CODING CHALLENGE ASSESSMENT **(DATA BRICKS)**

**Name: Vinutha S**

**Date: 06/01/2024 (Saturday)**

**1. Create a cluster & attach the notebook to the cluster and run all commands in the notebook &**

**creates a Data Frame from a Databricks dataset &**

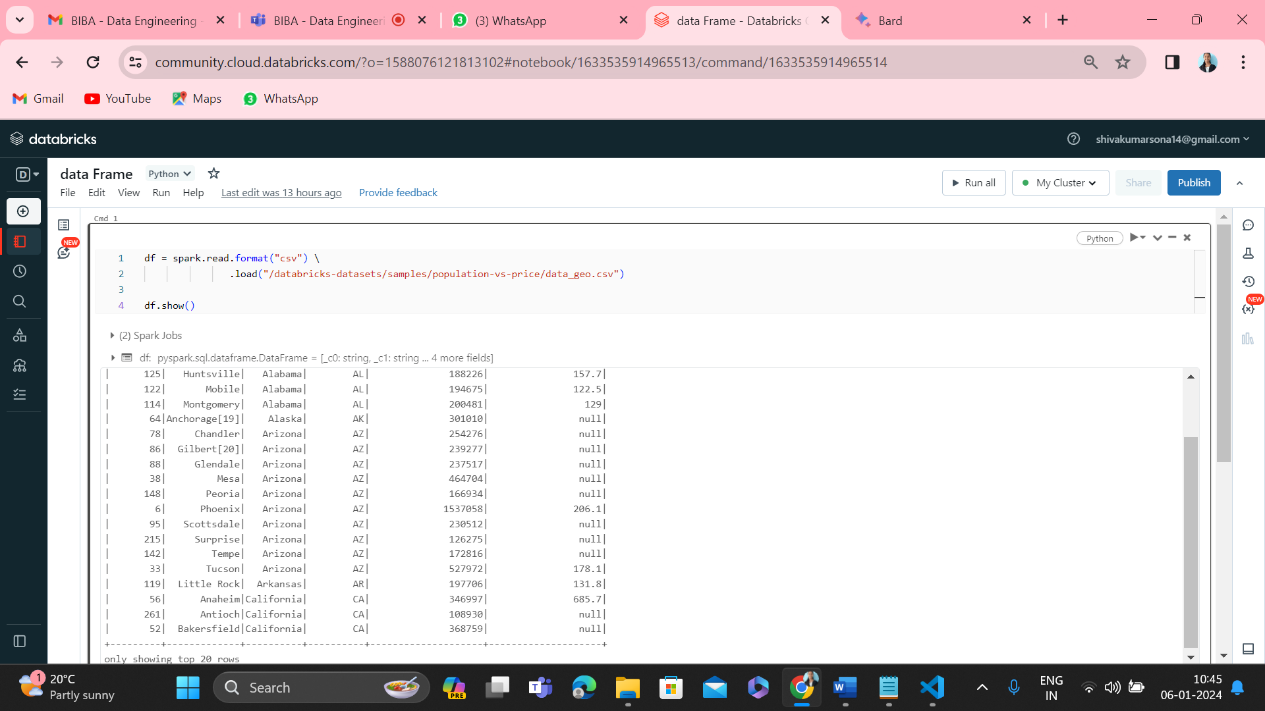
**Create a Visualizations in Databricks notebooks &**

**Rename, duplicate, or remove a visualization or data profile.**

* **Creating data Frame from Databricks Dataset:**

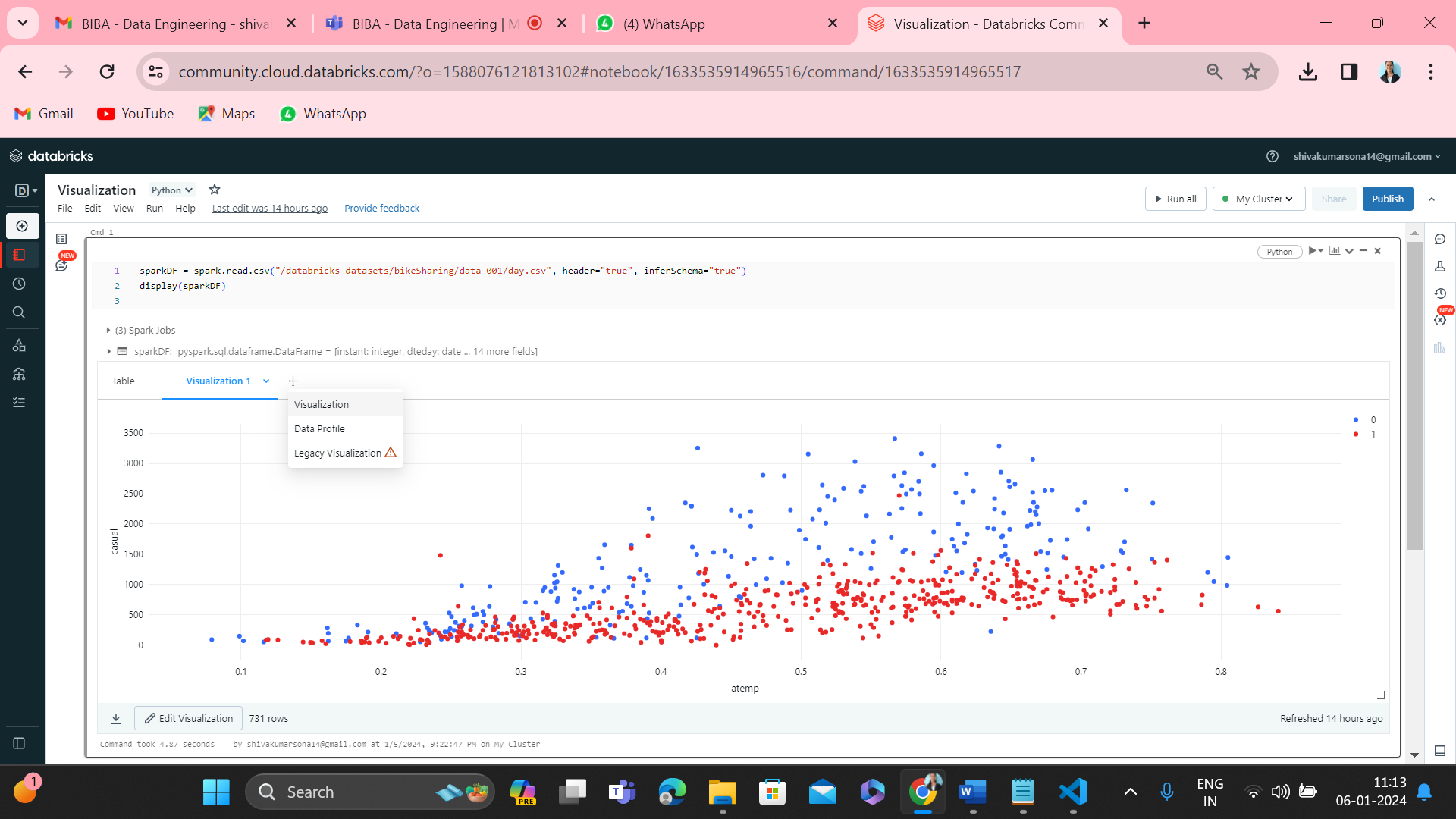
Spark Session: Data Frames are created using a Spark Session.

Created a new notebook and running the commands as shown below.

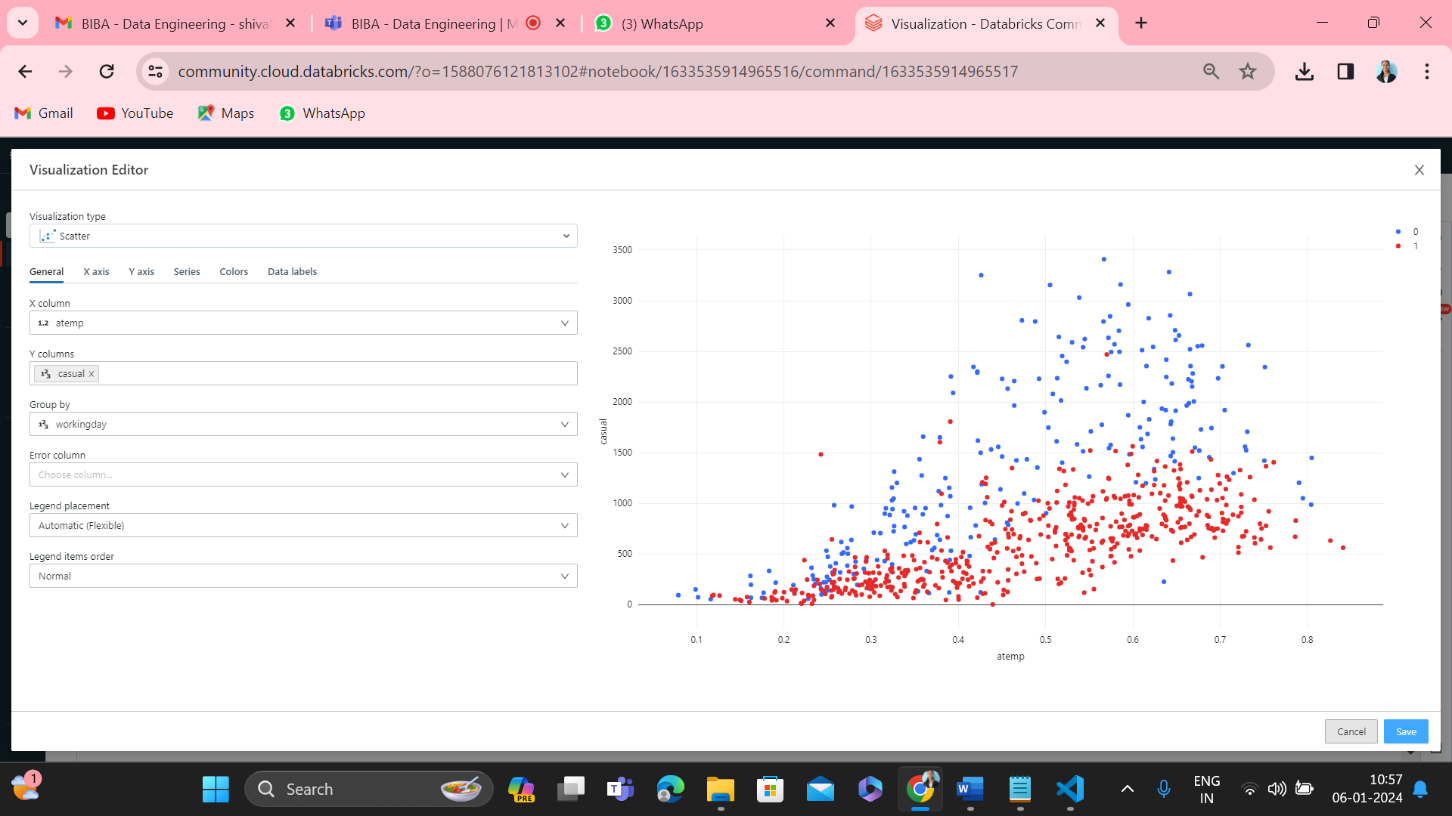
****

* **Creating a Visualizations in Databricks notebooks:**

Created a new notebook to visualize the things.

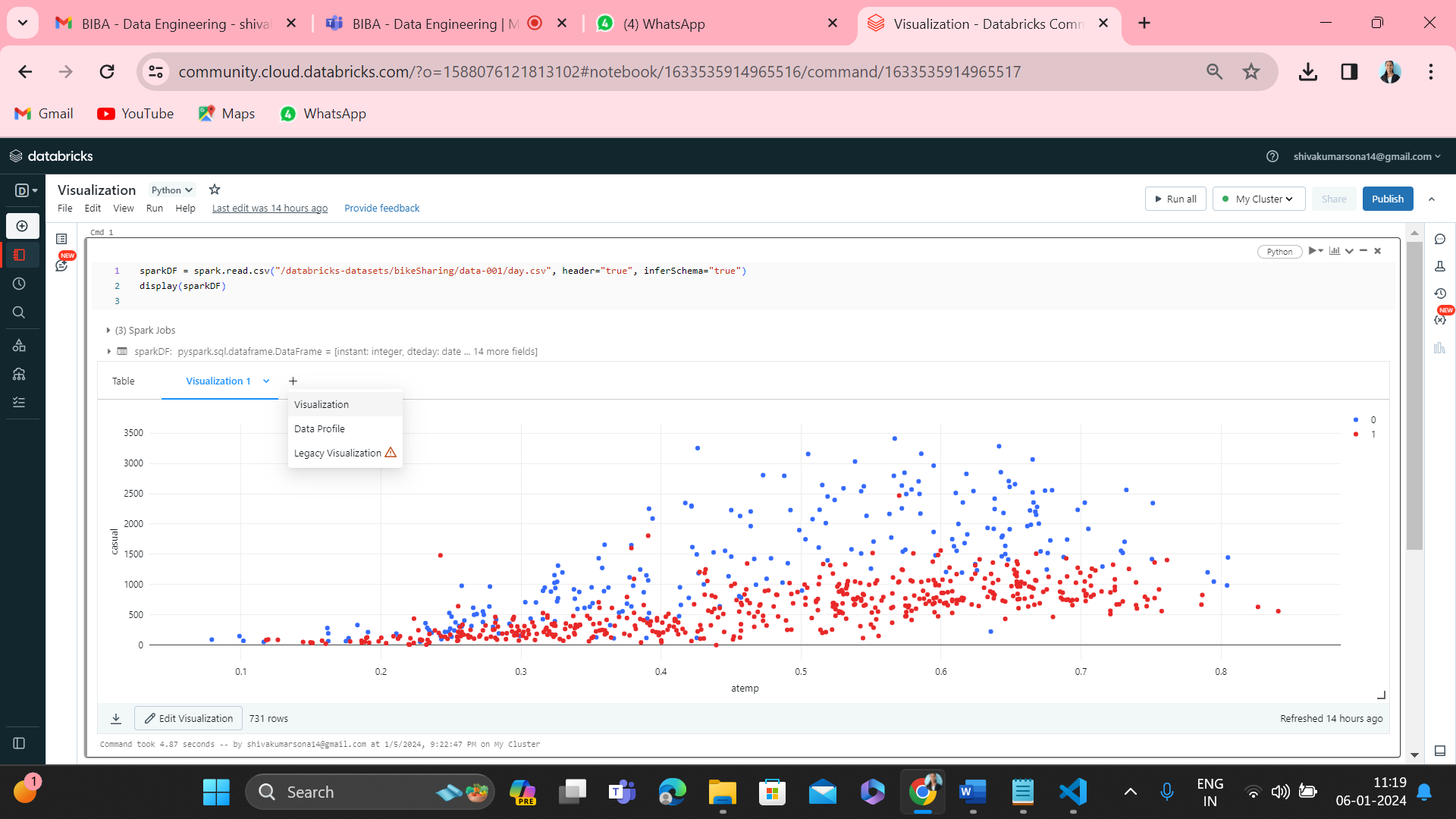


**After clicking visualization, we will get a visualization editor as shown below.**

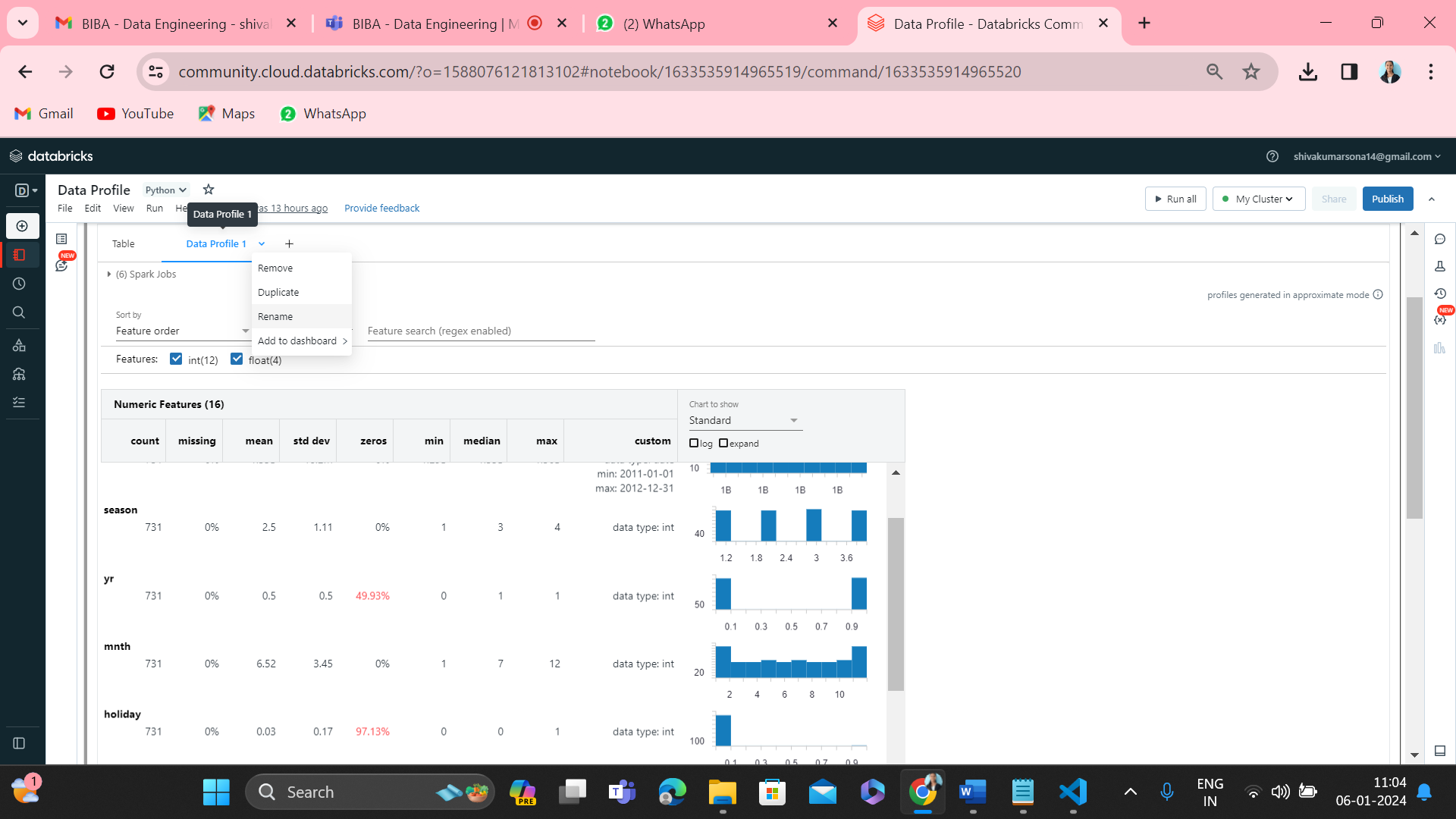
****

* **Rename, duplicate, or remove a visualization or data profile.**

To visualize the things, I ran the simple code and got table as shown below.



**TO Rename:** I Clicked the “rename” and changed the name from **Data Profile 1** to “**Data profile**”.



Like wise we can **remove** and create a **duplicate** just by clicking the names as shown in the above picture.

1. **Explain the 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.
* The Data Factory service allows you to create data pipelines that move and transform data and then run the pipelines on a specified schedule (hourly, daily, weekly, etc.).
* The data that is consumed and produced by workflows is time-sliced data, and we can specify the pipeline mode as scheduled (once a day) or one time.

**Simple Steps to create a data factory and copy activity**

1. To create a Data Factory with Microsoft Azure Portal, you will start by logging into the Azure portal. Click NEW on the left menu, click Data + Analytics, and then choose Data Factory.
2. Navigate to the ADF authoring UI.
3. Add a new pipeline.
4. Drag the Copy activity from the activities toolbox onto the pipeline canvas.
5. Configure the activity's properties, including source and sink datasets, linked services, copy behaviour, and other settings.
6. Save and publish the pipeline.

**Common Usage of Copy Activity**

1. Data ingestion from various sources into a centralized data lake or warehouse.
2. Data synchronization between databases and data stores.
3. Data loading for analytics and reporting platforms.
4. Data archiving and backup.