


Nama: Francisco NIM: 064002300044	 Praktikum Probabilitas dan Statistika	MODUL 8 Nama Dosen: Dedy Sugiarto
Hari/Tanggal: Rabu, 15 Mei 2024		Nama Asisten Labratorium 1. Adzriel Yusak Noah (064.20.11) 2. Mohammad Akhmal Firdaus (064.20.34)

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.

2. Alat dan Bahan

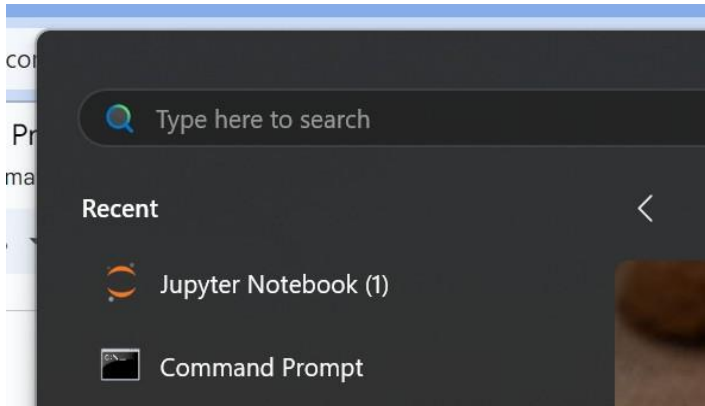
Hardware : Laptop/PC

Software : R Studio

3. Elemen Kompetensi

a. Latihan pertama – Praktikum

Buka Jupyter Notebook atau Google Colab di Browser



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:

```
import pandas as pd
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
```

Script

```
data_nama = pd.read_csv('C:/prakstatik/houseprice.csv')
data_nama.head(5)
```

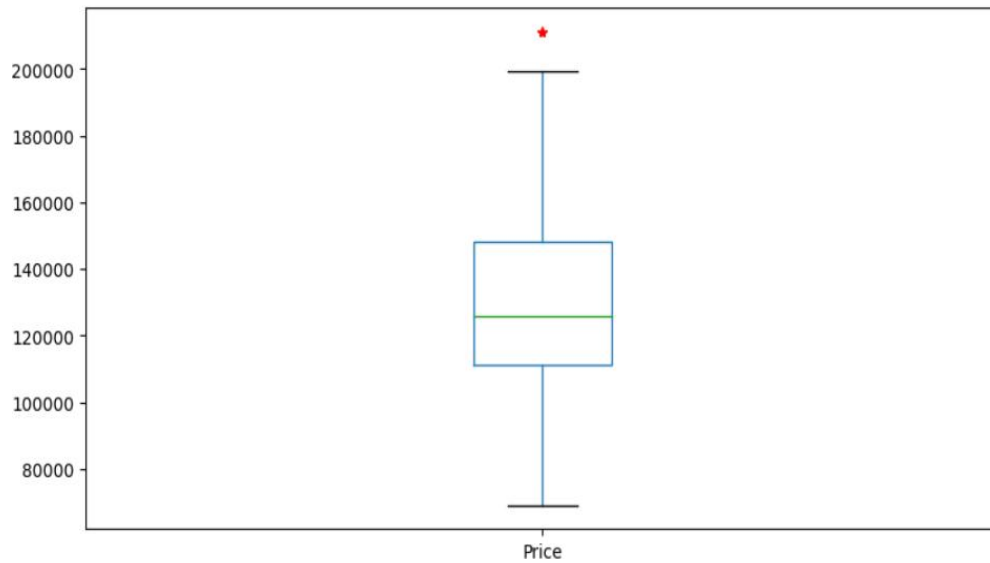
Output :

	Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood
0	114300	1790	2	2	2	No	East
1	114200	2030	4	2	3	No	East
2	114800	1740	3	2	1	No	East
3	94700	1980	3	2	3	No	East
4	119800	2130	3	3	3	No	East

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()
```

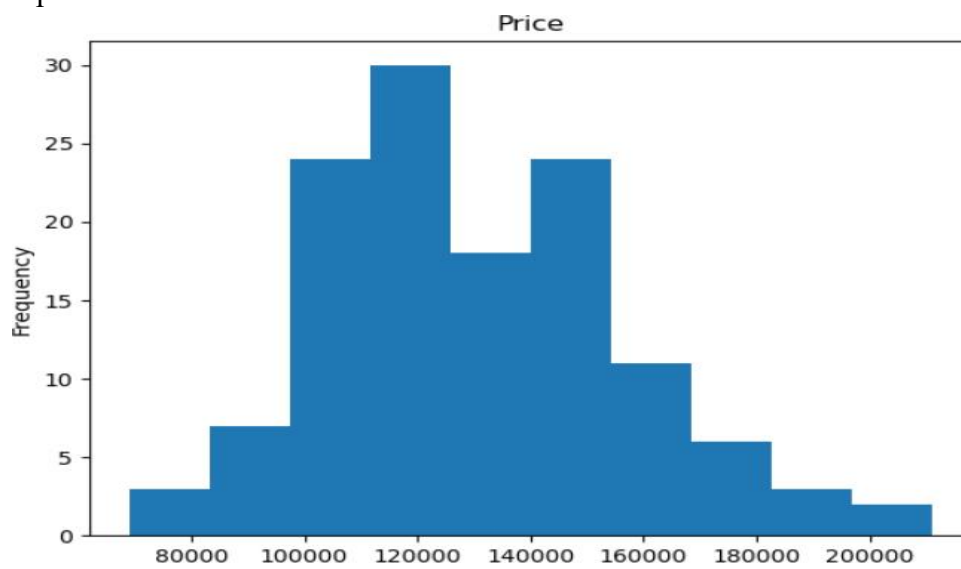
Output:



Histogram dari Price

```
plt.figure(figsize=(15,5))  
plt.subplot(121)  
data_nama['Price'].plot.hist(bins=10, title='Price')  
plt.show()
```

Output:

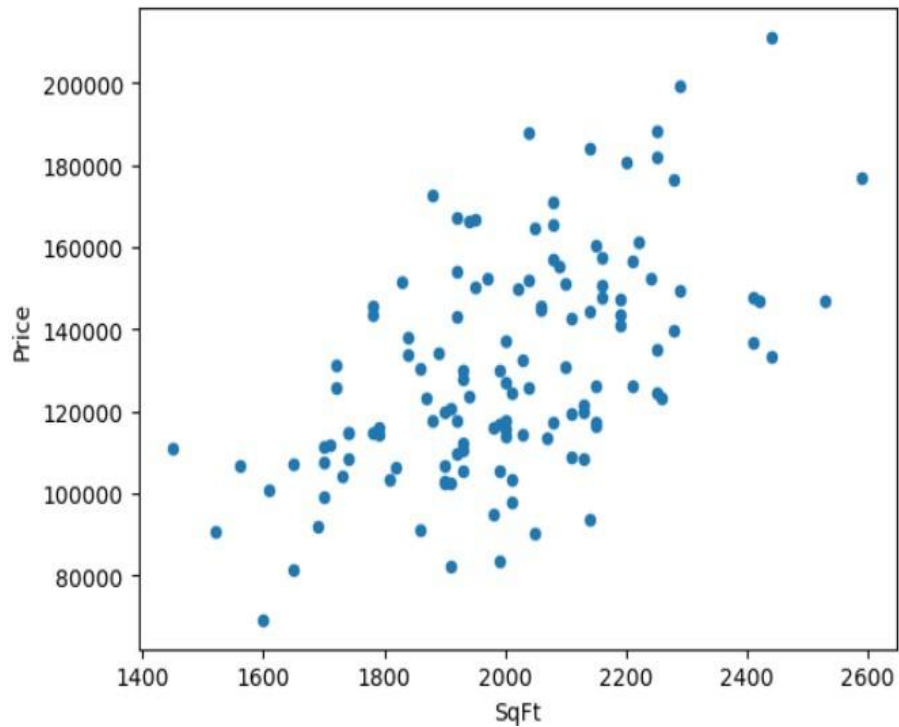


6. Scatter dari Price

```
col_list=['Price', 'SqFt']
numhouse = data_nama[data_nama.columns[data_nama.columns.isin(col_list)]]
numhouse.plot.scatter(x='SqFt', y='Price')AC
```

Output:

<Axes: xlabel='SqFt', ylabel='Price'>

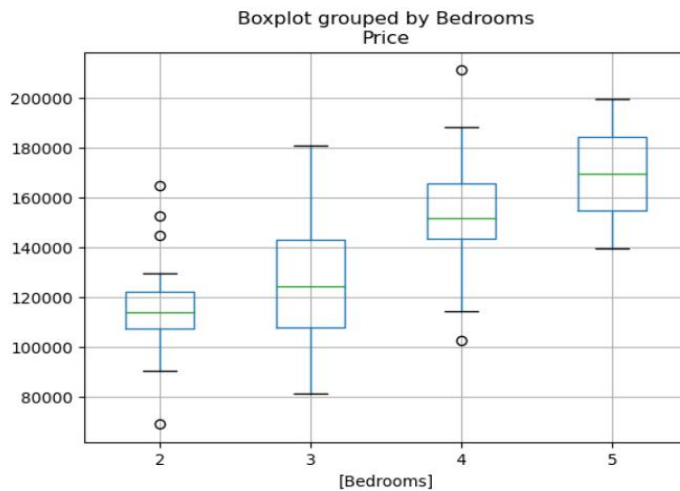


Group Boxplot Berdasarkan Bedrooms

```
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:

<Figure size 1500x500 with 0 Axes>



b. Latihan Kedua – Tugas

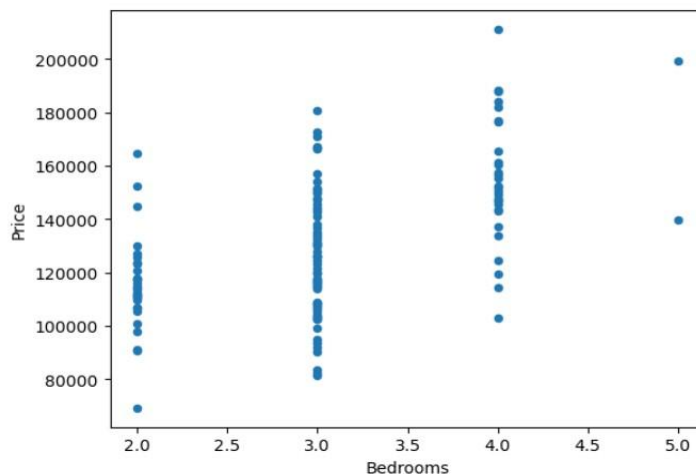
Buatlah Scatter Plot Harga Rumah Berdasarkan Bedrooms!

Script:

```
col_list = ['Bedrooms', 'Price']
numhouse = data_francisco[col_list]
numhouse.plot.scatter(x='Bedrooms', y='Price')
```

Output:

<Axes: xlabel='Bedrooms', ylabel='Price'>



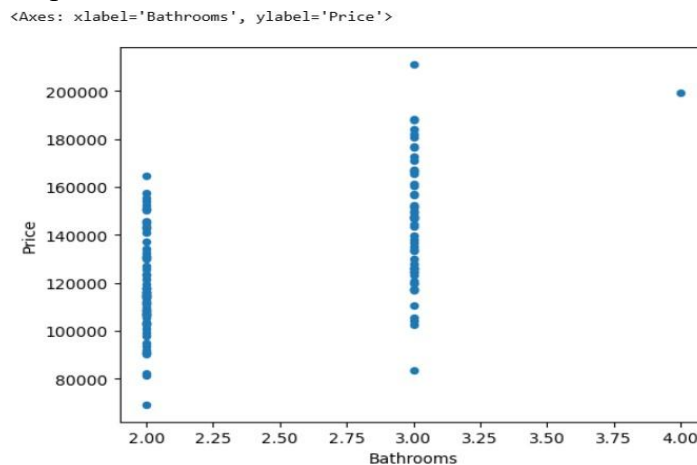
Penjelasan: dengan kita menggunakan rumus yang sudah ditunjukan dalam modul untuk collect list nya masukkan data price dengan bedrooms untuk memvisualisasikan serta menganalisis scatter plot berdasarkan harga dan bedrooms

Buatlah Scatter Plot Harga Rumah berdasarkan Bathrooms!

Script:

```
col_list = ['Bathrooms', 'Price']  
numhouse = data_ano[data_francisco.columns[data_francisco.columns.isin(col_list)]]  
numhouse.plot.scatter(x='Bathrooms', y='Price')
```

Output:



Penjelasan: sama halnya seperti sebelumnya hanya saja collect_listnya diubah dari bedrooms menjadi bathrooms tetapi untuk price tetap sama dimasukkan untuk menentukan scatter plot berdasarkan bathrooms dengan price.

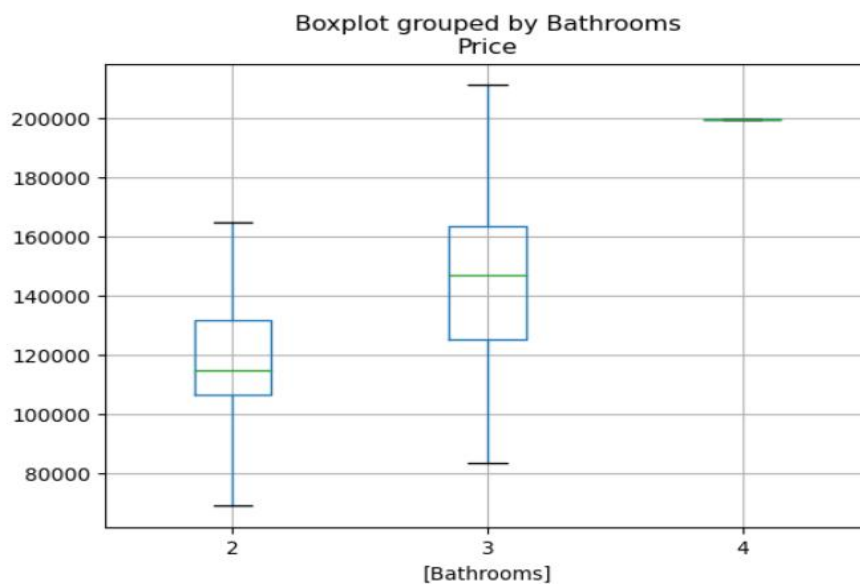
Buatlah Grouped Boxplot berdasarkan Bathrooms dengan Price sebagai perbandingannya!

Script:

```
col_list = ['Price', 'Bathrooms']  
numhouse = data_francisco[data_francisco.columns[data_francisco.columns.isin(col_list)]]  
plt.figure(figsize=(15, 5))  
numhouse.boxplot(by='Bathrooms')  
plt.show()
```

Output:

<Figure size 1500x500 with 0 Axes>



Penjelasan: untuk membuat analisis serta visualisasi boxplot grouped by bathrooms terlebih dahulu masukan data price dengan bathrooms untuk menentukannya lalu tentukan figsizenya sekiranya 15,5 maka program akan memvisualisasikan outputnya.

File Praktikum

Github Repository:

https://github.com/franafuk/Praktikum_stat

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) merupakan tahap awal yang penting dalam proses analisis data, karena memberikan pemahaman yang mendalam tentang data sebelum menerapkan model atau melakukan analisis yang lebih lanjut. juga sebuah pendekatan analisis data yang bertujuan untuk menganalisis dataset secara visual dan deskriptif untuk memahami karakteristik utama dari data tersebut.
2. EDA sangat penting karena memberikan pemahaman yang mendalam tentang data sebelum memulai analisis lebih lanjut. juga membantu dalam memahami struktur data, distribusi variabel, dan karakteristik utama lainnya.

Kesimpulan

- a. Dalam pengerjaan praktikum Statistika, tergolong mudah.
- b. Kita juga dapat mengetahui di modul ini bahwasannya kita dapat mengeksplorasi data analisis menggunakan python yang dimana proses tersebut penting dalam analisis data yang melibatkan pemahaman, visualisasi, dan pembersihan data menggunakan bahasa pemrograman Python. dengan mempelajari ini kita mampu memahami struktur dan karakteristik data dengan menggunakan berbagai teknik seperti statistik deskriptif dan visualisasi.

Cek List (✓)

No	Elemen Kompetensi	Penyelesaian	
		Selesai	Tidak Selesai
1.	Latihan Pertama	✓	
2.	Latihan Kedua	✓	

Formulir Umpan Balik

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Latihan Pertama	20 Menit	Menarik
2.	Latihan Kedua	7 Menit	Menarik

Keterangan:

3. Menarik
4. Baik
5. Cukup
6. Kurang