Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

Aim: Practical based on Pandas Data Structures

IDE:

What is Python Pandas?

Pandas is a powerful, open-source data analysis and manipulation package for Python. It provides data structures and functions needed to work on structured data seamlessly and efficiently.

What Is Pandas Used For?

Pandas is extensively used for:

- Data Cleaning: Handling missing values, duplications, and incorrect data formats.
- Data Manipulation: Filtering, transforming, and merging datasets.
- Data Analysis: Performing statistical analysis and aggregations.
- Data Visualization: Creating plots and charts to visualize data trends and patterns.
- Time Series Analysis: Handling and manipulating time series data.

Run the following command to install Pandas:

pip install pandas

import pandas as pd

print(pd.__version__)

Pandas Series

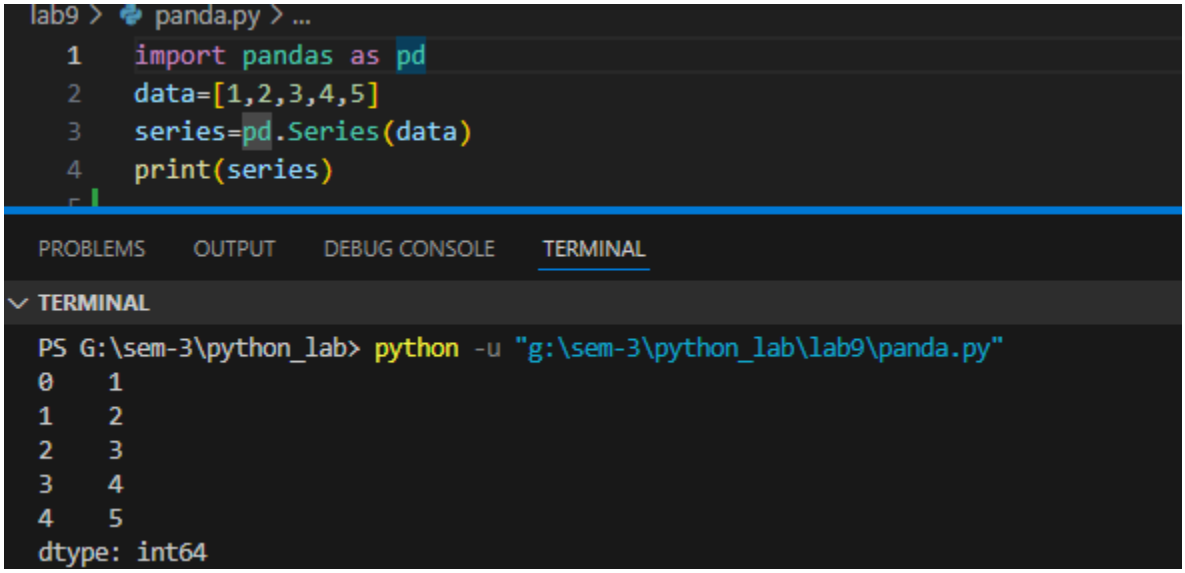
A Pandas Series is a one-dimensional labeled array capable of holding any data type. It is similar to a column in a spreadsheet or a SQL table.

Example:

```
import pandas as pd
# Creating a Series
data = [1, 2, 3, 4, 5]
series = pd.Series(data)
print(series)
```

Output:

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**



Basic Operations on Series

Perform various operations on Series, such as arithmetic operations, filtering, and statistical calculations.

Example:

```
# Arithmetic Operations
series2 = series + 10
print(series2)
# Filtering
filtered_series = series[series > 2]
print(filtered_series)
# Statistical Calculations
mean_value = series.mean()
print(mean_value)
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

```
lab9 > SeriesOper.py > ...
1   import pandas as pd
2   data=[1,2,3,4,5]
3   series=pd.Series(data)
4   series2=series+10
5   print(series2)
6
7   filtered=series[series>2]
8   print(filtered)
9
10  mean=series.mean()
11  print(mean)
12
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

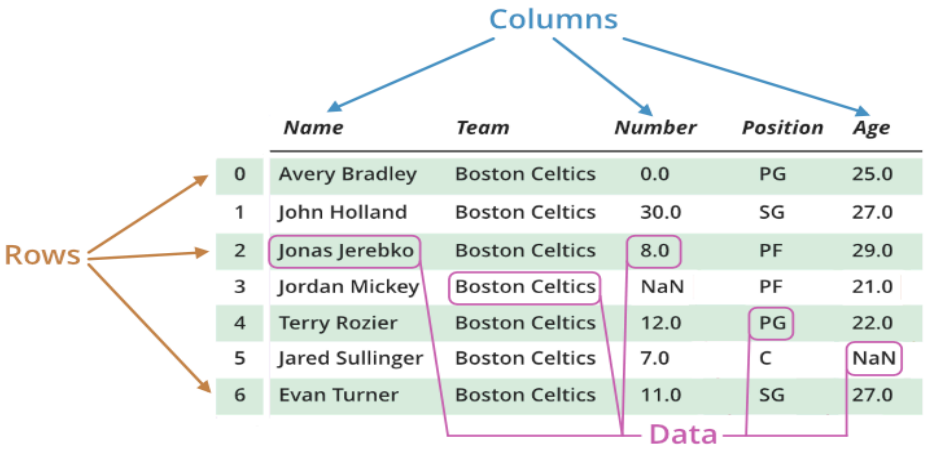
TERMINAL

```
PS G:\sem-3\python_lab> python -u "g:\sem-3\python_lab\lab9\SeriesOper.py"
0    11
1    12
2    13
3    14
4    15
dtype: int64
2    3
3    4
4    5
dtype: int64
3.0
```

Pandas Dataframe

Pandas DataFrame is two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns). A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns. Pandas DataFrame consists of three principal components, the data, rows, and columns.

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

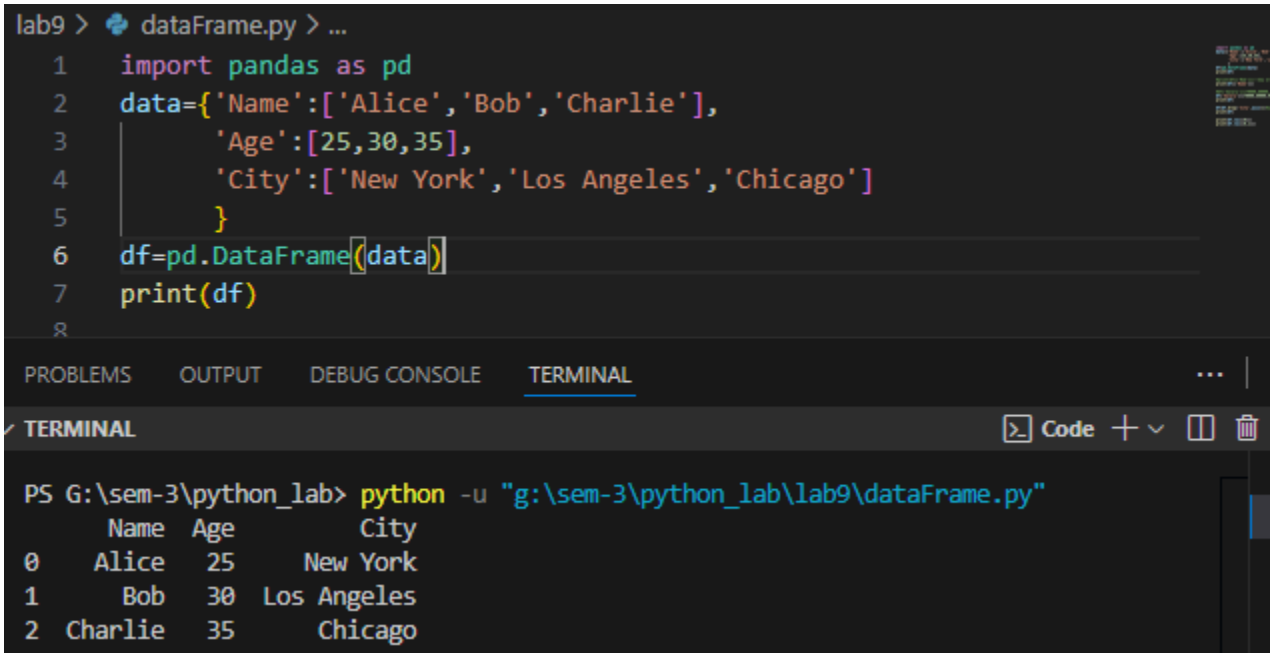


	Name	Team	Number	Position	Age
0	Avery Bradley	Boston Celtics	0.0	PG	25.0
1	John Holland	Boston Celtics	30.0	SG	27.0
2	Jonas Jerebko	Boston Celtics	8.0	PF	29.0
3	Jordan Mickey	Boston Celtics	NaN	PF	21.0
4	Terry Rozier	Boston Celtics	12.0	PG	22.0
5	Jared Sullinger	Boston Celtics	7.0	C	NaN
6	Evan Turner	Boston Celtics	11.0	SG	27.0

```
# Creating a DataFrame
data = {
   'Name': ['Alice', 'Bob', 'Charlie'],
   'Age': [25, 30, 35],
   'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
print(df)
```

Output

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**



Basic Operations on Dataframes

DataFrames support a wide range of operations for data manipulation and analysis.

# Accessing Columns (# select one column)
print(df[['Name']])
Output

Adding a New Column
df['Salary'] = [70000, 80000, 90000]
print(df)

# Dropping a Column
df = df.drop('City', axis=1)
print(df)
Output

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

```
lab9 > dataFrame.py > ...
  #print(df[['Nam']])--key error
  print(df[['Name']])

  #df['Salary']=[70000,80000,90000,90]--value error
  df['Salary']=[70000,80000,90000]
  print(df)

  df=df.drop('City',axis=1)#axis=0 = rows (default).axis=1 = columns.
  print(df)
```

```
      Name  Age         City  Salary
0    Alice   25     New York   70000
1      Bob   30  Los Angeles   80000
2  Charlie   35      Chicago   90000
      Name  Age  Salary
0    Alice   25   70000
1      Bob   30   80000
2  Charlie   35   90000
```

The DataFrame is like a table with rows and columns.
Pandas use the loc attribute to return one or more specified row(s)
# Return row 0:
print(df.loc[[0]])

#Return row 0 and 1:
#use a list of indexes:
print(df.loc[[0, 1]])
Output

```
  print(df.loc[[0]])
  print(df.loc[[0,1]])
```

```
    Name  Age  Salary
0  Alice   25   70000
    Name  Age  Salary
0  Alice   25   70000
1    Bob   30   80000
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

Named Indexes

With the index argument, you can name your own indexes.

Example:

Add a list of names to give each row a name:

```
import pandas as pd
data = {
  "calories": [420, 380, 390],
  "duration": [50, 40, 45]
}
df = pd.DataFrame(data, index = ["day1", "day2", "day3"])
print(df)
```

Output

Explanation of Key Pandas Functions

Reading and Writing Data:

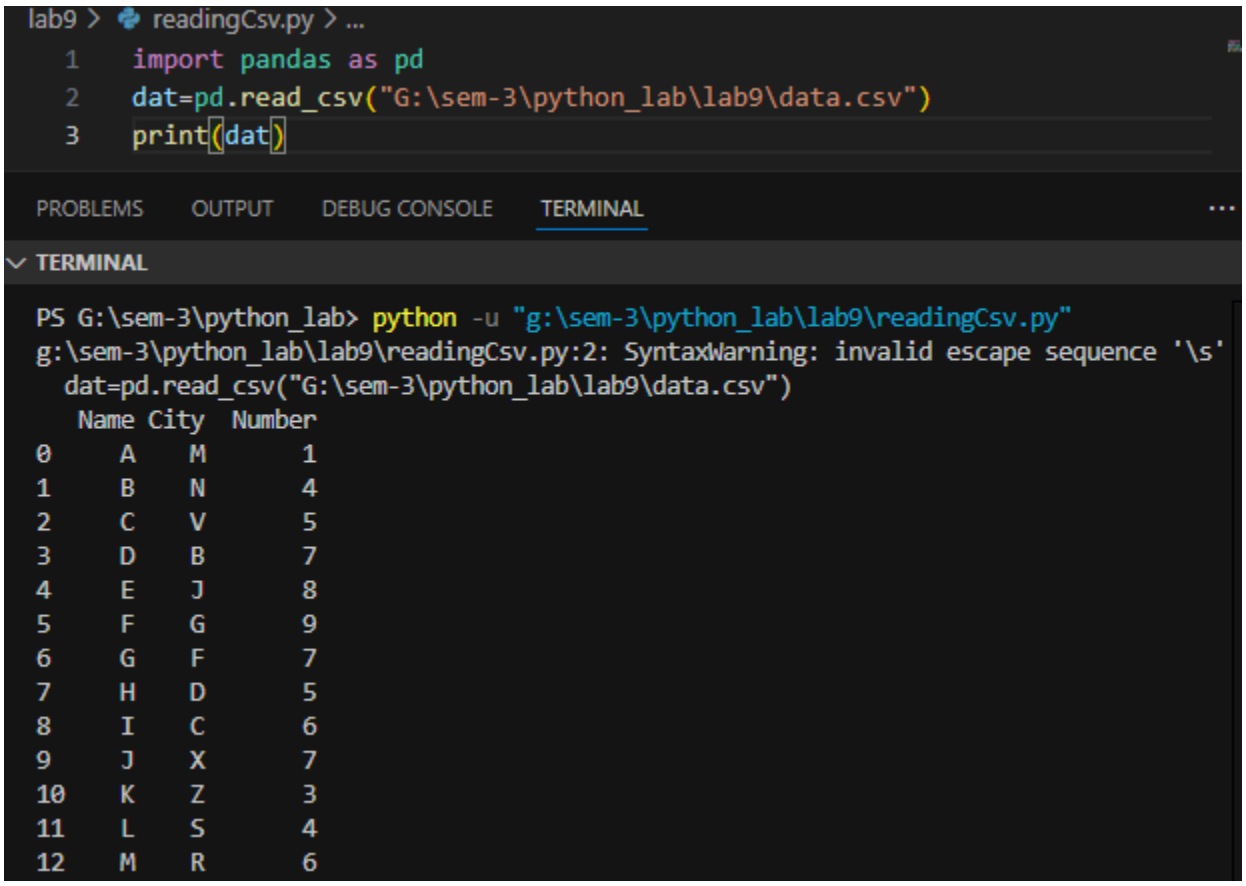
Reading Data: Read a CSV file into a DataFrame.

Example:

```
dat = pd.read_csv("data.csv")
print(dat)
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

Output



Writing Data: Write a DataFrame to a CSV file.
Note: Other Ways to Save Pandas DataFrames (to_excel(), to_json(), to_hdf(), to_sql(), to_pickle())

Example:

```
Biodata = {'Name': ['John', 'Emily', 'Mike', 'Lisa'],
        'Age': [28, 23, 35, 31],
        'Gender': ['M', 'F', 'M', 'F']
        }
df = pd.DataFrame(Biodata)
# Save the dataframe to a CSV file
df.to_csv('Biodata.csv', index=False)
```

Output

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

```
lab9 > writingCsv.py > ...
1  import pandas as pd
2  Biodata={'Name':['John','Emily','Mike','Lisa'],
3           'Age':[28,23,35,31],
4           'Gender':['M','F','M','F']
5           }
6  df=pd.DataFrame(Biodata)
7  print(df)
8  df.to_csv('Biodata.csv',index=False)
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL
TERMINAL                                  Code

PS G:\sem-3\python_lab> python -u "g:\sem-3\python_lab\lab9\writingCsv.py"
    Name  Age Gender
0   John   28      M
1  Emily   23      F
2   Mike   35      M
3   Lisa   31      F
```

Data Inspection:

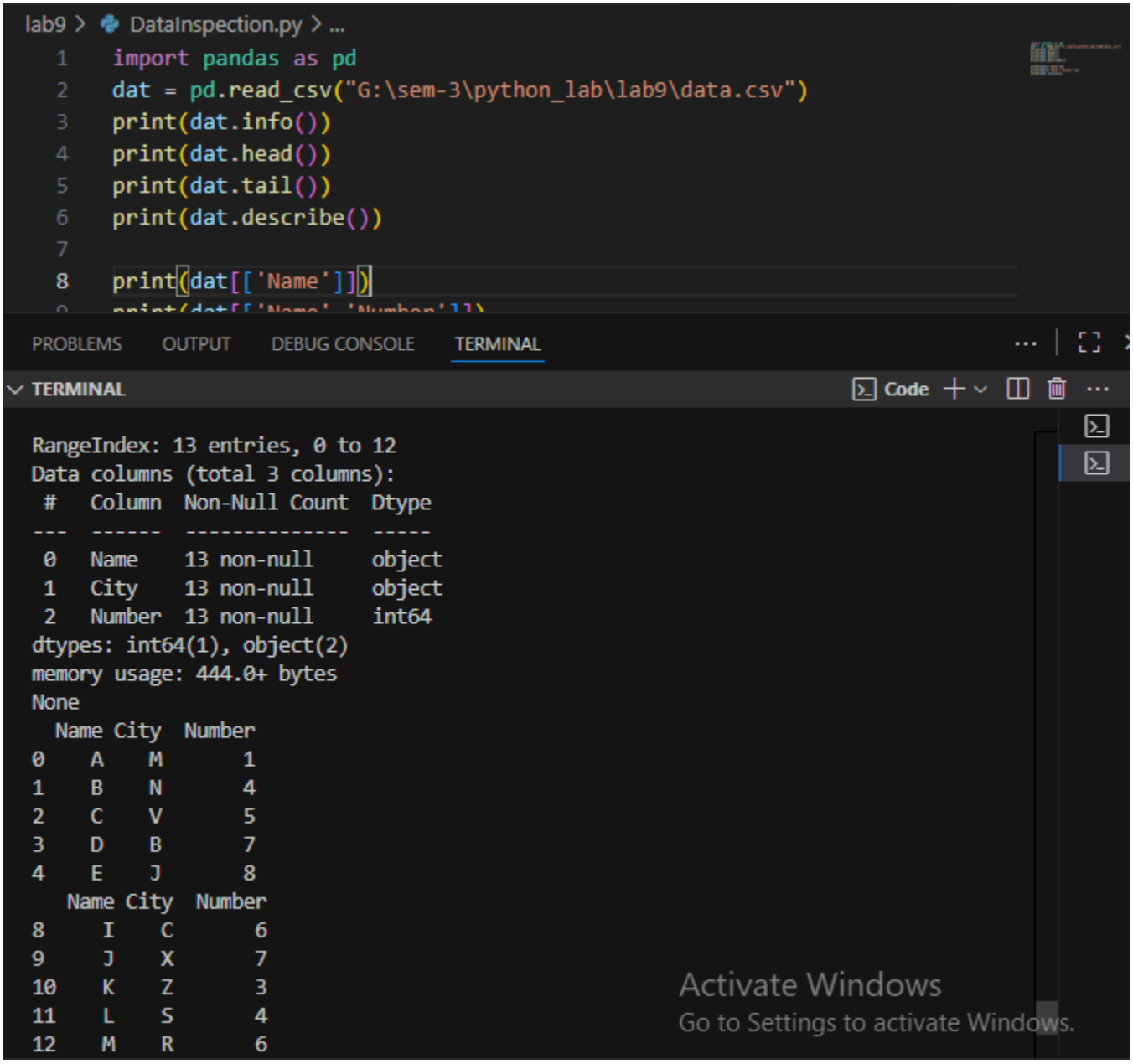
df.head(): Display the first few rows of the DataFrame.
df.tail(): Display the last few rows of the DataFrame.
df.info(): Display a summary of the DataFrame.
df.describe(): Provide descriptive statistics for numerical columns. (count: the number of non-null entries, mean: the mean value, std: the standard deviation, min: the minimum value, 25%, 50%, 75%: the lower, median, and upper quartiles, max: the maximum value)

Example:

```
dat = pd.read_csv("data.csv")
print(dat.info())
# shows first and last five rows
print(dat.head())
print(dat.tail())
print(dat.describe())
```

Output

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**



Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

```
         Number
count  13.000000
mean    5.538462
std     2.183857
min     1.000000
25%     4.000000
50%     6.000000
75%     7.000000
max     9.000000
```

Data Selection and Indexing:

dat[['A']]: Select a column.

dat[['A', 'B']]: Select multiple columns.

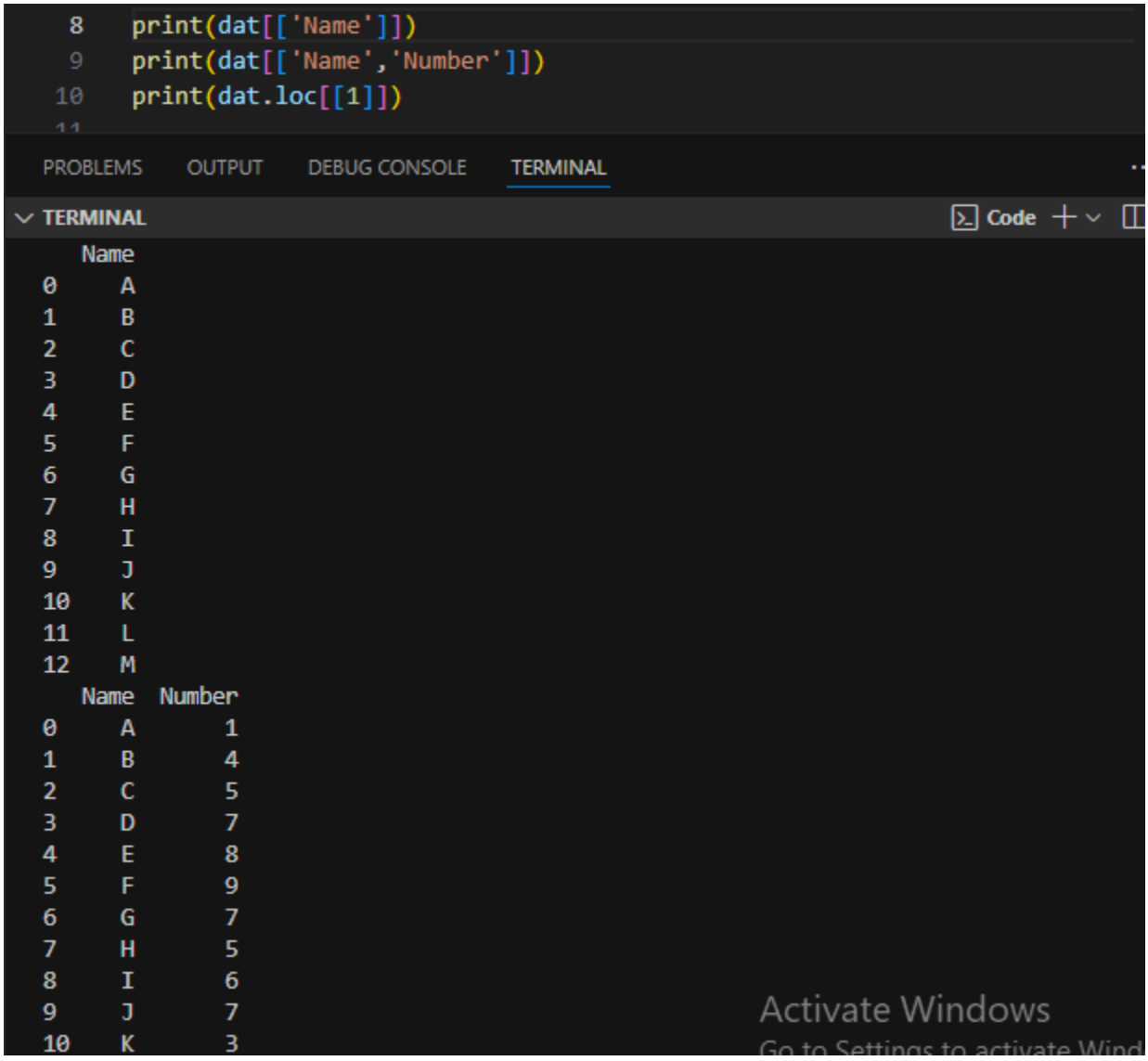
dat.loc[[0]]: Select a row by label.

Example:

```
print(dat[['Name']])
print(dat[['Name','Number']])
print(dat.loc[[1]])
```

Output

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**



Data Manipulation:

dat['A'] = dat['A'] * 2: Modify a column.

dat['F'] = dat['A'] + dat['B']: Create a new column based on existing columns.

dat.drop(columns=['A']): Drop a column.

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

dat.drop(index=[0]): Drop a row.

Task
Create a DataFrame with 5 numeric columns

```
data = {
   'A': [np.nan, 2, 3, 4, 5, 6, 7, 8, 9, 10],
   'B': np.random.normal(50, 15, 10),
   'C': np.random.rand(10) * 100,
   'D': np.linspace(1, 10, 10),
   'E': np.logspace(1, 2, 10)
}
df = pd.DataFrame(data)
```

Output

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

lab9 > Task.py > ...

```python
import numpy as np
import pandas as pd
#manipulation
data={
    'A':[np.nan,2,3,4,5,6,7,8,9,10],
    'B':np.random.normal(50,15,10),
    'C':np.random.rand(10)*100,
    'D':np.linspace(1,10,10),
    'E':np.logspace(1,2,10)
}
df=pd.DataFrame(data)
print(df)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

TERMINAL Code

```
PS G:\sem-3\python_lab> python -u "g:\sem-3\python_lab\lab9\Task.py"
      A          B          C     D           E
0   NaN  42.276080   4.740590   1.0   10.000000
1   2.0  60.166035  89.044420   2.0   12.915497
2   3.0  37.433499  79.576230   3.0   16.681005
3   4.0  70.481208  70.126616   4.0   21.544347
4   5.0  18.403087  92.011980   5.0   27.825594
5   6.0  54.247552  85.464963   6.0   35.938137
6   7.0  58.922126   9.708502   7.0   46.415888
7   8.0  45.758497  19.011180   8.0   59.948425
8   9.0  51.251011  63.013791   9.0   77.426368
9  10.0  68.682589  40.270201  10.0  100.000000
```

Post Lab Exercise:

a. Write a Pandas program to add, subtract, multiple and divide two Pandas Series.
b. Write a Pandas program to convert a dictionary to a Pandas series.
c. Write a Pandas program to create a series from a list, numpy array and dict
d. Write a Pandas program to stack two series vertically and horizontally

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

lab9 > PostLab.py > ...

```python
import pandas as pd
import numpy as np
#a. Write a Pandas program to add, subtract, multiple and divide two P
data1=[1,2,3,4,5]
data2=[1,2,3,4,5]
series1=pd.Series(data1)
series2=pd.Series(data2)
print(series1,"\n",series2)

seriesAdd=series1+series2
print("ADD")
print(seriesAdd)

seriesSub=series1-series2
print("Sub")
print(seriesSub)

seriesPro=series1*series2
print("Pro")
print(seriesPro)

seriesDiv=series1/series2
print("DIV")
print(seriesDiv)
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

```python

#b. Write a Pandas program to convert a dictionary to a Pandas series.
dict={'a': 100, 'b': 200, 'c': 300, 'd': 400}
series=pd.Series(dict)
print("Dictionary to Series:")
print(series)

#c. Write a Pandas program to create a series from a list, numpy array
list_data = [10, 20, 30, 40]
series_list = pd.Series(list_data)
print("Series from list:")
print(series_list)

arr_data = np.array([1, 2, 3, 4, 5])
series_array = pd.Series(arr_data)
print("\nSeries from NumPy array:")
print(series_array)

dict_data={'a': 10, 'b': 20, 'c': 30, 'd': 40}
series_dict=pd.Series(dict_data)
print("Dictionary to Series:")
print(series_dict)

#d. Write a Pandas program to stack two series vertically and horizont
s1 = pd.Series([1, 2, 3, 4])
s2 = pd.Series([5, 6, 7, 8])

```

```python
vertical_stack = pd.concat([s1, s2])
print("Vertical Stack:")
print(vertical_stack)

horizontal_stack = pd.concat([s1, s2], axis=1)
print("\nHorizontal Stack:")
print(horizontal_stack)
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

```
PS G:\sem-3\python_lab> python -u "g:\sem-3\python_lab\lab9\PostLab.py"
0    1
1    2
2    3
3    4
4    5
dtype: int64
 0    1
1    2
2    3
3    4
4    5
dtype: int64
ADD
0     2
1     4
2     6
3     8
4    10
```



```
dtype: int64
Sub
0    0
1    0
2    0
3    0
4    0
dtype: int64
Pro
0     1
1     4
2     9
3    16
4    25
dtype: int64
DIV
0    1.0
1    1.0
2    1.0
3    1.0
4    1.0
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

```
dtype: float64
Dictionary to Series:
a    100
b    200
c    300
d    400
dtype: int64
Series from list:
0    10
1    20
2    30
3    40
dtype: int64

Series from NumPy array:
0    1
1    2
2    3
3    4
4    5
dtype: int64
```



```
Dictionary to Series:
a    10
b    20
c    30
d    40
dtype: int64
Vertical Stack:
0    1
1    2
2    3
3    4
0    5
1    6
2    7
3    8
dtype: int64
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Engineering & Technology** **Department of Information and Communication Technology**	
Subject: Programming With Python (01CT1309)	**Aim:** Practical based on Pandas Data Structures	
Experiment No: 09	**Date:01/09/25**	**Enrollment No:92400133037**

GITHUB LINK

https://github.com/Heer972005/Python_Lab