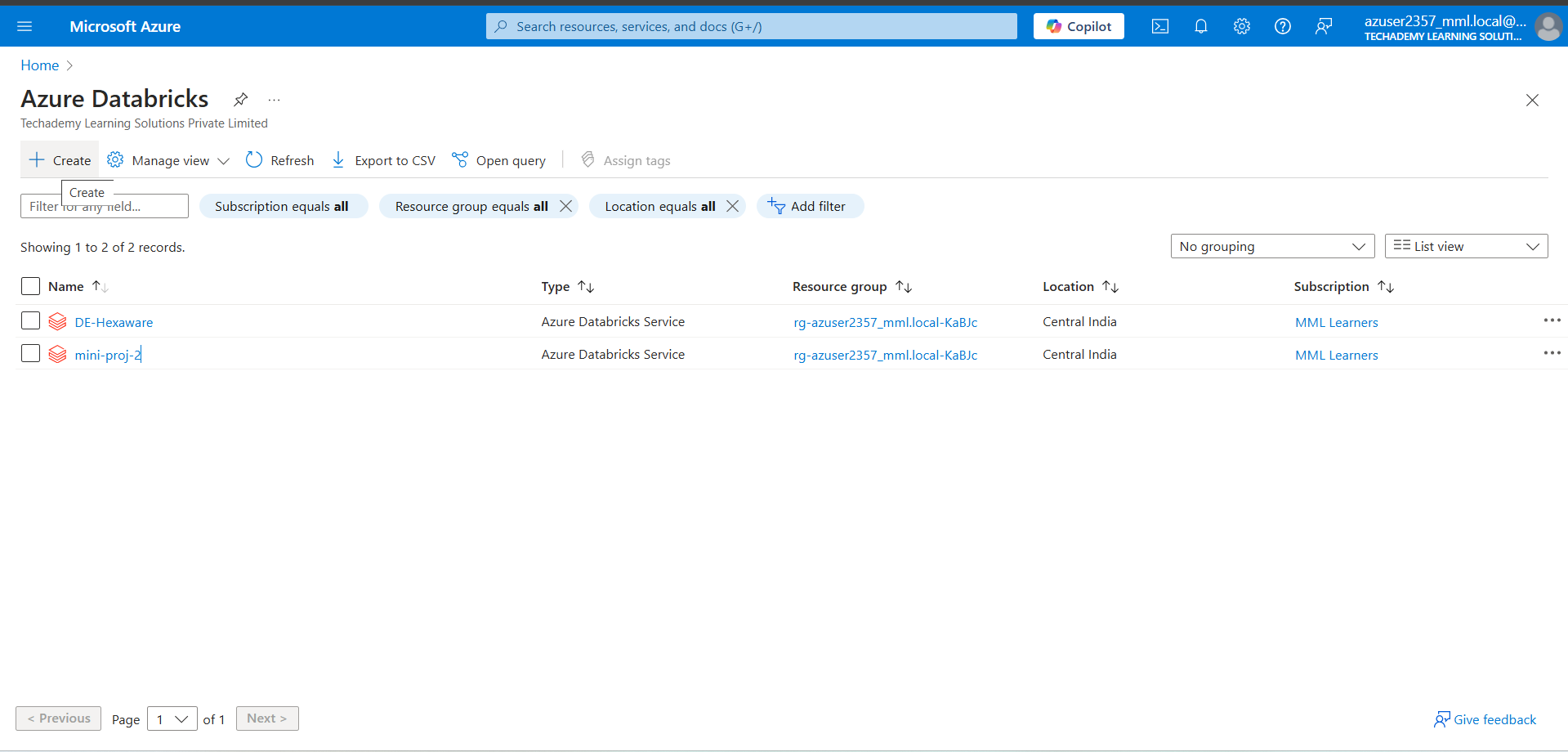
**Hexaware Technologies**

Azure – Coding Challenge

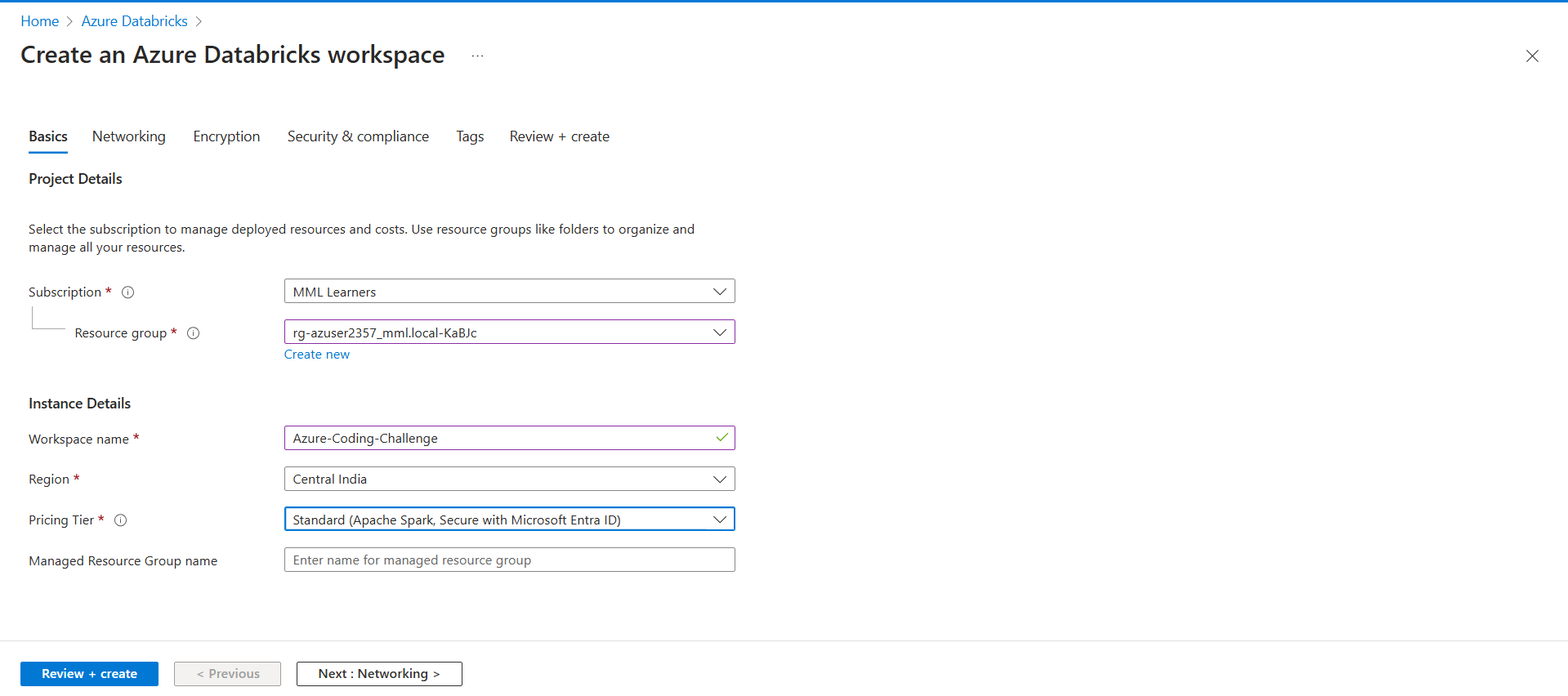
**A)Create a cluster &Attach the notebook to the cluster and run all commands in the notebook & creates a DataFrame from a Databricks dataset& Create a Visualizations in Databricks notebooks**

**Step 1: Create a Databricks Workspace**

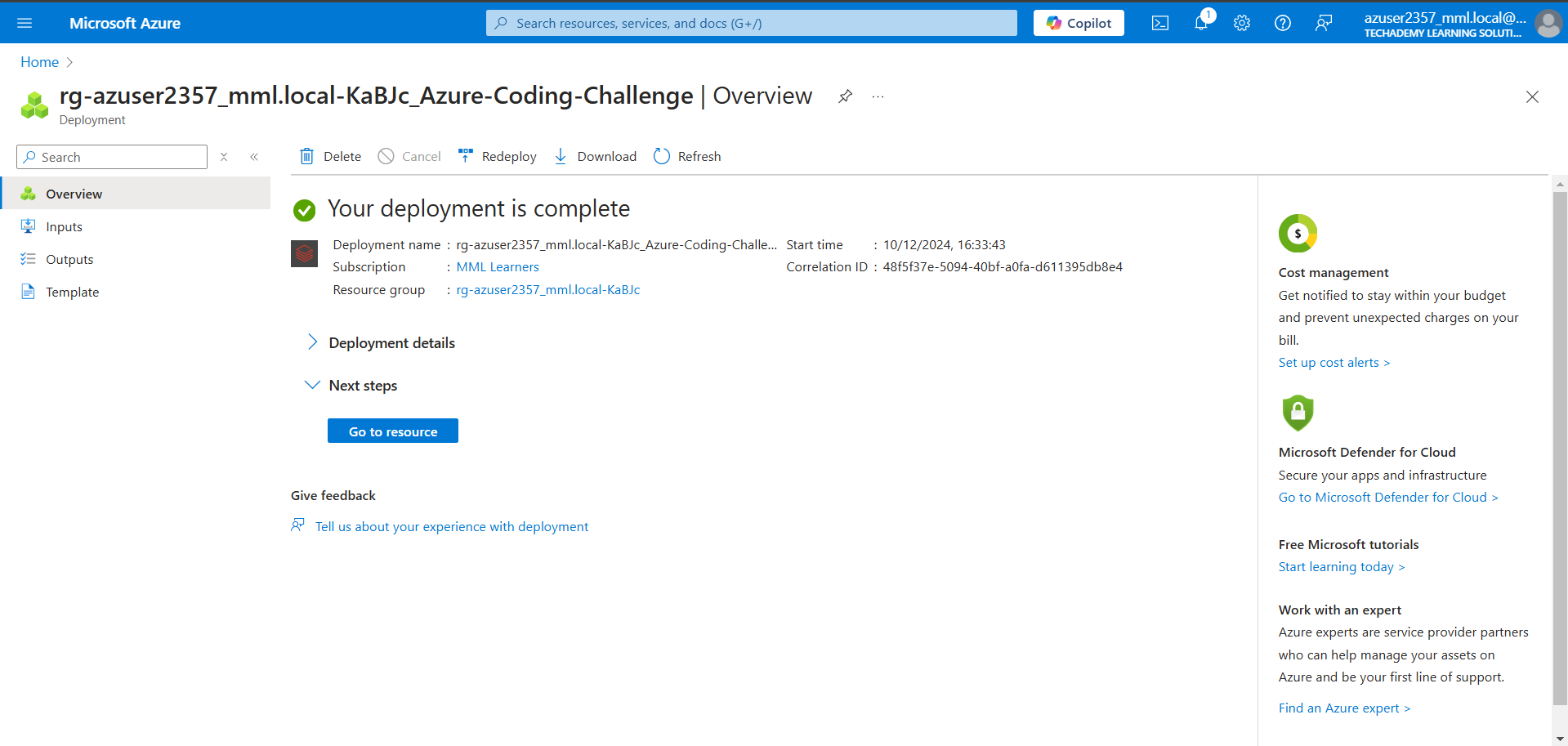
1. Log in to your Azure portal and search for Databricks in the search bar.
2. Click **Azure Databricks** and then click **Create**.



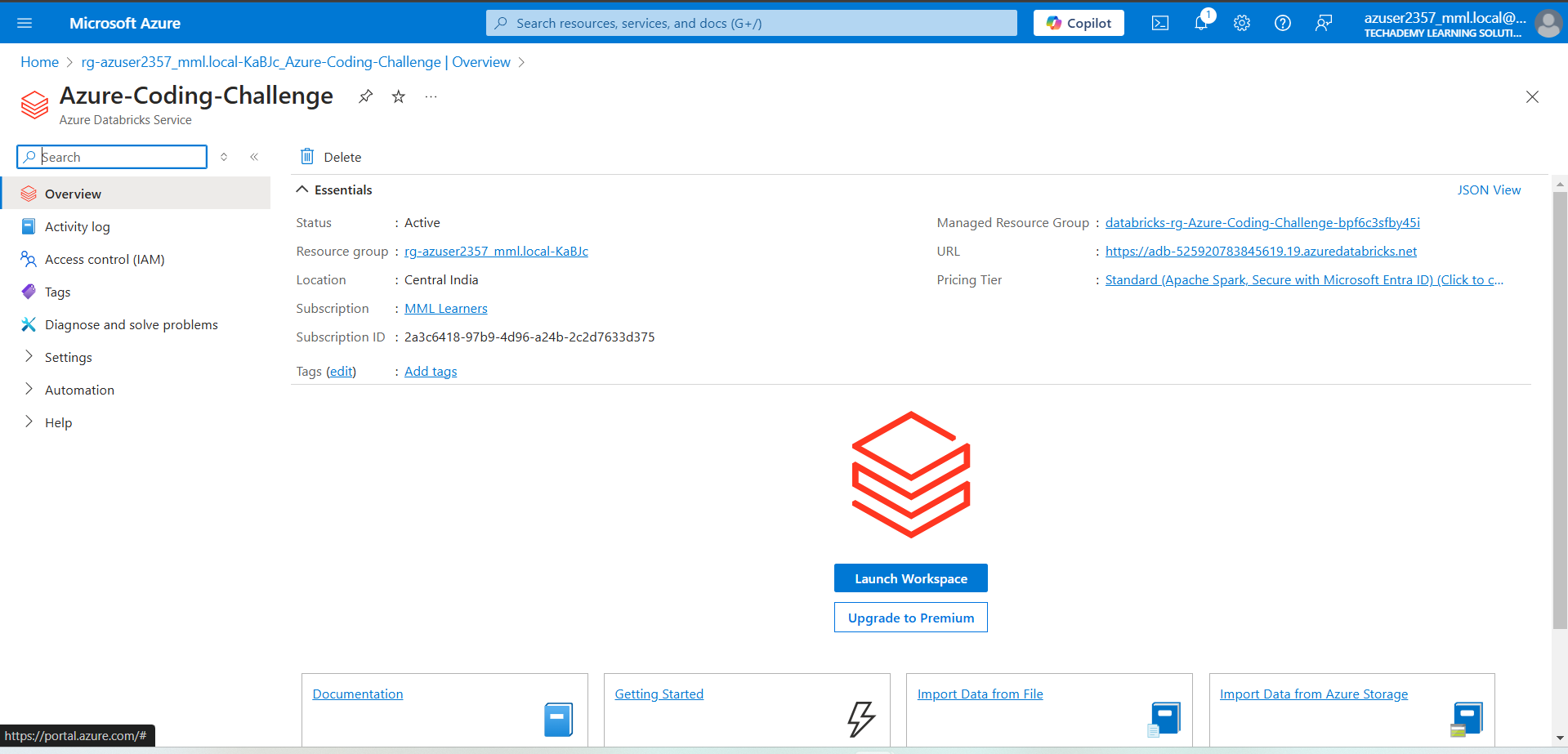
1. Fill in the required details:
   * **Workspace name**: Give your workspace a name.
   * **Region**: Choose a region close to you for better performance.
   * Pricing tier: Select the tier that fits your requirement (e.g., Standard or Premium).



1. Click **Review + Create**, and after the validation passes, click **Create** to deploy the workspace.
2. Once the deployment is complete, click **Go to resource**.

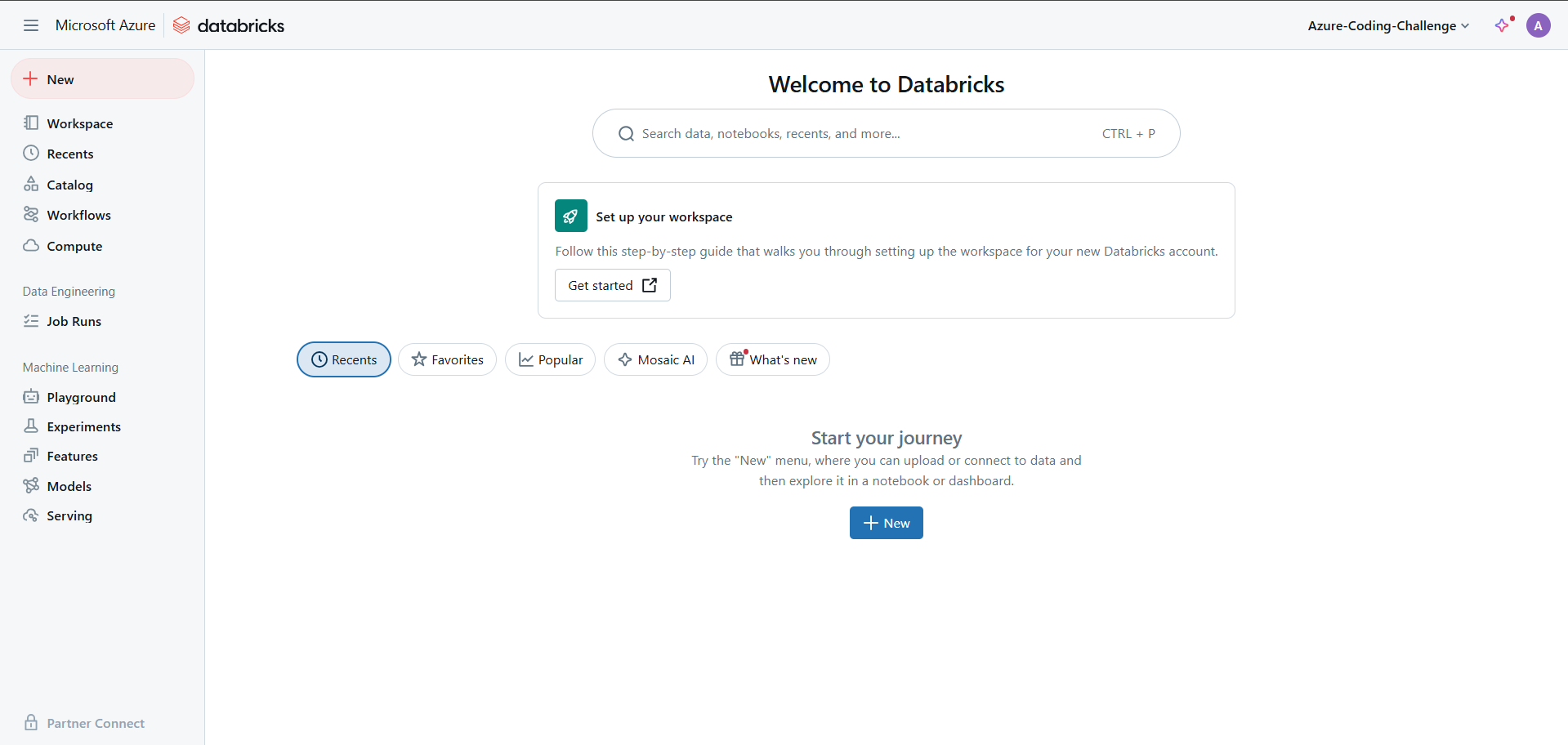


1. Click **Launch Workspace** to open your Databricks environment.

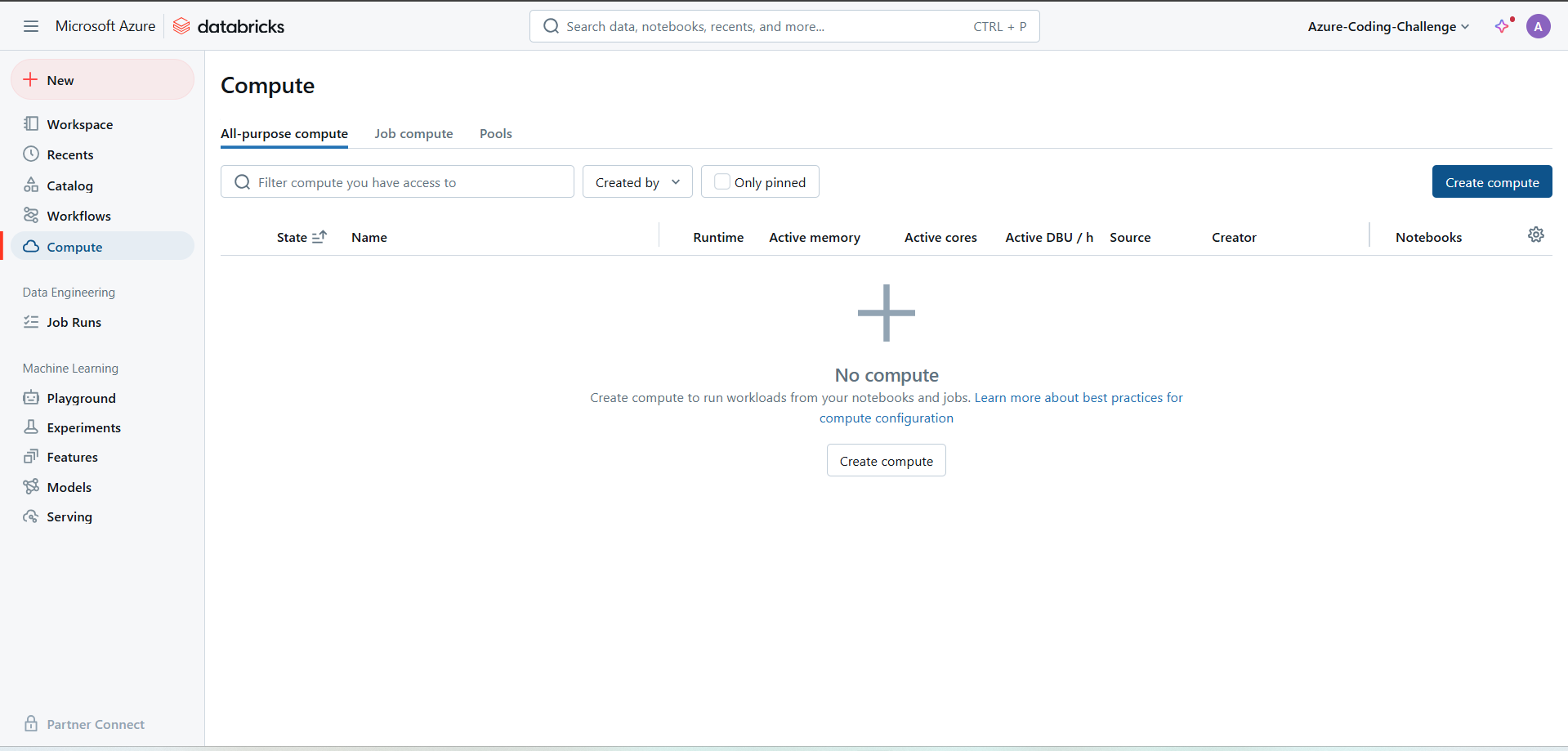


**Step 2: Creating a Cluster**

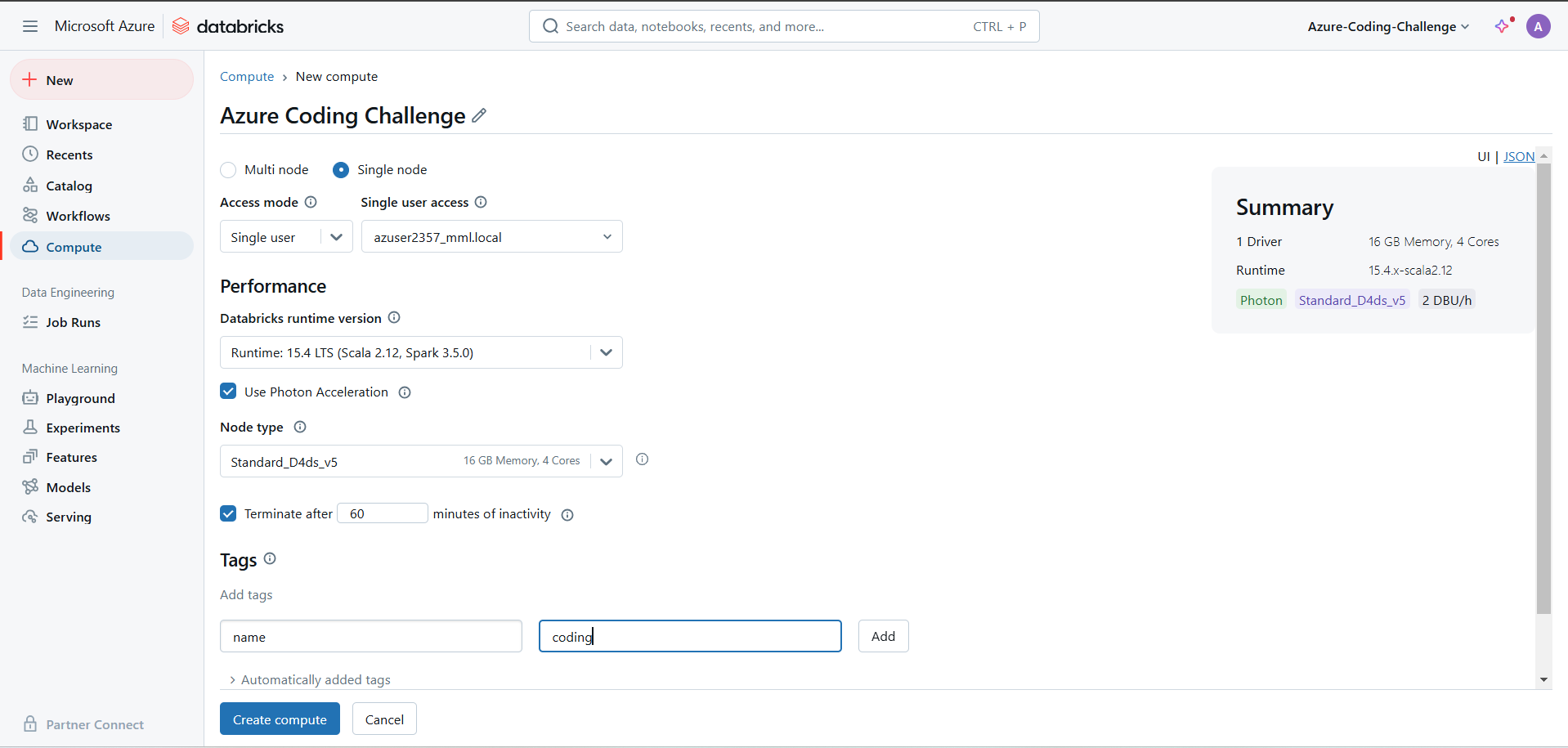
1. In your Databricks workspace, navigate to the **Compute** tab on the left-hand menu.



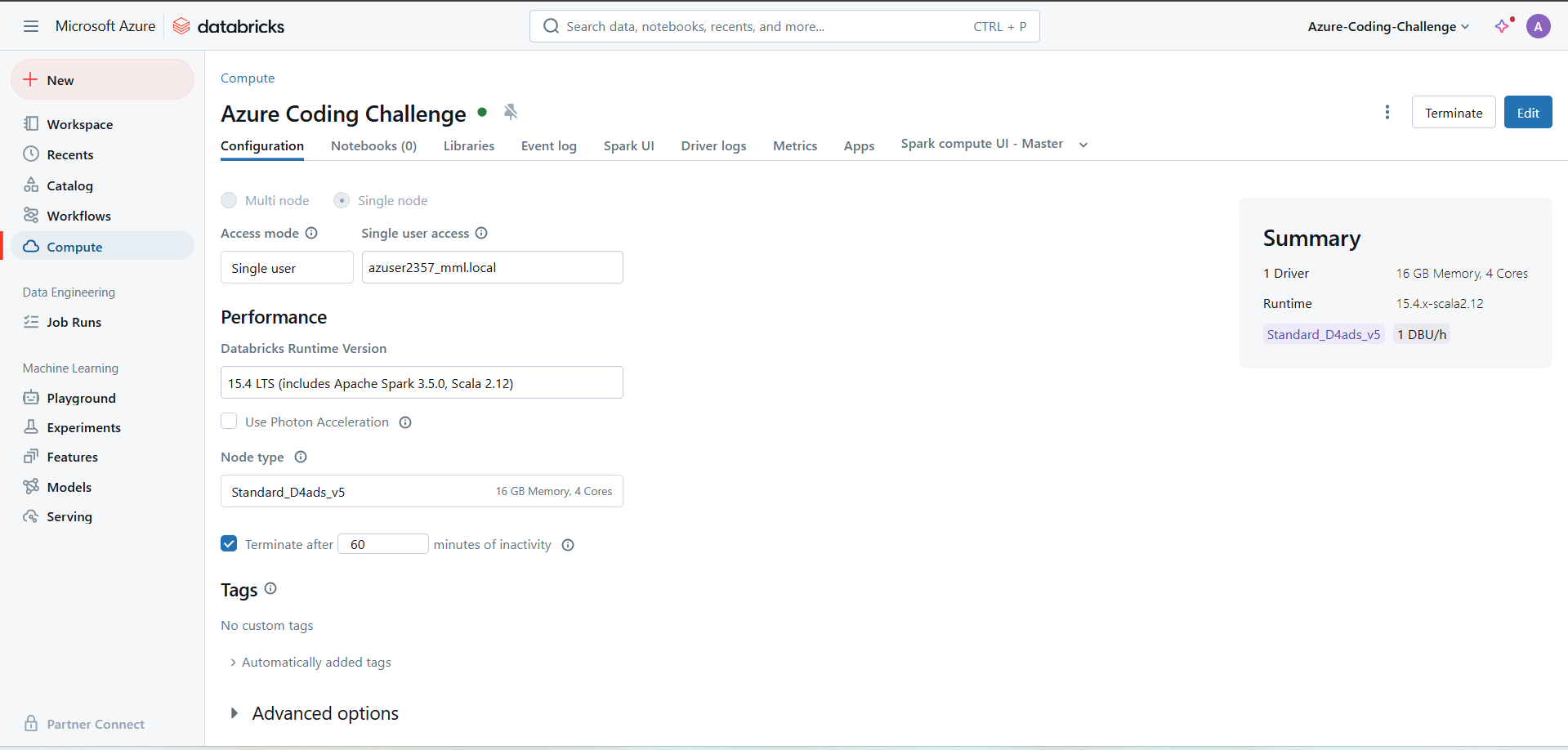
1. Click on the **Create Cluster** button.



1. Provide the following details:
   * **Cluster Name**: Give your cluster a name.
   * **Cluster Mode**: Select Single Node or Standard depending on your task.
   * **Databricks Runtime Version**: Choose the latest runtime.
   * **Node Type:** Select an appropriate instance.

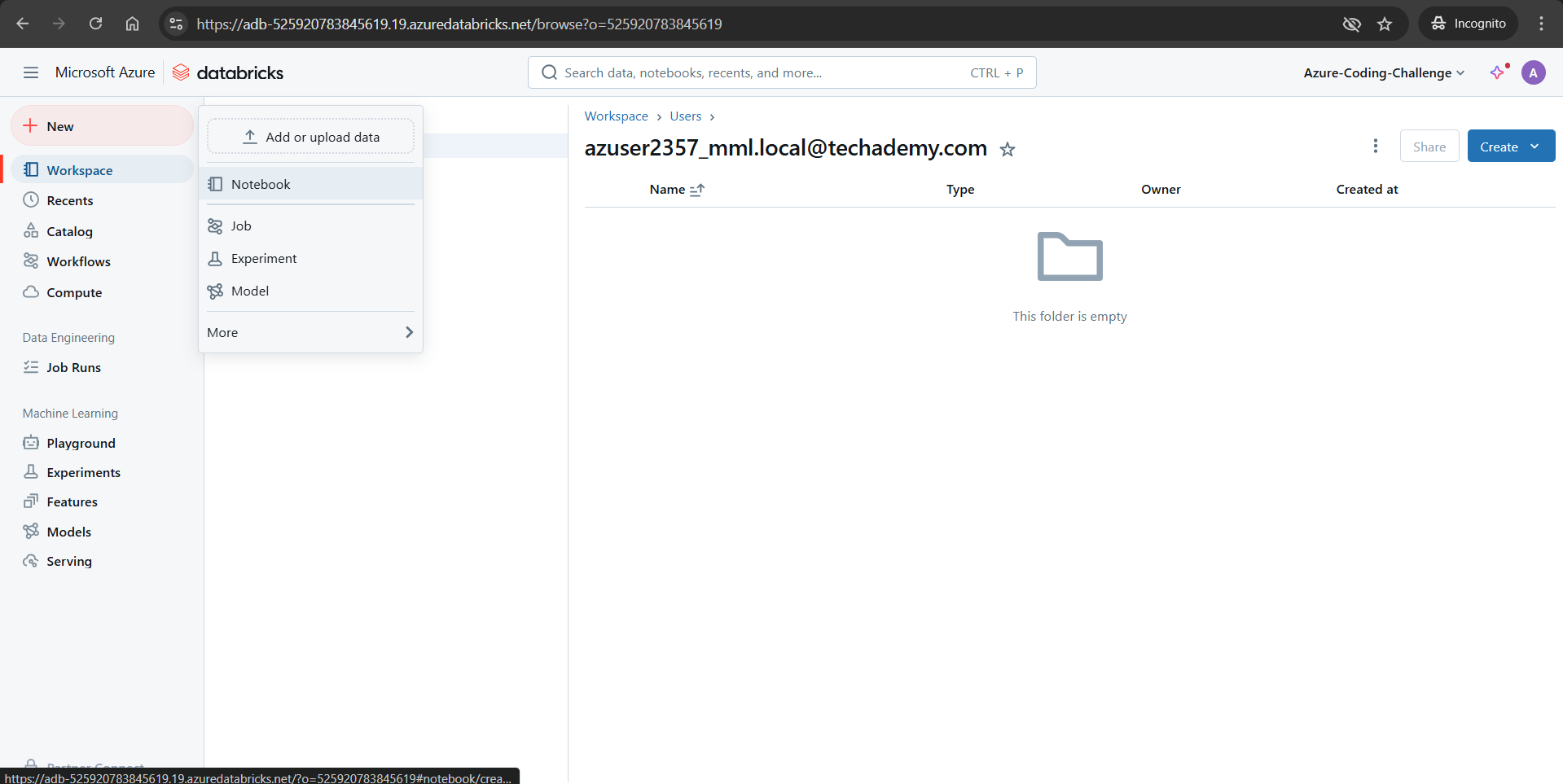


1. Click **Create Cluster** to start provisioning.

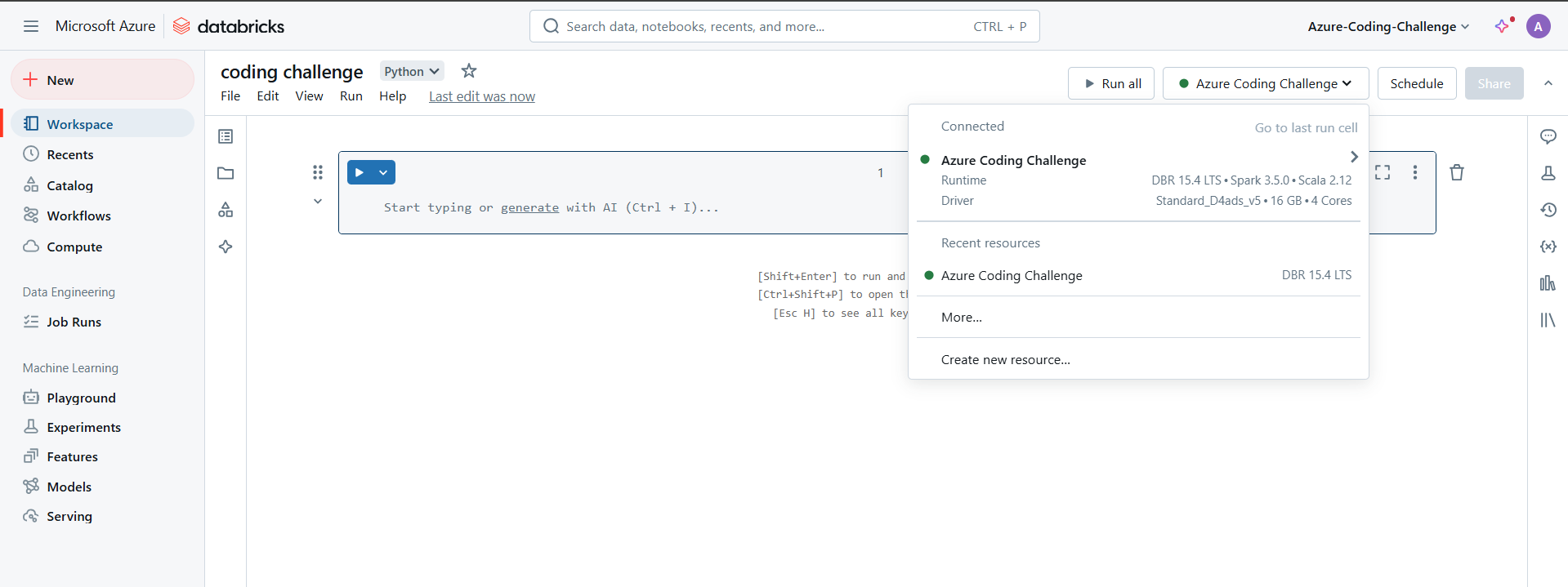
****

**Step 3: Creating and Attaching a Notebook**

1. In your Databricks workspace, click the **New** button on the top left of the screen.
2. From the dropdown, select **Notebook**.

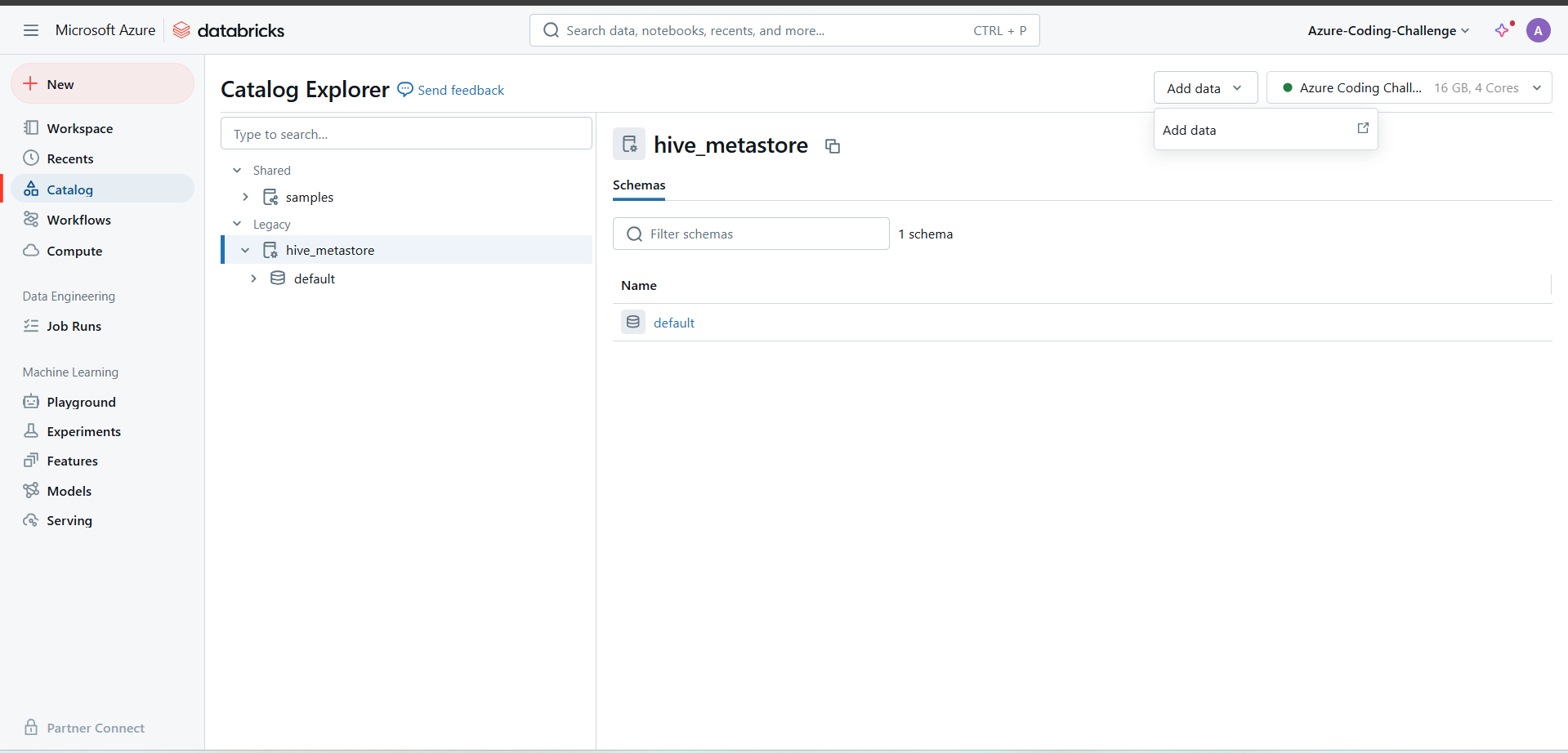


1. Provide a name for the notebook.
2. Choose the **Default Language**.
3. Click **Create** to open the notebook.
4. To attach the notebook to the cluster:
   * In the top left corner of the notebook, you'll see **Detached** next to the cluster name.
   * Click on **Detached** and select the cluster you created earlier. It will change to **Attached** once connected.

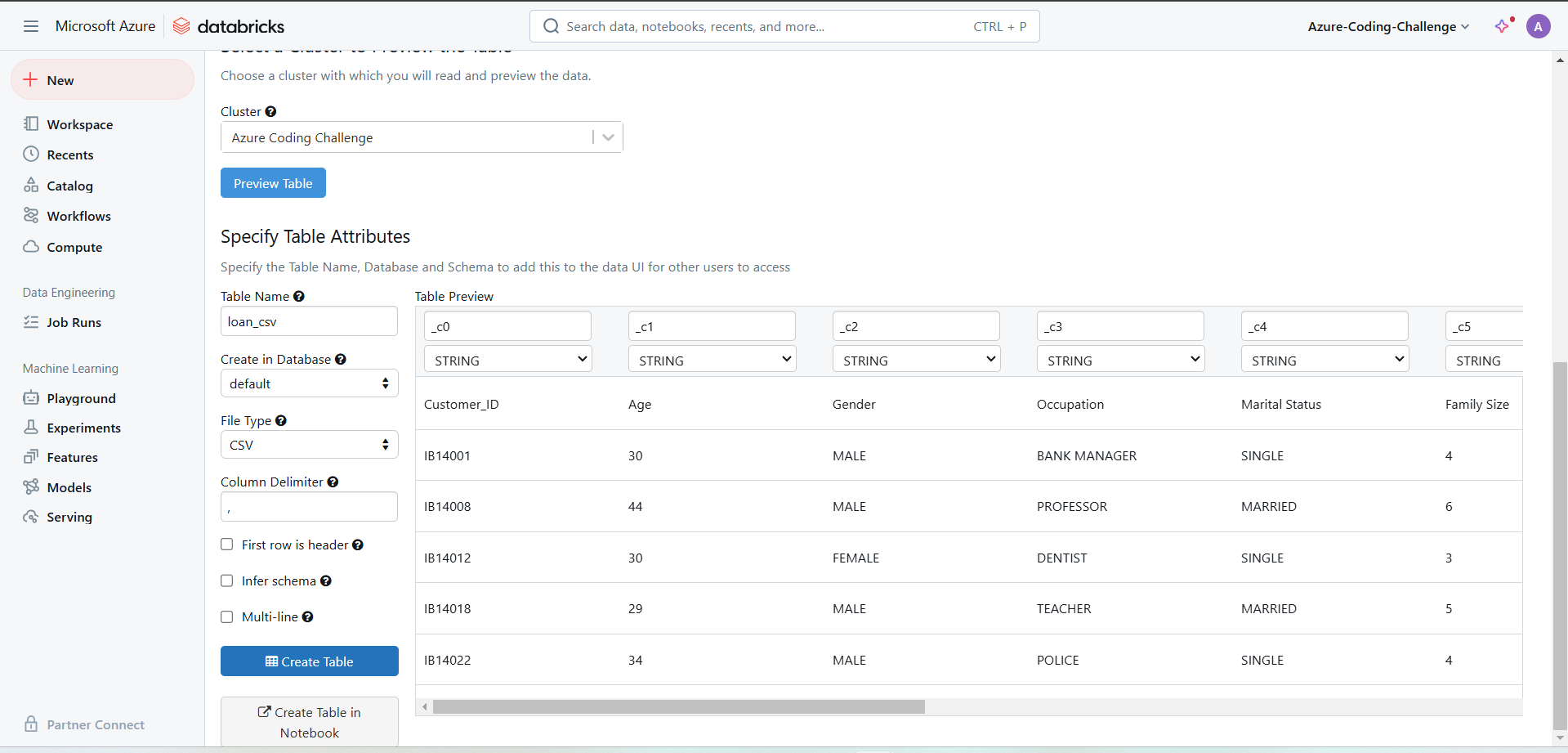


**Step 4: Adding a CSV File as a Delta Table in the Catalog**

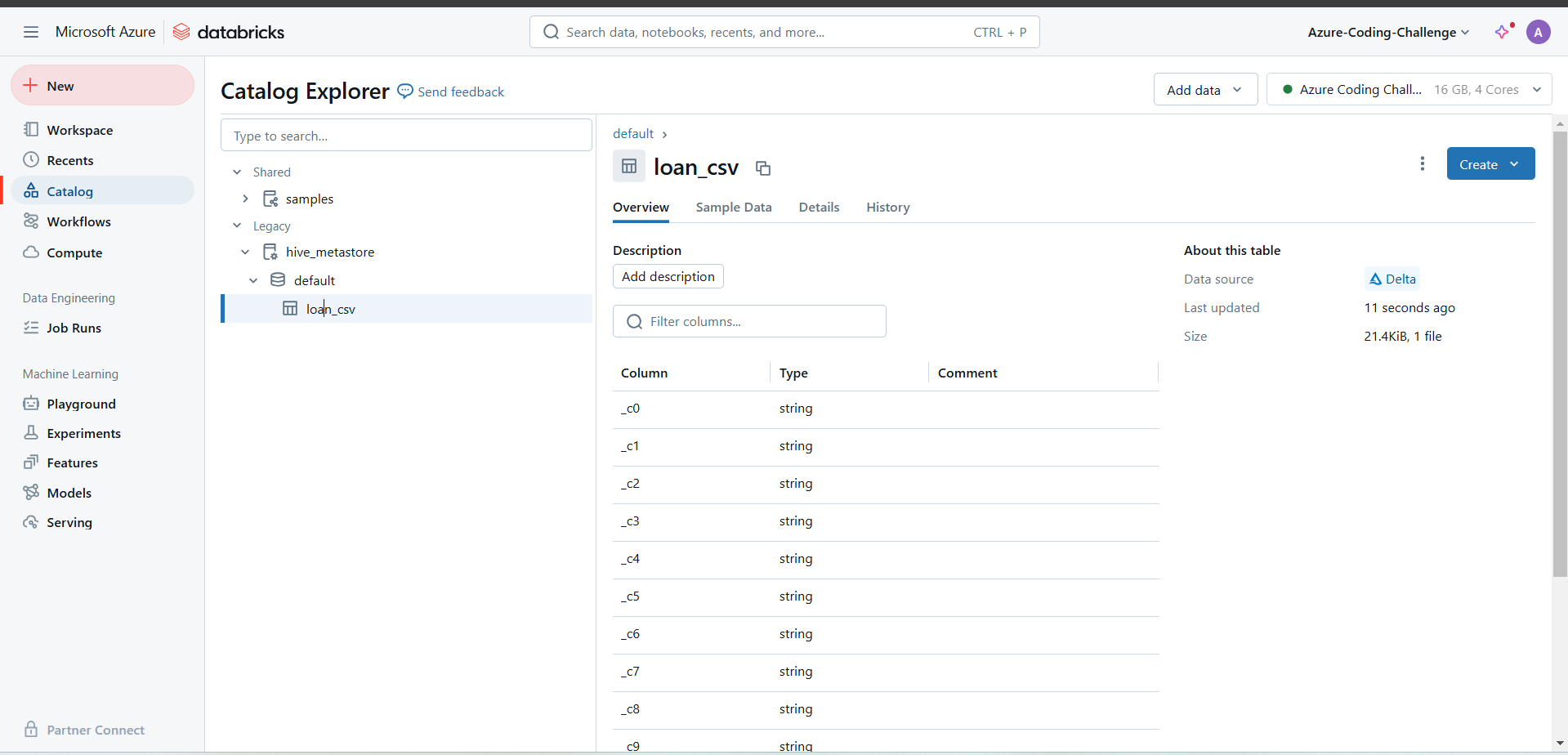
1. **Upload the CSV file**:
   * In your Databricks workspace, click on the **Catalog** tab on the left sidebar.
   * Click on **Add Data**.



* + Select **Upload File** and choose the CSV file from your local machine to upload it.

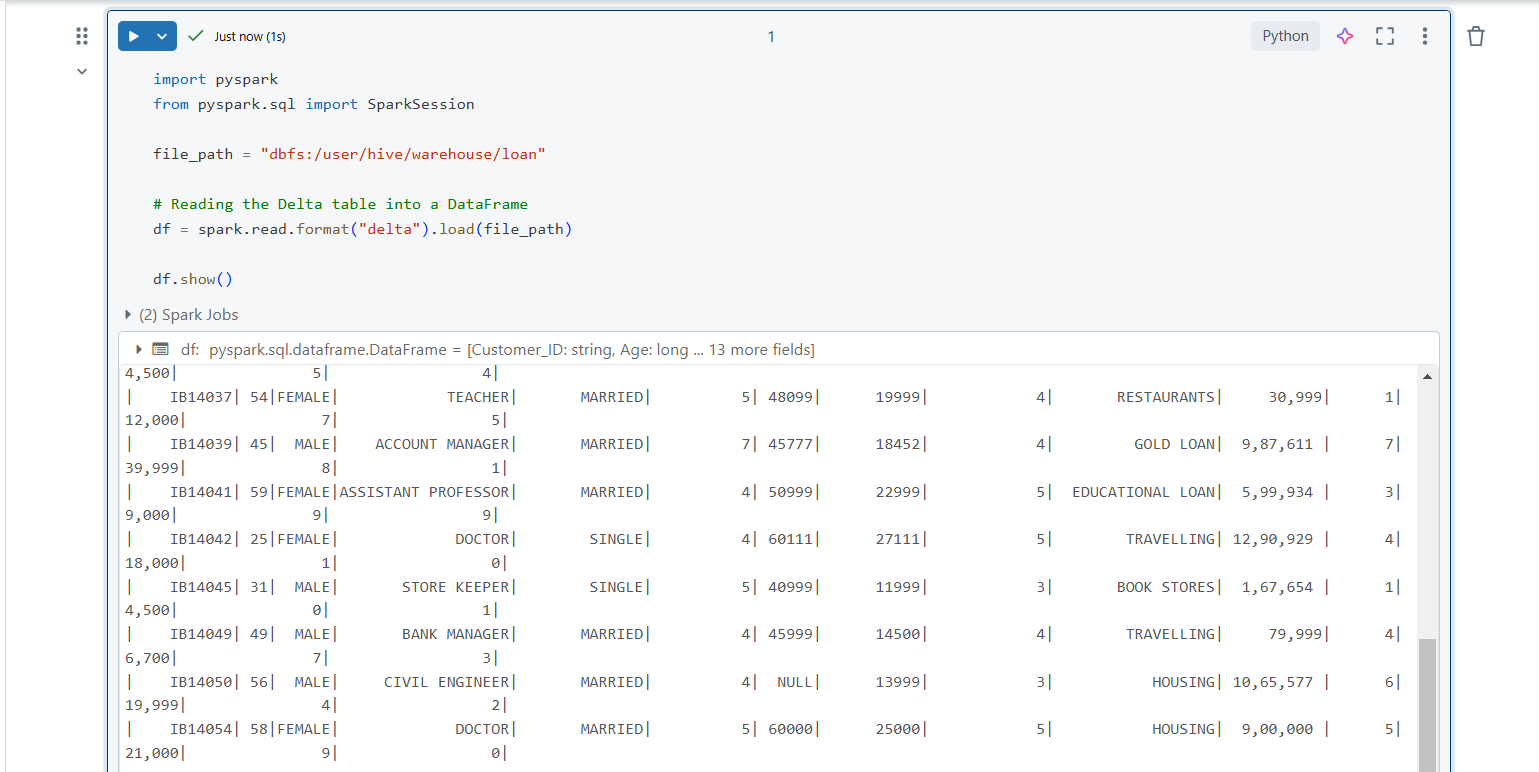


* + Click **Create** to create the Delta table.



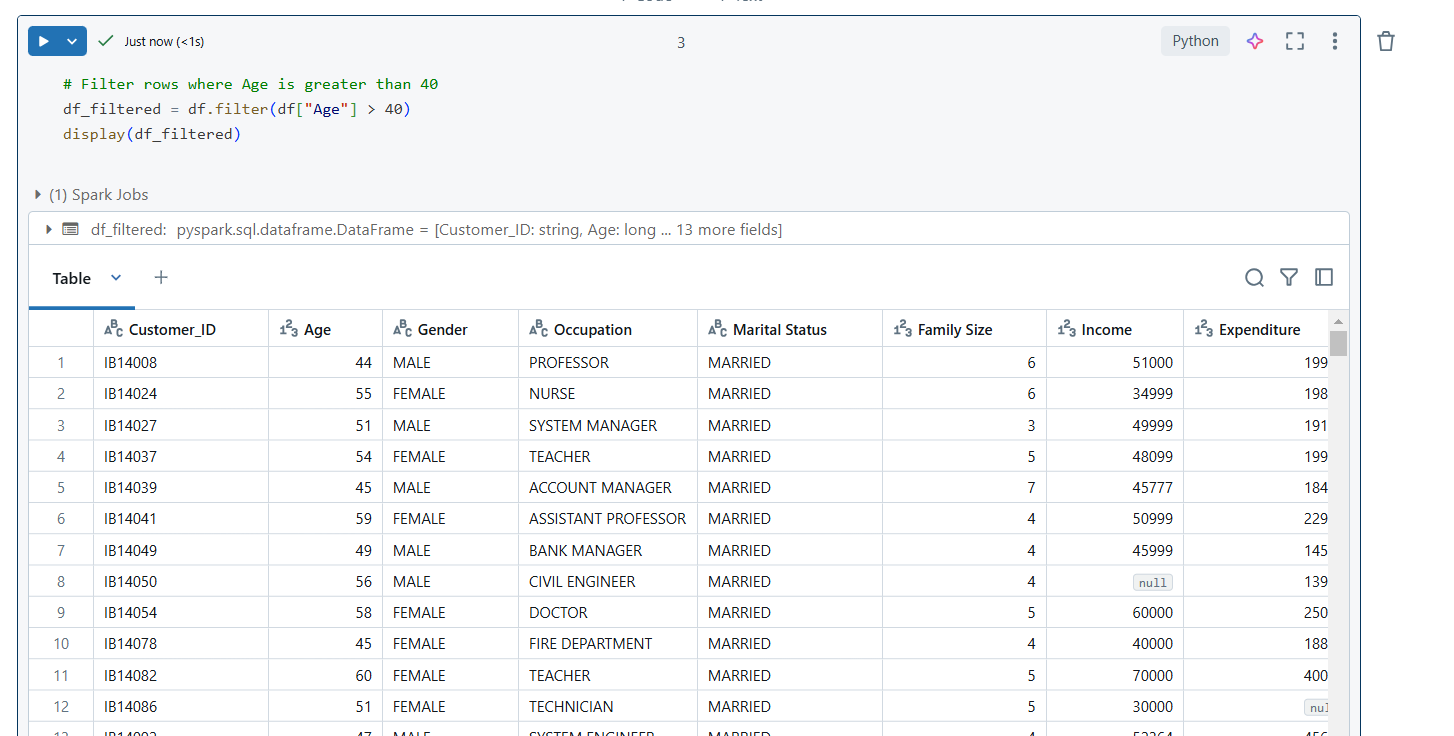
**Step 5: Creating a DataFrame from the Delta Table**

This will create a DataFrame (df) from the Delta table and display the first few rows to confirm

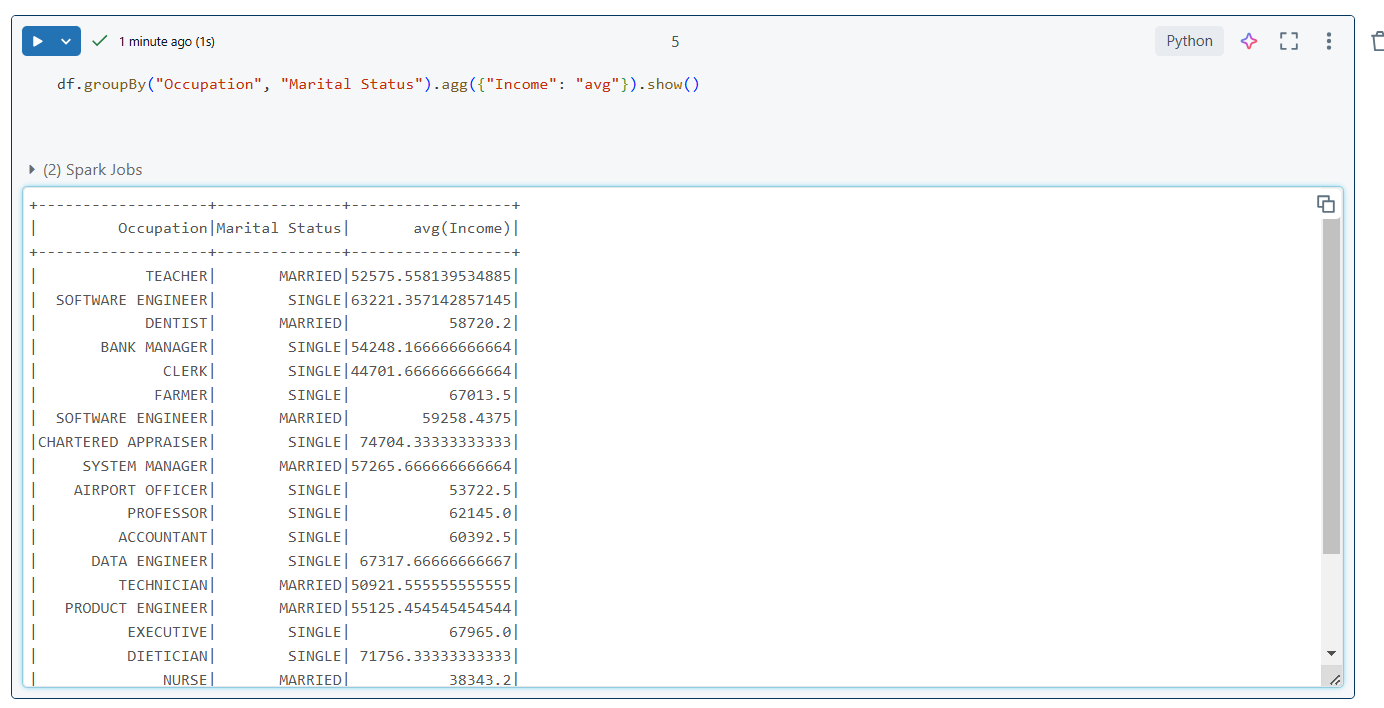


**Step 6: Commands for DataFrame (df) Operations**

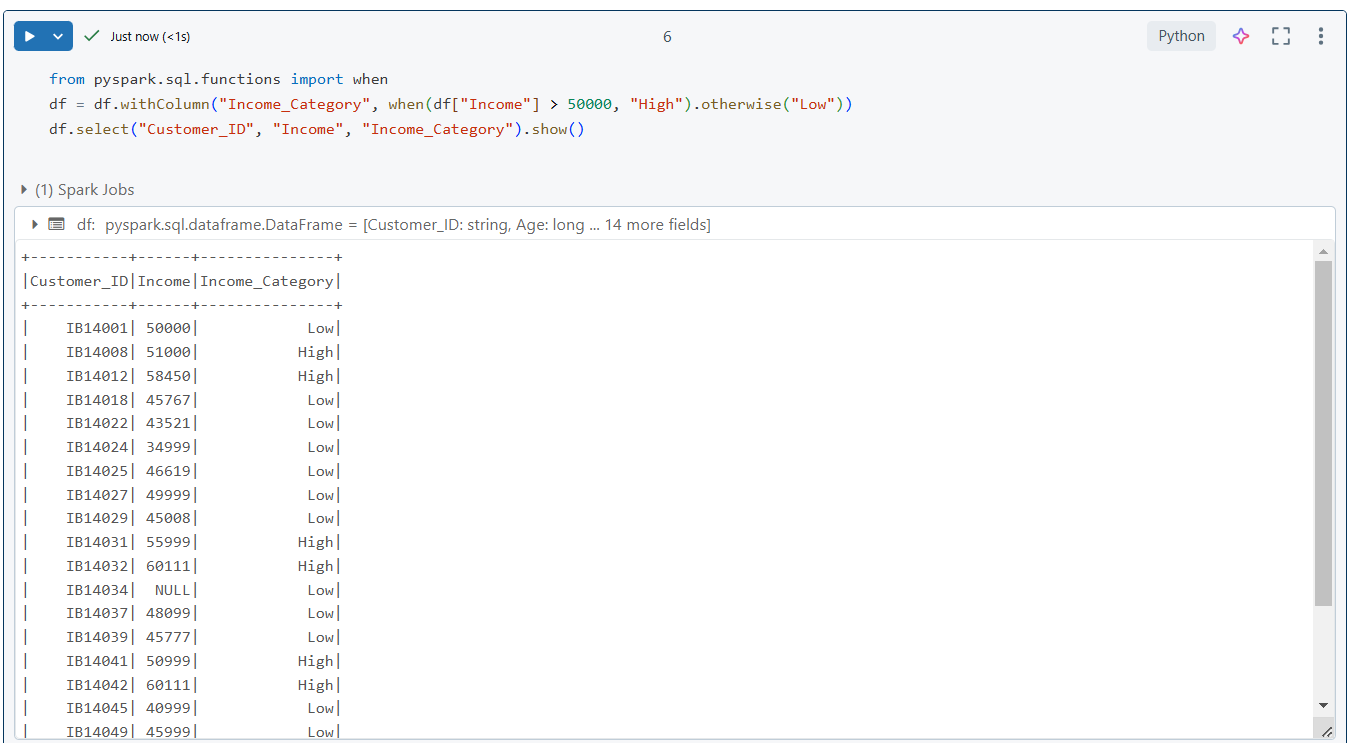
1. **Filter rows where Age is greater than 40**:



1. **Calculate the average Income for each Occupation and Marital Status**:



1. **Add a new column Income\_Category based on Income**:



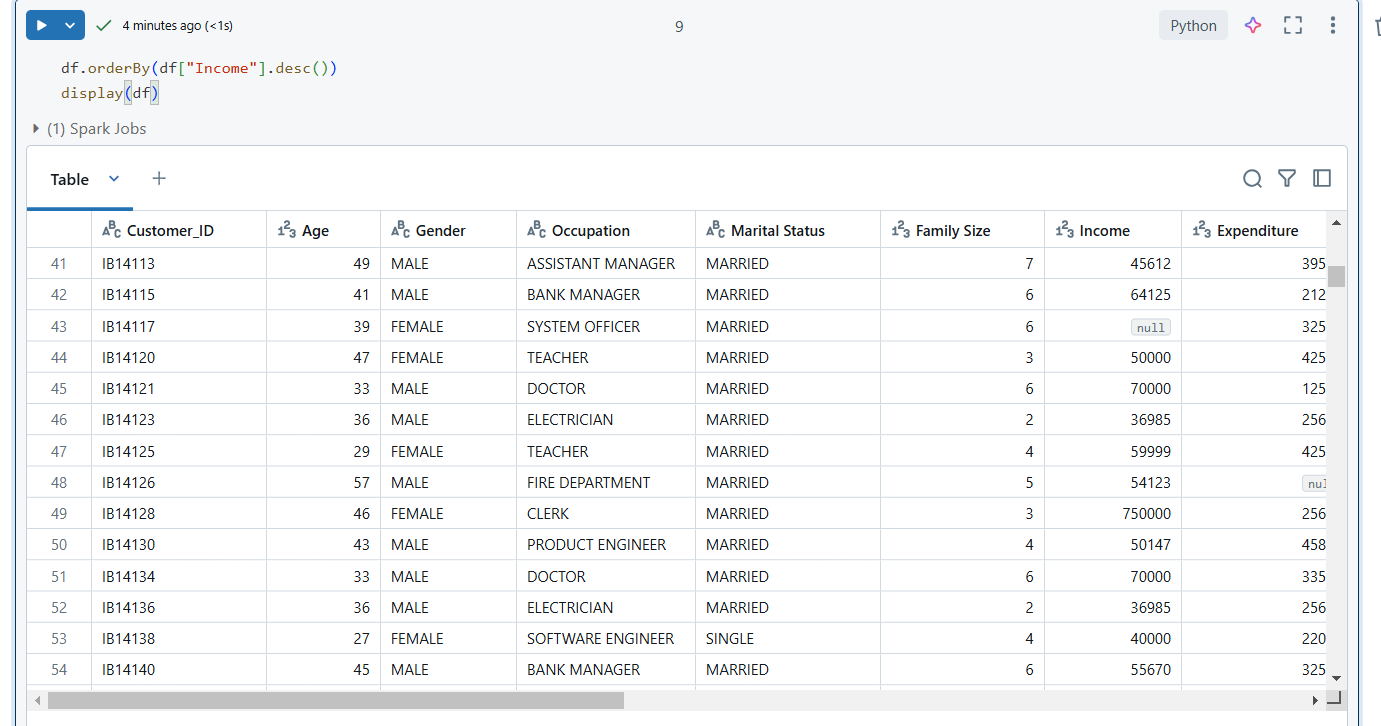
1. **Find the maximum Loan Amount for each Loan Category**:



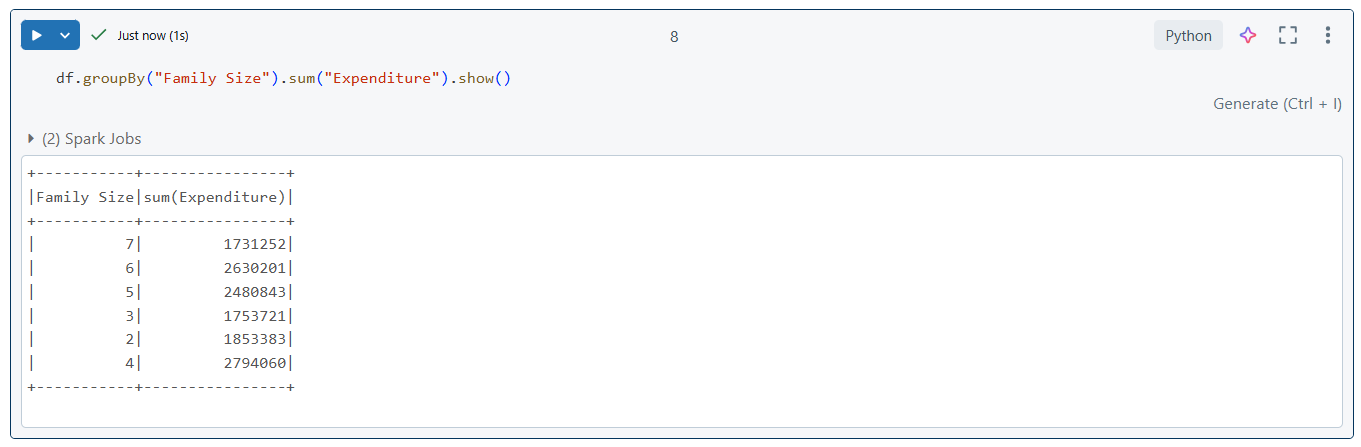
1. **Finding Overdue Records grater then 5**



1. **Sort the DataFrame by Income in descending order**:

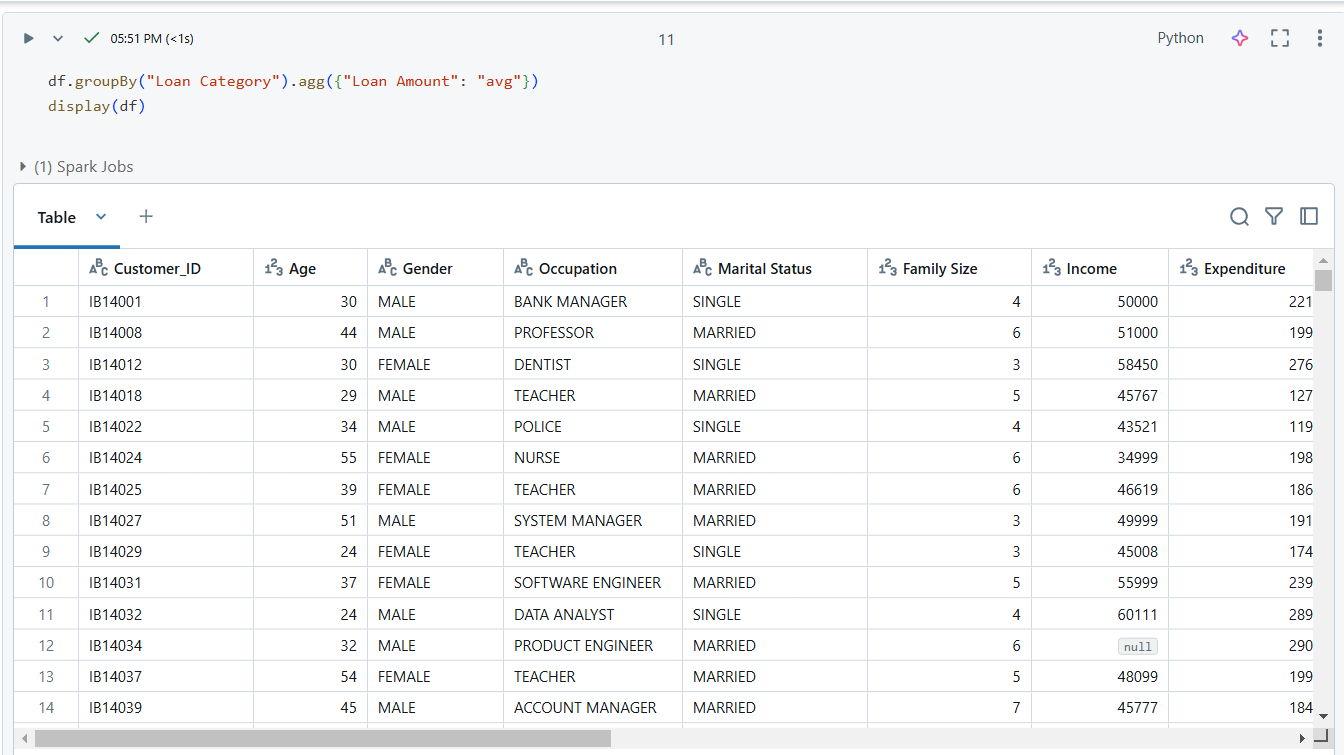


1. **Calculate the total Expenditure for each Family Size**:



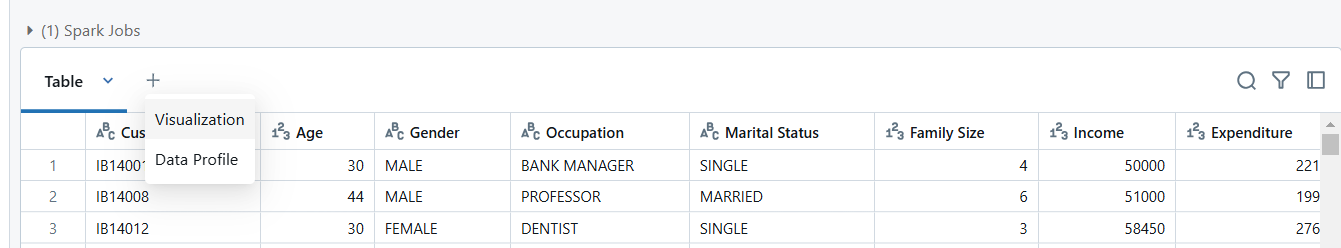
**Step 7: Creating Visualizations in Databricks**

1. **Start by selecting a cell in your notebook** where you want to create a visualization (you can use the output of any DataFrame operation, such as diasplay().
2. **Run a command to display the data you want to visualize**.

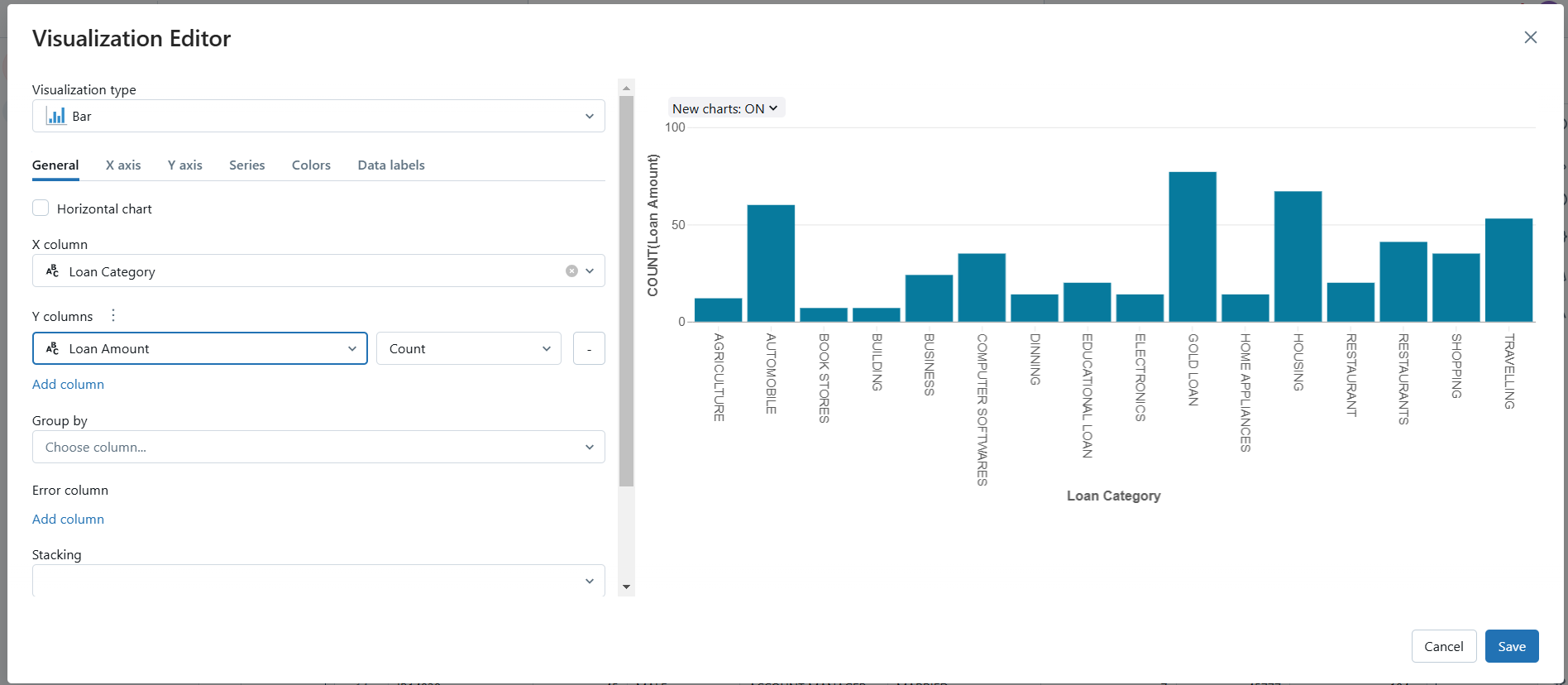


This will show the average loan amount per loan category.

1. **Click the “+” icon** at the top-right of the cell output
2. **Select “Visualization”**. A visualization pane will appear on the right side of the screen.



1. **Choose a chart type**. You can select from various chart types like bar charts, line graphs, pie charts, etc.
2. **Configure the chart**.
   * For a **bar chart**, for example, set Loan Category as the **X-axis** and avg(Loan Amount) as the **Y-axis**.
   * You can customize the visualization settings (titles, axis labels, etc.) for better clarity.



1. **Click on “Apply”** to render the visualization in the notebook.

