


<p><b>Nama:</b> <b>Vania Rahma Dewi</b></p> <p><b>NIM:</b> <b>064002200030</b></p>	 <p><b>Praktikum Statistika</b></p>	<p><b>MODUL 8</b></p> <p><b>Nama Dosen:</b> <b>Dedy Sugiarto</b></p>
<p><b>Hari/Tanggal:</b> <b>Senin, 7 Agustus 2023</b></p>		<p><b>Nama Asisten Labratorium</b></p> <p><b>1. Elen Fadilla Estri</b> <b>064002000008</b></p> <p><b>2. Rukhy Zaifa Aduhalim</b> <b>064002000041</b></p>

## 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

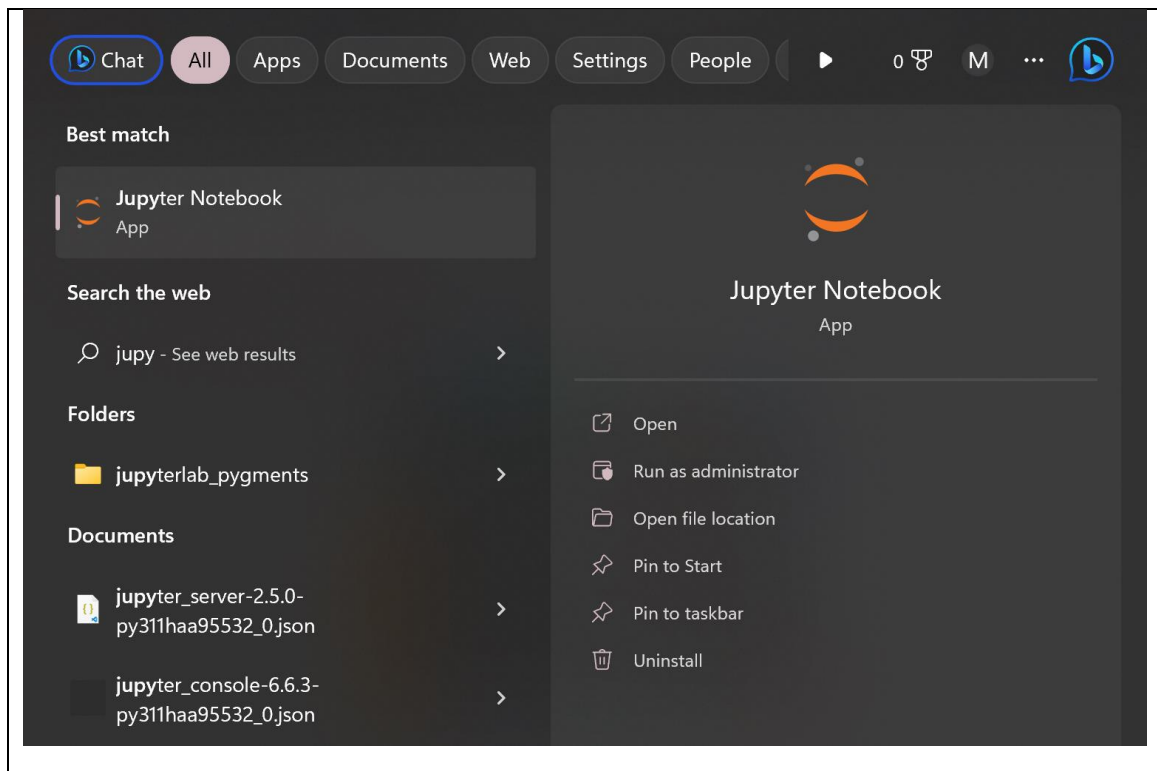
Hardware : Laptop/PC

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:

```
In [ ]: 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,
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, RandomForestC
```

### 3. Script

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

Output:

```
In [4]: data_vania = pd.read_csv('C:/Users/mulya/OneDrive/Documents/Kuliah (1)')
data_vania.head(5)
```

Out[4]:

	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

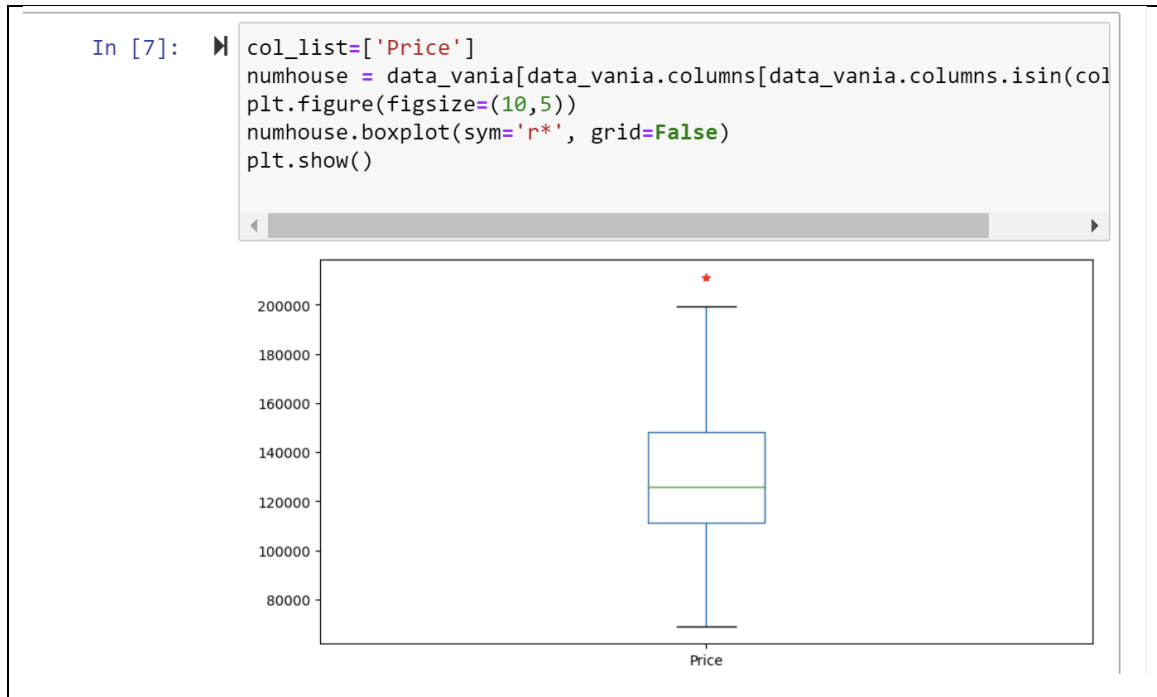
### 4. Boxplot dari harga rumah/Price

```
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:



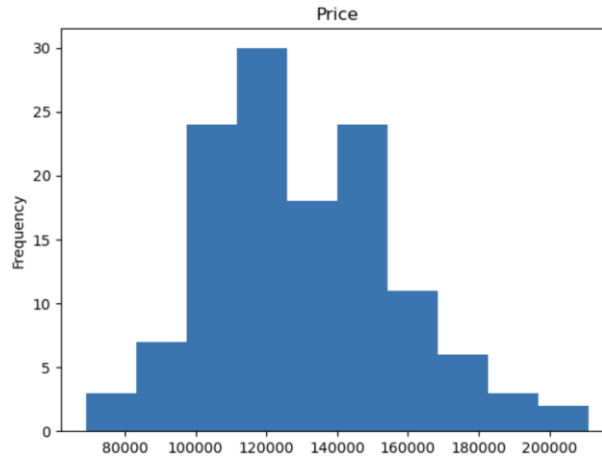
## 5. Histogram dari Price

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

Output:



```
In [8]: #hist dari price
plt.figure(figsize=(15,5))
plt.subplot(121)
data_vania['Price'].plot.hist(bins=10, title='Price')
plt.show()
```



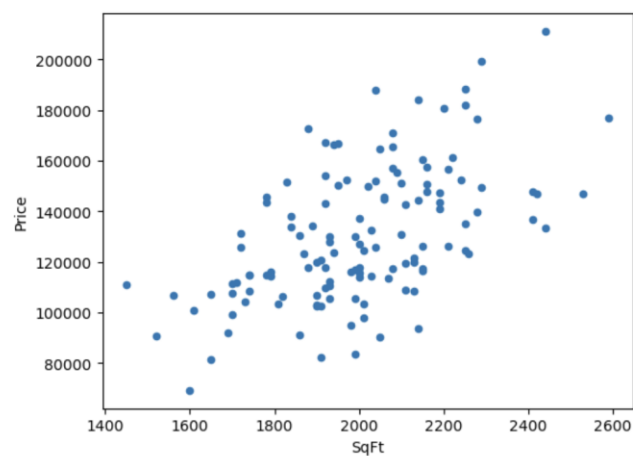
## 6. Scatter dari Price

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

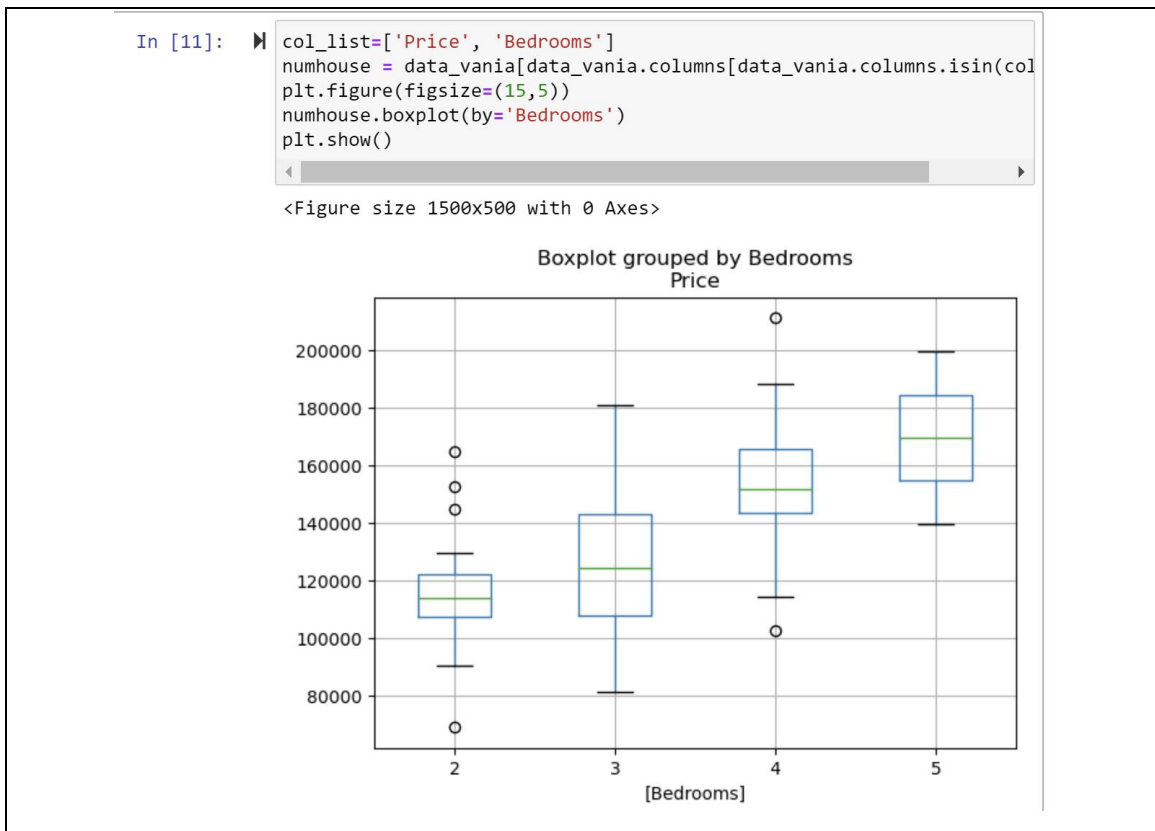
Out[10]: <Axes: xlabel='SqFt', ylabel='Price'>
```



## 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

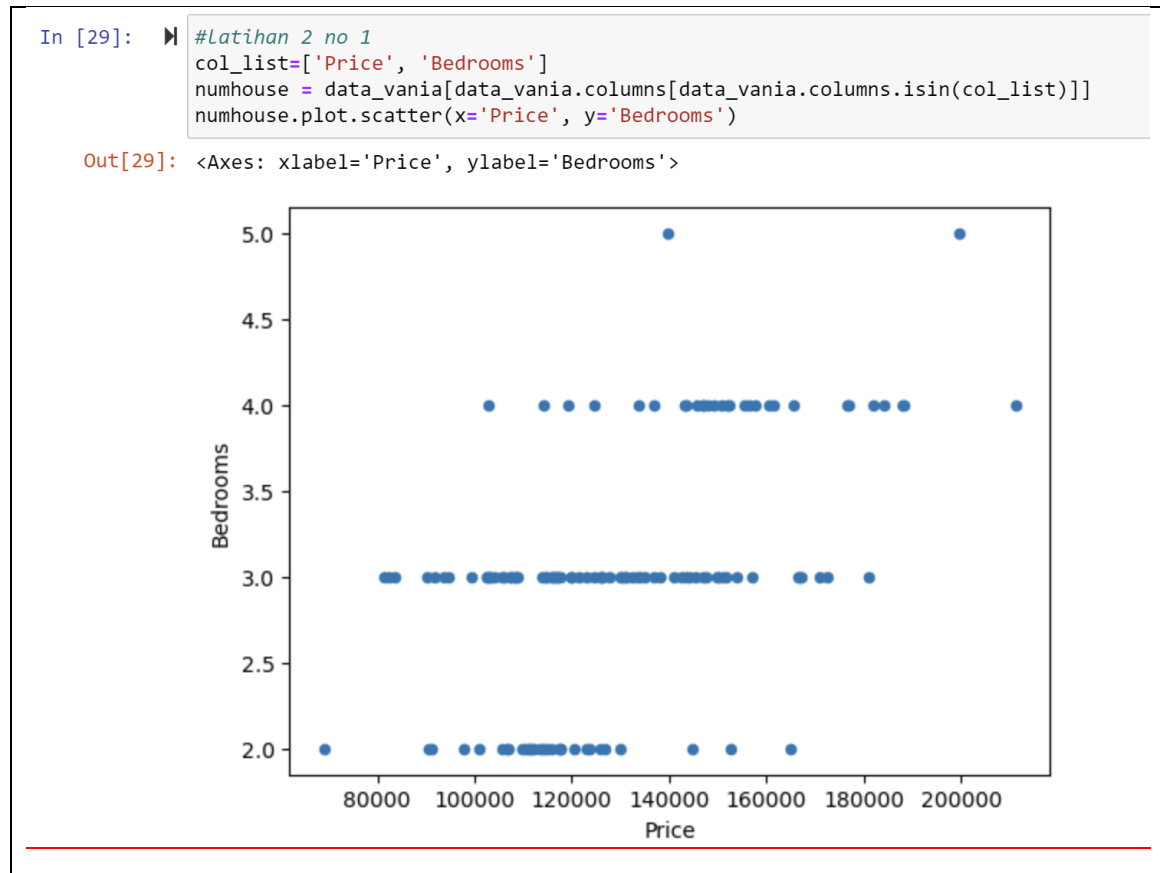
#### 1. Buatlah Scatter Plot Harga Rumah Berdasarkan Bedrooms!

Script:

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



Output:



Penjelasan: Diagram di atas merupakan Scatter Plot harga rumah/Price berdasarkan jumlah kamar tidurnya (Bedrooms). Area price berada di sumbu Y dan Bedrooms berada di sumbu X.

2. Buatlah Scatter Plot Harga Rumah berdasarkan Bathrooms!

Script:

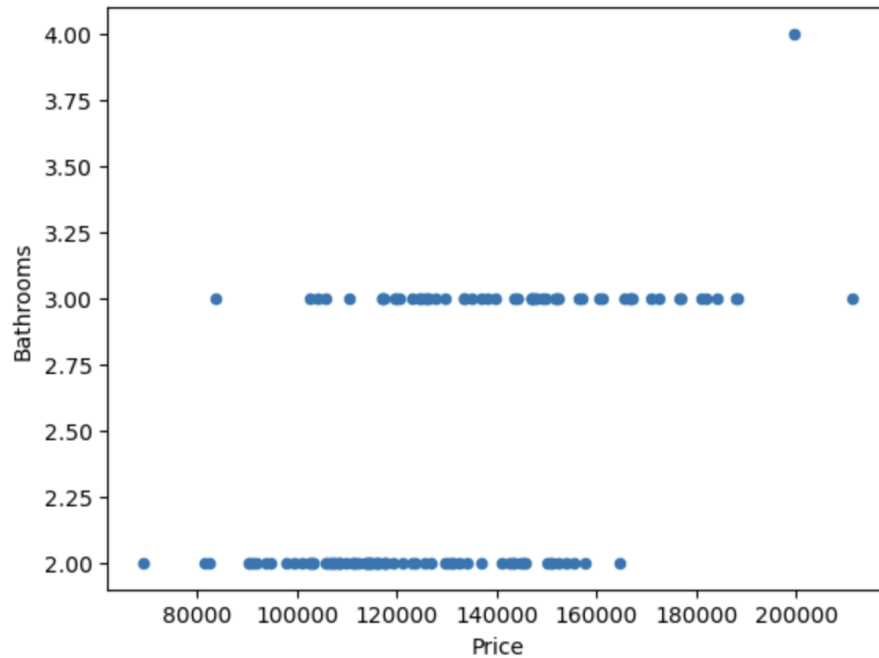
```
col_list=['Price', 'Bathrooms']
numhouse = data_vania[data_vania.columns[data_vania.columns.isin(col_list)]]
numhouse.plot.scatter(x='Bathrooms', y='Price')
```

Output:



```
In [30]: #Latihan 2 no 2
col_list=['Price', 'Bathrooms']
numhouse = data_vania[data_vania.columns[data_vania.columns.isin(col_list)]]
numhouse.plot.scatter(x='Price', y='Bathrooms')
```

Out[30]: <Axes: xlabel='Price', ylabel='Bathrooms'>



Penjelasan: Gambar di atas merupakan Scatter Plot harga rumah/Price berdasarkan jumlah kamar mandinya (Bathrooms). Price berada di sumbu Y sedangkan Bathrooms berada di sumbu X.

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

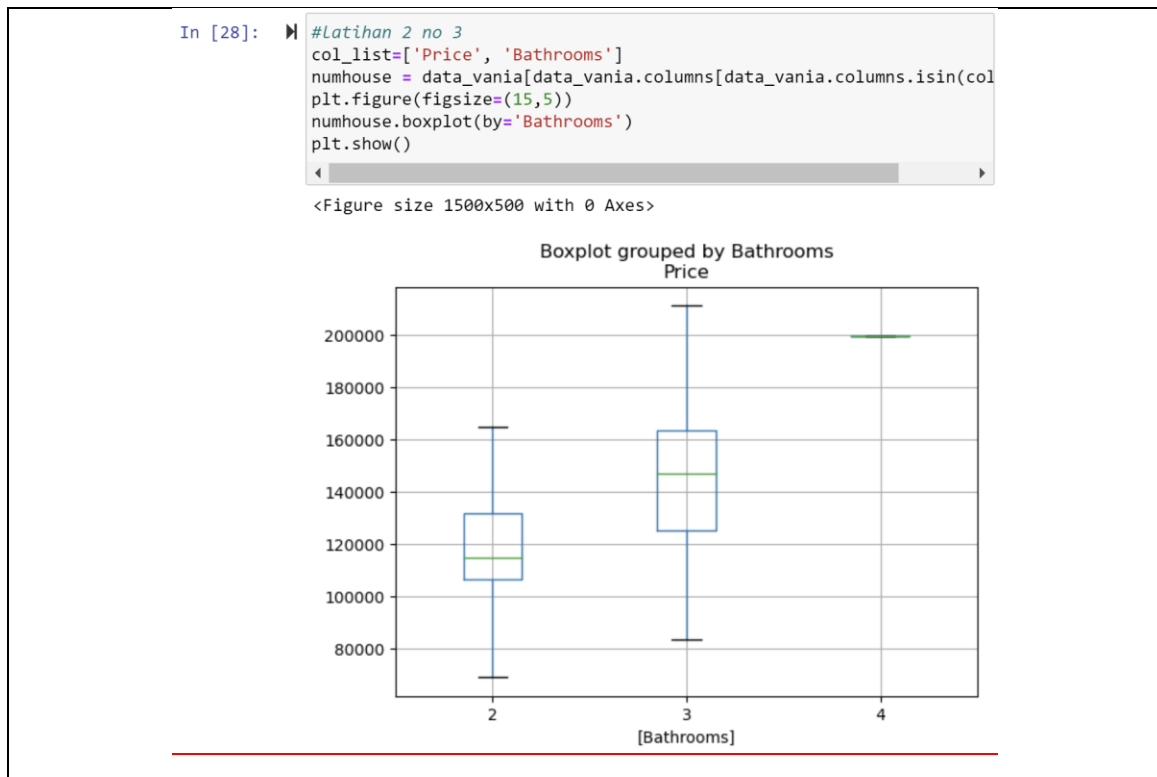
Script:

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

Output:







Penjelasan: Gambar di atas merupakan Grouped Boxplot dengan membandingkan harga rumah (price) dengan jumlah kamar mandi (bathrooms).

#### 4. File Praktikum

Github Repository:

#### 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 proses analisis dalam ilmu data dan statistic yang bertujuan untuk memahami dan menganalisis karakteristik dari data yang ada sebelum melakukan model atau hipotesis yang lebih lanjut.
2. Karena EDA membantu para analis data memahami data secara mendalam sebelum membangun model prediksi atau melakukan analisis yang lebih lanjut.



## 6. Kesimpulan

- Dalam pengerjaan praktikum Statistika, kita dapat memahami bagaimana cara melakukan eksplorasi data dengan menggunakan python pada jupyter notebook.
- Kita juga dapat mengetahui bahwa EDA (Exploratory Data Analysis) bertujuan untuk bertujuan untuk memahami dan menganalisis karakteristik dari data yang ada sebelum melakukan model atau hipotesis yang lebih lanjut.

## 7. Cek List (✓)

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

## 8. Formulir Umpan Balik

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Latihan Pertama	20 Menit	Menarik
2.	Latihan Kedua	15 Menit	Menarik

Keterangan:

- Menarik
- Baik
- Cukup
- Kurang