

# KATA PENGANTAR

Dengan mengucapkan puja dan puji syukur kehadirat Allah SWT, atas limpahan rahmat dan hidayah-Nya, kami dapat menyusun Modul Praktikum Mengolah Data Statistik Berbasis Excel dengan Python ini. Kami berharap modul ini dapat memotivasi dan menjadi acuan bagi pembaca yang ingin mempelajari Data Statistik menggunakan Excel dan Python.

Bahan ajar ini disusun berdasarkan pengalaman serta berpedoman pada berbagai sumber referensi, baik dari buku, tutorial, website, maupun karya ilmiah. Modul ini berisi teori, contoh studi kasus, dan penjelasan mendetail yang memudahkan pembaca dalam mempelajarinya. Dalam modul ini, kami akan membahas tentang cara mengolah data statistik berbasis Excel dengan menggunakan Python. Kami berharap, dengan contoh-contoh yang diberikan secara sederhana, pembaca dapat memahami dan mengerti materi yang disampaikan.

Kami menyampaikan permohonan maaf apabila terdapat kesalahan dalam penulisan atau penyampaian materi dalam modul ini. Kami sangat mengharapkan kritik dan saran yang membangun dari para pembaca untuk kemajuan modul ini dan perbaikan di masa mendatang. Semoga modul ini bermanfaat bagi semua yang mempelajarinya. Terima kasih.

Cirebon, Mei 2024

Tim Penyusun

# 

# DAFTAR ISI

[KATA PENGANTAR ii](#_Toc168037512)

[DAFTAR ISI iii](#_Toc168037513)

[BAB 1 PENDAHULUAN 1](#_Toc168037514)

[A. Latar Belakang 1](#_Toc168037516)

[B. Rumusan Masalah 2](#_Toc168037517)

[C. Tujuan 2](#_Toc168037518)

[BAB 2 CODE EDITOR 3](#_Toc168037519)

[A. Pengertian Code Editor 3](#_Toc168037521)

[B. Mengenal Visual Studio Code (VS Code) 3](#_Toc168037522)

[C. Mengenal Jupyter Notebook 4](#_Toc168037523)

[D. Langkah-langkah Penggunaan Jupyter Notebook dalam VS Code 4](#_Toc168037524)

[BAB 3 LIBRARY 6](#_Toc168037525)

[A. Pengertian Library 6](#_Toc168037527)

[B. Mengenal Pandas 6](#_Toc168037528)

[C. Mengenal Numpy 10](#_Toc168037529)

[D. Mengenal Matplotlib dan Seaborn 11](#_Toc168037530)

[E. Mengenal Tabulate 15](#_Toc168037531)

[F. Mengenal Openpyxl 16](#_Toc168037532)

[BAB 4 LATIHAN MENGOLAH DATA STATISTIK EXCEL DENGAN PYTHON 18](#_Toc168037533)

[A. Implementasi Library Pandas 18](#_Toc168037536)

[B. Implementasi Openpyxl dengan Pandas 30](#_Toc168037537)

[C. Fungsi yang Digunakan untuk Menambahkan, Menghapus, dan Mengupdate Data Statistik Excel 32](#_Toc168037538)

[D. Input dan Output Menambahkan Data Statistik Excel 35](#_Toc168037539)

[E. Input dan Output Menhapus Data Statistik Excel 36](#_Toc168037540)

[F. Input dan Output Mengupdate Data Statistik Excel 36](#_Toc168037541)

[BAB 5 PENUTUP 38](#_Toc168037542)

[A. Kesimpulan 38](#_Toc168037544)

[B. Saran 38](#_Toc168037545)

[DAFTAR PUSTAKA 39](#_Toc168037546)

[TENTANG PENULIS 40](#_Toc168037547)

# BAB 1

# PENDAHULUAN

## Latar Belakang

Python adalah bahasa pemrograman interpretatif yang dianggap mudah dipelajari serta berfokus pada keterbacaan kode. Dengan kata lain, Python diklaim sebagai bahasa pemrograman yang memiliki sintaksis yang sangat jelas, lengkap, dan mudah dipahami. Python mendukung berbagai paradigma pemrograman, termasuk pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Sebagai bahasa pemrograman multi-paradigma, Python dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi.

Jupyter adalah organisasi non-profit yang mengembangkan perangkat lunak interaktif dalam berbagai bahasa pemrograman. Salah satu produk utamanya adalah Jupyter Notebook, sebuah aplikasi web open-source yang memungkinkan pengguna untuk membuat dan berbagi dokumen interaktif yang berisi kode live, persamaan, visualisasi, dan teks naratif yang kaya. Ilustrasinya begini: sebelumnya, kita biasanya membagikan kode dan dokumen secara terpisah. Kode-kode disatukan dalam sebuah librari/aplikasi/proyek (seperti Visual Studio, Eclipse, dsb), sementara dokumen dibuat dengan penyunting kata. Dalam dokumen tersebut bisa ditampilkan cuplikan kode, hasil eksekusi, dan visualisasi lainnya dari program kita. Jupyter Notebook menyatukan semua ini dalam satu file interaktif, yang berisi teks/narasi, kode hidup, persamaan, hasil eksekusi, gambar statis, dan visualisasi grafis. Kelebihan lainnya adalah notebook dapat dijalankan ulang oleh siapapun yang membukanya, untuk mereproduksi eksekusi kode di dalamnya.

Pandas adalah library Python yang berfungsi untuk manipulasi dan analisis data, terutama data tabular seperti yang sering ditemui dalam spreadsheet atau relational database. Pandas, yang merupakan singkatan dari Python Data Analysis Library, adalah sebuah open source library dengan lisensi BSD yang menyediakan banyak alat untuk kebutuhan analisis data, dan manipulasi.

Excel adalah salah satu perangkat lunak pengolah data yang sangat populer dan sering digunakan untuk keperluan statistik, analisis data, dan pembuatan laporan. Dengan kemampuan yang luas dalam mengelola data tabular, Excel menyediakan berbagai fungsi dan fitur yang membantu dalam analisis data. Namun, untuk analisis data yang lebih kompleks dan pengolahan data yang lebih efisien, menggabungkan penggunaan Excel dengan kemampuan pemrograman Python dan library seperti Pandas menjadi solusi yang sangat efektif.

Dengan menggunakan Jupyter Notebook, kita dapat menyatukan seluruh proses tersebut dalam satu tempat. Kita dapat mengimpor data dari file Excel, memanipulasi dan menganalisis data tersebut menggunakan Pandas, serta mendokumentasikan seluruh proses dan hasilnya secara interaktif. Pendekatan ini tidak hanya meningkatkan efisiensi kerja, tetapi juga memudahkan reproduksi dan kolaborasi, karena seluruh alur kerja dapat disimpan dan dibagikan dalam satu notebook interaktif.

## Rumusan Masalah

* 1. Code Editor apa saja yang digunakan untuk mengolah data statistik pada Python?
  2. Library apa saja yang digunakan untuk mengolah data statistik pada Python?
  3. Bagaimana mengolah data statistik berbasis Excel dengan Python?

## Tujuan

Modul ini, diharapkan bagi pembaca dapat memahami berbagai macam rumusan masalah yang ada. Serta diharapkan bagi pembaca dapat melakukan praktikum sebagai bentuk latihan dari materi yang telah penulis sampaikan.

# BAB 2

# CODE EDITOR

## Pengertian Code Editor

Code editor adalah sebuah program komputer yang digunakan untuk menulis, mengedit, dan mengelola kode-kode untuk pengembangan perangkat lunak. Code editor biasanya dirancang untuk memberikan lingkungan yang nyaman bagi para pengembang untuk bekerja, dengan fitur-fitur seperti penyorotan sintaksis, pemeriksaan kesalahan (error checking), tata letak kode yang rapi, dan integrasi dengan alat pengembangan lainnya seperti debugger dan kontrol versi. Code Editor yang akan kami gunakan untuk mengolah data statistik pada Python ada dua yaitu Visual Studio Code (VS Code) dan Jupyter Notebook.

## Mengenal Visual Studio Code (VS Code)

Visual Studio Code (VS Code) adalah editor source code buatan Microsoft yang beroperasi pada komputer desktop dan kompatibel dengan sistem operasi Windows, macOS, dan Linux. Visual Studio Code merupakan editor kode sederhana dengan dukungan untuk operasi pengembangan seperti debugging, menjalankan tugas, dan kontrol versi. Cara menggunakan Visual Studio Code yaitu:

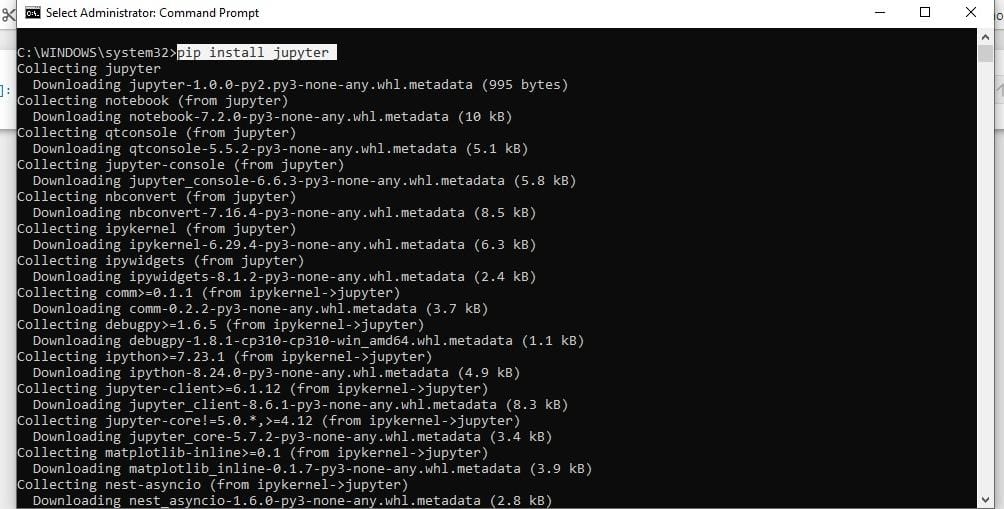
1. Download dan install VS Code untuk sistem operasi yang relevan dari website resmi, yaitu <https://code.visualstudio.com>.
2. Launch VS Code dan buka folder atau file terkait untuk memulai coding.
3. Tulis atau edit kode di area editor, misalnya dengan IntelliSense.
4. Tingkatkan fungsionalitas dengan meng-install extension yang sesuai.
5. Gunakan terminal terintegrasi untuk menjalankan dan men-debug kode.
6. Kelola version control dengan fitur Git bawaan.
7. Personalisasi VS Code dengan tema dan setting lainnya.
8. Lakukan kolaborasi dengan orang lain menggunakan fitur Live Share jika perlu.
9. Simpan perubahan dan tutup VS Code setelah selesai.

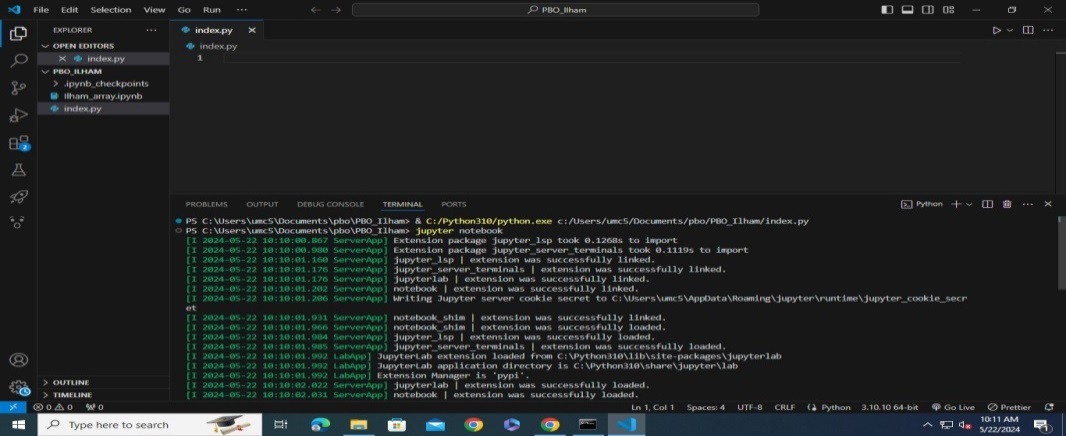
## Mengenal Jupyter Notebook

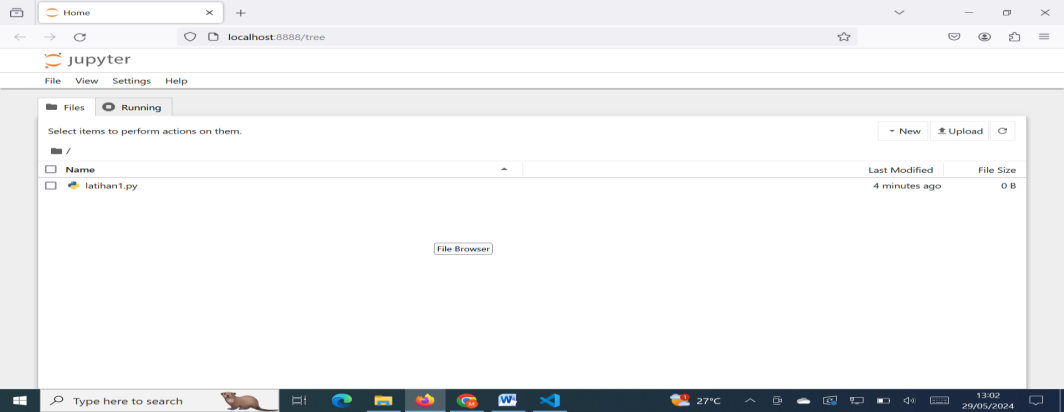
Jupyter Notebook (file yang berekstensi .ipynb) adalah dokumen interaktif yang dihasilkan oleh aplikasi Jupyter Notebook, yang berisi kode komputer dan elemen rich text seperti paragraf, persamaan matematika, gambar, dan tautan. Jupyter Notebook adalah aplikasi pengembangan interaktif berbasis web yang berjalan di browser, membuatnya mudah diakses dari berbagai perangkat tanpa perlu instalasi perangkat lunak tambahan. Dokumen Jupyter Notebook terdiri dari blok-blok kode yang disebut sel, yang dapat dieksekusi secara independen. Setiap sel dapat berisi satu atau beberapa baris kode yang bisa dijalankan terpisah dari sel lainnya, memberikan fleksibilitas tinggi untuk eksperimen, debugging, dan iterasi pengembangan. Selain kode, pengguna dapat menambahkan teks naratif menggunakan Markdown, menulis persamaan matematis dengan LaTeX, serta menyisipkan gambar dan tautan, sehingga seluruh proses analisis data, visualisasi, dan dokumentasi dapat dilakukan dalam satu tempat. Dengan fitur-fitur ini, Jupyter Notebook menjadi alat yang sangat berguna untuk pengembangan, analisis data, pembelajaran mesin, pendidikan, dan penelitian ilmiah, memungkinkan integrasi kode, visualisasi, dan dokumentasi dalam satu dokumen interaktif.

## Langkah-langkah Penggunaan Jupyter Notebook dalam VS Code

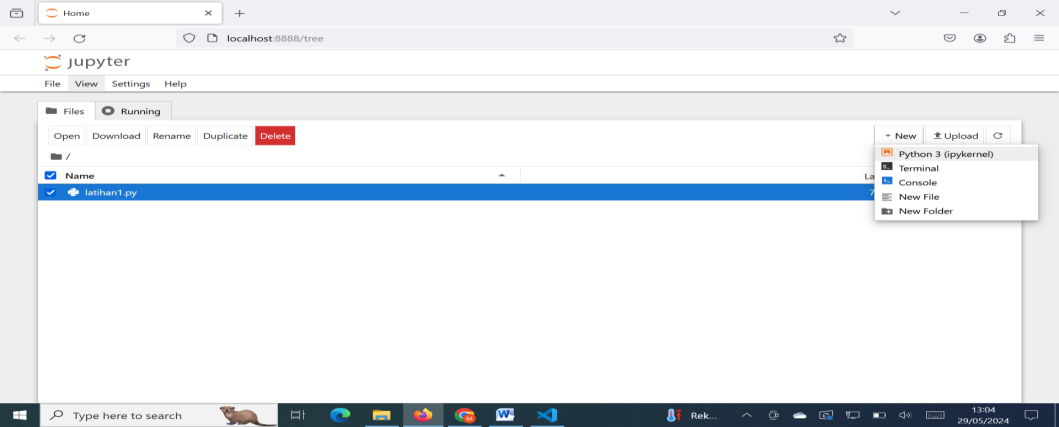
* 1. Buka CMD dan Select as administrator, ketik “pip install jupyter”

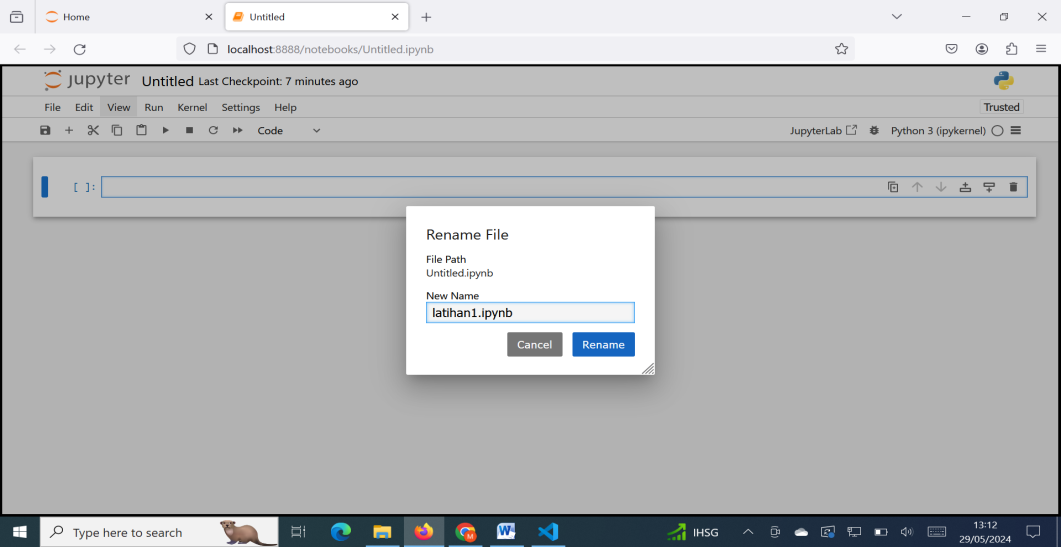


* 1. Buka terminal pada file python di vs code, lalu ketik "jupyter notebook" di terminal
  2. Lalu secara otomatis akan membuka web browser dan akanmenampilan seperti yang dibawah ini:



* 1. Klik New lalu pilih Python 3 (ipykernel)



* 1. Ubah nama File

# BAB 3

# LIBRARY

## Pengertian Library

Library dalam konteks pemrograman adalah kumpulan kode atau fungsi yang telah ditulis sebelumnya dan dapat digunakan kembali dalam pengembangan perangkat lunak. Library berisi serangkaian fungsi, prosedur, dan/atau objek yang dapat digunakan oleh program utama untuk mencapai tujuan tertentu.

Secara umum, library menyediakan berbagai fungsionalitas yang dapat digunakan oleh pengembang untuk memperluas kemampuan program tanpa harus menulis ulang kode dari awal. Ini memungkinkan pengembang untuk fokus pada logika khusus aplikasi mereka sambil memanfaatkan solusi yang sudah ada dan terbukti.

Library Python adalah kumpulan modul dan paket yang disediakan untuk bahasa pemrograman Python. Python memiliki ekosistem library yang sangat kaya dan luas, yang mencakup berbagai bidang seperti pengembangan web, analisis data, kecerdasan buatan, pemrosesan bahasa alami, pemrograman jaringan, dan banyak lagi. Beberapa library Python yang populer dan sering digunakan untuk mengolah data statistik antara lain: Pandas, Numpy, Matplotlib dan seaborn, Tabulate, Openpyxl.

## Mengenal Pandas

Pandas adalah library Python yang berfungsi untuk manipulasi dan analisis data, terutama data tabular seperti yang sering ditemui dalam spreadsheet atau relational database. Pandas merupakan singkatan dari Python Data Analysis Library. Pandas adalah sebuah open source python package/library dengan lisensi BSD yang menyediakan banyak perkakas untuk kebutuhan analisis data, manipulasi dan pembersihan data. Pandas mendukung pembacaan dan penulisan data dengan media berupa excel spreadsheet, CSV, dan SQL yang nantinya dijadikan sebagai objek python dengan rows dan columns.

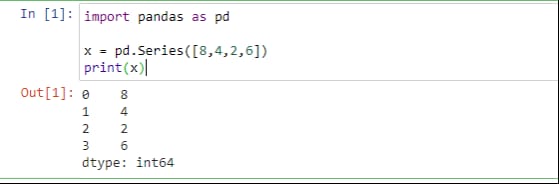
Pandas dapat digunakan bersamaan dengan library lain dalam data science. Karena dibuat menggunakan Numpy, artinya ada banyak struktur library yang digunakan atau direplika di dalam Pandas. Selain itu, data yang diproduksi oleh Pandas sering kali digunakan sebagai input plotting functions untuk Matplotlib, analisis statistik di SciPy, serta algoritma machine learning dalam Scikit-learn. Program library Pandas dalam Python sendiri dapat dijalankan menggunakan berbagai code editor, namun sangat disarankan untuk menggunakan Jupyter Notebook dan VS Code.

Objek Pandas Series dan Data Frame:

1. Series

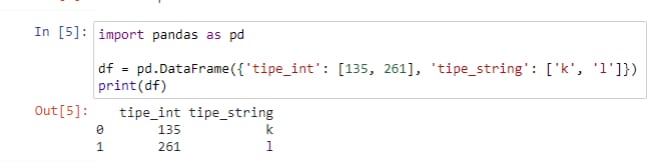
Series merupakan suatu objek satu dimensi yang dapat menyimpan berbagai jenis tipe data seperti integer, string, dan berbagai jenis tipe lainnya.

Contoh implementasi Series:



1. Data Frame

Dataframe adalah suatu objek 2 dimensi yang mana nantinya akan digunakan sebagai tempat menyimpan data dengan lebih terstruktur. Dataframe memiliki 2 indeks yaitu indeks baris dan indeks kolom.

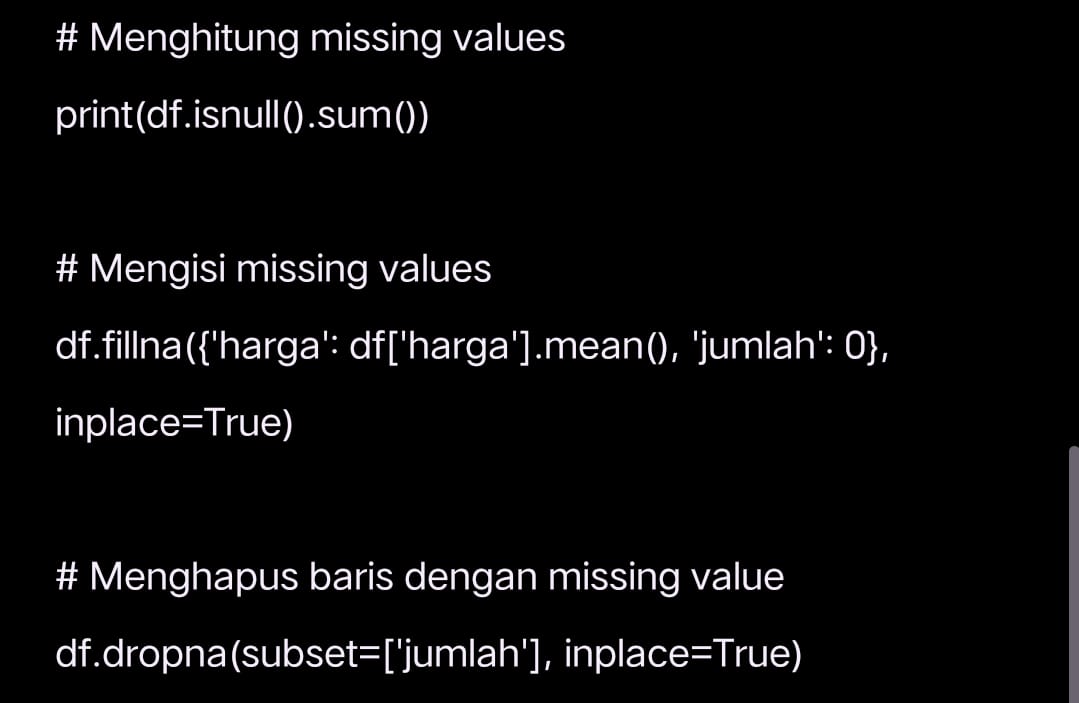


Pandas memiliki beberapa fitur utama:

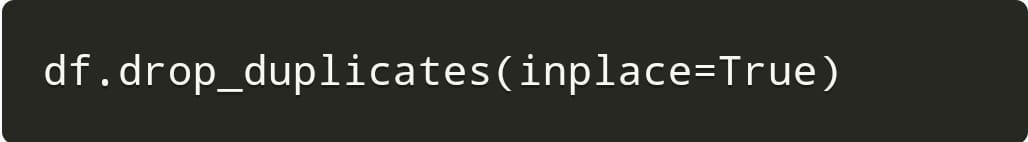
1. Struktur data, Pandas menyediakan dua struktur data utama, yaitu Series dan DataFrame. (struktur data sudah dijelaskan sebelumnya)
2. Data cleaning, Pandas menyediakan fungsi untuk mengatasi masalah ini, seperti dropna() untuk menghapus data yang hilang, fillna() untuk mengisi data yang hilang, dan drop\_duplicates() untuk menghapus data duplikat.

Fitur Utama Pandas bagian Data Cleaning yaitu:

Handling Missing Values:

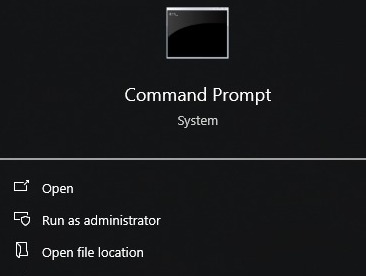


Removing Duplicate Values:

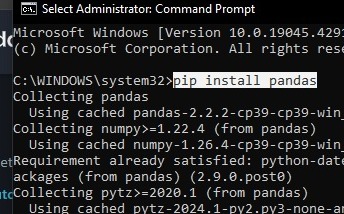


1. Transformasi data dengan Pandas, kita dapat dengan mudah mengubah bentuk data, menggabungkan beberapa set data, atau mengelompokkan data berdasarkan kriteria tertentu. Fungsi seperti pivot(), melt(), groupby(), dan merge() memudahkan proses transformasi data sesuai kebutuhan analisis.
2. Analisis data, kita bisa dengan cepat mendapatkan ringkasan statistik dari data dengan describe(), menghitung korelasi antar kolom dengan corr(), atau bahkan melakukan operasi yang lebih kompleks seperti agregasi dengan agg().
3. Kemudahan dalam impor/ekspor data merupakan salah satu kelebihan Pandas, kemampuannya untuk bekerja dengan berbagai format data. Kamu bisa dengan mudah mengimpor data dari CSV, Excel, SQL, dan banyak format lainnya menggunakan fungsi seperti read\_csv(). Begitu juga saat menyimpan data, Pandas menyediakan metode seperti to\_csv() yang memudahkan menyimpan hasil analisis.

Untuk menjalankan library pandas, langkah pertama yang paling terpenting adalah menginstall packagenya menggunakan command promt. buka command prompt dan pilih "Run as administrator."



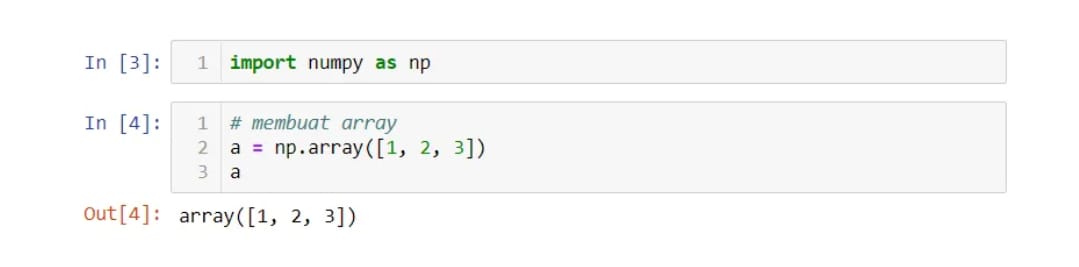
Kemudian ketik "pip install pandas" di kolom command promt. Lalu tunggu sampai proses download dan instalasi package library pandas berhasil (successfully).



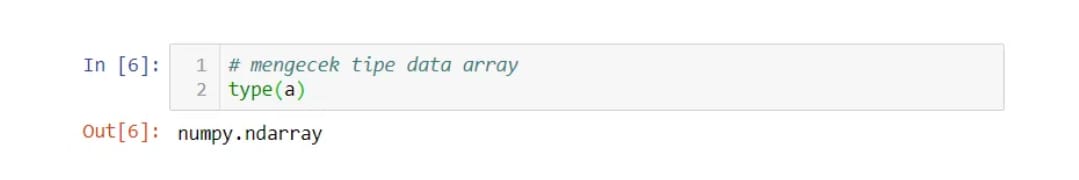
## Mengenal Numpy

NumPy (Numerical Python) adalah library Python yang fokus pada scientific computing. NumpPy memiliki kemampuan untuk membentuk objek N-dimensional array, yang mirip dengan list pada Python. Keunggulan NumPy array dibandingkan dengan list pada Python adalah konsumsi memory yang lebih kecil serta runtime yang lebih cepat.

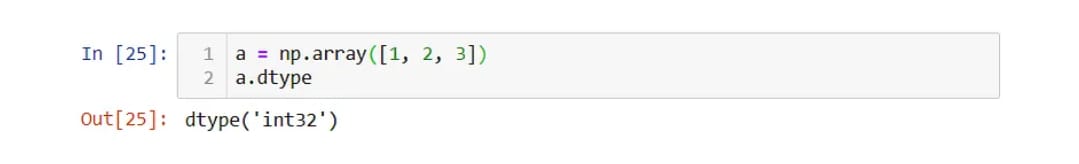
**Array pada NumPy**

Lakukan import terlebih dahulu library numpy as np. Penggunaan as disini, artinya kita menggantikan pemanggilan numpy dengan prefix np untuk proses berikutnya.

**Type Data pada NumPy**

Untuk melakukan pengecekan tipe pada array menggunakan fungsi type().

**Tipe Data Element pada NumPy**

Pengecekan tipe data element pada array menggunakan fungsi dtype

**Dimensi Array pada NumPy**

NumPy array memiliki keunggulan mendukung operasi pada data dimensional seperti Vektor dan Matriks. Untuk mengetahui jumlah dimensi pada data menggunakan fungsi ndim.

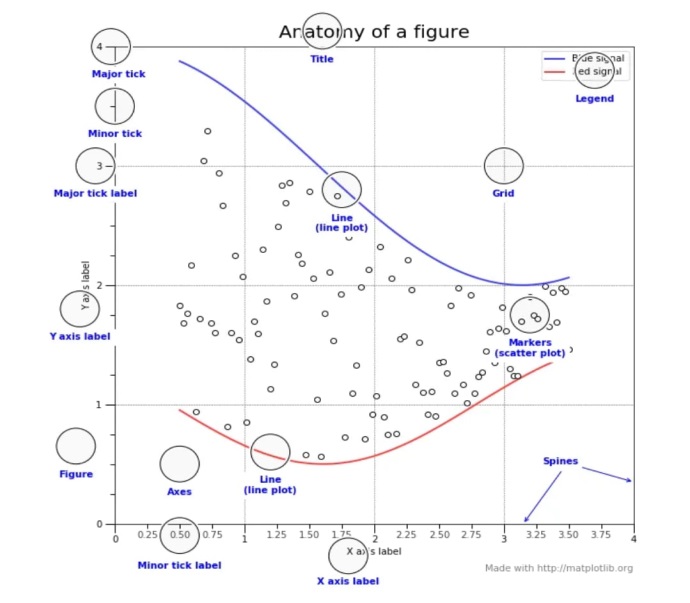
**Operasi Array pada NumPy**

NumPy memudahkan kita untuk operasi elementwise pada Vektor dan Matriks seperti penjumlahan, perkalian, pangkat, dan operasi lainnya.

## Mengenal Matplotlib dan Seaborn

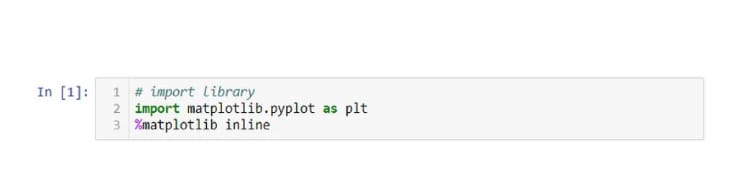
Matplotlib adalah library Python yang fokus pada visualisasi data seperti membuat plot grafik. Matplotlib pertama kali diciptakan oleh John D. Hunter dan sekarang telah dikelola oleh tim developer yang besar. Matplotlib dapat digunakan dalam skrip Python, Python dan IPython shell, server aplikasi web, dan beberapa toolkit graphical user interface (GUI) lainnya.

Gambar di bawah menunjukkan bagian-bagian dari visualisasi matplotlib dibuat oleh Nicolas P. Rougier.

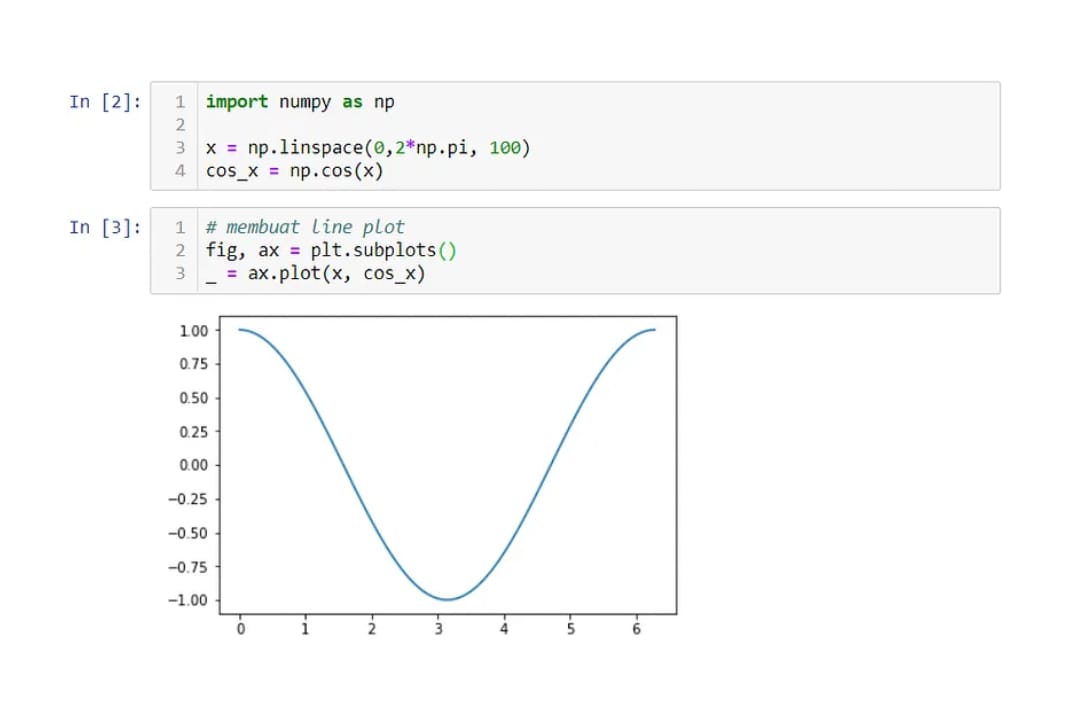


**Menggunakan matplotlib**

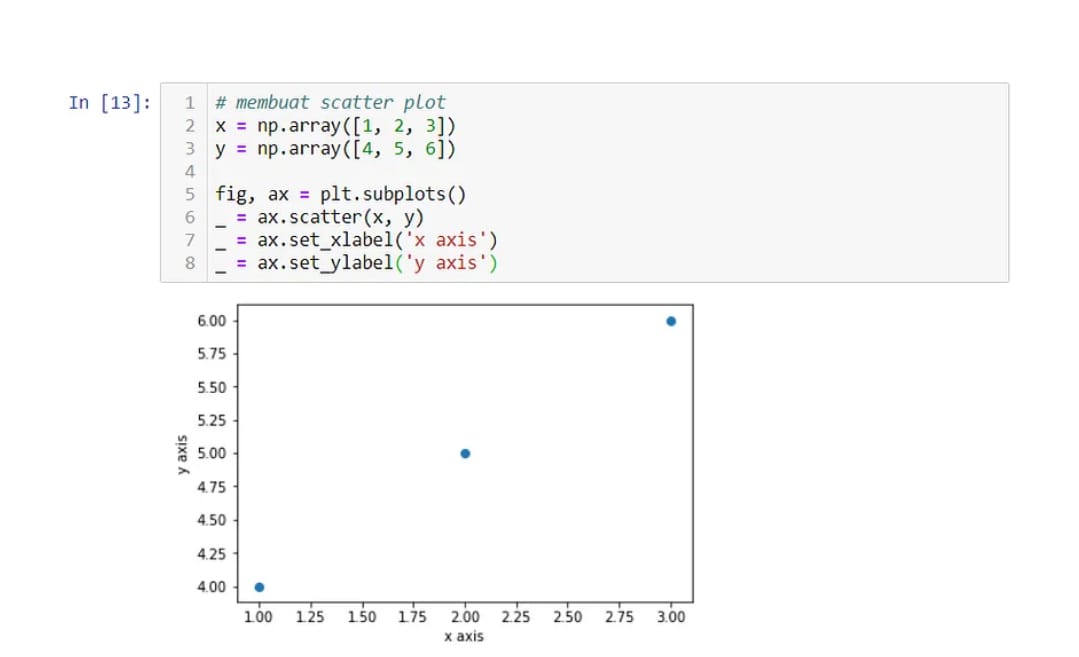
Untuk memulai menggunakan matplotlib, lakukan import terlebih dahulu library matplotlib.pyplot as plt. Penggunaan as disini, artinya kita menggantikan pemanggilan fungsi pyplot pada matplotlib dengan prefix plt untuk proses berikutnya. Disini terdapat magic command %matplotlib inline, untuk pengaturan pada backend matplotlib agar setiap grafik ditampilkan secara ‘inline’, yaitu akan ditampilkan langsung pada cell notebook.



**Membuat Line Plot**

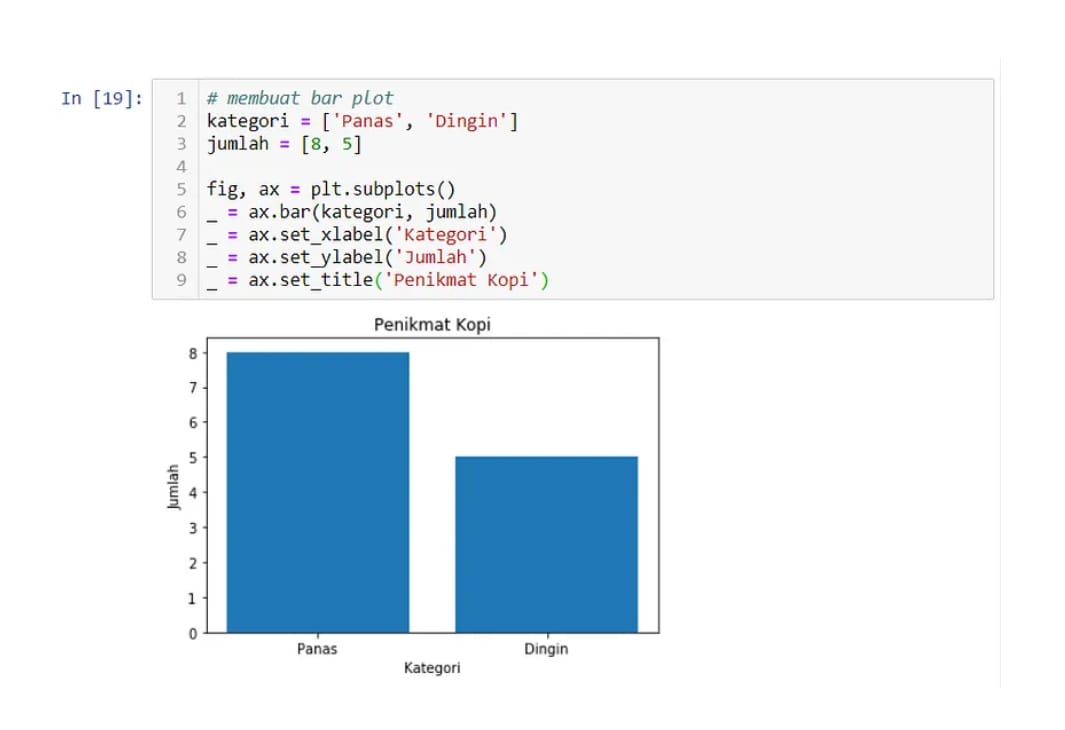
Tutorial kali ini akan membuat plot grafik line menggunakan gelombang cos. Kita akan menggunakan numpy untuk generate data gelombang cos dengan jumlah data 100 yang berjarak dari 0 sampai 2π.

**Membuat Scatter Plot**

Matplotlib menyediakan fungsi scatter() untuk mempermudah dalam visualisasi scatter plot.

**Membuat Bar Plot**

Matplotlib menyediakan fungsi bar() untuk mempermudah dalam visualisasi bar plot.



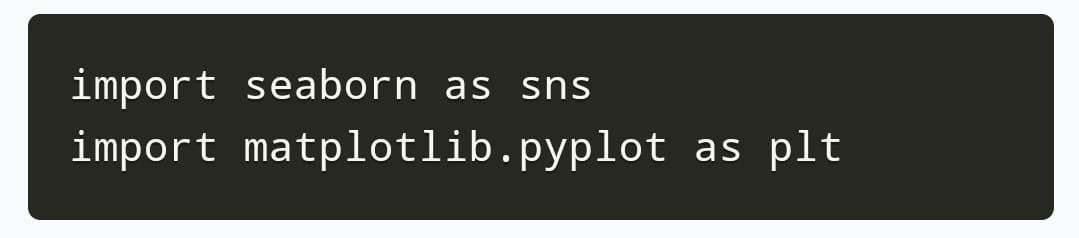
**Seaborn,** di sisi lain, adalah antarmuka tingkat tinggi untuk membuat grafik statistik. Itu dibangun di atas Matplotlib dan menyediakan antarmuka yang lebih sederhana dan intuitif untuk membuat plot statistik umum.

**Installasi Seaborn**

Setelah memastikan PIP berfungsi dengan baik, buka terminal atau command prompt dan ketik perintah berikut:

pip install seaborn

Untuk menggunakan Seaborn, kamu juga perlu mengimpor Matplotlib karena Seaborn dibangun di atasnya. Berikut cara mengimpornya:

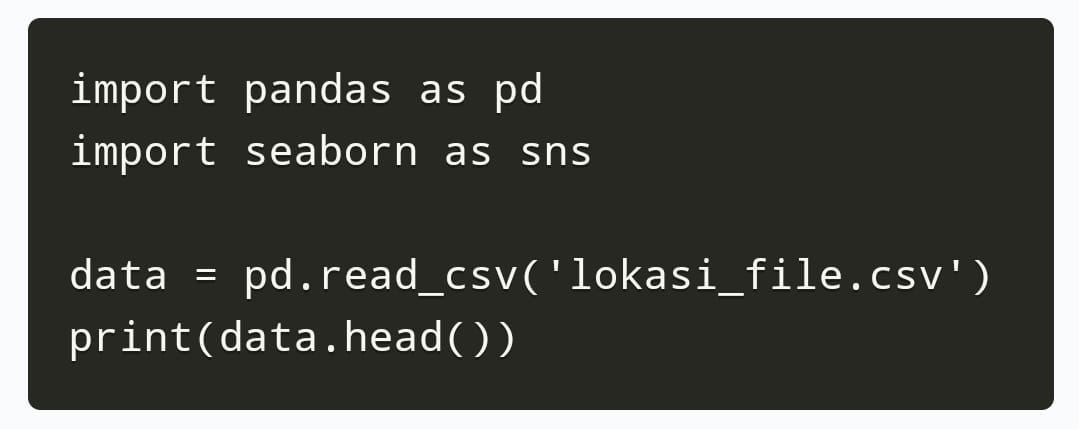


Menggunakan dataset sampel dari Seaborn

Untuk memuat dataset sampel, kamu bisa menggunakan fungsi load\_dataset(). Misalnya, untuk memuat dataset yang bernama 'tips', kamu bisa mengetik perintah berikut:

Fungsi head() dari Pandas digunakan untuk menampilkan lima baris pertama dari dataset.

Menggunakan Data yang Sudah Ada

Setelah data dimuat ke dalam DataFrame Pandas, kamu akan lebih mudah memvisualisasikannya dengan Seaborn. Berikut contoh cara memuat data dari file CSV:

## Mengenal Tabulate

Tabulate adalah perpustakaan sumber terbuka, artinya Anda dapat menggunakannya secara bebas tanpa khawatir akan plagiarisme. Atau, Tabulate adalah pembuatan tabel-tabel yang berisi data yang telah diberi kode sesuai dengan analisis yang dibutuhkan.

**Cara Menginstall Tabulate:**

Ikuti langkah-langkah berikut untuk menginstalnya:

1. Buka Terminal atau Command Prompt:

Tergantung pada sistem operasi Anda (Windows, macOS, Linux), buka terminal atau command prompt.

1. Activate Virtual Environment (Optional but Recommended):

Merupakan praktik yang baik untuk bekerja dalam lingkungan virtual untuk menghindari konflik dengan paket Python lainnya . Jika Anda tidak menggunakan lingkungan virtual, Anda dapat melewati langkah ini. Buat lingkungan virtual (ganti myenv dengan nama lingkungan pilihan Anda): python -m venv myenv

Aktifkan lingkungan virtual:

1. Di Windows : myenv\Scripts\aktifkan
2. Di macOS dan Linux: sumber myenv/bin/aktifkan
3. Instal tabulasi menggunakan pip:

Setelah Anda berada di lingkungan virtual (jika Anda menggunakannya), Anda dapat menginstal pustaka tabulasi menggunakan perintah berikut: "pip install tabulate" Ini akan mengambil versi terbaru perpustakaan tabulasi dari Indeks Paket Python (PyPI) dan menginstalnya di sistem Anda.

1. Verify Installation

Untuk memverifikasi bahwa instalasi berhasil, Anda dapat memulai sesi interaktif Python dengan mengetik python di terminal Anda. Kemudian, impor modul tabulasi dan periksa versinya: import tabulate print(tabulate.\_version\_)

Jika Anda melihat nomor versi tanpa kesalahan apa pun, berarti perpustakaan telah diinstal dengan benar.

1. Start Using ‘tabulate’

Anda sekarang siap menggunakan perpustakaan tabulasi dalam proyek Python Anda. Impor ke skrip atau buku catatan Anda menggunakan:

from tabulate import tabulate.

Anda telah berhasil menginstal pustaka tabulasi di sistem Anda dan sekarang siap membuat tabel berformat indah menggunakan Python.

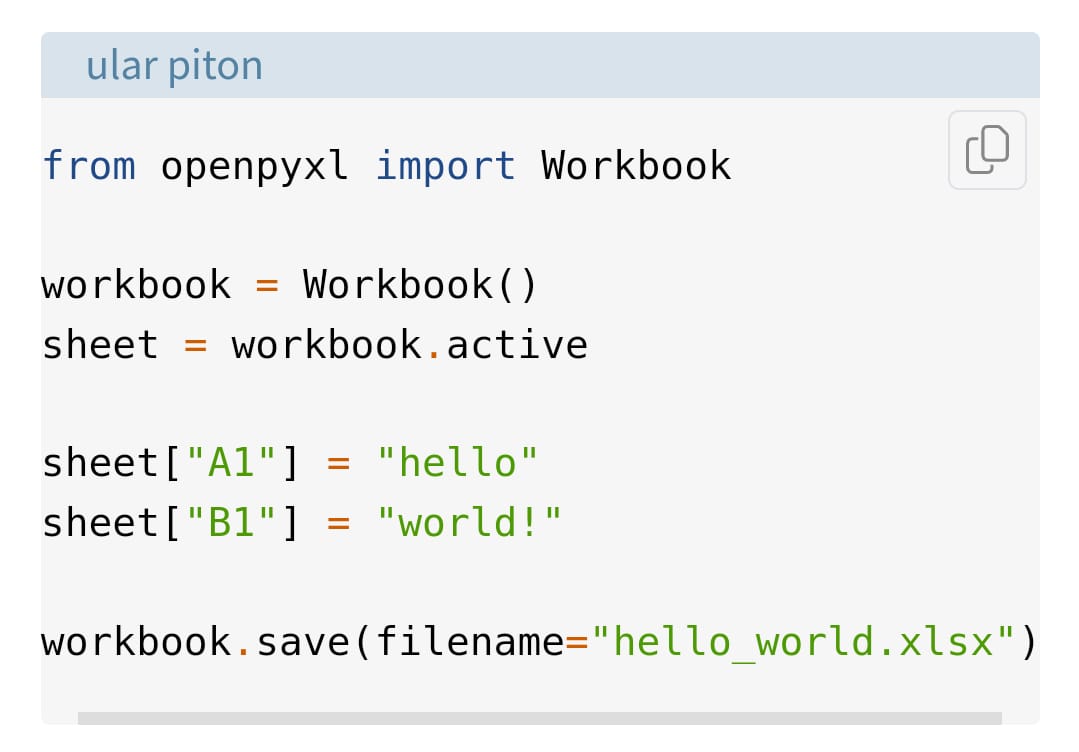
## Mengenal Openpyxl

Openpyxl adalah perpustakaan Python untuk membaca/menulis file Excel 2010 xlsx/xlsm/xltx/xltm. Openpyxl merupakan perpustakaan Python yang memungkinkan pengguna membaca file Excel dan menulis ke dalamnya.

Untuk tutorial ini, Anda harus menggunakan Python 3.7 dan openpyxl 2.6.2. Untuk menginstal paket, Anda dapat melakukan hal berikut:



Setelah Anda menginstal paketnya, Anda seharusnya dapat membuat spreadsheet super sederhana dengan kode berikut:



**Membaca Spreadsheet Excel Dengan openpyxl**

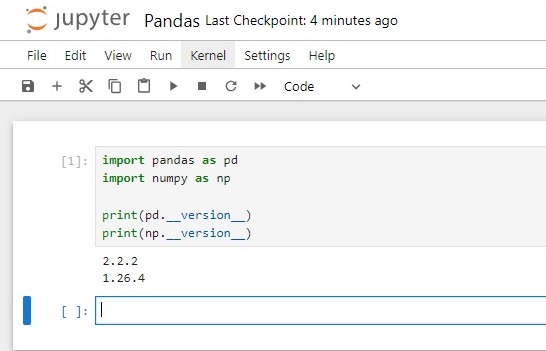
Ketika kita ingin membaca file spreadshee excel agar terbaca di openpyxl, kita harus menyimpan file dengan contoh nama : "sample.xlsx", dan letakkan di lokasi yang mudah di akses agar mudah terbaca.

# BAB 4

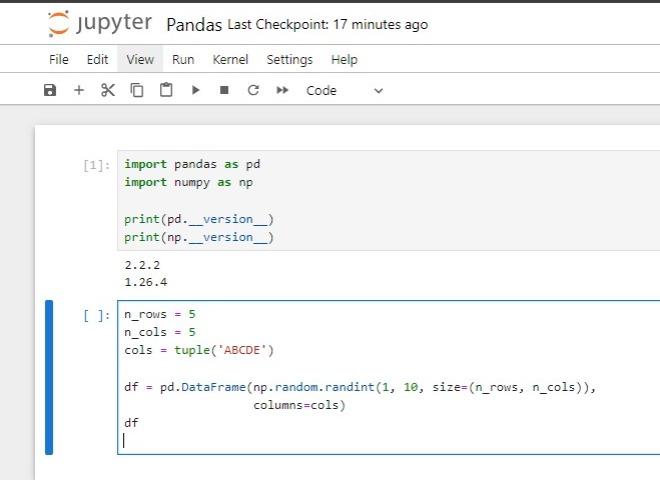
# LATIHAN MENGOLAH DATA STATISTIK

# EXCEL DENGAN PYTHON

## A. Implementasi Library Pandas



Setelah mengganti nama file di Jupyter, langkah berikutnya adalah mengimplementasikan library pandas. Pertama, pastikan pandas telah diunduh dan diinstal di perangkat Anda melalui command prompt. Setelah yakin pandas terinstal, impor library tersebut di Jupyter dengan mengetikkan `import pandas as pd`. Pastikan Anda menjalankan sel tersebut untuk memastikan tidak ada kesalahan. Anda bisa memeriksa versi pandas yang terinstal dengan menjalankan `pd.\_\_version\_\_` di Jupyter untuk memastikan bahwa instalasi berhasil.



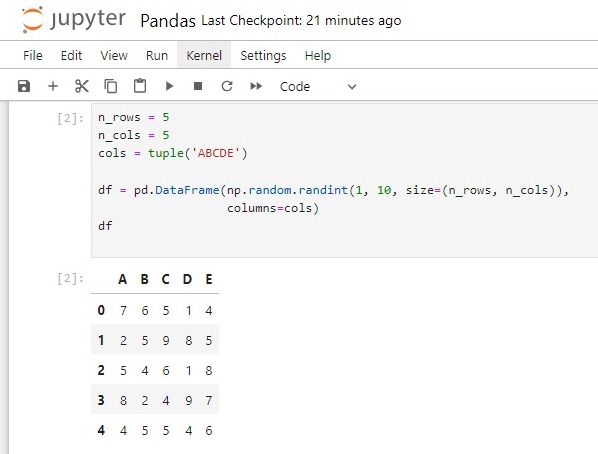
Untuk memulai pengenalan dengan library pandas, kita akan membuat data frame sederhana. Pertama, kita inisialisasikan variabel `row` dan `col`. Kemudian, kita menggunakan pendekatan variabel `row` dan `col` untuk membuat baris dan kolom. Di sini, kita menggunakan tuple 'ABCDE' sebagai inisiasi nama kolom. Yang penting, kita perlu mendefinisikan variabel dan membuat fungsi DataFrame dengan panjang kolom 5 dan panjang baris 10. Berikut langkah-langkahnya:

1. Inisialisasikan variabel `row` dan `col`.

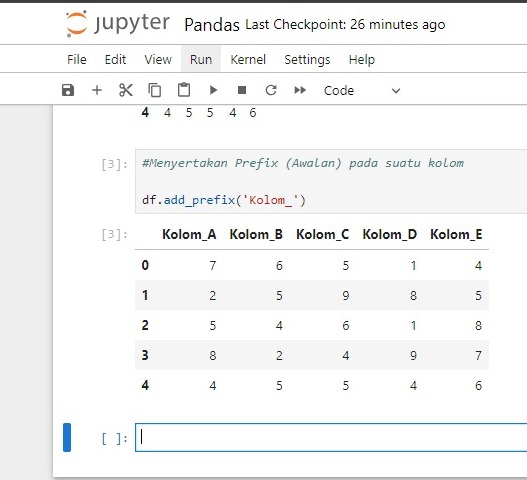
2. Gunakan variabel `row` dan `col` untuk membuat baris dan kolom.

3. Gunakan tuple 'ABCDE' sebagai inisiasi nama kolom.

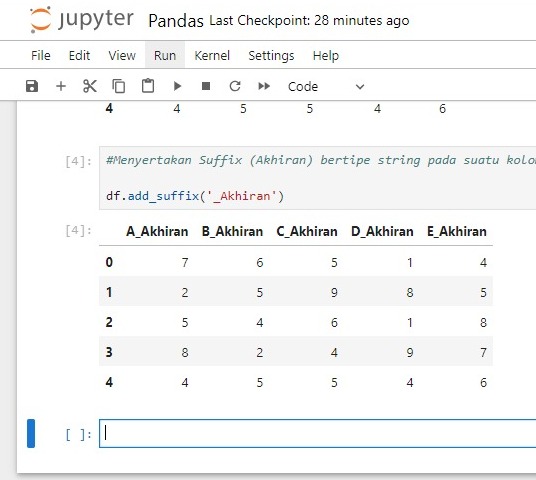
4. Buat DataFrame dengan panjang kolom 5 dan panjang baris 10.



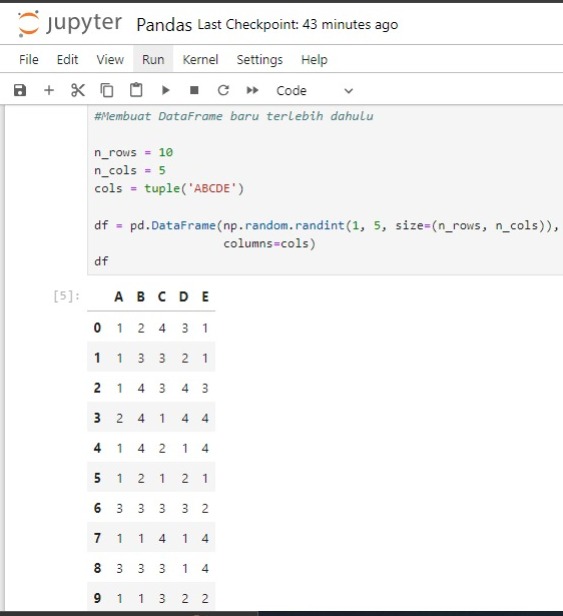
Setelah langkah-langkah sebelumnya, hasilnya akan seperti yang terlihat di gambar. Data frame telah dibuat dengan panjang kolom 5, diinisialisasi dengan A, B, C, D, E, dan panjang baris 10 dengan urutan indeks 0-9. Setiap kolom dan baris memiliki nilai pada setiap selnya. Nilai dari setiap sel dihasilkan dari perintah np.random.randint di kode sebelumnya.



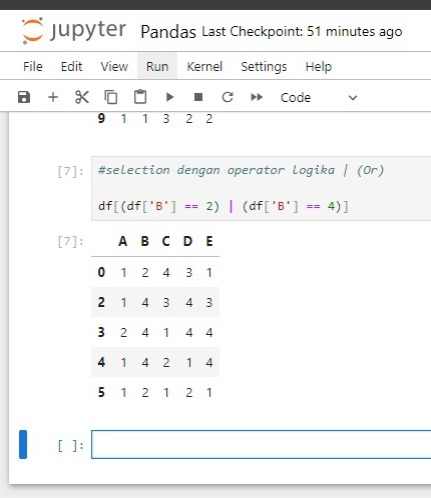
Selanjutnya, kita akan mencoba membuat prefix untuk suatu kolom dalam data frame. Prefix adalah awalan yang diberikan kepada nama kolom yang bertipe string. Prefix ini akan ditempatkan di awal inisialisasi kolom. Misalnya, jika kita awalnya membuat data frame dengan kolom bernama A, ketika kita menambahkan prefix, prefix tersebut akan ditempatkan di depan nama kolom yang sudah ada sebelumnya. Contoh pembuatan prefix pada suatu kolom dapat dilihat dalam gambar.



Selanjutnya, kita akan mencoba membuat suffix untuk suatu kolom dalam data frame. Suffix adalah akhiran yang diberikan kepada nama kolom yang bertipe string. Suffix ini akan ditempatkan di akhir inisialisasi kolom. Misalnya, jika kita awalnya membuat data frame dengan kolom bernama A, ketika kita menambahkan suffix, suffix tersebut akan ditempatkan di belakang nama kolom yang sudah ada sebelumnya. Contoh pembuatan suffix pada suatu kolom sangat sederhana dan hampir mirip dengan prefix.

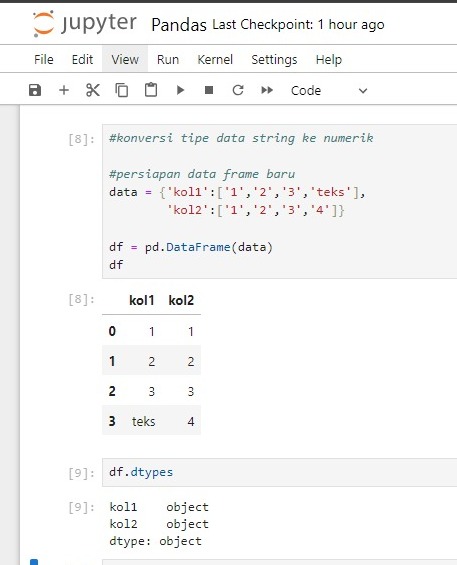


Kita akan mencoba melakukan seleksi pada library pandas, yang merupakan konsep terkait dengan logika matematika. Seleksi digunakan saat kita ingin memilih kolom atau baris dengan nilai sel tertentu yang sesuai dengan kebutuhan kita. Ini memungkinkan kita untuk memanipulasi dan menyortir data dengan cepat dan efisien. Untuk mencobanya, kita harus memiliki tabel data frame terlebih dahulu, yang telah dijelaskan sebelumnya. Hanya perlu menyesuaikan jumlah baris dan kolom sesuai kebutuhan.



Jika sudah memiliki tabel data frame, kita akan menggunakan syntax seperti pada contoh di gambar. kita menggunakan pendekatan logika or dengan lambang ( | ) di mana suatu data frame yang sudah kita definisikan sebagai df pada awal pembuatannya. di sini kita akan mencoba menyortir data frame df pada kolom B, di mana pada pada suatu cellNya memiliki value 2 (== 2) atau ( | ) value 4 (==4).

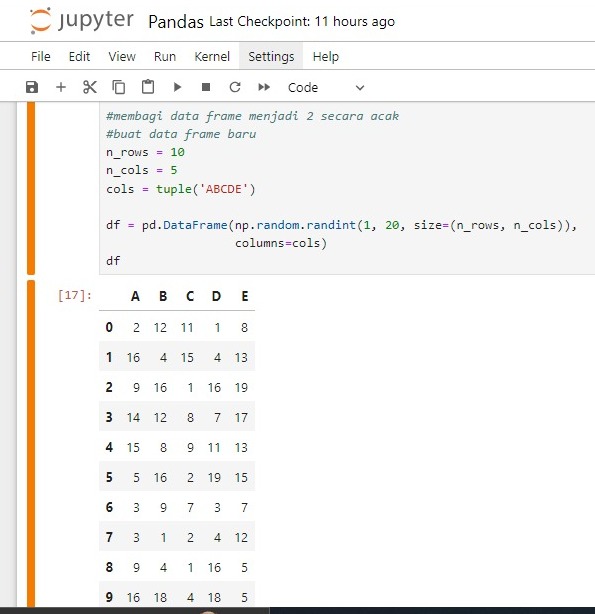
jika sudah maka hasilnya bisa dilihat seperti pada di gambar, yang sebelumnya kita memiliki banyak value pada suatu baris, kini kita memiliki baris di mana kolom yang kita tentukan hanya memuat value 2 atau 4 saja. di sini kita memilih kolom B dengan nilai 2 atau 4.



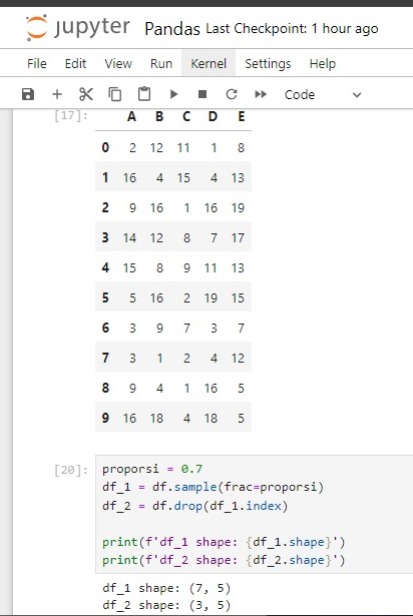
Kemudian kita akan belajar untuk mengkonversi tipe data string dari suatu cell menjadi tipe data berupa numerik. mula mula kita harus menyiapkan satu data frame terlebih dahulu. bisa diikuti pada contoh di gambar. Pada gambar yang tertera, kita telah membuat data frame dengan inisialisasi kolom yaitu kol1 dan kol2. dan kita sudah mendifinisikan tiap tiap nilai dari kol1 dan kol2. jika sudah, maka hasil tabel data frame yang dibuat bisa kita lihat. kemudian kita cek tipe data dari masing masing kolom yang ada dengan mengetik .dtypes di sini kita membuat data frame dengan inisialisasi df, jadi syntaxnya df.dtypes, pastikan sudah sesuai dengan inisialisasi variabel data frame yang anda buat. maka akan tampir tipe data dari kol1 dan kol2. pada gambar tertulis object, yang artinya value cell dari kol1 dan kol2 memiliki type data string. di sini kita akan belajar mencoba tipe data string tersebut menjadi numerik.



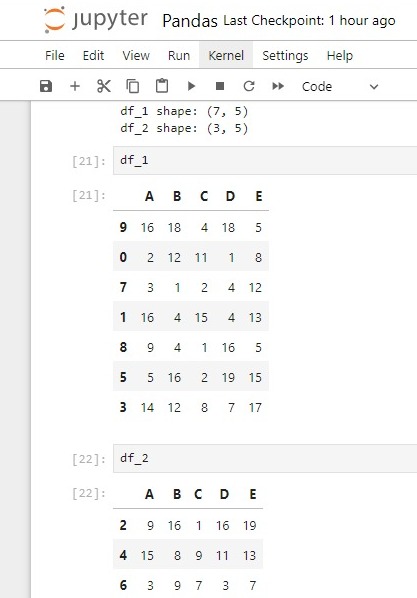
Untuk mengkonversi tipe data string ke numerik kita bisa menggunakan pendekatan astype, bisa dilihat pada gambar. Pertama kita mendifinisakan suatu variabel terlebih dahulu supaya fungsi pendekatan astype tersebut bisa kita eksekusi.Bisa dilihat pada gambar, kita akan merubah tipe data pada data frame dengan inisialisasi yang kita berikan, yakni df. Jadi kita dapat mengetikkan df.astype lalu kita bisa mengisi parameter dari astyple tersebut dengan nama kolom yang akan kita konversi tipe datanya, dan tipe data numerik apa yang akan kita berikan. di sini kita mencoba mengkoversi kol2 yang sebelumnya string, menjadi int. pada parameter astype. jika sudah bisa kita eksekusi kode tersebut. secara tampilan data frame tidak ada yang berubah, tapi ketika anda perika tipe datanya menggunakan dtypes, bisa dilihat tipe data pada kol 2 yang awalnya object (string) sekarang sudah dikonversi menjadi suatu tipe data integer (int). untuk konversi tipe numerik ini tidak hanya berlaku untuk konversi ke integer, bisa dikonversi ke tipe data float juga.



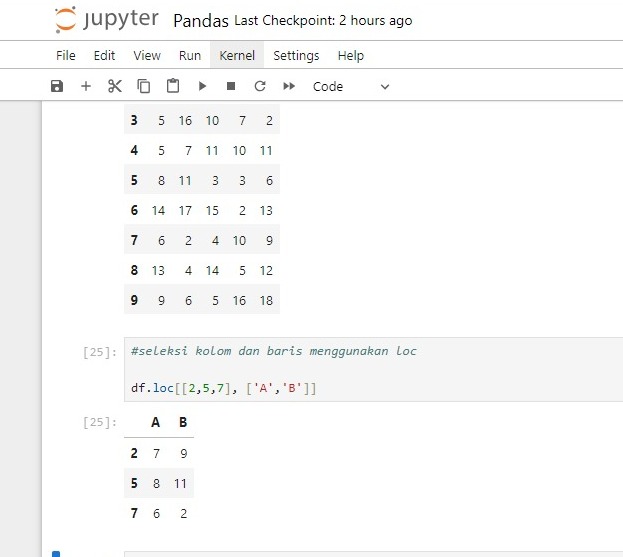
Berikutnya kita akan belajar cara membagi data frame menjadi 2 secara acak. pertama kita siapkan terlebih dahulu tabel data frame, untuk rentang kolom dan barisnya sesuaikan dengan kebutuhan kita. data yang telah kita siapkan ini nantinya akan kita bagi menjadi 2 tabel data frame secara acak.



Pertama kita cek terlebih dahulu data shape pada data frame yang akan kita bagi, pada data frame sebelumnya yang kita buat yaitu (10, 5) artinya jumlah baris kita adalah 10 dan kolomnya 5. lalu data frame df ini akan kita bagi menjadi 2 dengan proporsi 70% untuk data frame pertama, dan 30% untuk data frame kedua. caranya kita tetapkan proporsinya lalu kita masukkan ke dalam value variabel proporsi sebesar 0.7 lalu berikutnya kita panggil df.sample lalu menyertakan parmater frac (fraction=pecahan) dengan nilai proporsi, yakni 0.7 atau 70%. df.sample ini merupakan metode yang akan mengambil baris secara acak sejumlah 70% dari baris data yang dimiliki data frame df\_1 dan hasilnya akan dibentuk menjadi sebuah data frame baru yang akan kita tampung pada df\_1. Kemudian untuk mengisi df\_2 kita panggil dulu dfNya dan kita akan hapus atau df.drop berdasarkan df\_1.index, atau berdasarkan dataframe dataframe yang sudah diadopsi di df\_1. Lalu berikutnya kita print out shape dari kedua data framenya. kemudian eksekusi kodenya bisa kita lihat pada hasilnya, untuk df\_1 shapenya adalah (7,5) artinya jumlah barisnya ada 7 dan jumlah kolomnya tetap 5. Sedangkan untuk df\_2, shapenya adalah (3,5) artinya jumlah barisnya 3, dan jumlah kolomnya ada 5. dan hasil shape ini sudah sesuai dengan proporsi yang kita tetapkan yakni 0.7 / 70%. yaitu 70% untuk df\_1 dan 30% untuk df\_2.

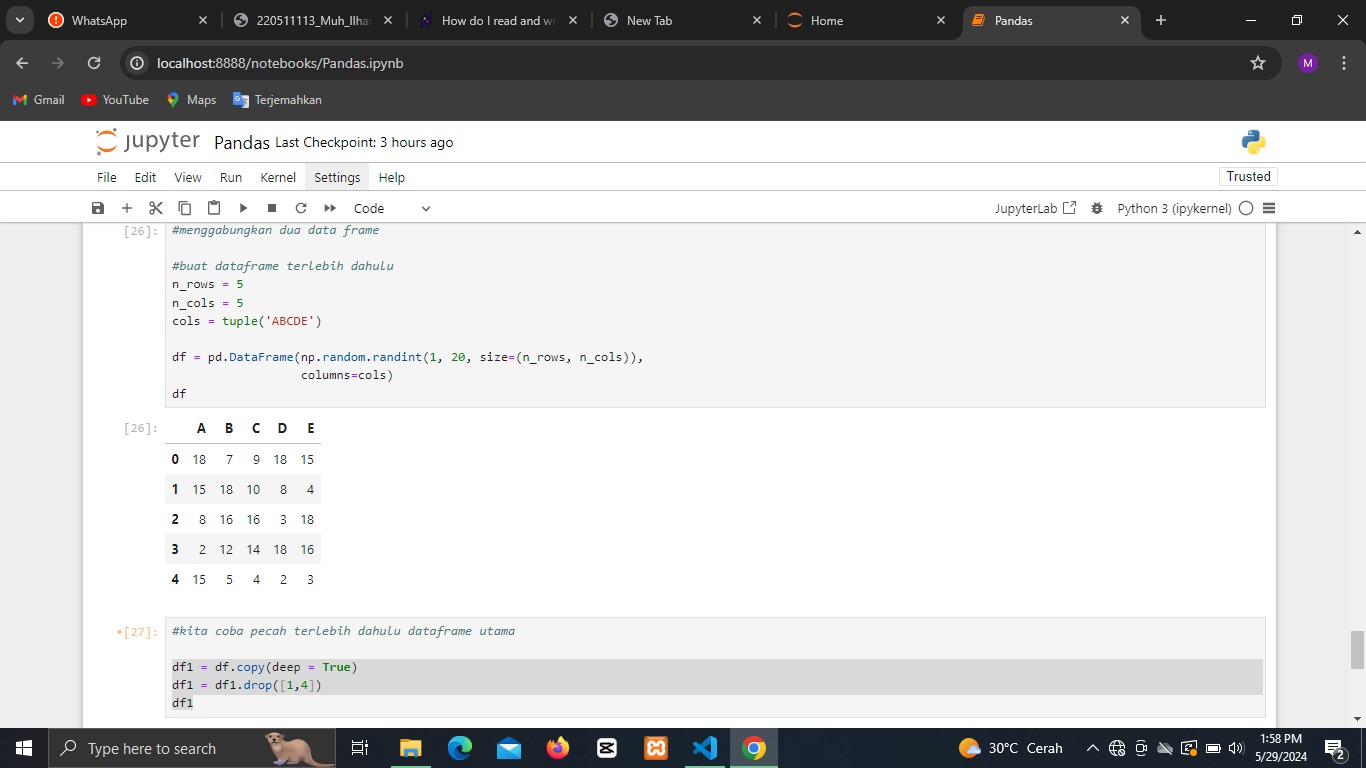


Lalu kita bisa cek data frame df\_1/df\_2 dengan cara seperti di gambar. Bisa kita lihat di sini bagaimana df\_1 maupun df\_2 terbentuk, kedua df ini terbentuk dari df sebelumnya. Yang perlu kita perhatikan adalah pada tiap tiap baris pada kedua df ini sifatnya acak. meski diadopsi dari df sebelumnya, tapi proses adopsinya yaitu acak, bukan sistematik. pada proses semacam ini seringkali dibutuhkan pada beberapa teknik dalam bidang data science maupun machine learning.

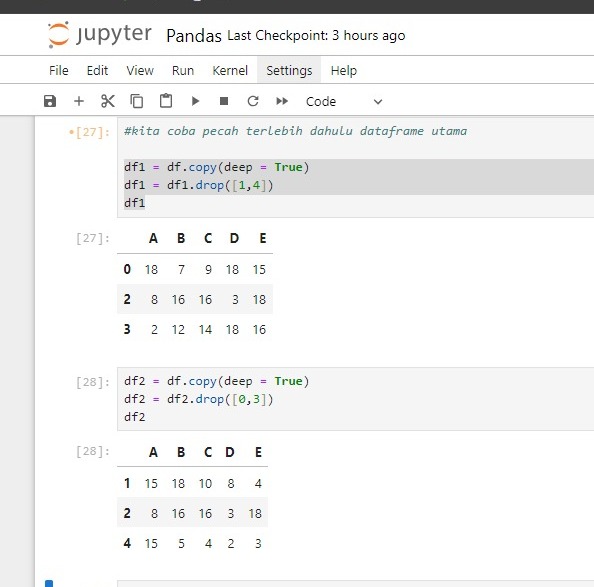


Berikutnya kita akan belajar menyeleksi baris dan kolom suatu data frame, pertam kita siapkan tabel data frame dengan rentang sesuai kebutuhan, kemudian untuk menyeleksi beberapa baris dan kolom kita bisa menggunakan pendekatan "loc".

Pertama kita panggil terlebih dahulu data frame kita, sebelumnya kita menginisialisasikan dataframe kita sebagai df, maka bisa kita ketikkan df.loc lalu diisi dengan parameter yang memuat baris, kemudian kolomnya. Pada gambar, kita akan menyeleksi baris 2,7,9 dan kolom A dan B. Maka hasilnya akan tampil data frame baru yang kita seleksi tersebut.

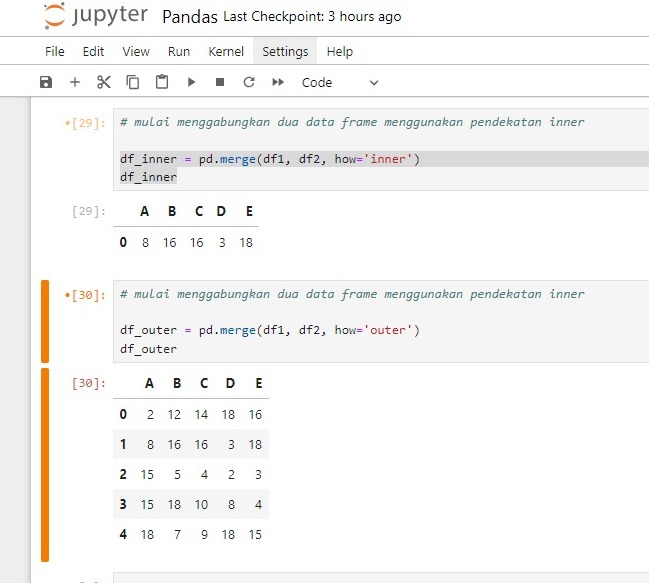


Sebelumnya kita telah belajar cara membagi suatu tabel data frame menjadi 2, kali ini kita aka belajar menggabungkan 2 data frame menjadi satu dengan pendekatan inner dan outer. pertama kita siapkan terlebih dahulu satu data frame, kemudian untuk melakukan percobaan penggabungan data frame. kita harus memecah data frame tersebut menjadi 2 terlebih dahulu. untuk cara memecah data framenya bisa dilihat pada contoh gambar, kita akan memecah menjadi df1 dan df2.



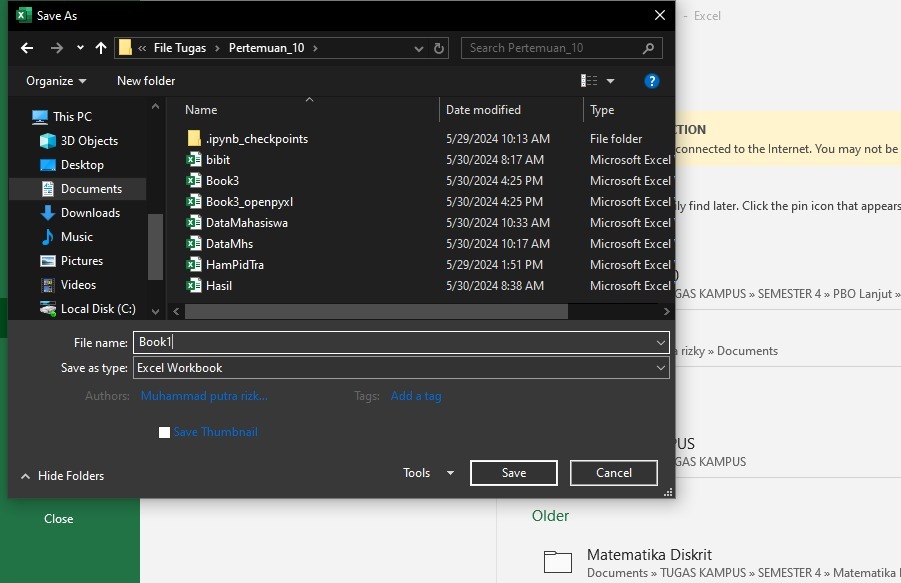
Pada df drop di df1 dan df2, artinya kita akan memecah df1/df2 tanpa menyisipkan baris maupun kolom yang nilainya dimasukan pada

parameter df.drop . bisa di lihat pada gambar.

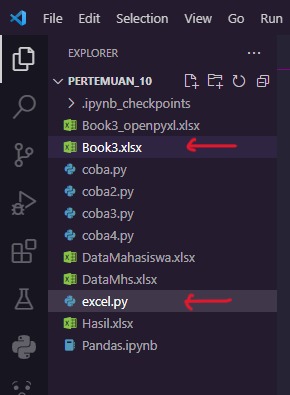


Terakhir kita akan mencoba menggabungkan kedua data frame yang sebelumnya telah kita bagi menjadi 2 menggunakan pendekatan pd.merge yang memiliki parameter how dan bernilai inner mauput outer. sebelum parameter how, kita akan menuliskan terlebih dahulu dataframe mana yang akan kita gabungkan.

Sebelum menjalankan fungsi pd.merge, kita definiskan suatu variabel yang akan memuat pendekatan merge tersebut terlbih dahulu. Untuk parameter how bernilai "inner" akan melakukan penggabungan antara data frame 1 dan data frame pada suatu baris dan kolomnya yang memiliki nilai cell yang sama. bisa dilihat pada gambar, berbeda dengan parameter how bernilai "outer", dia akan melakukan penggabungan secara keselurhan dari df1 dan df2, tapi jika ada 2 baris kolom yang memiliki nilai cell serupa, akan di eliminasi menjadi 1 elemen baris kode saja. jadi pendekatan ini memungkinkan pencegahan pada baris kolom yang memiliki nilai yang sama antar cellNya.

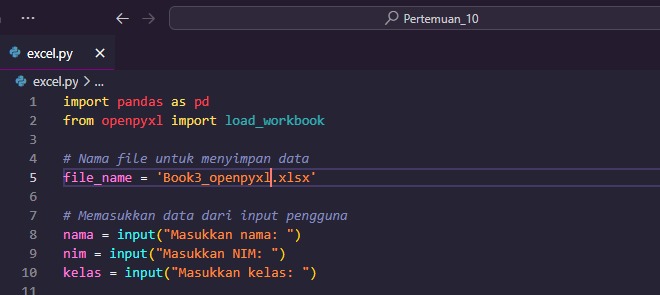


Pada tahap berikutnya kita akan belajar cara memanipulasi data excel menggunakan python x pandas, untuk code editor kita bisa menggunakan visual studio code maupun jupyter notebook. langkah pertama ialah kita buat dahulu file excel kosong yang tidak ada data di dalamnya, karna kita akan manipulasi data excel menggunakan python di vs code.

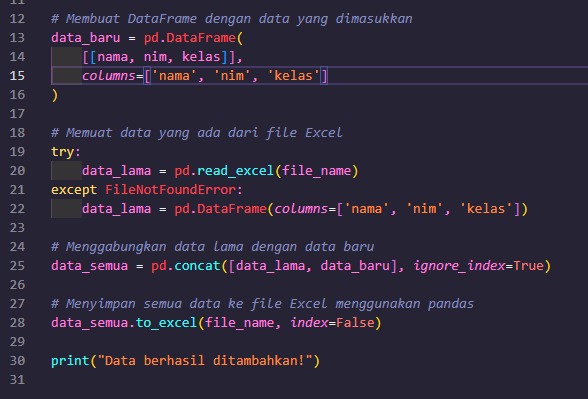


Pastikan file excel yang dibuat berada dalam satu direktori atau satu folder yang sama dengan file python kita, agar kita tidak bingung mencari /path file excel ketika hendak memanggilnya di vs code.

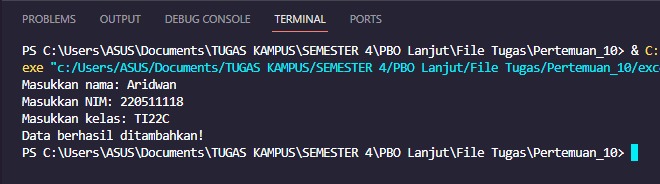
## B. Implementasi Openpyxl dengan Pandas



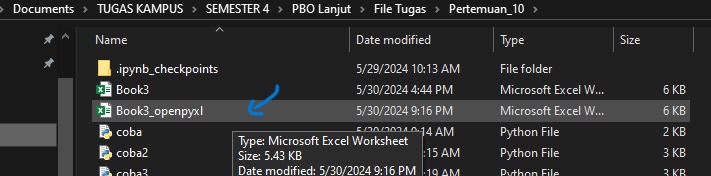
Lalu selanjutnya kita buka file python, kemudian seperti pada penjelasan di awal modul ini, kita harus mengimport library pandas terlebih dahulu dan openpyxl agar excel bisa memberikan hak akses untuk memanipulasi data di python. Selanjutnya kita akan membuat variabel untuk menyimpan nilai dari file excel kosong yang kita buat. bisa dilihat pada gambar, kita kali ini akan mencoba menambahkan data ke excel menggunakan input terminal python di vs code.



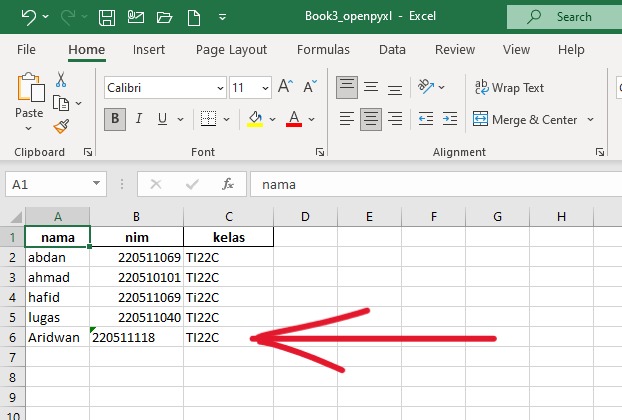
Berikutnya kita akan membuat sebuah variable yang akan memuat data frame, karena kita akan membuat data frame excel di python. untuk kolom kolomnya bisa dilihat pada gambar. lalu kita akan memanggil fungsi concat untuk digunakan dalam konteks program Anda untuk menambah data baru tanpa menimpa data lama. Lalu kita buat variable yang akan memuat perintah to.excel, perintah ini berfungsi untuk menyimpan data frame yang kita buat sebelumnya di file excel, bisa di lihat pada gambar.



Kemudian jika sudah mempraktikan seperti pada langkah langkah di gambar, anda bisa langsung menjalankan program tersebut. lalu masukkan data sesuai dengan inputannya.

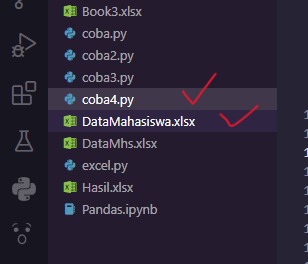


Jika data berhasil tersimpan, anda bisa membuka file excel yang sebelumnya sudah dibuat dan sudah dipindahkan menjadi satu direktori atau satu folder dengan file python kita.

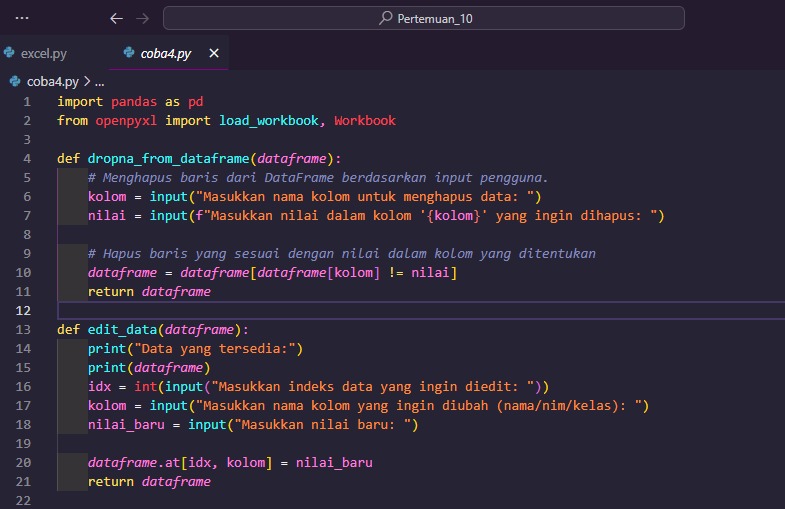


Kemudian anda bisa melihat, kita telah berhasil membuat kolom header dan memasukan data baru di excel menggunakan python.

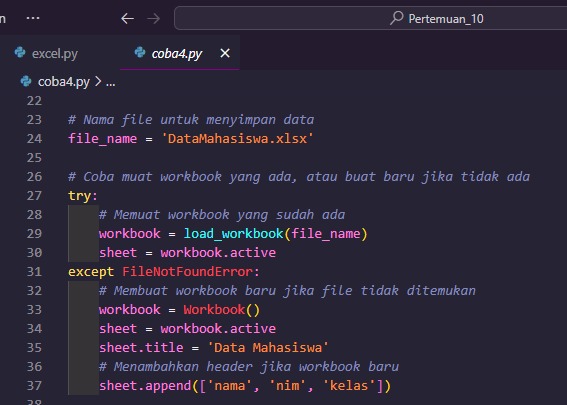
## C. Fungsi yang Digunakan untuk Menambahkan, Menghapus, dan Mengupdate Data Statistik Excel



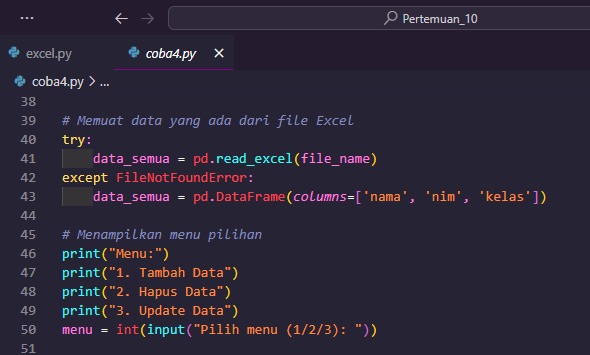
Sebelumnya kita telah belajar Menambah data pada file excel menggunakan python di Vs Code, kali ini kita akan mencoba belajar cara bagaimana agar kita bisa menghapus dan mengedit data pada file excel yang sudah ada. Pertama seperti biasa kita buat 1 file excel kosong dan 1 file python. pastikan kedua file ini berada pada folder atau direktori yang sama. contoh bisa dilihat pada gambar.



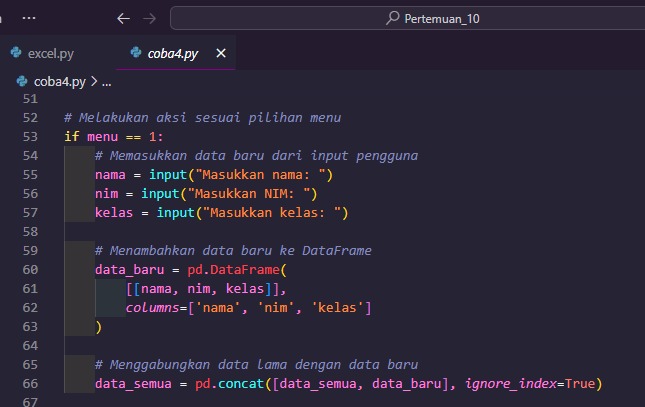
Pertama tama kita import terlebih dahulu library pandas dan openpyxl, karna di sini kita akan mengimplementasikan manipulasi data excel menggunakan python. untuk lebih jelas, perhatikan langkah pada gambar. Kita akan membuat fungsi dropna dengan parameter dataframe, fungsi dropna bertujuan untuk menghapus baris dari DataFrame berdasarkan nilai dalam kolom yang dimasukkan oleh pengguna. Lalu kita akan membuat variabel input kolom dan nilai, agar pengguna dapat menginput mana kolom dan nilai yang akan dihapus. Kemudian buat variable dataframe untuk memilih dan menghapus baris yang kolomnya mengandung nilai tertentu. Selanjutnya kita akan membuta fungsi edit\_data, fungsi ini bertujuan untuk mengedit data yang sudah ada dalam DataFrame. Lalu kita buat variabel input yang nantinya akan digunakan oleh user ketika ingin menentukan mana data indeks, kolom, dan nilai baru yang akan diganti.



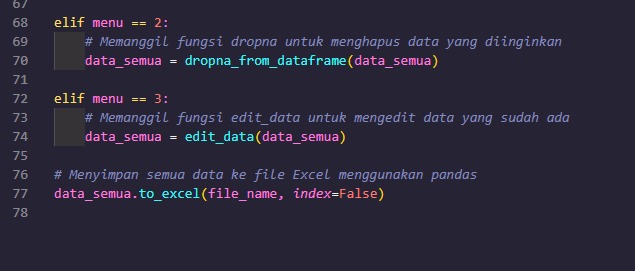
Lalu kita akan inisialisasikan variabel dengan nilainya yaitu nama file excel kita. pastika file excel kosong yang kita buat di awal berada di satu direktori yang sama dengan file python. dalam contoh di gambar, kita membuat exception handling, di mana kita akan try memuat file workbook yang ada dengan load\_workbook, dan jika workbook tidak ada, maka workbook baru pada excel akan dibuat dan diberi judul serta header untuk kolom data.



Lalu kita mencoba membuat try blok, agar memuat data dari excel ke dalam DataFrame "data\_semua", jika file tidak ditemukan pada excel, DataFrame kosong dengan kolom 'nama', 'nim', dan 'kelas' dibuat. itu sebabnya kita membuat file excel kosong di awal. Kemudian kita akan membuat menu pilihan yang akan ditampilkan kepada pengguna. variabel menu mengambil input pilihan dari pengguna untuk menentukan aksi yang akan dilakukan. apakah si pengguna ingin melakukan aksi tambah data, hapus data, maupun edit data excel.

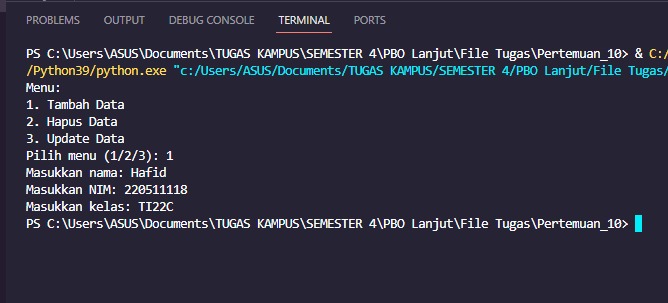


Kemudian kita membuat kondisi, jika pengguna memilih menu 1, maka akan melakukan aksi tambah data. proses ini melibatkan pengambilan data dari input pengguna pada 'nama','nim','kelas'. kemudian membuat DataFrame di data\_baru, dengan kata yang diinputkan oleh pengguna. lalu menggunakan fungsi concat untuk menggabungkan data\_baru dan data\_semua. pendekatan concat ini juga memungkinkan ketika user menambah data baru, data baru akan terUpdate dan tidak akan menimpa atau menghilangkan data lama.

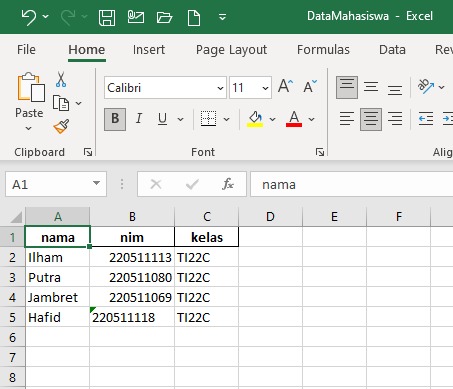


Lalu kita membuat kondisi kedua, jika pengguna memilih opsi 2 (Hapus Data), maka akan memanggil dan mengeksekusi fungsi dropna\_from\_dataframe yang sebelumnya dibuat, yang merupakan fungsi untuk menghapus data excel yang diinginkan pengguna. Selanjutnya kondisi ketiga, jika pengguna memilih opsi 3 (Edit Data), maka akan memanggil dan mengeksekusi fungsi edit\_data, yang sebelumnya dibuat, yang fungsi tersebut merupakan fungsi untuk mengedit data excel yang diinginkan pengguna. Kemudian terakhir kita menggunakan pendekatan to.excel untuk menyimpan DataFrame ("data\_semua") yang telah diperbarui oleh pengguna ke file excel dengan nama "file\_name"

## D. I nput dan Output Menambahkan Data Statistik Excel

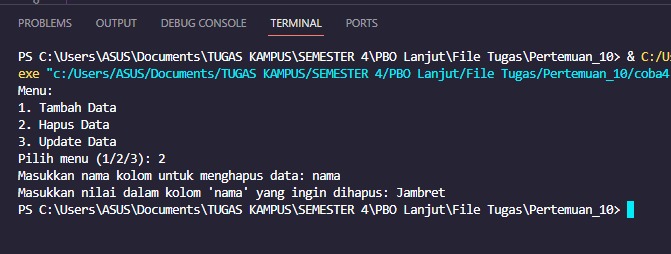


Kali ini kita akan melakukan uji coba pada program yang kita buat, pertama kita akan mencoba menambah data baru. Jika sudah melakukan input data baru, maka kita cek dengan membuka file excel yang kita buat.

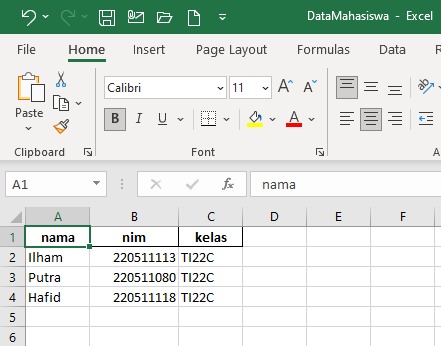


Bisa kita lihat, data yang kita input Nama "Hafid" di terminal, sekarang sudah bertambah dan masuk ke file excel.

## E. Input dan Output Menhapus Data Statistik Excel

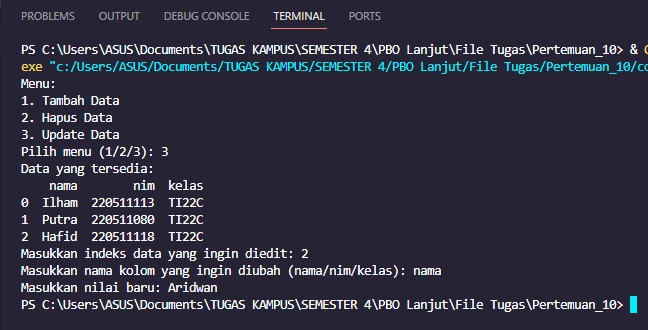


Setelah kita berhasil menginput data dan sudah masuk ke file excel, sekarang kita akan mencoba menghapus data yang kita inginkan di file excel.

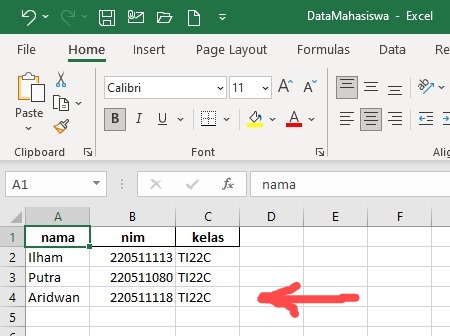


Seperti yang anda lihat, untuk kolom nama yang memiliki nilai/nama "Jambret" sudah terhapus dari file excel.

## F. Input dan Output Mengupdate Data Statistik Excel



Lalu yang terakhir kita akan mencoba untuk mengedit data pada file excel yang sebelumnya telah diinput. Pada gambar kita akan memilih baris mana (indeks), dan kolom mana yang akan diedit nilai/namanya. kemudian jika sudah memilih atau menyeleksi baris dan kolom yang akan diedit, kita akan memperbarui nama/nilai baru pada baris dan kolom yang kita seleksi.



Bisa dilihat pada file excel, nama dan baris kolom yang sebelumnya memiliki nama "Hafid" sekarang sudah diUpdate menjadi nama Aridwan.

# BAB 5

# PENUTUP

## A. Kesimpulan

Python adalah bahasa pemrograman yang mudah dipelajari dan berfokus pada keterbacaan kode. Diklaim memiliki sintaksis yang jelas, lengkap, dan mudah dipahami. Mendukung berbagai paradigma pemrograman, termasuk pemrograman berorientasi objek, imperatif, dan fungsional. Sebagai bahasa multi-paradigma, Python digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi. Jupyter, salah satu produk utamanya, adalah aplikasi web open-source yang memungkinkan pembuatan dan berbagi dokumen interaktif yang menggabungkan kode live, persamaan, visualisasi, dan teks naratif. Pandas, library Python, menyediakan alat untuk manipulasi dan analisis data tabular, seperti spreadsheet atau database relasional. Excel, perangkat lunak pengolah data populer, sering digunakan untuk statistik, analisis data, dan pembuatan laporan.

Namun, untuk analisis data yang lebih kompleks dan efisien, penggunaan Excel dengan Python dan library seperti Pandas menjadi solusi yang efektif. Dengan menggunakan Jupyter Notebook, seluruh proses dapat disatukan dalam satu tempat, memungkinkan impor data, manipulasi, analisis menggunakan Pandas, serta dokumentasi interaktif.

## B. Saran

Untuk meningkatkan kemampuan dalam mengolah data statistik menggunakan Python, ada beberapa saran yang bisa dipertimbangkan, perlu memperdalam pemahaman tentang bahasa pemrograman Python. Selanjutnya, eksplorasi dan kuasai library lain dalam ekosistem Python yang relevan dengan analisis data. Gunakan Jupyter Notebook untuk pengembangan interaktif yang efektif. Kolaborasi dengan rekan tim dan komunitas juga sangat bermanfaat. Dan pastikan untuk memelihara dan memperbarui lingkungan pengembangan Anda secara teratur. Dengan mengikuti saran ini, diharapkan dapat meningkatkan efisiensi dan kualitas dalam mengolah data statistik menggunakan Python.

# DAFTAR PUSTAKA

**Website:**

<https://revou.co/panduan-teknis/python-read-excel>

<https://dqlab.id/mengulik-pandas-python-library-powerful-dalam-pengolahan-data>

<https://revou.co/panduan-teknis/python-data-cleaning>

<https://revou.co/kosakata/visual-studio-code>

[https://medium.com/@yasirabd/pengenalan-numpy-pandas-matplotlib b90bafd36c0](https://medium.com/@yasirabd/pengenalan-numpy-pandas-matplotlib%20b90bafd36c0)

<https://revou.co/panduan-teknis/seaborn-python>

<https://360digitmg.com/blog/python-tabulate>

<https://revou.co/panduan-teknis/library-python#:~:text=Pandas%20adalah%20library%20Python%20yang,utama%2C%20yaitu%20Series%20dan%20DataFrame>.

**Jurnal:**

Ardianto, Dicky, and Arif Tri Widiyatmoko. "Color Detector in an Image using Python and Computer Vision Library." Journal of Intelligent Systems and Information Technology 1.1 (2024): 25-30.

Enterprise, Jubilee. Python untuk Programmer Pemula. Elex media komputindo, 2019.

Hendri, Rantau. Pembuatan Kode Program Dan Simulasi Skema Masakan ACD Menggunakan Jupyter Notebook Pada Pemrograman Python Anaconda. Diss. Politeknik LPP Yogyakarta, 2022.

Mait, Charolina Debora, et al. "Sistem Pendukung Keputusan Menggunakan Fuzzy Logic Tahani Untuk Penentuan Golongan Obat Sesuai Dengan Penyakit Diabetes." Jurnal Media Infotama 18.2 (2022): 344-353.

Setiabudidaya, Dedi. "Penggunaan Piranti Lunak Jupyter Notebook dalam Upaya Mensosialisasikan Open Science." (2018).

# TENTANG PENULIS

# 

# 