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:**	**Enrollment No: 92400133055**

GITHUB

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]
```

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:**	**Enrollment No: 92400133055**

series = pd.Series(data)
print(series)
Output:

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

```
PROBLEMS 1  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\lab_9.py"
0    1
1    2
2    3
3    4
4    5
dtype: int64
PS E:\SEM 3\PWP>
```

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:**	**Enrollment No: 92400133055**

Output

```python
import pandas as pd
data = [1, 2, 3, 4, 5]
series = pd.Series(data)
series2 = series + 10
print(series2)
# Filtering
filtered_series = series[series > 2]
print(filtered_series)
# Statistical Calculations
mean_value = series.mean()
print(mean_value)
```

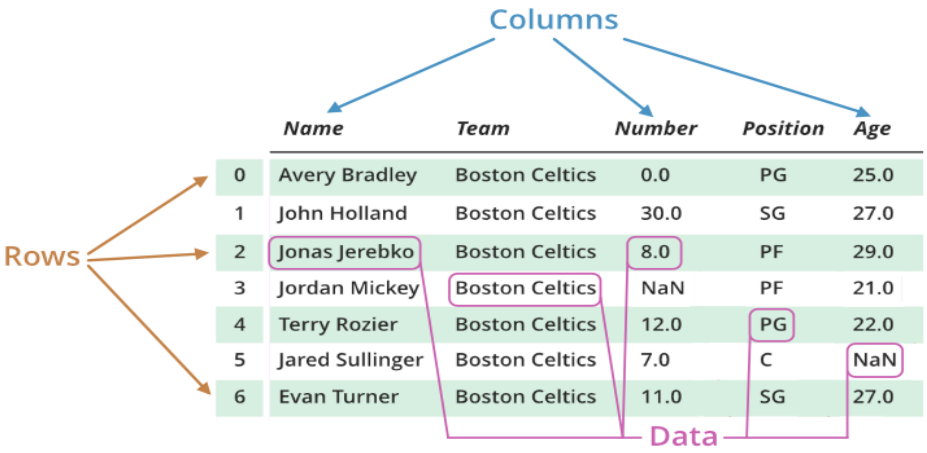
PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
2    3
3    4
4    5
dtype: int64
3.0
PS E:\SEM 3\PWP>
```

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:**	**Enrollment No: 92400133055**

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.




```
# 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:**	**Enrollment No: 92400133055**

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

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\lab_9.py"
      Name  Age         City
0    Alice   25     New York
1      Bob   30  Los Angeles
2  Charlie   35      Chicago
PS E:\SEM 3\PWP>
```

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

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:**	**Enrollment No: 92400133055**

```python
import pandas as pd
data = {
    'Name': ['Alice', 'Bob', 'Charlie'],
    'Age': [25, 30, 35],
    'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
# Accessing Columns (# select one column)
print(df[['Name']])
```

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\lab_9.py"
      Name
0    Alice
1      Bob
2  Charlie
PS E:\SEM 3\PWP>
```

Adding a New Column
df['Salary'] = [70000, 80000, 90000]
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:**	**Enrollment No: 92400133055**

```python
import pandas as pd
data = {
    'Name': ['Alice', 'Bob', 'Charlie'],
    'Age': [25, 30, 35],
    'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
df['Salary'] = [70000, 80000, 90000]
print(df)
```

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\lab_9.py"
      Name  Age         City  Salary
0    Alice   25     New York   70000
1      Bob   30  Los Angeles   80000
2  Charlie   35      Chicago   90000
PS E:\SEM 3\PWP>
```

# 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:**	**Enrollment No: 92400133055**

```python
import pandas as pd
data = {
    'Name': ['Alice', 'Bob', 'Charlie'],
    'Age': [25, 30, 35],
    'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
df['Salary'] = [70000, 80000, 90000]
print(df)
df = df.drop('City', axis=1)
print(df)
```

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\lab_9.py"
      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
PS E:\SEM 3\PWP>
```

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]])
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:**	**Enrollment No: 92400133055**

```python
import pandas as pd
data = {
    'Name': ['Alice', 'Bob', 'Charlie'],
    'Age': [25, 30, 35],
    'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
df['Salary'] = [70000, 80000, 90000]
print(df.loc[[0]])
```

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\lab_9.py"
    Name  Age      City  Salary
0  Alice   25  New York   70000
PS E:\SEM 3\PWP>
```

#Return row 0 and 1:
#use a list of indexes:
print(df.loc[[0, 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:**	**Enrollment No: 92400133055**

```
18  import pandas as pd
19  data = {
20      'Name': ['Alice', 'Bob', 'Charlie'],
21      'Age': [25, 30, 35],
22      'City': ['New York', 'Los Angeles', 'Chicago']
23  }
24  df = pd.DataFrame(data)
25  df['Salary'] = [70000, 80000, 90000]
26  print(df.loc[[0, 1]])
27
```

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\lab_9.py"
    Name  Age         City  Salary
0  Alice   25     New York   70000
1    Bob   30  Los Angeles   80000
PS E:\SEM 3\PWP>
```

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

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:**	**Enrollment No: 92400133055**

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

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\lab_9.py"
      calories  duration
day1       420        50
day2       380        40
day3       390        45
PS E:\SEM 3\PWP>
```

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)
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:**	**Enrollment No: 92400133055**

```python
import pandas as pd
data = {
  "calories": [420, 380, 390],
  "duration": [50, 40, 45]
}
df = pd.DataFrame(data, index = ["day1", "day2", "day3"])
dat = pd.read_csv(r"C:\Users\diyak\Downloads\data (1).csv")
print(dat)
```

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\lab_9.py"
   Name City  Number
0     A    M       1
1     B    N       4
2     C    V       5
3     D    B       7
4     E    J       8
5     F    G       9
6     G    F       7
7     H    D       5
8     I    C       6
9     J    X       7
10    K    Z       3
11    L    S       4
12    M    R       6
```

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())

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:**	**Enrollment No: 92400133055**

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

```
import pandas as pd
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(r'C:\Users\diyak\Downloads\Biodata.csv', index=False)
```

Saved:

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)

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:**	**Enrollment No: 92400133055**

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:**	**Enrollment No: 92400133055**

```
46   import pandas as pd
47   dat = pd.read_csv(r"C:\Users\diyak\Downloads\data (1).csv")
48   print(dat.info())
49   # shows first and last five rows
50   print(dat.head())
51   print(dat.tail())
52   print(dat.describe())
```

PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
RangeIndex: 13 entries, 0 to 12
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Name    13 non-null     object
 1   City    13 non-null     object
 2   Number  13 non-null     int64
dtypes: int64(1), object(2)
memory usage: 272.0+ bytes
None
  Name City  Number
0    A    M       1
1    B    N       4
2    C    V       5
3    D    B       7
4    E    J       8
```

Data Selection and Indexing:

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

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

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

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:**	**Enrollment No: 92400133055**

Example:

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

```
54  import pandas as pd
55  dat = pd.read_csv(r"C:\Users\diyak\Downloads\data (1).csv")
56  print(dat[['Name']])
57  print(dat[['Name','Number']])
58  print(dat.loc[[1]])
```

PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
0     A       1
1     B       4
2     C       5
3     D       7
4     E       8
5     F       9
6     G       7
7     H       5
8     I       6
9     J       7
10    K       3
11    L       4
12    M       6
  Name City  Number
1    B    N       4
```

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:**	**Enrollment No: 92400133055**

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.

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:**	**Enrollment No: 92400133055**

```
60  import pandas as pd
61  import numpy as np
62  data = {
63      'A': [np.nan, 2, 3, 4, 5, 6, 7, 8, 9, 10],
64      'B': np.random.normal(50, 15, 10),
65      'C': np.random.rand(10) * 100,
66      'D': np.linspace(1, 10, 10),
67      'E': np.logspace(1, 2, 10)
68  }
69  df = pd.DataFrame(data)
70  print(df)
```

PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
     A          B          C     D           E
0  NaN  60.626491   2.086771   1.0   10.000000
1  2.0  11.455842  34.111673   2.0   12.915497
2  3.0  65.292875  55.751457   3.0   16.681005
3  4.0  51.173129  67.205908   4.0   21.544347
4  5.0  61.416516  55.333786   5.0   27.825594
5  6.0  66.443029  66.436440   6.0   35.938137
6  7.0  19.057385  39.424948   7.0   46.415888
7  8.0  47.146305  30.311032   8.0   59.948425
8  9.0  51.911314   6.327910   9.0   77.426368
9 10.0  42.983490  49.914342  10.0  100.000000
```

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:**	**Enrollment No: 92400133055**

Post Lab Exercise:

a. Write a Pandas program to add, subtract, multiple and divide two Pandas Series.

```python
#Question 1
import pandas as pd
series1 = pd.Series([12,13,14,15,16])
series2 = pd.Series([1,2,3,4,5])
print("Addition: ",series1+series2)
print("Subtraction: ",series1-series2)
print("Multiplication: ",series1*series2)
print("Division: ",series1/series2)
```

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\post_lab_9.py"
Addition:  0    13
1    15
2    17
3    19
4    21
dtype: int64
Subtraction:  0    11
1    11
2    11
3    11
4    11
dtype: int64
Multiplication:  0    12
1    26
2    42
3    60
```

b. Write a Pandas program to convert a dictionary to a Pandas series.

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:**	**Enrollment No: 92400133055**

```python
#Question 2
import pandas as pd
data = {'a': 12,'b': 13,'c': 14,'d': 15}
print(data)
series = pd.Series(data)
print(series)
```

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\post_lab_9.py"
{'a': 12, 'b': 13, 'c': 14, 'd': 15}
a    12
b    13
c    14
d    15
dtype: int64
```

c. Write a Pandas program to create a series from a list, numpy array and dict

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:**	**Enrollment No: 92400133055**

```python
#Question 3
import numpy as np
import pandas as pd

data1 = [10, 20, 30, 40]
series1 = pd.Series(data1)
print("Series from list: ",series1)

data2 = np.array([1, 2, 3, 4, 5])
series2 = pd.Series(data2)
print("Series from numpy array: ",series2)

data3 = {'p': 10, 'q': 20, 'r': 30}
series3 = pd.Series(data3)
print("Series from dictionary: ",series3)
```

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\post_lab_9.py"
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
```

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:**	**Enrollment No: 92400133055**

```python
#Question 4
import pandas as pd
series1 = pd.Series([11,12,13,14])
series2 = pd.Series([15,16,17,18])

vertical_stack = pd.concat([series1, series2])
print("Vertical Stack: ",vertical_stack)
horizontal_stack = pd.concat([series1, series2], axis=1)
print("\nHorizontal Stack: ",horizontal_stack)
```

PROBLEMS 2 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS E:\SEM 3\PWP> python -u "e:\SEM 3\PWP\Class Tutorials\post_lab_9.py"
Vertical Stack:  0    11
1    12
2    13
3    14
0    15
1    16
2    17
3    18
dtype: int64

Horizontal Stack:      0   1
```