


| | | |
|--------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Nama: Aldi Surya Pranata</p> <p>NIM: 064002200039</p> |  <p>Praktikum Statistika</p> | <p>MODUL 8</p> <p>Nama Dosen: Dedy Sugiarto</p> |
| <p>Hari/Tanggal: Senin, 7 Agustus 2023</p> | | <p>Nama Asisten Labratorium</p> <p>1. Elen Fadilla Estri 064002000008</p> <p>2. Rukhy Zaifa Aduhalim 064002000041</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

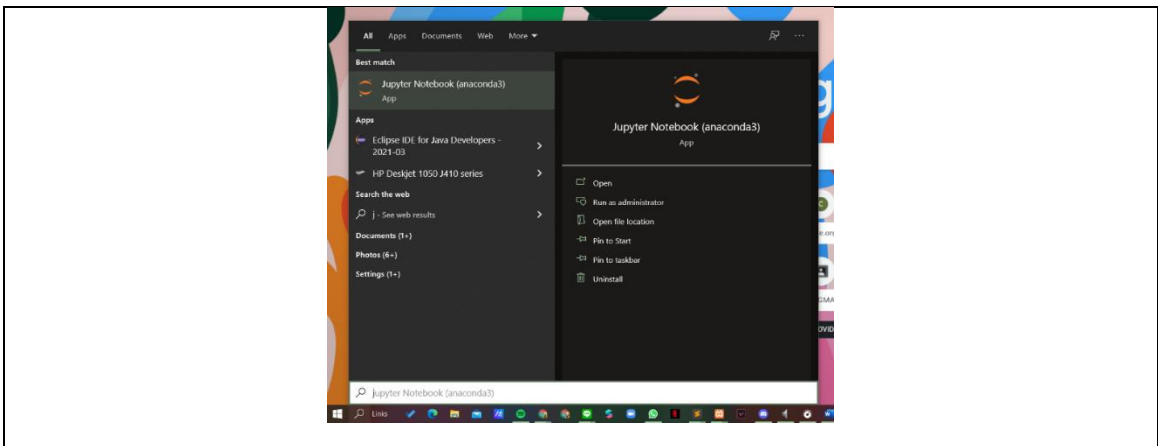
Hardware : Laptop/PC

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:

```
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
```

3. Script

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

Output:

```
data_aldi = pd.read_csv('C:/Users/ALDI PRANATA/Documents/prak1 stas/houseprice.csv', sep=";")
data_aldi.head(5)
```

```
[15]:
```

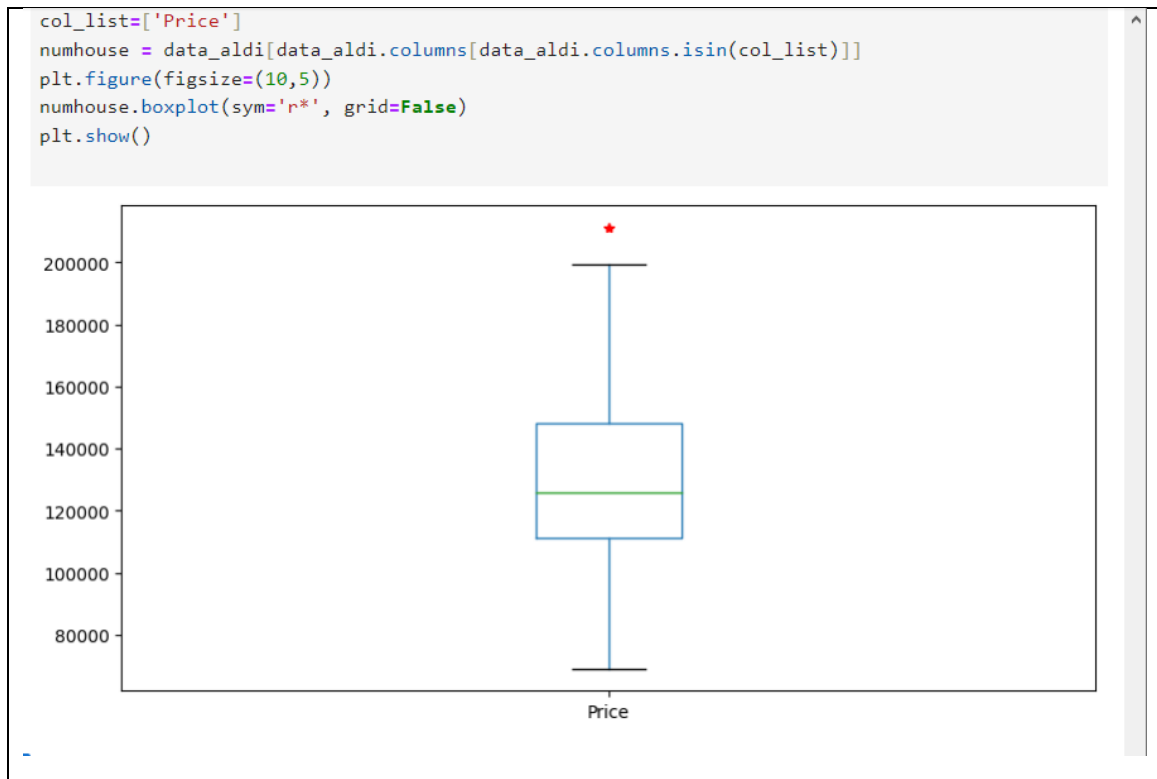
| | Price | SqFt | Bedrooms | Bathrooms | Offers | Brick | Neighborhood |
|---|--------|------|----------|-----------|--------|-------|--------------|
| 0 | 114300 | 1790 | 2 | 2 | 2 | No | East |
| 1 | 114200 | 2030 | 4 | 2 | 3 | No | East |
| 2 | 114800 | 1740 | 3 | 2 | 1 | No | East |
| 3 | 94700 | 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:



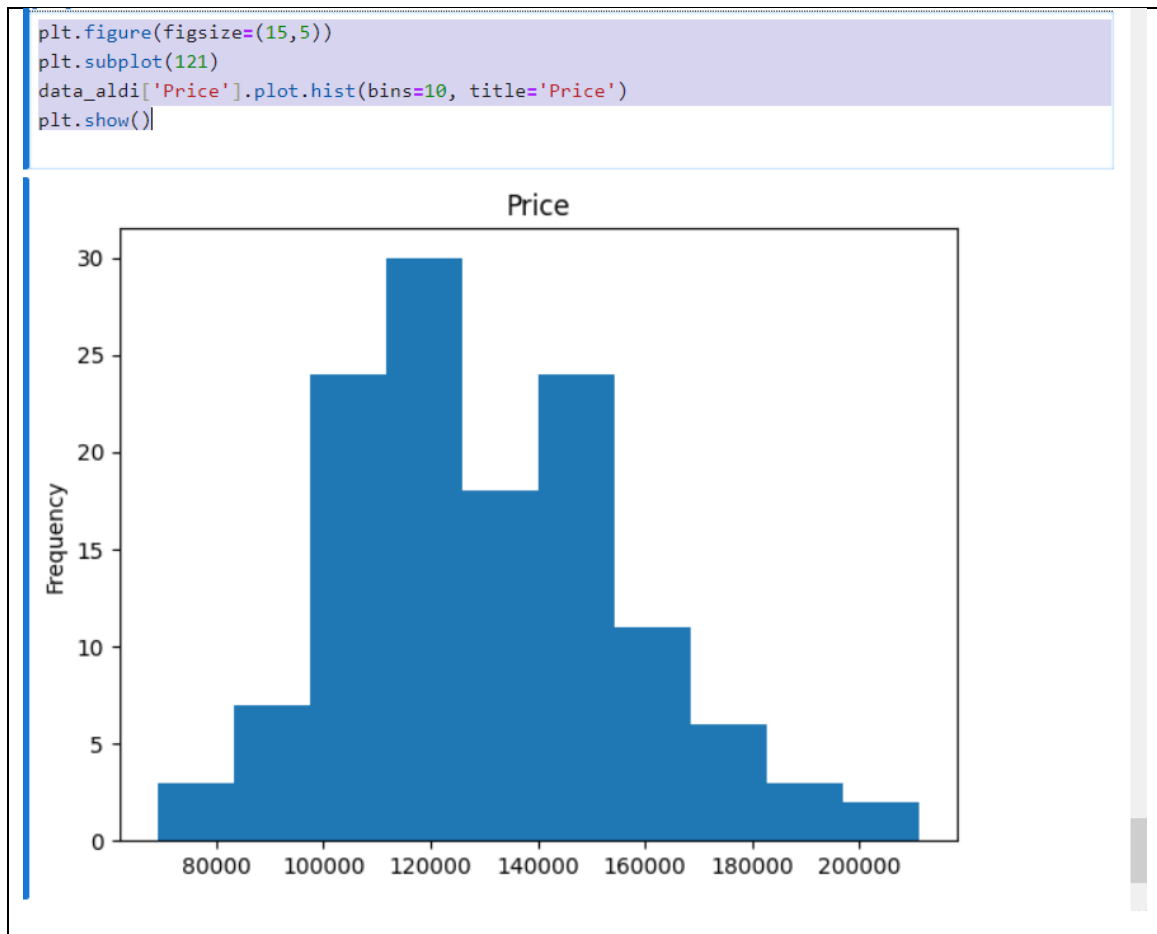


5. Histogram dari Price

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

Output:



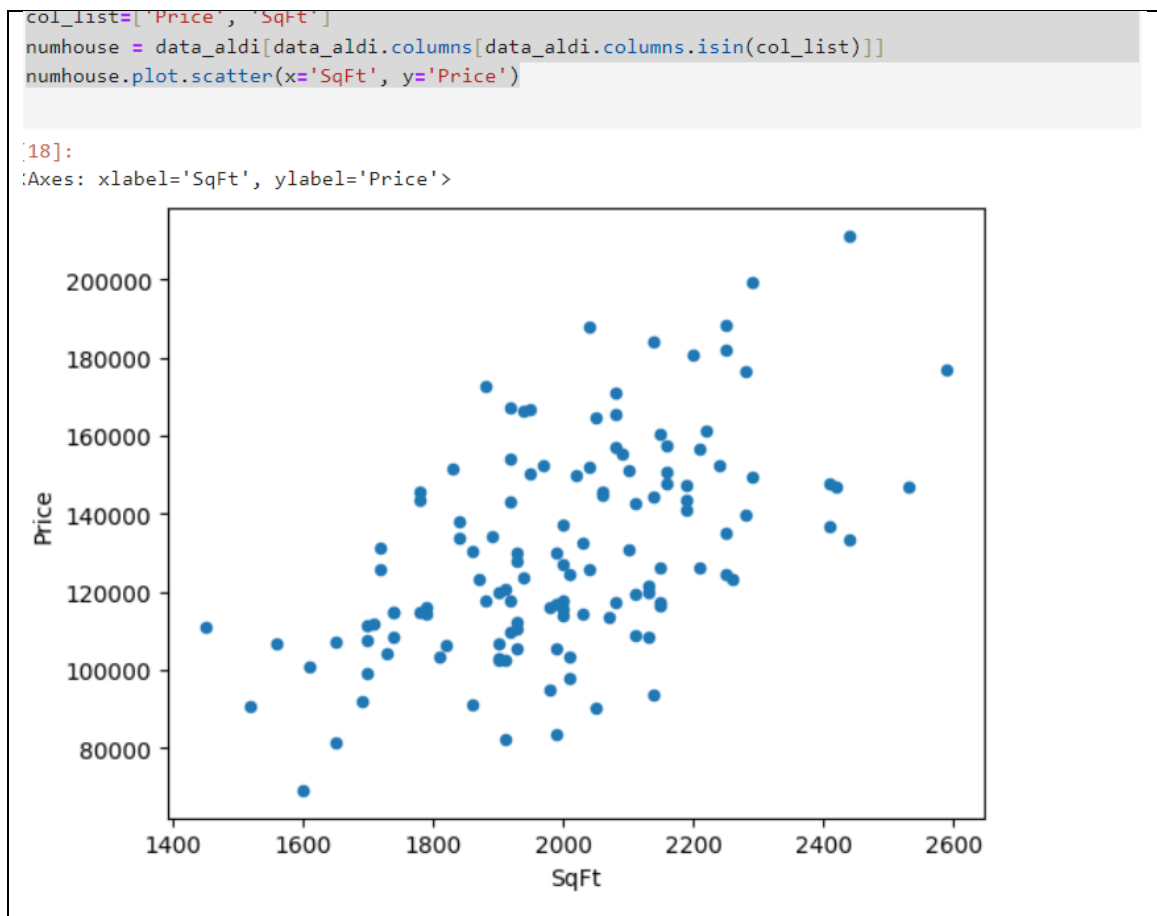


6. Scatter dari Price

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

Output:



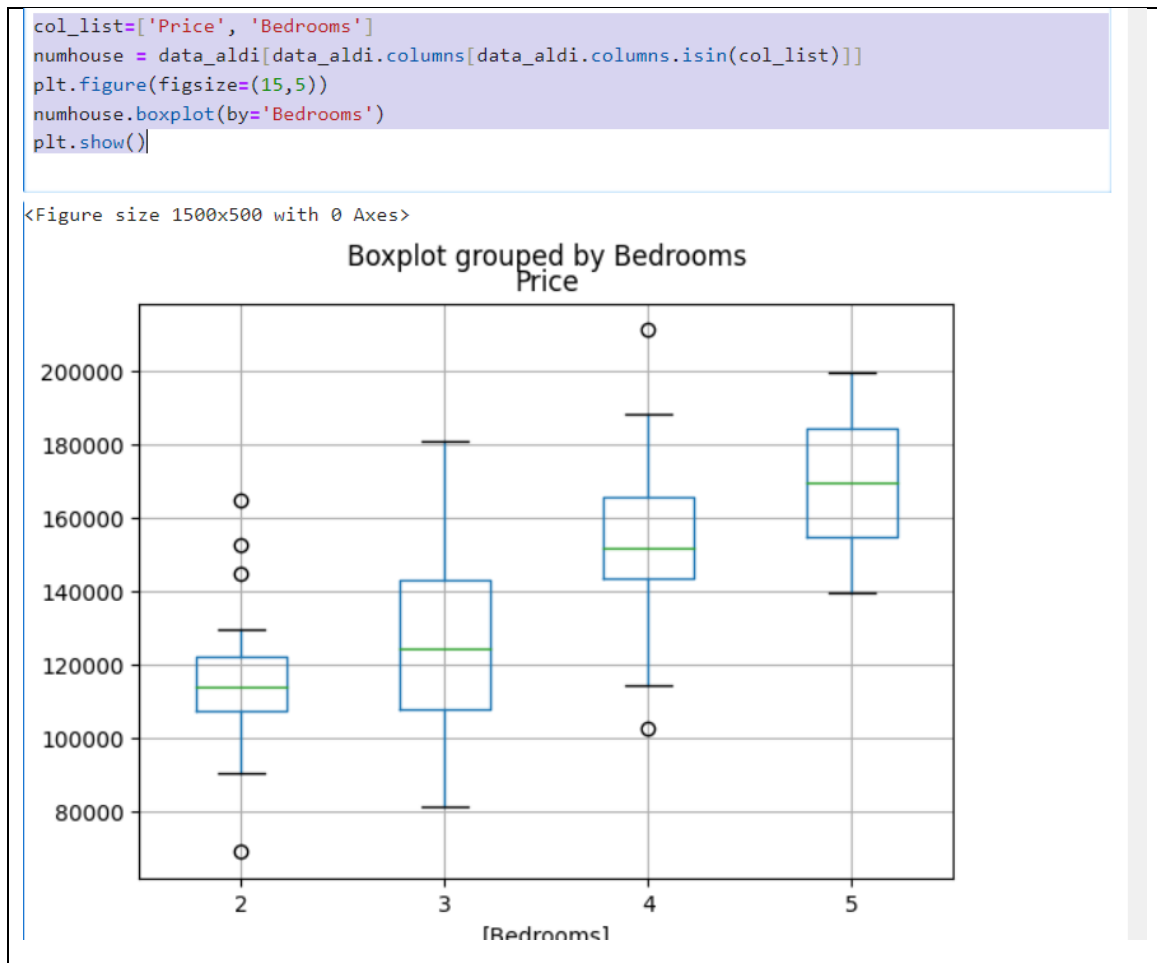


7. Group Boxplot Berdasarkan Bedrooms

```
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:





b. Latihan Kedua – Tugas

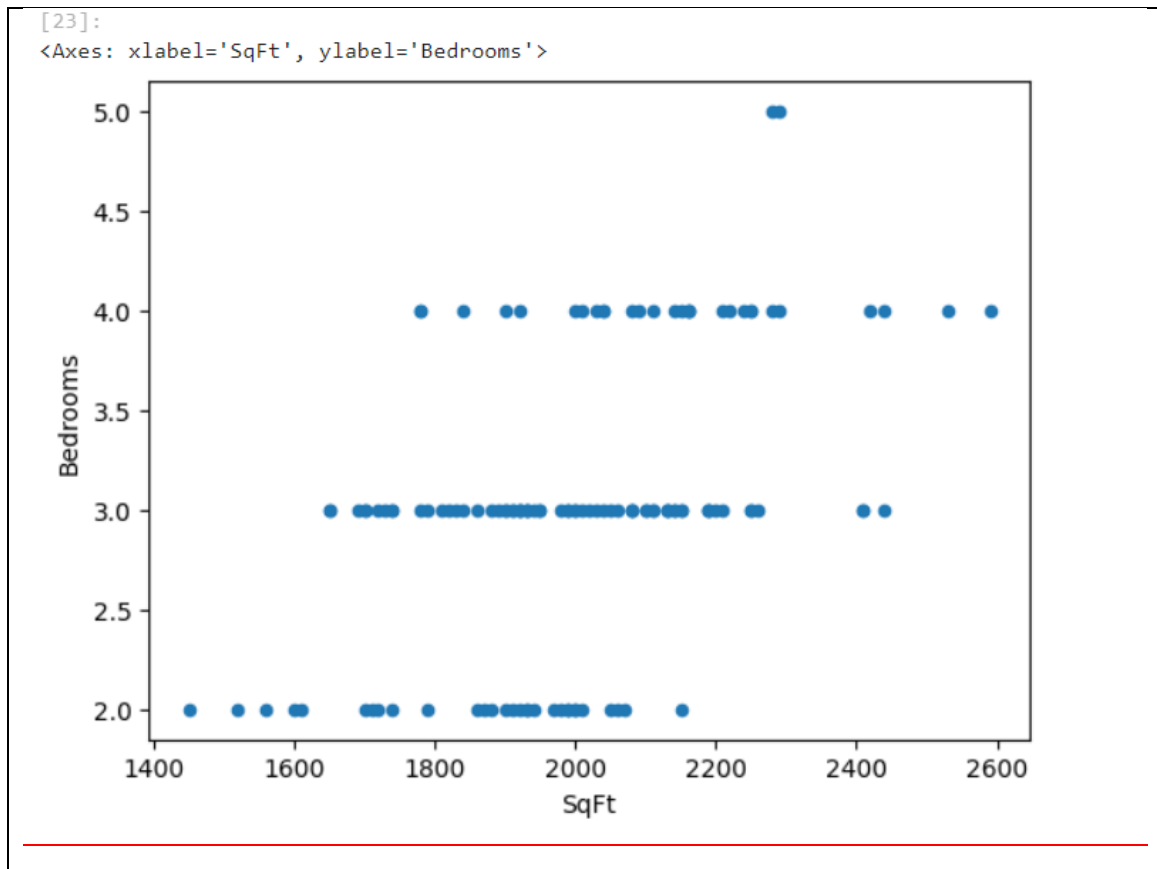
1. Buatlah Scatter Plot Harga Rumah Berdasarkan Bedrooms!

Script:

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

Output:





Penjelasan: membuat scatter plot yang menggambarkan hubungan antara ukuran rumah ('SqFt') dengan jumlah kamar tidur ('Bedrooms') dari dataset 'data_aldi'.

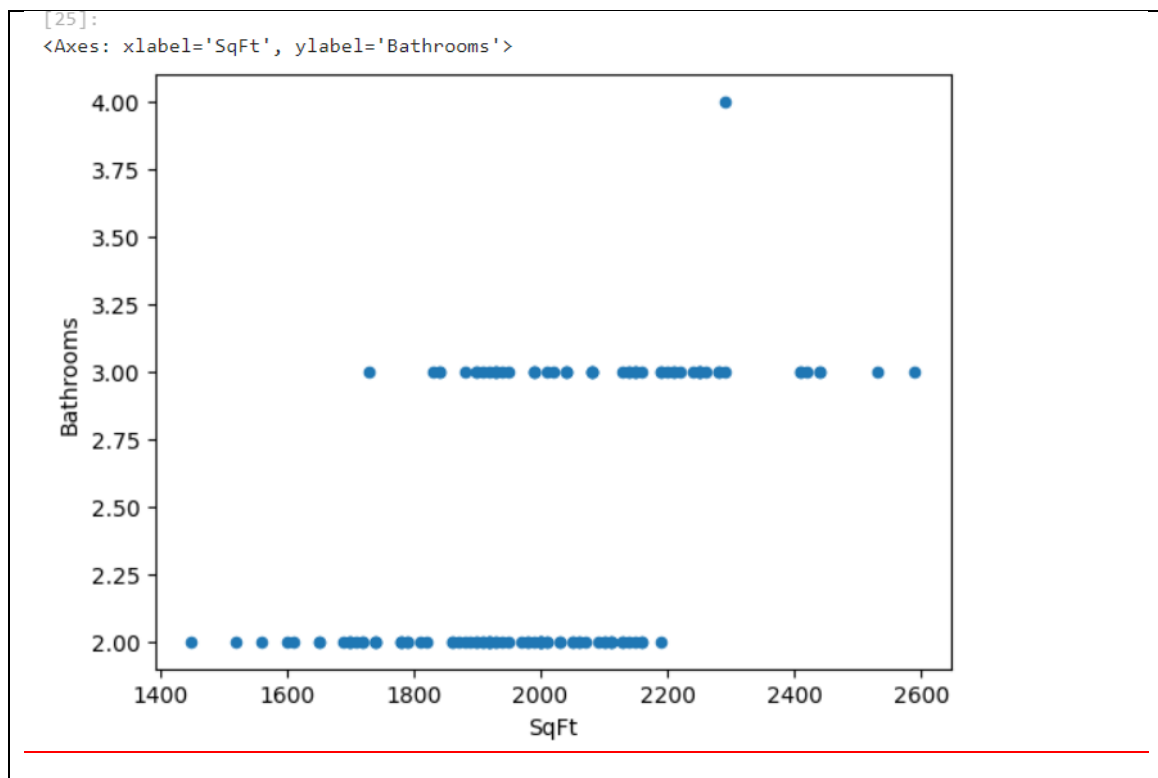
2. Buatlah Scatter Plot Harga Rumah berdasarkan Bathrooms!

Script:

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

Output:





Penjelasan: membuat scatter plot yang menggambarkan hubungan antara ukuran rumah ('SqFt') dengan jumlah kamar tidur ('Bathrooms') dari dataset 'data_aldi'.

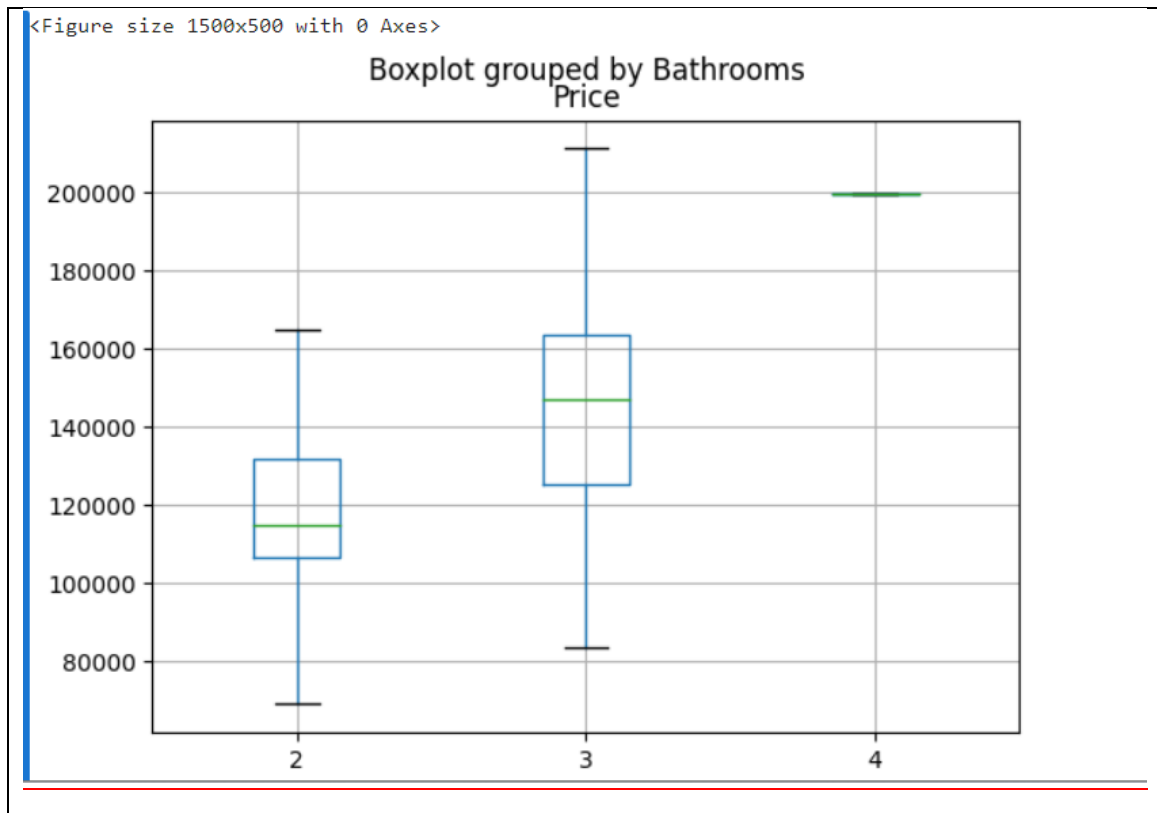
3. Buatlah Grouped Boxplot berdasarkan Bathrooms dengan Price sebagai perbandingannya!

Script:

```
[26]:
col_list=['Price', 'Bathrooms']
numhouse = data_aldi[data_aldi.columns[data_aldi.columns.isin(col_list)]]
plt.figure(figsize=(15,5))
numhouse.boxplot(by='Bathrooms')
plt.show()
```

Output:





Penjelasan: membuat sebuah boxplot dari data yang terkait dengan harga ('Price') dan jumlah kamar mandi ('Bathrooms') dari suatu dataset yang disebut 'data_aldi'.

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 adalah proses analisis awal yang dilakukan pada suatu set data untuk menggali wawasan, pola, hubungan, dan informasi penting dari data tersebut. Tujuannya adalah untuk memahami karakteristik data, mengidentifikasi outlier atau nilai-nilai ekstrem, mengeksplorasi distribusi, serta menemukan pola yang mungkin tidak terlihat secara langsung. EDA melibatkan penggunaan berbagai teknik visualisasi dan statistik deskriptif untuk mengenali tren, anomali, atau insight yang dapat membantu dalam pengambilan keputusan lebih lanjut.



2. Exploratory Data Analysis (EDA) diperlukan dalam melakukan analisis data karena memiliki beberapa manfaat yang penting:

1. **Memahami Data:** EDA membantu Anda untuk mengenal dan memahami data yang Anda miliki. Ini termasuk mengidentifikasi jenis data, mengenali tipe variabel (numerik atau kategorikal), dan memahami struktur dataset. Pemahaman ini menjadi dasar yang penting sebelum melakukan analisis lebih lanjut.
2. **Mengidentifikasi Anomali dan Kesalahan:** EDA memungkinkan Anda untuk mengidentifikasi adanya data yang tidak wajar atau nilai-nilai ekstrem (outliers). Jika ada kesalahan dalam data, seperti data yang hilang atau tidak valid, EDA membantu Anda menemukannya.
3. **Menemukan Pola dan Hubungan:** Melalui visualisasi dan perhitungan statistik dalam EDA, Anda dapat menemukan pola, korelasi, dan hubungan antara variabel. Ini bisa mengarah pada wawasan baru atau penemuan penting yang mungkin tidak terlihat secara kasual.
4. **Memilih Pendekatan Analisis yang Tepat:** Dengan memahami karakteristik data melalui EDA, Anda dapat memilih pendekatan analisis yang sesuai. Ini termasuk pemilihan metode statistik atau algoritma pembelajaran mesin yang cocok untuk masalah yang Anda hadapi.
5. **Pengambilan Keputusan yang Lebih Informatif:** EDA memberikan pemahaman lebih mendalam tentang data, sehingga pengambilan keputusan menjadi lebih terinformasi. Keputusan yang didasarkan pada pemahaman yang baik tentang data cenderung lebih akurat dan efektif.
6. **Mengoptimalkan Pemrosesan Data:** EDA dapat membantu Anda menentukan langkah-langkah pemrosesan data yang diperlukan sebelum analisis lebih lanjut. Ini bisa termasuk membersihkan data, mengisi nilai yang hilang, atau mentransformasi variabel.
7. **Komunikasi Hasil:** EDA memungkinkan Anda untuk menyajikan temuan Anda dengan cara yang lebih mudah dipahami oleh audiens yang mungkin tidak memiliki latar belakang teknis. Visualisasi yang dihasilkan dari EDA dapat membantu dalam menjelaskan hasil analisis kepada berbagai pemangku kepentingan.

6. Kesimpulan

a. Dalam pengerjaan praktikum Statistika, kita melakukan analisis mendalam terhadap data yang akan kita teliti. Langkah ini melibatkan berbagai teknik visualisasi dan statistik deskriptif untuk memahami karakteristik data, mengidentifikasi anomali, menemukan pola, serta memahami hubungan antara variabel. EDA membantu membentuk pemahaman awal tentang data yang akan dianalisis sebelum melangkah ke tahap analisis statistik yang lebih kompleks.

b. Kita juga dapat mengetahui wawasan yang mendasar tentang data yang dihadapi. EDA membantu mengungkap informasi yang mungkin tidak terlihat secara langsung dan memungkinkan kita untuk mengambil tindakan yang tepat berdasarkan pemahaman kita tentang data. Dengan melakukan EDA, kita dapat mengidentifikasi dan mengatasi masalah



seperti data yang hilang atau outlier, serta memilih pendekatan analisis yang paling sesuai untuk menggali wawasan lebih dalam. Selain itu, EDA juga memudahkan kita dalam berkomunikasi hasil analisis kepada orang lain dengan menggunakan visualisasi yang efektif.

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 | 15 Menit | Menarik |
| 2. | Latihan Kedua | 15 Menit | Menarik |

Keterangan:

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