


<b>Nama:</b> <b>Muhammad Ali Akbar</b> <b>Hidayatullah</b>  <b>NIM:</b> <b>064002300034</b>		<b>MODUL 12</b>  <b>Nama Dosen:</b> <b>Ratna Shofiati, S.Kom, M.</b> <b>Kom</b>
<b>Hari/Tanggal:</b> <b>Kamis 14 Desember 2023</b>	<b>Praktikum Algoritma &amp;</b> <b>Pemrograman</b>	<b>Nama Asisten Labratorium:</b> <b>1. Yuda Hadi Prasetyo</b> <b>– 065002100004</b> <b>2. Muhammad Hasan</b> <b>Husein –</b> <b>065002100009</b>

## Pengantar Data Science

### 1. Teori Singkat

Data Science adalah suatu disiplin ilmu yang khusus mempelajari data, khususnya data kuantitatif, baik yang terstruktur maupun tidak terstruktur. Banyak Bahasa pemrograman yang dimana dia mendukung untuk melakukan pengolahan data, diantaranya yaitu Bahasa Pemrograman R, Python, SQL, dan JavaScript, dan lain sebagainya. Python sendiri merupakan salah satu bahasa yang mendukung untuk melakukan pengolahan data, bahkan python sendiri menyediakan library untuk pengolahan data itu sendiri, salah satunya adalah library *pandas*. Untuk melakukan pengolahan data python sendiri merekomendasikan untuk menggunakan IDE yang disediakan untuk melakukan pengolahan data yaitu adalah Jupyter Notebook.

#### Pandas Data Frame

Struktur data dasar pandas dinamakan DataFrame, yaitu sebuah koleksi kolom berurutan dengan nama dan jenis, dengan demikian merupakan sebuah tabel yang tampak seperti database dimana sebuah baris tunggal mewakili sebuah contoh tunggal dan kolom mewakili atribut tertentu. Pandas data frame juga dapat disebut sebagai dictionary of list karena



bentuknya seperti list yang memiliki identifikasi *key-value* untuk mengidentifikasi setiap datanya.

### Contoh Program Data Science

```
import pandas as pd

data = {"Negara": ["Indonesia", "Jepang", "India", "China", "Amerika Serikat", "Brazil"],
        "Ibu Kota": ["Jakarta", "Tokyo", "New Delhi", "Beijing", "Washington, D.C.", "Brazilia"],
        "Benua": ["Asia", "Asia", "Asia", "Asia", "Amerika", "Amerika"],
        "Luas": [1905, 377, 3287, 9597, 9834, 8515],
        "Populasi": [264, 143, 1252, 1357, 329, 210] }

df = pd.DataFrame(data)
mean = df.groupby(['Benua']).mean()
std = df.groupby(['Benua']).std()

print(df)
print(mean)
print(std)
```

### Output

```

      Negara      Ibu Kota  Benua  Luas  Populasi
0  Indonesia      Jakarta   Asia  1905      264
1    Jepang      Tokyo     Asia   377      143
2     India  New Delhi     Asia  3287     1252
3     China    Beijing     Asia  9597     1357
4 Amerika Serikat Washington, D.C. Amerika  9834      329
5     Brazil    Brazilia  Amerika  8515      210

      Luas  Populasi
Benua
Amerika  9174.5      269.5
Asia     3791.5      754.0

      Luas  Populasi
Benua
Amerika  932.673844  84.145707
Asia     4048.705266  639.018518
```



## 2. Alat dan Bahan

Hardware : Laptop/PC

Software : Spyder (Anaconda Python)

## 3. Elemen Kompetensi

### a. Latihan pertama

Buatlah sebuah program yang dapat membaca data frame yang datanya diambil dari file CSV, buatlah minimal 10 data negara dan tampilkan Mean(Rata-rata) dan Standar Deviasinya.

### Contoh Output

```

Berikut Data Framenya:
      Ibu Kota   Benua   Luas   Populasi
Negara
Indonesia      Jakarta   Asia   1905     264
Jepang           Tokyo   Asia    377     143
India           New Delhi   Asia   3287    1252
China            Beijing   Asia   9597    1357
Amerika Serikat Washington DC Amerika 9834     329
Brazil           Brazilia Amerika 8515     210
Rusia            Moskow   Asia  17098     146
Meksiko          Meksiko City Amerika 1964     126
Nigeria          Abuja   Afrika  923     200
Jerman           Berlin   Eropa   357      83
Aljazair         Aljazair Afrika 2381      43
Inggris          London   Eropa   242      66

Berikut Data Mean:
      Luas   Populasi
Benua
Afrika  1652.0  121.500000
Amerika 6771.0  221.666667
Asia    6452.8  632.400000
Eropa    299.5   74.500000

Berikut Data Standard Deviation:
      Luas   Populasi
Benua
Afrika 1030.961687 111.015765
Amerika 4214.899406 102.001634
Asia    6906.996395  616.596546
Eropa    81.317280  12.020815

```



### Source Code

```
import pandas as pd

df = pd.read_csv('Negara.csv', index_col=0)

print(df.head())

mean = df.groupby(['Benua']).mean(numeric_only=True)
std = df.groupby(['Benua']).std(numeric_only=True)

print("Berikut Mean dari masing-masing benua :")
print(mean, '\n')

print("Berikut Data Standard Deviation benua :")
print(std, '\n')
```

### Output

```

                Ibu Kota    Benua    Luas    Populasi
Negara
Indonesia          Jakarta    Asia    1905         264
Jepang              Tokyo    Asia     377         143
India              New Delhi    Asia    3287        1252
China              Beijing    Asia    9597        1357
Amerika Serikat Washington DC Amerika    9834         329
Berikut Mean dari masing-masing benua:
                Luas    Populasi
Benua
Afrika      1652.0    121.500000
Amerika     6771.0    221.666667
Asia        6452.8    632.400000
Eropa        299.5     74.500000

Berikut Data Standard Deviation benua:
                Luas    Populasi
Benua
Afrika      1030.961687    111.015765
Amerika     4214.899406    102.001634
Asia        6906.996395    616.596546
Eropa         81.317280     12.020815
```



b. Latihan Kedua

Buatlah sebuah program yang dapat menulis file CSV yang berisi data mean dan standar deviasi dari hasil keluaran data frame Negara, Luas Area dan Total Populasi pada Latihan sebelumnya. Terdapat dua File output yaitu NegaraStandarDeviasi.csv dan juga NegaraMean.csv

Contoh Output

```

Berikut Data Framenya:
Negara      Ibu Kota  Benua  Luas  Populasi
Indonesia   Jakarta  Asia   1905   264
Jepang       Tokyo    Asia   377    143
India        New Delhi Asia   3287   1252
China        Beijing  Asia   9597   1357
Amerika Serikat Washington DC Amerika 9834   329
Brazil       Brasilia Amerika 8515   210
Rusia        Moscow   Asia   17098  146
Meksiko      Meksiko City Amerika 1964   126
Nigeria      Abuja    Afrika  923    200
Jerman       Berlin   Eropa   357    83
Aljazair     Aljazair Afrika  2381   43
Inggris      London   Eropa   242    66

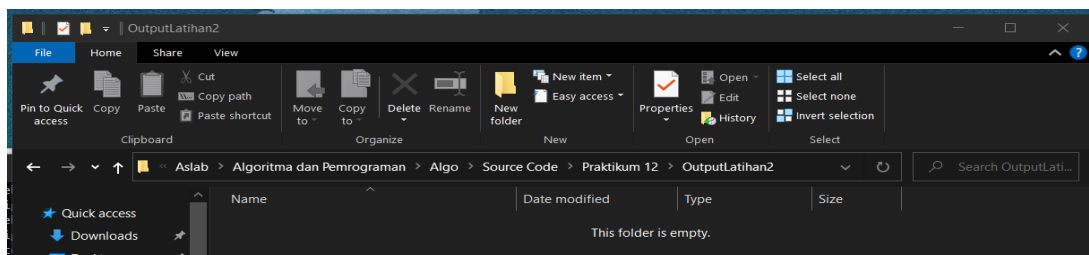
Berikut Data Mean:
Luas  Populasi
Benua
Afrika 1652.0 121.500000
Amerika 6771.0 221.666667
Asia 6452.8 632.400000
Eropa 299.5 74.500000

Berikut Data Standard Deviation:
Luas  Populasi
Benua
Afrika 1030.961687 111.015765
Amerika 4214.899406 102.001634
Asia 6906.996395 616.596546
Eropa 81.317280 12.020815

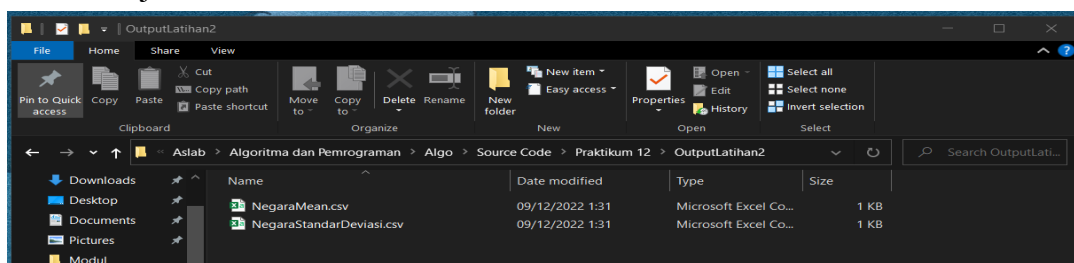
File Berhasil Dibuat

```

Sebelum dijalankan



Sesudah dijalankan



### Source Code

```
import pandas as pd

df = pd.read_csv('Negara.csv', index_col=0)

print(df.head())

mean = df.groupby(['Benua']).mean(numeric_only=True)
std = df.groupby(['Benua']).std(numeric_only=True)

print("\n")
print("Berikut Mean dari benua berikut:")
print(mean, '\n')

print("Berikut Data Standard Deviation benua tadi:")
print(std, '\n')

df.to_csv('Mean beberapa benua.csv', index=False)
df.to_csv('Data Stadar deviasi benua.csv', index=False)
```

### Output

```

Negara          Ibu Kota    Benua    Luas    Populasi
Indonesia      Jakarta     Asia    1905      264
Jepang          Tokyo       Asia     377      143
India           New Delhi    Asia   3287    1252
China           Beijing     Asia   9597    1357
Amerika Serikat Washington DC Amerika 9834      329

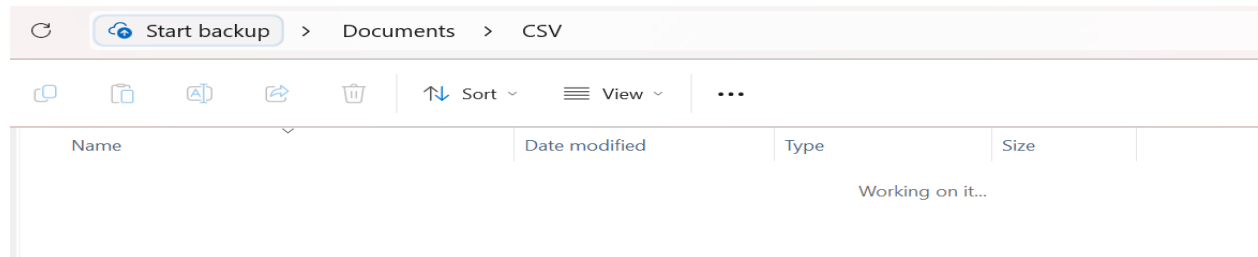
Berikut Mean dari benua berikut:
          Luas    Populasi
Benua
Afrika    1652.0    121.500000
Amerika   6771.0    221.666667
Asia      6452.8    632.400000
Eropa      299.5     74.500000

Berikut Data Standard Deviation benua tadi:
          Luas    Populasi
Benua
Afrika    1030.961687    111.015765
Amerika   4214.899406    102.001634
Asia      6906.996395    616.596546
Eropa      81.317280     12.020815

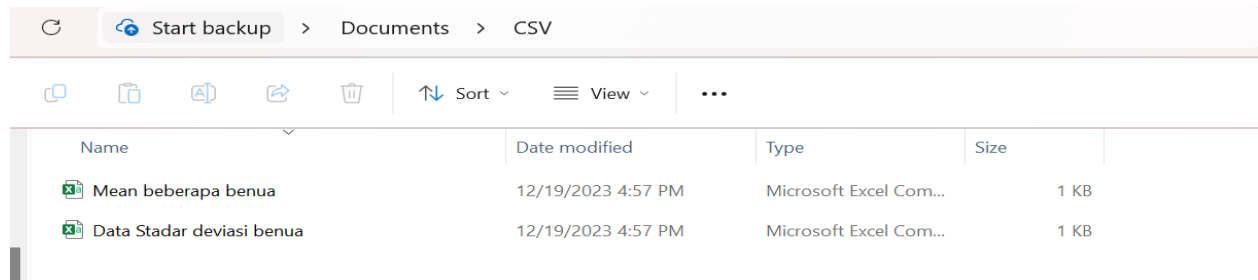
file berhasil dibuat
```



Sebelum dijalankan



Sesudah dijalankan



#### 4. File Praktikum

Github Repository:

#### 5. Soal Latihan

Soal:

1. Apa kegunaan data frame dalam library pandas pada bahasa pemrograman python dan jelaskan mengapa sebelum melakukan pengolahan data kita diharuskan menggunakan data frame?
2. Deskripsikan serta narasikan jalannya alur source code program yang sebelumnya telah kalian buat pada Elemen Kompetensi Latihan Kedua!

Jawaban:

1. kegunaannya yaitu untuk memanipulasi dan visualisasi data dan alasan mengapa kita harus menggunakan data frame sebelum melakukan pengolahan data yaitu karna data frame menyediakan struktur data untuk mewakili data
2. download file negara csv untuk membaca file tersebut. perintah pandas as pd juga penting, lalu groupby diperlukan untuk mengelompokkan data lalu untuk membuat file csv di python gunakan df.to\_csv(mean/data standard deviasi benua) otomatis program akan membuat file csv tersebut.



## 6. Kesimpulan

- Dalam pengerjaan program dengan bahasa pemrograman Python, kita harus benar-benar teliti dalam menginputkan suatu fungsi untuk menampilkan suatu keluaran pada layar dengan sesuai.
- Kita dapat mengetahui di modul ini yang mempelajari tentang materi pandas as pd ini sama seperti sebelumnya namun di modul ini kita harus membaca csv file dari python juga membuat file csv dari python jadi tujuannya supaya kita bisa lebih memahami materi ini secara jelas

## 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	30 Menit	Menarik
2.	Latihan Kedua	30 Menit	Menarik

Keterangan:

- Menarik
- Baik
- Cukup
- Kurang

