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

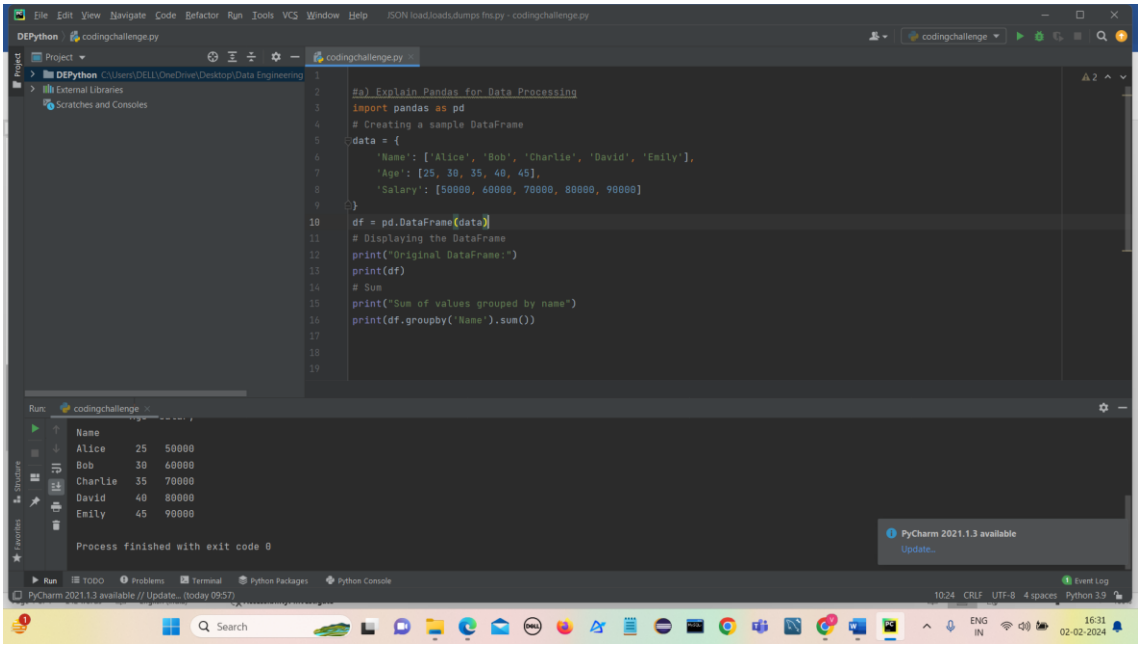
Pandas is an open-source Python library that provides high-performance, easy-to-use data structures, and data analysis tools for working with structured data. It is built on top of the NumPy library and is widely used for data manipulation, data cleaning, data exploration, and data analysis tasks in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

Data Structures:

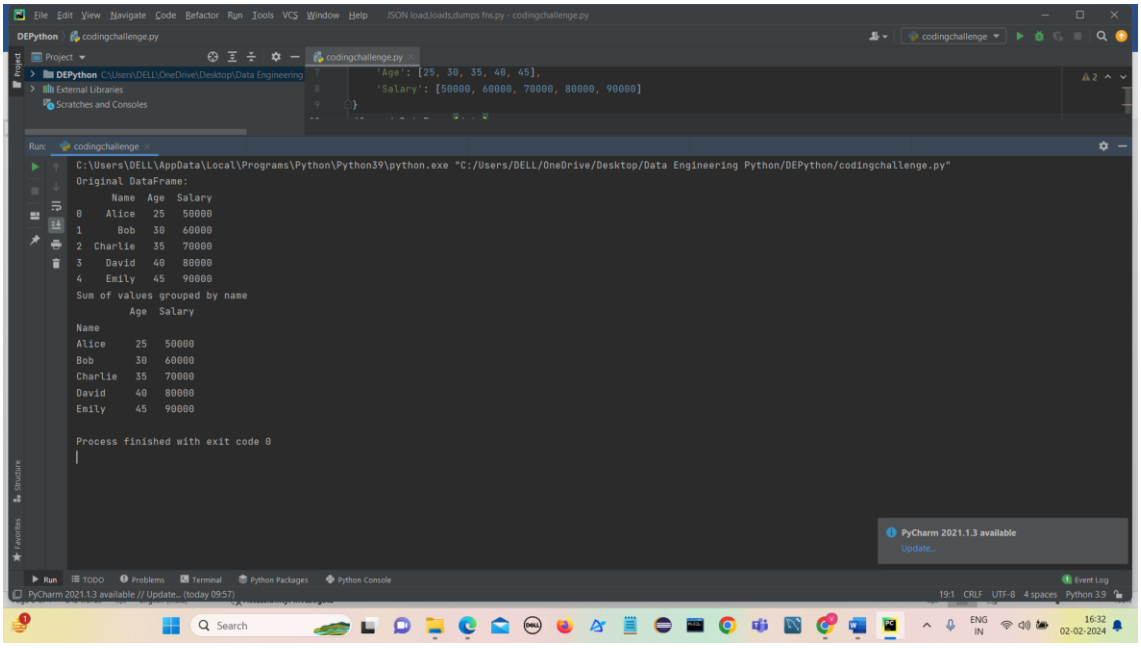
1) **Series:** It is a One-dimensional labeled array .It can hold any data type (e.g., integers, floats, strings, etc.). It is similar to a one-dimensional NumPy array but it has associated index with it
2) **DataFrame:** Two-dimensional labeled data structure with columns of different data types. It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



Output:

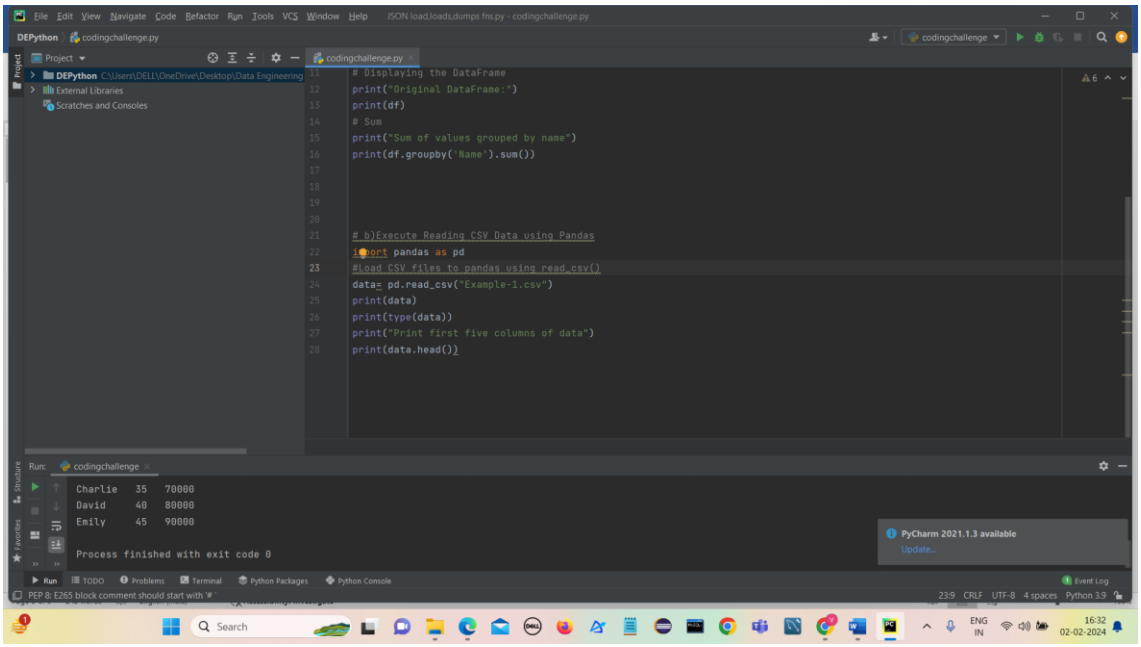


Explanation:

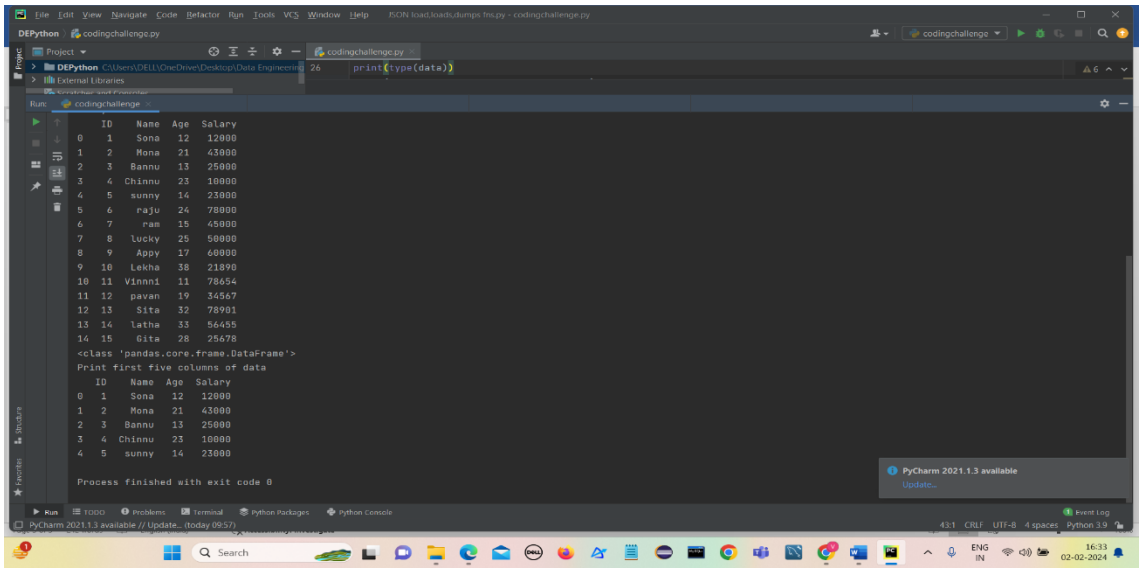
In this way data can be processed using pandas. The above code calculates the sum of all values grouped by Name.

b) Execute Reading CSV Data using Pandas

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



Output:



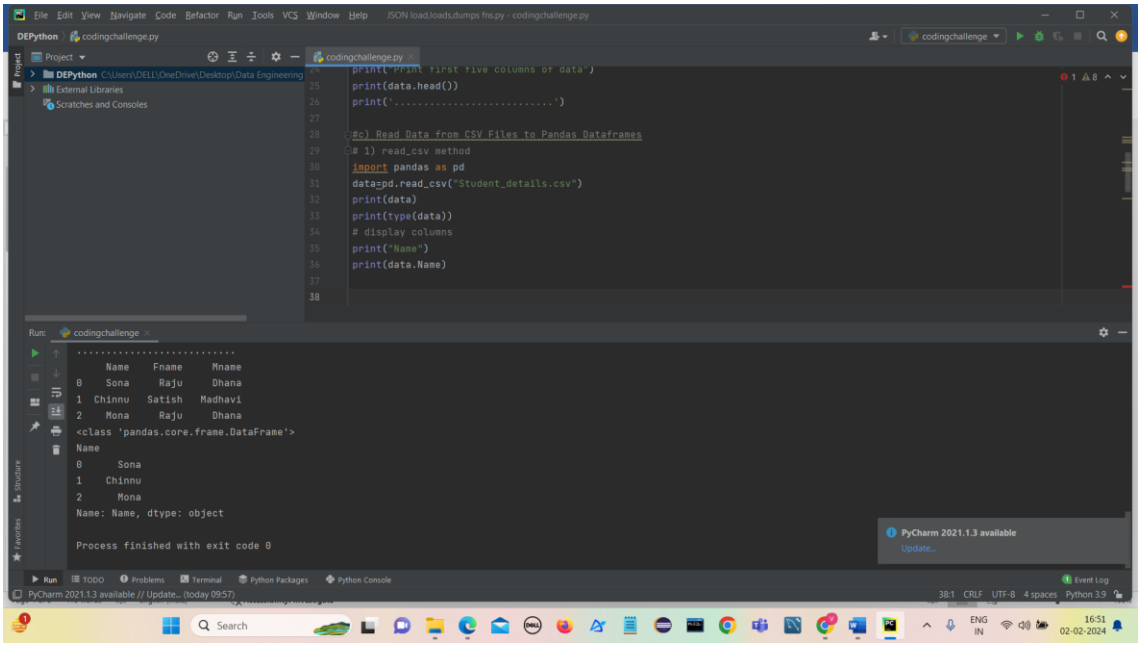
Explanation:

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

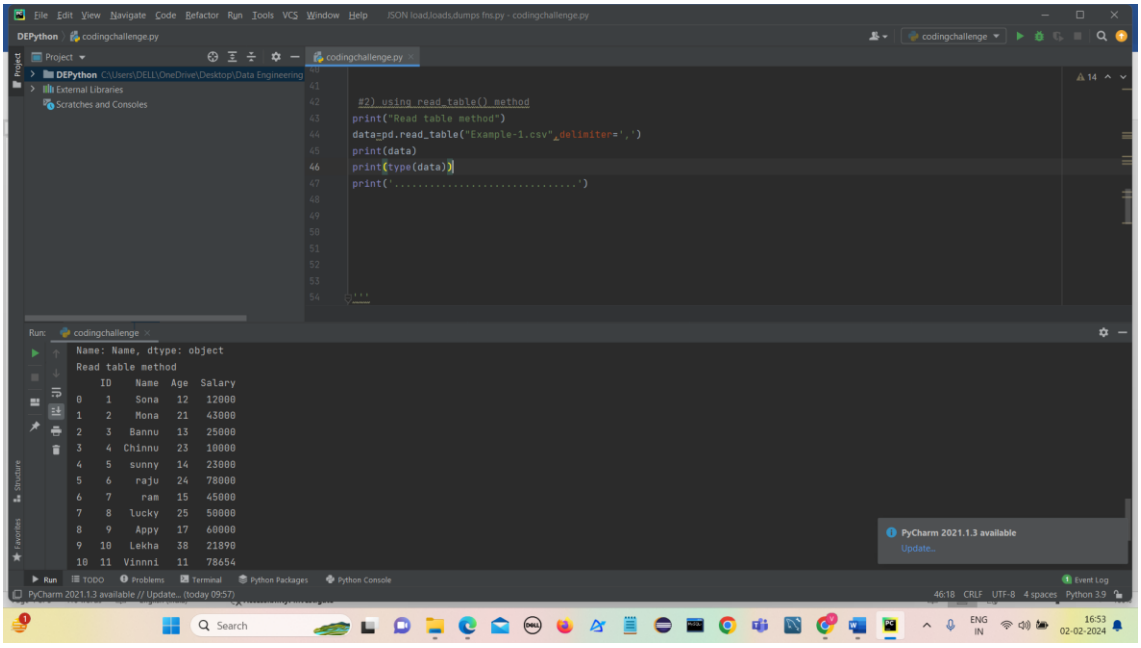
1) Using read_csv method



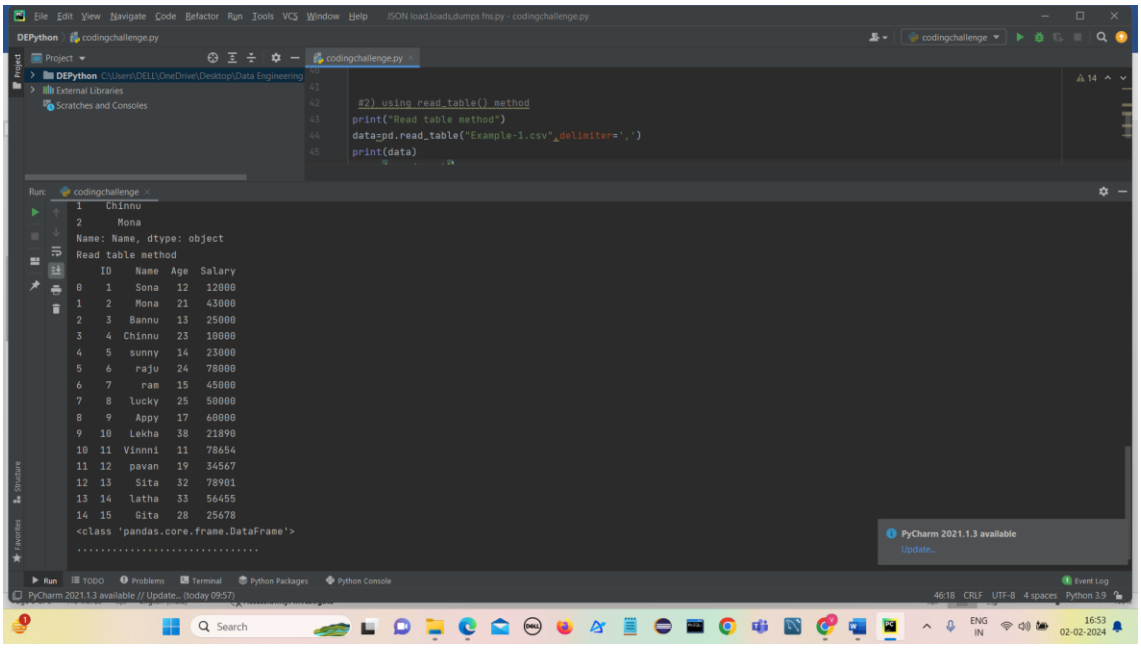
Explanation:

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

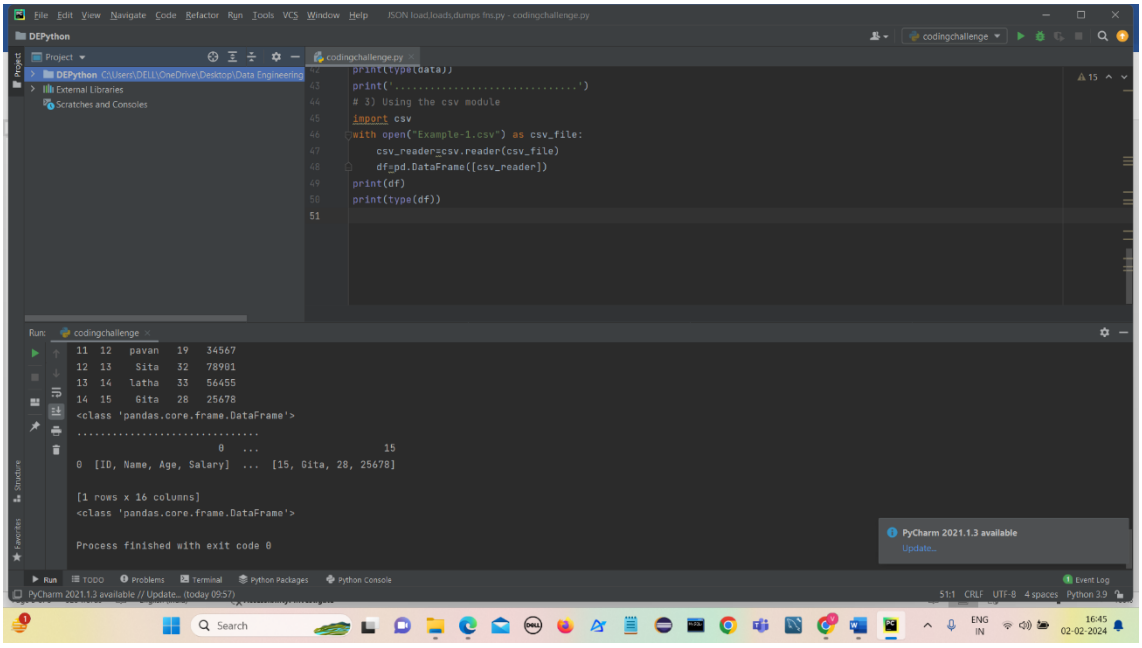


Output:



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



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

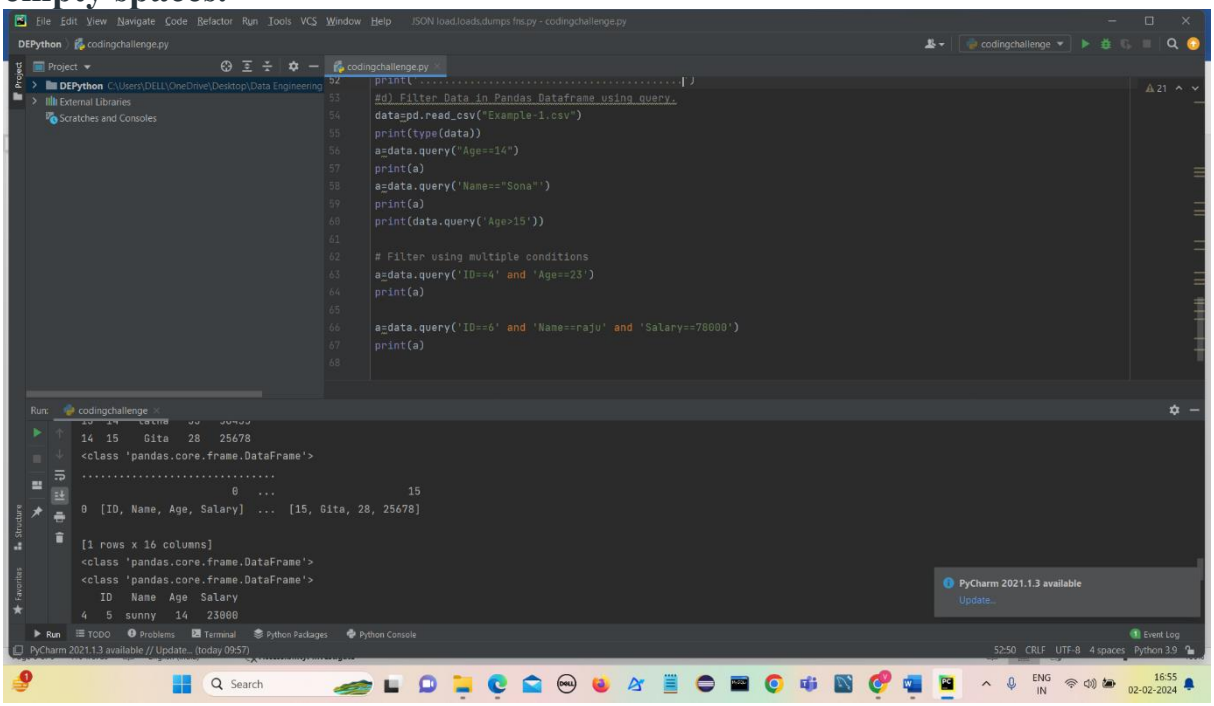
Syntax: DataFrame.query(expr, inplace=False, **kwargs)

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.**



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.

i) In the first query **data.query('Age==14')** the records whose age is is displayed.

ii) In the second query **data.query('Name==Sona')** the records with name Sona is displayed.

iii) 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 Salary78000 are displayed.

Output: