**Project 7: Data Warehousing with IBM Cloud Db2**

**Warehouse**

**Phase 5: Project Documentation & Submission.**

# PROJECT OBJECTIVE

Our project main ob to tackle the challenge of modernizing our data management by designing and implementing a robust data warehouse using IBM Cloud Db2 Warehouse. This project is driven by the need to harness the potential of the organization's data, coming from diverse sources, and to empower data architects with the tools needed for insightful data analysis and informed decisionmaking.

**The project encompasses the following key objectives:**

1. Data Warehouse Structure: We will define a flexible and scalable schema and structure for our data warehouse that can efficiently handle data from various sources.
2. Data Integration: The project will involve identifying key data sources within our organization and devising a strategy for integrating this data seamlessly into the data warehouse.
3. ETL (Extract, Transform, Load) Processes: To ensure data accuracy and relevance, we will develop robust ETL processes for extracting data from source systems, transforming it into a usable format, and loading it efficiently into the data warehouse. iv. Data Exploration: We aim to create an intuitive and interactive environment that allows data architects to explore and analyze the data effectively.

v. Actionable Insights: Ultimately, our goal is to deliver actionable insights from the data analysis, enabling our organization to make data-driven decisions.

# DESIGN THINKING

➢ **Data Warehouse Structure**

To define the data warehouse structure, we will take the following practical steps:

* Data Profiling: We will conduct thorough data profiling to gain insights into the data's characteristics, including data types, relationships, and quality.
* Entity-Relationship Diagram (ERD): We will create an ERD to visualize the data model, making it easier to design schemas and tables.

➢ **Data Integration**

For effective data integration, we will follow a realistic approach:

* Source Identification: We will identify all potential data sources within the organization, including databases, third-party APIs, and legacy systems.
* Data Extraction Strategy: We will determine data extraction methods and frequency, aligning them with our organization's data needs.
* Transformation Rules: We will establish clear data transformation rules and procedures to maintain data consistency and quality.

➢ **ETL Processes**

Pragmatic ETL processes will be designed and implemented as follows:

• ETL Tool Selection: We will select ETL tools or custom scripting languages based on our organization's resources and expertise.

➢ **Data Exploration**

To promote practical data exploration, we will adopt these strategies:

* User-Centric Design: Our user interface using flask will prioritize user-friendliness, making it easier for data architects to interact with the data warehouse.
* Query Building: We will provide user-friendly query-building capabilities to empower data architects to create custom queries.
* Data Visualization: Implementing data visualization components, such as interactive charts and graphs, will enhance the data exploration experience.
* **Actionable Insights**

Our approach to delivering actionable insights will be grounded in reality:

* Analysis Templates: We will create predefined analysis templates for common use cases to expedite insights generation.

# TOOLS AND TECHNOLOGIES PLANNED TO USE

**Data Modelling**

Creating a data model that represents the structure of our data warehouse by involving tables, relationships, and attributes.

We are going to use the IBM Data Architect tool which can provide assistance in data modeling.

**Setting Up IBM Cloud Db2 Warehouse**

Configuring our database instance, including setting up security, scalability, and performance options.

**Data Loading**

Loading the transformed data into your **IBM Db2** Warehouse instance. We will use IBM DataStage for this purpose.

Ensuring data quality and consistency during the loading process

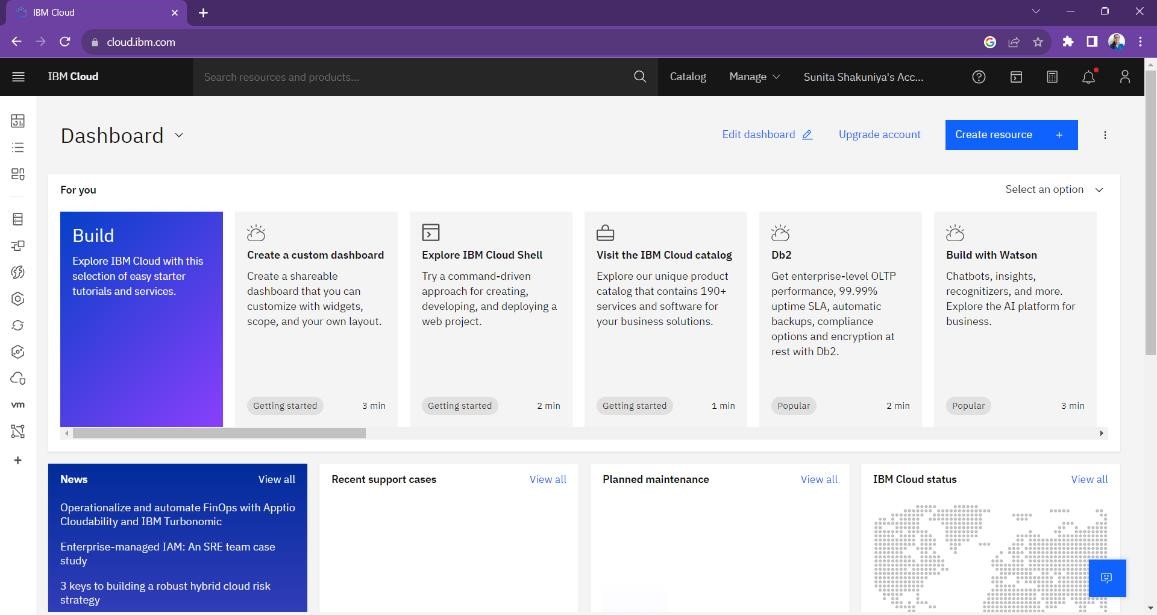
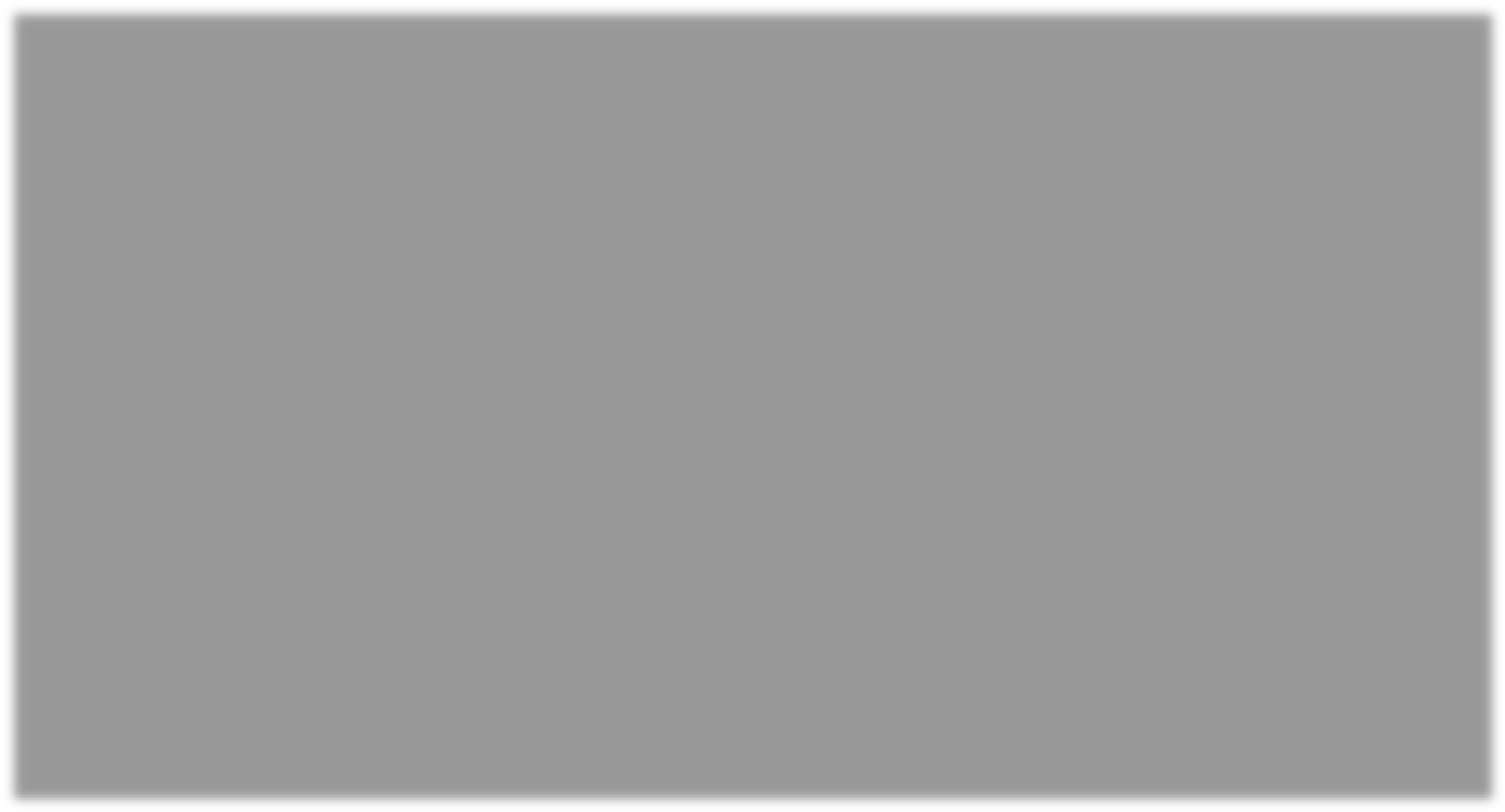
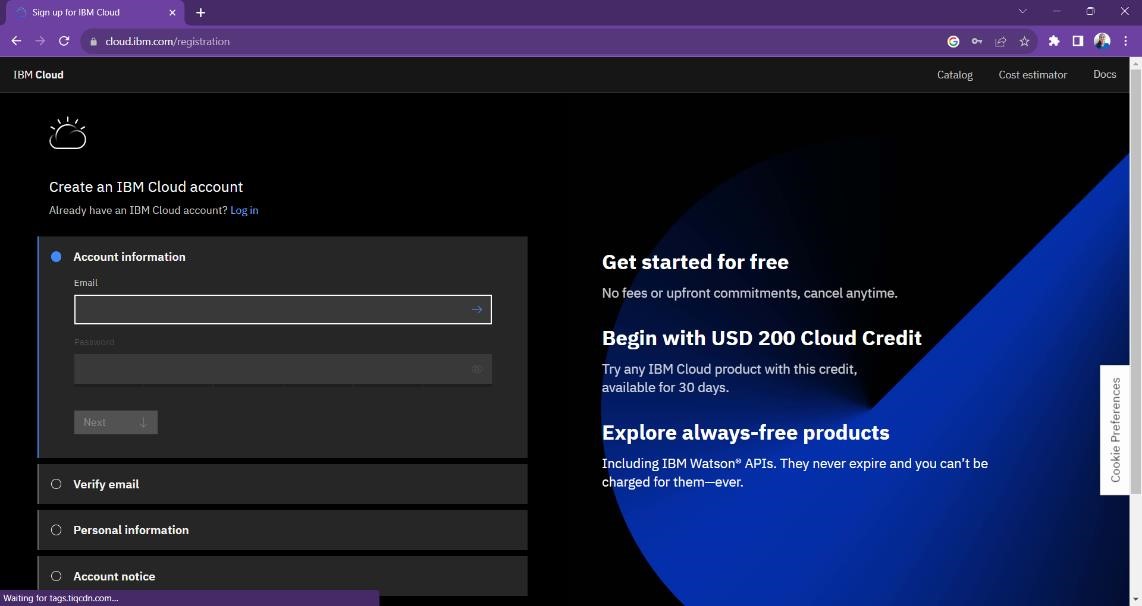
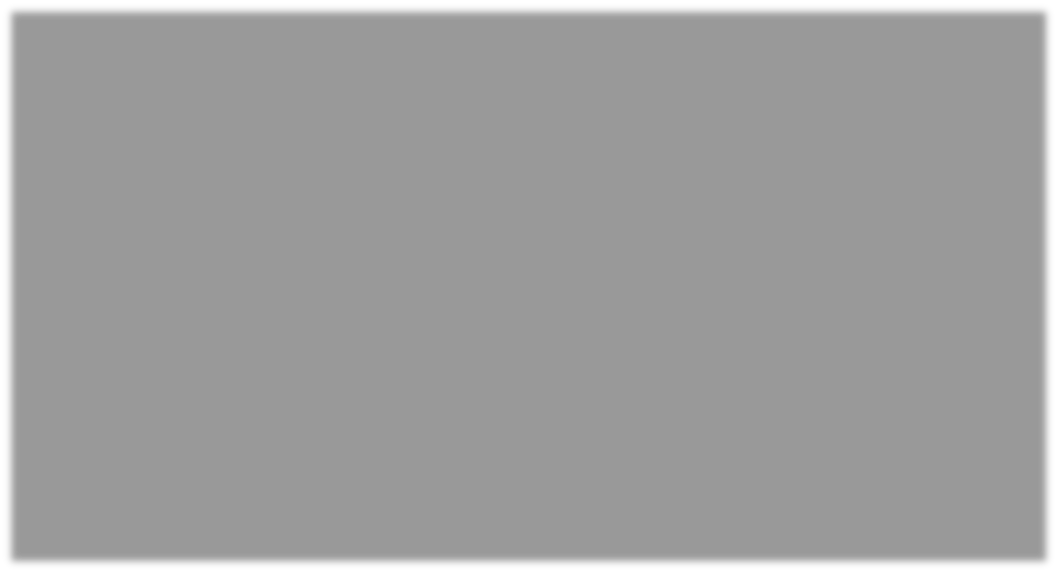
**Data Access and Analysis**

Analysis using **IBM Watson Studio** tool. By building dashboards, reports, and data analysis pipelines to empower your data architects and analysts.

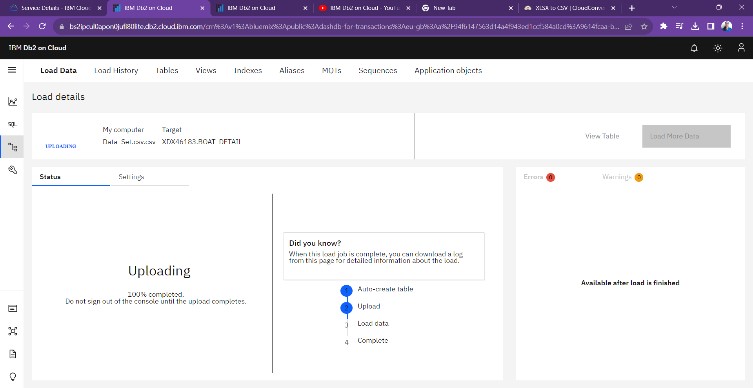
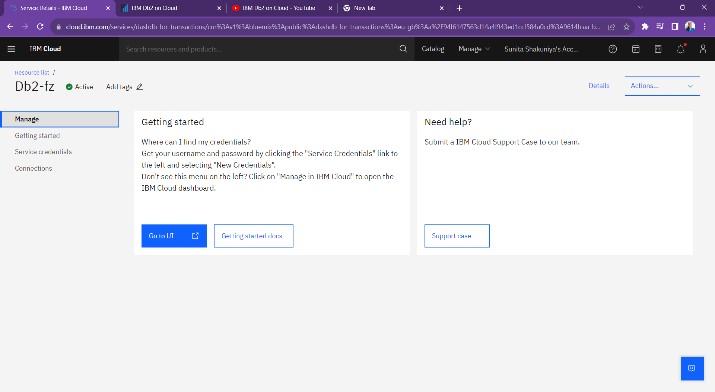
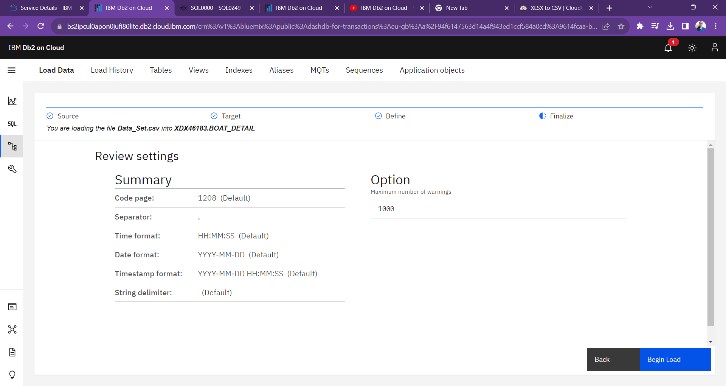
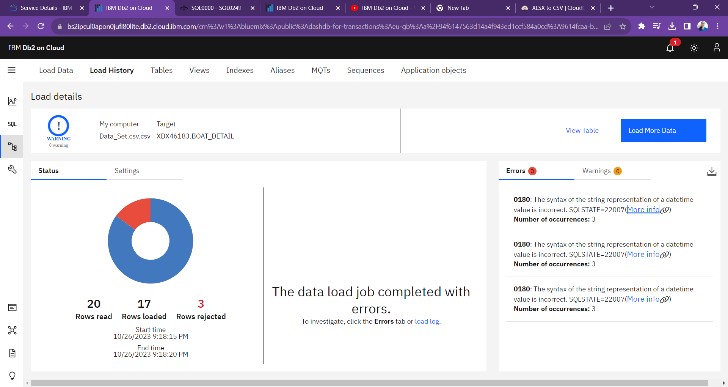
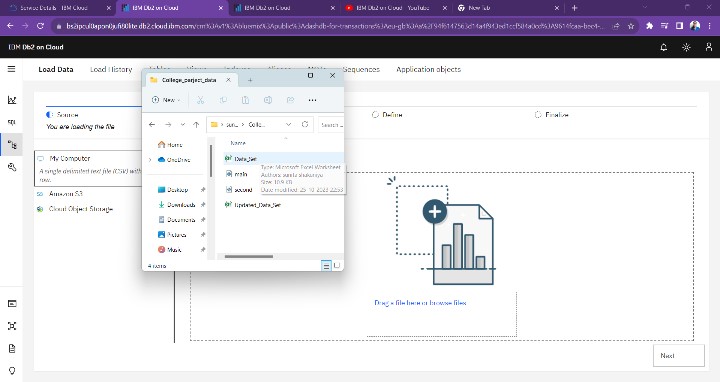
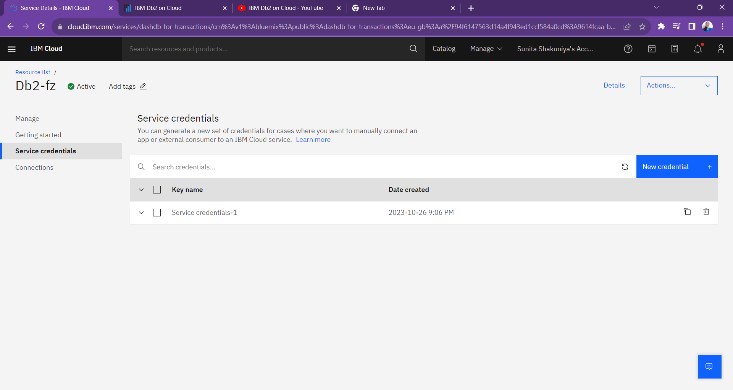
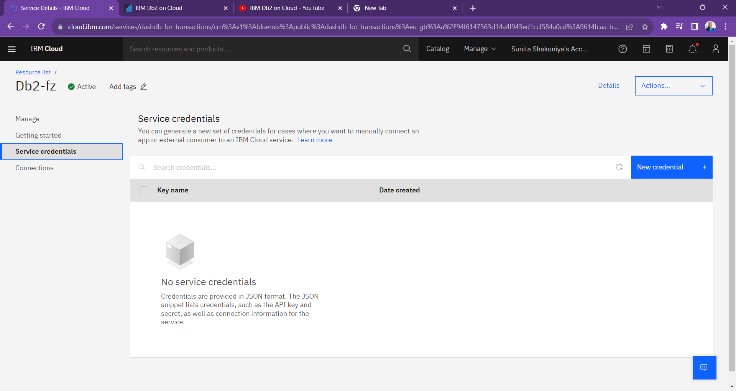
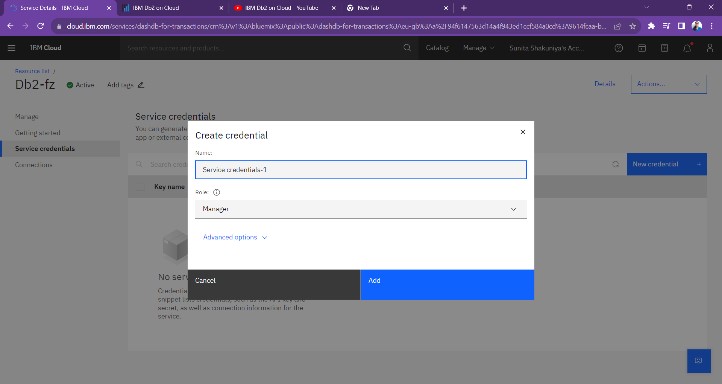
Utilize the capabilities of Db2 Warehouse, such as its in-database analytics and machine learning features, to perform advanced analyses.

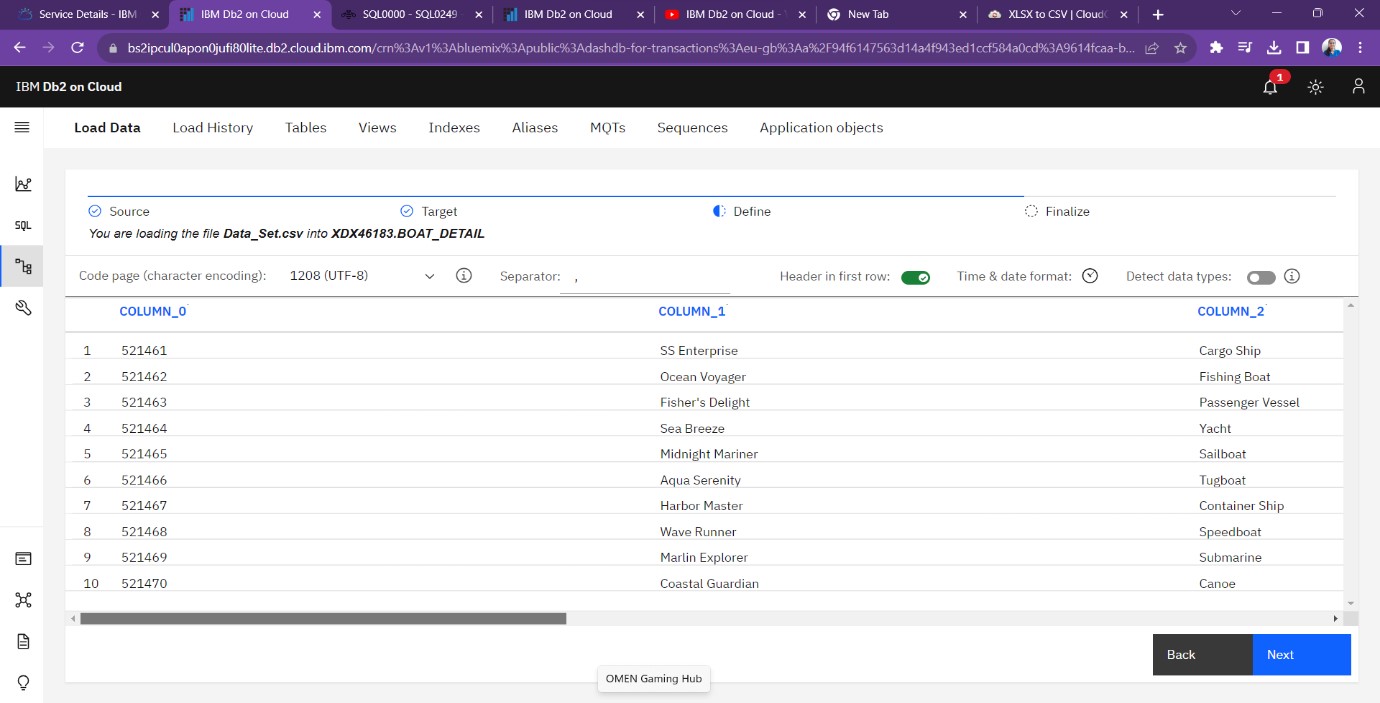
# TECHNICAL IMPLEMENTATION

## ➢ **Step 1:** Creating IBM cloud account

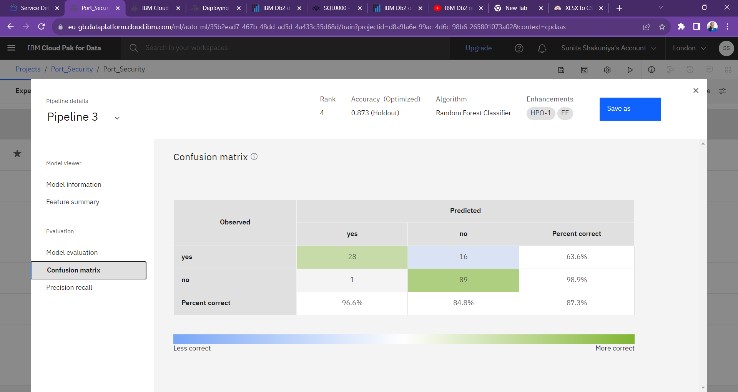
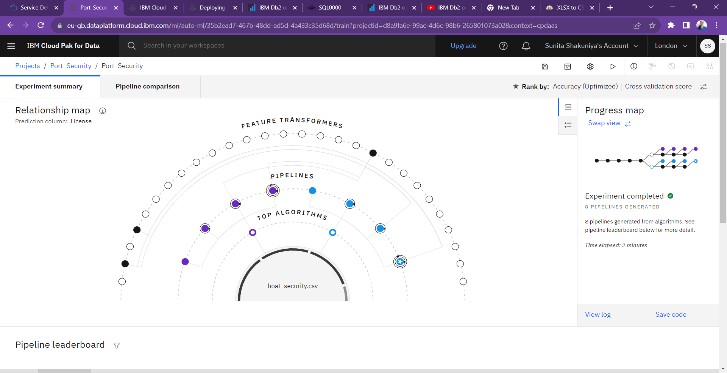
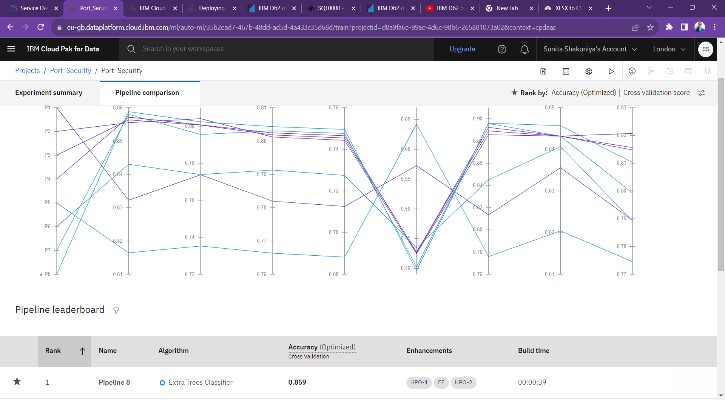
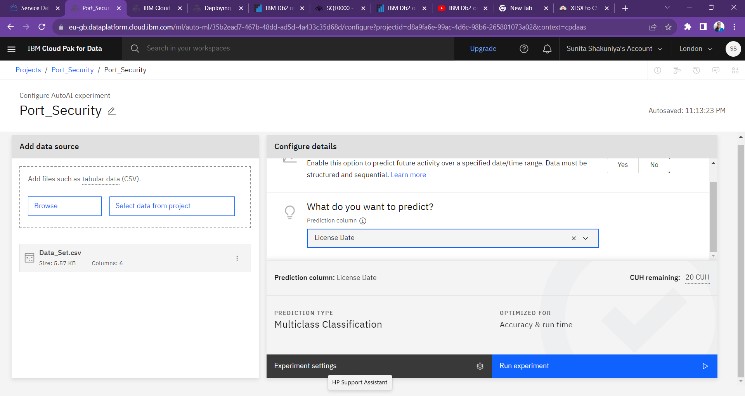
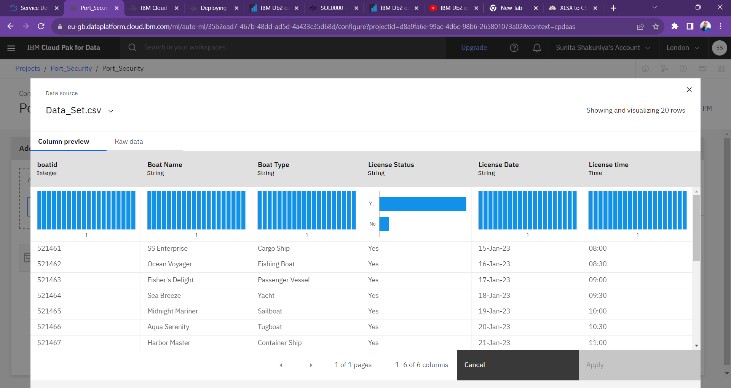
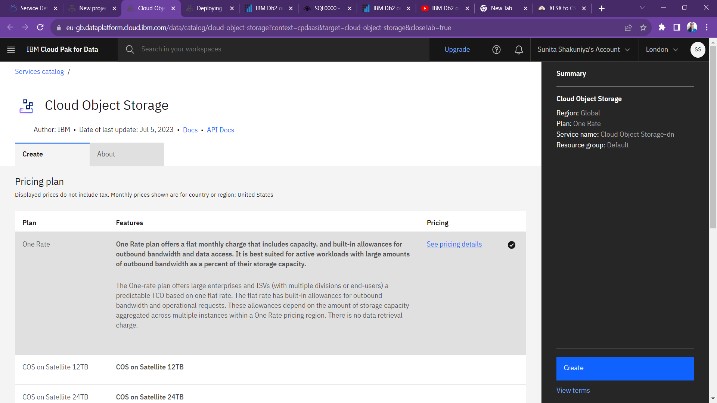
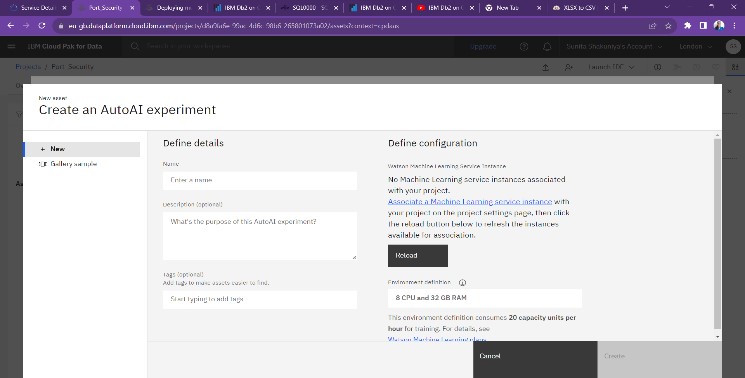
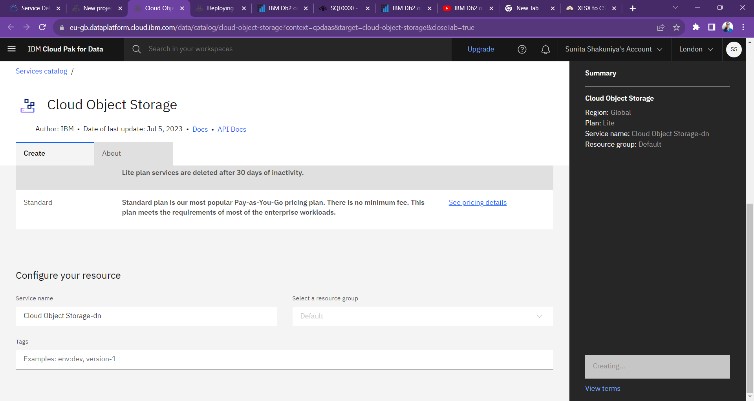
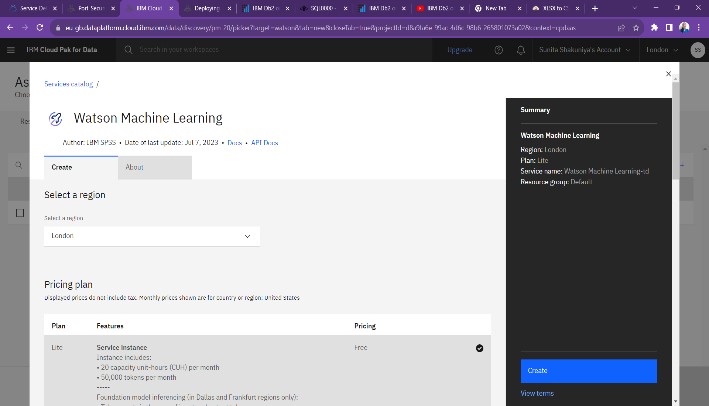
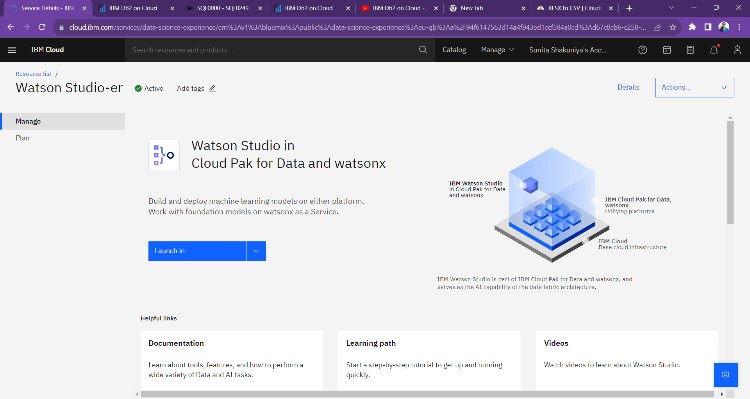


* **Step 2:** Setting up the db2 instance and experimenting.



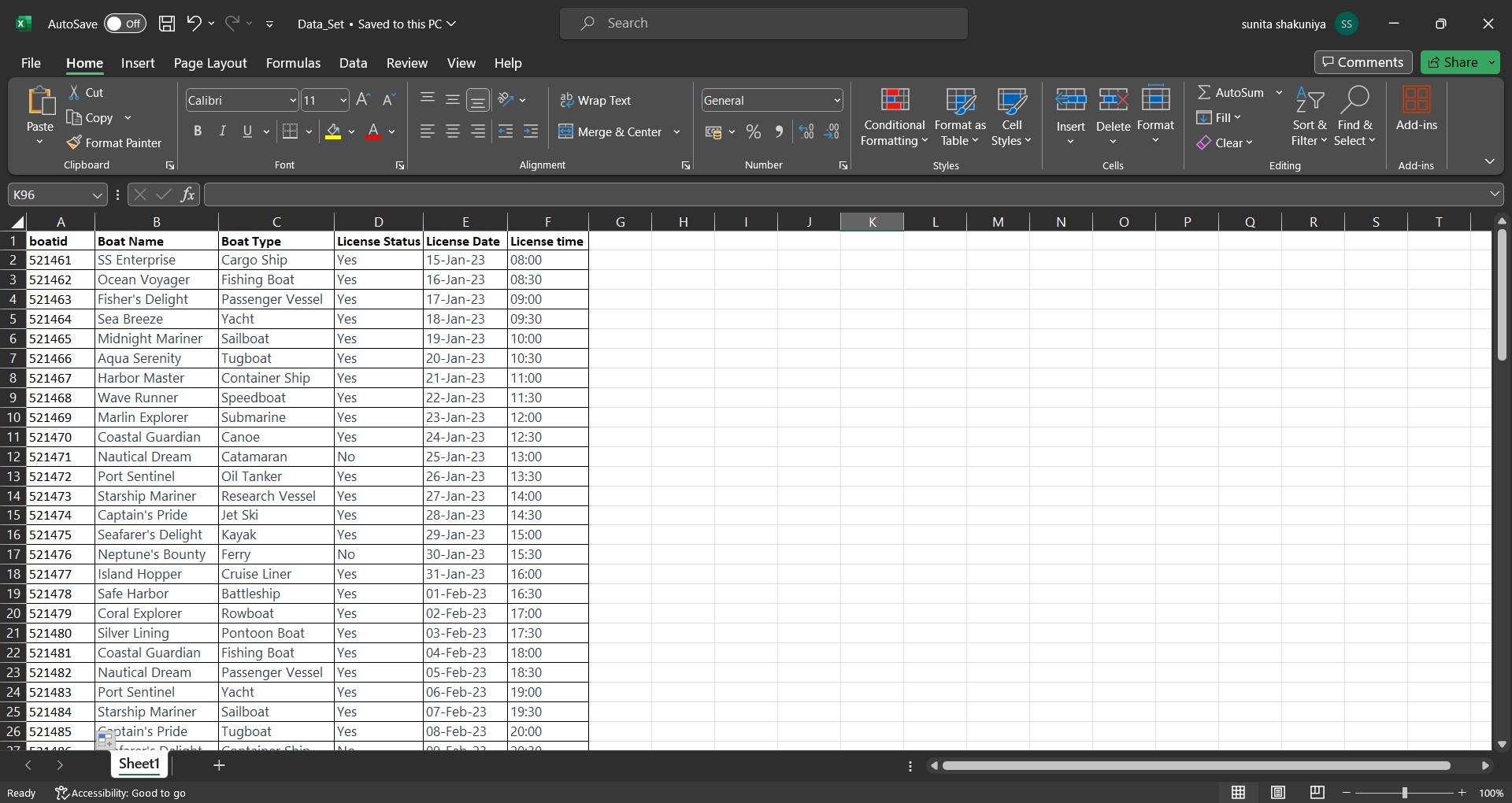


* **Step 3:** Setting up the Watson Studio instance and experimenting.

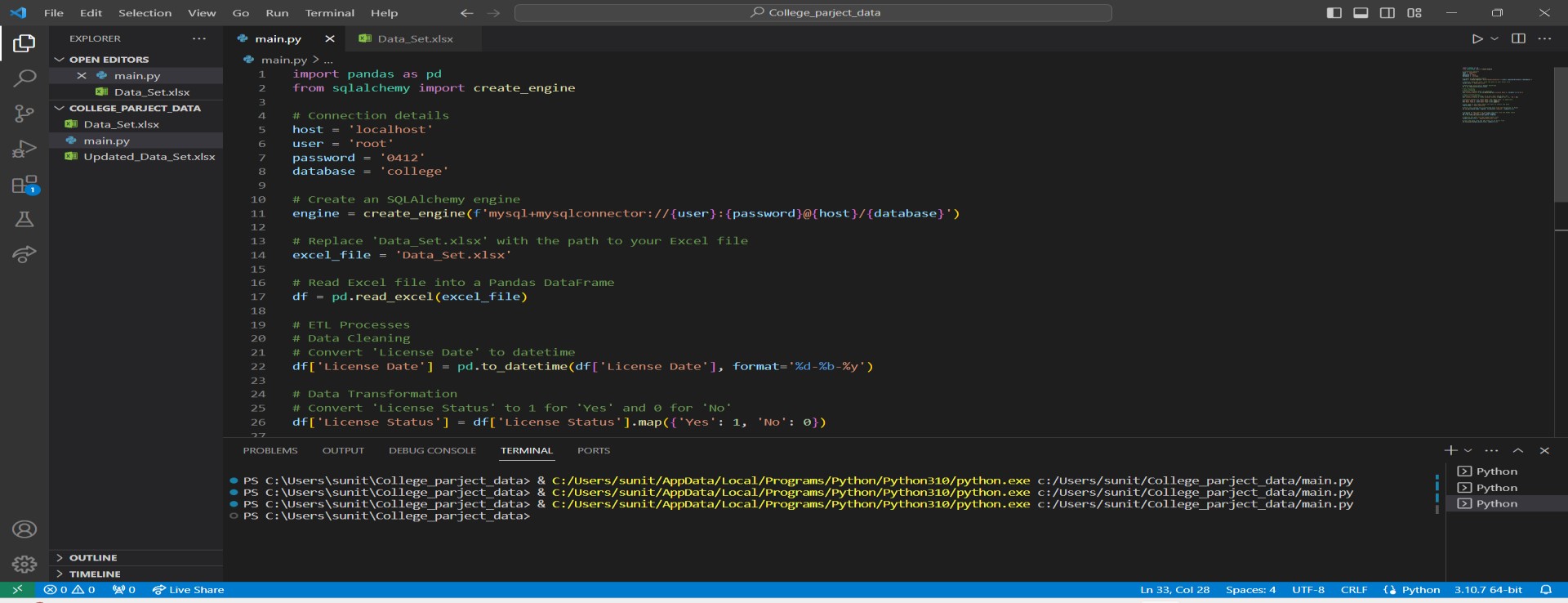


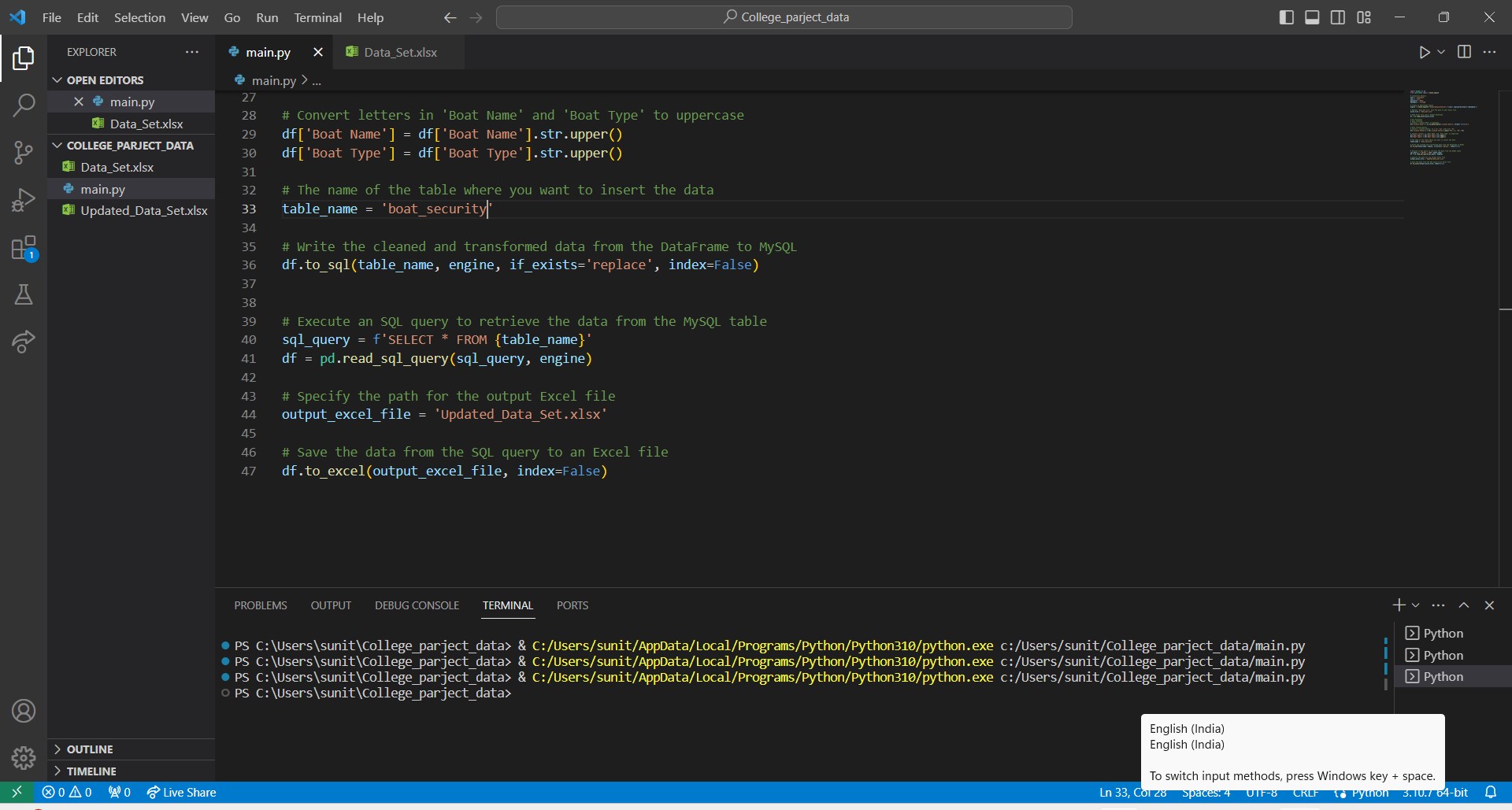
* **Step 4:** As we can perform ETL process over Watson studio so we further proceed with panadas library and MySql data base.

## • Creating Excel data set

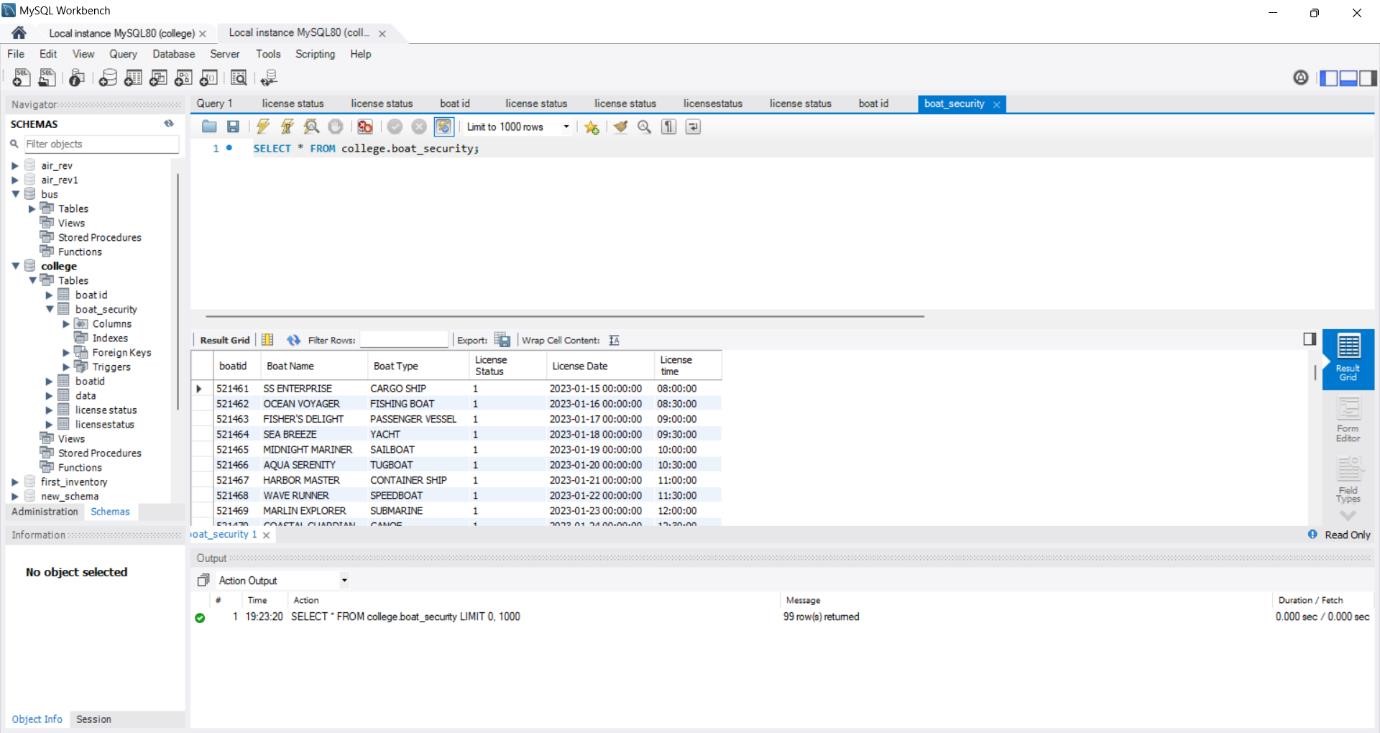


## • Coding implementation over Vscode

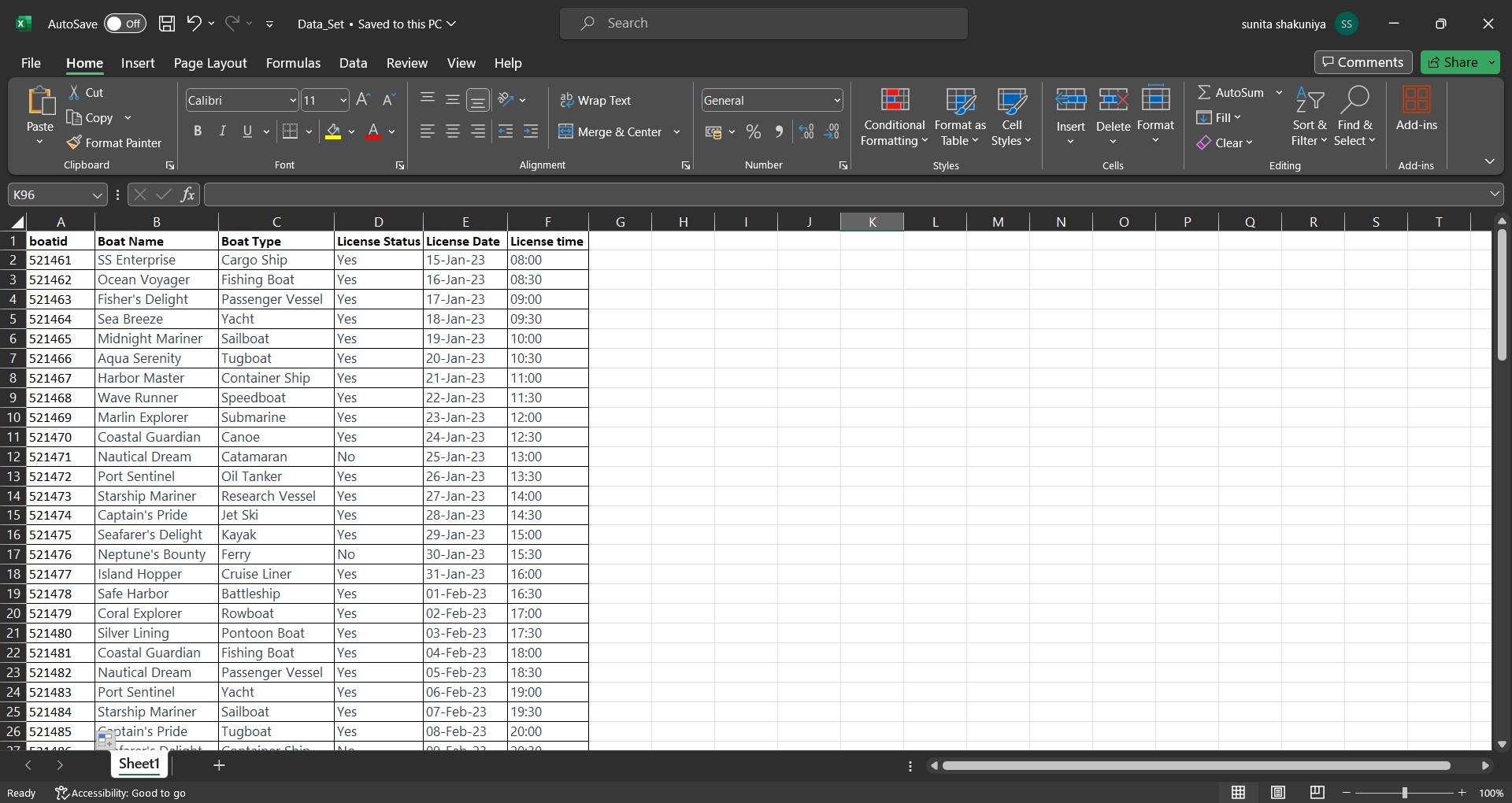




• MySQL Workbench



## • Updated excel sheet



➢ **Step 5**: Conclusion, we finally performed the ETL process using the pandas' library over vscode and Excel data and after running the code for applying ETL queries we got the updated data back in the same format.



1

. DB

2

2

. Watson studio

➢

Final Outcome 3. ETL process

4

. Pandas

5

. MySQL

## **Final Outcome**