Dominick's

DESIGN AND IMPLEMENTATION OF A DATA WAREHOUSE FOR A RETAIL STORE

Final Report

Submitted By
Arpita Deshmukh, Shilpa Chanshetti, Shruthi Shetty

ISTM 637 – 601

Access Credentials

I. SQL Server Login Credentials

User Name
Password:

II. Database Engine in the Microsoft SQL Server Management Studio
   a. Staging Database: 601Group5-staging-area
   b. Warehouse Database: 601Group5-dw-area

III. Analysis Services of the SQL Serve
   a. 601Group5-dw-area

1. Introduction

A Data Warehouse is basically a large storage of data which stores data integrated from various sources. It brings all the data in one place which also includes the historical data that helps in easy access to it in the future. Furthermore, it supports organizations in decision making & business analytics. By transforming the data and performing meaningful operations on it would help in generating insightful results.

Huge companies have multiple offices at various locations. Each of its branches generates a huge amount of data. This data is stored in the Data Warehouse. It helps the decision makers to produce a lucrative decision for the business. Similarly, retail stores collect a vast amount of data by selling products. There are a number of parameters that could affect the revenue of any store. To increase the sales, businesses perform analysis on the data to find the to the challenges faced by the organization.

The dataset that we have considered is Dominick's dataset. Dominick's Finer Foods (DFF) and Chicago Booth entered into a partnership in 1989 to 1994 for research in shelf management and pricing. However, the store was shut down due to low performance. Dominick's dataset consists of data over a period of about nine years which includes demographic data, data about different products, the price of the products, details about promotions etc. We would be analyzing this dataset to solve a few problems and provide discerning solutions.

2. Data Description
2.1. Understanding the Data

The Dominick's data contains the store-level scanner data that was collected for a span of over 7 years. This sales data is of Dominick Finer Foods across the United States almost 100 stores for approximately 3500 UPCs categorized into 29 different categories. Entire research data can be broadly divided into two different categories: General files and category specific files. The general files contain detailed data related to each category. The general files contain two parts: Customer-oriented data and demographics related data. The category files are also divided into two parts: UPC files and the Movement data.

The nature of the data is clear and be explained as follows: The data available with us contains a lot of missing values and is dirty data. Hence, it requires a lot of cleaning and manipulation before analysis. But we can clearly understand that the daily purchase data is recorded across all the DFF stores which helps understand the purchasing trend of the customers. The customer related data is basically divided into various categories so that it helps to understand the different customers based on the demographics. The data manual present here helps to understand the data in detail.

2.2.Metadata for all OLTP source files

Customer Count File:

The customer count file is about in-store traffic. The data informs about the visits by the customer and the daily in-store purchases for every store. This also tells about the total product sales, coupons redeemed for every product/transaction, date of the transaction, etc.

Store demographics File:

This contains store specific demographic data. The data is originated from the US govt. census data conducted for the Chicago metropolitan area. This data is processed to get demographics profile for each store of DFF. The data informs about the different age groups of customers, their household information, their income, their dependents in the house, information about the employment, etc. This helps to make a decision regarding the customer demographics to increase the sales by targeting customers belonging to a particular group and strategize to increase the profit for a particular product. The demographics file contains the following variables:

Category files - UPC File:

The category file has the list of all UPCs and their details regarding the size of each category. The information consists of the product name, description, size, number of items, UPC number, etc. The naming convention of the file is as "upcxxx", xxx is the three-letter acronym for the category. For example, "upcber" is the UPC file for beer. It will contain different types of beer. The variables of UPC files are as follows:

Movement Files- movement data by UPC:

The data in the movement files is about the weekly sales of the products which are of different categories. This data is for each UPC category. This information helps answer business questions related to holiday sales, demographic-level sales, weekly sales, etc. The information contained includes product price, profit, units sold, etc. The naming convention of the file is as "wxxx", xxx is the three-letter acronym for the category. For example, "wber" is the movement file for beer. It will contain all the sales data for different types of beer. The variables of movement files are as follows:

Week Decode Table:

This table helps get information about the week when the sale was recorded. This contains information about the dates, the week number they belong to and if it's a regular week or a holiday week. This data is helpful in identifying the trends in the regular weeks of the year and the holiday weeks. While identifying the trend, annually, semiannually, quarterly, etc, this table is useful.

2.3. ER Diagram
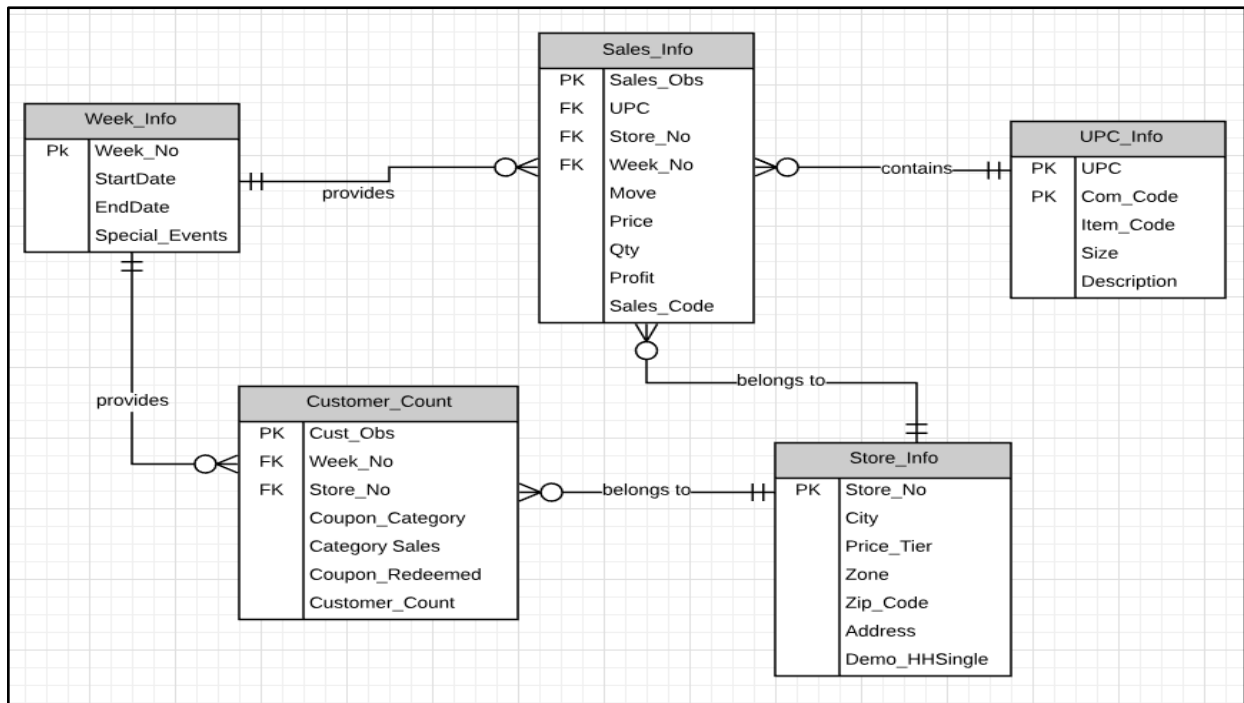
2.4. Domain Understanding

Jami, Ata, and Himanshu Mishra. "Downsizing and Supersizing: How Changes in Product Attributes Influence Consumer Preferences." Journal of Behavioral Decision Making (2013) talks about how a customer's behavior and judgments towards a product is influenced by changes in one or more attributes of the product while keeping others constant. There are many reasons for product attributes to changes. Some of them include, reduce the packaging size to offset any price increases due to the economic recession or to combine products and sell them as a 'bonus buy' to boost sales.

The paper demonstrates through six studies on how a change in one product attribute can change customer's preference towards an option that provides the highest possible change when that change is required by the customer, while the customer chooses the lowest possible change when the change is not required by the customer. For Example, in one of the studies, they examined the sales data of Kellog's Raisin Bran before and after the price changes were made. The results showed that the large packets of Kellog's sold less in the week after the price increase than in the week before that. This showed that price changes for the same product but for different sizes made a customer choose an option with a lower price increase or the option with a higher price reduction.

This research is helpful for marketers to analyze and predict consumer behavior to make any consumer-related decisions. Some of the decisions can be in the form of predicting the overall stock of the product in order to save money on storage. This is also helpful for producers and retailers to implement the right attribute changes so that it can be altered to the customer's expectations.

Nevo, Aviv and Hatzitaskos, Konstantinos, "Why Does the Average Price of Tuna Fall during Lent?" (2005) highlights the impact of holidays and seasonal demand for commodities. Whenever there are some festival and holidays, a huge change in the shopping pattern of all the customers is observed. There is a high demand for various types of things available. This may be due to sale, promotions, discounts, etc. A temporary rise in the sales of a few products is observed for that period. For example, turkey during Thanksgiving, cakes during Christmas, new year, etc. Such changes and variations need to be addressed and decisions should be taken accordingly.

The paper showcased the author's understanding of how a holiday period might affect the sales of the various product. The report also understands the sales pattern of the products over a period of time as per the occasion. This helps in getting information regarding such seasonal raises and what times of the stores would be busy.

This research clarifies how different holiday weeks might affect the sales and how can we take advantage of those weeks to increase the profit margin by a significant number. This in turn prepares in advance to be ready to manage the stocks and make sure we are not running out of anything and promotions of the products are in place.

The paper "Why do Manufacturers issue coupons" - focuses on the effects of coupons on sales growth. The topic of discussion in this research paper is how the sales may increase or decrease in the retail chains due to the usage of coupons and how lagged coupons correlate with the current sales. This correlation is a positive one which induces the customers to repurchase the product. It claims that spending money on coupons would be more advantageous than spending money on advertising. This paper discusses different reasons behind why coupons are introduced in any product.

First, the introduction of coupons on a new product makes the customer aware of the launch of the new product. Secondly, coupons help in increasing sales. Introducing coupons on the products with the highest sales as well as on the lowest sales help in boosting the sales of a particular product. We have captured this concept in our project as well. Finally, the third reason for introducing the coupons is that continuous application of coupons on the products changes the perception of the customers.

3. Business Question

The following business questions were studied. Out of which the top 5 were selected for creation of the data warehouse:

_BQ1_ : Analyze the average profit margin of Bath Tissues across all the stores to determine store with low-profit margin

_BQ2:_ What are the total sales of Beer during the holiday season?

_BQ3:_ Which pricing tier produces the highest Bottled Juice sales during Christmas Week?

_BQ4:_ Which Bath Soap product has a decreasing sales trend across different zones?

_BQ5:_ What category of coupons are redeemed the most for the entire duration across all the stores?

_BQ6:_ What are the trends in the sale of perishable items during the holiday weeks?

_BQ7:_ What are the sales of frozen food among the singles and working population across different regions?

_BQ8:_ How does bonus buy and Price Reduction work for Bottled Juices in different zones?

_BQ9:_ Which brand of Analgesics has the greatest sales over the last 1 month?

_BQ10:_ Which laundry detergent has the lowest sales for a store in the last 3 months?

On analyzing & prioritizing all the business questions, BQ1 through BQ5 have been selected for further analysis to create a data warehouse using the Kimball's approach

BQ1. Analyze the average profit margin of Bath Tissues across all the stores to determine store with low-profit margin
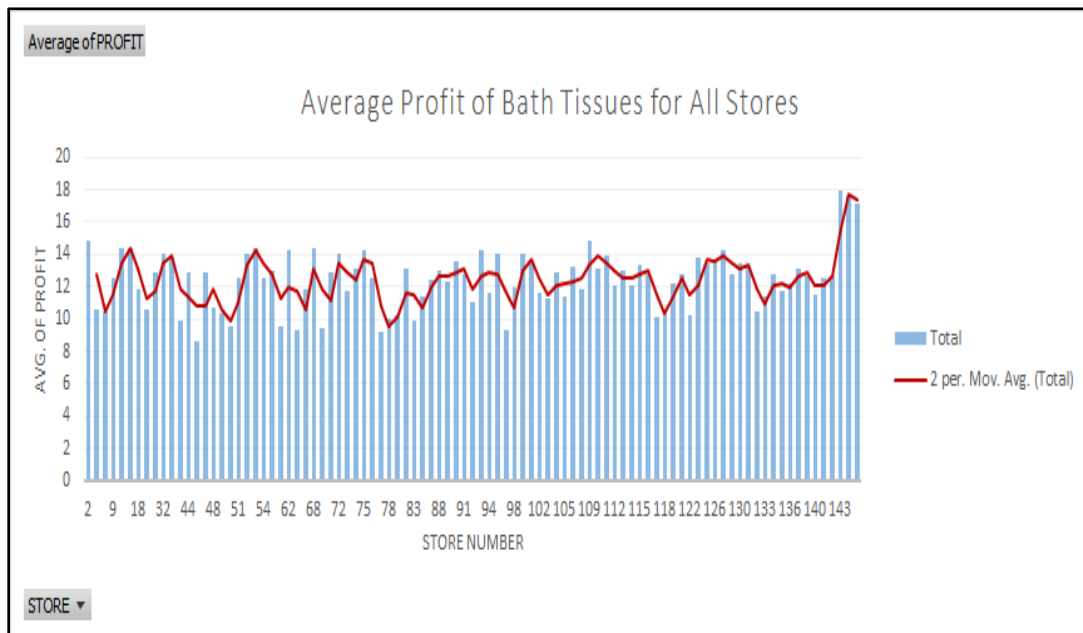
BQ2. What are the total sales of Beer during the holiday season?
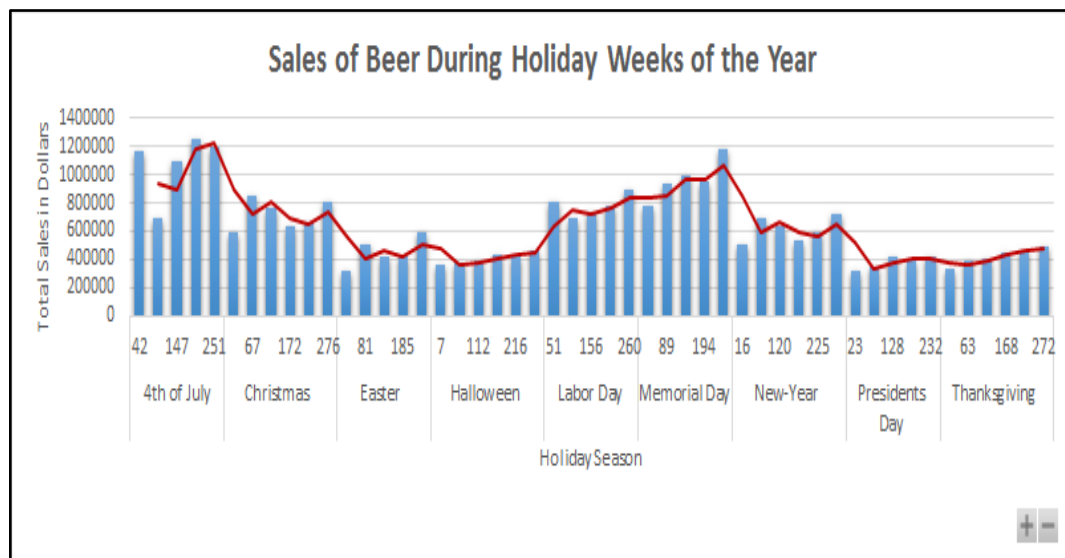
BQ3. Which pricing tier produces the highest bottled juice sales during Christmas Week?
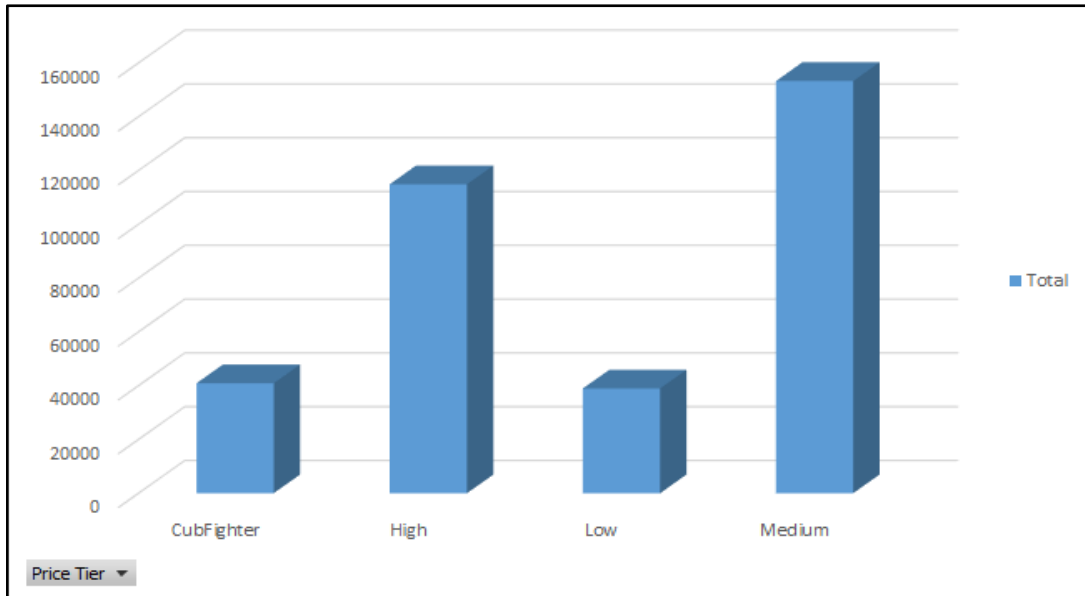
BQ4.Which Bath Soap product has a decreasing sales trend across different zones?
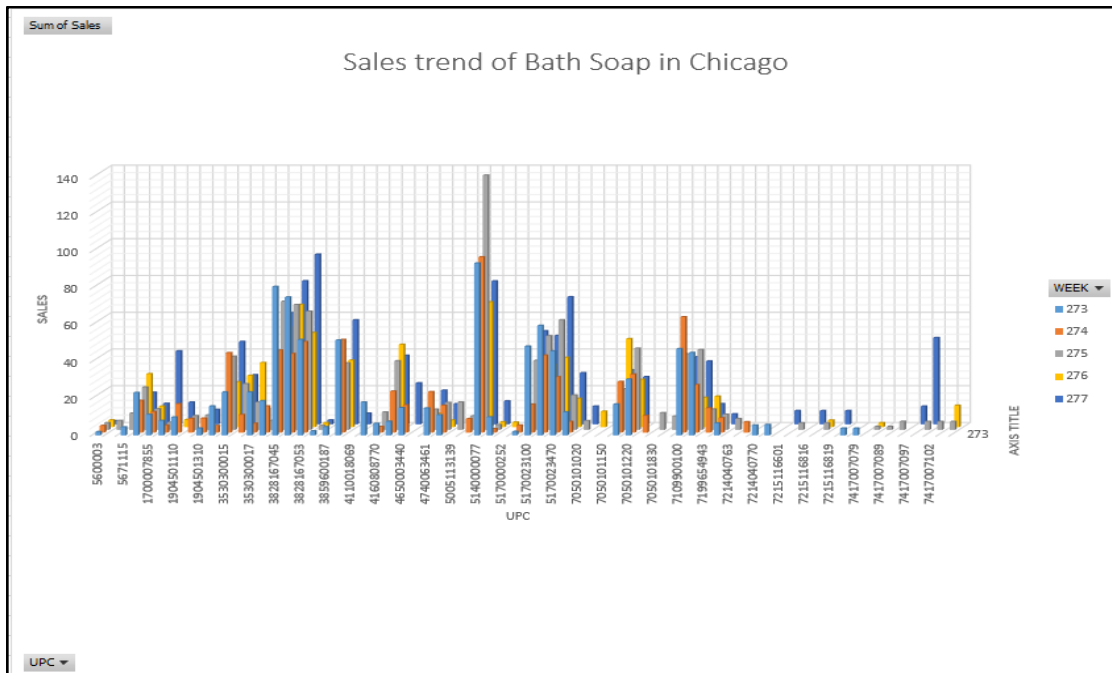
BQ5. What category of coupons are redeemed the most for the entire duration?
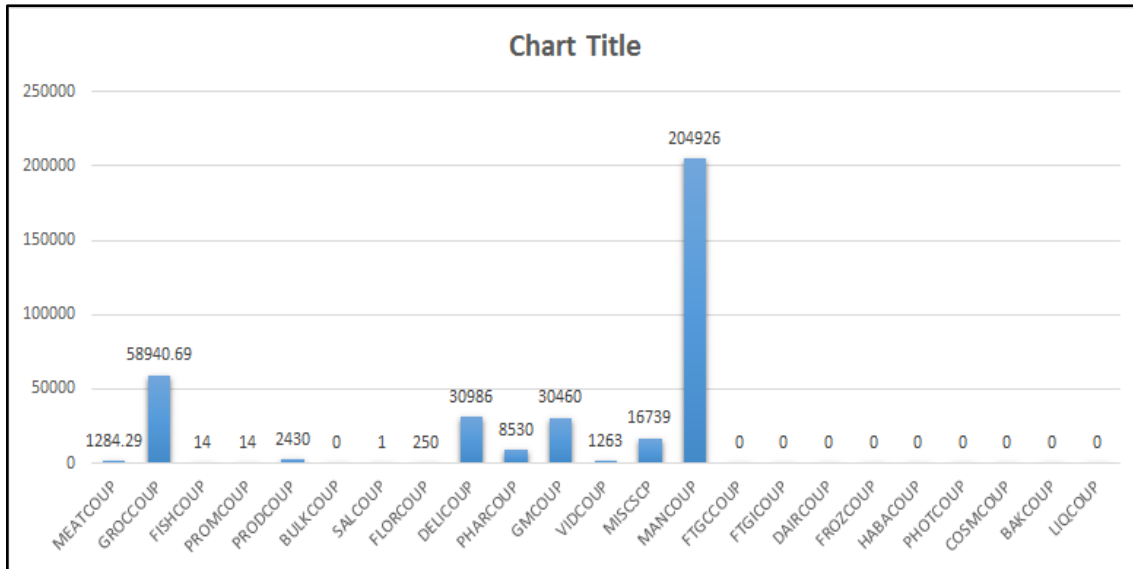
4. Independent Data Mart Design Using Kimball's Approach

4.1. Logical Design

Dimensional modeling developed by Ralph Kimball consist of set of techniques and methods which are used in designing of the data warehouse. This methodology focuses on the bottom-up approach i.e. implementing the data mart and later integrating these to create a data warehouse using a bus architecture consisting of conformed dimensions between the data marts.

Ralph Kimball states that the data warehouse should be created using a dimensional model (Star or snowflake schema). In dimensional approach, that data is partitioned into "facts" or "dimensions". Facts are generally numeric data and dimensions contain the reference information which provides context to the facts.

Dimensional modeling technique has been used to create STAR schema for designing the data warehouse for DFF. In this section, we have described various dimension tables and fact tables which have been created to address the business questions.

4.1.1. Dimension Tables

We have created four dimension tables: Product Dimension, Store Dimension, Category Dimension, and Time Dimension. Following is the detailed description of the tables.
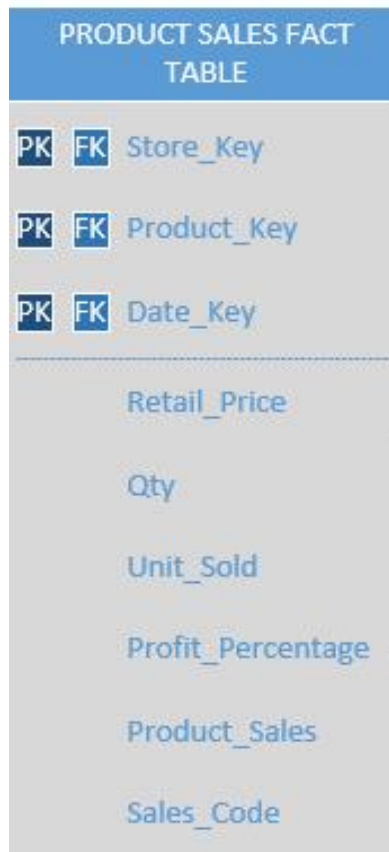
Product Dimension

The attributes of the product dimension table are as follows:

- Product_Key – This attribute is the surrogate key i.e. a unique identifier of the product dimension table
- Product Description – Description of the product available in the retail store
- UPC_No – The last five digits of the UPC Number identify the product, the remaining digits identify the manufacturer
- Category_Name – Name of the category defined in the data set

Store Dimension

The attributes of the store dimension table are as follows:

- Store_Key – This attribute is the surrogate key i.e. a unique identifier of the store dimension table
- Store_No – This attribute represents the numbers assigned to each store
- City – The city where the store is located
- Price_Tier – The price tier of a particular zone where the store is located
- Zone – This attribute is the identifier of the zone to which a store belongs

Promotion Dimension

The attributes of the promotional dimension table are as follows:

- Promotional_Key - This attribute is the surrogate key i.e. a unique identifier of the Promotion dimension table
- Coupon_Category_Name – Coupons redeemed corresponding to the category name

Date Dimension

The attributes of the date dimension table are as follows:

- Date_Key - This attribute is the surrogate key i.e. a unique identifier of the date dimension table
- Week_No – The identifier for the week in a particular year when sales occurred
- Month - Month of record of sales
- Year – Year when the sales were recorded
- Special_Event_Name – Name of the event during which the sales were recorded

4.1.2. Fact Tables

We have created 2 Fact tables called *Product Sales Fact Table* and *Promotional Sales Fact Table* to support our business questions. Below is a detailed description for the same.

Product Sales Fact Table

PRODUCT SALES FACT TABLE

PK FK Store_Key

PK FK Product_Key

PK FK Date_Key

Retail_Price

Qty

Unit_Sold

Profit_Percentage

Product_Sales

Sales_Code

Product Sales Fact Table

Keys:

- Store_Key - This attribute is the surrogate key i.e. a unique identifier for the stores which help in diving the data store-wise
- Product_Key - This attribute is the surrogate key i.e. a unique identifier for the products
- Date_Key - This attribute is the surrogate key i.e. a unique identifier for date

Measures:

- Retail_Price - Unit price of a batch of a particular product
- Qty – Size of that batch of a particular product

- Unit_Sold – Number of units sold for a particular product
- Profit_Percentage – Profit percent made by DFF on sale of the product

- Product_Sales – It represents the dollar value of the product sales
- *Product_Sales = (Retail_Price * Unit_Sold)/ Quantity*
- Sales_Code – This attribute indicates if the product was sold on promotion

PROMOTIONAL SALES FACT TABLE
PK FK Store_Key
PK FK Promotional_Key
PK FK Date_Key
Coupons_Redeemed

Promotional Sales Fact Table

Keys:

- Store_Key - This attribute is the surrogate key i.e. a unique identifier for the stores which help in diving the data store-wise
- Promotional_Key - This attribute is the surrogate key i.e. a unique identifier for promotions
- Date_Key - This attribute is the surrogate key i.e. a unique identifier for date

Measures:

- Coupon_Redeemed – This attribute indicates the sales by coupons

4.2. Dimension Matrix for Data Marts

Dimensions Data Marts	Product Dim	Store Dim	Date Dim	Promotional Dim
Product Sales	X	X	X	
Promotional Sales		X	X	X

4.3. Data Marts

Product Sales

The below star schema would be used to answer the Business Questions 1,2,3,4
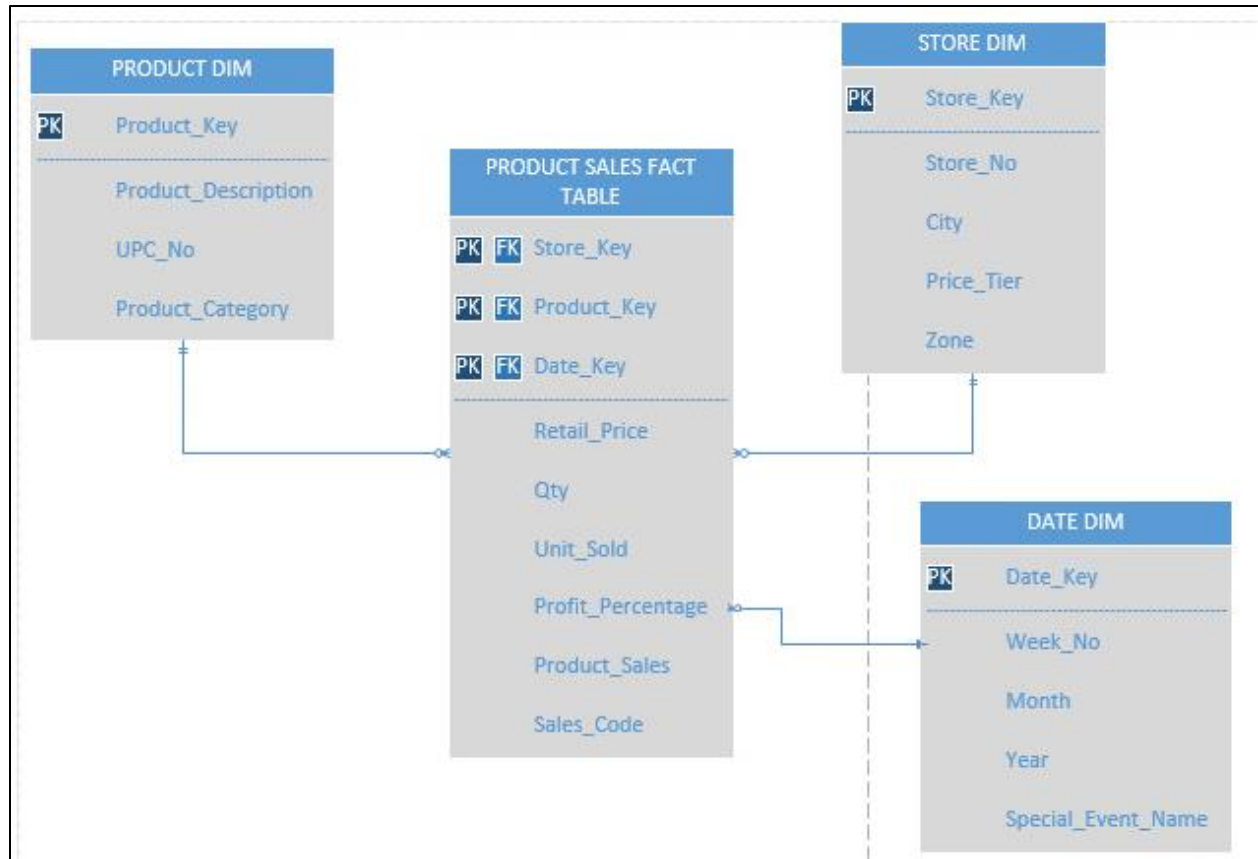
Star Schema for Product Sales Data Mart

Promotional Sales
The below star schema would be used to answer the Business Questions 5
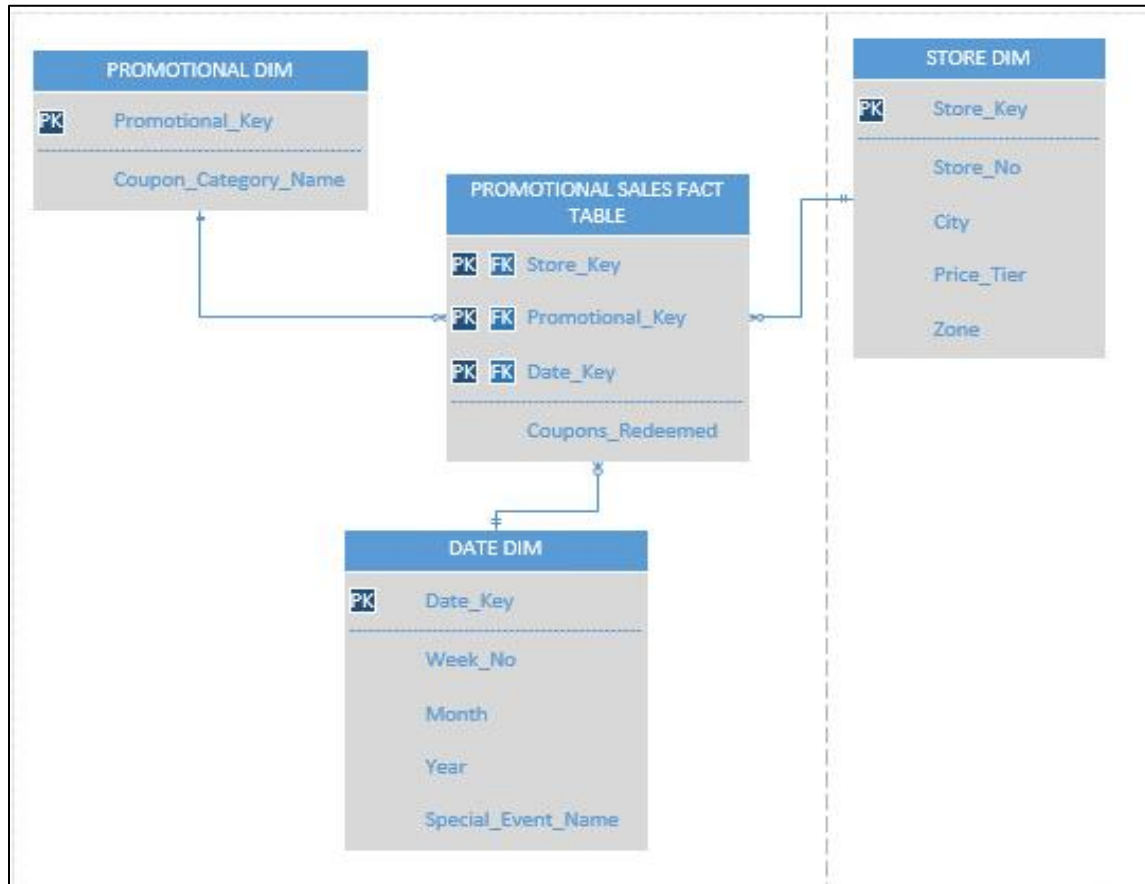
Star Schema for Promotional Sales Data Mart

4.4. Schema Justification for Business Question

BQ1. Analyze the average profit margin of Bath Tissues across all the stores to determine store with low-profit margin
This question is addressed by analyzing the by the Product Sales data mart. Here the question is addressed by analyzing the Bath tissue sales during all the stores to determine store with low profit as shown above. Here the product "Bath Tissue" is considered for analysis and the information related to that can be obtained from the PRODUCT_DIM dimension table, store related information could be found from STORE_DIM table and the profit related information could be taken from the profit percentage attribute in the PRODUCT_SALES fact table.

BQ2. What are the total sales of Beer during the holiday season?
This question is addressed by analyzing the by the Product Sales data mart. Here the question is addressed by analyzing the Beer sales during all the holiday weeks as shown above. Here the product "Beer" is considered for analysis and the information related to that can be obtained from the PRODUCT_DIM dimension table and Week related information could be taken from the

Week_no. attribute in the DATE_DIM dimension table and the sales related information could be taken from the product sales attribute in the PRODUCT_SALES fact table.

BQ3. Which pricing tier produces the highest bottled juice sales during Christmas Week?
This question is addressed by analyzing the by the Product Sales data mart. Here the question is addressed by analyzing the floral sales during all the Christmas week for different pricing tiers as shown above. Here the product "floral" is considered for analysis and the information related to that can be obtained from the PRODUCT_DIM dimension table and Week related information could be taken from the Week_no. attribute in the DATE_DIM dimension table and price tier information can be found from price_tier attribute in the STORE_DIM.

BQ4. Which Bath Soap product has a decreasing sales trend across different zones?
This question is addressed by analyzing the by the Product Sales data mart. Here the question is addressed by analyzing the Bath Soap product sales trend across different zones as shown above. Here the product "Bath Soap" is considered for analysis and the information related to that can be obtained from the PRODUCT_DIM dimension table and zone related information could be taken from the zone attribute in the STORE_DIM dimension table and the sales related information could be taken from the product sales attribute in the PRODUCT_SALES fact table.

BQ5. What category of coupons are redeemed the most for the entire duration across all the stores?
This question is addressed by analyzing the by the Coupon Sales data mart across all the stores. Here the question is addressed by analyzing the greatest number of coupons redeemed for the entire duration as shown above. Here the coupons redeemed are considered for analysis and the information related to that can be obtained from the PROMOTIONAL_DIM dimension table and duration related information could be taken from the DATE_DIM dimension table and store related information can be found in STORE_DIM dimension table.

5. Data Cleansing and Integration
    5.1. Data Quality Issues

Group	Quality	Issues Considered	Examples of Data Quality Problems
Relation to other data	Referential Integrity	Do records exist where expected? Do they contain unnecessary or inactive data? Are reference files/tables complete?	Date field in the CCoount contained garbage values. Many NULL values were found in Movement table
	Cardinality	Is the structure of relationships among entities	The structure and the data were found to be inconsistent

Structure of fields	Format	Do values follow consistent formatting standards?	Data field had garbage values
		and attributes maintained consistently?	
	Standard	Are data elements consistently defined and understood?	The data fields and structure were clearly defined in the manual but due to some inconsistencies some fields were not understood
	Consistent	Do values represent the same meaning across systems and files?	The files had the same data as defined in the manual
Content within data values	Complete	Is all necessary data present?	No, for some categories in Ccount the UPC or movement file was not found
	Accurate	Does the data accurately represent reality or a verifiable source?	It is found from a verifiable source, but the accuracy of the data cannot be determined
	Valid	Do data values fall within acceptable ranges defined by the business?	No, the store no was not valid in the Ccount and Movement tables
	Fit For Purpose	Is the information valuable to the business? Does the data convey information that can intelligently be consumed by the business?	Yes, once the data is cleaned and analyzed thoroughly, many useful business decisions can be derived.

5.2. ETL Plan

ETL stands for Extract, Transform & Load. ETL is a process of reading data from one or more data source to a destination. Extract is reading data from disparate sources of data and moving on to one or more destination databases. Data transformation involves types such as calculating derived values, merging information, date/time conversion, conversion of units etc. such as the procedure of Transformation involves data cleansing and converting them into proper format which would make the process of analyzing and querying easy. Final step of the ETL procedure is the Load which represents the insertion of data into the desired target database/data mart/data warehouse.

5.3. Data Loading to Date Warehouse

5.3.1. Target Data needed in the data warehouse

The table below shows the data source, staging area tables and warehouse tables

Data Source	Source File Name	Table in the Staging area	Table in the Warehouse
Ccount	CCount.csv	CCoupon, CCoupon_Category Derived	Promotional Sales Fact Table
UPC Files	Consolidated_UPC.xslx	UPC_Data	dim_product
Movement Files	Consolidated_Movement.xlsx	Movement	SalesFactTable
Week Decode File	Week_Decode.xlsx	Week_Decode	dim_date
StoreData	StoreData.xslx	Store_Data	dim_store

5.3.2. Data Sources

The table below shows the data sources and the source filename and the corresponding business question.

Data Source	Source File Name	Corresponding Business Question
Ccount	CCOUNT.csv	BQ3 BQ5
UPC Data	UPC_BAT.xlsx UPCTTI.xlsx UPCBER.xlsx UPCBJC.xlsx	BQ1 BQ2 BQ3 BQ4
Movement Data	DONE_WBAT.xlsx WTTI.xlsx DONE_WBER.xlsx DONE_WBJC.xlsx	BQ1 BQ2 BQ3 BQ4
Week_Decode	Week_Decode.xlsx	BQ2 BQ3
StoreData	StoreData.xlsx	BQ3

5.3.3. Data Mappings

Source Mapping from Excel to Staging Area:

Source File Name	Attributes in the Source File	Staging Area Table Name	Attributes in the Staging Area Table
Ccount.csv		CCoupon, CCouponCategory Derived,	Ccoupon – Ccoupon Category Derived - Coupon, Store, Week, Total_coupon_redeemed
Consolidated_UPC.csv	UPC_No, Product_Category, Product_Name	UPC_Data	UPC_No, Product_Category, Product_Name
Consolidated_Movement.csv	UPC Store Week Move Qty Price Sale Profit OK	Movement	UPC Store Week Move Qty Price Sale Profit Type Product_sales
StoreData.csv	Store No City Price Tier Zone	Store_Data	StoreNo City price Tier Zone
Week_Decode.csv	Week# Start End Special Events	Week_Decode	Start Year Month Week# Special Events

Source Mapping from Staging Area to Data Warehouse:

Staging Area Table Name	Staging Area Table Attributes	Data Warehouse Table	Data Warehouse Table Attributes	Mapping Function
Ccoupon				
Ccoupon Category Derived	Coupon Store Week Total_Coupon_Redeemed			
UPC_Data	UPC_No Product_Category Product_Name	dim_product	product_key upc_no product_category product_name	Surrogatekey Copy copy
Movement	UPC Store Week Move Qty Price Sale Profit Type Product_Sales	SalesFactTable	storekey productkey datekey retailPrice qty profit sales_code product_sales	
Store_Data	Store No City Price Tier Zone	dim_store	storekey store_no city price_tier zone_no	
Week_Decode	Start Year Month Week# Special Events	dim_date	date_key week_no month year special_event_name	

5.3.4. Data Extraction Rules

The first step of ETL is the extraction process and for this process, certain rules need to be defined.

- Out of all the source files (Ccount, consolidated_UPC, consolidated_Movement, StoreData, Week_Decode) except the Ccount file rest all files are in xlsx. While, Ccount is in csv format
- The data in Week_Decode is extracted from the manual of Dominick's Finer Food and recorded in an excel file with extension .xlsx
- The surrogate keys in the data warehouses have suffixes as _key
- The source files have been stored in the staging areas in separate tables which would make the process of extraction simpler

5.3.5. Data Transformation and Cleansing Rules

The Data extracted to be stored into the staging area is dirty. This data is stored is an undesired format and needs to be cleansed as per the requirements of the destination database i.e. the warehouse. Various transformation functions are applied to the data in the staging area to remove unwanted data, thus ensuring easy loading of the data into the data warehouse tables.

The data has been cleansed in the following way:
- **Removing NULL values**
Null values deviate the obtained results from the desired results. Hence, such values in the data set shouldn't be considered.

- **Removing Special Characters**
The data which is to be loaded into the warehouse has to be clean and so, cleansing the data by removing the special characters is important. For instance, special characters such as the attribute, product name, doesn't require the use of special characters and hence should be removed

- **Removing Unwanted Data**
The attributes which are not used to answer the business questions are removed

- **Creating fields using formula**
Certain required fields are created in the staging area table using the attributes. For instance, the field product_sales movement table is calculated using the formula *(Price * Move)/Qty*

- **Converting formats**
The data field in the source tables is combined using date, month and year. Hence, before the data is loaded in the staging table, the month (date datatype) and year (date datatype) is extracted from the Week_Decode.xlsx file. Extraction of this attribute help in the formation of

the date dimension.

- **Creation of Surrogate keys**

Surrogate keys are created in the dimension tables as unique identifiers and composition of these surrogate keys form the primary key in the fact table
The Surrogate Keys are as follows:
  - storekey: Unique identifier for the store dimension table i.e. dim_store
  - product_key: Unique identifier for the product dimension table i.e. dim_product
  - date_key: Unique identifier for the date dimension table i.e. dim_date

- **Removal of Unwanted Data**

The attributes that are not required to answer the business questions should be removed

- **Derived Columns**

Introduction of new columns were required in the transformation process. Such derived columns are needed to perform further calculations. The derived columns and the source are listed below,

Source Table Name	Source Column Name	Derived Column Name	Mapping Function
Consolidated_Movement.csv	PriceMoveQty	Product_Sales	(Price*Move)/Qty
Week_Decode.xlsx	Start	Day Month Year	Day(Start) Month(Start) Year(Date)

- **SSIS Functions**

Below are some of the SSIS functions that were used during the ETL process.

Function Name	Function's Use
AGGREGATE	Applying aggregate functions such as sum, average, count, group by etc. on groups in a data set
LOOKUP	This function performs lookups by joining the data in input columns to retrieve corresponding columns from the reference data set
MERGE JOIN	A Merge join is used to join datasets which are already sorted
UNPIVOT	This function transforms columns to rows.

- **Aggregate Plan**

The table given below shows the aggregation function involved in the warehouse.

Table Name	Attribute	Aggregate Function
Ccoupon_Category Derived	Store_No Week_No	Group_By

5.3.6. Procedures for Extraction and Loading

Data Extraction Procedures:

The following tables are created in the data staging area

a) **Ccount**

Source: Ccount.csv

Destination: CCoupon, CCoupon_Category_Derived,

  i.   The source file is extracted through the SSIS package
  ii. The unnecessary columns were dropped using SQL queries
  iii. After extraction, the necessary transformation procedure were performed such as unpivot to transpose the columns to rows and calculate the total coupons redeemed.

b) **UPC Files**

Source: UPC_BAT.xlsx, UPCTTI.xlsx, UPCBER.xlsx, UPCBJC.xlsx
Destination: UPC_Data

  i.   The required columns from these UPC Source files is extracted
  ii. Rest unnecessary columns are dropped at this time and an additional column called Product_Category is added
  iii. After extraction, transformations such as removal of special characters from the Product_Description column

c) **Movement Files**

Source: DONE_WBAT.xlsx, WTTI.xlsx, DONE_WBER.xlsx, DONE_WBJC.xlsx
Destination: Movement

  i.   The required columns from these Movement Source files is extracted
  ii. Rest unnecessary columns are dropped at this time and an additional column called Product_Sales is added which is created using the formula *(Price*Move)/Qty*
  iii. After extraction, transformations such as removing NULL values, selected only the store numbers between values 1 and 139

d) **Week_Decode**

Source: Week_Decode.xlsx
Destination: Week_Decode

  i.   The required columns from these Week_Decode Source file is extracted

ii.   Rest unnecessary columns are dropped at this time

iii.   After extraction, transformations such as creation of additional columns Year, Month using transformation process in SSIS

Data Loading Procedures:

The following tables are created in the warehouse

a) Dim_Date

Source: Week_Decode

Destination: Dim_Date

    i.   Performed the transformation to create columns Year & Month

    ii.   Added a surrogate key Date_Key

    iii.   Loaded the table into the data warehouse under Dim_Date

b) Dim_Product

Source: UPC_Data

Destination: Dim_Product

    i.   Created an additional column called Product_Category and removed unnecessary columns

    ii.   Added a surrogate key Product_Key

    iii.   Loaded the table into the data warehouse under Dim_Product

c) Dim_Store

Source: Store_Data

Destination: Dim_Store

    i.   Removed unnecessary columns from the source file

    ii.   Added a surrogate key Store_Key

    iii.   Loaded the table into the data warehouse under Dim_Store

d) Dim_Promotional

Source: CCoupon_Category_Derived

Destination: Dim_Promotional

    i.   Performed aggregate function to find total coupons redeemed per week, per store

    ii.   Added a surrogate key Promotional_Key

    iii.   Loaded the table into the data warehouse under Dim_Promotional

e) CCoupon

Source: CCount.xlsx

Destination: CCoupon

    i.   Applied queries to convert the negative values to positive values, and calculate the total coupons redeemed

    ii.   Loaded the table into the data warehouse

f) CCoupon_Category_Derived

Source: CCoupon

Destination: CCoupon_Category_Derived

    i.   Performed unpivot to convert the columns to rows

    ii.   Loaded the table into the data warehouse under CCoupon_Category_Derived

g) ProductSalesFactTable

Source: Movement Staging Table

Destination: ProductSalesFactTable

    i.   Derived the column such as retail price, qty, profit, and sales code

    ii.   Calculate the value for prodcuct_sales using the formula *(Price*Move)/Qty*

    iii.   Loaded the table into the data warehouse under ProductSales Fact table

h) PromotionalSalesFactTable

Source: CCoupon_Category_Derived

Destination: PromotionalFactTable

    i.   Performed aggregate function to get the total number of coupons redeemed and performed lookup on various dimension tables to obtain the key

    ii.   Loaded the table into the data warehouse under PromotionalSalesFactTable

5.3.7. ETL For Dimension Tables

The ETL process carried out for the dimension tables has been diagrammatically depicted below

Date Dimension Table
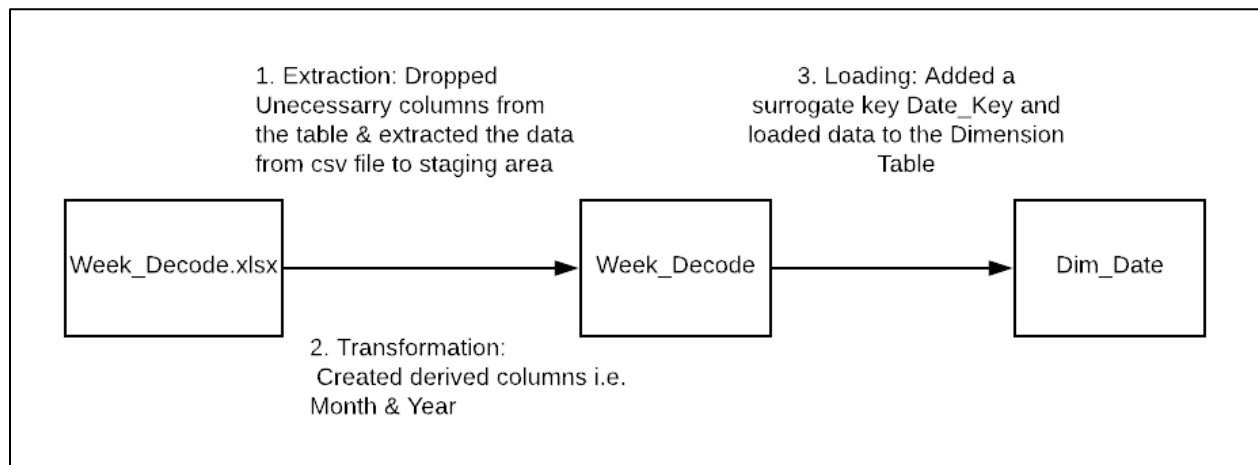
Product Dimension Table
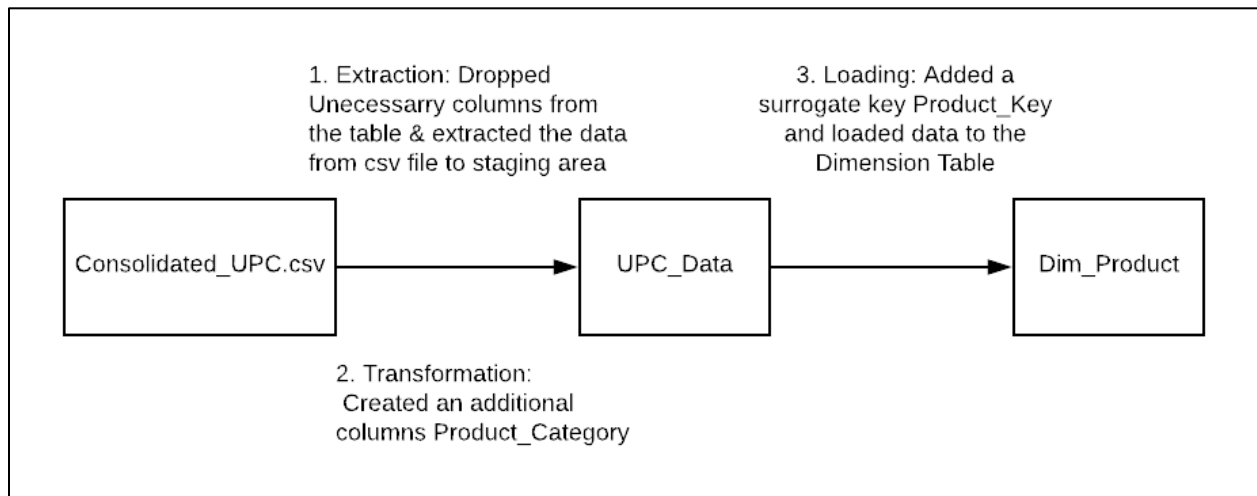
Store Dimension Table
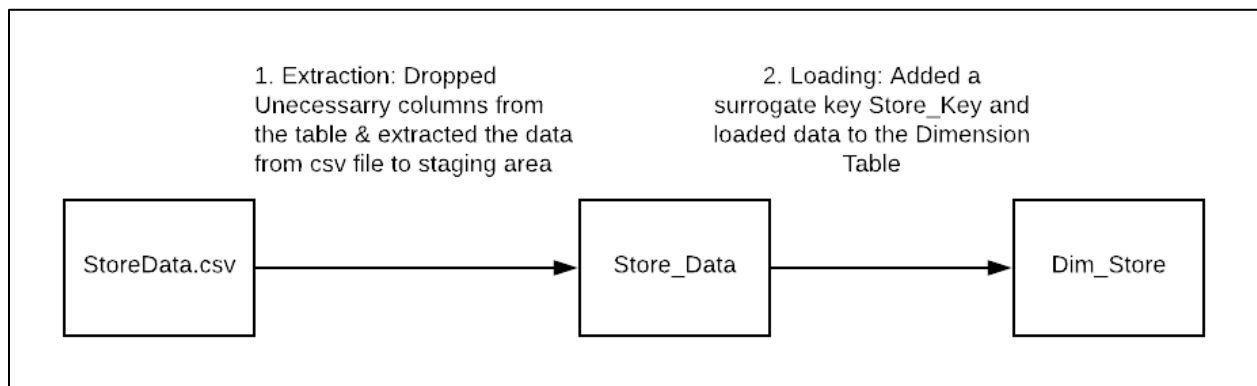
Promotional Dimension Table
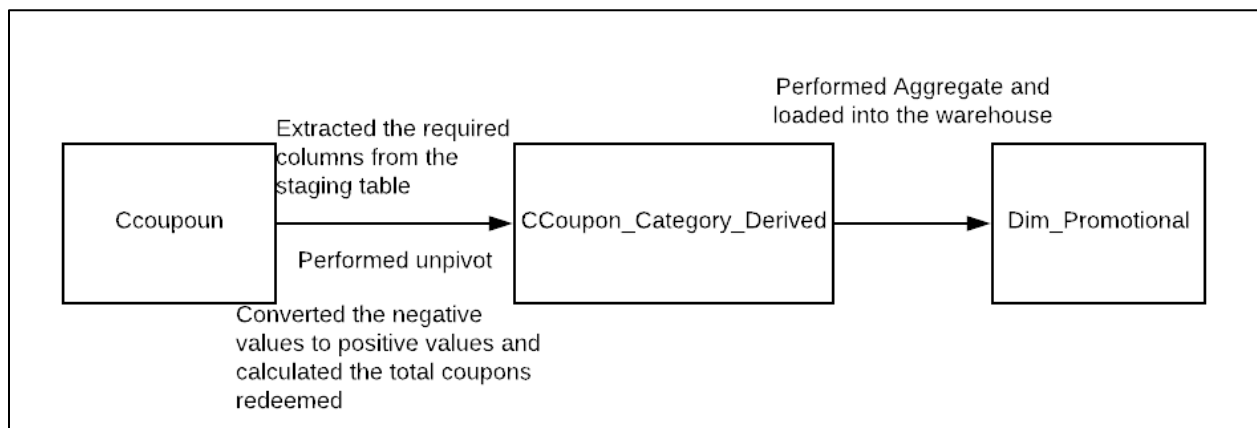
5.3.8. ETL For Fact Table

Product Sales Fact Table
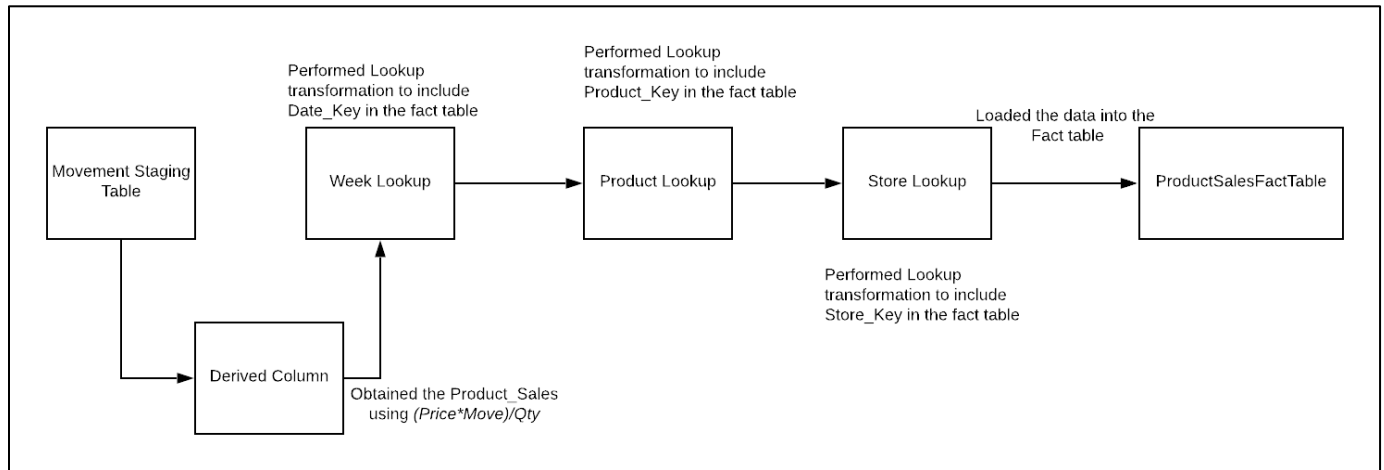
Promotional Sales Fact Table
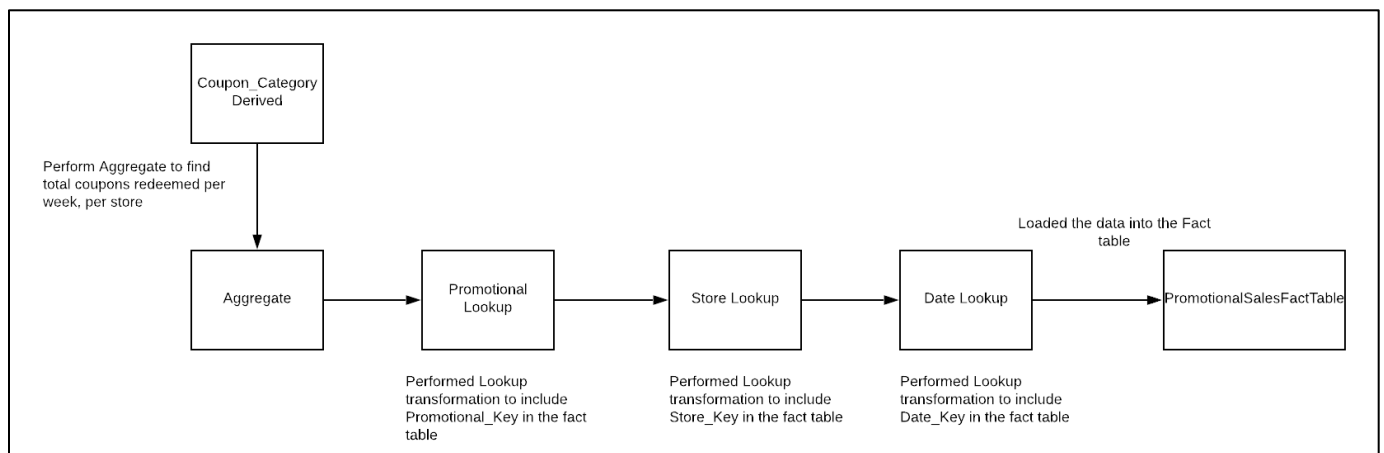
5.4. ETL Implementation

This section describes the actual implementation of the ETL processes in SSIS. The construction of each individual dimension and fact table has been shown right from the extraction of data from source files to the final loading of data into the data mart.

Product Dimension Table

Table containing product_description, UPC_no & category_name for each product. The surrogate key Product_Key identifies a product based on the UPC_no.

i. Extraction

Extracted the required columns (UPC_No, Product_Description) for the required products (beer, bath tissue, bath soap, bottled juice) from the UPC tables and loaded into the UPC_Data Staging table. Loaded the beer UPC initially and later appended the other products on top of existing data in the UPC_Data.
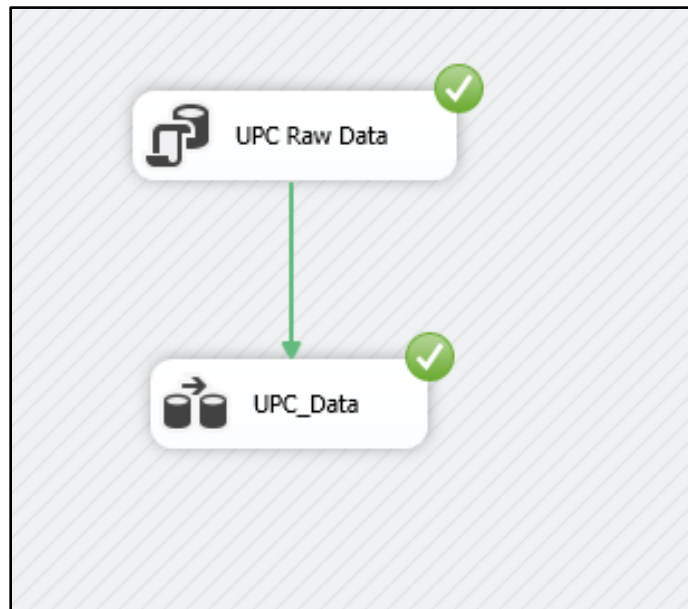
ii.   Transformation

Garbage data from product_description was removed and was renamed as Product_Name. Special characters & unnecessary data were removed. Added column called Category_Name
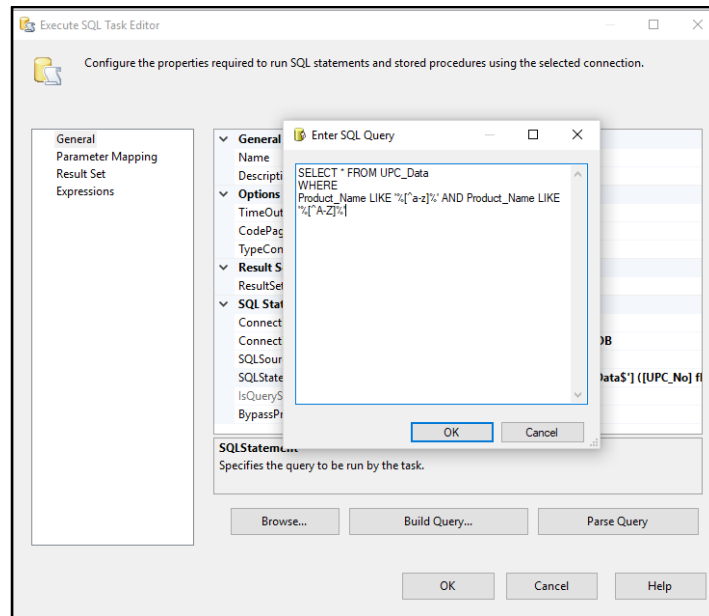
SQL Queries:

*SELECT * FROM UPC_Data*
WHERE
Product_Name LIKE '%[^a-z]%' AND Product_Name LIKE '%[^A-Z]%'

iii.   Loading

From the UPC_Data Staging table, Product_Key was added as a surrogate key & the data was loaded into Product_Dim table in the data warehouse.
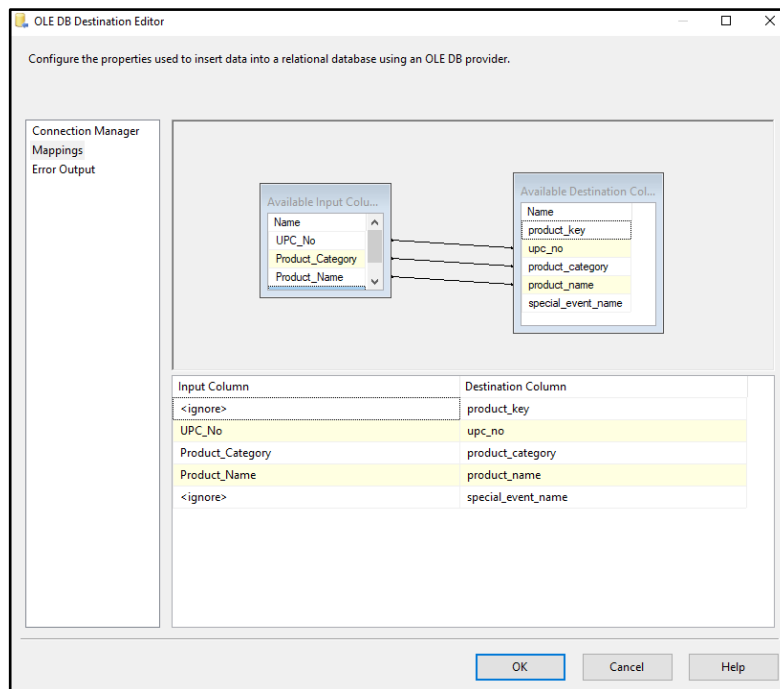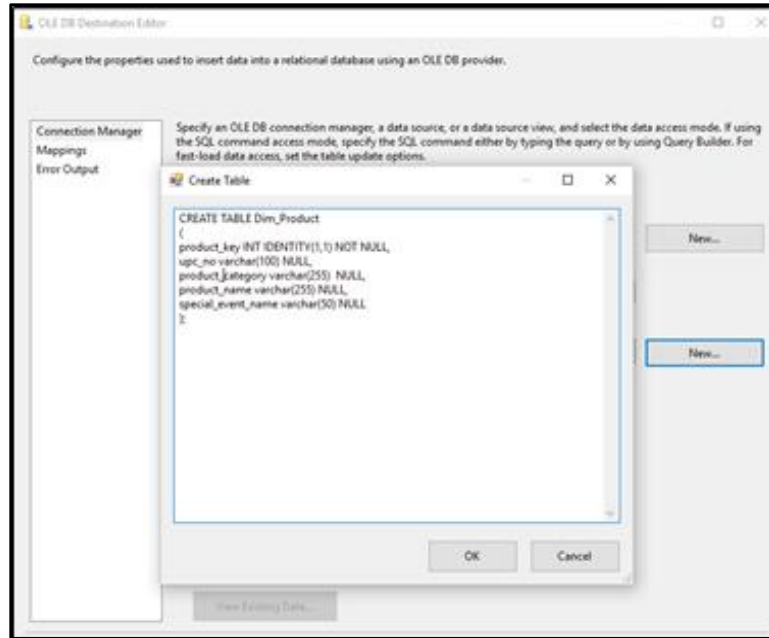
SQL Queries:

CREATE TABLE Dim_Product (
product_key INT IDENTITY(1,1) NOT NULL,
upc_no varchar(100) NULL,
product_category varchar(255) NULL,
product_name varchar(255) NULL,
special_event_name varchar(50) NULL);

	product_key	upc_no	product_category	product_description
1	1	5600003	BAT	GLYCERIN SOAP PINK G
2	2	5600004	BAT	GLYCERIN SOAP RED RA
3	3	5600005	BAT	GREEN APPLE GLY SP
4	4	5600006	BAT	STRAWBERRY GLY SP
5	5	5600007	BAT	FRESH LEMON GLY SP
6	6	5600008	BAT	FRESH APRICOT GLY SP
7	7	5600009	BAT	GLYCERIN SOAP WLD BL
8	8	5600010	BAT	MANDARIN GLY SP
9	9	5623204	BAT	25CT FISH BOWL DISPL
10	10	5623205	BAT	25CT FISH BOWL DISPL
11	11	5623206	BAT	25CT FISH BOWL DISPL
12	12	5623207	BAT	25CT FISH BOWL DISPL
13	13	5623209	BAT	25CT FISH BOWL DISPL
14	14	5623210	BAT	25CT FISH BOWL DISPL
15	15	5623230	BAT	30CT LITTLE SHAPES S
16	16	5623233	BAT	30CT LITTLE SHAPES S
17	17	5623236	BAT	30CT LITTLE SHAPES S
18	18	5623239	BAT	30CT LITTLE SHAPES S
19	19	5623242	BAT	30CT LITTLE SHAPES S

Store Dimension Table

Table containing Store_No, City, Price_Tier & Zone for each store. The surrogate key store_key identifies a store based on the store_no.
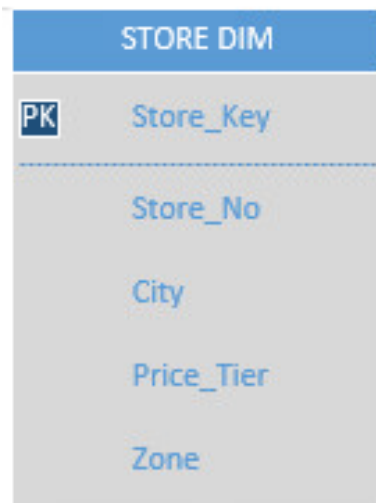
i.   Extraction

Extracted the required columns from StoreData.xlsx which was extracted from manual &
dropped other columns
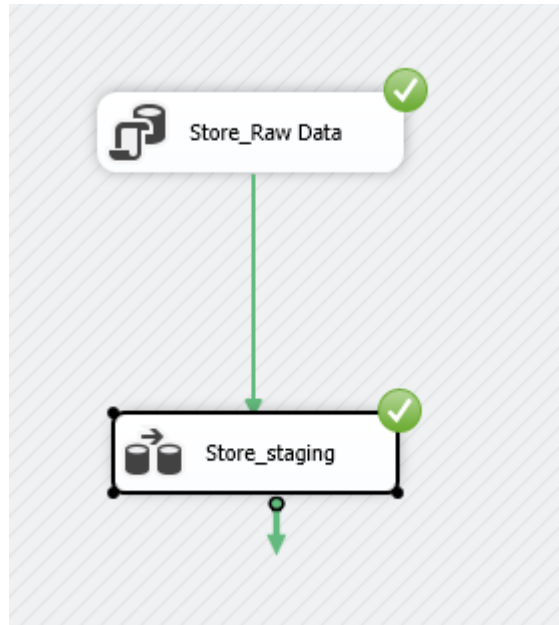
ii.   Loading

From the Store_Staging table, Store_Key was added as a surrogate key & the data was
loaded into Store_Dim table in the data warehouse.
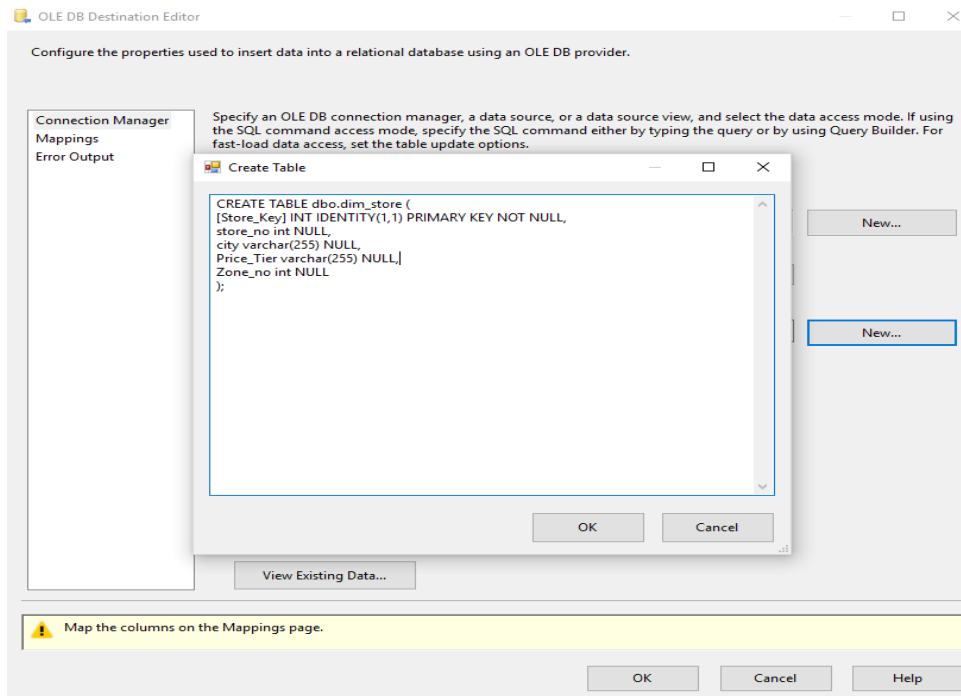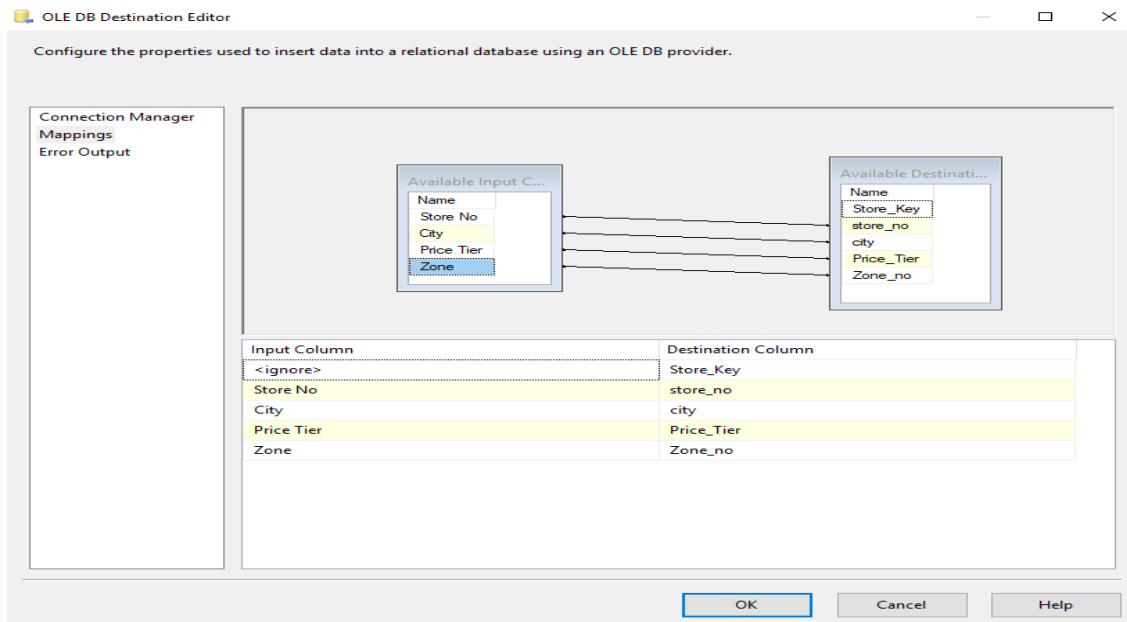
SQL Queries:

CREATE TABLE dbo.dim_store (
[storekey] INT IDENTITY(1,1) PRIMARY KEY NOT NULL,
store_no int NULL,
city varchar(255) NULL,
Price_Tier varchar(255) NULL,
Zone_no int NULL);

Date Dimension Table

Table containing Week_No, Month, Year, Special_Event_Name. The surrogate key date_key identifies a product based on the week.
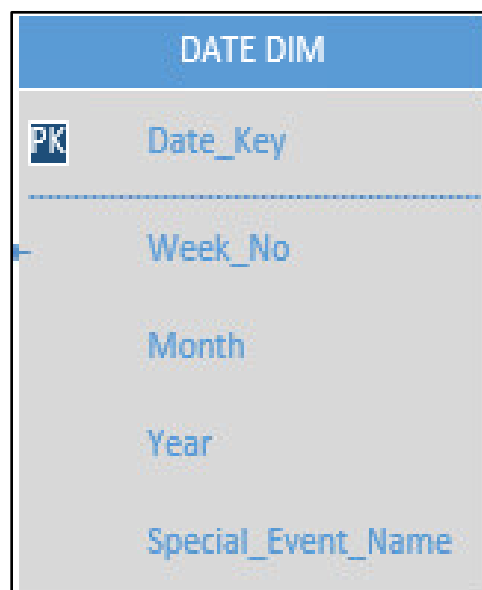
i.   Extraction

Extracted the required columns from Week_Decode.xlsx table & loaded into week staging table.
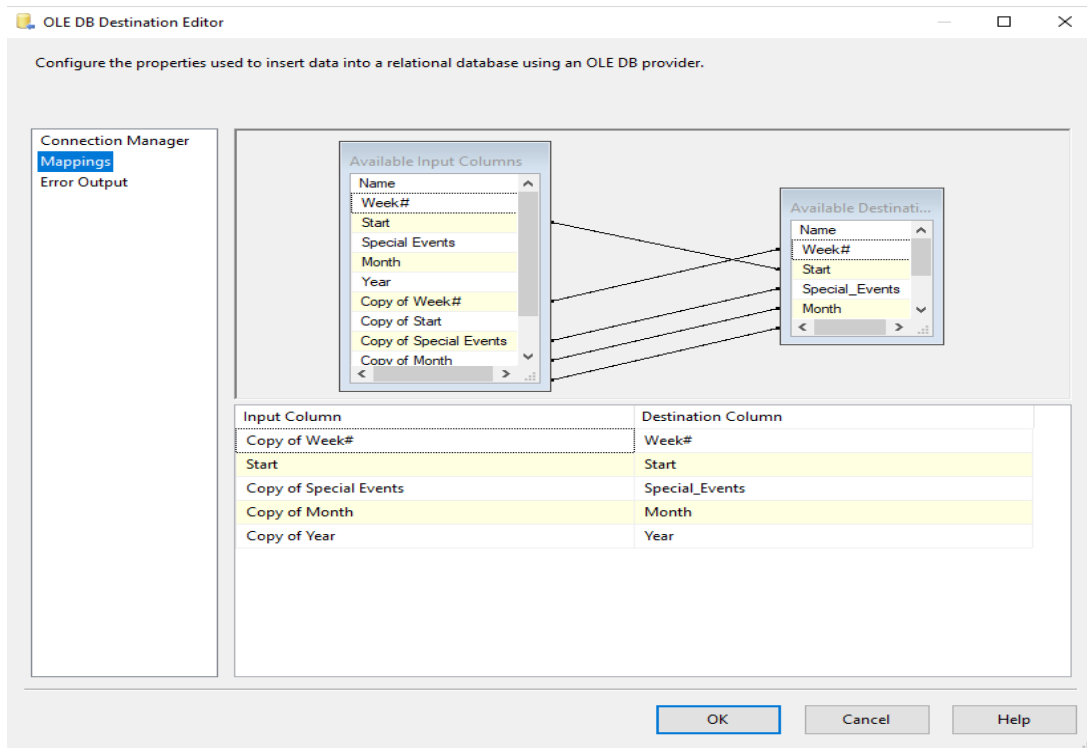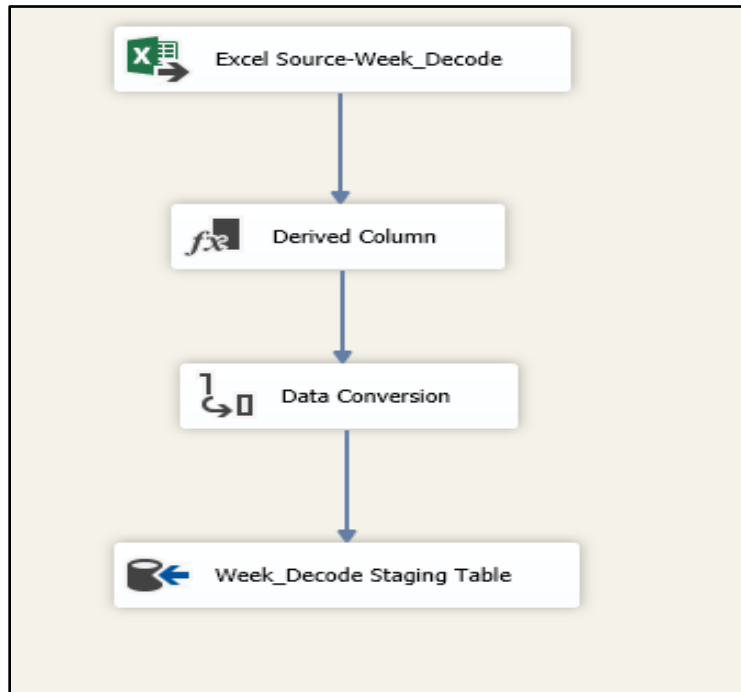
ii.   Transformation

Week, month, year were extracted from the date attribute.
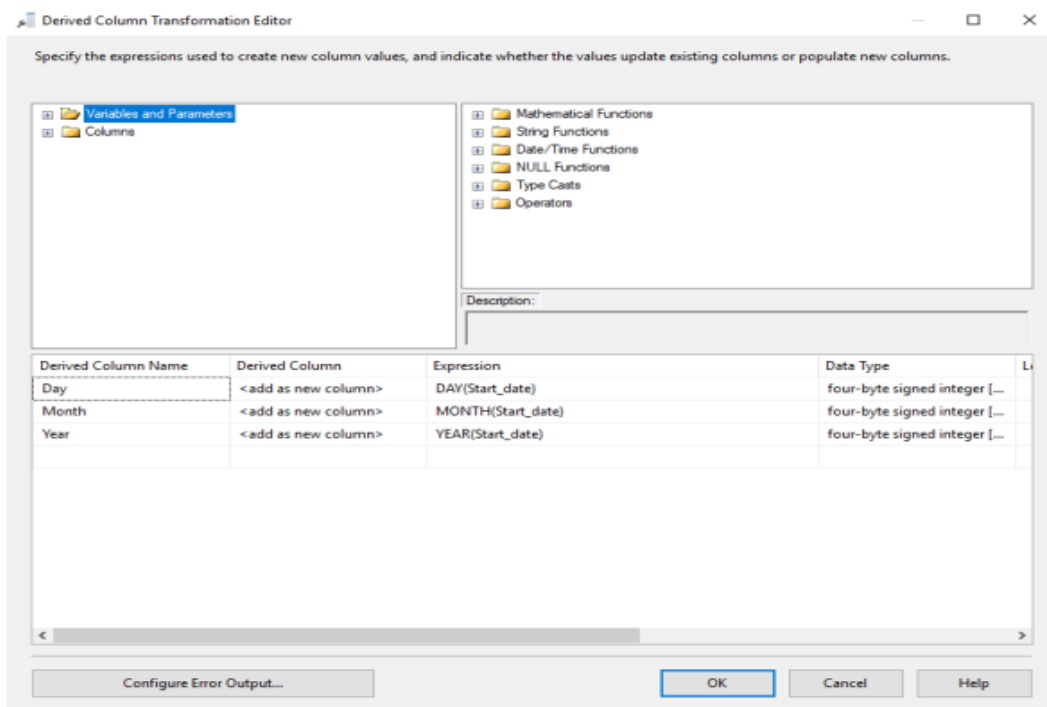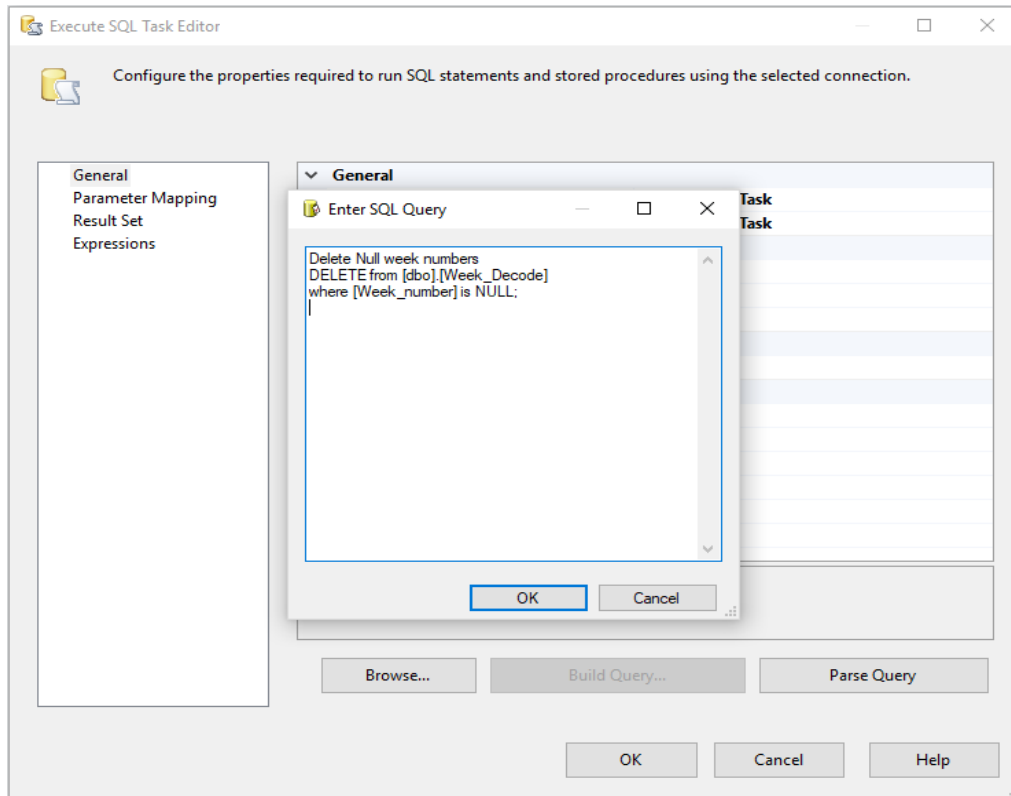
SQL Queries:

Delete Null week numbers
DELETE from [dbo].[Week_Decode]
where [Week_number] is NULL;

    iii.   Loading
From the Week_Decode table, date_key was added as a surrogate key & the data was loaded into Date_Dim table in the data warehouse.
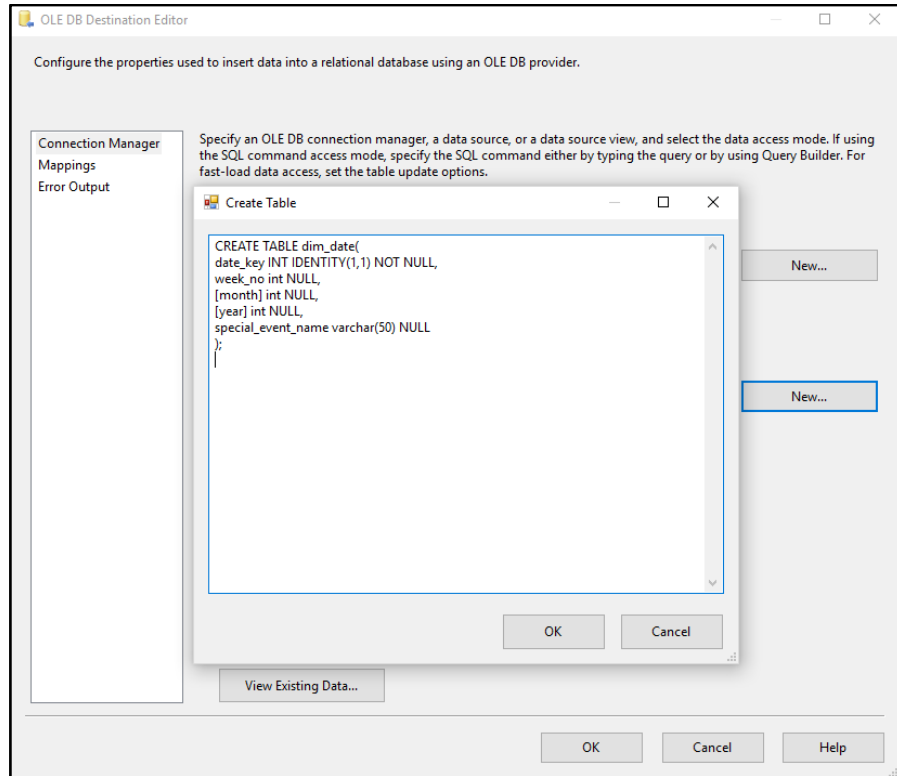
SQL Queries:

Date Dimension
CREATE TABLE dim_date(
date_key INT IDENTITY(1,1) NOT NULL,
week_no int NULL,
[month] int NULL,
[year] int NULL,
special_event_name varchar(50) NULL);
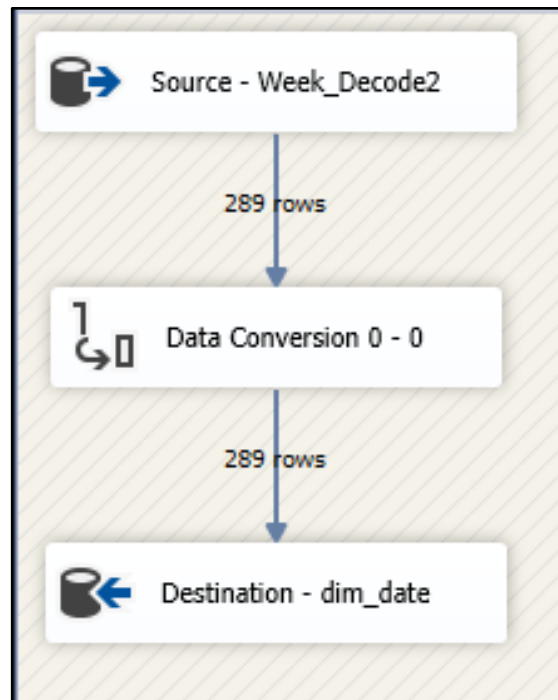
	date_key	week_no	month	year	special_event_name
1	2	1	9	1989	
2	3	2	9	1989	
3	4	3	9	1989	
4	5	4	10	1989	
5	6	5	10	1989	
6	7	6	10	1989	
7	8	7	10	1989	Halloween
8	9	8	11	1989	
9	10	9	11	1989	
10	11	10	11	1989	
11	12	11	11	1989	Thanksgiving
12	13	12	11	1989	
13	14	13	12	1989	
14	15	14	12	1989	
15	16	15	12	1989	Christmas
16	17	16	12	1989	New-Year
17	18	17	1	1990	
18	19	18	1	1990	
19	20	19	1	1990	

Promotion Dimension Table

Table containing coupon_category. The surrogate key promotional_key identifies based on the coupon_category_Name.

PROMOTIONAL DIM

PK   Promotional_Key

Coupon_Category_Name
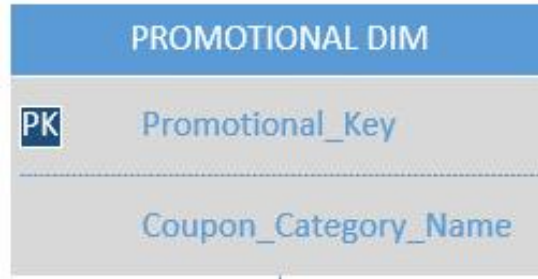
    i.   Extraction

Extracted the required columns from Ccount.csv table & loaded into staging table
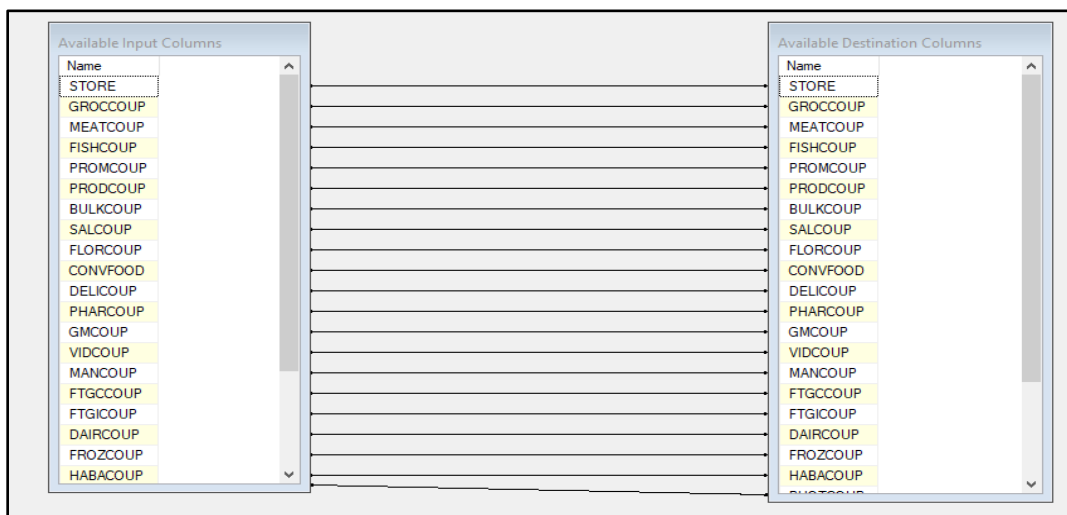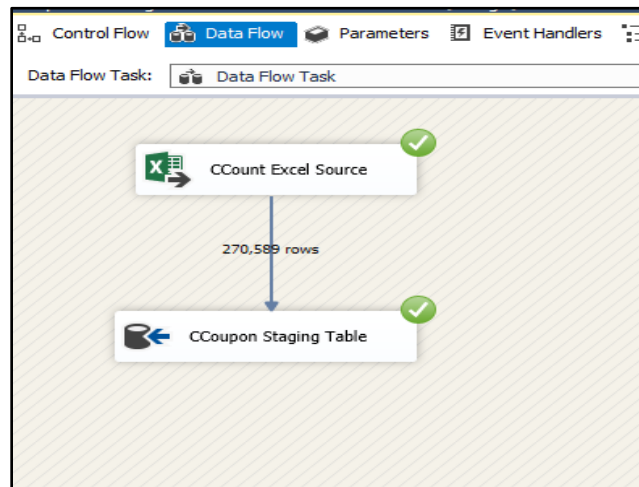
Results | Messages

	STORE	GROCCOUP	MEATCOUP	FISHCOUP	PROMCOUP	PRODCOUP	BULKCOUP	SALCOUP	FLORCOUP	CONVFOOD	DELICOUP	PHA
1	2	494.21	0	0	385	0	0	0	0	128.69	37	0
2	2	552.19	0	0	405	0	0	0	0	79.8	29.5	0
3	2	648.68	0	0	241	0	0	0	0	0	65	0
4	2	440.17	0	0	295	0	0	0	3	139.65	49.25	0
5	2	404.07	0	0	180	0	0	0	0	148.64	33.25	0
6	2	362.37	0	0	163	0	0	0	0	151.62	24.25	0
7	2	366.56	0	0	165	0	0	0	0	143.64	20.75	0
8	2	2.39	0	0	176	0	0	0	0	136.67	10.25	0
9	2	3.39	0	0	294.64	0.75	0	0	0	95.76	13.25	0
10	2	0.79	0	0	133.29	0	0	0	0	123.69	19.12	0
11	2	2.08	0	0	71	1.5	0	0	0	147.63	8	0
12	2	4.97	0	0	129	0	0	0	1.29	149.65	6	0
13	2	7.85	0	0	88	0	0	0	0	147.63	17.18	0
14	2	13.22	0	0	43	0	0	0	0	127.68	6	0
15	2	1917.3	193.03	0	224	672.29	0	0	0	104.75	223.8	0
16	2	2246.14	311.09	0	299	745.53	0	0	4	87.78	295.85	0
17	2	2138.87	262.72	0	164	735.53	0	0	0	156.62	388.4	1
18	2	1593.59	128.21	0	133	470.04	0	0	0	147.63	95.22	0

ii.   Transformation
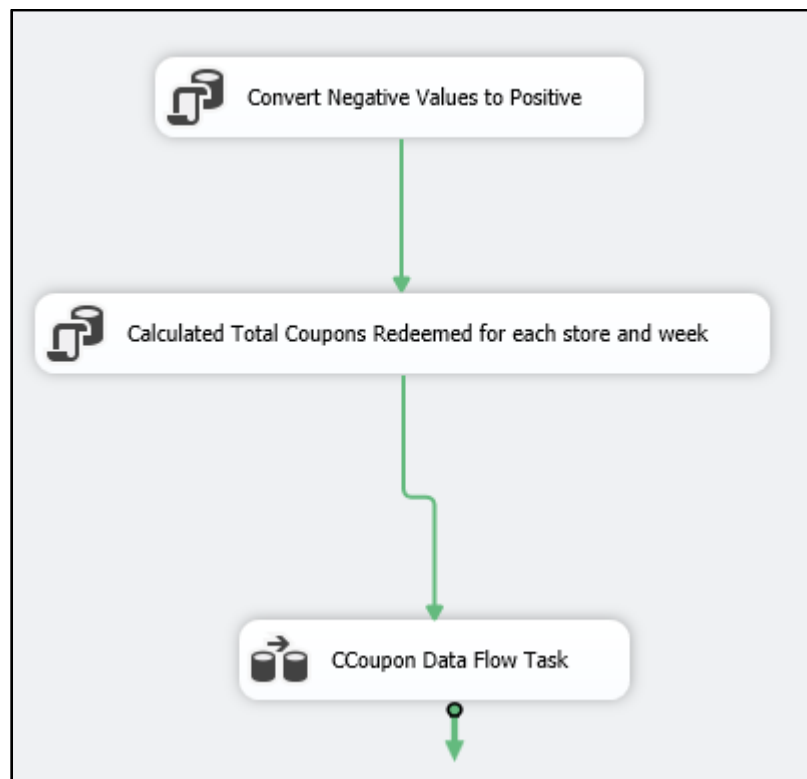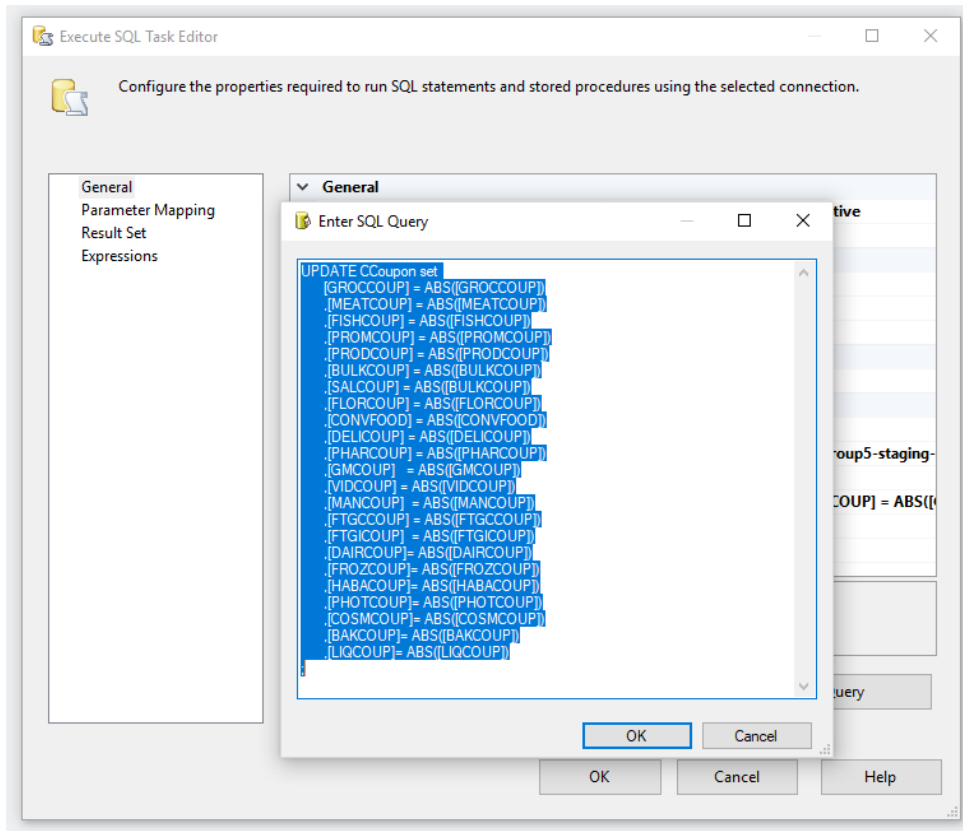
Converted the negative values to positive values for the coupons redeemed
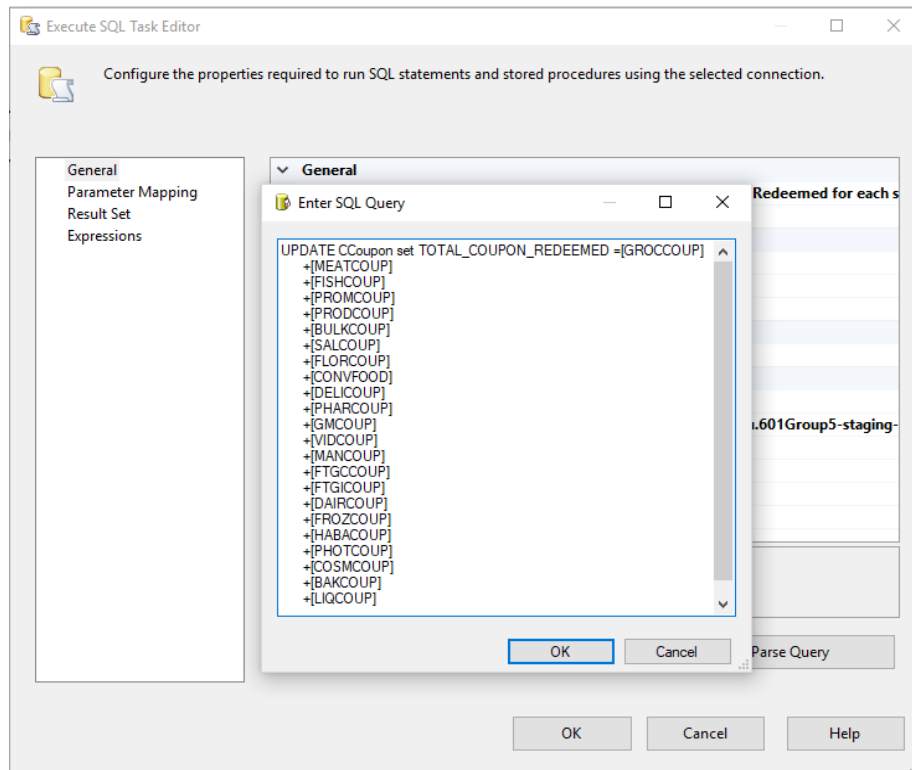
Calculated total coupon redeemed
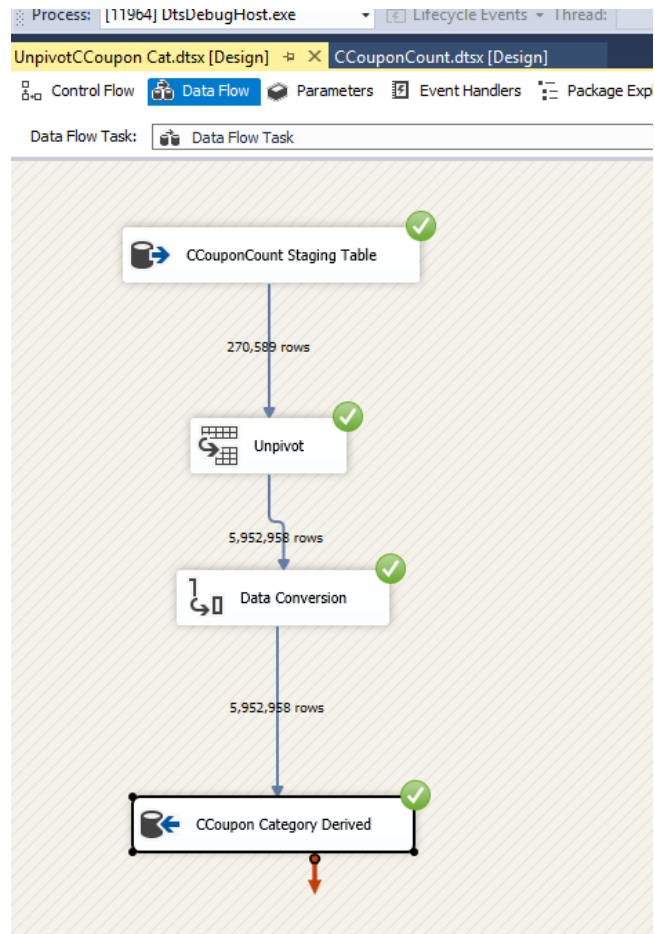
Transformed the coupon columns into rows

Data for CCoupon_Category_Derived

SQL Queries:

UPDATE CCoupon set
*   [GROCCOUP] = ABS([GROCCOUP])*
*   ,[MEATCOUP] = ABS([MEATCOUP])*
*   ,[FISHCOUP] = ABS([FISHCOUP])*
*   ,[PROMCOUP] = ABS([PROMCOUP])*
*   ,[PRODCOUP] = ABS([PRODCOUP])*
*   ,[BULKCOUP] = ABS([BULKCOUP])*
*   ,[SALCOUP] = ABS([BULKCOUP])*
*   ,[FLORCOUP] = ABS([FLORCOUP])*
*   ,[CONVFOOD] = ABS([CONVFOOD])*
*   ,[DELICOUP] = ABS([DELICOUP])*
*   ,[PHARCOUP] = ABS([PHARCOUP])*
*   ,[GMCOUP]   = ABS([GMCOUP])*
*   ,[VIDCOUP] = ABS([VIDCOUP])*
*   ,[MANCOUP] = ABS([MANCOUP])*
*   ,[FTGCCOUP] = ABS([FTGCCOUP])*
*   ,[FTGICOUP] = ABS([FTGICOUP])*
*   ,[DAIRCOUP]= ABS([DAIRCOUP])*
*   ,[FROZCOUP]= ABS([FROZCOUP])*
*   ,[HABACOUP]= ABS([HABACOUP])*
*   ,[PHOTCOUP]= ABS([PHOTCOUP])*
*   ,[COSMCOUP]= ABS([COSMCOUP])*
*   ,[BAKCOUP]= ABS([BAKCOUP])*
*   ,[LIQCOUP]= ABS([LIQCOUP]);*

UPDATE CCoupon set TOTAL_COUPON_REDEEMED =[GROCCOUP]
*   +[MEATCOUP]*
*   +[FISHCOUP]*
*   +[PROMCOUP]*
*   +[PRODCOUP]*
*   +[BULKCOUP]*
*   +[SALCOUP]*
*   +[FLORCOUP]*
*   +[CONVFOOD]*
*   +[DELICOUP]*
*   +[PHARCOUP]*
*   +[GMCOUP]*
*   +[VIDCOUP]*
*   +[MANCOUP]*
*   +[FTGCCOUP]*
*   +[FTGICOUP]*
*   +[DAIRCOUP]*
*   +[FROZCOUP]*
*   +[HABACOUP]*
*   +[PHOTCOUP]*
*   +[COSMCOUP]*
*   +[BAKCOUP]*
*   +[LIQCOUP]*

iii.   Loading
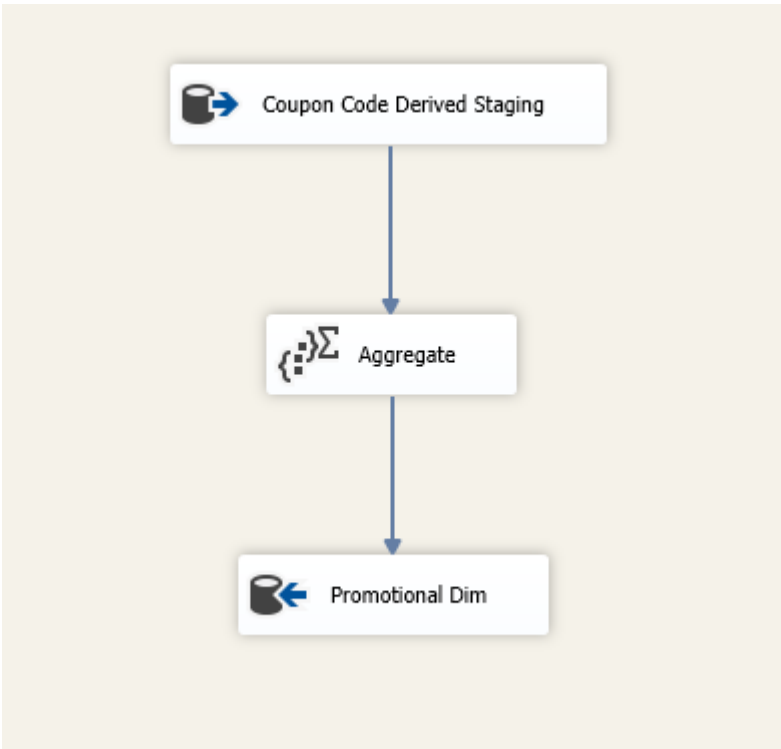
Loading per store per week coupons redeemed
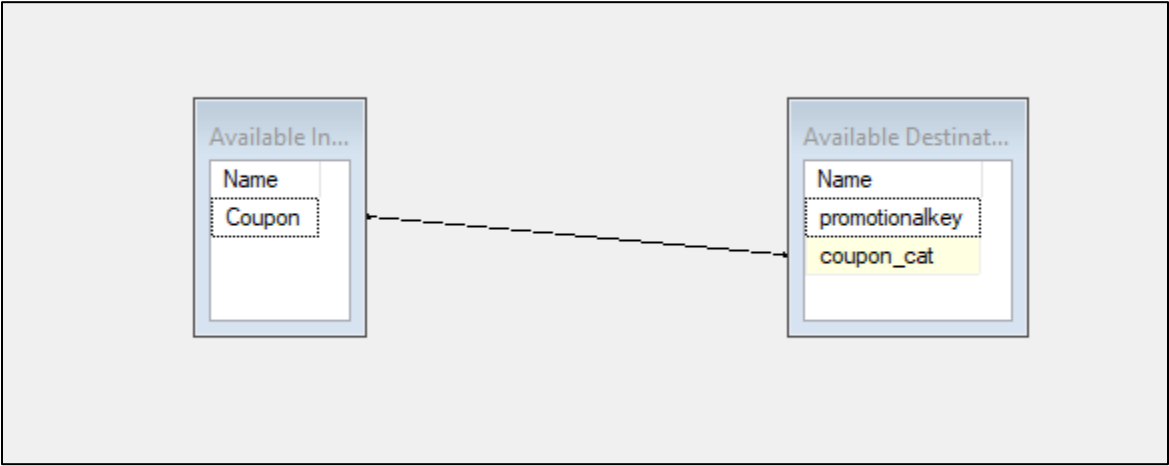
Mapping Tables from source to destination

	promotionalkey	coupon_cat
1	1	GMCOUP
2	2	VIDCOUP
3	3	PRODCOUP
4	4	BULKCOUP
5	5	FTGCCOUP
6	6	HABACOUP
7	7	FLORCOUP
8	8	FISHCOUP
9	9	FTGICOUP
10	10	MANCOUP
11	11	GROCCOUP
12	12	DELICOUP
13	13	SALCOUP
14	14	MEATCOUP
15	15	BAKCOUP
16	16	PHOTCOUP
17	17	COSMCOUP
18	18	DAIRCOUP
19	19	FROZCOUP

Product Sales Fact Table

Loaded the Movement Tables into staging database to get the values like retail_price, qty, profit, product_sales, and sales_code. To get the value of product_sales, created a derived function. Used this Movement table from staging database as the source for creating product_sales_fact table.

Derived Column Transformation Editor

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.

Variables and Parameters
Columns
- UPC
- STORE
- WEEK
- MOVE
- QTY
- PRICE
- SALE
- PROFIT
- TYPE
- Product_sales

Operators
- fx + (Add)
- fx + (Concatenate)
- fx - (Subtract)
- fx - (Negate)
- fx * (Multiply)
- fx / (Divide)
- fx % (Modulo)
- fx () (Parentheses)
- fx == (Equal)

Description:

Derived Column Name	Derived Column	Expression	Data Type	L
Product_Sales	<add as new column>	[PRICE] * [MOVE] / [QTY]	numeric [DT_NUMERIC]	

OLE DB Destination Editor

Configure the properties used to insert data into a relational database using an OLE DB provider.

Connection Manager
Mappings
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

Create Table

```
CREATE TABLE dbo.SalesFactTable (
[storekey] INT NOT NULL,
[productkey] INT NOT NULL,
[datekey] INT NOT NULL,
[RetailPrice] MONEY NULL,
[Qty] INT NULL,
[Profit] MONEY NULL,
[Product_Sales] MONEY NULL,
[Sales_code] VARCHAR(10) NULL
);
```

OK   Cancel

New...

New...

View Existing Data...

OK   Cancel   Help

Dim_Date Lookup

Dim_Product Lookup
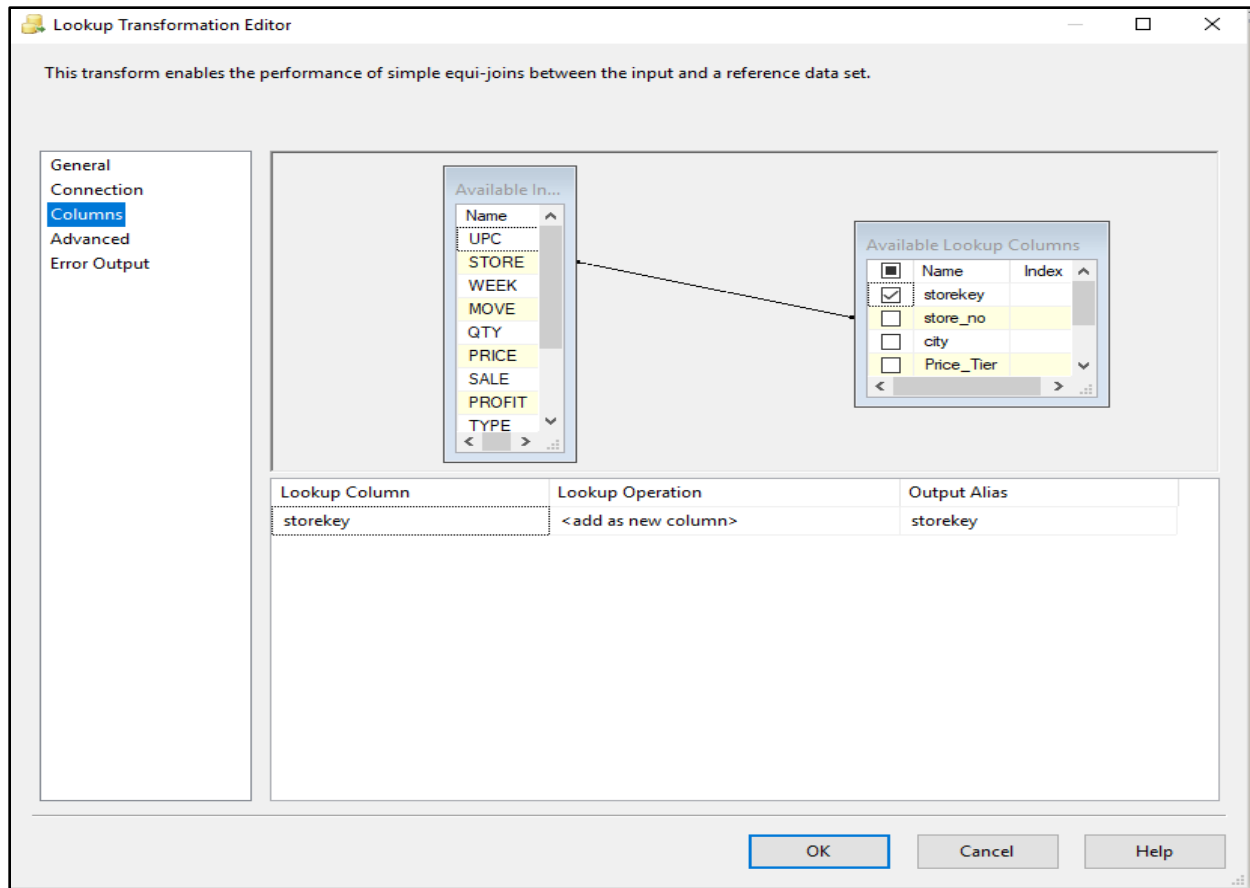
Dim_Store Lookup

SQL Queries:

Sales Fact Table
CREATE TABLE dbo.SalesFactTable (
[storekey] INT NOT NULL,
[productkey] INT NOT NULL,
[datekey] INT NOT NULL,
[RetailPrice] MONEY NULL,
[Qty] INT NULL,
[Profit] MONEY NULL,
[Product_Sales] MONEY NULL,
[Sales_code] VARCHAR(10) NULL
);
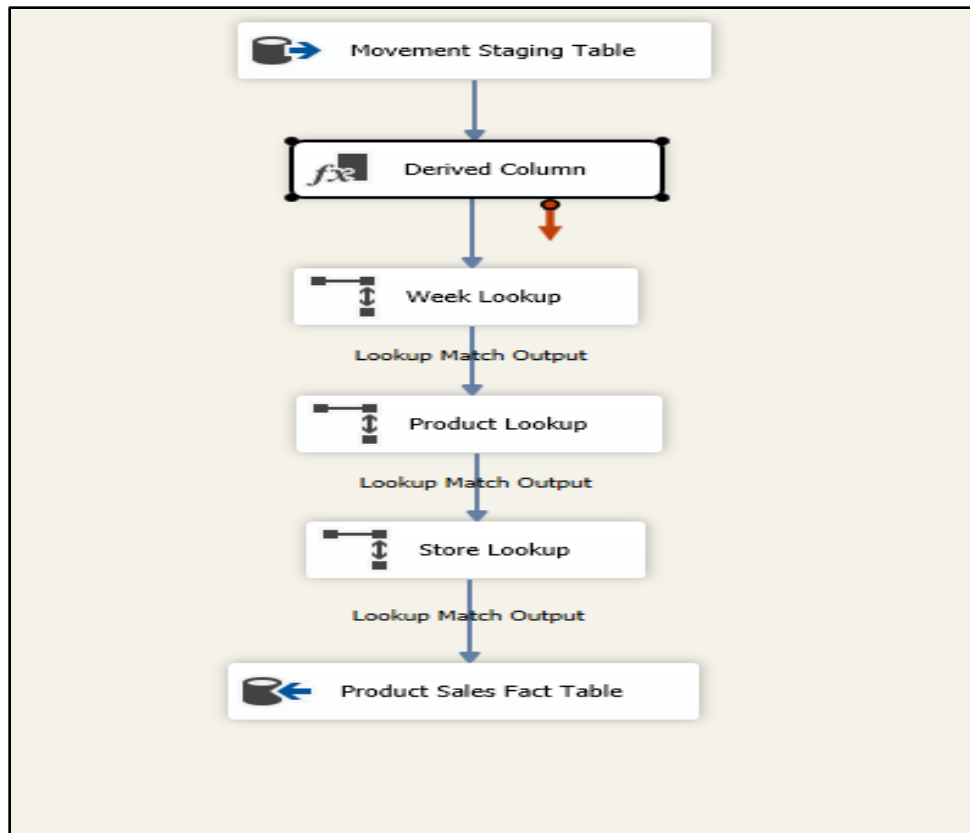
	storekey	productkey	datekey	retail_price	qty	profit	product_sales	sales_code
1	52	798	248	3.99	1	22.13	3.99	NULL
2	52	798	249	3.99	1	22.13	7.98	NULL
3	52	798	250	3.99	1	22.13	27.93	NULL
4	52	798	251	3.99	1	22.13	23.94	NULL
5	52	798	252	3.99	1	22.13	27.93	NULL
6	52	798	253	3.99	1	22.13	11.97	NULL
7	52	798	254	3.19	1	2.60	28.71	B
8	52	798	255	3.19	1	2.60	63.80	B
9	52	798	256	3.19	1	2.60	51.04	B
10	52	798	257	3.99	1	22.13	19.95	NULL
11	52	798	258	3.99	1	22.13	31.92	NULL
12	52	798	259	3.99	1	22.13	23.94	NULL
13	52	798	260	3.99	1	22.13	7.98	NULL
14	52	798	261	3.99	1	22.13	19.95	NULL
15	52	798	262	3.83	1	18.87	19.15	NULL
16	52	798	267	3.19	1	2.60	22.33	B
17	52	798	268	3.19	1	2.60	28.71	B
18	52	798	269	3.99	1	22.13	43.89	NULL
19	52	798	270	3.99	1	22.13	15.96	NULL

Promotional Sales Fact Table

Used cccoupon_category_derived as source to the fact table. Aggregated on the total coupons redeemed

Aggregate Function

Dim_Promotional Lookup

Dim Store Lookup
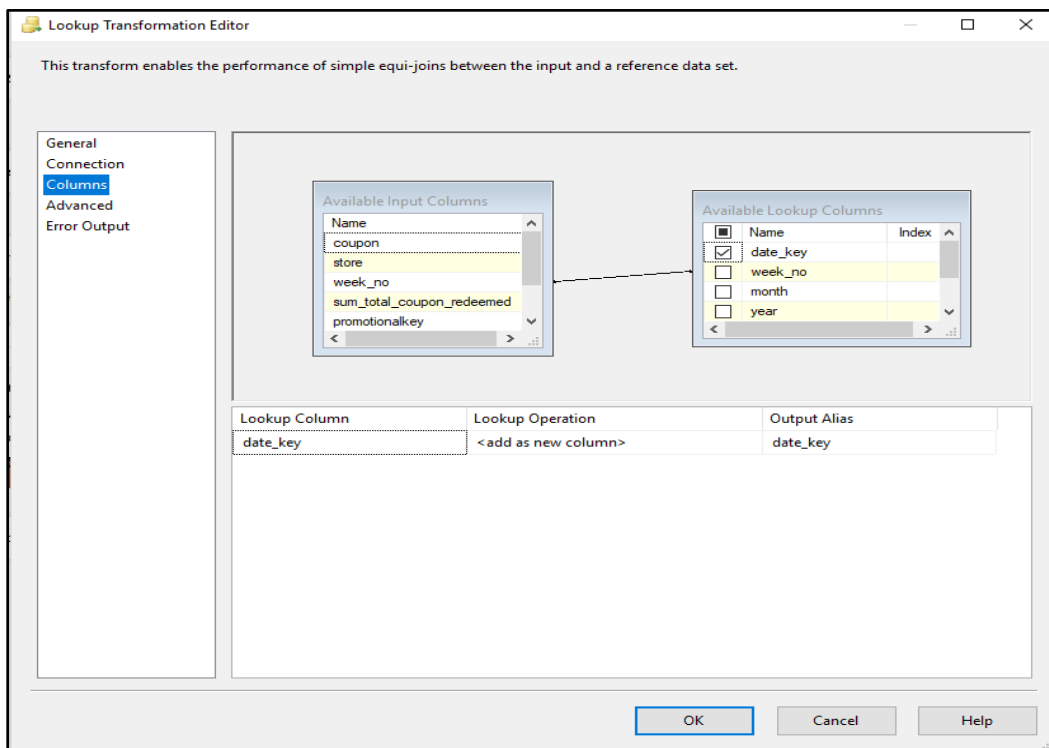
Dim Date Lookup
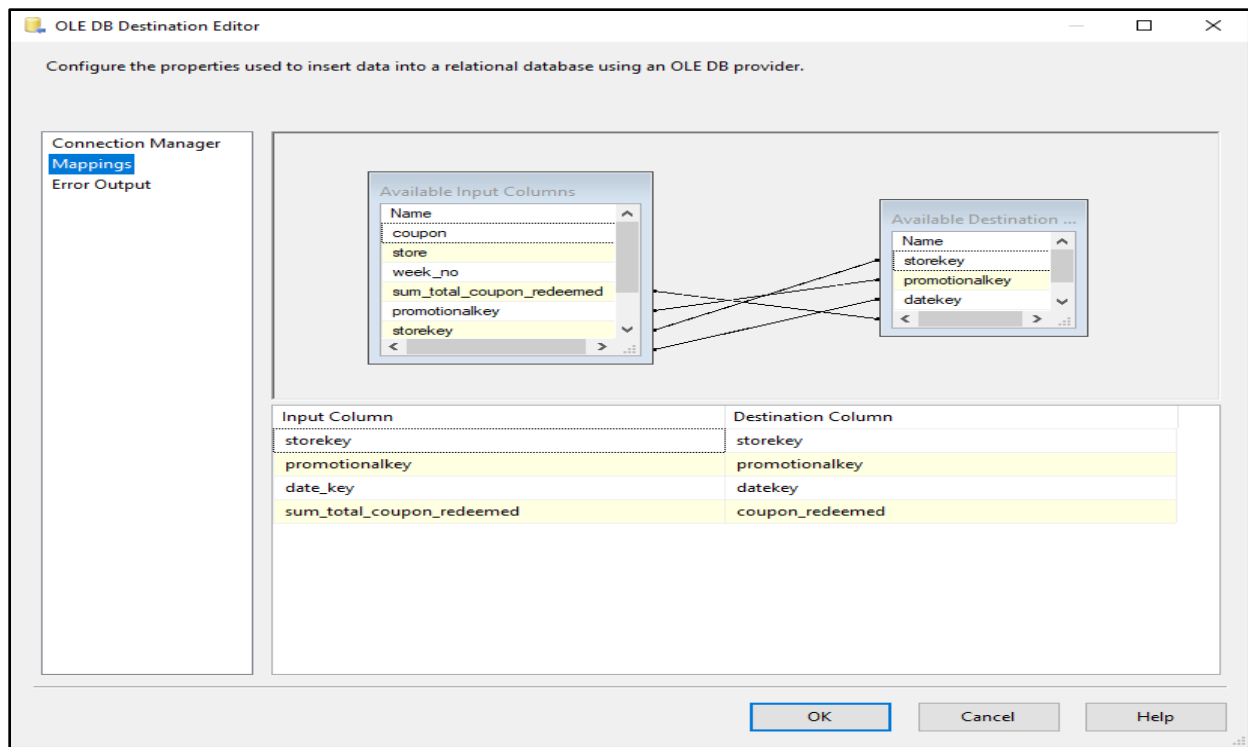
Mapping the respective keys

SQL Queries:

CREATE TABLE dbo.PromotionalFactTable (
[storekey] INT NOT NULL,
[promotional key] INT NOT NULL,
[datekey] INT NOT NULL,
[coupon_redeemed] DECIMAL(18, 0) NULL)
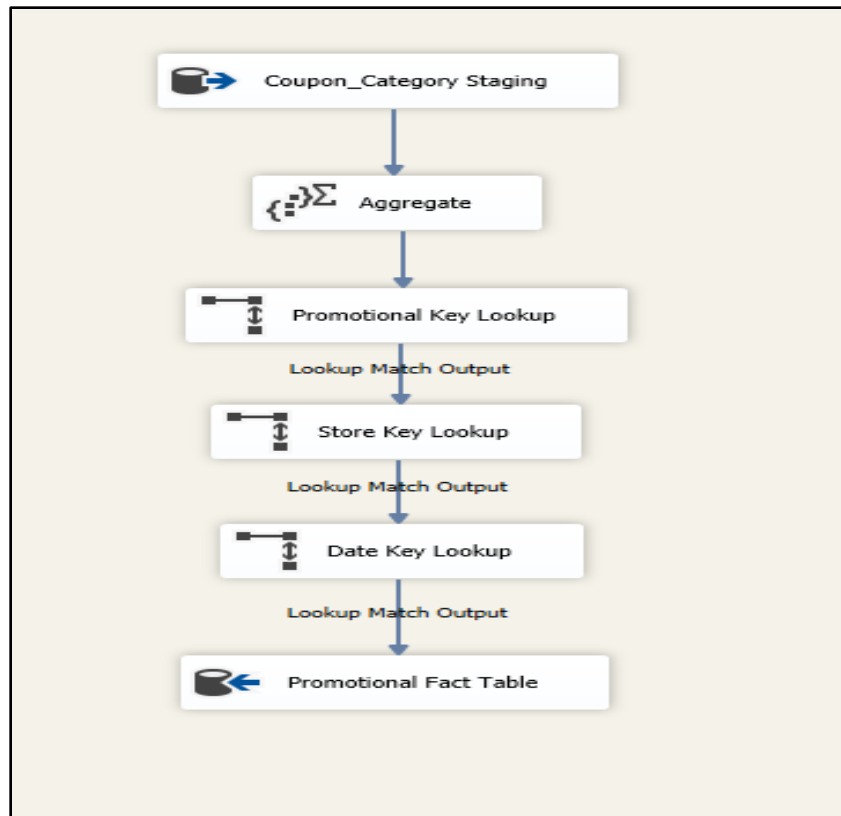
Promotional Fact Table Creation

Promotional Fact Table Data

6. BI Reporting

6.1. Reporting Plan

Business Intelligence (BI) is the process of retrieving ,analyzing and transforming data into meaningful information. It helps decision makers to identify goals to develop new business opportunities. There are many tools that help us visualize large data sets and produce meaningful results. For Dominik Fine Foods, we have used many reporting tools to present our analysis of the business questions.

Reporting Used	Business Question
SSRS	BQ 4
SSAS	BQ 1
SSRS+SSAS	BQ 3
Report Builder	BQ 2
SSRS	BQ 5

Business Question Answered with Reporting

6.2. Determining Target Reports to Answer the Business Questions

BQ1: Analyze the average profit margin of Bath Tissues across all the stores to determine store with low-profit margin

Reporting Tool used is **SSAS**

To generate this report , the dimensional and fact tables used are: Dim_Product and Dim_Store and ProductSalesFact. We chose the Product Category "Bath Tissues" which is defined in the table as "TTI". The Store_No is selected from the Dim_Store . The low profit margin is a derived column in the Fact table across the entire duration. We have plotted the average profit against the store no to determine the performance of bath tissue in each store. We have also used a chart type to help understand the business question better and provide better visualizations.

BQ2: What are the total sales of Beer during the holiday season?

Reporting Tool Used is **Report Builder**

To generate this report, the dimensional and fact tables used are: Dim_Product , Dim_Date and Product Sales Fact Table. From Dim_Product, the *product_category* "Beer" is selected which is defined in the table as 'BER'. From Dim_Time, the *week_no* and *special_event_name* is selected to obtain results for the holiday season. From Product Sales Fact Table, the *total_sales* are taken, which will be the sum of sales for that week. For better visualizations, we have taken a bar chart to help understand the trend of sale of beer during the holiday season.

BQ3. Which pricing tier produces the highest bottled juice sales during Christmas Week?

Reporting Tool Used is **SSRS+SSAS**

For this report, we created a cube using SSAS and then created a report using SSRS. To generate this report, the dimensional and fact tables used are: Dim_Product , Dim_Date, Dim_Store and Product Sales Fact Table. From Dim_Product, the *product_category* "Bottled Juices" is selected which is defined in the table as 'BJC'. From Dim_Time, the *special_event_name* is selected to obtain results for the Christmas Week. From Dim_Store, *Pricing_Tier* was chosen. From Product Sales Fact Table, the *total_sales* are taken, which will be the sum of sales for that week. For better visualizations, we have taken a bar chart to help understand the trend in sale of bottle juice among the different pricing tier.

BQ4.Which Bath Soap product has a decreasing sales trend across different zones?

Reporting Tool Used is **SSRS**

To generate this report, the dimensional and fact tables used are: Dim_Product , Dim_Date, Dim_Store and Product Sales Fact Table. From Dim_Product, the *UPC_No* is selected for product category Bath Soap which is defined as 'BAT'. From Dim_Time, the *Week_No* is selected to obtain results between certain weeks. From Dim_Store, *Zone_No* was chosen. From Product Sales Fact Table, the *total_sales* are taken, which will be the sum of sales for that UPC_No across different zones. For better visualizations, we have taken a histogram to help understand the trend in sale of a particular UPC in Bath Soap across different zones.

BQ5. What category of coupons are redeemed the most for the entire duration?

Reporting Tool Used is **SSRS**

To generate this report, the dimensional and fact tables used are: Dim_Store , Dim_Date, Dim_Promotional and Promotional Fact Table. From Dim_Time, the *Week_No* is selected to obtain results across all weeks. From Dim_Store, *Store_No* was chosen to fetch coupons across all stores. From Dim_Promotional, Coupon_Cat is used to get the different coupon category that have been issued. From Promotional Sales Fact Table, the *coupons_redeemed* are taken, which will be the sum of coupon_redeemed which is group by *coupon_cat* across all stores for the entire duration. For better visualizations, we have taken a bar chart to help understand the coupons redeemed for the entire duration.

6.3. Mapping Data Marts to Report Attributes

BQ1: Analyze the average profit margin of Bath Tissues across all the stores to determine store with low-profit margin
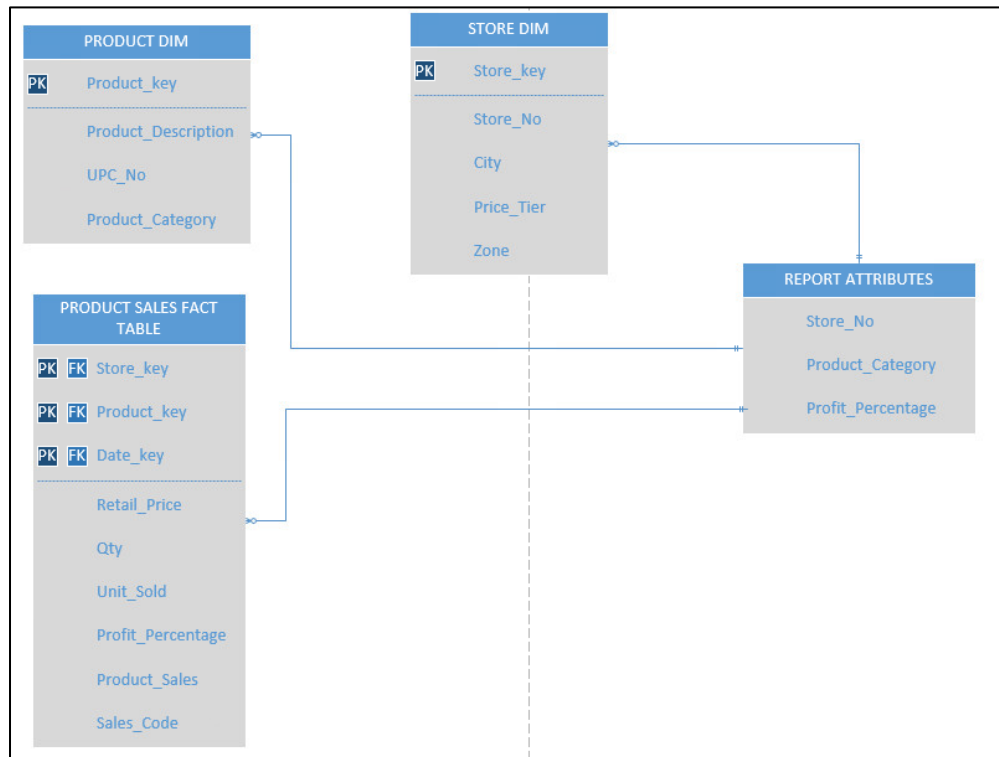
BQ2: What are the total sales of Beer during the holiday season?
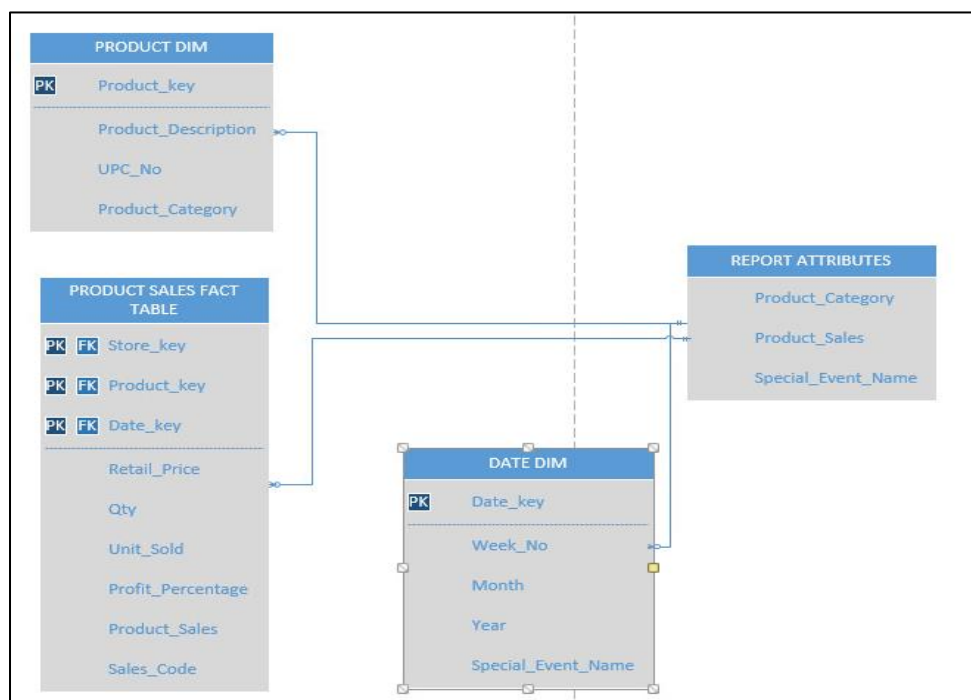
BQ3. Which pricing tier produces the highest bottled juice sales during Christmas Week?
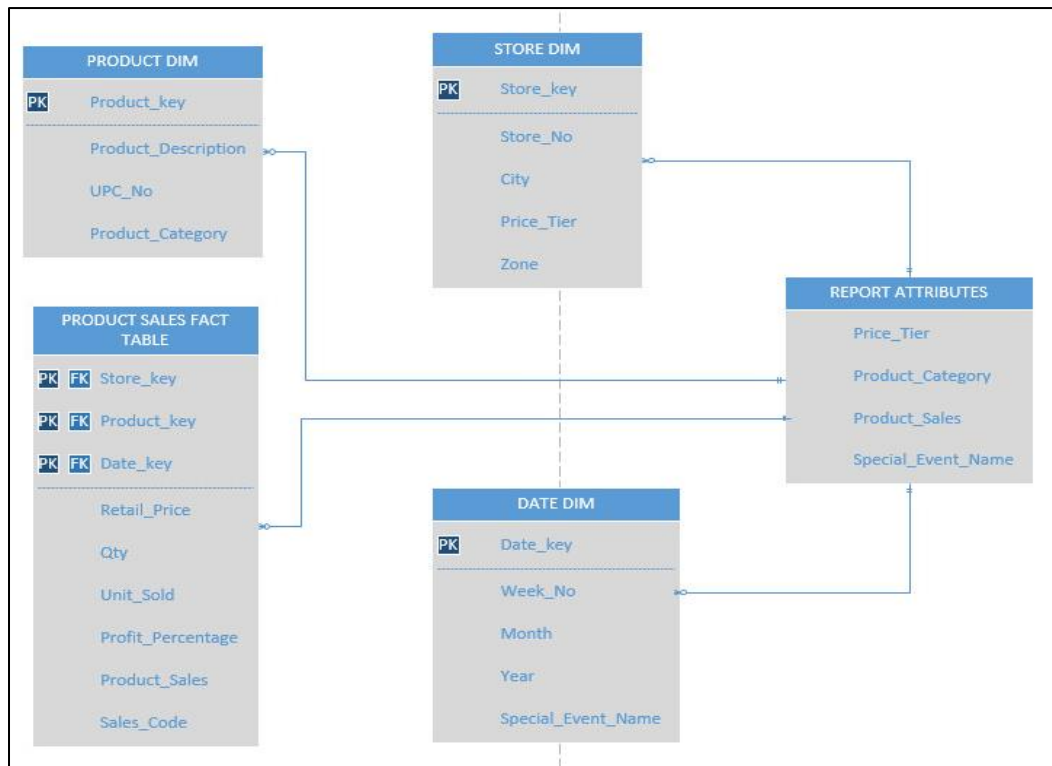
BQ4.Which Bath Soap product has a decreasing sales trend across different zones?
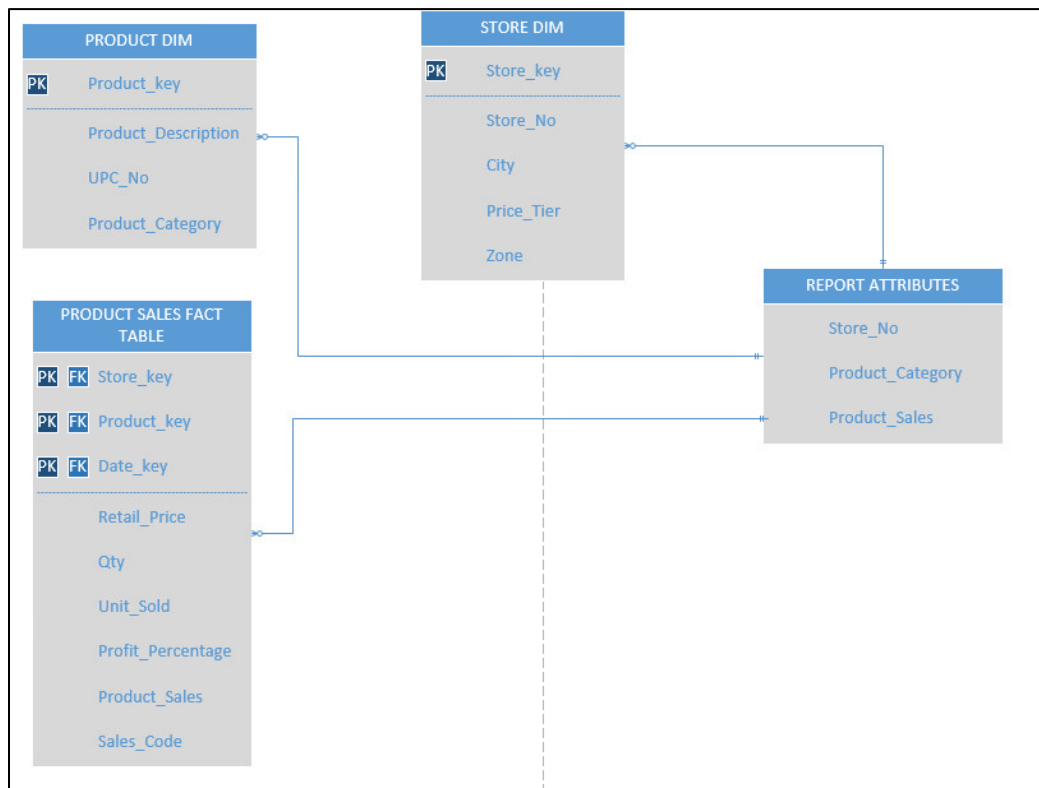
BQ5. What category of coupons are redeemed the most for the entire duration?
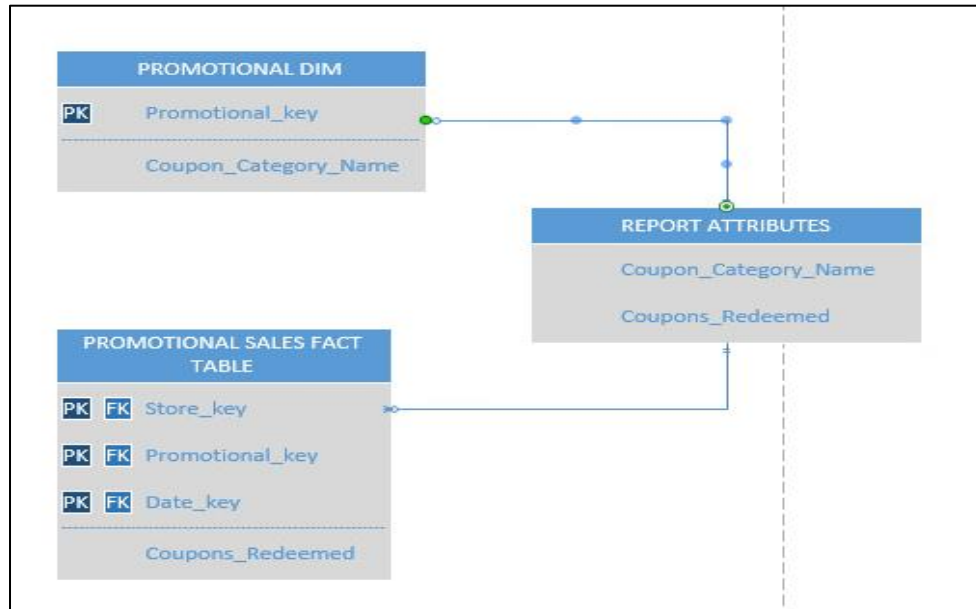
6.4. Cube Creation Using SSAS and SSRS for BQ3

BQ3. Which pricing tier produces the highest bottled juice sales during Christmas Week?

Creating of Data Source
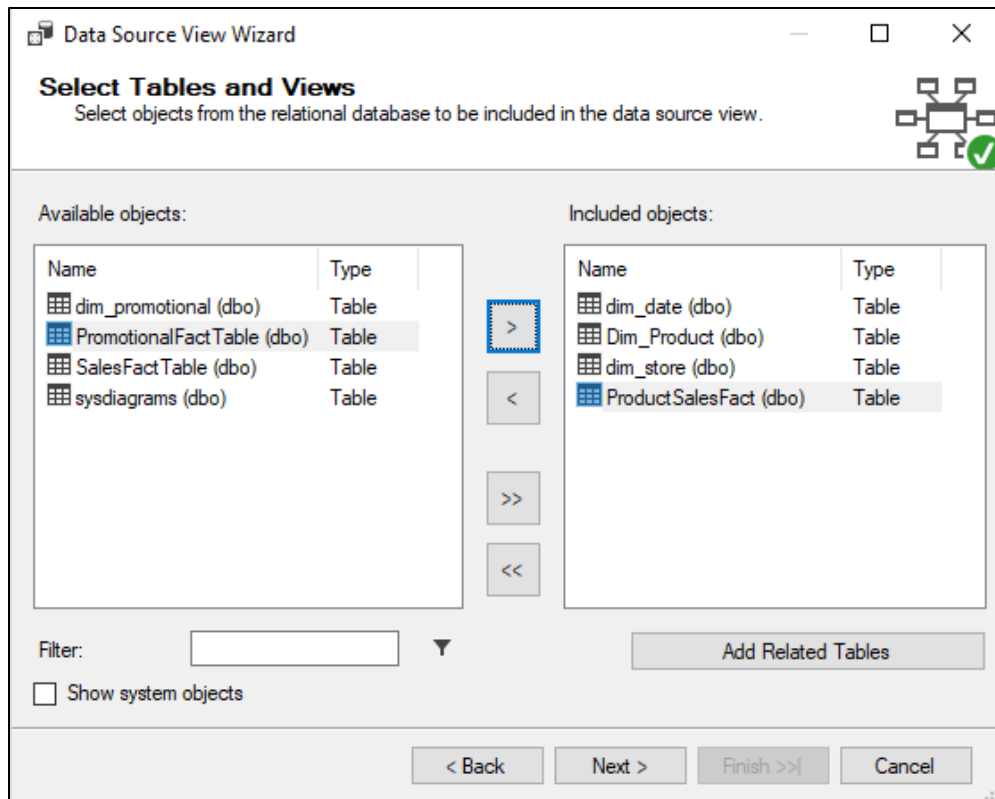
Creation of Data Source View
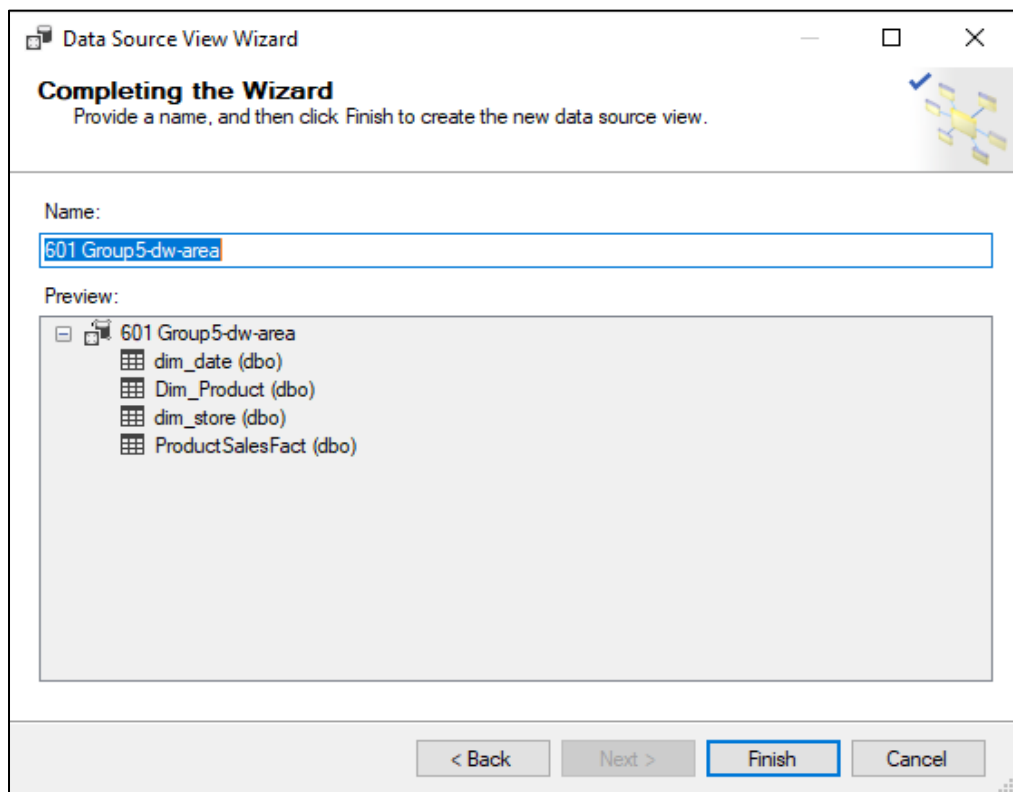
Completed the Creation of Data Source View
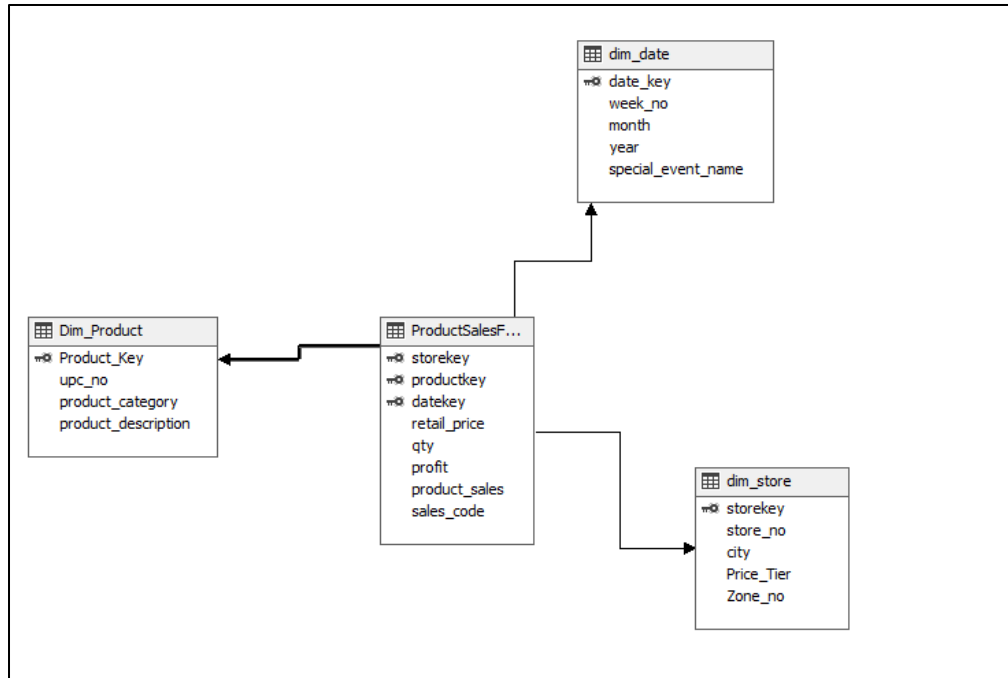
Data Source View Results
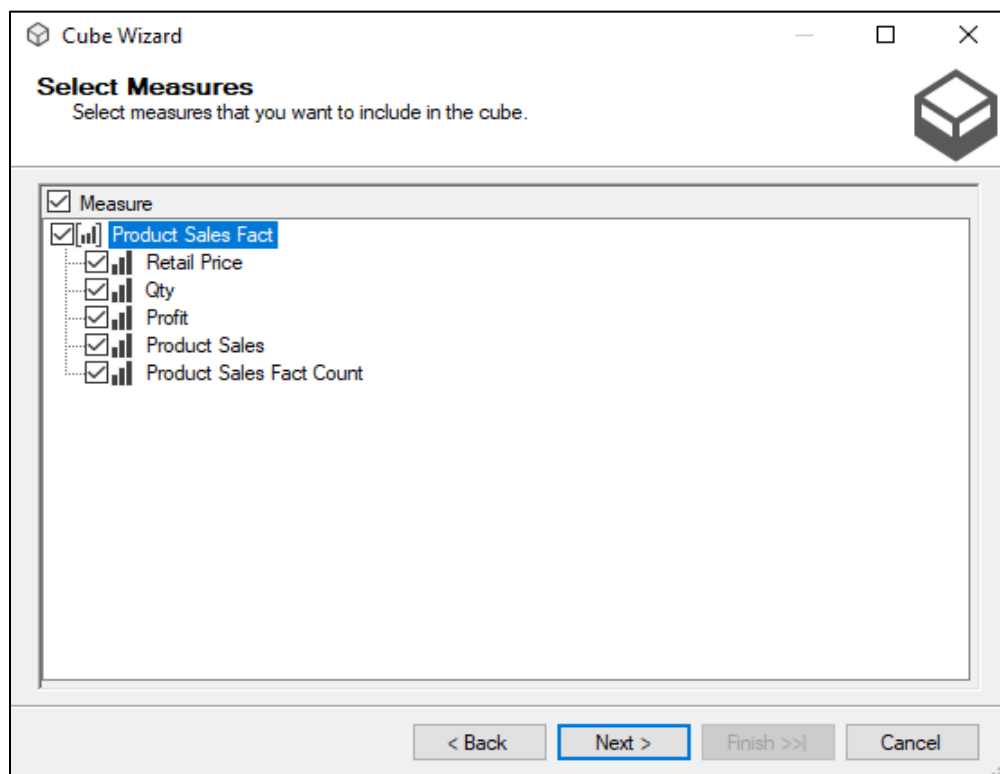
Selection of Measures
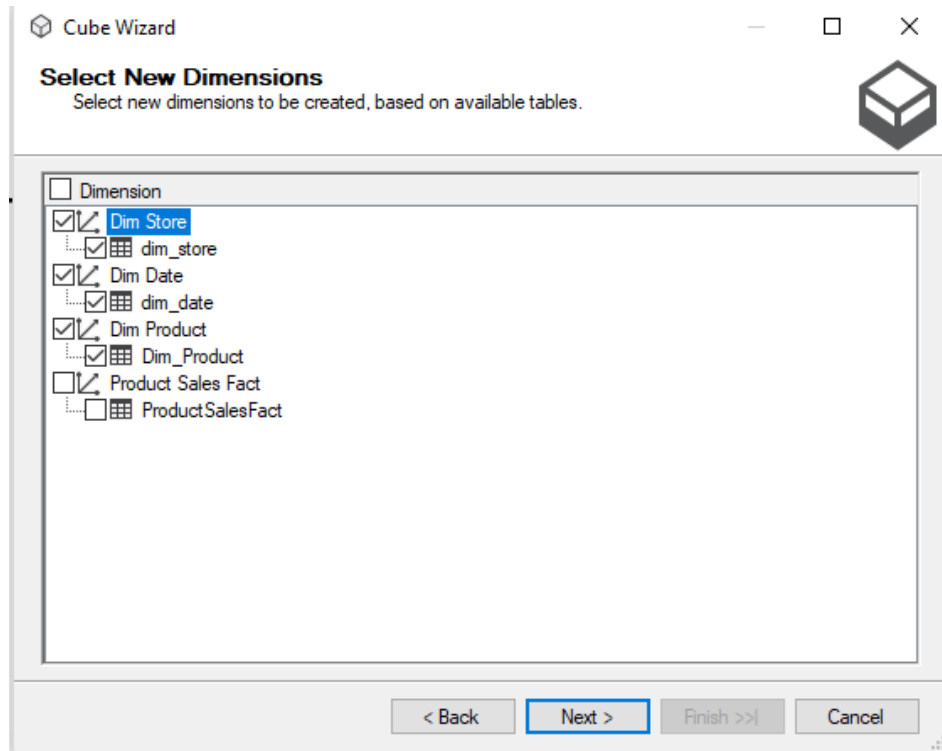
Selection of Dimensions
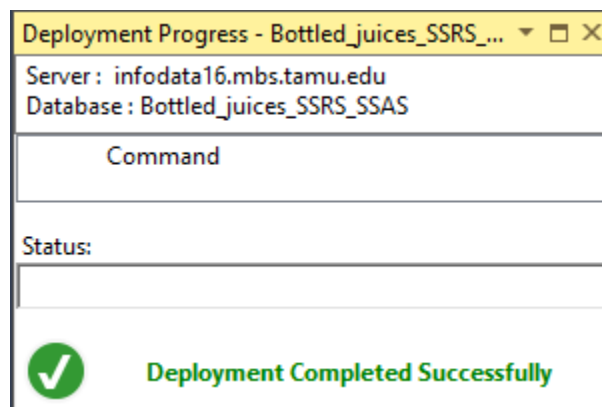
Successful Deployment of Cube
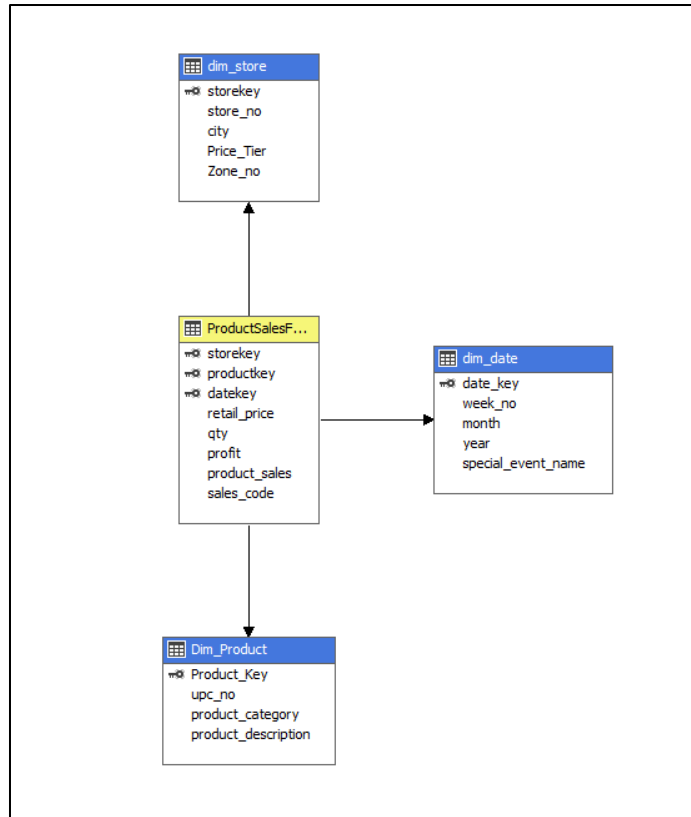
Cube Structure

Process of Creating the Report:
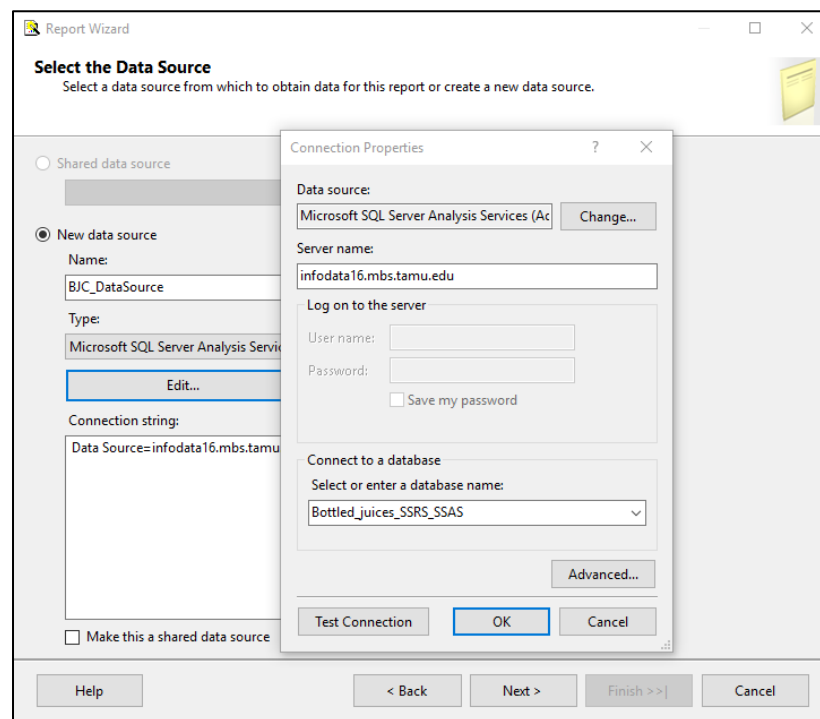
Selection of Data Source
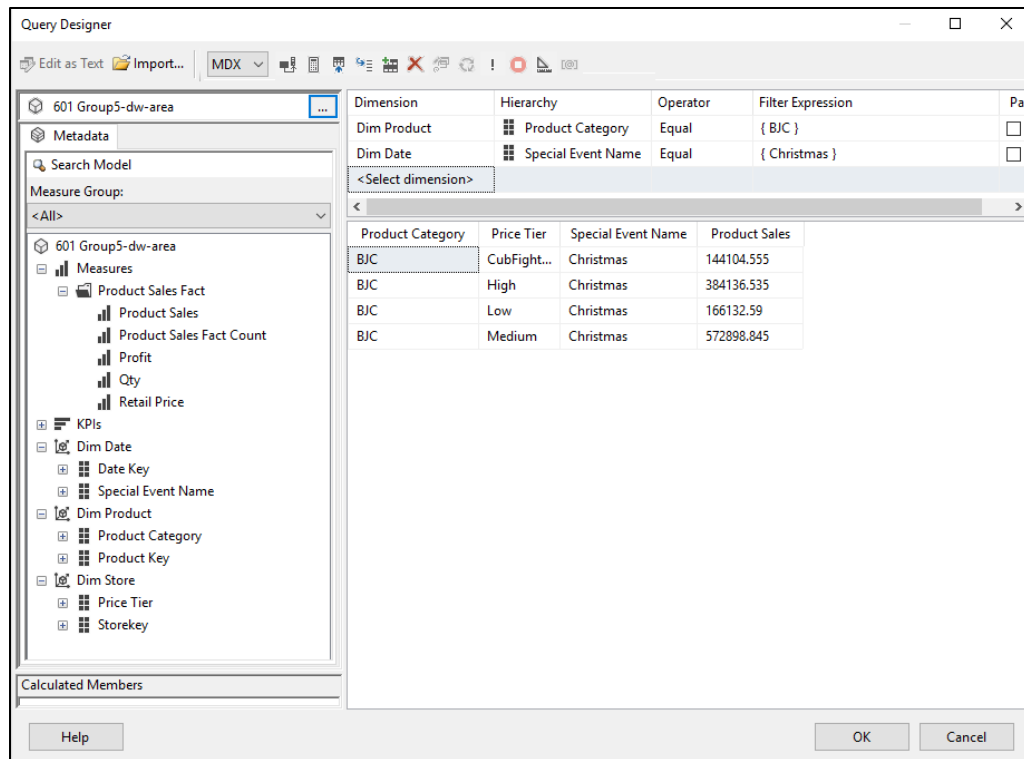
Selecting Report Attributes
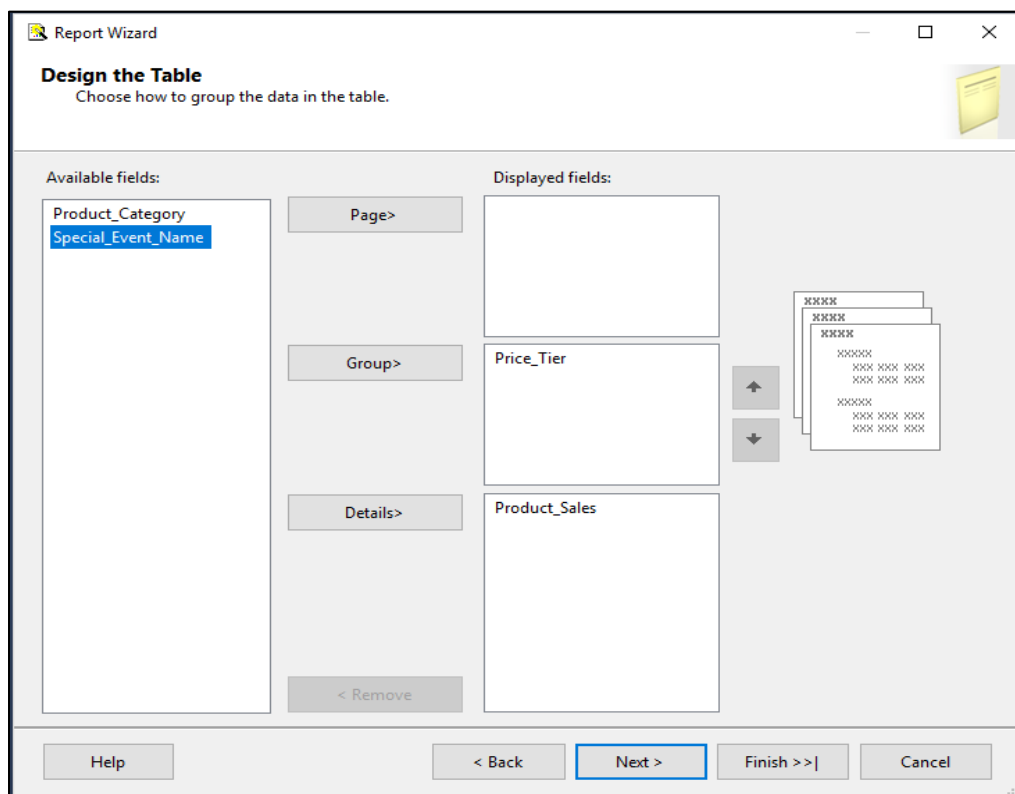
Creation of Report Structure

Price_Tier_that_produces_highest_b
ottled_juice_sales_during_Christmas

Price Tier	Product Sales
⊟ CubFighter	
	144104.555
⊟ High	
	384136.535
⊟ Low	
	166132.59
⊡ Medium	
	572898.845
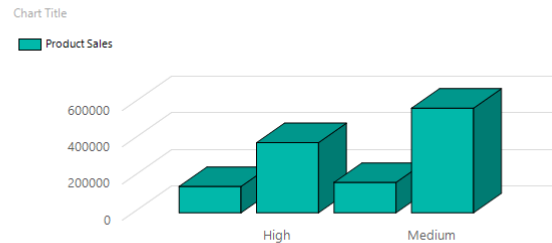
Report Deployed on the Web Server

<u>*Results from the Report:*</u>

We can see from the report that the product sales of bottled juices during Christmas is the highest in the "Medium" category of the pricing tier followed by "High" pricing tier. Also, this can be used as a good marketing strategy for comparing different products across the different pricing tiers and help prepare certain promotional strategies.

6.5. Cube Creation Using SSAS for BQ1

BQ1: Analyze the average profit margin of Bath Tissues across all the stores to determine store with low-profit margin
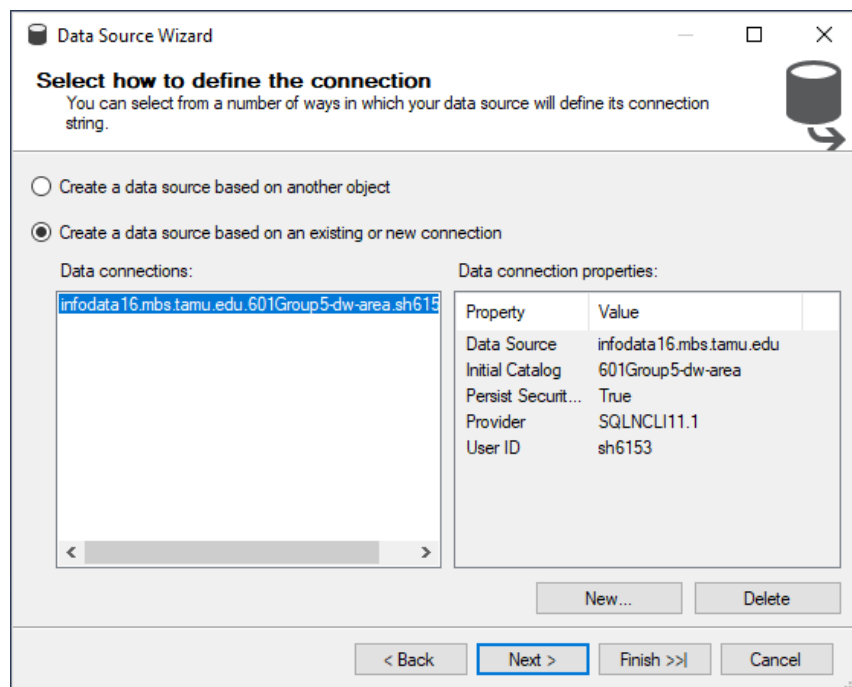
Creating a Data Source
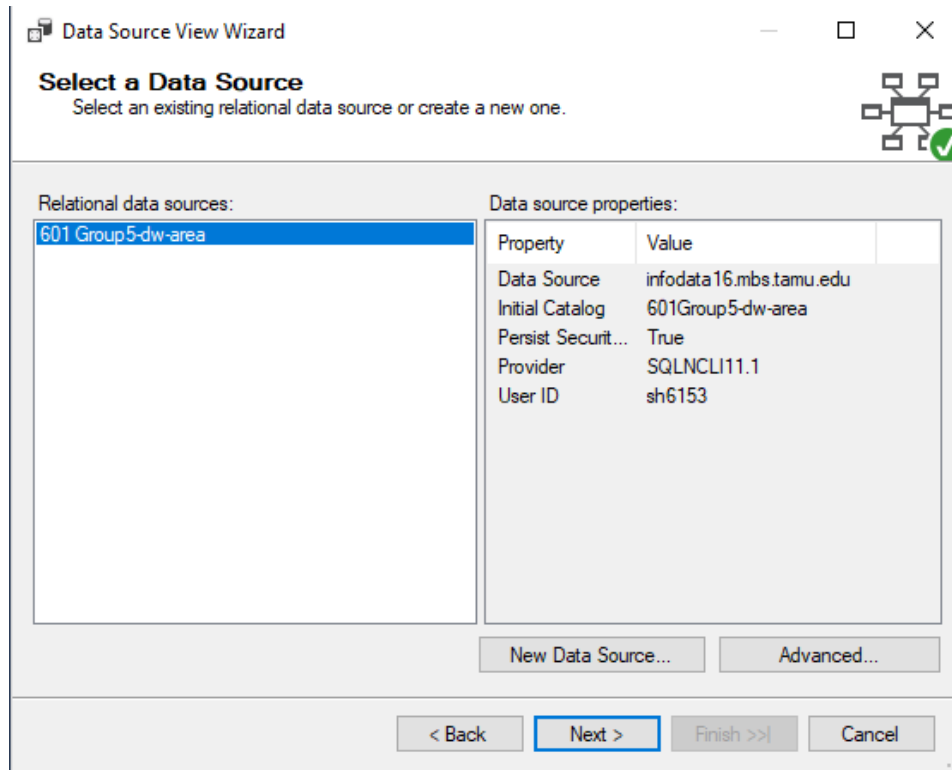
Selecting a Data Source
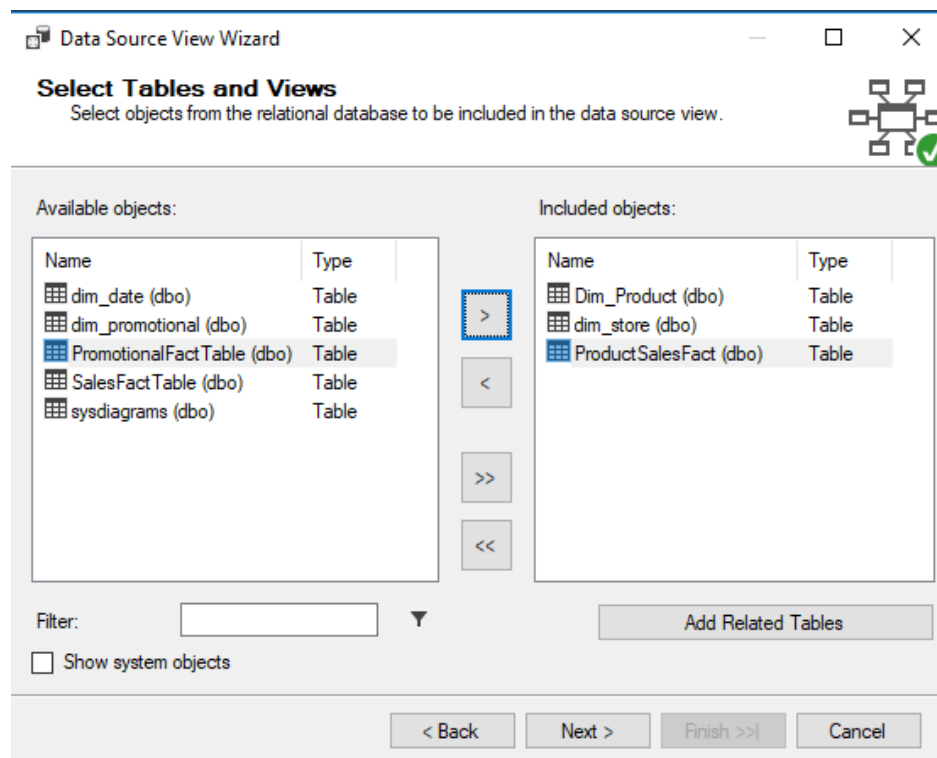
Creation of Data Source View
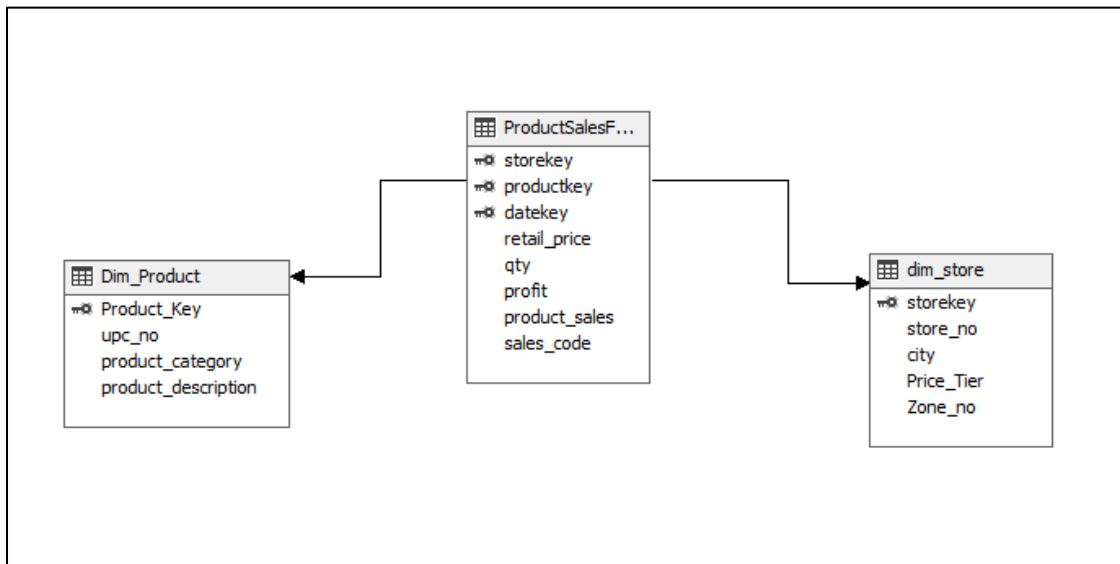
Data Source View Results

Creating a Cube
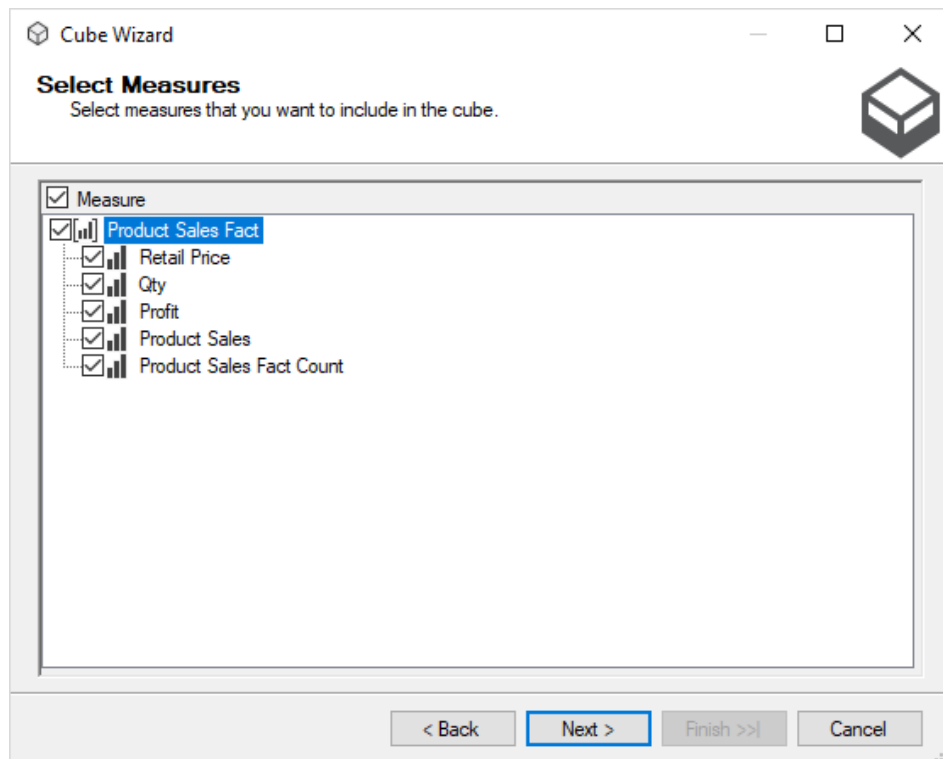
Selecting the Measures
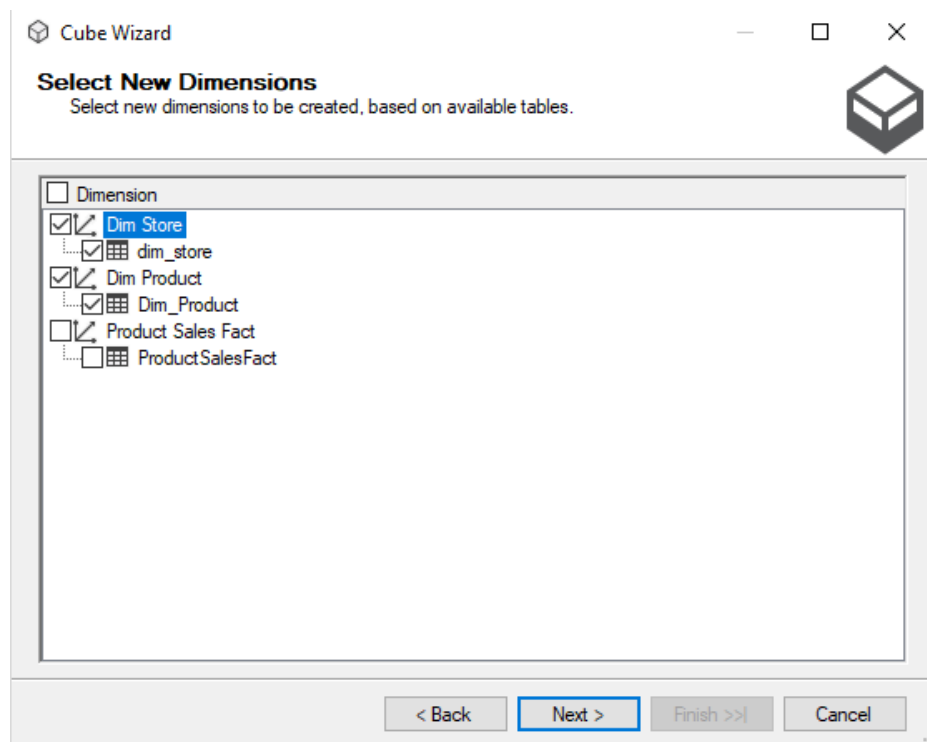
Selecting the Dimensions
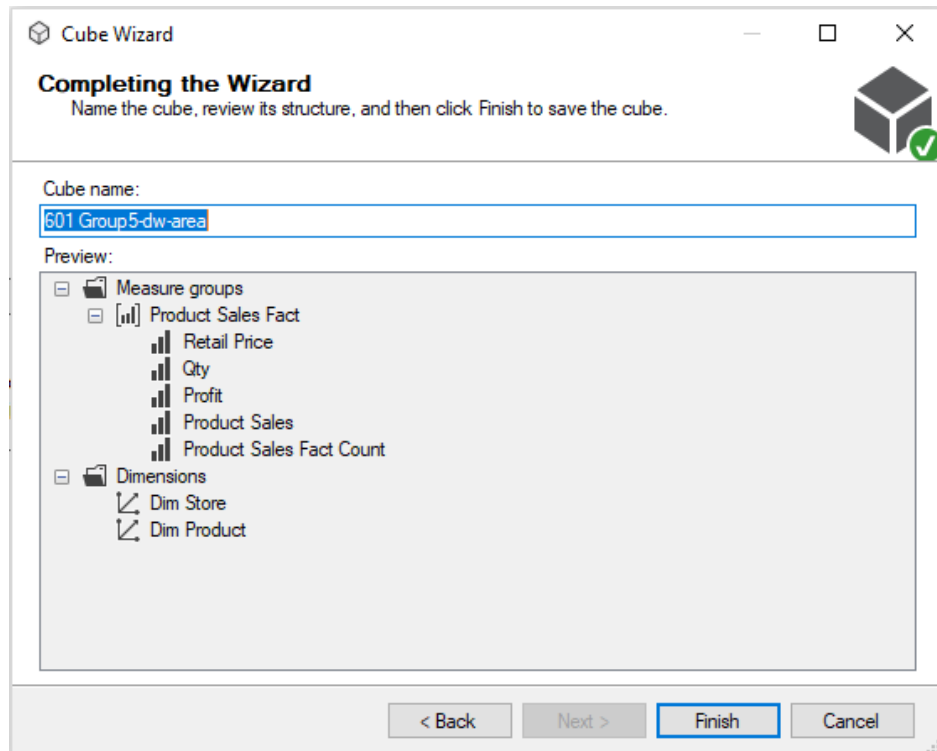
Completed Cube Creation
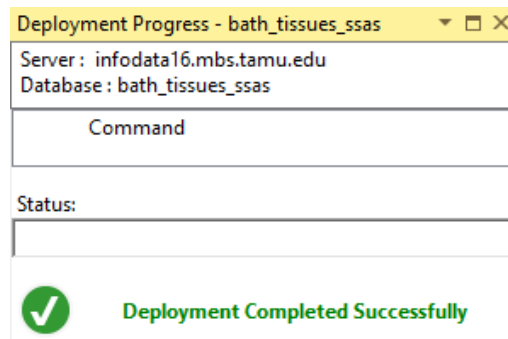
Deployed the Cube

Cube Structure

Cube Data

Creating a Report:

Creating the report structure

Store No	Profit
100	178420.4503
101	164756.455
102	131575.4903
103	127001.5968
104	151207.8397
105	134455.3114
106	157067.8484
107	151153.1757
109	188745.4482
110	158573.1236
111	168006.18
112	139760.9591
113	159861.9031

Report in Tabular Format
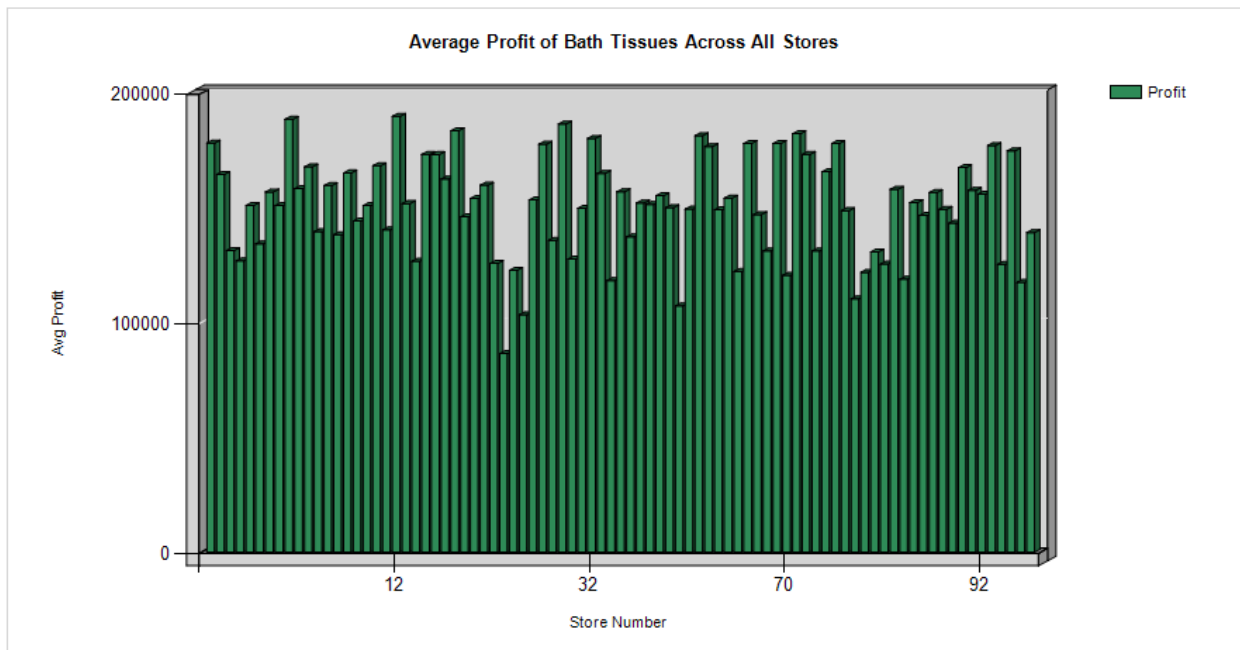
Graphical Report created on the Server

Results from the Report:

We can see from the report the product sales of bath tissues across the stores. This can help in maximizing the profit from the sale of tissues in a certain store. This can also help to strategize the inventory depending on the sales of the particular store. Store targeted promotions can also be made by looking at the average profit trend in each store.

6.6. Report Creation Using Report Builder for BQ2
BQ2: What are the total sales of Beer during the holiday season?
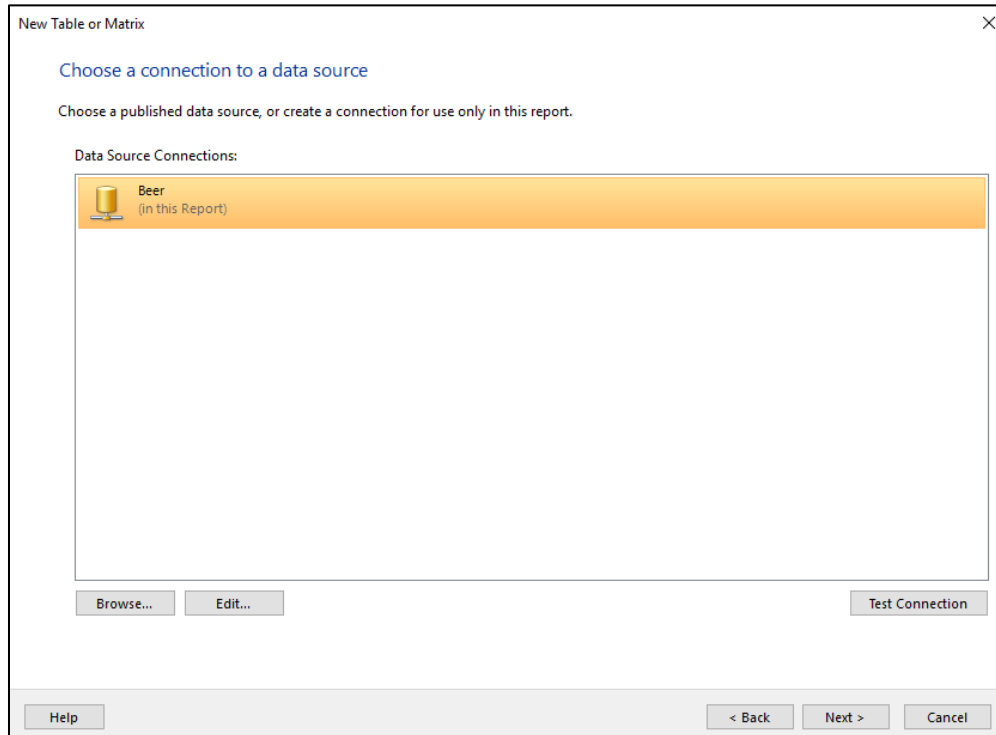
Creation of a Data Source
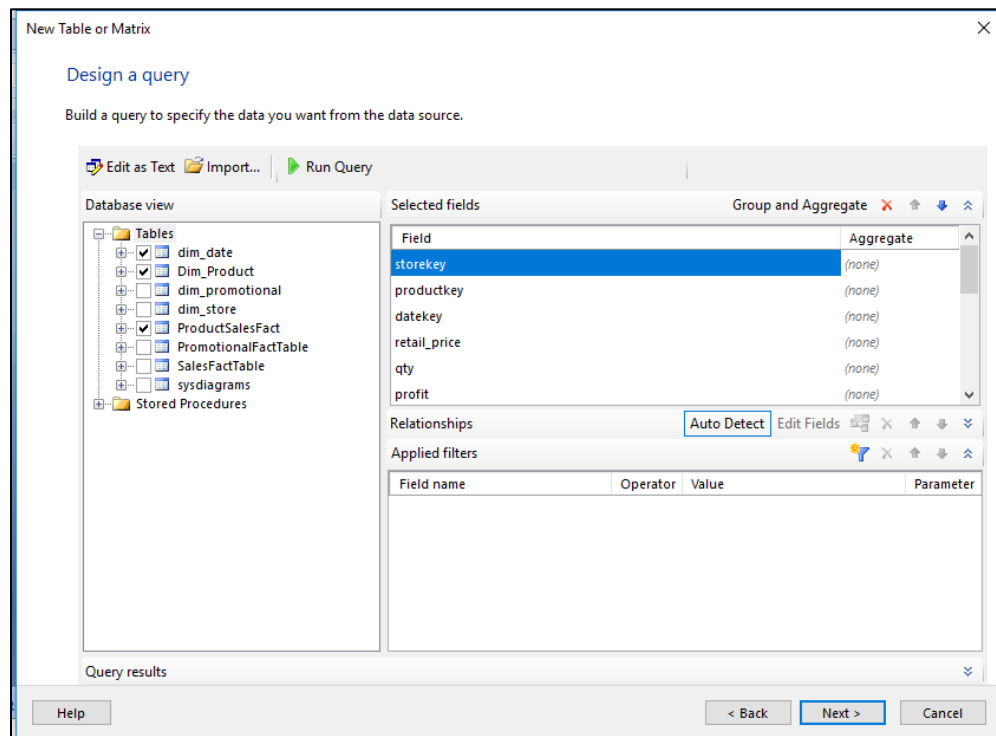
Creating the Query for the Report
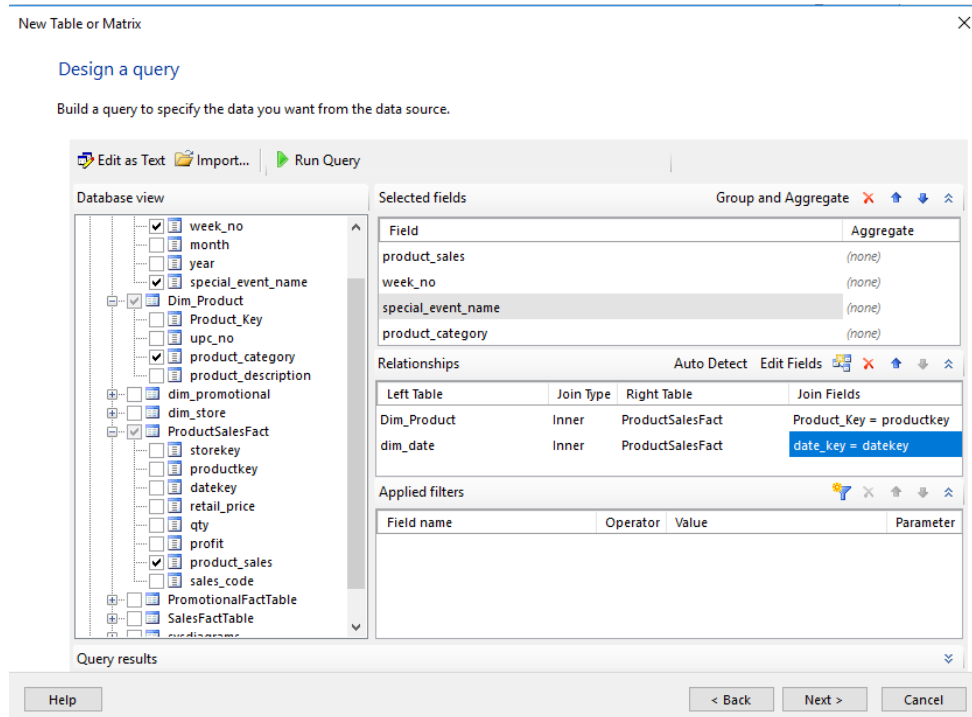
Select the fields and Relationships

SQL Query Used:

SELECT
  Dim_Product.product_category
  ,dim_date.week_no
  ,dim_date.special_event_name
  ,sum(ProductSalesFact.product_sales) as sum_product_sales
FROM
  Dim_Product
INNER JOIN ProductSalesFact
    ON Dim_Product.Product_Key = ProductSalesFact.productkey
INNER JOIN dim_date
    ON dim_date.date_key = ProductSalesFact.datekey
WHERE   dim_date.special_event_name   IN   ('Halloween',   'Thanksgiving','Christmas','New-Year','4th   of   July',   'Memorial   Day','Labor   Day','Presidents   Day','Easter')   AND
Dim_Product.product_category = 'BER'
GROUP BY dim_date.special_event_name, Dim_Product.product_category ,dim_date.week_no
ORDER BY dim_date.week_no

Test the Query

Creating the Report:

Creating the Report Structure

product category	special event name	week no	sum product sales
⊟ BER	⊟ 4th of July	95	563063.8400
		147	394775.2100
		199	484573.4100
		251	297416.8400
	⊟ Christmas	119	250055.1200
		172	349956.6000
		224	143310.1500
		276	182394.4600
	⊟ Easter	133	265257.6600
		185	203677.0000
		238	235662.1000
	⊟ Halloween	112	273100.3900
		164	268069.2500
		216	134634.0800

Final Report in Tabular Format
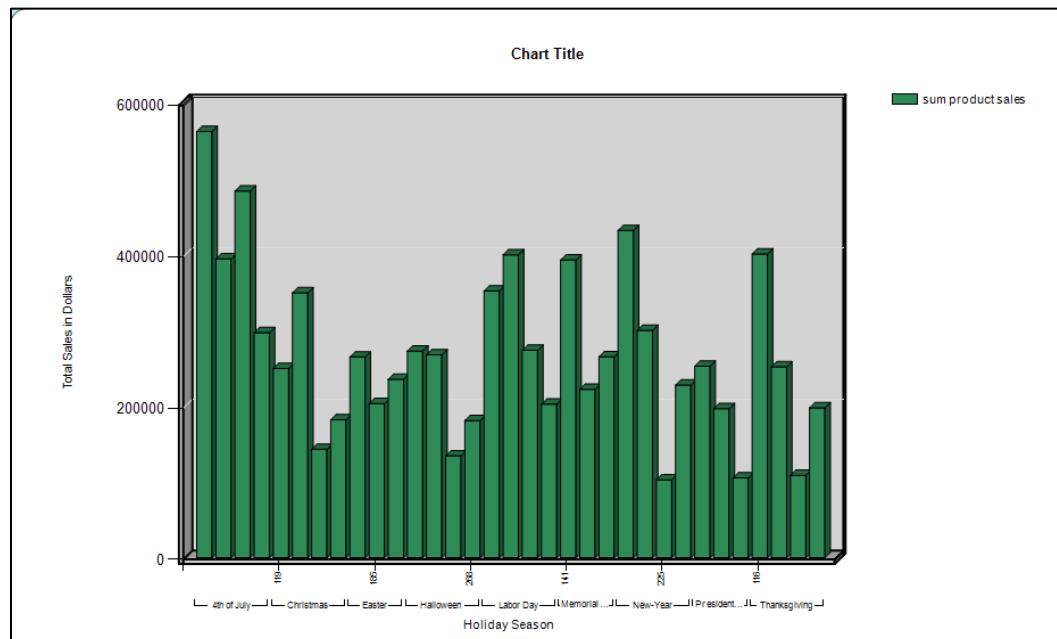
Final Report in Bar Chart

Results from the Report:

We can see from the report the product sales of Beer during the holiday season. This can also help to strategize the inventory depending on the sales of the beer during this peek period. Depending on the holiday season, promotions can also be planned for the beer categories depending on the sales of a particular beer product.

6.7. Report Creation Using Report Builder for BQ2
BQ4.Which Bath Soap product has a decreasing sales trend across different zones?

Selection of Data Source

***Design the Query***

Query Successfully Executed

SQL Query Used:

SELECT   Dim_Product.upc_no, sum(ProductSalesFact.product_sales) as Product_sales, dim_store.Zone_no
FROM   dim_date INNER JOIN
*   ProductSalesFact INNER JOIN*
*   dim_store ON ProductSalesFact.storekey = dim_store.storekey INNER JOIN*
*   Dim_Product ON ProductSalesFact.productkey = Dim_Product.Product_Key ON*
dim_date.date_key = ProductSalesFact.datekey
WHERE Dim_Product.product_category = 'BAT' and dim_date.week_no BETWEEN 273 AND 277
GROUP BY Dim_Product.upc_no, dim_store.Zone_no
ORDER BY Product_sales DESC, dim_store.Zone_no ASC
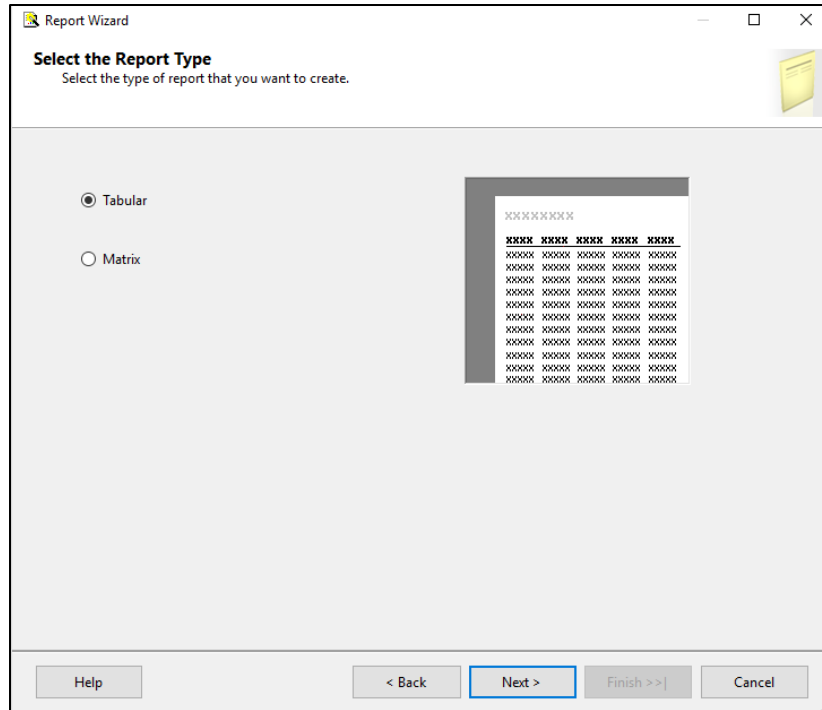
Selecting the Report Structure
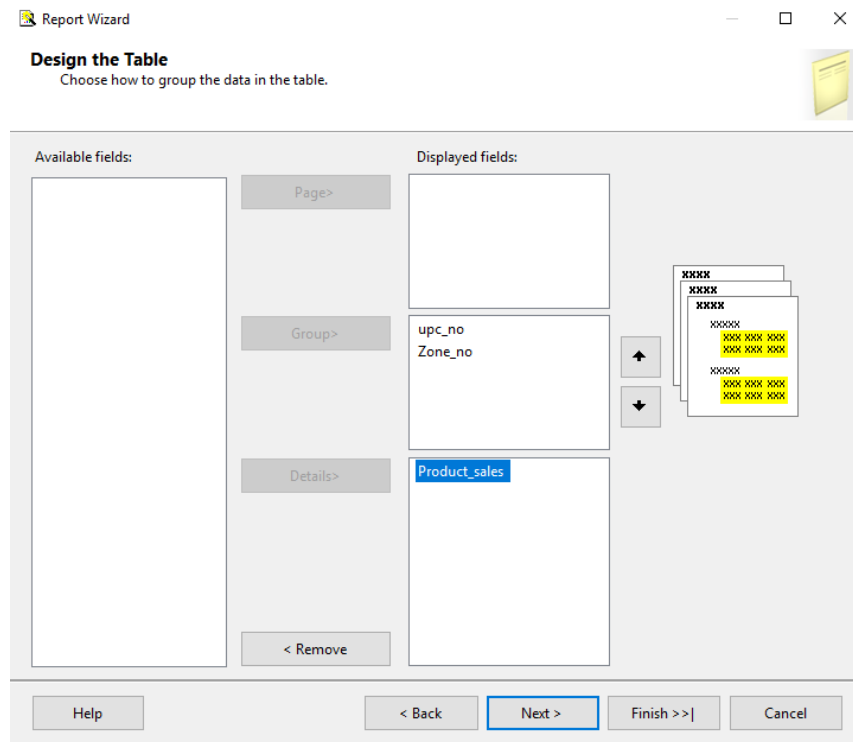
Selecting the Report Attributes
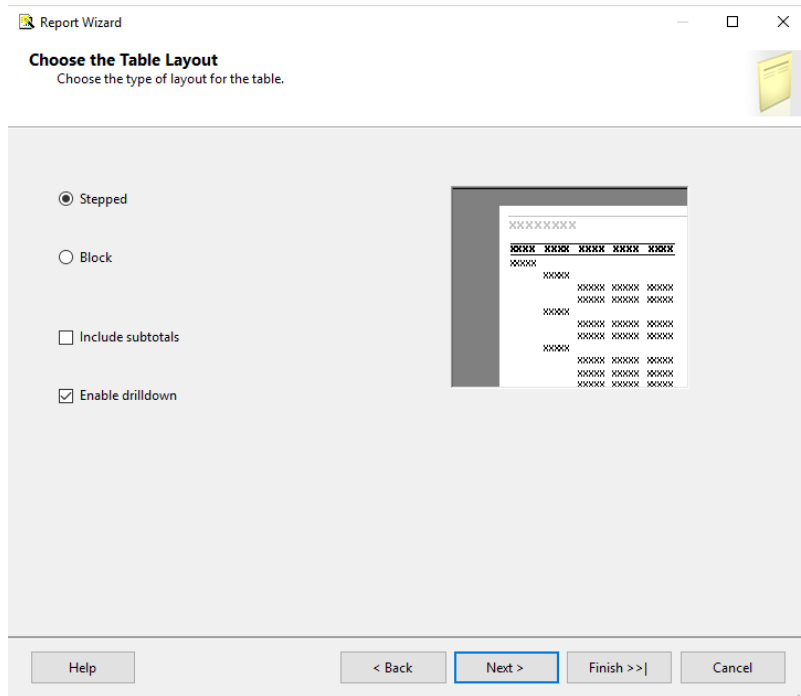
Selecting the Report Settings
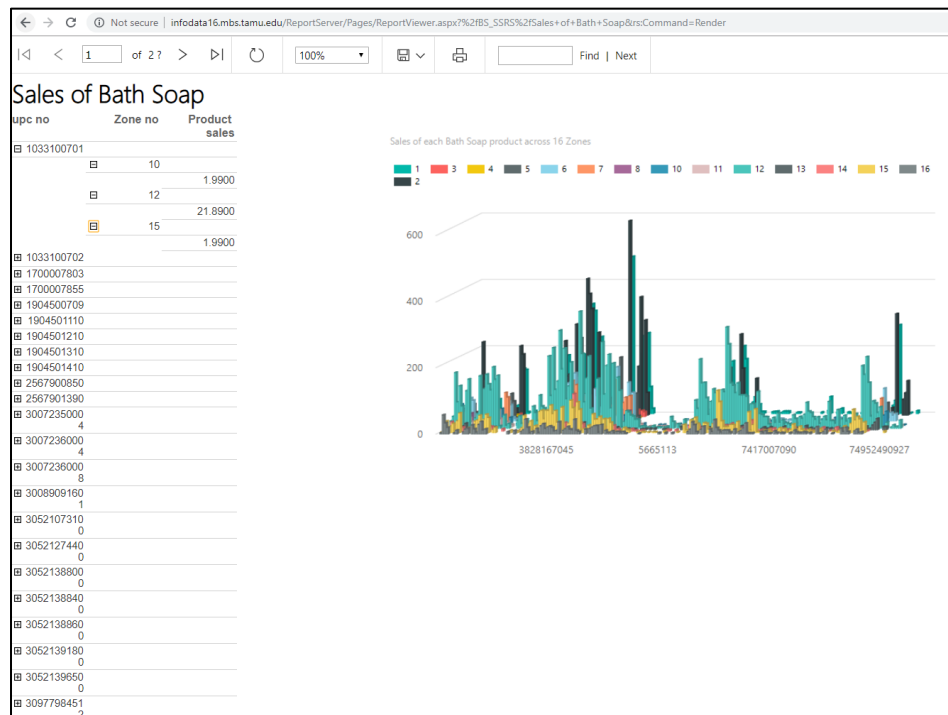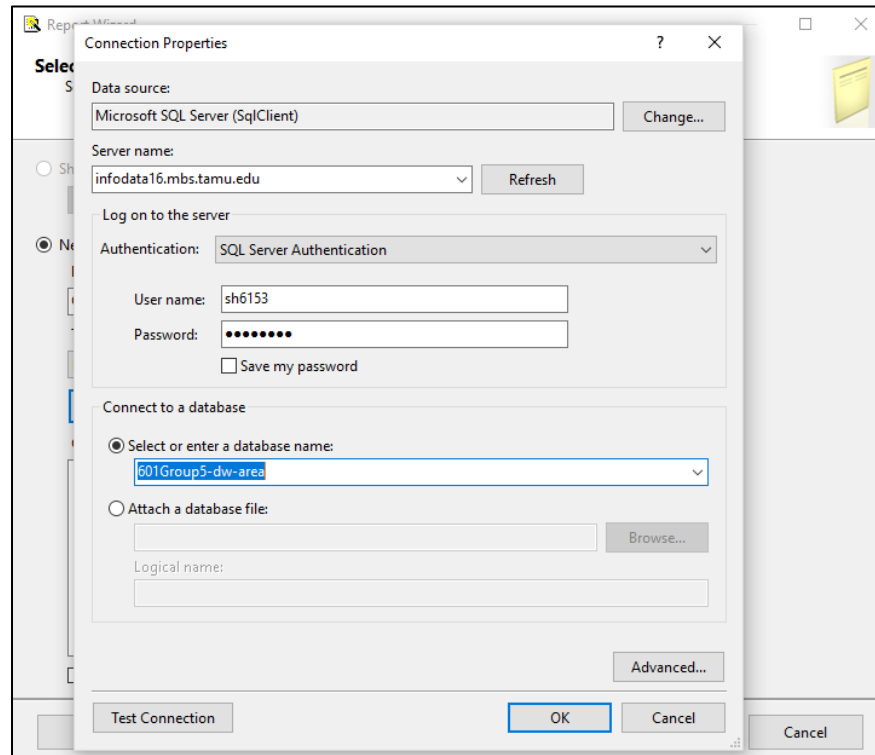

Final Report Deployed on the Web Server

Results from the Report:

We can see from the report the product sales of Bath Soap across different store zones. This can also help to strategize the brand of a bath soap zone-wise. It help to plan what brand of bath soap
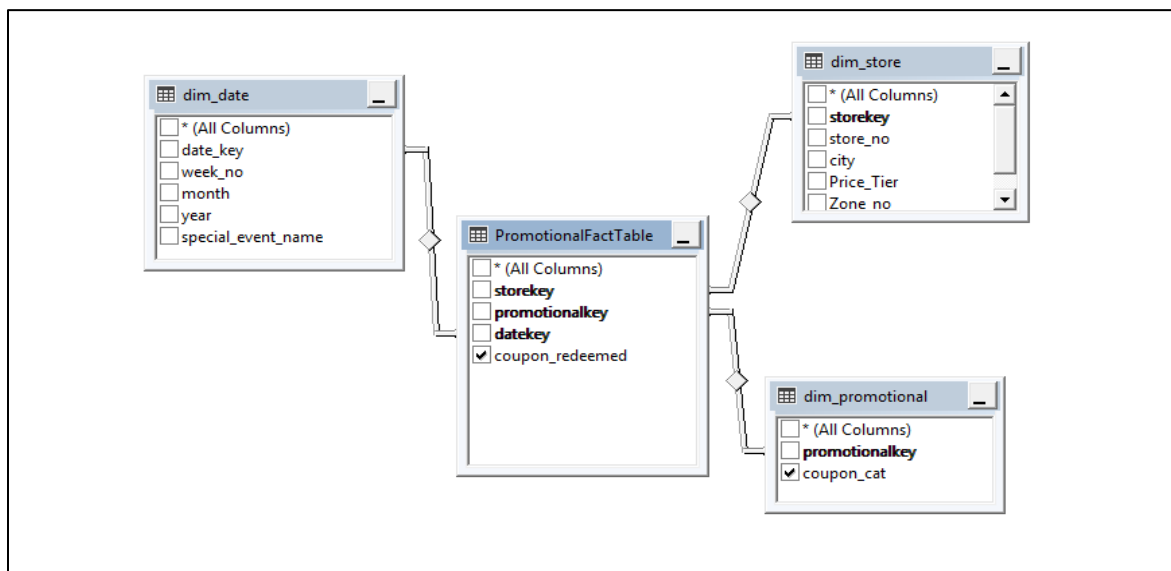
needs to be shipped , reducing the logistics and inventory. Coupons can also be issued based on a particular zone and promotions can be targeted at such zones.

6.8. Report Creation Using Report Builder for BQ2

BQ5. What category of coupons are redeemed the most for the entire duration?


Creating the Data Source

***Design the Query***

Query Successfully Executed

<u>SQL Query Used:</u>

SELECT Dim_Promotional.coupon_cat, sum(PromotionalFactTable.coupon_redeemed) as Coupon_Redeemed
FROM PromotionalFactTable INNER JOIN
Dim_Promotional ON Dim_Promotional.Promotional_Key = PromotionalFactTable.PromotionalKey INNER JOIN
Dim_Date ON PromotionalFactTable.datakey = Dim_Date.date_key INNER JOIN
Dim_Store ON PromotionalFactTable.storekey = Dim_Store.storekey
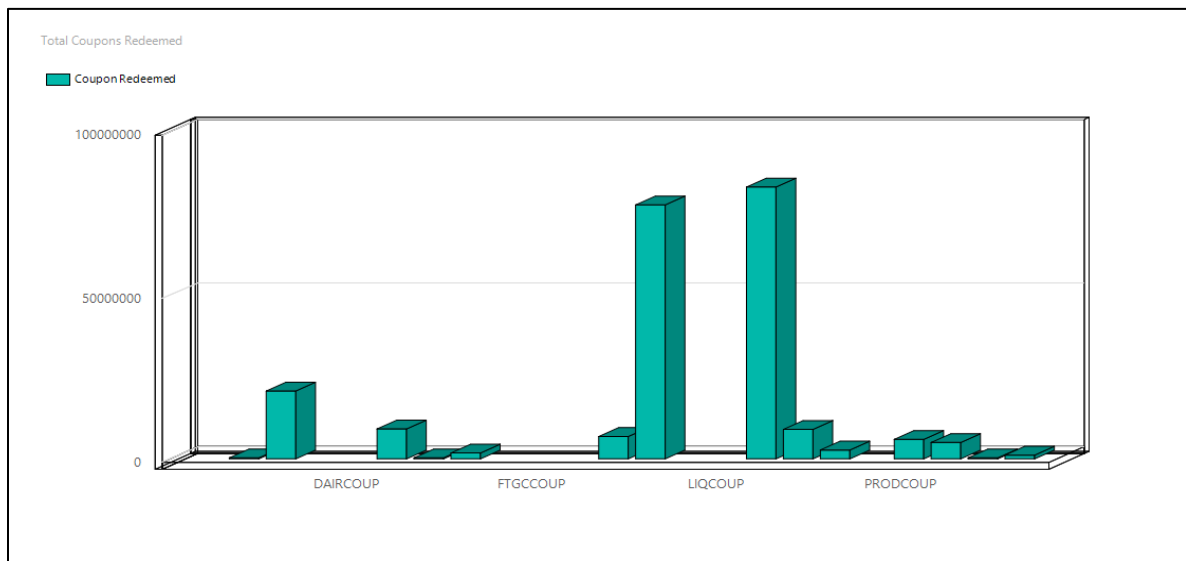*GROUP BY Dim_Promotional.coupon_cat*
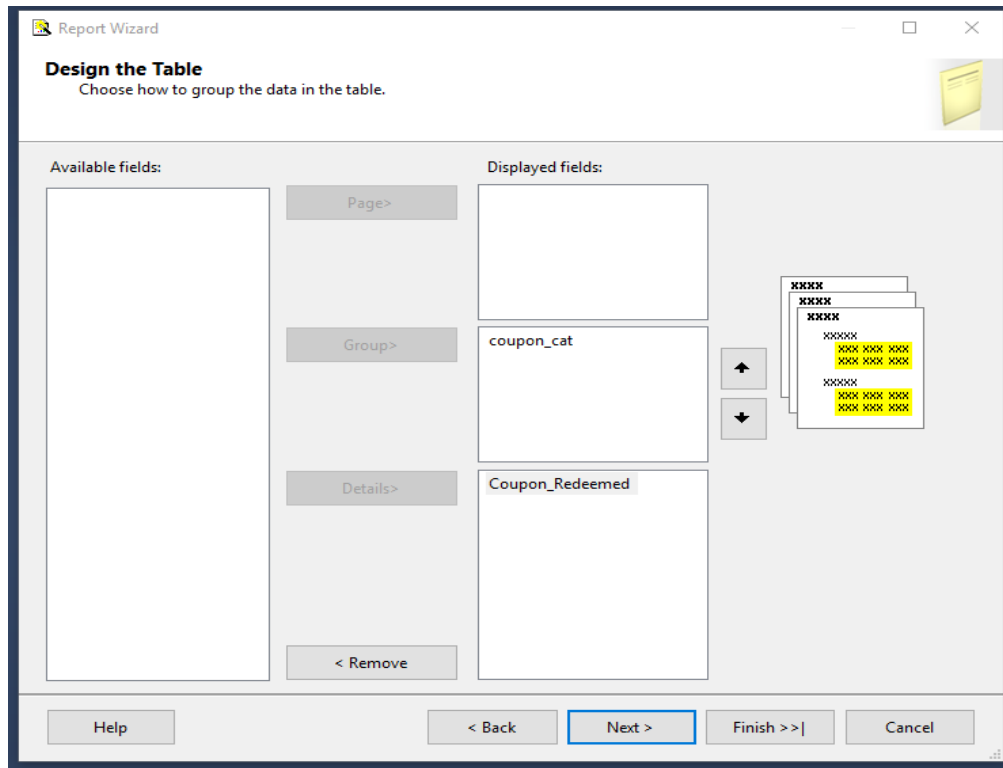
Report Structure and Attributes
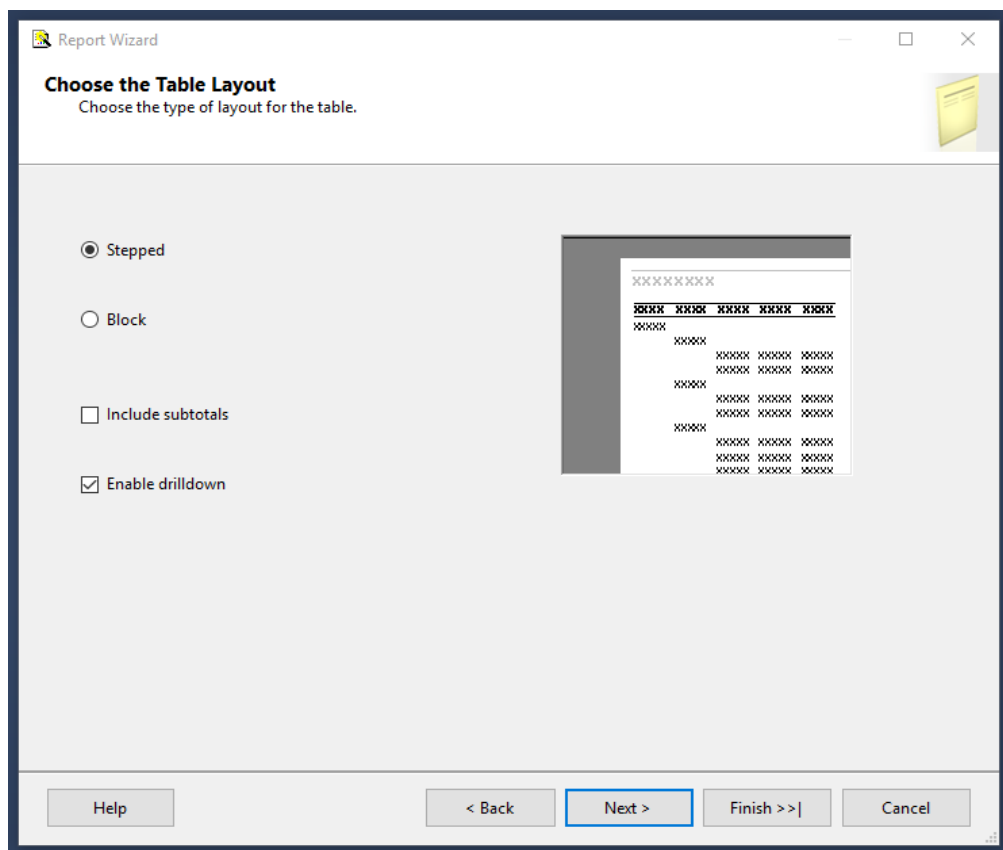
Selecting the Report Settings

coupon cat	Coupon Redeemed
⊟ BAKCOUP	
	0
⊟ BULKCOUP	
	27226
⊟ CONVFOOD	
	20759822
⊟ COSMCOUP	
	0
⊟ DAIRCOUP	
	0
⊟ DELICOUP	
	9166286
⊟ FISHCOUP	
	57094
⊟ FLORCOUP	
	1796366
⊟ FROZCOUP	
	0

Final Report deployed on web server in Tabular Format

Final Report deployed on web server in Bar Chart

Results from the Report:

We can see from the report the number of coupons redeemed for the entire duration. This can be used for strategizing promotional activities that involves specific coupons. We can also drill down to see which coupons were redeemed during a particular holiday season , which can help us promote certain coupons to achieve the maximum sales. This also allows to plan targeted promotion for a particular coupon category. We can also further drill down to see which coupons are redeemed at which store. This will give a picture to strategize sale in a particular store.

References:

1. Jami, Ata, and Himanshu Mishra. "Downsizing and Supersizing: How Changes in Product Attributes Influence Consumer Preferences." Journal of Behavioral Decision Making (2013).

2. Nevo, Aviv and Hatzitaskos, Konstantinos , "Why Does the Average Price of Tuna Fall during Lent?" (2005)

3. Nevo, Aviv and Catherine Wolfram , "Why do Manufacturers Issue Coupons? An Empirical Analysis of Breakfast Cereals", The RAND Journal of Economics (2002)

4. http://research.chicagobooth.edu/marketing/databases/dominicks/papers.aspx