

LAPORAN

Mengolah Data Tabel Pada Python



Disusun oleh :

Dyan Kartikasari 03411940000009

Fakultas Teknik Sipil, Perencanaan, dan Kebumihan

Institut Teknologi Sepuluh Nopember (ITS)

2020

Mengolah Data Tabel Menggunakan Python

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan *syntax* kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif.

Salah satu fitur yang tersedia pada python adalah sebagai bahasa pemrograman dinamis yang dilengkapi dengan manajemen memori otomatis. Seperti halnya pada bahasa pemrograman dinamis lainnya, python umumnya digunakan sebagai bahasa script meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa script. Python dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi. Python memiliki banyak *library* di dalamnya. *Library* pada python dapat digunakan untuk membantu menyelesaikan atau mengolah suatu data. Contoh *library* yang ada pada python adalah *numpy*, *matplotlib*, dan *pandas*(*Python Data Analysis Library*).

Berikut adalah cara mengaplikasikan *pandas* dan *matplotlib*. Data yang digunakan adalah data jumlah penduduk Surabaya pada tahun 2019 dalam format *Comma Separated Values* (.csv). Pada tugas kali ini adalah mengimpor data dengan menggunakan [Github.com](https://github.com) sebagai penghubung dengan *python*. Berikut adalah data yang digunakan.

	umur	laki-laki	perempuan	jumlah
1				
2	0-4	107043	102663	209706
3	5-9	107536	103436	210972
4	10-14	98976	95048	194024
5	15-19	108479	118284	226763
6	20-24	144661	147358	292019
7	25-29	137343	135581	272924
8	30-34	125978	126098	252076
9	35-39	116992	119206	236198
10	40-44	105843	108846	214689
11	45-49	95379	101564	196943
12	50-54	86854	94898	181752
13	55-59	72281	79841	152122
14	60-64	52247	52420	104667
15	65-69	35742	33603	69345
16	70-74	19089	20853	39942
17	75+	16545	25508	42053

Data tersebut dapat diolah menggunakan *python*. Berikut adalah langkah yang harus dilakukan untuk mengolah data tersebut.

1.

```
import pandas as pd
```

Import pandas as pd adalah mengimport *library* pandas pada *python* yang berfungsi untuk memanipulasi data, persiapan data, dan pembersihan data.

2.

```
data=pd.read_csv('https://raw.githubusercontent.com/DyanKartikasari/TugasAsistensiW5/master/DataKependudukanSurabaya2019.csv')
```

Membuat variabel data yang berfungsi sebagai nama file yang akan digunakan. Format `pd.read_csv` berfungsi untuk membaca data dari file csv yang tertera pada link tersebut.

3.

```
data.head()
```

Menggunakan `data.head()` untuk menampilkan 5 data paling atas pada keseluruhan data, sehingga output yang didapatkan sebagai berikut.

	umur	laki-laki	perempuan	jumlah
0	0-4	107043	102663	209706
1	5-9	107536	103436	210972
2	10-14	98976	95048	194024
3	15-19	108479	118284	226763
4	20-24	144661	147358	292019

4.

```
data.tail()
```

Menggunakan `data.tail()` untuk menampilkan 5 data paling bawah pada keseluruhan data, sehingga output yang didapatkan sebagai berikut.

	umur	laki-laki	perempuan	jumlah
11	55-59	72281	79841	152122
12	60-64	52247	52420	104667
13	65-69	35742	33603	69345
14	70-74	19089	20853	39942
15	75+	16545	25508	42053

5.

```
data.info()
```

Menggunakan `data.tail()` untuk menampilkan nomor index, *type* data yang digunakan dan menampilkan informasi yang berkaitan dengan data yang digunakan. Sehingga output yang didapatkan adalah sebagai berikut.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16 entries, 0 to 15
Data columns (total 4 columns):
umur          16 non-null object
laki-laki     16 non-null int64
perempuan    16 non-null int64
jumlah        16 non-null int64
dtypes: int64(3), object(1)
memory usage: 640.0+ bytes
```

6.

```
data.describe()
```

Menggunakan `data.describe()` untuk menampilkan statistika data seperti rata-rata, median, *maximum*, *minimum*, *quartile*, dll. Output yang didapatkan adalah sebagai berikut.

	laki-laki	perempuan	jumlah
count	16.000000	16.000000	16.000000
mean	89436.750000	91575.437500	181012.187500
std	39719.123318	38986.966639	78562.464808
min	16545.000000	20853.000000	39942.000000
25%	67272.500000	72985.750000	140258.250000
50%	102409.500000	102113.500000	203324.500000
75%	110607.250000	118514.500000	229121.750000
max	144661.000000	147358.000000	292019.000000

7.

```
data.groupby('jumlah').mean()
```

Menggunakan `data.groupby(jumlah).mean()` untuk mengelompokkan data yang diinginkan menjadi data rata-rata. Output yang didapatkan adaah sebagai berikut.

	laki-laki	perempuan
jumlah		
39942	19089	20853
42053	16545	25508
69345	35742	33603
104667	52247	52420
152122	72281	79841
181752	86854	94898
194024	98976	95048
196943	95379	101564
209706	107043	102663
210972	107536	103436
214689	105843	108846
226763	108479	118284
236198	116992	119206
252076	125978	126098
272924	137343	135581
292019	144661	147358

8.

```
import matplotlib.pyplot as plt
```

Import *matplotlib* as *plt* adalah mengimport *library* *matplotlib* pada *python* yang berfungsi untuk menampilkan grafik.

9.

```
plt.figure(figsize=(5,5))
plt.plot(data['laki-laki'], label='laki-laki')
plt.plot(data['perempuan'], label='perempuan')
plt.xlabel('data count 2019')
plt.ylabel('value')
plt.title('penduduk kota surabaya 2019')
plt.legend()
plt.show()
```

Pada script diatas diawali dengan `plt.figure(figsize=(5,5))` yang artinya membuat grafik dengan ukuran yang sudah ditentukan yaitu berukuran 5x5. Langkah kedua, `plt.plot(data['laki-laki'], label='laki-laki')` yang artinya memasukkan data laki-laki ke dalam grafik, dan diberi label 'laki-laki'. Langkah ketiga, `plt.plot(data['perempuan'],`

`label='perempuan')` yang artinya memasukkan data perempuan ke dalam grafik dan diberi label 'perempuan'. Langkah keempat, `plt.xlabel('data count 2019')` yang artinya memberi label pada sumbu x grafik, pada grafik tersebut diberi label 'data count 2019'. Langkah kelima, `plt.ylabel('value')` yang artinya memberi label pada sumbu y grafik, pada grafik tersebut diberi label 'value'. Langkah keenam, `plt.title('penduduk kota surabaya 2019')` yang artinya memberi judul pada grafik, pada grafik tersebut diberi nama 'penduduk kota surabaya 2019'. Langkah ketujuh, `plt.legend()` yang berfungsi untuk menampilkan legenda dari grafik yang dibuat, meliputi judul, label, warna, dan keterangan lainnya. Langkah terakhir yaitu `plt.show()` yang berfungsi untuk menampilkan grafik pada hasil script Python (Window). Sehingga output yang didapatkan sebagai berikut.

