**NAME – MD SARFARAZ ALAM**

**EMAIL- SAMALAMABD17@GMAIL.COM**

**AZURE CODING ASSESSMENT DATE-06-01-2024**

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.

**1. Create a Cluster:**

* Navigate to the Databricks workspace.
* Click on the "Clusters" tab in the sidebar.
* Click on "Create Cluster."
* Configure the cluster settings and click on "Create Cluster."

**2. Attach the Notebook to the Cluster:**

* Open the notebook you want to run.
* Click on the "Detached" dropdown at the top right.
* Choose your cluster from the list to attach the notebook.

**3. Run All Commands in the Notebook:**

* In the notebook, use the "Run All" option from the "Run" menu or toolbar.

**4. Create a DataFrame from a Databricks Dataset:**

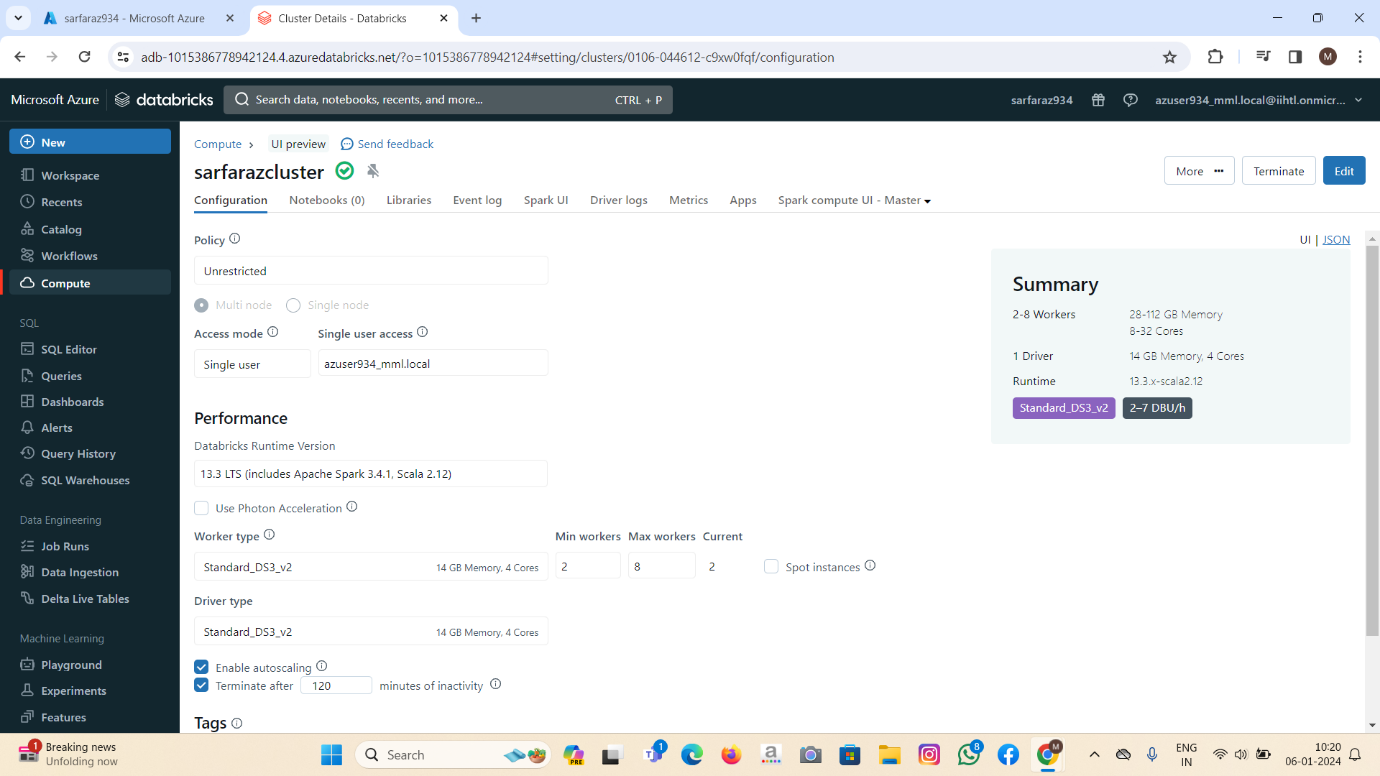
Assuming you have a dataset in DBFS (Databricks File System) or a Databricks Delta table:

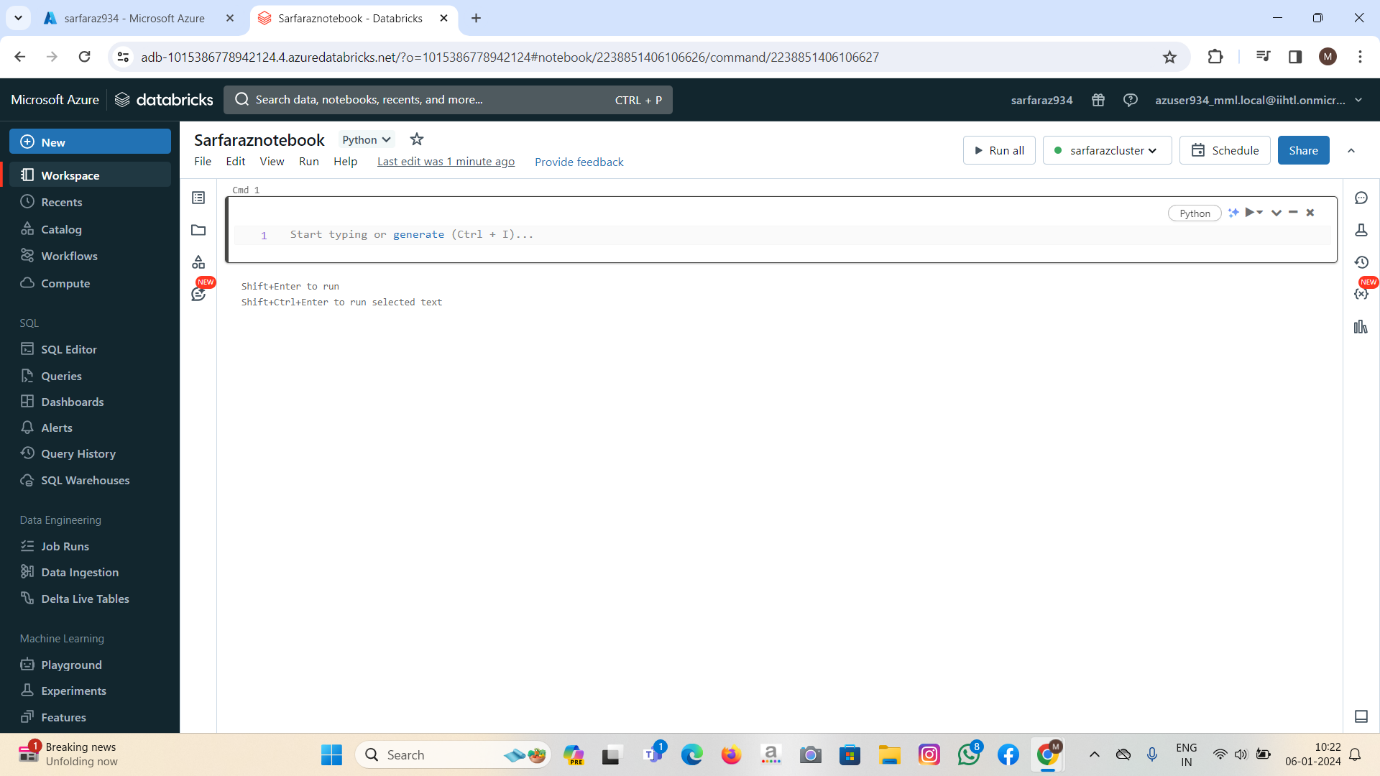
**5. Create Visualizations in Databricks Notebooks:**

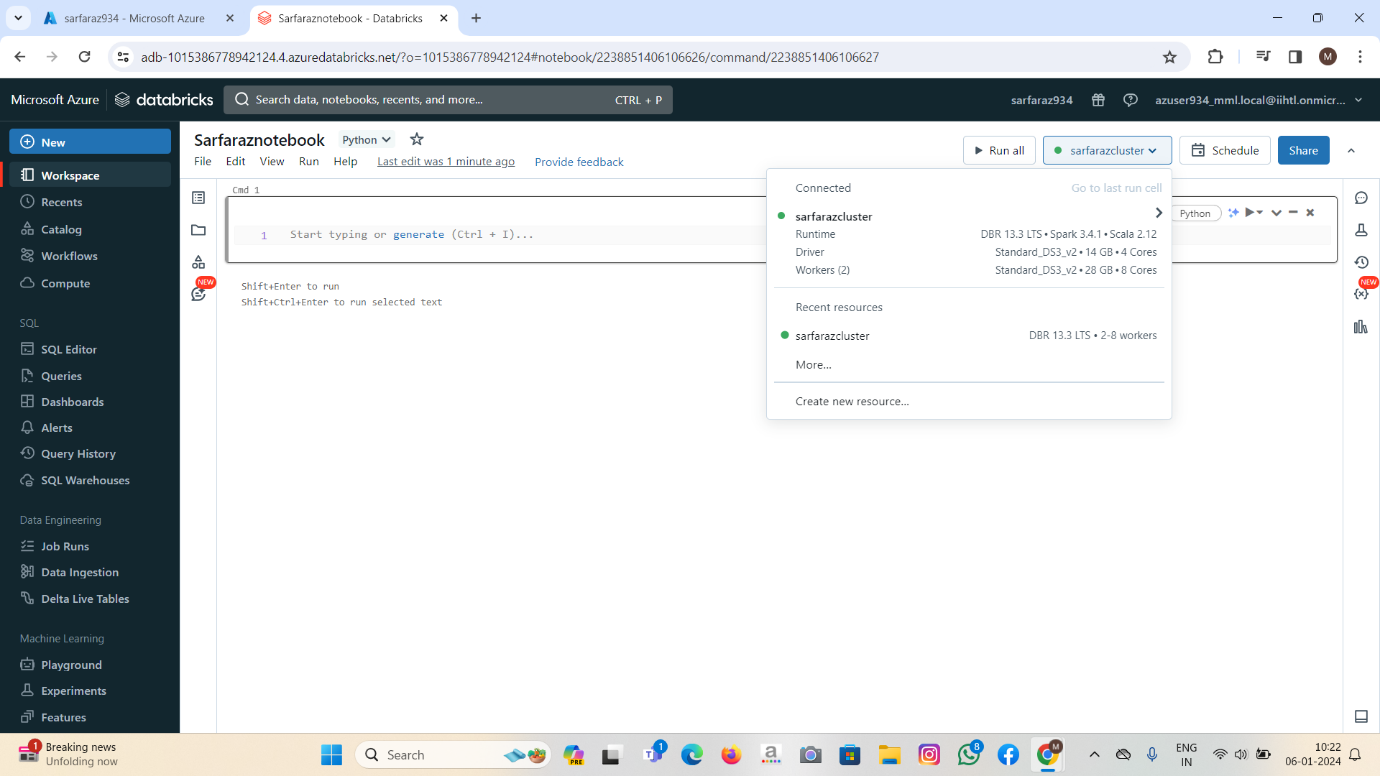
* Use **%sql** magic commands to create visualizations.

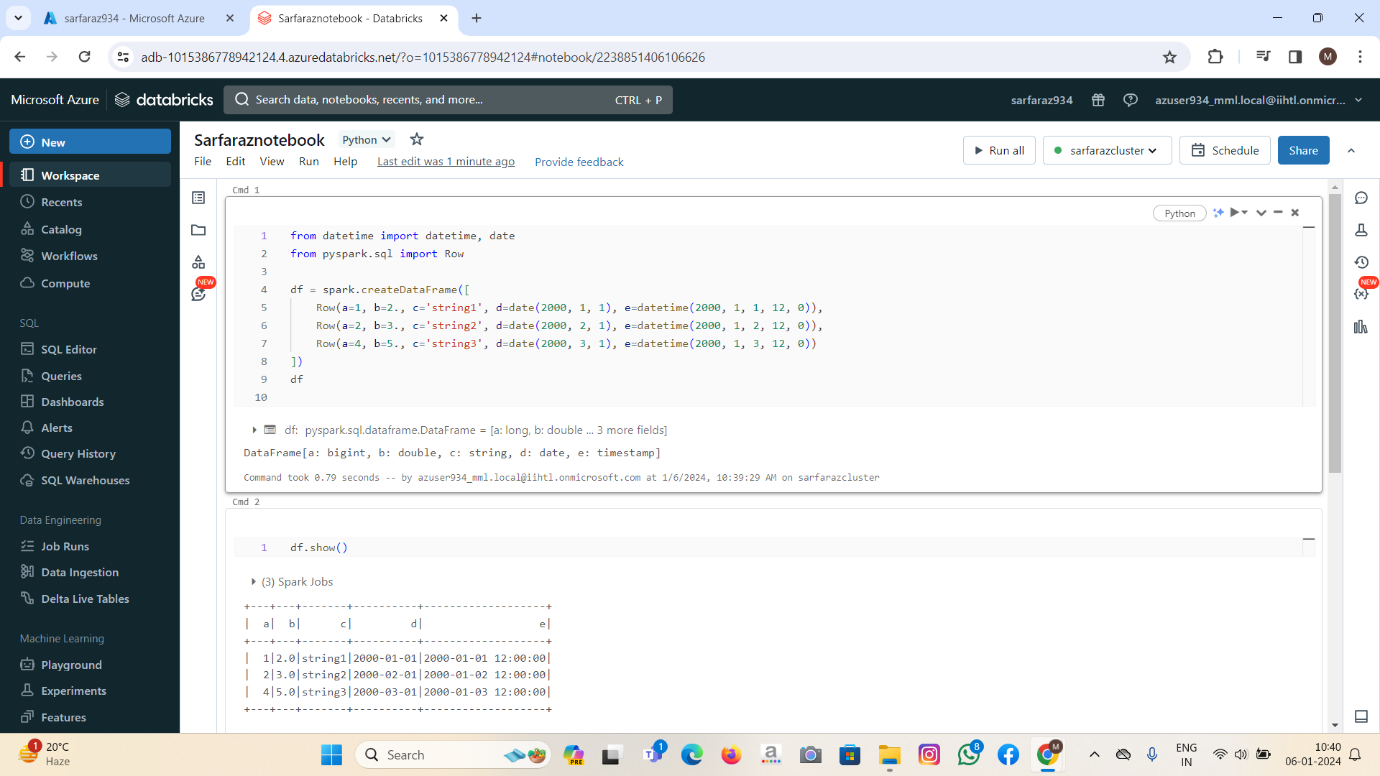
**6. Rename, Duplicate, or Remove a Visualization or Data Profile:**

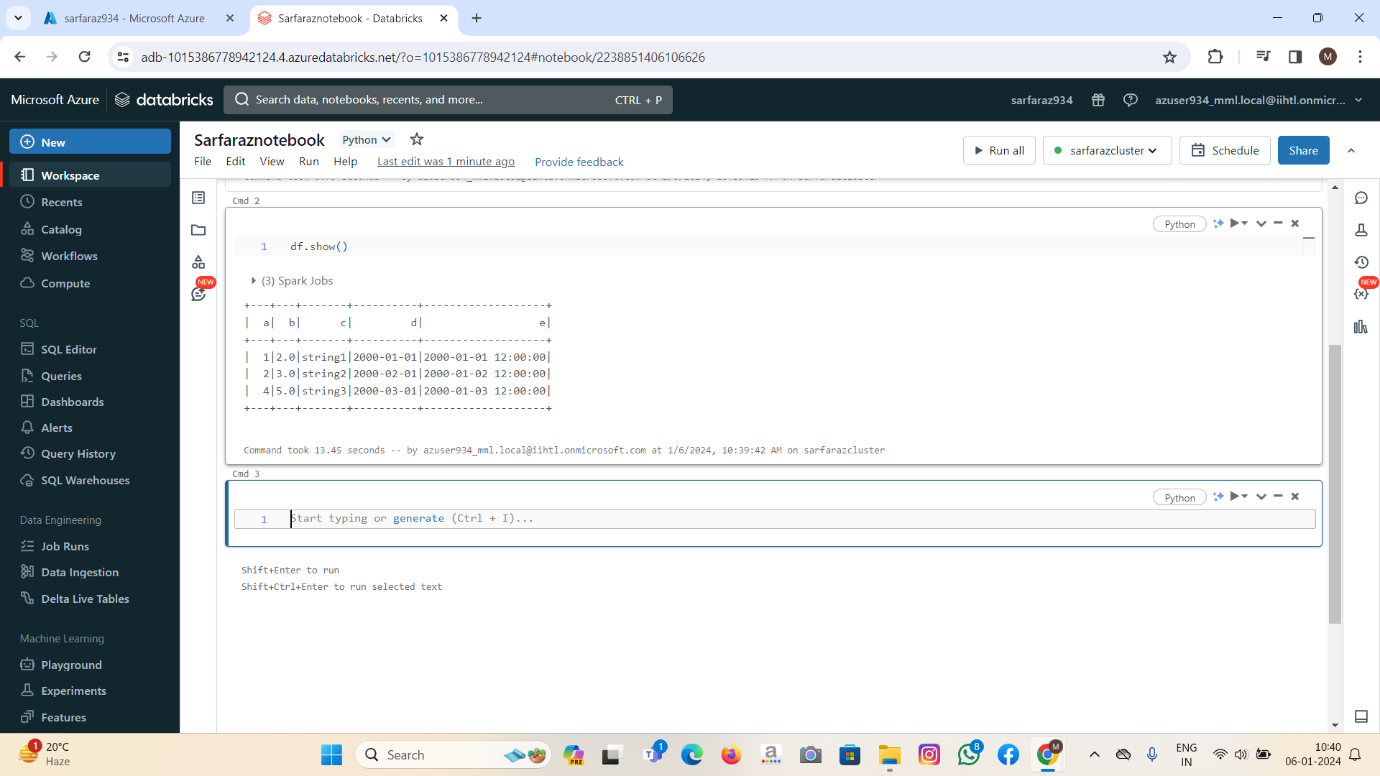
* After creating a visualization, we can interact with it.
* Click on the visualization, and options to rename, duplicate, or remove should be available.

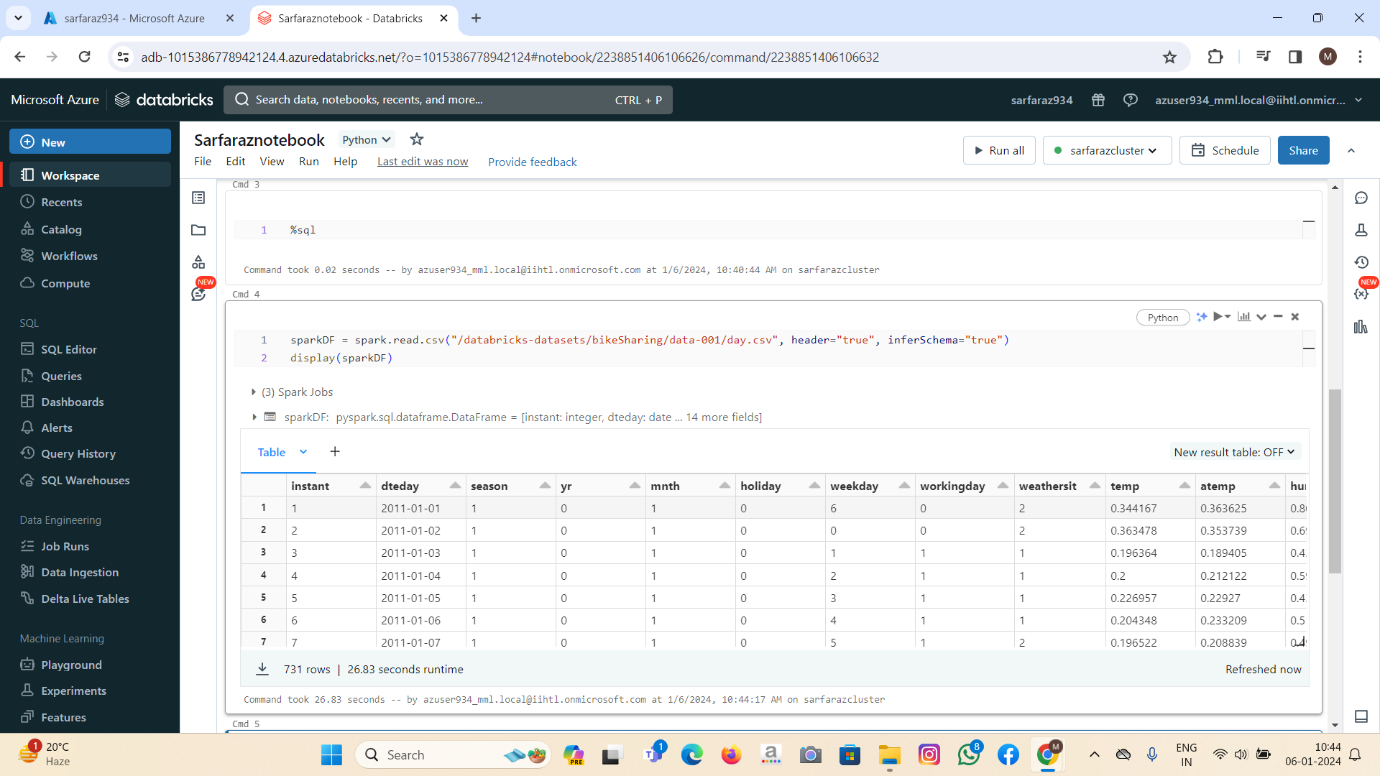


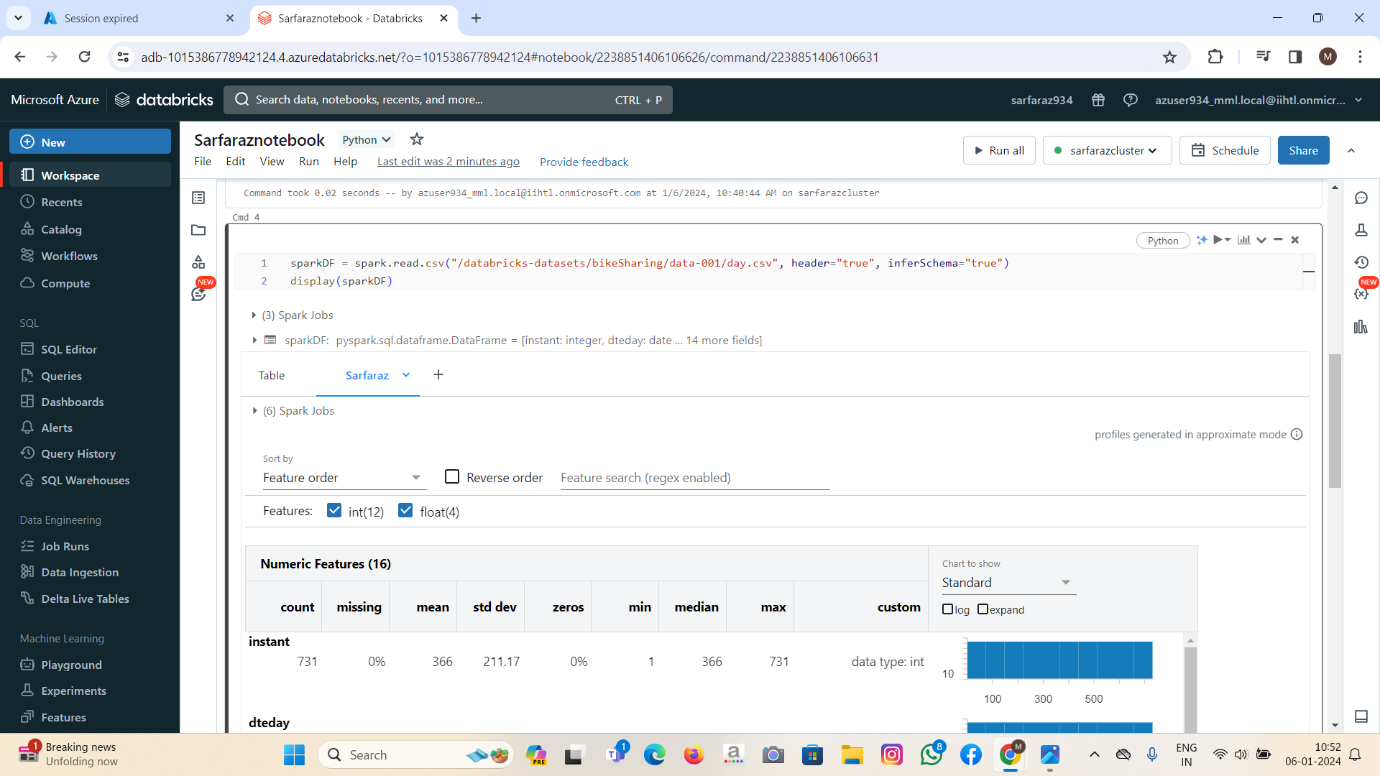
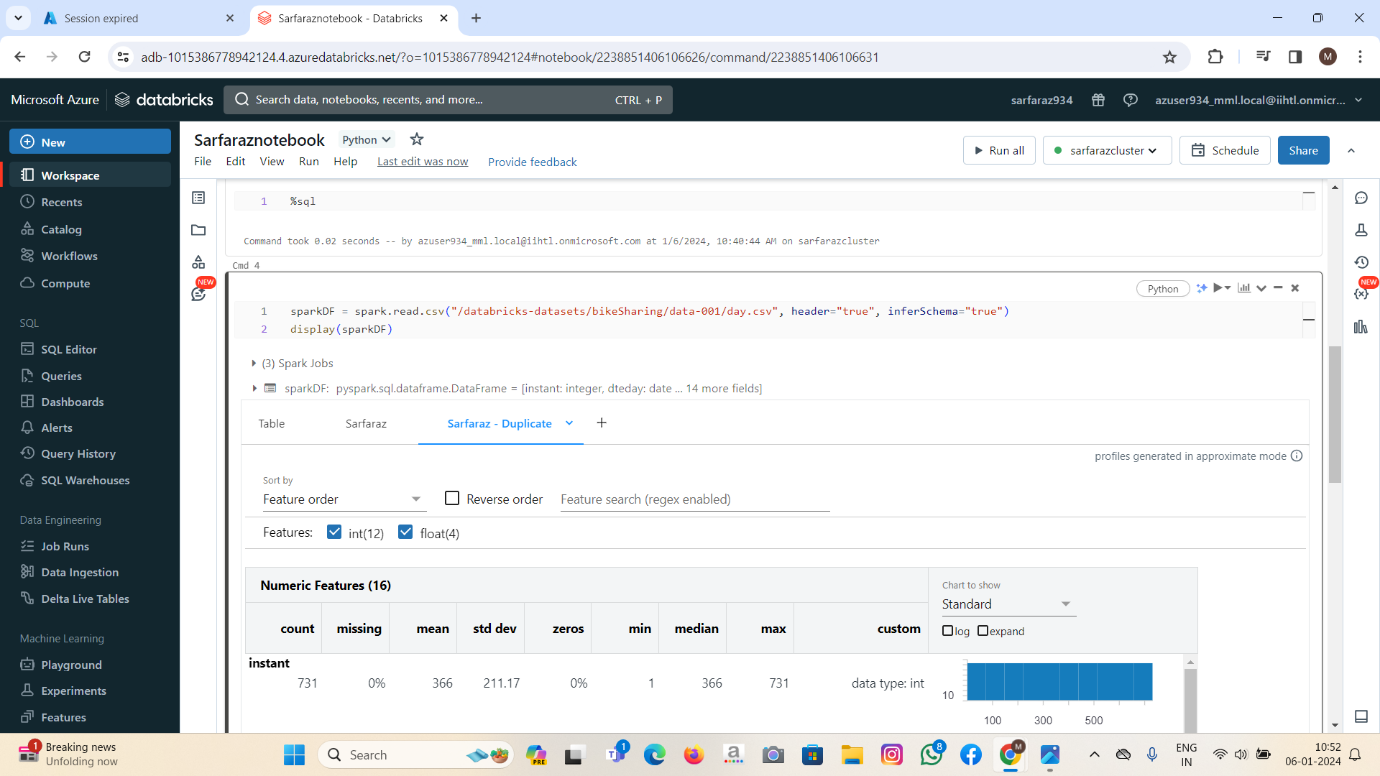


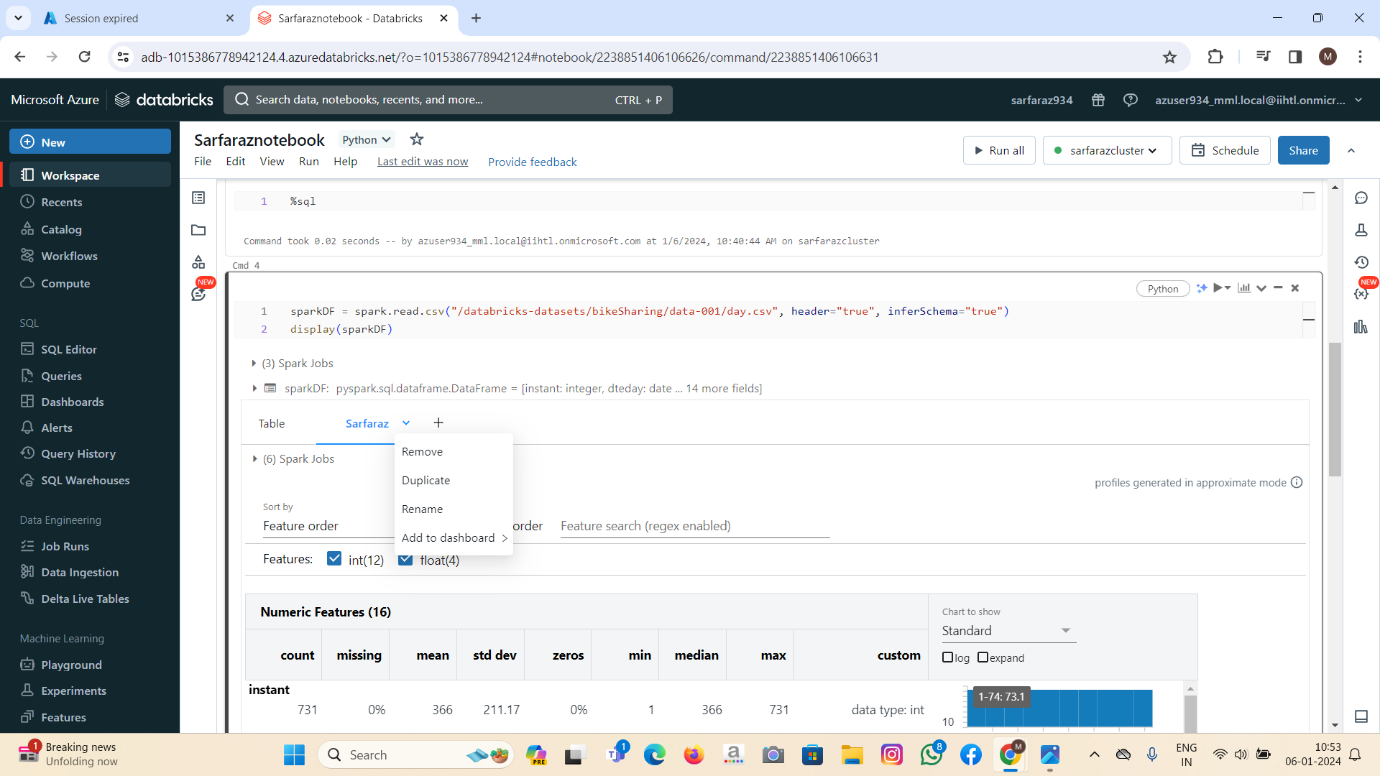
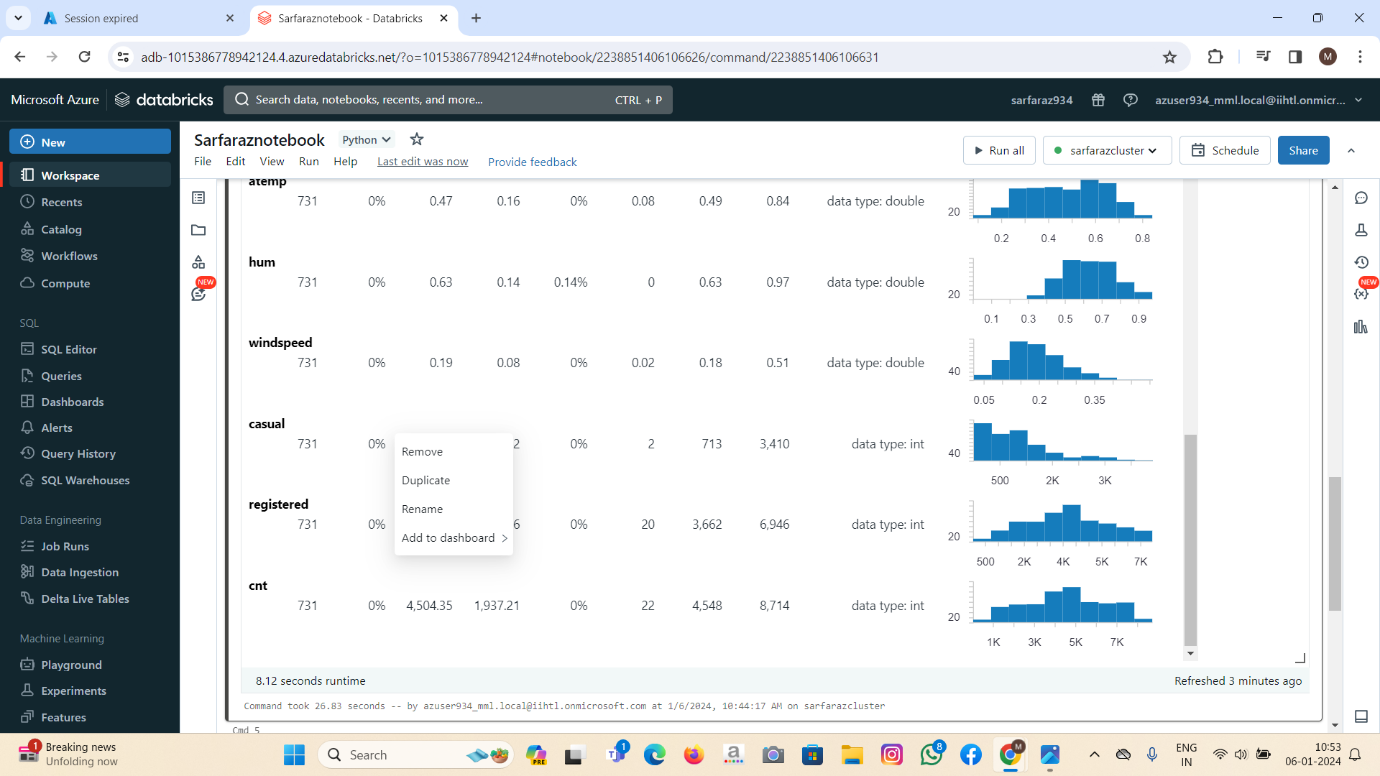












1. Explain the copy activity in Azure data factory.

In Azure Data Factory(ADF), the “Copy Data” activity is a fundamental building block that allows you to move data from one location to another. It is a part of the data movement and transformation capabilities provided by azure data factory. The Copy Data Activity is commonly used for tasks such as data migration, data integration, and data synchronization between different data stores.

Here are the key components and concepts related to the copy Data activity in Azure Data Factory:

1. Source Dataset:

* Represents the source data store from which data needs to be copied.
* Can be a file, database table, blob, on-premises data source, or any other supported data store.

1. Destination Dataset:

* Represents the destination data store where data will be copied.
* Similar to the source dataset, it can be any supported data store.

1. Mapping:

* Defines the relationship between the source and destination datasets.
* Specifies how data should be mapped, transformed, or modified during the copy operation.

1. **Integration Runtimes:**

* Specifies the compute infrastructure where the data movement and transformation will take place.
* Azure Data Factory supports different integration runtimes, including Azure, Self-hosted, and Azure-SSIS Integration Runtime.

1. **Data Movement and Transformation Activities:**

* Include various settings and configurations to control how data is moved and transformed during the copy operation.
* This can involve schema mapping, column mapping, data type conversions, and more.

1. **Monitoring and Logging:**

* Azure Data Factory provides monitoring capabilities to track the progress and status of the copy activities.
* You can view execution details, monitor performance, and identify any issues that may arise during the data movement process.

1. **Scheduling:**

* Copy activities can be scheduled to run at specific times or triggered by events.
* Scheduling options allow you to automate data movement tasks based on your business requirements.

1. **Error Handling:**

* Azure Data Factory includes features for handling errors during the copy process.
* You can configure retry policies, logging, and notifications to ensure that data movement tasks are robust and reliable.