

## *Machine Learning*

Python agaknya berbeda dengan MATLAB dalam mengolah data menjadi suatu grafik tertentu. Untuk dapat mengolah suatu data menjadi sebuah grafik, Python perlu mengimpor suatu 'fungsi' ke dalam *library*-nya. Di bawah ini, sajian gambar bagaimana script seharusnya dimulai untuk dapat menjalankan fungsi di bawahnya, sehingga dihasilkan grafik.

```

In [2]: import pandas as pd

In [12]: energy = pd.read_csv('https://www.stats.govt.nz/assets/Uploads/Environmental-economic-accounts/Environmental-economic-account
Out[13]: energy.head()

Out[13]:
  year resource variable units magnitude source data_value flag
0  2007   Biogas  Closing stock  Dollars  Thousands  Environmental Accounts  86690.00  R
1  2007   Biogas  Generation  Gigawatt hours  Actual  MBIE  224.00  F
2  2007   Biogas  Gigawatt hours  Number  Proportion  MBIE  0.01  F
3  2007   Biogas  Opening stock  Dollars  Thousands  Environmental Accounts  NaN  F
4  2007   Biogas  Other changes  Dollars  Thousands  Environmental Accounts  NaN  F

In [14]: energy.tail()

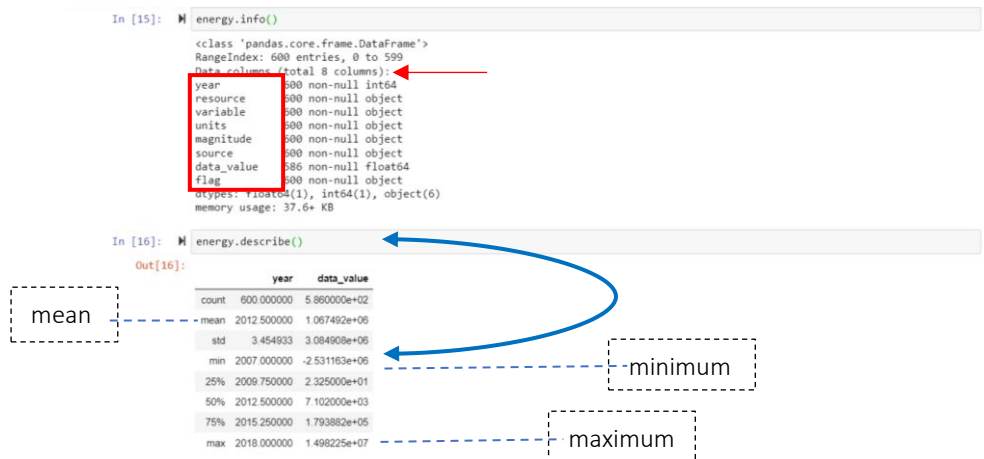
Out[14]:
  year resource variable units magnitude source data_value flag
595  2018   Wood  Generation  Gigawatt hours  Actual  MBIE  301.00  P
596  2018   Wood  Gigawatt hours  Number  Proportion  MBIE  0.01  P
597  2018   Wood  Opening stock  Dollars  Thousands  Environmental Accounts  118567.00  P
598  2018   Wood  Other changes  Dollars  Thousands  Environmental Accounts  -26061.00  P
599  2018   Wood  Resource rent  Dollars  Thousands  Environmental Accounts  5550.00  P
  
```

Gambar 1. Import pandas, langkah sebelum membaca dan mengola data csv file

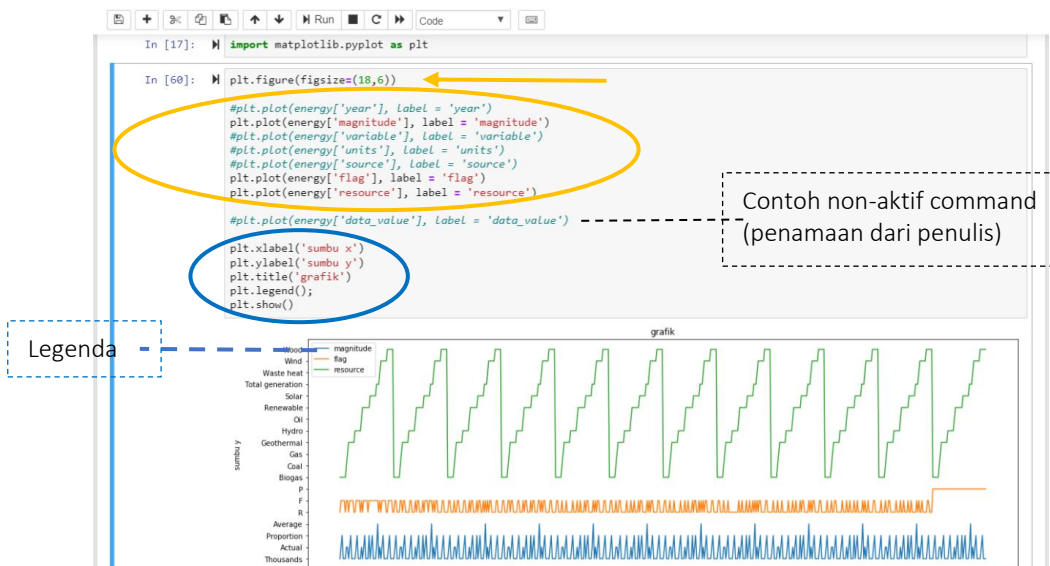
Baris pertama dalam script, akan tanpa hasil jika tidak diikuti dengan script lainnya. Maka, setelah menjalankan baris pertama, hal yang harus dilakukan adalah menginput script selanjutnya. Impor *pandas*, dalam kasus ini dengan menyertakan tautan menuju file data dengan format csv. Maka, pada sel di bawahnya tercantum sebagaimana bertanda kotak kuning pada gambar di atas.

Masih pada gambar (1), script bertanda kotak kuning tidak akan menampilkan sesuatu dan hanya akan menyambungkan script berikutnya dengan data yang dimuat dalam tautan di sel tersebut. Oleh karenanya, sel di bawahnya dituliskan script yang berfungsi untuk mengetahui isi data tersebut. Misalnya, untuk script bertanda panah hijau di atas, hijau tua untuk menampilkan lima data teratas (*head*), dan hijau muda untuk menampilkan lima data terbawah (*tail*).

Menampilkan lima teratas dan lima terbawah dari data yang tersedia, berguna untuk meringkas dan memudahkan user dalam membaca script saat akan menjalankan program. Python dapat memproses data lain tanpa harus menampilkan semua. Akan tetapi, user akan membutuhkan sampel data yang diolahnya. Untuk memenuhi kebutuhan user saat akan menganalisis data olahannya, selain menampilkan beberapa sampel teratas dan terbawah, dapat diketahui beberapa info dan deskripsi lain terkait data yang disajikan. Berikut ini adalah tampilan script yang perlu diinput, disertai dengan hasil *running cell*-nya (gambar 2).

Gambar 2. Input fungsi *info* dan *describe*, beserta Outputnya

*Info* dan *describe* adalah fungsi yang akan menampilkan detail dari data yang dimuat dalam link. Ditandai dengan kotak merah, pada gambar (2) menunjukkan variable apa saja yang dimuat dalam delapan kolom yang disebutkan di kalimat sebelumnya (panah merah). Hal itu merupakan hasil *running* dari *info*. Sedangkan untuk hasil *running* dari *describe* adalah berupa data statistik olahan dari data yang termuat dalam link, misalnya; *mean* (rata-rata), minimum data, dan maximum data. Untuk fungsi *describe* ini, penulis simpulkan hanya untuk mengolah data berupa angka. Sedangkan, pada data yang memiliki *value* bukan berupa angka, perlu diikuti fungsi lainnya untuk menampilkan hasil *running* variabelnya. Pada gambar di atas, hubungan input dan output dari *syntax describe* ditunjukkan dengan tanda panah biru.



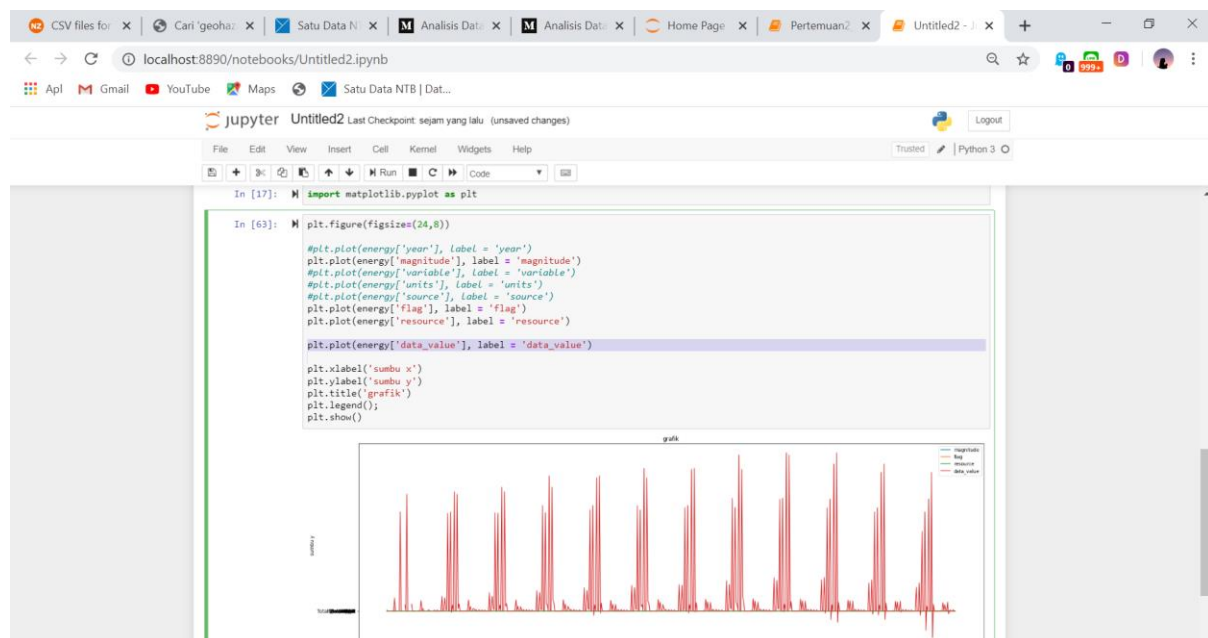
Gambar 3. Import matplotlib, langkah sebelum melakukan plotting (3 variabel) untuk suatu grafik

Selain library pandas yang diimpor ke Python, *matplotlib.pyplot* juga diperlukan dalam proses membuat grafik dari suatu data tertentu pada software Python. Berdasarkan gambar (3) di atas, seperti script impor yang telah dilakukan sebelumnya, tanpa diikuti suatu fungsi lainnya dalam script, maka *import* ini tidak akan menampilkan suatu apapun. Python hanya akan mengenali bahasa yang diimpor. Oleh karena itu, script selanjutnya akan dapat diproses setelah Python memahami bahasa plotting.

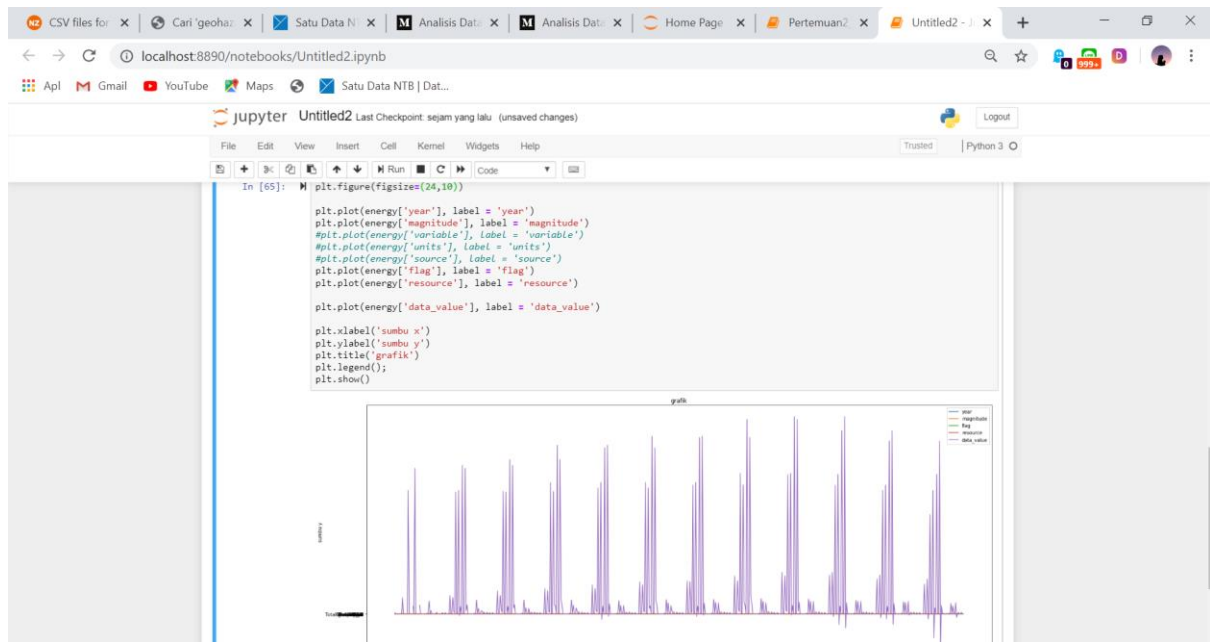
Yang ditunjuk dengan anak panah berwarna kuning pada gambar di atas, adalah script yang menentukan ukuran diagram kartesian yang akan digunakan untuk mengeplot data variable. Selanjutnya, yang berada di dalam lingkaran kuning pada gambar (3), adalah barisan variable apa yang akan diplot datanya, sehingga memunculkan suatu grafik tertentu. Dalam kasus ini, semula penulis menginginkan untuk mengeplot kedelapan variable yang ada untuk dijadikan grafik pada satu diagram kartesius. Namun, untuk beberapa kali percobaan, penulis sulit membaca hasil grafik. Alhasil, beberapa dari data variable yang tertulis di atas, diadakannya script *non-aktif* (hanya komentar), sehingga tidak dimunculkan saat plotting diproses. Caranya dengan membubuhkan tanda pagar (#).

Satu bagian yang tidak ketinggalan untuk dibahas, adalah bagian script pada sel yang masih sama, dimana pada gambar (3) ditandai dengan lingkaran berwarna biru. Bagian tersebut memuat script yang isinya adalah komando untuk melabeli nama grafik yang akan dibuat, parameter di sumbu x, serta sumbu y. Di akhir perintahnya, adalah komando untuk juga menyertakan legenda dari grafik yang ditampilkan. Secara menyeluruh, grafik yang diproses akan *terdisplay* dengan arahan dari komando *plt.show* seperti pada baris terakhir dalam gambar tersebut di atas.

Karena data yang diplot hanya tiga dari delapan variable yang ada, maka di atas adalah tampilan akhir dari gabungan grafik tiga variable terpilih. Berdasarkan legenda yang tercantum di sana, garis hijau dalam grafik adalah hasil plot variable 'source'. Untuk warna orange adalah hasil plot variable 'flag' dan biru untuk hasil plot variable 'magnitude'. Di bawah ini, penulis sertakan hasil plot dengan ragam variable yang dipilih untuk ikut diproses.



Gambar 4. Hasil grafik yang diplotting dengan 4 variabel terpilih



Gambar 5. Hasil grafik yang diplotting dengan 5 variabel terpilih

Secara garis besar, sesuai dengan pengamatan penulis terhadap grafik hasil olah data, setiap tahunnya tidak banyak perubahan yang terjadi dalam setiap variable yang diukur.