



Hands-On

Hands-On ini digunakan pada kegiatan Microcredential Associate Data Scientist 2021

Pertemuan 5

Pertemuan 5 (lima) pada Microcredential Associate Data Scientist 2021 menyampaikan materi mengenai Mengumpulkan Data, Menelaah Data dengan metode Statistik

Pengambilan Data dari API Kaggle

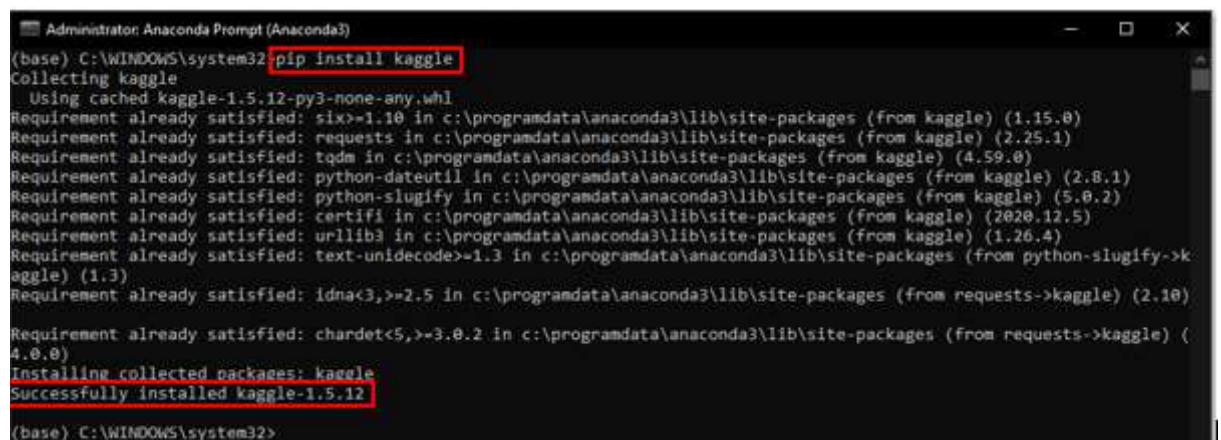
Salah satu portal yang menyediakan dataset untuk project Data Science adalah Kaggle (<https://www.kaggle.com/>). Pada latihan ini, silakan peserta mengunduh dataset mengenai bunga Iris dengan menggunakan kata kunci: "iris species" yang disediakan oleh UCI Machine Learning (UCIML)

1. Install Modul kaggle:

```
In [1]: # Install modul kaggle secara inline (di dalam notebook)
!pip install kaggle
```

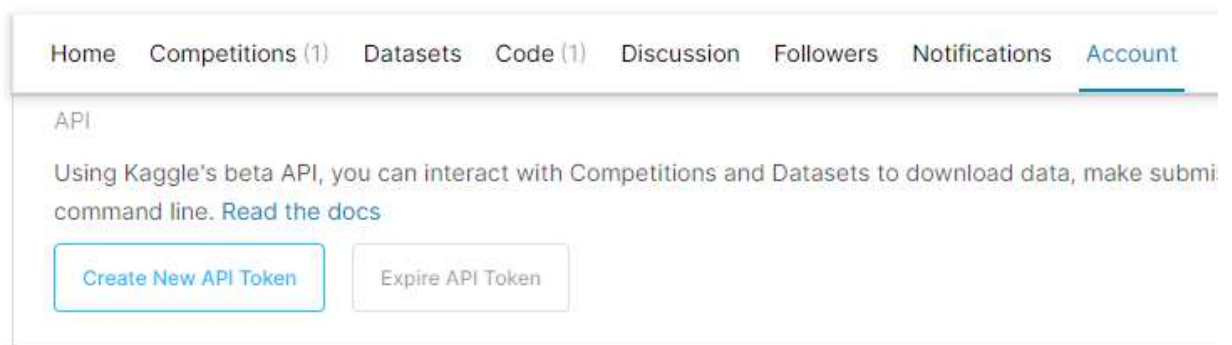
```
Collecting kaggle
  Downloading kaggle-1.5.12.tar.gz (58 kB)
Requirement already satisfied: six>=1.10 in c:\users\expertbook\anaconda3\lib\site-packages (from kaggle) (1.15.0)
Requirement already satisfied: certifi in c:\users\expertbook\anaconda3\lib\site-packages (from kaggle) (2020.12.5)
Requirement already satisfied: python-dateutil in c:\users\expertbook\anaconda3\lib\site-packages (from kaggle) (2.8.1)
Requirement already satisfied: requests in c:\users\expertbook\anaconda3\lib\site-packages (from kaggle) (2.25.1)
Requirement already satisfied: tqdm in c:\users\expertbook\anaconda3\lib\site-packages (from kaggle) (4.59.0)
Collecting python-slugify
  Downloading python_slugify-5.0.2-py2.py3-none-any.whl (6.7 kB)
Requirement already satisfied: urllib3 in c:\users\expertbook\anaconda3\lib\site-packages (from kaggle) (1.26.4)
Collecting text-unidecode>=1.3
  Downloading text_unidecode-1.3-py2.py3-none-any.whl (78 kB)
Requirement already satisfied: chardet<5,>=3.0.2 in c:\users\expertbook\anaconda3\lib\site-packages (from requests->kaggle) (4.0.0)
Requirement already satisfied: idna<3,>=2.5 in c:\users\expertbook\anaconda3\lib\site-packages (from requests->kaggle) (2.10)
Building wheels for collected packages: kaggle
  Building wheel for kaggle (setup.py): started
  Building wheel for kaggle (setup.py): finished with status 'done'
  Created wheel for kaggle: filename=kaggle-1.5.12-py3-none-any.whl size=73053 sha256=6b48273c10d518bc6948aad881820bc7bc319bcb440e60521482044108195e22
  Stored in directory: c:\users\expertbook\appdata\local\pip\cache\wheels\29\da\11\144cc25aebdaeb4931b231e25fd34b394e6a5725cbb2f50106
Successfully built kaggle
Installing collected packages: text-unidecode, python-slugify, kaggle
Successfully installed kaggle-1.5.12 python-slugify-5.0.2 text-unidecode-1.3
```

```
In [2]: # Install modul kaggle secara eksternal melalui anaconda prompt:
```

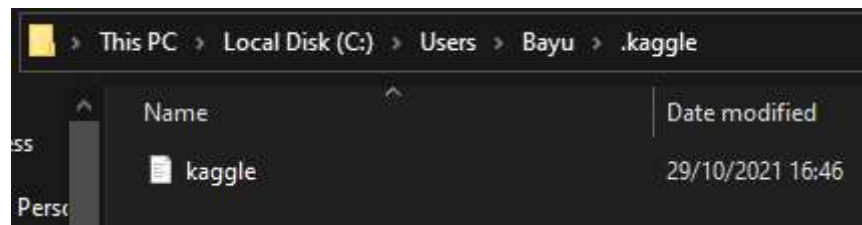


```
Administrator: Anaconda Prompt (Anaconda3)
(base) C:\WINDOWS\system32>pip install kaggle
Collecting kaggle
  Using cached kaggle-1.5.12-py3-none-any.whl
Requirement already satisfied: six>=1.10 in c:\programdata\anaconda3\lib\site-packages (from kaggle) (1.15.0)
Requirement already satisfied: requests in c:\programdata\anaconda3\lib\site-packages (from kaggle) (2.25.1)
Requirement already satisfied: tqdm in c:\programdata\anaconda3\lib\site-packages (from kaggle) (4.59.0)
Requirement already satisfied: python-dateutil in c:\programdata\anaconda3\lib\site-packages (from kaggle) (2.8.1)
Requirement already satisfied: certifi in c:\programdata\anaconda3\lib\site-packages (from kaggle) (2020.12.5)
Requirement already satisfied: urllib3 in c:\programdata\anaconda3\lib\site-packages (from kaggle) (1.26.4)
Requirement already satisfied: text-unidecode>=1.3 in c:\programdata\anaconda3\lib\site-packages (from python-slugify->kaggle) (1.3)
Requirement already satisfied: idna<3,>=2.5 in c:\programdata\anaconda3\lib\site-packages (from requests->kaggle) (2.10)
Requirement already satisfied: chardet<5,>=3.0.2 in c:\programdata\anaconda3\lib\site-packages (from requests->kaggle) (4.0.0)
Installing collected packages: kaggle
Successfully installed kaggle-1.5.12
(base) C:\WINDOWS\system32>
```

2. Create Token API kaggle:



1. Login Kaggle.com
2. Kemudian pada menu Profile --> Account
3. Klik Create New Api Token
4. Maka akan terdownload file kaggle.json



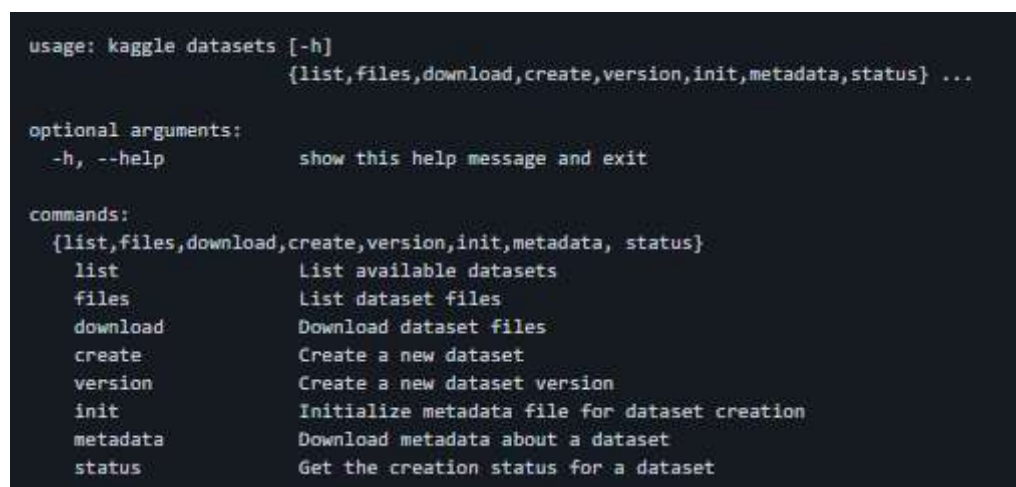
Kaggle API secara default mengasumsikan bahwa file kaggle.json tersebut berada di dalam folder:

- ~/.kaggle/ (Linux/Mac)
- C:\Users<Windows-username>.kaggle\ (Windows)

Jika folder tersebut belum ada:

1. Buat folder di direktori C:\Users<Windows-username>.kaggle\
2. letakkan file kaggle.json kedalam folder tersebut

3. Download Dataset dari Kaggle:



Dokumentasi Kaggle Commands selengkapnya [Disini \(https://github.com/Kaggle/kaggle-api#datasets\)](https://github.com/Kaggle/kaggle-api#datasets).

```
In [1]: # Mencari dataset yang tersedia di kaggle --> pilih data provider dari UCIML
!kaggle datasets list -s Iris
```

ref	size	lastUpdated	downloadCount	voteCount	usabilityRating	title
uciml/iris	4KB	2016-09-27 07:38:05	224131	2664	0.7941176	Iris Species
arshid/iris-flower-dataset	1010B	2018-03-22 15:18:06	40348	369	0.8235294	Iris Flower Dataset
vikrishnan/iris-dataset	999B	2017-08-03 16:00:44	2892	26	0.7647059	Iris Dataset
therohk/ireland-historical-news	52MB	2021-09-25 10:52:48				Irish Times - Waxy-Waxy News
chuckyin/iris-datasets	1KB	2017-03-10 09:35:43	1757	14	0.7352941	Iris datasets
rtatman/iris-dataset-json-version	1KB	2018-04-06 20:21:31			5615	43
parulpandey/palmer-archipelago-antarctica-penguin-data	11KB	2020-06-09 10:14:54			9895	114
conorrot/irish-weather-hourly-data	67MB	2020-06-29 20:15:18			1854	40
saurabh00007/iris.csv	1KB	2017-11-09 07:34:35	17030	57	0.4117647	Iris.csv
fleanend/birds-songs-numeric-dataset	25MB	2019-04-01 09:09:46			704	25
kamrankausar/iris-data	1KB	2017-11-30 10:26:01	1102	13	0.64705884	iris_data
jeffheaton/iris-computer-vision	5MB	2020-11-24 21:23:29	303	9	0.875	Iris Computer Vision
styven/iris-dataset	1KB	2017-11-04 14:10:12	787	8	0.29411766	Iris dataset
arslanali4343/iris-species	2KB	2020-07-02 06:09:09	49	13	0.5625	Iris Species
olgabelitskaya/flower-color-images	50MB	2020-10-01 22:48:07	8344	161	0.75	Flower Color Images
naureenmohammad/mmu-iris-dataset	30MB	2020-07-25 18:38:33	639	19	0.5625	MMU iris dataset
rutujavaidya/iris-dataset	1KB	2021-07-25 17:37:14	33	6	0.4117647	Iris Dataset
shantanuss/iris-flower-dataset	1KB	2020-01-18 19:43:18	191	3	0.9411765	IRIS flower dataset
ashishs0ni/iris-dataset	1KB	2018-08-05 14:26:19	584	7	0.64705884	Iris_dataset
jodx666/iris-is	981B	2018-02-16 09:11:35	339	6	0.3125	IRIS is

```
In [4]: # Download dan ekstrak dataset, secara default akan berada dalam satu direktori c
!kaggle datasets download uciml/iris --unzip
```

Downloading iris.zip to C:\Users\User\Microcred\Persiapan

```
0%|          | 0.00/3.60k [00:00<?, ?B/s]
100%|#####| 3.60k/3.60k [00:00<00:00, 665kB/s]
```

Atau bisa juga menggunakan link dari kaggle

Latihan (1)

Silahkan Download sebuah dataset menggunakan API Kaggle

```
In [2]: #Latihan (1)
!kaggle datasets list -s Iris
```

ref	size	lastUpdated	downloadCount	voteCount	title	usabilityRating
uciml/iris	4KB	2016-09-27 07:38:05	224131	2664	Iris Species	0.7941176
arshid/iris-flower-dataset	1010B	2018-03-22 15:18:06	40348	369	Iris Flower Dataset	0.8235294
vikrishnan/iris-dataset	999B	2017-08-03 16:00:44	2892	26	Iris Dataset	0.7647059
therohk/ireland-historical-news	52MB	2021-09-25 10:52:48			Irish Times - Waxy-Wany	2977 157 1.0
chuckyin/iris-datasets	1KB	2017-03-10 09:35:43	1757	14	Iris datasets	0.7352941
rtatman/iris-dataset-json-version	1KB	2018-04-06 20:21:31			Iris Dataset (JSON Version)	5615 43 0.75
parulpandey/palmer-archipelago-antarctica-penguin-data	11KB	2020-06-09 10:14:54			Palmer Archipelago (Antarctica) penguin data	9895 114 0.9705882
conorrot/irish-weather-hourly-data	67MB	2020-06-29 20:15:18			Irish Weather (hourly data)	1854 40 0.8235294
saurabh00007/iris.csv	1KB	2017-11-09 07:34:35	17030	57	Iris.csv	0.4117647
fleanend/birds-songs-numeric-dataset	25MB	2019-04-01 09:09:46			Birds' Songs Numeric Dataset	704 25 0.9411765
kamrankausar/iris-data	1KB	2017-11-30 10:26:01	1102	13	iris_data	0.64705884
jeffheaton/iris-computer-vision	5MB	2020-11-24 21:23:29	303	9	Iris Computer Vision	0.875
styven/iris-dataset	1KB	2017-11-04 14:10:12	787	8	Iris dataset	0.29411766
arсланali4343/iris-species	2KB	2020-07-02 06:09:09	49	13	Iris Species	0.5625
olgabelitskaya/flower-color-images	50MB	2020-10-01 22:48:07	8344	161	Flower Color Images	0.75
naureenmohammad/mmu-iris-dataset	30MB	2020-07-25 18:38:33	639	19	MMU iris dataset	0.5625
rutujavidya/iris-dataset	1KB	2021-07-25 17:37:14	33	6	Iris Dataset	0.4117647
shantanuss/iris-flower-dataset	1KB	2020-01-18 19:43:18	191	3	IRIS flower dataset	0.9411765
ashishs0ni/iris-dataset	1KB	2018-08-05 14:26:19	584	7	Iris_dataset	0.64705884
jodx666/iris-is	981B	2018-02-16 09:11:35	339	6	IRIS is	0.3125

PENGUNAAN LIBRARY PANDAS dan NUMPY

Pada materi ini, peserta sudah mendapatkan pemahaman mengenai data dan dataset. Penggunaan library pada Python memberikan kemudahan dalam proses data understanding. Beberapa library yang digunakan adalah library Pandas dan Numpy.

Latihan (2)

Lakukan import Library Pandas dan Library Numpy

```
In [3]: #Latihan(2)
        #Import Library Pandas

        import pandas as pd

        #Import Library Numpy

        import numpy as np
```

DATAFRAME

DataFrame adalah struktur data 2 dimensi yang berbentuk tabular (mempunyai baris dan kolom). Hampir semua data tidak hanya memiliki 1 kolom tetapi lebih dari 1 kolom, sehingga lebih cocok menggunakan pandas DataFrame untuk mengolahnya.

Penggunaan dataframe pada Python dengan menggunakan syntaks: df.

Latihan (3)

Panggil file (load dataset) dengan format .csv untuk dataset mengenai bunga Iris yang sudah peserta unduh dari Kaggle, dan akan disimpan di dalam dataframe df. Lalu tampilkan 5 baris awal dataset dengan function head()

```
In [4]: #Latihan(3)
        #Panggil file (load file bernama Iris.csv) dan simpan dalam dataframe Lalu tampilkan

        path = "Iris.csv"
        df = pd.read_csv(path)
```

```
In [5]: # menampilkan 5 baris awal dari dataset
df.head()
```

Out[5]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

Telaah Data

Pada telaah data, dapat dilakukan untuk mengetahui:

- tipe data dari setiap kolom
- deskripsi statistik data

Latihan (4)

Tampilkan tipe data dari kolom yang ada pada dataset

```
In [6]: #Latihan(4)
#Tampilkan tipe data dari kolom yang ada pada dataset

print(df.dtypes)
```

```
Id                int64
SepalLengthCm     float64
SepalWidthCm      float64
PetalLengthCm     float64
PetalWidthCm      float64
Species           object
dtype: object
```

Latihan (5)

Apakah tipe Data dari kolom berikut ini: (silakan diisi pada cell di bawah ini)


```
In [9]: #Latihan (5)
#Tipe Data dari kolom yang ada di dataset

#Kolom "Id" memiliki tipe data = <int64>
#Kolom "SepalLengthCm" memiliki tipe data = <float64>
#Kolom "SepalWidthCm" memiliki tipe data = <float64>
#Kolom "PetalLengthCm" memiliki tipe data = <float64>
#Kolom "PetalWidthCm" memiliki tipe data = <float64>
#Kolom "Species" memiliki tipe data = <object>
#Dengan dtype memiliki tipe data <object>
```

Latihan (6)

Hitunglah ukuran (jumlah baris dan kolom) dari dataset. Dengan menggunakan method function

```
In [7]: #Latihan (6)
#Hitung ukuran (jumlah baris dan kolom) dari dataset

df.sort_values(by=["Id"])
```

Out[7]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

Latihan (7)

Berapakah jumlah baris, dan jumlah kolom pada dataset? (silakan diisi pada cell di bawah ini)

```
In [11]: #Latihan (7)

#Jumlah Baris pada dataset adalah = <150 Baris>

#Jumlah kolom pada dataset adalah = <6 Kolom>
```

Latihan (8)

Tampilkan data yang hanya berisi kolom "Id" dan kolom "Species" dalam bentuk dataframe.

```
In [8]: #Latihan (8)
#Tampilkan data untuk kolom "Id" dan kolom "Species" dalam bentuk dataframe

df[["Id", "Species"]]
```

Out[8]:

	Id	Species
0	1	Iris-setosa
1	2	Iris-setosa
2	3	Iris-setosa
3	4	Iris-setosa
4	5	Iris-setosa
...
145	146	Iris-virginica
146	147	Iris-virginica
147	148	Iris-virginica
148	149	Iris-virginica
149	150	Iris-virginica

Latihan (9)

Tampilkan data dengan dataframe, dan data yang ditampilkan adalah data pada baris dengan indeks 0 (nol) sampai dengan indeks 9 (sembilan)

```
In [9]: #Latihan (9)
#Tampilkan data dengan dataframe, dan data yang ditampilkan adalah baris dengan i
df[["Id", "Species"]][:10]
```

Out[9]:

	Id	Species
0	1	Iris-setosa
1	2	Iris-setosa
2	3	Iris-setosa
3	4	Iris-setosa
4	5	Iris-setosa
5	6	Iris-setosa
6	7	Iris-setosa
7	8	Iris-setosa
8	9	Iris-setosa
9	10	Iris-setosa

Latihan (10)

Tampilkan data hanya kolom "Id" dan kolom "Species" dengan dataframe, dan yang ditampilkan adalah data pada baris dengan indeks 11 (sebelas) sampai dengan indeks 15 (limabelas)

```
In [10]: #Latihan (10)
#Tampilkan data hanya kolom "Id" dan kolom "Species", pada baris dengan indeks 11
df[["Id", "Species"]][11:16]
```

Out[10]:

	Id	Species
11	12	Iris-setosa
12	13	Iris-setosa
13	14	Iris-setosa
14	15	Iris-setosa
15	16	Iris-setosa

Latihan (11)

Pada DataFrame dapat menampilkan beberapa baris pertama/terakhir dari dataset yang di load. Gunakan Method head() dan tail().

Latihan: Tampilkan data pada 8 (delapan) baris pertama dari dataset. dengan dataframe.

```
In [11]: #Latihan (11)
#Tampilkan data pada 8 (delapan) baris pertama dari dataset, dengan dataframe

df.head(8)
```

Out[11]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa

Latihan (12)

Tampilkan data pada 3 (tiga) baris terakhir dari dataset, dengan dataframe.

```
In [12]: #Latihan (12)
#Tampilkan data pada 3 (tiga) baris terakhir dari dataset, dengan dataframe

df.tail(3)
```

Out[12]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

Deskripsi Statistik Data

DataFrame method describe() menampilkan statistik dasar setiap kolom data yang bertipe numerik, mencakup banyaknya data (count), rerata aritmetik (mean), simpangan baku (std), nilai terkecil (min), kuartil pertama (25%), kuartil kedua/median (50%), kuartil ketiga (75%), dan nilai terbesar (max).

Latihan (13)

Hitung korelasi dari dataset. Dengan menggunakan method function

```
In [13]: #Latihan (13)
#Hitung korelasi dataset
df.describe()
```

```
Out[13]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

```
In [14]: df.loc[:, 'SepalLengthCm':].corr()
```

```
Out[14]:
```

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
SepalLengthCm	1.000000	-0.109369	0.871754	0.817954
SepalWidthCm	-0.109369	1.000000	-0.420516	-0.356544
PetalLengthCm	0.871754	-0.420516	1.000000	0.962757
PetalWidthCm	0.817954	-0.356544	0.962757	1.000000

Latihan (14)

Berdasarkan pada perhitungan korelasi di Latihan (11), apakah yang dapat Bapak/Ibu simpulkan sementara? Silakan tuliskan simpulan sementara Bapak/Ibu pada cell di bawah ini.

```
In [16]: #Latihan (14)
#Simpulan Sementara Hasil Korelasi di Latihan (13)

### Korelasi Pearson antara kolom-kolom numerik

#Berdasarkan contoh diatas menggambarkan bahwa
#SepalLengthCm berkorelasi positif dengan SepalLengthCm
#SepalWidthCm berkorelasi negatif dengan SepalLengthCm, PetalLengthCm dan PetalWidthCm
#PetalLengthCm tidak ada korelasi linier dengan SepalLengthCm dan PetalWidthCm
#dan mereka berkorelasi positif dengan sesama tipe/kolom
```

Latihan (15)

Hitung korelasi untuk kolom berikut ini: PetalLengthCm, PetalWidthCm

```
In [17]: #Latihan (15)
#Hitung korelasi dataset untuk kolom PetalLengthCm, PetalWidthCm
df.loc[:, 'PetalLengthCm'].corr()
```

```
Out[17]:
```

	PetalLengthCm	PetalWidthCm
PetalLengthCm	1.000000	0.962757
PetalWidthCm	0.962757	1.000000

```
In [32]: df.loc[:, 'PetalLengthCm'].corr().sum()
```

```
Out[32]: PetalLengthCm    1.962757
PetalWidthCm    1.962757
dtype: float64
```

```
In [18]: #Berdasarkan contoh diatas menggambarkan bahwa
#PetalLengthCm tidak ada korelasi linier dengan PetalWidthCm dan sebaliknya
#PetalWidthCm tidak ada korelasi linier dengan PetalLengthCm dan sebaliknya
#Tetapi ketika korelasi keduanya dijumlahkan maka menjadikan kedua komponen tersebut
```

```
In [19]: df.loc[:, 'PetalWidthCm'].corr()
```

```
Out[19]:
```

	PetalWidthCm
PetalWidthCm	1.0

```
In [20]: df.loc[:, 'PetalWidthCm'].corr().sum()
```

```
Out[20]: PetalWidthCm    1.0
dtype: float64
```

```
In [22]: #Berdasarkan contoh diatas menggambarkan bahwa
#PetalWidthCm berkorelasi positif dengan PetalWidthCm
#dan data menunjukkan komponen tersebut menjadi korelasi positif dengan tingkat k
```

Latihan (16)

Method "describe" secara otomatis melakukan komputasi statistik untuk semua continuous variable. Secara default "describe" melakukan ignore terhadap variabel bertipe objek.

Komputasi statistik yang dilakukan terdiri dari: count, mean, std, min, max, 25%, 75%, max.

Latihan: Gunakan method describe pada dataset yang sudah di load untuk semua continuous variabel. (Dataset Iris.csv)

```
In [23]: #Latihan (16)
# Penggunaan Metode describe untuk komputasi statistik

df.describe(include="all")
```

Out[23]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
count	150.000000	150.000000	150.000000	150.000000	150.000000	150
unique	NaN	NaN	NaN	NaN	NaN	3
top	NaN	NaN	NaN	NaN	NaN	Iris-setosa
freq	NaN	NaN	NaN	NaN	NaN	50
mean	75.500000	5.843333	3.054000	3.758667	1.198667	NaN
std	43.445368	0.828066	0.433594	1.764420	0.763161	NaN
min	1.000000	4.300000	2.000000	1.000000	0.100000	NaN
25%	38.250000	5.100000	2.800000	1.600000	0.300000	NaN
50%	75.500000	5.800000	3.000000	4.350000	1.300000	NaN
75%	112.750000	6.400000	3.300000	5.100000	1.800000	NaN
max	150.000000	7.900000	4.400000	6.900000	2.500000	NaN

Latihan (17)

Gunakan method describe pada dataset yang sudah di load untuk data bertipe objek. (Dataset Iris.csv)

```
In [24]: #Latihan (17)
#Gunakan method describe pada dataset yang sudah di Load untuk data bertipe objek
df.describe()
```

Out[24]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

Latihan 18

Gunakan method describe pada dataset yang sudah di load untuk semua type data (continous variabel dan type object).

```
In [25]: #Latihan (18)
#Gunakan method describe pada dataset yang sudah di load untuk semua type data

df.describe(include="all")
```

```
Out[25]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
count	150.000000	150.000000	150.000000	150.000000	150.000000	150
unique	NaN	NaN	NaN	NaN	NaN	3
top	NaN	NaN	NaN	NaN	NaN	Iris-setosa
freq	NaN	NaN	NaN	NaN	NaN	50
mean	75.500000	5.843333	3.054000	3.758667	1.198667	NaN
std	43.445368	0.828066	0.433594	1.764420	0.763161	NaN
min	1.000000	4.300000	2.000000	1.000000	0.100000	NaN
25%	38.250000	5.100000	2.800000	1.600000	0.300000	NaN
50%	75.500000	5.800000	3.000000	4.350000	1.300000	NaN
75%	112.750000	6.400000	3.300000	5.100000	1.800000	NaN
max	150.000000	7.900000	4.400000	6.900000	2.500000	NaN

Latihan (19)

Hitunglah nilai mean dari dataset.

```
In [26]: #Latihan (19)
#Hitung nilai Mean dari dataset

df.mean()
```

```
Out[26]: Id          75.500000
SepalLengthCm      5.843333
SepalWidthCm       3.054000
PetalLengthCm      3.758667
PetalWidthCm       1.198667
dtype: float64
```

Latihan (20)

Hitung nilai mean dari dataset untuk kolom PetalLengthCm.


```
In [27]: #Latihan (20)
#Hitung nilai Mean untuk kolom PetalLengthCm

df.mean()['PetalLengthCm']
```

```
Out[27]: 3.75866666666666693
```

Latihan (21)

Carilah nilai minimal dari dataset untuk kolom SepalWidthCm.

```
In [28]: #Latihan (21)
#Cari nilai minimal untuk kolom SepalWidthCm

df.min()['SepalWidthCm']
```

```
Out[28]: 2.0
```

Method Groupby

Method groupby memungkinkan analisis dilakukan secara per kelompok nilai atribut tertentu.

Latihan (22)

Hitunglah nilai mean dari dataset untuk kolom SepalLengthCm per Species dengan menggunakan metode groupby.

```
In [29]: #Latihan (22)
#Hitung nilai mean dari dataset untuk SepalLengthCm per Species dengan metode groupby

df.groupby('Species')['SepalLengthCm'].mean()
```

```
Out[29]: Species
Iris-setosa      5.006
Iris-versicolor  5.936
Iris-virginica   6.588
Name: SepalLengthCm, dtype: float64
```

Method Value Count

value_counts() menghasilkan frekuensi setiap nilai unik di dalam kolom, dan yang tertinggi count-nya adalah merupakan modus pada kolom tersebut.

Latihan (23)

Hitunglah frekuensi pada kolom 'Species' dengan menggunakan metode value_counts().

```
In [30]: #Latihan (23)
#Hitung frekuensi pada kolom 'Species' dengan menggunakan metode value_counts()

df['Species'].value_counts()
```

```
Out[30]: Iris-setosa      50
Iris-virginica      50
Iris-versicolor      50
Name: Species, dtype: int64
```

Latihan (24)

Tampilkan perhitungan frekuensi pada kolom 'Species' dengan menggunakan metode value_counts() dalam bentuk dataframe.

```
In [31]: #Latihan (24)
#Perhitungan frekuensi pada kolom 'Species' dengan menggunakan metode value_counts()
df[["Id", "Species"]].value_counts()
```

```
Out[31]: Id  Species
1  Iris-setosa      1
95  Iris-versicolor  1
97  Iris-versicolor  1
98  Iris-versicolor  1
99  Iris-versicolor  1
..
51  Iris-versicolor  1
52  Iris-versicolor  1
53  Iris-versicolor  1
54  Iris-versicolor  1
150  Iris-virginica  1
Length: 150, dtype: int64
```

Latihan (25)

Hitunglah frekuensi pada kolom 'PetalLenghCm' dengan menggunakan metode value_counts() dan dalam bentuk dataframe.

```
In [32]: #Latihan (25)
# Hitung frekuensi pada kolom 'PetalLengthCm' dengan menggunakan metode value_counts()

df[["Id", "PetalLengthCm"]].value_counts()
```

```
Out[32]: Id    PetalLengthCm
1      1.4          1
95     4.2          1
97     4.2          1
98     4.3          1
99     3.0          1
      ..
51     4.7          1
52     4.5          1
53     4.9          1
54     4.0          1
150    5.1          1
Length: 150, dtype: int64
```