**Sunkari Manohar**

**Pyspark Coding Assessment**

**1.Implement processing Json and csv data with pyspark**

**Implementing csv data with pyspark:**

Here first read the csv file and then display the csv file

To read csv file used command as:

# Read CSV data

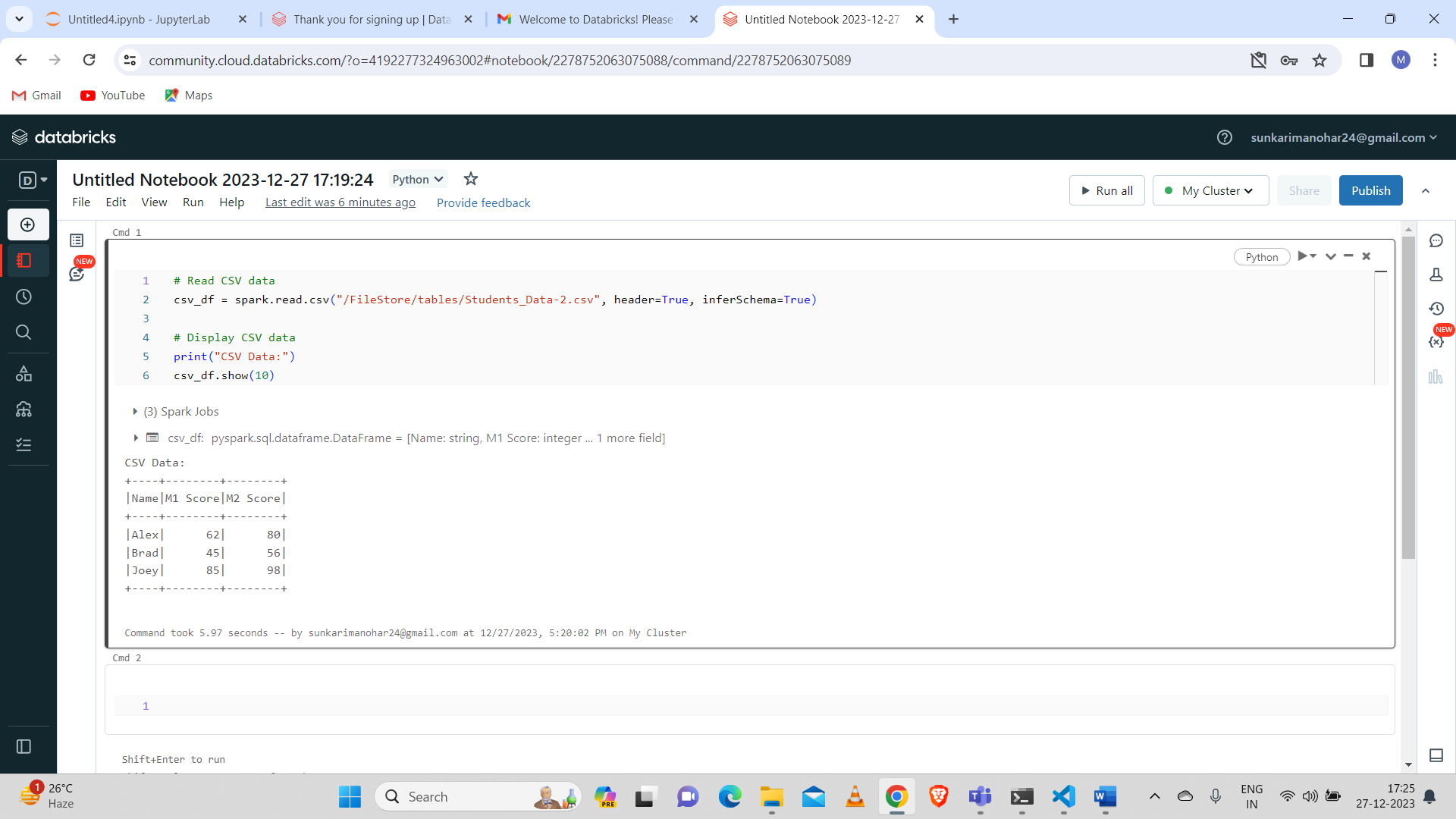
csv\_df = spark.read.csv("/FileStore/tables/Students\_Data-2.csv", header=True, inferSchema=True)

To display the csv file used command as:

# Display CSV data

print("CSV Data:")

csv\_df.show(10)

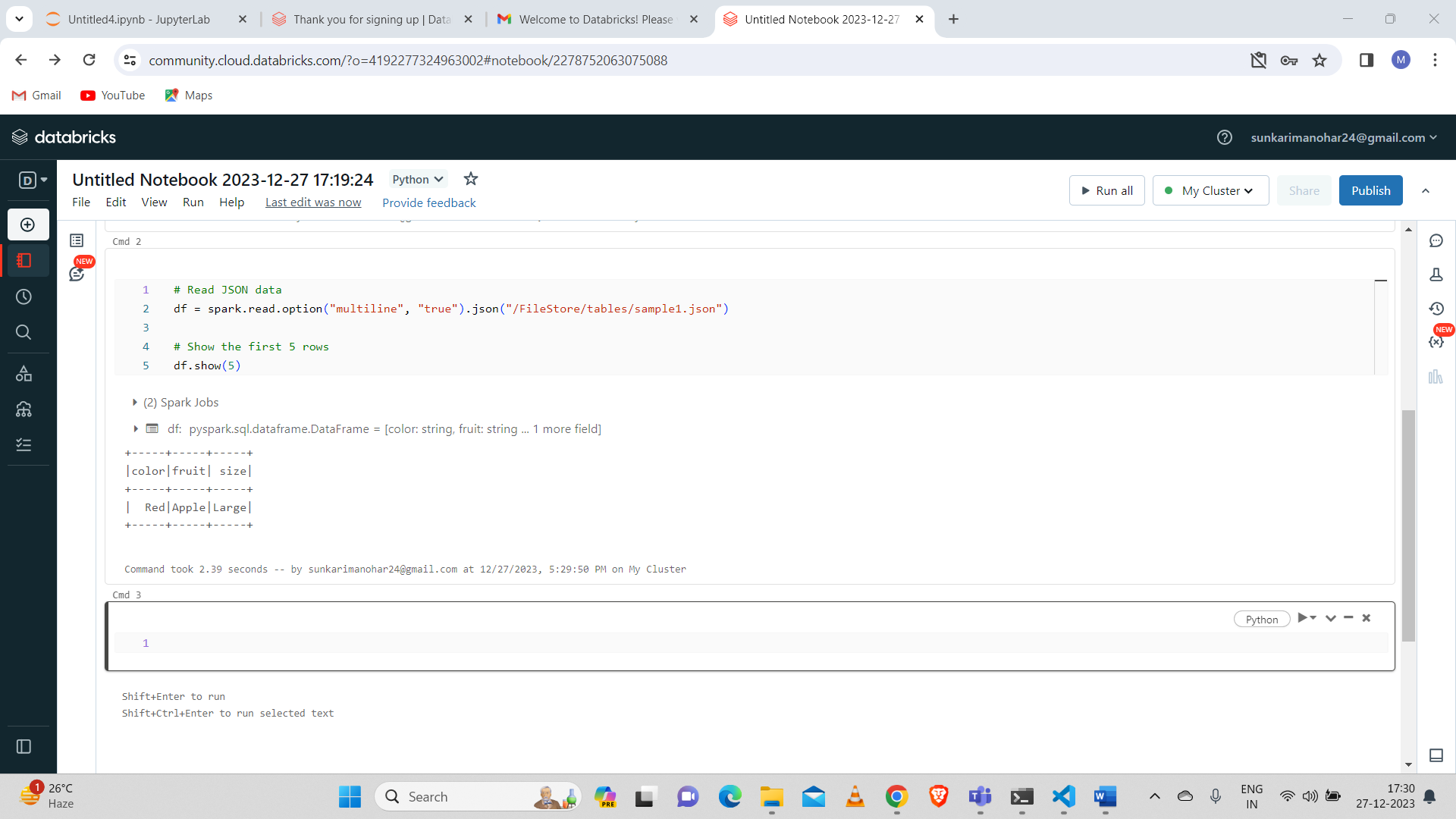


**Implement Json data with Pyspark:**

**For Json file to read:**

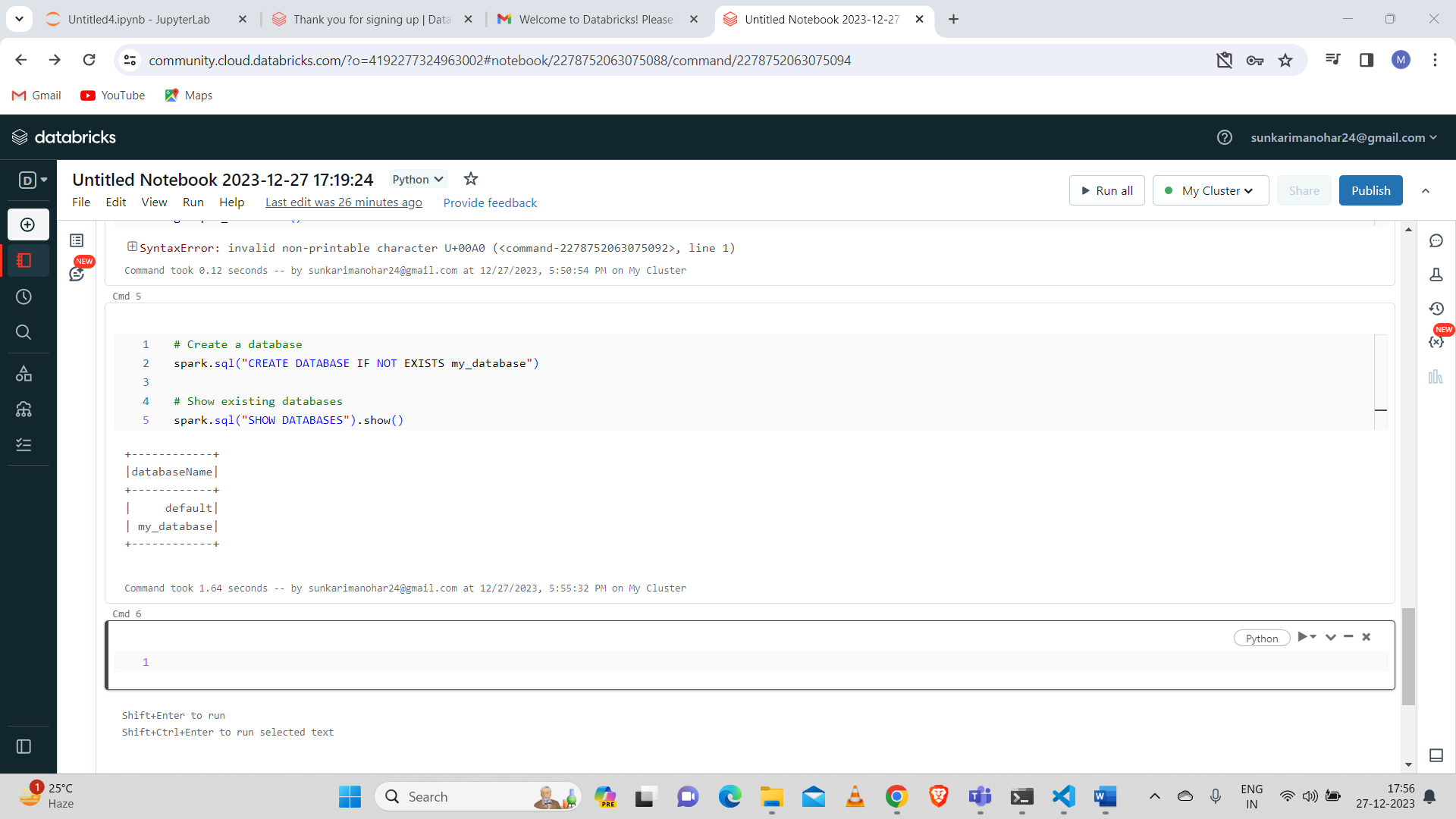
# Read JSON data

df = spark.read.option("multiline", "true").json("/FileStore/tables/sample1.json")



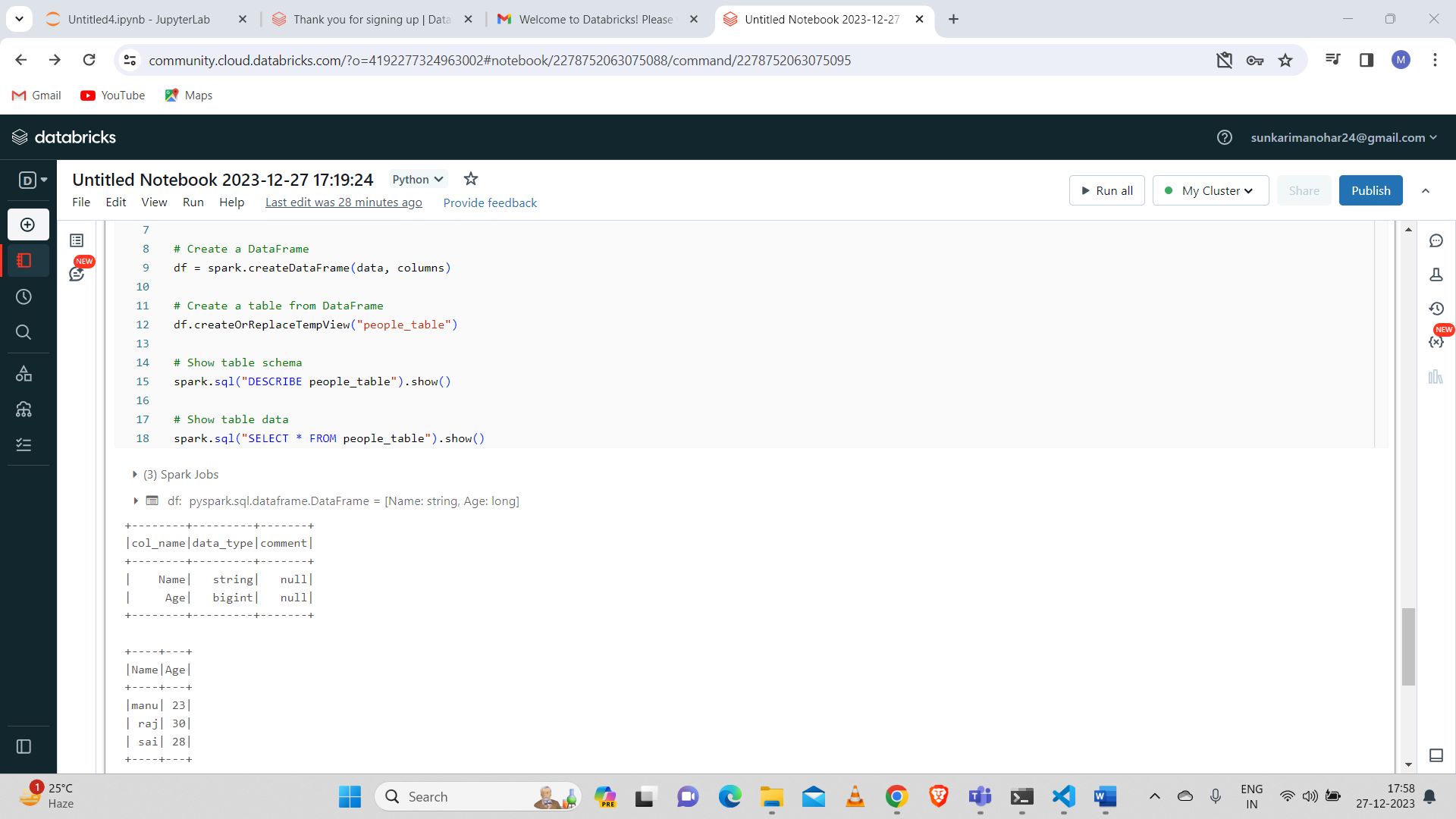
**3. Using Spark SQL – creating databases, tables**

First here created database called my\_database if the database doesnot exist.



And then using that database created data frames with two columns like “name” and “age”

And also created table named “people\_table” and also showed that table



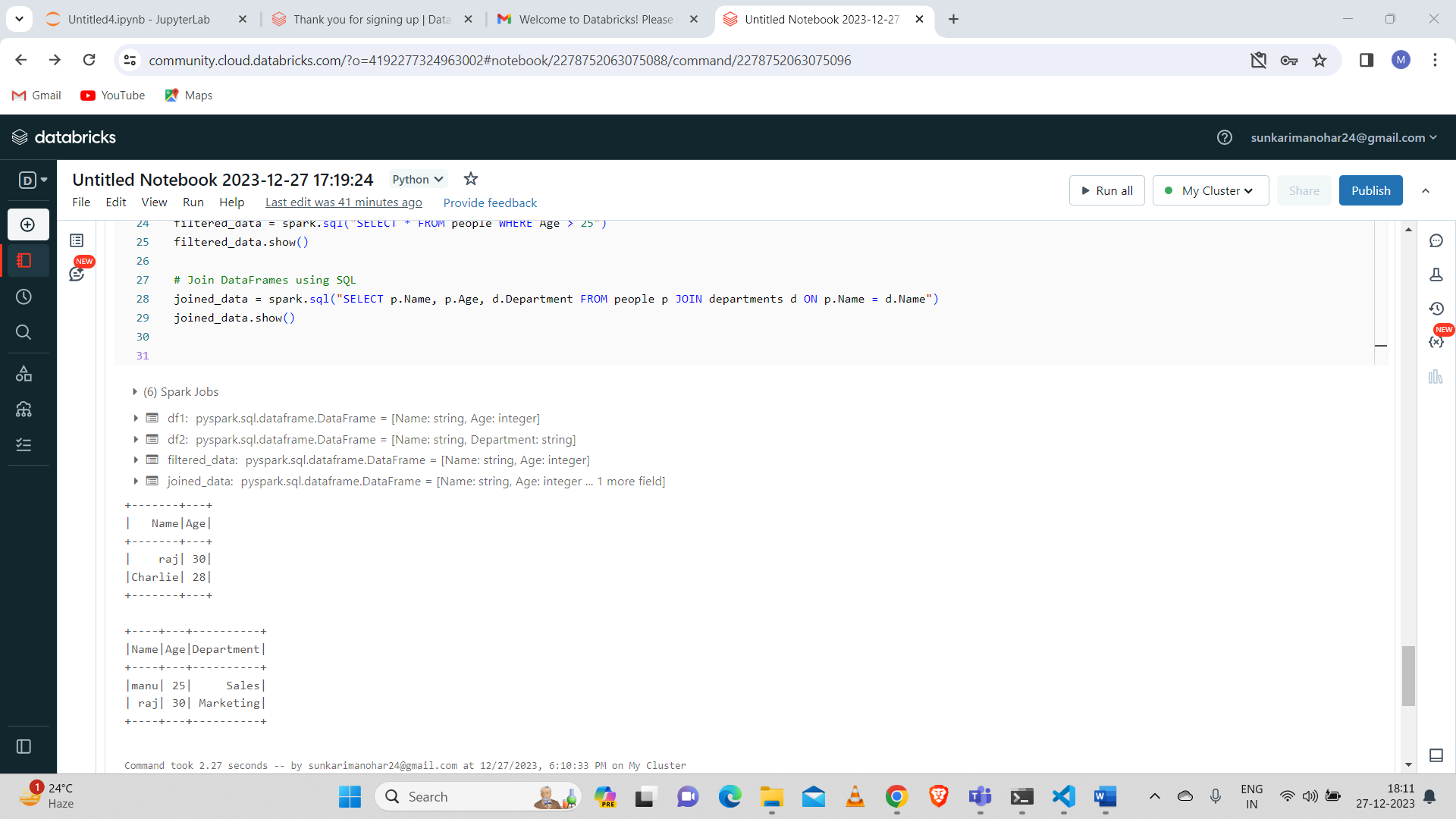
**4. Using Spark SQL transformations such as Filter, Join, Simple Aggregations, Group By**

Here I created two data frames named df1 and df2.

First data frame contains columns like “name” and “age”.

And second data frame contains columns like “name” and “department”.

Here I performed transformations like Filter and Join:



Here I done on Simple Aggregations and Group By aggregation:

