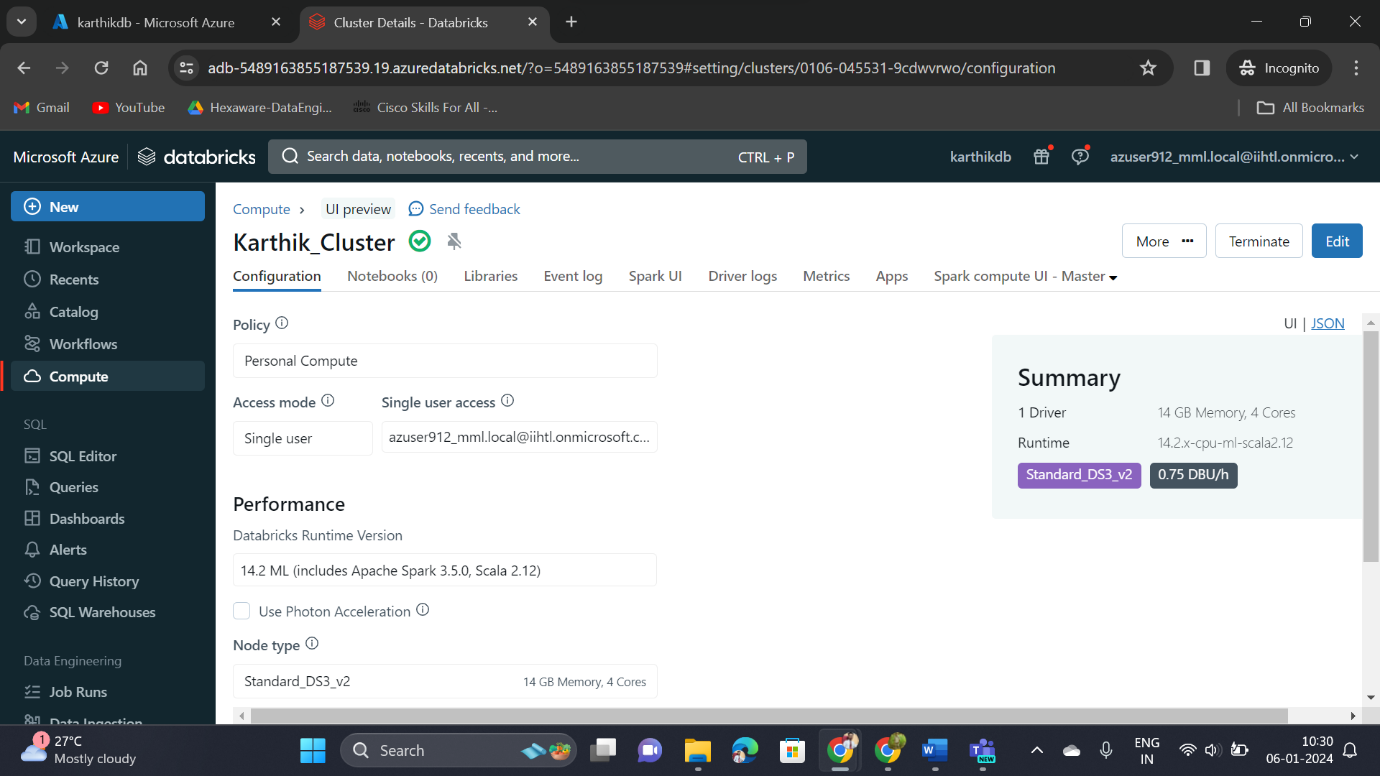
Name: Karthikeyan K R

Date: 06-01-2024

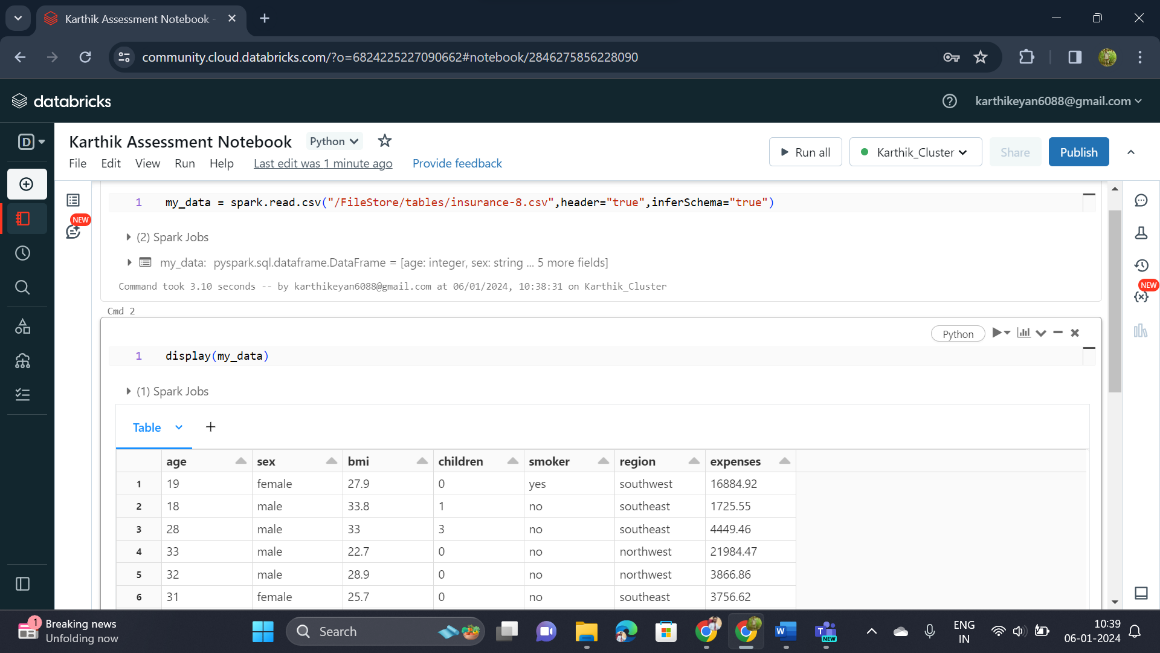
Topic: Azure Databricks

Type: Weekly Coding Assessment

1. Creating a cluster adding that to the notebook



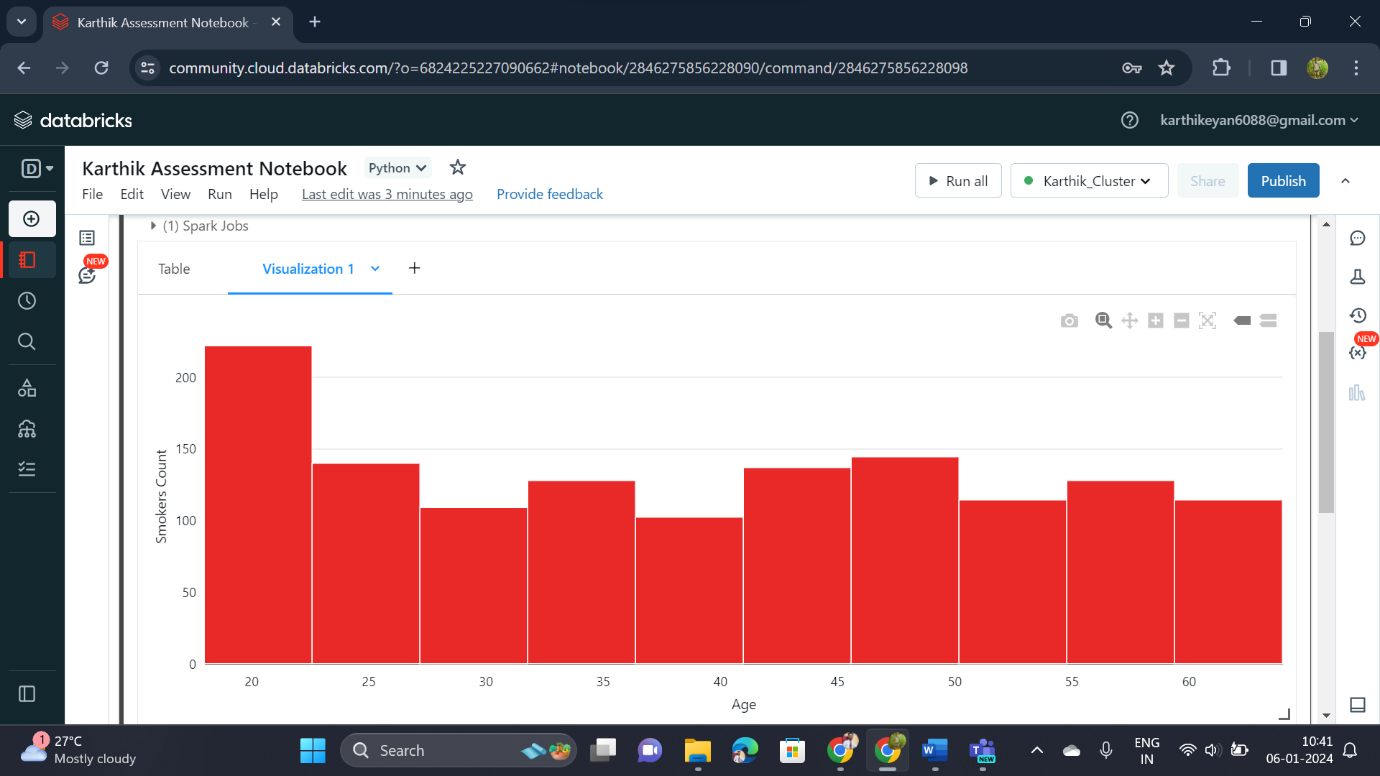
Then reading a csv file and converting into the data frame



Creating visualization and analysing the created one

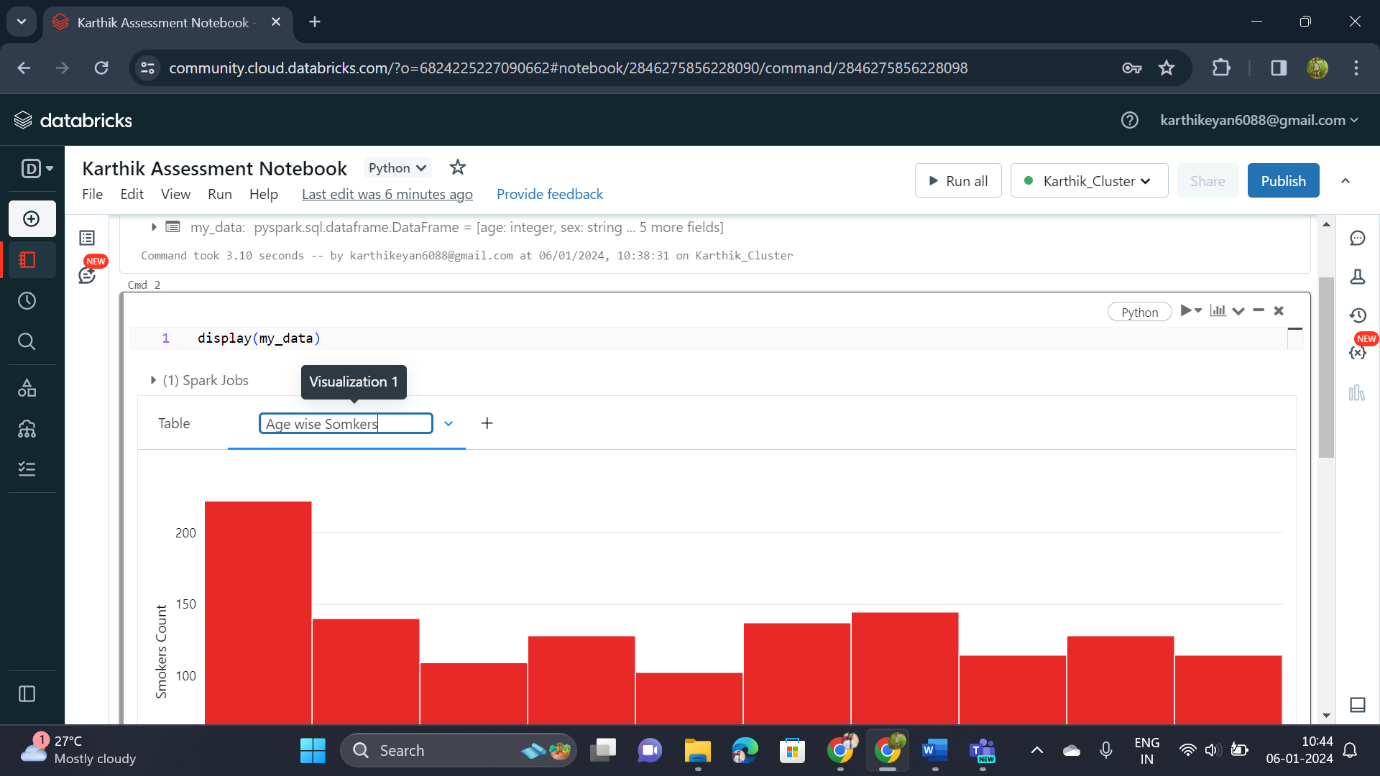
* To create visualization, click the Plus icon below the Spark Jobs
* Enter the x-axis and y-axis then select the visualization type





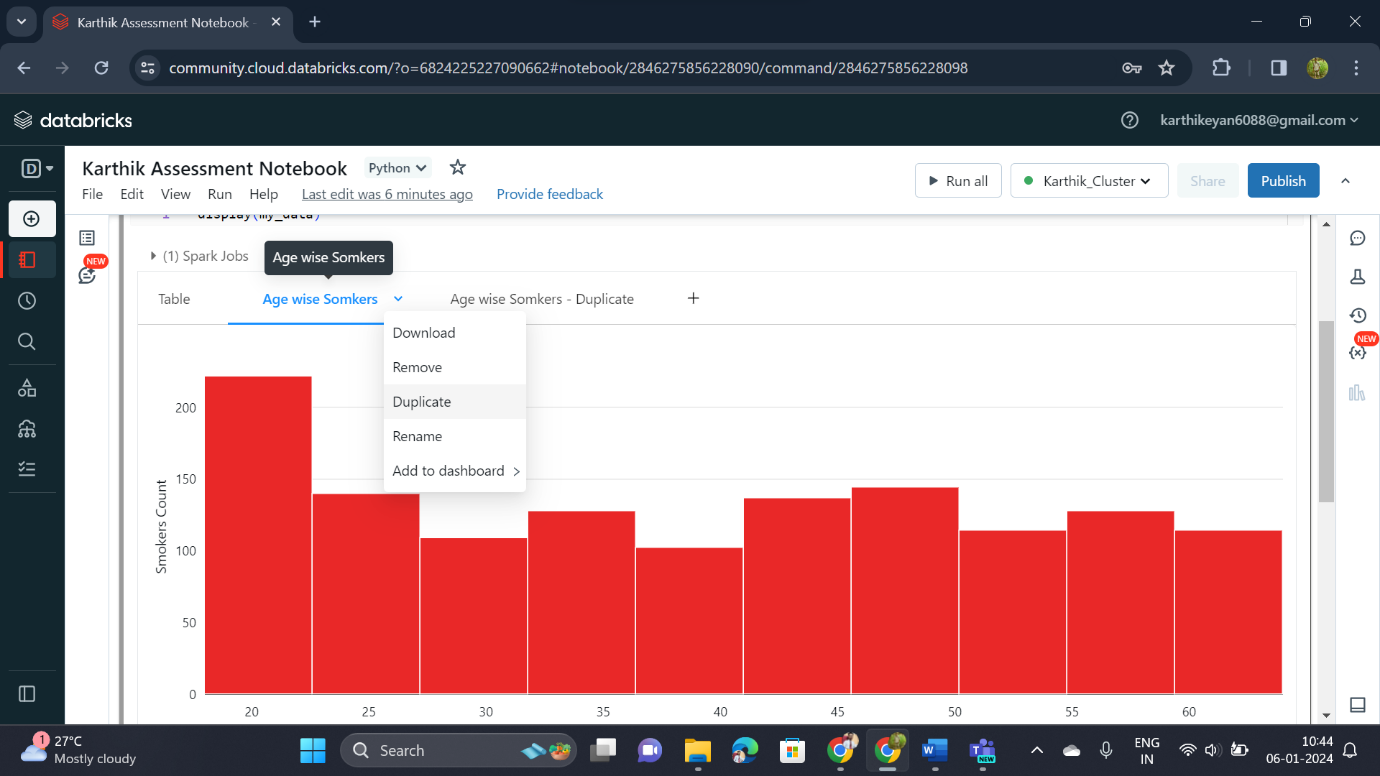
Renaming the visualization

* To rename the visualization, Click the drop down near visualization name tab



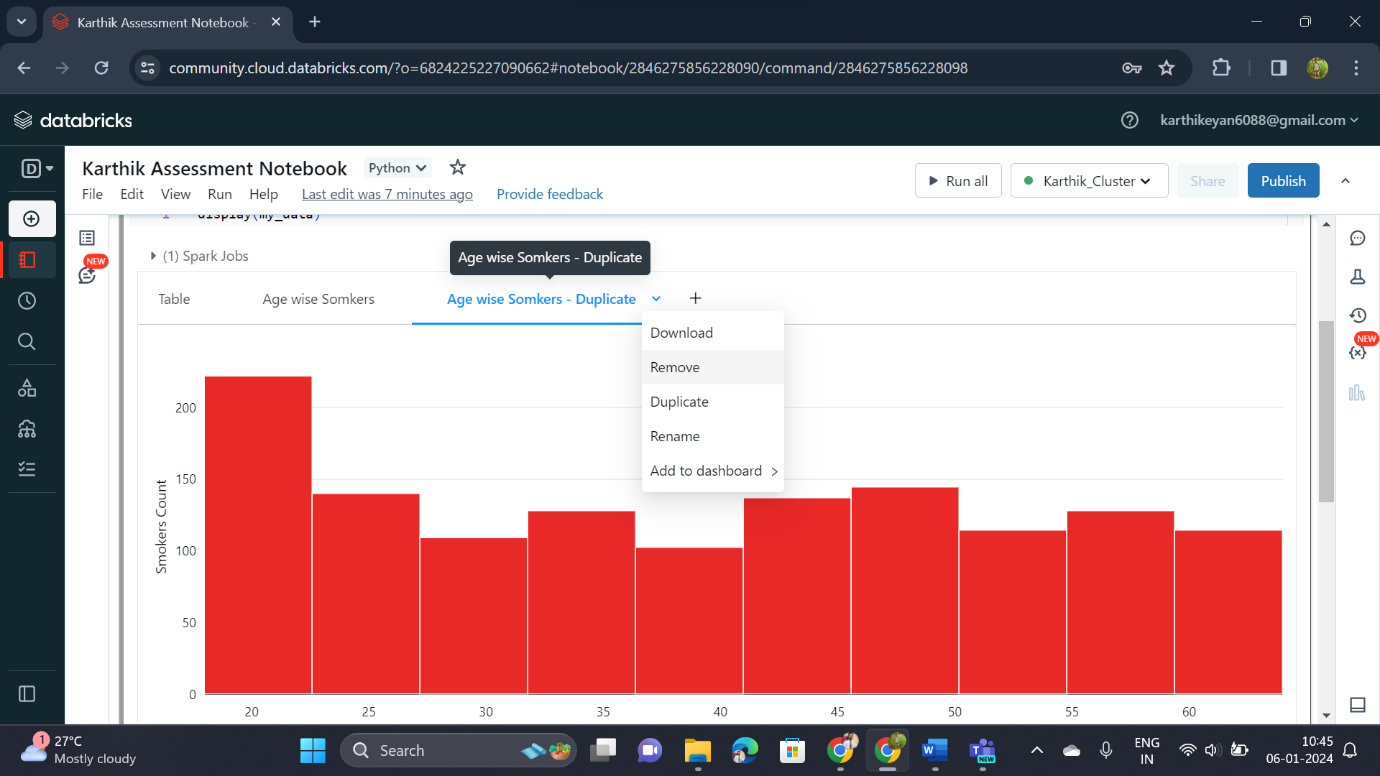
Duplicating the Visualization

* Inorder to duplicate it, click duplicate option in drop-down menu



Removing the visualization

* If we want remove the visualization the, then click the remove option from the drop-down menu



1. 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

1. **Data Pipelines:**

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

1. **Data Sources:**

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

1. **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.

1. **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.

1. **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.

1. **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:

1. **Data Copy Activity:**
   * Moves data from one location to another, like copying files from on-premises storage to Azure Blob Storage.
2. **SQL Activity:**
   * Runs a SQL script on a database to retrieve or manipulate data.