**Coding challenge**

**Python**

**Name: Rohan Vinayak Chaudhari**

**Batch: Data Engineering 1**

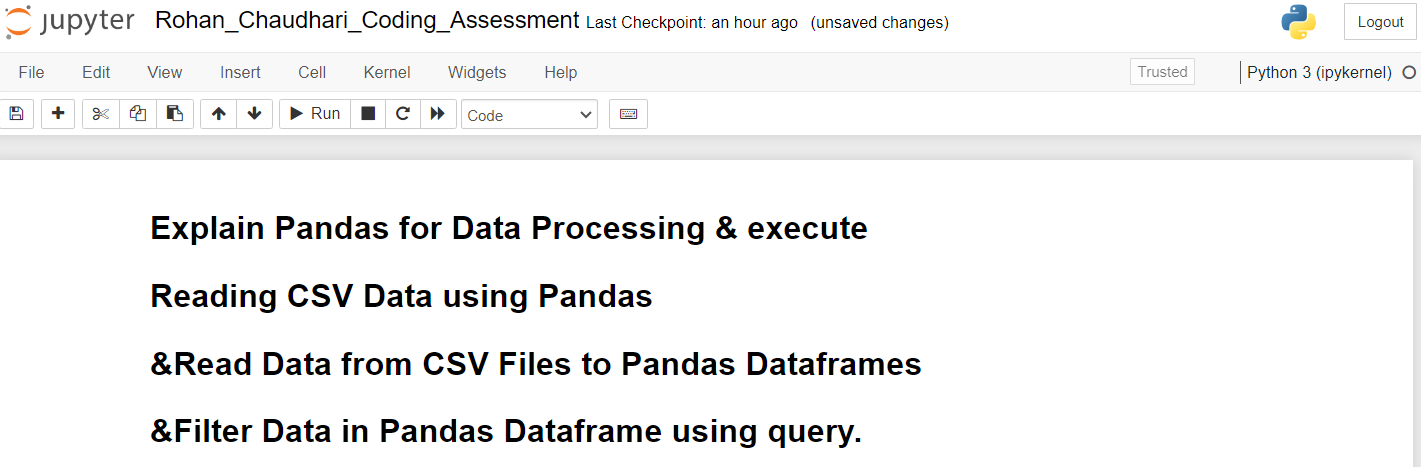
**Question1:**Explain Pandas for Data Processing & execute Reading CSV Data using Pandas  
 &Read Data from CSV Files to Pandas Dataframes  
 &Filter Data in Pandas Dataframe using query.

1)Pandas for data processing:

**Pandas** is an open-source data manipulation and analysis library for Python. It provides data structures such as Series and DataFrame for efficiently manipulating large datasets.Some Points include:

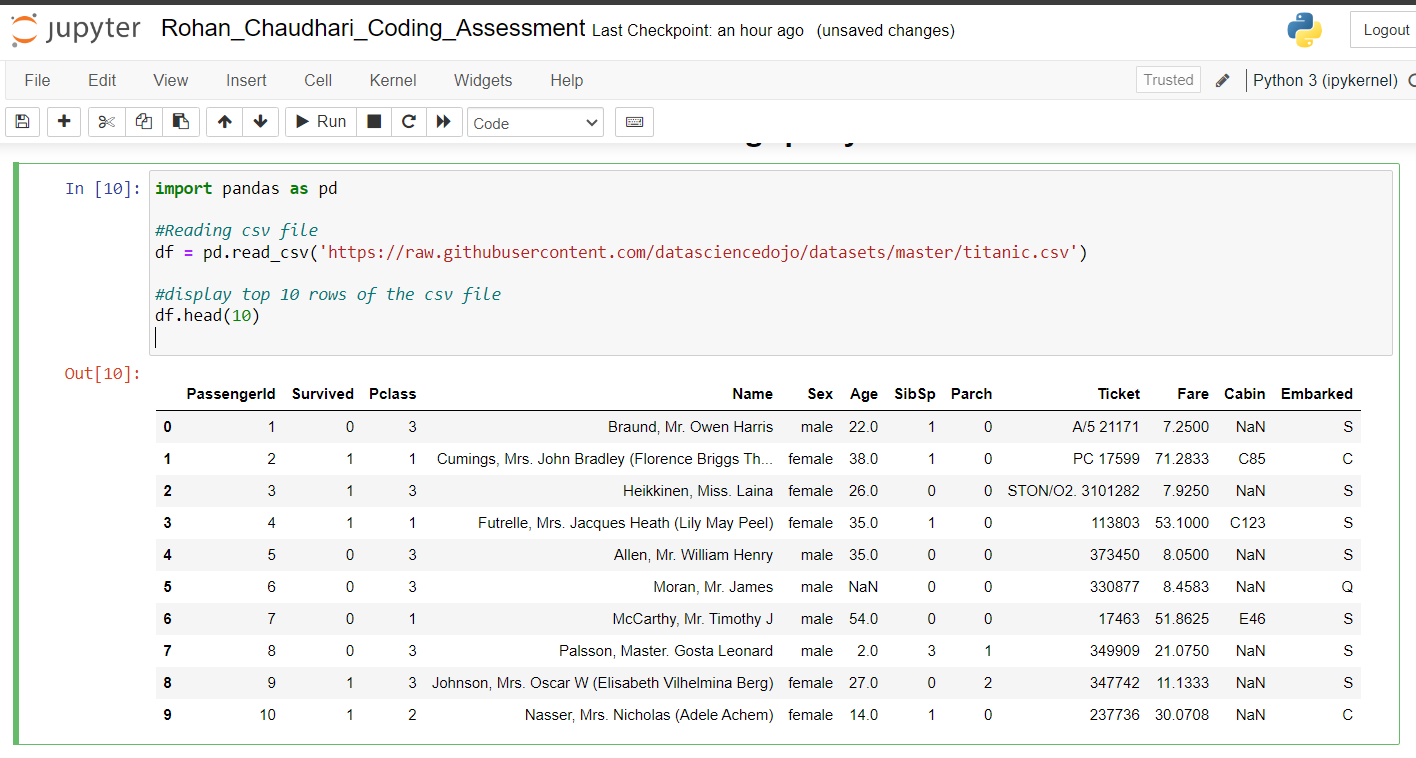
1. **DataFrame:** A two-dimensional table with rows and columns.Similar to sql tables.
2. **Series:** A one-dimensional labeled array containing of holding any data type.
3. **Data Cleaning:** Pandas provides functions for handling missing data and cleaning the dataset.
4. **Data Filtering and Selection:** Easy ways to filter, select, and manipulate data.
5. **Merging and Joining:** Combine different datasets using various methods.
6. **Aggregation:** Perform operations on data grouped by certain criteria.

**Question1:**

****

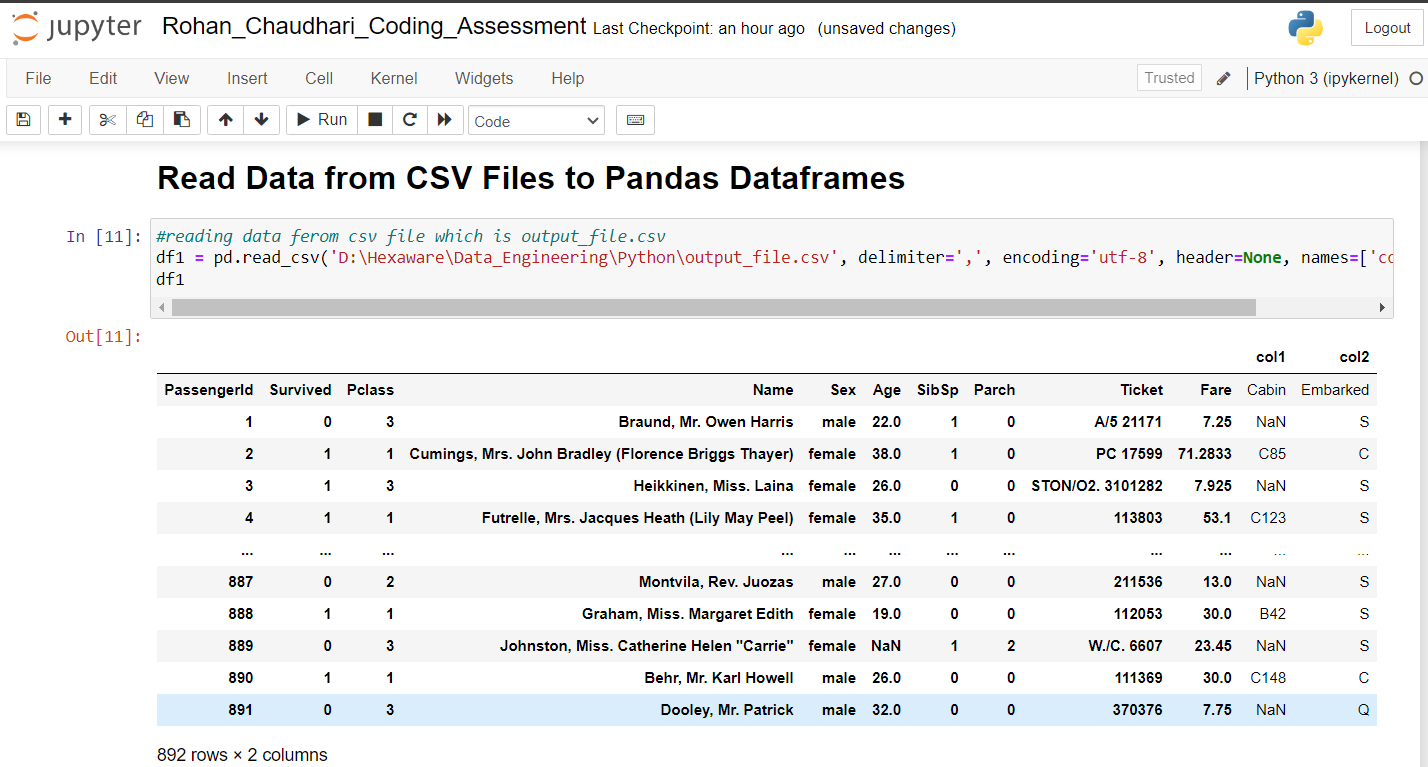
**2) Reading CSV Data using Pandas:**

Here I have taken a csv data name titanic where read the data using read\_csv and display the top 10 data from the csv file.



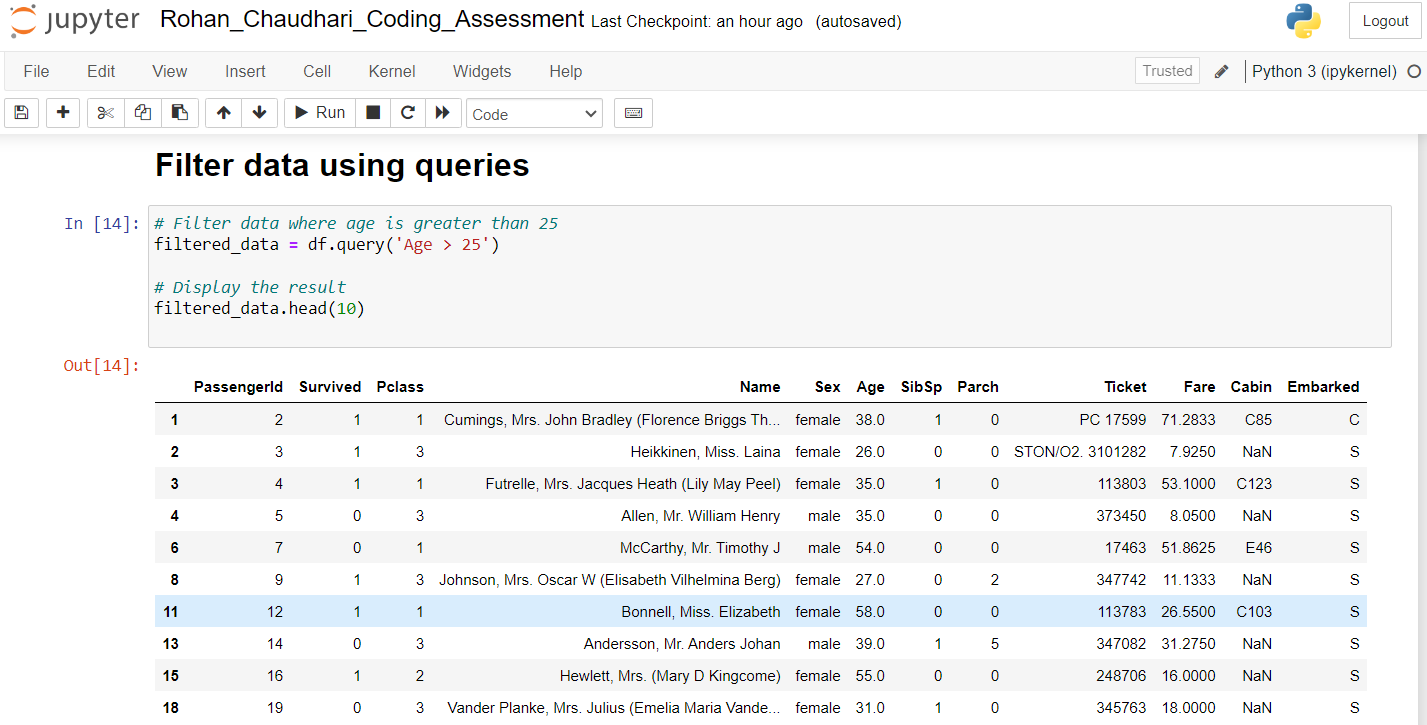
**3)Reading CSV data from file using pandas**

Here used the same dataset and read the data into dataframe from a already existing file.



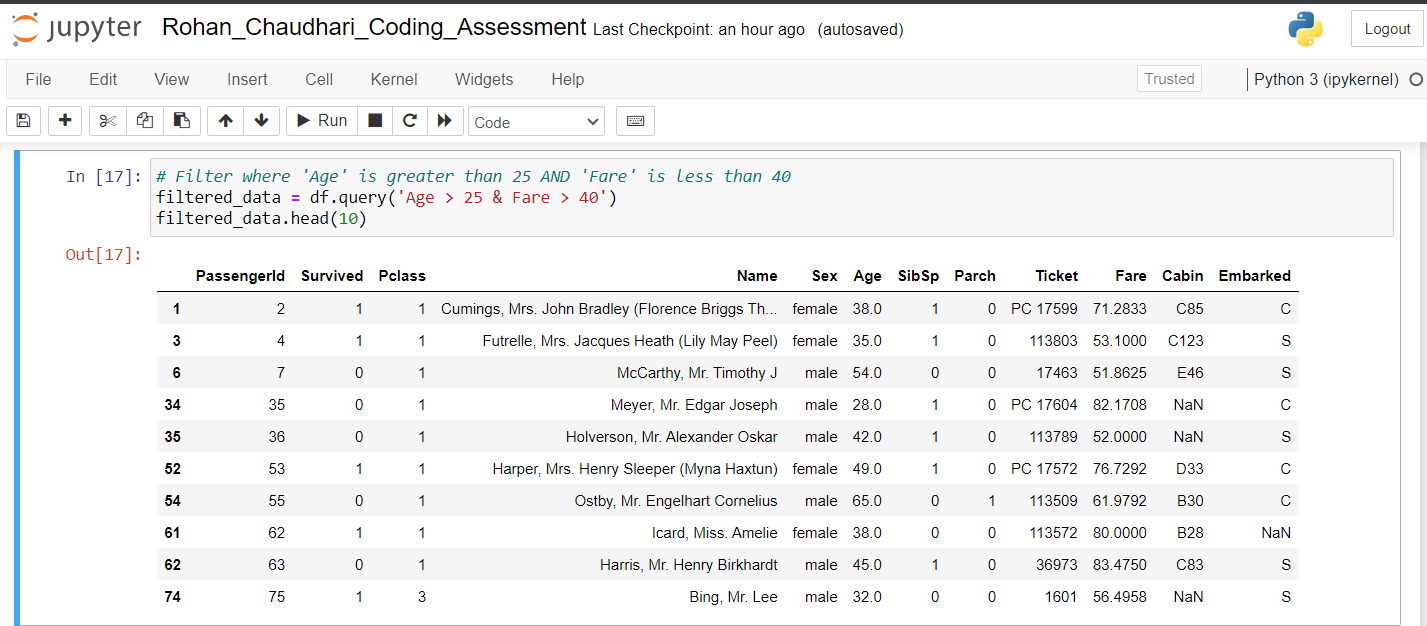
**4)Filter data using query**

Numeric Comparison:



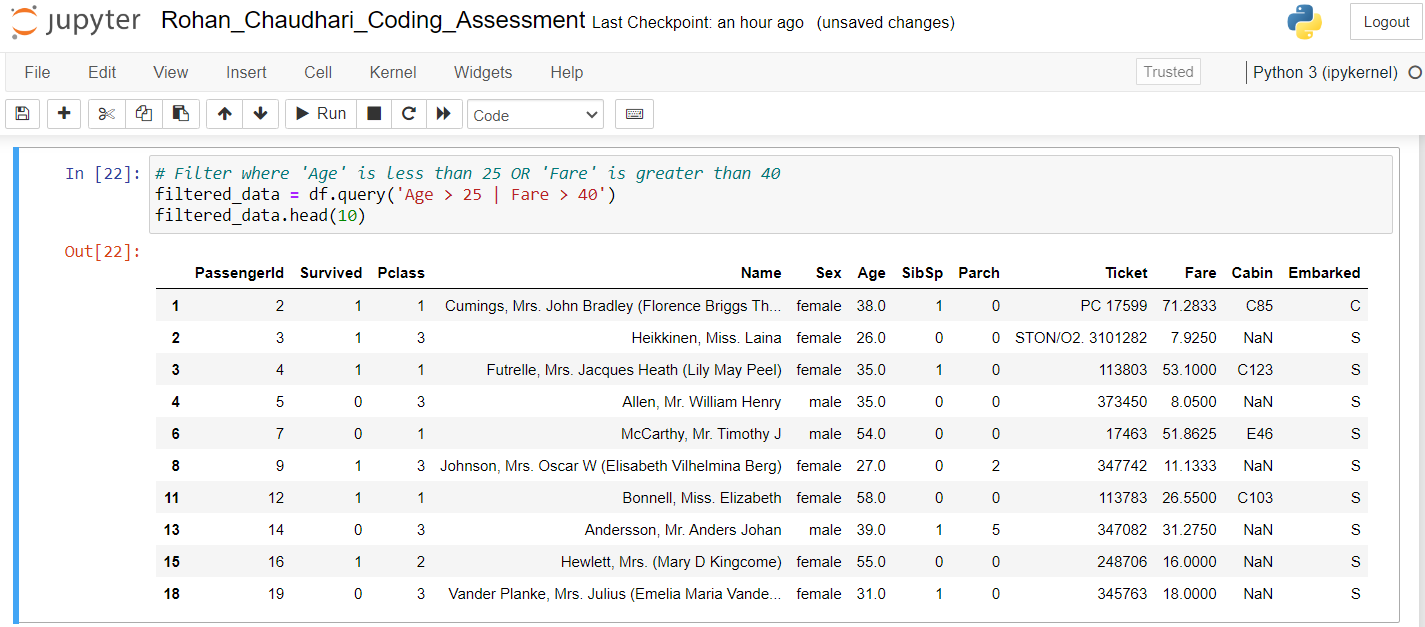
Logical And:

Extracted the data where age is greater than 25 and fare is grater than 40



Logical OR:

Extracted the data where age is greater than 25 or fare is greater than 25.



String Filtering:

Extracted the data where the name is similar to ‘McCarthy, Mr. Timothy J’

