Department of Engineering Sciences and Technology,

Second Year Btech in Computer Science

Project Based Learning-Python

Assignment - 13

Name - Paritosh kolwadkar

SRN – 31231313

Roll no – 39

Batch – D2

Problem statement : **Write a program to read data from a CSV file using Pandas. Display the first few rows (head()), information (info()), and basic statistics (describe()) of the DataFrame. Save the modified DataFrame to a new CSV file.**

Pre-requisites: Install the Pandas library:

```
pip install pandas
```

Basic understanding of CSV files and data manipulation.

Code:

```
# Import Pandas

import pandas as pd



# Read data from a CSV file

# Replace 'data.csv' with the path to your CSV file

df = pd.read_csv('data.csv')



# Display the first few rows of the DataFrame

print("First few rows of the DataFrame:")
```

```
print(df.head())


# Display information about the DataFrame

print("\nDataFrame Information:")

print(df.info())


# Display basic statistics for numerical columns

print("\nBasic Statistics:")

print(df.describe())


# Modify the DataFrame (example: adding a new column)

df['NewColumn'] = 'SampleData'


# Save the modified DataFrame to a new CSV file

df.to_csv('modified_data.csv', index=False)

print("\nModified DataFrame has been saved to 'modified_data.csv'.")
```

Explanation :

Read CSV File:

- **`pd.read_csv('data.csv')`: Reads the specified CSV file into a Pandas DataFrame.**

Display First Few Rows:

- **`df.head()`: Displays the first 5 rows (default) of the DataFrame.**

DataFrame Information:

- **`df.info()`: Provides a summary of the DataFrame, including column names, data types, and non-null counts.**

Basic Statistics:

- **`df.describe()`: Displays statistical summaries (mean, min, max, etc.) for numeric columns.**

Modify DataFrame:

- **Example: Adds a new column (`NewColumn`) with a sample value for demonstration purposes.**

Save to New CSV File:

- **`df.to_csv('modified_data.csv', index=False)`: Saves the modified DataFrame to a new CSV file without including the index.**
-

Output:

Given a sample CSV file `data.csv`:

```
Name,Age,Score
Alice,25,85
Bob,30,90
Charlie,35,95
```

Console Output:

```
First few rows of the DataFrame:
      Name  Age  Score
0    Alice   25     85
1      Bob   30     90
2  Charlie   35     95

DataFrame Information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3 entries, 0 to 2
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Name    3 non-null      object
```

```
 1   Age     3 non-null      int64
 2   Score   3 non-null      int64
dtypes: int64(2), object(1)
memory usage: 200.0+ bytes

Basic Statistics:
             Age      Score
count   3.000000   3.000000
mean   30.000000  90.000000
std     5.000000   5.000000
min    25.000000  85.000000
25%    27.500000  87.500000
50%    30.000000  90.000000
75%    32.500000  92.500000
max    35.000000  95.000000

Modified DataFrame has been saved to 'modified_data.csv'.
```

New CSV File (modified_data.csv):

```
Name,Age,Score,NewColumn
Alice,25,85,SampleData
Bob,30,90,SampleData
Charlie,35,95,SampleData
```

Output Explained:

- **The program demonstrates reading, analyzing, and modifying CSV data using Pandas.**
- **The modified DataFrame is saved to a new file, showing data manipulation capabilities.**