Python Coding Challenge

Q.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.

Ans.

Pandas for Data Processing:

Pandas Overview:

Pandas is an open-source data manipulation and analysis library for Python. It provides data structures like Series and DataFrame, which are powerful tools for working with structured data. Pandas is widely used in data preprocessing, cleaning, exploration, and analysis.

Key Components:

- **DataFrame**: A two-dimensional table with labeled axes (rows and columns).
- **Series**: A one-dimensional labeled array capable of holding any data type.
- **Index**: Labels for rows and columns in DataFrames.

Reading CSV Data using Pandas:

Reading CSV Files:

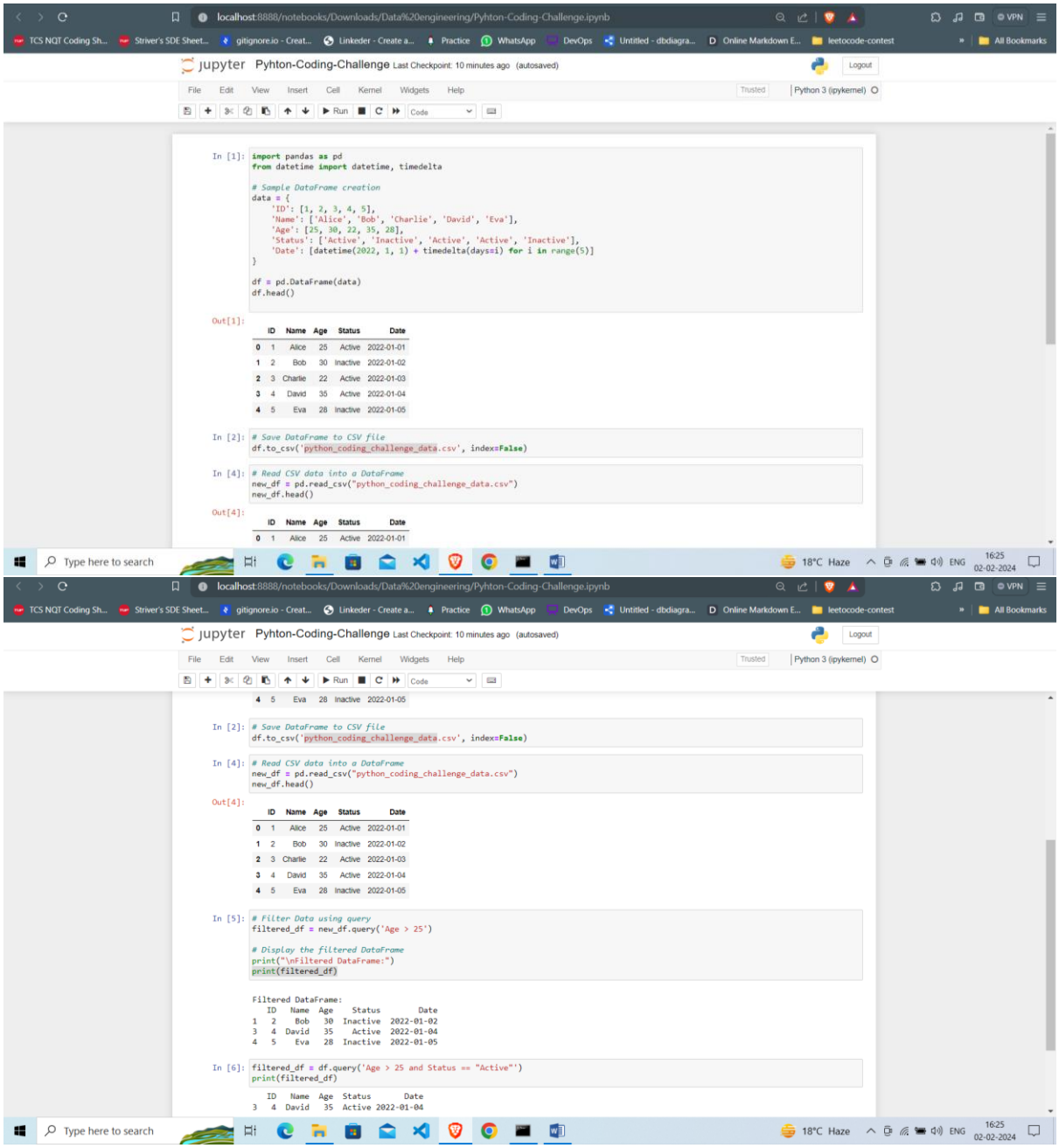
Pandas provides the read_csv() function to read data from CSV files into a DataFrame.

```
import pandas as pd

# Read CSV data into a DataFrame

df = pd.read_csv('your_file.csv')
```

Read Data from CSV Files to Pandas Dataframes:

Creating DataFrame from CSV:

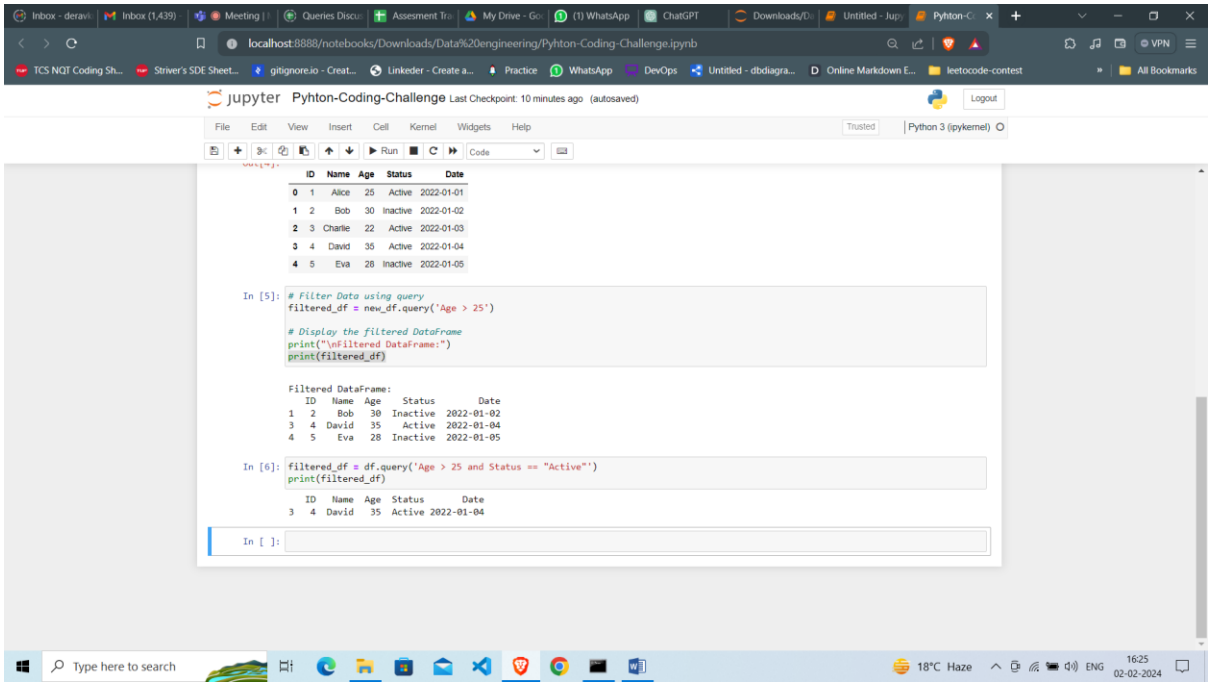
The read_csv() function reads CSV data and creates a DataFrame, allowing us to perform various operations on the data.



Filter Data in Pandas DataFrame using query:

Filtering Data with query:

Pandas provides the query() method to filter data based on a query expression.



In the above example, we filtered the dataframe based on attribute Age i.e. where age is greater then 30 and in other query we added an addition query i.e to include age over 30 and status as active. This filters the DataFrame to include only the rows that meet the specified criteria.