ISTM 637 - Data Warehousing Project

Group 5

Design And Implementation of a Data Warehouse for a Retail Store with Store Level Data

Section: 601

Names:

Shalin Gund
Abhishek Subramaniam
Rishabh Semwal

Table of Contents

Section 1: Introduction 3
Problems Addressed in the Group Project 3
Understanding the Data 4
Data Description 4
Metadata for OLTP Source Files 6
Entity-Relationship Diagram 8
Section 2: Subject Area Understanding 9
Business Questions 13
Section 3: Overview of Kimball's Methodology 26
Importance of Kimball's Methodology: 27
Section 4: DW Logical Design (STAR Schema Design) 27
Mapping Tables 34
Physical Design 37
Section 5: Data Cleaning and Integration 37
ETL Development Plan 37
Data Extraction 38
Data Transformation 39
Data Loading 39
Data Sources: 39
Target Data: 40
Data Mapping 40
Data Extraction Rules 45
Data Transformation Rules 45
Data Aggregation Plan 46
Organization of Data Staging Area 47
Procedure for Data Extraction and Loading 47
ETL for Dimension table 53
ETL for Fact Table 54
Section 6: Implementation of ETL: 55
1. Week Decode Dimension Table Workflow: 55
2. Week Decode Table Data Transformation: 59
3. Week Decode Table Data Loading: 60
4. Week Decode Table Dimension Creation: 61
5. Week Decode Table Dimension Loading: 62
6. Store Dimension Table Workflow: 64
7. Store Dimension Table Creation and Loading 70
8. Store Fact Table Creation and Loading: 71
9. Movement Dimension and Fact Table Workflow: 76

10. Product Staging and Dimension Table Workflow: 87
List of Temporary Tables: 92
References: 93

Section 1: Introduction

Dominick's Finer Foods was a famous grocery store chain that was present in the Chicago metropolitan area from the year 1918 to 2013, in the late 1980s and early 1990s, DFF decided to partner up with the Chicago Booth School to conduct store-level research on shelf management and pricing. Many experiments were conducted around more than 25 different categories throughout a chain of 100 stores, from the research, we have approximately nine years of store-level data on the sales which has more than 3,500 UPCs available in the database, and this dataset consists of two different categories of files, which are general files and category specific files.

Key Points About the DFF Data:

- **Time Frame**: the partnership between the Booth school and Dominick's Finer Foods ran from 1989 to 1994
- **Scale**: Covered a chain of 100 DFF's stores
- **Scope**: The experiment was performed among 25 different product categories
- **Products covered**: Information for over 3,500 UPCs (Universal Product Codes)
- **Unique Aspects**: Important retail margin information and Breadth of coverage among various categories

Problems Addressed in the Group Project

The group will be mainly focused on analyzing the rich dataset to address the issues related to retail management

1. **Pricing Strategies**: Understanding the impact of various pricing strategies on sales volume and profitability
2. **Inventory Management**: How can we optimize the stock levels and how can we improve the shelf space allocation
3. **Competitive Analysis**: We will try to understand how different brands perform among product categories
4. **Promotional Effectiveness**: Understanding how different promotional activities can influence consumer behavior across different product categories

5. **Margin Management**: Will analyze the relationship between profitability, sales volume, and pricing of the products

By using this historical data, the project can help us gain valuable insights about retail management strategies and consumer behavior which is important to modern retailers and marketers. The data allows us to give an in-depth analysis of promotional effectiveness, pricing strategies, inventory management, competitive dynamics, and product category management which ultimately leads to improved customer satisfaction and performance. The unique depth and breadth of the DFF dataset, which covers multiple product categories, offers a great opportunity to study the retail measures and dynamics in a very comprehensive manner.

Understanding the Data

The objective of our project is to develop a data warehouse for Dominick's Fine Food (DFF) to help answer some complex business questions. The data acquisition for this warehouse will be conducted on the data that has been collected by Chicago Booth as part of their research partnership. This section helps the user understand the various elements of datasets involved in the project, their relationship with one another, and their overall mapping. To understand the data better, we will go ahead and categorize the various pieces of datasets (CSV files) into modules and give a brief description of the information contained within.

Data Description

The dataset collected is one of a kind. It has store-level data for all the DFF outlets in Chicago for 10 years, covering over 3500 unique UPCs. The downloaded data is separated into mainly two types of files.

Category Name	File Name	Description
General Files	Customer Count Files	These files contain information about the daily incoming traffic at the store level. They also include relevant information about the revenue generated and the coupons redeemed on a department level within each of these DFF stores.
	Store Specific Demographics	The file contains store-specific demographic data. This can be considered as Standardized data as it has

		been originally provided by the U.S. government census and further worked upon to create custom profiles for unique DFF stores.
Category Specific Files	UPC Files	These files store data for UPCs sold by the DFF outlets. Each unique UPC file is named in the following naming conventions UPCXXX where XXX stands for the specific category. The file provides us with information regarding identifier codes, product name, size etc. regarding a particular UPC
	Movement Files	The movement files are divided at category level like the UPC files and include weekly sales data for each UPC in each store for over 7 years. The attributes included in this file are price, unit sold, profit margin, deal code, etc. The files are sorted by UPC, store, week
Additional Files	Dominick Store Locations	This file contains the store location information of all the DFF outlets.
	Map	Used in conjunction with the store location file and browsers. This data helps the user locate the information for a specific week number.
	Week's Decoder Table	This is calendar data which contains start dates and end dates of weeks instead of making the user perform various times series aggregations.

Metadata for OLTP Source Files

The manual and the website provide metadata for the key files that are listed below:

1. **Customer Count Files:** The customer count data tells us about the number of customers visiting the store and buying something. It also has information about total coupons redeemed and total dollar sales, these numbers are compiled daily from scanners and registers

   The data consists of variables given below:

   a. **File Type:** Numeric and Character
   b. **Length:** 6-8
   c. **Variables:** Different department items, sales and coupon data, date of observation, store code, week number, etc.

2. **Store-Specific Demographics**: The store-specific demographics data originally originated from U.S. government census data for the Chicago area. Market metrics created this data to build demographic profiles for each of the DFF stores, this data tells us about the description of all demographics files in the database.

   The data consists of variables given below:

   a. **Variables:** Age, ethnicity, education, income, household size, etc.
   b. **Description:** population under age 9, college graduates, blacks and Hispanics, etc

3. **UPC files:** UPC files contains a detailed description of each UPC in a category, the naming convention of files is as follows: upcxxx, where xxx stands as the three-letter acronym for the category

   UPC consists of the following features:

   a. **UPC:** The last five numbers of the UPC tell us about the product, the rest of the digits tell us about the manufacturer

   b. **Commodity Code**: This is the category of the DFF version, one file can contain multiple commodity codes

   c. **Item Code**: This is the way through which DFF tracks products across multiple upcs. If in case, there are similar items then one is a newer version of the other

d. **Description**: here you might find 5 special characters that are part of the description:

- ~: this used for discontinued items
- #: means the product is on sale
- <: used for trial size product
- $ and *: this does not mean anything

e. **Case**: total number of items in the cases coming from the manufacturers, not seen by the consumers

The data consists of variables given below:

- **Variables:** UPC, com_code, nitem, descrip, size, case.
- **Description:** UPC number, Dominick's commodity code, Dominick's item code, product name, product size, number of items in case
- **Type:** Numeric and character
- **length:** 8-20

4. **Movement Data by UPC:** These files contain the information related to sales at the store level for each UPC in a category, information is stored on week-to-week basis, naming convention of files is in the "wxxx" format where xxx stands for three - letter acronym for the category

Movement data consists of the features below:

- **UPC:** Key which is used to merge the UPC file

- **Profit**: indicates the gross margin that DFF makes during the sale of the UPC

- **Sales**: this variable is used to determine of a product was sold on a promotion on that week

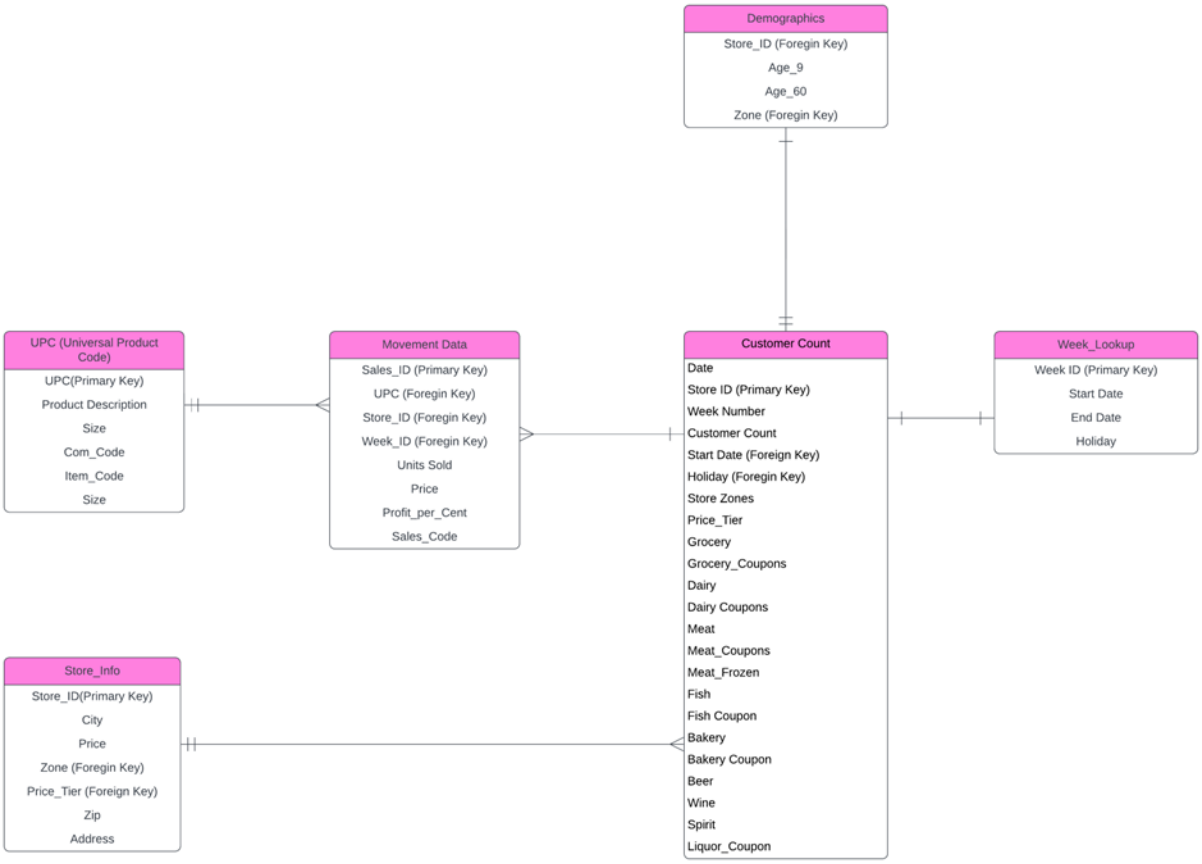
- **OK**: Flag set by the people to indicate that the data for that week is suspect

The data consists of variables given below:

- **Variable**: UPC, store, week, move, price, qty, sale, ok etc.
- **Description**: UPC number, store number, week number, retail price etc.
- **Type:** Numeric and characters

- **Length**: 3-8

Entity-Relationship Diagram



Section 2: Subject Area Understanding

[1] The research paper "The 'Shopping Basket': A Model for multicategory purchase incidence decisions" by Manchanda, Ansari, and Gupta (1999) introduces us to a novel approach to understanding the behavior of customer purchases items among various categories. The author showcases a multivariate probit that leads to interdependence in purchase decisions among a wide range of product categories

[1] The model is innovative as it takes care of both substitution and complementarity substitutions between various categories as well as specific household marketing mix activities, the study done by the author can be applied to DFF's business model in various ways:

1. **Multi-Category analysis:** The business model of DFF contains multiple categories, making the approach of this paper highly relevant, it gives us information on analyzing how a purchase in one category might influence the purchase among other categories

2. **Promotional Effectiveness:** By modeling cross-category effects, this research can help us understand how coupons and promotions in one category can affect the sales in the other categories, which is an important consideration in the retail market

3. **Multi-Category Analysis**: The dataset has multiple categories, making this a relevant approach. It gives a framework for analyzing how purchases in one category can affect the other categories

4. **Basket Analysis**: The understanding of "Shopping Basket" is important to retail operations and can give us insights into the store layout, inventory management, and cross-promotions

5. **Methodological Approach:** while the statistical techniques mentioned might be advanced, the approach of considering multiple categories together might be valuable for analyzing the business data model of DFF

6. **Managerial Implication**: The paper gives us information about category management and promotional planning, which is a very crucial thing for DFF's business model

In conclusion, the paper provides us with a framework that is required for analyzing purchase decisions among multiple categories. It can guide the decision-making of analytical approaches that consider the relationship between various product categories, which leads to more effective marketing strategies and analytical decision-making for a retail chain like Dominick's Finer Foods.

[2] A research paper by Ailawadi and Nesli (1998) titled "The Effect of Promotion on Consumption: Buying More and Consuming It Faster" gives us important insights into the complex relationship between consumer behavior and promotional activities. This study is very much inclined with the Dominick's Finer Foods (DFF) business model, as it showcases several key issues that play a crucial role in marketing strategy and retail management.

[2] The paper addresses how household consumption rates and purchasing behavior are affected by price promotions, using the scanner data from two product categories: ketchup and yogurt. This approach aligns well with the provided dataset, which covers multiple product categories among

various stores. The study's methodologies and findings can be applied to DFF's retail strategy in many ways:

1. **Impact of promotions on consumption:** The research shows that promotions not only increase sales but also help accelerate consumption rates. This is a very important insight for DFF to learn and understand the potential impact of their promotional activities on consumer behavior and sales patterns

2. **Stockpiling behavior**: This study also addresses the main issue of how promotions lead to stockpiling, which goes in line with DFF's inventory management and future sales forecasting. Understanding consumer behavior can help the company optimize its promotional strategies and inventory levels

3. **Short-term vs. long-term promotional effects**: This research helps in differentiating between immediate sales increase and long-term consumption effects. This is important because it helps in developing a balanced promotional strategy that takes care of both long-term consumer behavior changes and short-term revenue boosts.

4. **Household heterogeneity**: This study differentiates the household characteristics as this will help the company to segment the customers and make changes to their promotional offers effectively. This approach will help DFF in creating more efficient and targeted marketing strategies.

5. **Category-specific effects:** By analyzing the two different product categories, yogurt and ketchup separately, the paper showcases how promotional effects can vary significantly across various product categories, this is beneficial for DFF as they have a diverse product range and will help them understand the need of category-specific promotional strategies

6. **Data Utilization:** The study showcases how the data collected from scanner data, which is similar to the data that DFF has collected, can be used to gain deeper insights into understanding the consumers and their behavior. This proves the potential value of DFF's dataset for strategic decision-making

7. **Inventory and consumption modeling:** The paper showcases some methods for estimating household inventory and consumption, which can be adapted by DFF to improve its business model by improving demand forecasting and inventory management across various product categories.

In conclusion, the Ailawadi and Neslin paper provides a good framework for understanding and analyzing the promotional effects which is highly important for the DFF's business model. By

using and applying similar methodologies and considering these insights, DFF can improve its understanding of how it can optimize its promotional strategies, understand consumer behavior, and ultimately improve its retail performance across multiple product categories.

[3] The research paper by Levy et al. (1997) titled "The Magnitude of Menu Costs: Direct Evidence from Large U.S. Supermarket Chains", this article is based on the field of price stickiness and the cost of a menu in a retail environment.

[3] Levy and some of his friends conducted a user study of five large supermarket chains which included Dominick's Finer Foods, to quantify the costs related to changing prices, which is also known as menu costs. The researchers had unknown access to detailed data on the process of changing prices which allowed them to provide the initial direct and comprehensive measurement of these retail markets. The main aim of this research was to understand why prices might not change quickly as the economic condition might suggest, this study has a direct impact on DFF's retail strategy in many ways:

1. **Pricing Strategies:** This paper shows the cost which is related to the changing prices and is a fundamental aspect of DFF's business model, learning and understanding about the costs will make sure that the company makes informed decisions.

2. **Operational Insights**: The study also provides a detailed breakdown of the material costs, and the labor involved in the price change, offering important insights to DFF's management.

3. **Technology Impact**: The paper focuses on how the latest technology such as "Electronic Shelf Labels" could affect the cost of the menu, which is in line with the DFF's future technology investment decisions

4. **Competitive Environment**: Learning and studying about various business chains including DFF, helps in understanding industry-wide practices and understanding, helping to understand the position of DFF in the market

5. **Profitability Analysis**: By changing the menu costs, the paper gives information about DFF to assess the profitability of price changes, with a focus on small price adjustments.

6. **Economic Implications:** The findings have a very good implication for understanding inflation and making price adjustments in the retail sector, which is one of the most crucial things for DFF's long-term business planning.

This paper is highly relevant for DFF's business model as it gives direct evidence from DFF's own operation, this gives an insight into the importance of understanding about costs and complexities of price change and helps in learning about the reason behind the price change.

Business Questions

Priority Scale:

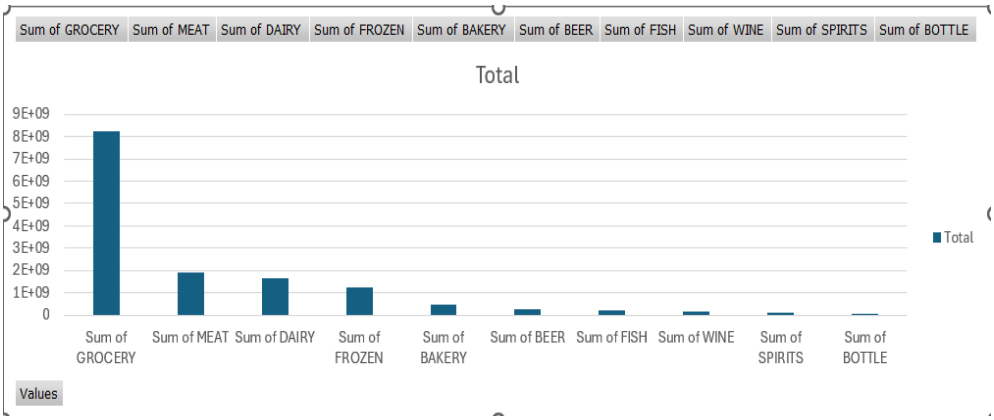
Priority	Description
Low Priority [1]	Minimal Impact on marketing strategies and sales
Somewhat Low Priority [2]	Limited influence in specific products
Medium Priority [3]	Provides good insights but affects smaller products
High Priority [4]	Useful for optimizing strategies and key product categories
Very High Priority [5]	Critical for core marketing and sales decisions

BQ1 : What are the all-time top-performing categories at DFF?	
Business Question Justification	The sales data at DFF is spread over various products and their subcategories. The business can analyze its income revenue to plan better. Knowing their top-selling products/categories helps them understand the influx of revenue better. They might even decide to constantly monitor the warehouses for these products since their demand is directly proportional to the sale. This information can also contribute towards demand forecasting, keeping DFF in business.

Pivot Table	

Values	
Sum of GROCERY	8256353813
Sum of MEAT	1901158042
Sum of DAIRY	1674633112
Sum of FROZEN	1267368525
Sum of BAKERY	470914346.5
Sum of BEER	267265053.3
Sum of FISH	251043187.4
Sum of WINE	159095615.3
Sum of SPIRITS	115163691.5
Sum of BOTTLE	8018235.96

Pivot Charts



Priority [5]

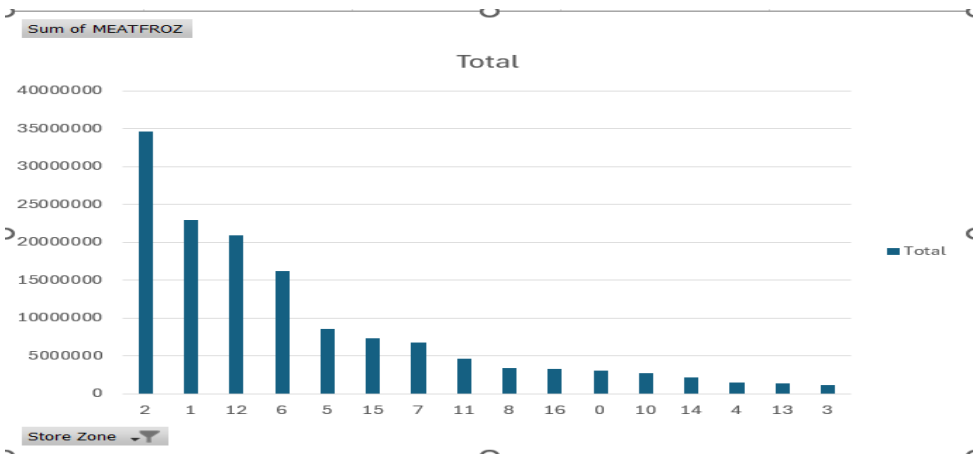
Justification: Understanding what the top-selling categories are is essential to maximize revenue and inventory management, knowing this will help us focus marketing efforts on the most profitable segments

BQ2 : Which were the Store Zones that contributed to maximum meat sales?	
Business Question Justification	One of DFF's most popular items is meat. The customer might prefer to buy meat in any format depending upon their convenience. The business is spread across various parts of the nation and handling perishable food items is a problem they might run into. Understanding the distribution of store zones with respect to meat consumption can help the business cater to a larger audience without hindrance. This will also help them to target their meat coupons and promotions at specific audiences, leading to better footfalls.

Pivot Table

Row Labels	Sum of MEATFROZ
2	34593844
1	22966460.15
12	20876058.26
6	16208157.96
5	8525366.98
15	7325902.49
7	6787092.18
11	4665585.16
8	3422879.08
16	3260289.69
0	3030481.43
10	2771006.61
14	2208477.78
4	1494381.02
13	1338486.68
3	1196322.18
Grand Total	**140670791.7**

Pivot Charts



Priority [4]

Justification: Meat is a popular product, and getting to know about the zones that perform well in meat sales, will help us in creating marketing strategies and managing the inventory for perishable goods

BQ3 : Which are the stores that cater to people with income below $15,000? What are the

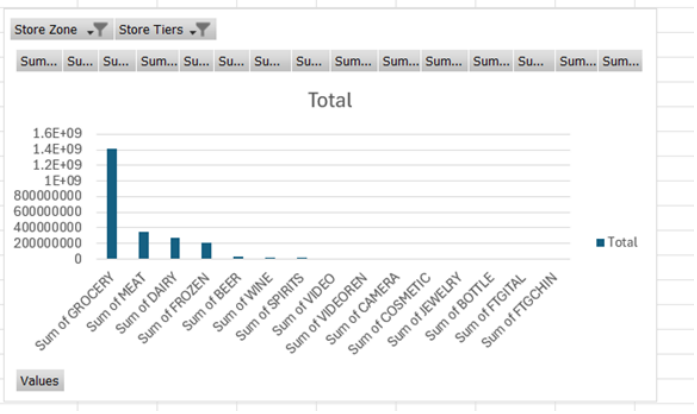
products that they opt for the most?	
Business Question Justification	The data within the files indicate that DFF caters to a large range of audience. This range falls in between varying kinds of income strata, living conditions, age, etc. DFF's marketing team can understand the needs of low-income households and strategically place coupons according to their needs. This will help the business understand the needs of society and address accessibility problems for those who are underprivileged.
Pivot Table	

Row Labels	Sum of POVERTY
2	**1.260890015**
4	0.037592397
5	0.025104603
9	0.028060483
28	0.035497431
44	0.030154849
45	0.022827393
47	0.032623559
48	0.018204313
49	0.028320971
50	0.042153494
54	0.021336421
56	0.027950834
64	0.038713642
65	0.02848541
74	0.06509299
76	0.133099825
81	0.031450382
84	0.024297424
86	0.15266297

Pivot Charts

Store Zone	2
Store Tiers	(Multiple Items)

Values	
Sum of GROCERY	1407445782
Sum of MEAT	343351681.1
Sum of DAIRY	277513209.3
Sum of FROZEN	212261686.5
Sum of BEER	35941002.12
Sum of WINE	21539522.91
Sum of SPIRITS	14706287.72
Sum of VIDEO	5749522.71
Sum of VIDEOREN	3234563.23
Sum of CAMERA	1667824.78
Sum of COSMETIC	1416319.17
Sum of JEWELRY	302863.17
Sum of BOTTLE	105460.1
Sum of FTGITAL	21112.19
Sum of FTGCHIN	8893.01



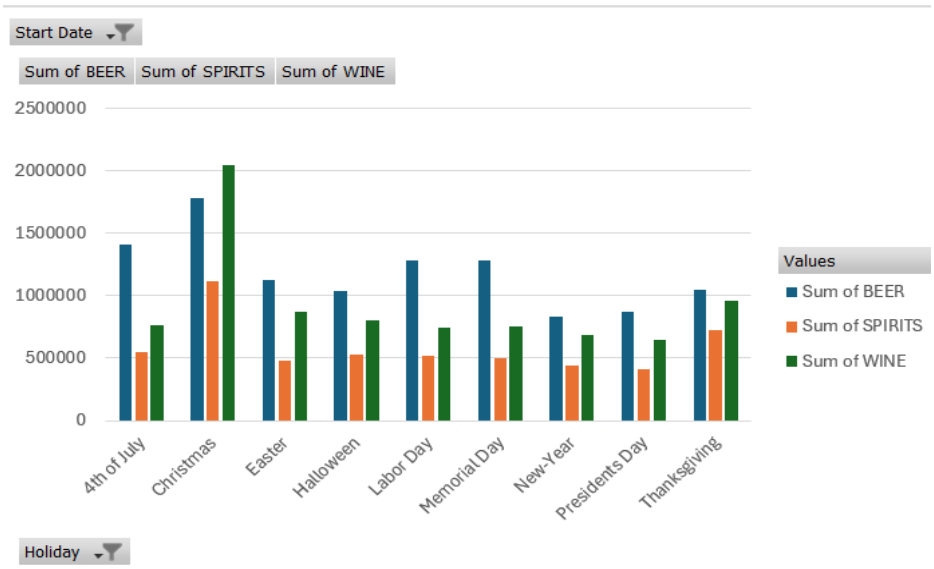
Priority [4]	**Justification:** Understanding low-income customers is important to gain a broad customer base. Targeted marketing for this segment can improve customer satisfaction and increase the number of sales for this segment of

	people.

BQ4 : How did the sale of alcoholic products fare during the holiday season for the period 1995-1996?	
Business Question Justification	The demand for alcoholic beverages surges during the holiday season. With festivities and parties in order, the consumers yearn for these products. Understanding the sales of these products during the holiday season from previous years will help the business understand the demand surge during these periods and make sure to stock up their freezers. Additionally, this might also be good from a promotion and coupon point of view. The reason is that the business can plan for its product placement and offers to maximize its footfalls during the holidays.
Pivot Table	

Start Date (Multiple Items)

Row Labels	Sum of BEER	Sum of SPIRITS	Sum of WINE
4th of July	1409994.72	549338.85	761805.17
Christmas	1785214.75	1113457.77	2043491.44
Easter	1123977.61	483080.91	874144.91
Halloween	1036278.26	525438.86	803740.62
Labor Day	1285847.52	523381.19	746589.17
Memorial Day	1282410.32	497625.57	751221.48
New-Year	828310.84	437892.92	689172.24
Presidents Day	872465	408129.15	650716.35
Thanksgiving	1052180.83	720600.76	963016.99
Grand Total	**10676679.85**	**5258945.98**	**8283898.37**

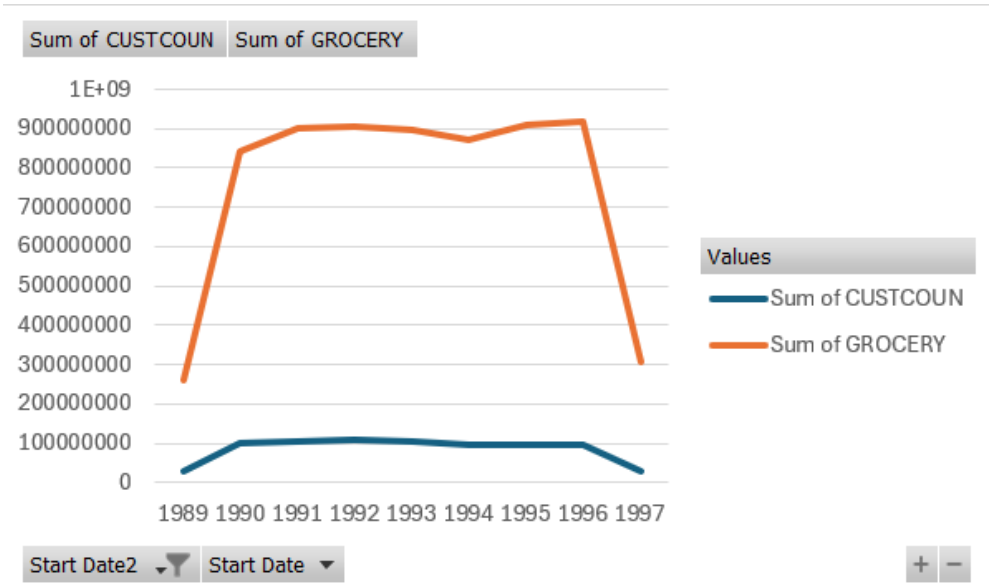
Pivot Charts	
Priority [3]	**Justification:** This might be important for seasonal marketing strategies, but the focus is on a limited period and it may limit its current relevance, however, it gives us insights on consumer behavior during festive seasons and high-demand times.

BQ5 : How has the consumer influx changed year on year? How has it affected the sales of Groceries?	
Business Question Justification	The business is home to more than a million consumers. The business would benefit from understanding the year-on-year change in the footfall for the business. The total customer count per year helps the business judge if their strategies are working as expected. The footfall is directly proportional to the revenue generated over various categories. By analyzing further into the groceries category (Highest Selling), the business will be more ready to understand the consumer demand to manage the perishable category of groceries.

Pivot Table

Row Labels	Sum of CUSTCOUN	Sum of GROCERY
1989	29072259.35	260115770.7
1990	99505559	842553804.9
1991	104100380	899213843.4
1992	107869416	907289874.5
1993	104831049	896891064.6
1994	97378230	870180585.1
1995	95376130	908381311.1
1996	96884374	918196285
1997	30286638	304754526.8
Grand Total	765304035.4	6807577066

Pivot Charts



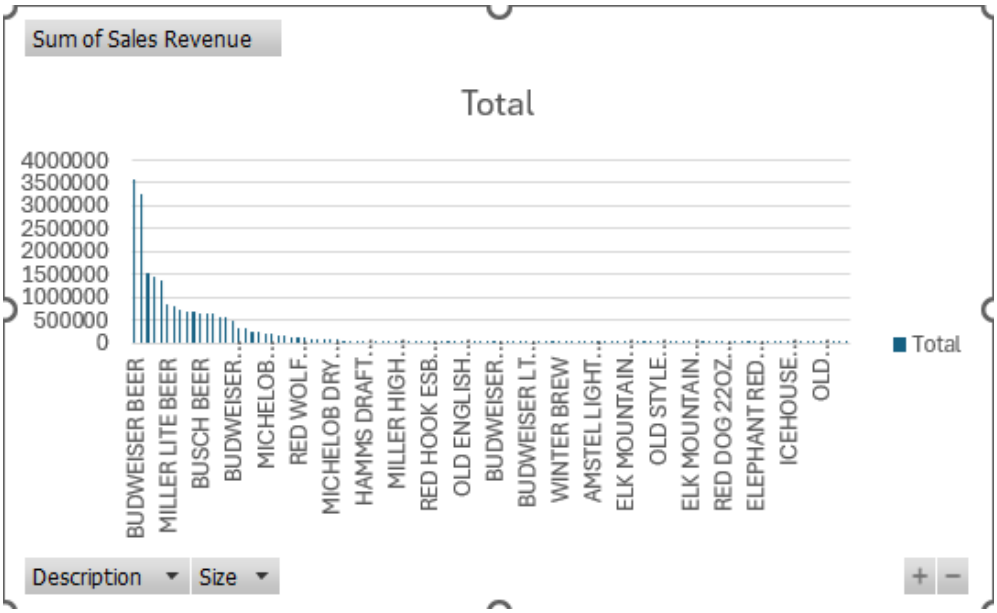
Priority [5]

Justification: Customer influx is something that directly has an impact on overall sales. Groceries are a staple category and understanding how this category is impacted by customer trends is crucial for business strategy

BQ6 : How has the sales revenue of Beer navigated with respect to various brands?	
Business Question Justification	Beer is a beverage that sells the most at DFF outlets. Understanding which beer sells the most will give the business a deeper understanding of consumer demands. On further investigation, knowing the most sold bottle will also help them in shelving and stocking the necessary brands.
Pivot Table	

Row Labels	Sum of Sales Revenue
BUDWEISER BEER	3561101.75
24/12O	2136324.92
12/12O	814713.75
6/12 O	573620.23
6/16 O	19372.77
8/7OZ	17070.08
MILLER GENUINE DRFT	3255889.83
MICHELOB GOLDEN DRAF	1531424.36
BUDWEISER LIGHT BEER	1452972.2
MILLER GENUINE DRAFT	1376691.98
MILLER LITE BEER	838653.93
MILLER HIGH LIFE BEE	819574.11
MICHELOB REGULAR BEE	715511.04
O'DOUL'S N/A LONGNEC	700105.3
MILLER GEN DRFT LT L	688278.36
BUSCH BEER	670091.34
MILLER SHARP'S N/A L	657235.31
MILLER HIGH LIFE PAR	655865.57
MILLER HIGH LIFE N.R	580697.36
BUDWEISER BEER N.R.B	553927.28
BUDWEISER BEER LONG	510538.28
BUSCH LIGHT BEER	341240.64
MILLER HIGH LIFE LNN	325314.62

Pivot Charts



Priority [3]	**Justification**: While beer is a good and high-selling product, groceries mostly have a broader demand. This question is important to understand beverage sales but may be secondary compared to other categories

BQ7 : Which are the products/categories that have been facing a year-on-year declining sales trend?

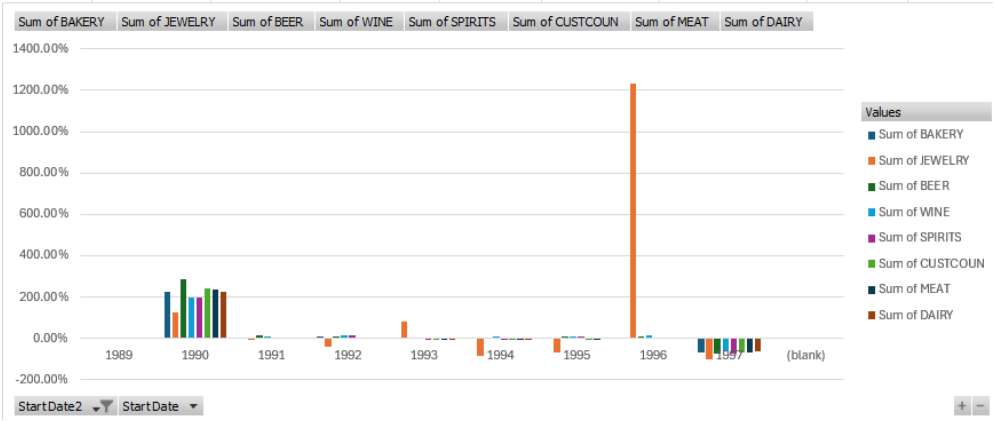
Business Question Justification

Analyzing data should help a business be in the game. Analyzing those products that have a declining trend of sales, can help make DFF offboard their unpopular products. From a marketing standpoint, if certain high-value products are facing some decline in sales, innovative promotional strategies can help them palace the products in a better manner.

Pivot Table

Row Labels	Sum of BAKERY	Sum of JEWELRY	Sum of BEER	Sum of WINE	Sum of SPIRITS	Sum of CUSTCOUN	Sum of MEAT	Sum of DAIRY
1989								
1990	225.07%	127.01%	288.34%	199.03%	195.94%	242.27%	234.52%	222.69%
1991	3.92%	-4.22%	15.74%	7.96%	3.57%	4.62%	3.87%	1.98%
1992	9.10%	-42.38%	7.28%	16.13%	15.67%	3.62%	0.21%	5.55%
1993	1.51%	78.39%	1.60%	1.05%	-1.16%	-2.82%	-3.46%	-4.81%
1994	0.54%	-83.55%	5.69%	7.10%	-1.06%	-7.11%	-3.33%	-1.34%
1995	0.49%	-70.34%	8.12%	9.24%	6.46%	-2.06%	-3.49%	2.04%
1996	3.75%	1235.29%	6.70%	11.99%	3.29%	1.58%	3.12%	4.63%
1997	-68.72%	-100.00%	-74.24%	-63.98%	-72.31%	-68.74%	-66.85%	-65.17%
(blank)	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!
Grand Total								

Pivot Charts



Priority [5]

Justification: Understanding important declining trends is vital for decision making, this can help DFF make the important changes, adjust the marketing efforts to revitalize the sales or discontinue the products

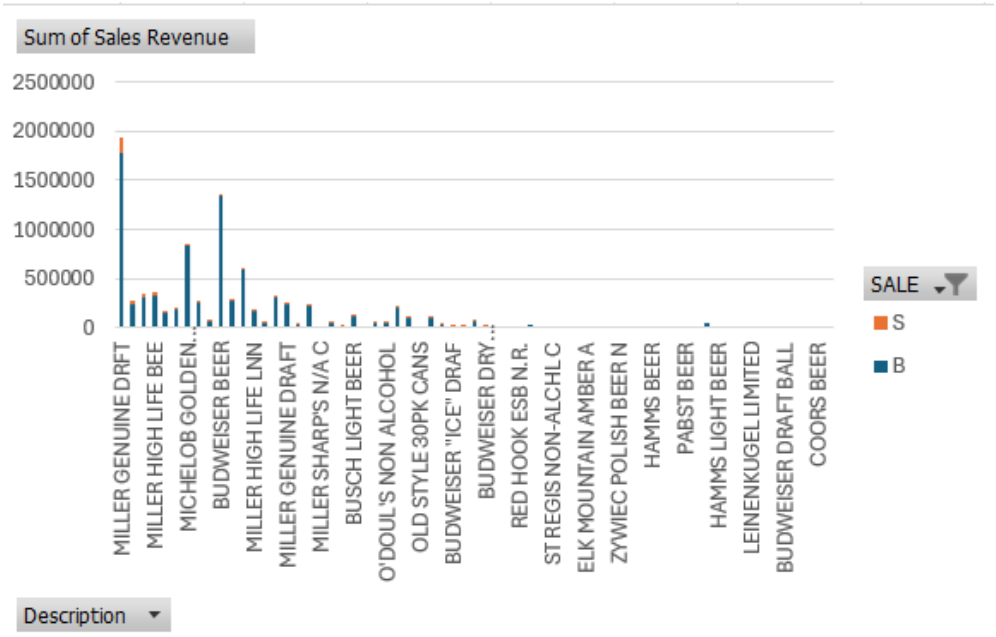
BQ8 : Which beer brand attracts the highest coupon redemption and buy back offers?

Business Question Justification	Studying the sales of beer means studying one of the most popular products sold at DFF. Although the sale could be perceived in any manner, thinking in terms of promotional codes and buy-back offers puts the promotional strategy to the test. Analyzing those brands and pushing for heavy collaborations can boost the organizations' sales and income revenue.

Pivot Table

Sum of Sales Revenue	Column Labels	
Row Labels	B	S
MILLER GENUINE DRFT	1785472.53	157961.9
MILLER HIGH LIFE N.R	241886.18	28689.73
MILLER HIGH LIFE PAR	313433.71	23662
MILLER HIGH LIFE BEE	332422.19	22011.19
BUDWEISER BEER N.R.B	147315.55	16846.76
MILLER SHARP'S N/A L	196714.23	9726.95
MICHELOB GOLDEN DRAF	832061.91	8012.43
MICHELOB REGULAR BEE	264507.48	7975.39
BUDWEISER LIGHT NR	74190.23	7666.13
BUDWEISER BEER	1335960.08	4845.17
O'DOUL'S N/A LONGNEC	280224.06	4105.83
BUDWEISER LIGHT BEER	590332.71	2762.2
MILLER HIGH LIFE LNN	166329.41	1827.31
BUDWEISER ICE DRAFT	44946.24	1445.89
MILLER GEN DRFT LT L	314736.05	1413.4
MILLER GENUINE DRAFT	238995.31	1263.57
MICHELOB DRY LONGNEC	29328.4	845.5
BUSCH BEER	214872.13	724.99
MILLER SHARP'S N/A C	2330.18	648.83

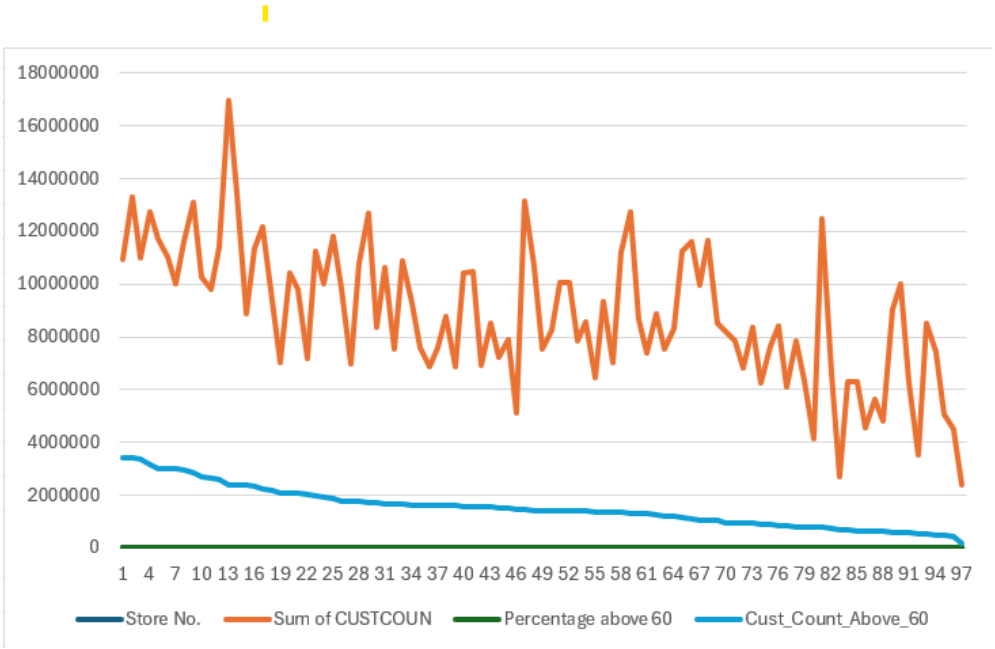
Pivot Charts



Priority [3]	**Justification**: Understanding coupon effectiveness is important but it focuses on specific product types, making it a very narrow business concern

BQ9 : Which are the DFF outlets that cater to senior citizens? What are the top coupons redeemed?	
Business Question Justification	Senior citizens contribute to a very large and loyal customer base at DFF. Understanding their needs and roadblocks to get their hands on the product, will help the team place these categories in a more streamlined manner. Thereby increasing the customer base and leading to a potential stream of revenue generation.
Pivot Table	

Store No.	Sum of CUSTCOU	Percentage above 6	Cust_Count_Above_6
306	10913986	0.312927976	3415291.551
32	13305564.15	0.254953032	3392293.917
74	10961628.25	0.307397856	3369581.027
98	12722785.02	0.24920053	3170524.775
73	11698947.8	0.257450782	3011903.26
18	11037935.17	0.272313368	3005777.306
113	10002495	0.299352545	2994272.339
8	11629092.4	0.252394035	2935113.548
102	13100154.7	0.216626232	2837837.154
107	10256046.51	0.261867453	2685724.779
71	9784525.83	0.26807087	2622946.351
101	11423675.26	0.225035218	2570729.257
303	16976940.31	0.142000922	2410741.178
12	13437868.68	0.178341405	2396528.381
9	8857354.05	0.269119018	2383682.42
75	11326894.55	0.207699493	2352590.254
114	12162686.82	0.182173296	2215716.74
95	9380363.47	0.230717504	2164214.045
53	6999696	0.300278681	2101859.482
83	10420296.12	0.200834686	2092756.897
111	9804868.94	0.210512842	2064050.83
72	7188290.49	0.283727688	2039517.04
123	11232774.02	0.176040942	1977428.121

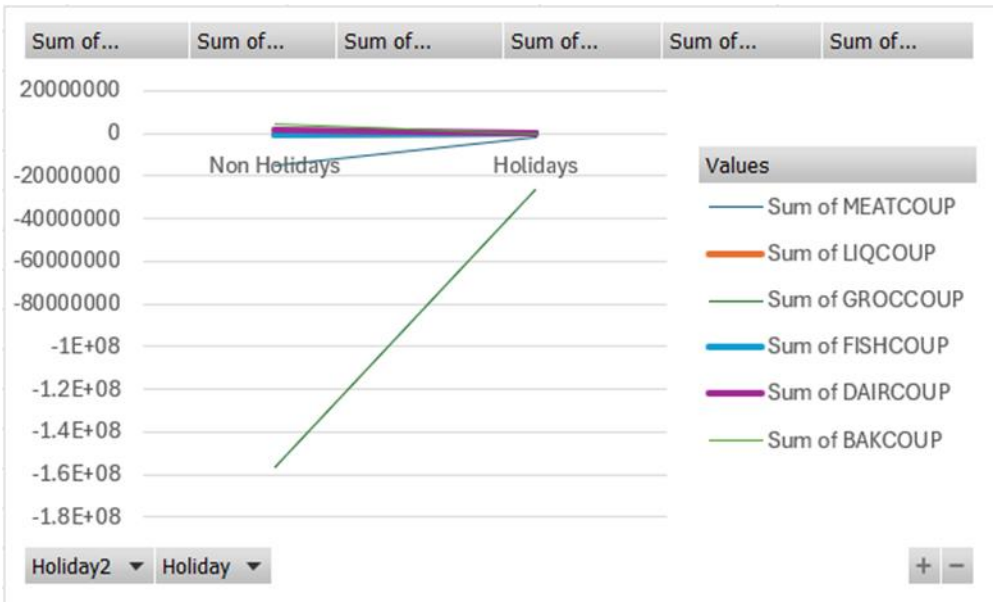
Pivot Charts	
Priority [4]	**Justification:** Senior citizens are important customers and understanding what they want, and coupon usage allows DFF to create targeted promotions, especially in demographic-specific outlets

BQ10 : How have the coupon sales been, especially during the holiday season?	
Business Question Justification	Coupons are the marquee promotional tool at DFF's disposal. By understanding the popular coupons, the marketing team at DFF can learn about consumer behavior. By understanding the periods when the coupon redemption goes up, the business can decide on the promotional strategy and push their more popular coupons and offers. Attracting customers via lucrative offers and promotional coupons will surely contribute towards DFF income revenue.

Pivot Table

Row Labels	Sum of MEATCOUP	Sum of LIQCOUP	Sum of GROCCOUP	Sum of FISHCOUP	Sum of DAIRCOUP	Sum of BAKCOUP
Non Holidays	**-15209467.81**	**-56942.61**	**-156244726.2**	**-671356.47**	**2081655.28**	**4690247.09**
Holidays	**-1752597.42**	**0**	**-26210086.47**	**-213711.52**	**0**	**0**
4th of July	-34357.16	0	-2223099.47	-375.55	0	0
Christmas	-102490.09	0	-2992973.43	-65810.53	0	0
Easter	-333203.56	0	-3472121.34	-95037.24	0	0
Halloween	-164052.3	0	-2066792.3	-1298.56	0	0
Labor Day	-32430.48	0	-2583254.64	-1714.58	0	0
Memorial Day	-201585.9	0	-2808833.58	-588.42	0	0
New-Year	-166512.08	0	-2533843.69	-13645.85	0	0
Presidents Day	-278123.76	0	-3693193.92	-21373.91	0	0
Thanksgiving	-439842.09	0	-3835974.1	-13866.88	0	0
Grand Total	**-16962065.23**	**-56942.61**	**-182454812.6**	**-885067.99**	**2081655.28**	**4690247.09**

Pivot Charts



Priority [4]

Justification: The holiday season mostly sees significant sales spikes, and understanding coupon redemption trends can help in optimizing promotional strategies during this crucial time, Targeting the right offers with well-timed promotions can significantly boost sales and the revenue

Section 3: Overview of Kimball's Methodology

Kimball's methodology is a comprehensive approach to designing, developing, and implementing data warehouse and business intelligence systems. It is based on a set of principles that define the structure and organization of data and focus on data analysis. Its approach highlights dimensional modeling and provides business value through iterative development.

It revolves around creating a flexible, user-friendly, and performance-oriented data warehouse. Kimball's model emphasizes using a bottom-up approach, i.e., designing a data warehouse by

creating data marts according to business problems that would eventually evolve into a cohesive data warehouse. This methodology is intended to deliver swift incremental business results within an enterprise data warehouse.

Kimball's methodology comprises steps that include:

1. Defining Business Requirements [4]: In this step, the team focuses on defining the objectives, scope, and overall planning to establish a concrete foundation for the entire project. The primary steps of this task are to carry out requirement gathering and understanding the business needs. Interviews and discussions are conducted with stakeholders of the project to identify reports and metrics required to support business decisions. This step is important as it ensures that the final product aligns with business goals.
2. Identifying the Business Process and Grain Definition: In this phase, different business processes are defined, and they will be based on building data marts or data warehouses. This step will also determine the level of granularity at which the data will be stored.
3. Technical Architecture Design: The team would design a suitable Data Warehouse structure based on the business requirements, IT environment, and data volume. This step focuses on selecting technologies that support the data warehouse.
4. Identification of Dimensions and Facts (Dimensional Modeling): In this step, the dimensions are defined, and measurable data (fact data) is identified accordingly. A set of dimension tables and fact tables is created to categorize and support businesses' analytical needs. A dimensional model is generated using either STAR Schema or SNOWFLAKE Schema but in Kimball's methodology, a STAR Schema is preferred where the dimension tables connect centralized facts.
5. Design and Development: Here, ETL (Extract, Transform, and Load) steps are carried out to handle common design problems in the analytical environment. According to the business question, reports and dashboards are also developed for the users to interact with the warehouse.
6. Deploy and Iterate: After deploying the system, feedback is collected from users to refine the data marts over time. This step is required to continuously improve the data warehouse as per new business requirements.

Importance of Kimball's Methodology:

Kimball's methodology is used to emphasize the importance of building integrated data marts that conform to the overall data warehouse architecture.

1. Consistency and Integration: This approach ensures that there is consistency in dimensions and facts across all the data marts, enabling smooth integration to build a data warehouse using different data marts.

2. Scalability: New data marts can be easily incorporated into the current framework by following a standardized approach and without hindering the overall architecture.
3. Performance Optimization: Through dimensional modeling, query performance can be optimized to handle analytical workloads [6].
4. Flexibility: Adding new dimensions and facts is easier in the dimensional modeling approach without disrupting the existing structure [6].
5. Cost Reduction: Even though the initial setup might require investment, the ease of scalability and flexibility can lead to significant long-term savings [6].

Section 4: DW Logical Design (STAR Schema Design)

After thoroughly studying the raw data collected by Chicago Booth and understanding the business context of the information retrieved, we recommend the usage of the Kimball methodology for the development of DFF's data warehouse. The Kimball approach is more flexible and simpler compared to the Inmon model, with the latter requiring the development of an enterprise-wide architecture. Kimball's methodology makes use of dimensional modeling to develop a data warehouse. The two main components of a multi-dimensional model (Kimball Warehouse) are dimensions and facts. Facts refer to the numerical value or measures that reside within the data and dimensions refer to the contextual information surrounding the facts. Ultimately, we aim to create a collection of robust independent data marts that help in answering business questions with contextual answers.

As mentioned in [4], the four main steps of developing a data warehouse using the Kimball methodology are:

1. Identification of the Business Problem
2. Definition of Level of Grain
3. Identification of Dimensions
4. Identification of Facts

Identification of the Business Problem

Under the requirements analysis phase, the team developed a set of business questions after performing due diligence. The business questions were drafted from the retail sales and marketing perspective. Depending upon priority and business significance the following questions were selected. The fully developed data warehouse should be able to answer the following questions:

ID	Business Questions
BQ3	Which are the stores that cater to people with income below $15,000? What are the products that they opt for the most?
BQ4	How did the sale of alcoholic products fare during the holiday season for the period 1995-1996?
BQ5	How has the consumer influx changed year on year? How has it affected the sales of Groceries?

BQ7	Which are the products/categories that have been facing a year-on-year declining sales trend?
BQ8	Which beer brand attracts the highest coupon redemption and buyback offers?

Identification of Level of Grain

After studying the data, we could understand that there was mainly sales data in two different levels of detail. Firstly, we had the 'ccount.csv' file which contained the store-level sales data for each day of the week. Next, we had the set of movement files which indicated the weekly sales data for a specific product. Therefore, to maintain uniformity and avoid any level of grain issues, we have decided to pursue the data in the weekly aggregated format across all the tables in the data warehouse.

Identification of Dimensions

The following are the dimension tables that will constitute an important component of the star schema to be developed.

Time_Dim

This dimension table mainly consists of the dates corresponding to a particular week number. Columns like Holiday_Flag and Holiday_Type help us with easy identification for time series analysis.

Attribute	Description
TimeDim_Key	Surrogate Key / Primary Key
Week	Week Number
Start_Date	Start Date of the Week
End_Date	End Date of the Week
Year	'YYYY' format of the year of the start_date
Holiday_Type	Type of Holiday weekend
Holiday_Flag	Indicator column for Holidays, by default set to 'Y'.

Store_Dim

The Store Dimension has basic information corresponding to the store number, region, week and poverty percentage. This gives context to the sales information pertaining to these outlets.

Attribute	Description
Store	Store Number
Week	Week Number
StoreDim_Key	Surrogate Key / Primary Key

Product_Dim

The product dimension helps in tracking the characteristics of individual UPCs.

Attribute	Description
Product_ID	Universal Product Code
Product_Description	Product Name
ProductDim_Key	Surrogate Key / Primary Key

Movement_Dim

Attribute	Description
MovementDim_Key	Surrogate Key / Primary Key
UPC	Universal Product Code
Week	Week Number
Sales_Code	Promotion Code

Identification of Facts

The following are the fact tables that will constitute the central component of the star schema to be developed.

Store_fact

The store fact table contains all the information related to the weekly customer traffic and weekly sales of various products like beer, wine, meat, frozen meat, fish, etc. at a store level. The surrogate keys act as foreign keys facilitating in bringing in contextual information from the dimension tables.

Attribute	Description
Customer_Count	Weekly Customer Traffic
Store	Store Number
Week	Week Number
Beer_Sales	Weekly Beer Sales in Dollars
Wine_Sales	Weekly Wine Sales in Dollars
Spirit_Sales	Weekly Spirit Sales in Dollars
Grocery_Sales	Weekly Grocery Sales in Dollars
Bakery_Sales	Weekly Bakery Sales in Dollars
Meat_Sales	Weekly Meat Sales in Dollars
Frozen_Meat_Sales	Weekly Frozen Meat Sales in Dollars
Fish_Sales	Weekly Fish Sales in Dollars
Poverty_Percentage	Average Poverty Percentage in that corresponding store
Poverty_Demographic	Customer Count of Consumers below the poverty line
TimeDim_Key	Surrogate Key / Foreign Key
StoreDim_Key	Surrogate Key / Foreign Key

Movement_Fact

The sales fact table contains all the information related to the weekly sales of a specific UPC. The surrogate keys act as foreign keys facilitating in bringing in contextual information from the dimension tables.

Attribute	Description
Move	Number of units sold
Quantity	Number of items bundled together
Price	Unit price of the Product
Total_Weekly_Sale	Total Weekly Sales = (Price * Move / Quantity)
Profit	Profit percentage on the sale of one unit
ProductDim_Key	Surrogate Key / Primary Key
MovementDim_Key	Surrogate Key / Primary Key

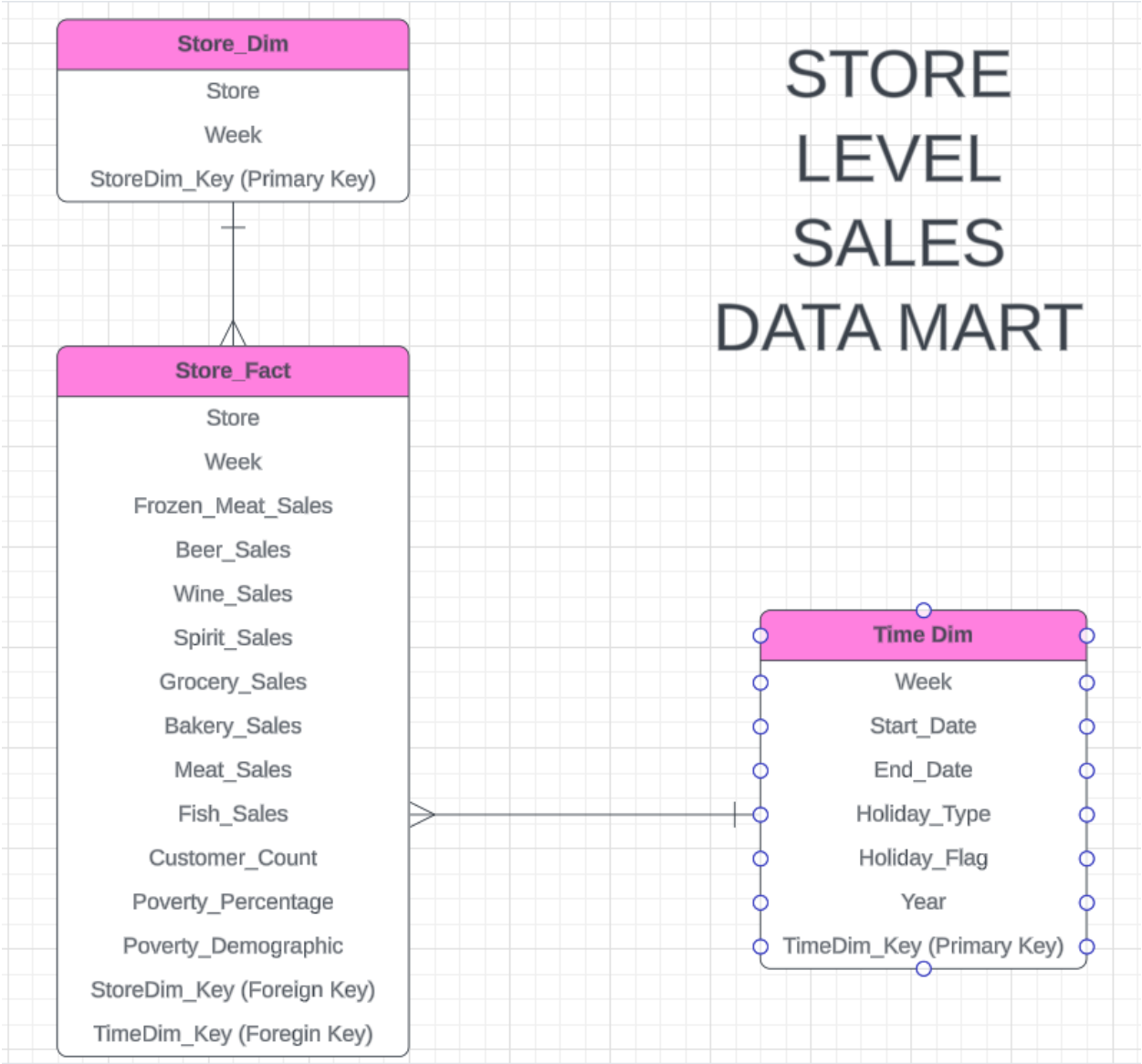
Now that all the dimensions and facts required for answering the business questions have been collected and structured, we now proceed to create a dimension versus data mart matrix, to define the relationship between them in a tabular format. The following are the independent data marts created:

1. Store Level Sales Data Mart
2. Product Level Sales Data Mart

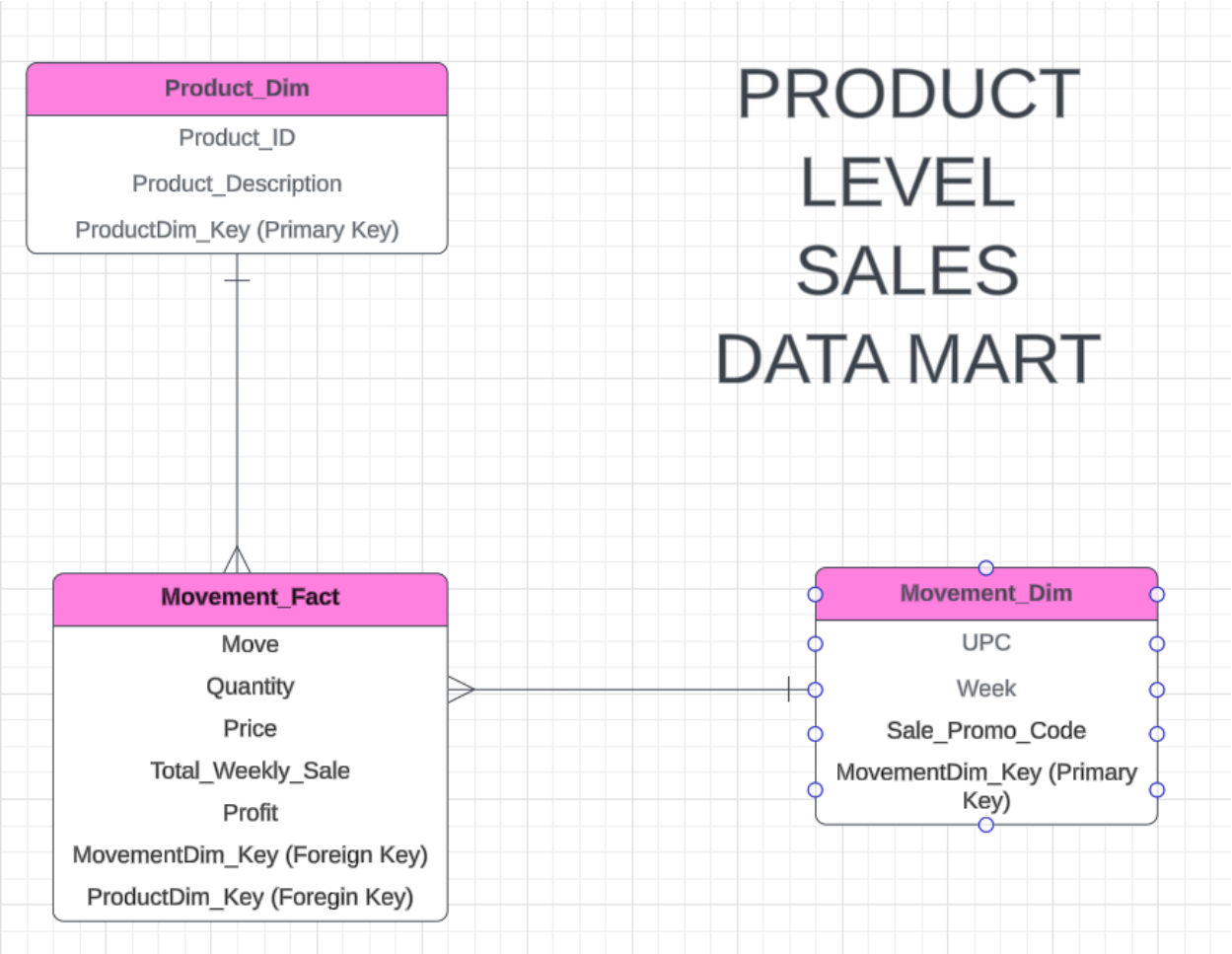
Data Mart Vs. Dimension Matrix:

Data Mart/Dimension Tables	**Time_Dim**	**Store_Dim**	**Product_Dim**	**Movement _Dim**
Store Level Sales Data Mart	✖	✖		
Product Level Sales Data Mart			✖	✖

Store Level Sales Data Mart:



Product Level Sales Data Mart:



Which are the stores that cater to people with income below $15,000? What are the products that they opt for the most?

Customer segmentation is an important analysis performed by audience-facing businesses. Profiling the customers depending on their gender, average income, region, and lifetime value gives the business an idea about the value brought in by each customer. The above question helps identify stores that cater to audiences with an average income of less than 15,000 USD.

The store-level sales data mart helps us in answering these business questions. The underlying star schema is mainly built of two elements. Firstly, we have the dimension tables in the form of Time_Dim and Store_Dim. These dimension tables have contextual information such as store number, region, week, and corresponding holidays. These are the dimension tables that are connected to a central fact table in the star schema. Store_Fact is the main fact table in this scenario. It contains numeric information like weekly sales data for various products and weekly customer traffic. It also contains a Poverty_Demogrpahic Column that comes in as the demographic information to identify the target audience.

The overall schema has store-level sales information aggregated weekly. Each row in the schema corresponds to a specific store and its weekly sales across an array of products and its associated

customer traffic and poverty demographic. This helps with ciphering the products that are mainly popular in these stores and helps with designing targeted marketing strategies.

How did the sale of alcoholic products fare during the holiday season for the period 1995-1996?

Alcoholic products have the highest markups in retail stores. Alcoholic products are among the marquee products at DFF and studying its consumer demand might help the organization build better sales and promotional strategies. In the above business question, we aim to highlight the weekly sales trend for 'Beer', 'Wine', and 'Spirits' during the holiday season for the calendar year 1995-96.

The store-level sales data mart helps us in answering this business question. The underlying star schema is mainly built of two elements. Firstly, we have the dimension tables in the form of Time_Dim and Store_Dim. These dimension tables have contextual information such as store number, region, week, and corresponding holidays. These are the dimension tables that are connected to a central fact table in the star schema. Store_Fact is the main fact table in this scenario. It contains numeric information like weekly sales data for various products and weekly customer traffic.

The store-level sales data mart helps us in querying the weekly sales of various alcoholic products. The time elements from the Time_Dim table. It helps us segregate the information as holiday vs. non-holiday. Columns like Holiday_type and Holiday_Flag will help us in further categorizing the data regarding various types of holiday weekends.

Each of these dimension tables is populated with surrogate keys, namely, StoreDim_Key and TimeDim_Key. This helps in referencing the rows in the fact table more easily.

How has the consumer influx changed year on year? How has it affected the sales of Groceries?

DFF's customer traffic is an important driver in its business. Having an innate understanding of their customers' demographic and their influx will help the organizations prepare for the season in a better way. The objective of this business question is to perform a basic time series analysis of customer count data to study the distribution of incoming customers.

Grocery is one of the top-selling categories at DFF. This could be considered as a template business question to study the customer influx against a category. This will help the business in estimating consumer traffic at various store locations spread across multiple price tiers.

For this business question, we will make use of the store-level sales data mart which has the customer count and the sales of grocery category across various stores for approximately 9 years with weekly data granularity.

Which are the products/categories that have been facing a year-on-year declining sales trend?

DFF has a wide range of categories with further classifications. Maintaining all these products in their warehouses and store locations is a behemoth of a task, which stores of these types are equipped to do. However, understanding the performance of various categories is key to understanding the demand fluctuation over the years.

The objective of this business question is to understand the demand fluctuation across various categories over the years. We make use of the store-level sales data mart which has the time dimensions included through the star schema. This helps the business in querying the weekly sales

of various categories and performing basic year-on-year analyses to understand the trend of sales. In the case of those products that highlight declining trends, DFF can consider onboarding these products or coming up with a promotional strategy to increase its demand.

Which beer brand attracts the highest coupon redemption and buyback offers?

Beer has the largest share in the alcoholic beverages market, accounting for roughly 45% of all sales. DFF has coupons and buyback offers covering the alcohol section. After studying the demand and supply fluctuations of alcoholic products, it is important to look at promotions redeemed by the buyers. Understanding the coupon redemptions and buyback offers to segment the beer brands for their promotional popularity will help drive the offers in a streamlined manner.

The product-level sales data mart focuses on the sales of specific products. This business question can serve as a template for other products and can be vertically scaled to ensure the addition of more products.

This data mart has a central 'movement_fact' table which contains quantifiable information regarding profit, sales, and coupon redemptions. The surrounding dimension tables namely 'Product_Dim' and 'Movement_Dim', provide contextual information such as store numbers, UPCs, and Sale codes. This information will help us aggregate the sales values and profits to get information on specific beers which attract more coupon redemptions.

Mapping Tables

Source Table to Staging Tables:

Source Data	Source Data Field	Mapping	Staging Table Type	Staging Table Name	Attribute
ccount.csv	GROCERY	Aggregation - SUM	Temp	Store_Stg	Grocery_Sales
	DAIRY	Aggregation - SUM	Temp	Store_Stg	Dairy_Sales
	MEAT	Aggregation - SUM	Temp	Store_Stg	Meat_Sales
	MEATFROZ	Aggregation - SUM	Temp	Store_Stg	Frozen_Meat_Sales
	FISH	Aggregation - SUM	Temp	Store_Stg	Fish_Sales
	BEER	Aggregation - SUM	Temp	Store_Stg	Beer_Sales
	WINE	Aggregation - SUM	Temp	Store_Stg	Wine_Sales
	SPIRITS	Aggregation - SUM	Temp	Store_Stg	Spirits_Sales

	BAKERY	Aggregation - SUM	Temp	Store_Stg	Bakery_Sales
	GROCERY	Aggregation - SUM	Temp	Store_Stg	Grocery_Sales
	CUSTCOUN	Aggregation - SUM	Temp	Store_Stg	Customer_Count
	STORE	Copy	Temp	Store_Stg	Store
	WEEK	Copy	Temp	Store_Stg	Week
Demo.csv	Poverty	Copy	Temp	Store_Stg	Poverty_Percent
UPCXXX.csv	UPC	Copy	Temp	Product__Stg	UPC
	DESCRIP	Copy	Temp	Product__Stg	Product_Description
		Surrogate Key	Temp	Product__Stg	ProductDim_Key
WXXX.csv		Surrogate Key	Temp	Movement_Stg	MovementDim_Key
	STORE	Copy	Temp	Movement_Stg	STORE
	UPC	Copy	Temp	Movement_Stg	UPC
	WEEK	Copy	Temp	Movement_Stg	WEEK
	MOVE	Copy	Temp	Movement_Stg	MOVE
	QTY	Copy	Temp	Movement_Stg	QTY
	PRICE	Copy	Temp	Movement_Stg	PRICE
	SALE	Copy	Temp	Movement_Stg	SALE
	PROFIT	Copy	Temp	Movement_Stg	PROFIT
Week Decode Table		Surrogate Key	Temp	Week_Lookup_Stg	TimeDim_Key
	WEEK	Copy	Temp	Week_Lookup_Stg	Week
	Start Date	Copy	Temp	Week_Lookup_Stg	Start_date
	End Date	Copy	Temp	Week_Lookup_Stg	End_Date
	Holiday	Copy	Temp	Week_Lookup_Stg	Holiday_Type

Staging Tables to Data Marts:

Source Data in Staging	**Staging Table Data Field**	**Mapping**	**Data Mart Table Type**	**Table Name**	**Attribute**
Store_Stg	Grocery_Sales	Copy	Fact	Store_Fact	Grocery_Sales
	Dairy_Sales	Copy	Fact		Dairy_Sales
	Meat_Sales	Copy	Fact		Meat_Sales

	Frozen_Meat_Sales	Copy	Fact		Frozen_Meat_Sales
	Fish_Sales	Copy	Fact		Fish_Sales
	Beer_Sales	Copy	Fact		Beer_Sales
	Wine_Sales	Copy	Fact		Wine_Sales
	Spirits_Sales	Copy	Fact		Spirits_Sales
	Bakery_Sales	Copy	Fact		Bakery_Sales
	Grocery_Sales	Copy	Fact		Grocery_Sales
	Customer_Count	Copy	Fact		Customer_Count
	Poverty_Percentage	Copy	Fact		Poverty_Percentage
	Poverty_Demographic	Rule: Poverty_Percentage * Customer Count	Fact		Poverty_Demographic
	Store	Copy	Dimension		Store
		Surrogate Key	Dimension		TimeDim_Key
		Surrogate Key	Dimension		StoreDim_Key
Movement_Stg		Surrogate Key	Dimension	Movement_Fact	MovementDim_Key
		Surrogate Key	Dimension		ProductDim_Key
	MOVE	Copy	Fact		MOVE
	QTY	Copy	Fact		QTY
	PRICE	Copy	Fact		UNIT_PRICE
		Total Weekly Sales = (Price * Move / Quantity)	Fact		TOTAL_WEEKLY_SALES
	PROFIT	Copy	Fact		PROFIT
Store_Stg	Store	Copy	Dimension	Store_Dim	Store
	Week	Copy	Dimension		Week
	StoreDim_Key	Surrogate Key	Dimension		StoreDim_Key
Product_Stg	UPC	Copy	Dimension	Product_Dim	Product_ID
	Product_Description	Copy	Dimension		Product_Description

	ProductDim_Key	Surrogate Key	Dimension		ProductDim_Key
Movement_Stg	UPC	Copy	Dimension	Movement_Dim	UPC
	STORE	Copy	Dimension		STORE
	WEEK	Copy	Dimension		WEEK
	SALE	Copy	Dimension		SALE_PROMO_CODe
		Surrogate Key	Dimension		MovementDim_Key
Week_Lookup_Stg	Week	Copy	Dimension	Time_Dim	Week
	Start_date	Copy	Dimension		Start_date
	End_Date	Copy	Dimension		End_Date
	Holiday_Type	Copy	Dimension		Holiday_Type
		Rule:	Dimension		Holiday_Flag
		Rule:	Dimension		Year
		Surrogate Key	Dimension		TimeDim_Key

Physical Design

The physical design of the data marts focuses on optimizing query performance, ensuring scalability, and maintaining data quality as the data grows. Data aggregation will involve pre-computing key metrics, such as total weekly sales and profit by product, store, and region. These aggregated values will be stored in dedicated summary tables to accelerate complex analytical queries and reporting tasks. Indexing strategies will include placing clustered indexes on primary keys (e.g., Product_ID, Store_ID, Week) to speed up record retrieval and non-clustered indexes on foreign keys to optimize join operations between fact and dimension tables. This will ensure that the system can efficiently handle increasing volumes of data and queries.

For data standardization, consistent formats for dates, numbers, and text fields will be enforced across all tables. This ensures data integrity and uniformity, facilitating smooth transformations and reducing data cleaning efforts. The storage plan will leverage scalable cloud-based platforms, such as Amazon Redshift or Snowflake, with partitioning strategies to separate data by time

periods (e.g., weekly partitions). This will allow for efficient storage management and faster query performance, with older data archived to less costly storage tiers to control costs while keeping relevant data readily accessible.

Section 5: Data Cleaning and Integration

ETL Development Plan

In the evolving realm of data warehouse development, the Extract, Transform, and Load (ETL) process plays an important role in converting raw data into a structured format that perfectly aligns with the complex business needs. This detailed process showcases extracting data from various sources which often deals with carefully handling large data. A well-defined ETL strategy is important for steering transformation tasks that are often time-intensive and intricate.

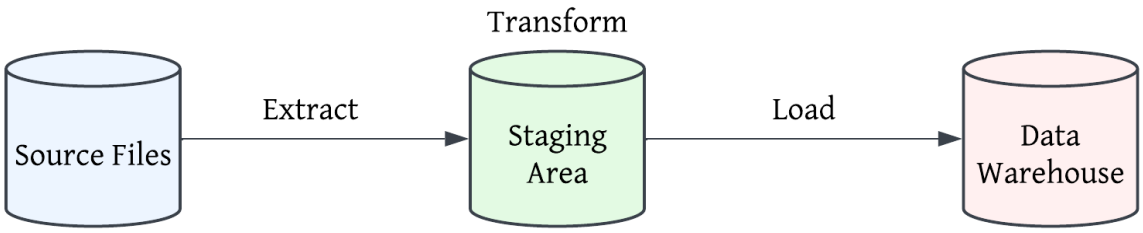
Our ETL methodology showcases three strategic phases, where each phase is designed to address various aspects of data processing. The initial phase consists of understanding the data source and selecting the important columns for extraction. This phase involves sourcing data from Excel files and other formats such as CSV or mainframe data. The extraction phase indicates the process of collecting the data.

Then comes the second phase which unfolds an in-depth analysis of the data which was extracted in the previous phase. This stage addresses a variety of challenges, some of which are data type mismatch, the presence of irrelevant data, and the presence of data that consists of errors. Here pre-established transformation rules and SQL queries play an important role, which helps in shaping the data to ensure uniformity and quality across all necessary fields and records. This phase is basically to refine the data and then prepare it for the final phase

The third phase is the final phase which marks the completion of the process, where all the errors have been removed, the data has been refined and the data is ready to be stored in final data mart tables. This data then is used for future reporting and analytical processes. Throughout this phase, a good amount of attention is given to data integrity and ensuring that the format and the data type remain consistent across the staging, fact tables, and the dimension tables.

This entire ETL process is completed and executed using Microsoft SQL Server Integration Services (SSIS), which is a powerful tool that supports the creation of different and unique packages for each ETL stage. The final step in this process includes the integration of the data and the use of surrogate keys, which makes sure in enabling efficient joins and makes sure the data loads seamlessly into the targeted fields that align with the business model requirements.

This whole process and strategic approach showcase our hard work in solving immediate business problems while building an adaptable and robust data infrastructure that can evolve based on advancements in analytics and business intelligence.



Data Extraction

This is the first step in every ETL process, here the data is sourced from specific file formats with Excel files (.xlsx) being converted into Comma Separated Values (.csv) to avoid the compatibility issue. SSIS was used to manage the data, where crucial columns were selected, and staging tables were created as a temporary point for storage before any further processing took place. Selecting the appropriate column is important as it sets the base for subsequent stages and enables smooth structuring and refinement of the data.

Data Transformation

The next phase in the process was to now utilize the extracted data as an input for the transformation process, by focusing mainly on the cleaning of the data and standardization. This involves applying transformation rules to organize the data and ensure its integrity. Key steps include:

- Generating surrogate keys
- Removing irrelevant data
- Creating derived attributes
- Removing not-so-important spaces or characters
- Conversion of data type
- Handling and removing the NULL values
- Removing duplicate values

Data Loading

This is the final phase of the process, this is where the data consistency checks are performed before the data gets loaded into data mart tables. Making sure that the quality of data was essential

to avoid any kind of inconsistency that may lead to disruption of downstream processes. This careful process guarantees that only clean, structured data is loaded into the final table.

Data Sources:

We had five data sources available from the DFF that were relevant to the problems we are solving regarding their business:

Source	Source File
Customer Count	Ccount.csv
Customer Demographics	Demo.csv
UPC Product	UPCXXX.csv
Movement	WXXX.csv
Time	Week Decode Table

Target Data:

Our Target data is populated with the help of staging tables:

Source files	Staging Tables	Data Mart Tables
Ccount.csv	Store_Stg	Store_dim and Store_fact
Demo.csv	Store_Stg	Store_dim and Store_fact
UPCXXX.csv	Product_Stg	Product_dim
WXXX.csv	Movement_stg	Movement_dim and Movement_fact
Week Decode Table	Week_Lookup_Stg	Time_dim

Data Mapping

Source File to Staging Table Mapping

Source Data	Source Data Field	Mapping	Staging Table Type	Staging Table Name	Attribute

ccount.csv	GROCERY	Aggregation - SUM	Temp	Store_Stg	Grocery_Sales
	DAIRY	Aggregation - SUM	Temp	Store_Stg	Dairy_Sales
	MEAT	Aggregation - SUM	Temp	Store_Stg	Meat_Sales
	MEATFROZ	Aggregation - SUM	Temp	Store_Stg	Frozen_Meat_ Sal es
	FISH	Aggregation - SUM	Temp	Store_Stg	Fish_Sales
	BEER	Aggregation - SUM	Temp	Store_Stg	Beer_Sales
	WINE	Aggregation - SUM	Temp	Store_Stg	Wine_Sales
	SPIRITS	Aggregation - SUM	Temp	Store_Stg	Spirits_Sales
	BAKERY	Aggregation - SUM	Temp	Store_Stg	Bakery_Sales
	CUSTCOUN	Aggregation - SUM	Temp	Store_Stg	Customer_Cou nt
	STORE	Aggregation - SUM	Temp	Store_Stg	Store
	WEEK	Aggregation - SUM	Temp	Store_Stg	Week
Demo.csv	Poverty	Copy	Temp	Store_Stg	Poverty_Perce nt
UPCXXX.cs v	UPC	Copy	Temp	Product__S tg	UPC
	DESCRIP	Copy	Temp	Product__S tg	Product_Descri ption
		Surrogate Key	Temp	Product__S	

				tg	
WXXX.csv		Surrogate Key	Temp	Movement_ stg	MovementDim _K ey
	UPC	Copy	Temp	Movement_ stg	UPC
	STORE	Copy	Temp	Movement_ stg	STORE
	WEEK	Copy	Temp	Movement_ stg	Week
	MOVE	Copy	Temp	Movement_ stg	Move
	QTY	Copy	Temp	Movement_ stg	Quantity
	PRICE	Copy	Temp	Movement_ stg	Price
	SALE	Copy	Temp	Movement_ stg	Sale_Promo_C od de
	PROFIT	Copy	Temp	Movement_ stg	Profit
Week Decode Table		Surrogate Key	Temp	Week_Loo kup_ Stg	TimeDim_Key
	WEEK	Copy	Temp	Week_Loo kup_ Stg	Week
	Start Date	Copy	Temp	Week_Loo kup_ Stg	Start_date
	Start Date	Copy	Temp	Week_Loo kup_ Stg	End_Date
	Holiday	Copy	Temp	Week_Loo kup_ Stg	Holiday_Type

Staging Table to Dimensions and Facts Mapping

Source Data in Staging	Source Data in Staging	Source Data in Staging	Data Mart Table Type	Table Name	Attribute
Store_stg	Grocery_Sales	Copy	Fact	Store_Fact	Store_Fact
	Grocery_Sales	Copy	Fact		Dairy_Sales
	Meat_Sales	Copy	Fact		Meat_Sales
	Frozen_Meat_Sales	Copy	Fact		Frozen_Meat_S ales
	Fish_Sales	Copy	Fact		Fish_Sales
	Beer_Sales	Copy	Fact		Beer_Sales
	Wine_Sales	Copy	Fact		Wine_Sales
	Spirits_Sales	Copy	Fact		Spirits_Sales
	Bakery_Sales	Copy	Fact		Bakery_Sales
	Grocery_Sales	Copy	Fact		Grocery_Sales
	Customer_Count	Copy	Fact		Customer_Count
	Poverty_Percent age	Copy	Fact		Poverty_Percent age
	Poverty_Demographic	Rule: Poverty_Percentage * Customer Count	Fact		Poverty_Demographic
	Store	Copy	Fact		Store
		Surrogate Key	Fact		TimeDim_Key

		Surrogate Key	Fact		StoreDim_Key
Movement_stg		Surrogate Key	Fact	Movement_Fact	MovementDim_ Key
		Surrogate Key	Fact		ProductDim_Key
	MOVE	Copy	Fact		MOVE
	QTY	Copy	Fact		QTY
	PRICE	Copy	Fact		UNIT_PRICE
		Total Weekly Sales = (Price * Move / Quantity)	Fact		TOTAL_WEEKLY_SALES
	PROFIT	Copy	Fact		PROFIT
Store_Stg	Store	Copy	Dimension	Store_Dim	Store
	Week	Copy	Dimension		Week
	StoreDim_Key	Surrogate Key	Dimension		StoreDim_Key
Product_Stg	UPC	Copy	Dimension	Product_Dim	Product_ID
	Product_Description	Copy	Dimension		Product_Description
	ProductDim_Key	Surrogate Key	Dimension		ProductDim_Key
Movement_stg	UPC	Copy	Dimension	Movement_Dim	UPC
	STORE	Copy	Dimension		STORE
	Week	Copy	Dimension		WEEK

	SALE_PROMO_CODE	Copy	Dimension		SALE_PROMO_CODE
	MovementDim_Key	Surrogate Key	Dimension		MovementDim_ Key
Week_Lookup_Stg	Week	Copy	Dimension	Time_Dim	Week
	Start_date	Copy	Dimension		Start_date
	End_Date	Copy	Dimension		End_Date
	Holiday_Type	Copy	Dimension		Holiday_Type
		Rule: This is a ‘Y’ or ‘N’ flag. The flag will be set to ‘Y’ for holidays	Dimension		Holiday_Flag
		Rule: format yyyy’ is extracted from the Start_date column.	Dimension		Year
		Surrogate Key	Dimension		TimeDim_Key

Data Extraction Rules

While building the data warehouse, making sure that accurate extraction of data consisting of all the required attributes is important for the effective integration of information from different sources or systems. For the Dominick Fast Food (DFF) data set, the initial phase of the ETL process involved carefully selecting and extracting the files which were necessary to answer the given business questions, this phase consisted of several types of data sources, such as:

- Comma-Separated Value (*CSV*) files: *demo.csv* and *Ccount.csv*

- Excel files (*xlsx, xls*): *weektables.xlsx*, which was created in the first step by transcribing data from the DFF data manual.
- SAS files (Statistical Files)

The main goal of the extraction process was to imbibe the data in the best way possible which led to the choice of having *CSV* as the file format. This is the reason why each selected file was converted to a *CSV* format. The data which was obtained from their files was then imported into SSIS using the import/export wizard during the SSIS package creation stage.

During the data extraction process, two important rules were followed:

- All files were converted into *CSV* format
- Only important columns and attributes that were related to the project were extracted

Data Transformation Rules

The data transformation phase is the second phase of the process which comes after completing the data extraction step. This phase plays a critical role in understanding the quality of the data that is being used in the project, making sure that the data is uniform and consistent across all the intermediate tables.

During this process, many different operations were performed to make sure that the data was formatted correctly, which made sure that the consistency, integrity, and overall quality were improved. Specific transformation rules and procedures were used to clean the data before loading it into the data marts. Some of these rules require the creation of new columns and modifying the existing ones such that it aligned with the required data types and formats. One important aspect of this process was the generation of the surrogate keys, where auto-generated values were assigned at the time when the dimension tables were created.

The applied transformation rules included:

- Creation of Surrogate Keys: Surrogate keys were created so that they could act as primary keys for the dimension tables before the data was loaded into the data mart. For example, keys such as *Week*, *MovementDim_Key*, *ProductDim_Key*, and *StoreKey* were generated to uniquely represent the week, movement, product, and store dimensions.

- Removing Inconsistent Records: Files such as *demo* and *ccount* had columns that had inconsistent values. These were taken care of using *SQL DELETE* operations to remove such inconsistent records

- Removing NULL Values: The *ccount.csv* file, after being imported into the staging tables, consisted of multiple NULL values that needed proper handling.

- Removing Extra spaces and characters : Some columns which are present in the ccount and demo data had issues where they consisted of unnecessary spaces or special characters such as (''). Transformation operations were applied to make sure that these extra values were eliminated to improve the consistency and quality across all records

- Derived Attributes: The demo file was used to generate derived columns such as total_weekly_sales, which was important for business analysis related to weekly sales.

- Conversion of Data: Changes were made to the data types within the temporary tables related to store, product, movement, and time before loading them into the staging tables to achieve consistent data type formatting for every attribute across the extracted columns

- Removing Duplicates: If any table consisted of duplicate data by aggregating their distinguishing attributes, they were removed with the help of the "sort" function in the SSIS. For example, duplicate values in the movement staging tables were removed by combining columns Store, UPC, and Week to identify each row uniquely.

Data Aggregation Plan

The final stage of the ETL process focuses on data aggregation, which can greatly improve the efficiency of data warehouse operations. Incorporating aggregation among the tables helps in reducing the number of data records compared to the original base tables. This approach makes sure that the performance is optimized by allowing most queries to run against the aggregate table instead of the base table which leads to faster query execution time.
In DFF's scenario, we chose to have a week level aggregation for each store in the store fact table. For this reason, aggregators were added in the data flow task for store fact creation. The data was grouped with respect to store and week and aggregated for their total sum. The final fact table now has data in a format that would make it easy for the user to query on a weekly granularity, at the same time can range it over store level as well.

Organization of Data Staging Area

The data staging area plays an important role and acts as the initial repository for all the extracted data that is originating from diverse sources. In this zone, data from different origins are combined and stored in separate staging tables. All data processing activities are performed among these tables before we advance to the next stage of the data warehousing process. For our project, two separate databases have been set up: the first one is set up for the data staging area, and the second

one is for the final data mart. In the staging area, data is loaded into tables that are specifically designed for each source file.

Store_stg: This table holds the data from the customer count file and consists of important columns such as store ID, week, and sales data for an array of categories (e.g. grocery, dairy, meat, frozen meat, fish, beer, wine, spirits, bakery). This table also includes customer count and poverty percentage data. Cleaning and transformation steps involved handling the missing values, standardizing data formats, and dropping invalid rows. A set of the table served as the source for populating the Store_Fact table.

Movement_stg: This table stores data from the movement files, which includes columns such as Store, UPC, week, units sold (move), quantity, price, and profit. Additional derived fields including total weekly sales (calculated as Price * Move / Quantity), were created during the transformation. We made sure that the data went under cleansing to handle the outliers and missing values, making sure that the data was consistent before loading into the Movement_Fact table.

Week_lookup_Stg: All columns from the week files were extracted into this table which includes the data from the start and end dates of weeks. This table required less cleaning but played an important role in defining the Time_Dim table. A rule was applied to extract the year from the start date column, and a holiday flag was set as 'Y' for holiday weeks.

Product_stg: Data from the UPC files, including product descriptions and identifiers, was loaded into this table. This staging table was then cleaned to remove special characters, standardize the product names and also make sure of the unique product ID's. The Product_stg table fed into the Product_Dim table.

Procedure for Data Extraction and Loading

Data Extraction Procedure:

store_stg:

- The extraction process here involved pulling data from the source file "ccount.csv" and "demo.csv", then populating the staging table with fields which included Store, Week, Grocery_Sales, Dairy_Sales, Meat_Sales, Frozen_Meat_Sales, Fish_Sales, Beer_Sales, Wine_Sales, Spirit_Sales, Bakery_Sales, Customer_Count, and Poverty_Percentage. Before the extraction took place. Many efforts were made to ensure all the fields were converted into appropriate date formats.

- The data extractions process was planned in such a manner that control flow tasks were used to drop any existing tables and create a new table in the

ISTM_GRP5_STAGING_AREA database. This was followed by a data flow task which consisted of a flat file source task, in our case, a csv file and a destination as the store_stg table. The SSIS toolbox's advanced editor aided in mapping the appropriate columns from the incoming source to the final staging table.

- Subsequently, the relevant file was loaded into the import wizard as a source file, paving the way for the creation of a comprehensive SSIS package which was designed to efficiently extract and integrate the data into the staging table. This process not only facilitates organized data management but also makes sure that information is accurate and readable within the designated database.

Week_Lookup_stg:

- The first step in the data integration process involved manually transcribing the data from the DFF manual into a week.txt file, by utilizing a tab-delimited format, the information was imported into Excel and then was further transformed into a CSV file for streamlined processing.

- To make sure there is systematic transfer of information, all values among the dataset were mapped from the source file to the dedicated staging table leveraging the capabilities of the SSIS import wizard. A connection to the Mays server was then established and the files were then loaded into the import wizard as a source.

- An SSIS package was created to extract and integrate the data into the staging table which included fields such as Week, Start_date, End_Date, and Holiday_Type.

Movement_Stg:

- Data Extraction WXXX.csv files where (XXX) represent the category code, involves retrieving information to specific columns such as Store, UPC, Week, Move, Quantity, Price, Sale, and Profit. These all were stored in the temporary table where data cleaning and transformation procedures were performed. All the fields were converted to their relevant data types and relevant cleaning procedures were implemented such as if the "Price" field was null, then it was set to 0.

- Additional data transformations were performed to convert the data types and were stored in the movement staging table. Duplicate data was removed using unique identifiers by combining the Store, UPC, and Week columns. The cleaned data was then stored in the staging table.

- The SSIS Import/Export Wizard was used to map values from the source file to the staging table. A server connection was established and an SSIS package was created to extract and also integrate data into the staging table. This table was important to track UPC, Week, Move, Quantity, Price, Sale, and Profit for each UPC in each store

Product_Stg:

- Data was extracted from UPCXXX.csv files (where XXX represents the category code). The extraction process involved pulling information from UPC, Product_Description, and other relevant product details. Similar transformations and cleaning according to the Movement table were performed after the data was loaded into the Product Temp table. The data was then cleaned, and all the duplicates were removed using the unique identifying columns Com_code and UPC.

- Import/Export Wizard will be used to map values from source files to the staging table. A connection was established and an SSIS package was created to extract and integrate the data into the staging table.

Data Loading Procedure

Store_Dim Loading:

- Source: Store_Stg staging table
- Procedure:
    - Establish a connection to the Mays server (infodata 16.mbps.tamu.edu)
    - Use control flow tasks to drop any existing tables and create a new table for staging.
    - Extract Store and Week from Store_Stg
    - Generate a surrogate key (StoreDim_Key)
    - Load data into Store_Dim table with the following attributes:
        - StoreDim_Key (Surrogate Key)
        - Store (Set as non-null)
        - Week
    - Create an SSIS package to extract and integrate the data into the dimension table

Product_Dim Loading:

- Source: Product_Stg staging table

- Procedure:
    - Establish a connection to the Mays server (infodata16.mbs.tamu.edu)
    - Use control flow tasks to drop any existing tables and create a new table for staging.
    - Extract UPC (as Product_ID) and Product_Description from Product_stg
    - Generate a surrogate key (Product_Dim_Key)
    - Load into the Product_Dim table with the following attributes:
        - ProductDim_Key (Surrogate Key)
        - Product_ID (UPC)
        - Product_Description
    - Created an SSIS package to extract and integrate the data into the dimension table

Movement_Dim Loading:

- Source: Movement_Stg staging table
- Procedure:
    - Establish a connection to the Mays server (infodata 16.mbps.tamu.edu)
    - Use control flow tasks to drop any existing tables and create a new table for staging.
    - Extract UPC, Week, and Sales_Promo_Code from Movement_Stg
    - Generate a surrogate key (MovementDim_Key)
    - Load data into Movement_Dim table with the following attributes:
        - MovementDim_Key (Surrogate Key)
        - STORE
        - UPC
        - WEEK
        - SALE_PROMO_CODE
    - Create an SSIS Package to extract and integrate the data into the dimension table

Time_Dim Loading:

- Source: Week_Lookup_Stg staging table
- Procedure:
    - Establish a connection to the Mays server
    - Use control flow tasks to drop any existing tables and create a new table for staging.

- Extract Week, Start_date, End_Date, and Holiday_Type from Week_Lookup_Stg
- Generate a surrogate key (TimeDim_Key)
- Extract year from start_date
- Calculate and load Holiday_Flag (set to 'Y' for holidays, 'N' otherwise)
- Load the data into Time_Dim table with the following attributes:
    - TimeDim_Key (Surrogate Key)
    - Week (set as unique and non-null)
    - Start_date
    - End_date
    - Year
    - Holiday_Type
    - Holiday_Flag
- Create an SSIS package to extract and integrate the data into the dimension table

Movement_fact Loading:

- Source: Movement staging table and Product staging table
- Procedure:
    - Extract MovementDim_Key, MOVE, QTY, UNIT_PRICE, PROFIT from the Movement Table and ProductDim_Key from the Product Staging Table
    - Add a Derived column attribute i.e. TOTAL_WEEKLY_SALES in the workflow and sort the columns according to the UPC codes.
    - Similarly, sort the columns from the Product staging table according to the UPC Codes and then join it with the Movement staging table to load the relevant columns into the Movement Fact Table.
    - Load the data into the Movement_fact table with the following attributes:
        - MovementDim_Key (Surrogate Key)
        - ProductDim_Key (Surrogate Key)
        - TOTAL_WEEKLY_SALES
        - MOVE
        - QTY
        - UNIT_PRICE

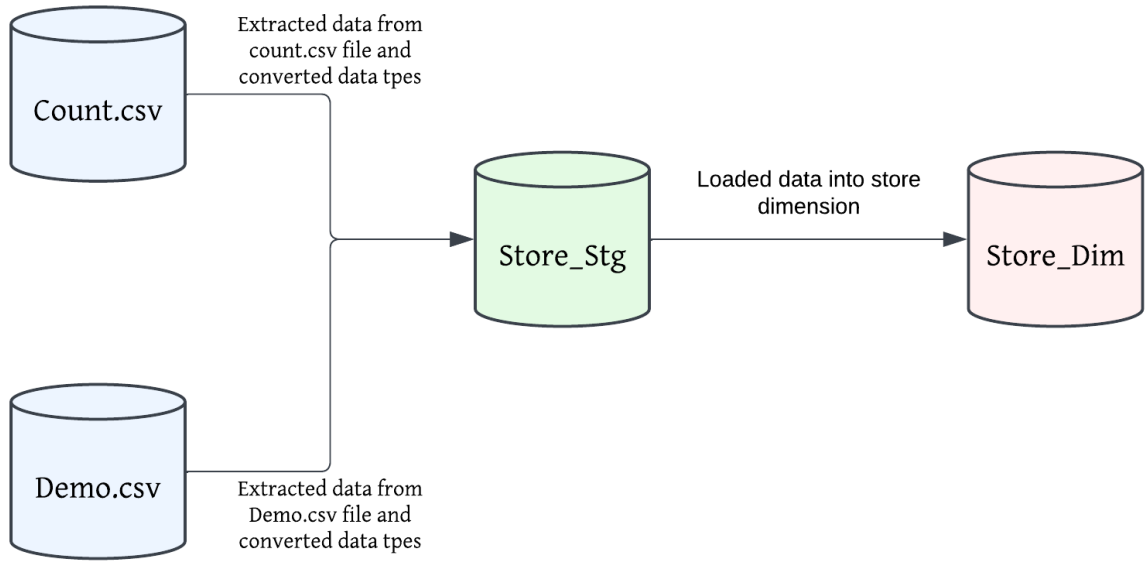
        - PROFIT
    - Create an SSIS package to extract and integrate the data into the fact table

Store_fact Loading:

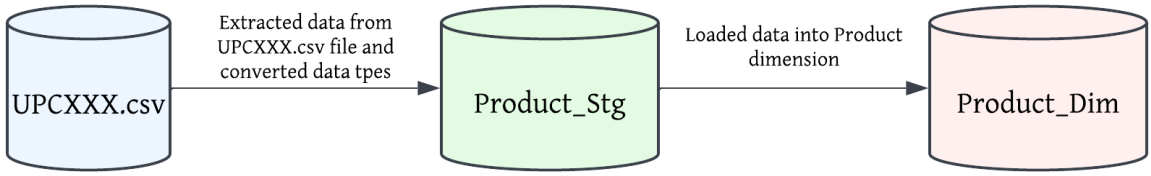
- Source: Store staging table, Store dimension table, Time dimension table
- Procedure:
    - Extract Grocery_Sales, Dairy_Sales, Meat_Sales, Frozen_Meat_Sales, Fish_Sales, Beer_Sales, Wine_Sales, Spirits_Sales, Bakery_Sales, Grocery_Sales, Customer_Count, Poverty_Percentage, Poverty_Demographic, Store, TimeDim_Key, StoreDim_Key from the Store Staging Table, Store_dim and Time_dim tables
    - Similarly sort the columns from the Product staging table according to the UPC Codes and then join it with the Movement staging table to load the relevant columns into the Movement Fact Table.
    - Load the data into the Movement_fact table with the following attributes:
        - MovementDim_Key (Surrogate Key)
        - ProductDim_Key (Surrogate Key)
        - TOTAL_WEEKLY_SALES
        - MOVE
        - QTY
        - UNIT_PRICE
        - PROFIT
    - Create an SSIS package to extract and integrate the data into the fact table

ETL for Dimension table

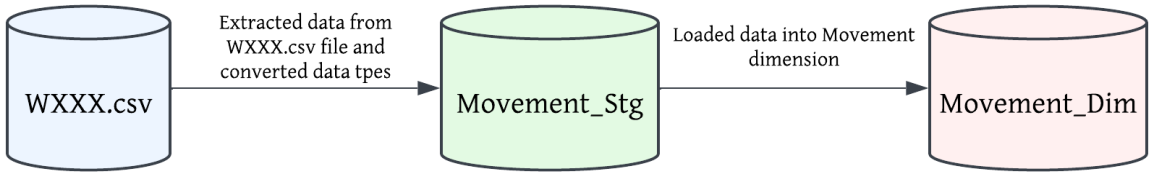
ETL for Store Dimension:



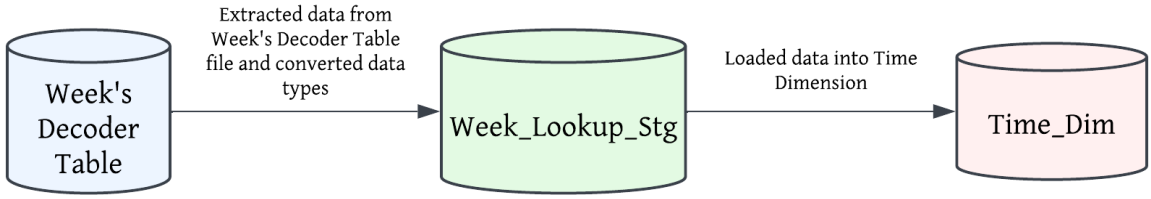
ETL for Product Dimension:



ETL for Movement Dimension:

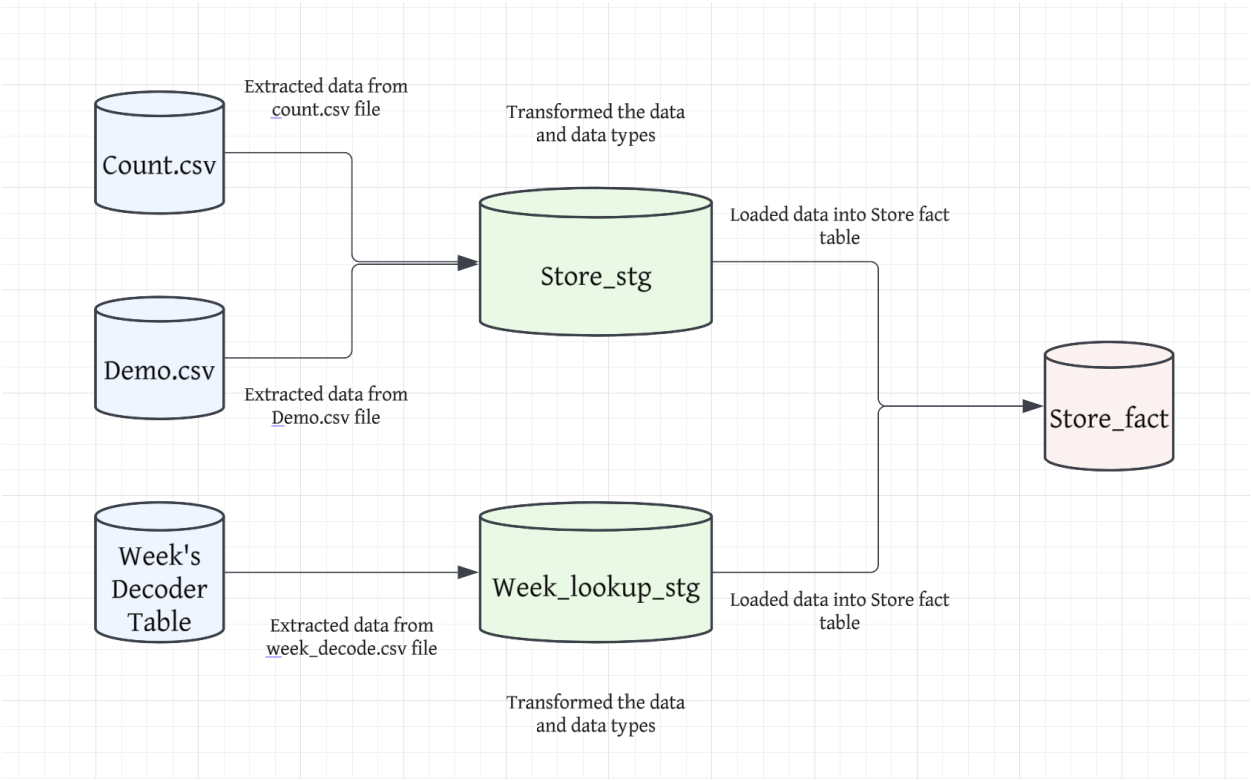


ETL for Time Dimension:

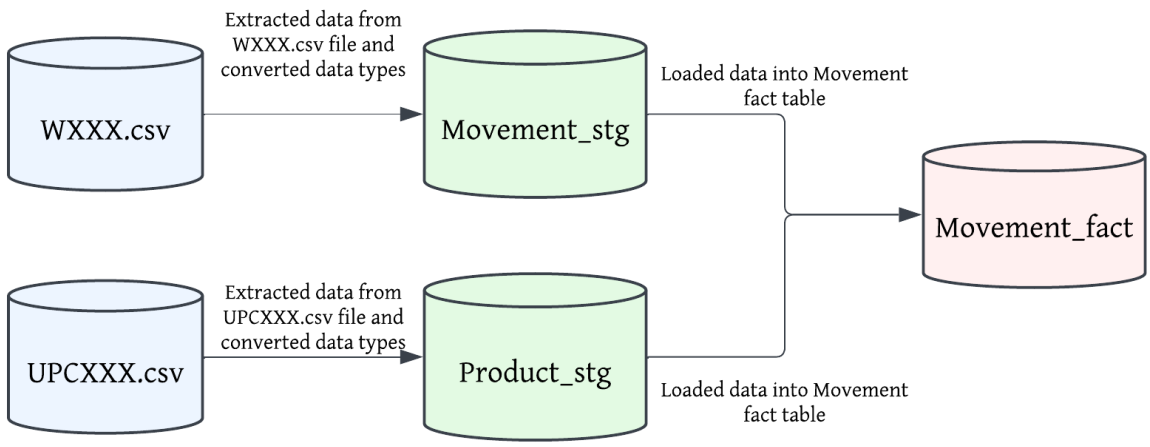


ETL for Fact Table

ETL for Store Fact Table:



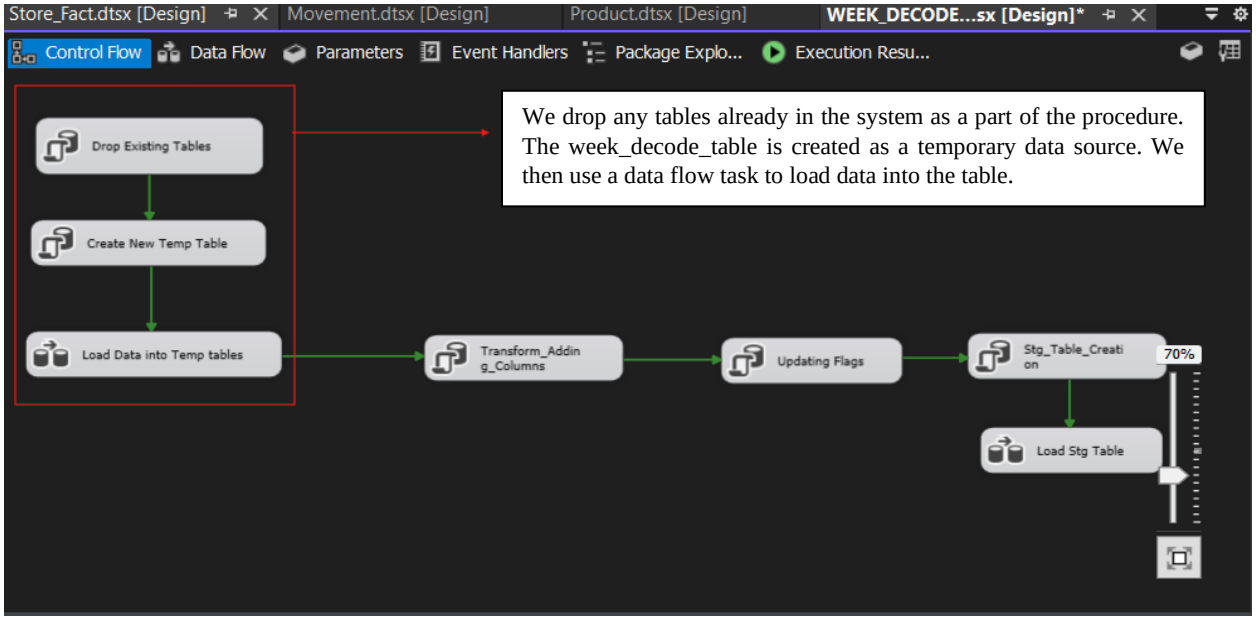
ETL for Movement Fact Table:



Section 6: Implementation of ETL:

1. Week Decode Dimension Table Workflow:

The Week_decode.csv file was extracted into a temporary table. The workflow was built using SSIS and created the necessary table with the data in the Staging Area which can be queried using SSMS.



Execute SQL Task Editor

Configure the properties required to run SQL statements and stored procedures using the selected connection.

General
Parameter Mapping
Result Set
Expressions

General	
Name	**Create New Temp Table**
Description	**Execute SQL Task**
Options	
TimeOut	**0**
CodePage	**1252**
TypeConversionMode	**Allowed**
Result Set	
ResultSet	**None**
SQL Statement	
ConnectionType	**OLE DB**
Connection	**infodata16.mbs.tamu.edu**
SQLSourceType	**Direct input**
SQLStatement	**USE ISTM637_GRP5_STAGING_AREA_2024**
IsQueryStoredProcedure	False
BypassPrepare	**True**

Name
Specifies the name of the task.

Browse... Build Query... Parse Query

OK Cancel Help

Execute SQL Task Editor

Configure the properties required to run SQL statements and stored procedures using the selected connection.

General
Parameter Mapping
Result Set
Expressions

General

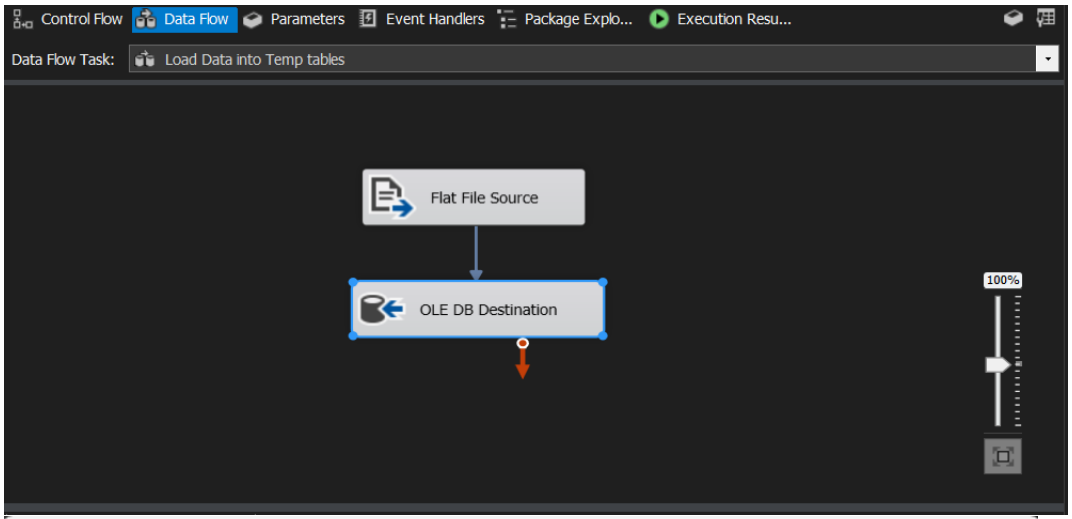
Enter SQL Query

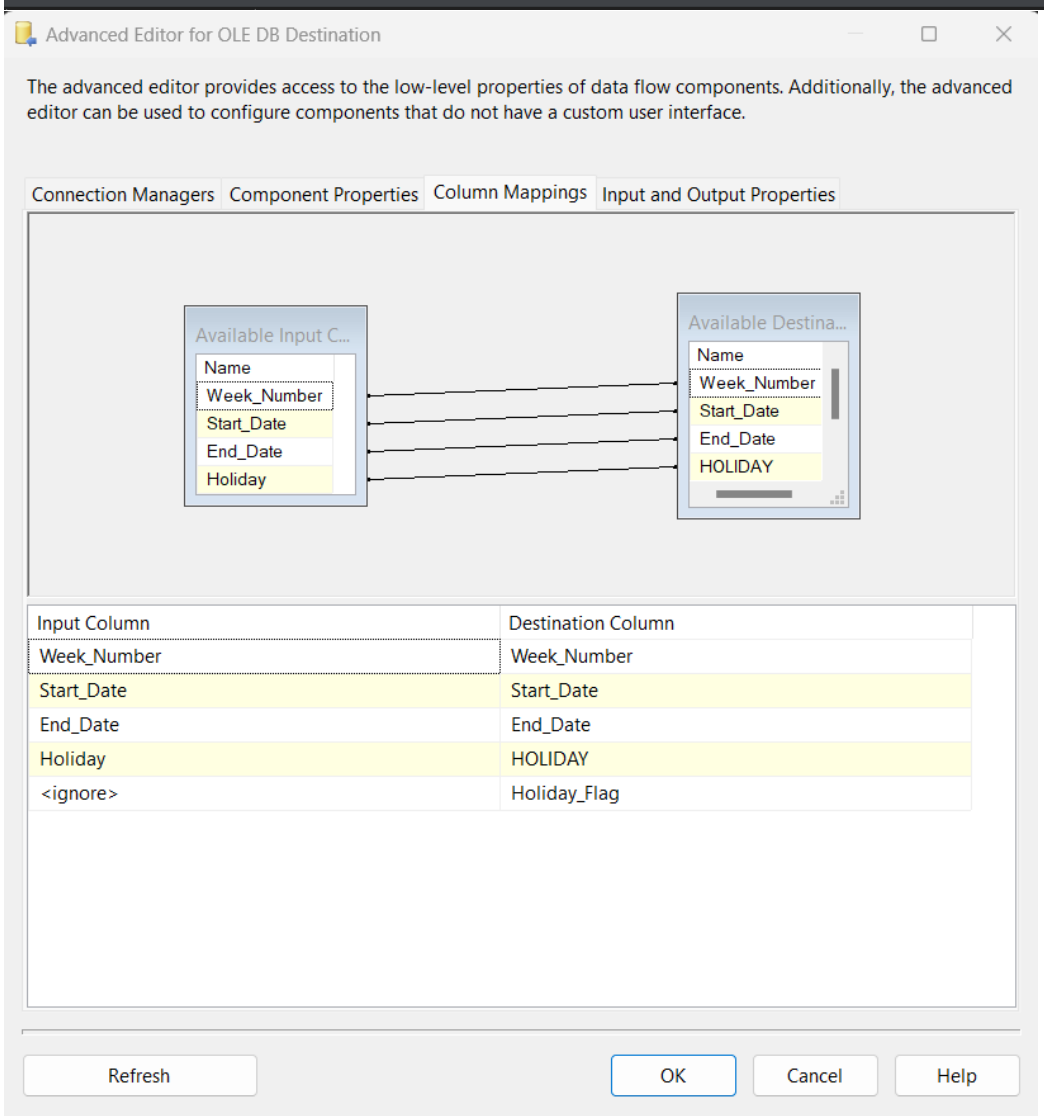
```
Use ISTM637_GRP5_STAGING_AREA_2024;

Create Table dbo.week_decode_table (Week_Number Int Not
Null,Start_Date Date  Not Null,End_Date Date  Not Null,
HOLIDAY Varchar(50)
);
```

OK Cancel

Browse... Build Query... Parse Query

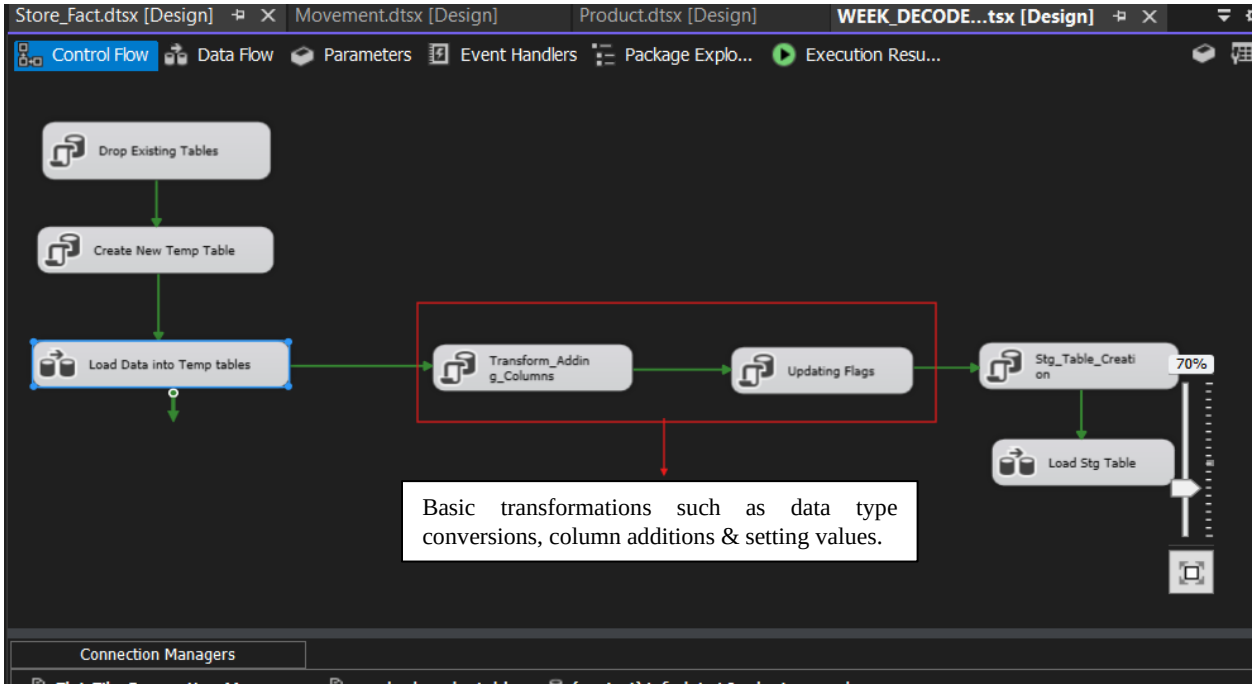
OK Cancel Help

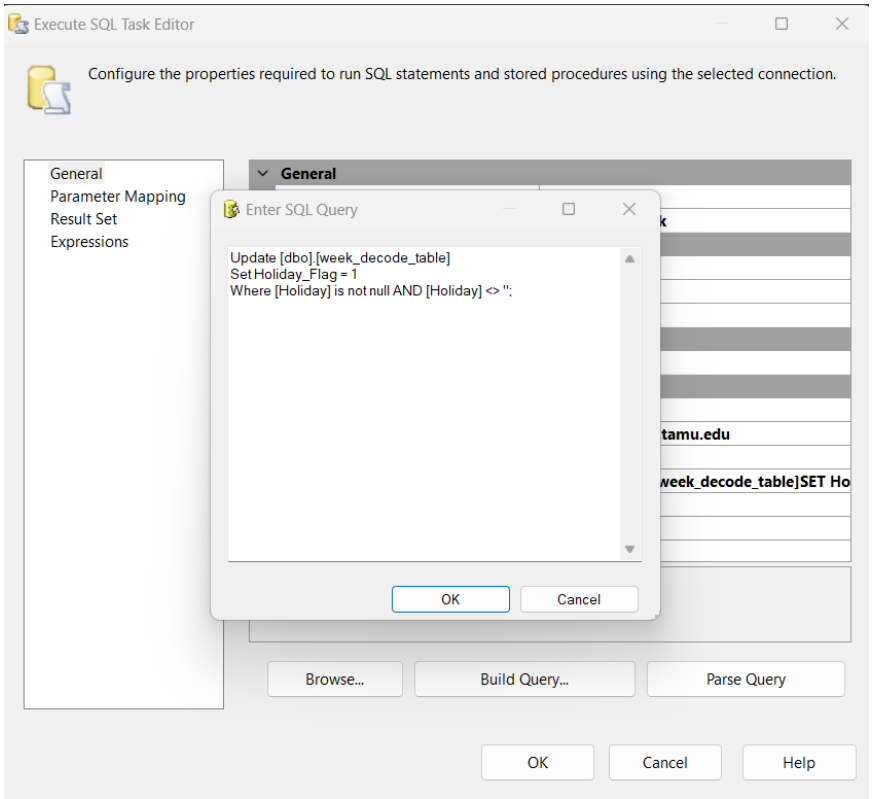


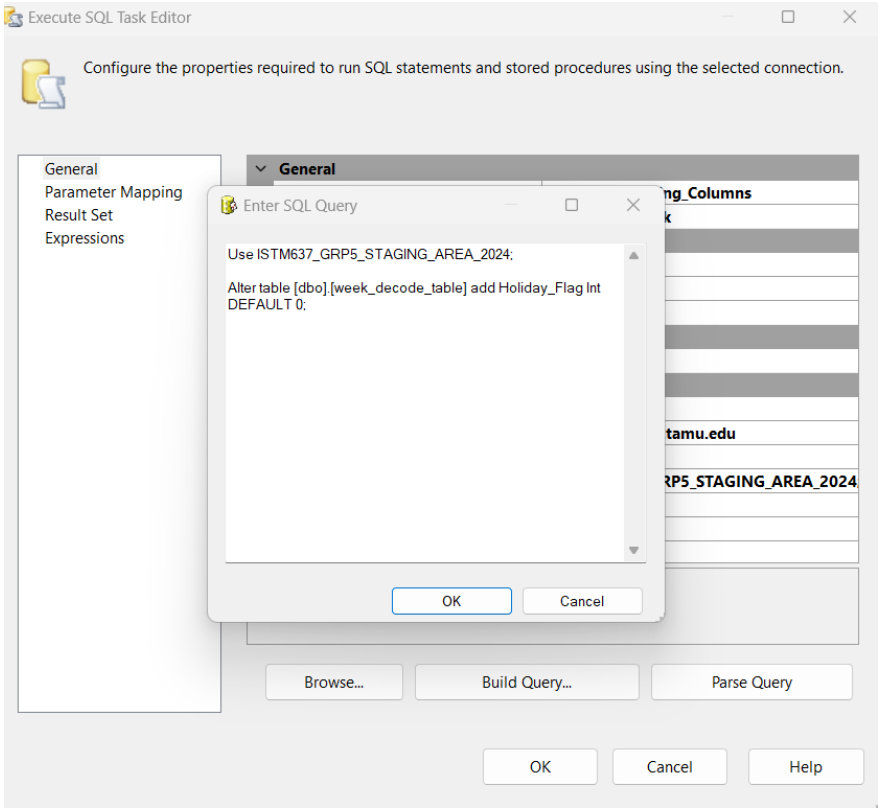


Input Column	Destination Column
Week_Number	Week_Number
Start_Date	Start_Date
End_Date	End_Date
Holiday	HOLIDAY
<ignore>	Holiday_Flag

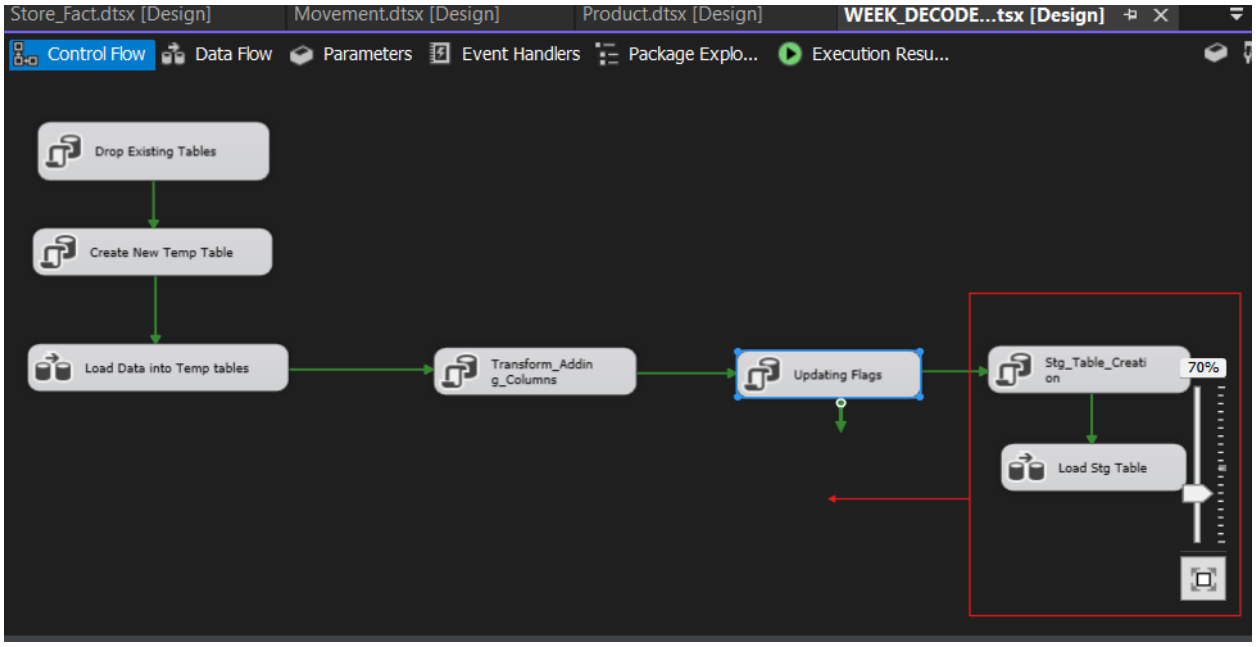
2. Week Decode Table Data Transformation:

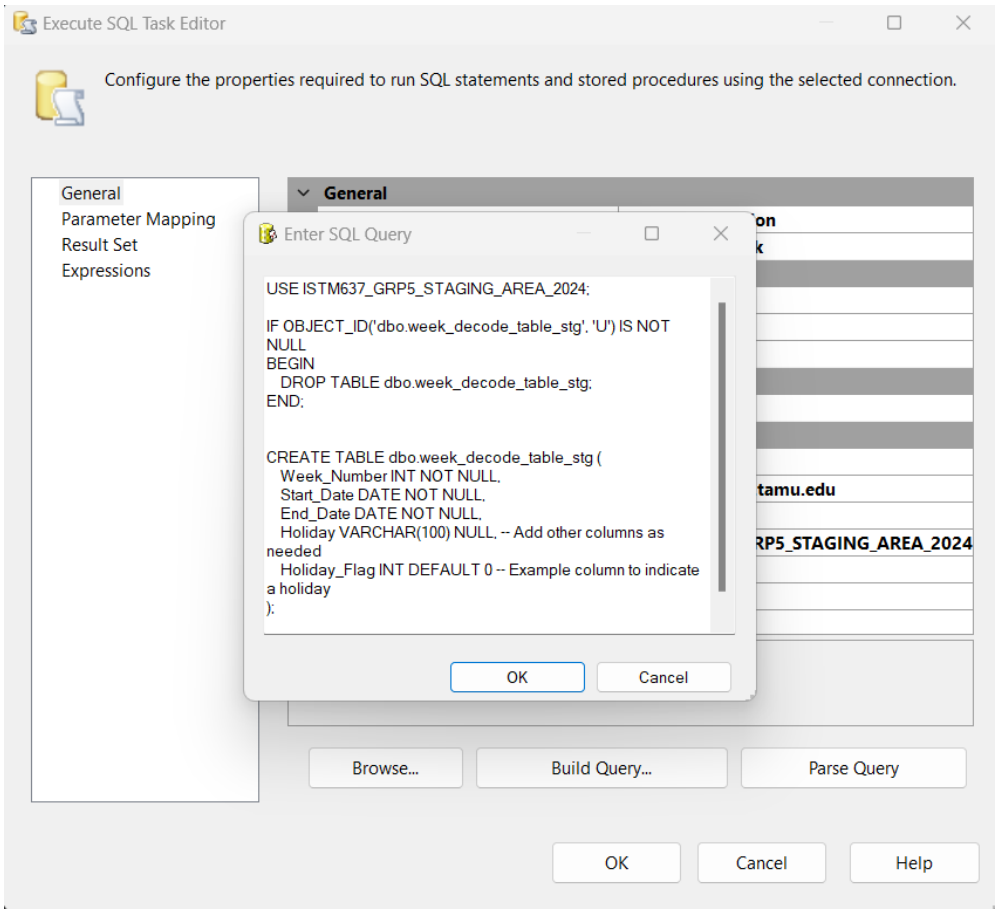




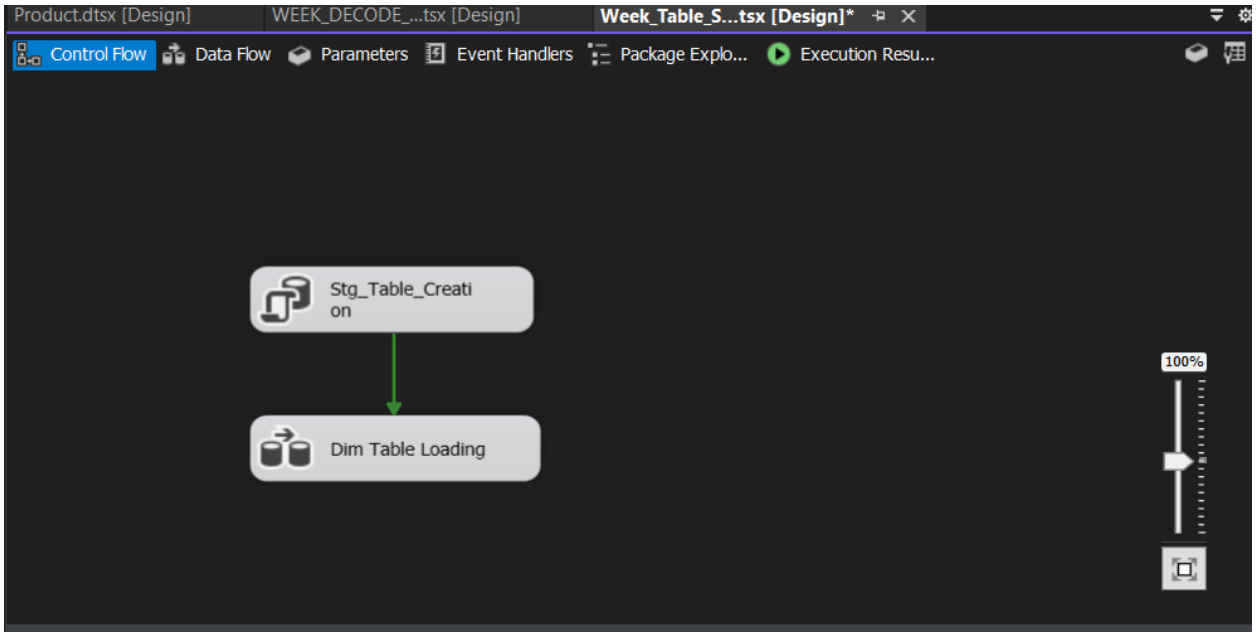


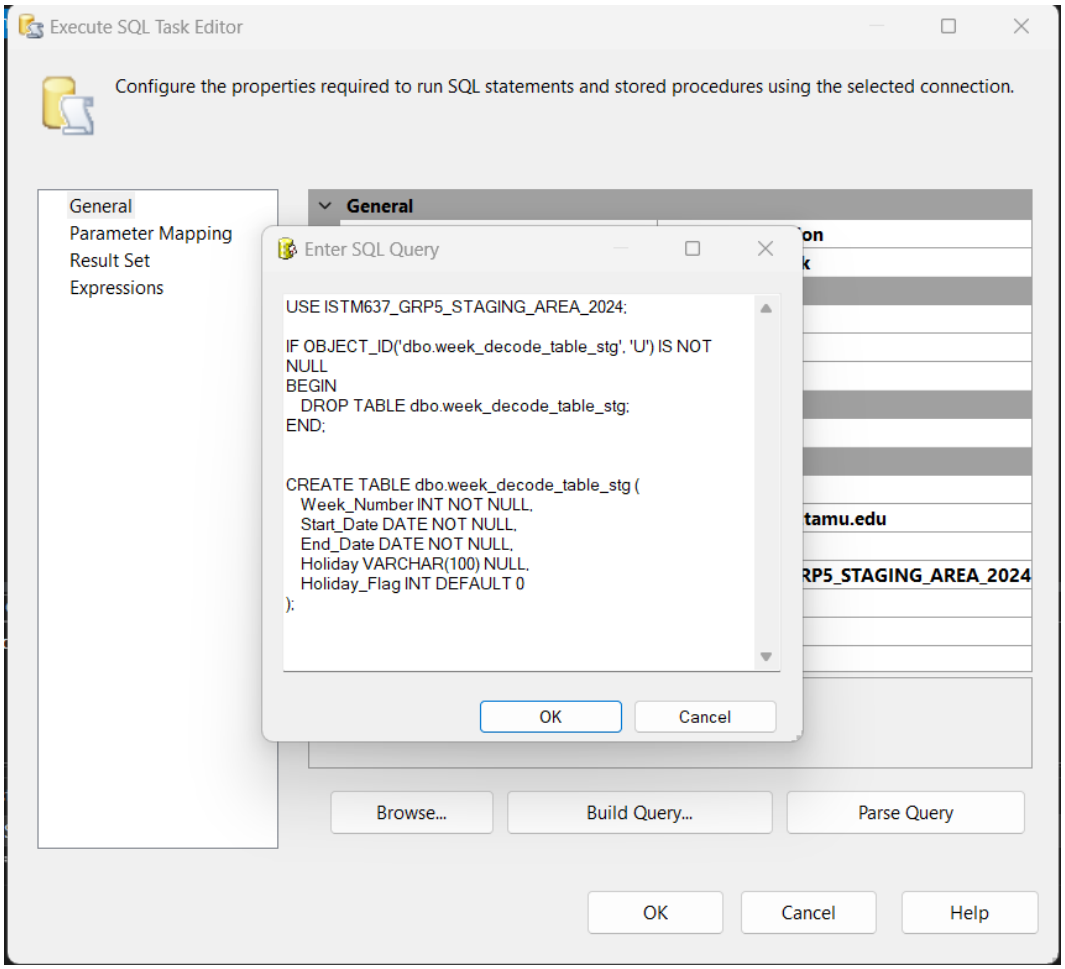
3. Week Decode Table Data Loading:



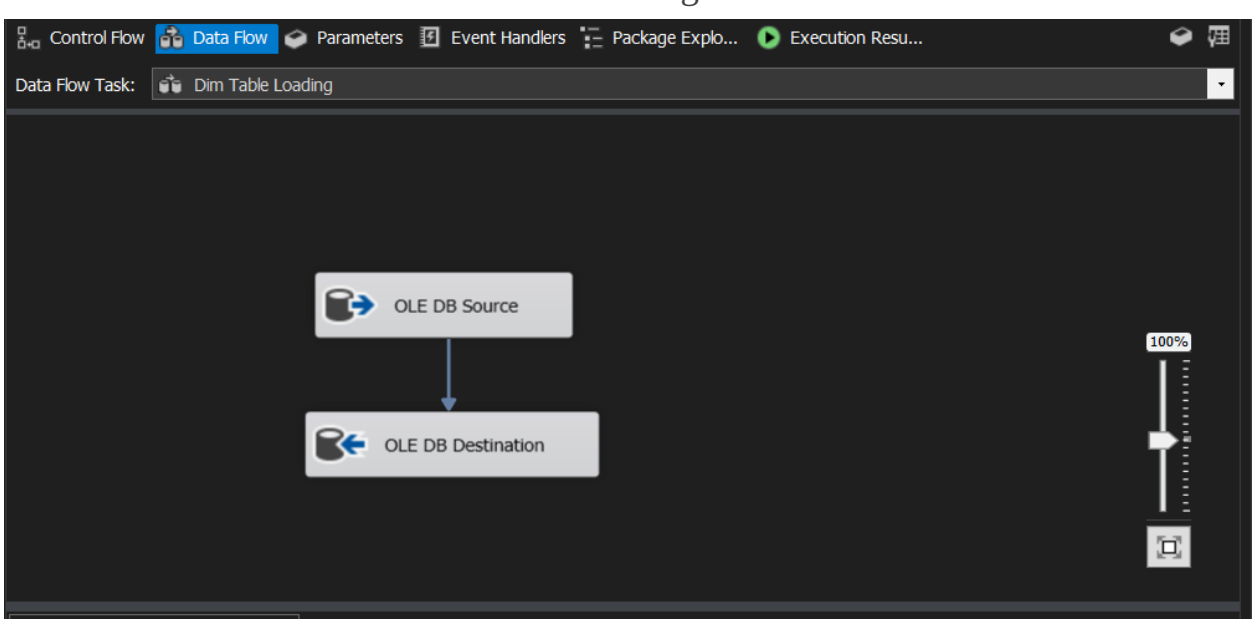


4. Week Decode Table Dimension Creation:





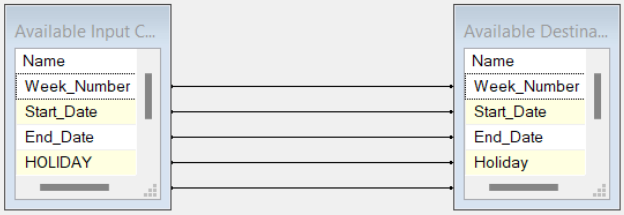
5. Week Decode Table Dimension Loading:



Advanced Editor for OLE DB Destination

The advanced editor provides access to the low-level properties of data flow components. Additionally, the advanced editor can be used to configure components that do not have a custom user interface.

Connection Managers | Component Properties | Column Mappings | Input and Output Properties

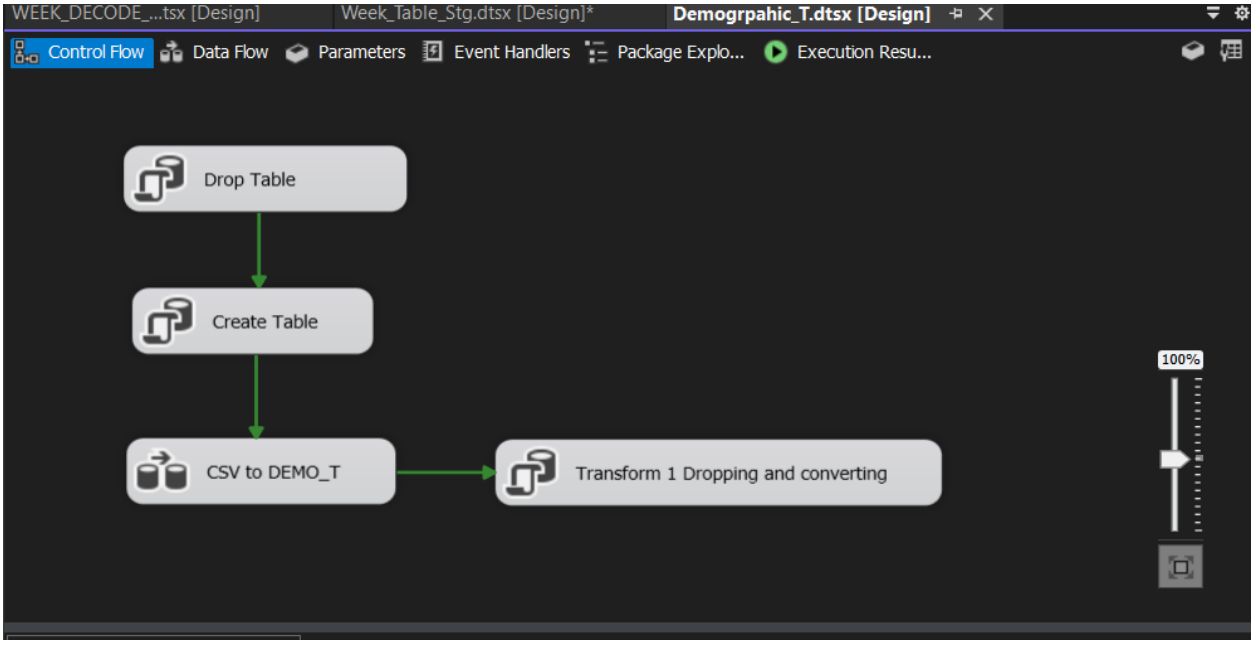


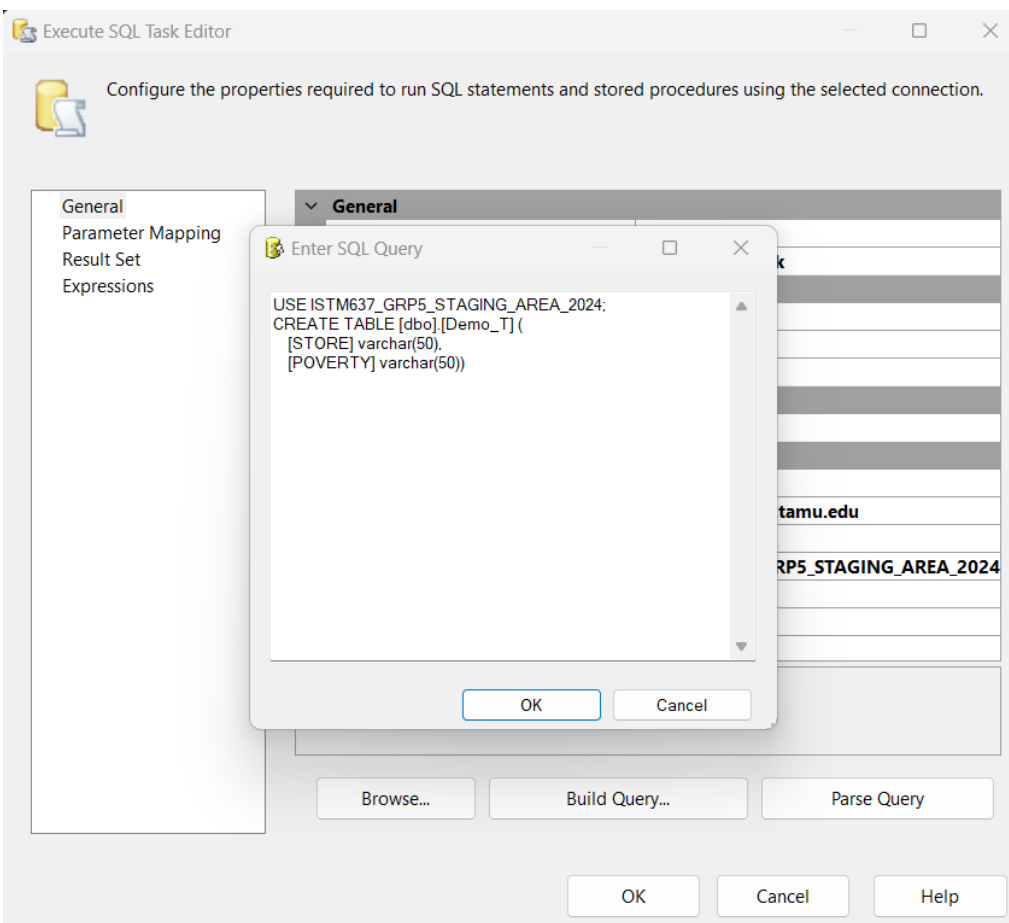
Input Column	Destination Column
Week_Number	Week_Number
Start_Date	Start_Date
End_Date	End_Date
HOLIDAY	Holiday
Holiday_Flag	Holiday_Flag

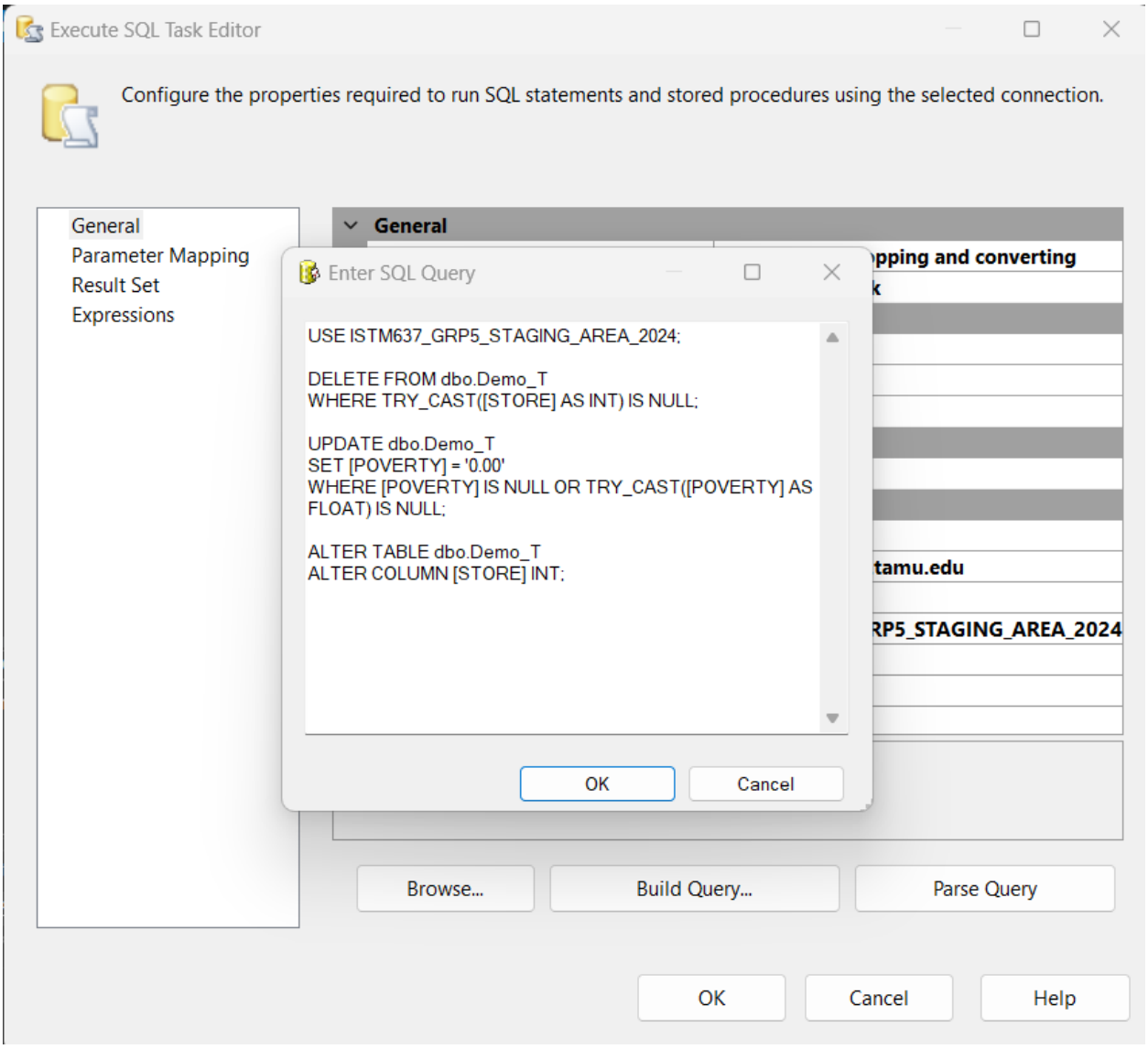
Refresh | OK | Cancel | Help

6. Store Dimension Table Workflow:

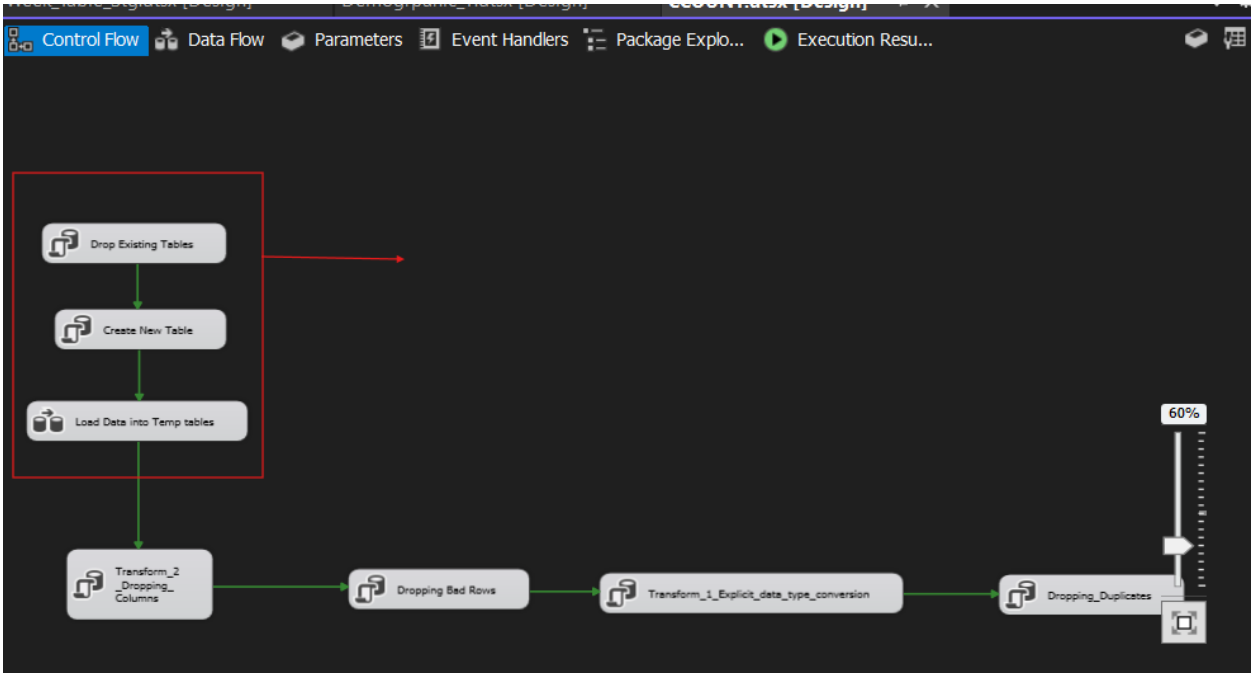
Demographic Table Data Extraction

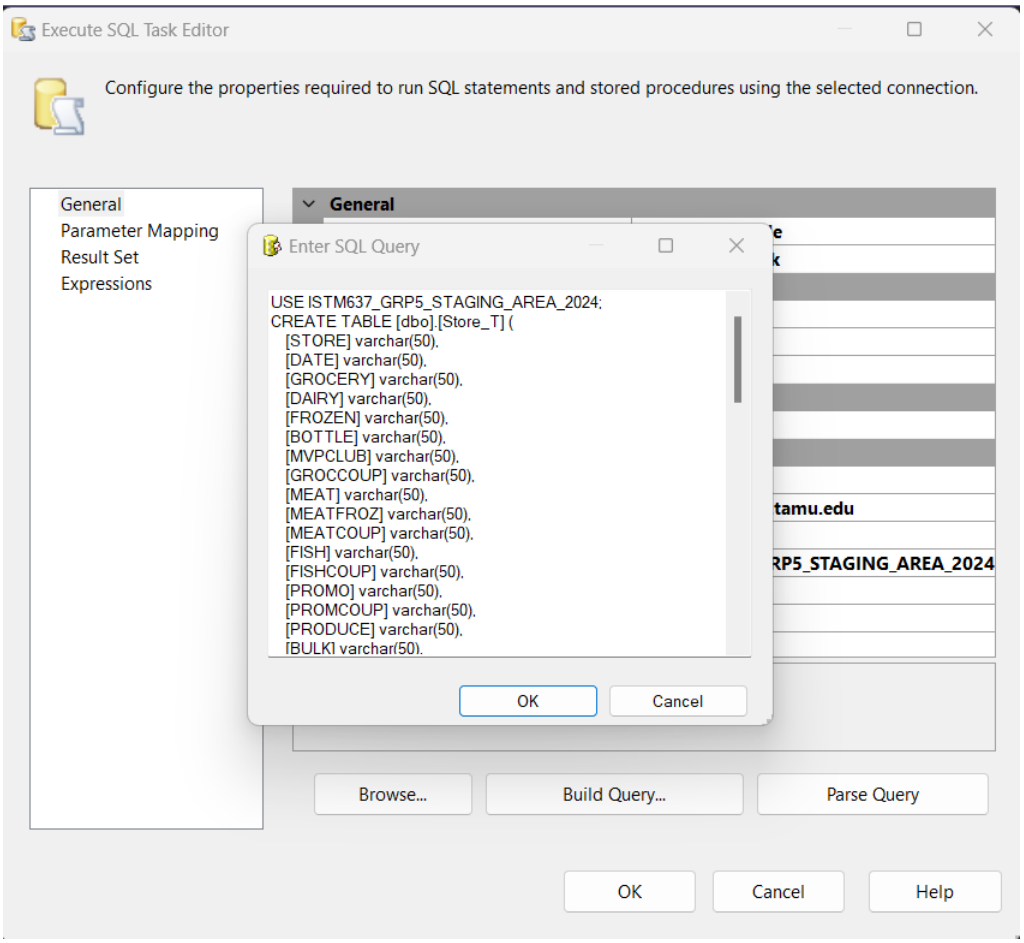




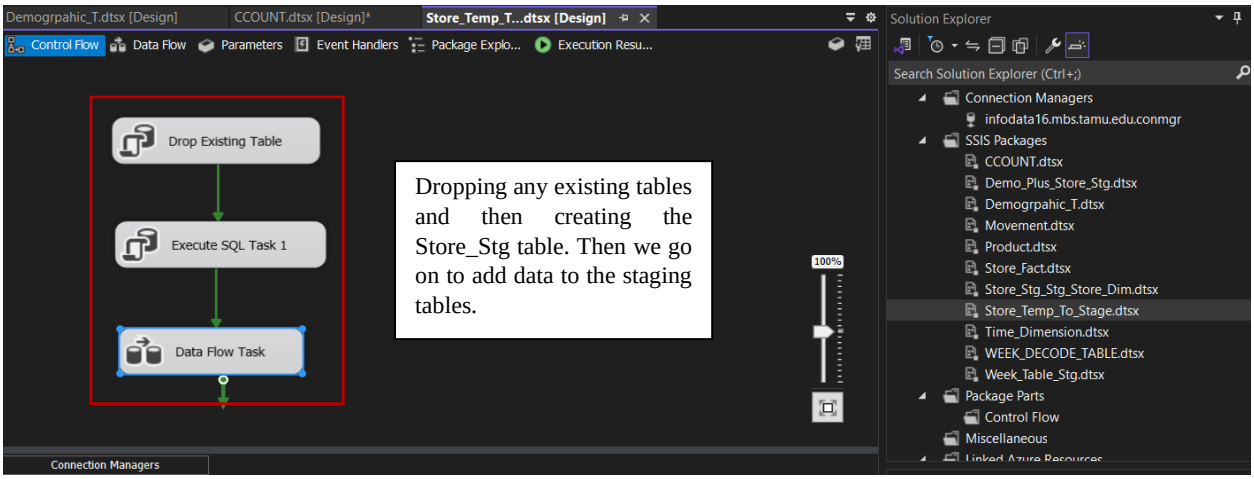


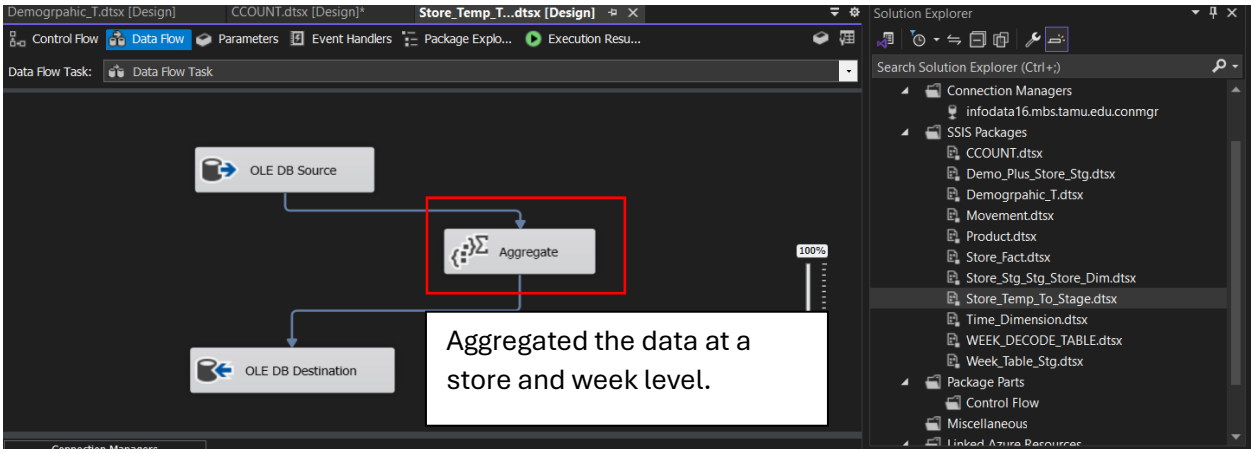
Data Extraction from Ccount.csv

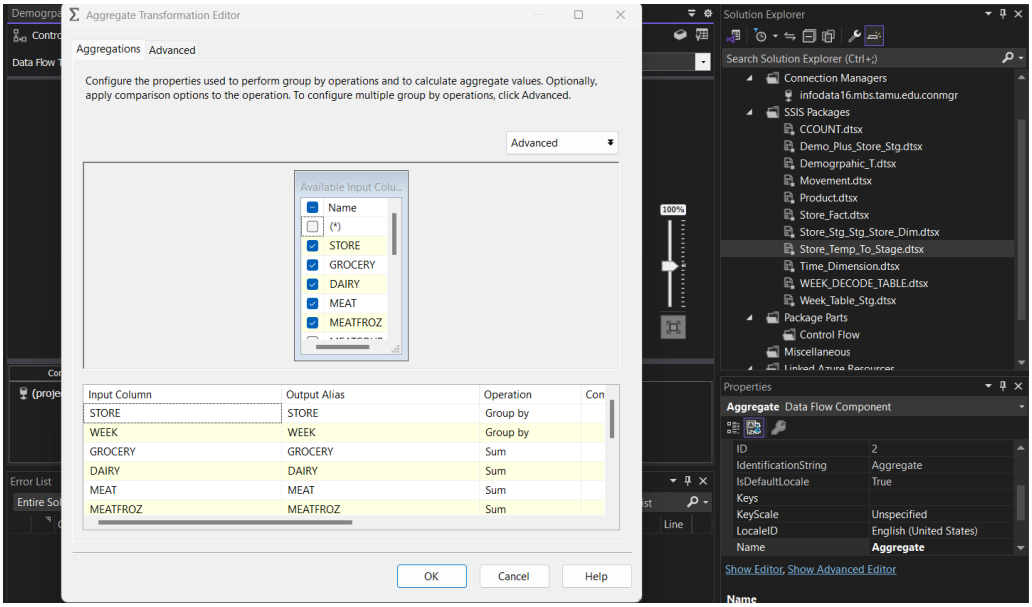




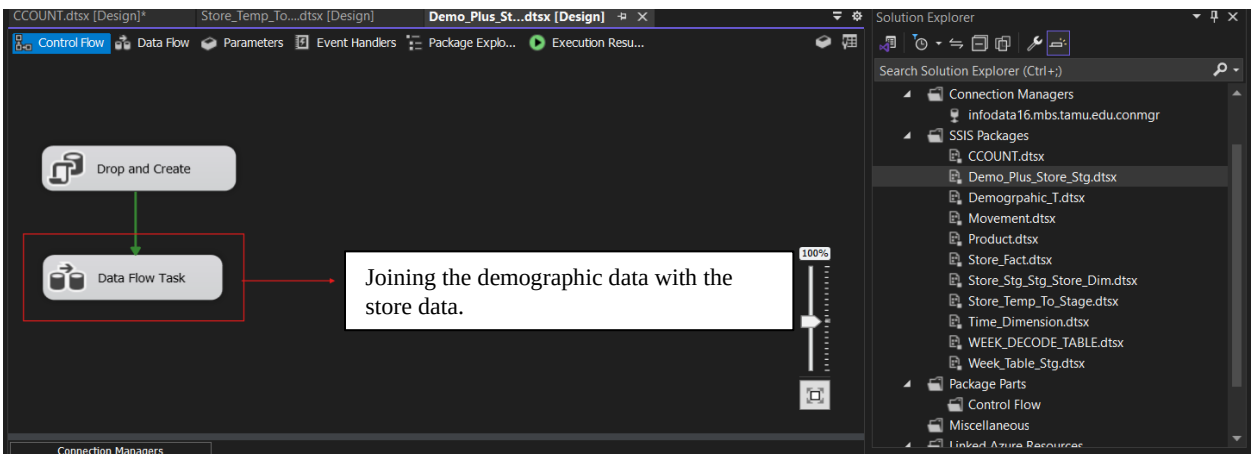
Loading Store Staging Table:

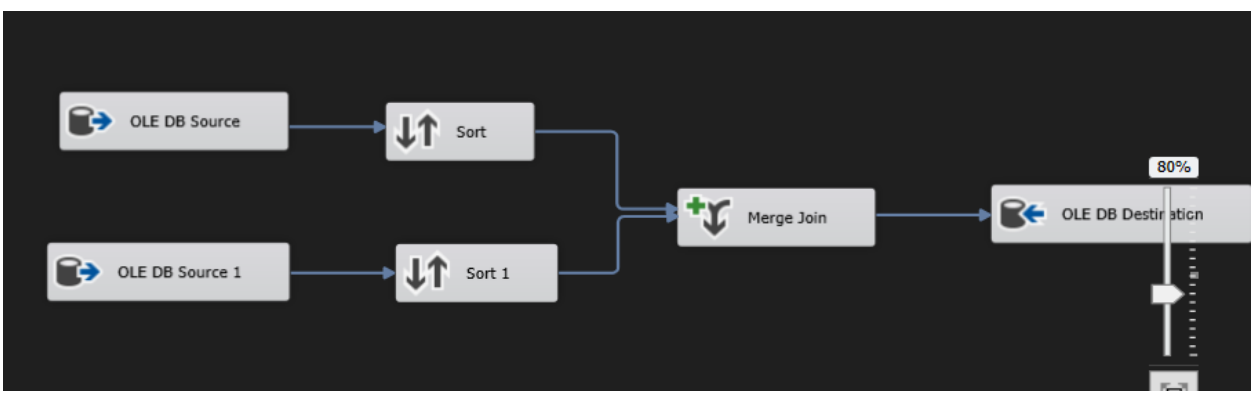

Demogrpahic_T.dtsx [Design]
CCOUNT.dtsx [Design]*
Store_Temp_T...dtsx [Design]
Control Flow
Data Flow
Parameters
Event Handlers
Package Explo...
Execution Resu...
Drop Existing Table
Execute SQL Task 1
Data Flow Task
Dropping any existing tables and then creating the Store_Stg table. Then we go on to add data to the staging tables.
100%
Connection Managers
Solution Explorer
Search Solution Explorer (Ctrl+;)
Connection Managers
infodata16.mbs.tamu.edu.conmgr
SSIS Packages
CCOUNT.dtsx
Demo_Plus_Store_Stg.dtsx
Demogrpahic_T.dtsx
Movement.dtsx
Product.dtsx
Store_Fact.dtsx
Store_Stg_Stg_Store_Dim.dtsx
Store_Temp_To_Stage.dtsx
Time_Dimension.dtsx
WEEK_DECODE_TABLE.dtsx
Week_Table_Stg.dtsx
Package Parts
Control Flow
Miscellaneous

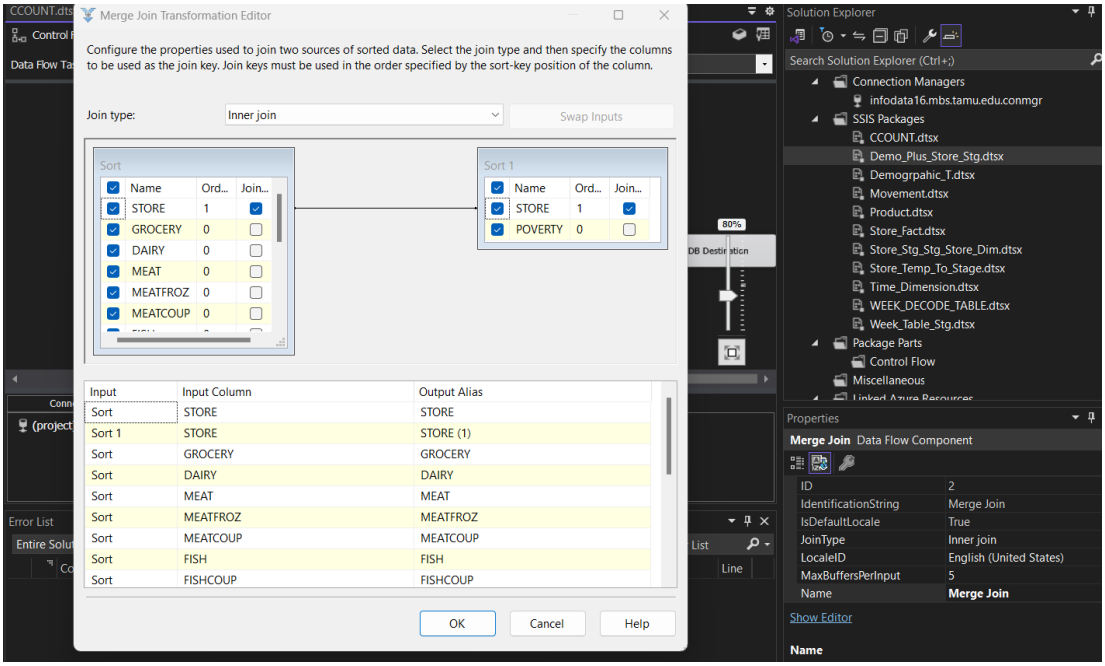

Demogrpahic_T.dtsx [Design]
CCOUNT.dtsx [Design]*
Store_Temp_T...dtsx [Design]
Control Flow
Data Flow
Parameters
Event Handlers
Package Explo...
Execution Resu...
Data Flow Task:
Data Flow Task
OLE DB Source
Aggregate
OLE DB Destination
Aggregated the data at a store and week level.
100%
Solution Explorer
Search Solution Explorer (Ctrl+;)
Connection Managers
infodata16.mbs.tamu.edu.conmgr
SSIS Packages
CCOUNT.dtsx
Demo_Plus_Store_Stg.dtsx
Demogrpahic_T.dtsx
Movement.dtsx
Product.dtsx
Store_Fact.dtsx
Store_Stg_Stg_Store_Dim.dtsx
Store_Temp_To_Stage.dtsx
Time_Dimension.dtsx
WEEK_DECODE_TABLE.dtsx
Week_Table_Stg.dtsx
Package Parts
Control Flow
Miscellaneous



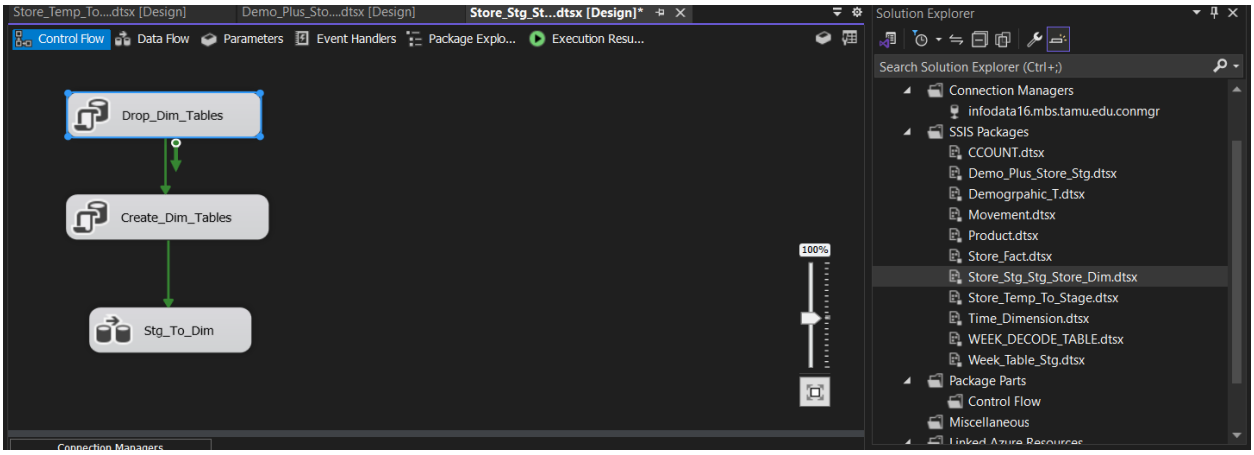
Joining Demographic Data with Store Staging Table:

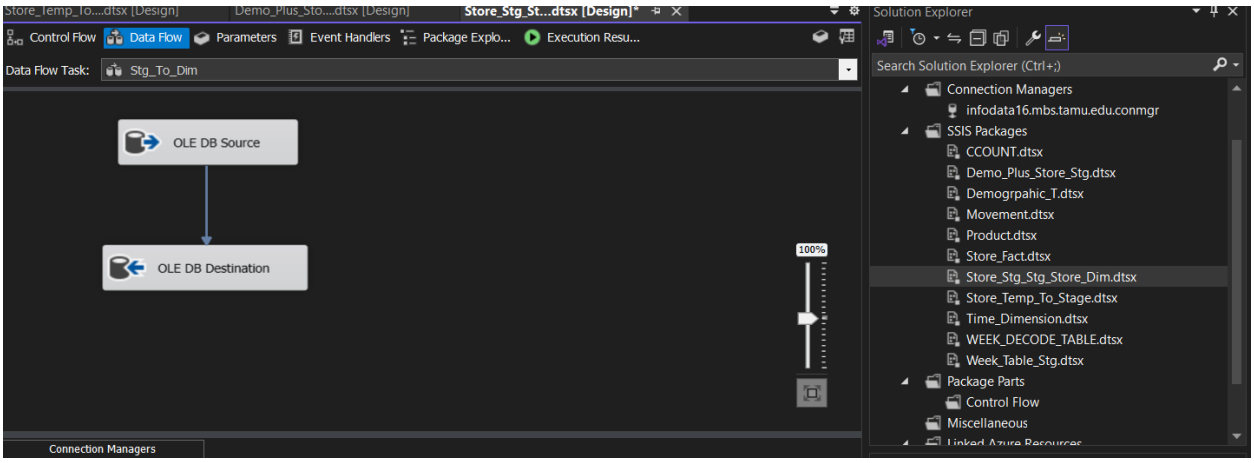


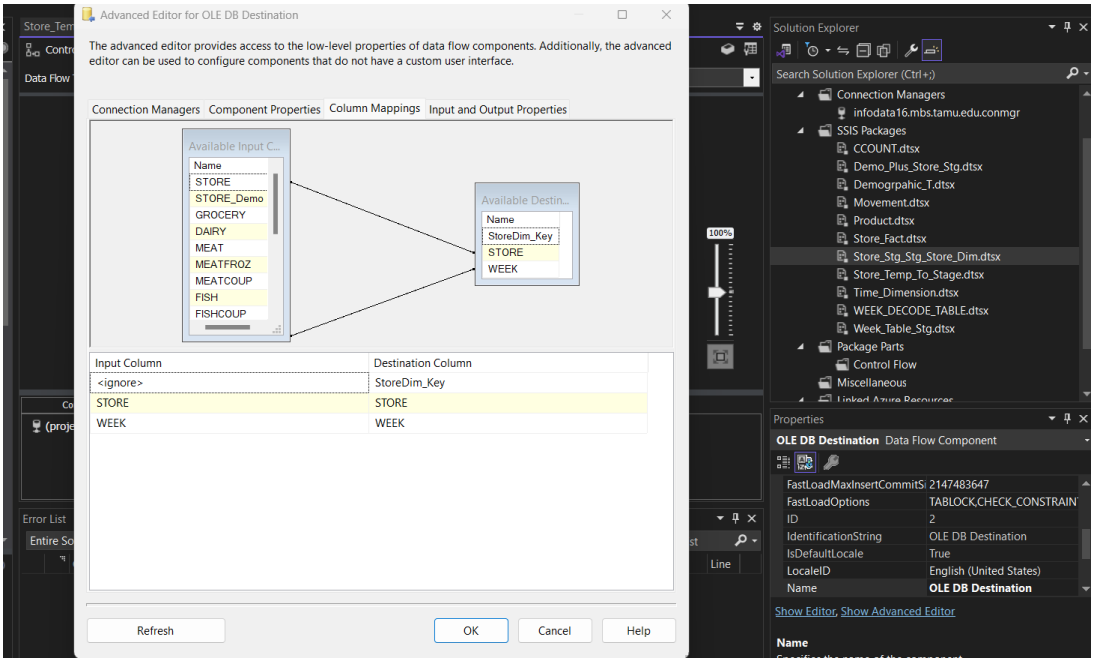




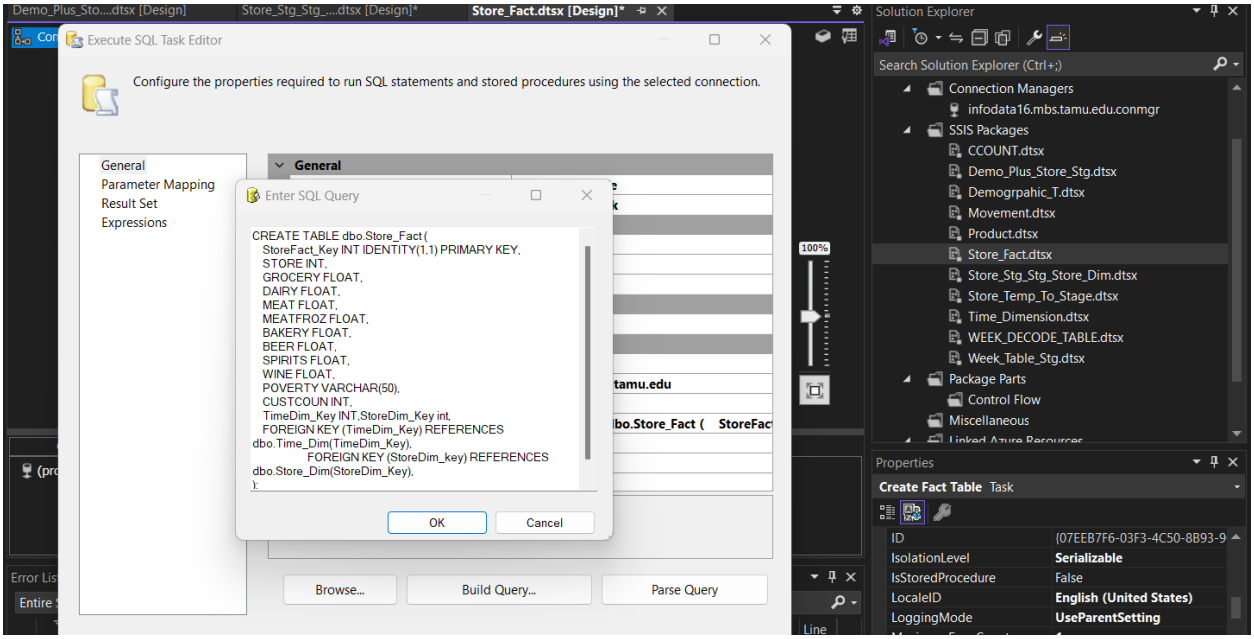
7. Store Dimension Table Creation and Loading

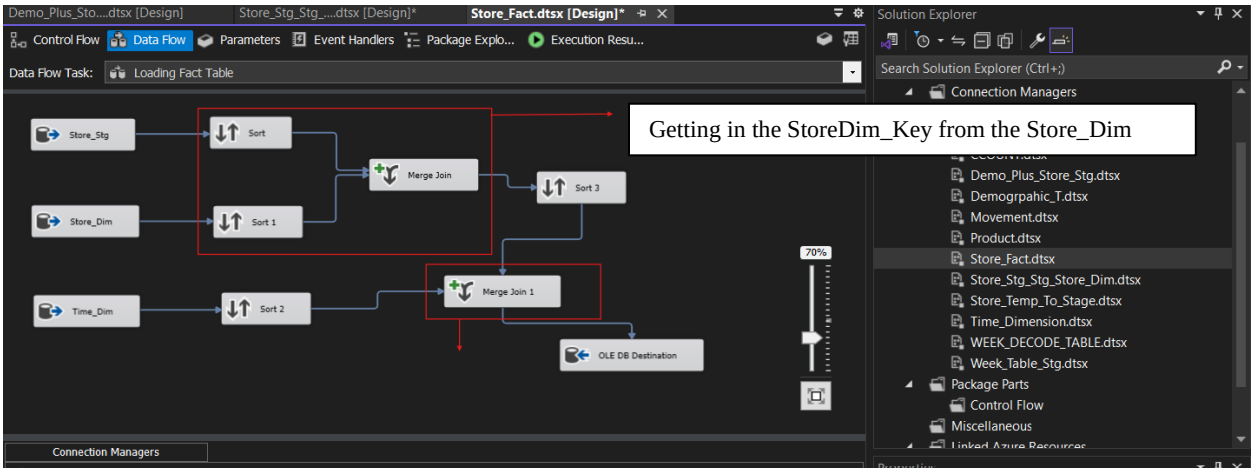




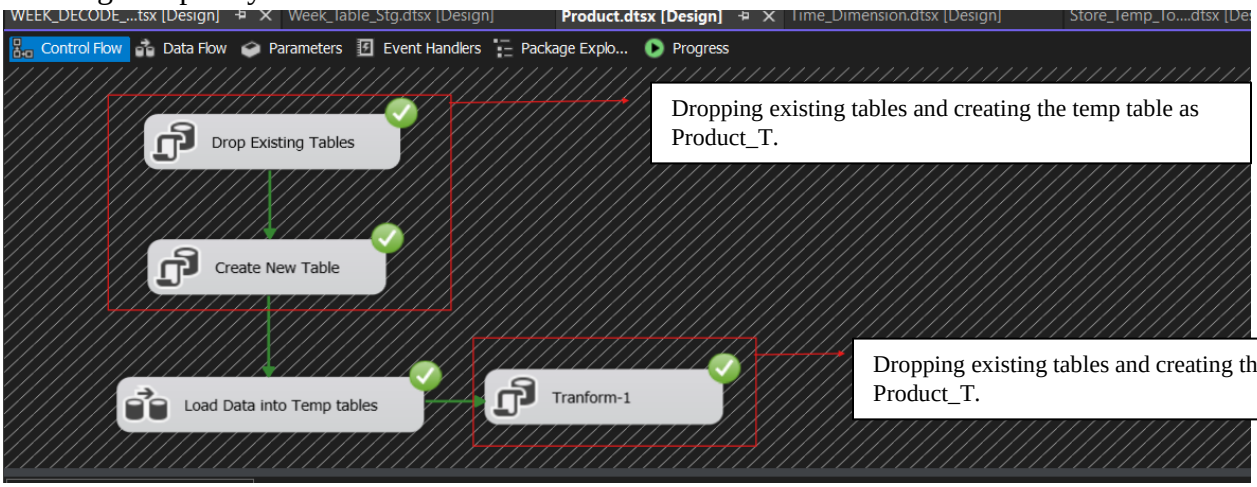


8. Store Fact Table Creation and Loading:

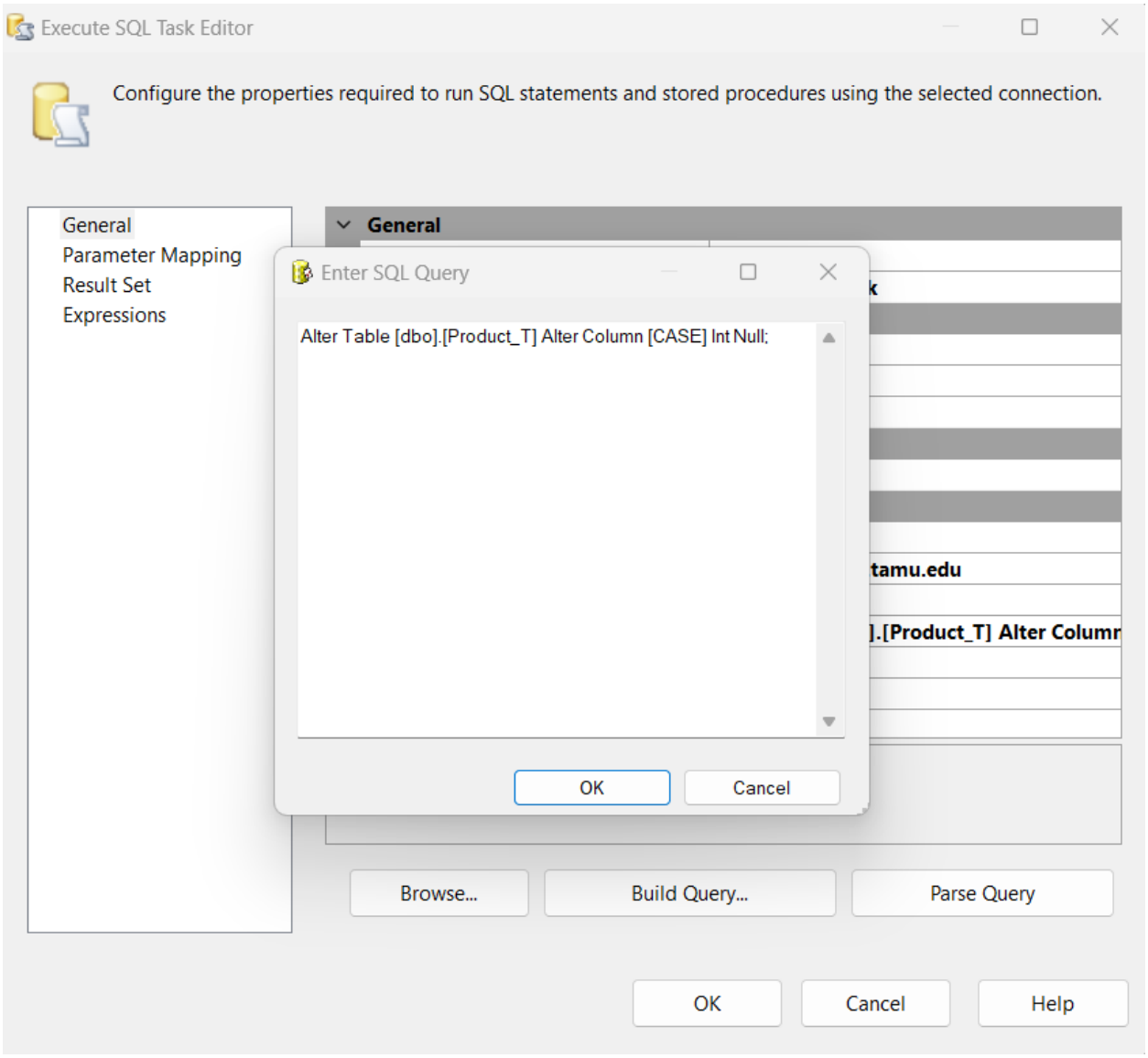


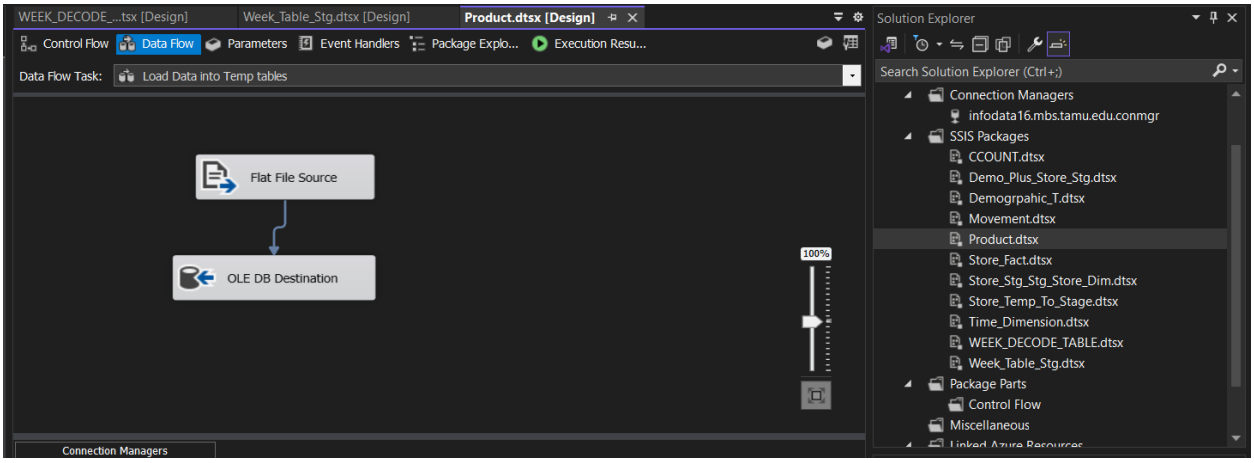


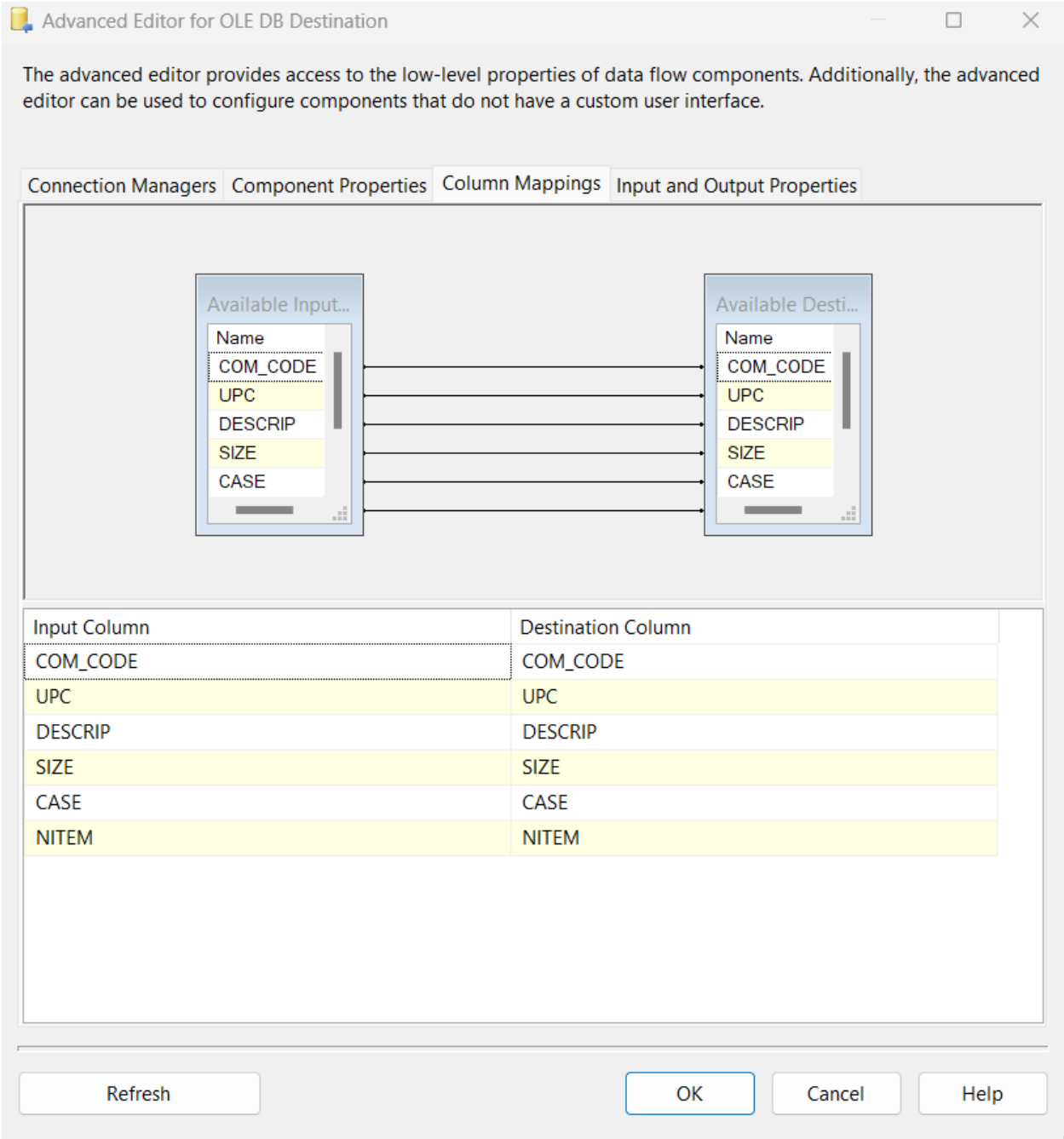
Creating Temporary Product Table



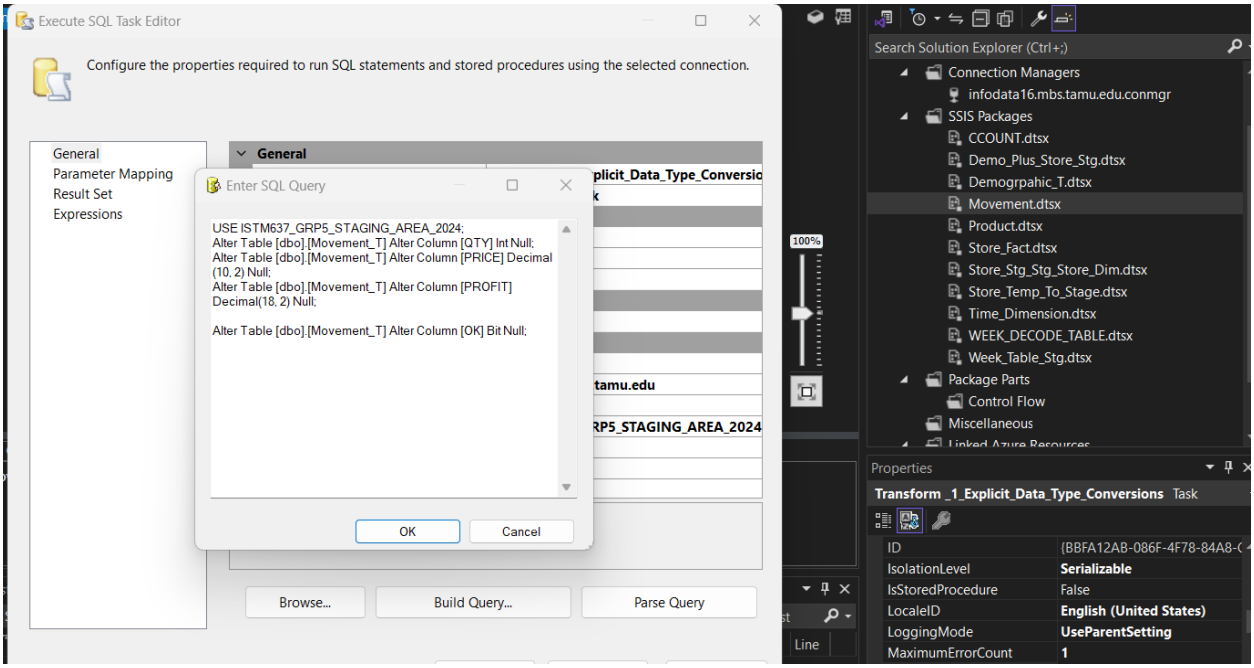
Dropping existing tables and creating the temp table as Product_T.

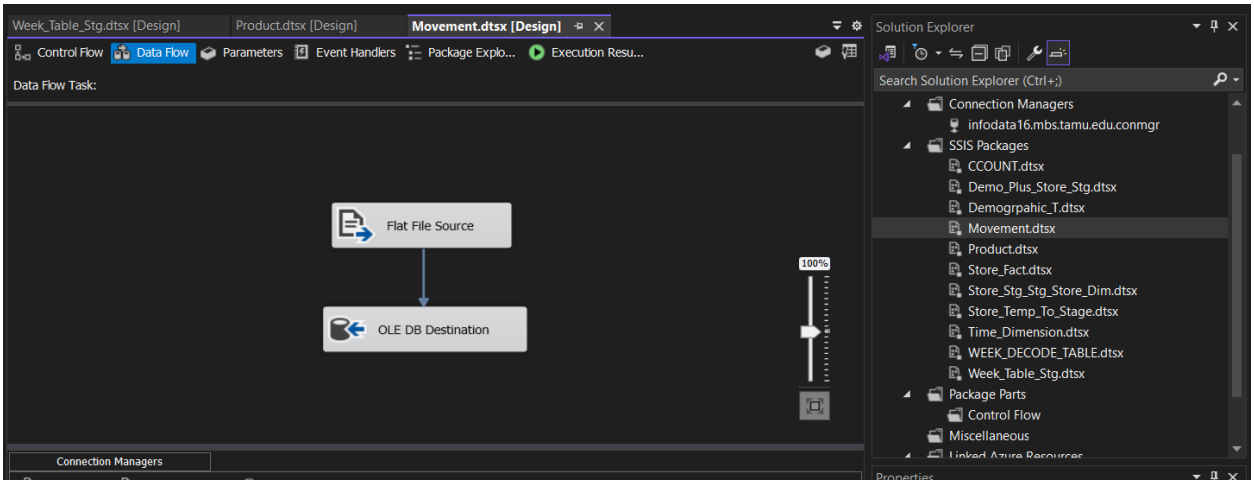

Execute SQL Task Editor
Configure the properties required to run SQL statements and stored procedures using the selected connection.
General
Parameter Mapping
Result Set
Expressions
General
Enter SQL Query
Alter Table [dbo].[Product_T] Alter Column [CASE] Int Null;
OK
Cancel
tamu.edu
].[Product_T] Alter Column
Browse...
Build Query...
Parse Query
OK
Cancel
Help

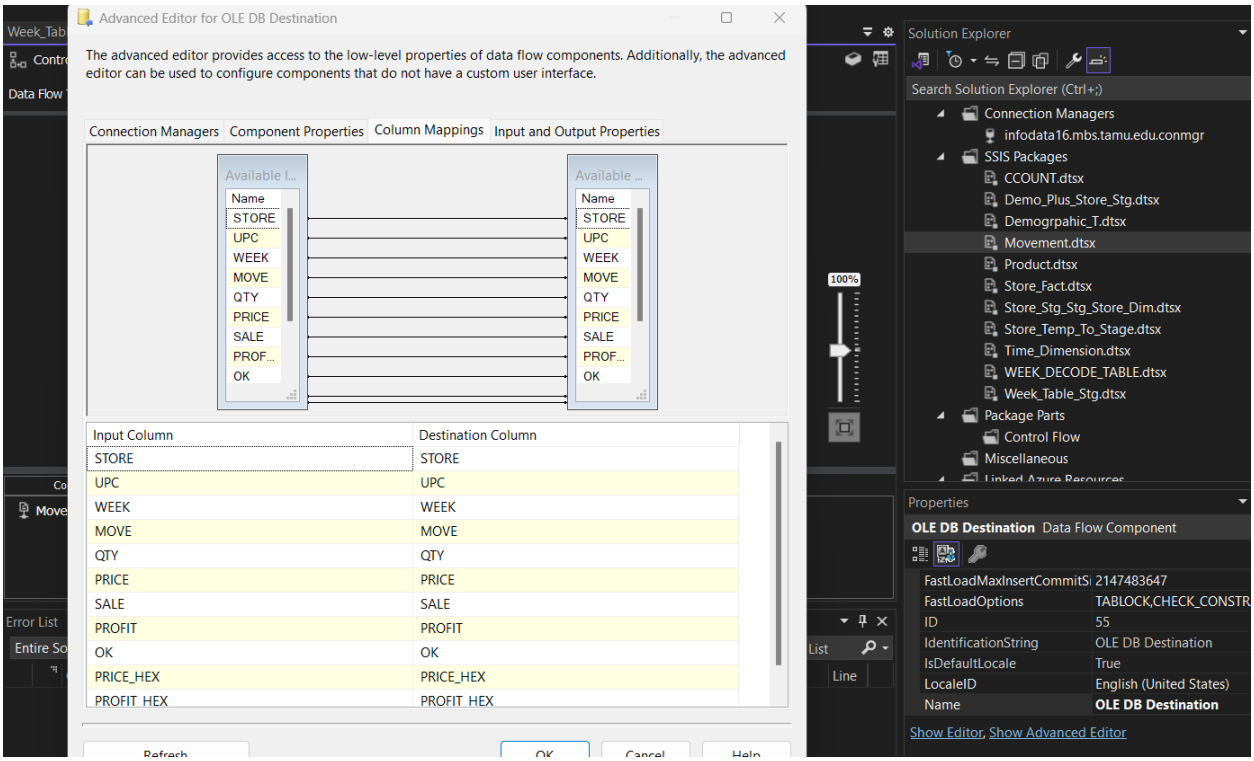

WEEK_DECODE_...tsx [Design]
Week_Table_Stg.dtsx [Design]
Product.dtsx [Design]
Control Flow
Data Flow
Parameters
Event Handlers
Package Explo...
Execution Resu...
Data Flow Task:
Load Data into Temp tables
Flat File Source
OLE DB Destination
100%
Connection Managers
Solution Explorer
Search Solution Explorer (Ctrl+;)
Connection Managers
infodata16.mbs.tamu.edu.conmgr
SSIS Packages
CCOUNT.dtsx
Demo_Plus_Store_Stg.dtsx
Demogrpahic_T.dtsx
Movement.dtsx
Product.dtsx
Store_Fact.dtsx
Store_Stg_Stg_Store_Dim.dtsx
Store_Temp_To_Stage.dtsx
Time_Dimension.dtsx
WEEK_DECODE_TABLE.dtsx
Week_Table_Stg.dtsx
Package Parts
Control Flow
Miscellaneous



Creating Temporary Movement Table

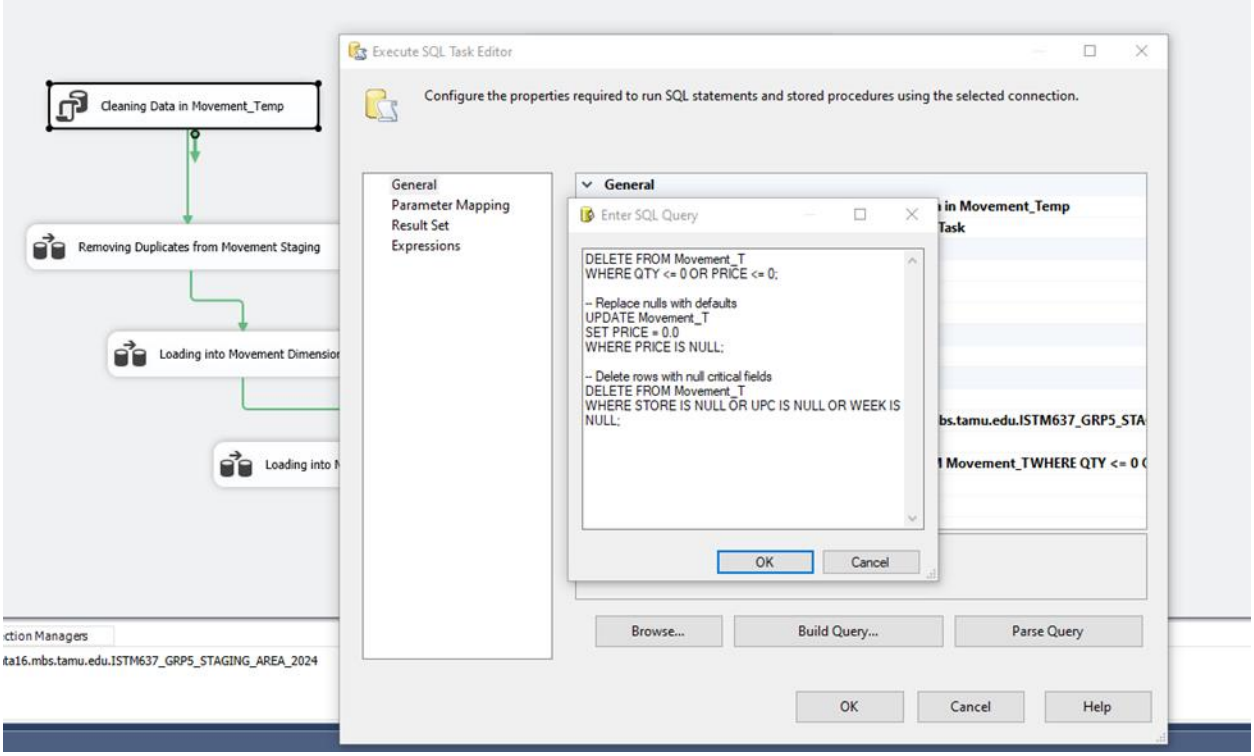

Execute SQL Task Editor
Configure the properties required to run SQL statements and stored procedures using the selected connection.
General
Parameter Mapping
Result Set
Expressions
General
Enter SQL Query
USE ISTM637_GRP5_STAGING_AREA_2024;
Alter Table [dbo].[Movement_T] Alter Column [QTY] Int Null;
Alter Table [dbo].[Movement_T] Alter Column [PRICE] Decimal (10, 2) Null;
Alter Table [dbo].[Movement_T] Alter Column [PROFIT] Decimal(18, 2) Null;
Alter Table [dbo].[Movement_T] Alter Column [OK] Bit Null;
OK
Cancel
Browse...
Build Query...
Parse Query
Search Solution Explorer (Ctrl+;)
Connection Managers
infodata16.mbs.tamu.edu.conmgr
SSIS Packages
CCOUNT.dtsx
Demo_Plus_Store_Stg.dtsx
Demogrpahic_T.dtsx
Movement.dtsx
Product.dtsx
Store_Fact.dtsx
Store_Stg_Stg_Store_Dim.dtsx
Store_Temp_To_Stage.dtsx
Time_Dimension.dtsx
WEEK_DECODE_TABLE.dtsx
Week_Table_Stg.dtsx
Package Parts
Control Flow
Miscellaneous
Properties
Transform _1_Explicit_Data_Type_Conversions Task
ID
IsolationLevel Serializable
IsStoredProcedure False
LocaleID English (United States)
LoggingMode UseParentSetting
MaximumErrorCount 1

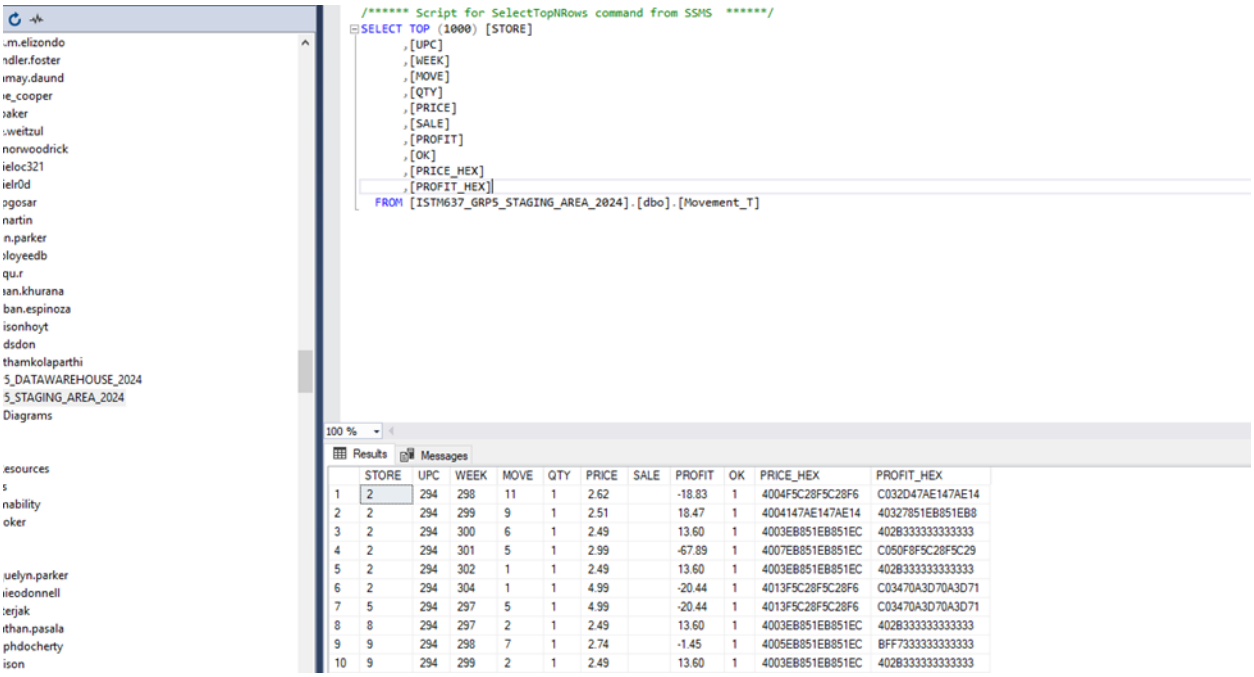

Week_Table_Stg.dtsx [Design]
Product.dtsx [Design]
Movement.dtsx [Design]
Control Flow
Data Flow
Parameters
Event Handlers
Package Explo...
Execution Resu...
Data Flow Task:
Flat File Source
OLE DB Destination
Connection Managers
Solution Explorer
Search Solution Explorer (Ctrl+;)
Connection Managers
infodata16.mbs.tamu.edu.conmgr
SSIS Packages
CCOUNT.dtsx
Demo_Plus_Store_Stg.dtsx
Demogrpahic_T.dtsx
Movement.dtsx
Product.dtsx
Store_Fact.dtsx
Store_Stg_Stg_Store_Dim.dtsx
Store_Temp_To_Stage.dtsx
Time_Dimension.dtsx
WEEK_DECODE_TABLE.dtsx
Week_Table_Stg.dtsx
Package Parts
Control Flow
Miscellaneous



9. Movement Dimension and Fact Table Workflow:

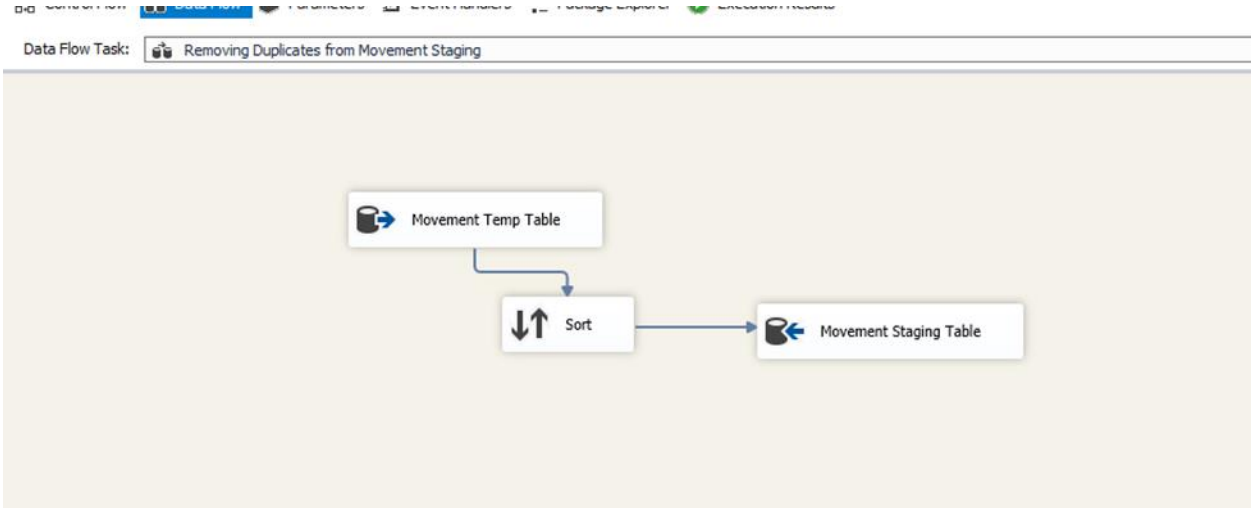
Cleaning Data in Movement Temp Table:

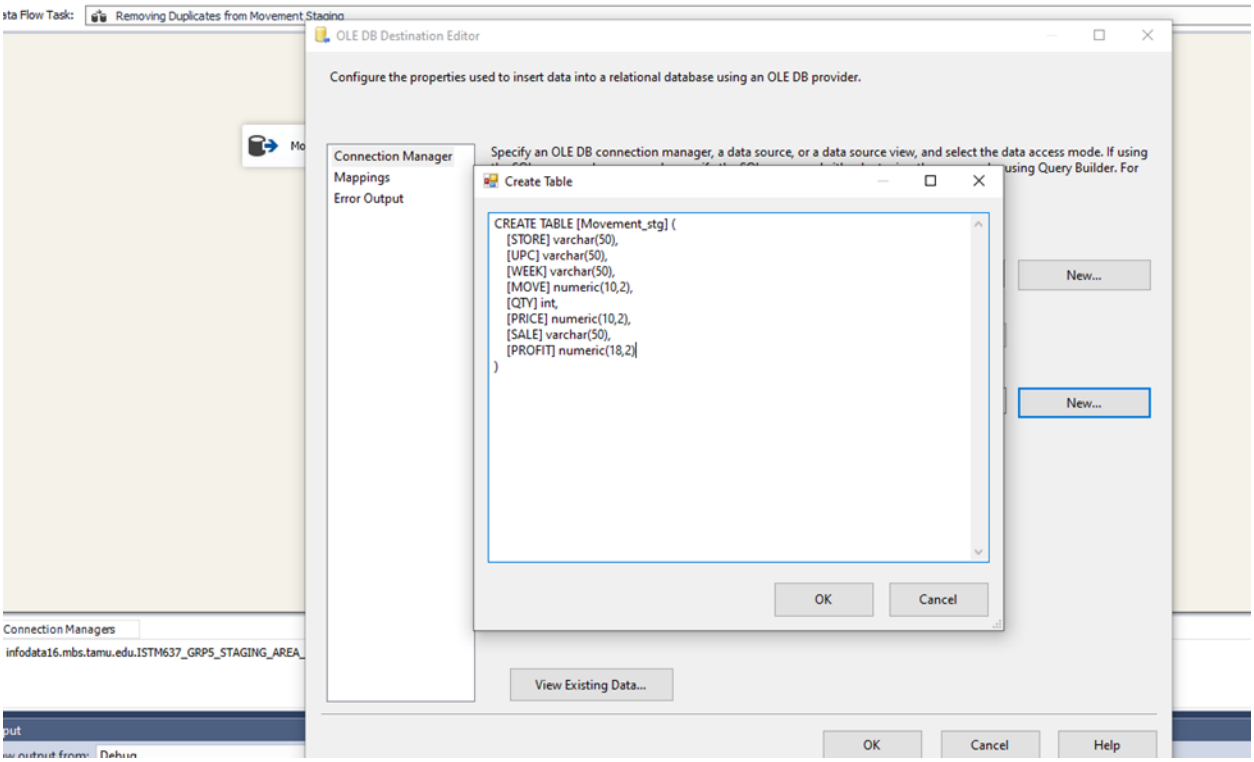


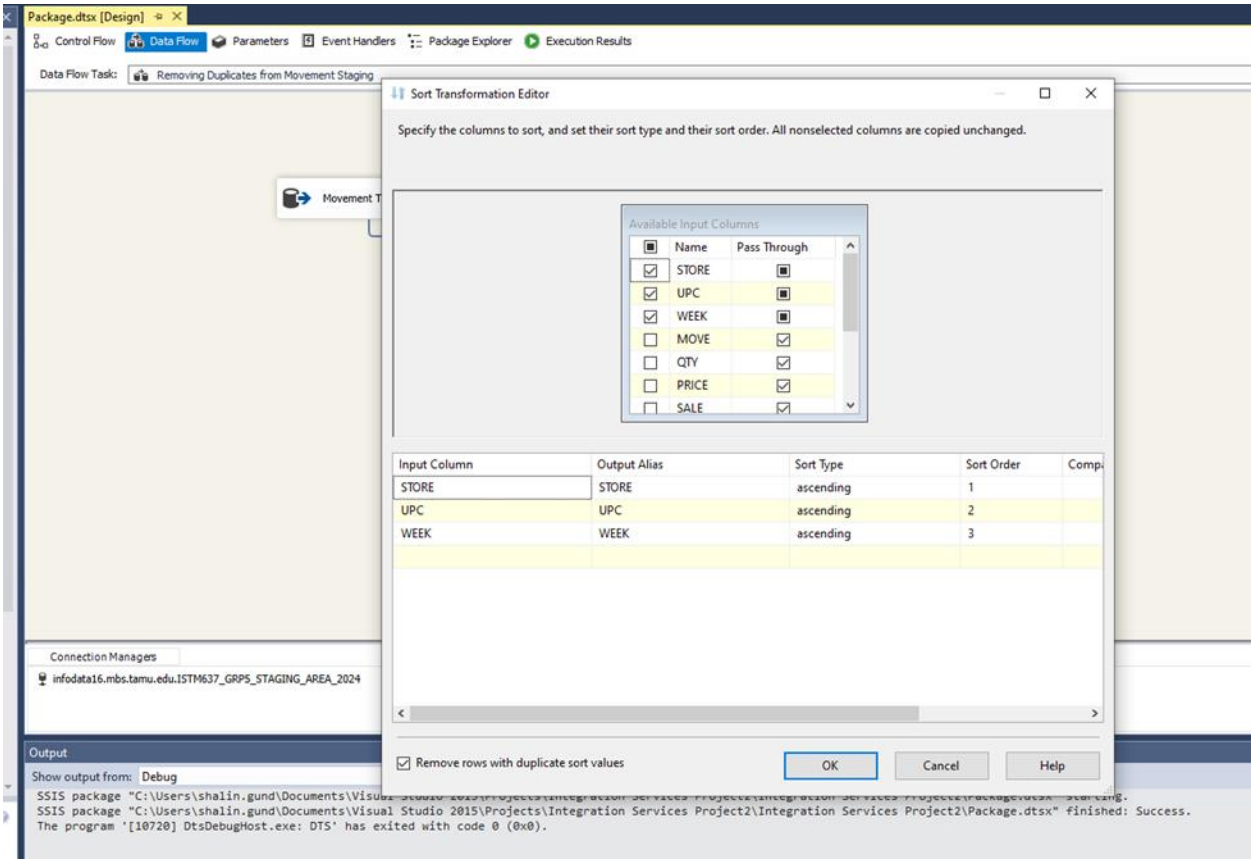


	STORE	UPC	WEEK	MOVE	QTY	PRICE	SALE	PROFIT	OK	PRICE_HEX	PROFIT_HEX
1	2	294	298	11	1	2.62		-18.83	1	4004F5C28F5C28F6	C032D47AE147AE14
2	2	294	299	9	1	2.51		18.47	1	4004147AE147AE14	40327851EB851EB8
3	2	294	300	6	1	2.49		13.60	1	4003EB851EB851EC	402B333333333333
4	2	294	301	5	1	2.99		-67.89	1	4007EB851EB851EC	C050F8F5C28F5C29
5	2	294	302	1	1	2.49		13.60	1	4003EB851EB851EC	402B333333333333
6	2	294	304	1	1	4.99		-20.44	1	4013F5C28F5C28F6	C03470A3D70A3D71
7	5	294	297	5	1	4.99		-20.44	1	4013F5C28F5C28F6	C03470A3D70A3D71
8	8	294	297	2	1	2.49		13.60	1	4003EB851EB851EC	402B333333333333
9	9	294	298	7	1	2.74		-1.45	1	4005EB851EB851EC	BFF7333333333333
10	9	294	299	2	1	2.49		13.60	1	4003EB851EB851EC	402B333333333333

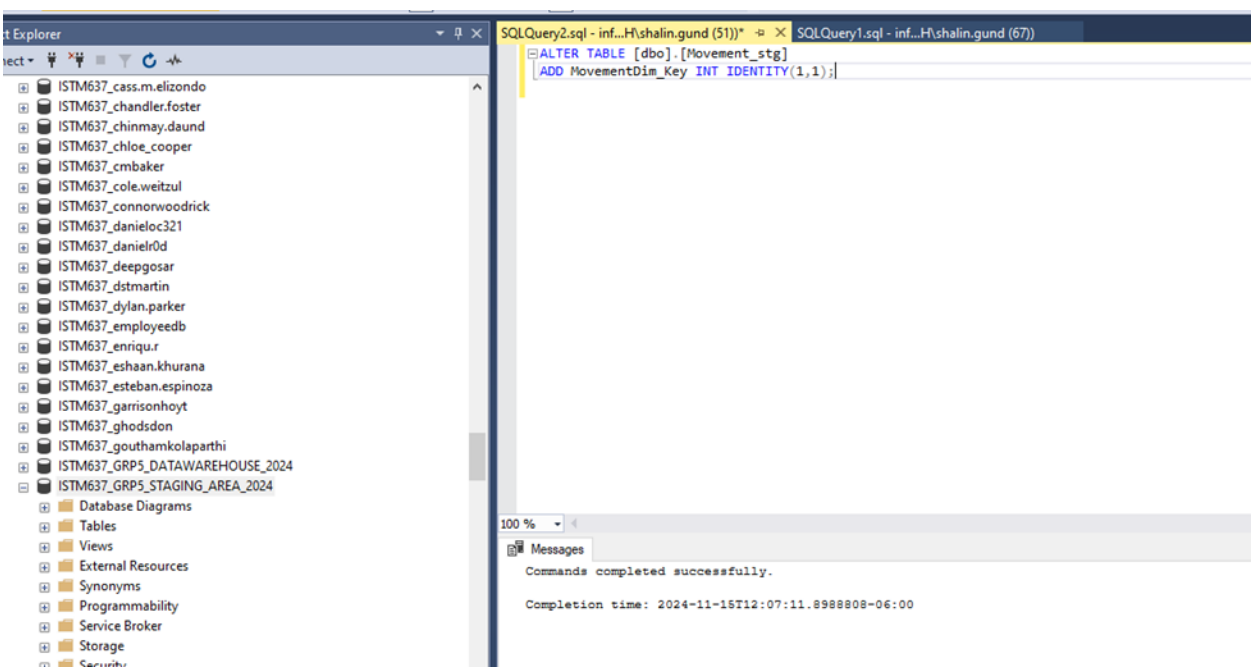
Removing duplicates from Movement Staging and Loading the data in the Movement Staging Table


Data Flow Task:
Removing Duplicates from Movement Staging
Movement Temp Table
Sort
Movement Staging Table

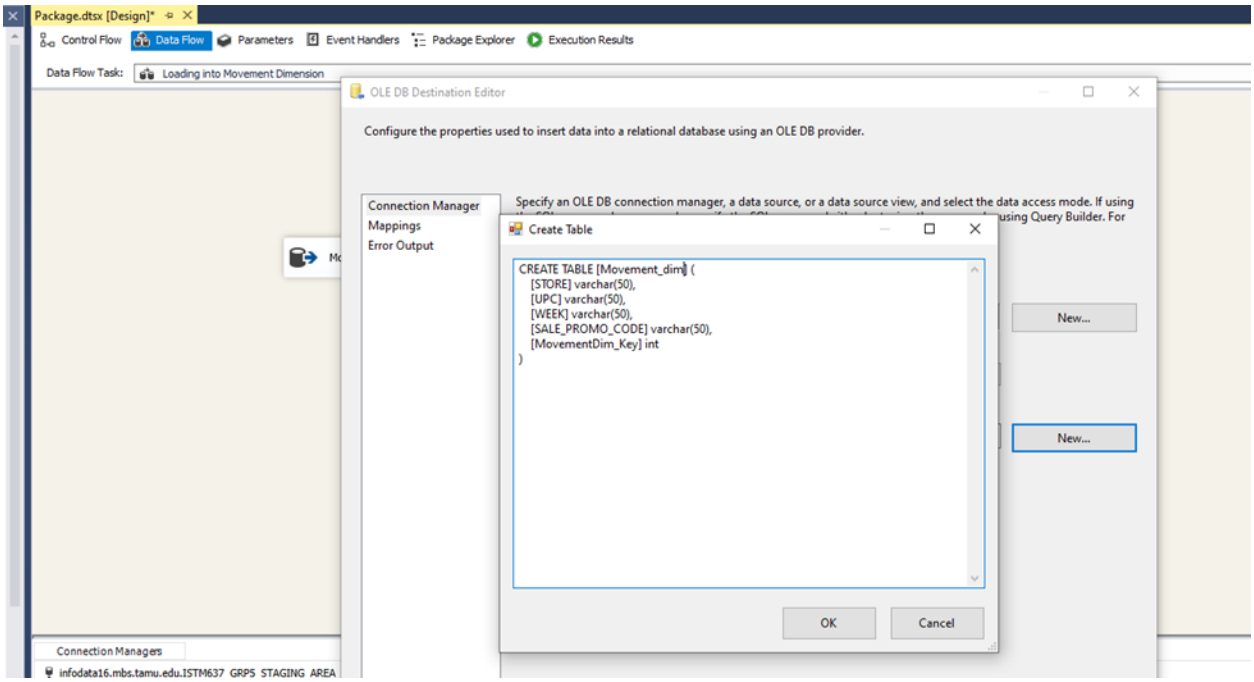

Removing Duplicates from Movement Staging
OLE DB Destination Editor
Configure the properties used to insert data into a relational database using an OLE DB provider.
Connection Manager
Mappings
Error Output
Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using
using Query Builder. For
Create Table
CREATE TABLE [Movement_stg] (
[STORE] varchar(50),
[UPC] varchar(50),
[WEEK] varchar(50),
[MOVE] numeric(10,2),
[QTY] int,
[PRICE] numeric(10,2),
[SALE] varchar(50),
[PROFIT] numeric(18,2)
)
OK
Cancel
New...
New...
View Existing Data...
OK
Cancel
Help
Connection Managers
infodata16.mbs.tamu.edu.ISTM637_GRPS_STAGING_AREA_

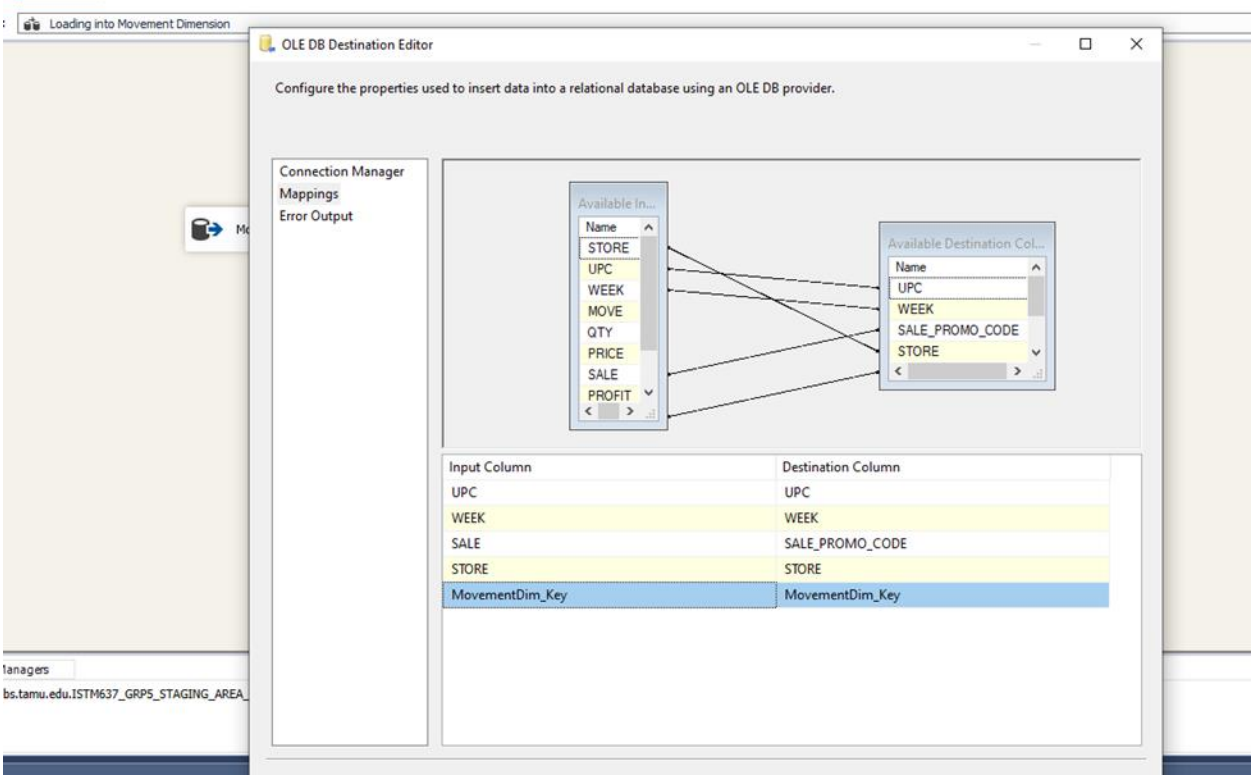


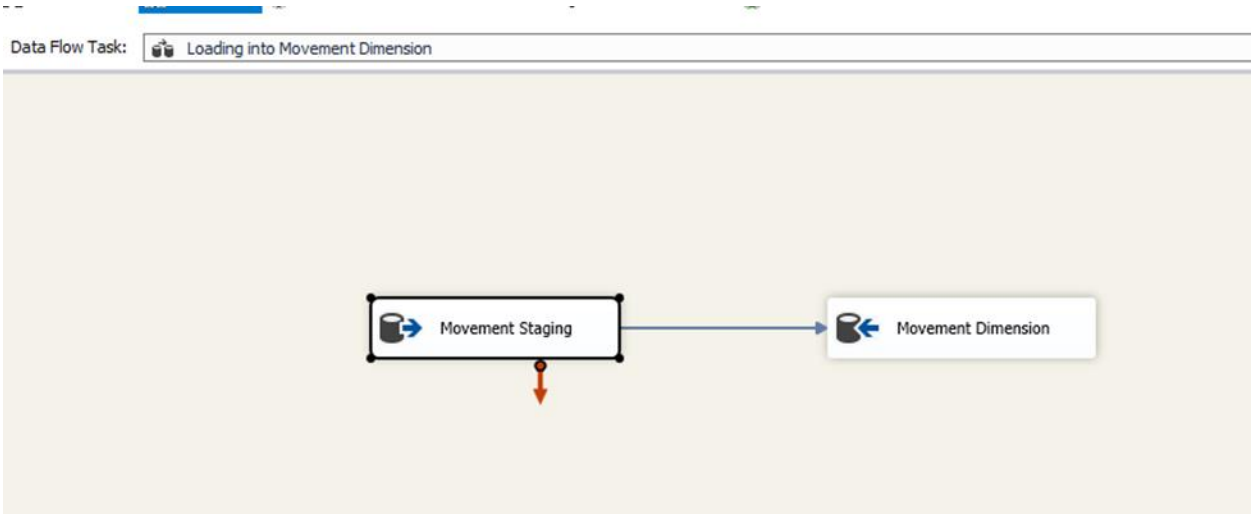
Adding MovementDim_Key (surrogate key):



Loading into Movement dimension Table:

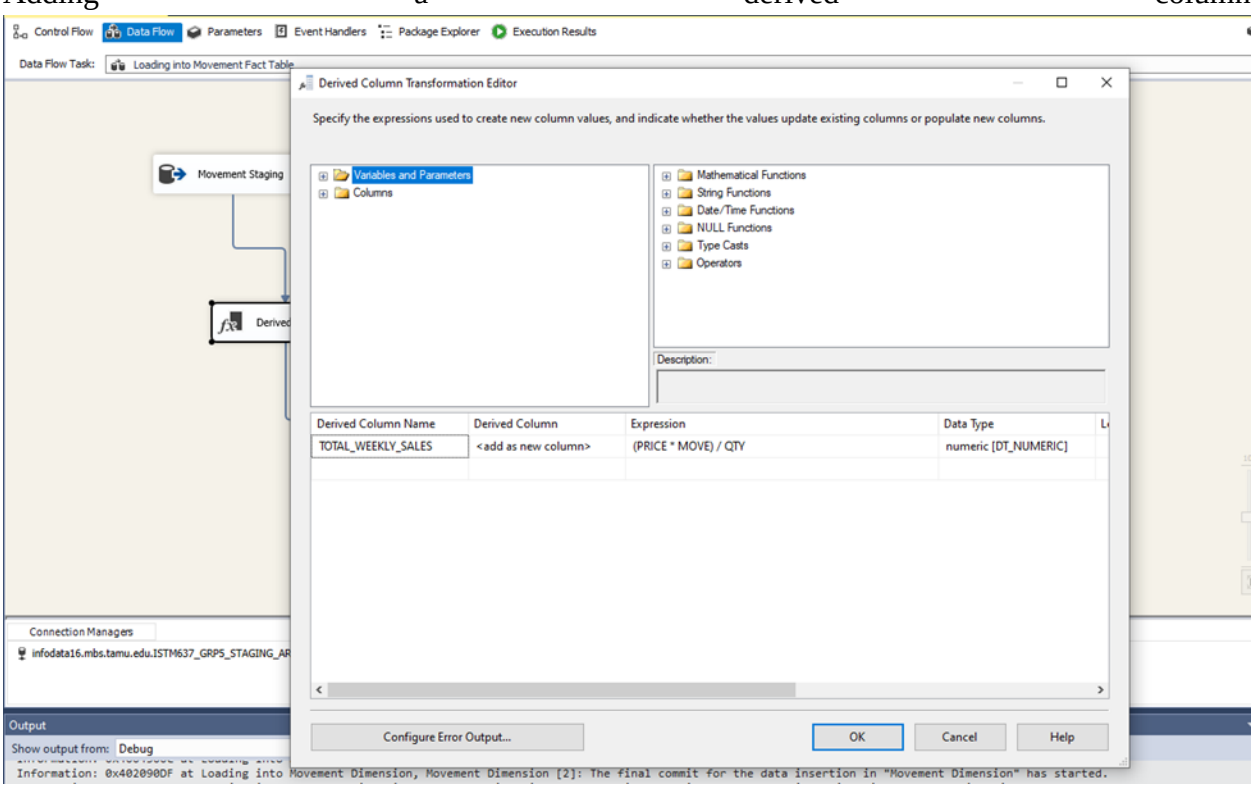




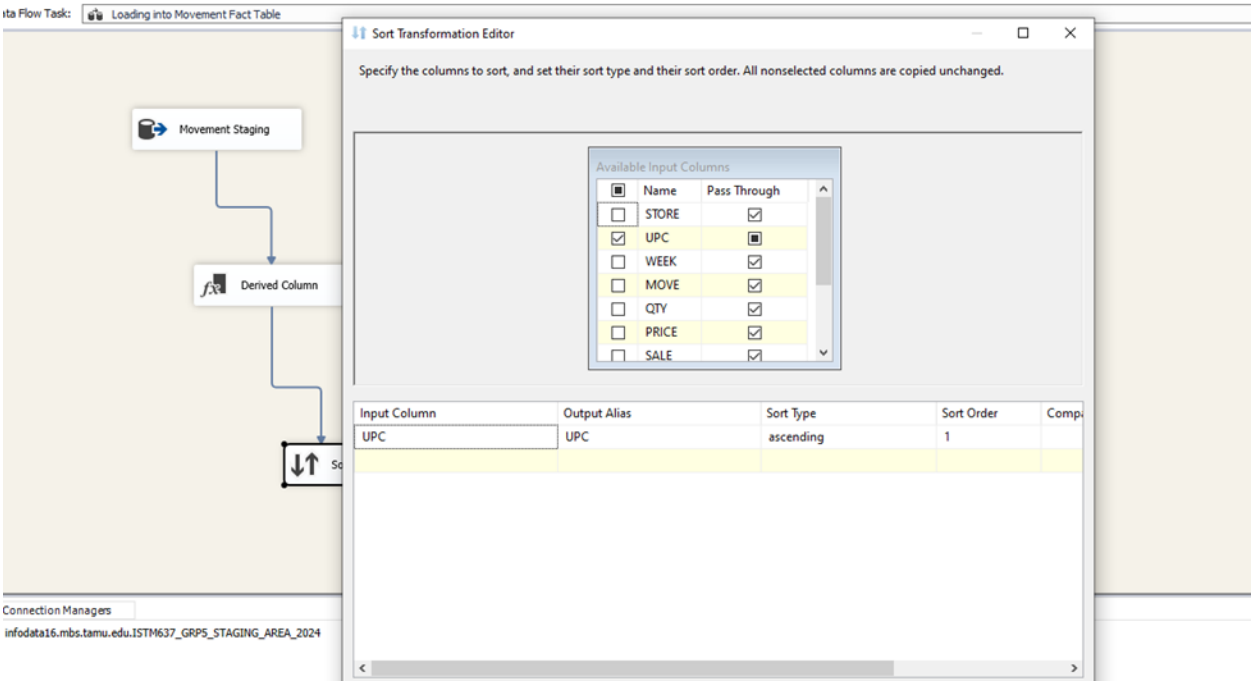


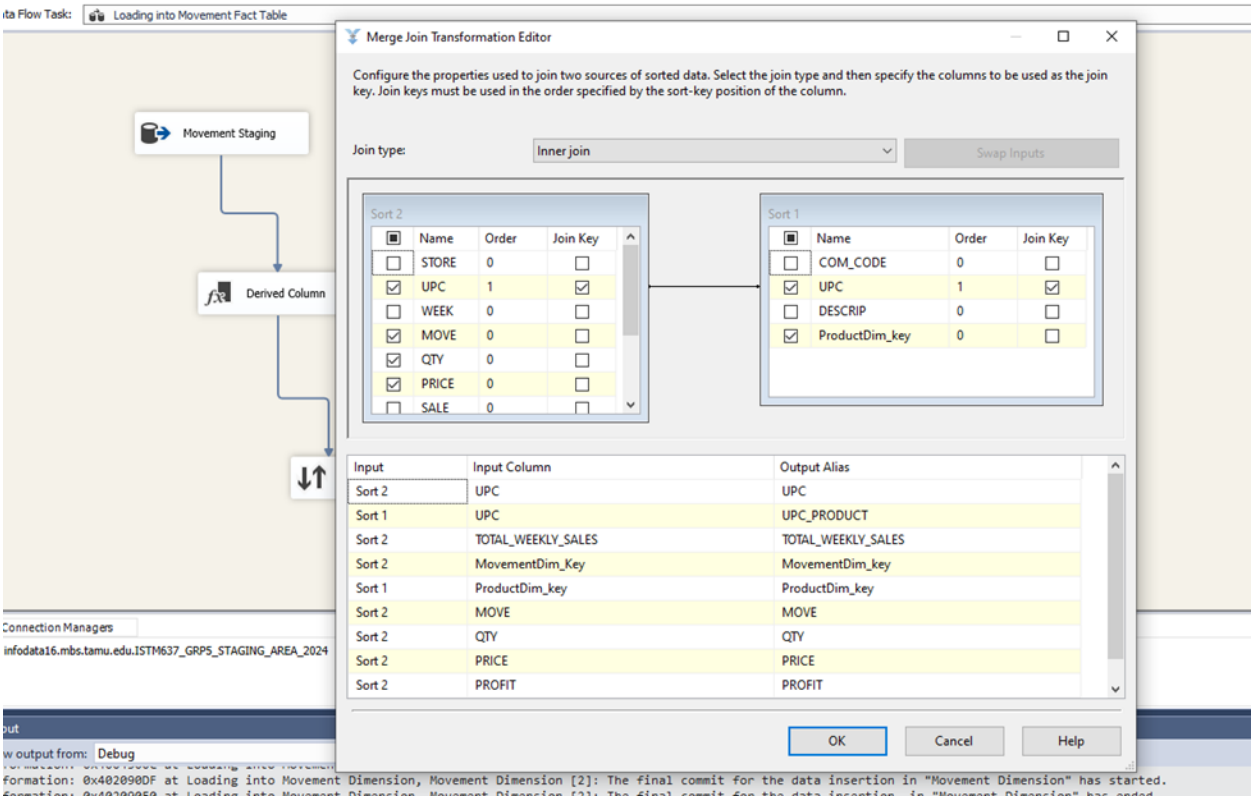
Loading into Movement Fact Table:

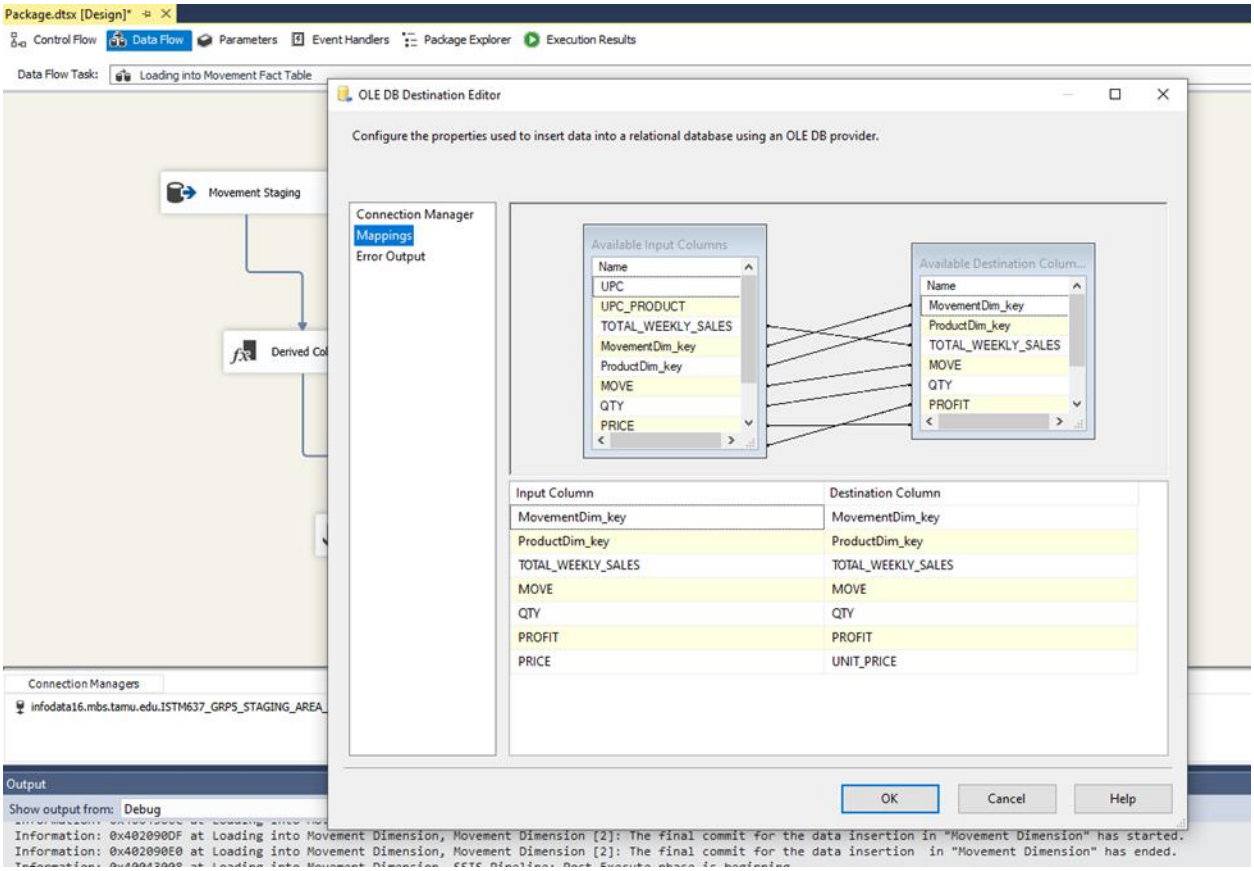
Adding a derived column

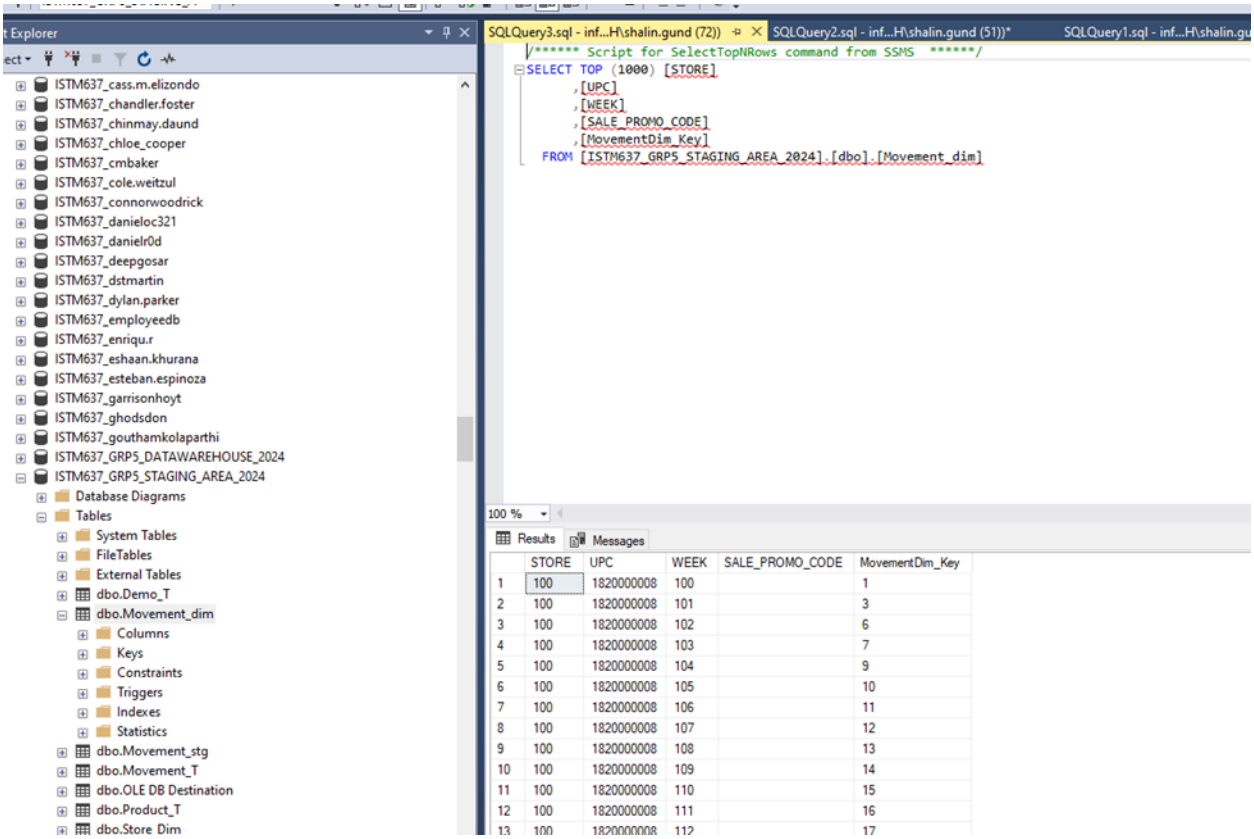


Sorting data using UPC






Package.dtsx [Design]*
Control Flow
Data Flow
Parameters
Event Handlers
Package Explorer
Execution Results
Data Flow Task:
Loading into Movement Fact Table
Movement Staging
Derived Col
Connection Managers
infodata16.mbs.tamu.edu.ISTM637_GRP5_STAGING_AREA_
Output
Show output from:
Debug
Information: 0x402090DF at Loading into Movement Dimension, Movement Dimension [2]: The final commit for the data insertion in "Movement Dimension" has started.
Information: 0x402090E0 at Loading into Movement Dimension, Movement Dimension [2]: The final commit for the data insertion in "Movement Dimension" has ended.
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
Available Input Columns
Name
UPC
UPC_PRODUCT
TOTAL_WEEKLY_SALES
MovementDim_key
ProductDim_key
MOVE
QTY
PRICE
Available Destination Colum...
Name
MovementDim_key
ProductDim_key
TOTAL_WEEKLY_SALES
MOVE
QTY
PROFIT
Input Column
Destination Column
MovementDim_key
MovementDim_key
ProductDim_key
ProductDim_key
TOTAL_WEEKLY_SALES
TOTAL_WEEKLY_SALES
MOVE
MOVE
QTY
QTY
PROFIT
PROFIT
PRICE
UNIT_PRICE
OK
Cancel
Help


Explorer
ISTM637_cass.m.elizondo
ISTM637_chandler.foster
ISTM637_chinmay.daund
ISTM637_chloe_cooper
ISTM637_cmbaker
ISTM637_cole.weitzul
ISTM637_connorwoodrick
ISTM637_danieloc321
ISTM637_danielr0d
ISTM637_deepgosar
ISTM637_dstmartin
ISTM637_dylan.parker
ISTM637_employeedb
ISTM637_enriqu.r
ISTM637_eshaan.khurana
ISTM637_esteban.espinoza
ISTM637_garrisonhoyt
ISTM637_ghodsdon
ISTM637_gouthamkolaparthi
ISTM637_GRP5_DATAWAREHOUSE_2024
ISTM637_GRP5_STAGING_AREA_2024
Database Diagrams
Tables
System Tables
FileTables
External Tables
dbo.Demo_T
dbo.Movement_dim
Columns
Keys
Constraints
Triggers
Indexes
Statistics
dbo.Movement_stg
dbo.Movement_T
dbo.OLE DB Destination
dbo.Product_T
dbo.Store_Dim
SQLQuery3.sql - inf...H\shalin.gund (72))
SQLQuery2.sql - inf...H\shalin.gund (51))*
SQLQuery1.sql - inf...H\shalin.gu
/****** Script for SelectTopNRows command from SSMS ******/
SELECT TOP (1000) [STORE]
,[UPC]
,[WEEK]
,[SALE_PROMO_CODE]
,[MovementDim_Key]
FROM [ISTM637_GRP5_STAGING_AREA_2024].[dbo].[Movement_dim]
100 %
Results
Messages
STORE UPC WEEK SALE_PROMO_CODE MovementDim_Key
1 100 1820000008 100 1
2 100 1820000008 101 3
3 100 1820000008 102 6
4 100 1820000008 103 7
5 100 1820000008 104 9
6 100 1820000008 105 10
7 100 1820000008 106 11
8 100 1820000008 107 12
9 100 1820000008 108 13
10 100 1820000008 109 14
11 100 1820000008 110 15
12 100 1820000008 111 16
13 100 1820000008 112 17

SQLQuery2.sql - inf...H\shalin.gund (66)) | SQLQuery1.sql - inf...H\shalin.gund (68))*

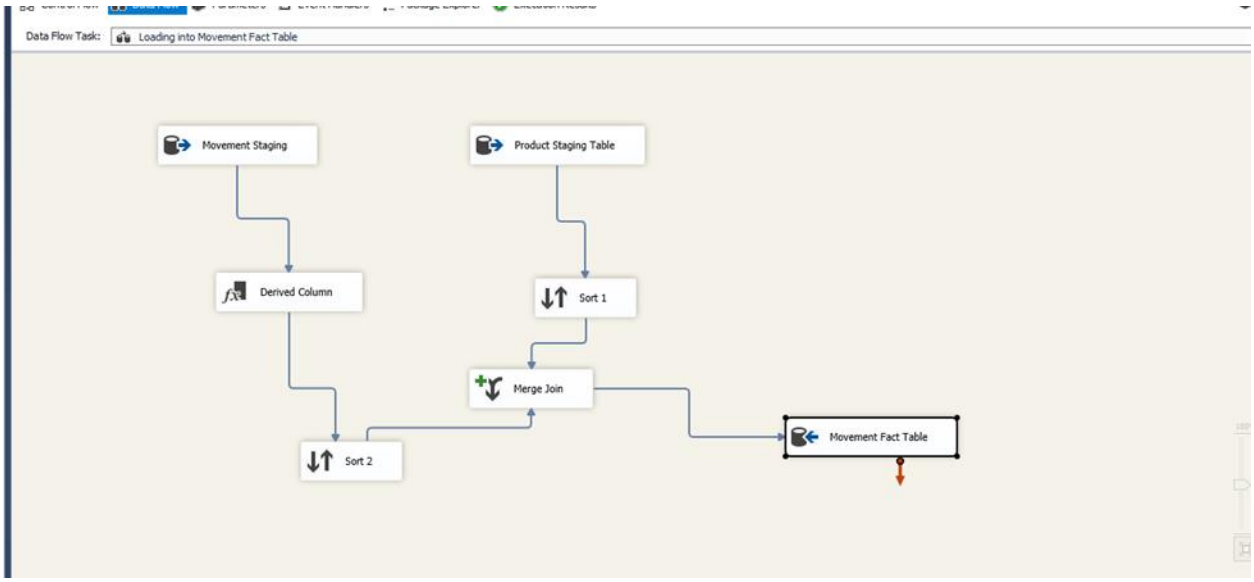
```sql
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [TOTAL_WEEKLY_SALES]
      ,[MovementDim_key]
      ,[ProductDim_key]
      ,[MOVE]
      ,[QTY]
      ,[UNIT_PRICE]
      ,[PROFIT]
  FROM [ISTM637_GRP5_STAGING_AREA_2024].[dbo].[Movement_fact]
```

100 %

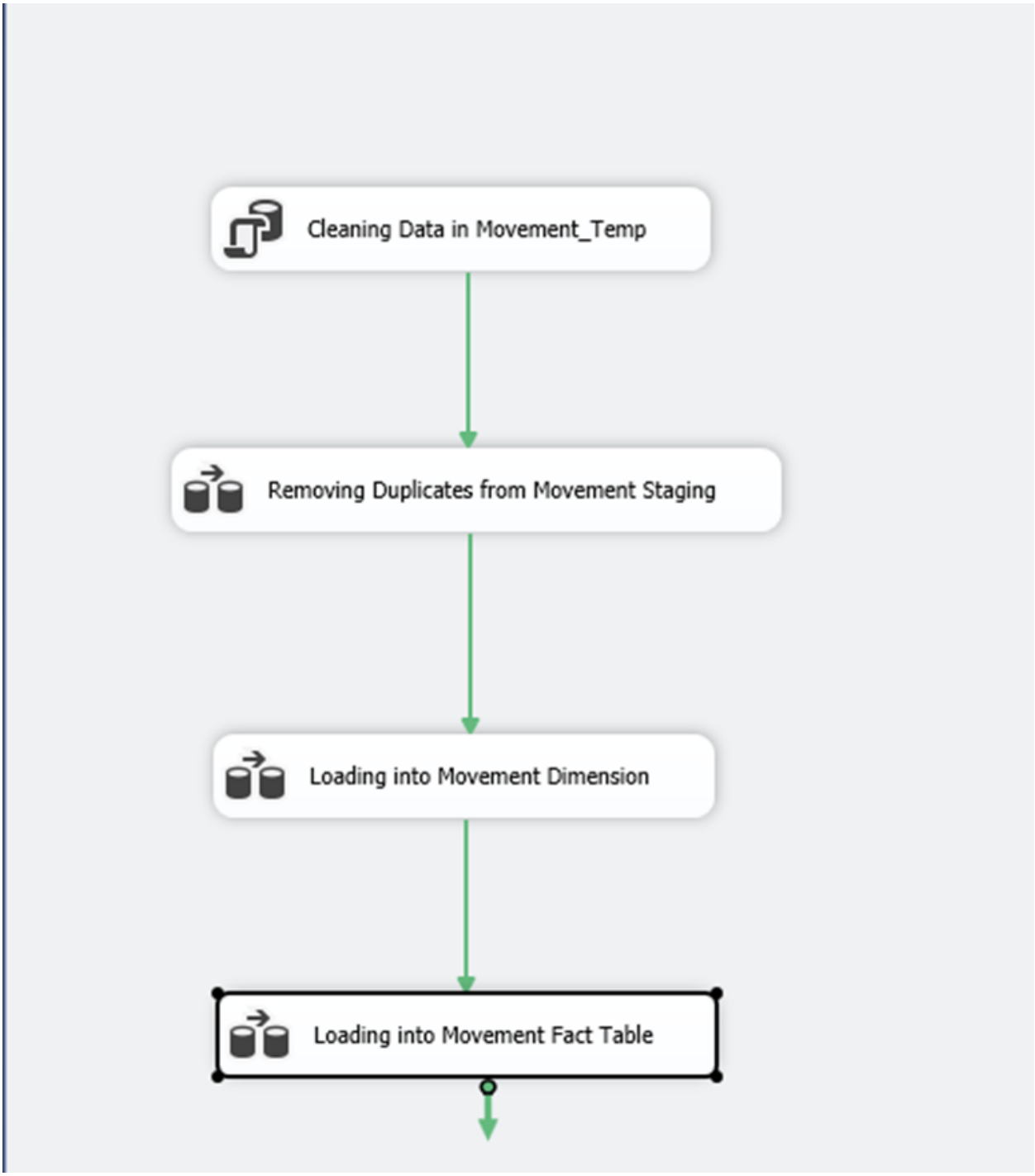
Results | Messages

	TOTAL_WEEKLY_SALES	MovementDim_key	ProductDim_key	MOVE	QTY	UNIT_PRICE	PROFIT
1	4.990000000000000	630849	1	1.00	1	4.99	30.12
2	4.990000000000000	938697	1	1.00	1	4.99	30.12
3	19.960000000000000	630759	1	4.00	1	4.99	30.12
4	4.990000000000000	938772	1	1.00	1	4.99	30.12
5	14.970000000000000	630744	1	3.00	1	4.99	30.12
6	4.990000000000000	938760	1	1.00	1	4.99	30.12
7	19.960000000000000	630738	1	4.00	1	4.99	30.12
8	4.990000000000000	938767	1	1.00	1	4.99	30.12
9	14.970000000000000	630730	1	3.00	1	4.99	30.12
10	4.990000000000000	938711	1	1.00	1	4.99	30.12
11	14.970000000000000	630767	1	3.00	1	4.99	30.12
12	4.990000000000000	938704	1	1.00	1	4.99	30.12
13	19.960000000000000	630723	1	4.00	1	4.99	30.12
14	4.990000000000000	938753	1	1.00	1	4.99	30.12

Loading Movement Fact Table

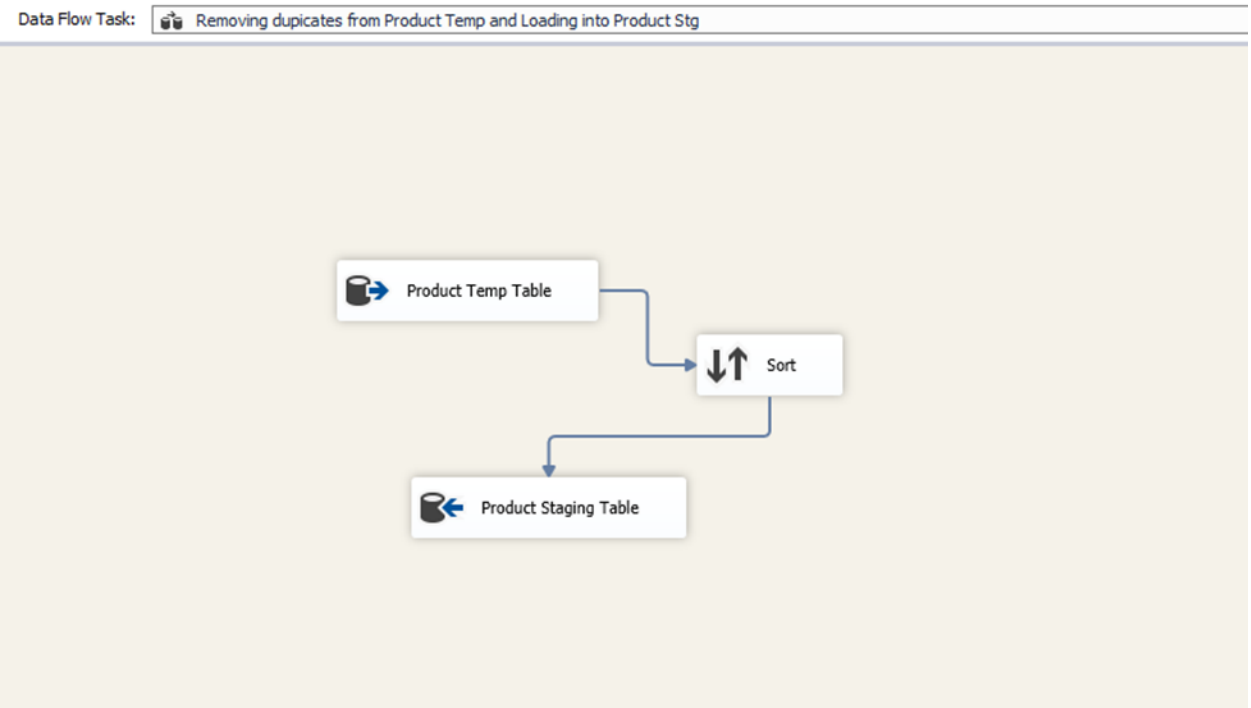

Data Flow Task:
Loading into Movement Fact Table
Movement Staging
Product Staging Table
Derived Column
Sort 1
Merge Join
Sort 2
Movement Fact Table

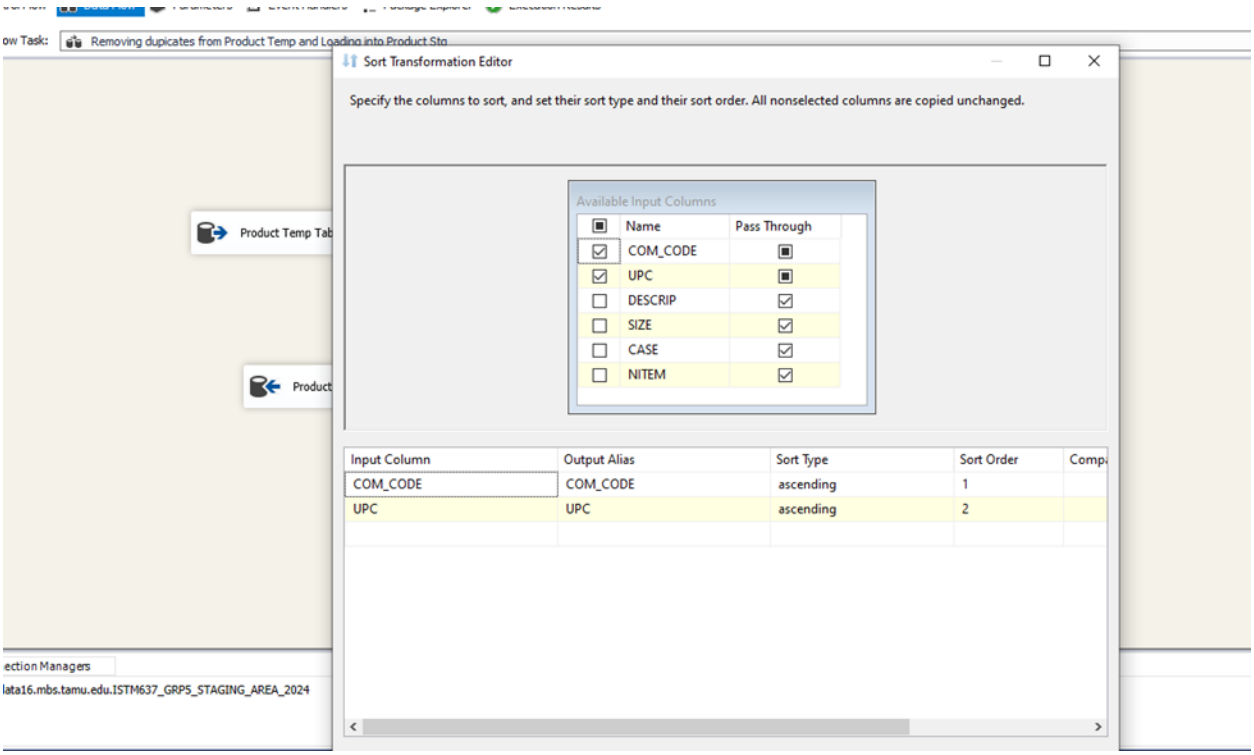
Overall Process:

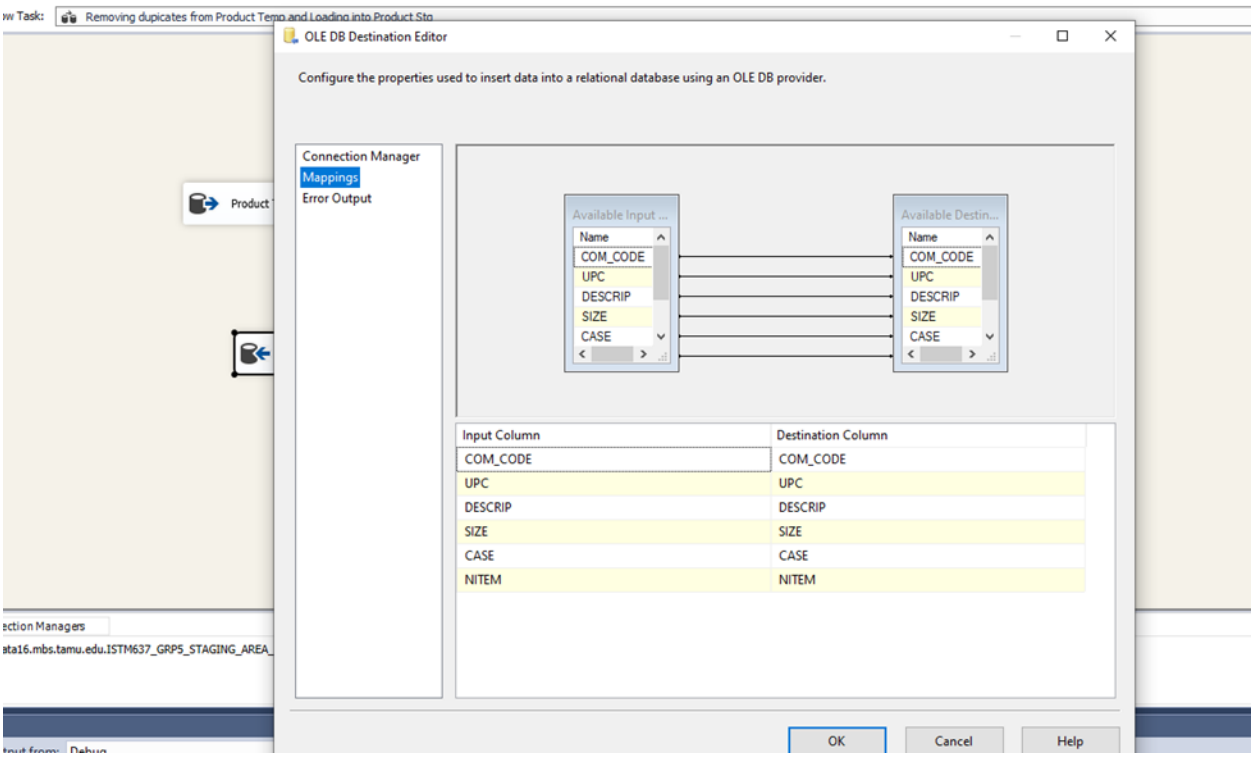


10.Product Staging and Dimension Table Workflow:

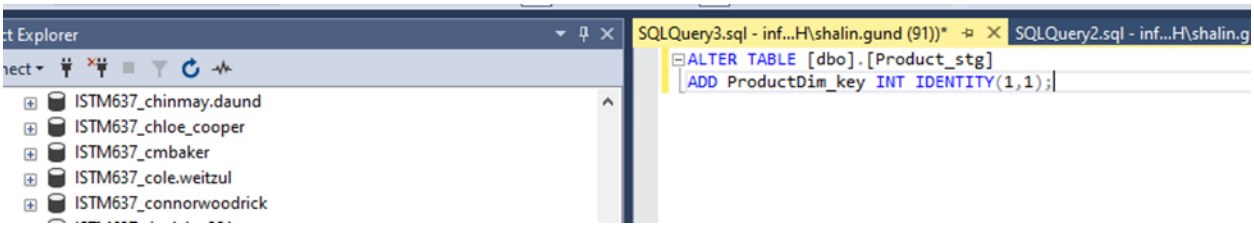
Removing duplicates from the Product temporary table and loading them in the Product staging table

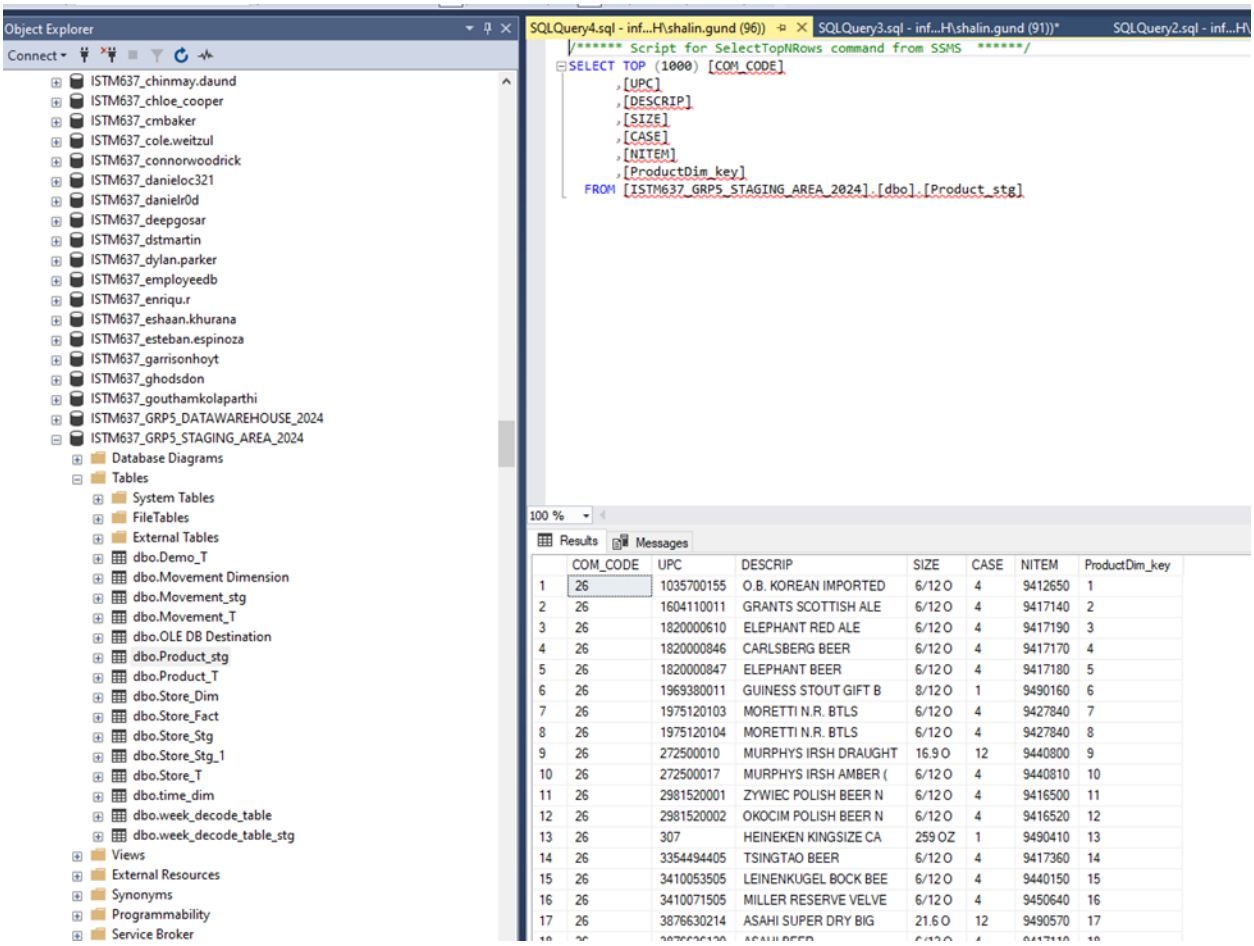



Removing dupicates from Product Temp and Loading into Product Sta
Sort Transformation Editor
Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.
Available Input Columns
Name
Pass Through
COM_CODE
UPC
DESCRIP
SIZE
CASE
NITEM
Input Column
Output Alias
Sort Type
Sort Order
COM_CODE
COM_CODE
ascending
1
UPC
UPC
ascending
2
Product Temp Tab
Product
ection Managers
lata16.mbs.tamu.edu.ISTM637_GRP5_STAGING_AREA_2024


Removing dupicates from Product Temp and Loading into Product Sta
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
Available Input ...
Available Destin...
Name
COM_CODE
UPC
DESCRIP
SIZE
CASE
Input Column
Destination Column
COM_CODE
COM_CODE
UPC
UPC
DESCRIP
DESCRIP
SIZE
SIZE
CASE
CASE
NITEM
NITEM
OK
Cancel
Help
ection Managers
ata16.mbs.tamu.edu.ISTM637_GRP5_STAGING_AREA_

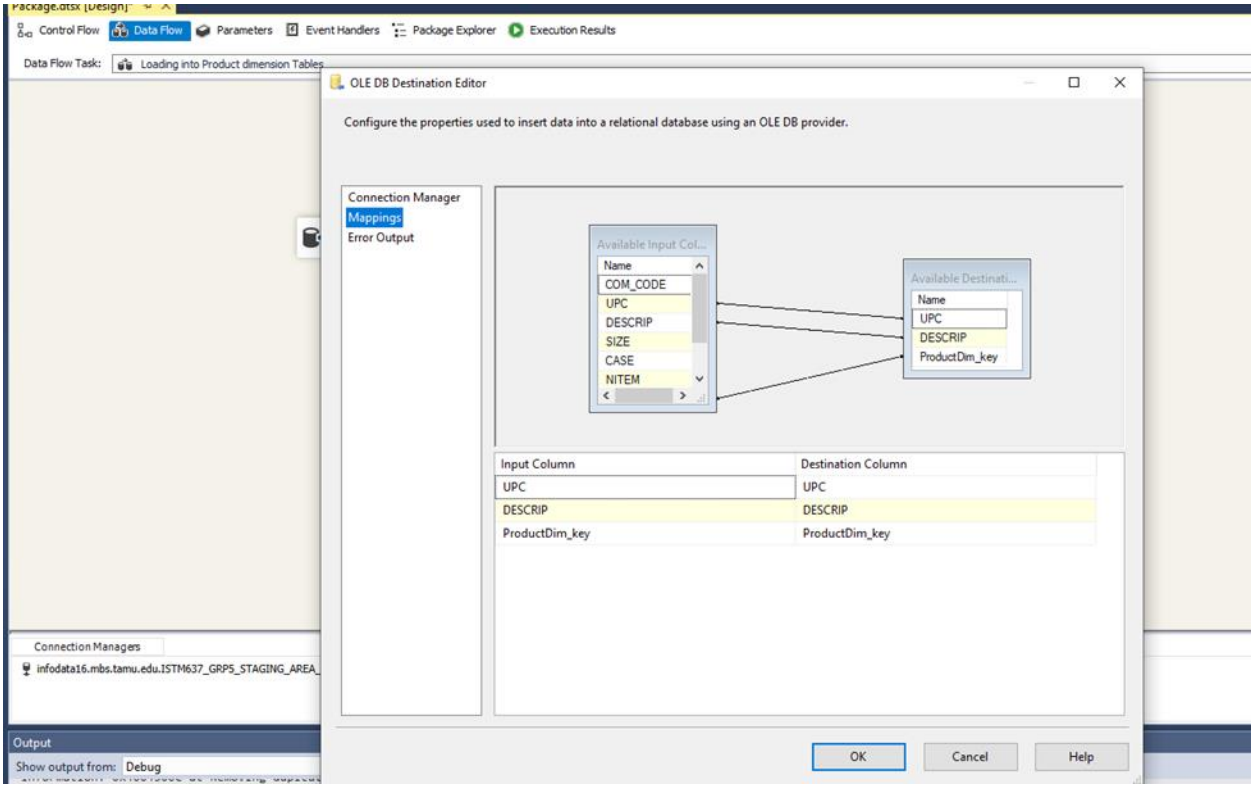
Creating ProductDim_key (Surrogate_key)

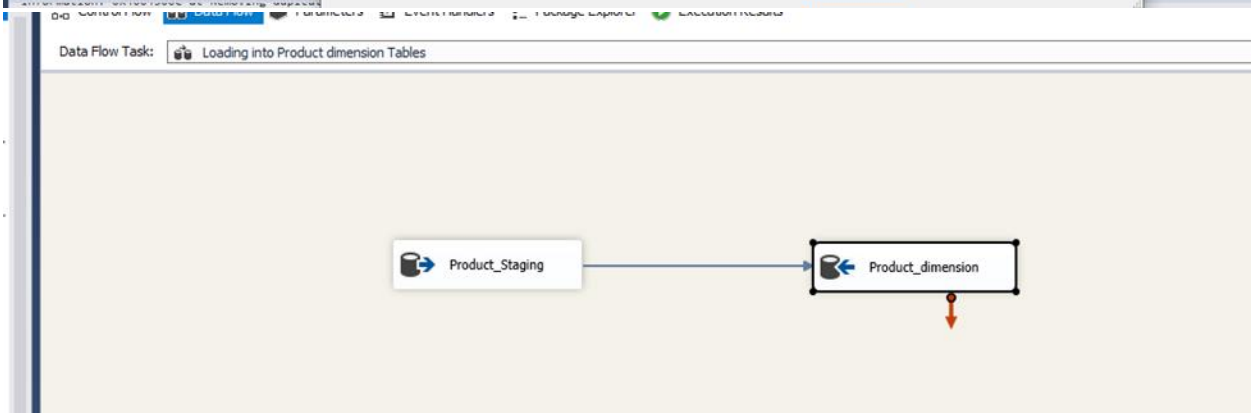




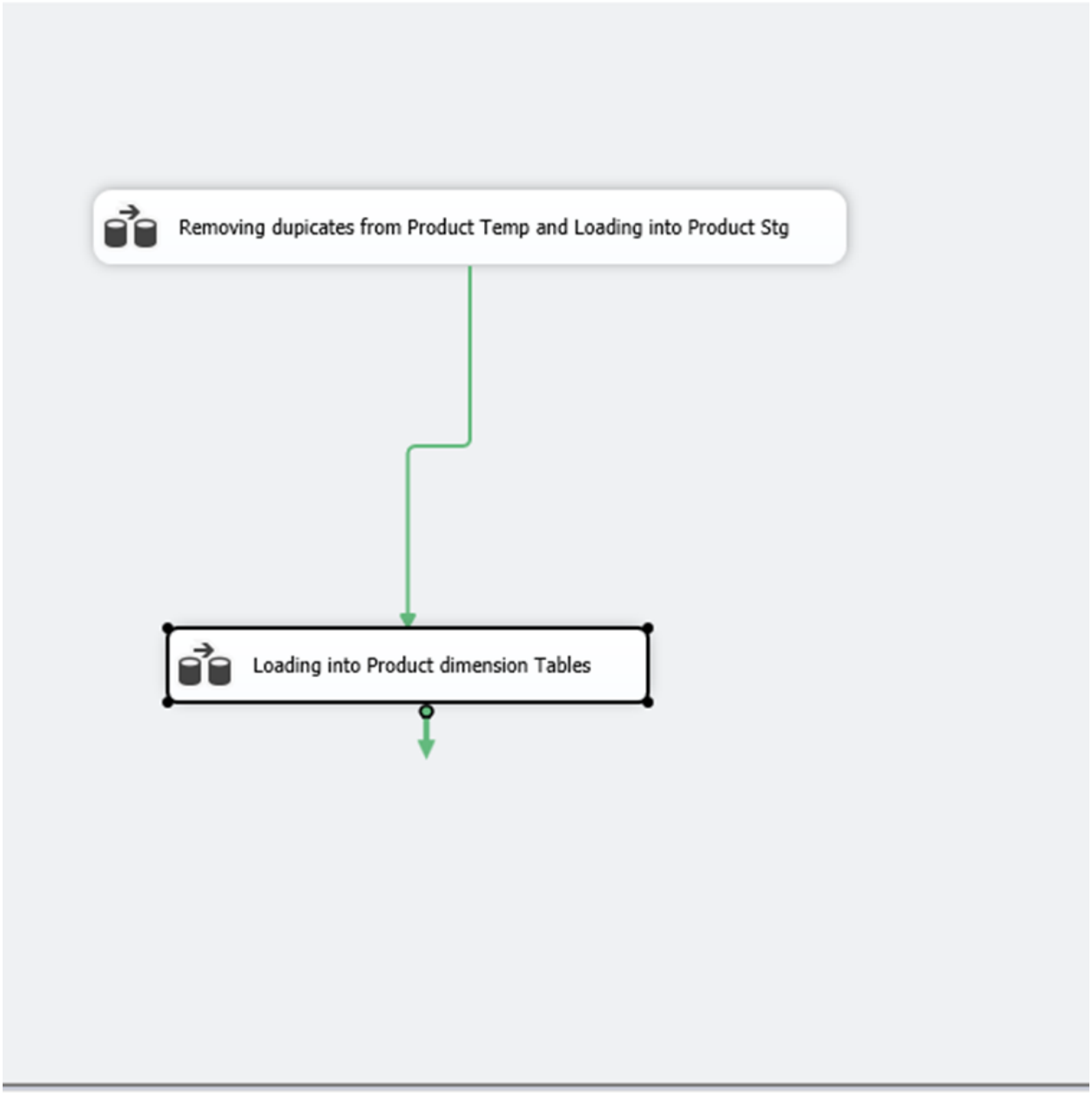
	COM_CODE	UPC	DESCRIP	SIZE	CASE	NITEM	ProductDim_key
1	26	1035700155	O.B. KOREAN IMPORTED	6/12 O	4	9412650	1
2	26	1604110011	GRANTS SCOTTISH ALE	6/12 O	4	9417140	2
3	26	1820000610	ELEPHANT RED ALE	6/12 O	4	9417190	3
4	26	1820000846	CARLSBERG BEER	6/12 O	4	9417170	4
5	26	1820000847	ELEPHANT BEER	6/12 O	4	9417180	5
6	26	1969380011	GUINESS STOUT GIFT B	8/12 O	1	9490160	6
7	26	1975120103	MORETTI N.R. BTLS	6/12 O	4	9427840	7
8	26	1975120104	MORETTI N.R. BTLS	6/12 O	4	9427840	8
9	26	272500010	MURPHYS IRSH DRAUGHT	16.9 O	12	9440800	9
10	26	272500017	MURPHYS IRSH AMBER (	6/12 O	4	9440810	10
11	26	2981520001	ZYWIEC POLISH BEER N	6/12 O	4	9416500	11
12	26	2981520002	OKOCIM POLISH BEER N	6/12 O	4	9416520	12
13	26	307	HEINEKEN KINGSIZE CA	259 OZ	1	9490410	13
14	26	3354494405	TSINGTAO BEER	6/12 O	4	9417360	14
15	26	3410053505	LEINENKUGEL BOCK BEE	6/12 O	4	9440150	15
16	26	3410071505	MILLER RESERVE VELVE	6/12 O	4	9450640	16
17	26	3876630214	ASAHI SUPER DRY BIG	21.6 O	12	9490570	17

Loading in Product Dimension Table


Package.dtsx [Design]
Control Flow
Data Flow
Parameters
Event Handlers
Package Explorer
Execution Results
Data Flow Task:
Loading into Product dimension Tables
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
Available Input Col...
Name
COM_CODE
UPC
DESCRIP
SIZE
CASE
NITEM
Available Destinati...
Name
UPC
DESCRIP
ProductDim_key
Input Column
Destination Column
UPC
UPC
DESCRIP
DESCRIP
ProductDim_key
ProductDim_key
OK
Cancel
Help
Connection Managers
infodata16.mbs.tamu.edu.ISTM637_GRP5_STAGING_AREA_
Output
Show output from:
Debug


Data Flow Task:
Loading into Product dimension Tables
Product_Staging
Product_dimension

Overall Process



List of Temporary Tables:

1. Movement_T: Temporary table to store all the movement data files before loading it into Movement_stg
2. Product_T: Temporary table to store all the product data files before loading into Product_stg

3. Store_T: Temporary table to store data related to demographics and store before loading it into Store_stg
4. Week_Decode_T: Temporary table to store time-related data before loading it into Week_Decode_stg

Section 5: Business Intelligence reporting

Reporting Plan

In this final stage of our project, we utilized many different types of business intelligence tools to create and generate the reports, making use of the data warehouse built for DFF. The main objective behind this was to address critical business problems and help the system users understand the business needs and the overall importance of data presented within the warehouse.

Throughout this project, we developed a good amount of knowledge in several visualization tools and key techniques employed are detailed below:

SSAS - Microsoft SQL Server Analysis Services: SSAS is an analytical data engine which helps in supporting decision making and business analytics. A common approach for deploying a data analysis solution in SSAS involves creating and deploying a multidimensional model as a database to the server instance.

SSRS - Microsoft SQL Server Reporting Services: SSRS is a server-based platform which is used to create and generate reports for visualizing organizational data. This technique is integrated with visual studio and SQL formatting tools, SSRS gives us a way to create drill-down and Ad-hoc reports tailored to the data warehouse in SQL Server.

Tableau: Tableau is a leading BI tool known for interactive visualizations and seamless data exploration. It is available on desktop and web, it offers a free version (Tableau Public) and advanced features in premium plans

To give an effective answer to the business questions, we used the following techniques with the help of visualizations and dashboards:

SSAS - Microsoft SQL Server Analysis Services: Deploying a multidimensional model on SSAS allowed us to use its analytical engine to support decision-making.

SSRS - Microsoft SQL Server Reporting Services: Through SSRS, we created server-based reports that help in visualizing the data and facilitate the generation of drill-down and Ad-hoc reports.
Tableau: Using Tableau's graphical abilities, we created and designed visualizations and dashboards that were user-friendly and accessible both on web browsers and the desktop

Together, these methods gave us a chance to extract meaningful insights, answer business queries and effectively communicate the value of the information stored in the data warehouse. Below, we detail the reporting methods used in each business question

BQ No.	Business Question (BQ)	Reporting Method Deployed
1	Which are the stores that cater to people with income below $15,000? What are the products that they opt for the most?	SSAS Cube
2	How did the sale of alcoholic products fare during the holiday season for the period 1995-1996?	SSRS
3	How has the consumer influx changed year on year? How has it affected the sales of Groceries?	Tableau
4	Which are the products/categories that have been facing a year-on-year declining sales trend?	SSRS with SSAS Cube
5	Which beer brand attracts the highest coupon redemption and buyback offers?	Tableau

Mapping Attributes from Independent Data Marts

BQ1: Which are the stores that cater to people with income below $15,000? What are the products that they opt for the most?

Attribute	Source Table	Filters	Corresponding Report Attribute
STORE	Store_Dim	NA	STORE
Store	Store Dim	None	Store
Customers Under Poverty Line	Store Fact	[Measures].[CUSTCOUN] * [Measures].[POVERT	Customers Under Poverty Line

		Y]	
Grocery Sales	Store Fact	None	Grocery Sales
Dairy Sales	Store Fact	None	Dairy Sales
Meat Sales	Store Fact	None	Meat Sales
Frozen Meat Sales	Store Fact	None	Frozen Meat Sales
Bakery Sales	Store Fact	None	Bakery Sales
Beer Sales	Store Fact	None	Beer Sales
Spirits Sales	Store Fact	None	Spirits Sales
Wine Sales	Store Fact	None	Wine Sales
Top Product 1 Name (Derived)	Derived from Store Fact	Uses the MAX function across all product measures to identify the highest sales category	Top Product 1 Name (calculated dynamically)
Top Product 1 Sales (Derived)	Derived from Store Fact	Uses the MAX function to identify the highest sales value across all product measures	Top Product 1 Sales (calculated dynamically)
Top Product 2 Name (Derived)	Derived from Store Fact	applies MAX logic excluding the Top Product 1 Name	Top Product 2 Name (calculated dynamically)
Top Product 2 Sales (Derived)	Derived from Store Fact	applies MAX logic excluding th on remaining sales values	Top Product 2 Sales (calculated dynamically)
Top Product 3 Name (Derived)	Derived from Store Fact	applies MAX logic excluding the Top Product 1/2 Name	Top Product 3 Name (calculated dynamically)
Top Product 3 Sales (Derived)	Derived from Store Fact	applies MAX logic excluding th on remaining sales values	Top Product 3 Sales (calculated dynamically)
Top Product 4 Name (Derived)	Derived from Store Fact	applies MAX logic excluding the Top Product 1/2/3 Name	Top Product 4 Name (calculated dynamically)

Top Product 4 Sales (Derived)	Derived from Store Fact	applies MAX logic excluding th on remaining sales values	Top Product 4 Sales (calculated dynamically)
Top Product 5 Name (Derived)	Derived from Store Fact	applies MAX logic excluding the Top Product 1/2/3/4 Name	Top Product 5 Name (calculated dynamically)
Top Product 5 Sales (Derived)	Derived from Store Fact	applies MAX logic excluding th on remaining sales values	Top Product 5 Sales (calculated dynamically)

BQ2: How did the sale of alcoholic products fare during the holiday season for the period 1995-1996?

Attribute	Source Table	Filters	Corresponding Report Attribute
STORE	Store_Dim	NA	STORE
Holiday	time_dim	Holiday_flag = 1	HOLIDAY
End_date	Time_dim	End date in the range 200 – 390	-
WINE	Store_Fact	NA	WINE
BEER	Store_Fact	NA	BEER
SPIRIT	Store_Fact	NA	SPIRIT

BQ3: How has the consumer influx changed year on year? How has it affected the sales of Groceries?

Attribute	Source Table	Filters	Corresponding Report Attribute
Customer_Count	Store_Fact	SUM	Custcoun
GROCERY	Store_Fact	SUM	Grocery
Start_date	time_dim	NA	Year(start_date)

BQ4: Which are the products/categories that have been facing a year-on-year declining sales trend?

Attribute	Source Table	Filters	Corresponding Report Attribute
Year	Time Dim	=Year(Fields!End_Date.Value) Extracted from End Date. Grouped using year.	Year
YoY % Change in Grocery Sales (Calculated Member)	Store Fact	IIF(IsNothing(Previous(Sum(Fields!GROCERY.Value))) OR Previous(Sum(Fields!GROCERY.Value)) = 0, 0, (Sum(Fields!GROCERY.Value) - Previous(Sum(Fields!GROCERY.Value))) / Previous(Sum(Fields!GROCERY.Value)) * 100)	YoY % Change in Grocery Sales
YoY % Change in Bakery Sales (Calculated Member)	Store Fact	IIF(IsNothing(Previous(Sum(Fields!BAKERY.Value))) OR Previous(Sum(Fields!BAKERY.Value)) = 0, 0, (Sum(Fields!BAKERY.Value) - Previous(Sum(Fields!BAKERY.Value))) / Previous(Sum(Fields!BAKERY.Value)) * 100)	YoY % Change in Bakery Sales
YoY % Change in Beer Sales	Store Fact	IIF(	YoY % Change in Beer Sales

(Calculated Member)		IsNothing(Previous(Sum(Fields!BEER.Value))) OR Previous(Sum(Fields!BEER.Value)) = 0, 0, (Sum(Fields!BEER.Value) - Previous(Sum(Fields!BEER.Value))) / Previous(Sum(Fields!BEER.Value)) * 100)	
YoY % Change in Wine Sales (Calculated Member)	Store Fact	IIF(IsNothing(Previous(Sum(Fields!WINE.Value))) OR Previous(Sum(Fields!WINE.Value)) = 0, 0, (Sum(Fields!WINE.Value) - Previous(Sum(Fields!WINE.Value))) / Previous(Sum(Fields!WINE.Value)) * 100)	YoY % Change in Wine Sales
YoY % Change in Spirit Sales (Calculated Member)	Store Fact	IIF(IsNothing(Previous(Sum(Fields!SPIRIT.Value))) OR Previous(Sum(Fields!SPIRIT.Value)) = 0, 0, (Sum(Fields!SPIRIT.Value) - Previous(Sum(Fields!SPIRIT.Value))) / Previous(Sum(Fields!SPIRIT.Value)) * 100)	YoY % Change in Spirit Sales
YoY % Change in Meat Sales (Calculated Member)	Store Fact	IIF(IsNothing(Previous(Sum(Fields!MEAT.Valu	YoY % Change in Meat Sales

		e))) OR Previous(Sum(Fields! MEAT.Value)) = 0, 0, (Sum(Fields! MEAT.Value) - Previous(Sum(Fields! MEAT.Value))) / Previous(Sum(Fields! MEAT.Value)) * 100)	
YoY % Change in Frozen Meat (Calculated Member)	Store Fact	IIF(IsNothing(Previous(S um(Fields!MEATFRO Z.Value))) OR Previous(Sum(Fields! MEATFROZ.Value)) = 0, 0, (Sum(Fields! MEATFROZ.Value) - Previous(Sum(Fields! MEATFROZ.Value))) / Previous(Sum(Fields! MEATFROZ.Value)) * 100)	YoY % Change in Frozen Meat
YoY % Change in Dairy Meat (Calculated Member)	Store Fact	IIF(IsNothing(Previous(S um(Fields!DAIRY.Val ue))) OR Previous(Sum(Fields! DAIRY.Value)) = 0, 0, (Sum(Fields!D AIRY.Value) - Previous(Sum(Fields! DAIRY.Value))) / Previous(Sum(Fields! DAIRY.Value)) * 100)	YoY % Change in Dairy Meat Sales

BQ5: Which beer brand attracts the highest coupon redemption and buyback offers?

Attribute	Source Table	Filters	Corresponding Report Attribute
Product_Description	Product_dim	NA	Descrip
SALE_PROMO_CODE	Movement_dim	Calculated field	SUM(Promo Redemption Count)

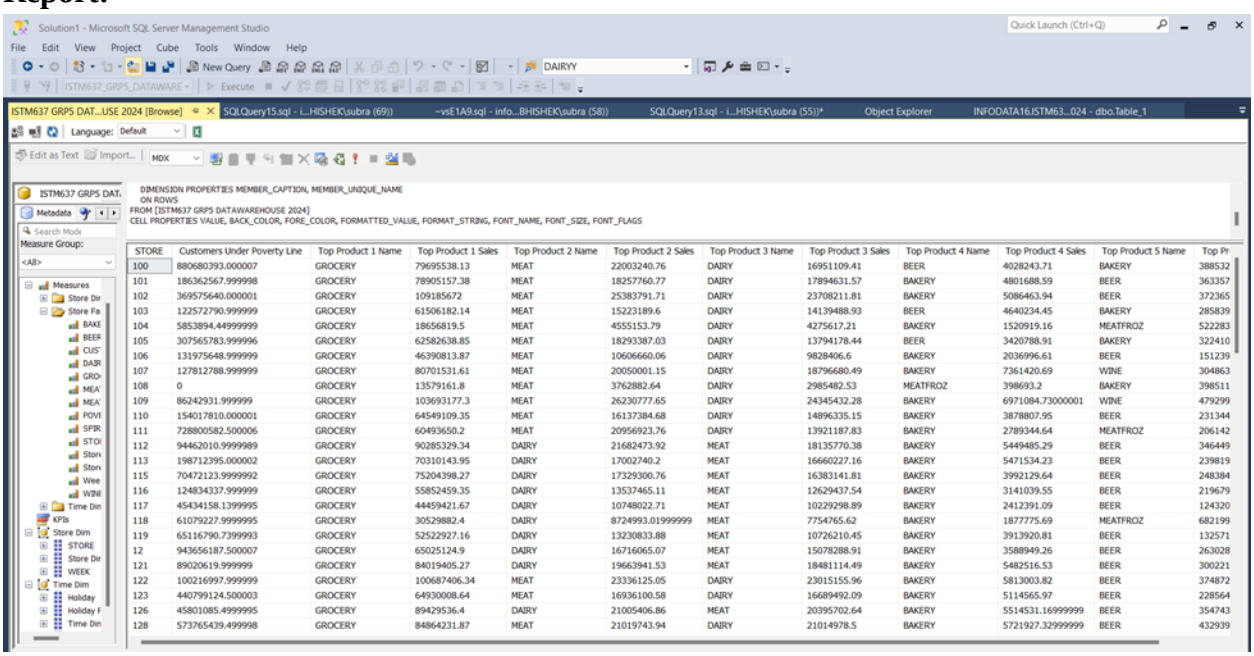
Reporting Implementation

Business Question 1: Which are the stores that cater to people with income below $15,000? What are the products that they opt for the most?

BI/Visualization Method: Analysis using SSAS Cube

Report Plan: This BQ involves analyzing multiple dimensions, such as Income Level, Store Location, and Product Categories. SSAS cubes were chosen for multidimensional data analysis, allowing the user to easily slice and dice data across these dimensions. SSAS cubes store pre-aggregated data, which makes querying faster compared to querying a relational database directly.

Report:



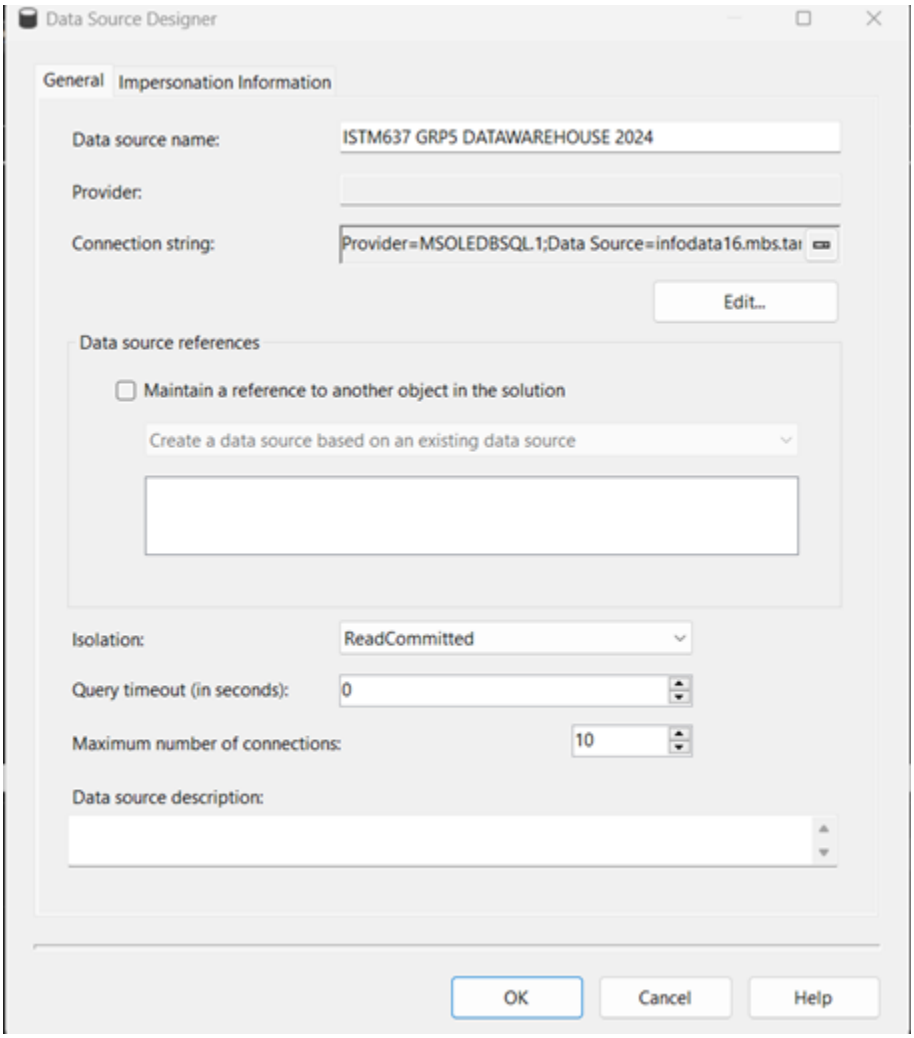
STORE	Customers Under Poverty Line	Top Product 1 Name	Top Product 1 Sales	Top Product 2 Name	Top Product 2 Sales	Top Product 3 Name	Top Product 3 Sales	Top Product 4 Name	Top Product 4 Sales	Top Product 5 Name	Top Pr
100	880680393.000007	GROCERY	79695538.13	MEAT	22003240.76	DAIRY	16951109.41	BEER	4028243.71	BAKERY	388532
101	186362567.999998	GROCERY	78905157.38	MEAT	18257760.77	DAIRY	17894631.57	BAKERY	4801688.59	BEER	363357
102	369575640.000001	GROCERY	109185672	MEAT	25383791.71	DAIRY	23708211.81	BAKERY	5086463.94	BEER	372365
103	122572790.999999	GROCERY	61506182.14	MEAT	15223189.6	DAIRY	14139488.93	BEER	4640234.45	BAKERY	285839
104	5853894.449999999	GROCERY	18656819.5	MEAT	4555153.79	DAIRY	4275617.21	BAKERY	1520919.16	MEATFROZ	522283
105	307565783.999996	GROCERY	62582638.85	MEAT	18293387.03	DAIRY	13794178.44	BEER	3420788.91	BAKERY	322410
106	131975648.999999	GROCERY	46390813.87	MEAT	10606660.06	DAIRY	9828406.6	BAKERY	2036996.61	BEER	151239
107	127812788.999999	GROCERY	80701531.61	MEAT	20050001.15	DAIRY	18796680.49	BAKERY	7361420.69	WINE	304863
108	0	GROCERY	13579161.8	MEAT	3762882.64	DAIRY	2985482.53	MEATFROZ	398693.2	BAKERY	398511
109	86242931.999999	GROCERY	103693177.3	MEAT	26230777.65	DAIRY	24345432.28	BAKERY	6971084.73000001	WINE	479299
110	154017810.000001	GROCERY	64549109.35	MEAT	16137384.68	DAIRY	14896335.15	BAKERY	3878807.95	BEER	231344
111	728800582.500006	GROCERY	60493650.2	MEAT	20956923.76	DAIRY	13921187.83	BAKERY	2789344.64	MEATFROZ	206142
112	94462010.9999989	GROCERY	90285329.34	DAIRY	21682473.92	MEAT	18135770.38	BAKERY	5449485.29	BEER	346449
113	198712395.000002	GROCERY	70310143.95	DAIRY	17002740.2	MEAT	16660227.16	BAKERY	5471534.23	BEER	239819
115	70472123.9999992	GROCERY	75204398.27	DAIRY	17329300.76	MEAT	16383141.81	BAKERY	3992129.64	BEER	248384
116	124834337.999999	GROCERY	55852459.35	DAIRY	13537465.11	MEAT	12629437.54	BAKERY	3141039.55	BEER	219679
117	45434158.1399995	GROCERY	44459421.67	DAIRY	10748022.71	MEAT	10229298.89	BAKERY	2412391.09	BEER	124320
118	61079227.9999995	GROCERY	30529882.4	DAIRY	8724993.01999999	MEAT	7754765.62	BAKERY	1877775.69	MEATFROZ	682199
119	65116790.7399993	GROCERY	52522927.16	DAIRY	13230833.88	MEAT	10726210.45	BAKERY	3913920.81	BEER	132571
12	943656187.500007	GROCERY	65025124.9	DAIRY	16716065.07	MEAT	15078288.91	BAKERY	3588949.26	BEER	263028
121	89020619.999999	GROCERY	84019405.27	DAIRY	19663941.53	MEAT	18481114.49	BAKERY	5482516.53	BEER	300221
122	100216997.999999	GROCERY	100687406.34	MEAT	23336125.05	DAIRY	23015155.96	BAKERY	5813003.82	BEER	374872
123	440799124.500003	GROCERY	64930008.64	MEAT	16936100.58	DAIRY	16689492.09	BAKERY	5114565.97	BEER	228564
126	45801085.4999995	GROCERY	89429536.4	DAIRY	21005406.86	MEAT	20395702.64	BAKERY	5514531.16999999	BEER	354743
128	573765439.499998	GROCERY	84864231.87	MEAT	21019743.94	DAIRY	21014978.5	BAKERY	5721927.32999999	BEER	432939

The analysis identifies that stores catering to customers with an income below $15,000 mainly see grocery items as the most purchased product category. Following groceries, meat ranks as the

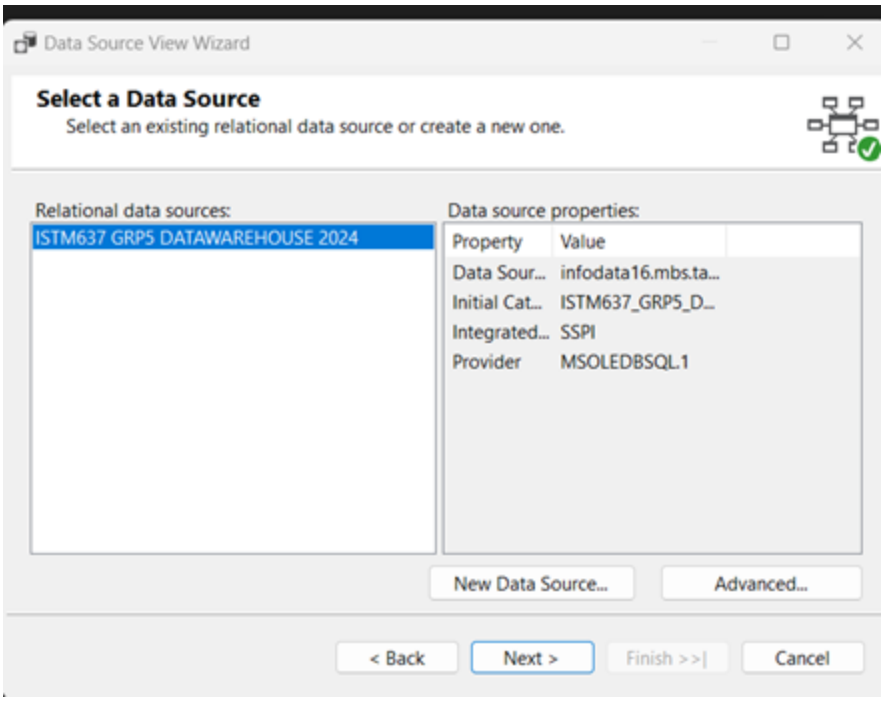
second most opted category, indicating its significance in daily consumption. Dairy products are the third preference, showcasing their necessity in regular diets. The remaining product categories, including bakery and alcoholic products, have relatively lower sales but still contribute to the overall shopping basket. This insight highlights the critical role of groceries, meat, and dairy in driving sales in these stores, providing an opportunity for focused inventory planning and marketing efforts to better serve this demographic.

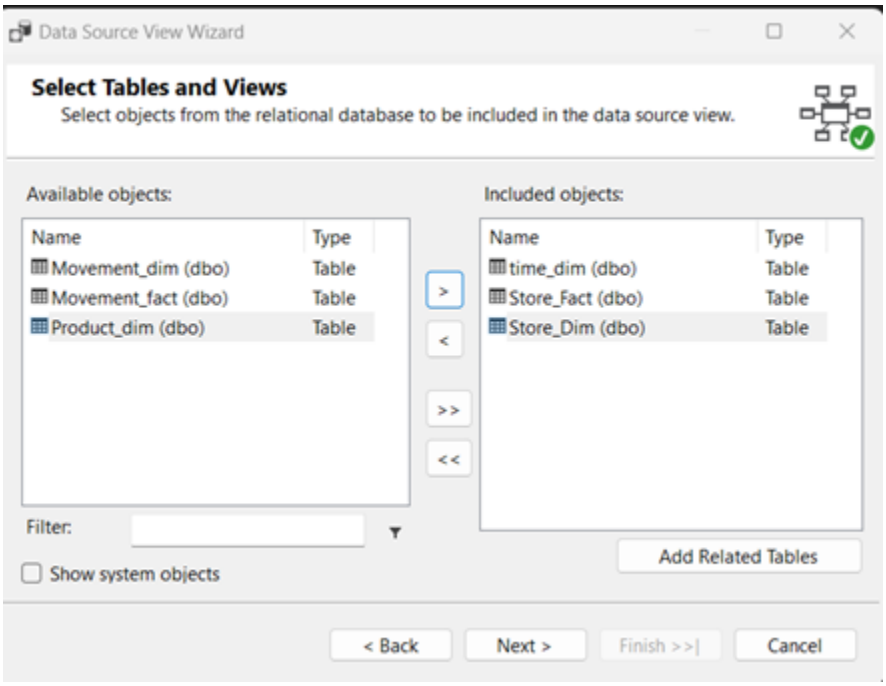
Implementation:

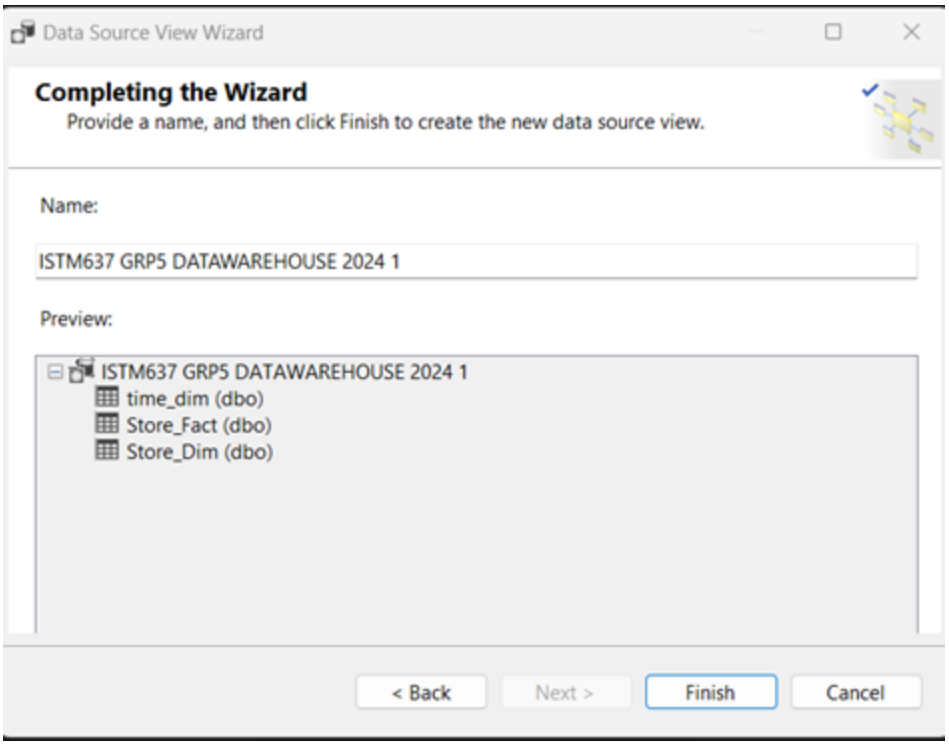
Establishing Connection to the data warehouse server:



Creating the Data Source View for Store level Data Mart: We are using Store Level Data Mart to retrieve solution for our query

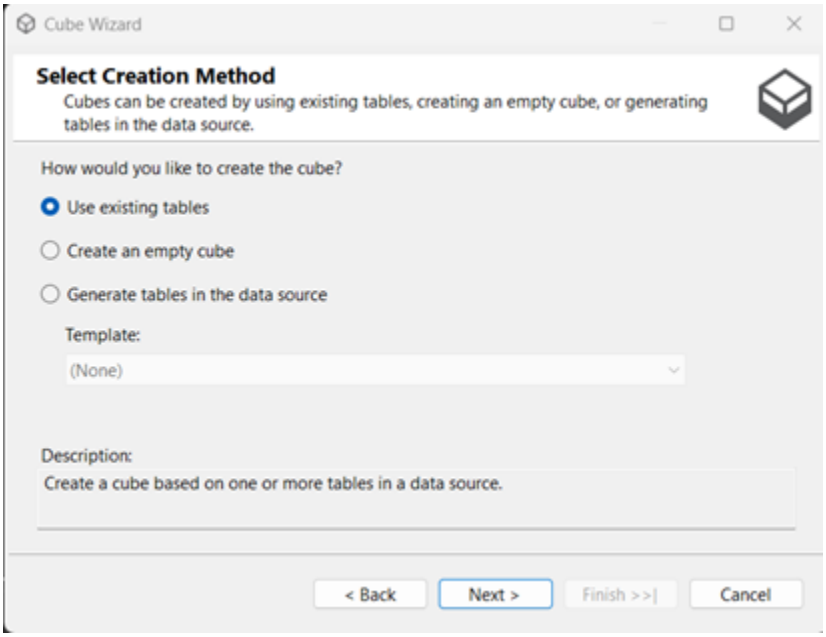

Data Source View Wizard
Select a Data Source
Select an existing relational data source or create a new one.
Relational data sources:
ISTM637 GRP5 DATAWAREHOUSE 2024
Data source properties:
Property
Value
Data Sour...
infodata16.mbs.ta...
Initial Cat...
ISTM637_GRP5_D...
Integrated...
SSPI
Provider
MSOLEDBSQL.1
New Data Source...
Advanced...
< Back
Next >
Finish >>|
Cancel

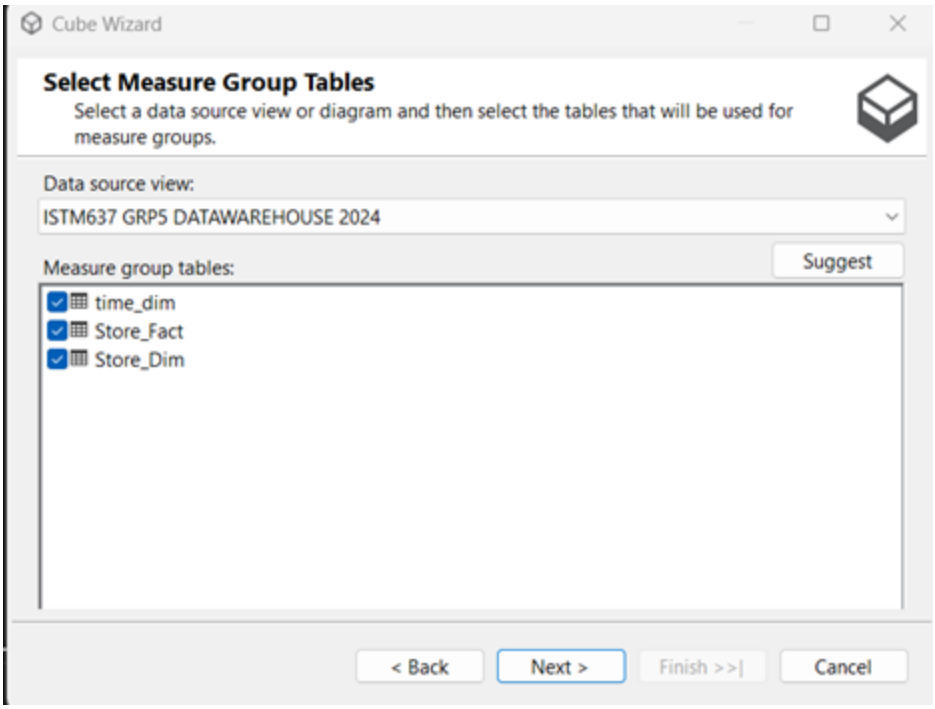

Data Source View Wizard
Select Tables and Views
Select objects from the relational database to be included in the data source view.
Available objects:
Name
Type
Movement_dim (dbo)
Table
Movement_fact (dbo)
Table
Product_dim (dbo)
Table
>
<
>>
<<
Included objects:
Name
Type
time_dim (dbo)
Table
Store_Fact (dbo)
Table
Store_Dim (dbo)
Table
Filter:
Show system objects
Add Related Tables
< Back
Next >
Finish >>|
Cancel

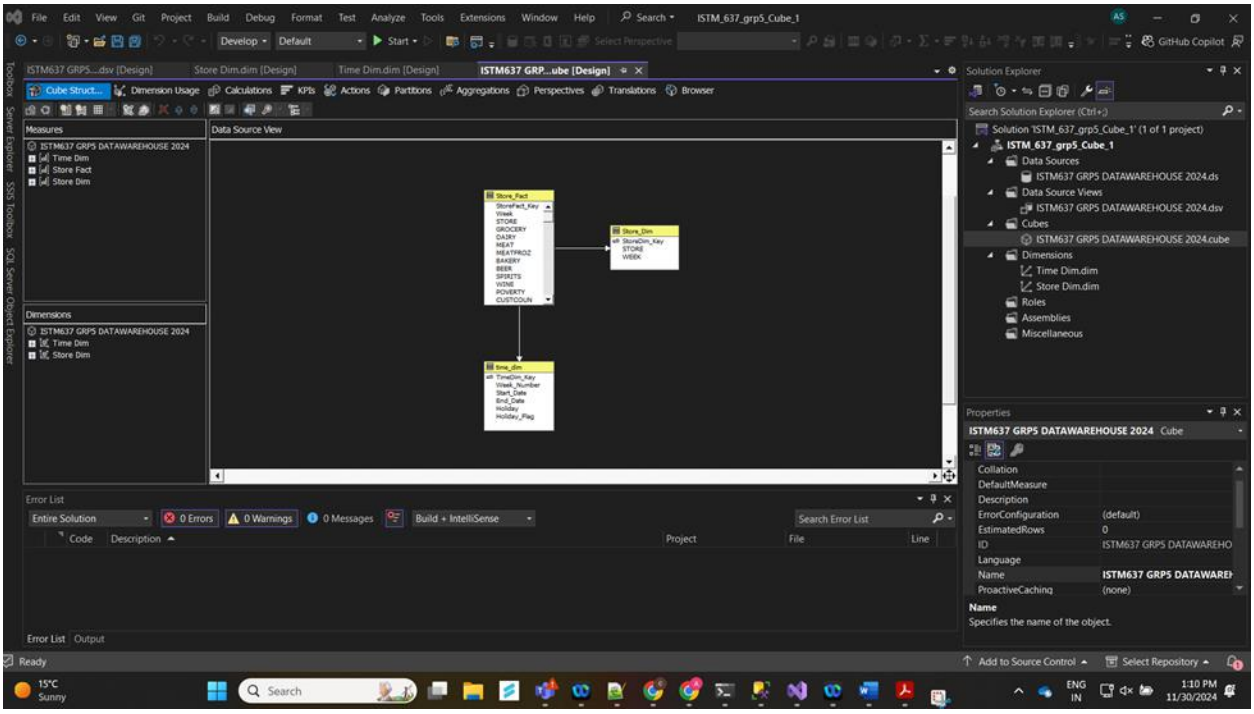


Creating the Cube:

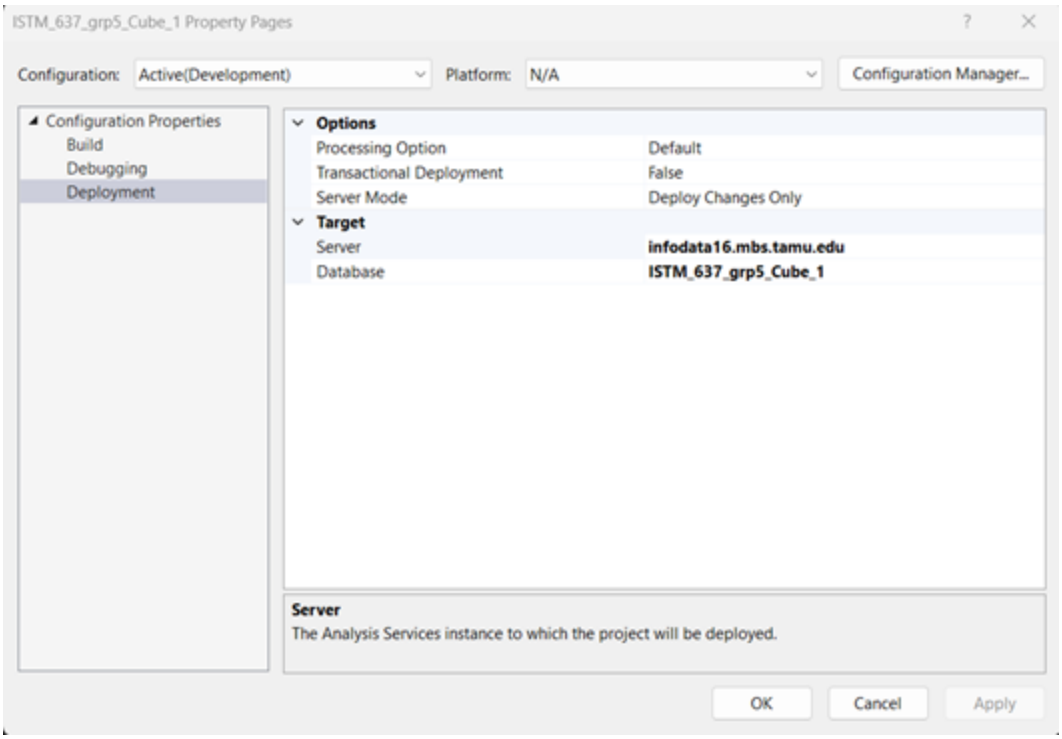
This is a main cube that has all the necessary data related to the stores and their sales. Executives and employees can readily connect to this cube and query for various other business questions related to store sales and their demographic distribution.

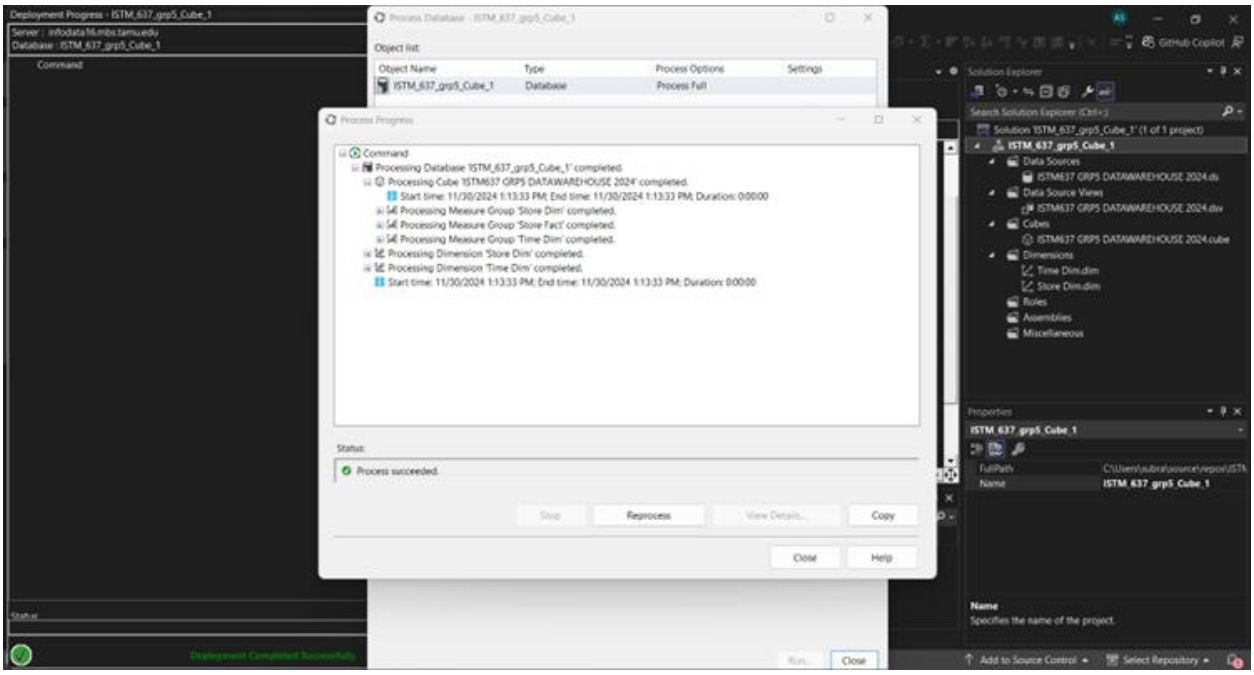






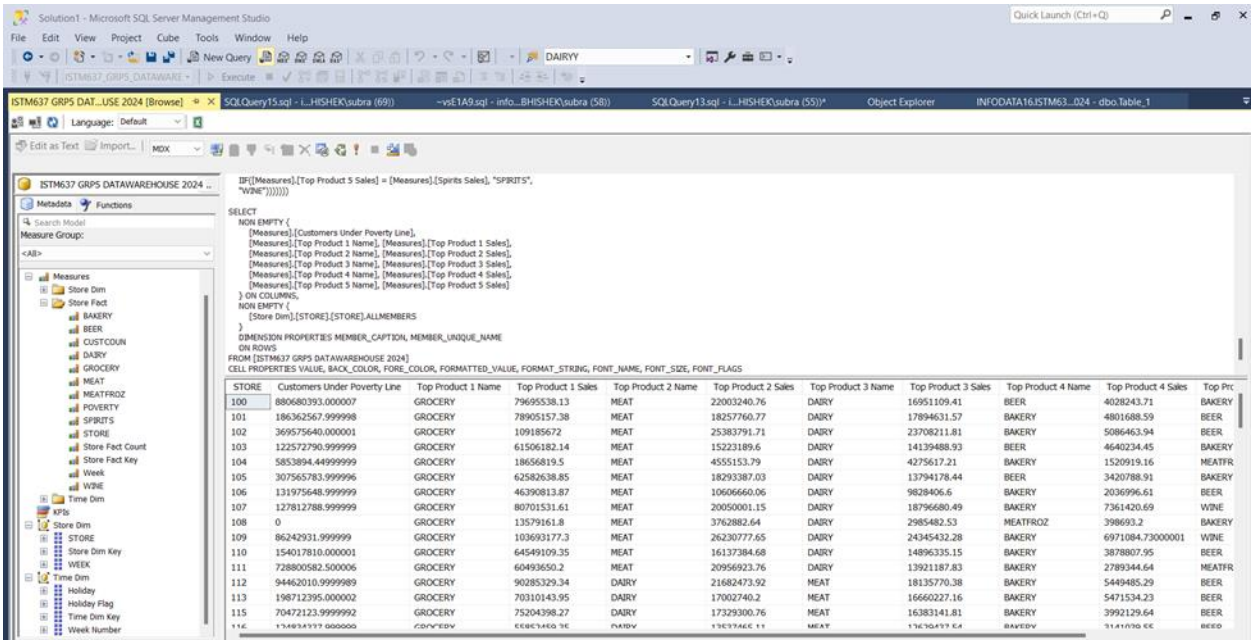
Deploying the Cube on SSAS:





The cube deployment was successful. To check the deployment status, the users can connect to the analysis services server ‘infodata16.mbs.tamu.edu’.

In order to retrieve the needed business questions, the team crafted a MDX query to retrieve the cube results in a tabular format.



STORE	Customers Under Poverty Line	Top Product 1 Name	Top Product 1 Sales	Top Product 2 Name	Top Product 2 Sales	Top Product 3 Name	Top Product 3 Sales	Top Product 4 Name	Top Product 4 Sales	Top Pro
100	880680393.000007	GROCERY	79695538.13	MEAT	22003240.76	DAIRY	16951109.41	BEER	4028243.71	BAKERY
101	186362567.999998	GROCERY	78905157.38	MEAT	18257760.77	DAIRY	17894631.57	BAKERY	4801688.59	BEER
102	369575640.000001	GROCERY	109185672	MEAT	25383791.71	DAIRY	23708211.81	BAKERY	5086463.94	BEER
103	122572790.999999	GROCERY	61506182.14	MEAT	15223189.6	DAIRY	14139488.93	BEER	4640234.45	BAKERY
104	5853894.44999999	GROCERY	18656819.5	MEAT	4555153.79	DAIRY	4275617.21	BAKERY	1520919.16	MEATFR
105	307565783.999996	GROCERY	62582638.85	MEAT	18293387.03	DAIRY	13794178.44	BEER	3420788.91	BAKERY
106	131975648.999999	GROCERY	46390813.87	MEAT	10606660.06	DAIRY	9828406.6	BAKERY	2036996.61	BEER
107	127812788.999999	GROCERY	80701531.61	MEAT	20050001.15	DAIRY	18796680.49	BAKERY	7361420.69	WINE
108	0	GROCERY	13579161.8	MEAT	3762882.64	DAIRY	2985482.53	MEATFROZ	398693.2	BAKERY
109	86242931.999999	GROCERY	103693177.3	MEAT	26230777.65	DAIRY	24345432.28	BAKERY	6971084.73000001	WINE
110	154017810.000001	GROCERY	64549109.35	MEAT	16137384.68	DAIRY	14896335.15	BAKERY	3878807.95	BEER
111	728800582.500006	GROCERY	60493650.2	MEAT	20956923.76	DAIRY	13921187.83	BAKERY	2789344.64	MEATFR
112	94462010.9999989	GROCERY	90285329.34	DAIRY	21682473.92	MEAT	18135770.38	BAKERY	5449485.29	BEER
113	198712395.000002	GROCERY	70310143.95	DAIRY	17002740.2	MEAT	16660227.16	BAKERY	5471534.23	BEER
115	70472123.9999992	GROCERY	75204398.27	DAIRY	17329300.76	MEAT	16383141.81	BAKERY	3992129.64	BEER

Business Question 2: How did the sale of alcoholic products fare during the holiday season for the period 1995-1996?

BI/Visualization Method: Analysis using SSRS

Report Plan:
SSRS is the ideal choice for analyzing the sales trends of alcoholic products during holiday seasons due to its powerful visualization and reporting capabilities. It enables users to aggregate data for a high-level overview while supporting detailed drill-down functionality, allowing for an in-depth examination of specific holiday periods and product categories. The ability to incorporate dynamic parameters, such as filtering by year, holiday type, and store region, enhances the flexibility of the analysis. This makes SSRS well-suited for identifying seasonal sales trends, evaluating the performance of specific products, and uncovering opportunities for optimization in inventory management and marketing.

Report:

Sale of Alcholic Products Over the

Not secure infodata16.mbs.tamu.edu/ReportServer/Pages/ReportViewer.aspx?%2fBusiness_C

1 of 2 ? 100%

Sale of Alcholic Products Over the Holiday

STORE	HOLIDAY	WINE	BEER	SPIRITS
⊞129		600062.26	388981.92	262108.11
⊞139		391081.5	443794.46	255744.48
⊞304		591738.7	991695.11	332602.87
⊞309		560804.68	796251.89	435272.26
⊞301		385534.64	765477.21	297573.73
⊞126		401158.05	376584.25	217685.99
⊞302		386087.01	746551.07	368247.67
⊞315		317415.14	671916.32	278582.88
⊞306		396178.27	677019.45	312365
⊞312		279579.95	625782.77	274947.95
⊞128		365557.12	412315.89	236752.53
⊞86		357345.29	379802.85	190312.75
⊞307		319774.1	725679.47	293608.55
⊞305		278487.48	658519.05	284360.37
⊞122		283448.86	364779.71	240950.91
⊞132		297101.2	296144.08	197760.7
⊞62		433861.26	264962.9	203586.21
⊞102		219733.09	362782.81	168839.24
⊞310		286197.67	511343.23	188280.71
⊞134		179842.33	323286.98	157469.82
⊞303		205630.46	566174.43	253520.69
⊞71		231549.64	242071.19	141277.67
⊞308		245253.06	490511.56	190763.03
⊞84		219301.91	204105.36	123622.96
⊞109		516058.37	289325.83	175197.45
⊞103		161587.13	465922.02	213190.6
⊞98		201540.55	337416.06	253494.18
⊞94		176803.48	194160.09	102405.06
⊞131		298103.59	339633.81	175849.78

Search

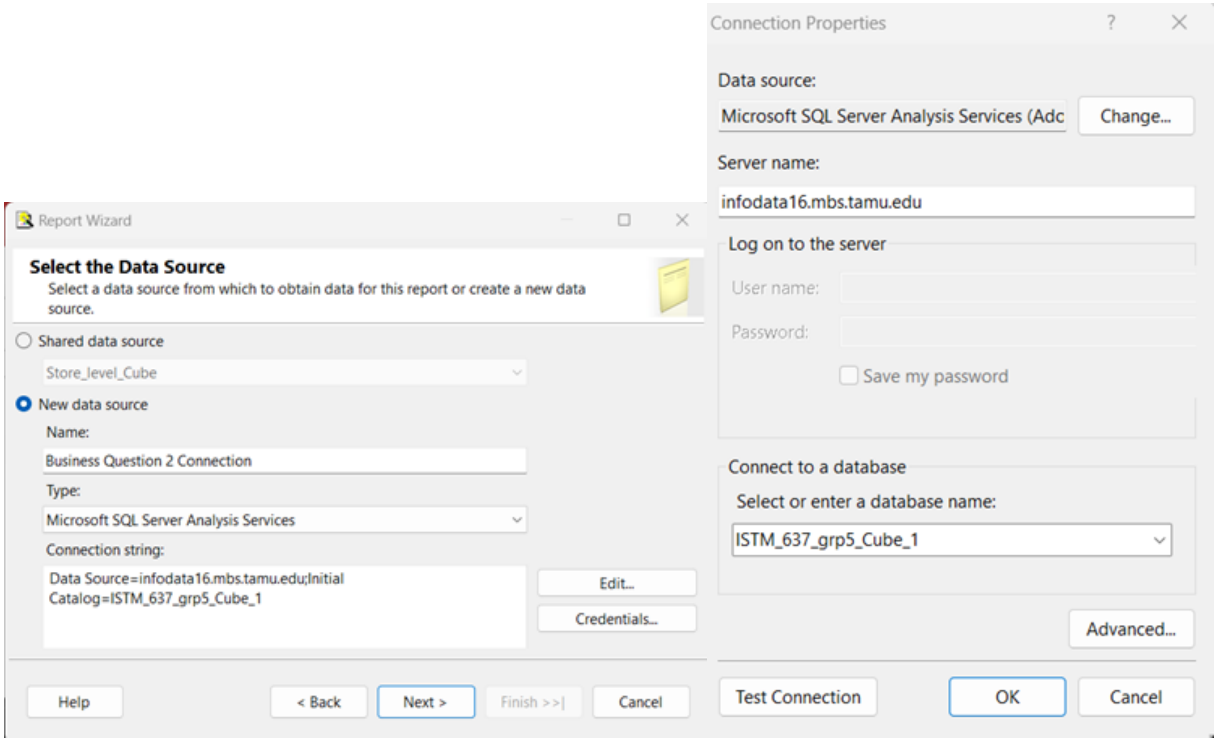
Sale of Alcholic Products Over the Holiday				
STORE	HOLIDAY	WINE	BEER	SPIRITS
⊟129		**600062.26**	**388981.92**	**262108.11**
	Christmas	138488.76	62498.87	63979.33
	Easter	53809.03	35092.21	21869.78
	Fourth of July	59108.35	60915.2	30861.51
	Halloween	48349.26	28493.27	19127.51
	Labor Day	63675.35	53618.09	27200.46
	Memorial Day	61414.29	56423.39	29517.16
	New Year	91951.19	48331.67	34527.84
	Presidents Day	33482.07	17696.43	12143.29
	Thanksgiving	49783.96	25912.79	22881.23
⊟139		**391081.5**	**443794.46**	**255744.48**
	Christmas	106916.48	72697.47	72351.53
	Easter	23339.14	27191.14	13481.83
	Fourth of July	36133.87	69449.27	28162.36
	Halloween	27363.19	35585.21	18330.27
	Labor Day	34728.99	61971.25	24855.25
	Memorial Day	36361.11	60127.72	24201.53
	New Year	72223.53	57739.03	39181.48
	Presidents Day	19659.44	22288.09	11564.76
	Thanksgiving	34355.75	36745.28	23615.47
⊟304		**591738.7**	**991695.11**	**332602.87**
	Christmas	124272.37	133141.73	80207.03
	Easter	52864.72	75480.59	25567.65
	Fourth of July	54141.61	157159.91	35035.06
	Halloween	40390.02	74987.76	23421.95
	Labor Day	50121.3	137400.05	30367.07
	Memorial Day	54493.77	151077.03	29410.82
	New Year	119251.76	135043.35	58026.63

Alcoholic beverages are among the marquee products at stores which are similar to DFF in nature. Holiday season is a peak sales period for categories under such products. Therefore, the team devised a report in SSRS to visualize the effect of seasonality on alcoholic beverages such as beer, wine, and spirits. The published report can be found on the official report server. The graph is by default sorted w.r.t stores that show a spike in sales during the holiday period. Store 129,139 and 304 are the top 3 stores from a sales perspective for alcoholic beverages. In all these stores, Christmas week is the busiest and creates a lot of sales for wine, beer and spirits. On the other hand, President's Day weekend receives the least amount of sales due to alcoholic beverages.

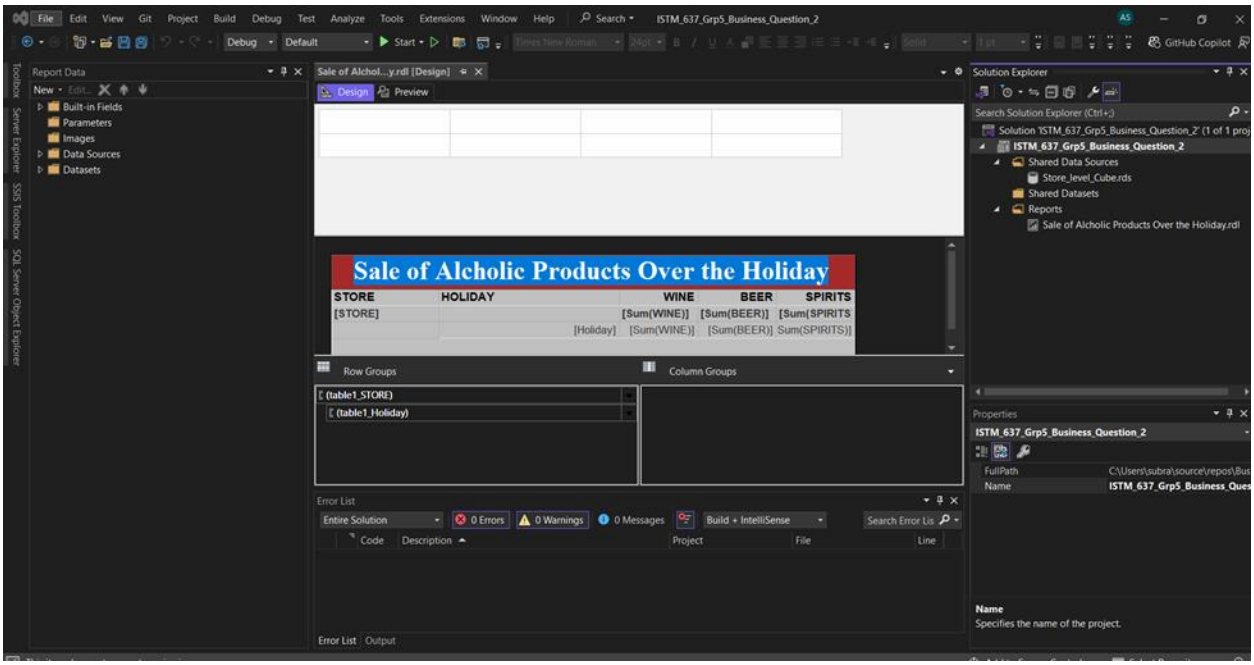
Lastly, the sales of alcoholic beverages hold a steady ground and ranges among similar figures when it comes to other holiday weekends. The report can be scrutinized further by adding dates and enabling further drilldown.

Implementation:

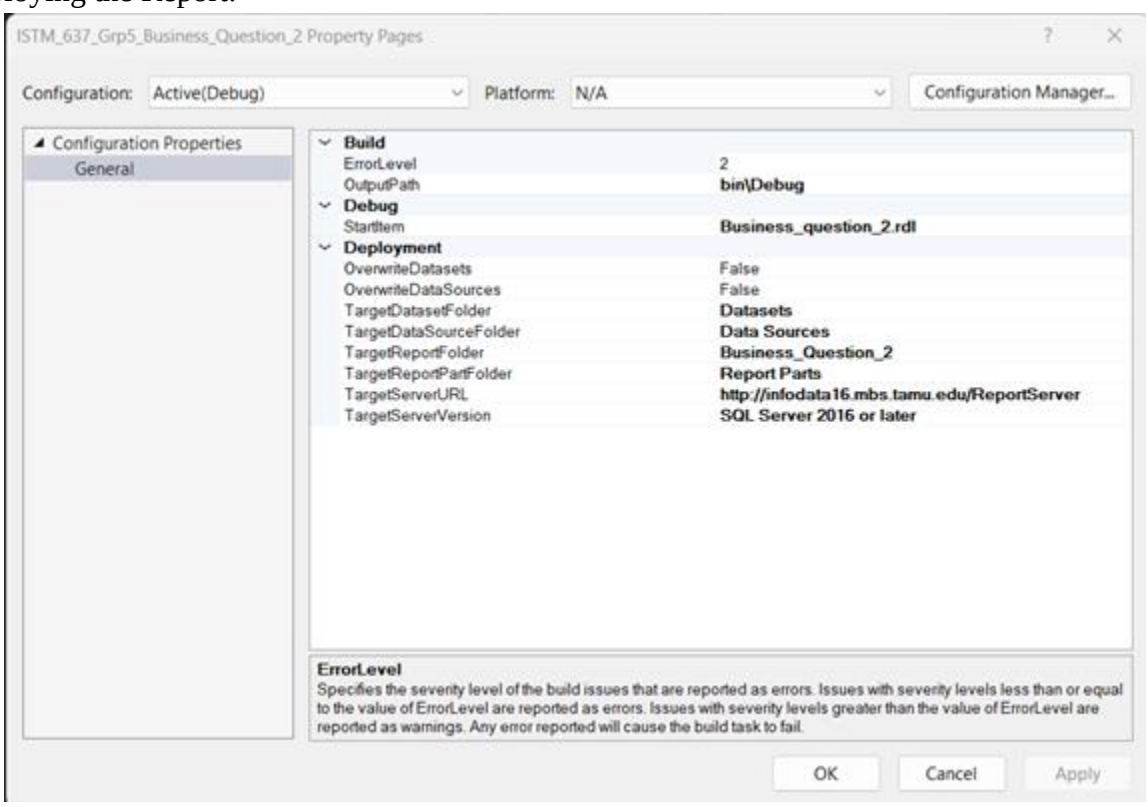
Establishing the connection to the data warehouse server:

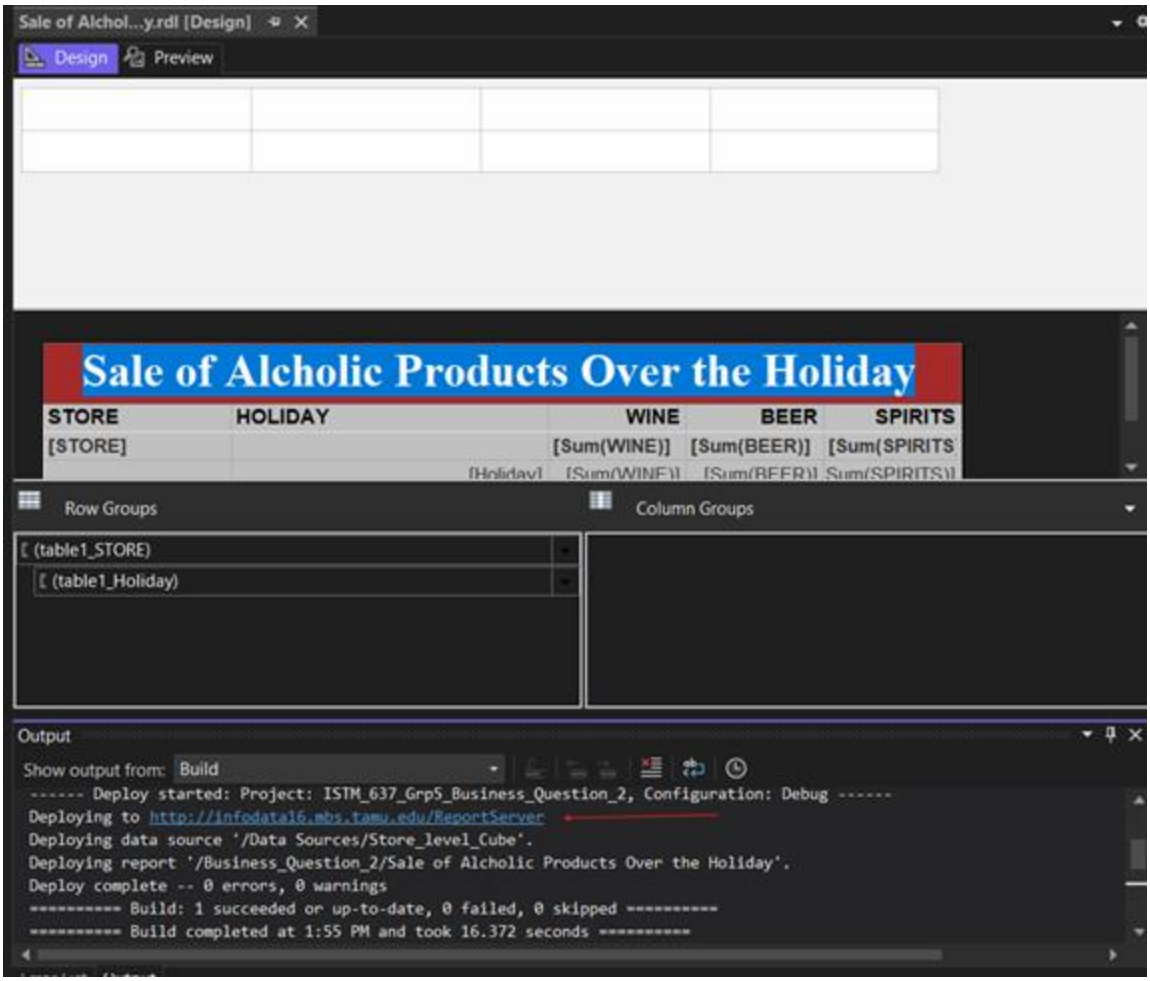


Creating the report in Designer Mode:



Deploying the Report:





Once the report has been deployed successfully, the link prompted on the screen will redirect the user to the desired report.

```
    Monday, November 27, 2023 7:02 PM        <dir> BQ5_SSRS
  Thursday, November 30, 2023 2:39 AM        <dir> BQ5-SSRS
Wednesday, November 29, 2023 12:46 AM        <dir> BQ5-Visualization
    Monday, November 27, 2023 9:24 PM        <dir> Bus_Q1_Minato
    Friday, November 29, 2024 4:03 PM        <dir> Business_Question_2
  Thursday, November 30, 2023 1:04 PM        <dir> BusinessQuestion4
Wednesday, November 29, 2023 11:10 PM        <dir> BusQ5_JaKaGop
    Sunday, November 26, 2023 9:07 PM        <dir> bw3
     Friday, January 19, 2024 4:23 PM        <dir> bzhang_ReportProject
```

[To Parent Directory]

Friday, November 29, 2024 4:03 PM	39466	Business_question_2
Friday, November 29, 2024 4:06 PM	53605	ISTM_637_Grp5_BQ2
Friday, November 29, 2024 4:03 PM	643	Report1
Friday, November 29, 2024 4:03 PM	53605	Report2
Saturday, November 30, 2024 1:55 PM	45012	Sale of Alcholic Products Over the Holiday

Sale of Alcholic Products Over the

Not secure infodata16.mbs.tamu.edu/ReportServer/Pages/ReportViewer.aspx?%2fBusiness_C

1 of 2 ? 100%

Sale of Alcholic Products Over the Holiday

STORE	HOLIDAY	WINE	BEER	SPIRITS
⊞129		600062.26	388981.92	262108.11
⊞139		391081.5	443794.46	255744.48
⊞304		591738.7	991695.11	332602.87
⊞309		560804.68	796251.89	435272.26
⊞301		385534.64	765477.21	297573.73
⊞126		401158.05	376584.25	217685.99
⊞302		386087.01	746551.07	368247.67
⊞315		317415.14	671916.32	278582.88
⊞306		396178.27	677019.45	312365
⊞312		279579.95	625782.77	274947.95
⊞128		365557.12	412315.89	236752.53
⊞86		357345.29	379802.85	190312.75
⊞307		319774.1	725679.47	293608.55
⊞305		278487.48	658519.05	284360.37
⊞122		283448.86	364779.71	240950.91
⊞132		297101.2	296144.08	197760.7
⊞62		433861.26	264962.9	203586.21
⊞102		219733.09	362782.81	168839.24
⊞310		286197.67	511343.23	188280.71
⊞134		179842.33	323286.98	157469.82
⊞303		205630.46	566174.43	253520.69
⊞71		231549.64	242071.19	141277.67
⊞308		245253.06	490511.56	190763.03
⊞84		219301.91	204105.36	123622.96
⊞109		516058.37	289325.83	175197.45
⊞103		161587.13	465922.02	213190.6
⊞98		201540.55	337416.06	253494.18
⊞94		176803.48	194160.09	102405.06
⊞131		298103.59	339633.81	175849.78

Search

Business Question 3: How has the consumer influx changed year on year? How has it affected the sales of Groceries?

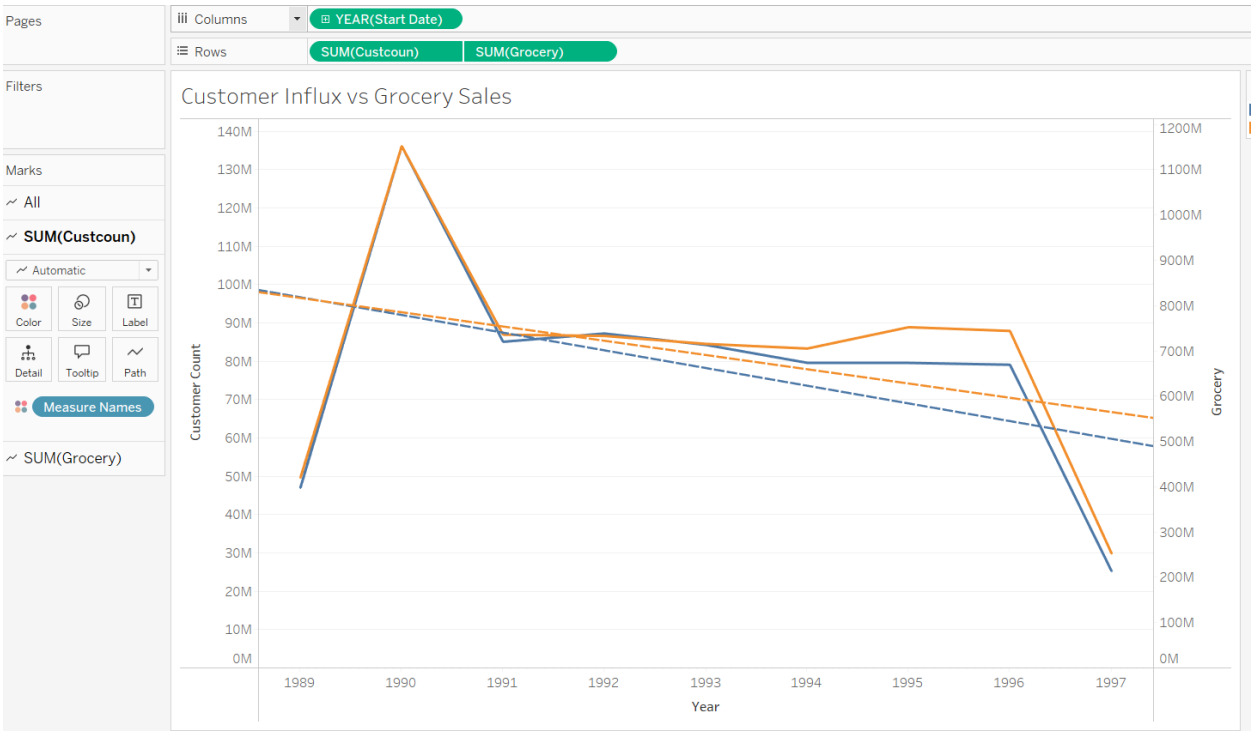
BI/Visualization Method: Tableau

Report Plan:

Tableau was chosen for this question due to its ability to visualize complex data trends and patterns effectively, especially over a timeline. The question involves a year-on-year analysis of consumer influx and its impact on grocery sales, which benefits from Tableau's interactive visualizations, such as line chart. These tools make it easier to explore trends and relationships in the data at a glance.

Moreover, Tableau's ability to filter, drill down, and combine multiple data views allows the user to analyze consumer influx trends by region, store, or price tier dynamically. This insight supports data-driven decision-making for inventory management and seasonal preparedness. Overall, Tableau's strength in visual analytics makes it an ideal choice for addressing this business question.

Report:



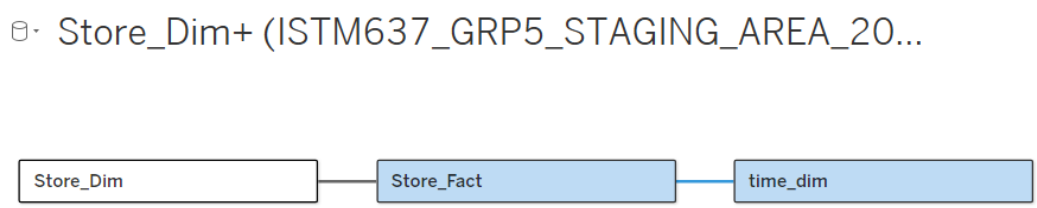
This reveals a declining trend in both customer influx and grocery sales over the years. This downward trajectory indicates a possible shift in consumer behavior or preferences, potentially

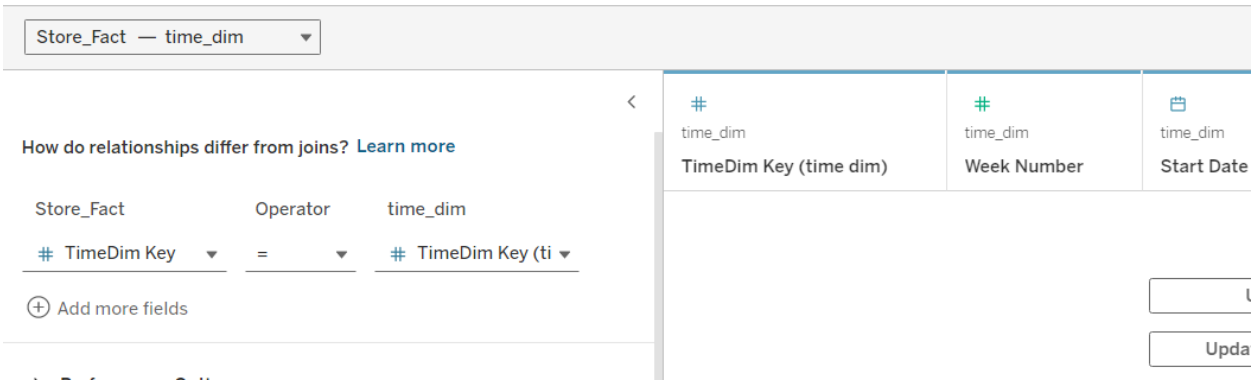
influenced by factors such as increased competition, changing economic conditions, or evolving market dynamics.

The year-on-year decline highlights the need for a shift in strategy to address this trend. Businesses could explore enhancing promotional efforts, tailoring product offerings, or improving the overall shopping experience to retain customers and boost grocery sales. Identifying specific time periods or regions where the decline is more pronounced can provide further insights for targeted actions.

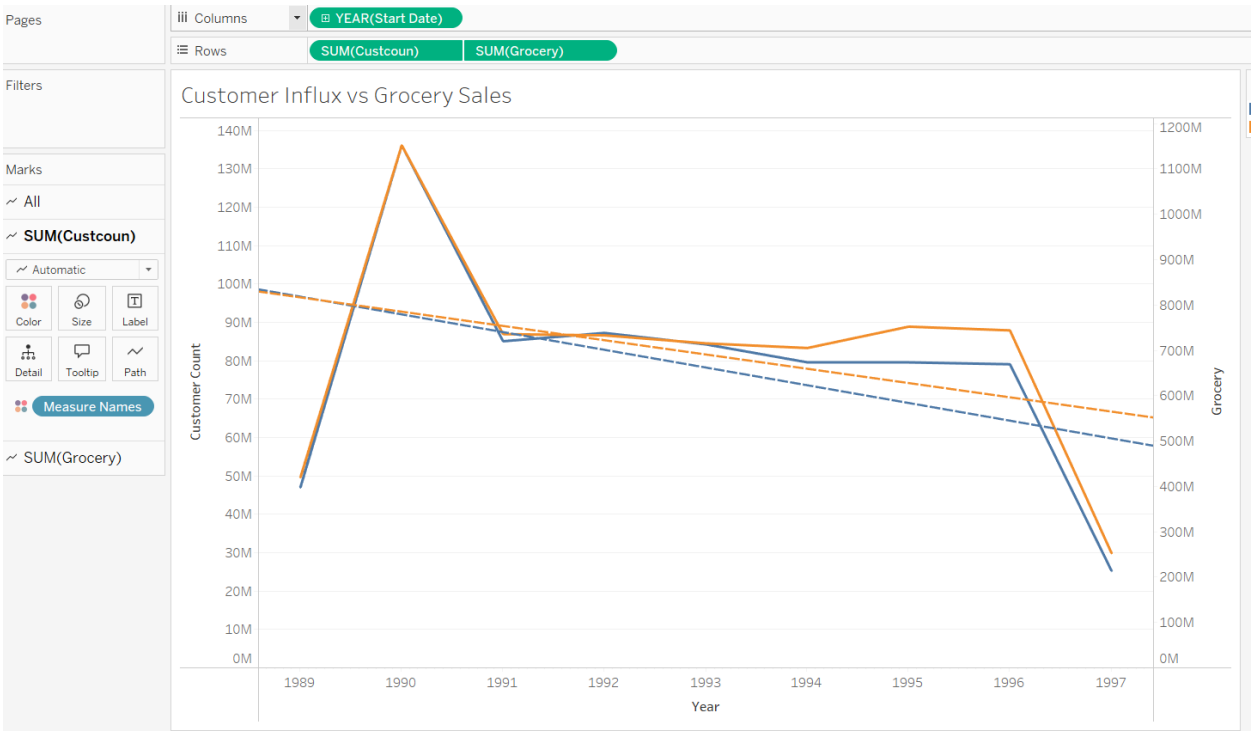
Implementation:

Connected the server in tableau and retrieved the data. Defined relationship in the “data source” section of tableau for clear analysis:





Plotted a line chart with Customer Count and Total Grocery Sales on Y-Axis and Years on X-Axis. A trendline was included to determine the year-on-year change in customer influx and grocery sales.



Business Question 4: Which are the products/categories that have been facing a year-on-year declining sales trend?

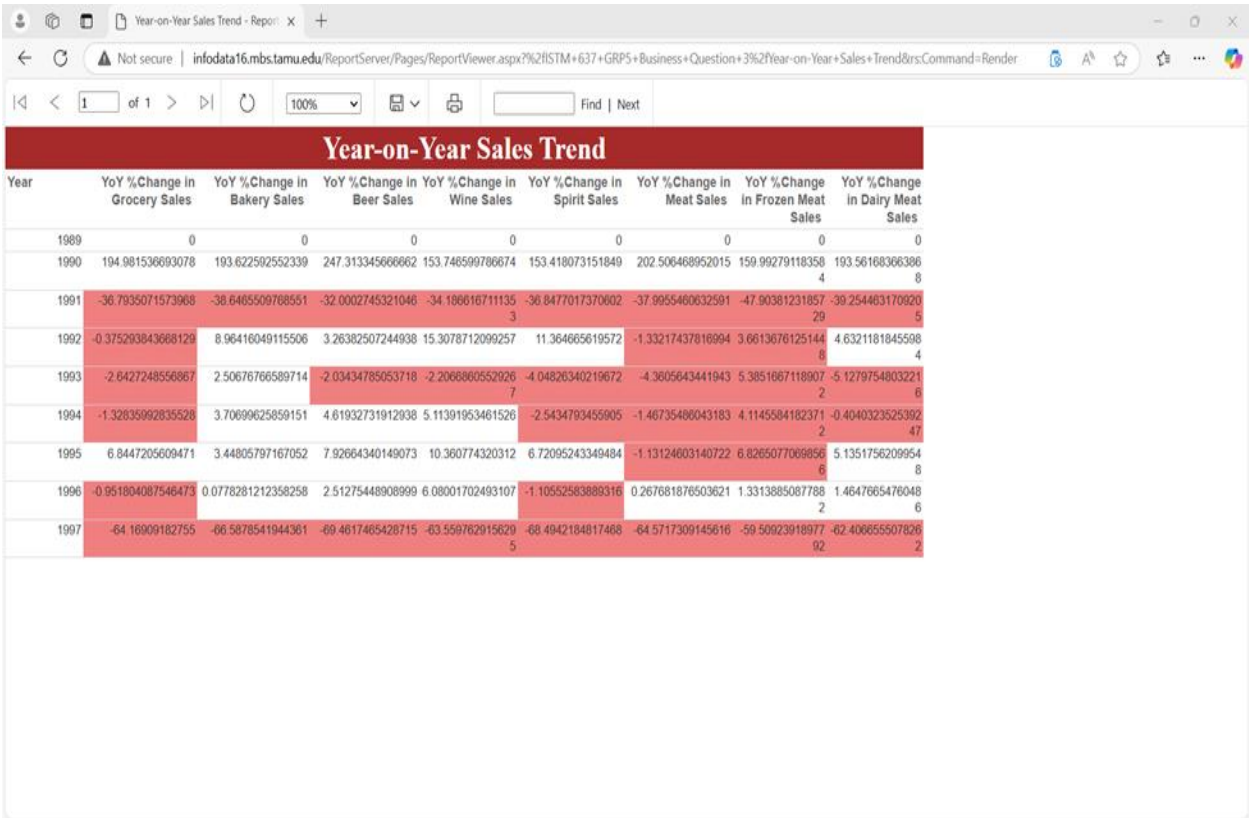
BI/Visualization Method: SSRS with SSAS Cube

Report Plan:

This combination is suitable for analyzing multidimensional data and presenting detailed, interactive report. SSAS cubes allow for pre-aggregated data to be queried efficiently, enabling quick identification of trends across multiple dimensions such as time, product categories, and regions.

SSRS complements this by providing customizable and drillable reports, which are ideal for visualizing year-on-year sales trends. Its ability to highlight declining categories with precision and provide insights at various levels of granularity makes it an effective tool for this business question. Together, SSRS and SSAS enable the user to explore trends interactively and derive insights, helping businesses address underperforming categories strategically.

Report:



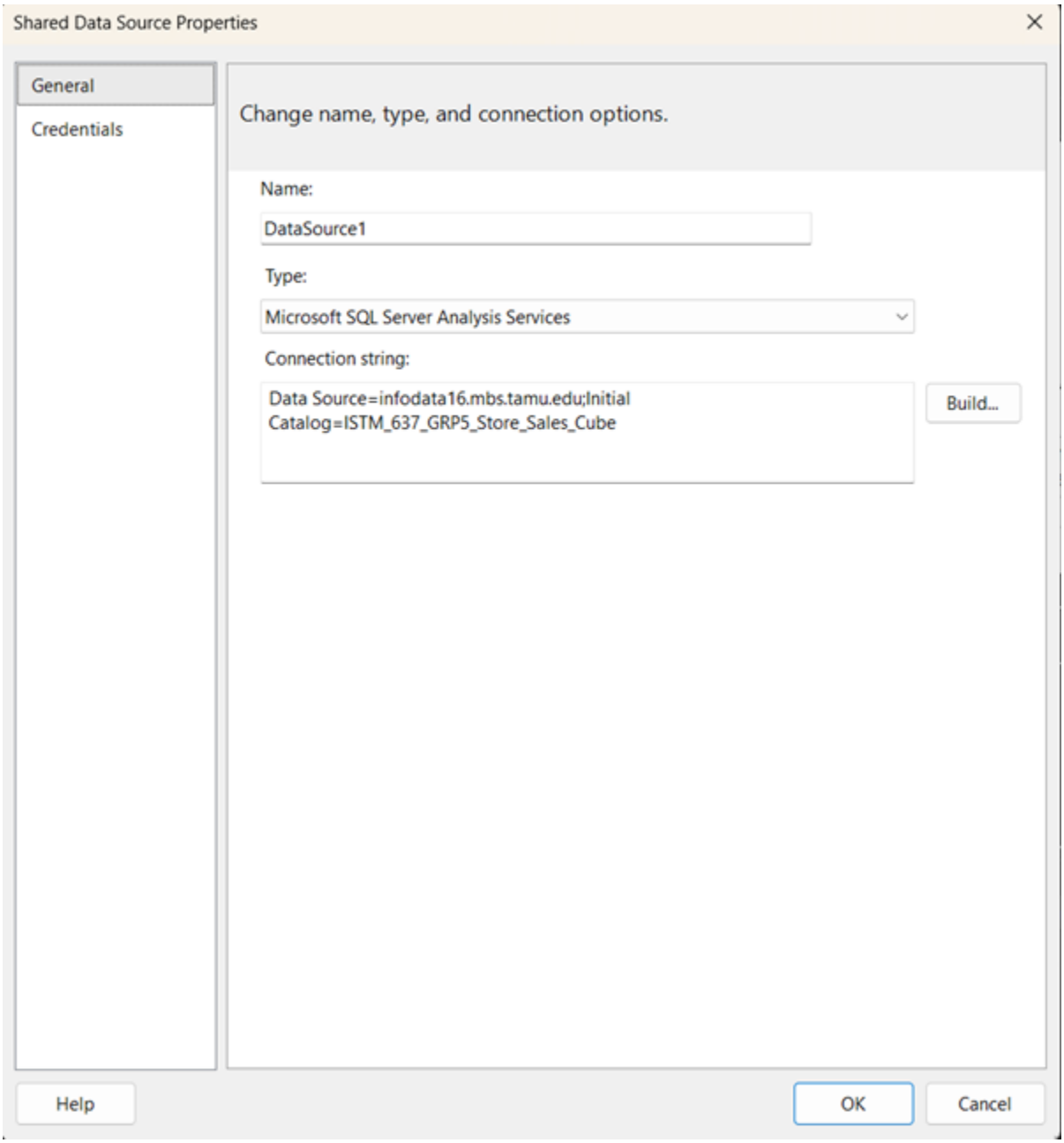
Year-on-Year Sales Trend

Year	YoY %Change in Grocery Sales	YoY %Change in Bakery Sales	YoY %Change in Beer Sales	YoY %Change in Wine Sales	YoY %Change in Spirit Sales	YoY %Change in Meat Sales	YoY %Change in Frozen Meat Sales	YoY %Change in Dairy Meat Sales
1989	0	0	0	0	0	0	0	0
1990	194.981536693078	193.622592552339	247.313345666662	153.746599786674	153.418073151849	202.506468952015	159.992791188584	193.561683663868
1991	-36.7935071573968	-38.6465509768551	-32.0002745321046	-34.1866167111353	-36.8477017370602	-37.9955460632591	-47.9038123185729	-39.2544631709205
1992	-0.375293843668129	8.96416049115506	3.26382507244938	15.3078712099257	11.364665619572	-1.33217437816994	3.66136761251448	4.63211818455984
1993	-2.6427248556867	2.50676766589714	-2.03434785053718	-2.20668605529267	-4.04826340219672	-4.3605643441943	5.38516671189072	-5.12797548032216
1994	-1.32835992835528	3.70699625859151	4.61932731912938	5.11391953461526	-2.5434793455905	-1.46735486043183	4.11455841823712	-0.404032352539247
1995	6.8447205609471	3.44805797167052	7.92664340149073	10.360774320312	6.72095243349484	-1.13124603140722	6.82650770698566	5.13517562099548
1996	-0.951804087546473	0.0778281212358258	2.51275448908999	6.08001702493107	-1.10552583889316	0.267681876503621	1.33138850877882	1.46476654760486
1997	-64.16909182755	-66.5878541944361	-69.4617465428715	-63.5597629156295	-68.4942184817468	-64.5717309145616	-59.5092391897792	-62.4066555078262

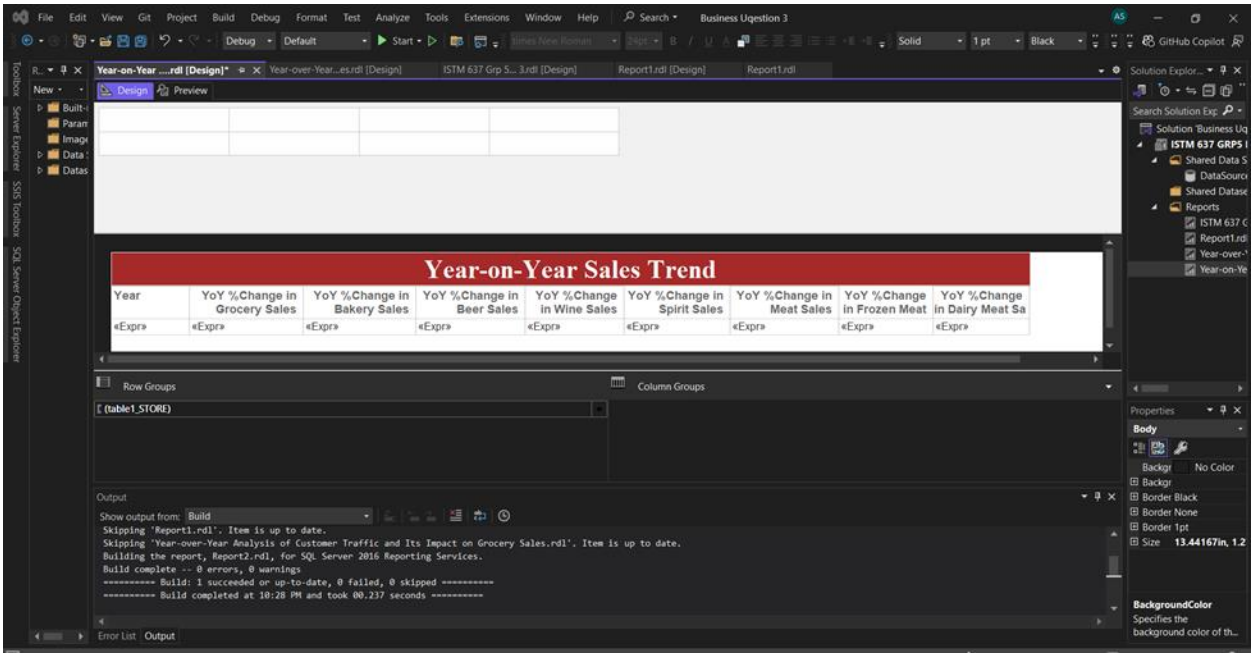
DFF is home to a collection of various products and categories. A smart retail store always assesses its revenue generating products and the products that are causing the business to lose money. By assessing the products that are facing a decline in sales, DFF can generate its future sales roadmap to focus only on products that help in catapulting the sales. The above visualization is an SSRS report derived from a SSAS cube. It displays the year-on-year sales trends for various products sold at DFF. Grocery sales took off in a spectacular way, but have faced declining sales ever since, taking a minimal hit in 1996 and continuing massively in 1997. Although bakery sales have not shot up significantly after 1991, yet there is a level of consistency in the sales trajectory. All the categories had a strong start in 1990 and faltered in 1991, experiencing a widespread decline in sales. 1995 saw most of these categories going into positive sales trends but experiencing fluctuations and fluctuation were faltering by the end of 1997. The graph helps us conclude that bakery is one among the saftest products in term of sales, as they do not face any major decline. On the other hand, meat sales and frozen meat sales should be investigated further to make a decision on these products, owing to their perishable nature.

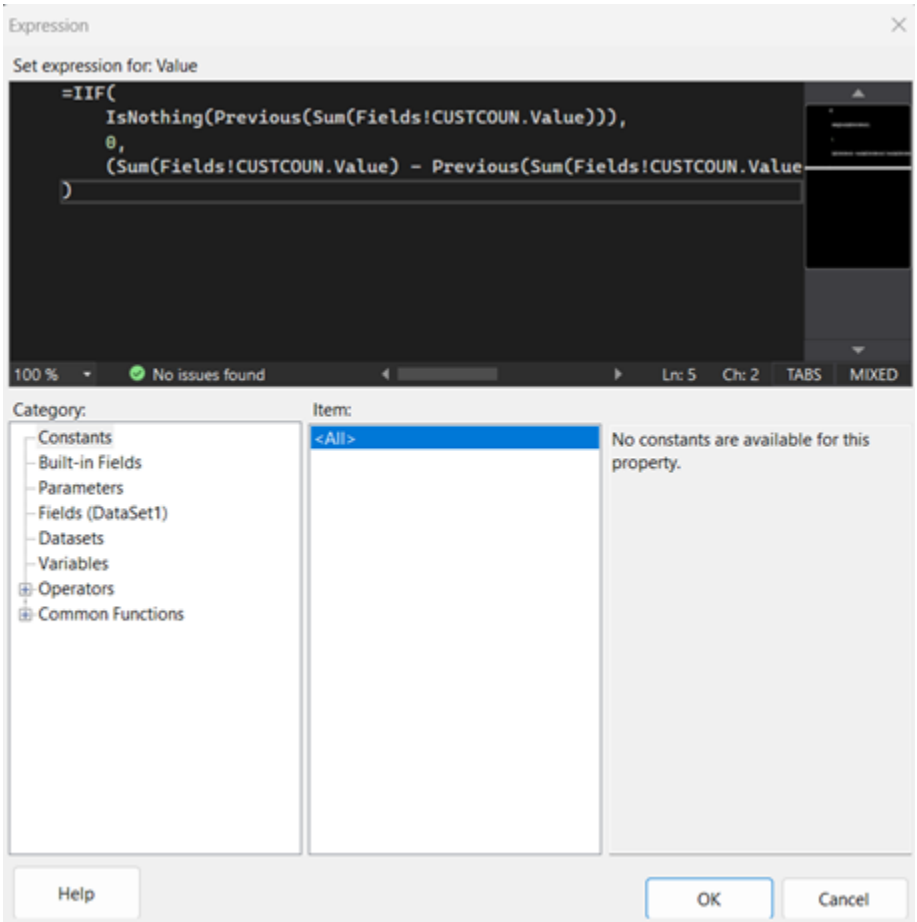
Implementation:

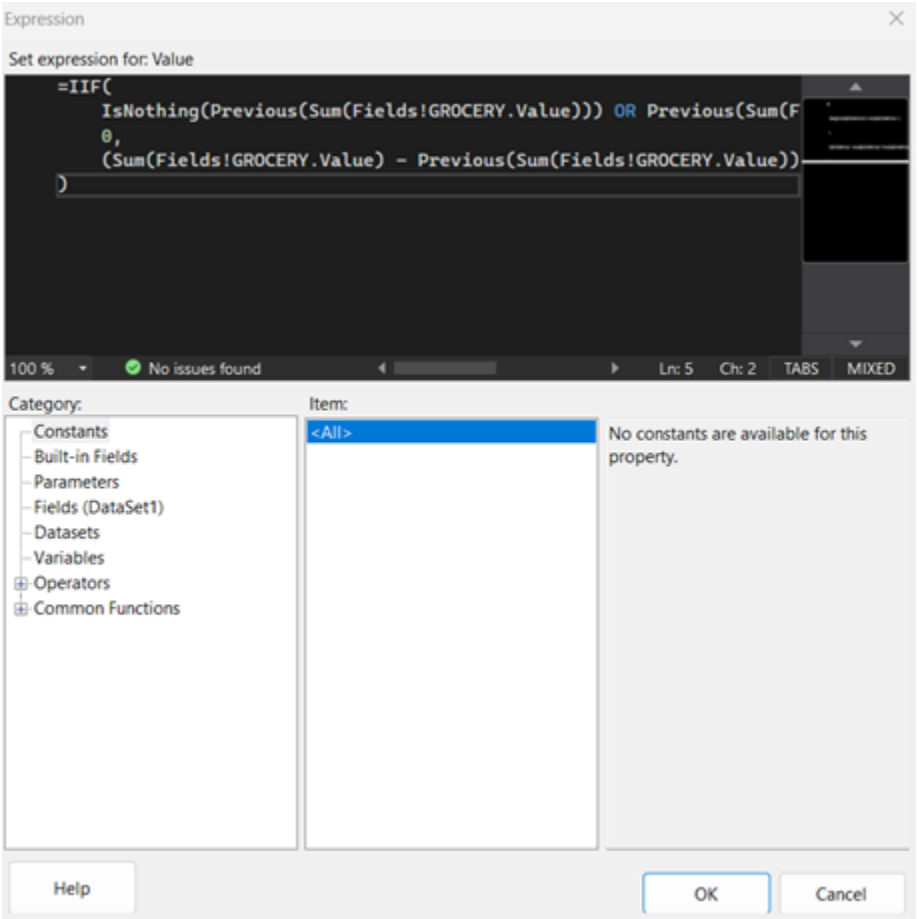
Establishing Cube Connection for Store level data:



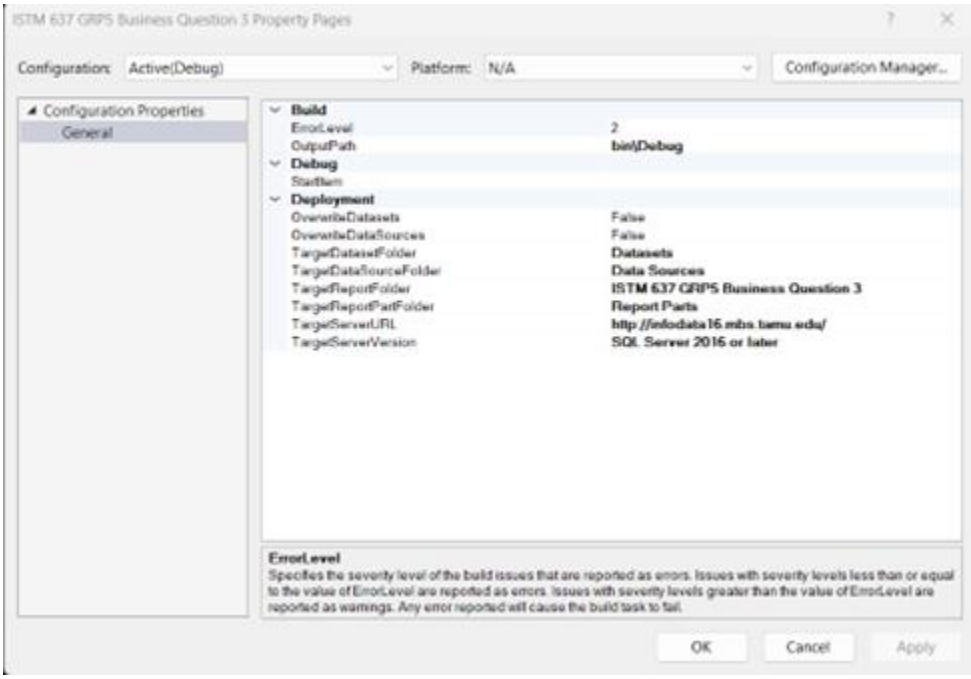
Designing the Report:

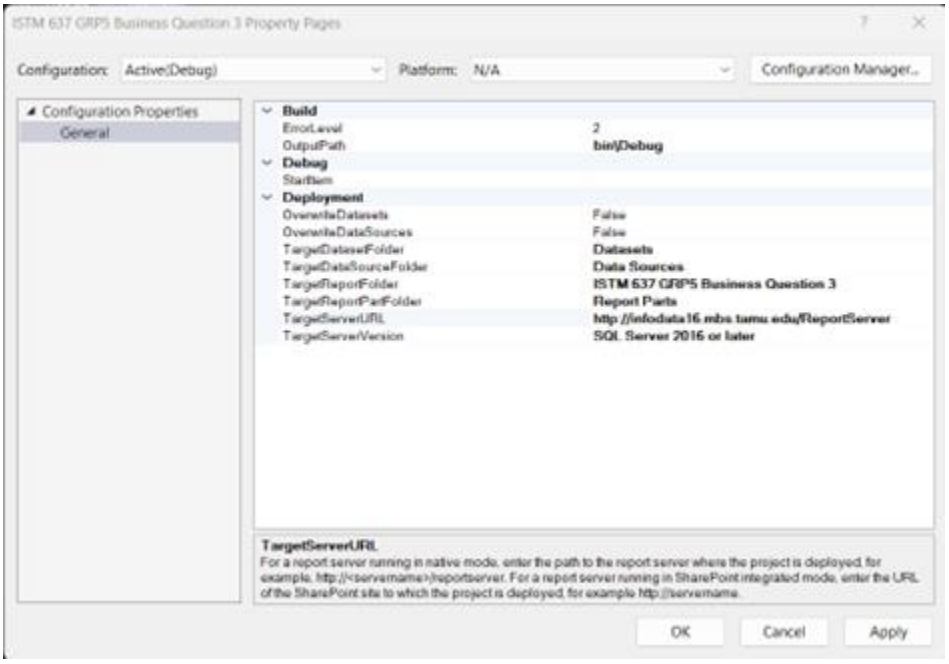


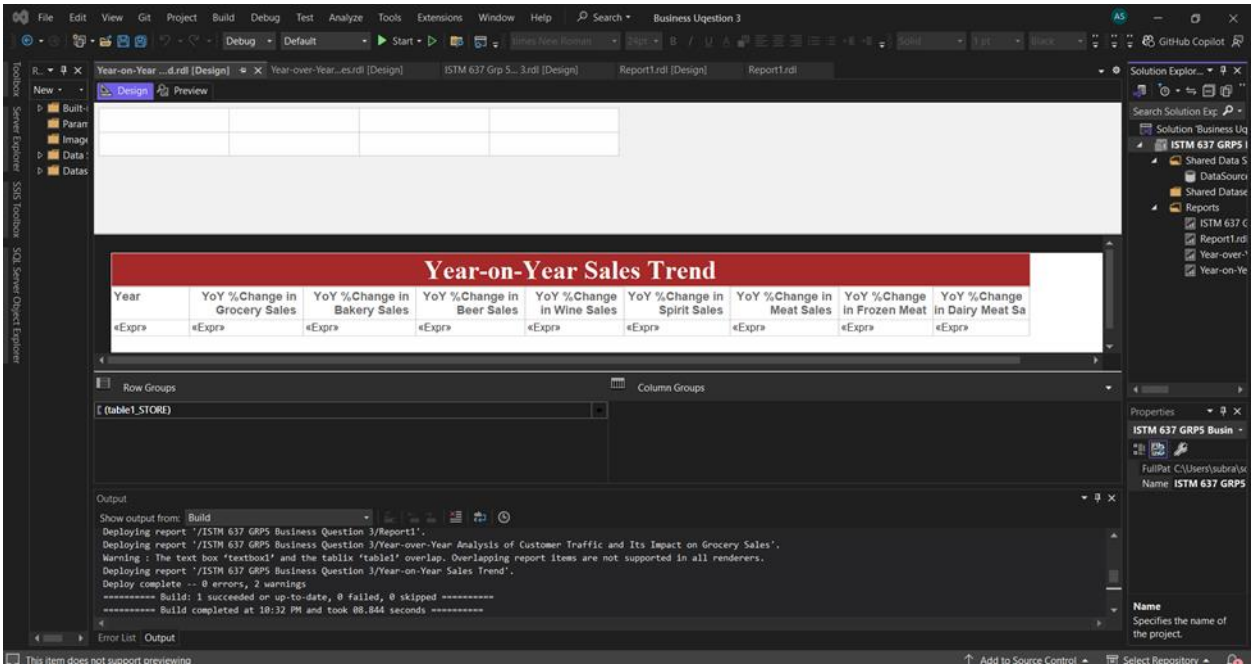


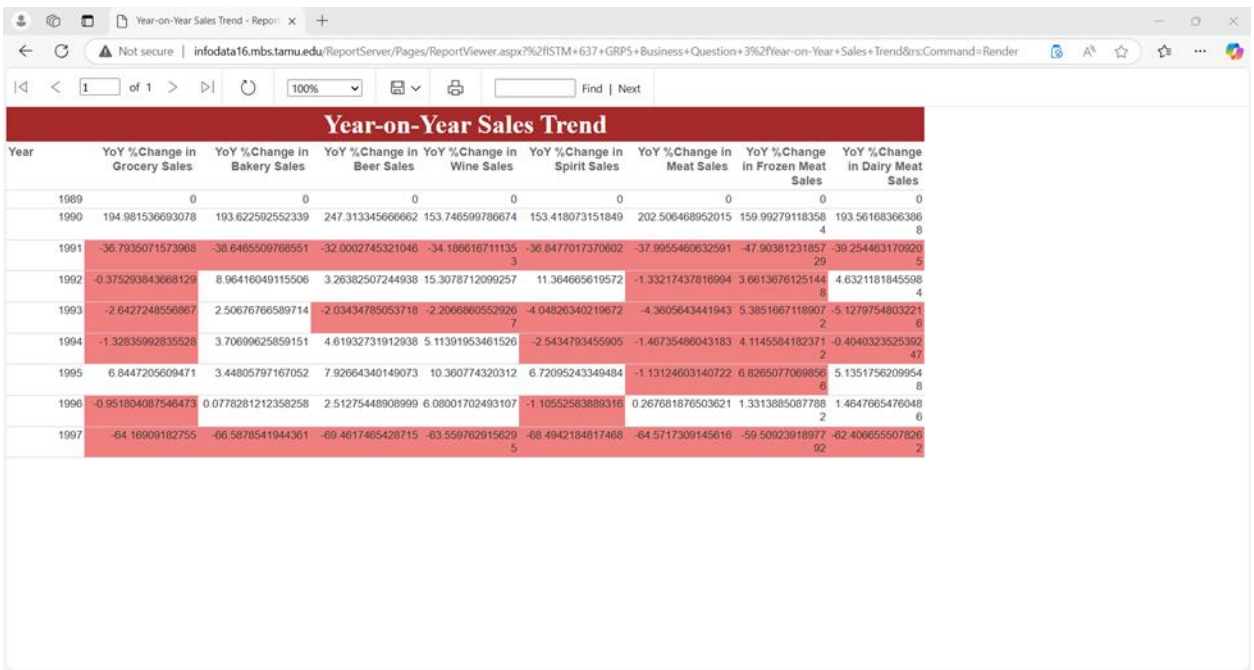

Expression
Set expression for: Value
=IIF(
IsNothing(Previous(Sum(Fields!GROCERY.Value))) OR Previous(Sum(F
0,
(Sum(Fields!GROCERY.Value) - Previous(Sum(Fields!GROCERY.Value))
)
100 %
No issues found
Ln: 5
Ch: 2
TABS
MIXED
Category:
Constants
Built-in Fields
Parameters
Fields (DataSet1)
Datasets
Variables
Operators
Common Functions
Item:
<All>
No constants are available for this property.
Help
OK
Cancel

Deploying the Report:









Year-on-Year Sales Trend

Year	YoY %Change in Grocery Sales	YoY %Change in Bakery Sales	YoY %Change in Beer Sales	YoY %Change in Wine Sales	YoY %Change in Spirit Sales	YoY %Change in Meat Sales	YoY %Change in Frozen Meat Sales	YoY %Change in Dairy Meat Sales
1989	0	0	0	0	0	0	0	0
1990	194.981536693078	193.622592552339	247.313345666662	153.746599786674	153.418073151849	202.506468952015	159.99279118358 4	193.56168366386 8
1991	-36.7935071573968	-38.6465509768551	-32.0002745321046	-34.186616711135 3	-36.8477017370602	-37.9955460632591	-47.90381231857 29	-39.254463170920 5
1992	-0.375293843668129	8.96416049115506	3.26382507244938	15.3078712099257	11.364665619572	-1.33217437816994	3.6613676125144 8	4.6321181845598 4
1993	-2.6427248556867	2.50676766589714	-2.03434785053718	-2.2066860552926 7	-4.04826340219672	-4.3605643441943	5.3851667118907 2	-5.1279754803221 6
1994	-1.32835992835528	3.70699625859151	4.61932731912938	5.11391953461526	-2.5434793455905	-1.46735486043183	4.1145584182371 2	-0.4040323525392 47
1995	6.8447205609471	3.44805797167052	7.92664340149073	10.360774320312	6.72095243349484	-1.13124603140722	6.8265077069856 6	5.1351756209954 8
1996	-0.951804087546473	0.0778281212358258	2.51275448908999	6.08001702493107	-1.10552583889316	0.267681876503621	1.3313885087788 2	1.4647665476048 6
1997	-64.16909182755	-66.5878541944361	-69.4617465428715	-63.559762915629 5	-68.4942184817468	-64.5717309145616	-59.50923918977 92	-62.406655507826 2

Business Question 5: Which beer brand attracts the highest coupon redemption and buyback offers?

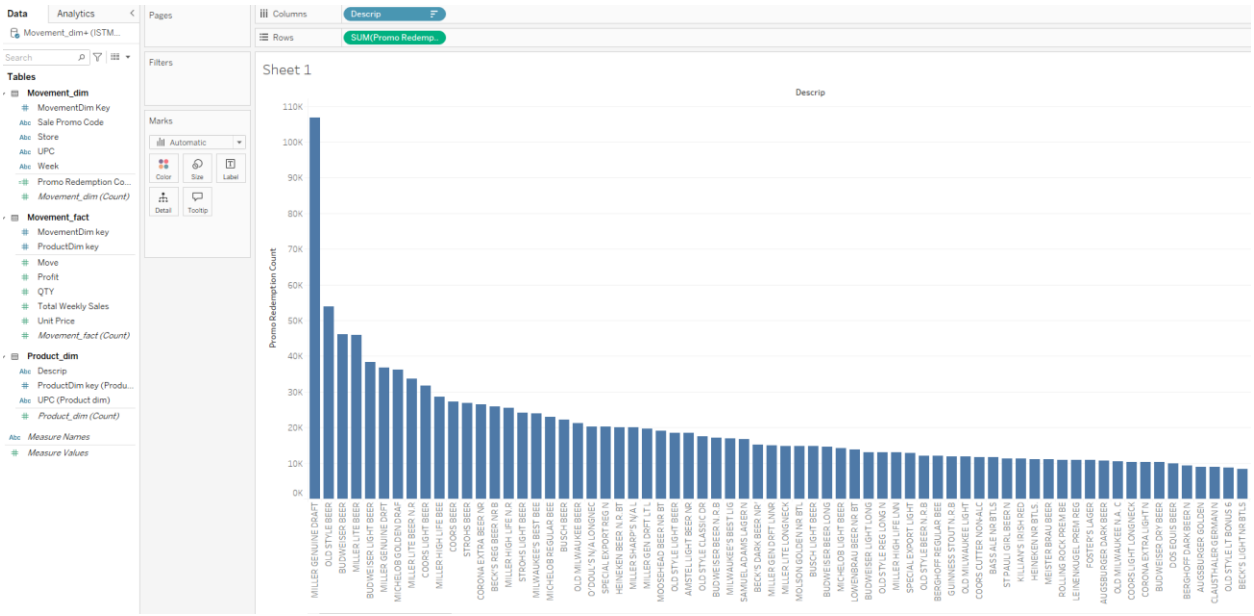
BI/Visualization Method: Tableau

Report Plan:

Tableau's robust data visualization capabilities allow for interactive and insightful exploration of the data. Tableau is ideal for quickly identifying patterns and trends, such as redemption rates and buyback activity, using its advanced visual analytics tools.

Tableau's ability to handle large datasets, integrate multiple data sources, and dynamically update visualizations ensures that users can drill down into specific brands or coupon types seamlessly. Its intuitive interface enables stakeholders to compare metrics like redemption rates across brands and observe correlations, making it an ideal choice for answering questions that require uncovering actionable insights through visual representation.

Report:

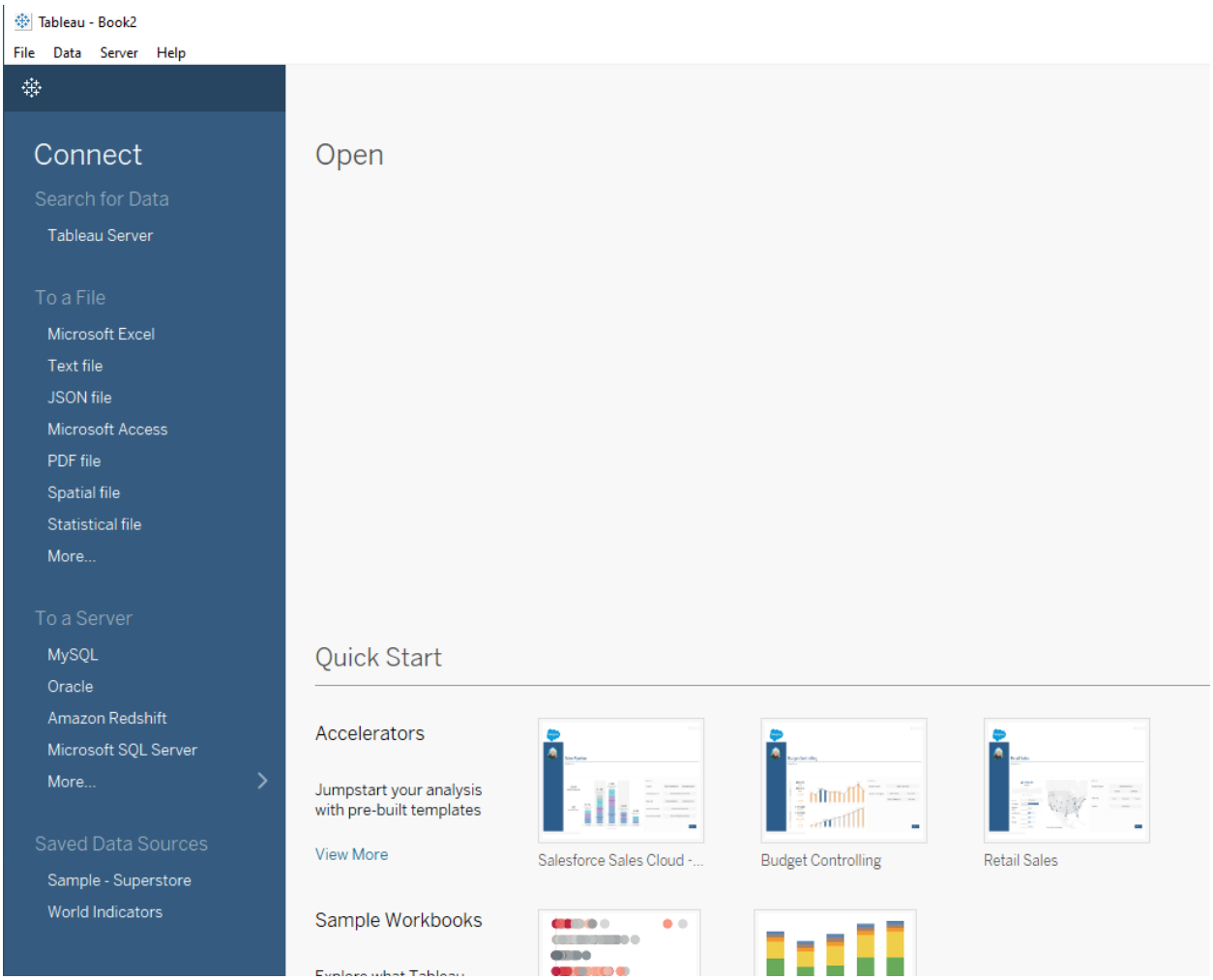


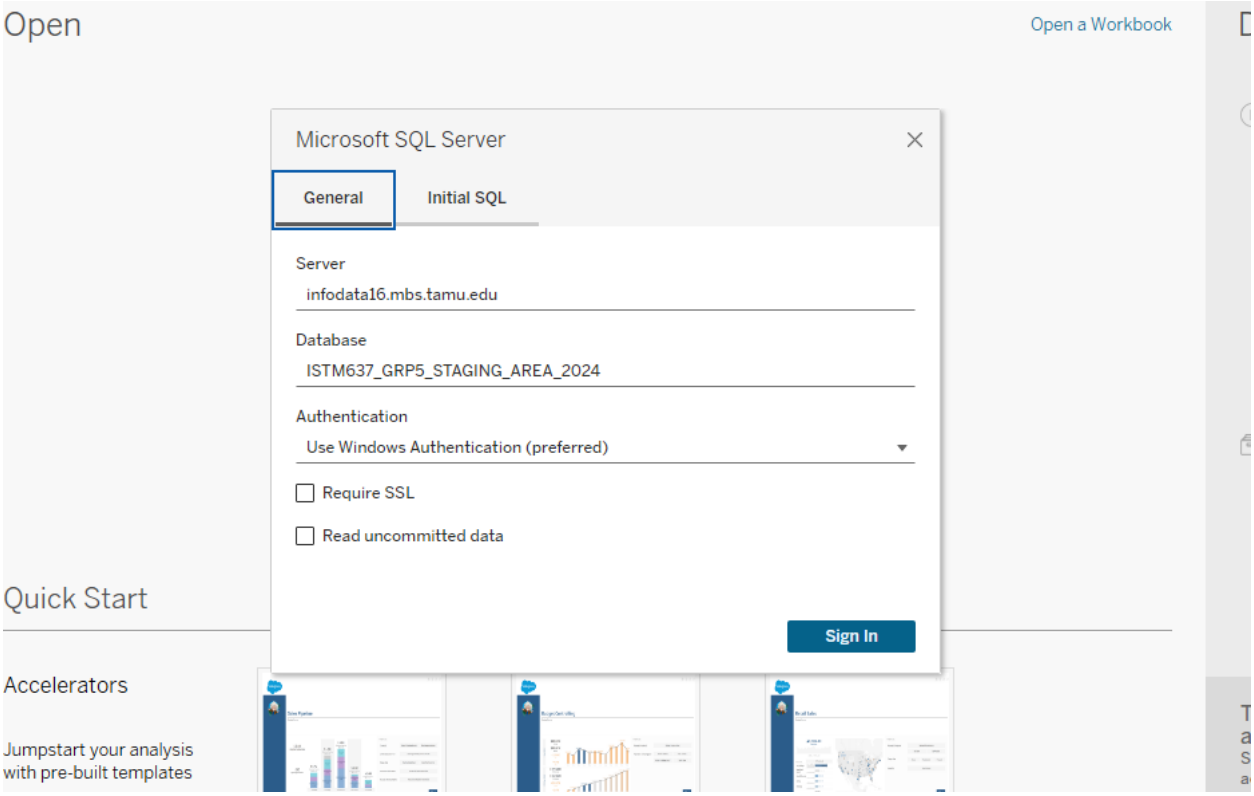
The analysis reveals that MILLER GENUINE DRAFT leads the pack with the highest coupon redemption and buyback offers. Following it in order of redemption and buyback activity are OLD STYLE BEER, BUDWEISER BEER, and other brands that follow in decreasing order of engagement.

This result highlights the popularity of MILLER GENUINE DRAFT in terms of consumer offers, indicating that it resonates most with the target market, driving higher coupon redemption and buyback participation. Brands like OLD STYLE BEER and BUDWEISER BEER also show strong performance, making them important competitors in the market for offers and promotions. This insight can be leveraged to refine promotional strategies and enhance customer engagement for these brands.

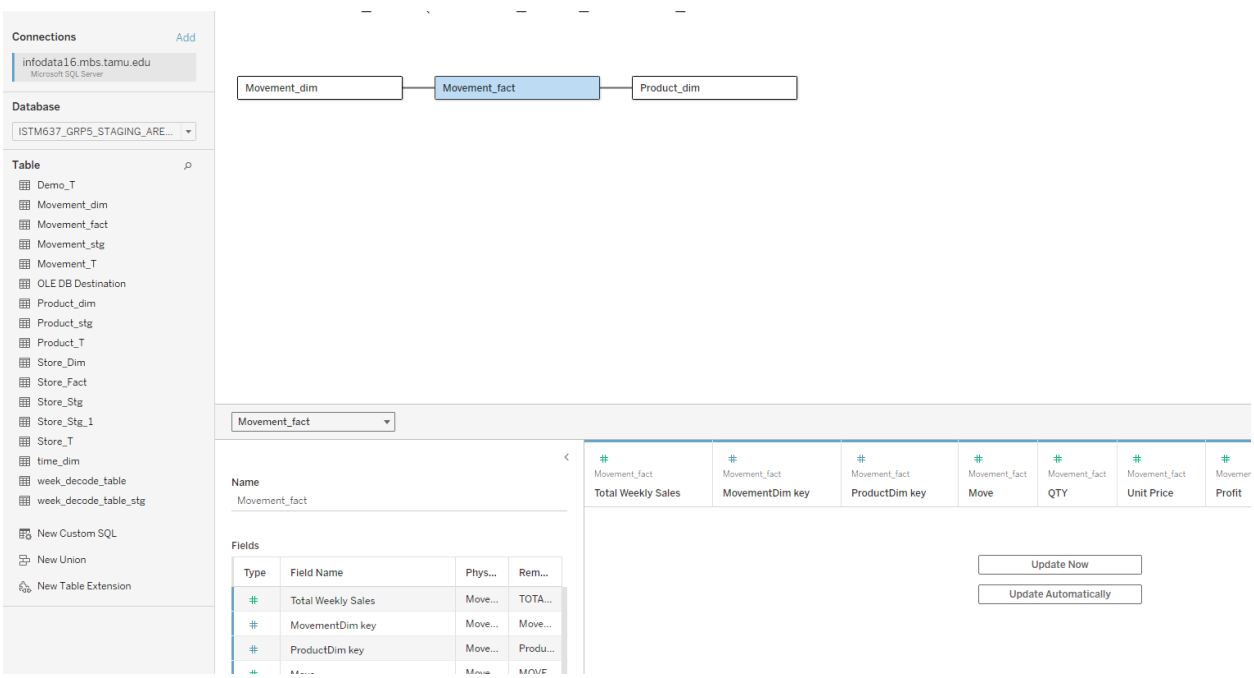
Implementation:

Connecting to the infodata16.mbs.tamu server from Tableau

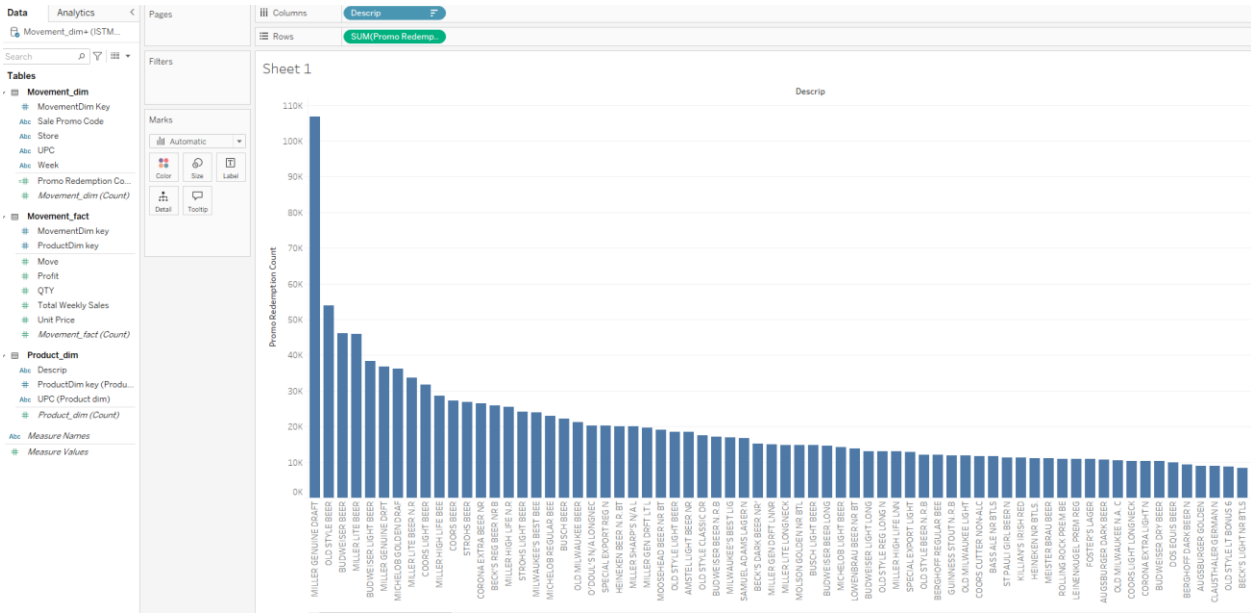




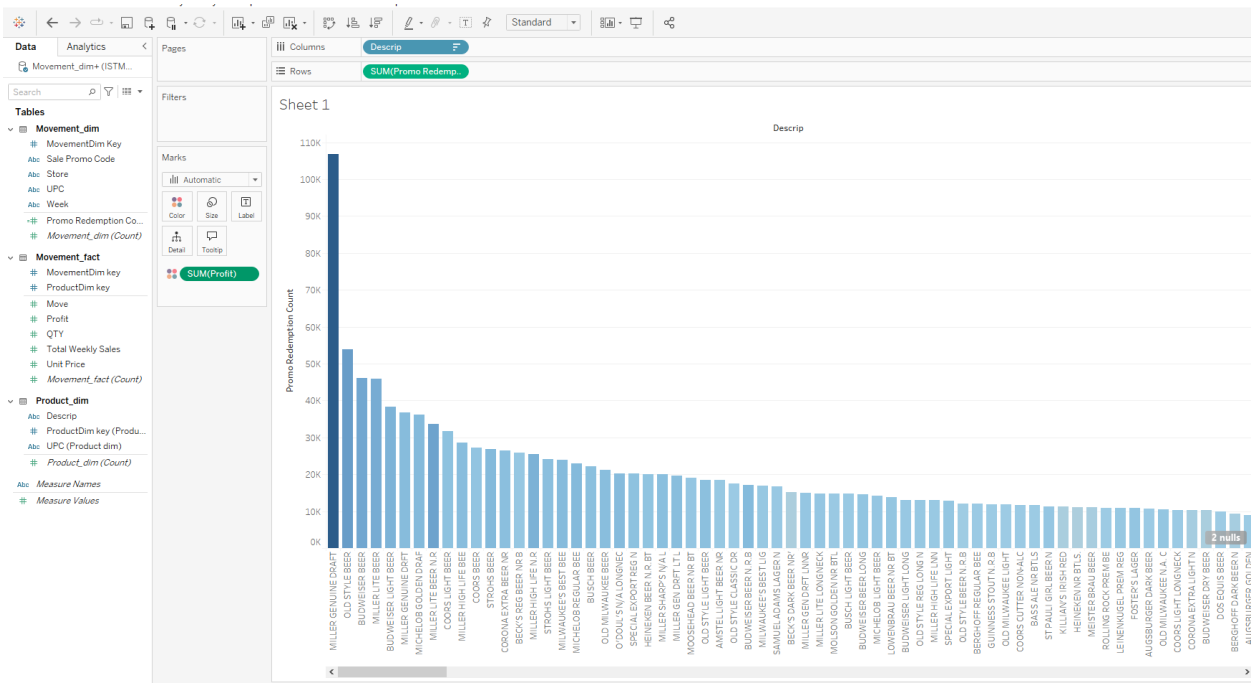
Defining relationship in Tableau for our Product Level Data Mart:



We created a calculated measure Promo Redemption Count that is the count of Sale promo code according to the beer brand category. We plotted the newly created calculated field across the beer brands to identify the top beer brands that attracts coupon redemptions and buyback offers.



We filtered the same analysis using profit to determine which brand gave us the highest profit with highest coupon redemption and buyback offer



References:

[1] - (Manchanda, P., Ansari, A., & Gupta, S., 1999, "The 'shopping basket': A model for multicategory purchase incidence decisions," Marketing Science, Vol. 18, No. 2, 1999) - Research Paper

[2] - (Ailawadi and Neslin, 1998, "The Effect of Promotion on Consumption: Buying More and Consuming It Faster," Journal of Marketing Research, Vol. 35, No. 3, 1998) - Research Paper

[3] - (Levy et al., 1997, "The Magnitude of Menu Costs: Direct Evidence from Large U.S. Supermarket Chains," Quarterly Journal of Economics, Vol. 112, No. 3, 1997) - Research Paper

[4] - Wikipedia - https://en.wikipedia.org/wiki/Kimball_lifecycle

[5] - DBT - https://docs.getdbt.com/blog/kimball-dimensional-model

[6] - Kimball - https://datascientest.com/en/kimball-method-what-is-it-how-do-i-use-it