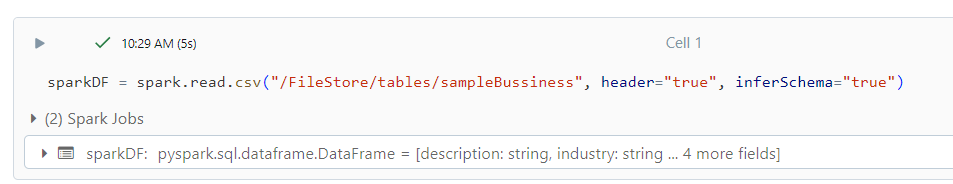
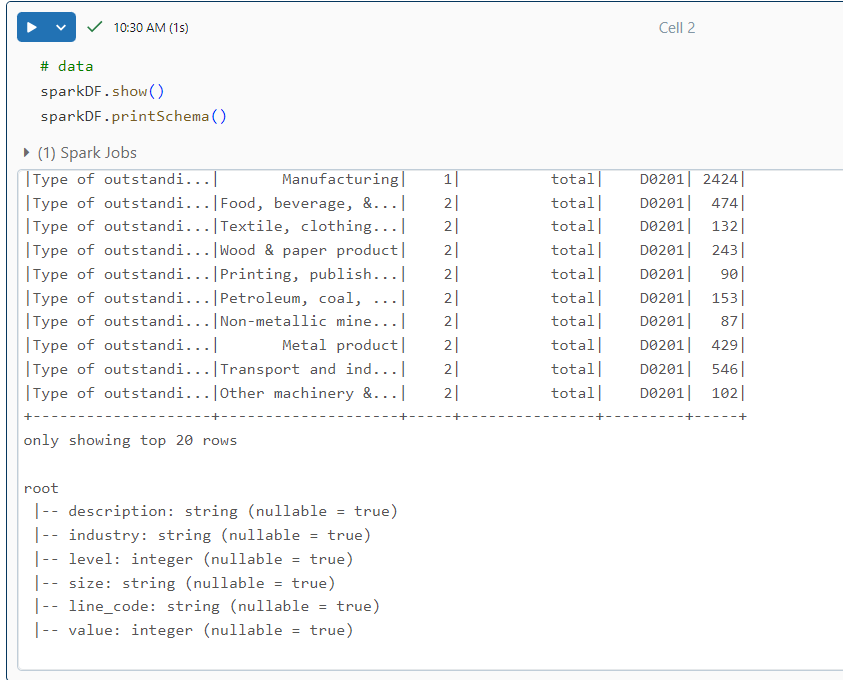
**1.Exploratory data analysis (EDA) in Databricks :**

To perform the EDA analysis in databricks first we read a csv file from the catalog present. We converted into a spark dataframe to make the analysis.

Here we are reading the CSV file and viewing the file along with its schema.



Commands to view the table and schema.



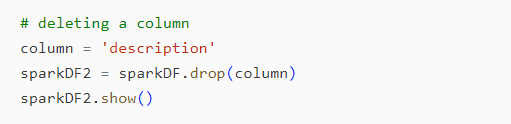
We are performing the data filtering by using group by, we are grouping the data by column “Industry” and we are expecting the sum of “level” from the data.

**Input\Output:**

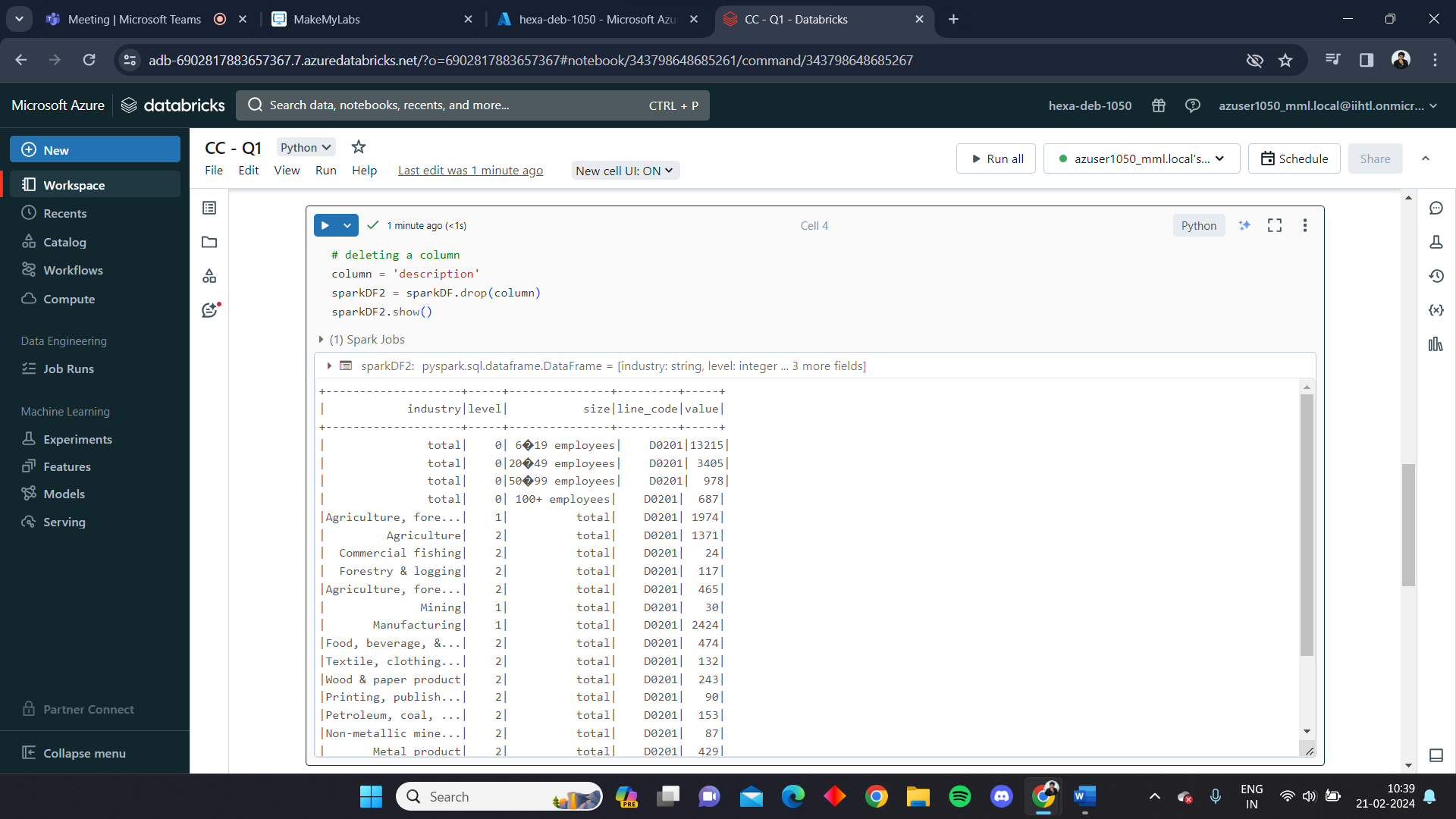


Here we are performing deletion operation by dropping the description column from the sparkDF and assigning it to the new sparkDF2 dataframe.

**Input:**



**Output:**

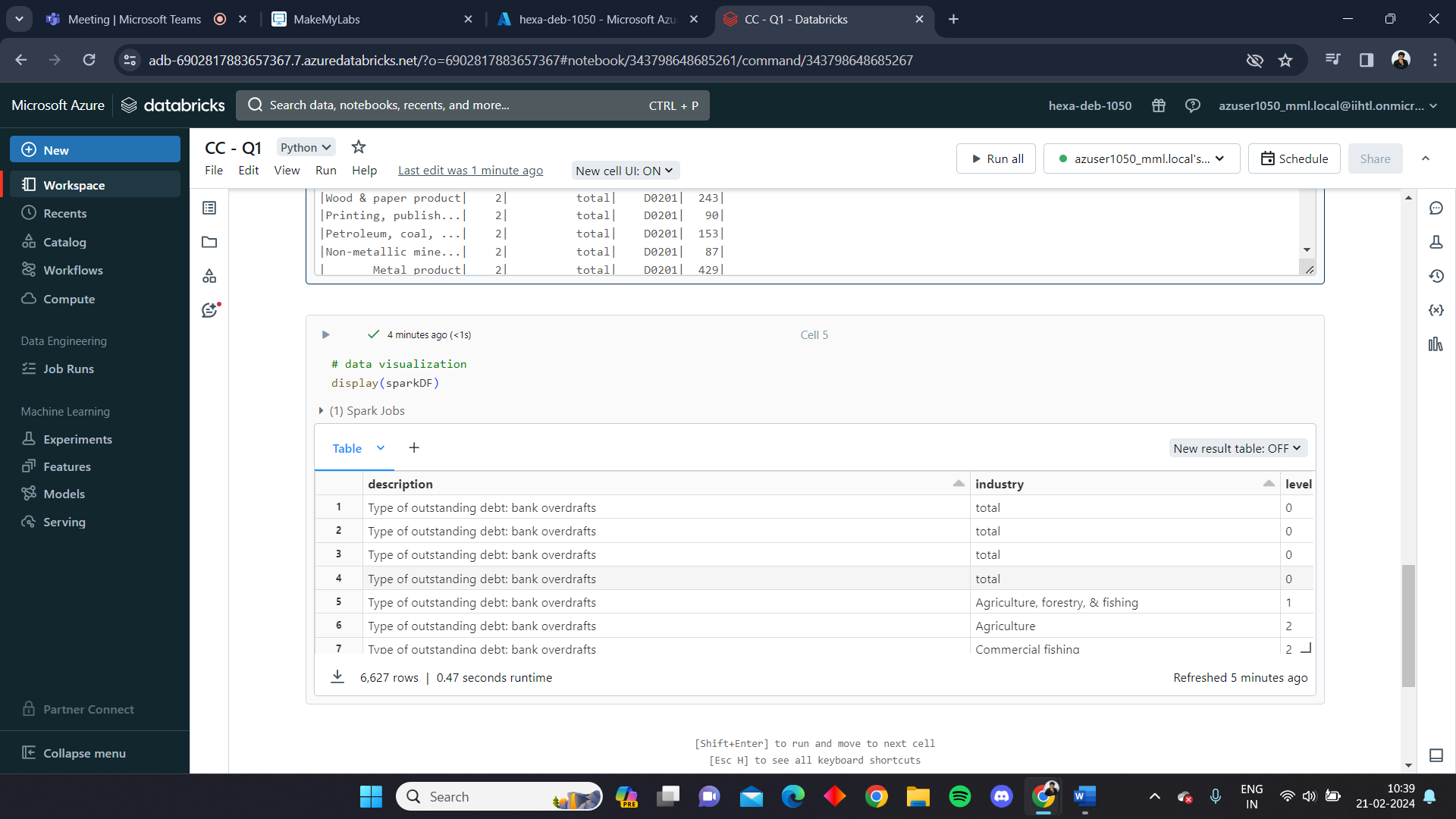


**Data Visualization:**

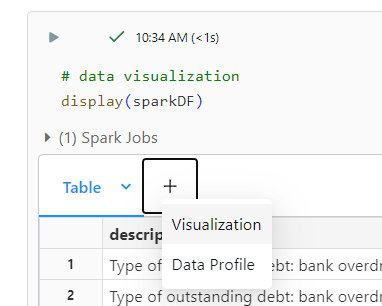
Data visualization is the presentation of data in a graphical or pictorial format to help people understand and interpret information more easily. Effective data visualization can reveal patterns, trends, and insights that might be challenging to discern from raw data alone. It plays a crucial role in data analysis, decision-making, and communication.

In Azure databricks we can visualize the data in many formats like graphs, pie charts, histograms and many more.

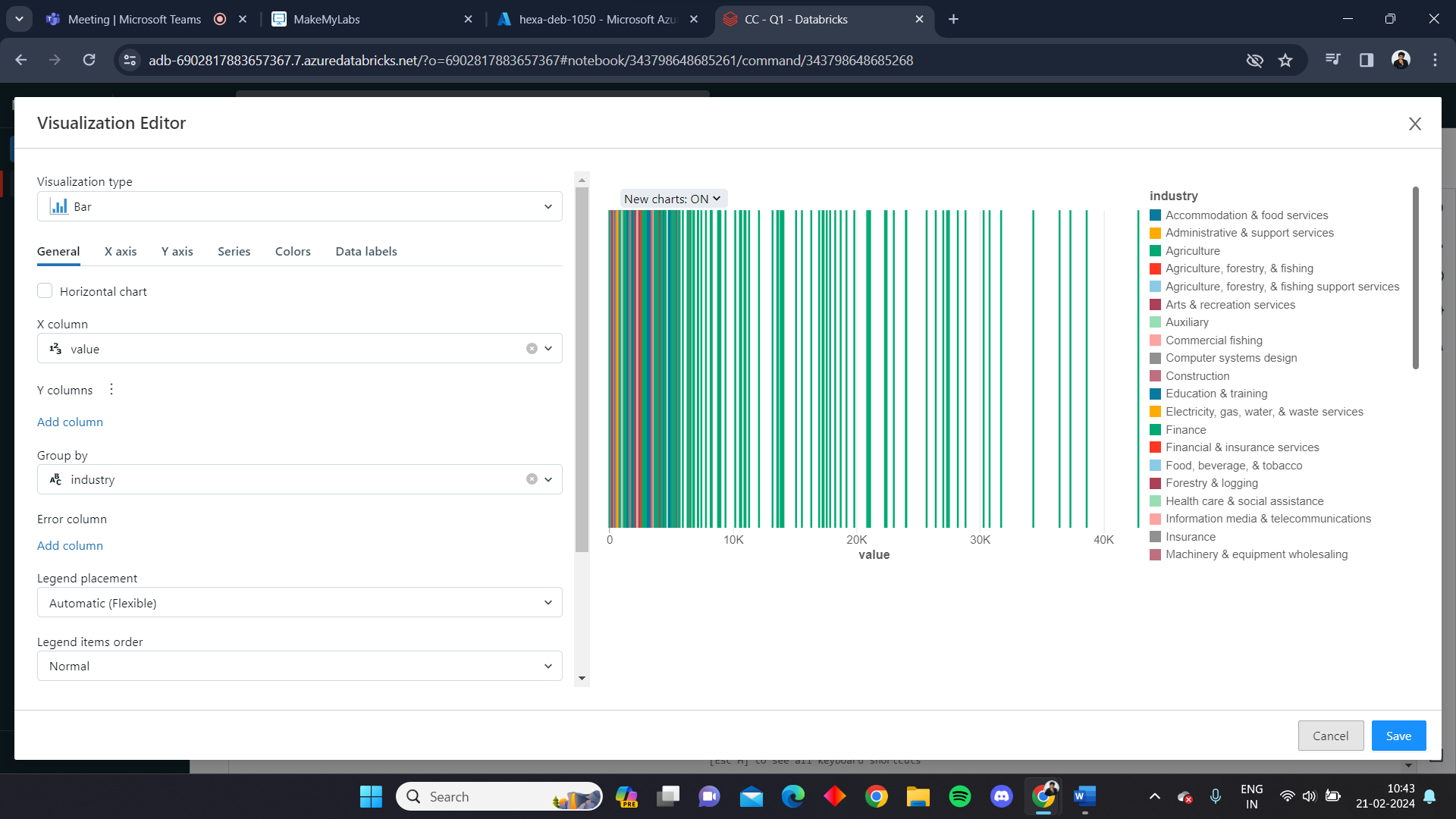
Here we are visualizing the data of the csv file



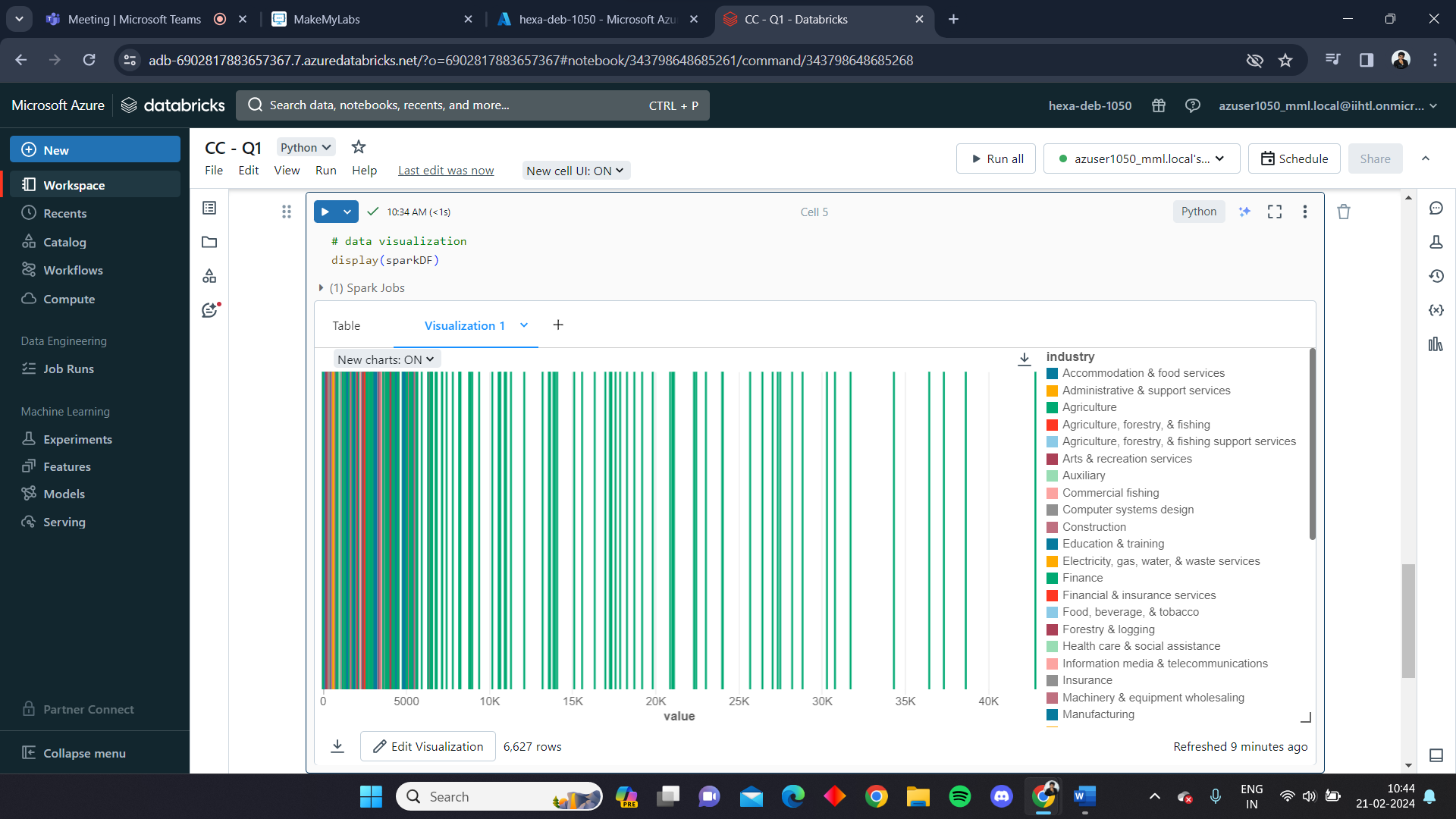
After running the display cell we can visualize the data using the visualization option



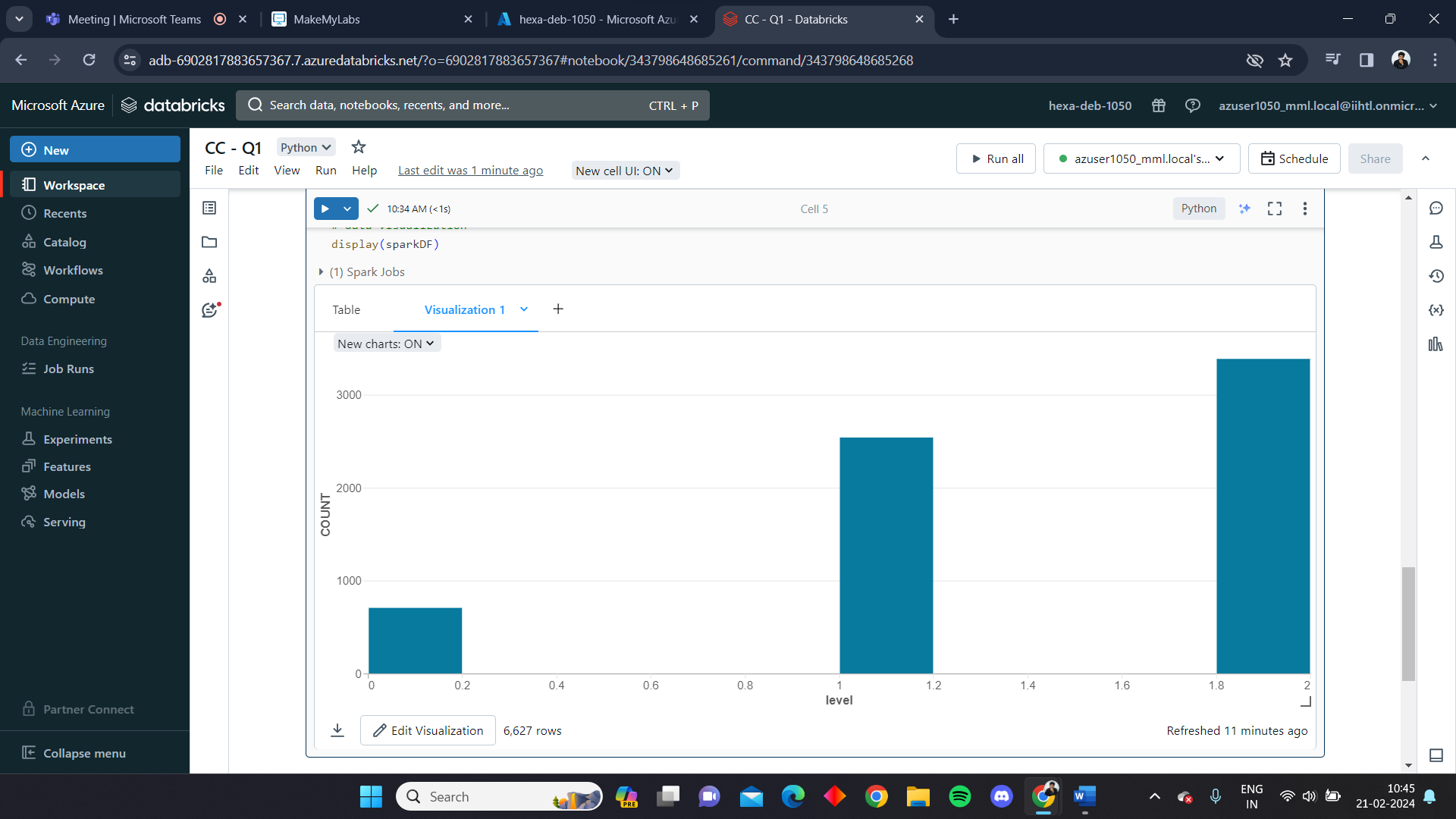
We can select the ways of visualization and we can also group them accordingly based on the requirement.



Here we are viewing the data in bar by value at x – axis, industry at y – axis and grouped by industry.



Here we are viewing the data in bar graph by level at x – axis, count at y – axis and grouped by industry.



Here we are viewing the data in bubble graph by industry at x – axis, level and size at y – axis and grouped by industry.

