**NAME : WANGU MAPOSA**

**STUDENT ID : BIDA23-043**

**EMAIL;** [**BIDA23-043@THUTO.BAC.AC.BW**](mailto:BIDA23-043@THUTO.BAC.AC.BW)

**ASSIGNMENT TITTLE : BI205-Data Warehousing assignment**

**COURSE: BUSINESS INTELLIGENCE AND DATA ANALYSIS**

**YEAR : YEAR 2 SEMESTER 2**

**COHORT : 2022-2027**

**DATE OF SUBMISSION ; 05/25/2025**

**TASK 1: REPORT**

**COMPANY**: First national bank (FNB)

**INDUSTRY**: financial services industry

FNB deals with a lot of transactions which include withdrawals, deposits, and card payments daily. Customer data, transaction data, marketing data as well as account data are collected from disparate sources like mobile banking and ATMs are stored for auditing, analysis, and reporting.

* **Customer Data:** Collected through account opening forms (online and in-branch), KYC (Know Your Customer) processes, customer service interactions, and digital banking profiles.
* **Transaction Data:** Generated from ATM withdrawals and deposits, point-of-sale (POS) systems, online payments, mobile banking transfers, and card swipes.
* **Marketing Data:** Gathered from website analytics, social media interactions, email campaign responses, customer surveys, and promotional offer redemptions.
* **Account Data:** Maintained through core banking systems, including balances, account activity, loan applications, and investment records.

Given this vast amount of data FNB needs a Datawarehouse to centralize data management ensuring that everyone in the company gets access to consistent, accurate and unified data. A Datawarehouse also allows for the analysis of historical data to discover trends and mitigate risks.

**INTRODUCTION**

In the modern business landscape, specifically the financial service industry, where data plays a crucial role in businesses, data-warehouses and data-marts are essential. A Data-warehouse is a centralized repository where all the time variant(historical ), subject oriented(e.g. sales, marketing) and integrated data from different sources is stored ,It serves as a storage system for enterprise-wide data ensuring that all departments use accurate and consistent data whereas in contrast a Data-mart is a subset of a data-warehouse which stores data for specific departments within the business catering to its unique needs.

Data-warehouses and Data-marts play a crucial role in business Intelligence by improving decision -making and strategic planning through accurate and consistent data. Additionally, they enhance data accessibility and support analytical processing, enabling organizations to gain valuable insights and gain a competitive edge.

This report aims to give a full explanation of the difference between a data warehouse and a data warehouse, to recommend the data warehouse for FNB and to ensure that the FNB stakeholders understand more about their need for the data warehouse than the data mart.

**COMPARISON OF THE DATAWAREHOUSE AND THE DATAMART**

**Size and Scope**

A data warehouse is large in scale as it manages vast amounts of data meant to serve the entire enterprise. For example, First National Bank (FNB) uses a data warehouse to centralize and analyse data from all departments such as sales, inventory, and marketing (Kimball and Ross 2013). Whereas a data mart is smaller and designed for specific departmental needs. The marketing department at FNB might use a data mart to store and analyse customer data to identify trends and develop targeted campaigns, without requiring enterprise-wide data (Inmon 2005).

**Data Integration**

Data warehouses integrate data from multiple sources across an organization, ensuring consistency and comprehensive reporting. FNB synthesizes transaction, customer, marketing, and account data in its warehouse to maintain a unified operational view (IBM 2022). Conversely, a data mart typically sources its data from the central data warehouse to serve specific departmental needs. For instance, FNB's Risk and Compliance department might use a data mart to monitor accounts, transaction patterns, and repayment histories for fraud detection and risk mitigation (Microsoft Azure 2023).

**Complexit**y

Data warehouses are complex to implement as they must manage large volumes of data from disparate sources while supporting complex analytical queries, requiring specialized infrastructure and IT expertise. FNB utilizes a sophisticated data warehouse (e.g., Snowflake) to process customer data for personalized recommendations (Snowflake 2023). Data marts, being department-specific, are simpler to implement and manage. A bank's finance department could deploy a data mart to track loan payments with relative ease compared to a full warehouse implementation (Kimball and Ross 2013).

**USE CASE**

A data warehouse supports reporting ,analysis and data mining for data in the warehouse for the whole insights ,to get insights that will help improve all the business processes for improves customer service and increase in profits. On the other hand, the data mart supports reporting and analysing for data for a specific department ,such as the marketing department to help improve the processes of that specific departments.

**Cost and Maintenance**

Implementing and maintaining a data warehouse requires significant investment in infrastructure, storage, and processing power. Financial institutions like FNB incur substantial costs for their warehouses to support compliance reporting and risk assessment (Inmon 2005). Data marts are more cost-effective due to their smaller scale and focused functionality. A risk and compliance data mart would be less expensive to deploy and maintain than an enterprise warehouse (Microsoft Azure 2023).

**Conclusion**

This report provides a comparison between a data warehouse and a Datamart ,though these two serve different purposes they compliment each other. For a larger company that deals with vast amounts of data such as FNB ,I highly recommend the adoption of a Data warehouse because even though it may be expensive to implement and maintain ,it provides a single source of truth to the entire company ensuring that the same ,accurate and consistent data is used to bring reliable insights for better decision making by stakeholders.

References

* IBM (2022) *Data Warehousing: Concepts and Architectures*. Available at: <https://www.ibm.com/cloud/learn/data-warehouse> (Accessed: 15 June 2024).
* Inmon, W.H. (2005) *Building the Data Warehouse*. 4th edn. Hoboken, NJ: Wiley.
* Kimball, R. and Ross, M. (2013) *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modelling*. 3rd edn. Indianapolis: Wiley.
* Microsoft Azure (2023) *Data Warehouse vs. Data Mart: Key Differences*. Available at: [https://azure.microsoft.com](https://azure.microsoft.com/) (Accessed: 15 June 2024).
* Snowflake (2023) *Modern Data Warehousing: Best Practices*. Available at: [https://www.snowflake.com](https://www.snowflake.com/) (Accessed: 15 June 2024).

**STAR SCHEMA DESIGN**

FNB must analyse the transaction data of customers across different branches and periods of time. The goal is to get a clear understanding of customer behaviours, branch performance, track any fraud activities ad well as to find the peak transaction times. By having different tables that contain important data about customers ,transactions and the FNB branches ,The star schema provides a straightforward structure that allows for easier understanding as well as cross referencing to better understand the relationships of all the tables. The following is the fact table along with its dimensions for FNB.

**FACT TABLE: FACT\_TRANSACTIONS TABLE**

|  |  |  |
| --- | --- | --- |
| Transaction\_ID | PK | Unique identifier of each transaction |
| Transaction type |  | Withdrawal, deposit etc. |
| Account\_ID | FK | Linked to the account dimension |
| Customer\_ID(FK) | FK | Linked to the customer dimension |
| Branch\_ID(FK) | FK | Linked to the branch dimension |
| Date (FK) | FK | Linked to the date dimension |
| Amount |  | Amount of money for the transaction |
| Transaction\_status |  | If the transaction failed or succeeded |

**DIMENSIONS TABLES:**

**CUSTOMER\_DIM TABLE**

|  |
| --- |
| CUSTOMER\_ID(PK) |
| FIRST\_NAME |
| LAST\_NAME |
| AGE |
| GENDER |
| PHONE\_NUMBER |

**ACCOUNT\_DIM TABLE**

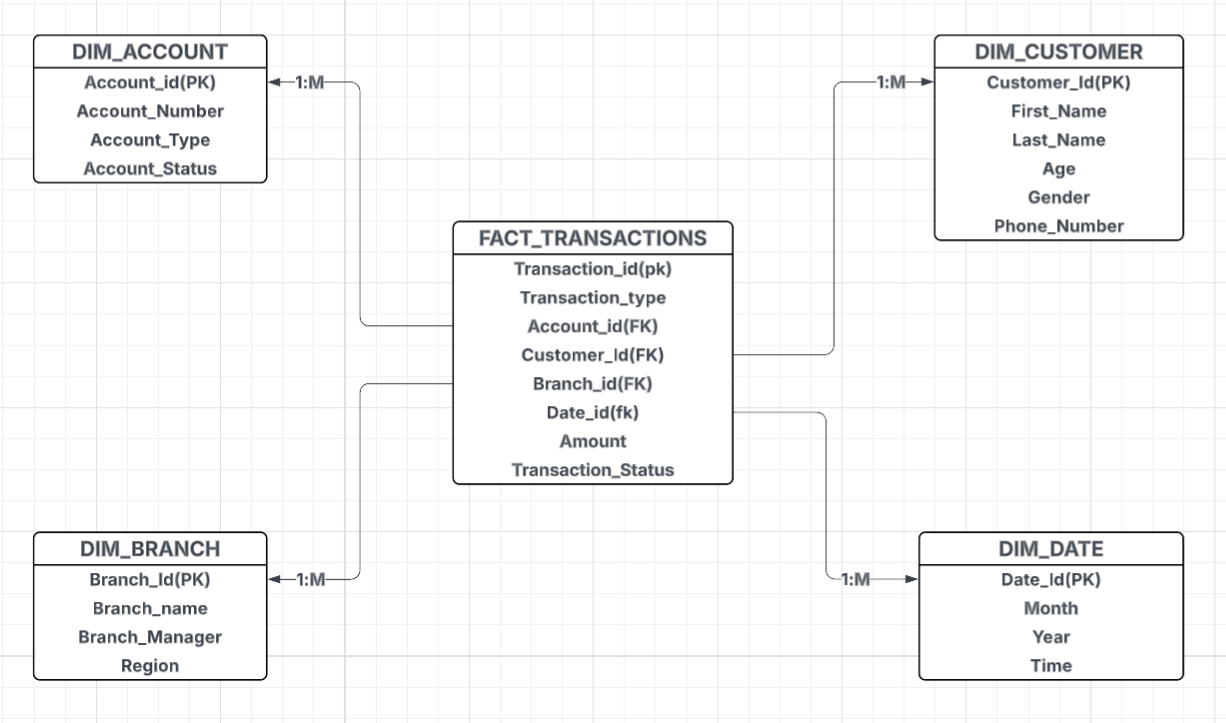
|  |
| --- |
| ACCOUNT\_ID (PK) |
| ACCOUNT\_NUMBER |
| ACCOUNT\_TYPE |
| ACCOUNT\_STATUS |

**BRANCH\_DIM TABLE**

|  |
| --- |
| BRANCH\_ID(PK) |
| BRANCH\_NAME |
| BRANCH\_MANAGER |
| REGION |

**DATE\_DIM TABLE**

|  |
| --- |
| DATE\_ID(PK) |
| MONTH |
| YEAR |
| TIME |



(The STAR SCHEMA DIAGRAM OF THE FNB FACT TABLE AND ITS DIMENSIONS)

**HOW THE STAR SCHEMA SUPPORTS THE BUSINESS ANALYSIS FOR FNB**

This schema allows FNB to:

**Analyse customer behaviours**: With the transaction data and customer data FNB can identify high value segments(those with high transaction amounts) as well as at-risk groups (those that have a declining activity), this information will then help stakeholders to come up with customer retention strategies for the at-risk groups and enable target marketing to those who are often perform transactions.

**Risk management and fraud detection:** The star schema structures the transactional data to be able to expose suspicious, anomaly and compliance risks and behaviours. The fact table (transactions) records all the financial events which makes it easier to identify anomaly behaviours, and by linking this table to the dimension tables such as the account and customer tables FNB can flag suspicious accounts and identify considerable risk customers.

**Multidimensional analysis:** This schema enables analysis across different business perspectives; customer analysis can be achieved by tracking behaviours because the customer dimension table is linked to the fact table. Branch performance, time-based trends as well as account insights can be analysed because they are linked to the fact table.

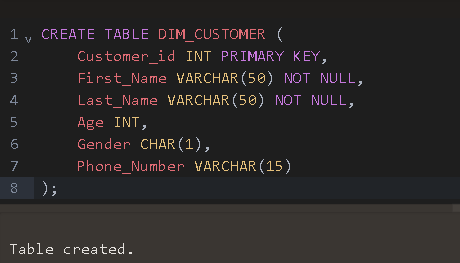
**Support historical analysis:** Through the date dimension which stores the transaction data from different periods of time, FNB will be able to track trends ,predict future outcomes of the company, and produce strategies that will improve the business processes.

**SQL DATA DEFINITIONS**

The following will be a demonstration of how all the tables are created using My SQL ,along with the output showing that the table has been successfully created.

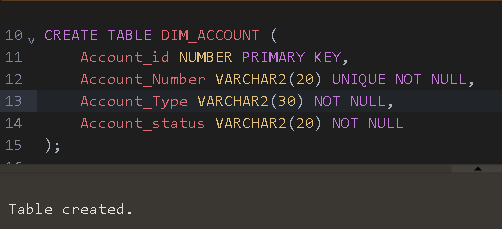
THE CUSTOMER DIMENTION TABLE

The diagram shoes the SQL statement for the creation of the table along with the output.



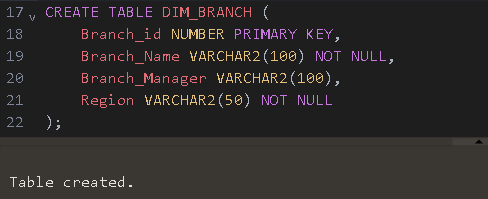
THE ACCOUNT DIMENSION TABLE

The diagram shoes the SQL statement for the creation of the Account\_Dim table along with the output.



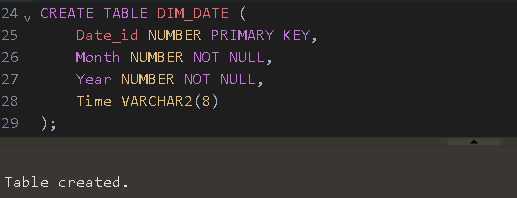
THE BRANCH DIMENSION TABLE

The diagram shoes the SQL statement for the creation of the Branch\_Dim table along with the output.



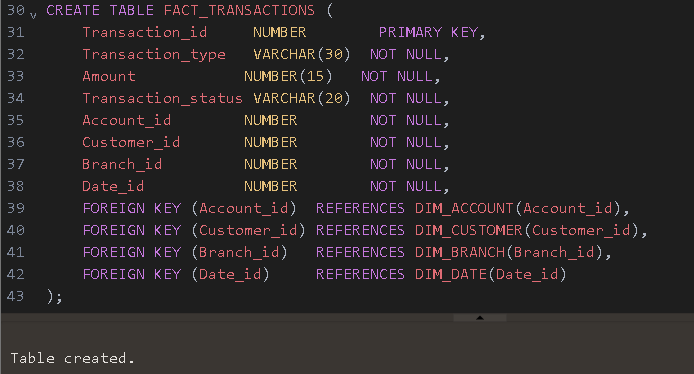
THE DATE DIMENSION TABLE

The diagram shoes the SQL statement for the creation of the Date\_Dim table along with the output.



THE FACT\_TRANSACTIONS TABLE

The diagram shoes the SQL statement for the creation of the FACT\_TRANSACTIONS table along with the output.



**BUSINESS INTELLIGENCE QUERIES**

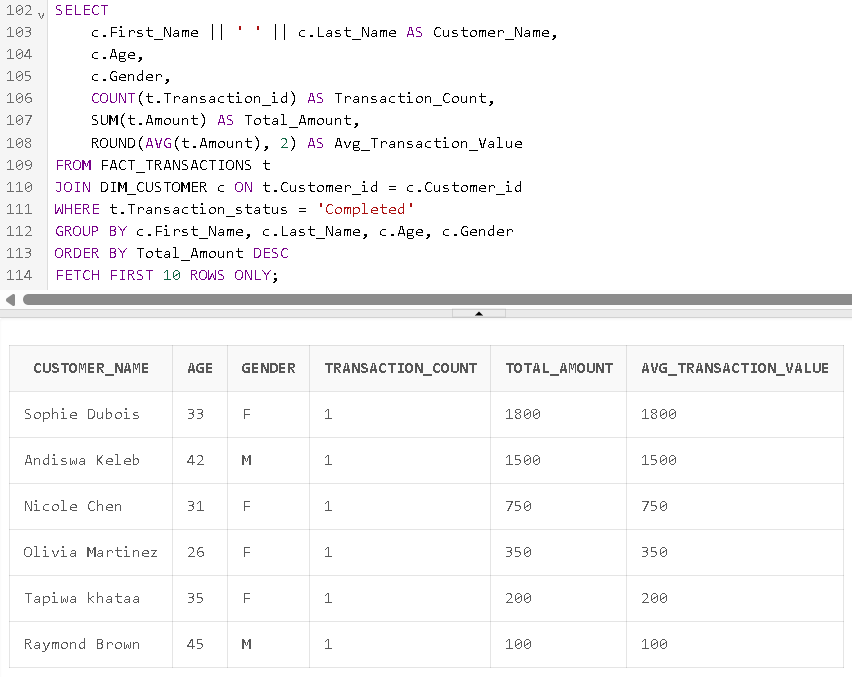
The following are the business intelligence queries that the star schema previously shown will support to help improve productivity at FNB, each query is accompanied by an explanation and questions that will be answered by the query.

**1.Analysis of Customer Transactions**

The query will help FNB to answer:

* Which customers are the Top customers in terms of transactions volumes and value?
* how can we improve the transaction processes to keep these business customers?

This query helps to analyse all customer transaction activities and identify high value customers to enable target marketing and customer retention strategies for customers who might churn.

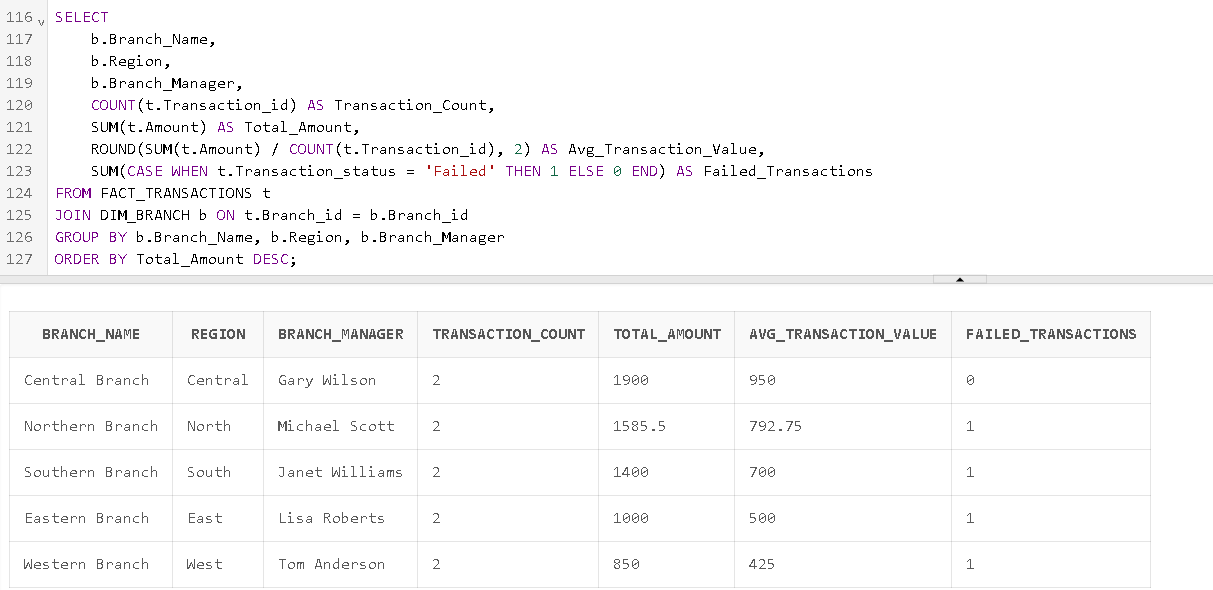


**2. Branch performance**

This query will provide answers for the following questions :

* How are all the FNB branches performing?
* Which branch is the best and which one has the poorest performance?

The branch performance query helps analyse the performance of the branches and simultaneously identify the best performing branch along with the least performing branch, based off the number of failed transactions of a branch and the average transaction value.

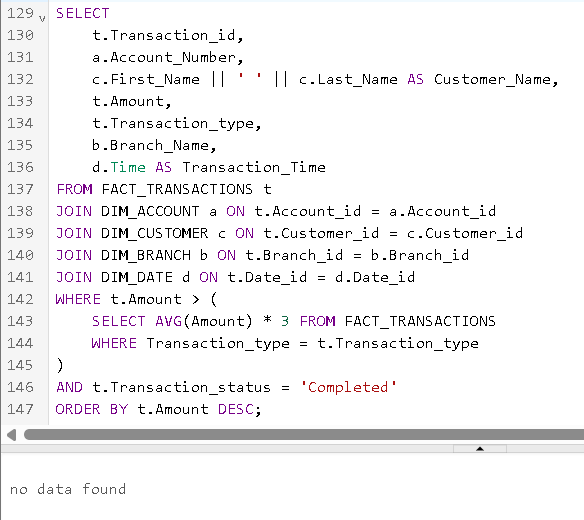


**3. Anomaly activity monitoring**

FNB will use this query to find :

* Which account transactions exceeds the normal thresholds of the average transaction value?

FNB will be able to use this query to identify potential fraud cases by identifying transactions that are high that may require investigation.

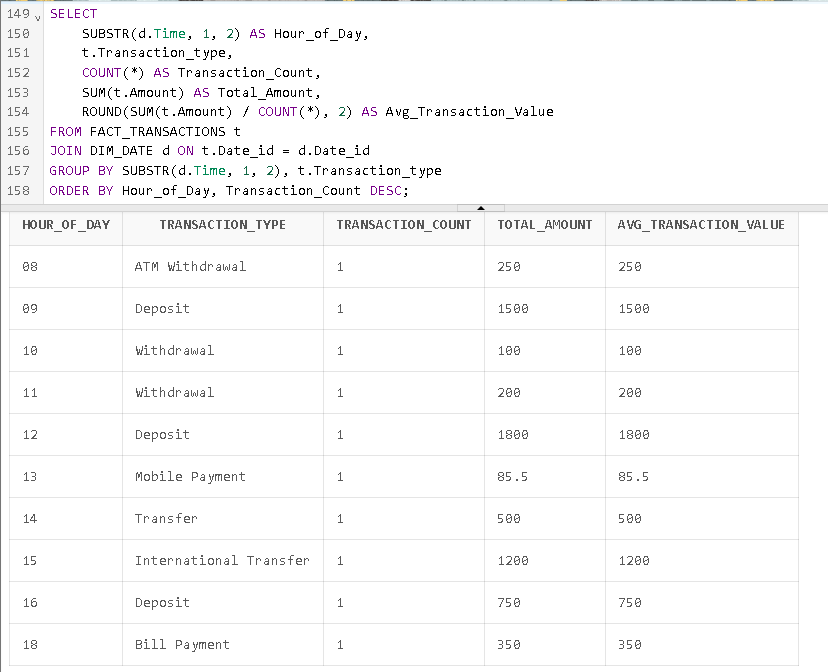


**4. Transactions Time-based analysis**

This query will provide answers to the following questions **:**

* What is the peak time for transactions?
* What type of transactions are performed during the peak transaction periods?

This query will help stakeholders to best allocate staff and resources to align with the identified peak transaction patterns to improve service quality.



**5. Account Type usage patterns**

This query will help answer the following questions **:**

* Which account type generates the most revenue?
* how do the account types perform in terms of usage

This query will give insights that will help develop strategies to optimise the different account types so that FNB can get generate more revenue from all the account types.

A screenshot of a computer

AI-generated content may be incorrect.

**6. service optimization analysis**

With this query FNB can answer

* When should more staff be allocated to manage the peak transaction times?

This query will help reduce the customer waiting times as well as find ways to improve the services by aligning staff with the identified peak times.

A screenshot of a computer

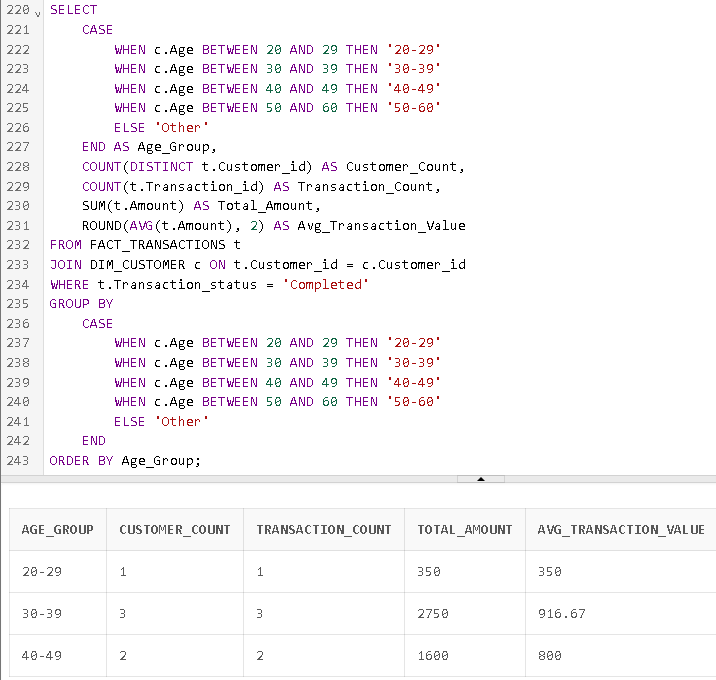
AI-generated content may be incorrect.

**7. Customer demographic analysis**

This query will help answer :

* Which age group has high value customers?
* Which age group has at risk customers?
* What type of accounts and transactions do all these age groups usually perform
* How can we better our services to cater for the different age groups?

This query will help enable targeted marketing based on the demographic categories.



**8. failed transactions analysis**

This query will help analyse and find:

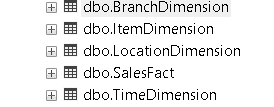
* What is the peak time of the failed transactions?
* What type of transactions usually fails?
* How can we fix this issue?

This query will help analyse and identify the weaknesses of the bank as well as ways to improve the services and processes of the transactions.



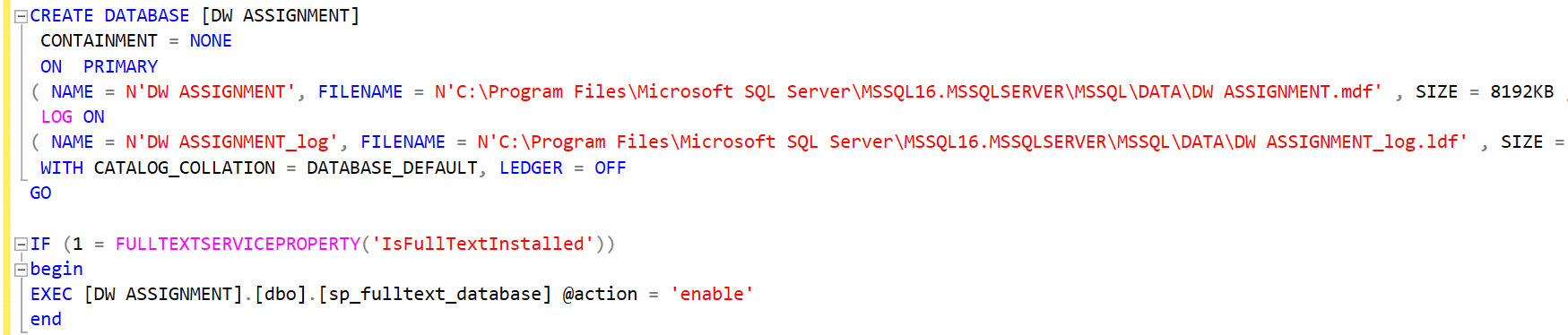
**TASK 2**

SQL server management studio was used to create a database with its objects . A total of five tables were created ,which includes the fact table along with its dimensional tables.



THE DATABASE

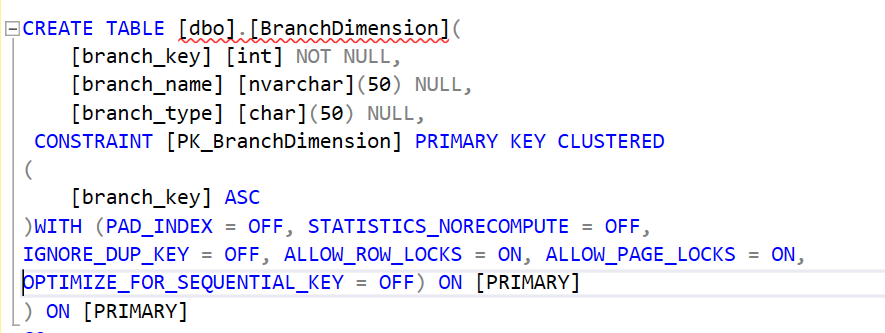
The following shows a screenshot of the code used to create the Database.

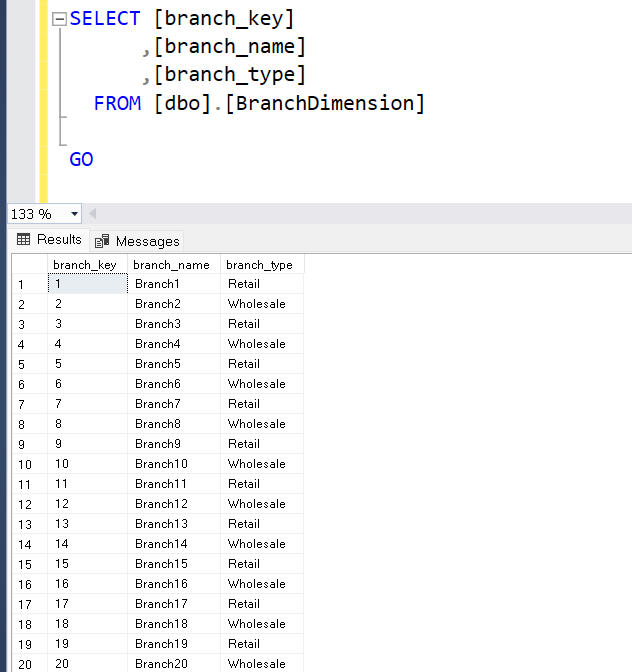


(fig 1.1)

**Branch dimension**

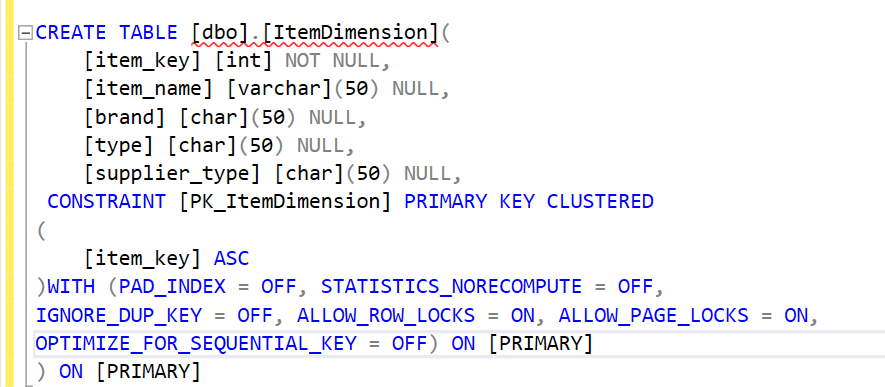
**Figs 1.2 and 1.3 below show the SQL statement for creating the branch dimension and the sample of the data loaded in this table.**

****(fig 1.2)

****(fig 1.3)

**Item dimension**

**Figs 1.4 and 1.5 below show the SQL statement for creating the Item dimension and the sample of the data loaded in this table.**

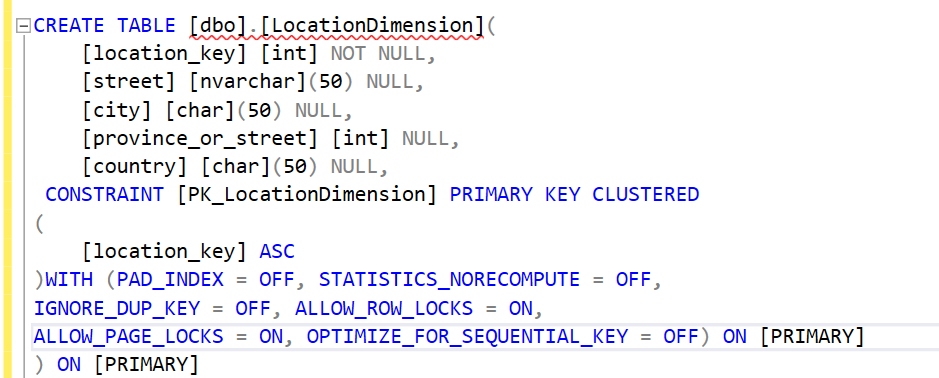
(fig 1.4)

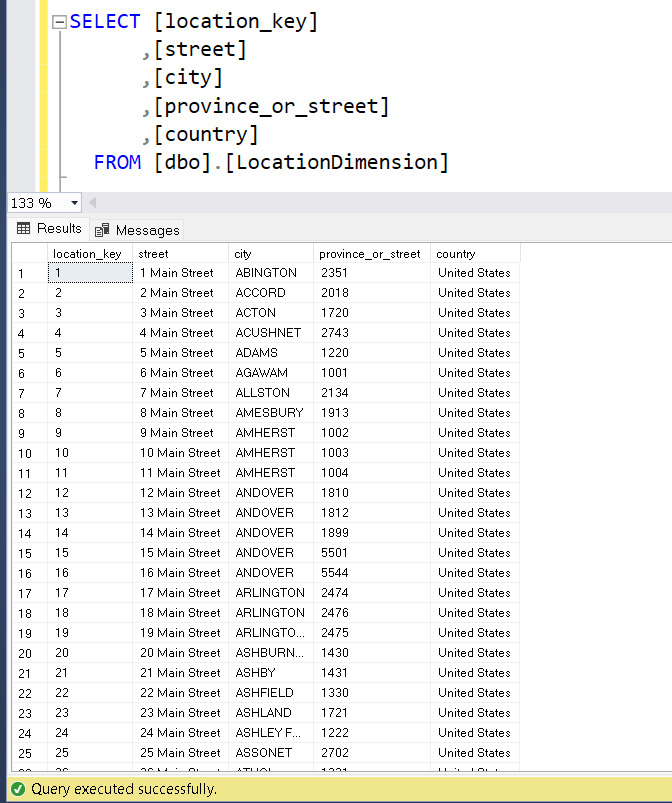
A screenshot of a computer

AI-generated content may be incorrect.(fig 1.5)

**Location dimension**

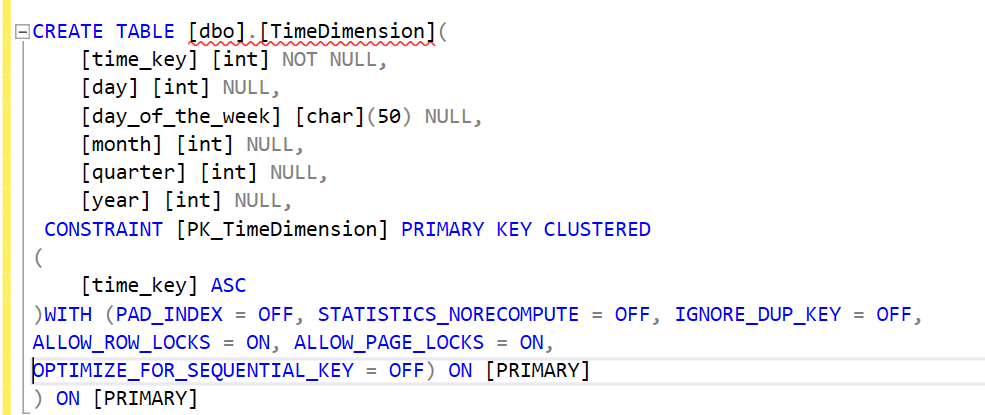
Figs 1.6 and 1.7 below show the SQL statement for creating the Location dimension and the sample of the data loaded in this table.

****(fig 1.6)

(fig1.7)

**Time dimension**

Figs 1.8 and 1.9 below show the SQL statement for creating the Time dimension and the sample of the data loaded in this table.

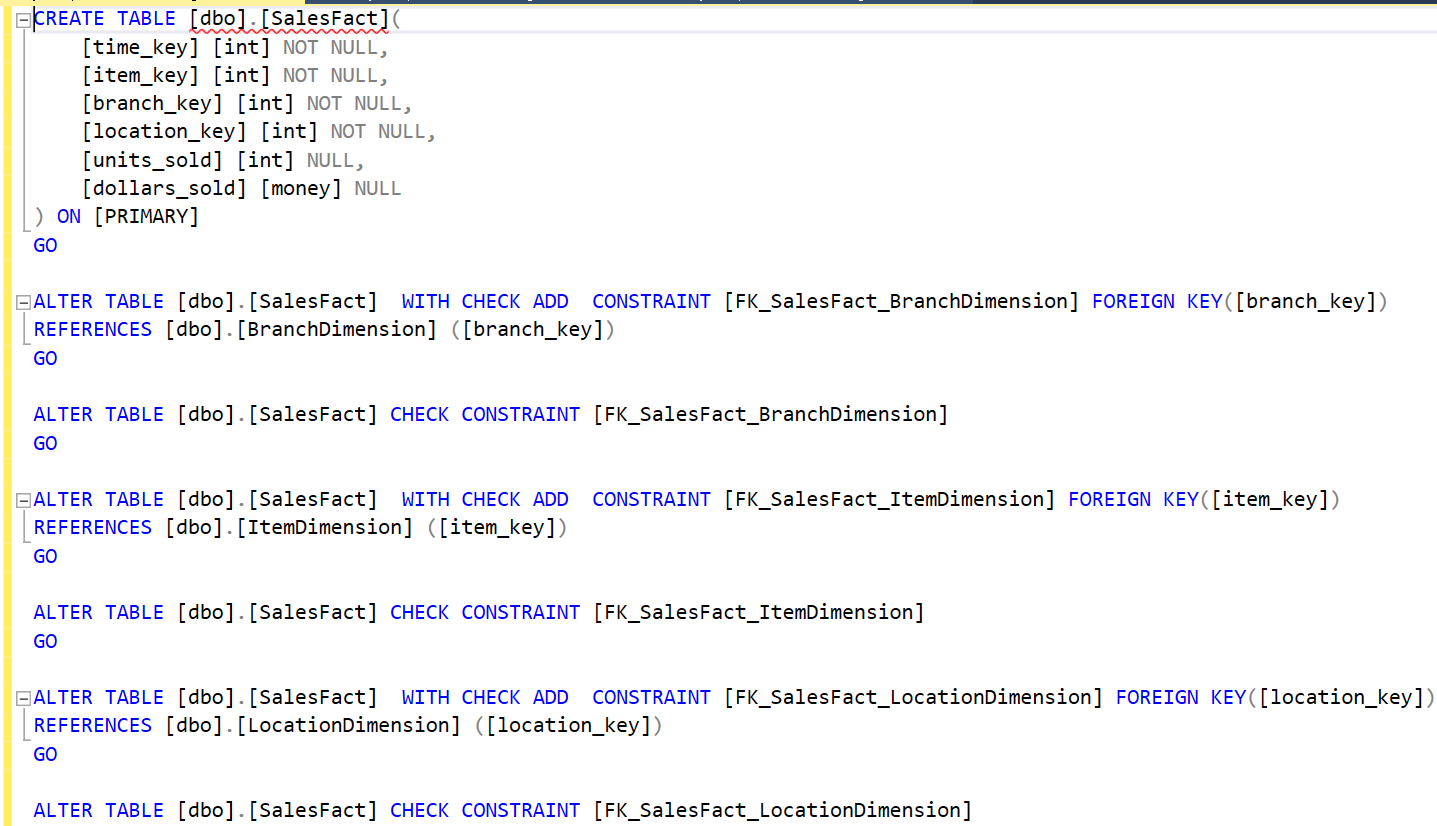
(fig 1.8)

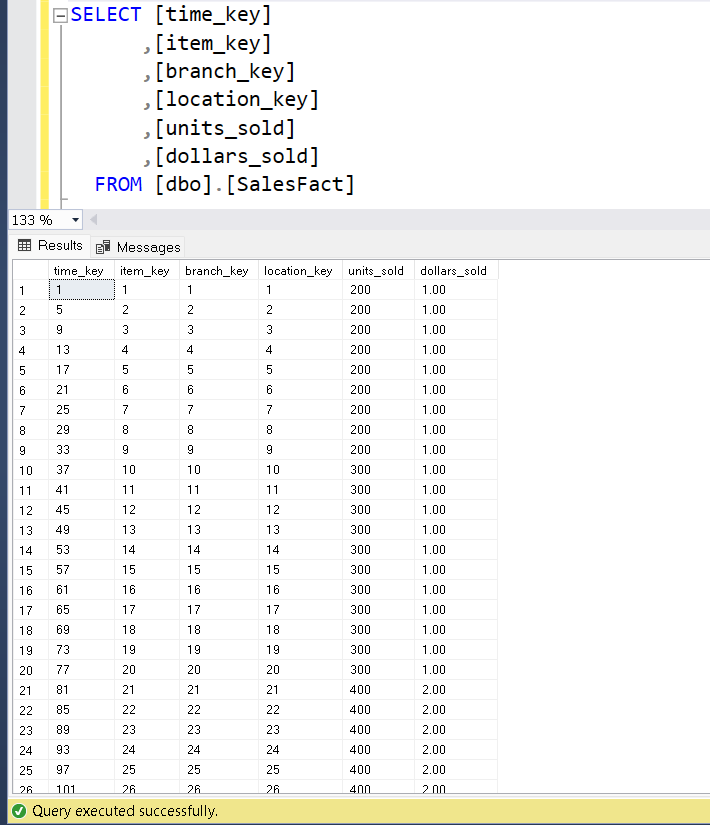
**A screenshot of a computer

AI-generated content may be incorrect.**(fig 1.9)

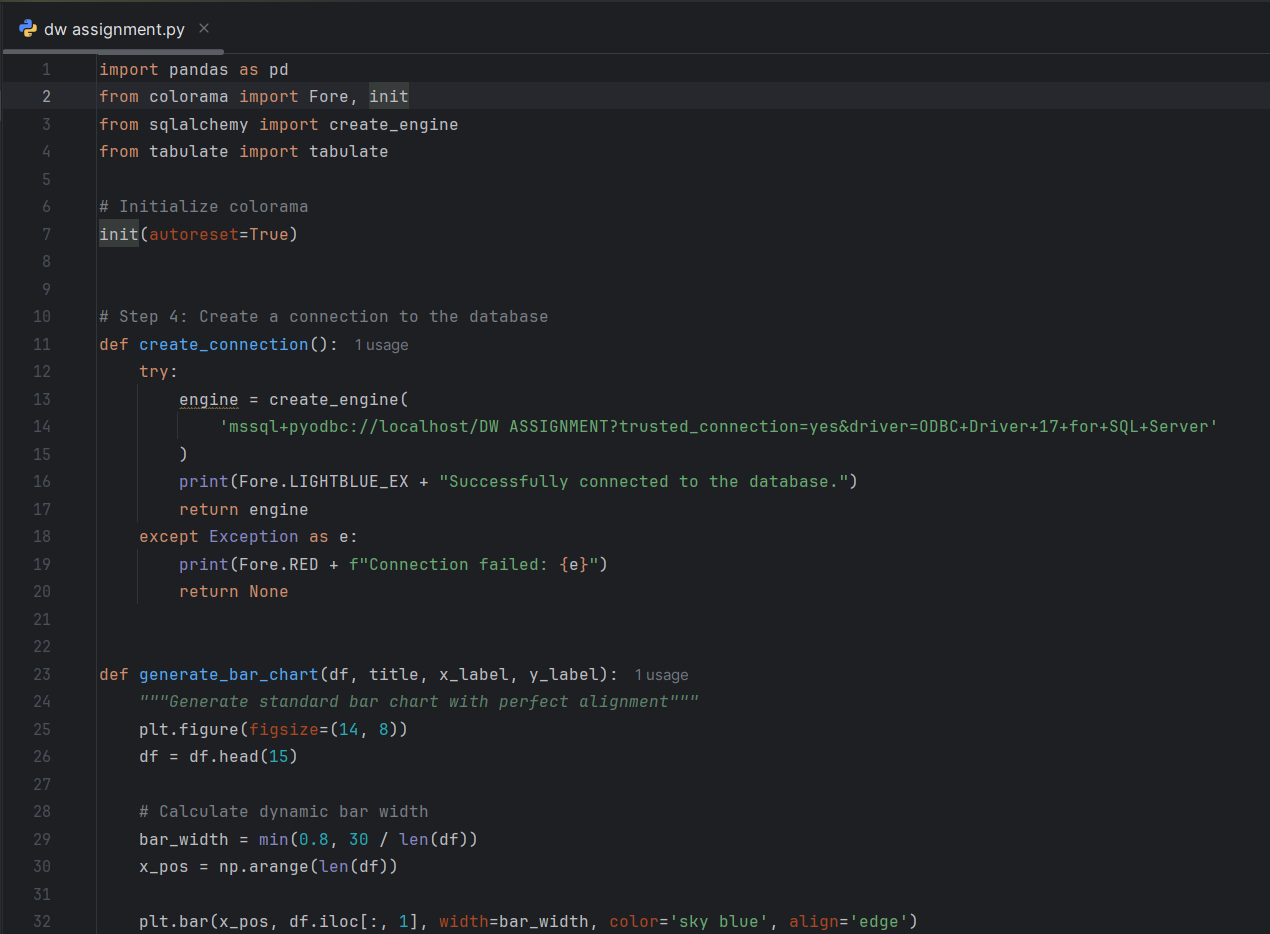
**Sales\_fact table**

Figs 2.0 and 2.1 below show the SQL statement for creating the Sales\_fact dimension and the sample of the data loaded in this table.

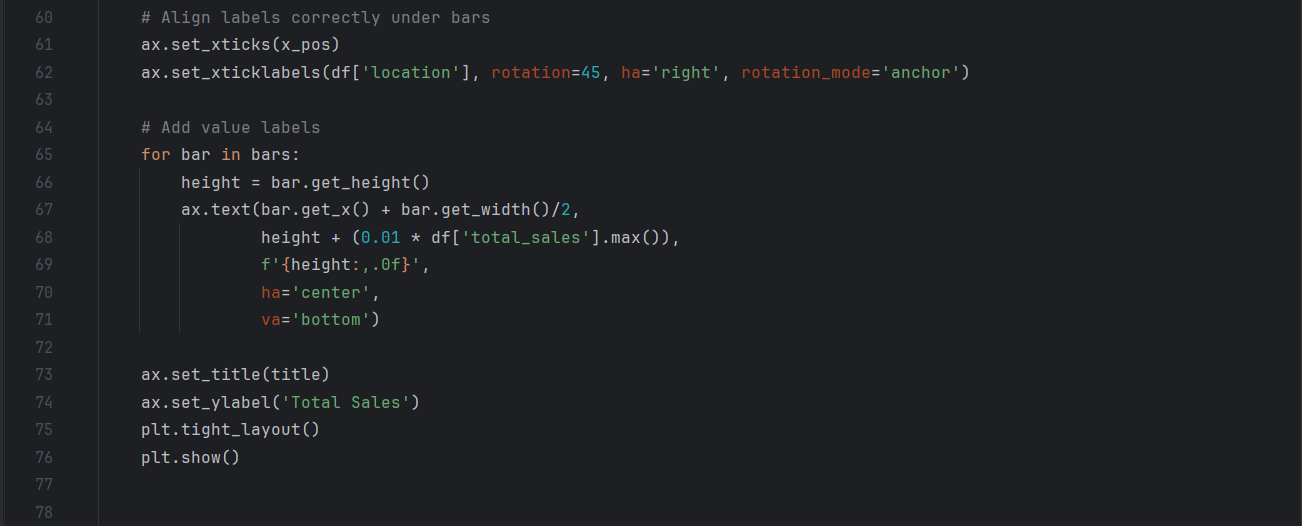
****(fig 2.0)

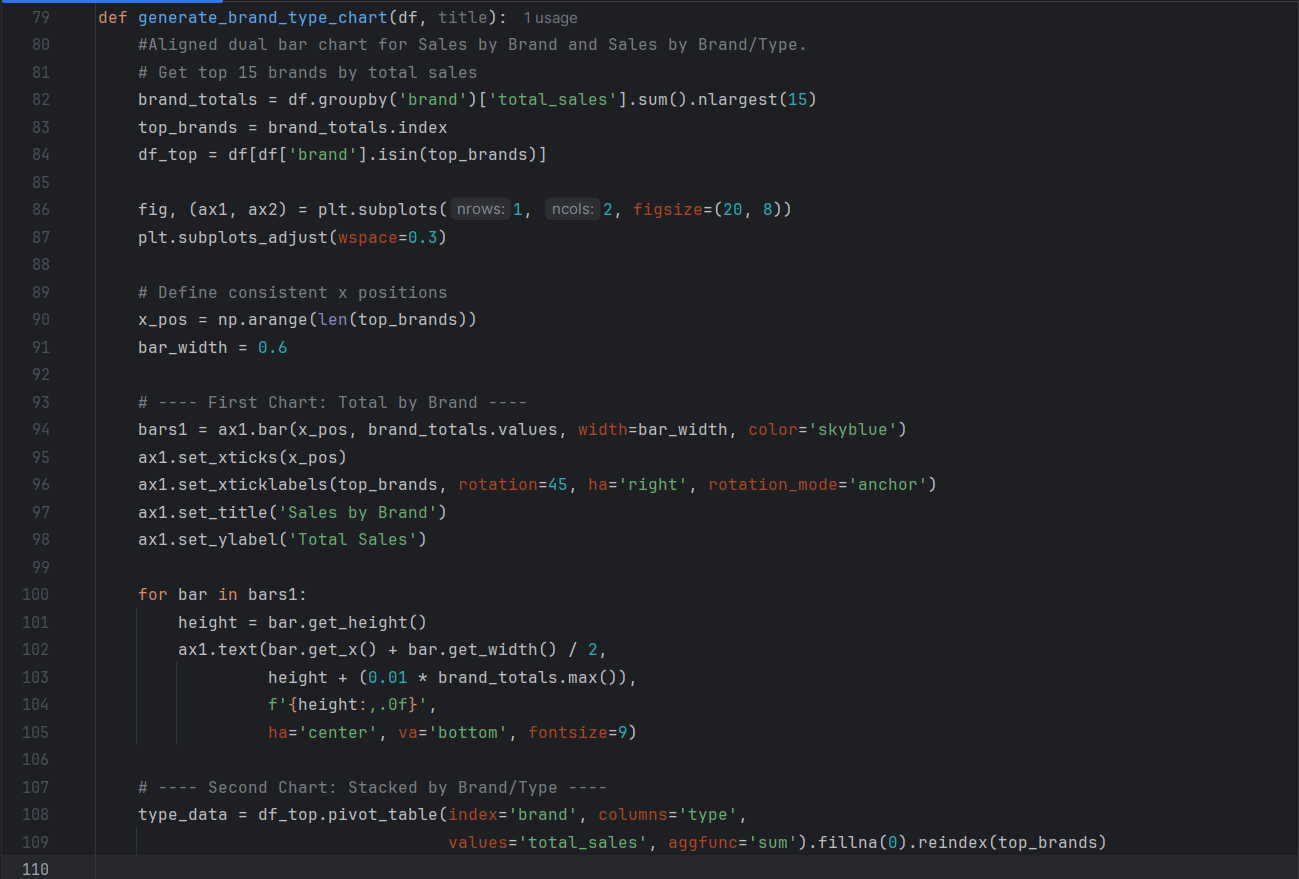
****(fig 2.1)

The code below was used in py-charm to connect with the data base in the SQL server management studio and to perform operations such as drill down and roll up on the data for analysis purposes. The code implements a user interface that allows users to select an option for what operation they want to perform on the data furthermore it also shows the output in coloured texts .Each operation shows a table as well as the bar graph for the query. Due to the vast amount of data ,the data is limited to only 15 entries so that the visualizations can make sense and can be clearly visible, however due to some complications for the second and third query the texts on the x axis is not aligned with the bars.

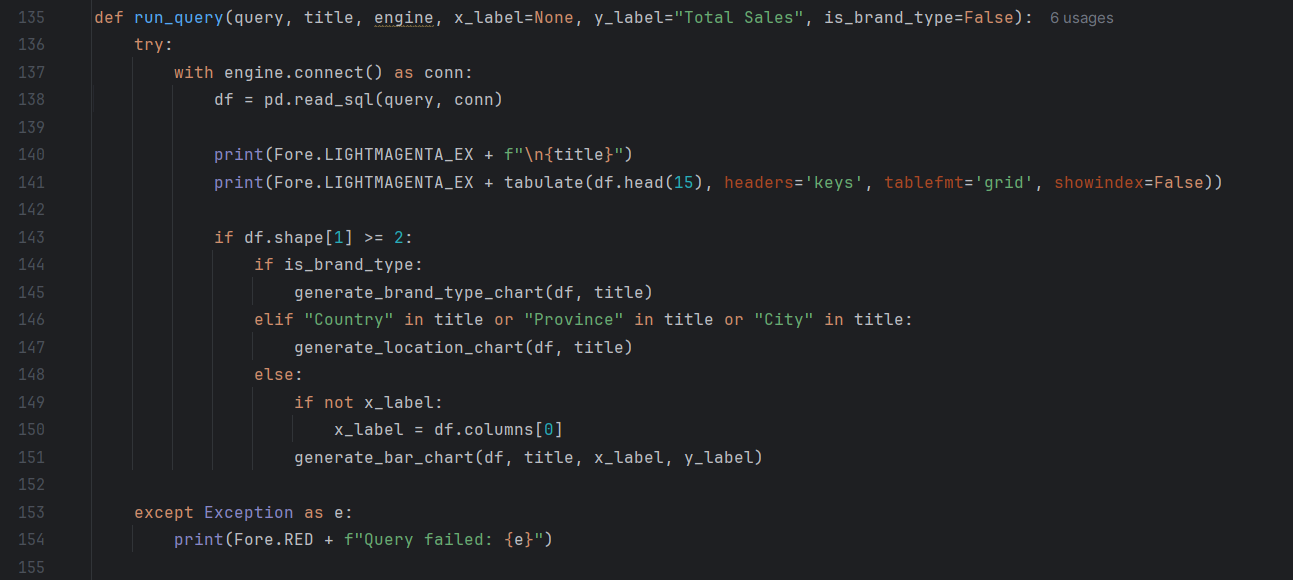


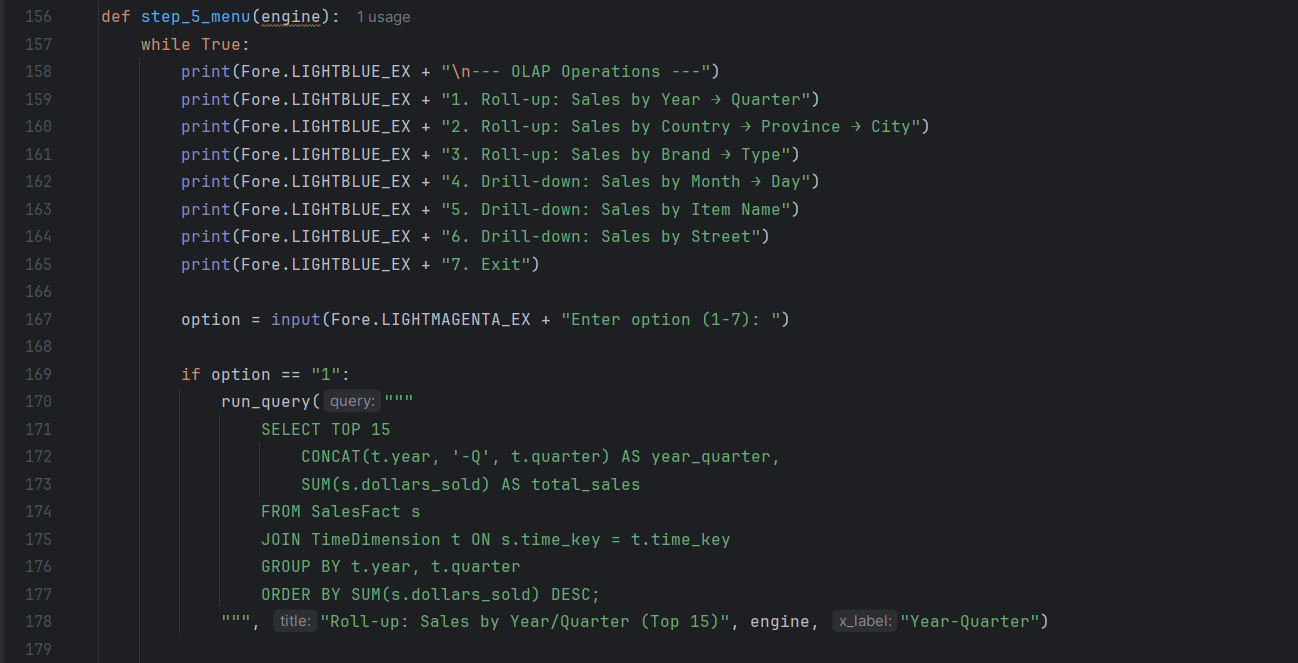






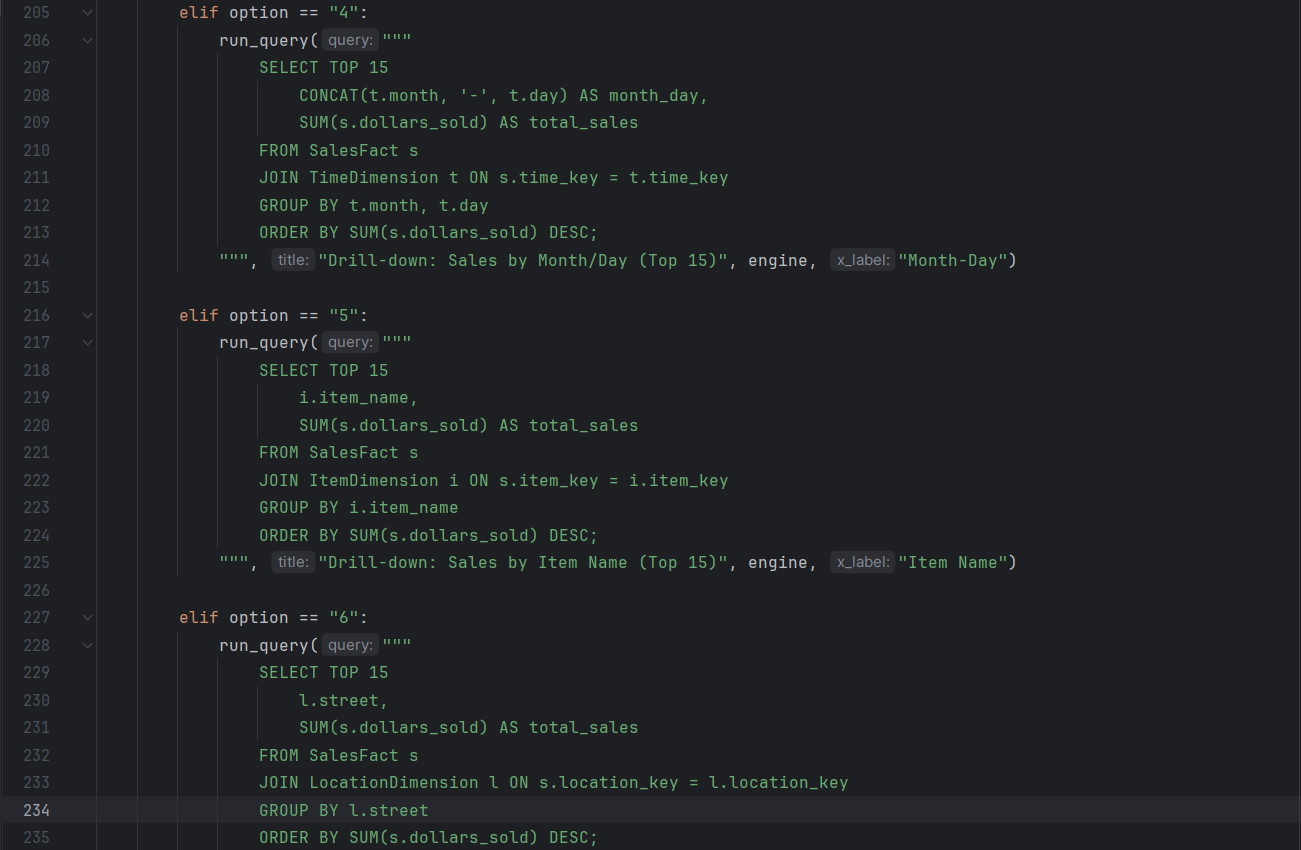






A screenshot of a computer program

AI-generated content may be incorrect.



A screenshot of a computer program

AI-generated content may be incorrect.

OUTPUT

The following is the output of the above code. This is the user face that users will use to interact with the data.

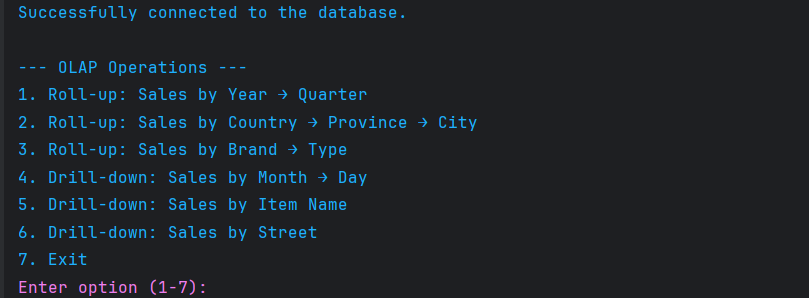
(fig 1.1)

Fig 1.2 shows the first option selected which rolls up the sales by the year/quarter this is the table.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.(fig 1.2)

Fig 1.3 shows us the corresponding bar graph to the first query.

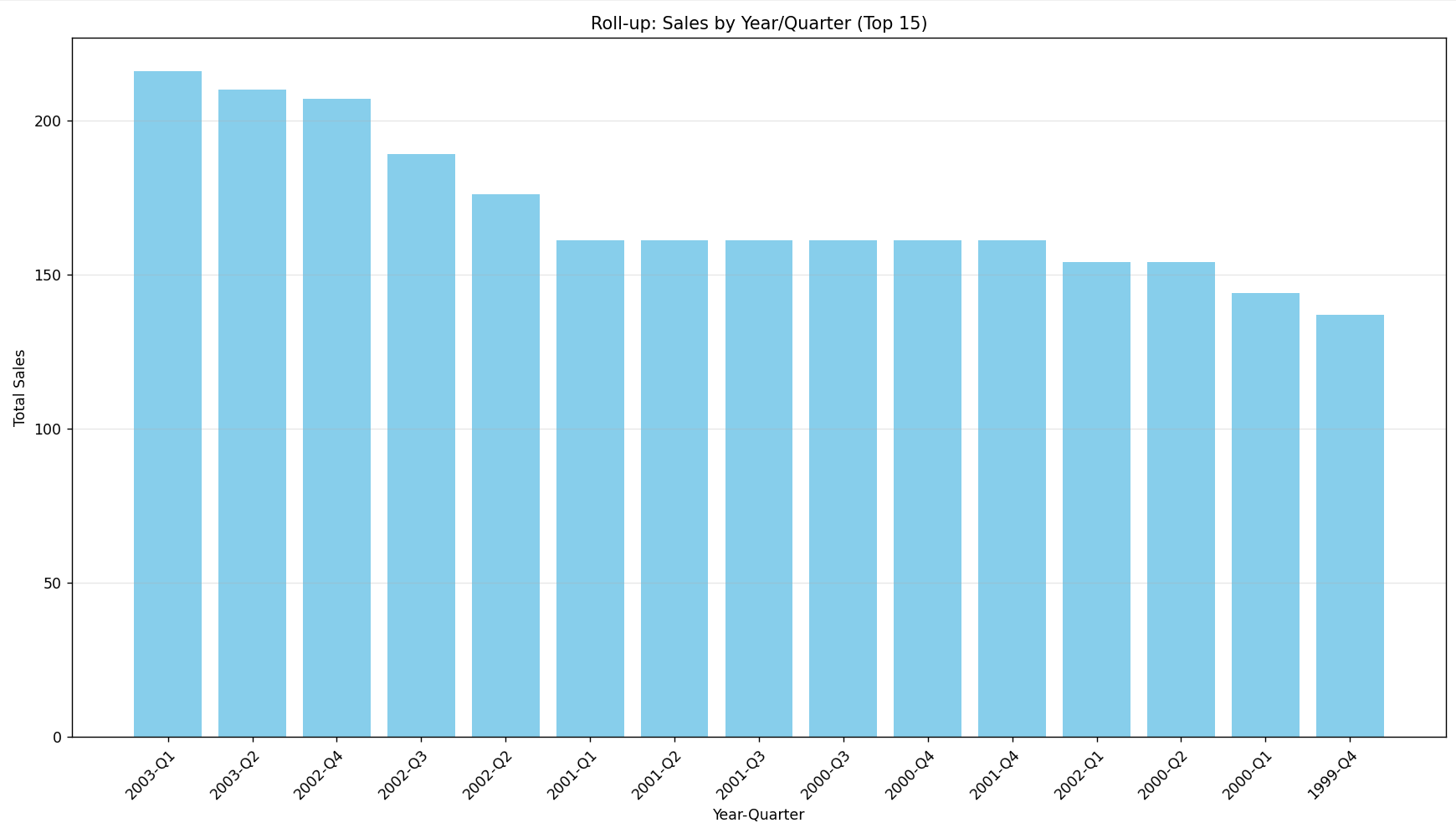
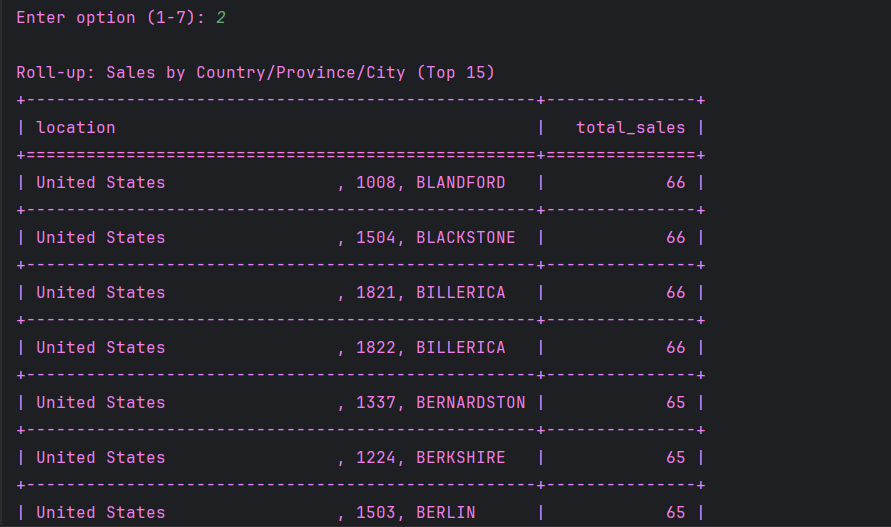
(fig 1.3)

Fig 1.4 shows the Table for the second query and Fig 1.5 shows the corresponding bar graph to this table.



A screenshot of a computer screen

AI-generated content may be incorrect.(fig 1.4)

A graph with blue and white stripes

AI-generated content may be incorrect.

(fig 1.5)

FIG 1.6 shows the table for the third query and Fig 1.7 shows the bar graph for this query.

A screenshot of a computer

AI-generated content may be incorrect.

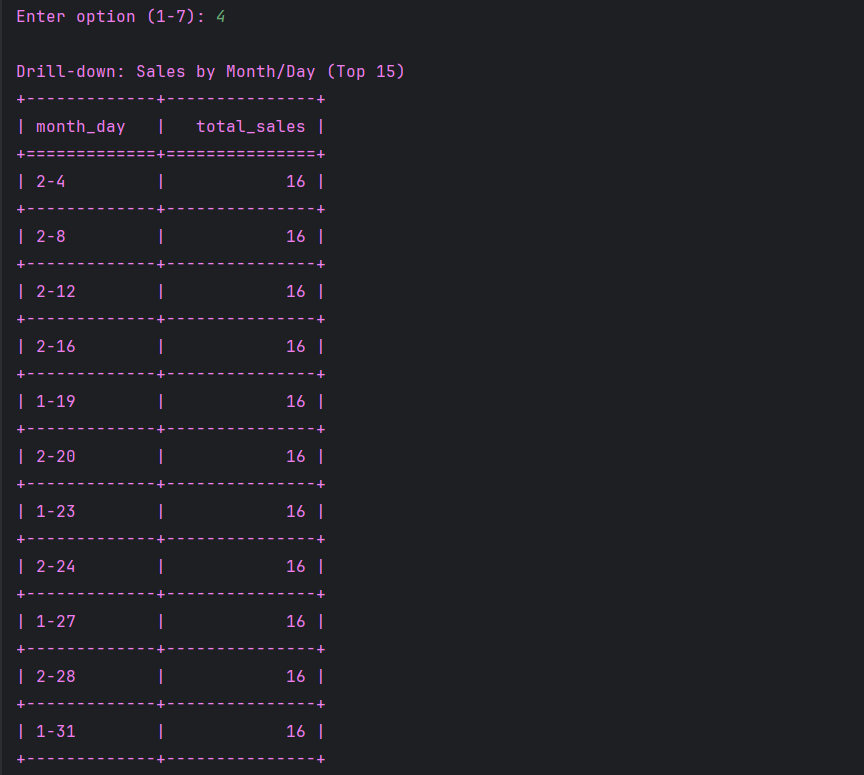
A screenshot of a computer

AI-generated content may be incorrect.(fig 1.6)

A screenshot of a graph

AI-generated content may be incorrect.(fig 1.7)

Fig 1.8 and fig 1.9 show the table as well as the bar graph for the fourth query.



A black background with numbers and lines

AI-generated content may be incorrect.(fig 1.8)

A graph with blue lines

AI-generated content may be incorrect.(fig 1.9)

Fig 2.1 and Fig 2.2 show the table for query five along with its table.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.(fig 2.1)

A graph with blue and white lines

AI-generated content may be incorrect.(fig 2.2)

FIG 2.3 and fig 2.4 shows the table for the sixth query along with its bar graph.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.(fig 2.3)

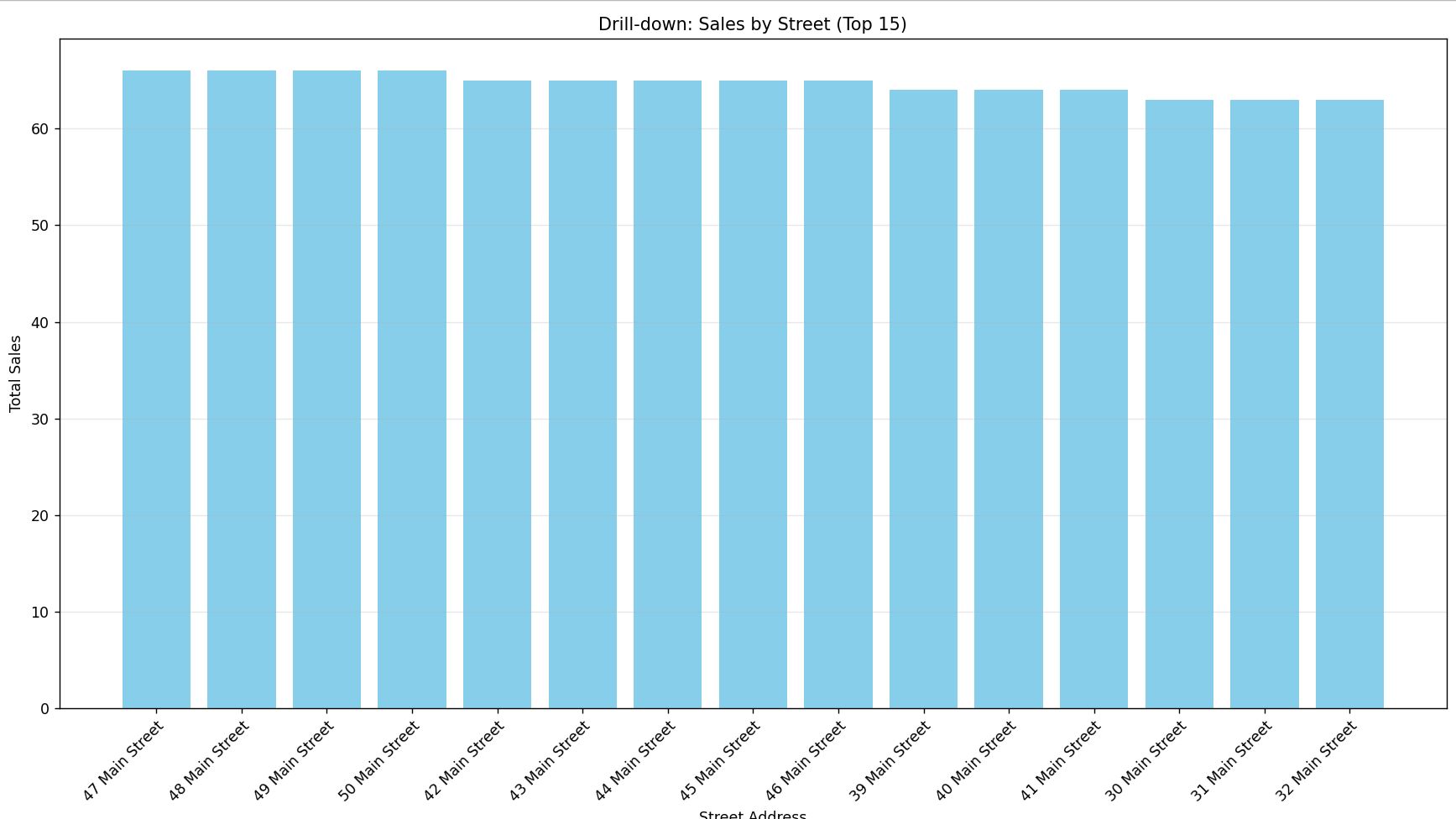
(fig 2.4)

Fig 5 shows us the last option on the user interface which exits the database.



(fig 2.5)