms' (jumlah kamar mandi).

3. Buatlah Grouped Boxplot berdasarkan Bathrooms dengan Price sebagai perbandingannya!

Script:

| #latihan 2 no 3  col\_list=['Price', 'Bathrooms']  numhouse = data\_Adrian[data\_Adrian.columns[data\_Adrian.columns.isin(col\_list)]]  plt.figure(figsize=(15,5))  numhouse.boxplot(by='Bathrooms')  plt.show() |
| --- |

Output:

|  |
| --- |

Penjelasan: Kode yang Anda berikan memiliki tujuan untuk membuat sebuah boxplot (diagram kotak) dari dua kolom dalam dataframe data\_Adrian, yaitu 'Bathrooms' (jumlah kamar mandi) dan 'Price' (harga).

1. **File Praktikum**

Github Repository:

|  |
| --- |

1. **Soal Latihan**

Soal:

1. Apa yang dimaksud Exploratory Data Analysis?
2. Mengapa EDA diperlukan melakukan dalam melakukan analisis data?

Jawaban:  
1. Exploratory Data Analysis (EDA) adalah sebuah metode analisis dalam ilmu data dan statistik yang digunakan untuk menggali dan memahami karakteristik data sebelum melakukan pembuatan model atau pengujian hipotesis lebih lanjut. Dengan melakukan EDA, kita dapat mengidentifikasi pola, relasi, dan sifat data secara visual dan deskriptif, sehingga dapat mengambil wawasan yang berharga dan mengarahkan langkah selanjutnya dalam analisis data secara lebih mendalam.  
2. Karena EDA membantu para analis data memahami data secara mendalam sebelum membangun model prediksi atau melakukan analisis yang lebih lanjut.

1. **Kesimpulan**
   1. Dalam pengerjaan praktikum Statistika, Kita belajar dan memahami Explorasi Data menggunakan Python.
   2. Kita juga dapat mengetahui cara Grouped Boxplot dan Scatter Plot
2. **Cek List (✔)**

| **No** | **Elemen Kompetensi** | **Penyelesaian** | |
| --- | --- | --- | --- |
| **Selesai** | **Tidak Selesai** |
| **1.** | Latihan Pertama | **…** |  |
| **2.** | Latihan Kedua | **…** |  |

1. **Formulir Umpan Balik**

| **No** | **Elemen Kompetensi** | **Waktu Pengerjaan** | **Kriteria** |
| --- | --- | --- | --- |
| **1.** | Latihan Pertama | … Menit | … |
| **2.** | Latihan Kedua | … Menit | … |

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang