TEXAS A&M UNIVERSITY
Mays Business School

Data Warehousing

Design and Implementation of a Data Warehouse for a Dominick's Finer Foods

By:

Sai Krishna Koppula

Table of Contents

Introduction

Dominick's Fine Foods (DFF) was a major grocery store chain primarily based in the Chicago metropolitan area. It was founded in 1918 by Dominick DiMatteo and grew over the decades into a large and popular supermarket chain. DFF became known for offering a range of fresh foods, groceries, and specialty products. Dominick's was also ahead of its time, pioneering innovative technologies like shopping carts with built-in computer screens that provided store directories, recipes, and advertisements—long before such tools became common in the market.

Over its nearly century-long history, DFF grew from a small family business into a major regional player, pioneering innovative grocery store formats. However, Dominick's underwent several acquisitions, notably by Fisher Foods and later Safeway Inc., which led to changes in store format and management. Despite many major efforts, Dominick's sales decreased which led to shut down of many stores and as time passed, the chain ultimately closed in 2014. (Reference: Wikipedia).

At its peak, DFF was known for its customer-friendly innovations, such as glass-fronted stores, exposed structural elements, and efficient POS systems. Dominick's Finer Foods entered into a partnership with Chicago Booth for store-level research into shelf management and pricing, conducting randomized experiments from 1989 to 1994 across 100 stores in Chicago.

The goal of this project is to design and develop a data warehouse using historical data to enable advanced data analysis and derive meaningful insights into retail operations and pricing strategies across Dominick's Fine Food (DFF) branches in the Chicago area. By leveraging OLAP (Online Analytical Processing) queries, the project aims to explore the retail data and address key business questions, ultimately enhancing marketing and sales strategies in this competitive market sector.

Understanding of the data

The collaboration between DFF & Chicago Booth produced a unique dataset that captures detailed information on more than 3,500 UPCs, including sales, retail margins, and shelf management across a wide variety of product categories. This dataset stands out due to its wide coverage and the depth of information it provides on retail margins.

This data set offers over seven years of store-level scanner data, organized into two main types of files: ***category-specific files***, which provide detailed data for individual product categories, and ***general files***, which include information relevant to all categories covered by the research. This unique dataset offers a rich source of insights for analyzing sales patterns, pricing, and shelf management practices across a broad spectrum of products.

Describe metadata for all the OLTP source files

Let us now explore the metadata structure of Dominick's dataset, and gain an overview of the key components and their relevance to the data analysis process.

Customer Count Files

The customer count file provides daily data on store traffic specific to each location. It records the number of customers who visited the store and made purchases. Additionally, this file includes total sales figures and the number of coupons redeemed, categorized by departments that the DFF defined.

Variable	Description	Type	Length	Variable	Description	Type	Length
DATE	Date of the Observation	Character	6	FTGITAL	Food-to-Go Italian Sales in Dollars	Numeric	8
Week	Week Number	Numeric	8	GM	General Merchandise Sales in Dollars	Numeric	8
Store	Store Code	Numeric	8	GMCOUP	General Coupons Redeemed	Numeric	8
BAKCOUP	Bakery Coupons Redeemed	Numeric	8	GROCCOUP	Grocery Coupons Redeemed	Numeric	8
BAKERY	Bakery Sales in Dollars	Numeric	8	GROCERY	Grocery Sales in Dollars	Numeric	8
BEER	Beer Sales in Dollars	Numeric	8	HABA	Health and Beauty Aids Sales in Dollars	Numeric	8
BOTTLE	Bottle Sales in Dollars	Numeric	8	HABACOUP	Health and Beauty Aids Coupons Redeemed	Numeric	8
BULK	Bulk Sales in Dollars	Numeric	8	JEWELRY	Jewelry Sales in Dollars	Numeric	8
BULKCOUP	Bulk Coupons Redeemed	Numeric	8	LIQCOUP	Liquor Coupons Redeemed	Numeric	8
CAMERA	Camera Sales in Dollars	Numeric	8	MANCOUP	Manufacturer Coupons Redeemed	Numeric	8
CHEESE	Cheese Sales in Dollars	Numeric	8	MEAT	Meat Sales in Dollars	Numeric	8
CONVFOOD	Conventional Foods Sales in Dollars	Numeric	8	MEATCOUP	Meat Coupons Redeemed	Numeric	8
COSMCOUP	Cosmetics Coupons Redeemed	Numeric	8	MEATFROZ	Meat-Frozen Sales in Dollars	Numeric	8
COSMETIC	Cosmetics Sales in Dollars	Numeric	8	MISCSCP	Misc. Coupons Redeemed	Numeric	8
CUSTCOUN	Customer Count	Numeric	8	MVPCLUB	MVP	Numeric	8
DAIRCOUP	Dairy Coupons Redeemed	Numeric	8	PHARCOUP	Pharmacy Coupons Redeemed	Numeric	8
DAIRY	Dairy Sales in Dollars	Numeric	8	PHARMACY	Pharmacy Sales in Dollars	Numeric	8
DELI	Deli Sales in Dollars	Numeric	8	PHOTCOUP	Photo Coupons Redeemed	Numeric	8
DELICOUP	Deli Coupons Redeemed	Numeric	8	PHOTOFIN	Photo	Numeric	8
DELIEXPR	Deli Express Sales in Dollars	Numeric	8	PRODCOUP	Produce Coupons Redeemed	Numeric	8
DELISELF	Deli Self Service Sales in Dollars	Numeric	8	PRODUCE	Produce Sales in Dollars	Numeric	8
FISH	Fish Sales in Dollars	Numeric	8	PROMCOUP	Promotion Coupons Redeemed	Numeric	8
FISHCOUP	Fish Coupons Redeemed	Numeric	8	PROMO	Promotion Sales in Dollars	Numeric	8
FLORAL	Floral Sales in Dollars	Numeric	8	SALADBAR	Salad Bar Sales in Dollars	Numeric	8
FLORCOUP	Floral Coupons Redeemed	Numeric	8	SALCOUP	Salad Coupons Redeemed	Numeric	8
FROZCOUP	Frozen Items Coupons Redeemed	Numeric	8	SPIRITS	Spirits Sales in Dollars	Numeric	8
FROZEN	Frozen Items Sales	Numeric	8	SSDELICP	Self Service Deli Sales in Dollars	Numeric	8
FTGCCOUP	Food-to-Go Coupons Redeemed	Numeric	8	VIDCOUP	Video Coupons Redeemed	Numeric	8
FTGCHIN	Food-to-Go Chinese Sales in Dollars	Numeric	8	VIDEO	Video Sales in Dollars	Numeric	8
FTGICOUP	Food-to-Go Coupons Redeemed	Numeric	8	VIDEOREN	Video Rentals (Dollar Amounts)	Numeric	8
				WINE	Wine Sales in Dollars	Numeric	8

Fig : Customer Count Metadata

Store-Specific Demographics

The demographics file contains information about the surrounding communities for all the stores. This data is obtained from the 1990 U.S. census for the Chicago metropolitan area.

Variable Name	Description	Variable Name	Description
age9	% Population under age 9	retired	% of Retired
age60	% Population over age 60	unemp	% of Unemployed
ethnic	% Blacks & Hispanics	wrkch5	% of working women with children under 5
educ	% College Graduates	wrkch17	% of working women with children 6 - 17
nocar	% With No Vehicles	nwrkch5	% of non-working women with children under 5
income	Log of Median Income	nwrkch17	% of non-working women with children 6 - 17
incsigma	Std dev of Income Distribution (Approximated)	wrkch	% of working women with children
hsizeavg	Average Household Size	nwrkch	% of non-working women with children
hsize1	% of households with 1 person	wrkwch	% of working women with children under 5
hsize2	% of households with 2 persons	wrkwnch	% of working women with no children
hsize34	% of households with 3 or 4 persons	telephn	% of households with telephones
hsize567	% of households with 5 or more persons	mortgage	% of households with mortgages
hh3plus	% of households with 3 or more persons	nwhite	% of population that is non-white
hh4plus	% of households with 4 or more persons	poverty	% of population with income under $15,000
hhsingle	% of households with 1 person	shopcons	% of Constrained Shoppers
hhlarge	% of households with 5 or more persons	shophurr	% of Hurried Shoppers
workwom	% Working Women with full-time jobs	shopavid	% of Avid Shoppers
sinhouse	% Detached Houses	shopstr	% of Shopping Stranges
density	Trading Area in Sq Miles per Capita	shopunft	% of Unfettered Shoppers
hval150	% of Households with Value over $150,000	shopbird	% of Shopper Birds
hval200	% of Households with Value over $200,000	shopindx	Ability to Shop (Car and Single Family House)
hvalmean	Mean Household Value (Approximated)	shpindx	Ability to Shop (Car and Single Family House)
single	% of Singles		

Fig : Store Specific Demographics Metadata

UPC Files

The UPC files list products by category, with the last five digits of the UPC identifying the product and the rest showing the manufacturer. DFF uses item codes to track products, but sometimes there are inconsistencies/errors. Special symbols in descriptions show if a product is on sale, a trial size, or discontinued.

Variable	Description	Type	Length
upc	UPC Number	Numeric	8
com_code	Dominick's Commodity Code	Numeric	8
nitem	Dominick's item code	Numeric	8
descrip	Product Name	Character	20
size	Product Size	Character	6
case	Number of items in a case	Numeric	8

Fig : UPC Metadata

Movement

The movement files track weekly sales for each product in a category at the store level, including price, quantity, and profit details.

Note: Promotions like discounts or coupons are noted, but the data isn't always consistently recorded. Some data may be flagged as unreliable and not be used in the analysis.

Variable	Description	Type	Length
upc	UPC Number	Numeric	8
store	Store Number	Numeric	3
week	Week Number	Numeric	3
move	Number of units sold	Numeric	8
price	Retail Price	Numeric	8
qty	Number of items bundled together	Numeric	3
profit	Gross margin	Numeric	8
sale	Sale code (B,C,S)	Character	8
ok	1 for valid data, 0 for trash	Numeric	3

Fig : Movement Metadata

Dominick's Stores

The store's date file identifies all the locations that were part of Dominick's research initiative.

Variable Name	Variable Description	Type
Store	Store Code	Numeric
City	Store located in the City	Character
Price Tier	Price tier of the Store	Character
Zone	Zone number of store	Numeric
Zip Code	Zip Code of the location	Zip Code
Address	Store Address	Character

Fig : Dominick Store Metadata

Weeks Decode Table

This file includes details about different special events occurring each week. It helps identify the dates of these events, making it valuable for analyzing sales patterns and associated trends.

Variable Name	Variable Description	Type
Week Number	Week Number	Numeric
Start Date	Start Date of the week	Character
End Date	End Date of the week	Character
Special Events	National Holiday Dates	Character

Fig : Week Decode Metadata

Entity Relationship Diagram (ERD)

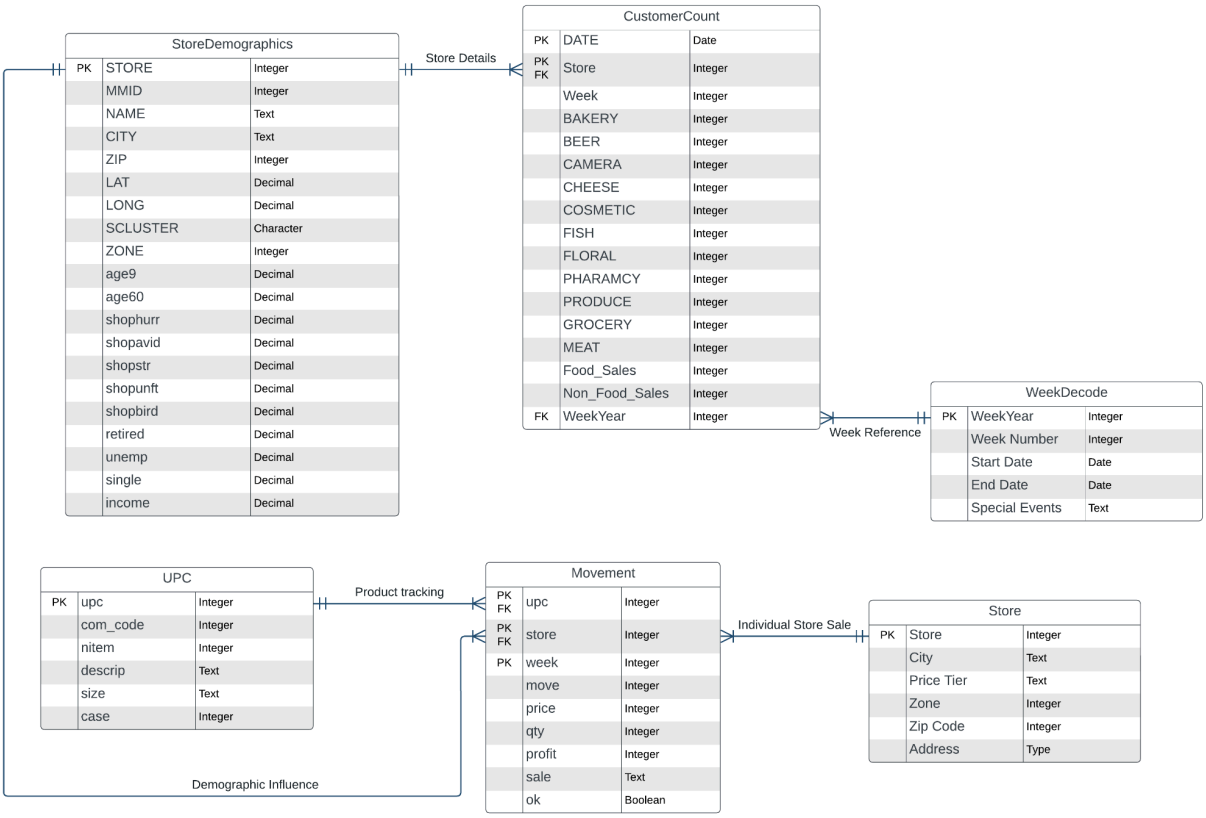


Fig: Entity Relationship Diagram between Tables

Subject Area Understanding

The retail industry is one of the competitive sectors which is always dynamic and plays a crucial role in the global economy. It encompasses a wide range of businesses, from large department stores to small local shops, each striving to meet the ever-evolving demands of consumers. In this fast-paced environment, understanding customer value, needs, and experiences is essential for retailers aiming to succeed and thrive. As consumer preferences shift and technological advancements reshape shopping behaviors, retailers must adapt their strategies to enhance customer satisfaction and loyalty while navigating the challenges posed by data management, pricing strategies, and demand forecasting [4].

Understanding customer value, needs, and experiences is a key factor in shaping business decisions and operations within the retail industry. According to Lemon et al., this understanding encompasses various dynamics related to customer experience, including buying behavior, customer satisfaction, loyalty, and relationship management. By tailoring product offerings to meet specific customer preferences, retailers can enhance satisfaction and encourage repeat purchases. The observation of customer purchases and preferences can help to identify pain points, creating a seamless shopping experience. Additionally, data-driven insights empower retailers to make informed decisions regarding inventory, pricing, and emerging trends, which fosters trust and stronger relationships that drive customer loyalty [5] .

However, offline retailers face significant challenges in effectively leveraging data. The maintenance of large datasets to comprehend the relationships between sales, product availability, and customer purchases is inherently difficult. The overwhelming volume of data can hide critical insights, complicating the implementation of effective predictive models that accurately reflect customer preferences and lifetime value. Privacy and security concerns further complicate matters, as consumers increasingly demand transparency and protection of their personal information. Retailers must navigate the delicate balance between utilizing data for marketing and maintaining customer trust [6].

Dynamic pricing strategies also pose challenges in the marketing, sales, and retail sectors. Frequent price changes can lead to confused and dissatisfied customers, undermining trust and brand loyalty. Accurate inventory and demand forecasting are crucial to avoid stockouts or excess inventory; however, limited data availability often hinders insights into customer preferences. Seasonal fluctuations can result in unpredictable demand spikes or drops, while poor communication between departments may cause misalignment in demand management. Additionally, rapid shifts in consumer preferences can render historical data less relevant. Addressing these challenges is essential for fostering strong customer relationships and ensuring a positive customer experience in an increasingly competitive marketplace [7].

10 Business Questions:

1. What is the proportion of sales made with coupons vs. without coupons in a given category?

   Rationale: Understanding the proportion of sales made with coupons vs. without coupons is important for companies because it provides insights into customer behavior and the effectiveness of promotions. It helps assess whether discounts drive sales and how much they impact profit margins. Companies can use this information to manage inventory, plan supply, and limit discount-based sales to maximize profit. It also aids in segmenting customers for targeted marketing, adjusting pricing strategies, and benchmarking against competitors. Overall, it helps companies optimize promotions and make informed decisions about pricing and customer engagement.

   The line chart presents an overall comparison of coupon-driven sales as a percentage of total sales across all stores, with filters applied for the year and specific months. This chart provides a comprehensive view of coupon usage trends over Q3 of 1996, highlighting variations in coupon percentages across different stores. Following this chart, the pivot table would display the coupon sales data segmented by store, allowing for detailed analysis of coupon versus non-coupon sales within each store, with filters applied for the year and specific months.

Fig: Coupon %of Sales w.r.t Total Sales of a given store

Store	Total Grocery Sales	Sum of Absolute Coupon Sales	Sum of Coupon % of Total Sales
2	1239719.32	7077.29	0.6%
8	2621730.44	24293.85	0.9%
9	1850709.83	15803.5	0.9%
12	1941118.34	9523.72	0.5%
14	1644757.16	8720.07	0.5%
18	2283402.4	17204.34	0.8%
21	1307143.32	11727.55	0.9%
28	1266490.73	8953.68	0.7%
32	2663881.21	17598.76	0.7%
33	1612977.63	6025.25	0.4%
40	1272804.57	15117.16	1.2%
44	2570571.14	16202.86	0.6%
47	1481271.1	13703.54	0.9%
51	1433005.92	12530.52	0.9%
52	2118797.83	11949.46	0.6%
53	1219840.23	8291.96	0.7%
54	1403137.25	23895.45	1.7%
56	1738369.28	13719.13	0.8%
59	1683681.92	14345.49	0.9%
62	2675460.36	6880.92	0.3%
67	1438167.69	13280.32	0.9%
68	1446699.36	7956.31	0.5%
70	1996326.46	20446.2	1.0%
71	2030953.27	33048.15	1.6%
72	1228473.6	8912.69	0.7%
73	1870439.05	19294.47	1.0%
74	2338850.96	23468.78	1.0%
75	1442780.45	5440.06	0.4%
76	1405449.3	10293.66	0.7%
77	2100731.44	14828.05	0.7%
78	1985719.95	19735.41	1.0%
80	2074580.75	16997.27	0.8%
81	1961765	14020.31	0.7%
83	1780286.01	16203.61	0.9%
84	2112160.81	18965.39	0.9%
86	2569316.41	15181.66	0.6%
88	1341403.57	11785.18	0.9%
89	1077088.77	7483.02	0.7%
90	1106916.07	7203.7	0.7%
91	1320590.11	12651.62	1.0%
92	511680.15	6545.12	1.3%
93	1588409.17	8224.61	0.5%
94	1489300.21	9832.13	0.7%
95	1523346.21	8526.93	0.6%
98	2225654.17	24459.93	1.1%

Fig: Sample Pivot table for Coupon % sales of Total Sales (1)

100	2103491.04	16718.38	0.8%
101	2010373.87	16083.25	0.8%
102	2684720.64	22270.4	0.8%
103	1331655.22	12746.74	1.0%
104	2231890.69	19571.98	0.9%
105	1264209.29	10491.8	0.8%
106	962149.48	8664.99	0.9%
107	2438674.01	18563.64	0.8%
109	2863091.63	10263.4	0.4%
110	1415661.75	11339.25	0.8%
111	1451881.51	10377.23	0.7%
112	2231938.34	13484.47	0.6%
113	2018428.58	9284.14	0.5%
114	1755179.3	13258.52	0.8%
115	2345176.39	54048.1	2.3%
116	1384341.05	13242.45	1.0%
119	1568116.69	9189.17	0.6%
121	2557839.32	16728.12	0.7%
122	2340486.64	13906.88	0.6%
123	1507424.33	13399.48	0.9%
124	1677332.28	11865.61	0.7%
126	2966633.3	63181.8	2.1%
128	2083877.47	15248.81	0.7%
129	2452726.27	13915.61	0.6%
130	1861246.71	13036.47	0.7%
131	1706219.84	14242.2	0.8%
132	2435177.02	24817.05	1.0%
133	2079160.32	16688.3	0.8%
134	1508635.81	8035.91	0.5%
135	2447580.47	37295.52	1.5%
136	1860158.62	6114.01	0.3%
137	3214542.41	11853.39	0.4%
139	2651424.01	15837.6	0.6%
140	755991.26	36777.78	4.9%
141	564102.36	63083.94	11.2%
142	923413.58	44782.49	4.8%
143	2488224.28	62245.57	2.5%
144	2303403.76	47960.04	2.1%
301	5466493.44	226531.79	4.1%
302	4337407.01	150154.29	3.5%
303	4339492.96	232915.46	5.4%
304	6094112.78	137458.8	2.3%
305	3856196.16	136509.02	3.5%
306	4820040.51	174403	3.6%
307	4976056.65	204724.95	4.1%
308	4329251.82	157720.73	3.6%
309	4452675.17	147629.97	3.3%
310	4249036.27	157862.16	3.7%

Fig: Sample Pivot table for Coupon % sales of Total (2)

2. Determine the distribution of the type of shoppers present across various zones?

Rationale: Determining the distribution of different types of shoppers across various zones is important for companies because it helps them tailor marketing efforts, adjust product offerings, and optimize store locations based on shopper preferences. It enables better inventory management by aligning stock levels with local demand, and enhances customer experiences through personalized services and promotions. This insight also supports strategic expansion decisions and helps companies stay competitive by identifying new opportunities. Overall, it allows businesses to target the right customers, increase sales, and improve profitability.

The bar chart displays the distribution of shopper types across various zones, filtered by store number. Each color represents a shopper type, with the x-axis showing percentages and the y-axis listing stores. This visualization highlights different shopper profiles in each store, enabling comparative analysis of shopper segmentation by zone.

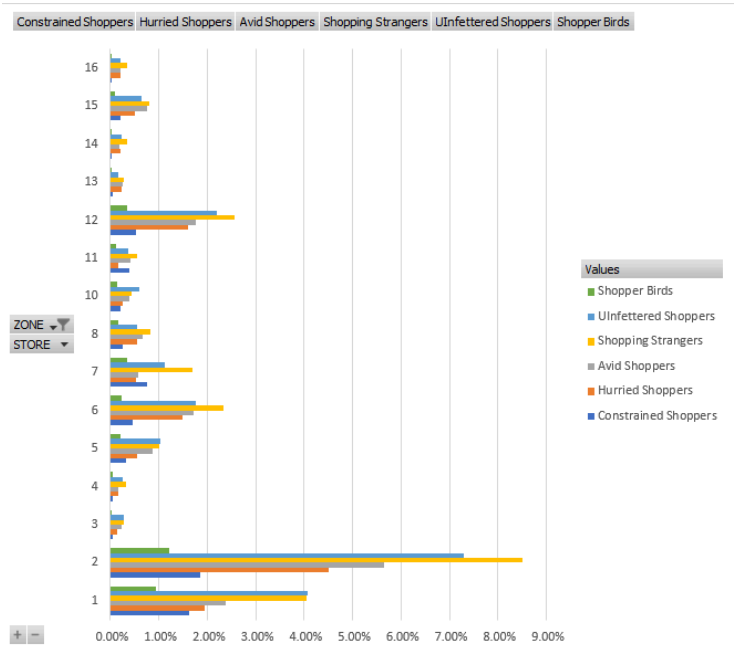


Fig: Distribution of type of shoppers present across various zones

Zone	Constrained Shoppers	Hurried Shoppers	Avid Shoppers	Shopping Strangers	UInfettered Shoppers	Shopper Birds
1	1.62%	1.94%	2.37%	4.04%	4.07%	0.95%
2	1.85%	4.51%	5.64%	8.50%	7.30%	1.21%
3	0.06%	0.14%	0.23%	0.27%	0.27%	0.03%
4	0.05%	0.16%	0.15%	0.32%	0.26%	0.06%
5	0.32%	0.56%	0.87%	1.01%	1.03%	0.21%
6	0.47%	1.49%	1.72%	2.32%	1.77%	0.22%
7	0.75%	0.53%	0.57%	1.69%	1.11%	0.34%
8	0.24%	0.55%	0.67%	0.83%	0.55%	0.16%
10	0.20%	0.25%	0.38%	0.43%	0.60%	0.14%
11	0.39%	0.17%	0.40%	0.54%	0.37%	0.12%
12	0.53%	1.60%	1.75%	2.57%	2.21%	0.34%
13	0.04%	0.24%	0.26%	0.28%	0.16%	0.02%
14	0.03%	0.21%	0.18%	0.34%	0.23%	0.01%
15	0.21%	0.51%	0.76%	0.79%	0.63%	0.09%
16	0.02%	0.21%	0.21%	0.35%	0.20%	0.01%
Grand Total	6.79%	13.07%	16.19%	24.29%	20.78%	3.89%

Fig: Pivot Table for distribution of type of shoppers present across various zones

3. Observe the sales of a given category in the respective week of every year?

   Rationale: Observing the sales of a given category in the same week each year is important for companies because it helps them identify trends and seasonal patterns, allowing for better forecasting and planning. By analyzing year-over-year sales, businesses can adjust inventory levels, optimize pricing, and plan marketing campaigns around peak sales periods. It also helps evaluate the impact of promotions or external factors, like holidays or events, on sales performance. Overall, this insight enables companies to make more informed decisions, improve sales strategies, and ensure they are prepared to meet customer demand.

   The line chart shows weekly beer sales across multiple years, with the x-axis representing the week of the year and the y-axis indicating total sales. Each line color represents a different year, allowing for a year-over-year comparison of beer sales trends across specific weeks. Peaks are visible in certain weeks, highlighting periods of increased sales that likely coincide with seasonal events or promotions. This chart enables an analysis of sales patterns over time to identify consistent trends and variations across years.

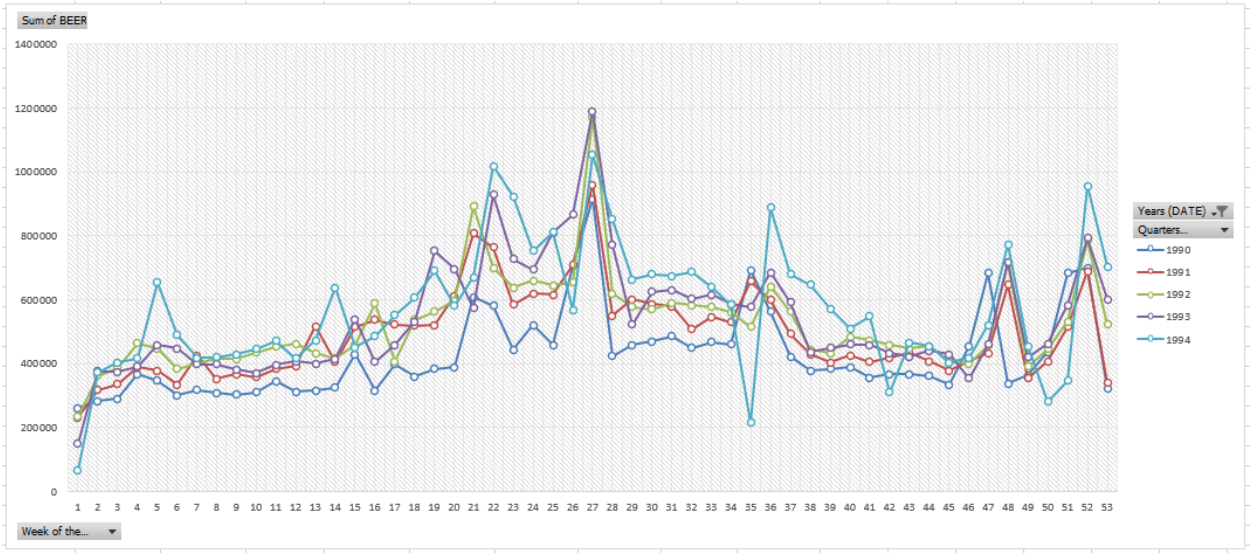


Fig: Sales of Beer Category in different stores over the years 1990 - 1994

Sum of BEER	Column Labels					
	1990	1991	1992	1993	1994	Grand Total
Row Labels						
1	260839	231351.71	232176.89	147666.72	64972.3	937006.62
2	281544.2	316569.31	360433.2	375796.65	371200.19	1705543.55
3	289109.41	334624.14	387024.24	374382.3	403570.59	1788710.68
4	365922.39	388938.64	463464.62	389762.34	416496.46	2024584.45
5	345611.31	378197.58	447810.67	457722.23	653063.68	2282405.47
6	299565.94	333067.36	382643.12	447631.22	488605.42	1951513.06
7	316938.48	424868.32	401766.01	398035.11	417082.71	1958690.63
8	307457.6	351605.15	415141.85	396695.22	418684	1889583.82
9	301544.73	366914.19	413822.55	380272.28	427890.93	1890444.68
10	309347.66	356513.1	434243.37	370759.23	445362.26	1916225.62
11	344171.51	383392.19	451679.03	395128.62	471084.37	2045455.72
12	310846.06	391904	462236.35	407460.49	415648.21	1988095.11
13	315476.4	515705.54	432223.29	399468.14	473378.79	2136252.16
14	323525.89	405561	415442.38	414313.21	637750.35	2196592.83
15	428465.09	514209.86	458443.79	537740.62	450022.9	2388882.26
16	315723.14	536467.78	587020.39	404565.65	484662.09	2328439.05
17	398173.33	521236.01	406080.95	457923.39	550884.88	2334298.56
18	358716.22	518474.65	535985.08	530442.68	605878.7	2549497.33
19	383462.74	520175.53	561607.53	752699.12	689627.21	2907572.13
20	388695.58	611985.67	596532.98	695820.02	579860.81	2872895.06
21	607710.19	808367.99	890969.67	573949.65	669919.97	3550917.47
22	581458.76	763093.86	700254.38	928096.61	1018683.29	3991586.9
23	443504.53	584827.3	636920.69	727145.82	922763.95	3315162.29
24	519752.37	619799.77	659747.25	694174.03	753580.14	3247053.56
25	457284.73	615827.48	645052.17	808686.55	812032.11	3338883.04
26	705200.99	709201.13	654045.58	866631.3	567545.35	3502624.35
27	916206.16	959844.49	1171206.56	1191154.16	1053315.36	5291726.73
28	423707.19	549200.83	618284.7	770906.44	851448.12	3213547.28
29	457459.95	599924.87	576295.86	521091.4	662736.64	2817508.72
30	469065.15	585401.82	569631.04	624778.21	679980.98	2928857.2
31	486145.74	578932.83	589677.67	630382.92	673359.57	2958498.73
32	449176.02	506873.37	581852.2	601964.48	687345.22	2827211.29
33	467611.51	546393.83	576420.84	615969.71	638378.1	2844773.99
34	459214.77	528462.68	559748.48	586158.64	584400.06	2717984.63
35	691556.41	656688.6	515172.36	576477.69	214820.11	2654715.17
36	562820.82	599339.42	641542.87	684775.28	888466.87	3376945.26
37	421246.76	492877.75	561523.22	592424.3	679014.14	2747086.17
38	377240.95	429504.34	442959.06	436811.28	645645.34	2332160.97
39	382808.95	403705.97	432691.04	449205.61	570725.84	2239137.41
40	389170.17	424800.11	483045.72	459201.93	508749.74	2264967.67

Fig: Pivot Table for Sales of Beer Category in different stores over the years 1990 - 1994 (1)

40	389170.17	424800.11	483045.72	459201.93	508749.74	2264967.67
41	354812.1	404165.23	472480.66	458786.78	547515.91	2237760.68
42	367146.33	418438.96	457723.97	432594.92	310051.42	1985955.6
43	365511.64	434593.68	446979.4	422178.54	464638.54	2133901.8
44	361183.58	405784.75	453472.47	440450.21	453939.27	2114830.28
45	333606.56	376418.98	403144.73	428475.59	402794.66	1944440.52
46	453817.68	408299.45	399566.06	353318.72	417136.53	2032138.44
47	683372.26	432260.25	456627.51	459361.77	520302.81	2551924.6
48	337963.24	647098.38	712385.29	718159.32	772426.21	3188032.44
49	362391.98	352761.73	389654.06	419046.86	454401.27	1978255.9
50	435368.46	405823.67	445981.57	462444.94	280667.45	2030286.09
51	682273.15	515528.18	531467.38	582309.74	346847.98	2658426.43
52	700091.95	687920.69	784922.67	793547.71	956791.96	3923274.98
53	319890.58	339216.96	522513.17	599972.84	703658.7	2485252.25
Grand Total	**22670908.31**	**26193141.08**	**27829738.59**	**28644919.19**	**30179810.46**	**135518517.6**

Fig: Pivot Table for Sales of Beer Category in different stores over the years 1990 - 1994 (2)

4. How do food and non-food product sales compare across all stores, segmented by year and week number?

   Rationale: Analyzing the sales performance of food categories (e.g., grocery, fish, meat, deli, produce) and non-food categories (e.g., video, cameras, photos, pharmacy, health and beauty) provides valuable insights into consumer preferences. This helps DFF adjust inventory levels for top-selling items and experiment with different product mix strategies. Ultimately, it allows DFF to better meet customer needs, boost sales, and foster customer loyalty.

   The following first Bar Chart presents an overall comparison of food vs. non-food sales across all stores with filters applied for year and specific week, while the second Bar Chart provides a filtered view, highlighting a selection of stores for enhanced visibility and focused analysis. Following these charts, the pivot table displays the sales data for stores, segmented into food and non-food categories, with filters applied for year and specific week.

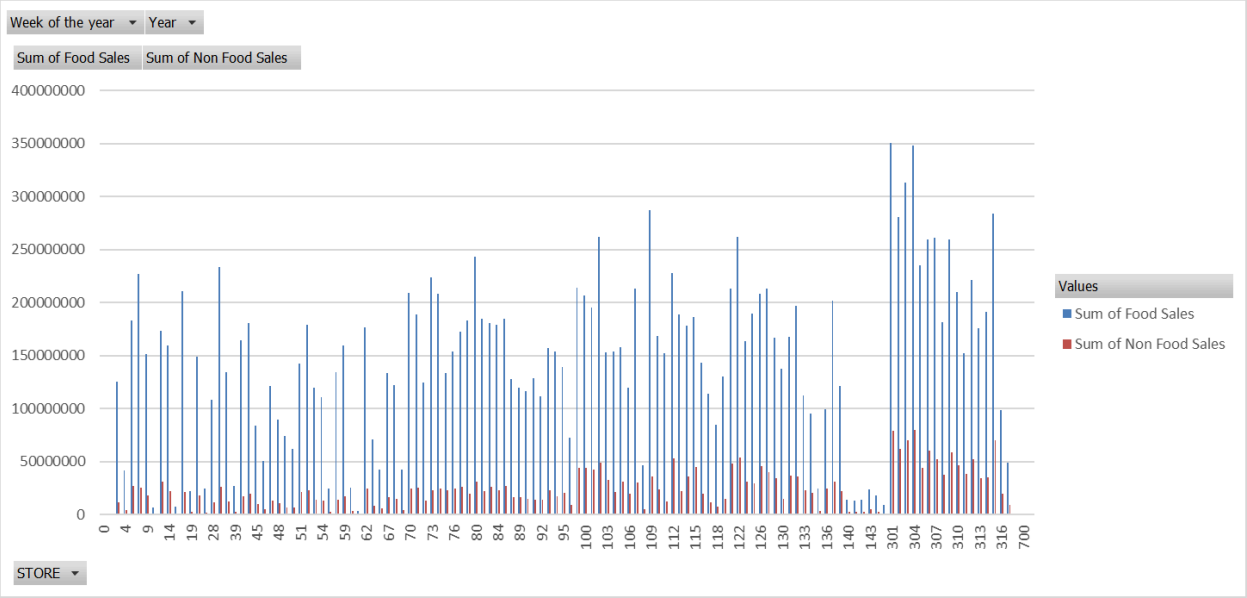


Fig: Bar Chart Comparing Food vs. Non-Food Sales Across All Stores

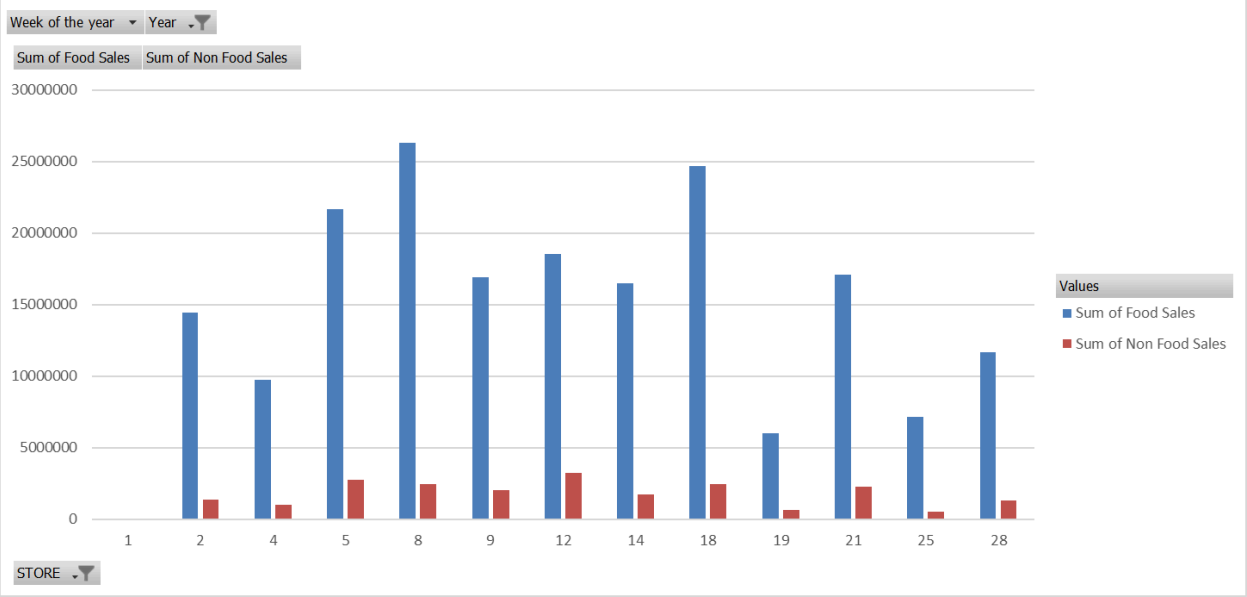


Fig: Filtered Bar Chart Highlighting Selected Stores for Food vs. Non-Food Sales

	A	B	C
1			
2	Week of the year	(All)	
3	Year	(All)	
4			
5	Row Labels	Sum of Food Sales	Sum of Non Food Sales
6	0	130201	20358
7	1	54116.58	9665.26
8	2	125788155.4	12053180.2
9	4	42123344.6	4174005.63
10	5	182790997.4	27055913.65
11	8	227306806.2	25711369.28
12	9	151098922.3	18240853.67
13	10	7027941	660777
14	12	173001301.4	31356429.62
15	14	159513513.7	22477112.69
16	16	7284527	780768
17	18	210740190.1	21502045.75
18	19	22306696.83	2378771.26
19	21	149035033.5	18217999.64
20	25	24327533.29	1724806.42
21	28	108015109.5	11912650.95
22	32	233233468.2	26621018.07
23	33	134285014.1	12419038.03
24	39	26809451.65	2763991.58
25	40	164217805.9	17740767.68
26	44	180441786.1	19732100.82
27	45	84402889.64	9892819.95
28	46	50648522.04	5307018.21
29	47	121314932.1	13043787.4
30	48	89775077.67	10719246.77
31	49	74242219.35	7108161.16
32	50	61831141.86	6667731.65
33	51	142134257.2	21144088.73
34	52	179014313	22867657.18
35	53	119374980.1	13958150.3
36	54	110671948.5	13181600.12
37	55	24733112.6	2655636.37
38	56	134502084.5	14038336.25
39	59	159779409.8	17547535.78
40	60	25468179.01	3446371.06
41	61	3222751	370844
42	62	176491366.6	24605648.45
43	64	71088555.46	8519931.85
44	65	42754315.77	5834030.57
45	67	133280655.2	16335654.18
46	68	122429250.6	14814421.34
47	69	42428693.48	4321570.76
48	70	208876168.2	24827676.09
49	71	188520874	25647163.75
50	72	124429562.5	13238866.78
51	73	223570965.7	23415906.1
52	74	208630739.2	24924836.57
53	75	133655896.2	23319310.48
54	76	154271050.7	24574168.92
55	77	172848676.7	26114028.66
56	78	183315706.4	19919748.12
57	80	243290585.4	31289688.88
58	81	184964901.6	22309376.97
59	83	180631359	26727650.28
60	84	179095915.8	23240362.1
61	86	184562949.2	27044047.64
62	88	127839691.4	16672482.8
63	89	119432698.9	16473968.84
64	90	116712280.2	15100479.68
65	91	128898986.6	14495885.74
66	92	111312943.9	14153531.42
67	93	157299732.5	23061220.94
68	94	154288893.1	17611149.01
69	95	138907320.8	20529525.35
70	97	72680890.39	9231755.1
71	98	213774432.3	43883707.19
72	100	206883044.5	44605491.21
73	101	194983334.1	42779929.51
74	102	262117057.8	49211597.59
75	103	153223971.1	32950742.78
76	104	154139053.9	21713056.21
77	105	158233512.3	31250117.95
78	106	120144693.2	19546290.87
79	107	213531886.5	30370811.42
80	108	46805911.42	5511765.16
81	109	287334310.9	36266785.44
82	110	168151607	24119372.11
83	111	152464941.9	12862511.46
84	112	227629046.4	53479685.92
85	113	188660716.5	22025860.32
86	114	178110843.9	36046765.46
87	115	186735152.6	44713876.78
88	116	143433065.1	19655299.74
89	117	114079386.6	12052225.65
90	118	84432740.76	7817664.84
91	119	130459247.9	15061518.37
92	121	213197692.4	48032037.11
93	122	261616460.6	54339775.4
94	123	163518960.7	31102622.98

Fig: Sales Data Pivot Table for Food and Non-Food Categories (1)

93	122	261616460.6	54339775.4
94	123	163518960.7	31102622.98
95	124	189960885.9	29947677.78
96	126	208454186.1	45560774.78
97	128	212797090.1	40076796.42
98	129	167251909.8	34804127.46
99	130	137953720.3	14669654.34
100	131	167621583.5	36833931.23
101	132	197221583.1	36128022.34
102	133	112849119.8	22818940.67
103	134	95139217.44	20762569.7
104	135	24974921.51	3975885.78
105	136	99199713.21	24864127.78
106	137	201538550.7	31303736.4
107	139	121515892.1	22244256.73
108	140	13840656.76	2687430.9
109	141	13640608.32	2761975.64
110	142	14370900.09	2539577.57
111	143	23633670.62	5052111.03
112	144	18597672.45	2627429.98
113	146	9214400.82	1365706.64
114	301	350673942.1	79534297.29
115	302	280629937.8	61806448.4
116	303	312741097.3	70014434.41
117	304	348369189.4	79809888.99
118	305	235163528.8	44105248.36
119	306	259412437.8	60365398.09
120	307	260861078.3	52633765.56
121	308	181341707.4	37704620.41
122	309	259860064.1	59001373.31
123	310	209856138.6	46282693.6
124	311	152115528	38120237.08
125	312	221316139.3	52142802.79
126	313	176036424.6	34655630.35
127	314	191022212.6	35211720.87
128	315	283537999.1	69934017.92
129	316	98926154.83	19676558.14
130	318	49133435.36	9040923.29
131	333	19	3
132	700	23	6
133	964	1138.33	273.54
134	**Grand Total**	**18121594908**	**2968321282**

Fig: Sales Data Pivot Table for Food and Non-Food Categories (2)

5. Which stores achieve the highest total sales in a given quarter?

   Rationale: Identifying the store with the highest total sales allows DFF to benchmark performance across all locations, revealing best practices in areas like product placement, customer service, product offerings, and understanding local demand. This also helps DFF allocate resources more effectively to meet customer needs, improve inventory management, and focus on enhancing performance at other stores. Additionally, recognizing the success factors can support business expansion by opening new stores in similar high-performing locations.

The following first Bar Chart illustrates the sales performance across all quarters for all stores with filters applied by year, while the second Bar Chart offers a filtered view of selected stores for enhanced visibility and analysis. Below these charts, the pivot table displays sales data segmented by quarter, with filters applied by year.

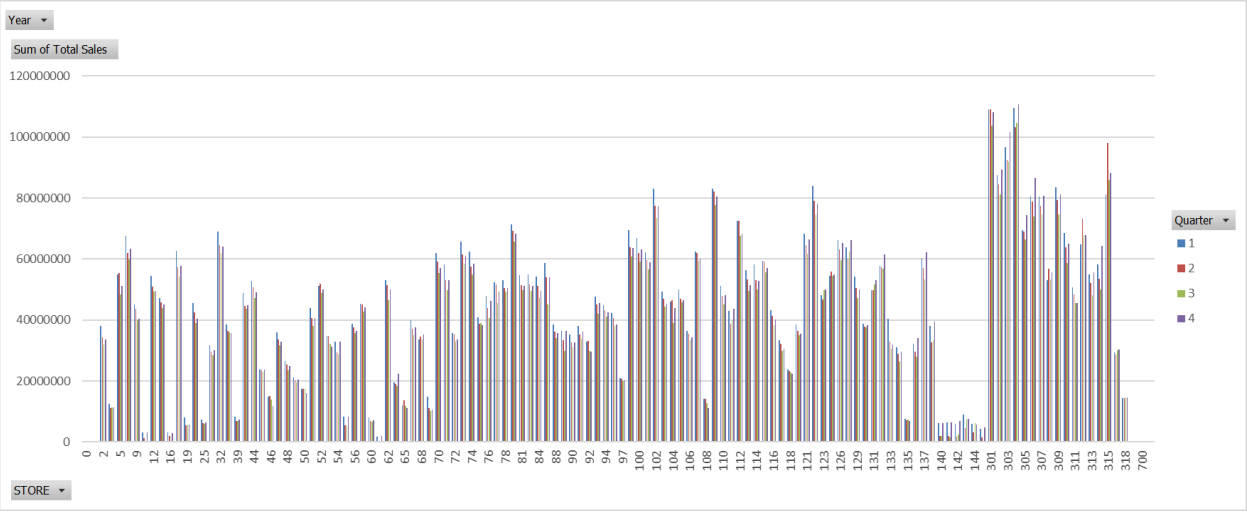


Fig: Bar Chart of Sales Performance Across All Quarters for All Stores

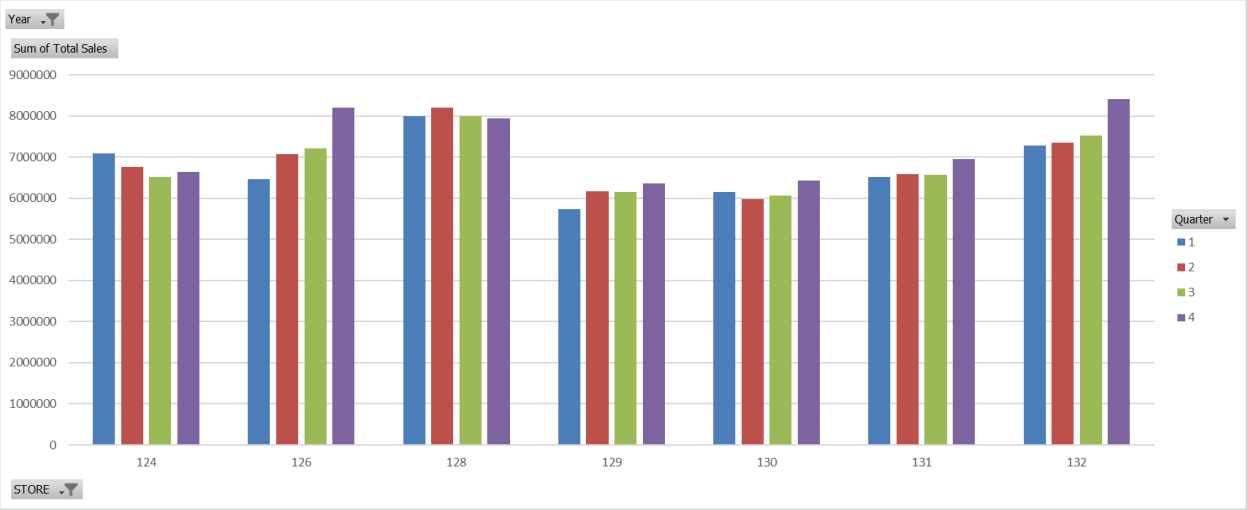


Fig: Filtered Bar Chart of Selected Stores' Sales Performance by Quarter

Year	(All)				
Sum of Total Sales	Column Labels				
Store Number	1	2	3	4	Grand Total
0	1	150558			150559
1	63781.84				63781.84
2	38015224.05	34319325.2	31892674.82	33614111.57	137841335.6
4	12490470.12	11221429.79	11248149.12	11337301.2	46297350.23
5	54933639.2	55358971.96	48301092.86	51253207	209846911
8	67590199.16	62079194.8	59926625.82	63422155.73	253018175.5
9	45179059	43692067.48	40063452.74	40405196.76	169339776
10	3131525	1392481		3164712	7688718
12	54593056.67	50872410.83	49301398.81	49590864.66	204357731
14	47185909.88	45740654.85	43905774.57	45158287.08	181990626.4
16	3051077	2070509		2943709	8065295
18	62762344.44	57358139.2	54355288.54	57766463.63	232242235.8
19	8030406.42	5544397.09	5465081.1	5645583.48	24685468.09
21	45513849.43	42427896.42	38985330.7	40325956.56	167253033.1
25	7428866.27	6197205.02	6088327.7	6337940.72	26052339.71
28	31738442.64	29591703.18	28479262.64	30118352.02	119927760.5
32	69043499.63	64565923.22	62058037.74	64187025.71	259854486.3
33	38534879.2	36557858.81	35873070.54	35738243.57	146704052.1
39	8309785.3	6927563.59	6971385.04	7364709.3	29573443.23
40	48835178.3	44641089.88	43713215.08	44769090.3	181958573.6
44	52904881.39	50842698.71	47224301.05	49202005.76	200173886.9
45	23867760.78	23506896.98	23134821.17	23786230.66	94295709.59
46	14941954.57	15181099.19	13937516.2	11894970.29	55955540.25
47	35970357.62	33715945.6	31663458.96	33008957.31	134358719.5
48	26539171.61	25384079.4	23641853.99	24929219.44	100494324.4
49	21123222.97	20154883.39	19576672.01	20495602.14	81350380.51
50	17557227.61	17544709.19	17384080.75	16012855.96	68498873.51
51	43963602.19	40630642.78	37982255.7	40701845.25	163278345.9
52	51138818.59	51878994.64	48921367.63	49942789.34	201881970.2
53	34788655.08	34888276.19	32275667.62	31380531.5	133333130.4
54	32839439.84	29407500.79	28670973.73	32935634.28	123853548.6
55	8420369.29	5444052.96	5182031.49	8342295.23	27388748.97
56	38872284.25	37608413.79	35624966.58	36434756.11	148540420.7
59	45259962.88	45150322.25	42815446.78	44101213.67	177326945.6
60	8002830.78	6917070.6	6794260.01	7200388.68	28914550.07
61	1700753	12511		1880331	3593595
62	53141005.39	51369823.39	46620862.36	49965323.91	201097015.1
64	19678446.74	19140592.89	18446041.63	22343406.05	79608487.31
65	11982915.8	13688615.61	11728986.15	11187828.78	48588346.34
67	39965888.53	37038699.49	34957206.47	37654514.85	149616309.3
68	33744485.04	34575378.36	33727232.12	35196576.4	137243671.9
69	14869883.3	11196856.38	10093748.11	10589776.45	46750264.24
70	62065980.13	59263279.95	55373584.97	57000999.23	233703844.3
115	59417619.18	59167398.22	55730064.2	57133947.73	231449029.3
116	43195758.37	41429582.61	38305531.79	40157492.11	163088364.9
117	33441347.57	32241969.31	29791272.1	30657023.27	126131612.3
118	23822438.96	23279085.85	22751180.7	22397700.09	92250405.6
119	38555673.68	36512245.26	35024567.18	35428280.19	145520766.3
121	68429108.71	64579295.7	61675833.01	66545492.04	261229729.5
122	83979866.53	79128182.84	74592792.31	78255394.31	315956236
123	48196268.71	46799483.39	49703715.95	49922115.61	194621583.7
124	54476024.35	55886283.83	54478704.78	55067550.72	219908563.7
126	66129293.02	63094098.17	59543590.1	65247979.57	254014960.9
128	63945897	60191305.76	62522631.28	66214052.48	252873886.5
129	54227866.83	50568292.12	47290459.63	49969418.67	202056037.3
130	38770194.92	37779597.26	37702581.24	38371001.2	152623374.6
131	49887429.08	49798036	51592212.82	53177836.81	204455514.7
132	57746675.37	57261964.29	56928045.78	61412920.01	233349605.5
133	40366060.17	32937694.72	30472356.61	31891948.95	135668060.5
134	31049914.3	28836263.6	26348699.58	29666909.66	115901787.1
135	7585996.97	7139287.37	7376006.42	6849516.53	28950807.29
136	32214069.97	29759117.22	28016227.1	34074426.7	124063841
137	60290552.7	57108702.04	53281932.85	62161099.52	232842287.1
139	38146449.92	32726333.1	33345727.39	39541638.4	143760148.8
140	6224441.39	1935503.44	2057967.16	6310175.67	16528087.66
141	6373247.7	2031778.4	1490240.74	6507317.12	16402583.96
142	5908862.78	1790255.11	2384740.37	6826619.4	16910477.66
143	9012921.28	4565287.56	7387250.02	7720322.79	28685781.65
144	5894186.83	3177999.49	6318816.68	5834099.43	21225102.43
146	4355843.66	1481477.25		4742786.55	10580107.46
301	109200875.6	109211840.1	103724720.2	108070803.5	430208239.4
302	87487077.15	84537539.12	81092596.55	89319173.37	342436386.2
303	96767487.46	92421462.68	91914024.2	101652557.4	382755531.7
304	109521466.7	103277395.3	104546616	110833600.6	428179078.4
305	69574991.2	68947791.16	66338757.88	74407236.94	279268777.2
306	80513101.01	78868335.14	73849218.3	86547181.39	319777835.8
307	80520367.73	77480937.18	74709194.39	80784344.57	313494843.9
308	53107166.12	56789799.07	53394080.24	55755282.39	219046327.8
309	83573099.33	79446399.75	74727839.43	81114098.92	318861437.4
310	68576319.78	63910799.72	58605522.76	65046189.95	256138832.2
311	50719332.38	48329598.53	45564830.79	45622003.41	190235765.1
312	64759919.04	73135497.97	67786304.95	67777220.14	273458942.1
313	54938042.8	52031663.96	48115840.15	55606508.08	210692055
314	58268280.82	53536544.2	50026553.7	64402554.71	226233933.4
315	81215825.12	98060254.4	86005856	88190081.5	353472017
316	29523833.2	28704894.27	30063533.36	30310452.14	118602713
318	14506731.31	14337165.35	14754760.82	14575701.17	58174358.65
333		12	10		22
700				29	29
964	302.48			1109.39	1411.87
Grand Total	5516501149	5262393389	4997161269	5313860382	21089916189

Fig: Sales Data Pivot Table Segmented by Quarter with Yearly Filters

6. What is the annual customer inflow for each store allowing for insights into customer behavior and store performance?

   Rationale: Understanding customer inflow helps DFF identify peak shopping times, ensuring sufficient staffing, inventory, and optimized store operations. Additionally, it allows for adjustments in store layout to enhance the customer experience, such as adding more checkout lanes or improving product placements. For stores with lower customer inflow, targeted marketing promotions and loyalty programs can be introduced to boost engagement.

   The following Line Chart provides a visual representation of customer count across all stores for easy analysis and visualization. Below it, the pivot table displays customer count data segmented by year, with filters applied for both year and week.

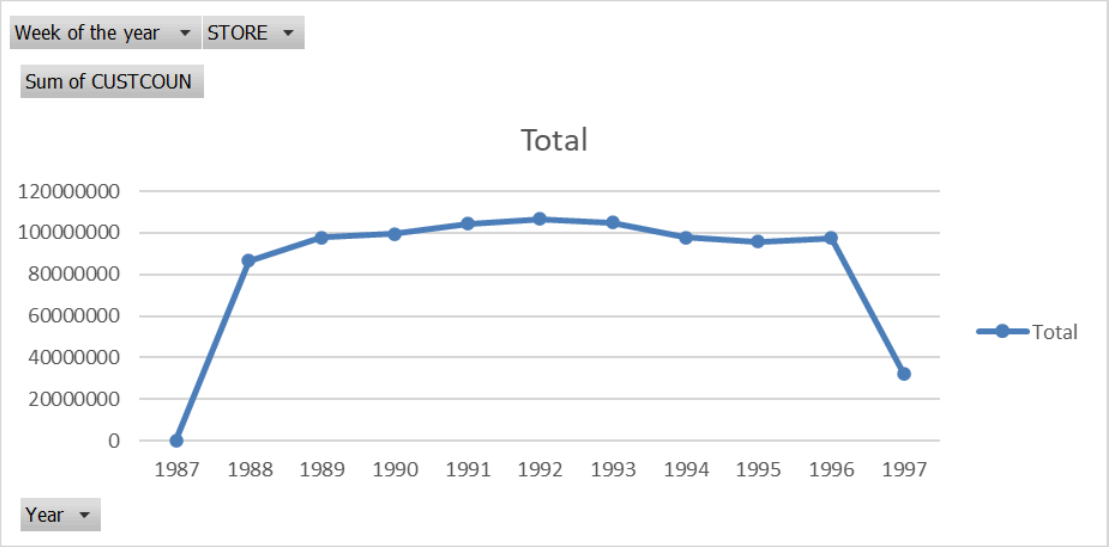


Fig: Line Chart of Customer Count Across All Stores with Year and Week Filters

Week of the year	(All)
STORE	(All)

Row Labels	Sum of CUSTCOUN
1987	215841.73
1988	86645981.93
1989	97861853.65
1990	99384416
1991	104340958
1992	106684354
1993	104875533
1994	97906188
1995	95740992
1996	97392644
1997	32244775
Grand Total	**923293537.3**

Fig: Customer Count Data Pivot Table Segmented by Year with Year and Week Filters

7. What are the top-selling departments (e.g., Bakery, Meat, Deli) across stores, and how do they vary by store and week?

   Rationale: This question helps the company understand which departments or product categories, like Bakery, Meat, or Deli, are driving the most sales and contributing to

overall profitability. By identifying the top-performing areas, the company can allocate resources more effectively and plan strategically to maximize profits. It also reveals any departments or stores that may be underperforming, by showing where sales are falling short. This allows the company to identify inefficiencies and apply targeted promotions to boost sales, helping both strong and weak stores improve performance and profitability.

The following bar chart illustrates the average sales of each product category during a certain period which can be filtered based on Store Number, Week Number, and Year.

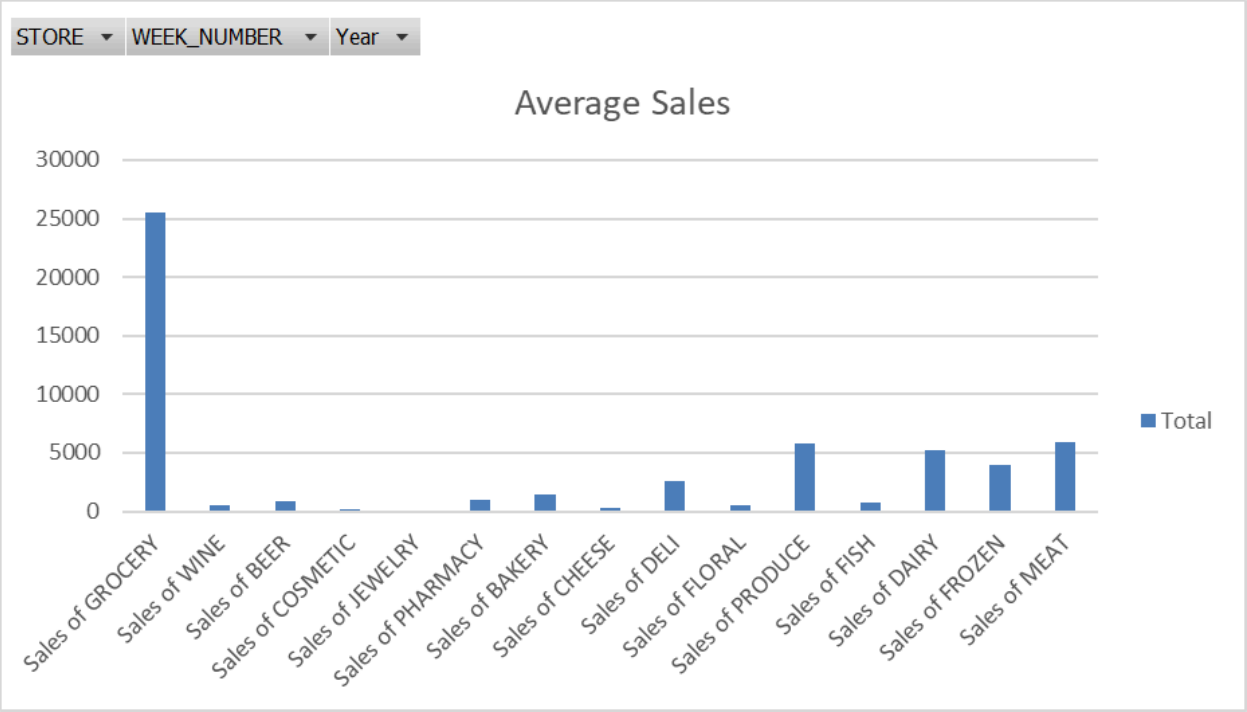


Fig: Average Sales of each product category with Store Number, Week, and Year Filters

Values	
Sales of GROCERY	25472.5278
Sales of WINE	490.849493
Sales of BEER	824.564652
Sales of COSMETIC	165.774232
Sales of JEWELRY	17.8417515
Sales of PHARMACY	977.5967
Sales of BAKERY	1451.06538
Sales of CHEESE	298.698676
Sales of DELI	2555.8164
Sales of FLORAL	496.221523
Sales of PRODUCE	5760.66726
Sales of FISH	753.338585
Sales of DAIRY	5166.60515
Sales of FROZEN	3910.1081
Sales of MEAT	5865.46699

Fig: Average Sales Data Pivot Table

8. Which stores or zones are driving the highest and lowest sales for a specific product?

   Rationale: Observing where a product from a category is being sold (based on store/zone) and in what quantities is important for companies because it helps them understand regional demand and customer preferences. This insight allows businesses to optimize inventory distribution, ensuring that the right products are available in the right locations. It also helps in adjusting marketing efforts and promotions to match local demand, reducing the risk of overstocking or stockouts. Overall, it enables companies to meet customer needs more effectively, improve sales performance, and boost operational efficiency.

   The bar chart illustrates profit distribution across various stores or zones for specific products, with each color representing a different store. The x-axis lists individual products, while the y-axis shows the profit amounts. This enables a comparison of which stores or zones drive the highest and lowest profits for each product.

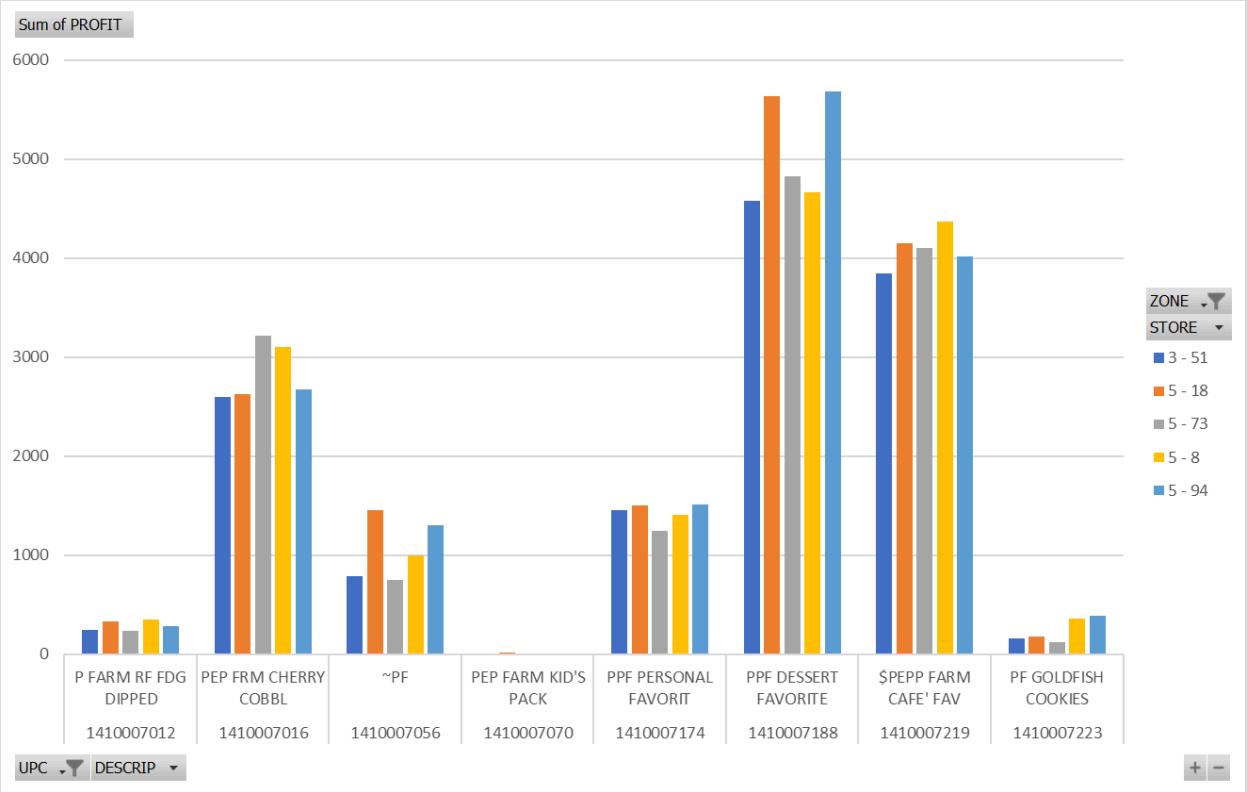


Fig: Sales of Different UPC's in Different Stores

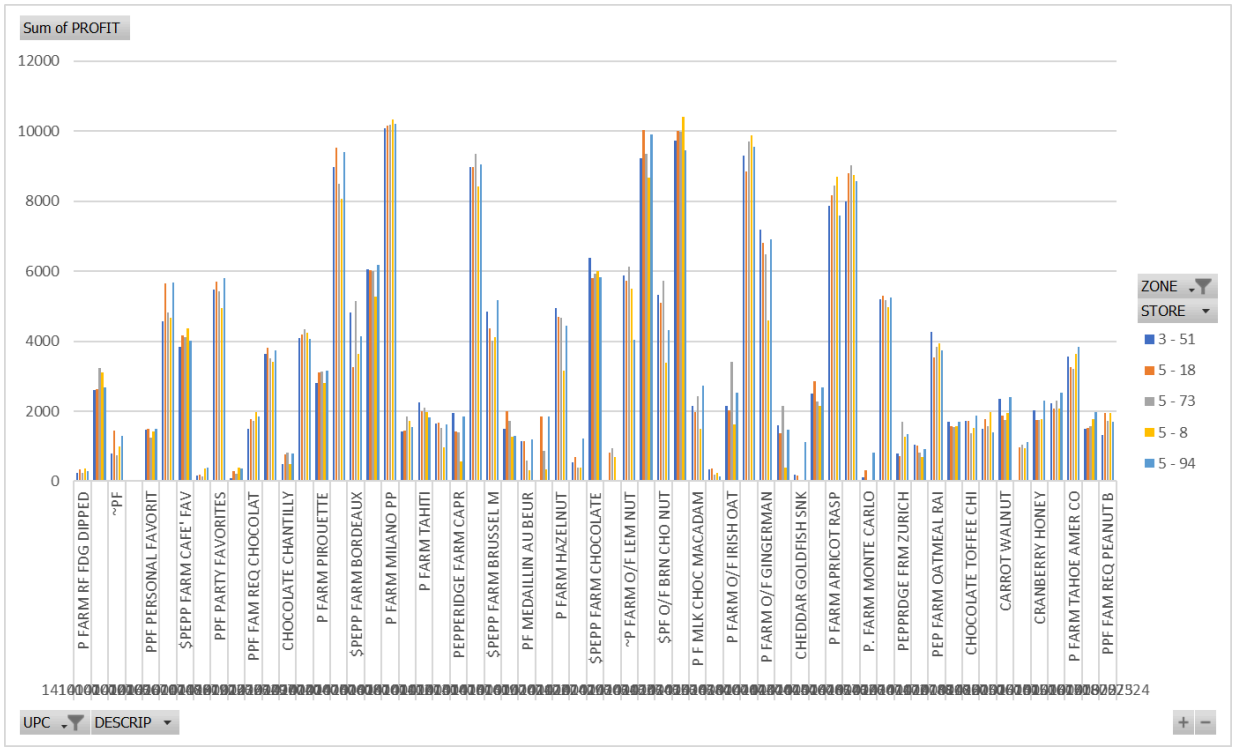


Fig: Sales of Different UPC's in Different Stores

Sum of PROFIT		ZONE	STORE															
UPC	DESCRIP	1	10	11	12	13	14	15	16	2	3	4	5	6	7	8	Grand Total
1410007012	P FARM RF FDG DIPPED	3691.54	219.62	501.21	2160.18	266.16	414.66	803.8	351.02	4267.05	244.56	260.69	1206.44	2084.68	1375.51	566.86	18413.98
1410007016	PEP FRM CHERRY COBBL	44081.79	3069.46	5879.86	27062.83	2819.82	3052.71	8669.2	3239.11	61067.04	2602.82	2755.61	11620.21	19500.46	17433.38	4020.26	216874.56
1410007056	~PF	18501.77	1199.71	2317.37	10389.7	1075.32	1227.6	2093.3	1458.12	24114.87	791.41	1049.89	4504.66	8841.78	5528.73	1864.78	84959.01
1410007070	PEP FARM KID'S PACK	16.84								16.84			16.84	16.84			67.36
1410007174	PPF PERSONAL FAVORIT	19441.78	2048.04	2900.92	12614.86	1347.55	1358.35	3918.75	1324.53	31487.12	1458.89	1540.65	5664.71	8476.52	6276.44	1736.61	101595.72
1410007188	PPF DESSERT FAVORITE	76951.52	5861.07	10623.61	46044.39	5073.79	5919.09	14913.86	5865.84	119302.97	4577.98	5324.32	20818.48	39250.85	28491.77	8781.5	397801.04
1410007219	$PEPP FARM CAFE' FAV	61871.31	4236.34	8373.81	36494.71	2760.73	3962.28	10938.98	4516.47	85389.04	3843.55	4066.14	16645.52	25085.96	22242.89	5163.26	295590.99
1410007223	PF GOLDFISH COOKIES	2531.55	124.39	164.31	2456.22	210.35	508.85	689.51	420.57	5660.16	159.29	112.43	1059.74	2246.52	1050.05	381.48	17775.42
1410007233	PPF PARTY FAVORITES	81646.24	7144.04	11504.39	48643.47	5511.21	5898.92	15526.03	5888.65	126007.75	5469.69	5495.19	21835.68	42373.61	29228.4	10066.19	422239.46
1410007234	PF GOLDFISH COOKIES	3679.28	138.37	411.44	3350.12	474.13	538.4	947.74	482.49	6977.23	89.85	333.57	1266.25	2380.61	1087.18	327.59	22484.25
1410007297	PPF FAM REQ CHOCOLAT	25585.36	3289.38	3778.3	15093.79	1143.49	1978.75	4367.31	2007.76	40223.7	1506.79	1343.96	7328.56	13343.89	9307.31	3468.81	133767.16
1410007402	P FARM SOUTHPORT ASS	51751.84	6584.57	7850.45	32415.38	1694.28	3989.57	9445.19	4036.47	92058.12	3627.01	2988.26	14456.98	27880.29	20160.29	7977.27	286915.97
1410007403	MILANO SINGLE SERVE		224.21							864.81							1089.02
1410007404	CHOCOLATE CHANTILLY	13306.04	920.22	506.07	7721.31	833.03	912.96	1945.27	966.14	17353.4	493.71	699.88	2843.96	5079.45	5454.47	934.53	59970.44
1410007405	P FARM CHAMPAGNE ASS	59247.72	7620.14	8363.15	36549.45	1761.89	4274.44	11780.83	4271.89	103733.48	4078.04	3600.63	16824.92	31286.17	21384.95	8367.08	323144.78
1410007406	P FARM PIROUETTE	41989.3	5293.19	6629.76	27521.09	1503.48	3531.26	8021.59	3426.5	78157.64	2807.34	2648.76	12201.27	23076.25	16585.77	8284.42	241677.62
1410007407	CHESSMAN SINGLE SERV		256.24							832.78							1089.02
1410007408	P FARM CHOCOLATE PIR	125540.29	11602.5	18018.13	80266.83	6843.71	10044.23	24924.83	9817.6	213968.91	8964.48	8269.97	35521.76	68307.06	46224.45	16838.98	685153.73
1410007410	$PEPP FARM BORDEAUX	81935.1	7466.49	10501.86	45125.5	2225.53	5497.43	12653.17	5277.71	111707.07	4827.54	4324.35	16197.25	31594.14	30218.91	8771.38	378323.43
1410007411	PEPPERIDGE FARM LIDO	91584.4	10185.63	13363.54	54933.71	3112.52	6809.91	15396.06	6555.47	139840.75	6052.25	5883.91	23485.42	40727.21	34934.21	9736.66	462601.65
1410007412	P FARM MILANO PP	148353.1	16505.5	20420.44	90754.67	8460.22	10376.72	30591.93	10398.79	256022.3	10082.04	9756.74	40880.19	80968.44	51593.15	24139.68	809303.91
1410007413	CAMPBELLS SOUP OYSTE	26428.34	3473.54	3547.42	16368.49	239.79	2382.54	4297.82	1884.19	43862.69	1412.85	1191.39	6567.97	12371.01	10543.46	3206.96	137778.46
1410007414	P FARM TAHITI	31814.1	3627.29	4392.12	19044.66	532.86	2505.6	5554.73	2239.26	48210.42	2245.13	2025.68	7881.88	13771.44	12396.8	2787.07	159029.04
1410007415	P FARM NASSAU	23411.93	2285.03	2770.28	14631.24	266.42	2305.29	4166.11	1897.48	35653.57	1641.21	1270.62	5774.56	9837.99	11415.45	506.21	117833.39
1410007416	PEPPERIDGE FARM CAPR	26881.44	2866.72	3352.35	15920.74	373.02	2540.55	4383.9	1974.59	40941.26	1941.63	1254.88	5208.3	12123.76	12687.52	3455.49	135906.15
1410007419	P FARM GENEVA COOKIE	134576.52	12485.54	19297.12	81493.89	7053.05	9820.03	25488.73	9982.26	213272.46	8980.4	8547.72	35805.65	67733.23	49839.65	18504.01	702880.26
1410007420	$PEPP FARM BRUSSEL M	71127.91	5152.98	11425.02	44661.37	2865	5825.53	13149.89	6079.2	117483.62	4845.7	5261.62	17669.32	38704.13	30876.15	12858.83	387986.27
1410007421	P F CHOC CHUNK PECAN	17671.44	1310.82	2546.16	9524.85	52.7	1310.82	3525.12	1175.28	24765.71	1487.51	885.07	6278.15	7205.45	6082.87	1085.04	84906.99
1410007422	PF MEDAILLIN AU BEUR	15078.07	693.47	929.25	9840.03	579.67	1353.76	2086.57	1564.92	24321.82	1148.21	802.84	3216.14	6470.35	5779.79	818.15	74683.04
1410007424	P FARM CAPPUCINO COO	25212.17	1946.7	3256.35	12944.79	213.16	2237.76	1239.55	1938.62	38019.29		1077.96	4911.71	9606.52	11441.01	1504.02	115549.61
1410007429	P FARM HAZELNUT	68688.86	6521.06	11117.41	41288.23	2992.2	5100.62	10790.83	5281.1	112391.99	4957.57	3918.02	16937.27	34651.17	26698.57	10464.45	361799.35
1410007430	P. FARM AQUARIUM PAC	10027.33	420.81	1816.34	6338.98	559.02	372.68	1728.6	106.48	15772.93	550.29	356.07	2668.82	3011.26	4072.23	1254.55	49056.39
1410007431	$PEPP FARM CHOCOLATE	85606.34	9618.35	11981.84	49785.73	2372.21	6166.44	15550.36	5185.98	134869.8	6376.96	5604.57	23562.86	38542.61	32047.21	10888.67	438159.93
1410007433	P FARM O/F OATMEAL R	10537.02	642.72	1918.8	2987.01			1026.48		8701.29			2441.4	1433.25	4568.07	55.38	34311.42
1410007434	~P FARM O/F LEM NUT	88298.51	10307.92	12383.52	51903.01	3466.48	6294.73	16433.89	6516.16	135045.4	5889.98	5367.45	21390.13	42045.71	33115.25	12169.1	450627.24
1410007435	P FARM ORANGE MILANO	141025.48	15144.62	20172.25	87001.39	7590.16	10337.93	28043.58	10416.82	234894.49	9231.8	9051.93	37949.59	74607.15	51730.38	19955.17	757152.74
1410007436	$PF O/F BRN CHO NUT	80784.35	9626.66	11394.55	47132.47	2842.06	5644.25	14762.48	5658.24	120068.89	5322.62	5232.16	18518.84	36806.79	31739.58	10053.91	405587.85
1410007438	P FARM O/F SUGAR COO	145572.86	15073.78	19376.62	89769.76	6653.72	10348.73	29387.81	10496.59	234764.7	9728.55	9530.18	39836.3	73438.93	50905.08	22767.71	767651.32
1410007440	P F MLK CHOC MACADAM	26613.81	2323.85	3148.74	20925.32	1001.07	2519.47	6469.17	2293.47	51568.12	2142.83	1171.31	8639	14917.62	8945.27	2667.12	155346.17
1410007441	P FARM FUDGE STRIPE	3975.51	173.75	628.82	2381.92	221.9	389.37	812.26	372.16	5459.3	349.21	382.45	919.19	2459.55	1391.59	419.22	20336.2
1410007443	P FARM O/F IRISH OAT	22773.41	2129.89	1208.94	21806.77	79.92	2777.54	6351.06	2625.64	42012.89	2147.9	2748.43	9602.1	12419.46	6704.01	5799.73	141187.69
1410007444	P FARM OLD FASH SHOR	140751.22	14711.92	20354.53	81642.32	4393.12	8983.72	27050.09	9226.42	207758.9	9309.29	9057.79	37974.36	61154.62	50235.03	17135.25	699738.58
1410007445	P FARM O/F GINGERMAN	116927.38	8983.15	16380.41	74436.92	7106.54	10319.34	21681.11	9582.3	181881.93	7192.14	8223.3	24787	52056.25	44956.95	14472.96	598987.68

Fig: Pivot Table for Sales of Different UPC's in Different Stores (1)

1410007415	P FARM NASSAU	23411.93	2285.03	2770.28	14631.24	266.42	2305.29	4166.11	1897.48	35653.57	1641.21	1270.62	5774.56	9837.99	11415.45	506.21	117833.39
1410007416	PEPPERIDGE FARM CAPR	26881.44	2866.72	3352.35	15920.74	373.02	2540.55	4383.9	1974.59	40941.26	1941.63	1254.88	5208.3	12123.76	12687.52	3455.49	135906.15
1410007419	P FARM GENEVA COOKIE	134576.52	12485.54	19297.12	81493.89	7053.05	9820.03	25488.73	9982.26	213272.46	8980.4	8547.72	35805.65	67733.23	49839.65	18504.01	702880.26
1410007420	$PEPP FARM BRUSSEL M	71127.91	5152.98	11425.02	44661.37	2865	5825.53	13149.89	6079.2	117483.62	4845.7	5261.62	17669.32	38704.13	30876.15	12858.83	387986.27
1410007421	P F CHOC CHUNK PECAN	17671.44	1310.82	2546.16	9524.85	52.7	1310.82	3525.12	1175.28	24765.71	1487.51	885.07	6278.15	7205.45	6082.87	1085.04	84906.99
1410007422	PF MEDAILLIN AU BEUR	15078.07	693.47	929.25	9840.03	579.67	1353.76	2086.57	1564.92	24321.82	1148.21	802.84	3216.14	6470.35	5779.79	818.15	74683.04
1410007424	P FARM CAPPUCINO COO	25212.17	1946.7	3256.35	12944.79	213.16	2237.76	1239.55	1938.62	38019.29		1077.96	4911.71	9606.52	11441.01	1504.02	115549.61
1410007429	P FARM HAZELNUT	68688.86	6521.06	11117.41	41288.23	2992.2	5100.62	10790.83	5281.1	112391.99	4957.57	3918.02	16937.27	34651.17	26698.57	10464.45	361799.35
1410007430	P. FARM AQUARIUM PAC	10027.33	420.81	1816.34	6338.98	559.02	372.68	1728.6	106.48	15772.93	550.29	356.07	2668.82	3011.26	4072.23	1254.55	49056.39
1410007431	$PEPP FARM CHOCOLATE	85606.34	9618.35	11981.84	49785.73	2372.21	6166.44	15550.36	5185.98	134869.8	6376.96	5604.57	23562.86	38542.61	32047.21	10888.67	438159.93
1410007433	P FARM O/F OATMEAL R	10537.02	642.72	1918.8	2987.01			1026.48		8701.29			2441.4	1433.25	4568.07	55.38	34311.42
1410007434	~P FARM O/F LEM NUT	88298.51	10307.92	12383.52	51903.01	3466.48	6294.73	16433.89	6516.16	135045.4	5889.98	5367.45	21390.13	42045.71	33115.25	12169.1	450627.24
1410007435	P FARM ORANGE MILANO	141025.48	15144.62	20172.25	87001.39	7590.16	10337.93	28043.58	10416.82	234894.49	9231.8	9051.93	37949.59	74607.15	51730.38	19955.17	757152.74
1410007436	$PF O/F BRN CHO NUT	80784.35	9626.66	11394.55	47132.47	2842.06	5644.25	14762.48	5658.24	120068.89	5322.62	5232.16	18518.84	36806.79	31739.58	10053.91	405587.85
1410007438	P FARM O/F SUGAR COO	145572.86	15073.78	19376.62	89769.76	6653.72	10348.73	29387.81	10496.59	234764.7	9728.55	9530.18	39836.3	73438.93	50905.08	22767.71	767651.32
1410007440	P F MLK CHOC MACADAM	26613.81	2323.85	3148.74	20925.32	1001.07	2519.47	6469.17	2293.47	51568.12	2142.83	1171.31	8639	14917.62	8945.27	2667.12	155346.17
1410007441	P FARM FUDGE STRIPE	3975.51	173.75	628.82	2381.92	221.9	389.37	812.26	372.16	5459.3	349.21	382.45	919.19	2459.55	1391.59	419.22	20336.2
1410007443	P FARM O/F IRISH OAT	22773.41	2129.89	1208.94	21806.77	79.92	2777.54	6351.06	2625.64	42012.89	2147.9	2748.43	9602.1	12419.46	6704.01	5799.73	141187.69
1410007444	P FARM OLD FASH SHOR	140751.22	14711.92	20354.53	81642.32	4393.12	8983.72	27050.09	9226.42	207758.9	9309.29	9057.79	37974.36	61154.62	50235.03	17135.25	699738.58
1410007445	P FARM O/F GINGERMAN	116927.38	8983.15	16380.41	74436.92	7106.54	10319.34	21681.11	9582.3	181881.93	7192.14	8223.3	24787	52056.25	44956.95	14472.96	598987.68
1410007446	P FARM O/F MOLASSES	25249.97	1816.44	3207.02	15695.12	202.58	1944.52	3937.03	1320.91	38446.29	1587	1571.88	5396.65	11485.94	9284.66	4670.98	125816.99
1410007448	CHEDDAR GOLDFISH SNK	8220.32	286.6	261.9	4995.47	1388.09	288.09	627.07	340.47	9747.78	190.74	52.38	1293.85	2795.86	2011.65	1444.2	33944.47
1410007457	SELECTION DE CHOIX P	38133.88	2195.91	3473.55	23085.7	2348.13	3094.49	7310.39	2902.88	54878.63	2512.23	2786.91	9994.76	19247.81	14465.47	4450.49	190881.23
1410007462	P FARM APRICOT RASP	122298.24	12582.21	17688.78	73539.65	6314.15	9084.19	23827.83	9322.49	201084.98	7860.56	8036.64	32902.93	60403.21	45966.97	14602.15	645514.98
1410007464	P FARM STRAWBERRY CO	133295.69	13185.13	19113.07	84477.91	7837.91	9812.7	28157.91	9860.84	210173.36	7995.23	8020.13	35128.44	66965.07	45826.74	14843.14	694693.27
1410007471	P. FARM MONTE CARLO	10043.63	115.89	386.24	2974.32			540.82	656.71	9643.34	115.89	347.67	1120.27	2626.74	1661.09	733.97	30966.58
1410007472	P FARM DOUBLE CHOC M	76498.26	6662.35	10596.84	46367.24	5085.28	5310.53	15661.02	5421.95	118749.45	5208.42	5223.47	20712.9	40079.97	26689.25	9776.49	398043.42
1410007478	PEPPRDGE FRM ZURICH	23403.78	2165.03	2695.43	12182.44	159.88	1401.95	3407.1	1085.72	37147.91	792.96	1427.04	5009.75	10966.56	9619.91	2607.44	114072.9
1410007481	~P FARM CARM PEC SOF	16189.36	1494.97	900.25	8242.65	904.24	773.28	2160.19	760.01	20571.46	1039.48	749.46	3451.72	5589.53	5484.39	1633.39	69944.38
1410007482	PEP FARM OATMEAL RAI	63948.29	3636.58	8105.68	37089.51	3952.24	4222.67	12115.03	4237.66	83479.75	4262.71	3708.99	15066.4	25748.54	21355.33	7123.24	298052.62
1410007505	PECAN SHORTBREAD	22815.39	2770.29	3553.74	13561.24	575.54	1400.88	3746.13	1289.28	36849.64	1688.76	1459.39	6393.34	11222.09	8669.2	3558.7	119553.61
1410007506	CHOCOLATE TOFFEE CHI	25443.39	2312.26	3505.8	14291.6	522.26	1635.42	3475.03	1768.66	37047.67	1715.4	1374.3	6478.83	11248.92	9883.78	3745.22	124448.54
1410007514	PPF FAM REQ OATMEAL	23372.64	2612.91	3081.3	14581.54	1462.43	1727.85	3938.61	2026.56	38072.1	1495.92	1337.74	6708.03	11192.5	7579.55	3281.5	122471.18
1410007515	CARROT WALNUT	31693.64	3613.02	4325.05	19105.39	1965.84	1973.16	6044.27	2055.73	50725.74	2350.23	2242.77	7961.42	14989.03	11429.56	3226.25	163701.1
1410007516	WALNUT DATE	15495.64	867.98	2290.59	9284.2	822.56	1098.86	2290.27	1274.04	23730.76		1018.39	4044.62	6495.28	5953.39	1588.31	76254.89
1410007517	CRANBERRY HONEY	28529.94	1914.68	4234.77	19197.74	2024.11	2501.06	4437.35	2881.65	49054.55	2030.71	2057.36	7534.28	13308.83	12150.09	2757.79	154614.91
1410007518	OATMEAL RAISIN	34228	3794.2	4839.78	20614.34	2146.02	2791.8	5738.01	2878.07	52000.93	2235.6	2295.95	8995.48	15428.76	12145.71	4317.99	174450.64
1410007522	P FARM TAHOE AMER CO	54325.09	3826.16	7949.04	33479.91	3762.6	4076.36	11434.73	3994.67	80476.07	3550.35	3160.49	13965.73	28304.83	19480.2	6558.95	278345.18
1410007523	PPF FAM REQ CINN SUG	23673.14	2761.52	3216.22	15895	1571.59	2248.94	5213.73	1709.35	44299.85	1484.24	1559.59	6831.23	14299.74	8257.74	4423.5	137445.38
1410007524	PPF FAM REQ PEANUT B	18396.24	2814.1	1304.69	9077.17		1779.46	2966.63	1755.26	39185.52	1328.94	1329.36	7321.36	12153.63	5558.02	3566.62	108537
otal		3060745.3	294937.91	420237.2	1859166.2	141616.73	226999.04	568636.64	224345.2	4826140.3	198024.39	193176.3	799231.37	1482441.8	1130222.5	393166.66	15819087.6

Fig: Pivot Table for Sales of Different UPC's in Different Stores (2)

9. Determine the sales of a given category during the weeks of special events.

   Rationale: Special events often drive higher customer traffic and increased spending, so tracking sales during these times can provide insights into customer buying behavior, allowing the company to plan better for future events by optimizing inventory, staffing, and promotional strategies. So, this business question is essential for analyzing the impact of promotions, holidays, and other special events on specific product categories. Understanding how sales in categories like Meat, HABA, Bakery, Pharmacy, etc. fluctuate during these periods helps the company evaluate the effectiveness of optimizing inventory, allocate resources effectively, and tailor product offerings to meet store-specific customer preferences.

   The following bar chart illustrates the total sales of product categories each year during special events such as Halloween, Thanksgiving, Year, etc., and can be filtered by Special Events and Store Number.

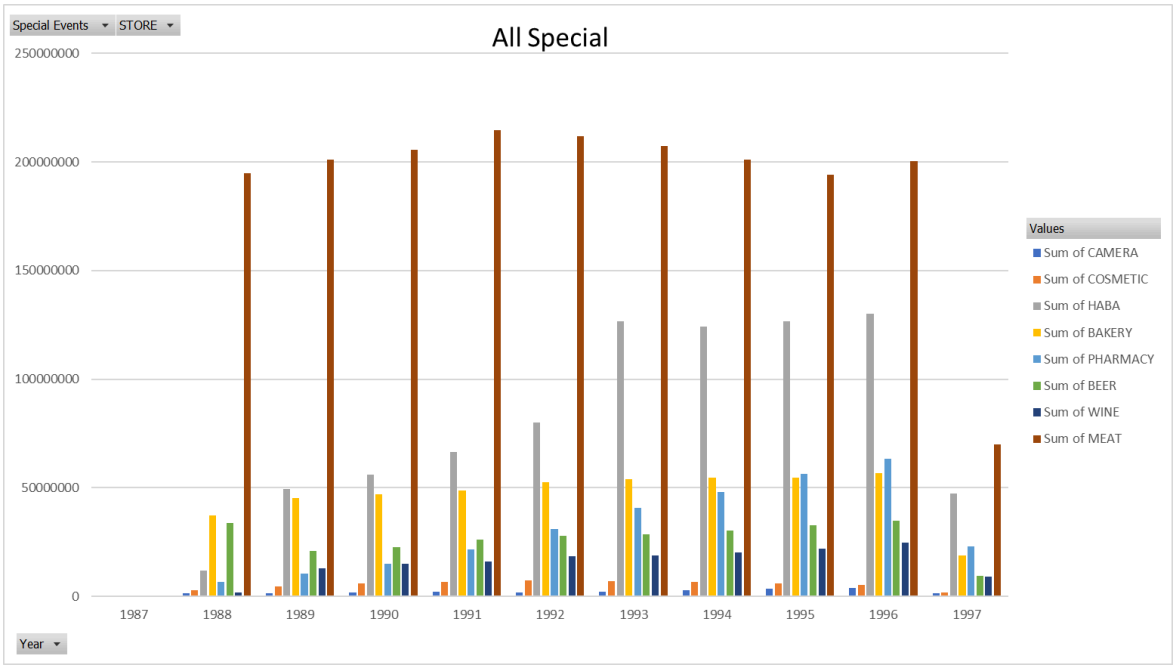


Fig: Total Sales of product categories each year during special events with Special Events and Store Number Filters.

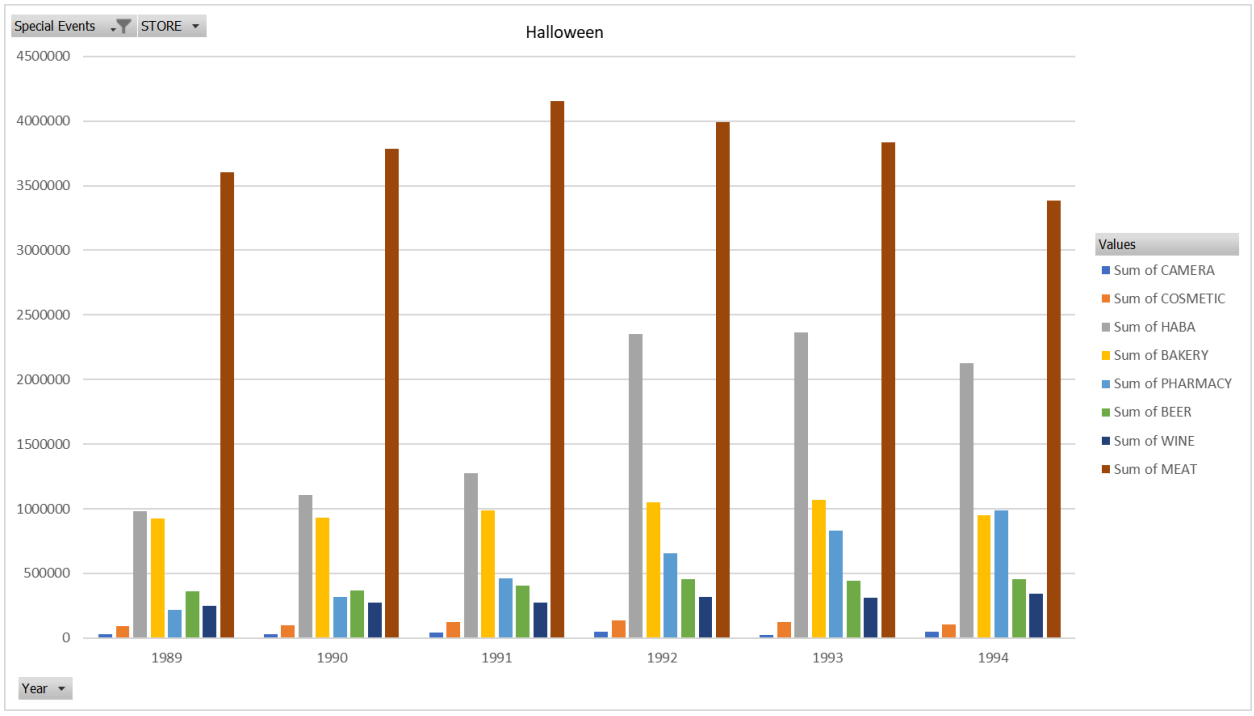


Fig: Total Sales of product categories each year during Halloween of all the stores.

Row Labels	Sum of CAMERA	Sum of COSMETIC	Sum of HABA	Sum of BAKERY	Sum of PHARMACY	Sum of BEER	Sum of WINE	Sum of MEAT
1987	0	0	0	4592.18	0	13233.77	0	28062.32
1988	1481564.3	2793802.15	11985757.8	37420096.72	6594730.15	33730113.69	1674788.85	194781008.1
1989	1544838.38	4540557.01	49452303.08	45280061.73	10371910.77	20922872.25	13086642.23	200998663
1990	1742050.45	5888512.67	56227637.53	47023454.91	15149334.83	22670908.31	14875793	205725762.3
1991	2037994.35	6719181	66373751.06	48903740.14	21761895.21	26193141.08	16019131.14	214743722.2
1992	1902048.53	7296986.16	79909990.24	52591025.21	31162555.89	27829738.59	18408560.45	211983261
1993	2196794.62	6951340.65	126595176.8	54031688.41	40907630.48	28644919.19	18831875.95	207491147.4
1994	2702151.08	6592177.3	124139587.2	54536231.36	47947360.44	30179810.46	20162115.58	201025816.7
1995	3499491.37	5999165.44	126668536.1	54839225.12	56454507.56	32729201.96	22058673.43	194138303.7
1996	4057030.77	5287156.35	130298047.8	56894129.81	63443242.1	34890088.41	24690145.67	200349886.6
1997	1453730.75	1662694.3	47505120.55	18802321.33	23070362.94	9457990.13	9288865.48	69878855.94
Grand Total	**22617694.6**	**53731573.03**	**819155908.1**	**470326566.9**	**316863530.4**	**267262017.8**	**159096591.8**	**1901144489**

Fig: Total Sales Data Pivot Table for each product category during each year

10. Determine the number of units sold in each store, by product category, within a city.

Rationale: This helps the company understand local demand and sales trends, allowing for adjustments in inventory and promotions. From a logistics perspective, this data enables optimized stock replenishment and more efficient transportation scheduling, reducing overstocking, understocking, and unnecessary shipments. By accurately forecasting demand, the company can streamline delivery routes, minimize transportation costs, and ensure timely product availability, improving both customer satisfaction and operational efficiency.

The below line chart illustrates the total number of units being sold for a product from each store by filtering based on the city or product itself.

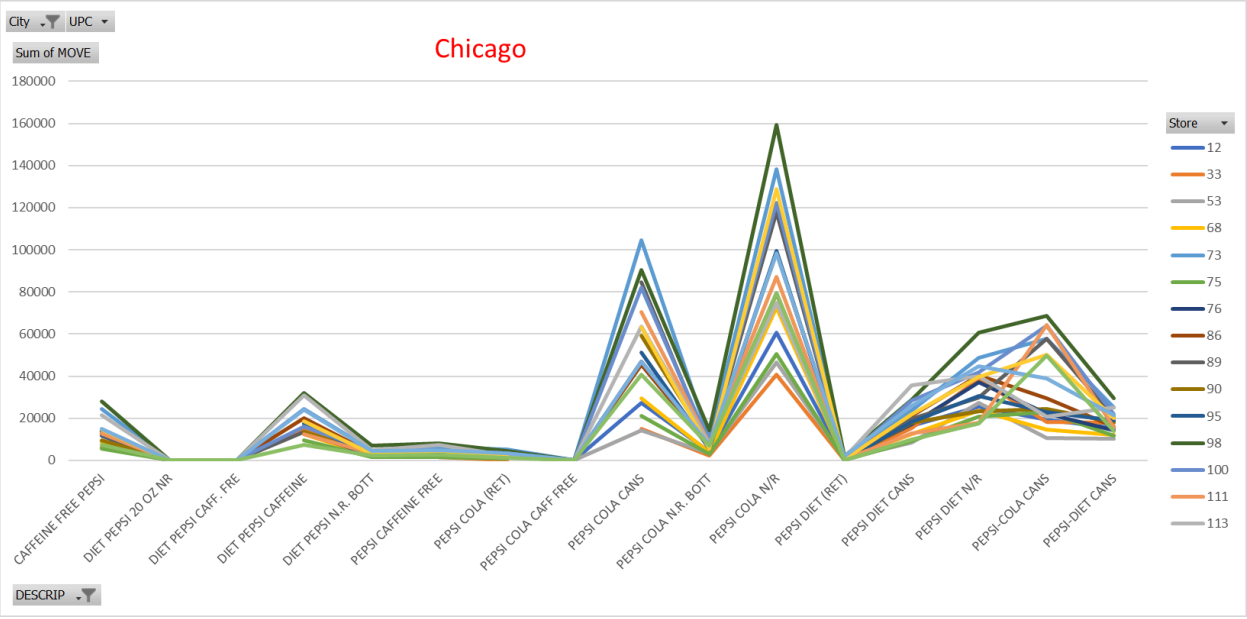


Fig: Total number of units sold for a product from each store, with filters for city and UPC

Sum of MOVE	Store																		
DESCRIP	12	33	53	68	73	75	76	86	89	90	95	98	100	111	113	123	128	130	Grand Total
CAFFEINE FREE PEPSI	8911	6396	13010	8859	24300	5670	11894	13990	12228	9398	13814	27956	12737	12741	21418	14559	14903	7371	240155
DIET PEPSI 20 OZ NR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DIET PEPSI CAFF. FRE	0	0			0			0	0			0	0		0		0	0	0
DIET PEPSI CAFFEINE	15178	14280	16696	12312	24389	9639	12879	20068	13642	13627	17516	31914	15919	12447	31000	18551	24232	7401	311690
DIET PEPSI N.R. BOTT	2929	2239	2459	2061	4938	1535	2984	3270	2443	3829	2411	7037	4594	3261	4903	3180	4440	2339	60852
PEPSI CAFFEINE FREE	3038	1697	2289	2668	6127	1697	3612	4241	4718	3697	4221	7967	4939	4647	7509	4531	4838	2608	75044
PEPSI COLA (RET)	1578	5	1438	1568	5058	814	3048	2229	2736	1939	1890	4566	3211	2505	2373	2328	3251	1087	41624
PEPSI COLA CAFF FREE	0		0		0		0	0				0	0		0		0	0	0
PEPSI COLA CANS	27336	14924	14355	29508	104366	20993	46969	45511	84511	59206	51094	90388	82108	70363	63578	63304	46828	40587	955929
PEPSI COLA N.R. BOTT	4116	2307	3217	3924	10436	2947	6808	6327	6488	7685	5597	14190	8174	8230	7597	5481	7209	7058	117791
PEPSI COLA N/R	60720	40628	46528	72577	138302	50440	119694	98680	117891	78978	99609	159294	122234	87166	74966	128870	98537	79411	1674525
PEPSI DIET (RET)	444	1	725	461	1173	229	1129	1111	520	291	246	1946	470	351	1070	642	1952	71	12832
PEPSI DIET CANS	16323	15297	8356	12454	23694	9356	16994	21218	19842	17905	18136	28687	28440	12924	35486	21889	25830	10096	342927
PEPSI DIET N/R	25334	38469	27202	23631	48557	20688	36987	40365	29982	23401	30486	60815	41936	17887	39852	39519	44566	17479	607156
PEPSI-COLA CANS	19237	18352	10588	14729	57615	22795	22409	29375	57861	24401	22910	68715	64026	64418	20656	50077	38830	49754	656748
PEPSI-DIET CANS	15193	17807	10318	12206	25155	11644	14389	16049	20299	18232	18808	29477	22044	15890	25276	19975	21646	13730	328138
Grand Total	200337	172402	157181	196958	474110	158447	299796	302434	373161	262589	286738	532952	410832	312830	335684	372906	337062	238992	5425411

Fig: Pivot table with total units sold for each product category by store, filtered by city.

5 Business Questions selected by DFF:

Below are the 5 business questions that DFF has selected. Refer to the questions 1, 4, 6, 7, and 9 in the above section for the justification and rationale for each question.

1. What is the proportion of sales made with coupons vs. without coupons in a given category?
2. How do food and non-food product sales compare across all stores, segmented by year and week number?
3. What are the top-selling departments (e.g., Bakery, Meat, Deli) across stores, and how do they vary by store and week?
4. Determine the sales of a given category during the weeks of special events.
5. What is the annual customer inflow for each store allowing for insights into customer behavior and store performance?

Data Warehousing: Why & How?

Data warehousing is very essential in today's business world, where information extracted from meaningful insights in the ocean of data bombarded requires a seamless and organized information system. Read further to understand the logical design and physical design for a Data Warehousing project for Dominick's Fine Foods, a well-recognized retail chain,developed by using Ralph Kimball's Dimensional Modeling methodology. We are doing this to come up with a scalable and insightful data warehouse using the star schema approach. This process will help confirm that the system will be able to answer key business questions and support the decision-making process.

Logical design in the data warehousing process is one of the most important steps because it provides the design on how to mold and organize raw data into meaningful dimensions and fact tables. In this stage, the focus is to define the logical modeling of data based on meeting business needs, ensuring that the data is more accessible and analyzable. We keep a central fact table connected to several dimension tables using a star schema for efficient querying and reporting. The primary focus is on the star schema, although snowflake schemas are also considered where normalization, if necessary, promotes better performance.

We will define the relationships between different dimensions like stores, products, customers, and time with fact tables to ensure the data model supports the business questions posed for DFF. Design of a clear and comprehensive schema using visual tools such as Visio and LucidChart that are subsequently implemented in SQL Server 2016 will be discussed in the following sections. Further, mapping tables will also be developed to ensure the smooth transformation of data from source systems to staging and presentation layers.

The logical design will provide the basis on which all data analysis and business intelligence in the future are carried out at DFF to support the organization in making proper data-driven decisions aimed at enhancing profitability and operational efficiency.

Overview of Kimball’s Methodology

The purpose of this Kimball methodology is to build data warehouses that are business driven. It is developed by Ralph Kimball and focuses on creating dimensional models and independent data marts to meet specific business needs. The dimensional models and independent data marts are part of the Kimball principles that are useful to transform complex data into structured information that can be queried easily and derive insights from the data. There are a sequence of steps from gathering business requirements to building reports to help to build the models, marts and build data warehouses that align with the business goals. The advantages include a modular approach, managing large volumes of data, reliability and data optimization. The modular approach is where independent data marts come into picture, that is, business functions have independent data marts that integrate and retrieve the data more efficiently. In short, Kimball methodology improves decision-making and strategic planning for organizations

The following are the detailed steps in the Kimball methodology

1. **Business Requirements:** The first step involves understanding the company's business functions, data needs, and the types of reports required to derive insights and answer key business questions. This can be achieved by conducting interviews and Joint Application Development (JAD) sessions, which help clarify the business questions that need to be addressed. The accurate information needs to be gathered during this phase which is critical to ensure that the data warehouse is designed to meet real business needs effectively.

2. **Dimensional Modeling:** In this step, a dimensional model is designed to optimize data structuring for analytical queries. Dimensional modeling involves creating fact tables (store measurable data) and dimension tables (provide context for the facts), establishing a way to connect and analyze data. This step is crucial in the Kimball methodology because it transforms raw data into a structured format that makes querying and analysis more efficient.

3. **Logic Design:** This phase involves designing the logical structure of the data warehouse. It includes defining the relationships between fact and dimension tables, organizing business rules, and establishing the ETL (Extract, Transform, Load) processes. Logical design acts as the basic picture for the physical design and the overall functioning of the data warehouse.

It is important to follow the Kimball methodology for creating independent data marts because independent data marts align with specific organizational needs and functions. As a result, the data marts are modular and developed making it more scalable and adaptable. The dimension modeling integrates multiple data marts into a single data warehouse for efficient data integration. The query performance, results fetching is improved and faster leading to business insights. The maintenance and updating of the independent data marts is easy as it does not depend on other data marts. The functioning of the warehouse is smooth and cost effective. Overall, Kimball's methodology provides a structured way to build optimized, scalable, and business-aligned data marts.

DW Logical Design (Star Schema Design)

Data Mart Matrix

Data Marts	Store Dimension	Date Dimension	Demographic Dimension	Product Dimension
Fact_Sale	X	X		X
Fact_Customer_ Count	X	X	X	

A star schema is a type of database schema that is commonly used in data warehousing and business intelligence systems. The schema resembles a star: a central fact table surrounded by related dimension tables. Below is a star schema design for our DFF database.

Dimension Tables

Dim_Store

The Dim_Store table contains detailed information about each store location, including its unique store identifier Store # (Primary Key). It tracks the store's geographical details such as City, Zone, Zip code, and Address, which are used for store-based analysis (Top-selling depts and Food vs. non-food), and customer inflow.

Dim_Store	
Primary Key	Store_Key
	Store #
	City
	Price Tier
	Zone
	ZipCode
	Address

Dim_Date

The Dim_Date table provides a structured view of date-related attributes, including each record's unique identifier, Date_ID (Primary Key). It includes fields such as Week_Start_Date, Week_End_Date, Year, Month, Quarter, and Week #, allowing for time-based analysis (Top-selling depts and food vs. non-food). This dimension table supports filtering and aggregating data by different periods, enabling trends and seasonal analysis.

Dim_Date	
Primary Key	Date_Key
	Date_ID
	Year
	Month
	Quarter
	WeekYear
	Week_of_the_Year
	Special Events

Dim_Demo

The Dim_Demo table stores demographic-related attributes, including each store's MMID (Primary Key) as the unique identifier. It contains fields such as Scluster, SHPCONS, SHPHUR, SHPAVID, SHPKSTR, SHPUNFT, and SHPBIRD, which capture various consumer behavior patterns and preferences. This dimension enables the analysis of customer segments and shopping habits to better understand market trends and target specific groups.

Dim_Demo	
Primary Key	Demo_Key
	MMID
	SHPCONS
	SHPHUR
	SHPAVID

	SHPKSTR
	SHPUNFT
	SHPBIRD

Dim_Product

The Dim_Product table contains details about each product, with Product_ID as the unique identifier. It includes attributes such as Product_Desc (a description of the product) and Product_Group (a classification or category for grouping similar products). This dimension is used for product-level analysis, helping link sales and other metrics to specific items or product categories.

Dim_Product	
Primary Key	Product_Key
	Product_ID
	Product_Description
	Product_Group

Fact Tables

Fact_Sales

The Fact_Sale table records transactional data related to product sales, with foreign keys linking to the Store #, Date_ID, and Product_ID (Primary Key) dimensions. It captures the Sales amount for each product sold at a particular store on a specific date. This fact table enables a detailed analysis of sales performance by store, time, and product category.

Fact_Sales	
Surrogate Key	Sales_ID
Foreign Key	Store #
Foreign Key	Date_ID
Foreign Key	Product_ID
	Sales

Fact_Customer_Count

The Fact_Customer_Count table tracks the number of customers visiting each store, with foreign keys referencing Store #, Date_ID, and MMID (demographic profile). The CustCount field captures the total number of customers for each combination of store, date, and demographic segment. This fact table supports analysis of customer traffic patterns and demographic insights across different periods and locations.

Fact_Customer_Count	
Surrogate Key	Record_ID
Foreign Key	Store #
Foreign Key	Date_ID
Foreign Key	MMID
	CustCount

Star Schema Representation

Sales Data Mart

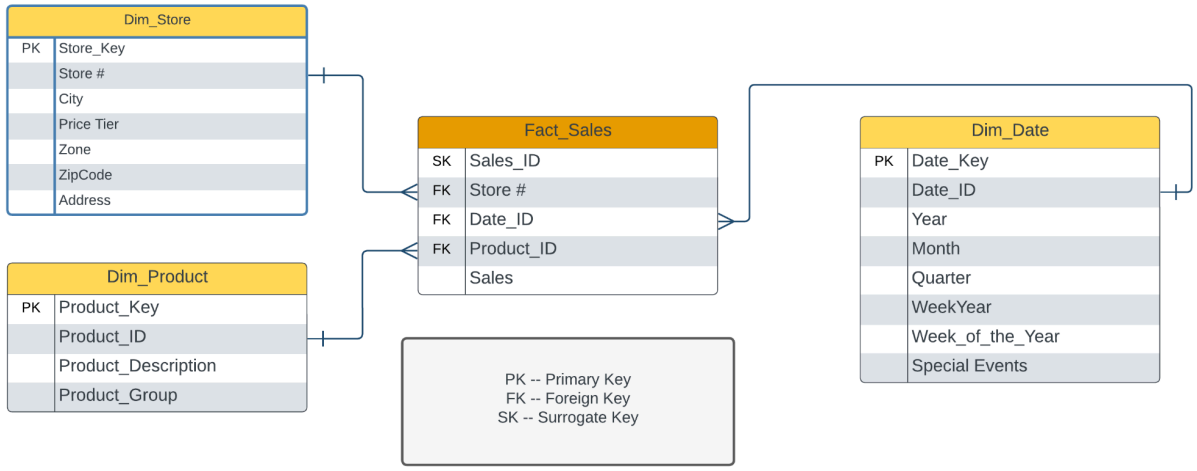


Fig: Sales Dart Mart

Customer Count Data Mart

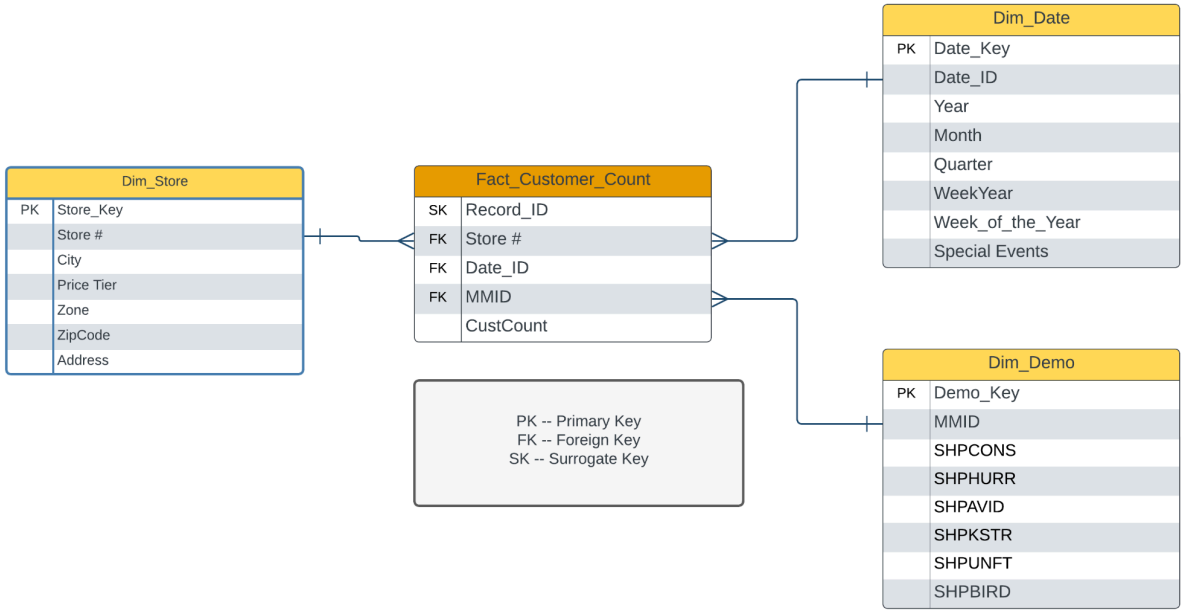


Fig: Customer Count Dart Mart

Data Warehouse Schema

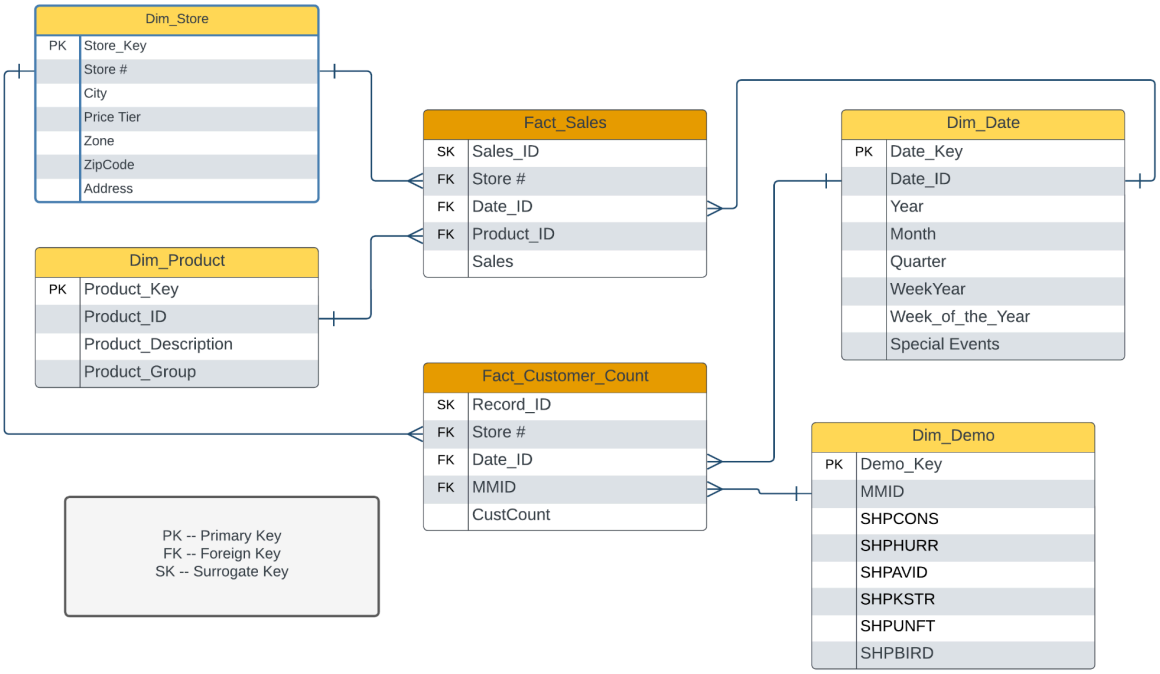


Fig: Data Warehouse Schema

Justification of each schema with our logical design and with data explaining how it supports the chosen BQs

1. What is the proportion of sales made with coupons vs. without coupons in a given category?

   Justification: The data we need to answer this question is the sales made in each category by using a coupon and not using a coupon. This data is available in the Fact_Sales table in the column Sales which can be further granularly examined by using the columns Store #, Date_ID, Product_ID. The Product_ID will determine the category of sales and that will help us group data by category to find the sales made using coupons and sales made by not using coupons.

   The pivot table presents comparison of coupon-driven sales as a percentage of total sales across all stores, with filters applied for the year and specific months. These filters can be applied and by exploring the data granularity using the columns Date_ID or Store #. This chart provides a comprehensive view of coupon usage trends over Q3 of 1996 (Which can be applied as filters).

Store	Total Grocery Sales	Sum of Absolute Coupon Sales	Sum of Coupon % of Total Sales
2	1239719.32	7077.29	0.6%
8	2621730.44	24293.85	0.9%
9	1850709.83	15803.5	0.9%
12	1941118.34	9523.72	0.5%
14	1644757.16	8720.07	0.5%
18	2283402.4	17204.34	0.8%
21	1307143.32	11727.55	0.9%
28	1266490.73	8953.68	0.7%
32	2663881.21	17598.76	0.7%
33	1612977.63	6025.25	0.4%
40	1272804.57	15117.16	1.2%
44	2570571.14	16202.86	0.6%
47	1481271.1	13703.54	0.9%
51	1433005.92	12530.52	0.9%
52	2118797.83	11949.46	0.6%
53	1219840.23	8291.96	0.7%
54	1403137.25	23895.45	1.7%
56	1738369.28	13719.13	0.8%
59	1683681.92	14345.49	0.9%
62	2675460.36	6880.92	0.3%
67	1438167.69	13280.32	0.9%
68	1446699.36	7956.31	0.5%
70	1996326.46	20446.2	1.0%
71	2030953.27	33048.15	1.6%
72	1228473.6	8912.69	0.7%
73	1870439.05	19294.47	1.0%
74	2338850.96	23468.78	1.0%
75	1442780.45	5440.06	0.4%
76	1405449.3	10293.66	0.7%
77	2100731.44	14828.05	0.7%
78	1985719.95	19735.41	1.0%
80	2074580.75	16997.27	0.8%
81	1961765	14020.31	0.7%
83	1780286.01	16203.61	0.9%
84	2112160.81	18965.39	0.9%
86	2569316.41	15181.66	0.6%
88	1341403.57	11785.18	0.9%
89	1077088.77	7483.02	0.7%
90	1106916.07	7203.7	0.7%
91	1320590.11	12651.62	1.0%
92	511680.15	6545.12	1.3%
93	1588409.17	8224.61	0.5%
94	1489300.21	9832.13	0.7%
95	1523346.21	8526.93	0.6%
98	2225654.17	24459.93	1.1%

Fig: Sample Pivot table for Coupon % sales of Total Sales (1)

100	2103491.04	16718.38	0.8%
101	2010373.87	16083.25	0.8%
102	2684720.64	22270.4	0.8%
103	1331655.22	12746.74	1.0%
104	2231890.69	19571.98	0.9%
105	1264209.29	10491.8	0.8%
106	962149.48	8664.99	0.9%
107	2438674.01	18563.64	0.8%
109	2863091.63	10263.4	0.4%
110	1415661.75	11339.25	0.8%
111	1451881.51	10377.23	0.7%
112	2231938.34	13484.47	0.6%
113	2018428.58	9284.14	0.5%
114	1755179.3	13258.52	0.8%
115	2345176.39	54048.1	2.3%
116	1384341.05	13242.45	1.0%
119	1568116.69	9189.17	0.6%
121	2557839.32	16728.12	0.7%
122	2340486.64	13906.88	0.6%
123	1507424.33	13399.48	0.9%
124	1677332.28	11865.61	0.7%
126	2966633.3	63181.8	2.1%
128	2083877.47	15248.81	0.7%
129	2452726.27	13915.61	0.6%
130	1861246.71	13036.47	0.7%
131	1706219.84	14242.2	0.8%
132	2435177.02	24817.05	1.0%
133	2079160.32	16688.3	0.8%
134	1508635.81	8035.91	0.5%
135	2447580.47	37295.52	1.5%
136	1860158.62	6114.01	0.3%
137	3214542.41	11853.39	0.4%
139	2651424.01	15837.6	0.6%
140	755991.26	36777.78	4.9%
141	564102.36	63083.94	11.2%
142	923413.58	44782.49	4.8%
143	2488224.28	62245.57	2.5%
144	2303403.76	47960.04	2.1%
301	5466493.44	226531.79	4.1%
302	4337407.01	150154.29	3.5%
303	4339492.96	232915.46	5.4%
304	6094112.78	137458.8	2.3%
305	3856196.16	136509.02	3.5%
306	4820040.51	174403	3.6%
307	4976056.65	204724.95	4.1%
308	4329251.82	157720.73	3.6%
309	4452675.17	147629.97	3.3%
310	4249036.27	157862.16	3.7%

Fig: Sample Pivot table for Coupon % sales of Total (2)

2. How do food and non-food product sales compare across all stores, segmented by year and week number?

   Justification :

   To address the analysis of sales data for various products categorized as Food and Non-Food, it is essential to aggregate the relevant information. The sales figures can be found in the Fact_Sales table, specifically within the Sales column. This aggregation will be performed by grouping the products according to the Product_Group column from the Dim_Product table. For a more granular level of analysis, we will also consider additional columns, including Store_ID, Date_ID, and Week_Number. This approach will provide a comprehensive view of sales performance across different product categories and time period**s.**

   The pivot table displays the sales data for stores, segmented into food and non-food categories, with filters applied for year and specific week.

	A	B	C
1			
2	Week of the year	(All)	
3	Year	(All)	
4			
5	Row Labels	Sum of Food Sales	Sum of Non Food Sales
6	0	130201	20358
7	1	54116.58	9665.26
8	2	125788155.4	12053180.2
9	4	42123344.6	4174005.63
10	5	182790997.4	27055913.65
11	8	227306806.2	25711369.28
12	9	151098922.3	18240853.67
13	10	7027941	660777
14	12	173001301.4	31356429.62
15	14	159513513.7	22477112.69
16	16	7284527	780768
17	18	210740190.1	21502045.75
18	19	22306696.83	2378771.26
19	21	149035033.5	18217999.64
20	25	24327533.29	1724806.42
21	28	108015109.5	11912650.95
22	32	233233468.2	26621018.07
23	33	134285014.1	12419038.03
24	39	26809451.65	2763991.58
25	40	164217805.9	17740767.68
26	44	180441786.1	19732100.82
27	45	84402889.64	9892819.95
28	46	50648522.04	5307018.21
29	47	121314932.1	13043787.4
30	48	89775077.67	10719246.77
31	49	74242219.35	7108161.16
32	50	61831141.86	6667731.65
33	51	142134257.2	21144088.73
34	52	179014313	22867657.18
35	53	119374980.1	13958150.3
36	54	110671948.5	13181600.12
37	55	24733112.6	2655636.37
38	56	134502084.5	14038336.25
39	59	159779409.8	17547535.78
40	60	25468179.01	3446371.06
41	61	3222751	370844
42	62	176491366.6	24605648.45
43	64	71088555.46	8519931.85
44	65	42754315.77	5834030.57
45	67	133280655.2	16335654.18
46	68	122429250.6	14814421.34
47	69	42428693.48	4321570.76
48	70	208876168.2	24827676.09
49	71	188520874	25647163.75
50	72	124429562.5	13238866.78
51	73	223570965.7	23415906.1
52	74	208630739.2	24924836.57
53	75	133655896.2	23319310.48
54	76	154271050.7	24574168.92
55	77	172848676.7	26114028.66
56	78	183315706.4	19919748.12
57	80	243290585.4	31289688.88
58	81	184964901.6	22309376.97
59	83	180631359	26727650.28
60	84	179095915.8	23240362.1
61	86	184562949.2	27044047.64
62	88	127839691.4	16672482.8
63	89	119432698.9	16473968.84
64	90	116712280.2	15100479.68
65	91	128898986.6	14495885.74
66	92	111312943.9	14153531.42
67	93	157299732.5	23061220.94
68	94	154288893.1	17611149.01
69	95	138907320.8	20529525.35
70	97	72680890.39	9231755.1
71	98	213774432.3	43883707.19
72	100	206883044.5	44605491.21
73	101	194983334.1	42779929.51
74	102	262117057.8	49211597.59
75	103	153223971.1	32950742.78
76	104	154139053.9	21713056.21
77	105	158233512.3	31250117.95
78	106	120144693.2	19546290.87
79	107	213531886.5	30370811.42
80	108	46805911.42	5511765.16
81	109	287334310.9	36266785.44
82	110	168151607	24119372.11
83	111	152464941.9	12862511.46
84	112	227629046.4	53479685.92
85	113	188660716.5	22025860.32
86	114	178110843.9	36046765.46
87	115	186735152.6	44713876.78
88	116	143433065.1	19655299.74
89	117	114079386.6	12052225.65
90	118	84432740.76	7817664.84
91	119	130459247.9	15061518.37
92	121	213197692.4	48032037.11
93	122	261616460.6	54339775.4
94	123	163518960.7	31102622.98

Fig: Sales Data Pivot Table for Food and Non-Food Categories (1)

93	122	261616460.6	54339775.4
94	123	163518960.7	31102622.98
95	124	189960885.9	29947677.78
96	126	208454186.1	45560774.78
97	128	212797090.1	40076796.42
98	129	167251909.8	34804127.46
99	130	137953720.3	14669654.34
100	131	167621583.5	36833931.23
101	132	197221583.1	36128022.34
102	133	112849119.8	22818940.67
103	134	95139217.44	20762569.7
104	135	24974921.51	3975885.78
105	136	99199713.21	24864127.78
106	137	201538550.7	31303736.4
107	139	121515892.1	22244256.73
108	140	13840656.76	2687430.9
109	141	13640608.32	2761975.64
110	142	14370900.09	2539577.57
111	143	23633670.62	5052111.03
112	144	18597672.45	2627429.98
113	146	9214400.82	1365706.64
114	301	350673942.1	79534297.29
115	302	280629937.8	61806448.4
116	303	312741097.3	70014434.41
117	304	348369189.4	79809888.99
118	305	235163528.8	44105248.36
119	306	259412437.8	60365398.09
120	307	260861078.3	52633765.56
121	308	181341707.4	37704620.41
122	309	259860064.1	59001373.31
123	310	209856138.6	46282693.6
124	311	152115528	38120237.08
125	312	221316139.3	52142802.79
126	313	176036424.6	34655630.35
127	314	191022212.6	35211720.87
128	315	283537999.1	69934017.92
129	316	98926154.83	19676558.14
130	318	49133435.36	9040923.29
131	333	19	3
132	700	23	6
133	964	1138.33	273.54
134	**Grand Total**	**18121594908**	**2968321282**

Fig: Sales Data Pivot Table for Food and Non-Food Categories (2)

3. What is the annual customer inflow for each store allowing for insights into customer behavior and store performance?

   Justification:

   To address the analysis regarding customer counts, we need to utilize the CustCount column from the Fact_Customer_Count table. This data will be aggregated by grouping based on the Store_ID and Week_Number columns, sourced from the Dim_Store and Dim_Date tables, respectively. This approach will allow us to conduct a more detailed analysis of customer counts across different stores and time periods.

The pivot table displays customer count data segmented by year, with filters applied for both year and week.

Week of the year	(All)
STORE	(All)

Row Labels	Sum of CUSTCOUN
1987	215841.73
1988	86645981.93
1989	97861853.65
1990	99384416
1991	104340958
1992	106684354
1993	104875533
1994	97906188
1995	95740992
1996	97392644
1997	32244775
Grand Total	**923293537.3**

Fig: Customer Count Data Pivot Table Segmented by Year with Year and Week Filters

4. What are the top-selling departments (e.g., Bakery, Meat, Deli) across stores, and how do they vary by store and week?

   Justification: To answer this question, we need to analyze the sales data at the product category level, focusing on departments such as Bakery, Meat, and Deli. The data required is present in the Fact_Sales table, which records sales amounts for each product (Sales) and is linked to the Dim_Product table through Product_ID, where Product_Desc can categorize products into departments like Bakery, Meat, and Deli. By combining this data with the Dim_Store and Dim_Date tables (linked via Store # and Week #, respectively), we can aggregate sales by store and week.

   The pivot table below displays the average sales by department, with filters for Store Number and Week Number, giving stakeholders a clear understanding of department-level performance over time.

Values	
Sales of GROCERY	25472.5278
Sales of WINE	490.849493
Sales of BEER	824.564652
Sales of COSMETIC	165.774232
Sales of JEWELRY	17.8417515
Sales of PHARMACY	977.5967
Sales of BAKERY	1451.06538
Sales of CHEESE	298.698676
Sales of DELI	2555.8164
Sales of FLORAL	496.221523
Sales of PRODUCE	5760.66726
Sales of FISH	753.338585
Sales of DAIRY	5166.60515
Sales of FROZEN	3910.1081
Sales of MEAT	5865.46699

Fig: Average Sales Data Pivot Table

5. Determine the sales of a given category during the weeks of special events.

   Justification: To answer this question, we need to analyze sales data for specific product categories during designated special events, as these periods often see increased customer traffic and spending. The required data can be extracted from the Fact_Sales table, where sales figures for each product category can be identified using the Product_ID linked to the Dim_Product table. By combining this with the Dim_Date table, we can filter sales records to isolate those occurring during special events like Halloween, Thanksgiving, and New Year.

   The pivot table display total sales for each product category during special events, with filters available for specific events and Store Numbers, providing a comprehensive overview of category performance during high-traffic periods.

Row Labels	Sum of CAMERA	Sum of COSMETIC	Sum of HABA	Sum of BAKERY	Sum of PHARMACY	Sum of BEER	Sum of WINE	Sum of MEAT
1987	0	0	0	4592.18	0	13233.77	0	28062.32
1988	1481564.3	2793802.15	11985757.8	37420096.72	6594730.15	33730113.69	1674788.85	194781008.1
1989	1544838.38	4540557.01	49452303.08	45280061.73	10371910.77	20922872.25	13086642.23	200998663
1990	1742050.45	5888512.67	56227637.53	47023454.91	15149334.83	22670908.31	14875793	205725762.3
1991	2037994.35	6719181	66373751.06	48903740.14	21761895.21	26193141.08	16019131.14	214743722.2
1992	1902048.53	7296986.16	79909990.24	52591025.21	31162555.89	27829738.59	18408560.45	211983261
1993	2196794.62	6951340.65	126595176.8	54031688.41	40907630.48	28644919.19	18831875.95	207491147.4
1994	2702151.08	6592177.3	124139587.2	54536231.36	47947360.44	30179810.46	20162115.58	201025816.7
1995	3499491.37	5999165.44	126668536.1	54839225.12	56454507.56	32729201.96	22058673.43	194138303.7
1996	4057030.77	5287156.35	130298047.8	56894129.81	63443242.1	34890088.41	24690145.67	200349886.6
1997	1453730.75	1662694.3	47505120.55	18802321.33	23070362.94	9457990.13	9288865.48	69878855.94
Grand Total	**22617694.6**	**53731573.03**	**819155908.1**	**470326566.9**	**316863530.4**	**267262017.8**	**159096591.8**	**1901144489**

Fig: Total Sales Data Pivot Table for each product category during each year

Mapping Tables

Stg_Store

Source layer to Stage layer:

Source Data	Source Data Field	Mapping	Staging Table Type	Staging Table Name	Attributes
dominicks_store_descriptions.csv	Store	Copy	Relation	stg_store	Store
	City				City
	Price Tier				Price Tier
	Zone				Zone
	Zip Code				ZipCode
	Address				Address

Note: We created the file dominicks_store_descriptions.csv from the pdf report provided for the use case.

Dim_Store

Stage layer to Dimension layer:

Source Data in Staging	Staging table Data Field	Mapping	Data Mart Table type	Table Name	Attribute
stg_store		Surrogate Key	Dimension Table	Dim_Store	Store_Key
	Store	Copy			Store
	City				City
	Price Tier				Price Tier
	Zone				Zone
	ZipCode				ZipCode
	Address				Address

Stg_Demo

Source layer to Stage layer:

Source Data	Source Data Field	Mapping	Staging Table Type	Staging Table Name	Attributes
DEMO.CSV	MMID	Copy	Relation	stg_demo	MMID
	SHPCONS				SHPCONS
	SHPHUR				SHPHUR
	SHPAVID				SHPAVID
	SHPKSTR				SHPKSTR
	SHPUNFT				SHPUNFT
	SHPBIRD				SHPBIRD

Dim_Demo

Stage layer to Dimension layer:

Source Data in Staging	Staging table Data Field	Mapping	Data Mart Table type	Table Name	Attribute
stg_demo		Surrogate Key	Dimension Table	Dim_Demo	Demo_Key
	MMID	Copy			MMID
	SHPCONS				SHPCONS
	SHPHUR				SHPHUR
	SHPAVID				SHPAVID
	SHPKSTR				SHPKSTR
	SHPUNFT				SHPUNFT
	SHPBIRD				SHPBIRD

Stg_Weeks_Decode_Table

Source layer to Stage layer:

Source Data	Source Data Field	Mapping	Staging Table Type	Staging Table Name	Attributes
File1: weeks_decode_table.csv	Start Date	Copy	Relation	**Table1:** stg_weeks_decode_table **Table2:** stg_ccount	Start Date
	End Date				End Date
	Special Events				Special Events
	Week				Week
	Year				Year
	Year&Week				Year&Week

Note:

- We created the file weeks_decode_table.csv from the pdf report provided for the use case.

Stg_Ccount

Source layer to Stage layer:

Source Data	Source Data Field	Mapping	Staging Table Type	Staging Table Name	Attributes
CCount.csv	Store #	Copy	Relational	stg_ccount	Store #
	Date_ID				Date_ID
	Grocery				Grocery
	Dairy				Dairy
	Frozen				Frozen
	Meat				Meat
	Custcoun				Custcoun

Dim_Date

Stage layer to Dimension layer:

Source Data in Staging	Staging table Data Field	Mapping	Data Mart Table type	Table Name	Attribute
Table1: stg_weeks_decode_table **Table2:** stg_ccount		Surrogate Key	Dimension Table	Dim_Date	Date_Key
	Table2.Date_ID	Copy + Transformation (Join Table 1 & Table 2 by Date over WeekYear)			Date_ID
	Year				Year
	Month				Month
	Quarter				Quarter
	WeekYear				WeekYear
	Week_of_the_Year				Week_of_the_Year
	Table1.Special Events				Table1.Special Events

Stg_Product

Source layer to Stage layer

Source Data	Source Data Field	Mapping	Staging Table Type	Staging Table Name	Attributes
Multiple upcxxx.csv files where xxx refers to three-letter acronym used for file Product_Index_csv (created based on manual)	Surrogate Key (Generated uniquely for each record)	copy	Relational	stg_product	Product_ID
	xxx source file for the respective product				Product_Description
	Product_Group (from the file)				Product_Group

Dim_Product

Stage Layer to Dimension Layer

Source Data in Staging	Staging table Data Field	Mapping	Data Mart Table type	Table Name	Attribute
Product_stg		Surrogate Key	Dimension Table	Dim_Product	Product_Key
	Product_ID	copy			Product_ID
	Product_Description				Product_Description
	Product_Group				Product_Group

Fact_Sales

Source to Stage Layer: We have loaded the required stage tables stg_ccount before. We are going to use the same stage table here as well.

Stage layer to Dimension layer:

Source Data in Staging	Staging table Data Field	Mapping	Data Mart Table type	Table Name	Attribute
stg_ccount		Surrogate Key	Fact Table	Fact_Sales	Sales_ID
	Store #	Copy			Store #
	Date_ID				Date_ID
	Product_ID	Unpivot Transformation			Product_ID
	Sales				Sales

Fact_Customer_Count

Source to Stage Layer: We have loaded the required stage tables stg_demo & stg_ccount before. We are going to use the same stage table here as well.

Stage layer to Dimension layer:

Source Data in Staging	Staging table Data Field	Mapping	Data Mart Table type	Table Name	Attribute
stg_ccount_demo		Surrogate Key	Fact Table	Fact_Customer_Count	Record_ID
	Store #	Copy			Store #
	Date_ID				Date_ID
	MMID				MMID
	CustCoun				CustCoun

Note: Temporary/More stage tables can be used to combine data from one file to another.

Physical Design

The physical design of a data warehouse focuses on how data is stored, organized, and retrieved.The aim is to optimize the logical design in a way that aligns with the specific physical architecture of the database and hardware environment. It involves structuring the data for efficient storage, retrieval, and management while considering performance, scalability, and space requirements. The tables are structured based on how the data will be queried, with indexes applied to key columns to speed up search and retrieval. Partitioning can be used to divide large tables into smaller sections, making it easier and faster to access specific data without scanning the entire table. For results that are derived but are accessed frequently, materialized views store precomputed summaries, reducing the need to process large amounts of data repeatedly.

The goal of our physical design discussed below is to create a system that can handle large data volumes and complex queries while maintaining fast access and scalability. This design ensures that the data warehouse operates efficiently and reliably in high-demand environments.

1. **Data Aggregation Plan:** The objective of our data aggregation plan is to improve query performance by precomputing and storing summary or aggregated data. Fact tables are designed to store data at different levels of granularity, ensuring that aggregations are readily available without the need for real-time computation. Although this approach may introduce data redundancy, it significantly reduces query times by providing quick access to pre-aggregated data, especially in high-demand environments. We are using multiple levels of aggregation to suit different reporting needs like Daily sales by product category, Weekly customer visits by store, etc. having the aggregated level computations performed and stored in fact tables, the results are readily available for reporting purposes.

2. **Indexing Plan:** An indexing plan is necessary to ensure that our data warehouse operates efficiently and handles large queries quickly. By using indexes, we reduce the time it takes to search, sort, and filter through large volumes of data. Without these indexes, our system would have to scan entire tables, which slows down query performance. With B-tree indexes on key columns (date and store_id) we speed up range queries like from one date to another. Clustered indexes on fact tables allow us to physically organize data for quicker access, and foreign key indexes make joins between tables faster. Partitioned indexes help us target specific sections of data, further improving performance.

3. **Data Standardization Plan:** Our data standardization plan aims to ensure consistency in data formats, types, and structures across the data warehouse. This consistency supports efficient queries and helps us avoid discrepancies in our data. We will implement standard data types across all tables. For example, we will use VARCHAR for

textual data, DATE or TIMESTAMP for date fields, and INTEGER or BIGINT for numeric values. We will also standardize how we handle date and time fields. All date fields will be stored in a consistent format, such as YYYY-MM-DD. To maintain consistency in key data entities like customers, products, and stores, we will implement a master data management (MDM) system. This MDM will serve as a single source of truth, enforcing standard codes, formats, and naming conventions. During our ETL process, we will focus on data cleansing to remove duplicates, invalid records, and inconsistencies such as missing values. We will establish validation rules to enforce our data quality standards.

4. **Data Storage Plan:** The objective of our data storage plan is to ensure that data is stored efficiently, is easily accessible when needed, and is protected against loss or unauthorized access. A well-defined storage plan enhances data management, supports analytics efforts, and ensures the data warehouse can scale to meet future demands. As the volume of data currently available can be handled in a single table for now, We can use the common storage process. But, as the volume increases over time, data partitioning can be implemented. This involves dividing large tables into smaller, more manageable sections based on criteria like Date Range, Zone based sales information. It is important to identify data archival policies like after how long the data can be archived taking into account business requirements. Another important thing to consider is our backup frequency. We will schedule regular backups, which will include weekly jobs to do delta backup of data.

With the four plans—data aggregation, indexing, data standardization, and data storage—we can create a data warehouse physical design that performs well and is reliable. Integrating these plans ensures our data is stored effectively, easily accessible, consistent, and secure as discussed above.

ETL Plan

Data Tables	Source File Name
Products	Prod_Index.csv
Demographics	demo.csv
Date	ccount.csv & weeks_decode_table.csv
Store	dominicks_store_descriptions.csv

Staging Tables

Stg_Demo

DW Stage Table	DW Stage Column	Datatype
dbo.stg_demo	MMID	VARCHAR(50)
	STORE	VARCHAR(50)
	SHPCONS	VARCHAR(50)
	SHPHURR	VARCHAR(50)
	SHPAVID	VARCHAR(50)
	SHPKSTR	VARCHAR(50)
	SHPUNFT	VARCHAR(50)
	SHPBIRD	VARCHAR(50)

Stg_Store

DW Stage Table	DW Stage Column	Datatype
dbo.stg_store	STORE	VARCHAR(50)
	CITY	VARCHAR(50)
	PRICE TIER	VARCHAR(50)
	ZONE	VARCHAR(50)
	ZIP CODE	VARCHAR(50)

	ADDRESS	VARCHAR(50)

Stg_Product

DW Stage Table	DW Stage Column	Datatype
dbo.stg_product	PRODUCT_ID	VARCHAR(50)
	PRODUCT_DESCRIPTION	VARCHAR(50)
	PRODUCT_GROUP	VARCHAR(50)

Stg_Ccount

DW Stage Table	DW Stage Column	Datatype
dbo.stg_ccount	STORE	VARCHAR(50)
	DATE	VARCHAR(50)
	GROCERY	VARCHAR(50)
	DAIRY	VARCHAR(50)
	FROZEN	VARCHAR(50)
	FISH	VARCHAR(50)
	GROCCOUP	VARCHAR(50)
	DAIRCOUP	VARCHAR(50)
	FROZCOUP	VARCHAR(50)
	FISHCOUP	VARCHAR(50)

Stg_Weeks_Decode_Table

DW Stage Table	DW Stage Column	Datatype
dbo.stg_weeks_decode_table	START DATE	VARCHAR(50)
	END DATE	VARCHAR(50)

	SPECIAL EVENTS	VARCHAR(50)
	WEEK	VARCHAR(50)
	YEAR	VARCHAR(50)
	YEAR & WEEK	VARCHAR(50)

Dimension Tables

Dim Demo

DW Target Table	DW Target Column	Datatype	Column Type
dbo.dim_demo	Demo_Key	INT IDENTITY(1,1)	Surrogate Key, Primary Key
	MMID	NUMERIC(38,0)	
	SHPCONS	NUMERIC(38,4)	
	SHPHURR	NUMERIC(38,4)	
	SHPAVID	NUMERIC(38,4)	
	SHPKSTR	NUMERIC(38,4)	
	SHPUNFT	NUMERIC(38,4)	
	SHPBIRD	NUMERIC(38,4)	

Dim Store

DW Target Table	DW Target Column	Datatype	Column Type
dbo.dim_store	Store_Key	INT IDENTITY(1,1)	Surrogate Key, Primary Key
	Store	NUMERIC(10,0)	
	CITY	VARCHAR(50)	
	PRICE TIER	VARCHAR(50)	
	ZONE	NUMERIC(10,0)	
	ZIP CODE	NUMERIC(10,0)	

	ADDRESS	VARCHAR(50)	

Dim Product

DW Target Table	DW Target Column	Datatype	Column Type
dbo.dim_product	Product_Key	INT IDENTITY(1,1)	Surrogate Key, Primary Key
	PRODUCT_ID	VARCHAR(50)	
	PRODUCT_DESCRIPTION	VARCHAR(50)	
	PRODUCT_GROUP	VARCHAR(50)	

Dim Date

DW Target Table	DW Target Column	Datatype	Column Type
dbo.dim_date	Date_Key	INT IDENTITY(1,1)	Surrogate Key, Primary Key
	DATE	DATE	
	YEAR	NVARCHAR(100)	
	MONTH	NVARCHAR(100)	
	QUARTER	NVARCHAR(100)	
	WEEKOFTHEYEAR	NVARCHAR(100)	
	SPECIAL EVENTS	NVARCHAR(100)	
	WEEKYEAR	NVARCHAR(100)	

Fact Tables

Fact Sales

DW Target Table	DW Target Column	Datatype	Column Type
dbo.fact_sales	SALES_ID	INT IDENTITY(1,1)	Surrogate Key,

			Primary Key
	STORE	INT	Foreign Key
	DATE	DATE	Foreign Key
	PRODUCT_ID	VARCHAR(50)	Foreign Key
	SALES	NUMERIC(18,0)	

Fact Customer Count

DW Target Table	DW Target Column	Datatype	Column Type
dbo.fact_customer_count	RECORD_ID	INT IDENTITY(1,1)	Surrogate Key, Primary Key
	MMID	NUMERIC(38,0)	Foreign Key
	STORE	NUMERIC(10,0)	Foreign Key
	DATE	DATE	Foreign Key
	CUSTCOUN	INT	

Data Mappings for Data Elements

Source to Staging

Source	Source Attribute	Staging Table	Staging Table Attributes	Mapping Function
demo.csv	MMID	stg_demo	MMID	Direct Copy
	STORE		STORE	Direct Copy
	SHPCONS		SHPCONS	Direct Copy
	SHPHURR		SHPHURR	Direct Copy
	SHPAVID		SHPAVID	Direct Copy
	SHPKSTR		SHPKSTR	Direct Copy
	SHPUNFT		SHPUNFT	Direct Copy

	SHPBIRD		SHPBIRD	Direct Copy
dominicks_store_descriptions.csv	STORE	stg_store	STORE	Direct Copy
	CITY		CITY	Direct Copy
	PRICE TIER		PRICE TIER	Direct Copy
	ZONE		ZONE	Direct Copy
	ZIP CODE		ZIP CODE	Direct Copy
	ADDRESS		ADDRESS	Direct Copy
Prod_Index.csv	PRODUCT_ID	stg_product	PRODUCT_ID	Direct Copy
	PRODUCT_DESCRIPTION		PRODUCT_DESCRIPTION	Direct Copy
	PRODUCT_GROUP		PRODUCT_GROUP	Direct Copy
ccount.csv	STORE	stg_ccount_temp	STORE	Direct Copy
	DATE		DATE	Direct Copy
	GROCERY		GROCERY	Direct Copy
	DAIRY		DAIRY	Direct Copy
	FROZEN		FROZEN	Direct Copy
	FISH		FISH	Direct Copy
	GROCCOUP		GROCCOUP	Direct Copy
	DAIRCOUP		DAIRCOUP	Direct Copy
	FROZCOUP		FROZCOUP	Direct Copy
	FISHCOUP		FISHCOUP	Direct Copy
weeks_decode_table.csv	START DATE	stg_weeks_decode_table	START DATE	Direct Copy
	END DATE		END DATE	Direct Copy
	SPECIAL EVENTS		SPECIAL EVENTS	Direct Copy
	WEEK		WEEK	Direct Copy

	YEAR		YEAR	Direct Copy
	YEAR & WEEK		YEAR & WEEK	Direct Copy

Staging to Data Marts

Source data in staging	**Staging Table Data Field**	**Data Mart Table Type**	**Table Name**	**Attribute**	**Mapping**
stg_demo		Dimension Table	dim_demo	Demo_Key	Surrogate Key
	MMID			MMID	Direct Copy
	SHPCONS			SHPCONS	Direct Copy
	SHPHURR			SHPHURR	Direct Copy
	SHPAVID			SHPAVID	Direct Copy
	SHPKSTR			SHPKSTR	Direct Copy
	SHPUNFT			SHPUNFT	Direct Copy
	SHPBIRD			SHPBIRD	Direct Copy
stg_store		Dimension Table	dim_store	Store_Key	Surrogate Key
	Store			Store	Direct Copy
	CITY			CITY	Direct Copy
	PRICE TIER			PRICE TIER	Direct Copy
	ZONE			ZONE	Direct Copy
	ZIP CODE			ZIP CODE	Direct Copy
	ADDRESS			ADDRESS	Direct Copy
stg_product		Dimension Table	dim_product	Product_Key	Surrogate Key
	PRODUCT_ID			PRODUCT_ID	Direct Copy
	PRODUCT_DESCRIPTI			PRODUCT_DESCRIPTIO	Direct Copy

	ON			N	
	PRODUCT_GROUP			PRODUCT_GROUP	Direct Copy
Stg_ccount_temp		Dimension Table	dim_date	Date_Key	Surrogate Key
	DATE			DATE	Direct Copy
				YEAR	Derived Attribute
				MONTH	Derived Attribute
				QUARTER	Derived Attribute
				WEEKOFTHEYEAR	Derived Attribute
				WEEK YEAR	Derived Attribute
stg_weeks_decode_table	SPECIAL EVENTS			SPECIAL EVENTS	Direct Copy
stg_ccount_temp		Fact Table	fact_sales	Sales_ID	Surrogate Key
	STORE			STORE	Direct Copy
	DATE			DATE	Direct Copy
	GROCERY			PRODUCT_ID	Unpivot Transformation (The product and sales columns are visually separated for presentation purposes. This does not indicate that they originate from different sets of columns.)
	DAIRY				
	FROZEN				
	FISH				
	GROCCOUP			SALES	
	DAIRCOUP				
	FROZCOUP				
	FISHCOUP,				

	etc. (There are more columns)				
		Fact Table	fact_customer_count	Record_ID	Surrogate Key
stg_demo	MMID			MMID	Direct Copy
stg_ccount_temp	STORE			STORE	Direct Copy
	DATE			DATE	Direct Copy
	CUSTCOUN			CUSTCOUN	Direct Copy

ETL Rules

The below information outlines the process to be followed to extract, clean, and transform data for our data warehousing project. The goal is to prepare high-quality data that is consistent and reliable for analytics and reporting.

Data Extraction Rules

Our data extraction focuses on creating a uniform and structured staging environment that accurately mirrors the source files and facilitated later data processing steps.

- **File Format Standardization:** All data sources(including files that were in incompatible formats) were converted to CSV for consistency to ensure compatibility with the staging environment.
- **Data Replication in Staging Area:** The source files were replicated exactly into tables in the staging area to serve as a secure, reliable space for preliminary processing.
- **Attribute Filtering and Integration:** Once data is in the staging area, we filtered necessary attributes from each table. These filtered attributes were then loaded into appropriate dimension tables, integrating relevant data for analytical purposes.

Data Cleaning Rules

Data cleaning was conducted prior to transformation to ensure that the data was consistent, readable, and structured properly for integration into data marts.

- **Elimination of NULL Values:** NULL values were addressed by either removing incomplete records or imputing default values, as appropriate for the data context.

- **Surrogate Key Creation:** Surrogate keys were generated to uniquely identify rows in fact and dimension tables, supporting referential integrity and simplifying joins in data marts.
- **Trimming Extra Characters and Spaces:** Unnecessary characters and spaces in text fields were removed to standardize formats and prevent data misinterpretation.
- **Removal of Unreadable Characters:** Identified and eliminated any non-standard characters that could impact data parsing and readability in the database.

Data Transformation Rules

After data cleaning, we transformed the data to enhance its usability for analytical reporting. This step included data type adjustments, derived columns, unpivoting operations, and other structural modifications.

- **Data Type Conversion:**
  - Strings representing dates were converted into standard date formats.
  - Strings with numeric values were cast into appropriate numerical data types for accurate calculations.
- **Derived Columns:**
  - Created new columns for improved time-based analysis
    - Week of the Year**:** Derived from the date column to support weekly insights.
    - Year**:** Extracted from the date for annual aggregations.
    - Month**:** Added to facilitate monthly reporting.
    - Quarter**:** Derived for quarterly analysis.
- **Unpivoting Data:**
  - Applied unpivot transformations to convert wide tables into a long format, enabling better alignment with data mart requirements and improving analytical flexibility. This unpivoting step allowed for efficient storage and simplified query execution for multi-dimensional analyses.
- **Look-Up Operations:**
  - Applied look-up operations to enrich data by linking to other tables for consistent, standardized values. These operations ensured that IDs, codes, and other references were transformed into meaningful values.
  - **Example:** Product Look-Up - Mapped product codes to product descriptions.

Data Loading Rules:

The data loading phase involved moving the cleaned and transformed data from the staging area to the data warehouse's final data marts, with considerations for future scalability and analytical requirements. This section outlines our data loading rules and considerations for future enhancements.

- **Loading to Final Data Marts**
  - After completing data cleaning and transformation, we loaded the data into well-defined data marts. Each data mart was created to support specific analytical needs, structured to accommodate the final attributes necessary for reporting and querying across dimensions (such as time, product, location, and customer).

- **Considerations for Aggregations**
  - Although our current data marts store data at the most granular level, we designed the architecture with future aggregation needs in mind. This includes storing dimensional data and measures in ways that allow for easy roll-ups and drill-downs.
  - Future Aggregation Needs: We plan to add pre-aggregated tables to support faster analytics and reporting for commonly requested metrics, such as monthly sales totals, quarterly revenue, and customer purchase frequency.

- **Optimization for Query Performance**
  - **Indexing Key Columns:** We indexed frequently used columns, such as surrogate keys in dimension tables, to improve query performance. These indexes support faster retrieval in joins and help optimize analytics.

- **Referential Integrity and Key Constraints**
  - **Foreign Key Constraints:** We implemented foreign key constraints between fact and dimension tables to ensure referential integrity, guaranteeing that all fact data references a valid dimension record.
  - **Surrogate Keys in Fact Tables:** Surrogate keys were assigned to the fact tables, which serve as unique identifiers, ensuring consistent data relationships across multiple dimensions.

- **Data Verification Post-Load**
  - After loading the data, verification checks were conducted to confirm data integrity:
    - **Row Counts:** Verified row counts between the staging area and the final tables to ensure complete data migration.
    - **Data Accuracy:** Performed spot-checks on critical metrics (e.g., totals, averages) to confirm accuracy and consistency between the source data and final tables.

- **Error Logging and Handling**
  - Developed a logging system to track errors encountered during data loading, enabling efficient troubleshooting.

Summary

The extraction, cleaning, and transformation processes were designed to ensure that the data loaded into data marts was consistent, accurate, and structured for analytical use. Through rigorous data standardization, cleaning, and transformations—including unpivoting where appropriate—we have achieved a solid, high-quality data foundation suitable for robust analysis and reporting.

The data loading phase focused on ensuring data accuracy, performance optimization, and scalability. With future aggregations in mind, the data marts were structured for flexibility, allowing for both detailed and summarized data to support a wide range of reporting needs. By indexing key columns, partitioning where needed, and implementing thorough data verification checks, we ensured that the data warehouse was not only reliable but also optimized for future analytical demands.

Data Extraction, Transformation & Loading Procedure

As we explained above, stage tables are replications of files. Let us see the source file for each table.

- stg_demo → Demo.csv – Provided in the dataset
- stg_store → Dominic Store Descriptions.csv – Copied from the manual
- stg_product → Prod_Index.csv – Created from the manual and business requirement analysis
- stg_ccount_temp → ccount.csv - Provided in the dataset
- stg_weeks_decode_table → weeks_decode_table.csv - Copied from the manual

Now let us look at the procedure for loading data from stage to dim/fact tables

- **Dim_product:**
  - Created product dimensional table in the ISTM_637_602_2_Sales_DM database with a surrogate key
  - Loaded data by copying from stg_product table with a surrogate key
- **Dim_store:**
  - Created store dimensional table in the ISTM_637_602_2_Sales_DM & ISTM_637_602_2_Customer_DM databases with a surrogate key
  - Loaded data by copying from stg_store table
  - Transformed the data types for the columns - Store, Zone, ZipCode

- **Dim_Date:**
  - Created date dimensional table in the ISTM_637_602_2_Sales_DM & ISTM_637_602_2_Customer_DM databases with a surrogate key
  - Loaded data by copying data from stg_ccount & stg_weeks decode tables by joining them and also deriving the columns like, week, month, quarter, year from the date column
  - Transformed the data types for the columns - Date

- **Dim_Demo:**
  - Created demograpic dimensional table in the ISTM_637_602_2_Customer_DM database with a surrogate key
  - Loaded data by copying from stg_demo
  - Transformed the data types for the columns - MMID, SHPCONS, SHPHURR, SHPAVID, SHPKSTR, SHPUNFT, SHPBIRD

- **Fact_sales:**
  - Created Sales Fact table in the ISTM_637_602_2_Sales_DM database with a surrogate key
  - Loaded data by copying from stg_ccount by unpivoting the data based on the columns - Store, Date, Product_ID, Sales

- **Fact_Customers_Count:**
  - Created Fact Customers Count table in the ISTM_637_602_2_Customer_DM database with a surrogate key
  - Loaded data by copying from stg_ccount by filtering only to required columns
  - Transformed the data types for the columns - MMID, Store, Date, CustCoun

ETL Implementation

Staging Tables

Stg_store

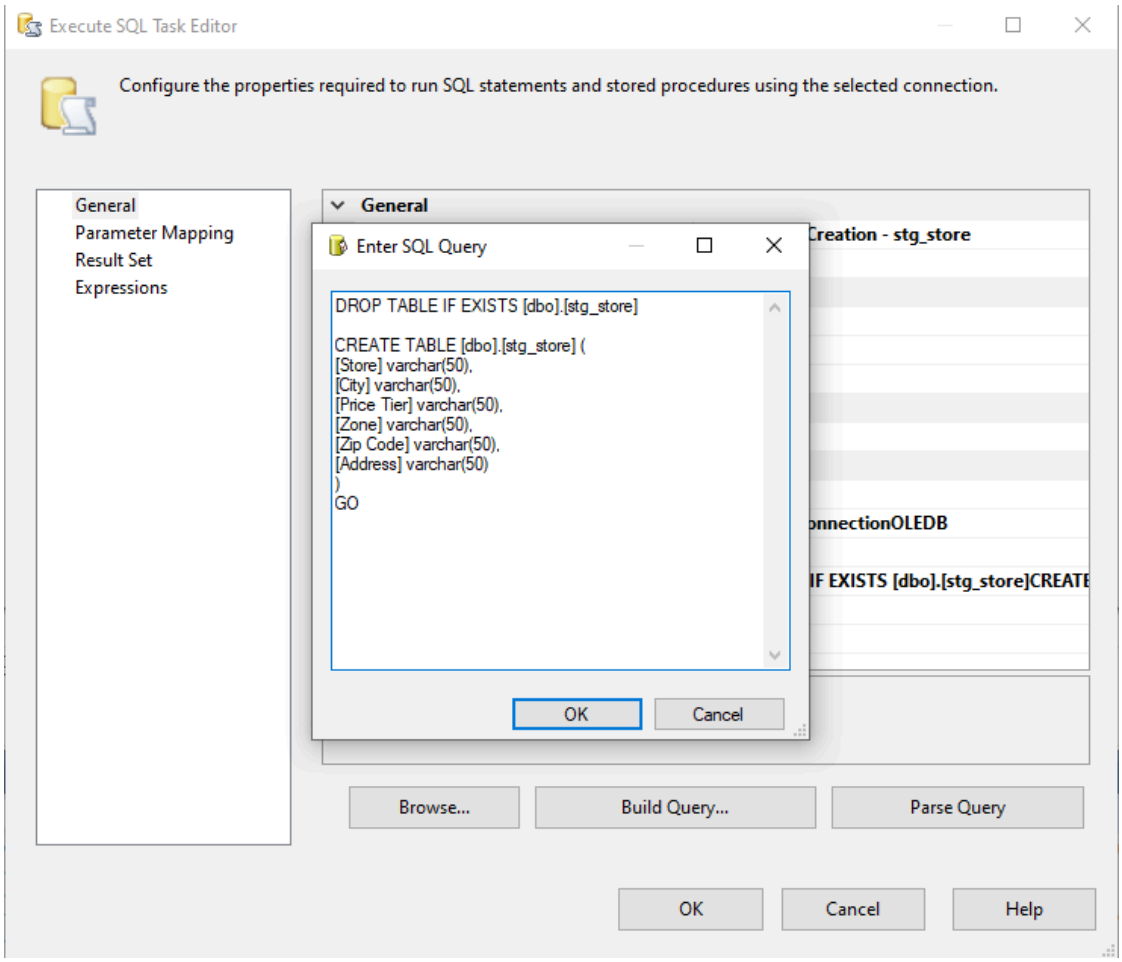


Fig: stg_store Table Creation

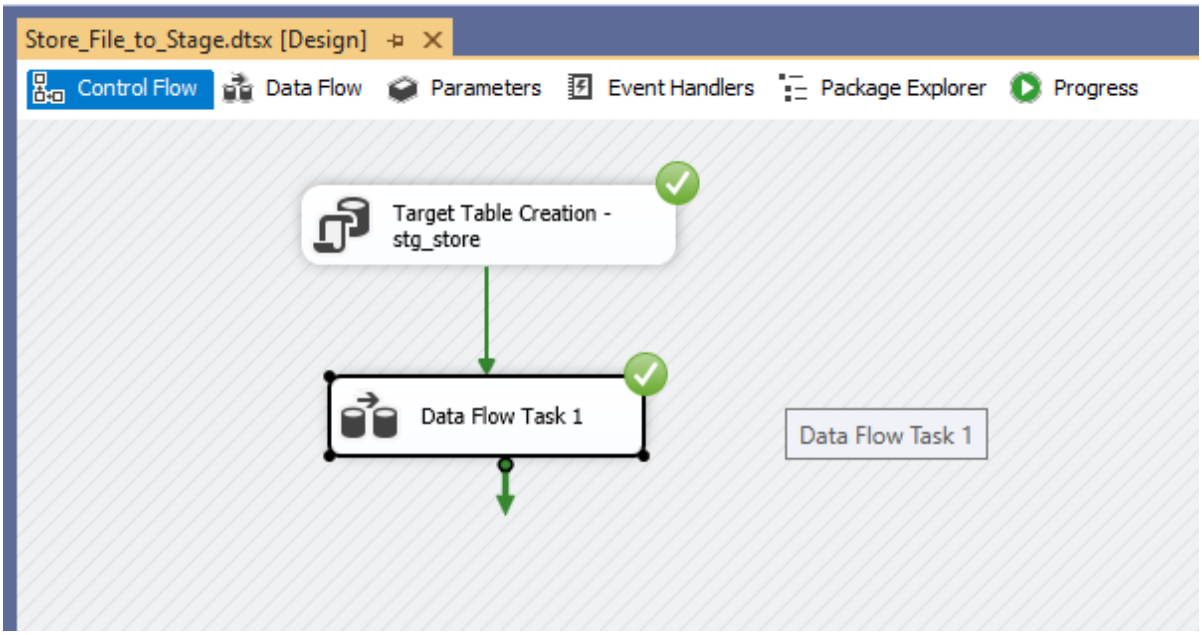


Fig: stg_store Control Flow

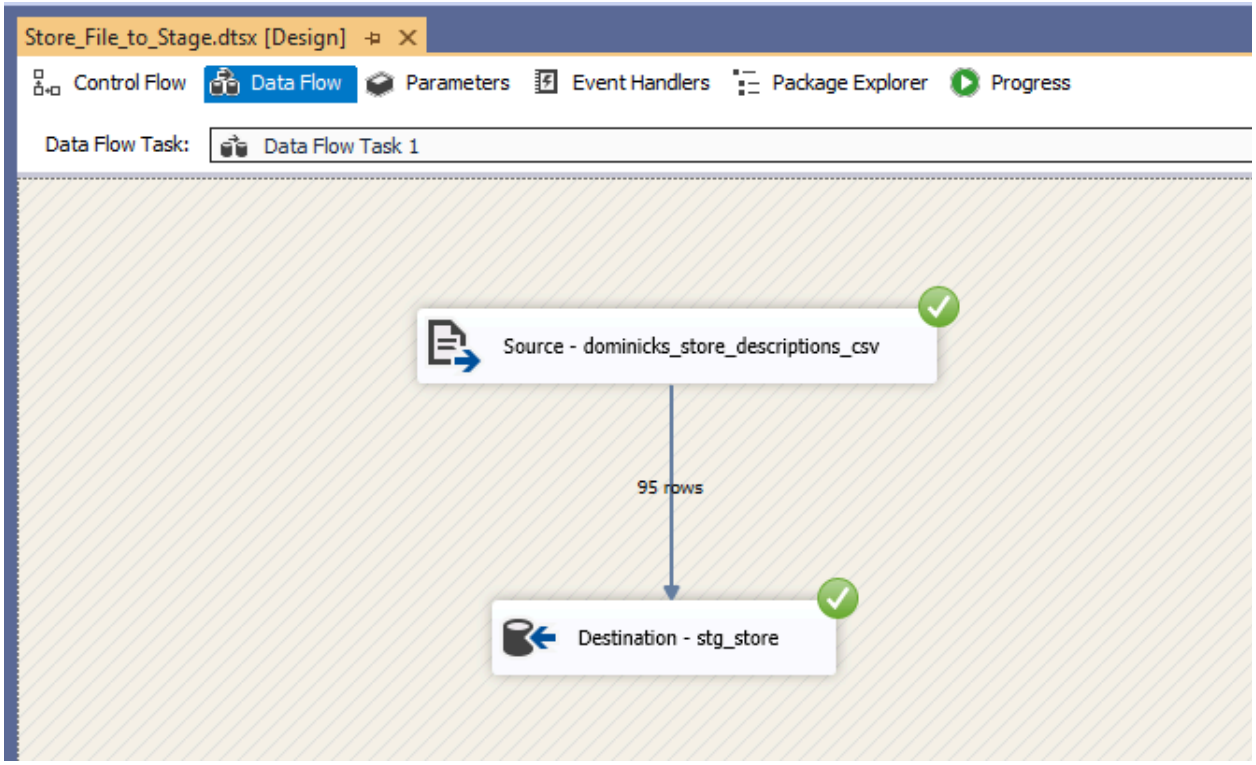
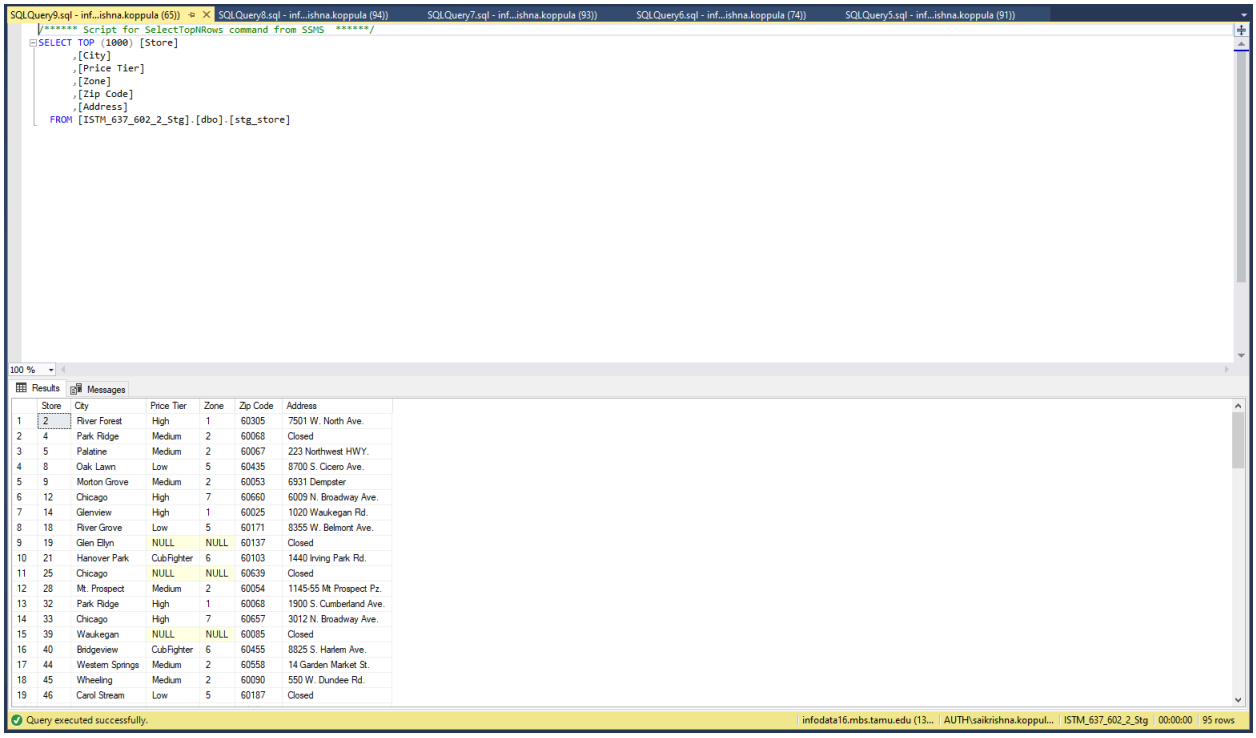


Fig : stg_store Data Flow



	Store	City	Price Tier	Zone	Zip Code	Address
1	2	River Forest	High	1	60305	7501 W. North Ave.
2	4	Park Ridge	Medium	2	60068	Closed
3	5	Palatine	Medium	2	60067	223 Northwest HWY.
4	8	Oak Lawn	Low	5	60435	8700 S. Cicero Ave.
5	9	Morton Grove	Medium	2	60053	6931 Dempster
6	12	Chicago	High	7	60660	6009 N. Broadway Ave.
7	14	Glenview	High	1	60025	1020 Waukegan Rd.
8	18	River Grove	Low	5	60171	8355 W. Belmont Ave.
9	19	Glen Ellyn	NULL	NULL	60137	Closed
10	21	Hanover Park	CubFighter	6	60103	1440 Irving Park Rd.
11	25	Chicago	NULL	NULL	60639	Closed
12	28	Mt. Prospect	Medium	2	60054	1145-55 Mt Prospect Pz.
13	32	Park Ridge	High	1	60068	1900 S. Cumberland Ave.
14	33	Chicago	High	7	60657	3012 N. Broadway Ave.
15	39	Waukegan	NULL	NULL	60085	Closed
16	40	Bridgeview	CubFighter	6	60455	8825 S. Harlem Ave.
17	44	Western Springs	Medium	2	60558	14 Garden Market St.
18	45	Wheeling	Medium	2	60090	550 W. Dundee Rd.
19	46	Carol Stream	Low	5	60187	Closed

Fig: stg_store Sample Data Retrieval

Stg_demo

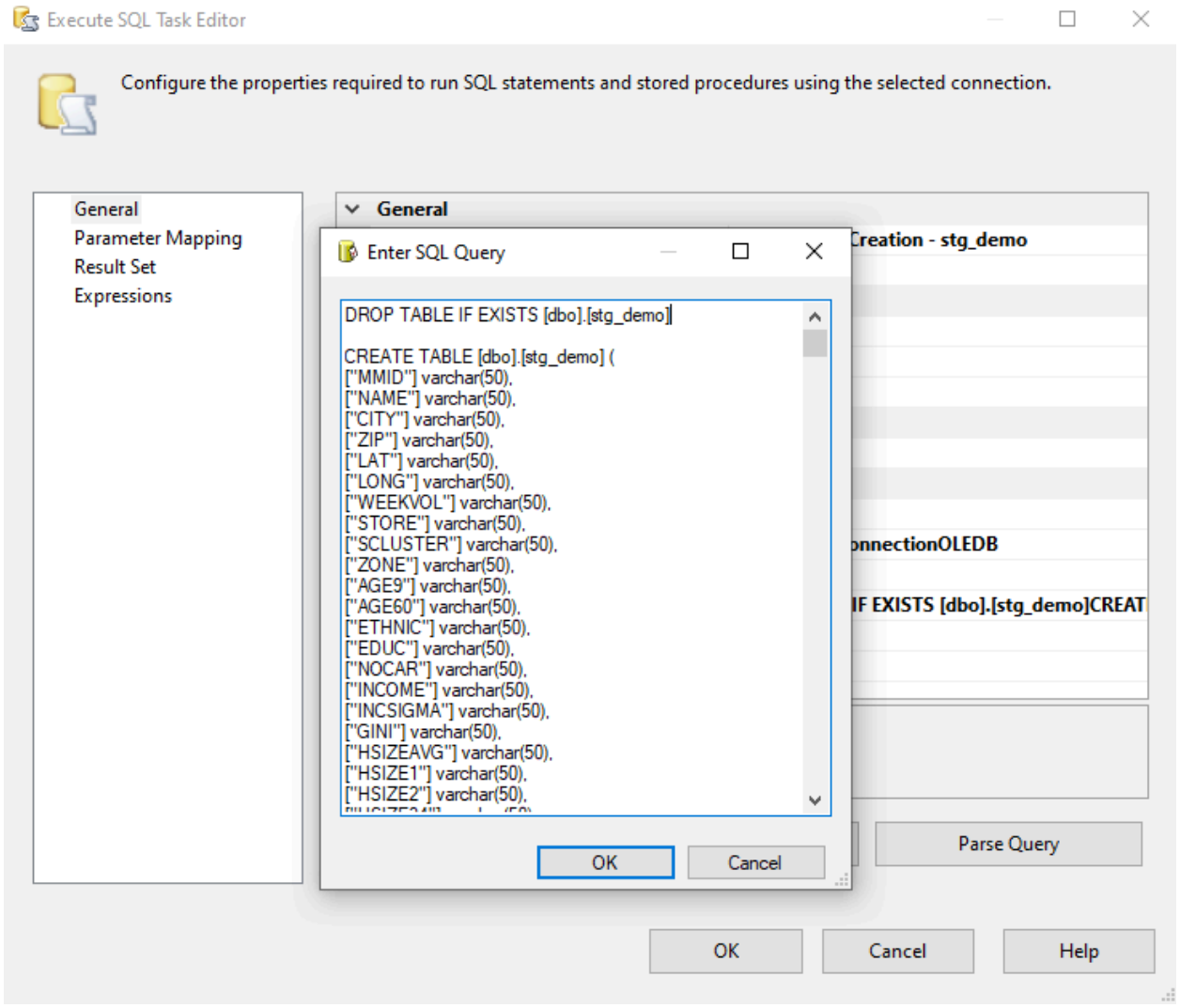


Fig: stg_demo Table Creation

Fig: stg_demo Control Flow

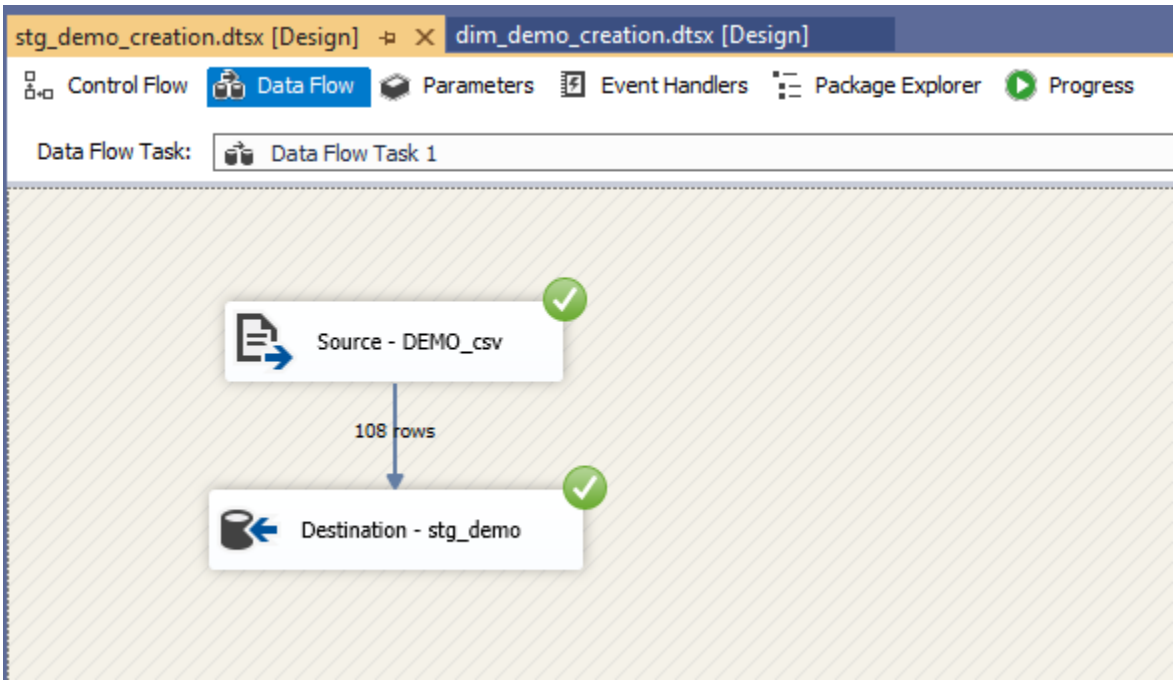
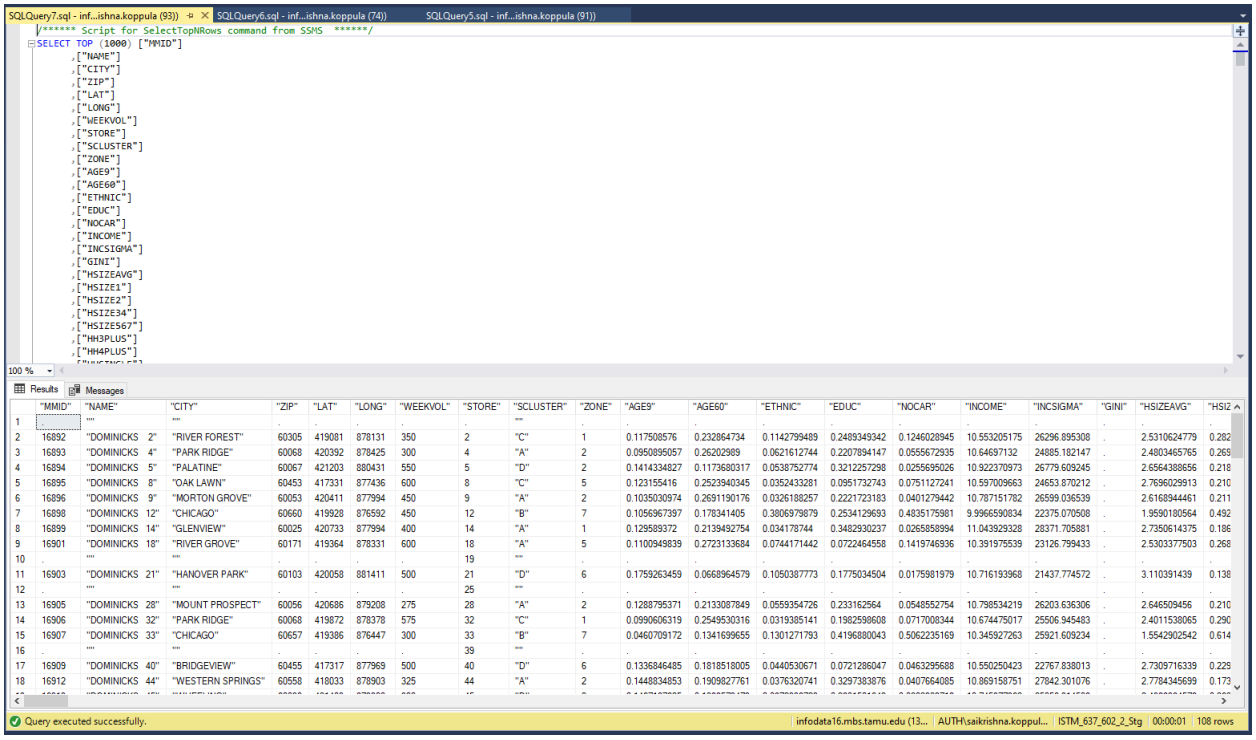


Fig: stg_demo Data Flow



	"MMID"	"NAME"	"CITY"	"ZIP"	"LAT"	"LONG"	"WEEKVOL"	"STORE"	"SCLUSTER"	"ZONE"	"AGE9"	"AGE60"	"ETHNIC"	"EDUC"	"NOCAR"	"INCOME"	"INCSIGMA"	"GINI"	"HSIZEAVG"	"HSIZ
1		""	""	.	.	.	.	.	""	.	.	.	.	.	.	.	.	.	.	.
2	16892	"DOMINICKS 2"	"RIVER FOREST"	60305	419081	878131	350	2	"C"	1	0.117508576	0.232864734	0.1142799489	0.2489349342	0.1246028945	10.553205175	26296.895308	.	2.5310624779	0.282
3	16893	"DOMINICKS 4"	"PARK RIDGE"	60068	420392	878425	300	4	"A"	2	0.0950895057	0.26202989	0.0621612744	0.2207894147	0.0555672935	10.64697132	24885.182147	.	2.4803465765	0.269
4	16894	"DOMINICKS 5"	"PALATINE"	60067	421203	880431	550	5	"D"	2	0.1414334827	0.1173680317	0.0538752774	0.3212257298	0.0255695026	10.922370973	26779.609245	.	2.6564388656	0.218
5	16895	"DOMINICKS 8"	"OAK LAWN"	60453	417331	877436	600	8	"C"	5	0.123155416	0.2523940345	0.0352433281	0.0951732743	0.0751127241	10.597009663	24653.870212	.	2.7696029913	0.210
6	16896	"DOMINICKS 9"	"MORTON GROVE"	60053	420411	877994	450	9	"A"	2	0.1035030974	0.2691190176	0.0326188257	0.2221723183	0.0401279442	10.787151782	26599.036539	.	2.6168944461	0.211
7	16898	"DOMINICKS 12"	"CHICAGO"	60660	419928	876592	450	12	"B"	7	0.1056967397	0.178341405	0.3806979879	0.2534129693	0.4835175981	9.9966590834	22375.070508	.	1.9590180564	0.492
8	16899	"DOMINICKS 14"	"GLENVIEW"	60025	420733	877994	400	14	"A"	1	0.129589372	0.2139492754	0.034178744	0.3482930237	0.0265858994	11.043929328	28371.705881	.	2.7350614375	0.186
9	16901	"DOMINICKS 18"	"RIVER GROVE"	60171	419364	878331	600	18	"A"	5	0.1100949839	0.2723133684	0.0744171442	0.0722464558	0.1419746936	10.391975539	23126.799433	.	2.5303377503	0.268
10	.	""	""	.	.	.	.	19	""	.	.	.	.	.	.	.	.	.	.	.
11	16903	"DOMINICKS 21"	"HANOVER PARK"	60103	420058	881411	500	21	"D"	6	0.1759263459	0.0668964579	0.1050387773	0.1775034504	0.0175981979	10.716193968	21437.774572	.	3.110391439	0.138
12	.	""	""	.	.	.	.	25	""	.	.	.	.	.	.	.	.	.	.	.
13	16905	"DOMINICKS 28"	"MOUNT PROSPECT"	60056	420686	879208	275	28	"A"	2	0.1288795371	0.2133087849	0.0559354726	0.233162564	0.0548552754	10.798534219	26203.636306	.	2.646509456	0.210
14	16906	"DOMINICKS 32"	"PARK RIDGE"	60068	419872	878378	575	32	"C"	1	0.0990606319	0.2549530316	0.0319385141	0.1982598608	0.0717008344	10.674475017	25506.945483	.	2.4011538065	0.290
15	16907	"DOMINICKS 33"	"CHICAGO"	60657	419386	876447	300	33	"B"	7	0.0460709172	0.1341699655	0.1301271793	0.4196880043	0.5062235169	10.345927263	25921.609234	.	1.5542902542	0.614
16	.	""	""	.	.	.	.	39	""	.	.	.	.	.	.	.	.	.	.	.
17	16909	"DOMINICKS 40"	"BRIDGEVIEW"	60455	417317	877969	500	40	"D"	6	0.1336846485	0.1818518005	0.0440530671	0.0721286047	0.0463295688	10.550250423	22767.838013	.	2.7309716339	0.229
18	16912	"DOMINICKS 44"	"WESTERN SPRINGS"	60558	418033	878903	325	44	"A"	2	0.1448834853	0.1909827761	0.0376320741	0.3297383876	0.0407664085	10.869158751	27842.301076	.	2.7784345699	0.173

Fig: stg_demo Sample Data Retrieval

Stg_ccount_temp

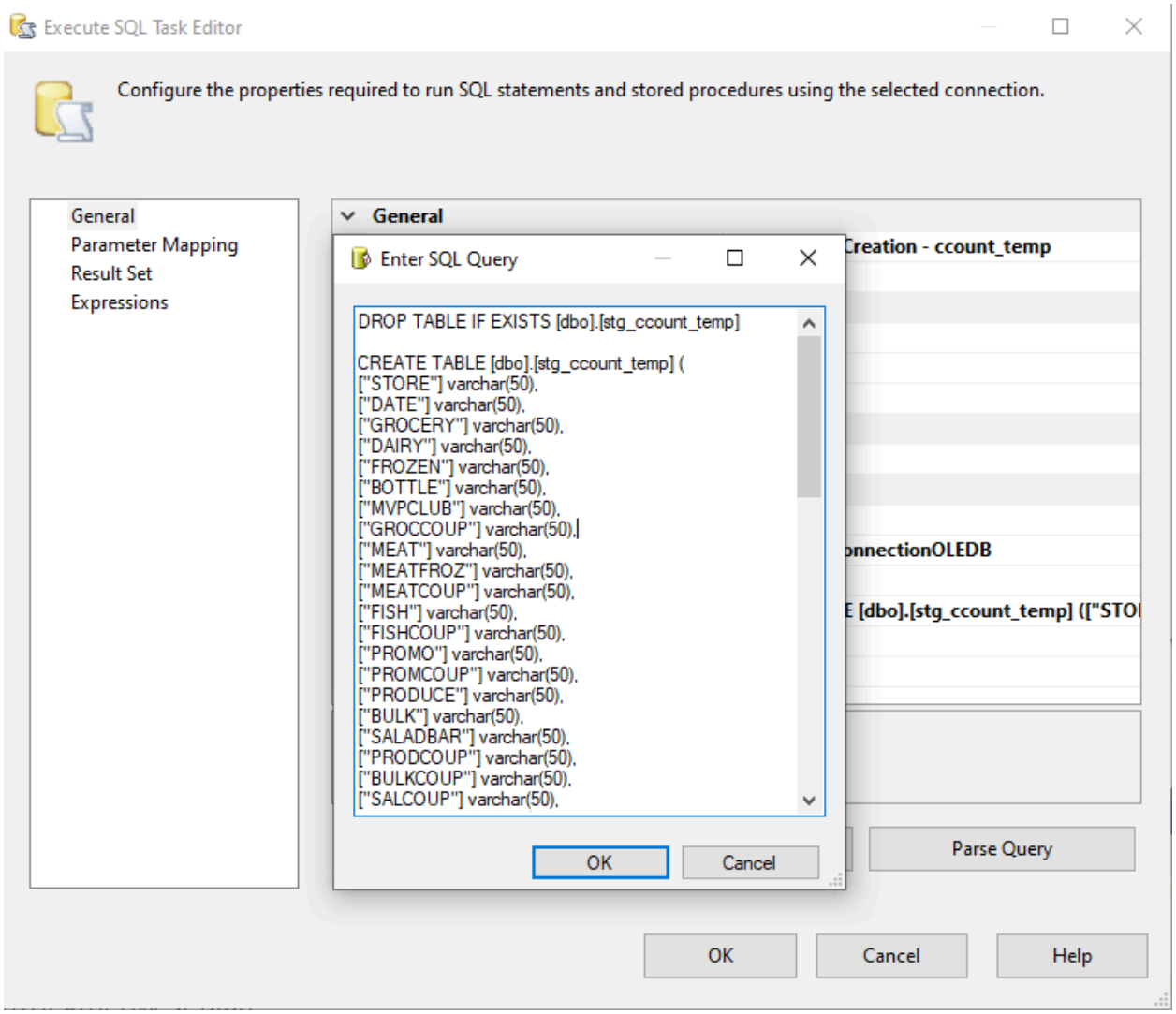


Fig: stg_ccount_temp Table Creation

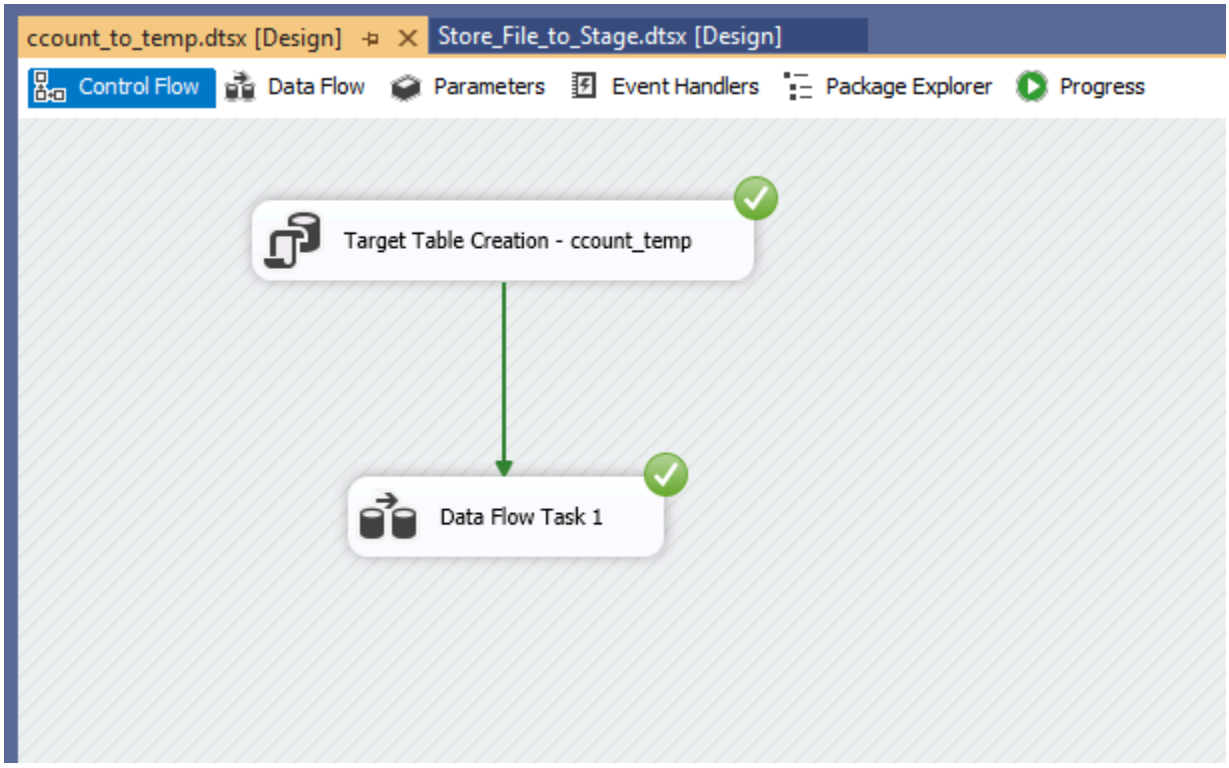


Fig: stg_ccount_temp Control Flow

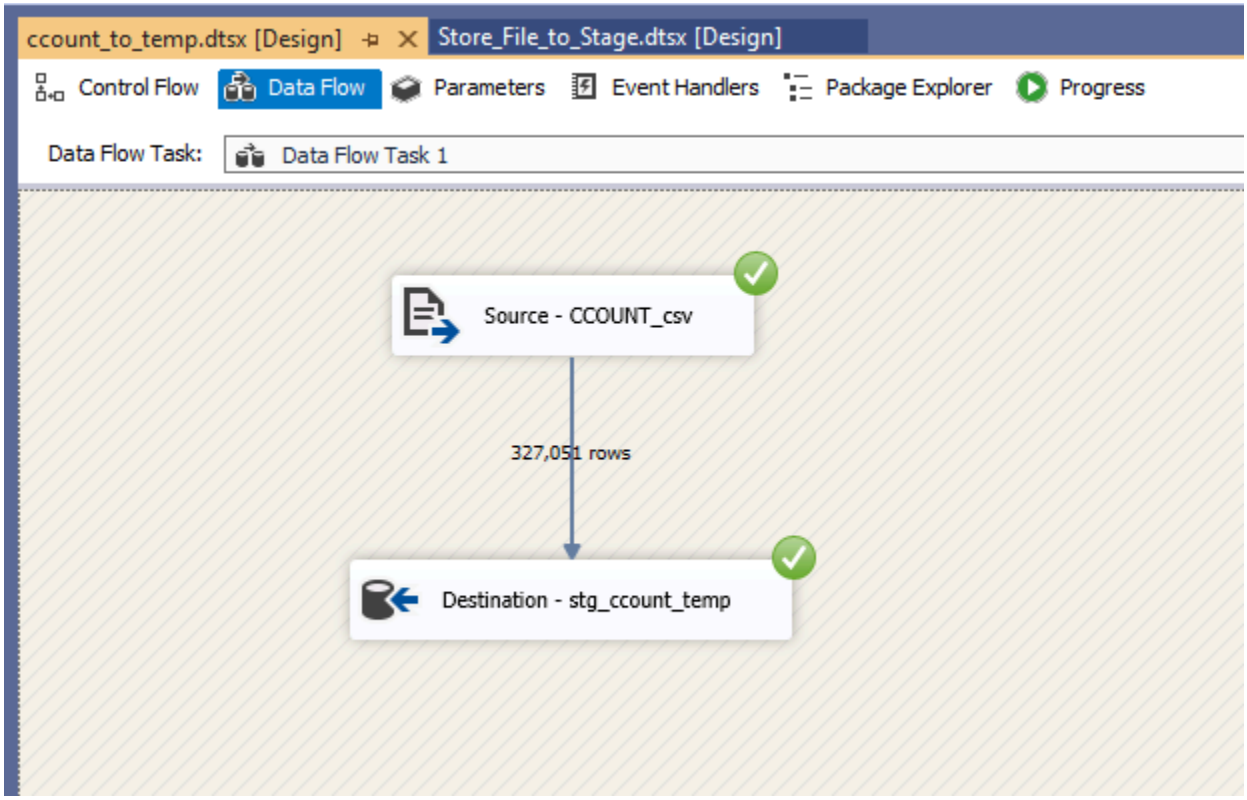
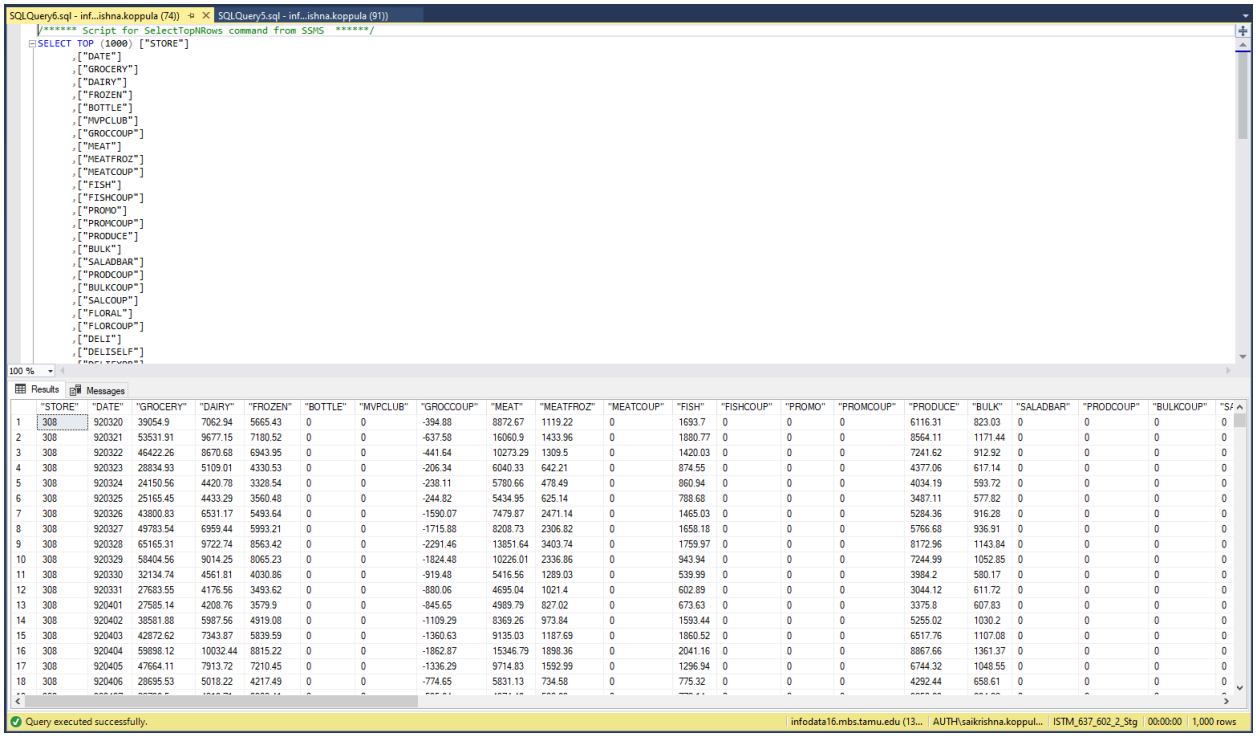


Fig: stg_ccount_temp Data Flow



	"STORE"	"DATE"	"GROCERY"	"DAIRY"	"FROZEN"	"BOTTLE"	"MVPCLUB"	"GROCCOUP"	"MEAT"	"MEATFROZ"	"MEATCOUP"	"FISH"	"FISHCOUP"	"PROMO"	"PROMCOUP"	"PRODUCE"	"BULK"	"SALADBAR"	"PRODCOUP"	"BULKCOUP"	"SA
1	308	920320	39054.9	7062.94	5665.43	0	0	-394.88	8872.67	1119.22	0	1693.7	0	0	0	6116.31	823.03	0	0	0	0
2	308	920321	53531.91	9677.15	7180.52	0	0	-637.58	16060.9	1433.96	0	1880.77	0	0	0	8564.11	1171.44	0	0	0	0
3	308	920322	46422.26	8670.68	6943.95	0	0	-441.64	10273.29	1309.5	0	1420.03	0	0	0	7241.62	912.92	0	0	0	0
4	308	920323	28834.93	5109.01	4330.53	0	0	-206.34	6040.33	642.21	0	874.55	0	0	0	4377.06	617.14	0	0	0	0
5	308	920324	24150.56	4420.78	3328.54	0	0	-238.11	5780.66	478.49	0	860.94	0	0	0	4034.19	593.72	0	0	0	0
6	308	920325	25165.45	4433.29	3560.48	0	0	-244.82	5434.95	625.14	0	788.68	0	0	0	3487.11	577.82	0	0	0	0
7	308	920326	43800.83	6531.17	5493.64	0	0	-1590.07	7479.87	2471.14	0	1465.03	0	0	0	5284.36	916.28	0	0	0	0
8	308	920327	49783.54	6959.44	5993.21	0	0	-1715.88	8208.73	2306.82	0	1658.18	0	0	0	5766.68	936.91	0	0	0	0
9	308	920328	65165.31	9722.74	8563.42	0	0	-2291.46	13851.64	3403.74	0	1759.97	0	0	0	8172.96	1143.84	0	0	0	0
10	308	920329	58404.56	9014.25	8065.23	0	0	-1824.48	10226.01	2336.86	0	943.94	0	0	0	7244.99	1052.85	0	0	0	0
11	308	920330	32134.74	4561.81	4030.86	0	0	-919.48	5416.56	1289.03	0	539.99	0	0	0	3984.2	580.17	0	0	0	0
12	308	920331	27683.55	4176.56	3493.62	0	0	-880.06	4695.04	1021.4	0	602.89	0	0	0	3044.12	611.72	0	0	0	0
13	308	920401	27585.14	4208.76	3579.9	0	0	-845.65	4989.79	827.02	0	673.63	0	0	0	3375.8	607.83	0	0	0	0
14	308	920402	38581.88	5987.56	4919.08	0	0	-1109.29	8369.26	973.84	0	1593.44	0	0	0	5255.02	1030.2	0	0	0	0
15	308	920403	42872.62	7343.87	5839.59	0	0	-1360.63	9135.03	1187.69	0	1860.52	0	0	0	6517.76	1107.08	0	0	0	0
16	308	920404	59898.12	10032.44	8815.22	0	0	-1862.87	15346.79	1898.36	0	2041.16	0	0	0	8867.66	1361.37	0	0	0	0
17	308	920405	47664.11	7913.72	7210.45	0	0	-1336.29	9714.83	1592.99	0	1296.94	0	0	0	6744.32	1048.55	0	0	0	0
18	308	920406	28695.53	5018.22	4217.49	0	0	-774.65	5831.13	734.58	0	775.32	0	0	0	4292.44	658.61	0	0	0	0

Fig: stg_ccount_temp Data Control Retrieval

Stg_product

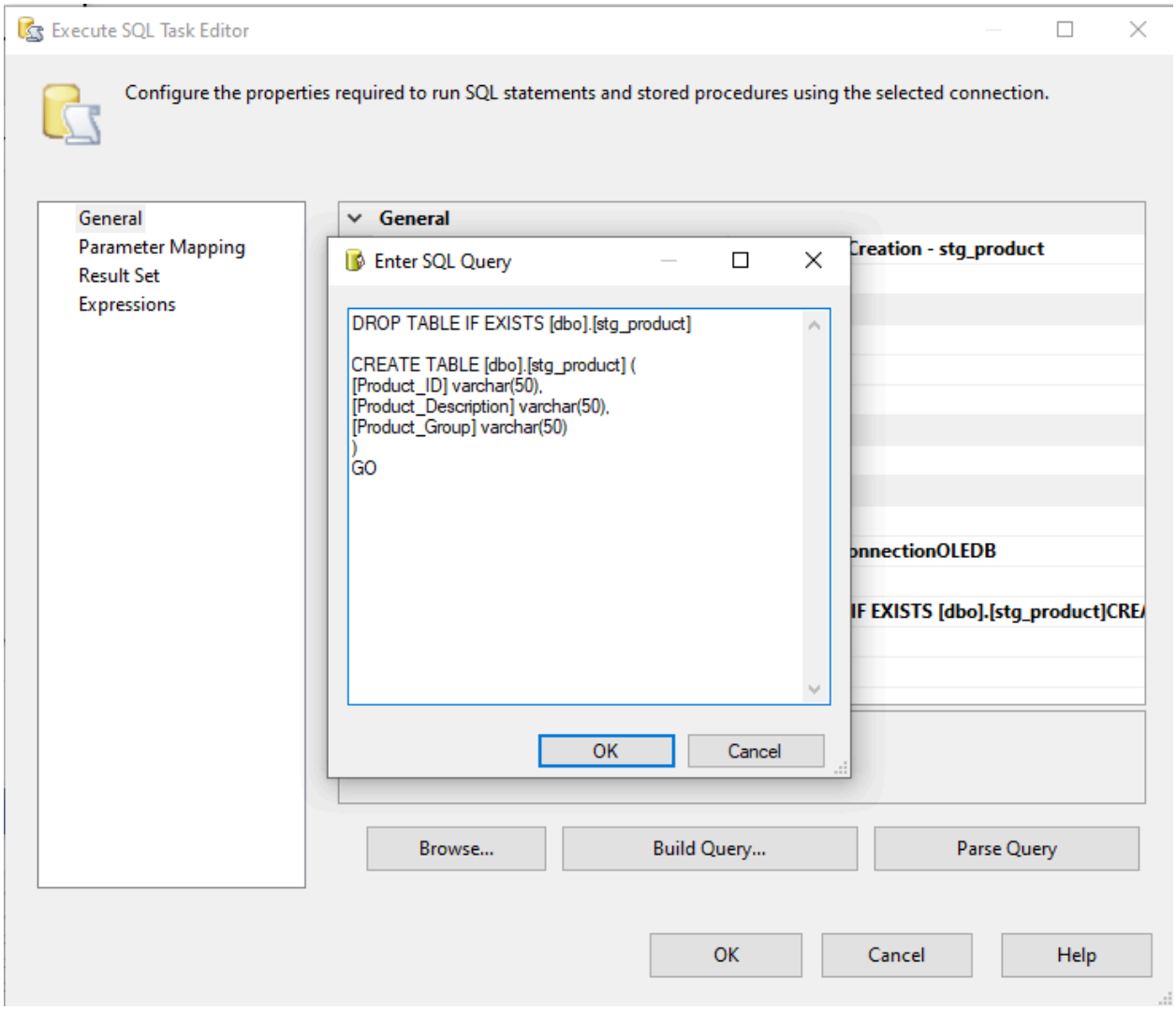


Fig: stg_product Control Flow

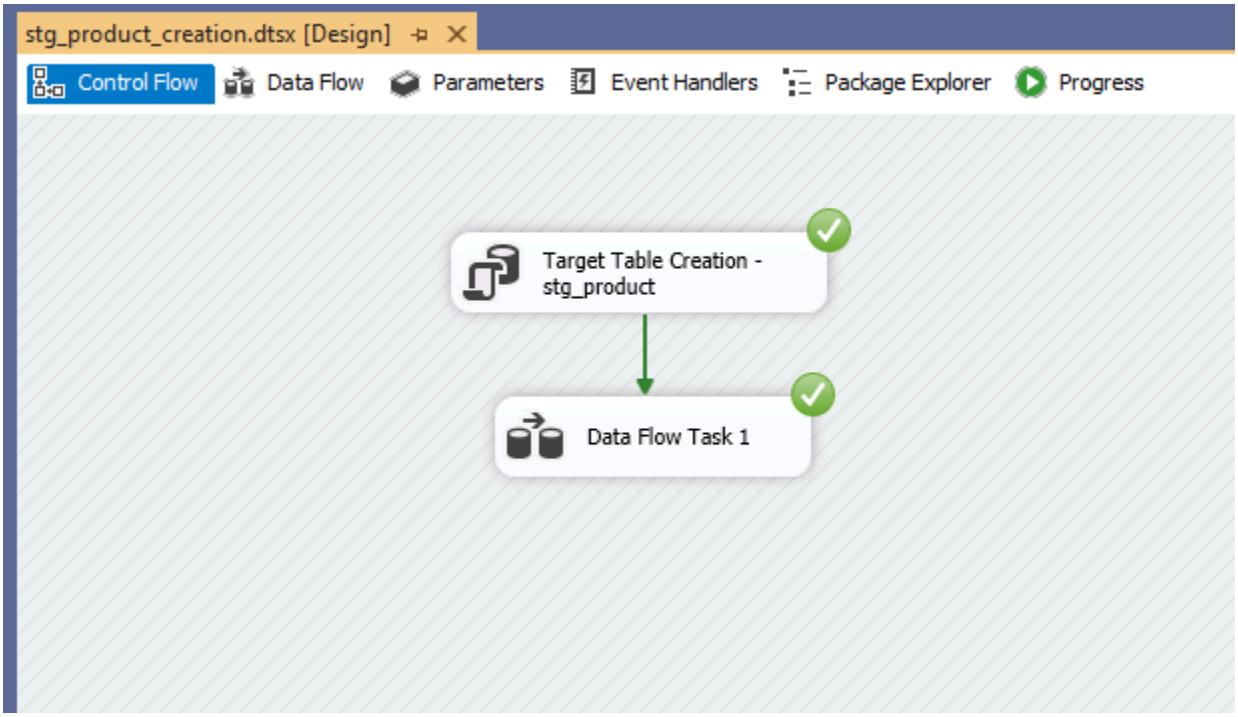


Fig: stg_product Control Flow

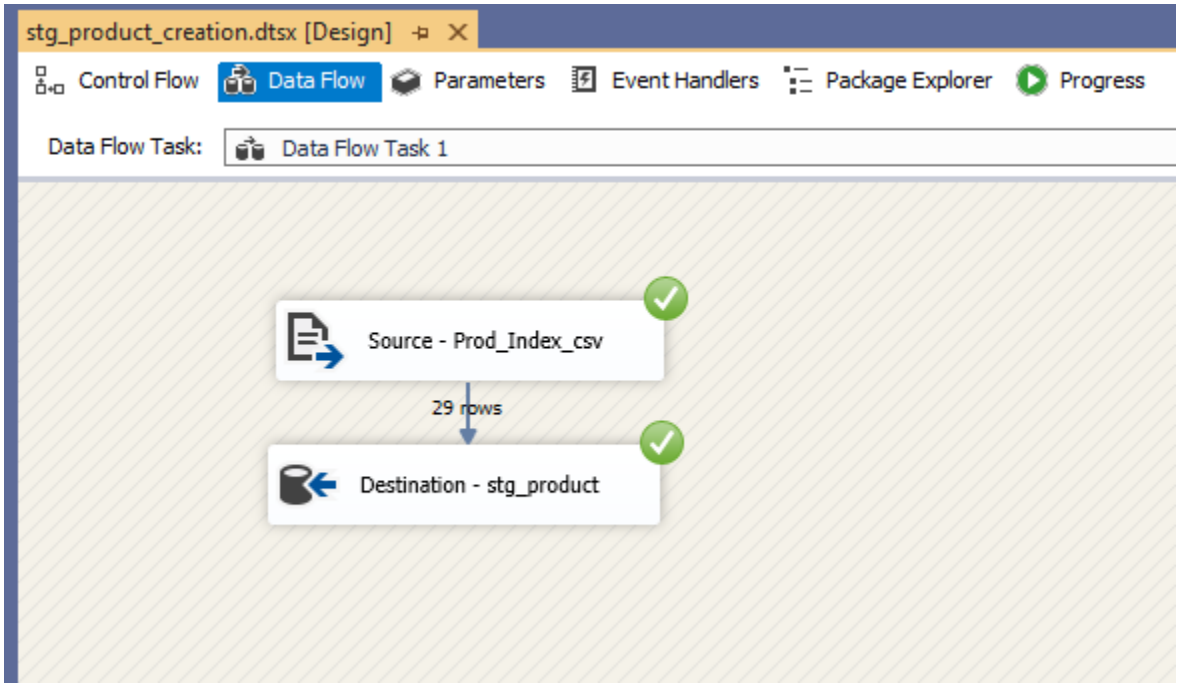
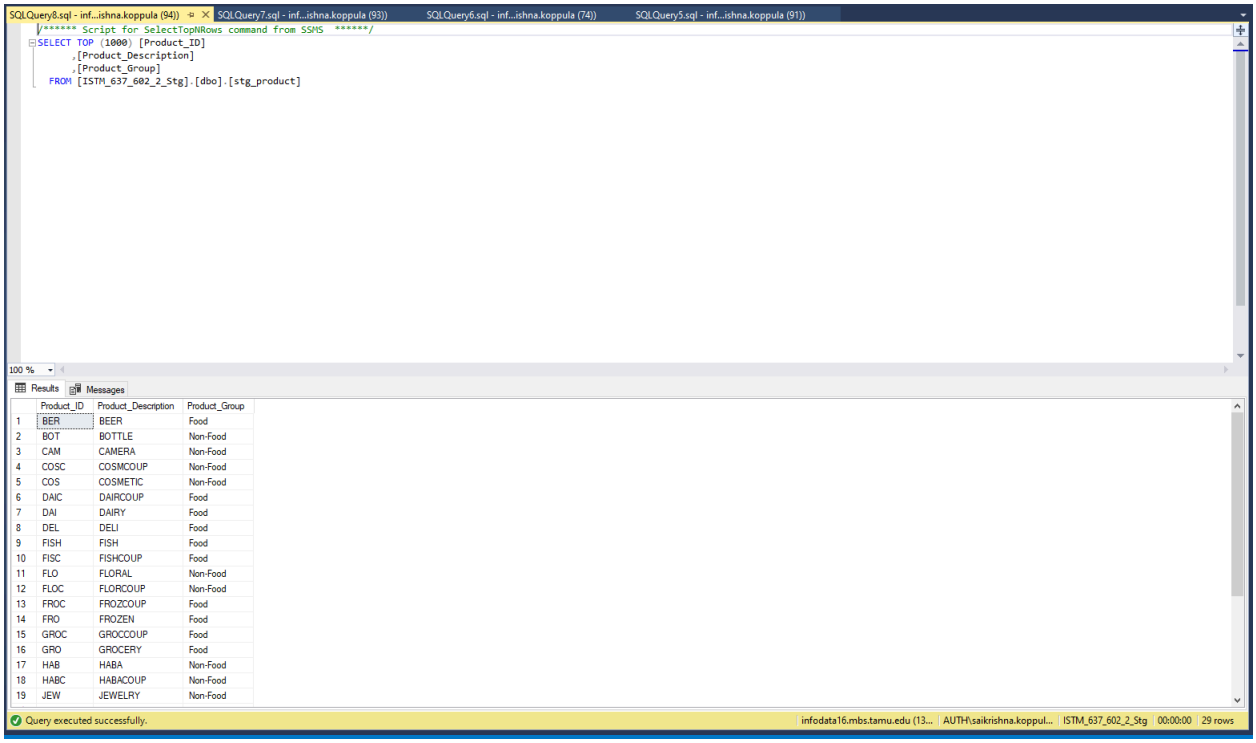


Fig: stg_product Data Flow



	Product_ID	Product_Description	Product_Group
1	BER	BEER	Food
2	BOT	BOTTLE	Non-Food
3	CAM	CAMERA	Non-Food
4	COSC	COSMCOUP	Non-Food
5	COS	COSMETIC	Non-Food
6	DAIC	DAIRCOUP	Food
7	DAI	DAIRY	Food
8	DEL	DELI	Food
9	FISH	FISH	Food
10	FISC	FISHCOUP	Food
11	FLO	FLORAL	Non-Food
12	FLOC	FLORCOUP	Non-Food
13	FROC	FROZCOUP	Food
14	FRO	FROZEN	Food
15	GROC	GROCCOUP	Food
16	GRO	GROCERY	Food
17	HAB	HABA	Non-Food
18	HABC	HABACOUP	Non-Food
19	JEW	JEWELRY	Non-Food

Fig: stg_product Sample Data Retrieval

Stg_weeks_decode_table

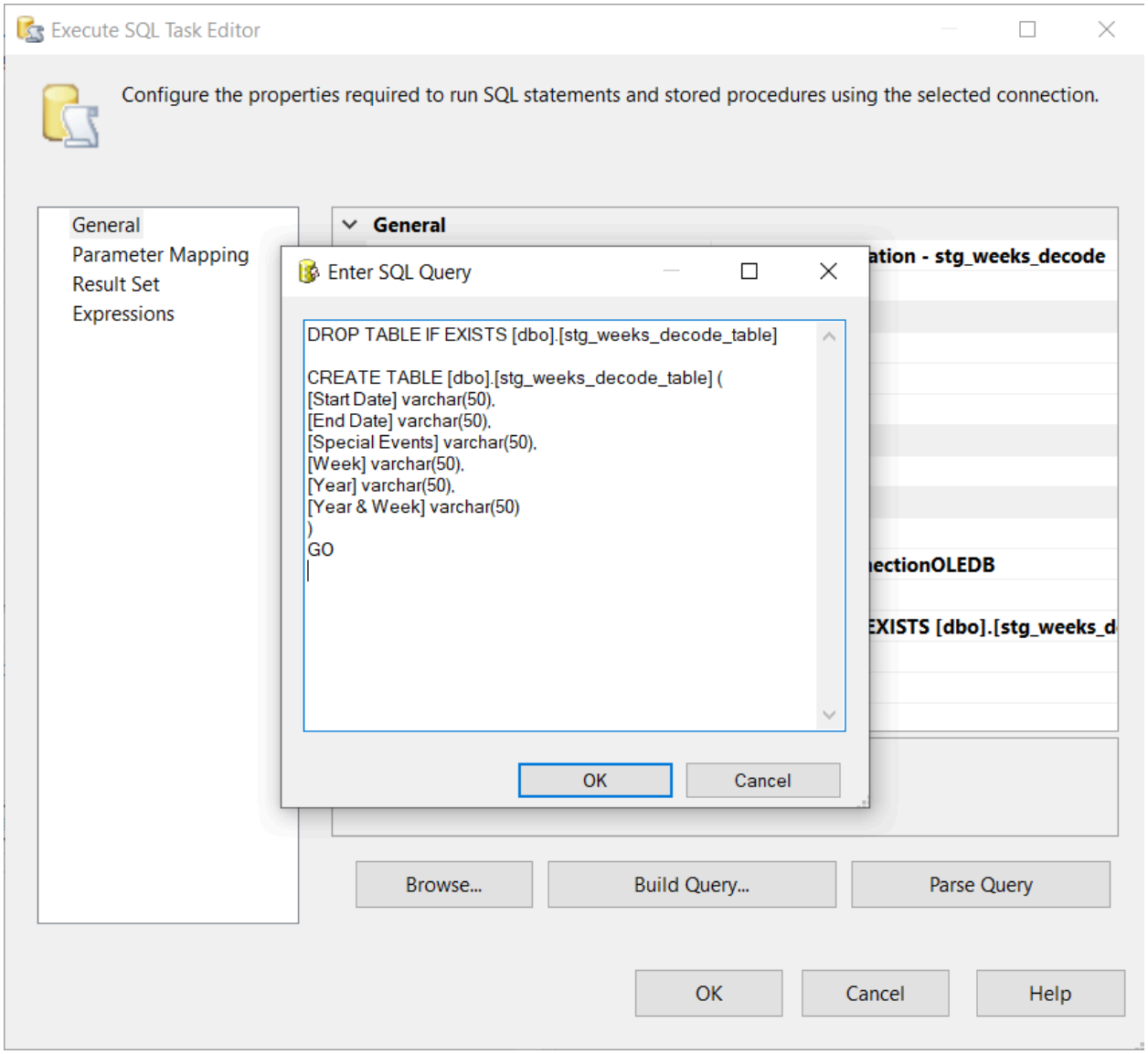


Fig: stg_weeks_decode_table Table Creation

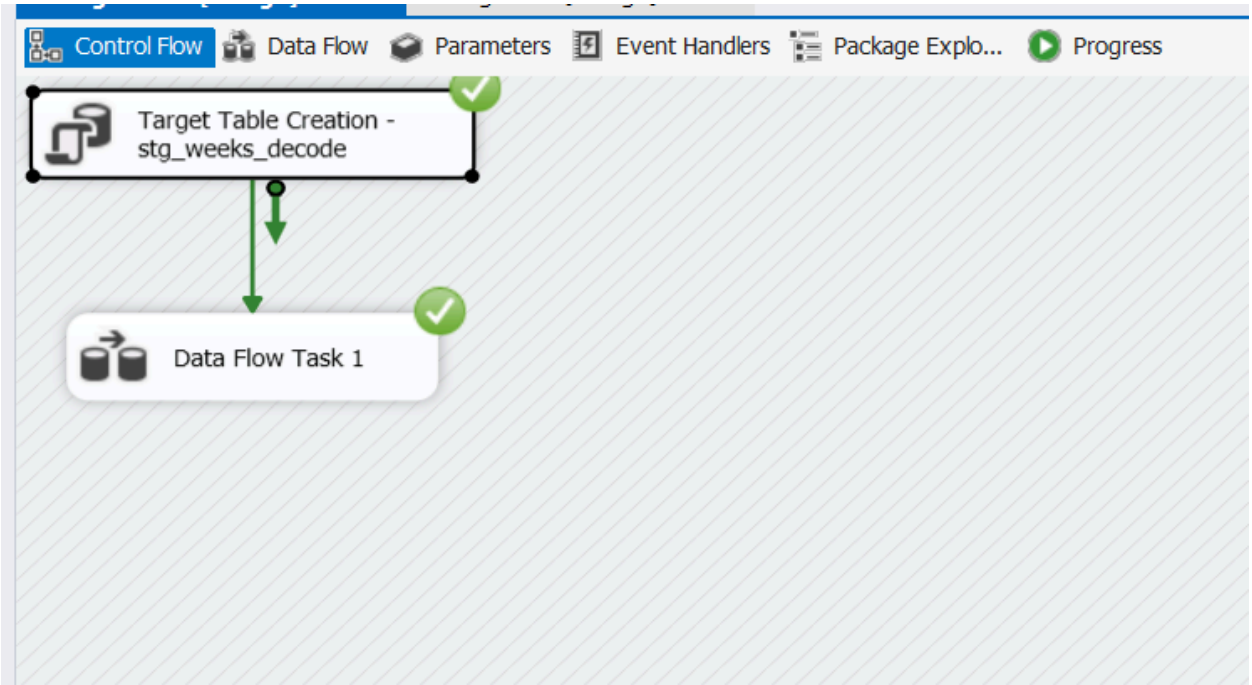


Fig: stg_weeks_decode_table Control Flow

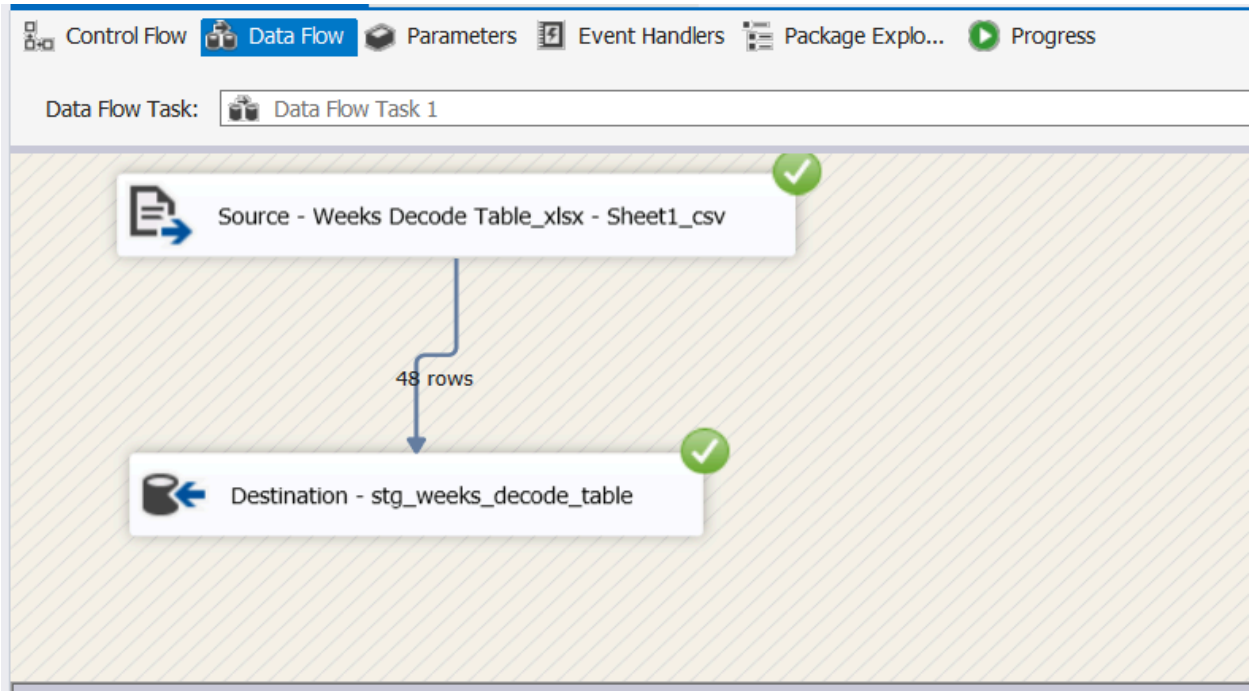


Fig: stg_weeks_decode_table Data Flow

```
SELECT TOP (1000) [Start Date]
      ,[End Date]
      ,[Special Events]
      ,[Week]
      ,[Year]
      ,[Year & Week]
  FROM [ISTM_637_602_2_Stg].[dbo].[stg_weeks_decode_table]
```

100 %

Results | Messages

	Start Date	End Date	Special Events	Week	Year	Year & Week
1	10/26/89	11/01/89	Halloween	43	1989	198943
2	11/23/89	11/29/89	Thanksgiving	47	1989	198947
3	12/21/89	12/27/89	Christmas	51	1989	198951
4	12/28/89	01/03/90	New-Year	52	1989	198952
5	02/15/90	02/21/90	Presidents Day	7	1990	19907
6	03/22/90	03/28/90	Easter	12	1990	199012
7	05/24/90	05/30/90	Memorial Day	21	1990	199021
8	06/28/90	07/04/90	4th of July	26	1990	199026
9	08/30/90	09/05/90	Labor Day	35	1990	199035
10	10/25/90	10/31/90	Halloween	43	1990	199043
11	11/22/90	11/28/90	Thanksgiving	47	1990	199047
12	12/20/90	12/26/90	Christmas	51	1990	199051
13	02/14/91	02/20/91	Presidents Day	7	1991	19917
14	03/28/91	04/03/91	Easter	13	1991	199113
15	05/23/91	05/29/91	Memorial Day	21	1991	199121
16	07/04/91	07/10/91	4th of July	27	1991	199127
17	08/29/91	09/04/91	Labor Day	35	1991	199135

Query executed successfully.

Fig: stg_weeks_decode_table Sample Data Retrieval

Dimension Tables

Dim_Store

The dominicks_store_descriptions.csv file is created according to the manual and business requirements. It includes detailed store information such as Store, City, Price Tier, Zone, Zip Code, and Address.

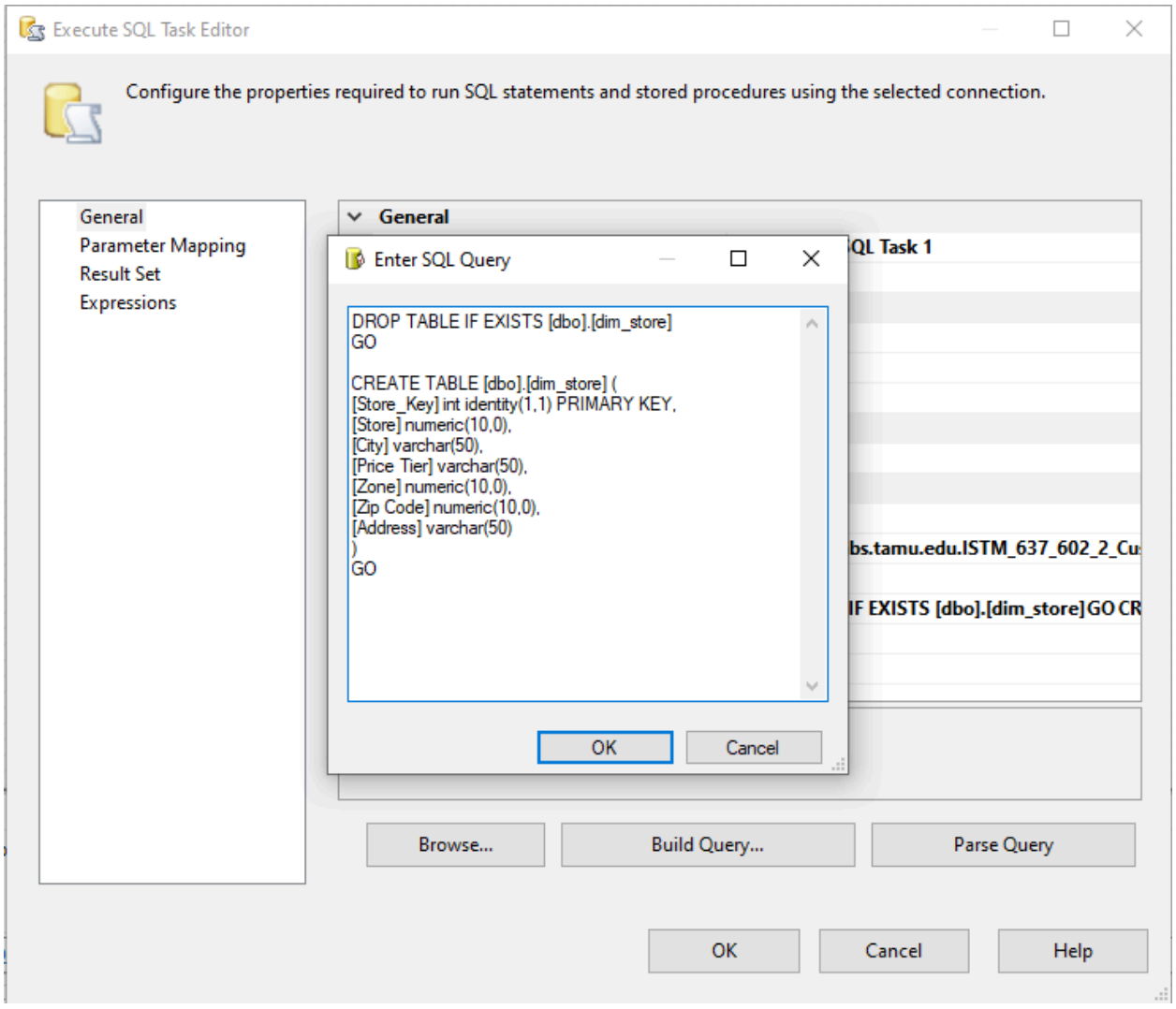


Fig: dim_store Table creation

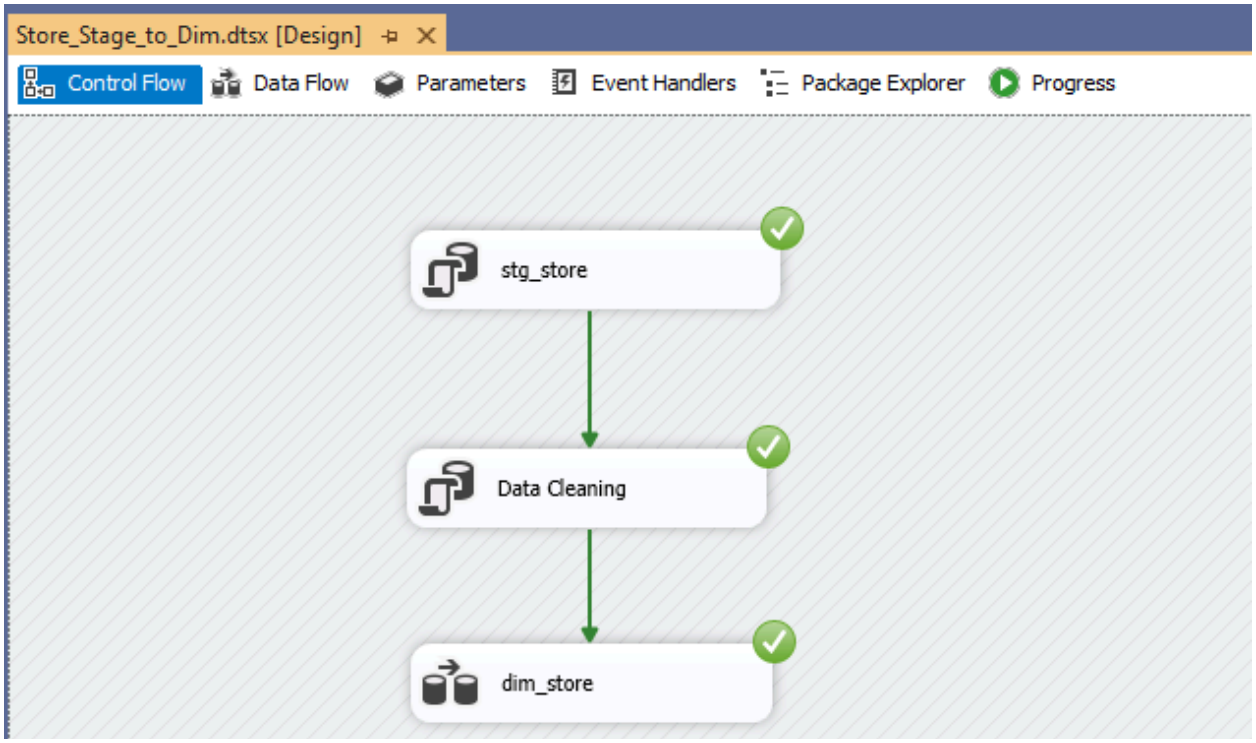


Fig: dim_store Control Flow

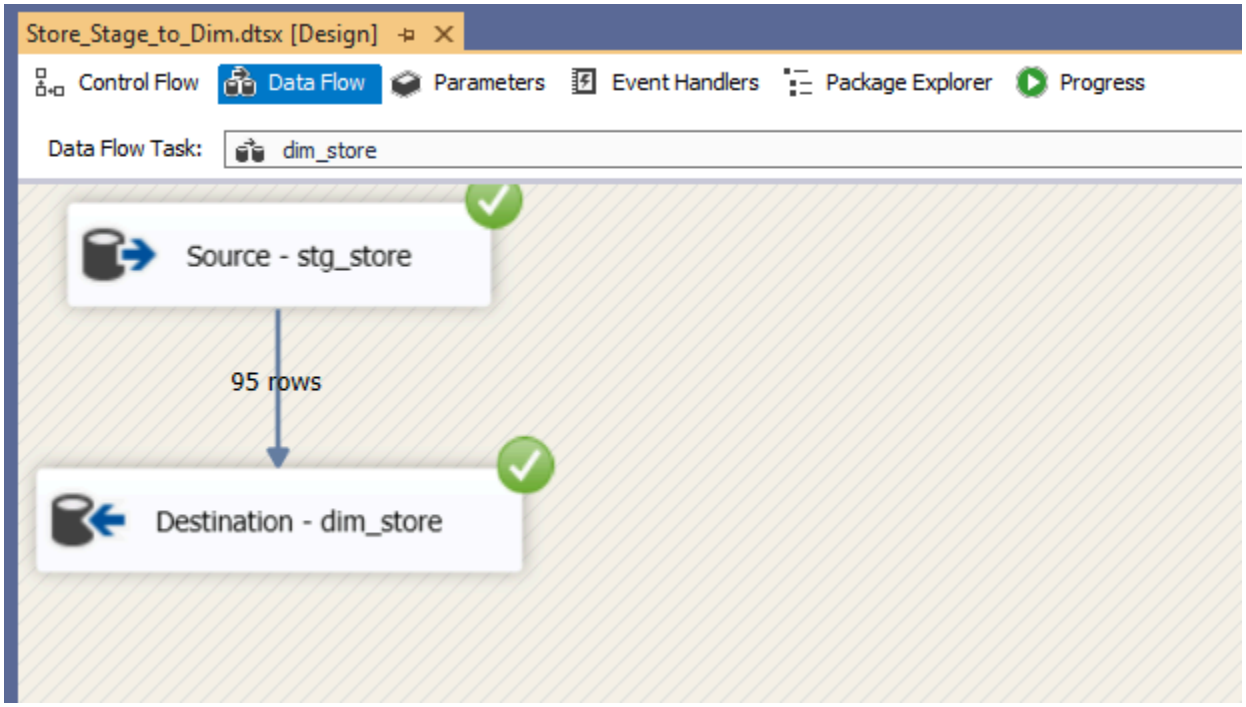


Fig: dim_store Data Flow

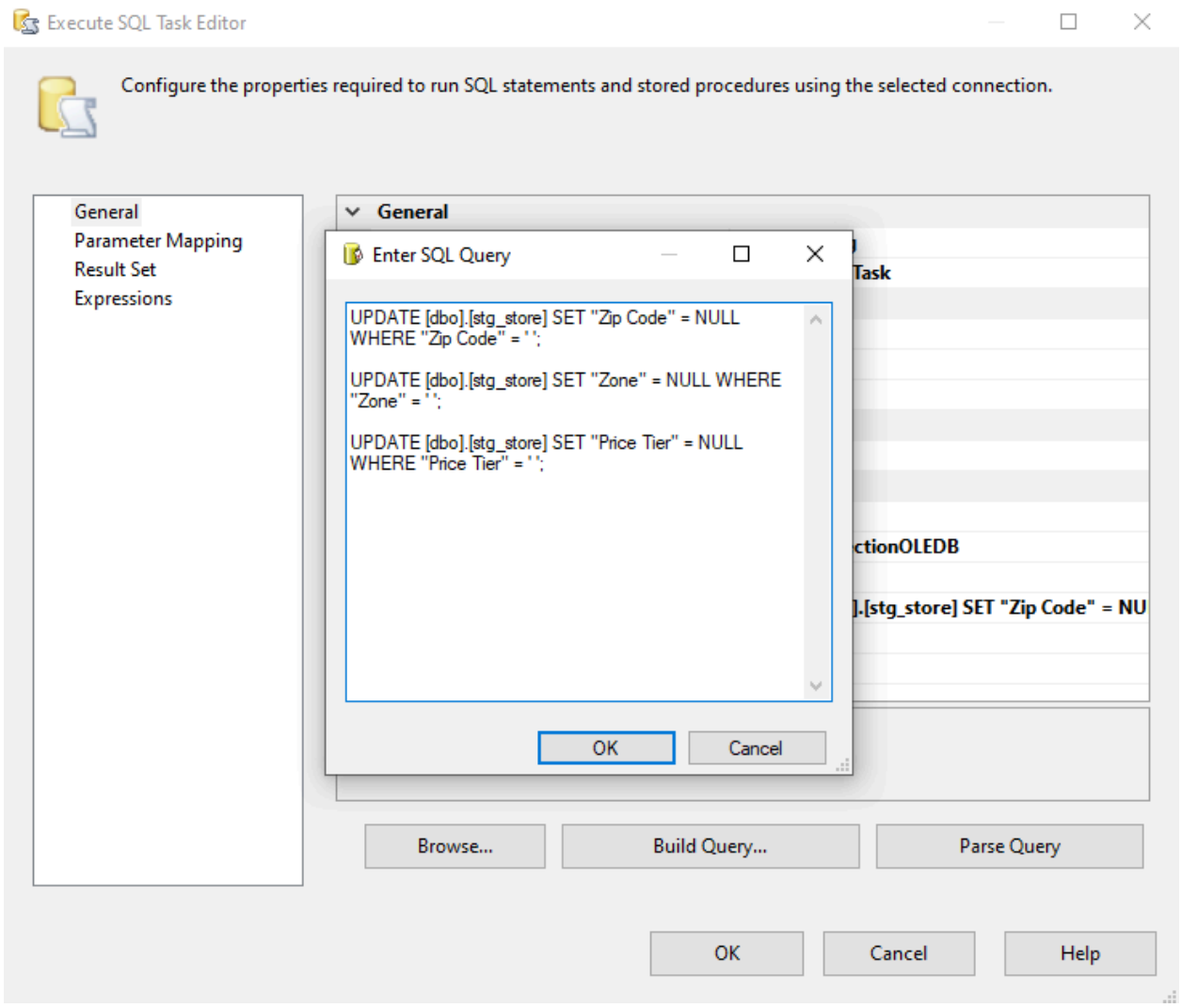


Fig: dim_store - Query for Data Cleaning

SQLQuery16.sql - i...amsibezawada (99)) | SQLQuery15.sql - i...amsibezawada (72)) | SQLQuery14.sql -

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [Store_Key]
      ,[Store]
      ,[City]
      ,[Price Tier]
      ,[Zone]
      ,[Zip Code]
      ,[Address]
  FROM [ISTM_637_602_2_Sales_DM].[dbo].[dim_store]
```

100 %

Results | Messages

	Store_Key	Store	City	Price Tier	Zone	Zip Code	Address
1	1	2	River Forest	High	1	60305	7501 W. North Ave.
2	2	4	Park Ridge	Medium	2	60068	Closed
3	3	5	Palatine	Medium	2	60067	223 Northwest HWY.
4	4	8	Oak Lawn	Low	5	60435	8700 S. Cicero Ave.
5	5	9	Morton Grove	Medium	2	60053	6931 Dempster
6	6	12	Chicago	High	7	60660	6009 N. Broadway Ave.
7	7	14	Glenview	High	1	60025	1020 Waukegan Rd.
8	8	18	River Grove	Low	5	60171	8355 W. Belmont Ave.
9	9	19	Glen Ellyn	NULL	NULL	60137	Closed
10	10	21	Hanover Park	CubFighter	6	60103	1440 Irving Park Rd.
11	11	25	Chicago	NULL	NULL	60639	Closed
12	12	28	Mt. Prospect	Medium	2	60054	1145-55 Mt Prospect Pz.
13	13	32	Park Ridge	High	1	60068	1900 S. Cumberland Ave.
14	14	33	Chicago	High	7	60657	3012 N. Broadway Ave.
15	15	39	Waukegan	NULL	NULL	60085	Closed
16	16	40	Bridgeview	CubFighter	6	60455	8825 S. Harlem Ave.
17	17	44	Western Springs	Medium	2	60558	14 Garden Market St.
18	18	45	Wheeling	Medium	2	60090	550 W. Dundee Rd.
19	19	46	Carol Stream	Low	5	60187	Closed

Query executed successfully.

Fig: dim_store Sample Data Retrieval

Dim_Demo

The Demo.csv file is utilized to populate essential columns, such as MMID and shopper categories, to support and address specific business questions.

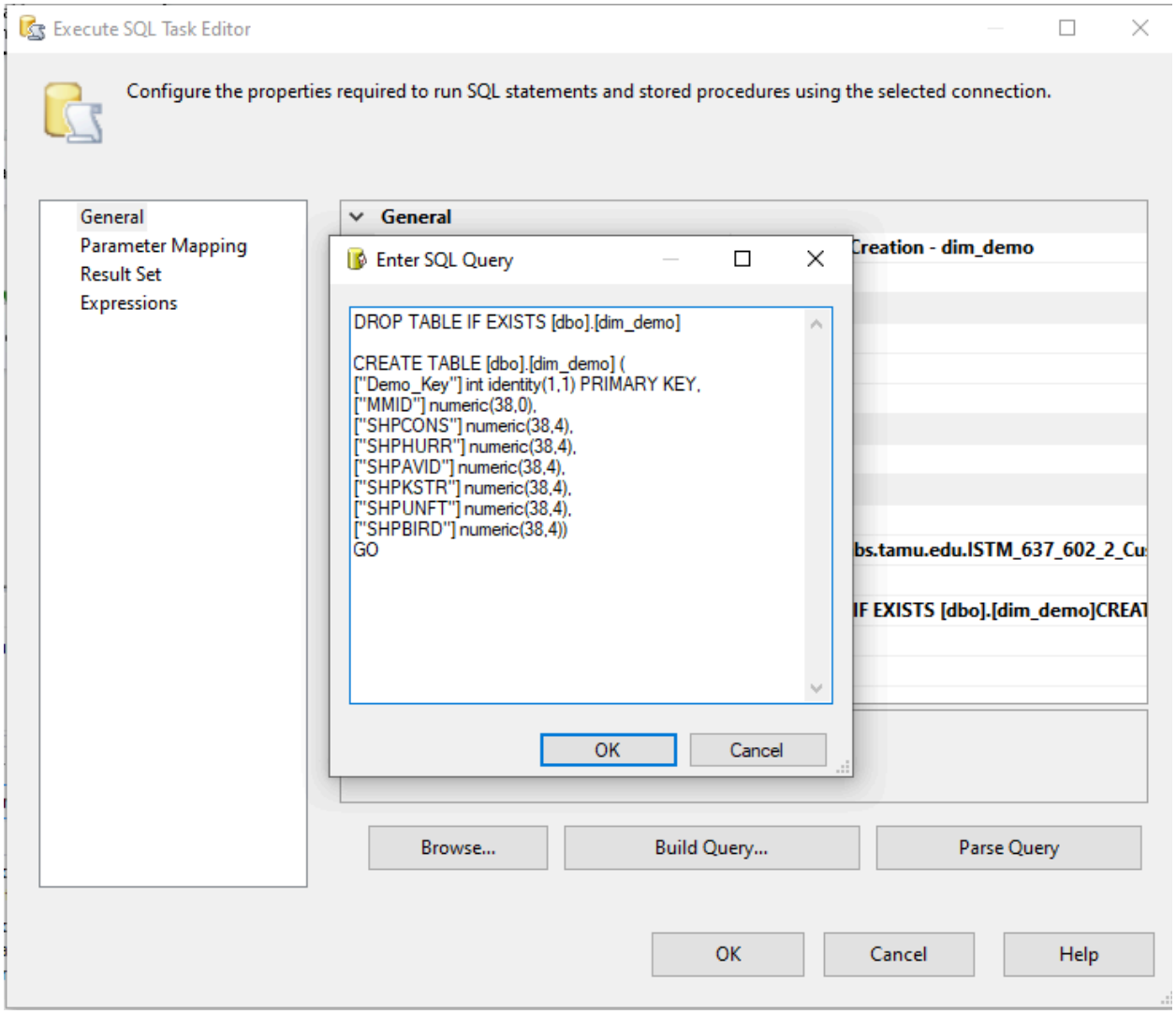


Fig: dim_demo Table Creation

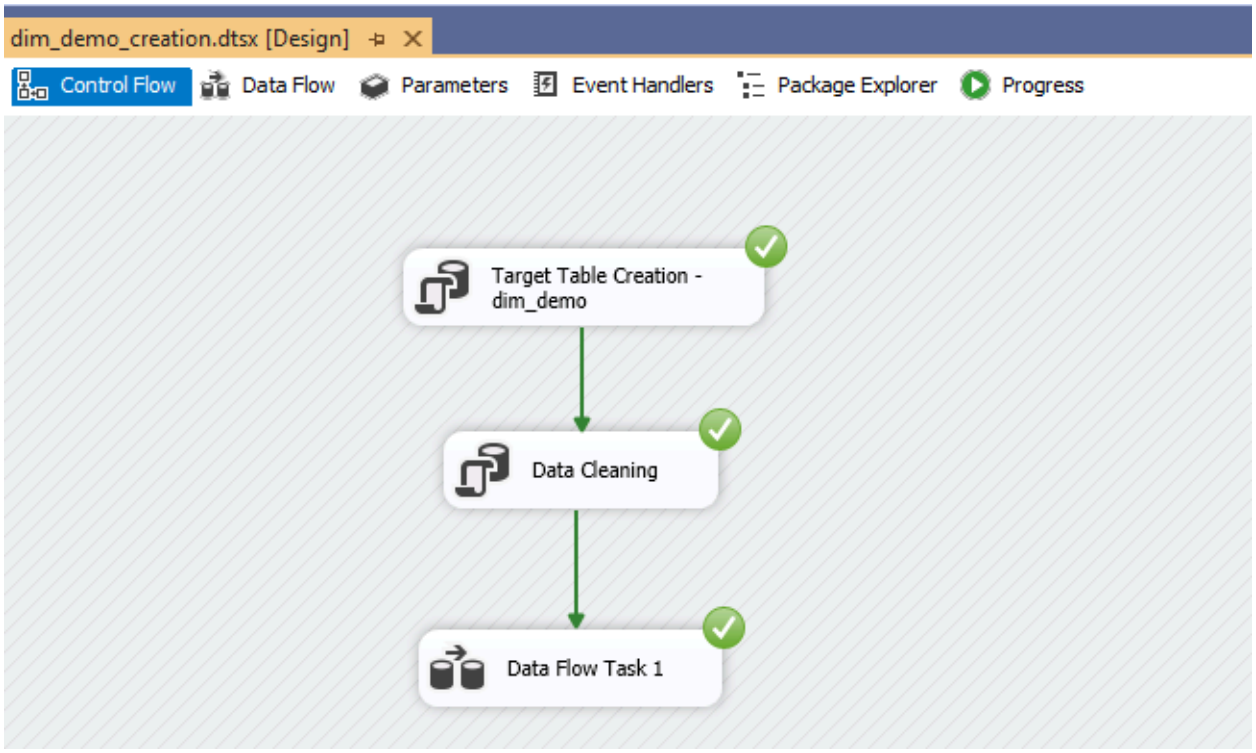


Fig: dim_demo Control Flow

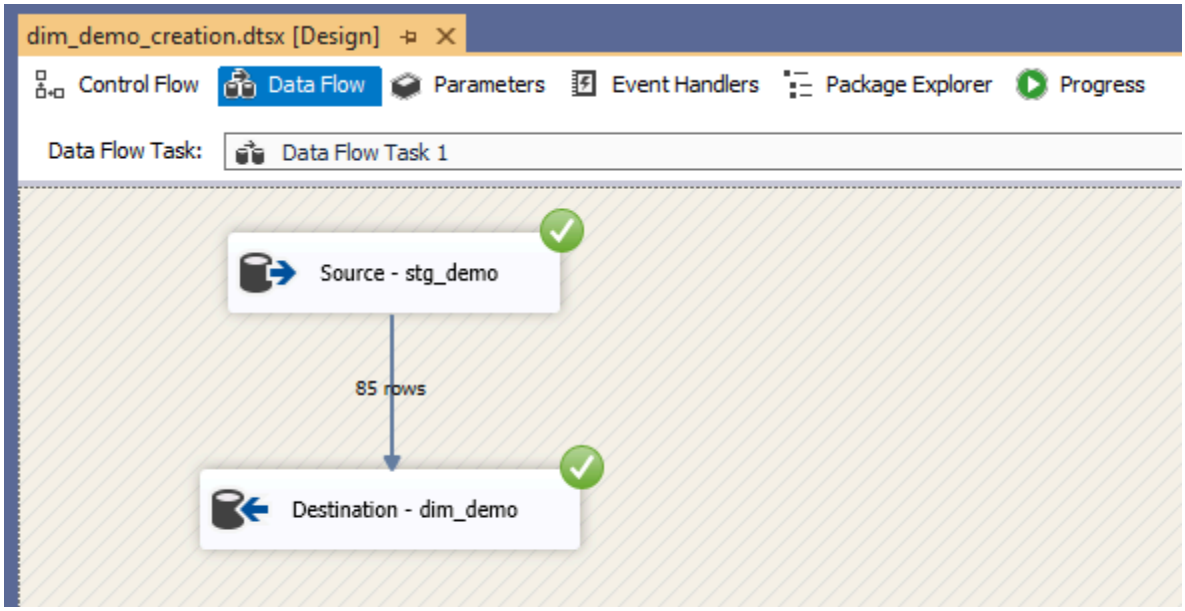


Fig: dim_demo Data Flow

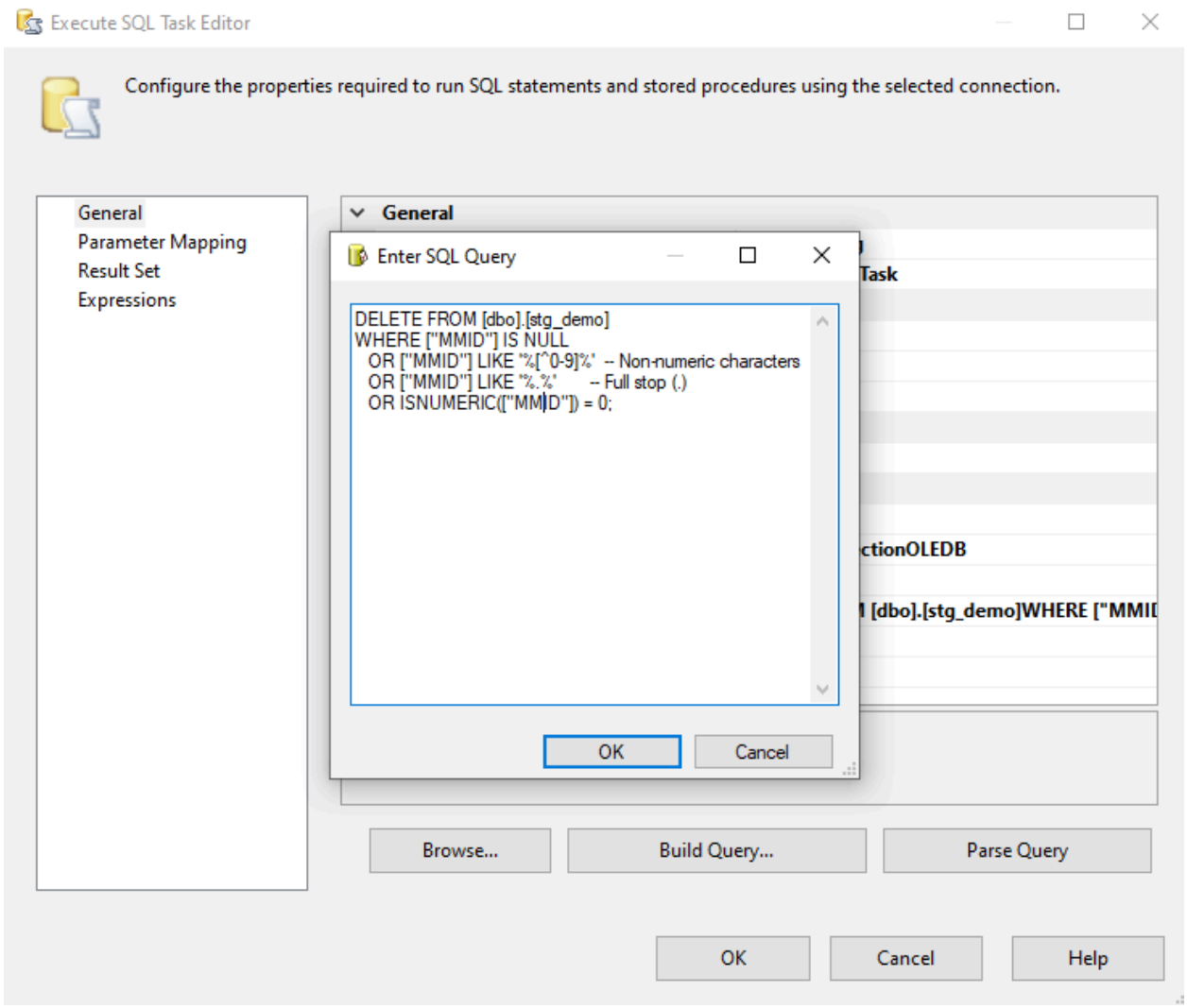


Fig: dim_demo - Query in Data Cleaning

SQLQuery18.sql - i...amsibezawada (55)) | SQLQuery17.sql - i...amsibezawada (56)) | SQLQuery16.sql - i...amsib

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) ["Demo_Key"]
      ,["MMID"]
      ,["SHPCONS"]
      ,["SHPHURR"]
      ,["SHPAVID"]
      ,["SHPKSTR"]
      ,["SHPUNFT"]
      ,["SHPBIRD"]
  FROM [ISTM_637_602_2_Customer_DM].[dbo].[dim_demo]
```

100 %

Results | Messages

	"Demo_Key"	"MMID"	"SHPCONS"	"SHPHURR"	"SHPAVID"	"SHPKSTR"	"SHPUNFT"	"SHPBIRD"
1	1	16892	0.0910	0.1210	0.1812	0.2322	0.2843	0.0900
2	2	16893	0.0553	0.1183	0.1628	0.3169	0.3124	0.0340
3	3	16894	0.0325	0.1906	0.1680	0.3301	0.2633	0.0152
4	4	16895	0.0762	0.1351	0.2174	0.2161	0.3027	0.0522
5	5	16896	0.0433	0.1586	0.1536	0.2364	0.3604	0.0475
6	6	16898	0.2298	0.0459	0.1104	0.3630	0.1820	0.0686
7	7	16899	0.0245	0.2154	0.1386	0.2164	0.3616	0.0432
8	8	16901	0.1154	0.1023	0.2156	0.2310	0.2619	0.0736
9	9	16903	0.0335	0.2269	0.2502	0.2958	0.1852	0.0081
10	10	16905	0.0488	0.1658	0.1630	0.2659	0.3180	0.0382
11	11	16906	0.0565	0.1147	0.1584	0.2906	0.3111	0.0684
12	12	16907	0.1248	0.0263	0.0613	0.5577	0.1721	0.0576
13	13	16909	0.0784	0.1303	0.2380	0.2775	0.2480	0.0275
14	14	16912	0.0413	0.2046	0.1770	0.2052	0.3199	0.0517
15	15	16913	0.0344	0.1574	0.1756	0.3922	0.2239	0.0162
16	16	16915	0.0494	0.1686	0.2196	0.3126	0.2243	0.0254
17	17	16916	0.0268	0.1293	0.1384	0.4518	0.2416	0.0118
18	18	16917	0.0456	0.1719	0.1896	0.3044	0.2271	0.0611
19	19	16918	0.0618	0.1264	0.2258	0.3214	0.2293	0.0350

Fig: dim_demo Sample Data Retrieval

Dim_Date

The Weeks Decode table from Dominick's user manual serves as a reference for joining with the stg_ccount table to derive the Special Events column. Additionally, the month, year, quarter, and week values are extracted from the date attribute in stg_ccount and stored in the Date Dimension table.

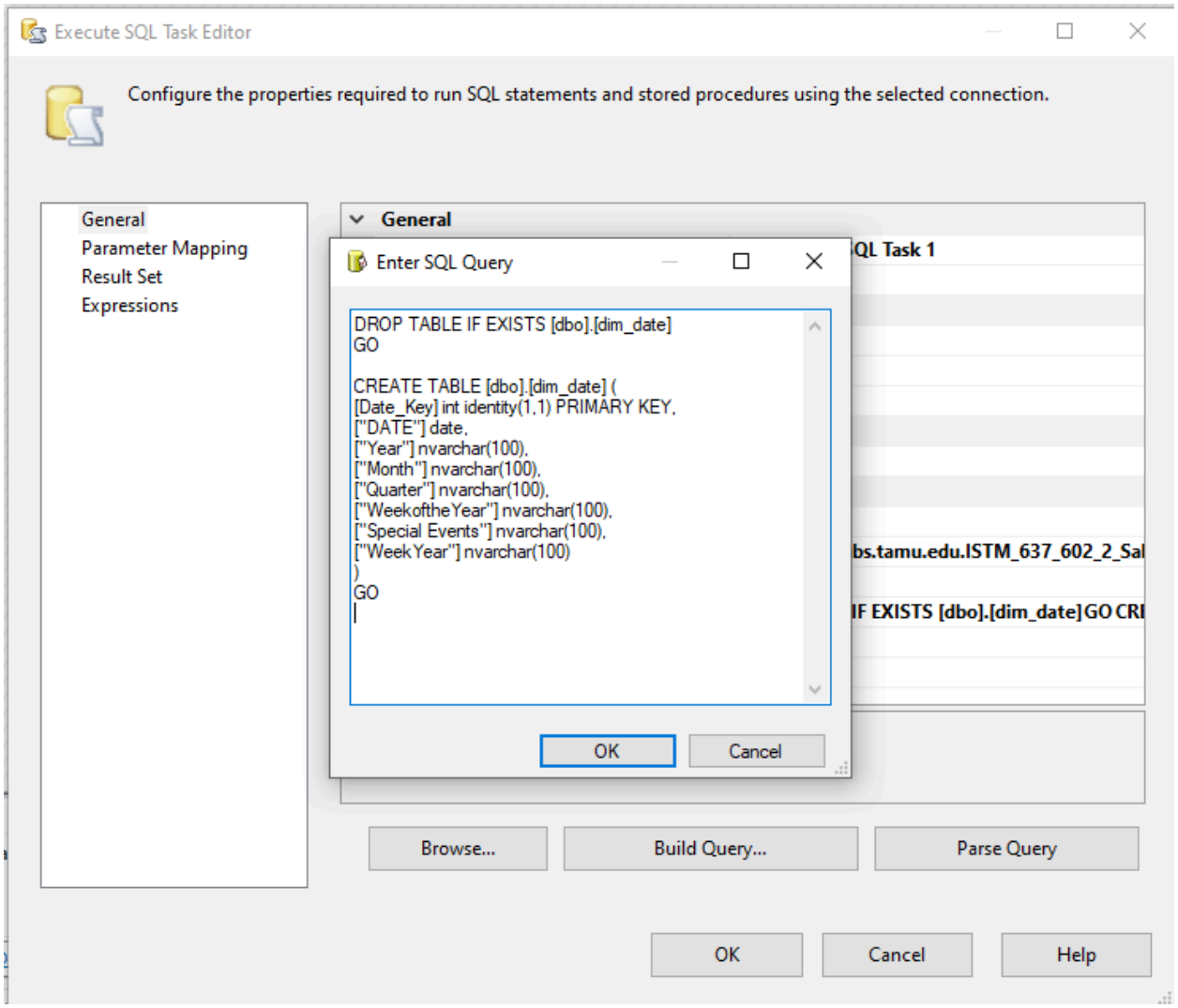


Fig: dim_date Table Creation

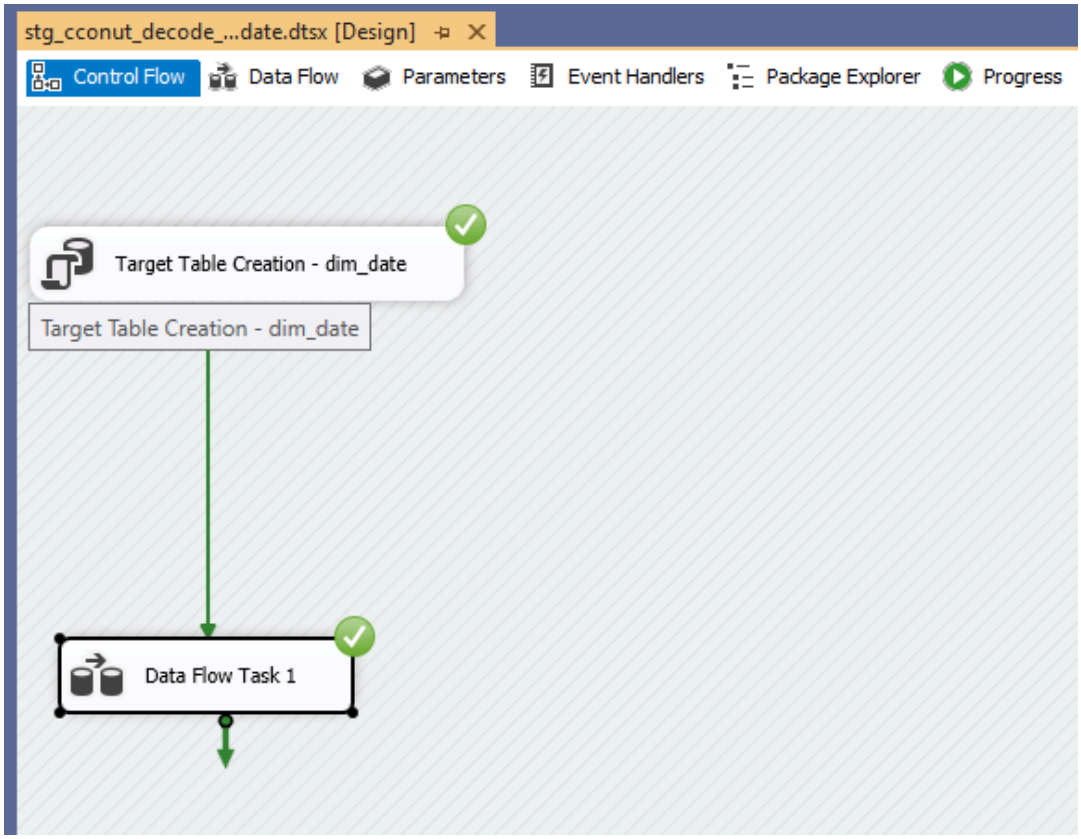


Fig: dim_date Control Flow

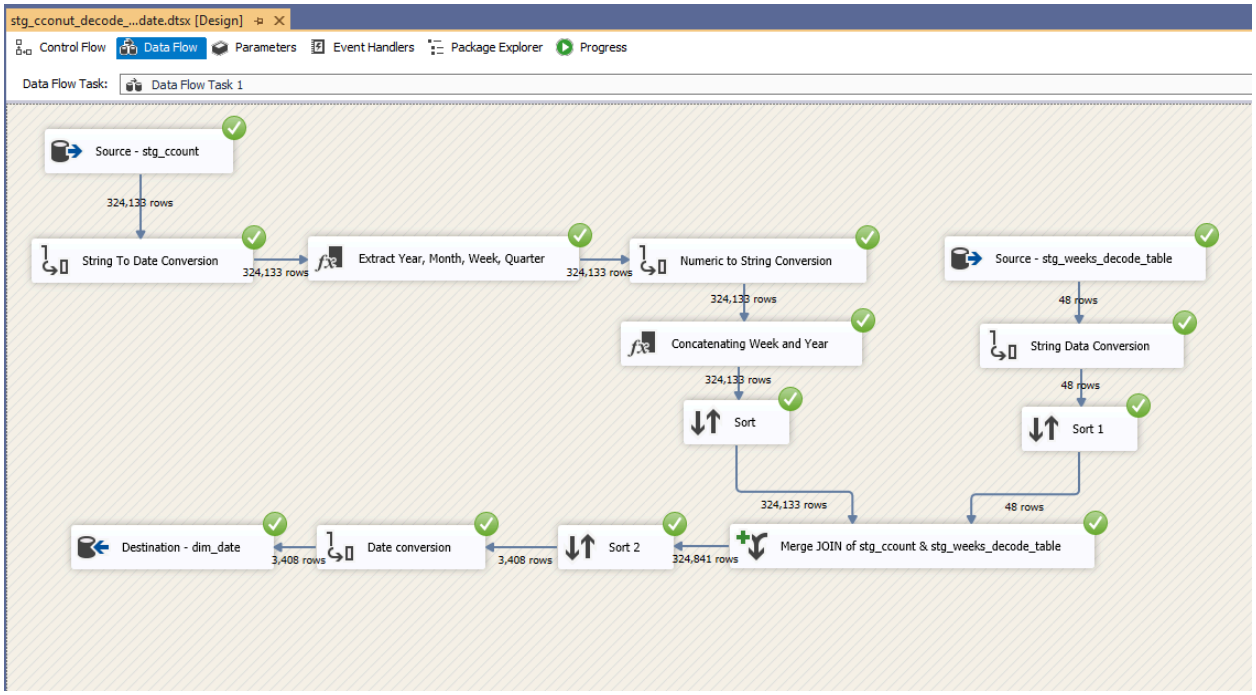


Fig: dim_date Data Flow

SQLQuery17.sql - i...amsibezawada (56)) | SQLQuery16.sql - i...amsibezawada (99)) | SQLQuery15.sql - i...

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [Date_Key]
      ,["DATE"]
      ,["Year"]
      ,["Month"]
      ,["Quarter"]
      ,["WeekoftheYear"]
      ,["Special Events"]
      ,["WeekYear"]
  FROM [ISTM_637_602_2_Sales_DM].[dbo].[dim_date]
```

100 %

Results | Messages

	Date_Key	"DATE"	"Year"	"Month"	"Quarter"	"WeekoftheYear"	"Special Events"	"WeekYear"
720	720	1989-12-14	1989	12	3	50	NULL	198950
721	721	1989-12-15	1989	12	3	50	NULL	198950
722	722	1989-12-16	1989	12	3	50	NULL	198950
723	723	1989-12-17	1989	12	3	51	Christmas	198951
724	724	1989-12-18	1989	12	3	51	Christmas	198951
725	725	1989-12-19	1989	12	3	51	Christmas	198951
726	726	1989-12-20	1989	12	3	51	Christmas	198951
727	727	1989-12-21	1989	12	3	51	Christmas	198951
728	728	1989-12-22	1989	12	3	51	Christmas	198951
729	729	1989-12-23	1989	12	3	51	Christmas	198951
730	730	1989-12-24	1989	12	3	52	New-Year	198952
731	731	1989-12-25	1989	12	3	52	New-Year	198952
732	732	1989-12-26	1989	12	3	52	New-Year	198952
733	733	1989-12-27	1989	12	3	52	New-Year	198952
734	734	1989-12-28	1989	12	3	52	New-Year	198952
735	735	1989-12-29	1989	12	3	52	New-Year	198952
736	736	1989-12-30	1989	12	3	52	New-Year	198952
737	737	1989-12-31	1989	12	3	53	NULL	198953
738	738	1990-01-01	1990	1	0	1	NULL	19901

Query executed successfully.

Fig: dim_demo Sample Data Retrieval

Dim_Product

The prod_index.csv file is generated according to the manual and specific business requirements, mapping products to their descriptions based on data from the Movement and UPC files.

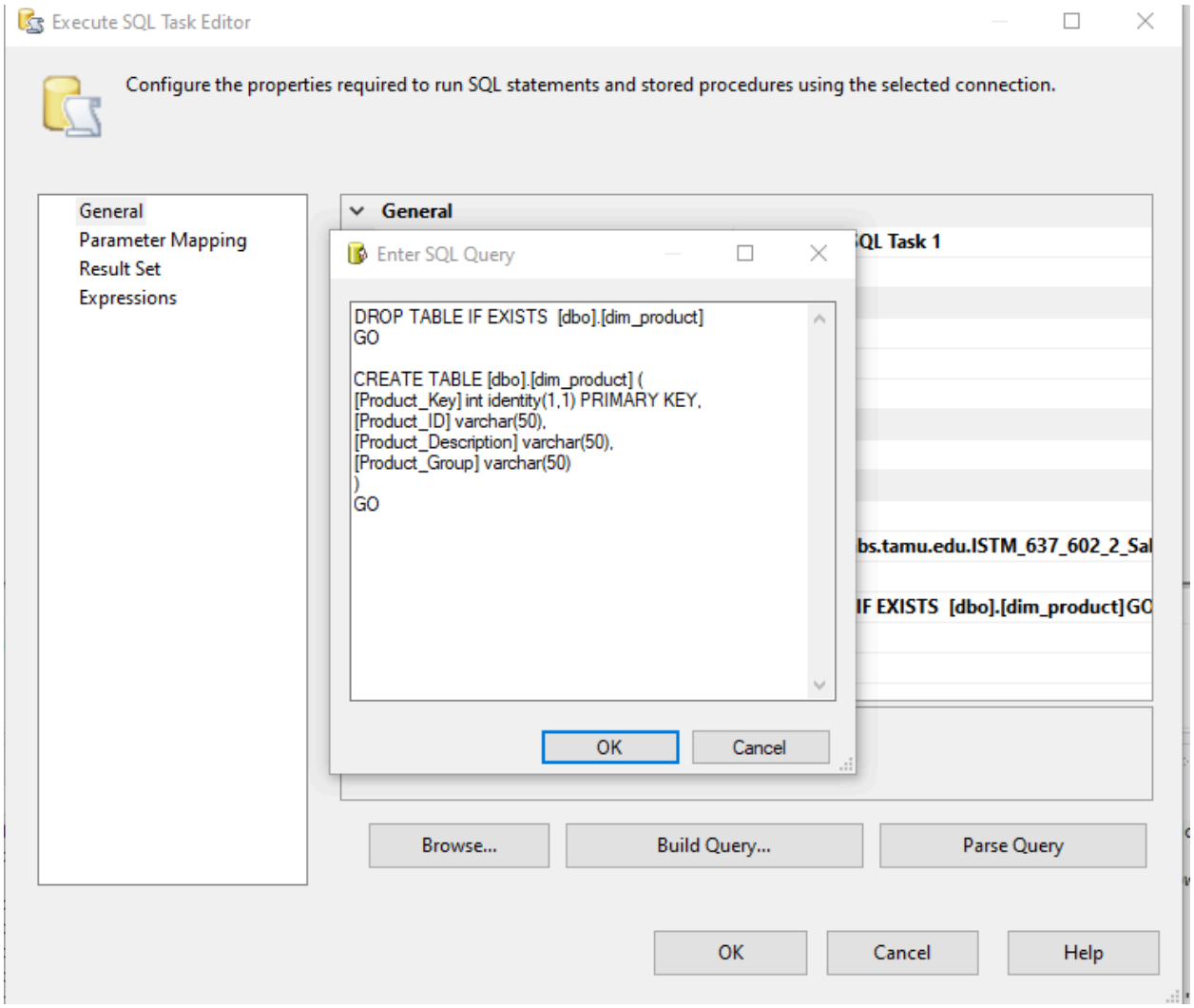


Fig: dim_product Table Creation

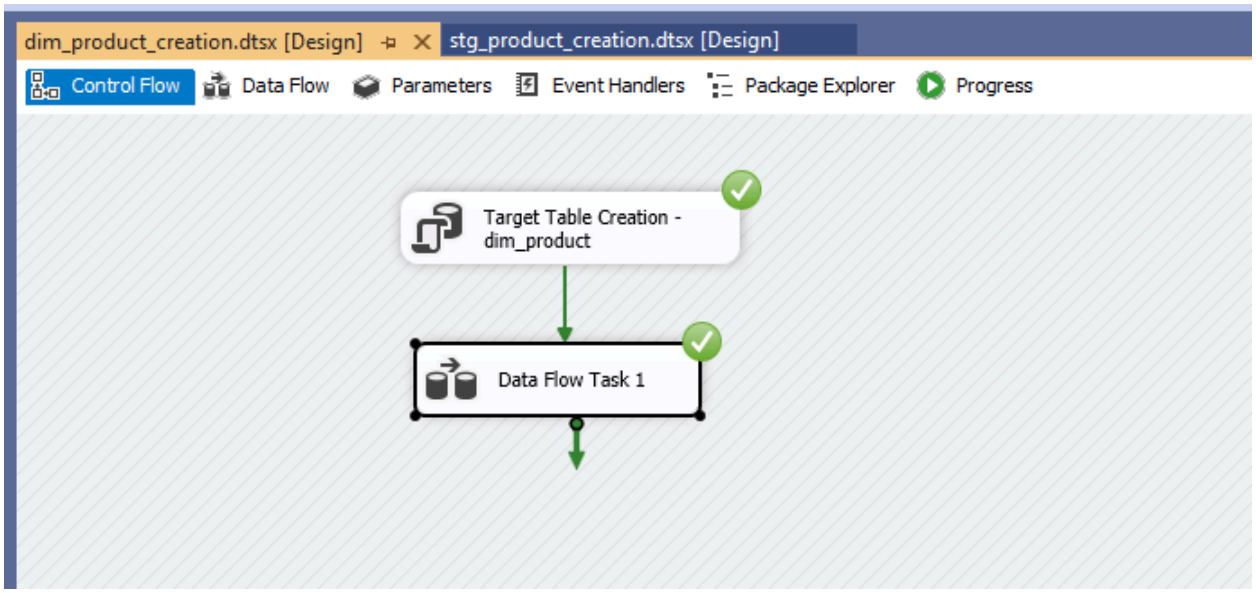


Fig: dim_product Control Flow

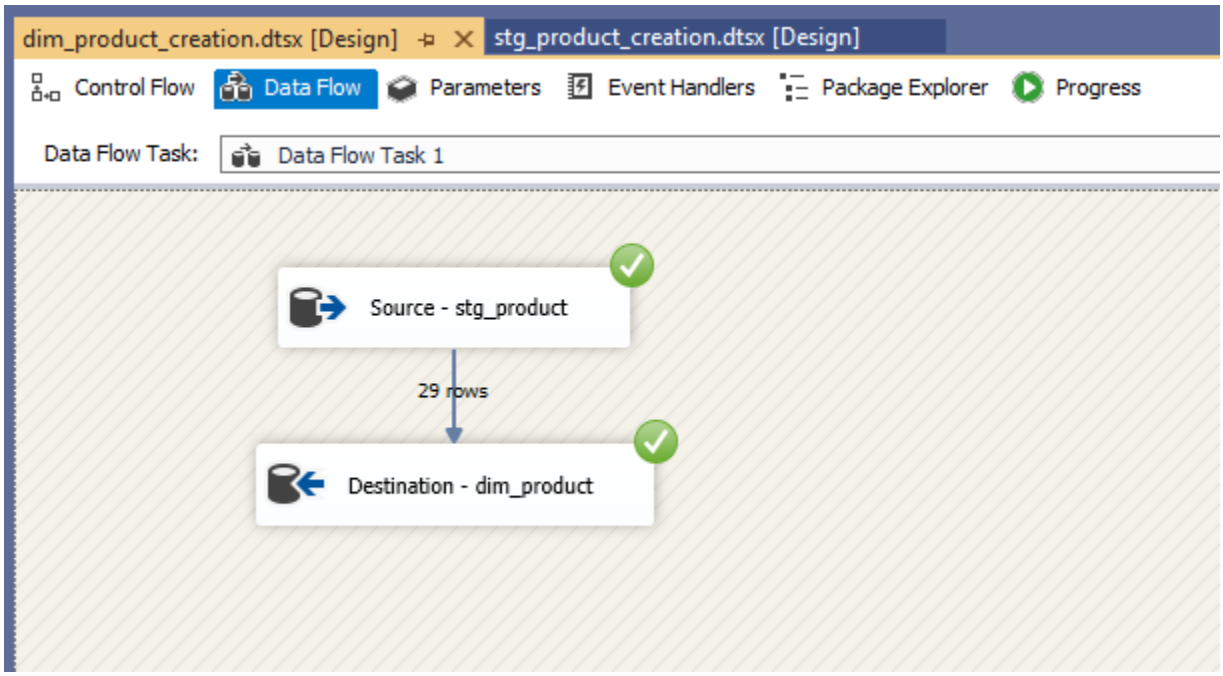


Fig: dim_product Data Flow

SQLQuery28.sql - i...amsibezawada (66)) | SQLQuery27.sql - i...amsibezawada (61))

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [Product_Key]
      ,[Product_ID]
      ,[Product_Description]
      ,[Product_Group]
  FROM [ISTM_637_602_2_Sales_DM].[dbo].[dim_product]
```

100 %

Results | Messages

	Product_Key	Product_ID	Product_Description	Product_Group
1	1	BER	BEER	Food
2	2	BOT	BOTTLE	Non-Food
3	3	CAM	CAMERA	Non-Food
4	4	COSC	COSMCOUP	Non-Food
5	5	COS	COSMETIC	Non-Food
6	6	DAIC	DAIRCOUP	Food
7	7	DAI	DAIRY	Food
8	8	DEL	DELI	Food
9	9	FISH	FISH	Food
10	10	FISC	FISHCOUP	Food
11	11	FLO	FLORAL	Non-Food
12	12	FLOC	FLORCOUP	Non-Food
13	13	FROC	FROZCOUP	Food
14	14	FRO	FROZEN	Food
15	15	GROC	GROCCOUP	Food
16	16	GRO	GROCERY	Food
17	17	HAB	HABA	Non-Food
18	18	HABC	HABACOUP	Non-Food
19	19	JEW	JEWELRY	Non-Food

Query executed successfully.

Fig: dim_product Sample Data Retrieval

Fact Tables

Fact_Sales

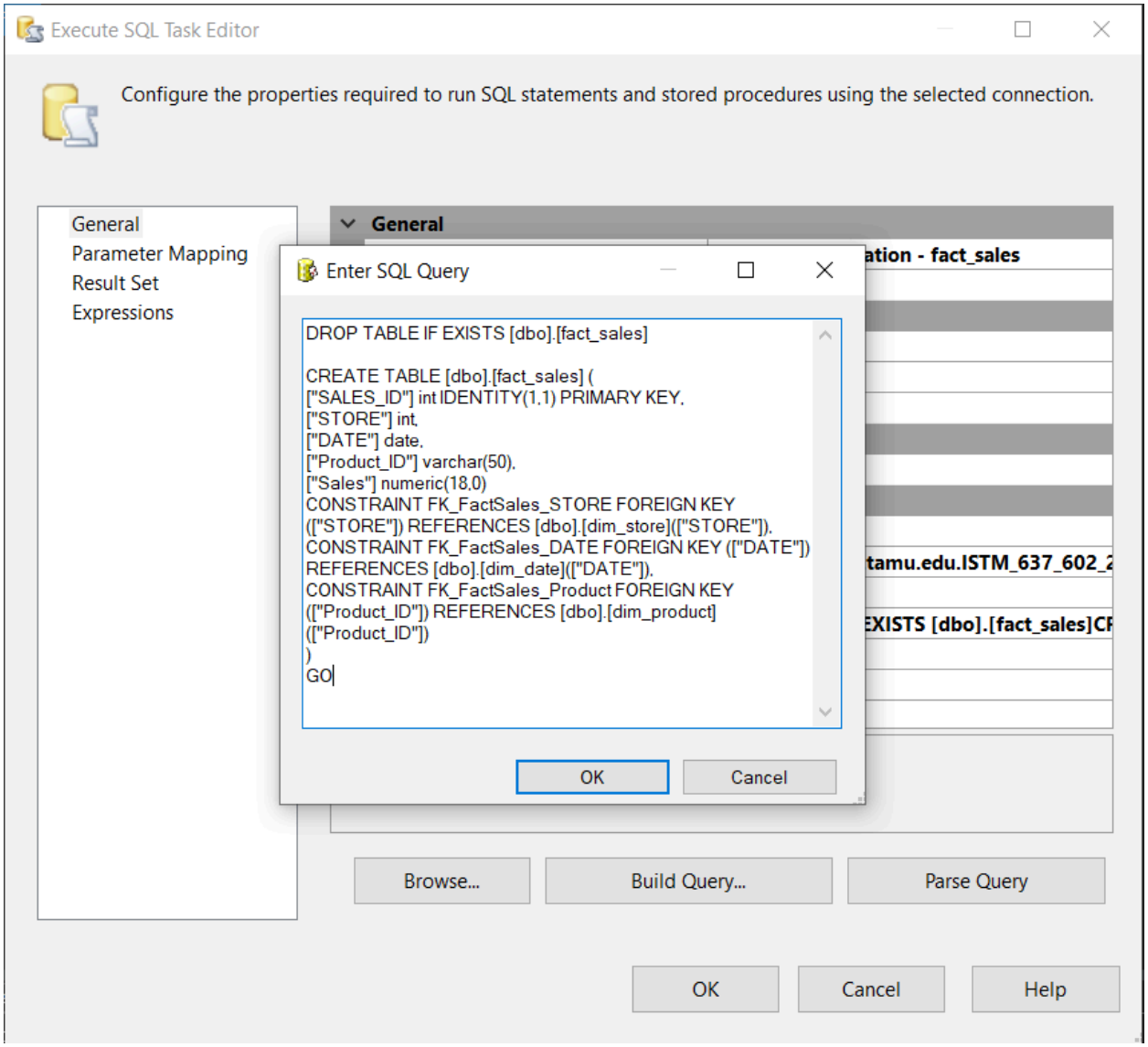


Fig: fact_sales Table Creation

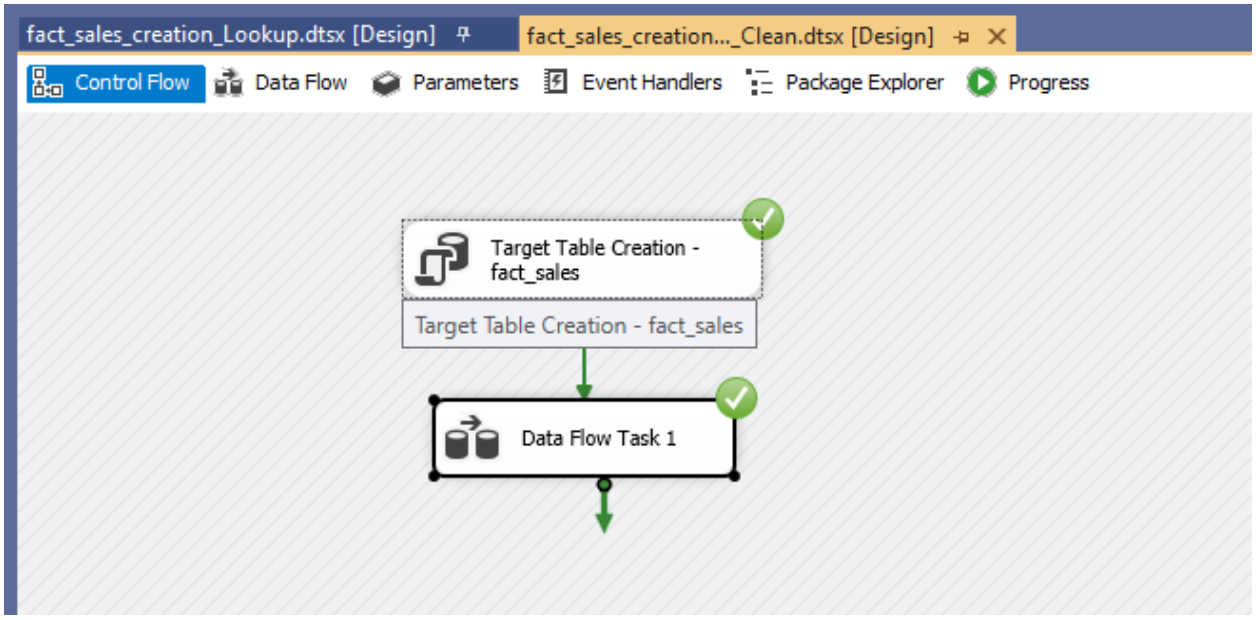


Fig: fact_sales Control Flow

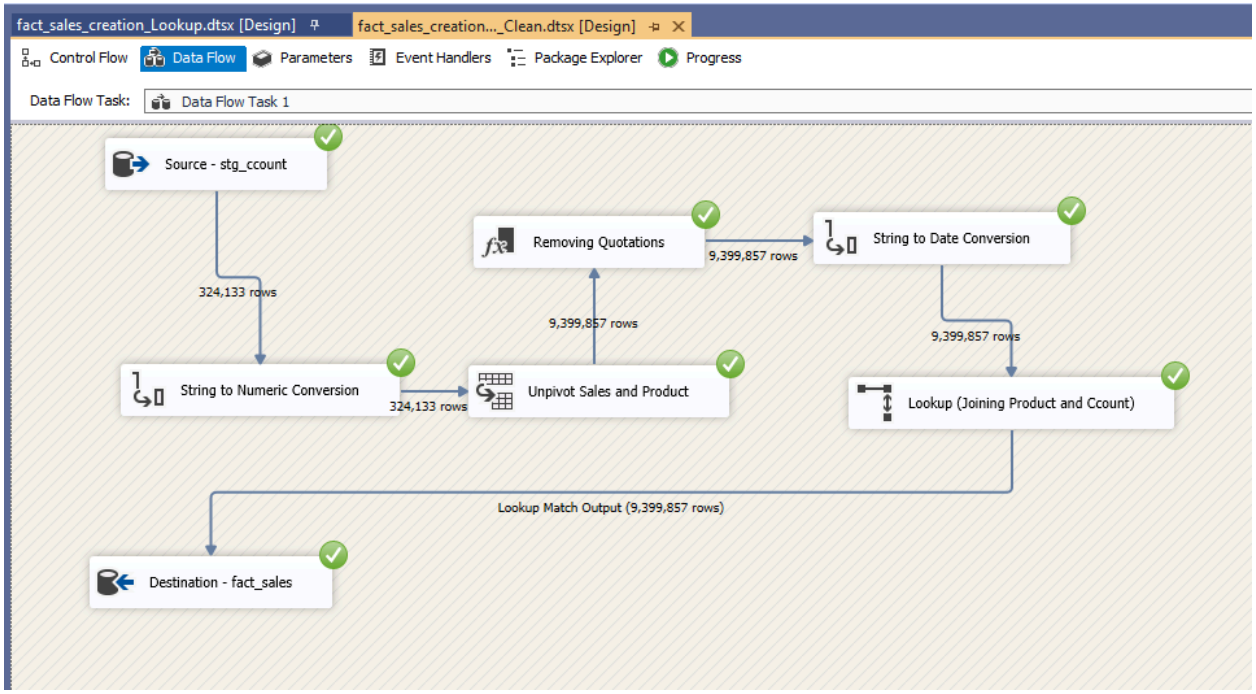


Fig: fact_sales Data flow

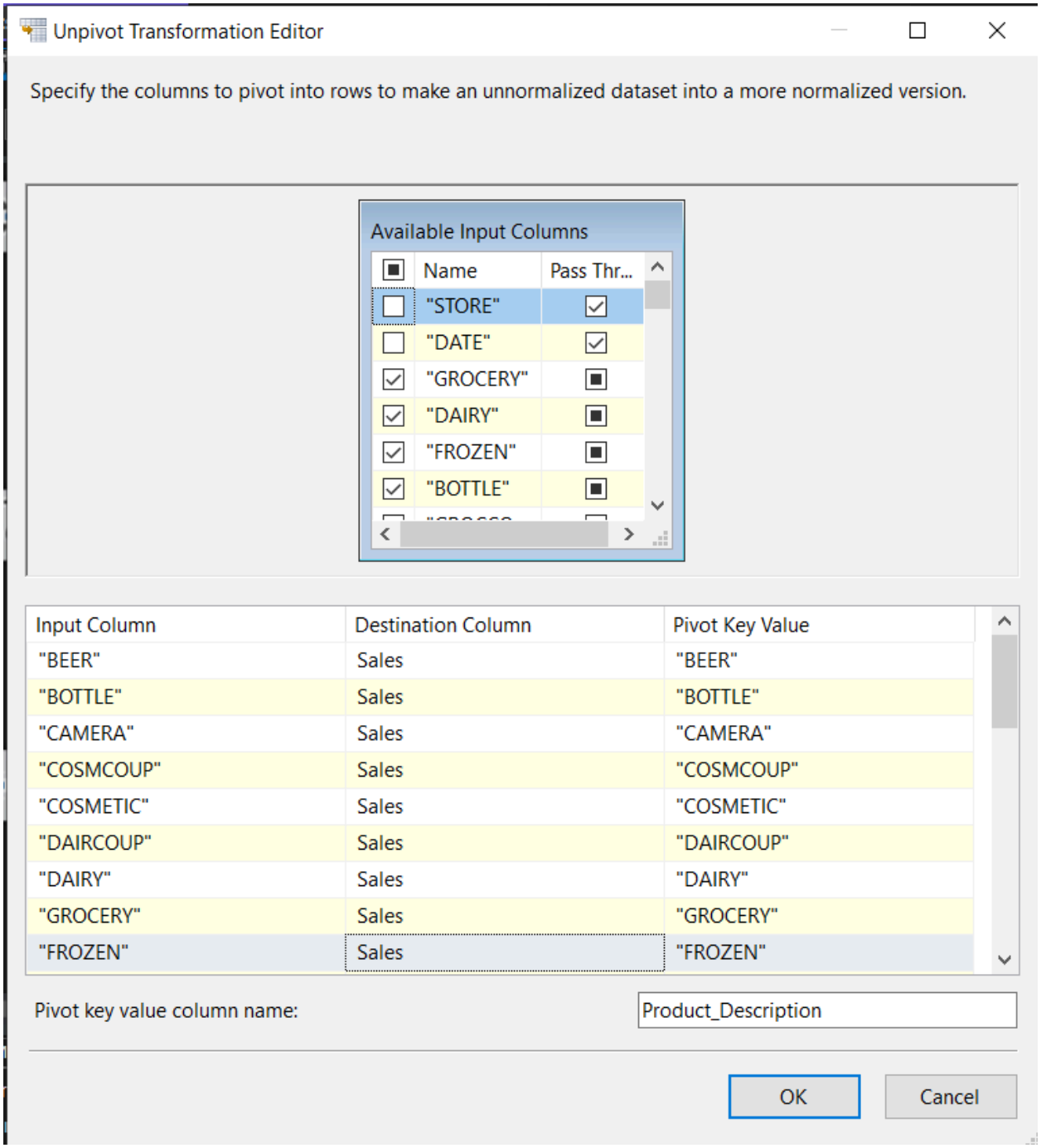


Fig: fact_sales Unpivot Transformation

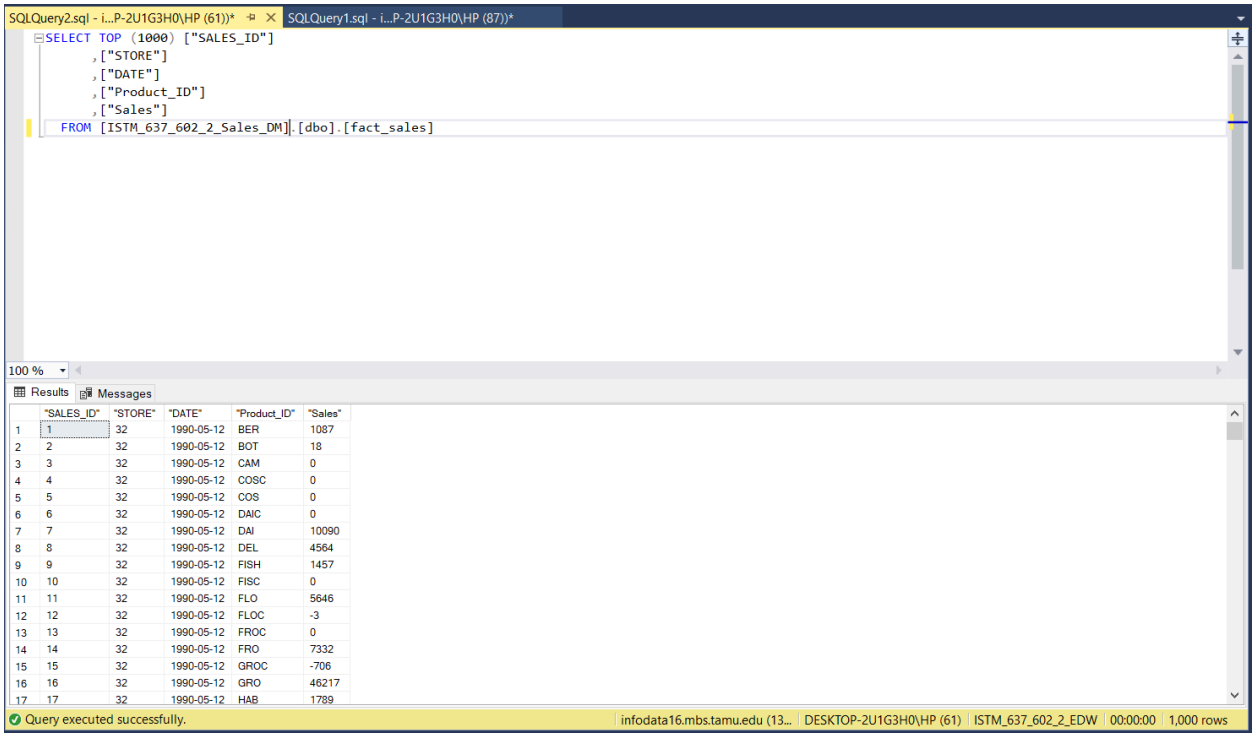


Fig: fact_sales Sample Data Retrieval

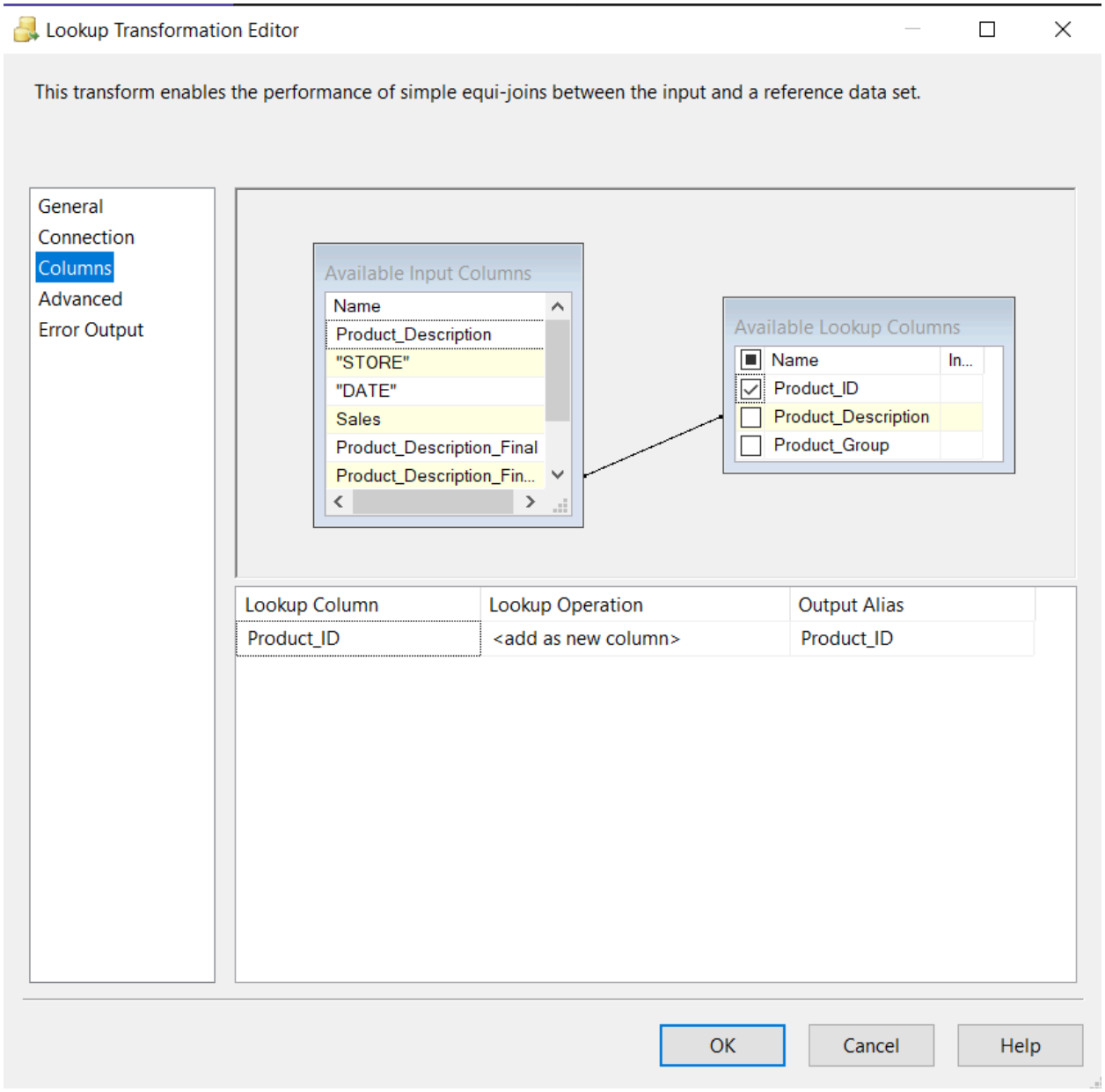


Fig: fact_sales Lookup Transformation

Fact_Customer_Count

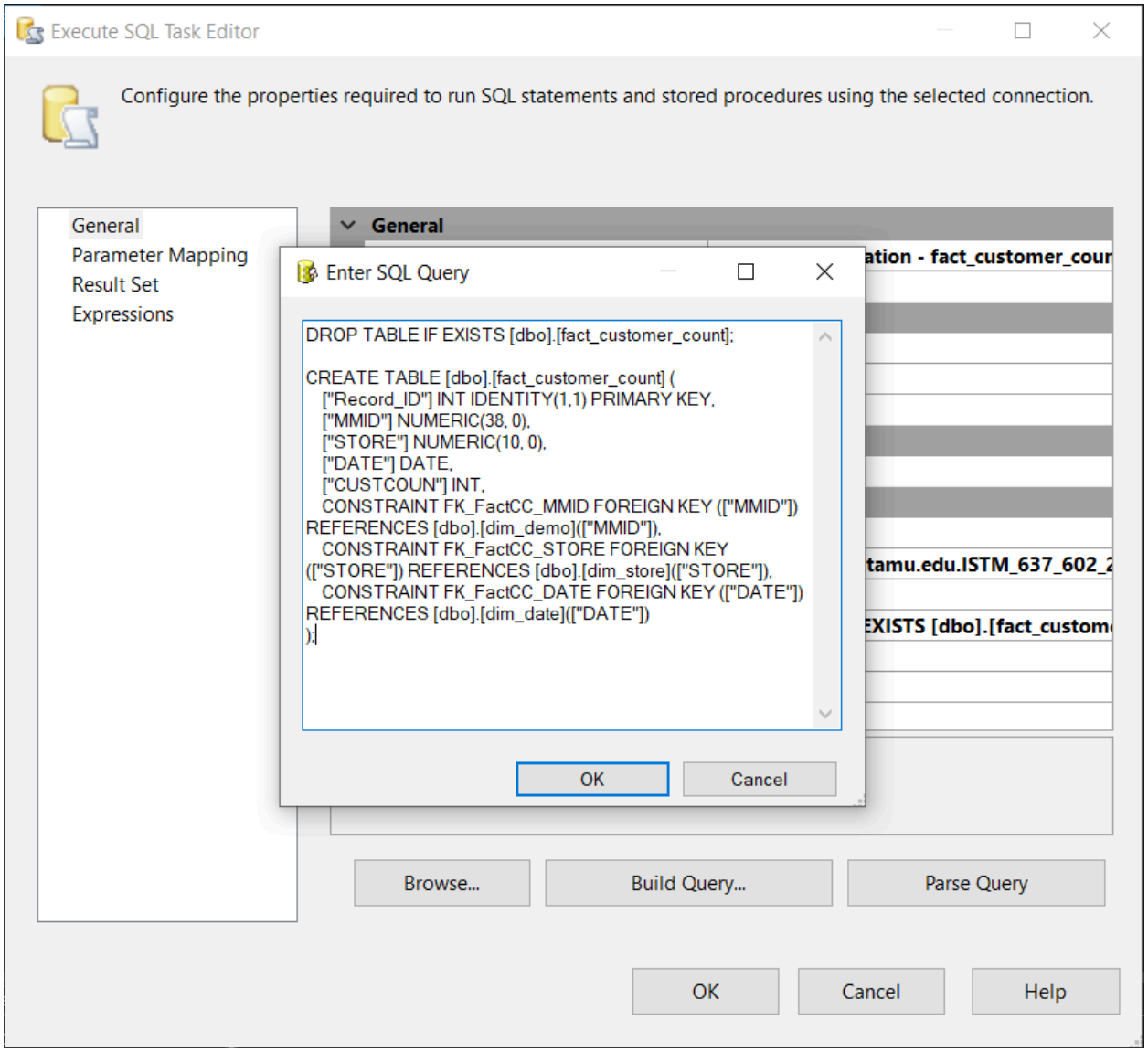


Fig: fact_customer_count Table Creation

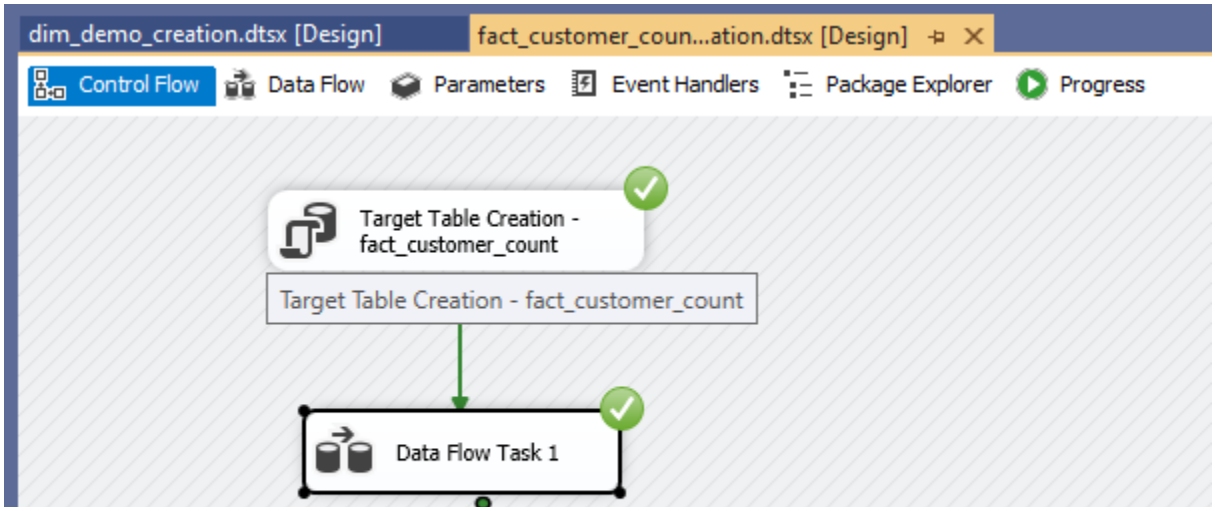


Fig: fact_customer_count Control Flow

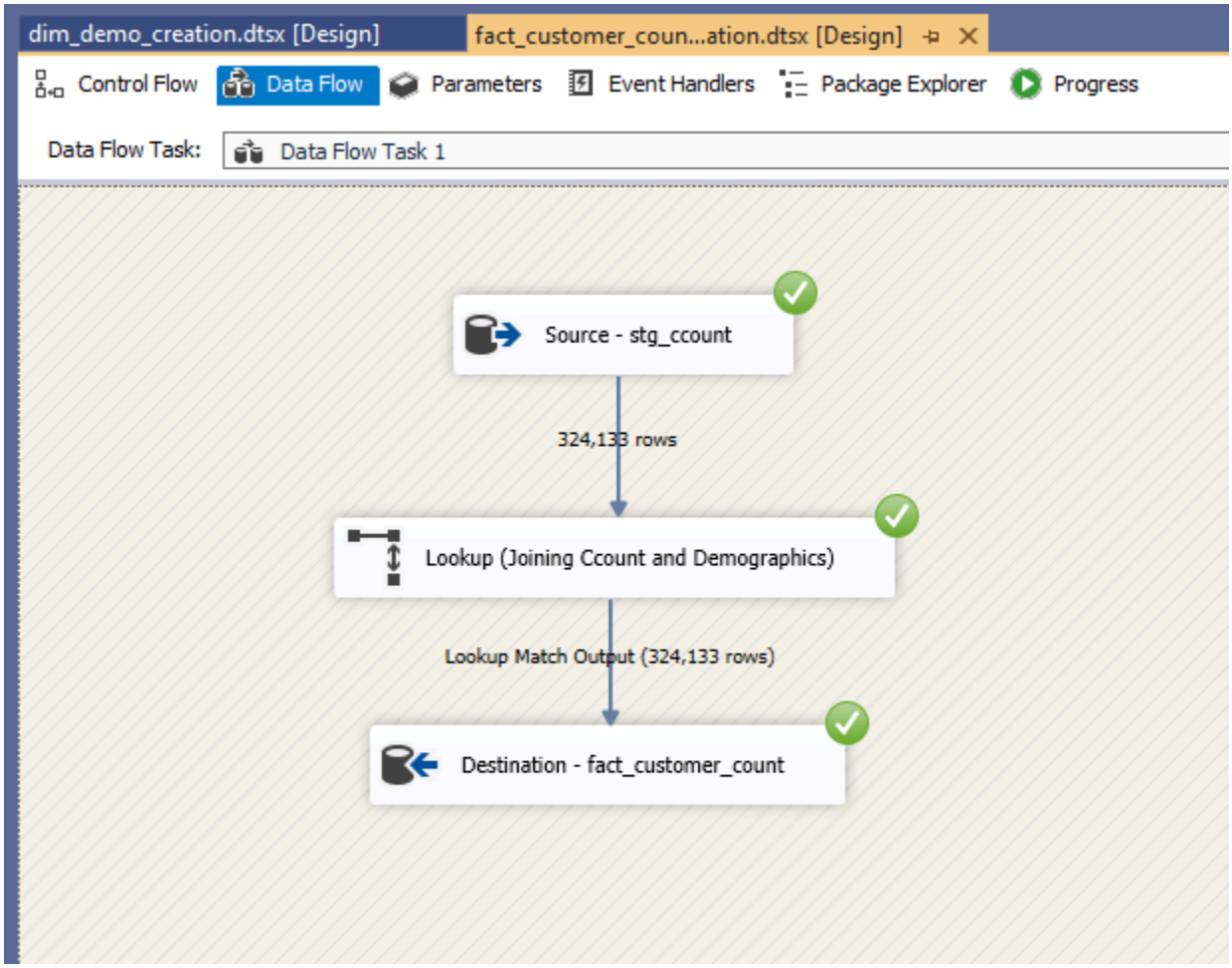


Fig: fact_customer_count Data Flow

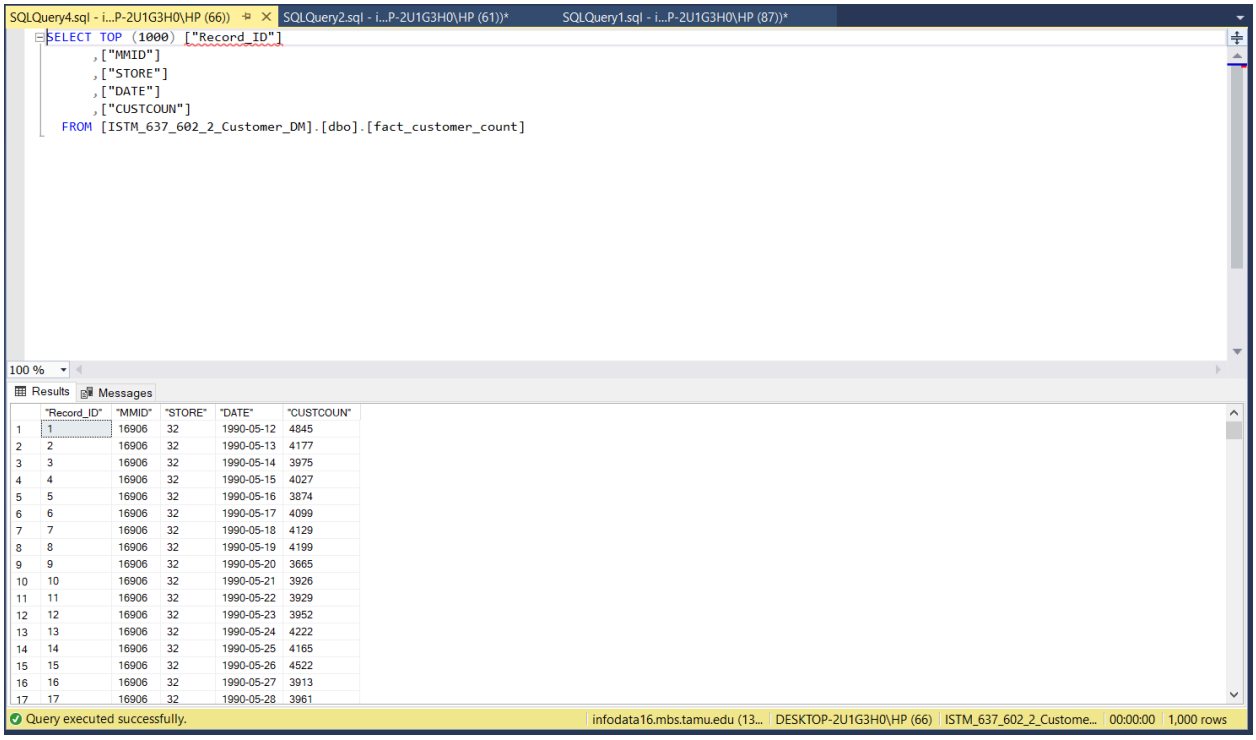


Fig: fact_customer_count Sample Data Retrieval

Data Granularity for Independent Data marts

For both data marts, we have defined five levels of data granularity: day, month, year, quarter, and WeekYear. These levels are used to analyze product sales and their categories in the Fact_Sales data mart, as well as to examine customer counts in the Fact_Customer_Count data mart.

SQL STATEMENTS USED

```
DELETE FROM [dbo].[stg_ccount_temp] where ["DATE"] LIKE '%[a-Z]%';
DELETE FROM [dbo].[stg_ccount_temp] where ["DATE"] NOT LIKE
'%[1-9]%';
DELETE FROM [dbo].[stg_ccount_temp] WHERE ["DATE"] NOT LIKE
'%[0-9]%';
DELETE FROM [dbo].[stg_demo] WHERE ["MMID"] IS NULL
  OR ["MMID"] LIKE '%[^0-9]%'  -- Non-numeric characters
  OR ["MMID"] LIKE '%.%'       -- Full stop (.)
  OR ISNUMERIC(["MMID"]) = 0;

UPDATE [dbo].[stg_ccount_temp]  SET ["DATE"] =
REPLACE(["DATE"],'"','');
UPDATE [dbo].[stg_store] SET "Zip Code" = NULL WHERE "Zip Code" = '
';
UPDATE [dbo].[stg_store] SET "Zone" = NULL WHERE "Zone" = ' ';
UPDATE [dbo].[stg_store] SET "Price Tier" = NULL WHERE "Price Tier" =
' ';
```

Dimension Tables Creations

```
DROP TABLE IF EXISTS [dbo].[dim_demo]

CREATE TABLE [dbo].[dim_demo] (
[Demo_Key] int IDENTITY(1,1) PRIMARY KEY,
["MMID"] numeric(38,0),
["SHPCONS"] numeric(38,4),
["SHPHURR"] numeric(38,4),
["SHPAVID"] numeric(38,4),
["SHPKSTR"] numeric(38,4),
["SHPUNFT"] numeric(38,4),
["SHPBIRD"] numeric(38,4)
)
```

```
DROP TABLE IF EXISTS [dbo].[dim_product]
CREATE TABLE [dbo].[dim_product] (
[Product_Key] int IDENTITY(1,1) PRIMARY KEY,
[Product_ID] varchar(50),
[Product_Description] varchar(50),
[Product_Group] varchar(50)
)

DROP TABLE IF EXISTS [dbo].[dim_store]
CREATE TABLE [dbo].[dim_store] (
[Store_Key] int IDENTITY(1,1) PRIMARY KEY,
[Store] numeric(10,0) PRIMARY KEY,
[City] varchar(50),
[Price Tier] varchar(50),
[Zone] numeric(10,0),
[Zip Code] numeric(10,0),
[Address] varchar(50)
)

DROP TABLE IF EXISTS [dbo].[dim_date]
CREATE TABLE [dbo].[dim_date] (
[Date_Key] int IDENTITY(1,1) PRIMARY KEY,
["DATE"] date,
["Year"] nvarchar(100),
["Month"] nvarchar(100),
["Quarter"] nvarchar(100),
["WeekoftheYear"] nvarchar(100),
["Special Events"] nvarchar(100),
["WeekYear"] nvarchar(100)
```

```
)
```

Fact Tables Creations

```
DROP TABLE IF EXISTS [dbo].[fact_sales]

CREATE TABLE [dbo].[fact_sales] (
["SALES_ID"] int IDENTITY(1,1) PRIMARY KEY,
["STORE"] int,
["DATE"] date,
["Product_ID"] varchar(50),
["Sales"] numeric(18,0)
)

DROP TABLE IF EXISTS [dbo].[fact_customer_count]
CREATE TABLE [dbo].[fact_customer_count] (
["Record_ID"] INT IDENTITY(1,1) PRIMARY KEY,
["MMID"] numeric(38,0),
["STORE"] numeric(10,0),
["DATE"] date,
["CUSTCOUN"] int
)
```

Temporary Tables Removed from staging area

[ISTM_637_602_2_Stg].[sbo].[stg_ccount_tmp]

Business Intelligence Reporting

Reporting Plan

After completing the ETL process to load cleansed, transformed, and aggregated data into a centralized data warehouse, we now shift our focus to designing and delivering Business Intelligence (BI) reports. These reports will translate raw data into actionable insights, addressing key business questions and enabling informed decision-making.

To ensure alignment with organizational goals and stakeholder expectations, we will first develop a comprehensive reporting plan. This plan will leverage the prepared data from independent data marts to produce insights that seamlessly support strategic objectives.

The primary objectives of this reporting plan are:

- Transform data from independent marts into meaningful, business-ready insights.
- Deliver intuitive and interactive reports tailored to different stakeholder groups.
- Address strategic and operational needs using tools such as SSRS, SSAS, Power BI, Tableau, and others.
- Develop reusable templates to ensure consistency and ease of use.

This plan will provide detailed guidelines for:

- Mapping report attributes to the data marts.
- Selecting the appropriate BI tools and technologies for various reporting requirements.
- Designing report templates to address business questions effectively.
- Designing reports based on the templates.

To effectively meet the diverse business reporting needs, we plan to use a variety of reporting systems. By integrating multiple systems, the organizations can cater to different user groups and scenarios, ranging from in-depth data exploration to high-level performance tracking. Below are some reporting systems that we could use for our reporting.

SQL Server Reporting Services (SSRS): SQL Server Reporting Services is a robust reporting platform designed for creating static, paginated reports. It excels in producing detailed operational reports, such as sales invoices, inventory summaries, and financial statements, where precise formatting is critical. With its strong integration with SQL Server databases, SSRS ensures high performance and data accuracy. Additionally, it supports automated report scheduling and distribution, making it a valuable tool for consistent delivery of time-sensitive information.

SQL Server Analysis Services (SSAS): SQL Server Analysis Services can be used for more advanced data analysis. It enables the development of multidimensional cubes and tabular models that simplify complex data exploration. SSAS is particularly useful for drilling down into large datasets to uncover hidden patterns and trends. SSAS can be used to enhance the performance of BI tools like Power BI and Tableau.

Tableau: Tableau is known for its ability to transform data into compelling visual narratives. It is ideal for creating interactive dashboards that allow users to explore data dynamically. Tableau's interface allows analysts to build advanced visualizations without requiring extensive coding knowledge. Its ability to connect to multiple data sources, both on-premises and in the cloud, makes it versatile for diverse business use cases.

Report Builder: Report Builder is a lightweight and user-friendly tool for creating straightforward reports. Designed for ad hoc reporting, it empowers non-technical users to generate simple, on-demand reports. Working seamlessly with SSRS, Report Builder ensures consistency in report output while requiring minimal technical expertise. It is useful for generating quick insights with minimal setup time, making it an excellent choice for business users seeking fast access to data and reports.

Power BI: Power BI is a powerful business intelligence platform that supports interactive dashboards and real-time data visualization. Power BI's seamless integration with Microsoft's ecosystem, including Excel and Azure, enhances its functionality for users familiar with these tools. With its cloud accessibility, Power BI allows stakeholders to access reports anytime, anywhere, supporting agile decision-making processes.

Report Templates

With so many reporting systems available to us, it is important to maintain consistency to enable organizations to work effectively. This is where reporting templates come into play. Report templates are pre-defined layouts that establish a consistent structure for reports, ensuring clarity and ease of interpretation across various reporting needs. They serve as blueprints for designing reports, standardizing the inclusion of all essential elements. Using report templates makes the reporting process more efficient, with a uniform presentation style that stakeholders can easily navigate and understand.

Visualizations play a central role in report templates by representing data through charts, graphs, and tables. Templates also include interactive elements, such as filters and parameters, allowing users to customize the report view based on specific criteria like date range, region, or product category. This interactivity enhances the user experience by providing more relevant and actionable insights. Incorporating well-defined templates ensures that all reports align with organizational standards and branding while addressing unique reporting requirements.

One of the primary benefits of using report templates is maintaining consistency in design and functionality across all reports. This consistency ensures that end-users can quickly adapt to new reports without requiring extensive onboarding or training. There are different report templates for various business use cases. Some examples include Sales Performance Report Templates and Financial Analysis Report Templates. A Sales Performance Report Template includes sections such as tracking monthly sales trends, identifying top-performing regions, and summarizing total revenue. A Financial Report Template, on the other hand, focuses on key metrics like net profit, revenue growth, and cost distribution across business units.

Let us now take a look at how we can answer each business question with different reporting systems and reporting templates available to us.

Target Reports - 5 Business Questions selected by DFF

Below are the 5 business questions that DFF has selected. Refer to the questions 1, 4, 6, 7, and 9 in the section - 10 Business Questions for justification and rationale for each question.

The below are the questions selected for specific report template

S. No	Business Question	Report Templates
1	What is the proportion of sales made with coupons vs. without coupons in a given category?	Tableau with SSAS cube (Option 4)
2	How do food and non-food product sales compare across all stores, segmented by year and week number?	Tableau (Option 3)
3	What are the top-selling departments (e.g., Bakery, Meat, Deli) across stores, and how do they vary by store and week?	SSRS (Option 2)
4	Determine the sales of a given category during the weeks of special events.	Tableau (Option 3)
5	What is the annual customer inflow for each store allowing for insights into customer behavior and store performance?	SSAS Cube (Option1)

Mapping Attributes from independent data marts

1. What is the proportion of sales made with coupons vs. without coupons in a given category?

Attribute Name	**Attribute's Dimension/Fact Table**	**Filters**	**Corresponding Report Attribute**
Store	dim_store	All	Store
Total Sales	fact_sales	NA	Sum([Coupon Sales])/SUM([Non Coupon Sales]) *100

2. How do food and non-food product sales compare across all stores, segmented by year and week number?

Attribute Name	**Attribute's Dimension/Fact Table**	**Filters**	**Corresponding Report Attribute**
Store	dim_store	All	Store
Total Sales	fact_sales	NA	Sales
Product_Group	dim_product	Food, Non-Food	Product_Group
Month	dim_date	All	Month
Year	dim_date	1989 and 1992	YEAR("Date")

3. What are the top-selling departments (e.g., Bakery, Meat, Deli) across stores, and how do they vary by store and week?

Attribute Name	**Attribute's Dimension/Fact Table**	**Filters**	**Corresponding Report Attribute**
Product_ID	dim_product	All	Product_ID
Average Sales	fact_sales	NA	Average Sales

4. Determine the sales of a given category during the weeks of special events.

Attribute Name	**Attribute's Dimension/Fact Table**	**Filters**	**Corresponding Report Attribute**
Year	dim_date	All	Year
Total Sales	fact_sales	NA	Sales
Product Description	dim_product	Beer, Camera, Cosmetic, HABA, Meat, Pharmacy, Wine	Product Description
Special Events	dim_date	4th of July, Halloween	Special Events

5. What is the annual customer inflow for each store allowing for insights into customer behavior and store performance?

Attribute Name	**Attribute's Dimension/Fact Table**	**Filters**	**Corresponding Report Attribute**
Year	dim_date	1987-1997	Year
Total Customer count	fact_customer_count	NA	Customer Inflow
Week of the Year	dim_date	All	Weekofthe Year
Store	dim_store	All	Store

Note: The filters seen in the above tables are the initial filters set on the report. These can be changed by the user dynamically when using the report to review the data and insights for different regions, time range, Stores & Zones.

Report Implementation

Question 1 : What is the proportion of sales made with coupons vs. without coupons in a given category?

Chosen Reporting System: Tableau with SSAS Cube

Justification: Tableau's powerful visualization capabilities combined with the analytical depth of SSAS cubes allow for detailed comparison and exploration of coupon vs. non-coupon sales proportions. SSAS cubes enable quick querying of large datasets, and Tableau effectively visualizes these insights using charts like pie charts or bar charts to depict proportions.

Report Template:

- Header: Report title
- Filters: Date Range, Store, and Product Category.
- Visualizations: Line chart with % of coupon sales as a percentage of non-coupon sales.
- Summary: Highlights the percentage of sales from those without coupons.

Data Source Creation

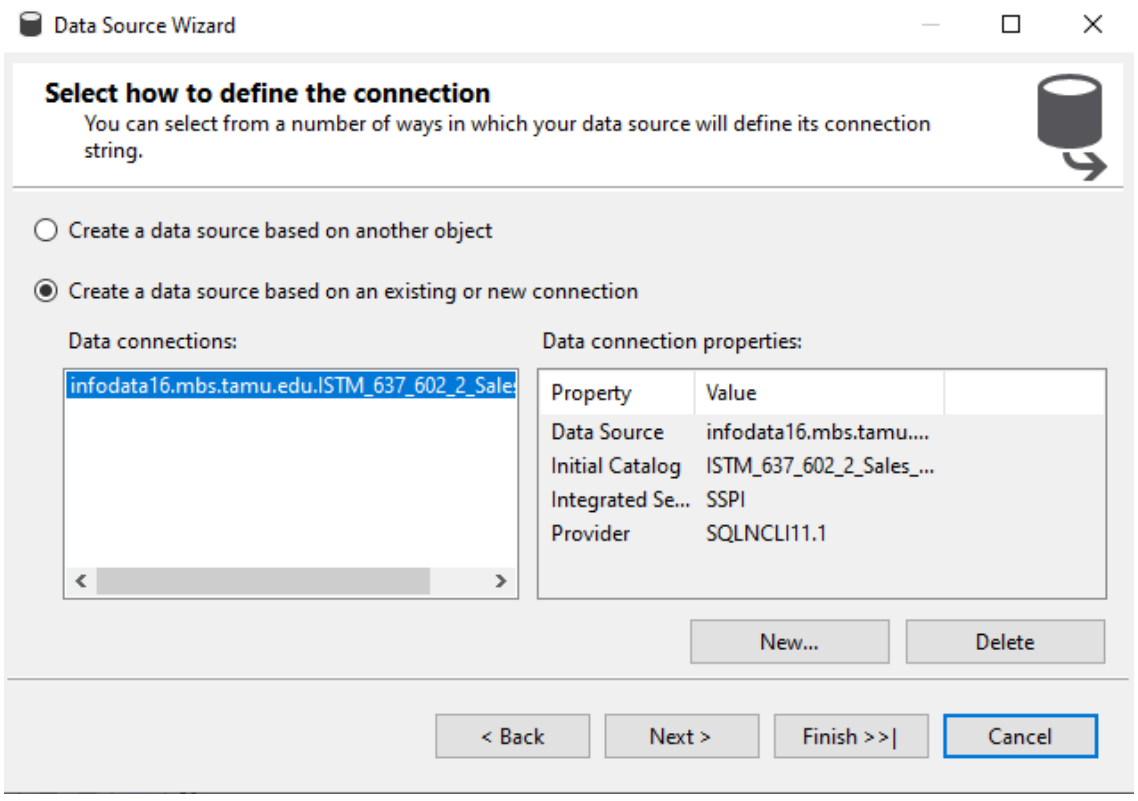


Fig: Creating the Data Source

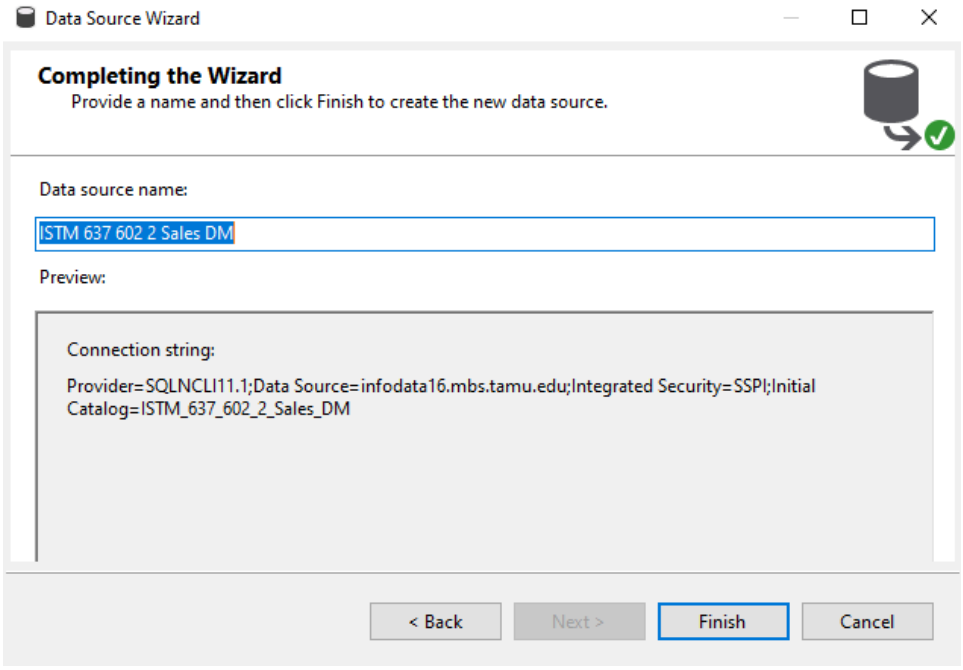


Fig: Connection String

Data Source View Creation

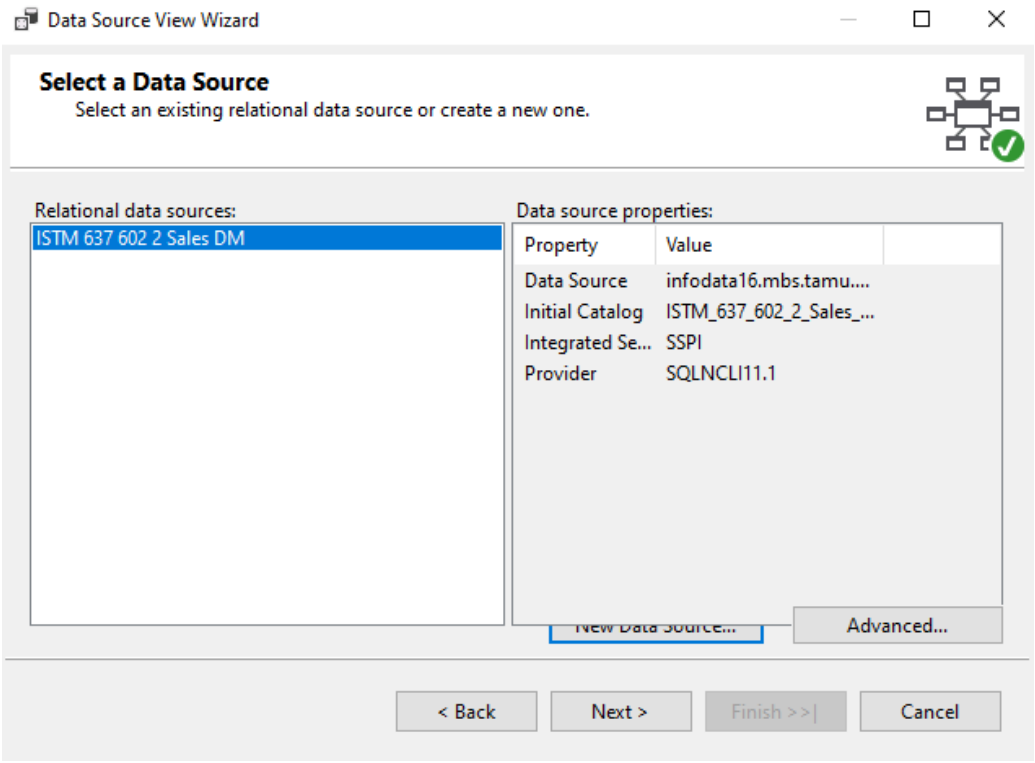


Fig: Data Source View Creation

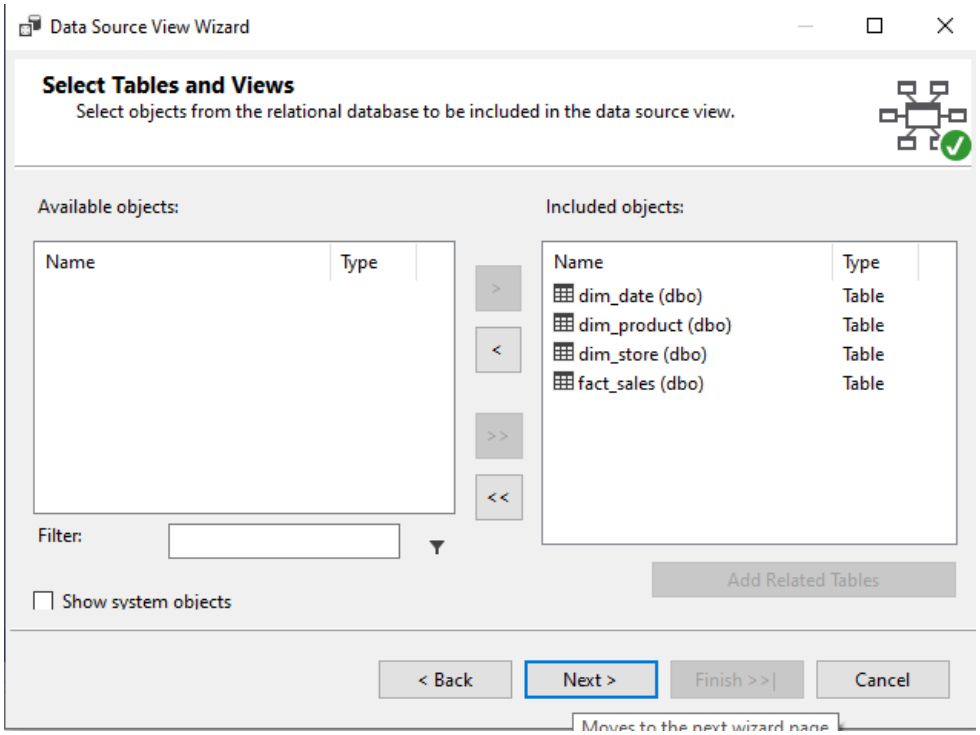


Fig: Table Selections

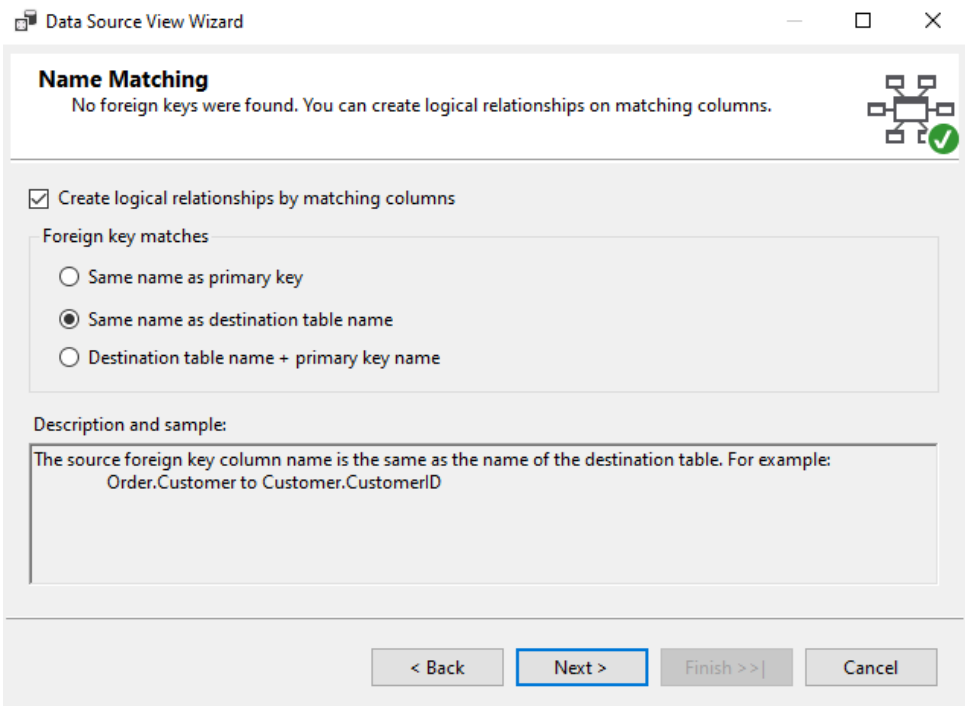


Fig: Relationship Matching between tables

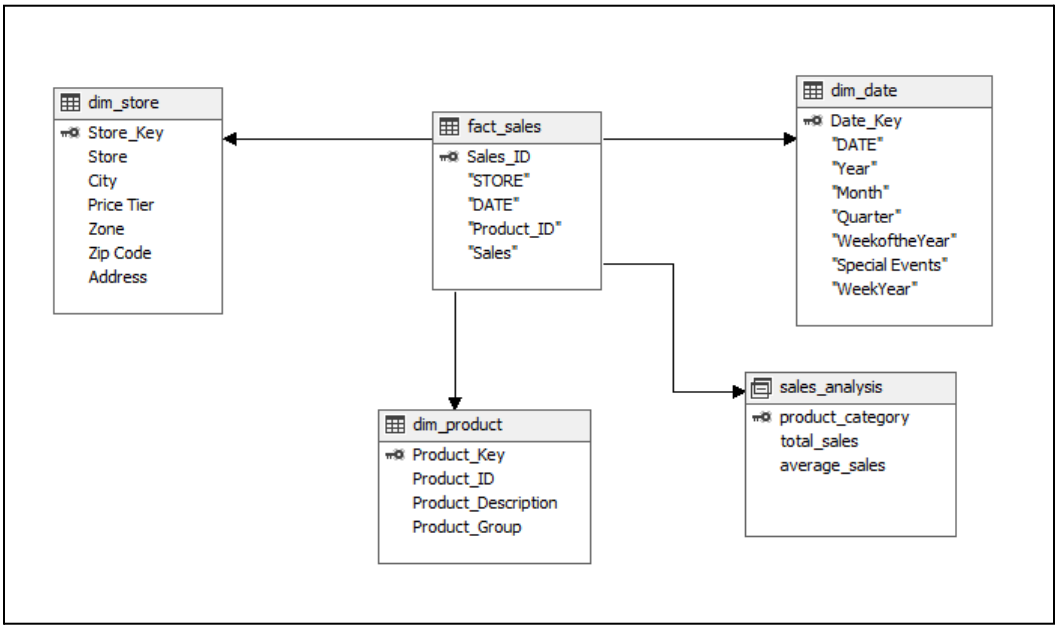


Fig: Data Source View Schema

Cube Creation

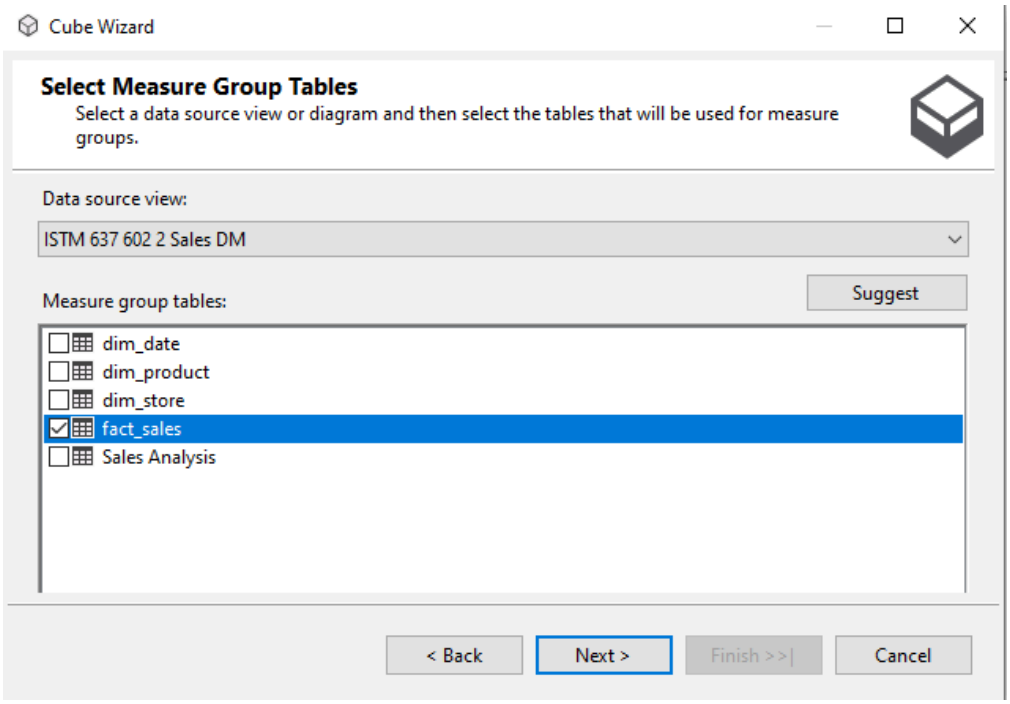


Fig: Selecting the Measure Group

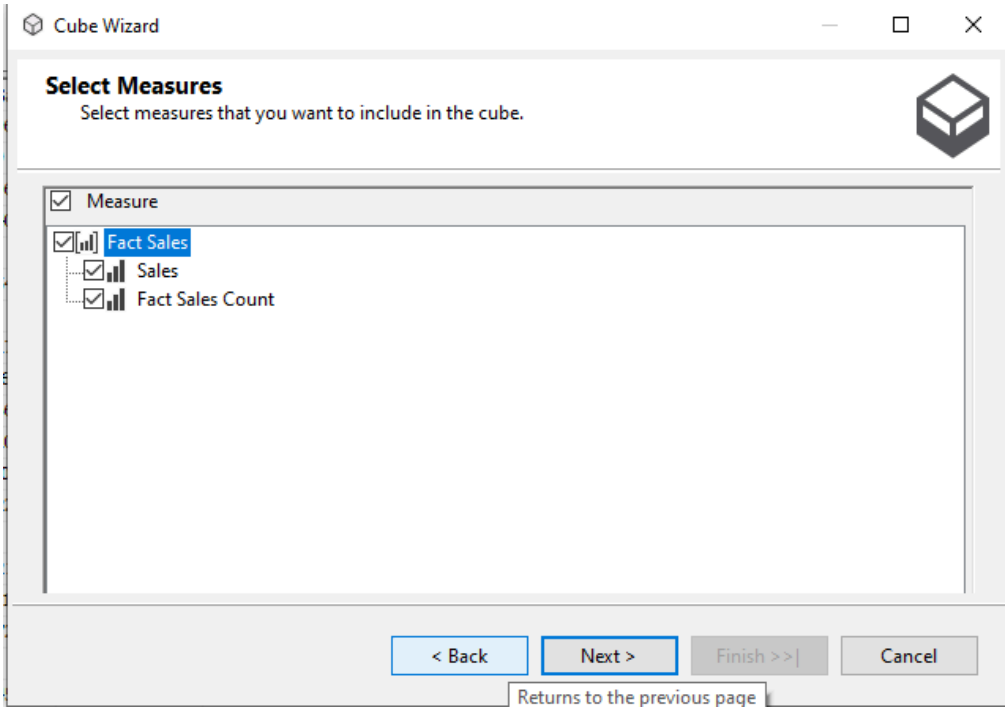


Fig: Selecting the Measures

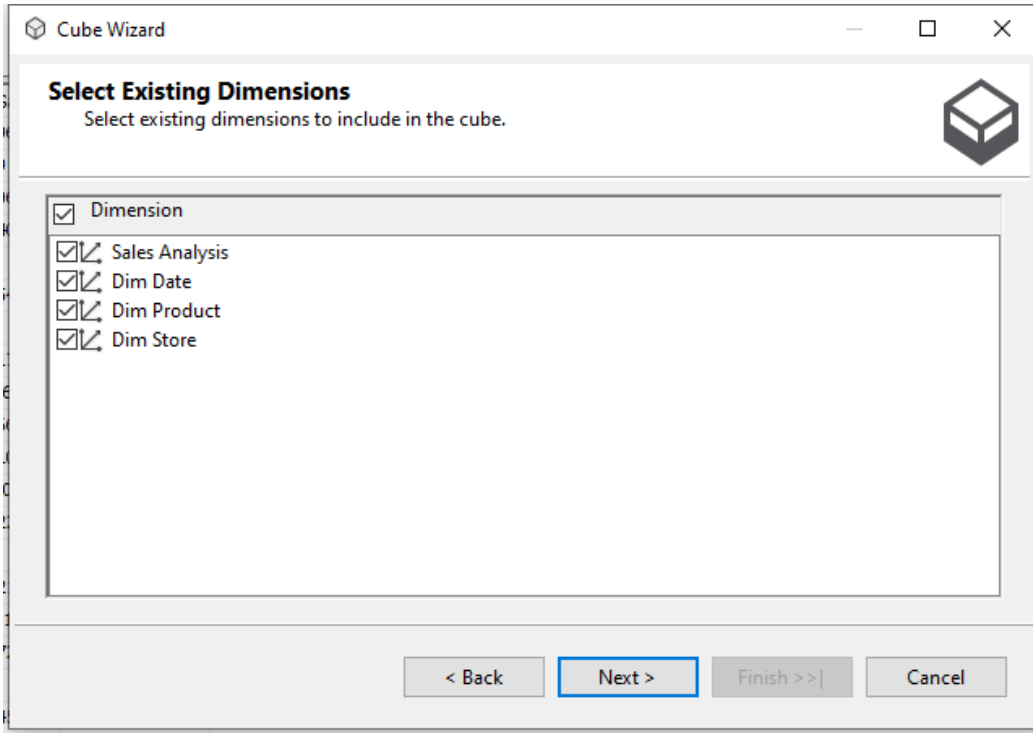


Fig: Selecting the Dimensions

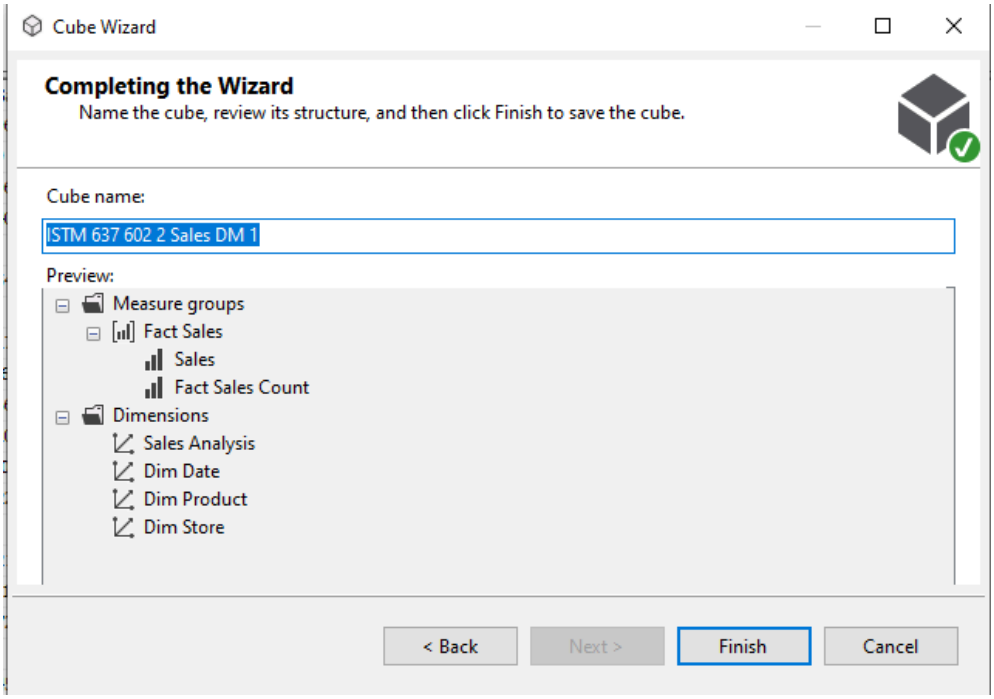


Fig: Cube Preview

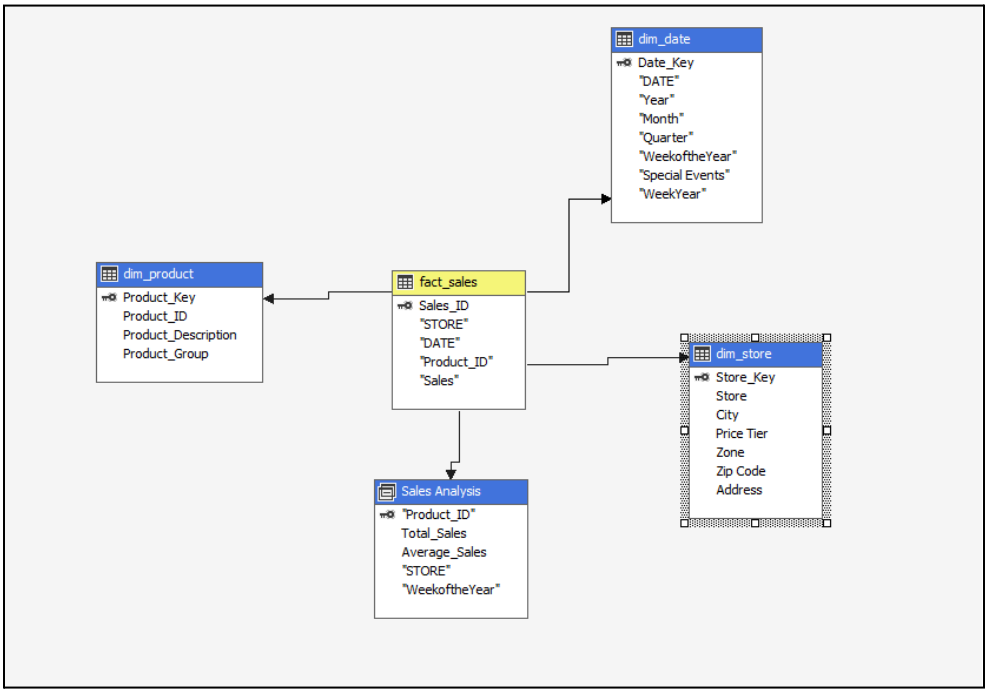


Fig: Cube Schema View

Deployment Configuration

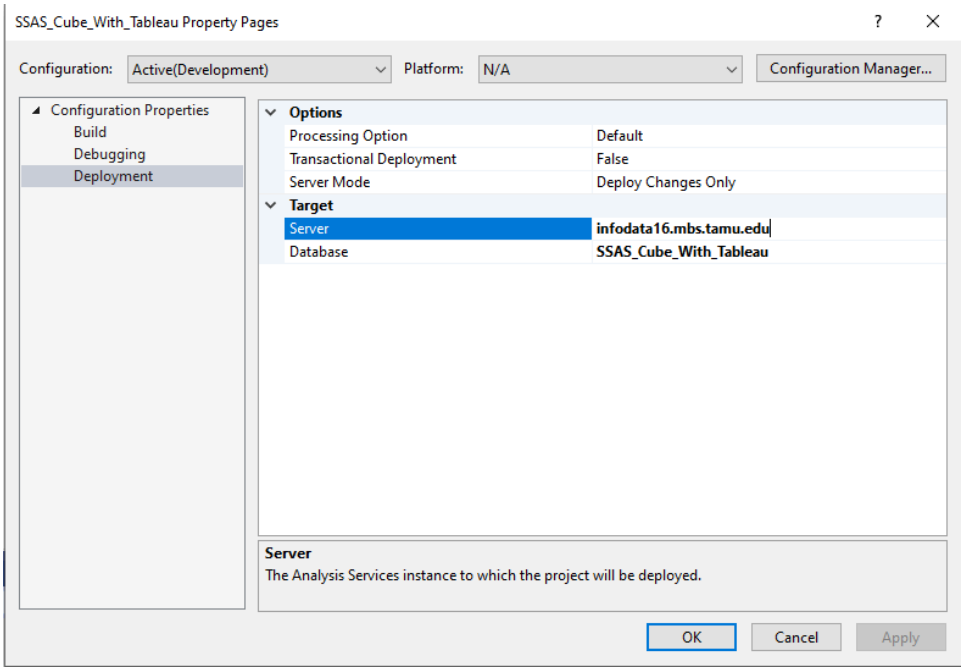


Fig: Server Name Configuration

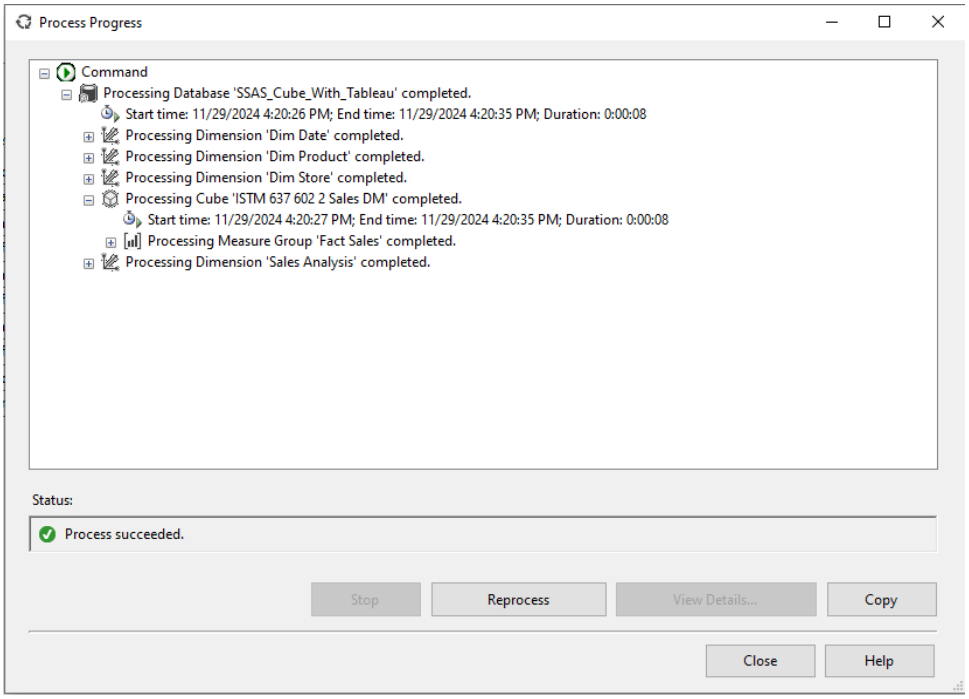


Fig: Successfully Processed and Deployed

Hierarchies Creation:

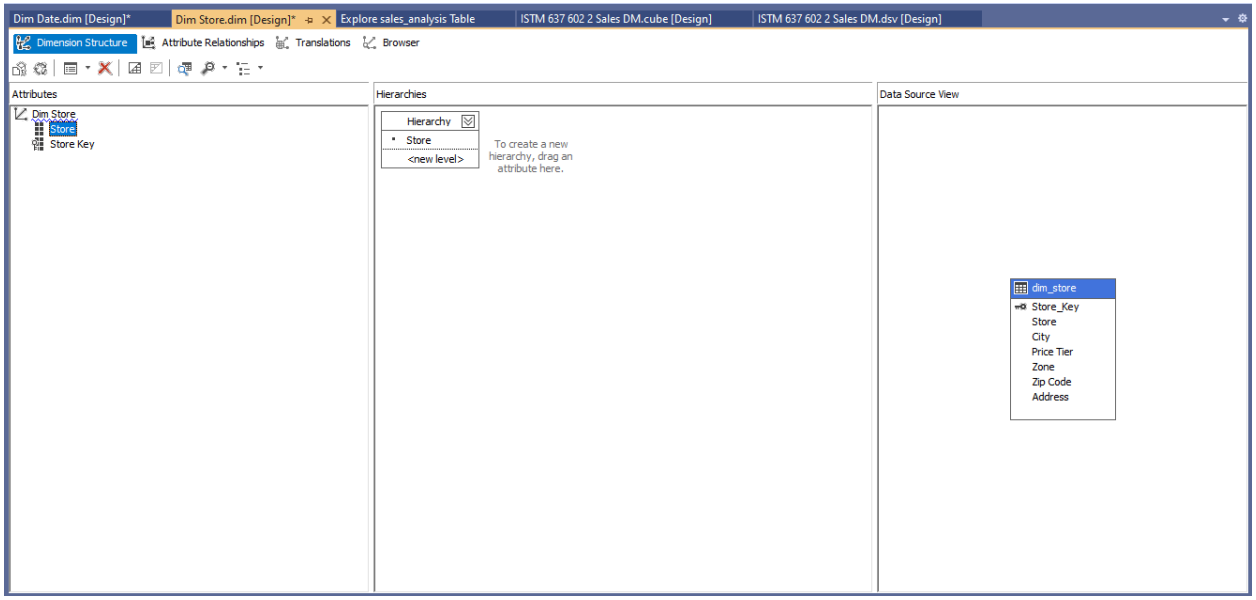


Fig: Store Hierarchy Creation

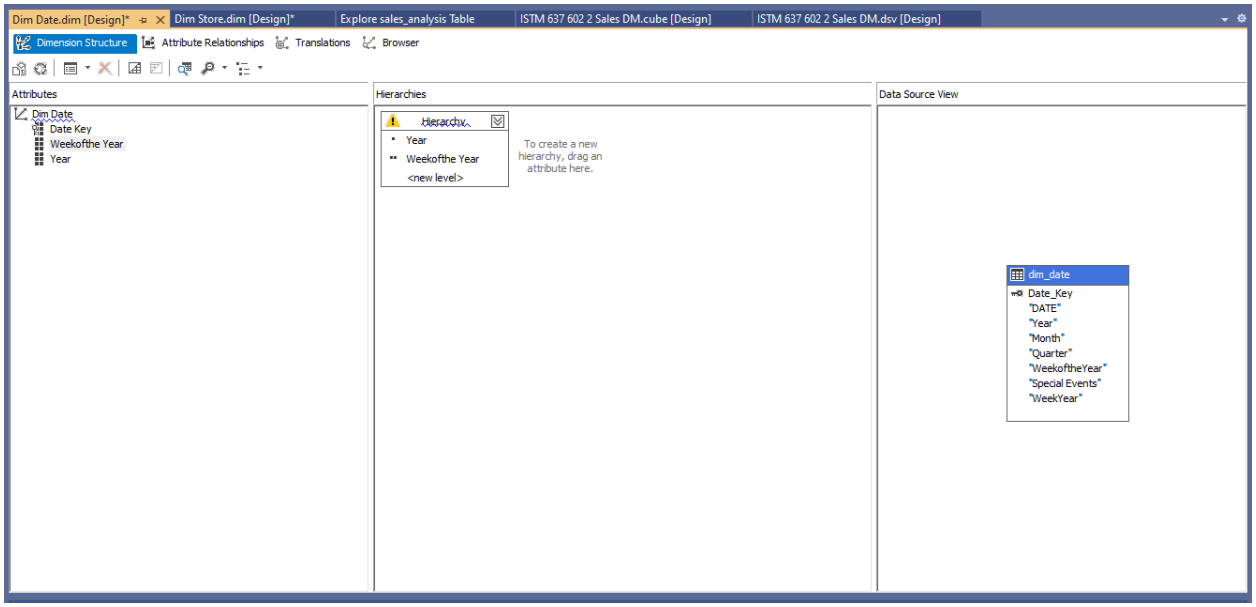


Fig: Date Hierarchy Creation

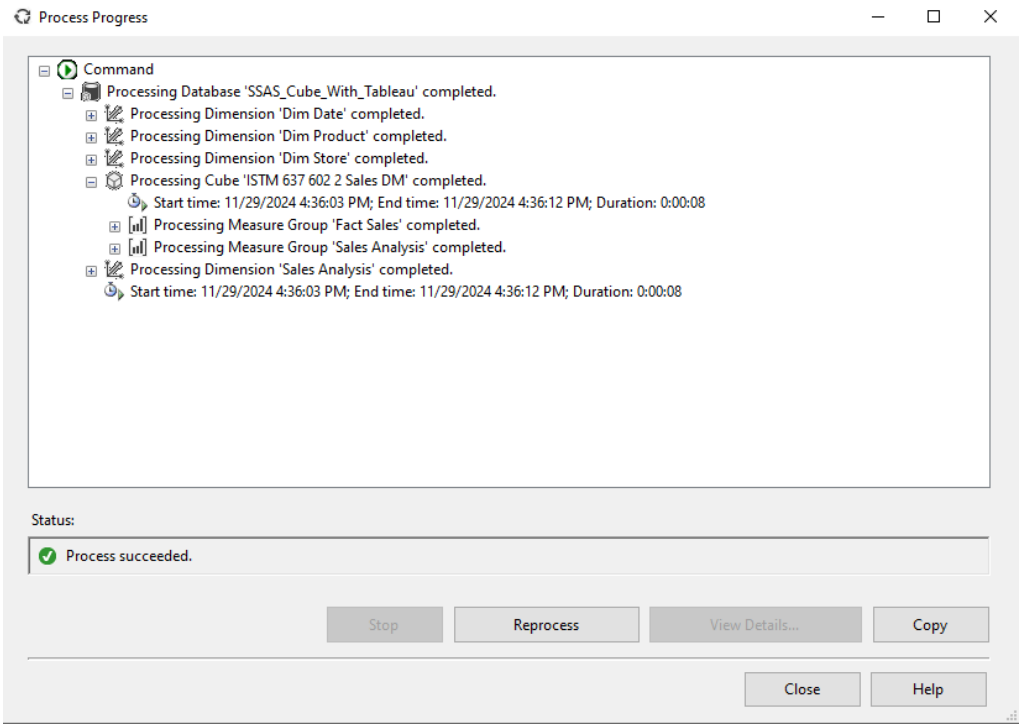


Fig: Successfully Processed and Deployed

Final Report Browsing

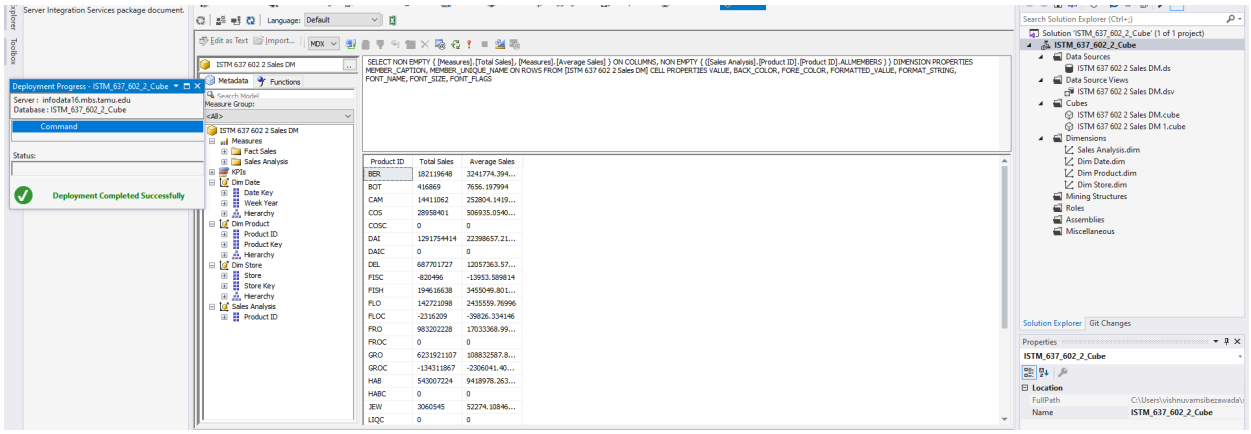


Fig: Report Created Successfully

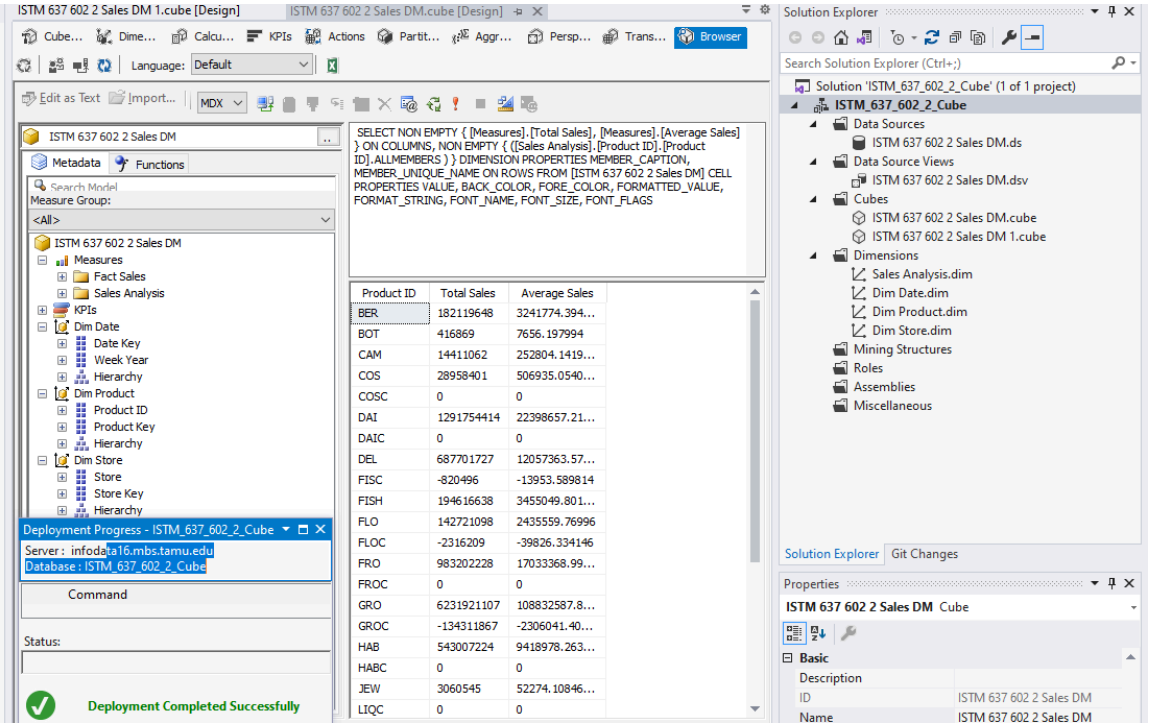


Fig: Better View of Reports with Results, Hierarchy and Tables

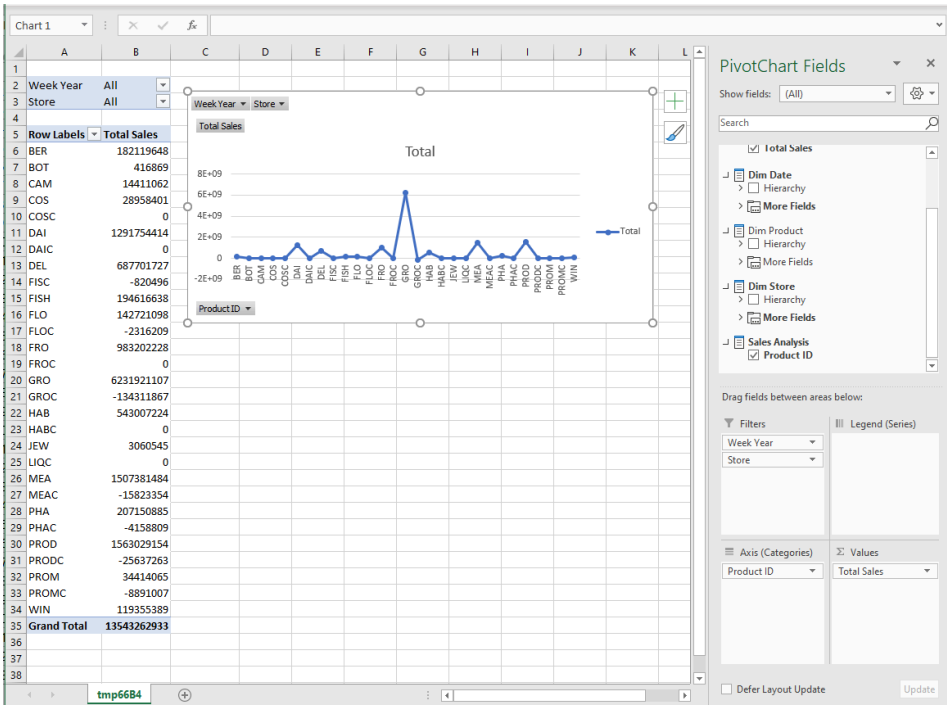


Fig: Exported the data to the Excel for Data Visualization

Note: The above graph consists of total sales of all the categories for all stores and all years

Data Visualization in SSMS

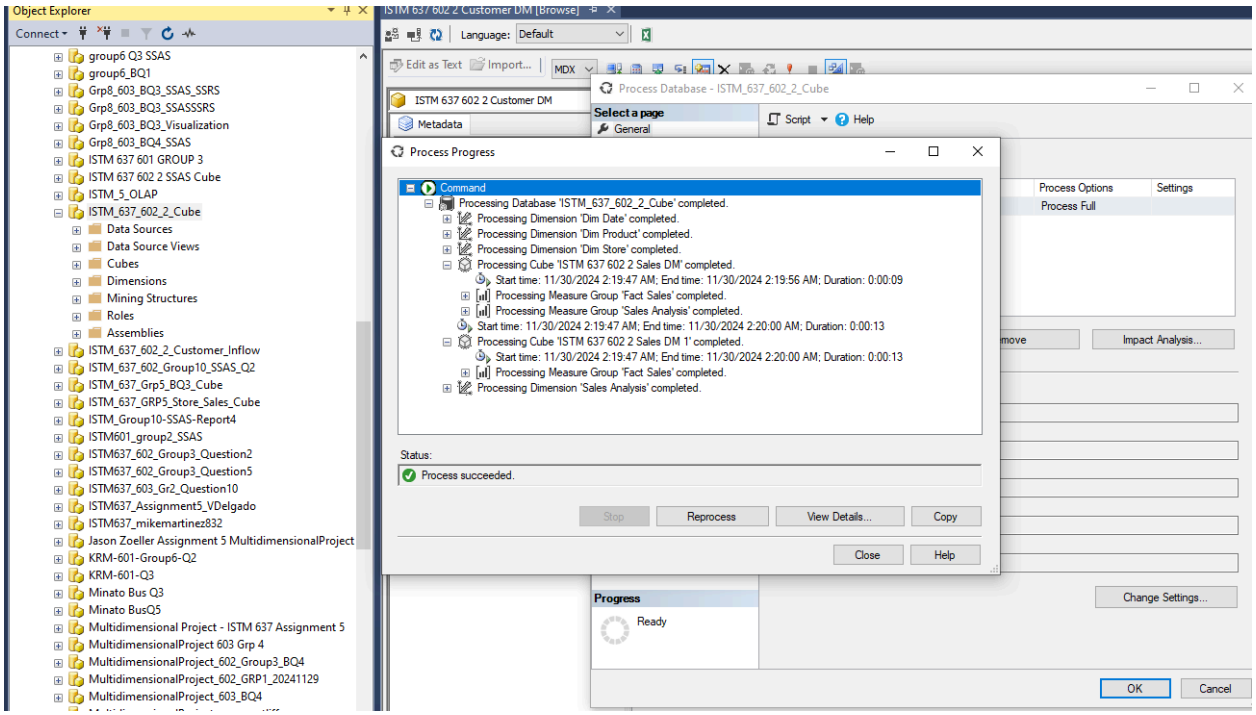


Fig: Successfully Processed the Cube in SSMS

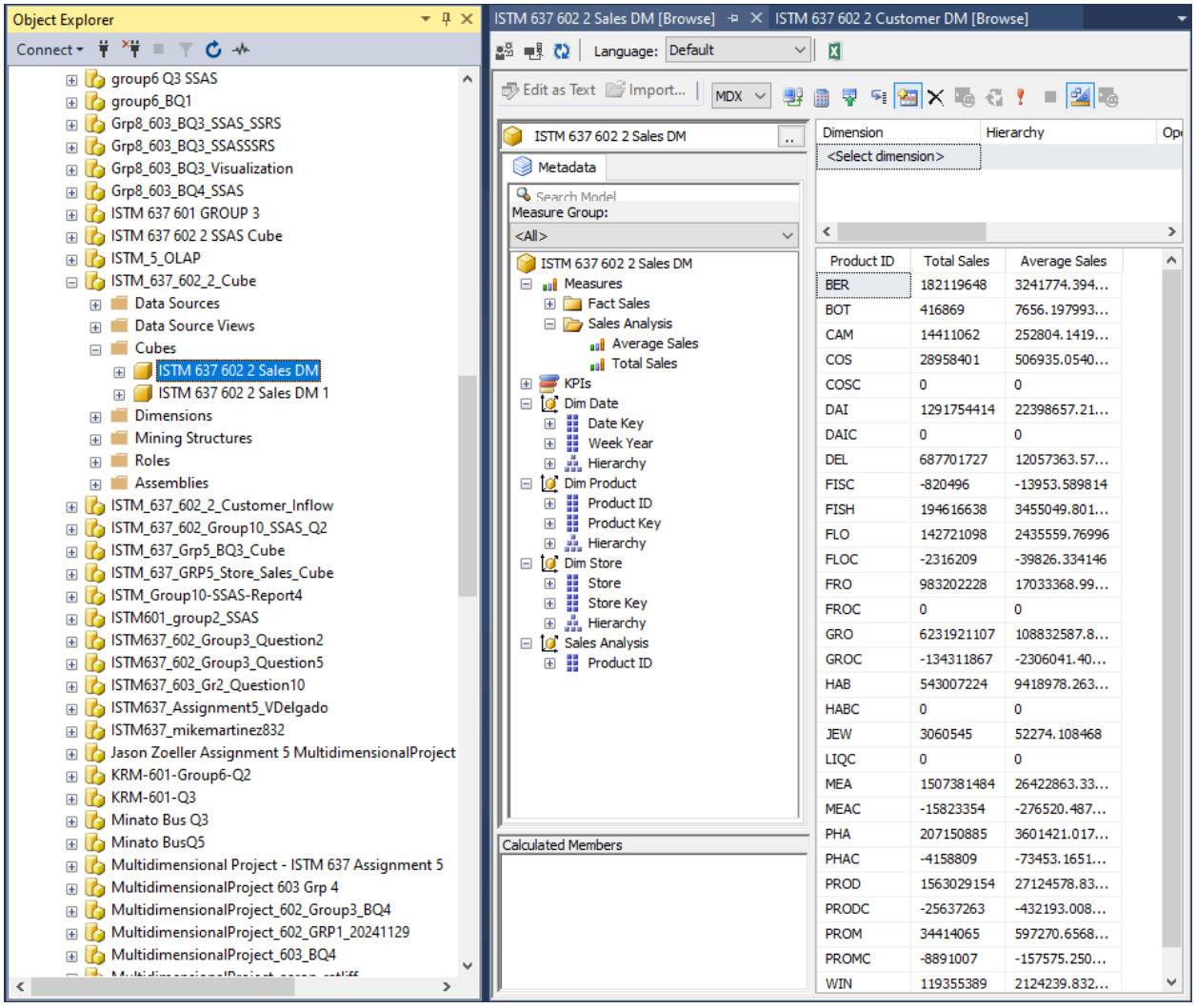


Fig: Better View of Reports with Results, Hierarchy and Tables in SSMS

SSAS Cube with Tableau

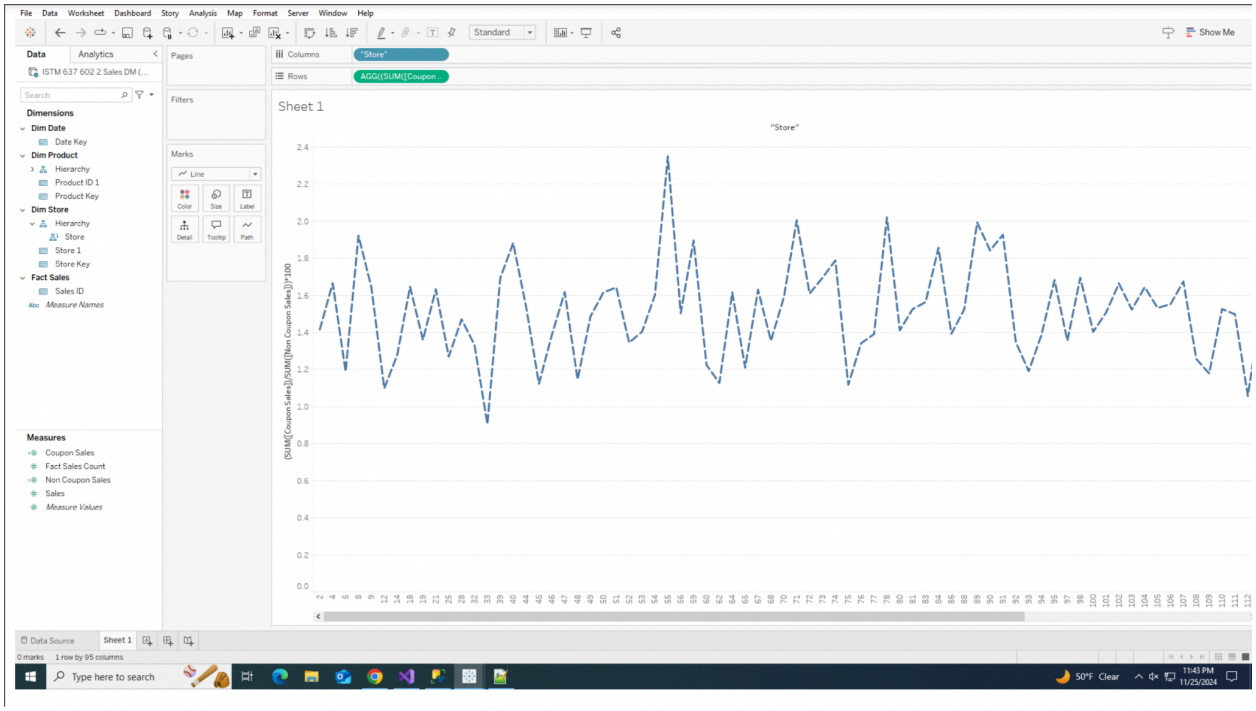


Fig: Final Report Implemented with Tableau to answer the Business Question

Question 2 : How do food and non-food product sales compare across all stores, segmented by year and week number?

Chosen Reporting System: Tableau

Justification: Tableau's dynamic dashboard features make it ideal for comparing sales data segmented by year and week/month. Its interactive filters allow stakeholders to drill down into specific stores or products, making it suitable for presenting comparisons clearly and concisely.

Report Template:

- Header: Report Title
- Filters: Year, Month, Product Category
- Visualizations: A grouped bar chart showing sales of food and non-food items for each store
- Summary: Total sales for food vs. non-food items across all stores

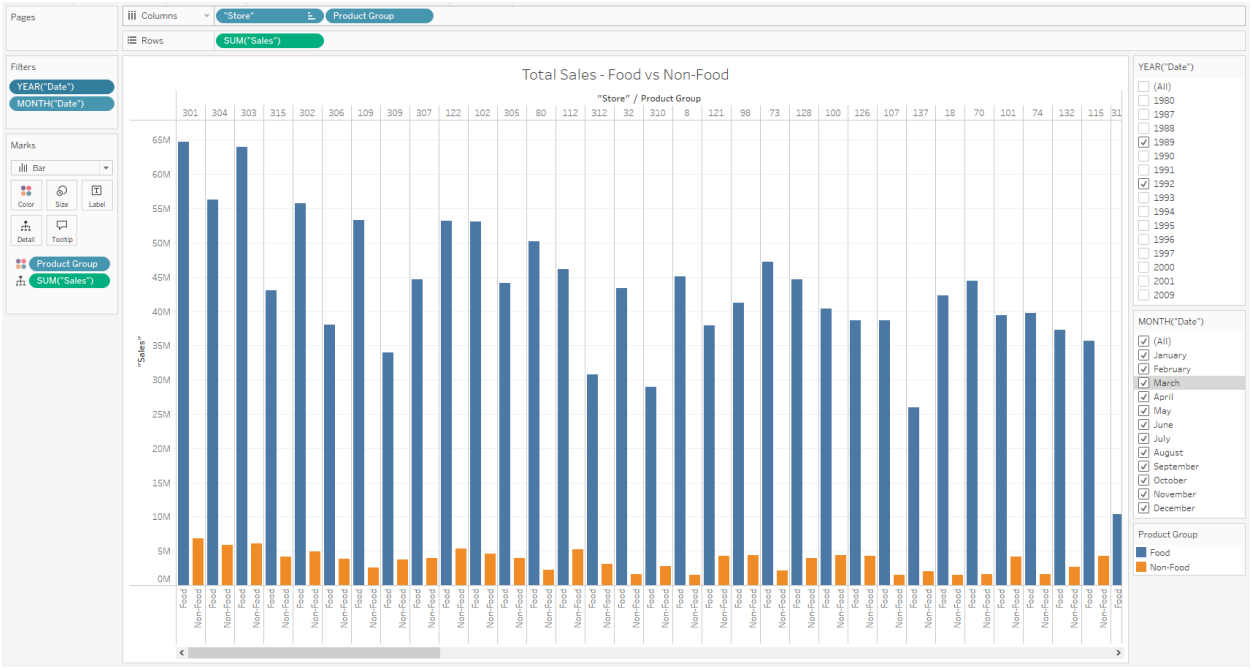


Fig: Total Sales of Food Vs Non Food filtered by Year and Store

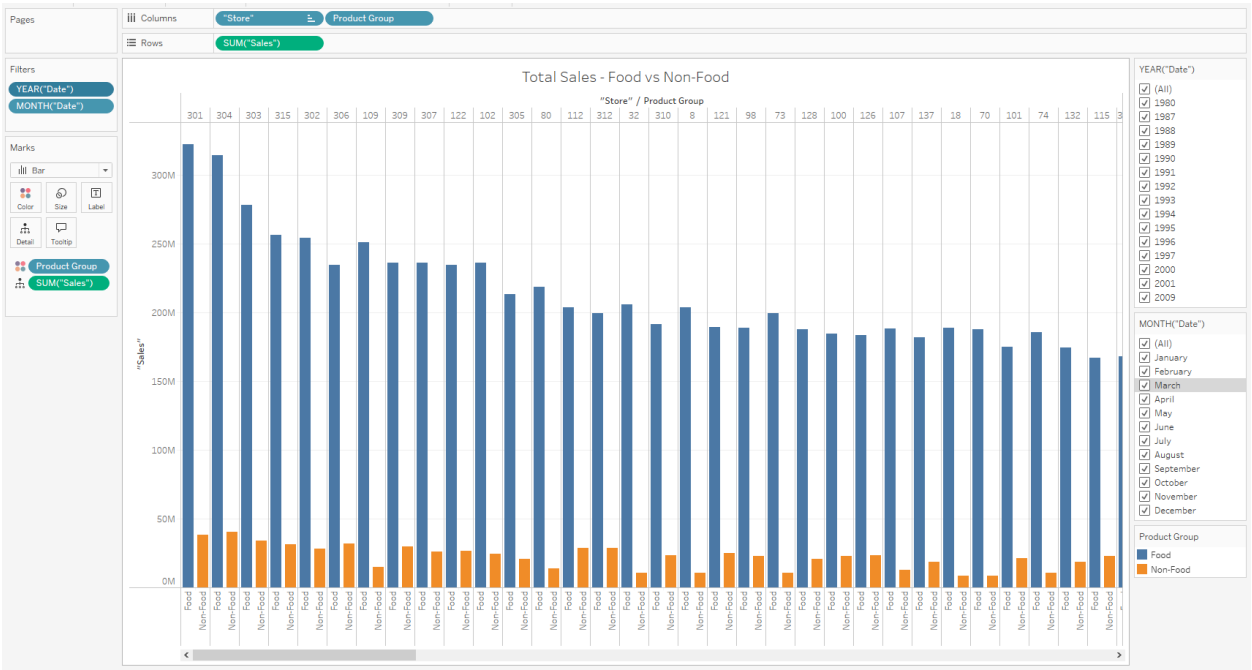


Fig: Total Sales of Food Vs Non Food

Question 3 : What are the top-selling departments (e.g., Bakery, Meat, Deli) across stores, and how do they vary by store and week?

Chosen Reporting System: SSRS

Justification: SSRS is well-suited for generating detailed, paginated reports that list the top-selling departments and their variations by store and week. It supports tabular layouts that allow for detailed breakdowns while offering robust parameter filtering.

Report Template:

- Header: Report Title
- Filters: Store, Product Category, and Week Range.
- Details: A tabular report listing sales by Product Category ranked by performance.
- Visualizations: Optional bar chart visualizing weekly sales performance by department.
- Summary: Top departments by total sales.

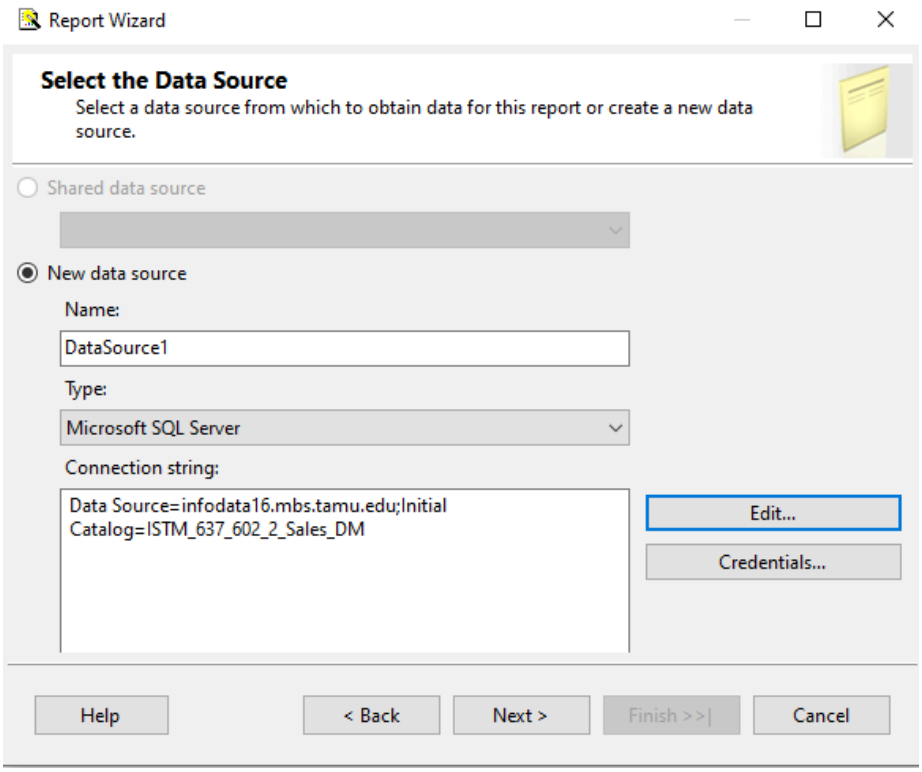


Fig: Creating the Data Source

Creating Report without filters

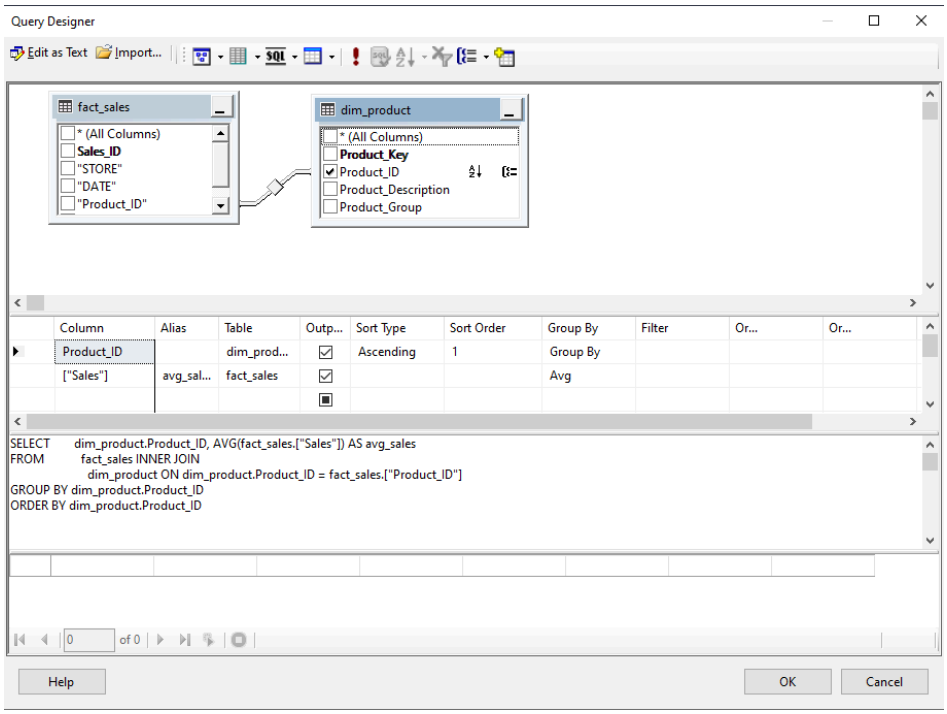


Fig: Using Query Designer to fetch the data

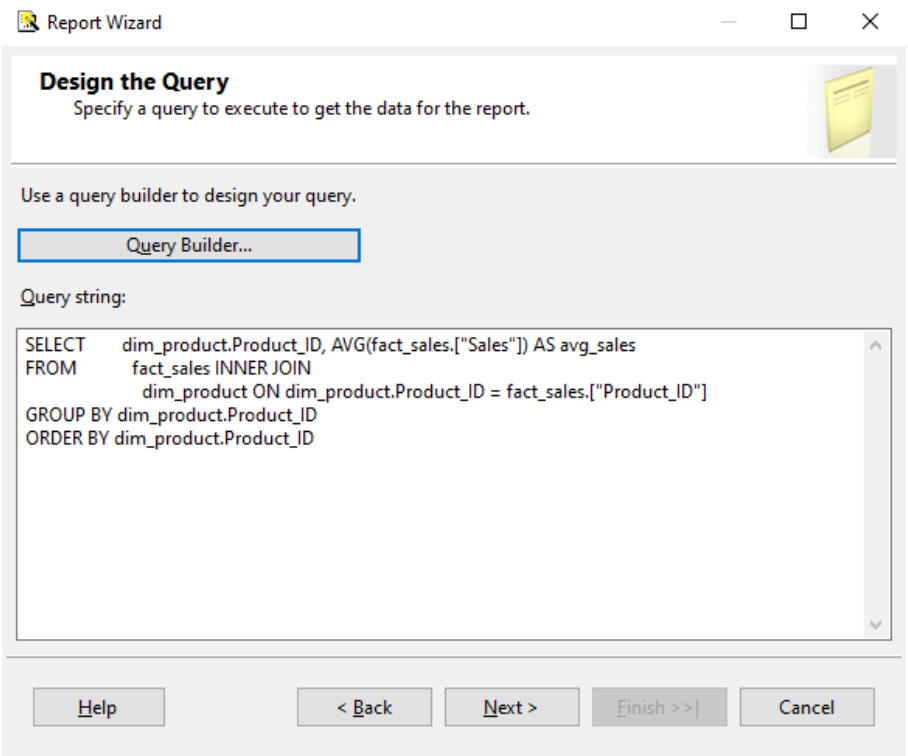


Fig: Query to extract the data

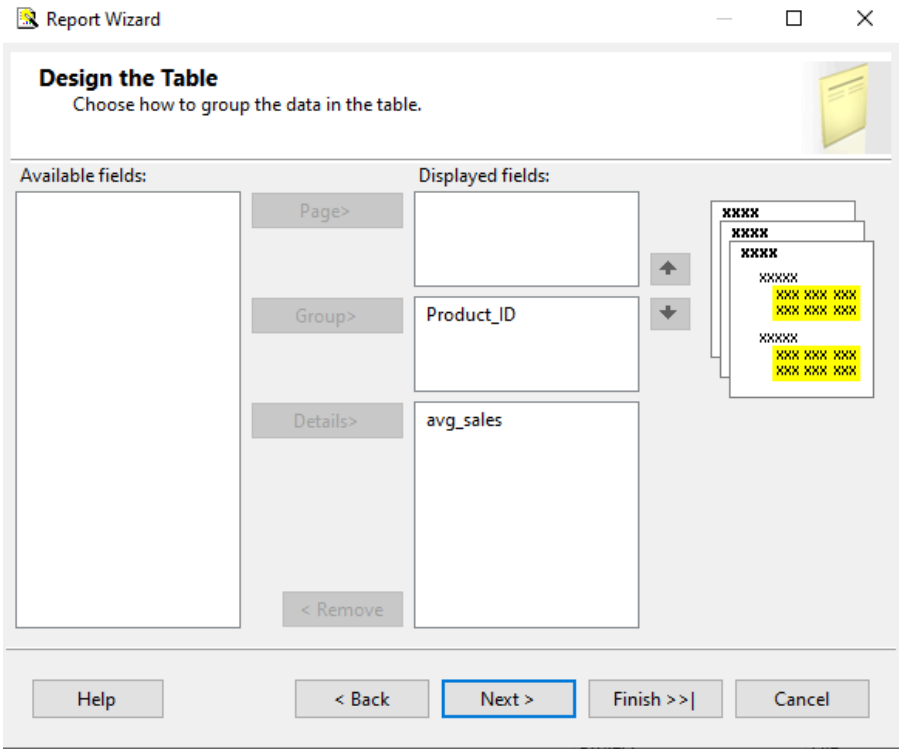


Fig: Adding the necessary fields to structure the report

Results

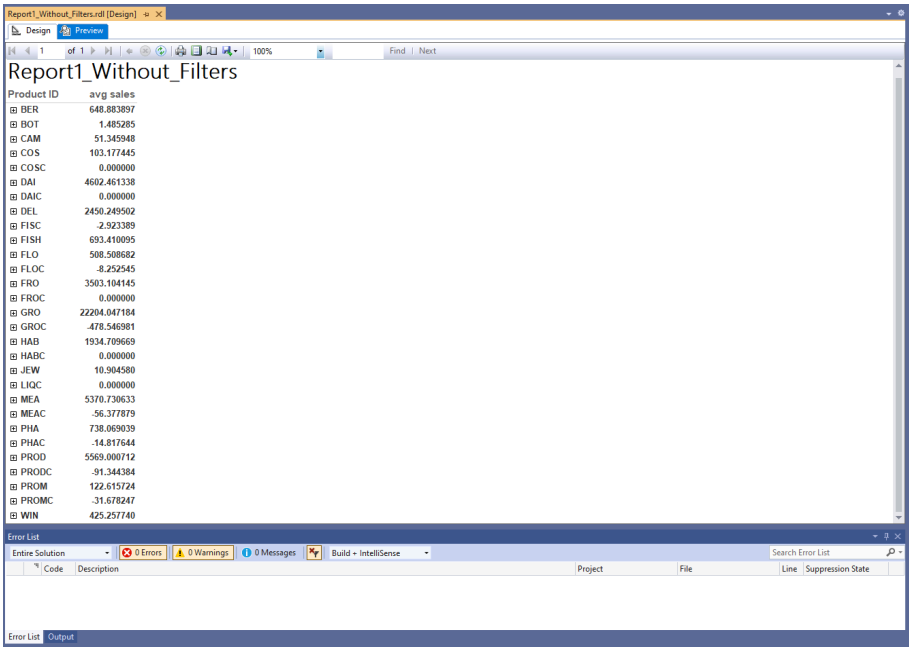


Fig: Report without any Filters

Deployment of Report

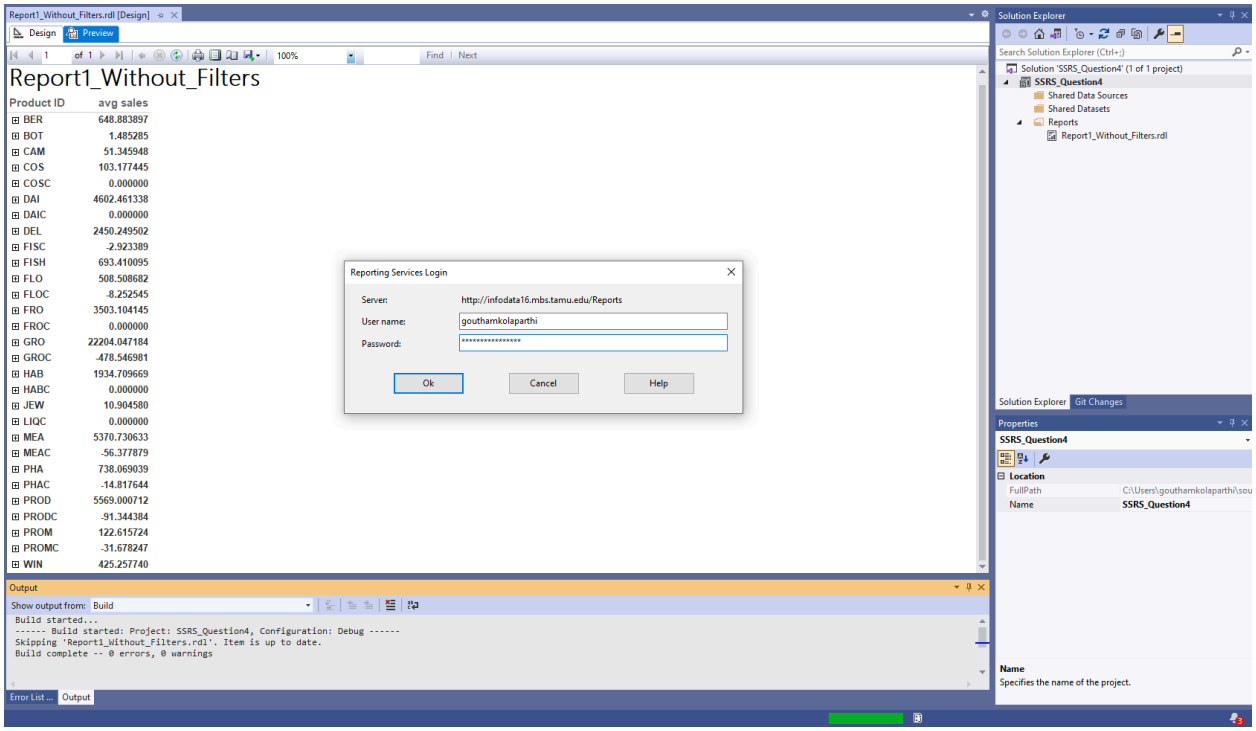


Fig: Deploying the report into the server

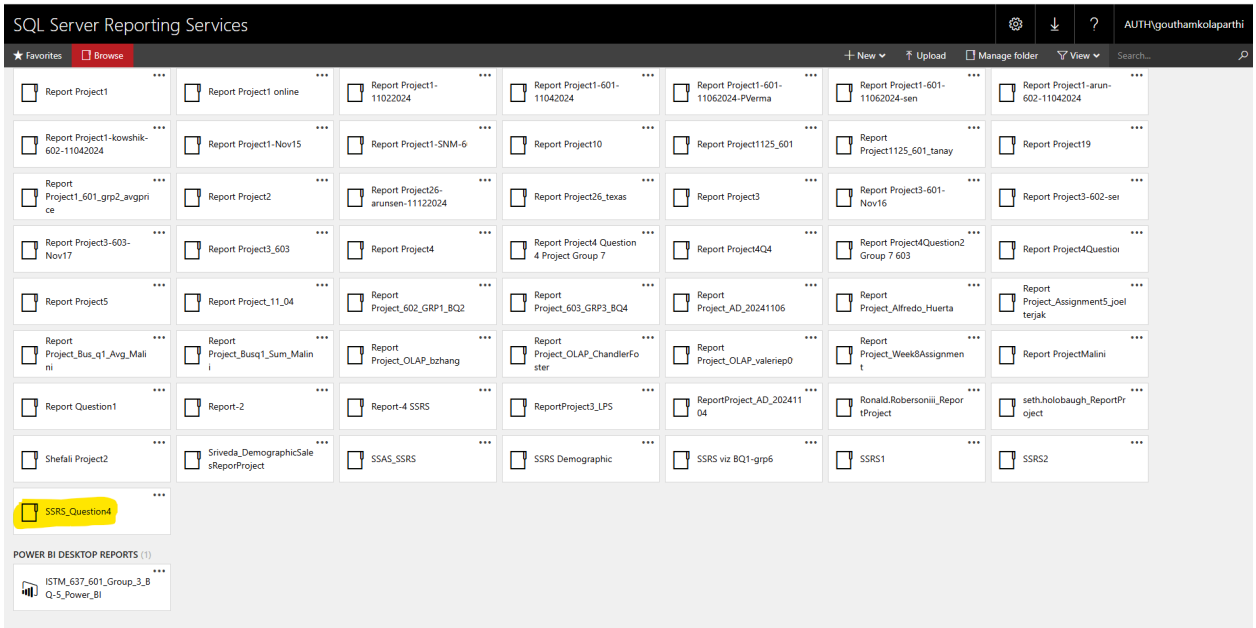


Fig: Deployed Report Folder (SSRS Question 4) Location in server

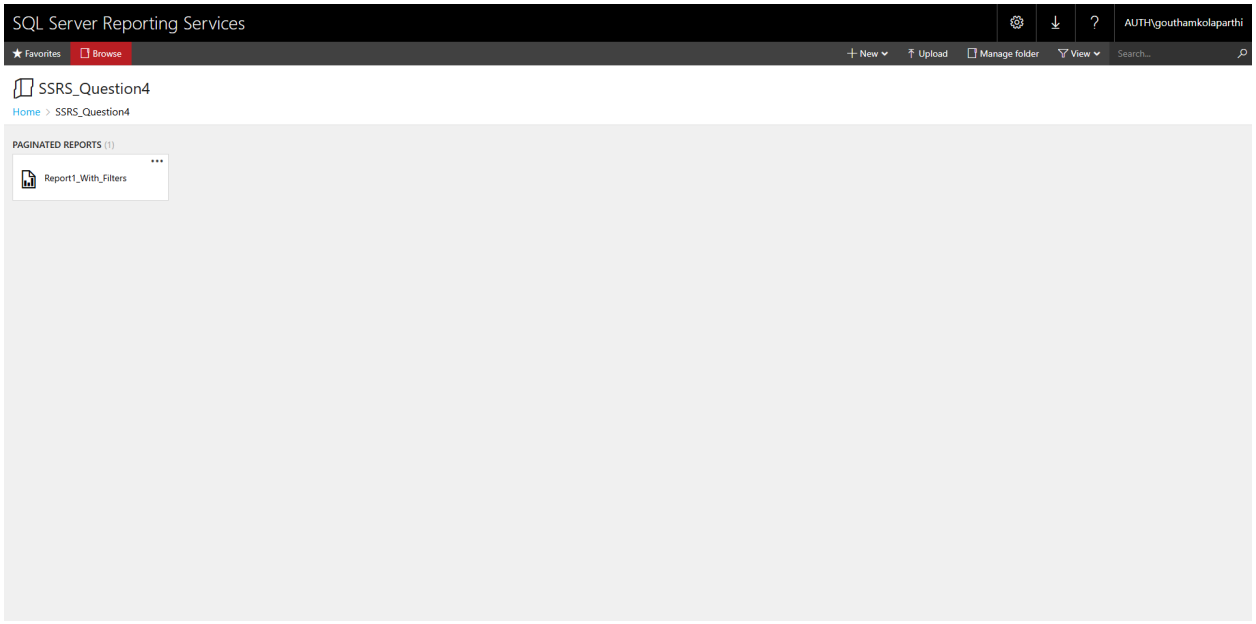


Fig: Deployed Report Location in Folder SSRS Question 4

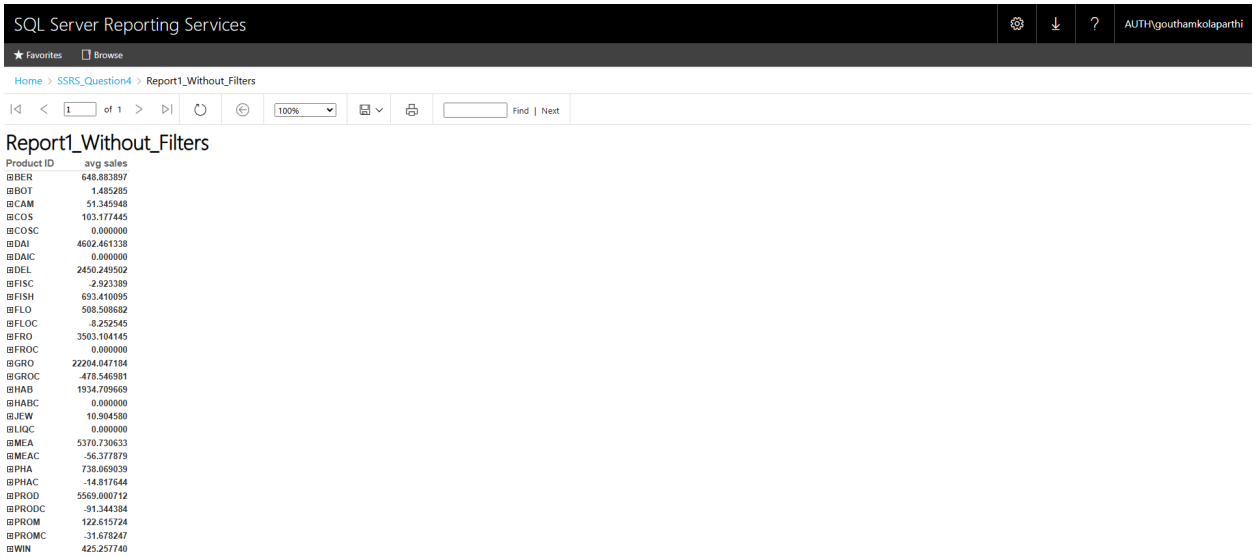


Fig: Data Sample Viewing in the server

Creating Report with filters : Store as filter

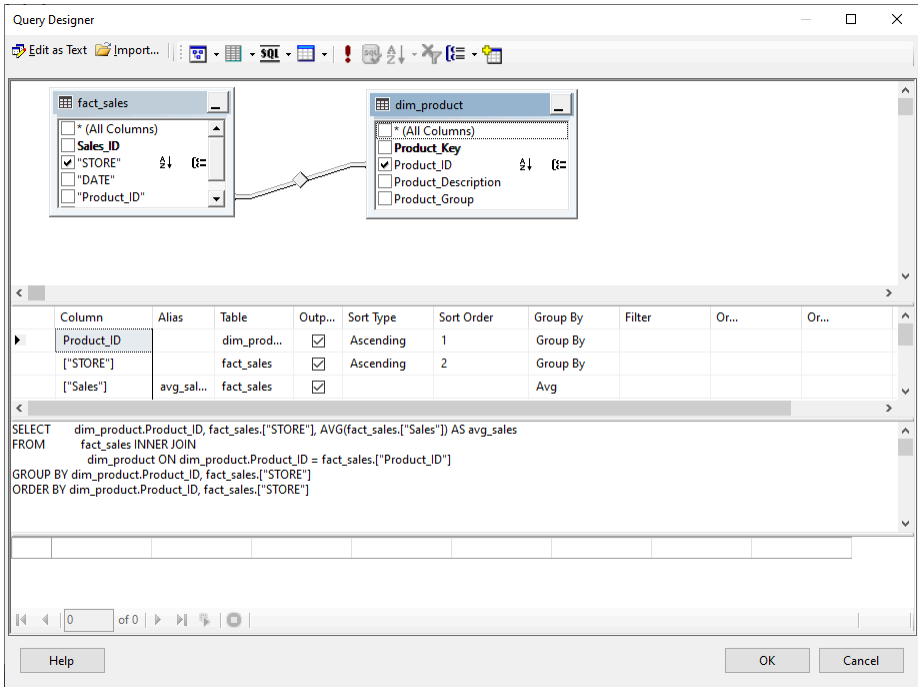


Fig: Using Query Designer to fetch the data

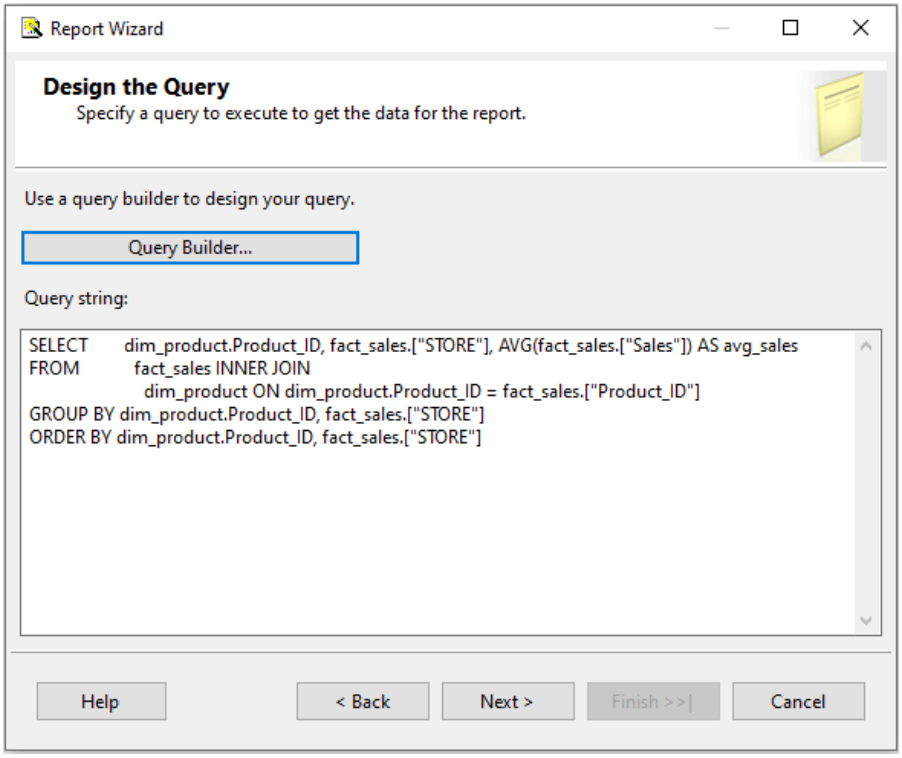


Fig: Query to extract the data

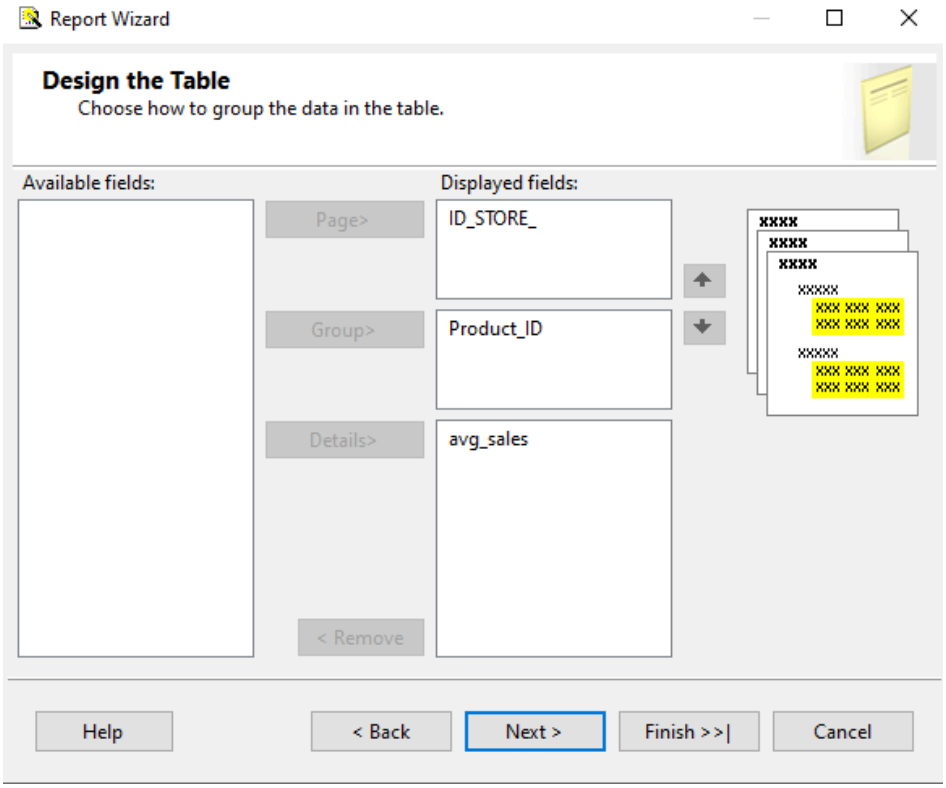


Fig: Adding the necessary fields to structure the report

Results

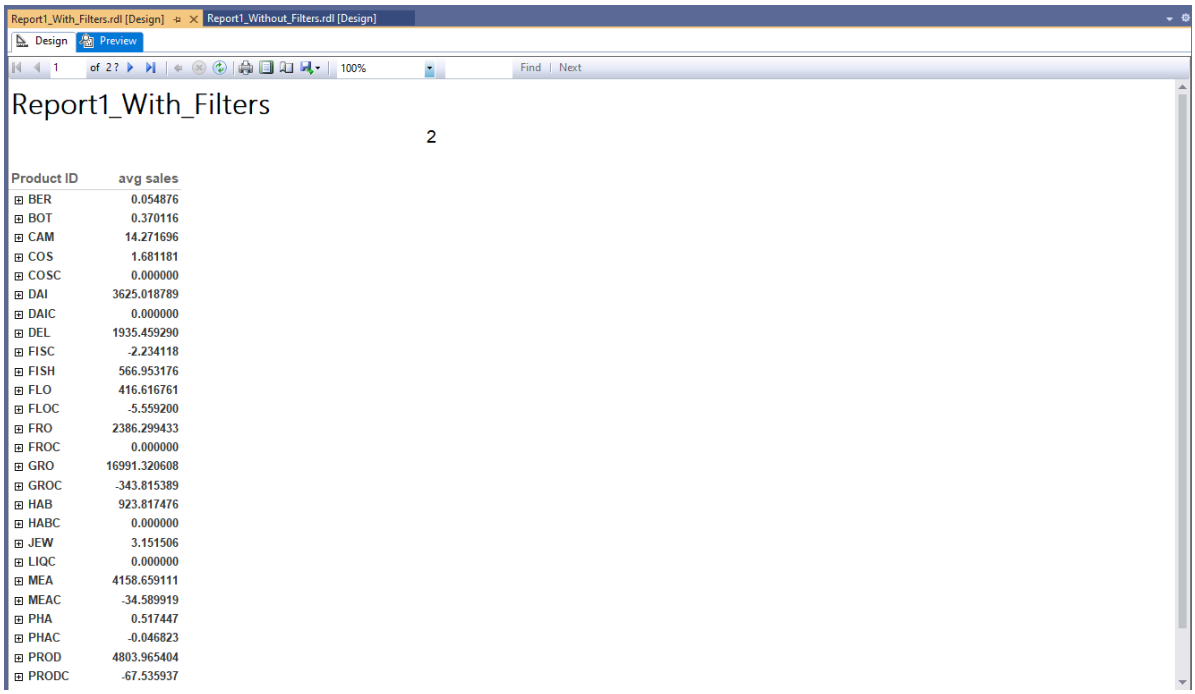


Fig: Report with store Filters

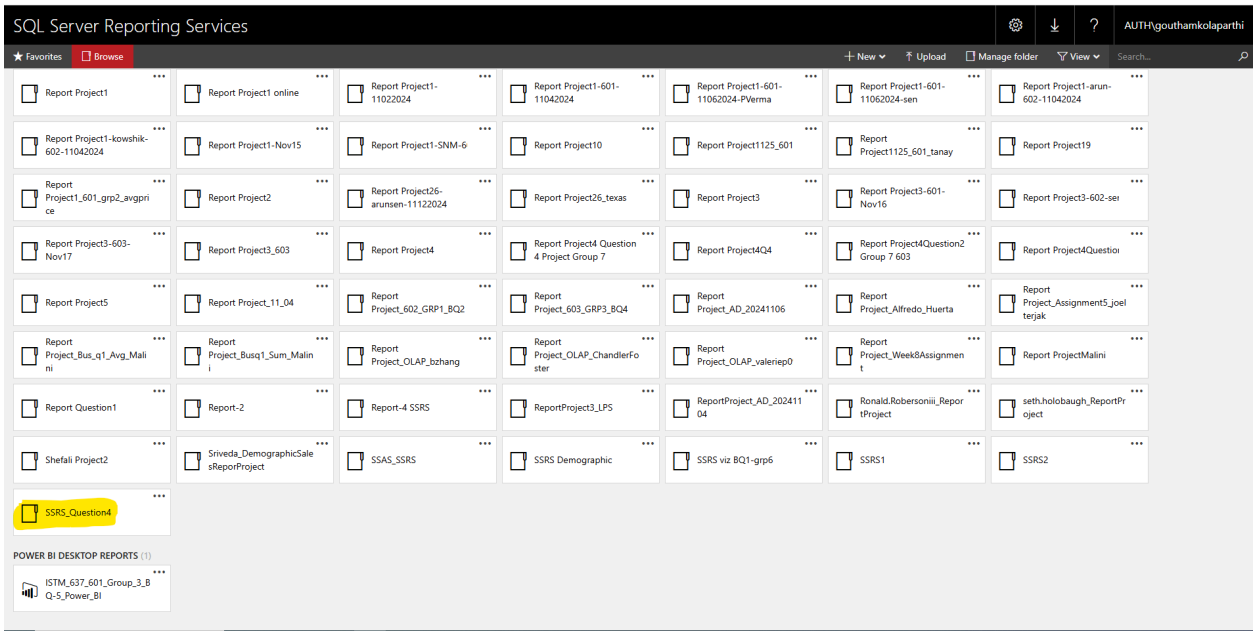


Fig: Deployed Report Folder (SSRS Question 4) Location in server

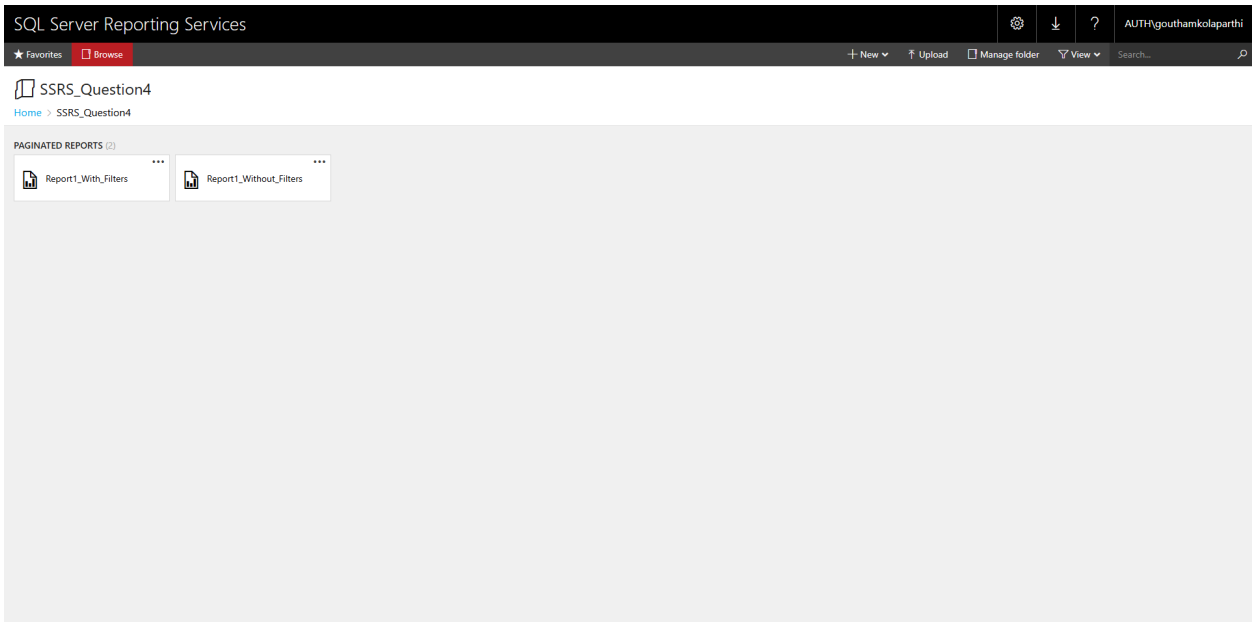
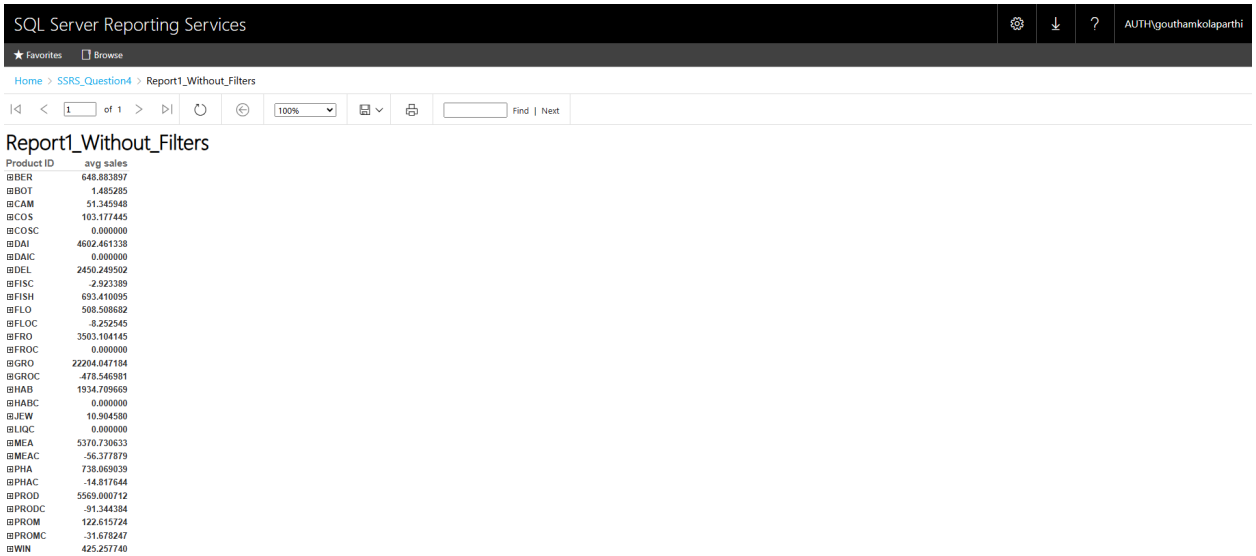


Fig: Deployed Report Location in Folder SSRS Question 4



Report1_Without_Filters

Product ID	avg sales
BER	648.883897
BOT	1.485285
CAM	51.345948
COS	103.177445
COSC	0.000000
DAI	4602.461338
DAIC	0.000000
DEL	2450.249502
FISC	-2.923389
FISH	693.410095
FLO	508.508682
FLOC	-8.252545
FRO	3503.104145
FROC	0.000000
GRO	22204.047184
GROC	-478.546981
HAB	1934.709669
HABC	0.000000
JEW	10.904580
LIQC	0.000000
MEA	5370.730633
MEAC	-56.377879
PHA	738.069039
PHAC	-14.817644
PROD	5569.000712
PRODC	-91.344384
PROM	122.615724
PROMC	-31.678247
WIN	425.257740

Fig: Data Sample Viewing in the server

Row Labels	Average of avg sales
	1708.1701
BER	648.883897
BOT	1.485285
CAM	51.345948
COS	103.177445
COSC	0
DAI	4602.461338
DAIC	0
DEL	2450.249502
FISC	-2.923389
FISH	693.410095
FLO	508.508682
FLOC	-8.252545
FRO	3503.104145
FROC	0
GRO	22204.04718
GROC	-478.546981
HAB	1934.709669
HABC	0
JEW	10.90458
LIQC	0
MEA	5370.730633
MEAC	-56.377879
PHA	738.069039
PHAC	-14.817644
PROD	5569.000712
PRODC	-91.344384
PROM	122.615724
PROMC	-31.678247
WIN	425.25774
Grand Total	1685.662866

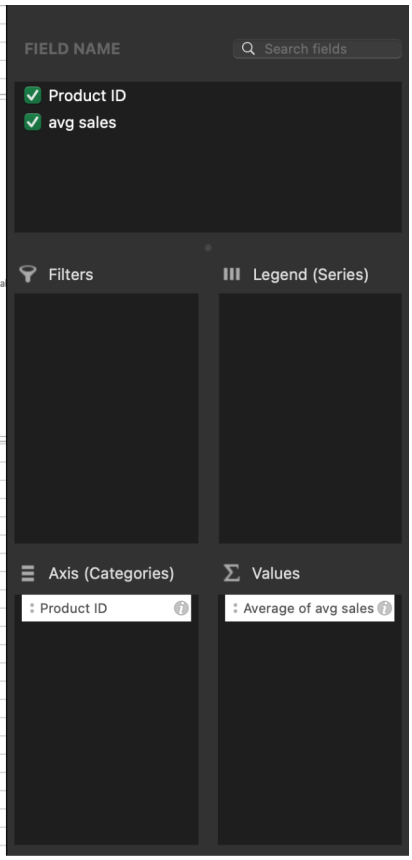


Fig: Exported the data to the Excel for Data Visualization

Question 4 : Determine the sales of a given category during the weeks of special events.

Chosen Reporting System: Tableau

Justification: Tableau's flexibility in handling event-based data through custom filtering options makes it an excellent choice for analyzing sales during specific weeks. Interactive dashboards allow stakeholders to explore event-related sales trends with ease.

Report Template:

- Header: Report Title.
- Filters: Event Type, Week, and Product Category.
- Visualizations: Grouped Bar chart to track sales over event weeks
- Summary: Total sales during different special events

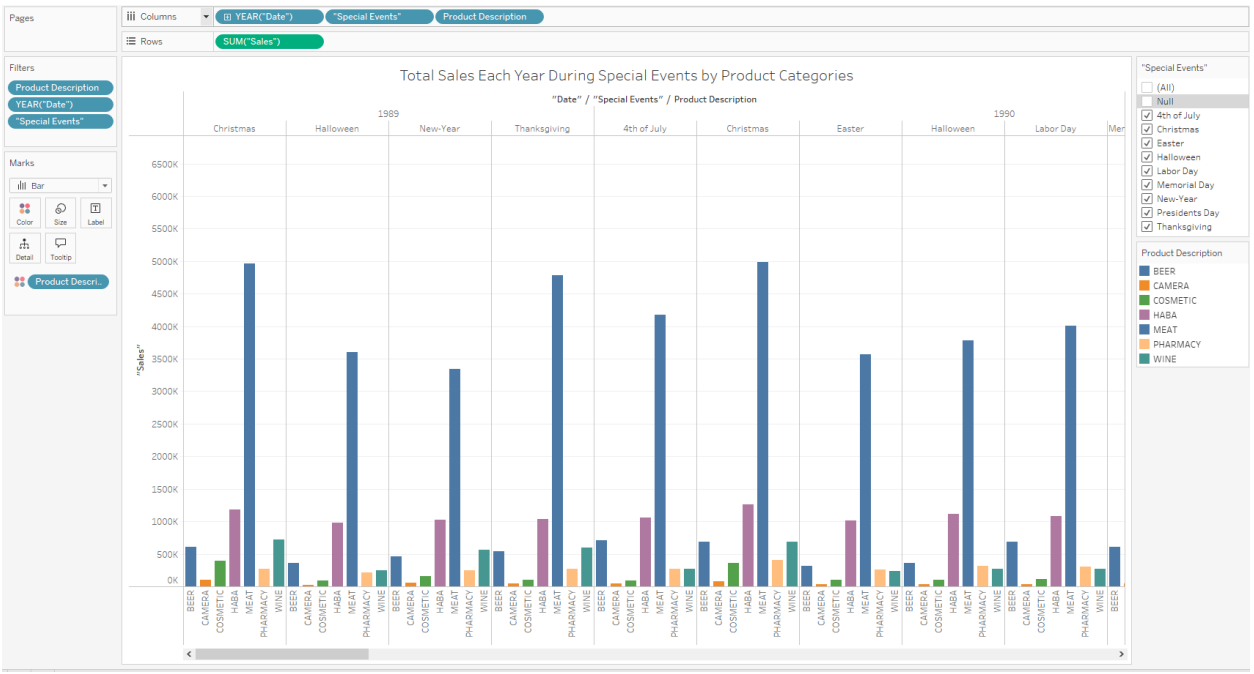


Fig: All Special Events Data

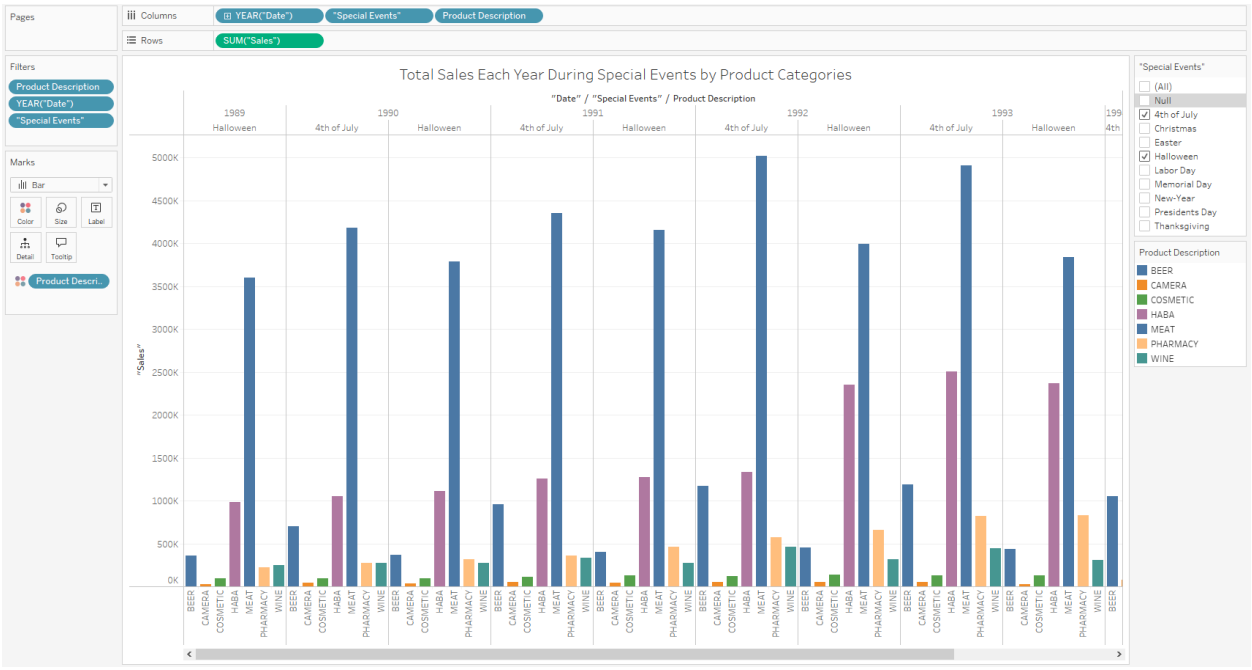


Fig: Filtered Special Events (4th of July and Halloween) Data

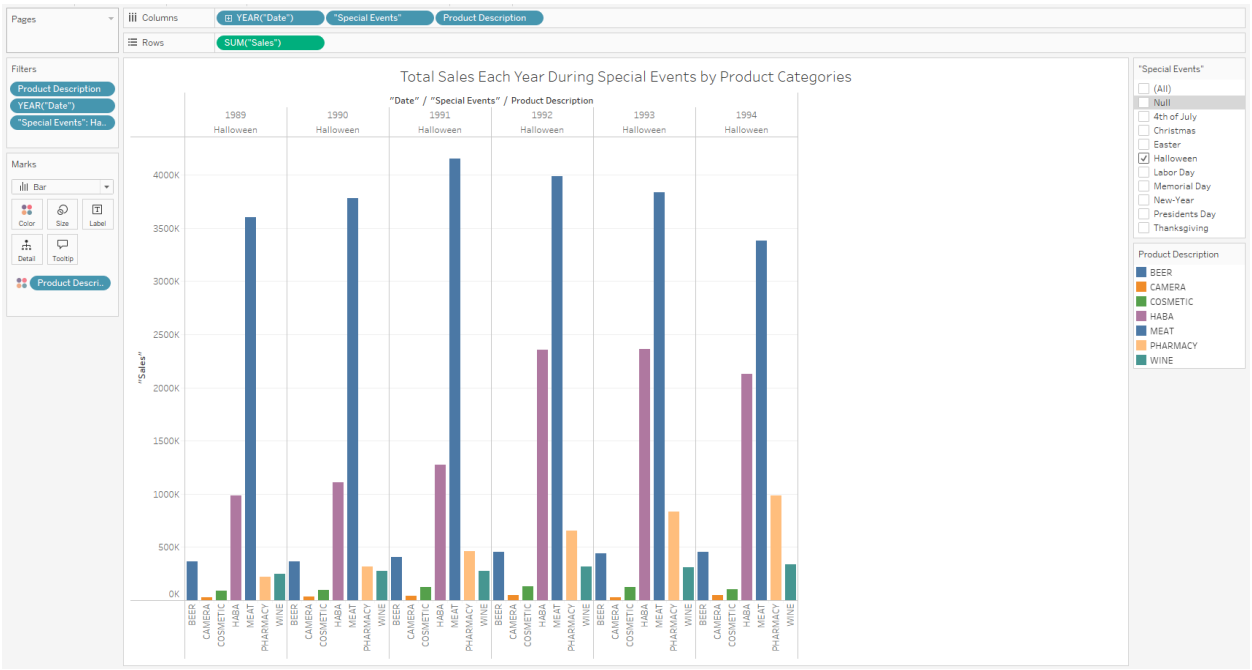


Fig: Filtered Special Events (Halloween) Data

Question 5 : What is the annual customer inflow for each store allowing for insights into customer behavior and store performance?

Justification:

SSAS cubes enable detailed multidimensional analysis of customer inflow data, aggregated by year and store. It is ideal for presenting high-level metrics while allowing for deeper drill-down into customer behavior.

Report Template:

- Header: Report Title specifying annual customer inflow analysis.
- Filters: Year, Store, Week of the Year
- Visualizations: Line chart showing annual trends on inflow distribution across stores.
- Summary: Highlights high-performing stores and customer behavior insights.

Data Source Creation

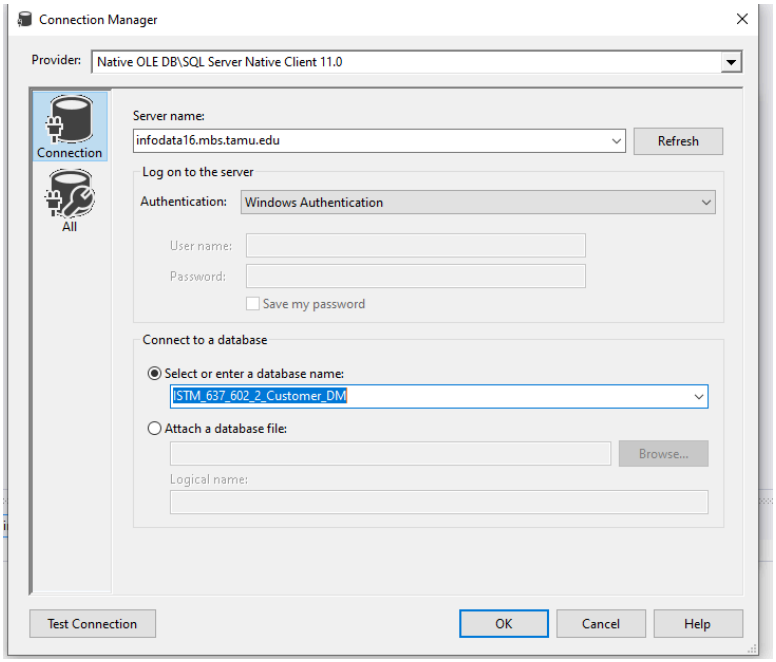


Fig: Data Source Creation

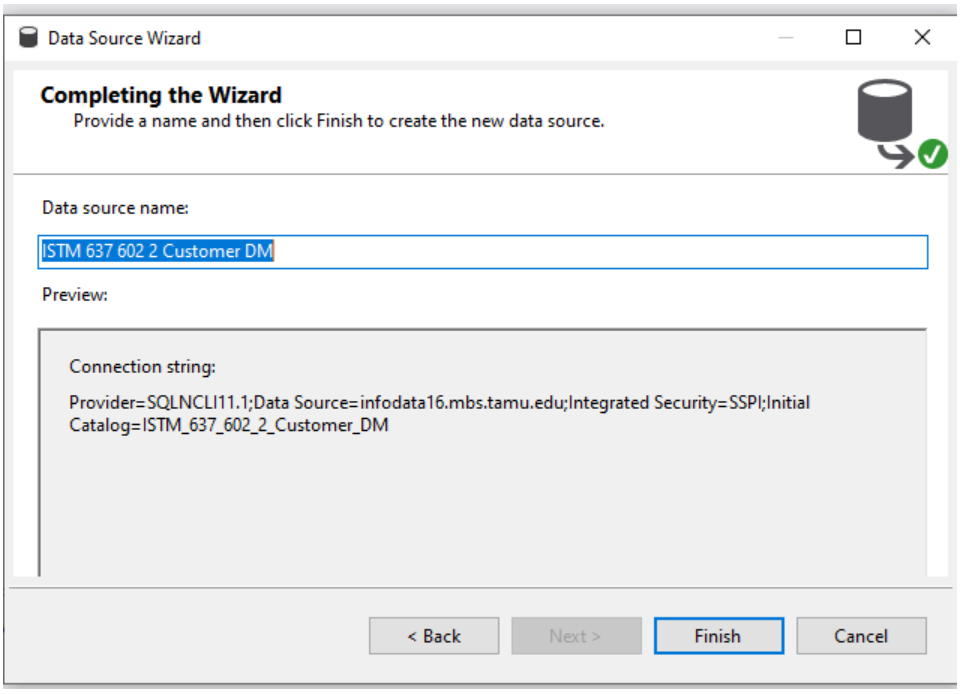


Fig: Connection String

Data Source View Creation

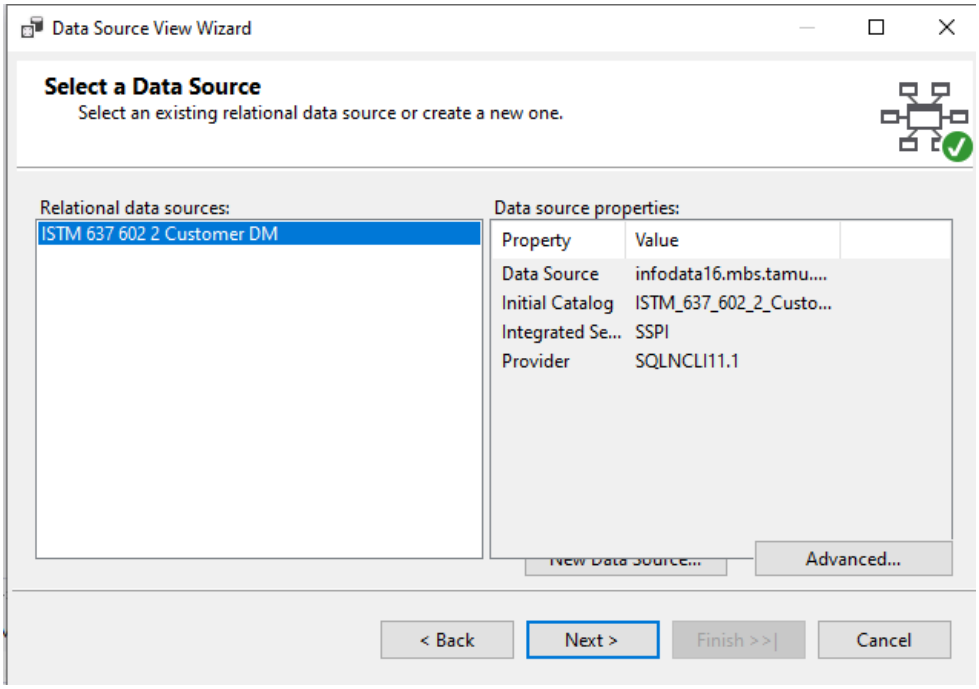


Fig: Data Source View Creation

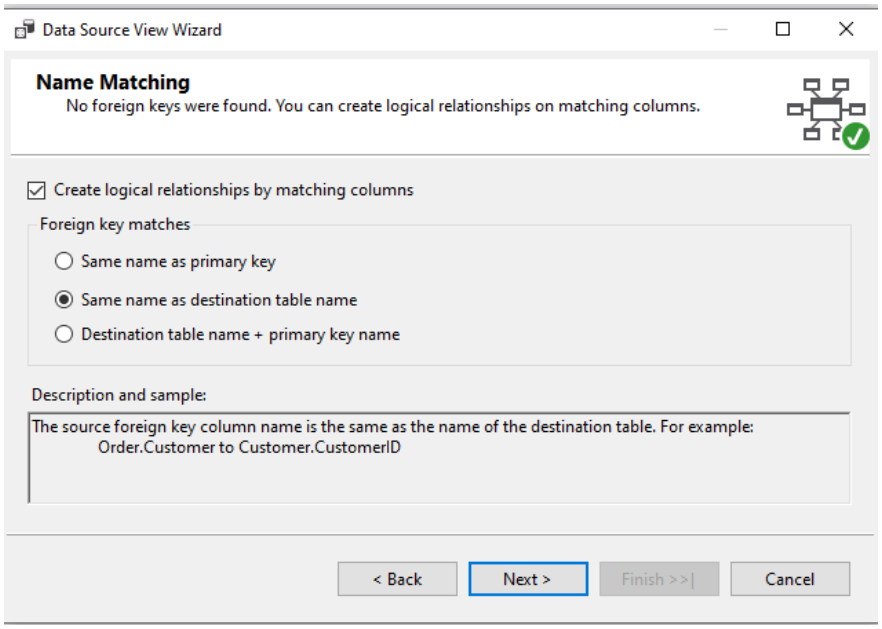


Fig: Relationship Matching between tables

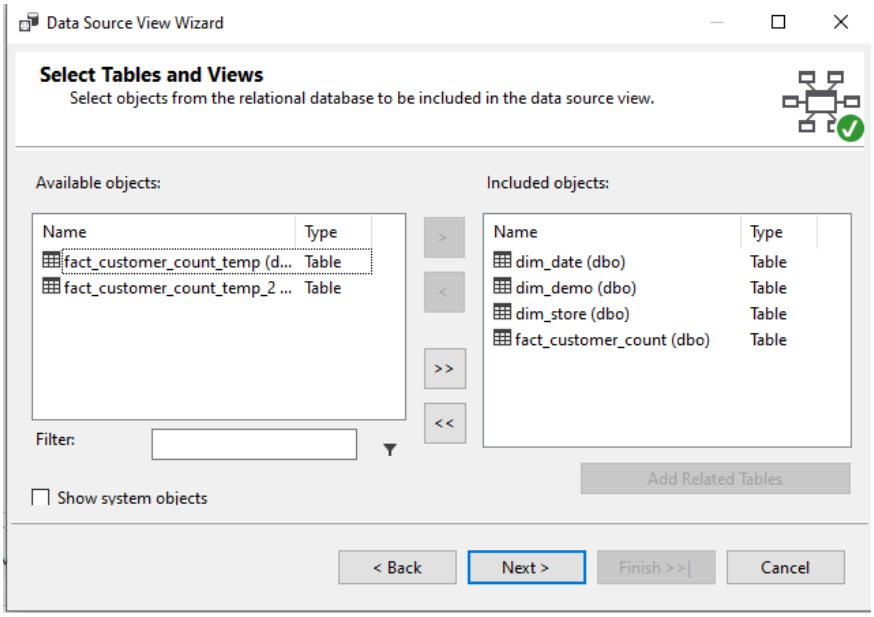


Fig: Table Selections

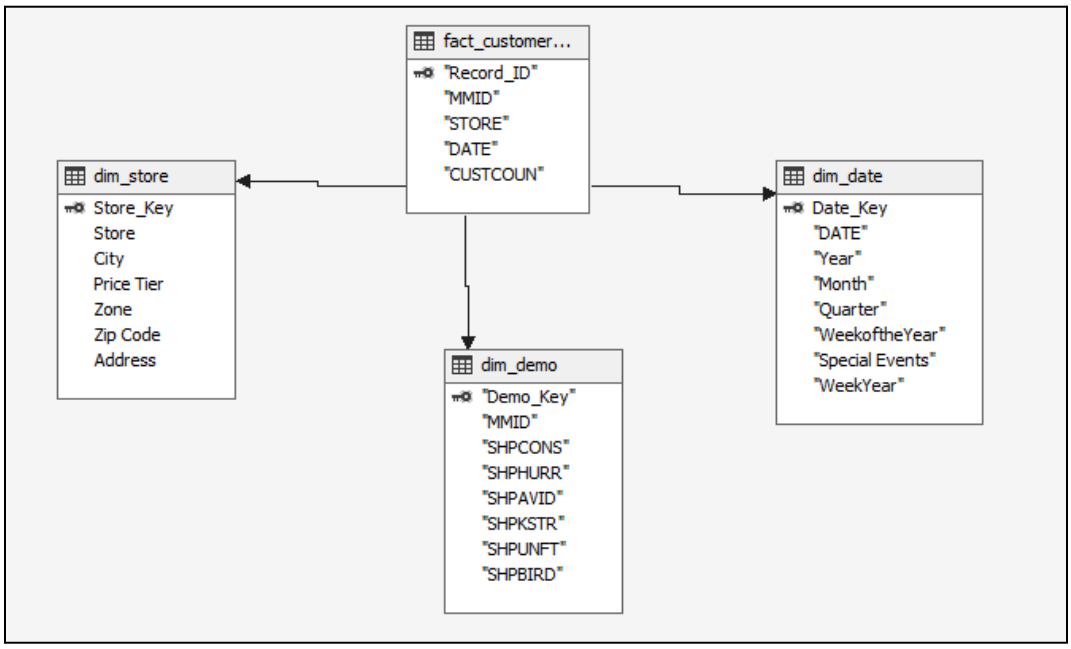


Fig: Data Source View Schema

Cube Creation

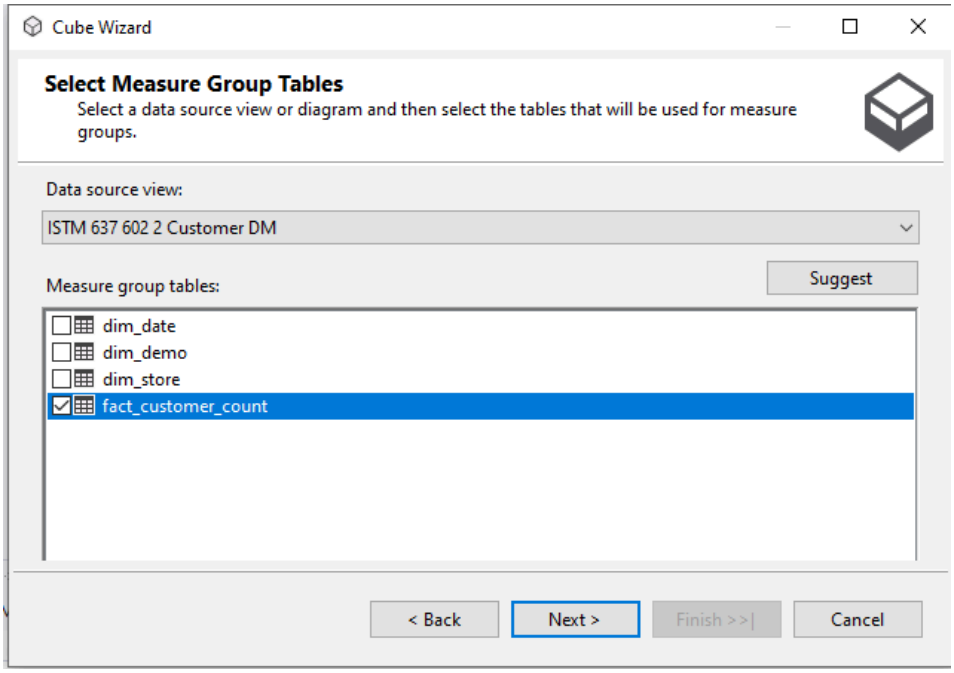


Fig: Selecting the Measure Group

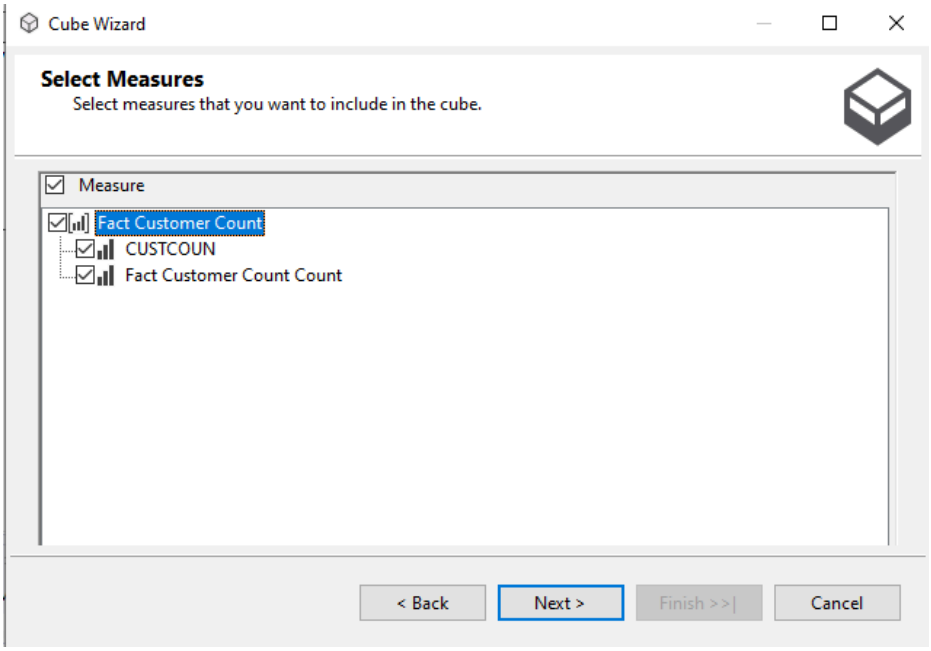


Fig: Selecting the Measures

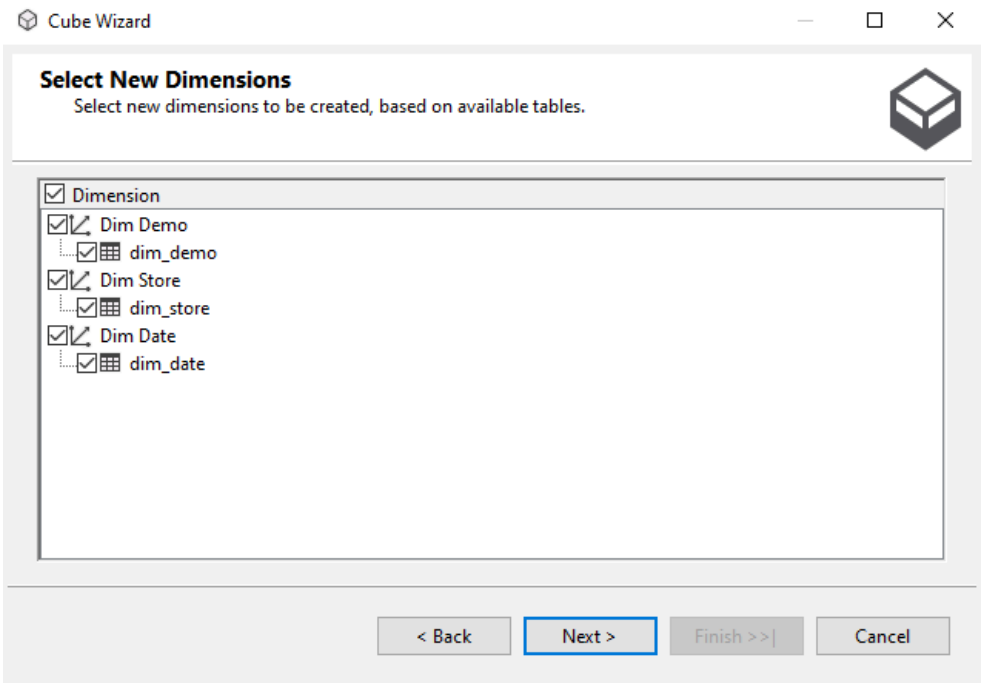


Fig: Selecting the Dimensions

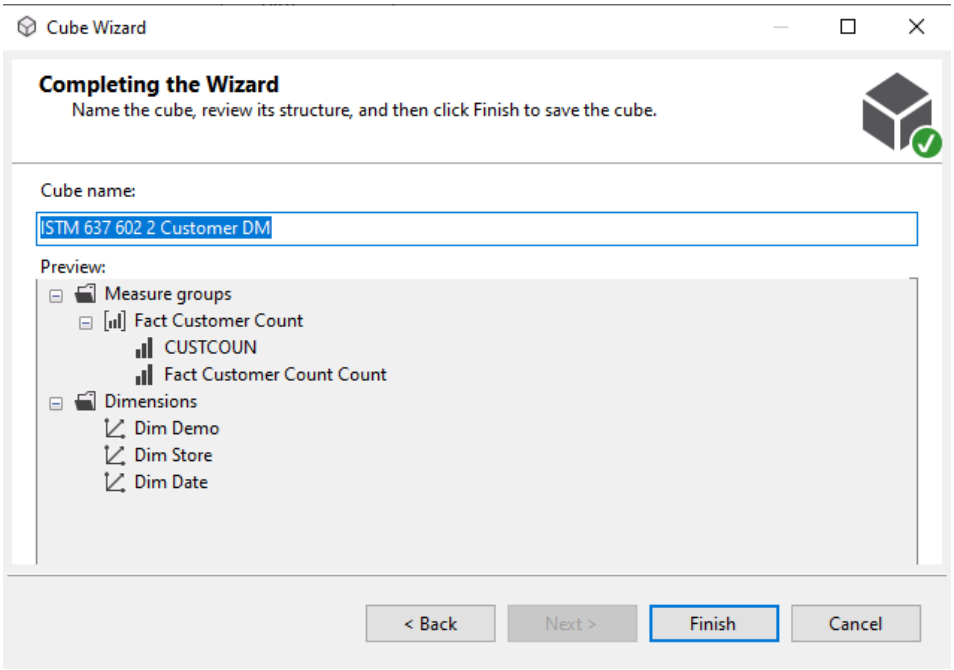


Fig: Cube Preview

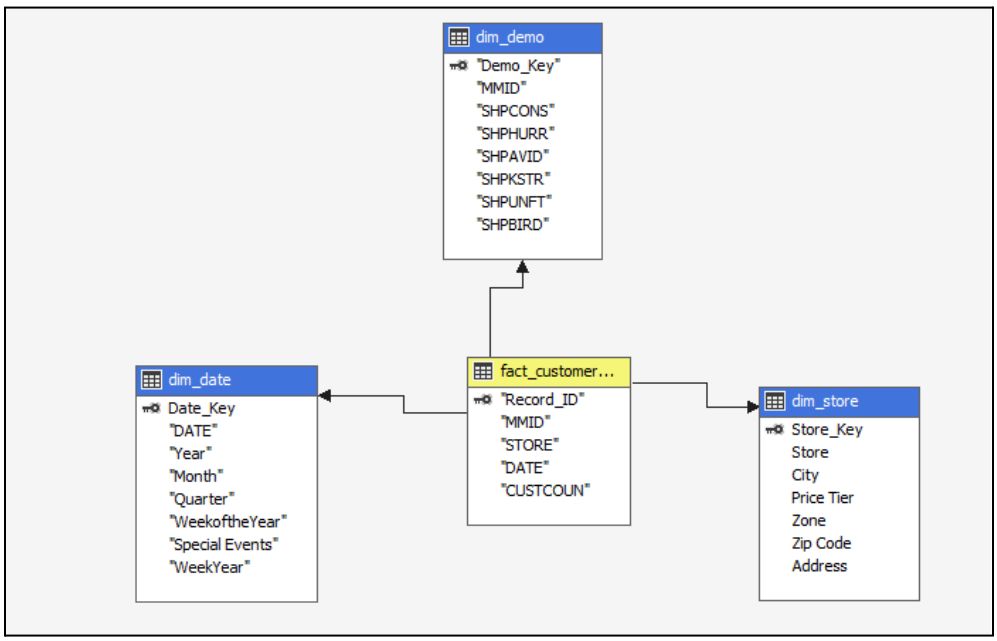


Fig: Cube Schema View

Hierarchies Creation:

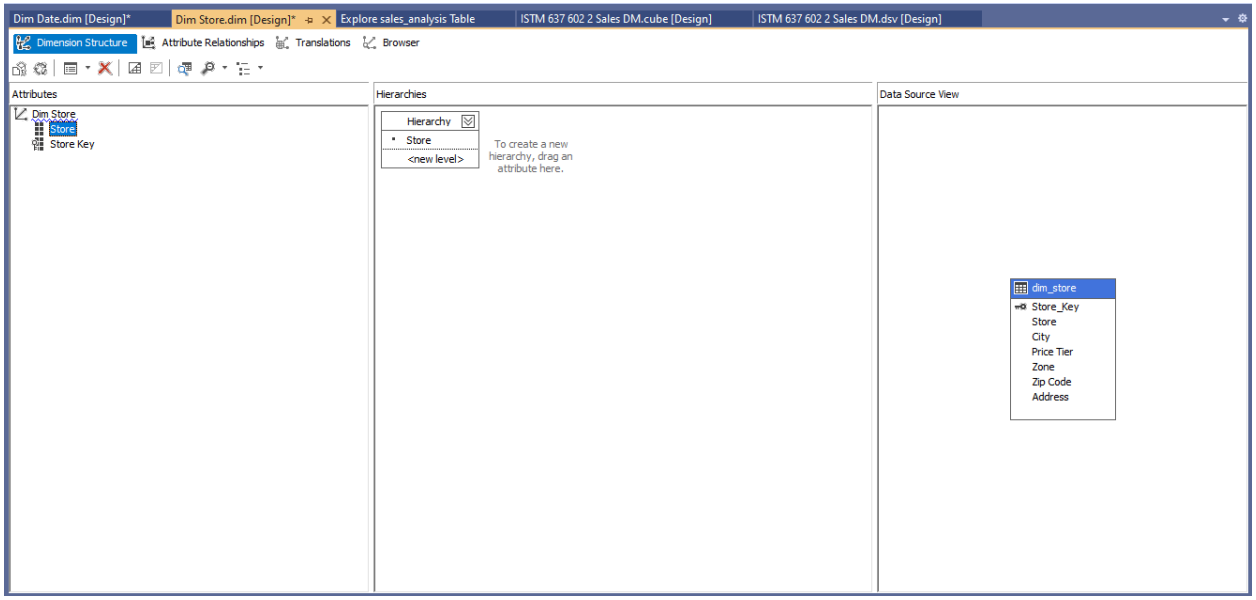


Fig: Store Hierarchy Creation

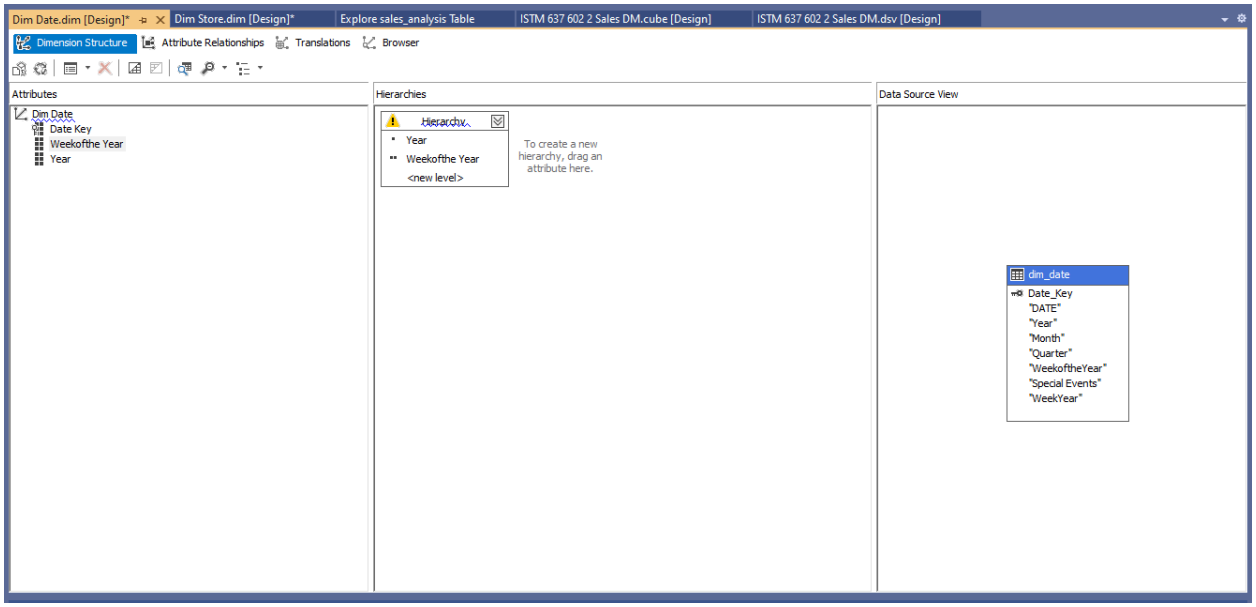


Fig: Date Hierarchy Creation

Final Report Browsing

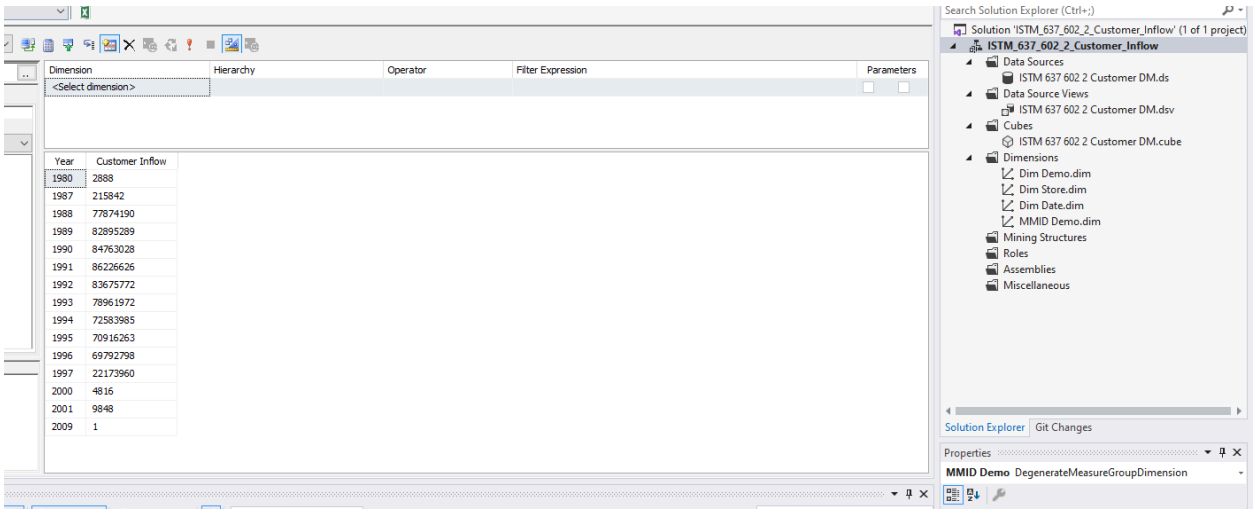


Fig: Report Created Successfully

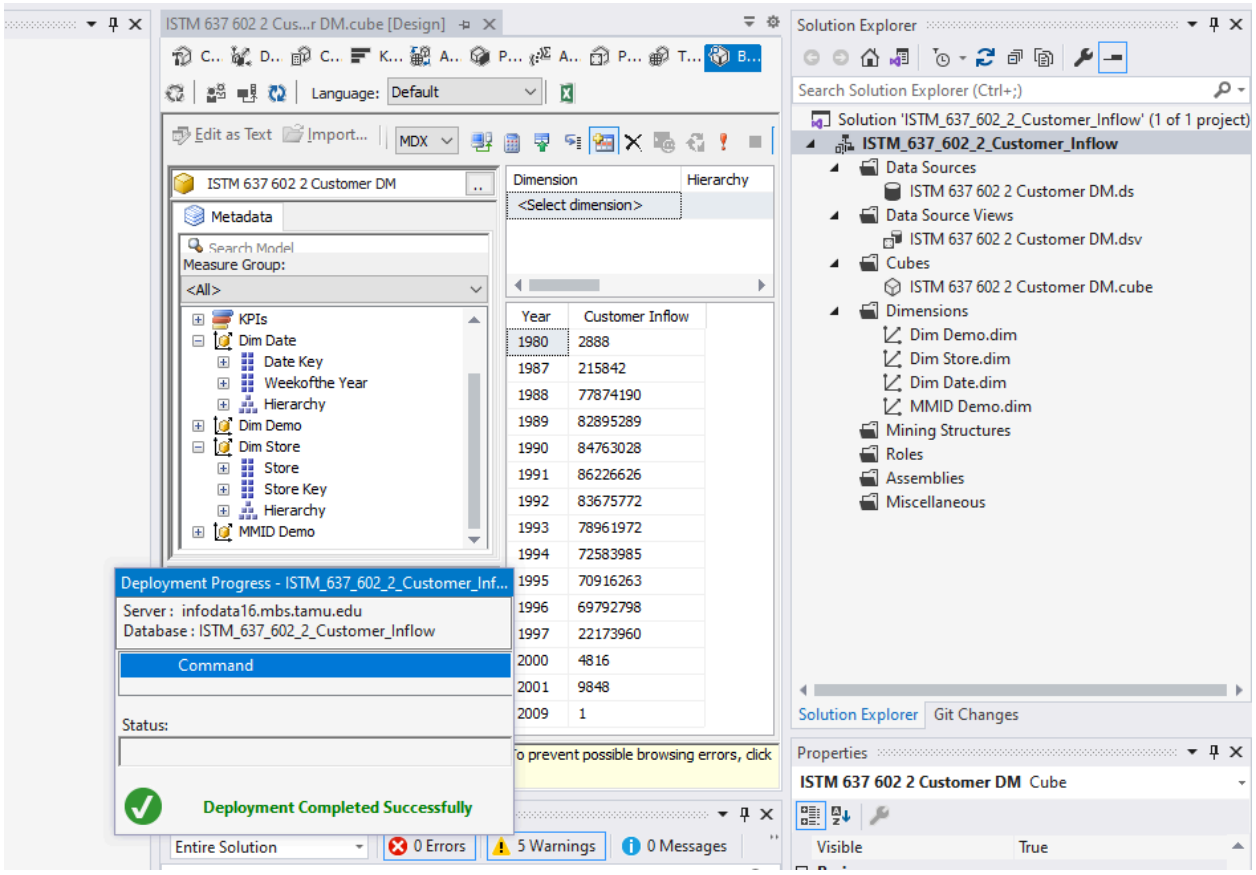


Fig: Better View of Reports with Results, Hierarchy and Tables

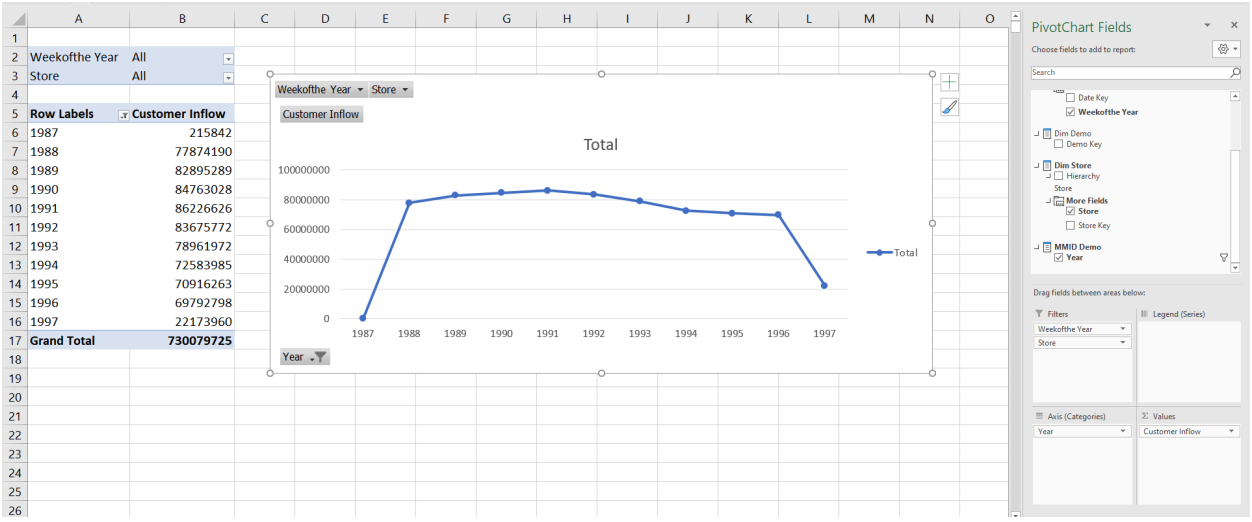


Fig: Exported the data to the Excel for Data Visualization

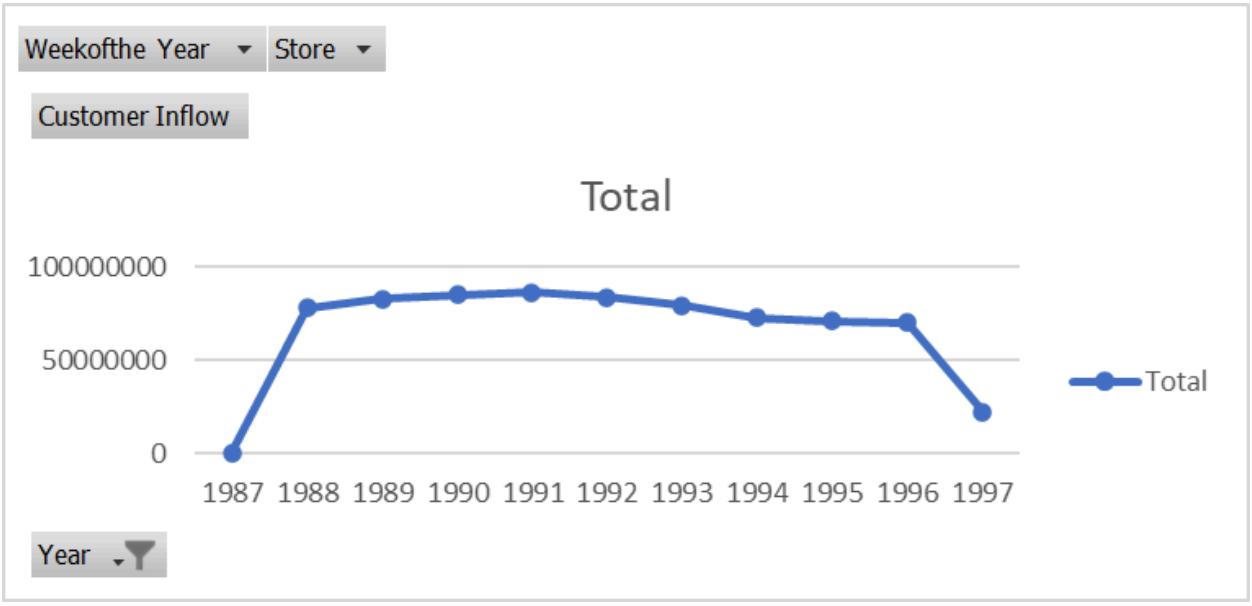


Fig: Excel Chart of data generated from the report

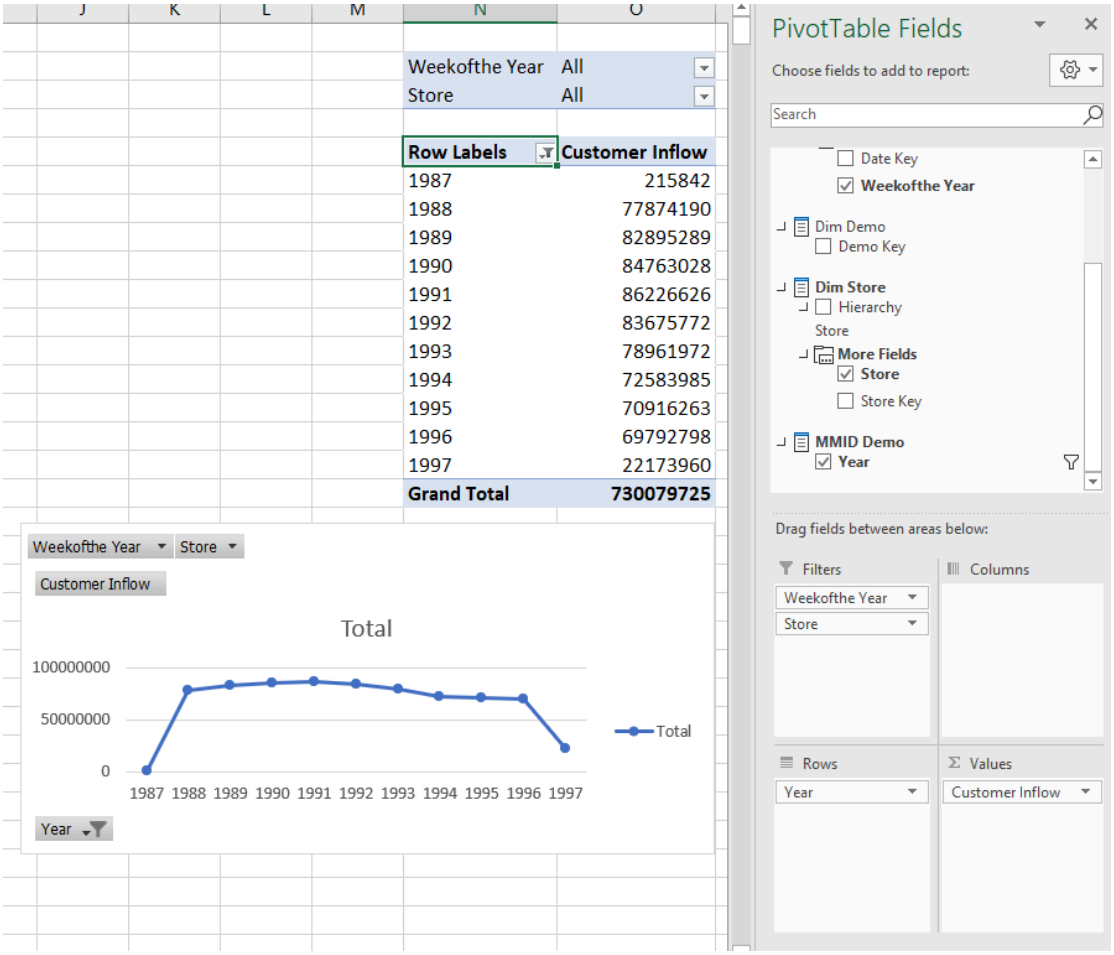


Fig: Better View of the Exported data to the Excel for Data Visualization

Data Visualization in SSMS

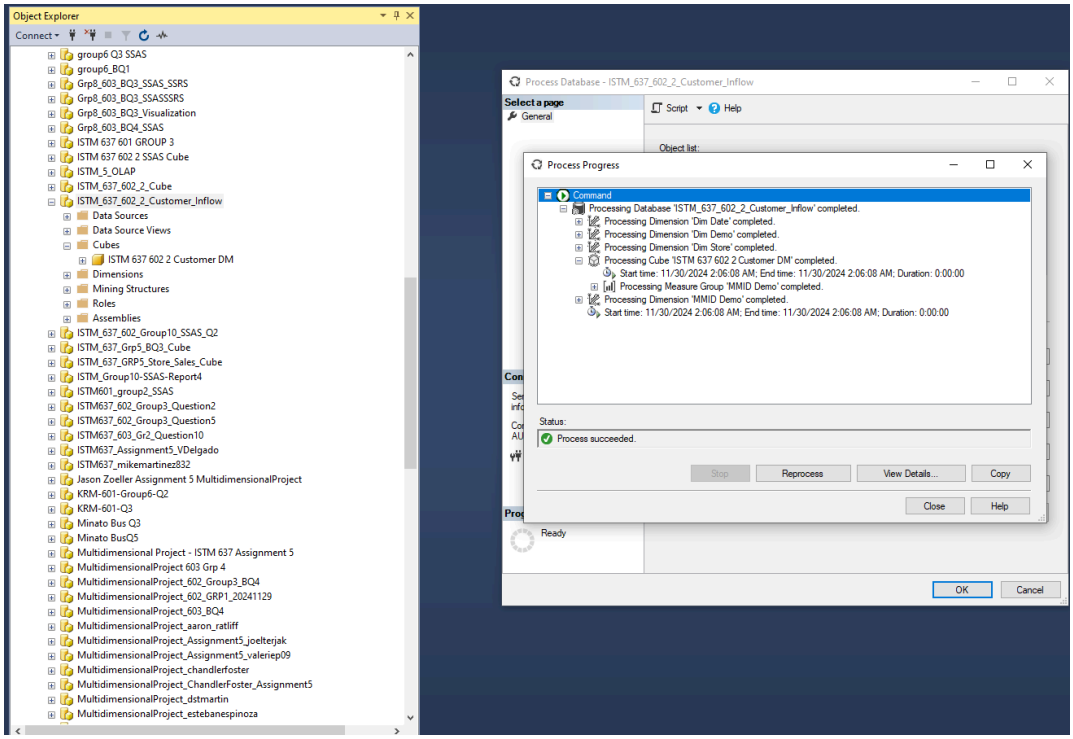


Fig: Successfully Processed the Cube in SSMS

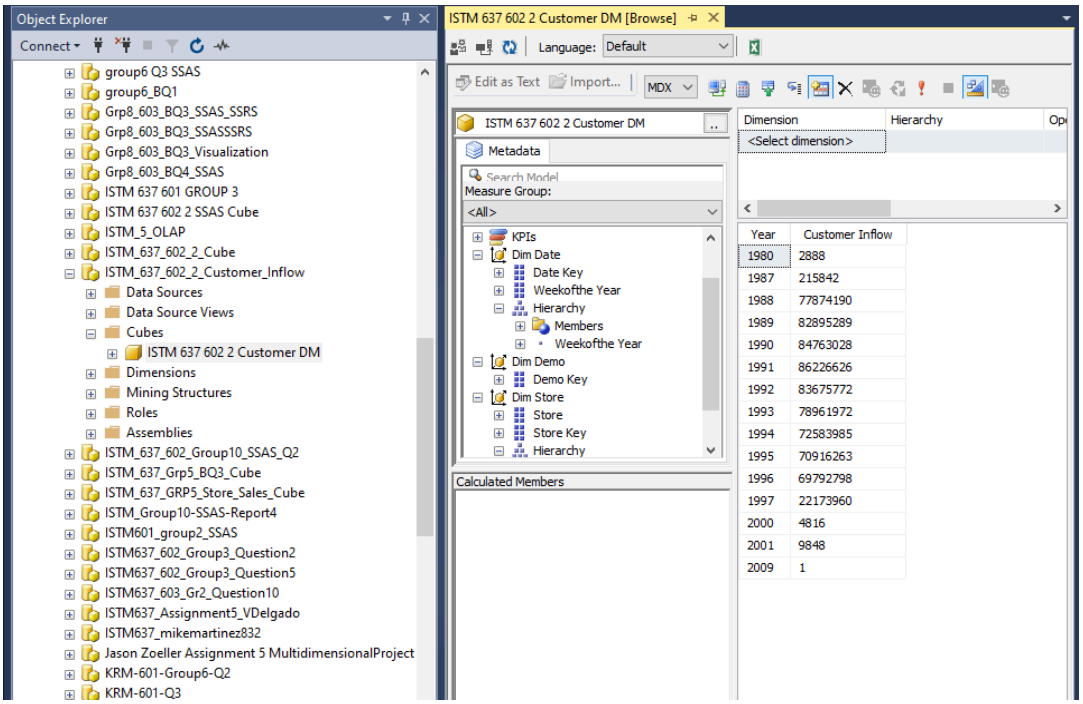


Fig: Better View of Reports with Results, Hierarchy and Tables in SSMS

Location of Data Marts and SSAS Cube

Data Marts

- ISTM_637_602_2_Sales_DM
- ISTM_637_602_2_Customer_DM

SSAS Cube

- ISTM_637_602_2_Cube
- ISTM_637_602_2_Customer_Inflow

Conclusion

In conclusion, this report has detailed the design, implementation, and analysis of a data warehousing solution for Dominick's Fine Food (DFF). By following Kimball's methodology, we developed a logical and physical design that integrates data from multiple sources, ensuring that the data warehouse is optimized for both analytical querying and reporting.

The five business questions selected by DFF were effectively addressed through well-structured data marts that allowed for insightful reporting. The design and implementation of the ETL process, along with the careful creation of dimension and fact tables, ensured that data was accurately extracted, transformed, and loaded, providing reliable data for decision-making.

Through the implemented reports, we gained valuable insights into sales trends, customer behavior, and store performance, which are essential for driving strategic decisions. These insights not only reflect the current performance of DFF but also offer a strong foundation for future analysis and continuous improvement.

Overall, the development of this data warehouse and its associated reports enhances DFF's ability to make data-driven decisions, providing a clear path toward optimizing business operations and improving performance across all stores.

References

1. https://www.company-histories.com/Dominicks-Finer-Foods-Inc-Company-History.html
2. https://en.wikipedia.org/wiki/Dominick's
3. https://www.chicagotribune.com/2017/02/27/dominicks-demise-led-to-boom-but-now-grocers-face-reality-new-rivals-online-shift/
4. Lemon, K. N., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. Journal of Marketing, 80(6), 69-96. https://doi.org/10.1509/jm.15.0420
5. Kumar, Viswanathan, and Werner Reinartz. "Creating enduring customer value." *Journal of marketing* 80.6 (2016): 36-68.
6. Bradlow, Eric T., et al. "The role of big data and predictive analytics in retailing." *Journal of retailing* 93.1 (2017): 79-95.
7. Kopalle, Praveen K., et al. "Dynamic pricing: Definition, implications for managers, and future research directions." *Journal of Retailing* 99.4 (2023): 580-593.
8. Dataset : https://www.chicagobooth.edu/research/kilts/research-data/dominicks
9. https://www.jmp.com/en_us/home.html
10. ISTM 637 Class Lectures and Material
11. Sen, Arun, and Atish P. Sinha. "A comparison of data warehousing methodologies." Communications of the ACM 48.3 (2005): 79-84
12. Sen, Arun, Keshavamurthy Ramamurthy, and Atish P. Sinha. "A model of data warehousing process maturity." IEEE Transactions on Software Engineering 38.2 (2011): 336-353.
13. https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dw-bi-lifecycle-method/
14. https://datascientest.com/en/kimball-method-what-is-it-how-do-i-use-it
15. https://www.astera.com/type/blog/data-warehouse-concepts/
16. https://www.holistics.io/books/setup-analytics/kimball-s-dimensional-data-modeling/