**SEIS630 Database Final Project, Part 2**

[**Super Mart Grocery Sales System Database**](https://uofstthomasmn-my.sharepoint.com/:f:/g/personal/verm5201_stthomas_edu/EpSMQIjlljZBumuIQNHtlMkBxtSwZbOUv5Yb0KKhx3OmTQ?e=EIN8ko)

**Pragya Verma - 101276204**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Introduction:**

This project is an effort to create a database schema for the Super Mart supermarket sales management system, which will result in an error-free, secure, informative, and speedy management system. This database is effective in running supermarket grocery stores, and the database's users will be store staff. The project also expands to include the creation of a dashboard for further investigation.

The project will deal with the generation of normalized tables as well as the data importation into them. Product, Category, Customer, Order, OrderDetail, and Branch are likely to be the tables.

The project was chosen because I am interested in grocery sales management systems for future company ideas, and it will offer me a good knowledge of how the system works inside and out, as well as the challenges I may have in the future and how I may fix them. This project also allows me to gain hands-on experience with many tools, including MS SQL Server, MS SQL Server Management Studio, Oracle Data Modeler, and tableau.

**a) Exploring:**

1 - Collect and explore the Super Mart supermarket sales dataset.

2 - Designing the Super Mart grocery sales database system using the E/R approach over SQL Server Management Studio.

3 - Logical and Relational diagrams will be designed using oracle Data Modeler.

4 - Go through the normalization process to create a set of tables.

5 - SQL Server database will be used to create and maintain the database schema and normalized tables.

6 - The population of the database from CSV to SQL Server table can be done in bulk upload/ import option using SQL Server Management Studio or by SQL Server insert statement.

7- Formulate SQL queries for basic and deep searches.

8 - Constraints and relationships will be clearly defined.

9 - Executing those queries using SQL Server.

10 - Connect the database to any data visualization tool – Tableau

**b) Discovering:**

***1) Data collection***

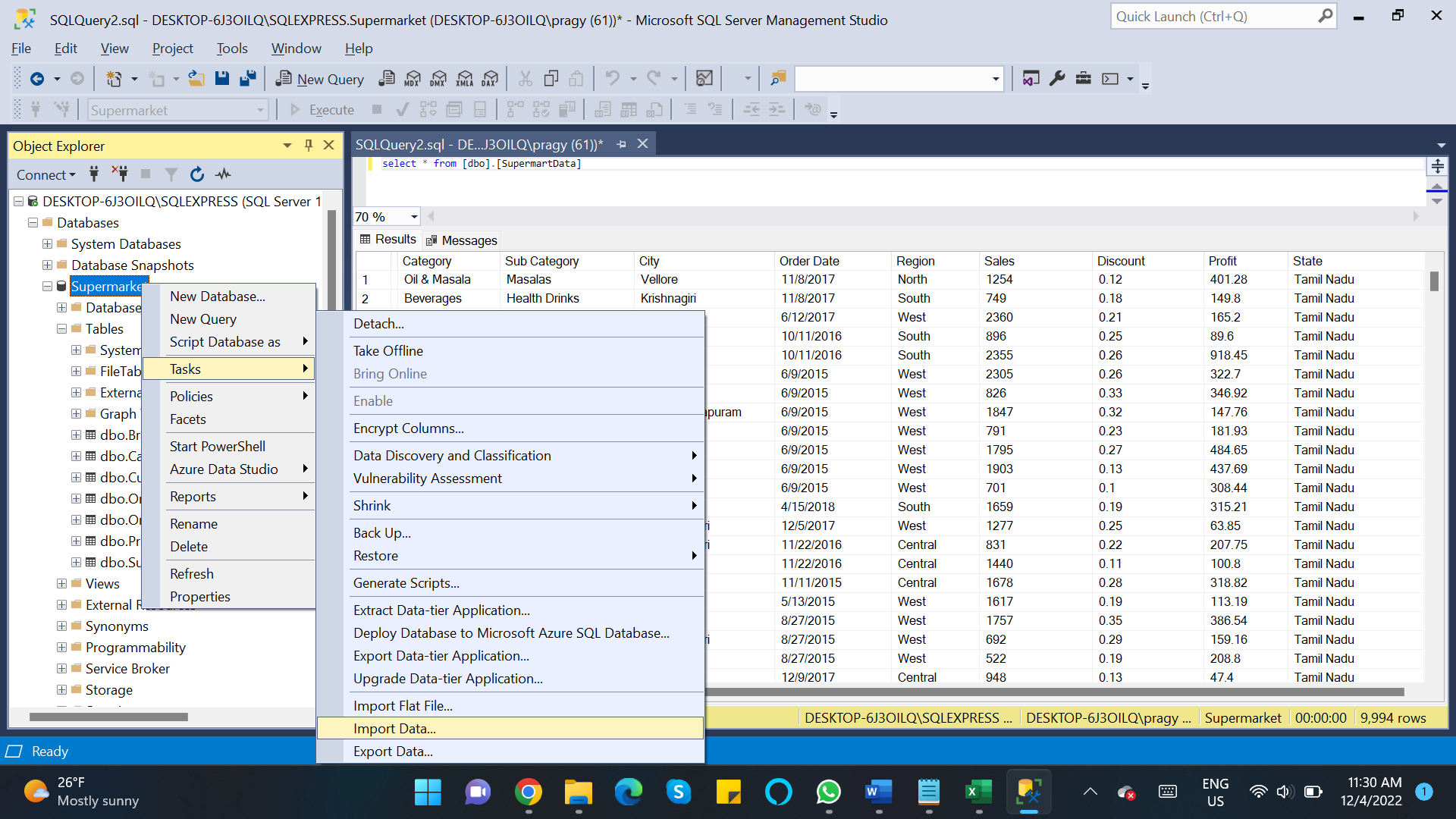
The data was gathered from Kaggle's Open dataset. This is an open dataset that may be utilized for further analysis. The dataset’s content is as follows:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Order ID** | **Customer Name** | **Category** | **Sub Category** | **City** | **Order Date** | **Region** |
| **Sales** | **Discount** | **Profit** | **State** |  |  |  |

These columns can be separated into six tables. This column partitioning is done to avoid data redundancy and eliminate anomalies produced by errors during Insertions, Updates, and Deletions. The goal is to organize and normalize the dataset.

Below is the table structure:

|  |  |
| --- | --- |
| **Product** | ProductId, ProductName, CategoryId |
| **Category** | CategoryId, CategoryName |
| **Customer** | CustomerId, CustomerName |
| **Order** | OrderId, OrderName, CustomerId, BranchId, OrderDate, Sales, Discount, Profit |
| **OrderDetail** | OrderDetailId, OrderId, ProductId |
| **Branch** | BranchId, Region, State, City |

 **c) Building:**

* I have used SSMS database to create the tables and upload the data into these six tables.
* For creating the ER diagram, I have used SQL Server Management Studio Database Diagram option.
* Logical and Relational diagrams will be designed using oracle Data Modeler
* For writing the DDL (Create and Alter Table) and DML (Insert, Update, Delete) queries to enter the data in the table, I have used SSMS.
* For data analysis of the dataset, I have used Tableau. Tableau can be connected with the SQL Server Management Studio and data tables can be used for data analysis.

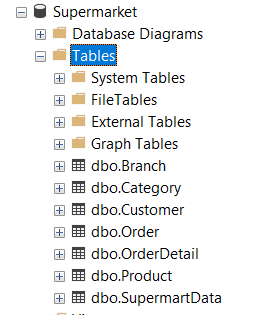
**Learnings:**

The top learnings applied while doing and completing this project are as follows:

**1) Normalization:**

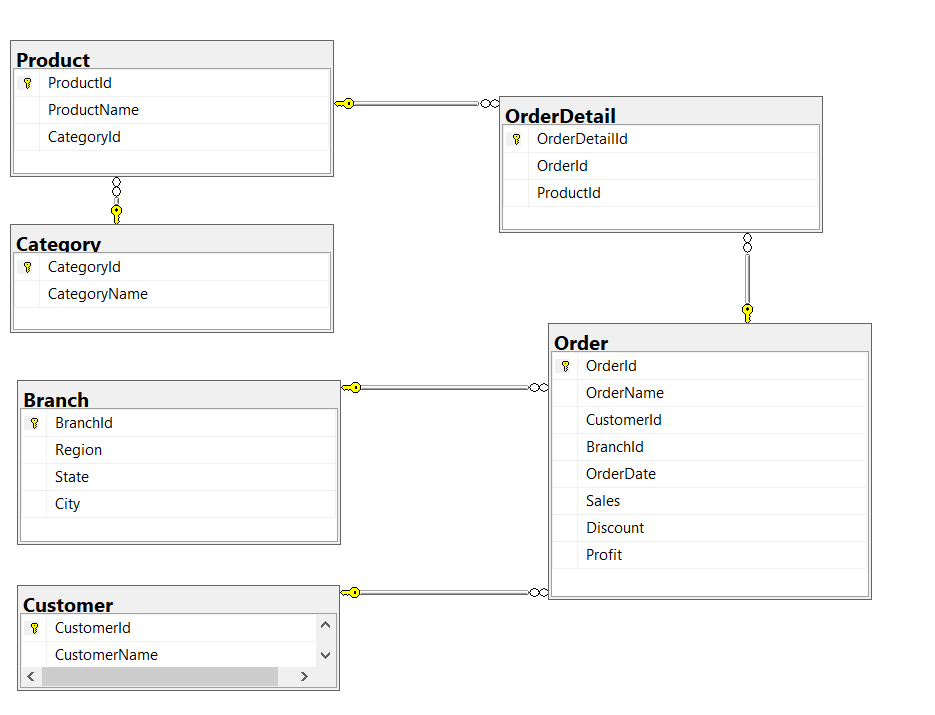
I have a large dataset that contains a lot of redundancy. I generated 6 Normalized tables, loaded data from the Supermarket dataset into those 6 tables, and created the database ER diagram.

By dividing the dataset into six tables, I removed redundancy and made the dataset more consistent for further updates.



**2 - ER Diagram**

I have created the ER modeling of the database table structure below which shows their mapping keys and relationship.



**3) Data Modeling and DDL commands and adding Constraints:**

I generated the data model with SSMS and wrote SQL's Data Definition Language commands like **CREATE, ALTER table and TTRUNCATE** that we learned in class. I also used Constraints like PRIMARY Key and FOREIGN Key.

***Create table Queries***

*CREATE TABLE [dbo].[Product](*

*[ProductId] [int] IDENTITY(1,1) NOT NULL,*

*[ProductName] [varchar](255) NULL,*

*[CategoryId] [int] NOT NULL,*

*CONSTRAINT [PK\_Product] PRIMARY KEY CLUSTERED*

*(*

*[ProductId] ASC*

*)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, IGNORE\_DUP\_KEY = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [PRIMARY]*

*) ON [PRIMARY]*

*ALTER TABLE [dbo].[Product] WITH CHECK ADD CONSTRAINT [FK\_Product\_Category] FOREIGN KEY([CategoryId])*

*REFERENCES [dbo].[Category] ([CategoryId])*

*ALTER TABLE [dbo].[Product] CHECK CONSTRAINT [FK\_Product\_Category]*

**4 - DML & DQL commands:**

Once the tables are created, we must populate the tables with the data. I have used DML commands like **INSERT, and UPDATE** commands to populate and update the data in tables.

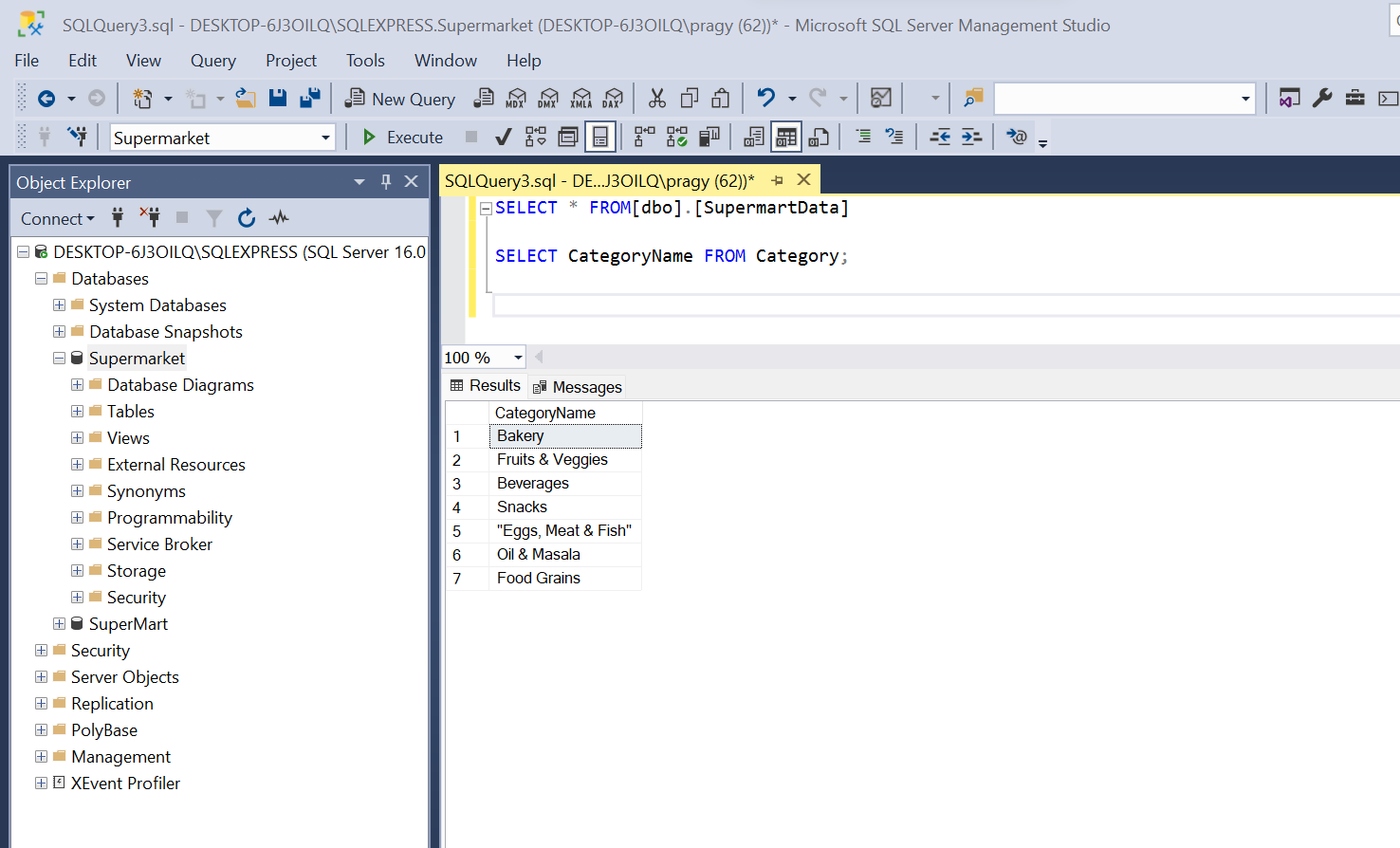
I instered unique values of "Category" column from SupermartData table into Category table.

*INSERT INTO Category (CategoryName)*

*SELECT DISTINCT Category*

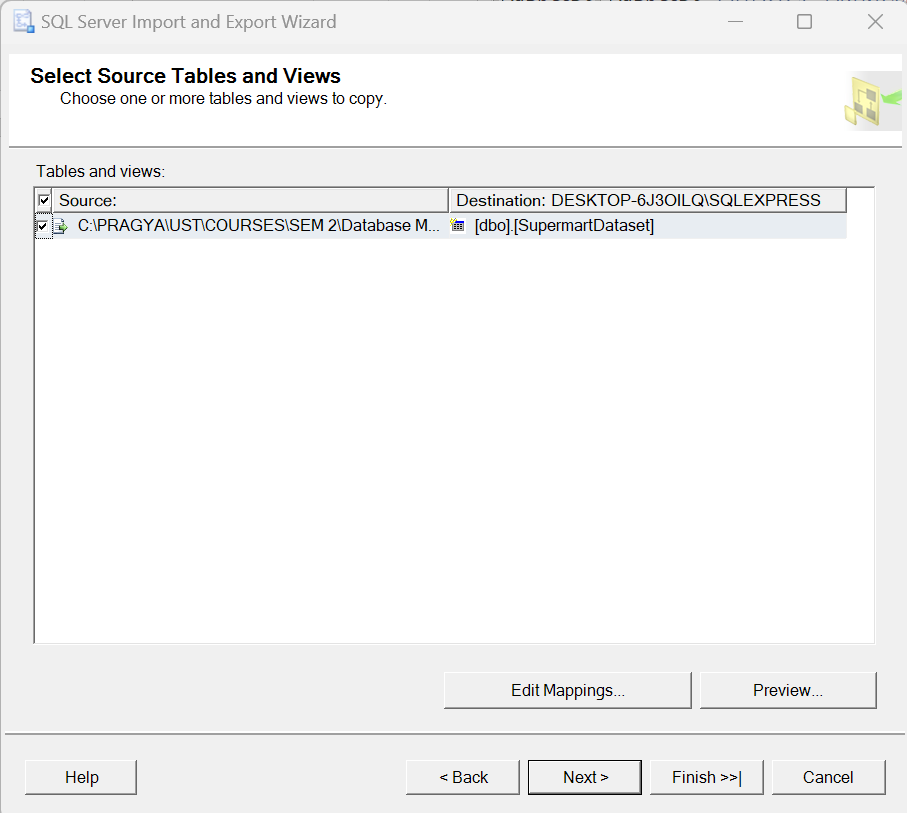
*FROM SupermartData*

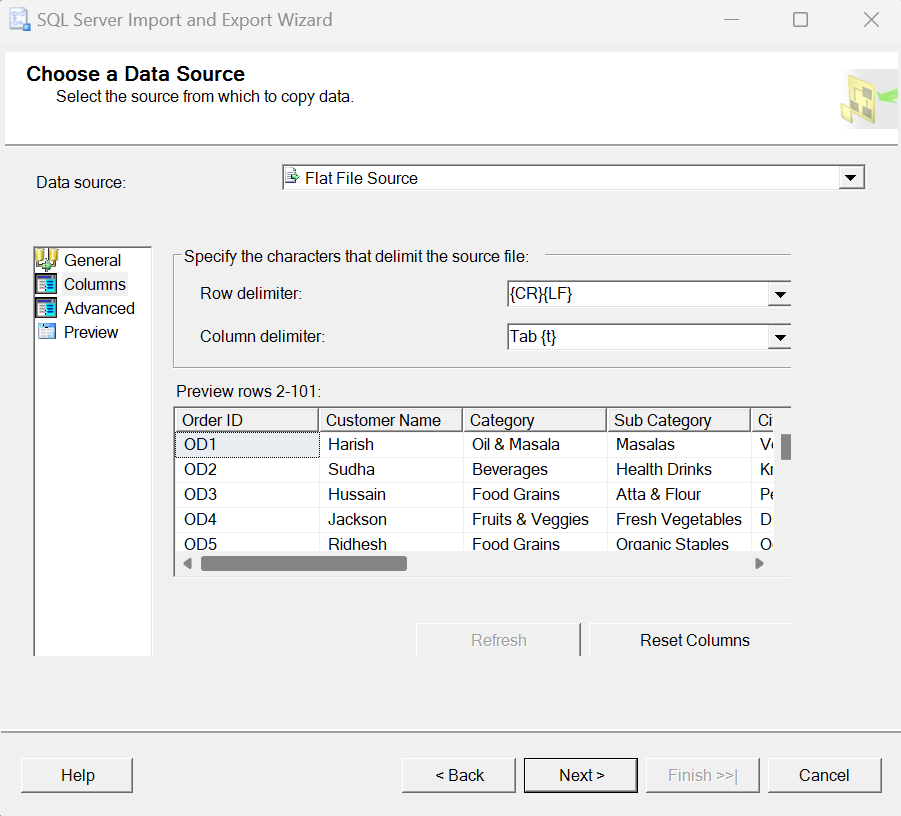
After the data is populated, we must query the data using DQL commands like SELECT clauses.

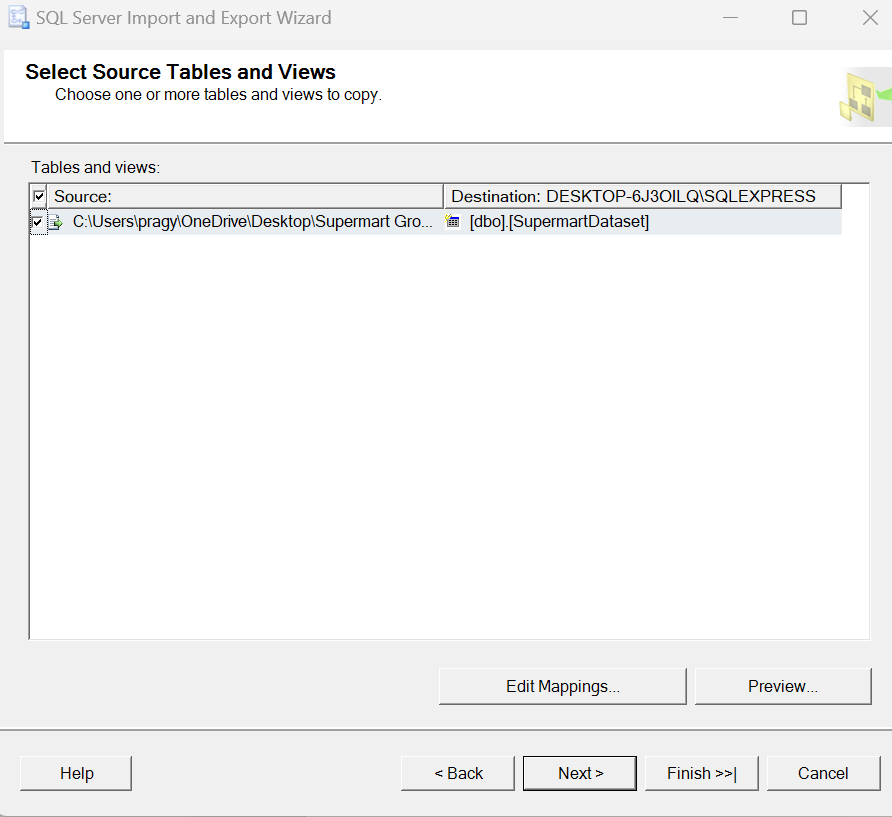


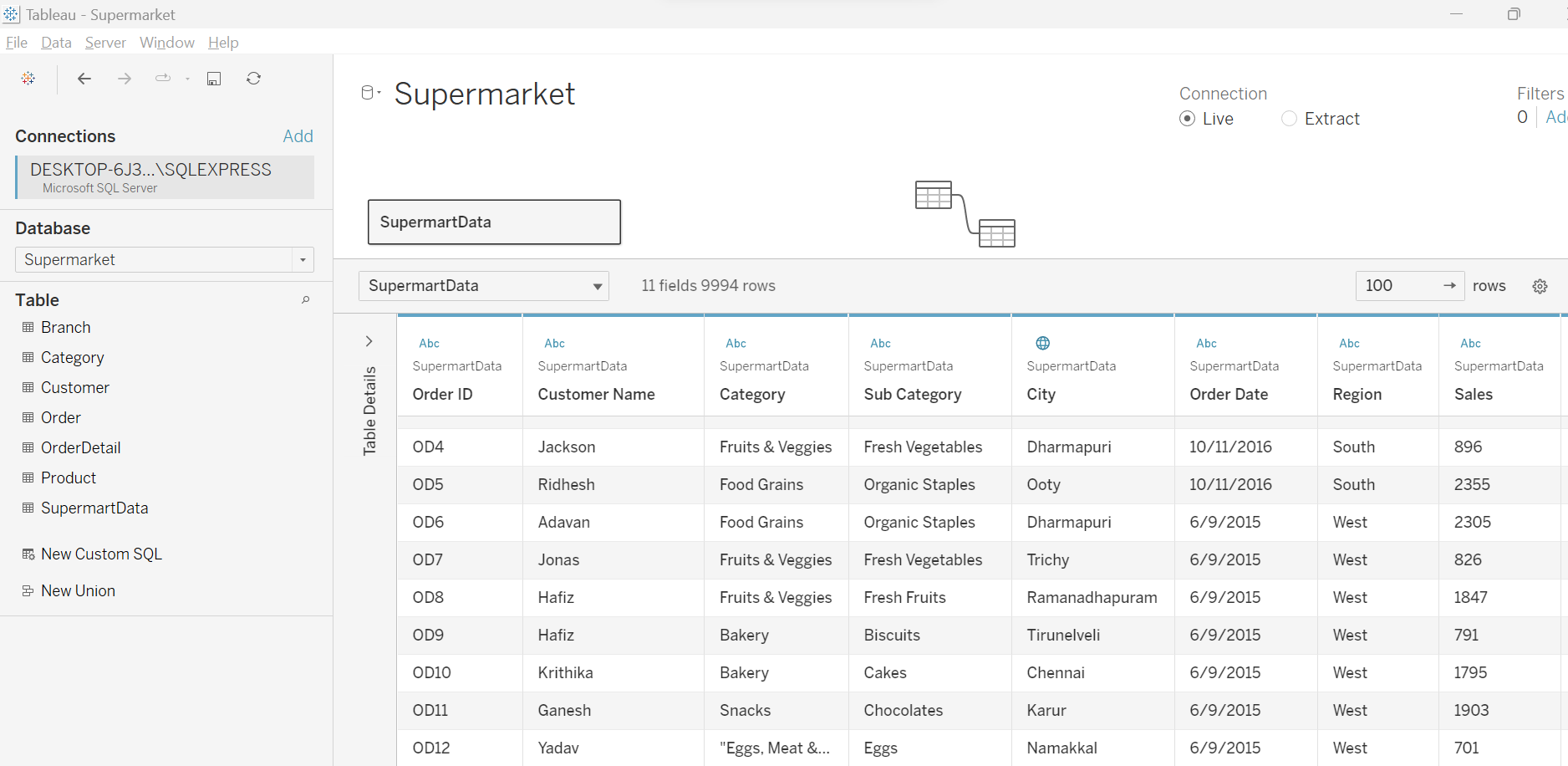
**5) Databases connection with Business Intelligence Tool**

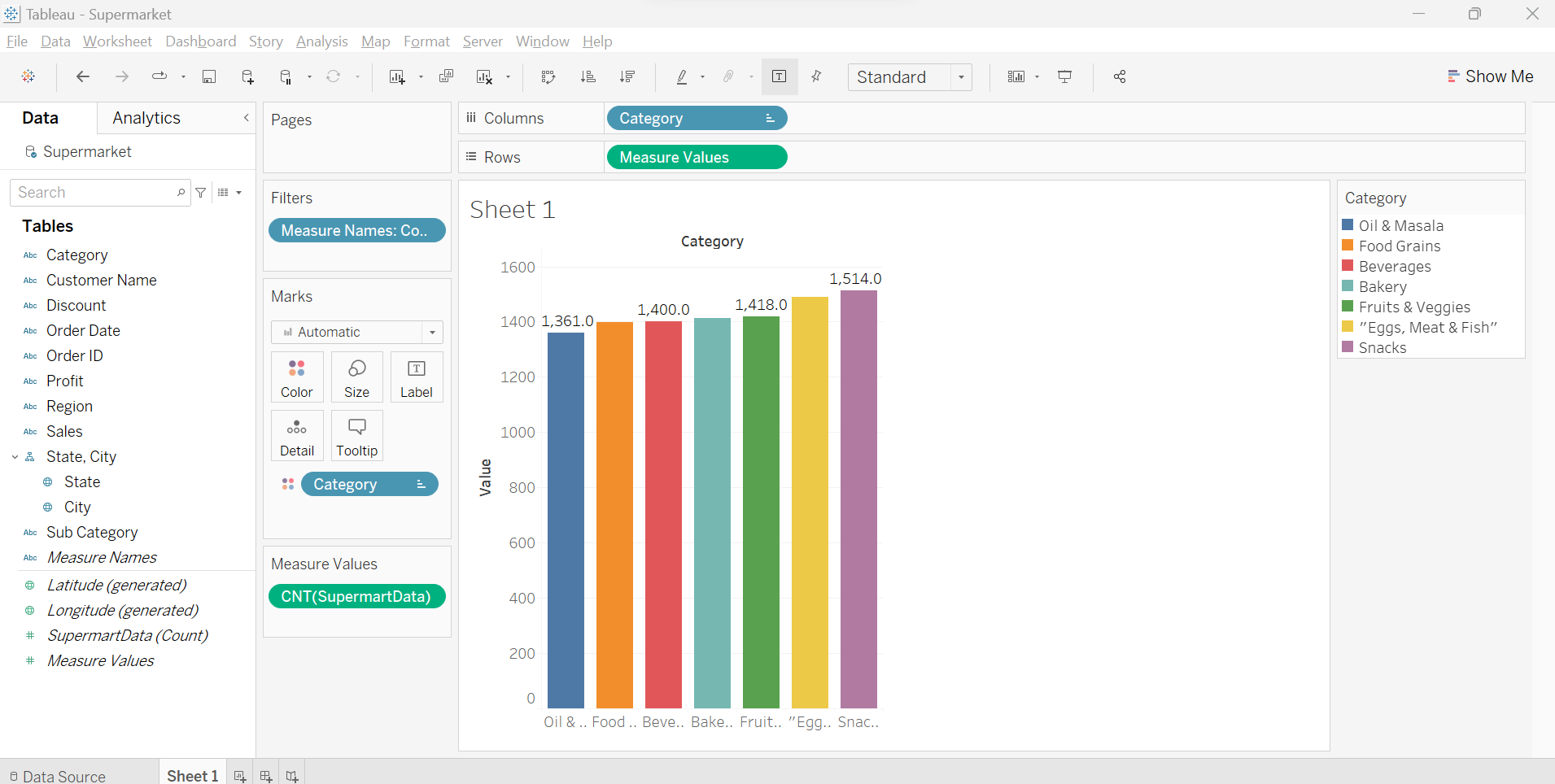
Now we have learned that we can connect external data visualization and business intelligence (BI) tools to our database and use them to further manage, visualize, and analyze data. In this project, I have connected the SSMS database with Tableau.











**Correctness and completeness.**

In the class, we studied Data modeling, DDL, DML, Normalization of the data and ER diagrams. We studied the connection of the database with third-party tools. So, as per my understanding, this project is fulfilling all the required tasks to work on and is correct.

**Conclusion:**

Database Management and Design is an interesting course. Though I had never worked with databases previously, my prior experience was restricted to using Excel and generating reports using the Excel functions. After this course, my database knowledge and skills improved significantly after taking this course.

In this class, I learned about the topics like Database tables, the relationship between the tables, Database queries, Normalization, Data Modelling, ER diagrams, and now I have a better understanding of these topics which I can implement in my future career goals. Also, the advanced topics discussed in the class like Django, Snowflake, Data Warehouses, Business Intelligence, and Big Data were really new to me and I believe they would surely help me in my knowledge building and its utilization in my further learning in the Data Science domain.

In the class, we covered Oracle Cloud platform and I feel that if had more time, I would have created databases over the cloud.

Finally, I would like to express my gratitude to Professor Abe for allowing us to work on projects as per our choice. His knowledge of databases is tremendous and he inspires others in this field.

**Note:**

I am still working on the project but still have to make some changes:

1 - I have to make the tables more normalized which I feel that I need a little more time to work on.

2 – Need to work on the logical and relational diagram on Oracle Modeler to have clarity on the types of relationships between tables. The SSMS Draw diagram function does not show the types of relationships between the tables.