Operasi Python Menggunakan Data Analysis Library (Pandas)

Phyton memiliki 3 library terbaik dalam pengolahan data science. Dalam pengolahan data tugas kali ini menggunakan 2 library ,yaitu Pandas dan Numpy. Dalam pengolahan data dengan python lebih efisiensi,keterbacaan kode dan kecepatan telah membuat pyton menjadi bahasa pemogramaan yang banyak digunakan.Pyton menjadi pilihan untuk para data scientist dan machine learning engineer untuk mengembangkan model dan berbagai aplikasi terkait data science.berikut penejelasan mengenai Pandas dan Numpy library :

1. Pandas

Pandas kependekan dari Python Data Analysis Library. Pandas merupakan sebuah *Open source package library* dengan lisensi BSD yang menyediakan banyak perkakas untuk kebutuhan data analisis, maipulasi,dan pembersihan data.Pandas mendukung pembacaan dan penulisan data dengan media berupa *Excel Spreadsheet*,CSV,dan SQL yang kemudian akan dijadikan sebagai objek python dengan *rows* dan *colums* yang disebut *data frame* seperti halnya pada tabel statistik.

Ada tiga jenis struktur data di library ini:

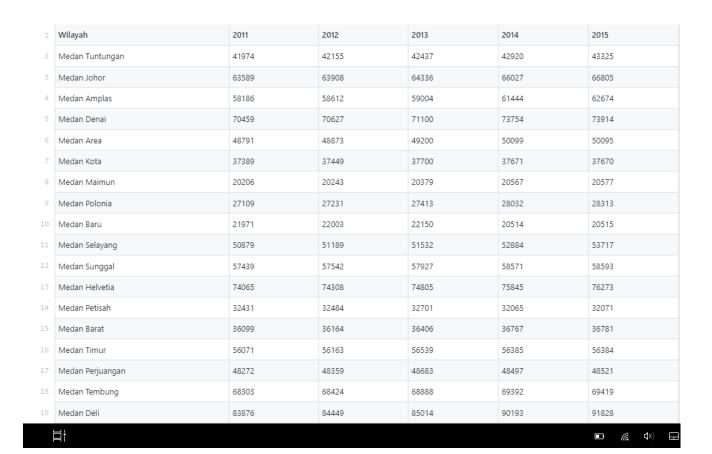
- 1. Series: single-dimensional, array homogeny
- 2. DataFrame: two-dimensional dengan kolom yang diketik secara heterogen
- 3. Panel: three-dimensional, array size-mutable

Sebagai contoh,kita menggunakan library Panda Python yang disingkat (pd) dapat digunakan untuk melakukan beberapa perhitungan statistik deskriptif.

```
In [1]: M import pandas as pd

In [10]: M data=pd.read_csv('https://raw.githubusercontent.com/Yosiaginting/tugas-phyton02/master/data%20penduduk%20medan.csv')
```

Adapun hasil raw table data "jumlah penduduk untuk wilayah Medan dengan rentang tahun 2011-2015" yang didapatkan dari GITHUB adalah sebagai berikut:



Berikut beberapa Tools yang digunakan untuk melihat/mengecek data:

♦ head(n): Berfungsi untuk melihat data sebanyak n pada kolom awal.Dimana data yang ditampilkan berupa data awal dari 0 s.d 4 (memuat 5 data awal dari 0). Berikut Ouput yang didapatkan pada data ini,sebagai berikut:



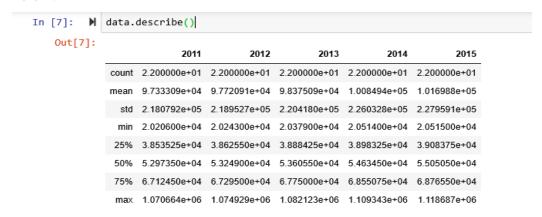
❖ tail(n): Berfungsi untuk melihat data sebanyak n pada kolom akhir. Dimana data ynag ditampilkan berupa data akhir dari 17 s.d 21 (data 5 terakhir). berikut Output yang didaptkan pada dat ini, sebagai berikut:

In [4]: N	dat	a.tail()						
Out[4]:		Wilayah	2011	2012	2013	2014	2015	
	17	Medan Deli	83876	84449	85014	90193	91828	
	18	Medan Labuhan	55068	55309	55679	58910	59447	
	19	Medan Marelan	71786	72645	73131	79180	82115	
	20	Medan Belawan	46701	46794	47105	49626	49650	
	21	Medan	1070664	1074929	1082123	1109343	1118687	

info(): Nomor index beserta tipe datanya. Tipe data yang digunakan. Berikut Output yang didaptakn pada data ini ,sebagai berikut:

```
In [5]:
         data.info()
            <class 'pandas.core.frame.DataFrame'>
            RangeIndex: 22 entries, 0 to 21
            Data columns (total 6 columns):
            Wilayah
                       22 non-null object
            2011
                       22 non-null int64
            2012
                       22 non-null int64
            2013
                       22 non-null int64
            2014
                       22 non-null int64
            2015
                       22 non-null int64
            dtypes: int64(5), object(1)
            memory usage: 1.2+ KB
```

describe(): Menunjukkan rangkuman statistik seperti rata-rata, median, dll pada kolom.



❖ Dilanjutkan dengan menggunakan "Data groupby ('Jumlah penduduk di wilayah Medan dan sekitarnya dalam rentang tahun 2011 sampai dengan 2015").mean()" yang artinya adalah mengelompokkan data yang diinginkan menjadi data rata-rata sehingga ditambahkan "mean" pada akhir formula. Output yang didaptkan adalah

	data.gro		,	(/		
Out[8]:		2011	2012	2013	2014	
	2015					
	20515	21971	22003	22150	20514	
	20577	20206	20243	20379	20567	
	28313	27109	27231	27413	28032	
	32071	32431	32484	32701	32065	
	36781	36099	36164	36406	36767	
	37670	37389	37449	37700	37671	
	43325	41974	42155	42437	42920	
	48521	48272	48359	48683	48497	
	49650	46701	46794	47105	49626	
	50095	48791	48873	49200	50099	
	53717	50879	51189	51532	52884	

Untuk menampilkan sebuh grafik berdasarkan data yang di input,harus mengimpor library lain yang akan mendukung pembentukan grafik yaitu dengana mengimpor data ke matplotlib. Matplotlib juga merupakan bagian dari paket inti SciPy dan ditawarkan di bawah lisensi BSD. Ini adalah library ilmiah Python populer yang digunakan untuk menghasilkan visualisasi yang sederhana dan kuat. Anda dapat menggunakan kerangka kerja Python untuk ilmu data untuk menghasilkan grafik, chart, histogram, dan bentuk dan gambar lain yang kreatif — tanpa perlu khawatir menulis banyak baris kode. Sebagai contoh, mari kita lihat bagaimana perpustakaan Matplotlib dapat digunakan untuk membuat bar chart sederhana.

Mari memulainya dengan mengimport library:

```
In [11]: | import matplotlib.pyplot as plt
In [17]: | plt.figure(figsize=(10,5))
    plt.plot(data['2011'],label = '2011')
    plt.plot(data['2012'],label='2012')
    plt.plot(data['2013'],label = '2013')
    plt.plot(data['2014'],label = '2014')
    plt.plot(data['2015'],label='2015')
    plt.xlabel('wilayah')
    plt.ylabel('jumlah penduduk')
```

Penjelasan dari data yang di input adalah dimama (figsize) menjelaskan mengenai skala dari grafik 2 dimensi yang akan ditampilkan,pada gradik ini menampilkan jumla populasi penduduk di setiap wilayah tersebut.

