Name : Parth Nandedkar
Date : 02 Feb 2024
Topics : Python Coding Challenge Q1
Batch : Data Engineering Batch-1

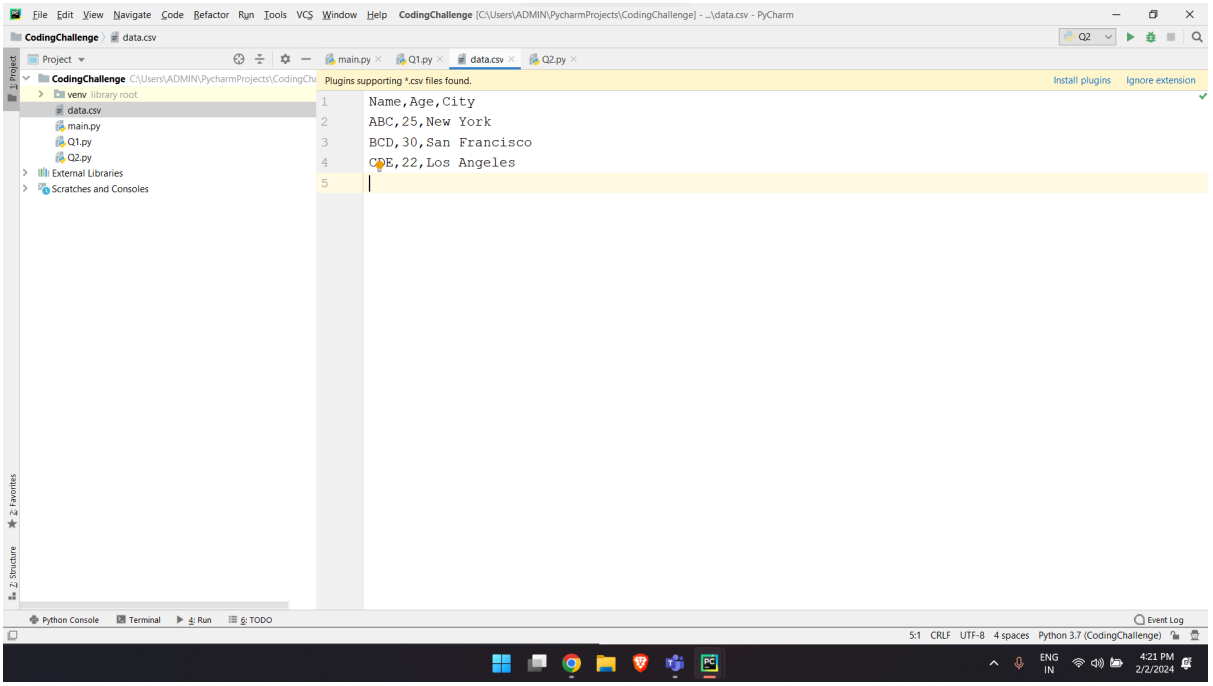
Q. *Explain Pandas for Data Processing & execute Reading CSV Data using Pandas & Read Data from CSV Files to Pandas Dataframes & Filter Data in Pandas Dataframe using query.*

Pandas simplifies the data processing workflow by providing a set of functions for common operations.The library allows for quick exploration, manipulation, and transformation of data, making it an essential tool in the data processing.

Pandas has defined classes which are useful to create objects such as dataframes so that we can work on data.

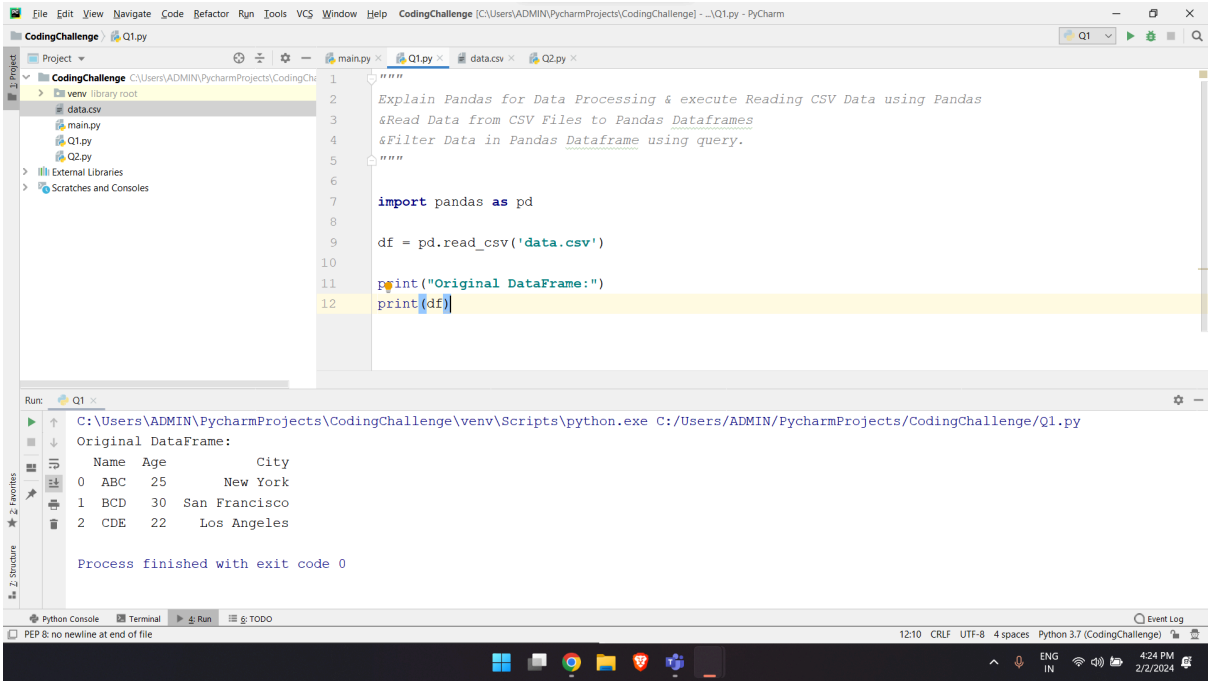
For Reading CSV files I have implemented following example,

CSV file which we created for demo named as ‘data.csv’ :



I have given the basic data Name,Age,City of 3 people.

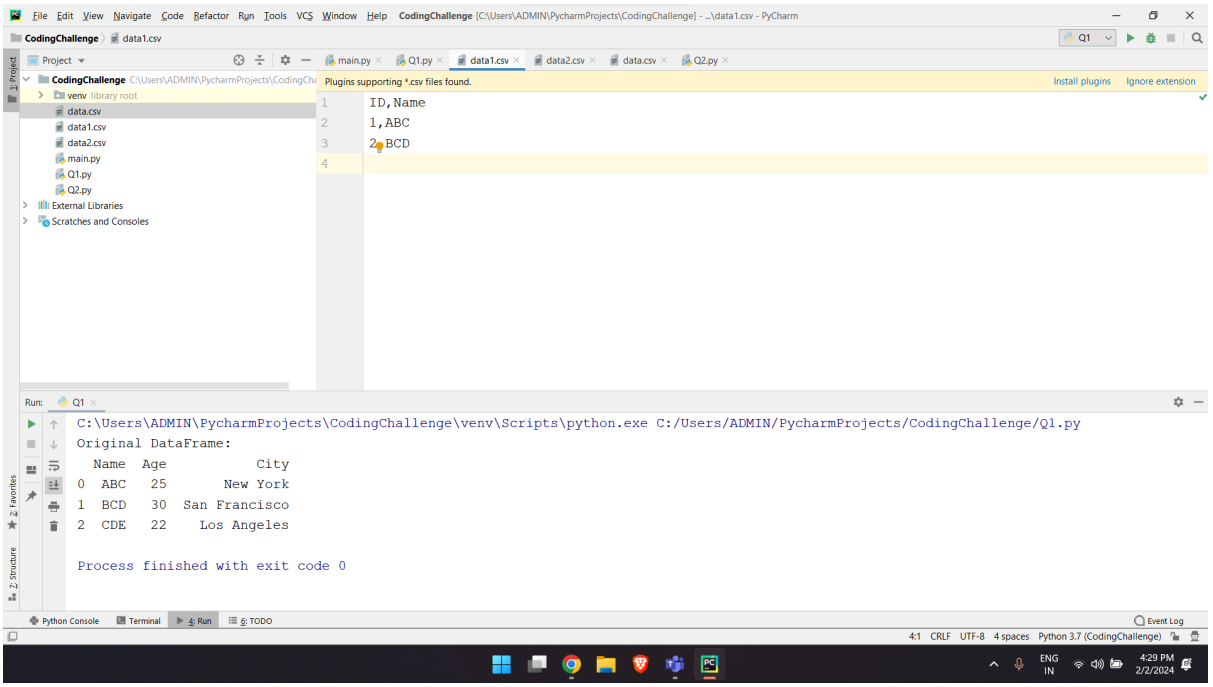
The following is screenshot of reading that data using Pandas.

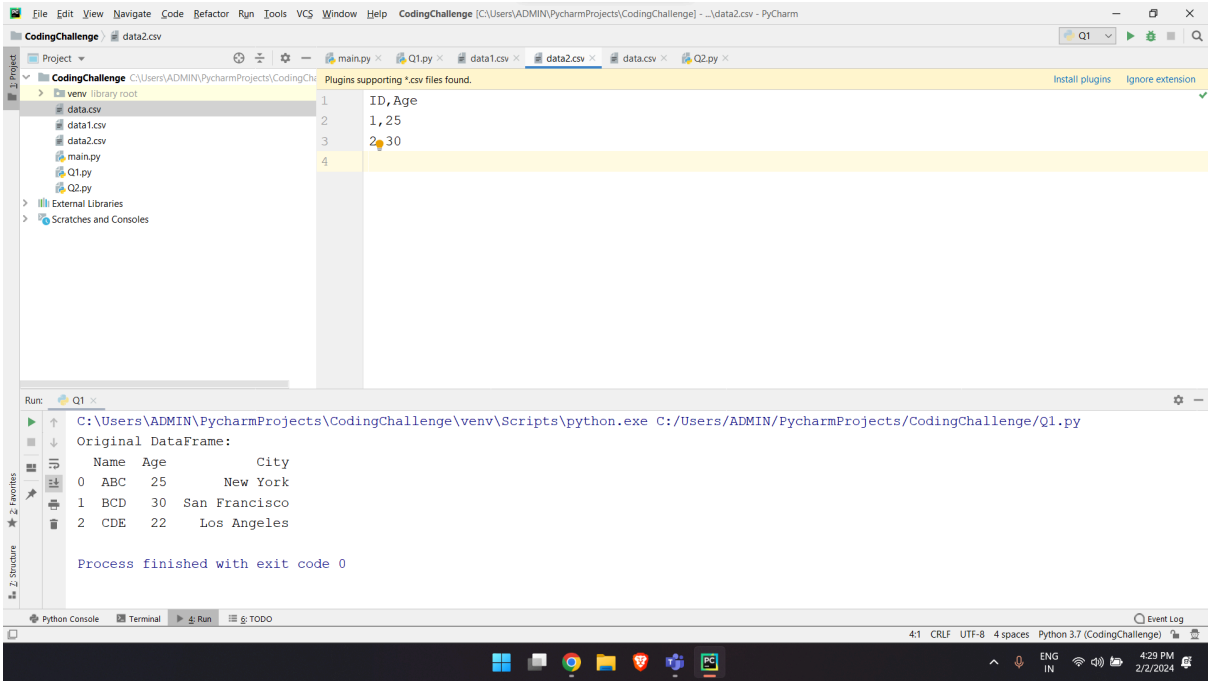


Here the **pd.read_csv()** method is used to read the file , **'data.csv'** is the name of the file.

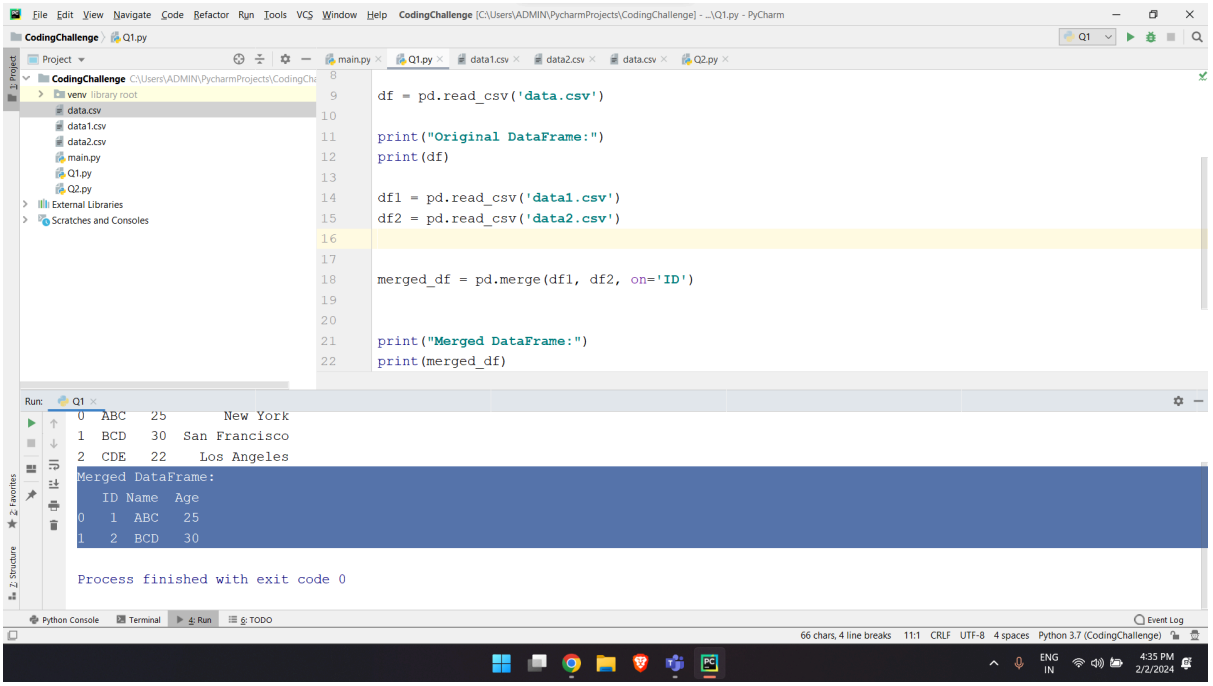
Filtering Data :

For filtering purpose I have taken 2 more CSV files **‘data1.csv’** and **‘data2.csv’**



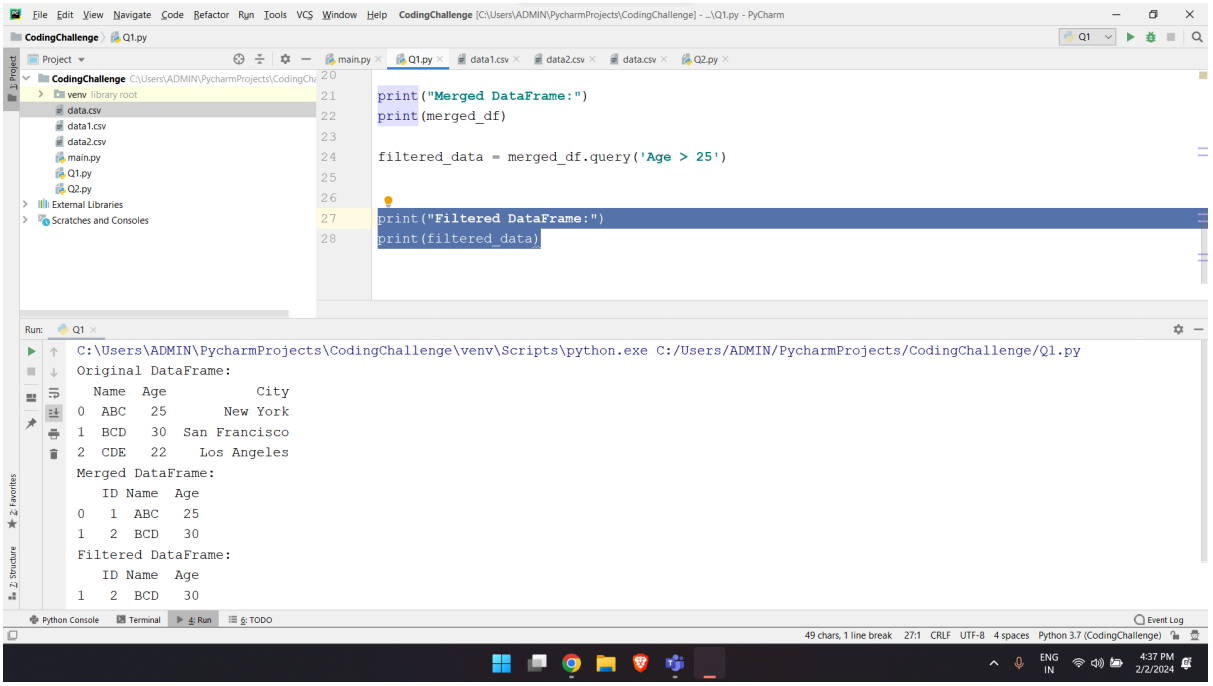


Now we can use query operations using **pd.merge()**
So we will connect these two CSV files to get connected output:



On parameter ON = 'ID' we have joined the CSV files so that we can get filtered data using it.

For filtering I have given a condition as **age > 25** so that we can get filtered output :



merged_df.query('Age >= 25'): Uses the query method to filter the DataFrame based on the condition **'Age >= 25'.**

print(filtered_data): Displays the filtered DataFrame.

Summery :

We have used CSV files to create dataframes in Pandas.dataframe object and we used methods like merge to query the data and merged_df.query() to provide the condition for filtering the data.