# Big Data Analysis with IBM Cloud Database

## Phase-4  Development Part 2

Project 5 : Big Data Analysis


## Problem Statement:

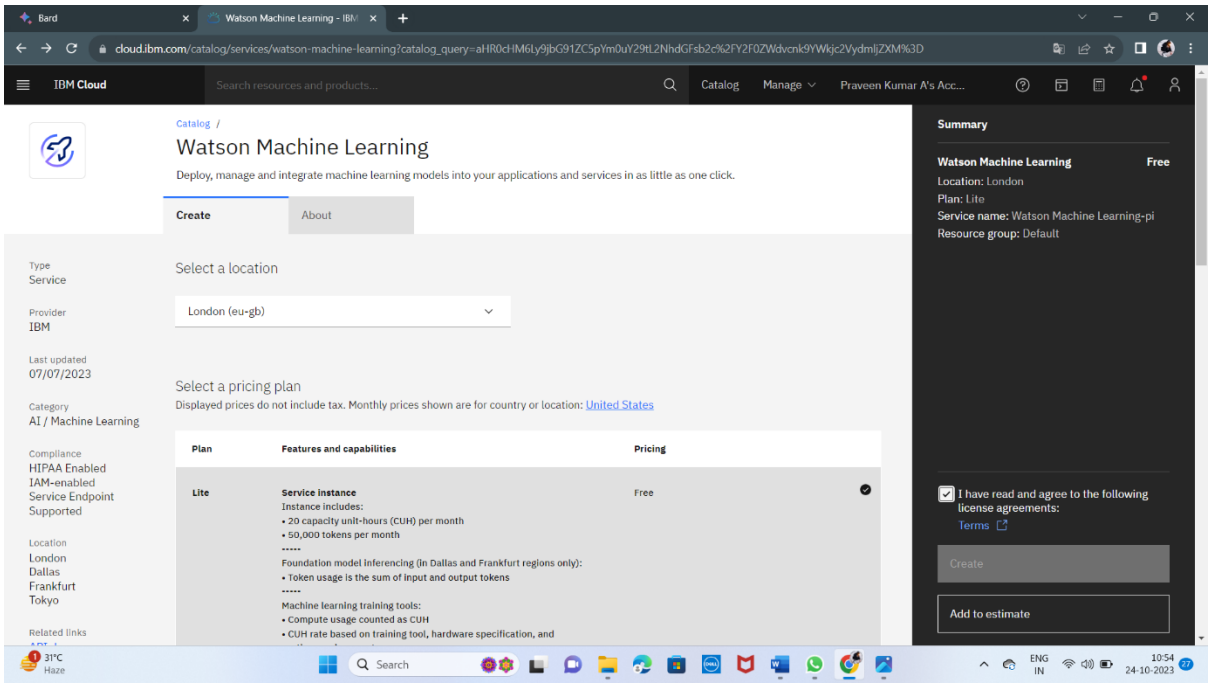
Dive into the world of big data analysis with IBM Cloud Databases. Uncover hidden insights from vast datasets, from climate trends to social patterns. Visualize your findings and derive valuable business intelligence. Embark on data-driven adventures, exploring the endless possibilities of big data!

**To continue building big data analysis solution using IBM Cloud Databases, we can apply more complex analysis techniques and visualize the results.**

# Machine Learning:

Machine learning algorithms can be used to learn from the data in your climate dataset and to make predictions. For example, you could use a machine learning algorithm to predict the average temperature in a given area in the future.

To use machine learning with IBM Cloud Databases, you can use **IBM Cloud Watson Machine Learning**. IBM Cloud Watson Machine Learning is a suite of machine learning services that can be used to build and deploy predictive models.



## Visualization:

Once you have performed your analysis, we can visualize the results to communicate our findings to others. We can use a variety of tools and software applications to create visualizations. For example, you could use the Python library Matplotlib to create graphs and charts.

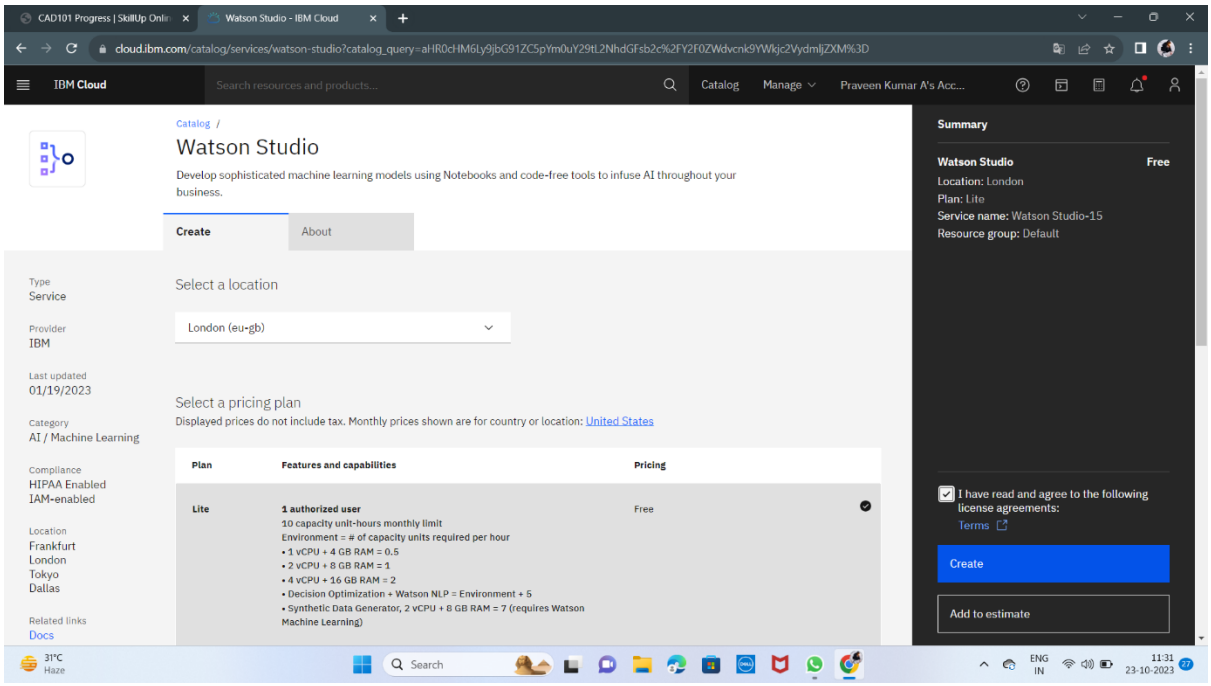
## IBM Watson Studio:

IBM Watson Studio is a unified platform for data scientists, business analysts, and developers to collaborate on data preparation, machine learning, and data visualization. **IBM Watson Studio** provides a variety of tools and services that can help you to perform advanced analysis and to visualize your results.

To create an IBM Watson Studio project, follow these steps:

- Go to the IBM Cloud console and sign in to your account.
- Click Catalog and then click Watson Studio.

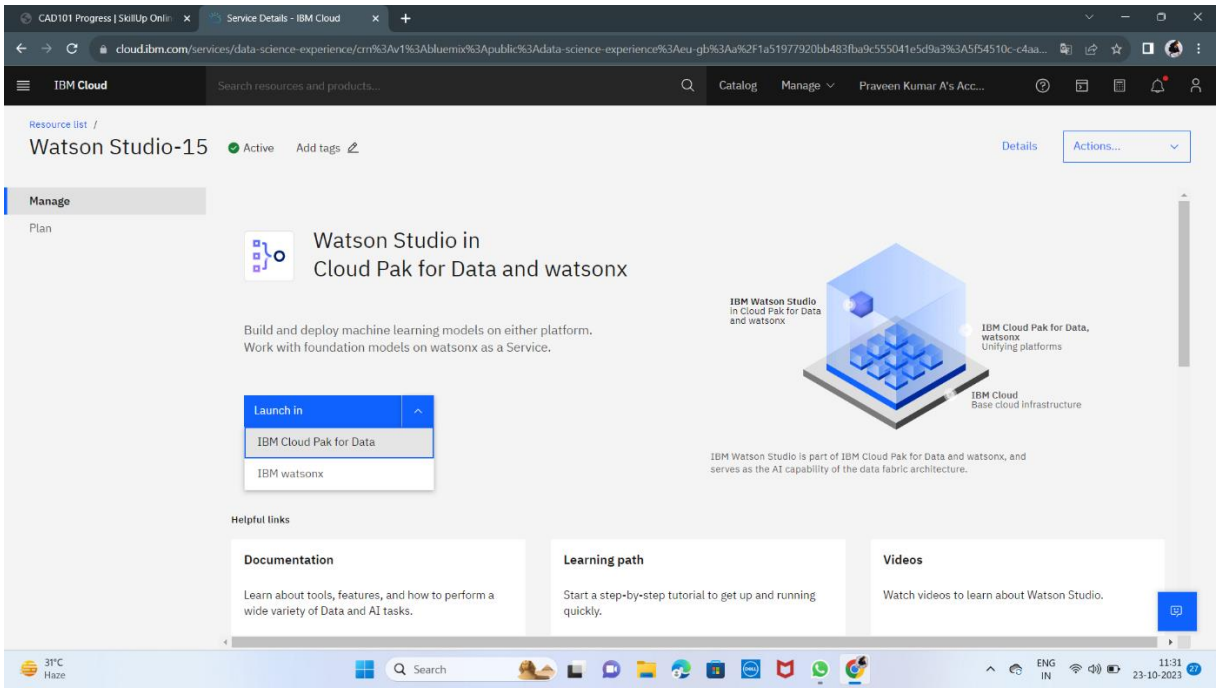
- Click Create project.
- Enter a name for your project and select a region.
- Click Create.




To add data to your project, you can:

- Upload files from your local computer.
- Connect to a database or cloud storage service.
- Use a Watson Studio service to generate data.

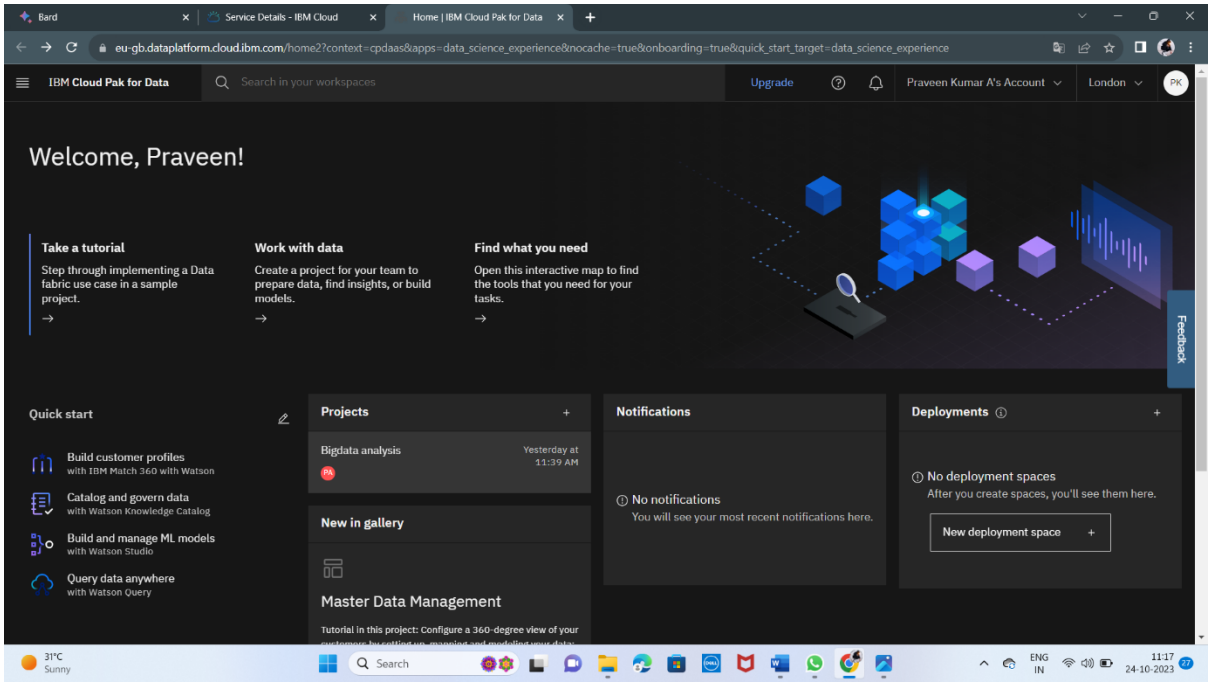
After Creating the IBM Watson Studio. Launch in the IBM Cloud Pak for data to visualize the results in the form of graphs and charts.




To launch the IBM Cloud Pak for Data to visualize the result in IBM Watson Studio, follow these steps:

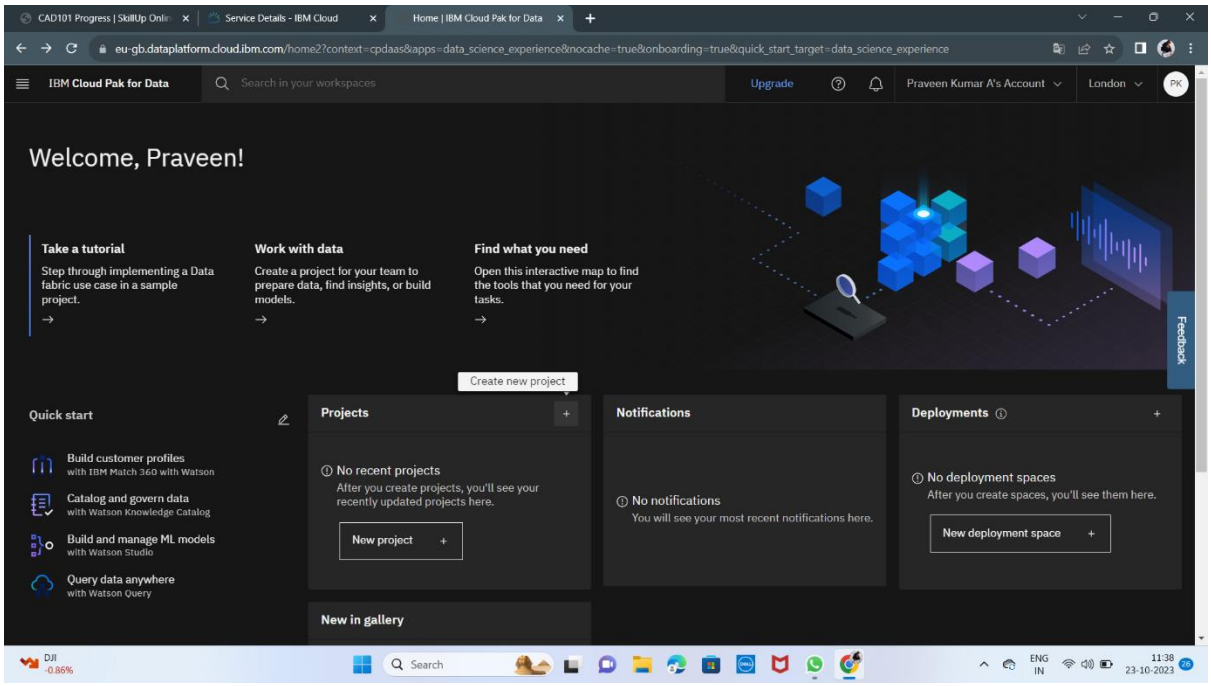
- Go to the Catalog tab in IBM Watson Studio.
- Click IBM Cloud Pak for Data.
- Click Launch.

A new window will open with the IBM Cloud Pak for Data user interface.

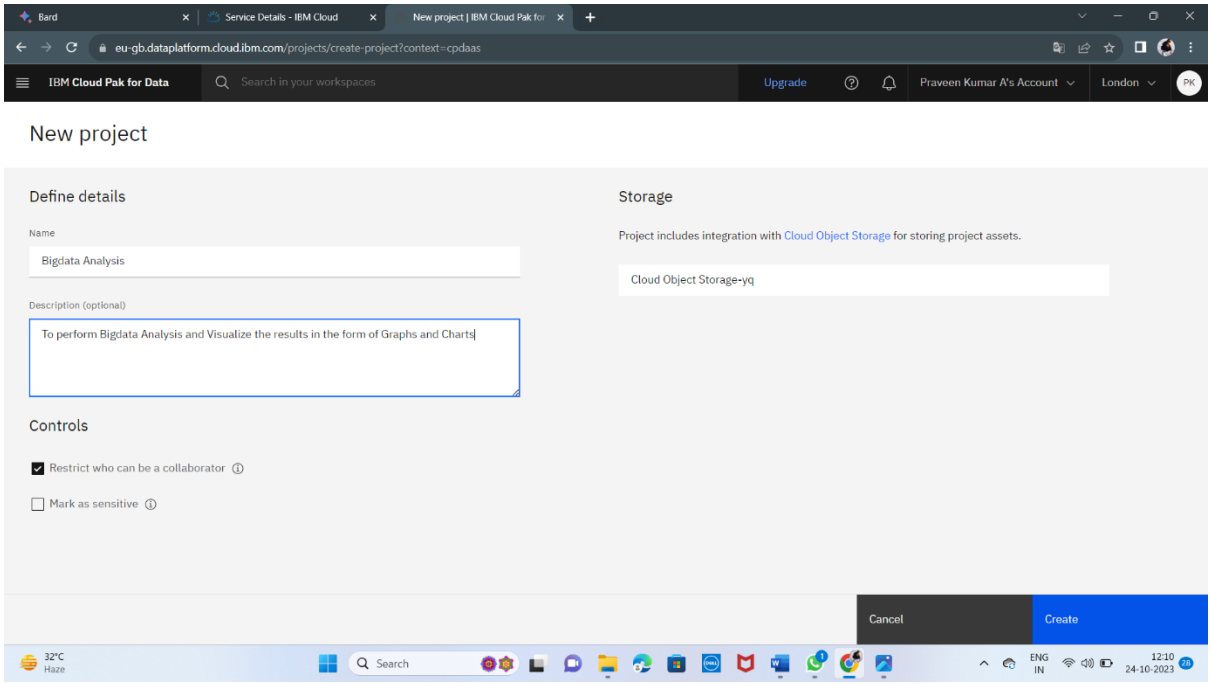



To create a new project in IBM Cloud Pak for Data:

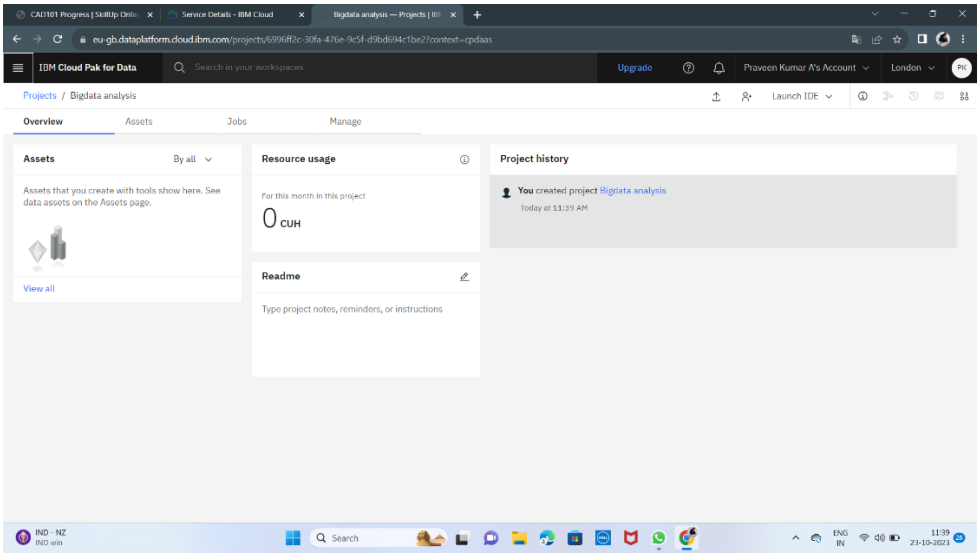
- Open the IBM Cloud Pak for Data console.
- Click Projects in the left navigation pane.
- Click Create project.



- In the Create project dialog, enter a name and description for your project.
- Select a project type. You can choose from a variety of project types.
- Click Create.

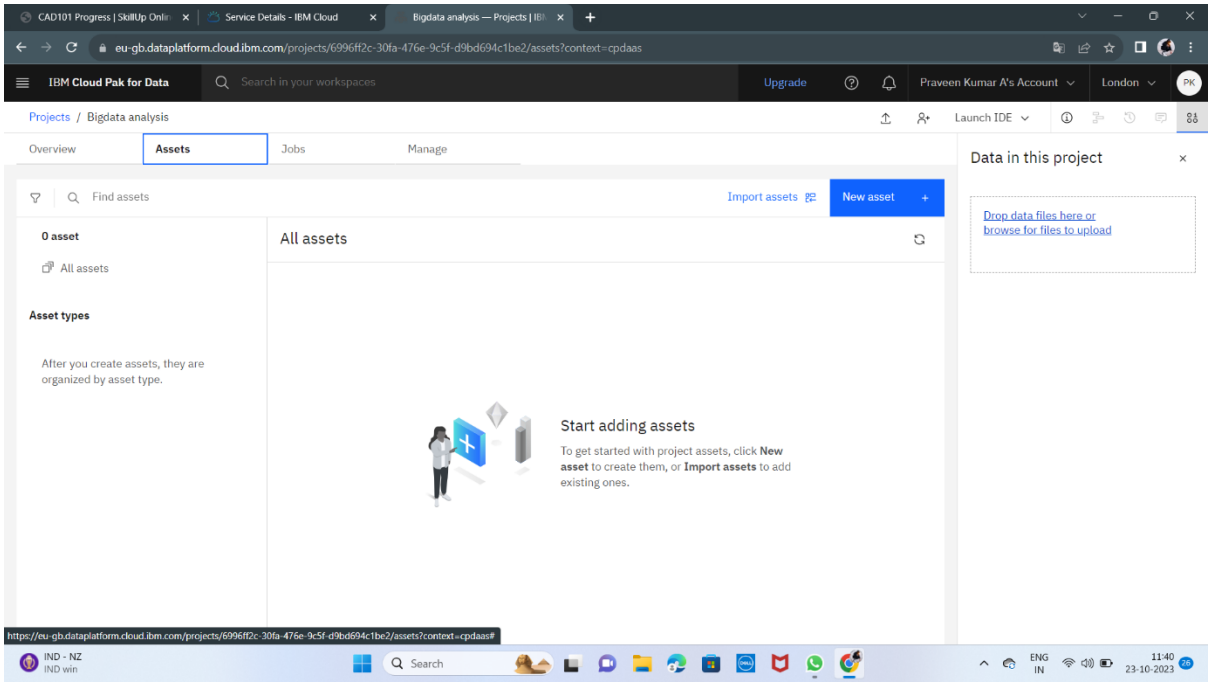



Once the Project is created it will be redirected to the project dashboard.

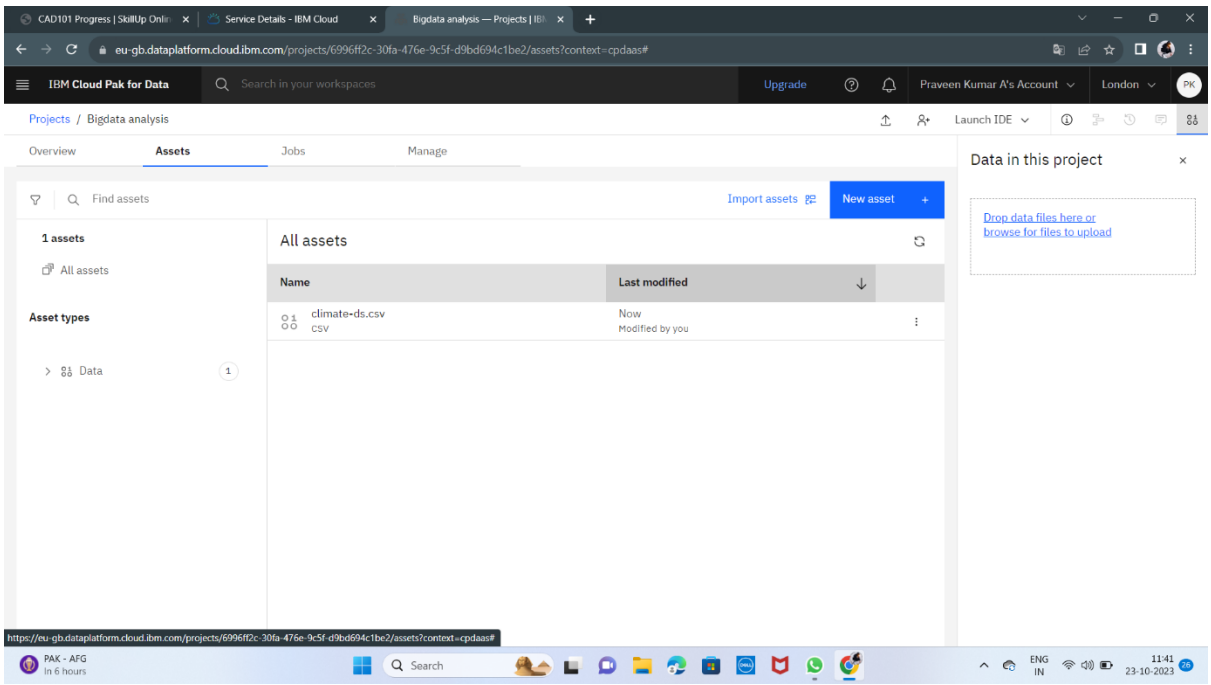


To add data to your project, you can:

- Upload files from your local computer.

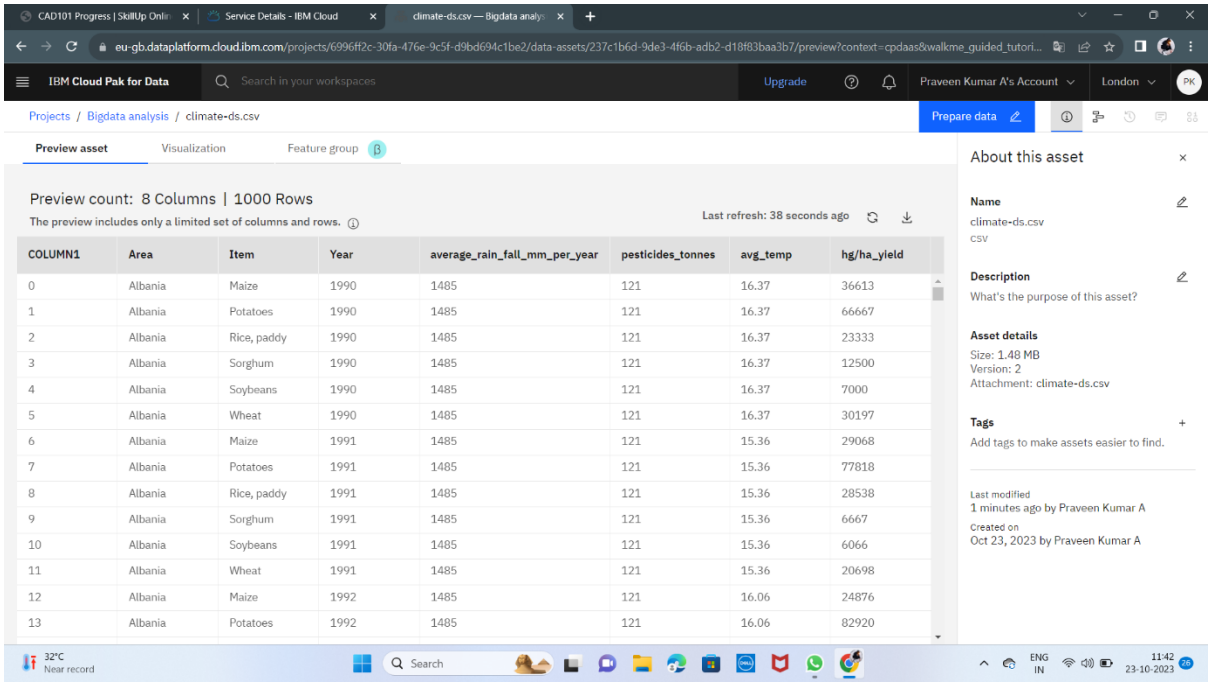



- Connect to a database or cloud storage service.
- Use a Watson Studio service to generate data.



Once you have added data to your project, you can start building and deploying machine learning models, exploring and visualizing data.

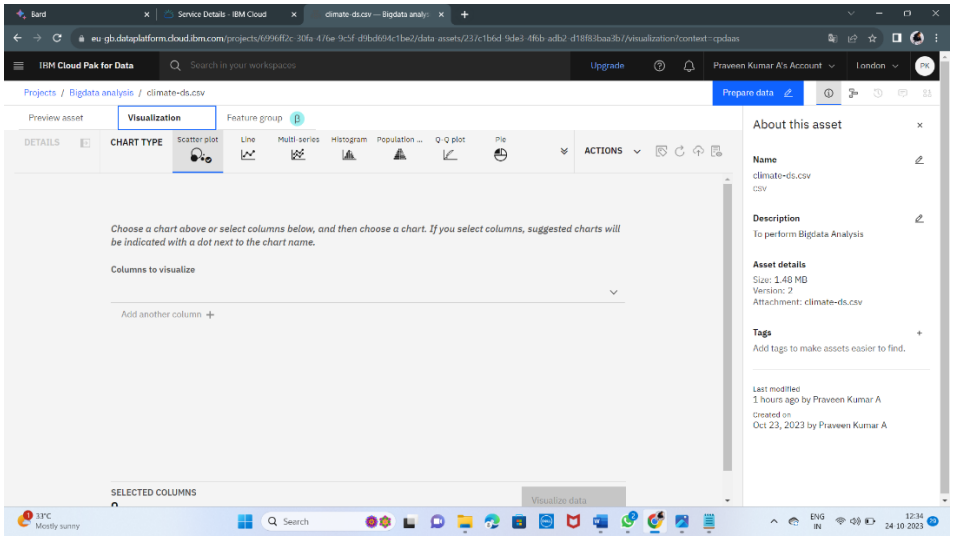
## The preview asset of our Climate Dataset:



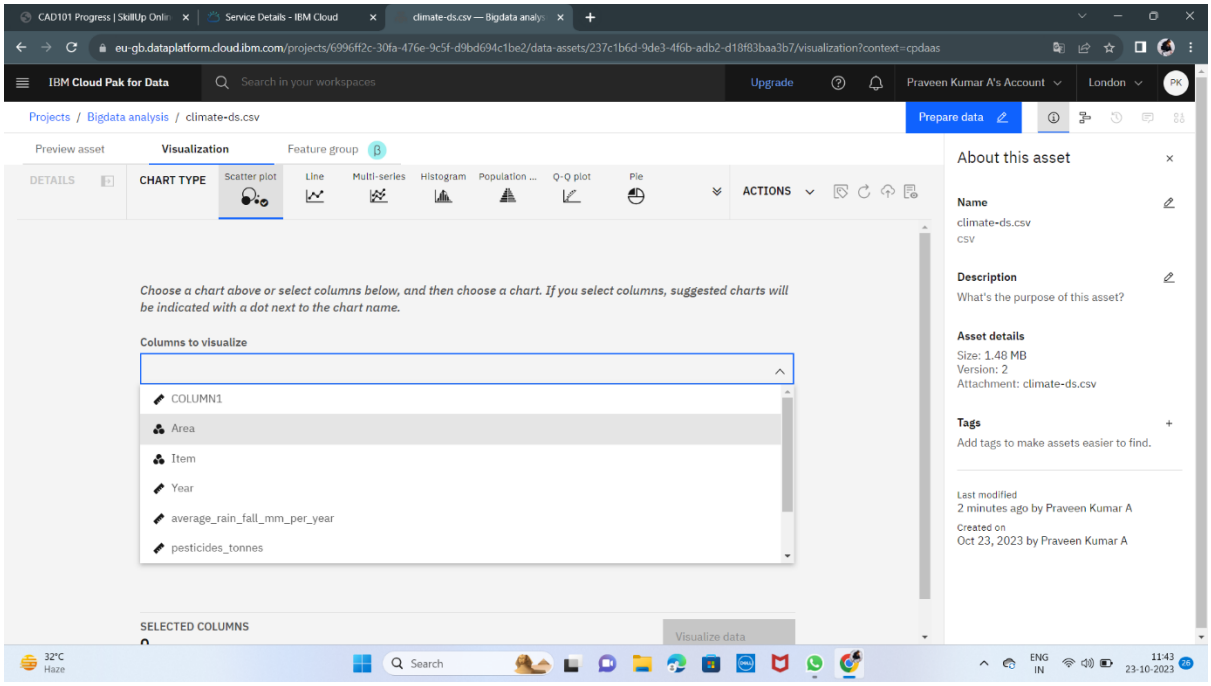

| COLUMN1 | Area | Item | Year | average_rain_fall_mm_per_year | pesticides_tonnes | avg_temp | hg/ha_yield |
|---|---|---|---|---|---|---|---|
| 0 | Albania | Maize | 1990 | 1485 | 121 | 16.37 | 36613 |
| 1 | Albania | Potatoes | 1990 | 1485 | 121 | 16.37 | 66667 |
| 2 | Albania | Rice, paddy | 1990 | 1485 | 121 | 16.37 | 23333 |
| 3 | Albania | Sorghum | 1990 | 1485 | 121 | 16.37 | 12500 |
| 4 | Albania | Soybeans | 1990 | 1485 | 121 | 16.37 | 7000 |
| 5 | Albania | Wheat | 1990 | 1485 | 121 | 16.37 | 30197 |
| 6 | Albania | Maize | 1991 | 1485 | 121 | 15.36 | 29068 |
| 7 | Albania | Potatoes | 1991 | 1485 | 121 | 15.36 | 77818 |
| 8 | Albania | Rice, paddy | 1991 | 1485 | 121 | 15.36 | 28538 |
| 9 | Albania | Sorghum | 1991 | 1485 | 121 | 15.36 | 6667 |
| 10 | Albania | Soybeans | 1991 | 1485 | 121 | 15.36 | 6066 |
| 11 | Albania | Wheat | 1991 | 1485 | 121 | 15.36 | 20698 |
| 12 | Albania | Maize | 1992 | 1485 | 121 | 16.06 | 24876 |
| 13 | Albania | Potatoes | 1992 | 1485 | 121 | 16.06 | 82920 |

To visualize,

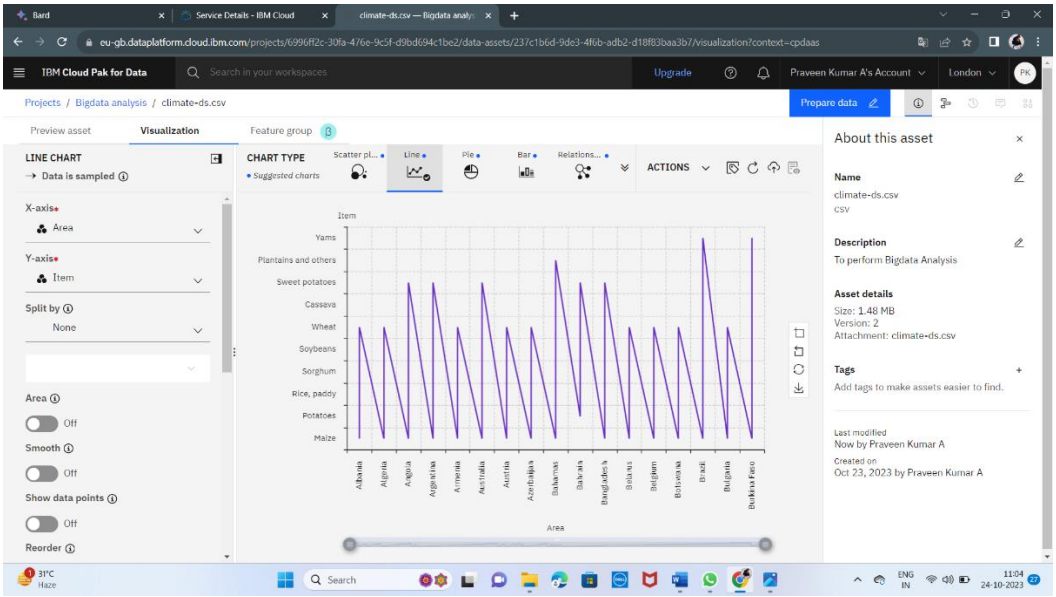
- Click Visualization tab

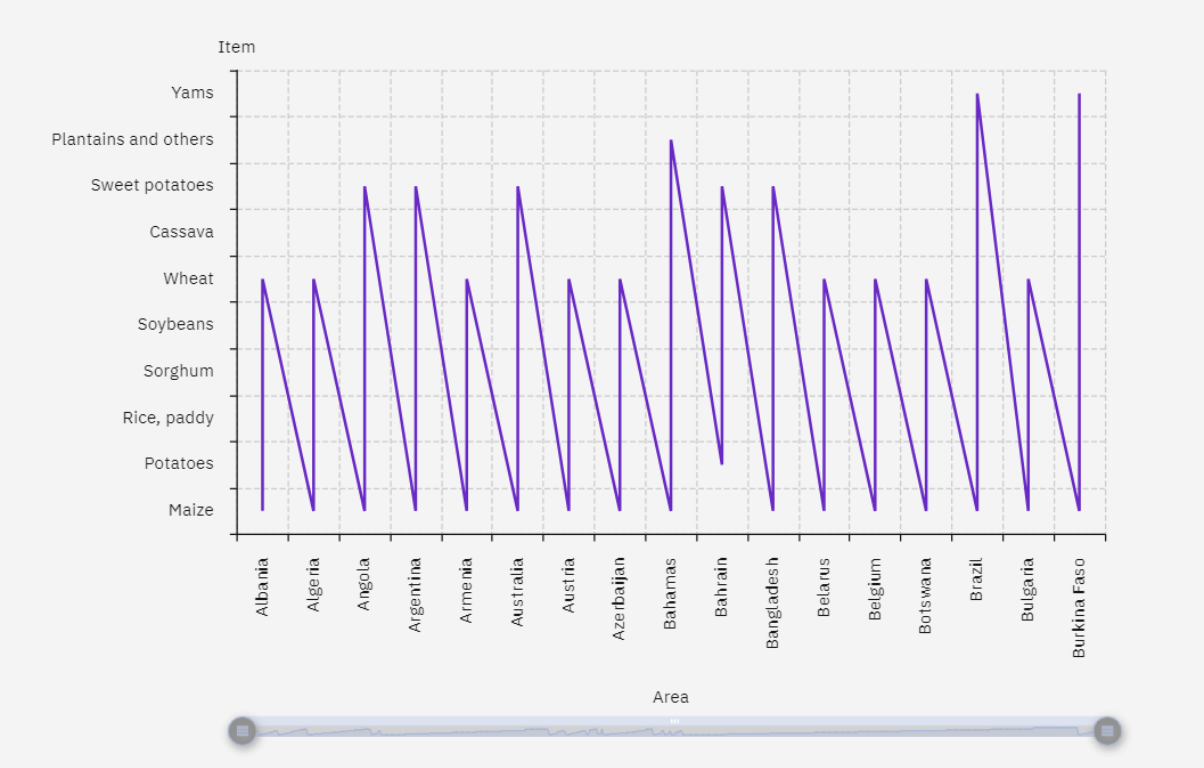


- Choose a chart above or select columns below, and then choose a chart. If you select columns, suggested charts will be indicated with a dot next to the chart name.

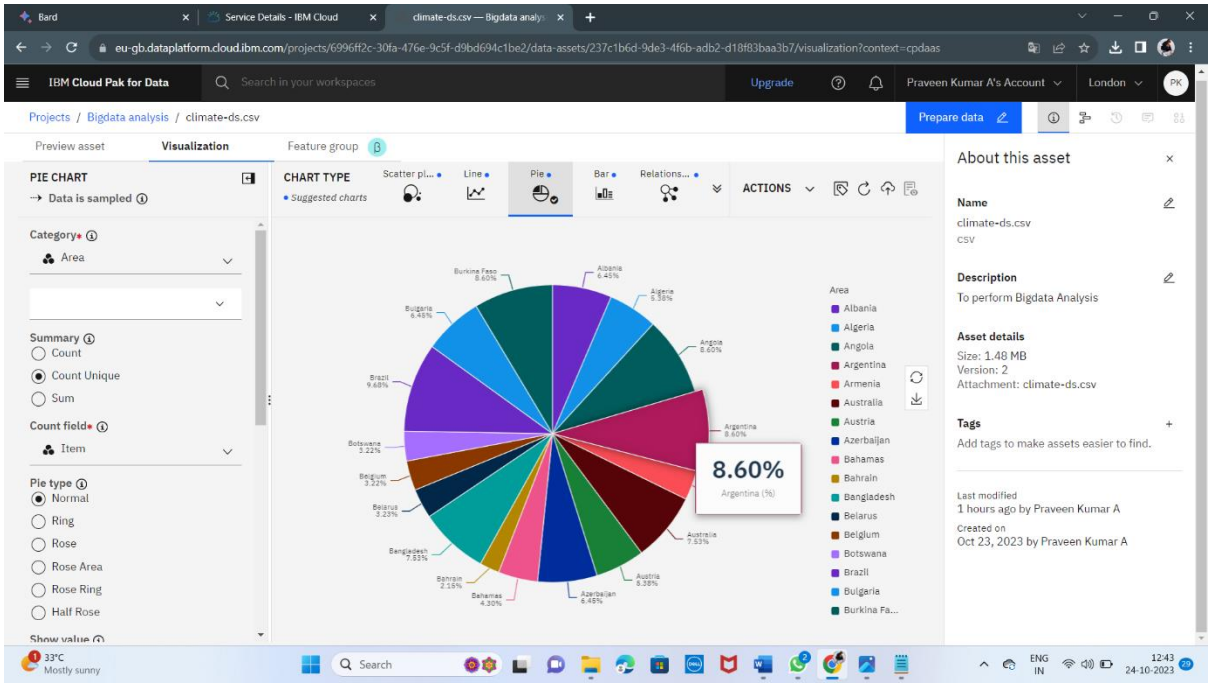


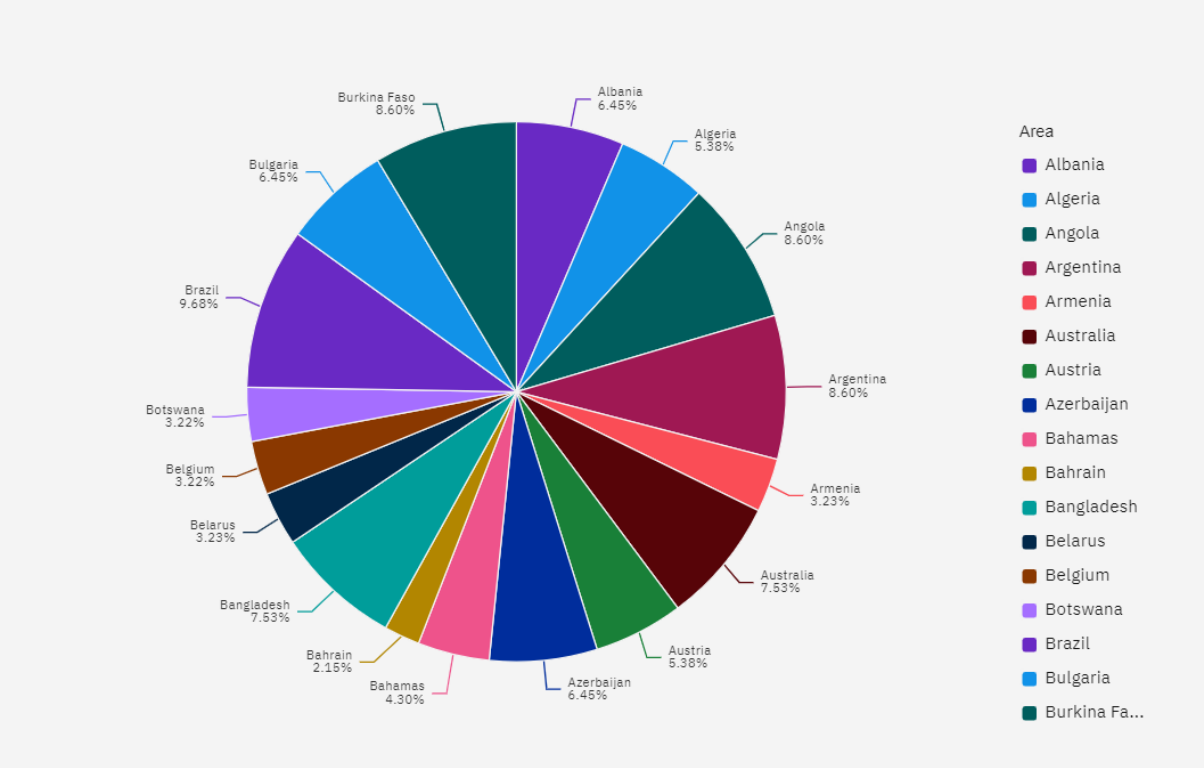

Data Visualization of columns **AREA** and **ITEM** in Line chart.



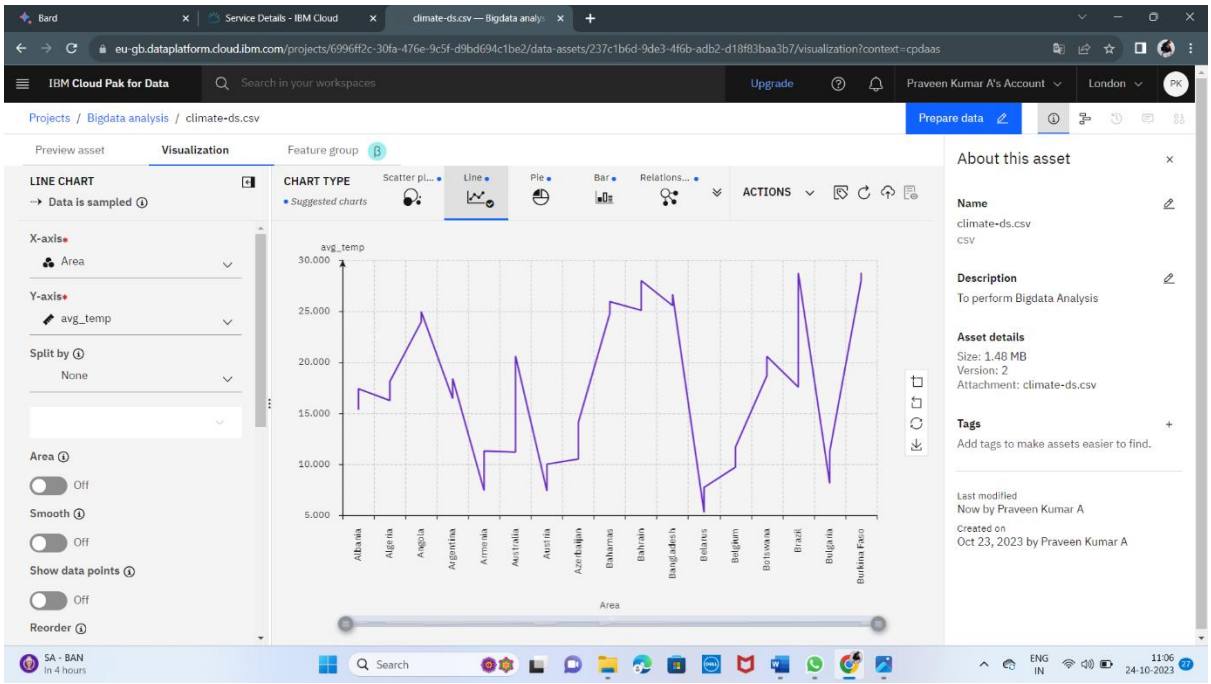



Data Visualization of columns **AREA** in PIE chart.

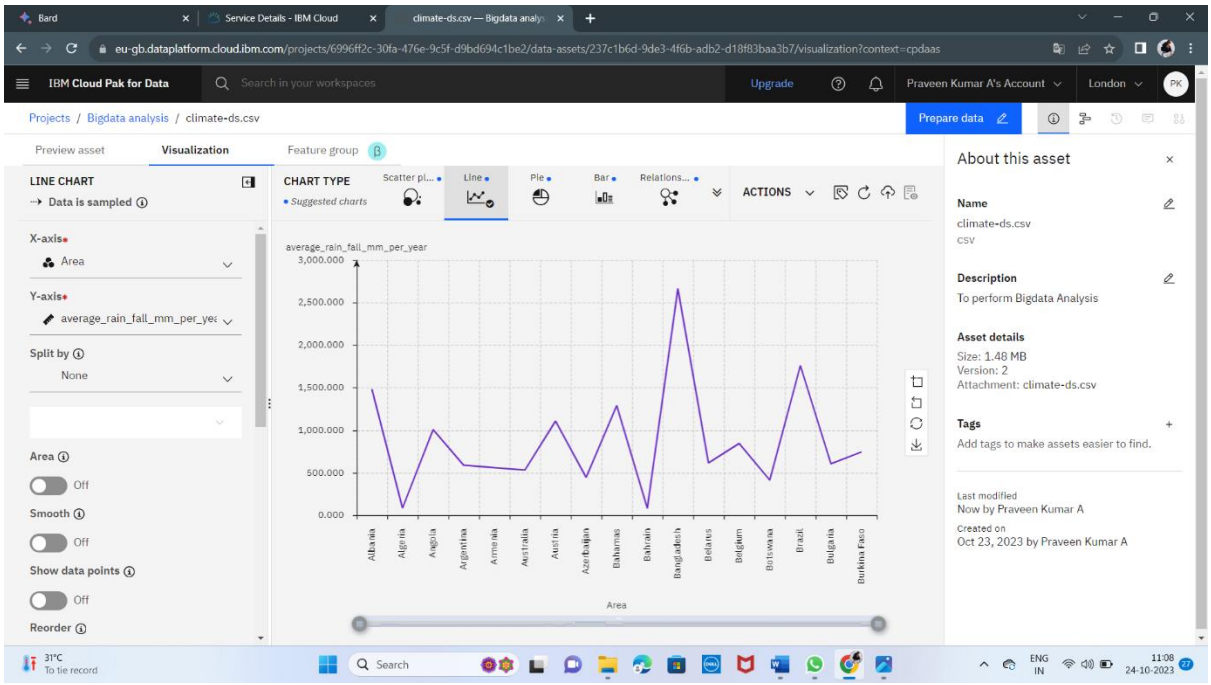


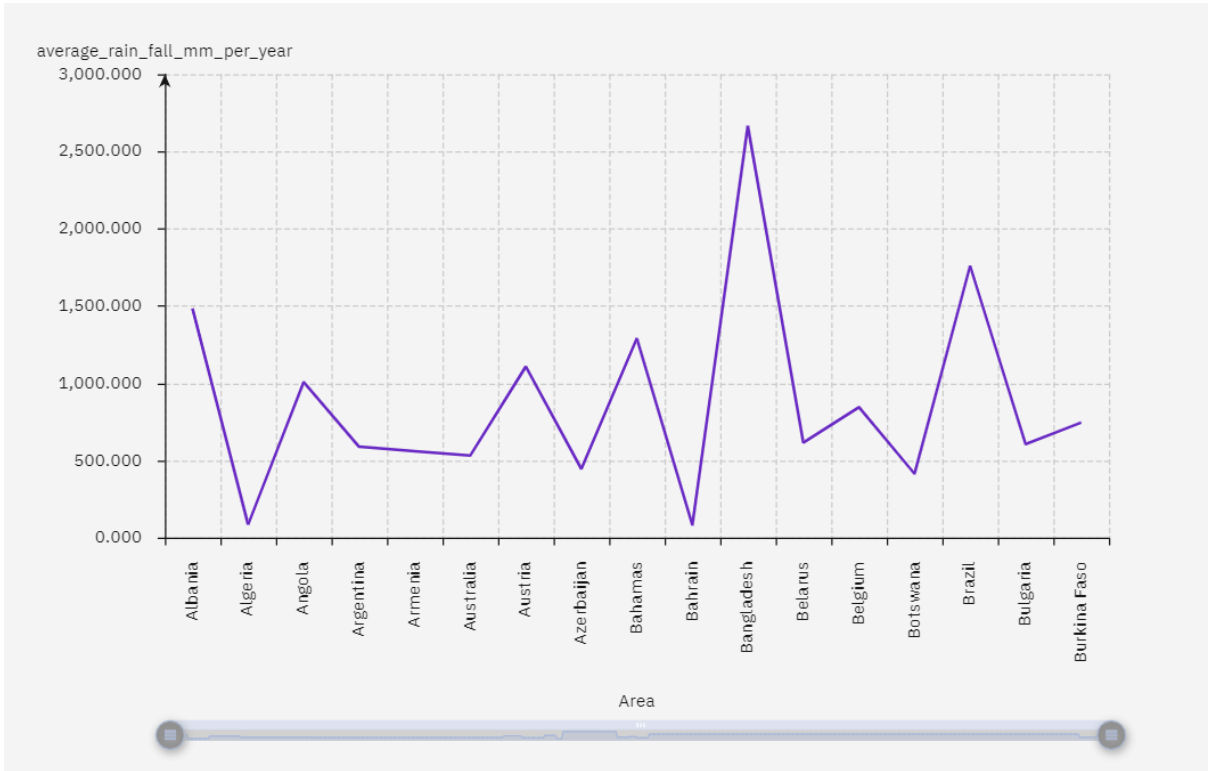



Data Visualization of columns **AREA** and **AVG_TEMP** in Line chart.

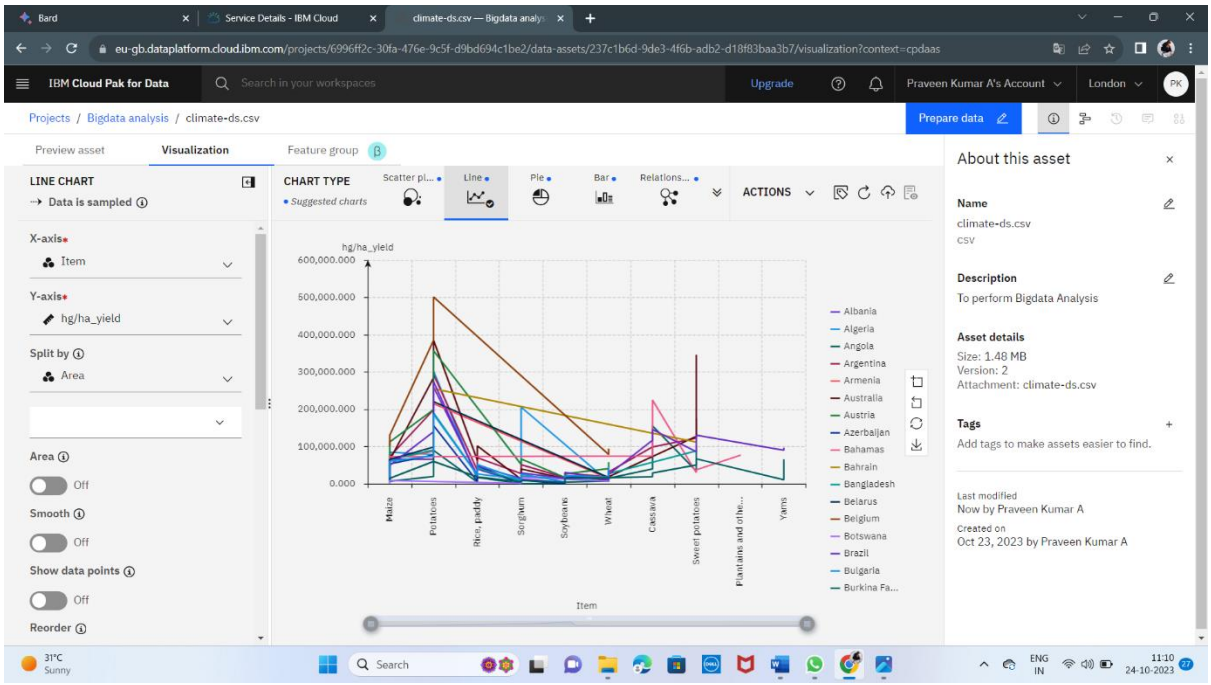


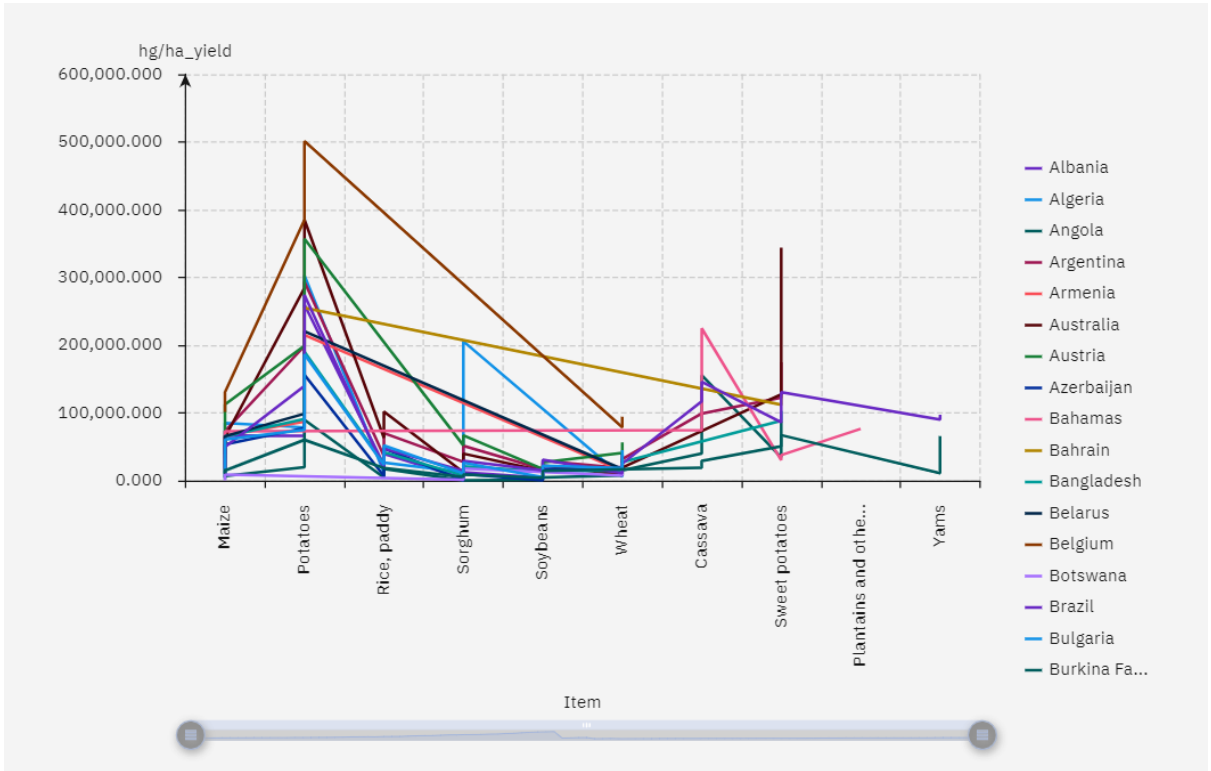
Data Visualization of columns **AREA** and **AVERAGE_RAIN_FALL_MM_PER_YEAR** in Line chart.



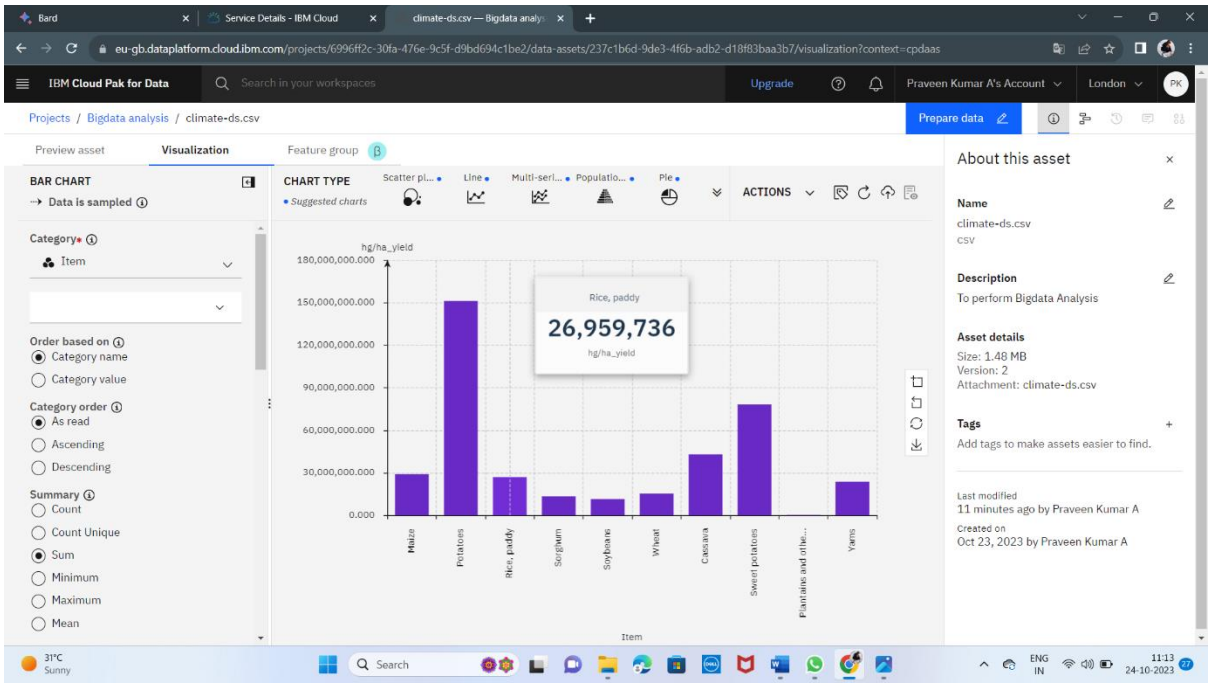



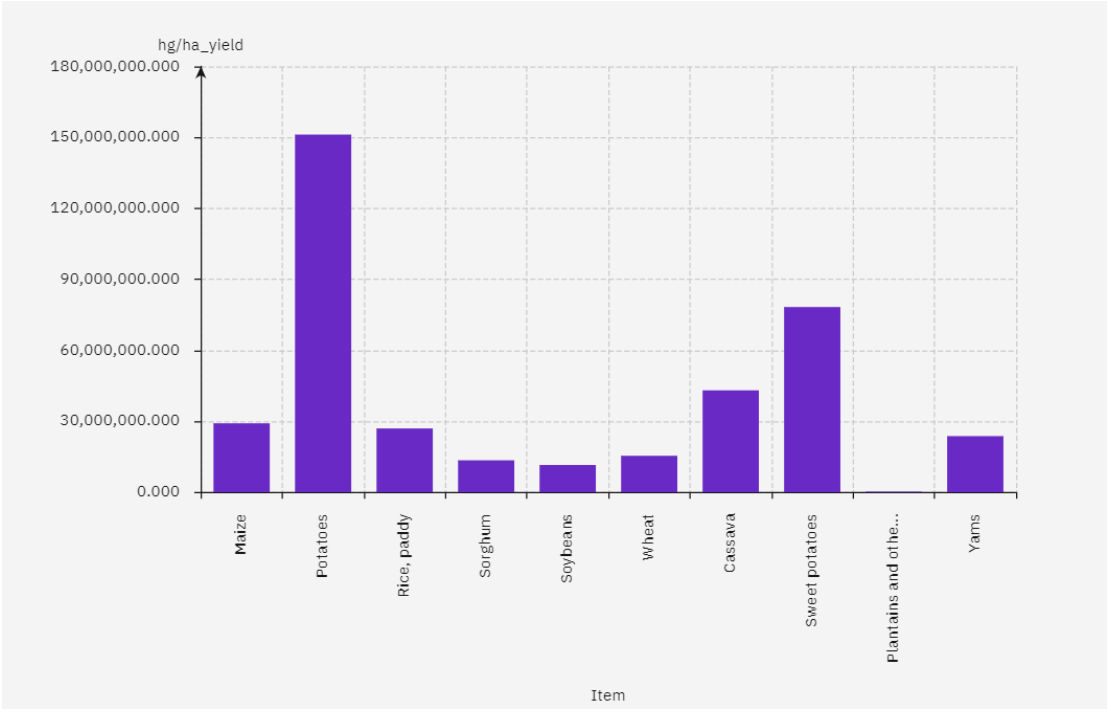
Data Visualization of columns **ITEM** and **HG_HA_YIELD** Split by **AREA** in Line chart.



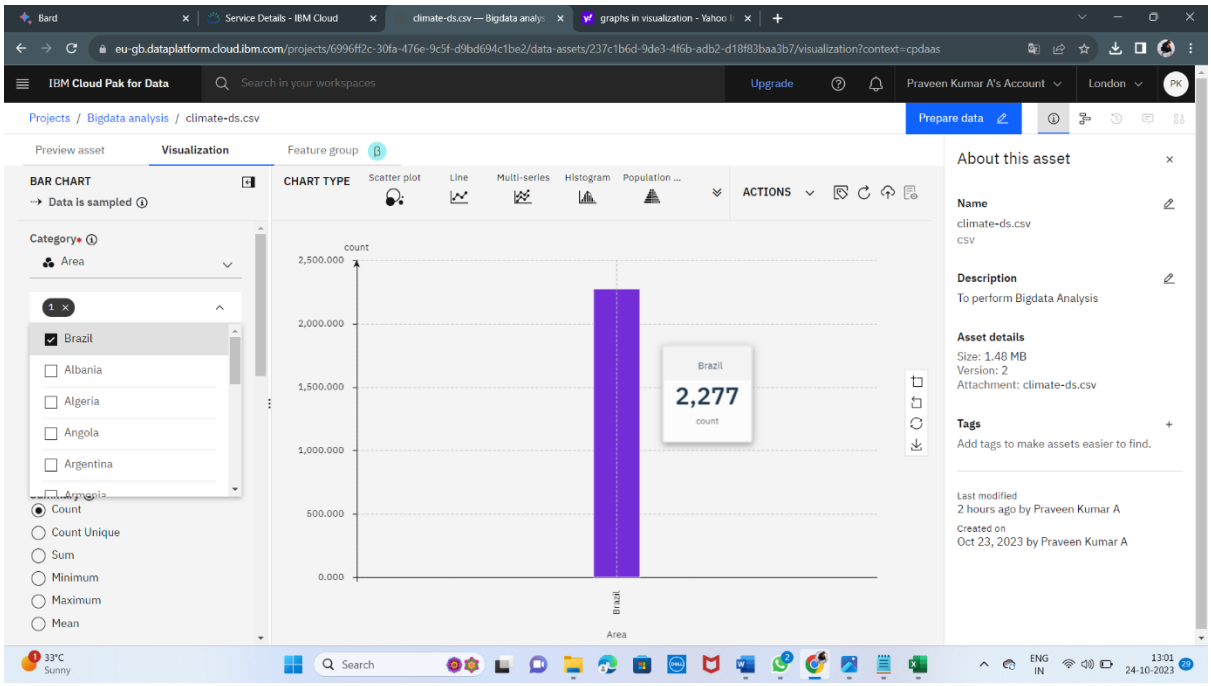



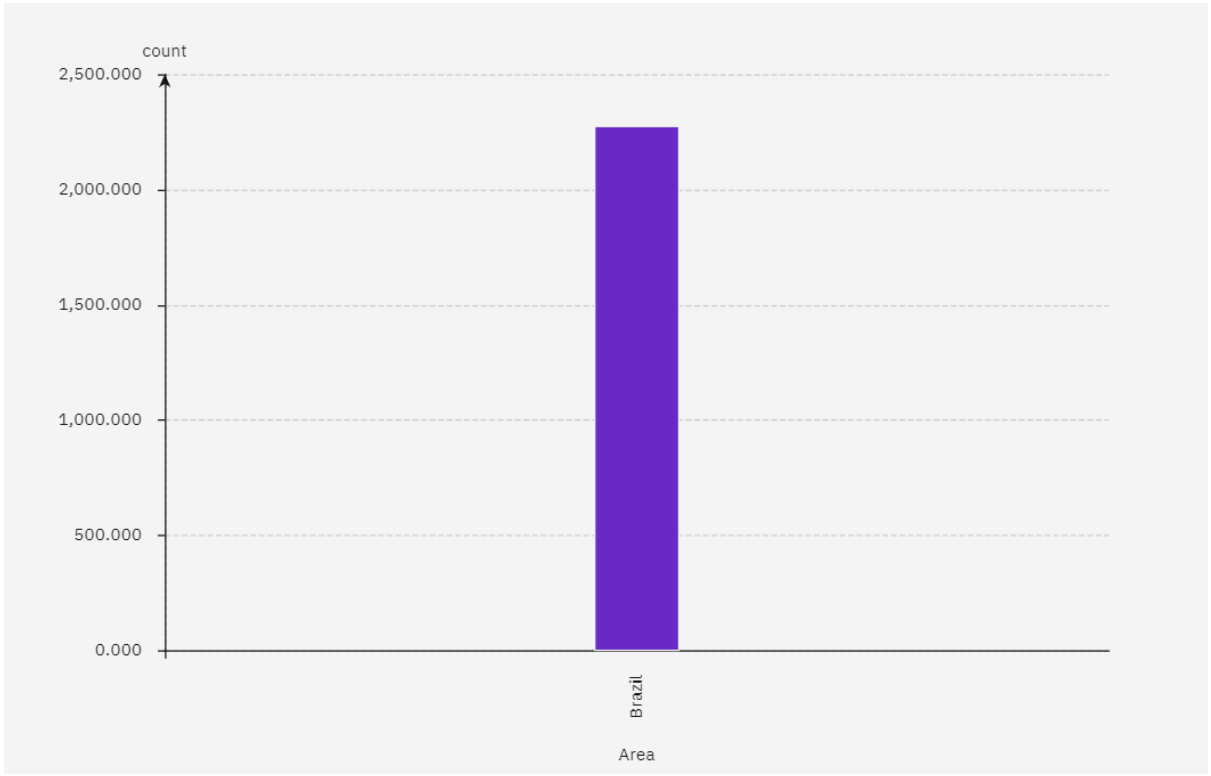
Data Visualization of columns **ITEM** and **HG_HA_YIELD** in BAR chart.



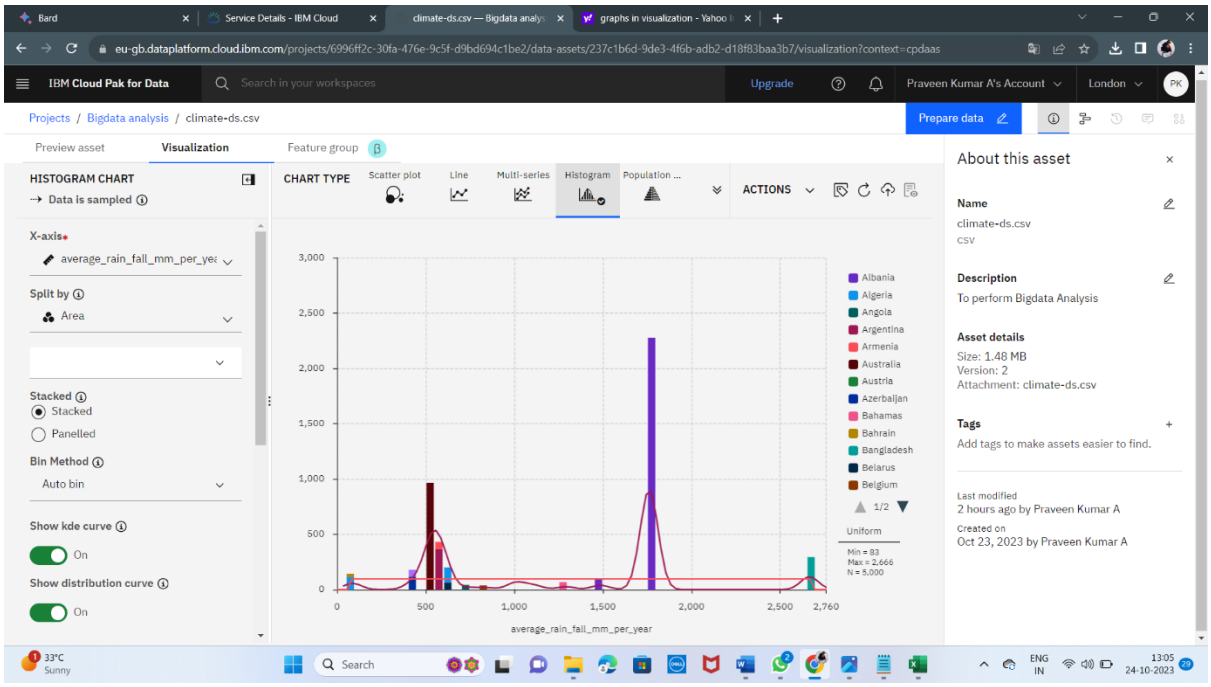
# Data Visualization of Specific **AREA** in BAR chart.

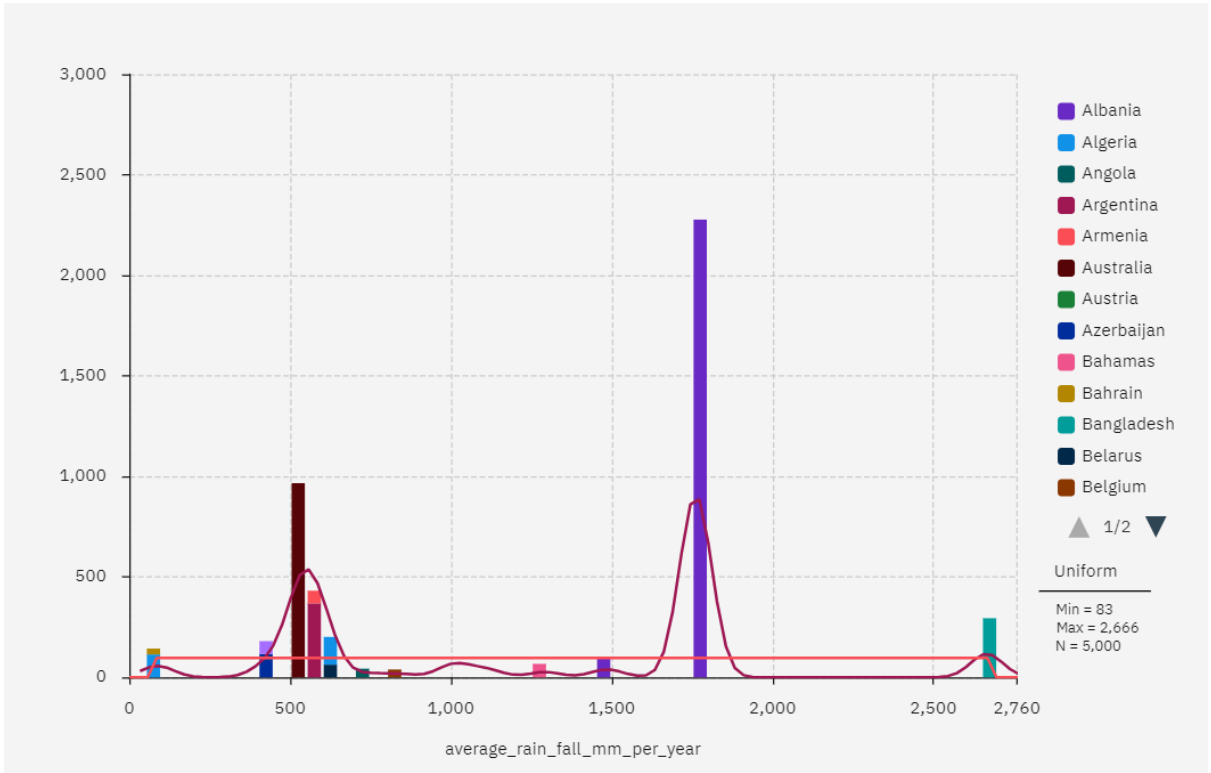


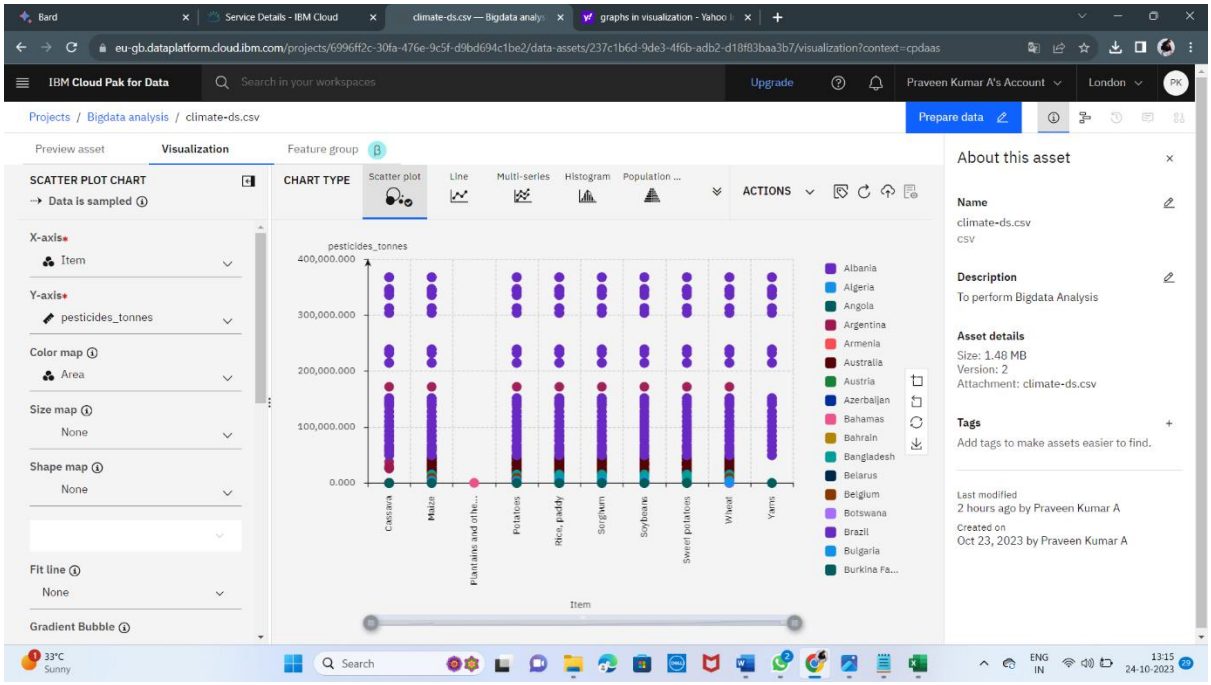
Data Visualization of column

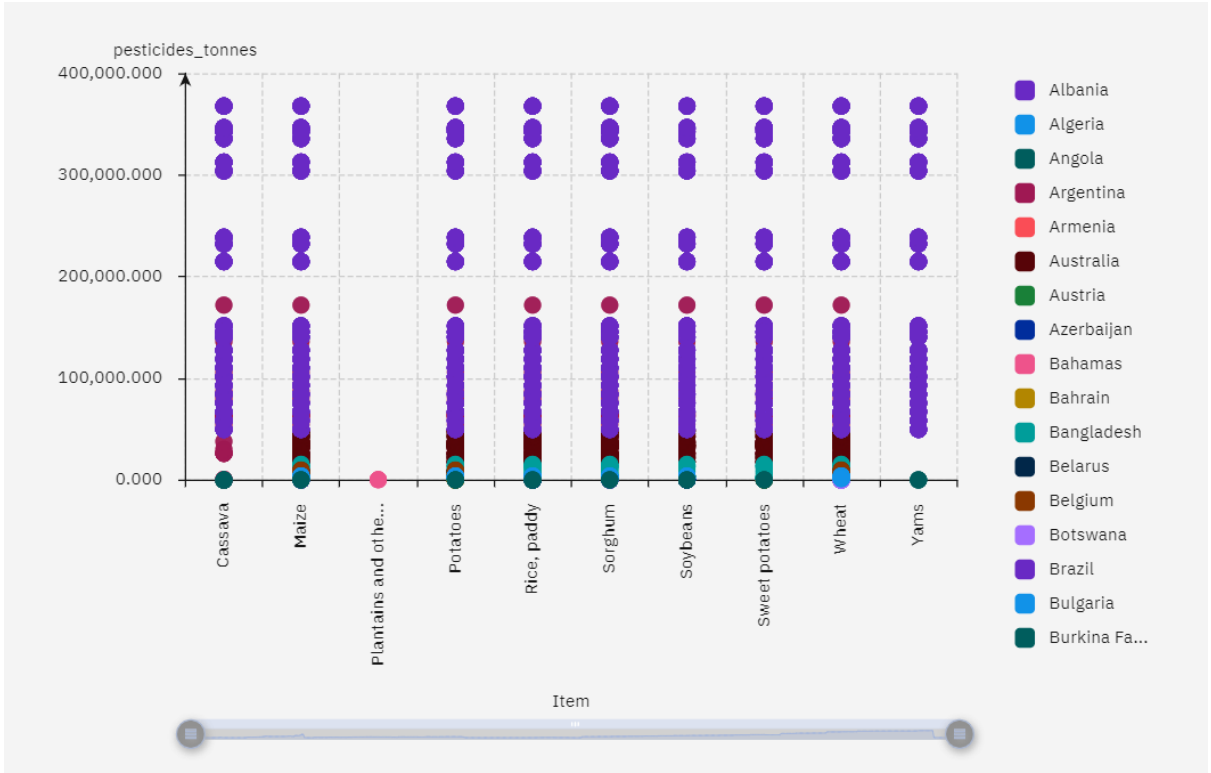
**AVERAGE_RAIN_FALL_MM_PER_YEAR** Split by **AREA** in Histogram chart.



Data Visualization of columns **ITEM** and **PESTICIDES_TONNES** Colour map by AREA in Scatter Plot chart.



**Team Leader name :**

Praveen Kumar A (Reg no :111421104091)

**Team Members :**

Vijayakumar C (Reg no :111421104118)

Praveen joel (Reg no :111421104090)

Yashwanth Kumar S (Reg no :111421104123)

Nighil Ananth V (Reg no :111421104072)