Research Methodology

This study employed a comparative analysis approach to evaluate the performance of Tableau and Power BI dashboards.

Data Collection

The Superstore dataset was obtained from Kaggle and Tableau websites.

Data Preparation

1. Data cleaning: Handling missing values and outliers.
2. Data transformation: Converting data types and aggregating data.
3. Data visualization: Creating dashboards using Tableau and Power BI.

Dashboard Creation

1. Tableau: Utilized Tableau Desktop for dashboard creation.
2. Power BI: Used Power BI Desktop for dashboard creation.

Performance Evaluation

The data sets offer details on meta data about the sales, profit, and other fascinating key factors of a massive superstore.

Data Warehousing and Dataset Overview

The Superstore dataset contains sales, orders, and product information from various regions worldwide, including:

- Canada
- America
- Asia
- Africa
- India

- Cameroon
- Indonesia
- Mexico
- Japan
- Belgium
- France
- England
- United States

Dataset Characteristics

- Comma-Separated Values (CSV) file
- 51,290 rows of data
- Flat file with separate tables for orders and returns
- Order details: 2 files
- Returns details: 1 file

Data Preparation and Warehousing

To handle large datasets, proper cleaning and preparation are necessary. Tableau and Power BI may struggle with processing large amounts of data.

Proposed Research Methodology

To facilitate data analysis, this study employs a data warehousing approach to clean and prepare the data. We utilize data mining techniques to extract relevant information, ensuring accurate analysis.

Data Processing and Manipulation

Using a business scenario, we will demonstrate data importation, processing, and manipulation with Tableau Desktop and Power BI Desktop. Both tools offer free downloads and support live or desktop server connections.

Data Cleaning and Preparation

Although Tableau offers automatic data cleaning, we opt for the data warehousing method (Extract, Transform, Load - ETL) to preserve crucial customer information. This approach ensures data integrity and relevance.

Research Methodology Steps

Figure 12 illustrates the process:

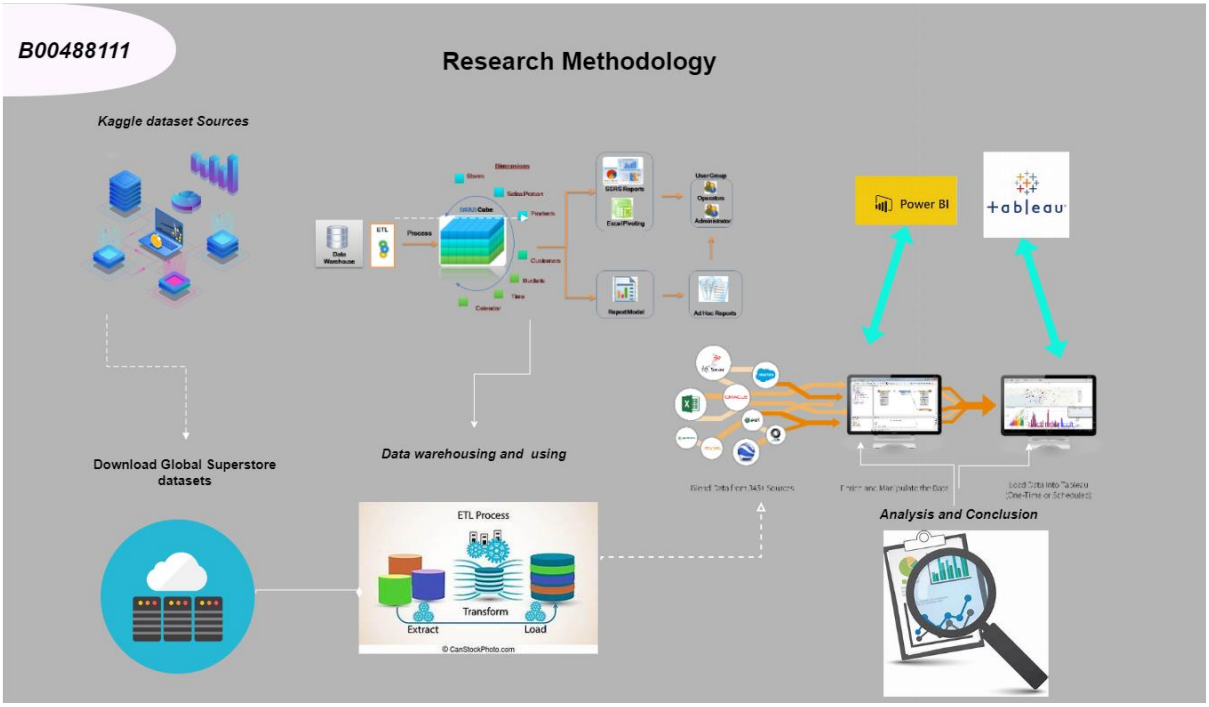


Figure 1: Research methodology

1. Data Extraction: Kaggle dataset download
2. Data Transformation: Data warehousing (ETL)
3. Data Loading: Into Tableau and Power BI
4. Data Analysis: Using business scenarios
5. Results Comparison: Tableau vs Power BI

Dashboard Performance Analysis

Data Preparation

Cleaning and normalizing data is crucial before importing it into visualization tools for business performance monitoring. Although Tableau handles more datasets than Power BI, improper data cleaning can lead to errors.

Data Import and Analysis

After data normalization, we import datasets from the DataMart created in our Datawarehouse processing. Extracted data is organized into sheets for analysis.

Research Objective

Investigate key aspects of Tableau and Power BI using a business scenario.

Research Objective

This study aims to investigate the differences in performance and data analysis capabilities between Tableau and Power BI, despite their similarities in presenting information on dashboards.

Literature Gap

Existing research has explored these tools, but new features and updates from vendors remain unexamined. This study fills this gap.

Methodology

Data Mining using ETL Processes

Table 2 outlines our plan of action:

1. Search datasets on Kaggle
2. Download datasets as CSV files
3. Convert files to Excel 2003-2007 format
4. Connect to CSVSOCSQL01\SQL2012 database engine
5. Create B0048811SalesDataMartMSc data warehouse

6. Create Staging Area and DataMart schemas

7. Create tables with SQL queries in DataMart schema

8. Import data into Staging Area using Import Wizard

9. Analyse datasets and fetch information

10. Run SQL queries to export datasets

Data Visualization and Analysis

Tableau Desktop

1. Import data

2. Create new sheets

3. Create columns and measures

4. Create visuals

5. Format visuals

1. Import data

2. Edit in Power Query Editor

3. Create columns and measures

4. Create visuals

5. Format visuals

Results and Evaluation

1. Compare results

2. Evaluate performance differences

3. Draw conclusions

Tools Used

1. Tableau Desktop (Version 2022.1)

2. Power BI Desktop (Version 2.104.702.0 32-bit)

3. Microsoft SQL Server 2012

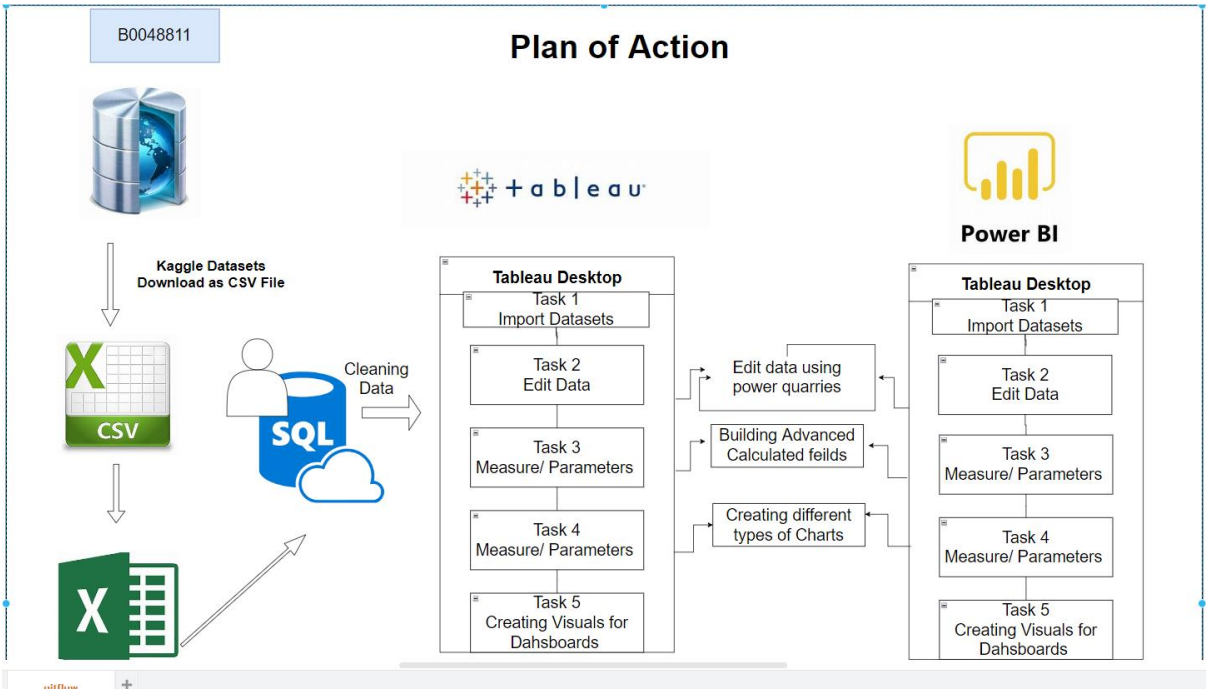


Table 1: Plan of Action

Proposed Methods

Data mining, as defined by Friedman, involves extracting valuable patterns from large, complex datasets (Luo, 2022). This study employs data mining techniques to analyze sales data from a Global Superstore.

Dataset Description

The dataset, sourced from a SQL database, is saved in CSV format and contains:

- Global Superstore sales data

- 2 sheets (Global Superstore returns 2016)

- 1 sheet (Global Superstore Returns 2016)

Data Cleaning and Preparation

To manage large datasets, a single data source is essential. We utilize:

- SQL Server for data storage

- CSV files for data exchange

- Power BI and Tableau for visualization

Data Analysis and Visualization

Our methodology involves:

1. Data cleaning and preparation

2. Data analysis using advanced metrics and calculated fields

3. Visualization using Power BI and Tableau

4. Drawing valuable conclusions from dashboard insights

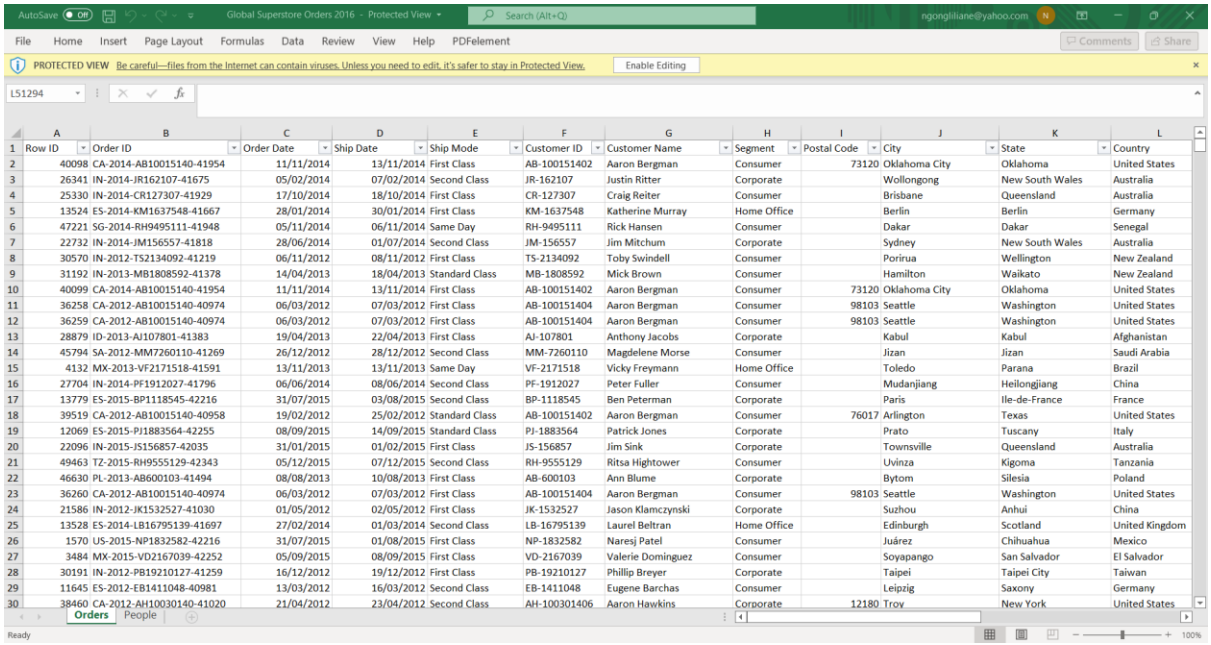
Tools Used

- Tableau Desktop (Version 2022.1)

- Power BI Desktop (Version 2.104.702.0 32-bit)

- Microsoft SQL Server 2012

Changes made:



Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Postal Code	City	State	Country
40098	CA-2014-AB10015140-41954	11/11/2014	13/11/2014	First Class	AB-100151402	Aaron Bergman	Consumer	73120	Oklahoma City	Oklahoma	United States
26341	IN-2014-JR162107-41675	05/02/2014	07/02/2014	Second Class	JR-162107	Justin Ritter	Corporate		Wollongong	New South Wales	Australia
25330	IN-2014-CR127307-41929	17/10/2014	18/10/2014	First Class	CR-127307	Craig Reiter	Consumer		Brisbane	Queensland	Australia
13524	ES-2014-KM1637548-41667	28/01/2014	30/01/2014	First Class	KM-1637548	Katherine Murray	Home Office		Berlin	Berlin	Germany
47221	SG-2014-RH9495111-41948	05/11/2014	06/11/2014	Same Day	RH-9495111	Rick Hansen	Consumer		Dakar	Dakar	Senegal
22732	IN-2014-JM156557-41818	28/06/2014	01/07/2014	Second Class	JM-156557	Jim Mitchum	Corporate		Sydney	New South Wales	Australia
30570	IN-2012-TS2134092-41219	06/11/2012	08/11/2012	First Class	TS-2134092	Toby Swindell	Consumer		Porirua	Wellington	New Zealand
31192	IN-2013-MB1808592-41378	14/04/2013	18/04/2013	Standard Class	MB-1808592	Mick Brown	Consumer		Hamilton	Waikato	New Zealand
40099	CA-2014-AB10015140-41954	11/11/2014	13/11/2014	First Class	AB-100151402	Aaron Bergman	Consumer	73120	Oklahoma City	Oklahoma	United States
36258	CA-2012-AB10015140-40974	06/03/2012	07/03/2012	First Class	AB-100151404	Aaron Bergman	Consumer	98103	Seattle	Washington	United States
36259	CA-2012-AB10015140-40974	06/03/2012	07/03/2012	First Class	AB-100151404	Aaron Bergman	Consumer	98103	Seattle	Washington	United States
28879	ID-2013-AJ107801-41383	19/04/2013	22/04/2013	First Class	AJ-107801	Anthony Jacobs	Corporate		Kabul	Kabul	Afghanistan
45794	SA-2012-MM7260110-41269	26/12/2012	28/12/2012	Second Class	MM-7260110	Magdelene Morse	Consumer		Jizan	Jizan	Saudi Arabia
4132	MX-2013-VF2171518-41591	13/11/2013	13/11/2013	Same Day	VF-2171518	Vicky Freymann	Home Office		Toledo	Parana	Brazil
27704	IN-2014-PF1912027-41796	06/06/2014	08/06/2014	Second Class	PF-1912027	Peter Fuller	Consumer		Mudanjiang	Heilongjiang	China
13779	ES-2015-BP1118545-42216	31/07/2015	03/08/2015	Second Class	BP-1118545	Ben Peterman	Corporate		Paris	Ile-de-France	France
39519	CA-2012-AB10015140-40958	19/02/2012	25/02/2012	Standard Class	AB-100151402	Aaron Bergman	Consumer	76017	Arlington	Texas	United States
12069	ES-2015-PJ1883564-42255	08/09/2015	14/09/2015	Standard Class	PJ-1883564	Patrick Jones	Corporate		Prato	Tuscany	Italy
22096	IN-2015-JS156857-42035	31/01/2015	01/02/2015	First Class	JS-156857	Jim Sink	Corporate		Townsville	Queensland	Australia
49463	TZ-2015-RH9555129-42343	05/12/2015	07/12/2015	Second Class	RH-9555129	Ritsa Hightower	Consumer		Uvinza	Kigoma	Tanzania
46630	PL-2013-AB600103-41494	08/08/2013	10/08/2013	First Class	AB-600103	Ann Blume	Corporate		Bytom	Silesia	Poland
36260	CA-2012-AB10015140-40974	06/03/2012	07/03/2012	First Class	AB-100151404	Aaron Bergman	Consumer	98103	Seattle	Washington	United States
21586	IN-2012-JK1532527-41030	01/05/2012	02/05/2012	First Class	JK-1532527	Jason Klamczynski	Corporate		Suzhou	Anhui	China
13528	ES-2014-LB16795139-41697	27/02/2014	01/03/2014	Second Class	LB-16795139	Laurel Beltran	Home Office		Edinburgh	Scotland	United Kingdom
1570	US-2015-NP1832582-42216	31/07/2015	01/08/2015	First Class	NP-1832582	Naresj Patel	Consumer		Juárez	Chihuahua	Mexico
3484	MX-2015-VD2167039-42252	05/09/2015	08/09/2015	First Class	VD-2167039	Valerie Dominguez	Consumer		Soyapango	San Salvador	El Salvador
30191	IN-2012-PB19210127-41259	16/12/2012	19/12/2012	First Class	PB-19210127	Phillip Breyer	Corporate		Taipei	Taipei City	Taiwan
11645	ES-2012-EB1411048-40981	13/03/2012	16/03/2012	Second Class	EB-1411048	Eugene Barchas	Consumer		Leipzig	Saxony	Germany
38460	CA-2012-AH10030140-41020	21/04/2012	23/04/2012	Second Class	AH-100301406	Aaron Hawkins	Corporate	12180	Troy	New York	United States

Proposed Methods

Data mining, as defined by Friedman, involves extracting valuable patterns from large, complex datasets (Luo, 2022). This study employs data mining techniques to analyse sales data from a Global Superstore.

Dataset Description

The dataset, sourced from a SQL database, is saved in CSV format and contains:

- Global Superstore sales data

- 2 sheets (Global Superstore returns 2016)

- 1 sheet (Global Superstore Returns 2016)

Data Cleaning and Preparation

To manage large datasets, a single data source is essential. We utilize:

- SQL Server for data storage

- CSV files for data exchange

- Power BI and Tableau for visualization

Data Analysis and Visualization

Our methodology involves:

1. Data cleaning and preparation
2. Data analysis using advanced metrics and calculated fields
3. Visualization using Power BI and Tableau
4. Drawing valuable conclusions from dashboard insights

Tools Used

- Tableau Desktop (Version 2022.1)
- Power BI Desktop (Version 2.104.702.0 32-bit)
- Microsoft SQL Server 2012

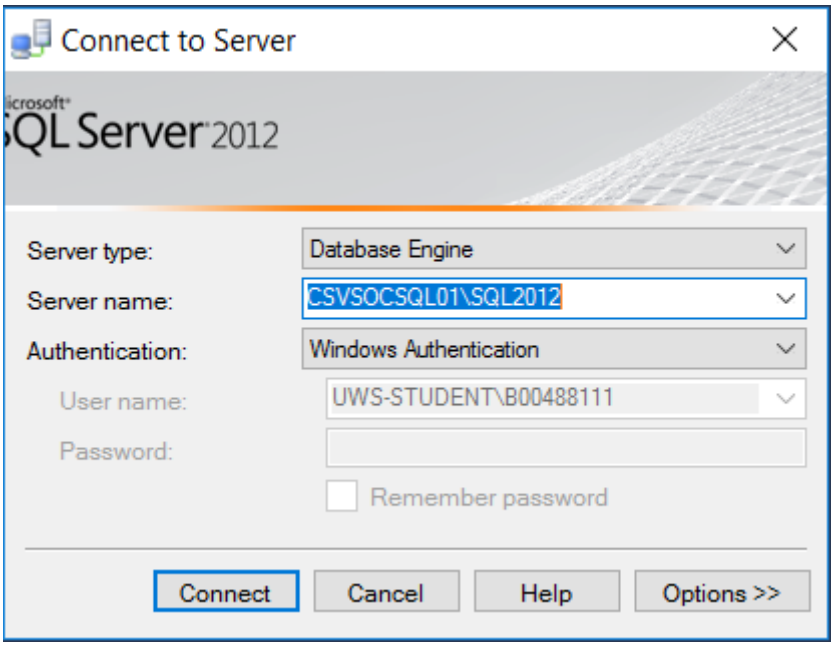


Figure 2: *Screenshot of Database Engine Connection*

Database Setup and Schema Design

After connecting to the database engine, I created:

1. Database: B0048811SalesDataMartMSc
2. Schemas:
   - Staging Area (for data import)
   - DataMart (for data transformation and loading)

Dimension Creation

In the DataMart schema, I created four dimensions using SQL codes:

1. DimCustomer
2. DimProduct
3. DimOrders
4. DimFactsGlobalSales

SQL Statements

Below are the SQL statements used to create dimensions and populate tables with extracted data:

```
-- Create database
CREATE DATABASE B0048811SalesDataMartMSc
USE B0048811SalesDataMartMSc;
go
CREATE SCHEMA StagingArea;
CREATE SCHEMA DataMart;
```

```
-- Create schemas
CREATE SCHEMA StagingArea;
CREATE SCHEMA DataMart;

-- Create dimensions
CREATE TABLE DataMart.DimCustomer (
  CustomerID INT PRIMARY KEY,
  CustomerName VARCHAR(255),
  Address VARCHAR(255),
  City VARCHAR(100),
  Country VARCHAR(100)
);
CREATE TABLE DataMart.DimProduct (
  ProductID INT PRIMARY KEY,
  ProductName VARCHAR(255),
  Category VARCHAR(100),
  Price DECIMAL(10, 2)
);

CREATE TABLE DataMart.DimOrders (
  OrderID INT PRIMARY KEY,
  OrderDate DATE,
  CustomerID INT,
  ProductID INT,
  Quantity INT
);

CREATE TABLE DataMart.DimFactsGlobalSales (
  SalesID INT PRIMARY KEY,
```

```
  OrderID INT,
  ProductID INT,
  SalesAmount DECIMAL(10, 2),
  SalesDate DATE
);

-- Populate tables with extracted data
INSERT INTO DataMart.DimCustomer (CustomerID, CustomerName, Address, City, Country)
SELECT CustomerID, CustomerName, Address, City, Country

INSERT INTO DataMart.DimProduct (ProductID, ProductName, Category, Price)
SELECT ProductID, ProductName, Category, Price
FROM StagingArea.Products;

INSERT INTO DataMart.DimOrders (OrderID, OrderDate, CustomerID, ProductID, Quantity)
SELECT OrderID, OrderDate, CustomerID, ProductID, Quantity
FROM StagingArea.Orders;

INSERT INTO DataMart.DimFactsGlobalSales (SalesID, OrderID, ProductID, SalesAmount, SalesDate)
SELECT SalesID, OrderID, ProductID, SalesAmount, SalesDate
```

FROM StagingArea.Sales;

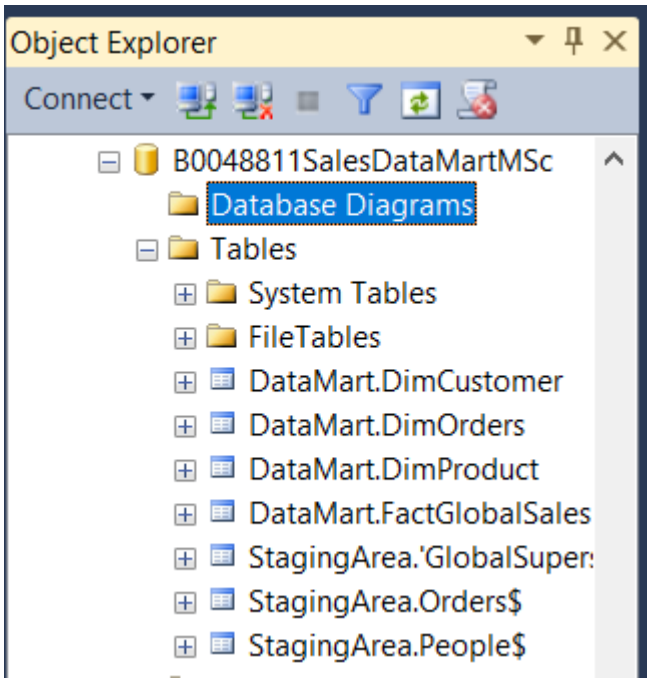


Figure 3: *Database, DataMart, and Schema's*

Importing Data into Staging Area

To import data into the Staging Area:

1. Right-click on the database (B0048811SalesDataMartMSc)
2. Select "Tasks" from the dropdown menu
3. Choose "Import Data"
4. The Import Wizard will appear

Selecting Data Source

In the Import Wizard:

1. Choose "Excel" as the data source

2. Select the Excel file containing the data

3. Proceed to the "Import Data" option

Successful Importation

The image below illustrates the successful importation of datasets into the Staging Area:

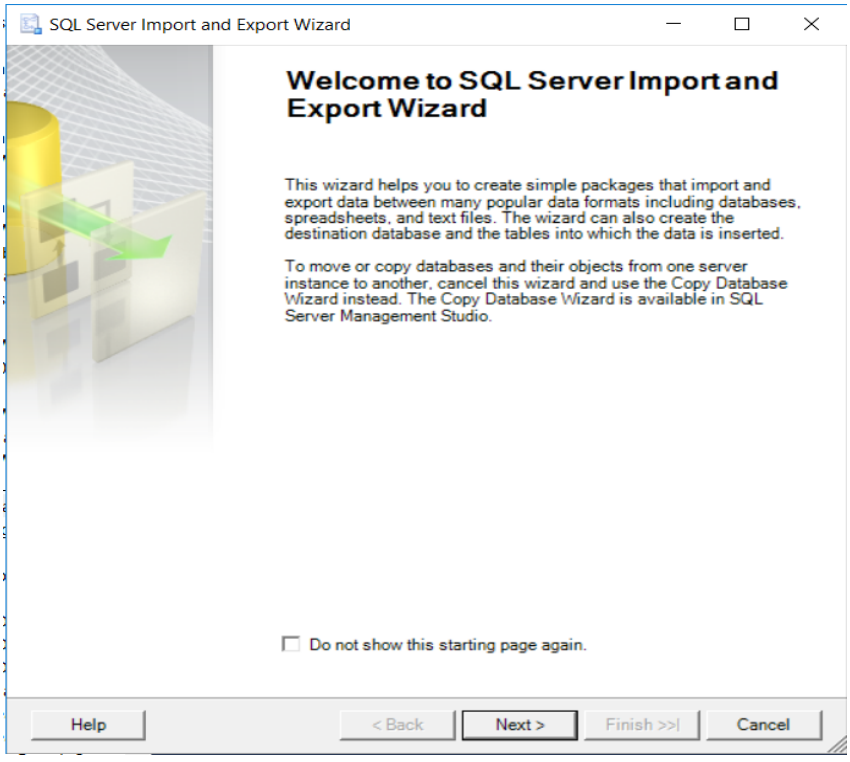


Figure 4: *Data Importation Using Import Wizard*

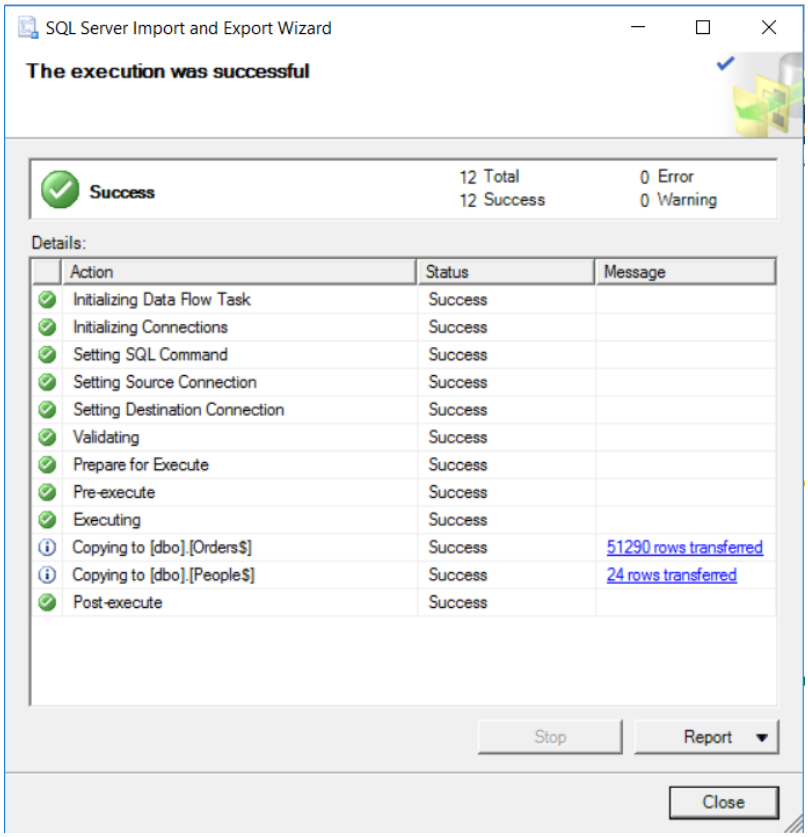


Figure 5: *Successfully Importing Datasets*

Transferring Data to Staging Area

To transfer imported data into the Staging Area schema:

```
-- Transfer table (orders$) into schema (StagingArea)
USE B0048811SalesDataMartMSc
ALTER SCHEMA StagingArea TRANSFER dbo.SampleSuperstore$;
GO

-- Transfer schema into schema (StagingArea)
ALTER SCHEMA StagingArea TRANSFER StagingArea.Orders$;
GO
```

```
ALTER SCHEMA StagingArea TRANSFER dbo.People$;
GO
ALTER SCHEMA StagingArea TRANSFER StagingArea.People$;
GO
ALTER SCHEMA StagingArea TRANSFER dbo.['GlobalSuperstoreReturns2016$'];
GO
ALTER SCHEMA StagingArea TRANSFER StagingArea.['GlobalSuperstoreReturns2016$'];
GO
```

Creating Dimensions and Extraction

Dimension: Customer

```
-- Create DimCustomer table in DataMart schema
USE B0048811SalesDataMartMSc
CREATE TABLE DataMart.DimCustomer (
    customerID NVARCHAR(255),
    customerName NVARCHAR(255),
    country NVARCHAR(255),
    region NVARCHAR(255),
    segment NVARCHAR(255),
    CONSTRAINT pkcustomerID PRIMARY KEY (customerID)
);
GO

-- Insert values into DimCustomer table
INSERT INTO DataMart.DimCustomer
SELECT DISTINCT
    [CustomerID],
```

```
    [CustomerName],
    [Country],
    [region],
    [Segment]
FROM
    [B0048811SalesDataMartMSc].[StagingArea].[Orders$];
```

Viewing DimCustomer Table

```
-- View all records in DimCustomer table
USE B0048811SalesDataMartMSc
GO
SELECT *
FROM DataMart.DimCustomer;
```

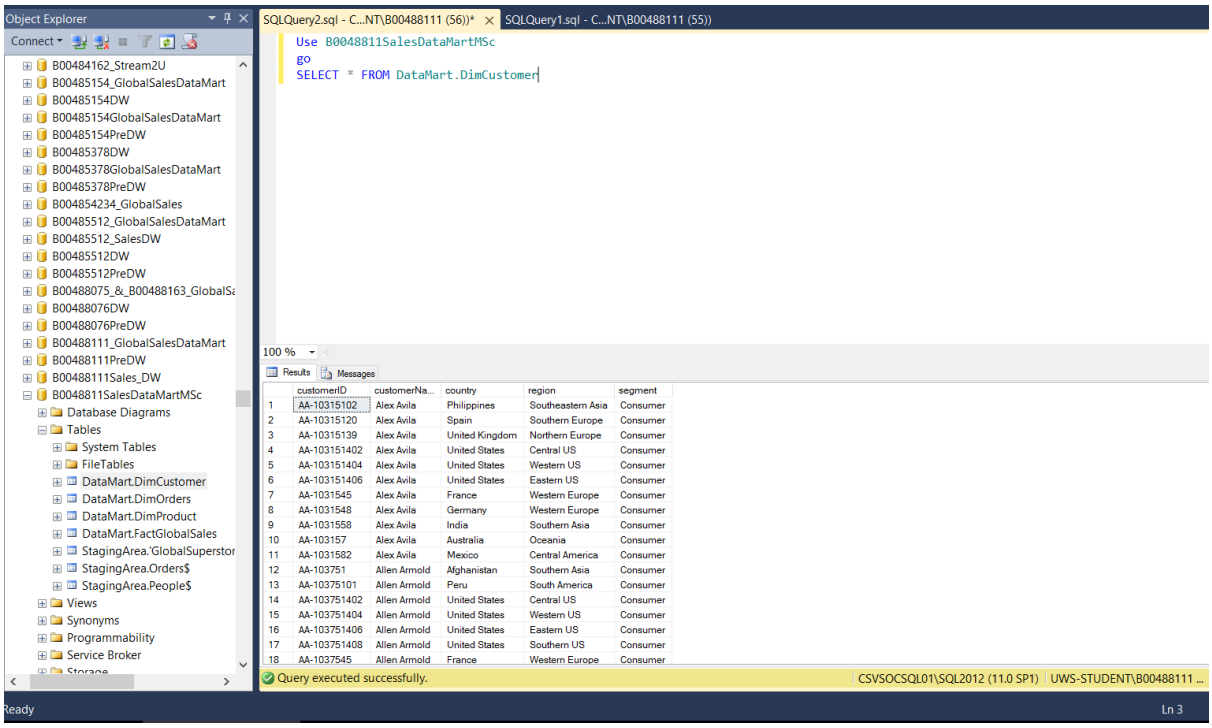


Figure 6: *Dimension Customer Table*

Dimension: Product

-- Create DimProduct table in DataMart schema

USE B0048811SalesDataMartMSc

CREATE TABLE DataMart.DimProduct (

productID NVARCHAR(255),

productName NVARCHAR(255),

category NVARCHAR(255),

subCategory NVARCHAR(255),

CONSTRAINT pkproductID PRIMARY KEY (productID)

);

GO

-- Insert values into DimProduct table

INSERT INTO DataMart.DimProduct

SELECT DISTINCT

[ProductID],

[ProductName],

[Category],

[SubCategory]

FROM

[B0048811SalesDataMartMSc].[StagingArea].[Orders$];

GO

``"

Viewing DimProduct Table

sql

-- View all records in DimProduct table

USE B0048811SalesDataMartMSc

GO

*SELECT **

FROM DataMart.DimProduct;

```

*Query Results:*

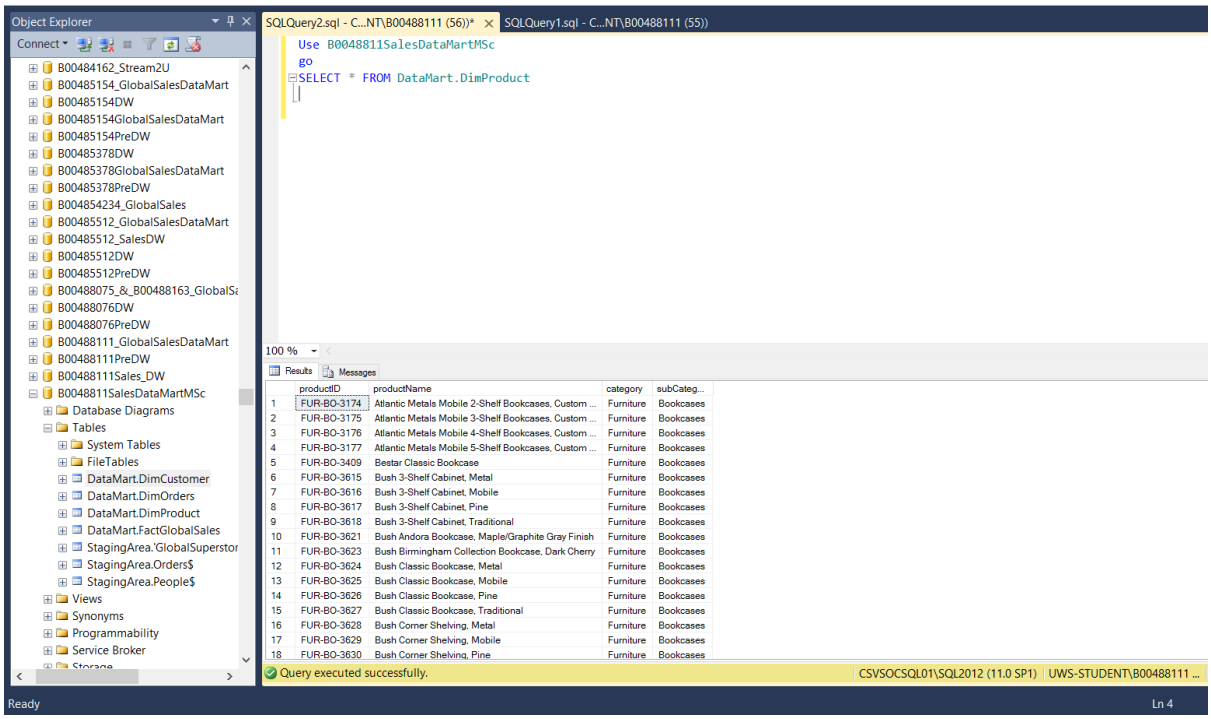



***Figure 7:*** *Dimension Product Table content*

Here's a rewritten version with improved clarity, structure, and formatting:

**Dimension: Orders**

*-- Create DimOrders table in DataMart schema*

*USE B0048811SalesDataMartMSc*

*CREATE TABLE DataMart.DimOrders (*

*orderID NVARCHAR(255) PRIMARY KEY,*

*orderDate DATETIME*
```

```
);
GO

-- Insert values into DimOrders table
INSERT INTO DataMart.DimOrders
SELECT DISTINCT
    [OrderID],
    [OrderDate]
FROM
    [B0048811SalesDataMartMSc].[StagingArea].[Orders$];
GO
```

Viewing DimOrders Table

```
-- View all records in DimOrders table
USE B0048811SalesDataMartMSc
GO
SELECT *
FROM DataMart.DimOrders;
```

Query Results:

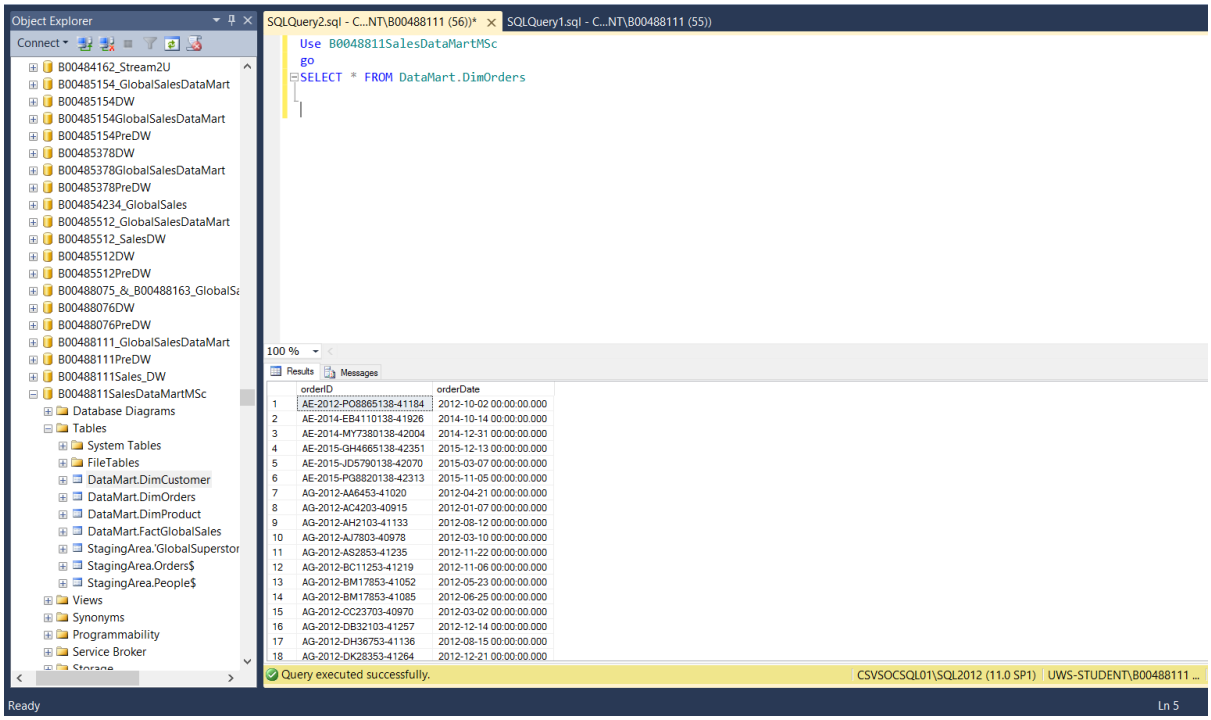


Figure 8: *Dimension Orders table details*

Facts Table: Global *Sales*

-- Create FactGlobalSales table in DataMart schema

USE B0048811SalesDataMartMSc

CREATE TABLE DataMart.FactGlobalSales (

customerID NVARCHAR(255),

productID NVARCHAR(255),

orderID NVARCHAR(255),

Sales MONEY,

Quantity FLOAT,

Discount FLOAT,

Profit MONEY,

ShippingCost FLOAT,

CONSTRAINT FK_CustomerID FOREIGN KEY (CustomerID) REFERENCES DataMart.DimCustomer(CustomerID),

CONSTRAINT FK_ProductID FOREIGN KEY (ProductID) REFERENCES DataMart.DimProduct(ProductID),

CONSTRAINT FK_OrderID FOREIGN KEY (OrderID) REFERENCES DataMart.DimOrders(OrderID)

);

GO

- View content of FactGlobalSales table

USE B0048811SalesDataMartMSc

GO

SELECT *

FROM DataMart.FactGlobalSales;

```"

**Query Results:**

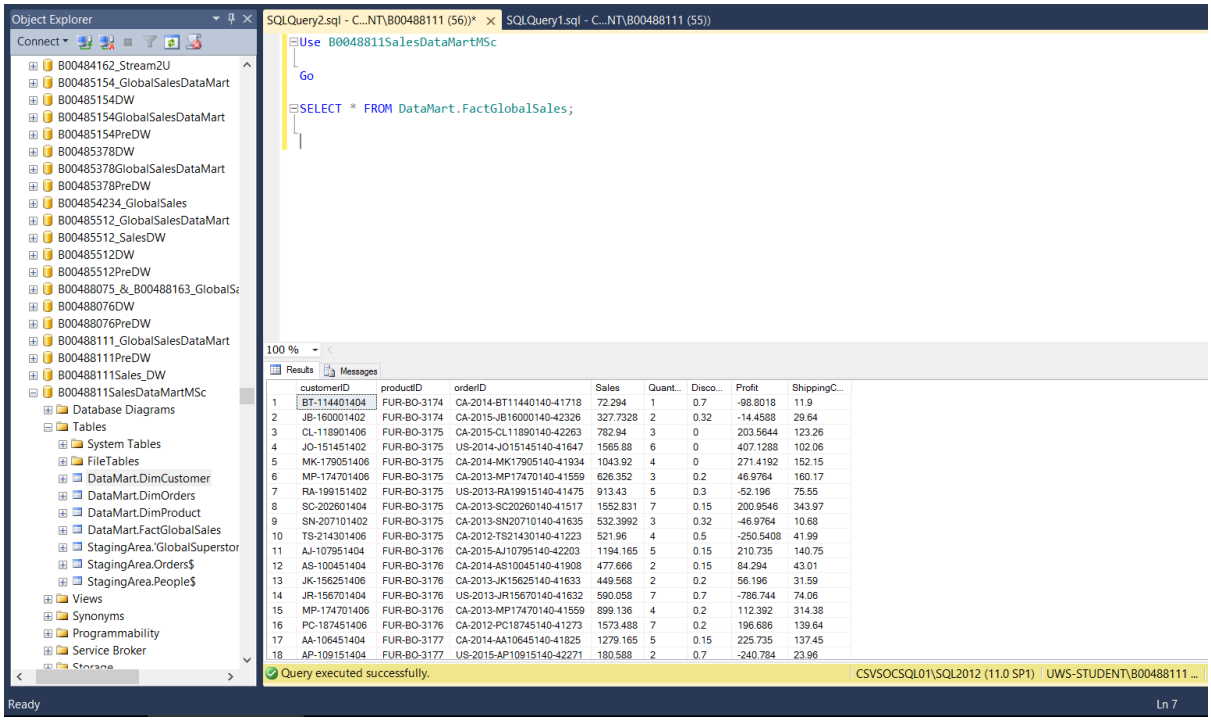



***Figure 9:*** *Facts Table Content with Extracted Data*

Final Steps: Star Schema Diagram and Data Export

To visualize the DataMart's schema and prepare data for analysis, I:
```

1. Created a Star Schema diagram

2. Exported data to an Excel file for analysis in Tableau Desktop and Power BI Desktop

Creating Star Schema Diagram

To create the diagram:

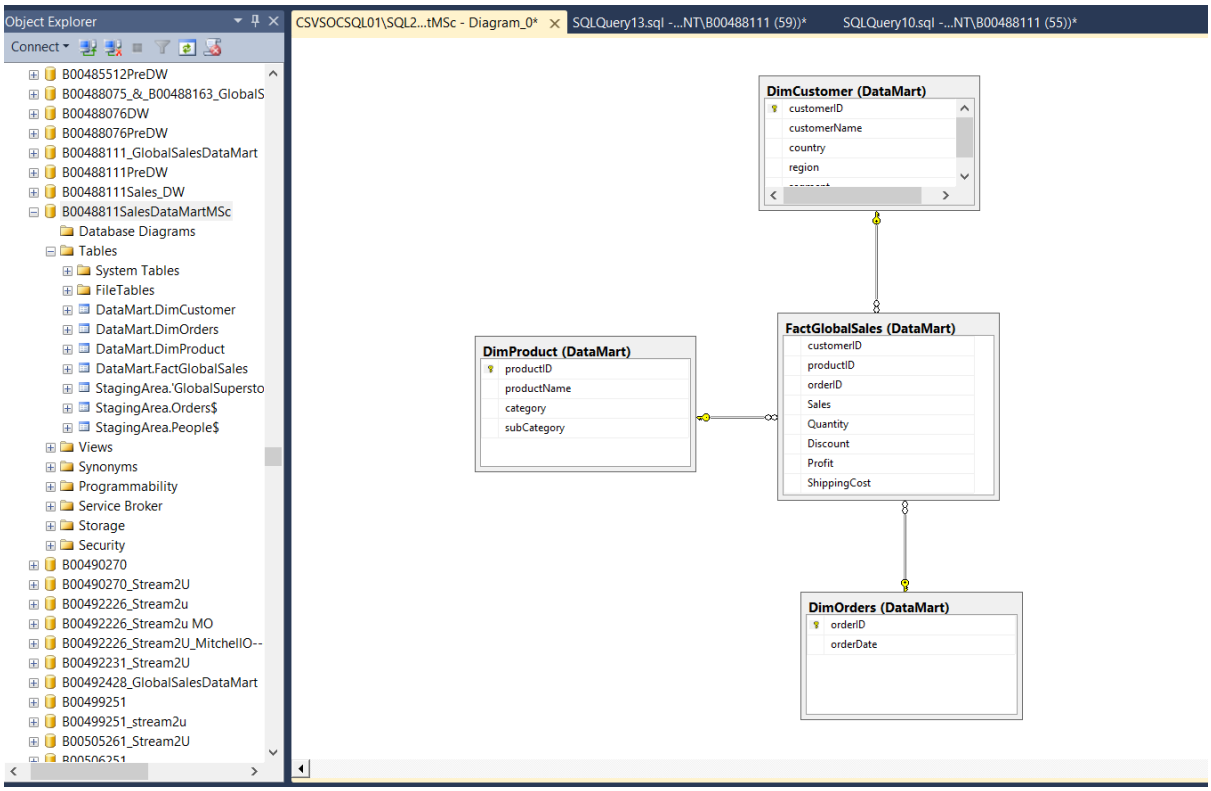
1. Selected "Database Diagram" option on the database

2. Chose "New Database Diagram" option

3. Selected required tables from the pop-up menu

4. Added referential integrity option to ensure proper key administration

Star Schema Diagram:



Exporting Data to Excel

I exported the DataMart data to an Excel file:

1. Selected the database
2. Right-clicked and chose "Tasks" > "Export Data"
3. Chose Excel file format and destination (desktop)

Data Ready for Analysis

The exported data is now ready for analysis in:

1. Tableau Desktop
2. Power BI Desktop

Next steps:

1. Connect to the Excel file in Tableau Desktop and Power BI Desktop
2. Create visualizations and dashboards to gain insights from the data

OLAP Queries and Cubes

The following OLAP (Online Analytical Processing) queries utilize relational aggregations to demonstrate rollup relationships on the newly created tables from the extracted datasets.

Query 1: Profit and Sales Analysis

```
SELECT
    C.CustomerName,
    SUM(F.Sales) AS TotalSales,
    SUM(F.Profit) AS TotalProfit
FROM
```

DataMart.FactGlobalSales F

INNER JOIN

DataMart.DimCustomer C ON F.CustomerID = C.CustomerID

GROUP BY

C.CustomerName

WITH ROLLUP

This query:

- Joins FactGlobalSales and DimCustomer tables
- Groups sales and profit data by customer name
- Uses ROLLUP to display subtotals and grand totals

OLAP Concepts

OLAP queries utilize various dimensions to analyze data:

- ROLLUP: aggregates data to display subtotals and grand totals
- CUBES: multidimensional data structures for analysis
- SLICING: creates a new cube by selecting specific dimensions
- DICING: creates a sub-cube by filtering specific dimensions

Benefits

OLAP queries enable:

- Multidimensional analytical queries
- Financial reporting and forecasting
- Data analysis and visualization

OLAP Query: Profit and Sales Analysis by Customer

```
USE B0048811SalesDataMartMSc
GO

SELECT
   F.Profit,
   F.Sales
FROM
   DataMart.DimCustomer C
INNER JOIN
  DataMart.FactGlobalSales F ON C.customerID = F.customerID
GROUP BY
CUBE (F.Profit, F.Sales);
```

Query Explanation:

This OLAP query:

- Joins DimCustomer and FactGlobalSales tables
- Groups profit and sales data by customer using CUBE aggregation
- Displays aggregated values for profit and sales

CUBE Aggregation:

The CUBE aggregation generates all possible combinations of the specified columns, including:

- Total profit and sales
- Profit by customer

- Sales by customer
- Grand total

Benefits:

This query enables:

- Multidimensional analysis of profit and sales
- Customer-level insights
- Rolled-up aggregations