|  |  |
| --- | --- |
| Name | Enayathulla A |
| Date | 06/01/2024 |
| Type | Weekly Coding Test |
| Topic | Q1) Cluster creation, running commands on  notebook, Data frame creation, Creating  visualization & rename, duplicate, or remove a visualization or data profile.  Q2) Copying data on Azure Data Factory. |

**Q1) 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**

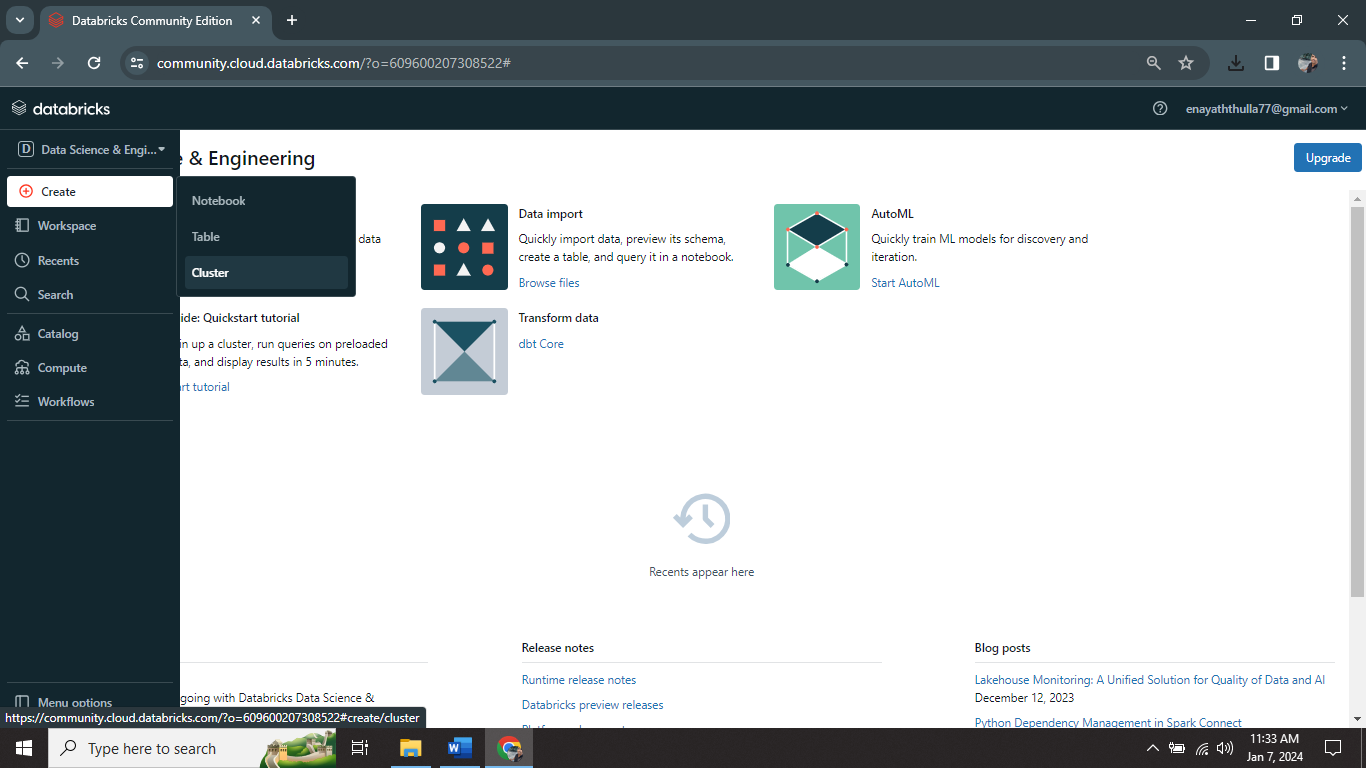
**Databricks notebooks & Rename, duplicate, or remove a visualization or data profile.**

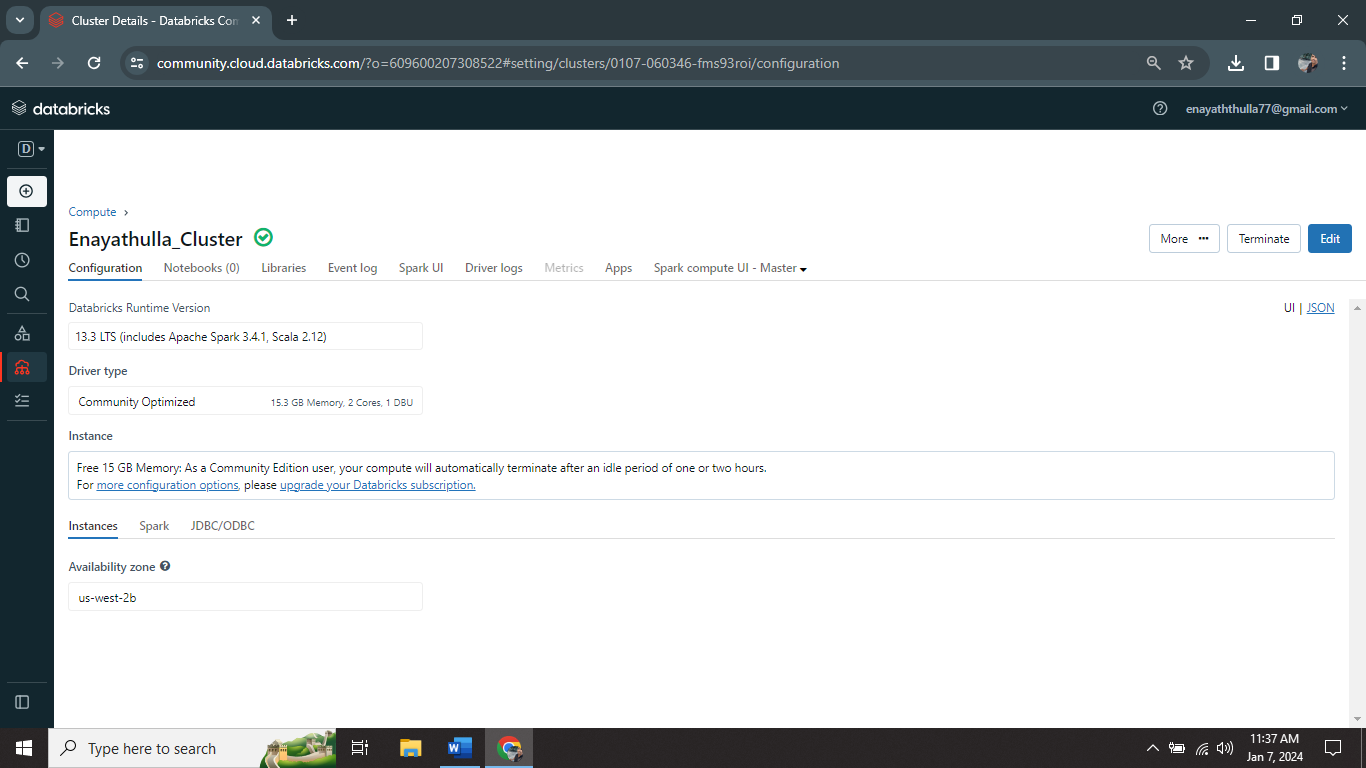
**Ans:**

**a) Create a cluster & Attach the notebook to the cluster and run all commands in the**

**Notebook.**

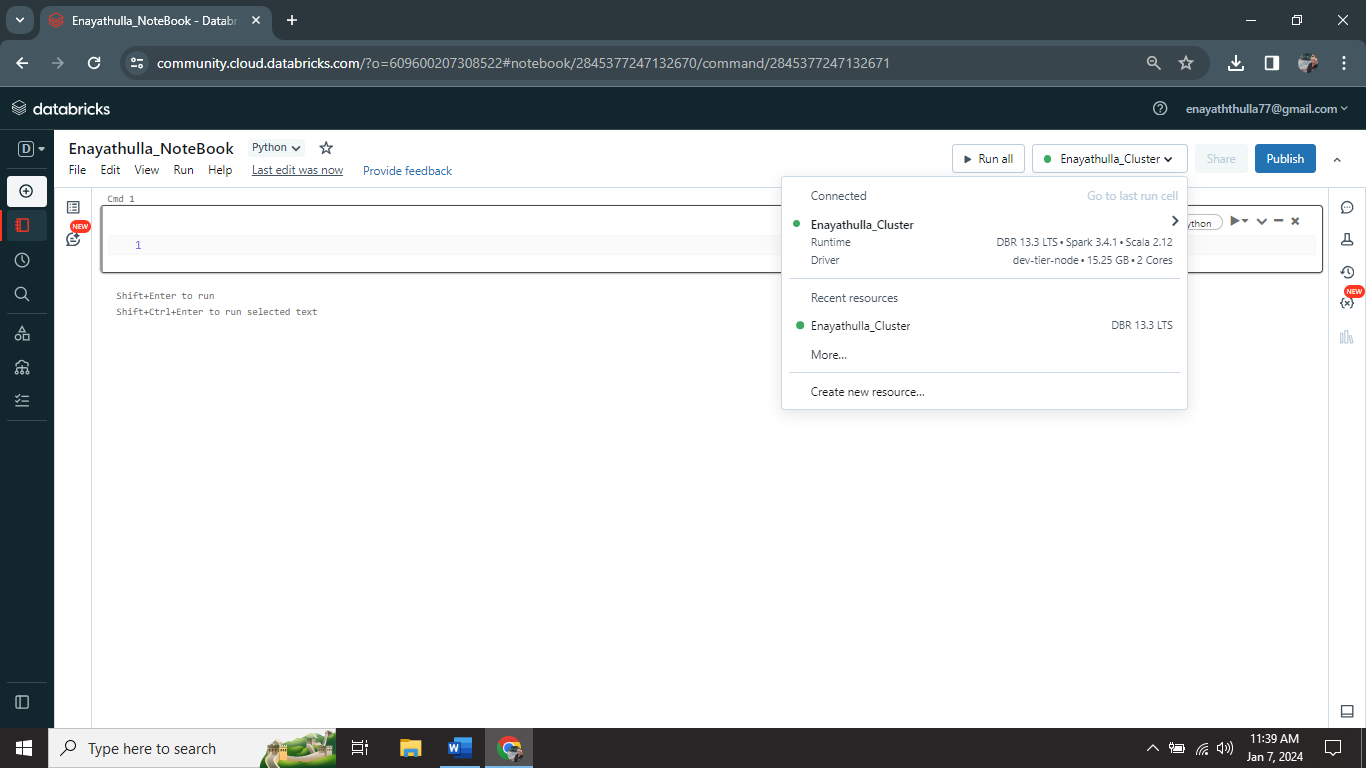
**Cluster Creation:**



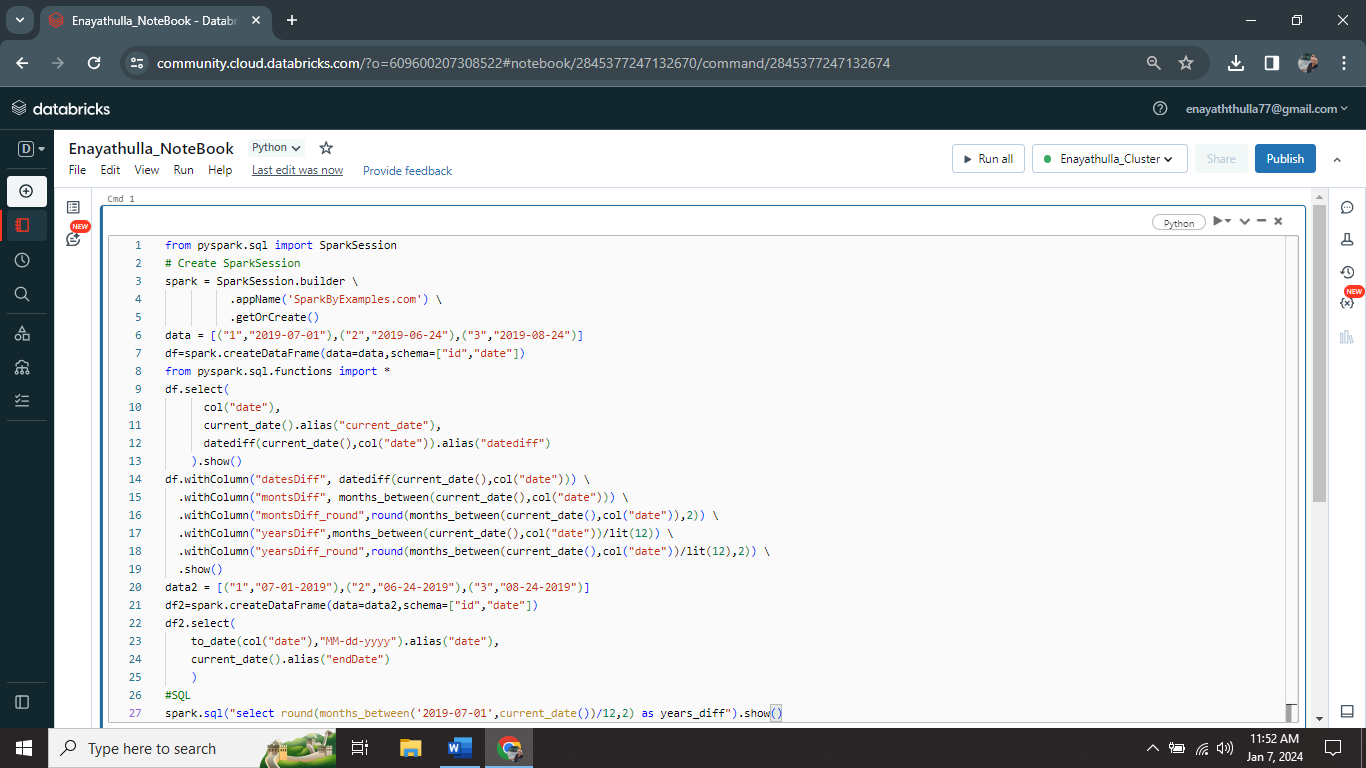


I created one cluster (Enayathulla\_Cluster) in Databricks.

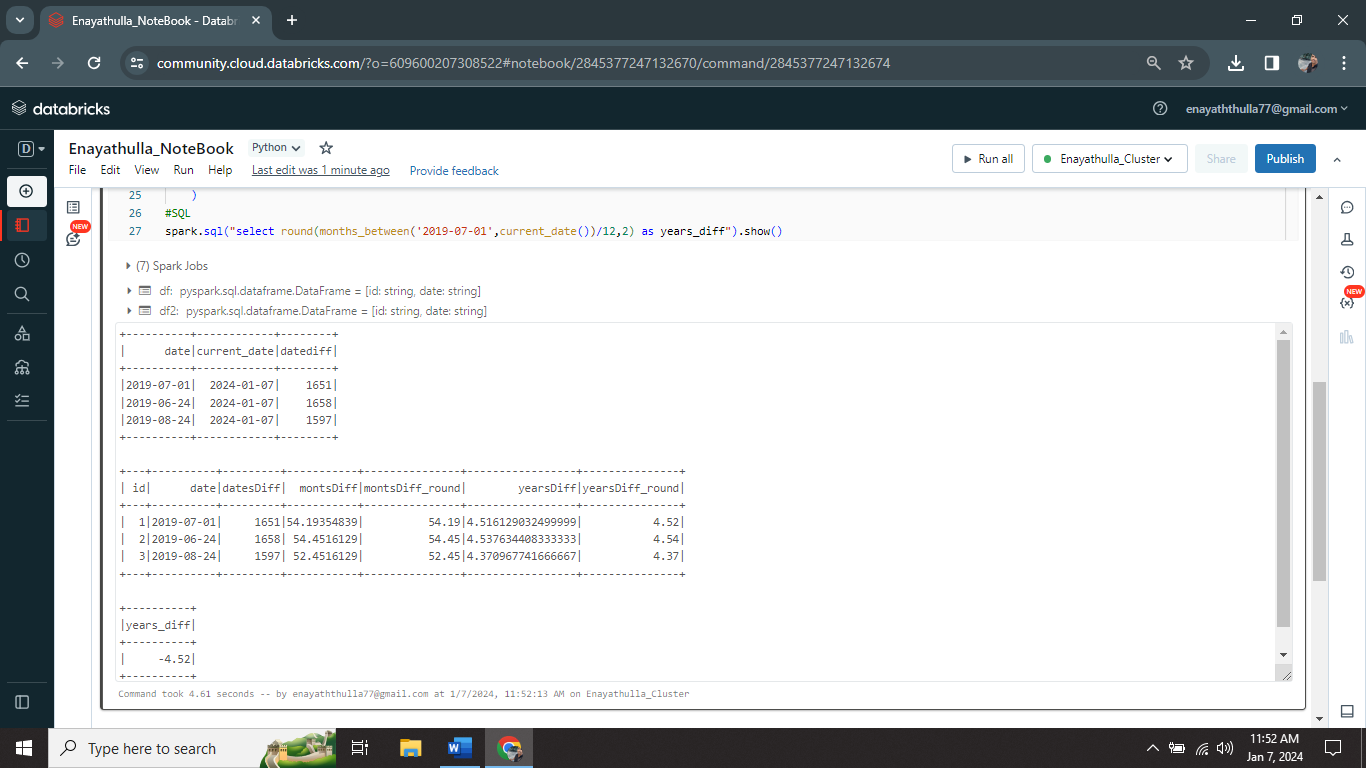
**NoteBook creation:**



I created one notebook(Enayathulla\_NoteBook) on Enayathulla\_Cluster.

**Code run:**

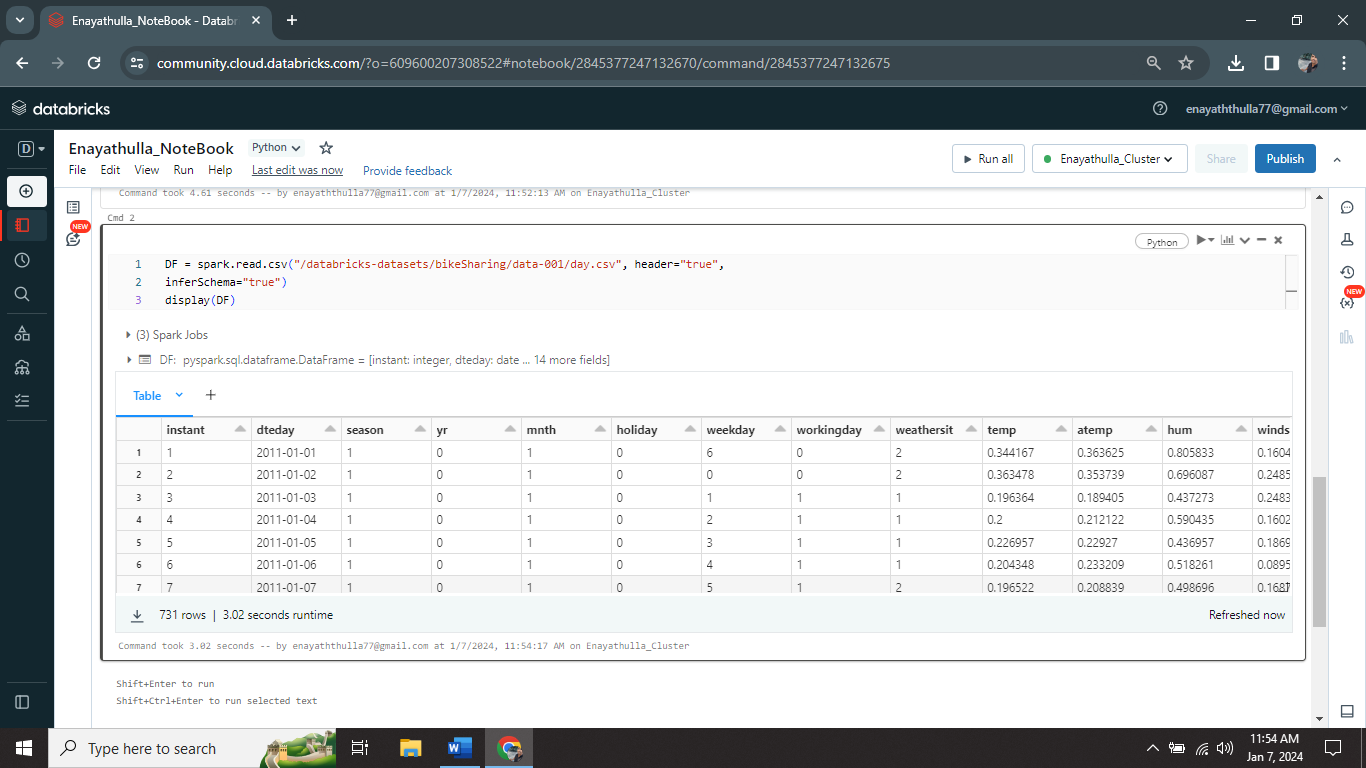
I successfully ran one sample code on my notebook.



This is the output I got from sample code.

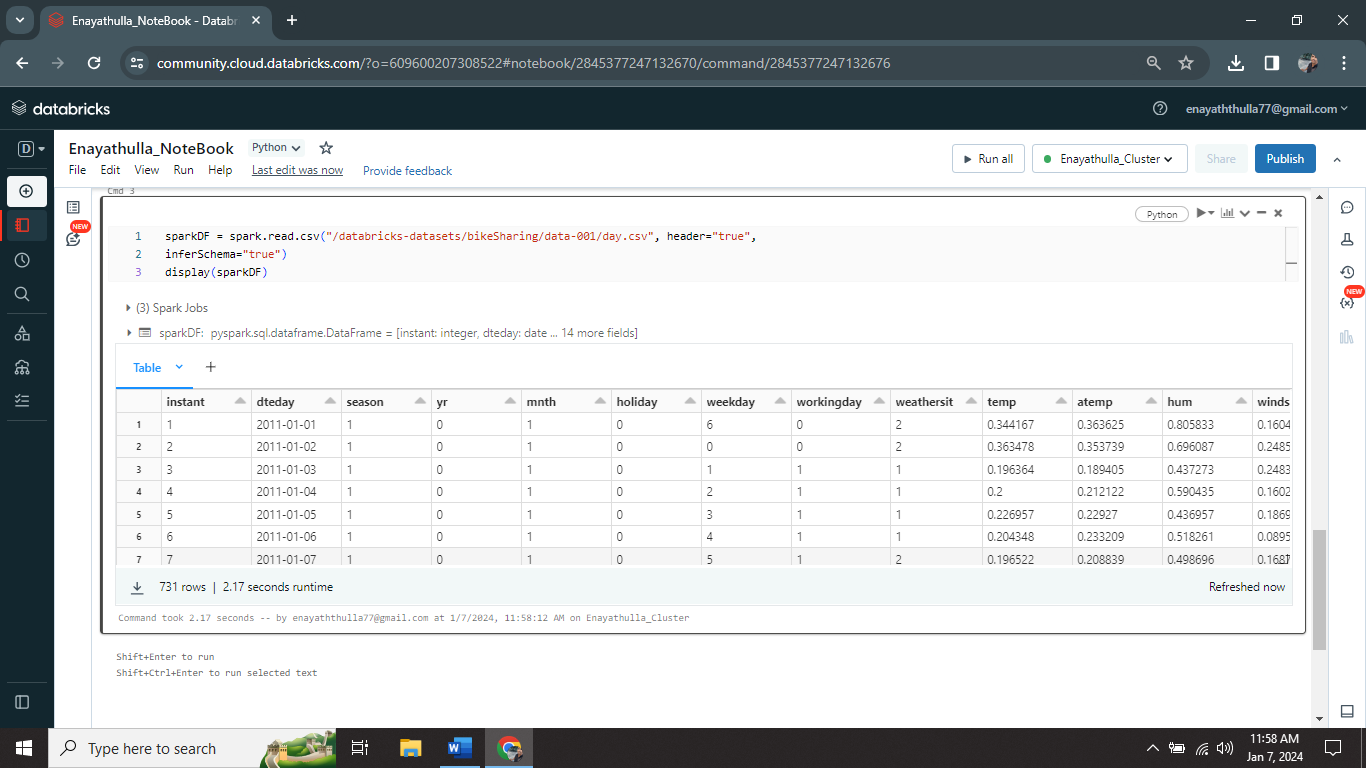
**b) creates a DataFrame from a Databricks dataset.**

**Creation of DataFrame:**

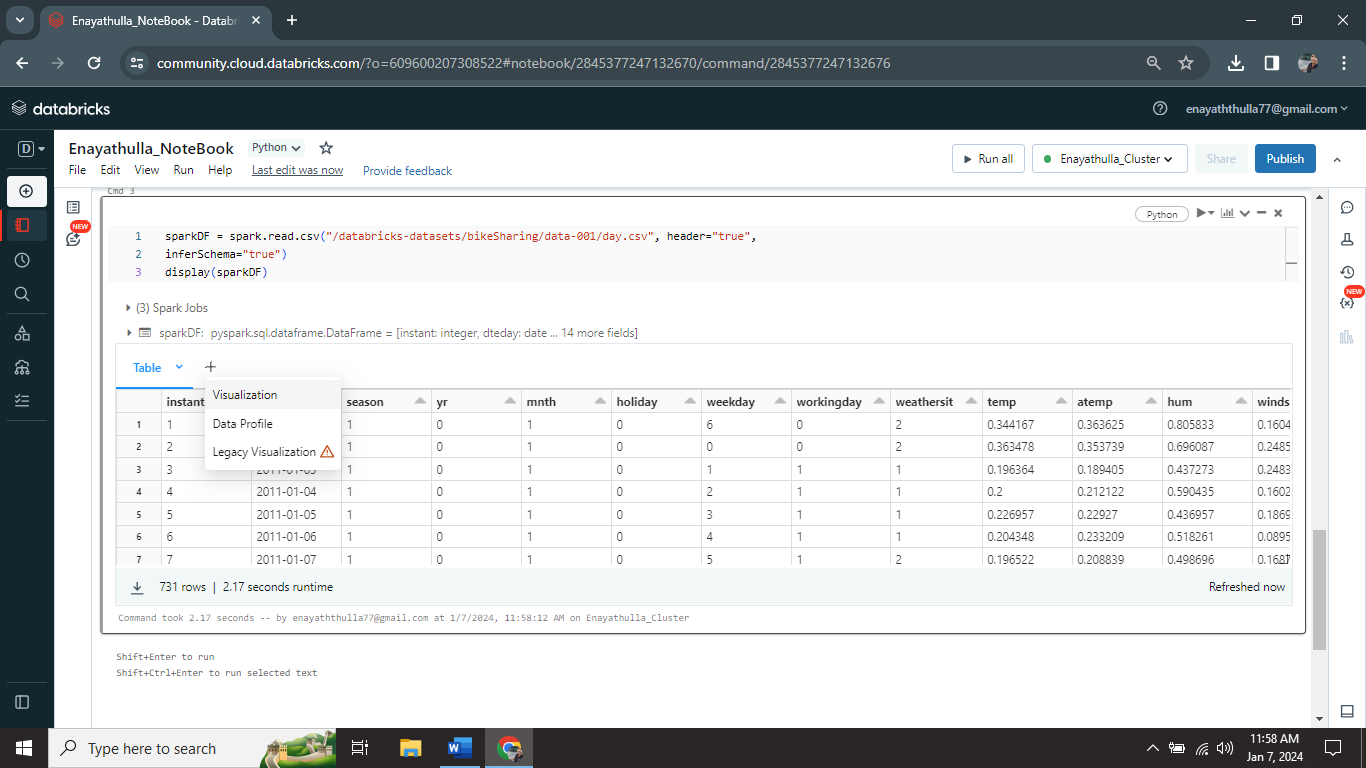


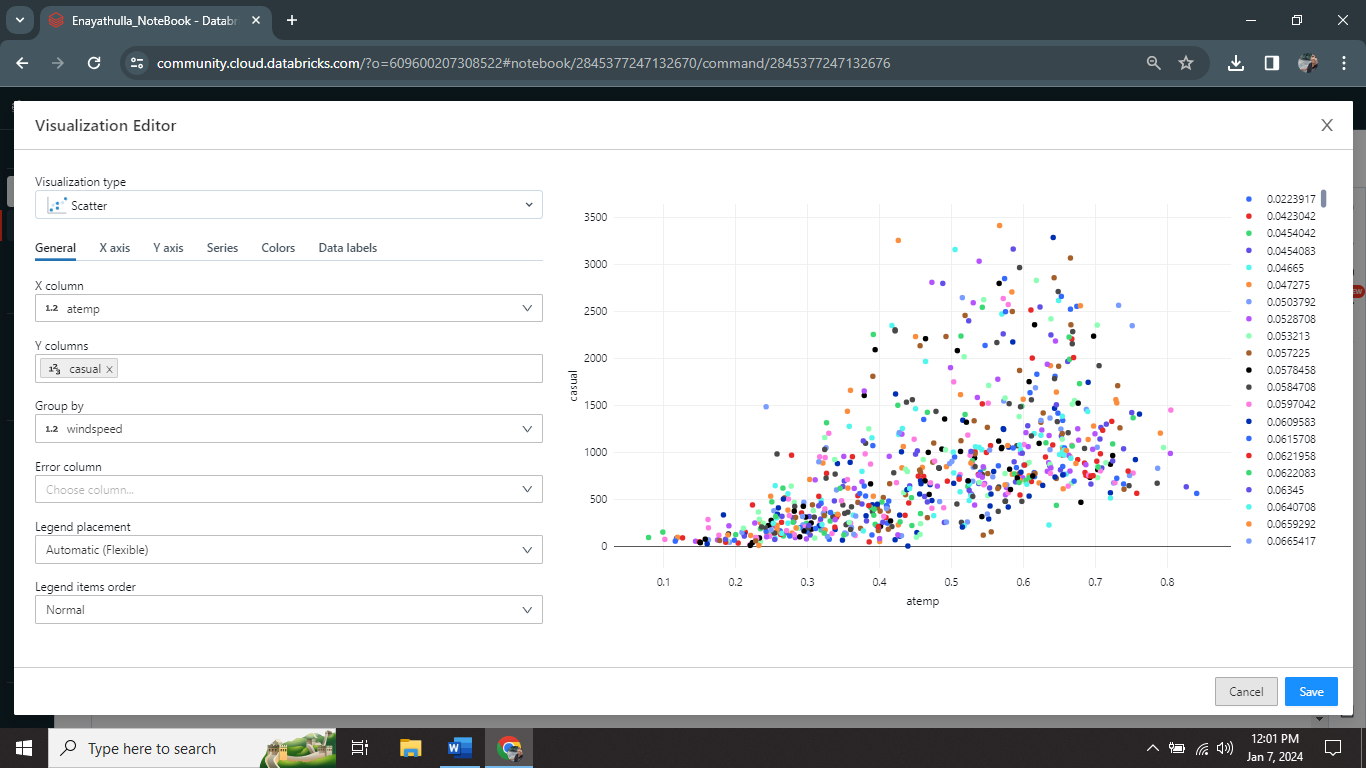
This is the sample DataFrame (DF).

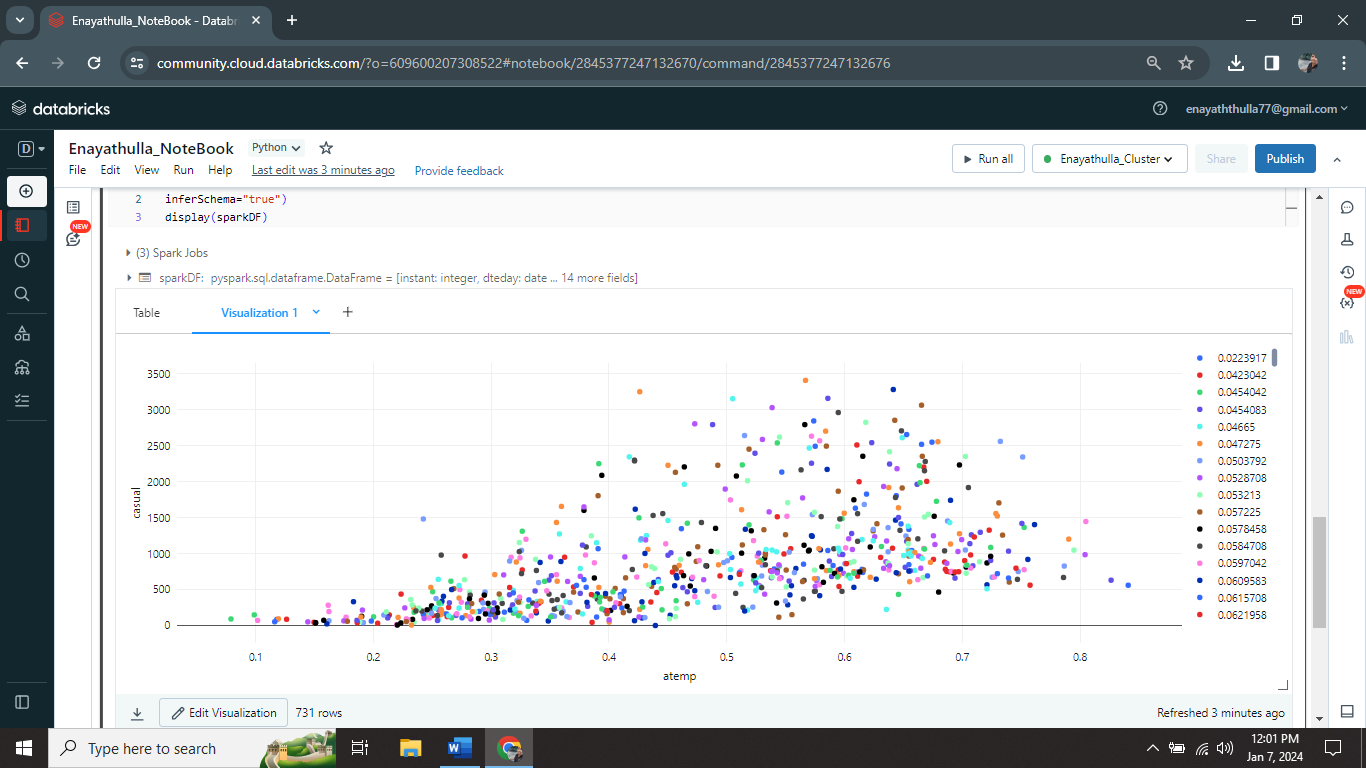
**c) Create a Visualizations in Databricks notebooks & Rename, duplicate, or remove a visualization or data profile.**



**Creation of Visualization:**

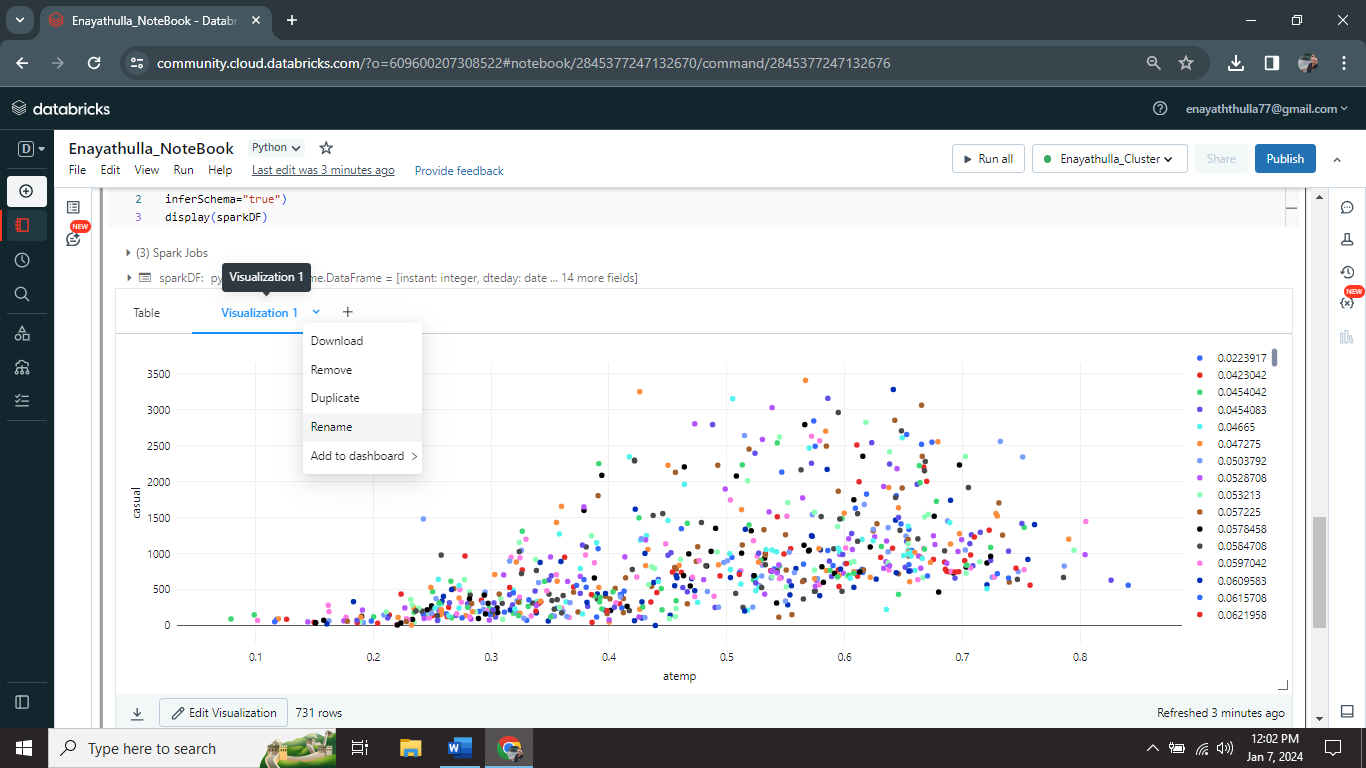




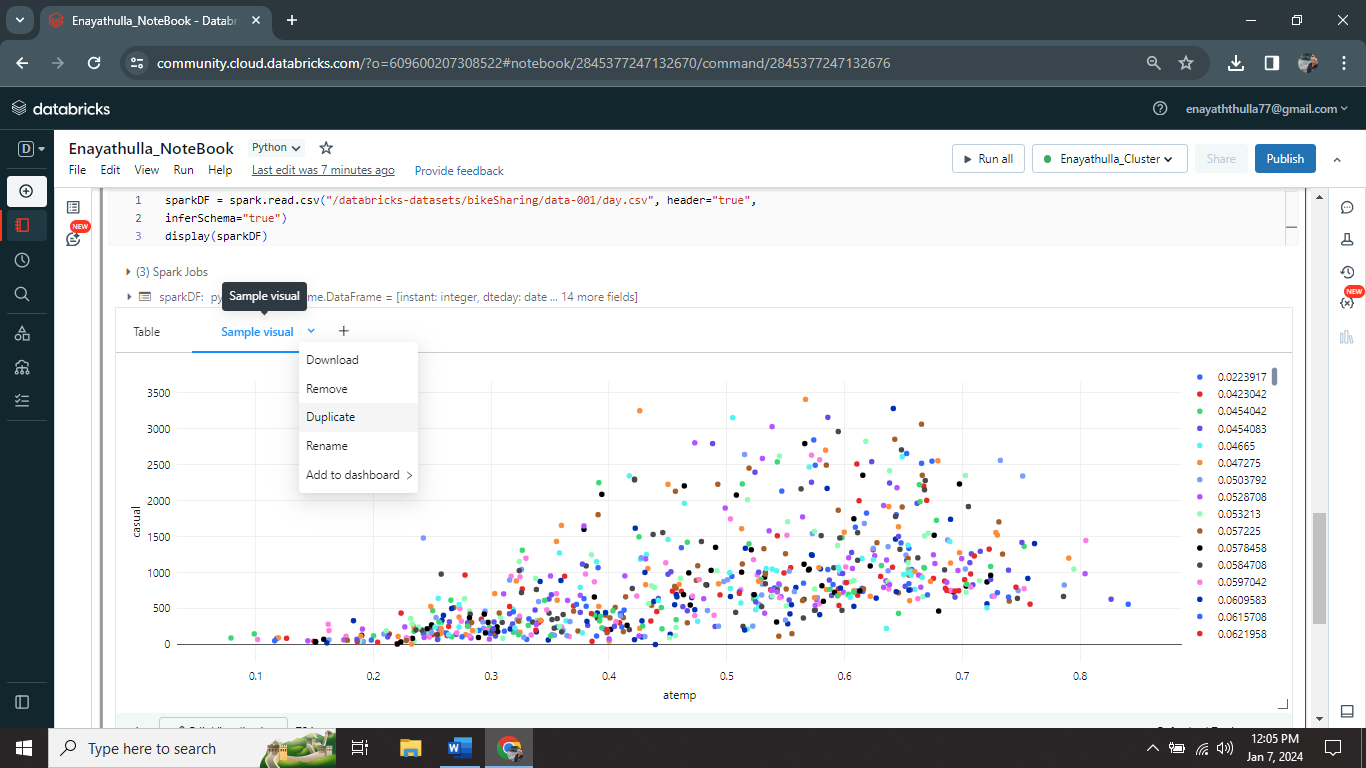


I successfully created Visualization 1.

**Rename of Virtualization:**

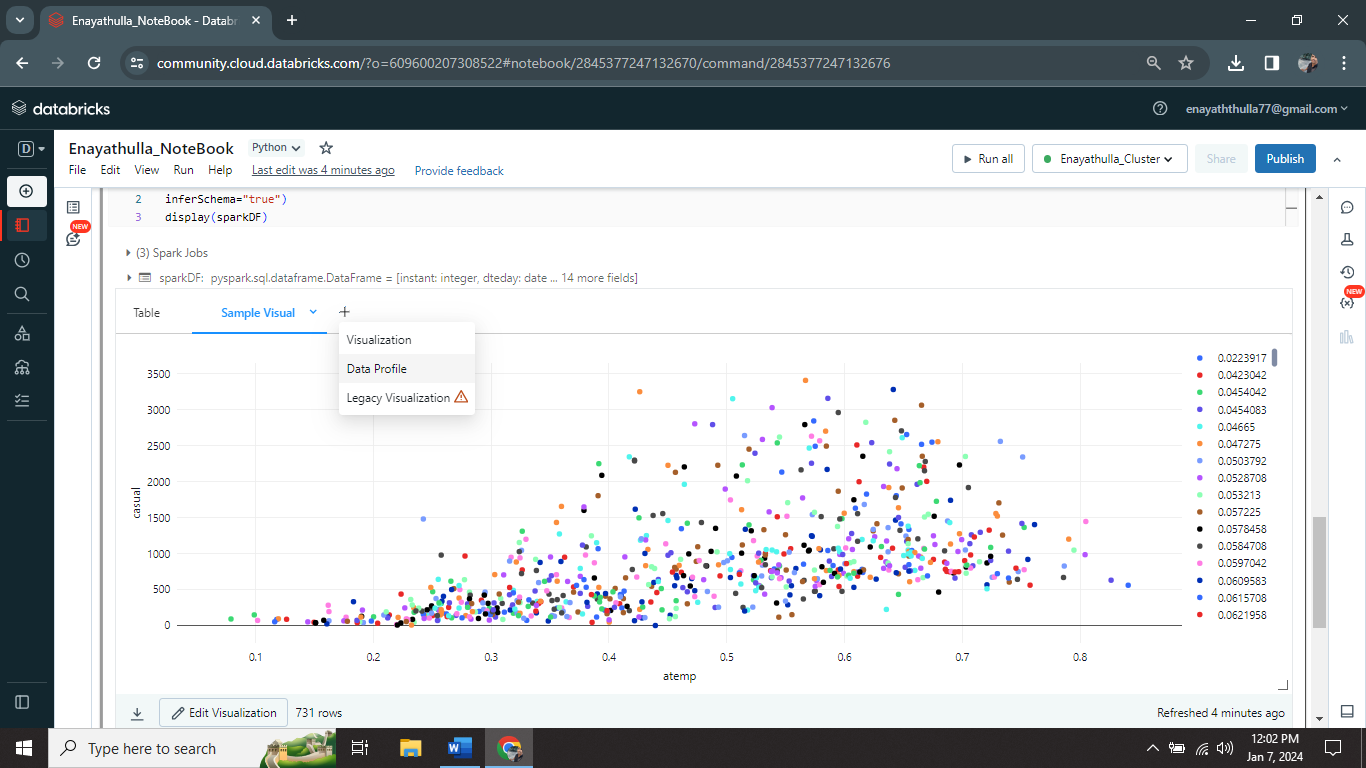


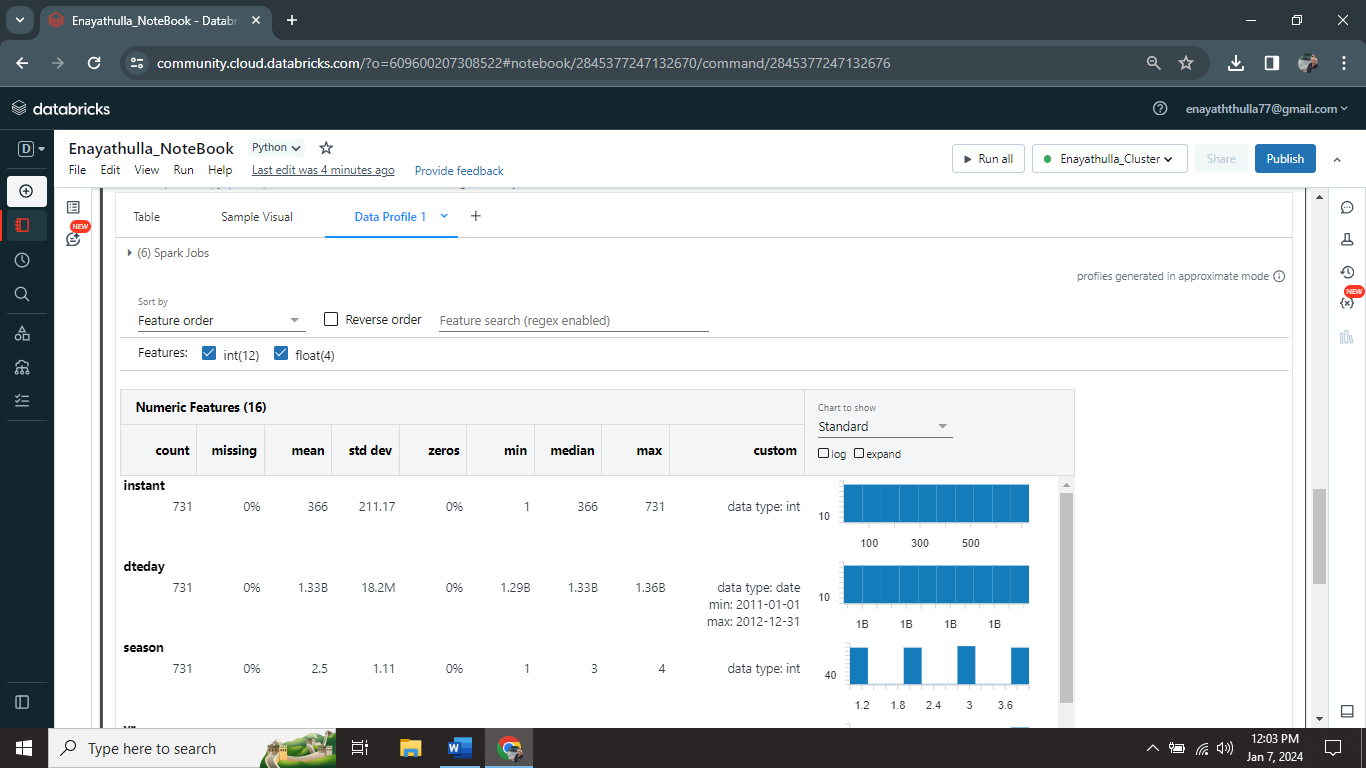
I renamed visualization 1 to Sample Visual using Rename option.

**Duplicate of Virtualization:**

I successfully duplicate my visualization using Duplicate tab.

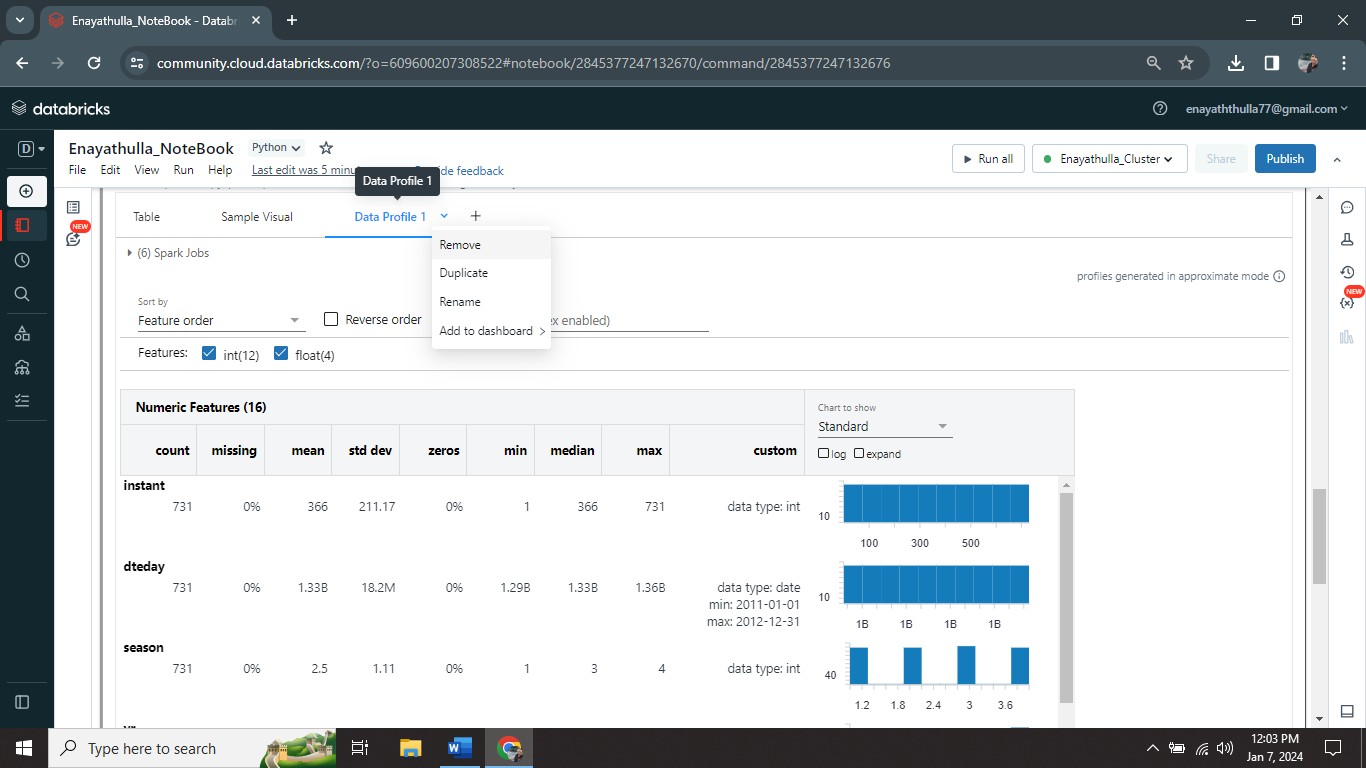
**Creation of Data Profile:**



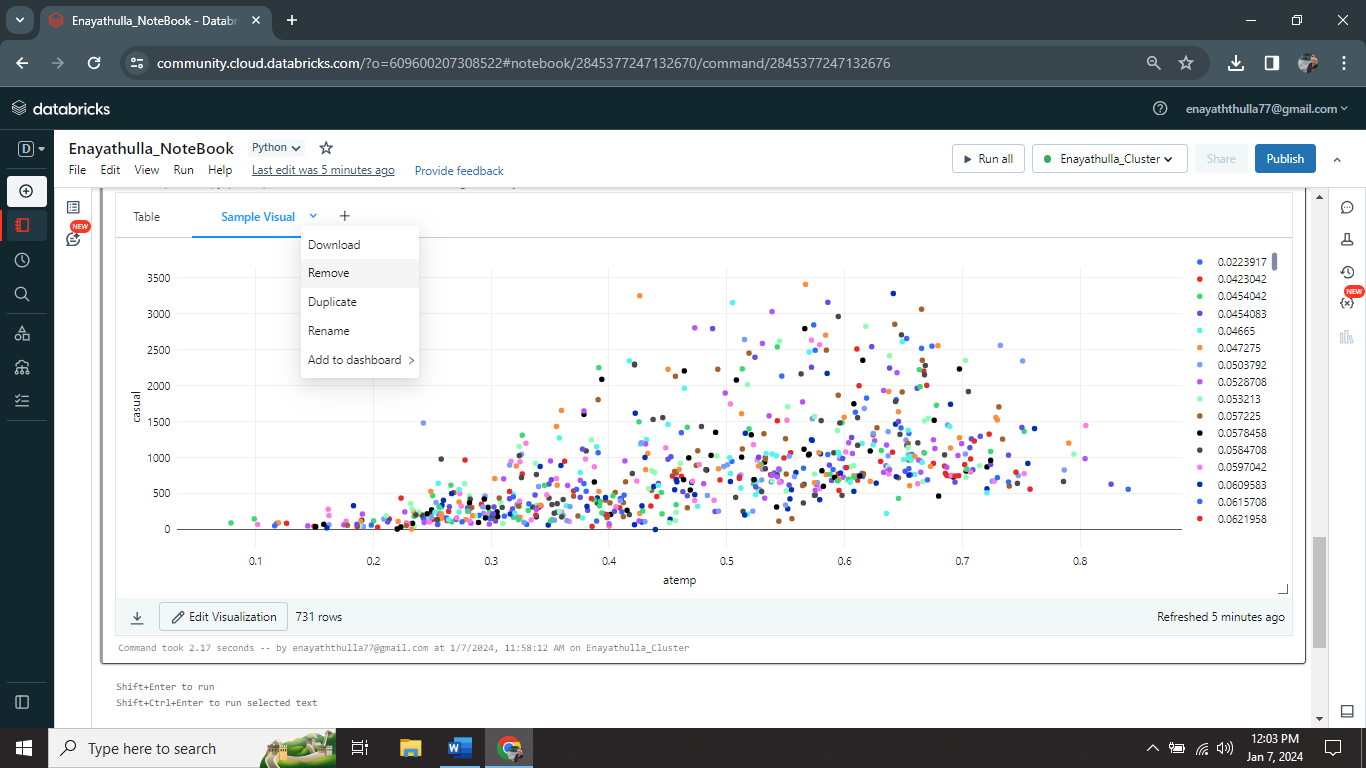


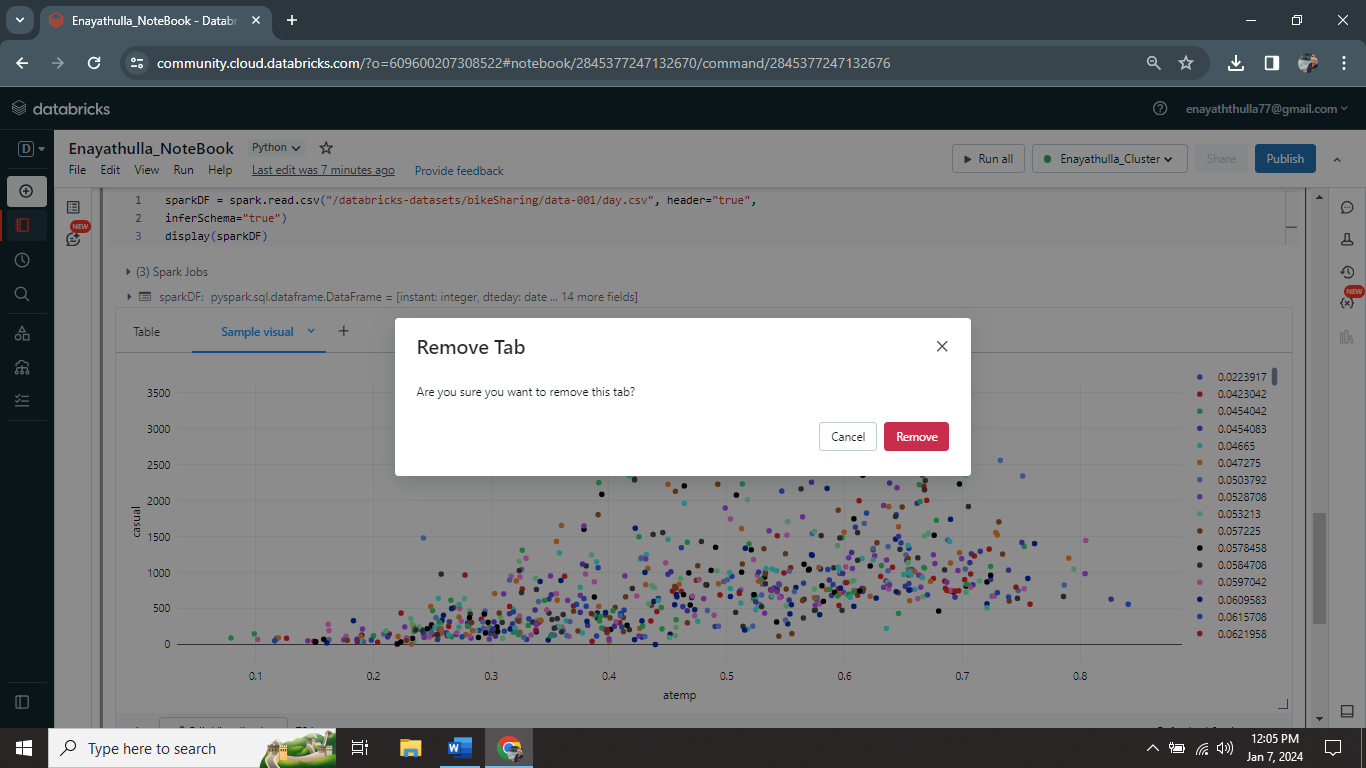
I successfully created Data Profile 1.

**Remove Data Profile:**



I removed Data Profile using Remove option.

**Remove Visualization:**



I successfully removed visualization using remove option.

**Q2) Explain the copy activity in Azure data factory.**

**Ans:**

**Copy Activity:**

The Copy Activity is a key element of Azure Data Factory (ADF) that facilitates data movement from a source to a destination. It is an essential component of building ADF data pipelines that facilitate the extraction, transformation, and loading (ETL) of many data sources.

**Azure Data Factory:**

* Azure Data Factory is a cloud-based data integration service provided by Microsoft Azure.
* It helps organizations to collect, transform, and move their data from various sources to different destinations.

**Important Concepts of Azure Data Factory:**

**Data Pipelines:**

A pipeline is a set of data-driven activities that can perform various operations on your data.

**Data Sources:**

It supports connecting to various data sources, such as databases, on-premises systems, cloud services, and more.

**Data Destinations:**

It supports various data destinations where your data will be stored or processed, such as databases, data warehouses, and even other cloud services.

**Data Transformations:**

we can be able apply transformations to our data as it moves through the pipeline. This might include cleaning, aggregating, or enriching the data to make it more useful.

**Scheduled and Triggered Execution:**

Data pipelines can be scheduled to run at specific times or triggered by events. This flexibility allows us to automate our data workflows.

**Monitoring and Management:**

Azure Data Factory provides monitoring capabilities, allowing us to track the progress of your data pipelines and identify any issues.

**Azure Data Factory Activity:**

An activity in Azure Data Factory represents a unit of work in a data pipeline.

It is a specific operation or task that is part of a larger data workflow.

Activities perform actions such as moving data, transforming data, copying data, and running external scripts or applications.

For, consider an activity as a single step or task that needs to be completed to achieve a specific goal within your data integration process. Each activity is designed to perform a specific action on your data

**Examples of Azure Data Factory Activity:**

**Data Copy Activity:**

Moves data from one location to another, like copying files from on-premises storage to Azure Blob Storage.

**SQL Activity:**

Runs a SQL script on a database to retrieve or manipulate data.