Lab – 25

Student Name: Navneet P

Student ID: AF0411619

Topic-Working with Pandas Dataframes

Functions used in the assignment:

1. **Pandas.DataFrame** – A 2-dimensional, size-mutable, and labeled data structure in Pandas, similar to an Excel spreadsheet or SQL table.

2. **to_string()** – A method to convert a DataFrame into a string format for easy display or output, with options for formatting.

3. **read_csv()** – A Pandas function to read a CSV file into a DataFrame, allowing for efficient data manipulation.

4. **head()** – A method that returns the first 5 rows of a DataFrame (or a specified number of rows).

5. **tail()** – A method that returns the last 5 rows of a DataFrame (or a specified number of rows).

6. **loc[]** – A method to access rows and columns in a DataFrame by labels (index) or a boolean array.

7. **to_csv()** – A method to export a DataFrame to a CSV file, allowing data to be saved in a readable format.

Q1. Create sample DataFrame using pandas with example.

Solution:

```python
import pandas as pd  # Import the pandas library for data manipulation

# Creating the first DataFrame with subject names as index and student names as columns
print("Dataframe 1\n--------------------")

subjects = ["Physics", "Chemistry", "Maths"]  # List of subjects as index
marks = [[89, 90, 91, 99], [79, 96, 95, 99], [77, 80, 91, 99]]  # List of marks for 3 subjects (each row for a subject)
students = ["Adarsh", "John", "Samuel", "Cheng-Lu"]  # List of students as column names

# Creating a DataFrame from the marks, with subjects as row index and students as column headers
df = pd.DataFrame(marks, index=subjects, columns=students)
print(df)  # Display the first DataFrame

print("\nDataframe 2\n--------------------")

# Creating the second DataFrame manually with player names and their runs
df2 = pd.DataFrame({
    "Most Runs(Indian)": ["S Tendulkar", "V Kohli", "R Dravid", "R Sharma"],  # Names of cricket players
    "Runs": [34357, 27041, 24064, 19276]  # Corresponding number of runs scored by each player
})

# Printing the second DataFrame without displaying the index column
print(df2.to_string(index=False))  # 'index=False' ensures that the index column is not displayed

```

Output:

```
Dataframe 1
--------------------
           Adarsh  John  Samuel  Cheng-Lu
Physics        89    90      91        99
Chemistry      79    96      95        99
Maths          77    80      91        99

Dataframe 2
--------------------
Most Runs(Indian)  Runs
      S Tendulkar 34357
          V Kohli 27041
         R Dravid 24064
         R Sharma 19276
```

Q2. Read 8362_data.csv print complete data and display columns[student name,attendance,mcq,lab]
Solution:

```python
import pandas as pd  # Import the pandas library for data manipulation

# Read the CSV file "8362_data.csv" into a DataFrame 'data'
# This file is expected to contain columns like "Student Name", "Attendance %", "FD MCQ %", and "TD Lab %"
data = pd.read_csv("8362_data.csv")

# Access specific columns from the 'data' DataFrame and store them in variables
names = data["Student Name"]  # Extract the "Student Name" column
attd = data["Attendance %"]   # Extract the "Attendance %" column
mcq = data["FD MCQ %"]        # Extract the "FD MCQ %" column (representing percentage of MCQs completed)
lab = data["TD Lab %"]        # Extract the "TD Lab %" column (representing percentage of Labs completed)

# Create a new DataFrame 'new_data' with more meaningful column names
new_data = pd.DataFrame({
    "Student Name": names,        # Assign the "Student Name" column
    "Attendance": attd,           # Rename "Attendance %" to "Attendance"
    "MCQs completed": mcq,        # Rename "FD MCQ %" to "MCQs completed"
    "Labs completed": lab         # Rename "TD Lab %" to "Labs completed"
})

# Convert the new DataFrame 'new_data' to a string representation for display (without the index column)
display_data = new_data.to_string(index=False)

# Print the formatted data
print(display_data)

```

Output:

Student Name	Attendance	MCQs completed	Labs completed
Dashmeet Singh	95	87	88
Mr Ketan	66	17	36
Harsh panchal	40	17	48
Rahil Ahmad khan	79	46	61
vishal singh	13	0	4
Kishan Mishra	31	18	18
Shubham Kumar	60	72	83
Sonal Garg	66	46	65
RATAN SRIVASTAV	71	83	82
Rohit verma	89	85	85
Yashika Gupta	80	86	84
Dishant Kumar Moga	75	83	85
Abhishek .	73	65	66
Aryan Verma	88	90	88
Ankush .	94	79	87
Riya Singla	78	76	80
Rajneesh Singh	51	19	7
Khushi Chaudhary	13	0	11
Chauhan Vandana Ramdayal	87	88	85
Bittu Samui	43	36	14
Vikranth Singh	12	17	31
Divyanshi Dyori	20	16	27
Kunal Bisht	12	16	7
Navneet P	96	94	87
Abhinandan Kumar	93	92	80
Aafrin Alam	56	89	84

Q3. Display data using head() Function,tail() Function and Slicing data[4:21]
Solution:

```python
import pandas as pd  # Import the pandas library for data manipulation

# Function to return a separator line of 90 dashes
def line():
    return "-" * 90

# Load the CSV data from "8362_data.csv" into a DataFrame 'data'
data = pd.read_csv("8362_data.csv")

# Extract specific columns from the 'data' DataFrame
names = data["Student Name"]     # Extract the "Student Name" column
attd = data["Attendance %"]      # Extract the "Attendance %" column
mcq = data["FD MCQ %"]           # Extract the "FD MCQ %" column (MCQ completion percentage)
lab = data["TD Lab %"]           # Extract the "TD Lab %" column (Lab completion percentage)

# Create a new DataFrame 'new_data' with extracted columns and more user-friendly names
new_data = pd.DataFrame({
    "Student Name": names,        # Column for student names
    "Attendance": attd,           # Column for attendance percentages
    "MCQs completed": mcq,        # Column for MCQ completion percentages
    "Labs completed": lab         # Column for Lab completion percentages
})

# Display the last 4 rows of the DataFrame
# The line() function adds a separator of dashes for better formatting
print(f"\nBottom 4\n{line()}\n", new_data.tail(4).to_string(index=False))

# Display the first 4 rows of the DataFrame
print(f"\nTop 4\n{line()}\n", new_data.head(4).to_string(index=False))

# Display the rows from index 4 to 21 (excluding top 4 and bottom 4)
print(f"\nRest Students\n{line()}\n", new_data.loc[4:21].to_string(index=False))

```

Output:

```
Bottom 4
-------------------------------------------------------------------------------------
     Student Name  Attendance  MCQs completed  Labs completed
     Kunal Bisht          12              16               7
       Navneet P          96              94              87
Abhinandan Kumar          93              92              80
     Aafrin Alam          56              89              84

Top 4
-------------------------------------------------------------------------------------
      Student Name  Attendance  MCQs completed  Labs completed
   Dashmeet Singh          95              87              88
        Mr Ketan           66              17              36
    Harsh panchal          40              17              48
Rahil Ahmad  khan          79              46              61
```

```
Rest Students
-------------------------------------------------------------------------------------
            Student Name  Attendance  MCQs completed  Labs completed
           vishal singh          13               0               4
          Kishan Mishra          31              18              18
          Shubham Kumar          60              72              83
             Sonal Garg          66              46              65
        RATAN SRIVASTAV          71              83              82
            Rohit verma          89              85              85
          Yashika Gupta          80              86              84
     Dishant Kumar Moga          75              83              85
            Abhishek .           73              65              66
           Aryan Verma           88              90              88
              Ankush .           94              79              87
           Riya Singla           78              76              80
         Rajneesh Singh          51              19               7
        Khushi Chaudhary         13               0              11
Chauhan Vandana Ramdayal         87              88              85
            Bittu Samui          43              36              14
          Vikranth Singh         12              17              31
         Divyanshi Dyori         20              16              27
```

Q4.Export data set using .to_csv() only selected columns[student name,attendance,mcq,lab]
Solution:

```python
import pandas as pd  # Import the pandas library for data manipulation

# Read the CSV file "8362_data.csv" into a DataFrame 'data'
data = pd.read_csv("8362_data.csv")

# Extract specific columns from the 'data' DataFrame and store them in variables
names = data["Student Name"]     # Extract the "Student Name" column
attd = data["Attendance %"]      # Extract the "Attendance %" column
mcq = data["FD MCQ %"]           # Extract the "FD MCQ %" column (percentage of MCQs completed)
lab = data["TD Lab %"]           # Extract the "TD Lab %" column (percentage of Labs completed)

# Create a new DataFrame 'new_data' with meaningful column names and extracted data
new_data = pd.DataFrame({
    "Student Name": names,        # Assign the "Student Name" column
    "Attendance": attd,           # Rename "Attendance %" to "Attendance"
    "MCQs completed": mcq,        # Rename "FD MCQ %" to "MCQs completed"
    "Labs completed": lab         # Rename "TD Lab %" to "Labs completed"
})

# Save the new DataFrame 'new_data' to a CSV file named "Filtered Data.csv" without the index column
new_data.to_csv("Filtered Data.csv", index=False)

```

Output:

Student Name	Attendance	MCQs completed	Labs completed
Dashmeet Singh	95	87	88
Mr Ketan	66	17	36
Harsh panchal	40	17	48
Rahil Ahmad khan	79	46	61
vishal singh	13	0	4
Kishan Mishra	31	18	18
Shubham Kumar	60	72	83
Sonal Garg	66	46	65
RATAN SRIVASTAV	71	83	82
Rohit verma	89	85	85
Yashika Gupta	80	86	84
Dishant Kumar Moga	75	83	85
Abhishek .	73	65	66
Aryan Verma	88	90	88
Ankush .	94	79	87
Riya Singla	78	76	80
Rajneesh Singh	51	19	7
Khushi Chaudhary	13	0	11
Chauhan Vandana Ramdayal	87	88	85
Bittu Samui	43	36	14
Vikranth Singh	12	17	31
Divyanshi Dyori	20	16	27
Kunal Bisht	12	16	7
Navneet P	96	94	87
Abhinandan Kumar	93	92	80
Aafrin Alam	56	89	84