**DATABRICKS ASSESSMENT**

|  |  |
| --- | --- |
| **NAME** | KAILASH S |
| **DATE** | 06-01-24 |

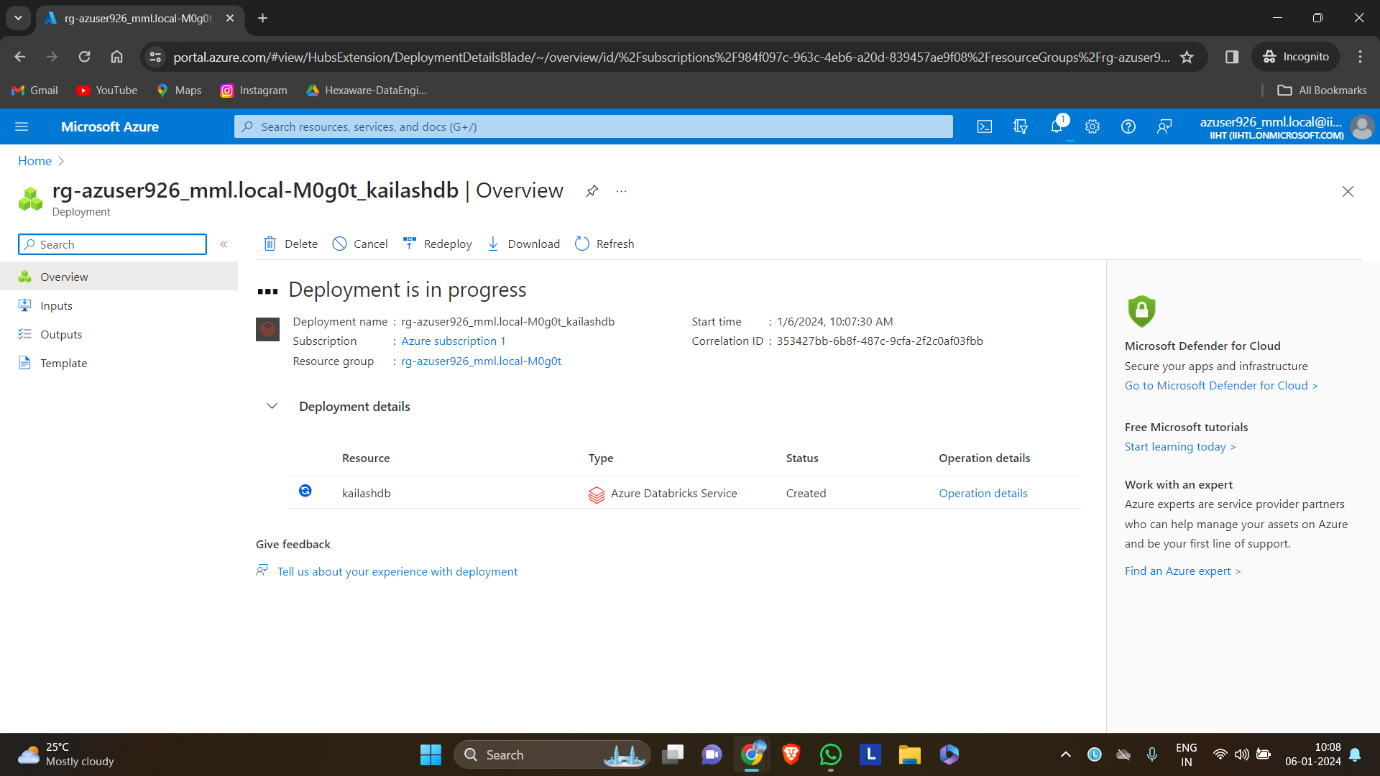
**QUESTIONS:**

1.Create a cluster &attach the notebook to the cluster and run all commands in the notebook&creates a DataFrame from a Databricks dataset&Create a Visualizations in Databricks notebooks  
&rename, duplicate, or remove a visualization or data profile.  
2. Explain the copy activity in Azure data factory.

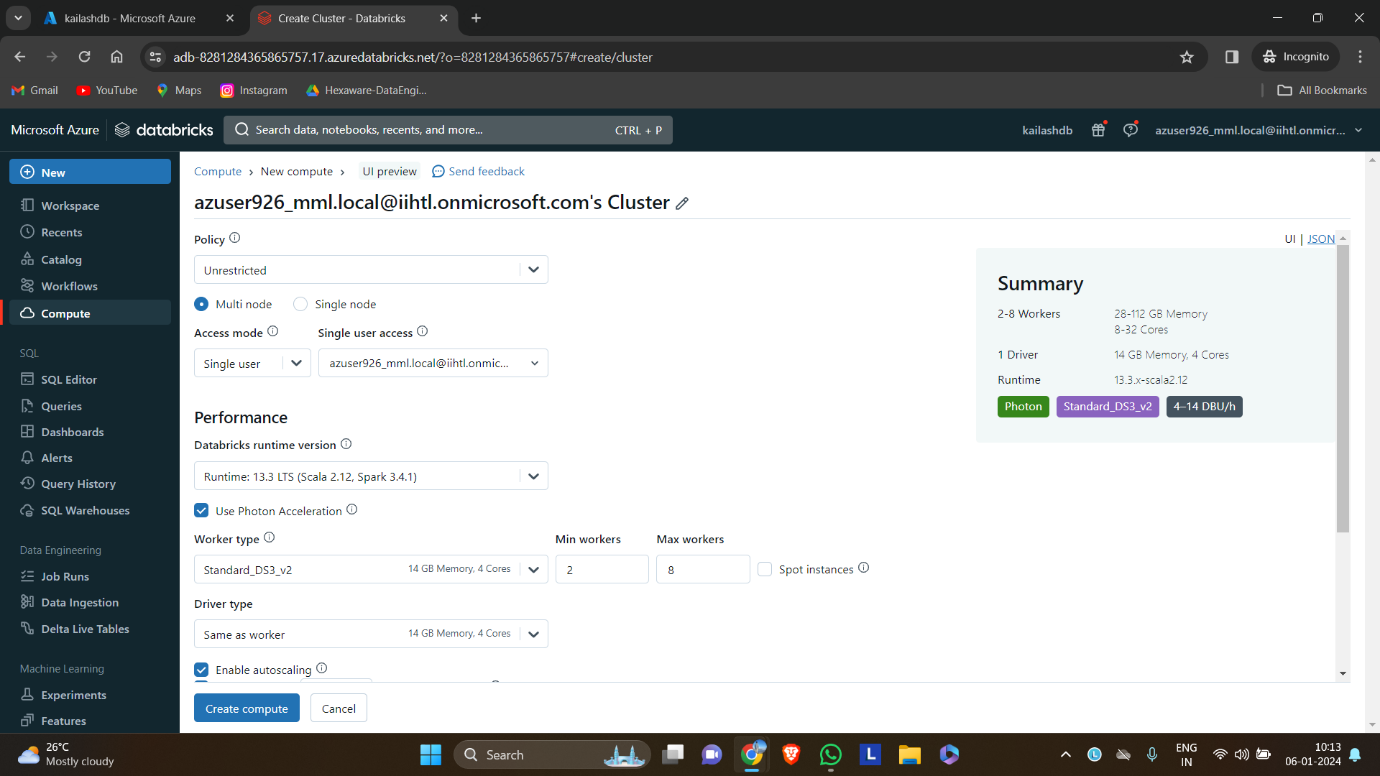
**SOLUTION:**

**1.**

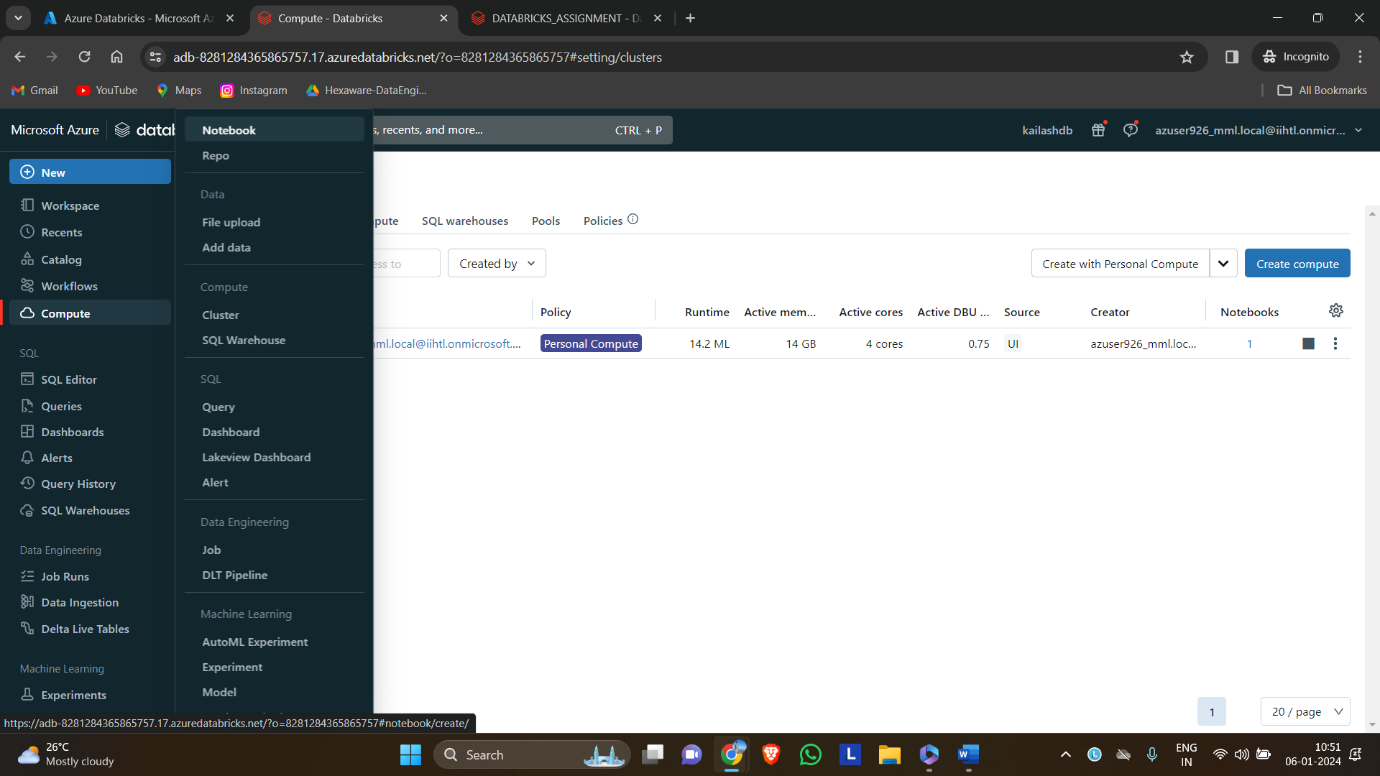
**STEP-1:** Create an azure data bricks workspace**.**

****

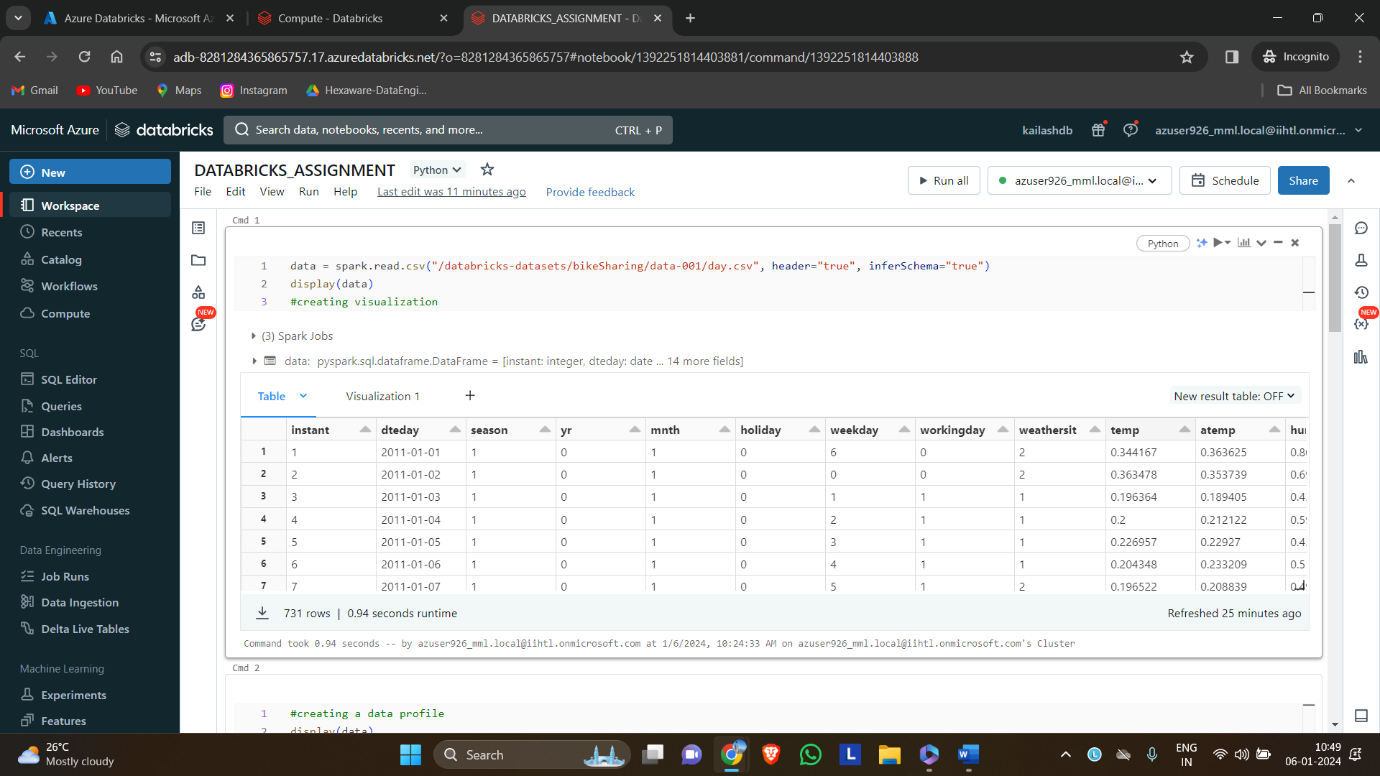
**STEP-2:** Select compute from the left side menu bar and select create compute to create a new cluster.

****

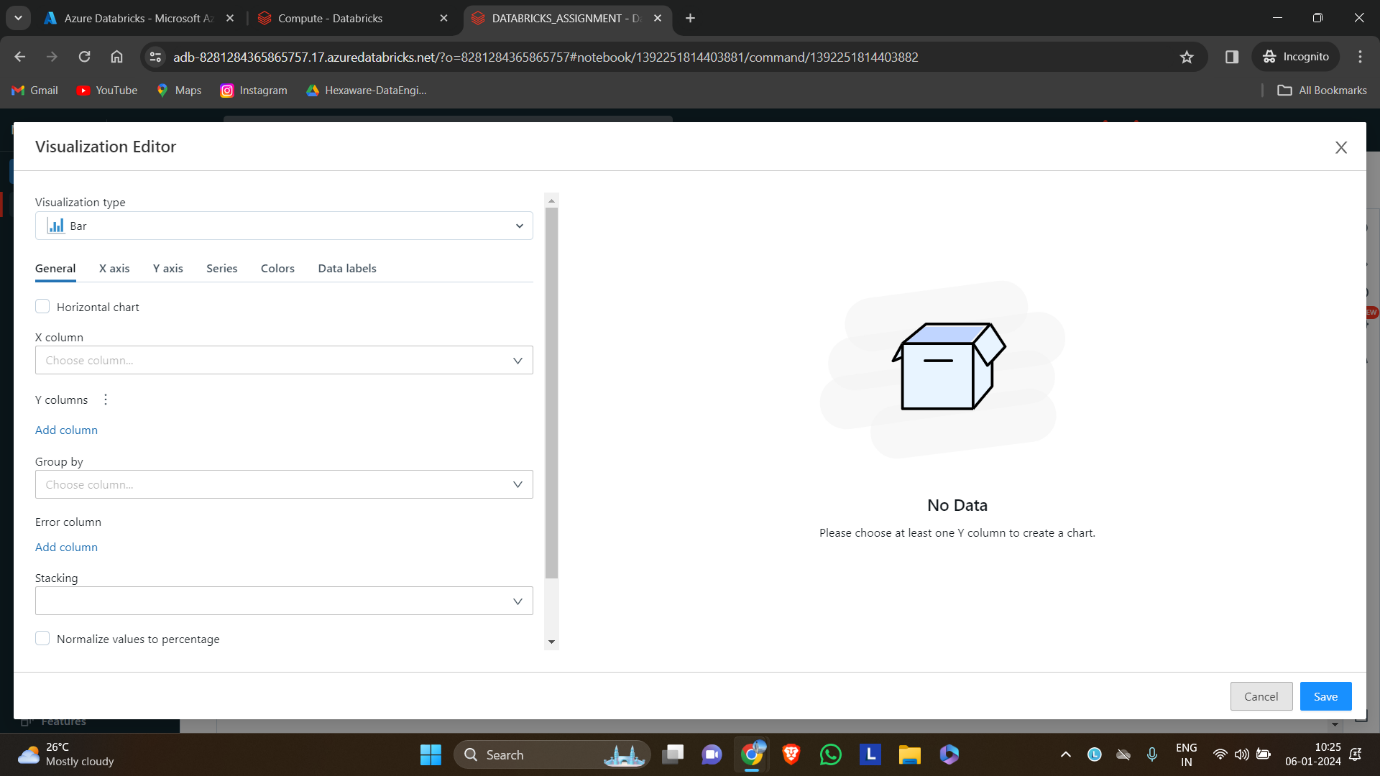
**Step-3:** After creating the cluster, create notebook. Select new in the menu bar and create a new notebook.



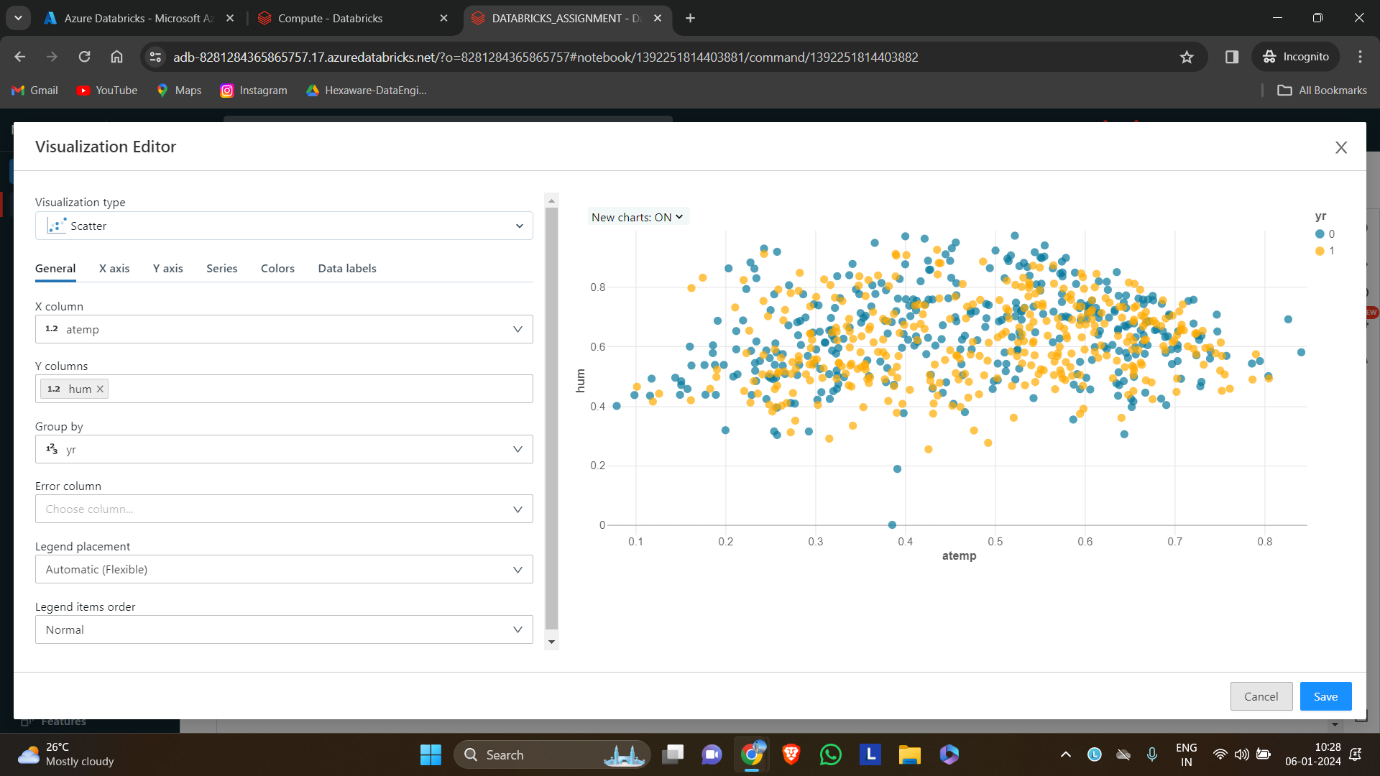
**Step-4:** Creating a new dataframe from the databricks dataset.

****

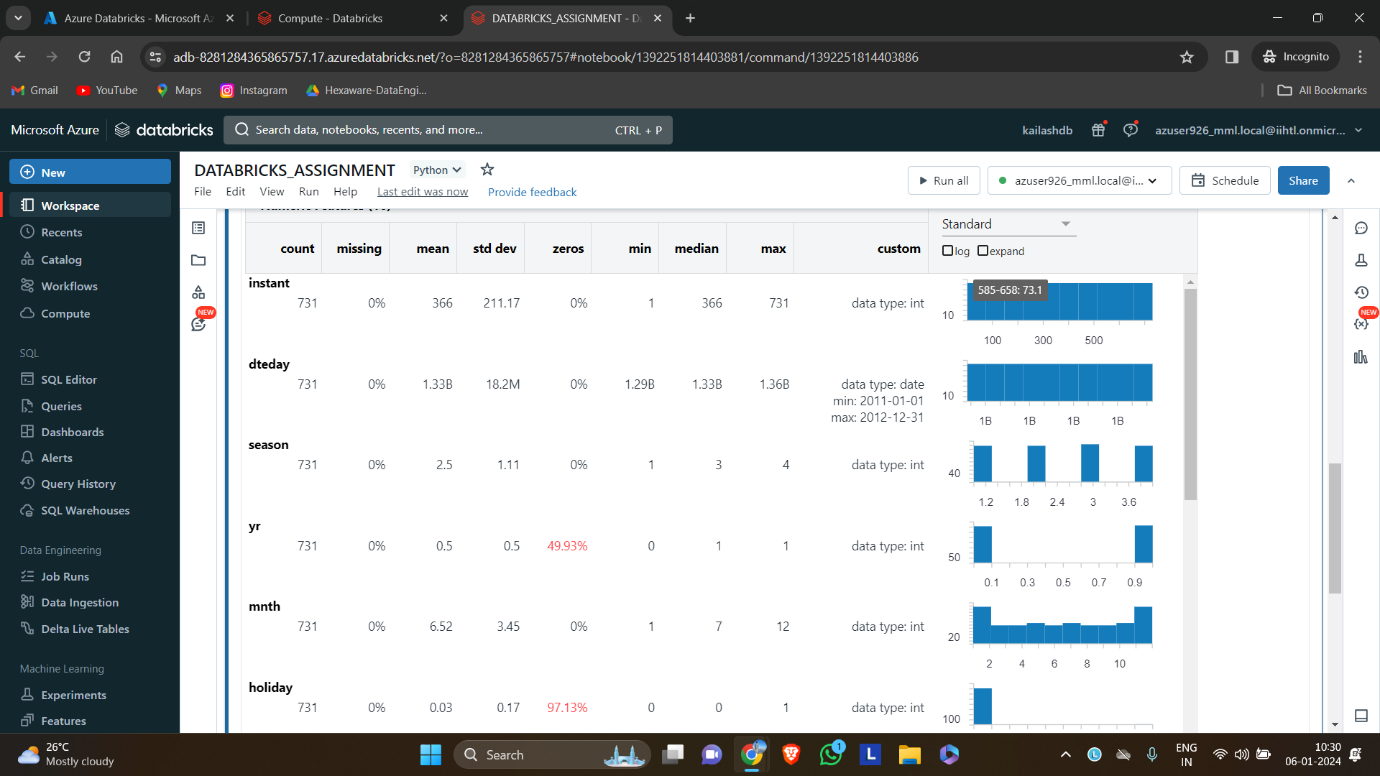
**Step-5:** After viewing the table, we can create visualization by selecting the visualization tab near the table.

****

**Step-6:** We can select various types of visualizations in the visualization editor. Here scatter plot is selected.

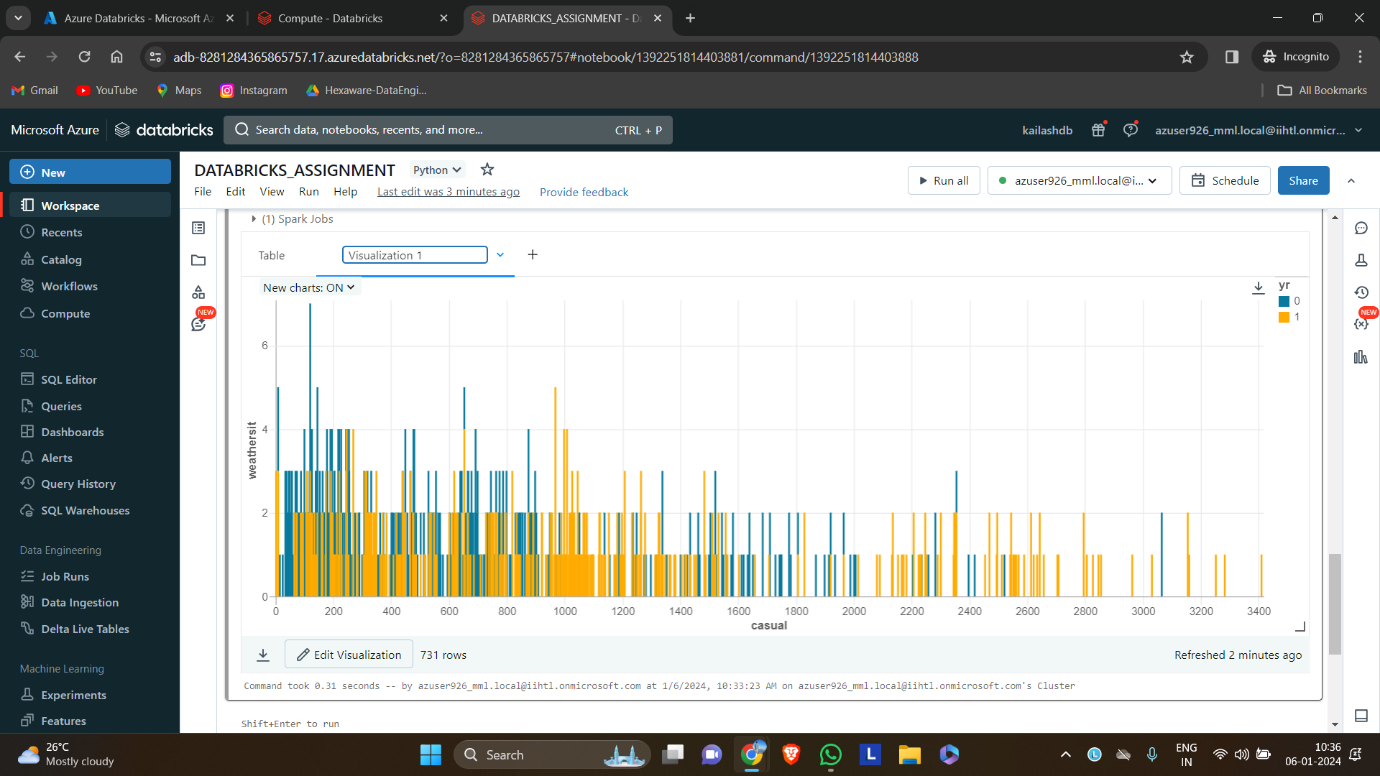
****

**Step-7:** In similar to the visualization, we can create the data profile of the table by clicking the + button near the tabs.

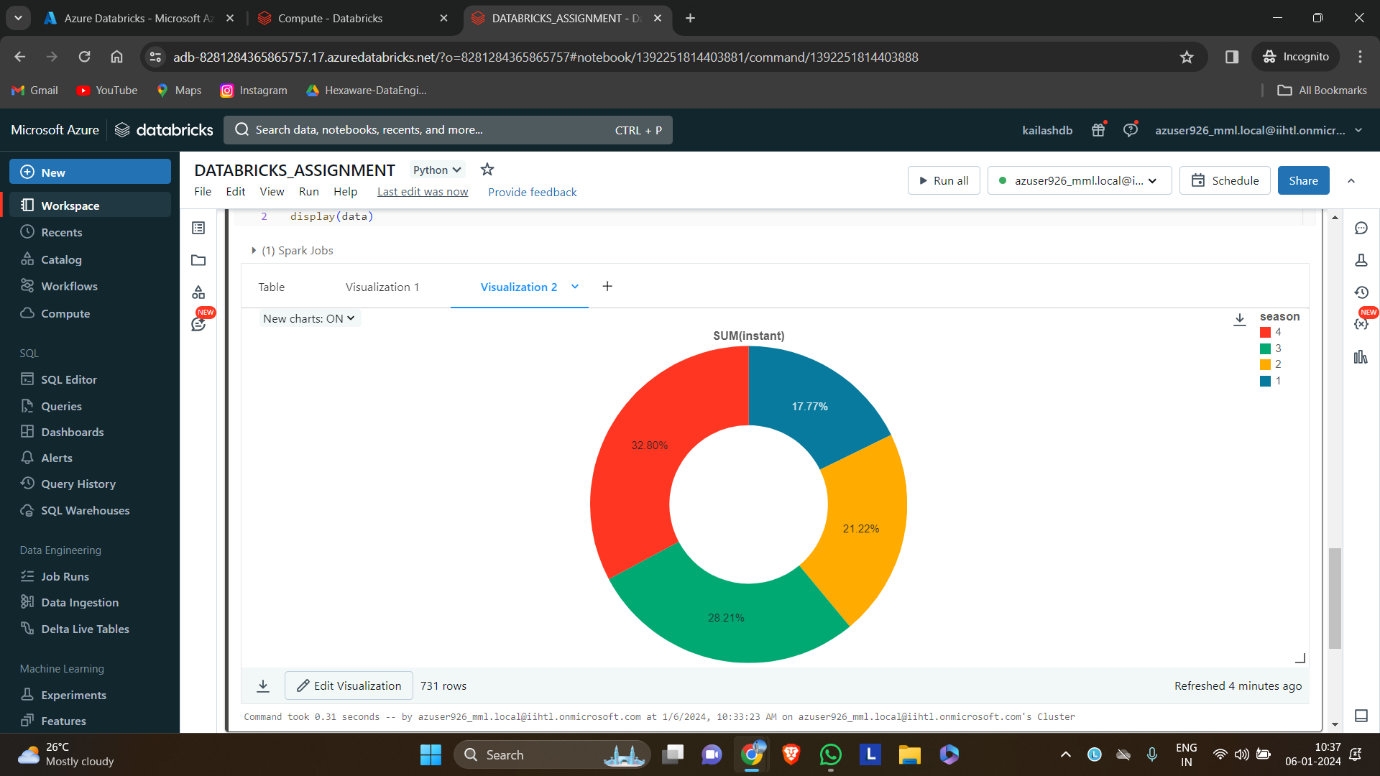
****

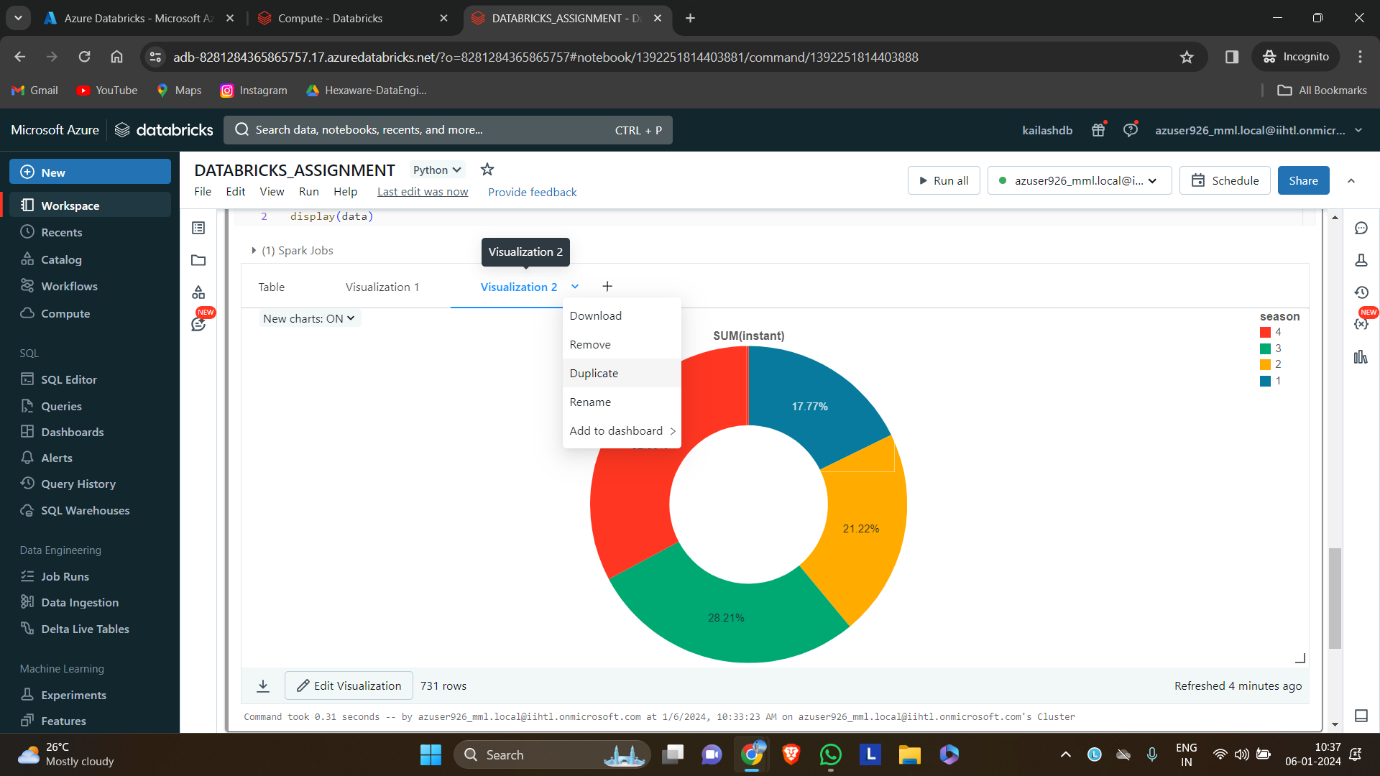
**Step-8:** By selecting the dropdown menu in the tabs we RENAME,REMOVE OR DUPLICATE the visuals or data profile.

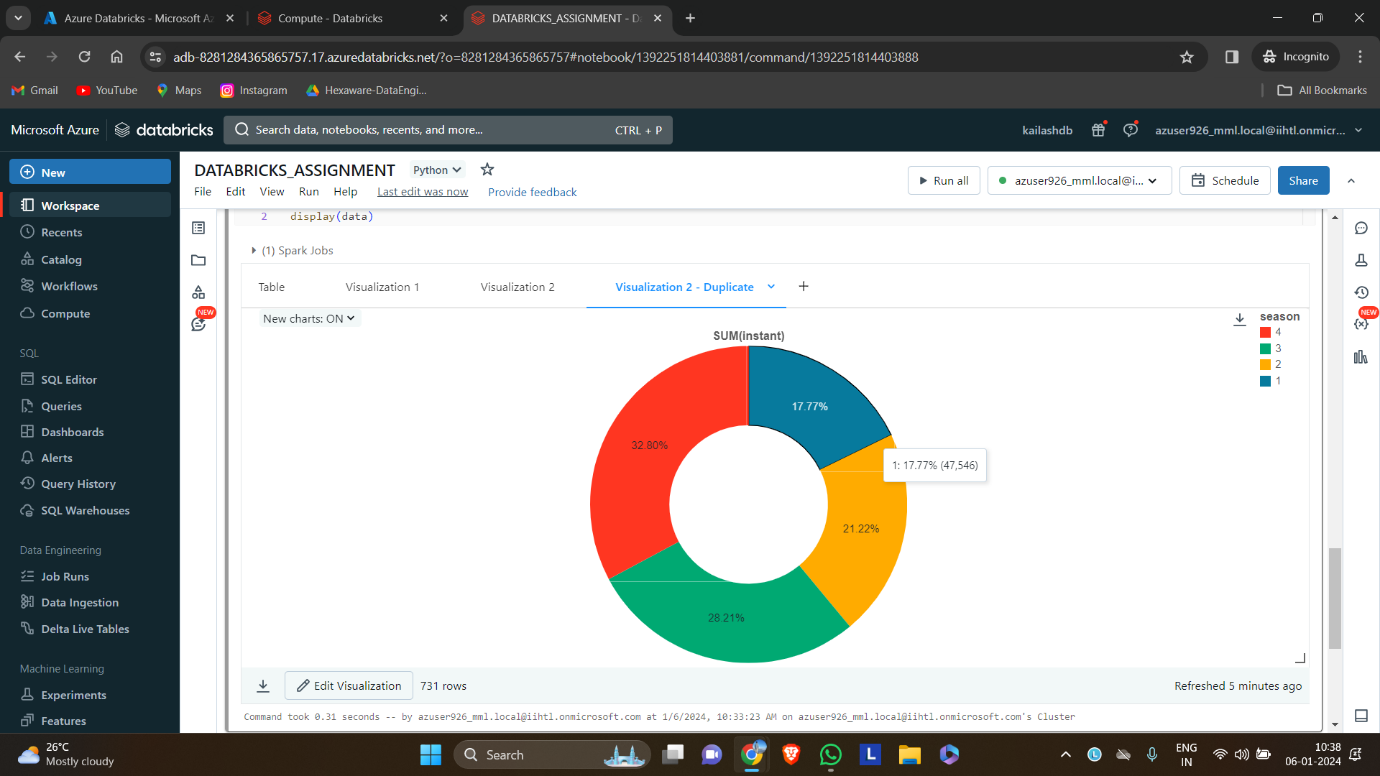
**RENAME:**

****

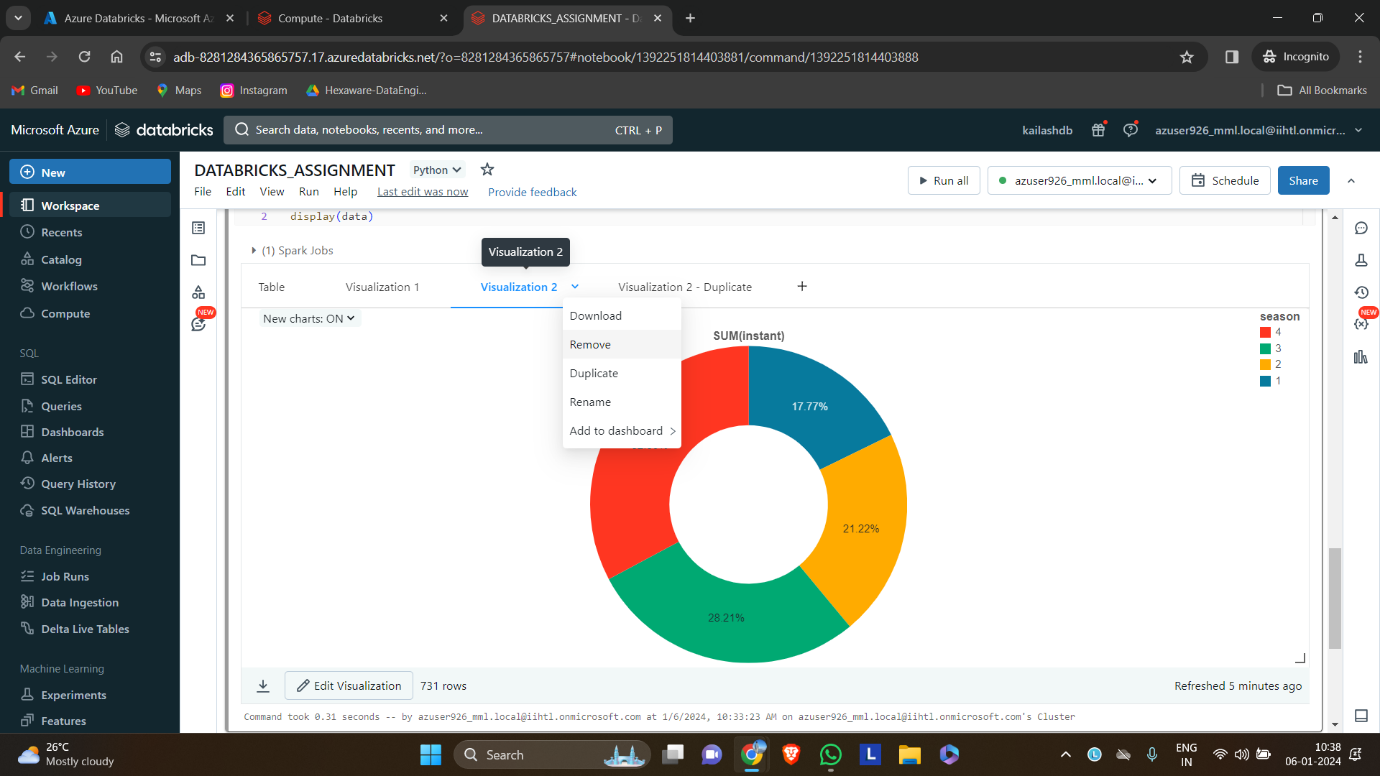
**DUPLICATE:**

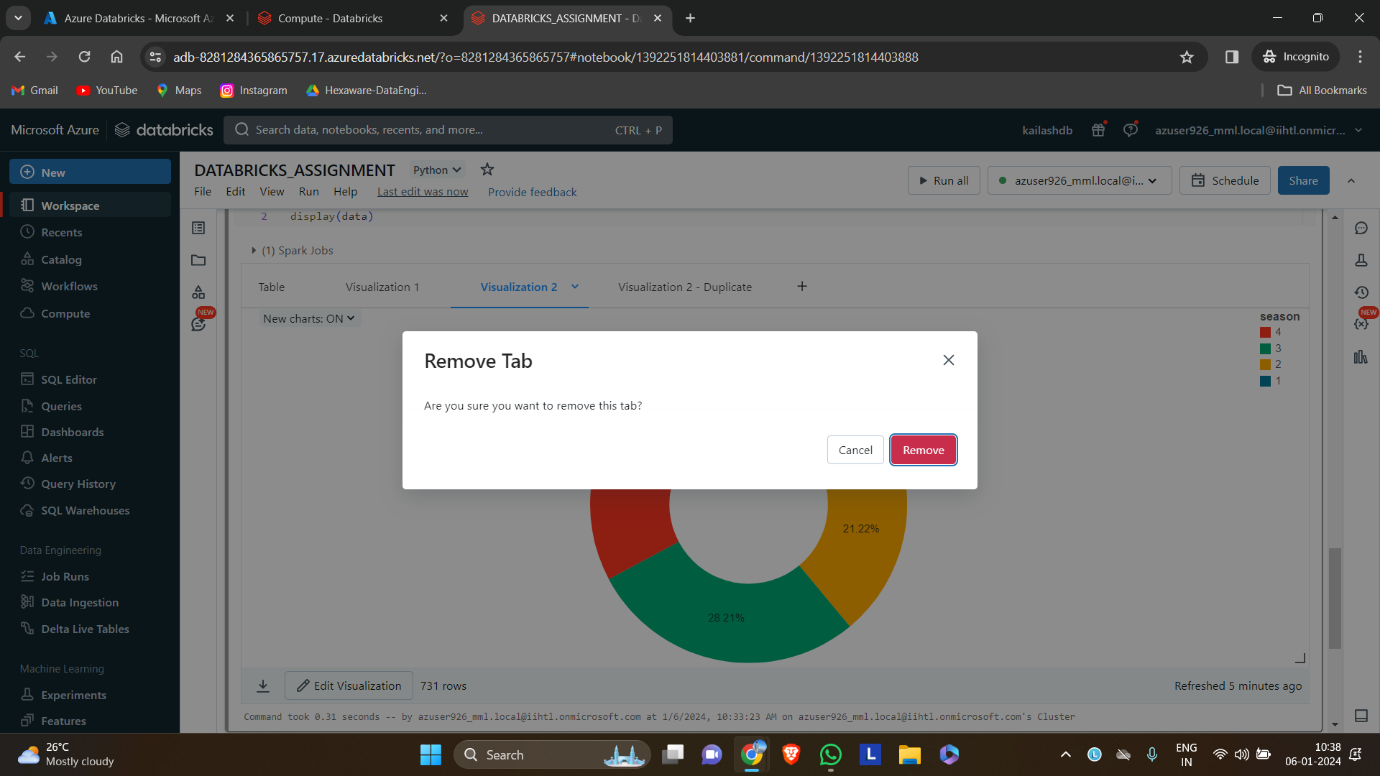
****

****

****

**REMOVE:**

****

****

**2. Copy activity in Azure data factory.**

**Azure data factory:**

Azure Data Factory is a cloud-based data integration service that allows you to create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation. It does not store any data itself. It also allows you to monitor and manage workflows using both programmatic and UI mechanisms.

**Azure Storage Account:**

An Azure Storage Account is a secure account, which provides you access to services in Azure Storage. The storage account is like an administrative container, and within that, we can have several services like blobs, files, queues, tables, disks, etc. And when we create a storage account in Azure, we will get the unique namespace for our storage resources.

**Requirements:**

* Azure cloud account
* Azure storage account (2)
* Azure data factory

**PROCEDURE:**

**STEP 1:** create a two azure storage accounts in the azure portal. During creation of the storage accounts make one as datalake gen2 storage.

**STEP 2:** Next create Azure data factory studio. After creating the data factory, launch the studio.

**STEP 3:** In the studio select, Ingestion tab.

**STEP 4:** Select built in copy task and give the time schedule for running the pipeline.

**STEP 5:** After that in the source data store, select the storage account to be copied and create a connection with the pipeline. Also select the folder to be copied.

**STEP 6:** Same like above, Select the destination storage account and folder in the destination data store and create a connection with the pipeline

**STEP 7:** After reviewing the details, the copy activity can be initiated.

**STEP 8:** Now we can see the copied files in the new storage account.

**STEP 9:** We can check the status of the pipeline in the factory resources.