Laporan Praktikum Pertemuan 2

**Data Science Lanjut**

Dibuat Oleh

Nama : Muhamad Faisal Halim

NIM : 19.240.0163

Kelas : -

Mata Kuliah : Data Science Lanjut

Mahasiswa Pertukaran Mahasiswa.

**Universitas Muhammadiyah Kalimantan Timur**

~ STMIK Widya Pratama Pekalongan

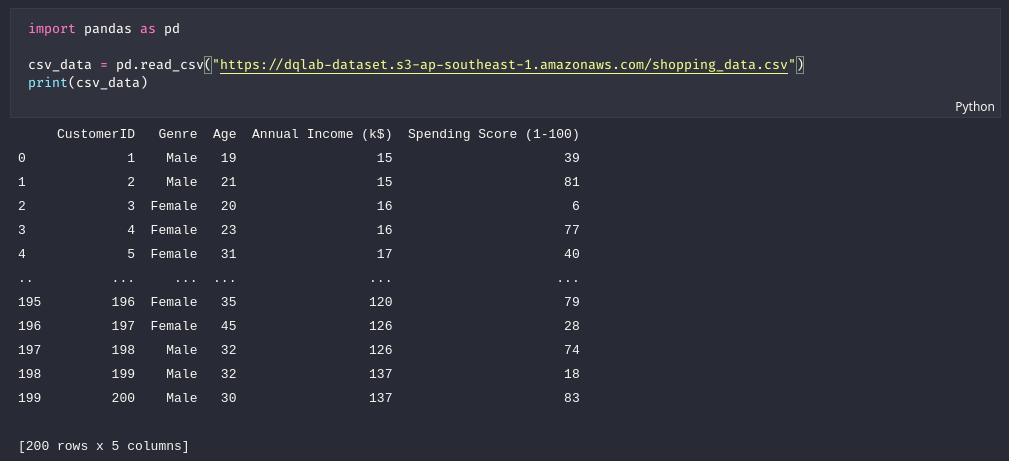
**Note**

Data pada praktikum ini disamakan dengan data yang ada pada contoh yang diberikan di Openlearning UMKT.

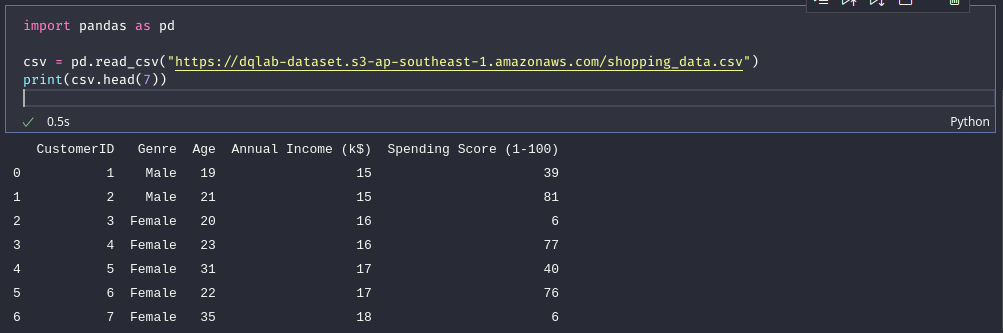
**Praktikum**

**Membaca File Pada Python Menggunakan Pandas**

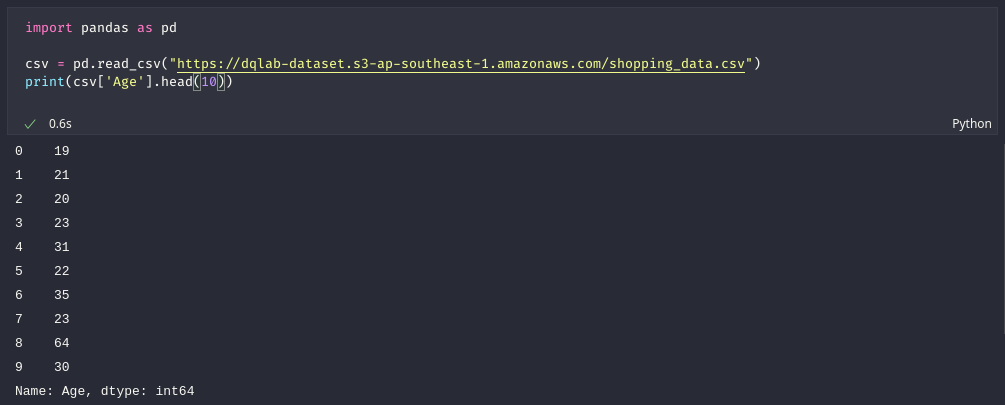
pada praktikum ini, kita akan mencoba membaca file dengan pandas pada python. Pandas adalah salah satu library untuk bisa membaca file dan proses awal dari analisis data. file yang dapat di baca oleh pandas antara lain. csv, txt, tsv dll.



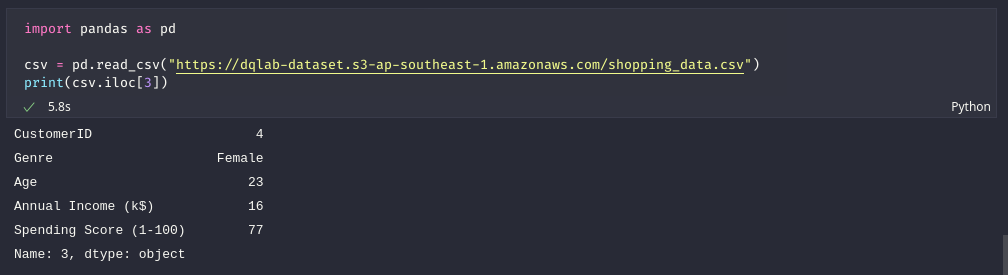
masih menggunakan pandas. terkadang kita memiliki data yang jumlahnya luarbiasa banyak, hal itu akan menyebabkan proses dan loading yang lama. jadi untuk memastikan data terbaca dengan aman, kita bisa menggunakan fungsi .head() dari pandas. secara default fungsi .head() pada pandas akan menampilkan 5 data dari jumlah kesluruhan data yang dibaca.



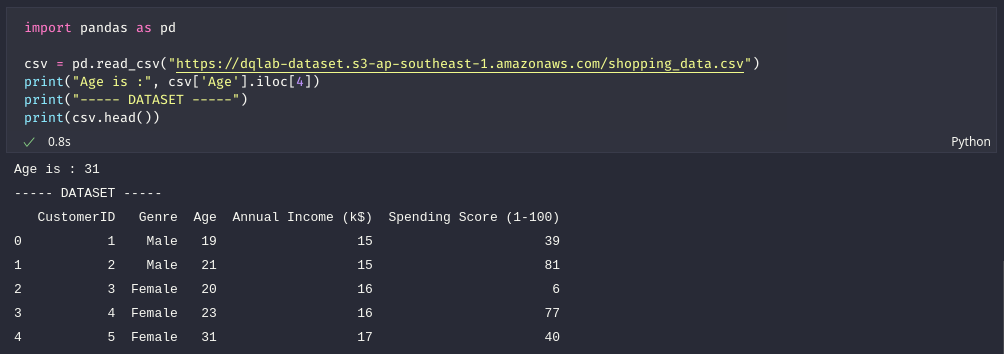
kita juga dappat menampilkan hanya 1 kolom dari table file csv kita, untuk contohnya seperti dibawah ini.



selain itu kita juga dapat mengkombinasikanya dengan .head() untuk hanya menampilkan sebagian data saja. dan selain dapat menampilkan hanya 1 kolom saja, panda juga bisa membantu kita mengintip 1 data dari baris tertentu.



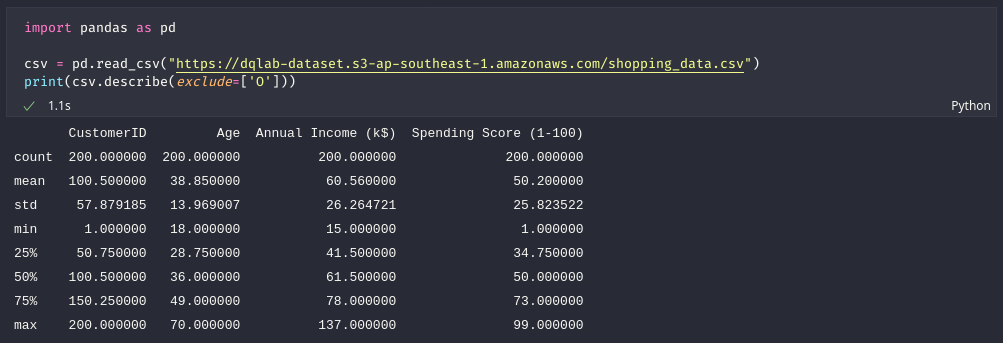
selain itu, kita juha bisa menkombinasikan baris dan kolomnya, jadi kita bisa mengambil 1 baris data dari kolom tertentu. perhatikan pada .iloc[4], angka 4 diambil berdasarkan nomor index data. dari kode dibawah ini kita berhasil menampilkan umur dari data index 4 (data ke 5)



selain itu kita dapat menampilkan data dari range tertentu, misal kita menampilkan data dari index 5 ke index 10. kita bisa menggunakan print(csv.iloc[5:10])

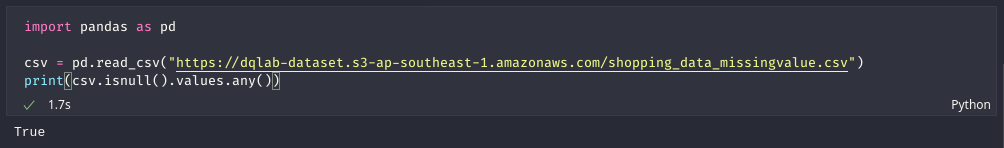
**Menampilkan Informasi statistik**.

pada praktik ini kita akan memanfaatkan fungsi .describe() pada pandas. fungsi ini digunakan untuk melihat beberapa detail statistik dasar seperti persentil, mean, std, dll. dari dataFrame atau serangkaian nilai numeric.



Melakukan Pengecekan Data NULL / NAN pada data

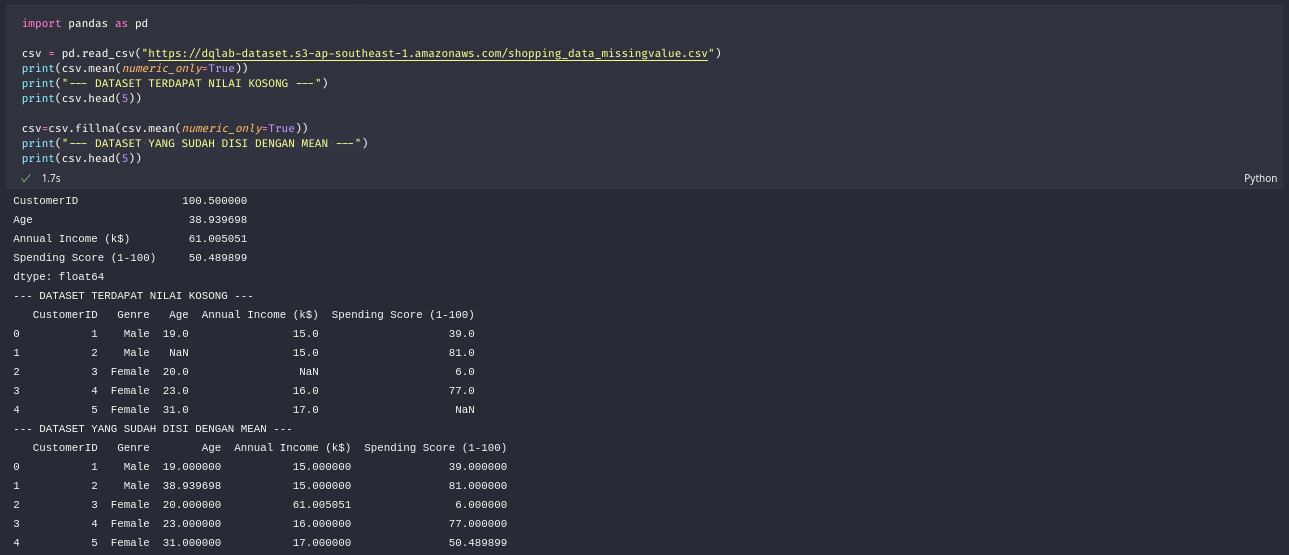
dengan memanfaatkan fungsi yang disediakan oleh pandas kita dapat melakukan pengecekan data jika data itu memiliki nilai null atau nan. tentusaja akan sangat mempermudah ketika kita memiliki ribuan atau lebih data, tentusaja jika dilakukan manual akan sangat memakan waktu dan tenaga.



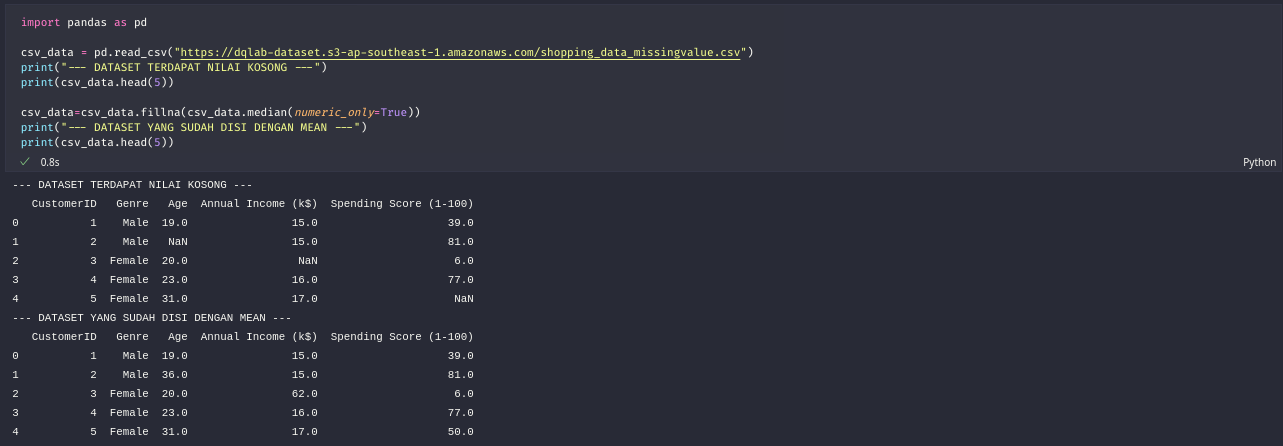
jika dilihat dari dokumentasinya, fungsi ini akan mengembalikan boolean yang menunjukkan jika nilainya NA. Nilai NA, seperti None atau numpy.NaN (NaN / Not a Number), dipetakan ke nilai True. Segala sesuatu yang lain dipetakan ke nilai False. Karakter seperti string kosong ' ' atau numpy.inf tidak dianggap sebagai nilai NA (kecuali jika Anda menyetel pandas.options.mode.use\_inf\_as\_na = True).

lalu untuk mengatasi kekosongan data diatas kita dapat mengisina dengan MEAN atau MEDIAN, seperti pada contoh yang ada. kode praktik ada pada halama selanjutnya.

Mean

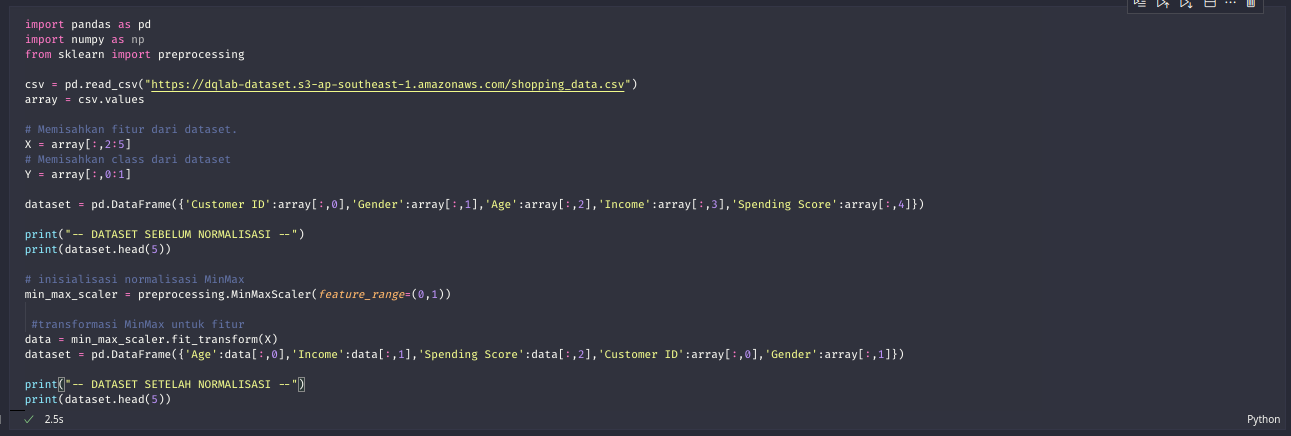


Median



**Praktik Normalisasi Menggunakan Scikit Learn Pada Python**

Scikit Learn merupakan library pada python yang digunakan untuk machine learning dan data science. Salah satu library yang selalu menjadi favorit dan komunitasnya sangat kuat. Scikit-learn sendiri tidak hanya untuk analytics saja, namun juga untuk pre-processing, feature selection, dan proses analysis lainnya



Hasil Output

