**BIG DATA ANALYSIS USING IBM CLOUD**

**Project’s Objective:**

The objective of this project is to clearly understand the data analytics concept and how to use the IBM cloud for find the hidden trends and patterns inside the dataset and IBM Watson Studio for visualization like creating graphs and charts. This project also includes how to use the python libraries like numpy (numerical python), pandas (framing the data), matplotlib (plotting the graph) for analyze the data.

**Design Thinking:**

Design thinking is the iterative process of how to perform the project.

**Data Selection:** Identify the datasets to be analyzed, such as climate data or social media trends data but our team choose the climate data of analyze.

**Database Setup:** Using the feature code in the skillup course like Introduction to cloud to freely open our IBM cloud and use the IBM db2 and IBM Watson Studio for analyze, store the dataset and visualization the dataset respectively.

**Data Exploration:** In IBM db2 develop queries and scripts to explore the datasets,extract relevant information and identify patterns.

**Analysis Techniques:** Applying the appropriate machine learning techniques to find the hidden patterns in the dataset. Our team use the K-Means Clustering algorithm to find the hidden pattern in the dataset.

**Visualization:** Using the IBM Watson Studio to visualize the dataset like graphs and charts to understand about the hidden trends and patterns in the dataset.

**How our team performs in the Project.**

**Step 1-** At first, clearly understand the concepts in each phase.

**Step 2-** After understanding, discuss about the concepts with our team members to perform it.

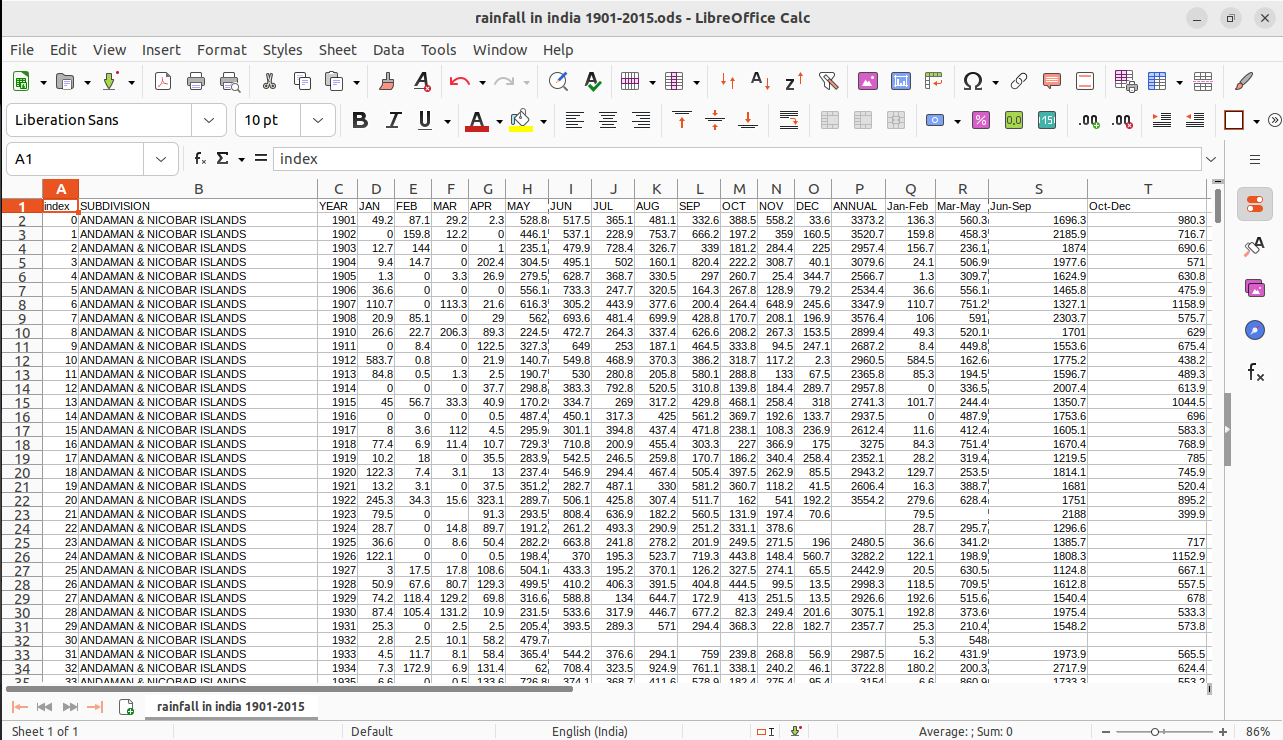
**Step 3-** Separate our works to each other and perform their prepective tasks very well.

**Step 4-** Finally review our each works and collaborate our works to make it as a single project.

**Step 5-** Again, review the whole project and ask my mentor to review about our project and corrects mistake if any occurs and then our team submit it.

**Dataset Analysis:**

We choose the climate data for analyze purpose. We collect the dataset from the kaggle.com to download the dataset. The dataset contains various irrelvant data like noisy using the python libraries like pandas to remove the noisy data and also use the IBM db2 to remove the noisy data using the queries and scripts in SQL section.

****

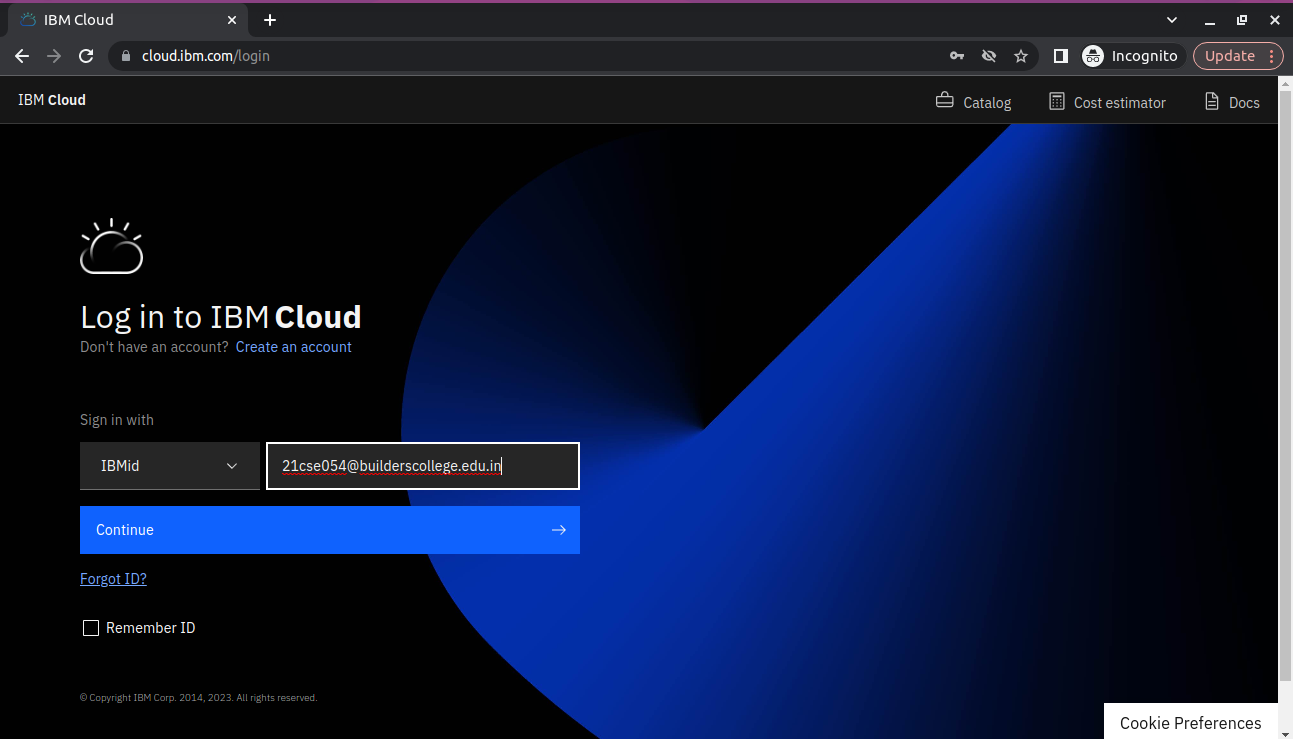
**Setup Database and upload the dataset into db2:**

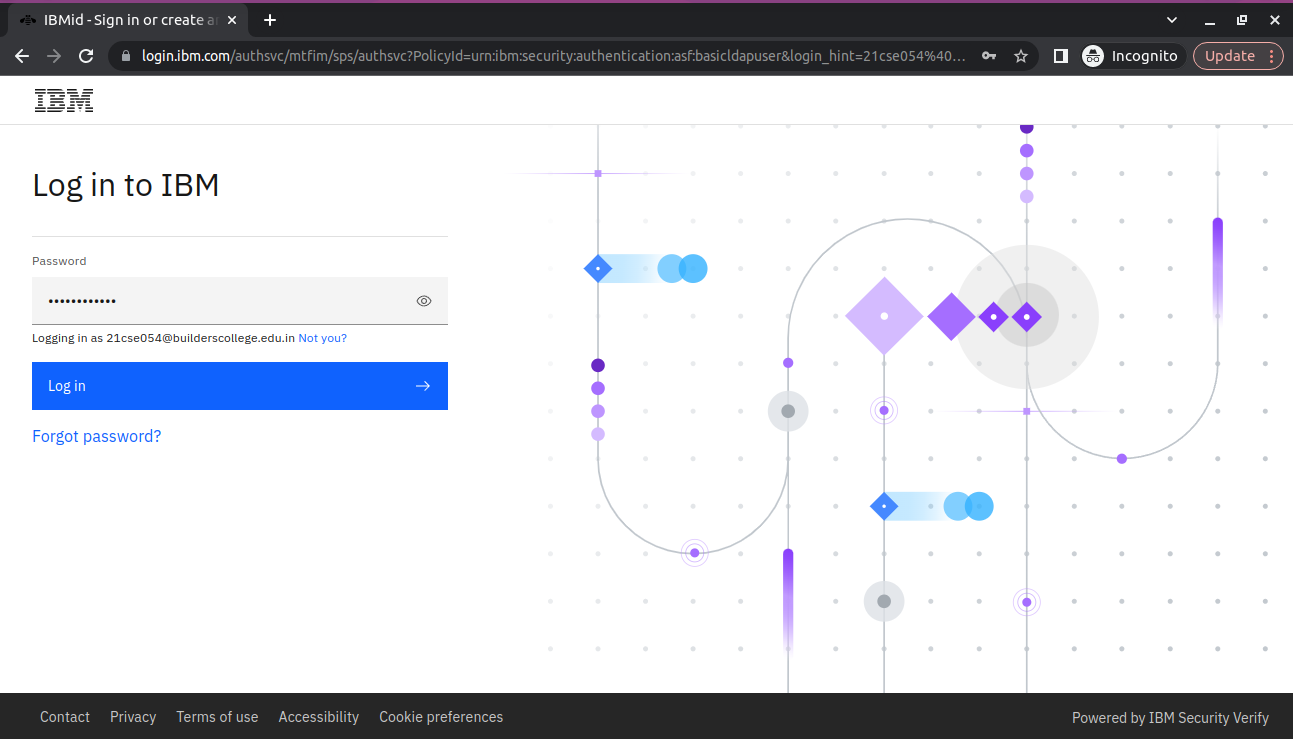
**Steps followed to create IBM Cloud Account.**

**Step 1 -** Go to [myclass.skillup.online](http://myclass.skillup.online/) platform and search Introduction to cloud.

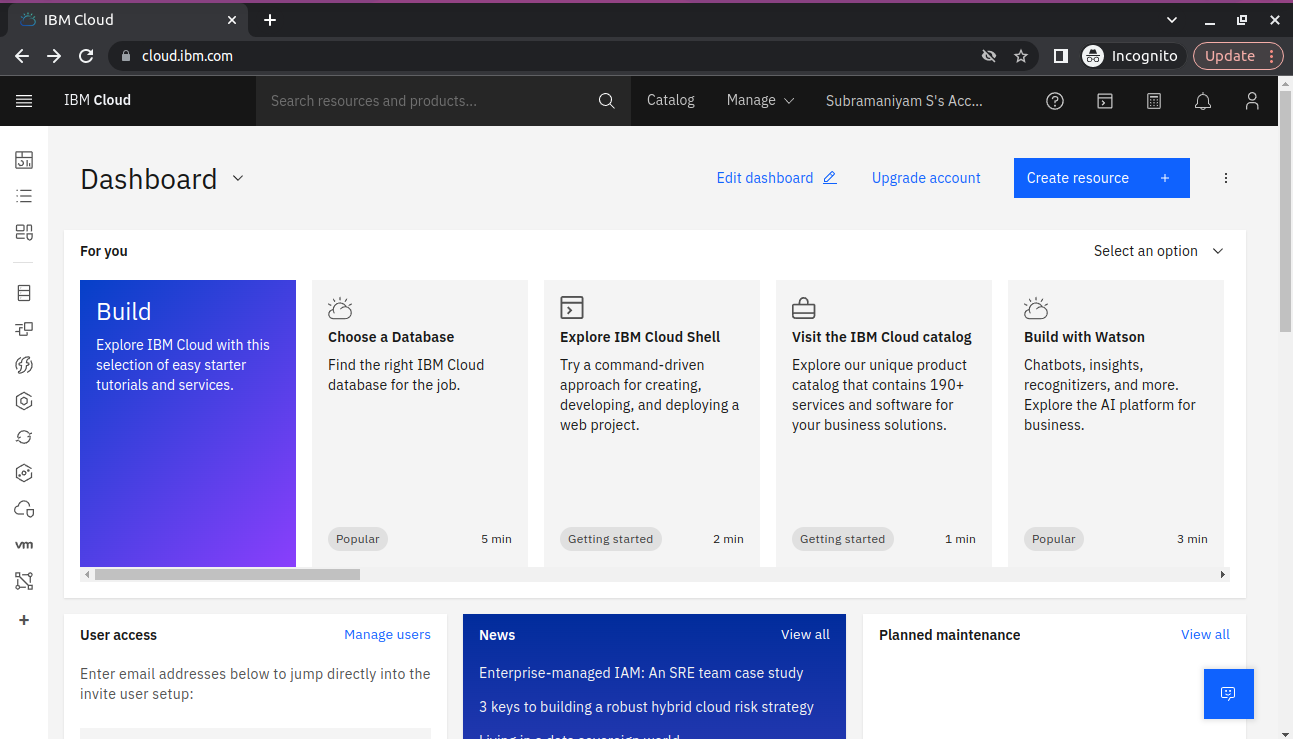
**Step 2 -** Then Go to Module 1 and click the obtain feature code to obtain the feature code. Then click to activate the account.

**Step 3 -** Then you will be redirected to the [cloud.ibm.com](http://cloud.ibm.com/) and then type our registered email id and password.

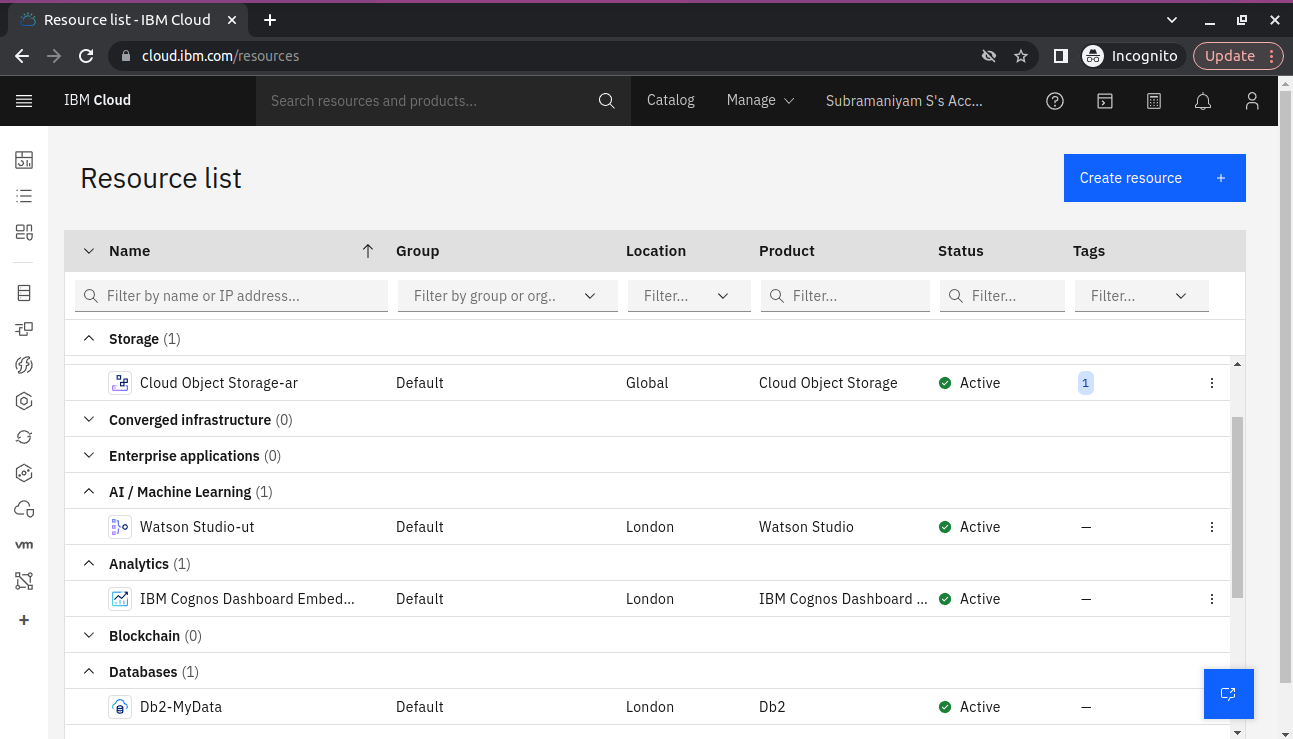




**Step 4 -** Then you will be redirected to the dashboard.

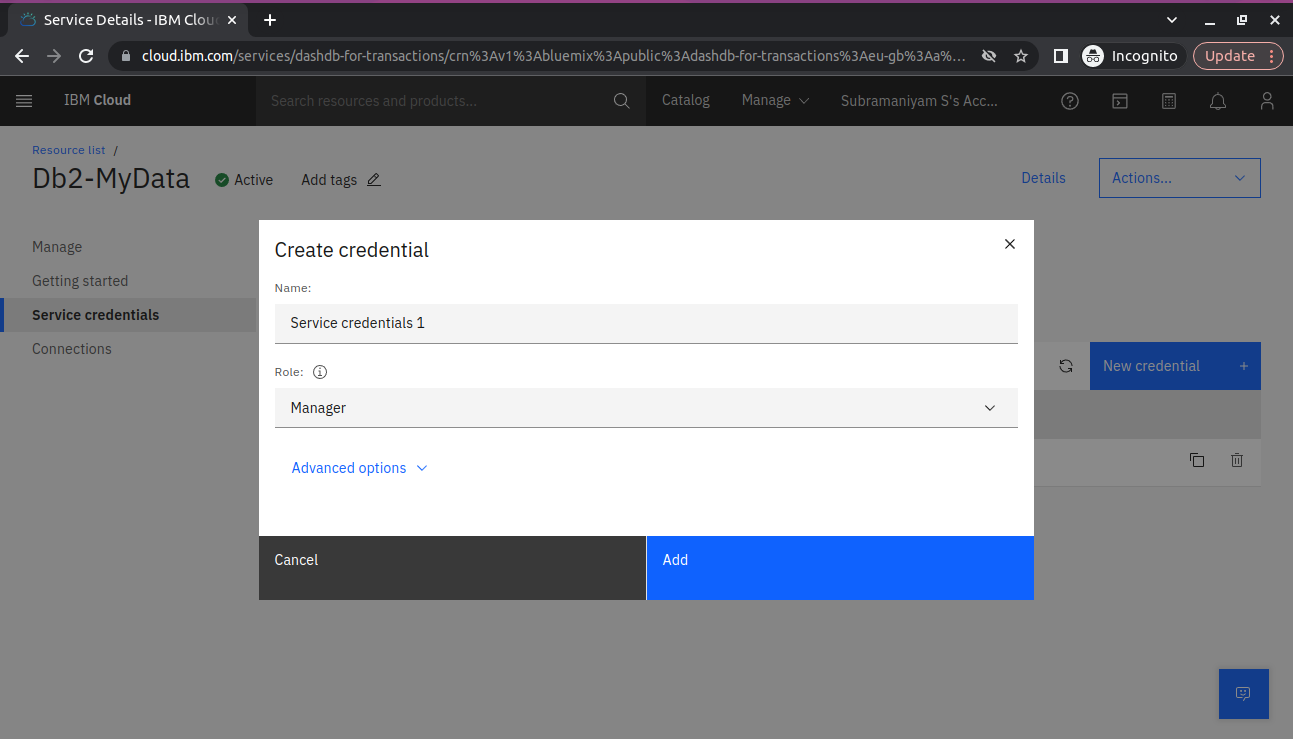


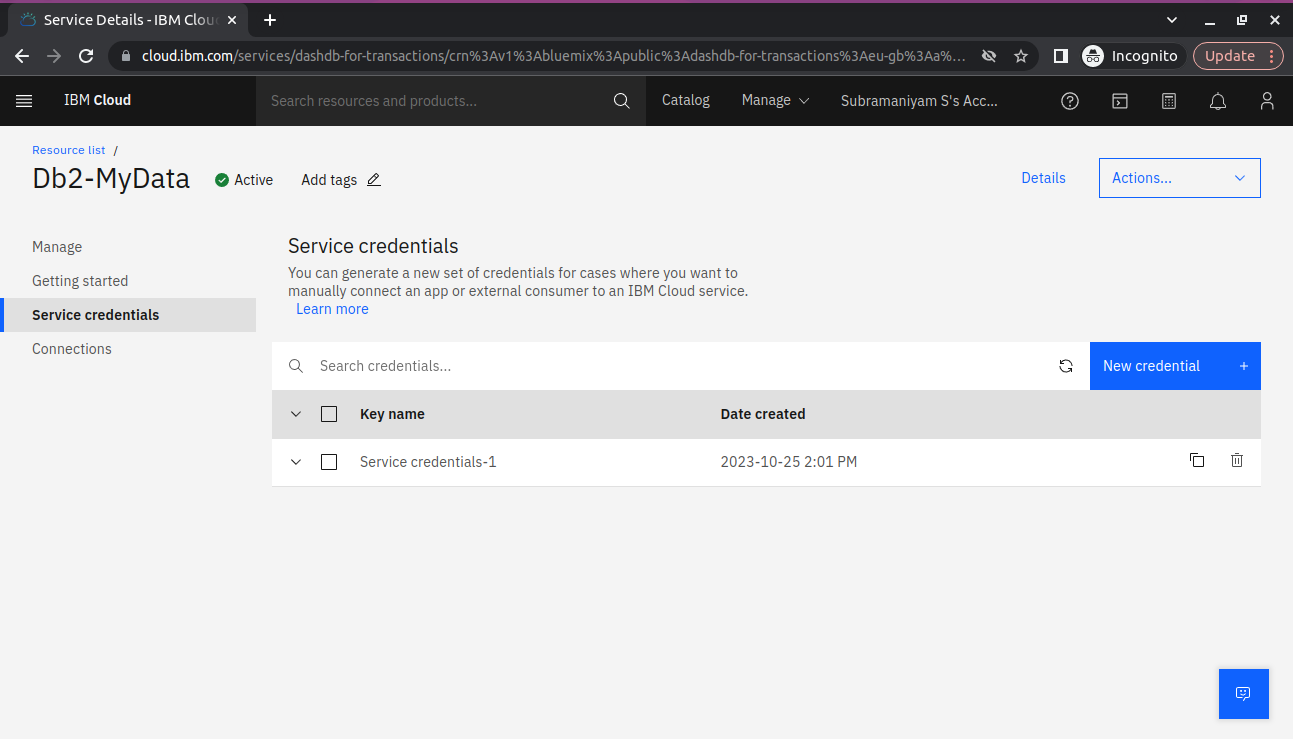
**Step 5 -** Then you go to catalog and add db2 to our resources. After successful creation go to resource list and see that.



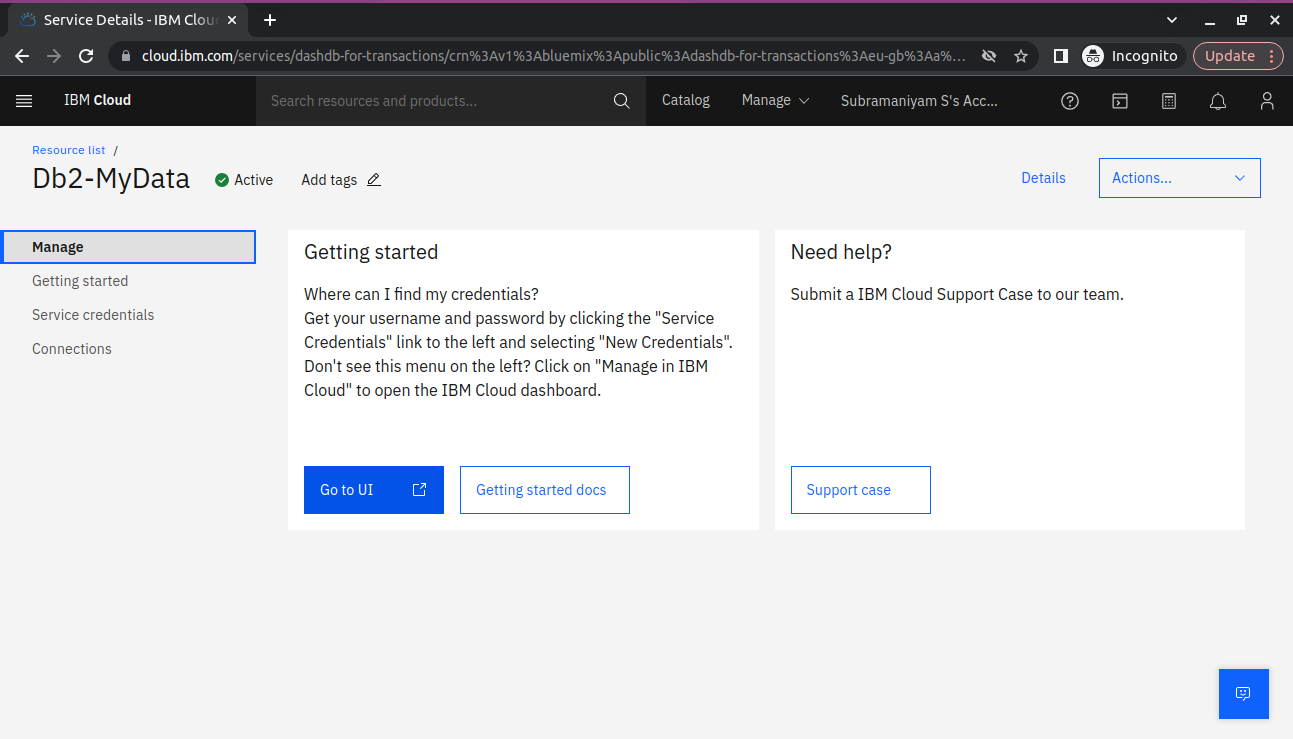
**Data exploration and Analysis:**

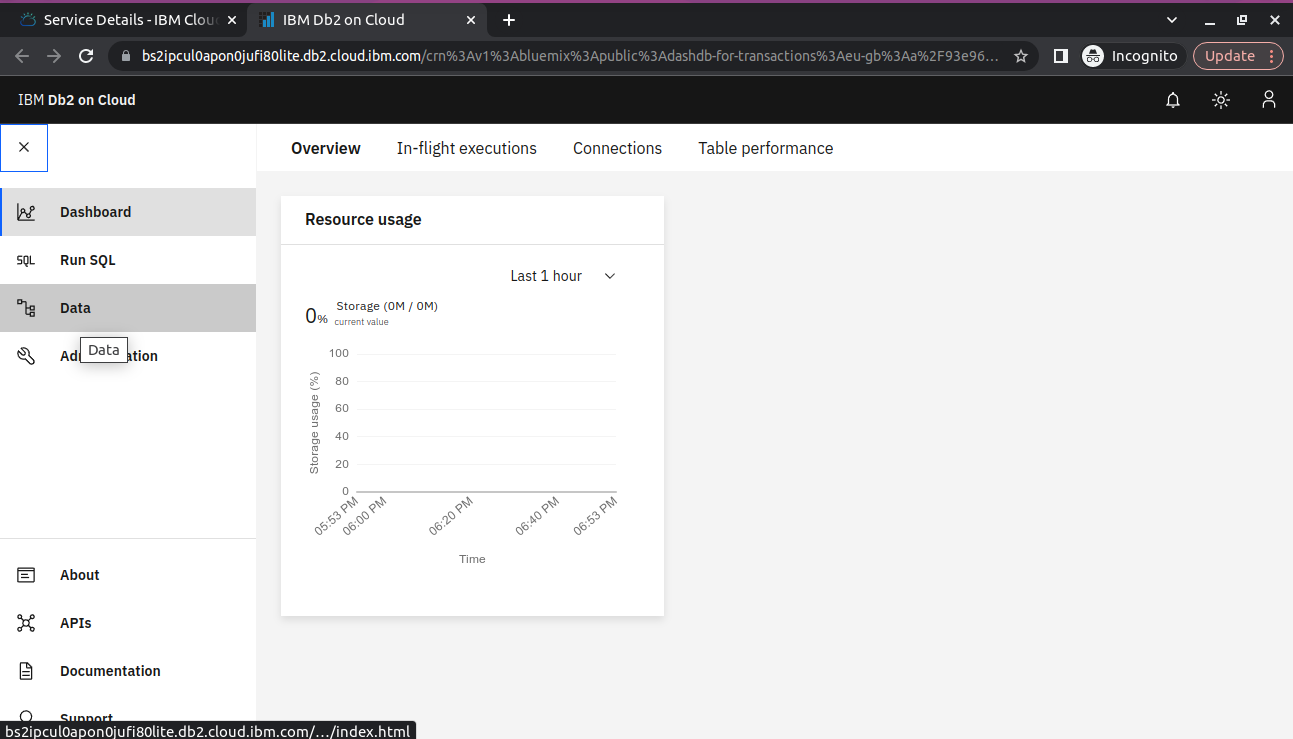
**Step 6 –** Then click the database you created in the resource poll and click the service credentials and create new services.



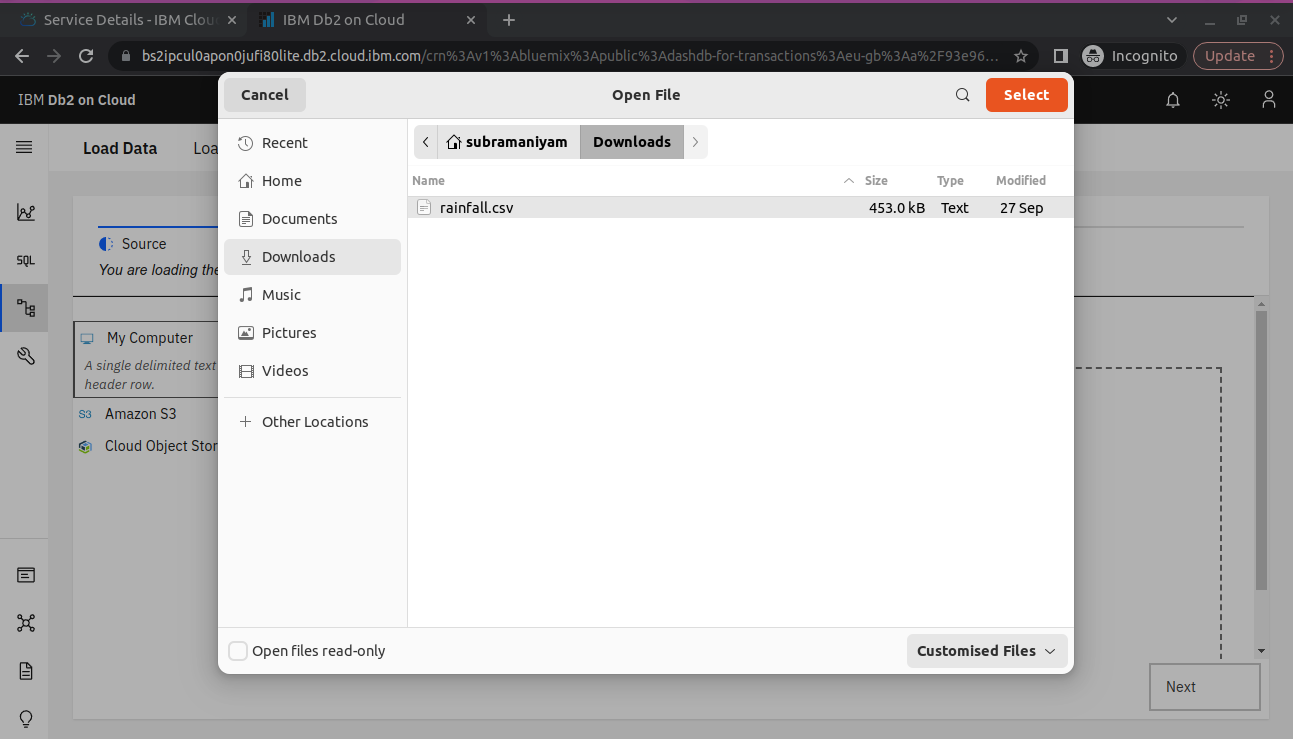
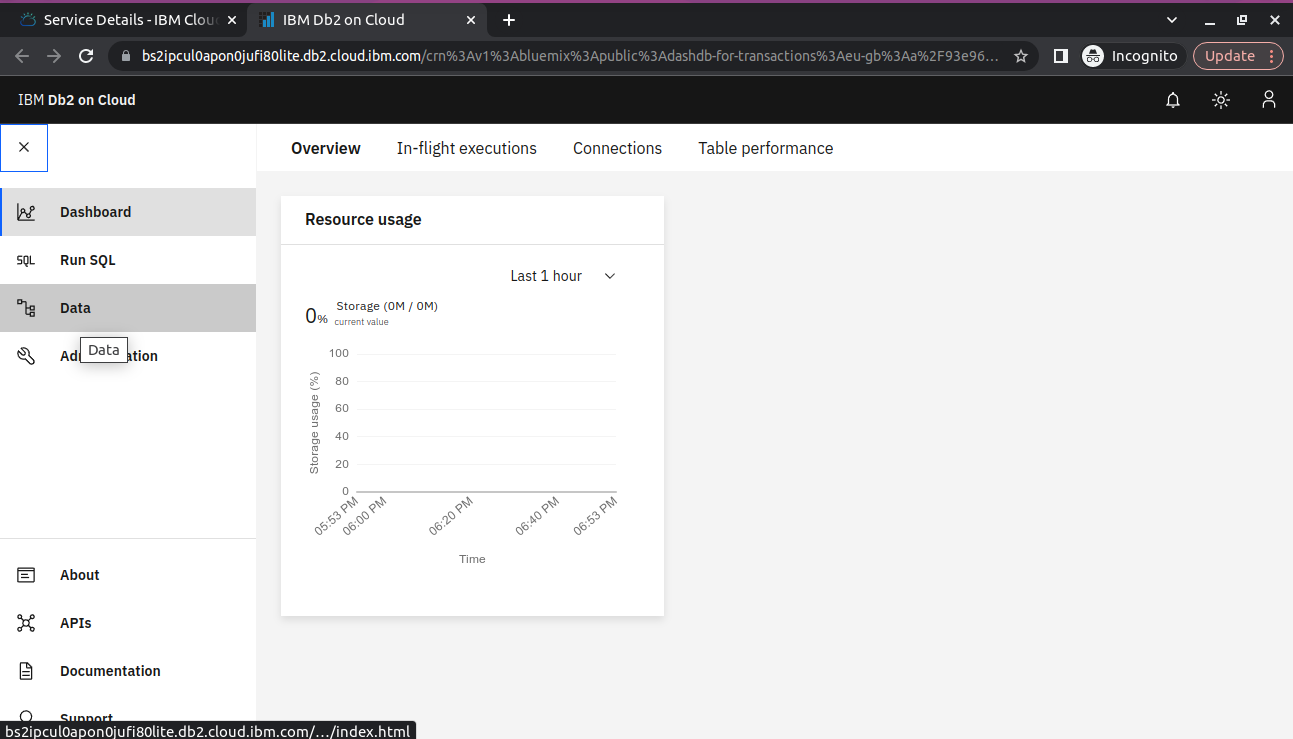


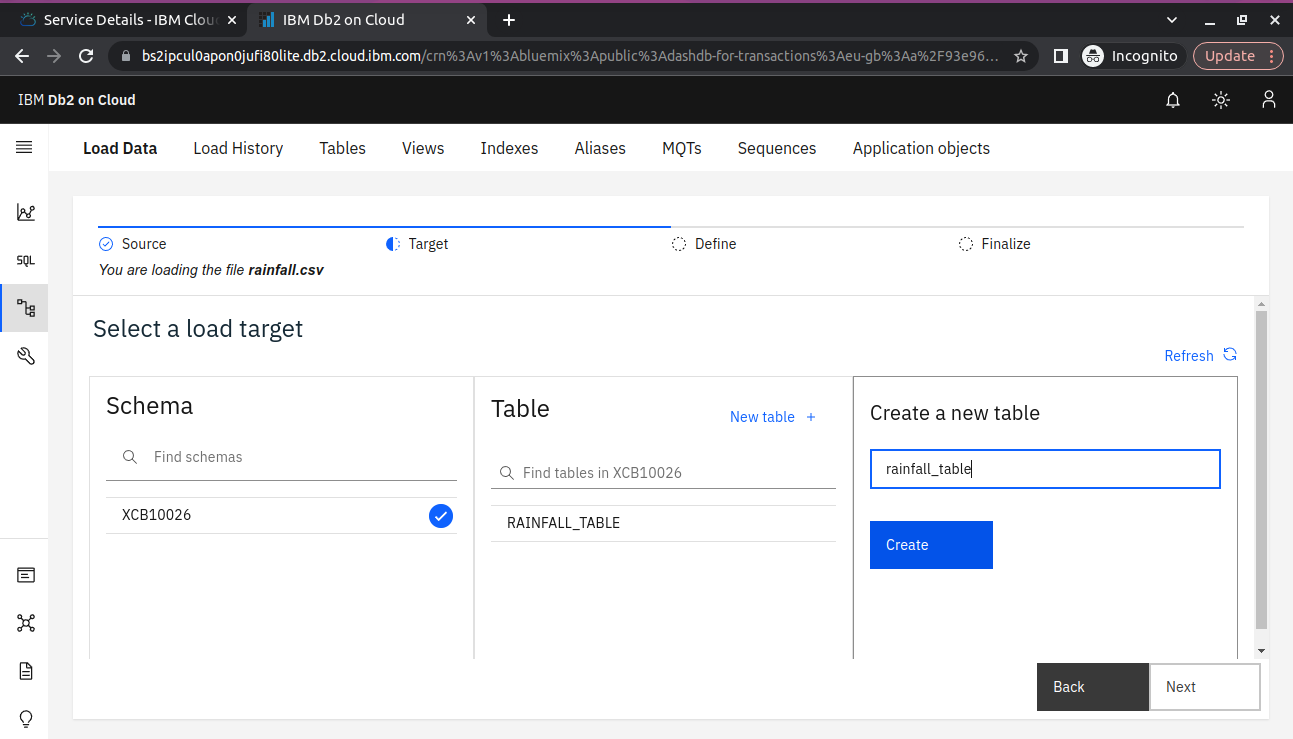
**Step 7 –** Then go to manage section and click go to UI.

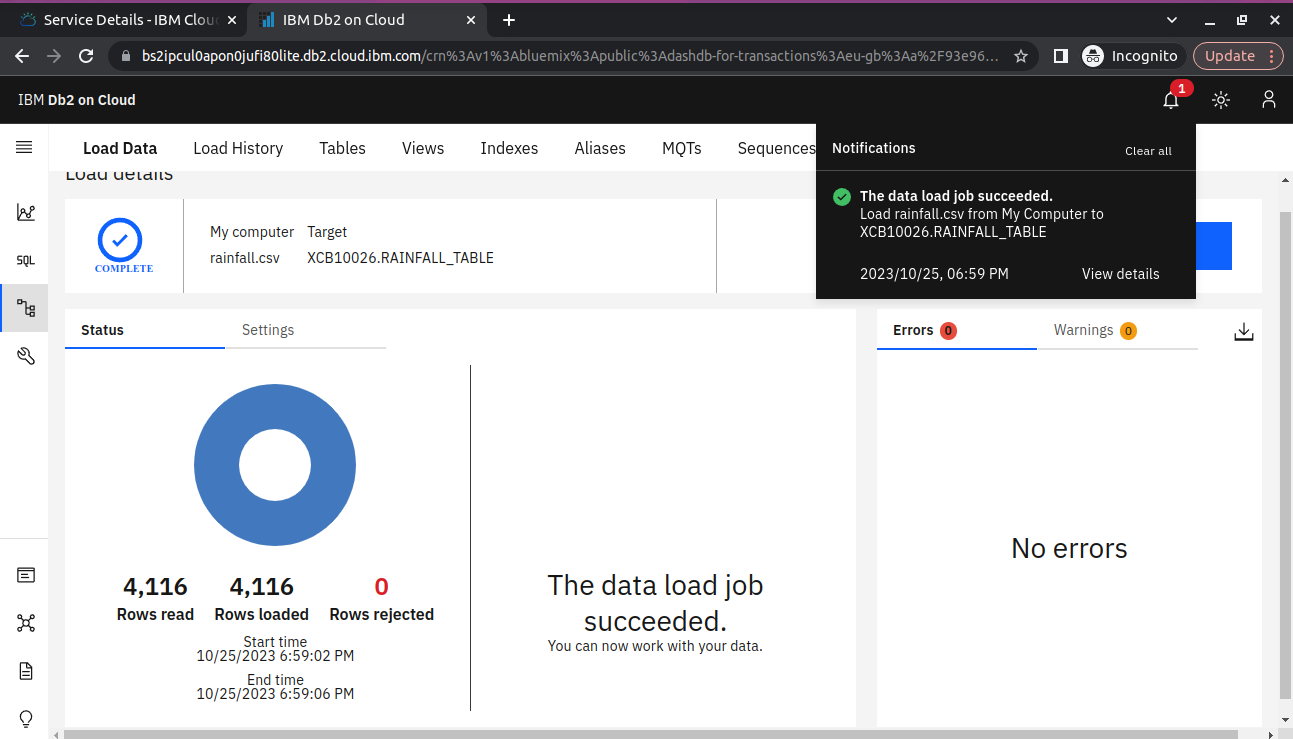




**Step 8 –** Go to data section and upload the dataset that is csv file.

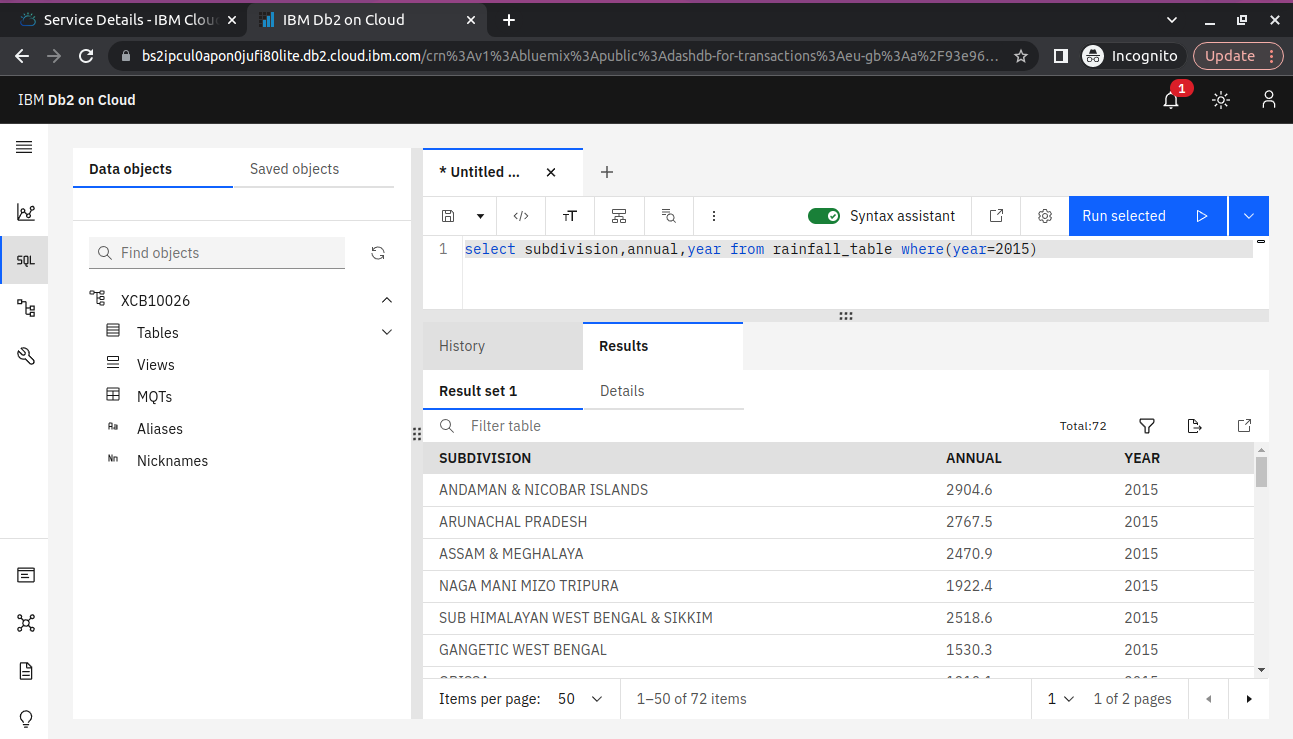




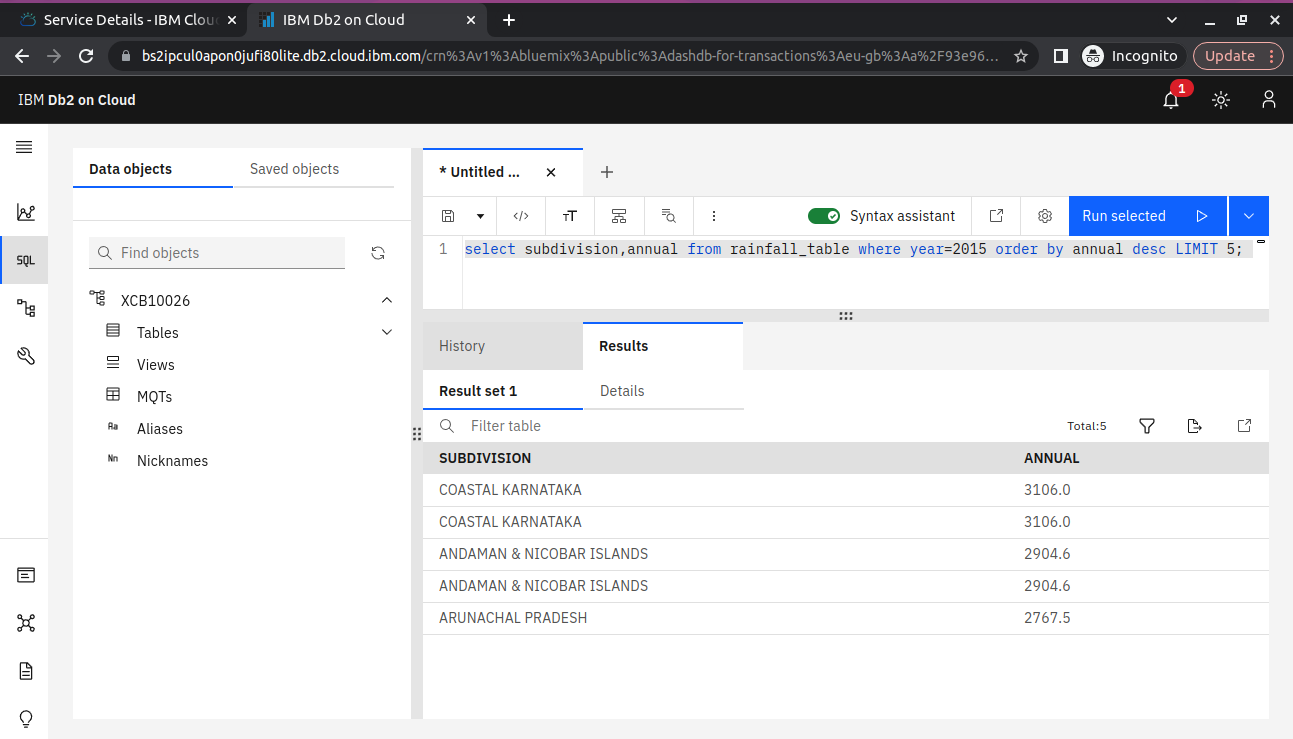


Finally, dataset can be uploaded successfully in db2.

**Step 9-** Using SQL queries to analyze the data.



Again, using the SQL queries to analyze the dataset to find the hidden trends and patterns.

Finally, we analyze the dataset using the IBM db2.

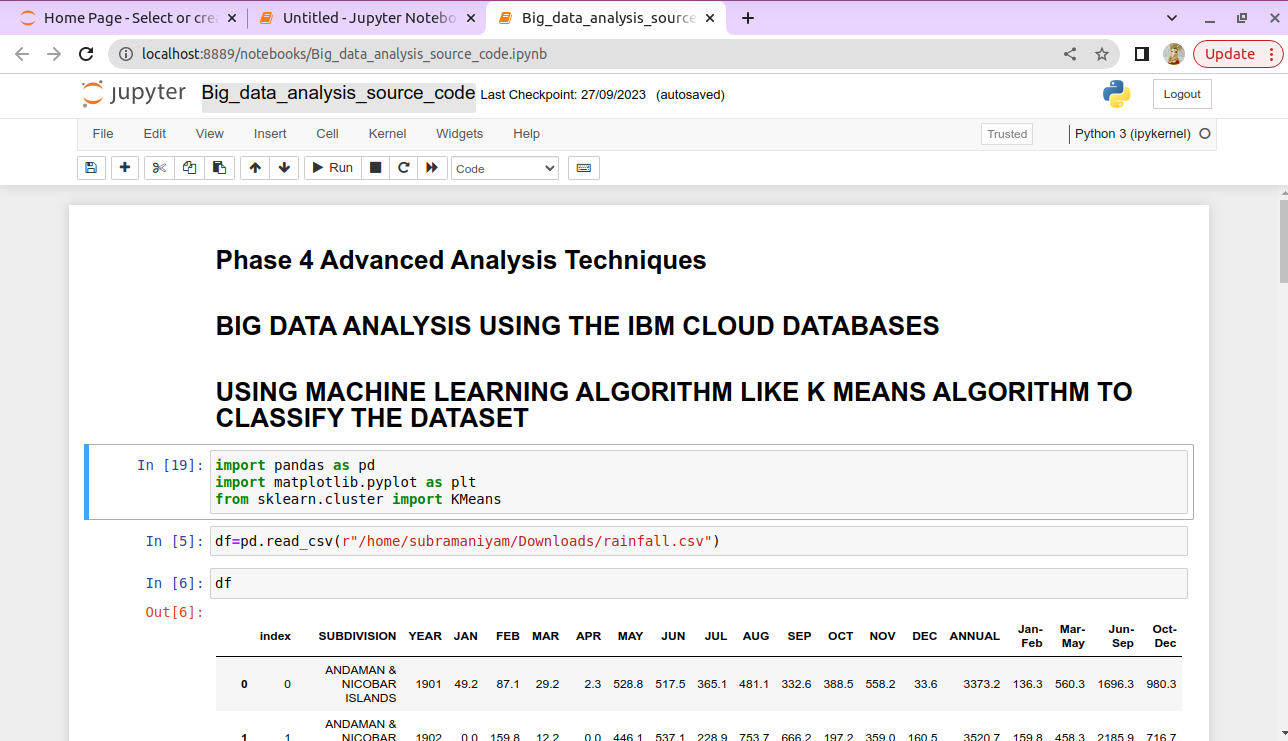
**Advanced Analytics Techniques:**

**Follow the below steps for Advanced Analytics Techniques:**

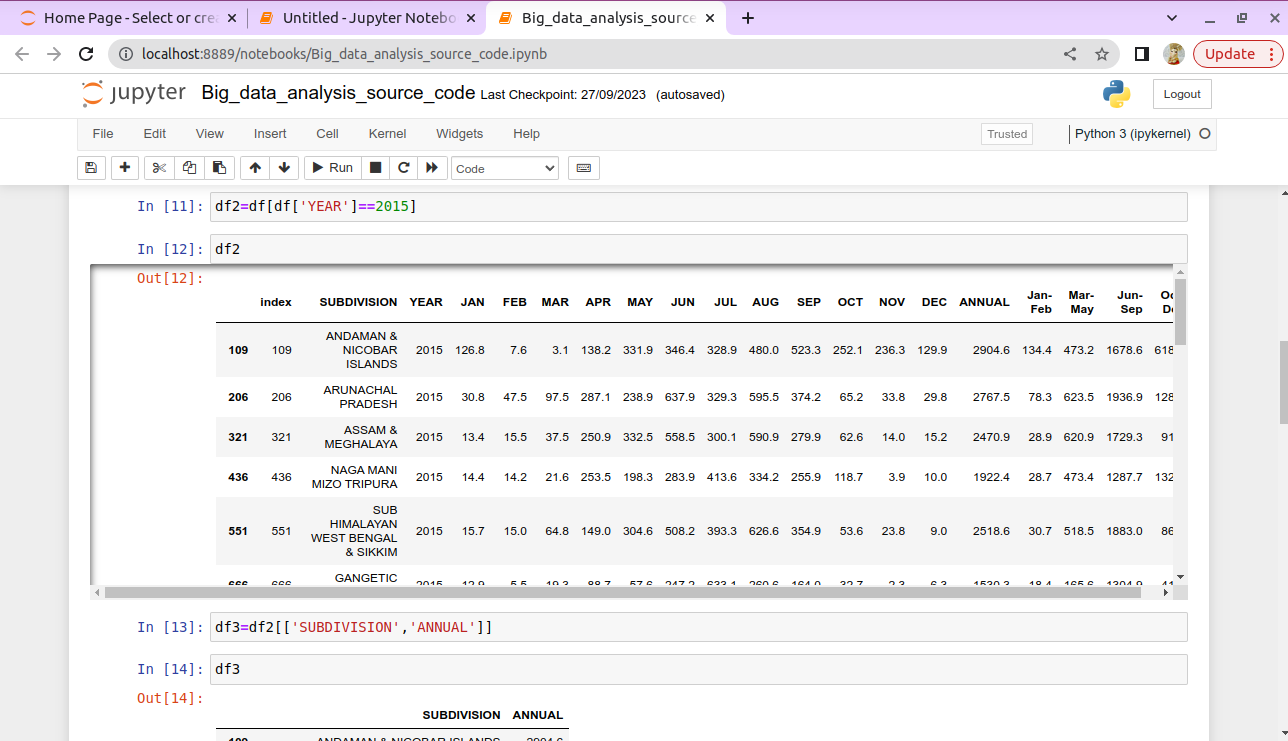
**NOTE:** We are going to use the Machine Learning Algorithm like K Means

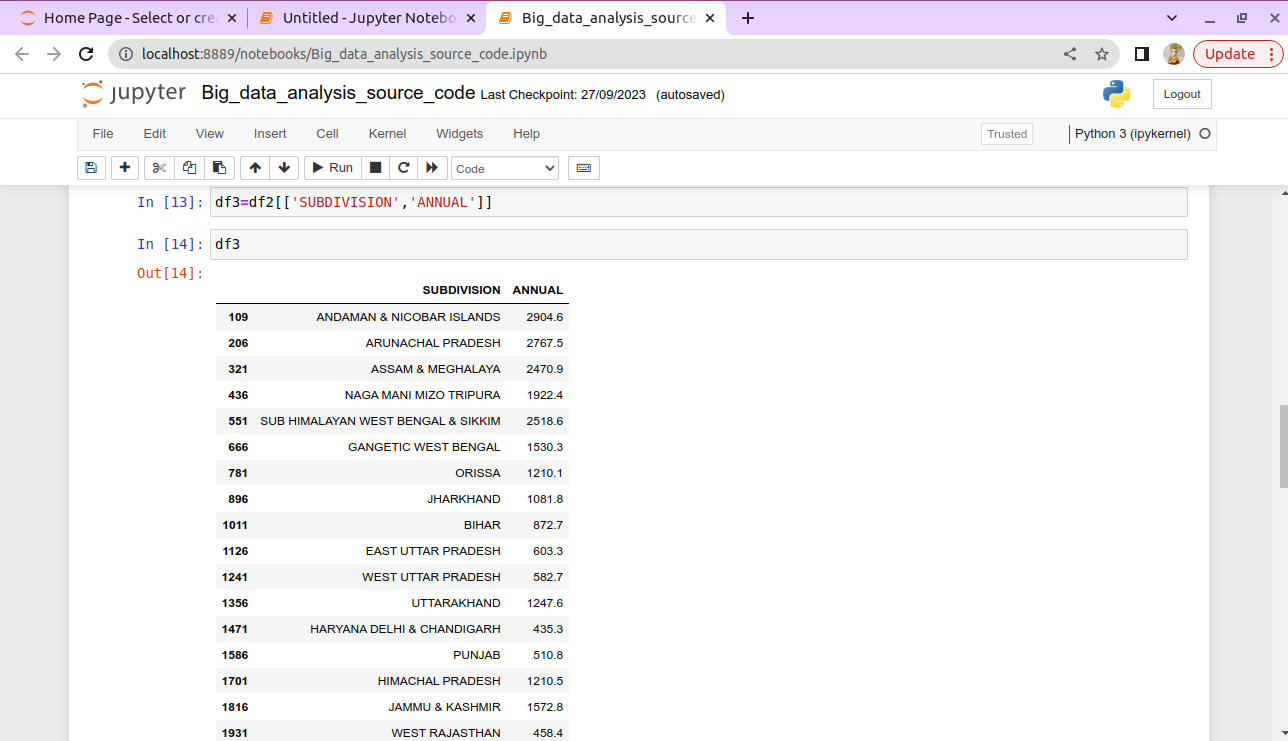
Clustering Algorithm for analysis.

**Step 1 –** Import the necessary libraries and the dataset in Jupyter Notebook.

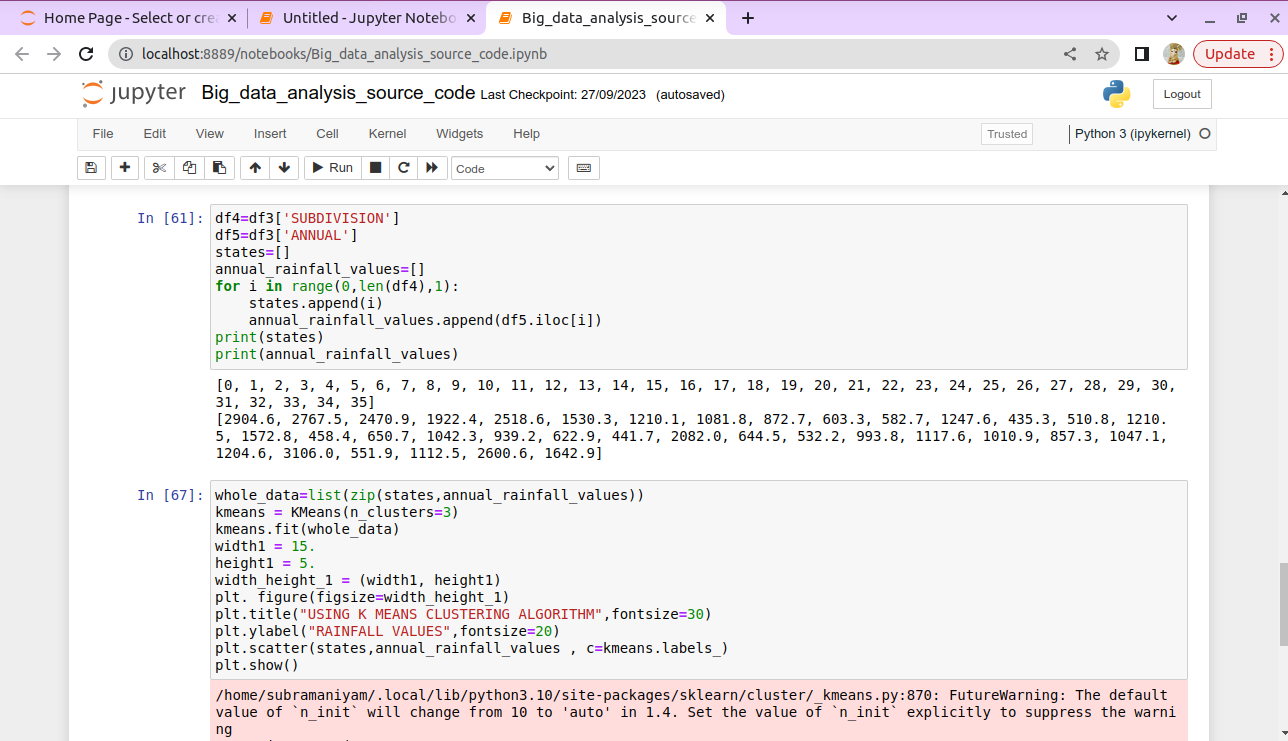


**Step 2 –** Clean and remove the noisy data in the dataset using python script.

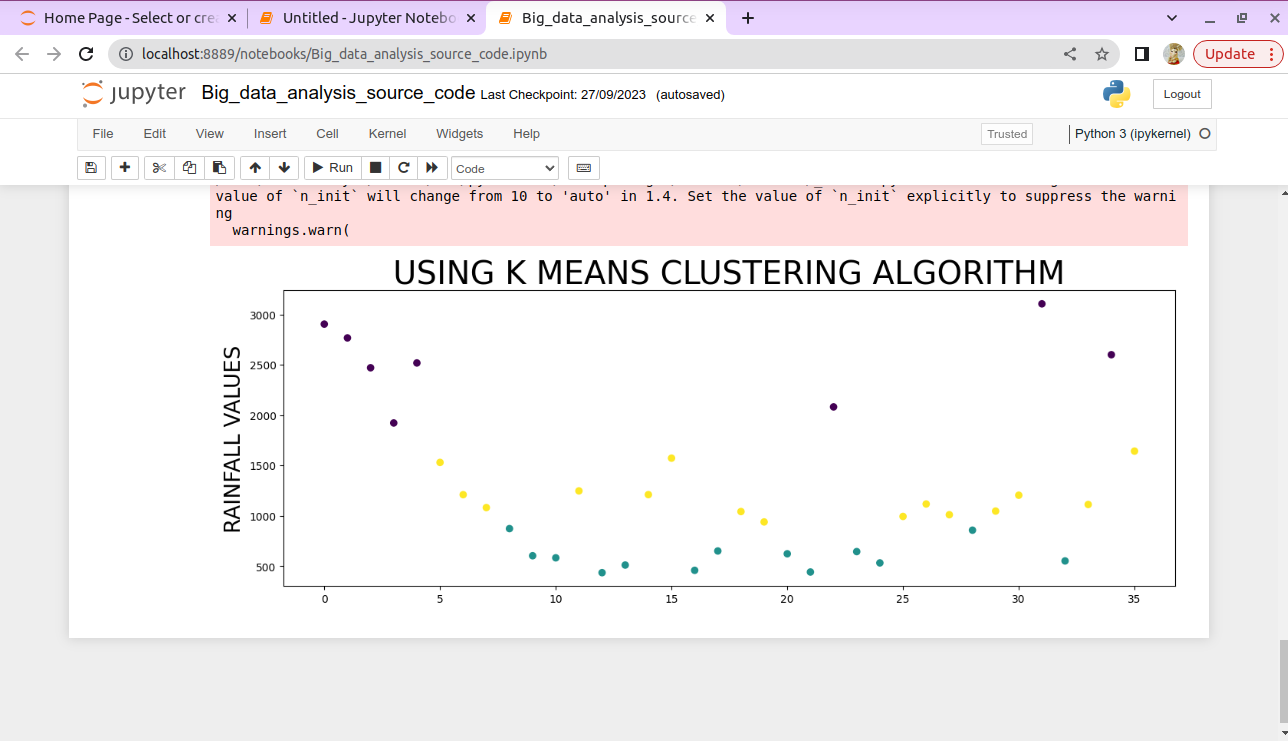




**Step 3 –** Store the necessary values into the empty array for plotting purpose.



**Step 4 –** After storing the values in array using K Means Clustering Algorithm to plot the graph and analyze the results.



**Visualization:** Create visualizations to showcase the analysis results. Use tools

like Matplotlib, Plotly, or IBM Watson Studio for creating graphs and charts.

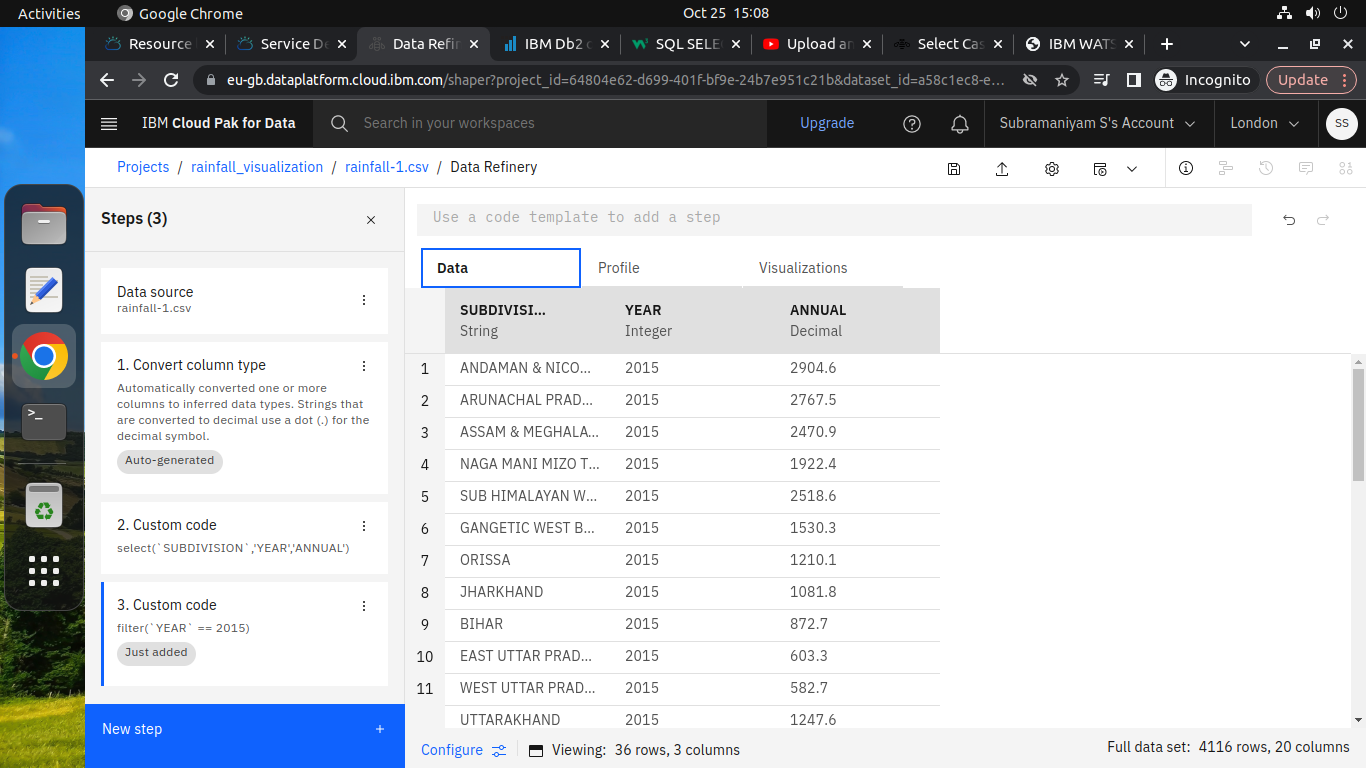
**Follow the below steps for Advanced Analytics Techniques:**

**NOTE:** We are going to use the IBM Watson Studio for creating graphs and charts.

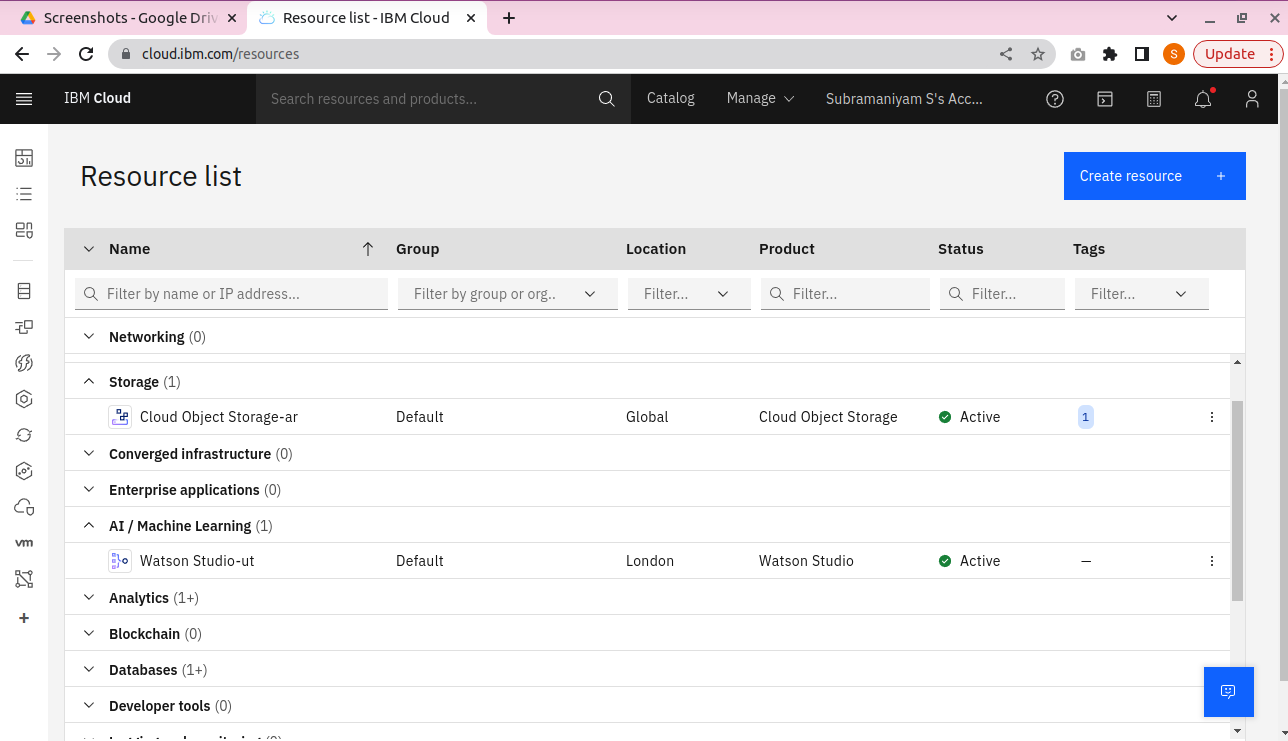
**Step 1 –** Open our cloud account and create the IBM Watson Studio then go to the

Resource Poll and click Artificial Intelligence and Machine Learning and choose

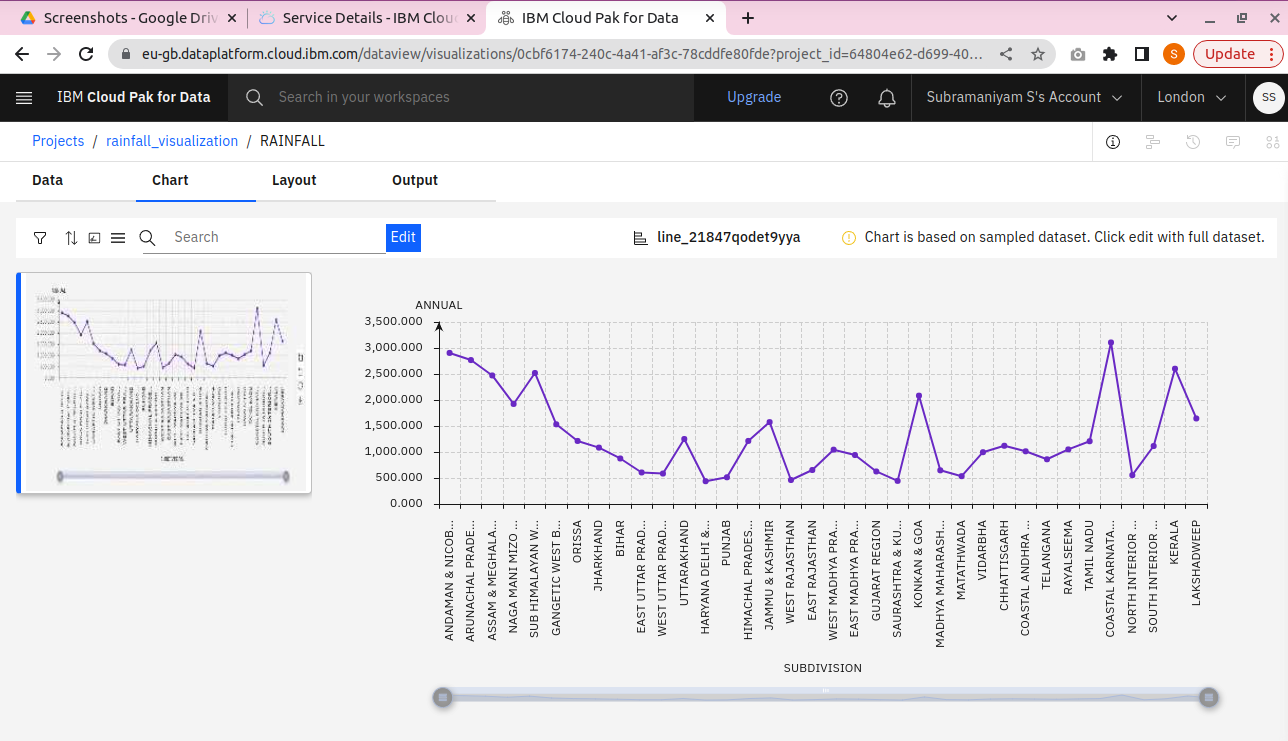
IBM Watson Studio.



**Step 2 –** Load the Dataset and put some queries to refine the data for our visualization.



**Step 3 –** Finally, using the refine script to perform the visualization.



**Conclusion:**

In this project, we clearly understand the big data analytics topics and how to perform analytics work using IBM db2 and visualization works using IBM Watson Studio and Python.