

 <b>Marwadi</b> University Marwadi Chandrana Group	<b>NAAC</b>  <b>A+</b>	
<b>Subject: Programming With Python (01CT1309)</b>	<b>Aim:</b> Practical based on Data Loading, Storage and File Formats	
<b>Experiment No: 22</b>	<b>Date:</b>	<b>Enrollment No: 92400133181</b>

**Aim:** Practical based on Data Loading, Storage and File Formats

#### **IDE:**

load, manipulate, and store data using Python (over reading and writing CSV, JSON, and Excel files)

#### Library Installation

```
pip install pandas openpyxl
```

#### **Sample Data:**

Create a folder for this experiment and add the following sample data files:

**sample\_data.csv** (Name,Age,City)

Alice,30,New York

Bob,25,Los Angeles

Charlie,35,Chicago)

**sample\_data.json** ([

{"Name": "David", "Age": 28, "City": "San Francisco"},

{"Name": "Eve", "Age": 22, "City": "Seattle"}

])

**sample\_data.xlsx** (you can create this using Excel with similar data)\\

#### Loading Data from CSV

Read the CSV file and perform basic data manipulation.

```
import pandas as pd
```

```
# Load data from CSV
```

 <b>Marwadi</b> University Marwadi Chandrana Group	<b>NAAC</b>  <b>A+</b>	<b>Marwadi University</b> <b>Faculty of Engineering &amp; Technology</b> <b>Department of Information and Communication Technology</b>
<b>Subject: Programming With Python (01CT1309)</b>	<b>Aim:</b> Practical based on Data Loading, Storage and File Formats	
<b>Experiment No: 22</b>	<b>Date:</b>	<b>Enrollment No: 92400133181</b>

```
csv_file_path = 'sample_data.csv'
```

```
df_csv = pd.read_csv(csv_file_path)
```

```
# Display the DataFrame
```

```
print("CSV Data:")
```

```
print(df_csv)
```

```
# Basic data manipulation: Filter by age
```

```
filtered_data = df_csv[df_csv['Age'] > 30]
```

```
print("\nFiltered Data (Age > 30):")
```

```
print(filtered_data)
```

Loading Data from JSON

Read the JSON file and manipulate the data.

```
# Load data from JSON
```

```
json_file_path = 'sample_data.json'
```

```
df_json = pd.read_json(json_file_path)
```

```
# Display the DataFrame
```

```
print("\nJSON Data:")
```

```
print(df_json)
```

```
# Basic data manipulation: Find the average age
```

 <b>Marwadi</b> University <small>Marwadi Chandrana Group</small>	<b>NAAC</b>  <b>A+</b>	<b>Marwadi University</b> <b>Faculty of Engineering &amp; Technology</b> <b>Department of Information and Communication Technology</b>
<b>Subject: Programming With Python (01CT1309)</b>	<b>Aim:</b> Practical based on Data Loading, Storage and File Formats	
<b>Experiment No: 22</b>	<b>Date:</b>	<b>Enrollment No: 92400133181</b>

```
average_age = df_json['Age'].mean()
```

```
print("\nAverage Age:", average_age)
```

Loading Data from Excel

Read the Excel file and display its contents.

```
# Load data from Excel
```

```
excel_file_path = 'sample_data.xlsx'
```

```
df_excel = pd.read_excel(excel_file_path)
```

```
# Display the DataFrame
```

```
print("\nExcel Data:")
```

```
print(df_excel)
```

```
# Basic data manipulation: Count the number of entries
```

```
entry_count = df_excel.shape[0]
```

```
print("\nNumber of entries in Excel file:", entry_count)
```

Writing Data to Different Formats

Save manipulated DataFrames to new files in different formats.

```
# Save filtered CSV data to a new file
```

```
filtered_data.to_csv('filtered_data.csv', index=False)
```

```
print("\nFiltered data saved to 'filtered_data.csv'.")
```

```
# Save DataFrame to a new JSON file
```

 <b>Marwadi</b> University <small>Marwadi Chandrana Group</small>	<b>NAAC</b> 	<b>Marwadi University</b> <b>Faculty of Engineering &amp; Technology</b> <b>Department of Information and Communication Technology</b>
<b>Subject: Programming With Python (01CT1309)</b>	<b>Aim:</b> Practical based on Data Loading, Storage and File Formats	
<b>Experiment No: 22</b>	<b>Date:</b>	<b>Enrollment No: 92400133181</b>

```
df_json.to_json('new_data.json', orient='records', lines=True)
```

```
print("JSON data saved to 'new_data.json'.")
```

```
# Save DataFrame to a new Excel file
```

```
df_excel.to_excel('new_data.xlsx', index=False)
```

```
print("Excel data saved to 'new_data.xlsx'.")
```



**Subject: Programming With Python (01CT1309)**

**Aim:** Practical based on Data Loading, Storage and File Formats

**Experiment No: 22**

**Date:**

**Enrollment No: 92400133181**

```
1 import pandas as pd
2
3 # Load data from CSV
4 csv_file_path = r"C:\Users\devah\Desktop\sample_data.csv"
5 df_csv = pd.read_csv(r"C:\Users\devah\Desktop\sample_data.csv")
6
7 # Display the DataFrame
8 print("CSV Data:")
9 print(df_csv)
10
11 # Basic data manipulation: Filter by age
12 filtered_data = df_csv[df_csv['Age'] > 30]
13 print("\nFiltered Data (Age > 30):")
14 print(filtered_data)
15
16 # Load data from JSON
17 json_file_path = r"C:\Users\devah\Desktop\new_data.json"
18 df_json = pd.read_json(r"C:\Users\devah\Desktop\new_data.json")
19
20 # Display the DataFrame
21 print("\nJSON Data:")
22 print(df_json)
23
24 # Basic data manipulation: Find the average age
25 average_age = df_json['Age'].mean()
26 print("\nAverage Age:", average_age)
27
28 # Load data from Excel
29 excel_file_path = r"C:\Users\devah\Desktop\sample_data.xlsx"
30 df_excel = pd.read_excel(r"C:\Users\devah\Desktop\sample_data.xlsx")
31
32 # Display the DataFrame
33 print("\nExcel Data:")
34 print(df_excel)
35
```



**Subject: Programming With Python (01CT1309)**

**Aim:** Practical based on Data Loading, Storage and File Formats

**Experiment No: 22**

**Date:**

**Enrollment No: 92400133181**

```
35
36     # Basic data manipulation: Count the number of entries
37     entry_count = df_excel.shape[0]
38     print("\nNumber of entries in Excel file:", entry_count)
39
40     # Save filtered CSV data to a new file
41     filtered_data.to_csv('filtered_data.csv', index=False)
42     print("\nFiltered data saved to 'filtered_data.csv'.")
43
44     # Save DataFrame to a new JSON file
45     df_json.to_json('new_data.json', orient='records', lines=True)
46     print("JSON data saved to 'new_data.json'.")
47
48     # Save DataFrame to a new Excel file
49     df_excel.to_excel('new_data.xlsx', index=False)
50     print("Excel data saved to 'new_data.xlsx'.")
```

**Output:**

.csv file:



**Subject: Programming With Python (01CT1309)**

**Aim:** Practical based on Data Loading, Storage and File Formats

**Experiment No: 22**

**Date:**

**Enrollment No: 92400133181**

```
Console 1/A X
.... df_csv = pd.read_csv(r"C:\Users\devah\Desktop\sample_data.csv")
.... 
.... # Display the DataFrame
.... print("CSV Data:")
.... print(df_csv)
.... 
.... # Basic data manipulation: Filter by age
.... filtered_data = df_csv[df_csv['Age'] > 30]
.... print("\nFiltered Data (Age > 30):")
.... print(filtered_data)
.... 

CSV Data:
      Name    Age        City   Unnamed: 3
0    Alice  30.0    New York       NaN
1     Bob  25.0  Los Angeles       NaN
2  Charlie  35.0    Chicago       NaN
3      NaN    NaN           NaN       NaN

Filtered Data (Age > 30):
      Name    Age        City   Unnamed: 3
2  Charlie  35.0    Chicago       NaN
```

```
In [17]:
.... filtered_data.to_csv('filtered_data.csv', index=False)
.... print("\nFiltered data saved to 'filtered_data.csv'.")
.... 

Filtered data saved to 'filtered_data.csv'.
```

.json file:



**Marwadi University**  
**Faculty of Engineering & Technology**  
**Department of Information and Communication Technology**

**Subject: Programming With Python (01CT1309)**

**Aim:** Practical based on Data Loading, Storage and File Formats

**Experiment No: 22**

**Date:**

**Enrollment No: 92400133181**

```
Resulting DataFrame:  
    Name: "David"  Age: 28  City: "San Francisco"  
0  Name: "Eve"   Age: 22  City: "Seattle"  
  
Dtypes:  
Name:          object  
"David":       object  
Age:           object  
28:            int64  
City:          object  
"San Francisco": object  
dtype: object
```

.xlsx file:

```
In [20]:  
....: excel_file_path = r"C:\Users\devah\Desktop\sample_data.xlsx"  
....: df_excel = pd.read_excel(r"C:\Users\devah\Desktop\sample_data.xlsx")  
....:  
....: # Display the DataFrame  
....: print("\nExcel Data:")  
....: print(df_excel)  
....:  
  
Excel Data:  
      Name  Age        City  
0    Alice  30  New York  
1     Bob   25  Los Angeles  
2  Charlie  35     Chicago
```



**Subject: Programming With Python (01CT1309)**

**Aim:** Practical based on Data Loading, Storage and File Formats

**Experiment No: 22**

**Date:**

**Enrollment No: 92400133181**

```
In [22]:  
...: df_excel.to_excel('new_data.xlsx', index=False)  
...: print("Excel data saved to 'new_data.xlsx'.")  
...:  
...:  
Excel data saved to 'new_data.xlsx'.  
In [22]:
```

Count the number of entries:

```
In [23]:  
...: entry_count = df_excel.shape[0]  
...: print("\nNumber of entries in Excel file:", entry_count)  
  
Number of entries in Excel file: 3
```

## Post Lab:

**Write a code snippet to check the data types of each column in a DataFrame.**

**Code:**

```
1 import pandas as pd  
2  
3 df = pd.DataFrame({  
4     'Name': ['A', 'B', 'C'],  
5     'Age': [20, 25, 30],  
6     'Score': [88.5, 92.3, 79.8],  
7     'Passed': [True, False, True]  
8 })  
9  
10 print(df.dtypes)  
11
```

**Output:**



**Subject: Programming With Python (01CT1309)**

**Aim:** Practical based on Data Loading, Storage and File Formats

**Experiment No: 22**

**Date:**

**Enrollment No: 92400133181**

```
In [24]: %runfile 'C:/Users/devah/Documents/PWP/Check data types in DataFrame PWP Lab 22.py' --wdir
Name      object
Age       int64
Score     float64
Passed    bool
dtype: object
```

**Write a code snippet that demonstrates how to fill missing values with the mean of a column.**

**Code:**

```
1 import pandas as pd
2 import numpy as np
3
4 df = pd.DataFrame({
5     'Age': [25, 30, np.nan, 40, np.nan]
6 })
7
8 df[ 'Age' ] = df[ 'Age' ].fillna(df[ 'Age' ].mean())
9
10 print(df)
11
```

**Output:**

```
In [25]: %runfile 'C:/Users/devah/Documents/PWP/Fill missing values with mean of column PWP Lab 22.py' --wdir
          Age
0  25.000000
1  30.000000
2  31.666667
3  40.000000
4  31.666667
```

 <b>Marwadi</b> University <small>Marwadi Chandarana Group</small>	<b>NAAC</b>  <b>A+</b>	<b>Marwadi University</b> <b>Faculty of Engineering &amp; Technology</b> <b>Department of Information and Communication Technology</b>
<b>Subject: Programming With Python (01CT1309)</b>	<b>Aim:</b> Practical based on Data Loading, Storage and File Formats	
<b>Experiment No: 22</b>	<b>Date:</b>	<b>Enrollment No: 92400133181</b>

**GitHub:**

<https://github.com/mallaadisrinivasu132035-code/python.git>