**VISUALIZATION IN**

**MICROSOFT**

**AZURE**

**S.R.TAANUSRI**

**30.12.2023**

1. Compute Summary Statistics:

Use functions like describe() or summary() to generate statistical summaries of the data within your Data Frame.

1. Collect Data:

Collect the summary statistics data from PySpark Data Frame to the driver node.

1. Visualize:

Use external Python visualization libraries such as Matplotlib, Seaborn, or others to create visualizations based on the collected summary statistics.

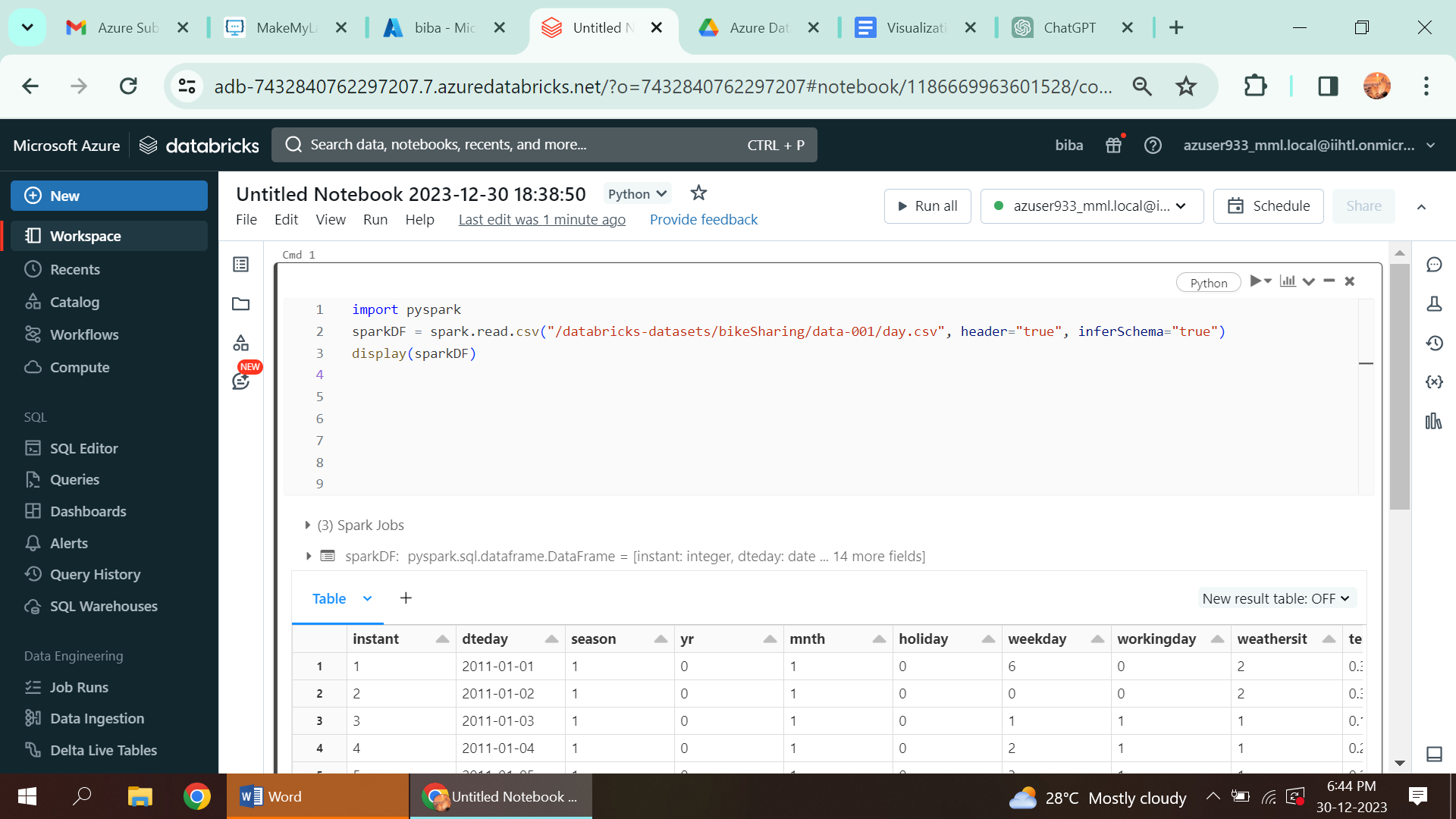
1. Categorical Data Visualization:

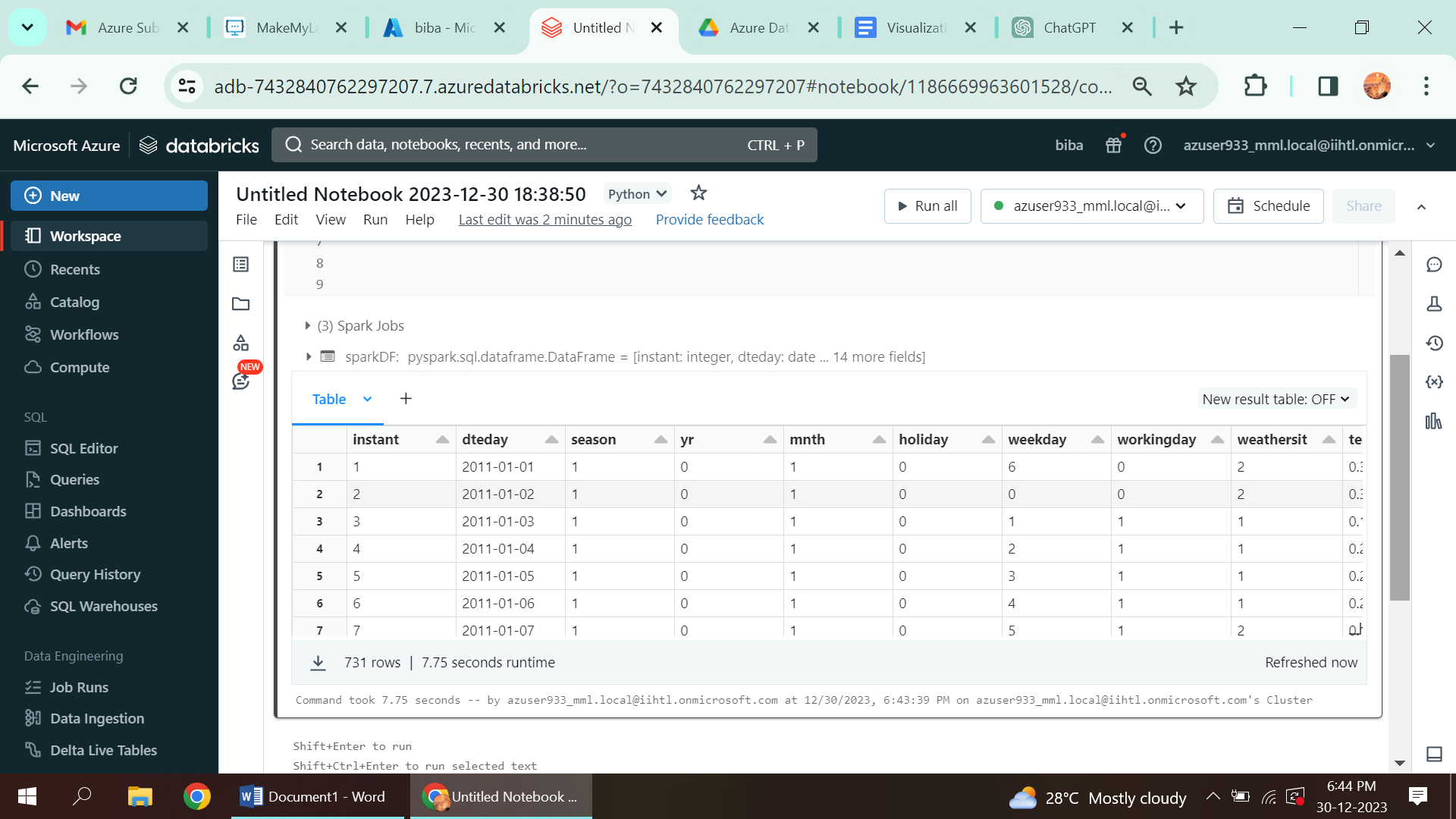
For categorical columns, visualize the frequency distribution using bar plots or pie charts. You can count the occurrences of each category using PySpark's groupBy() and count() functions.

1. Documentation and Interpretation:

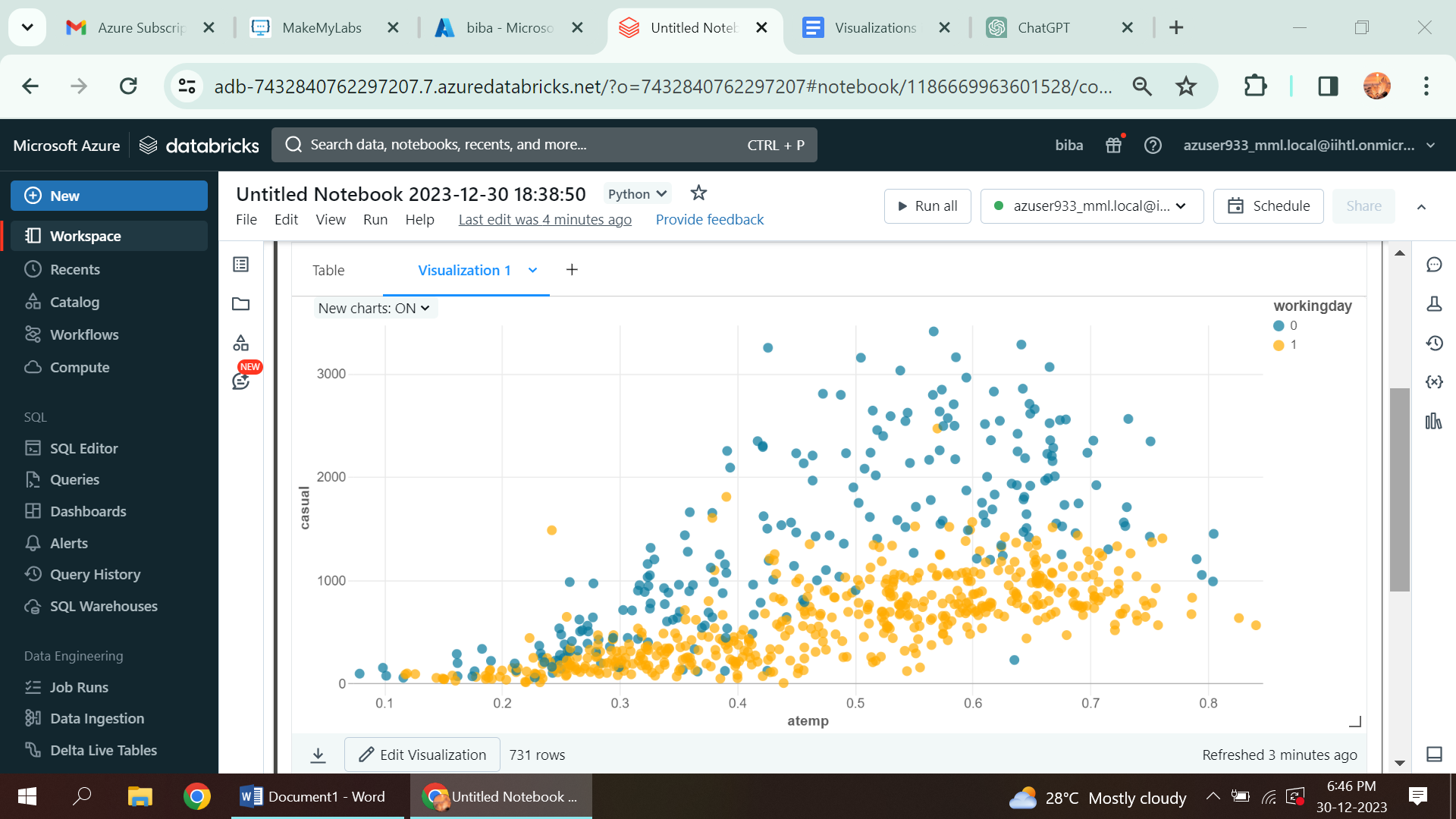
Document the visualization summaries by including titles, labels, legends, and descriptions to make them understandable and interpretable for others who might be analyzing the data.

1. Create a new visualization

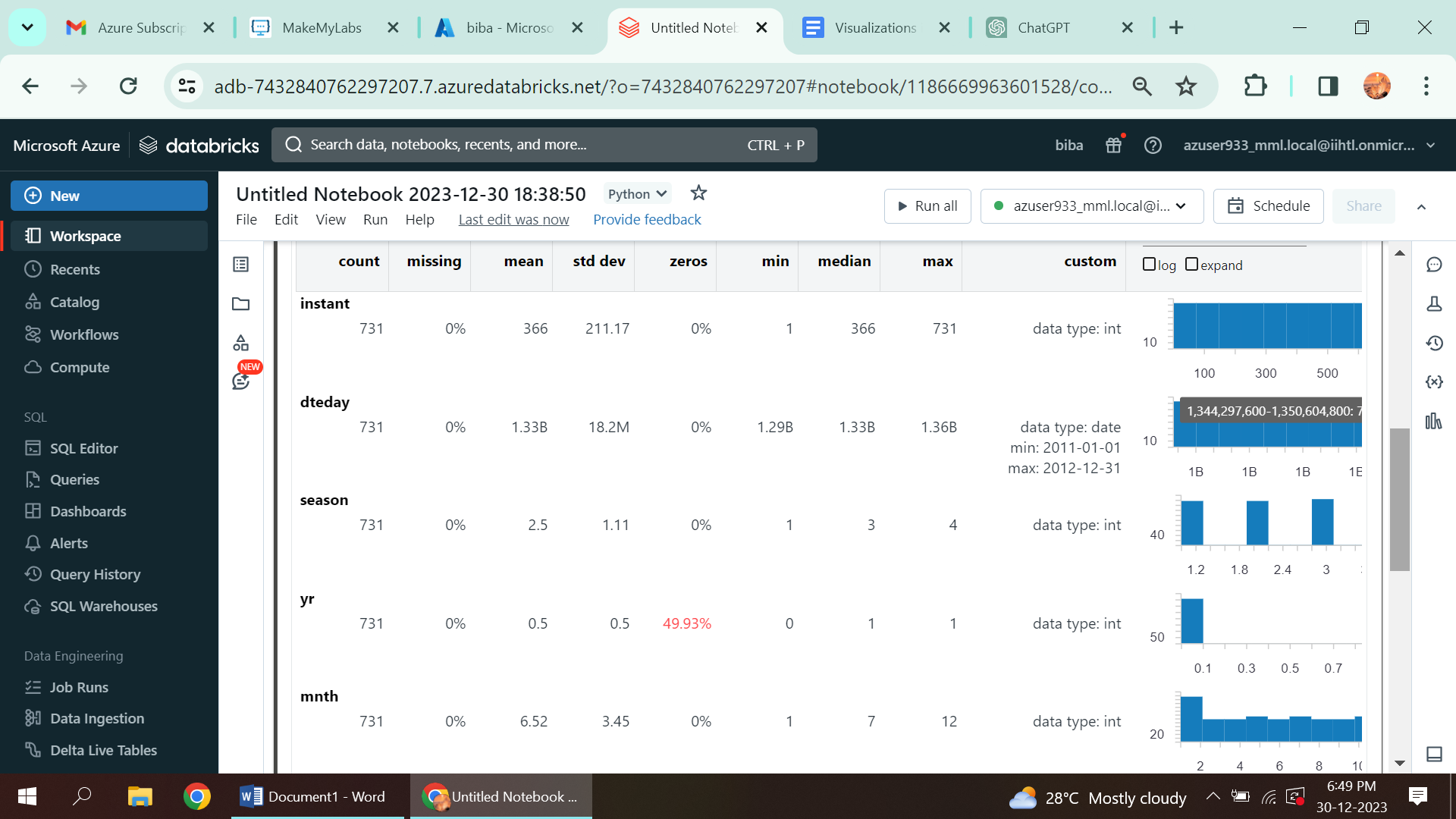


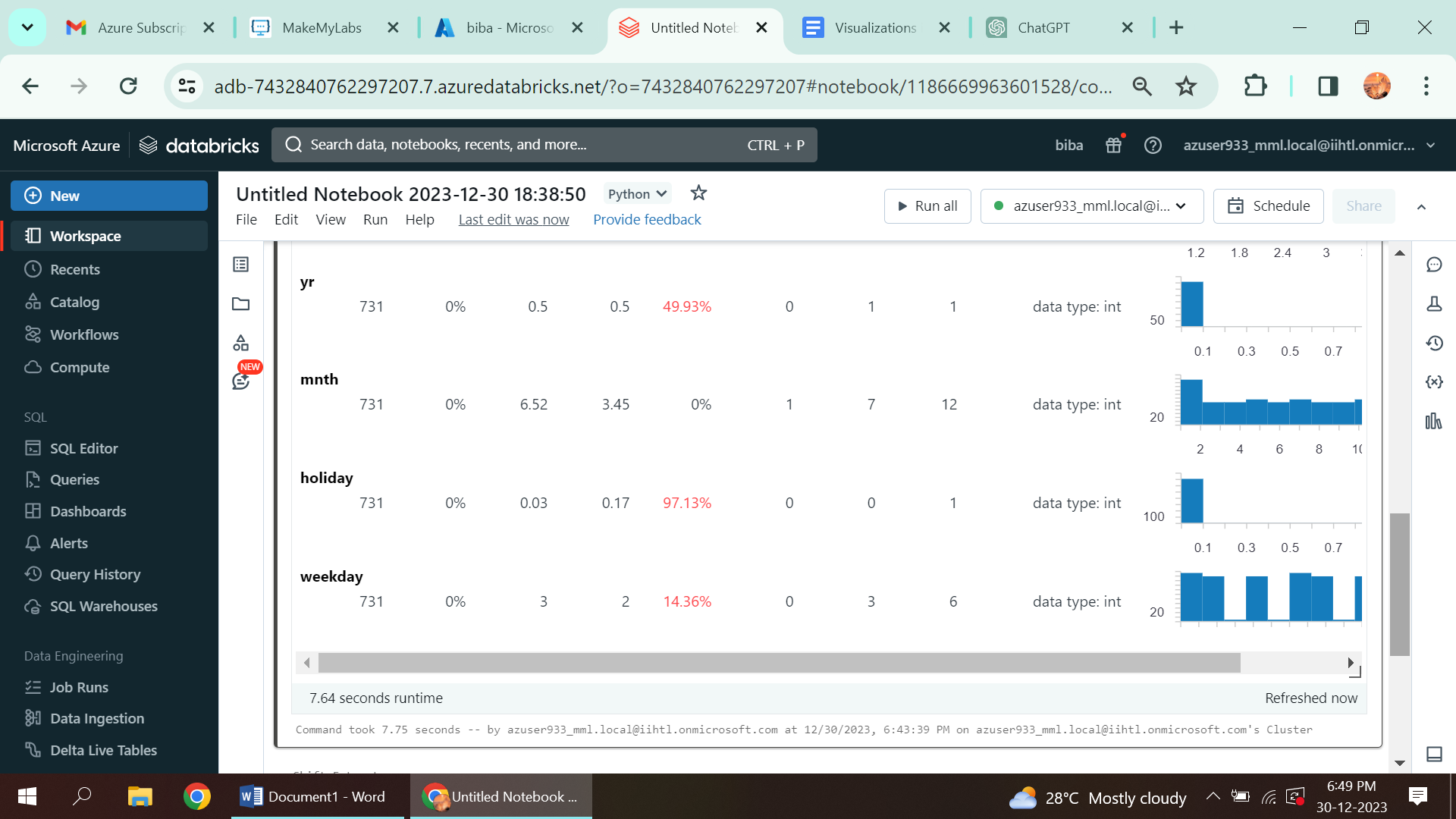


VISUALIZATION

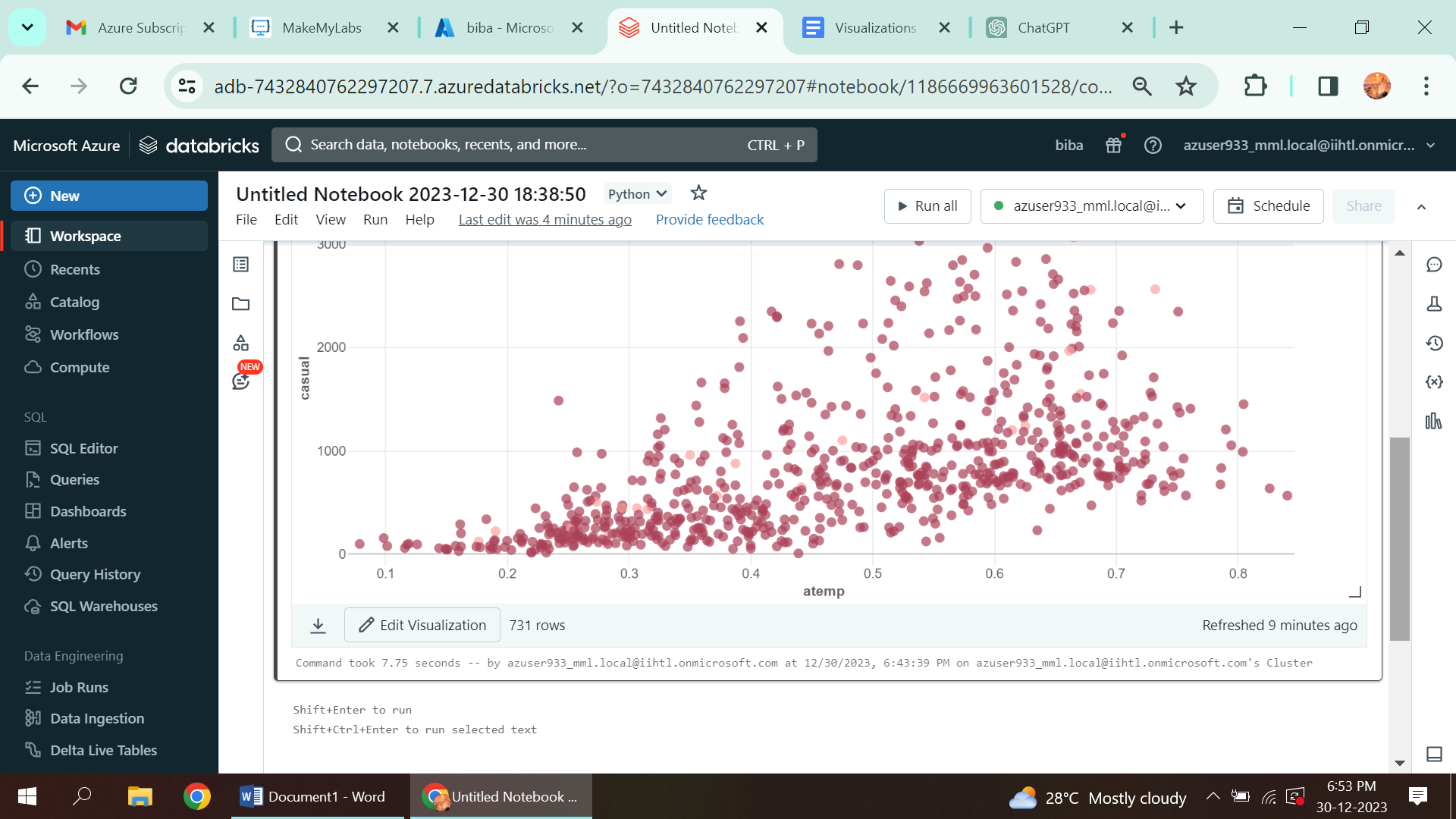


2. Create a new data profile





3.Editing the visualization



4.Changing the color of visualization and renaming the data profile name

