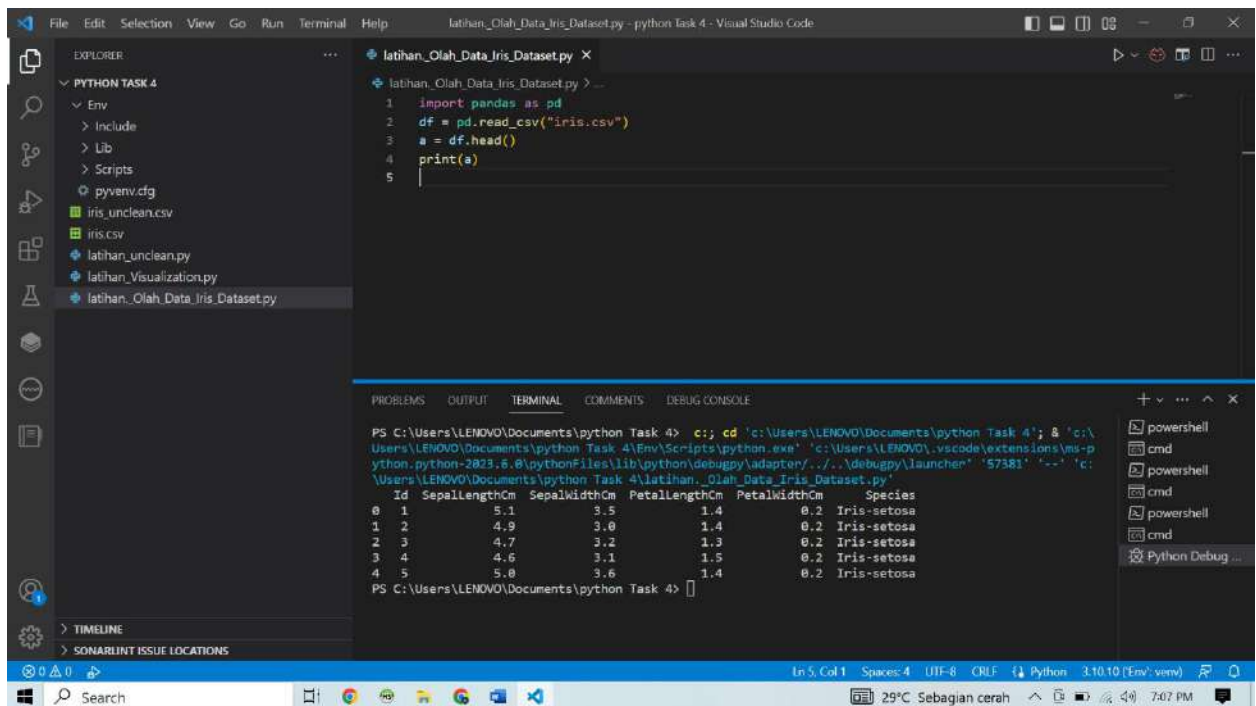


Nama : Ambar Wati
NIM : 20.01.013.001
Kelas : C
Mata Kuliah : Pemrograman Python

Task 5

❖ Olah Data Iris Dataset

1. Panggil file (iris.csv) lalu tampilkan 5 baris awal dataset dengan function head()



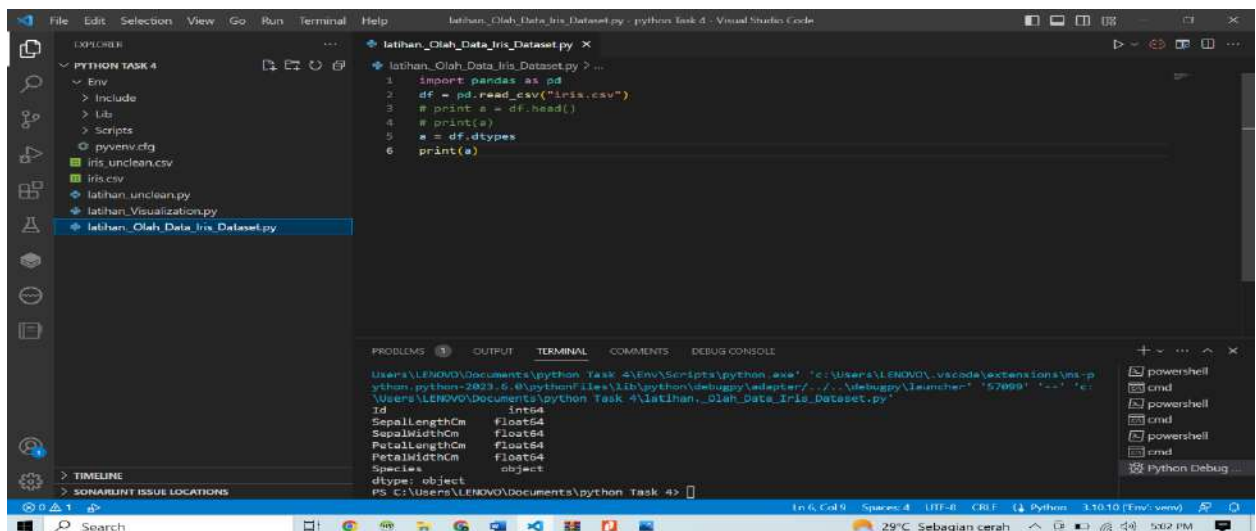
The screenshot shows a Visual Studio Code window with a file explorer on the left, a code editor in the center, and a terminal at the bottom. The code editor contains a Python script named `latihan_Olah_Data_Iris_Dataset.py` with the following code:

```
1 import pandas as pd
2 df = pd.read_csv("iris.csv")
3 a = df.head()
4 print(a)
5
```

The terminal shows the command prompt output, which includes the file path and the first 5 rows of the Iris dataset:

```
PS C:\Users\LENOVO\Documents\python Task 4> c:\; cd 'c:\Users\LENOVO\Documents\python Task 4'; & 'c:\Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe' 'c:\Users\LENOVO\.vscode\extensions\ms-pyhton.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57381' '--' 'c:\Users\LENOVO\Documents\python Task 4\latihan_Olah_Data_Iris_Dataset.py'
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
PS C:\Users\LENOVO\Documents\python Task 4>
```

2. Tampilkan tipe data dari kolom yang ada pada dataset menggunakan dtypes



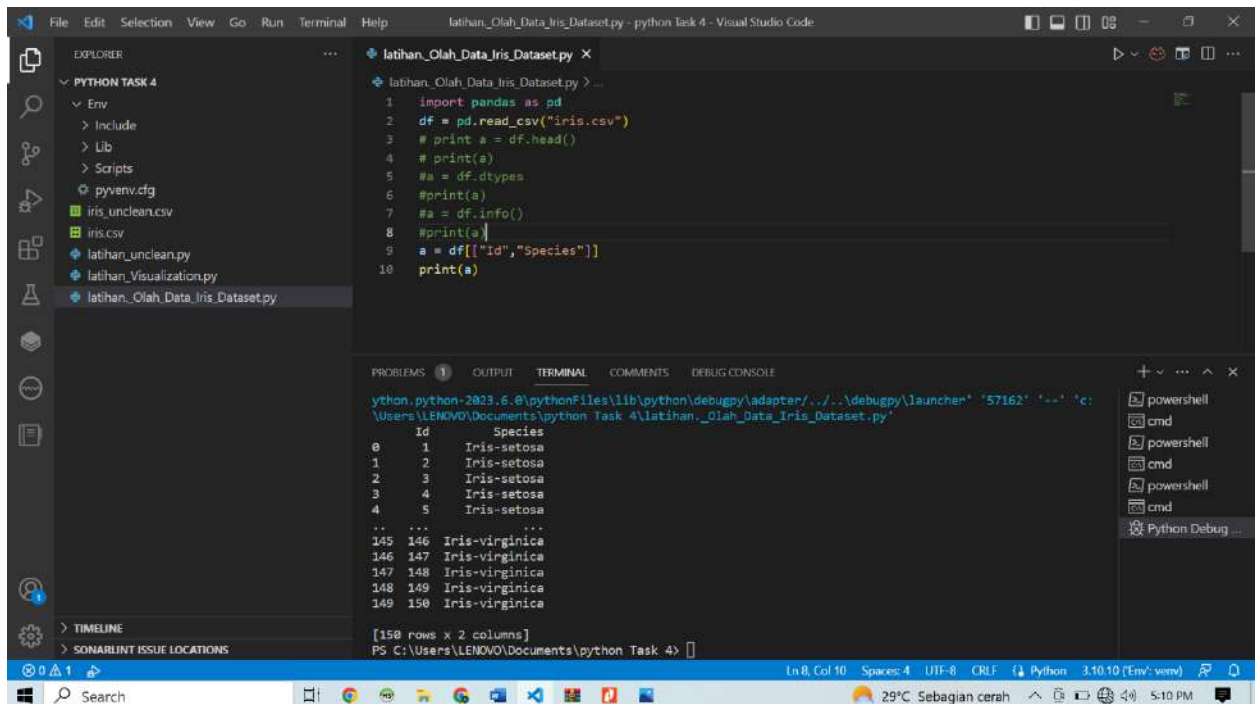
The screenshot shows a Visual Studio Code window with a file explorer on the left, a code editor in the center, and a terminal at the bottom. The code editor contains a Python script named `latihan_Olah_Data_Iris_Dataset.py` with the following code:

```
1 import pandas as pd
2 df = pd.read_csv("iris.csv")
3 # print(a = df.head())
4 # print(a)
5 a = df.dtypes
6 print(a)
```

The terminal shows the command prompt output, which includes the file path and the data types of the columns in the Iris dataset:

```
Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe 'c:\Users\LENOVO\.vscode\extensions\ms-pyhton.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57099' '--' 'c:\Users\LENOVO\Documents\python Task 4\latihan_Olah_Data_Iris_Dataset.py'
Id      int64
SepalLengthCm float64
SepalWidthCm  float64
PetalLengthCm float64
PetalWidthCm  float64
Species      object
dtype: object
PS C:\Users\LENOVO\Documents\python Task 4>
```

3. Tampilkan data untuk kolom “id” dan kolom “Species” dalam bentuk dataframe



The screenshot shows the Visual Studio Code interface with a Python file named `latihan_Olah_Data_Iris_Dataset.py`. The script reads an `iris.csv` file into a pandas DataFrame `df`, prints the first 5 rows, and then prints the columns 'Id' and 'Species' for the first 150 rows. The terminal output shows the first 5 rows of the dataset, which are all `Iris-setosa`.

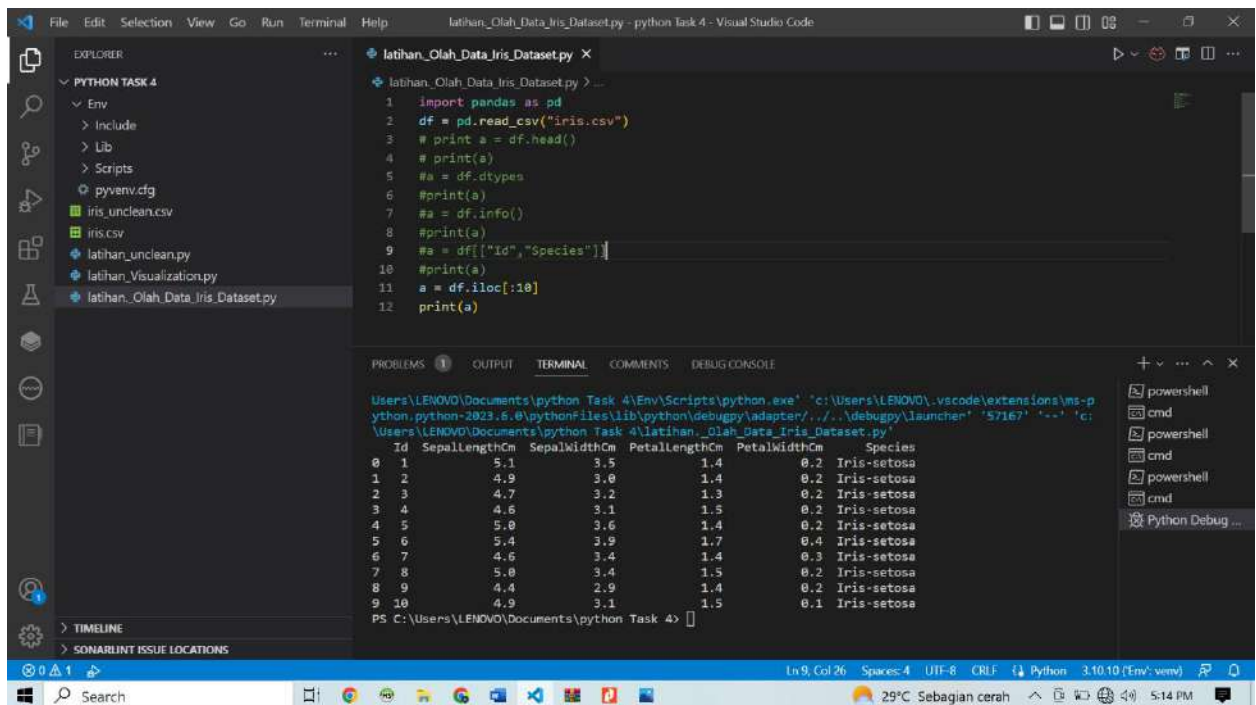
```
1 import pandas as pd
2 df = pd.read_csv("iris.csv")
3 # print a = df.head()
4 # print(a)
5 #a = df.dtypes
6 #print(a)
7 #a = df.info()
8 #print(a)
9 a = df[["Id", "Species"]]
10 print(a)
```

```
ython.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher '57162' '--' 'c:\Users\LENOVO\Documents\python Task 4\latihan_Olah_Data_Iris_Dataset.py'

   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]
PS C:\Users\LENOVO\Documents\python Task 4>
```

4. Tampilkan data baris indeks ke-0 (no1) sampai dengan indeks ke-9



The screenshot shows the Visual Studio Code interface with the same Python file. The script is modified to print the first 10 rows of the dataset using `df.iloc[:10]`. The terminal output shows the first 10 rows of the dataset, including the 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm', and 'Species' columns.

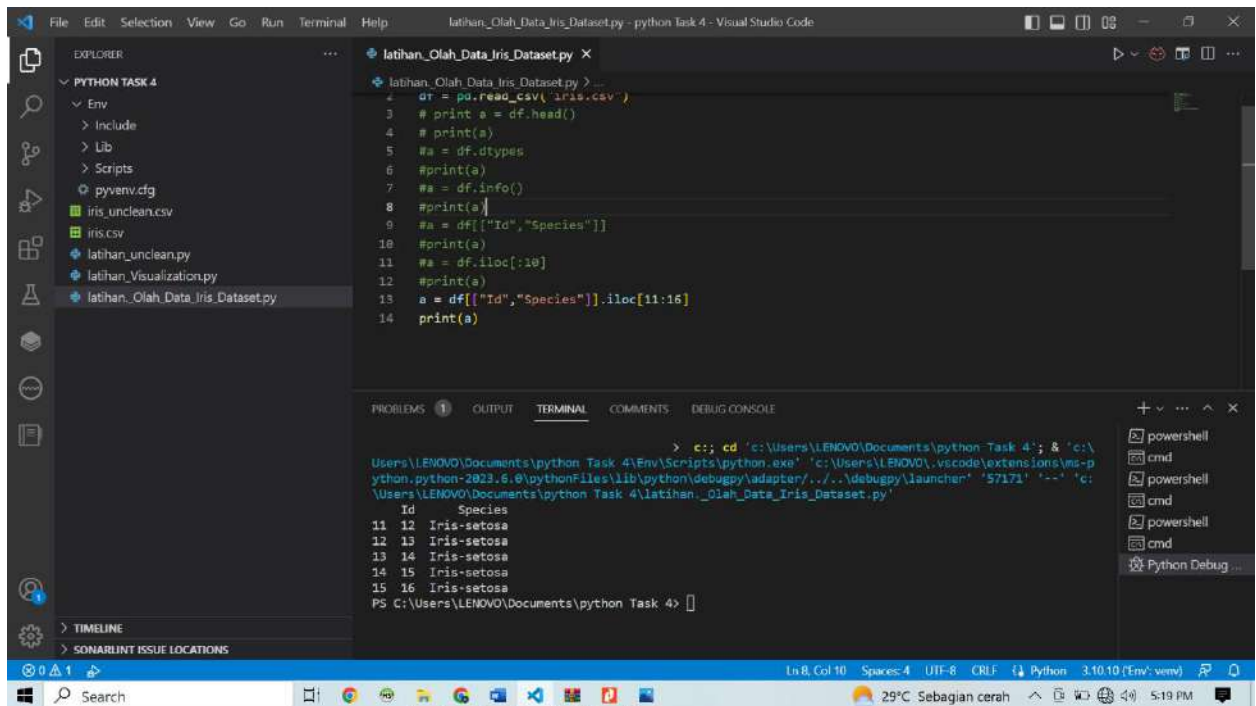
```
1 import pandas as pd
2 df = pd.read_csv("iris.csv")
3 # print a = df.head()
4 # print(a)
5 #a = df.dtypes
6 #print(a)
7 #a = df.info()
8 #print(a)
9 #a = df[["Id", "Species"]]
10 #print(a)
11 a = df.iloc[:10]
12 print(a)
```

```
Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe 'c:\Users\LENOVO\.vscode\extensions\ms-p
ython.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57167' '--' 'c:\
Users\LENOVO\Documents\python Task 4\latihan_Olah_Data_Iris_Dataset.py'

   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

PS C:\Users\LENOVO\Documents\python Task 4>
```

5. Tampilkan data hanya kolom “Id” dan kolom “Species”, dan yang di tampilkan adalah data indeks ke-11 sampai dengan indeks ke-15



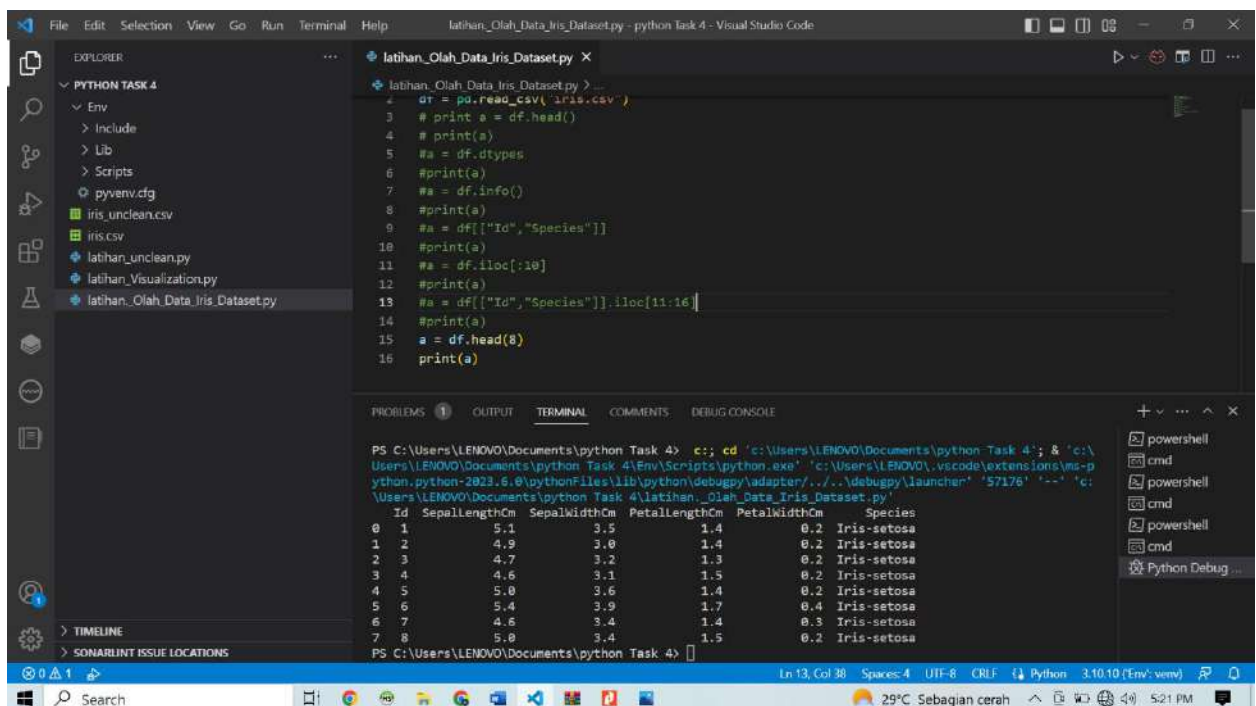
The screenshot shows a Visual Studio Code editor with a Python file named `latihan_Olah_Data_Iris_Dataset.py`. The script reads a CSV file, prints the first 10 rows, and then prints rows 11 to 15, specifically showing the 'Id' and 'Species' columns. The terminal output shows the execution of the script, displaying the first 10 rows and then the selected rows 11 to 15.

```
1 df = pd.read_csv('iris.csv')
2 # print a = df.head()
3 # print(a)
4 # print(a)
5 #a = df.dtypes
6 #print(a)
7 #a = df.info()
8 #print(a)
9 #a = df[["Id", "Species"]]
10 #print(a)
11 #a = df.iloc[:10]
12 #print(a)
13 a = df[["Id", "Species"]].iloc[11:16]
14 print(a)
```

```
> c:; cd 'c:\Users\LENOVO\Documents\python Task 4'; & 'c:\Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe' 'c:\Users\LENOVO\.vscode\extensions\ms-python.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57171' '--' 'c:\Users\LENOVO\Documents\python Task 4\latihan_Olah_Data_Iris_Dataset.py'

   Id  Species
11 12  Iris-setosa
12 13  Iris-setosa
13 14  Iris-setosa
14 15  Iris-setosa
15 16  Iris-setosa
PS C:\Users\LENOVO\Documents\python Task 4>
```

6. Tampilkan data 8 baris pertama



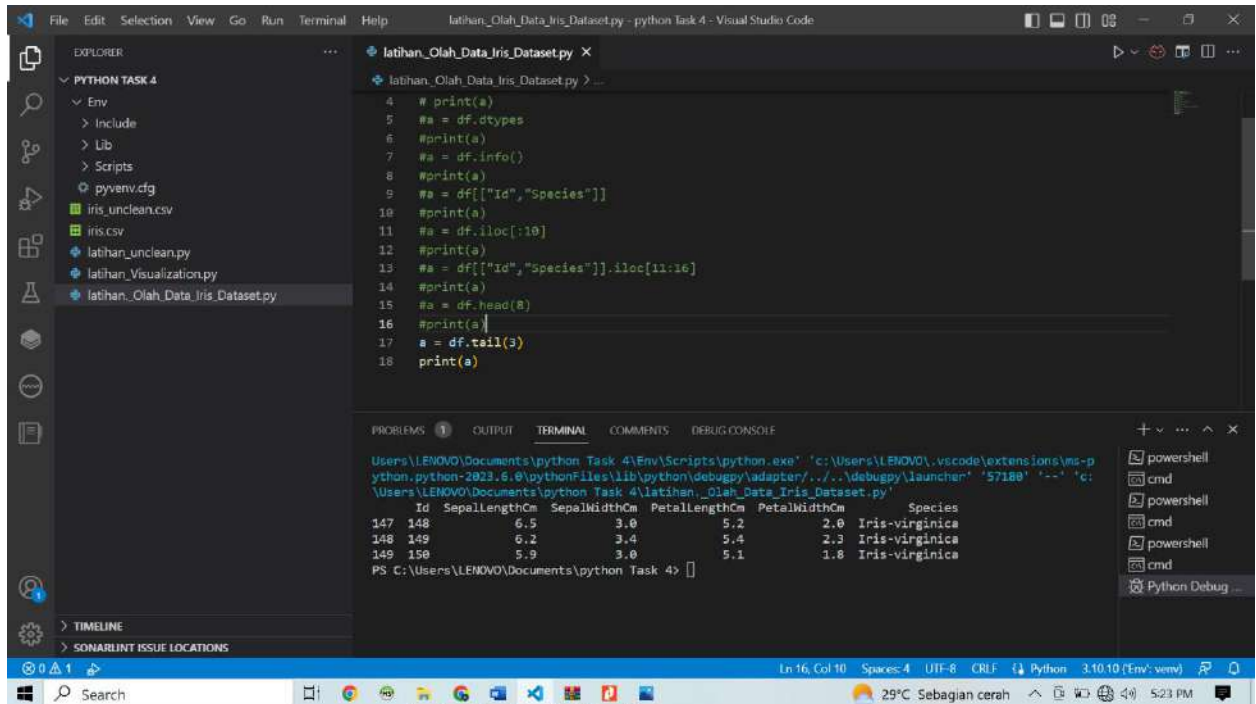
The screenshot shows a Visual Studio Code editor with a Python file named `latihan_Olah_Data_Iris_Dataset.py`. The script reads a CSV file, prints the first 8 rows, and then prints the first 8 rows, specifically showing the 'Id' and 'Species' columns. The terminal output shows the execution of the script, displaying the first 8 rows and then the selected rows 11 to 15.

```
1 df = pd.read_csv('iris.csv')
2 # print a = df.head()
3 # print(a)
4 # print(a)
5 #a = df.dtypes
6 #print(a)
7 #a = df.info()
8 #print(a)
9 #a = df[["Id", "Species"]]
10 #print(a)
11 #a = df.iloc[:10]
12 #print(a)
13 #a = df[["Id", "Species"]].iloc[11:16]
14 #print(a)
15 a = df.head(8)
16 print(a)
```

```
PS C:\Users\LENOVO\Documents\python Task 4> c:; cd 'c:\Users\LENOVO\Documents\python Task 4'; & 'c:\Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe' 'c:\Users\LENOVO\.vscode\extensions\ms-python.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57176' '--' 'c:\Users\LENOVO\Documents\python Task 4\latihan_Olah_Data_Iris_Dataset.py'

   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
PS C:\Users\LENOVO\Documents\python Task 4>
```

7. Tampilkan data 3 terakhir



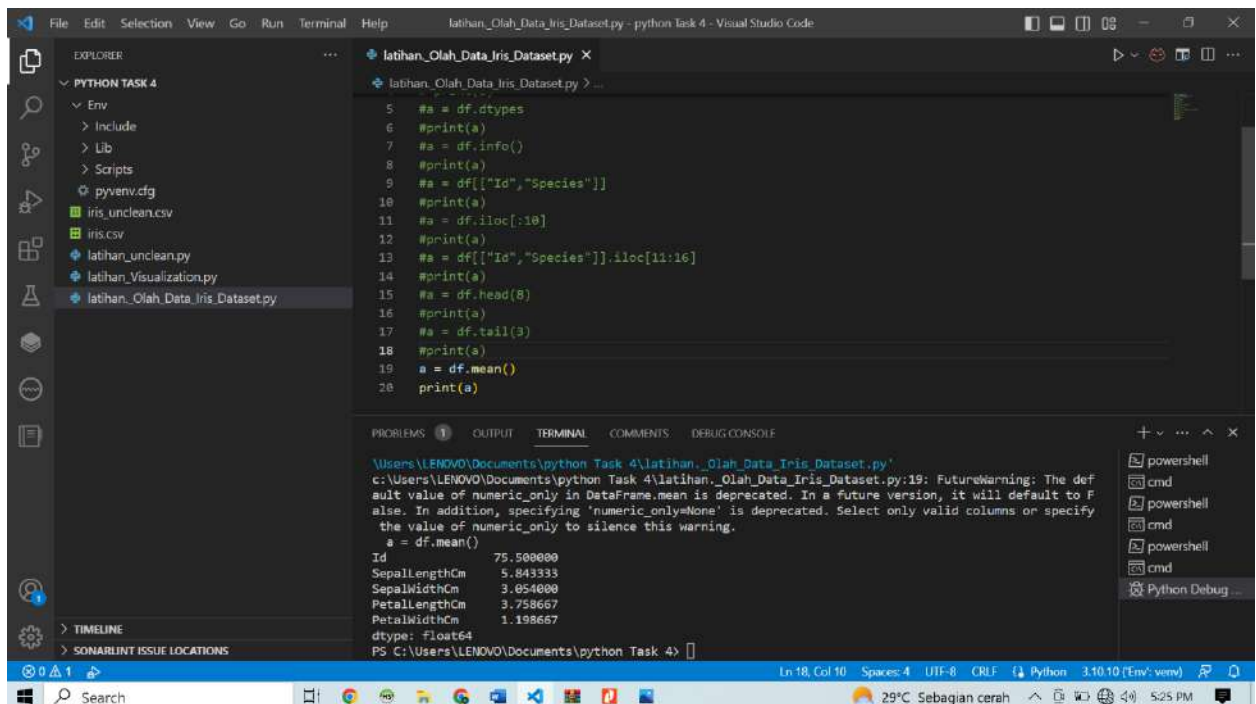
The screenshot shows a Visual Studio Code window with a Python script named `latihan_Olah_Data_Iris_Dataset.py` open. The script contains the following code:

```
4 # print(a)
5 #a = df.dtypes
6 #print(a)
7 #a = df.info()
8 #print(a)
9 #a = df[["Id", "Species"]]
10 #print(a)
11 #a = df.iloc[:10]
12 #print(a)
13 #a = df[["Id", "Species"]].iloc[11:16]
14 #print(a)
15 #a = df.head(8)
16 #print(a)
17 a = df.tail(3)
18 print(a)
```

The terminal output shows the result of the `df.tail(3)` operation, displaying the last three rows of the dataset:

	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

8. Hitung nilai mean dari dataset



The screenshot shows a Visual Studio Code window with a Python script named `latihan_Olah_Data_Iris_Dataset.py` open. The script contains the following code:

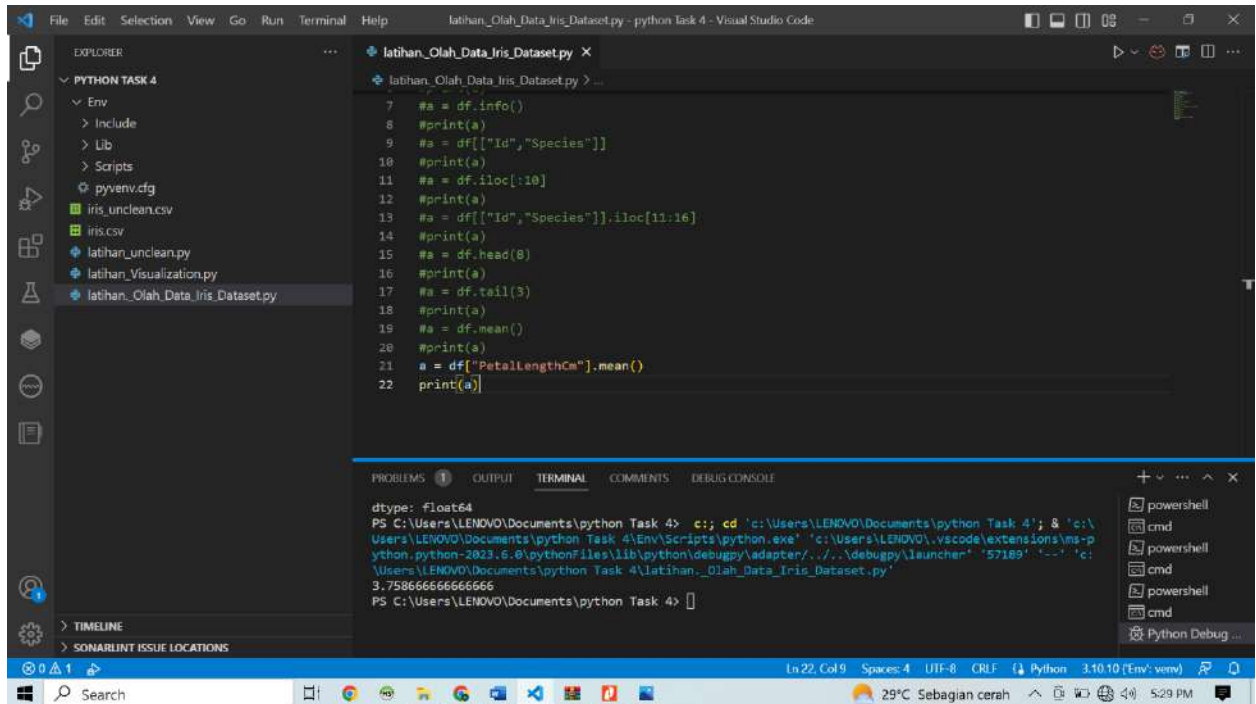
```
5 #a = df.dtypes
6 #print(a)
7 #a = df.info()
8 #print(a)
9 #a = df[["Id", "Species"]]
10 #print(a)
11 #a = df.iloc[:10]
12 #print(a)
13 #a = df[["Id", "Species"]].iloc[11:16]
14 #print(a)
15 #a = df.head(8)
16 #print(a)
17 #a = df.tail(3)
18 #print(a)
19 a = df.mean()
20 print(a)
```

The terminal output shows the result of the `df.mean()` operation, displaying the mean values for each column:

	Id	SepallengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
a	75.500000	5.843333	3.054000	3.758667	1.198667

The output also includes a warning message: "FutureWarning: The default value of numeric_only in DataFrame.mean is deprecated. In a future version, it will default to False. In addition, specifying 'numeric_only=None' is deprecated. Select only valid columns or specify the value of numeric_only to silence this warning."

9. Hitung nilai mean untuk kolom PetalLengthCm



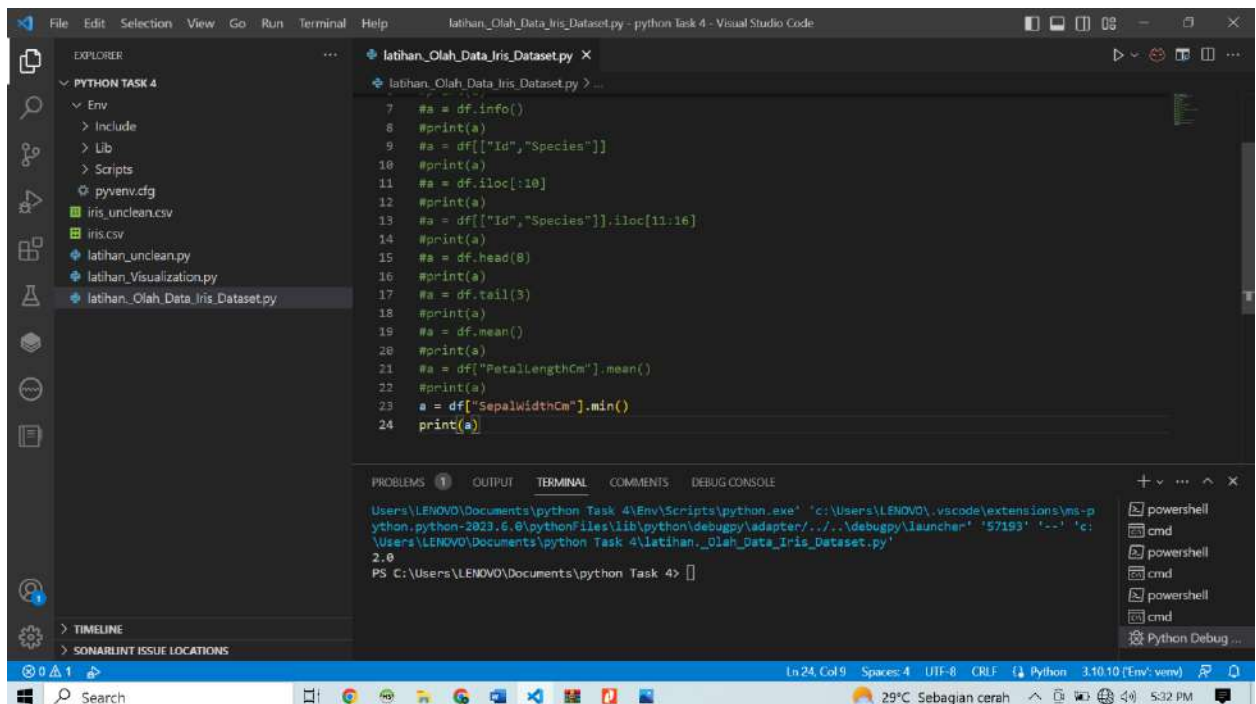
The screenshot shows the Visual Studio Code interface with a Python file named `latihan_Olah_Data_Iris_Dataset.py` open. The file contains the following code:

```
7 #a = df.info()
8 #print(a)
9 #a = df[["Id","Species"]]
10 #print(a)
11 #a = df.iloc[:10]
12 #print(a)
13 #a = df[["Id","Species"]].iloc[11:16]
14 #print(a)
15 #a = df.head(8)
16 #print(a)
17 #a = df.tail(3)
18 #print(a)
19 #a = df.mean()
20 #print(a)
21 a = df["PetalLengthCm"].mean()
22 print(a)
```

The terminal output shows the execution of the script, resulting in the following output:

```
dtype: float64
PS C:\Users\LENOVO\Documents\python Task 4> c:\; cd 'c:\Users\LENOVO\Documents\python Task 4'; & 'c:\Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe' 'c:\Users\LENOVO\.vscode\extensions\ms-pyhton.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57189' '-...' 'c:\Users\LENOVO\Documents\python Task 4\latihan_Olah_Data_Iris_Dataset.py'
3.7586666666666666
PS C:\Users\LENOVO\Documents\python Task 4> 
```

10. Cari nilai minimal untuk kolom SepalWidthCm



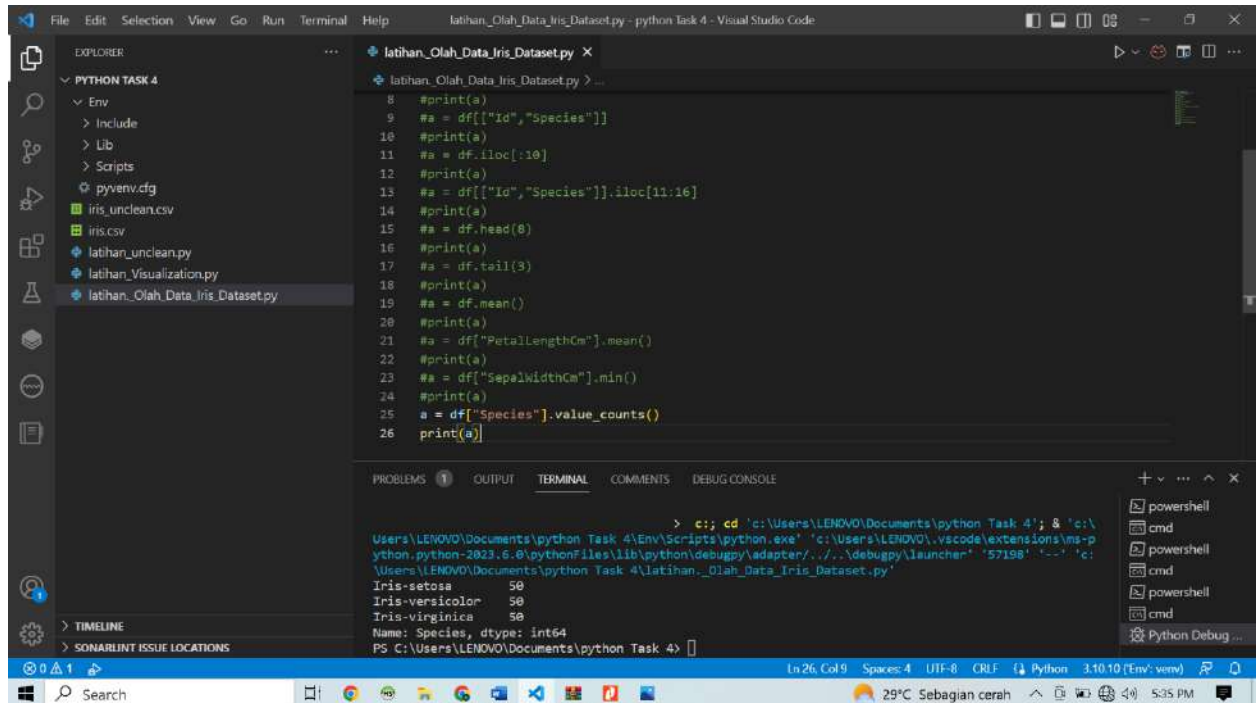
The screenshot shows the Visual Studio Code interface with a Python file named `latihan_Olah_Data_Iris_Dataset.py` open. The file contains the following code:

```
7 #a = df.info()
8 #print(a)
9 #a = df[["Id","Species"]]
10 #print(a)
11 #a = df.iloc[:10]
12 #print(a)
13 #a = df[["Id","Species"]].iloc[11:16]
14 #print(a)
15 #a = df.head(8)
16 #print(a)
17 #a = df.tail(3)
18 #print(a)
19 #a = df.mean()
20 #print(a)
21 #a = df["PetalLengthCm"].mean()
22 #print(a)
23 a = df["SepalWidthCm"].min()
24 print(a)
```

The terminal output shows the execution of the script, resulting in the following output:

```
Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe' 'c:\Users\LENOVO\.vscode\extensions\ms-pyhton.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57193' '-...' 'c:\Users\LENOVO\Documents\python Task 4\latihan_Olah_Data_Iris_Dataset.py'
2.0
PS C:\Users\LENOVO\Documents\python Task 4> 
```

11. Hitung frekuensi pada kolom Species dengan menggunakan metode value_counts()



```
File Edit Selection View Go Run Terminal Help latihan_Olah_Data_Iris_Dataset.py - python task 4 - Visual Studio Code

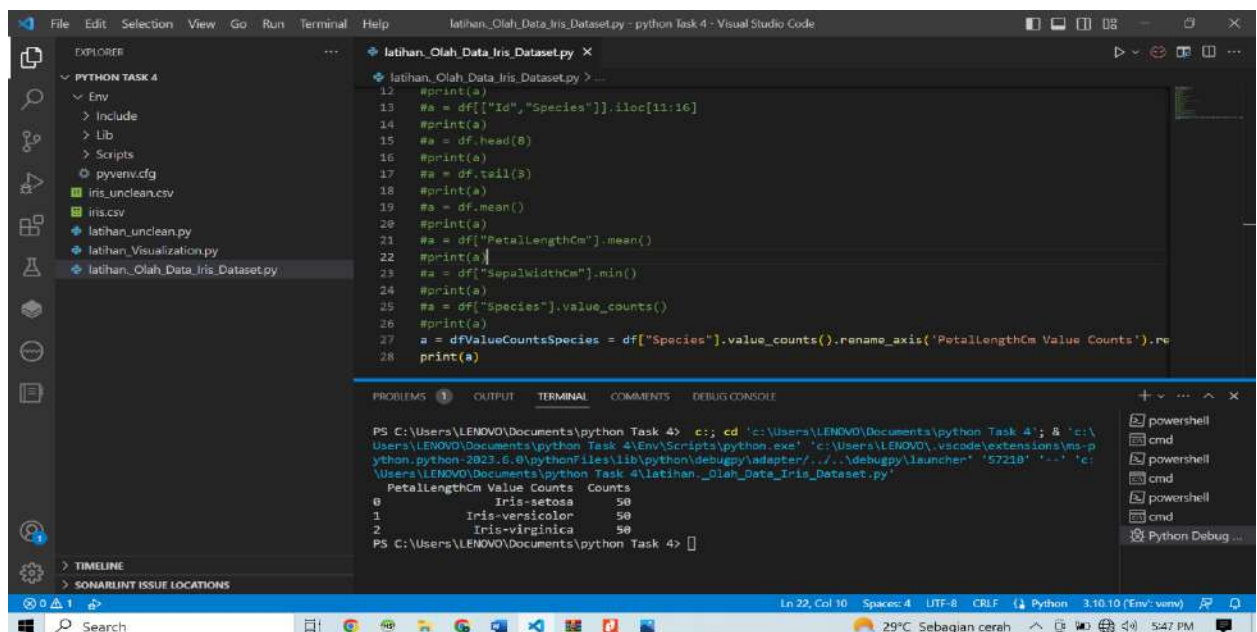
EXPLORER
PYTHON TASK 4
  Env
    Include
    Lib
    Scripts
  pyvenv.cfg
  iris_unclean.csv
  iris.csv
  latihan_unclean.py
  latihan_Visualization.py
  latihan_Olah_Data_Iris_Dataset.py

latihan_Olah_Data_Iris_Dataset.py
8 #print(a)
9 #a = df[["Id","Species"]]
10 #print(a)
11 #a = df.iloc[:10]
12 #print(a)
13 #a = df[["Id","Species"]].iloc[11:16]
14 #print(a)
15 #a = df.head(8)
16 #print(a)
17 #a = df.tail(3)
18 #print(a)
19 #a = df.mean()
20 #print(a)
21 #a = df[["PetalLengthCm"].mean()
22 #print(a)
23 #a = df[["SepalWidthCm"].min()
24 #print(a)
25 a = df["Species"].value_counts()
26 print(a)
```

```
PS C:\Users\LENOVO\Documents\python Task 4> c:\; cd 'c:\Users\LENOVO\Documents\python Task 4'; & 'c:\Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe' 'c:\Users\LENOVO\.vscode\extensions\ms-python.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57198' '-...' 'c:\Users\LENOVO\Documents\python Task 4\latihan_Olah_Data_Iris_Dataset.py'

Iris-setosa      50
Iris-versicolor  50
Iris-virginica   50
Name: Species, dtype: int64
PS C:\Users\LENOVO\Documents\python Task 4>
```

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



```
File Edit Selection View Go Run Terminal Help latihan_Olah_Data_Iris_Dataset.py - python task 4 - Visual Studio Code

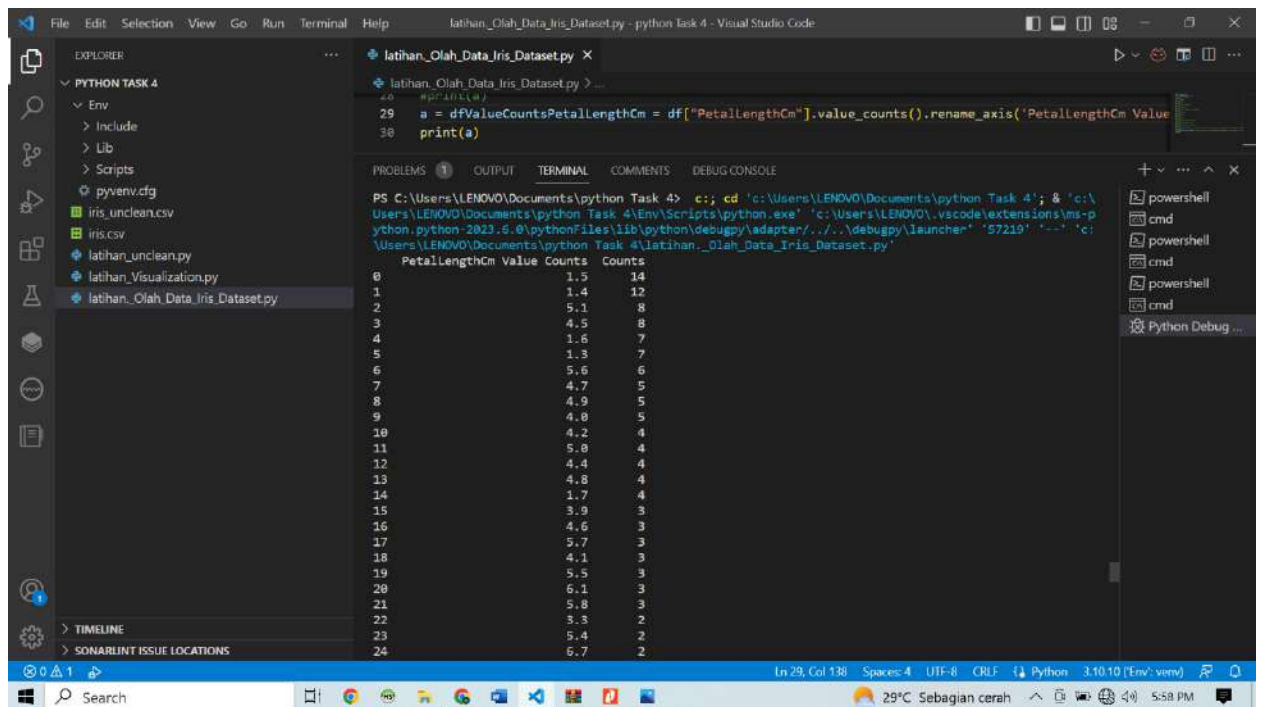
EXPLORER
PYTHON TASK 4
  Env
    Include
    Lib
    Scripts
  pyvenv.cfg
  iris_unclean.csv
  iris.csv
  latihan_unclean.py
  latihan_Visualization.py
  latihan_Olah_Data_Iris_Dataset.py

latihan_Olah_Data_Iris_Dataset.py
12 #print(a)
13 #a = df[["Id","Species"]].iloc[11:16]
14 #print(a)
15 #a = df.head(8)
16 #print(a)
17 #a = df.tail(3)
18 #print(a)
19 #a = df.mean()
20 #print(a)
21 #a = df[["PetalLengthCm"].mean()
22 #print(a)
23 #a = df[["SepalWidthCm"].min()
24 #print(a)
25 #a = df["Species"].value_counts()
26 #print(a)
27 a = dfValueCountsSpecies = df["Species"].value_counts().rename_axis('PetalLengthCm Value Counts').re
28 print(a)
```

```
PS C:\Users\LENOVO\Documents\python Task 4> c:\; cd 'c:\Users\LENOVO\Documents\python Task 4'; & 'c:\Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe' 'c:\Users\LENOVO\.vscode\extensions\ms-python.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57218' '-...' 'c:\Users\LENOVO\Documents\python Task 4\latihan_Olah_Data_Iris_Dataset.py'

PetalLengthCm Value Counts  Counts
0      Iris-setosa           50
1      Iris-versicolor       50
2      Iris-virginica        50
PS C:\Users\LENOVO\Documents\python Task 4>
```

13. Hitung frekuensi pada kolom PetalLengthCm dengan menggunakan value_counts () dalam bentuk dataframe



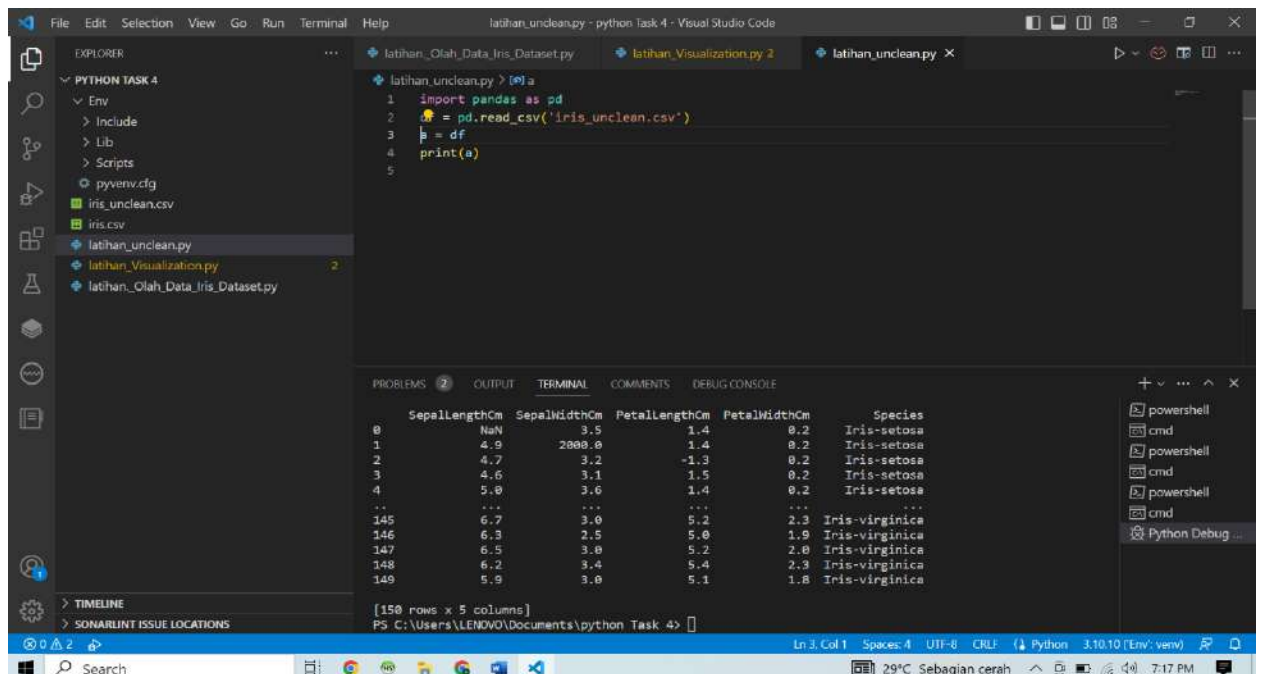
```
latihan_Olah_Data_Iris_Dataset.py
29 a = dfValueCountsPetalLengthCm = df["PetalLengthCm"].value_counts().rename_axis('PetalLengthCm Value
30 print(a)
```

```
PS C:\Users\LENOVO\Documents\python Task 4> c:\; cd 'c:\Users\LENOVO\Documents\python Task 4'; & 'c:\
Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe' 'c:\Users\LENOVO\.vscode\extensions\ms-p
ython.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57219' '--' 'c:\
Users\LENOVO\Documents\python Task 4\latihan_Olah_Data_Iris_Dataset.py'
```

PetalLengthCm Value Counts	Counts
1.5	14
1.4	12
5.1	8
4.5	8
1.6	7
1.3	7
5.6	6
4.7	5
4.9	5
4.0	5
4.2	4
5.0	4
4.4	4
4.8	4
1.7	4
3.9	3
4.6	3
5.7	3
4.1	3
5.5	3
6.1	3
5.8	3
3.3	2
5.4	2
6.7	2

❖ Olah Data Iris Dataset (Unclean)

1. Tampilkan dataset

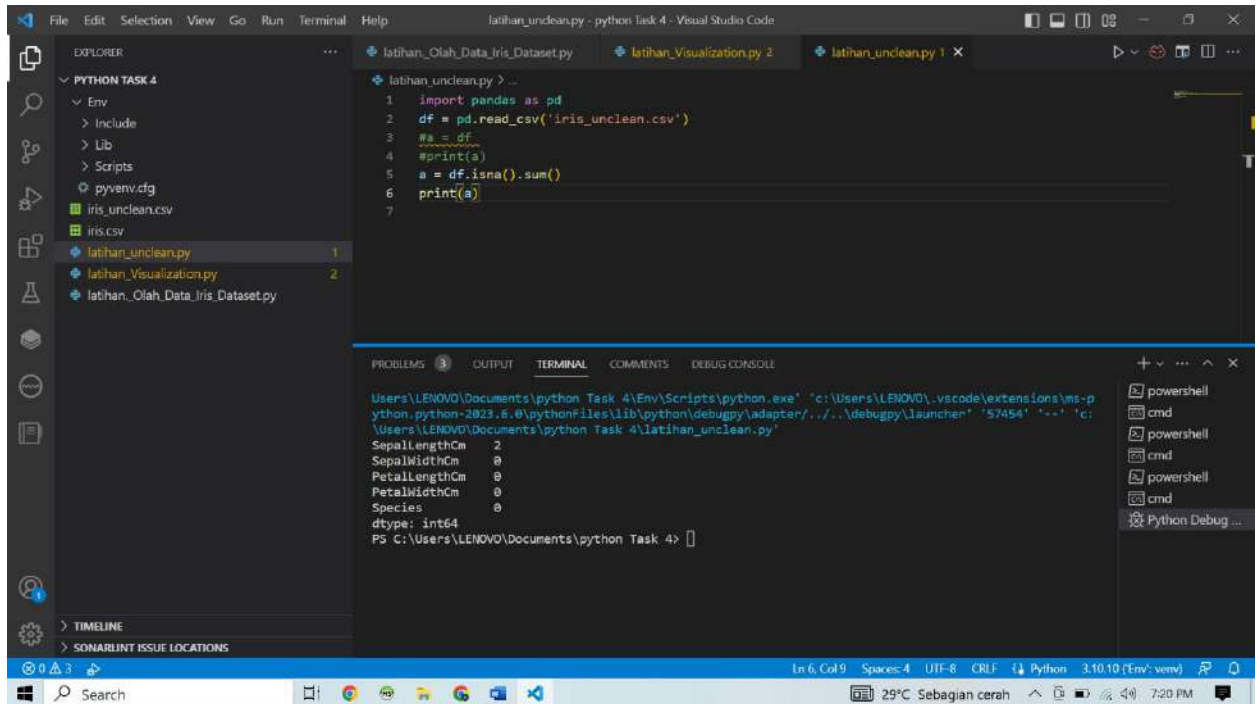


```
latihan_Unclean.py
1 import pandas as pd
2 df = pd.read_csv('iris_unclean.csv')
3 a = df
4 print(a)
```

```
PS C:\Users\LENOVO\Documents\python Task 4> c:\; cd 'c:\Users\LENOVO\Documents\python Task 4'; & 'c:\
Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe' 'c:\Users\LENOVO\.vscode\extensions\ms-p
ython.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57219' '--' 'c:\
Users\LENOVO\Documents\python Task 4\latihan_Unclean.py'
```

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	NaN	3.5	1.4	0.2	Iris-setosa
1	4.9	2000.0	1.4	0.2	Iris-setosa
2	4.7	3.2	-1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

2. Hitung jumlah nilai null pada dataset



The screenshot shows the Visual Studio Code interface with a Python file named `latihan_unclean.py` open. The file contains the following code:

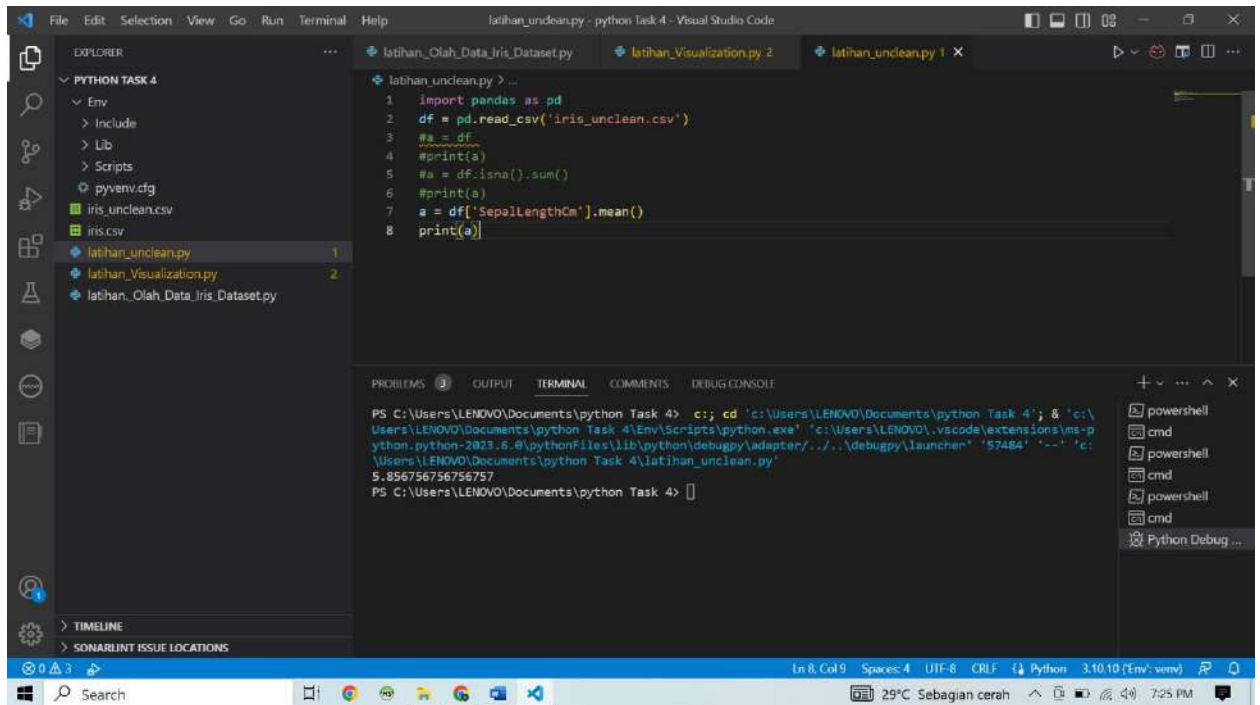
```
1 import pandas as pd
2 df = pd.read_csv('iris_unclean.csv')
3 #a = df
4 #print(a)
5 a = df.isna().sum()
6 print(a)
```

The Explorer panel on the left shows the project structure for 'PYTHON TASK 4', including files like `iris_unclean.csv`, `latihan_unclean.py`, `latihan_Visualization.py`, and `latihan_Olah_Data_Iris_Dataset.py`.

The Terminal panel at the bottom shows the output of the script:

```
Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe "c:\Users\LENOVO\.vscode\extensions\ms-p
python.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher" '57454' '--' 'c:
\Users\LENOVO\Documents\python Task 4\latihan_unclean.py'
SepallengthCm      2
SepallengthCm      0
SepallengthCm      0
SepallengthCm      0
SepallengthCm      0
Species            0
dtype: int64
PS C:\Users\LENOVO\Documents\python Task 4>
```

3. Cari nilai mean dari SepaLengthCm



The screenshot shows the Visual Studio Code interface with the same Python file `latihan_unclean.py` open. The code has been updated to calculate the mean of the `SepaLengthCm` column:

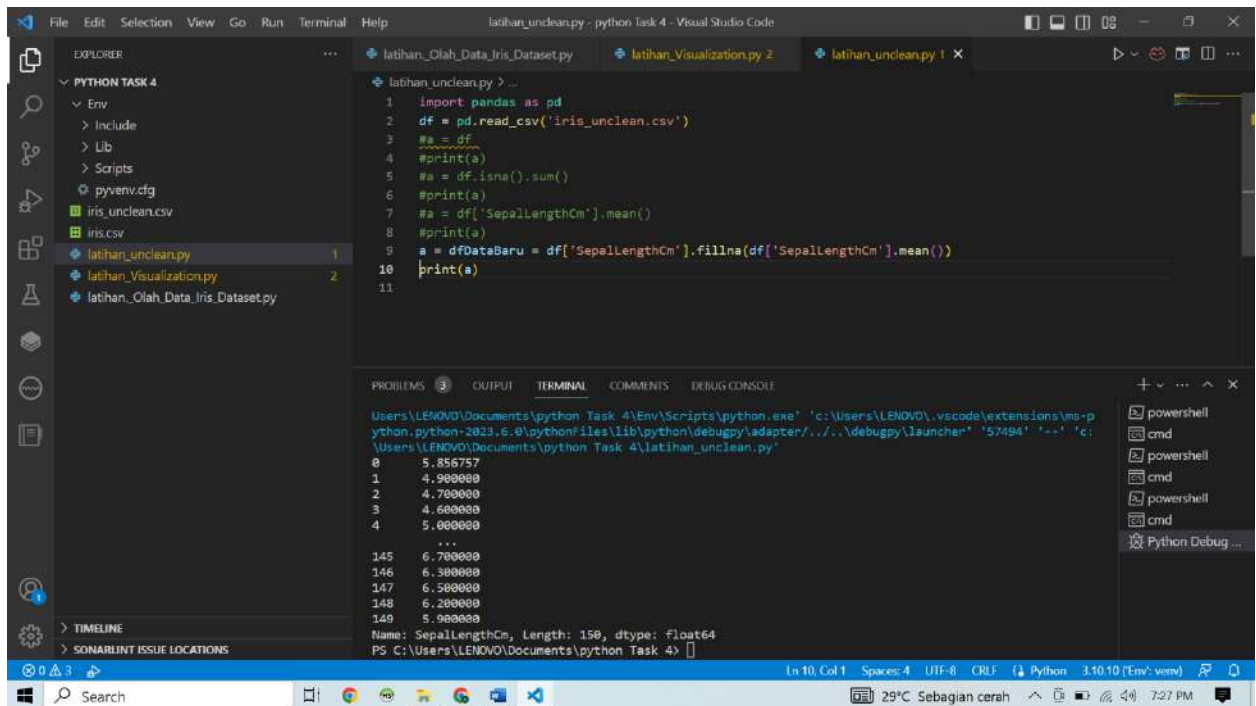
```
1 import pandas as pd
2 df = pd.read_csv('iris_unclean.csv')
3 #a = df
4 #print(a)
5 #a = df.isna().sum()
6 #print(a)
7 a = df['SepaLengthCm'].mean()
8 print(a)
```

The Explorer panel on the left remains the same, showing the project structure for 'PYTHON TASK 4'.

The Terminal panel at the bottom shows the output of the script:

```
PS C:\Users\LENOVO\Documents\python Task 4> c:\Users\LENOVO\Documents\python Task 4> & 'c:\
Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe' 'c:\Users\LENOVO\.vscode\extensions\ms-p
python.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57484' '--' 'c:
\Users\LENOVO\Documents\python Task 4\latihan_unclean.py'
5.856756756756757
PS C:\Users\LENOVO\Documents\python Task 4>
```

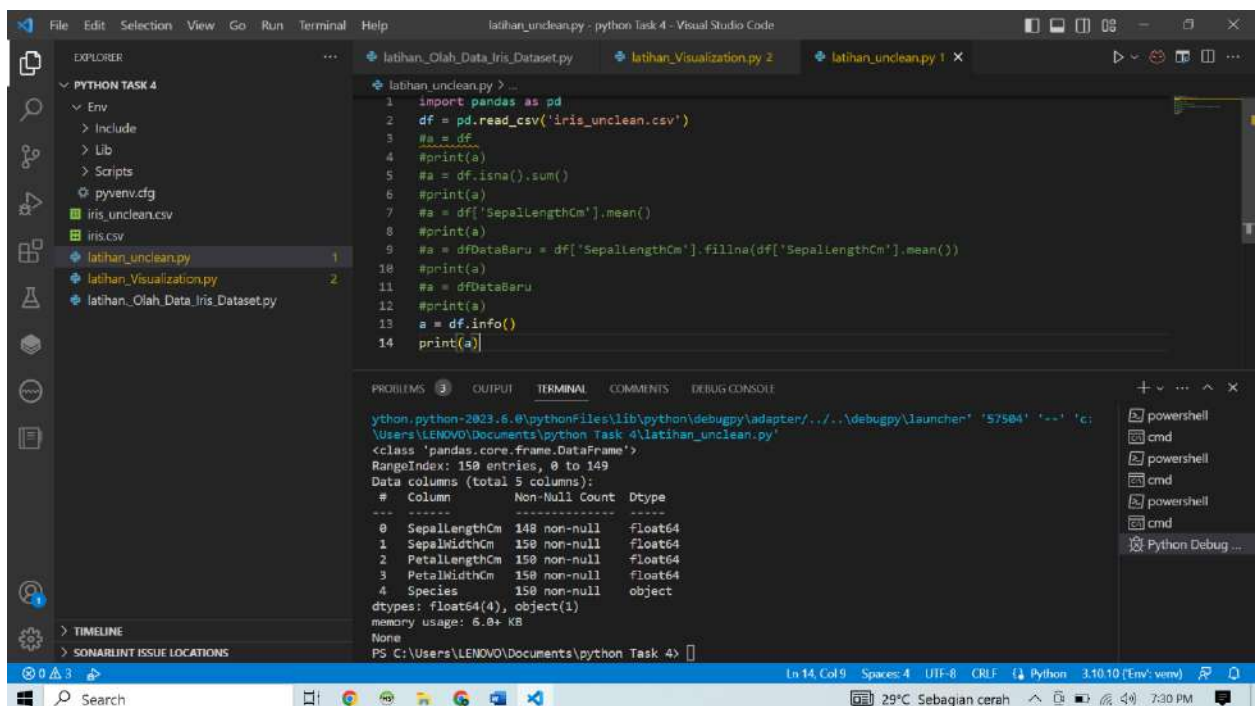

4. Mengganti missing value dengan mean(), kemudian masukkan ke variable



```
1 import pandas as pd
2 df = pd.read_csv('iris_unclean.csv')
3 #a = df
4 #print(a)
5 #a = df.isna().sum()
6 #print(a)
7 #a = df['SepallLengthCm'].mean()
8 #print(a)
9 a = dfDataBaru = df['SepallLengthCm'].fillna(df['SepallLengthCm'].mean())
10 print(a)
11
```

```
Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe 'c:\Users\LENOVO\.vscode\extensions\ms-p
ython.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57494' '--' 'c:
\Users\LENOVO\Documents\python Task 4\latihan_unclean.py'
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: SepallLengthCm, Length: 150, dtype: float64
PS C:\Users\LENOVO\Documents\python Task 4>
```

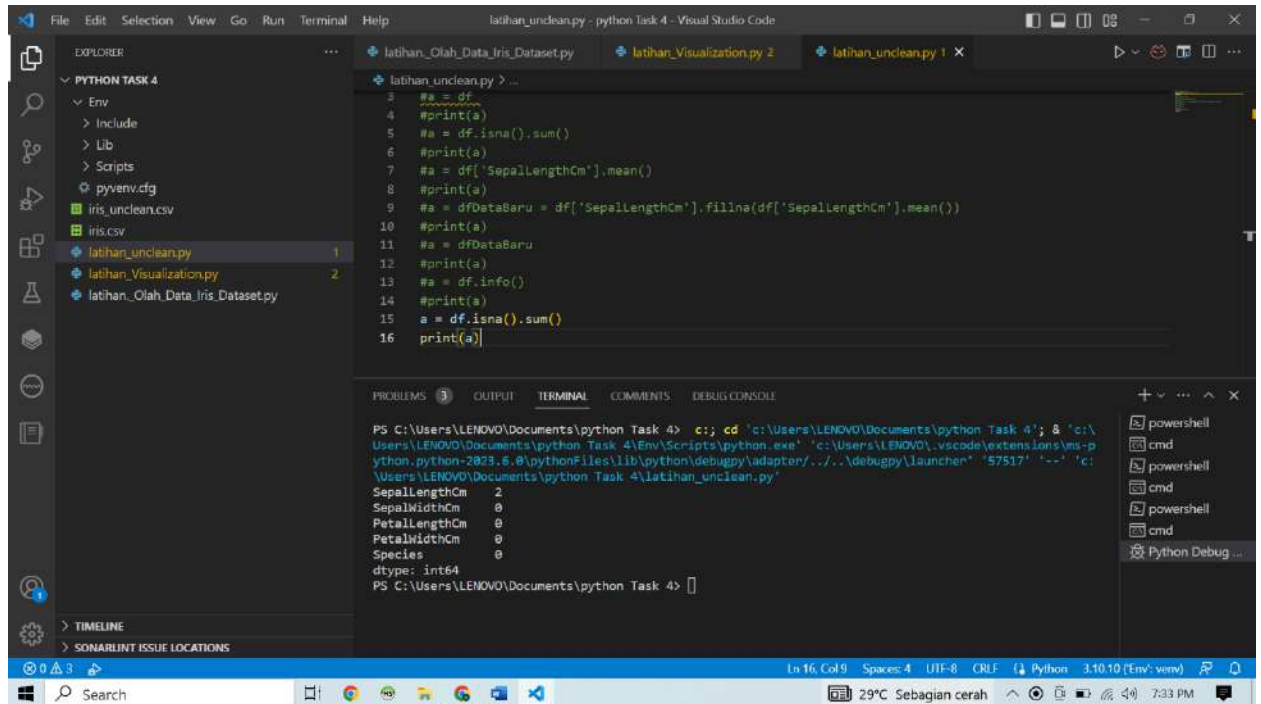
5. Cek jumlah baris dan kolom



```
1 import pandas as pd
2 df = pd.read_csv('iris_unclean.csv')
3 #a = df
4 #print(a)
5 #a = df.isna().sum()
6 #print(a)
7 #a = df['SepallLengthCm'].mean()
8 #print(a)
9 #a = dfDataBaru = df['SepallLengthCm'].fillna(df['SepallLengthCm'].mean())
10 #print(a)
11 #a = dfDataBaru
12 #print(a)
13 a = df.info()
14 print(a)
```

```
ython.python-2023.6.0\pythonFiles\lib\python\debugpy\adapter\..\..\debugpy\launcher' '57584' '--' 'c:
\Users\LENOVO\Documents\python Task 4\latihan_unclean.py'
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ---
0   SepallLengthCm  148 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: 6.8+ KB
None
PS C:\Users\LENOVO\Documents\python Task 4>
```

6. Hitung jumlah nilai null pada Dataframe baru



The screenshot shows the Visual Studio Code interface with a Python file named `latihan_unclean.py` open. The file contains a script to load the Iris dataset, check for null values, and fill them with the mean of the 'SepallengthCm' column. The terminal window at the bottom shows the command to run the script and the resulting output.

```
latihan_unclean.py > ...
3  #a = df
4  #print(a)
5  #a = df.isna().sum()
6  #print(a)
7  #a = df['SepallengthCm'].mean()
8  #print(a)
9  #a = dfDataBaru = df['SepallengthCm'].fillna(df['SepallengthCm'].mean())
10 #print(a)
11 #a = dfDataBaru
12 #print(a)
13 #a = df.info()
14 #print(a)
15 a = df.isna().sum()
16 print(a)
```

```
PS C:\Users\LENOVO\Documents\python Task 4> c:; cd 'c:\Users\LENOVO\Documents\python Task 4'; & 'c:\Users\LENOVO\Documents\python Task 4\Env\Scripts\python.exe' 'c:\Users\LENOVO\.vscode\extensions\ms-python.python-2023.6.0\pythonFiles\lib\python\debugpy\adaptor\...\debugpy\launcher' '57517' '--' 'c:\Users\LENOVO\Documents\python Task 4\latihan_unclean.py'
SepallengthCm    2
SepallwidthCm    0
PetallengthCm    0
PetallwidthCm    0
Species          0
dtype: int64
PS C:\Users\LENOVO\Documents\python Task 4>
```