

Python Coding Assessment

1)

Explain Pandas for Data Processing

Execute Reading CSV Data using Pandas

Read Data from CSV Files to Pandas Dataframes

Filter Data in Pandas Dataframe using query.

a) Explain Pandas for Data Processing

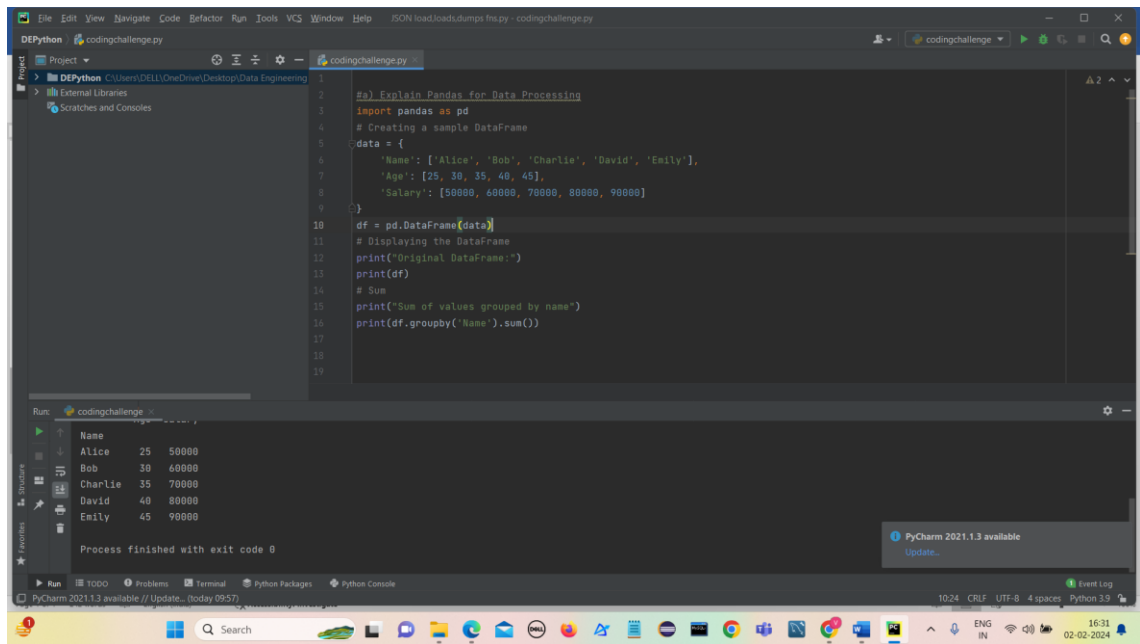
Pandas is an open-source Python library that provides high-performance, easy-to-use data structures, and data analysis tools for working with structured data. It is built on top of the NumPy library and is widely used for data manipulation, data cleaning, data exploration, and data analysis tasks in Python.

Data Structures:

- 1) **Series:** It is a One-dimensional labeled array .It can hold any data type (e.g., integers, floats, strings, etc.). It is similar to a one-dimensional NumPy array but it has associated index with it
- 2) **DataFrame:** Two-dimensional labeled data structure with columns of different data types. It can be a spreadsheet or a SQL table. Each column in a DataFrame is a Series.

Example:

We can process the data in many ways. We can find the sum of columns using grouping, filter data based on condition etc.,



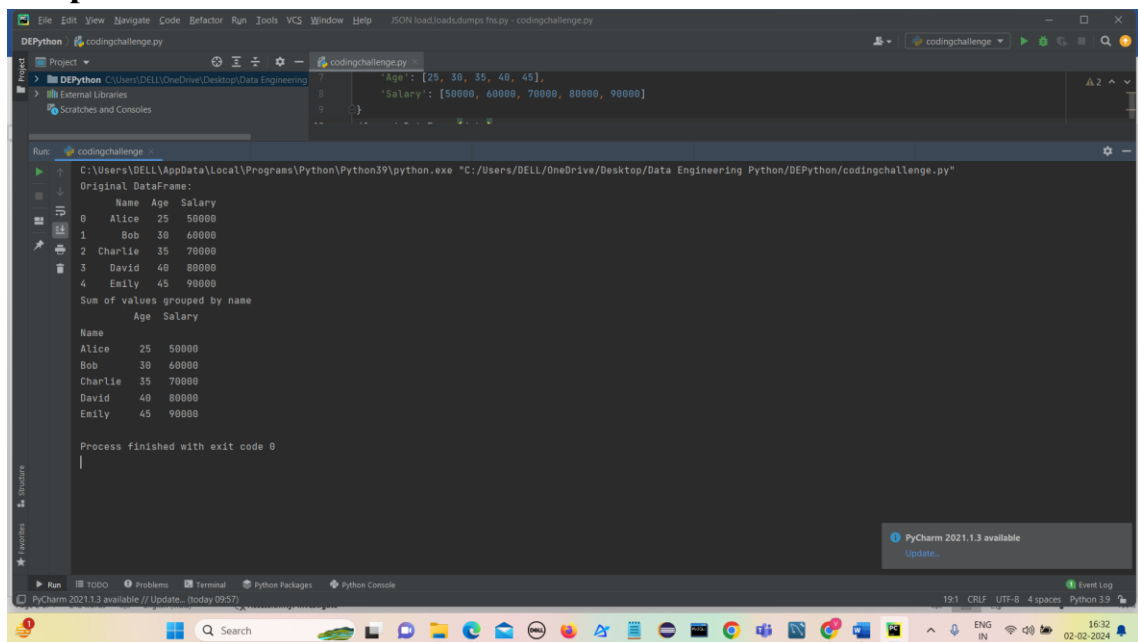
```
1 #a) Explain Pandas for Data Processing
2
3 import pandas as pd
4 # Creating a sample DataFrame
5
6 data = {
7     'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Emily'],
8     'Age': [25, 30, 35, 40, 45],
9     'Salary': [50000, 60000, 70000, 80000, 90000]
10 }
11
12 df = pd.DataFrame(data)
13 # Displaying the DataFrame
14 print("Original DataFrame:")
15 print(df)
16 # Sum
17 print("Sum of values grouped by name")
18 print(df.groupby('Name').sum())
19
```

Run console output:

Name	Age	Salary
Alice	25	50000
Bob	30	60000
Charlie	35	70000
David	40	80000
Emily	45	90000

Process finished with exit code 0

Output:



```
1 #a) Explain Pandas for Data Processing
2
3 import pandas as pd
4 # Creating a sample DataFrame
5
6 data = {
7     'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Emily'],
8     'Age': [25, 30, 35, 40, 45],
9     'Salary': [50000, 60000, 70000, 80000, 90000]
10 }
11
12 df = pd.DataFrame(data)
13 # Displaying the DataFrame
14 print("Original DataFrame:")
15 print(df)
16 # Sum
17 print("Sum of values grouped by name")
18 print(df.groupby('Name').sum())
19
```

Run console output:

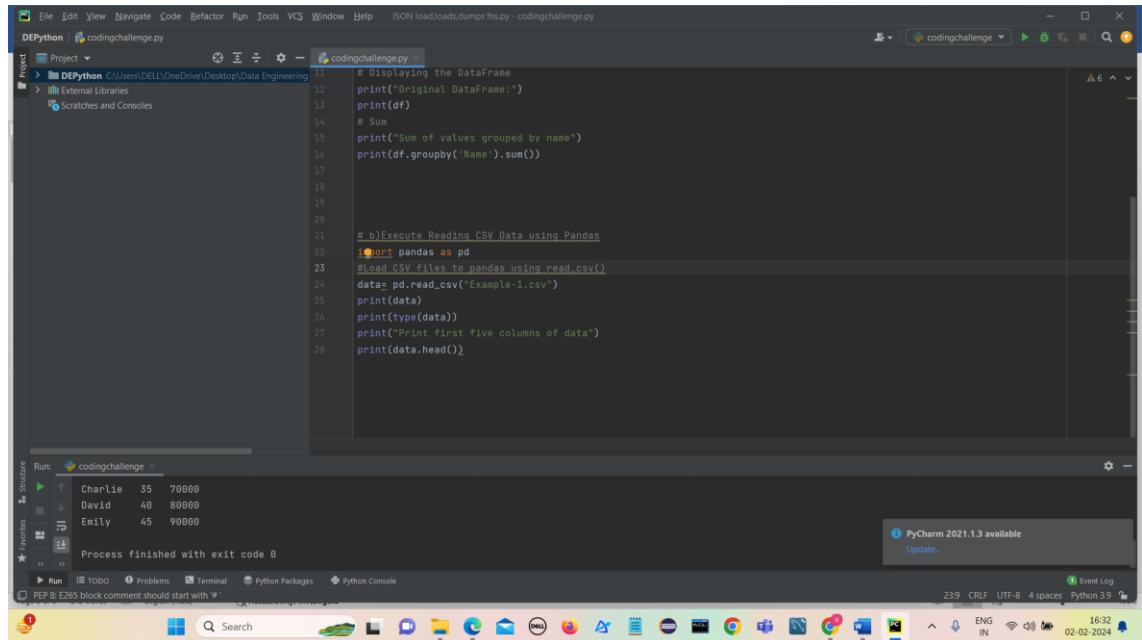
```
C:\Users\DELL\AppData\Local\Programs\Python\Python39\python.exe "C:/Users/DELL/OneDrive/Desktop/Data Engineering Python/DEPython/codingchallenge.py"
Original DataFrame:
   Name  Age  Salary
0  Alice   25   50000
1   Bob   30   60000
2 Charlie   35   70000
3  David   40   80000
4  Emily   45   90000
Sum of values grouped by name
   Name  Age  Salary
Name
Alice   25   50000
Bob     30   60000
Charlie  35   70000
David   40   80000
Emily   45   90000
Process finished with exit code 0
```

Explanation:

In this way data can be processed using pandas. The above code calculates the sum of all values grouped by Name.

b) Execute Reading CSV Data using Pandas

- Import pandas library
- Load CSV files to pandas using read_csv()
- Print the data



The screenshot shows the PyCharm IDE with a Python script named `codingchallenge.py`. The script includes comments and code for reading a CSV file. The output window at the bottom shows the result of the script execution, displaying the first five rows of the CSV data.

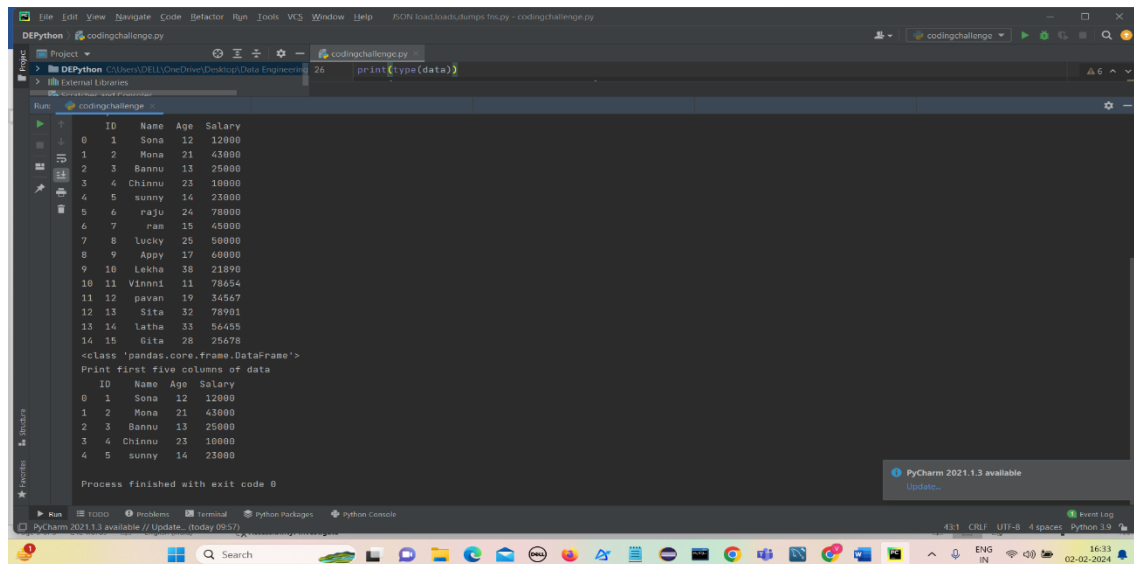
```
11 # Displaying the DataFrame
12 print("Original DataFrame:")
13 print(df)
14 # Sum
15 print("Sum of values grouped by name")
16 print(df.groupby('Name').sum())
17
18
19
20
21 # b)Execute Reading CSV Data using Pandas
22 import pandas as pd
23 #load CSV files to pandas using read_csv()
24 data= pd.read_csv('example-1.csv')
25 print(data)
26 print(type(data))
27 print("Print first five columns of data")
28 print(data.head())
```

Run: codingchallenge

↑	Charlie	35	70000
↓	David	40	80000
	Emily	45	90000

Process finished with exit code 0

Output:



The screenshot shows the PyCharm IDE with the same Python script. The output window at the bottom shows the result of the script execution, displaying the first five rows of the CSV data.

```
26 print(type(data))
```

Run: codingchallenge

ID	Name	Age	Salary
0	1	Sona	12 12000
1	2	Mona	21 43000
2	3	Bannu	13 25000
3	4	Chinnu	23 10000
4	5	sunny	14 23000

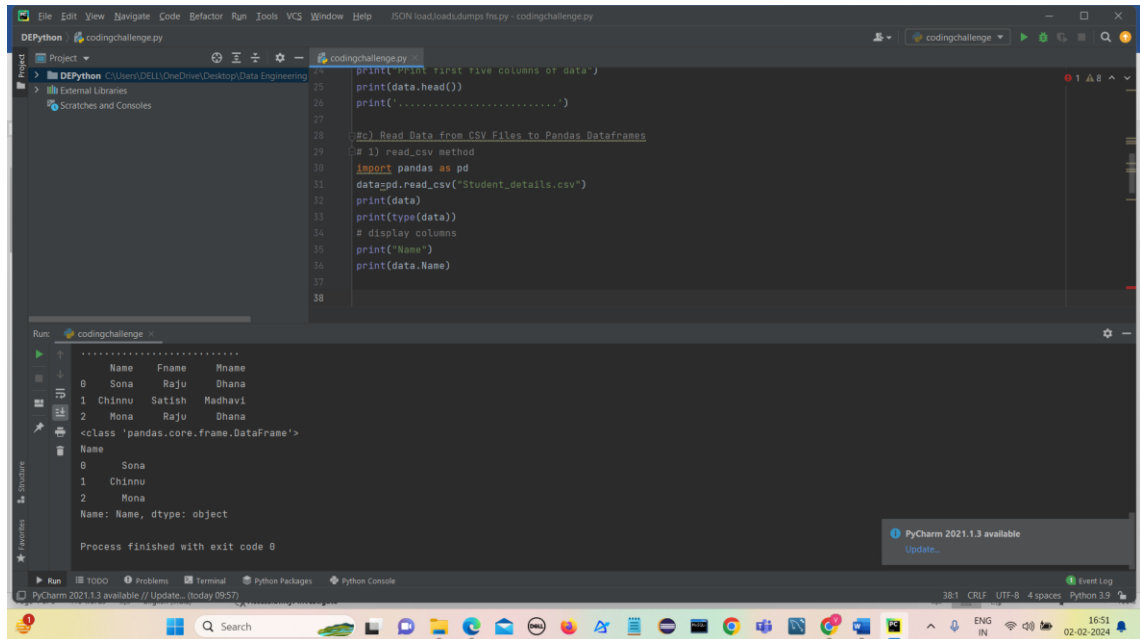
Process finished with exit code 0

Explanation:

In the above code the csv data in Example-1.csv is converted into pandas using `read_csv()` method
`head()` method is used to find first five records in the dataset.

c) Read Data from CSV Files to Pandas Dataframes

1) Using read_csv method



The screenshot shows the PyCharm IDE with a Python file named `codingchallenge.py`. The code uses `pd.read_csv()` to load data from `Student_details.csv`. The output in the Run console shows a DataFrame with columns `Name`, `Fname`, and `Mname`.

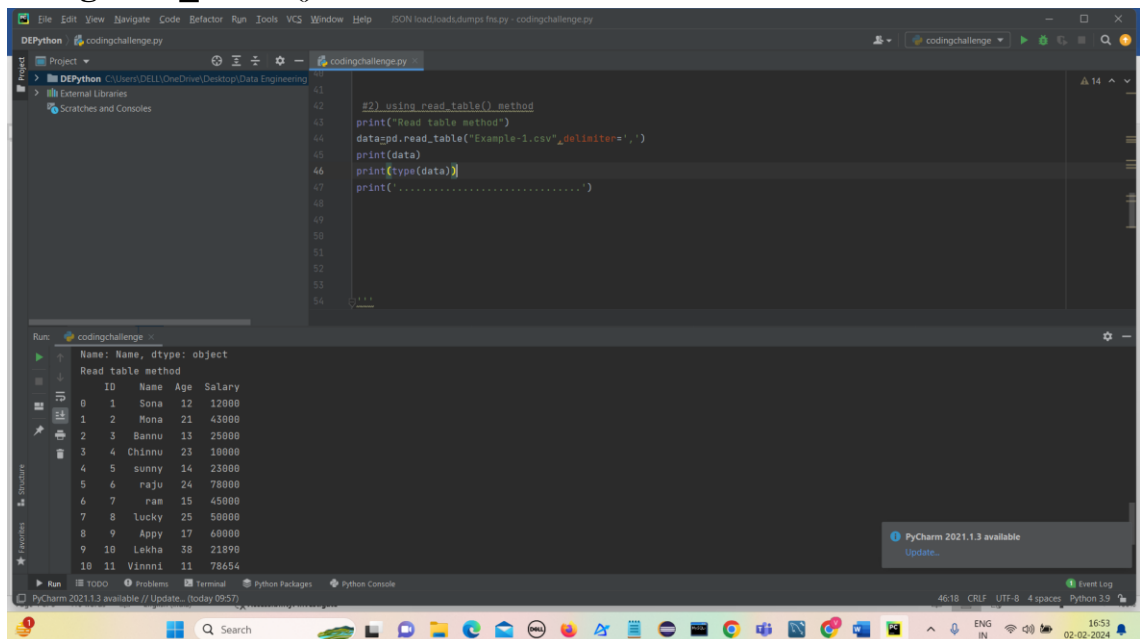
```
codingchallenge.py
14 print('Print first five columns of data.')
15 print(data.head())
16 print('.....')
17
18 #2) Read Data from CSV Files to Pandas Dataframes
19 # 1) read_csv method
20 import pandas as pd
21 data=pd.read_csv("Student_details.csv")
22 print(data)
23 print(type(data))
24 # display columns
25 print("Name")
26 print(data.Name)
27
28
```

```
Run: codingchallenge
.....
      Name  Fname  Mname
0    Sona   Raju   Dhana
1  Chinnu  Satish Madhavi
2    Mona   Raju   Dhana
<class 'pandas.core.frame.DataFrame'>
Name
0    Sona
1  Chinnu
2    Mona
Name: Name, dtype: object
Process finished with exit code 0
```

Explanation:

In this code `read_csv()` method is used to read csv data into pandas dataframes. We can also display columns of data using `data.column_name`.

2) Using read_tables() method

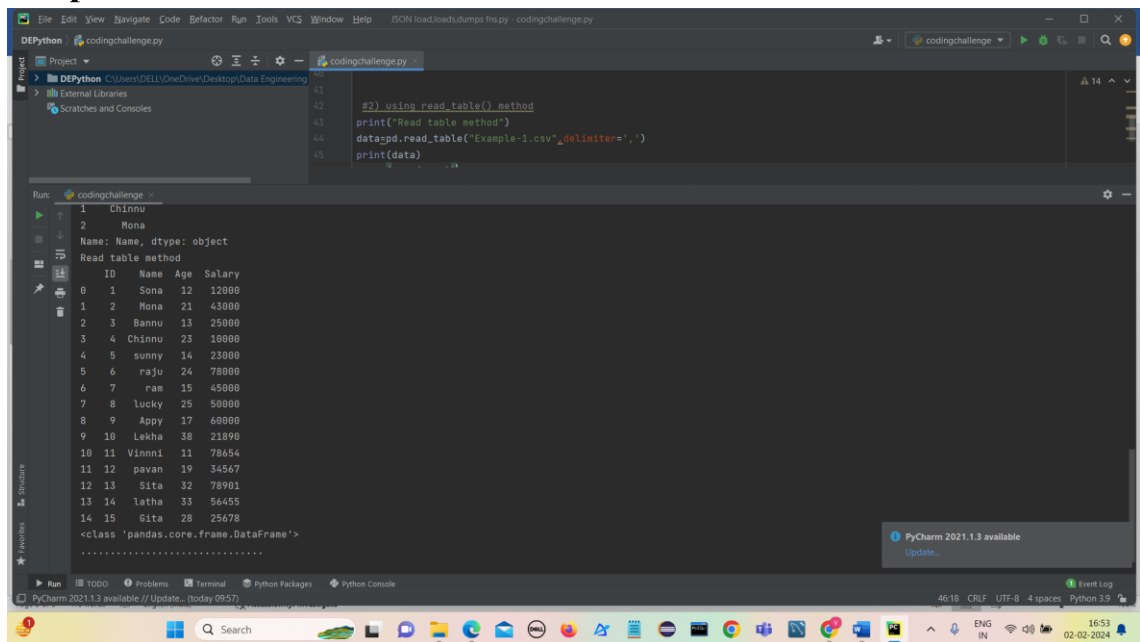


The screenshot shows the PyCharm IDE with a Python file named `codingchallenge.py`. The code uses `pd.read_table()` to load data from `Example-1.csv`. The output in the Run console shows a DataFrame with columns `ID`, `Name`, `Age`, and `Salary`.

```
codingchallenge.py
41
42 #2) using read-table() method
43 print("Read table method")
44 data=pd.read_table("Example-1.csv",delimiter=',')
45 print(data)
46 print(type(data))
47 print('.....')
48
49
50
51
52
53
54
```

```
Run: codingchallenge
Name: Name, dtype: object
Read table method
   ID  Name  Age  Salary
0   1  Sona   12   12000
1   2  Mona   21   43000
2   3  Bannu  13   25000
3   4  Chinnu  23   10000
4   5  sunny   14   23000
5   6  raju   24   78000
6   7   ram   15   45000
7   8  lucky   25   50000
8   9  Appy   17   60000
9  10  Lekha   38   21890
10 11  Vinnni  11   78654
```

Output:

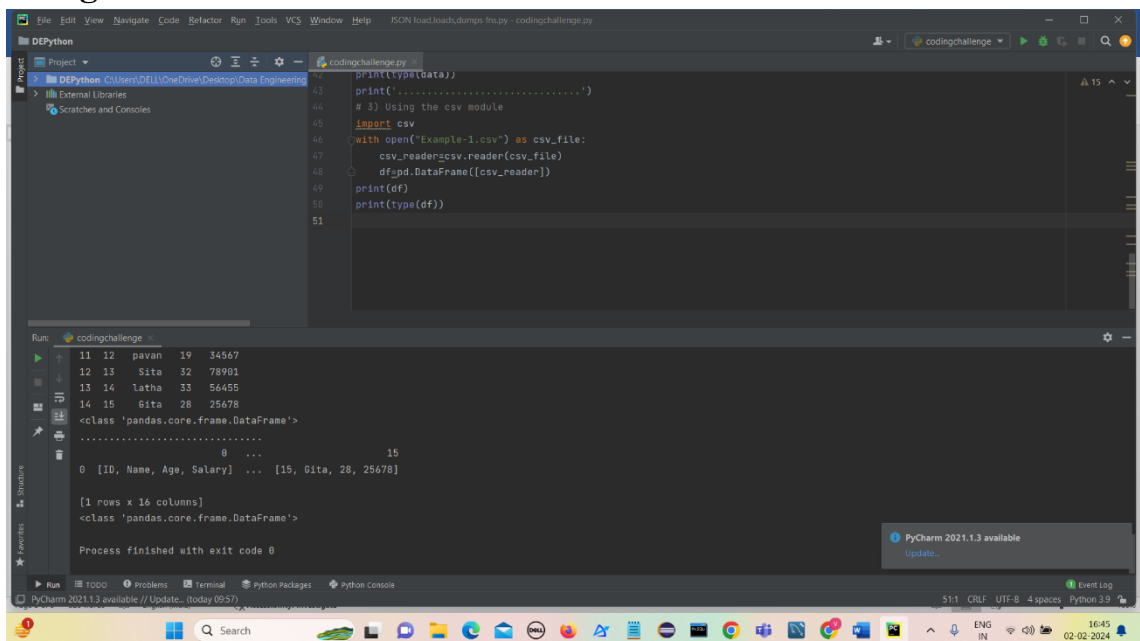


The screenshot shows the PyCharm IDE with a Python script named `codingchallenge.py`. The script uses the `read_table` method to read CSV data into a pandas DataFrame. The output window displays the following data:

```
1 Chinnu
2 Mona
Name: Name, dtype: object
Read table method
  ID  Name  Age  Salary
0   1   Sona  12   12000
1   2   Mona  21   43000
2   3   Bannu  13   25000
3   4   Chinnu  23   10000
4   5   sunny  14   23000
5   6   raju   24   78000
6   7   ram    15   45000
7   8   lucky  25   50000
8   9   Appy   17   60000
9  10   Lekha  38   21890
10  11  Vinnu  11   78654
11  12  pavan  19   34567
12  13   Sita  32   78901
13  14  latha  33   56455
14  15   Gita  28   25678
<class 'pandas.core.frame.DataFrame'>
```

In this code `read_table` method is used to read csv data into pandas dataframe and the delimiter used here is “,”.

3) Using csv module



The screenshot shows the PyCharm IDE with a Python script named `codingchallenge.py`. The script uses the `csv` module to read CSV data into a pandas DataFrame. The output window displays the following data:

```
11 12 pavan 19 34567
12 13 Sita 32 78901
13 14 latha 33 56455
14 15 Gita 28 25678
<class 'pandas.core.frame.DataFrame'>
.....
0 ... 15
0 [ID, Name, Age, Salary] ... [15, Gita, 28, 25678]

[1 rows x 16 columns]
<class 'pandas.core.frame.DataFrame'>
Process finished with exit code 0
```

In this code `csv` module is used to read csv data into pandas dataframes. First open the csv file, then using `csv.reader()` read the csv file and then convert that csv reader object into pandas dataframe object. Then print the data.

d) Filter Data in Pandas Dataframe using query.

It is used to filter dataframe

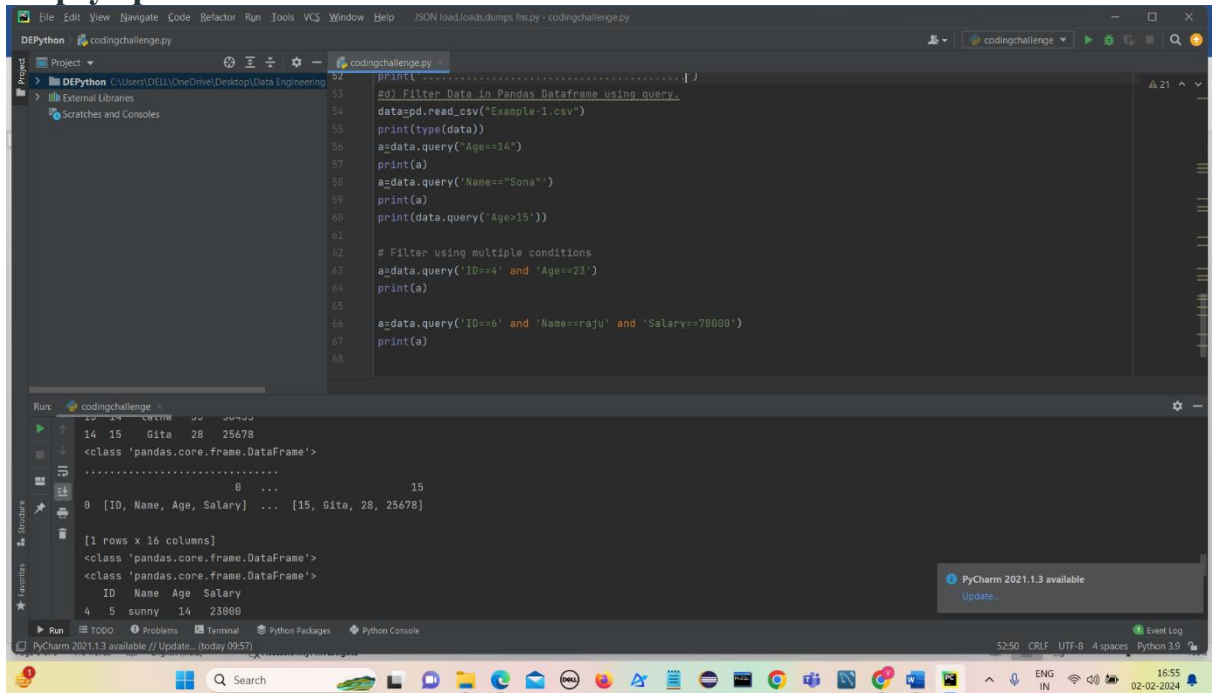
Syntax: DataFrame.query(expr, inplace=False, **kwargs)

Parameters:

- **expr:** Expression in string form to filter data.
- **inplace:** Make changes in the original data frame if True
- **kwargs:** Other keyword arguments.

Return type: Filtered Data frame

Dataframe.query() method only works if the column name doesn't have any empty spaces.



```
52 print('.....')
53 #d) Filter Data in Pandas Dataframe using query.
54 datapd.read_csv("Example-1.csv")
55 print(type(data))
56 a=data.query('Age==14')
57 print(a)
58 a=data.query('Name=="Sona"')
59 print(a)
60 print(data.query('Age>15'))
61
62 # Filter using multiple conditions
63 a=data.query('ID==4' and 'Age==25')
64 print(a)
65
66 a=data.query('ID==6' and 'Name=="raju"' and 'Salary==78000')
67 print(a)
68
```

Run: codingchallenge

```
14 15 Gita 28 25678
<class 'pandas.core.frame.DataFrame'>
.....
0 [ID, Name, Age, Salary] ... [15, Gita, 28, 25678]

[1 rows x 4 columns]
<class 'pandas.core.frame.DataFrame'>
<class 'pandas.core.frame.DataFrame'>
ID Name Age Salary
4 5 sunny 14 23000
```

Explanation:

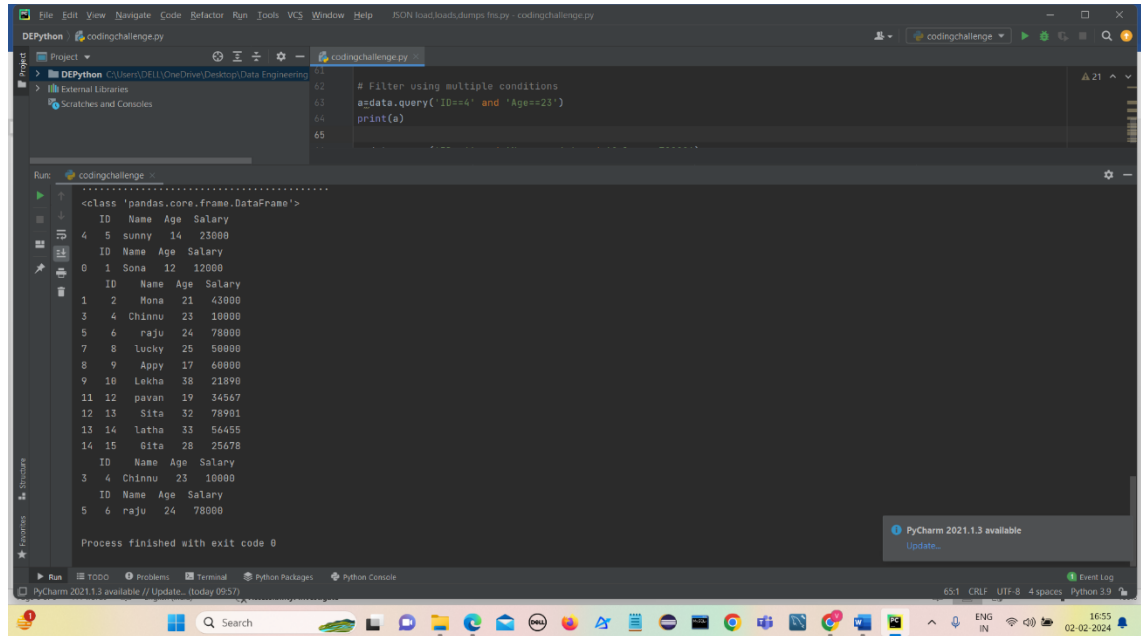
In the above code first read the csv file using **read_csv()** method.

Use the function **query()** to filter the records based on the condition.

- In the first query **data.query('Age==14')** the records whose age is is displayed.
- In the second query **data.query('Name=="Sona"')** the records with name Sona is displayed.
- In the third query **data.query('Age>15')** the records with age greater than 15 are displayed.

- iv) In the fourth query records with id 4 and age 23 are displayed
- v) In the last query records with id 6, Name Raju and Salary 78000 are displayed.

Output:



The screenshot shows a Python IDE with a file named `codingchallenge.py`. The code in the editor is as follows:

```
51  
52  
53 # Filter using multiple conditions  
54 a=data.query('ID==4' and 'Age==23')  
55 print(a)
```

The Run console displays the output of the code, showing a pandas DataFrame with columns `ID`, `Name`, `Age`, and `Salary`. The output is as follows:

```
<class 'pandas.core.frame.DataFrame'>  
ID Name Age Salary  
4 5 sunny 14 23000  
ID Name Age Salary  
0 1 Sona 12 12000  
ID Name Age Salary  
1 2 Mona 21 43000  
3 4 Chinnu 23 10000  
5 6 rajju 24 78000  
7 8 Lucky 25 50000  
8 9 Appy 17 60000  
9 10 Lekha 18 21000  
11 12 pavani 19 34567  
12 13 Sita 12 78901  
13 14 Latha 13 56455  
14 15 Gita 28 25678  
ID Name Age Salary  
3 4 Chinnu 23 10000  
ID Name Age Salary  
5 6 rajju 24 78000
```

The process finished with exit code 0. A notification for PyCharm 2021.1.3 is visible in the bottom right corner.