**Azure Coding Challenge - 1**

**Name:** V Venkata Sri Prasad

**Batch:** Data Engineering

**Date:** 21/02/2024

Email : venkatvanniya2002@gmail.com

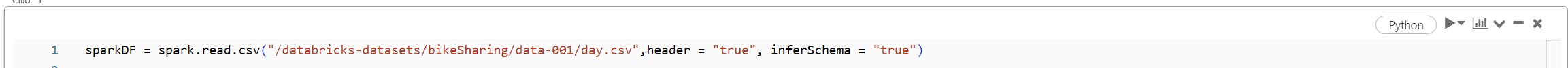
Question:

Exploratory data analysis (EDA) in Databricks &Visualizing data in Databricks using display keyword

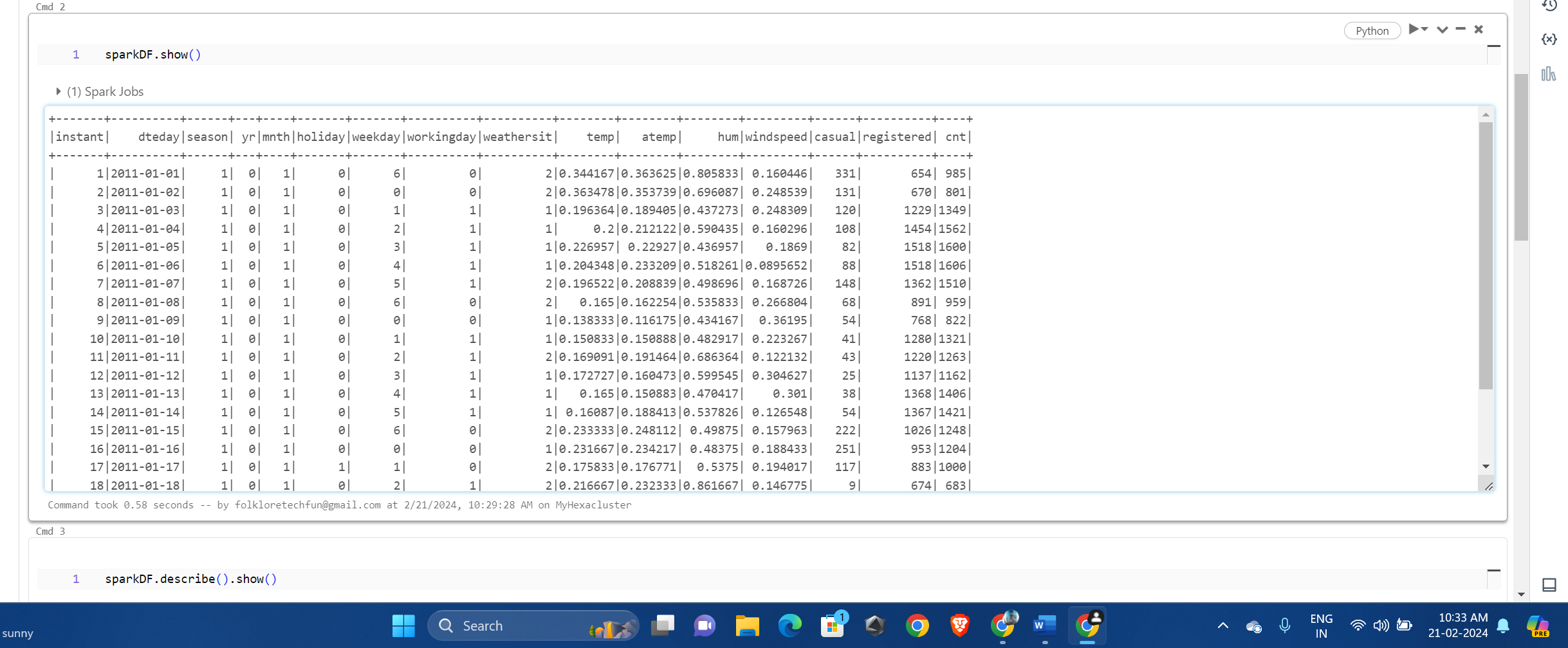
**Exploratory Data Analysis (EDA):**

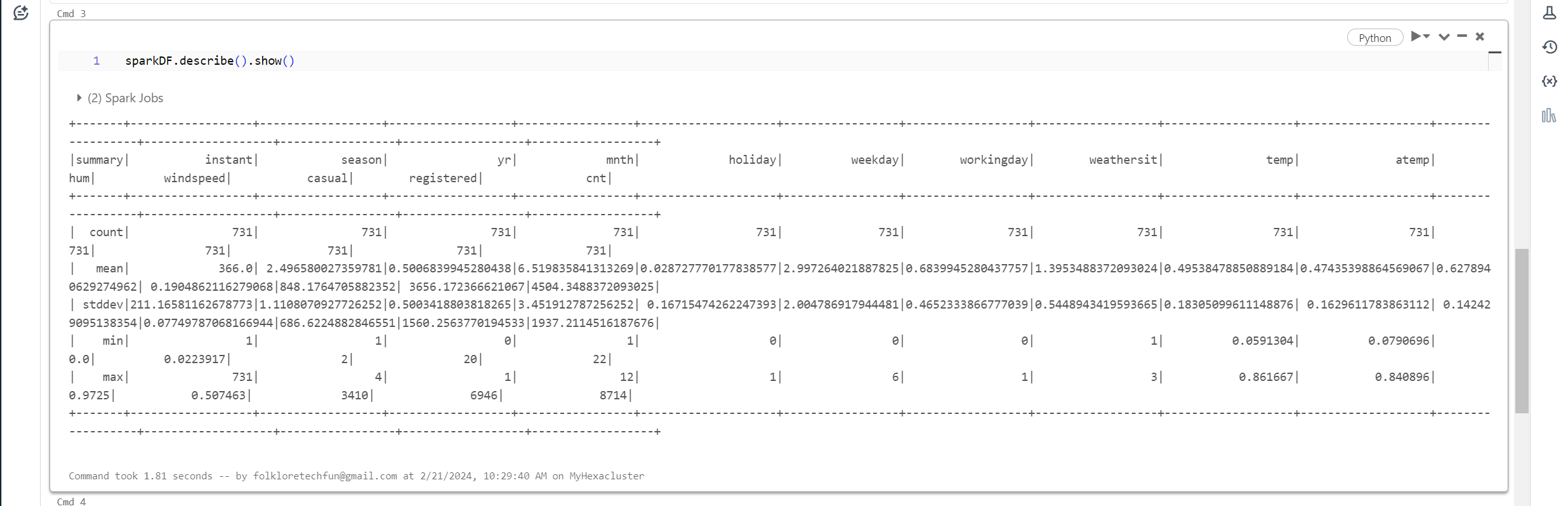
EDA in Databricks refers to the process of analyzing and exploring a dataset to understand its characteristics, uncover patterns, detect anomalies, and formulate hypotheses. Databricks provides a powerful platform for performing EDA using Spark, which allows for scalable data processing and analysis. Here's a basic outline of how you can perform EDA in Databricks:

Data Loading: Load your dataset into Databricks. You can load data from various sources such as CSV files, databases, data lakes, etc. For example:

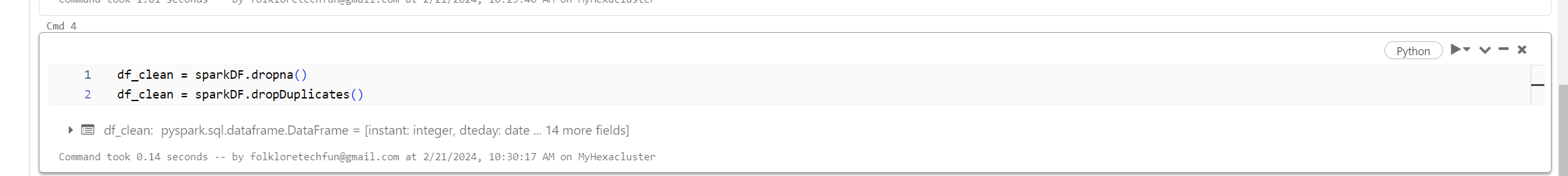


Data Understanding: Get an overview of your dataset using functions like display(), show(), describe(), etc.

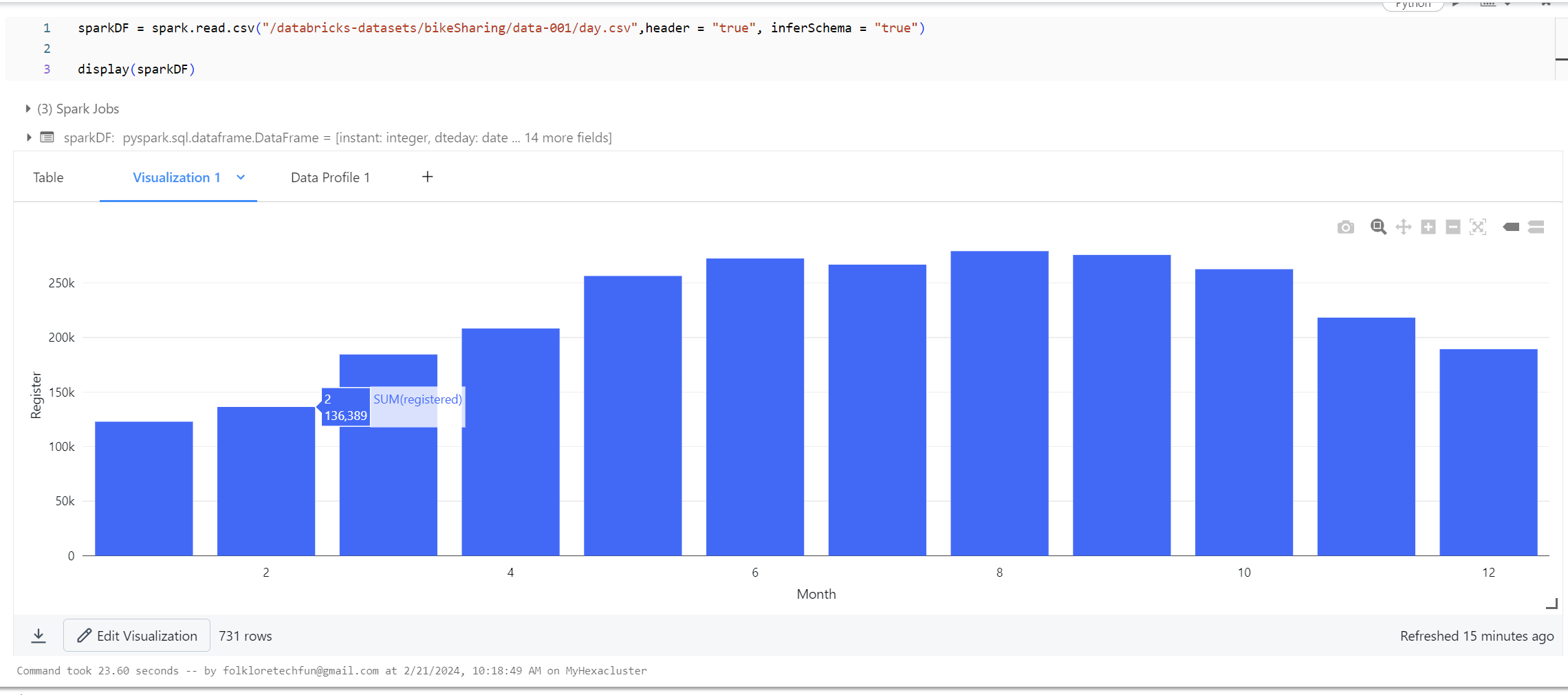


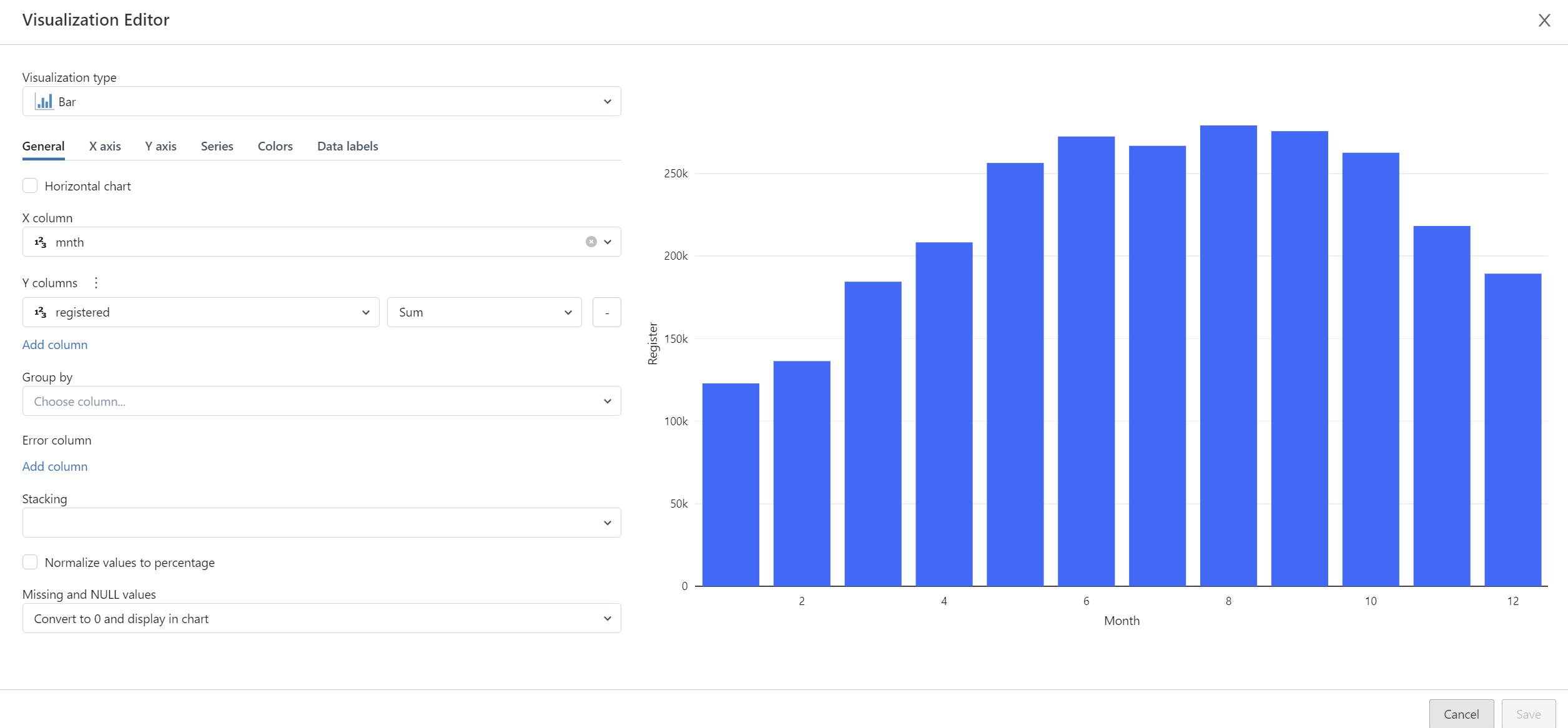


Data Cleaning: Perform data cleaning tasks such as handling missing values, removing duplicates, correcting data types, etc.

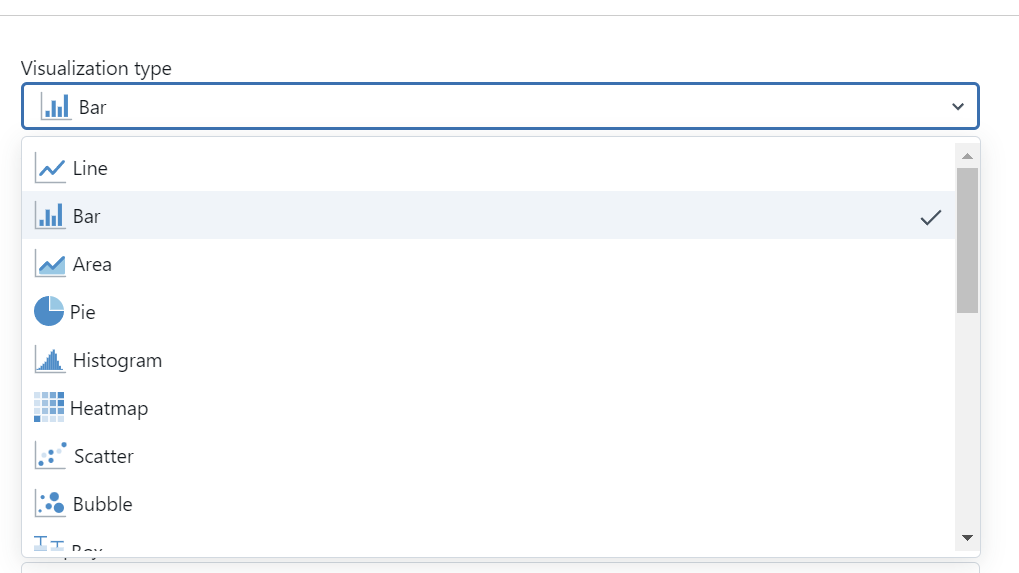


Data Visualization: Visualize your data to gain insights using the display() keyword in Databricks. This keyword allows you to create interactive visualizations directly within the Databricks notebook.

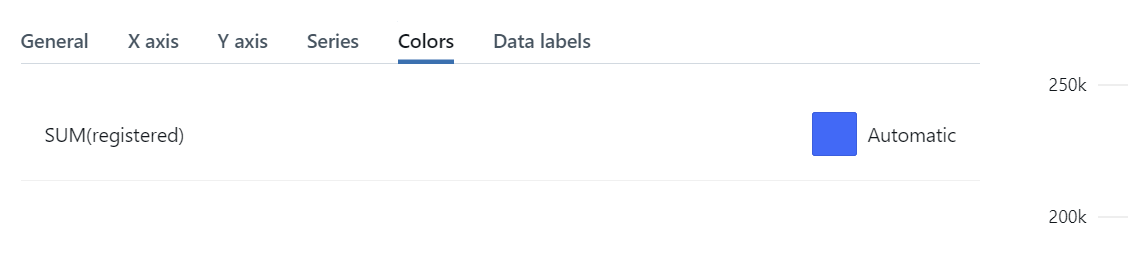




We can select graph type which is suitable for visualisation



Several features like colour , x axis , y axis are provided



Data Profile:

