

NIM : 2111521008
Nama : Septia Amanda

Tanggal : Jumat, 04 November 2020
Asisten : 1. Bobby Darmawan
2. Dwisuci Insani Karimah
3. Iqbal Fitrahul Ramadhan
4. Iqbal Manazil Yuni
5. Muhammad Afif
6. M. Rayhan Rizaldi

Mata Kuliah : Praktikum Data Mining
Modul : 01
Kelas : 02

Resume dan Tugas “Instalasi Python dan Pengolahan Data di Excel ”

Masukkan screenshot instalasi python, serta screenshot data excel yang sudah diolah dan lakukan analisis pengolahan datanya.

A. Resume Materi

1. Pengertian Python

Python adalah bahasa pemrograman yang banyak digunakan dalam aplikasi web, pengembangan perangkat lunak, ilmu data, dan machine learning (ML). Developer menggunakan Python karena efisien dan mudah dipelajari serta dapat dijalankan di berbagai platform.

2. Manfaat Python

Manfaat Python mencakup:

- Developer dapat dengan mudah membaca dan memahami program Python karena memiliki sintaksis dasar seperti dalam bahasa Inggris.

- Python membuat developer lebih produktif karena mereka dapat menulis program Python menggunakan baris kode yang lebih sedikit dibandingkan dengan bahasa lainnya.
- Python memiliki pustaka standar besar berisi kode yang dapat digunakan kembali untuk hampir semua tugas. Sehingga, *developer* tidak perlu menulis kode dari awal.
- *Developer* dapat dengan mudah menggunakan Python dengan bahasa pemrograman populer lain seperti Java, C, dan C++.
- Komunitas aktif Python memiliki jutaan developer pendukung di seluruh dunia. Jika Anda mengalami kesulitan, Anda bisa mendapatkan dukungan cepat dari komunitas.
- Tersedia banyak sumber daya bermanfaat di internet jika Anda ingin mempelajari Python. Contohnya, Anda dapat dengan mudah menemukan video, tutorial, dokumentasi, dan panduan developer.
- Python portabel untuk beragam sistem operasi komputer seperti Windows, macOS, Linux, dan Unix.

3. Kegunaan Python

- **Matematika:** menyelesaikan permasalahan matematika seperti aljabar, kalkulus, dan trigonometri.
- **Penulisan skrip sistem:** membuat perintah secara otomatis dan menyelesaikan pekerjaan yang memakan waktu.
- **Software development:** *bug tracking* dan *testing software*.
- **Web development:** *URL routing*, memastikan keamanan *website*, memproses dan mengirim data.
- **Data analysis:** melakukan kalkulasi statistik, visualisasi data, dan menganalisis data.
- **Machine learning:** membuat algoritma untuk modul pembelajaran.

4. Fitur Python

Fitur bahasa pemrograman Python berikut menjadikan Python unik:

- **Bahasa yang diinterpretasikan**

Python adalah bahasa yang diinterpretasikan, artinya menjalankan kode secara langsung baris demi baris. Jika ada kesalahan dalam kode program, maka akan berhenti. Oleh karena itu, pemrogram dapat dengan cepat menemukan kesalahan dalam kode.

- **Bahasa yang mudah digunakan**

Python menggunakan kata-kata seperti dalam bahasa Inggris. Tidak seperti bahasa pemrograman lain, Python tidak menggunakan kurung kurawal. Python menggunakan indentasi.

- **Bahasa diketik secara dinamis**

Pemrogram tidak perlu menyatakan tipe variabel saat menulis kode karena Python menentukannya pada waktu aktif. Sehingga, Anda dapat menulis program Python dengan lebih cepat.

- **Bahasa dengan instruksi yang dapat dipahami manusia (high-level)**

Python lebih mirip dengan bahasa manusia dibandingkan bahasa pemrograman lainnya. Oleh karena itu, pemrogram tidak perlu khawatir mengenai fungsionalitasnya seperti arsitektur dan manajemen memori.

- **Bahasa berorientasi objek**

Python menganggap semua hal sebagai objek, namun juga mendukung tipe lain pemrograman seperti pemrograman terstruktur dan fungsional.

5. Kelebihan dan Kekurangan Bahasa Pemrograman Python

a) Kelebihan

1. Mudah dipelajari, Bahasa ini memiliki sintaks-sintaks yang cukup sederhana dan gampang dimengerti.
2. Mudah diaplikasikan, dalam mengembangkan sebuah produk, baik itu situs, software, aplikasi situs, maupun video game.
3. Dukungan IoT(Internet of Things)

4. Fleksible, dapat dijalankan di hampir semua system operasi (Linux, Windows, Mac OS X, dll).
5. Meningkatkan Produktivitas
6. Free dan Open Source

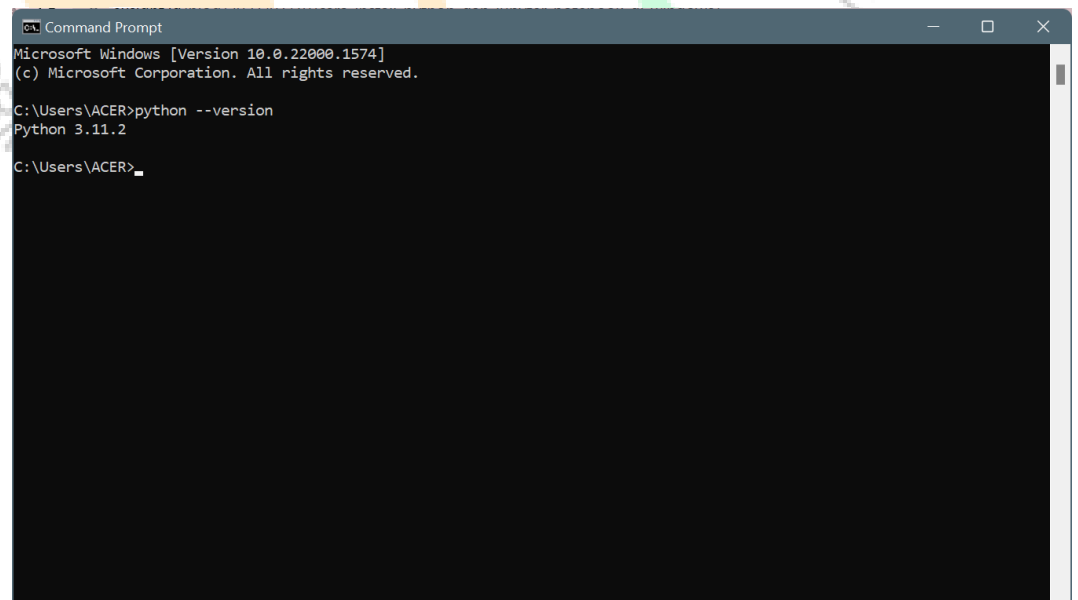
b) Kekurangan

1. Kurangnya dukungan multiprosesor
2. Lebih sedikit developers yang berpengalaman
3. Tidak ideal untuk memori intensive task
4. Kurang populer untuk mobile app development
5. Banyak Batasan desain
6. memiliki Batasan kecepatan

B. Instalasi Python dan Jupyter

1. Instalasi Python

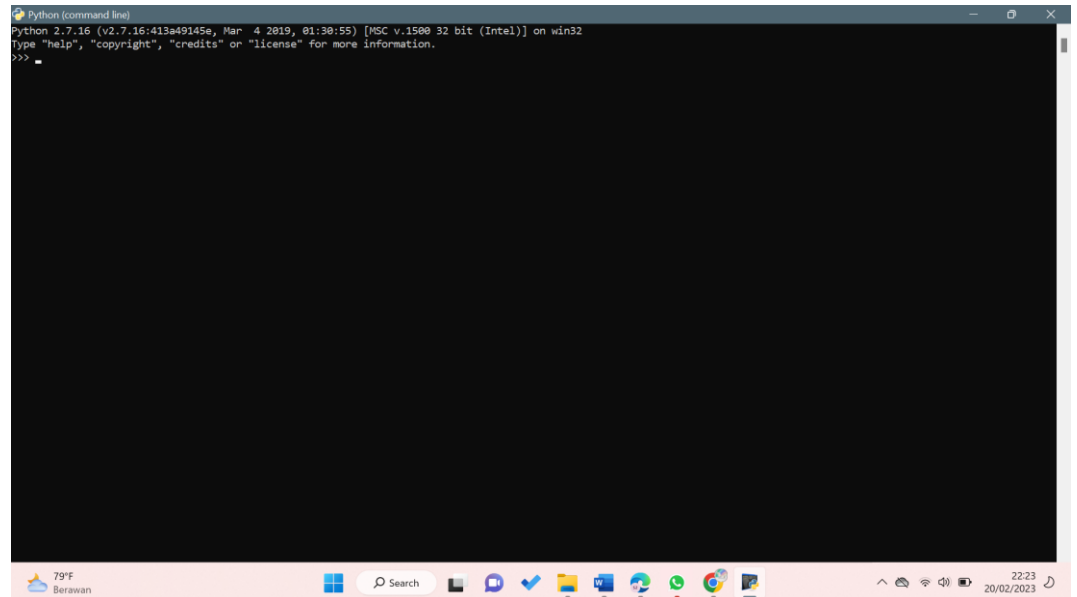
Berikut adalah hasil instalasi Python pada computer saya. Jika ingin menjalankan Python dapat dilakukan dengan cara menginstall Jupyter Notebook sebagai tool yang populer digunakan untuk melakukan pengolahan data.



```
Command Prompt
Microsoft Windows [Version 10.0.22000.1574]
(c) Microsoft Corporation. All rights reserved.

C:\Users\ACER>python --version
Python 3.11.2

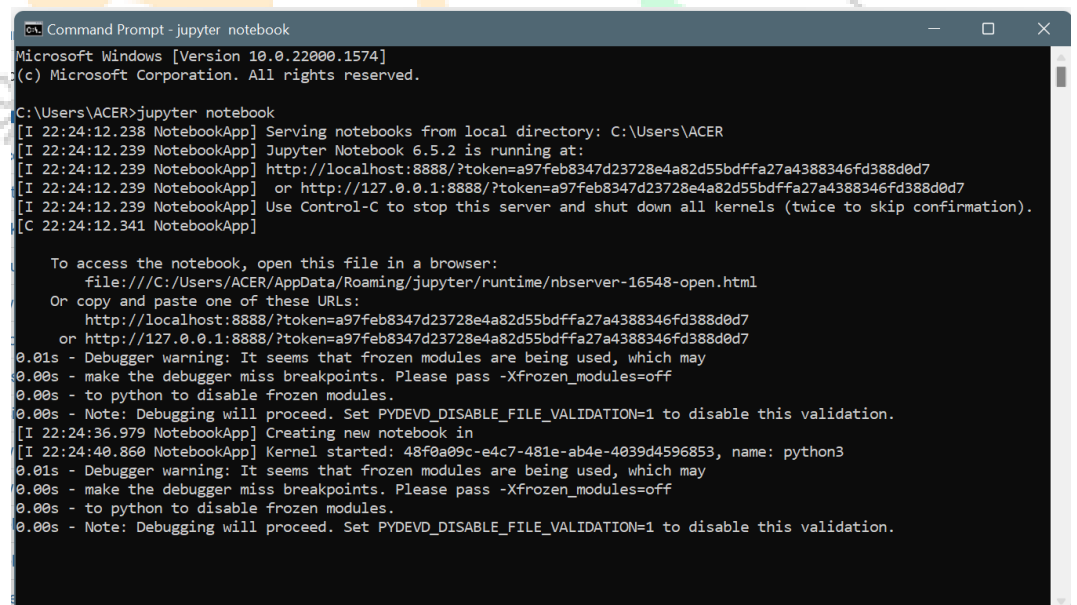
C:\Users\ACER>
```



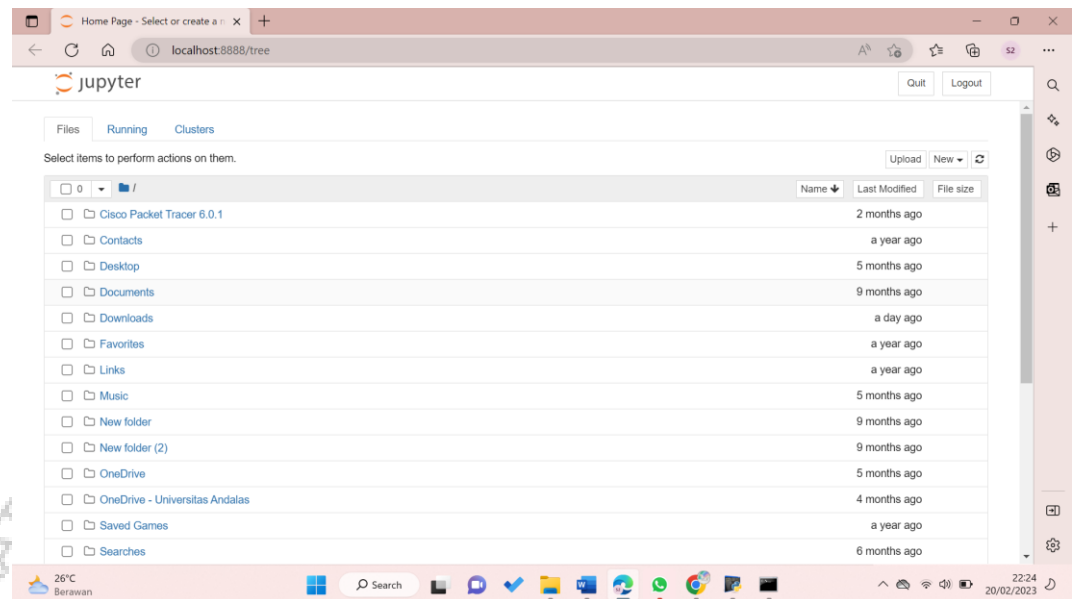
2. Instalasi Jupyter Notebook menggunakan brew (MacOS)

Berikut adalah hasil instalasi Jupyter Notebook menggunakan brew (MacOS).

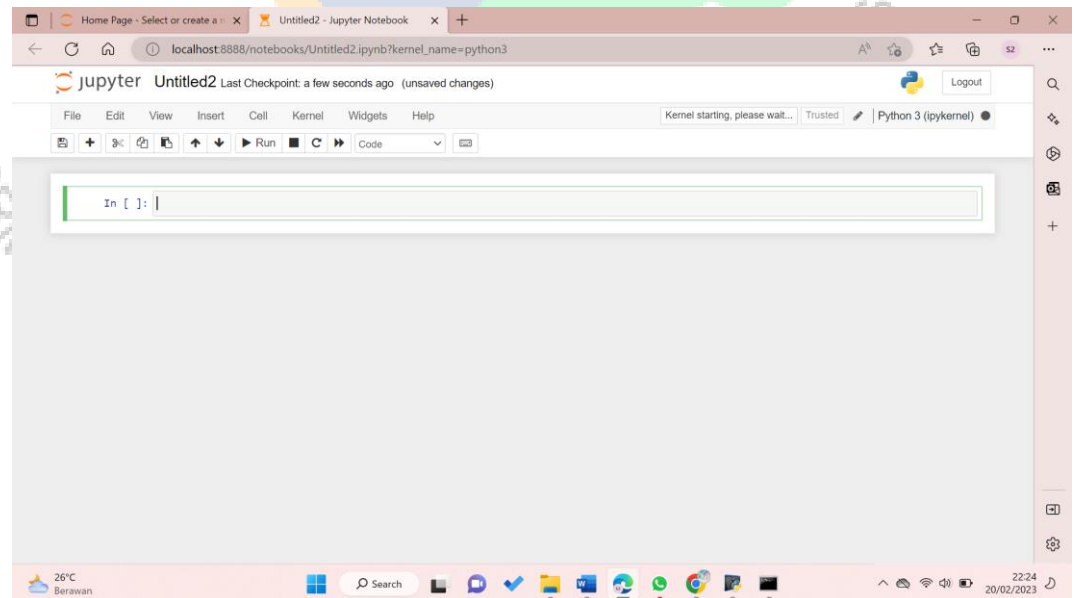
Untuk membuka hasil instalasi Jupyter Notebook adalah dengan Command Prompt dan menginputkan jupyter notebook. Maka, program akan jalan dan jupyter akan ditampilkan pada browser yang ada di computer anda.



Berikut adalah hasil jalan program dan tampilan jupyter notebook yang telah berhasil dibuka.



Selanjutnya, jika user ingin membuka file baru. Maka, user dapat klik menu new yang sudah tersedia lalu memilih Python 3. Maka, program sudah siap untuk digunakan.



C. Pengolahan Data

Caranya sebagai berikut :

1. Instalasi Pandas

2. Menampilkan Data

Selanjutnya, jika Pandas sudah terinstall. Maka, kita dapat melanjutkan dengan perintah menampilkan data.

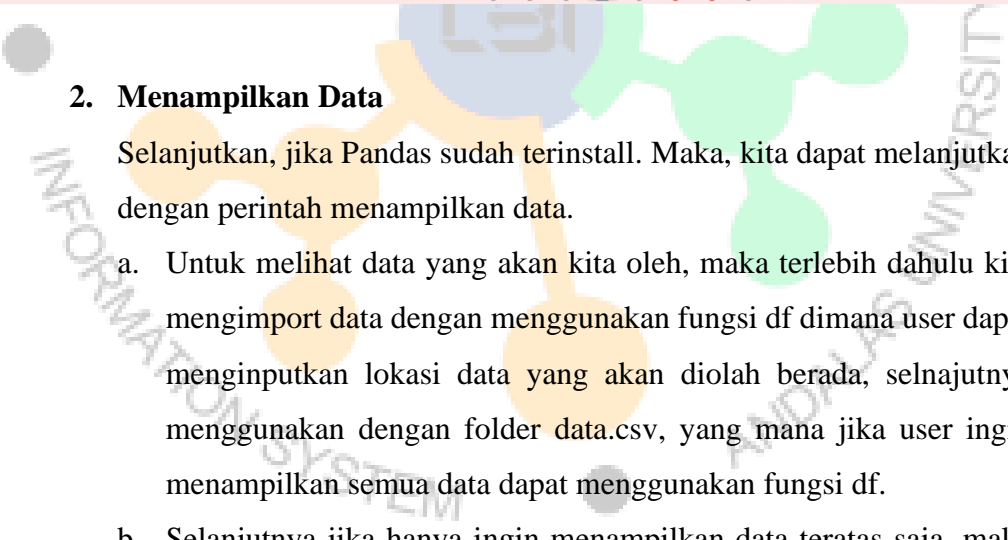
- Untuk melihat data yang akan kita olah, maka terlebih dahulu kita mengimport data dengan menggunakan fungsi `df` dimana user dapat menginputkan lokasi data yang akan diolah berada, selanjutnya menggunakan dengan folder `data.csv`, yang mana jika user ingin menampilkan semua data dapat menggunakan fungsi `df`.

Selanjutkan, jika Pandas sudah terinstall. Maka, kita dapat melanjutkan dengan perintah menampilkan data.

- Untuk melihat data yang akan kita olah, maka terlebih dahulu kita mengimport data dengan menggunakan fungsi `df` dimana user dapat menginputkan lokasi data yang akan diolah berada, selanjutnya menggunakan dengan folder `data.csv`, yang mana jika user ingin menampilkan semua data dapat menggunakan fungsi `df`.
- Selanjutnya jika hanya ingin menampilkan data teratas saja, maka kita dapat menggunakan fungsi `head`.

- a. Untuk melihat data yang akan kita olah, maka terlebih dahulu kita mengimport data dengan menggunakan fungsi `df` dimana user dapat menginputkan lokasi data yang akan diolah berada, selanjutnya menggunakan dengan folder `data.csv`, yang mana jika user ingin menampilkan semua data dapat menggunakan fungsi `df`.

b. Selanjutnya jika hanya ingin menampilkan data teratas saja, maka kita dapat menggunakan fungsi head.



The screenshot shows a Jupyter Notebook with the following code cells:

```
In [2]: import pandas as pd
In [3]: from pandas import DataFrame
In [4]: df = pd.read_csv("D:\SISTEM INFORMASI UA\21\MATKUL\SEMESTER 4\5. Data Mining\Tugas\data.csv", delimiter=';')
In [5]: df
```

The output of the last cell is a DataFrame with the following columns: nama_sentra, jenis_produk, jumlah_unit_usaha, alamat, kelurahan, kecamatan, wilayah, tenaga_kerja, nilai_investasi, kapasitasproduksi, satuan, and l. The first few rows of the DataFrame are:

	nama_sentra	jenis_produk	jumlah_unit_usaha	alamat	kelurahan	kecamatan	wilayah	tenaga_kerja	nilai_investasi	kapasitasproduksi	satuan	l
0	Jatinegara Lio	Meubel	45	Jl.Raya Bekasi Gg. Lio	Cakung	Cakung	Jakarta Timur	240.0	8.000000e+08	80000.0	kg	-6.1
1	PIK Penggilingan	Kopi/Jok	7	Jl.Raya Penggilingan	Cakung	Cakung	Jakarta Timur	105.0	3.500000e+08	NaN	kg	-6.1
2	PIK Penggilingan	Tas/Sepatu	45	Jl.PIK penggilingan	Penggilingan	Cakung	Jakarta Timur	21.0	7.000000e+07	NaN	kg	-6
3	PIK Penggilingan	Percetakan Aneka	51	Jl.PIK penggilingan Blok E Jatim	Penggilingan	Cakung	Jakarta Timur	726.0	2.420000e+09	NaN	kg	-6
4	Rawa Sari	Tahu dan Tempe	100	Jl.Pramuka Sari	Cempaka Putih	Cempaka Putih	Jakarta Pusat	300.0	1.000000e+09	100000.0	kg	-6.1
...
70	Sunter Barat	Tahu dan Tempe	136	Jl.Sunter Barat	Kemayoran	Tanjung Pahlawan	Jakarta Barat	36.0	1.200000e+08	NaN	kg	-6.1

3. Menghapus Sel Kosong

Untuk lanjut mengolah data, maka kita dapat menghapus data, baik itu yang error, data kosong, ataupun data dengan format yang salah.

- Pertama, cek data kosong dengan menggunakan fungsi DataFrame yang dilanjutkan dengansitaks `df.isnull().sum().sum()`
- Pada output terdapat 66 data yang kosong atau tidak memiliki nilai.
- Untuk itu, data yang kosong ini harus dihapuskan. Untuk menghapus fungsi yang kosong tersebut menggunakan dropna NaN. Dimana sintaks yang digunakan adalah

```
df = df.dropna()
df.isnull().sum().sum()
```

Maka data dengan sel kosong tersebut sudah dihapus dengan hasil output dari data kosong tersebut adalah 0.

- Selanjutnya, kita dapat membersihkan format dan tulisan dengan menggunakan Upper Case yang bertujuan untuk memudahkan dalam penganalisisan data.


```
75 rows x 13 columns

In [6]: df.isnull().sum().sum()
Out[6]: 66

In [7]: df = df.dropna()
df.isnull().sum().sum()
Out[7]: 0

In [8]: df.apply(lambda x: x.astype(str).str.upper())
Out[8]:
```

	jumlah_unit_usaha	alamat	kelurahan	kecamatan	wilayah	tenaga_kerja	nilai_investasi	kapasitas_produksi	satuan	latitude	longitude
0	45	JL RAYA BEKASI GG. LIO	CAKUNG	CAKUNG	JAKARTA TIMUR	240.0	800000000.0	80000.0	KG	-6.102.255	106.928.986
1	100	JL PRAMUKA SARI	CEMPAKA PUTIH	CEMPAKA PUTIH	JAKARTA PUSAT	300.0	1000000000.0	100000.0	KG	-6.190.226	106.867.767
2	25	JL.H. DAWA JALIL	CILANDAK	CILANDAK	JAKARTA SELATAN	75.0	250000000.0	25000.0	KG	-6.292.149	106.792.839
3	286	JL.TIPAR CAKUNG	SEMPER BARAT	CILINCING	JAKARTA UTARA	60.0	200000000.0	20000.0	KG	-6.121.443	106.915.848
4	80	JL.KALI BARU	KALI BARU	CILINCING	JAKARTA UTARA	558.0	1860000000.0	186000.0	KG	-6.162.279	106.881.348

4. Memilih Data

Selanjutnya, jika data sudah teratur. Maka, user dapat memilih dan memisahkan data mana saja yang dirasa perlu.

Pada, data ini saya memilih Wilayah sebagai pilihan.

- Saya mengeluarkan output yang bersangkutan dnegan wilayah dengan sintaks dari nama-nama table tersebut yang kemudian dipisahkan dari data yang tidak termasuk kedalam wilayah lalu menyisakan output yang bersangkutan dengan wilayah saja.

```
In [15]: df[['nama_sentra', 'alamat', 'kelurahan', 'wilayah', 'latitude', 'longitude']]
Out[15]:
```

	nama_sentra	alamat	kelurahan	wilayah	latitude	longitude
0	Jatinegara Lio	Jl.Raya Bekasi Gg. Lio	Cakung	Jakarta Timur	-6.102.255	106.928.986
4	Rawa Sari	Jl.Pramuka Sari	Cempaka Putih	Jakarta Pusat	-6.190.226	106.867.767
5	Cilandak	Jl.H. Dawa Jalil	Cilandak	Jakarta Selatan	-6.292.149	106.792.839
7	Semper	Jl.Tipar Cakung	Semper Barat	Jakarta Utara	-6.121.443	106.915.848
8	Kalibaru	Jl.Kali Baru	Kali Baru	Jakarta Utara	-6.162.279	106.881.348
12	Klender	Jl.Pahlawan Revolusi I	Pondok Bambu	Jakarta Timur	-6.200.747	10.691.114
17	Lenteng Agung	Jl.Sorton 100	Lenteng Agung	Jakarta Selatan	-6.315.991	106.843.231
20	Kebon Pala	Jl.Tanah Rendah Kebon Pala 1	kebon Pala	Jakarta Timur	-6.272.538	106.863.472
21	Rawa Bening	Jl.PD. Pasar Jaya St. Jatinegara Rawa Bening	Rawa Bunga	Jakarta Timur	-6.220.242	106.871.239
22	Bidara Cina	Jl. Setia	Kampung Melayu	Jakarta Timur	-6.183.657	106.814.514
23	Kampung Rawa	Jl.Kampung Rawa Selatan	Johar Baru	Jakarta Pusat	-6.180.002	10.686.013
26	Semanan	Jl.KP. Malang Kel. Semanan	kalideres	Jakarta Barat	-6.169.961	106.707.237
27	Kramat Pela	Jl.PT. Rongan I Gang II	Kebayoran Baru	Jakarta Selatan	-6.244.664	106.792.694
28	Kebayoran Lama	Jl.Gg. H. Mas'yud	Cipulir	Jakarta Selatan	-6.244.458	106.792.572
33	Kebon Kosong	Jl.Utan Panjang	Utan Panjang	Jakarta Pusat	-6.162.775	10.685.498
34	Kel.Serdano	Jl.Irian	Serdano	Jakarta Pusat	-6.1.561	106.862.823

5. Kesimpulan

Yang termasuk data wilayah adalah

(nama_sentra, alamat, kelurahan, wilayah, latitude, dan longitude)

D. Daftar Pustaka

1. <https://algorit.ma/blog/data-science/apa-itu-python-2022/>
2. <https://aws.amazon.com/id/what-is/python/>
3. <https://www.ekrut.com/media/4-kelebihan-bahasa-pemrograman-python>

