

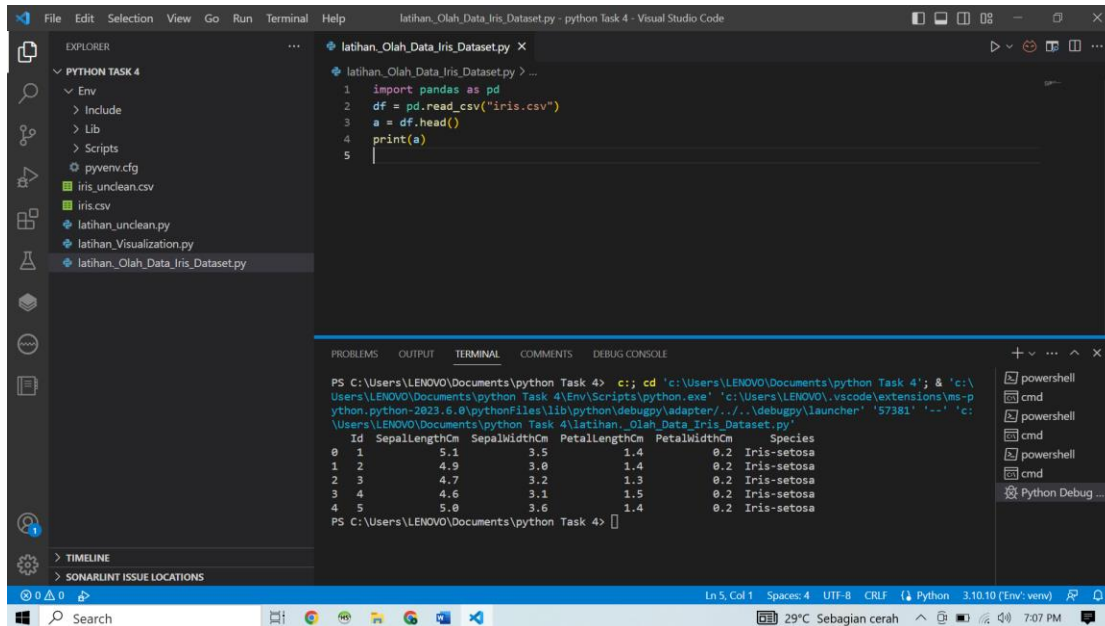
Nama : Hadijatol Kadri

Kelas : A

NIM : 20.01.013.006

A. Data set Iris.csv

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



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

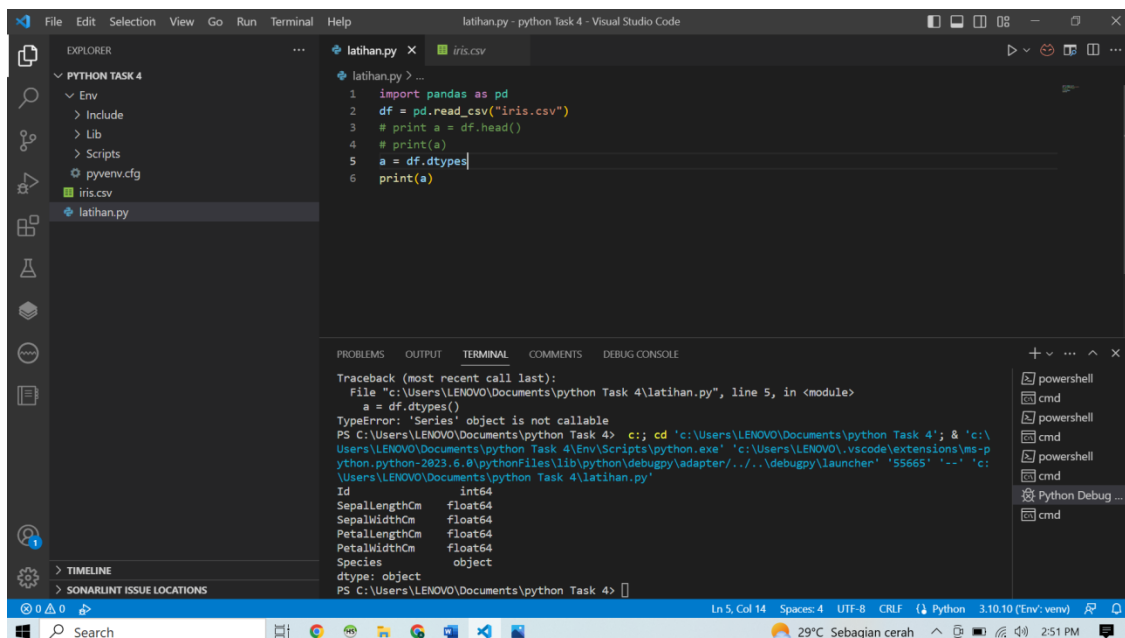
EXPLORER
PYTHON TASK 4
  Env
  Include
  Lib
  Scripts
  pyvenv.cfg
  iris_unclean.csv
  iris.csv
  latihan_unclean.py
  latihan_Visualization.py
  latihan_Olah_Data_Iris_Dataset.py

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

PROBLEMS OUTPUT TERMINAL COMMENTS DEBUG CONSOLE
PS C:\Users\LENOVO\Documents\python Task 4> 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' '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>

Ln 5, Col 1  Spaces: 4  UTF-8  CRLF  Python 3.10.10 (Env: venv)
```

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



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

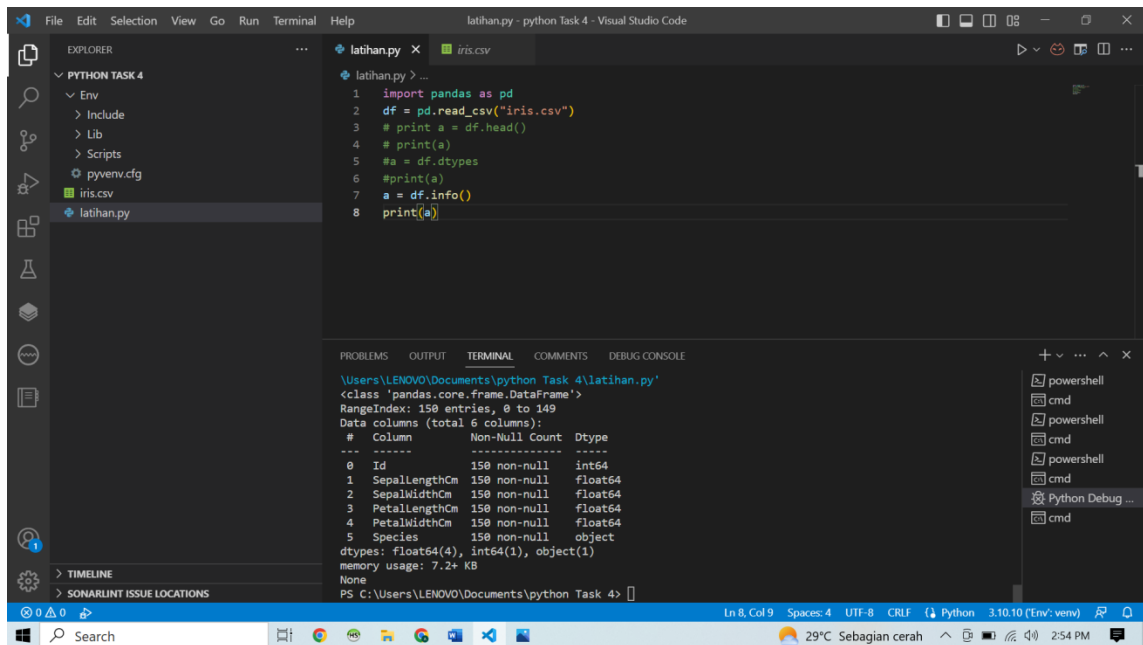
EXPLORER
PYTHON TASK 4
  Env
  Include
  Lib
  Scripts
  pyvenv.cfg
  iris.csv
  latihan.py

latihan.py
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)

PROBLEMS OUTPUT TERMINAL COMMENTS DEBUG CONSOLE
Traceback (most recent call last):
  File "c:\Users\LENOVO\Documents\python Task 4\latihan.py", line 5, in <module>
    a = df.dtypes()
TypeError: 'Series' object is not callable
PS C:\Users\LENOVO\Documents\python Task 4> 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' '55665' '--' 'c:\Users\LENOVO\Documents\python Task 4\latihan.py'
Id                int64
SepalLengthCm     float64
SepalWidthCm      float64
PetalLengthCm     float64
PetalWidthCm      float64
Species           object
dtype: object
PS C:\Users\LENOVO\Documents\python Task 4>

Ln 5, Col 14  Spaces: 4  UTF-8  CRLF  Python 3.10.10 (Env: venv)
```

3. Hitungkan ukuran (jumlah baris pada kolom) dari dataset



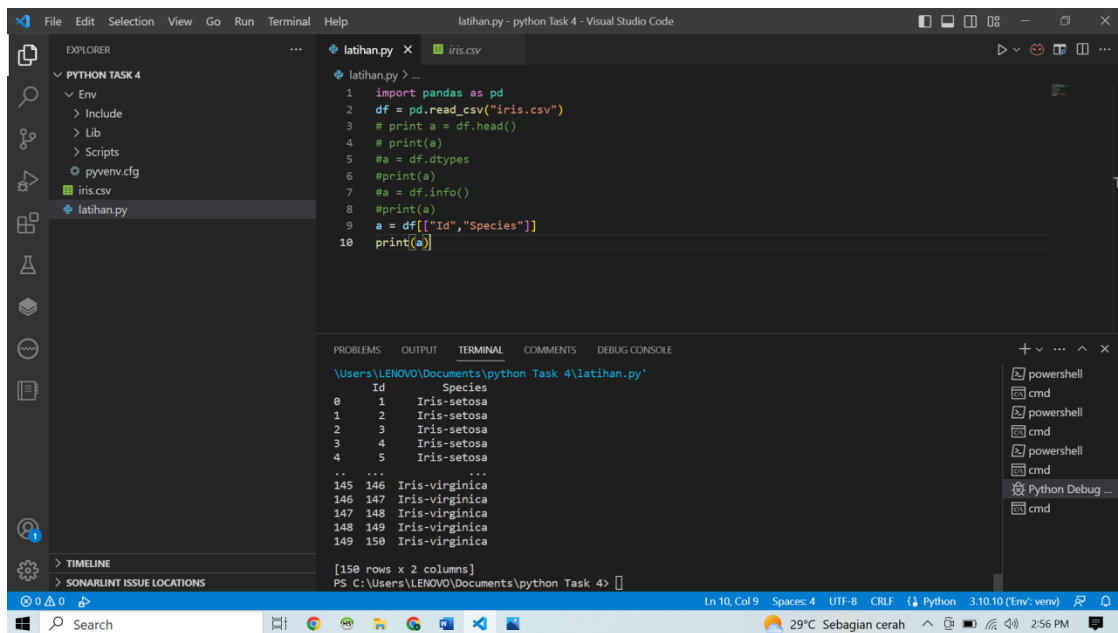
The screenshot shows a Visual Studio Code window with a Python file named `latihan.py` open. The code in the file is as follows:

```
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)
```

The terminal output shows the result of running the script:

```
\Users\LENOVO\Documents\python Task 4\latihan.py'
<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
None
PS C:\Users\LENOVO\Documents\python Task 4>
```

4. Tampilkan data untuk kolom “Id” dan kolom “species” dalam bentuk dataframe



The screenshot shows a Visual Studio Code window with a Python file named `latihan.py` open. The code in the file is as follows:

```
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)
```

The terminal output shows the result of running the script:

```
\Users\LENOVO\Documents\python Task 4\latihan.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>
```

5. Tampilkan data baris indeks ke-0 sampai dengan indeks ke Sembilan

The screenshot shows a Visual Studio Code window with a Python script named `latihan.py` open. The script reads an `iris.csv` file into a pandas DataFrame and prints the first 10 rows. The terminal output shows the first 10 rows of the dataset, including columns for `Id`, `SepalLengthCm`, `SepalWidthCm`, `PetalLengthCm`, `PetalWidthCm`, and `Species`.

```
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)
```

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

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

The screenshot shows the same Visual Studio Code window, but the script has been modified to print data from index 11 to 15, specifically the `Id` and `Species` columns. The terminal output shows the first 10 rows of the dataset, followed by an error message: `KeyError: ["Id", "Species"]`. The error message indicates that the columns `Id` and `Species` are not found in the DataFrame.

```
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)
13 a = df[["Id", "Species"]].iloc[11:16]
14 print(a)
```

self._raise_if_missing(keyarr, indexer, axis_name)
File "c:\Users\LENOVO\Documents\python Task 4\Env\lib\site-packages\pandas\core\indexes\base.py", line 6139, in _raise_if_missing
raise KeyError(f"({not_found}) not in index")
KeyError: ["Id", "Species"]

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

7. Tampilkan data 8 baris pertama

The screenshot shows a Visual Studio Code window with a Python file named `latihan.py` open. The script imports `pandas` as `pd`, reads the `iris.csv` file into a DataFrame `df`, and prints the first 8 rows. The terminal output shows the first 8 rows of the dataset.

```
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)
13 # a = df[["Id", "Species"]].iloc[11:16]
14 #print(a)
15 a = df.head(8)
16 print(a)
```

	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. Tampilkan data 3 baris pertaamaa

The screenshot shows a Visual Studio Code window with a Python file named `latihan.py` open. The script imports `pandas` as `pd`, reads the `iris.csv` file into a DataFrame `df`, and prints the last 3 rows. The terminal output shows the last 3 rows of the dataset.

```
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)
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)
```

	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

9. Hitung nilai mean dari dataset

```
1 # Import pandas as pd
2 import pandas as pd
3 # Load the iris.csv file into a DataFrame
4 df = pd.read_csv('iris.csv')
5 # Print the first 10 rows of the DataFrame
6 #print(df.head(10))
7 # Print the data types of the columns
8 #print(df.dtypes)
9 # Print the information of the DataFrame
10 #print(df.info())
11 # Select the 'petalLengthCm' column
12 #petalLengthCm = df['petalLengthCm']
13 # Print the first 10 rows of the selected column
14 #print(petalLengthCm.head(10))
15 # Print the data types of the selected column
16 #print(petalLengthCm.dtypes)
17 # Print the information of the selected column
18 #print(petalLengthCm.info())
19 # Calculate the mean of the 'petalLengthCm' column
20 a = petalLengthCm.mean()
21 print(a)
```

fyding 'numeric_only=None' is deprecated. Select only valid columns or specify the value of numeric_on
ly to silence this warning.
a = df.mean()
Id 75.500000
SepalLengthCm 5.843333
SepalWidthCm 3.854000
PetalLengthCm 3.758667
PetalWidthCm 1.198667
dtype: float64
PS C:\Users\LENOVO\Documents\python Task 4>

10. Hitung nilai mean untuk kolom petalLengthCm

```
1 # Import pandas as pd
2 import pandas as pd
3 # Load the iris.csv file into a DataFrame
4 df = pd.read_csv('iris.csv')
5 # Print the first 10 rows of the DataFrame
6 #print(df.head(10))
7 # Print the data types of the columns
8 #print(df.dtypes)
9 # Print the information of the DataFrame
10 #print(df.info())
11 # Select the 'petalLengthCm' column
12 #petalLengthCm = df['petalLengthCm']
13 # Print the first 10 rows of the selected column
14 #print(petalLengthCm.head(10))
15 # Print the data types of the selected column
16 #print(petalLengthCm.dtypes)
17 # Print the information of the selected column
18 #print(petalLengthCm.info())
19 # Calculate the mean of the 'petalLengthCm' column
20 a = petalLengthCm.mean()
21 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' '55854' '--' 'c:
Users\LENOVO\Documents\python Task 4\latihan.py'
3.7586666666666666
PS C:\Users\LENOVO\Documents\python Task 4>

11. Cari nilai minimal untuk kolom SepalwidthCm


```
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('Species Value Counts').reset_index()
28 print(a)
```

IndentationError: unexpected indent

```
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\launcher' '55936' '-...' 'c:\Users\LENOVO\Documents\python Task 4\latihan.py'
```

Species Value Counts	counts
0 Iris-setosa	50
1 Iris-versicolor	50
2 Iris-virginica	50

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

```
27 a = dfValueCountsSpecies = df["Species"].value_counts().rename_axis('Species Value Counts').reset_index()
28 #print(a)
29 a = dfValueCountsPetalLengthCm = df["PetalLengthCm"].value_counts().rename_axis('PetalLengthCm Value Counts').reset_index()
30 print(a)
```

PetalLengthCm Value Counts	counts
0 1.5	14
1 1.4	12
2 5.1	8
3 4.5	8
4 1.6	7
5 1.3	7
6 5.6	6
7 4.7	5
8 4.9	5
9 4.0	5
10 4.2	4
11 5.0	4
12 4.4	4
13 4.8	4
14 1.7	4
15 3.9	3
16 4.6	3
17 5.7	3
18 4.1	3
19 5.5	3
20 6.1	3
21 5.8	3
22 3.3	2
23 5.4	2
24 6.7	2
25 5.3	2
26 5.9	2