**LAPORAN DATA WRANGLING C**

**INTRODUCTION TO PANDAS AND MATPLOTLIB**

**EXERCISE WEEK MODUL 4 DAN 5**



**DISUSUN OLEH:**

Mohamad Ibnu Fajar Maulana (21083010106)

**DOSEN PENGAMPU:**

KARTIKA MAULIDA HINDRAYANI, S.Kom, M.Kom

**PROGRAM STUDI SAINS DATA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”**

**JAWA TIMUR**

**2022**

# **DAFTAR ISI**

[**DAFTAR ISI** i](#_Toc129417920)

[**Exercise Modul Week 4** 1](#_Toc129417921)

[**Input:** 1](#_Toc129417922)

[**Output:** 4](#_Toc129417923)

[**Exercise Modul Week 5** 5](#_Toc129417924)

[**Input-Output:** 5](#_Toc129417925)

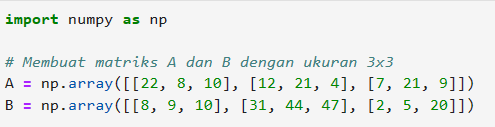
# **Exercise Modul Week 4**

**106\_Mohamad Ibnu   
Fajar Maulana**

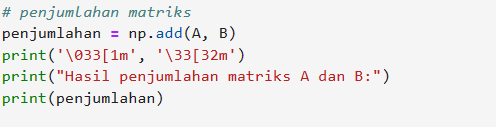
1. Dengan menggunakan data kedua matriks (matriks A dan B) yang degenerate acak pada Array Operations, lakukan penambahan, perkalian, pembagian, 3\*A-2\*B, tambahkan 100 pada scalar matriks A, A dipangkatkan 3, dan akar pangkat dari A

## **Input:**

Pada kasus diatas yaitu mengoperasikan matriks A dan B disini saya menggunakan kedua matriks tersebut berdimensi 3 atau matriks , bisa dilihat pada gambar dibawah ini:



* Jangan lupa untuk mengimport library python *numpy* yang nanti digunakan untuk menghitung komputasi numerik, selanjutnya lakukan koding untuk penjumlahannya seperti gambar dibawah ini.

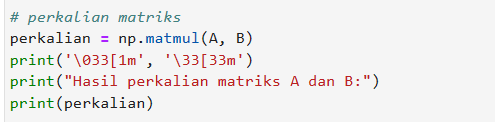


* Pada gambar diatas menggunakan *np.add* yang merupakan fungsi pada library NumPy di Python yang digunakan untuk melakukan operasi penjumlahan antara dua array atau lebih secara elemen-wise. Artinya, fungsi ini akan menambahkan setiap elemen pada array pertama dengan elemen yang sesuai pada array kedua (dan seterusnya, jika terdapat lebih dari dua array), dan menghasilkan array baru yang berisi hasil penjumlahan tersebut. Kemudian setelah melakukan penjumlahan,

**106\_Mohamad Ibnu   
Fajar Maulana**

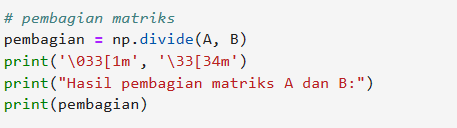
lakukanlah perkalian seperti gambar bawah ini:

**106\_Mohamad Ibnu   
Fajar Maulana**



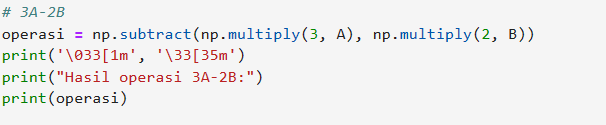
* Pada gambar diatas menggunakan *np.matmul* yang digunakan untuk melakukan perkalian matriks. Fungsi ini menghitung hasil perkalian matriks antara dua array.

Kemudian lakukan pembagian seperti gambar dibawah ini:



* Pada gambar diatas menggunakan *np.divide* merupakan fungsi pada library NumPy di Python yang digunakan untuk melakukan operasi pembagian antara dua array atau lebih secara elemen-wise. Artinya, fungsi ini akan membagi setiap elemen pada array pertama dengan elemen yang sesuai pada array kedua (dan seterusnya, jika terdapat lebih dari dua array), dan menghasilkan array baru yang berisi hasil pembagian tersebut serta dapat menghitung pembagian skalar berapapun.

Kemudian lakukan operasi matriks seperti gambar dibawah ini:



* *np.subtract* dan *np.multiply* adalah fungsi pada library NumPy di Python yang digunakan untuk melakukan operasi pengurangan dan perkalian antara dua array atau lebih secara elemen-wise. *np.subtract* dan *np.multiply* juga dapat digunakan untuk mengurangi atau mengalikan setiap elemen dalam array dengan skalar.

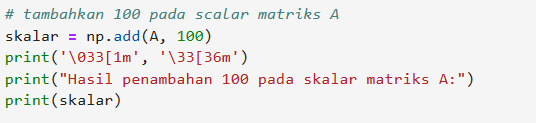
**106\_Mohamad Ibnu   
Fajar Maulana**

* np.subtract digunakan untuk mengurangi setiap elemen pada array pertama dengan elemen yang sesuai pada array kedua (dan seterusnya, jika terdapat lebih dari dua array), dan menghasilkan array baru yang berisi hasil pengurangan tersebut.

**106\_Mohamad Ibnu   
Fajar Maulana**

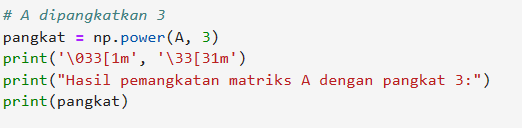
* np.multiply digunakan untuk mengalikan setiap elemen pada array pertama dengan elemen yang sesuai pada array kedua (dan seterusnya, jika terdapat lebih dari dua array), dan menghasilkan array baru yang berisi hasil perkalian tersebut.

Setelah melakukan operasi matriks, selanjutnya lakukanlah penambahan 100 pada skalar A, seperti gambar dibawah ini:



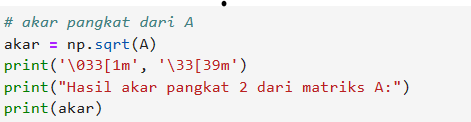
* Pada gambar diatas menggunakan *np.add* untuk menambahkan setiap elemen pada array pertama dengan elemen yang sesuai pada array kedua (dan seterusnya, jika terdapat lebih dari dua array).

Kemudian matriks A dipangkatkan 3, seperti gambar dibawah ini:



* Pada gambar diatas menggunakan *np.power*(A, 3)untuk memangkatkan matriks A kedalam bentuk pangkat 3.

Selanjutnya, menghitung akar pangkat dari matriks A, seperti gambar dibawah ini:



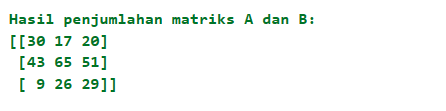
* Pada gambar diatas menggunakan *np.sqrt* untuk mencari nilai akar pangkat dari matriks A.

**106\_Mohamad Ibnu   
Fajar Maulana**

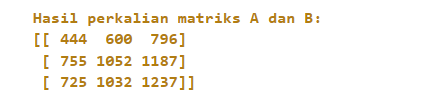
## **Output:**

**106\_Mohamad Ibnu   
Fajar Maulana**

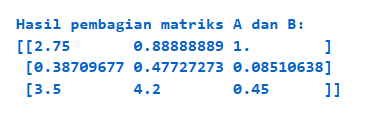
* Penjumlahan Matriks A dan B



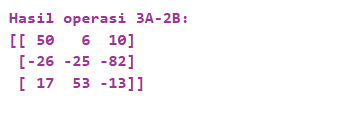
* Perkalian Matriks A dan B



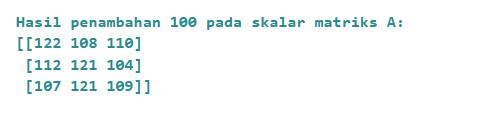
* Pembagian Matriks A dan B



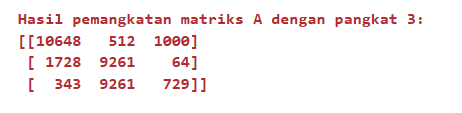
* Hasil Operasi Matriks



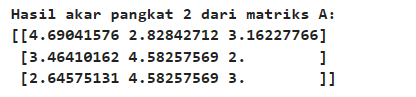
* Penambahan 100 pada skalar Matriks A



* Matriks A dipangkatkan 3



* Akar pangkat Matriks A



**106\_Mohamad Ibnu   
Fajar Maulana**

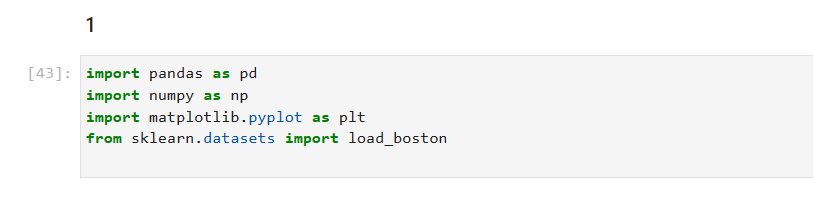
# **Exercise Modul Week 5**

**106\_Mohamad Ibnu   
Fajar Maulana**

1. Load Pustaka/Library yang dibutuhkan.
2. Tampilkan Boston Housing dataset (.csv file)..
3. Periksa 10 baris pertama. Temukan total baris dari dataset tersebut
4. Buat DataFrame yang lebih kecil dengan kolom yang tidak terdapat CHAS, NOX, B, dan LSTAT.
5. Periksa 7 baris terakhir dari DataFrame yang baru.
6. Gambarkan visualisasi histogram dari semua variable/kolom pada DataFrame yang baru.
7. Gambarkan semua histogram menggunakan for loop. Tambahkan nama yang unik pada plot
8. Buat scatter plot dari crime rate versus price.
9. Buat Plot log10(crime) versus price.
10. Kalkulasikan statistic yang berguna, seperti mean rooms per dwelling, median age, mean distances to five Boston employment centers, and the percentage of houses with a low price  
    (< $20,000).

## **Input-Output:**

1. Pada gambar dibawah ini saya mengimport library python *pandas, numpy matplotlib.pyplot, dan sklearn.*



* Pada gambar diatas melakukan import ***pandas*** sebagai alat untuk membaca, memanipulasi, dan menganilisis data dalam bentuk tabel/dataframe atau series.
* Kemudian fungsi ***numpy*** digunakan untuk menghitung komputasi numerik yang nantinya akan membantu dari segi perhitungan.
* Selanjutnya fungsi ***matplotlib.pyplot***digunakan untuk visualisasi data. Library ini menyediakan berbagai jenis plot dan grafik, seperti bar chart, line chart, scatter plot, dan lain-lain. Dengan menggunakan Matplotlib, kita bisa membuat visualisasi data yang menarik dan mudah dipahami.

**106\_Mohamad Ibnu   
Fajar Maulana**

* ***sklearn.datasets*** adalah submodul dari library Scikit-learn (sklearn) yang menyediakan beberapa dataset yang umum digunakan dalam pemodelan dan analisis data. Dataset ini sering digunakan sebagai sampel dalam eksperimen dan penelitian.Kemudian mengimport ***load\_boston*** merupakan salah satu fungsi dari submodul sklearn.datasets yang digunakan untuk memuat dataset Boston Housing. Dataset ini berisi informasi tentang harga rumah di kota Boston dan faktor-faktor yang mempengaruhi harga tersebut, seperti jumlah kamar tidur, tingkat kejahatan di daerah tersebut, aksesibilitas ke jalan raya, dan lain-lain.

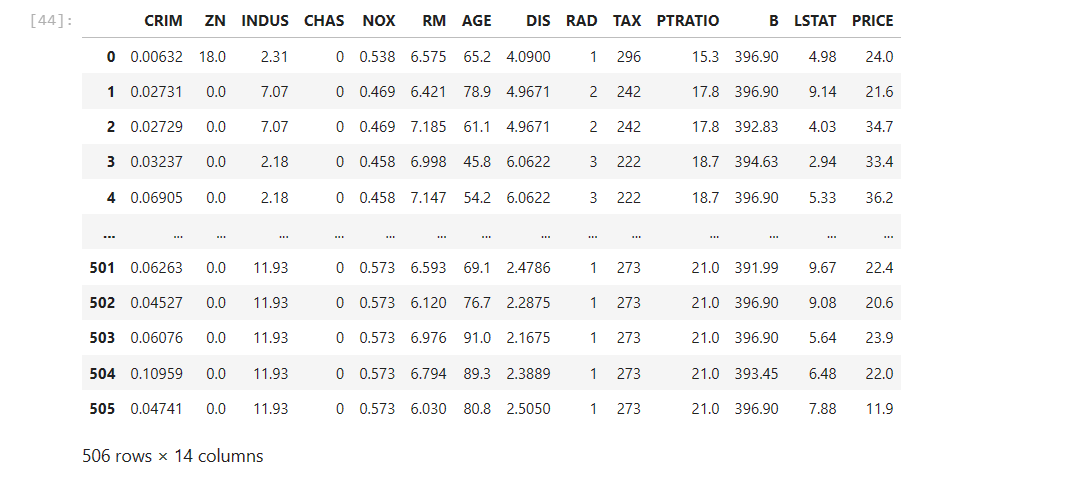
**106\_Mohamad Ibnu   
Fajar Maulana**

Library-library diatas dapat membantu pemrosesan data atau kasus yang diberikan untuk soal nomor 2-10.

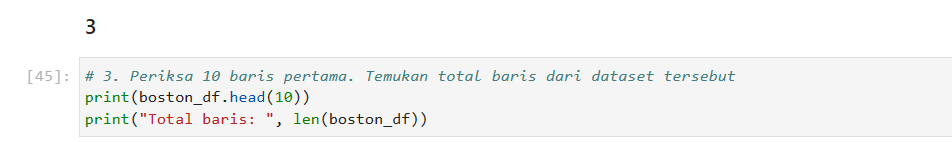
1. Lakukan koding seperti gambar dibawah ini, untuk menampilkan dataframe Boston\_Housing



* Kodingan tersebut mengimport *pandas as pd* yang menghasilkan output file csv seperti gambar dibawah ini:



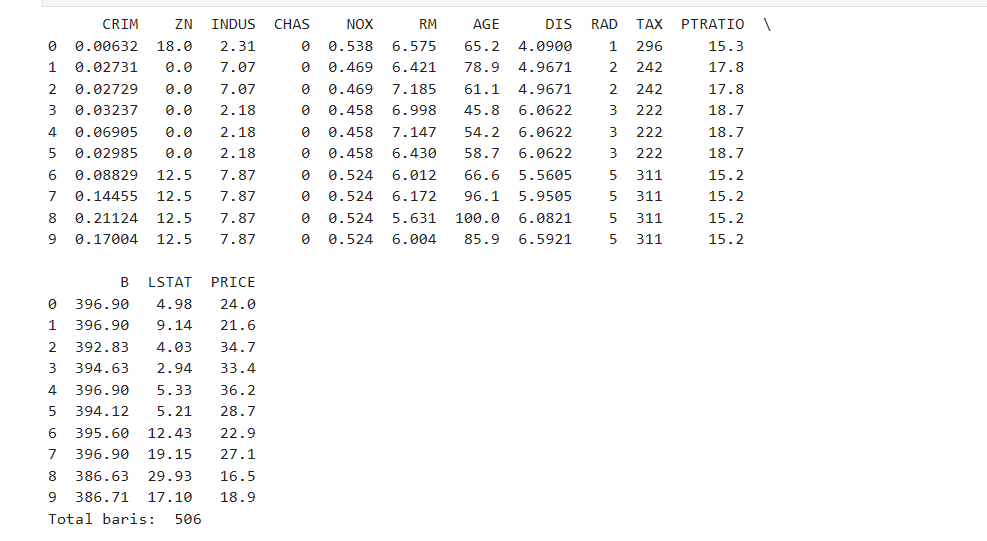
1. Selanjutnya untuk memriksa 10 baris pertama serta menemukan total baris, lakukan koding seperti berikut ini:



**106\_Mohamad Ibnu   
Fajar Maulana**

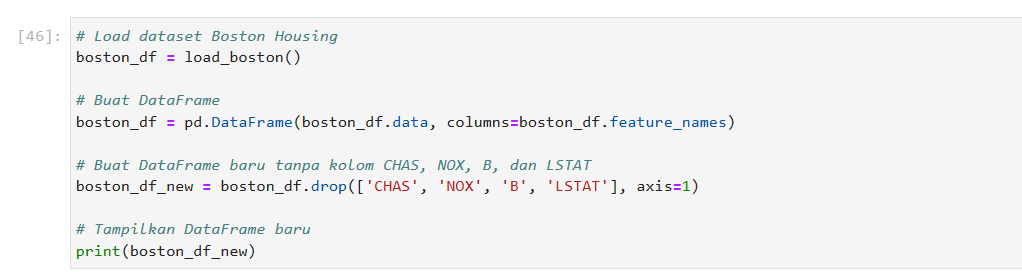
* Kodingan tersebut menggunakan *df.head* untuk melihat 10 baris pertama pada dataframe kemudiang menghitung panjangnya baris dengan fungsi *len* dan menghasilkan output seperti gambar dibawah ini:

**106\_Mohamad Ibnu   
Fajar Maulana**

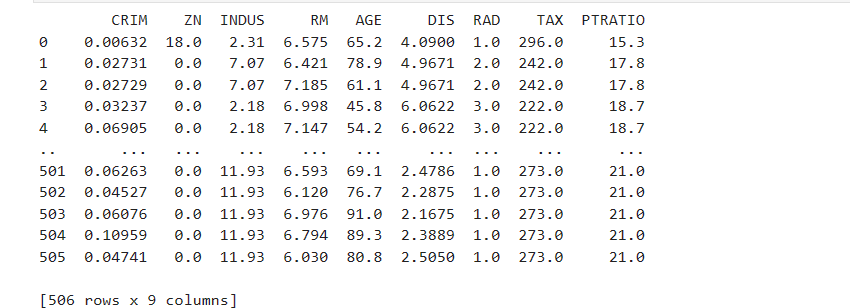


* Gambar diatas menampilkan 10 baris pertama dan total baris yang berjumlah 506

1. Pada kodingan dibawah ini membuat DataFrame yang lebih kecil dengan kolom yang tidak terdapat CHAS, NOX, B, dan LSTAT



* Pada gambar diatas menggunakan library *pandas*, *df.drop* digunakan untuk menghapus baris atau kolom dari dataframe yang tidak diperlukan. Kemudian menghasilkan output seperti gambar dibawah ini:



* Pada gambar diatas sudah tidak ada lagi kolom CHAS, NOX, B, dan LSTAT

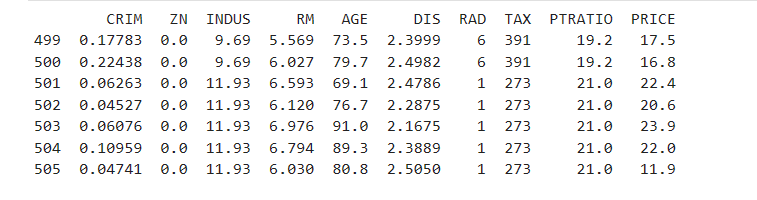
**106\_Mohamad Ibnu   
Fajar Maulana**

1. Memeriksa 7 baris terakhir pada DataFrame yang baru

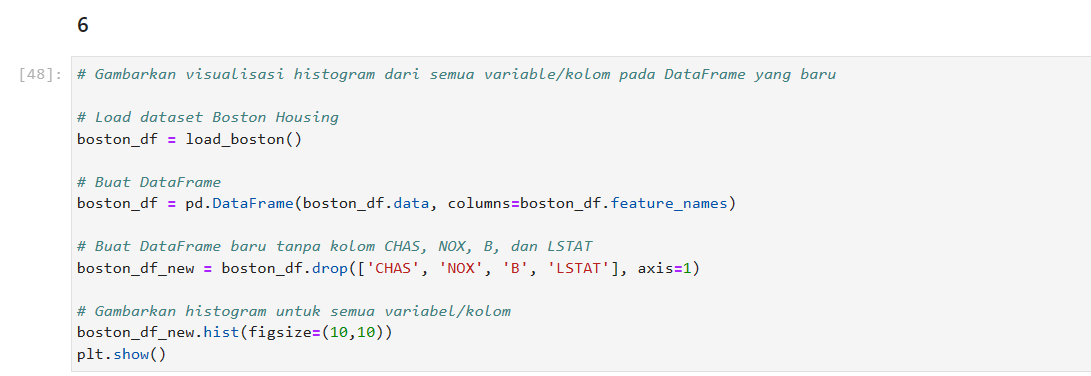
**106\_Mohamad Ibnu   
Fajar Maulana**

* Pada gambar diatas menggunakan library *pandas*, *df.tail* untuk memeriksa baris terakhir kemudian memasukan angka 7 dalam tanda kurung untuk memeriksa 7 baris terakhir dari DataFrame yang baru. Kemudian menghasilkan output:





1. Menggambarkan visualisasi histogram dari semua variable/kolom pada DataFrame yang baru dengan koding seperti gambar dibawah ini:

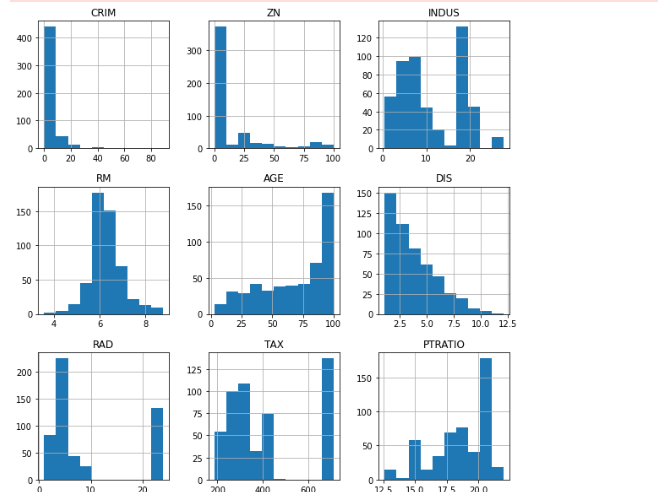


* Pada gambar diatas tidak lupa melakukan load dataset Boston\_Housing dengan menggunakan library *skleran.datasets import load\_boston* untuk memberikan informasi DataFrame Boston\_housing
* Selanjutnya, boston.data adalah sebuah array pada dataset load\_boston pada library scikit-learn yang berisi nilai-nilai numerik yang berhubungan dengan rumah-rumah di Boston pada tahun 1970-an. Dataset ini digunakan sebagai contoh dalam pemodelan regresi.
* Kemudian menghapuskan beberapa kolom yang tidak perlu dengan menggunakan library *pandas*, *df.drop*.

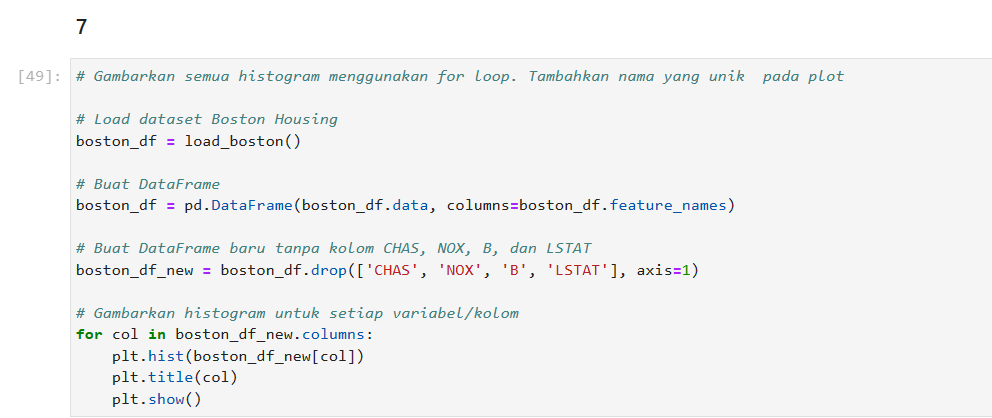
**106\_Mohamad Ibnu   
Fajar Maulana**

* Setelah itu untuk memvisualisasikan gunakan library python *import matplotlib.pyplot as plt* serta *plt.show* untuk menampilkan histogram. Seperti gambar dibawah ini:

**106\_Mohamad Ibnu   
Fajar Maulana**

**

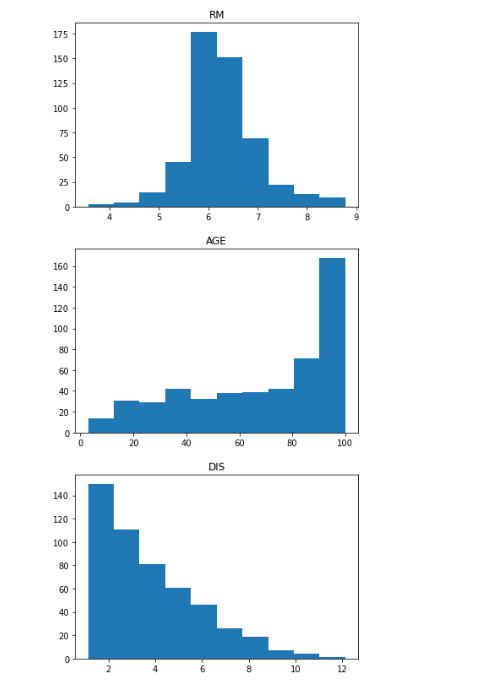
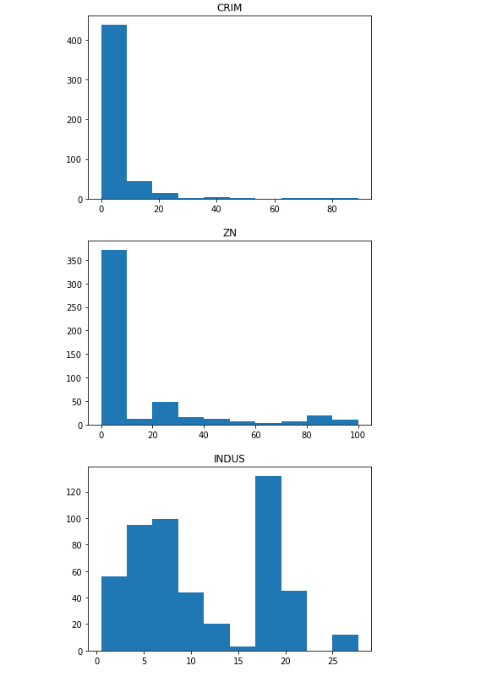
1. Menggambarkan semua histogram menggunakan for loop dan menambahkan nama yang unik pada plot, dengan koding seperti dibawah ini:

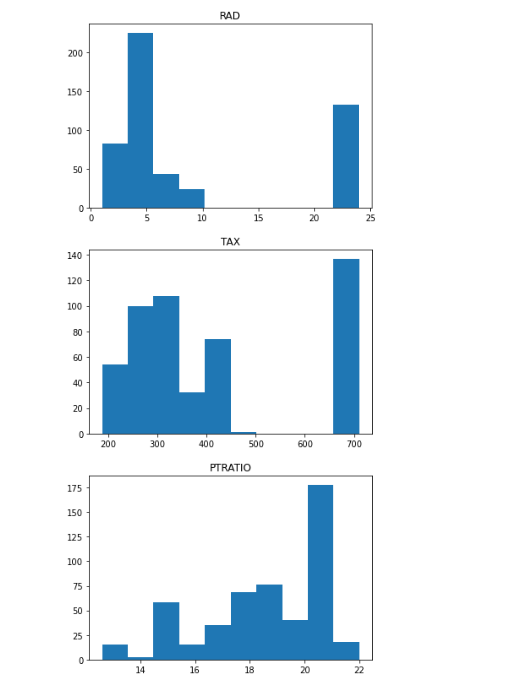


* Pada gambar diatas beberapa koding ada yang saya tidak perlu jelaskan ulang, silahkan baca nomor 6 dan hanya berfokus pada penjelasan *for loop* serta penambahan namaa unik
* Pada gambar diatas menggunakan *for loop* untuk memanggil semua histogram
* Kemudian menggunakan library python *matplotlib as plt*, *plt.hist* untuk memvisualisasikan data dalam bentuk histogram, selanjutnya *plt.title* untuk menambahkan nama unik(judul) untuk masing masing kolom pada histogram dan yang terakhir *plt.show* untuk menampilkan hasil eksekusi yakni histogram seperti gambar dibawah ini:

**106\_Mohamad Ibnu   
Fajar Maulana**

**106\_Mohamad Ibnu   
Fajar Maulana**

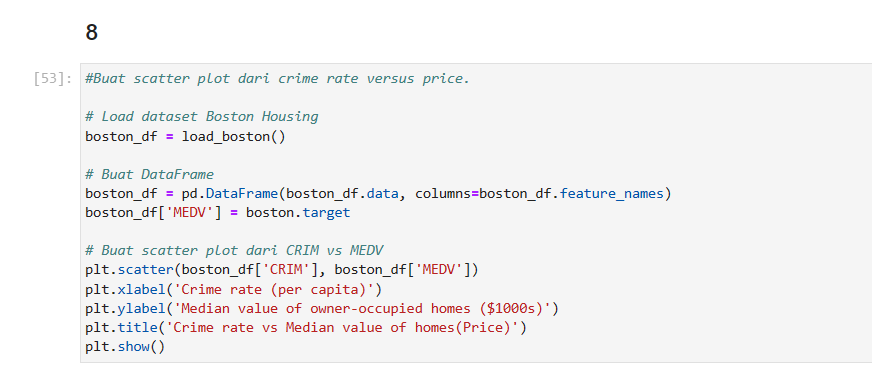
**



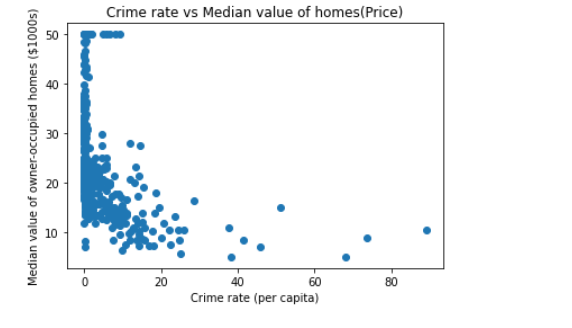
**106\_Mohamad Ibnu   
Fajar Maulana**

1. Membuat scatterplot dari crime rate versus price

**106\_Mohamad Ibnu   
Fajar Maulana**



* Pada gambar diatas beberapa koding ada yang saya tidak perlu jelaskan ulang, silahkan baca nomor 6. Disini saya berfokus menjelaskan scatterplot.
* Gambar diatas menggunakan boston.data dan boston.target digunakan untuk memprediksi nilai nilai median pada rumah boston\_housing
* Pada kodingan diatas menggunakan library python *matplotlib as plt,* *plt.scatter* untuk menampilkan penyebaran data dari Crime Rate atau per-kapita dengan Median Value of Homes atau price(harga rumah). Dengan *plt.xlabel* memberikan label pada sumbu yang menafsirkan Crime Rate dan dengan *plt.ylabel* memberikan label pada sumbu yang menafsirkan Median value of owner-occupied homes ($1000s). kemudian *plt.tittle* untuk memberikan judul pada scatterplot Crime rate vs Median Value of Homes(Price), dan terakhir untuk menampilkan scatterplot menggunakan *plt.show*



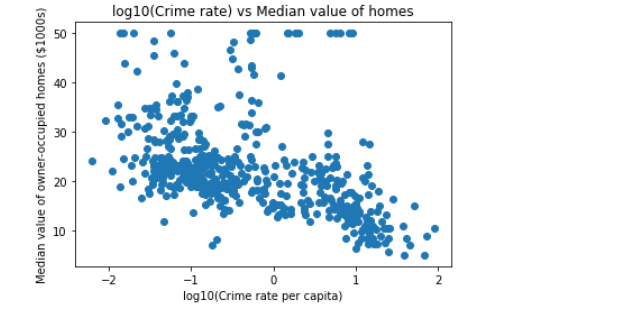
**106\_Mohamad Ibnu   
Fajar Maulana**

1. Membuat Plot log10(crime) versus price.

**106\_Mohamad Ibnu   
Fajar Maulana**



* Pada gambar diatas hampir semua kodingan sudah saya jelaskan, silahkan simak kembali nomor 6&8 agar memahami kodingan tersebut.
* Disini saya akan menjelaskan apa itu log10(crime), log10(crime) adalah hasil dari penghitungan logaritma basis 10 dari variabel crime rate per capita (CRIM) pada dataset Boston Housing. Tujuan dari menghitung log10(crime) adalah untuk mengubah distribusi data menjadi lebih normal dan mengurangi efek outlier pada data.
* Kemudian menghasilkan output log(CRIM) vs Price sebagai berikut ini;



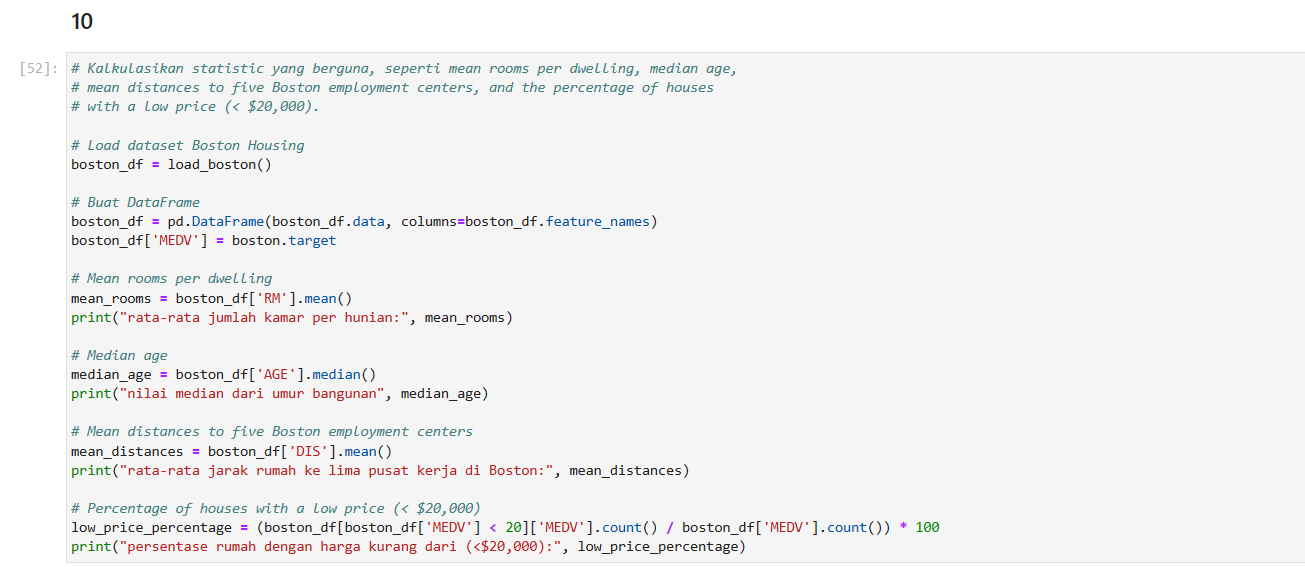
* NOTE: Pada gambar 8-9 sering kali kita melihat **MEDV,** apa itu **MEDV?**

MEDV (median value of owner-occupied homes in $1000's) adalah salah satu dari 14 variabel atau fitur pada dataset Boston Housing yang dikembangkan oleh Harrison dan Rubinfeld (1978). MEDV adalah nilai median dari harga rumah yang dihuni oleh pemilik rumah dalam skala ribuan dolar ($1000).

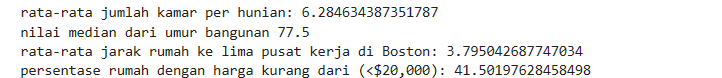
**106\_Mohamad Ibnu   
Fajar Maulana**

1. Kalkulasikan statistik yang berguna, seperti mean rooms per dwelling, median age, mean distances to five Boston employment centers, and the percentage of houses with a low price  
   (< $20,000).

**106\_Mohamad Ibnu   
Fajar Maulana**



* Pada gambar diatas perhatikan kodingannya dan simak baik-baik, karena saya tidak akan menjelaskannya lebih lanjut, sebab pada kodingan diatas sudah terlihat jelas penjelasannya. Saya hanya menjelaskan output dari kodingan tersebut:



* Pada kodingan diatas memiliki maksud menghitung statistik
* Statistik yang dihitung adalah:
  + Mean rooms per dwelling: rata-rata jumlah kamar per hunian
  + Median age: nilai median dari umur bangunan
  + Mean distances to five Boston employment centers: rata-rata jarak rumah ke lima pusat kerja di Boston
  + Percentage of houses with a low price (< $20,000): persentase rumah dengan harga kurang dari $20,000.

**106\_Mohamad Ibnu   
Fajar Maulana**