

**Name: Pradip Bochare**

### **Python Coding Challenge**

#### **Explain Pandas for Data Processing:**

Pandas is a popular open-source data manipulation and analysis library for the Python programming language. It provides data structures and functions needed to manipulate and analyze structured data, primarily in the form of tabular data like spreadsheets and SQL tables.

#### **Key components of Pandas**

1. **DataFrame:**

The central data structure in Pandas is the DataFrame, a two-dimensional table with labeled axes (rows and columns). Columns can have different data types (integers, floats, strings) and can be heterogeneous.

2. **Series:**

A one-dimensional labeled array capable of holding any data type. A single column of a DataFrame is essentially a Series.

3. **Index:**

Pandas uses the Index to label the rows and columns of a DataFrame. It helps in selecting, slicing, and manipulating data.

4. **Data Cleaning and Transformation:**

Pandas provides numerous functions for handling missing data, filtering, and cleaning data. It allows reshaping and pivoting data, merging and concatenating datasets, and handling duplicates.

5. **Data Selection and Indexing:**

Pandas offers powerful tools for selecting, indexing, and filtering data. It supports both label-based and position-based indexing.

## 6. Grouping and Aggregation:

Pandas enables grouping of data based on one or more criteria and then applying a function to each group independently. Functions like `groupby()` and `agg()` facilitate these operations.

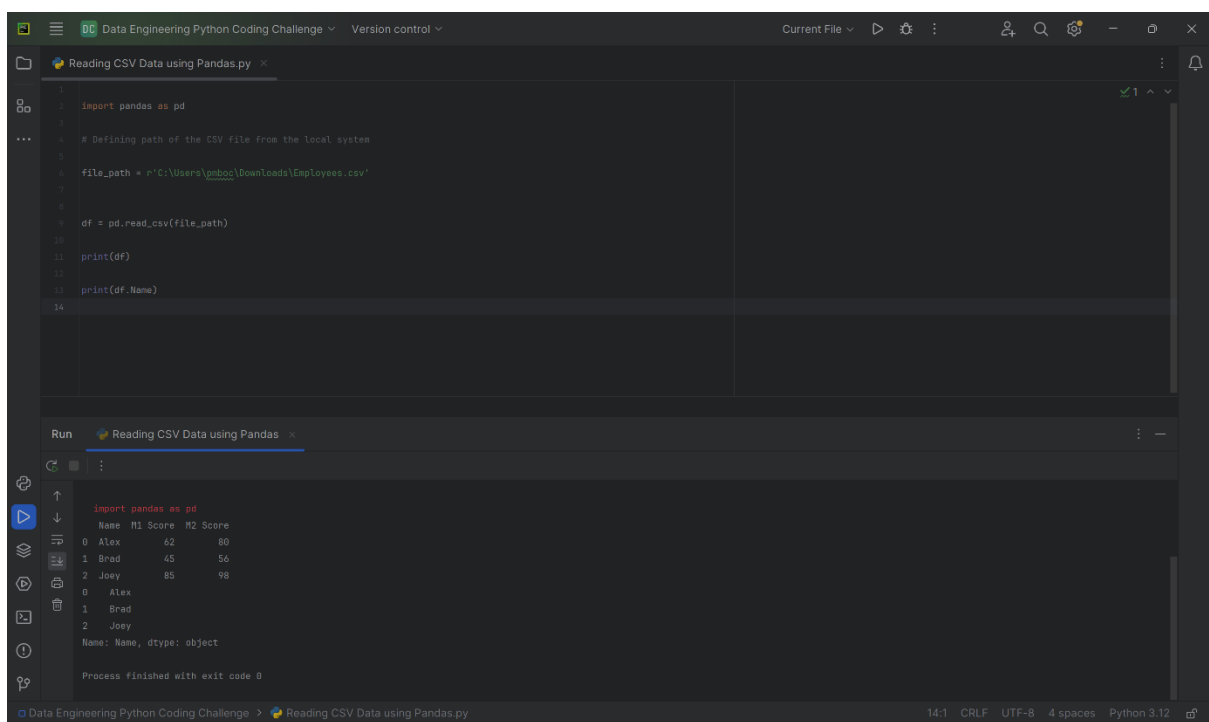
Pandas supports reading and writing data in various formats, such as CSV, Excel, SQL databases

### Execute Reading CSV Data using Pandas

To read CSV file using Pandas, we use the '`pandas.read_csv()`' function

First, we need to import pandas module after that give file path of the csv file we want to read.

After that we can read csv file by `pd.read_csv(file path)`

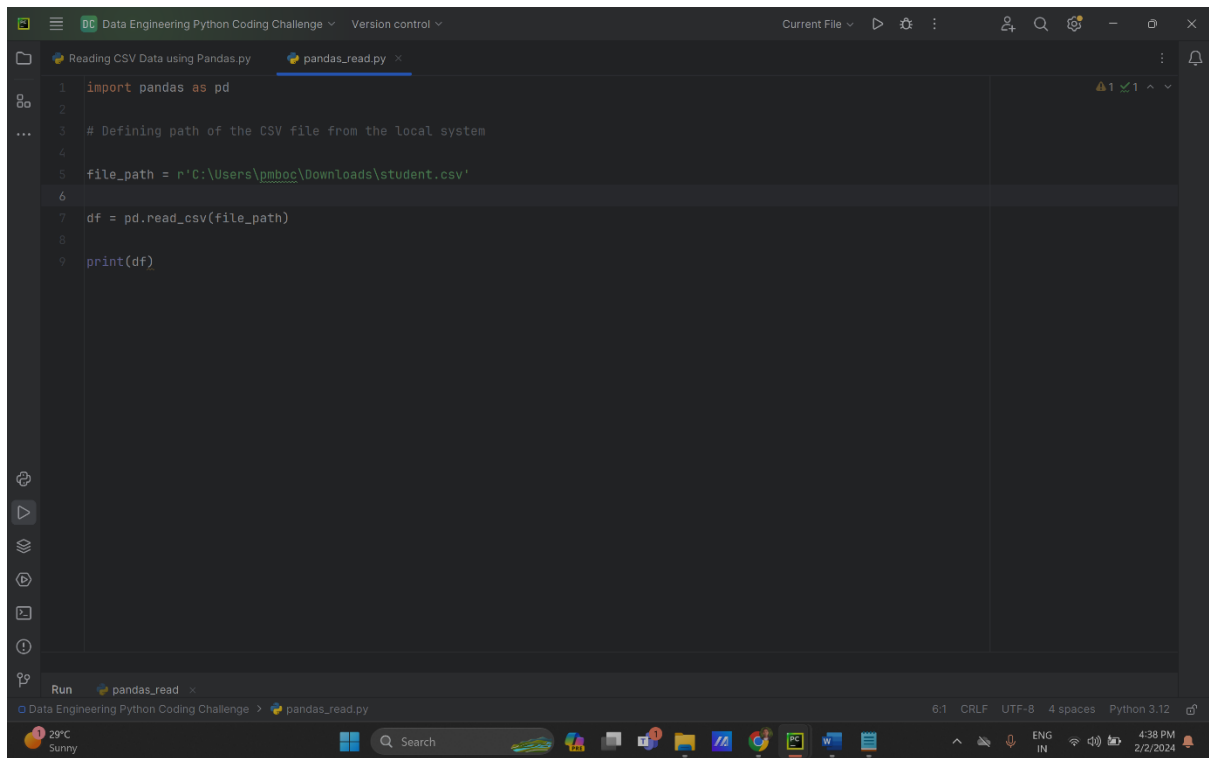


```
1 import pandas as pd
2
3 # Defining path of the CSV file from the local system
4 file_path = r"C:\Users\gmboc\Downloads\Employee.csv"
5
6 df = pd.read_csv(file_path)
7
8 print(df)
9
10 print(df.Name)
```

Run Reading CSV Data using Pandas

```
import pandas as pd
Name M1 Score M2 Score
0 Alex 62 80
1 Brad 65 56
2 Joey 85 98
0 Alex
1 Brad
2 Joey
Name: Name, dtype: object
Process finished with exit code 0
```

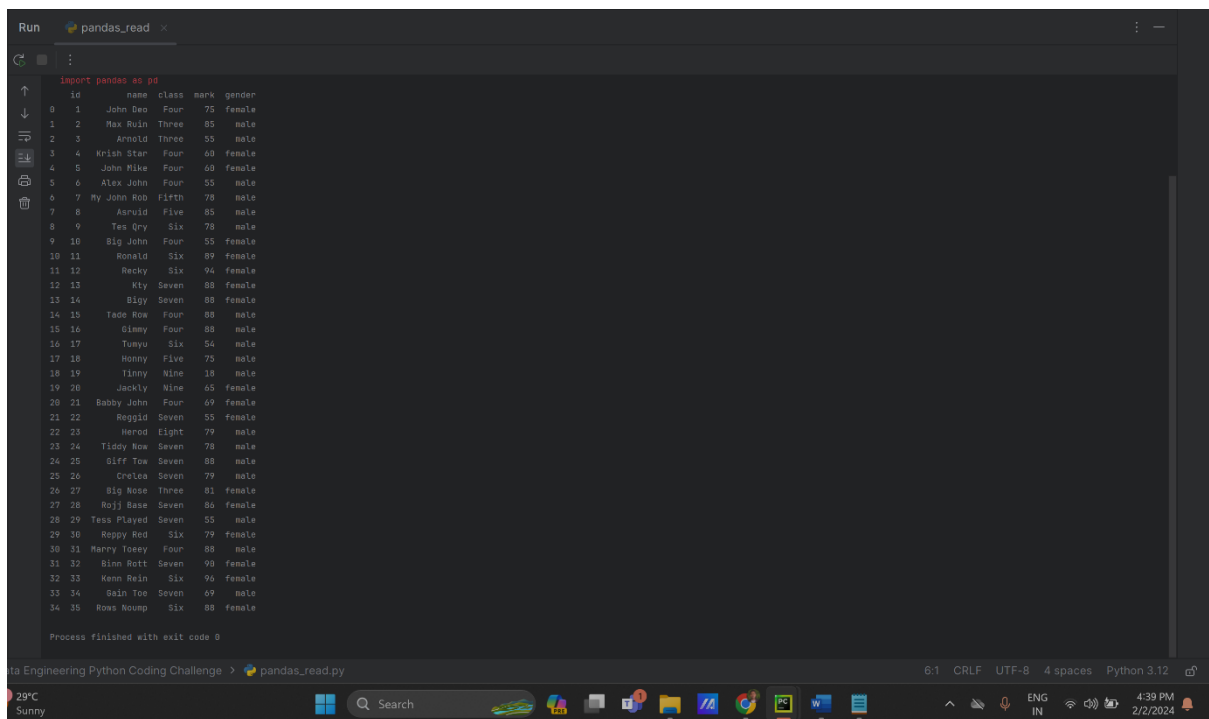
## Read Data from CSV Files to Pandas Dataframes



```
1 import pandas as pd
2
3 # Defining path of the CSV file from the local system
4
5 file_path = r'C:\Users\pmboc\Downloads\student.csv'
6
7 df = pd.read_csv(file_path)
8
9 print(df)
```

The screenshot shows a code editor window titled "Data Engineering Python Coding Challenge" with a file named "pandas\_read.py". The code imports pandas as pd, defines a file path, and uses pd.read\_csv to load the data into a DataFrame, which is then printed. The status bar at the bottom indicates the file encoding is UTF-8, 4 spaces, and Python 3.12.

## Result



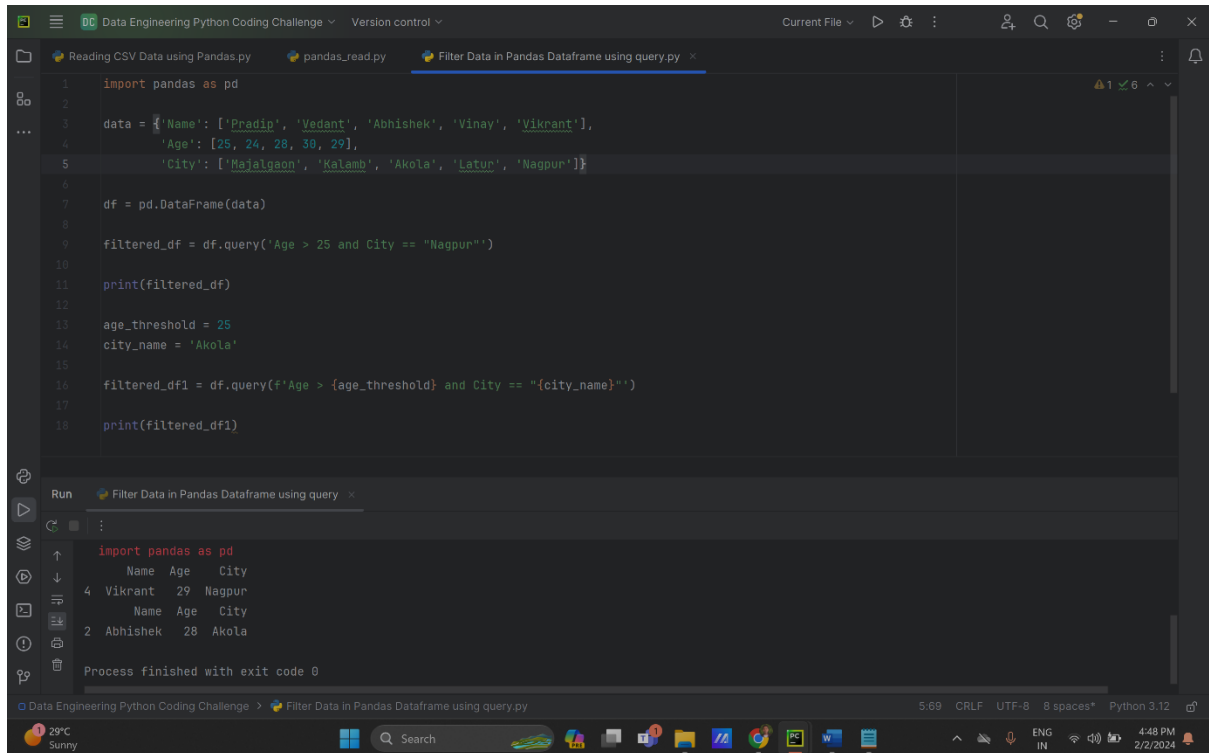
```
Run pandas_read
import pandas as pd
id      name  class  mark  gender
0 1    John  Des    Four    75  female
1 2    Max  Ruin   Three   85  male
2 3    Arnold Three  55  male
3 4    Krish Star   Four   60  female
4 5    John Mike   Four   60  female
5 6    Alex John  Four   55  male
6 7    My John Rob Fifth  78  male
7 8    Asruid Five  85  male
8 9    Ten Gry    Six   78  male
9 10   Big John  Four   55  female
10 11   Ronald  Six   89  female
11 12   Recky   Six   94  female
12 13      Kty  Seven  88  female
13 14   Bigy  Seven  88  female
14 15   Tade Row   Four  88  male
15 16   Glimy  Four  88  male
16 17   Tumyu  Six   54  male
17 18   Honny  Five  75  male
18 19   linny  Nine  18  male
19 20   Jackly  Nine  65  female
20 21   Baby John Four  60  female
21 22   Reggid Seven  55  female
22 23   Herod  Eight  79  male
23 24   Tiddy Row Seven  78  male
24 25   Giff Tow  Seven  88  male
25 26   Crelea Seven  79  male
26 27   Big Nose Three  81  female
27 28   Robj Base Seven  86  female
28 29   Tess Played Seven  55  male
29 30   Regpy Red  Six   79  female
30 31   Harry Toey Four  88  male
31 32   Rinn Ratt Seven  98  female
32 33   Kane Rein  Six   96  female
33 34   Gain Toe Seven  69  male
34 35   Rows Nounp Six  88  female
Process finished with exit code 0
```

The screenshot shows the output of the Python script, displaying a DataFrame with 35 rows of student data. The columns are id, name, class, mark, and gender. The data is displayed in a scrollable list format. The status bar at the bottom indicates the file encoding is UTF-8, 4 spaces, and Python 3.12.



## Filter Data in Pandas Dataframe using query.

In pandas we can filter Dataframe using 'query ()' method which allows you to express complex filtering conditions using SQL like syntax



```
1 import pandas as pd
2
3 data = {'Name': ['Pradip', 'Vedant', 'Abhishek', 'Vinay', 'Vikrant'],
4         'Age': [25, 24, 28, 30, 29],
5         'City': ['Majalgaon', 'Kalamb', 'Akola', 'Latun', 'Nagpur']}
6
7 df = pd.DataFrame(data)
8
9 filtered_df = df.query('Age > 25 and City == "Nagpur"')
10
11 print(filtered_df)
12
13 age_threshold = 25
14 city_name = 'Akola'
15
16 filtered_df1 = df.query(f'Age > {age_threshold} and City == "{city_name}"')
17
18 print(filtered_df1)
```

Run Filter Data in Pandas Dataframe using query

```
import pandas as pd
      Name Age  City
4  Vikrant  29  Nagpur
      Name Age  City
2  Abhishek  28  Akola
```

Process finished with exit code 0

5:59 CRLF UTF-8 8 spaces\* Python 3.12



## Execute with one example Lambda Functions in Python

A lambda function is a small anonymous function.

A lambda function can take any number of arguments, but can only have one expression.

The screenshot shows a PyCharm IDE window titled "Data Engineering Python Coding Challenge". The editor displays a file named "lambda function.py" with the following code:

```
1 usage
2 def square(x):
3     return x ** 2
4
5 # lambda function
6 square_lambda = lambda x: x ** 2
7
8 print(square(5))
9 print(square_lambda(5))
10
11 # Using lambda with map to square each element
12 numbers = [1, 2, 3, 4, 5]
13 squared_numbers = list(map(lambda x: x**2, numbers))
14
15 print(squared_numbers)
```

The Run console at the bottom shows the execution output:

```
"C:\Program Files\Python312\python.exe" "C:\Users\pmboc\PycharmProjects\Data Engineering Python Coding Challenge\lambda function.py"
25
25
[1, 4, 9, 16, 25]
Process finished with exit code 0
```

The status bar at the bottom indicates the file encoding is UTF-8, 4 spaces, and Python 3.12.



## Read JSON Strings to Python dicts or lists

In python we can use 'json' module to convert JSON strings to python dictionaries or lists. This module provides methods for encoding and decoding JSON data.

'json.loads()' is used to parse the json string into python dictionary or list. Resulting python\_dict variable contains the equivalent python data structure.

The screenshot shows the PyCharm IDE with a file named 'json string to python list.py'. The code defines two JSON strings and uses the `json.loads()` function to convert them into Python objects. The first string represents a dictionary with keys 'name', 'age', and 'city'. The second string represents a list of integers. The output of the script is displayed in the Run console.

```
1 import json
2
3 # json string
4 json_string = '{"name": "Pradip", "age": 24, "city": "Pune"}'
5
6 python_dict = json.loads(json_string)
7
8 print(python_dict)
9
10 import json
11
12 # json string representing a list
13 json_string_list = '[1, 2, 3, 4, 5]'
14
15 python_list = json.loads(json_string_list)
16
17 print(python_list)
18
```

Run console output:

```
"C:\Program Files\Python312\python.exe" "C:\Users\pmboc\PycharmProjects\Data Engineering Python Coding Challenge\json string to python list.py"
{'name': 'Pradip', 'age': 24, 'city': 'Pune'}
[1, 2, 3, 4, 5]
Process finished with exit code 0
```

❖ Instead of json string we can use JSON file also here we use ‘`json.load`’ function to read directly from file

The screenshot shows the PyCharm IDE with a file named 'json file.py'. The code defines the path to a JSON file, opens it, and uses the `json.load()` function to read the data. The output of the script is displayed in the Run console.

```
1 import json
2
3 json_file_path = r'C:\Users\pmboc\Downloads\sample1.json'
4
5
6 with open(json_file_path, 'r') as file:
7     python_object = json.load(file)
8
9
10 print(python_object)
11
```

Run console output:

```
"C:\Program Files\Python312\python.exe" "C:\Users\pmboc\PycharmProjects\Data Engineering Python Coding Challenge\json file.py"
{'fruit': 'Apple', 'size': 'Large', 'color': 'Red'}
Process finished with exit code 0
```

