

Nama: Satya Ananta Azis
NIM: 2112500901

Berikut hasil dari outputnya
Membaca file csv pada Pandas

```
import pandas as pd
df = pd.read_csv('Datartrwdki1317.csv')
print(df)
```

	tahun	nama_provinsi	nama_kabupaten_kota	nama_kecamatan	\
0	2013	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU UTR	
1	2013	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU UTR	
2	2013	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU UTR	
3	2013	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU SLT	
4	2013	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU SLT	
...
1330	2017	PROVINSI DKI JAKARTA	JAKARTA TIMUR	CIPAYUNG	
1331	2017	PROVINSI DKI JAKARTA	JAKARTA TIMUR	CIPAYUNG	
1332	2017	PROVINSI DKI JAKARTA	JAKARTA TIMUR	CIPAYUNG	
1333	2017	PROVINSI DKI JAKARTA	JAKARTA TIMUR	CIPAYUNG	
1334	2017	PROVINSI DKI JAKARTA	JAKARTA TIMUR	CIPAYUNG	
	nama_kelurahan	jumlah_rt	jumlah_rw		
0	P. PANGGANG	29	5		
1	P. KELAPA	31	5		
2	P. HARAPAN	15	3		
3	P. UNTUNG JAWA	9	3		
4	P. TIDUNG	29	4		
...		
1330	MUNJUL	75	8		
1331	SETU	44	6		
1332	BAMBU APUS	65	5		
1333	LUBANG BUAYA	117	12		
1334	CEGER	39	5		

Melihat/mengecek data

```
df.head()
```

	tahun	nama_provinsi	nama_kabupaten_kota	nama_kecamatan	nama_kelurahan	jumlah_rt	jumlah_rw
0	2013	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU UTR	P. PANGGANG	29	5
1	2013	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU UTR	P. KELAPA	31	5
2	2013	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU UTR	P. HARAPAN	15	3
3	2013	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU SLT	P. UNTUNG JAWA	9	3
4	2013	PROVINSI DKI JAKARTA	KAB.ADM.KEP.SERIBU	KEP. SERIBU SLT	P. TIDUNG	29	4

```
df.tail()
```

	tahun	nama_provinsi	nama_kabupaten_kota	nama_kecamatan	nama_kelurahan	jumlah_rt	jumlah_rw
1330	2017	PROVINSI DKI JAKARTA	JAKARTA TIMUR	CIPAYUNG	MUNJUL	75	8
1331	2017	PROVINSI DKI JAKARTA	JAKARTA TIMUR	CIPAYUNG	SETU	44	6
1332	2017	PROVINSI DKI JAKARTA	JAKARTA TIMUR	CIPAYUNG	BAMBU APUS	65	5
1333	2017	PROVINSI DKI JAKARTA	JAKARTA TIMUR	CIPAYUNG	LUBANG BUAYA	117	12
1334	2017	PROVINSI DKI JAKARTA	JAKARTA TIMUR	CIPAYUNG	CEGER	39	5

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1335 entries, 0 to 1334
Data columns (total 7 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   tahun              1335 non-null   int64
 1   nama_provinsi       1335 non-null   object
 2   nama_kabupaten_kota 1335 non-null   object
 3   nama_kecamatan      1335 non-null   object
 4   nama_kelurahan      1335 non-null   object
 5   jumlah_rt          1335 non-null   int64
 6   jumlah_rw          1335 non-null   int64
dtypes: int64(3), object(4)
memory usage: 73.1+ KB
```

df.describe()

	tahun	jumlah_rt	jumlah_rw
count	1335.000000	1335.000000	1335.000000
mean	2015.000000	114.695131	10.233708
std	1.414744	51.643694	4.078786
min	2013.000000	9.000000	2.000000
25%	2014.000000	79.000000	7.000000
50%	2015.000000	110.000000	10.000000
75%	2016.000000	145.000000	13.000000
max	2017.000000	280.000000	26.000000

Menampilkan data unik(distinct dalam SQL) kolom kabupaten_kota :

```
import pandas as pd
df = pd.read_csv('Dataatrwdki1317.csv')
available_kabkota = df['nama_kabupaten_kota'].unique()
print(available_kabkota)
```

['KAB.ADM.KEP.SERIBU' 'JAKARTA PUSAT' 'JAKARTA UTARA' 'JAKARTA BARAT'
'JAKARTA SELATAN' 'JAKARTA TIMUR']

Group By: split-apply-combine

```
import pandas as pd
df = pd.read_csv('Dataatrwdki1317.csv')
count_kabkota = df.groupby(['nama_kabupaten_kota']).size().reset_index(name='counts')
print(count_kabkota)
```

	nama_kabupaten_kota	counts
0	JAKARTA BARAT	280
1	JAKARTA PUSAT	220
2	JAKARTA SELATAN	325
3	JAKARTA TIMUR	325
4	JAKARTA UTARA	155
5	KAB.ADM.KEP.SERIBU	30

Untuk penjumlahan suatu kolom terhadap kolom tertentu, semisal kita ingin total pada kolom jumlah_rt terhadap kolom nama_kabupaten_kota, nama_kecamatan, nama_kelurahan

```
rt = df.groupby(['nama_kabupaten_kota', 'nama_kecamatan', 'nama_kelurahan'])['jumlah_rt'].sum().reset_index()
print(rt)
```

	nama_kabupaten_kota	nama_kecamatan	nama_kelurahan	jumlah_rt
0	JAKARTA BARAT	CENGKARENG	CENGKARENG BARAT	906
1	JAKARTA BARAT	CENGKARENG	CENGKARENG TIMUR	1132
2	JAKARTA BARAT	CENGKARENG	DURI KOSAMBI	843
3	JAKARTA BARAT	CENGKARENG	KAPUK	1110
4	JAKARTA BARAT	CENGKARENG	KEDAUNG KALI ANGKE	494
...
262	KAB.ADM.KEP.SERIBU	KEP. SERIBU SLT	P. TIDUNG	145
263	KAB.ADM.KEP.SERIBU	KEP. SERIBU SLT	P. UNTUNG JAWA	45
264	KAB.ADM.KEP.SERIBU	KEP. SERIBU UTR	P. HARAPAN	75
265	KAB.ADM.KEP.SERIBU	KEP. SERIBU UTR	P. KELAPA	155
266	KAB.ADM.KEP.SERIBU	KEP. SERIBU UTR	P. PANGGANG	145

[267 rows x 4 columns]

Iris.csv

Cek Instalasi

```
import pandas
import sklearn
import scipy
import numpy
import matplotlib

pandas.__version__, sklearn.__version__, scipy.__version__, numpy.__version__, matplotlib.__version__

('1.5.3', '1.2.2', '1.11.4', '1.23.5', '3.7.1')
```

Loading package

```
import matplotlib.pyplot as plt
import numpy
from pandas.plotting import scatter_matrix
from sklearn import model_selection
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
from sklearn import datasets
```

LOADING DATA DARI PACKAGE SKLEARN

```
iris?
```

Type: Bunch
String form: {'data': array([[5.1, 3.5, 1.4, 0.2], [4.9, 3. , 1.4, 0.2], [4.7, 3.2, 1.3, 0.2], <...> length (cm)', 'petal width (cm)'], 'filename': 'iris.csv', 'data_module': 'sklearn.datasets.data'})
Length: 8
File: /usr/local/lib/python3.10/dist-packages/sklearn/utils/_bunch.py
Docstring: Container object exposing keys as attributes.
Bunch objects are sometimes used as an output for functions and methods.
They extend dictionaries by enabling values to be accessed by key, 'bunch["value_key"]', or by an attribute, 'bunch.value_key'.
Examples

>>> from sklearn.utils import Bunch
>>> b = Bunch(a=1, b=2)
>>> b['b']
2

Menampilkan Nama Field dalam dataset iris

```
[13] iris?
```

```
iris.keys()
dict_keys(['data', 'target', 'target_names', 'DESCR', 'feature_names'])
print(iris.DESCR)

Index(['sepal.length', 'sepal.width', 'petal.length', 'petal.width',
       'variety'],
      dtype='object')
```

Menampilkan deskripsi dataset

```
print(iris.DESCR)

**Data Set Characteristics:**

: Number of Instances: 150 (50 in each of three classes)
: Number of Attributes: 4 numeric, predictive attributes and the class
: Attribute Information:
  - sepal length in cm
  - sepal width in cm
  - petal length in cm
  - petal width in cm
  - class:
    - Iris-Setosa
    - Iris-Versicolour
    - Iris-Virginica

: Summary Statistics:

=====
              Min  Max   Mean  SD   Class Correlation
=====
sepal length:  4.3  7.9   5.84   0.83    0.7826
sepal width:   2.0  4.4   3.05   0.43   -0.4194
petal length:   1.0  6.9   3.76   1.76    0.9490 (high!)
petal width:   0.1  2.5   1.20   0.76    0.9565 (high!)
=====
```

Convert dataset ke dalam Pandas DataFrame

```
dataset = pandas.DataFrame(data= numpy.c_[iris['data'], iris['target']],
                           columns= iris['feature_names'])
dataset['class'] = numpy.c_[list(map((lambda x: iris.target_names[x]), iris.target))]
```

Eksplorasi dataset

```
print (dataset.shape)

(150, 5)
```

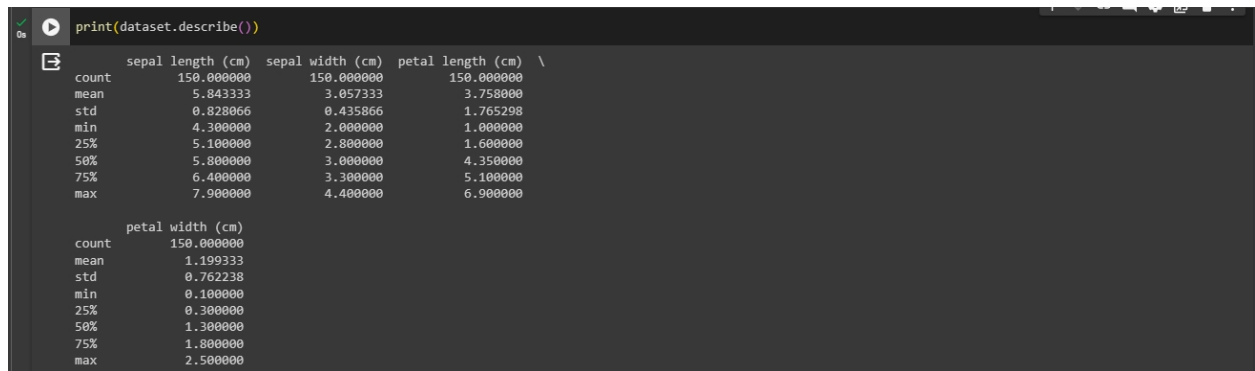
Fungsi head

```
print(dataset.head(5))

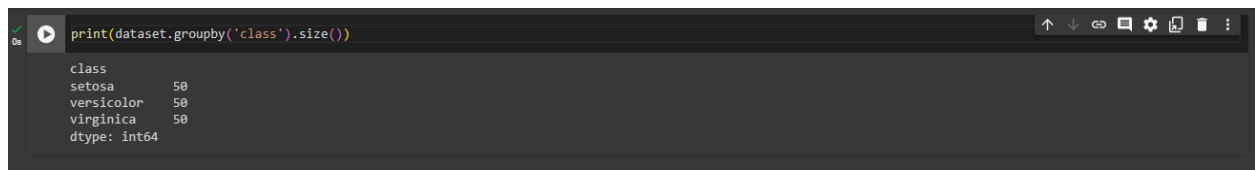
   sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (cm)  \
0                5.1                3.5                1.4                0.2
1                4.9                3.0                1.4                0.2
2                4.7                3.2                1.3                0.2
3                4.6                3.1                1.5                0.2
4                5.0                3.6                1.4                0.2

   class
0  setosa
1  setosa
2  setosa
3  setosa
4  setosa
```

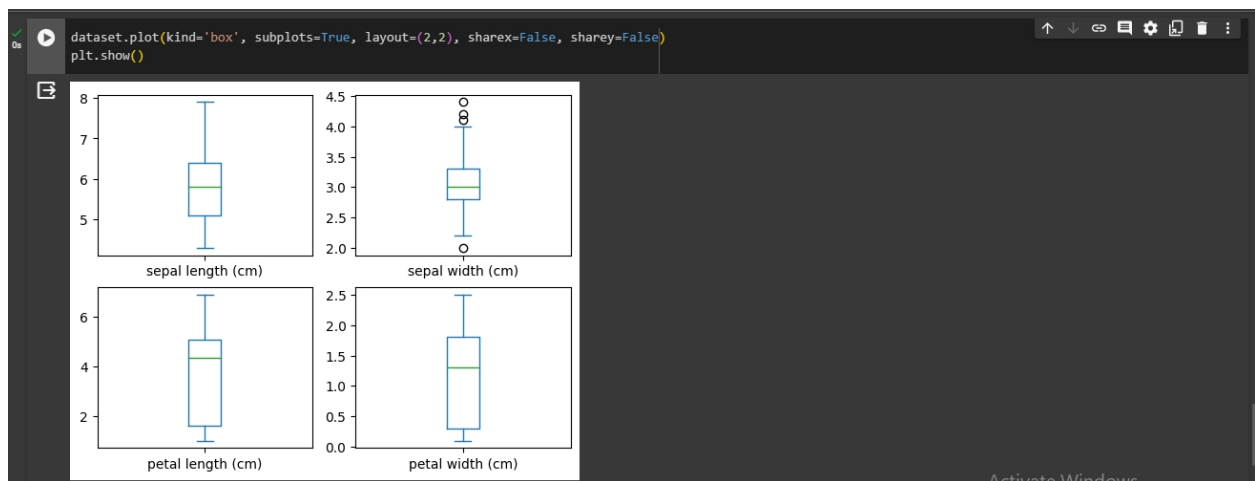
Statistical summary

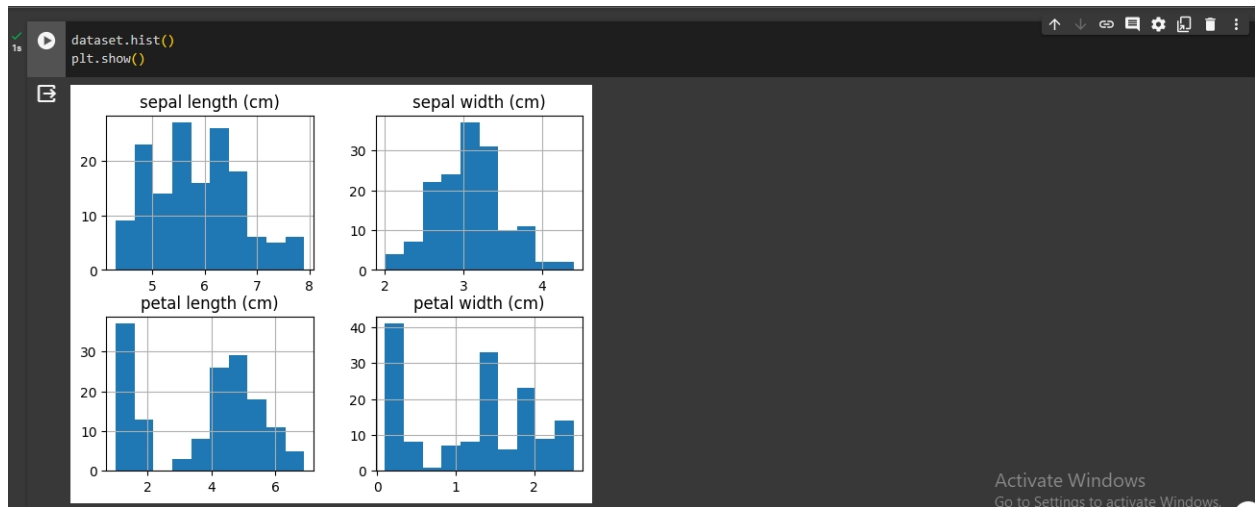


Distribusi data



Visualisasi Data dalam bentuk boxplot





Interaksi Antar Variabel dengan scatterplot

