ISTM 637
Data Warehousing

Final Report- Group 9

Dominick's Fine Food

Submitted By
Rishi Shah
Pratyush Saxena
Sidhant Vyas
Tanmay Kakkad

Table of Contents

Section 1: Introduction

1.1 DFF and problems for the project

The phrase data warehousing refers to the use of data analysis and reporting on historical and current data to develop insightful ideas to aid decision-making. We assist management in making use of these insights by creating analytics reports utilizing Online Analytical Processing Technology (OLAP). We can increase system efficiency and enhance a company's earnings and production.

Dominick's Finer Foods is a retail shop company centered in Illinois that was started in 1918. Dominick's sales declined over time, culminating in the closure of more stores owing to low sales and the company's general bad performance. The company eventually closed all of its locations and exited the market in 2014, as sales declined and revenue plummeted. As a result, Dominick's dataset included a vast amount of product sales data, customer-related data, store-related data, and demographics. The study described in this paper includes data from 1989 to 1996.

Some problems that we faced during this project:

- **Data Size**: Our data was massive, over 4.76 GB. This made it hard to handle and process using regular tools. We needed to find better ways to deal with such a large amount of information.
- **Combining Different Data**: We had information spread across multiple files. Making all this data work together to tell a single story was challenging.
- **Inconsistent Data Points**: Some of our data had a lot of variation, making it tough to group or analyze. We had to find ways to deal with these inconsistencies.
- **Changing Trends Over Time**: Since our data spanned several years, it had different trends and patterns depending on the time period. We needed to keep this in mind when looking at the bigger picture.

Our rest of the report contains ERD's, Pivot tables and business questions that we formed to answer some of the pain points regarding DFF.

1.2 Understanding the data

The dataset we are working with is derived from the James M. Klits Center at the University of Chicago Booth School of Business. It contains historical data spanning nine years, from 1989 to 1994, collected during joint research conducted by Chicago Booth and Dominick's Finer Foods. This research aims to enhance shelf management and pricing in the retail industry.

The dataset is substantial, comprising over 4.76 GB of data, and requires significant preprocessing to clean and prepare it for meaningful analysis. It encompasses nine years of store-level data for over 3,500

Universal Product Codes (UPCs) sold across approximately 100 stores in the United States, with a concentration of stores in the Chicago area. The product line is categorized into 29 distinct categories.

The data is structured into four files, falling into two categories: general files and category files, all available in .csv format for sales analysis of Dominick's Finer Foods (DFF).

1. **Customer Count Files:** These files contain information about in-store traffic recorded weekly via scanners at each DFF store. They provide details on sales for various product categories, including Cheese, Cosmetic, Floral, and Dairy, at the store level on a weekly basis.

   Additionally, this data includes information about product purchases made using coupons.

2. **Store-Specific Demographics:** This file contains store-wise customer demographic information, including age groups, household income, number of dependent members, employment status, and retired status of customers.

   This demographic data is vital for later stages of the project, such as data warehousing and formulating store-specific strategies targeting different demographics.

3. **UPC Files:** These files map each product to a unique UPC and provide additional product-related information. The mapping of UPC codes to product descriptions can be valuable for formulating product-specific strategies.

4. **Movement Files:** These files offer insights into category-wise weekly product movements within DFF. They are crucial for understanding profit margins, which, in turn, inform strategies to reduce losses and increase profits.

   Analyzing weekly sales data can provide an advantage to DFF's inventory department in preparing for peak and off-peak seasons.

In addition to these data sources, a "Weeks Decode Table" facilitates week-to-date mappings, which are valuable for analysis and formulating business strategies.

While extensive and rich in information, this dataset presents the challenge of handling and cleaning the data to derive meaningful insights for retail analysis.

1.3 Metadata description for all the OLTP Source files

1. **CCOUNT:** Below are the descriptions for each attribute in the 'ccount' file. The file includes details about in-store customer visits, as well as data pertaining to sales and total coupons redeemed for purchasing.

Variable	Description	Type	Length

DATE	Date of the Observation	Character	6
Week	Week Number	Numeric	8
Store	Store Code	Numeric	8
BAKCOUP	Bakery Coupons Redeemed	Numeric	8
BAKERY	Bakery Sales in Dollars	Numeric	8
BEER	Beer Sales in Dollars	Numeric	8
BOTTLE	Bottle Sales in Dollars	Numeric	8
BULK	Bulk Sales in Dollars	Numeric	8
BULKCOUP	Bulk Coupons Redeemed	Numeric	8
CAMERA	Camera Sales in Dollars	Numeric	8
CHEESE	Cheese Sales in Dollars	Numeric	8
CONVFOOD	Conventional Foods Sales in Dollars	Numeric	8
COSMCOUP	Cosmetics Coupons Redeemed	Numeric	8
COSMETIC	Cosmetics Sales in Dollars	Numeric	8
CUSTCOUN	Customer Count	Numeric	8
DAIRCOUP	Dairy Coupons Redeemed	Numeric	8
DAIRY	Dairy Sales in Dollars	Numeric	8
DELI	Deli Sales in Dollars	Numeric	8
DELICOUP	Deli Coupons Redeemed	Numeric	8
DELIEXPR	Deli Express Sales in Dollars	Numeric	8
DELISELF	Deli Self Service Sales in Dollars	Numeric	8
FISH	Fish Sales in Dollars	Numeric	8
FISHCOUP	Fish Coupons Redeemed	Numeric	8

FLORAL	Floral Sales in Dollars	Numeric	8
FLORCOUP	Floral Coupons Redeemed	Numeric	8
FROZCOUP	Frozen Items Coupons Redeemed	Numeric	8
FROZEN	Frozen Items Sales	Numeric	8
FTGCCOUP	Food-to-Go Coupons Redeemed	Numeric	8
FTGCHIN	Food-to-Go Chinese Sales in Dollars	Numeric	8
FTGICOUP	Food-to-Go Coupons Redeemed	Numeric	8
FTGITAL	Food-to-Go Italian Sales in Dollars	Numeric	8
GM	General Merchandise Sales in Dollars	Numeric	8
GMCOUP	General Coupons Redeemed	Numeric	8
GROCCOUP	Grocery Coupons Redeemed	Numeric	8
GROCERY	Grocery Sales in Dollars	Numeric	8
HABA	Health and Beauty Aids Sales in Dollars	Numeric	8
HABACOUP	Health and Beauty Aids Coupons Redeemed	Numeric	8
JEWELRY	Jewelry Sales in Dollars	Numeric	8
LIQCOUP	Liquor Coupons Redeemed	Numeric	8
MANCOUP	Manufacturer Coupons Redeemed	Numeric	8
MEAT	Meat Sales in Dollars	Numeric	8
MEATCOUP	Meat Coupons Redeemed	Numeric	8
MEATFROZ	Meat-Frozen Sales in	Numeric	8

	Dollars		
MISCSCP	Misc. Coupons Redeemed	Numeric	8
MVPCLUB	MVP	Numeric	8
PHARCOUP	Pharmacy Coupons Redeemed	Numeric	8
PHARMACY	Pharmacy Sales in Dollars	Numeric	8
PHOTCOUP	Photo Coupons Redeemed	Numeric	8
PHOTOFIN	Photo	Numeric	8
PRODCOUP	Produce Coupons Redeemed	Numeric	8
PRODUCE	Produce Sales in Dollars	Numeric	8
PROMCOUP	Promotion Coupons Redeemed	Numeric	8
PROMO	Promotion Sales in Dollars	Numeric	8
SALADBAR	Salad Bar Sales in Dollars	Numeric	8
SALCOUP	Salad Coupons Redeemed	Numeric	8
SPIRITS	Spirits Sales in Dollars	Numeric	8
SSDELICP	Self Service Deli Sales in Dollars	Numeric	8
VIDCOUP	Video Coupons Redeemed	Numeric	8
VIDEO	Video Sales in Dollars	Numeric	8
VIDEOREN	Video Rentals (Dollar Amounts)	Numeric	8
WINE	Wine Sales in Dollars	Numeric	8

2. **DEMOGRAPHICS:** Below is the description for each attribute in the Demography file, which contains census data for the Chicago metropolitan area. This dataset includes information about customer demographics, age groups, households, and purchasing behavior in different locations.

Variable Name	Description	Variable Name	Description
age9	% Population under age 9	single	% of Singles
age60	% Population over age 60	retired	% of Retired
ethnic	% Blacks & Hispanics	unemp	% of Unemployed
educ	% College Graduates	wrkch5	% of working women with children under 5
nocar	% With No Vehicles	wrkch17	% of working women with children 6 - 17
income	Log of Median Income	nwrkch5	% of non-working women with children under 5
incsigma	Std dev of Income Distribution (Approximated)	nwrkch17	% of non-working women with children 6 - 17
hsizeavg	Average Household Size	wrkch	% of working women with children
hsize1	% of households with 1 person	nwrkch	% of non-working women with children
hsize2	% of households with 2 persons	wrkwch	% of working women with children under 5
hsize34	% of households with 3 or 4 persons	wrkwnch	% of working women with no children
hsize567	% of households with 5 or more persons	telephn	% of households with telephones
hh3plus	% of households with 3 or more persons	mortgage	% of households with mortgages
hh4plus	% of households with 4 or more persons	nwhite	% of population that is non-white
hhsingle	% of households with 1 person	poverty	% of population with income under $15,000
hhlarge	% of households with 5 or more persons	shopcons	% of Constrained Shoppers

workwom	% Working Women with full-time jobs	shophurr	% of Hurried Shoppers
sinhouse	% Detached Houses	shopavid	% of Avid Shoppers
density	Trading Area in Sq Miles per Capita	shopstr	% of Shopping Strangers
hval150	% of Households with Value over $150,000	shopunft	% of Unfettered Shoppers
hval200	% of Households with Value over $200,000	shopbird	% of Shopper Birds
hvalmean	Mean Household Value (Approximated)	shopindx	Ability to Shop (Car and Single Family House)

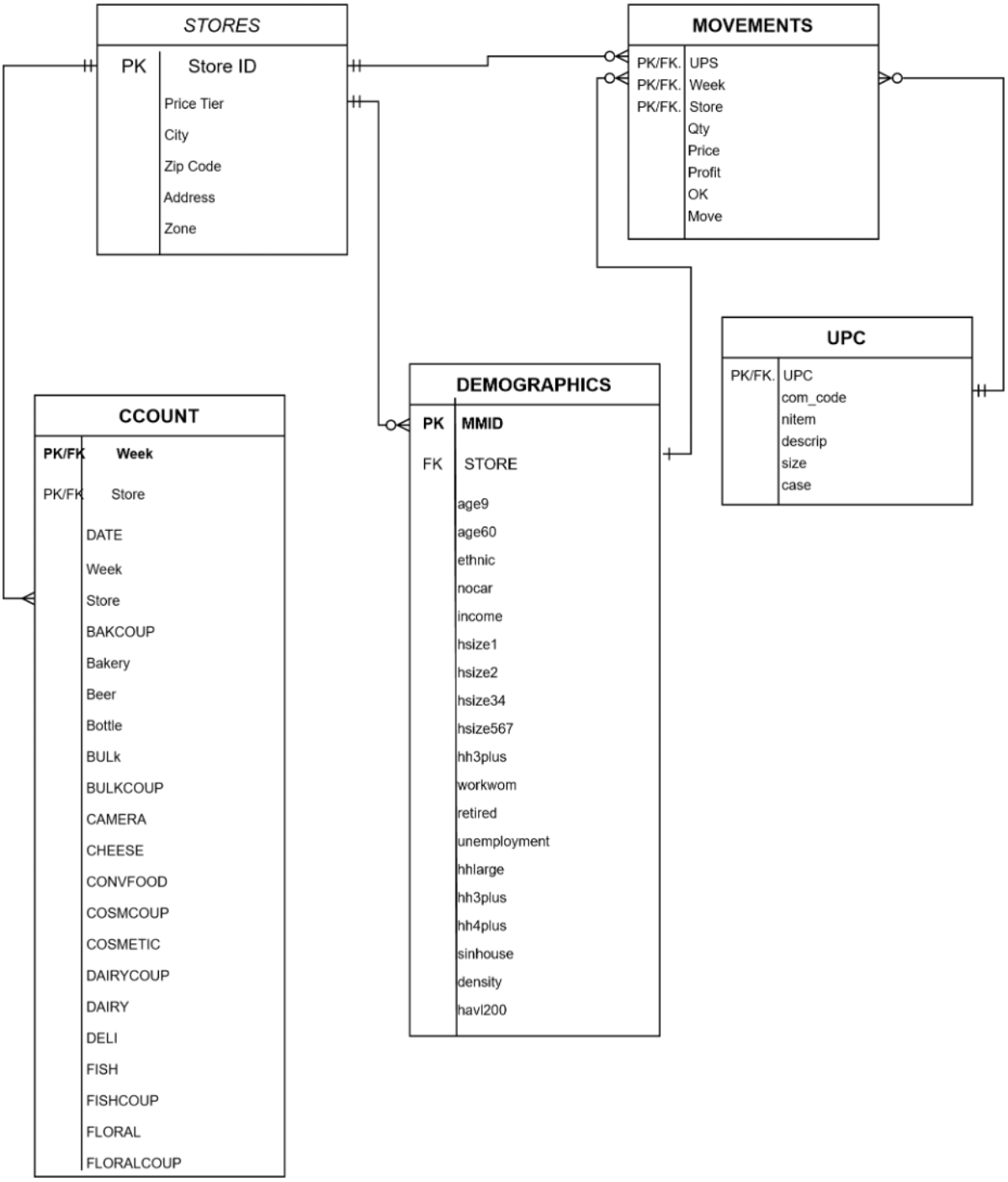
3. **UPC Files:** Below, you'll discover the attribute details within the UPC file, including the UPC number, Dominick's Commodity Code, Dominick's item code, Product Name, Product Size, and Case.

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

4. **Movement Files:** Below is the description of each attribute found in the Movement file, which includes metadata such as UPC number, week number, the quantity of units sold, price, profit, sale, and the 'ok' indicator.

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

Entity-Relation Diagram


STORES
PK
Store ID
Price Tier
City
Zip Code
Address
Zone
MOVEMENTS
PK/FK.
UPS
PK/FK.
Week
PK/FK.
Store
Qty
Price
Profit
OK
Move
UPC
PK/FK.
UPC
com_code
nitem
descrip
size
case
DEMOGRAPHICS
PK
MMID
FK
STORE
age9
age60
ethnic
nocar
income
hsize1
hsize2
hsize34
hsize567
hh3plus
workwom
retired
unemployment
hhlarge
hh3plus
hh4plus
sinhouse
density
havl200
CCOUNT
PK/FK
Week
PK/FK
Store
DATE
Week
Store
BAKCOUP
Bakery
Beer
Bottle
BULk
BULKCOUP
CAMERA
CHEESE
CONVFOOD
COSMCOUP
COSMETIC
DAIRYCOUP
DAIRY
DELI
FISH
FISHCOUP
FLORAL
FLORALCOUP

Section 2 Business Questions And Analysis

1.1 Research and Understanding

After analyzing the data for DFF and reading papers about the retail industry, there are certain challenges in the retail industry that are also relevant for DFF:

1. **Retail Performance Indicators:**

   DFF's data provides information on store performance like weekly volume and profit. This shows the difficulty of measuring store performance, identifying failing stores, and developing plans to improve their chances of success.

   Preuss, Björn & Argiolas, Matteo. (2016)

2. **Product Choices and Inventory Management:**

   Ordering and maintaining the right product choices depending on the demand is a big task in the retail industry. Keeping the right amount of each product in stock based on accurate forecasting and analyzing historical sales of each product is key.

   (Luther, 1993)

3. **Turnover rate and sales strategies:**

   Again, a major challenge in the retail domain. For example, The WFEC.CSV dataset showcases weekly movement, quantity, price, sales, and profit data for specific UPC codes at the store level. This data underscores the challenge of predicting inventory turnover rates and brainstorming sales strategies based on real-time data.

   (Breivik, 2019)

4. **Data integration and Data Warehousing:**

   It is a big challenge to integrate data from many different places, data such as product details, store demographics, and sales data, into one central data warehouse for useful insights.

   Girsang, Ganda & Arisandi, Geri & Elysisa, Calista & Michelle, & Saragih, Melva. (2019)

5. **Demographic driven marketing:**

   The idea of demographic-driven product matching is particularly significant for a retail chain such as DFF which operates many outlets in possibly diverse areas. Products that are tailored to the requirements and tastes of the local demographic can increase sales and improve customer satisfaction.

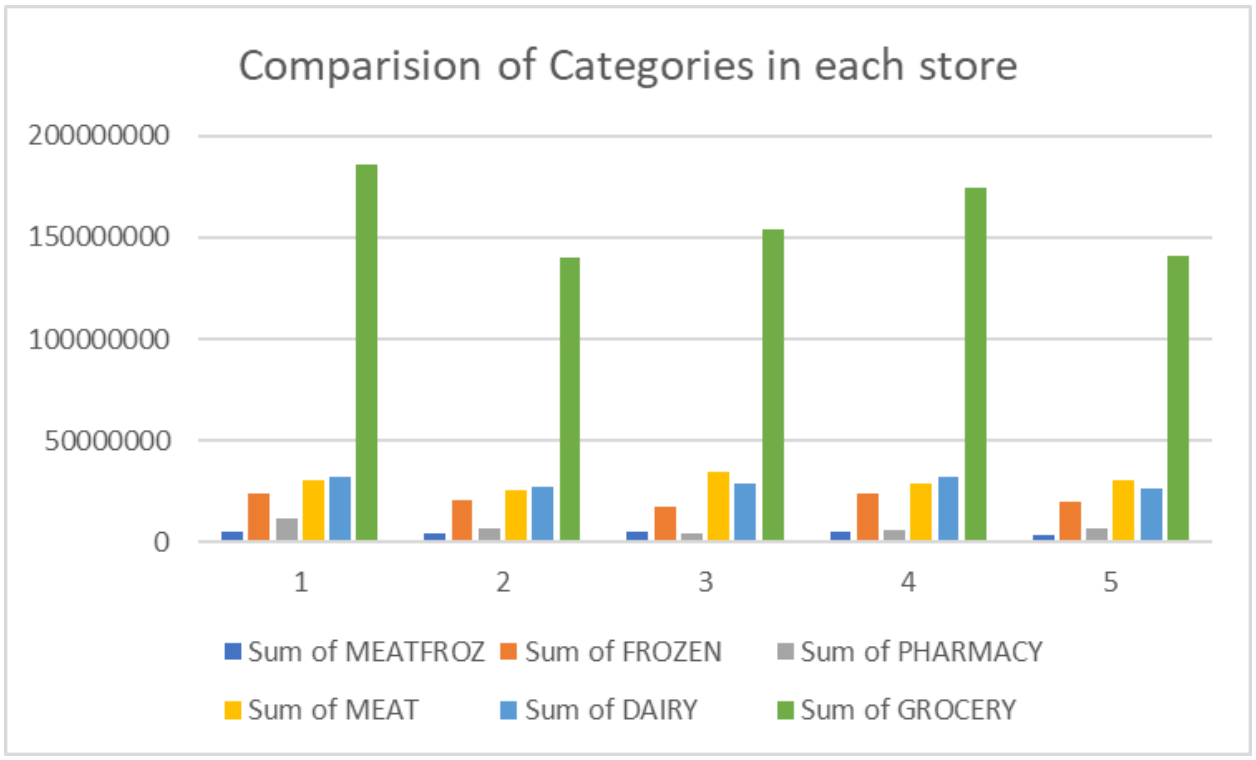
   (Fairlie, 2023)

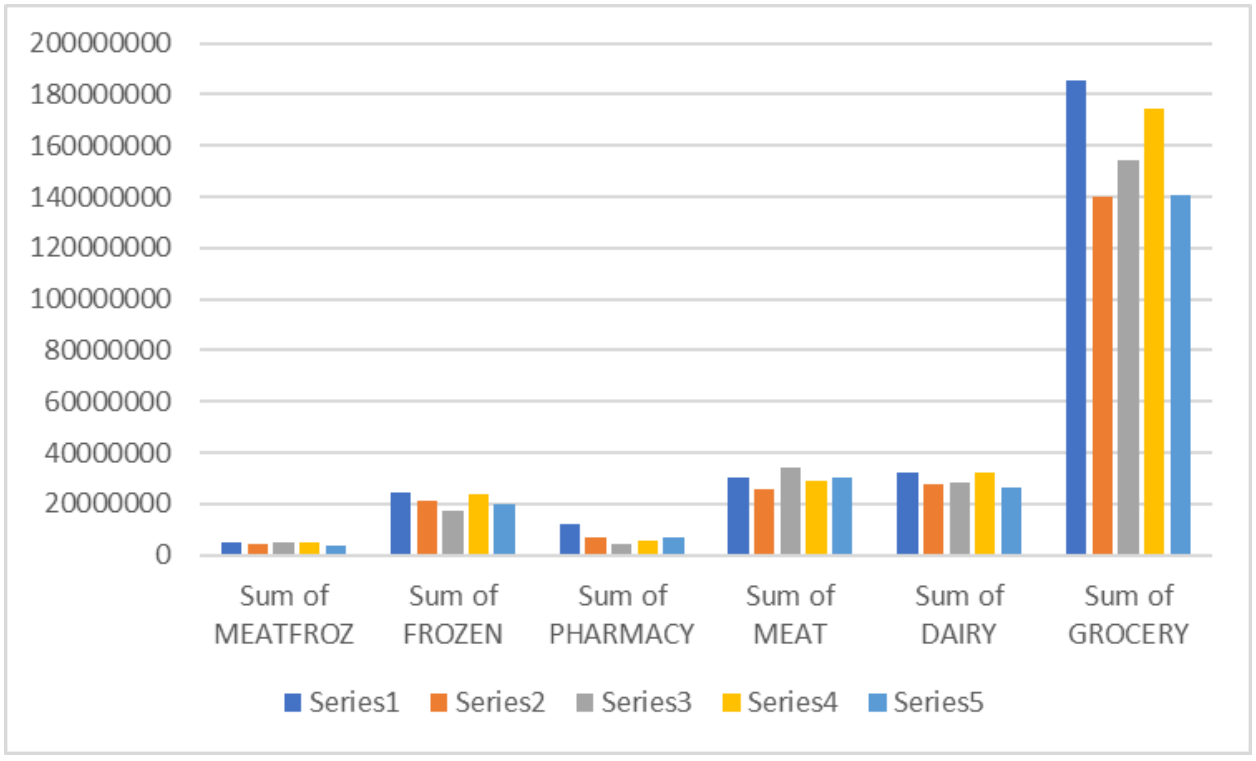
1.2 Business Questions

1. In the 5 most successful stores, what is the contribution and distribution of different categories to the sales of the stores?

Five stores with the most sales have been chosen here. Some important categories have been chosen, representing sales in those categories.

Store Number	Sum of MEATFROZ	Sum of FROZEN	Sum of PHARM	Sum of MEAT	Sum of DAIRY	Sum of GROCE	Sum of Total Sales
301	4891261.6	24395103.51	11884658.31	30568699.23	32402710.09	185527951.3	$42,66,61,906.93
302	4200915.11	20947788.01	7155962.07	25751849.24	27541333.91	139748864.5	$33,87,72,994.16
303	5027360.81	17633041.16	4629475.48	34533535.17	28694849.56	153990851.4	$38,07,66,921.18
304	5022176.52	23994384.66	5928998.9	28989389.11	32089857.81	174660751.8	$42,38,70,473.51
315	3651971.34	19672394.3	7111235.81	30532512.5	26428170.12	140672544.9	$35,20,82,807.55


Comparision of Categories in each store
200000000
150000000
100000000
50000000
0
1
2
3
4
5
Sum of MEATFROZ
Sum of FROZEN
Sum of PHARMACY
Sum of MEAT
Sum of DAIRY
Sum of GROCERY


200000000
180000000
160000000
140000000
120000000
100000000
80000000
60000000
40000000
20000000
0
Sum of MEATFROZ
Sum of FROZEN
Sum of PHARMACY
Sum of MEAT
Sum of DAIRY
Sum of GROCERY
Series1
Series2
Series3
Series4
Series5

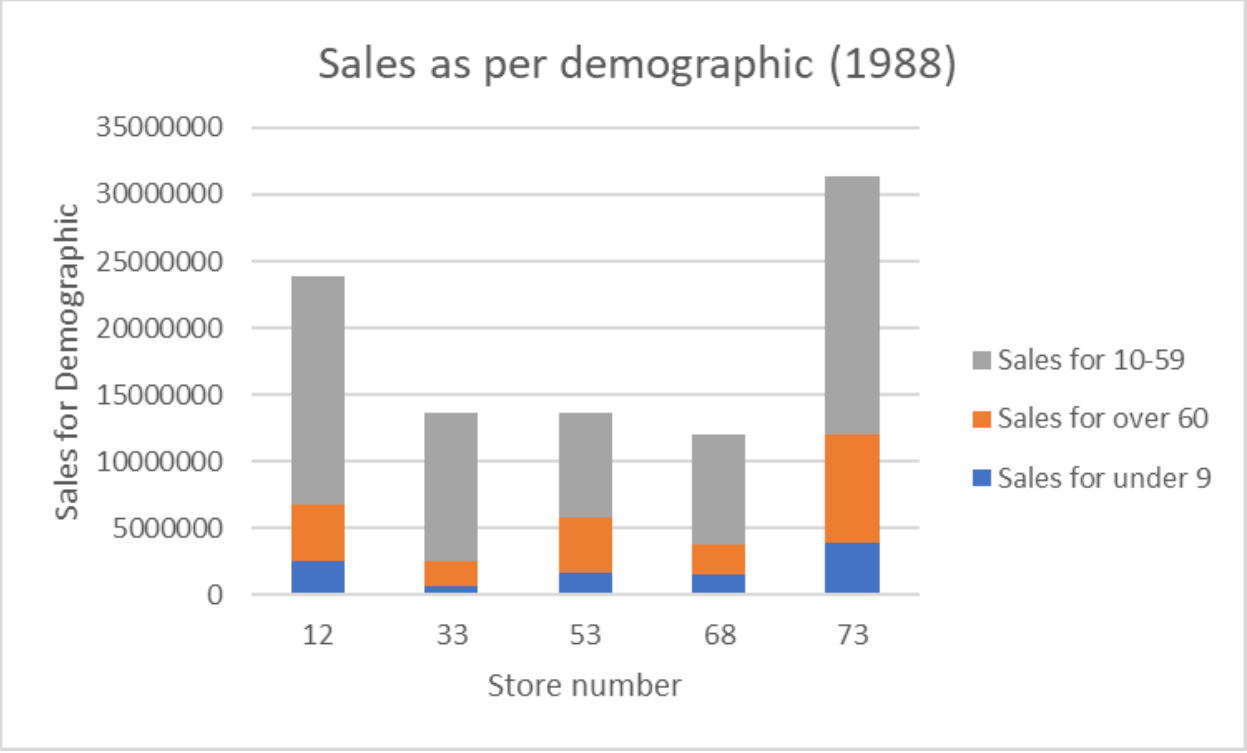
Explanation:

This will help us understand what product categories make the successful stores different from the other ones so as to take advantage of those trends and replicate it into other stores to increase the sales and profits
The patterns in higher sales of certain products and categories can be an underlying reason for a store to be more successful than others.

2. What are store wise demographics across the total company sales for each product?

Assumption: Comparison between stores should be for same geographical area, such as city

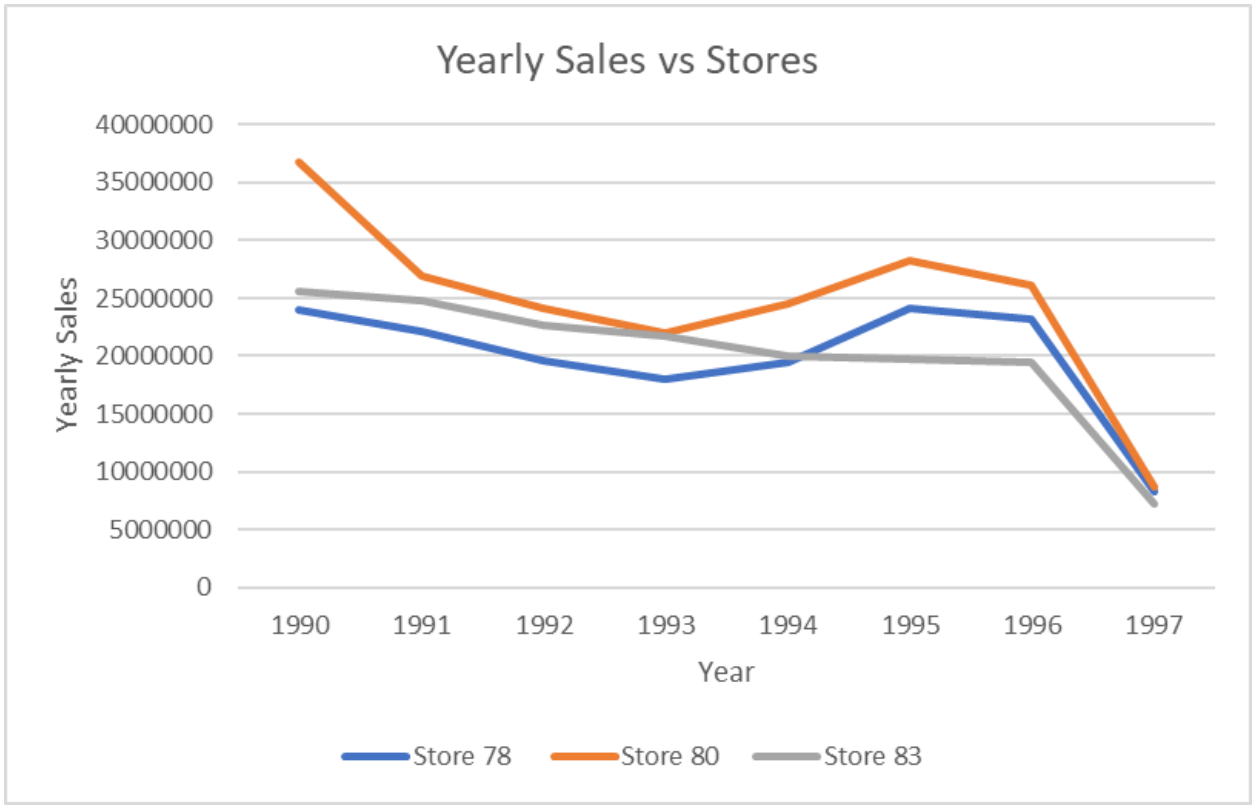
Explanation: This business question would help the company understand the contribution of different demographics in the age groups to the total sales of different stores in a similar geographic location. For the chart, the stores in the city of Chicago have been chosen. This insight could help us identify the stores that are more appealing to a specific demographic age group so as to help DFF manage product inventory catered to that particular age group better according to each store.



Store	Year	Yearly Sales	% Under 9	% Above 60	% Age 10-59	Sales for under 9	Sales for over 60	Sales for 10-59
12	1988	23808383	0.10569674	0.178341405	0.715961855	2516468.461	4246020.475	17045894.06
33	1988	13668837	0.046070917	0.134169966	0.819759117	629735.8576	1833947.389	11205153.75
53	1988	13668837	0.120839139	0.300278681	0.57888218	1651730.497	4104460.344	7912646.159
68	1988	11959736	0.130970406	0.181417756	0.687611838	1566371.481	2169708.472	8223656.047
73	1988	31307862	0.124465022	0.257450782	0.618084196	3896733.739	8060233.555	19350894.71

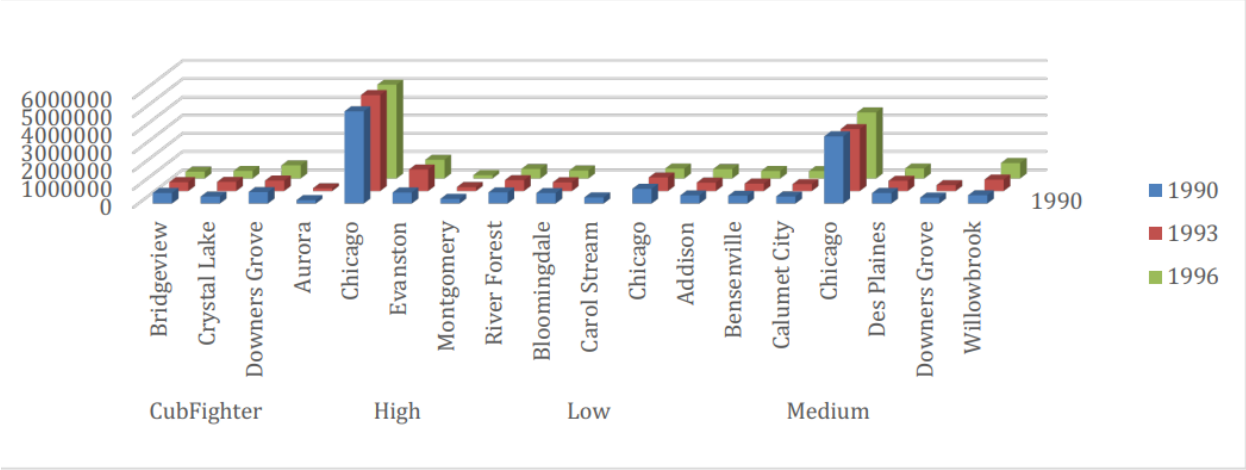
3. Which DFF stores have experienced significant changes in sales performance over the years, and what actions have been taken to address these changes?

Explanation: Monitoring sales performance at individual stores and understanding the actions taken in response can help DFF maintain consistency in performance across branches. Moreover, learning from successful and unsuccessful techniques implemented in a few branches can help reform the policies for other branches. Yearly sales have been compared for three different stores here to gauge the trend.



	A	B	C	D
1		Yearly Sales		
2	Year	Store 78	Store 80	Store 83
3	1990	24032686.81	36792937.31	25529548.43
4	1991	22171219.96	26949324.16	24781530.16
5	1992	19620416.55	24087558.68	22681969.52
6	1993	17999314.92	22038051.46	21774441.05
7	1994	19467081.4	24533276.41	19981276.05
8	1995	24116661.32	28256476.01	19693797.9
9	1996	23136112.96	26184521.08	19412981.54
10	1997	8263455.27	8640254.02	7153637.7
11				

4. What do the sales patterns for bakery products look like over a span of three years, categorized into low, medium, and high-price tier stores across multiple cities?

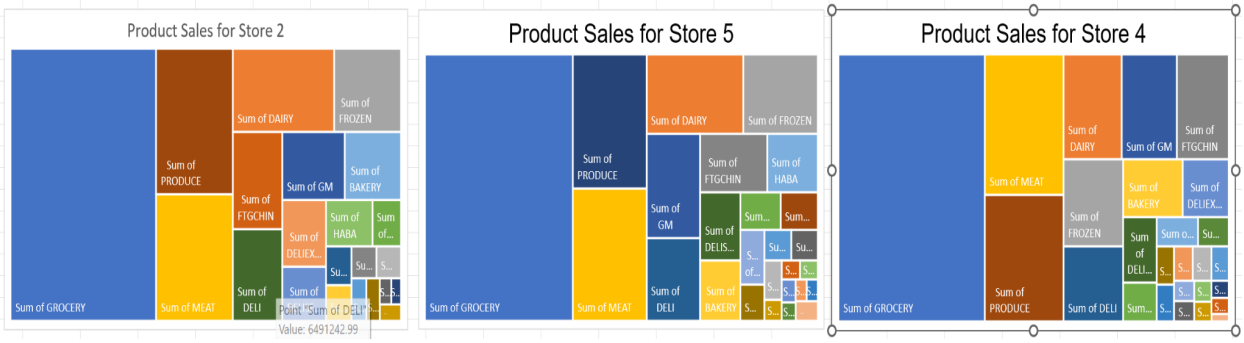


High	**6745781.16**	**7436234.37**	**6891231.84**	**21073247.37**
Aurora	191481.22	152740.09		344221.31
Chicago	5073742.74	5280422.25	5160627.35	15514792.34
Evanston	607552.26	1184031.99	1028903.94	2820488.19
Montgomery	260295.02	219294.22	182379.53	661968.77
River Forest	612709.92	599745.82	519321.02	1731776.76
Low	**1724088.33**	**1226557.66**	**981506.38**	**3932152.37**
Bloomingdale	577365.85	478940.12	437200.72	1493506.69
Carol Stream	337415.84			337415.84
Chicago	809306.64	747617.54	544305.66	2101229.84
Medium	**6334888.3**	**6175244.55**	**6364736.66**	**18874869.51**
Addison	447868.07	460456.46	515920.81	1424245.34
Bensenville	431892.01	409033.09	411613.06	1252538.16
Calumet City	392233.66	381428.94	407154.09	1180816.69
Chicago	3698765.84	3408436.87	3636043.25	10743245.96
Des Plaines	581424.56	562344.09	545947.57	1689716.22
Downers Grove	327247.18	322840.44		650087.62
Willowbrook	455456.98	630704.66	848057.88	1934219.52
Grand Total	**16406839.98**	**16405923.52**	**15753898.04**	**48566661.54**

Explanation: The posed inquiry equips managers to tackle concerns regarding store performance, specifically in relation to pricing tiers (low, medium, high, and budget-friendly). Upon data examination, it becomes evident that sales of bakery items exhibit a stable trajectory across the four pricing tiers in various cities. However, it is apparent that no notable expansion occurs even within consistently performing outlets like those in Chicago (high and medium pricing tiers). The span of every three years offers an adequate timeframe to gauge growth rates. Consequently, this query aids in evaluating sales performance for products that demonstrate substantial and enduring market demand, spanning various municipalities and pricing categories. Bakery goods have enjoyed a lasting presence in the market, making them an ideal subject for assessing customer trends in diverse urban centers.

5. What are the top-performing product categories in terms of sales revenue for each DFF branch over the years?

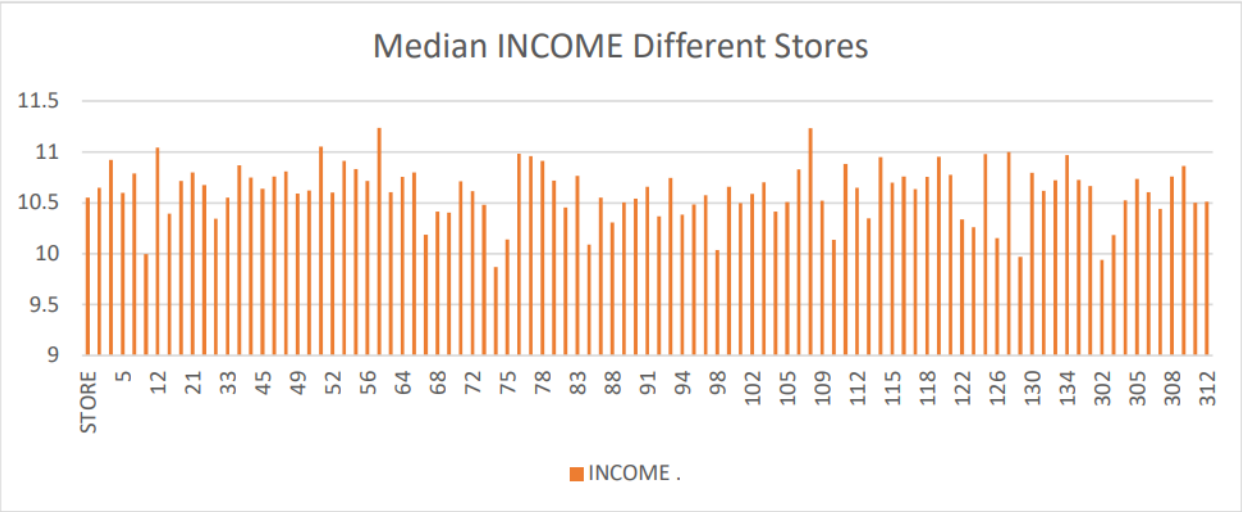
Store Code	Sum of GROCERY	Sum of DAIRY	Sum of FROZEN	Sum of BOTTLE	Sum of MEAT	Sum of MEATFROZ	Sum of FISH	Sum of PROMO	Sum of PRODUCE	Sum of BULK	Sum of SALADBAR	Sum of FLORAL	Sum
0	57056	14677	10719	0	17523	0	710	0	13108	1563	869	707	
1	24325.63	5698.45	3677.99	12.8	6839.53	670.17	587.26	94.12	4938.27	657.97	261.65	211.94	
2	56973550.59	12156170.94	8002733.99	1042.38	13945632.37	914000.55	1902656.84	354332.76	16109339.36	429527.34	821353.64	1398544.3	64
3	0	0	0	0	0	0	0	0	0	0	0	0	
4	19479836.94	3035574.58	2563399.64	7179.51	5495732	292916.54	651293.92	151022.66	4941003.94	204587.75	18617.36	329187.47	22
5	82416716.74	16072907.23	12286261.56	18656.29	20440463.97	1739659.82	3157550.58	437510.87	20707755.61	2892369.13	1679859.95	1789120.72	92
6	0	0	0	0	0	0	0	0	0	0	0	0	
7	0	0	0	0	0	0	0	0	0	0	0	0	
8	105216042.2	19560836.91	15558883.26	-379.33	27059501.23	2349356.4	2362416.3	550051.24	21789575.44	3237831.36	1261578.8	1845382.98	123
9	63648369.29	13250678.23	9706555.45	7685.34	16209543.75	1159221.88	2610262.52	394778.14	18523349.92	2981199.02	2299163.68	1633363.44	76
10	3483643	594573	425105	0	842489	0	77513	9505	702462	0	0	34010	
12	74943358.04	15820007.04	11524995.1	2132.19	17022384.01	1776635	2476040.01	351850.09	20550730.2	2719576.21	1365322.11	1573634.66	70
14	70375743.85	14423698.55	10768335.22	1594.28	15161039.82	1082311.91	1917622.14	338175.25	19098420.9	1043915.44	1733558.43	2021861.05	73
16	3558868	613325	478829	0	908768	0	86292	15856	737531	0	0	41894	
18	96673975.62	18088429.11	13390134.97	10966.84	23711982.54	1803184.66	2433443.01	489204.65	21798136.67	2764367.79	1113879.89	1765708.71	117
19	11966970.24	1417598.53	1121724.49	35.75	2478387.23	146690.31	655625.39	60029.38	2673458.81	120379.88	2524.02	91760.74	11
21	69986543.48	13086895.64	10481303.89	-12445.61	15793919.05	1949710.51	1550756.92	408932.39	13972954.92	2158103.24	594810.22	1077974.41	83
25	11479543.73	1680518.06	912997.55	2464.48	4148876.19	155732.2	230721.27	71628.7	2433685.42	139800.52	665.02	60695.64	14
28	50170475.47	10844210.37	8301303.41	1451.71	12021005.69	924185.17	1147145.77	247285.61	12356726.88	1183723.9	81852	1241216.6	51
32	100962802.4	20490671.1	14719502.32	10510.19	23014566.32	2123160.04	2953291.61	514487.37	28036047.81	3681720.42	3272855.79	2860609.36	12
33	57030265.7	14251539.39	9976360.42	4709.66	10684798.08	969876.56	1756802.53	207327.11	19184830.54	288046.5	1209883.5	1236490.82	56
39	12669515.92	1974382.97	1296338.45	-181.71	4542208.25	186632.64	586702.71	82611.8	3063457.03	276073.24	2008.47	93361.96	14
40	78407300.9	13838707.97	10546652.53	-19238.6	19614233.54	2049701.88	1546752.96	525324.88	14334633.49	2376585.12	697623.37	1155920.21	90
44	82860985.15	17511754.66	13890861.58	-13687.06	16048154.64	1989871.38	2208059.24	458605.22	22673657	2399007.66	1042754.86	1718172.44	8
45	40092275.69	8006051.42	5879364.35	951.98	10283088.95	779245	850697.33	223914.51	8983924.06	978547.28	37475.68	700841.03	41
46	25985769.09	3330672.05	2512021.07	6221.65	6177245.25	306073.27	636424.26	140523.32	5482562.15	268846.91	42658.42	347742.18	25
47	57116573.23	11141851.54	8433552.45	20369.01	16268614.83	974588.6	1680864.88	332568.69	12314304.82	1686834.91	53415.88	745051.11	55
48	43237409.48	8611488.28	6261364.68	30.5	10087630.9	771391.82	1115726.56	239912.47	11149992.15	363853.96	115610.5	1125454.12	38
49	34634788.26	6571371.98	4899158.01	-727.13	8825399.91	559825.85	692308.18	220542.35	8703395.61	610211.83	33822.82	291004.28	32
50	29832271.65	5393063.63	3991966.89	10315.99	7180120.67	503865.31	309939.21	202979.6	6274914.91	981863.92	492341.99	457173.77	30
51	64926705.86	12134093.03	9471654.51	9344.99	14223066.51	1196322.18	1570425.42	447955.83	15454205.02	2272079.47	1032794.4	1607399.13	75
52	77726686.51	16471856.15	12227982	10667.79	16356603.41	1476605.45	3488727.87	426628.84	23799566.77	2948795.99	3760417.76	2935744.72	8
53	58973692.22	11980781.15	8436034.23	-2295.09	10597396.57	706392.71	1705669.64	308364.01	14717387.47	558740.95	540419.52	1035617.27	40
54	51090257.87	11737364.72	8612661.97	-9793.6	10837430	958701.22	1288168.05	278276.47	13131552.57	661770.39	867857.48	1151136.88	49
55	12373255.31	2144950.07	1688439.66	418.18	2638986.19	257694.7	230792.88	120821.77	2220695.19	278346.13	7125.88	121815.45	11



Explanation:This question helps DFF identify which product categories are driving the highest revenue in each store. It can guide inventory management and promotional strategies to maximize profits.

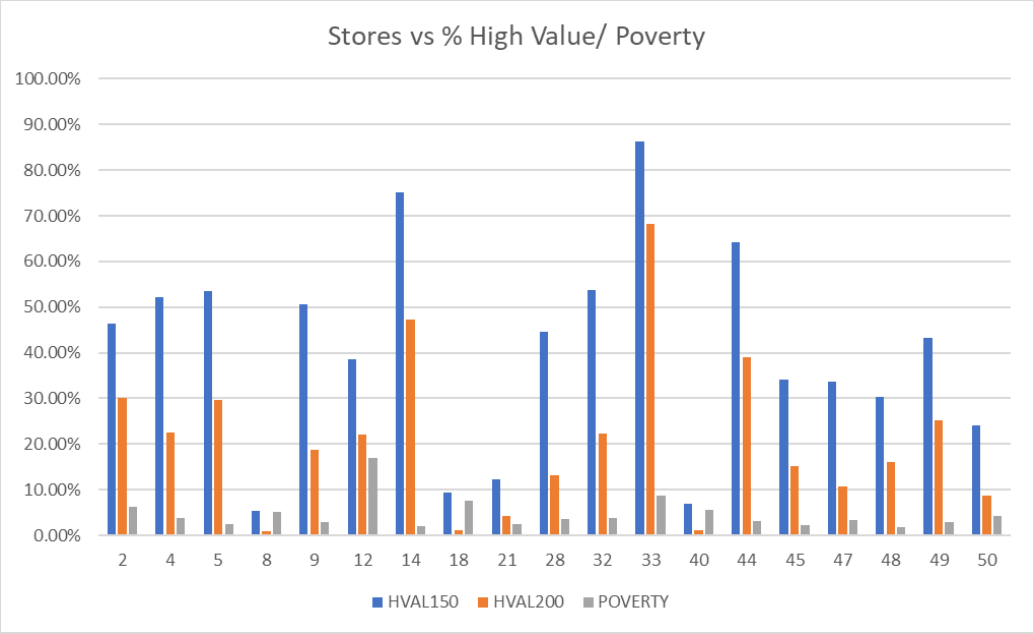
6. What impact does the socioeconomic status of a region have on the choice of retail locations and strategies?

Explanation: By delving into customer behaviors and preferences, companies can finely adjust their product offerings, pricing structures, and promotional activities, all with the goal of optimizing sales performance. These insights facilitate the development of precise advertising campaigns and individualized customer interactions. Furthermore, data-supported decisions about inventory management can mitigate problems such as stock shortages and overstock issues. In sum, harnessing this data can significantly boost revenue figures and enhance customer contentment, leading to a prosperous business outcome.



7. Which stores attract people who earn below the poverty line and have high value income thresholds?

Explanation: This data would help DFF increase/decrease the prices of certain products exclusively at certain stores to strategically increase profit margins. Also, this data would help DFF understand which stores have what proportion of high value customers and target those customers with higher margin expensive products/categories.



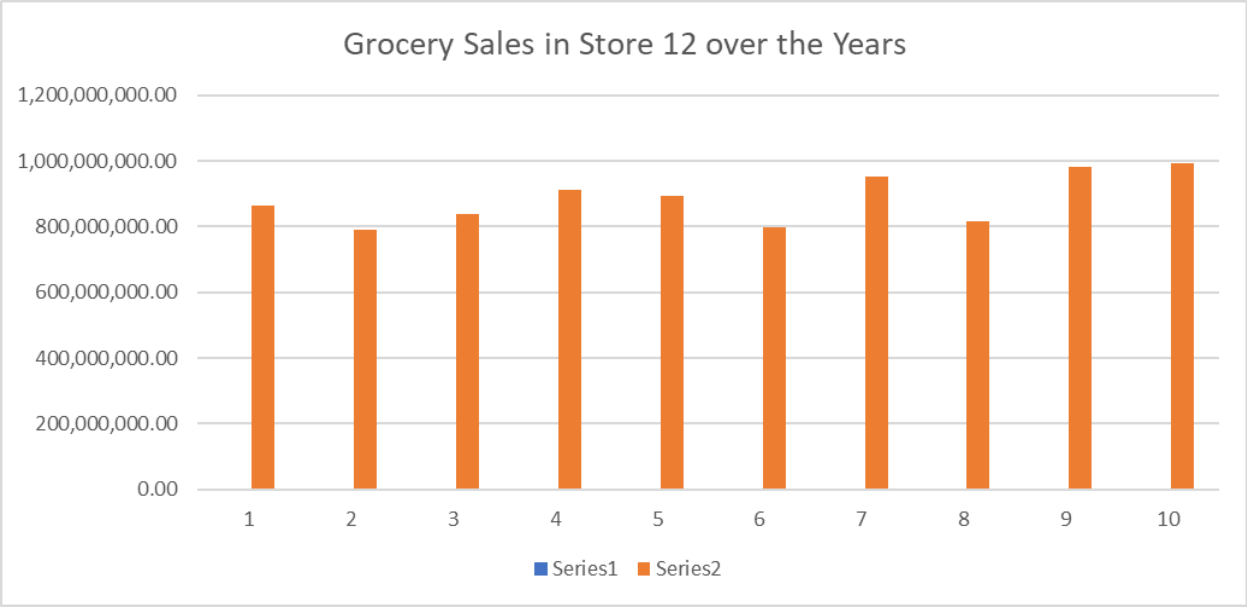
8. What is the surge in categories of products like Wine used for celebrations in the holiday season (over the years)?

Explanation: This data would help us understand the peak demand for a few categories of products during the holiday period to meet those demands. This insight would help us manage inventory and supply chains better to extract higher profits during the busier holiday season.

The chart data shows that sales for wine were much higher in the month of April. The Easter festival in the month could be a contributing factor to this trend.

A	B	C	D	E	F	G
STORE	DATE	Month	**Monthly Sales of Wine**	Day	Total Sa	WINE
4	1/1/1990	April	**2192.8**	Monday	15584.56	103.97
4	1/2/1990	April		Tuesday	24005.29	54.05
4	1/3/1990	April		Wednesda	25520.82	107.43
4	1/4/1990	April		Thursday	29980.31	102.34
4	1/5/1990	April		Friday	30749.51	190.36
4	1/6/1990	April		Saturday	44507.15	231.52
4	1/7/1990	April		Sunday	30231.25	82.18
4	1/8/1990	April		Monday	24911.17	74.87
4	1/9/1990	April		Tuesday	20586.91	98.14
4	1/10/1990	April		Wednesda	20875	67.48
4	1/11/1990	April		Thursday	26992.01	163.96
4	1/12/1990	April		Friday	32337.7	205.64
4	1/13/1990	April		Saturday	43881.43	262.98
4	1/14/1990	April		Sunday	29333.12	110.63
4	1/15/1990	April		Monday	25568.06	109.88
4	1/16/1990	April		Tuesday	21796.75	123.83
4	1/17/1990	April		Wednesda	23526.86	103.54
4	1/18/1990	May	**3038.5**	Thursday	32277.29	165.95
4	1/19/1990	May		Friday	39908.34	168.11
4	1/20/1990	May		Saturday	45742.96	224.26
4	1/21/1990	May		Sunday	32509.9	182.9
4	1/22/1990	May		Monday	27543.94	105.35
4	1/23/1990	May		Tuesday	22976.75	89.49
4	1/24/1990	May		Wednesda	28875.76	193.18
4	1/25/1990	May		Thursday	32777.66	209.12
4	1/26/1990	May		Friday	42565.68	272.57
4	1/27/1990	May		Saturday	55988.46	231.27
4	1/28/1990	May		Sunday	35958.38	218.88
4	1/29/1990	May		Monday	26469.55	129.37
4	1/30/1990	May		Tuesday	24016.8	128.11
4	1/31/1990	May		Wednesda	23063.88	124.02
4	2/1/1990	May		Thursday	33561.34	167.36
4	2/2/1990	May		Friday	40105.85	170.29
4	2/3/1990	May		Saturday	49576.33	258.27
4	2/4/1990	May		Sunday	31004.66	150.91
4	2/5/1990	May		Monday	24909.31	77.89

9. What are trends in categories such as grocery or dairy over the years?



Product category and product sales data over the years inform marketing and sales strategies. It aids in demand forecasting, optimizing inventory, setting prices, targeting customers, and refining promotions. Analyzing trends, product placement, and customer behavior enhances the shopping experience and ultimately drives profitability and efficiency in the grocery retail industry.

10. Patterns in customer count or footfalls throughout the year?

Explanation: Analyzing trends in customer foot traffic across different seasons can offer important insights for marketing and sales strategies. Understanding when customer numbers are at their highest or lowest helps retailers plan promotions and stock inventory more effectively. This data is valuable for enhancing customer experiences and optimizing resource allocation. Ultimately, it enables businesses to better align marketing and sales strategies with customer behavior, improving sales and customer satisfaction.

Prioritizing Business Questions

1. What are the top-performing product categories in terms of sales revenue for each DFF branch over the years?
    a. Rationale: This question directly impacts revenue and profitability, making it a top priority. Identifying high-performing categories can guide inventory and marketing strategies.
2. Which DFF stores have experienced significant changes in sales performance over the years, and what actions have been taken to address these changes?
    a. Rationale: Addressing changes in sales performance is crucial for business sustainability. Learning from these changes and implementing successful actions can benefit the entire chain.
3. What is the surge in categories of products like Wine used for celebrations in the holiday season (over the years)?
    a. Rationale: Understanding holiday season trends is essential for optimizing inventory and promotions during peak demand periods, ensuring profitability.
4. Which stores attract people who earn below the poverty line and have high-value income thresholds?
    a. Rationale: Targeting specific customer segments can lead to tailored marketing and pricing strategies, optimizing profit margins.
5. What do the sales patterns for bakery products look like over a span of three years, categorized into low, medium, and high-price tier stores across multiple cities?

    a. Rationale: Analyzing sales patterns across different pricing tiers provides insights for product assortment and marketing strategies, particularly for bakery products.
6. What impact does the socioeconomic status of a region have on the choice of retail locations and strategies?
    a. Rationale: Understanding how location and strategy are influenced by socioeconomic status is vital for optimizing store placement and marketing efforts.
7. What are the bakery sales trends in a 3-year succession range across the low, medium, high-price tier categorization of stores in various cities?
    a. Rationale: Bakery product sales trends across pricing tiers and cities can inform inventory management and marketing campaigns.
8. What is the contribution and distribution of different categories to the sales of the 5 most successful stores?
    a. Rationale: Analyzing product category contributions to successful stores helps identify strategies to replicate in other locations, driving sales and profits.
9. What are store-wise demographics across the total company sales for each product in the same geographic area, such as a city?
    a. Rationale: Understanding the contribution of demographics to store sales in a specific location allows for better-targeted marketing and inventory management.
10. What are the top-performing product categories in terms of sales revenue for each DFF branch over the years?
    a. Rationale: While important, this question has a lower priority as it focuses on product performance within branches. Other questions offer more comprehensive insights into trends and demographics, which can be applied to optimize sales strategies.

Section 3: Independent Conformable Data Mart Design (Kimball)

1.1 Introduction

In this project, we have designed a data warehousing infrastructure for Dominick Fine Foods. With this design, we are creating robust and effective data marts, utilizing the Kimball approach and a dimensional model schema in the HOLAP fashion. The role of SQL Server is to help our data repository. It increases its dependability and effectiveness. Our design approach emphasizes a bottom-up strategy, with a strong focus on building versatile data marts. We employ the STAR schema method, making it easier to navigate and access data within the warehouse, especially for intensive data queries. With this approach, we've designed our data marts to effectively address our business needs.

Overview of Kimball's Methodology

Founded by Ralph Kimball, this methodology is known for creating, configuring, and maintaining data warehouses. It supports a bottom-up technique, highlighting the importance of data warehousing and the role of various data marts in a business data warehouse.
Following are the key steps of Kimball's Methodology:

Step 1: Picking a Business Function to Model

Kimball's approach revolves around comprehending essential business processes and the questions a data warehouse should answer. It starts with identifying and gathering data from various sources, including transactional systems, and then loading this data into a staging area using ETL (Extract, Transform, Load) software.

Step 2: Select the grain

In this context, "grain" refers to the level of detail at which data should be stored in the primary fact table. Kimball advises going for the finest level of granularity, which means data cannot be divided further. For instance, if Dominick's Finer Foods wants detailed data, it'd be things like sales of specific products in specific stores on specific days. Starting at this level ensures data is available for in-depth analysis.

Step 3: Identify the dimensions that relate to each fact table row

Selecting dimensions becomes simpler when you've chosen the right level of granularity. Dimensions should align with different business processes. The aim is to create a comprehensive set of dimensions covering all possible descriptions for designing fact tables.

Step 4: Choosing Numeric Data for Fact Tables

Fact tables get filled with numerical data relevant to specific business operations and questions. This data must match the chosen granularity for accurate querying and analysis. If there's a mismatch in granularity, a separate fact table might be needed.

3.2 Importance

Kimball's Methodology is a crucial approach when it comes to creating separate sections of data. It has several advantages that make it an excellent choice for businesses:

Efficiency and Speed: This approach allows you to analyze data quickly by focusing on what your business does and using star schemas.

Lower Initial Costs: When you start with Kimball, you plan your data warehouse at the beginning, and the cost remains the same as you go along. It's budget-friendly.

Quick Setup: Data setups following Kimball's methodology are super quick. Everything is ready to go with pre-made data models and structures.

Generalist Team: You don't need a team of specialists. A group of all-around experts can make it work.

Focused Data Integration: With Kimball, you get data set up specifically for your business area. It's like a tailored suit for your data, ensuring you receive the most accurate analysis.

Data Matrix

Data Mart	Dimension			
	Demo_Dim	Category_Dim	Date_Dim	Store_Dim
Sales_Fact	x	x	x	x

Section 4

4.1 Dimensional Modeling

Dimensional Tables

a. Product Dimension (productDim)

The ProductDim Table consists of various attributes related to the products sold by Dominick's finer Foods. This will be useful in answering the questions about various products and product categories.

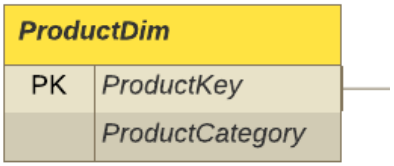


Table Attributes

ProductKey: This is the surrogate key used as the primary key for this table.
Category: Describes the category under which the product falls.

b. Demographic Dimension (DemoDim)

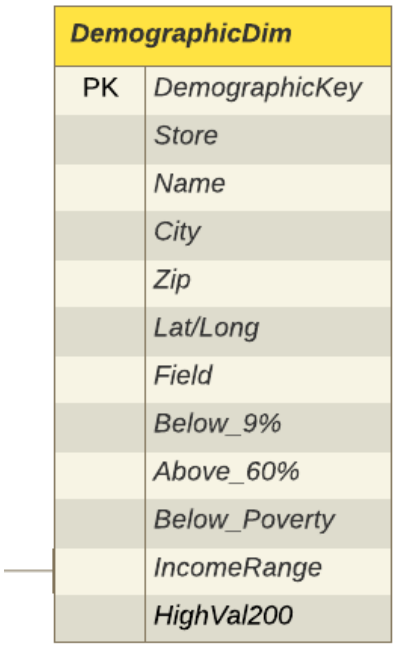


Table Attributes

DemographicKey: This is the surrogate key used as the primary key for this table.
Store: This is the Store ID for the store in Dominick's Finer Food Chain
Name: This is the Store name
City: This is the city in which the store is located
Zip: This is the ZIP code in which the store is located
Lat/Long: This is the information about the Lat/Long of the store
Field: This is the field information about the store
Below_9%: Defines the percentage of customers below age of 9
Above_60%: Defines the percentage of customers above the age of 60
Below_Poverty: Defines the percentage of customers below poverty level
Income_Range: Defines the income range of the customers
HighVal200:

c. TimeDimension Table (TimeDim)

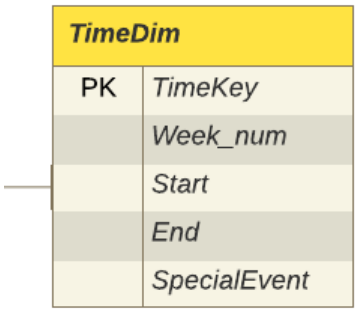


Table Attributes

TimeKey: This is the surrogate key used as the primary key for this table.
Week_num: This is the week number for the data
Start: This is the start date of the week
End: This is the end date of the week
SpecialEvent: The information about the special events during the year

d. Store Dimension Table (StoreDim)

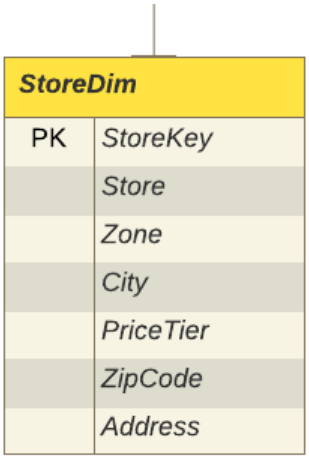


Table Attributes

StoreKey: This is the surrogate key used as the primary key for this table.
Store: This is the Store ID for the store in Dominick's Finer Food Chain
Zone: This is the zone area of the store

ZipCode: This is the ZIP code for the store
City: This is the name of the city in which the store is located
PriceTier: This is the derived price tier of the products in each store
Address: This is the address in which the store is located

e. SalesFact Table

The SalesFact Table consists of the aggregated data for sales of each product category sold by Dominick's Finer Foods. Since we are using a Star Schema, the foreign keys from various dimensional tables are used as composite primary key for the SalesFact table.

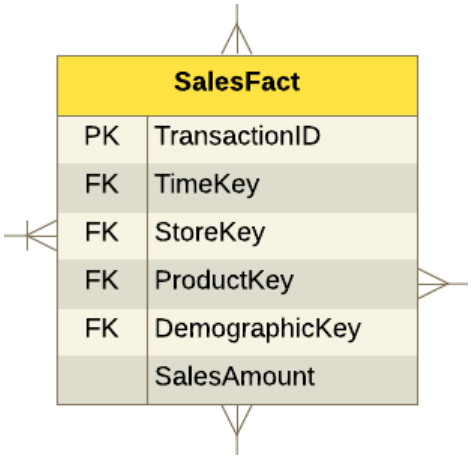


Table Attributes

TransactionID: Used as a unique identifier and a surrogate key for the SalesFact table.

TimeKey: This is the primary key for the DateDim dimension table and used as a foreign key for the fact table.

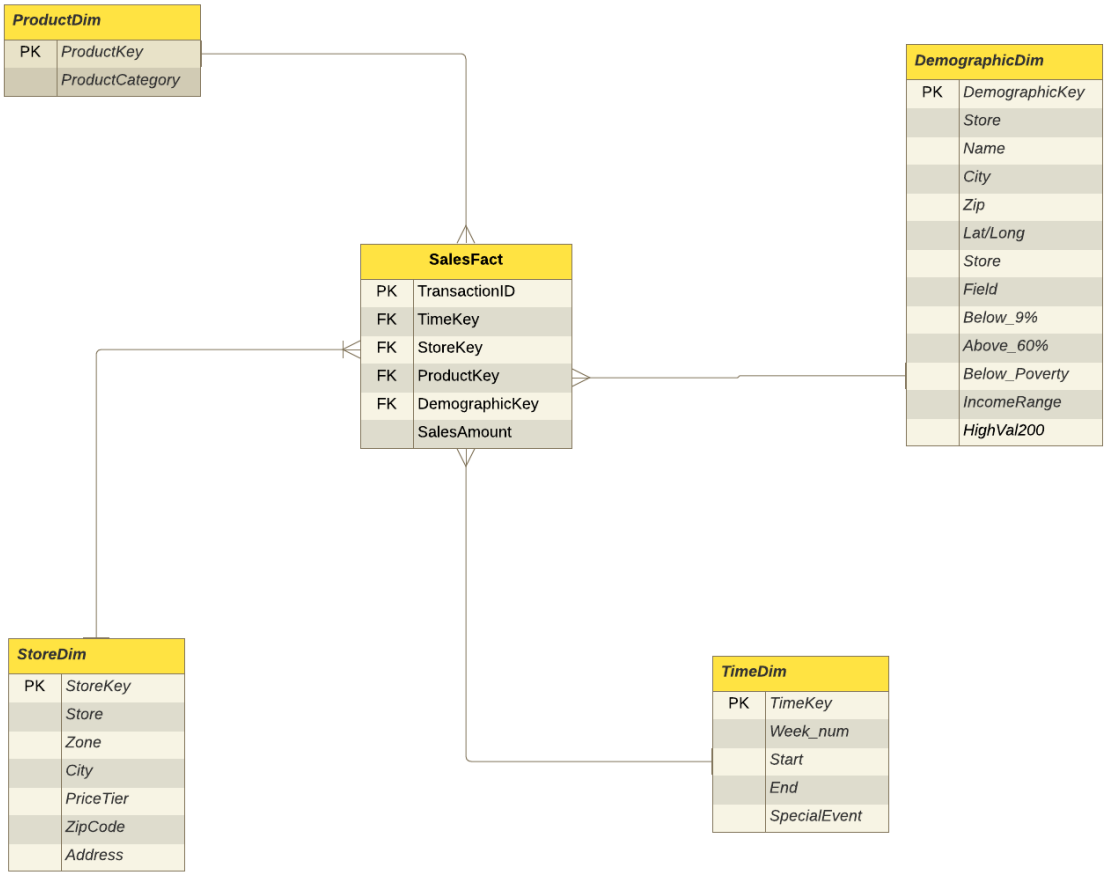
StoreKey: This is the primary key for the StoreDim dimension table and used as a foreign key for the fact table.

ProductKey: This is the primary key for the ProductDim dimension table and used as a foreign key for the fact table.

DemographicKey: This is the primary key for the DemographicDim dimension table and used as a foreign key for the fact table.

SalesAmount: This is the aggregate sales across all product categories for each store in the Dominick's Finer Foods retail chain.

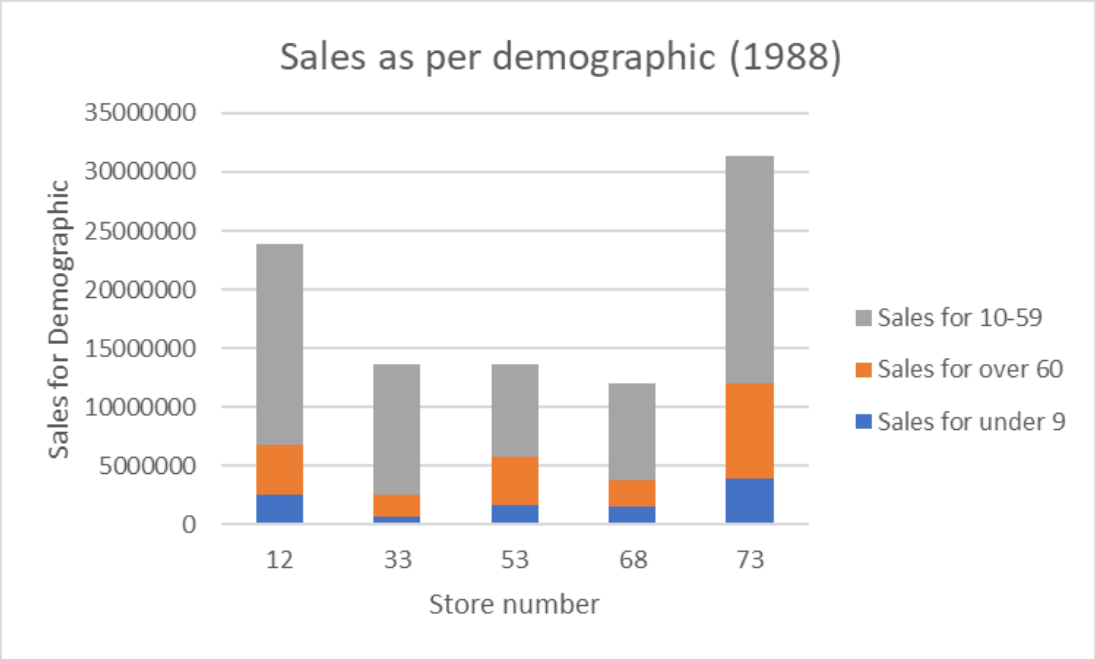
4.2 Star Schema



4.3 Selected Business Questions and logic

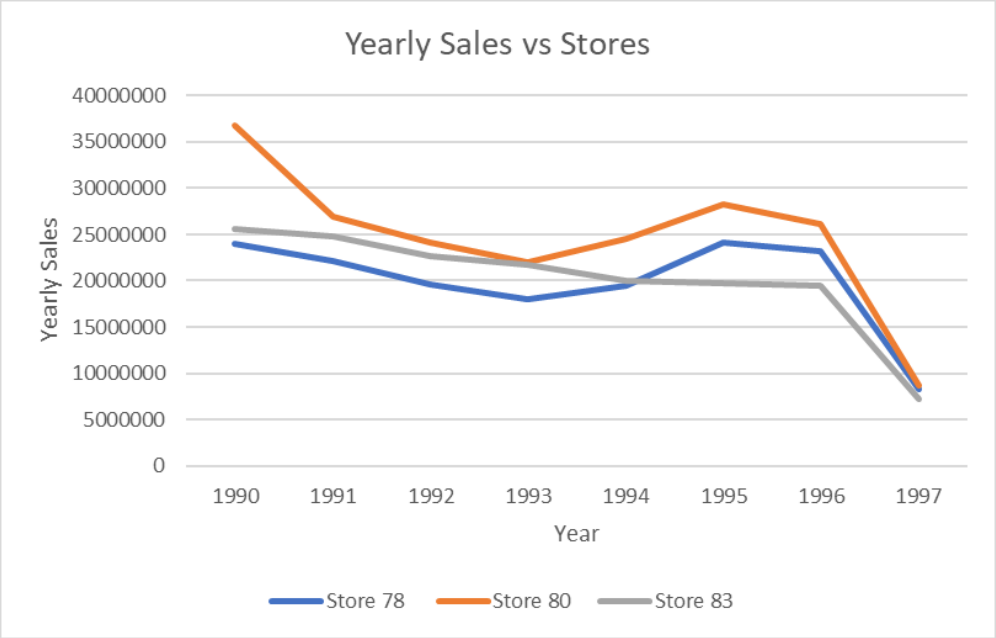
1. What are store wise demographics across the total company sales for each store?

Reason: This business question can be solved by analyzing contribution of different demographics in terms of age of population to the total company sales across each DFF store.
The information for the demographic of customer age can be obtained from the Age column in the demographic dimension and sales can be obtained from the SalesFact dimension. This analysis will help us understand the contribution of different age groups to sales in different DFF stores.



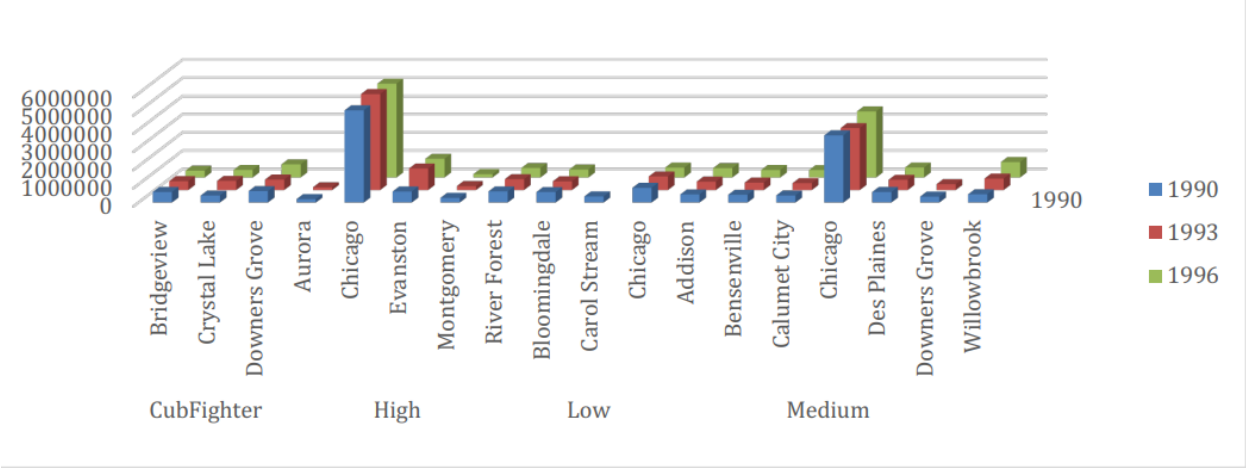
2. Which DFF stores have experienced significant changes in sales performance over the years, and what actions have been taken to address these changes?

Reason This business question can help us understand the trend in sales over the years for different DFF stores.
This business question can be answered by analyzing the sales across each stores over the years and gauging the trends in the sales according to location. The sales data can be aggregated over the years from Week_num column in TimeDim and using sales amounts from SalesFact table. A pattern like good sales performance in most stores in an year with an exception to a few stores can help us understand the room for improvement in those DFF stores.



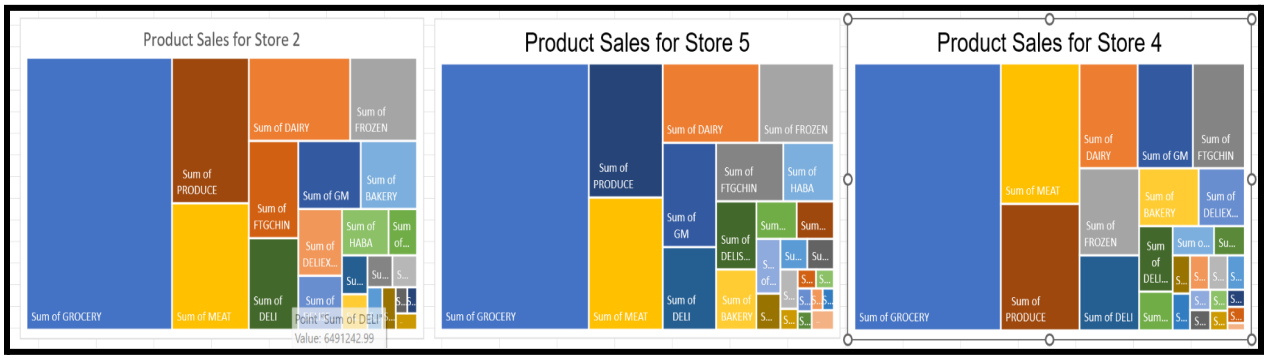
3. What do the sales patterns for bakery products look like over a span of three years, categorized into low, medium, and high-price tier stores across multiple cities?

   Reason: This business question can help us understand the sales of bakery products across different DFF stores categorized into different tiers according to their geographic location. The store tier can be obtained from PriceTier column in the StoreDim. The sales for those stores in different Tier city can than be obtained from SalesAmount column in the SalesFact table for bakery products.



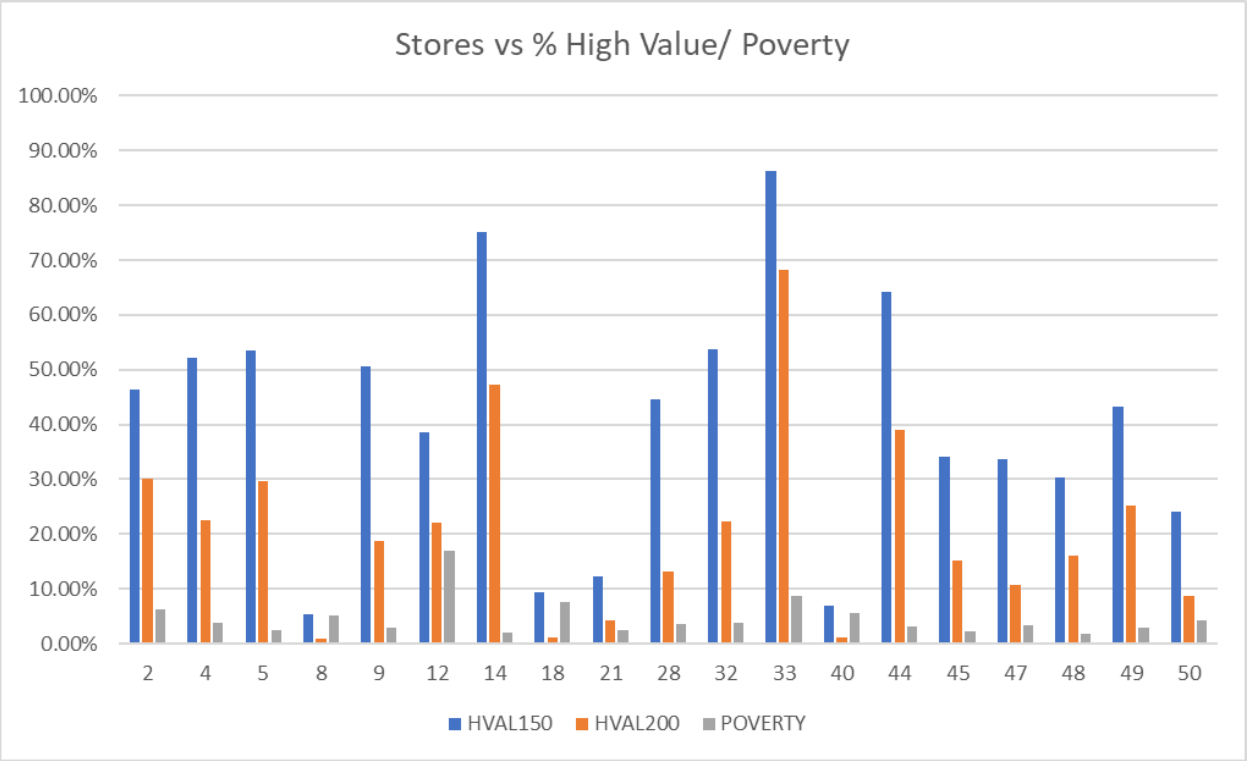
4. What are the top-performing product categories in terms of sales revenue for each DFF branch over the years?

   Reason: This business question can be answered by analyzing the total sales for each category of product over time period. Sales for each category are aggregated across all the DFF stores. This analysis can be performed by fetching ProductCategory from ProductDim and then gathering sales data for each of those product categories from SalesAmount column in the SalesFact table in the star schema.



5. Which stores attract people who earn below the poverty line and have high value income thresholds?

   Reason: The table for the income and proportion of High Value shoppers and below poverty customers can be obtained from the Demographic dimension. The Income can be obtained from the IncomeRange column in DemographicDim which can be used to aggregate the sales of each demographic category of shoppers across the total sales from the SalesFact table across different DFF stores.



4.4 Source to Data Staging Table

Source Table	**Source Data Field**	**Mapping**	**Staging Table Type**	**Staging Table Name**	**Attribute**
UPCXXX.csv	COM_CODE	Surrogate key used to connect product values in SalesFactTable	Staging table	**ProductStg**	ProductKey
CCOUNT.csv	PRODUCT_CATEGORY	Direct Mapping			Category
		Surrogate key used to connect demographic values in SalesFactTable		**DemographicStg**	Demographic Key
DEMO.csv	Store	Direct Mapping	Staging		Store
DEMO.csv	Name	Direct Mapping	Staging		Name

DEMO.csv	City	Direct Mapping	Staging		City
DEMO.csv	Zip	Direct Mapping	Staging		Zip
DEMO.csv	Field	Direct Mapping	Staging		Field
DEMO.csv	Below_9%	Direct mapping	Staging		Below_9%
DEMO.csv	Above_60%	Direct mapping	Staging		Above_60%
DEMO.csv	Below_Poverty				
DEMO.csv	LAT, LONG	Derive from the LAT and LONG attributes	Staging		State
DEMO.csv	INCOME	Direct mapping	Staging		IncomeRange
DEMO.csv	HVAL200	Direct Mapping	Staging		HighVal200
DEMO.csv	INCOME, INCSIGMA	Using the STD deviation from INCSIGMa and INCOME values, define the percentage of people below level of poverty	Staging		Below_Poverty
		Surrogate key used to connect date values in SalesFactTable	Staging	**TimeStg**	TimeKey
CCOUNT.csv	Week_Num	Extract day from Date column	Staging		Day
CCOUNT.csv	Start	Derived from date column	Staging		Day
CCOUNT.csv	End	Derived from date column	Staging		Day
CCOUNT.csv	Special Event	Derived from date column	Staging		Date
		Surrogate key used to connect store values in SalesFactTable	Staging	**StoreStg**	StoreKey
CCOUNT.csv	STORE	Direct mapping	Staging		StoreID

DEMO.csv	ZIPCode	Direct mapping of ZIP code	Staging		Location
DEMO.csv	CITY	Direct mapping of City	Staging		City
DEMO.csv	Address	Direct mapping of Address	Staging		Address
DEMO.csv	PriceTier	Derive from the PriceTier	Staging		PriceTier
DEMO.csv	Zone	Direct mapping of Zone	Staging		Zone
			FactTable	**SalesStg**	TransactionID
TimeDim	TimeKey	Foreign Key that refers to Surrogate key in DateDim dimension table			TimeKey
StoreDim	StoreKey	Foreign Key that refers to Surrogate key in StoreDim dimension table			StoreKey
ProductDim	ProductKey	Foreign Key that refers to Surrogate key in ProductDim dimension table			ProductKey
DemographicDim	Demographic Key	Foreign Key that refers to Surrogate key in DemographicDim dimension table			Demographic Key
CCOUNT.csv	Sales Amount	This is the aggregate sales for each store in the Dominick's food chain summed across all product categories			SalesAmount

4.5 Staging table to Presentation server (Data warehouse)

Source	**Source Data**	**Mapping**	**Staging**	**Staging Table**	**Attribute**

Table	Field		Table Type	Name	
ProductStg	ProductKey	Direct Mapping	Staging table	**ProductDim**	ProductKey
	ProductCategory	Direct Mapping			ProductCategory
DemographicStg	DemographicKey	Direct Mapping		**DemographicDim**	DemographicKey
DemographicStg	Store	Direct Mapping	Staging		Store
DemographicStg	Name	Direct Mapping	Staging		Name
DemographicStg	City	Direct Mapping	Staging		City
DemographicStg	Zip	Direct Mapping	Staging		Zip
DemographicStg	Field	Direct Mapping	Staging		Field
DemographicStg	Below_9%	Direct Mapping	Staging		Below_9%
DemographicStg	Above_60%	Direct Mapping	Staging		Above_60%
DemographicStg	Below_Poverty	Direct Mapping			
DemographicStg	LAT, LONG	Direct Mapping	Staging		State
DemographicStg	INCOME	Direct Mapping	Staging		IncomeRange
DemographicStg	HVAL200	Direct Mapping	Staging		HighVal200
DemographicStg	INCOME, INCSIGMA	Direct Mapping	Staging		Below_Poverty
TimeStg	TimeKey	Direct mapping	Staging	**TimeDim**	TimeKey
	Week_Num	Direct mapping	Staging		Day

	Start	Direct mapping	Staging		Day
	End	Direct mapping	Staging		Day
	Special Event	Direct mapping	Staging		Date
StoreStg	StoreKey	Direct mapping	Staging	**StoreDim**	StoreKey
	STORE	Direct mapping	Staging		StoreID
	ZIPCode	Direct mapping	Staging		Location
	CITY	Direct mapping	Staging		City
	Address	Direct mapping	Staging		Address
	PriceTier	Direct mapping	Staging		PriceTier
	Zone	Direct mapping of Zone	Staging		Zone
		Direct mapping	FactTable	**SalesFact**	TransactionID
TimeDim	TimeKey	Direct mapping			TimeKey
StoreDim	StoreKey	Direct mapping			StoreKey
ProductDim	ProductKey	Direct mapping			ProductKey
Demographic Dim	Demographic Key	Direct mapping			Demographic Key
CCOUNT.csv	Sales Amount	Direct mapping			SalesAmount

4.6 Physical Design

a. Data Aggregation Plan

It is critical that we maintain thorough data records in order to meet the criteria of our queries, particularly those relating to sales patterns and product categories. This gives us more freedom in data aggregation and enables for more detailed analysis. For instance, when going into bakery product sales trends across multiple years, individual transaction data is vital. However, we'll compute necessary aggregates in advance, like quarterly or yearly summaries to optimize performance.

b. Data Retrieval Plan

With the large amount of data and the relationships between our tables, data indexing becomes essential. Columns with high uniqueness and primary keys will utilize b-tree indexing to ensure quick data access. Also, because our structure includes several joins between our fact and dimension tables, indexing foreign key columns will improve query performance.

c. Data Standardization Plan

In a setting with a lot of data like this, a simplified approach is essential. We'll use a systematic name scheme: dimension tables will have a "DIM" postfix, whereas the fact table will have a "FACT" postfix. This helps in quick identification and lowers the possibility of error. Mandatory fields will be carefully looked into to ensure that no null values exist, and unique primary keys will remain intact to ensure data integrity. To maintain data reliability, consistent data types across tables will be applied.

d. Data Storage Plan

As our data environment is built to manage both historical records and periodic new data, storage planning is important. We've divided our data environment into marts, each corresponding to a different aspect of business, such as sales, stores, product categories, and customer insights. Regular updates, particularly for sales figures and customer data, may be required on a weekly basis, requiring a simplified ETL method. We also have to be prepared for the addition of new product lines or locations. Our storage strategy will meet our present demands while also being open to future development.

Section 4: Data Cleaning and Integration

ETL Development Plan

5.1 Data Processes

ETL will be one of the most important elements of this project. We have to make sure that the data we need to answer our business questions is properly aggregated, transformed and made appropriate for analysis. Due to the volume and complexity of the data in this project, we need a standardized and structured ETL plan for a successful and working data warehouse.

The three phases we will use are as follows:

Phase 1: Data Extraction

In this phase, we need to identify and source the data which is relevant to our particular business questions. The mapping table from Report 2 will help us greatly in this.
We will extract data from various tables, based on our star schema. The most important part of this will be to ensure that all the columns from the mapping table are included.

Phase 2: Data Transformation

Our mapping table has shown us that there is a need for data quality and consistency. It is the transformation phase, where we make sure of data quality and consistency. We will use SQL queries. Transformation and extraction rules to make sure that data is consistent and uniform. Areas to address include missing values, inconsistent naming conventions, etc.

Phase 3: Data Loading

Now that we have the data which is clean and consistent, we will load the clean data into our final data marts. The data marts are built to answer the business questions and support the reporting needs. To make sure that data is correctly associated and relationships between tables are correctly maintained, we will make use of joins and surrogate keys.

We will maintain strict data integrity and format standards. We will use SSIS to create packages for each stage of ETL. The process is meant to be iterative and we aim to have a robust data warehouse ready after our ETL process.

1.1 Data Extraction

In the extraction phase, we have extracted data from the source files which were in CSV format which is compatible with our ETL tools. After that, based on the star schema, we extracted specific columns, which we need to answer the business questions, loading them SSIS using the Import/Export Wizard. The foundation for our staging tables has been set, which helps in ensuring the integrity of all the data.

1.2 Data Transformation

The transformation phase is where we will clean our extracted data and make sure that it is consistent. We will do this by applying some transformation rules like:

Data Conversion: We will convert all data into standardized and appropriate formats.

Removing Garbage values: Remove any data which is not contributing to the final output needed.

Removing or Converting NULL values: Remove NULL values or replace them with default values.

Create Surrogate Keys: Surrogate keys are created to ensure referential integrity amongst dimension and fact tables.

Derived Columns: Creating new columns based on the existing data to support detailed analysis.

1.3 Data Loading:

Data loading is the final phase where data is loaded into the data marts for in depth analysis. Before loading the data into the data marts, data consistency and integrity will be checked again. The data is loaded into the data mart with a focus on maintaining the structure and relationships as per the mapping table.

5.2 Data Sources and Target Data:

Following data sources(files) were used from DFF:

Data Source	**Source Files**
Demographics	Demo.csv
Customer Count	Ccount.csv
Store data	Stores.csv
Week data	Weektables.csv

Target Data for data warehouse:

Data Source	**Source File**	**Staging Area tables**	**Datamart tables**
Demographics	Demo.csv	Demo_stg	DemoDim, SalesFact Table
Customer Count	Ccount.csv	Ccount_stg	ProductDim, SalesFact Table
Week data	Weektables.csv	Week_stg	TimeDim, SalesFact Table
Store data	Store.csv	Store_stg,	StoreDim, SalesFact Table

5.3 DATA MAPPING FOR DATA ELEMENTS FROM SOURCE TO STAGING AND STAGING TO DATA WAREHOUSE

Source Data to Staging Area Table:

Source File	Source Attributes	Staging Table	Staging Table Attributes	Mapping Function
Weektable.csv	WeekNumber	Week_stg	Week_num	Direct copy
	Start		Start	Direct copy
	End		End	Direct copy
	SpecialEvents		SpecialEvents	Direct copy
Stores.csv	Store	Store_stg	Store	Direct copy
	City		City	Direct copy
	Address		Address	Direct copy
	Zipcode		Zipcode	Direct copy
	Zone		Zone	Direct copy
	Price_Tier		Price_Tier	Direct copy
Ccount.csv	STORE	Ccount_stg	STORE	Direct copy
	DATE		DATE	Direct copy
	WEEK		WEEK	Direct copy
	CONVFOOD		CONVFOOD	Direct copy
	PRODUCE		PRODUCE	Direct copy
	GROCERY		GROCERY	Direct copy
	MEATFROZ		MEATFROZ	Direct copy
	BAKERY		BAKERY	Direct copy
	BOTTLE		BOTTLE	Direct copy
	CHEESE		CHEESE	Direct copy
	FROZEN		FROZEN	Direct copy
	MEAT		MEAT	Direct copy
	FISH		FISH	Direct copy
	DAIRY		DAIRY	Direct copy
	DELI		DELI	Direct copy
	MEATCOUP		MEATCOUP	Direct copy
	PHARMACY		PHARMACY	Direct copy
Demo.csv	NAME	Demo_stg	NAME	Direct copy

	CITY		CITY	Direct copy
	ZIP		ZIP	Direct copy
	STORE		STORE	Direct copy
	ZONE		ZONE	Direct copy
	Age9		Age9	Direct copy
	Age60		Age60	Direct copy
	Income		Income	Direct copy
	HVAL200		HVAL200	Direct copy
	Poverty		Poverty	Direct copy

Staging Tables to Presentation Server (Warehouse) Tables:

Staging Table	**Staging Table Attributes**	**Warehouse Table**	**Warehouse Table Attributes**	**Mapping Function**
Week_stg	Week#	DimTime	WeekNumber	Direct copy
	Start		Start	Direct copy
	End		End	Direct copy
	SpecialEvents		SpecialEvents	Direct copy
Store_stg	Store	DimStore	StoreNumber	Direct copy
	City		StoreCity	Direct copy
	Address		StoreAddress	Direct copy
	Zipcode		StoreZip	Direct copy
	Zone		StoreZone	Direct copy
Ccount_stg	CONVFOOD	DimProduct	ProdCategory	Derived attribute
	PRODUCE		ProdCategory	Derived attribute
	GROCERY		ProdCategory	Derived attribute
	MEATFROZ		ProdCategory	Derived attribute
	BAKERY		ProdCategory	Derived attribute
	BOTTLE		ProdCategory	Derived attribute
	CHEESE		ProdCategory	Derived attribute

	FROZEN		ProdCategory	Derived attribute
	MEAT		ProdCategory	Derived attribute
	FISH		ProdCategory	Derived attribute
	DAIRY		ProdCategory	Derived attribute
	DELI		ProdCategory	Derived attribute
	MEATCOUP		ProdCategory	Derived attribute
	PHARMACY		ProdCategory	Derived attribute
Demo_stg	NAME	DemoDim	Name	Direct copy
	CITY		City	Direct copy
	ZIP		Zip	Direct copy
	STORE		Store	Direct copy
	ZONE		Zone	Direct copy
	Age9		Age9	Direct copy
	Age60		Age60	Direct copy
	Income		Income	Direct copy
	HVAL200		HVAL200	Direct copy
	Poverty		Poverty	

Data Extraction Rules:

The extraction of consistent and relevant data for our data warehouse is a critical step in integrating different data sources. We concentrated on identifying and extracting data that directly supports the business questions. We used the following data sources for this project:

- Comma Separated Value (CSV) files: These were the most common file formats in our project, and they included files like Weektables.csv, Stores.csv, and Ccount.csv.
- Excel files (xlsx, xls): To extract demographic data, we used Excel files such as Demo.xlsx.

- Text files (txt): Data such as special event information was present in text format and we included that in our data extraction process.

To streamline the extraction process, all files were converted into CSV format and data was imported into SSIS using the Import/Export Wizard.

We followed the following data extraction rules:

- Uniform File Format: To streamline the extraction process, all extracted data files were converted and saved in the csv file format.
- Surrogate Key Formatting: We used a defined format to create surrogate keys after extraction which is <Dimension_Name><Key>. This helps in maintaining consistency across the warehouse
- Selective Extraction: Only the needed columns were extracted to avoid redundant data.

5.5 Data Transformation Rules:

This process is a key step after data extraction. Data transformation is done to ensure that data is clean and consistent. Data has to be accurate before loading it into the data marts. A number of data transformation rules were applied:

- Maintaining/Removing NULL Values: Our extracted data includes multiple NULL values, especially from files such as Weektables.csv and Stores.csv. In order to protect data integrity, we either removed them or replaced them with default or calculated values.
- Data Conversion: Data has been standardized to make sure that all data is uniform across all the attributes. This is done to ensure data consistency.
- Surrogate Keys: Surrogate keys were created to maintain referential integrity across dimension and fact tables. They act as unique identifiers for each record in our dimension tables.
- Removing Garbage Values: We removed any data that is not necessary pertaining to our business questions.
- Derived Columns: For more detailed analysis, we created additional columns based on existing data.

5.6 Data Aggregation

We will use data aggregation in the final phase of our ETL process to help our data warehouse run faster and more efficiently.

Currently, our data warehouse doesn't use aggregated tables, however, this can change as the amount of data rises. This is something that could concern the SalesFactTable. If we develop a summary version of this table called SalesFactAggregate, it may speed up our indexing, especially if we have a large amount of data.

The important step is deciding what type of data (for example: total sales) could be summarized, which would give us an advantage. Then we create summary tables to hold the aggregated data. The summary tables need to be updated regularly. Data aggregation makes the data warehouse faster and more efficient.

5.7 Organization of Data Staging Area

Staging Area is where we load the data initially from different sources. Here the data from different sources is stored in different staging tables, before being loaded into the data warehouse.

We created two databases for our project: one for the data staging area and one for the final data mart. The staging tables that we created are as follows:

- **Week_stg:** This staging table contains all of the data from Weektables.csv. The table contains every column from this file. While this table did not need a lot of cleansing or transformation, it is important as it contains all the data regarding dates and times.
- **Stores_stg:** This contains data from the Stores.csv file. We loaded all columns, and this table will have derived columns to include the Price_Tier information, which is critical for our data warehouse.
- **Demo_stg:** The Demo_stg table contains data from the Demo.csv file. Only selected columns were chosen. This table was cleaned and transformed multiple times to maintain data consistency.
- **Ccount_stg:** Data from the Ccount.csv file is used to fill the Ccount_stg table. Because of the nature of its data, this table required the most cleaning.

5.8 Data Extraction and Loading

5.8.1 Data Extraction

1. **Week_stg:**
    - Data was carefully transcribed into Week.txt, imported into Excel, and then produced into a CSV file (Weektable.csv) using the tab-delimited format.
    - Prior to extraction, every field was transformed into the corresponding data format.
    - The SSIS Import/Export wizard was used to map all of the data's values from the source file to the staging table.
    - In order to construct a destination for import, a connection to the Mays server (infodata16.mbs.tamu.edu) was first established. Then, the relevant file was loaded into the import wizard as a source in order to create a package to extract the data in SSIS.
    - No alterations were made to this staging table, as all columns and attributes were extracted precisely as they were.

2. **Stores_stg:**
    - Store, City, Address, ZipCode, and Zone were among the columns from the source file (Stores.csv), from which data were extracted.
    - But before that, each field was transformed into the appropriate data format.
    - The SSIS Import/Export wizard, as seen in the screenshot below, was used to map all of the data's values from the source file to the staging table.
    - In order to construct a destination for import, a connection to the Mays server (infodata16.mbs.tamu.edu) was first established. Then, the relevant file was loaded into the import wizard as a source in order to create a package to extract the data in SSIS.

3. **Demo_stg:**
    - Name, City, Zip, Store, Zone, Age60, Age9, Income, HVAL200, and Poverty were among the columns extracted from the source files' Demo.csv file.
    - Prior to extraction, every field was transformed into the appropriate data format.
    - The SSIS Import/Export wizard was used to map all of the data's values from the source file to the staging table.
    - In order to construct a destination for import, a connection to the Mays server (infodata16.mbs.tamu.edu) was first established. Then, the relevant file was loaded into the import wizard as a source in order to create a package to extract the data in SSIS.

4. **Ccount_stg:**
    - This staging table contains fields for CONVFOOD, PRODUCE, GROCERY, MEATFROZ, BAKERY, BOTTLE, CHEESE, FROZEN, MEAT, FISH, DAIRY, DELI, MEATCOUP, and PHARMACY that are pulled from the source file (Ccount.csv)
    - As this table requires the most modification, prior to extraction, every field was transformed into the appropriate data format.
    - The SSIS Import/Export wizard was used to map all of the data's values from the source file to the staging table.
    - In order to construct a destination for import, a connection to the Mays server (infodata16.mbs.tamu.edu) was first established. Then, the relevant file was loaded into the import wizard as a source in order to create a package to extract the data in SSIS.

5.8.2 Data Loading:

1. **DemoDim:**

   Using SSIS, the data Loading Procedure for the Demographic Dimension Table involves creating a staging table (Demo_stg), cleansing and transforming the data, and loading it into the DimDemo dimension table. The process includes handling data quality issues, adding a surrogate key as the primary key, implementing error handling, and deploying the SSIS package.

2. **StoreDim:**

   The data Loading Procedure for the Store Dimension Table using SSIS encompasses the creation of a staging table (Stores_stg), data cleansing and transformation, and subsequent loading into the DimStore dimension table. The process involves addressing data quality concerns, introducing a surrogate key as the primary key, incorporating error handling mechanisms, deploying the SSIS package.

3. **TimeDim:**

   The Data Loading Procedure for the Time Dimension Table (DimTime) using SSIS encompasses loading data from the Weektable.csv file to the staging table (Week_stg). The process involves data cleansing, transformation, and loading into

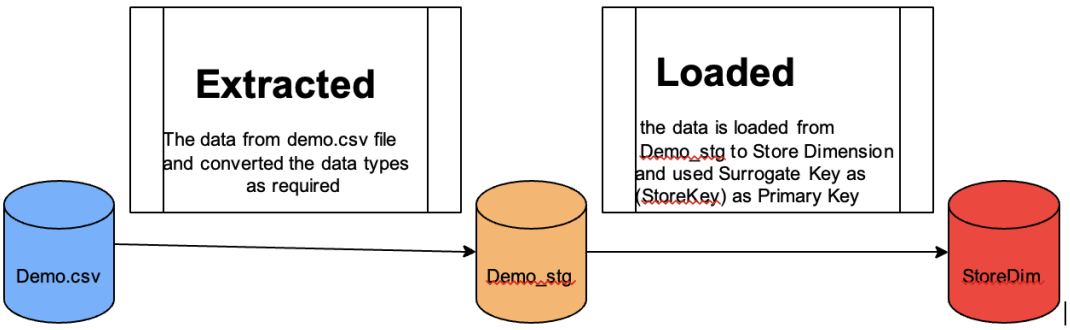
DimTime, including the addition of a surrogate key as the primary key. It addresses data quality, implements error handling, deploys the SSIS package.

4. **ProductDim:**

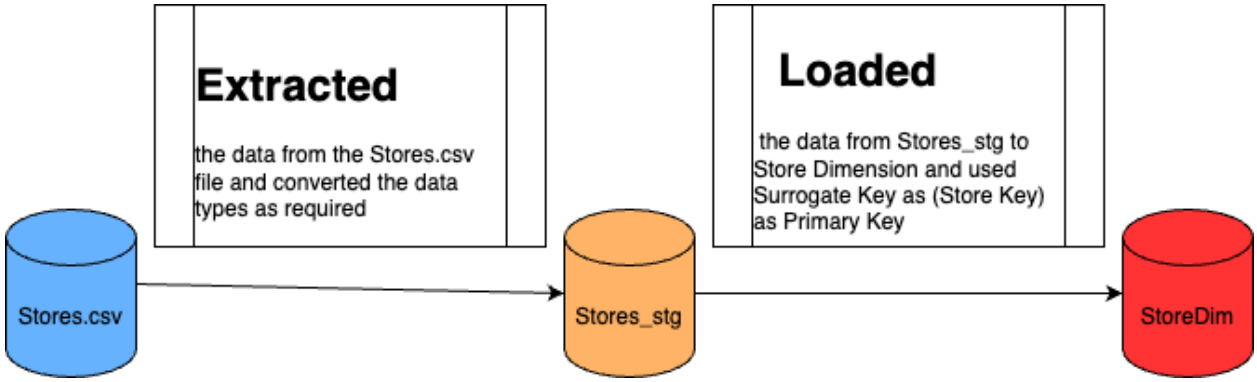
   The Data Loading Procedure for the Product Dimension Table (DimProduct) using SSIS includes loading data from the Ccount.csv source file to the staging table (Ccount_stg). The process involves data cleansing, transformation, and loading into the DimProduct dimension table. It incorporates the addition of a surrogate key as the primary key, addressing data quality issues, implementing error handling, and deploying the SSIS package.

5.9 ETL for Dimension Table

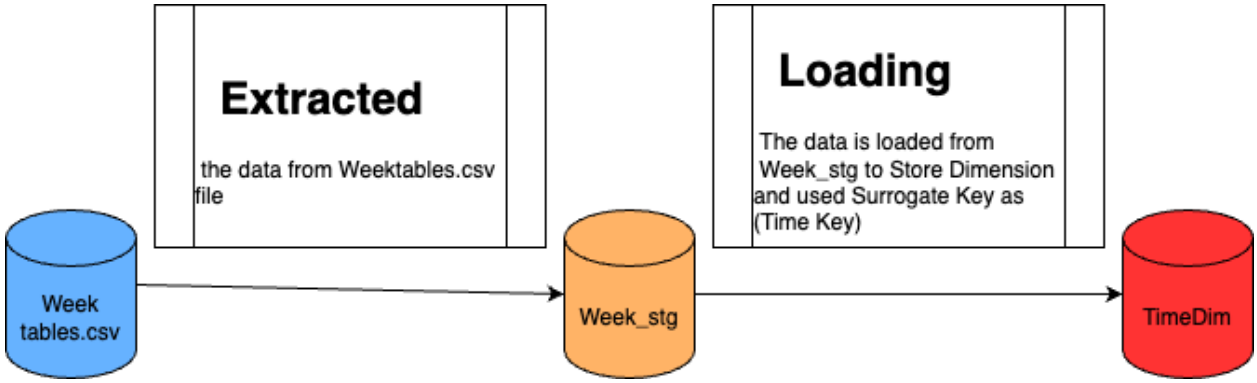
1. **DemoDim:**



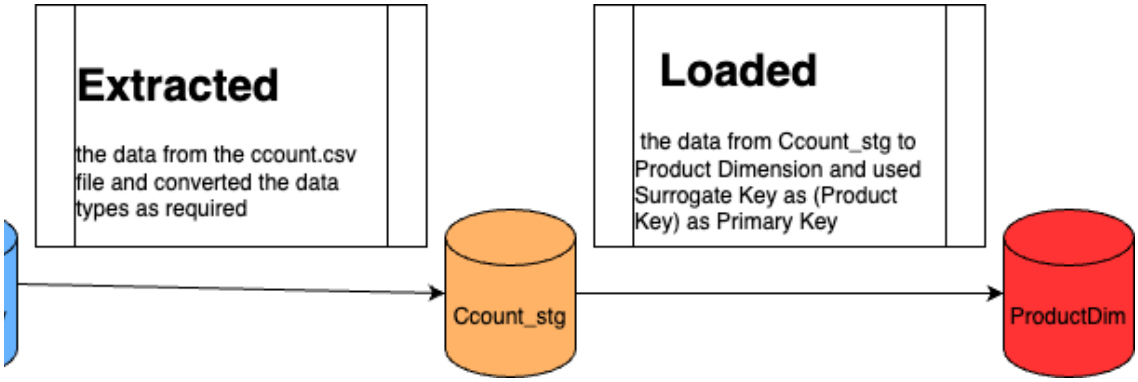
StoreDim:



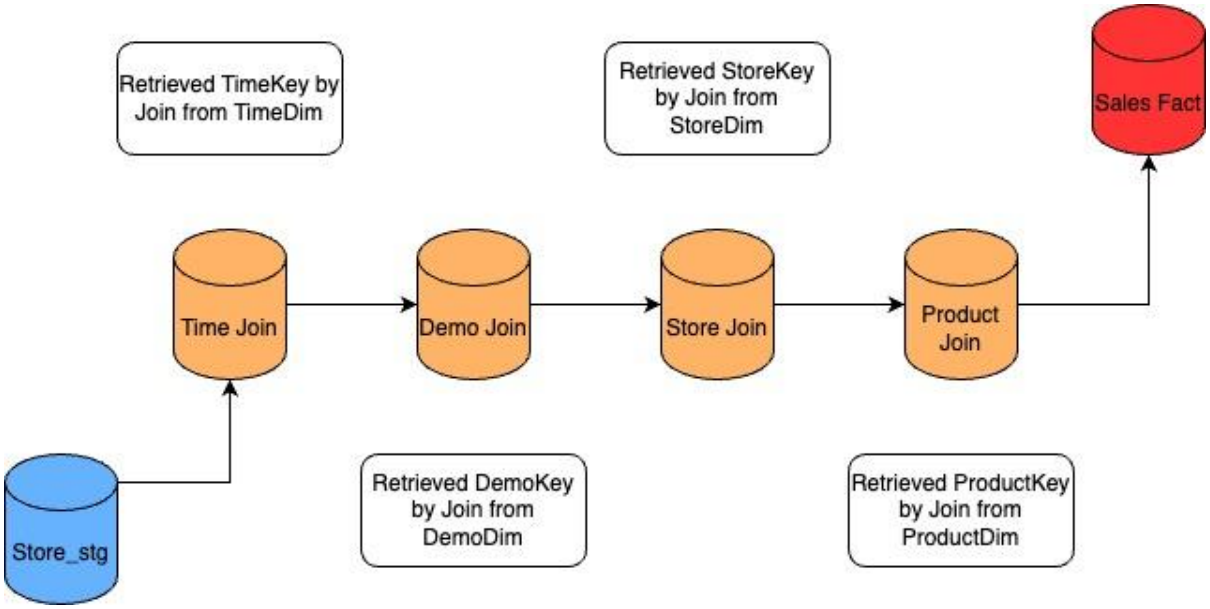
TimeDim:



ProductDim:



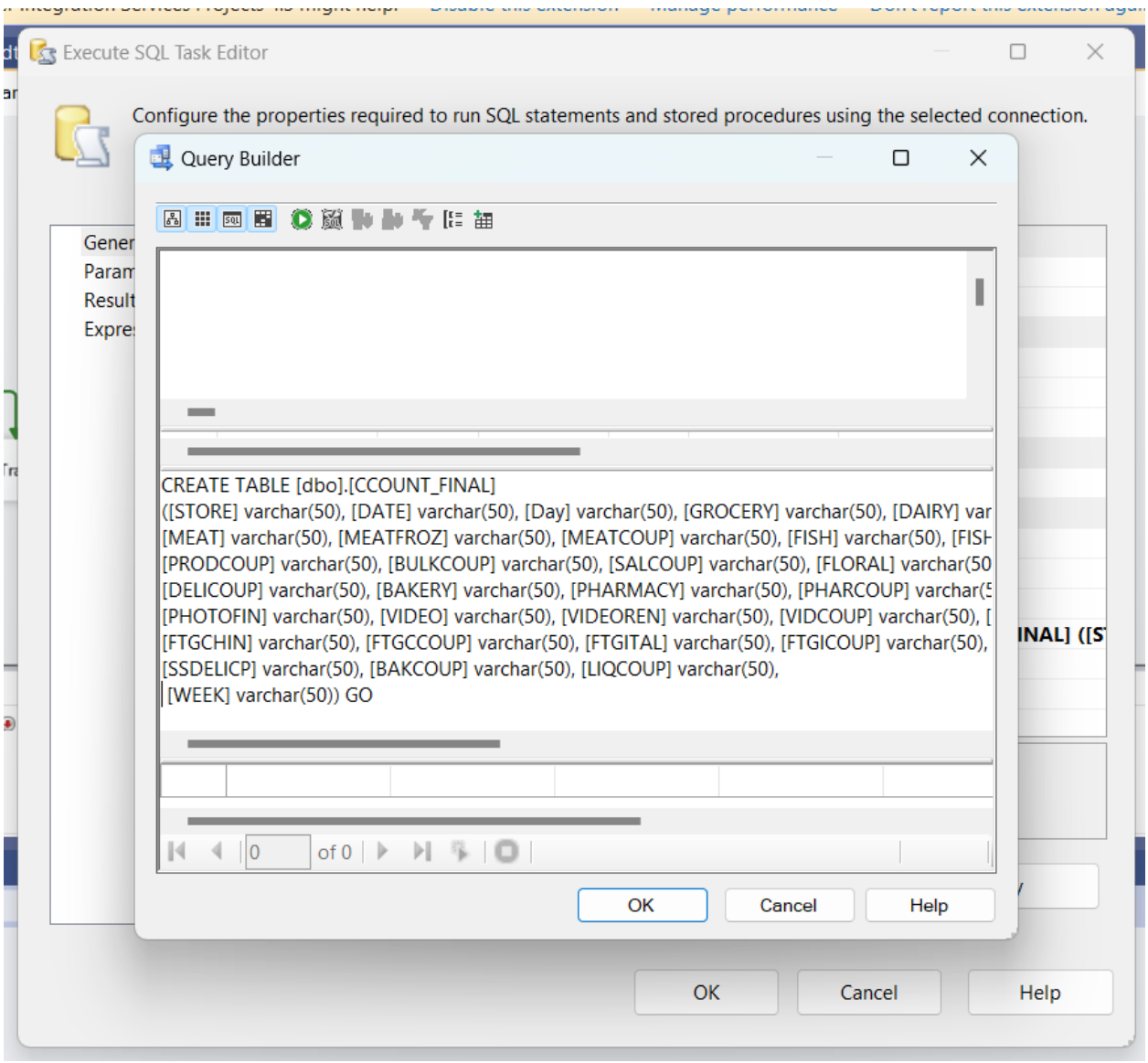
5.10 ETL for Fact Table

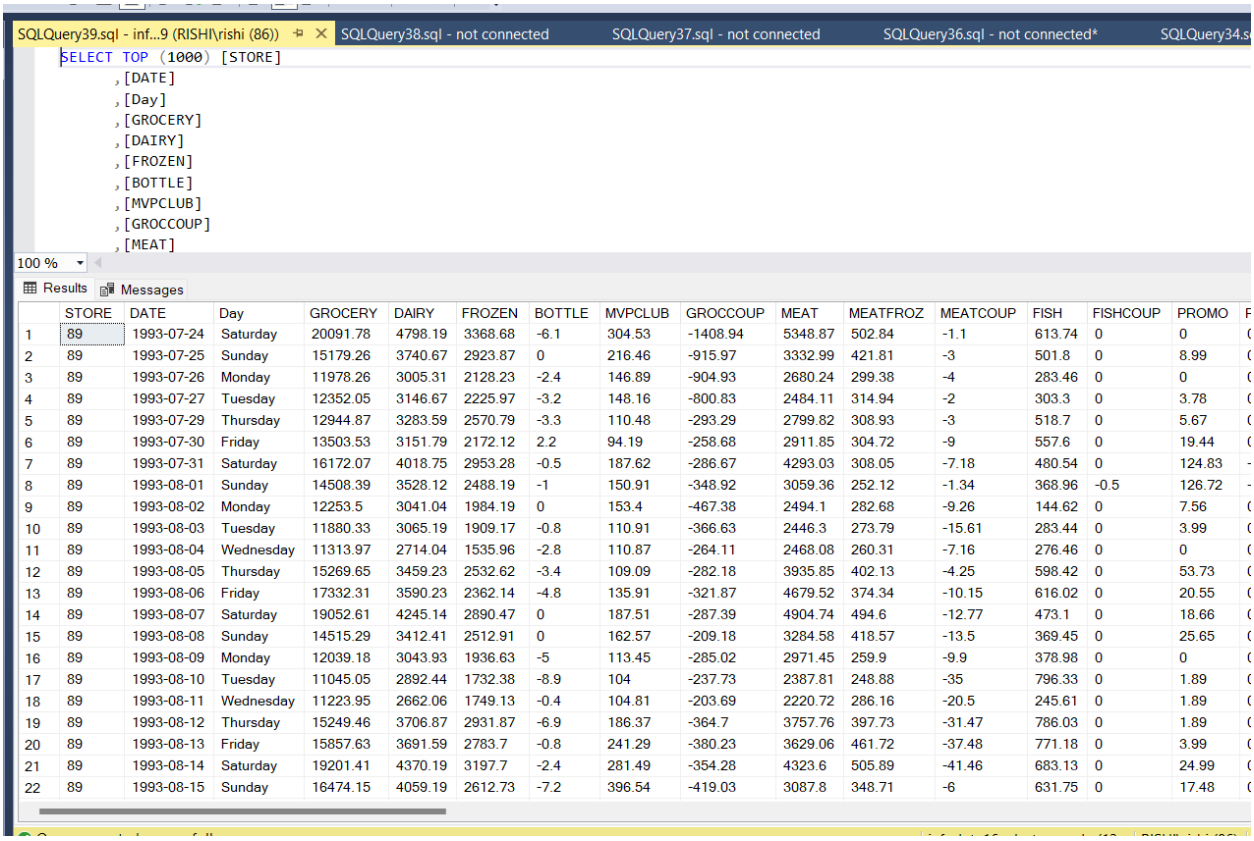


Section 6 - Implementation of ETL

6.1 Extraction of Staging Tables

1. **Ccount staging table**


Execute SQL Task Editor
Configure the properties required to run SQL statements and stored procedures using the selected connection.
Query Builder
CREATE TABLE [dbo].[CCOUNT_FINAL]
([STORE] varchar(50), [DATE] varchar(50), [Day] varchar(50), [GROCERY] varchar(50), [DAIRY] var
[MEAT] varchar(50), [MEATFROZ] varchar(50), [MEATCOUP] varchar(50), [FISH] varchar(50), [FISH
[PRODCOUP] varchar(50), [BULKCOUP] varchar(50), [SALCOUP] varchar(50), [FLORAL] varchar(50
[DELICOUP] varchar(50), [BAKERY] varchar(50), [PHARMACY] varchar(50), [PHARCOUP] varchar(5
[PHOTOFIN] varchar(50), [VIDEO] varchar(50), [VIDEOREN] varchar(50), [VIDCOUP] varchar(50), [
[FTGCHIN] varchar(50), [FTGCCOUP] varchar(50), [FTGITAL] varchar(50), [FTGICOUP] varchar(50),
[SSDELICP] varchar(50), [BAKCOUP] varchar(50), [LIQCOUP] varchar(50),
[WEEK] varchar(50)) GO
0
of 0
OK
Cancel
Help
OK
Cancel
Help



	STORE	DATE	Day	GROCERY	DAIRY	FROZEN	BOTTLE	MVPCLUB	GROCCOUP	MEAT	MEATFROZ	MEATCOUP	FISH	FISHCOUP	PROMO
1	89	1993-07-24	Saturday	20091.78	4798.19	3368.68	-6.1	304.53	-1408.94	5348.87	502.84	-1.1	613.74	0	0
2	89	1993-07-25	Sunday	15179.26	3740.67	2923.87	0	216.46	-915.97	3332.99	421.81	-3	501.8	0	8.99
3	89	1993-07-26	Monday	11978.26	3005.31	2128.23	-2.4	146.89	-904.93	2680.24	299.38	-4	283.46	0	0
4	89	1993-07-27	Tuesday	12352.05	3146.67	2225.97	-3.2	148.16	-800.83	2484.11	314.94	-2	303.3	0	3.78
5	89	1993-07-29	Thursday	12944.87	3283.59	2570.79	-3.3	110.48	-293.29	2799.82	308.93	-3	518.7	0	5.67
6	89	1993-07-30	Friday	13503.53	3151.79	2172.12	2.2	94.19	-258.68	2911.85	304.72	-9	557.6	0	19.44
7	89	1993-07-31	Saturday	16172.07	4018.75	2953.28	-0.5	187.62	-286.67	4293.03	308.05	-7.18	480.54	0	124.83
8	89	1993-08-01	Sunday	14508.39	3528.12	2488.19	-1	150.91	-348.92	3059.36	252.12	-1.34	368.96	-0.5	126.72
9	89	1993-08-02	Monday	12253.5	3041.04	1984.19	0	153.4	-467.38	2494.1	282.68	-9.26	144.62	0	7.56
10	89	1993-08-03	Tuesday	11880.33	3065.19	1909.17	-0.8	110.91	-366.63	2446.3	273.79	-15.61	283.44	0	3.99
11	89	1993-08-04	Wednesday	11313.97	2714.04	1535.96	-2.8	110.87	-264.11	2468.08	260.31	-7.16	276.46	0	0
12	89	1993-08-05	Thursday	15269.65	3459.23	2532.62	-3.4	109.09	-282.18	3935.85	402.13	-4.25	598.42	0	53.73
13	89	1993-08-06	Friday	17332.31	3590.23	2362.14	-4.8	135.91	-321.87	4679.52	374.34	-10.15	616.02	0	20.55
14	89	1993-08-07	Saturday	19052.61	4245.14	2890.47	0	187.51	-287.39	4904.74	494.6	-12.77	473.1	0	18.66
15	89	1993-08-08	Sunday	14515.29	3412.41	2512.91	0	162.57	-209.18	3284.58	418.57	-13.5	369.45	0	25.65
16	89	1993-08-09	Monday	12039.18	3043.93	1936.63	-5	113.45	-285.02	2971.45	259.9	-9.9	378.98	0	0
17	89	1993-08-10	Tuesday	11045.05	2892.44	1732.38	-8.9	104	-237.73	2387.81	248.88	-35	796.33	0	1.89
18	89	1993-08-11	Wednesday	11223.95	2662.06	1749.13	-0.4	104.81	-203.69	2220.72	286.16	-20.5	245.61	0	1.89
19	89	1993-08-12	Thursday	15249.46	3706.87	2931.87	-6.9	186.37	-364.7	3757.76	397.73	-31.47	786.03	0	1.89
20	89	1993-08-13	Friday	15857.63	3691.59	2783.7	-0.8	241.29	-380.23	3629.06	461.72	-37.48	771.18	0	3.99
21	89	1993-08-14	Saturday	19201.41	4370.19	3197.7	-2.4	281.49	-354.28	4323.6	505.89	-41.46	683.13	0	24.99
22	89	1993-08-15	Sunday	16474.15	4059.19	2612.73	-7.2	396.54	-419.03	3087.8	348.71	-6	631.75	0	17.48

Fig: Validation of Ccount data in SSMS for Ccount_stg

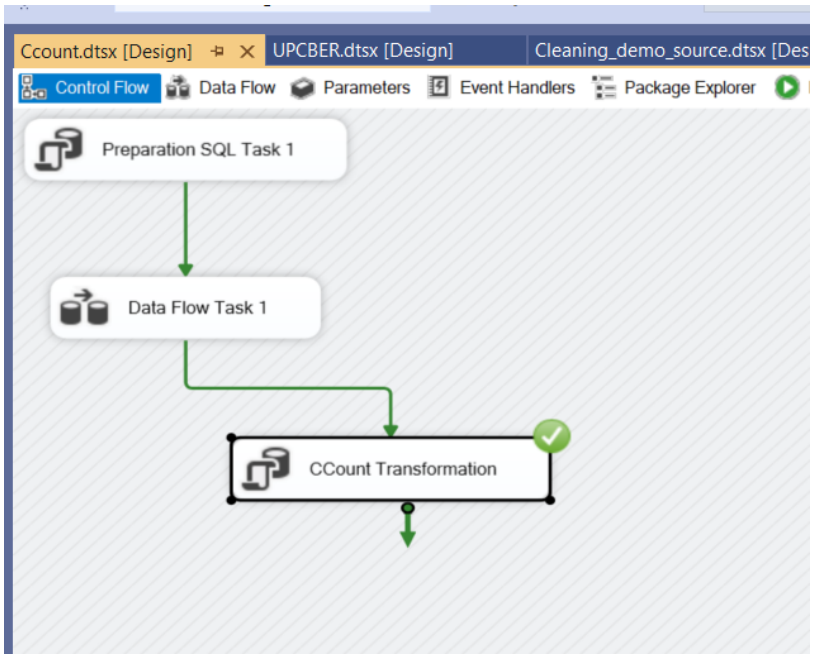


Fig: SQL Task Editor to transform Ccount Staging table

2. Demographic Staging table

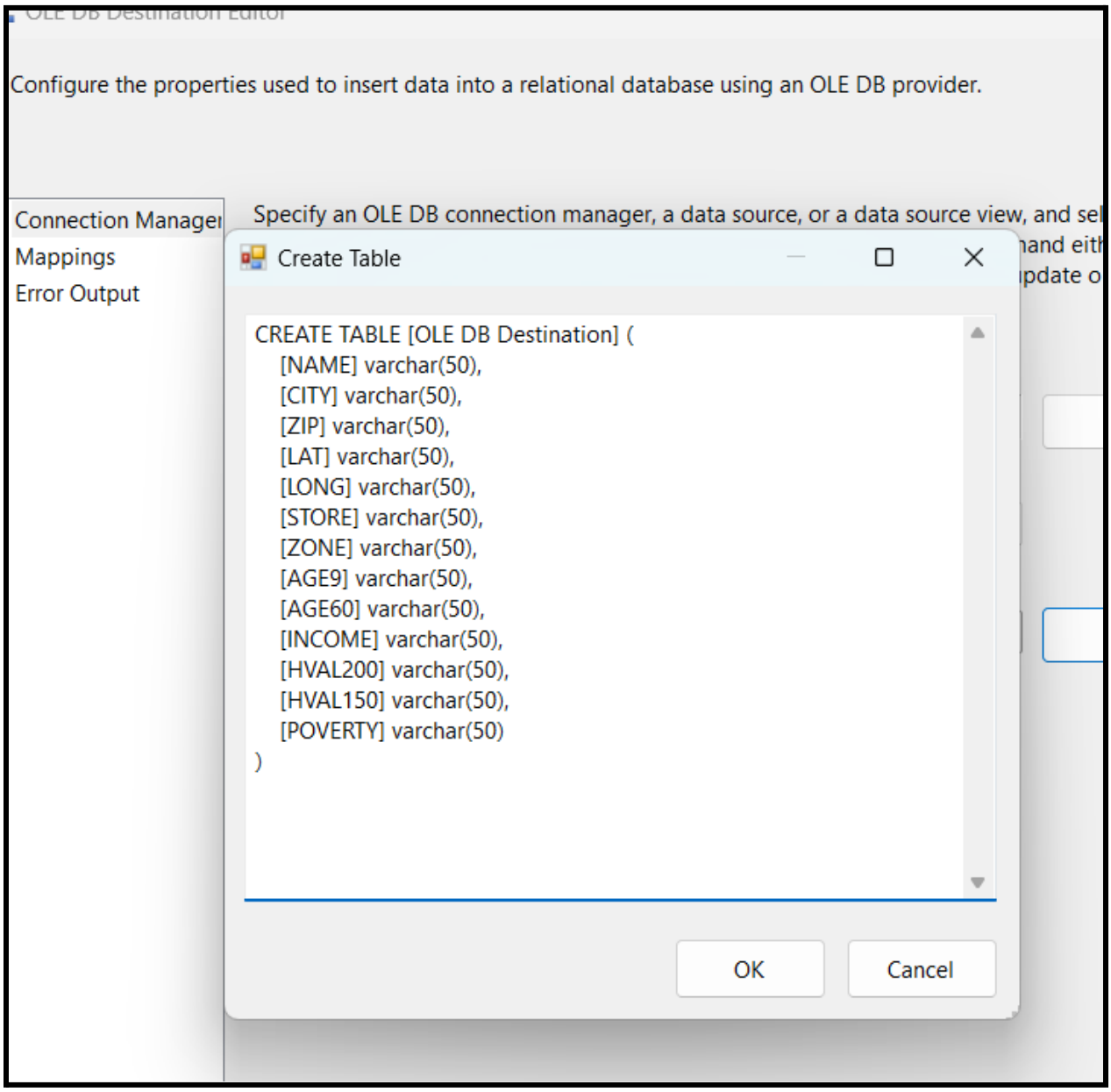


Fig: Loading Demo_stg using SSIS

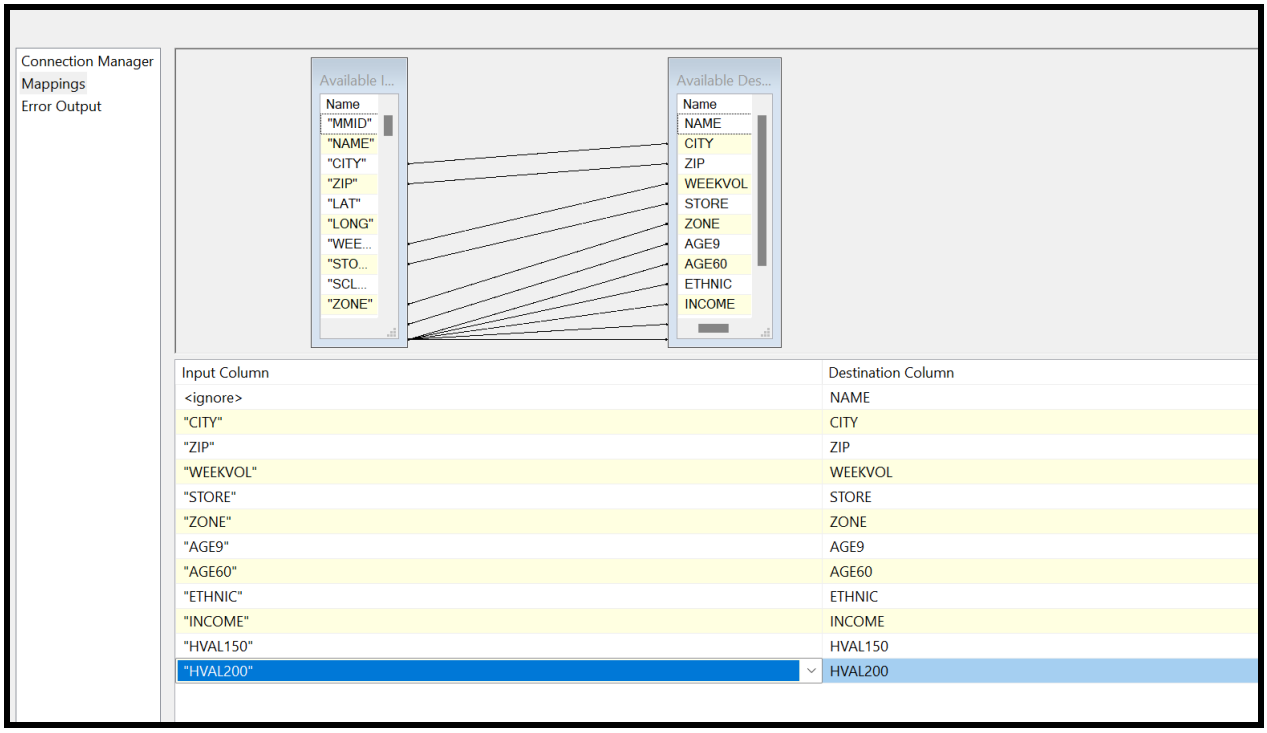
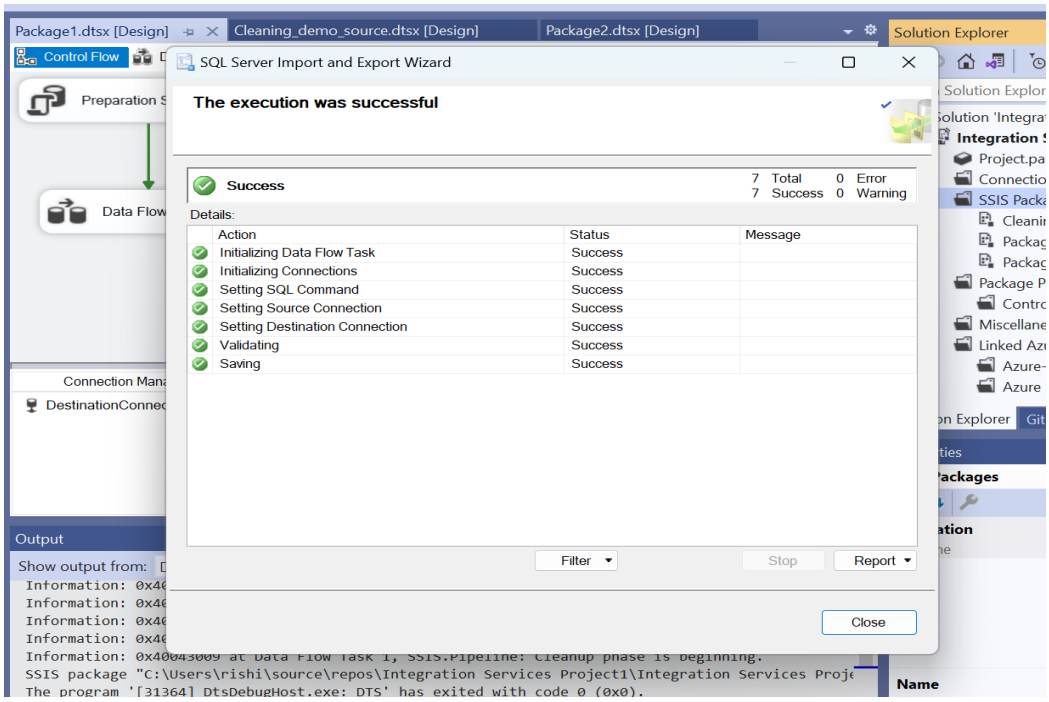
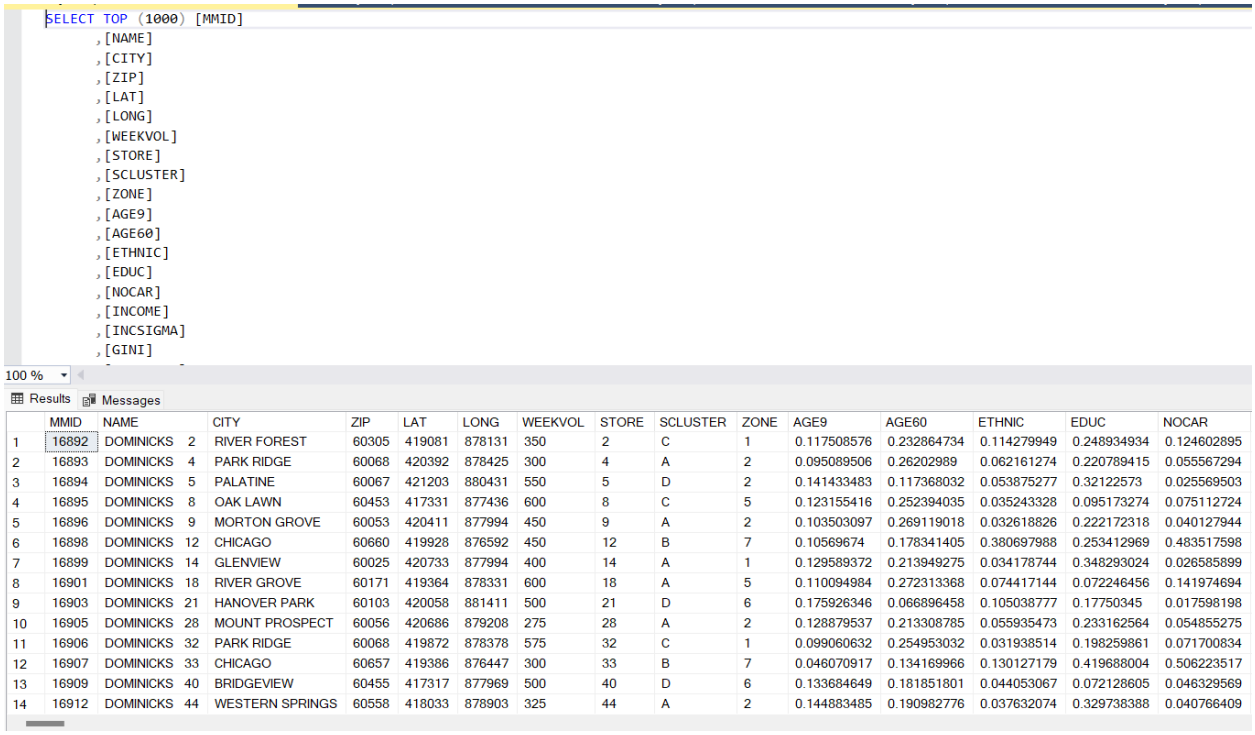


Fig: Source file to Staging table mapping

Fig: Package Execution for loading Demo_stg





```
SELECT TOP (1000) [MMID]
      ,[NAME]
      ,[CITY]
      ,[ZIP]
      ,[LAT]
      ,[LONG]
      ,[WEEKVOL]
      ,[STORE]
      ,[SCLUSTER]
      ,[ZONE]
      ,[AGE9]
      ,[AGE60]
      ,[ETHNIC]
      ,[EDUC]
      ,[NOCAR]
      ,[INCOME]
      ,[INCSIGMA]
      ,[GINI]
```

100 %

Results | Messages

	MMID	NAME	CITY	ZIP	LAT	LONG	WEEKVOL	STORE	SCLUSTER	ZONE	AGE9	AGE60	ETHNIC	EDUC	NOCAR
1	16892	DOMINICKS 2	RIVER FOREST	60305	419081	878131	350	2	C	1	0.117508576	0.232864734	0.114279949	0.248934934	0.124602895
2	16893	DOMINICKS 4	PARK RIDGE	60068	420392	878425	300	4	A	2	0.095089506	0.26202989	0.062161274	0.220789415	0.055567294
3	16894	DOMINICKS 5	PALATINE	60067	421203	880431	550	5	D	2	0.141433483	0.117368032	0.053875277	0.32122573	0.025569503
4	16895	DOMINICKS 8	OAK LAWN	60453	417331	877436	600	8	C	5	0.123155416	0.252394035	0.035243328	0.095173274	0.075112724
5	16896	DOMINICKS 9	MORTON GROVE	60053	420411	877994	450	9	A	2	0.103503097	0.269119018	0.032618826	0.222172318	0.040127944
6	16898	DOMINICKS 12	CHICAGO	60660	419928	876592	450	12	B	7	0.10569674	0.178341405	0.380697988	0.253412969	0.483517598
7	16899	DOMINICKS 14	GLENVIEW	60025	420733	877994	400	14	A	1	0.129589372	0.213949275	0.034178744	0.348293024	0.026585899
8	16901	DOMINICKS 18	RIVER GROVE	60171	419364	878331	600	18	A	5	0.110094984	0.272313368	0.074417144	0.072246456	0.141974694
9	16903	DOMINICKS 21	HANOVER PARK	60103	420058	881411	500	21	D	6	0.175926346	0.066896458	0.105038777	0.17750345	0.017598198
10	16905	DOMINICKS 28	MOUNT PROSPECT	60056	420686	879208	275	28	A	2	0.128879537	0.213308785	0.055935473	0.233162564	0.054855275
11	16906	DOMINICKS 32	PARK RIDGE	60068	419872	878378	575	32	C	1	0.099060632	0.254953032	0.031938514	0.198259861	0.071700834
12	16907	DOMINICKS 33	CHICAGO	60657	419386	876447	300	33	B	7	0.046070917	0.134169966	0.130127179	0.419688004	0.506223517
13	16909	DOMINICKS 40	BRIDGEVIEW	60455	417317	877969	500	40	D	6	0.133684649	0.181851801	0.044053067	0.072128605	0.046329569
14	16912	DOMINICKS 44	WESTERN SPRINGS	60558	418033	878903	325	44	A	2	0.144883485	0.190982776	0.037632074	0.329738388	0.040766409

Fig: Validation of data in SSMS for Demo_stg

3. Staging table 3 - Store_stg table

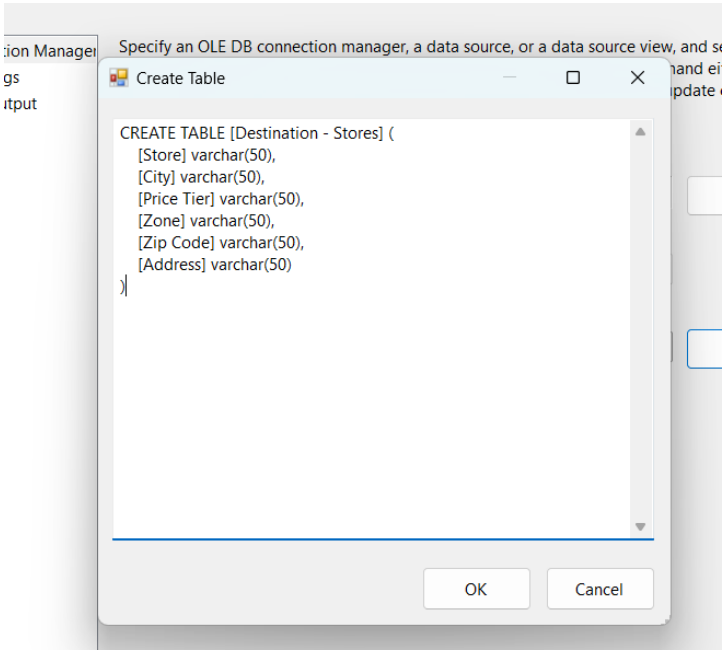


Fig - SQL query to create Store_stg table

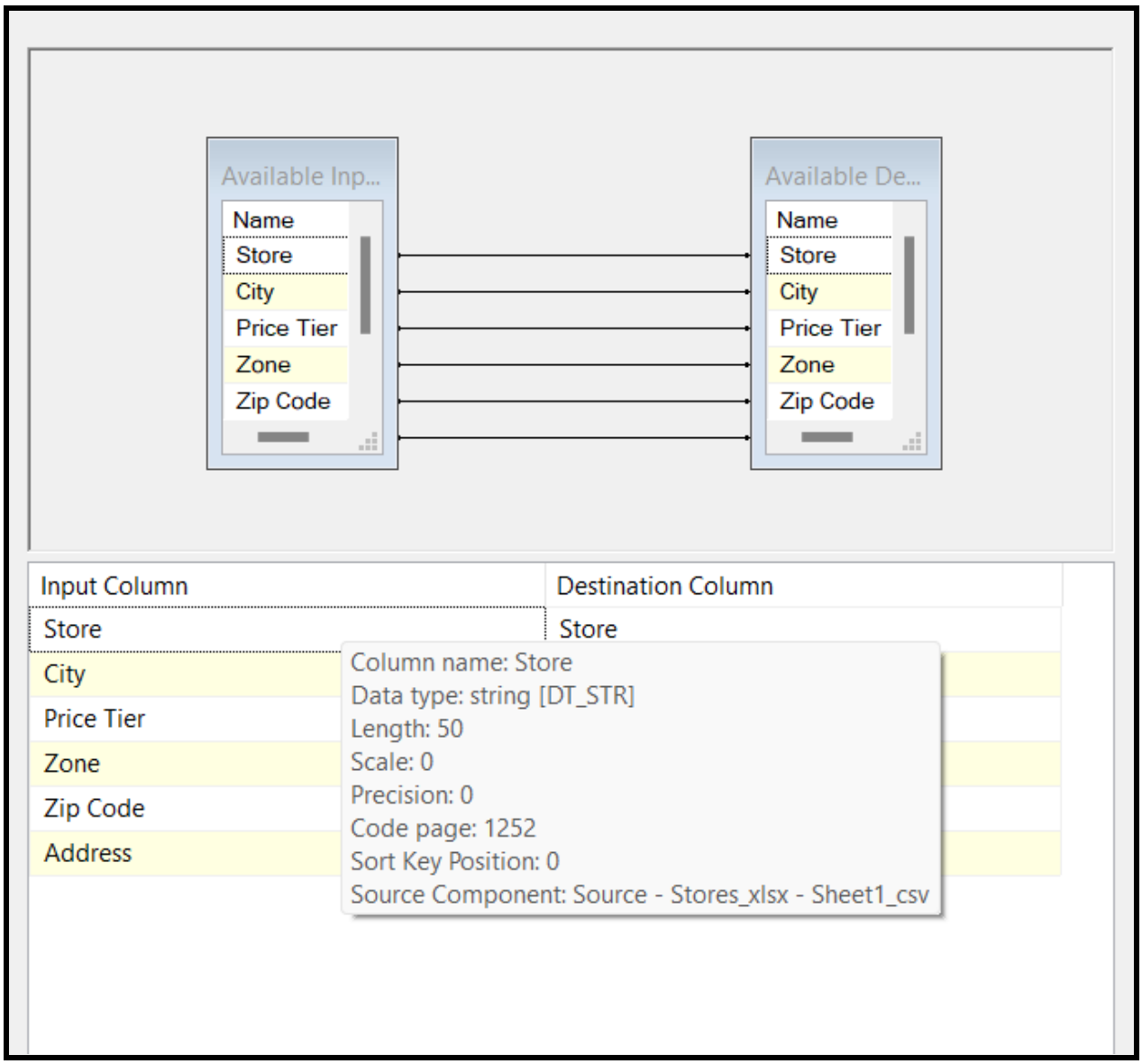
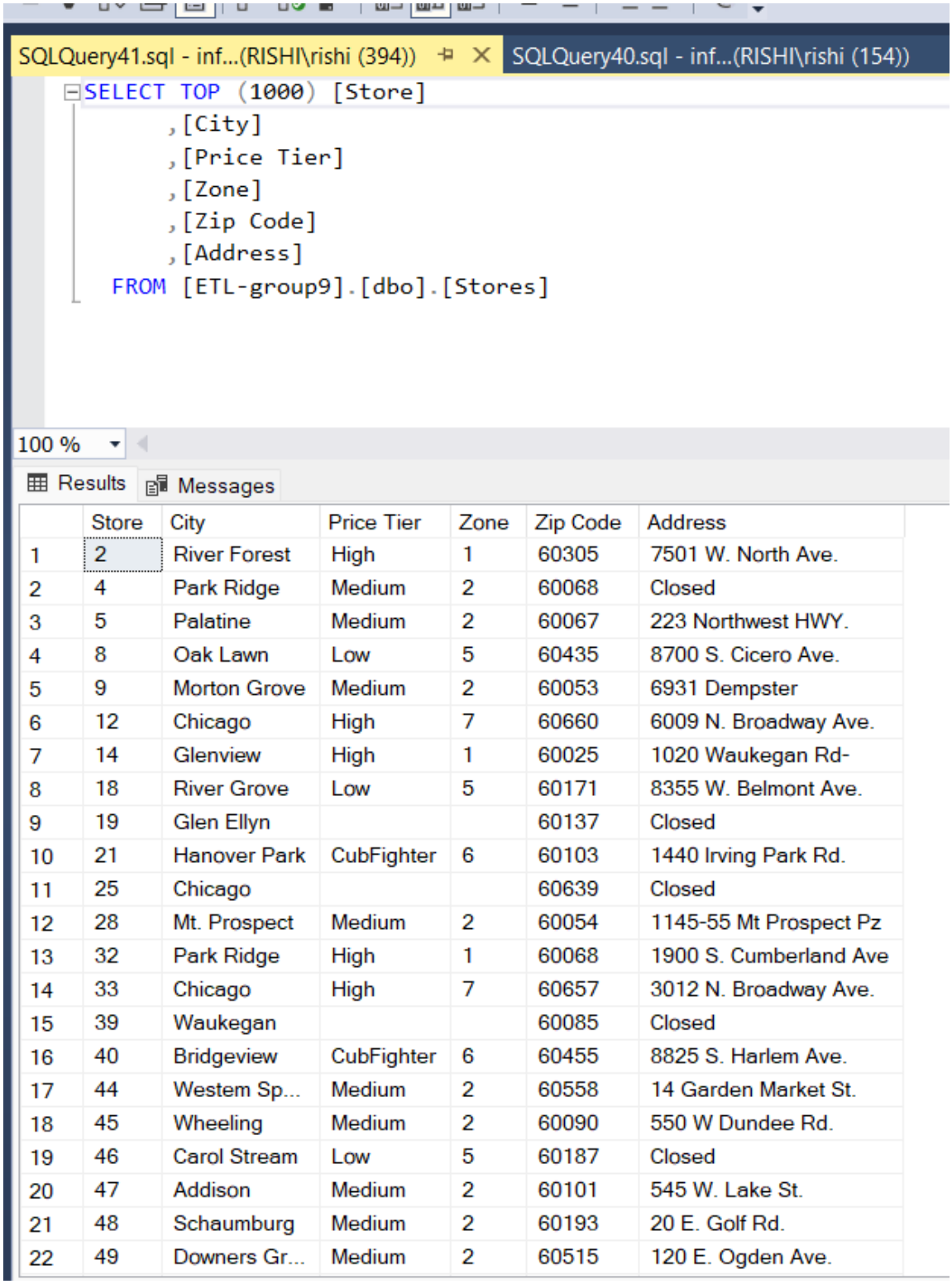


Fig: Mapping for Store_stg table



	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			60137	Closed
10	21	Hanover Park	CubFighter	6	60103	1440 Irving Park Rd.
11	25	Chicago			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			60085	Closed
16	40	Bridgeview	CubFighter	6	60455	8825 S. Harlem Ave.
17	44	Westem Sp...	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
20	47	Addison	Medium	2	60101	545 W. Lake St.
21	48	Schaumburg	Medium	2	60193	20 E. Golf Rd.
22	49	Downers Gr...	Medium	2	60515	120 E. Ogden Ave.

Fig: Validation for Store_stg using SSMS

4. Staging table 4 - Week_stg

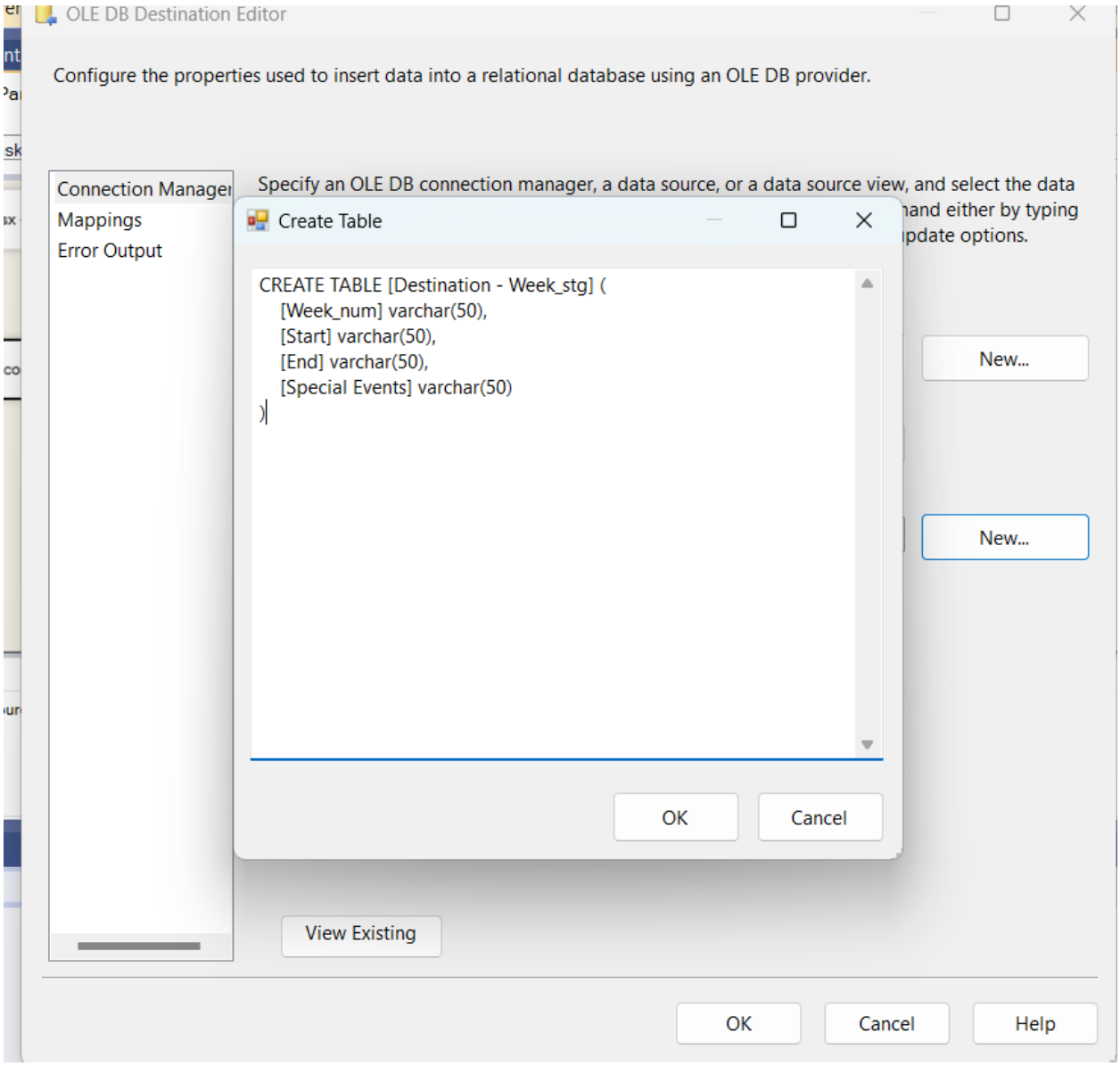


Fig: SQL statement to create Week_stg table

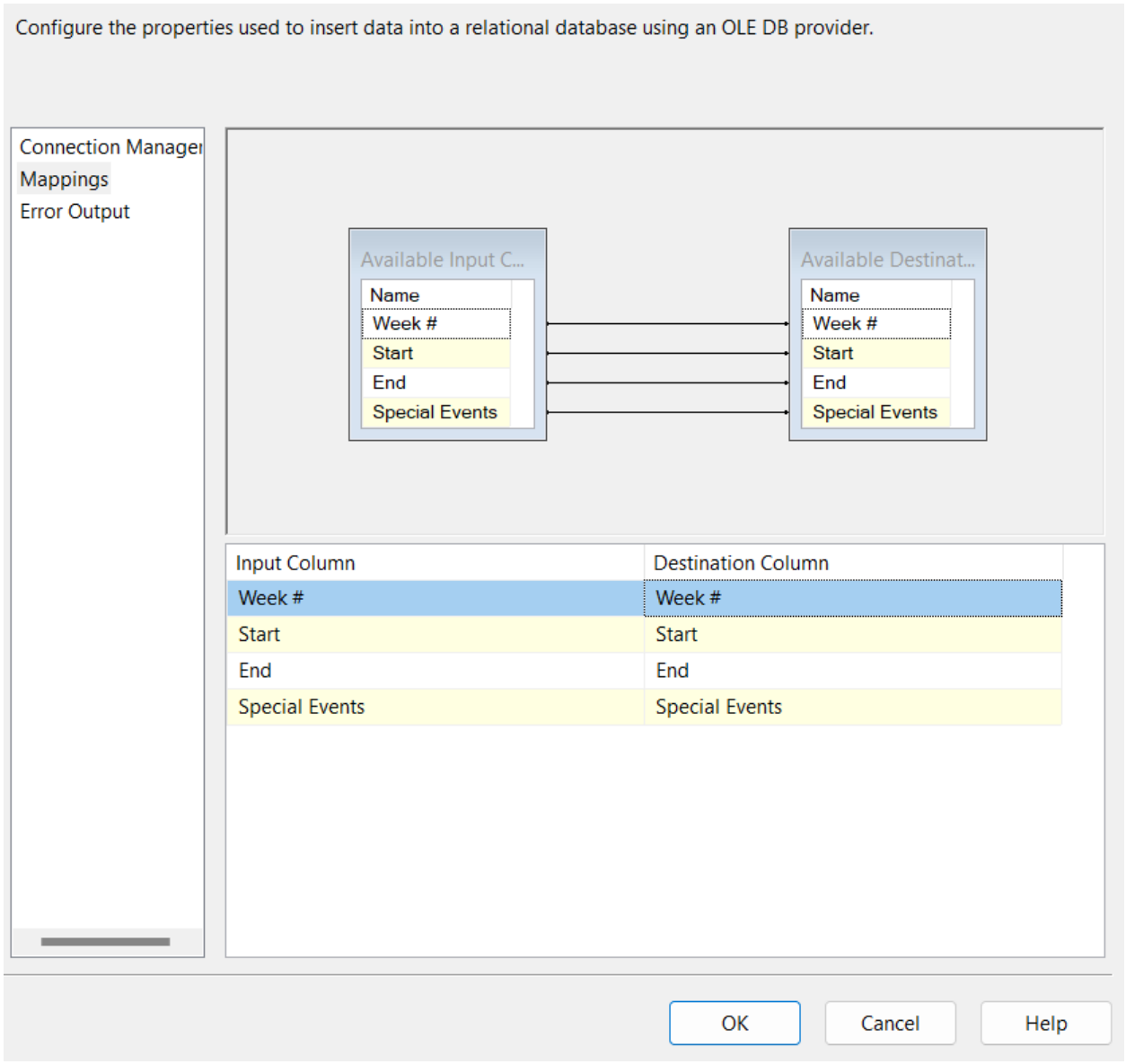
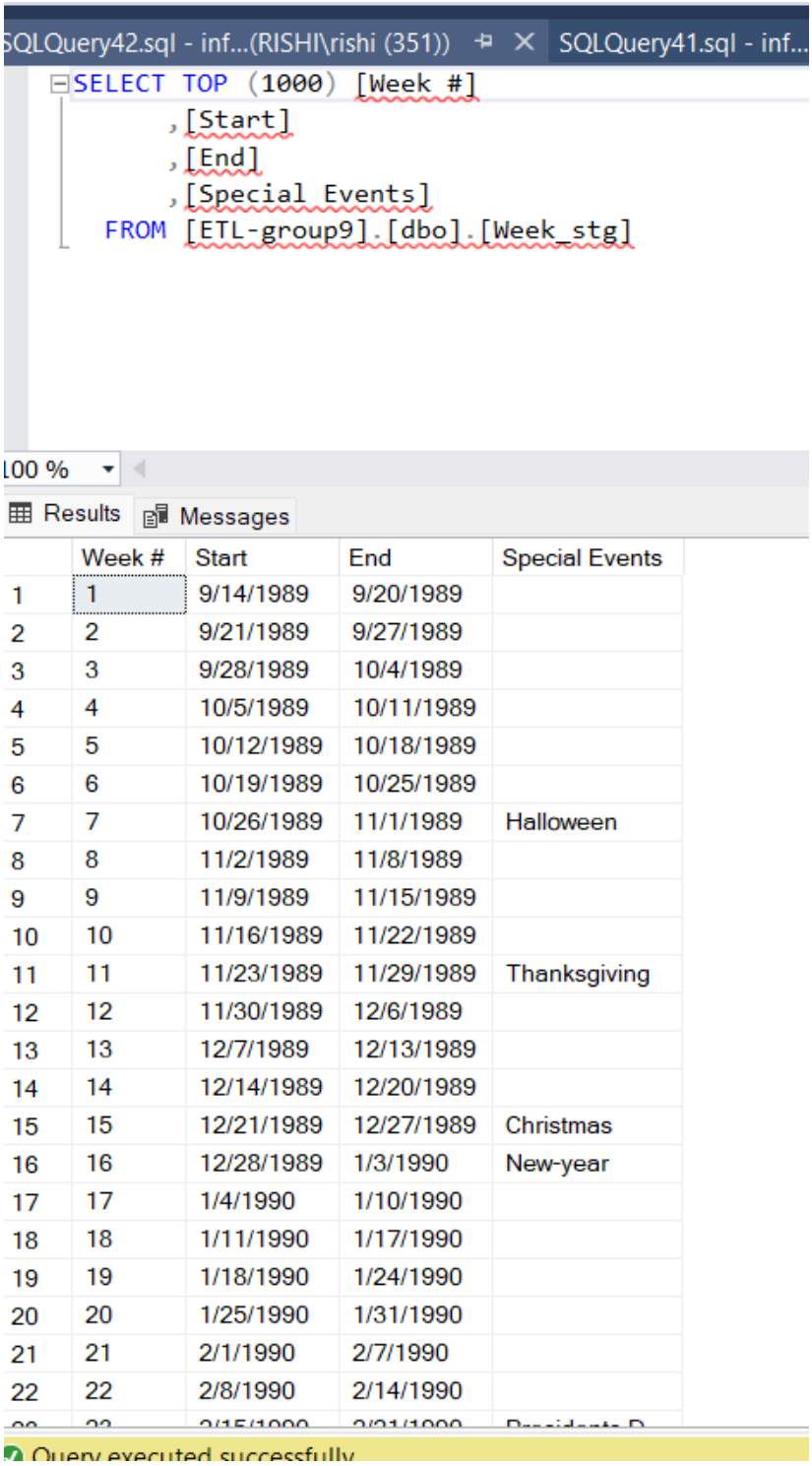


Fig: Mapping for Week_stg table



	Week #	Start	End	Special Events
1	1	9/14/1989	9/20/1989	
2	2	9/21/1989	9/27/1989	
3	3	9/28/1989	10/4/1989	
4	4	10/5/1989	10/11/1989	
5	5	10/12/1989	10/18/1989	
6	6	10/19/1989	10/25/1989	
7	7	10/26/1989	11/1/1989	Halloween
8	8	11/2/1989	11/8/1989	
9	9	11/9/1989	11/15/1989	
10	10	11/16/1989	11/22/1989	
11	11	11/23/1989	11/29/1989	Thanksgiving
12	12	11/30/1989	12/6/1989	
13	13	12/7/1989	12/13/1989	
14	14	12/14/1989	12/20/1989	
15	15	12/21/1989	12/27/1989	Christmas
16	16	12/28/1989	1/3/1990	New-year
17	17	1/4/1990	1/10/1990	
18	18	1/11/1990	1/17/1990	
19	19	1/18/1990	1/24/1990	
20	20	1/25/1990	1/31/1990	
21	21	2/1/1990	2/7/1990	
22	22	2/8/1990	2/14/1990	

Fig: Validation of Week_stg table using SSMS

6.2 Transformation and Cleaning in Staging Area after Extraction

This stage basically requires us to clean all the data in the staging tables and perform the necessary transformations on these staging tables so as to use them safely for Dimension tables.

Table-1: Cleaning Demographic_stg:

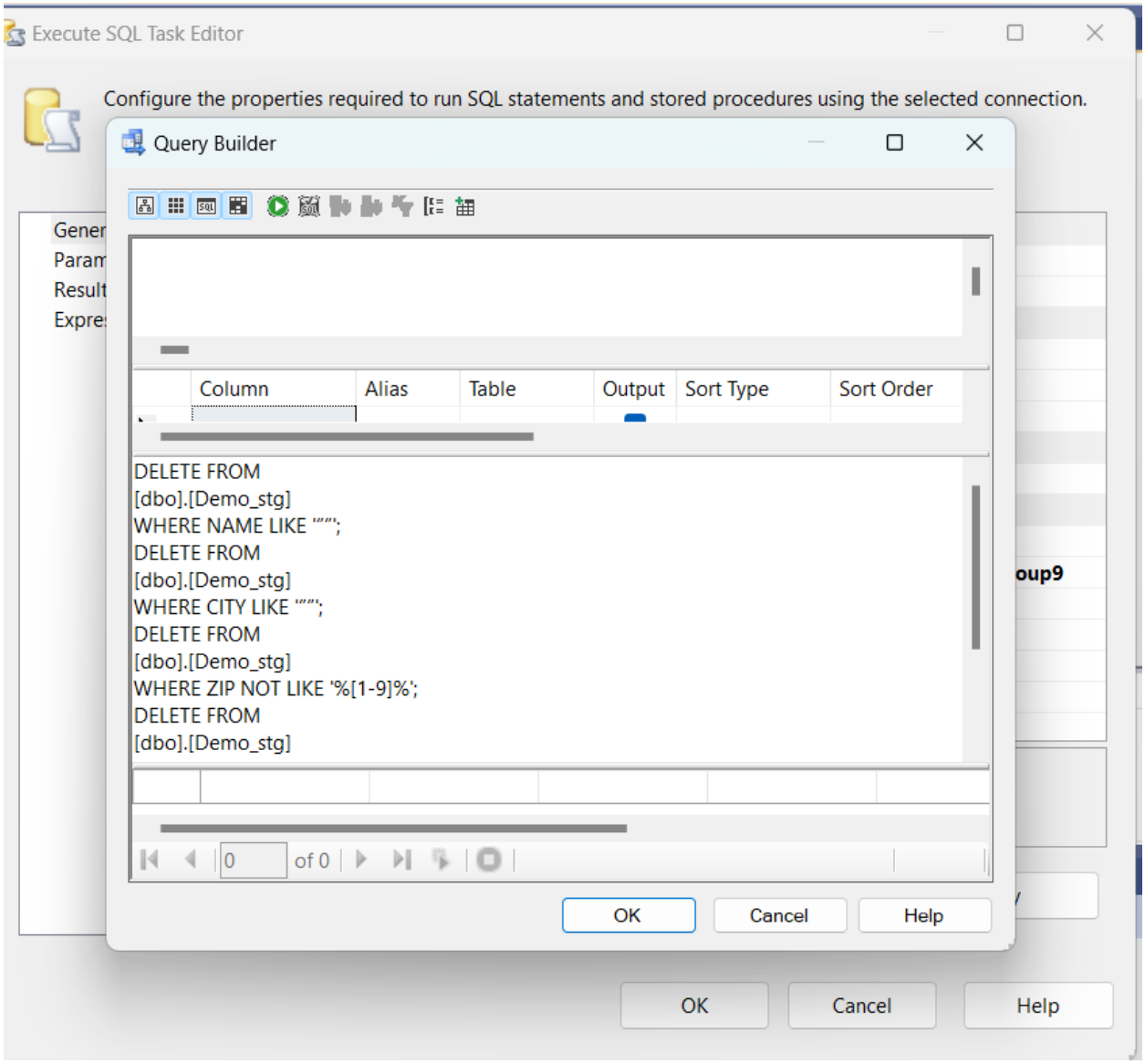


Fig: Query to clean Demo_stg table

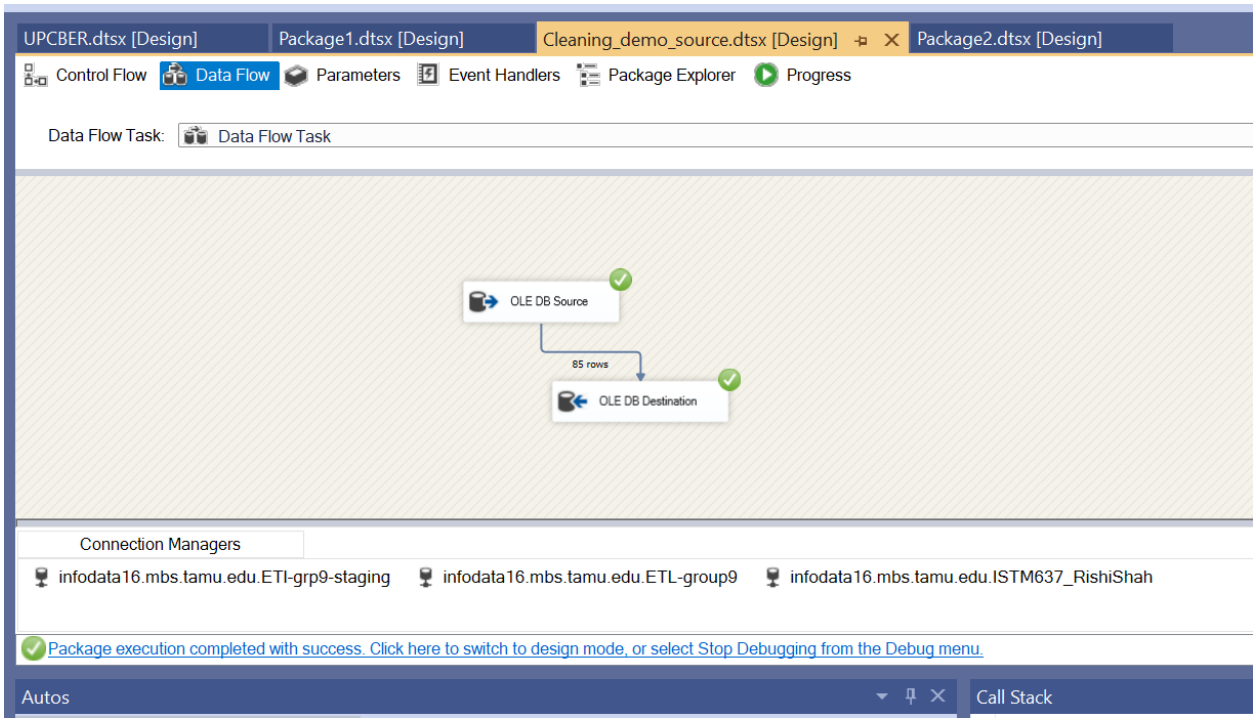


Fig: Completion of Cleaning Demo package

Query to clean Demo_stg	DELETE FROM [dbo].[Demo_stg] WHERE NAME LIKE '"""'; DELETE FROM [dbo].[Demo_stg] WHERE CITY LIKE '"""'; DELETE FROM [dbo].[Demo_stg] WHERE ZIP NOT LIKE '%[1-9]%'; DELETE FROM [dbo].[Demo_stg] WHERE STORE NOT LIKE '%[1-9]%'; DELETE FROM [dbo].[Demo_stg] WHERE ZONE NOT LIKE '%[1-9]%'; ALTER TABLE [dbo].[Demo_stg] ALTER COLUMN [STORE] INT; ALTER TABLE [dbo].[Demo_stg] ALTER COLUMN [ZONE] INT;

Table-2: Cleaning Ccount_stg

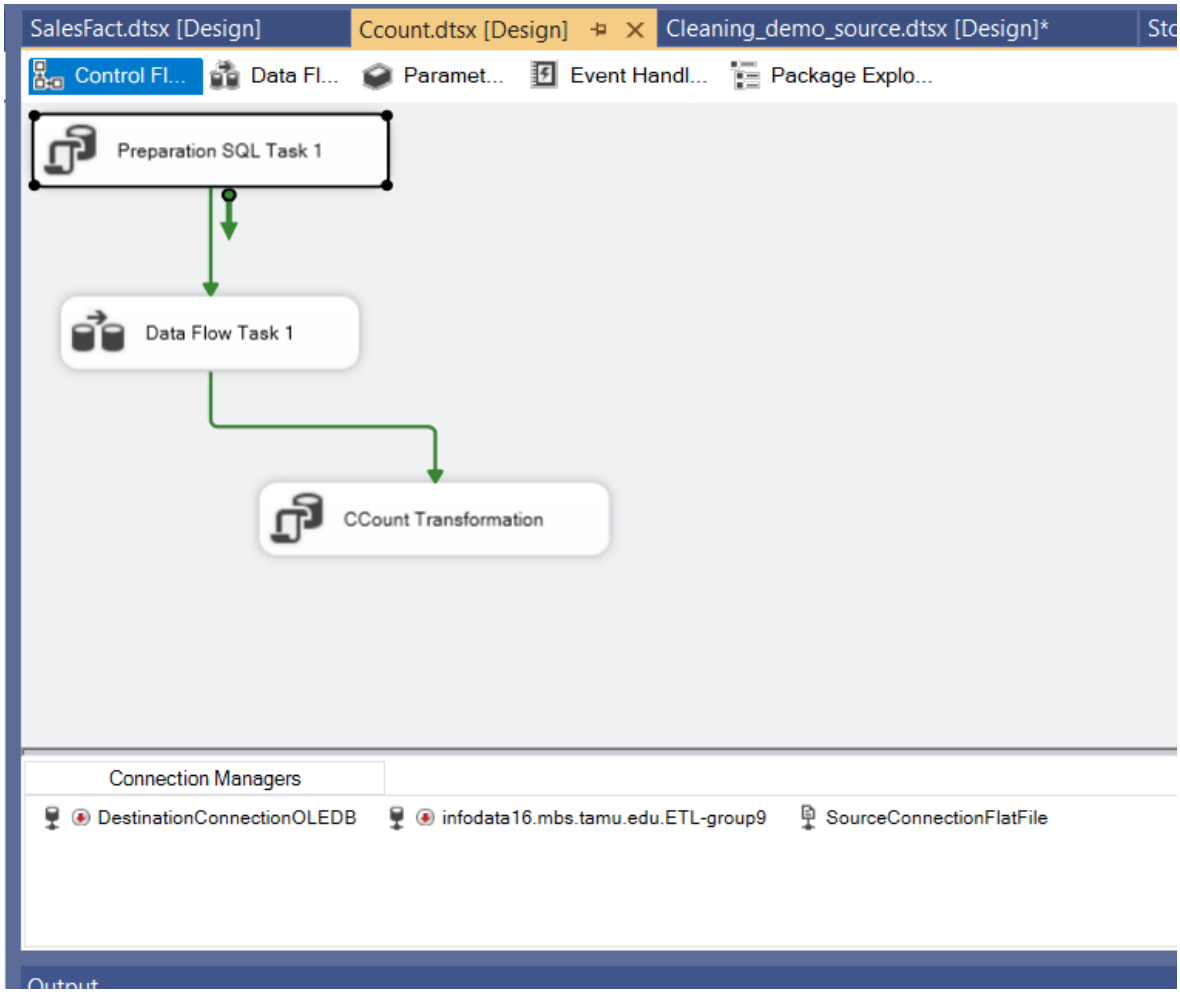


Fig: SSIS package to clean Ccount file

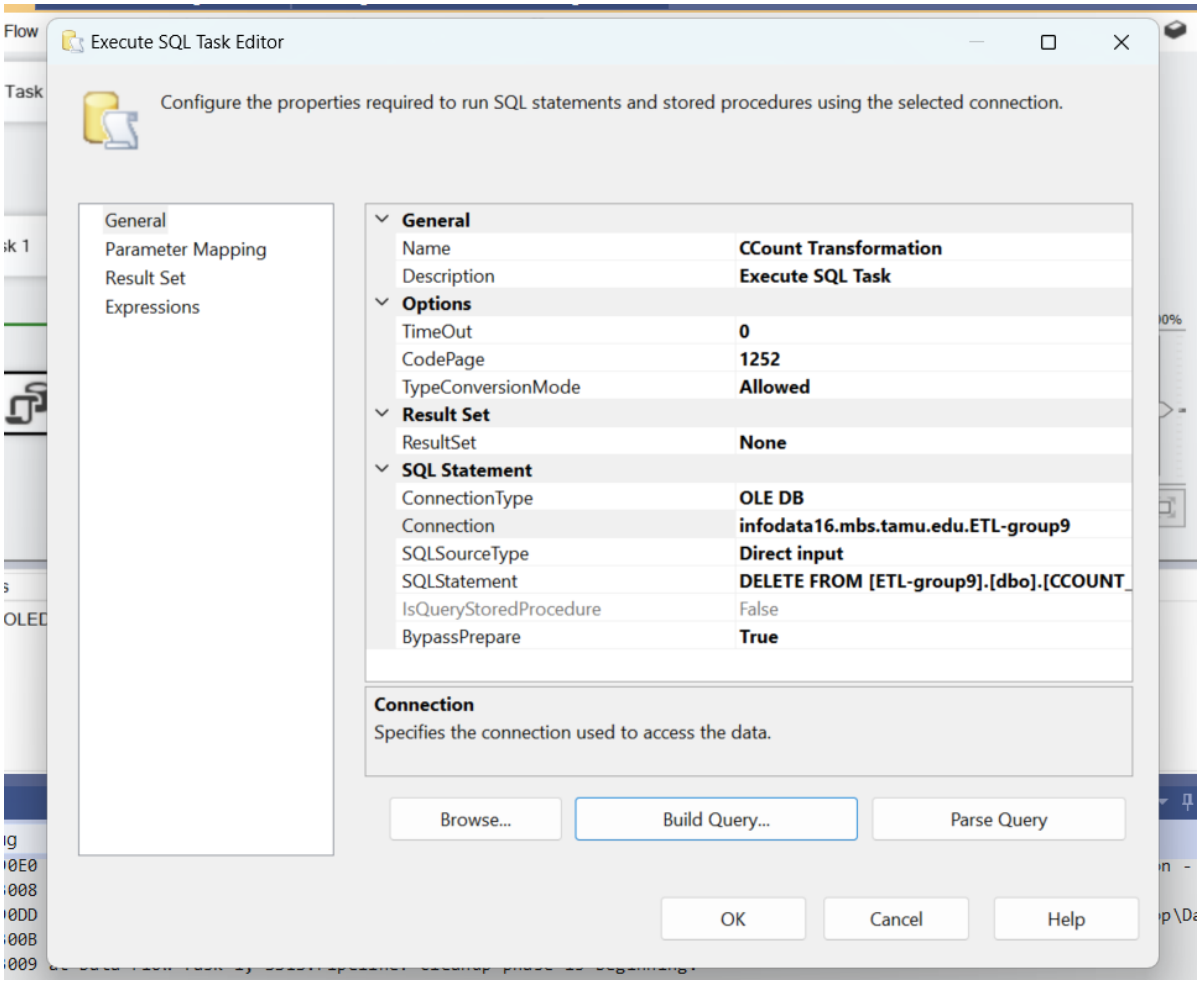


Fig: Query editor to clean Ccount_stg

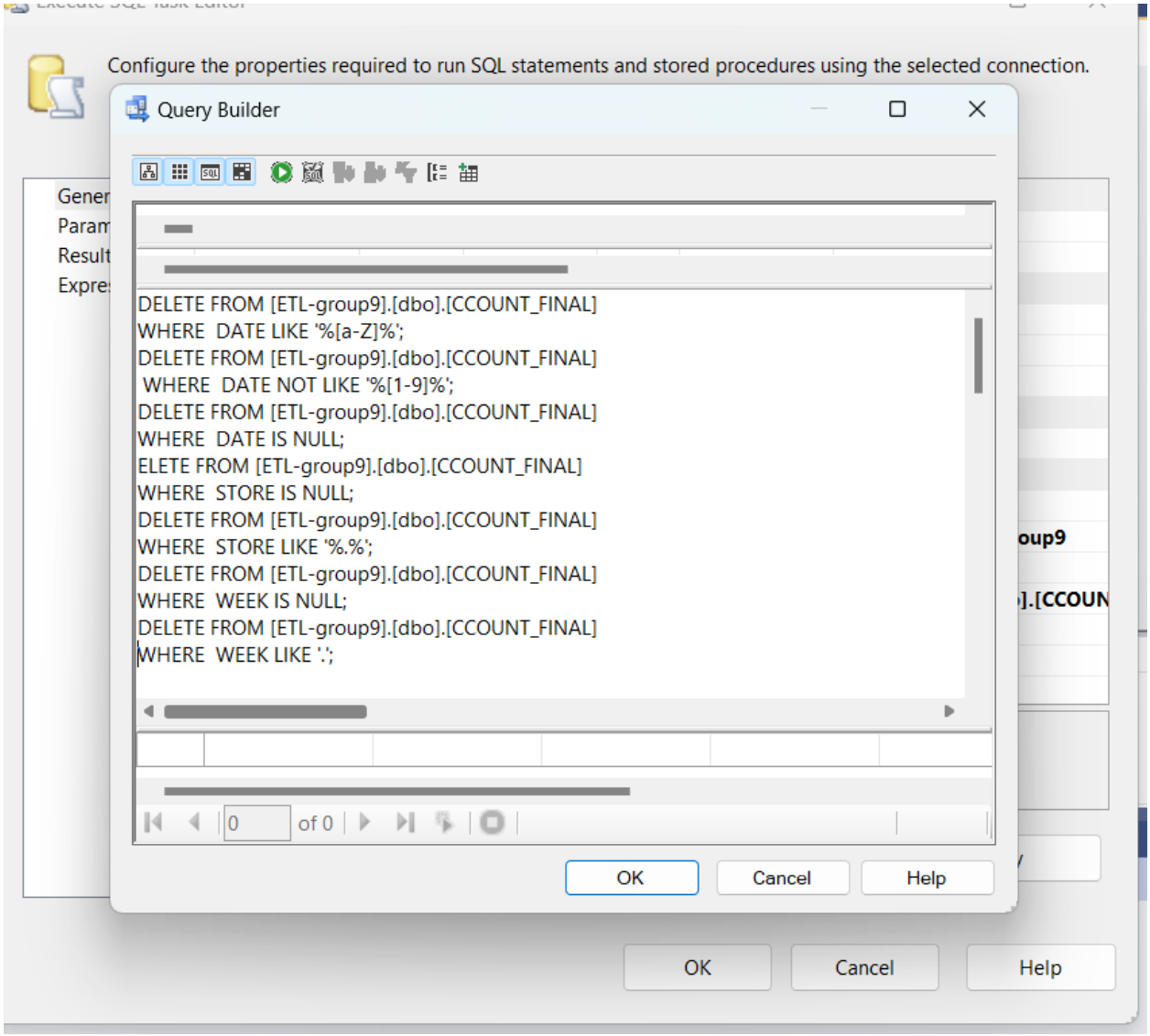


Fig: SQL query to clean ccount_stg

SQL Statement	DELETE FROM [ETL-group9].[dbo].[CCOUNT_FINAL] WHERE DATE LIKE '%[a-Z]%'; DELETE FROM [ETL-group9].[dbo].[CCOUNT_FINAL] WHERE DATE NOT LIKE '%[1-9]%'; DELETE FROM [ETL-group9].[dbo].[CCOUNT_FINAL]

	WHERE DATE IS NULL; DELETE FROM [ETL-group9].[dbo].[CCOUNT_FINAL] WHERE STORE IS NULL; DELETE FROM [ETL-group9].[dbo].[CCOUNT_FINAL] WHERE STORE LIKE '%.%'; DELETE FROM [ETL-group9].[dbo].[CCOUNT_FINAL] WHERE WEEK IS NULL; DELETE FROM [ETL-group9].[dbo].[CCOUNT_FINAL] WHERE WEEK LIKE '.'; ALTER TABLE [ETL-group9].[dbo].[CCOUNT_FINAL] ALTER COLUMN [GROCERY] FLOAT; ALTER TABLE [ETL-group9].[dbo].[CCOUNT_FINAL] ALTER COLUMN [BAKERY] FLOAT; ALTER TABLE [ETL-group9].[dbo].[CCOUNT_FINAL] ALTER COLUMN [FISH] FLOAT; ALTER TABLE [ETL-group9].[dbo].[CCOUNT_FINAL] ALTER COLUMN [MEATFROZ] FLOAT; ALTER TABLE [ETL-group9].[dbo].[CCOUNT_FINAL] ALTER COLUMN [BOTTLE] FLOAT; ALTER TABLE [ETL-group9].[dbo].[CCOUNT_FINAL] ALTER COLUMN [CHEESE] FLOAT;

	ALTER TABLE [ETL-group9].[dbo].[CCOUNT_FINAL] ALTER COLUMN [MEAT] FLOAT; ALTER TABLE [ETL-group9].[dbo].[CCOUNT_FINAL] ALTER COLUMN [DELI] FLOAT; ALTER TABLE [ETL-group9].[dbo].[CCOUNT_FINAL] ALTER COLUMN [DAIRY] FLOAT; ALTER TABLE [ETL-group9].[dbo].[CCOUNT_FINAL] ALTER COLUMN [PHARMACY] FLOAT;

6.3 DATA LOADING

Dimension Tables Creation

1. **DemographicDim**

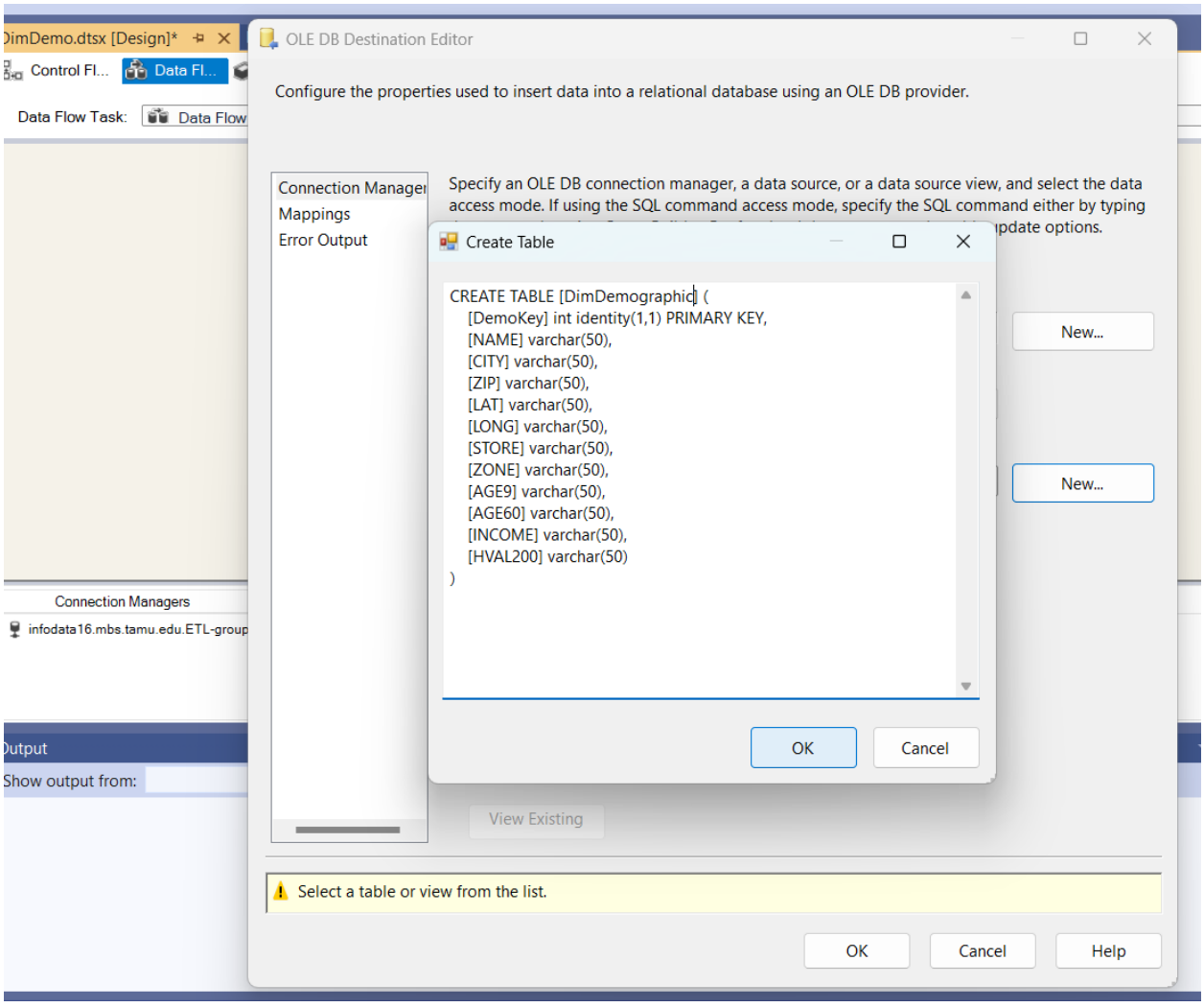


Fig: Query statement to create DemographicDim

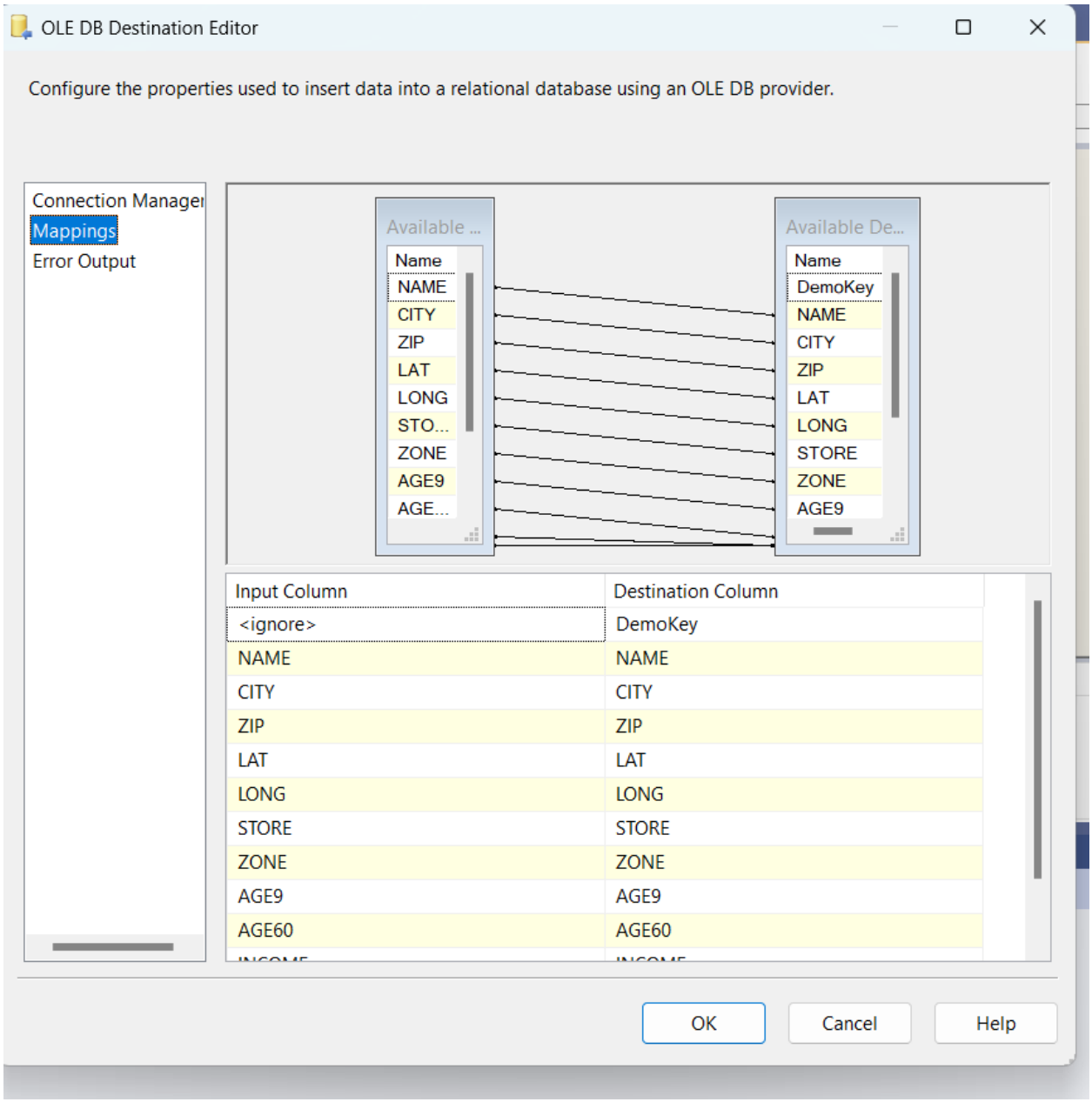
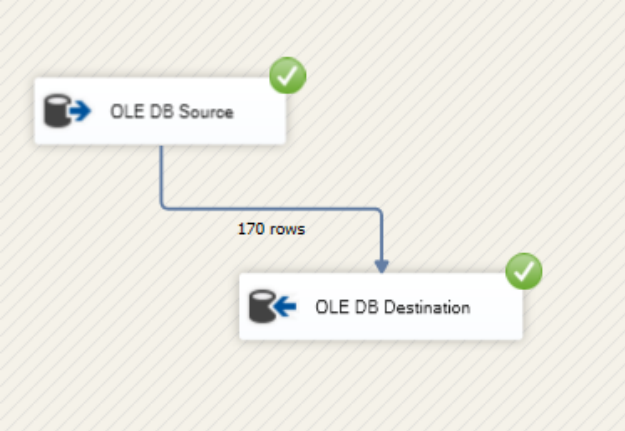


Fig: Mappings for DemographicDim



2. StoreDim

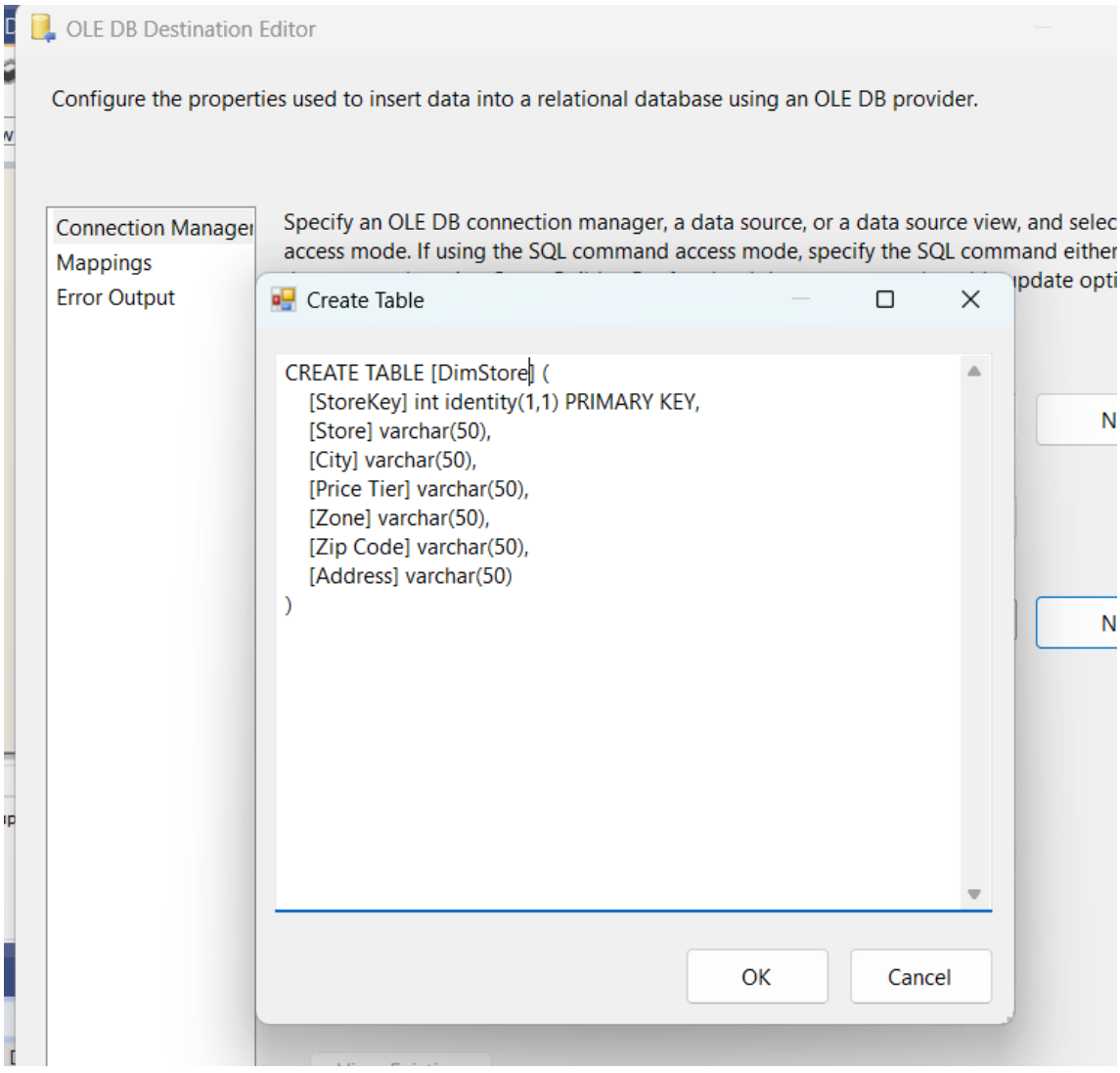


Fig: SQL query to create StoreDim

3. TimeDim

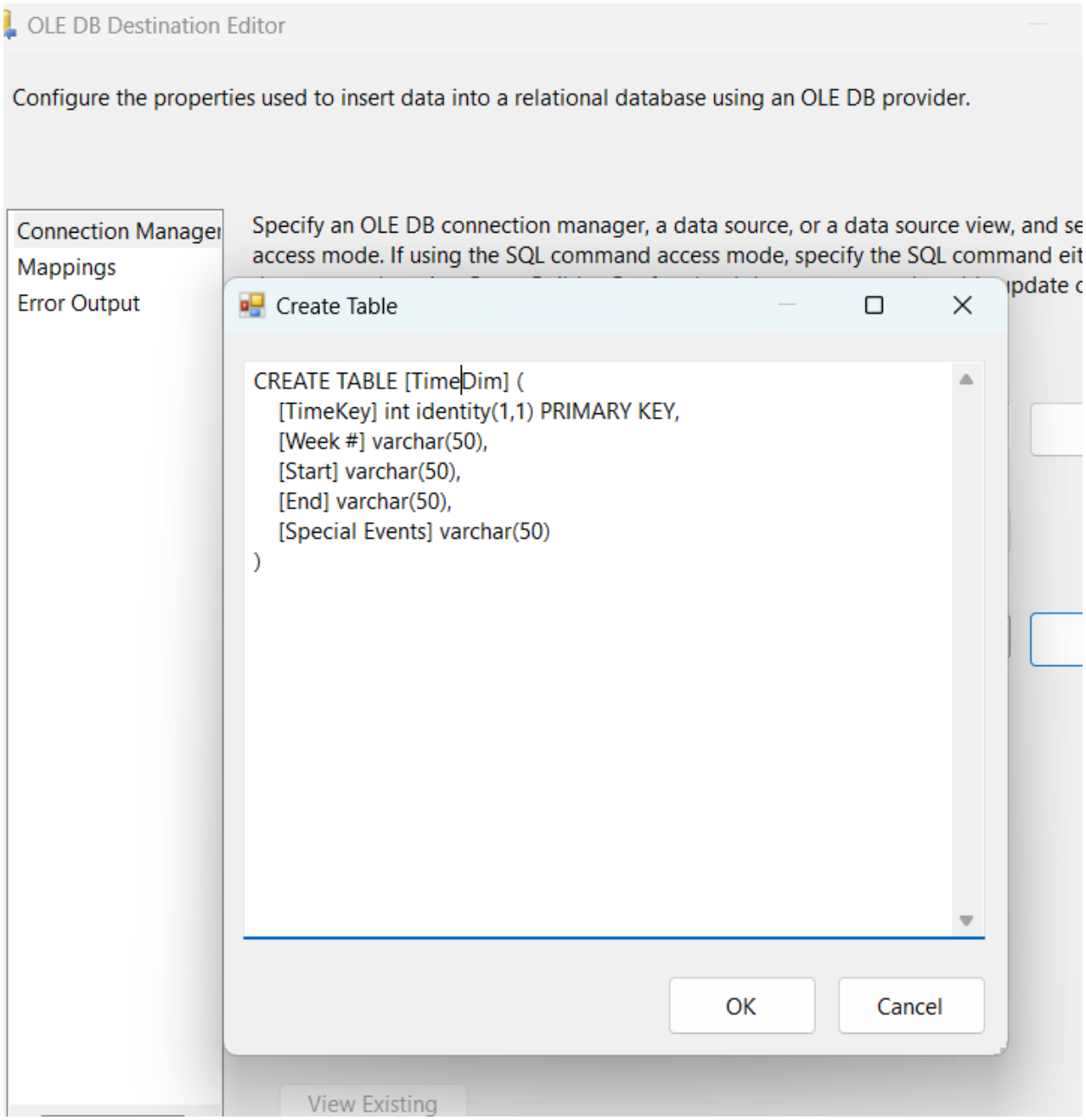


Fig: SQL Statement to create TimeDim

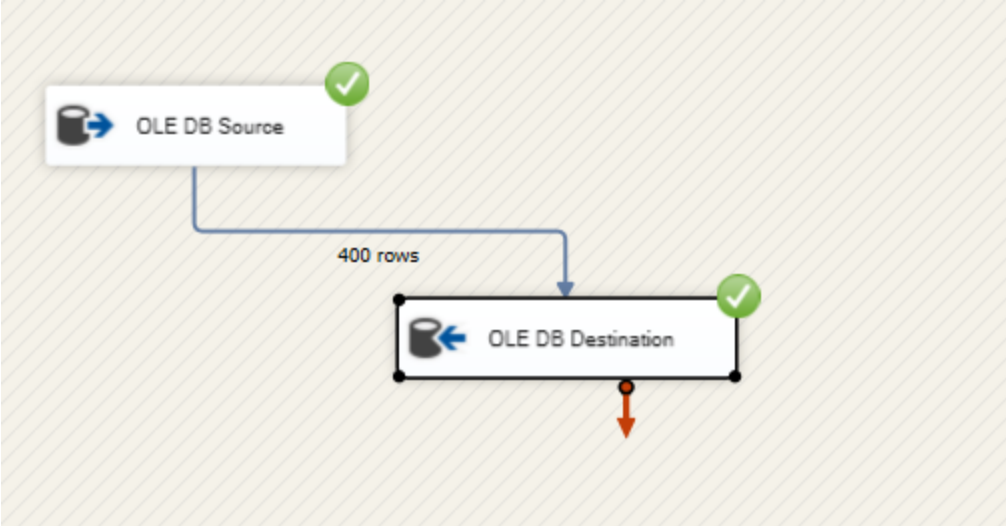


Fig: Execution of SSIS package for TimeDim

4. ProductDim

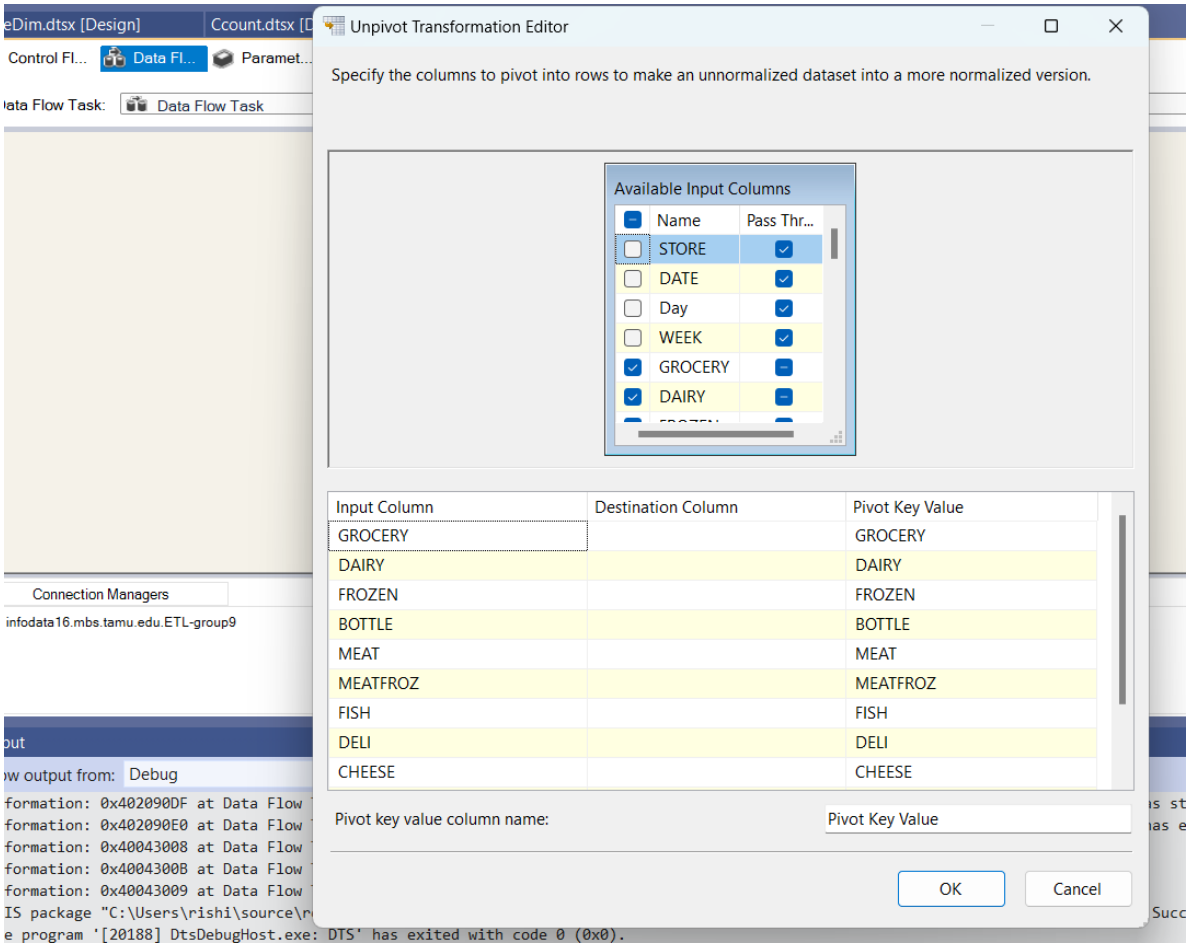


Fig: Mappings for ProductDim

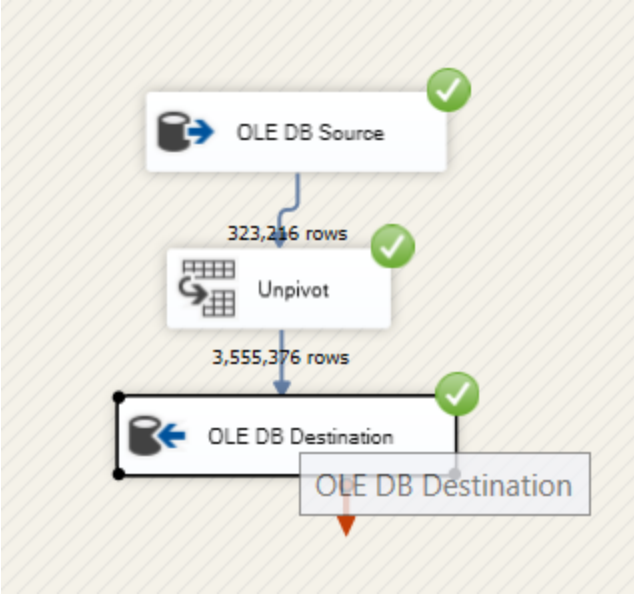
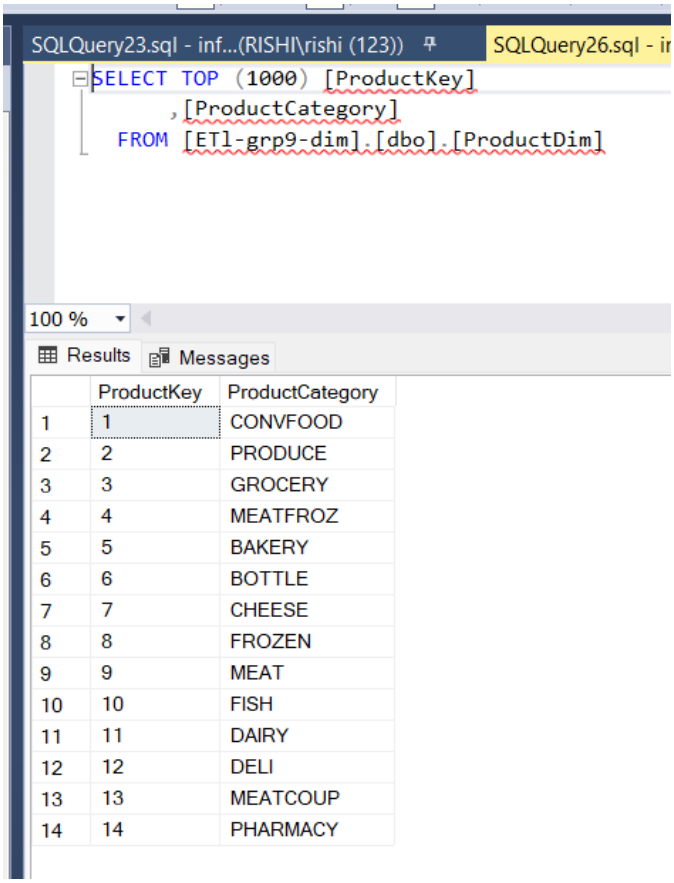


Fig: Execution of SSIS package for ProductDim



	ProductKey	ProductCategory
1	1	CONVFOOD
2	2	PRODUCE
3	3	GROCERY
4	4	MEATFROZ
5	5	BAKERY
6	6	BOTTLE
7	7	CHEESE
8	8	FROZEN
9	9	MEAT
10	10	FISH
11	11	DAIRY
12	12	DELI
13	13	MEATCOUP
14	14	PHARMACY

Fig: ProductDim output in SSMS

5. **SALESFACT table:**

SQL Statement:

```
SELECT unpcc.Sales, tim.TimeKey, dim.DemoKey, st.StoreKey, prod.ProductKey
FROM unpivot_ccount as unpcc
JOIN ProductDim prod ON unpcc.Category = prod.ProductCategory
JOIN DimStore st ON unpcc.STORE = st.StoreKey
JOIN DemographicDim dim ON unpcc.STORE = dim.STORE
JOIN TimeDim tim ON unpcc.WEEK = tim.Week_num;
```

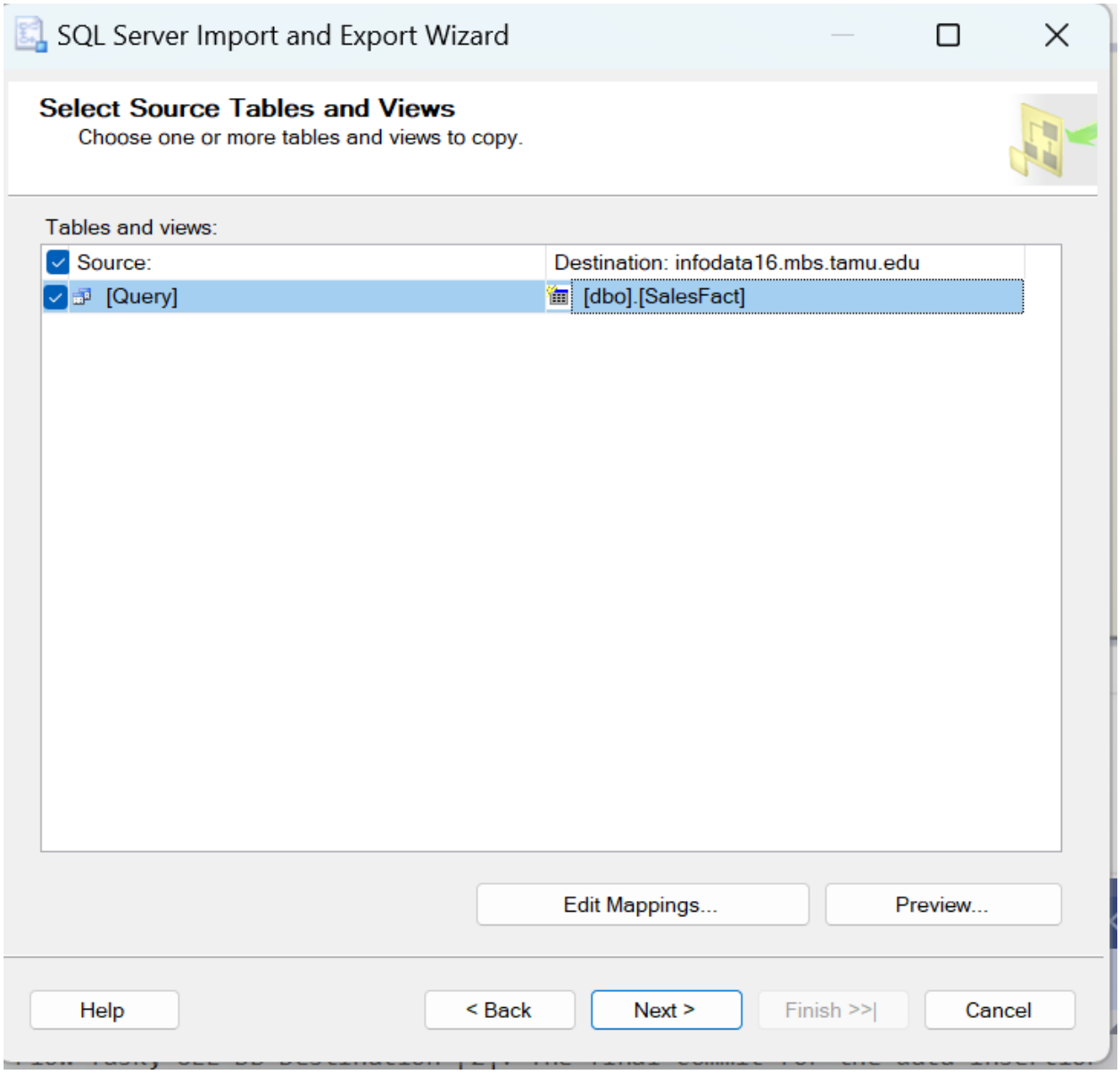


Fig: Import/Export wizard to create Sales Fact table

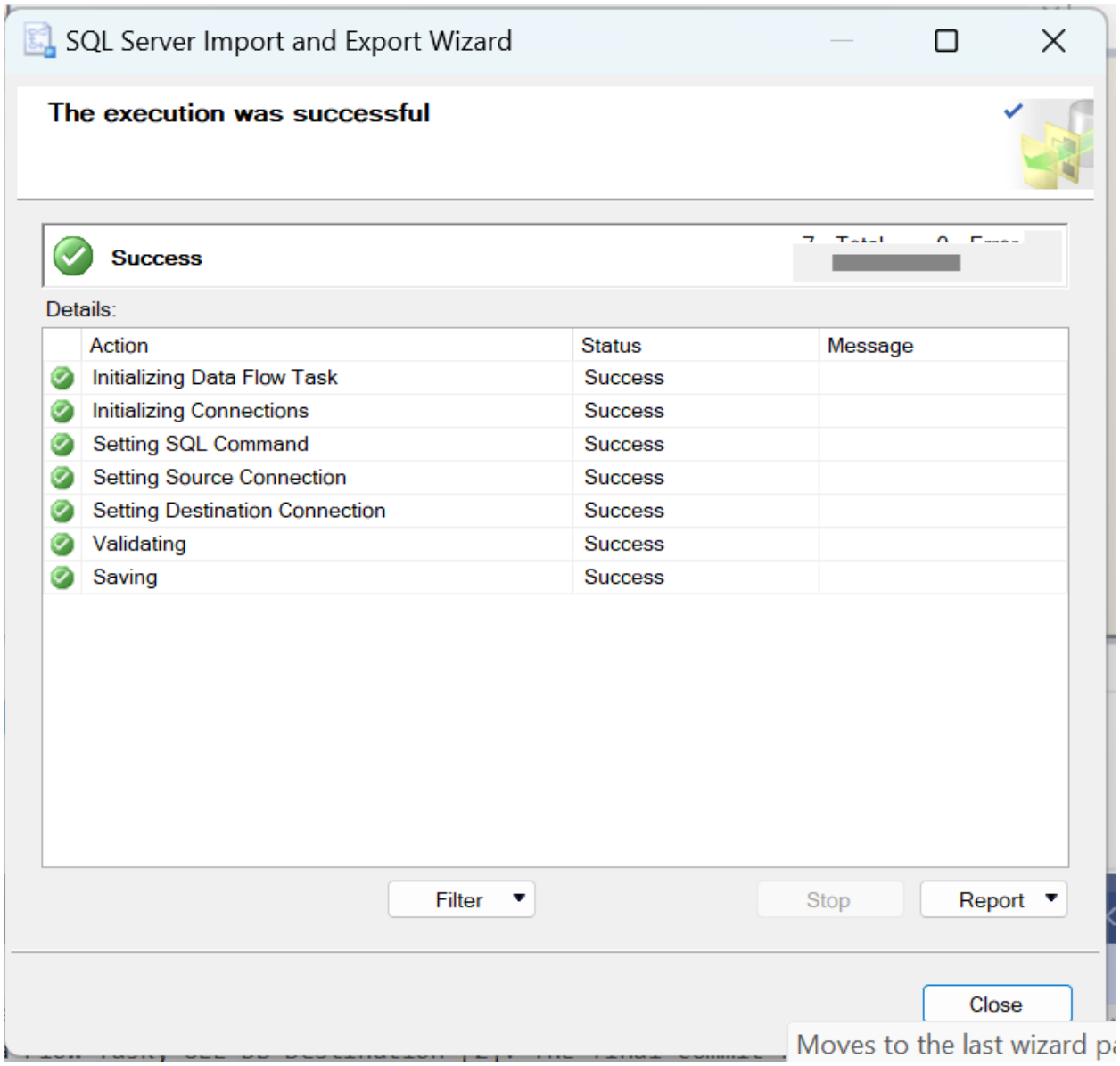


Fig: Completion of wizard to create fact table

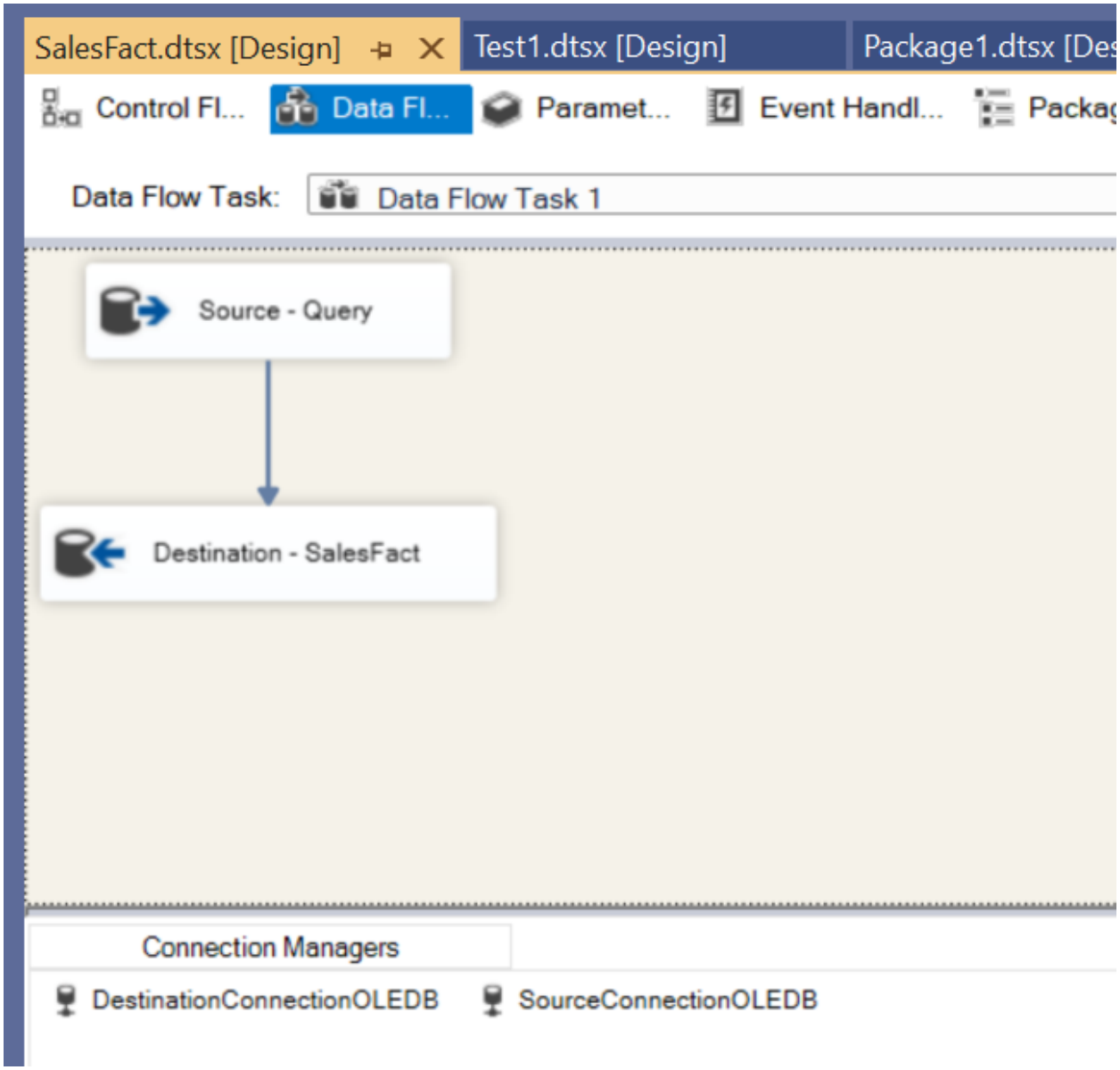
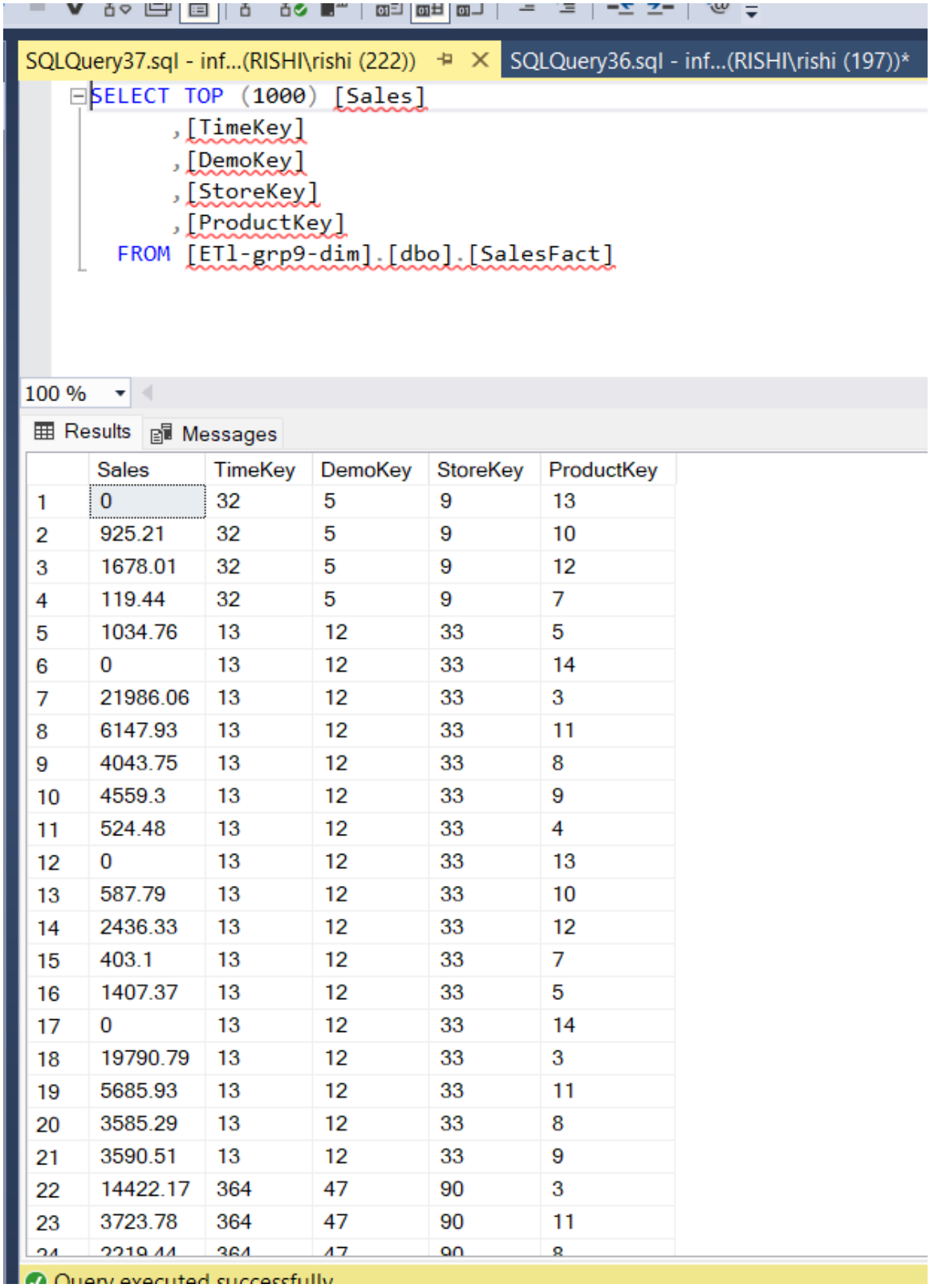


Fig: Query statement to join tables for Fact table



	Sales	TimeKey	DemoKey	StoreKey	ProductKey
1	0	32	5	9	13
2	925.21	32	5	9	10
3	1678.01	32	5	9	12
4	119.44	32	5	9	7
5	1034.76	13	12	33	5
6	0	13	12	33	14
7	21986.06	13	12	33	3
8	6147.93	13	12	33	11
9	4043.75	13	12	33	8
10	4559.3	13	12	33	9
11	524.48	13	12	33	4
12	0	13	12	33	13
13	587.79	13	12	33	10
14	2436.33	13	12	33	12
15	403.1	13	12	33	7
16	1407.37	13	12	33	5
17	0	13	12	33	14
18	19790.79	13	12	33	3
19	5685.93	13	12	33	11
20	3585.29	13	12	33	8
21	3590.51	13	12	33	9
22	14422.17	364	47	90	3
23	3723.78	364	47	90	11
24	2219.44	364	47	90	8

Fig: SalesFact table output in SSMS

6.4 List of temp Tables

Temp Tables	Queries to remove temp tables
Store_stg	DROP TABLE Store_stg;
Week_stg	DROP TABLE Week_stg;
Ccount_stg	DROP TABLE Ccount_stg;
Demo_stg	DROP TABLE Demo_stg;

References

1. https://www.guru99.com/star-schema-in-data-warehouse-modeling.html
2. https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/

Section 5: Business Intelligence Reporting

Business Questions, Tools and Rationale

1. What are store wise demographics across the total company sales for each store?

   Tools used: SSAS alone.

   Rationale: This business question tries to discover how different age groups contribute to the business of DFF. By looking at data in the age column in the demographic dimension and correlating it with the sales data from SalesFact, we can get an understanding of how different age groups contribute to the business of DFF.

   We can see that Age group 10-59 contributes the most in every store. After learning about different age demographics, DFF can perform targeted marketing for specific age groups which contribute the most. This can lead to increased sales and increased customer satisfaction rates.

2. Which DFF stores have experienced significant changes in sales performance over the years, and what actions have been taken to address these changes?

   Tools used: Combination of SSAS and SSRS.

   Rationale: This business question aims to identify the sales trends for stores over a period of time. We will correlate Sales data from SalesFact with time data from Week_num in TimeDim to look at the sales of stores over a number of years. This will help in examining which stores have had the most change in sales, whether positive or negative. This analysis will help us understand the trend and patterns over the years for different stores in DFF to help us make better data driven decisions.

   If we compare 3 stores- 78, 80 and 83, we can see that store 80 has the highest drop in sales from 1990 to 1997. By analyzing which stores are seeing maximum changes, DFF can devise strategies which can help stores with massive drop in sales, bounce back.

3. What do the sales patterns for bakery products look like over a span of three years, categorized into low, medium, and high-price tier stores across multiple cities?

   Tools used: SSRS alone

Rationale: This business question aims to look at sales of bakery products in different tiered stores in different cities. By looking at sales data in SalesFact and looking at store categories (high, medium, low), from the PriceTier column in StoreDim table, we can see which tiered stores in which cities have the highest bakery sales.

If we compare stores in cities, we find that high price stores in Chicago and Evanston have the highest bakery sales in 1990, 1993 and 1996. This data helps in understanding the market dynamics of different economic segments and make changes accordingly. This analysis helps us understand the impact of the different stores in the different PriceTier brackets to the contribution in different product categories and specifically bakery products.

4. What are the top-performing product categories in terms of sales revenue for each DFF branch over the years?

   Tools used: PowerBi

   Rationale: This question aims to identify the highest selling product categories in terms of sales, across DFF stores. We will take sales data from SalesFact and correlate it with the Product category in ProductDim.

   If we look at stores- 2, 4 and 5, we can see that for all 3 stores, the grocery sales are the highest for all of them. Looking at the sales of particular product categories, will help DFF focus on inventory management and resource allocation. This analysis helps us understand the contribution of different product categories to the overall sales in different stores so as to help us visualize the best and worst performing product categories.

5. Which stores attract people who earn below the poverty line and have high value income thresholds?

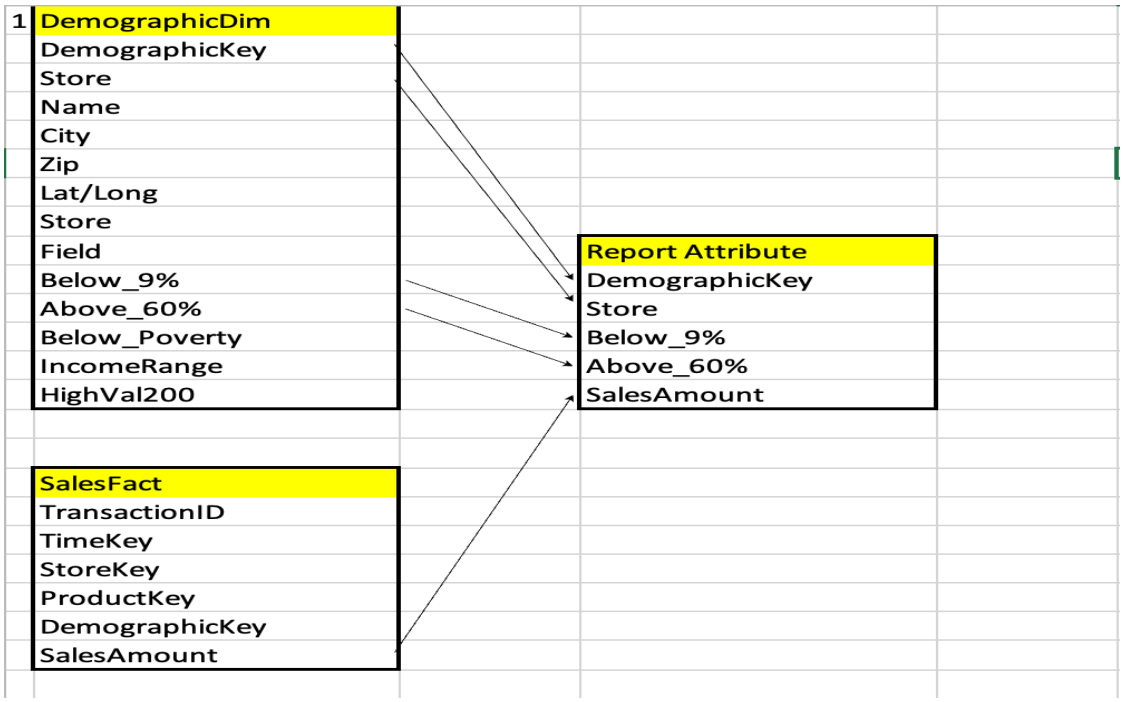
   Tools used: PowerBi

   Rationale: This question is mainly to understand the socio economic status of customers of DFF. We want to see which stores are attracting customers which are below the poverty line and have high income thresholds. We will use IncomeRange from DemographicDim and correlate it with sales data from SalesFact.

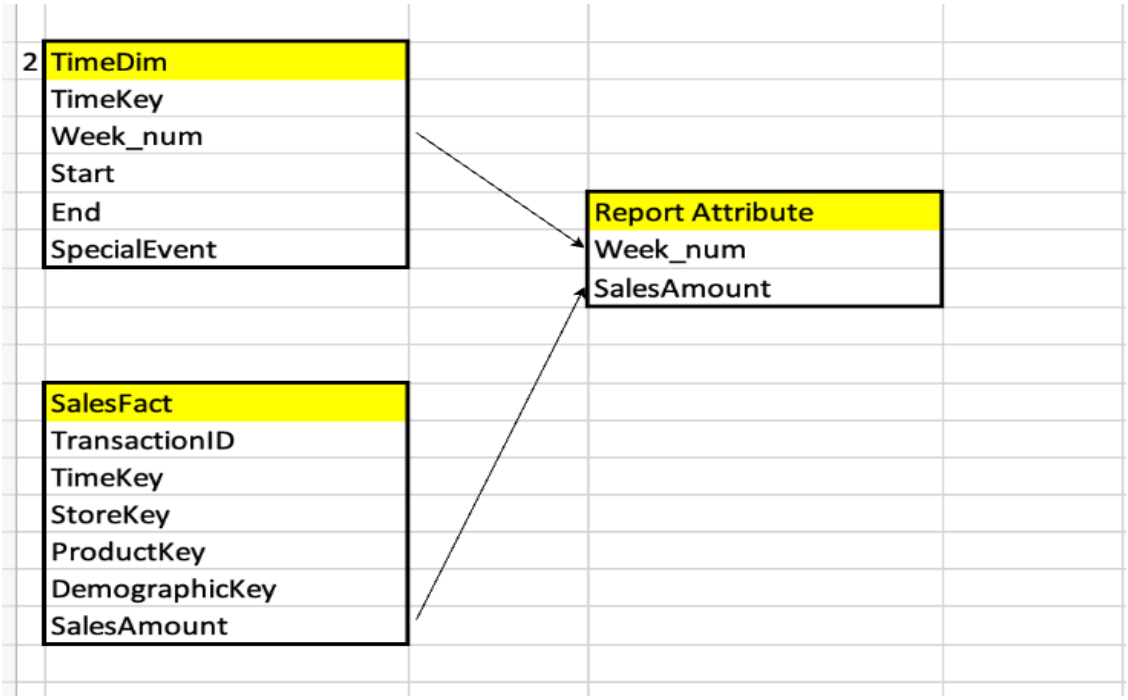
   We can see that store 12 has the highest percentage of customers who are below the poverty line. After visualizing we can also see that stores with more customers having income over 150,000 and 200,000 have higher sales too. Understanding the income profile of customers helps in understanding the customers better. This analysis would also help us decide the marketing strategies better as well since different products would appeal to different customer profiles and hence a more sophisticated marketing approach can be created from these insights.

MAPPING ATTRIBUTES FROM INDEPENDENT DATA MARTS

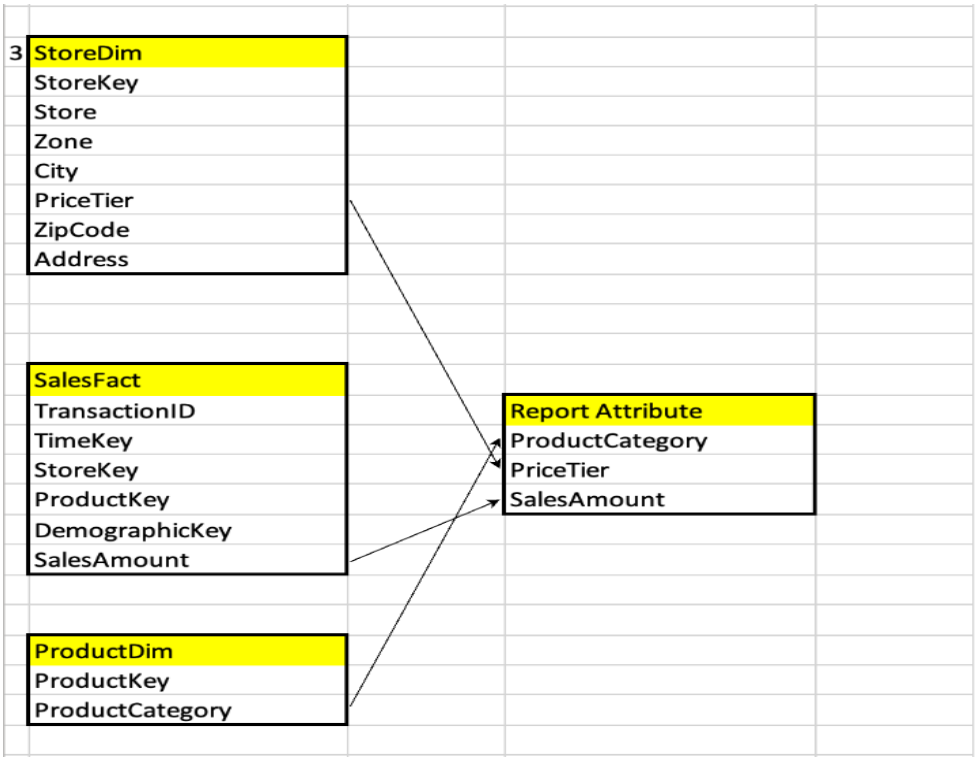
Q1.What are store wise demographics across the total company sales for each store?



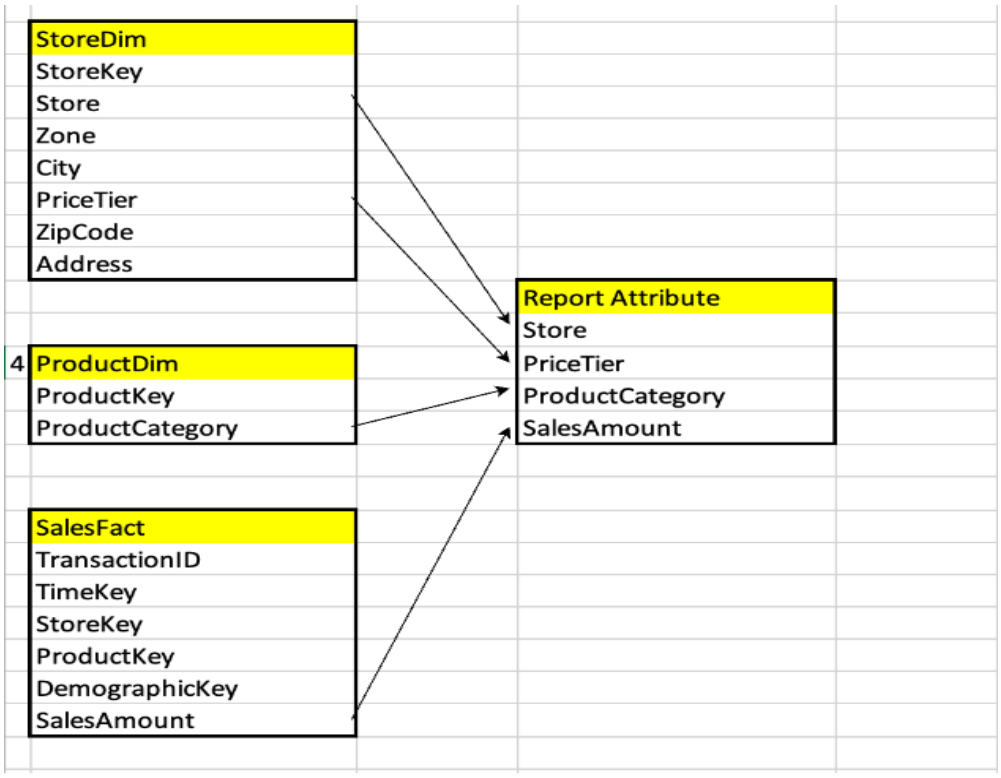
Q2.Which DFF stores have experienced significant changes in sales performance over the years, and what actions have been taken to address these changes?



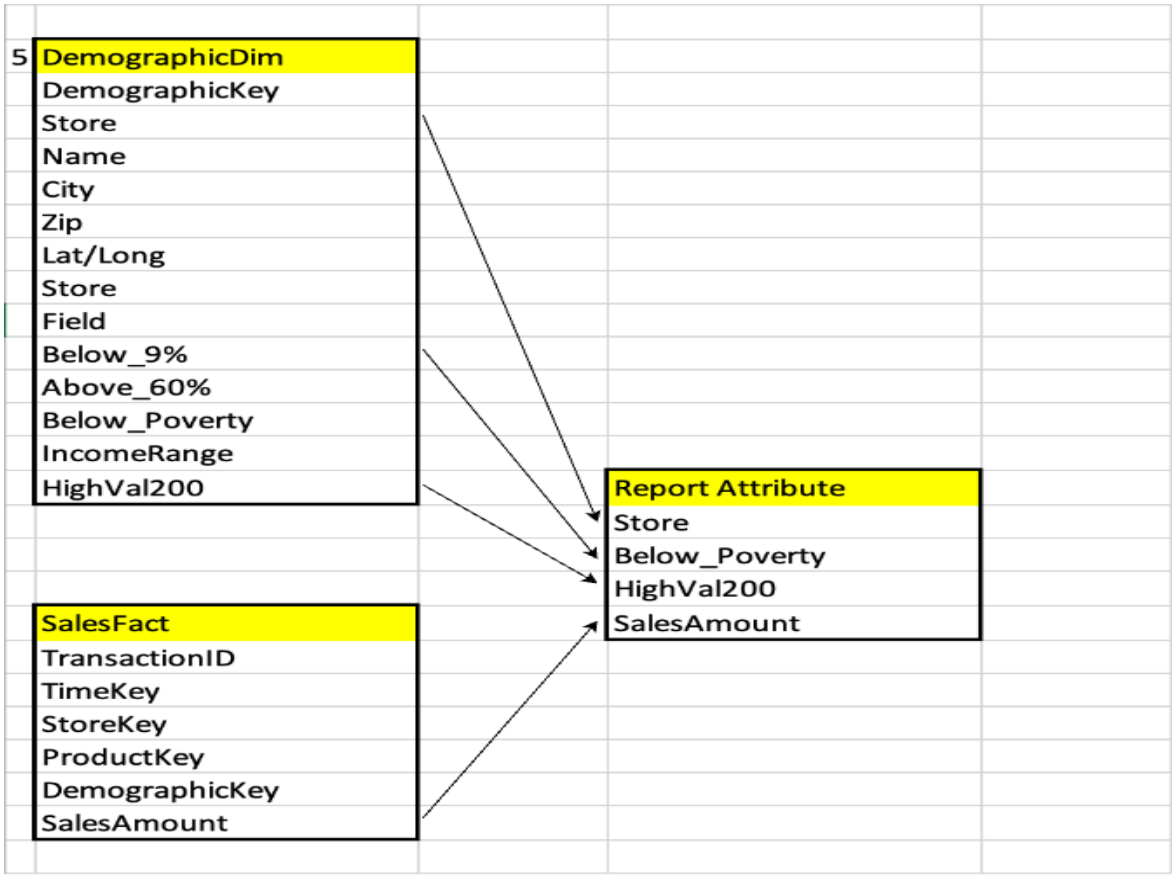
Q3. What do the sales patterns for bakery products look like over a span of three years, categorized into low, medium, and high-price tier stores across multiple cities?



Q4. What are the top-performing product categories in terms of sales revenue for each DFF branch over the years?



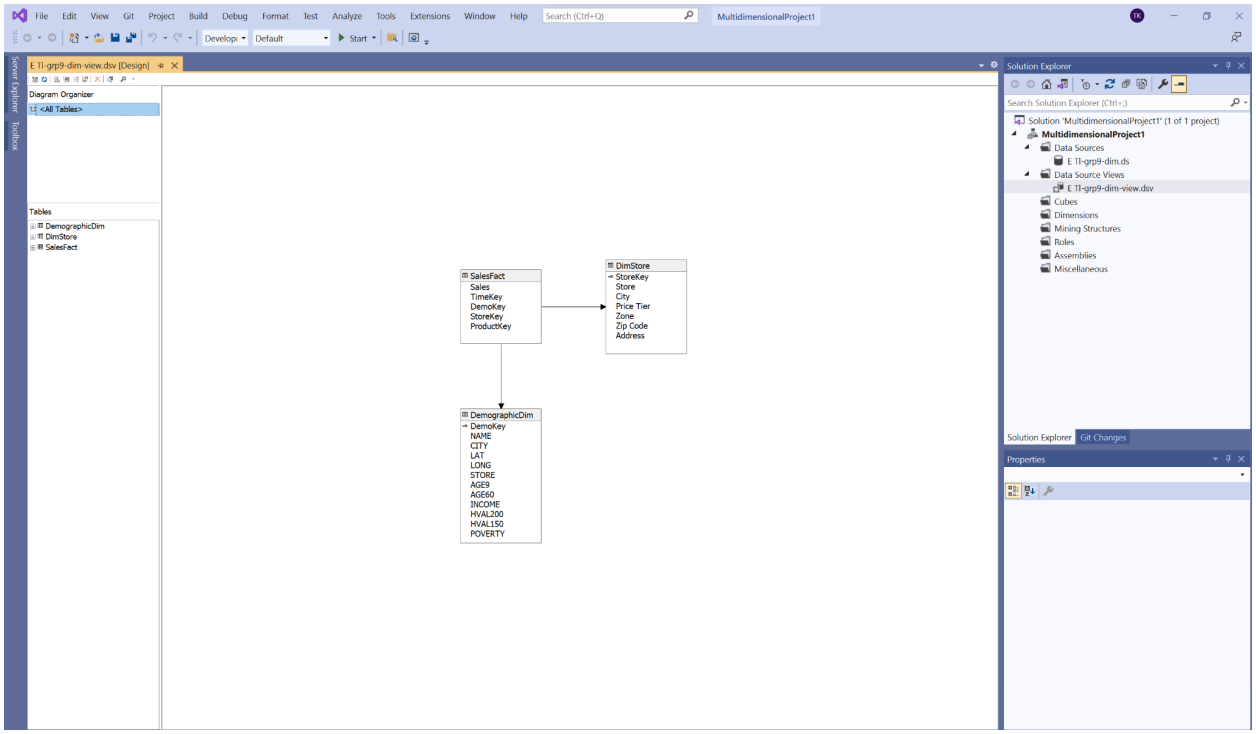
Q5. Which stores attract people who earn below the poverty line and have high value income thresholds?



Reports:

SSAS

BQ1: What are store wise demographics across the total company sales for each store?

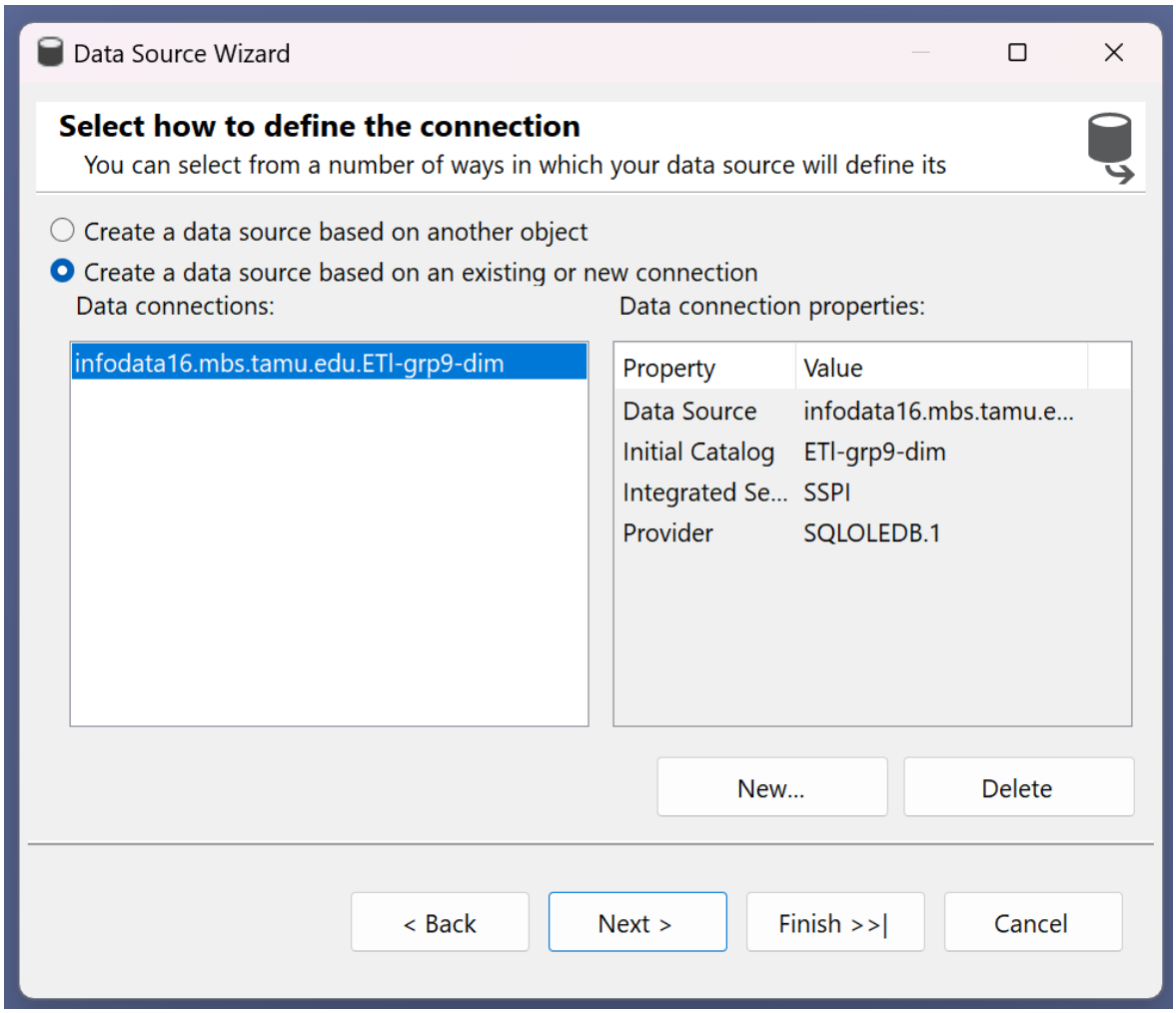

File Edit View Git Project Build Debug Format Test Analyze Tools Extensions Window Help
MultidimensionalProject1
E TI-grp9-dim-view.dsv [Design]
Diagram Organizer
<All Tables>
Tables
DemographicDim
DimStore
SalesFact
SalesFact
Sales
TimeKey
DemoKey
StoreKey
ProductKey
DimStore
StoreKey
Store
City
Price Tier
Zone
Zip Code
Address
DemographicDim
DemoKey
NAME
CITY
LAT
LONG
STORE
AGE9
AGE60
INCOME
HVAL200
HVAL150
POVERTY
Solution Explorer
Solution 'MultidimensionalProject1' (1 of 1 project)
MultidimensionalProject1
Data Sources
E TI-grp9-dim.ds
Data Source Views
E TI-grp9-dim-view.dsv
Cubes
Dimensions
Mining Structures
Roles
Assemblies
Miscellaneous
Solution Explorer
Git Changes
Properties

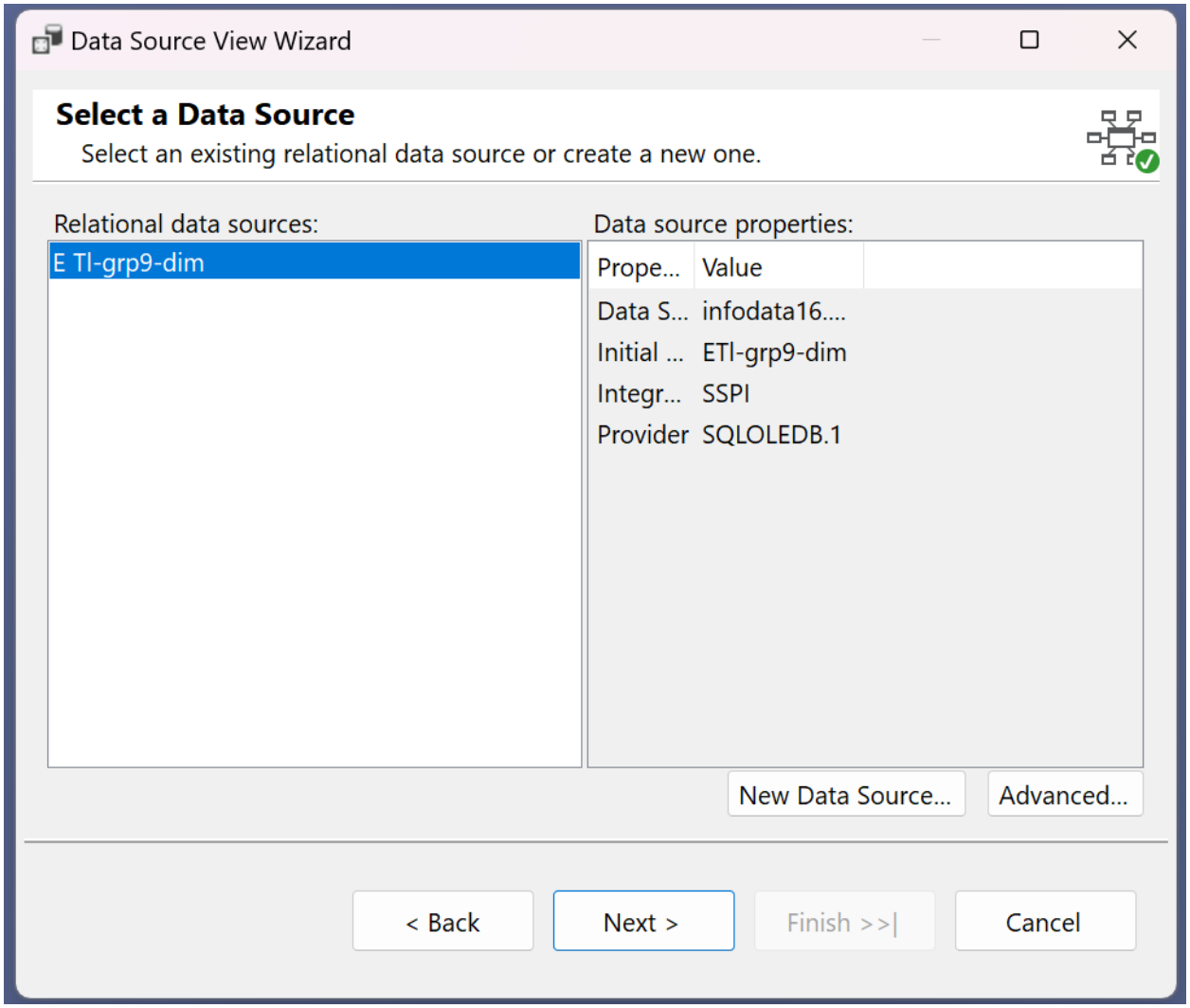
Explore ETl-grp9-SSAS Table | E Tl-grp9-dim-view.dsv [Design]*

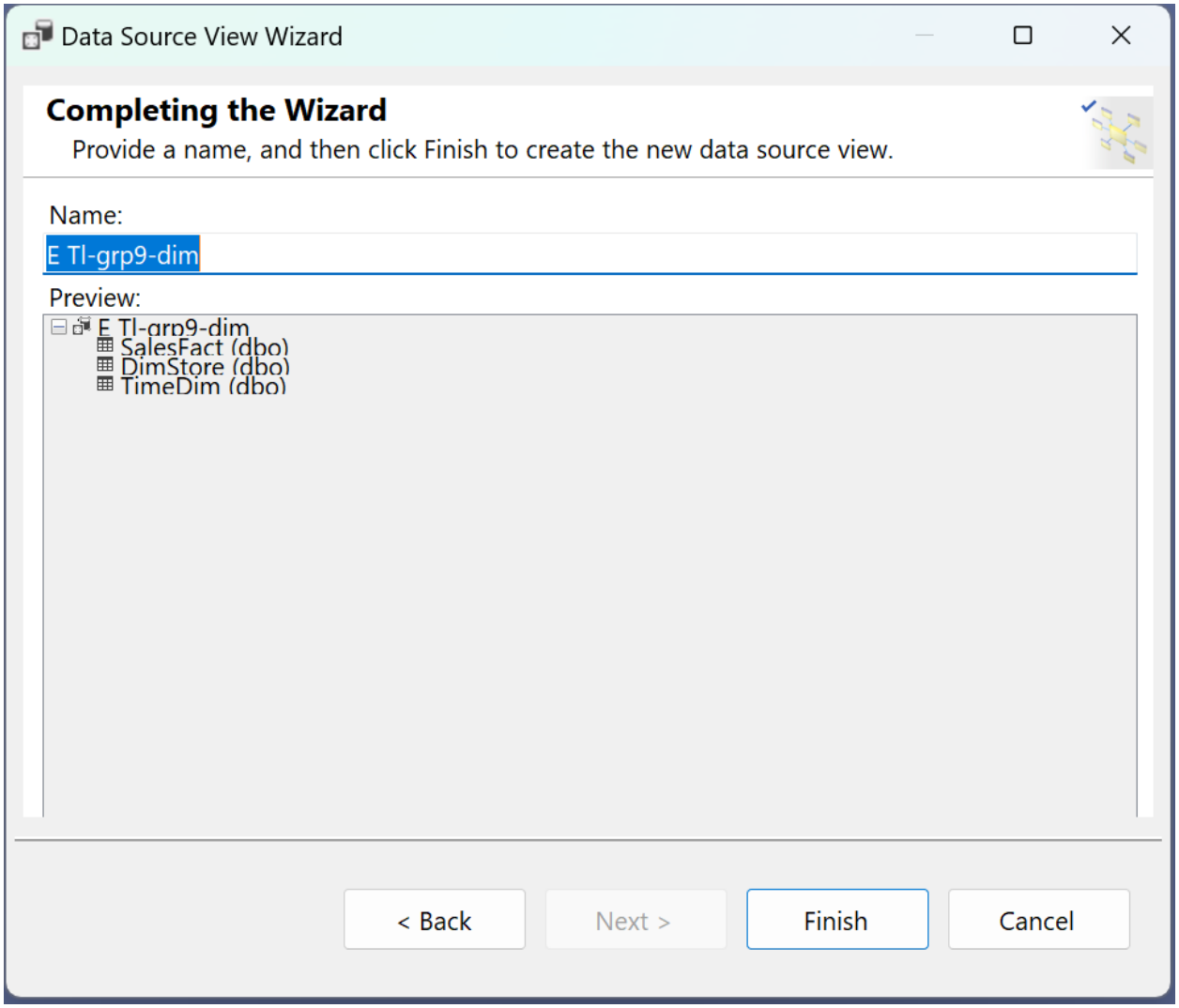
Table

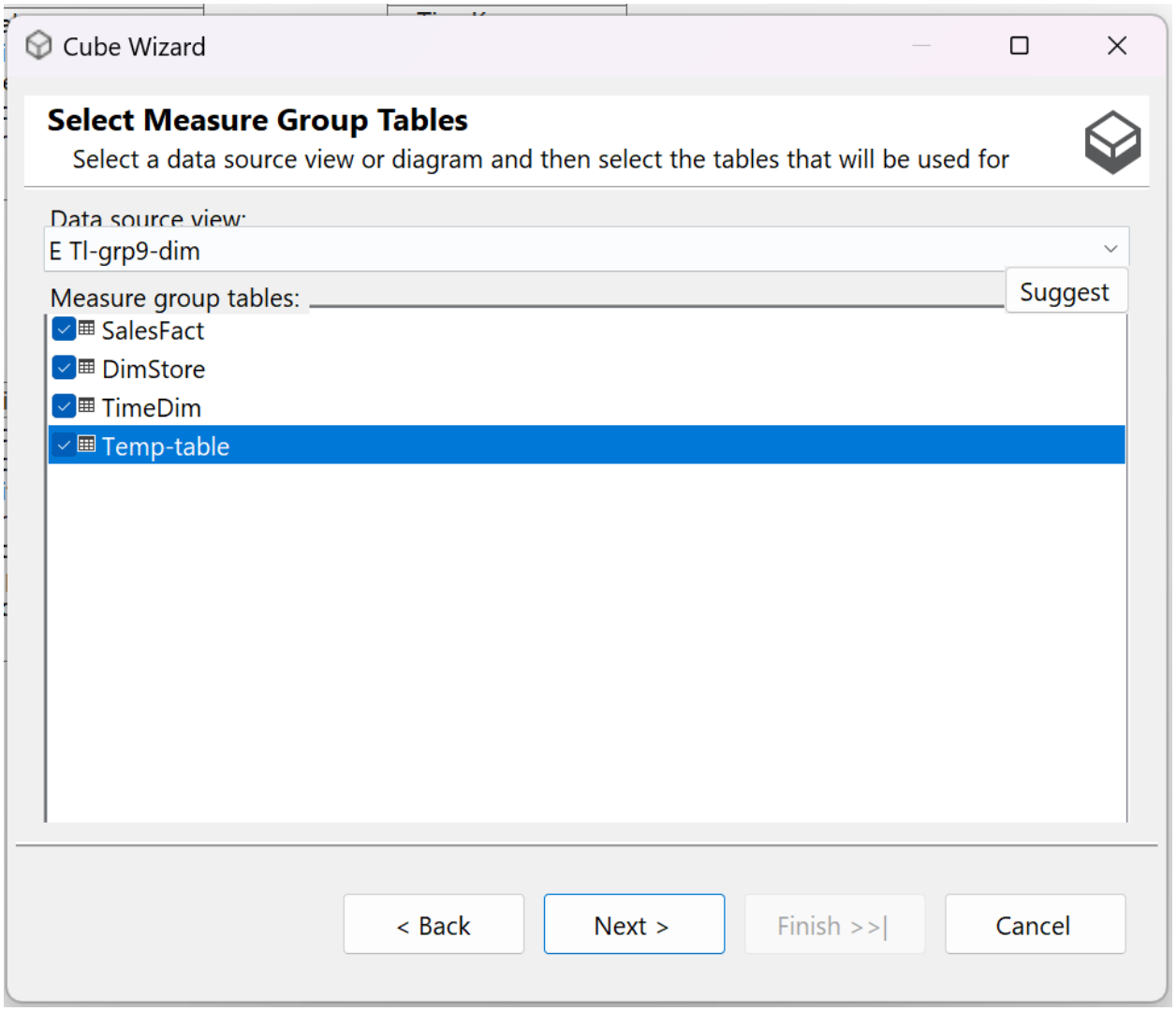
Sales	DemoKey	AGE9	AGE60	STORE
0	5	0.103503097	0.269119018	9
925.21	5	0.103503097	0.269119018	9
1678.01	5	0.103503097	0.269119018	9
119.44	5	0.103503097	0.269119018	9
1034.76	12	0.046070917	0.134169966	33
0	12	0.046070917	0.134169966	33
21986.06	12	0.046070917	0.134169966	33
6147.93	12	0.046070917	0.134169966	33
4043.75	12	0.046070917	0.134169966	33
4559.3	12	0.046070917	0.134169966	33
524.48	12	0.046070917	0.134169966	33
0	12	0.046070917	0.134169966	33
587.79	12	0.046070917	0.134169966	33
2436.33	12	0.046070917	0.134169966	33
403.1	12	0.046070917	0.134169966	33
1407.37	12	0.046070917	0.134169966	33
0	12	0.046070917	0.134169966	33
19790.79	12	0.046070917	0.134169966	33
5685.93	12	0.046070917	0.134169966	33
3585.29	12	0.046070917	0.134169966	33
3590.51	12	0.046070917	0.134169966	33
14422.17	47	0.129722571	0.225219573	90
3723.78	47	0.129722571	0.225219573	90
2219.44	47	0.129722571	0.225219573	90
4987.54	47	0.129722571	0.225219573	90
426.92	47	0.129722571	0.225219573	90
-7	47	0.129722571	0.225219573	90
363.71	47	0.129722571	0.225219573	90
1673.88	47	0.129722571	0.225219573	90
92.79	47	0.129722571	0.225219573	90
1380.04	47	0.129722571	0.225219573	90
0	47	0.129722571	0.225219573	90
15008.22	47	0.129722571	0.225219573	90
3596.64	47	0.129722571	0.225219573	90

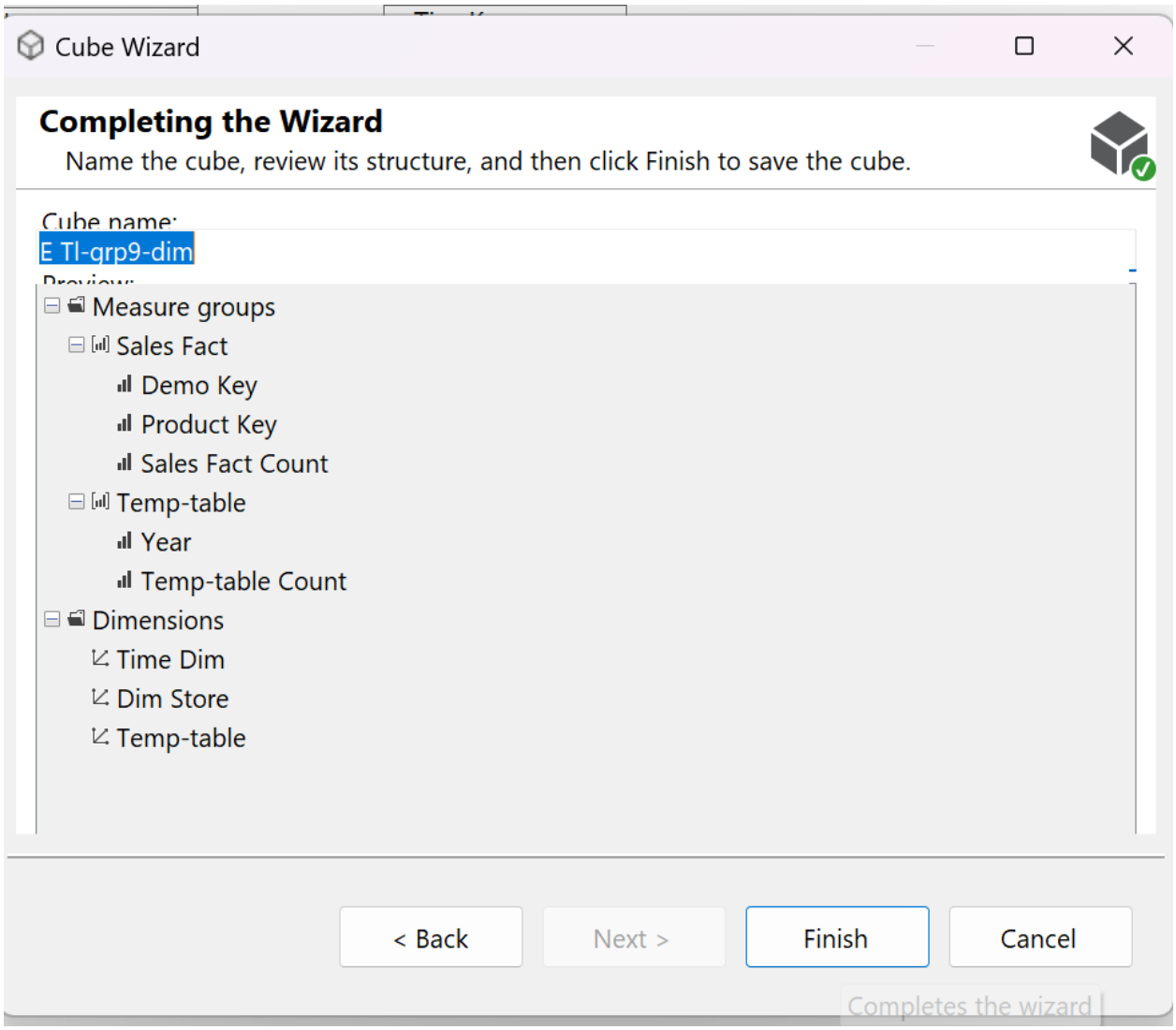
BQ 2.
SSAS Cube+SSRS

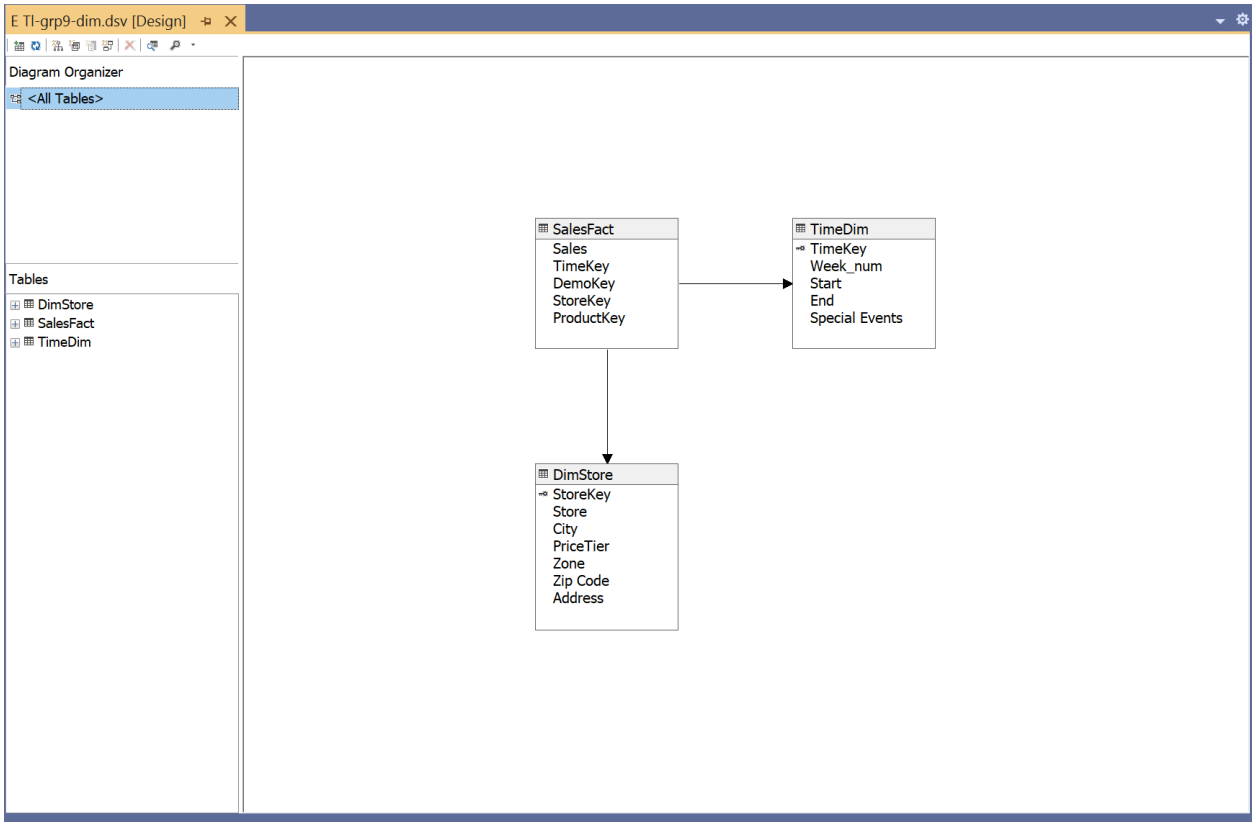


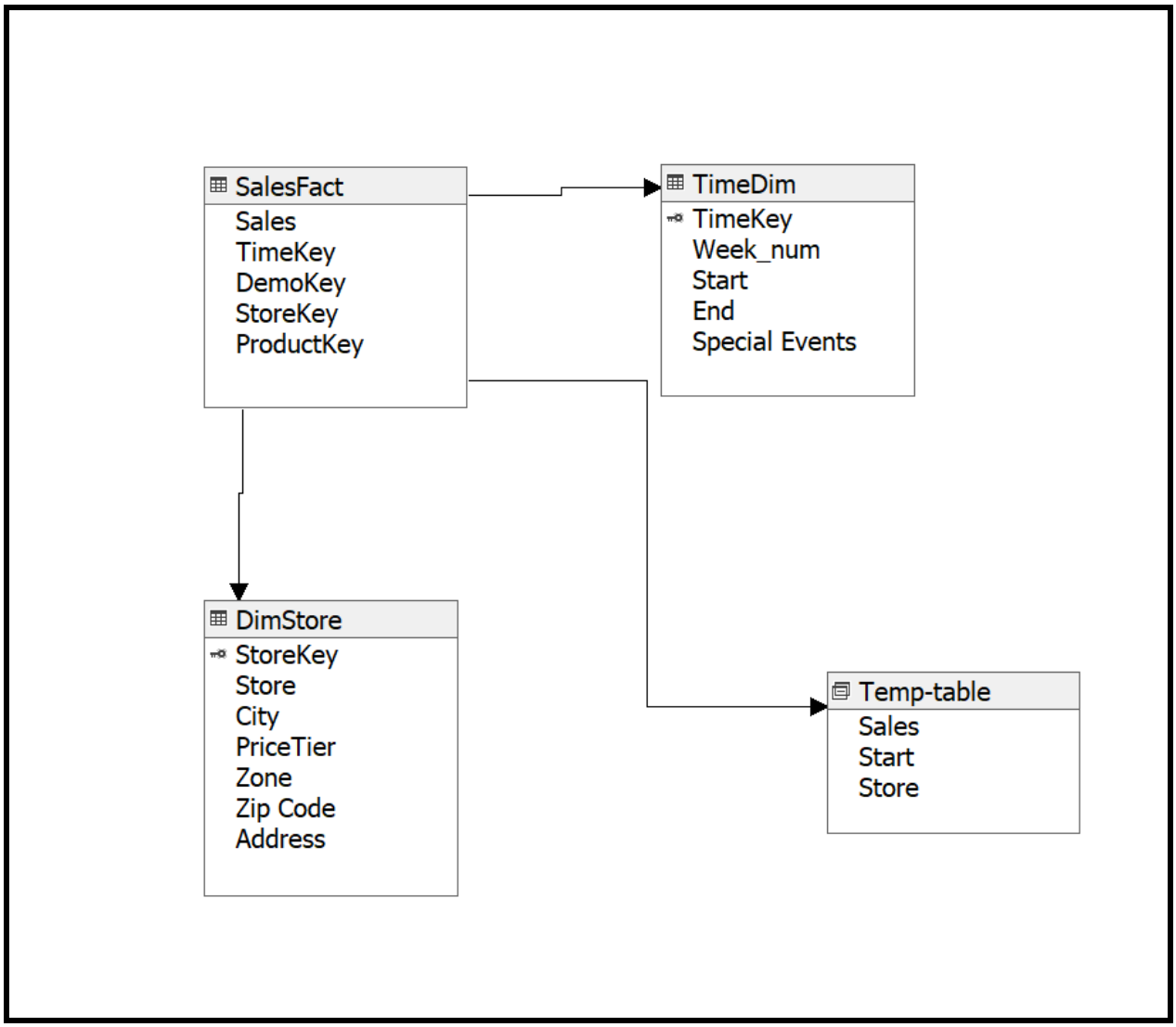

Data Source View Wizard
Select a Data Source
Select an existing relational data source or create a new one.
Relational data sources:
E TI-grp9-dim
Data source properties:
Prope...
Value
Data S...
infodata16....
Initial ...
ETl-grp9-dim
Integr...
SSPI
Provider
SQLOLEDB.1
New Data Source...
Advanced...
< Back
Next >
Finish >>|
Cancel

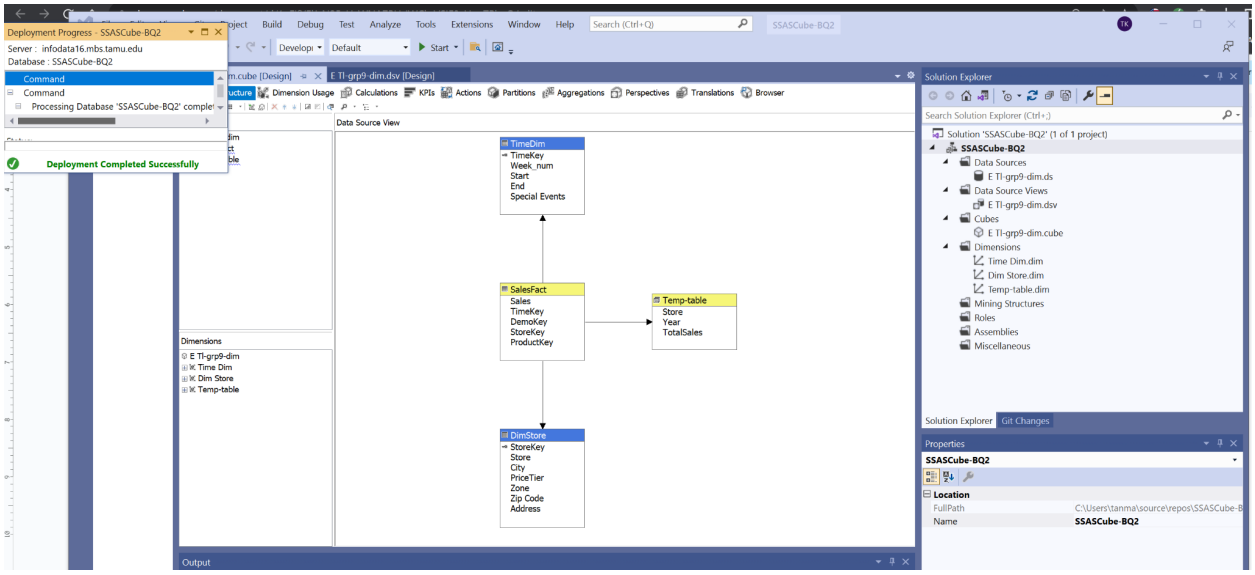

Data Source View Wizard
Completing the Wizard
Provide a name, and then click Finish to create the new data source view.
Name:
E TI-grp9-dim
Preview:
E TI-grp9-dim
SalesFact (dbo)
DimStore (dbo)
TimeDim (dbo)
< Back
Next >
Finish
Cancel

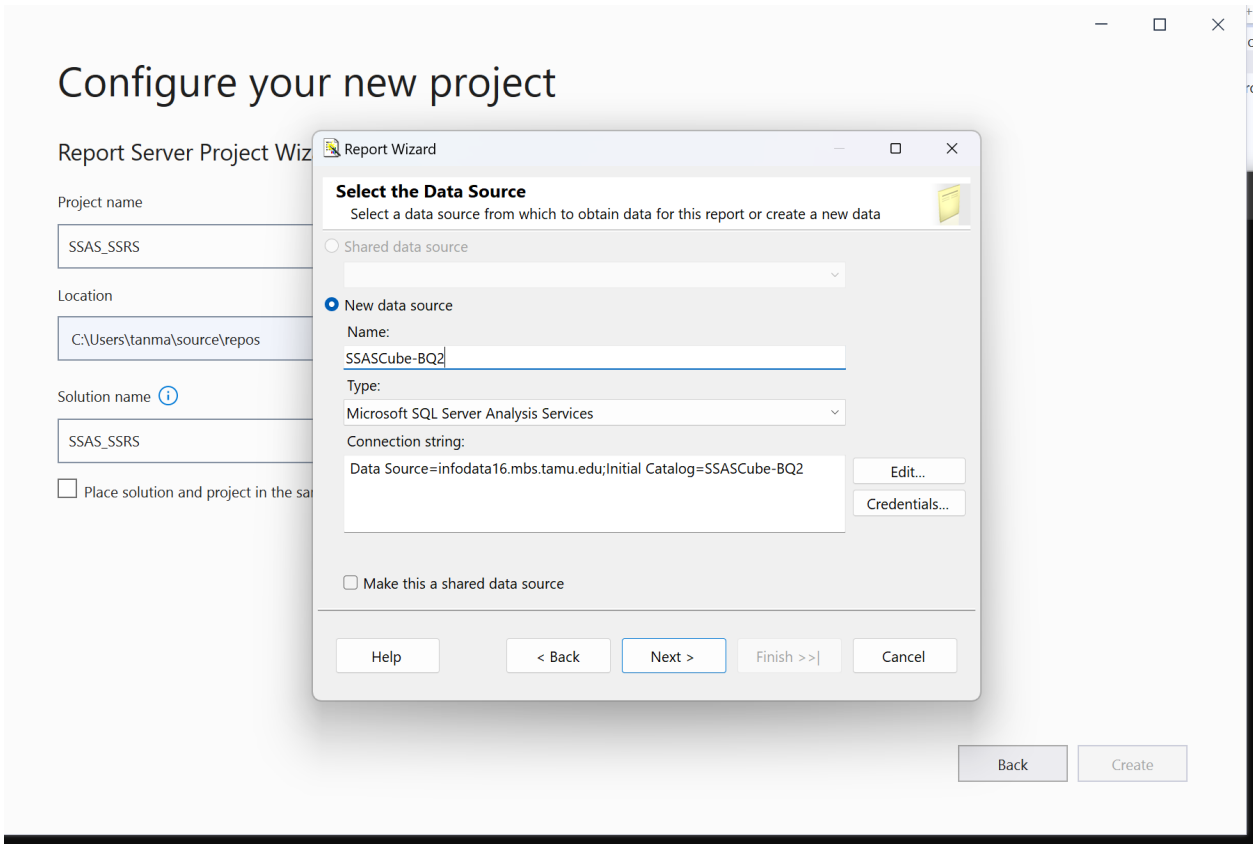

Cube Wizard
Select Measure Group Tables
Select a data source view or diagram and then select the tables that will be used for
Data source view:
E Tl-grp9-dim
Measure group tables:
Suggest
SalesFact
DimStore
TimeDim
Temp-table
< Back
Next >
Finish >>|
Cancel

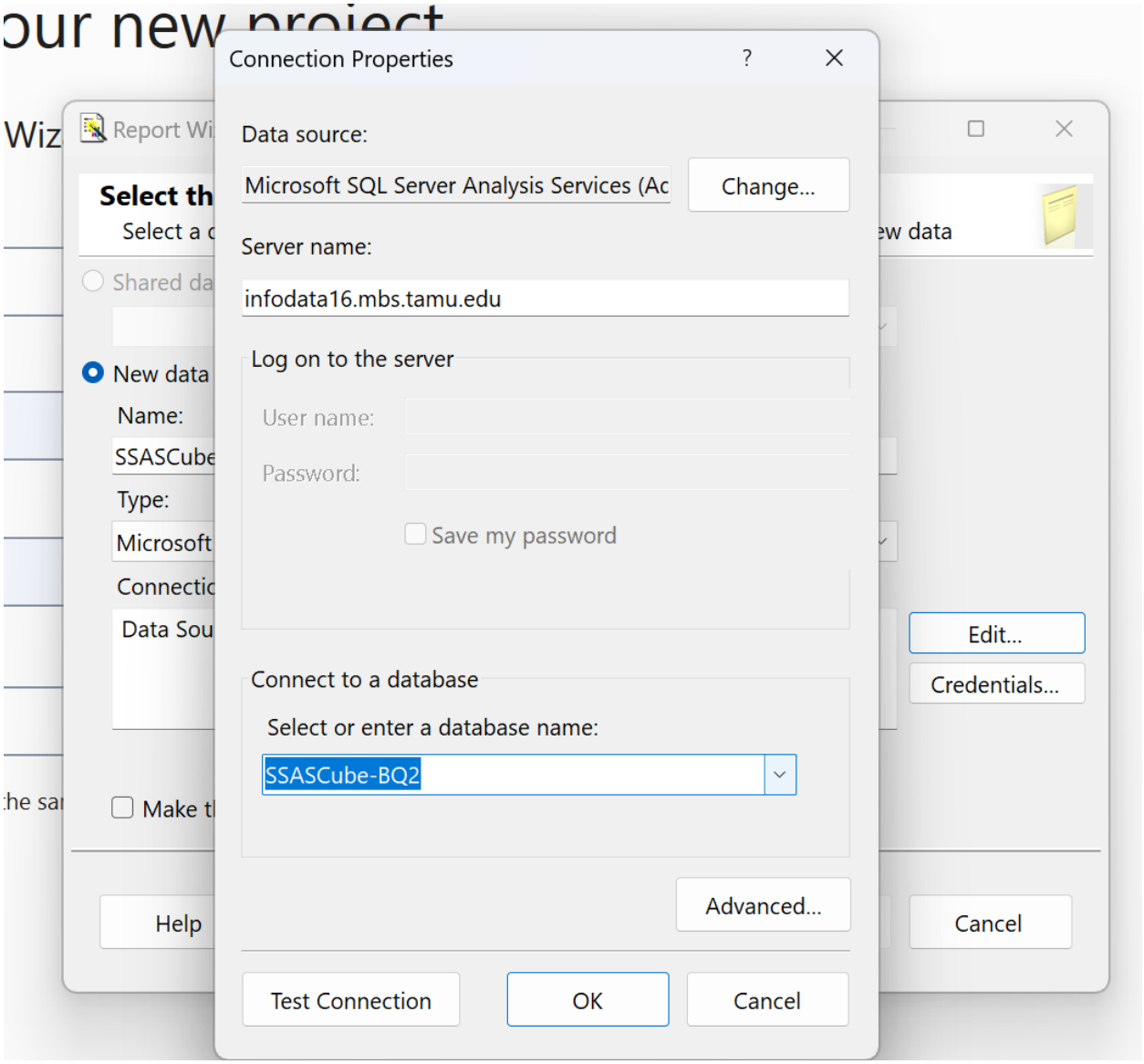

Cube Wizard
Completing the Wizard
Name the cube, review its structure, and then click Finish to save the cube.
Cube name:
E TI-grp9-dim
Measure groups
Sales Fact
Demo Key
Product Key
Sales Fact Count
Temp-table
Year
Temp-table Count
Dimensions
Time Dim
Dim Store
Temp-table
< Back
Next >
Finish
Cancel
Completes the wizard

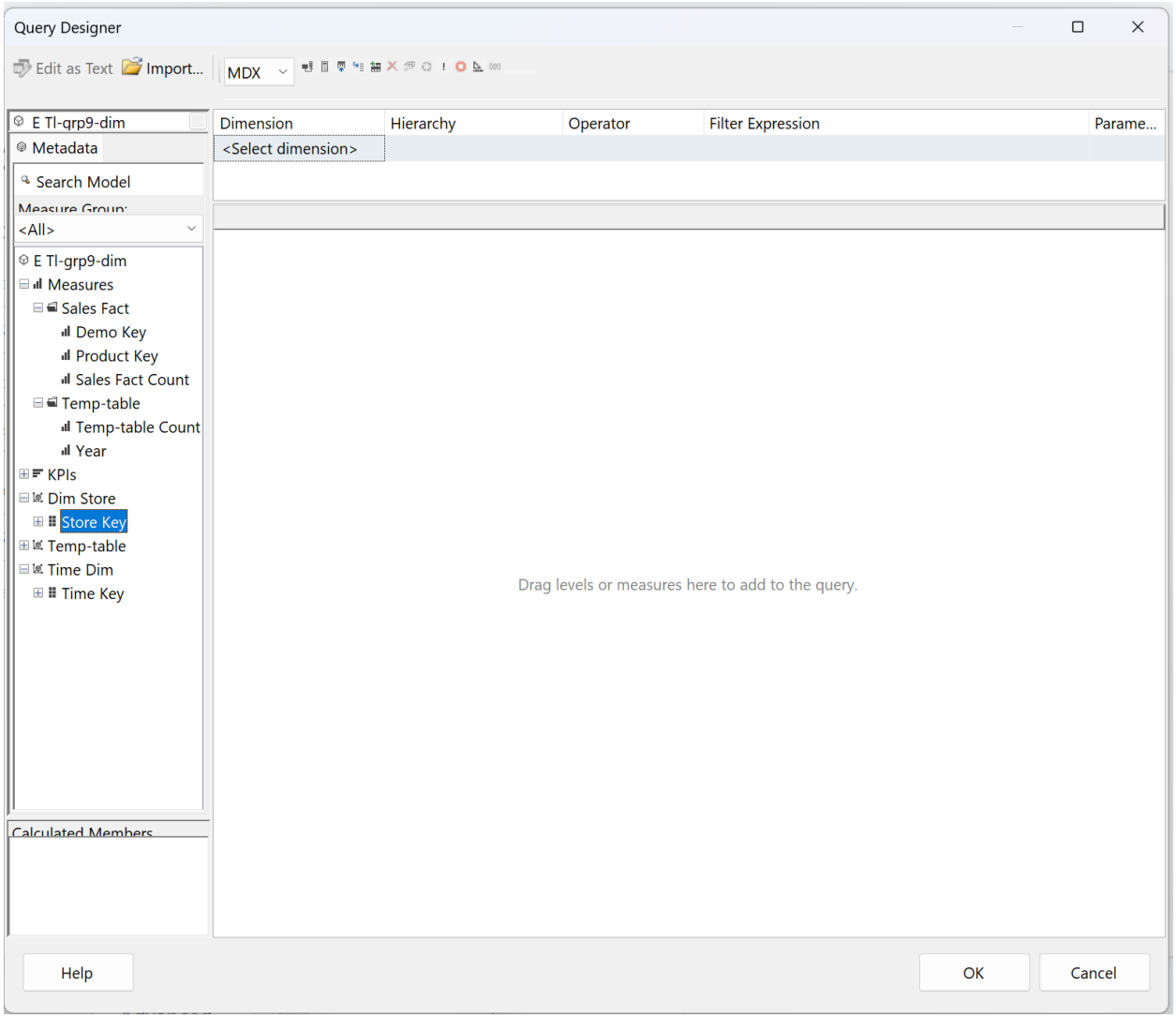

E TI-grp9-dim.dsv [Design]
Diagram Organizer
<All Tables>
Tables
DimStore
SalesFact
TimeDim
SalesFact
Sales
TimeKey
DemoKey
StoreKey
ProductKey
TimeDim
TimeKey
Week_num
Start
End
Special Events
DimStore
StoreKey
Store
City
PriceTier
Zone
Zip Code
Address


SalesFact
Sales
TimeKey
DemoKey
StoreKey
ProductKey
TimeDim
TimeKey
Week_num
Start
End
Special Events
DimStore
StoreKey
Store
City
PriceTier
Zone
Zip Code
Address
Temp-table
Sales
Start
Store


Deployment Progress - SSASCube-BQ2
Server : infodata16.mbs.tamu.edu
Database : SSASCube-BQ2
Command
Command
Processing Database 'SSASCube-BQ2' complet
Deployment Completed Successfully
Data Source View
TimeDim
TimeKey
Week_num
Start
End
Special Events
SalesFact
Sales
TimeKey
DemoKey
StoreKey
ProductKey
Temp-table
Store
Year
TotalSales
DimStore
StoreKey
Store
City
PriceTier
Zone
Zip Code
Address
Dimensions
E TI-grp9-dim
Time Dim
Dim Store
Temp-table
Solution Explorer
Solution 'SSASCube-BQ2' (1 of 1 project)
SSASCube-BQ2
Data Sources
E TI-grp9-dim.ds
Data Source Views
E TI-grp9-dim.dsv
Cubes
E TI-grp9-dim.cube
Dimensions
Time Dim.dim
Dim Store.dim
Temp-table.dim
Mining Structures
Roles
Assemblies
Miscellaneous
Properties
SSASCube-BQ2
Location
FullPath
Name
SSASCube-BQ2
Output


Configure your new project
Report Server Project Wiz
Project name
SSAS_SSRS
Location
C:\Users\tanma\source\repos
Solution name
SSAS_SSRS
Place solution and project in the sa
Report Wizard
Select the Data Source
Select a data source from which to obtain data for this report or create a new data
Shared data source
New data source
Name:
SSASCube-BQ2
Type:
Microsoft SQL Server Analysis Services
Connection string:
Data Source=infodata16.mbs.tamu.edu;Initial Catalog=SSASCube-BQ2
Edit...
Credentials...
Make this a shared data source
Help
< Back
Next >
Finish >>|
Cancel
Back
Create

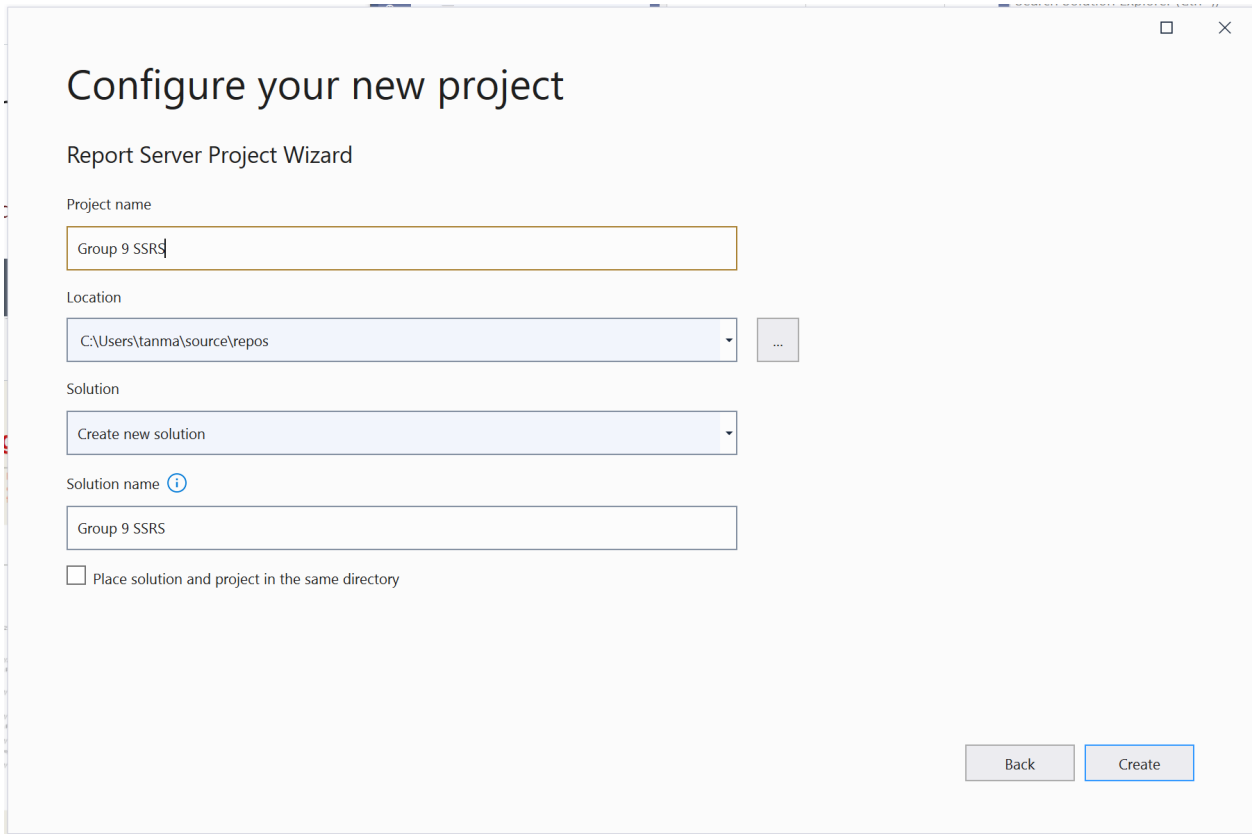

Connection Properties
Data source:
Microsoft SQL Server Analysis Services (Ac
Change...
Server name:
infodata16.mbs.tamu.edu
Log on to the server
User name:
Password:
Save my password
Connect to a database
Select or enter a database name:
SSASCube-BQ2
Advanced...
Test Connection
OK
Cancel
Report Wi
Select th
Select a c
Shared da
New data
Name:
SSASCube
Type:
Microsoft
Connectic
Data Sou
Make t
Help
Edit...
Credentials...
Cancel

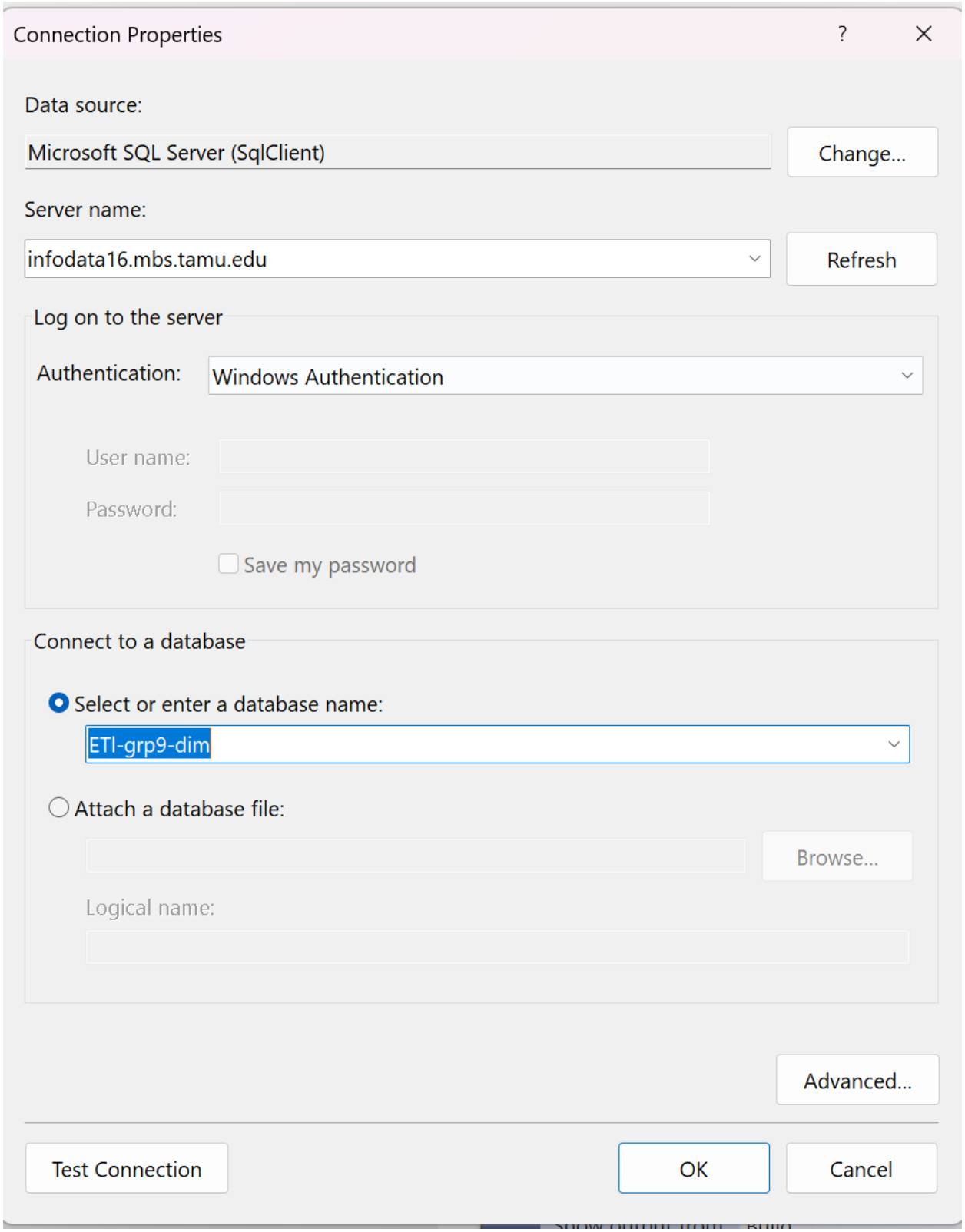

Query Designer
Edit as Text
Import...
MDX
E TI-grp9-dim
Metadata
Search Model
Measure Group:
<All>
E TI-grp9-dim
Measures
Sales Fact
Demo Key
Product Key
Sales Fact Count
Temp-table
Temp-table Count
Year
KPIs
Dim Store
Store Key
Temp-table
Time Dim
Time Key
Calculated Members
Dimension
Hierarchy
Operator
Filter Expression
Parame...
<Select dimension>
Drag levels or measures here to add to the query.
Help
OK
Cancel

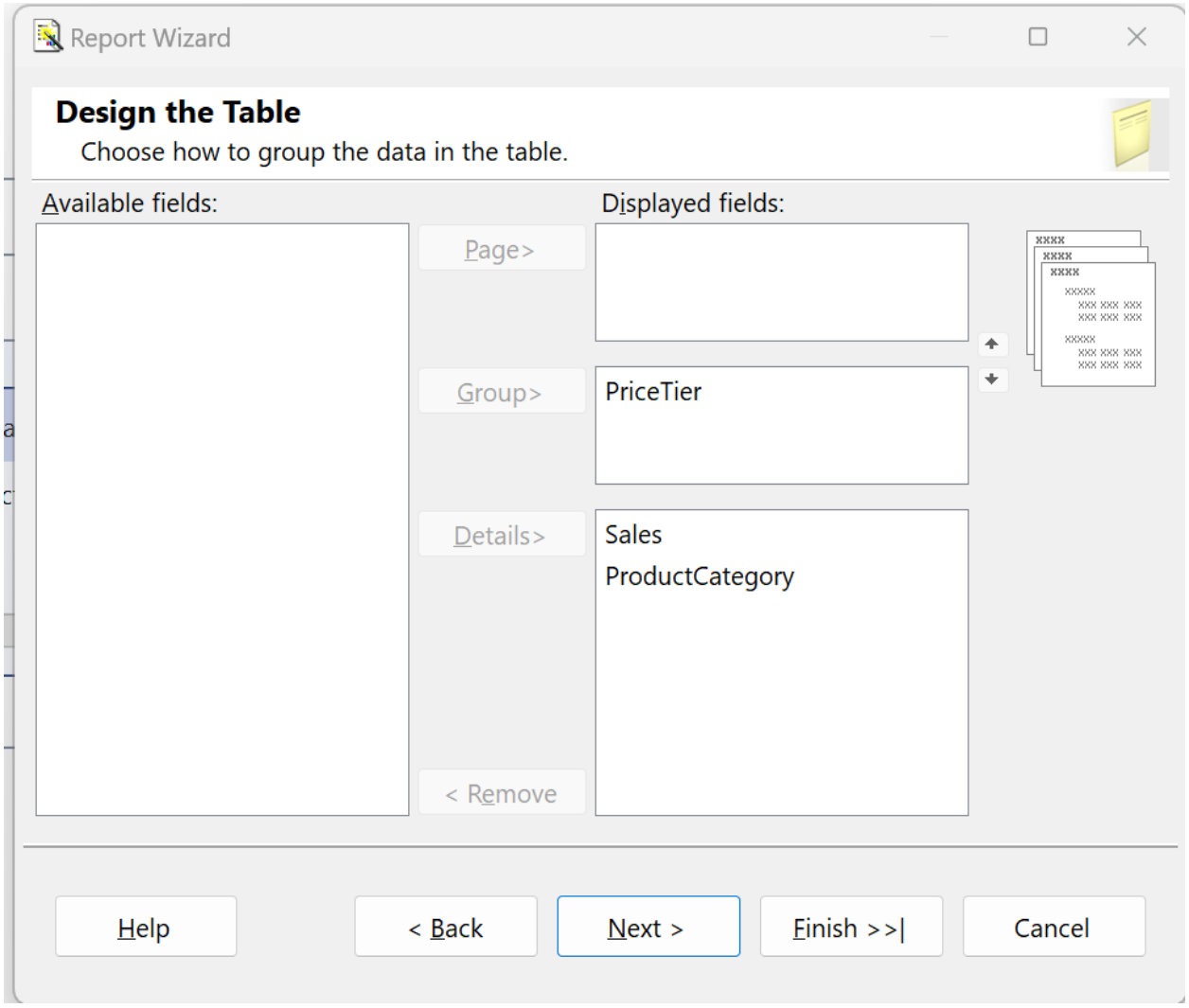
	Store	Year	TotalSales
1	76	1993	16742451.25
2	109	1992	17943893.37
3	92	1994	10167031.49
4	9	1995	18803898.5
5	115	1997	5857348.82
6	115	1991	13879826.87
7	86	1995	12322270.53
8	76	1996	18621555.42
9	134	1991	14195543.12
10	137	1996	13179069.02
11	92	1997	3819716.57
12	109	1989	5274638.03
13	115	1994	18025426
14	129	1992	11607480.33
15	64	1993	11176537.68
16	137	1993	14764643.76
17	119	1989	5712519.78
18	62	1990	20015420.13

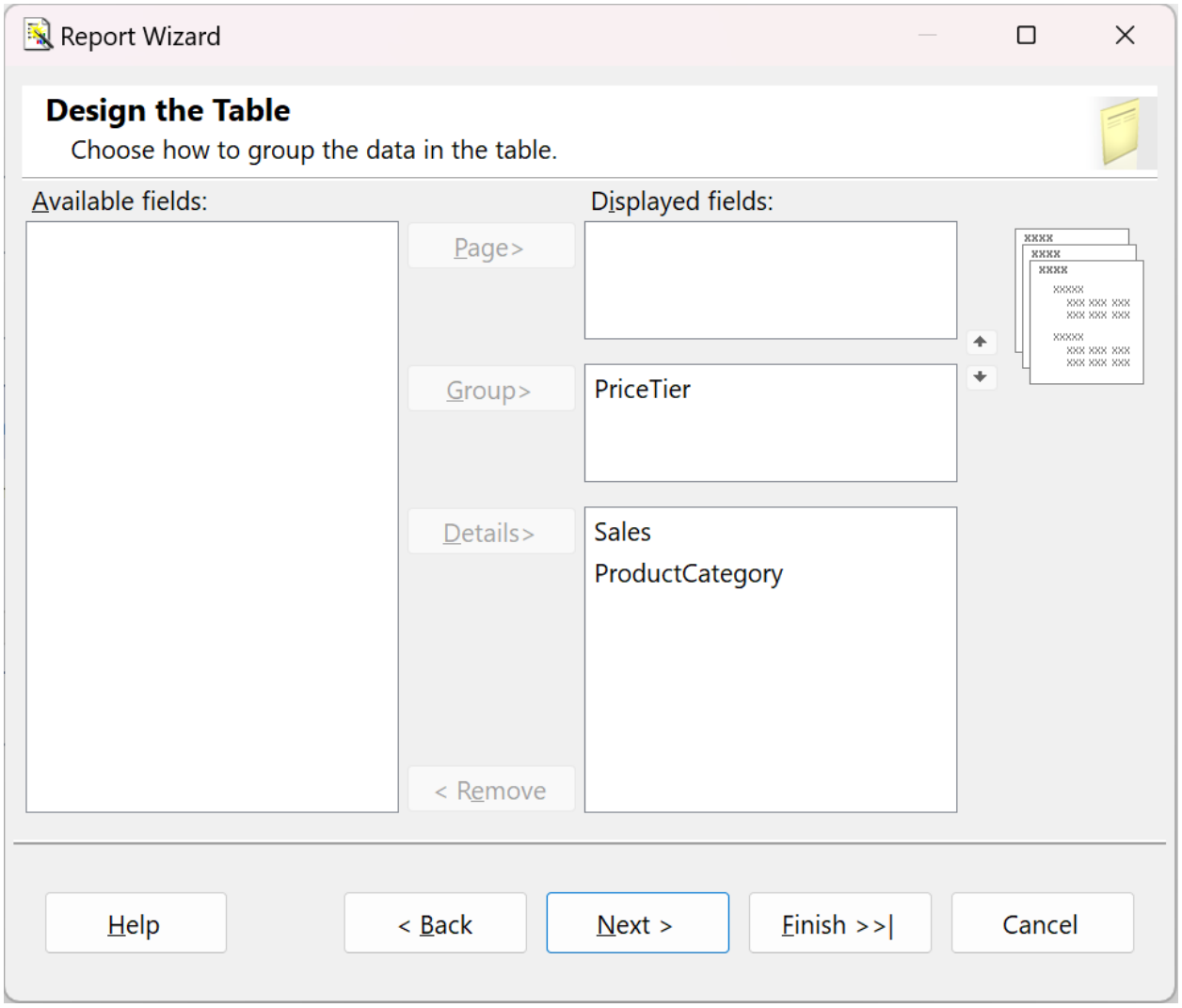
SSRS Q3

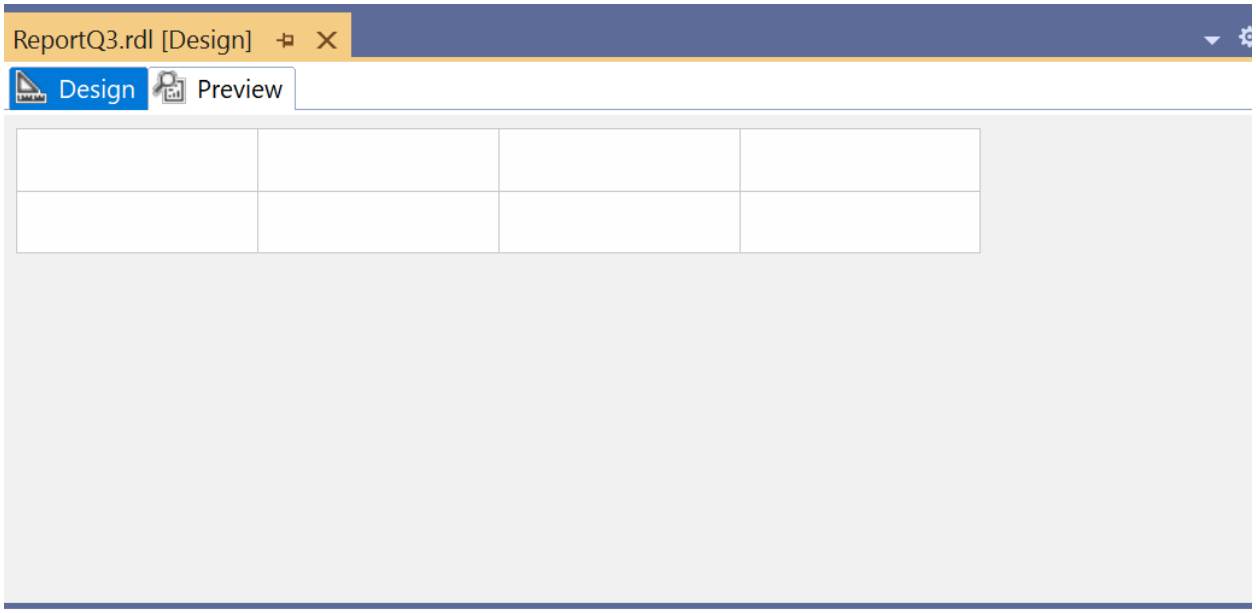
What do the sales patterns for bakery products look like over a span of three years, categorized into low, medium, and high-price tier stores across multiple cities?


Configure your new project
Report Server Project Wizard
Project name
Group 9 SSRS
Location
C:\Users\tanma\source\repos
...
Solution
Create new solution
Solution name
Group 9 SSRS
Place solution and project in the same directory
Back
Create


Connection Properties
?
Data source:
Microsoft SQL Server (SqlClient)
Change...
Server name:
infodata16.mbs.tamu.edu
Refresh
Log on to the server
Authentication:
Windows Authentication
User name:
Password:
Save my password
Connect to a database
Select or enter a database name:
ETI-grp9-dim
Attach a database file:
Browse...
Logical name:
Advanced...
Test Connection
OK
Cancel


Report Wizard
Design the Table
Choose how to group the data in the table.
Available fields:
Displayed fields:
Page>
Group>
PriceTier
Details>
Sales
ProductCategory
< Remove
Help
< Back
Next >
Finish >>|
Cancel


Report Wizard
Design the Table
Choose how to group the data in the table.
Available fields:
Displayed fields:
Page>
Group>
Details>
< Remove
PriceTier
Sales
ProductCategory
Help
< Back
Next >
Finish >>|
Cancel

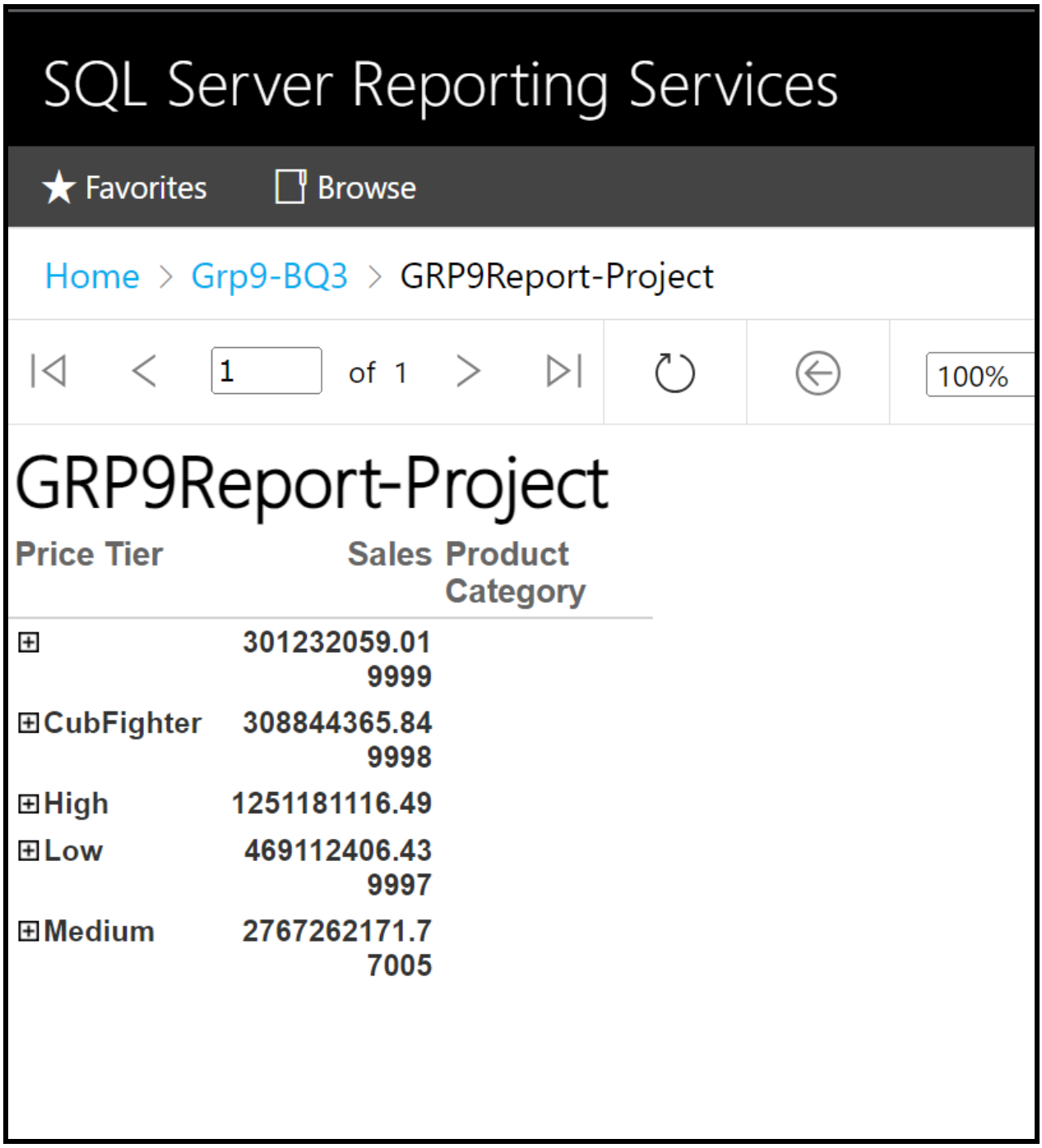


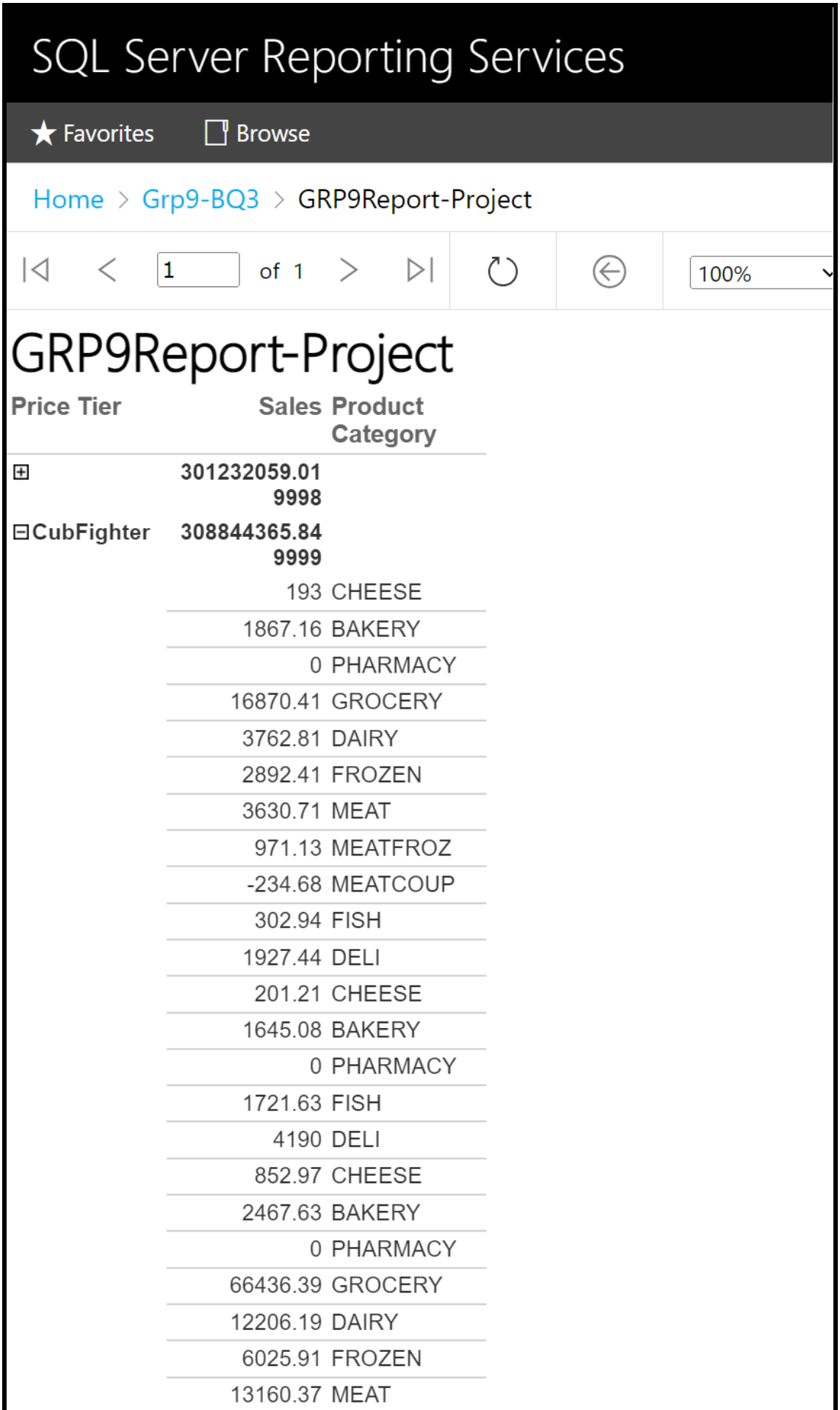
ReportQ3

Price Tier	Sales	Product Cat
[PriceTier]		
	[Sales]	[ProductCatego

infodata16.mbs.tamu.edu/ReportServer - /

Tuesday, November 28, 2023 11:59 PM	<dir>	601_Group8_BQ1
Monday, November 27, 2023 12:08 AM	<dir>	601_Group8_BQ4_Bundle_Sales
Tuesday, November 28, 2023 6:09 PM	<dir>	601_group8_Report1
Thursday, November 30, 2023 1:18 AM	<dir>	601-Group1ReportProjectAnalysisBQ3
Thursday, November 30, 2023 12:40 AM	<dir>	602_GR5_BQ_Meat
Tuesday, November 28, 2023 11:10 AM	<dir>	602_Group4_Project_Report-Q1
Thursday, November 30, 2023 2:26 AM	<dir>	602_Group4_Report-Q1-SSRS
Wednesday, November 29, 2023 11:46 PM	<dir>	602_Group4_Report-Q3
Thursday, November 30, 2023 1:03 AM	<dir>	602_Group4_Report-Q4
Monday, November 27, 2023 4:47 PM	<dir>	602_Group8_Reports
Tuesday, November 28, 2023 2:55 PM	<dir>	602-GR5-report
Saturday, November 25, 2023 2:14 PM	<dir>	603_Group4-DW_Project_Report
Tuesday, November 28, 2023 8:06 PM	<dir>	603-03-Report-1
Tuesday, November 28, 2023 12:15 PM	<dir>	603-group5
Thursday, November 30, 2023 12:16 AM	<dir>	603-GRP-3-BQ1
Thursday, November 30, 2023 4:04 AM	<dir>	603-GRP-3-BQ2
Thursday, November 30, 2023 1:41 PM	<dir>	603-GRP3-BQ3
Thursday, November 30, 2023 4:50 AM	<dir>	603-GRP-3-BQ4
Thursday, November 30, 2023 5:49 PM	<dir>	ARUSHIINCOMEREPORT-PROJECT
Tuesday, November 28, 2023 11:55 PM	<dir>	BQ_Visuazilation_SSRS
Tuesday, November 21, 2023 8:05 PM	<dir>	BQ2_603_grp1_SSRS
Friday, November 24, 2023 12:12 AM	<dir>	BQ3_603_grp1_SSAS_SSRS
Thursday, November 30, 2023 1:27 AM	<dir>	BQ3-Viz
Monday, November 27, 2023 7:02 PM	<dir>	BQ5_SSRS
Thursday, November 30, 2023 2:39 AM	<dir>	BQ5-SSRS
Wednesday, November 29, 2023 12:46 AM	<dir>	BQ5-Visualization
Monday, November 27, 2023 9:24 PM	<dir>	Bus_Q1_Minato
Thursday, November 30, 2023 1:04 PM	<dir>	BusinessQuestion4
Wednesday, November 29, 2023 11:10 PM	<dir>	BusQ5_JaKaGop
Sunday, November 26, 2023 9:07 PM	<dir>	bw3
Sunday, November 26, 2023 12:55 AM	<dir>	C1G2_SSRS
Sunday, November 26, 2023 10:24 PM	<dir>	C1G2_SSRS_SSAS
Tuesday, November 21, 2023 8:11 PM	<dir>	Data_Sources
Thursday, November 30, 2023 1:10 AM	<dir>	Datasets
Wednesday, November 29, 2023 12:26 AM	<dir>	GR5_v2
Thursday, November 23, 2023 3:12 PM	<dir>	Group11_report_4
Wednesday, November 29, 2023 11:26 PM	<dir>	Group12_BQ10_CouponRedemption
Wednesday, November 29, 2023 10:40 PM	<dir>	Group12_BusinessQuestion1
Thursday, November 30, 2023 2:30 AM	<dir>	Group12_SSAS_SSRS_Report
Wednesday, November 29, 2023 11:21 PM	<dir>	Group1-601Report_ProjectBQ3
Monday, November 27, 2023 11:22 PM	<dir>	Group1-601ReportProject
Saturday, November 25, 2023 2:26 PM	<dir>	Group6_BQ1_SSRS
Saturday, November 25, 2023 6:41 PM	<dir>	Grp8_603_BQ1_SSRS
Sunday, November 26, 2023 4:43 PM	<dir>	Grp8_603_BQ2_SSRS
Saturday, November 25, 2023 6:09 PM	<dir>	Grp8_603_BQ3_SSAS_SSRS_v1
Sunday, November 26, 2023 9:11 PM	<dir>	Grp8_603_BQ3_SSRS
Sunday, November 26, 2023 3:25 PM	<dir>	Grp8_603_Bq3_Vizualization_SSRS
Thursday, November 30, 2023 9:47 PM	<dir>	Grp9-BQ3
Wednesday, November 29, 2023 12:17 AM	<dir>	HarshReport_-_Project
Wednesday, November 29, 2023 6:40 PM	<dir>	HarshReport-Project
Thursday, November 16, 2023 1:16 PM	<dir>	ISTM_637_601_Group_3
Thursday, November 30, 2023 9:23 AM	58341	ISTM_637_601_Group_3_BQ-5_Power_BI.pbix
Tuesday, November 28, 2023 4:13 AM	<dir>	ISTM_637_601_GROUP10_BQ3
Thursday, November 30, 2023 2:10 PM	<dir>	ISTM637_602_Grp3_Question5
Thursday, November 30, 2023 1:17 PM	<dir>	ISTM637_Group3_Question1
Friday, November 24, 2023 5:42 PM	<dir>	ISTM-637-602-Group10-Question6


SQL Server Reporting Services
Favorites
Browse
Home
Grp9-BQ3
GRP9Report-Project
1
of 1
100%
GRP9Report-Project
Price Tier
Sales
Product Category
301232059.019999
CubFighter
308844365.849998
High
1251181116.49
Low
469112406.439997
Medium
2767262171.77005

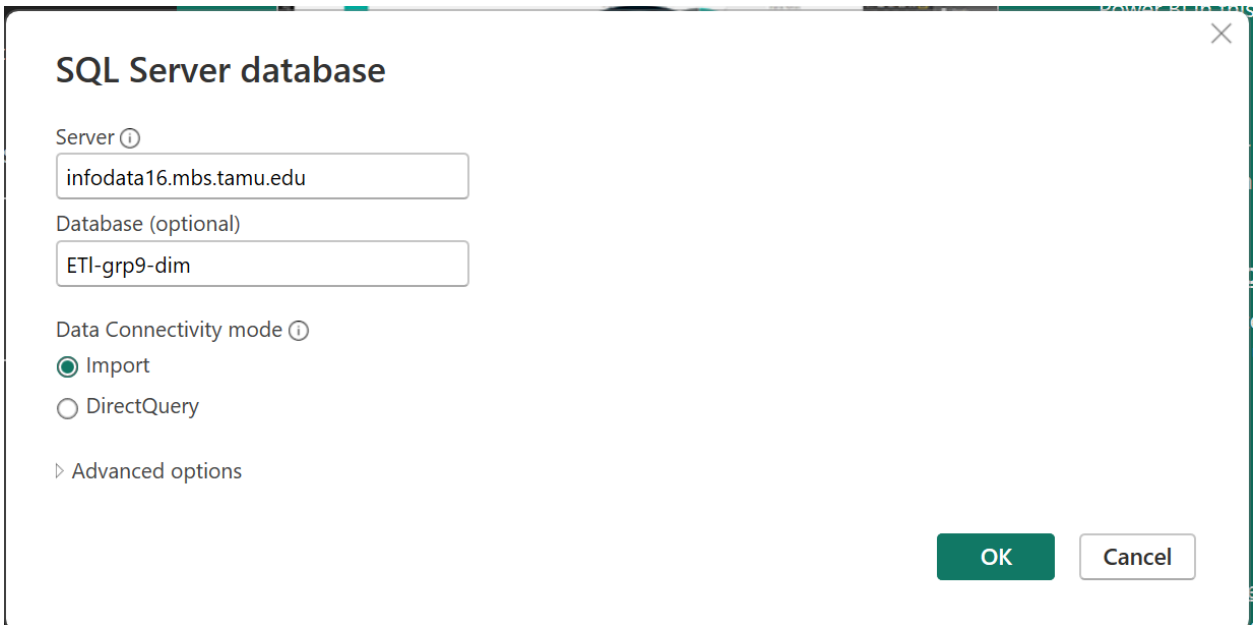


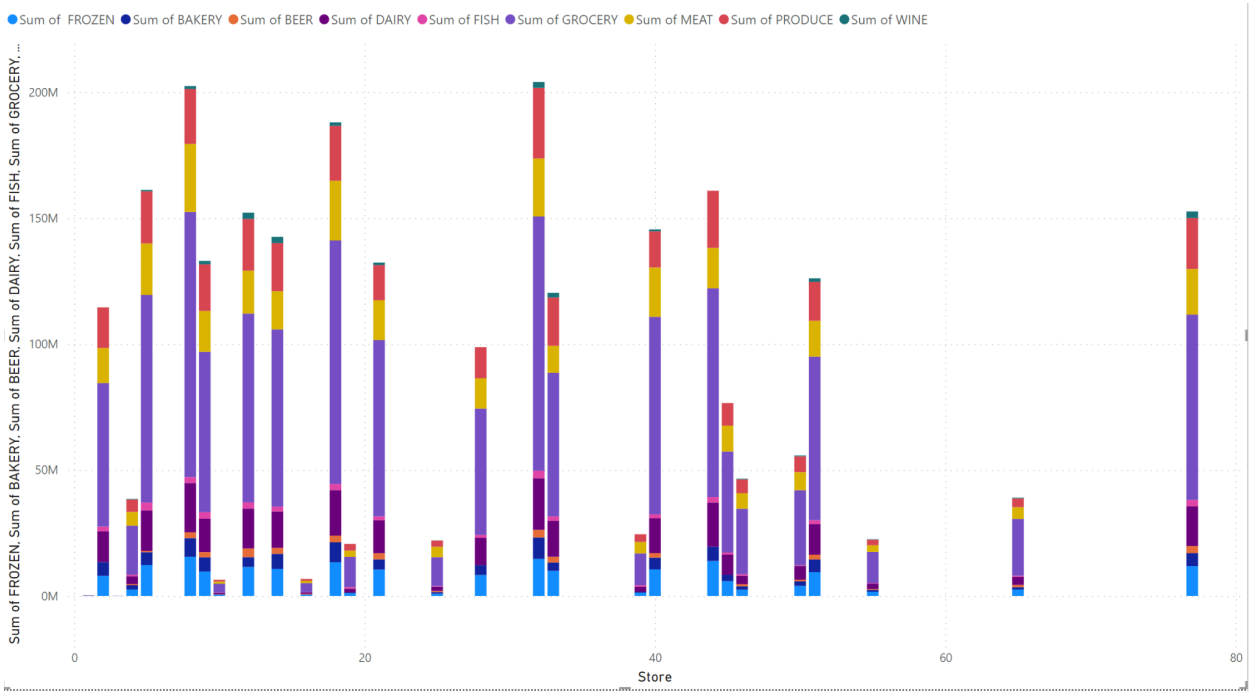
GRP9Report-Project

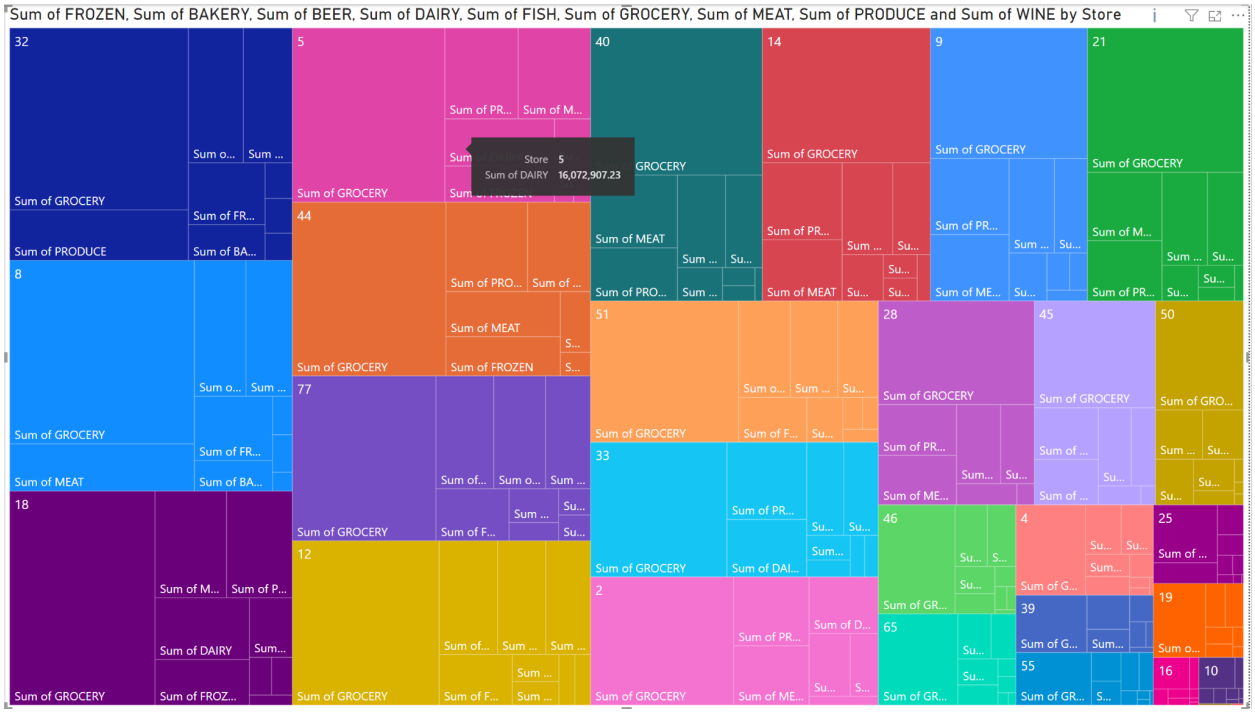
Price Tier	Sales	Product Category
⊞	**301232059.019998**	
⊟**CubFighter**	**308844365.849999**	
	193	CHEESE
	1867.16	BAKERY
	0	PHARMACY
	16870.41	GROCERY
	3762.81	DAIRY
	2892.41	FROZEN
	3630.71	MEAT
	971.13	MEATFROZ
	-234.68	MEATCOUP
	302.94	FISH
	1927.44	DELI
	201.21	CHEESE
	1645.08	BAKERY
	0	PHARMACY
	1721.63	FISH
	4190	DELI
	852.97	CHEESE
	2467.63	BAKERY
	0	PHARMACY
	66436.39	GROCERY
	12206.19	DAIRY
	6025.91	FROZEN
	13160.37	MEAT

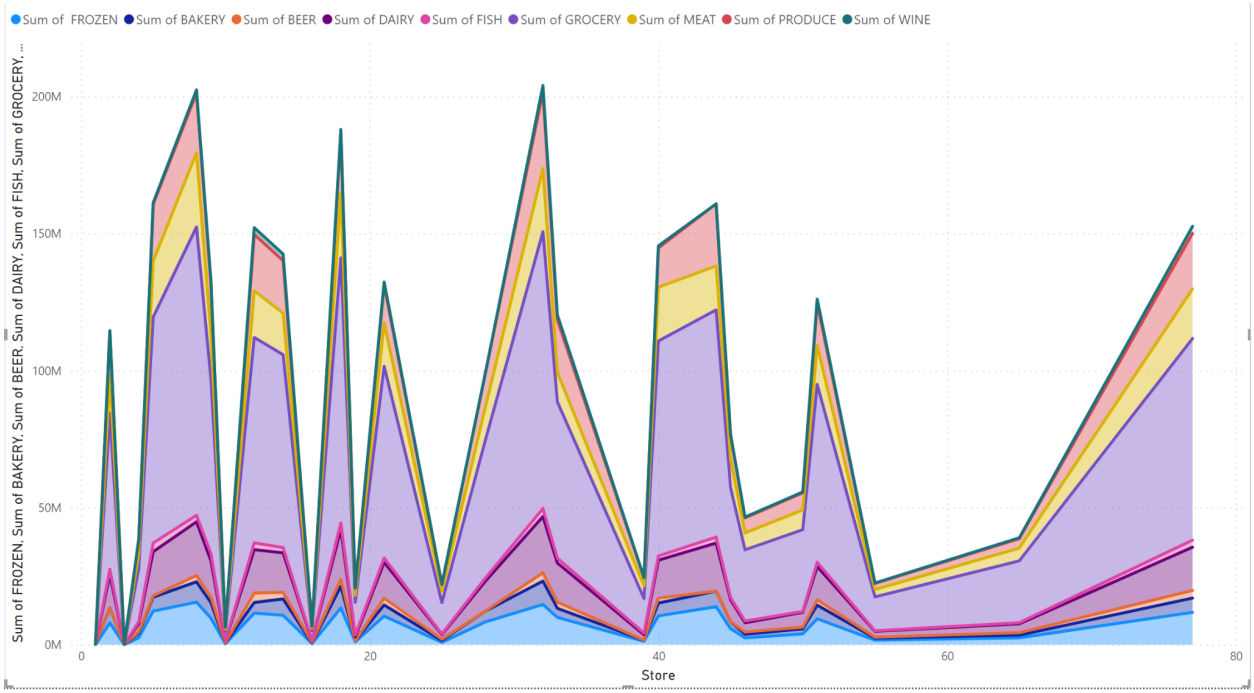
PowerBI

Q4 What are the top-performing product categories in terms of sales revenue for each DFF branch over the years?



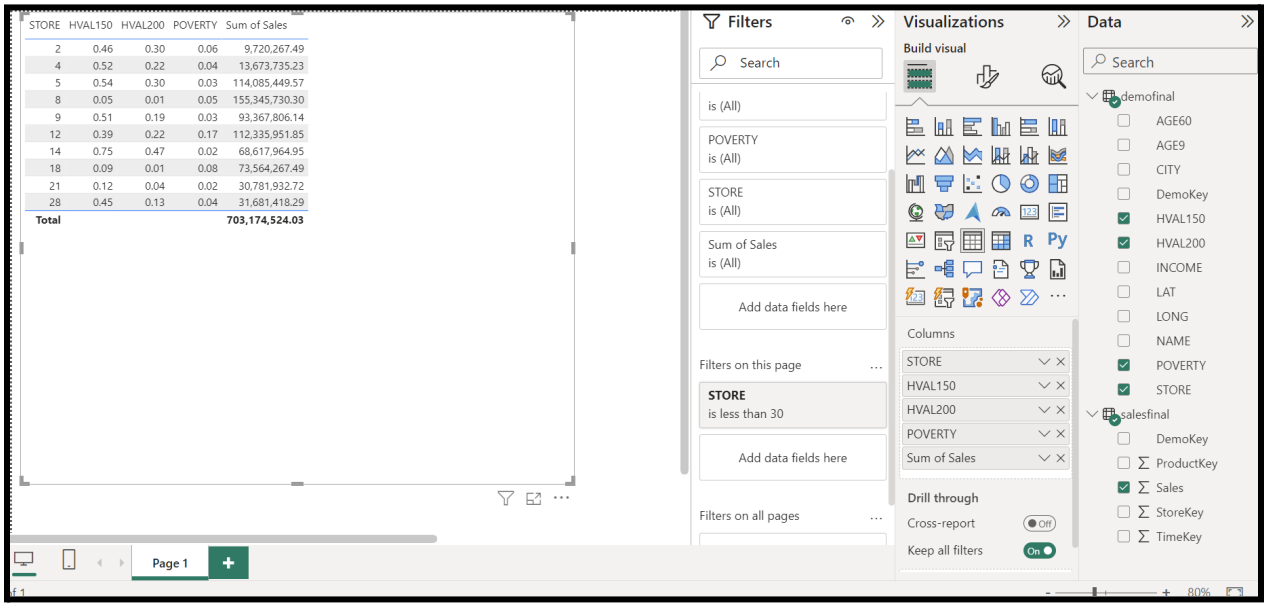


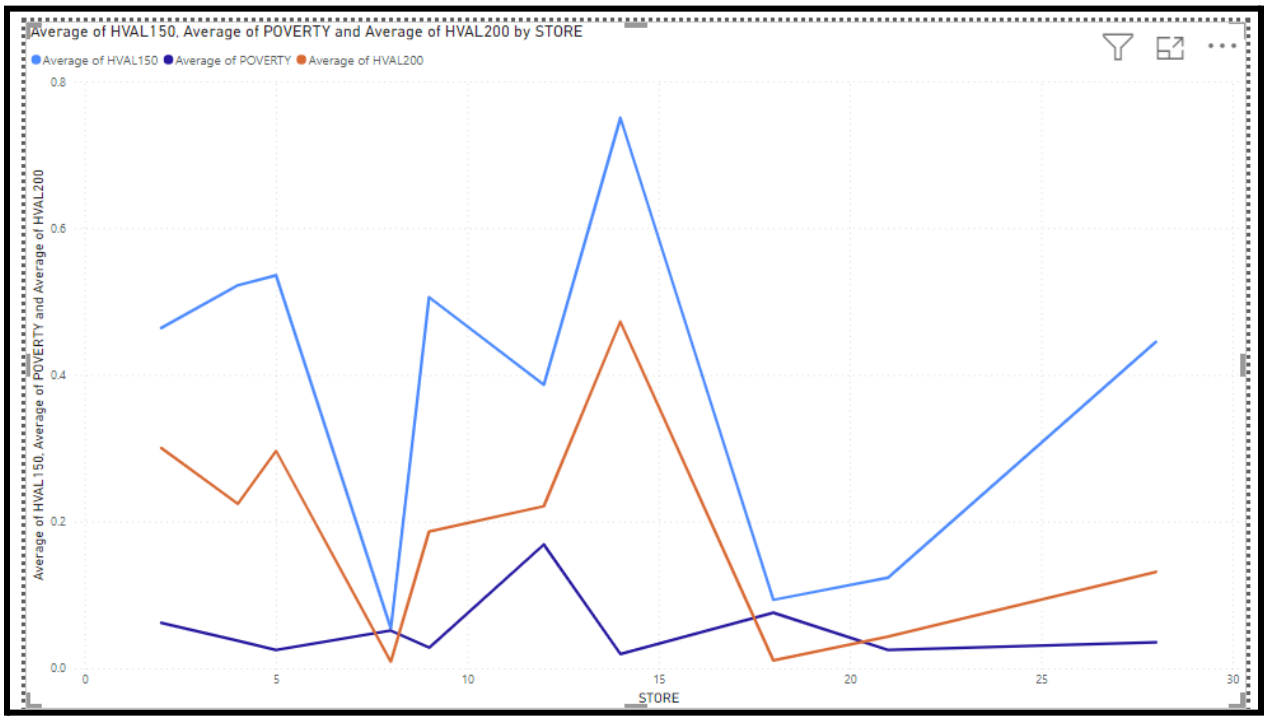

Sum of FROZEN, Sum of BAKERY, Sum of BEER, Sum of DAIRY, Sum of FISH, Sum of GROCERY, Sum of MEAT, Sum of PRODUCE and Sum of WINE by Store
Store 5
Sum of DAIRY 16,072,907.23

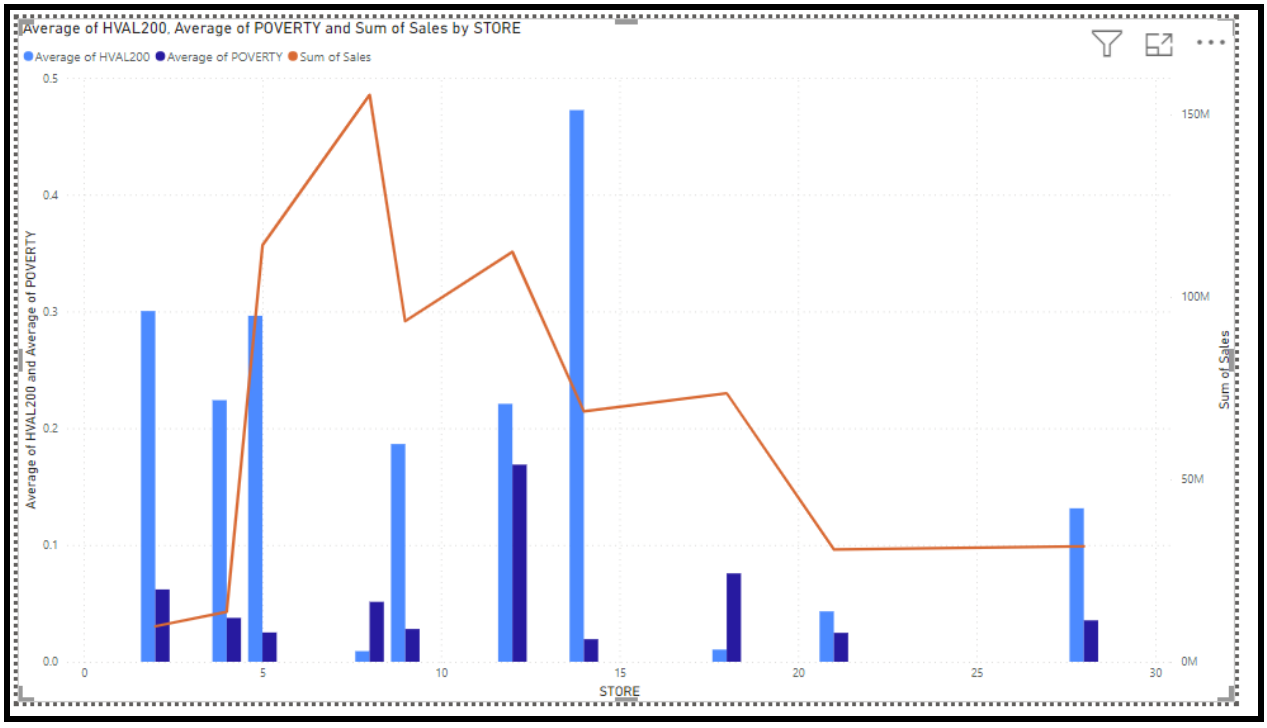

Sum of FROZEN
Sum of BAKERY
Sum of BEER
Sum of DAIRY
Sum of FISH
Sum of GROCERY
Sum of MEAT
Sum of PRODUCE
Sum of WINE
Sum of FROZEN, Sum of BAKERY, Sum of BEER, Sum of DAIRY, Sum of FISH, Sum of GROCERY, ...
200M
150M
100M
50M
0M
0
20
40
60
80
Store

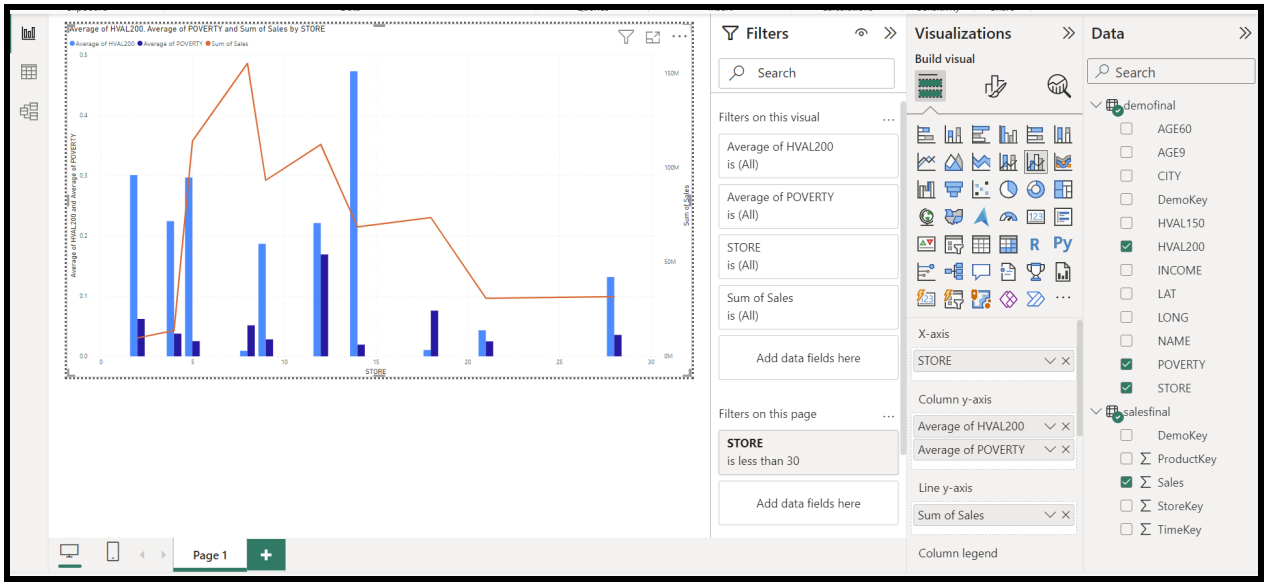
5. PowerBI

Which stores attract people who earn below the poverty line and have high value income thresholds?






Average of HVAL200, Average of POVERTY and Sum of Sales by STORE
Average of HVAL200
Average of POVERTY
Sum of Sales
Average of HVAL200 and Average of POVERTY
Sum of Sales
STORE
0.5
0.4
0.3
0.2
0.1
0.0
150M
100M
50M
0M
0
5
10
15
20
25
30


Filters
Search
Filters on this visual
Average of HVAL200 is (All)
Average of POVERTY is (All)
STORE is (All)
Sum of Sales is (All)
Add data fields here
Filters on this page
STORE is less than 30
Add data fields here
Visualizations
Build visual
X-axis
STORE
Column y-axis
Average of HVAL200
Average of POVERTY
Line y-axis
Sum of Sales
Column legend
Data
Search
demofinal
AGE60
AGE9
CITY
DemoKey
HVAL150
HVAL200
INCOME
LAT
LONG
NAME
POVERTY
STORE
salesfinal
DemoKey
ProductKey
Sales
StoreKey
TimeKey
Page 1

Section 6: References

Preuss, Björn & Argiolas, Matteo. (2016)

(Luther, 1993)
(Breivik, 2019)

Girsang, Ganda & Arisandi, Geri & Elysisa, Calista & Michelle, & Saragih, Melva. (2019)

(Fairlie, 2023)

https://www.guru99.com/star-schema-in-data-warehouse-modeling.html

https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/