

```
!pip install kaggle
```

```
Requirement already satisfied: kaggle in  
/usr/local/lib/python3.7/dist-packages (1.5.12)  
Requirement already satisfied: certifi in  
/usr/local/lib/python3.7/dist-packages (from kaggle) (2021.10.8)  
Requirement already satisfied: python-dateutil in  
/usr/local/lib/python3.7/dist-packages (from kaggle) (2.8.2)  
Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-  
packages (from kaggle) (4.62.3)  
Requirement already satisfied: python-slugify in  
/usr/local/lib/python3.7/dist-packages (from kaggle) (5.0.2)  
Requirement already satisfied: six>=1.10 in  
/usr/local/lib/python3.7/dist-packages (from kaggle) (1.15.0)  
Requirement already satisfied: requests in  
/usr/local/lib/python3.7/dist-packages (from kaggle) (2.23.0)  
Requirement already satisfied: urllib3 in  
/usr/local/lib/python3.7/dist-packages (from kaggle) (1.24.3)  
Requirement already satisfied: text-unidecode>=1.3 in  
/usr/local/lib/python3.7/dist-packages (from python-slugify->kaggle)  
(1.3)  
Requirement already satisfied: idna<3,>=2.5 in  
/usr/local/lib/python3.7/dist-packages (from requests->kaggle) (2.10)  
Requirement already satisfied: chardet<4,>=3.0.2 in  
/usr/local/lib/python3.7/dist-packages (from requests->kaggle) (3.0.4)
```

```
#Latihan (1)
```

```
#Langkah nya seperti contoh diatas
```

```
!kaggle datasets list -s Iris
```

```
!kaggle datasets download uciml/iris --unzip
```

```
#Latihan(2)
```

```
#Import Library Pandas
```

```
import pandas as pd
```

```
#Import Library Numpy
```

```
import numpy as np
```

```
#latihan(3)
```

```
#Panggil file (load file bernama Iris.csv) dan simpan dalam dataframe  
Lalu tampilkan 5 baris awal dataset dengan function head()
```

```
dataset = pd.read_csv("Iris.csv")
```

```
dataset.head(6)
```

```
   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						

#latihan(4)

#Tampilkan tipe data dari kolom yang ada pada dataset

```
print(dataset.dtypes)
```

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

#Latihan (5)

#Tipe Data dari kolom yang ada di dataset

```
print(dataset.dtypes)
```

#Kolom "Id" memiliki tipe data = <isikan jawaban di sini>

```
print('Kolom "Id" memiliki tipe data =',(dataset['Id'].dtypes))
```

#Kolom "SepalLengthCm" memiliki tipe data = <isikan jawaban di sini>

```
print('Kolom "SepalLengthCm" memiliki tipe data =',
(dataset['SepalLengthCm'].dtypes))
```

#Kolom "Species" memiliki tipe data = <isikan jawaban di sini>

```
print('Kolom "Species" memiliki tipe data =',
(dataset['Species'].dtypes))
```

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

```
Kolom "Id" memiliki tipe data = int64
```

```
Kolom "SepalLengthCm" memiliki tipe data = float64
```

```
Kolom "Species" memiliki tipe data = object
```

```
#Latihan (6)
#Hitung ukuran (jumlah baris dan kolom) dari dataset
dataset.shape
```

```
(150, 6)
```

```
#Latihan (7)
```

```
#Jumlah Baris pada dataset adalah = <isikan jawaban di sini>
print('Jumlah Baris pada dataset adalah ',len(dataset.index))
```

```
#Jumlah kolom pada dataset adalah = <isikan jawaban di sini>
print('Jumlah kolom pada dataset adalah ',len(dataset.columns))
```

```
Jumlah Baris pada dataset adalah = 150
Jumlah kolom pada dataset adalah = 6
```

```
#Latihan (8)
```

```
#Tampilkan data untuk kolom "Id" dan kolom "Species" dalam bentuk
dataframe
```

```
dataset[['Id','Species']]
```

	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

```
[150 rows x 2 columns]
```

```
#Latihan (9)
```

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

```
dataset.loc[:9]
```

	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					
8	9	4.4	2.9	1.4	0.2 Iris-
setosa					
9	10	4.9	3.1	1.5	0.1 Iris-
setosa					

#Latihan (10)

#Tampilkan data hanya kolom "Id" dan kolom "Species", pada baris dengan indeks 0 (nol) sampai dengan indeks 9 (sembilan)

```
dataset[['Id', 'Species']][:10]
```

	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 (11)

#Tampilkan data pada 8 (delapan) baris pertama dari dataset, dengan dataframe

```
dataset.head(9)
```

	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
8 9          4.4          2.9          1.4          0.2 Iris-
setosa

```

#Latihan (12)

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

```
dataset.tail(3)
```

```

      Id  SepalLengthCm  ...  PetalWidthCm      Species
147  148           6.5  ...           2.0  Iris-virginica
148  149           6.2  ...           2.3  Iris-virginica
149  150           5.9  ...           1.8  Iris-virginica

```

[3 rows x 6 columns]

#Latihan (13)

#Hitung korelasi dataset

```

correlations = dataset.corr()
print(correlations)

```

```

              Id  SepalLengthCm  ...  PetalLengthCm
PetalWidthCm
Id          1.000000      0.716676  ...      0.882747
0.899759
SepalLengthCm 0.716676      1.000000  ...      0.871754
0.817954
SepalWidthCm -0.397729     -0.109369  ...     -0.420516      -
0.356544
PetalLengthCm 0.882747      0.871754  ...      1.000000
0.962757
PetalWidthCm  0.899759      0.817954  ...      0.962757
1.000000

```

[5 rows x 5 columns]

#latihan (14)

#Simpulan Sementara Hasil Korelasi di latihan (13)

```
print('Data Korelasi : -1 = korelasi negatif, 0 = tidak ada korelasi
linear, +1 = korelasi positif')
```

Data Korelasi : -1 = korelasi negatif, 0 = tidak ada korelasi linear, +1 = korelasi positif

#Latihan (15)

#Hitung korelasi dataset untuk kolom PetalLengthCm, PetalWidthCm

```
print(correlations[['PetalLengthCm','PetalWidthCm']])
```

	PetalLengthCm	PetalWidthCm
Id	0.882747	0.899759
SepalLengthCm	0.871754	0.817954
SepalWidthCm	-0.420516	-0.356544
PetalLengthCm	1.000000	0.962757
PetalWidthCm	0.962757	1.000000

#Latihan (16)

Penggunaan Metode describe untuk komputasi statistik

```
dataset.describe()
```

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

#Latihan (17)

#Gunakan method describe pada dataset yang sudah di load untuk data bertype objek

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

	Id	SepalLengthCm	...	PetalWidthCm	Species
count	150.000000	150.000000	...	150.000000	150
unique	NaN	NaN	...	NaN	3
top	NaN	NaN	...	NaN	Iris-virginica
freq	NaN	NaN	...	NaN	50
mean	75.500000	5.843333	...	1.198667	NaN
std	43.445368	0.828066	...	0.763161	NaN

min	1.000000	4.300000	...	0.100000	NaN
25%	38.250000	5.100000	...	0.300000	NaN
50%	75.500000	5.800000	...	1.300000	NaN
75%	112.750000	6.400000	...	1.800000	NaN
max	150.000000	7.900000	...	2.500000	NaN

[11 rows x 6 columns]

#Latihan (18)

#Gunakan method describe pada dataset yang sudah di load untuk semua type data

dataset.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Id               150 non-null    int64
1   SepalLengthCm    150 non-null    float64
2   SepalWidthCm     150 non-null    float64
3   PetalLengthCm    150 non-null    float64
4   PetalWidthCm     150 non-null    float64
5   Species          150 non-null    object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

#Latihan (19)

#Hitung nilai Mean dari dataset

dataset.mean()

```
Id                75.500000
SepalLengthCm     5.843333
SepalWidthCm      3.054000
PetalLengthCm     3.758667
PetalWidthCm      1.198667
dtype: float64
```

#Latihan (20)

#Hitung nilai Mean untuk kolom PetalLengthCm

dataset["PetalLengthCm"].mean()

3.7586666666666693

#Latihan (21)

#Cari nilai minimal untuk kolom SepalWidthCm

dataset["SepalWidthCm"].min()

2.0

#Latihan (22)

#Hitung nilai mean dari dataset untuk SepalLengthCm per Species dengan metode groupby

```
dataset.groupby(['SepalLengthCm', 'Species']).mean()
```

		Id	...	PetalWidthCm
SepalLengthCm	Species		...	
4.3	Iris-setosa	14.000000	...	0.100000
4.4	Iris-setosa	30.333333	...	0.200000
4.5	Iris-setosa	42.000000	...	0.300000
4.6	Iris-setosa	20.500000	...	0.225000
4.7	Iris-setosa	16.500000	...	0.200000
4.8	Iris-setosa	25.400000	...	0.200000
4.9	Iris-setosa	21.250000	...	0.125000
	Iris-versicolor	58.000000	...	1.000000
	Iris-virginica	107.000000	...	1.700000
5.0	Iris-setosa	29.625000	...	0.287500
	Iris-versicolor	77.500000	...	1.000000
5.1	Iris-setosa	27.125000	...	0.312500
	Iris-versicolor	99.000000	...	1.100000
5.2	Iris-setosa	30.000000	...	0.166667
	Iris-versicolor	60.000000	...	1.400000
5.3	Iris-setosa	49.000000	...	0.200000
5.4	Iris-setosa	17.400000	...	0.320000
	Iris-versicolor	85.000000	...	1.500000
5.5	Iris-setosa	35.500000	...	0.200000
	Iris-versicolor	79.600000	...	1.180000
5.6	Iris-versicolor	77.200000	...	1.300000
	Iris-virginica	122.000000	...	2.000000
5.7	Iris-setosa	17.500000	...	0.350000
	Iris-versicolor	85.800000	...	1.220000
	Iris-virginica	114.000000	...	2.000000
5.8	Iris-setosa	15.000000	...	0.200000
	Iris-versicolor	81.333333	...	1.133333
	Iris-virginica	120.000000	...	2.066667
5.9	Iris-versicolor	66.500000	...	1.650000
	Iris-virginica	150.000000	...	1.800000
6.0	Iris-versicolor	78.000000	...	1.425000
	Iris-virginica	129.500000	...	1.650000
6.1	Iris-versicolor	75.500000	...	1.325000
	Iris-virginica	131.500000	...	1.600000
6.2	Iris-versicolor	83.500000	...	1.400000
	Iris-virginica	138.000000	...	2.050000
6.3	Iris-versicolor	72.666667	...	1.466667
	Iris-virginica	124.500000	...	1.983333
6.4	Iris-versicolor	63.500000	...	1.400000
	Iris-virginica	125.600000	...	2.060000

6.5	Iris-versicolor	55.000000	...	1.500000
	Iris-virginica	120.250000	...	2.000000
6.6	Iris-versicolor	67.500000	...	1.350000
6.7	Iris-versicolor	77.000000	...	1.533333
	Iris-virginica	133.200000	...	2.220000
6.8	Iris-versicolor	77.000000	...	1.400000
	Iris-virginica	128.500000	...	2.200000
6.9	Iris-versicolor	53.000000	...	1.500000
	Iris-virginica	134.333333	...	2.233333
7.0	Iris-versicolor	51.000000	...	1.400000
7.1	Iris-virginica	103.000000	...	2.100000
7.2	Iris-virginica	122.000000	...	1.966667
7.3	Iris-virginica	108.000000	...	1.800000
7.4	Iris-virginica	131.000000	...	1.900000
7.6	Iris-virginica	106.000000	...	2.100000
7.7	Iris-virginica	124.000000	...	2.200000
7.9	Iris-virginica	132.000000	...	2.000000

[57 rows x 4 columns]

#Latihan (23)

#Hitung frekuensi pada kolom 'Species' dengan menggunakan metode value_counts()

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

```
Iris-virginica    50
Iris-setosa       50
Iris-versicolor   50
Name: Species, dtype: int64
```

#Latihan (24)

#Perhitungan frekuensi pada kolom 'Species' dengan menggunakan metode value_counts() dalam bentuk dataframe

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

```
Iris-virginica    50
Iris-setosa       50
Iris-versicolor   50
Name: Species, dtype: int64
```

#Latihan (25)

Hitung frekuensi pada kolom 'PetalLenghCm' dengan menggunakan metode value_counts()

```
dataset['PetalLengthCm'].value_counts()
```

```
1.5    14
1.4    12
5.1     8
```

4.5	8
1.3	7
1.6	7
5.6	6
4.0	5
4.9	5
4.7	5
4.8	4
1.7	4
4.4	4
4.2	4
5.0	4
4.1	3
5.5	3
4.6	3
6.1	3
5.7	3
3.9	3
5.8	3
1.2	2
1.9	2
6.7	2
3.5	2
5.9	2
6.0	2
5.4	2
5.3	2
3.3	2
4.3	2
5.2	2
6.3	1
1.1	1
6.4	1
3.6	1
3.7	1
3.0	1
3.8	1
6.6	1
6.9	1
1.0	1

Name: PetalLengthCm, dtype: int64