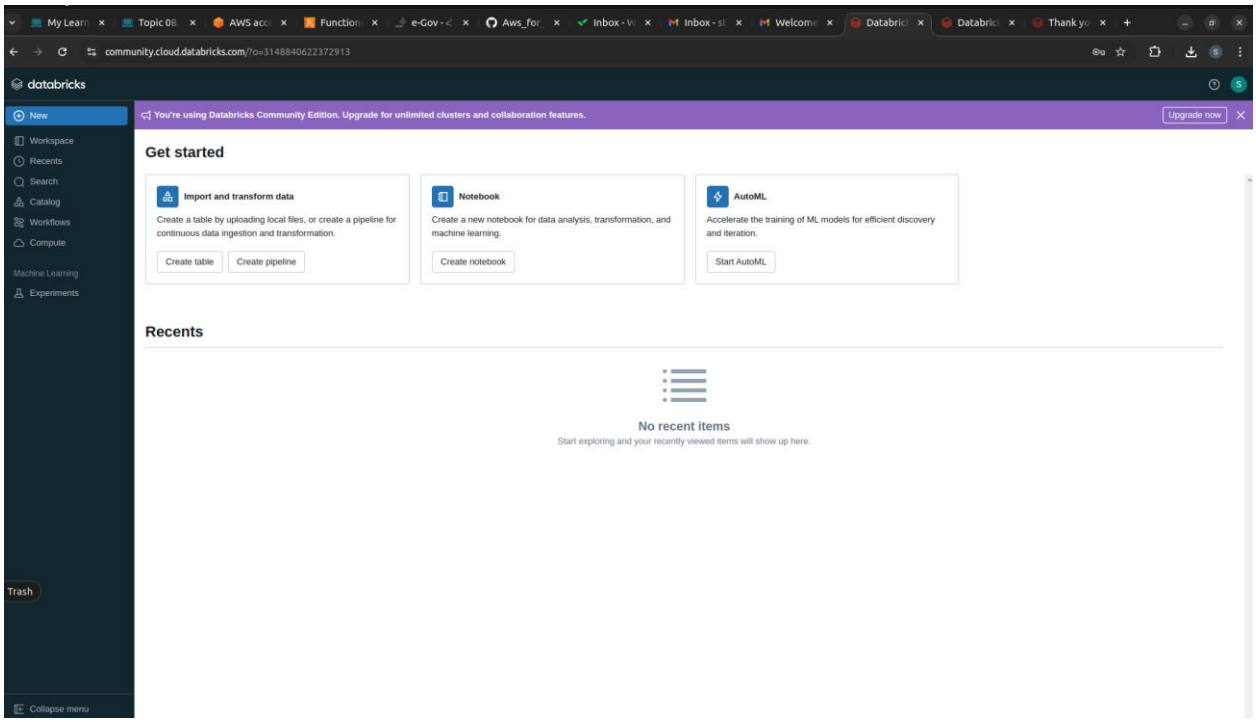
EPAM Systems, RD Dep., RD Dep.

DATABRICKS OVERVIEW

DQE

Account created.



Dbc file imported in Databricks.

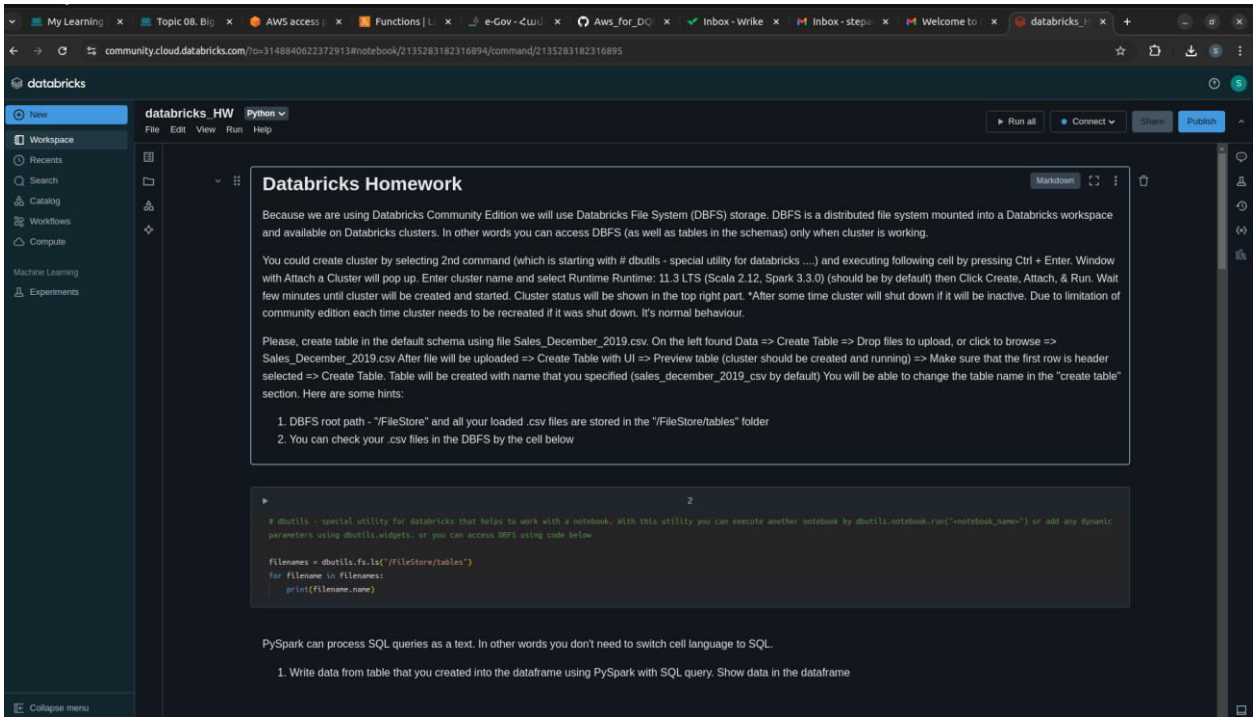
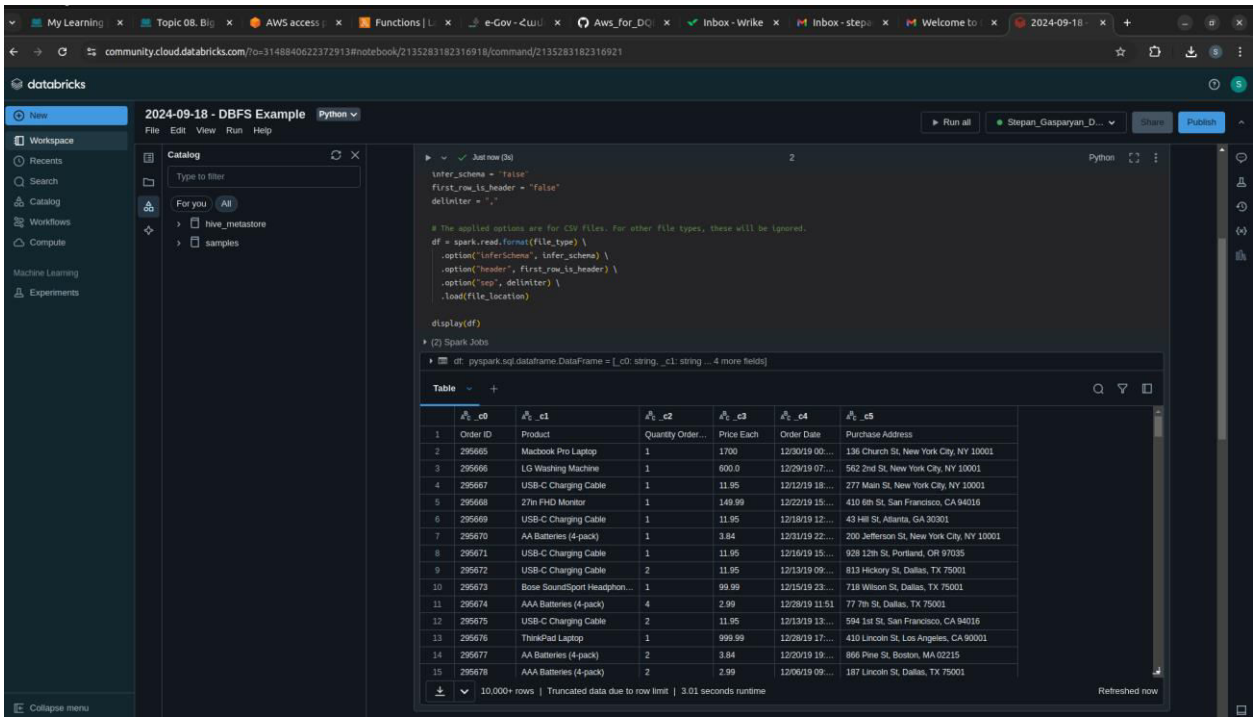
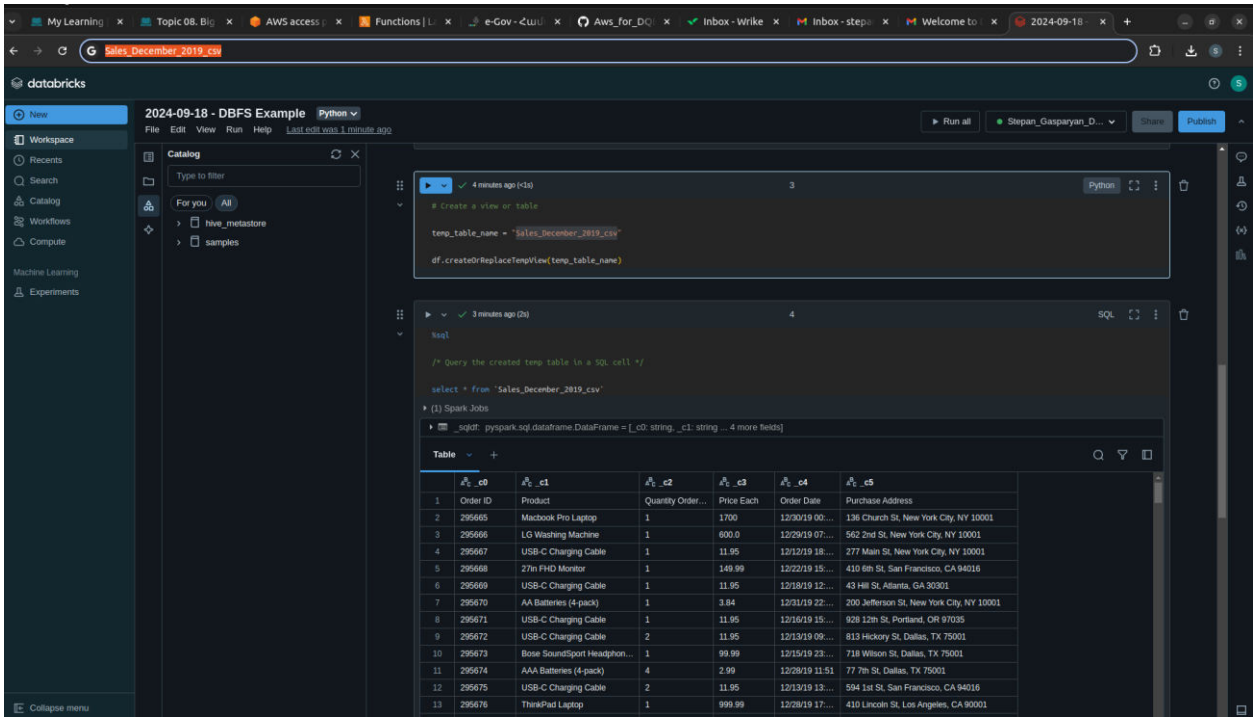


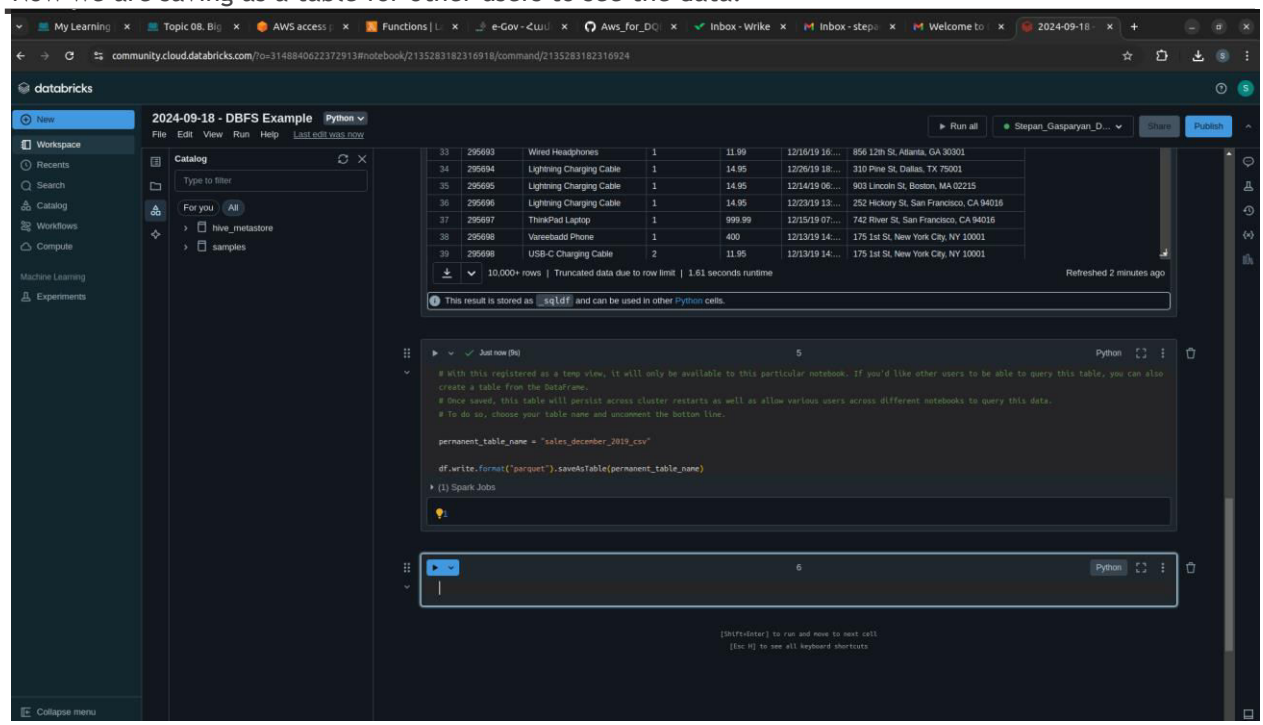
Table created from provided csv file .



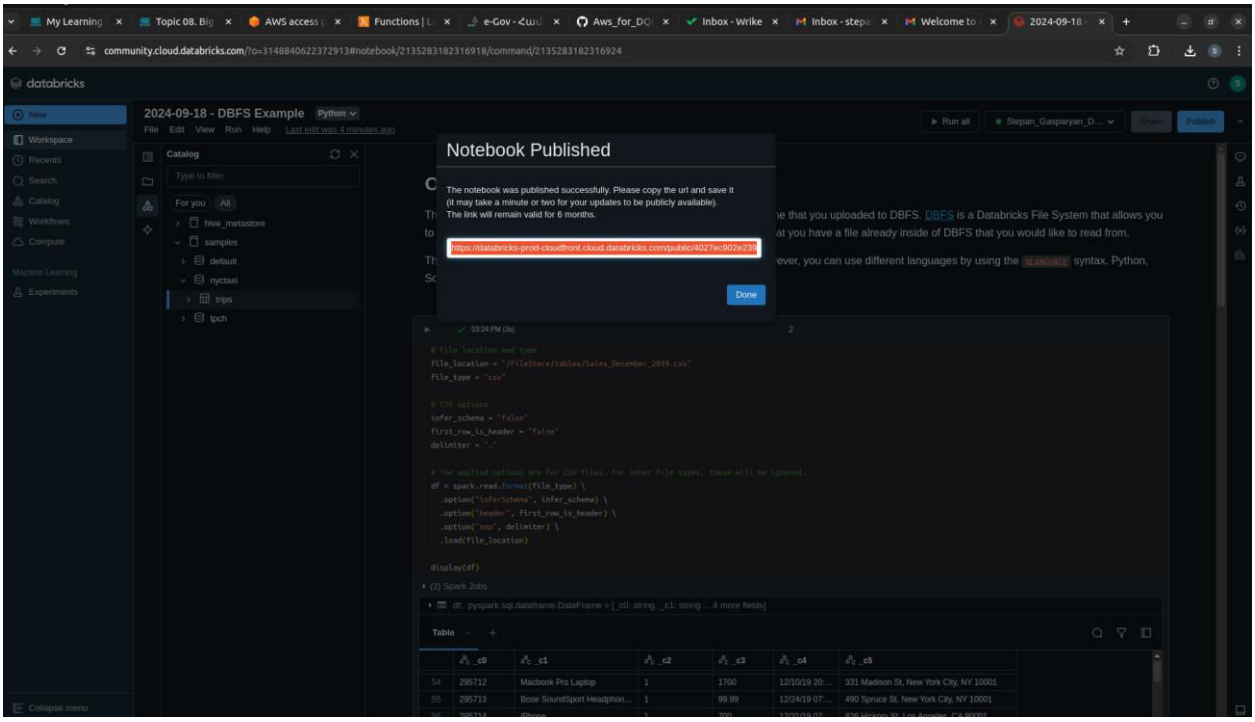
TempView created with the name Sales_December_2019_csv ,see data inside.



Now we are saving as a table for other users to see the data.



Finally I published the notebook.



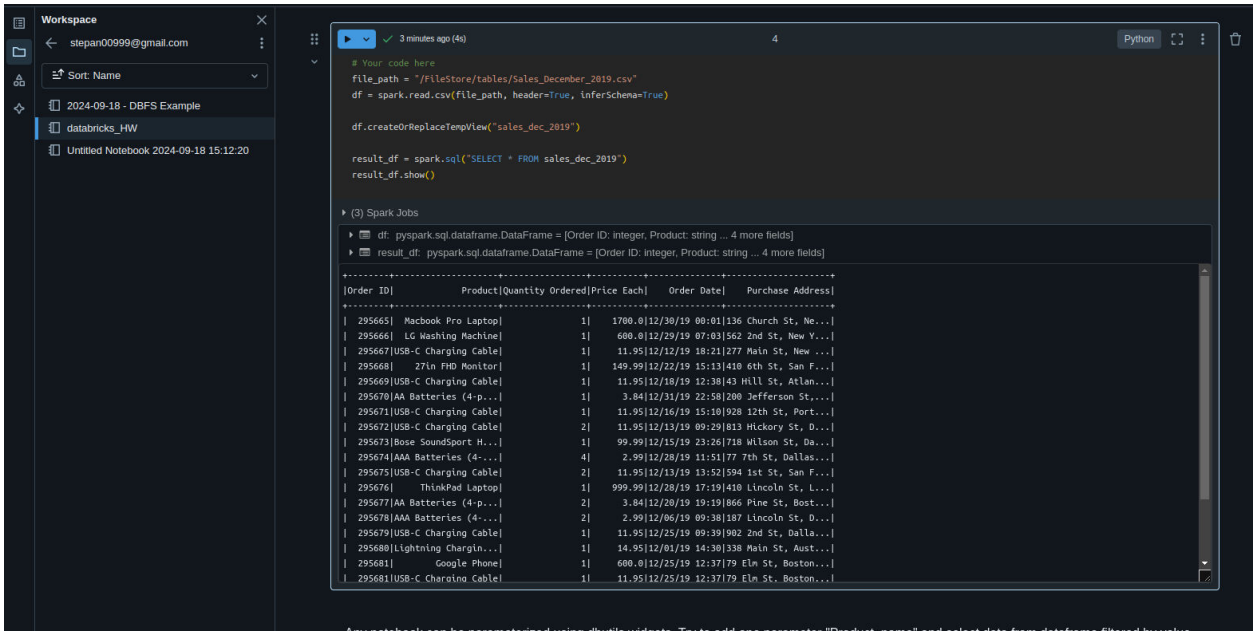
Link to notebook -

https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/3148840622372913/2135283182316918/4761590133845256/latest.html

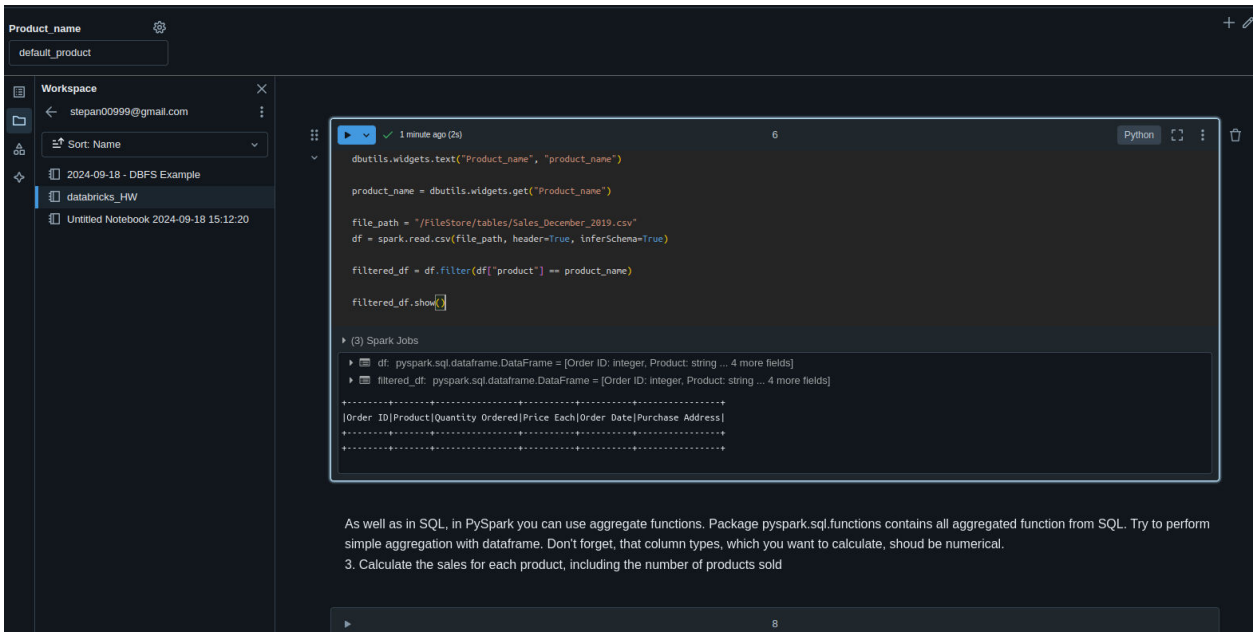
Now lets do HW file tasks.

1. We can see that we got the file in our file system.

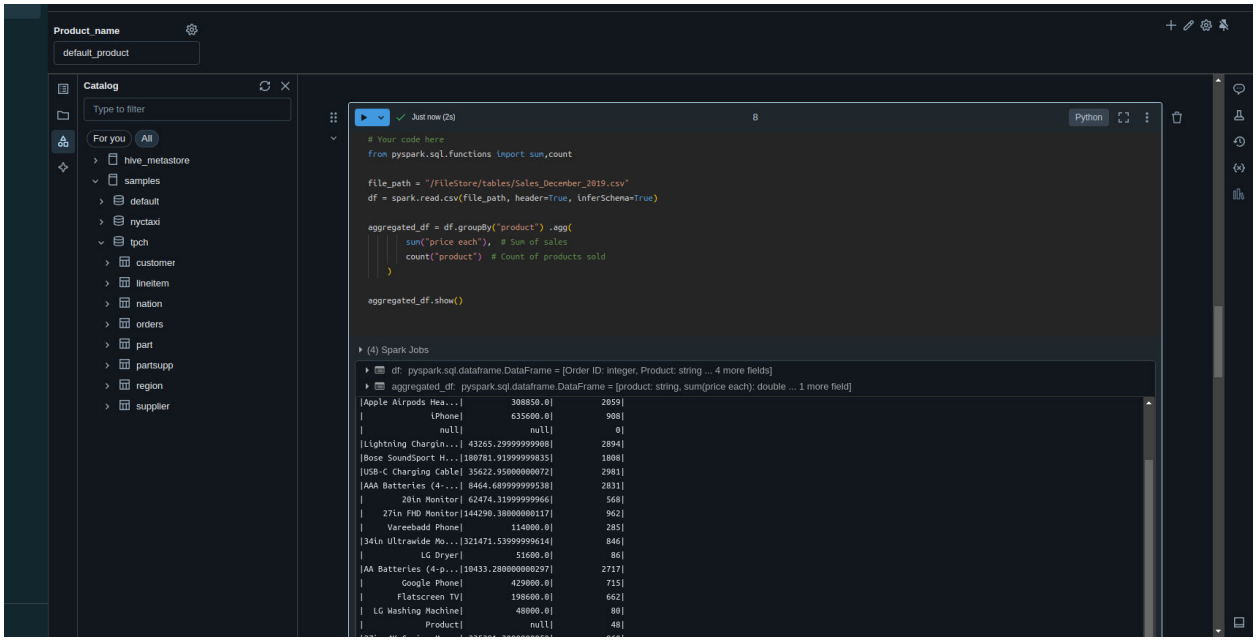
2. First we need to read our file then create a view. Next we use sqk inside a spark sql to extract all data.



3. Db util created with product_name value and used in df filtering.

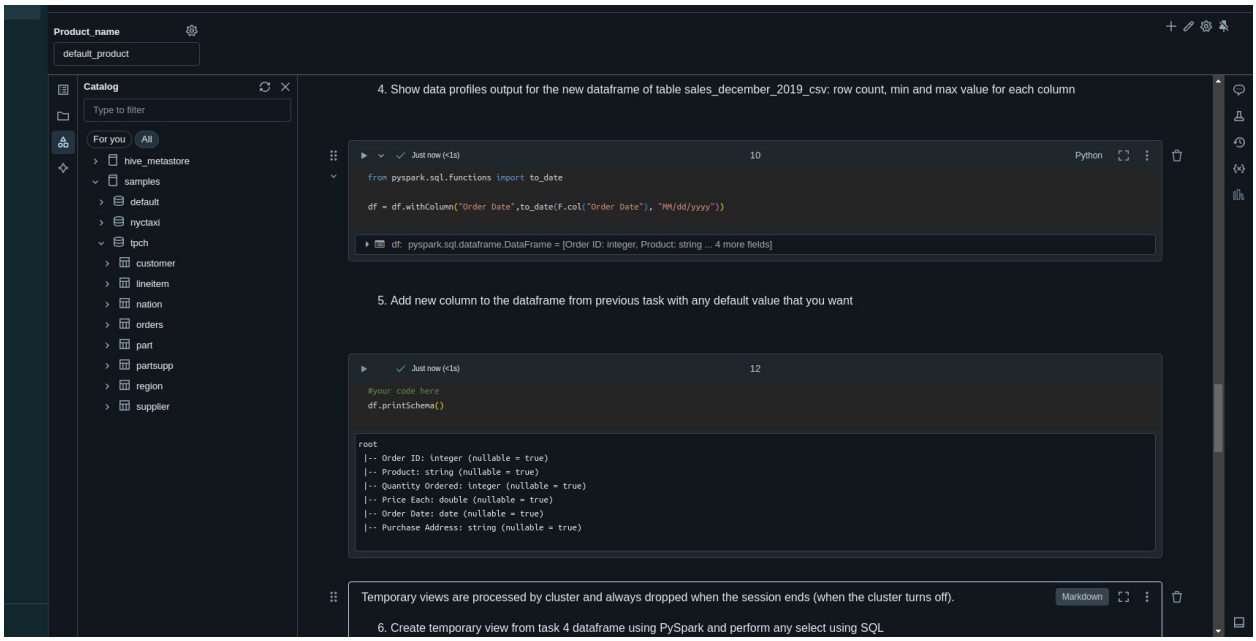


4. Grouped by product table:

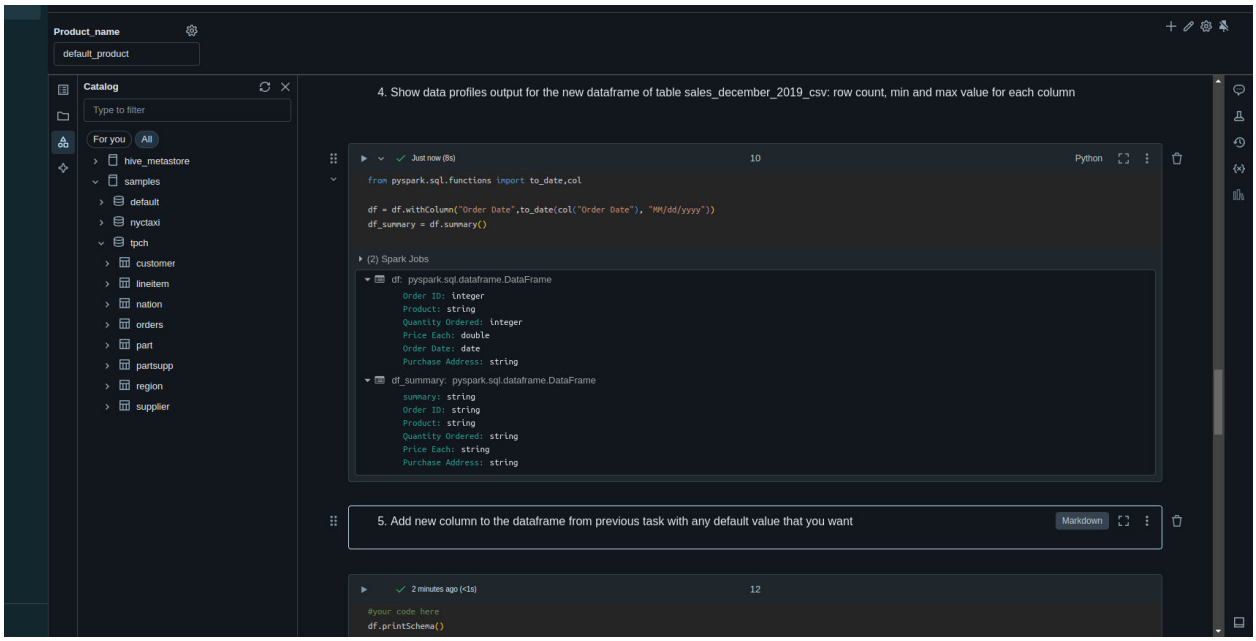


.

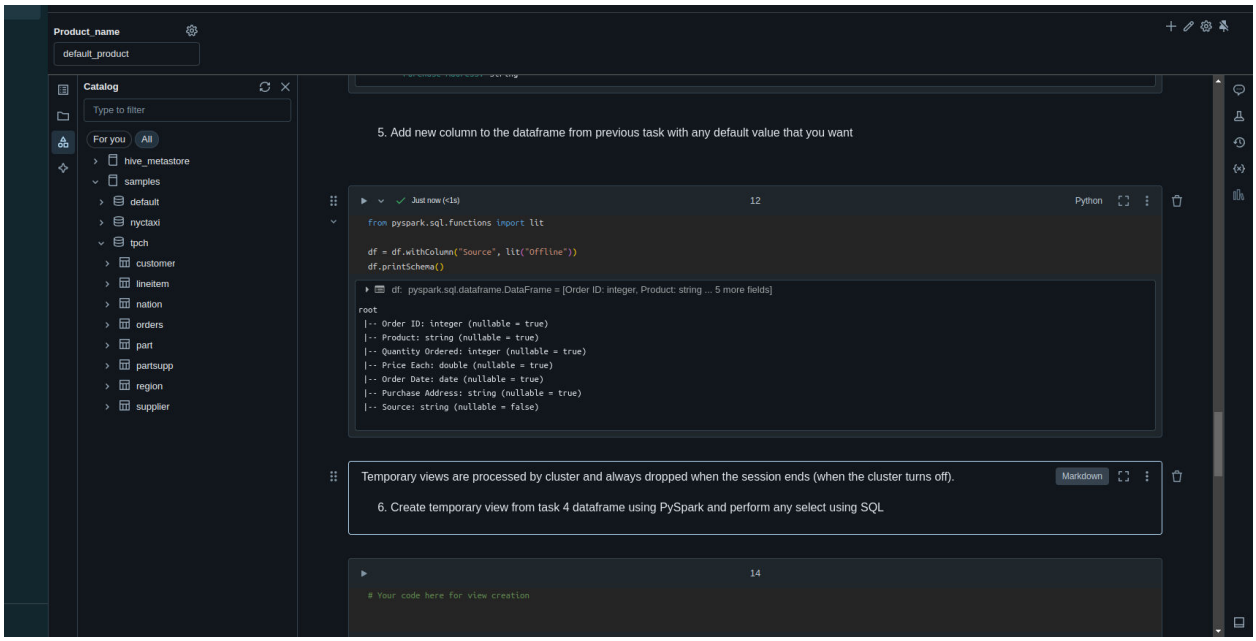
5. The date column stored as string , I have changed the datatype to DATE:



See result

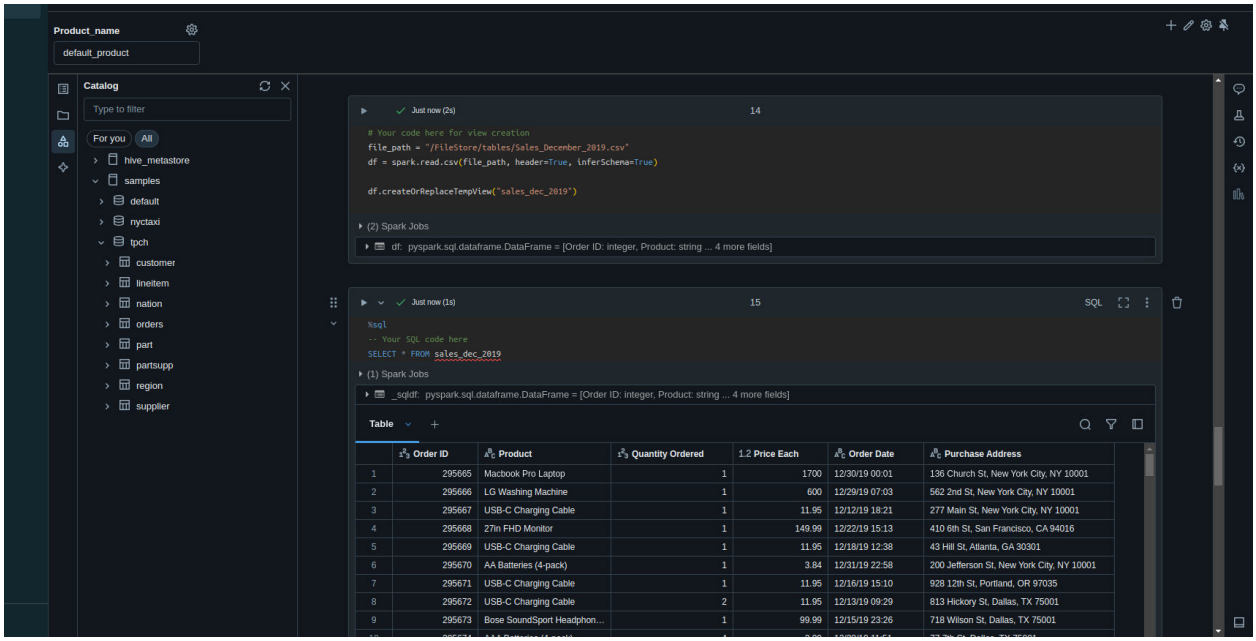


6.New column added with the default value:

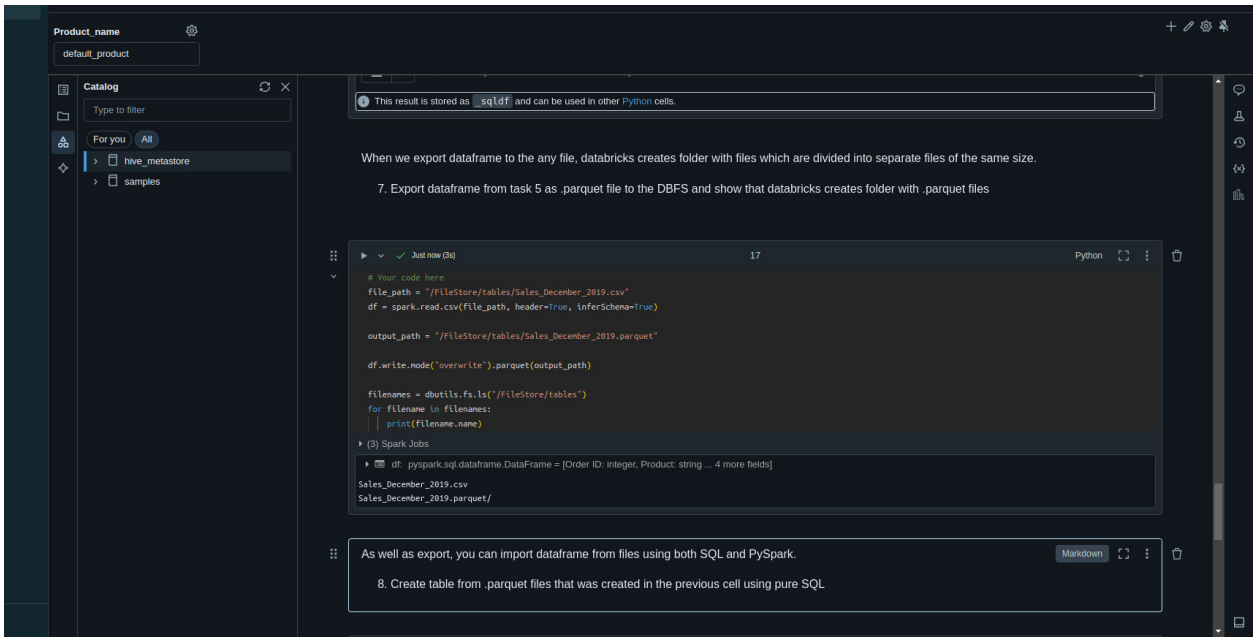


.

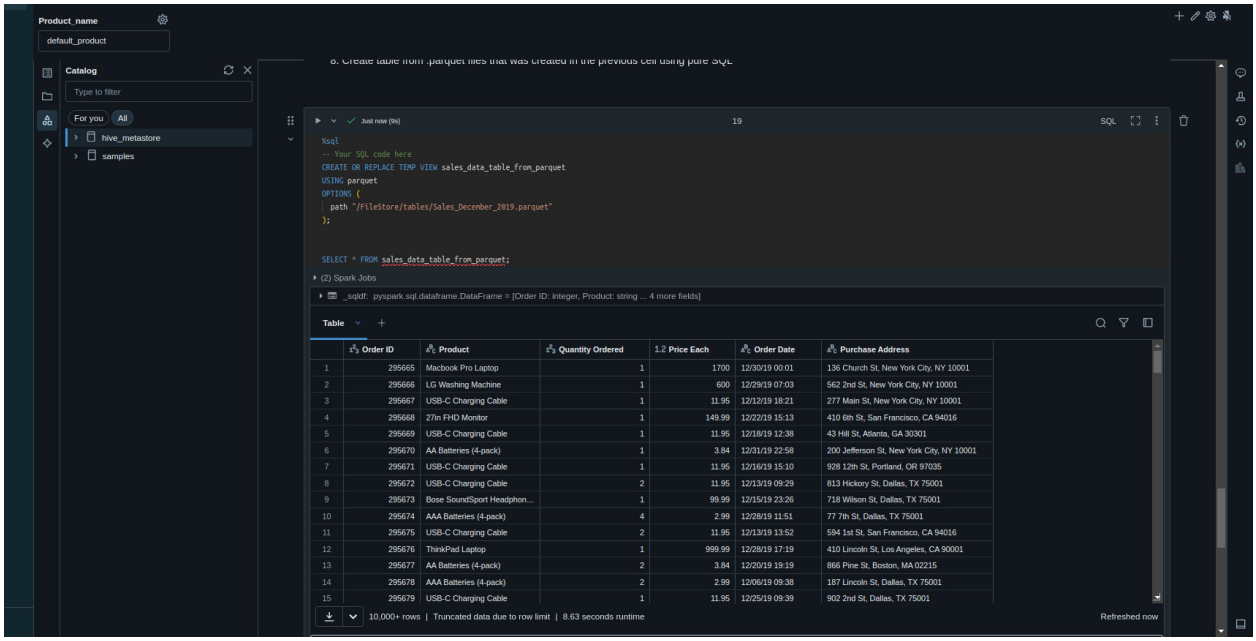
7. New table view created from csv file in the top block and then we just query that view using SQL.



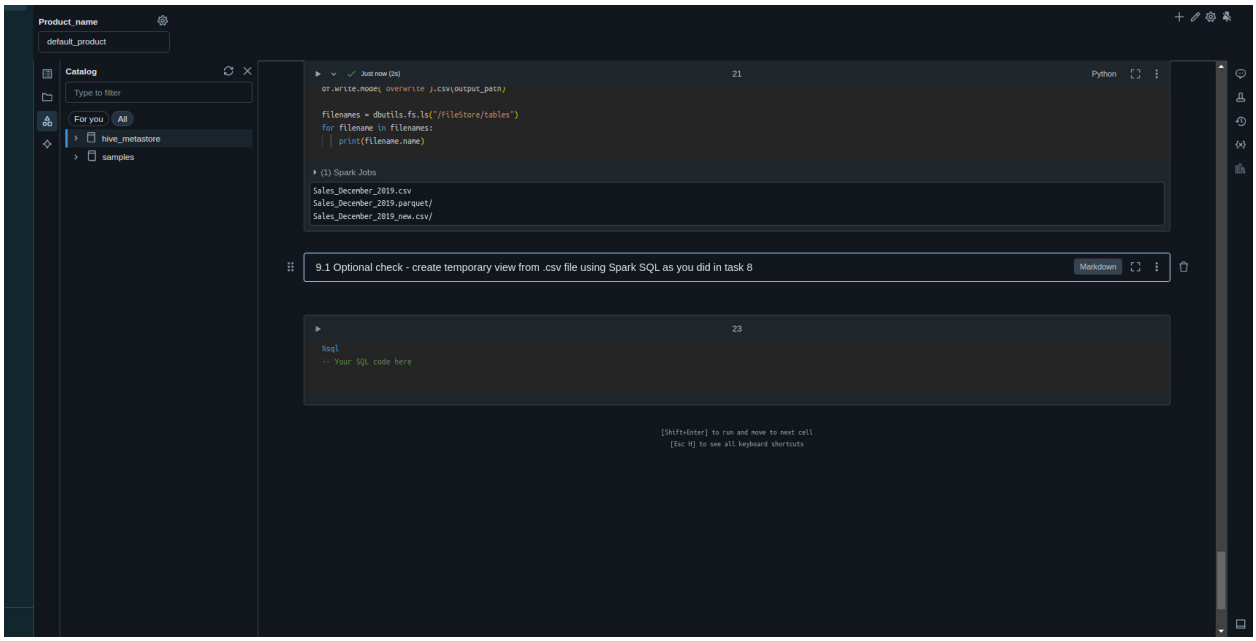
8. New parquet file created from csv:



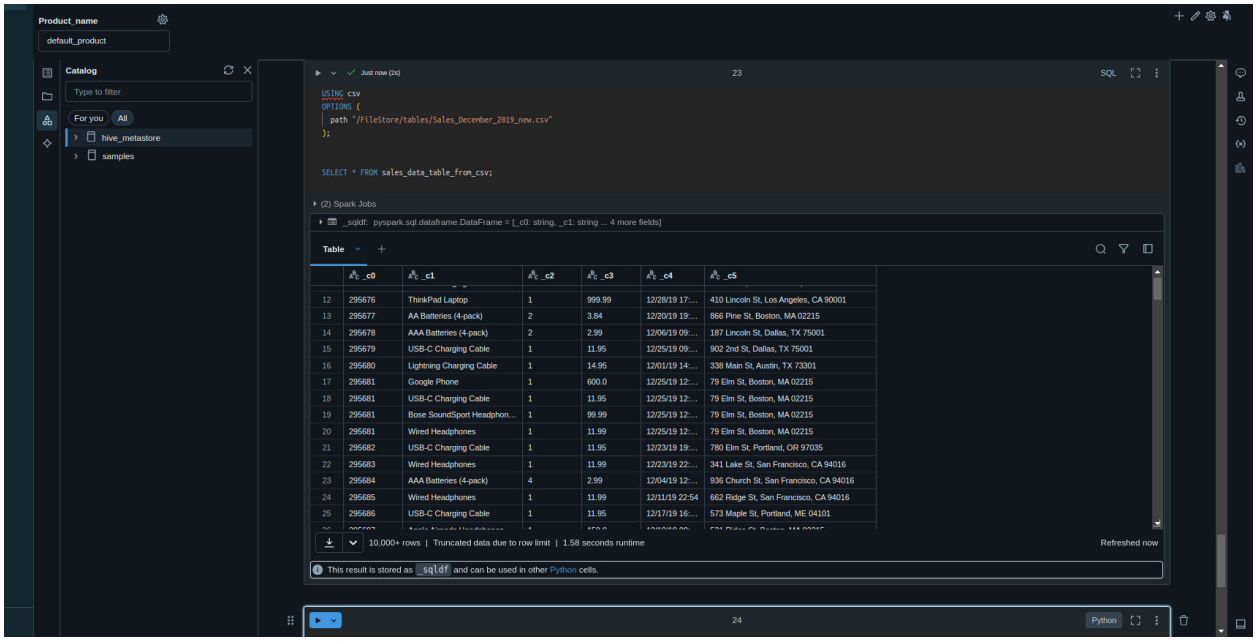
9. Created temp table using parquet file:



10. New csv file created from previous dataframe.



11. New temp table from new csv file we stored in file system.



Homework notebook publishet at link :

https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/3148840622372913/2135283182316894/4761590133845256/latest.html