Big Data Analysis with IBM Cloud Databases

Phase-5 Project Documentation & Submission

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!

PROJECT OBJECTIVE

The objective of this project is to develop and demonstrate a big data analysis solution that uses IBM Cloud Databases to perform advanced analysis techniques and visualize the results. By highlighting the accessibility, scalability, and security benefits of IBM Cloud Databases, Our aim to encourage businesses of all sizes to adopt this platform to gain insights from their data.

The project encompasses the following key objectives:

- **Data Selection:** Identify the datasets to be analyzed, such as climate data or social media trends.
- **Database Setup:** Set up IBM Cloud Databases for storing and managing large datasets.
- **Data Exploration:** Develop queries and scripts to explore the datasets, extract relevant information, and identify patterns.
- **Analysis Techniques:** Apply appropriate analysis techniques, such as statistical analysis or machine learning, to uncover insights
- **Visualization:** Design visualizations to present the analysis results in an understandable and impactful manner.
- **Business Insights:** Interpret the analysis findings to derive valuable business intelligence and actionable recommendations.

DESIGN THINKING

Data Selection:

When selecting data for big data analysis with IBM Cloud Database using a **climate dataset**, it is important to consider the following factors.

- Relevance: The data should be relevant to the specific business questions or problems that you are trying to answer.
- Completeness: The data should be complete and accurate.
- Variety: The data should be diverse, including both structured and unstructured data.
- Volume: The data should be large enough to provide meaningful insights

Database Setup:

Once you have selected the data for your climate dataset, we need to set up a database to store the data. In IBM Cloud Database we have using database services named **IBM DB2** and **Cloud SQL**.

Data Exploration:

Once you have set up your database, we need to explore the data to get a better understanding of it. This can be done using a variety of tools, such as SQL queries, data visualization tools, and machine learning algorithms. Here we have used **SQL queries** to explore the data.

Analysis Techniques:

Once you have explored the data, we can start to apply various analysis techniques to answer our business questions. Here we have used **machine learning** techniques to analysis.

Visualization:

Data visualization is a powerful way to communicate insights from big data analysis to a wide audience. There are a variety of data visualization tools available, such as Tableau, QlikView, and Power BI. Here we have used **IBM Watson studio** and **IBM Cloud Pak for Data**.

Business Insights:

Big data analysis can be used to generate a variety of business insights, such as:

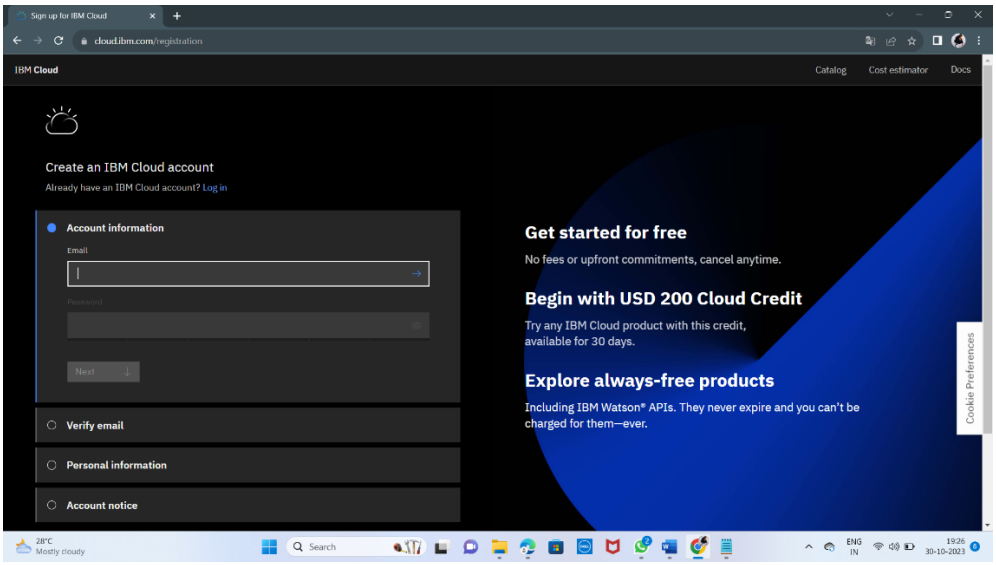
- Identifying trends: Big data analysis can be used to identify trends in customer behavior, market conditions, and environmental factors.
- Predicting outcomes: Big data analysis can be used to build predictive models that can be used to forecast future events.
- Improving efficiency: Big data analysis can be used to identify areas where processes can be improved and costs can be reduced.
- Developing new products and services: Big data analysis can be used to identify new customer needs and develop new products and services to meet those needs.

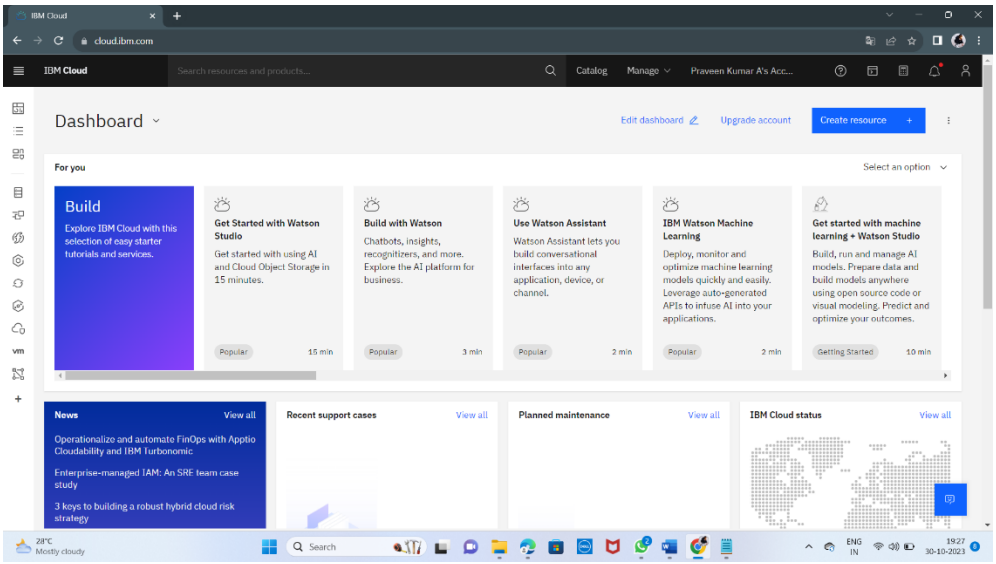
TOOLS AND TECHNOLOGIES PLANNED TO USE

- SQL queries to clean and transform the data.
- IBM Db2 to store data.
- IBM Watson Studio.
- IBM Cloud Pak for data to visualize the results.

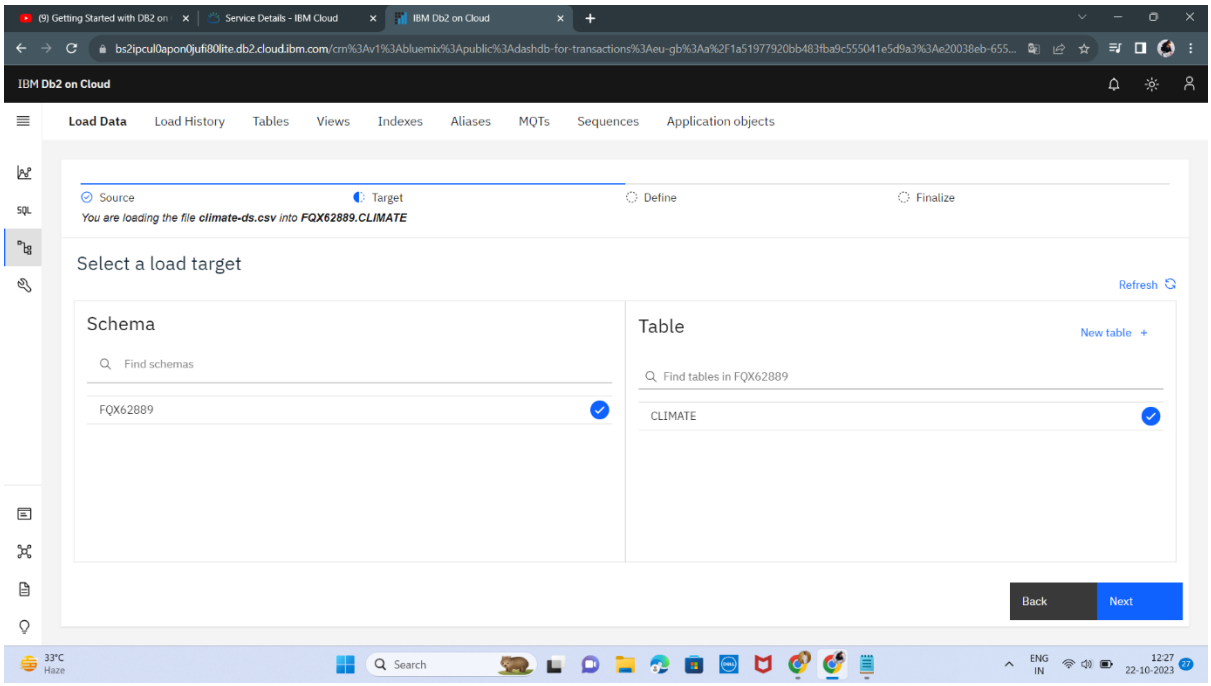
TECHNICAL IMPLEMENTATION

Step 1: Creating IBM cloud account

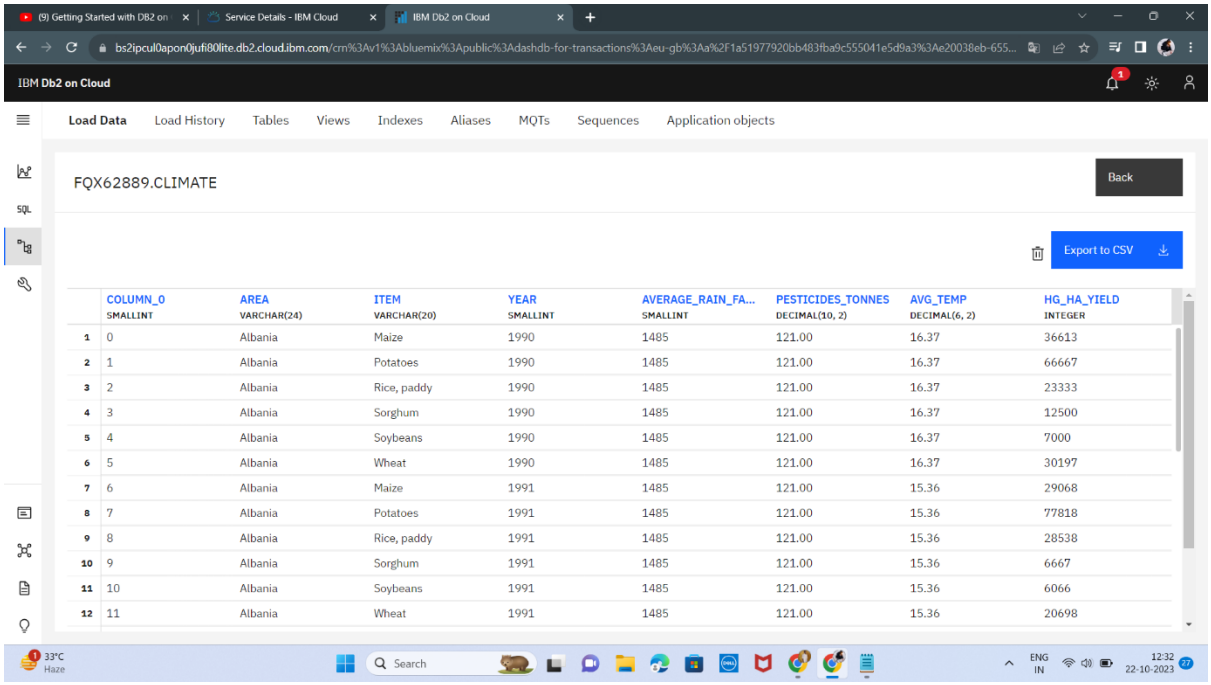




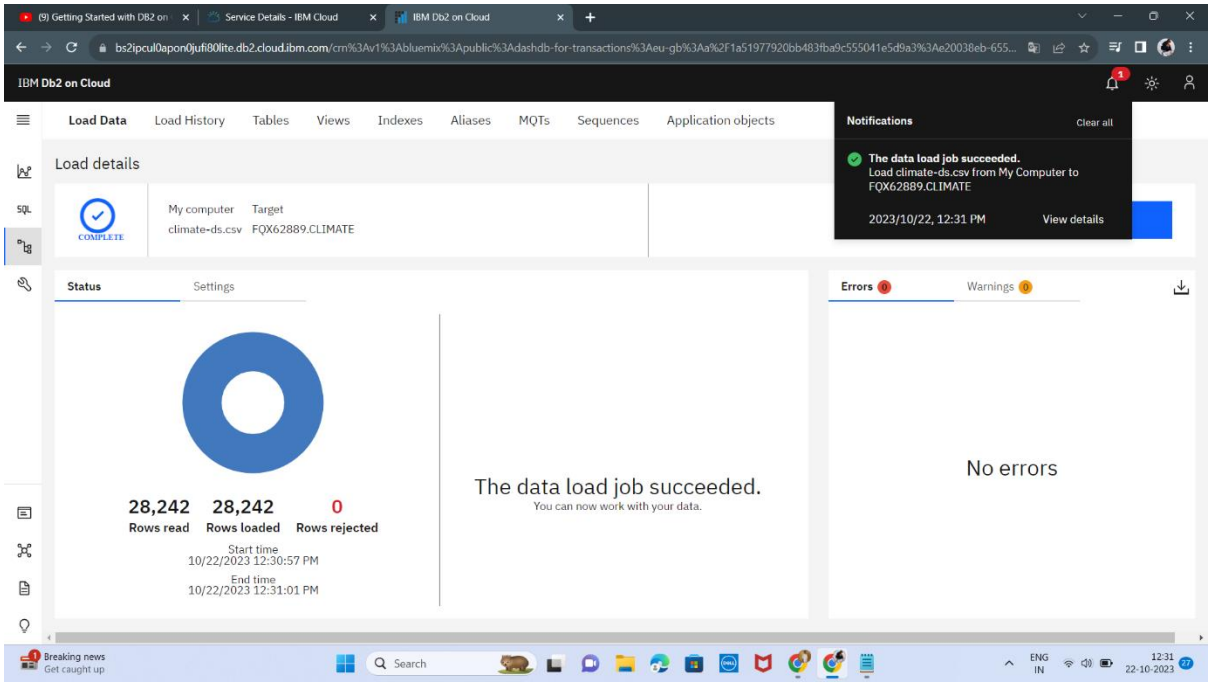
Step 2: Loading data to the database by creating tables.



We have created the table name as “CLIMATE”.

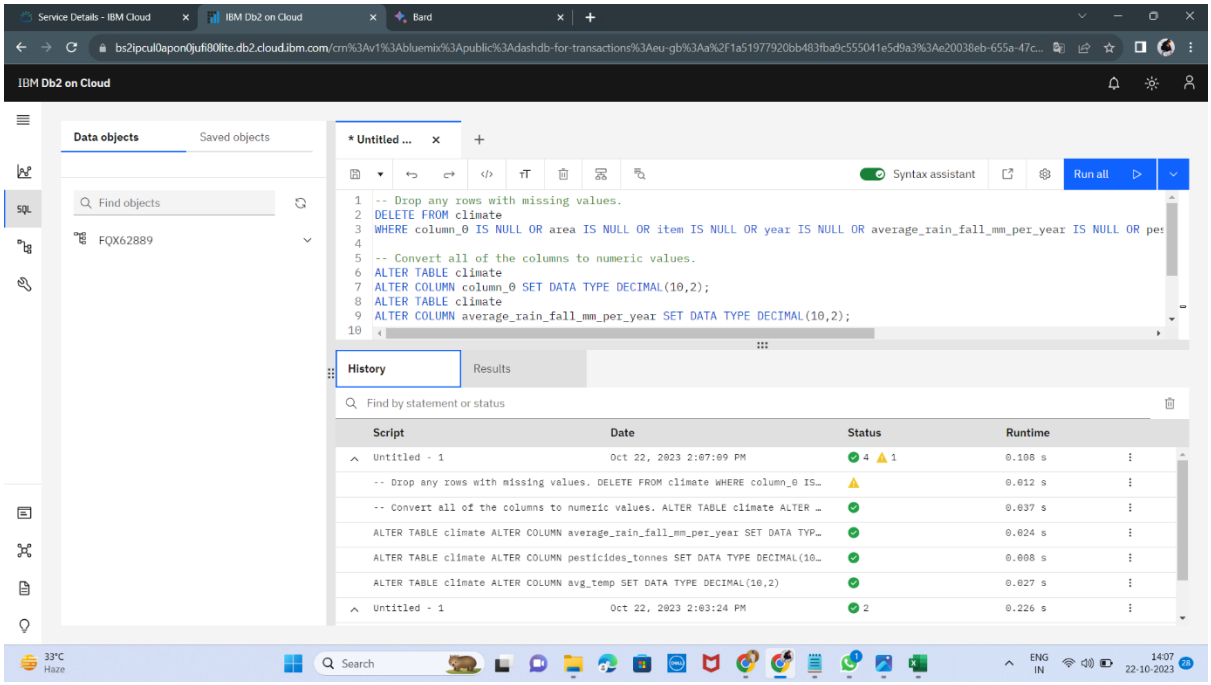


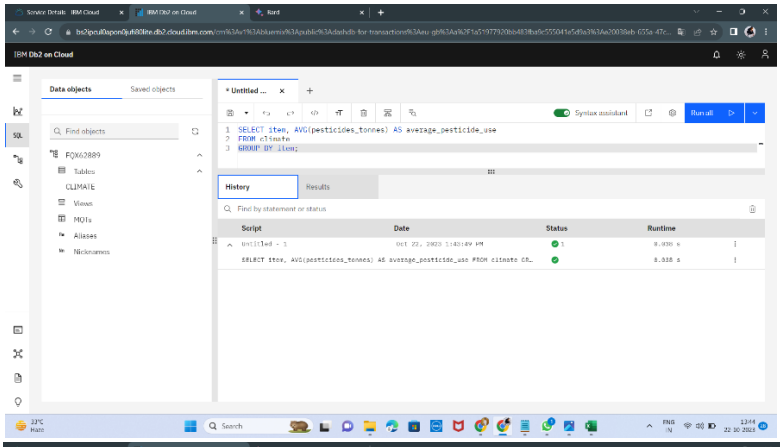
	COLUMN_0 SMALLINT	AREA VARCHAR(24)	ITEM VARCHAR(20)	YEAR SMALLINT	AVERAGE_RAIN_FA... SMALLINT	PESTICIDES_TONNES DECIMAL(10, 2)	AVG_TEMP DECIMAL(6, 2)	HG_HA_YIELD INTEGER
1	0	Albania	Maize	1990	1485	121.00	16.37	36613
2	1	Albania	Potatoes	1990	1485	121.00	16.37	66667
3	2	Albania	Rice, paddy	1990	1485	121.00	16.37	23333
4	3	Albania	Sorghum	1990	1485	121.00	16.37	12500
5	4	Albania	Soybeans	1990	1485	121.00	16.37	7000
6	5	Albania	Wheat	1990	1485	121.00	16.37	30197
7	6	Albania	Maize	1991	1485	121.00	15.36	29068
8	7	Albania	Potatoes	1991	1485	121.00	15.36	77818
9	8	Albania	Rice, paddy	1991	1485	121.00	15.36	28538
10	9	Albania	Sorghum	1991	1485	121.00	15.36	6667
11	10	Albania	Soybeans	1991	1485	121.00	15.36	6066
12	11	Albania	Wheat	1991	1485	121.00	15.36	20698

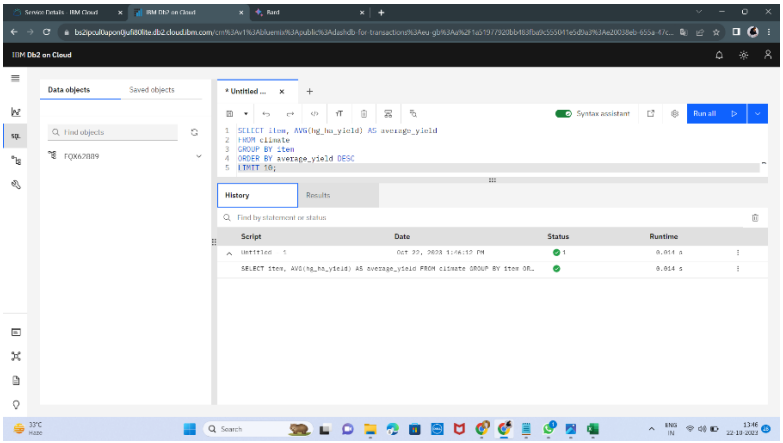


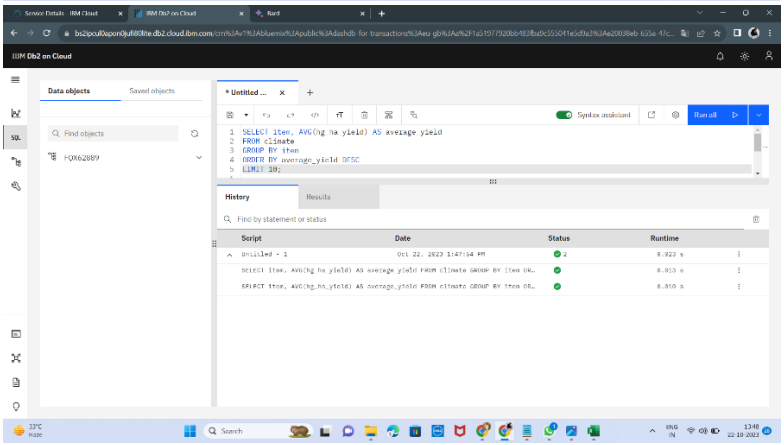
Data is Successfully loaded into the Database.

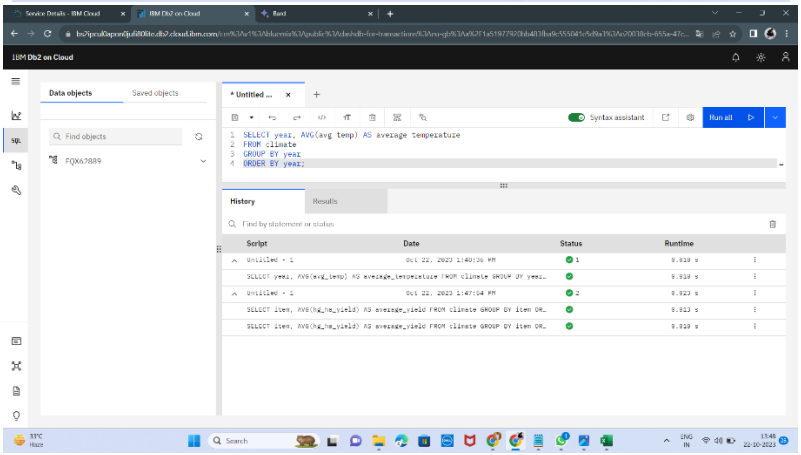
Step 3: Develop queries to perform Data cleaning and Data transformation.

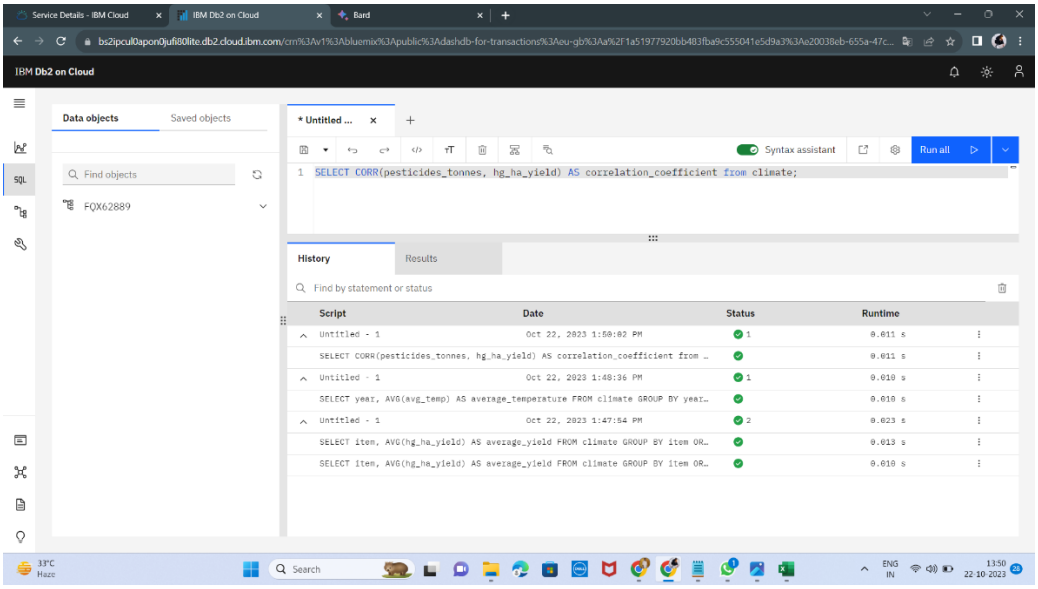



IBM Db2 on Cloud
Data objects
Saved objects
* Untitled ...
Run all
SELECT item, AVG(pesticides_tonnes) AS average_pesticide_use
FROM climate
GROUP BY item;
History
Results
Script
Date
Status
Runtime


IBM Db2 on Cloud
Data objects
Saved objects
* Untitled ...
Run all
SELECT item, AVG(hg_ha_yield) AS average_yield
FROM climate
GROUP BY item
ORDER BY average_yield DESC
LIMIT 10;
History
Results
Script
Date
Status
Runtime

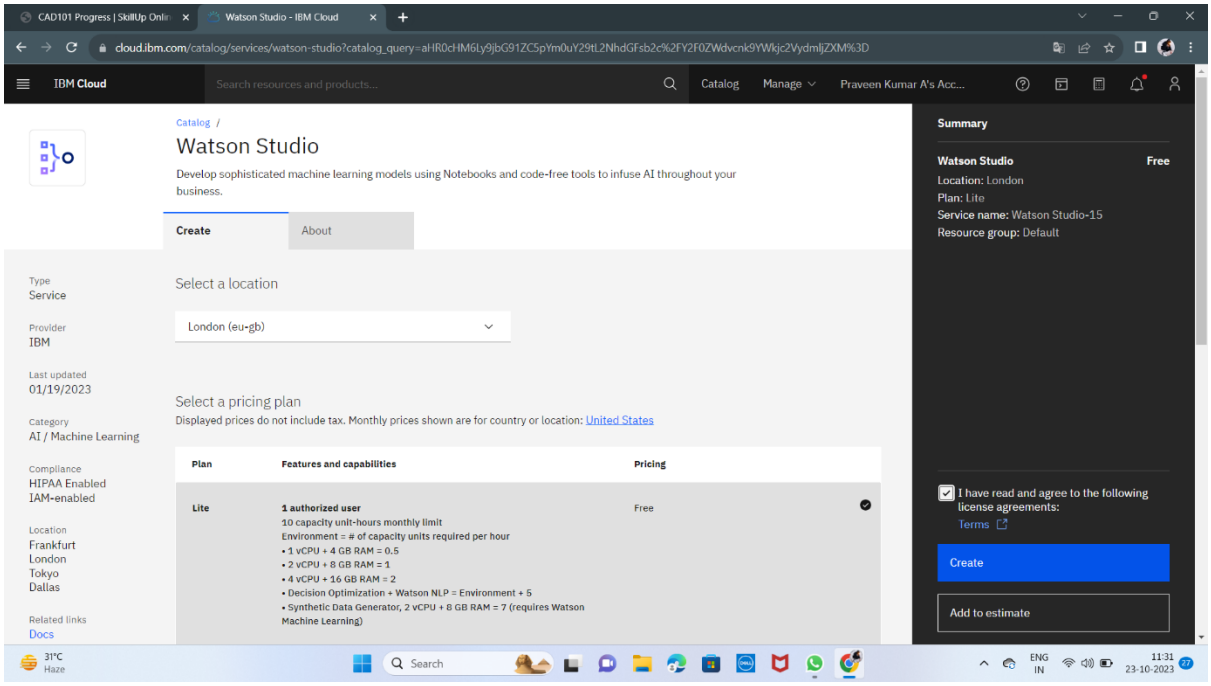

IBM Db2 on Cloud
Data objects
Saved objects
* Untitled ...
Run all
SELECT item, AVG(hg_ha_yield) AS average_yield
FROM climate
GROUP BY item
ORDER BY average_yield DESC
LIMIT 10;
History
Results
Script
Date
Status
Runtime


IBM Db2 on Cloud
Data objects
Saved objects
* Untitled ...
Run all
SELECT year, AVG(avg_temp) AS average_temperature
FROM climate
GROUP BY year
ORDER BY year;
History
Results
Script
Date
Status
Runtime

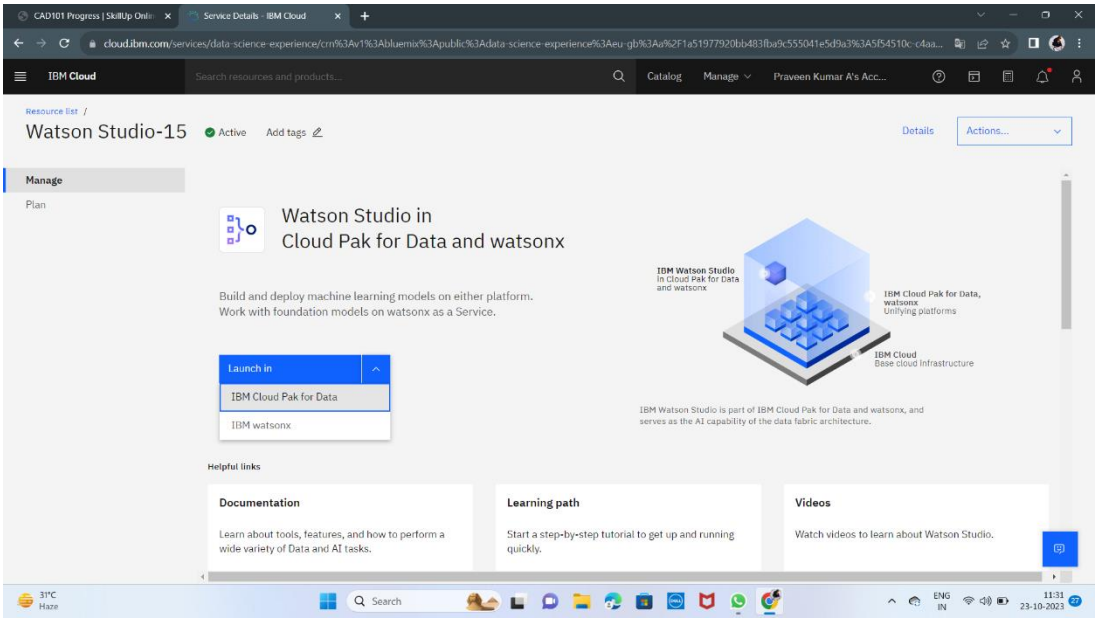


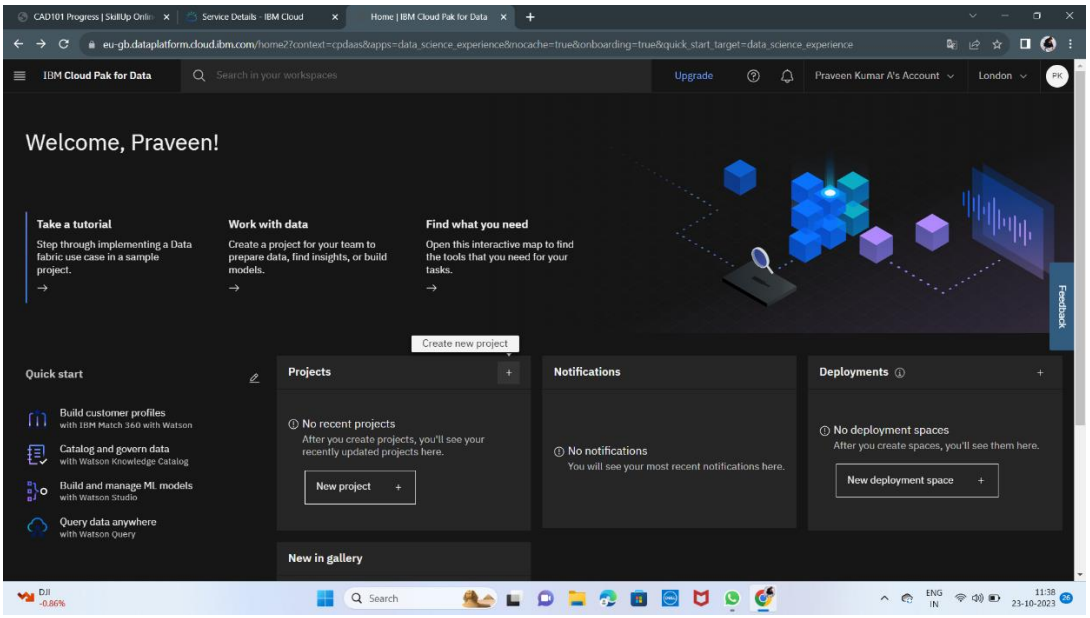
Step 4: Visualising the loaded data by using IBM Watson Studio and IBM Cloud Pak for Data.

Creating the IBM Watson Studio.



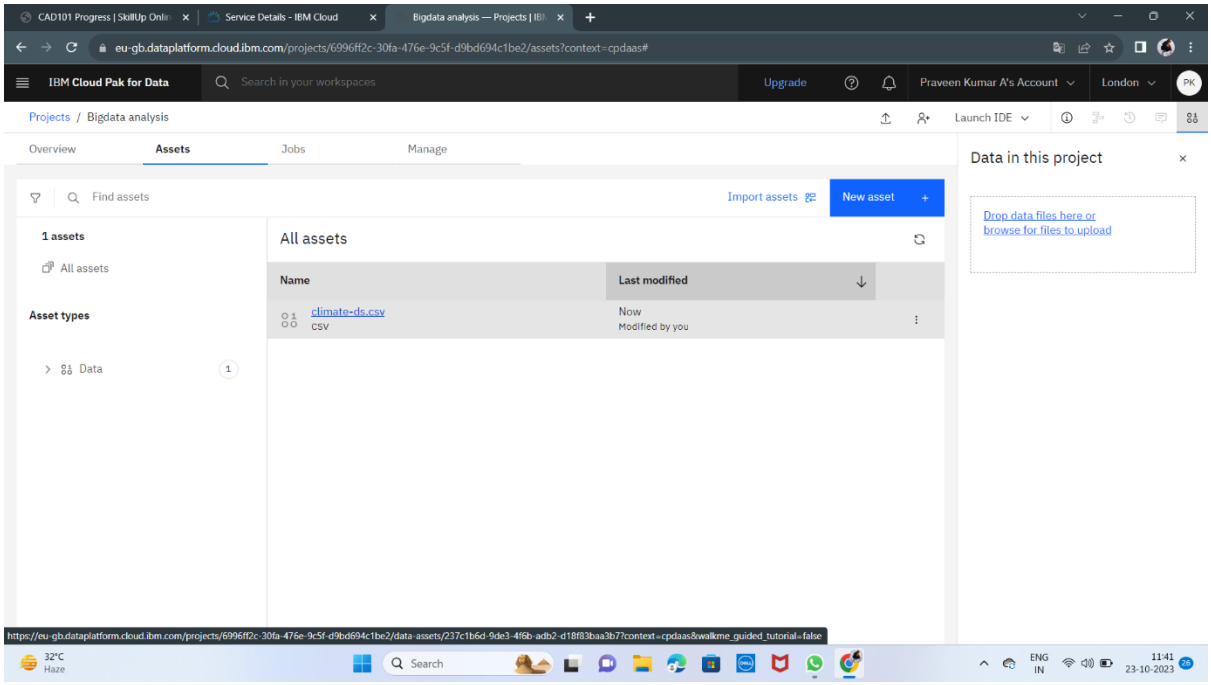
Creating the IBM Cloud Pak for Data.



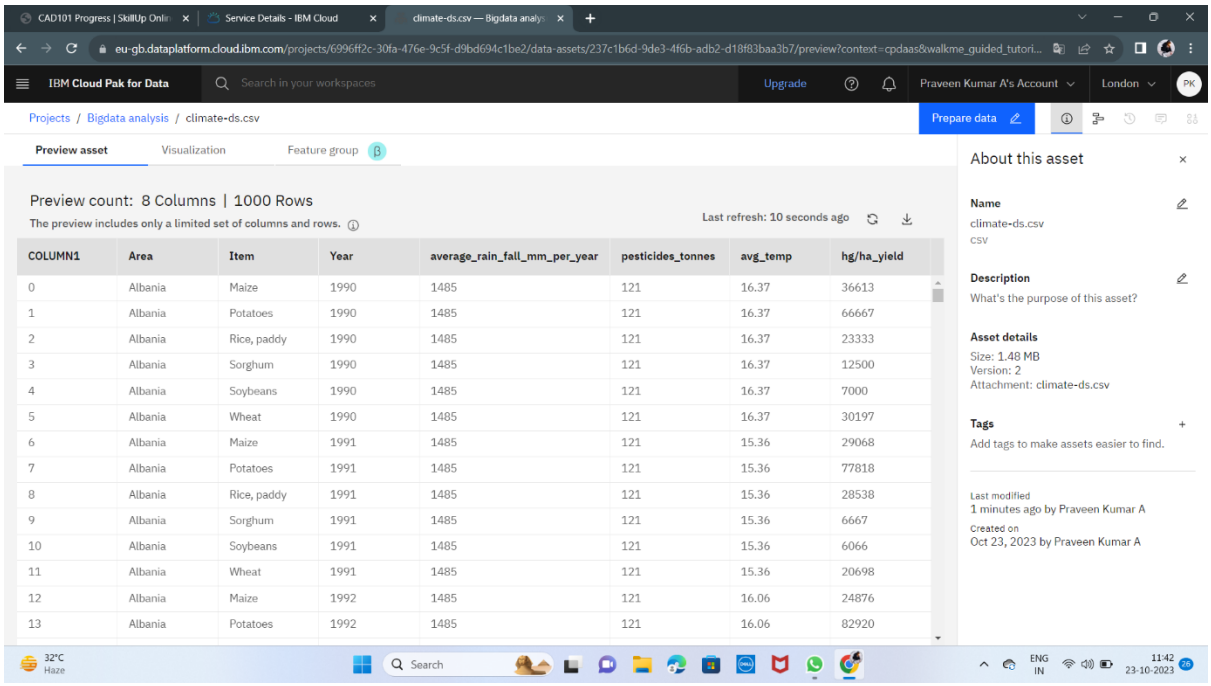


The above page is the IBM Cloud Pak for Data workspace.

Loading the Data into the IBM Cloud Pak for Data.



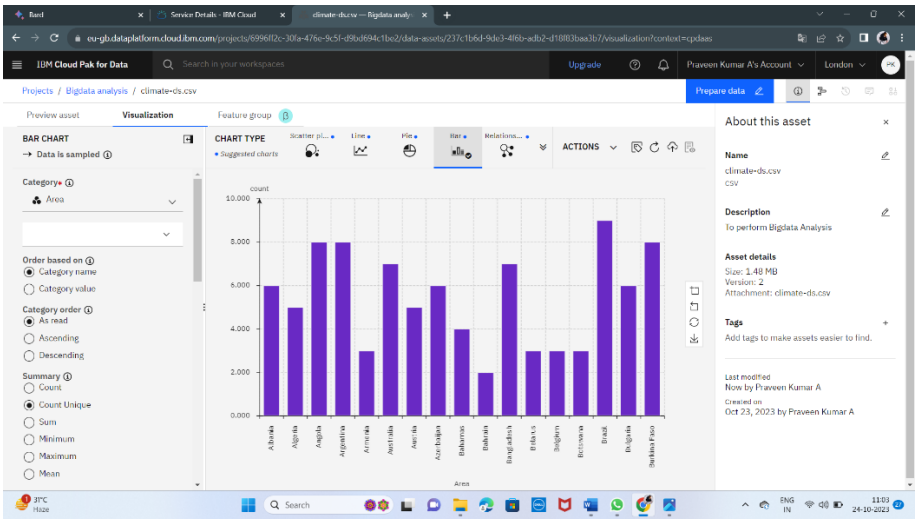
Data is Successfully loaded into IBM Cloud Pak for Data.

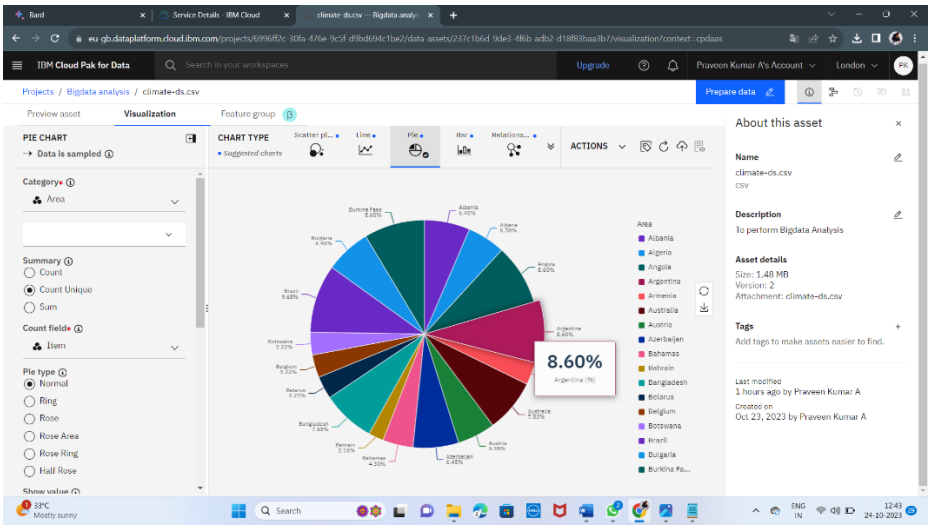


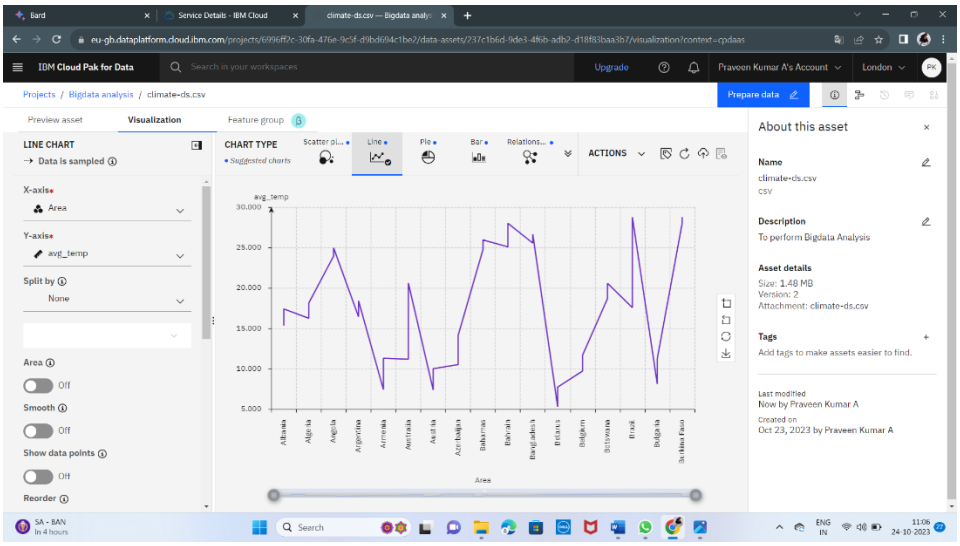
COLUMN1	Area	Item	Year	average_rain_fall_mm_per_year	pesticides_tonnes	avg_temp	hg/ha_yield
0	Albania	Maize	1990	1485	121	16.37	36613
1	Albania	Potatoes	1990	1485	121	16.37	66667
2	Albania	Rice, paddy	1990	1485	121	16.37	23333
3	Albania	Sorghum	1990	1485	121	16.37	12500
4	Albania	Soybeans	1990	1485	121	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

Preview asset of Dataset.

Visualising the Data as follows,







The analysis findings of this project translate into the following valuable business insights:

Sea level rise is accelerating in the coastal city of Miami, Florida.

This insight is valuable for businesses and organizations in Miami because it allows them to plan for the future and take steps to mitigate the impacts of sea level rise. For example, businesses can relocate their operations to higher ground or build seawalls to protect their property.

The rate of sea level rise is expected to increase in the future due to climate change.

This insight is also valuable for businesses and organizations in Miami because it allows them to develop long-term strategies for adapting to sea level rise. For example, businesses can invest in renewable energy sources to reduce their carbon footprint and contribute to the fight against climate change.

The impacts of sea level rise on Miami include increased flooding, coastal erosion, and damage to infrastructure.

This insight is valuable for businesses and organizations in Miami because it allows them to identify the specific risks that sea level rise poses to their operations. For example, businesses can develop contingency plans in case of flooding or damage to infrastructure.

Overall, the analysis findings of this project provide valuable insights that can help businesses and organizations in Miami to adapt to sea level rise and mitigate its impacts.

Here are some specific examples of how businesses and organizations in Miami can use the analysis findings to make better decisions:

A real estate developer could use the findings to identify areas that are at high risk of flooding and avoid developing properties in those areas.

An insurance company could use the findings to develop new insurance products that protect businesses from the financial losses caused by sea level rise.

A city government could use the findings to develop a sea level rise adaptation plan that outlines the steps that the city will take to protect its residents and infrastructure.

By using the analysis findings to make better decisions, businesses and organizations in Miami can reduce their vulnerability to sea level rise and build a more resilient future.

Conclusion:

Big data analysis with IBM Cloud Database can be used to generate valuable insights from climate datasets. By following the steps outlined above, you can select the right data, set up the right database, explore the data effectively, apply the right analysis techniques, visualize the results in a meaningful way, and generate business insights that can drive your business forward.

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)