ISTM 637 - PROJECT REPORT 4

GROUP 5 - SECTION 601
HARSH PARIKH
MOHIT BHAT
RUSHI SHAH

TABLE OF CONTENTS

SECTION 1: INTRODUCTION **4**
DETAILS ABOUT THE DATA 5
ER DIAGRAM 6
SECTION 2: BUSINESS QUESTIONS AND THEIR EXPLANATIONS **7**
BUSINESS QUESTIONS 7
SELECTED BUSINESS QUESTIONS 24
SECTION 3: INDEPENDENT DATA MART DESIGN **24**
STAR SCHEMA REPRESENTATION 25
REASONING AND STEPS FOR DATA MART DESIGN: 29
PHYSICAL DESIGN 32
SECTION 4: DATA CLEANING AND INTEGRATION **34**
DATA QUALITY ISSUES 34
ETL PLAN 34
TARGET DATA 35
DATA SOURCES 35
DATA MAPPING 37
DATA EXTRACTION RULES 39
DATA TRANSFORMATION AND CLEANSING RULES 40
PLAN FOR AGGREGATE TABLES 41
ORGANIZATION OF DATA STAGING AREA 41
PROCEDURE FOR DATA EXTRACTION AND LOADING 42
DATA EXTRACTION 42
DATA LOADING PROCEDURE 44
ETL FOR DIMENSION TABLES 46
ETL FOR FACT TABLES 48
ETL IMPLEMENTATION 49
DATA MART 1: STORE DATA MART (STORE LEVEL SALES) 49
DimTime 49
DimStore 54
FactStore 60
DATA MART 2: SALES DATA MART (SALE PER STORE, PER CATEGORY) 65
FINAL_MOVEMENT_STAGING 65
dimProduct 71
factSales 76
DATA GRANULARITY 81
LIST OF TEMPORARY TABLES IN STAGING AREA 81

SECTION 5: BUSINESS INTELLIGENCE REPORTING **83**
REPORTING PLAN 83
MAPPING ATTRIBUTES FROM INDEPENDENT DATA MARTS 84
REPORT IMPLEMENTATION 87
Business Question 1 87
Business Question 2 96
Business Question 3 103
Business Question 4 120
Business Question 5 124
REFERENCES 128

SECTION 1: INTRODUCTION

Dominick's Finer Foods is a chain of grocery stores based in Chicago. With focus on customer satisfaction and premium shopping experience, this 100-store chain has outlets which combine both food stores and drugstores and deal with 3,500 products across 25 different categories.

Data warehousing is the area of technology that deals with collecting, storing and managing large volumes of data from various sources within the organization.
The data warehouse acts as a central repository and the stored data can be efficiently analyzed and used for business intelligence and decision-making purposes.

Our project aims to help DFF make strategic decisions and grow their business by implementing a Data Warehouse. The dataset that we have contains 9 years of store-level data, and we will leverage this data to gain insights into customer preferences, purchasing behaviors, and popular products. In addition, by analyzing historical data, DFF can optimize its inventory levels by reducing overstocking and stockouts.

The report explains the data files available to us, sheds light on the problems faced in the retail industry and derives 10 business questions that would help DFF optimize key business metrics such as sales, inventory and profit.

Some of the issues we faced on the initial stage of the project were handling messy data with erroneous rows, missing values, incomplete files and processing exceptionally large amounts of data using the Excel software.

However, we were able to overcome these challenges through various processes and techniques in data understanding, data migration, data modeling and data transformations. We were able to investigate our business questions and conclude various insights which we hope to delve further into in the next part of the project.

ISTM 637 - Group 5 Report

DETAILS ABOUT THE DATA

The Data Source:

The data used for the study is provided by James M. Kilts Center, University of Chicago Booth School of Business.The Dominick's database covers store-level scanner data collected at Dominick's Finer Foods over a period of more than seven years. The data is made available in SAS format and is converted into CSV files by Jens Mehrhoff. Variables have been modified and truncated in the process but yield the same end results.

Data Description:

The data is vast and unique in its breadth of coverage. The database covers approximately 9 years of data spread across all the stores. It has details on more than 3500 universal product codes (UPCs) spanning across around 25 categories. In addition, there are several supporting files that provide seasonal, demographic and store location data to help support the different metrics with more attributes.

1) General Files:

The general files contain information across all the categories and about the overall performance of products and holistic attributes across the categories. There are two files in this category: The customer count file and store-specific demographics

a. Customer Count File: The file contains daily information of store level information like number of customers visited, purchased, total dollar sales, coupons redeemed etc. across more than 25 categories of products.

b. Store Specific Demographics: The file contains store specific demographic data. The data has been originally provided by the U.S. government census and helps create demographic profile for each DFF store

2) Category Specific Files:

a. UPC File: There are different files for each 29 category containing one record for each UPC. These files are named as UPCXXX where XXX stands for the specific category. The file provides us with information regarding identifier codes, product name, size etc. regarding a particular UPC

b. Movement Files: The movement files are divided at category level like the UPC files and include weekly sales data for each UPC in each store for over 5 6 years. The attributes included in this file are price, unit sold, profit margin, deal code, etc. The files are sorted by UPC, store, week.

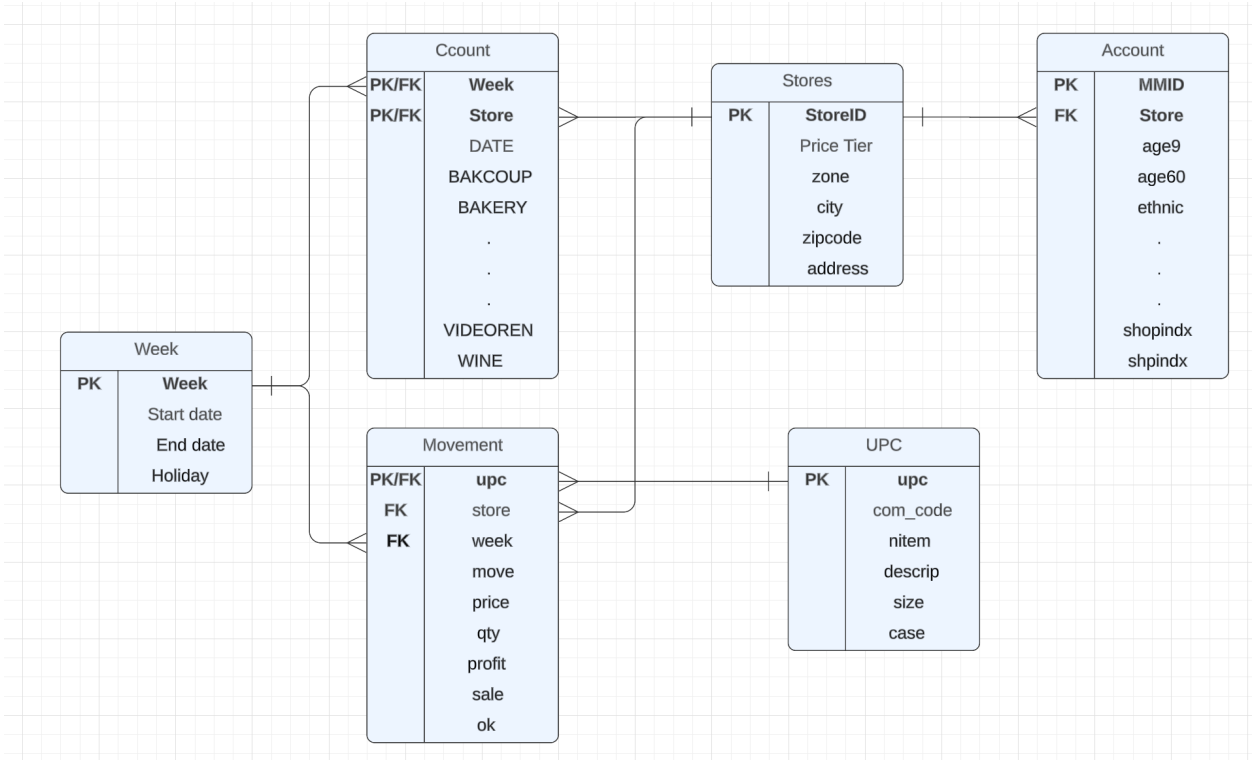
Other important data:

1) Dominick Store locations: This file has all store data like the city, zonal information, the price tier and the address information
2) Map of Dominick stores: Above information is supplemented by map of Dominick Stores

3) Week Decode table: This table includes information of the week like start date, end date, special events marked on those days , mapped to a week number used in the other files mentioned above.

Below is an ER Diagram to help us understand the fields of the data files and how they are related to each other.

ER DIAGRAM



SECTION 2: BUSINESS QUESTIONS AND THEIR EXPLANATIONS

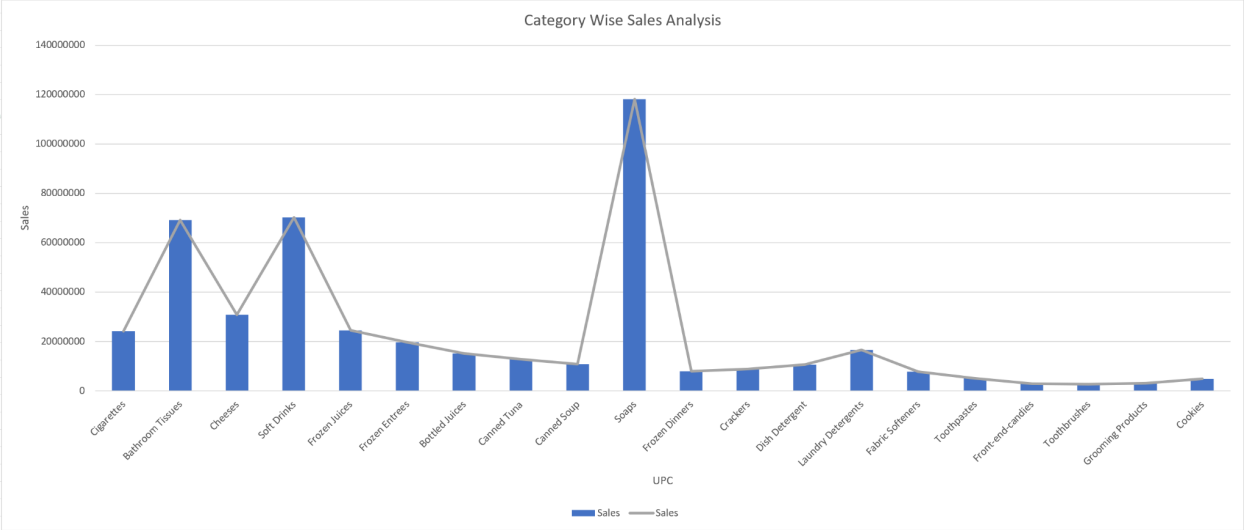
BUSINESS QUESTIONS

1. **What are the various possibilities to maximize earnings based on high sales categories?**

   Rationale for choosing this question: We wanted to find out categorize various product categories into high volume high sale, low volume high sale and so on categories. This would give us a great understanding into which categories to analyze further helping out DFF to strategize on what to work on.

UPC	Sales
Cigarettes	24065839.93
Bathroom Tissues	69108558.22
Cheeses	30870767.66
Soft Drinks	70174689.99
Frozen Juices	24443094.67
Frozen Entrees	19581687.17
Bottled Juices	15206398.37
Canned Tuna	12793199.94
Canned Soup	10815938.42
Soaps	118119905.9
Frozen Dinners	7936312.293
Crackers	8814714.233
Dish Detergent	10596756.53
Laundry Detergents	16581462.35
Fabric Softeners	7732656.27
Toothpastes	5013996.16
Front-end-candies	2881547.483
Toothbrushes	2751590.497
Grooming Products	2972138.453
Cookies	4820511.798

Data Snapshot



Sales by Category Chart

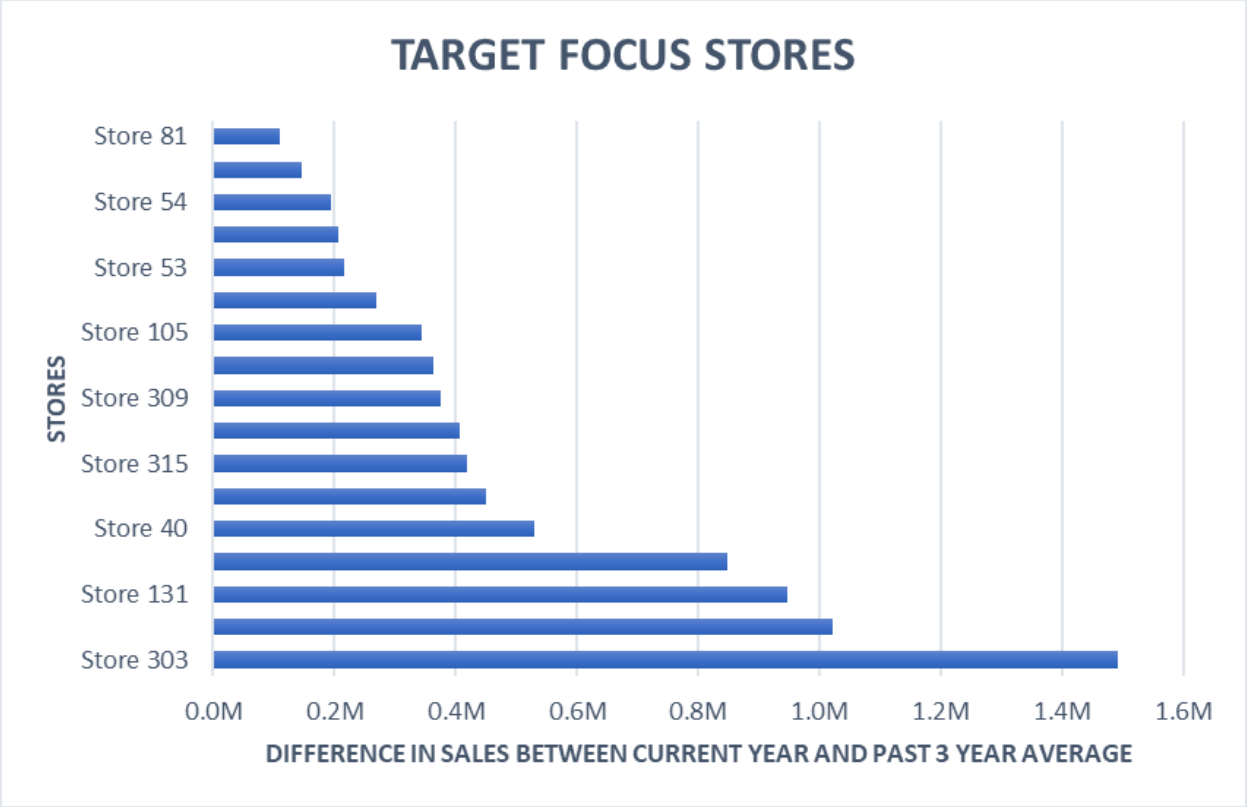
Findings: We found various product categories belonging to different baskets. For example, cigarettes are a low volume high profit category. Soaps are a high volume, low profit category. Grooming products are a low volume low profit category. Moreover, sales for soaps, soft drinks, bathroom tissues, and cigarettes was exceptionally high.
Justification for prioritization: This question will provide a high level overview to the business stakeholders on the most important KPI - sales. Accordingly they can allocate resources on those categories and maximize earnings.

2. **What are the Target Focus Stores (Stores performing badly as compared to the past 3 years)?**

   Rationale for choosing this question: Identifying stores that have performed well historically but are not performing well presently can help DFF dive deeper into understanding what's causing this and fix it which will lead to increasing profits.

STORE	PRESENT YEAR SALES	P3Y AVERAGE SALES	DIFFERENCE
Store 303	12810493.95	14300899.73	1490405.78
Store 301	14899633.83	15921390.01	1021756.177
Store 131	7260842.04	8206758.36	945916.32
Store 304	18901712.64	19749327.66	847615.0233
Store 40	4844505.74	5372806.34	528300.6
Store 73	7096598.45	7546351.777	449753.3267
Store 315	17510110.06	17928673.99	418563.9333
Store 312	14783207.91	15189169.63	405961.72
Store 309	14513576.22	14888770.35	375194.13
Store 77	8368554.6	8730875.387	362320.7867
Store 105	4471101.95	4815523.357	344421.4067
Store 103	5550354.53	5819145.98	268791.45
Store 53	4066495.25	4282633.97	216138.72
Store 314	15860358.13	16067457.67	207099.54
Store 54	4414009.04	4607208.367	193199.3267
Store 89	4339490.89	4486519.41	147028.52
Store 81	6934915.96	7044953.933	110037.9733

Data Snapshot



Stores where Current Year Sales are lesser than Past 3 year Average

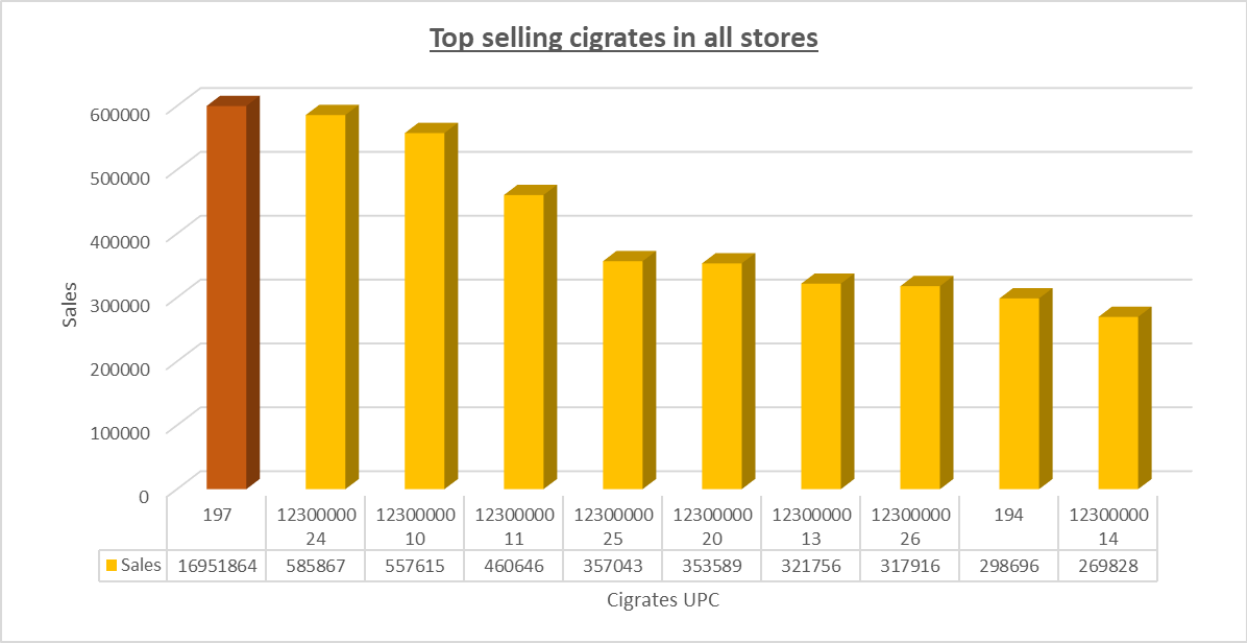
Findings: We found 8 stores with their 4 month sales in the year 1997 to be $200k or more less than their previous 3 year's first 4 month average. These form our target focus stores that DFF can work on diving deeper into.
Justification for prioritization: This question not only helps realize a dip in performance by certain stores but will also help DFF focus more on understanding the gaps between customer requirements and their offerings. They can also try to figure out the reasons for downfall and work on it which will help them achieve high performance in all of their stores.

3. **What are the top selling Cigarette SKUs in the Cigarette Product Category for DFF?**

Rationale for choosing this question: Cigarettes are a highly profitable product category for DFF, hence, finding the top performers in this category would help DFF focus on strategizing its product placement and inventory management to meet the demand for it.

Stores	Sales
197	16951864
1230000024	585867
1230000010	557615
1230000011	460646
1230000025	357043
1230000020	353589
1230000013	321756
1230000026	317916
194	298696
1230000014	269828

Data Snapshot



Top Selling Cigarette SKUs across all stores

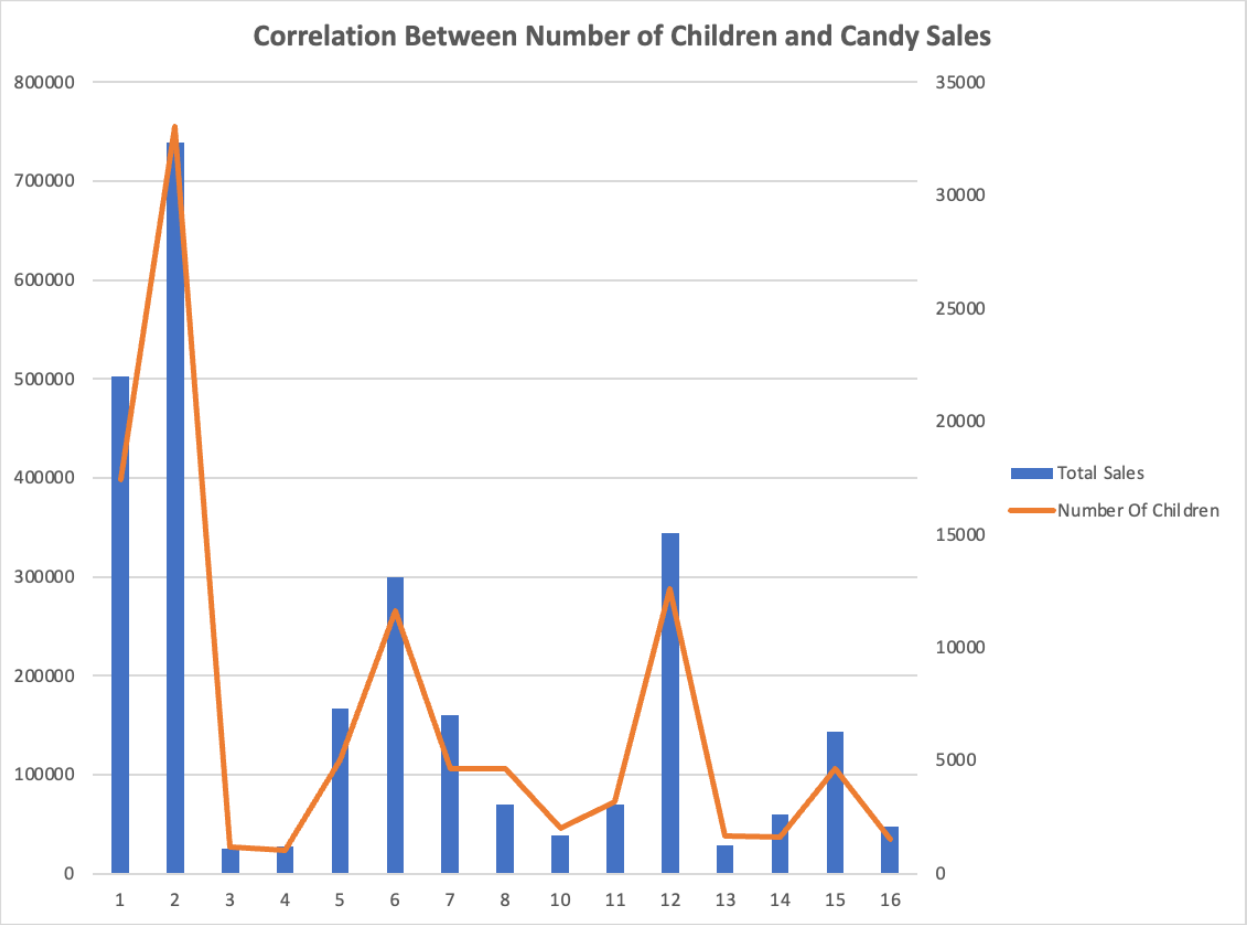
Findings: We found that the '197' SKU cigarettes is a standout amongst the cigarette products sold. Its sales are orders above the sales of other cigarette products, hence DFF should focus on selling this product the most in this category.
Justification for prioritization: Since we found out that Cigarettes have a really high profit margin through the profit vs sales analysis, we believe leveraging cigarette sales the right way can help the business gain more profits.Hence it was highly essential to understand top selling SKUs within the cigarette category.

4. **What is the correlation between candy sales and the number of children in a given demographic area?**

Rationale for choosing this question: Analyzing the correlation between candy sales and number of children in the zones would help DFF customize their product offerings. By aligning inventory with demand patterns, DFF can optimize stock level, ensuring the availability of candies.

Row Labels	Total Sales	Number Of Children
1	502746.1683	17422.18416
2	738401.315	33017.00873
3	25074.12167	1163.767469
4	27027.06667	1041.220106
5	167218.7367	5033.813191
6	299620.5083	11635.89581
7	160532.5317	4616.595956
8	69487.305	4660.33953
10	39062.23833	1991.549186
11	69421.17333	3162.57098
12	344571.6183	12614.75028
13	28675.115	1674.42993
14	59642.605	1613.540161
15	143657.4717	4621.837807
16	47752.625	1504.20822
Grand Total	**2722890.6**	**105773.7115**

Data Snapshot



Correlation between footfall of children in stores vs Candy Sales

Finding: As we can see above, the number of children and total sales of candies are highly correlated. DFF can manage their inventory based on these metrics and enhance customer satisfaction, boost candy sales, and position themselves strategically in the competitive grocery market.

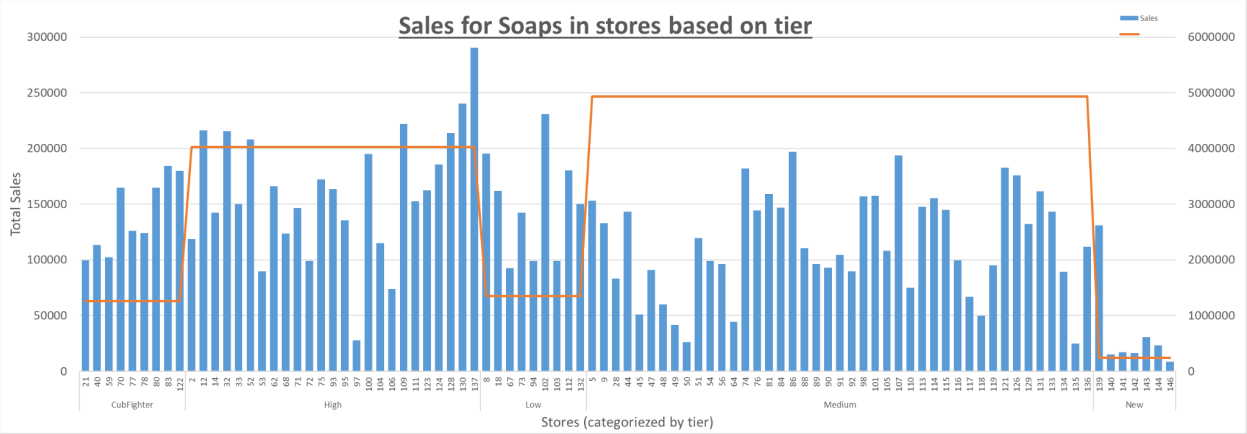
Justification for prioritization: Children being really addicted to candies can be a great sector to earn more out of. Understanding correlation between the sales and children can not only help forecast candy sales and thereby revenue but also help in a more optimized inventory management and marketing strategy.

5. **How does the sales of soaps compare across various price tiers?**

   Rationale for choosing this question: Soaps having the highest sales made it one of the product categories to focus on. Understanding the sales based on the sale tier will help the business to leverage those sale tiers to augment profit percent.

Tier	Store	Sales	
	21	99388	1259315
	40	113410	1259315
	59	102307	1259315
	70	164886	1259315
CubFighte	77	126002	1259315
	78	123996	1259315
	80	164862	1259315
	83	184462	1259315
	122	180003	1259315
	2	118675	4024697
	12	216082	4024697
	14	142429	4024697
	32	215450	4024697
	33	149936	4024697
	52	207862	4024697
	53	89609	4024697
	62	165936	4024697
	68	123767	4024697
	71	146221	4024697
	72	99268	4024697
	75	172307	4024697
High	93	163593	4024697
	95	135385	4024697
	97	27737	4024697
	100	195104	4024697
	104	114850	4024697
	106	73731	4024697
	109	221906	4024697
	111	152340	4024697
	123	162518	4024697
	124	185761	4024697
	128	213722	4024697
	130	240134	4024697
	137	290375	4024697
	8	195335	1351495
	18	162071	1351495
	67	92647	1351495
	73	142427	1351495
Low	94	99074	1351495
	102	230661	1351495
	103	98982	1351495
	112	180314	1351495
	132	149983	1351495
	5	152887	4935102
	9	132927	4935102
	28	83159	4935102
	44	143181	4935102
	45	50917	4935102
	47	90963	4935102
	48	59942	4935102
	49	41494	4935102
	50	26086	4935102
	51	119679	4935102
	54	98992	4935102
	56	96276	4935102
	64	44575	4935102
	74	182063	4935102
	76	144392	4935102
	81	158958	4935102
	84	146907	4935102
	86	196829	4935102
	88	110496	4935102
	89	96256	4935102
	90	92922	4935102
Medium	91	104366	4935102
	92	89526	4935102
	98	157088	4935102
	101	157280	4935102
	105	108075	4935102
	107	193741	4935102
	110	74969	4935102
	113	147699	4935102
	114	155398	4935102
	115	144789	4935102
	116	99510	4935102
	117	66978	4935102
	118	49566	4935102
	119	95107	4935102
	121	182784	4935102
	126	175855	4935102
	129	132074	4935102
	131	161574	4935102
	133	143019	4935102
	134	89379	4935102
	135	24834	4935102
	136	111591	4935102
	139	130871	241297
	140	14969	241297
	141	16991	241297
New	142	16338	241297
	143	30455	241297
	144	23125	241297
	146	8547	241297

Data Snapshot



Soap Sales based on Price Tier

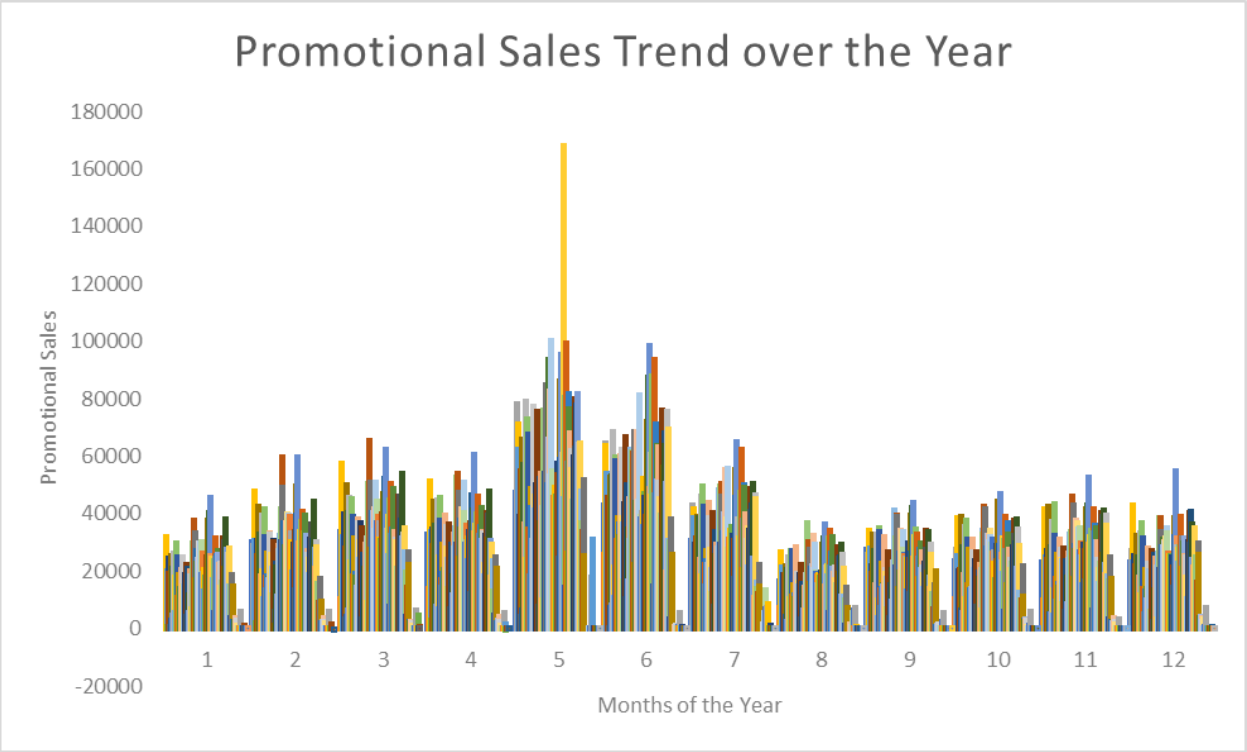
Findings: There is an inverse correlation between price tier and bath soap sales.
Justification for prioritization: The sales vs profit analysis clearly portrayed a huge sale of soap but a bleak profit margin comparably. This required additional exploration of which tiers should be uplifted to increase profits and hence greater revenue.

6. **In which months of the year do maximum promotional sales take place?**

Rationale for choosing this question: Promotions are drivers of sales for retail stores. Hence we wanted to find out what was the trend of promotional sales over the different months of the year for the entire span of the data.

Sum of PROMO	Column Labels					
Row Labels	2	4	5	8	9	10
1	21675.32	6417.69	19949.28	32190.13	20212.36	956
2	30063.46	13612.31	28877.56	47916.56	30812.36	1898
3	33868.21	10261	32706.65	57640.34	32205.14	2215
4	32889.46	13829.83	36091.11	51586.75	33816.46	2995
5	47601.95	20574.66	78185.99	71333.54	62214.02	1441
6	43196.14	25777.1	64637.05	63620.49	53990.69	
7	30774.07	17927.01	43141.68	41507.75	38601.4	
8	16414.57	7512.33	23074.15	26780.49	21457.25	
9	27548.87	9324.09	25938.13	34302.33	27949.84	
10	23820.05	10333.36	27606.89	38689.69	25645	0
11	23262.34	10194.51	30081.82	41645.61	24836.48	0
12	23218.32	5258.77	27220.56	42837.56	23037.14	0
Grand Total	**354332.76**	**151022.66**	**437510.87**	**550051.24**	**394778.14**	**9505**

Data Snapshot



Promotional sales Trend

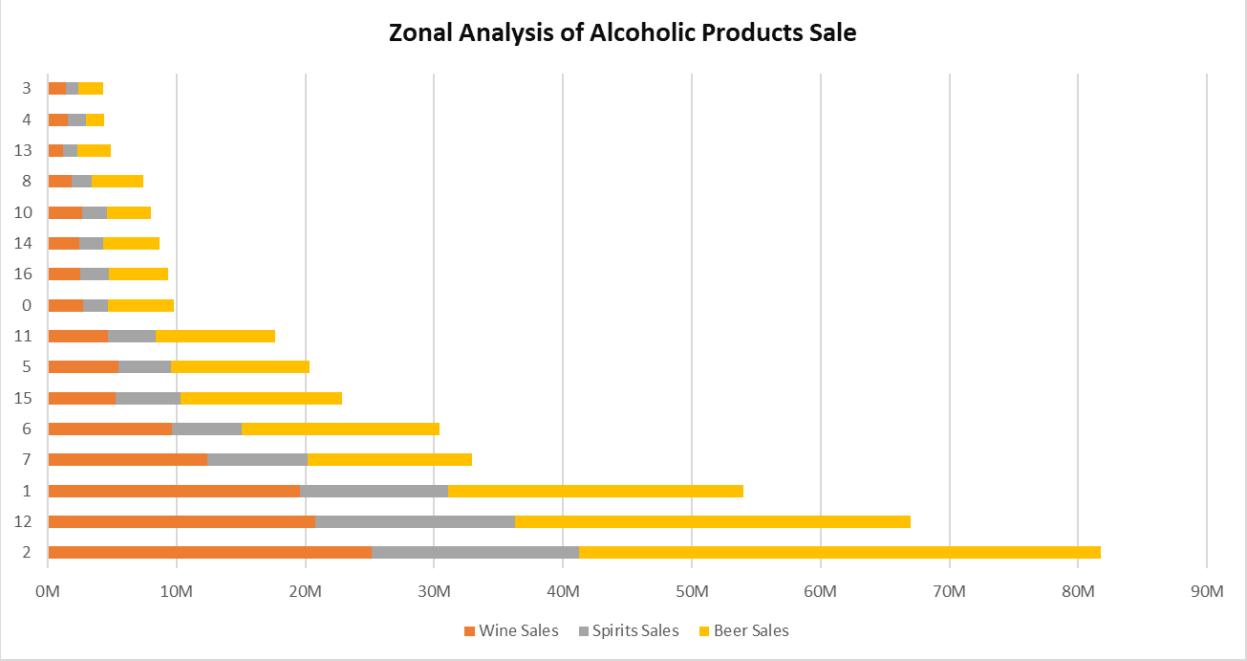
Findings: We found that May and June have high promotional sales for the DFF hence it can strategically make use of these months to test out various promotions. We can also deep dive into customer footfall during these times of the year.
Justification for prioritization: Majorly promotions drive sales over time. Hence it is essential for DFF to gain insights on when is the right time to release promotional offers so they can gain most out of it. This can clearly help them schedule campaigns and tactics.

7. **What is the sales of Alcoholic beverages (Wine, Beer, and Spirits) across different zones?**

Rationale for choosing this question: Alcoholic beverages are a highly profitable retail category in which consumers do not pay much heed to the price point while spending money. Hence, strategically pricing these products based on the sales observed in various locations would help DFF maximize its profitability from the category.

ZONE	Wine Sales	Spirits Sales	Beer Sales	Total Sales
2	25144593.69	16082163.18	40509899.38	81736656.25
12	20781068.34	15512968.62	30651957.87	66945994.83
1	19603347.61	11466059.47	22946919.47	54016326.55
7	12411146.46	7767241.13	12762363.52	32940751.11
6	9674940.17	5375045.97	15377750.61	30427736.75
15	5300938.01	5025993.36	12527375.15	22854306.52
5	5482526.12	4071210.76	10740130.6	20293867.48
11	4665312.95	3708842.2	9271388.78	17645543.93
0	2719750.34	1994584.78	5089896.71	9804231.83
16	2498813.73	2247388.64	4626374.06	9372576.43
14	2456584.03	1851242.06	4356177.84	8664003.93
10	2681330.25	1951822.49	3373034.1	8006186.84
8	1871414.56	1533940.53	4034015.64	7439370.73
13	1204162.89	1102005.81	2579822.57	4885991.27
4	1541021.96	1439679.01	1430033.33	4410734.3
3	1419959.41	934555.32	1931983.09	4286497.82

Data Snapshot



Zonal Analysis of Alcohol sales

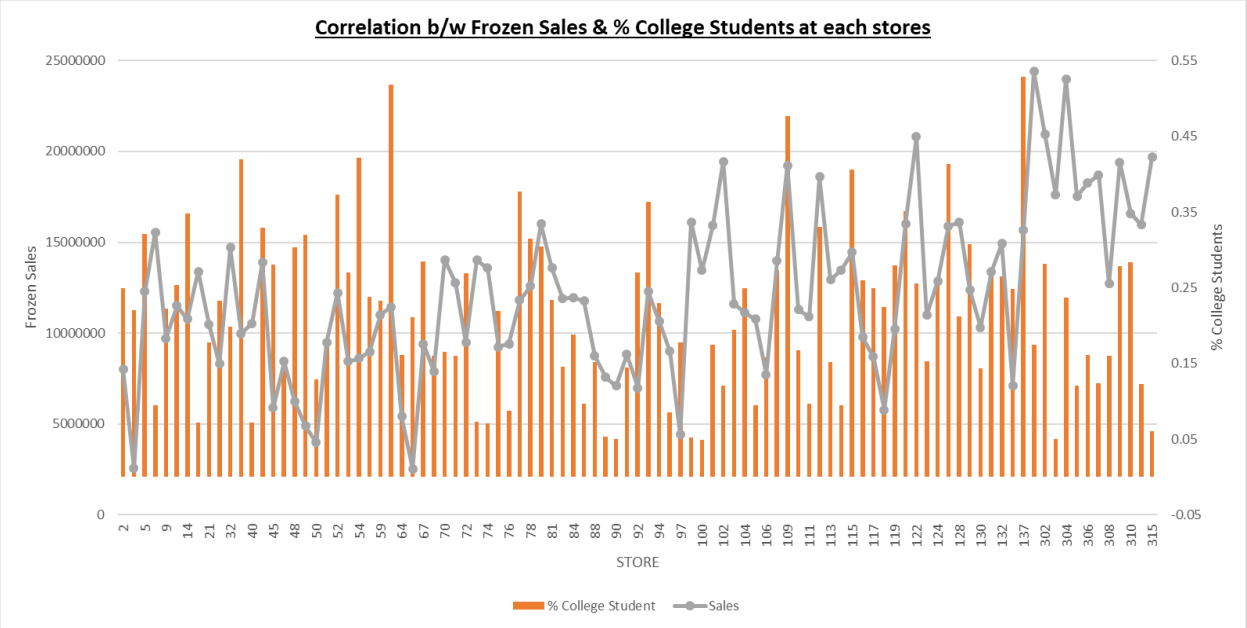
Findings: We were able to articulate the zones where alcoholic beverages were sold the most which will help DFF strategize the pricing strategy and inventory management for the stores in these zones.
Justification for prioritization: Alcohol being a prime product category generates greater revenue and thereby profit. Gaining zone wise and product wise clarity on performance can assist in better allocation of resources and inventory while also optimizing pricing strategies.

8. **Is there a correlation between Frozen Foods Sales and Stores where college students have a high footfall?**

   Rationale for choosing this question: It is common knowledge that college students have a high tendency to buy frozen foods over any other consumer category. Hence, defining the correlation between their footfall in a store and the sale of frozen foods can help DFF in determining product placement for this product category across different stores in the region.

STORE	% Colle	Sales
2	25%	7999097.99
4	22%	2563399.64
5	32%	12286261.56
8	10%	15558883.26
9	22%	9706555.45
12	25%	11514083.1
14	35%	10768335.22
18	7%	13390134.97
21	18%	10481303.89
28	23%	8301303.41
32	20%	14719502.32
33	42%	9976360.42
40	7%	10546652.53
44	33%	13890861.58
45	28%	5879364.35
47	14%	8433552.45
48	30%	6261364.68
49	32%	4899158.01
50	13%	3991966.89
51	17%	9471654.51
52	37%	12227982
53	27%	8436034.23
54	42%	8612661.97
56	24%	8973656.32
59	23%	11009567.18
62	52%	11455825.67
64	16%	5414491.56
65	21%	2536048.53
67	28%	9414022.79
68	16%	7894328.97
70	17%	14037190.57
71	16%	12774058.35
72	27%	9492133.63
73	7%	14026951.45
74	7%	13581163.41
75	22%	9237411.85
76	9%	9423444.81
77	38%	11824220.89
78	31%	12592416.11
80	30%	16015095.11
81	23%	13588978.23
83	15%	11894674.15
84	19%	11958386.91
86	10%	11772263.29
88	15%	8753410.27
89	5%	7583287.11
90	5%	7104957.15
91	14%	8829504.05
92	27%	6983085.58
93	36%	12288004.29
94	23%	10664073.78
95	9%	9012556.53
97	18%	4417613.36
98	5%	16110954.59
100	5%	13445782.16
101	17%	15944694.07
102	12%	19453133.02
103	19%	11623766.89
104	25%	11130537.63
105	9%	10781164.23
106	16%	7730975.71
107	27%	13983947.55
109	48%	19228457.34
110	17%	11304920.63
111	10%	10902572.99
112	33%	18593293.38
113	15%	12941742.58
114	9%	13475679.11
115	41%	14456162.5
116	26%	9785759.39
117	25%	8691062.51
118	22%	5777619.76
119	28%	10224934.66
121	35%	16020517.98
122	26%	20814299.9
123	15%	10991034.79
124	26%	12841612.45
126	41%	15898309.4
128	21%	16094915.68
129	31%	12398756.12
130	14%	10329702.62
131	27%	13395320.95
132	26%	14929332.65
134	25%	7130102.82
137	53%	15676418.69
301	17%	24395103.51
302	28%	20947788.01
303	5%	17633041.16
304	24%	23994384.66
305	12%	17551297.45
306	16%	18256801.52
307	12%	18718928.56
308	16%	12724601.96
309	28%	19394636.38
310	28%	16586660.94
312	12%	15970504.99
315	6%	19672394.3

Data Snapshot



Correlation Chart

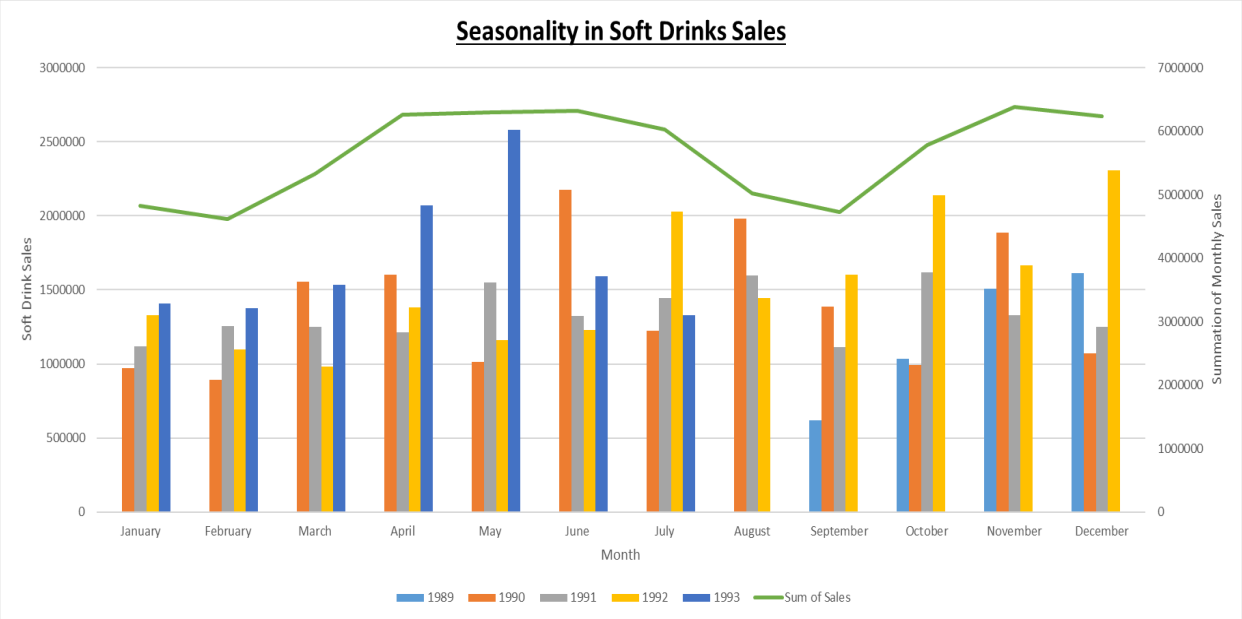
Findings: We found a high correlation between frozen foods sold and the percentage footfall of college students across various stores. Hence, DFF can strategize placing these products in these stores to meet the demand for it.
Justification for prioritization: Frozen food and fast food categories are common among youngsters i.e. college students. Therefore, DFF can optimize their product placement and sales by targeting those areas with frozen food where there is a higher number of college students. College students are a small subset of the entire customer base and hence this is given low priority.

9. What is the seasonality of soft drinks sales?

Rationale for choosing this question: As per the profit vs sales analysis, soft drinks is another product category of interest as it has high sales and low profit.
Moreover, soft drinks are a highly perishable product category hence understanding the months where demand is high will help DFF optimize shelf management of this product category.

Sum of Sales / Row Labels	Column Labels 1989	1990	1991	1992	1993	Grand Total
January		971495.71	1116970.5	1329310.113	1406121.813	4823898.137
February		894024.4825	1255122.765	1095576.147	1375677.587	4620400.981
March		1556592.41	1252190.33	982497.9667	1531800.097	5323080.804
April		1602188.945	1213702.35	1380673.973	2070065.353	6266630.622
May		1013991.727	1548923.57	1161286.647	2580167.43	6304369.375
June		2178199.718	1326022.47	1226577.383	1592200.527	6323000.098
July		1222714.61	1444559.49	2027675.167	1329119.3	6024068.567
August		1981278.19	1597077.043	1442161.28		5020516.513
September	620185.33	1387158.92	1114191.457	1601194.53		4722730.237
October	1035711.95	993422.18	1618291.163	2138318.367		5785743.66
November	1509582.67	1886412.43	1329804.037	1664054.52		6389853.657
December	1611367.62	1069571.095	1251106.103	2305406.68		6237451.498
Grand Total	**4776847.57**	**16757050.42**	**16067961.28**	**18354732.77**	**11885152.11**	**67841744.15**

Data Snapshot



Seasonality of Soft Drinks Sales

Findings: Not only do the summer months of May and June have high sales but also the winter festive months of November and December have high sales of soft drinks. In addition, their purchase percentage is low in the initial months of each year, which could be attributed to the health-focused resolutions taken by the consumers at the start of the year.

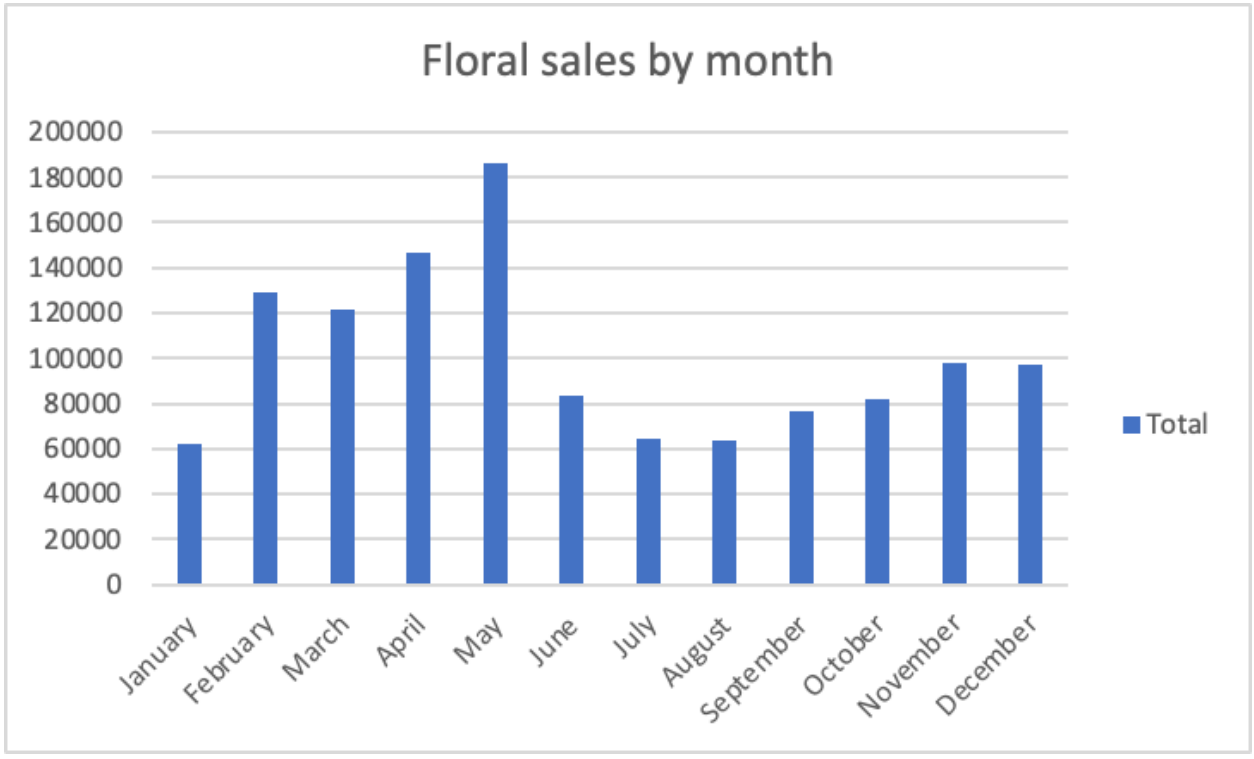
Justification for prioritization: Understanding customers is one of the most important aspects in performing a good business. Focusing on soft drinks which have a good sales and decent profit margin can be beneficial to DFF. Understanding trends and patterns in their usage can help them to manage these products better while also serving customers better. There isn't a lot of variation in the chart and hence this can be given a lesser priority.

10. **What is the impact of festive occasions and other significant events, on monthly floral sales?**

Rationale for choosing this question: The impact of festive occasions on monthly floral sales is justified due to seasonal demand, emotional significance, cultural norms, market trends, and strategic planning.

Row Labels	Sum of Sales
January	4823898.137
February	4620400.981
March	5323080.804
April	6266630.622
May	6304369.375
June	6323000.098
July	6024068.567
August	5020516.513
September	4722730.237
October	5785743.66
November	6389853.657
December	6237451.498
Grand Total	**67841744.15**

Data Snapshot



Monthly trend of Floral sales

Findings: Festive events like Valentine's Day and cultural celebrations prompt increased flower purchases, driven by emotional gestures and societal traditions. Understanding these patterns can help DFF to tailor floral offerings, marketing strategies, and inventory management effectively.

Justification for prioritization: Major events drive the sales of festive products generally. Events like Valentine’s day give the business an opportunity to target more customers with appropriate products and hence increase sales. But this is prioritized low since events like these occur once a year and hence can be given lesser priority.

SELECTED BUSINESS QUESTIONS

Selected Business Questions for further analysis:

1. What are the top selling Cigarette SKUs in the Cigarette Product Category for DFF?
2. In which months of the year do maximum promotional sales take place?
   What are the various possibilities to maximize earnings based on high sales categories?
3. What is the correlation between candy sales and the number(percentage) of children in a given demographic area?
4. What is the impact of festive occasions and other significant events, on monthly floral sales?

SECTION 3: INDEPENDENT DATA MART DESIGN

Using Kimball's methodology we decided to keep two independent data marts for our two fact tables.

	Dimension		
Data Mart	dimProduct	dimTime	dimStore
Sales Data Mart	x	x	x
Store Data Mart		x	x

Kimball's Data Marts Matrix

STAR SCHEMA REPRESENTATION

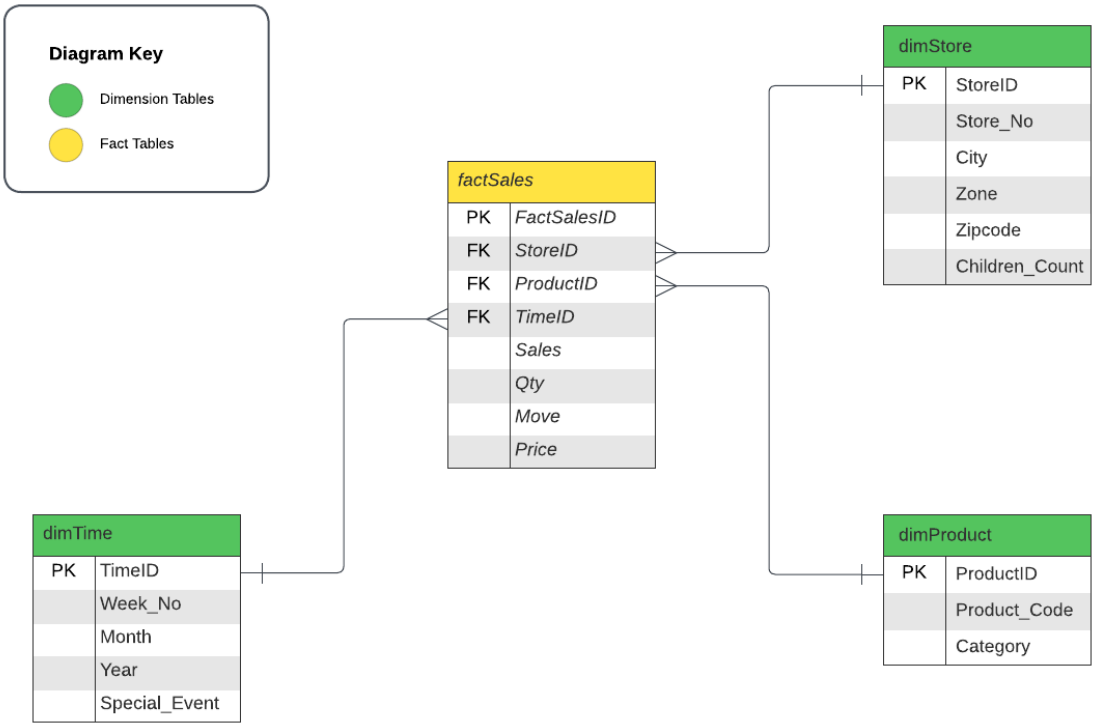


Fig: Star Schema Representation of Sales Data Mart

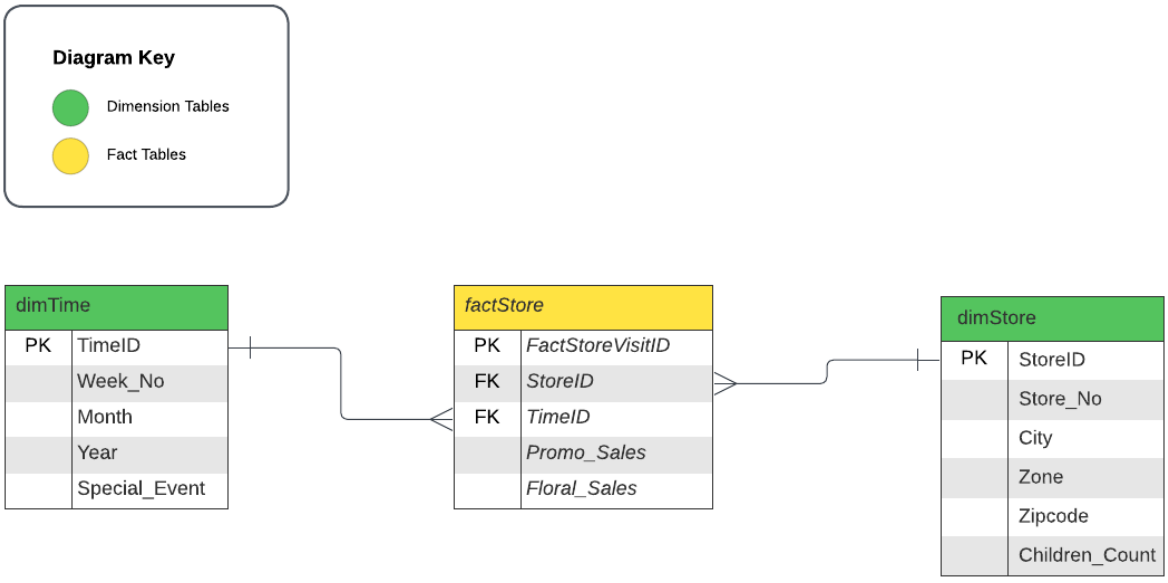


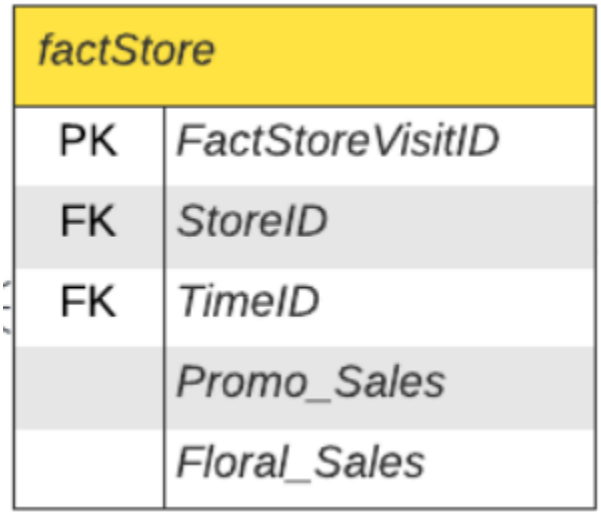
Fig: Star Schema Representation of Store Data Mart

Fact Tables

1. **factSales**: This table contains Sales Data of each product. It is on a weekly granularity.

FactSalesID: Surrogate key generated by the Data Warehouse to ensure uniqueness of each row in the table.
StoreID: Foreign key to connect with the dimStore table.
ProductID: Foreign key to connect with the dimProduct table.
TimeID: Foreign key to connect with the dimTime table.
Sales: Measure of the sales made for each product for each week in each store.
Qty: Measure of the number of items bundled together.
Move: Measure of the number of units sold.
Price: Measure of the retail price.

2. **factStore**: This fact table stores a lot of information on the Store level.



FactStoreID: Surrogate key generated by the Data Warehouse to ensure uniqueness of each row in the table.
StoreID: Foreign key to connect with the dimStore table.
TimeID: Foreign key to connect with the dimTime table.
Promotional_Sales: Dollar value of sales made through promotions by the store.
Floral_Sales: Dollar value of sales made through selling floral items by the store.

Dimension Tables

1. **dimTime**: This table stores the week to time mapping for the data.

dimTime	
PK	TimeID
	Week_No
	Month
	Year
	Special_Event

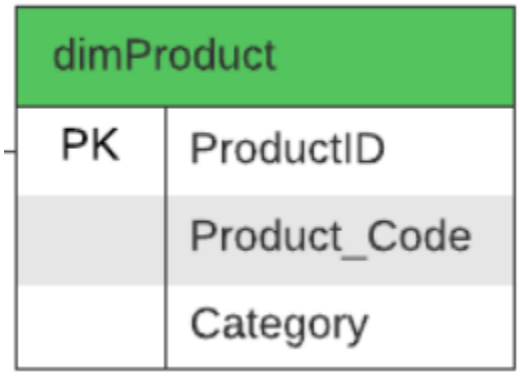
TimeID: Surrogate Key generated by the Data Warehouse to ensure uniqueness of each row in the table.
Week_No: Identifies the number of the week ranging from 9/14/89 to 5/14/97
Month: Denotes the month of the year corresponding to the week.
Year: Denotes the year.
Special_Event: Describes if that week corresponds to a special event.

2. **dimStore**: This table stores the Store level descriptive information.

dimStore	
PK	StoreID
	Store_No
	City
	Zone
	Zipcode
	Children_Count

StoreID: Surrogate Key generated by the Data Warehouse to ensure uniqueness of each row in the table.
Store_No: This is the store number assigned to that particular store by DFF.
City: City in which that store is located.
Zone: Zone in which that store is located.
Zipcode: Zipcode of the store.
Children_Count: Percentage share of children in the store.

3. **dimProduct**: This table stores the Product level descriptive information.



ProductID: Surrogate Key generated by the Data Warehouse to ensure uniqueness of each row in the table.
Product_Code: Universal Product Code.
Category: Denotes the umbrella category to which the product belongs.

REASONING AND STEPS FOR DATA MART DESIGN:

1. **What are the top-selling Cigarette SKUs in the Cigarette Product Category for DFF?**
   Dominick's Finer Foods will greatly benefit from these insights because they'll get a comprehensive overview of their sales performance within the Cigarette product category. This will help them serve their customers better by optimizing inventory management, targeted marketing, and strategic decision-making.

   Schema 1 is highly based on sales information which will assist in analysis. The dimProduct contains all the product related information required including product code, name and category. Similarly the factSales provides all the sales related metrics such as

sales, price and quantity. Hence the schema supports this business question and provides greater insights into product performance.

2. **In which months of the year do maximum promotional sales take place?**
   Getting a greater insight into the promotional sales performance over the year. DFF can optimize their promotional strategies and align them with the months that exhibit the greatest promotional sales. They can develop better promotional campaigns, enhance customer engagement and thereby maximize promotional Return on Investment.

   Schema 2 focusses on store data. The dimTime will provide time related information such as month. The factStore will provide data on promo sales which can then be aggregated on a monthly basis. Hence the schema supports this business question and provides greater insights into promotional sales.

3. **What are the various possibilities to maximize earnings based on high sales categories?**
   DFF can strategize effectively based on which categories provide maximum sales. This can help them optimize pricing, allocate resources more efficiently, enhance product placement thereby improve the overall earnings for the business.

   Schema 1 is highly based on sales information which will assist in analysis. The factSales provides all the sales related metrics such as sales, move, price, and quantity. Hence the schema supports this business question and provides greater insights into revenue and profits.

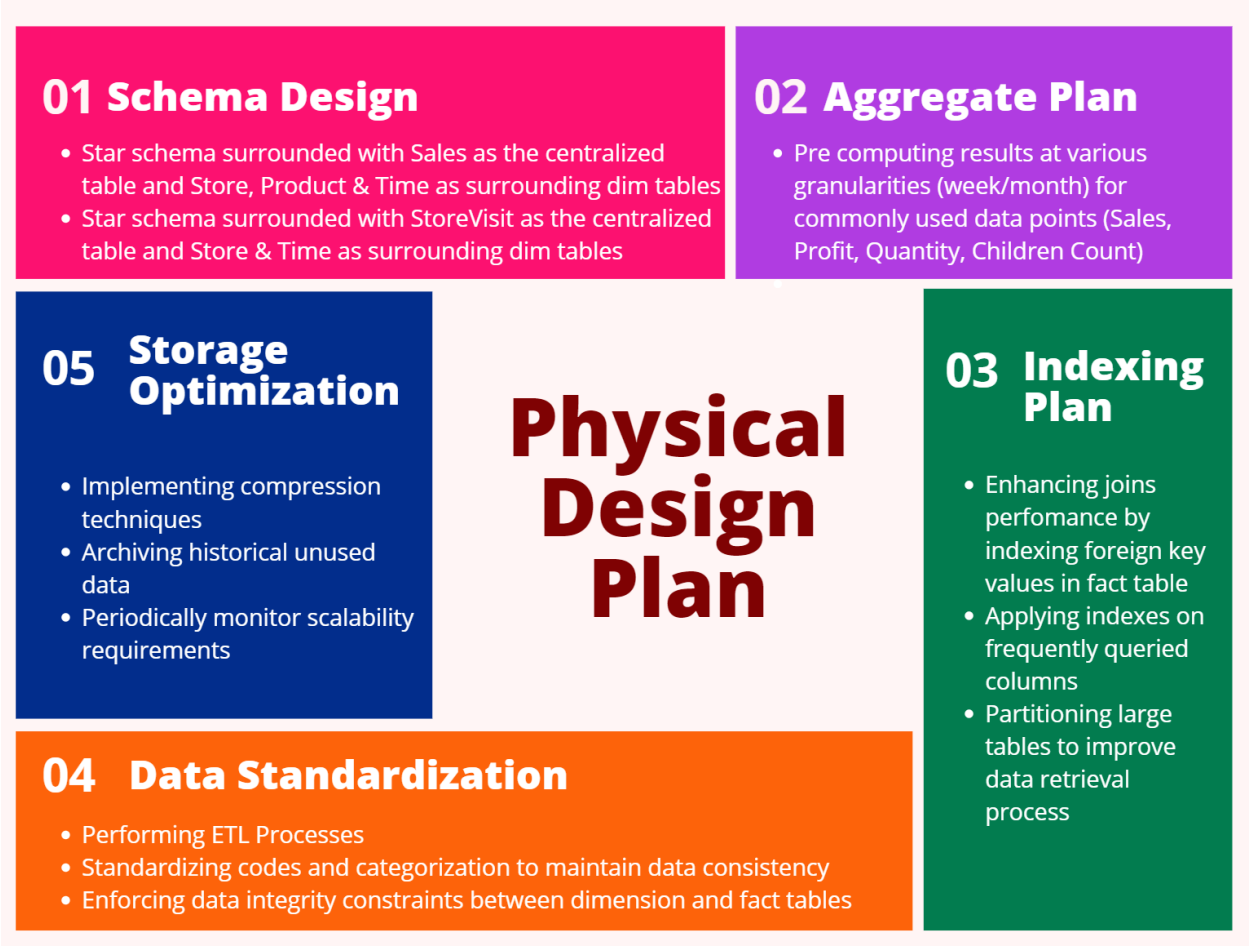
4. **What is the correlation between candy sales and the number (percentage) of children in a given demographic area?**
   Understanding the correlation between percentage of children and candy sales, DFF can enable targeted marketing, tailor product offerings, and enhance product placements. This will help provide a better user experience and improve sales.

   Schema 2 captures a lot of relevant information for this business question. dimStore contains store and its demographic information. Sales and quantity information can be pulled from the factSales table. Additionally, factStore provides children count which will help in creating a correlation between that and candy sales. Hence the schema supports this business question and provides greater insights into demographic correlations.

5. **What is the impact of festive occasions and other significant events on monthly floral sales?**
   Understanding how festive occasions and other events drive customer behavior and impact business, DFF can optimize floral inventory and perform better marketing. This way they can capitalize on seasonal demand, ensure inventory stock levels, and maximize sales.

   Within Schema 2, dimTime focusses on month and special events occurring over the year. Whereas the factStore will capture floral sales. Hence the schema supports this business question and provides greater insights into monthly floral sales thereby improving seasonal marketing and inventory planning.

PHYSICAL DESIGN



While planning for an efficient physical schema, it was essential to keep in mind future scope and upgrade possibilities. We followed a strategic approach prioritizing system performance, data consistency, integrity and scalability. Schema 1 focusses highly on sales related data, capturing transactions and providing a comprehensive analysis of sales performance with metrics such as sales, profit, quantity, and price. Schema 2 focuses on store data. It additionally captures metrics like promo sales, floral sales and number of children giving insights into store demographics and performance. This schema is designed to support footfall, promotional effectiveness, etc. Choosing a star schema for the physical design plan will not only help us deliver high performance analytics but also be scalable as we grow.

Employing data aggregations techniques such as pre computing results, designing clustered and non clustered indexes on primary and foreign keys, enforcing integrity constraints, and partitioning will together help us build a robust framework. This design plan facilitates quick

data retrieval, optimized query performance, and maintains data integrity. Periodic checks and monitoring will be highly essential to keep the data warehouse responsive and effective over time with constantly evolving business requirements.

SECTION 4: DATA CLEANING AND INTEGRATION

DATA QUALITY ISSUES

Following were the data quality issues encountered while performing the ETL process:

1) Columns such as Date were not in appropriate format, with many values being inconsistent with the standard date formats.
2) Presence of both positive and negative values for attributes like coupon, which do not make sense in the business context.
3) In files Demo.csv and Cccount.csv, there were multiple inconsistent values. These values were removed by using SQL operations.
4) Multiple columns in the source files had missing values, which had to be handled separately.
5) Other issues included data with no meaning such as '.' and negative values for columns like Store and Week.

ETL PLAN

ETL is a crucial part of the data warehouse design and implementation. Using ETL, all the raw data can be converted into useful insights that can be used to address business questions. Although there are different approaches to performing this operation, such as ELT(Extract Load Transform), we implemented ETL in 3 phases as described below.

We adopted an iterative approach to implementing the ETL plan.
Firstly, we loaded the data as is from the source files into the temporary tables, without filtering columns or performing any data type conversions. Once the data was loaded, we analyzed the data, identifying missing values, garbage data, and any potential formatting issues. In the next iteration, we started with data cleaning, creating SQL queries and rules to make the data consistent. All these transformations were performed while loading data into staging from the temporary tables. Finally, in the third iteration, we loaded the cleaned data into the final data mart tables. The purpose of temporary tables was to just load the entire data and analyze the various scenarios to deal with as part of data cleaning/pre-processing.

In performing the entire ETL process, we ensured that the right data types were used and there was no loss of data and the integrity of data was maintained across all the staging, dimension and

fact tables. All the operations were performed in Microsoft SQL Server Integration Services(SSIS) using distinct packages corresponding to each of the ETL stages.

TARGET DATA

The target data was obtained by mapping the various data sources to create the staging tables and final data marts as follows:

Data Mart Tables	**Data Source**	**Source File Name**	**Staging Area Tables**
dimStore	Store	Stores.csv	STORES_STAGING
dimTime	Week Mappings	Time.csv	TIME_STAGING
dimProduct	Movement files	Wxxx.csv	FINAL_MOVEMENT_STAGING
factStore	Customer count Store Time	Ccount.csv Stores.csv Time.csv	CCOUNT_FINAL_STAGING STORES_STAGING TIME_STAGING
factSales	Movement data Store Time	Wxxx.csv Stores.csv Time.csv	FINAL_MOVEMENT_STAGING STORES_STAGING TIME_STAGING

DATA SOURCES

Following source files were used from the dataset:

Data Source	**Source File**
Customer Count	Ccount.csv
Demographics	Demo.csv

Time	Time.csv
Bottle Juice Movement	DONE-WBJC.csv
Cheese Movement	Done-WCHE.csv
Cigarettes Movement	Done-WCIG.csv
Cookies Movement	DONE-WCOO.csv
Crackers Movement	Done-WCRA.csv
Canned Soup Movement	WCSO-Done.csv
Dish Detergent Movement	WDID-Done.csv
Front-End-Candies Movement	WFEC.csv
Frozen Dinners Movement	WFRD.csv
Frozen Entrees Movement	WFRE-Done.csv
Frozen Juices Movement	WFRJ.csv
Fabric Softener Movement	WFSF-done.csv
Grooming Products Movement	WGRO.csv
Laundry Detergents Movement	wlnd.csv
Soft Drinks Movement	wsdr.csv
Soap Movement	WSOA.csv
Toothbrushes Movement	WTBR_done.csv
Canned Tuna Movement	WTNA_done.csv
Toothpastes Movement	WTPA_done.csv
Toilet Papers Movement	WTTI.csv

DATA MAPPING

Mapping Table 1 - Source Data to Staging Tables

Source data	Source data field	Mapping	Staging Table Name	Attribute
DEMO.csv	STORE	Copy	DEMO_STAGING	Store_No
	AGE9	Copy		Children_Count
CCOUNT.csv	WEEK_NO	Copy	CCOUNT_FINAL_STAGING	Week
	PROMO	Copy		Promo_Sales
	FLORAL	Copy		Floral_Sales
	STORE	Copy		Store_No
WXXX.csv	MOVE	ok=1 (valid data)	FINAL_MOVEMENT_STAGING	Move
	QTY	ok=1 (valid data)		Qty
	PRICE	ok=1 (valid data)		Price
	SALES	(Price * Move) / Qty		Sales
	UPC	ok=1 (valid data)		Product_Code
	WEEK_NO	ok=1 (valid data)		Week_No
	STORE	ok=1 (valid data)		Store_No
	N/A	Derived Column		Category
TIME.csv	WEEK	Copy	TIME_STAGING	Week

	SPECIAL_EVENTS	Copy		Special_Events
	MONTH	Copy		Month
	YEAR	Copy		Year
	START	Copy		Start_Date
	END	Copy		End_Date

Mapping Table 2 - Staging Tables To Presentation Server Tables

Source data in Staging	**Staging data field**	**Mapping**	**Data Mart Table Type**	**Table Name**	**Attribute**
DEMO_STAGING	Store_Number	Copy	DIMENSION	dimStore	Store_No
	Children_count	Join with STORE_STAGING			Children_count
CCOUNT_FINAL_STAGING	Promo_Sales	Aggregation	FACT	factStore	Promo_Sales
	Floral_Sales	Aggregation			Floral_Sales
FINAL_MOVEMENT_STAGING	Price	Copy	FACT	factSales	Price
	Move	Copy			Move
	Qty	Copy			Qty
	Sales	Copy			Sales
	ProductID	Foreign key from dimProduct			ProductID (FK)
	ProductID	Surrogate key	DIMENSION	dimProduct	ProductID (PK)

	Product_Code	Copy			Product_Code
	Category	Copy			Product_Category
TIME_STAGING	TimeID	Foreign Key from dimTime	FACT	factStore	TimeID (FK)
	TimeID	Foreign Key from dimTime	FACT	factSales	TimeID (FK)
	TimeID	Copy	DIMENSION	dimTime	TimeID (PK)
	Week_No	Copy			Week_No
	Special_Event	Copy			Special_Events
	Month	Copy			Month
	Year	Copy			Year

DATA EXTRACTION RULES

While building the data warehouse, extracting the right data with the right number of attributes is crucial to integrating data from different sources. In the given dataset for DFF, we selected the files which were required to address our business questions.

The idea in the extraction phase was to make the format consistent for all the source files prior to integrating them. The format we decided was csv. So, we first converted all the individual files into the csv file format prior to extraction.

Below are the rules we followed as part of extraction:

1) Make all the source file format consistent, that is, in csv.
2) All the columns in the selected source files in the initial extraction process would be extracted as is into the temporary tables.
3) The naming convention would be to keep the name of the temporary tables as the name of the source file with a '_TMP' added to the table name.
   Ex. Temporary table for Demo.csv source file was DEMO_TMP
   Temporary table for Ccount.csv was CCOUNT_TMP

4) The columns/fields required to build the data marts and address the business questions would be filtered while loading the data from the temporary tables to the staging tables.

The extraction was performed using the Import/Export wizard of the Microsoft SQL Server Integration Services (SSIS).

DATA TRANSFORMATION AND CLEANSING RULES

- **Remove Invalid Data:** All the invalid data was removed, by filtering out records where 'ok' column has a value of 0
- **Data Cleaning:** Data was cleaned by removing dirty data and data with no meaning, such as '.' and negative values for Store and Week columns.
- **Data Type Conversion:** All the data was initially loaded as varchar data type by default. The numeric columns were transformed by doing the data type conversions for the following attributes:

TABLE	COLUMNS
CCOUNT_STAGING	WEEK, FLORAL_SALES, PROMO_SALES
TIME_TMP	WEEK, MONTH, YEAR
MOVEMENT_STAGING	QTY, MOVE, PRICE, PROFIT

- **Derived Columns:** SALES column is a derived column calculated by using the following formula:
  Sales = (Price * Move)/Qty
- **Lookup:** All the primary keys were added as foreign keys to the fact table using lookup transformation function
- **Union:** Data from all movement temporary tables were combined and loaded into a single staging table MOVEMENT_STAGING using the UNION ALL operator.
- **Creating Surrogate Keys:** Following surrogate keys were created for dimension tables**:**

Table	**Surrogate Key**

dimStore	StoreID
dimTime	TimeID
dimProduct	ProductID

PLAN FOR AGGREGATE TABLES

Data aggregation can be utilized to improve execution time of the queries. Aggregated values in the fact tables facilitate reporting procedures and enhance performance of queries as the metrics required for analysis are precomputed at some level. Both the fact tables have some form of aggregated data, which will help us address the business questions that we have devised.
Our data warehouse consists of the following aggregations:

factSales

- The sales value is aggregated based on product category, by calculating sum of sales
- The sales value is aggregated for cigarettes based on store, using sum function

factStore

- The floral sales and promotional sales are summed on the basis of week number and store in the ‘factStore’ fact table

ORGANIZATION OF DATA STAGING AREA

Staging area forms the base for all the transformations. Data extracted from all source files are first loaded into the temporary tables.
We have created two databases for our data warehouse: one for the staging area and one for the data marts. The database serving as the staging area, ‘focus-group-staging’, has all the temporary tables and the staging tables. Here, the data is first loaded directly into the temporary tables from the source files without manipulation and filtering. Then, all the cleaning and transformations

were performed while loading the data into the staging tables, which reside in the staging database.

PROCEDURE FOR DATA EXTRACTION AND LOADING

DATA EXTRACTION

In the extraction process, connection was first established to the Mays server (infodata16.mbs.tamu.edu) to gain access to the staging database.
The Import/Export wizard of SSIS was used to create a package for the import, which has a combination of SQL preparation and Data Flow task.

1) CCOUNT_FINAL_STAGING

- In this import, Ccount.csv file was selected as the flat file source.
  The data was loaded into the temporary table CCOUNT_TMP. All the attributes were loaded in the process, with the name of columns being the same as that in the csv file.
- In the next stage, columns were filtered and only the required ones were imported into the staging table, CCOUNT_FINAL_STAGING
- Data type conversions were performed during the process

CCOUNT.CSV	**CCOUNT_FINAL_STAGING**
STORE_NO	STORE_NO
FLORAL_SALES	FLORAL_SALES
PROMO_SALES	PROMO_SALES
WEEK	WEEK

2) DEMO_STAGING

- In this import, the Demo.csv file was selected as the source flat file.
  The data was loaded into the temporary table DEMO_TMP.

- The structure of the data was preserved during the load, with the name of columns being the same as that in the csv file.
- Fields were filtered and only the required columns were loaded into the staging table

DEMO.CSV	DEMO_STAGING
STORE	STORE_NO
AGE9	CHILDREN_COUNT

- The children_count column was converted to **int** during the loading process of DEMO_STAGING from DEMO_TMP

3) STORES_STAGING

- The data for this table was taken from Dominick's manual
- We manually created stores.csv using the information available in the manual
- Connection was first established to the Mays server (infodata16.mbs.tamu.edu) to gain access to the staging database
- The Import/Export wizard of SSIS was used to create the package for the import, which has a combination of SQL preparation and Data Flow task.

- The table created as a result of the operation, STORES_STAGING served as the input for the dimStore dimension table and contains information about each store

STORES.csv	STORES_STAGING
STORE_NO	STORE_NO
CITY	CITY
ZONE	ZONE
ZIPCODE	ZIPCODE

4) FINAL_MOVEMENT_STAGING

- In this extraction process, temporary tables were created for each movement file and data was extracted from the files into these tables.
- The data from each temporary table was then combined using the UNION operation and loaded into the FINAL_MOVEMENT_STAGING table.

WXXX.CSV	**FINAL_MOVEMENT_STAGING**
STORE	STORE_NO
WEEK_NO	WEEK_NO
MOVE	MOVE
QTY	QTY
PRICE	PRICE
SALES	SALE
UPC	PRODUCT_CODE

5) **TIME_STAGING**

- In this import, the Time.csv file was selected as the source flat file.
- The data was loaded directly into the staging table, TIME_STAGING.

DATA LOADING PROCEDURE

1) **dimStore**
    - Created the dimStore table using the Execute SQL task of SSIS.
    - A surrogate key, StoreID, was created in dimStore table
    - Data was loaded from the STORE_STAGING table into dimStore table
    - DEMO_STAGING was joined with STORE_STAGING to load the Children_Count column in the dimStore table

2) **dimTime**
    - Created the dimTime table using the Execute SQL task of SSIS.

- A surrogate key, TimeID, was created in the dimTime table.
- Data was loaded from the TIME_STAGING table into the dimTime table using the Import/Export wizard.

3) factStore

- Created the factStore table using the Execute SQL task of SSIS.
- A surrogate key, FactStoreVisitID, was created in the factStore table.
- Floral_Sales and Promo_Sales were aggregated in the table based on week and store number.
- Lookup functions were used to load the foreign keys:
    - Lookup function1 to load StoreID from dimStore
    - Lookup function2 to load TimeID from dimTime

4) dimProduct

- Created the dimProduct table using the Execute SQL task of SSIS.
- A surrogate key, ProductID was created in the table.
- Data was loaded from FINAL_MOVEMENT_STAGING into this table.
- Only selected columns, Product_Code and Category were loaded into this table.

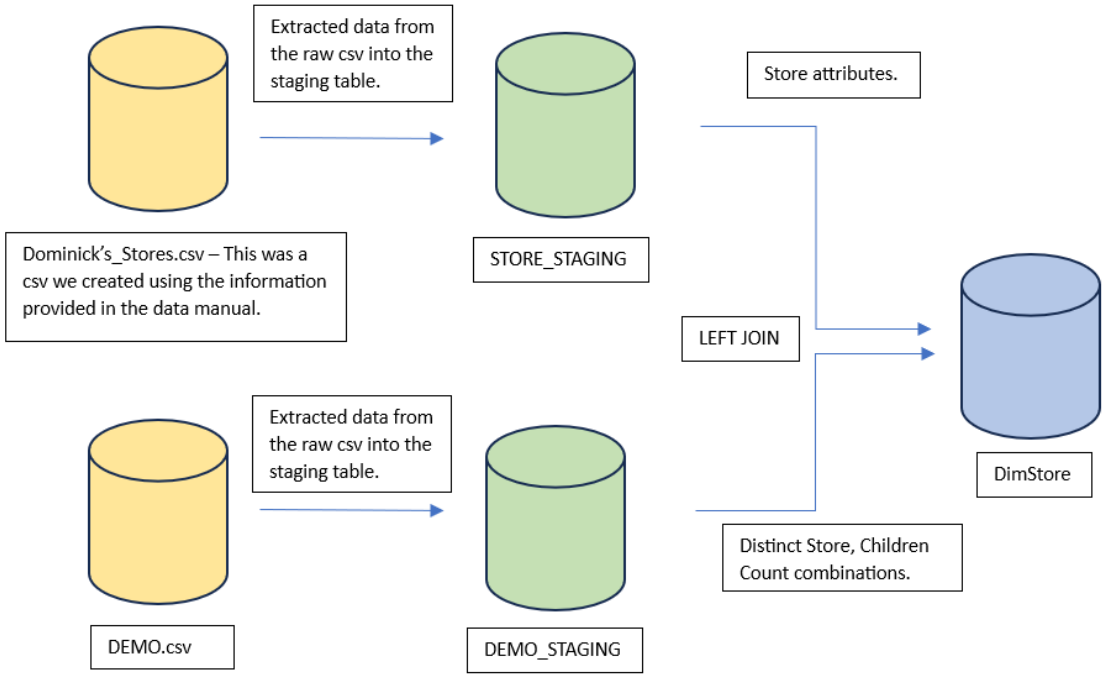
5) factSales

- Created the dimProduct table using the Execute SQL task of SSIS.

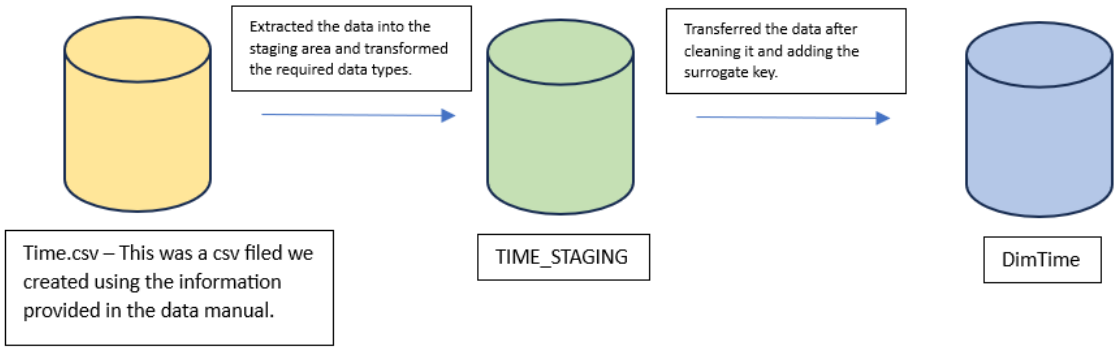
- A surrogate key, FactSalesID was created in this table.
- Columns StoreID, ProductID and TimeID were added as foreign keys from the respective tables using lookup operation.
- Columns Sales, Qty, Move and Price were loaded from the FINAL_MOVEMENT_STAGING table.

ETL FOR DIMENSION TABLES

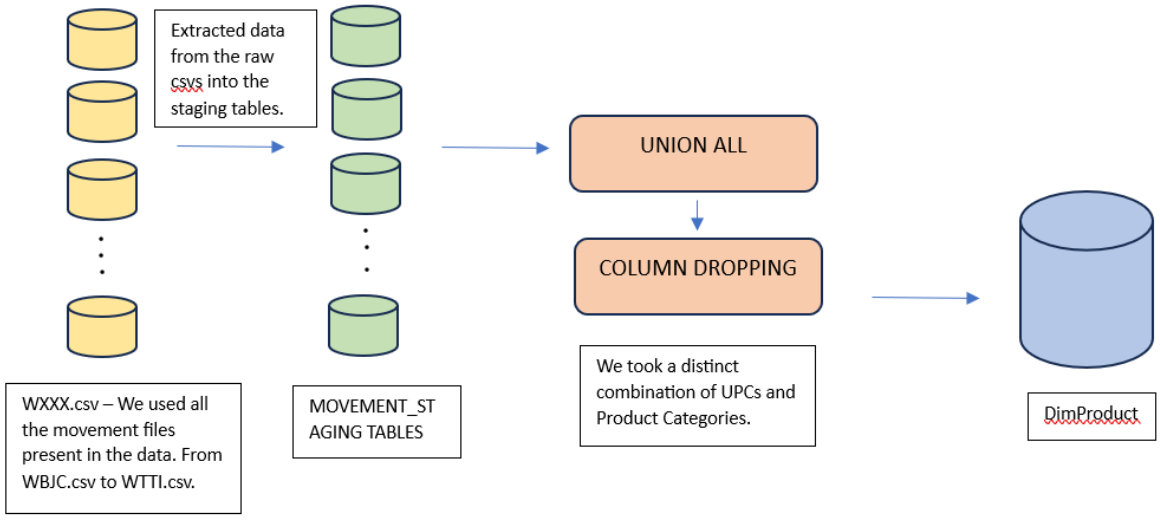
1. DimStore



2. DimTime

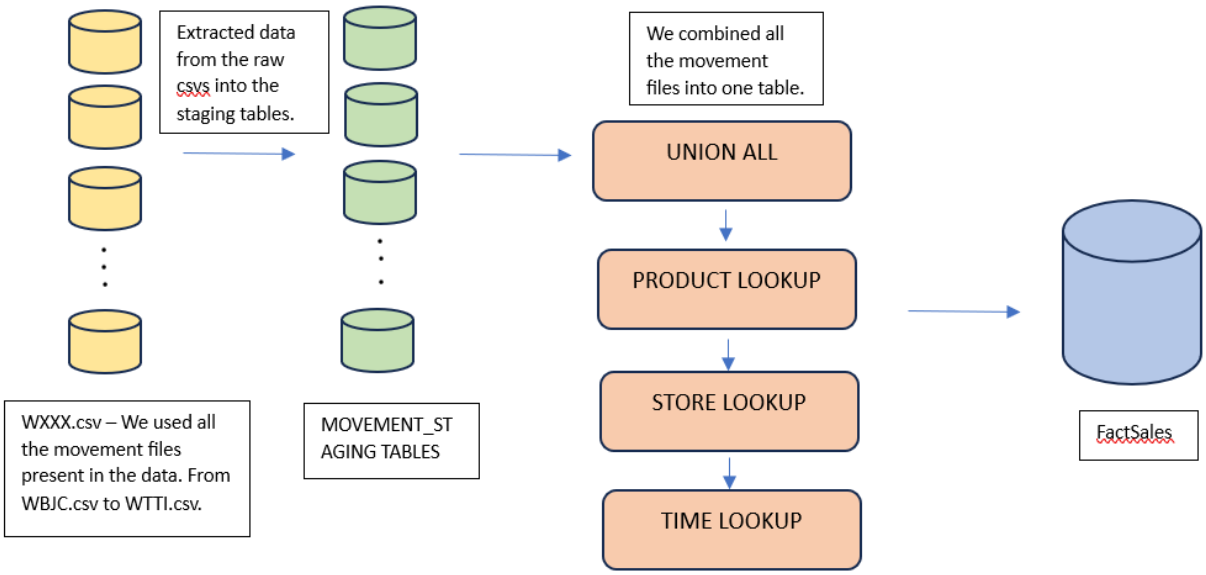


3. DimProduct

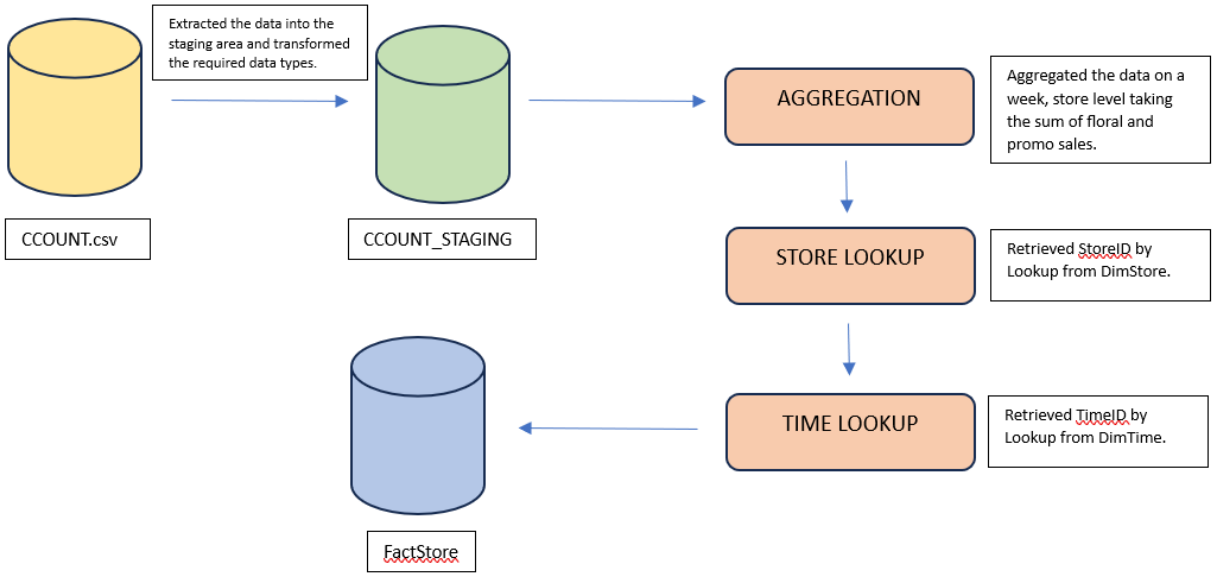


ETL FOR FACT TABLES

1. FactSales



2. FactStore



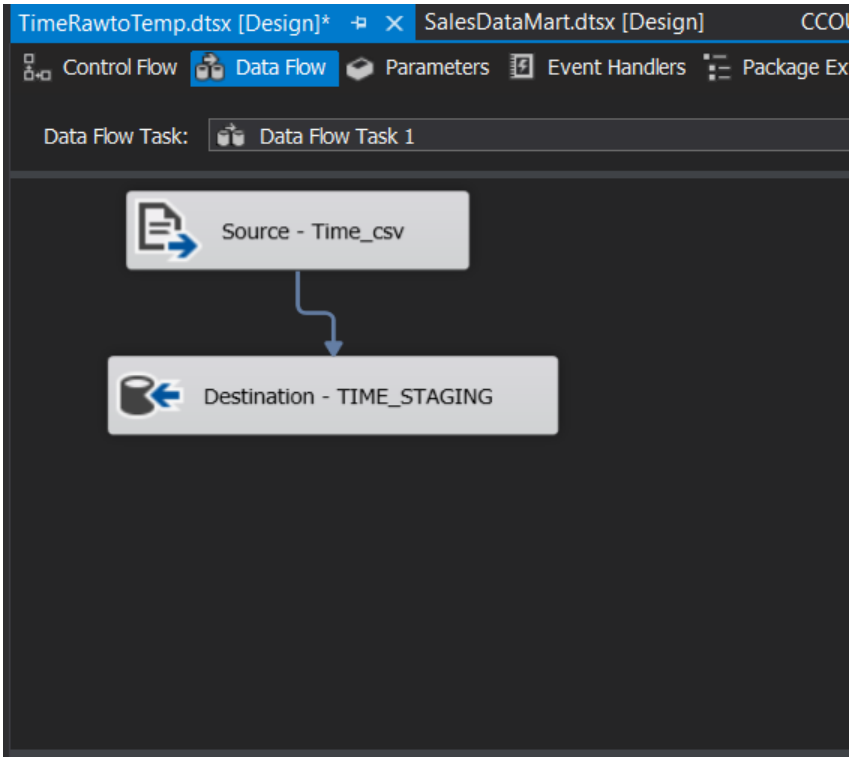
ETL IMPLEMENTATION

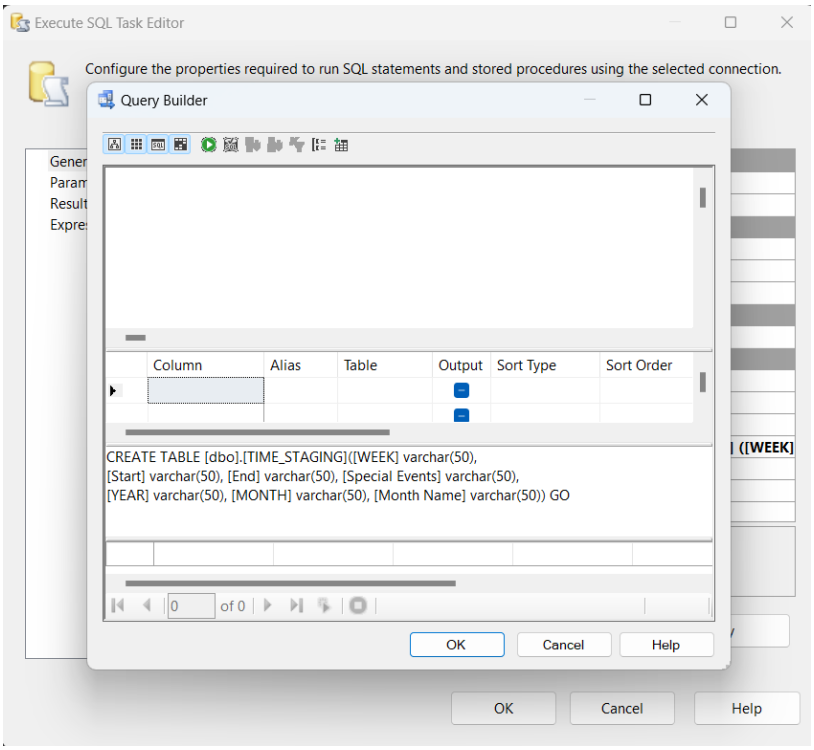
DATA MART 1: STORE DATA MART (STORE LEVEL SALES)

DimTime

EXTRACTION

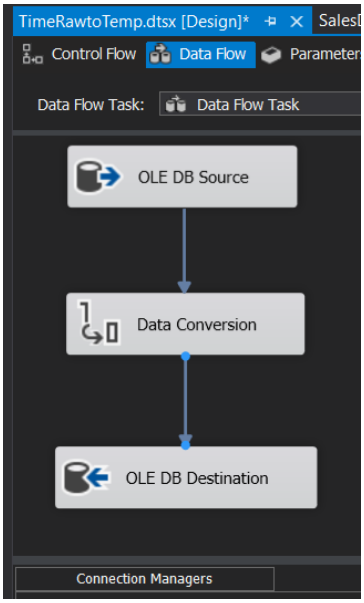
The Time.xlsx was created using the data manual of Dominick's website and then converted into CSV format for the ease of loading into SSIS.

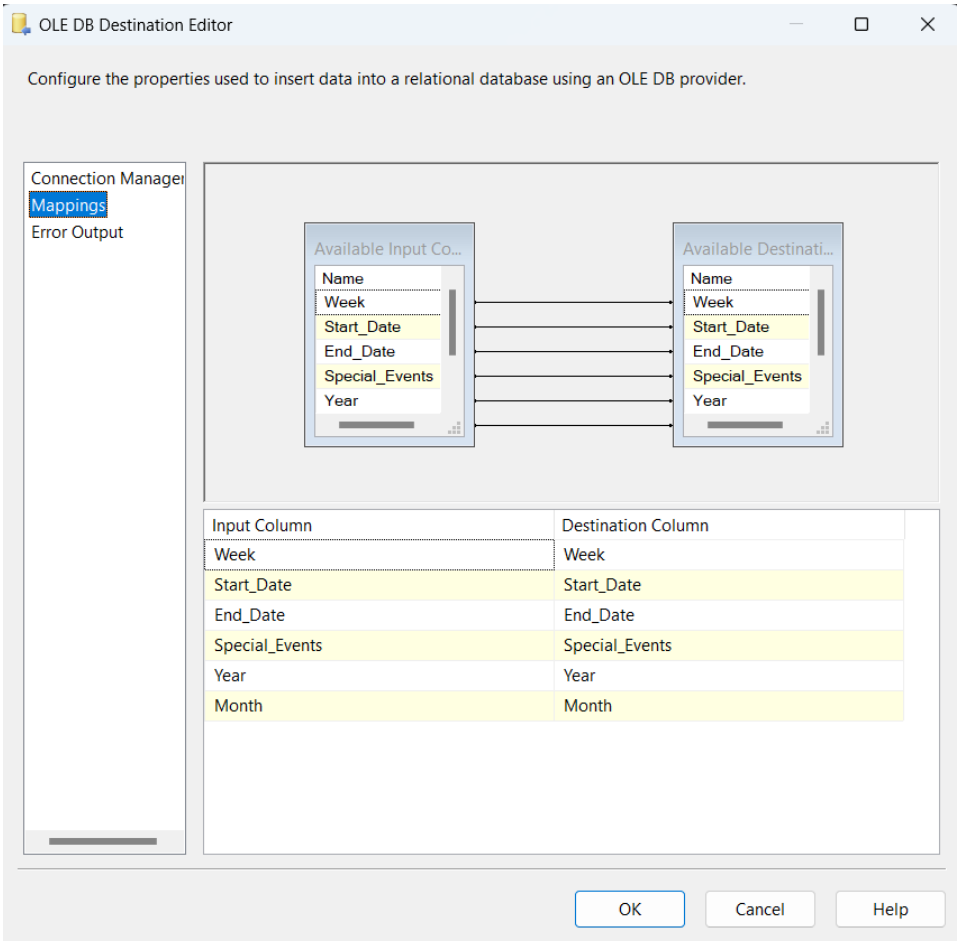




TRANSFORMATION

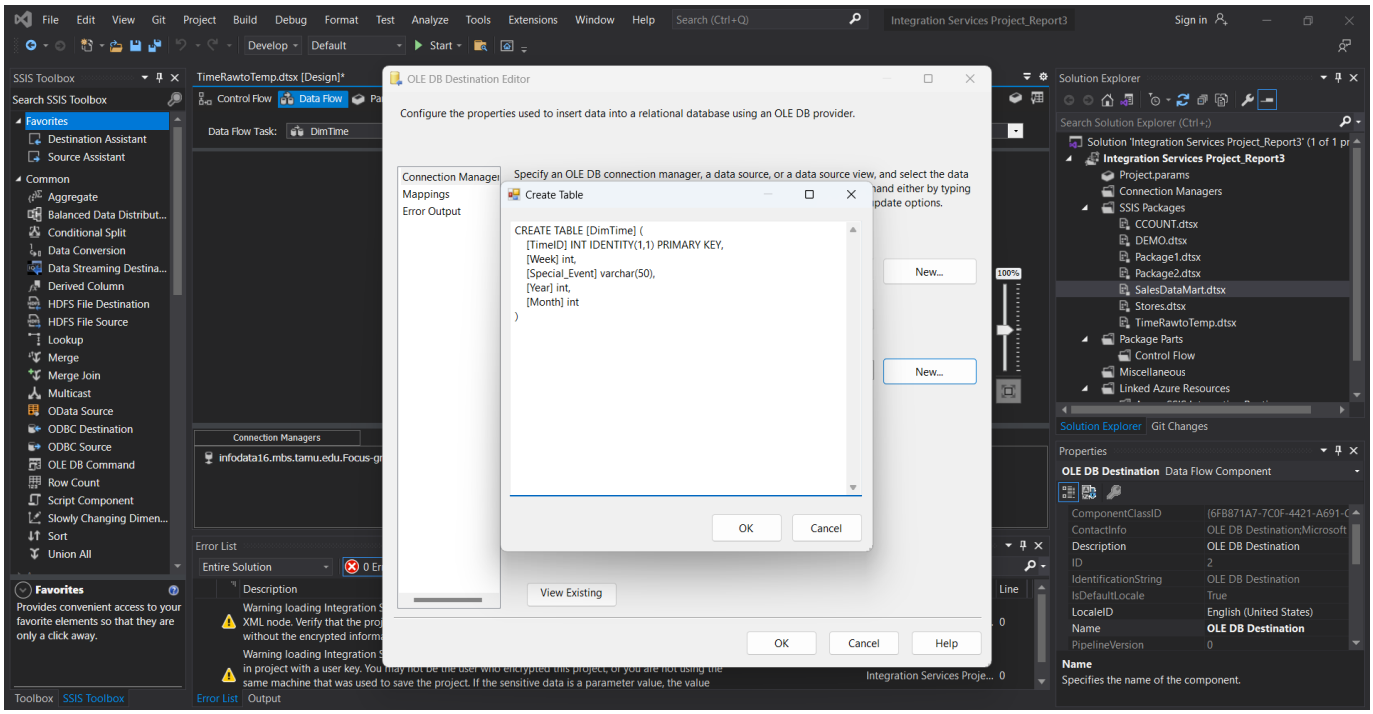
The required columns were selected in the staging table and some of the data types of certain columns were changed.





LOADING

Finally, the data from the staging table is loaded into the Data Mart DimTime table with the new surrogate key TimeID generated by the database.



DimTime Table in the Data Mart

The following query gives us a view of the data in the DimTime table.

SQLQuery2.sql - in...4N8L2\Harsh (365)) | SQLQuery1.sql - in...4N8L2\Ha

```
SELECT TOP (1000) [TimeID]
      ,[Week_No]
      ,[Special_Event]
      ,[Year]
      ,[Month]
  FROM [Focus-group-Sales-Data-Mart].[dbo].[DimTime]
```

100 %

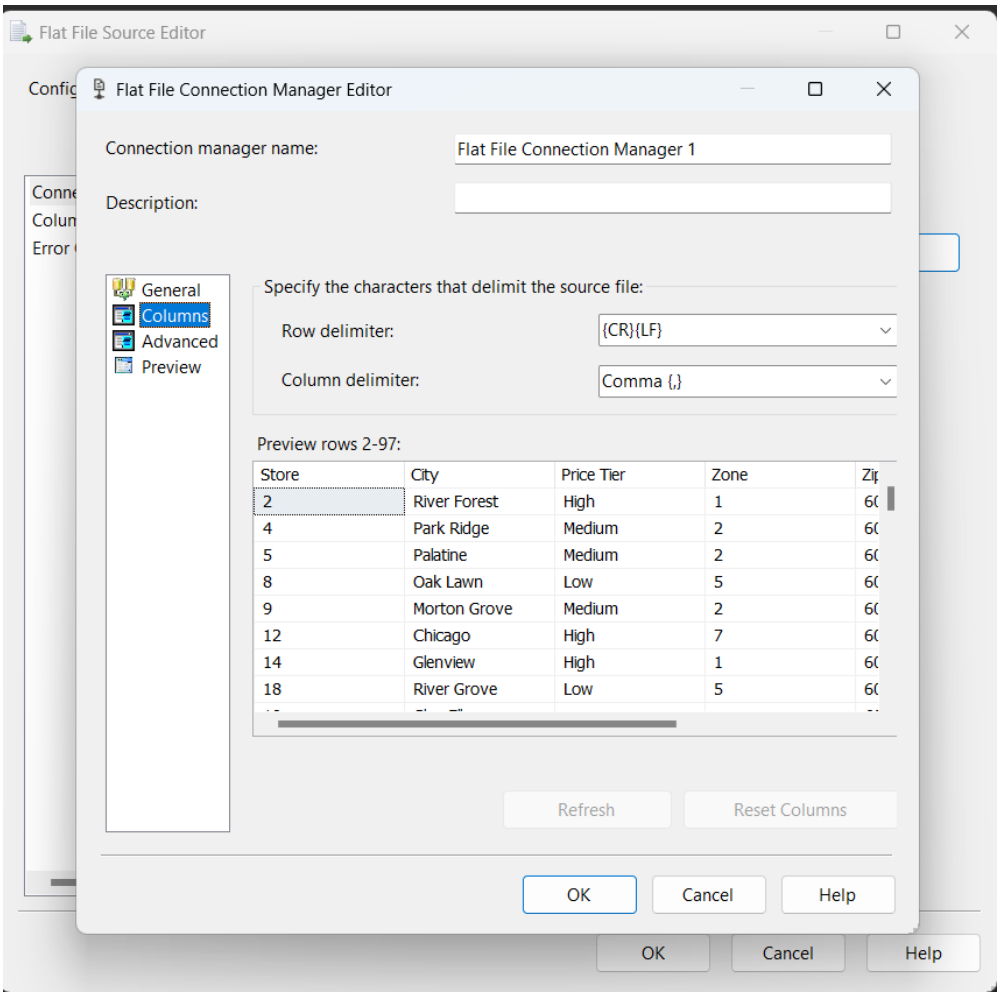
Results | Messages

	TimeID	Week_No	Special_Event	Year	Month
1	1	1		1989	9
2	2	2		1989	9
3	3	3		1989	9
4	4	4		1989	10
5	5	5		1989	10
6	6	6		1989	10
7	7	7	Halloween	1989	10
8	8	8		1989	11
9	9	9		1989	11
10	10	10		1989	11
11	11	11	Thanksgiving	1989	11
12	12	12		1989	11
13	13	13		1989	12
14	14	14		1989	12
15	15	15	Christmas	1989	12
16	16	16	New-Year	1989	12
17	17	17		1990	1
18	18	18		1990	1
19	19	19		1990	1
20	20	20		1990	1
21	21	21		1990	2
22	22	22		1990	2
23	23	23	Presidents Day	1990	2
24	24	24		1990	2
25	25	25		1990	3
26	26	26		1990	3
27	27	27		1990	3
28	28	28	Easter	1990	3

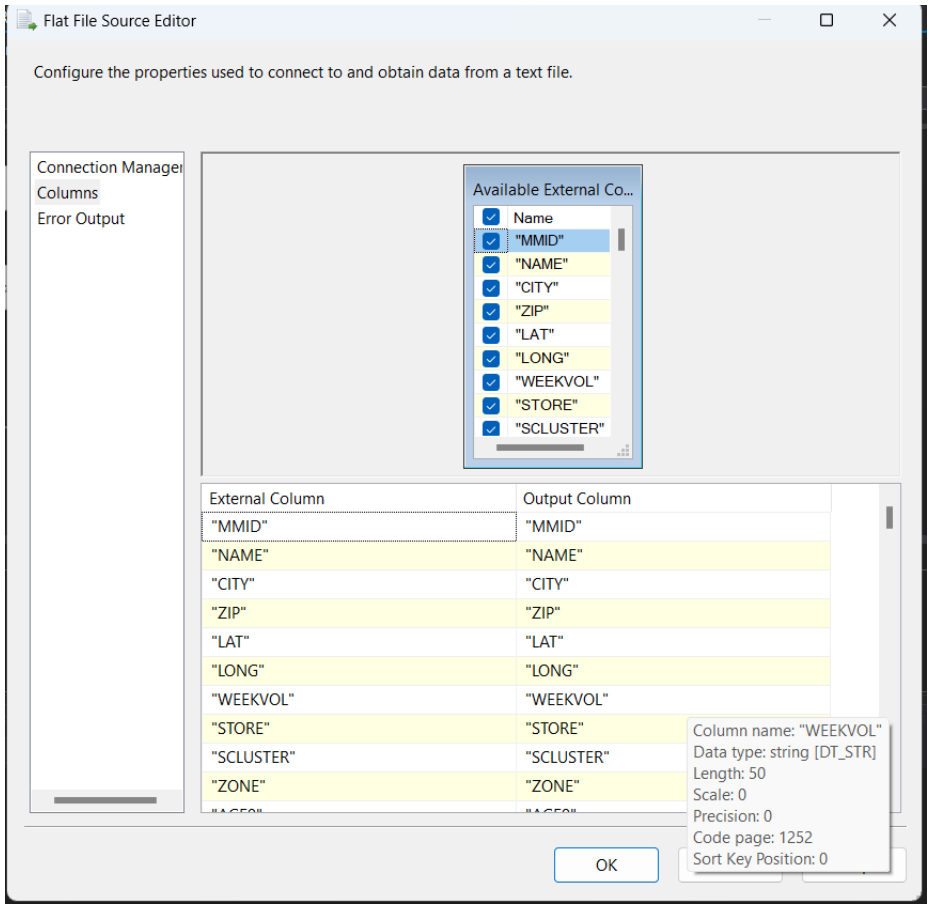
DimStore

EXTRACTION

The Stores.csv was created using data provided in the data manual containing columns such as City, Zip Code, Address of the stores.



We also extracted the DEMO.csv file as we wanted to get the Children_Count attribute per store in our DimStore Table.

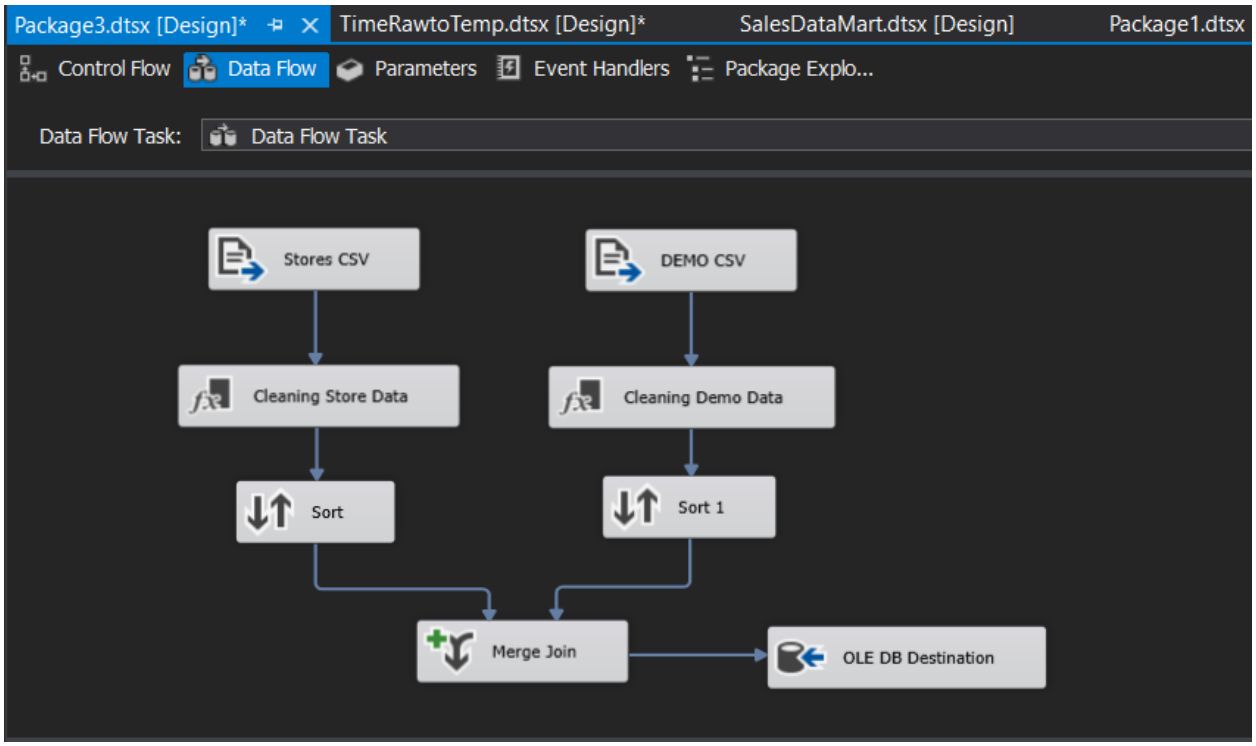


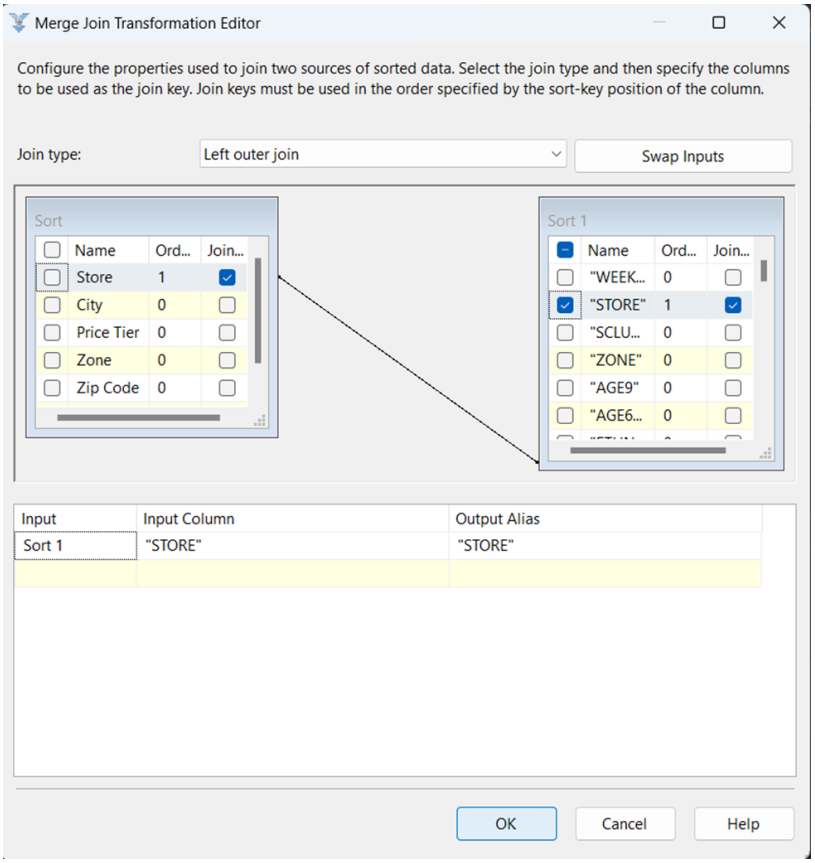
TRANSFORMATION

We first cleaned the Store and Demo data using the Derived Column tool in SSIS. We used queries like:

```
SET Store_No = REPLACE(REPLACE(REPLACE(Store_No, '"', ''), '%', ''), '<', '');
```

To clean our data and make sure it was ready to be joined.
After that we joined the Store Data with the Demo data to get the Children_Count attribute in our DimStore table eventually.

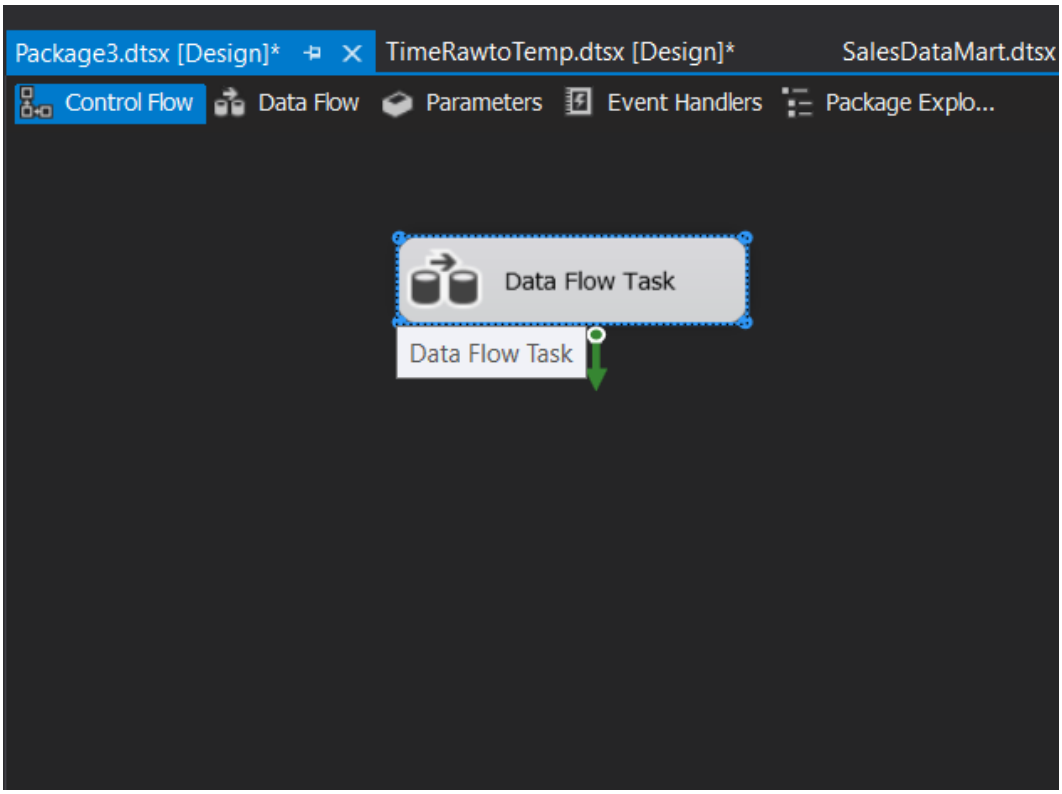

Package3.dtsx [Design]*
TimeRawtoTemp.dtsx [Design]*
SalesDataMart.dtsx [Design]
Package1.dtsx
Control Flow
Data Flow
Parameters
Event Handlers
Package Explo...
Data Flow Task:
Data Flow Task
Stores CSV
DEMO CSV
Cleaning Store Data
Cleaning Demo Data
Sort
Sort 1
Merge Join
OLE DB Destination

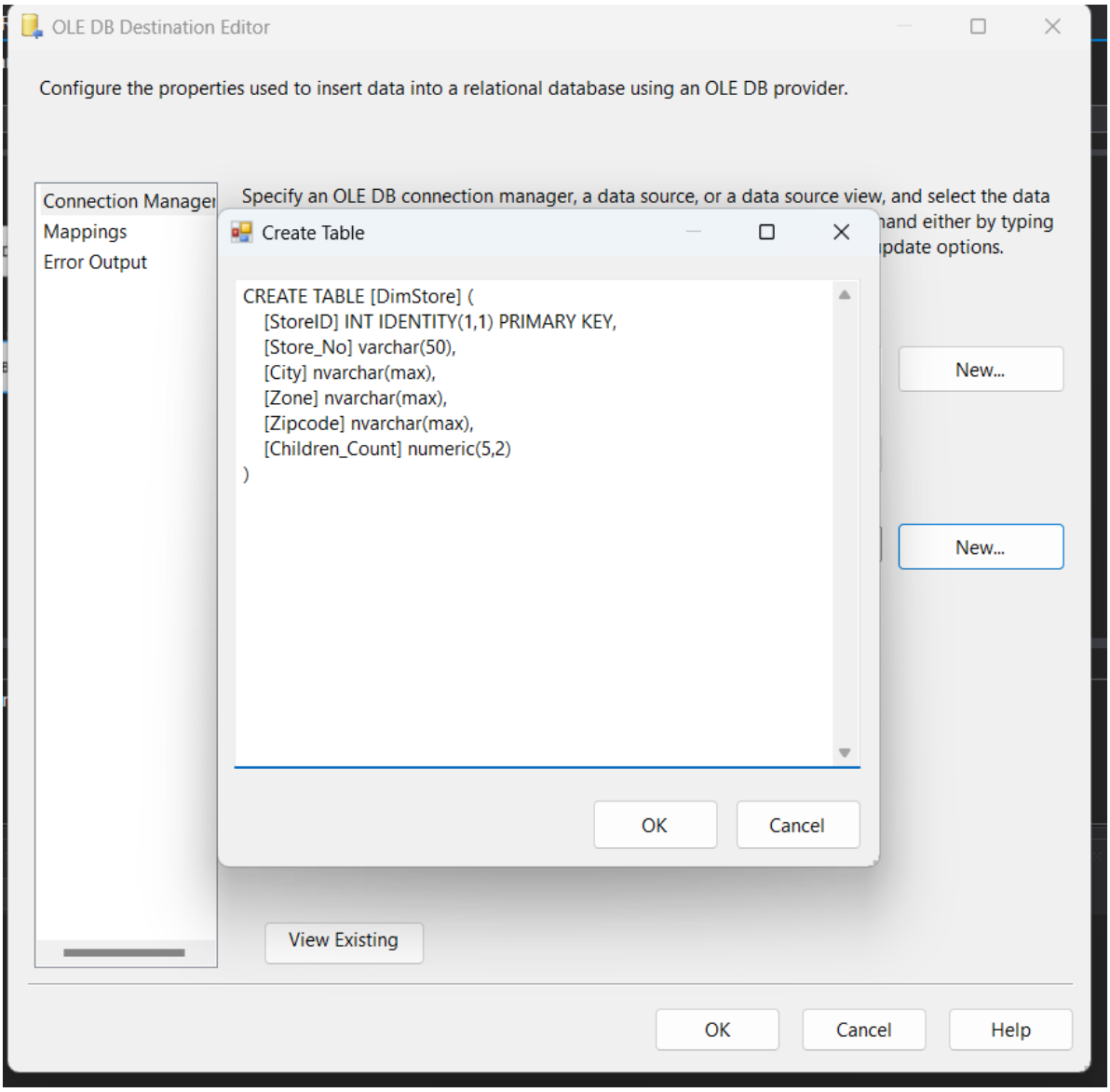

Merge Join Transformation Editor
Configure the properties used to join two sources of sorted data. Select the join type and then specify the columns to be used as the join key. Join keys must be used in the order specified by the sort-key position of the column.
Join type:
Left outer join
Swap Inputs
Sort
Name
Ord...
Join...
Store
City
Price Tier
Zone
Zip Code
Sort 1
"WEEK...
"STORE"
"SCLU...
"ZONE"
"AGE9"
"AGE6...
Input
Input Column
Output Alias
Sort 1
"STORE"
"STORE"
OK
Cancel
Help

The output of the join was stored in the Stores_Staging table in the staging area.

LOADING

The final output was loaded into the DimStore table along with the StoreID surrogate key generated by the database.




OLE DB Destination Editor
Configure the properties used to insert data into a relational database using an OLE DB provider.
Connection Manager
Mappings
Error Output
Specify an OLE DB connection manager, a data source, or a data source view, and select the data
and either by typing
update options.
New...
New...
Create Table
CREATE TABLE [DimStore] (
[StoreID] INT IDENTITY(1,1) PRIMARY KEY,
[Store_No] varchar(50),
[City] nvarchar(max),
[Zone] nvarchar(max),
[Zipcode] nvarchar(max),
[Children_Count] numeric(5,2)
)
OK
Cancel
View Existing
OK
Cancel
Help

DimStore Table in the Data Mart

The following query will give us a view of the DimStore table in the data mart. The data is at a store level with various attributes belonging to each store.

SQLQuery2.sql - in...4N8L2\Harsh (365)) SQLQuery1.sql - in...4N8L2\Harsh (2

```
SELECT TOP (1000) [StoreID]
      ,[Store_No]
      ,[City]
      ,[Zone]
      ,[Zipcode]
      ,[Children_Count]
  FROM [Focus-group-Sales-Data-Mart].[dbo].[DimStore]
```

100 %

Results Messages

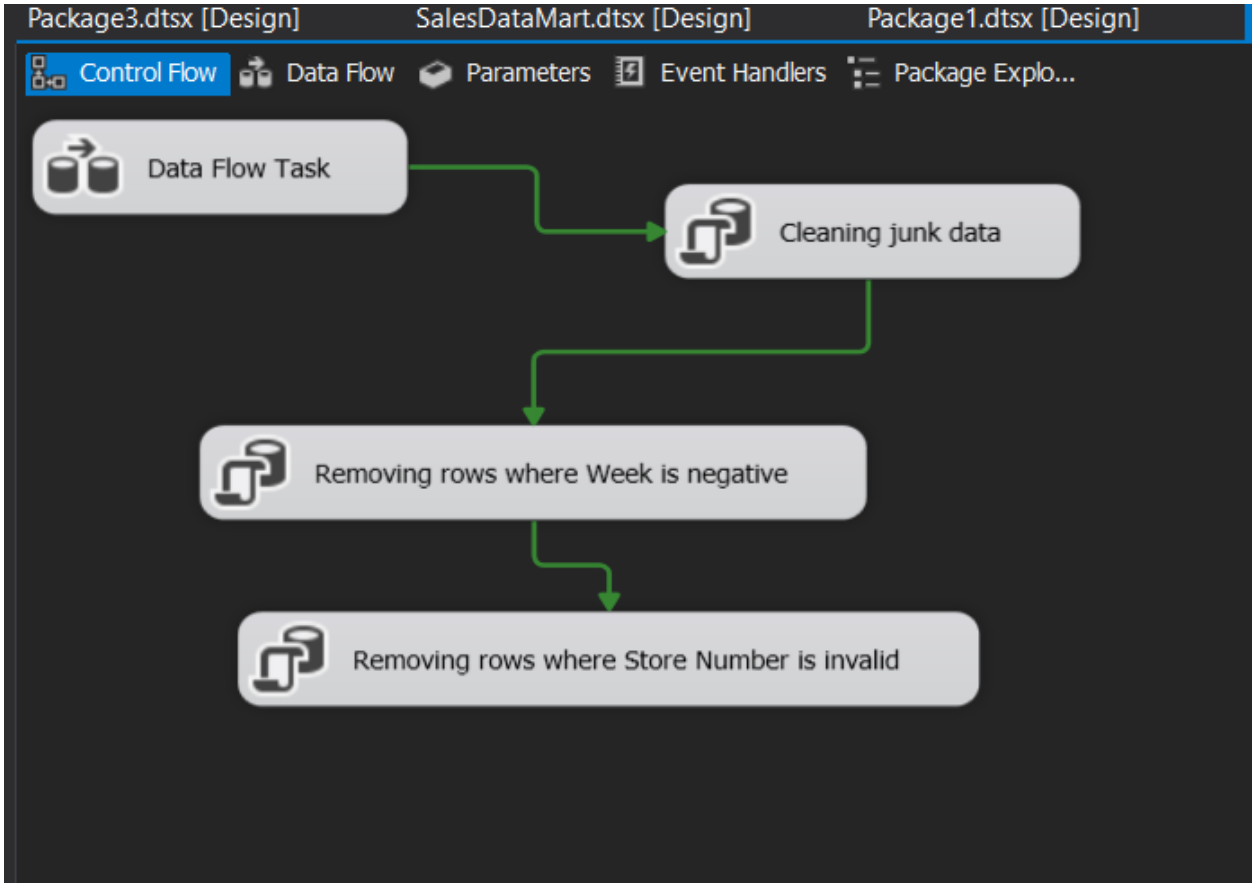
	StoreID	Store_No	City	Zone	Zipcode	Children_Count
43	43	75	Chicago	7	60640	0.11
44	44	76	Chicago	2	60618	0.14
45	45	77	Vernon Hills	6	60061	0.17
46	46	78	Downers Grove	6	60516	0.15
47	47	80	Arlington Hei...	6	60005	0.13
48	48	81	Mt. Prospect	2	60056	0.11
49	49	83	Lansing	6	60438	0.12
50	50	84	Orland Park	2	60462	0.16
51	51	86	Chicago	2	60618	0.14
52	52	88	Bensenville	2	60106	0.13
53	53	89	Chicago	2	60632	0.15
54	54	90	Chicago	10	60617	0.12
55	55	91	Oak Lawn	2	60453	0.11
56	56	92	Hazel Crest	2	60429	0.14
57	57	93	Evanston	1	60202	0.11
58	58	94	Bloomingdale	5	60108	0.16
59	59	95	Chicago	1	60634	0.11
60	60	97	Aurora	8	60506	0.16
61	61	98	Chicago	12	60638	0.11
62	62	100	Chicago	11	60698	0.18
63	63	101	Des Plaines	12	60016	0.11
64	64	102	Merrionette P...	15	655	0.14
65	65	103	Bolingbrook	15	60439	0.18
66	66	104	St. Charles	8	60174	0.16
67	67	105	Melrose Park	12	60160	0.14
68	68	106	Montgomery	8	60538	0.18
69	69	107	Westchester	2	60153	0.11

ISTM 637 - Group 5 Report

FactStore

EXTRACTION

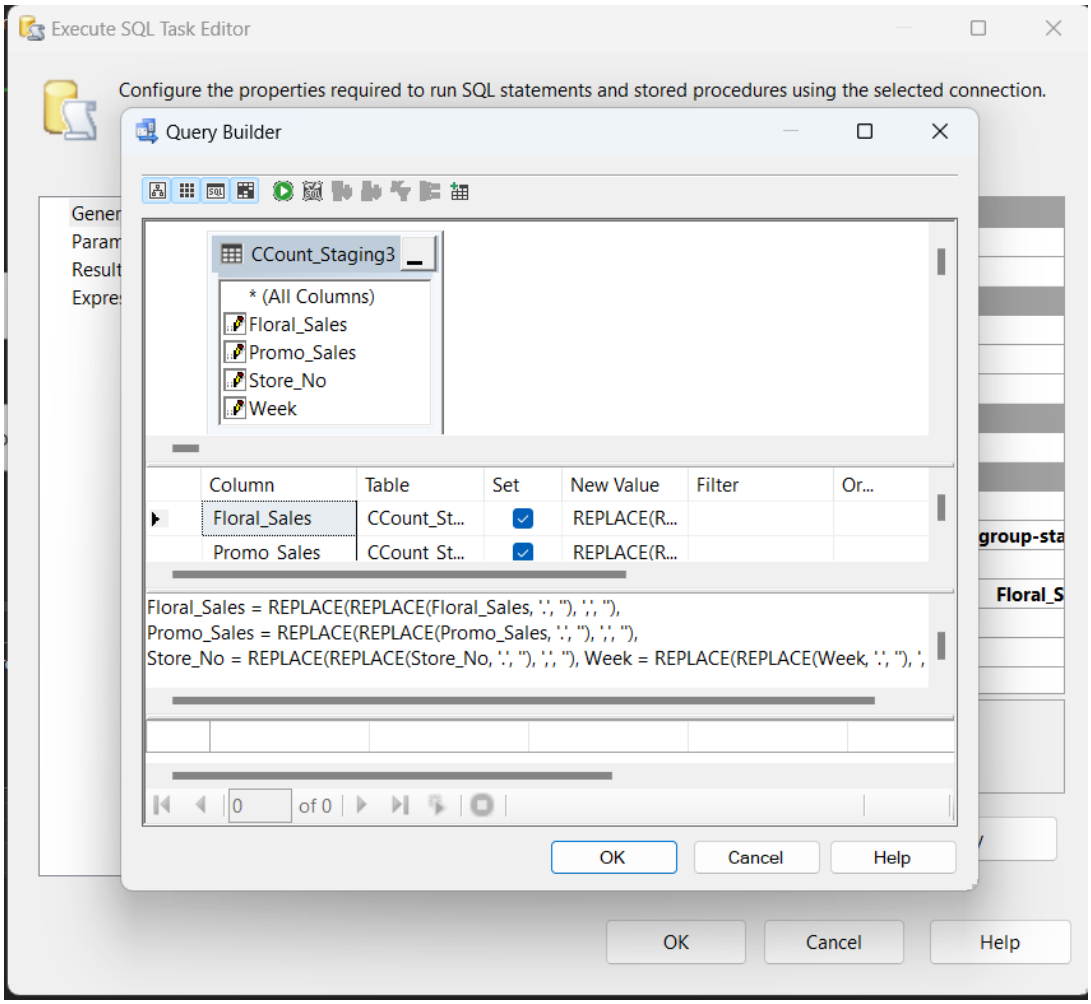
We used the CCOUNT.csv as our source to extract the Store Level Sales data. The columns that we extracted from this file were Store, Week, PromoSales and FloralSales.



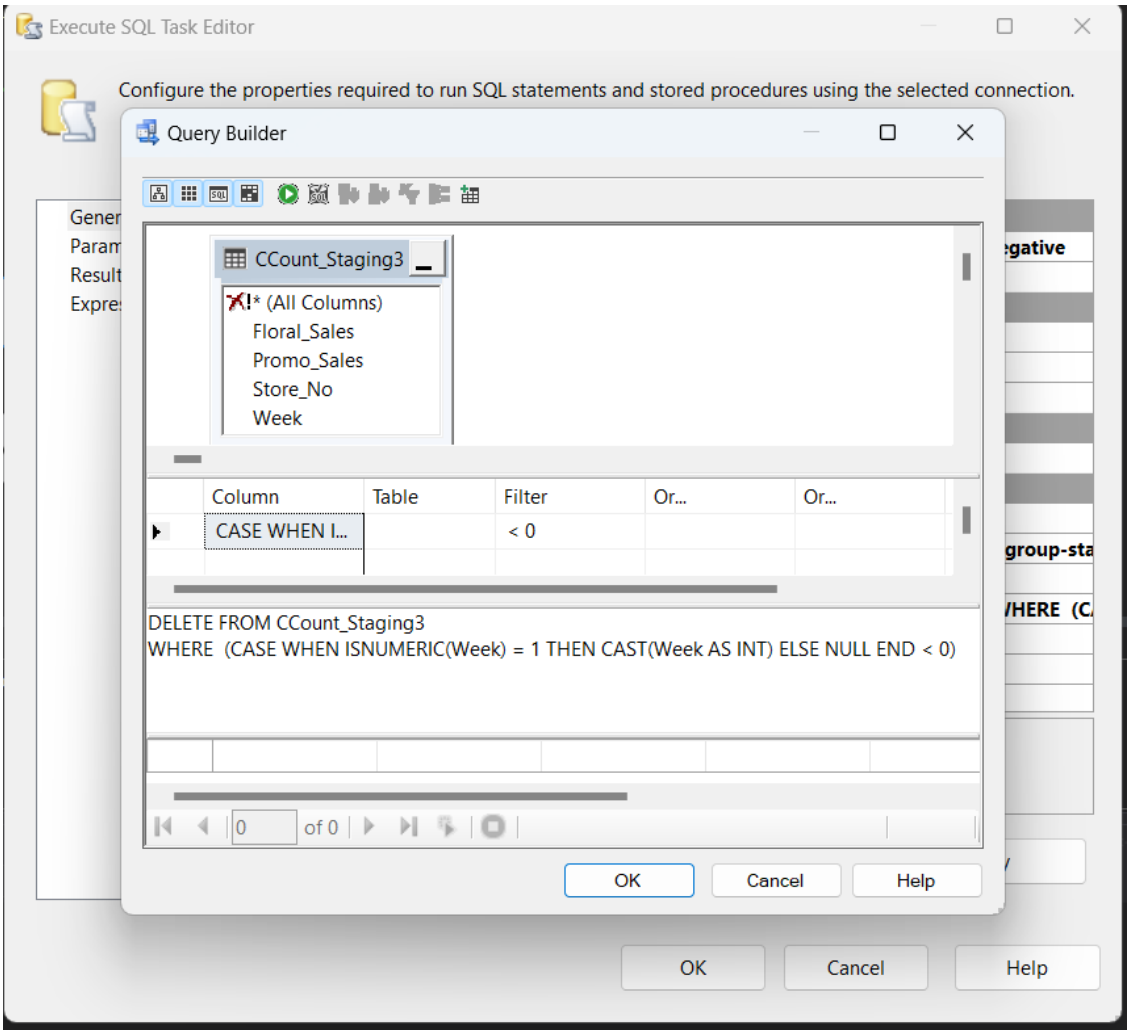
TRANSFORMATION

A lot of transformations were performed such as removing rows where week is negative and removing rows where the store number was invalid. Some example scripts are:

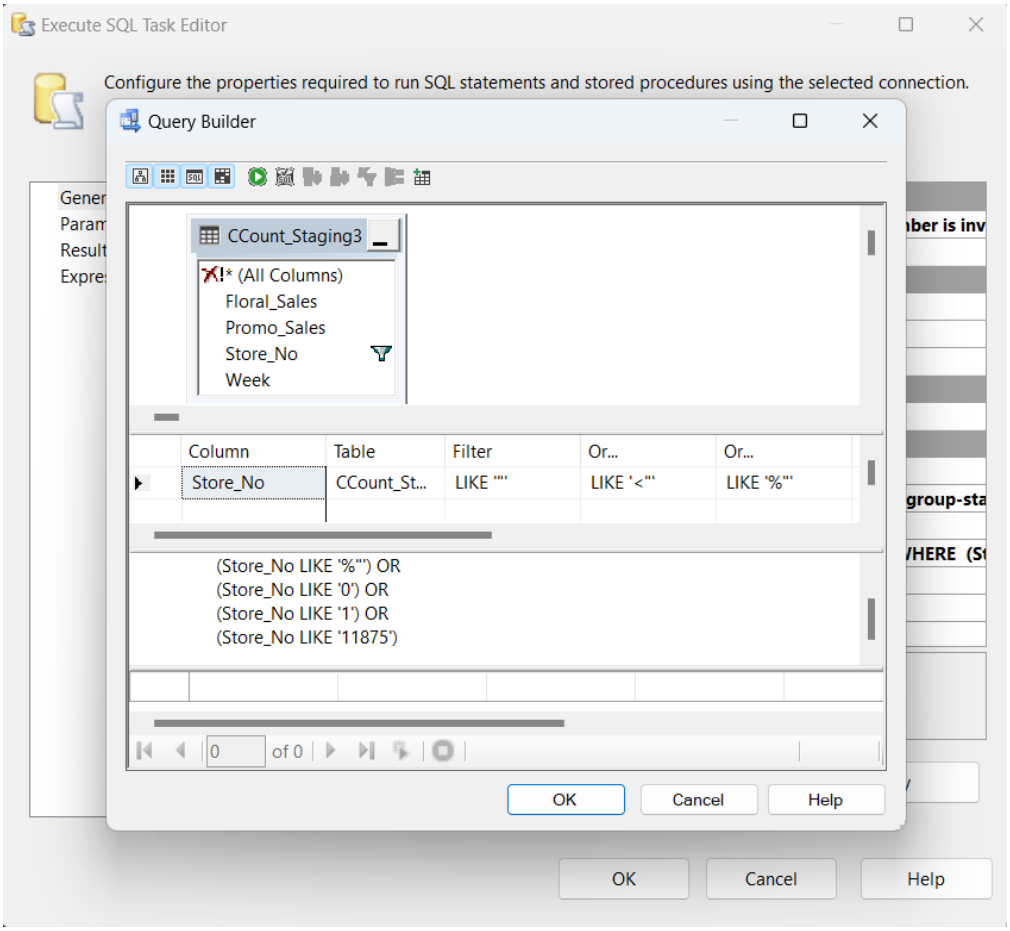
```
UPDATE CCount_Staging
SET
Floral_Sales = REPLACE(REPLACE(Floral_Sales, '.', ''), ',', ''),
Promo_Sales = REPLACE(REPLACE(Promo_Sales, '.', ''), ',', ''),
Store_No = REPLACE(REPLACE(Store_No, '.', ''), ',', ''), Week = REPLACE(REPLACE(Week, '.', ''), ',', '')
```



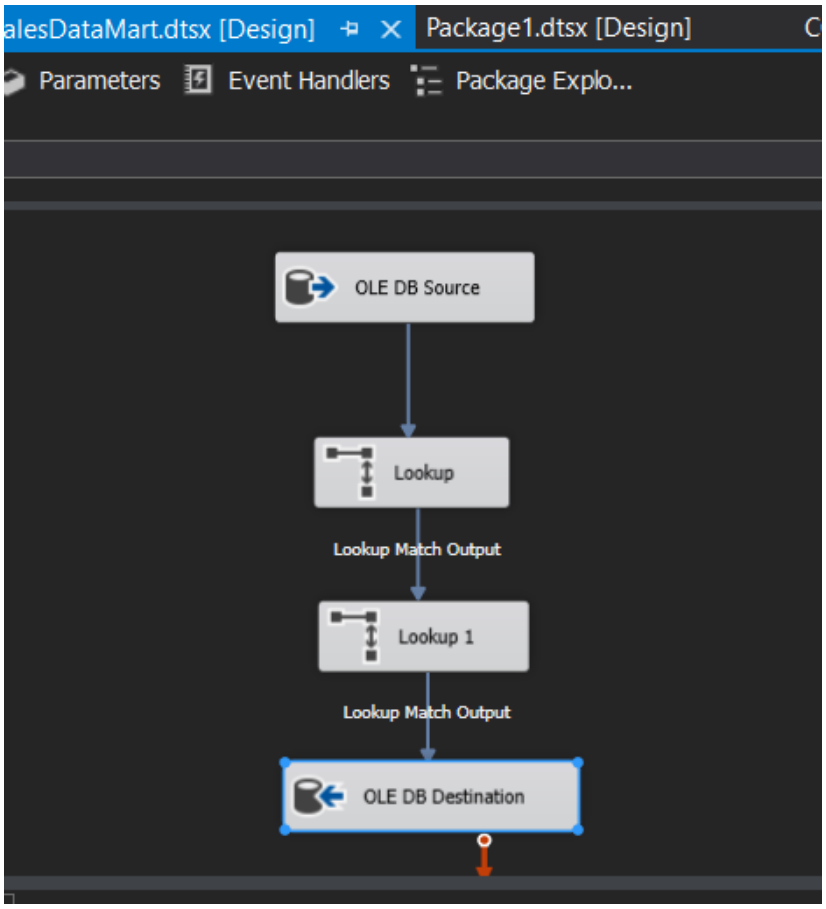

```
DELETE FROM CCount_Staging
WHERE  (CASE WHEN ISNUMERIC(Week) = 1 THEN CAST(Week AS INT) ELSE NULL
END < 0)
```




```
DELETE FROM CCount_Staging
WHERE  (Store_No LIKE "") OR
            (Store_No LIKE '<"') OR
            (Store_No LIKE '%"') OR
            (Store_No LIKE '0') OR
            (Store_No LIKE '1') OR
            (Store_No LIKE '11875')
```



Later on, after the initial cleaning was done, a second level of transformation was performed by using the Lookup tool in SSIS to bring the StoreID and TimeID columns in the FactStore table.



LOADING

The final data was loaded into the FactStore table in the Data Mart along with the database generated primary key FactStoreVisitID.

The following SQL Query was used to create the FactStore table:

```
CREATE TABLE [FactStore] (
   [FactStoreVisitID] INT IDENTITY(1,1) PRIMARY KEY,
   [Floral_Sales] NUMERIC(18,2),
   [Promo_Sales] NUMERIC(18,2),
   [StoreID] INT,
   [TimeID] INT,
   FOREIGN KEY ([StoreID]) REFERENCES DimStore([StoreID]),
   FOREIGN KEY ([TimeID]) REFERENCES DimTime([TimeID]));
```

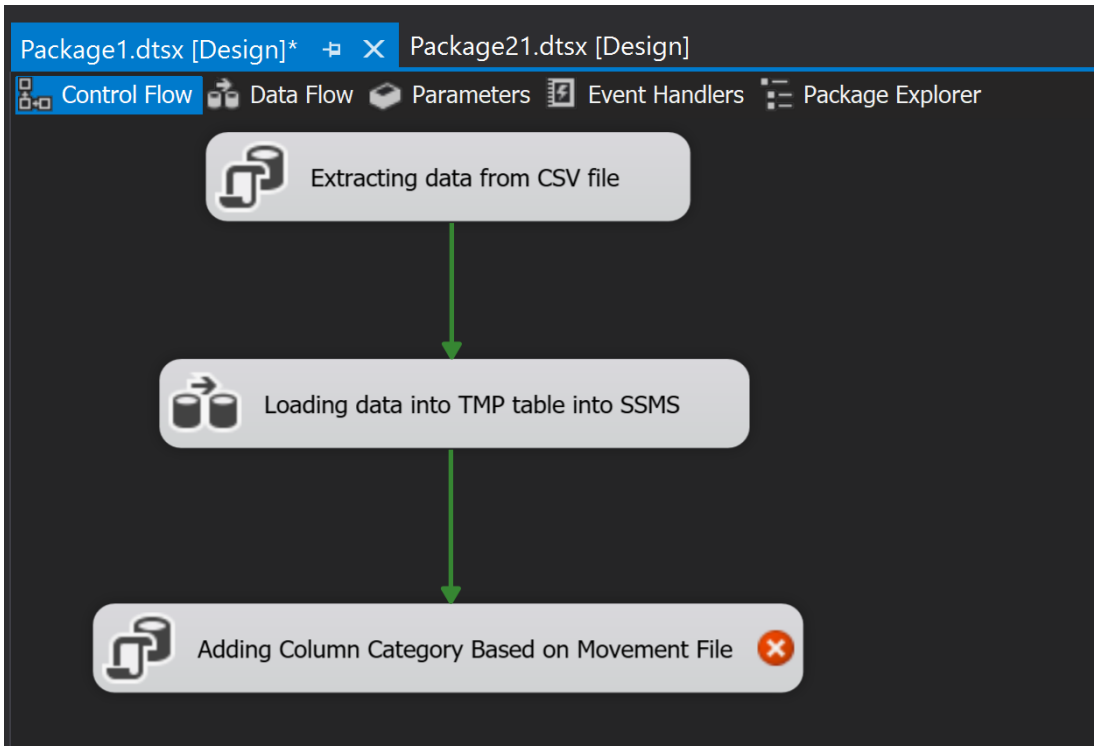
ISTM 637 - Group 5 Report

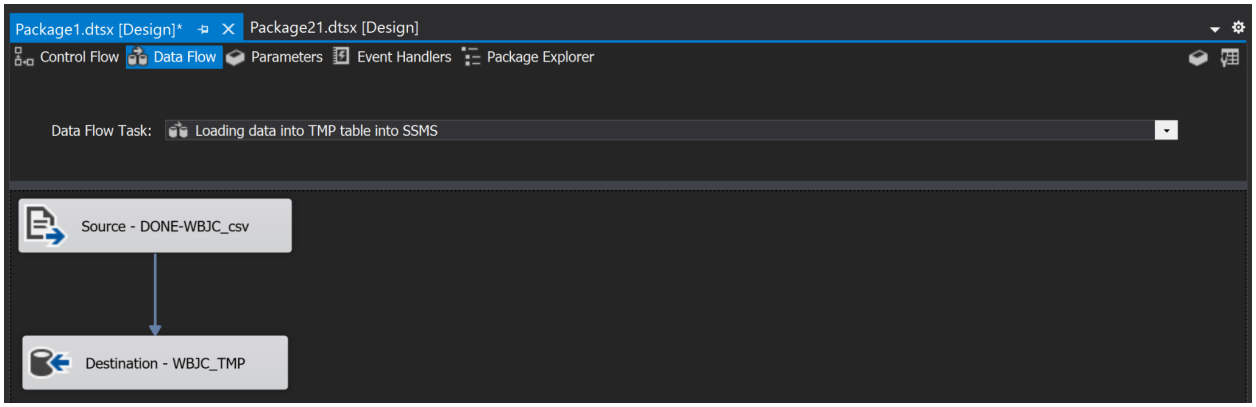
DATA MART 2: SALES DATA MART (SALE PER STORE, PER CATEGORY)

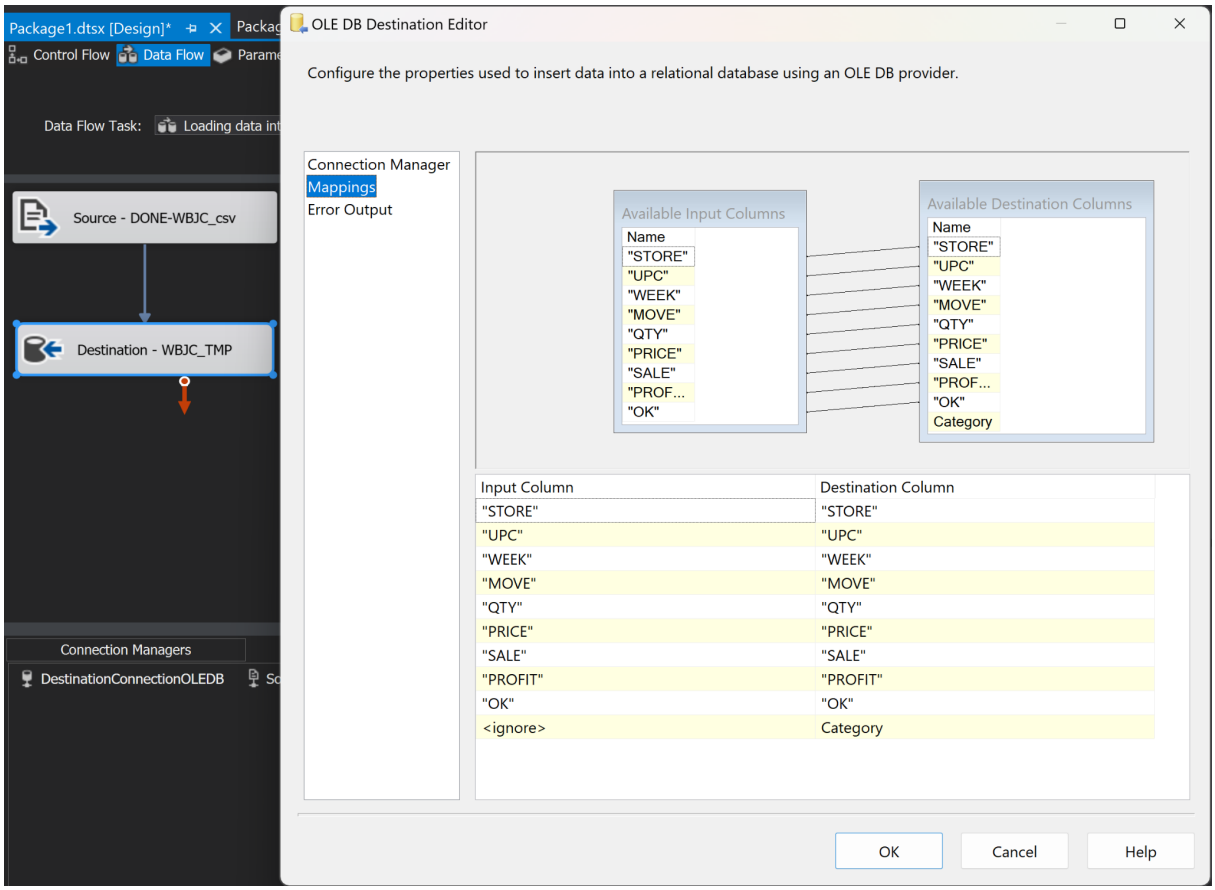
FINAL_MOVEMENT_STAGING

EXTRACTION

Firstly we imported data from each of the 20 distinct movement files named: WXXX.csv
For eg: WBJC.csv, WCIG.csv. We created a new import export package and correctly assigned data source and data destination.





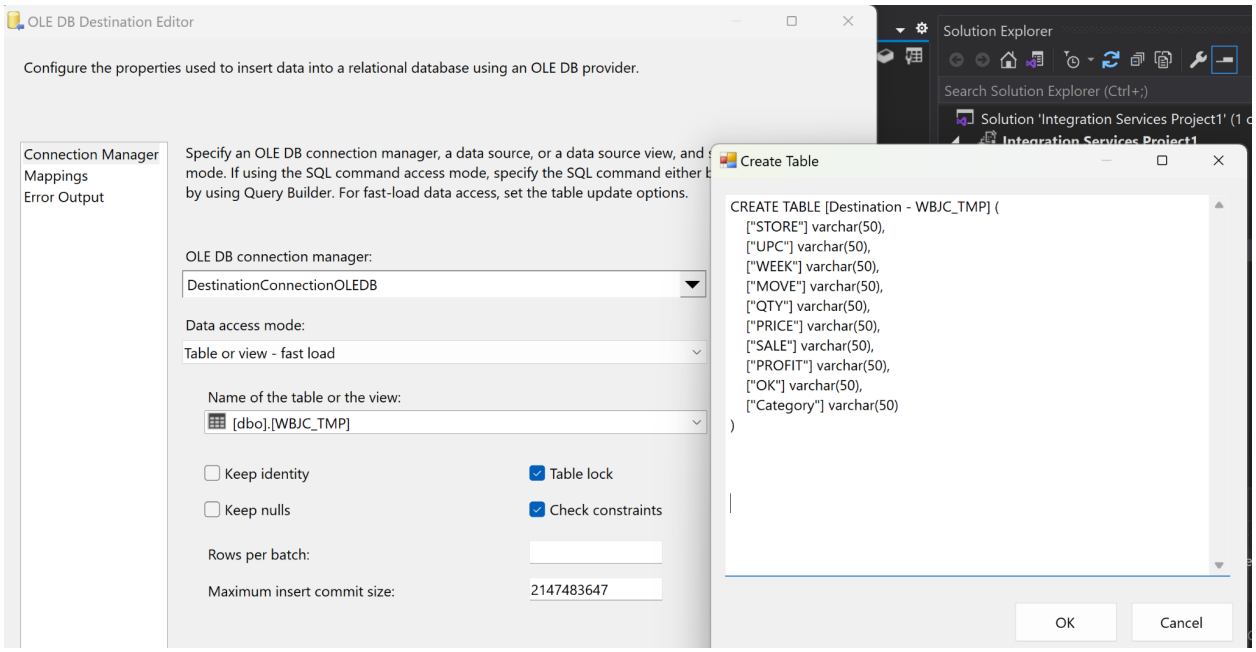


TRANSFORMATION

Once the data was extracted, we performed a Union ALL Operation to combine all 20 files since they hold the same type of data but for different categories.

We also performed a few conversions and data validation checks on it. Like:

- Adding Category Column for each Movement file



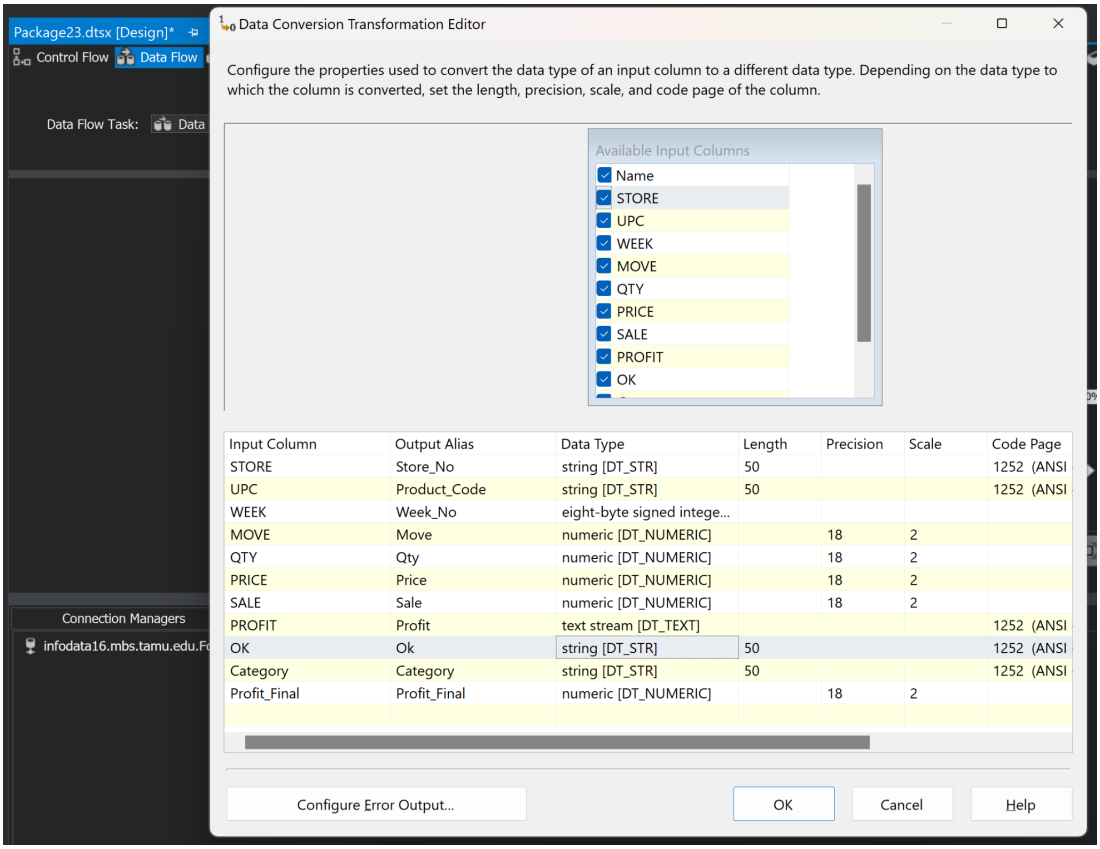
- Replacing empty strings with null
  UPDATE [dbo].[FINAL_MOVEMENT_STAGING]
  SET [QTY] = NULLIF(QTY, '');
  Similarly we handled all empty strings and converted them to NULLs

- Deleted all rows with ok = 0 because it was all trash data
  DELETE FROM [dbo].[FINAL_MOVEMENT_STAGING]
  WHERE [OK] = '0';

- Replaced everything in sale column with NULL
  UPDATE [dbo].[FINAL_MOVEMENT_STAGING]
  SET [SALE] = NULL;

- Deleting rows with bad data such as inappropriate characters
  - UPDATE [Focus-group-staging_area].[dbo].[FINAL_MOVEMENT_STAGING]
    SET
    [QTY] = CASE WHEN [QTY] = '.' THEN NULL ELSE [QTY] END
  - delete from [dbo].[FINAL_MOVEMENT_STAGING]
    where [Ok] not in ('1','0')
  - We performed several such transformations to eliminate and manage bad data

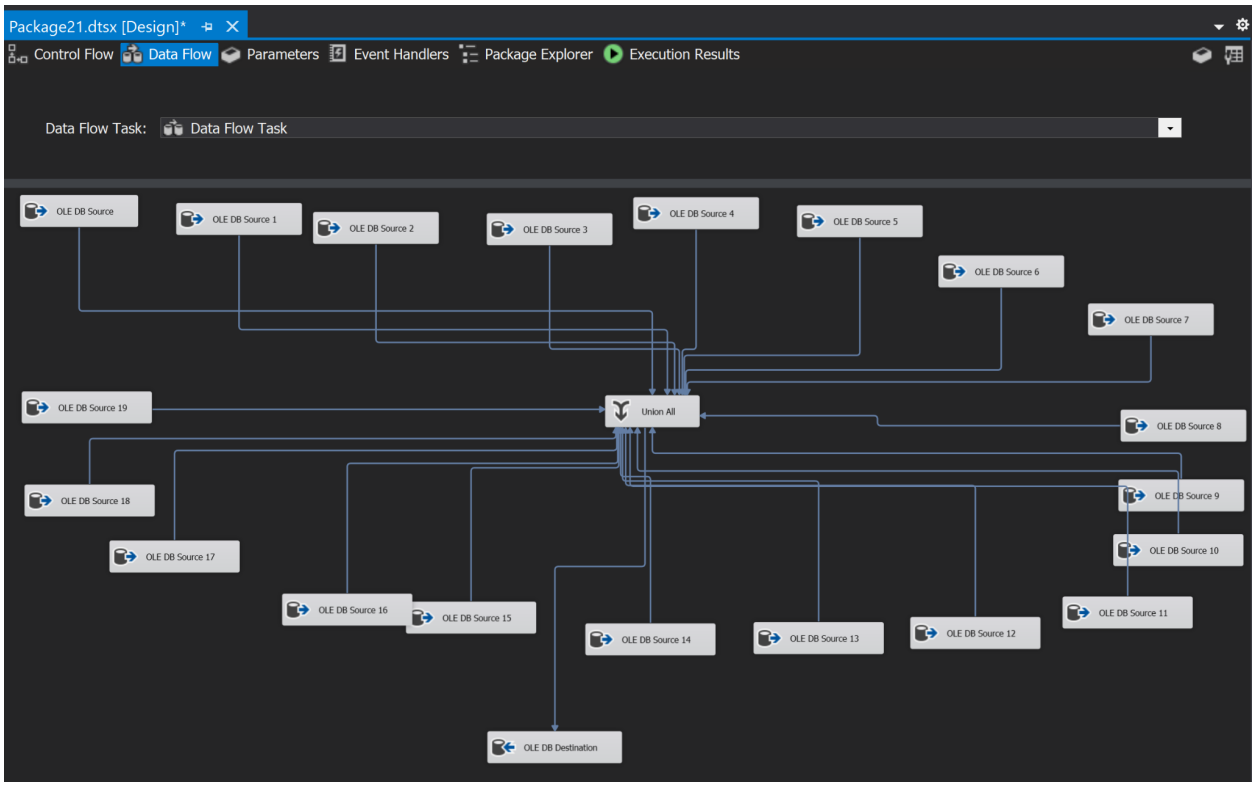
- Data type Conversion

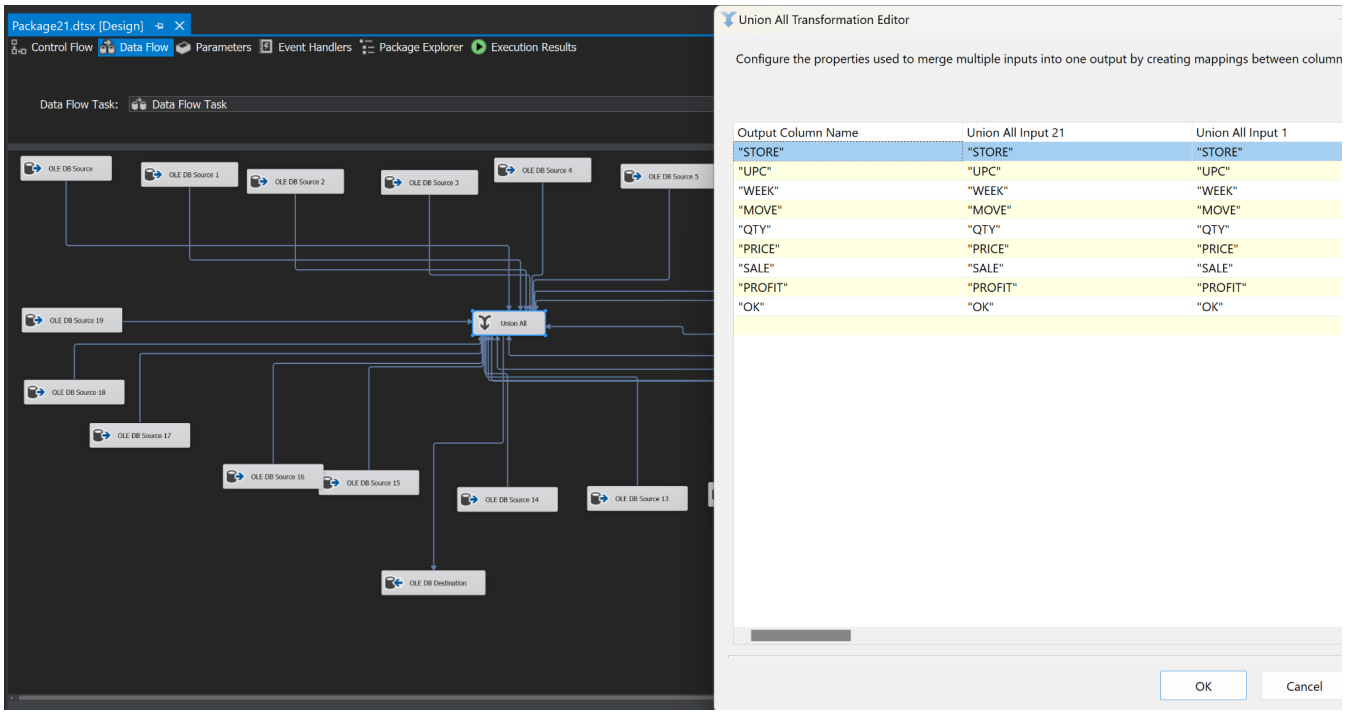


- Calculated sale based on the other 3 quantities in batches because the data size was too large

```
DECLARE @BatchSize INT = 10000;  -- Adjust the batch size as needed
DECLARE @RowsAffected INT = 1;

WHILE @RowsAffected > 0
BEGIN
   UPDATE TOP (@BatchSize) [dbo].[FINAL_MOVEMENT_STAGING]
   SET [SALE] =
      CASE
         WHEN TRY_CAST([QTY] AS DECIMAL(18, 2)) <> 0
```

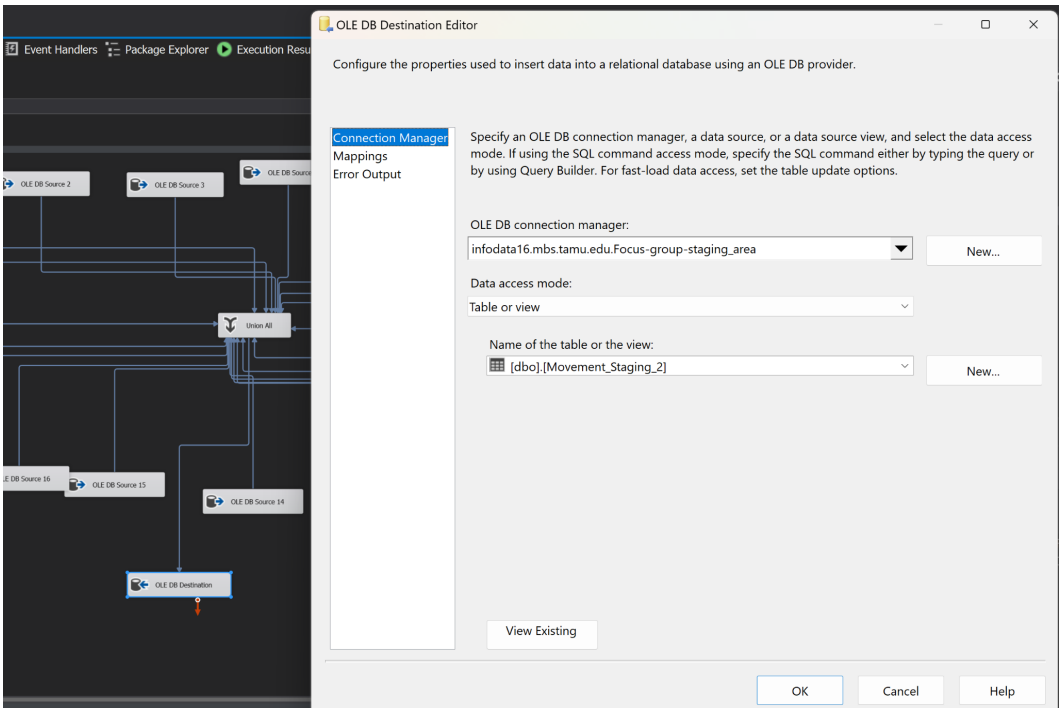
```
        THEN (CAST([PRICE] AS DECIMAL(18, 2)) * CAST([MOVE] AS
DECIMAL(18, 2))) / TRY_CAST([QTY] AS DECIMAL(18, 2))

        ELSE NULL -- or another default value
    END;

  SET @RowsAffected = @@ROWCOUNT;
END;
```





LOADING

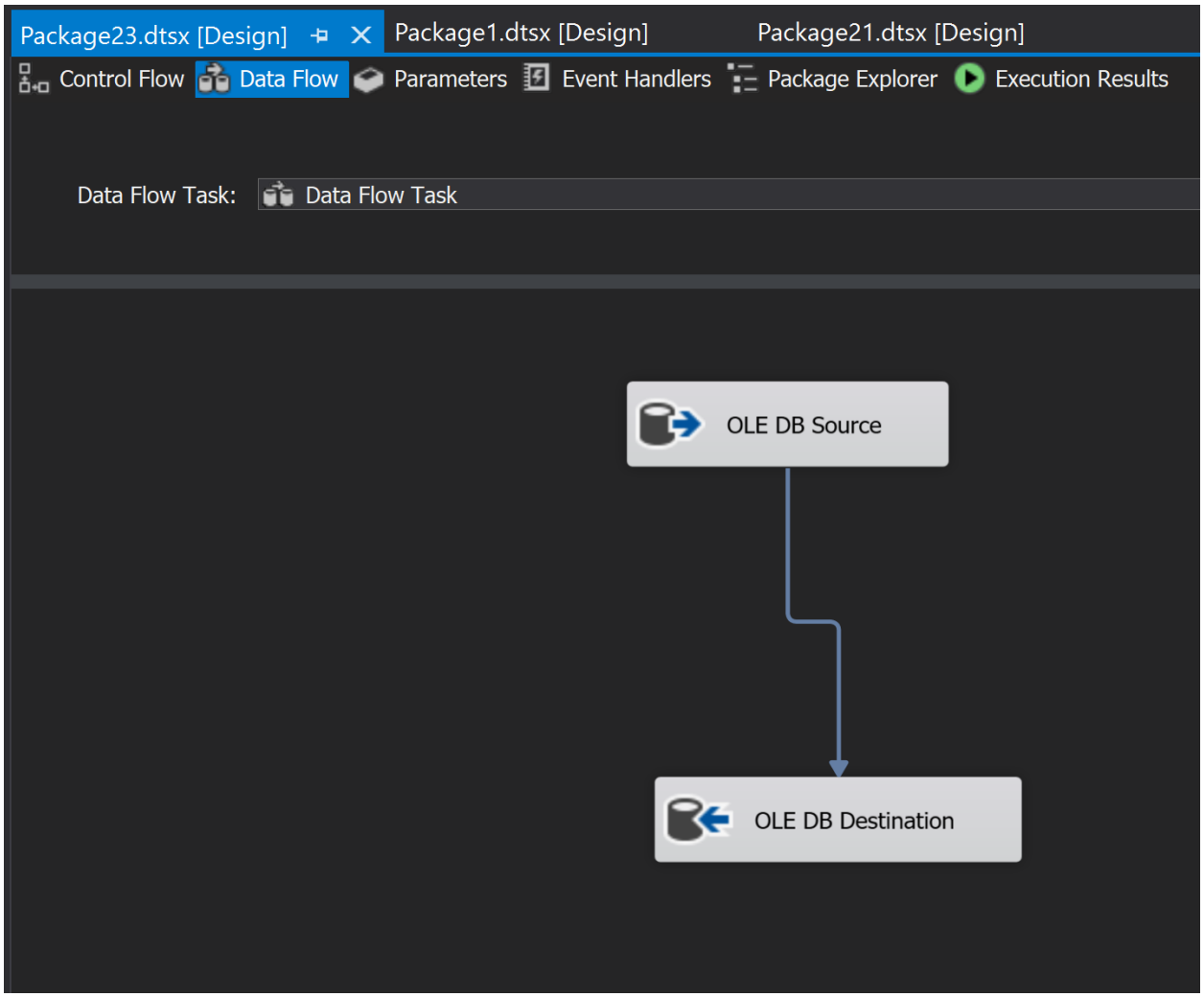
After performing all necessary transformations, We finally Loaded that data into our staging table.



dimProduct

EXTRACTION

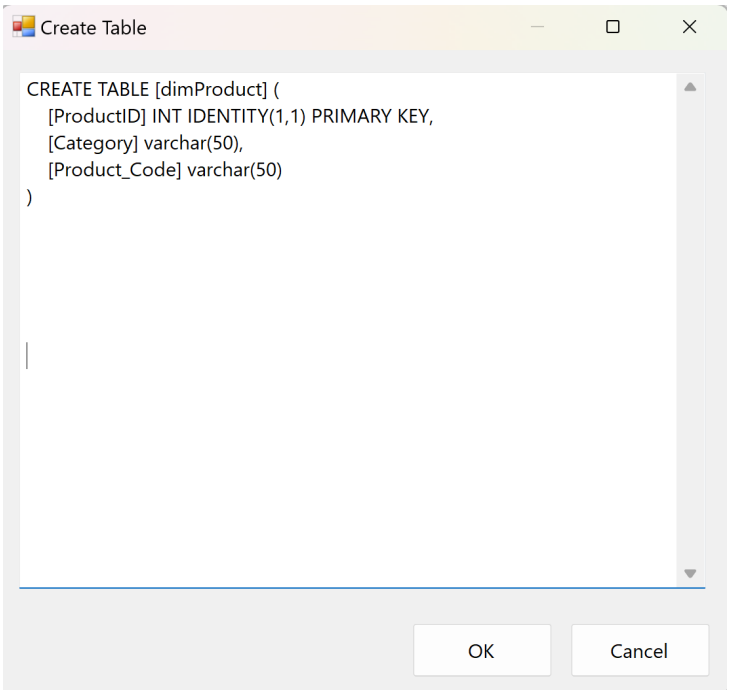
Once all transformations are completed in the FINAL_MOVEMENT_STAGING table, we extract data out of it and further load it into dimProduct.

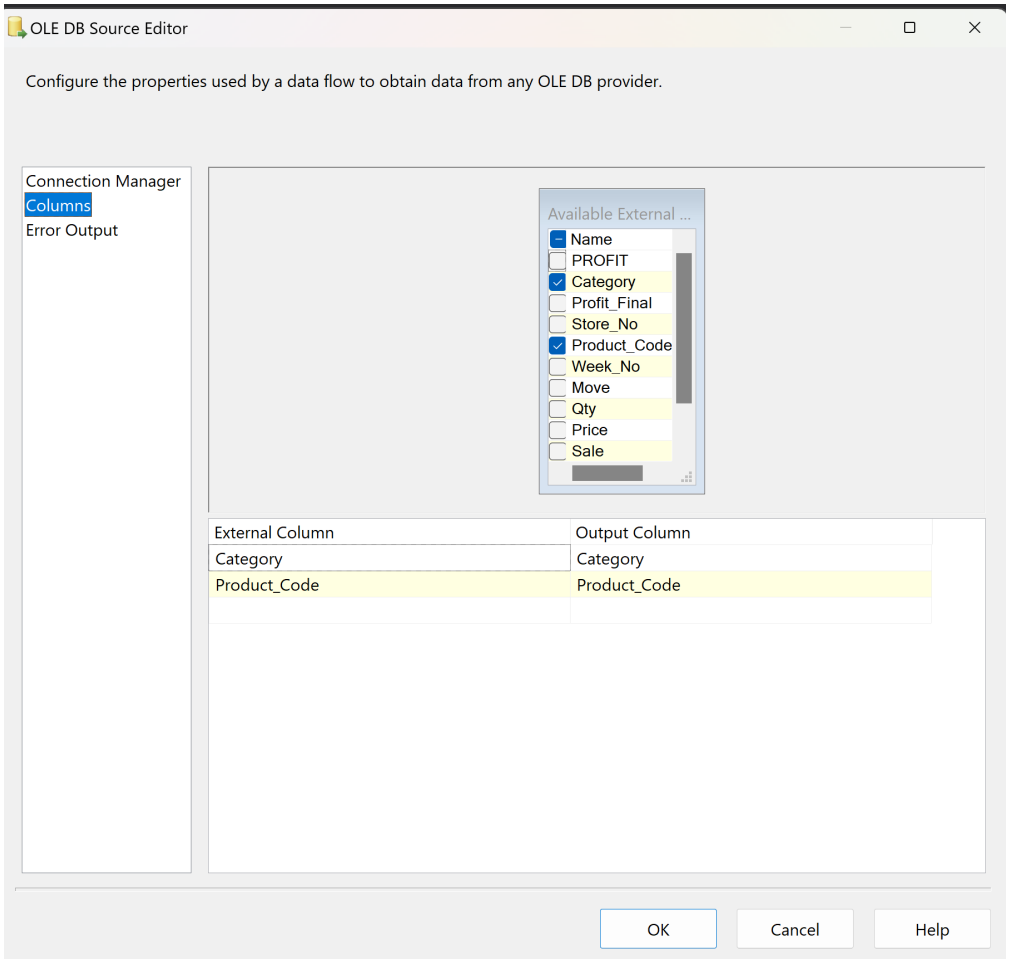


TRANSFORMATION

Here we map the relevant columns and also create a surrogate key ProductID. We will only select and map those columns that we need for dimProduct.

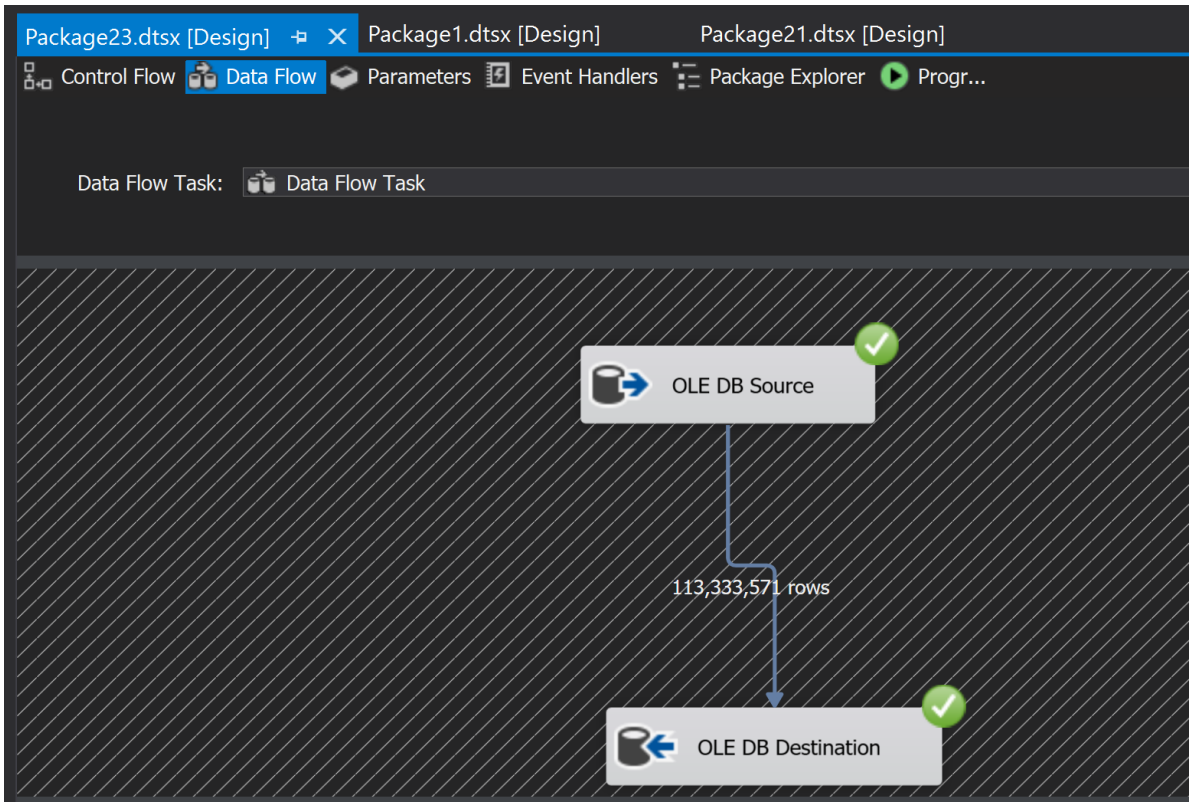
ISTM 637 - Group 5 Report





LOADING

Finally we load the selected data into dimProduct and we have our dimension table ready with required fields and values.



- dbo.dimProduct
  - Columns
    - ProductID (PK, int, not null)
    - Category (varchar(50), null)
    - Product_Code (varchar(50), null)
  - Keys
  - Constraints
  - Triggers
  - Indexes

ISTM 637 - Group 5 Report

90 %

Results Messages

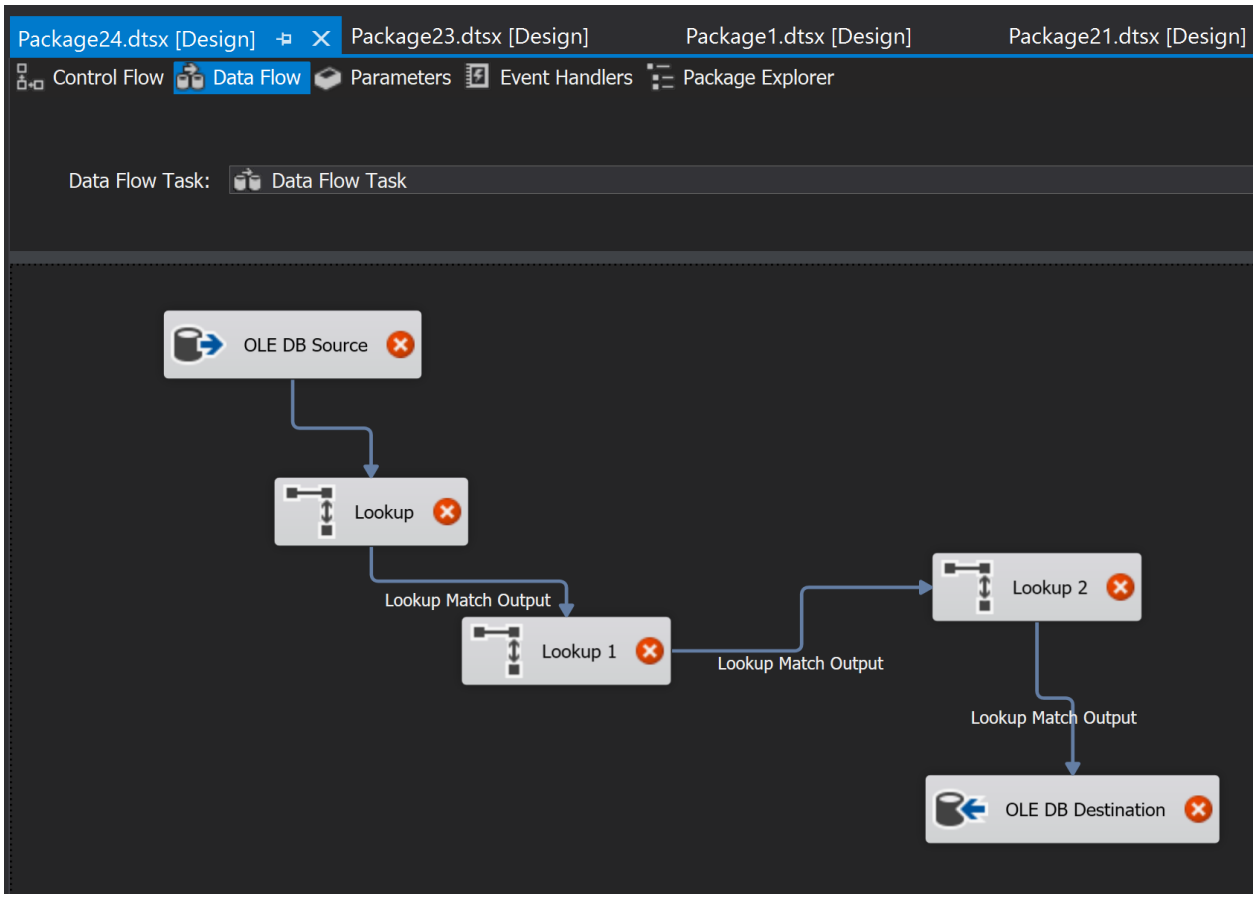
	ProductID	Category	Product_Code
1	1	Cigarettes	1230011036
2	2	Cigarettes	1230011036
3	3	Cigarettes	1230011036
4	4	Cigarettes	1230011036
5	5	Cigarettes	1230011036
6	6	Cigarettes	1230011036
7	7	Cigarettes	1230011036
8	8	Cigarettes	1230011036
9	9	Cigarettes	1230011036
10	10	Cigarettes	1230011036
11	11	Cigarettes	1230011036
12	12	Cigarettes	1230011036
13	13	Cigarettes	1230011036
14	14	Cigarettes	1230011036
15	15	Cigarettes	1230011036
16	16	Cigarettes	1230011036
17	17	Cigarettes	1230011036
18	18	Cigarettes	1230011036
19	19	Cigarettes	1230011036
20	20	Cigarettes	1230011036
21	21	Cigarettes	1230011036
22	22	Cigarettes	1230011036
23	23	Cigarettes	1230011036
24	24	Cigarettes	1230011036
25	25	Cigarettes	1230011036
26	26	Cigarettes	1230011036
27	27	Cigarettes	1230011036
28	28	Cigarettes	1230011036
29	29	Cigarettes	1230011036
30	30	Cigarettes	1230011036
31	31	Cigarettes	1230011036
32	32	Cigarettes	1230011036
33	33	Cigarettes	1230011036
34	34	Cigarettes	1230011036

Query executed successfully. infodata16.mbs.tamu.edu (13...

factSales

EXTRACTION

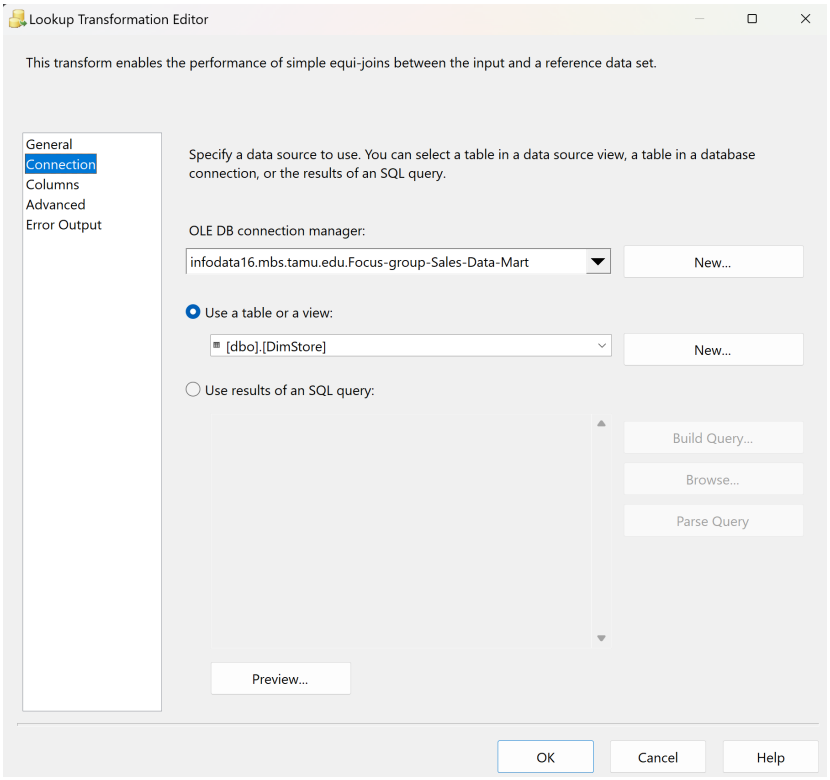
We will take the FINAL_MOVEMENT_STAGING as the source table since it contains most of our Fact table quantities.

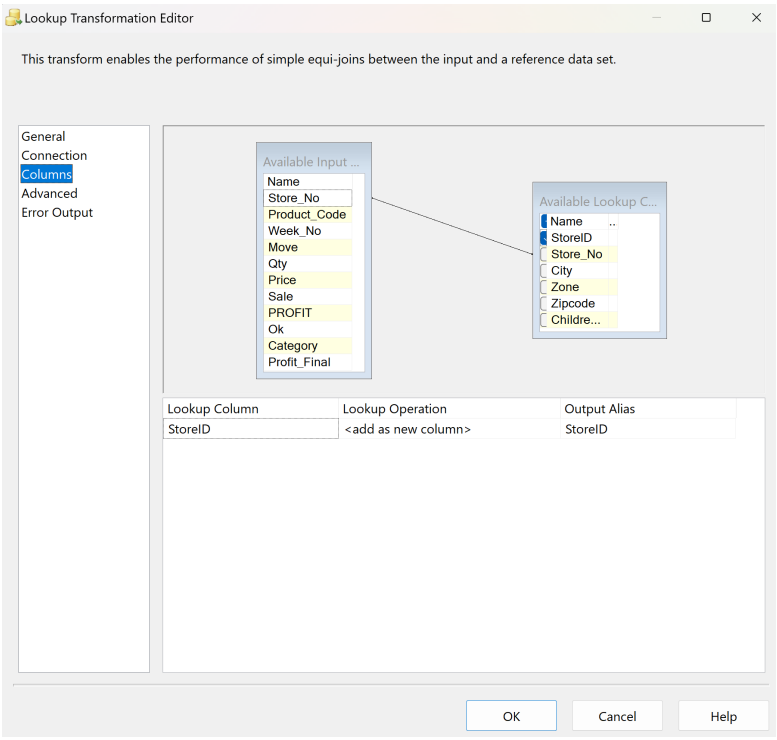


TRANSFORMATION

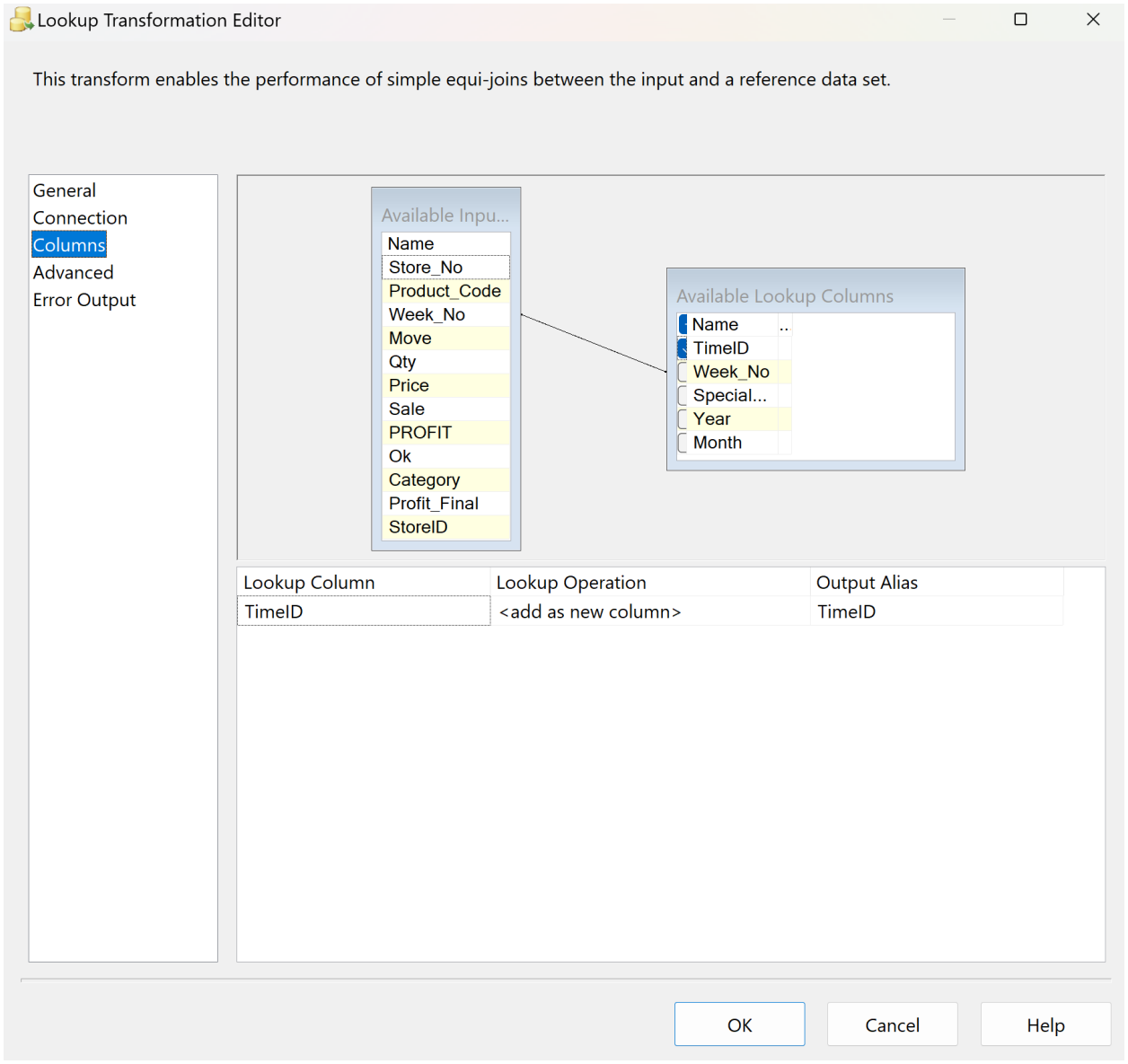
Since the fact table contains several parent-child relationships, we have used lookups to join data from all the dim tables with the corresponding key.

Lookup (Source - dimStore):

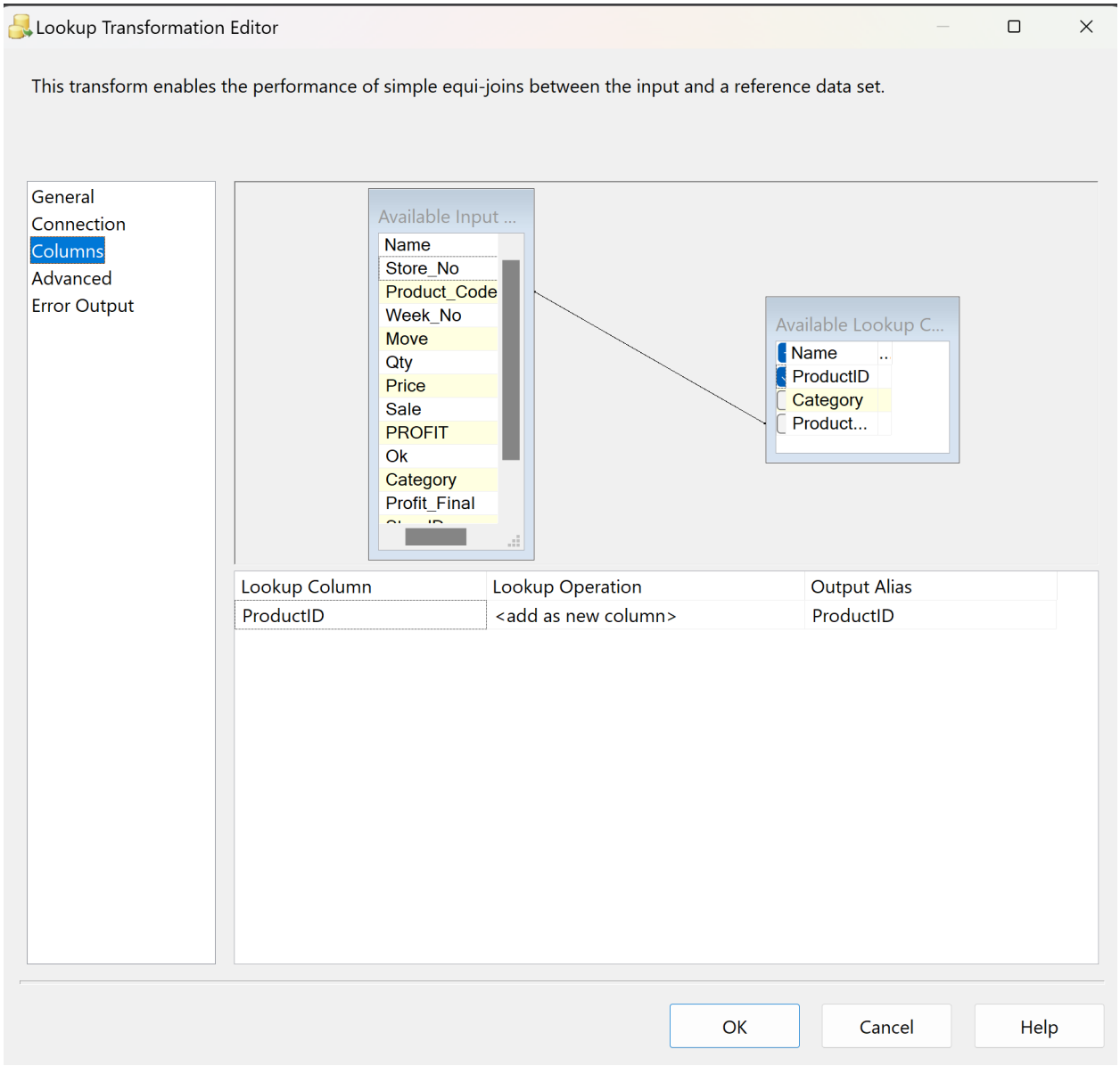




Lookup1 (Source - dimTime):

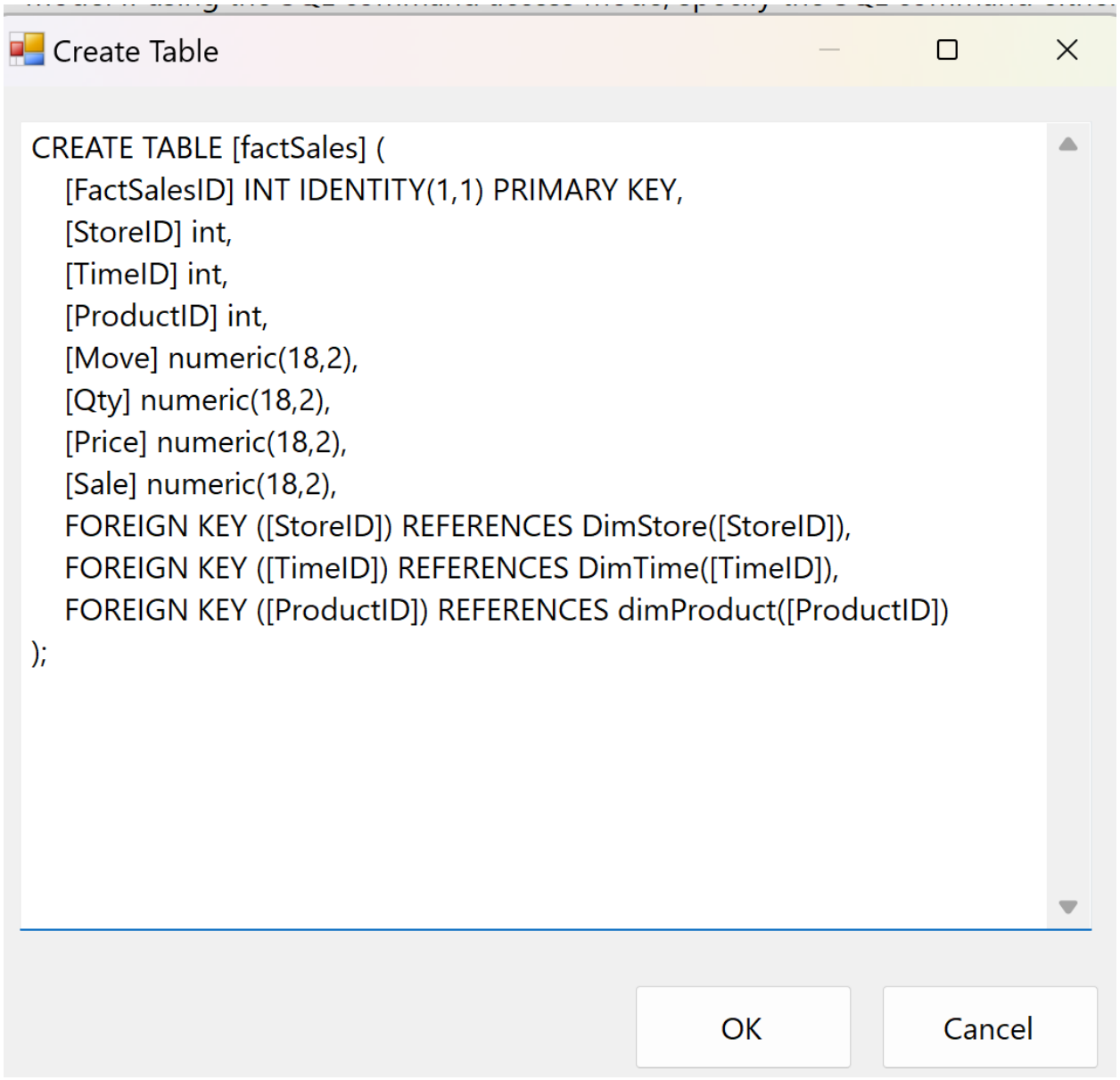


Lookup2 (Source - dimProduct):


Lookup Transformation Editor
This transform enables the performance of simple equi-joins between the input and a reference data set.
General
Connection
Columns
Advanced
Error Output
Available Input ...
Name
Store_No
Product_Code
Week_No
Move
Qty
Price
Sale
PROFIT
Ok
Category
Profit_Final
Available Lookup C...
Name
ProductID
Category
Product...
Lookup Column
Lookup Operation
Output Alias
ProductID
<add as new column>
ProductID
OK
Cancel
Help

LOADING

Once we have joined all the data and have made the correct mappings, we finally load that into a new table named: factSales with the desired Primary foreign key pairings and surrogate key generation.



DATA GRANULARITY

Tables in Data Mart	**Granularity**
dimStore	Store Level
dimTime	Week Level
dimProduct	Product Level
factStore	Store and Week
factSales	Week and Product

LIST OF TEMPORARY TABLES IN STAGING AREA

Temporary Table	**Description**
CCOUNT_TMP	Used to store raw CCOUNT data
DEMO_TMP	Used to store raw demographic data
store_tmp	Used to load raw store data
TIME_TMP	Used to load raw time data
WBJC_TMP	Used to load Bottled Juice Movement Data
WCHE_TMP	Used to load Cheese Movement Data
WCIG_TMP	Used to load Cigarettes Movement Data
WCOO_TMP	Used to load Cookies Movement Data
WCRA_TMP	Used to load Crackers Movement Data
WCSO_TMP	Used to load Canned Soup Movement Data
WDID_TMP	Used to load Dish Detergent Movement Data

WFEC_TMP	Used to load Front-End-Candies Movement Data
WFRD_TMP	Used to load Frozen Dinners Movement Data
WFRE_TMP	Used to load Frozen Entrees Movement Data
WFRJ_TMP	Used to load Frozen Juices Movement Data
WFSF_TMP	Used to load Fabric Softener Movement Data
WGRO_TMP	Used to load Grooming Products Movement Data
WLND_TMP	Used to load Laundry Detergents Movement Data
WSDR_TMP	Used to load Soft Drinks Movement Data
WSOA_TMP	Used to load Soap Movement Data
WTBR_TMP	Used to load Toothbrushes Movement Data
WTNA_TMP	Used to load Canned Tuna Movement Data
WTPA_TMP	Used to load Toothpastes Movement Data
WTTI_TMP	Used to load Toilet Papers Movement Data

SECTION 5: BUSINESS INTELLIGENCE REPORTING

REPORTING PLAN

In this section of our project, we leveraged various business intelligence tools to build the reports. These reports were built using data from the data warehouse we built for DFF to answer the business questions and thus, help the users of the system in their decision making process.

Following tools were used for reporting:

1. **SSAS – Microsoft SQL Server Analysis Services**

   SQL Server Analysis Services is an analytical tool that is used in Online Analytical Processing. SSAS enables users to analyze and explore data in a structured manner, using multidimensional models.

2. **SSRS – Microsoft SQL Server Reporting Services**

   SQL Server Reporting Services is a Microsoft server-based reporting platform used for the building, management and delivery of interactive, tabular, graphical and spatial reports. SSRS is commonly used to generate dynamic reports and dashboards, facilitating data-driven decision making.

3. **Tableau**

   Tableau is a powerful data visualization and business intelligence tool, used to connect to various sources, visualize and understand their data. Since we are using SQL Server, Tableau can be used to connect to the database server and extract the data for reporting purposes.

The following questions were utilized to answer the business questions effectively using visualizations and dashboards:

Business Question Number	Question	Reporting Method Deployed
1	What are the top-selling Cigarette SKUs in the Cigarette Product Category for DFF?	Tableau
2	In which months of the year do maximum promotional sales take place?	Independent Data Mart using SSRS
3	What are the various possibilities to maximize earnings based on high sales categories?	Tableau
4	What is the correlation between candy sales and the number (percentage) of children in a given demographic area?	Cubes from SSAS
5	What is the impact of festive occasions and other significant events on monthly floral sales?	Cubes from SSRS + SSAS

MAPPING ATTRIBUTES FROM INDEPENDENT DATA MARTS

BUSINESS QUESTION 1: In which months of the year do maximum promotional sales take place?

Attribute Name	Attribute's Dimension/ Fact table	Filters	Corresponding Report Attribute
Time ID	dimTime/factStore	NA	NA
Promo_Sales	factStore	NA	Promotional Sales
Month	dimTime	NA	Month
Year	dimTime	NA	Year

BUSINESS QUESTION 2: What is the correlation between candy sales and the number (percentage) of children in a given demographic area?

Attribute Name	**Attribute's Dimension/ Fact table**	**Filters**	**Corresponding Report Attribute**
Candy Sales	factSales	Category = 'Front end sales'	Candy Sales
Children Count	dimStore	NA	Children Count

BUSINESS QUESTION 3: What is the impact of festive occasions and other significant events on monthly floral sales?

Attribute Name	**Attribute's Dimension/ Fact table**	**Filters**	**Corresponding Report Attribute**
Time ID	dimTime	NA	NA
Floral_Sales	factStore	NA	Floral Sales
Month	dimTime	NA	Month
Year	dimTime	NA	Year

BUSINESS QUESTION 4: What are the top-selling Cigarette SKUs in the Cigarette Product Category for DFF?

Attribute Name	**Attribute's Dimension/ Fact table**	**Filters**	**Corresponding Report Attribute**
Product_Code	dimProduct	Order by Sales Top	Product Code

		10	
Product ID	dimProduct/factSales	NA	NA
Category	dimProduct	‘Cigarettes’	Category
Store_No	dimStore	NA	Store Number
Sales	factSales	NA	Sales

BUSINESS QUESTION 5: What are the various possibilities to maximize earnings based on high sales categories?

Attribute Name	**Attribute’s Dimension/ Fact table**	**Filters**	**Corresponding Report Attribute**
Sales	factSales	NA	Sales
Product ID	dimProduct/factSales	NA	NA
Category	dimProduct	NA	Category

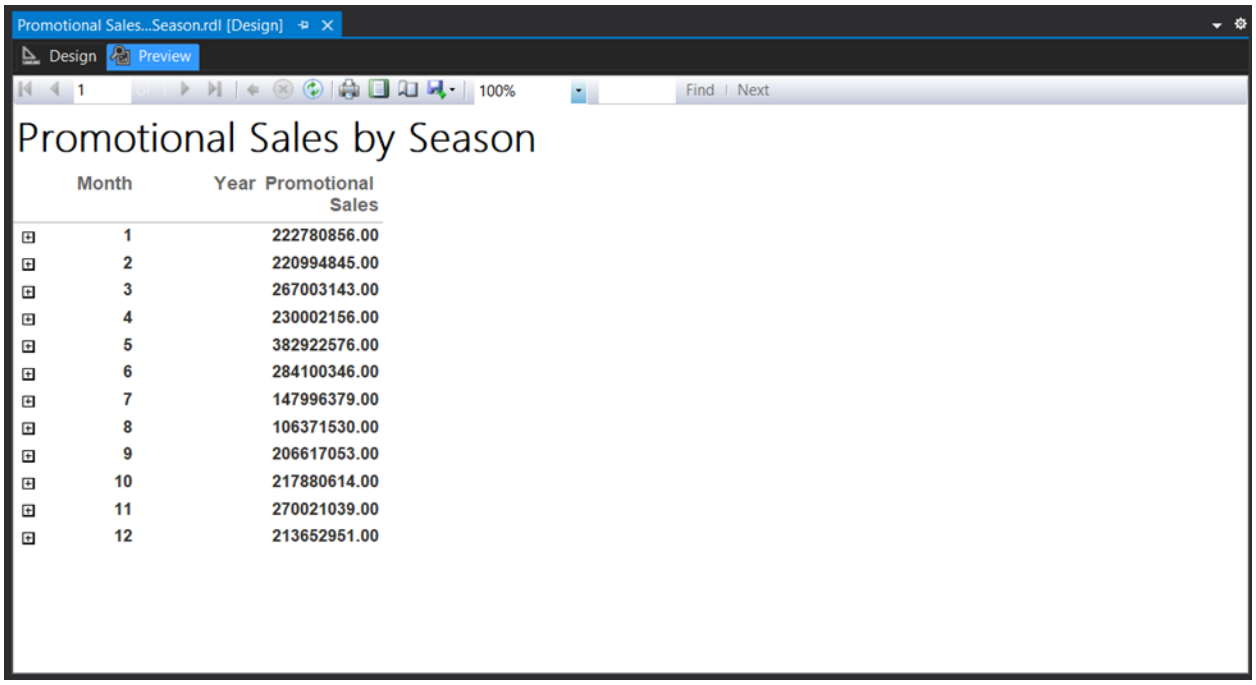
REPORT IMPLEMENTATION

Business Question 1

In which months of the year do maximum promotional sales take place?

BI/Visualization Method: Reporting from the independent Data Mart using SSRS.

Report:



Month	Year Promotional Sales
1	222780856.00
2	220994845.00
3	267003143.00
4	230002156.00
5	382922576.00
6	284100346.00
7	147996379.00
8	106371530.00
9	206617053.00
10	217880614.00
11	270021039.00
12	213652951.00

Rationale:

Typically, any commercial business would want to analyze their sales spread over a time-series to check for any spikes in sales and study the effects of seasonality on the purchasing habit of consumers. This question is based on the seasonality of promotional sales to find out months where there were maximum promotional sales taking place.

The value proposition of this business question lies in the company then strategizing when to dial up and dial down promotions to maximize profits using this information of the past 10 years. This can serve to be an effective strategy for coming up with new promotions in the future.

The beauty of this question lies in the simplicity of it as well as the amount of insight and actionable information it provides.

Reporting Plan:

The given business question was analyzed using the SSRS tabular method given the straightforward nature of the dimensions and measures involved in the question. There are two dimensions involved – month and year. Primarily, the user would want to view the information grouped by months and see a sum of promotional sales across the same months of different years.

However, there is also a drill down option in the report to see the corresponding years for a drive down analysis.

Implementation Steps:

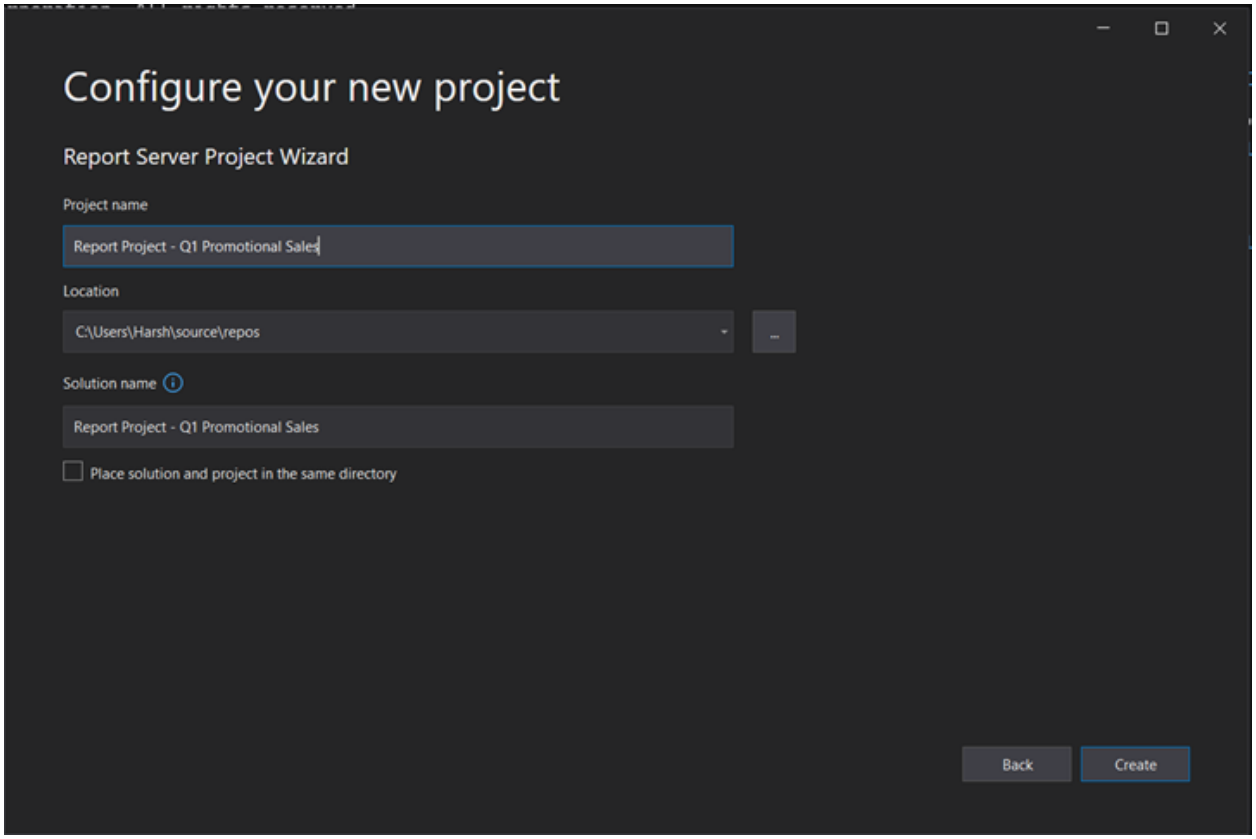


Figure: Configuring a new Reporting Server Project Wizard

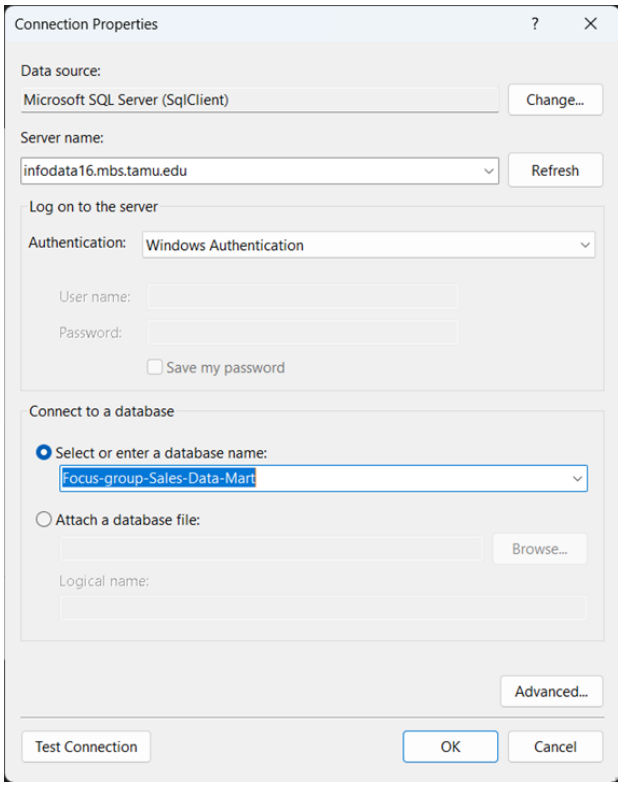


Figure: Configuring the connection properties to our server and database

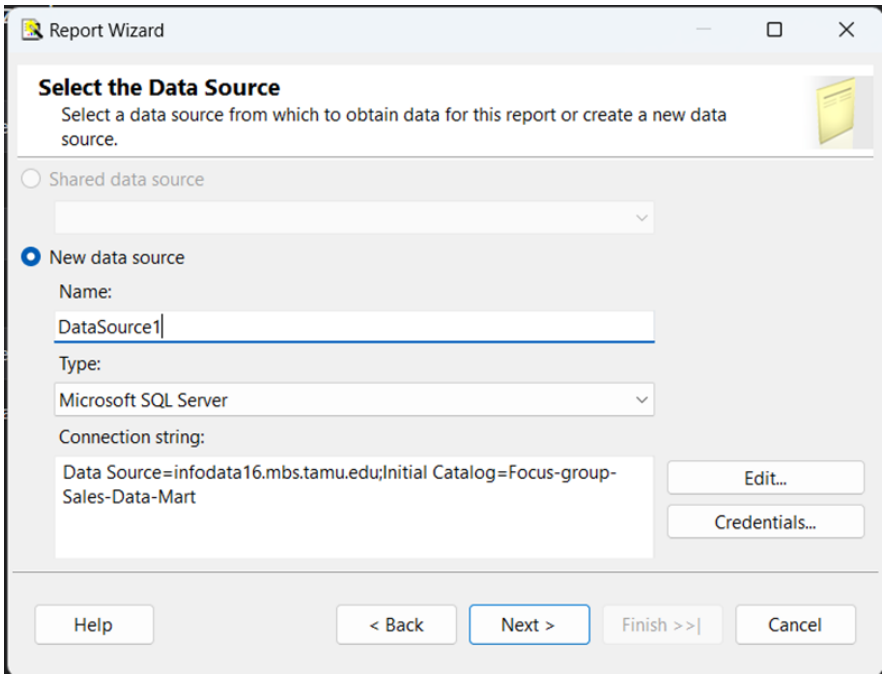


Figure: Finalizing the credentials in the report wizard

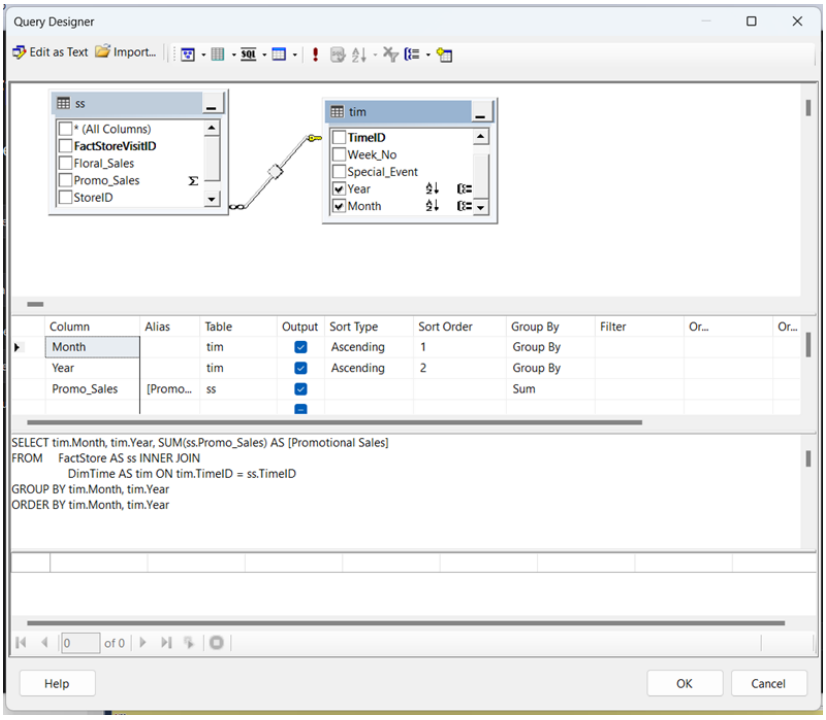


Figure: Designing the query to pull the required data from FactStore and DimTime. This involves aggregating promotional sales by month and year and joining the fact and dimension tables.

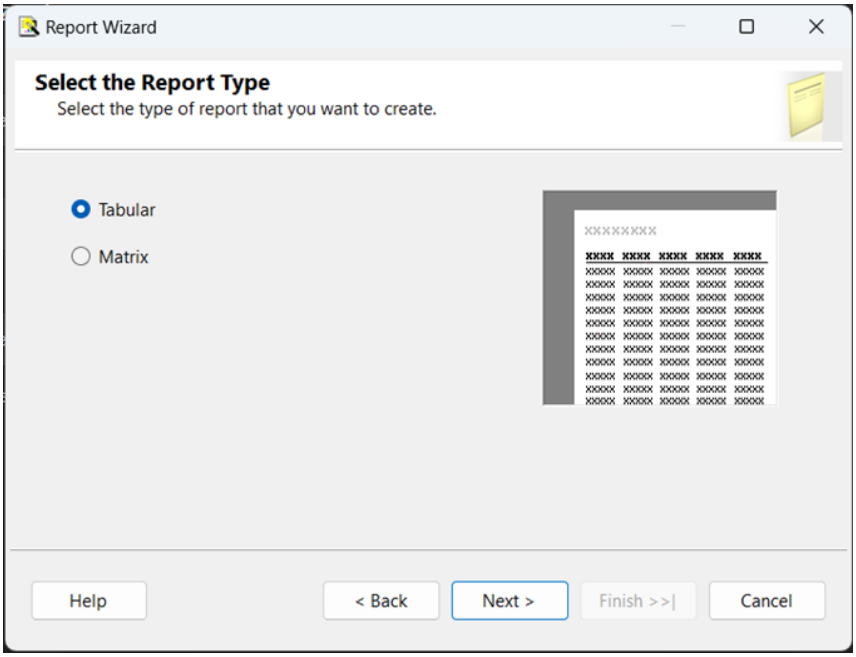


Figure: Choosing a tabular format to depict our report as that is the best way to show our data.

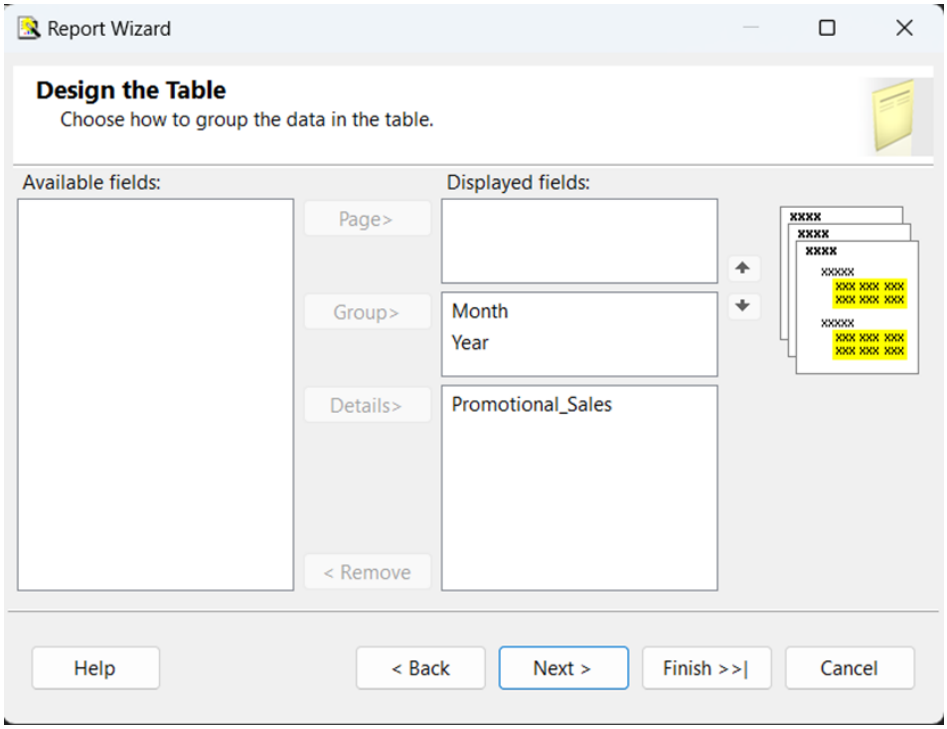


Figure: Choosing the grouping levels and details of our report.

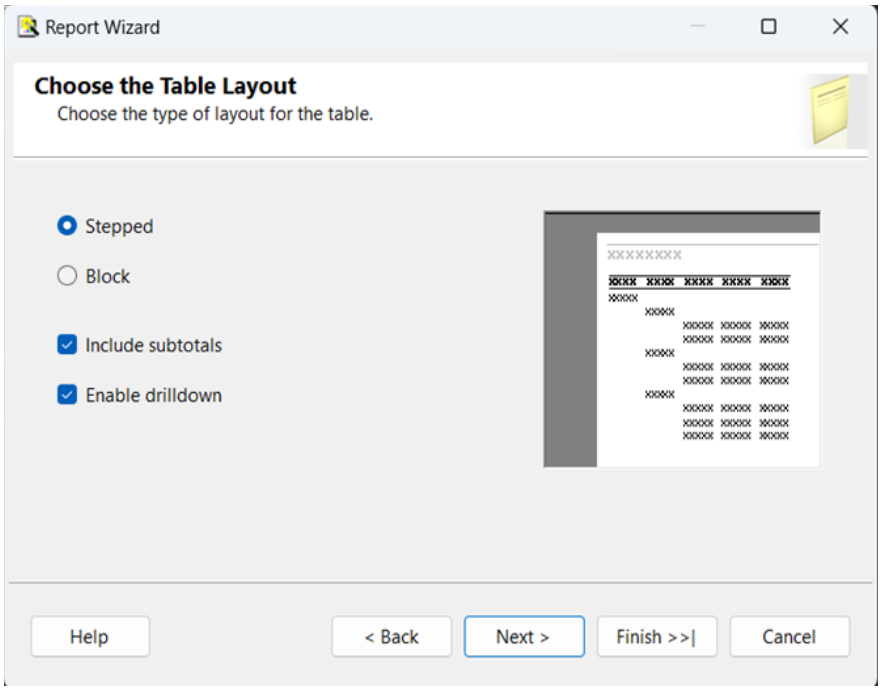


Figure: Choosing to include subtotals and enabling drilldown for deeper analysis.

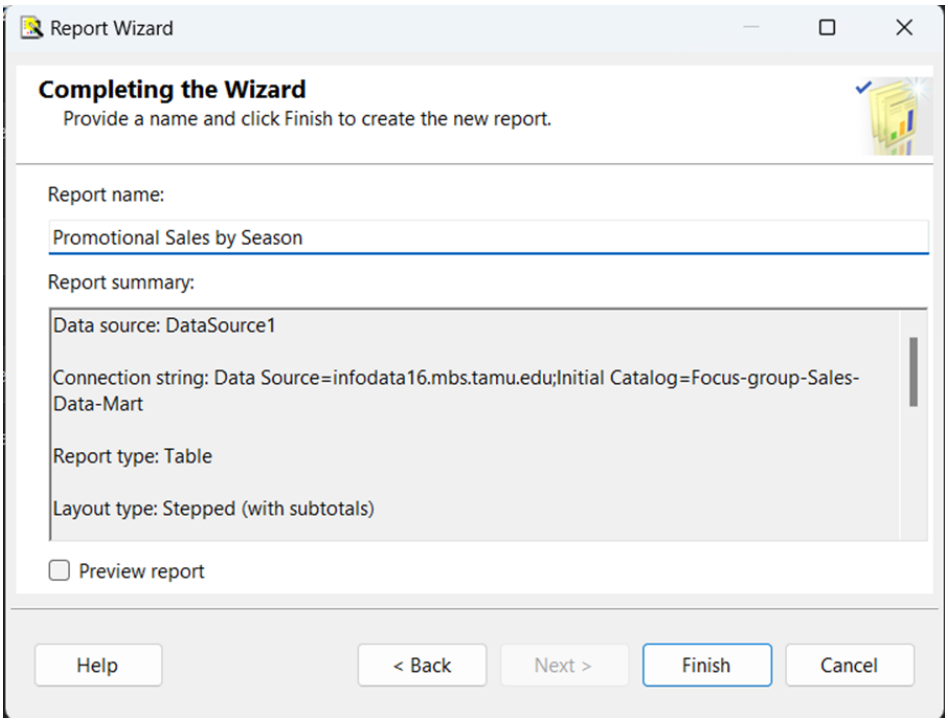


Figure: Finalizing the details.

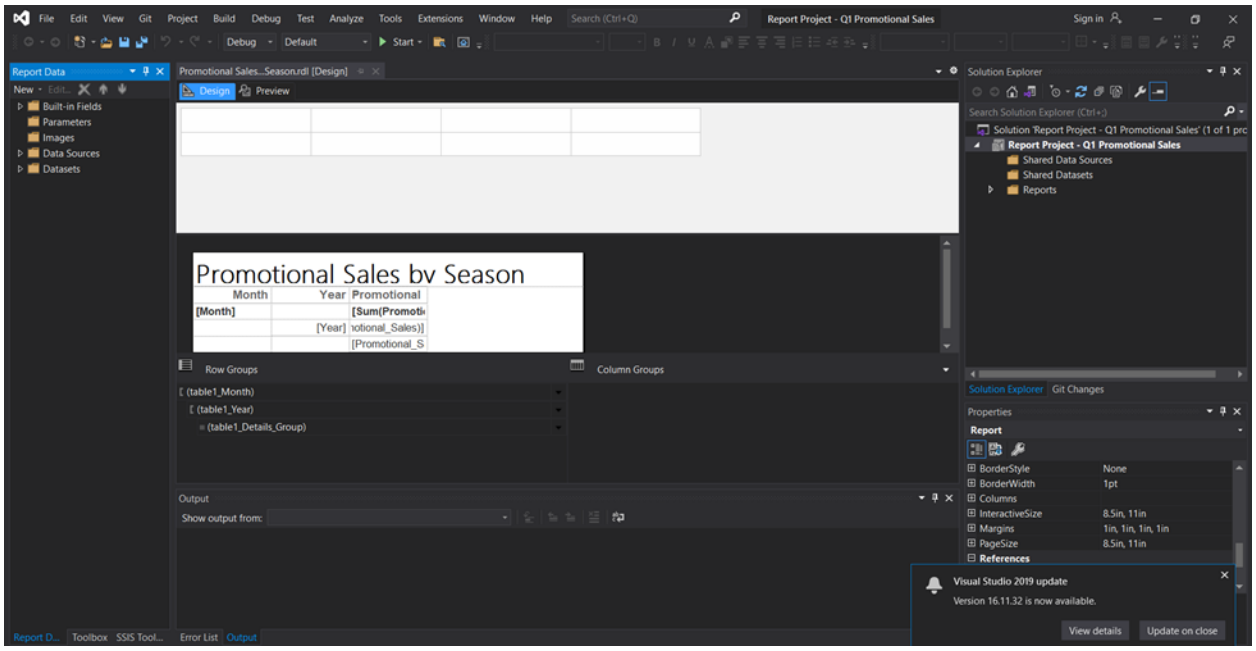


Figure: This gives us the design view of our report.

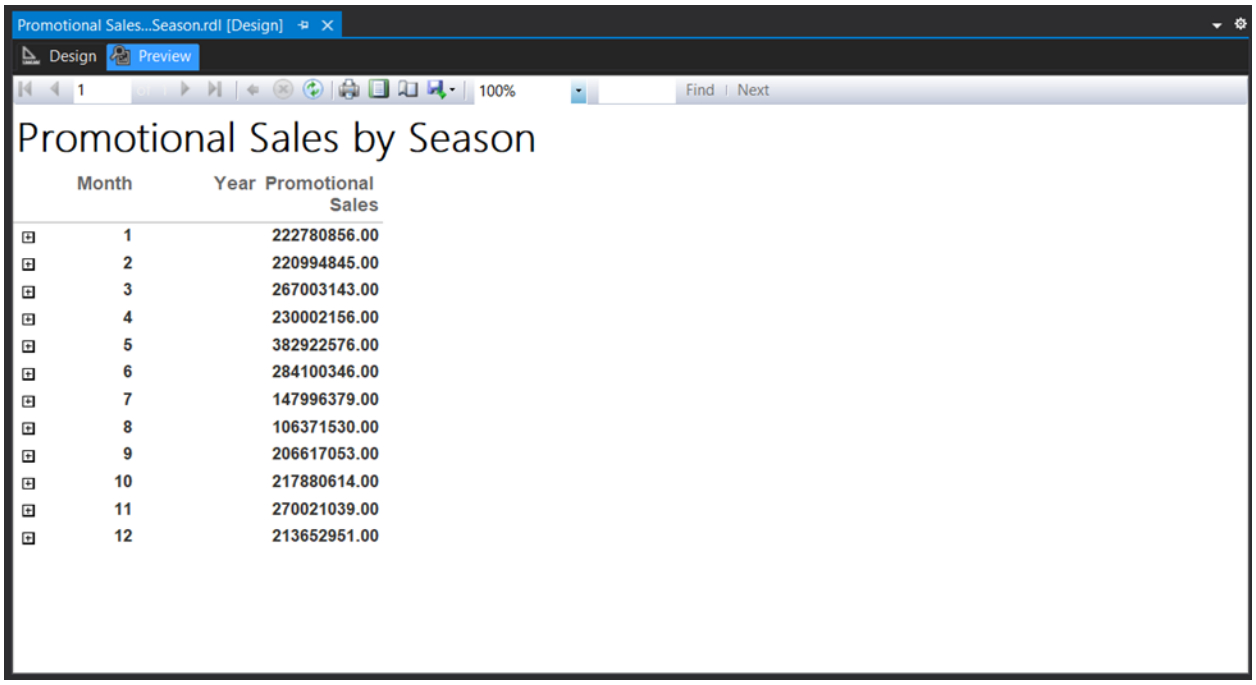


Figure: Preview of SSRS Report Showing the Seasonal Trend of Promotional Sales.

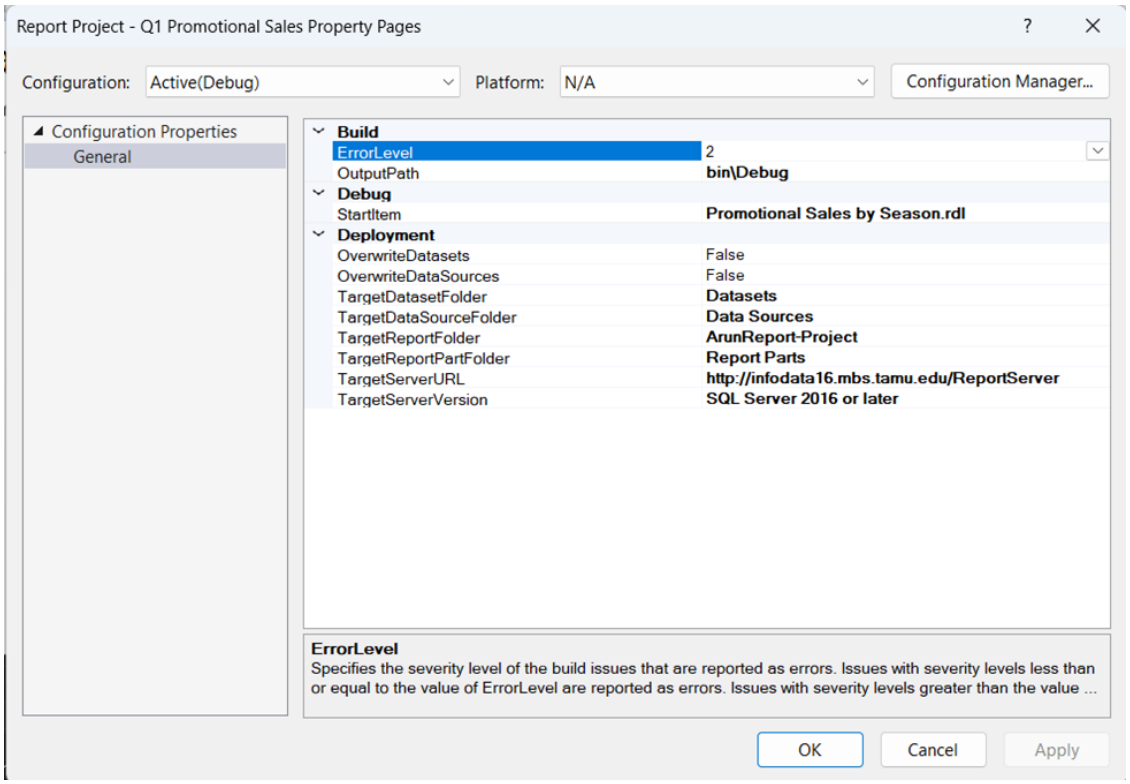


Figure: Changing the Target Server URL to our desired server.

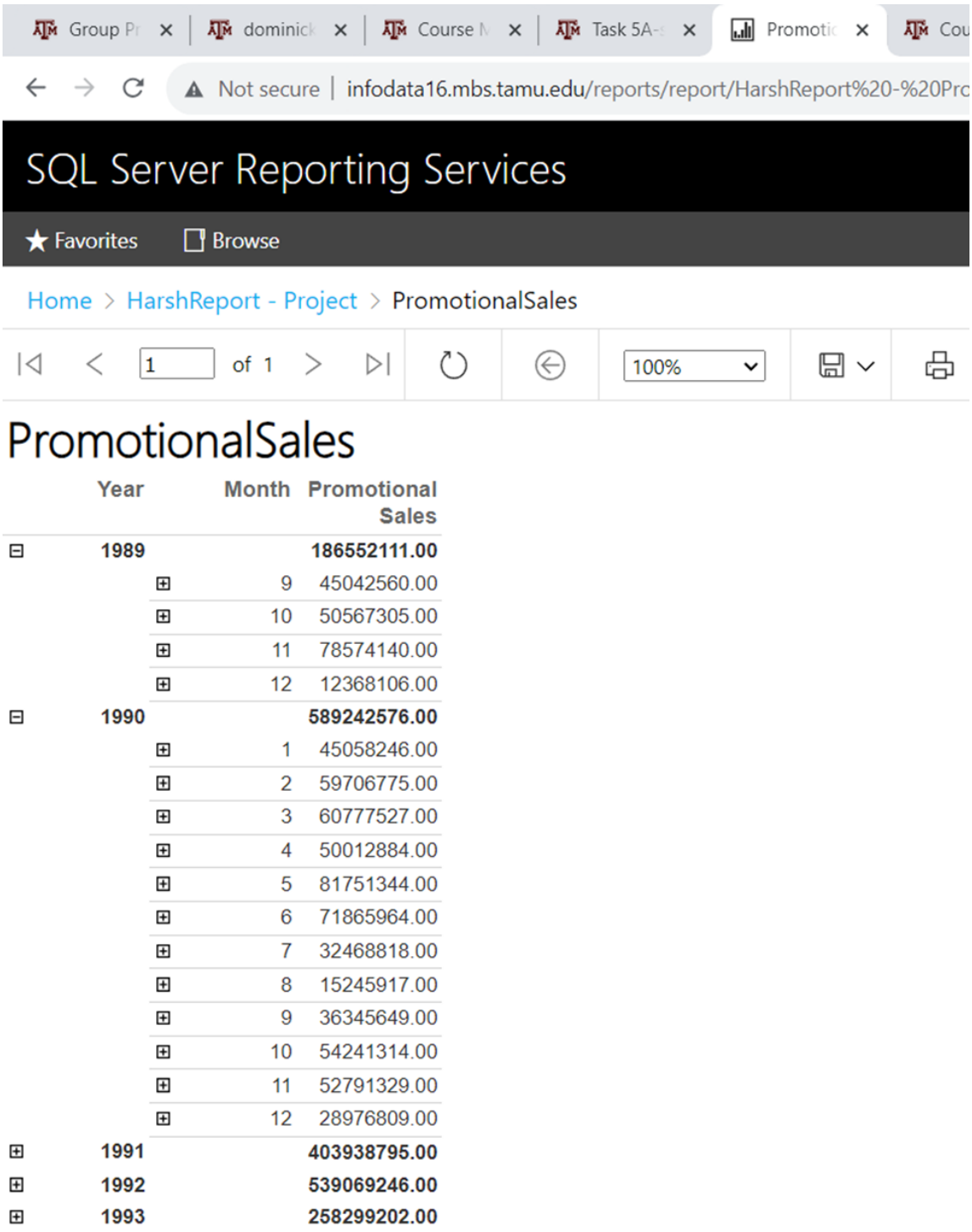


Figure: Deployed SSRS Report on the Target Server URL.

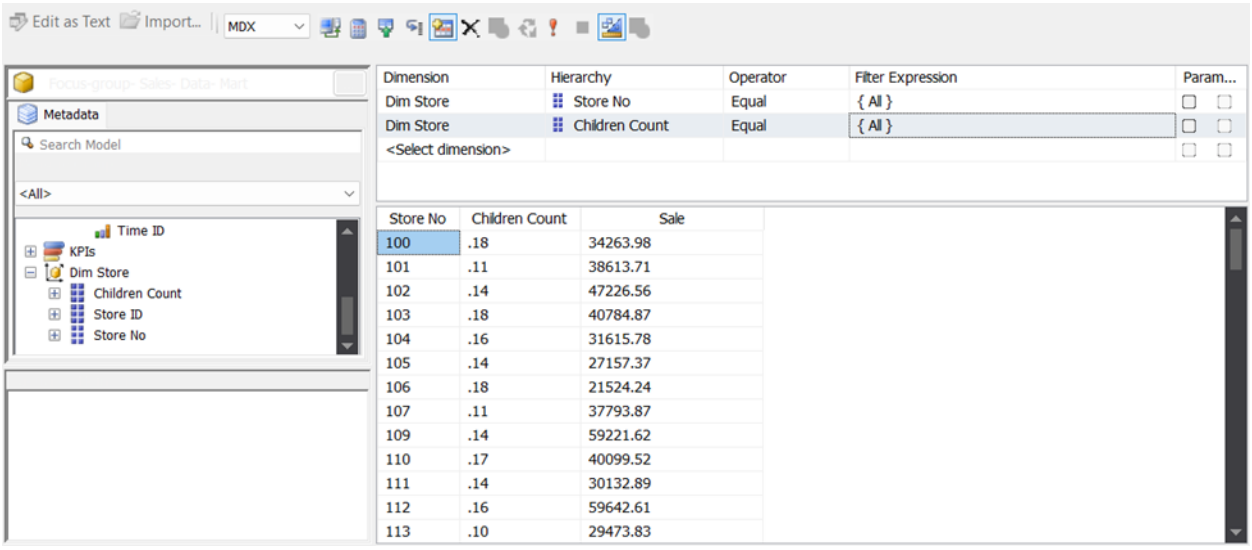
ISTM 637 - Group 5 Report

Business Question 2

What is the correlation between candy sales and the number(percentage) of children in each demographic area?

BI/Visualization Method: Report using Multidimensional analysis cubes in SSAS.

Report:



Store No	Children Count	Sale
100	.18	34263.98
101	.11	38613.71
102	.14	47226.56
103	.18	40784.87
104	.16	31615.78
105	.14	27157.37
106	.18	21524.24
107	.11	37793.87
109	.14	59221.62
110	.17	40099.52
111	.14	30132.89
112	.16	59642.61
113	.10	29473.83

Rationale:

Understanding the correlation between percentage of children and candy sales, DFF can enable targeted marketing, tailor product offerings, and enhance product placements. This can improve user experience and boost sales in this category.

Report Plan:

Using SSAS a cube of the FactSales and DimStore tables was created to provide an efficient query structure. Later on, the query was designed by pulling the store number, children count and candy sales columns.

Implementation Steps:

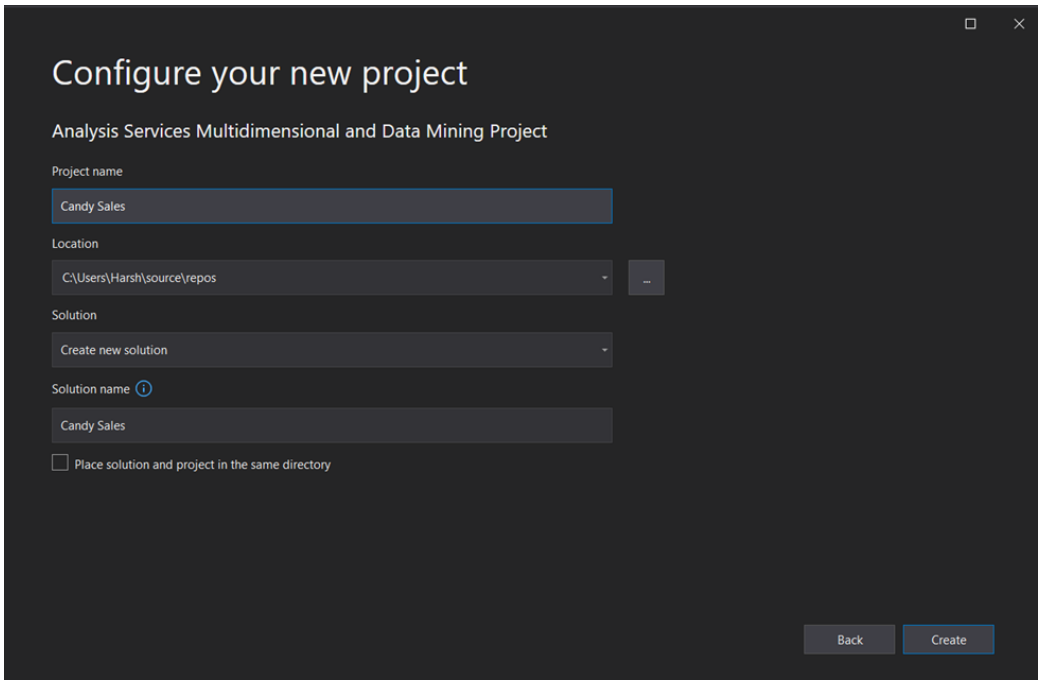


Figure: Configuring the Analysis Services Multidimensional and Data Mining Project.

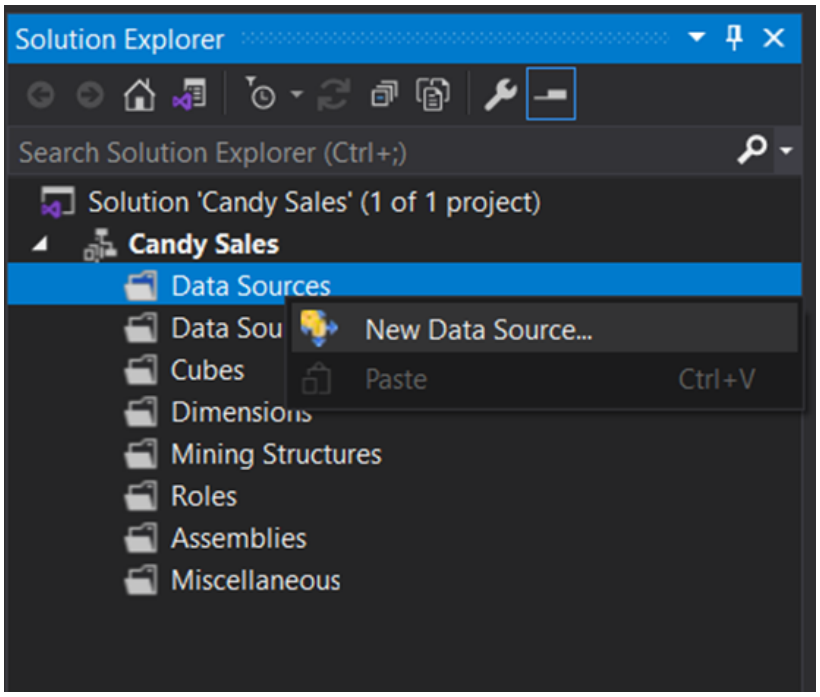


Figure: Configuring a new data source.

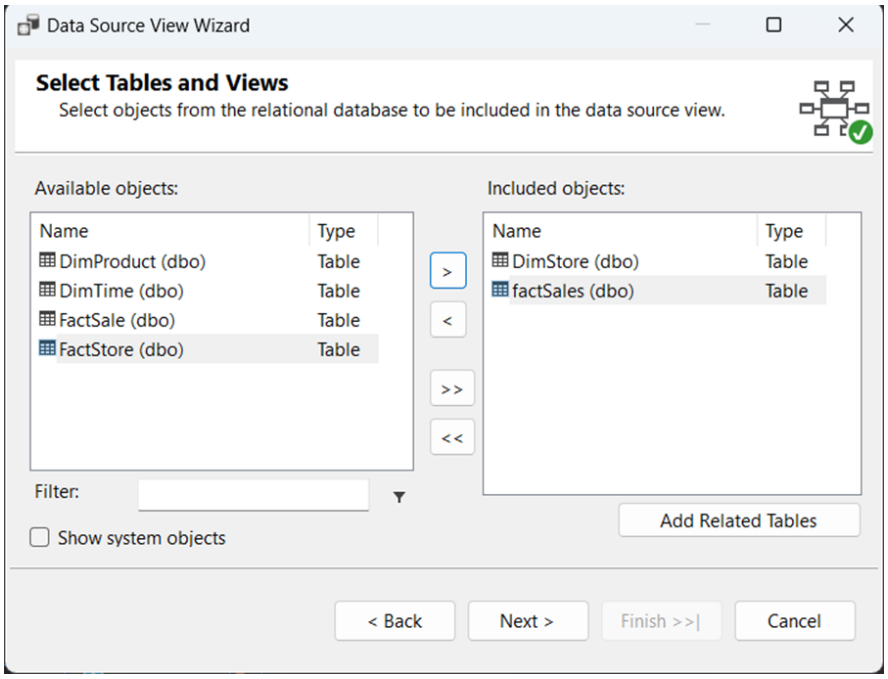


Figure: Selecting the required fact and dimension tables from the independent data mart.

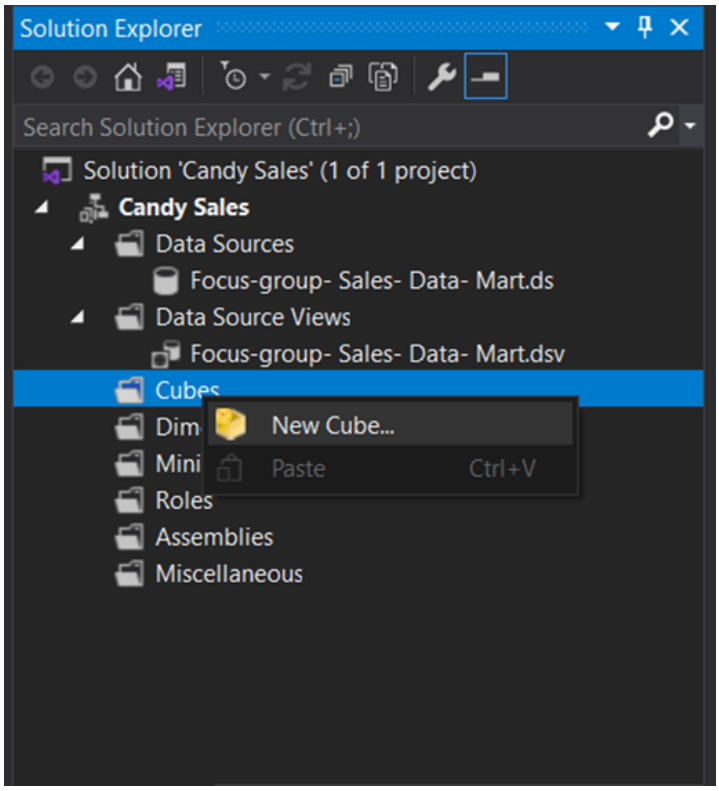


Figure: Selecting a new cube.

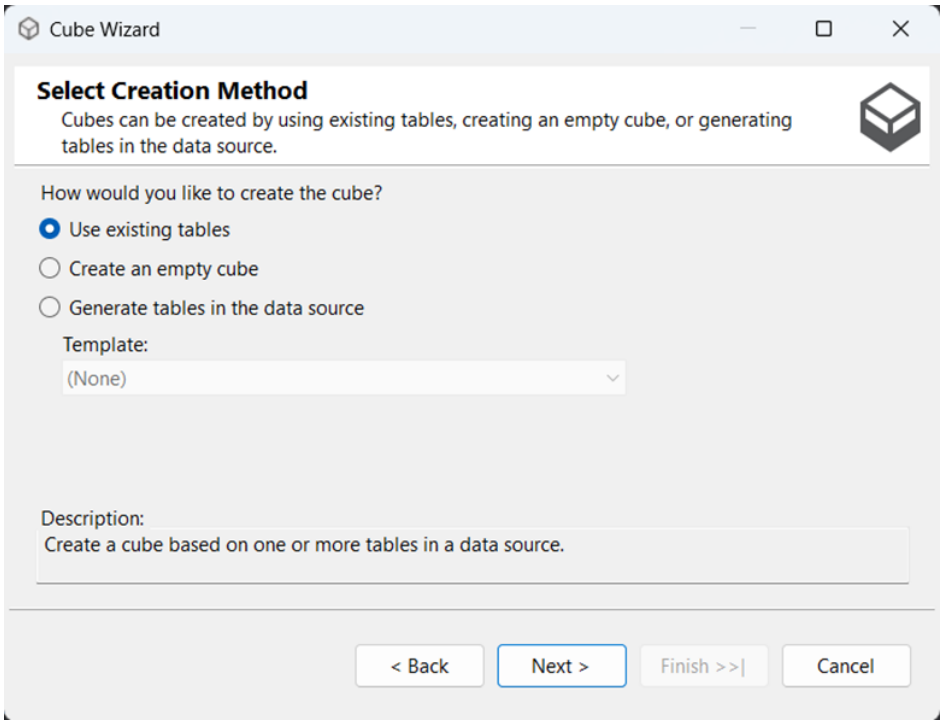


Figure: Using existing tables to create the cube.

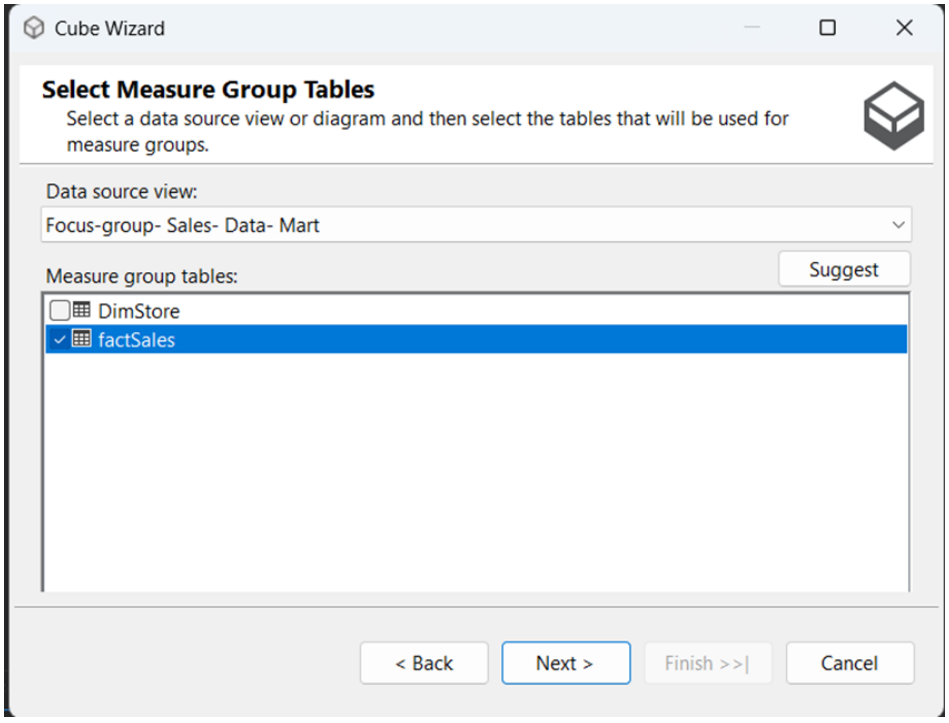


Figure: Selecting the appropriate fact tables into the measure group tables.

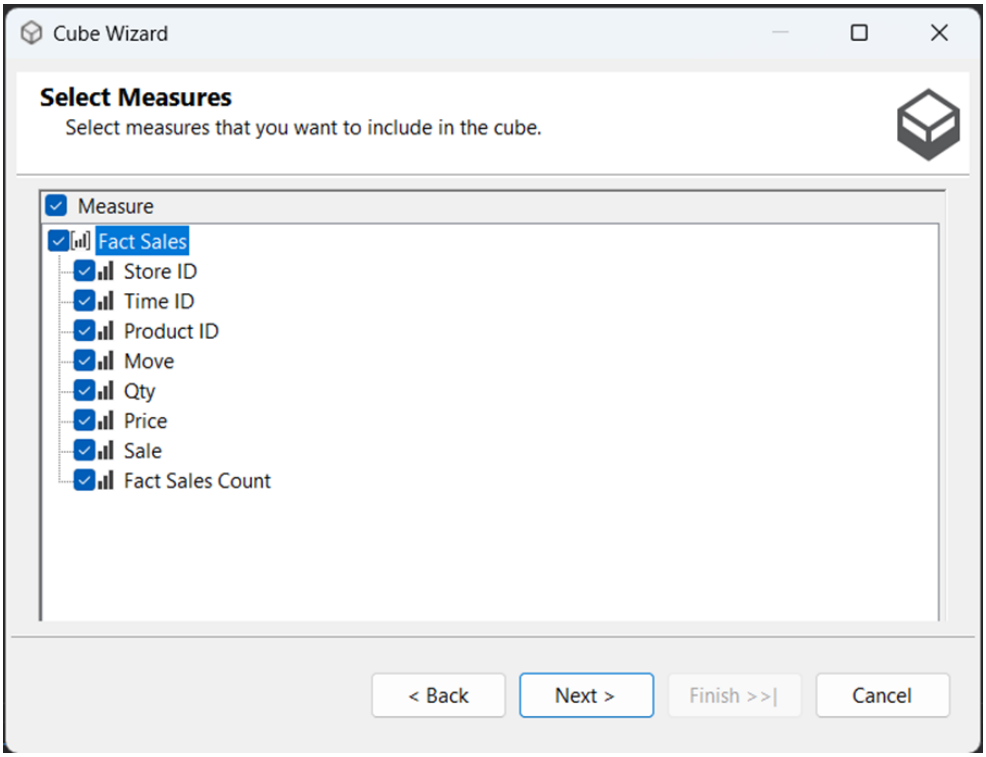


Figure: Selecting the appropriate measures from the measure group tables.

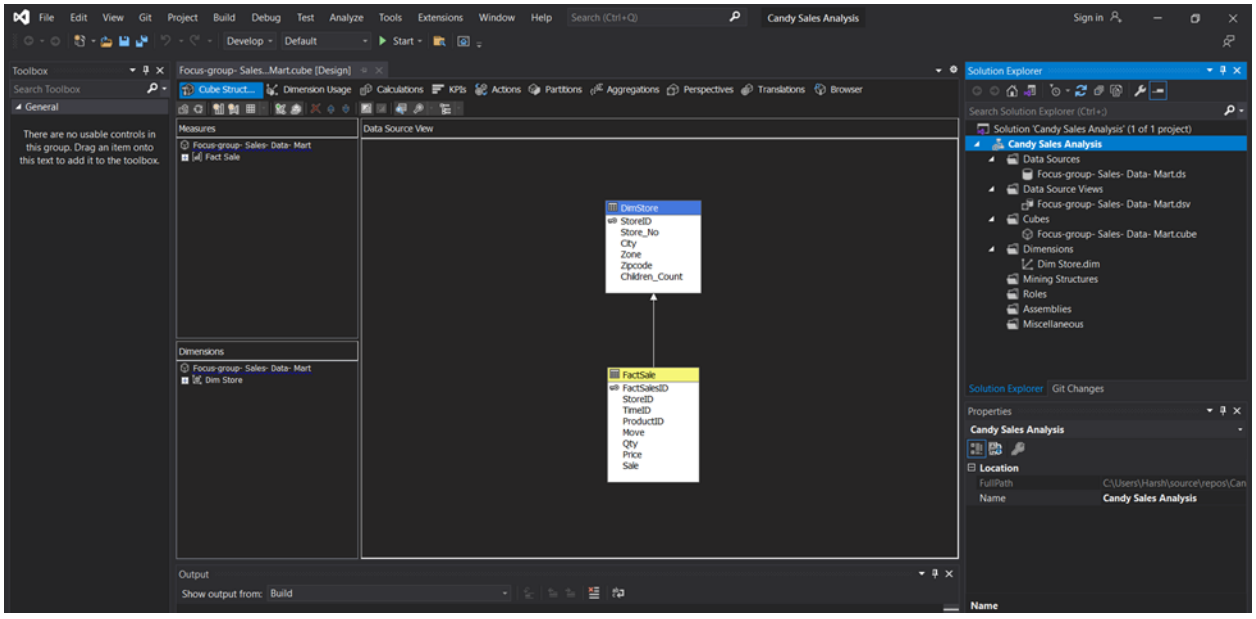


Figure: A view of the final cube.

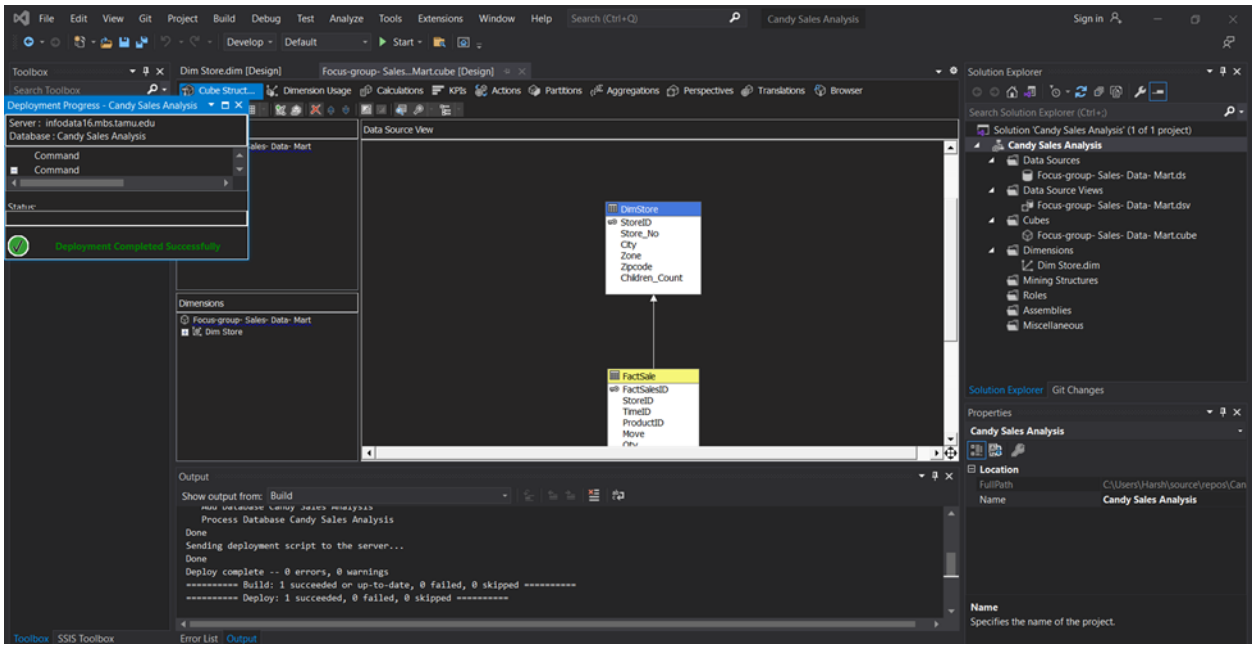


Figure: Successfully deploying the cube to the appropriate server.

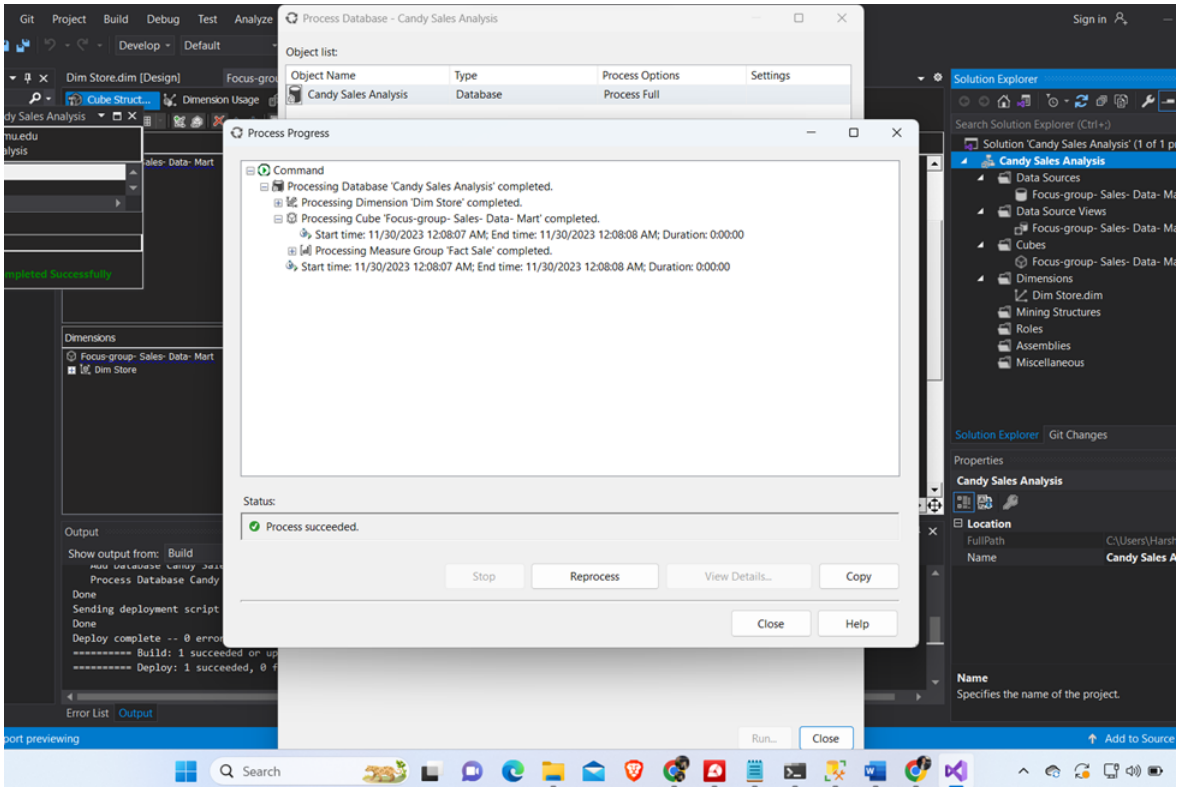


Figure: Successfully executing the process.

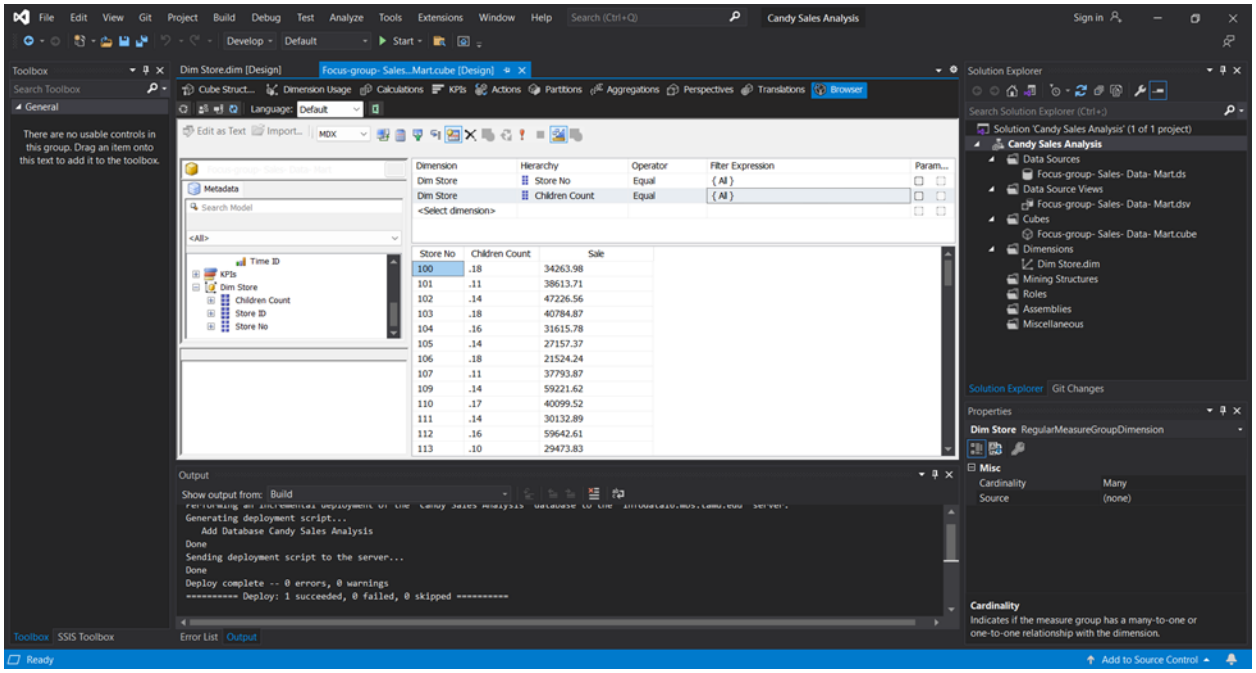


Figure: We now have the final report of Candy Sales and Children Demographic of each store belonging to DFF

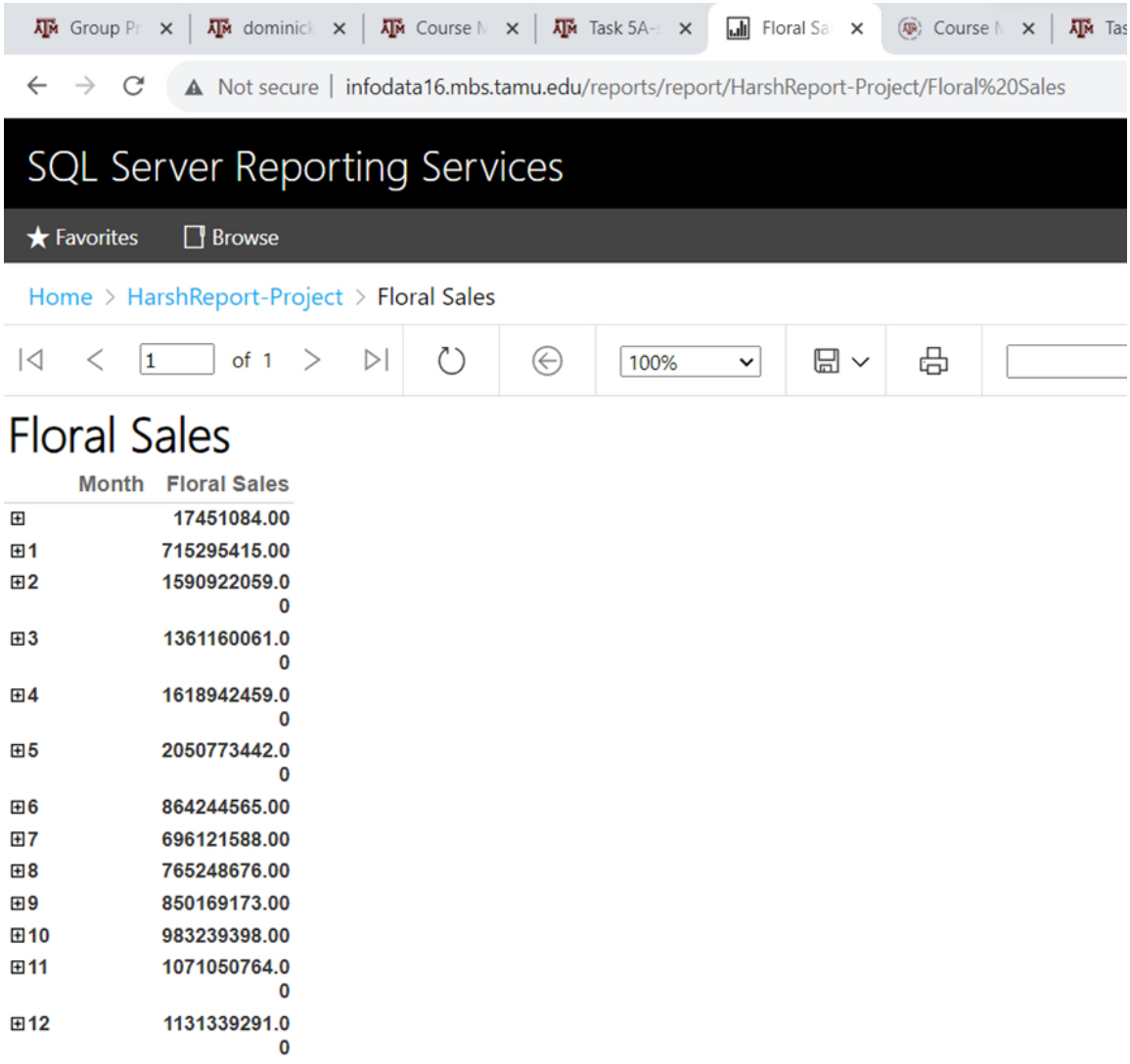
ISTM 637 - Group 5 Report

Business Question 3

What is the impact of festive occasions and other significant events on monthly floral sales?

BI/Visualization Method: Report using Multidimensional Analysis Cube from SSAS and Reporting using SSRS.

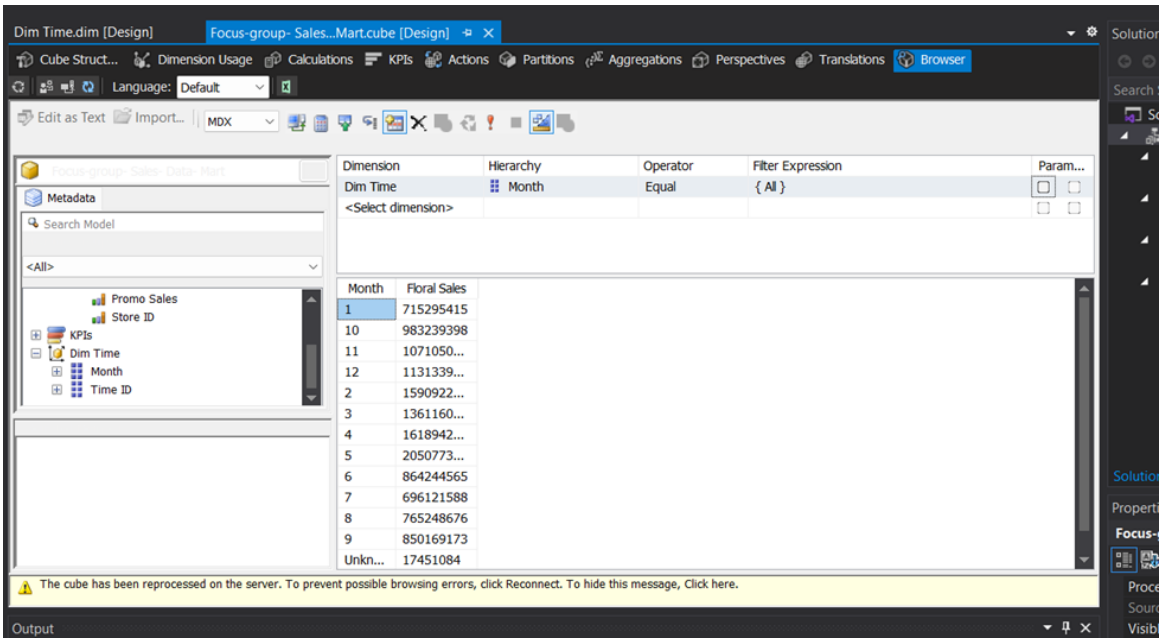
Report:



Floral Sales

Month	Floral Sales
⊞	17451084.00
⊞1	715295415.00
⊞2	1590922059.00
⊞3	1361160061.00
⊞4	1618942459.00
⊞5	2050773442.00
⊞6	864244565.00
⊞7	696121588.00
⊞8	765248676.00
⊞9	850169173.00
⊞10	983239398.00
⊞11	1071050764.00
⊞12	1131339291.00

ISTM 637 - Group 5 Report



Rationale:

The retail landscape is a system based on the inputs of supply and outputs of demand. The factors that affect the supply are seasonality of a product and those that affect the demand are the emotional significance of a product in a consumer's mind, cultural norms, and market trends. This question will have a multifaceted solution proposition for Dominick's Fast Foods Corporation as it will solve demand as well as supply issues related to the seasonality of floral goods supply and demand.

Reporting Plan:

The given business question was analyzed using the Multidimensional Analysis Cube from SSAS and Reporting using SSRS. This helped in viewing the analysis through a cube as well as through a direct report. The use of the cube was to accelerate the query performance on the data to generate better insights.

Implementation Steps:

SSRS:

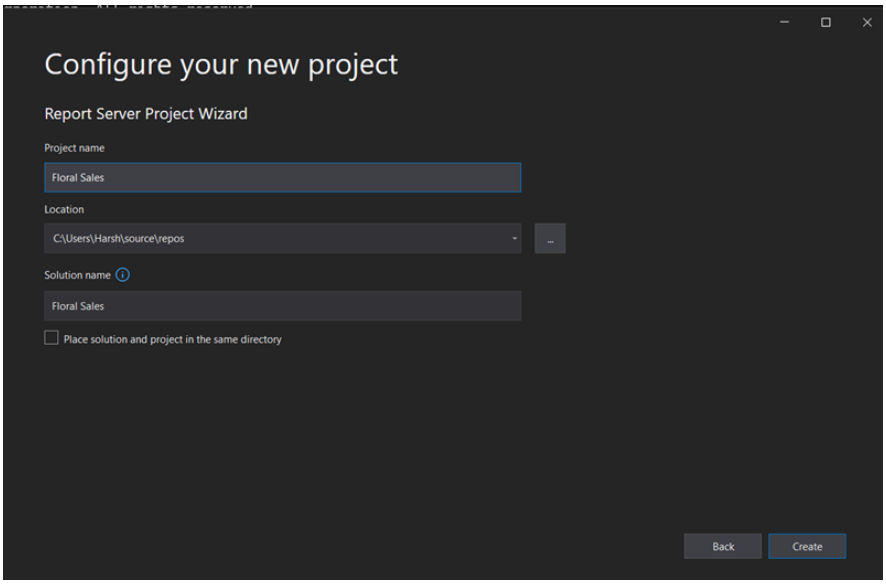


Figure: Configuring the new Project Wizard.

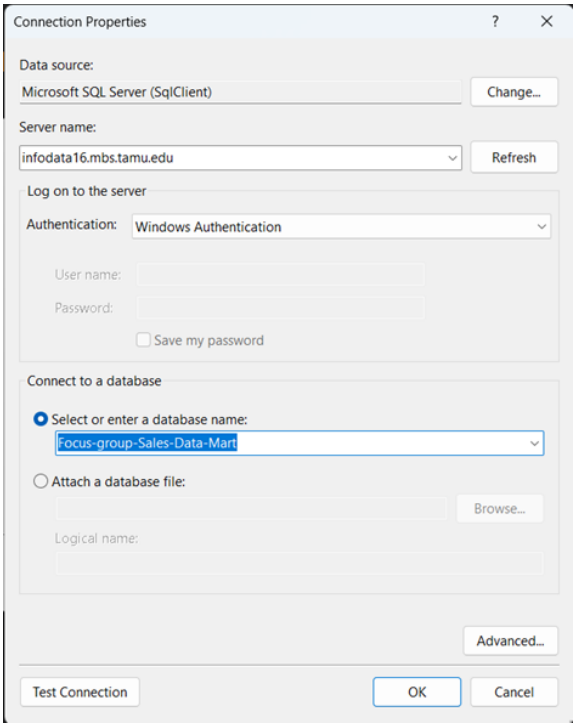


Figure: Configuring the connection to the server and database.

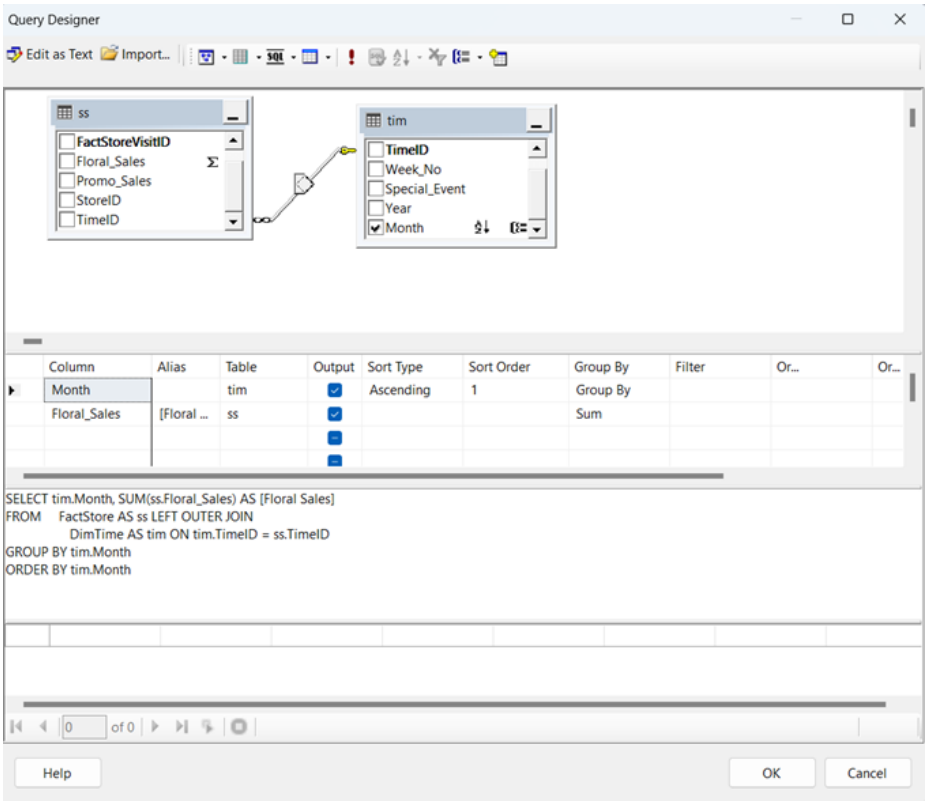


Figure: Designing the query with the appropriate joins and aggregations.

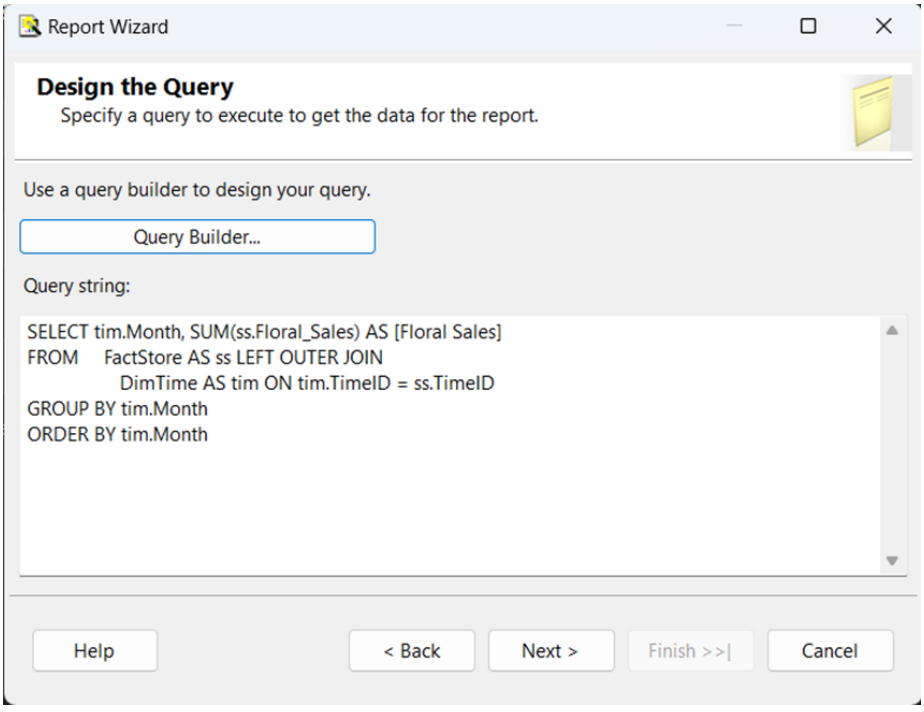


Figure: Finalizing the query.

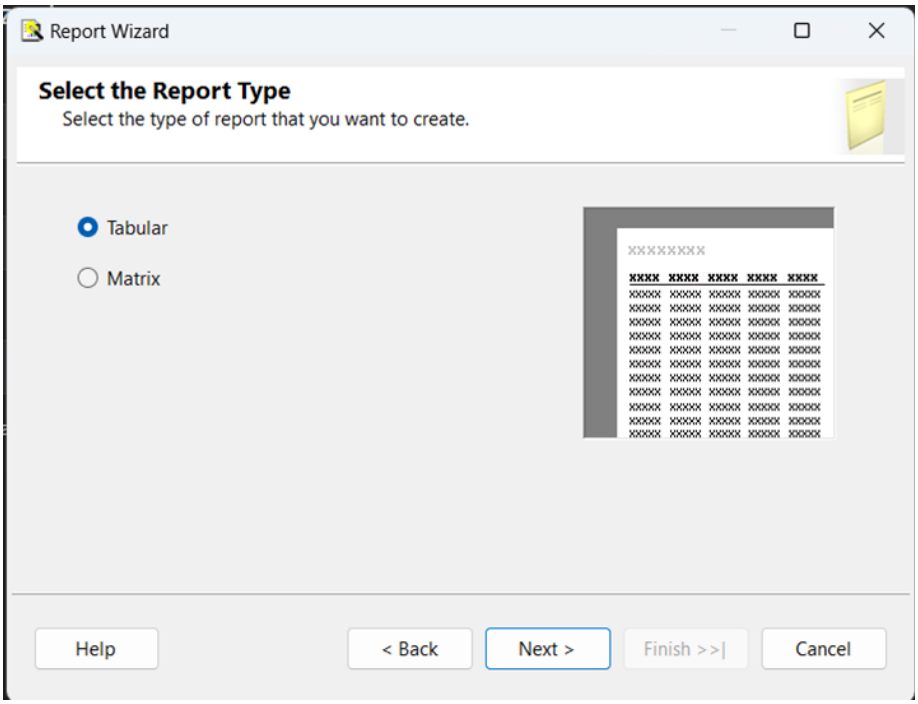


Figure: Selecting the tabular format for the report.

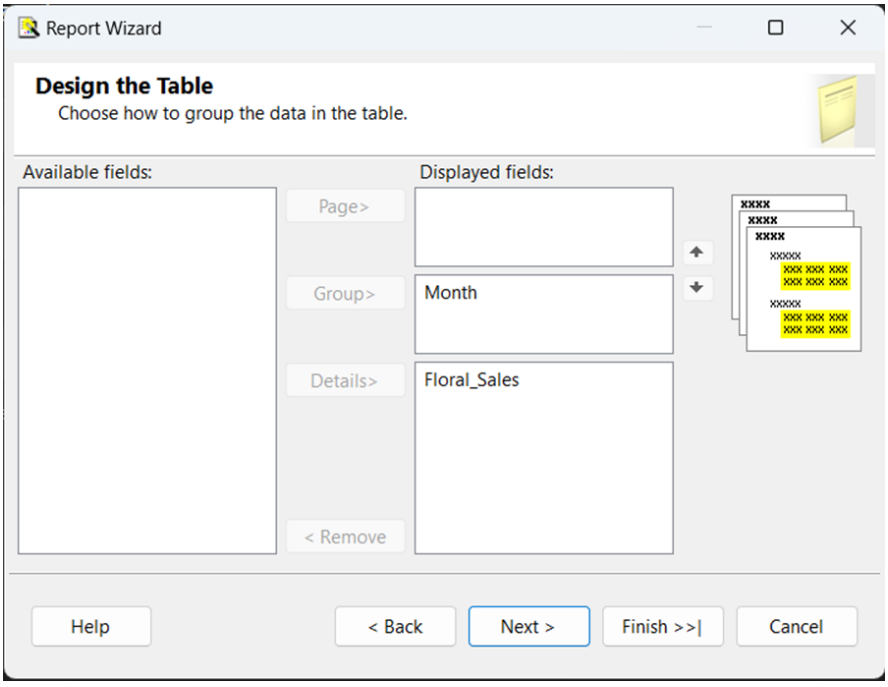


Figure: Splitting the fields into groupings and details.

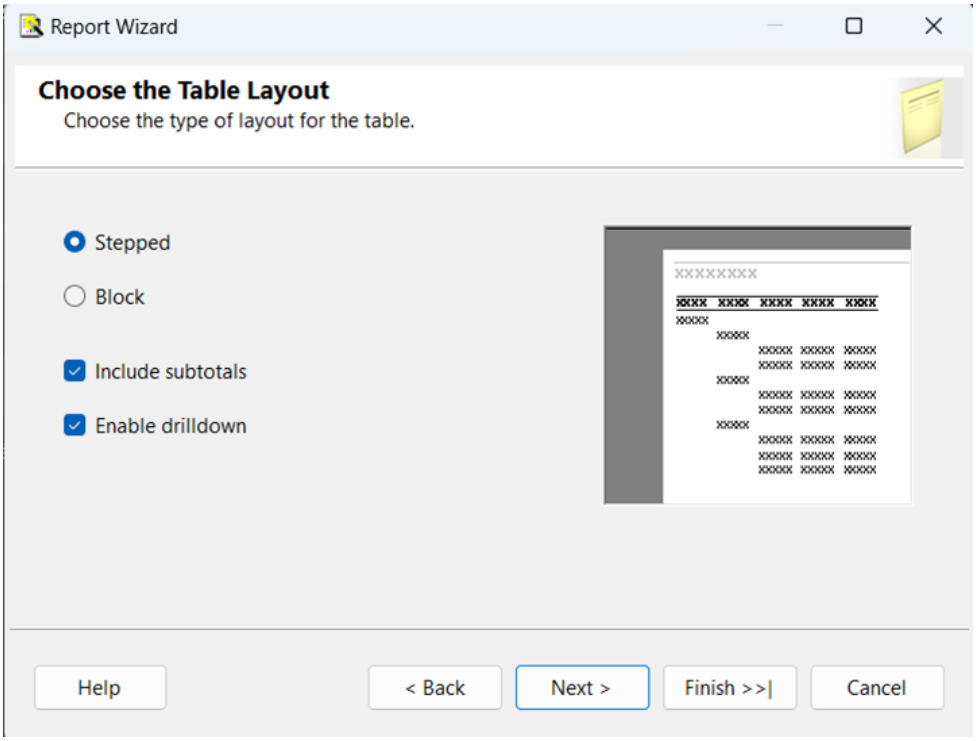


Figure: Enabling Subtotals and Drilldown for further analysis.

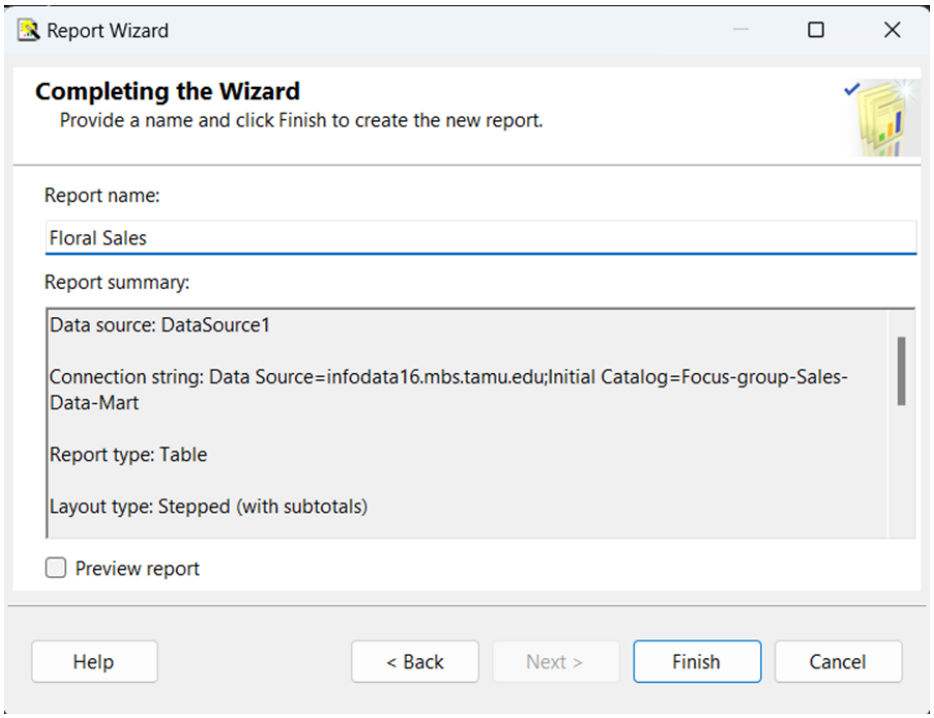


Figure: Finalizing the details.

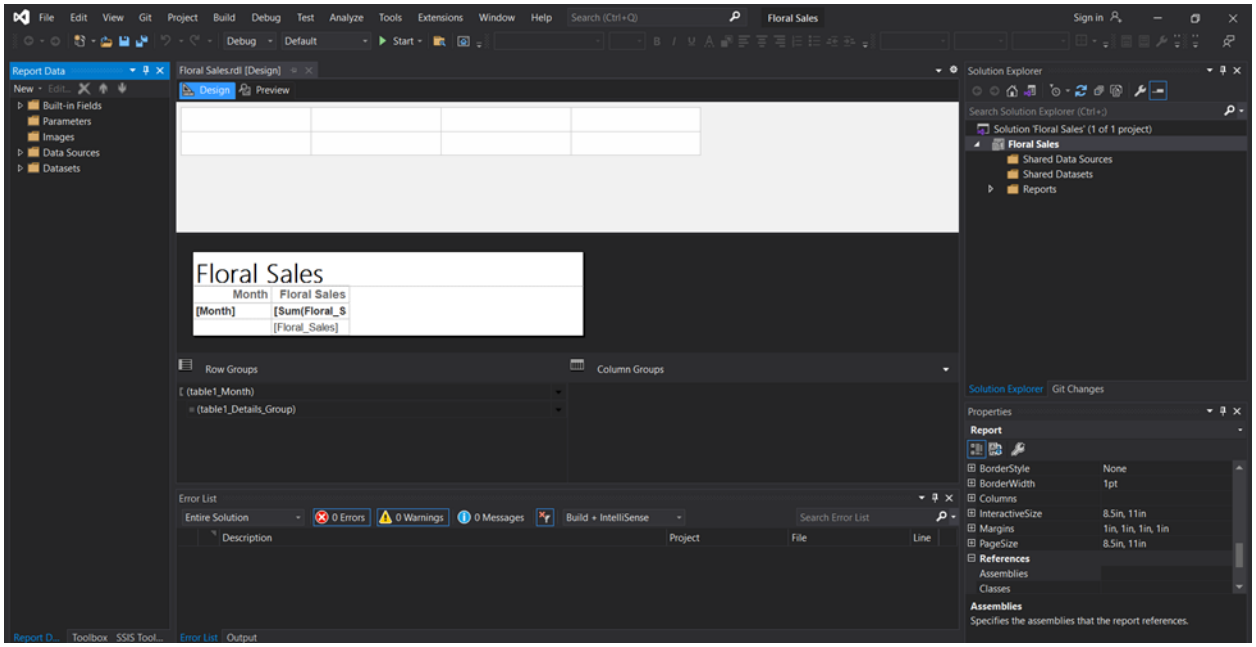


Figure: Design View of the SSRS Report.

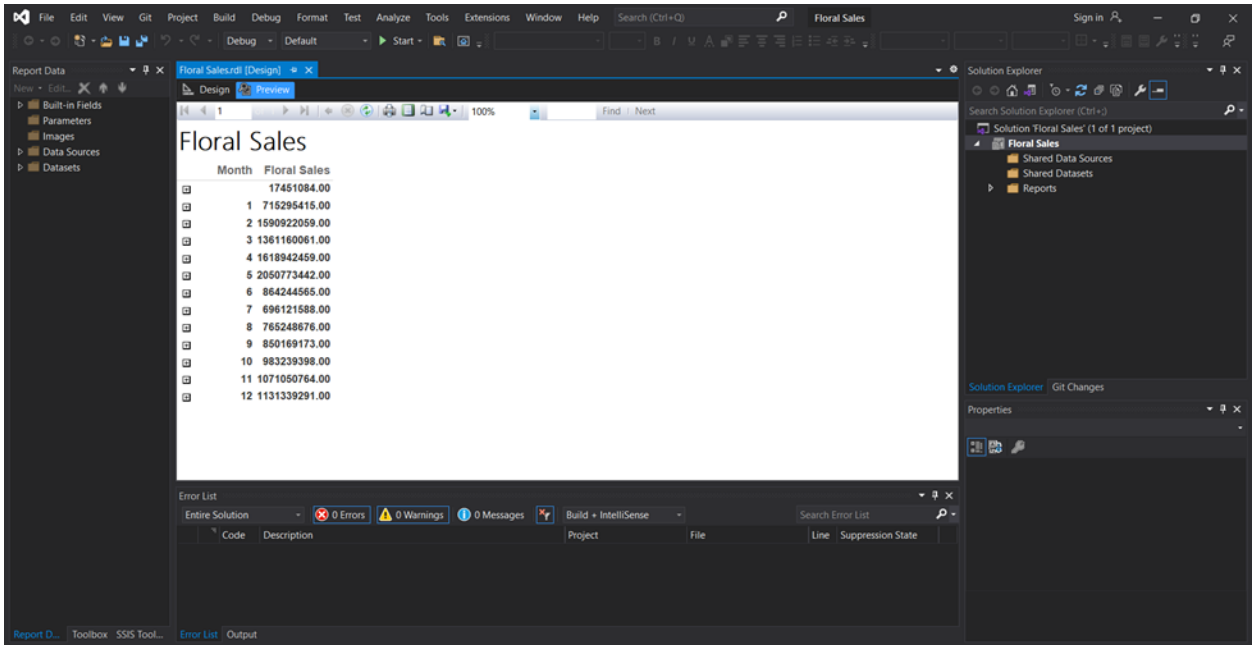


Figure: Report Preview.

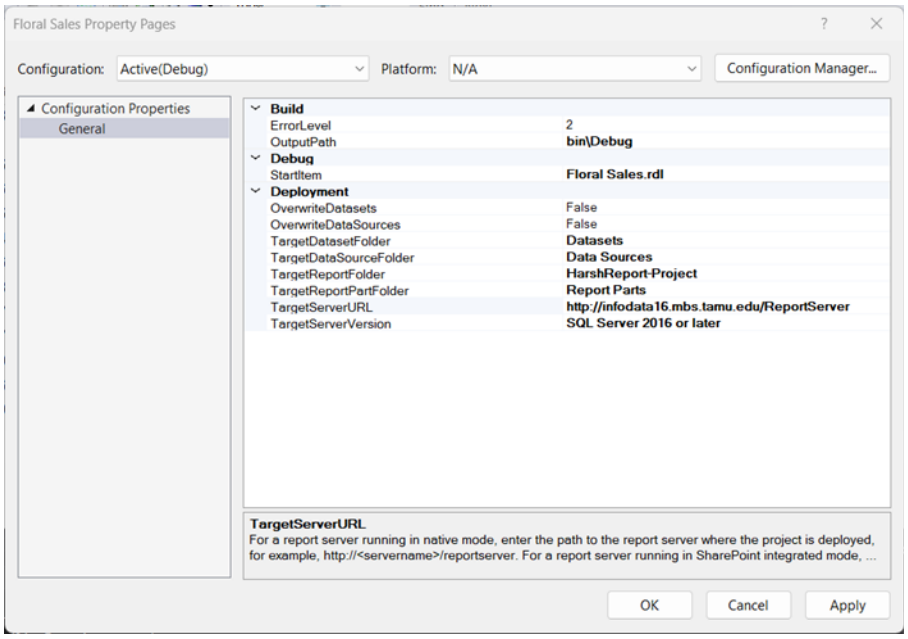


Figure: Editing the target server url to replace it with the correct one.

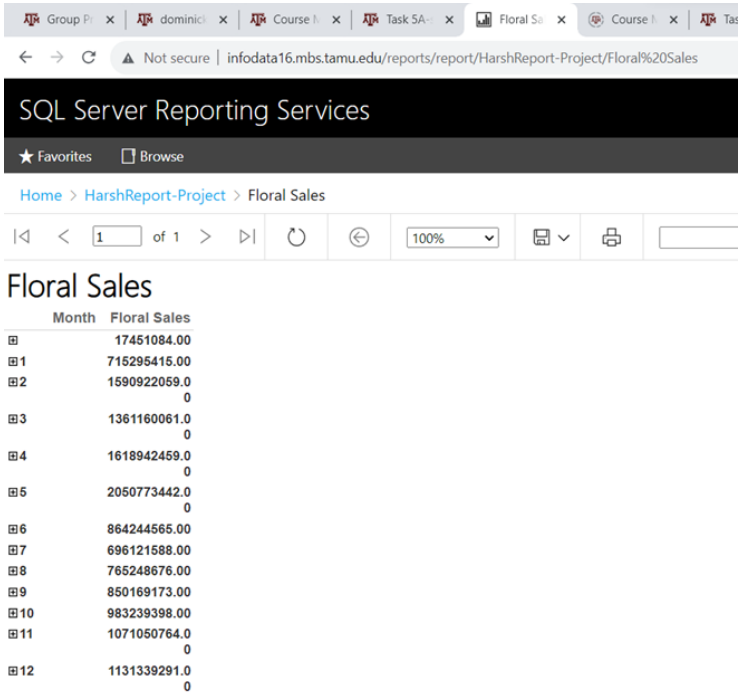


Figure: Final SSRS Report

SSAS:

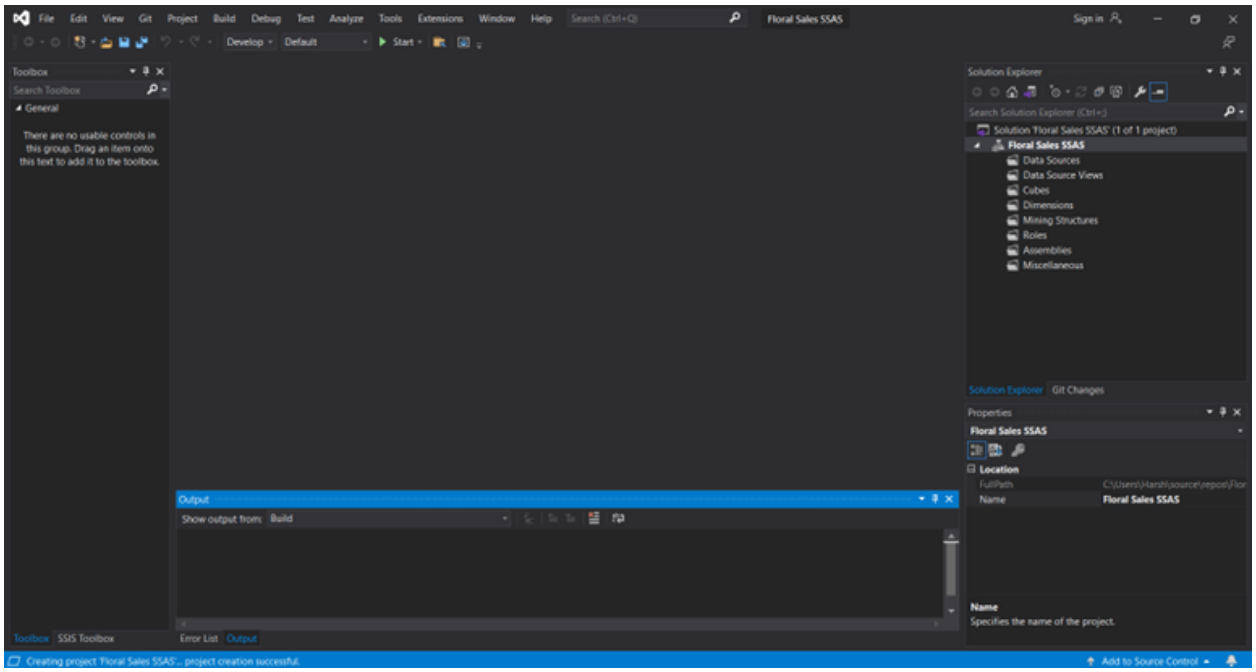


Figure: Opening the Analysis Project.

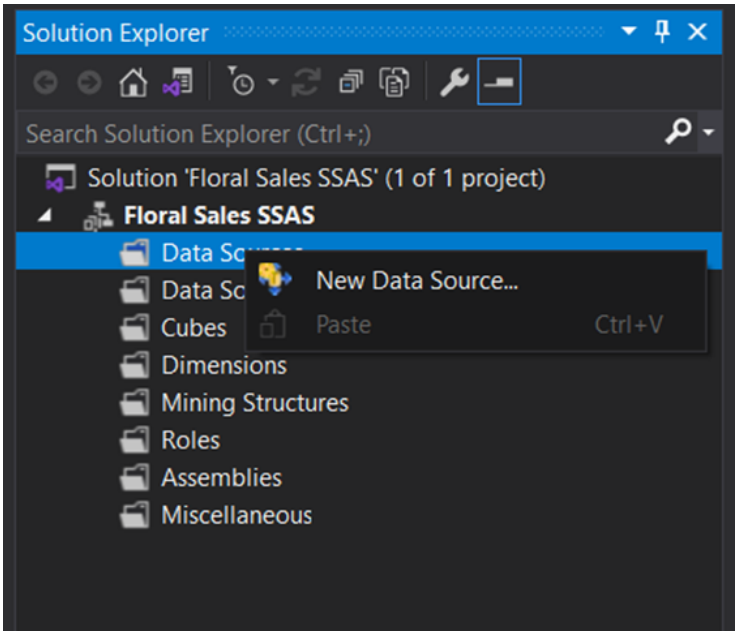


Figure: Configuring the Data Source.

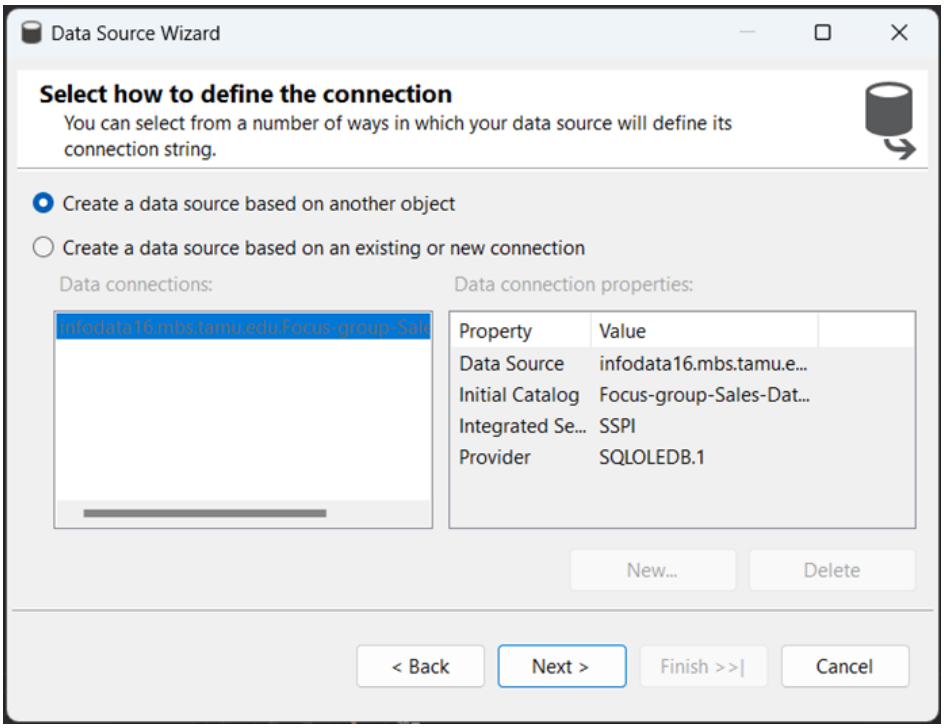


Figure: Establishing a connection with the infodata16 server.

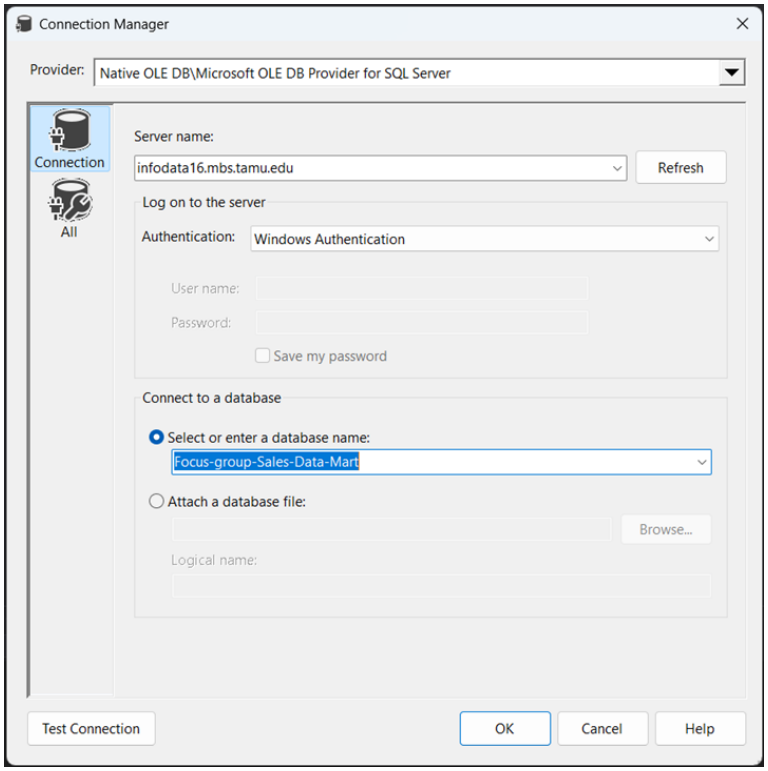


Figure: Selecting our specific Database.

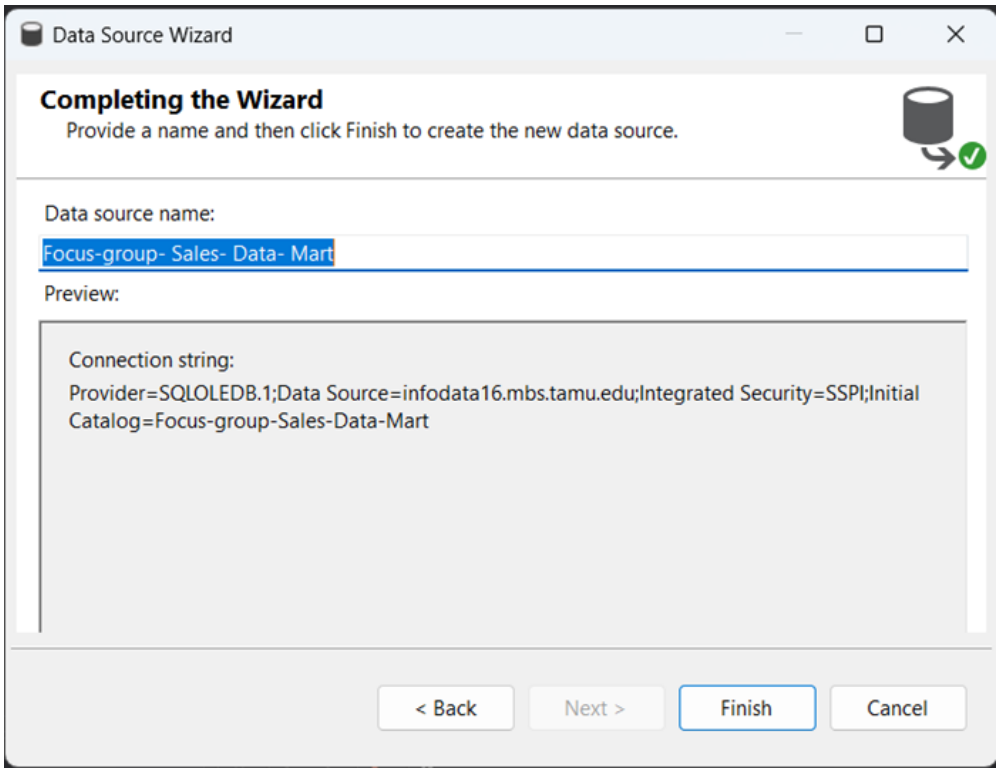


Figure: Finalizing details.

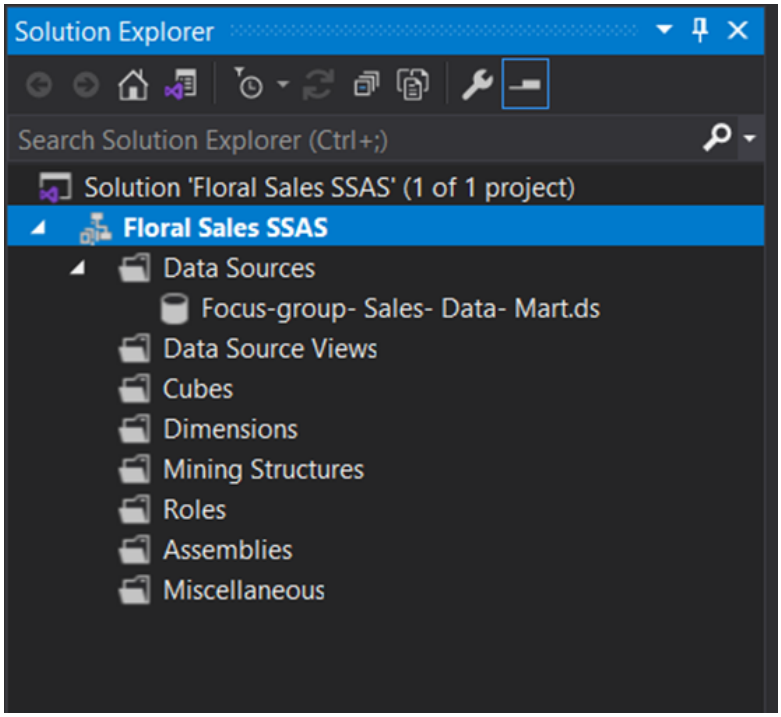


Figure: Data Source is configured.

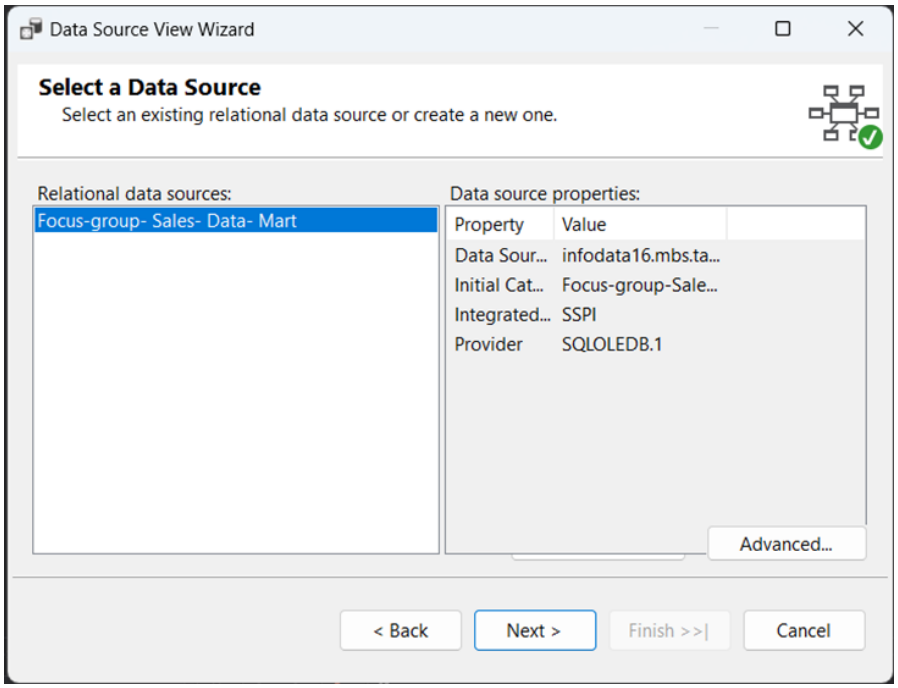


Figure: Configuring Data Source Views.

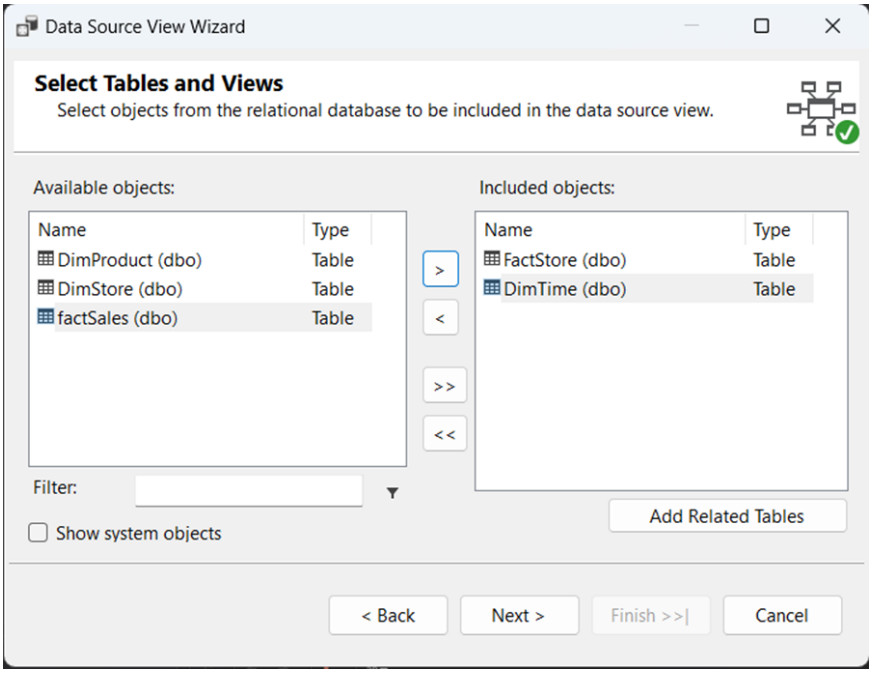


Figure: Selecting the required Fact and Dimension tables.

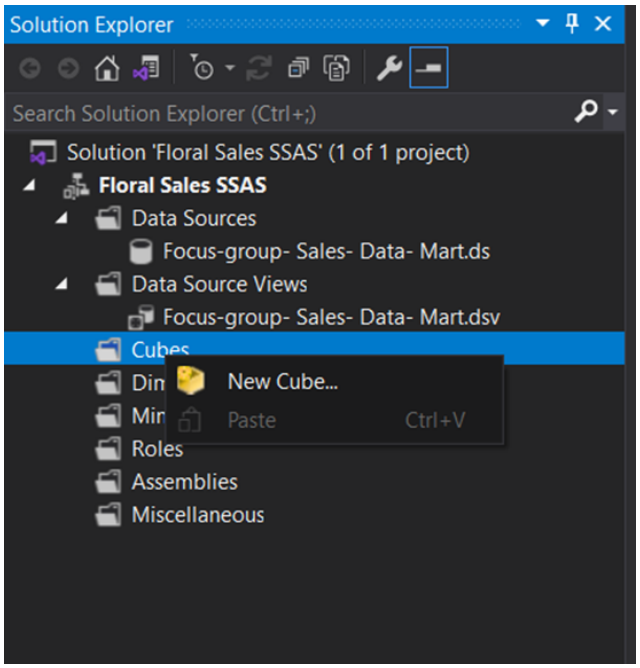


Figure: Configuring a new Cube.

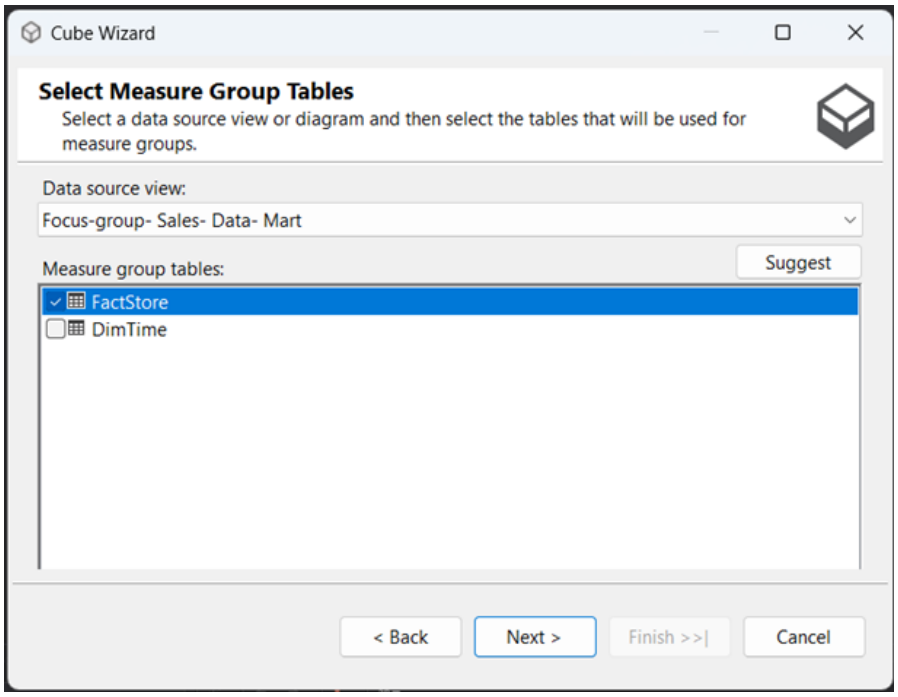


Figure: Defining the Measure Group tables - Factstore.

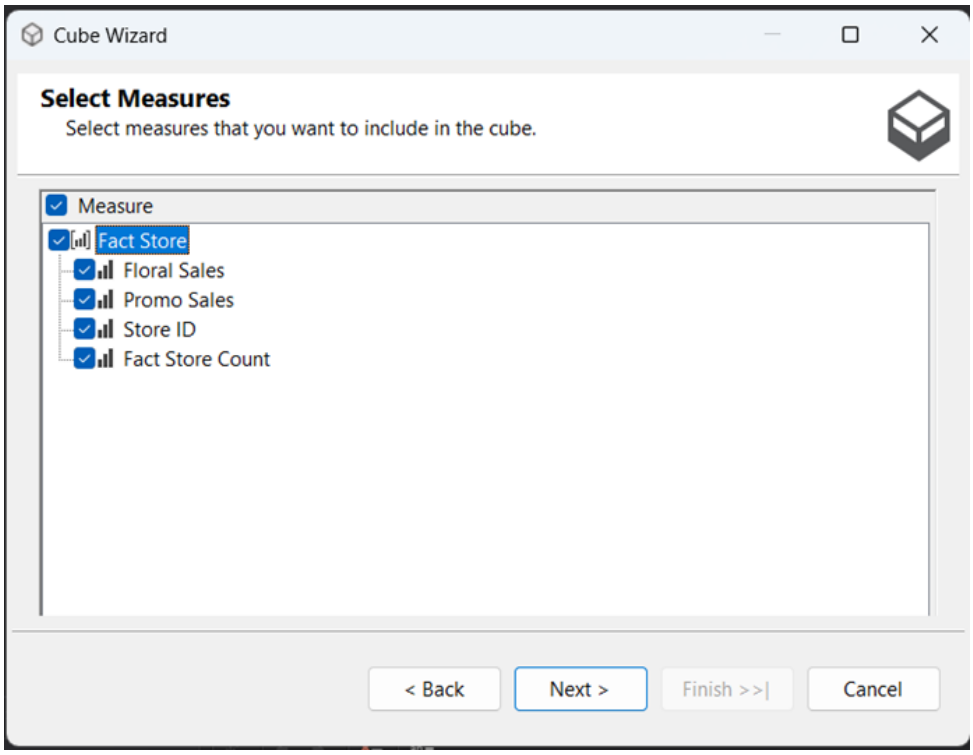


Figure: Selecting the relevant measures.

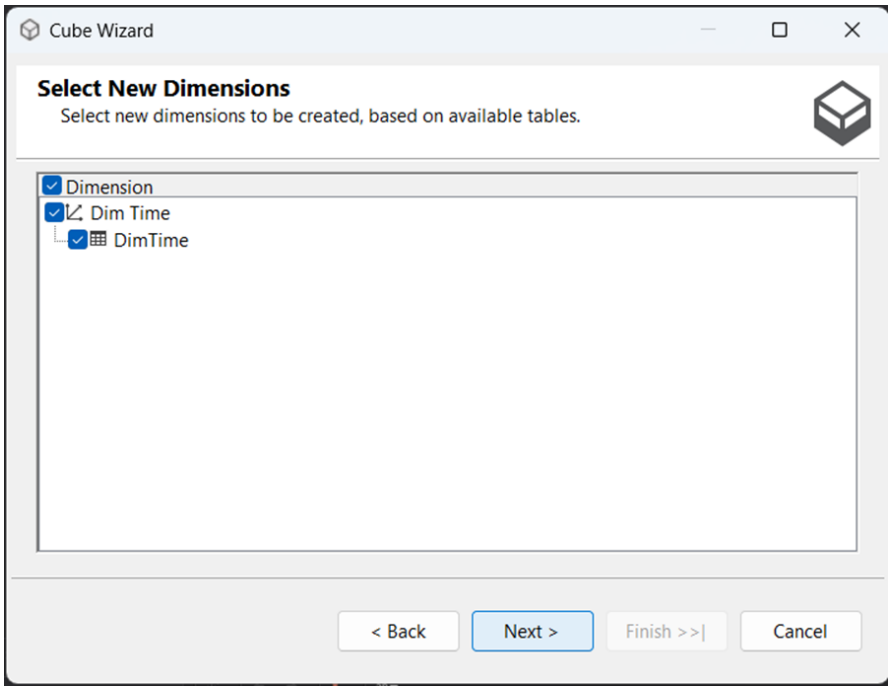


Figure: Selecting the Dimension tables.

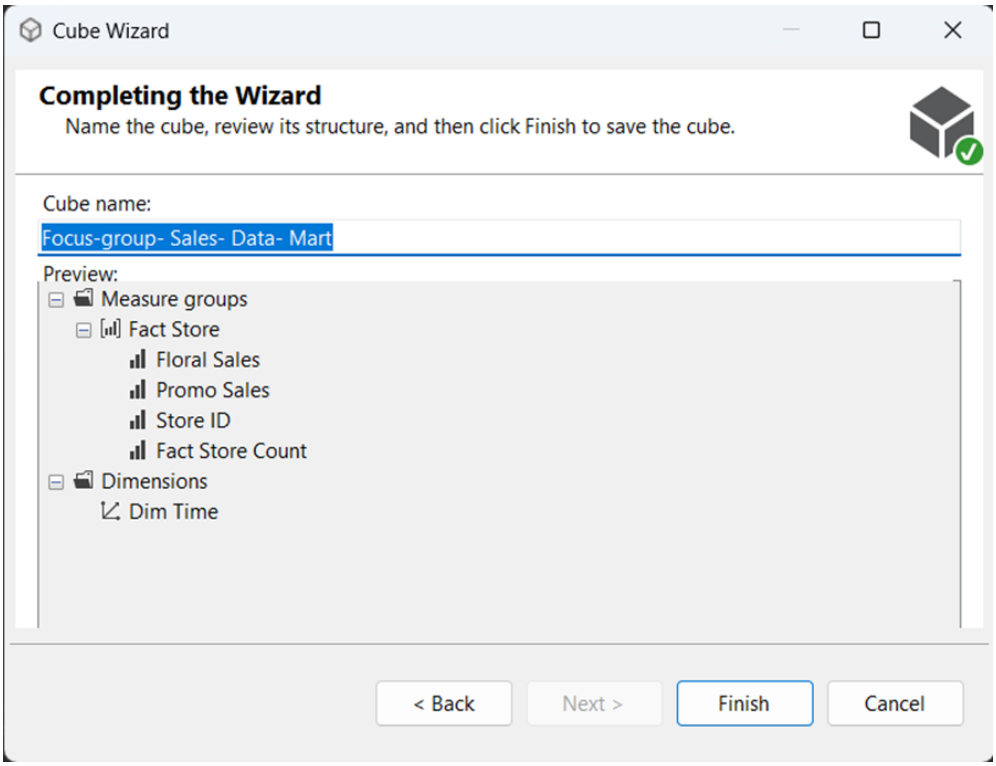


Figure: Finalizing the Measure and Dimension groups.

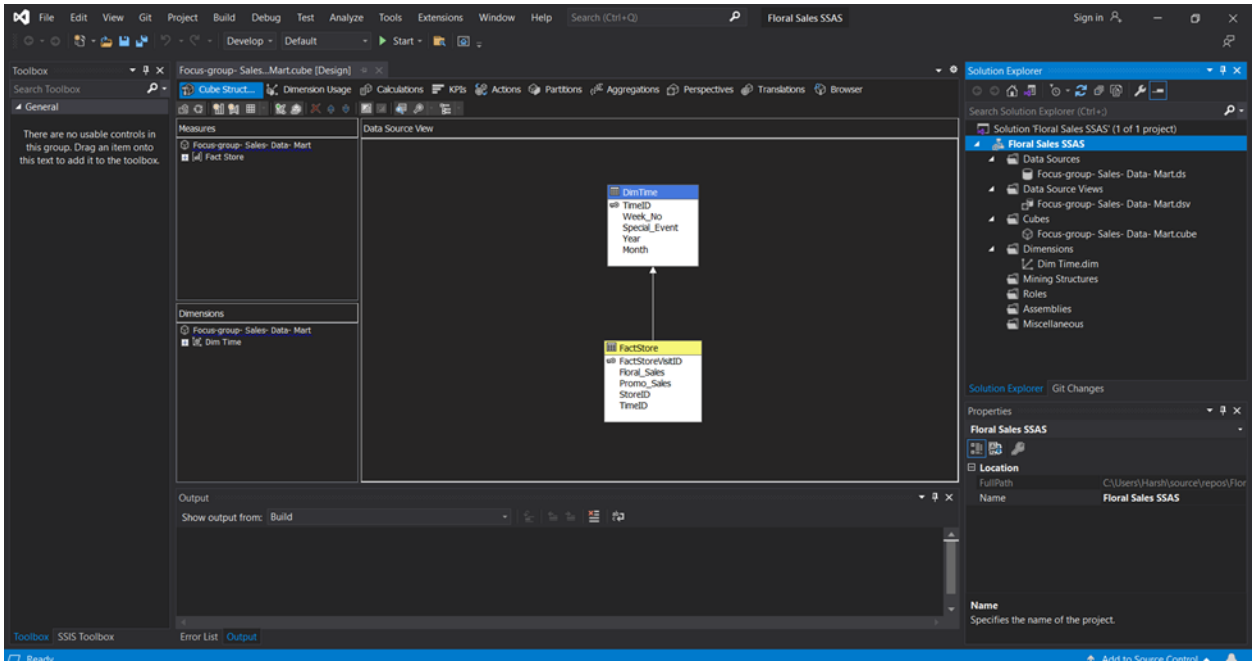


Figure: Our cube containing the Factstore and DimTime views.

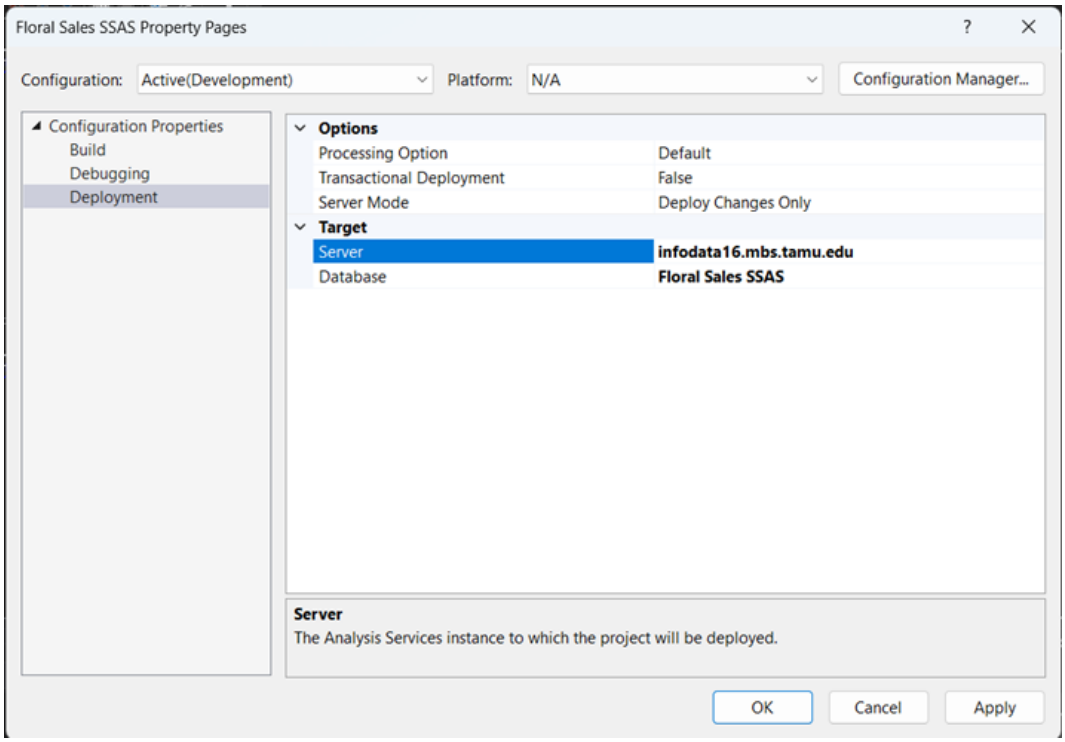


Figure: Configuring the correct server to deploy the report.

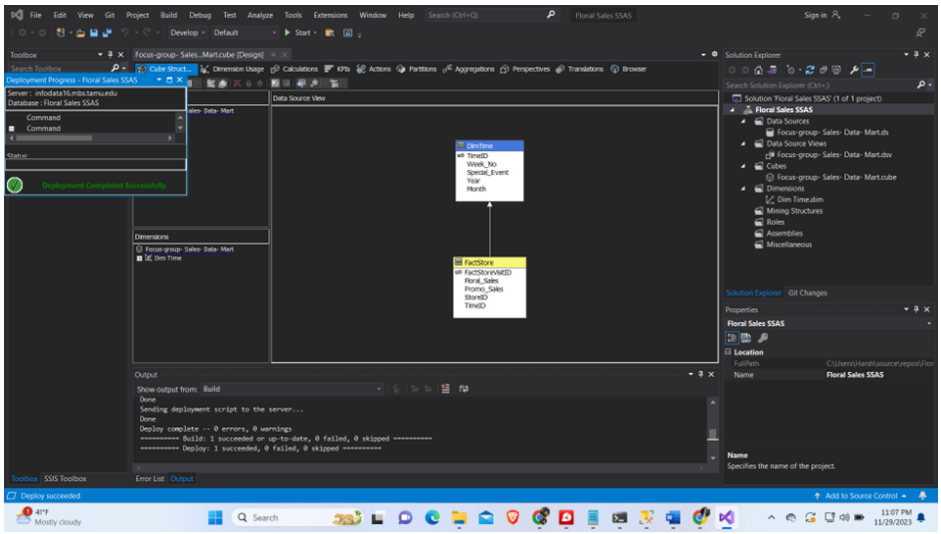


Figure: The deployment is successfully executed.

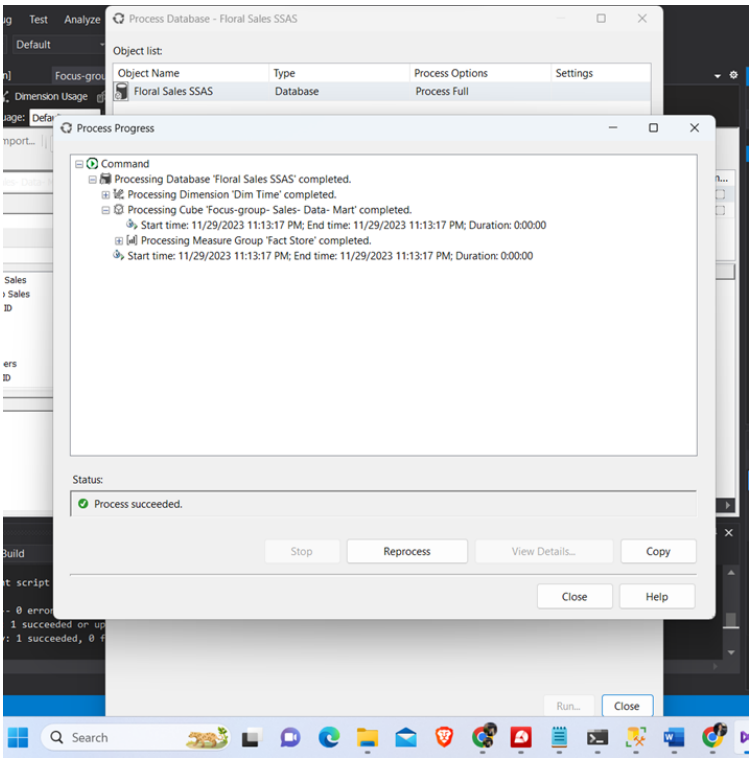


Figure: The process is executed successfully.

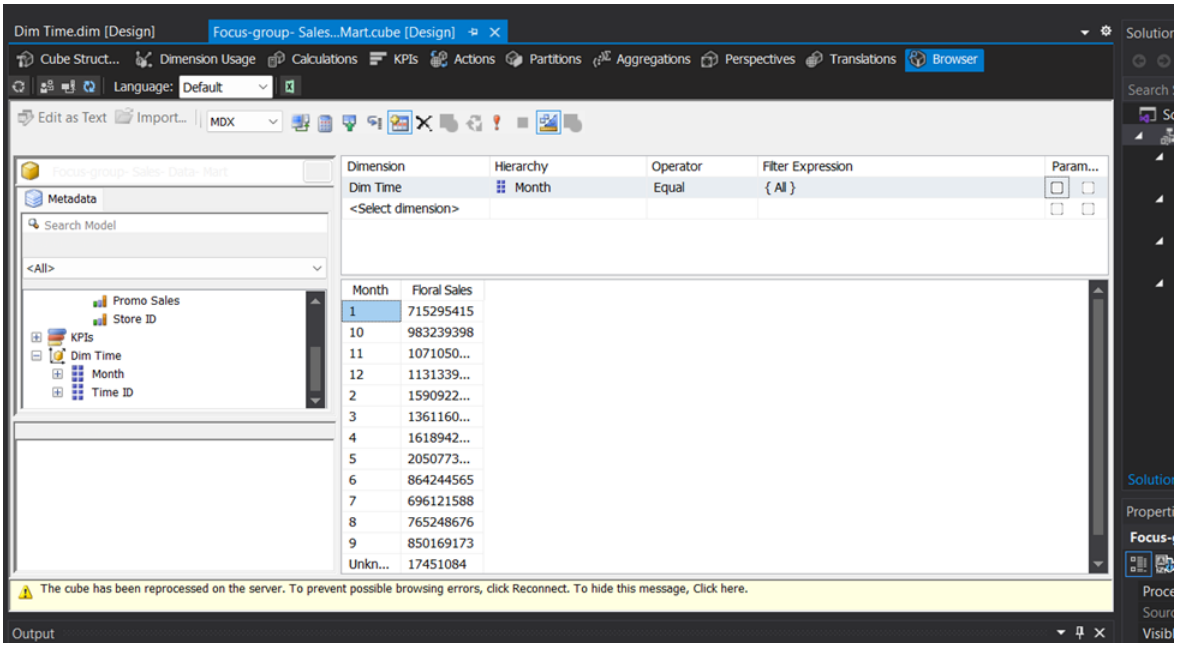


Figure: The final report has been configured by including the relevant dimensions and measures from the respective tables.

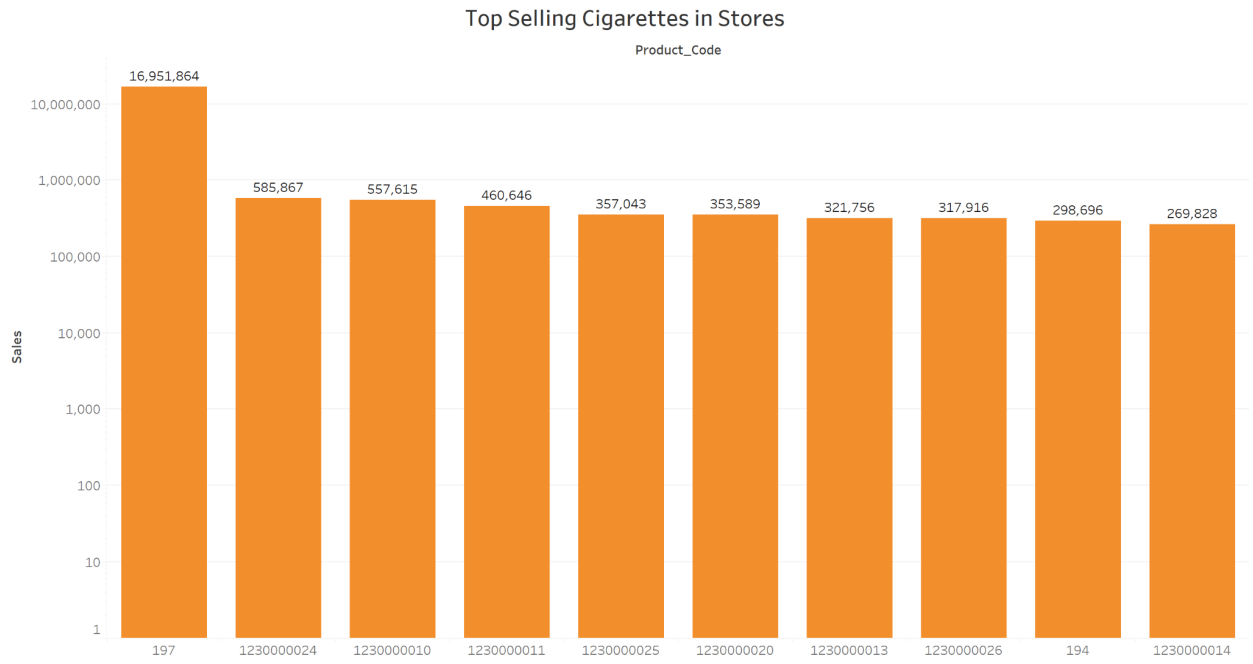
ISTM 637 - Group 5 Report

Business Question 4

What are the top-selling Cigarette SKUs in the Cigarette Product Category for DFF?

BI/Visualization Method: Report using Tableau

Report:



Rationale:

Dominick's Finer Foods will benefit from detailed insights into their cigarette sales performance, enabling optimized inventory management and targeted marketing. This comprehensive understanding will enhance customer service and guide strategic decision-making, thereby improving overall business efficiency and profitability.

Reporting Plan:

The given business question was analyzed using Tableau given the straightforward nature of the dimensions and measures involved in the question. There are two dimensions involved – product_code and sales. Primarily, the user would want to view the top selling cigarette SKUs and see a sum of their overall sales in stores.

Implementation Steps:

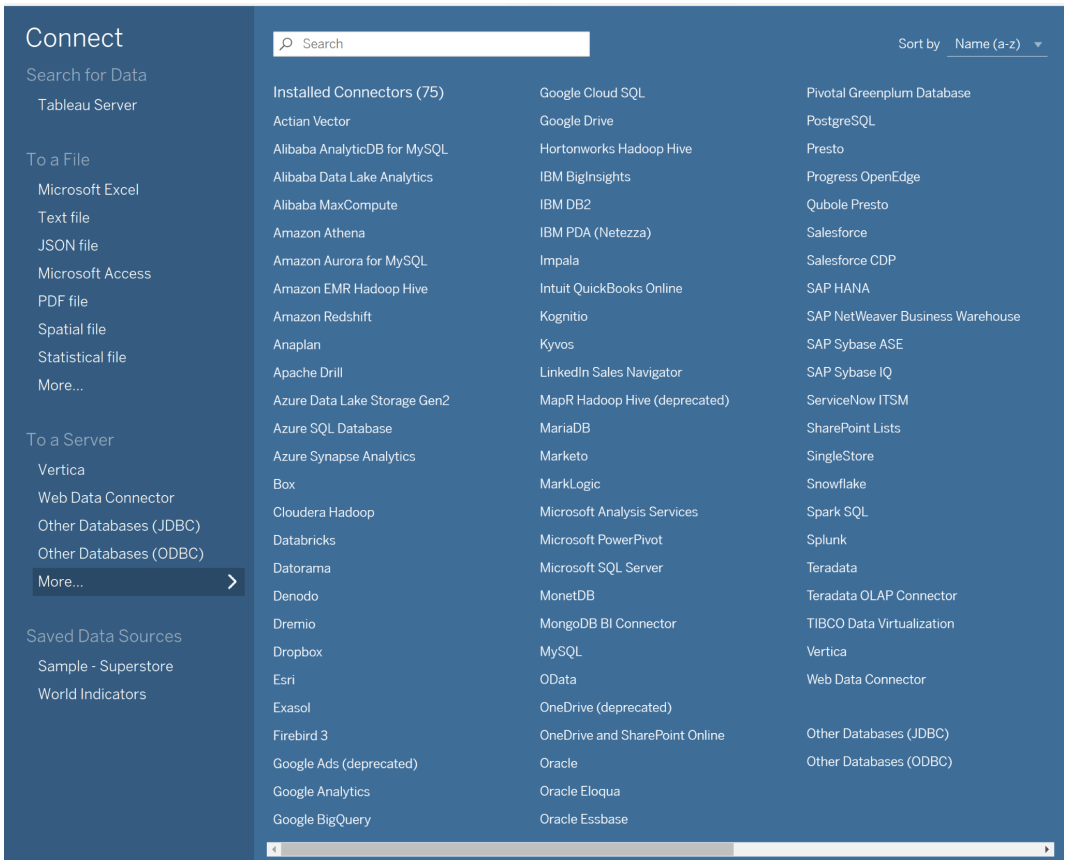


Figure: Opening Tableau and selecting data source

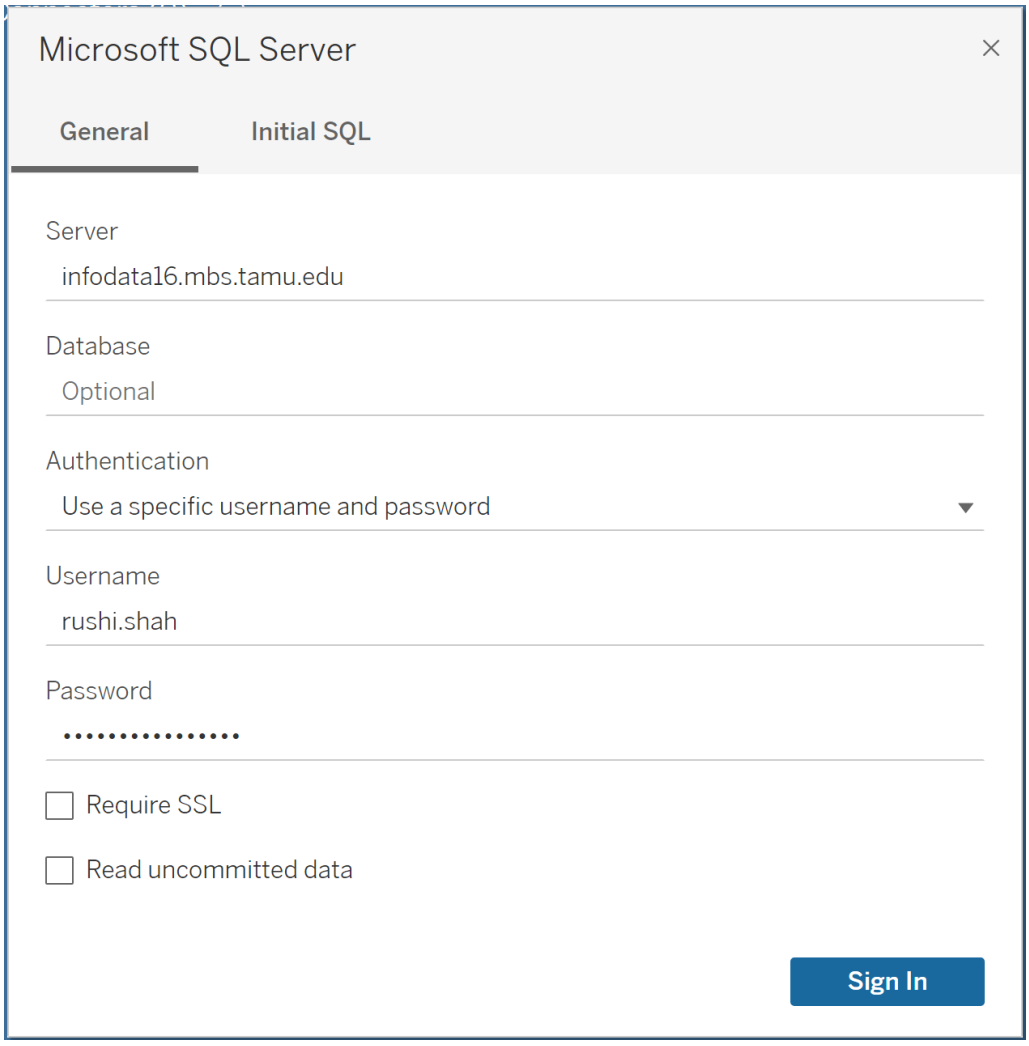


Figure: Selecting Microsoft SQL Server and entering credentials to make a connection with the Data Warehouse

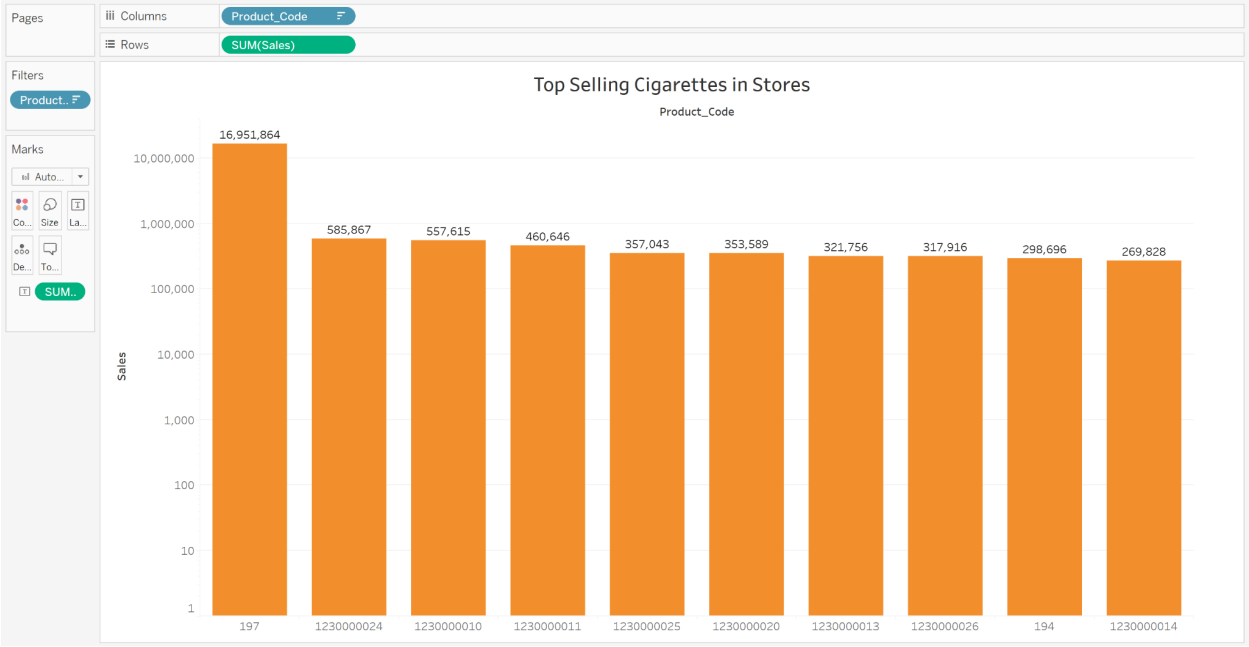


Figure: Once we have all the data, we created a visualization to fetch further insights and analyze top selling cigarette products in stores.

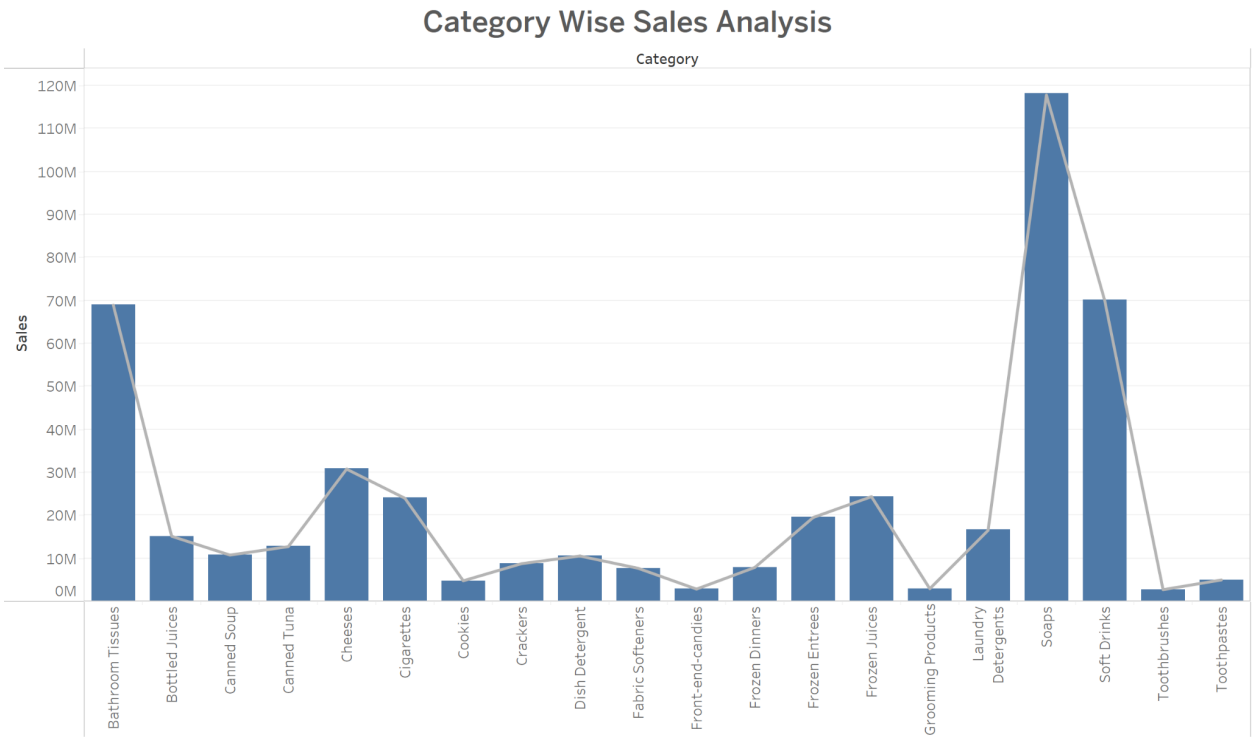
ISTM 637 - Group 5 Report

Business Question 5

What are the various possibilities to maximize earnings based on high sales categories?

BI/Visualization Method: Report using Tableau

Report:



Rationale: We wanted to find out categorize various product categories into high volume high sale, low volume high sale and so on categories. This would give us a great understanding into which categories to analyze further helping out DFF to strategize on what to work on. Figuring out high sales categories will help them improve their sales and marketing strategies thereby boosting their total revenue.

Reporting Plan:

The given business question was analyzed using Tableau given the straightforward nature of the dimensions and measures involved in the question. There are two dimensions involved –

category and sales. Primarily, the user would want to view the top selling categories and see a sum of their overall sales in stores.

Implementation Steps:

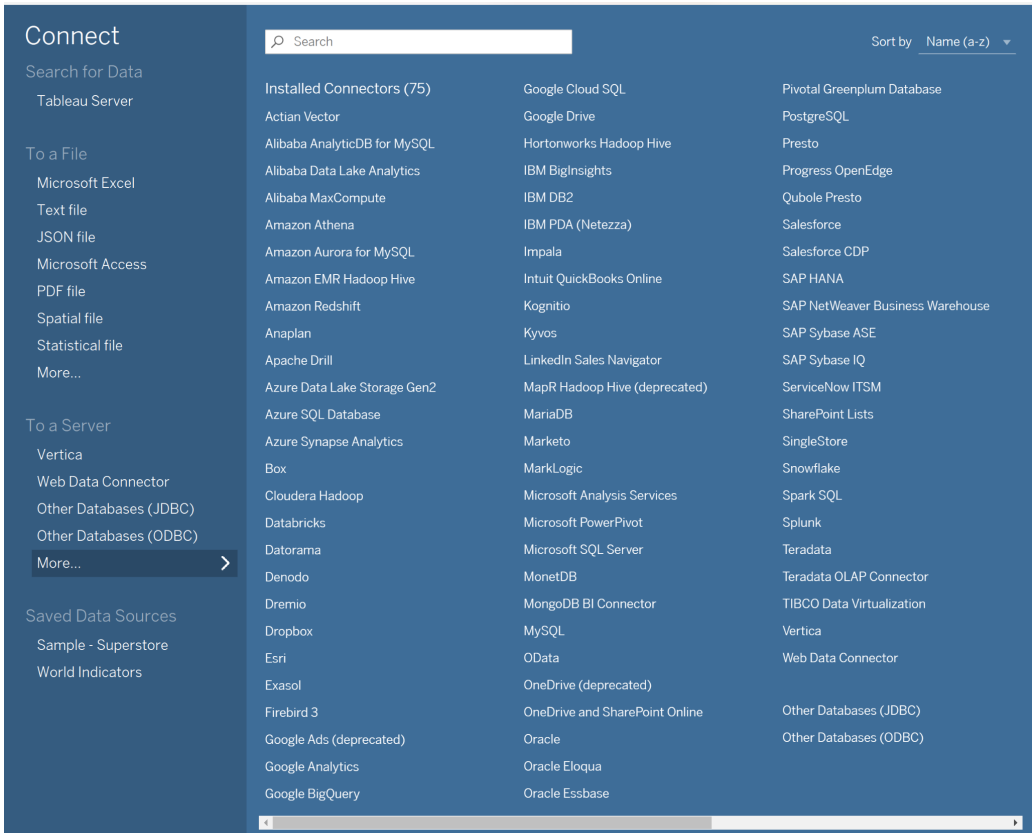


Figure: Opening Tableau and selecting data source

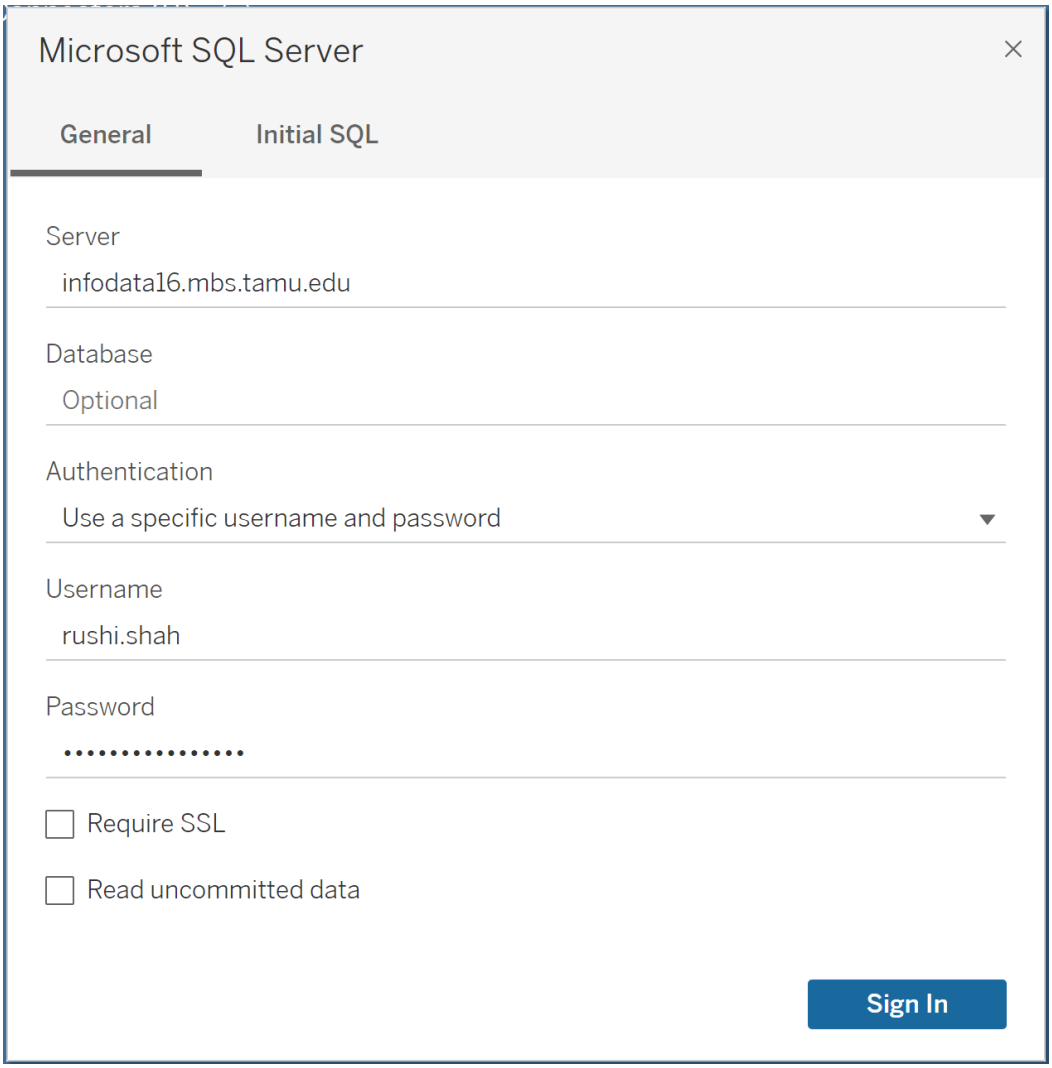


Figure: Selecting Microsoft SQL Server and entering credentials to make a connection with our Data Warehouse

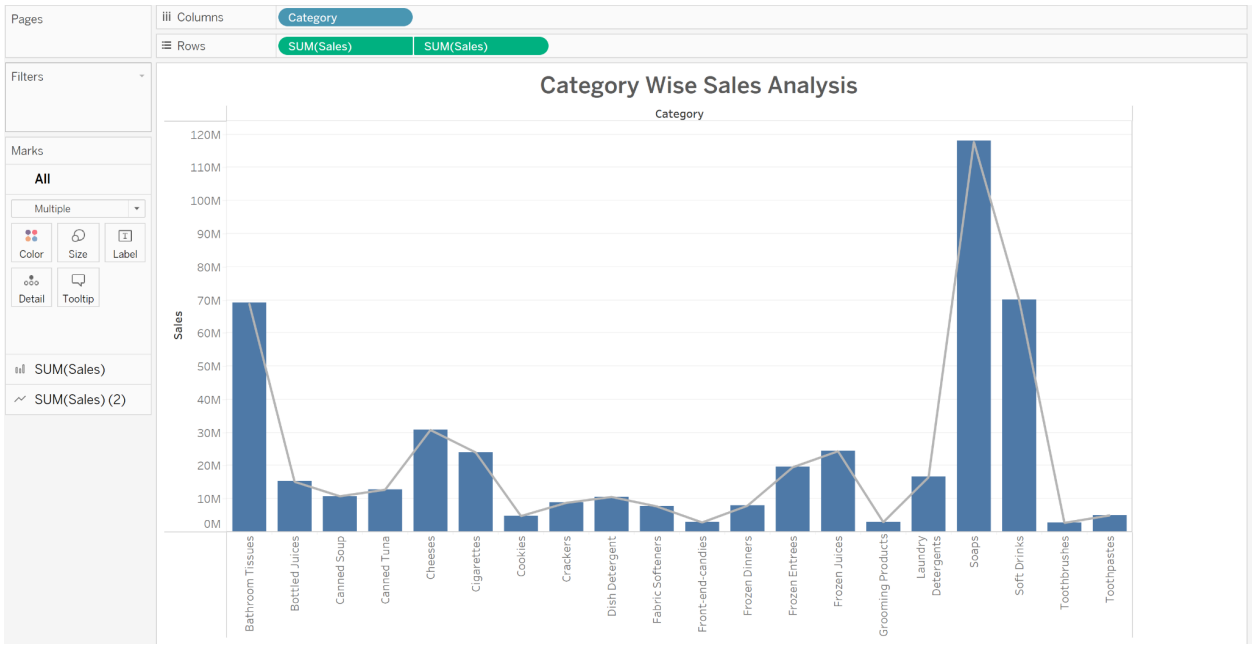


Figure: Once we have all the data, we created a visualization to fetch further insights and analyze category wise sales.

REFERENCES

- https://www.javatpoint.com/ssrs
- https://www.guru99.com/ssas-tutorial.html
- https://www.analyticsvidhya.com/blog/2017/07/data-visualisation-made-easy/
- https://help.tableau.com/current/pro/desktop/en-us/examples_sqlserver.htm
- https://www.mssqltips.com/sqlservertip/2704/developing-a-ssrs-report-using-a-ssas-data-source/