**CODING ASSESSMENT - 3**

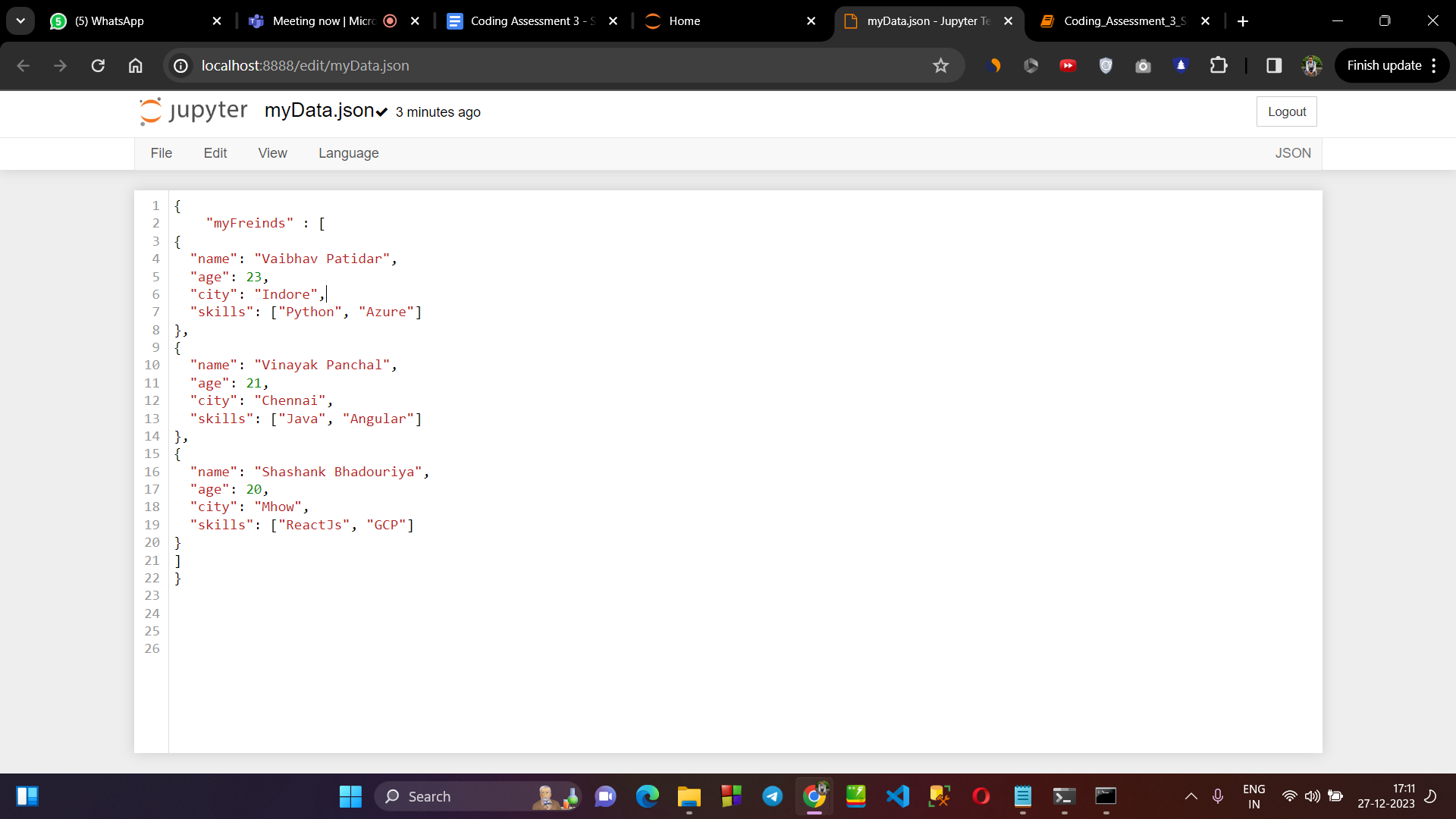
VAIBHAV PATIDAR

27-12-2023

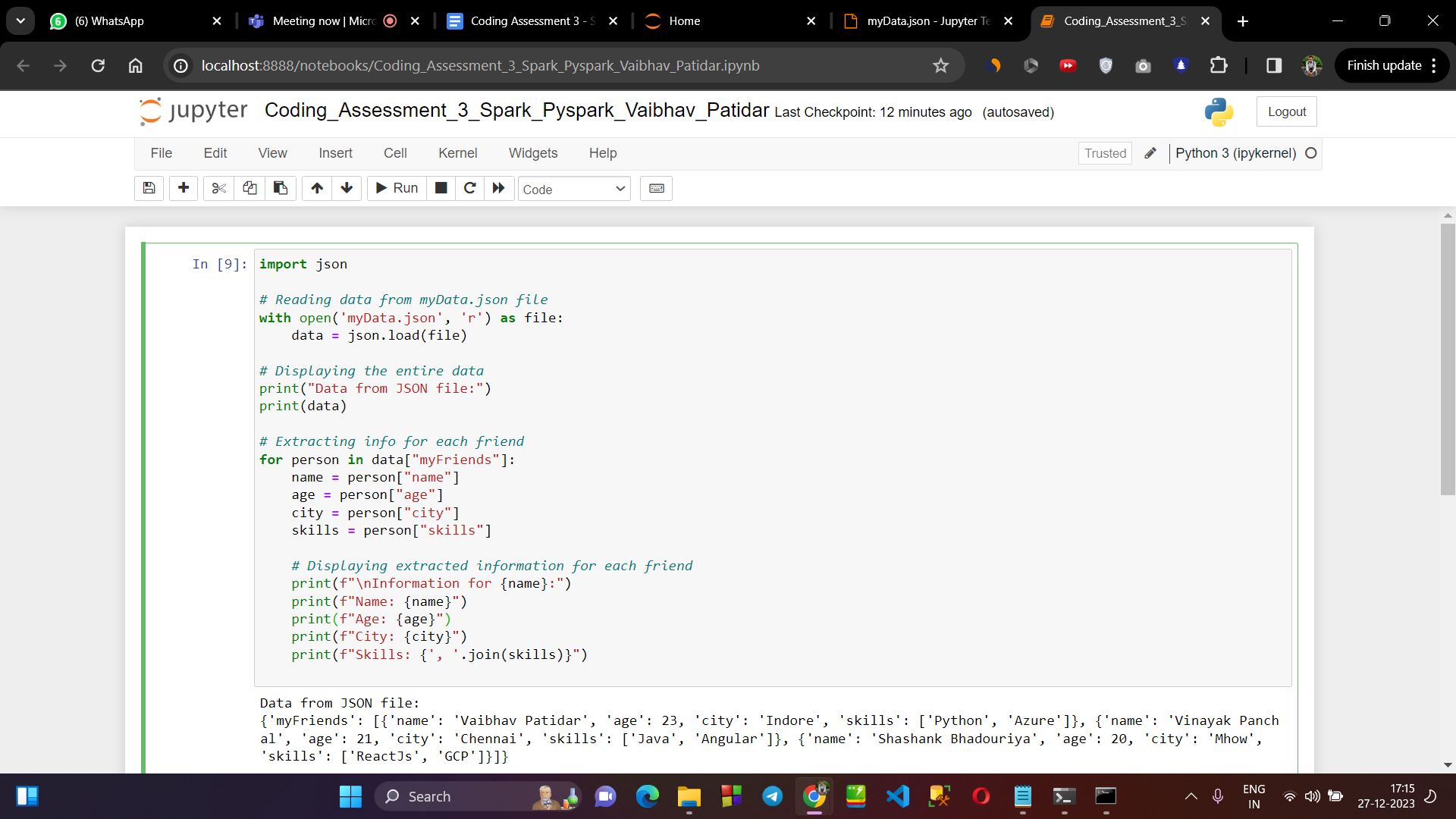
**PART 1 : Implement Processing JSON and CSV data with PySpark**

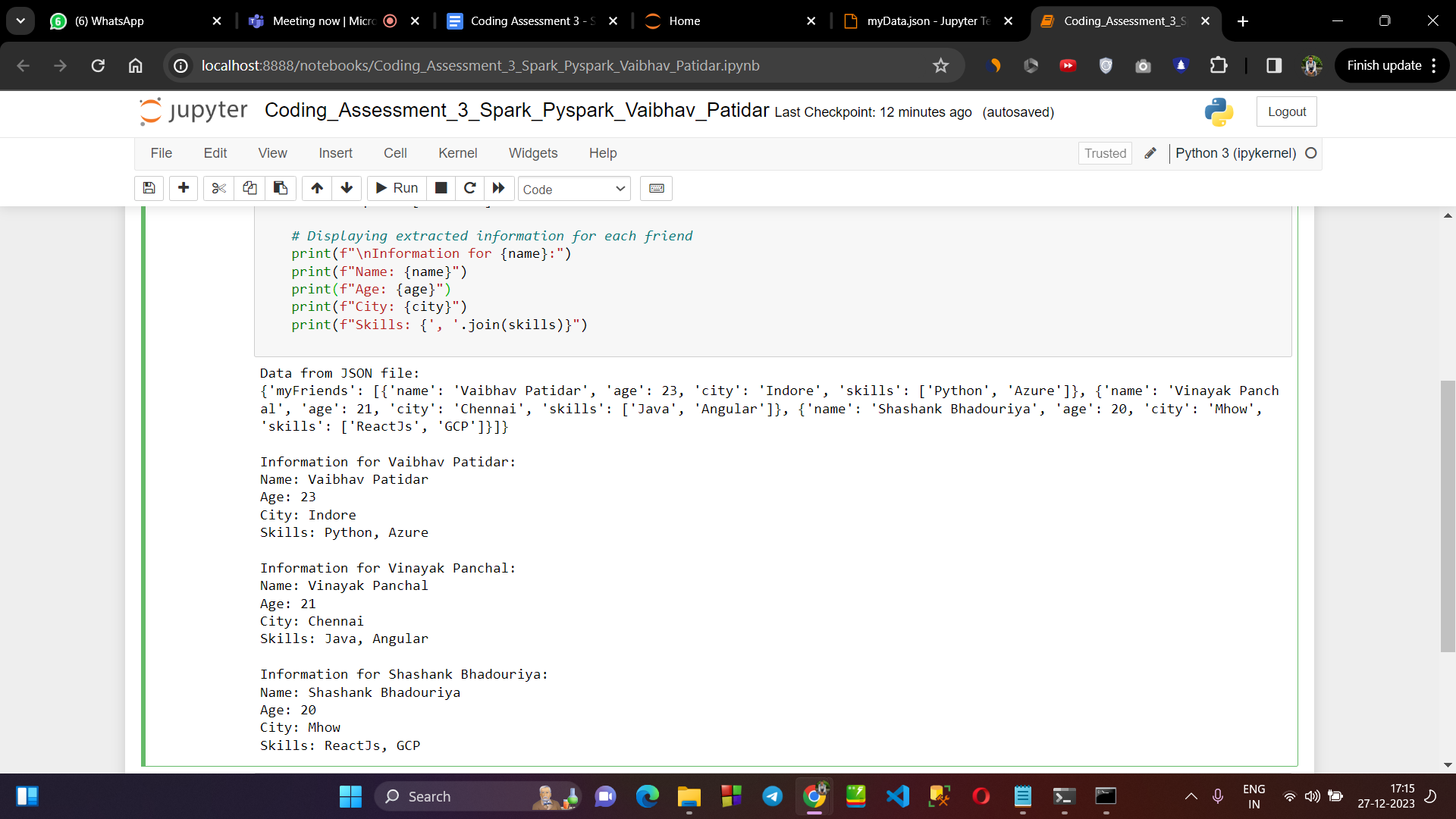
**→ CREATING A JSON FILE :**

* Here, i have created a json file named as “myData.json” which contains the info of my friends:



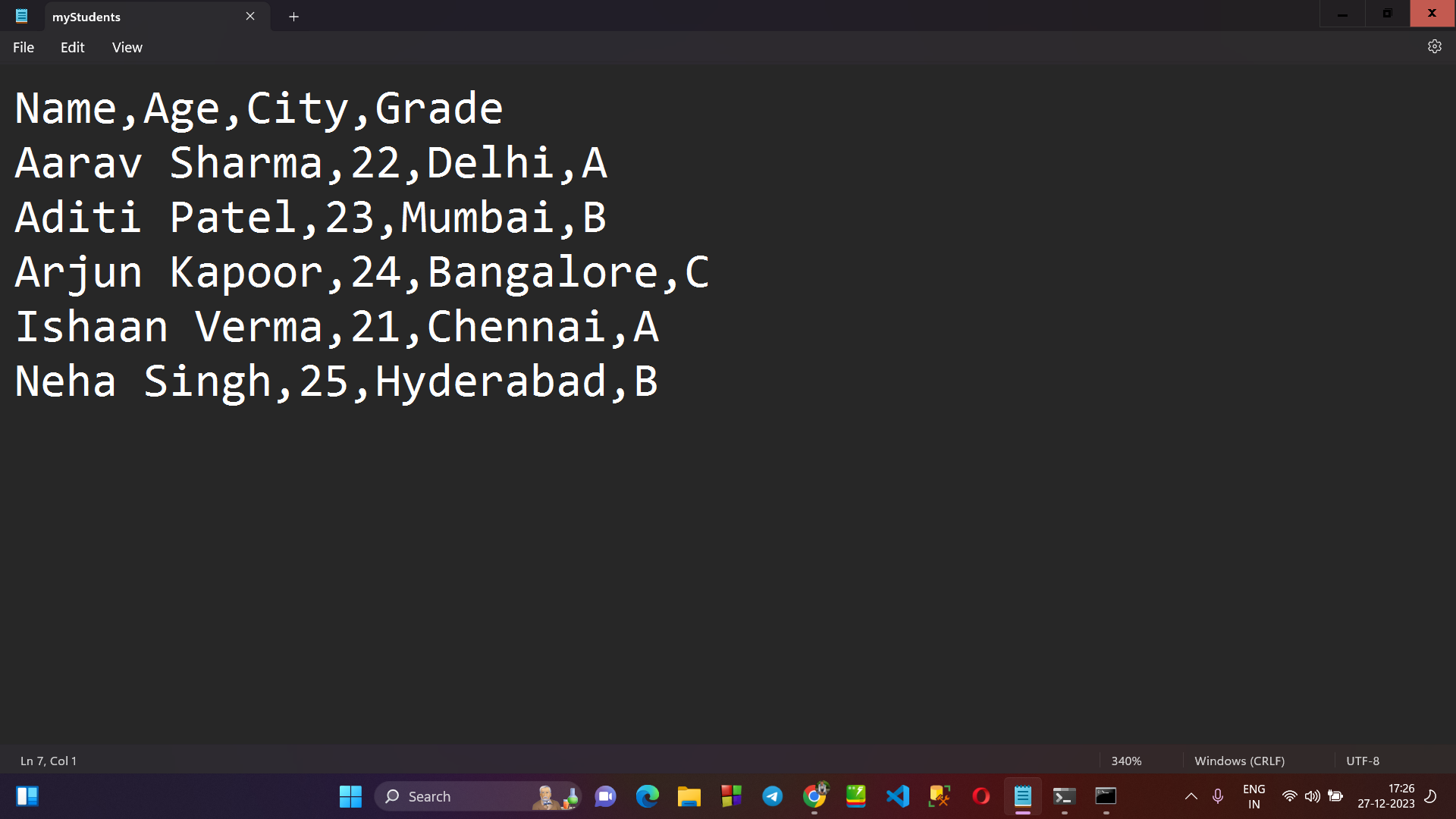
**→ READING THE DATA OF THIS JSON FILE :**

****

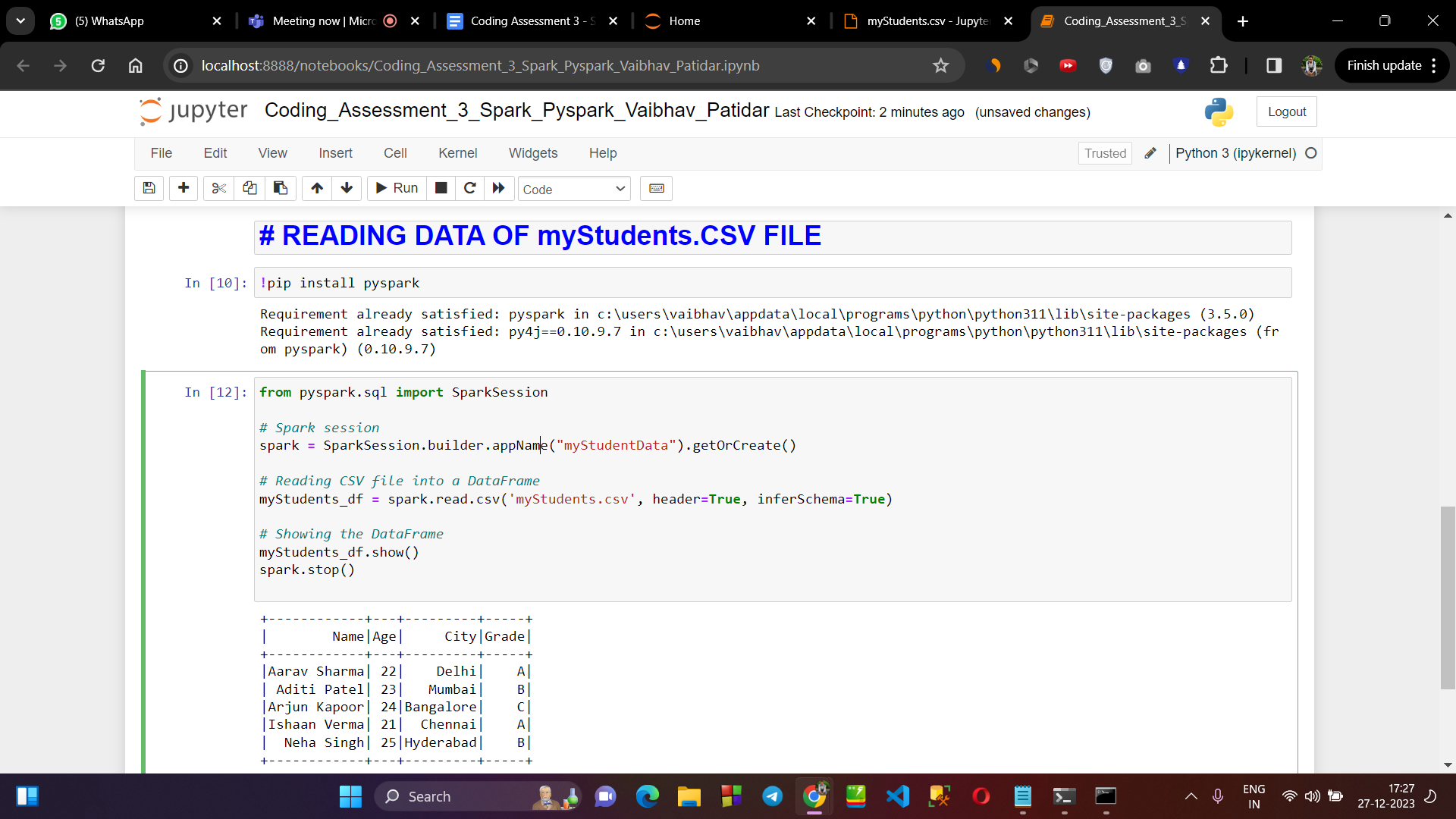
****

**→ CREATING A CSV FILE :**

* Here, I have created a csv file containing data of 5 students.

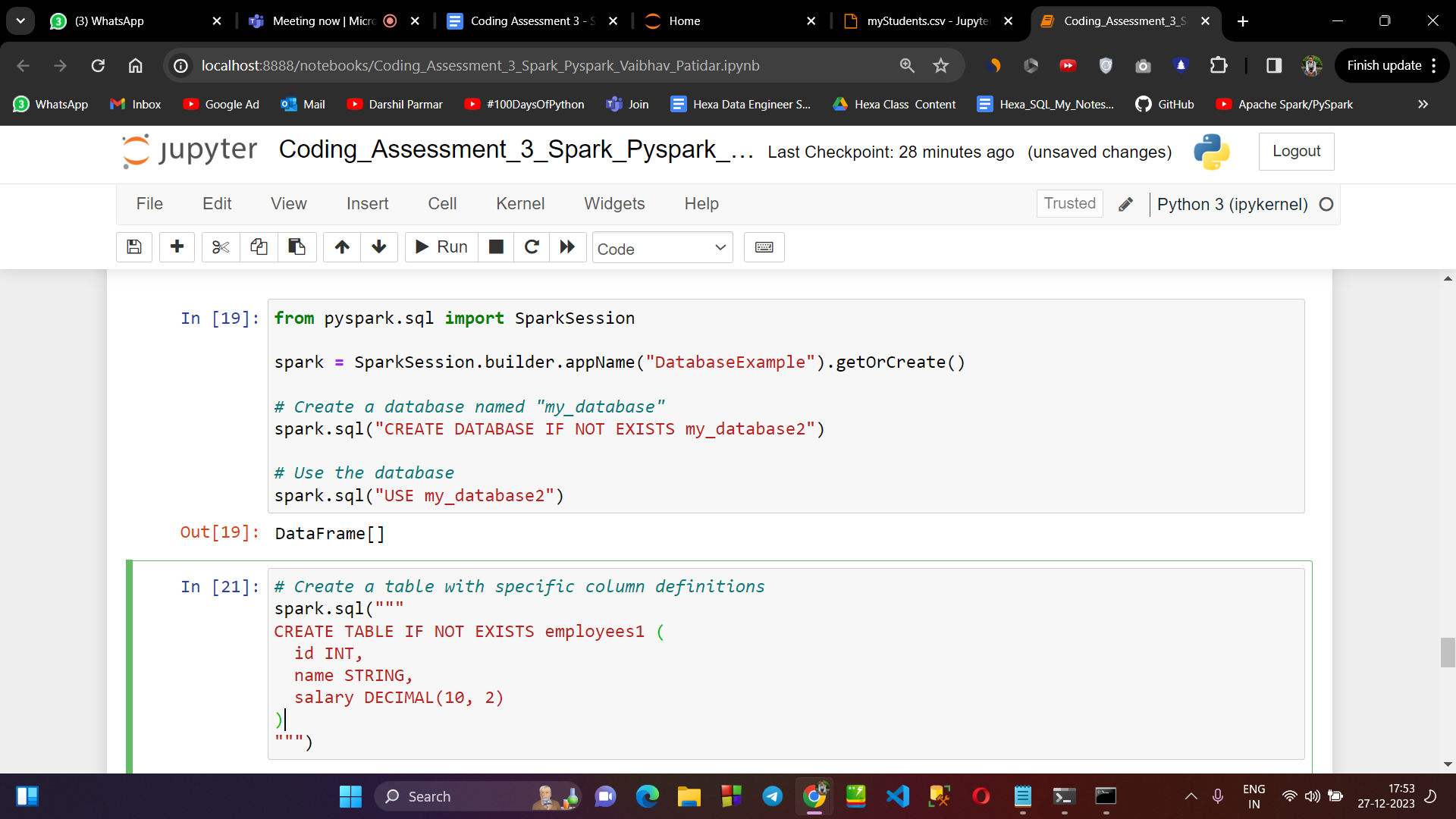


**→ READING THE DATA OF THIS .CSV FILE :**

****

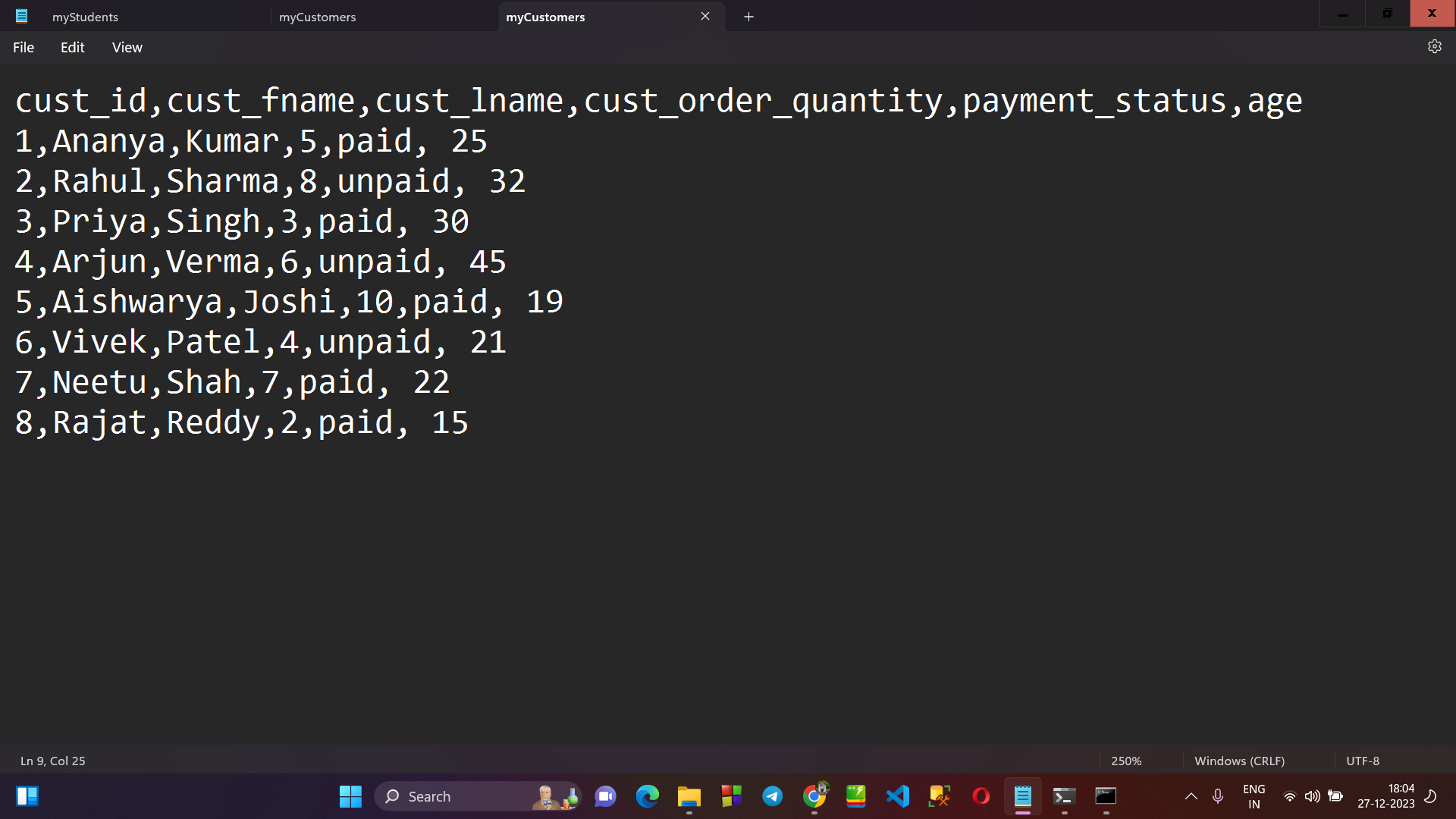
**\*\*\*\*\*\*\*\*\*\*\*\*\***

**PART 3 : CREATE DATABASE AND TABLE USING SPARK SQL**

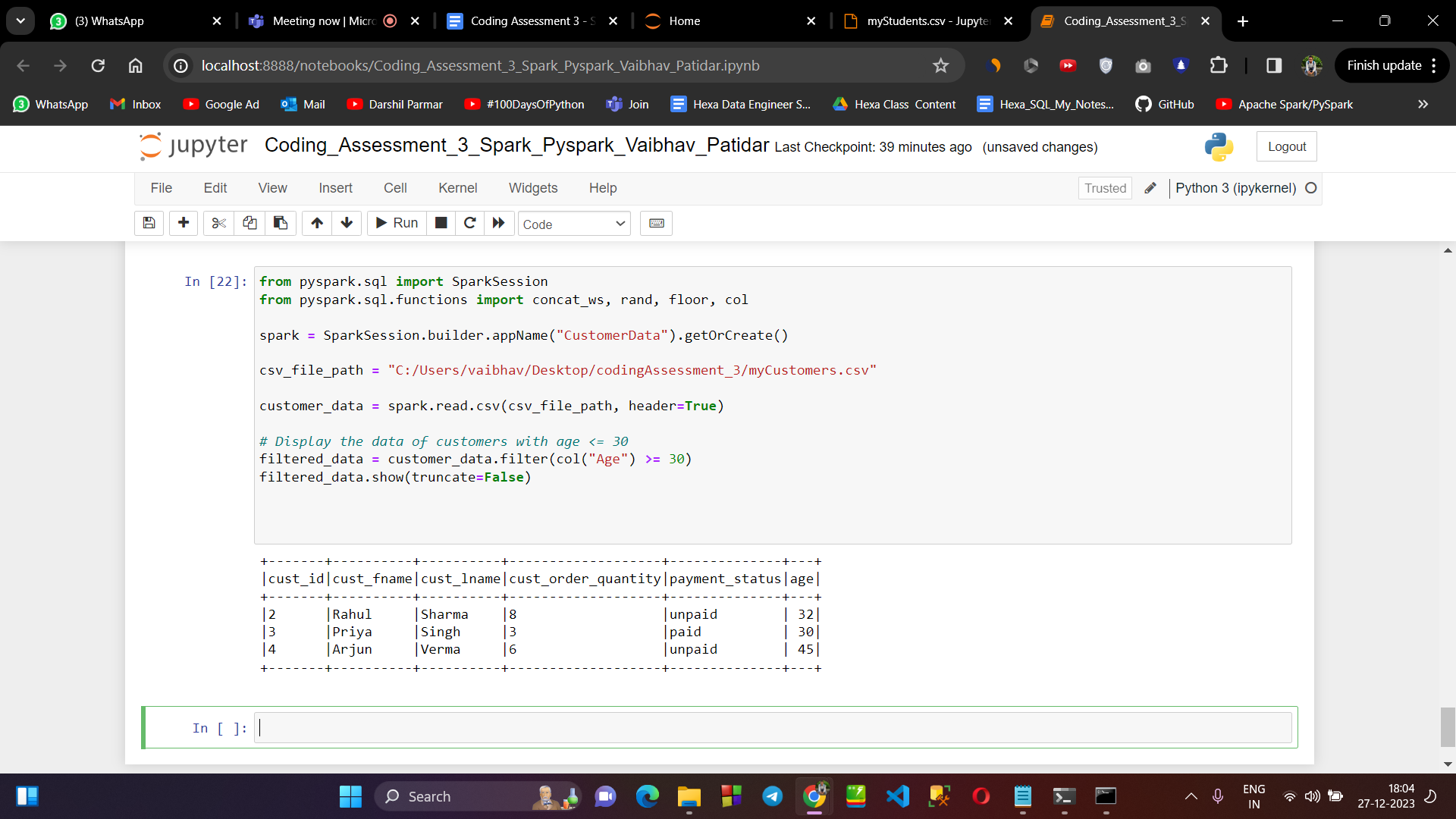
****

**PART 4 : PERFORMING Transformations such as Filter, Simple Aggregations, GroupBy.**

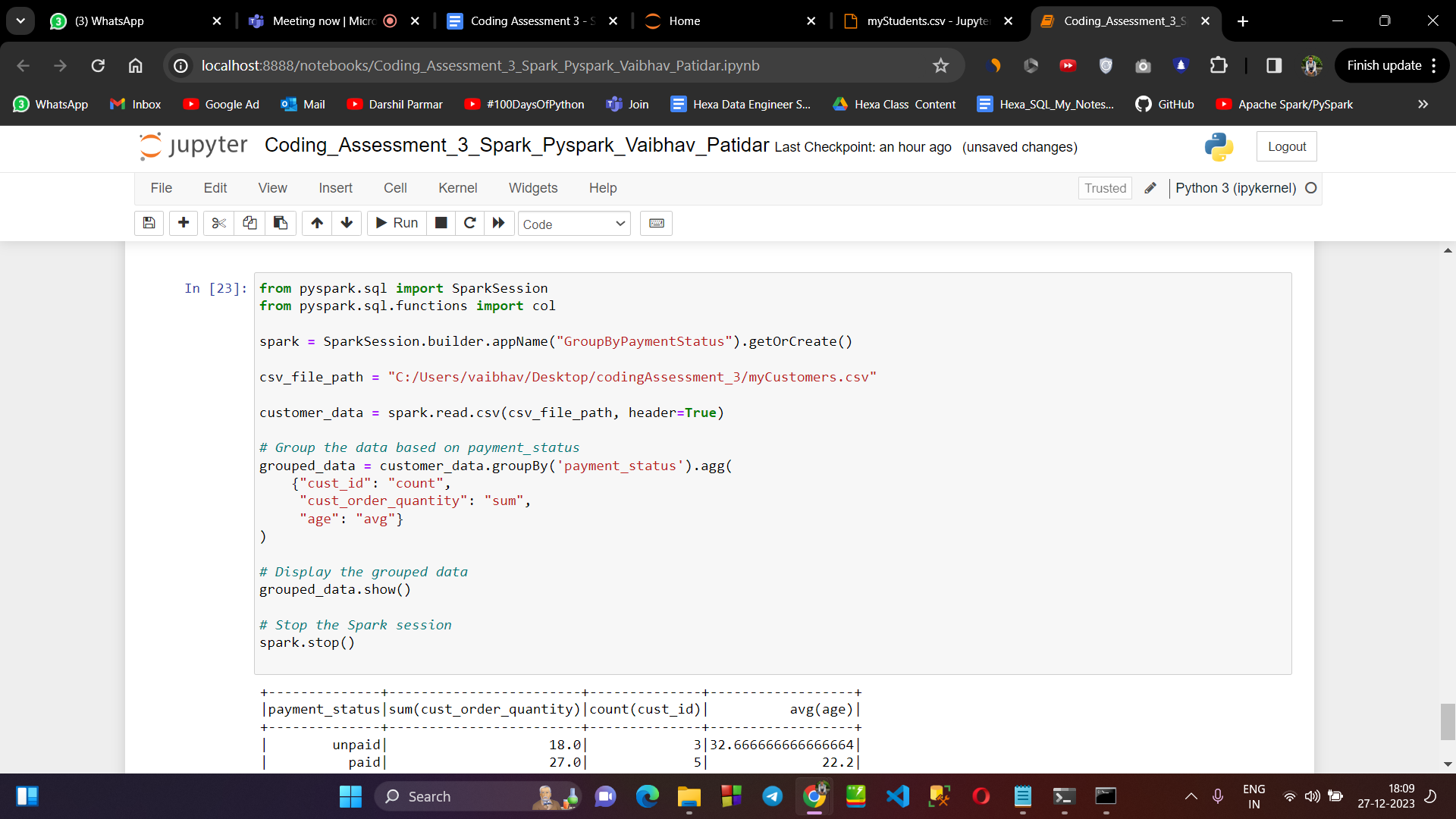
**Creating a csv file containing info of 8 customers :**

****

**→ FILTER METHOD :**

****

**→ GROUP BY AND AGGREGATE :**

****

**PART 2 : Explain ETL (Extract, Transform, Load) with PySpark**

* ETL stands for “Extract, Transform & Load”.
* **Extract:** Retrieve data from various sources.
* **Transform:** Apply operations (filter, aggregate, etc.) to convert and clean the data.
* **Load:** Store the processed data into a target destination.