


Nama: Putra Mario Santoso NIM: 065002200037	 Praktikum Statistika	MODUL 8 Nama Dosen: Dedy Sugiarto
Hari/Tanggal: Senin, 7 Agustus 2023		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.

2. Alat dan Bahan

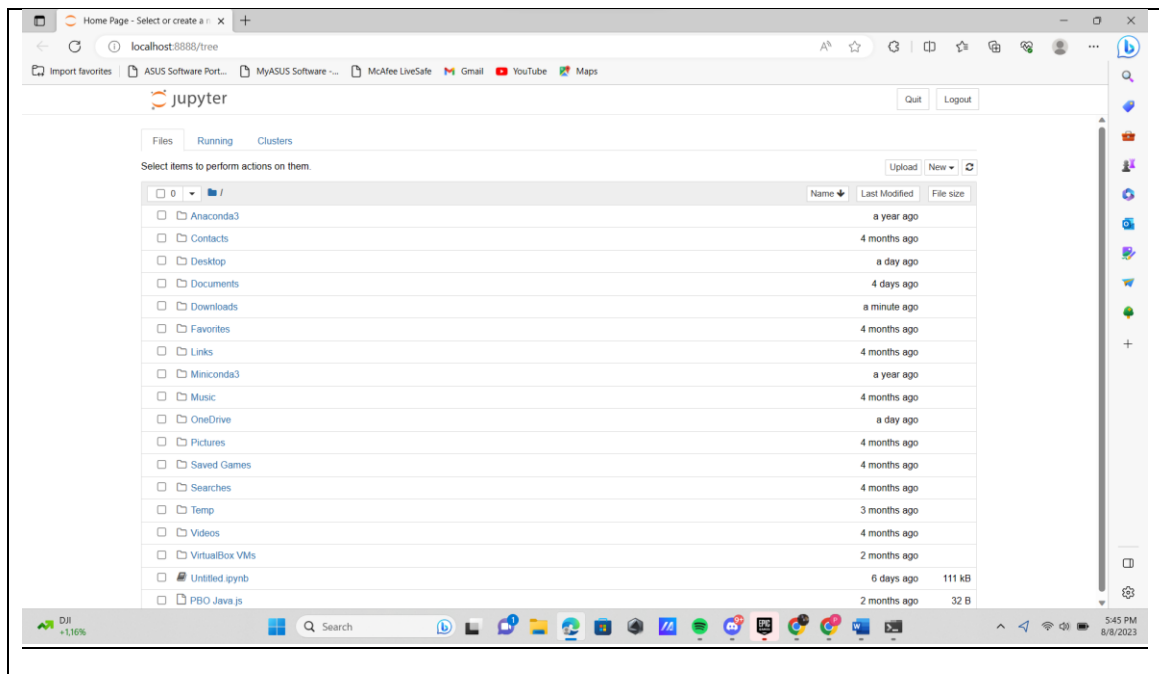
Hardware : Laptop/PC

Software : R Studio

3. Elemen Kompetensi

a. Latihan pertama – Praktikum

1. Buka Jupyter Notebook atau Google Colab di Browser



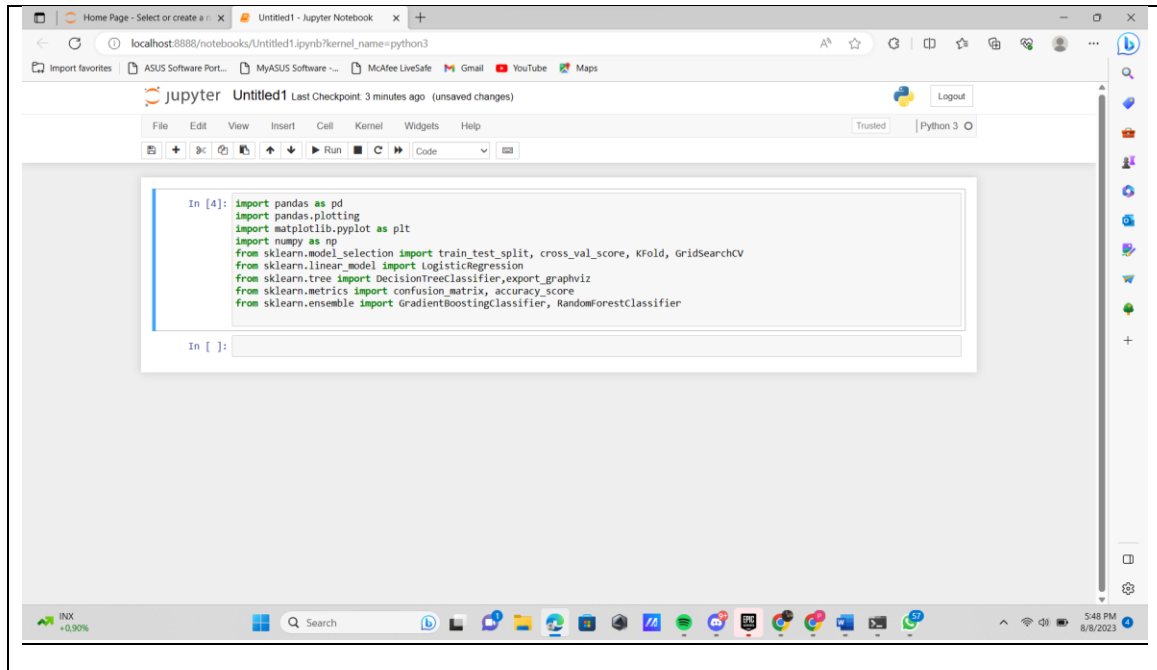
2. Lalu jalankan script berikut dan berikan output (gunakan nama variable data dengan nama masing-masing)

```
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:



The screenshot shows a Jupyter Notebook window titled 'Untitled1' running on a local host. The code cell contains the following imports:

```
In [4]: import pandas as pd
import pandas.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
```

The notebook interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help), a toolbar with icons for file operations and execution, and a sidebar with a file explorer and a search bar. The Windows taskbar at the bottom shows the system clock as 5:48 PM on 8/8/2023.

3. Script

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

Output:



```
In [6]: data_rio = pd.read_csv('C:/Users/rio santoso/Documents/Prak Statis/houseprice.csv')
data_rio.head(5)
```

Out[6]:

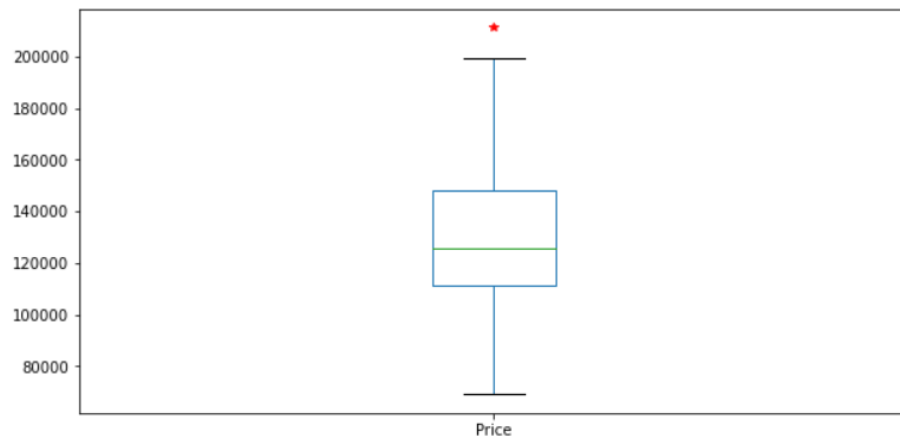
	Price	SqFt	Bedrooms	Bathrooms	Offers	Brick	Neighborhood
0	114300	1790	2	2	2	No	East
1	114200	2030	4	2	3	No	East
2	114800	1740	3	2	1	No	East
3	94700	1980	3	2	3	No	East
4	119800	2130	3	3	3	No	East

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:

```
In [8]: col_list=['Price']
numhouse = data_rio[data_rio.columns[data_rio.columns.isin(col_list)]]
plt.figure(figsize=(10,5))
numhouse.boxplot(sym='r*', grid=False)
plt.show()
```



5. Histogram dari Price

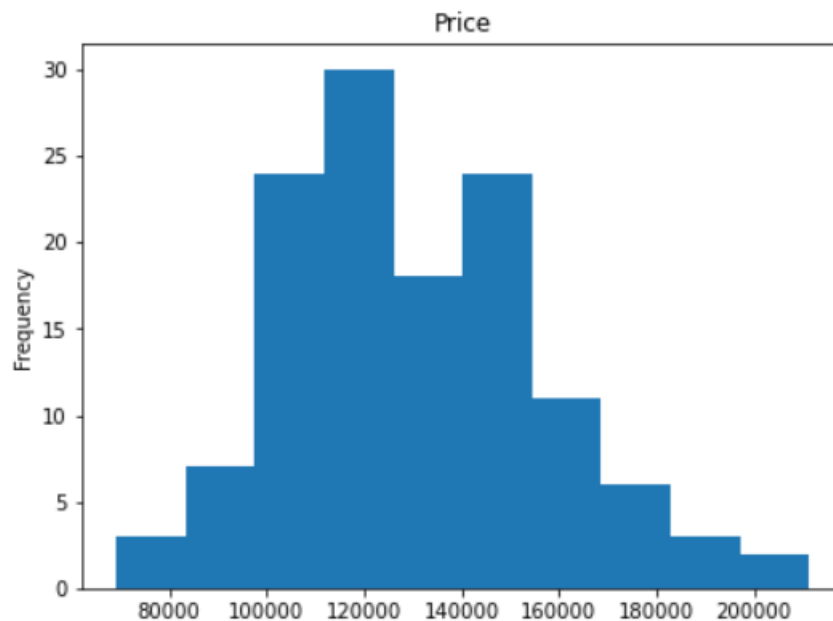
```
plt.figure(figsize=(15,5))
```



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

Output:

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



6. Scatter dari Price

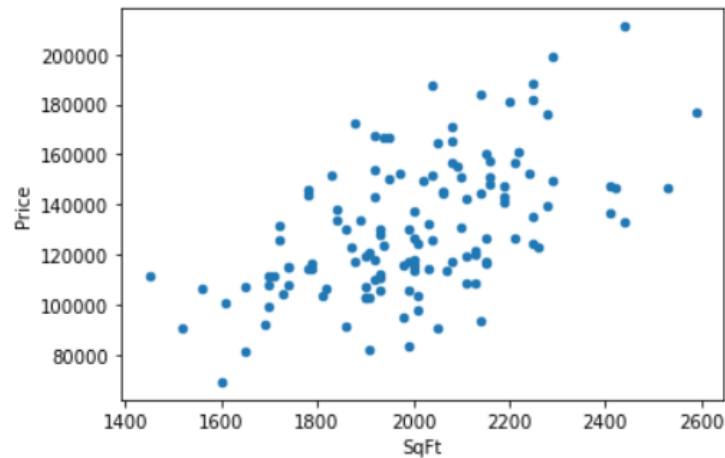
```
col_list=['Price', 'SqFt']
numhouse = data_nama[data_nama.columns[data_nama.columns.isin(col_list)]]
numhouse.plot.scatter(x='SqFt', y='Price')
```

Output:



```
In [10]: col_list=['Price', 'SqFt']  
numhouse = data_rio[data_rio.columns[data_rio.columns.isin(col_list)]]  
numhouse.plot.scatter(x='SqFt', y='Price')
```

Out[10]: <AxesSubplot:xlabel='SqFt', ylabel='Price'>

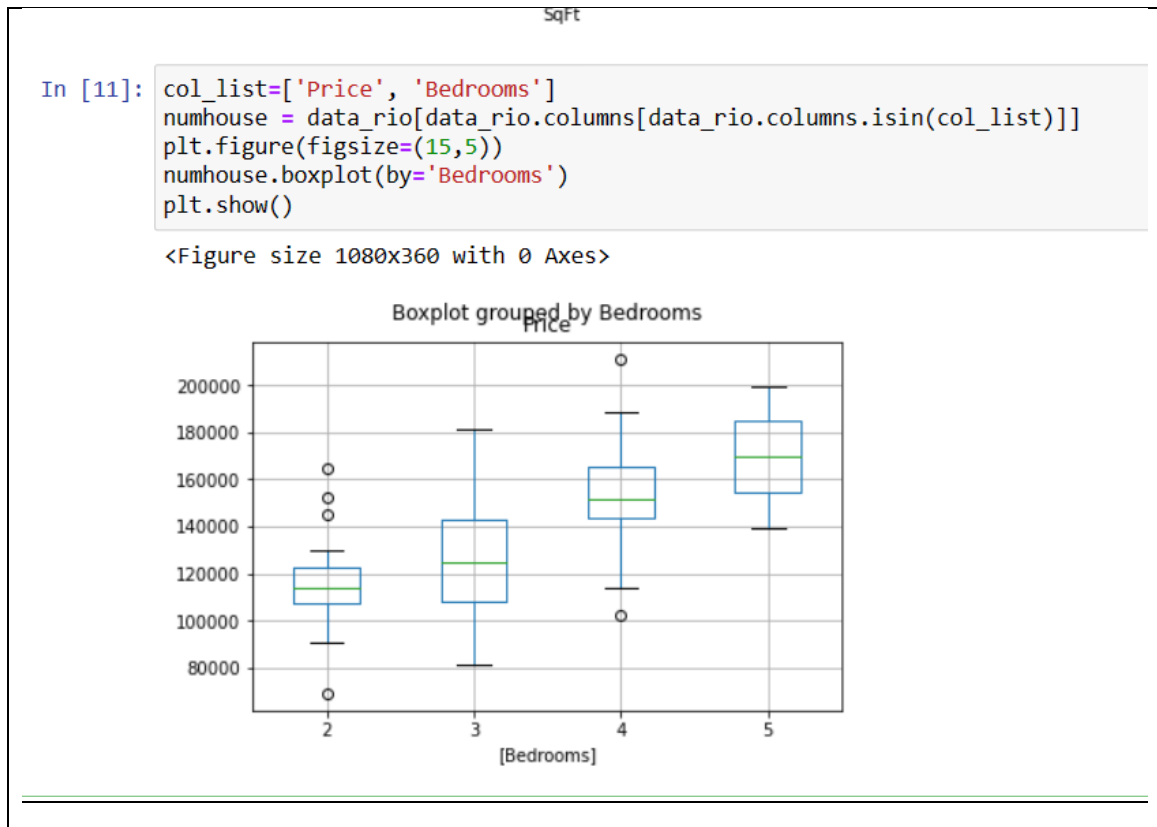


7. Group Boxplot Berdasarkan Bedrooms

```
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:





b. Latihan Kedua – Tugas

1. Buatlah Scatter Plot Harga Rumah Berdasarkan Bedrooms!

Script:

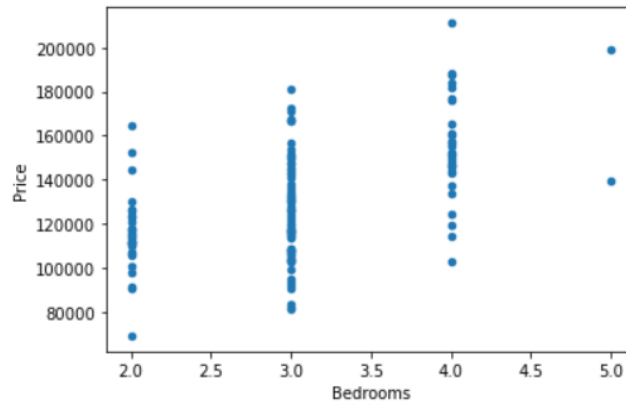
```
col_list=['Price', 'Bedrooms']
numhouse = data_rio[data_rio.columns[data_rio.columns.isin(col_list)]]
numhouse.plot.scatter(x='Bedrooms', y='Price')
```

Output:



```
In [12]: col_list=['Price', 'Bedrooms']  
numhouse = data_rio[data_rio.columns[data_rio.columns.isin(col_list)]]  
numhouse.plot.scatter(x='Bedrooms', y='Price')
```

Out[12]: <AxesSubplot:xlabel='Bedrooms', ylabel='Price'>



Penjelasan: berikut menampilkan harga rumah dengan berdasarkan colomn bedrooms dan menggunakan scatter plot

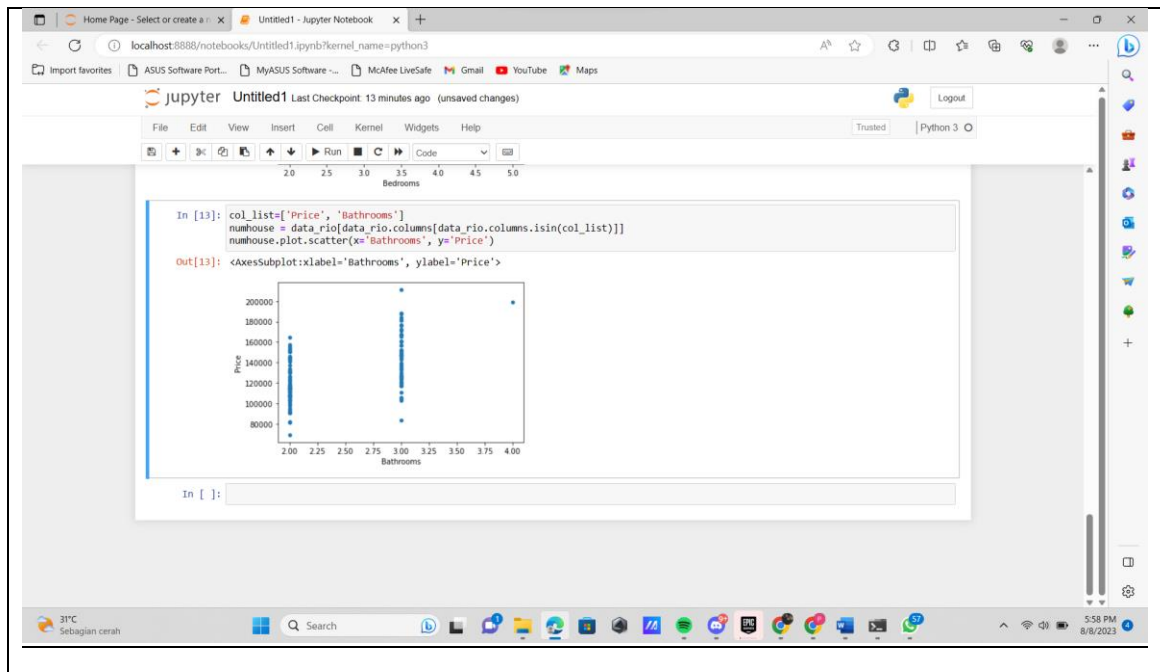
2. Buatlah Scatter Plot Harga Rumah berdasarkan Bathrooms!

Script:

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

Output:





Penjelasan: Diagram scatter plot ini menampilkan harga rumah berdasarkan pada kolom Bathrooms

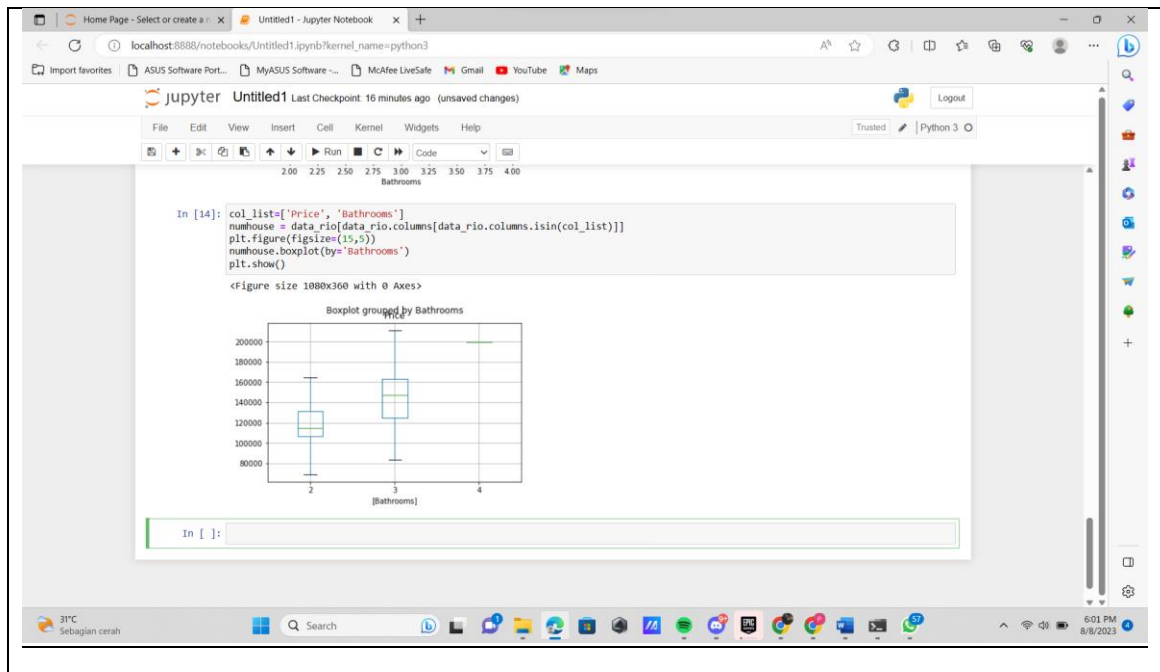
3. Buatlah Grouped Boxplot berdasarkan Bathrooms dengan Price sebagai perbandingannya!

Script:

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

Output:





Penjelasan: jadi jika dibuat menggunakan boxplot berdasarkan price dan perbandingannya menggunakan bathrooms akan menampilkan output seperti diatas

4. File Praktikum

Github Repository:

<https://github.com/yorioooo/Praktikum-Statistika>

5. 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 suatu pendekatan dalam analisis data yang bertujuan untuk memahami karakteristik, pola, dan hubungan dalam dataset secara visual dan deskriptif. EDA digunakan untuk menjelajahi data sebelum melakukan analisis yang lebih mendalam atau pemodelan statistik.
2. EDA adalah langkah kritis dalam proses analisis data yang membantu membangun pemahaman yang kuat tentang dataset, mengidentifikasi pola dan anomali, serta membimbing langkah-langkah analisis lebih lanjut.

6. Kesimpulan



- a. Dalam pengerjaan praktikum Statistika, kita dapat belajar bagaimana cara menginput file csv ke dalam aplikasi jupyter dan bagaimana menggunakan aplikasi tersebut
- b. Kita juga dapat mengetahui cara menggunakan boxplot, scatter plot dan diagram batang dalam mencari suatu masalah, misalnya membuat grouped boxplot berdasarkan suatu kolom dengan perbandingan kolom yang lain

7. Cek List (□)

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

8. Formulir Umpan Balik

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Latihan Pertama	15 Menit	Baik
2.	Latihan Kedua	15 Menit	Baik

Keterangan:

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

