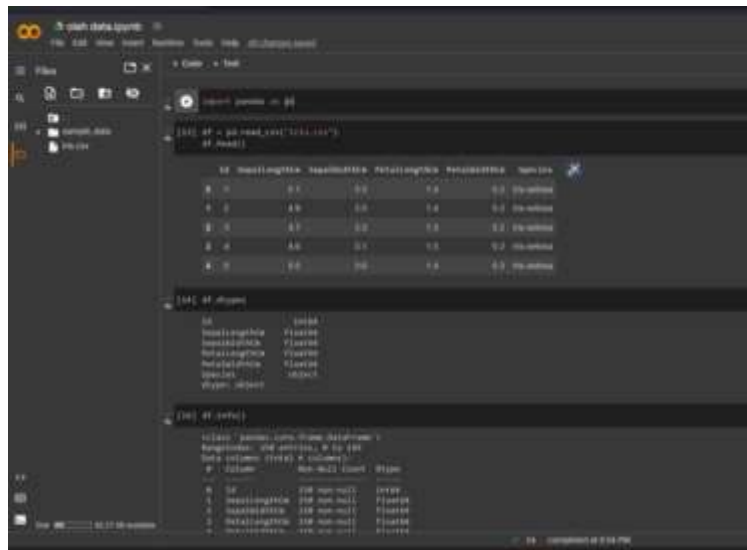


Nama : Novitasari

Nim : 20.01.013.012

Kelas : Python\_D

## 1. Mengolah data iris Database



```
import pandas as pd

df = pd.read_csv('iris.csv')
df.head()

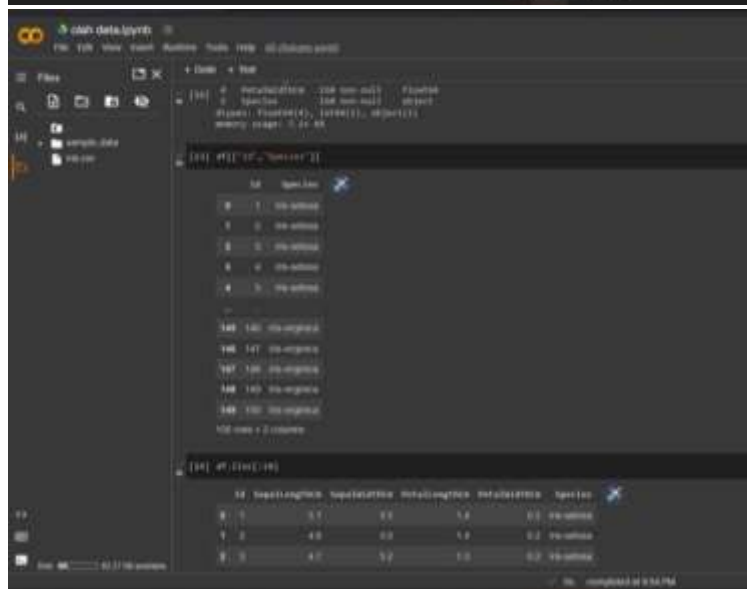
df.info()

df.describe()
```

	sepal.length	sepal.width	petal.length	petal.width	species
0	5.1	3.5	1.4	0.2	Setosa
1	4.9	3.0	1.4	0.2	Setosa
2	5.0	3.2	1.3	0.2	Setosa
3	4.8	3.1	1.6	0.2	Setosa
4	5.1	3.6	1.4	0.2	Setosa

```
>>> df.columns
Out[1]: Index(['sepal.length', 'sepal.width', 'petal.length', 'petal.width', 'species'], dtype=object)
```

```
>>> df.info()
Out[1]:
Int64Index: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   column             non-null count  dtype
 0   sepal.length       150 non-null float64
 1   sepal.width        150 non-null float64
 2   petal.length       150 non-null float64
 3   petal.width        150 non-null float64
 4   species            150 non-null object
dtypes: float64(4), object(1)
memory usage: 9.2+ KB
```



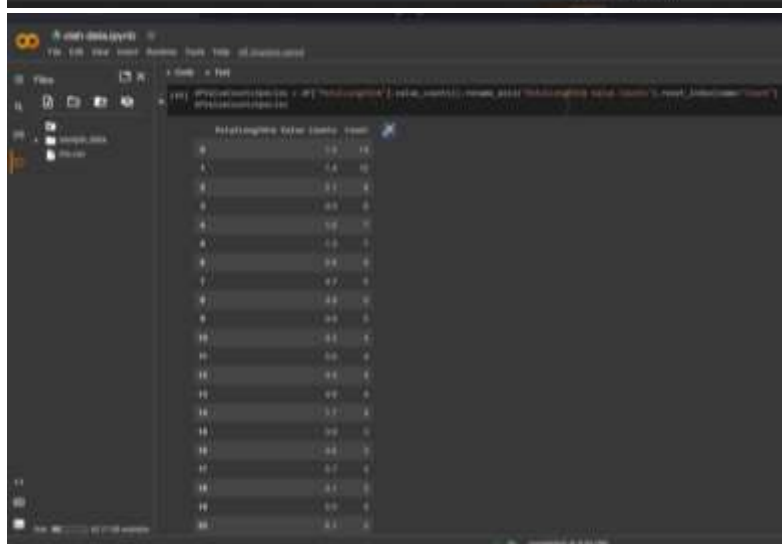
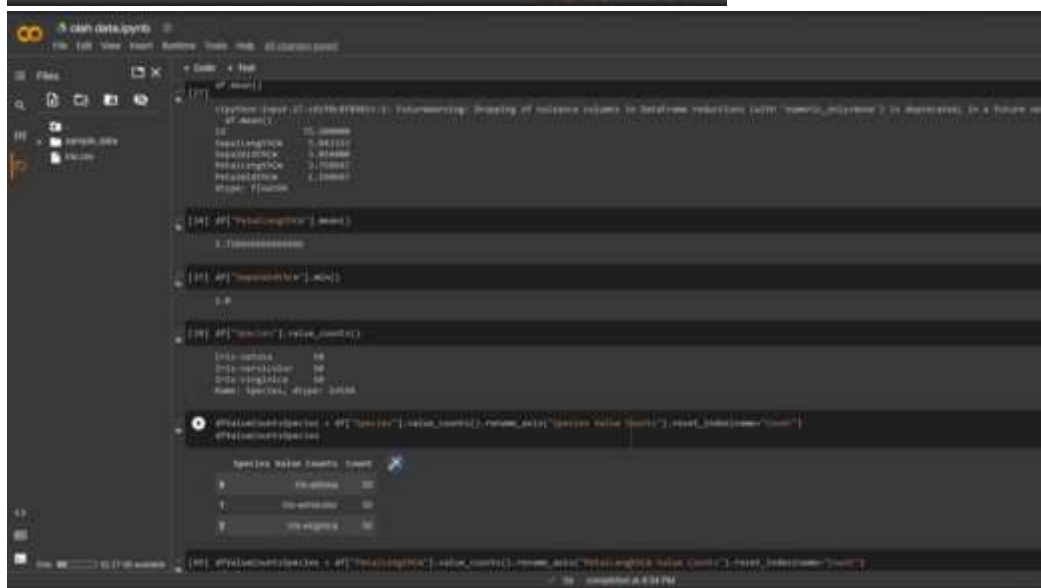
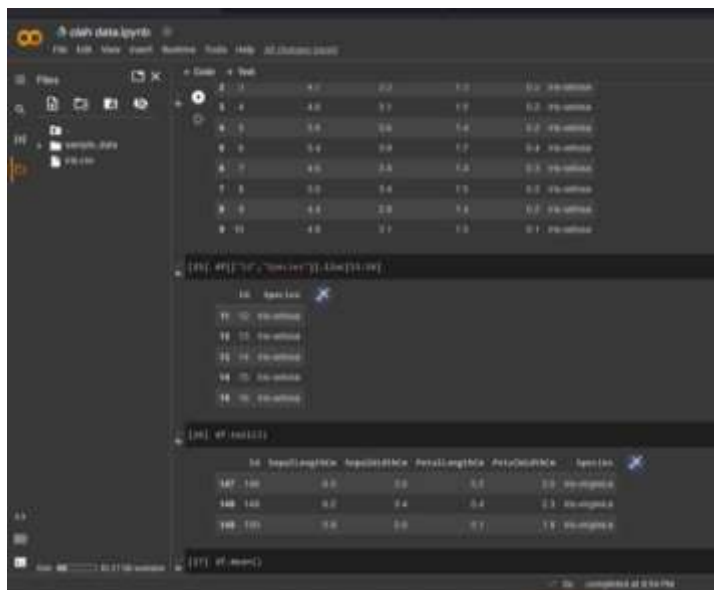
```
df[df['species'] == 'Setosa']

df.groupby('species').mean()
```

	sepal.length	sepal.width	petal.length	petal.width
Setosa	5.006	3.428	1.462	0.246
versicolour	6.762	3.773	4.353	1.192
virginica	7.003	3.293	5.552	2.026

```
df.groupby('species').agg({'sepal.length': 'mean', 'sepal.width': 'mean', 'petal.length': 'mean', 'petal.width': 'mean'})
```

species	sepal.length	sepal.width	petal.length	petal.width
Setosa	5.006	3.428	1.462	0.246
versicolour	6.762	3.773	4.353	1.192
virginica	7.003	3.293	5.552	2.026





olah data iris(unclean).ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- iris\_unclean.csv

```
[5] dfDataBaru = df['SepallengthCm'].fillna(df['SepallengthCm'].mean())
```

dfDataBaru

0	5.856757
1	4.900000
2	4.700000
3	4.600000
4	5.000000
...	...
145	6.700000
146	6.300000
147	6.500000
148	6.200000
149	5.900000

Name: SepallengthCm, Length: 150, dtype: float64

```
df2 = pd.DataFrame({'SepallengthCm': dfDataBaru, 'SepalWidthCm': df['SepalWidthCm'],  
                  'PetallengthCm': df['PetallengthCm'], 'PetalWidthCm': df['PetalWidthCm'],  
                  'Species': df['Species']})
```

df2

Disk 82.27 GB available

0s completed at 10:58 PM

olah data iris(unclean).ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- iris\_unclean.csv

```
df2
```

	SepallengthCm	SepalWidthCm	PetallengthCm	PetalWidthCm	Species
0	5.856757	3.5	1.4	0.2	Iris-setosa
1	4.900000	2000.0	1.4	0.2	Iris-setosa
2	4.700000	3.2	-1.3	0.2	Iris-setosa
3	4.600000	3.1	1.5	0.2	Iris-setosa
4	5.000000	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...
145	6.700000	3.0	5.2	2.3	Iris-virginica
146	6.300000	2.5	5.0	1.9	Iris-virginica
147	6.500000	3.0	5.2	2.0	Iris-virginica
148	6.200000	3.4	5.4	2.3	Iris-virginica
149	5.900000	3.0	5.1	1.8	Iris-virginica

150 rows x 5 columns

Disk 82.27 GB available

0s completed at 10:58 PM

olah data iris(unclean).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- iris\_unclean.csv

```
[148] 6.200000    3.4    0.4    2.3  Iris-virginica
[149] 5.900000    3.0    0.1    1.8  Iris-virginica
150 rows x 5 columns
```

```
[146] df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  --
 0   SepalLengthCm 150 non-null    float64
 1   SepalWidthCm  150 non-null    float64
 2   PetalLengthCm 150 non-null    float64
 3   PetalWidthCm  150 non-null    float64
 4   Species        150 non-null    object  
dtypes: float64(4), object(1)
memory usage: 8.8+ KB
```

```
df2.isna().sum()
SepalLengthCm    0
```

0s completed at 10:58 PM

olah data iris(unclean).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

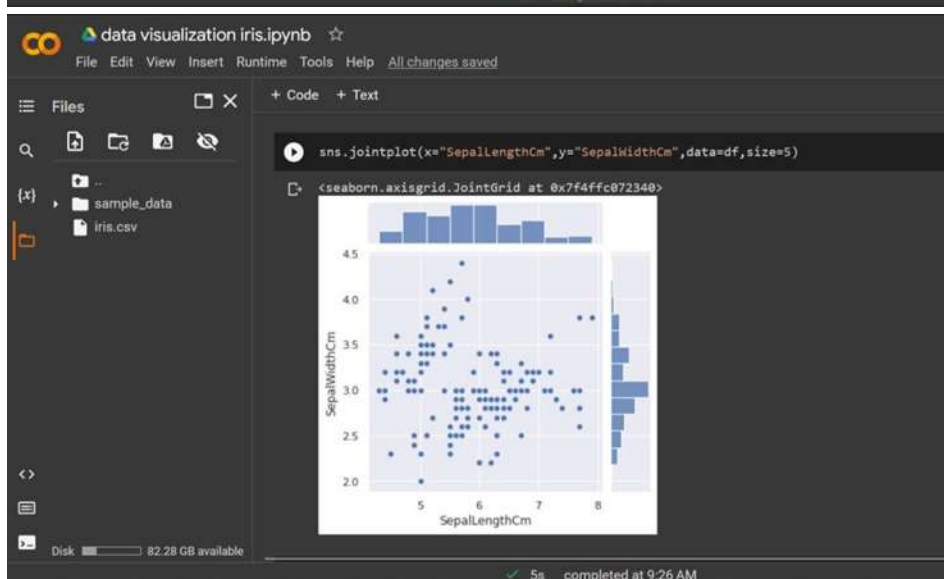
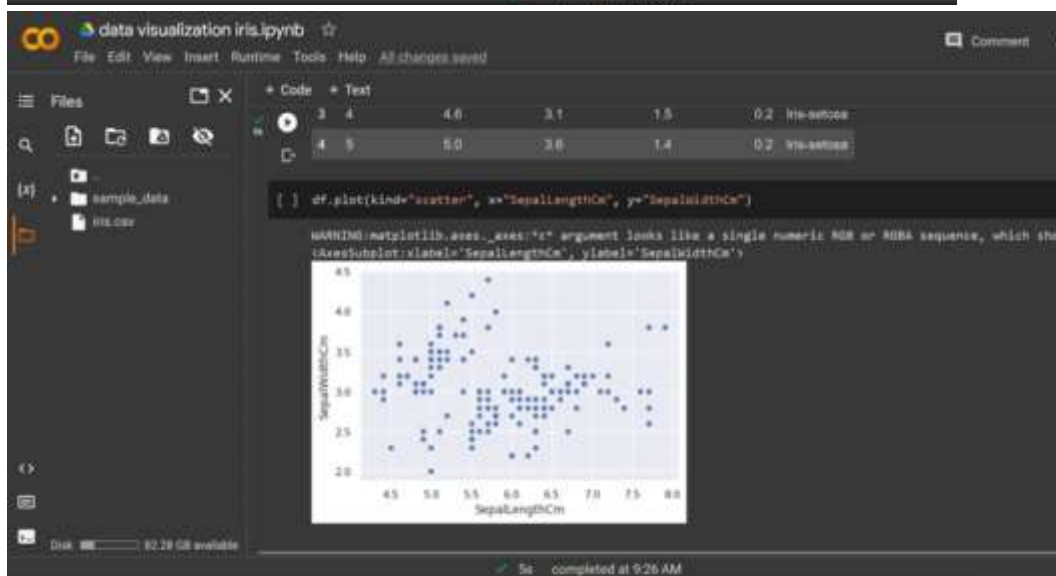
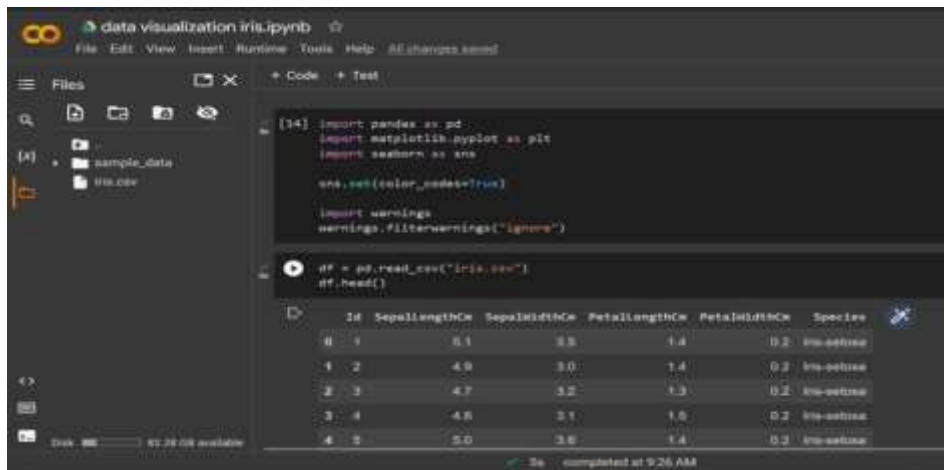
- sample\_data
- iris\_unclean.csv

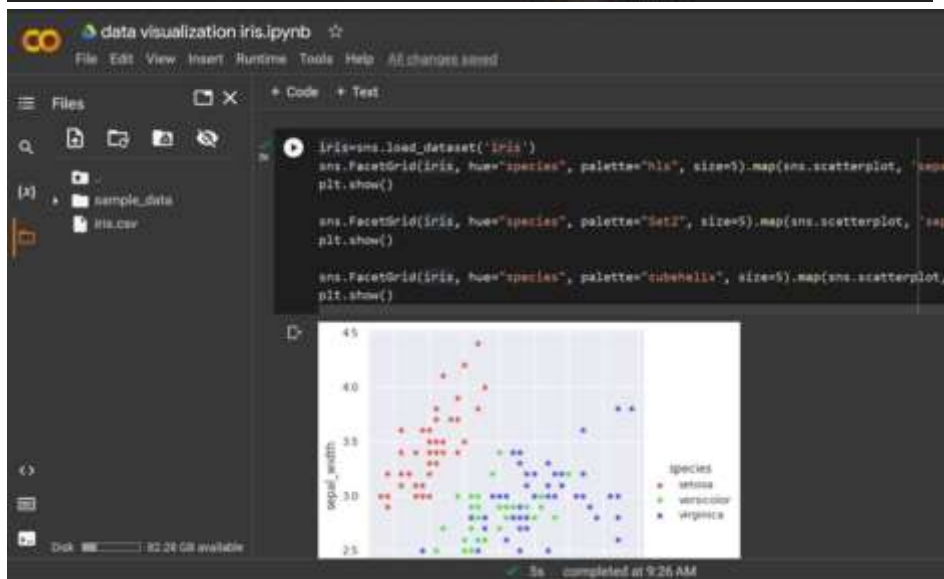
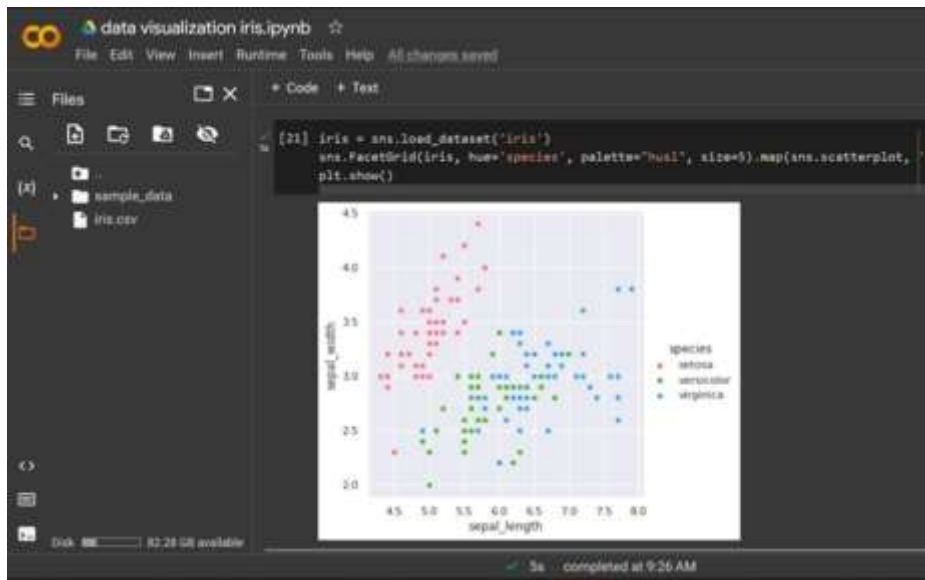
```
[16] 2   PetalLengthCm 150 non-null    float64
      3   PetalWidthCm  150 non-null    float64
      4   Species        150 non-null    object  
dtypes: float64(4), object(1)
memory usage: 0.0+ KB
```

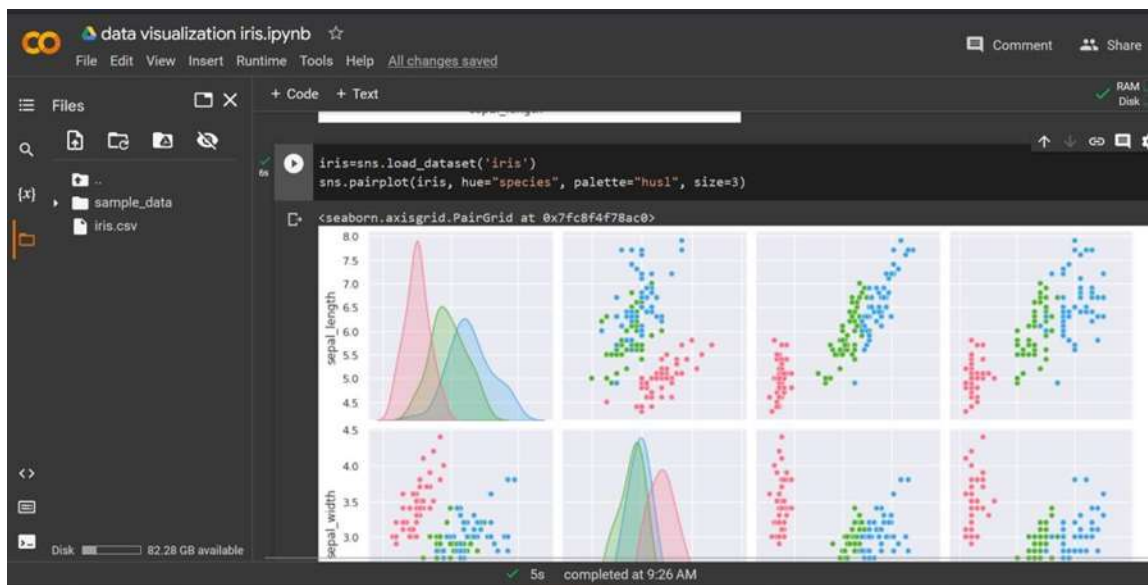
```
df2.isna().sum()
SepalLengthCm    0
SepalWidthCm     0
PetalLengthCm    0
PetalWidthCm     0
Species          0
dtype: int64
```

0s completed at 10:58 PM

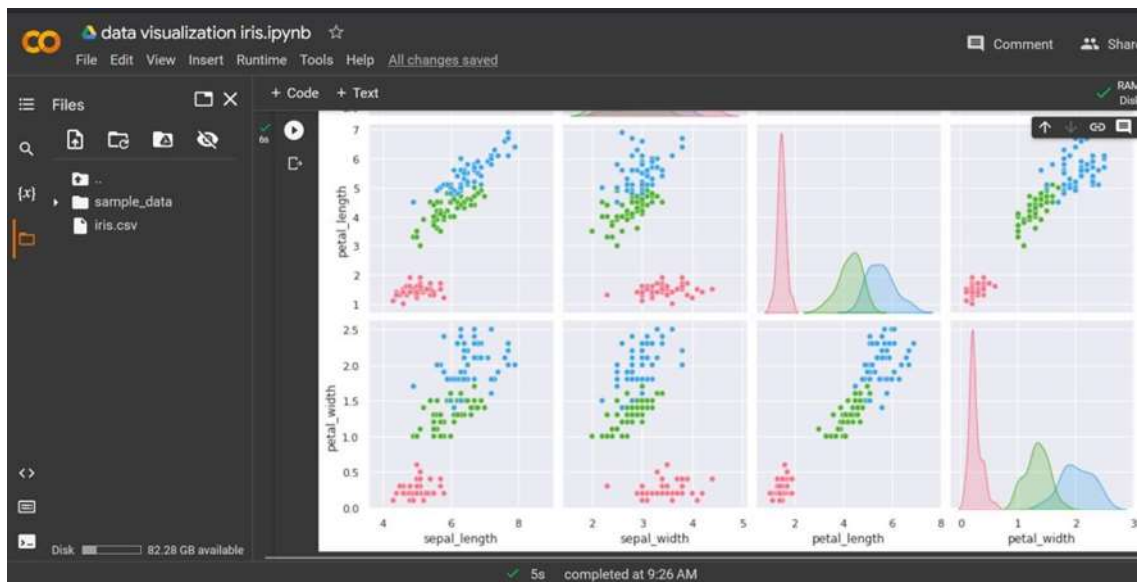
### 3. Data Visualization Iris











#### 4. visualisasi data (pandas dataframe)

- Series

```
#series
import pandas as pd
data = [1,2,3,4]
series1 = pd.Series(data)
series1

0    1
1    2
2    3
3    4
dtype: int64

[6] type(series1)

pandas.core.series.Series

[10] series1 = pd.Series(data, index=["a","b","c","d"])
series1

a    1
b    2
c    3
d    4
dtype: int64
```

The figure displays a Jupyter Notebook titled "visualisasi data (pandas dataframe).ipynb". The interface shows a file explorer on the left with a folder named "sample\_data" containing "sampledata.csv". The main area contains code cells for creating a pandas Series. The first code cell defines a list "data = [1, 2, 3, 4]" and creates a Series "series1 = pd.Series(data)". The output shows the Series with integer values and dtype "int64". The second code cell shows the type of "series1" as "pandas.core.series.Series". The third code cell shows the Series with string indices "a", "b", "c", "d" and integer values, with dtype "int64". The status bar at the bottom indicates "83.20 GB available" and "completed at 12:35 PM".

- DataFrame

The screenshot shows a Jupyter Notebook titled "visualisasi data (pandas dataframe).ipynb". The left sidebar displays a file explorer with a folder named "sample\_data" containing a file "sampledataok.csv". The main code area contains two code cells. The first cell creates a DataFrame from a list: 

```
import pandas as pd
data = [1,2,3,4]
df = pd.DataFrame(data)
df
```

. The output is a table with a single column of values 1, 2, 3, and 4. The second cell creates a DataFrame from a dictionary: 

```
dictionary = {'buah': ['Apel', 'Jeruk', 'Lemon'], 'jumlah': [10, 5, 12]}
df = pd.DataFrame(dictionary)
df
```

. The output is a table with two columns: "buah" and "jumlah", containing the specified fruit names and their counts.

buah	jumlah
0	1
1	2
2	3
3	4

	buah	jumlah
0	Apel	10
1	Jeruk	5
2	Lemon	12

The screenshot shows the same Jupyter Notebook interface. The first code cell is the same as in the previous image. The second code cell creates a DataFrame from a list of lists with custom index and columns: 

```
import pandas as pd
data = [['Berti', 90, 85, 95, 90.5],
        ['Qorygore', 80, 85, 90, 86.6],
        ['Bimo', 70, 75, 80, 78.5]]
index = [0, 1, 2]
kolom = ["Nama", "Tugas", "UTS", "UAS", "Rata-Rata"]
df = pd.DataFrame(data, index=index, kolom=kolom)
df
```

. The output is a table with five columns: "Nama", "Tugas", "UTS", "UAS", and "Rata-Rata", and three rows of student data.

	buah	jumlah
0	Apel	10
1	Jeruk	5
2	Lemon	12

	Nama	Tugas	UTS	UAS	Rata-Rata
0	Berti	90	85	95	90.5
1	Qorygore	80	85	90	86.6
2	Bimo	70	75	80	78.5

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

Code

```
[19]
      Name Tugas UTS UAS Rata-Rata
0      Bert  90  85  95      90.5
1  Gorygore  80  85  90      88.6
2      Bimo  70  75  80      78.5
```

```
[30]
import pandas as pd
nama = ['Berti', 'Gorygore', 'Bimo']
tugas = [90, 80, 70]
uts = [85, 85, 75]
uas = [95, 90, 80]
ratarata = [90.5, 88.6, 78.5]
df2 = pd.DataFrame({'Nama':nama, 'Tugas':tugas, 'UTS':uts, 'UAS':uas, 'Rata-Rata':ratarata})
df2
```

	Nama	Tugas	UTS	UAS	Rata-Rata
0	Berti	90	85	95	90.5
1	Gorygore	80	85	90	88.6
2	Bimo	70	75	80	78.5

0% completed at 12:35 PM

- Merge

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

Code

```
[31]
#merge
#data pertama
import pandas as pd
nama = ['Berti', 'Rynde', 'Ari']
tugas = [95, 90, 75]
jurusan = ['IP', 'SI', 'KA']
df3 = pd.DataFrame({'Nama':nama, 'Tugas':tugas, 'Jurusan':jurusan})
df3
```

	Nama	Tugas	Jurusan
0	Berti	95	IP
1	Rynde	90	SI
2	Ari	75	KA

```
[32]
#data kedua
import pandas as pd
nama = ['Berti', 'Rynde', 'Ryia']
uts = [85, 84, 70]
jurusan = ['IP', 'SI', 'SI']
df4 = pd.DataFrame({'Nama':nama, 'UTS':uts, 'Jurusan':jurusan})
df4
```

0% completed at 12:36 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

+ Code + Text

```
uts = [85,84,76]
jurusan = ['IF','SI','SI']
df4 = pd.DataFrame({'Nama':nama,'UTS':uts,'Jurusan':jurusan})
df4
```

	Nama	UTS	Jurusan
0	Berli	85	IF
1	Ryndes	84	SI
2	Rylo	76	SI

```
[33] df3.merge(df4)
```

	Nama	Tugas	Jurusan	UTS
0	Berli	95	IF	85
1	Ryndes	90	SI	84

```
[34] df3.merge(df4, on='Nama', how='left')
```

Disk 62.20 GB available

0s completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

+ Code + Text

```
[34] df3.merge(df4, on='Nama', how='left')
```

	Nama	Tugas	Jurusan_x	UTS	Jurusan_y
0	Berli	95	IF	85.0	IF
1	Ryndes	90	SI	84.0	SI
2	Arin	75	KA	NaN	NaN

```
df3.merge(df4, on='Nama', how='right')
```

	Nama	Tugas	Jurusan_x	UTS	Jurusan_y
0	Berli	95.0	IF	85	IF
1	Ryndes	90.0	SI	84	SI
2	Rylo	NaN	NaN	76	SI

```
[37] df3.merge(df4, on='Nama', how='outer')
```

Disk 62.20 GB available

0s completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

Code

```
df3.merge(df4, on='Nama', how='right')
```

	Nama	Tugas	Jurusan_x	UTS	Jurusan_y
0	Berli	95.0	IF	85	IF
1	Ryndes	90.0	SI	84	SI
2	Rylo	NaN	NaN	70	SI

Code

```
[37]: df3.merge(df4, on='Nama', how='outer')
```

	Nama	Tugas	Jurusan_x	UTS	Jurusan_y
0	Berli	95.0	IF	85.0	IF
1	Ryndes	90.0	SI	84.0	SI
2	Acin	75.0	KA	NaN	NaN
3	Rylo	NaN	NaN	70.0	SI

0s completed at 12:35 PM

- Join

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

Code

```
#join
data pertama
nama = ['Berli', 'Ryndes', 'Acin']
tugas = [95, 90, 75]
jurusan = ['IF', 'SI', 'KA']
df3 = pd.DataFrame({'Nama':nama, 'Tugas':tugas, 'Jurusan':jurusan}, index=['1.1', '1.2', '1.3'])
df3
```

	Nama	Tugas	Jurusan
1.1	Berli	95	IF
1.2	Ryndes	90	SI
1.3	Acin	75	KA

Code

```
[44]: data kedua
nama = ['Berli', 'Ryndes', 'Rylo']
uts = [85, 84, 70]
jurusan = ['IF', 'SI', 'SI']
df4 = pd.DataFrame({'Nama':nama, 'UTS':uts, 'Jurusan':jurusan}, index=['1.2', '1.3', '1.4'])
df4
```

0s completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

sample\_data  
sampledataok.csv

+ Code + Text

df3

	Nama B	UTS	Jurusan B
1.2	Bert	85	IF
1.3	Ryndes	84	SI
1.4	Rylo	70	SI

df3.join(df4, how='inner')

	Nama	Tugas	Jurusan	Nama B	UTS	Jurusan B
1.2	Ryndes	90	SI	Bert	85	IF
1.3	Ann	75	KA	Ryndes	84	SI

[46] df3.join(df4, how='right')

	Nama	Tugas	Jurusan	Nama B	UTS	Jurusan B
1.2	Ryndes	90.0	SI	Bert	85	IF
1.3	Ann	75.0	KA	Ryndes	84	SI

Disk 88% 81.20 GB available

completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

sample\_data  
sampledataok.csv

+ Code + Text

	Nama	Tugas	Jurusan	Nama B	UTS	Jurusan B
1.2	Ryndes	90.0	SI	Bert	85	IF
1.3	Ann	75.0	KA	Ryndes	84	SI
1.4	NaN	NaN	NaN	Rylo	70	SI

[47] df3.join(df4, how='left')

	Nama	Tugas	Jurusan	Nama B	UTS	Jurusan B
1.1	Bert	95	IF	NaN	NaN	NaN
1.2	Ryndes	90	SI	Bert	85.0	IF
1.3	Ann	75	KA	Ryndes	84.0	SI

[48] df3.join(df4, how='outer')

	Nama	Tugas	Jurusan	Nama B	UTS	Jurusan B
1.1	Bert	95.0	IF	NaN	NaN	NaN
1.2	Ryndes	90.0	SI	Bert	85.0	IF

Disk 88% 81.20 GB available

completed at 12:35 PM

- Concatenate

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

```
#concatenate
pd.concat([df3,df4], sort=False)
```

FutureWarning: Passing non boolean values for sort is deprecated and will be removed in a future version. Use `sort=False` instead.

	Jurusan	Jurusan B	Nama	Nama B	Tugas	UTS
1.1	IF	NaN	Bert	NaN	85.0	NaN
1.2	SI	NaN	Rynde	NaN	90.0	NaN
1.3	KA	NaN	Ain	NaN	75.0	NaN
1.2	NaN	IF	NaN	Bert	NaN	85.0
1.3	NaN	SI	NaN	Rynde	NaN	94.0
1.4	NaN	SI	NaN	Rylo	NaN	70.0

```
[53]: #import data csv
import pandas as pd
sample = pd.read_csv("sampledataok.csv")
sample
```

Disk 82.20 GB available

0s completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

```
[53]: #import data csv
import pandas as pd
sample = pd.read_csv("sampledataok.csv")
sample
```

Unnamed: 0	nama_youtuber	jenis_kelamin	umur	kategori	subscriber
0	Raditya Dika	L	34	Komedi	7000000
1	Statement Prod	L	29	Daily Vlog	120000
2	Arief Muhammad	L	28	Daily Vlog	3000000
3	Annisia Aziza	P	25	Food Travel	600000
4	Sarah Viloid	P	23	Gamer	2000000
5	MLI	L	30	Komedi	800000
6	Chandra Liow	L	26	Sketsa	3000000

```
[54]: sample.head()
```

Unnamed: 0	nama_youtuber	jenis_kelamin	umur	kategori	subscriber
0	Raditya Dika	L	34	Komedi	7000000
1	Statement Prod	L	29	Daily Vlog	120000
2	Arief Muhammad	L	28	Daily Vlog	3000000
3	Annisia Aziza	P	25	Food Travel	600000
4	Sarah Viloid	P	23	Gamer	2000000

Disk 82.20 GB available

0s completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

Code

```
[54] sample.head()
```

	Unnamed: 0	nana_youtuber	jenis_kelamin	umur	kategori	subscriber
0	0	Radiya Dika	L	34	Komedi	7000000
1	1	Statement Prod	L	29	Daily Vlog	120000
2	2	Anief Muhammad	L	28	Daily Vlog	3000000
3	3	Anissa Aziza	P	25	Food Travel	600000
4	4	Sarah Vlod	P	23	Gamer	2000000

```
[55] sample.tail()
```

	Unnamed: 0	nana_youtuber	jenis_kelamin	umur	kategori	subscriber
2	2	Anief Muhammad	L	28	Daily Vlog	3000000
3	3	Anissa Aziza	P	25	Food Travel	600000
4	4	Sarah Vlod	P	23	Gamer	2000000

Disk 82.20 GB available

0s completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

Code

```
sample.tail()
```

	Unnamed: 0	nana_youtuber	jenis_kelamin	umur	kategori	subscriber
2	2	Anief Muhammad	L	28	Daily Vlog	3000000
3	3	Anissa Aziza	P	25	Food Travel	600000
4	4	Sarah Vlod	P	23	Gamer	2000000
5	5	MLJ	L	30	Komedi	800000
6	6	Chandra Lioe	L	26	Sketsa	3000000

```
[56] sample.tail(3)
```

	Unnamed: 0	nana_youtuber	jenis_kelamin	umur	kategori	subscriber
4	4	Sarah Vlod	P	23	Gamer	2000000
5	5	MLJ	L	30	Komedi	800000
6	6	Chandra Lioe	L	26	Sketsa	3000000

Disk 82.20 GB available

0s completed at 12:35 PM



visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

```
[57] sample.shape
```

```
[7, 4]
```

```
[58] sample.mean()
```

```
<ipython-input-58-86cf36839a2>:1: FutureWarning: Dropping of nuisance columns in Data
sample.mean()
Unnamed: 0    3.000000e+00
umur          2.785714e+01
subscriber    2.300000e+06
dtype: float64
```

```
sample.median()
```

```
<ipython-input-59-8b21c98c6364>:1: FutureWarning: Dropping of nuisance columns in Data
sample.median()
Unnamed: 0      3.0
umur           25.0
subscriber  2000000.0
dtype: float64
```

Disk 82.20 GB available

completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

```
[60] sample.std()
```

```
<ipython-input-60-d251cfc37f6f>:1: FutureWarning: Dropping of nuisance columns in Data
sample.std()
Unnamed: 0    2.160247e+00
umur          3.625308e+00
subscriber    2.346174e+06
dtype: float64
```

```
sample.max()
```

```
Unnamed: 0      8
nama_youtuber    Statement Prod
jenis_kelamin    P
umur            34
kategori         Sketsa
subscriber       7000000
dtype: object
```

```
[62] sample.min()
```

```
Unnamed: 0      0
nama_youtuber    Annisa Aliza
jenis_kelamin    L
```

Disk 82.20 GB available

completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

```
[63] sample.count()
```

```
Unnamed: 0      7
nama_youtuber    7
jenis_kelamin    7
umur             7
kategori         7
subscriber       7
dtype: int64
```

```
sample.describe()
```

	Unnamed: 0	umur	subscriber
count	7.000000	7.000000	7.000000e+00
mean	3.000000	27.857143	2.300000e+06
std	2.160247	3.625308	2.346174e+06
min	0.000000	23.000000	1.200000e+05
25%	1.500000	25.500000	7.000000e+05
50%	3.000000	28.000000	2.000000e+06

Disk 82.20 GB available

completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

Code

```
75% 4.500000 29.500000 3.000000e+06
max 6.000000 34.000000 7.000000e+06
```

```
[65] sample = sample.rename(columns={'nama_youtuber':'Youtuber'})
sample
```

	Unnamed: 0	Youtuber	jenis_kelamin	umur	kategori	subscriber
0	0	Raditya Dika	L	34	Komedi	7000000
1	1	Statement Prod	L	29	Daily Vlog	120000
2	2	Arief Muhammad	L	28	Daily Vlog	3000000
3	3	Annisa Aziza	P	25	Food Travel	600000
4	4	Sarah Viloid	P	23	Gamer	2000000
5	5	MLI	L	30	Komedi	800000
6	6	Chandra Liow	L	26	Sketsa	3000000

```
[66] sample = sample.drop(columns=['jenis_kelamin'])
```

completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

Code

```
[66] sample = sample.drop(columns=['jenis_kelamin'])
sample
```

	Unnamed: 0	Youtuber	umur	kategori	subscriber
0	0	Raditya Dika	34	Komedi	7000000
1	1	Statement Prod	29	Daily Vlog	120000
2	2	Arief Muhammad	28	Daily Vlog	3000000
3	3	Annisa Aziza	25	Food Travel	600000
4	4	Sarah Viloid	23	Gamer	2000000
5	5	MLI	30	Komedi	800000
6	6	Chandra Liow	26	Sketsa	3000000

```
[67] sample.iloc[:,2]
```

0	34
1	29
2	28

completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

Code

```
[67] 1 29
      2 28
      3 25
      4 23
      5 30
      6 26
      Name: umur, dtype: int64
```

```
[68] sample.iloc[0:3,2]

      0 34
      1 29
      2 28
      Name: umur, dtype: int64
```

```
[69] sample.iloc[:,:]
```

	Unnamed: 0	Youtuber	umur	kategori	subscriber
0	0	Raditya Dika	34	Komedi	7000000
1	1	Statement Prod	29	Daily Vlog	120000
2	2	Arief Muhammad	28	Daily Vlog	3000000

0s completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

Code

```
[69] 1 1 Statement Prod 29 Daily Vlog 120000
      2 2 Arief Muhammad 28 Daily Vlog 3000000
      3 3 Annisa Aziza 25 Food Travel 600000
      4 4 Sarah Vlog 23 Gamer 2000000
      5 5 MLI 30 Komedi 800000
      6 6 Chandra Lioe 26 Sketsa 3000000
```

```
sample.iloc[3:,2:]
```

	umur	kategori	subscriber
3	25	Food Travel	600000
4	23	Gamer	2000000
5	30	Komedi	800000
6	26	Sketsa	3000000

```
[74] sample.loc[:, "Youtuber"]
```

0s completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

```
[74] sample.loc[:, "Youtuber"]
```

	Youtuber
0	Raditya Dika
1	Statement Prod
2	Arief Muhammad
3	Annisa Aziza
4	Sarah Viloid
5	MLI
6	Chandra Liow

Name: Youtuber, dtype: object

```
[84] sample.loc[0:3, "Youtuber"]
```

	Youtuber
0	Raditya Dika
1	Statement Prod
2	Arief Muhammad
3	Annisa Aziza

Name: Youtuber, dtype: object

```
[86] sample['subscriber']=1
```

sample

completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

```
sample['subscriber']=1
```

sample

	Unnamed: 0	Youtuber	umur	kategori	subscriber
0	0	Raditya Dika	34	Komedi	1
1	1	Statement Prod	29	Daily Vlog	1
2	2	Arief Muhammad	28	Daily Vlog	1
3	3	Annisa Aziza	25	Food Travel	1
4	4	Sarah Viloid	23	Game	1
5	5	MLI	30	Komedi	1
6	6	Chandra Liow	26	Sketsa	1

```
[87] sample.sort_values(by="kategori")
```

	Unnamed: 0	Youtuber	umur	kategori	subscriber
1	1	Statement Prod	29	Daily Vlog	1

completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

[87] sample.sort\_values(by='kategori')

Unnamed: 0	Youtuber	umur	kategori	subscriber	
1	1	Statement Prod	29	Daily Vlog	1
2	2	Arief Muhammad	28	Daily Vlog	1
3	3	Anissa Aziza	25	Food Travel	1
4	4	Sarah Vilod	23	Gamer	1
5	0	Raditya Dika	34	Komedi	1
6	5	MLI	30	Komedi	1
6	6	Chandra Liow	26	Sketsa	1

[88] sample.sort\_values(by='kategori', ascending=False)

Unnamed: 0	Youtuber	umur	kategori	subscriber	
6	6	Chandra Liow	26	Sketsa	1
0	0	Raditya Dika	34	Komedi	1

Disk 82.29 GB available 0s completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample\_data
- sampledataok.csv

[89] sample['umur'] > 28

Unnamed: 0	Youtuber	umur	kategori	subscriber	
6	6	Chandra Liow	26	Sketsa	1
0	0	Raditya Dika	34	Komedi	1
5	5	MLI	30	Komedi	1
4	4	Sarah Vilod	23	Gamer	1
3	3	Anissa Aziza	25	Food Travel	1
1	1	Statement Prod	29	Daily Vlog	1
2	2	Arief Muhammad	28	Daily Vlog	1

[89] sample['umur'] > 28

0	True
1	True
2	False
3	False
4	False
5	True
6	False

Name: umur, dtype: bool

Disk 82.20 GB available 0s completed at 12:35 PM

visualisasi data (pandas dataframe).ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

sample\_data

sampledata.csv

+ Code + Text

[90] filter1 = sample['umur'] > 28  
filterbaru = sample[filter1]  
filterbaru

Unnamed: 0	Youtuber	umur	kategori	subscriber	
0	0	Radhy Oka	34	Komedi	1
1	1	Statement Prod	29	Daily Vlog	1
5	5	MLI	30	Komedi	1

[91] filter2 = (sample['umur'] > 27) & (sample['kategori'] == 'Daily Vlog')  
filterbaru2 = sample[filter2]  
filterbaru2

Unnamed: 0	Youtuber	umur	kategori	subscriber	
1	1	Statement Prod	29	Daily Vlog	1
2	2	Anel Muhammad	28	Daily Vlog	1

Disk 82.20 GB available

0s completed at 12:35 PM