

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

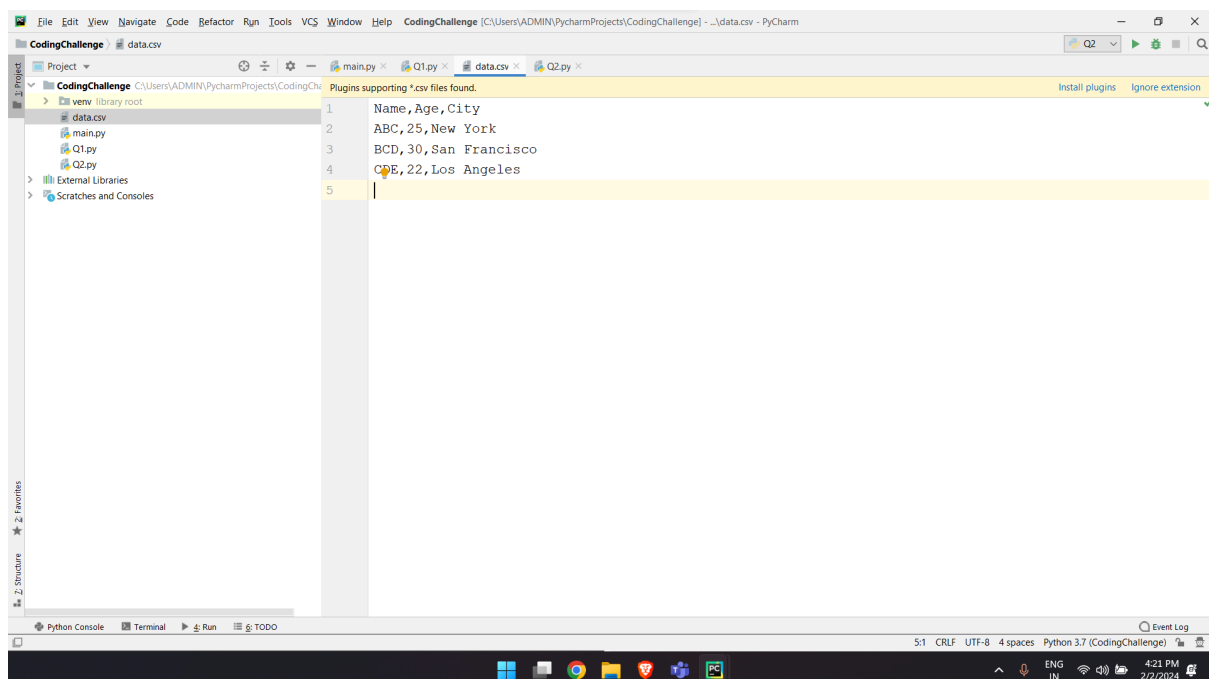
Q. Explain Pandas for Data Processing & execute Reading CSV Data using Pandas & Read Data from CSV Files to Pandas Dataframes & Filter Data in Pandas Dataframe using query.

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.

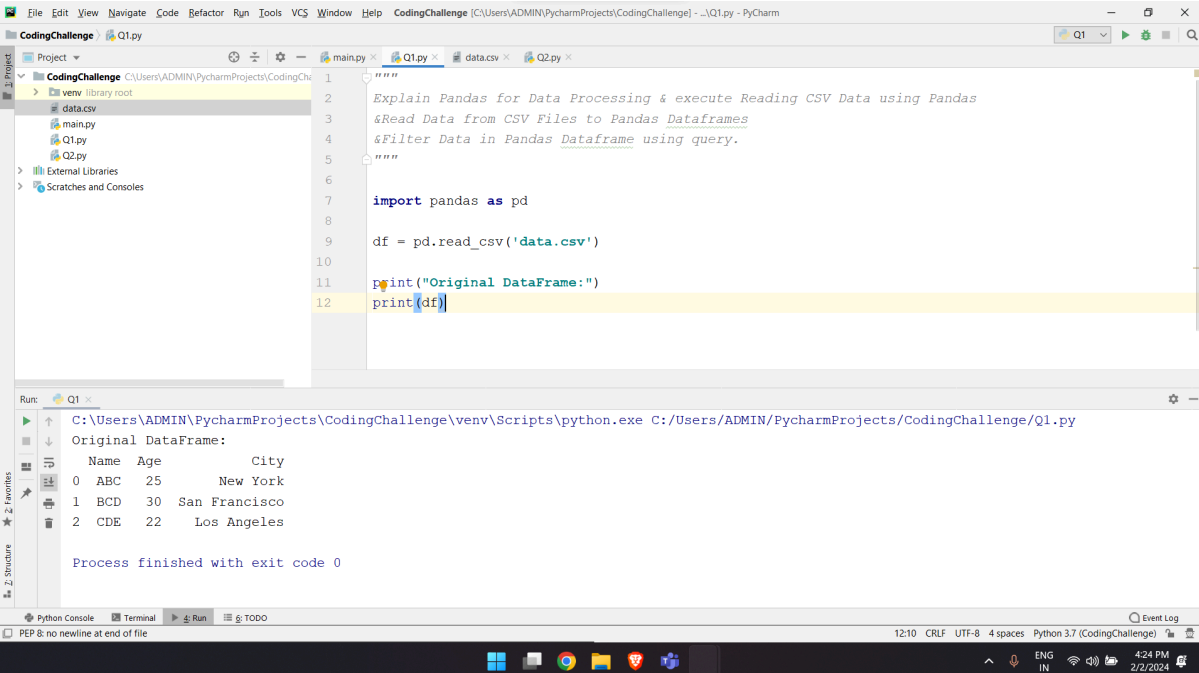
For Reading CSV files I have implemented following example,

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.



The screenshot shows the PyCharm IDE interface. The main editor window displays a Python script in `Q1.py` that reads a CSV file named `data.csv` using `pd.read_csv()` and prints the resulting DataFrame. The script includes a docstring with instructions: "Explain Pandas for Data Processing & execute Reading CSV Data using Pandas", "Read Data from CSV Files to Pandas Dataframes", and "Filter Data in Pandas Dataframe using query.".

```
1 """
2 Explain Pandas for Data Processing & execute Reading CSV Data using Pandas
3 Read Data from CSV Files to Pandas Dataframes
4 Filter Data in Pandas Dataframe using query.
5 """
6
7 import pandas as pd
8
9 df = pd.read_csv('data.csv')
10
11 print("Original DataFrame:")
12 print(df)
```

The Run console at the bottom shows the output of the script. It displays the command used to run the file, the printed DataFrame, and a confirmation that the process finished with exit code 0.

```
Run: C:\Users\ADMIN\PycharmProjects\CodingChallenge\venv\Scripts\python.exe C:/Users/ADMIN/PycharmProjects/CodingChallenge/Q1.py
Original DataFrame:
   Name  Age  City
0  ABC   25  New York
1  BCD   30  San Francisco
2  CDE   22  Los Angeles

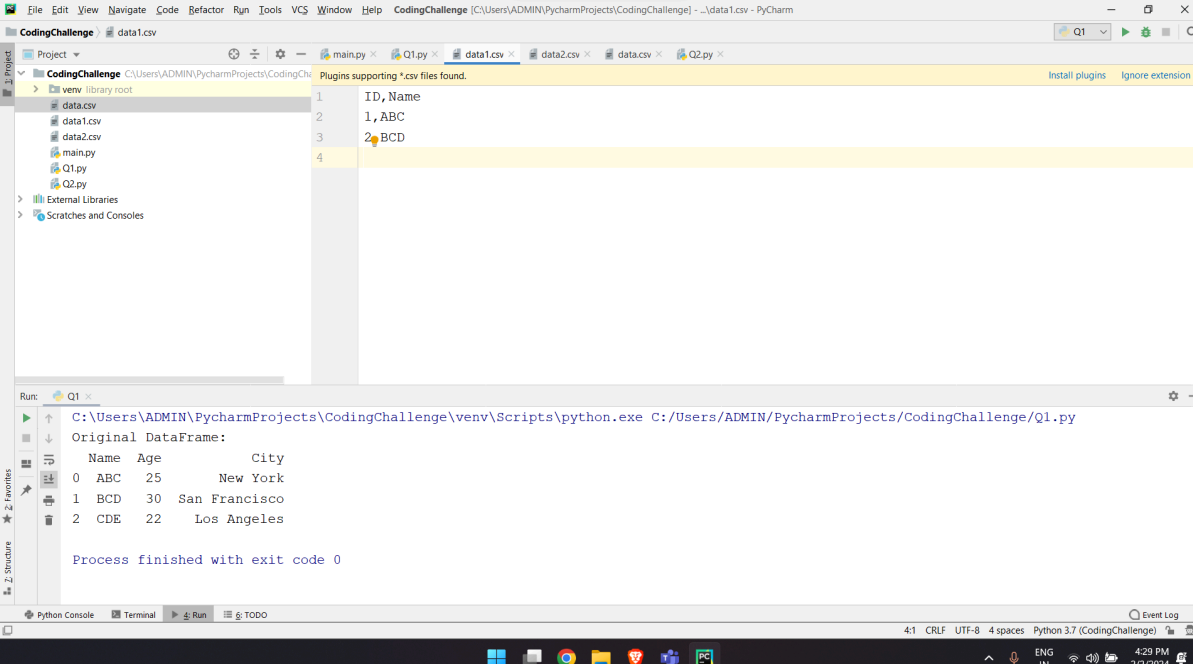
Process finished with exit code 0
```

The status bar at the bottom indicates the file encoding is UTF-8, the editor uses 4 spaces for indentation, and the Python version is 3.7.

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'



```
1 ID, Name
2 1, ABC
3 2, BCD
4
```

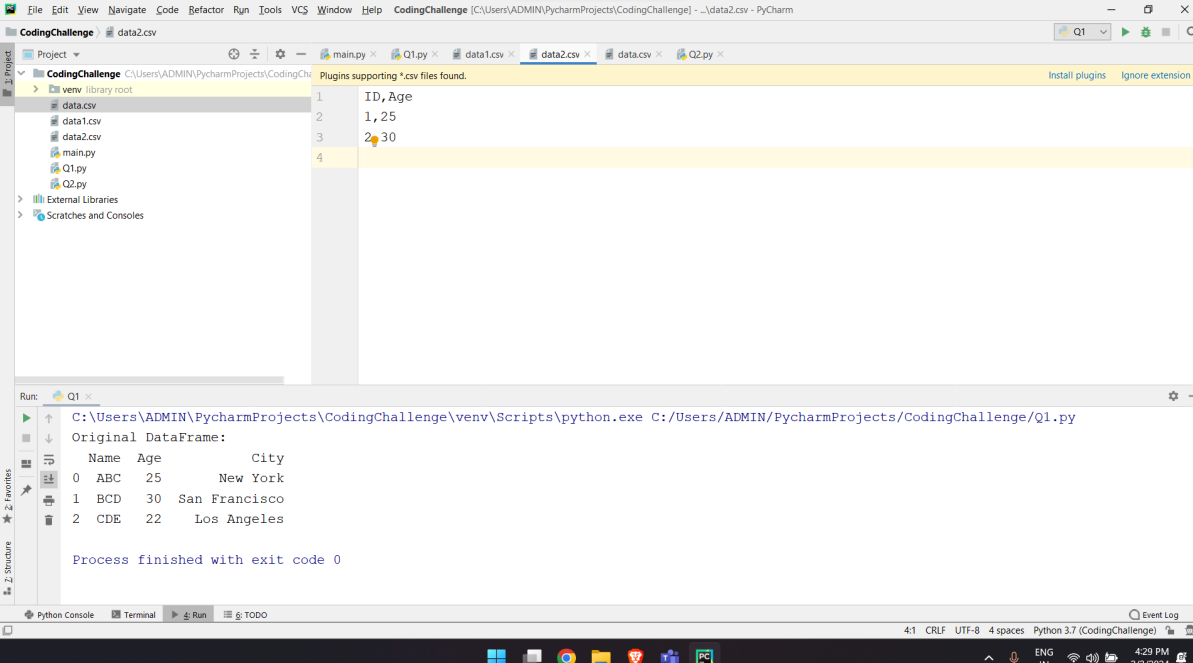
Run: Q1

```
C:\Users\ADMIN\PycharmProjects\CodingChallenge\venv\Scripts\python.exe C:/Users/ADMIN/PycharmProjects/CodingChallenge/Q1.py
```

Original DataFrame:

	Name	Age	City
0	ABC	25	New York
1	BCD	30	San Francisco
2	CDE	22	Los Angeles

Process finished with exit code 0



```
1 ID, Age
2 1, 25
3 2, 30
4
```

Run: Q1

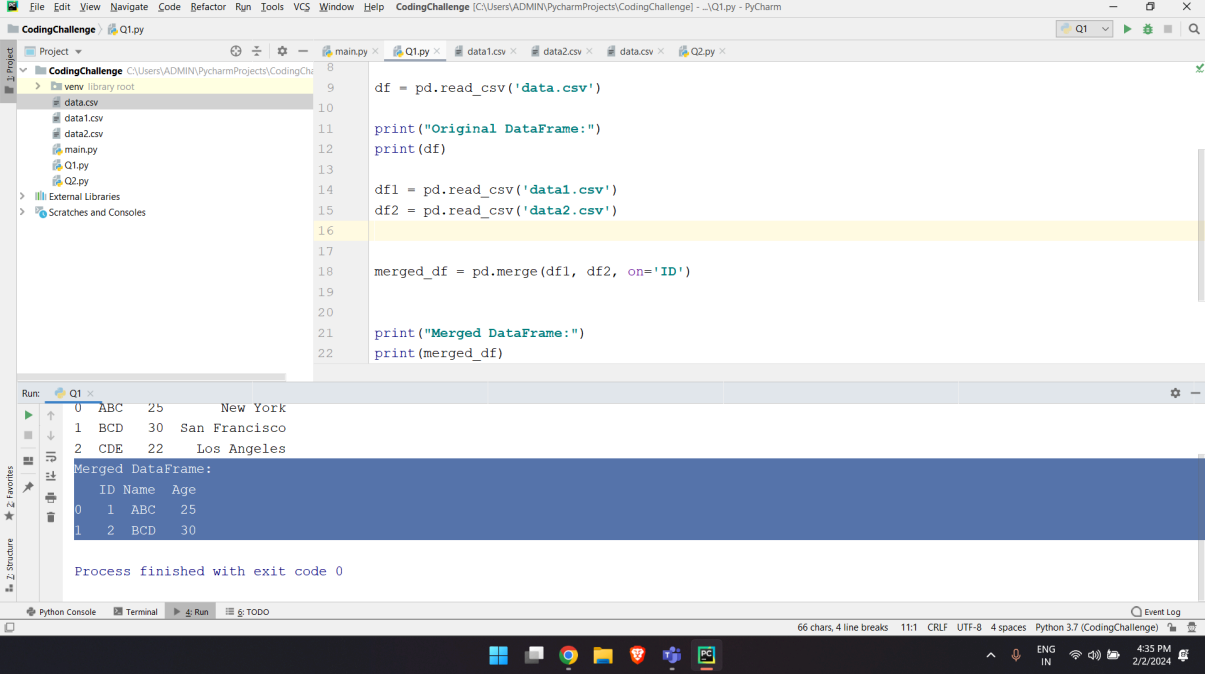
```
C:\Users\ADMIN\PycharmProjects\CodingChallenge\venv\Scripts\python.exe C:/Users/ADMIN/PycharmProjects/CodingChallenge/Q1.py
```

Original DataFrame:

	Name	Age	City
0	ABC	25	New York
1	BCD	30	San Francisco
2	CDE	22	Los Angeles

Process finished with exit code 0

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



```
File Edit View Navigate Code Refactor Run Tools VCS Window Help CodingChallenge [C:\Users\ADMIN\PycharmProjects\CodingChallenge] - Q1.py - PyCharm
Project: CodingChallenge
venv library root
data.csv
data1.csv
data2.csv
main.py
Q1.py
Q2.py
External Libraries
Scratches and Consoles

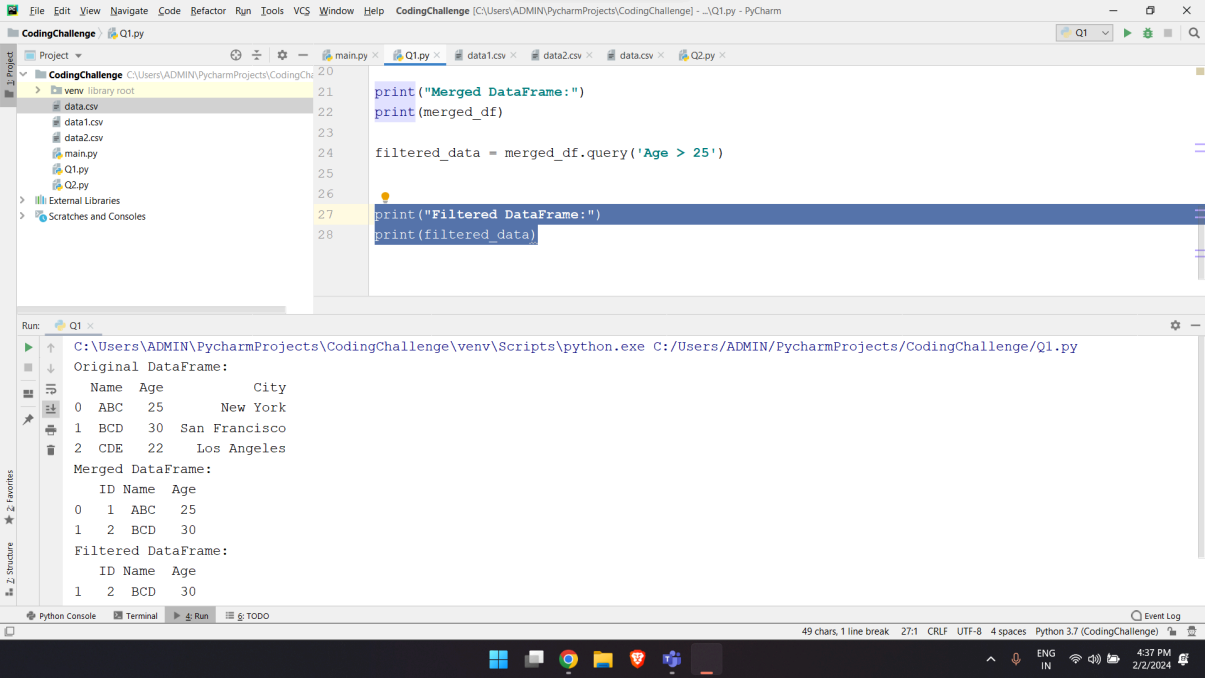
main.py
8 df = pd.read_csv('data.csv')
9
10 print("Original DataFrame:")
11 print(df)
12
13 df1 = pd.read_csv('data1.csv')
14 df2 = pd.read_csv('data2.csv')
15
16
17 merged_df = pd.merge(df1, df2, on='ID')
18
19
20 print("Merged DataFrame:")
21 print(merged_df)
22

Run: Q1
0 ABC 25 New York
1 BCD 30 San Francisco
2 CDE 22 Los Angeles
Merged DataFrame:
ID Name Age
0 1 ABC 25
1 2 BCD 30
Process finished with exit code 0

Python Console Terminal Run TODO 66 chars, 4 line breaks 11:1 CRLF UTF-8 4 spaces Python 3.7 (CodingChallenge) 4:35 PM 2/2/2024
```

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 :



```
File Edit View Navigate Code Refactor Run Tools VCS Window Help CodingChallenge [C:\Users\ADMIN\PycharmProjects\CodingChallenge] - Q1.py - PyCharm
Project: CodingChallenge
venv library root
data.csv
data1.csv
data2.csv
main.py
Q1.py
Q2.py
External Libraries
Scratches and Consoles

main.py
20 print("Merged DataFrame:")
21 print(merged_df)
22
23 filtered_data = merged_df.query('Age > 25')
24
25
26 print("Filtered DataFrame:")
27 print(filtered_data)
28

Run: Q1
C:\Users\ADMIN\PycharmProjects\CodingChallenge\venv\Scripts\python.exe C:/Users/ADMIN/PycharmProjects/CodingChallenge/Q1.py
Original DataFrame:
Name Age City
0 ABC 25 New York
1 BCD 30 San Francisco
2 CDE 22 Los Angeles
Merged DataFrame:
ID Name Age
0 1 ABC 25
1 2 BCD 30
Filtered DataFrame:
ID Name Age
1 2 BCD 30

Python Console Terminal Run TODO 49 chars, 1 line break 27:1 CRLF UTF-8 4 spaces Python 3.7 (CodingChallenge) 4:37 PM 2/2/2024
```

`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.