Cloud Application Development
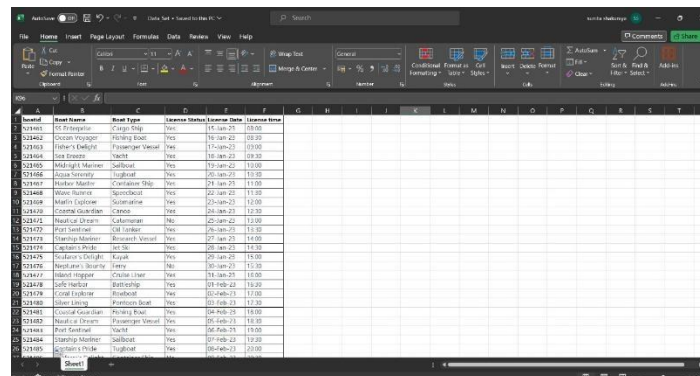
Data Warehousing with IBM Cloud Db2 Warehouse

Phase4

To Do:

Building the data warehouse by implementing ETL processes and enabling data exploration. Implement ETL processes to extract, transform, and load data into the data warehouse. Enable data architects to explore and analyse data within Db2 Warehouse using SQL queries and analysis techniques.

Implementation:

Figure 1 DATA CREATION

Figure 2 OPEN VISUAL STUDIO
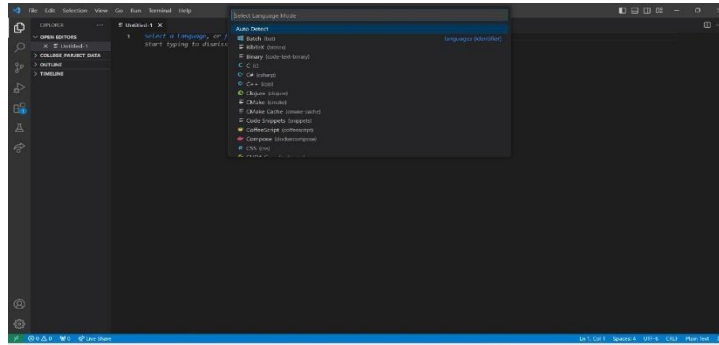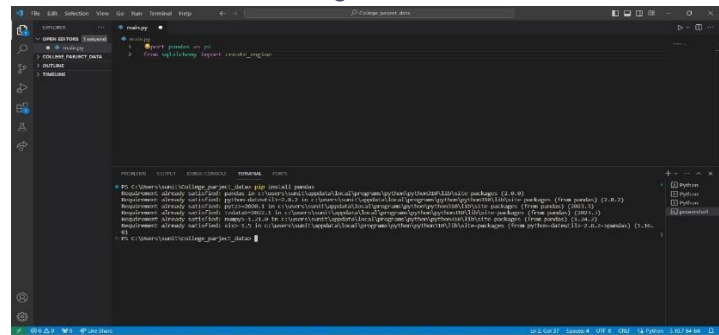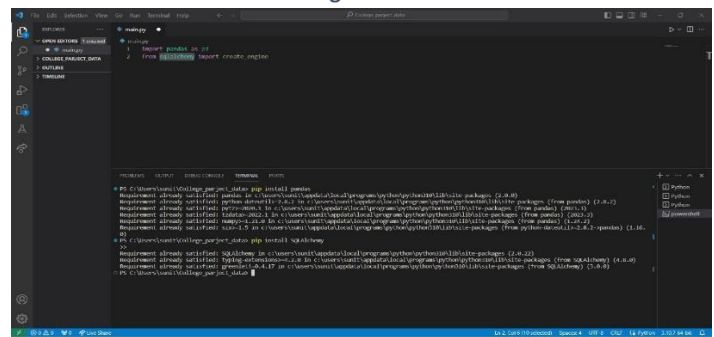
Figure 2.1
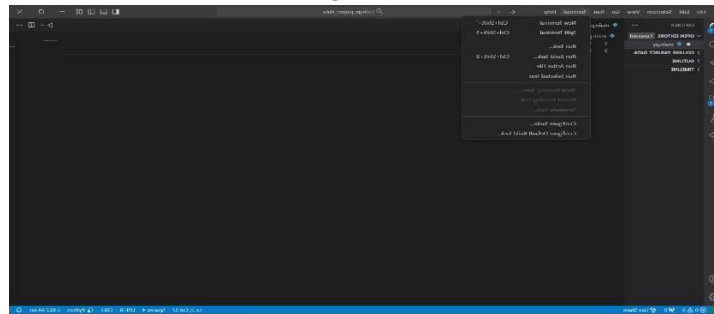
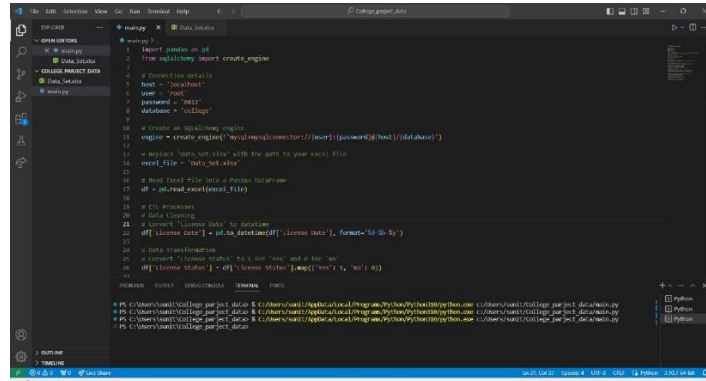**Figure 2.2**

Figure 2.3

Figure 2.4

Figure 2.5 PANDAS CODING
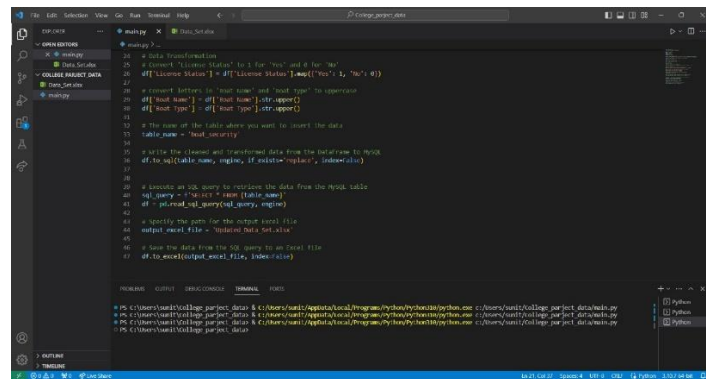
Figure 2.6 PANDAS CODING END
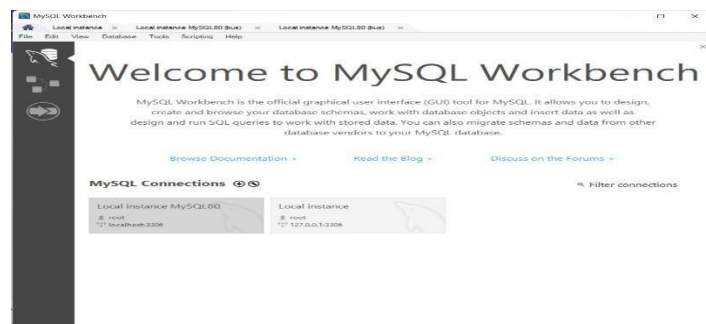
Figure 3 OPEN MYSQL
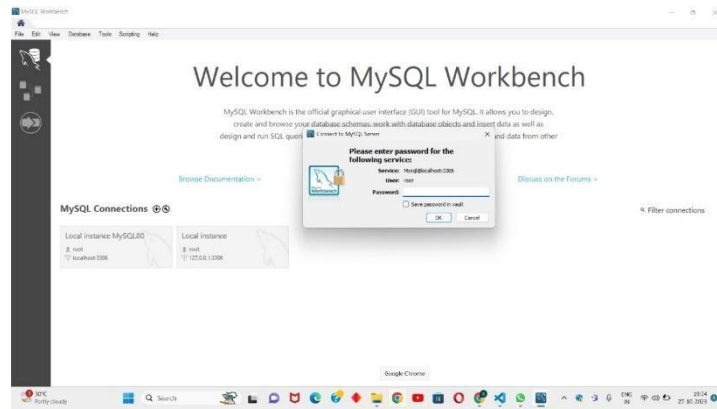
Figure 3.1
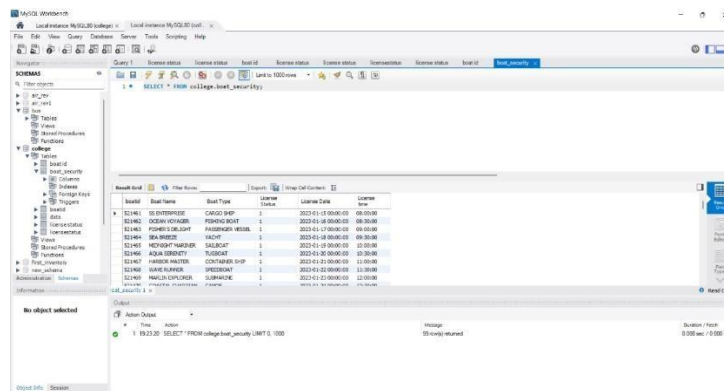
Figure 3.2 AFTER RUNNING THE CODE IN MYSQL SERVER
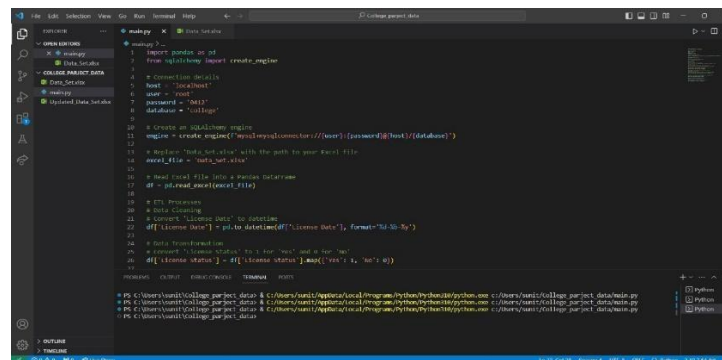
Figure 3.3 UPDATED DATA FILE GENERATED (SEE LEFT SIDE)

Figure 3.4 ENDING

Figure 4 UPDATED DATA AFTER RUNNING ETL PROCESS OVERR THA DATA USING PANDAS

Project Goals:

Building the Data Warehouse:

The primary goal was to create a data warehouse infrastructure using IBM Db2 Warehouse.

Implementing ETL Processes:

We aimed to establish efficient ETL processes to extract, transform, and load data into the data warehouse.

Eg :

Extract data from a source (e.g., CSV file)

INSERT INTO TargetTable (Column1, Column2, Column3)

SELECT SourceColumn1, SourceColumn2, SourceColumn3

FROM SourceCSV;

Enabling Data Exploration:

The project aimed to provide data architects with the tools and capabilities to explore and analyze data within Db2 Warehouse using SQL queries and analysis techniques.

Basic SQL Query:

Retrieve data from a table

SELECT Column1, Column2

FROM WarehouseTable

WHERE Condition = 'Value';

Joining Tables:

Join multiple tables for more complex queries

SELECT W.ColumnA, T.ColumnX

FROM WarehouseTable W

INNER JOIN AnotherTable T ON W.ID = T.ID;

Aggregation and Analysis:

Perform aggregate functions for analysis

SELECT Year, SUM(Sales) AS TotalSales

FROM SalesData

GROUP BY Year;

Project Milestones and Achievements

1. Data Warehouse Implementation

Successfully deployed IBM Db2 Warehouse, providing a scalable platform for data storage and management.

2. ETL Process Implementation

Designed and implemented ETL processes that automate data extraction from various sources, perform necessary transformations, and load data into the warehouse.

Achieved data integration across different systems, ensuring a unified and consistent data source.

3. Enabling Data Exploration

Provided data architects with access to Db2 Warehouse, including necessary permissions and credentials.

Facilitated the use of SQL queries and data analysis techniques, empowering data architects to explore the data effectively.

Conclusion

In conclusion, our hands-on experience with IBM cloud, DB2 and Watson Studio has been both enlightening and informative. These platforms have offered us a valuable glimpse into the world of cloud computing and data science. However, it is worth noting that we encountered certain limitation in performing ETL (Extract, Transform, Load) processes over cloud services.

However, it has become evident that proceeding with ETL (Extract, Transform, Load) processes over cloud services presents certain limitations and challenges. As a results, we have made the strategic decision to transition towards using Pandas and MySQL for our ETL processes. This shift allows us to work with greater flexibility and control, ensuring seamless data transformations and management. By combining the capabilities of Pandas and MySQL, we are better equipped to tackle the intricacies of data processing and analysis, setting the stage for future development and data-driven projects. This decision marks a pivotal moment in out journey, as we adapt and evolve in response to the ever-changing landscape of data science and technology.