Dominick's Finer Foods

DESIGN AND IMPLEMENTATION OF A DATA WAREHOUSE FOR A RETAIL STORE

REPORT 4: BI REPORT DESIGN AND IMPLEMENTATION FOR DFF

TOTALLY INTEGRATED TERM REPORT

ISTM 637-603 Group-9
POOJA CHOUREY, SAI SHIVA GANESAN, SUSHANT PATIL
EMAIL: saishivaganesan@tamu.edu

"An Aggie does not lie, cheat or steal or tolerate those who do."

Credentials for accessing the DW, staging, reports and analysis services

SQL Server Details:

Username: *****

Password: *****

Staging Area: **<603><Group-9>staging-area**

Data Warehouse Area: **<603><Group-9>-dw-area**

SSAS cube:

- **603-Group9-SSAS-Q3**
- **603-Group9-SSRSwithSSAS-Q4**

Folder on [redacted] :

- **603-Group9-SSRS-Q1**
- **603-Group9-SSRS-Q2**
- **603-Group9-SSRSandSSAS-Q4**

Table of Contents

1. Introduction

Dominick's Finer Food (DFF) was a retail chain based out of Chicago. DFF had a great presence across the Chicago area. In 1998, they had close to 120 stores and a very decent market share across the retail space in the Chicago area. Dominick's Finer Food entered an agreement with Chicago Booth to conduct store-level research into shelf management and pricing of retail products. As such, we now have store-level data on different products, their categories, the demographics in which they were sold, etc. for 9 years from 1988-1997.

Our task was to clean and analyze this dataset to come up with business questions that provide us insights for maximizing Dominick's Finer Foods sales. Following were the challenges that befell us while we were analyzing the aforementioned dataset:

- The datasheets CCOUNT, MOVEMENT sheets, STORE, UPC sheets were not cleaned. They had blank or special character values that did not conform with other data.
- The sales data of different categories present in CCOUNT was not aggregated to a desirable level for ex: everyday data was provided in the dataset and for a monthly or yearly analysis, data had to be manually filtered.
- Store data which had zones and price tiers, was not mapped to different product categories. We had to merge data from different sheets to proceed further with our analysis.
- Important data related to store locations, zones, price tier had to be extracted from a file of different file format.
- In conclusion, data across UPC, MOVEMENT, CCOUNT, STORES, DEMOGRAPHICS is not integrated. So pitting product categories against each other and analyzing their performance in profits/sales proved to be difficult.

2. Data Description

2.1 Understanding of the Data

A large set of data has been given for DFF. The data is available at various level of details. In some files, the data has daily sales values and in some files the sales data is aggregated on weekly basis. One file has sales data on a broad variety of products whereas for some categories of products, there are separate files available. The details around special weeks is also available to us. Furthermore, we have a file containing data

about the DFF store locations along with the zone and price tier in which they fall. The individual files have been described below:

Customer Count File

This file contains data on daily sales of different products. These products are being sold across the various stores of DFF. A daily count of customers visiting a specific store is also available. The data contains information on the coupons getting redeemed at a store as well. We also have the week number to which that specific date belongs.

Movement Files

We have movement files based on categories. There is a file for each of the product category. These files follow a similar format which includes details around every UPC of a product sold for over 5 years in every store on weekly basis. The information on number of units sold in the stores, the price of each of the product and the profit earned on every sale is also available.

Dominick Store Locations

This table includes store number, city, zip code and address of the different stores of DFF. Along with these details, we have the price tier to which that store belongs. The different price tiers are: CubFighter, High, Low and Medium. Also, we have the zone in which a store belongs. The zone numbers start from 1 and range till 16.

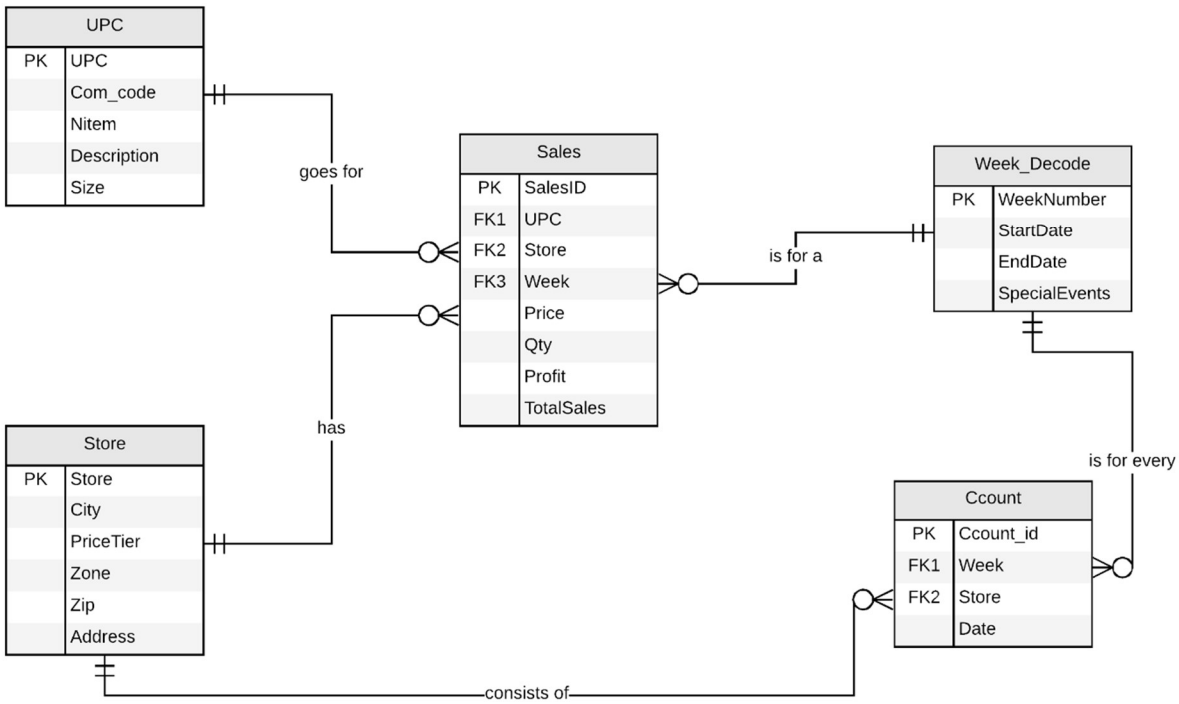
Week Decode Table

This table gives us information on the week number associated with the special occasions such as Halloween, New Year, Thanksgiving etc. starting from the year 1989 till 1997. The start date and end date of each of the weeks is also present in this table.

UPC Files

These files contain data on UPC belonging to each category. The data includes UPC number, product name, description, size etc.

2.2 Entity – Relationship Diagram



2.3 Metadata for OLTP source files

Ccount File

This file contains information on the number of customers visiting DFF stores on a daily basis and the coupons getting used on every sale.

Variable	Description	Type	Length
Date	Date of the Observation	Character	6
Week	Week Number	Numeric	8
Store	Store Code	Numeric	8
BAKCOUP	Bakery Coupons Redeemed	Numeric	8
BAKERY	Bakery Sales in Dollars	Numeric	8
BEER	Beer Sales in Dollars	Numeric	8
BOTTLE	Bottle Sales in Dollars	Numeric	8
BULK	Bulk Sales in Dollars	Numeric	8

BULKCOUP	Bulk Coupons Redeemed	Numeric	8
CAMERA	Camera Sales in Dollars	Numeric	8
CHEESE	Cheese Sales in Dollars	Numeric	8
CONVFOOD	Conventional Foods Sales in Dollars	Numeric	8
COSMCOUP	Cosmetics Coupons Redeemed	Numeric	8
COSMETIC	Cosmetics Sales in Dollars	Numeric	8
CUSTCOUN	Customer Count	Numeric	8
DAIRCOUP	Dairy Coupons Redeemed	Numeric	8
DAIRY	Dairy Sales in Dollars	Numeric	8
DELI	Deli Sales in Dollars	Numeric	8
DELICOUP	Deli Coupons Redeemed	Numeric	8
DELIEXPR	Deli Express Sales in Dollars	Numeric	8
DELISELF	Deli Self Service Sales in Dollars	Numeric	8
FISH	Fish Sales in Dollars	Numeric	8
FISHCOUP	Fish Coupons Redeemed	Numeric	8
FLORAL	Floral Sales in Dollars	Numeric	8
FLORCOUP	Floral Coupons Redeemed	Numeric	8
FROZCOUP	Frozen Items Coupons Redeemed	Numeric	8
FROZEN	Frozen Items Sales	Numeric	8
FTGCCOUP	Food-to-Go Coupons Redeemed	Numeric	8
FTGCHIN	Food-to-Go Chinese Sales in Dollars	Numeric	8
FTGICOUP	Food-to-Go Coupons Redeemed	Numeric	8
FTGITAL	Food-to-Go Italian Sales in Dollars	Numeric	8
GM	General Merchandise Sales in Dollars	Numeric	8
GMCOUP	General Coupons Redeemed	Numeric	8
GROCCOUP	Grocery Coupons Redeemed	Numeric	8

GROCERY	Grocery Sales in Dollars	Numeric	8
HABA	Health and Beauty Aids Sales in Dollars	Numeric	8
HABACOUP	Health and Beauty Aids Coupons Redeemed	Numeric	8
JEWELRY	Jewelry Sales in Dollars	Numeric	8
LIQCOUP	Liquor Coupons Redeemed	Numeric	8
MANCOUP	Manufacturer Coupons Redeemed	Numeric	8
MEAT	Meat Sales in Dollars	Numeric	8
MEATCOUP	Meat Coupons Redeemed	Numeric	8
MEATFROZ	Meat-Frozen Sales in Dollars	Numeric	8
MISCSCP	Misc. Coupons Redeemed	Numeric	8
MVPCLUB	MVP	Numeric	8
PHARCOUP	Pharmacy Coupons Redeemed	Numeric	8
PHARMACY	Pharmacy Sales in Dollars	Numeric	8
PHOTCOUP	Photo Coupons Redeemed	Numeric	8
PHOTOFIN	Photo	Numeric	8
PRODCOUP	Produce Coupons Redeemed	Numeric	8
PRODUCE	Produce Sales in Dollars	Numeric	8
PROMCOUP	Promotion Coupons Redeemed	Numeric	8
PROMO	Promotion Sales in Dollars	Numeric	8
SALADBAR	Salad Bar Sales in Dollars	Numeric	8
SALCOUP	Salad Coupons Redeemed	Numeric	8
SPIRITS	Spirits Sales in Dollars	Numeric	8
SSDELICP	Self Service Deli Sales in Dollars	Numeric	8
VIDCOUP	Video Coupons Redeemed	Numeric	8
VIDEO	Video Sales in Dollars	Numeric	8

VIDEOREN	Video Rentals (Dollar Amounts)	Numeric	8
WINE	Wine Sales in Dollars	Numeric	8

UPC Files

These files contain information on all the UPCs in a product category. The files follow a naming convention of UPCxxx where xxx is the category acronym.

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

Movement Files

The movement files have data on weekly sales of all the UPCs in a product category. The data is available for over 5 years.

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

Dominick's Stores

This table has data on Dominick's stores: store number, city, price tier, zone, zip code and address. There are 4 different price tiers and 16 zones.

Variable	Description	Type	Length
Store Number	Store Number	Numeric	3
City	City in which store is located	Character	20
Price Tier	Price Tier of store	Character	10
Zone	Zone to which store belongs	Numeric	3
Zip Code	Zip Code of the area	Numeric	5
Address	Address of the store	Character	30

Week's Decode Table

This is a SAS file which contains week number, start date, end date and the special occasion in that week like Halloween, Thanksgiving, Christmas etc.

Variable	Description	Type	Length
Week number	Week Number	Numeric	3
Start date	Start date of week	Character	8
End date	End date of week	Character	8
Special occasion	Special occasion	Character	15

2.4 Domain Understanding

Toro-González, Daniel, Jill J. McCluskey, and Ron C. Mittelhammer. "Beer Snobs do Exist: Estimation of Beer Demand by Type." Journal of Agricultural and Resource Economics (JARE) (2014).

This paper talks about the demand in beer by category using the Dominick finer foods dataset available in Chicago booth business website. The author makes use of the Dominick scanner data that covers up to 9 years of beer purchases across Dominick stores in Chicago to arrive at his conclusions. The three beer categories that the paper focuses on are – mass produced beers, crafted beers and imported beers. Based on the observations from the supermarket scanner data, some of the conclusions made are as follows.

Different category beers, although they are of the same product, one category of beer can never be a substitute for the other. For example, a person who has developed a taste for craft beer would rather not drink than drink a mass produced or imported beer. Packaging of beers also was found to be different across every category. From the analyzed data, mass

produced beers seemed to have the maximum market share followed by crafted and imported beers. Average income across households located near these stores also seemed to have played a major role towards purchase of a specific category of beer. The author finally inferred that beer was considered to be a normal good that was highly insensitive to price changes. This specially did hold true for mass produced beers.

Macé, Sandrine. "The Impact and Determinants of Nine-Ending Pricing in Grocery Retailing." *Journal of Retailing* 88.1 (2012)

Author Sandrine Mace tried to validate the much talked about effects of prices ending with the digit "9" on grocery sales in the retail space in this paper. The author's research validates the impact of Nine-Ending prices on grocery retailing but results also show that this impact is context dependent. The author also finds that Nin-Ending prices can have a detrimental
effect on sales for premium brands whereas it can be more effective in increasing sales for small/medium size brands that belong to a weaker market (Less budget share, less prices).
It was found out in the research that impact of Nine-Ending prices varies according to store location, clientele income, number of working women that visit the store, items, categories, etc.

The traditional belief is that prices that end in "9" have a positive impact on the sales of that product or good. However, educated people possess the cognitive ability to truly see the economic advantage a price tag offers them. If they don't see any value, they won't buy that product. In contrast, working women tend to round down prices because they have time constraints. So, there is better chance of working women buying products with prices ending in higher 9. Families with low income have budget and price constraints. Hence, they scrutinize prices more stringently which in turn nullifies the effect of pricing strategies. Their decisions are not influenced by Nine-Ending prices and so stores which attract lower income families should dial down on their Nine-Ending prices. Lastly, older people tend to rely more on level as their ability to process information(price) is low. In such scenarios, higher Nine ending prices can affect the sales positively.

Huang, Tao, Robert Fildes, and Didier Soopramanien. "The Value of Competitive Information in Forecasting FMCG Retail Product Sales and the Variable Selection Problem." European Journal of Operational Research (2014).

Sales forecasting is important for retail businesses to manage their inventory. The authors have proposed new methods which they believe would prove to be more effective in

forecasting retail UPC sales. Forecasting will be accomplished by incorporating competitive information including prices and promotions. Forecasting results were generally poor before because of large number of competitive explanatory variables. The authors explain that under such circumstances, time series models can easily become over-fitted. As such, it will forecast poorly. The authors propose a method which will have two stages. In the first stage, we will have to refine competitive information i.e. we identify the most relevant explanatory variables using variable selection methods, or pool information across all variables through factor analysis to construct small number of diffusion indexes. In the later stage, we will have to specify Autoregressive Distributed Lag (ADL) following a modelling strategy with the explanatory variables and constructed diffusion indexes.

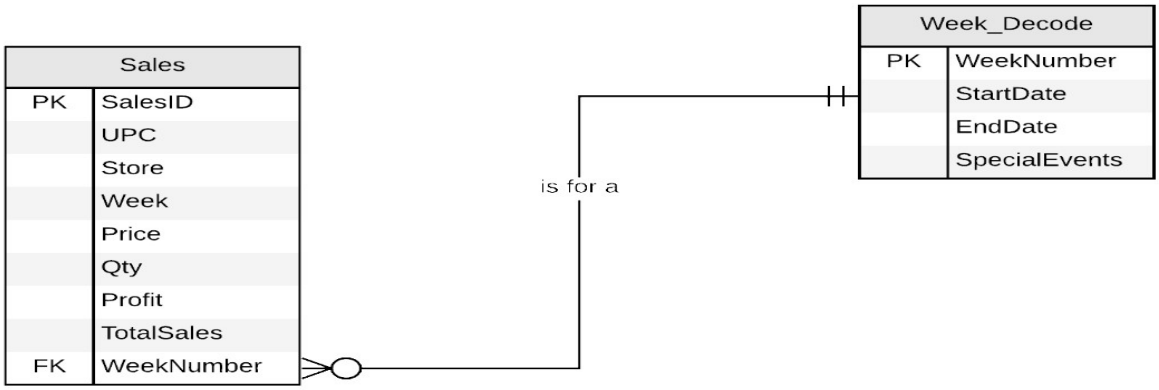
The ADL and ADL-DI model proposed by the authors significantly outperform the models constructed with price and promotion information. The improvements in forecasting become more substantial as the forecast horizon increases. However, there remains a scope for improvement. The authors have included majority of products in each category when implementing variable selection methods. This uncertainty can be reduced if short list of factors can be produced based on market knowledge of category managers. The authors have also not accounted for the effect of advertising. This situation can be improved by incorporating advertising information despite previous studies finding the effects of advertising to be fragile. The benefits of embracing this model building approach has paid off.

3. Business Questions

Question 1:

What has been the trend in sales of cookies over the years during thanksgiving?

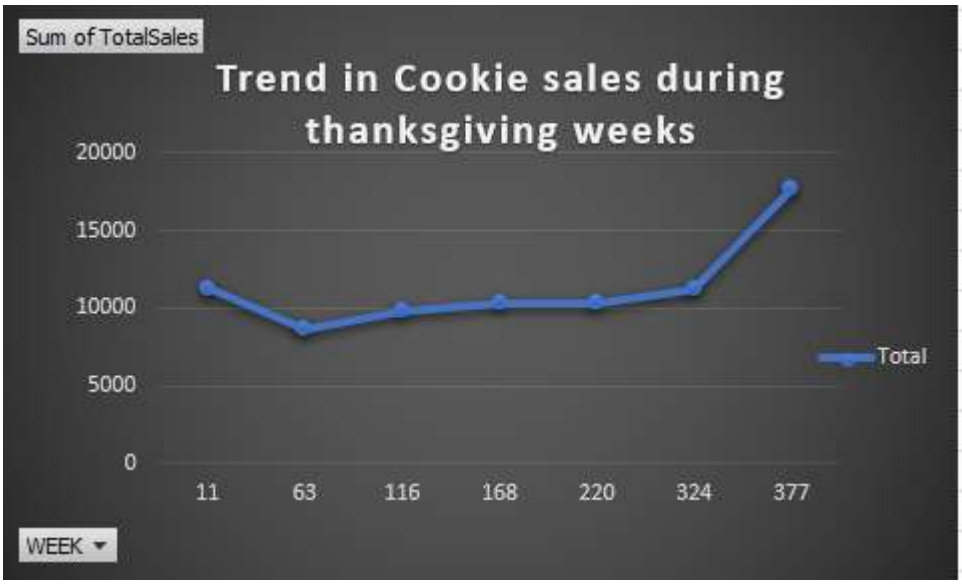
Entity-Relationship Diagram:



Pivot Table:

Row Labels	Sum of TotalSales
11	11343.76
63	8692.67
116	9838.42
168	10356.27
220	10318.92
324	11264.38
377	17704.84
Grand Total	**79519.26**

Pivot Chart:



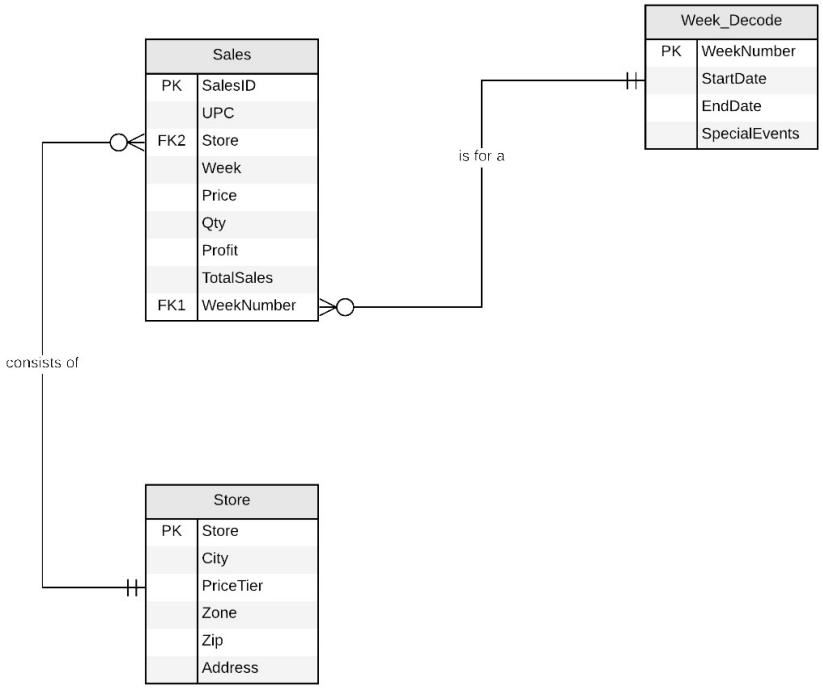
Justification:

Cookies are one of the most sought-after food items during the time of Thanksgiving. The above business question would help DFF in finding the trend in cookie sales across the last few years. This graph shows total cookies sales amount from all the stores of DFF during the Thanksgiving week. From the graph, we can see that there has been a considerable hike in cookie sales during the last year. The strategy behind this improvement can be identified and applied across other food products to improve their thanksgiving sales.

Question 2:

Which zone had the maximum number of beer sales during New Year week of 1997?

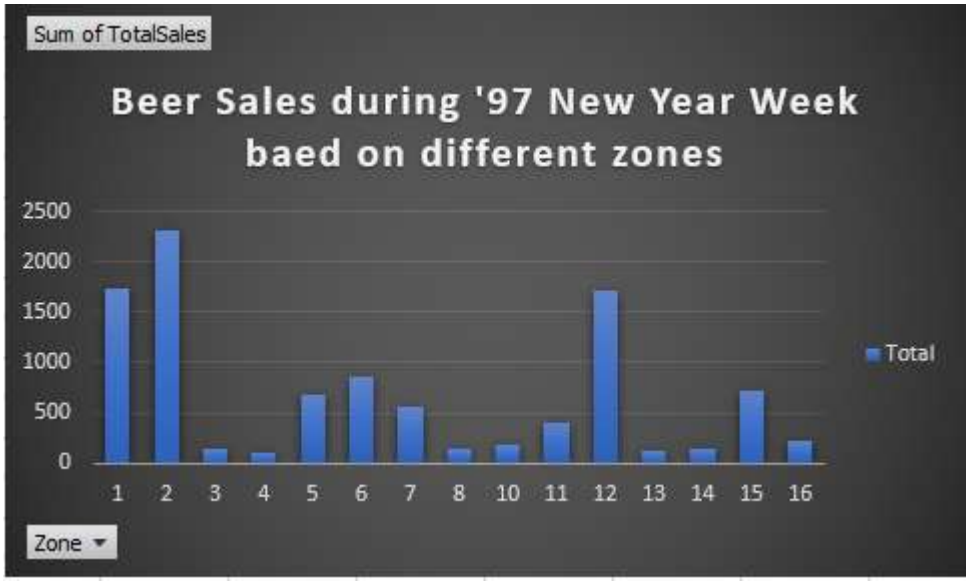
Entity-Relationship Diagram:



Pivot Table:

Row Labels	Sum of TotalSales
1	1725.96
2	2311.51
3	136.66
4	100.74
5	684.66
6	857.57
7	562.11
8	145.02
10	190.14
11	405.66
12	1708.97
13	133.16
14	144.65
15	724.08
16	214
Grand Total	**10044.89**

Pivot Chart:



Justification:

During New Year week, it is important for any retail store to keep a track of their beer sales as it may sell the most during this time of the year. The above visualization shows the total beer sales across every zone of Dominick finer food in the year 1997. As we can see from the graph, zone 1 and 2 have seen the most money out of beer sales during the New Year week. Using this, further analysis can be carried out on how to promote beer sales across other zones.

Question 3:

What are the best performing foods through the years? On which days, the sales are higher for such products?

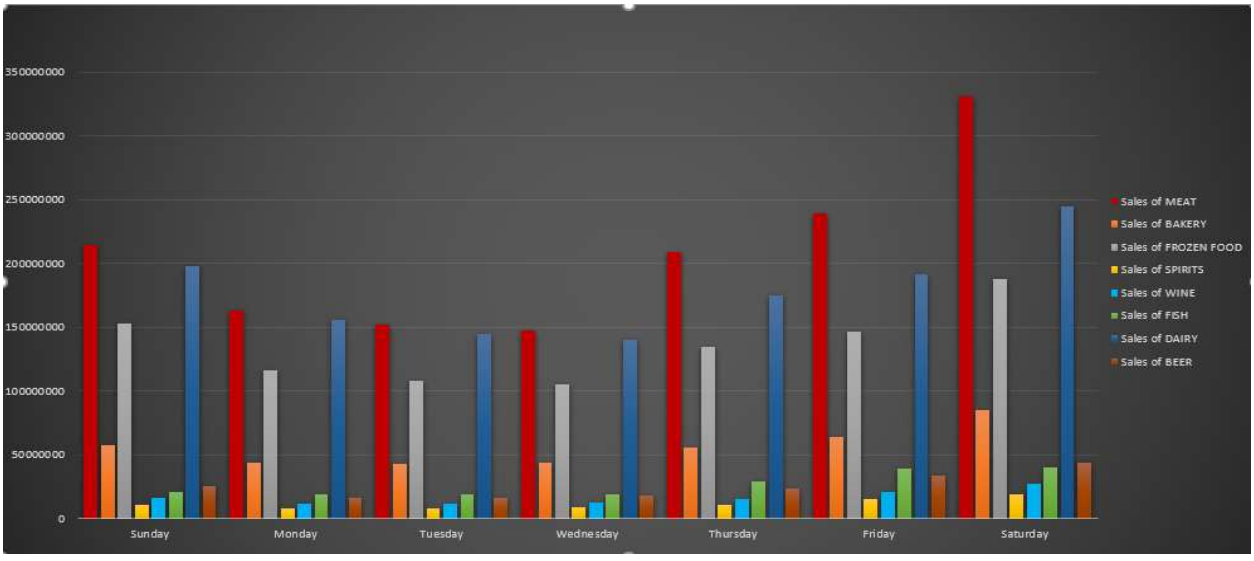
Entity-Relationship Diagram:

CCount	
PK	Ccount_id
	Date
	Store
	Week
	Cheese
	Meat
	Bakery
	Frozen
	Spirits
	Wine
	Fish
	Dairy
	Beer

Pivot Table:

Row Labels	Sales of MEAT	Sales of BAKERY	Sales of FROZEN F	Sales of SPIRITS	Sales of WINE	Sales of FISH	Sales of DAIRY	Sales of BEER
Sunday	214049077.9	57891857.74	152531445.4	10417953.61	15869472.65	21011097.38	198148554.5	25793436.92
Monday	162932844.8	43864555.3	116197751.3	7988505.25	11396728.57	19022552.97	155201687.1	16496354.81
Tuesday	151605022.9	43216878.34	108286179.1	8291044.86	11937689.11	19072171.2	144176718.2	16272820.6
Wednesday	147212339.5	44059343.34	105035900.4	8726563.23	12531626.26	19399985.39	139765868.6	17743565.59
Thursday	208493309.5	55944559.77	134140092.3	10916237.34	15719956.35	29101677.18	174419823.9	23132556.24
Friday	239342489.7	64044071.73	146576513.2	14842676.07	20653181.84	38961788.18	191367979.9	33407505.06
Saturday	331099974.3	84714273.2	187275265.9	18878832.34	27270045.12	40155356.5	244161385.6	43505805.92
Grand Total	**1454735059**	**393735539.4**	**950043147.8**	**80061812.7**	**115378699.9**	**186724628.8**	**1247242018**	**176352045.1**

Pivot Chart:



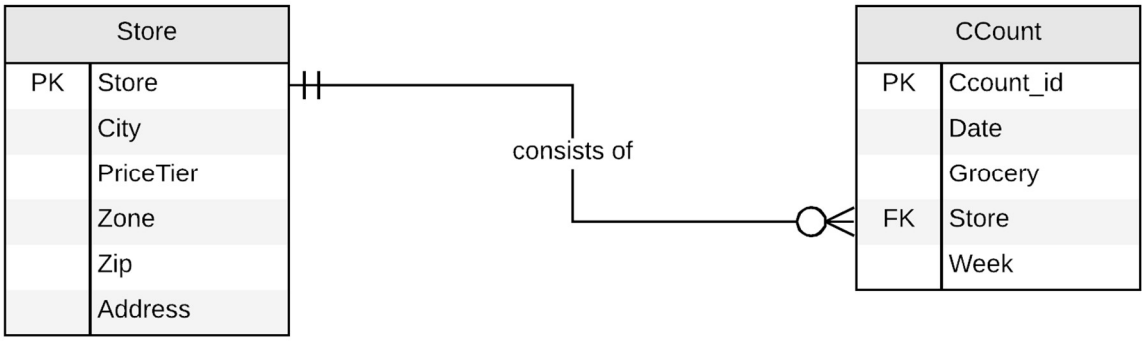
Justification:

Dominick's Finer Foods will want to know who their best performer is and when are that product's sales the highest. The data here clearly shows that DFF sells Meat the most and that too on weekends. As such, they can use this data to plan inventory refilling for meat so that they have enough stock for their target days of the week. This will only ensure higher sales for DFF.

Question 4:

How did grocery sales for cubfighter, low, medium, and high price tier fare for years 1988-1997?

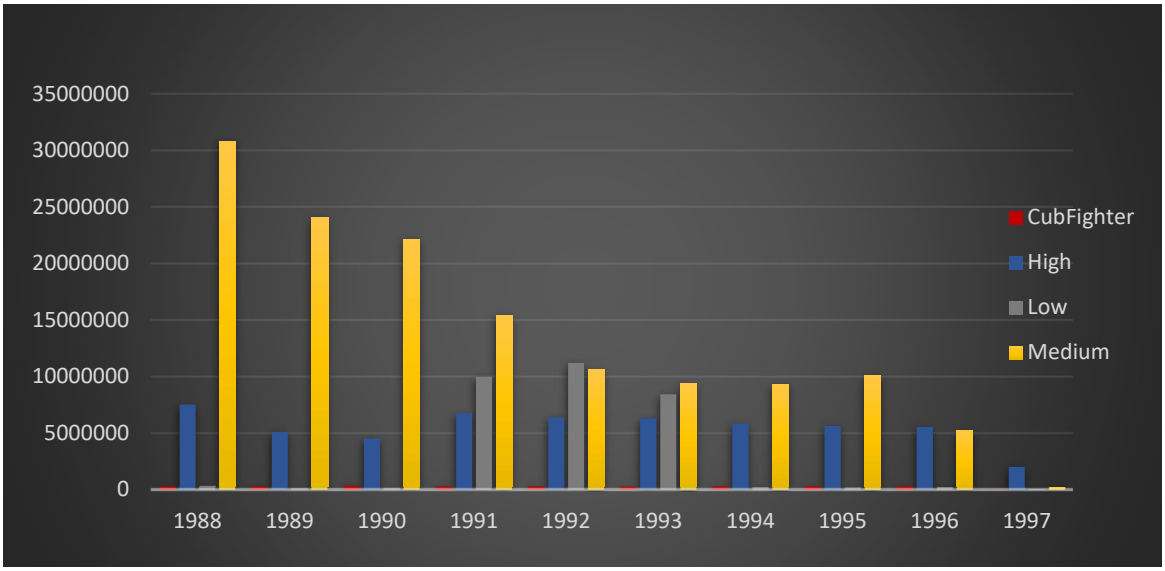
Entity-Relationship Diagram:



Pivot Table:

GROCERY SALES	Column Labels				
Row Labels	CubFighter	High	Low	Medium	Grand Total
1988	277958.16	7526345.11	336549	30867375.26	39008227.53
1989	291345.88	5096710.38	151186.65	24063752.75	29602995.66
1990	325035.88	4529635.42	179853.14	22143226.12	27177750.56
1991	268831.51	6754290.7	9993246.38	15437235.98	32453604.57
1992	276659.35	6388795.18	11216559.96	10604760.33	28486774.82
1993	240135.23	6309004.16	8434566.15	9396200.71	24379906.25
1994	270894.4	5852820.9	211281.84	9346457.28	15681454.42
1995	303896.54	5629297.75	196106.32	10123702.33	16253002.94
1996	294679.84	5542590.8	224305.21	5240704.38	11302280.23
1997	15673.98	1998464.46	63242.33	233116.64	2310497.41
Grand Total	**2565110.77**	**55627954.86**	**31006896.98**	**137456531.8**	**226656494.4**

Pivot Chart:



Justification:

Since Dominick Finer Foods are primarily a retail chain, keeping their grocery sales up is one of their main objectives. The pivot chart above clearly shows that Grocery Sales in general have been decreasing over the years among all price tiers. The best performing category is Medium price tier here. The management can look into this data and figure out a strategy to improve their grocery sales. While doing so, they can now specifically target some specific price tiers and employ strategies that work for Medium price tier.

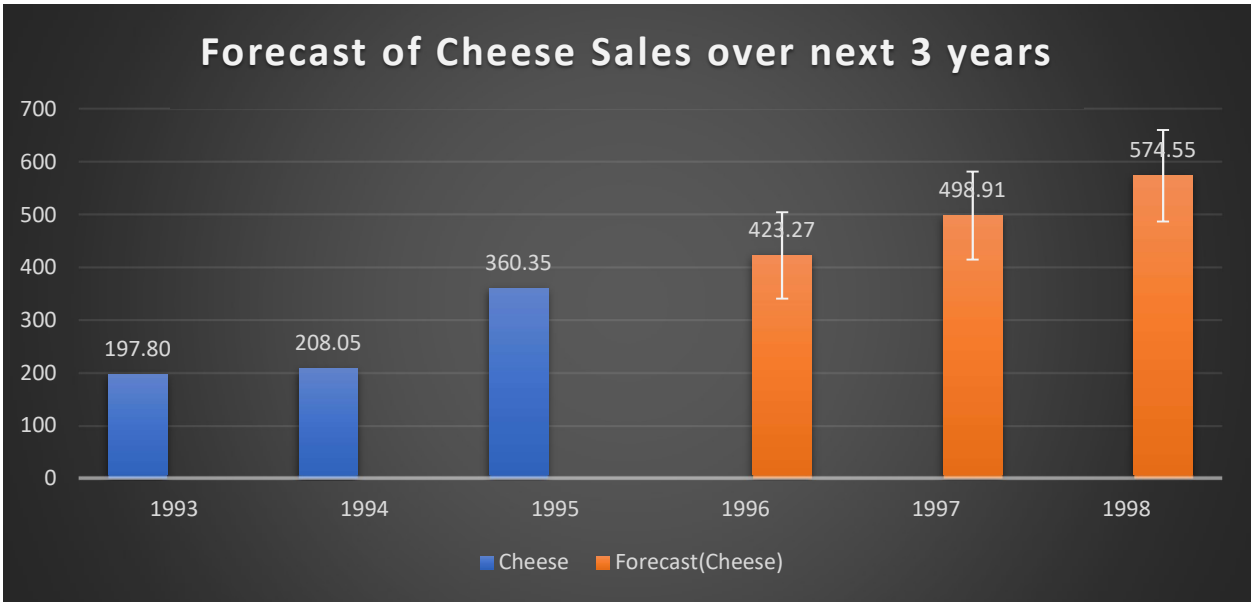
Question 5:

How will the cheese sales be over next three years based on the data available from past 3 years?

Entity-Relationship diagram:

CCount	
PK	Ccount_id
	Date
	Cheese
	Store
	Week

Pivot Chart:



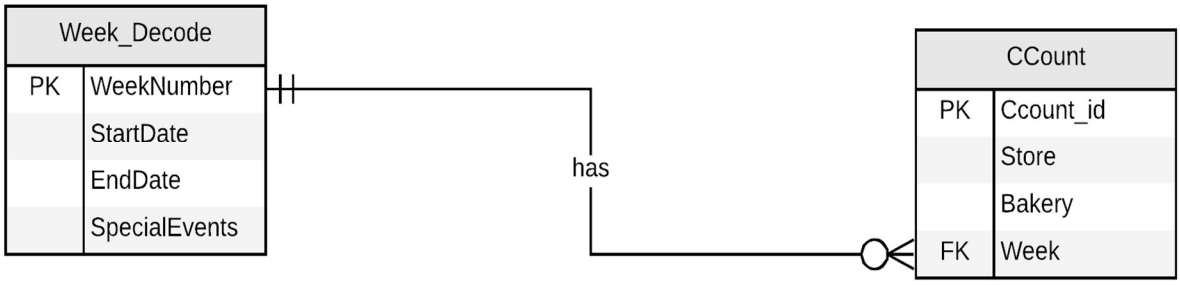
Justification:

The above plot shows how the sales of cheese would be over next three years based on the sales from last 3 years. Forecasting helps companies in predicting the sales nature of a product so that they can come up with new sales strategies, ultimately increasing the company's profit.

Question 6:

Compare the bakery sales during Easter and 4th of July 1996 across all the stores?

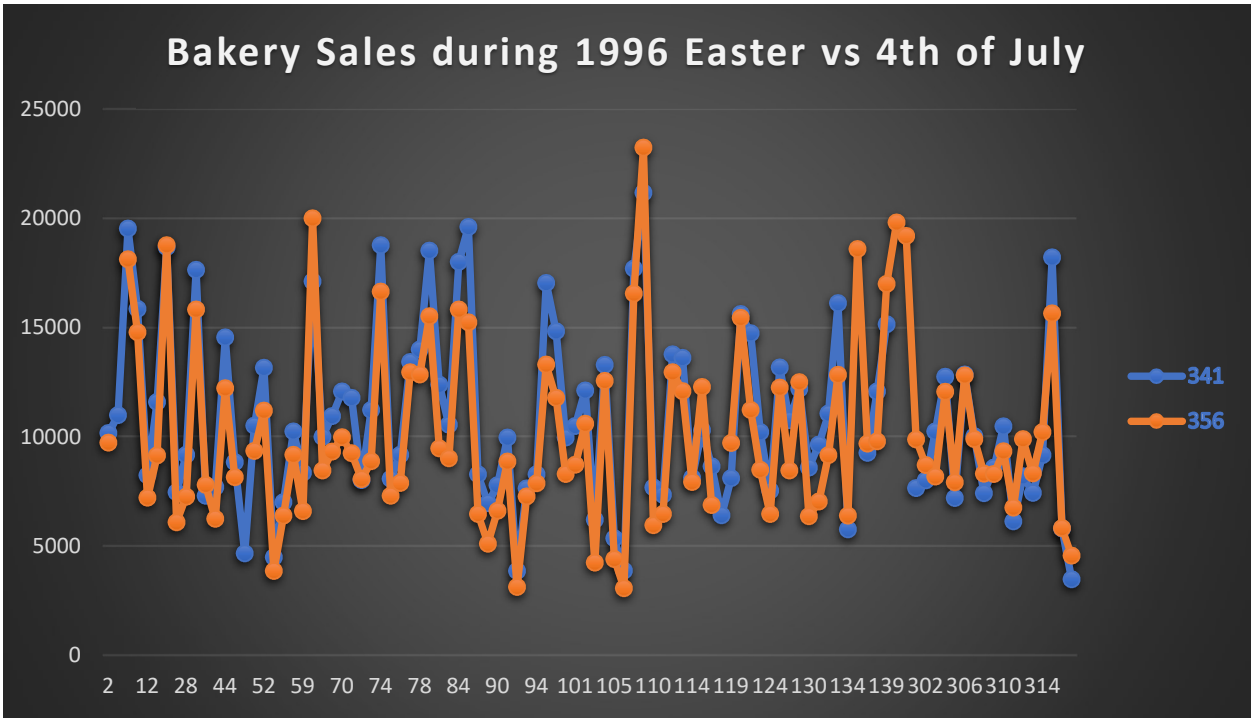
Entity-Relationship diagram:



Pivot Table:

Sum of Bakery	Column Labels		
Row Labels	341	356	Grand Total
2	10176.74	9735.89	19912.63
5	11001.41		11001.41
8	19546.77	18137.65	37684.42
9	15864.33	14791.91	30656.24
12	8244.52	7212.41	15456.93
14	11595.32	9146.15	20741.47
18	18688.86	18780.28	37469.14
21	7462.9	6087.57	13550.47
28	9177.05	7256.81	16433.86
32	17650.19	15833.65	33483.84
33	7302.14	7808.85	15110.99
40	7697.79	6252.45	13950.24
44	14555.44	12236.77	26792.21
47	8844.09	8157.82	17001.91
48	4662.44		4662.44
51	10509.7	9359.69	19869.39
52	13161.51	11200.36	24361.87
53	4496.03	3861.72	8357.75
54	7014.16	6396.67	13410.83

Pivot Chart:



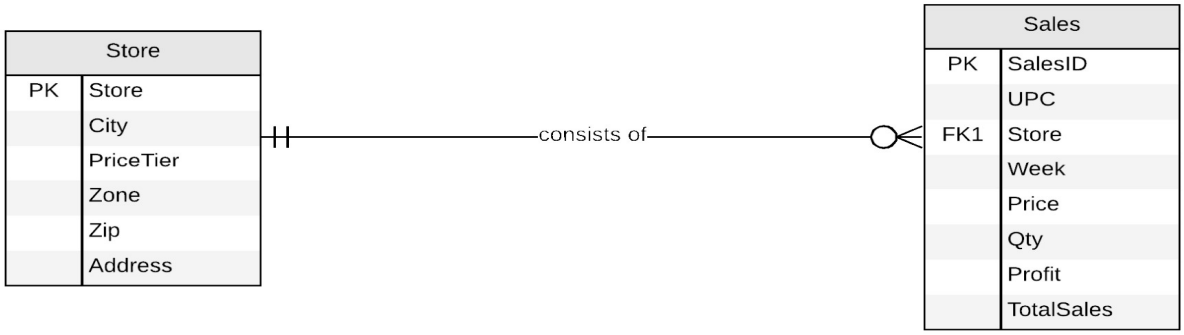
Justification:

Bakery sales are generally higher during special occasions. Here, we have a comparison of bakery sales across all the stores during two special occasions in the year 1996: Easter and 4th of July. This plot helps us in identifying the store which had maximum and minimum sales so that the following year some marketing strategy can be applied to improve the sales.

Question 7:

Which price tier has been performing the best in grooming category?

Entity-Relationship diagram:



Pivot Table:

Row Labels	Sum of TotalSales
CubFighter	320480.0217
Low	353176.8917
Medium	1361388.645
High	822404.105
Grand Total	**2857449.663**

Pivot Chart:

Justification:

Grooming products are something that are available across every price tier. Based on their price tier and quality, every grooming product may have its own strong set of users. From the graph, we can see that there is a significant difference in total sales of medium and high price tier products compared to cubfighter and low-price tier products. DFF can further use this analysis to round in on the worst performing UPC's in the lower two tiers and take a business decision on either improving the product sales or curbing the number of products available for sale.

Question 8:

Which products sell the most and the least at store number 75 on an average per day?

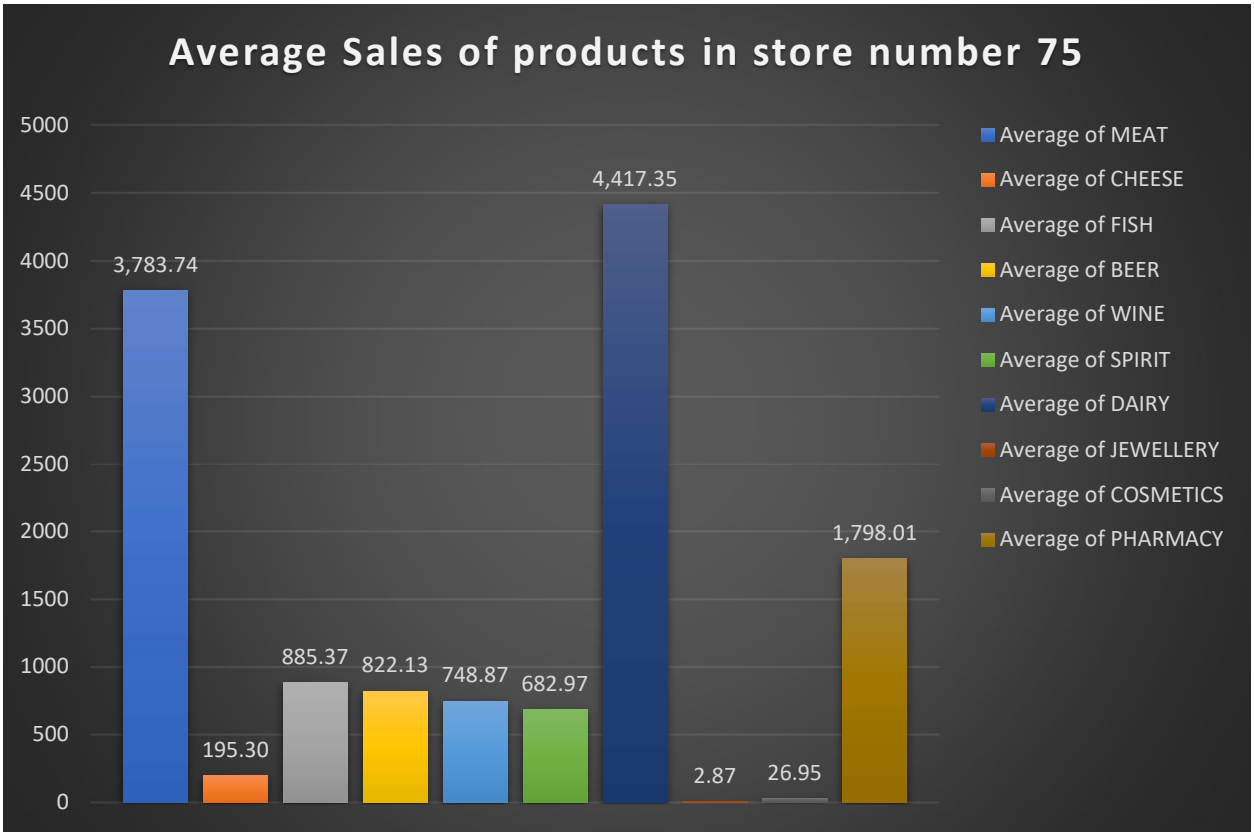
Entity-Relationship diagram:

CCount	
PK	Ccount_id
	Date
	Store
	Week
	Cheese
	Jewellery
	Pharmacy
	Cosmetics
	Spirit
	Wine
	Fish
	Dairy

Pivot Table:

Row Labels	Average of MEAT	Average of CHEESE	Average of FISH	Average of BEER	Average of WINE	Average of SPIRIT	Average of DAIRY	Average of JEWELLERY	Average of COSMETICS	Average of PHARMACY
75	3783.737168	195.2992344	885.3693077	822.1270037	748.8720696	682.9740623	4417.352674	2.868626374	26.94723077	1798.013582
Grand Total	**3783.737168**	**195.2992344**	**885.3693077**	**822.1270037**	**748.8720696**	**682.9740623**	**4417.352674**	**2.868626374**	**26.94723077**	**1798.013582**

Pivot Chart:



Justification:

The above plot shows how different products sell on an average at a certain store. It is useful in identifying the product which sells the most and the product which sells the least. Further analysis can be carried out on how the sales strategy has been deployed which will further aid in deciding whether to discontinue some of the UPCs of the least selling product or to change the market strategy for that product. Similarly, the more UPCs can be introduced for the most selling product. This type of data can be analyzed for other stores as well.

Question 9:

How much is the cigarette sale and profit for a specific UPC (1100000111) over all the years?

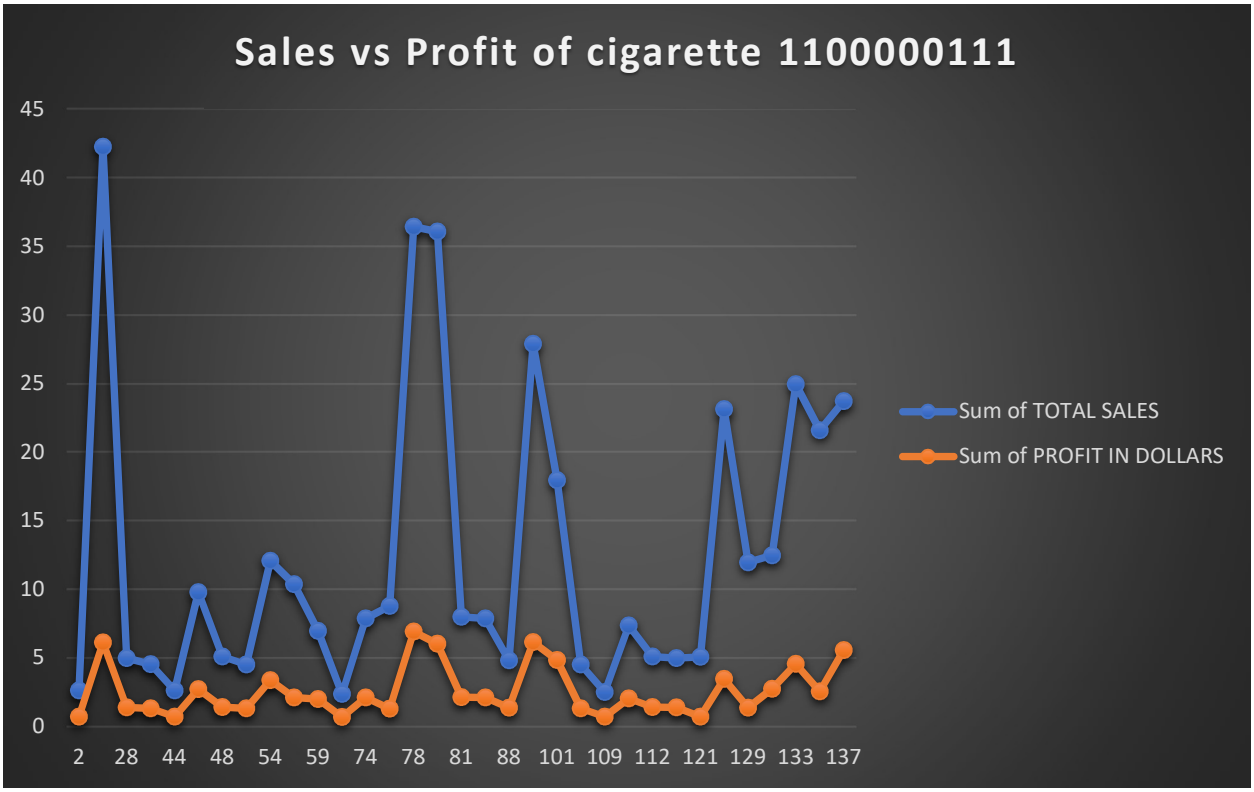
Entity-Relationship diagram:

Sales	
PK	SalesID
	UPC
	Store
	Week
	Price
	Qty
	Profit
	TotalSales

Pivot Table:

Row Labels	Sum of TOTAL SALES	Sum of PROFIT IN DOLLARS
2	2.59	0.6911
5	42.28	6.1036
28	4.96	1.3548
32	4.5	1.2888
44	2.59	0.6911
47	9.77	2.6887
48	5.08	1.37
51	4.49	1.2872
54	12.06	3.338
56	10.36	2.0733
59	6.93	1.9608
72	2.33	0.6566

Pivot Chart:



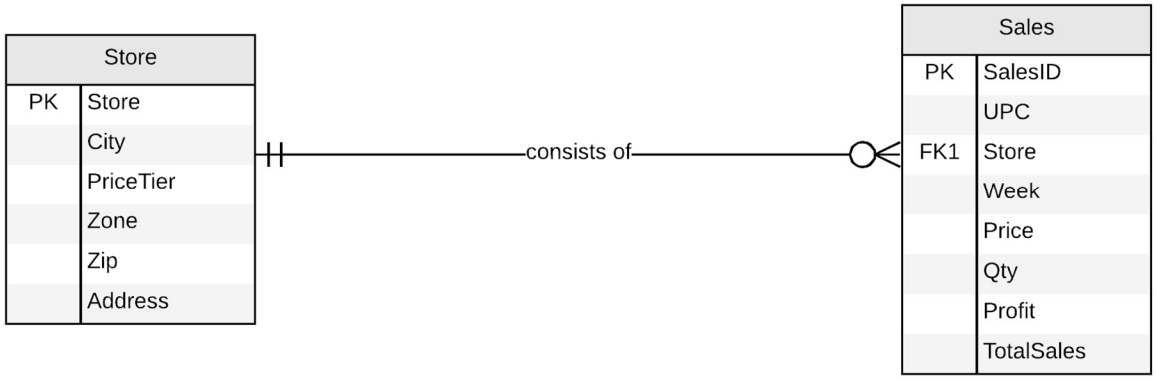
Justification:

The above plot shows the sales and profit for a specific UPC of cigarette. There are many UPCs under cigarette category and not all sell so well. This type of analysis helps in identifying whether this UPC, 1100000111, should be continued across stores or it should be withdrawn.

Question 10:

How is Analgesics performing in different Price Tiers and Zones?

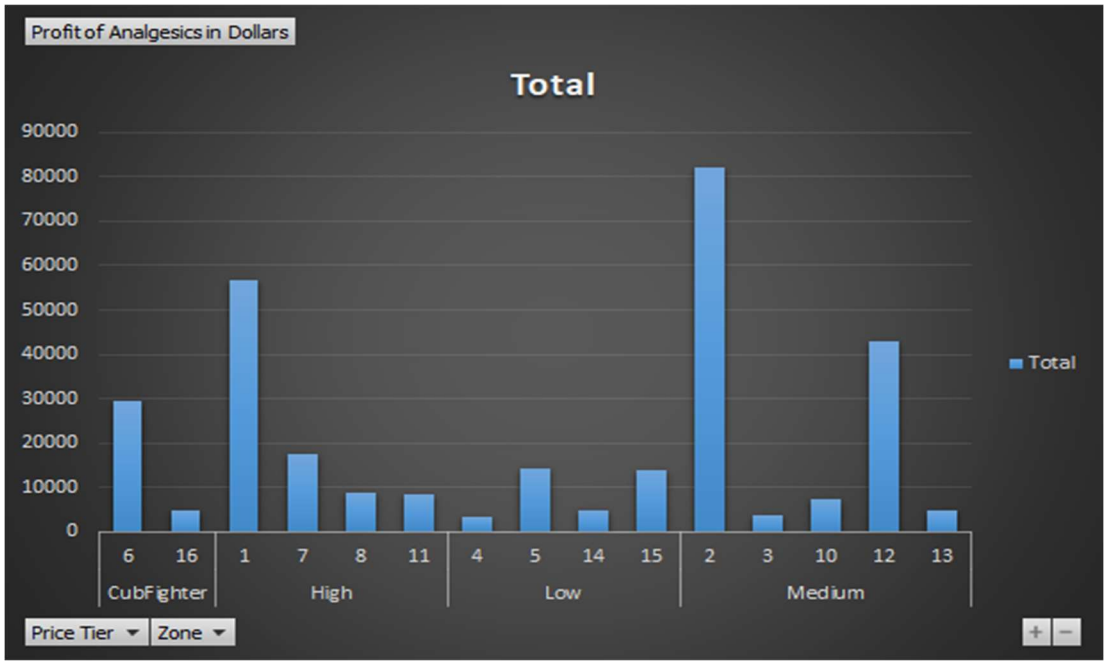
Entity-Relationship Diagram:



Pivot Table:

Row Labels	Profit of Analgesics in Dollars
CubFighter	**34595.2764**
6	29617.7423
16	4977.5341
High	**91797.1983**
1	56569.1161
7	17644.2293
8	9004.4584
11	8579.3945
Low	**36207.1725**
4	3314.9833
5	14221.0083
14	4892.2424
15	13778.9385
Medium	**140971.6902**
2	82168.0114
3	3729.1719
10	7419.1444
12	42856.78582
13	4798.5767
Grand Total	**303571.3374**

Pivot Chart:



Justification:

DFF can view the performance of any product type like Analgesics. They can determine how Analgesics is faring in different price tiers and different zones. It is safe to say that Analgesics performs best in the domain of Medium price tier and specifically in zone 2. DFF can apply strategy employed in zone 2 to maximize their sales for analgesics in other zones and price tiers.

4. Independent Data Marts design using Kimball's approach

We have used the dimensional modeling technique to create the star schemas for Dominick's Finer Foods data warehouse. We are following Kimball's approach to independent data marts. Below we describe the various dimension tables and fact tables created for answering our business questions.

4.1 Dimensional Modeling

4.1.1 Dimension matrix for Data Marts

Dimension / Data Mart	Dim_Time	Dim_Product	Dim_Store
Product Sales	Yes	Yes	Yes
Customer Count	Yes		Yes

Fig. Dimension matrix

4.1.2 Dimension Tables

There are three dimension tables in our design: Time dimension, Product dimension and Store dimension. All three tables have been briefly described below:

- *Dim_Time*

  The time dimension table stores the time specific attributes. It has reference data about the measures seen in the ccount fact table and the product sales fact table. Both the fact tables have been described in fact table section of this report.

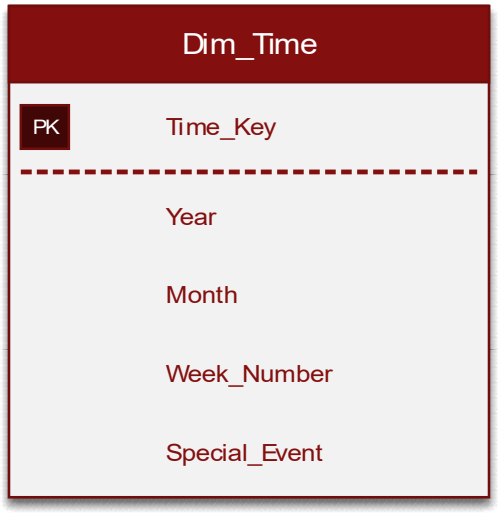


Fig. Time dimension table

The attributes of Dim_Time dimension table have been described as follows:

Time_Key: The surrogate key for unique identification in time dimension table.

Year: The identifier for the year in which sales have been made.

Month: The identifier for the month of the year in which sales have been made.

Week_Number: The identifier for the week of the year in which sales have been made.

Special_Event: The identifier for special events during the year.

- *Dim_Product*

The product dimension table has information about the various products available in the Dominick's Finer Foods retail stores. It has the reference data about measures seen in product sales fact table. The product sales fact table has been described in fact table section of this report.

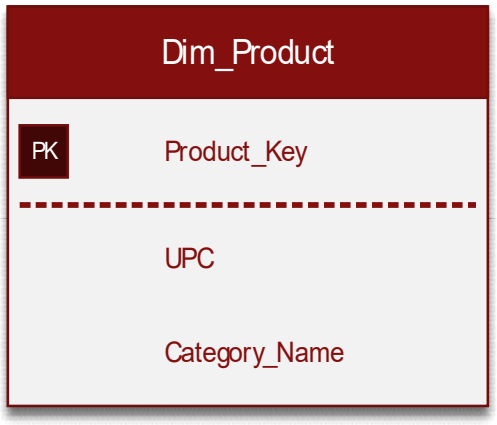


Fig. Product dimension table

The attributes of Dim_Product dimension table have been described as follows:

Product_Key: The surrogate key for unique identification in product dimension table.

UPC: The last five digits of the UPC number identify the product, the remaining digits identify the manufacturer.

Category_Name: Name of a category as given in Dominick's Finer Foods.

- *Dim_Store*

The store dimension table has information about the various stores of Dominick's Finer Foods retail store chain. It has reference data about the measures seen in the ccount fact table and the product sales fact table. Both the fact tables have been described in fact table section of this report.

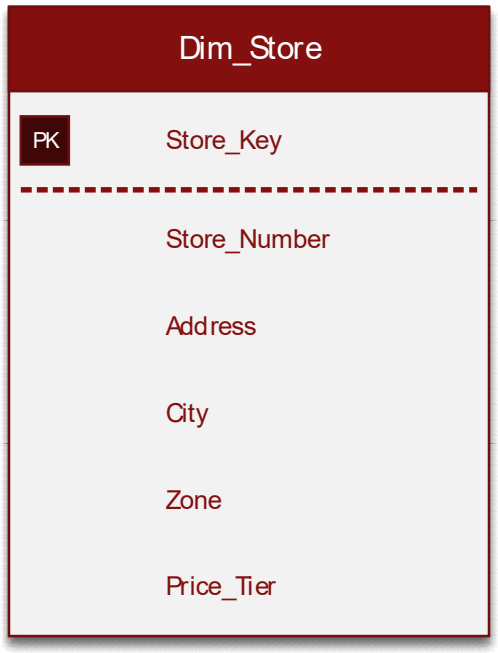


Fig. Store dimension table

The attributes of Dim_Product dimension table have been described as follows:

Store_Key: The surrogate key for unique identification in store dimension table.

Store_Number: The number assigned to a specific store.

Address: The address of a store.

City: The city in which a store is located.

Zone: The identifier for zone in which a store belongs.

Price_Tier: The price tier in which a zone falls where the store is located.

4.1.3 Fact Tables

There are two fact tables in our design: Product Sales fact table and Ccount fact table. All three tables have been briefly described below:

- *Product Sales Fact Table*
  This table contains the quantitative as well as aggregated information on sales of various products across the Dominick's Finer Foods stores during different time.

Fig. Product Sales fact table

Fact Table Keys:

Product_Key: The surrogate key for unique identification of products.

Store_Key: The surrogate key for unique identification of stores and aids in division of data by store like zone, price tier etc.

Time_Key: The surrogate key for unique identification of time and aids in division of data by time like for a year, during special event etc.

Fact Table Measures:

Qty: Size of the bundle

Move: The number of actual items sold

Price: The price of a bundle of the product

Total_Sales: This is a derived attribute and indicates the total dollar sales of the product across various dimensions.

*Total_Sales = Unit_Price * Unit_Sold / Quantity*

- *Ccount Fact Table*

  This table contains the quantitative as well as aggregated information on the sales of two specific products across the stores of Dominick's Finer Foods.

Fig. Ccount fact table

Fact Table Keys:

Time_Key: The surrogate key for unique identification of time and aids in division of data by time like for a year, during special event etc.

Store_Key: The surrogate key for unique identification of store and aids in division of data by store like store number, city etc.

Fact Table Measures:

Cheese_Sales: The dollar value of cheese sales across various dimensions.

Bakery_Sales: The dollar value of bakery sales across various dimensions.

4.2 Star Schema

The above created fact and dimension tables are integrated together to form the two star schemas that result in two data marts – Product sales data mart and Customer count data mart. Both these data marts are adequate to answer the business questions framed for the Dominick finer foods data warehouse project.

Product Sales Data Mart

The product sales data mart consists of the fact product sales and all three dimensions – store, product and time. This data mart is designed to answer business questions Q1 (trend of cookie sales), Q2 (Zone with maximum beer sales) and Q10 (Performance of Analgesics across different zones).

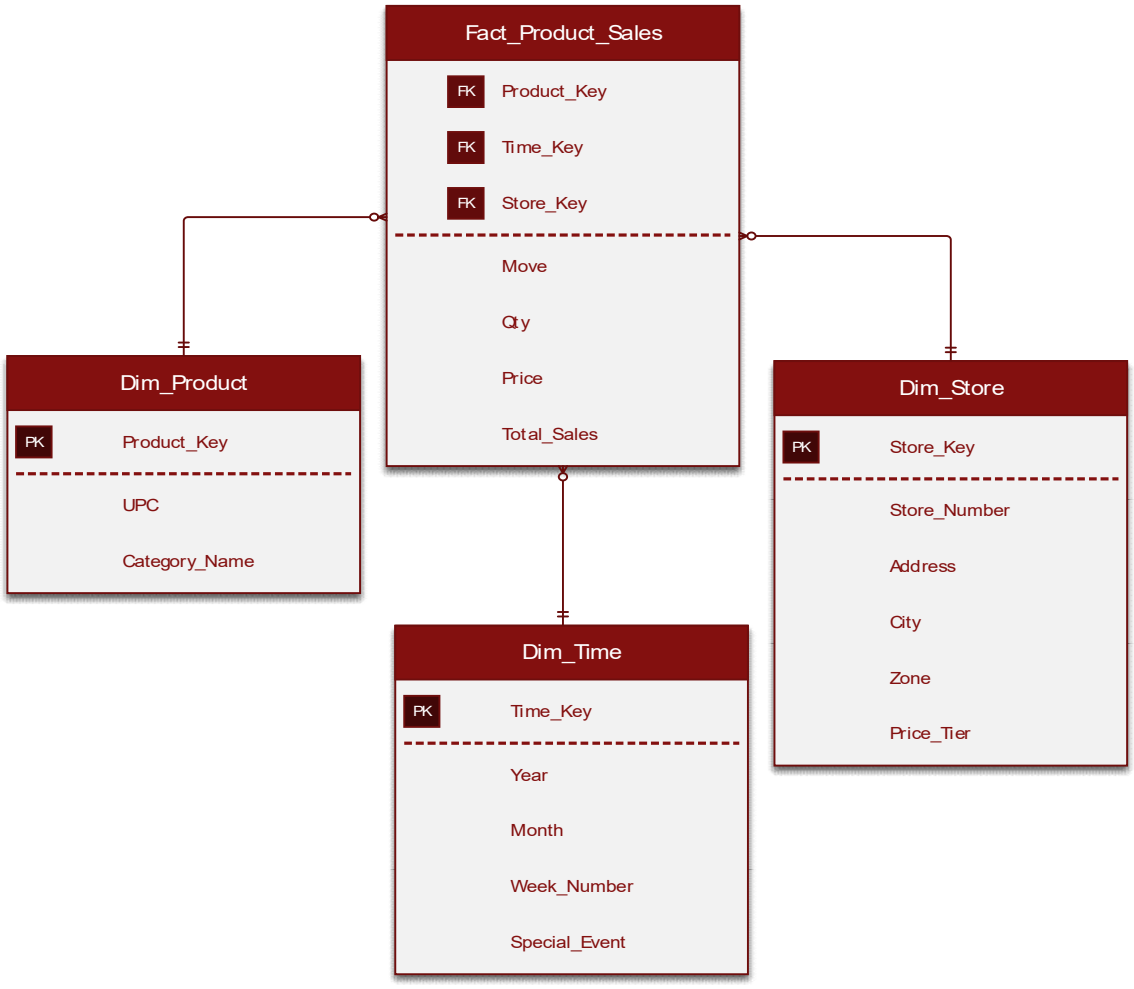


Fig. Product Sales Data Mart

Customer Count Data Mart

The customer count data mart comprises of the fact table (Fact_CCount) and two dimension tables – Dim_Store and Dim_Time. This data mart is used to answer business questions Q5 (Cheeses sales over the past three years) and Q6 (comparison of bakery sales across stores during Easter and 4th July).

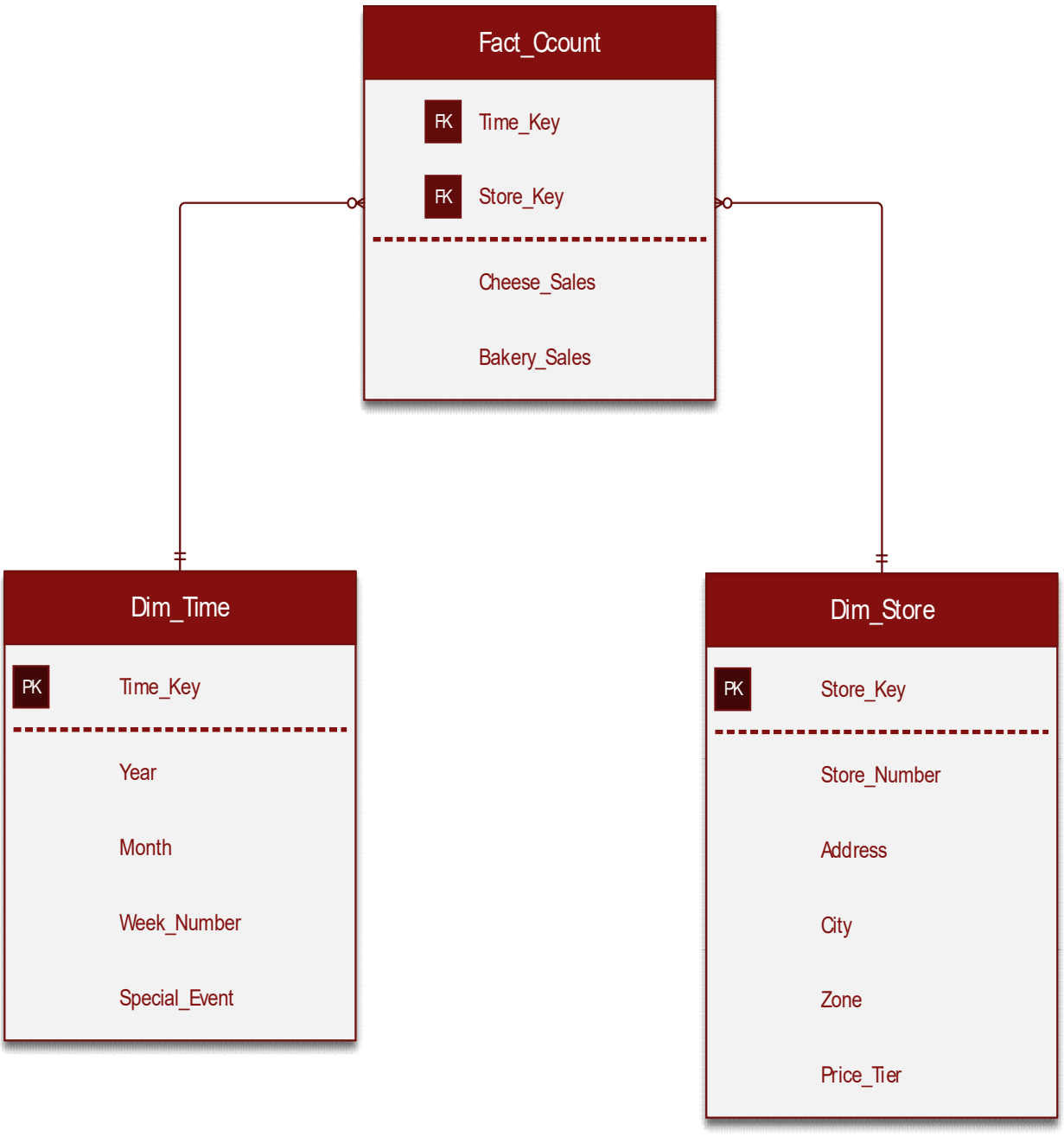


Fig. Customer Count Data Mart

Design justification for each business question will be provided in later sections of this report.

4.3 Mapping Tables

Mapping table for Product Sales Data Mart

Store Dimension

Source Table	Source Table Attribute	Mapping Function	DW Dimension Table	DW Dimension Attribute
Dominick's Store Demographics			Dim_Store	Store_Key
	Store	Copy from source		Store_Number
	Address	Copy from source		Address
	City	Copy from source		City
	Zone	Copy from source		Zone
	Price_Tier	Copy from source		Price_Tier

Time Dimension

Source Table	Source Table Attribute	Mapping Function	DW Dimension Table	DW Dimension Attribute
Dominicks Week Decode table			Dim_Time	Time_Key
	Start or end date	Derived from start or end date		Year
	Start or end date	Derived from start or end date		Month
	Week#	Copy from source		Week_Number
	Special Events	Copy from source		Special_Event

Product Dimension

Source Table	Source Table Attribute	Mapping Function	DW Dimension Table	DW Dimension Attribute
UPC Table			Dim_Product	Product_Key
	UPC	Copy from source		UPC
		Derived based on the category of the UPC file being loaded		Category_Name

Product Sales Fact

Source Table	Source Table Attribute	Mapping Function	DW Fact Table	DW Fact Attribute
Product Dimension	Product_Key	Foreign key from product dimension	Fact_Product_Sales	Product_Key
Store Dimension	Store_Key	Foreign key from store dimension		Store_Key
Time Dimension	Time_Key	Foreign key from time dimension		Time_Key
Movement table	Qty	Copy from source		Qty
	Move	Copy from source		Move
	Price	Copy from source		Price
		Total Sales = price*move/qty		Total_Sales

Mapping table for Customer count data mart

Store Dimension

Source Table	Source Table Attribute	Mapping Function	DW Dimension Table	DW Dimension Attribute
Dominick's Store Demographics			Dim_Store	Store_Key
	Store	Copy from source		Store_Number
	Address	Copy from source		Address
	City	Copy from source		City
	Zone	Copy from source		Zone
	Price_Tier	Copy from source		Price_Tier

Time Dimension

Source Table	Source Table Attribute	Mapping Function	DW Dimension Table	DW Dimension Attribute
Dominicks Week Decode table			Dim_Time	Time_Key
	Start or end date	Derived from start or end date		Year
	Start or end date	Derived from start or end date		Month
	Week#	Copy from source		Week_Number
	Special Events	Copy from source		Special_Event

Customer Count Fact

Source Table	Source Table Attribute	Mapping Function	DW Fact Table	DW Fact Attribute
Store Dimension	Store_Key	Foreign key from store dimension	Fact_Ccount	Store_Key
Time Dimension	Time_Key	Foreign key from time dimension		Time_Key
Ccount table	Cheese	Copy from source		Cheese_Sales
	Bakery	Copy from source		Bakery_Sales

4.4 Justification for Business Questions

Q.1 What has been the trend in sales of cookies over the years during thanksgiving week?

Ans. The question will help the organization to see how sales in cookies varies during the thanksgiving week over the years. There seems to be a surge in the sales of cookies during thanksgiving week across all DFF stores. The goal will be to identify an increase in sales during special occasions like Thanksgiving. As such, the management can plan promotional campaigns and inventory management to maximize the organization's revenue in such periods. The Product Sales Data Mart will help us to answer this question. Dimension tables like Dim_Store, Dim_Time, Dim_Product will help us map cheese sales across different stores, time(Weeks/Special Occasions) and products. The fact table Fact_Product_Sales will give us the necessary sales data(Quantity, Unit_Sold, Price, Sales).

Q.2 Which zone had the maximum beer sale for the new year week of 1997?

Ans. The question tries to identify the zone, in which Beer products did the best in new year week of 1997. This can help us identify a zone which is crucial for our Beer Sales. We can then identify reasons behind the success of beer sales in that zone and target other zones with the same practices to boost our beer sales in other regions as well. Furthermore, this question is important as beer or liquor sales go up during New Year's week and so it is important to track these sales figures during such periods so that the organization can maximize their sales on such occasions. This business question is answered by Product Sales Data Mart. The fact table Fact_Product_Sales will give us the total sales figure for beer products (derived from the dimension table Dim_Product). These figures will be mapped across stores and zones with respect to New Year's Week. The stores and zones will be derived from Dim_Store while the time period of New Year's week will be derived from the dimension table Dim_Time.

Q.5 How will the cheese sales be in the next 3 years based on the data of the past years?

Ans. Forecasting is a key to making to impactful business decisions before time. This question will help us forecast cheese sales for the next three years. These forecast results can then be considered as targets and as such, strategies can be devised to achieve better cheese sales. The Customer Count Data Mart will help us forecast cheese sales for the next three years. The fact table Fact_Ccount will give us the cheese sales details. These sales figures will be mapped across stores derived from dimension table Dim_Store and years which will be derived from the dimension table Dim_Time. We can then predict the trend in sales for the next three years based on the trend seen in the last three years.

Q.6 Compare the bakery sales during Easter and 4th of July for the year 1996 across all stores?

Ans. Bakery products do well during festive occasions. We aim to identify how bakery products perform on two such special occasions – the 4th of July and Easter. The sales data across all stores will be considered so that we can derive what works for certain regions and what doesn't. By taking a holistic look at the trends and strategies that are employed, we could then work on different campaigns or strategies to better these sales. The Customer Count Data Mart will help us answer this question. The fact table Ccount will give us the necessary Bakery sales data. This data will then be mapped across stores and zones which will derived from the dimension table Dim_Store. The dimension table Dim_Time will provide us the time period to identify Bakery sales on Easter and 4th of July.

Q.10 How is Analgesics performing in different price tiers and zones?

Ans. The purpose of this question is to identify how Analgesics faring in different price tiers and different zones. DFF can then look at the strategies applied for a price tier in a particular zone. As such, same strategies can be tailored to be applied in other price tiers and zones. This can help DFF increase their Analgesics sales across all price tiers and zones. The fact table Fact_Product_Sales will give us the sales data for analgesic products (derived from dimension table Dim_Product). This data will be mapped across price tier

and zones both of which are derived from the dimension table Dim_Store. The dimension table Dim_Time will help us arrange this data with respect to time (years).

4.5 Physical Design

Data aggregate plan:

Storing, maintaining, and accessing data in the most atomic form for large data sets can result in long delays. This will make operational transactions lengthy time wise. So, we have designed fact table that will have sum of Sales, Quantity, Units Sold and Price of Units. These aggregated values will help us answer our questions faster. Furthermore, we have also defined month and week, year and special events from the Week Decode table. We will then map these aggregated values across different time periods and geographies to analyze business trends.

Indexing plan:

To ensure that navigation through tables is fairly easy, we have used indexing. Indexing will help prevent scanning of entire table. We can simple use relations to give context to data and arrange it in the best way possible. We are using unique non-null attributes as Primary Keys in our dimension tables. The aggregated quantities will be mapped to these dimension tables with the help of these keys that will be present as foreign keys in our fact table.

Data standardization plan:

Since Dominic Finer Foods have lots of stores across the United States of America, it is safe to assume that DFF has to deal with humungous datasets. Storing data that is consistent and has integrity is the key to success in building a data warehouse. To accomplish this, we standardized the naming conventions of our tables. These standards are important as they give clarity for implementing design of a datawarehouse. We have used the convention Fact_<tableName> to indicate our fact tables. Whereas, Dimension tables are indicated by the convention DIM_<tableName>.

Storage plan:

Storage is a very crucial part of any data warehousing project and as such plays a very important role. We plan to put a great emphasis on data storage. We will derive the data from excel files and will send it to the staging area. Since the data is dirty, we will clean it to give it proper context. Following cleaning, we will store it in a temporary table. We will validate keys in dimension tables. We will convert the source product keys into data warehouse product keys. We will create DW tables during the staging process. We will then copy these tables into warehouse database. The process will contain indexing. Certain changes will be made to the initial storage plan to optimize the storage.

5. Data Cleaning and Integration

5.1 Data Quality Issues

Group	Quality	Issue Considered	Data Quality Problem in DFF
Relation to Data	Referential Integrity	Do records exist where expected? Do they contain unnecessary or inactive data? Are reference files/tables complete?	The DFF data was not of good quality because there were a lot of null values and some incomprehensible values. The columns which could have been considered as primary keys also had repetitive values
	Cardinality	Is the structure of relationships among entities and attributes maintained consistently?	The structure of relationships among entties and attributes were very inconsistent
Structure of Fields	Format	Do values follow consistent formatting standards?	The values do not follow a consistent format
	Standard	Are data elements consistently defined and understood?	The manual for the data provides a good explanation of data. However, there were some fields which were not easy to understand
	Consistent	Do values represent the same meaning across systems and files?	Data had same meaning across systems as per the explanation in the manual
Content within Data Values	Complete	Is all necessary data present?	Data is missing from some of the files, even for the business questions.
	Accurate	Does the data accurately represent reality or a verifiable source?	The data represents reality as it belongs to DFF. However, some of the values like negative values and null values in some places did not make any sense
	Valid	Do data values fall within acceptable ranges defined by the business?	The data present in store file had some of the store numbers which had not been given in the manual
	Fit for purpose	Is the information valuable to the business? Does the data convey information that can intelligently be consumed by the business?	The information is valuable to the business because it contains a lot of data on sales, products and customers

5.2 ETL Plan

ETL plan is a step by step guide for extracting, transforming and loading of data for data warehouses.

- Extraction is the processing of extracting data from sources like excel files and flat files.
- Transforming is the processing of transforming the data into something which will be readable, and which will adhere to our set standards. Transforming includes cleaning which eliminates null or invalid values from a data set.
- Loading is the process of transferring data from our staging area into our data warehouse after it is cleaned.

ETL is used to create data marts, data warehouses and to migrate data from one database to another database. Following is the ETL plan for data warehouse implementation.

5.2.1 Target Data needed in the Data Warehouse

Data Needed
Product Data for Analgesics, Cookies, Beer
Movement Data for Analgesics, Cookies, Beer
Total Sales Data for Cheese and Bakery
Start Date, End Date, Week, Special Events for Week
Store Data

5.2.2 Data Sources Used

Data Sources
CCount.csv
UPCANA.csv, UPCCOO.csv, UPCBER.csv
DONE-WANA.csv, DONE-WBER.csv, DONE-WCOO.csv
Week_Decode.csv
Store.csv

5.2.3 Mapping Tables

Mapping from Source to Staging Area

Source File Name	Source File Attributes	Staging Area table	Staging Area Attributes
Store.csv	store number	dbo.Store	Store
	City		City
	Price tier		Price tier
	Zone		Zone
	Zip		Zip Code
	Address		Address
Week_decode.csv	week number	dbo.Week	Week #
	start date		Start
	end date		End
	special events		Special Events
CCount.csv	Store	dbo.CCount	Store
	Date		Date
	Week		Week
	Bakery		Bakery
	Cheese		Cheese
UPCANA.csv, UPCBER.csv, UPCCOO.csv	UPC	dbo.Product	UPC

	Descrip		CATEGORY
DONE-ANA.csv, DONE-BER.csv, DONE-COO.csv	STORE	dbo.Movement_New_Staging	STORE
	UPC		UPC
	WEEK		WEEK
	MOVE		MOVE
	QTY		QTY
	PRICE		PRICE
	Total_Sales		Total_Sales

Mapping from Staging area to Data Warehouse area

Staging Area Table Name	Staging Area Attribute	Data Warehouse Table	Data Warehouse Table Attributes	Mapping Function
dbo.Week		dbo.Dim_Time	Time_Key	Surrogate Key
	Week #		Week	YEAR(start_date)
	Start		Month	MONTH(start_date)
	End		Year	
	Special Events		Special_Event	
dbo.Store		dbo.Dim_Store	Store_Key	Surrogate Key
	Store		Store_Number	
	City		City	
	Price tier		Price_tier	
	Zone		Zone	
	Zip Code		Zip_Code	
	Address		Address	
dbo.Product		dbo.Dim_Product	Product_Key	Surrogate key
	UPC		UPC	
	CATEGORY		CATEGORY	
dbo.Movement_New_Staging	MOVE PRICE QTY Total_Sales	dbo.Fact_product_sales	MOVE PRICE QTY Total_Sales	
dbo.Dim_Time	Time_Key		Time_Key	

dbo.Dim_Product	Product_Key		Product_Key	Composite Primary Key
dbo.Dim_Store	Store_Key		Store_Key	
dbo.AggCcount	Cheese_Sales Bakery_Sales	dbo.Fact_Ccount	Cheese_Sales Bakery_Sales	
dbo.Dim_Time	Time_Key		Time_Key	Composite Primary Key
dbo.Dim_Store	Store_Key		Store_Key	

5.2.4 Data Extraction Rules

Data Extraction is the process of extracting data from flat files and excel files for cleaning and processing. Extraction is the first process in ETL (Extraction, Transformation, Loading). Extraction has to be given utmost importance if we are to build an efficient data warehouse.

The extraction of data will be carried out by extracting data from flat files and excel files through SQL Server Integration Services of Visual studio and it will be stored in Microsoft SQL Server Management Studio. The data will be initially stored in a staging database wherein it will undergo the cleaning process before being transferred to the warehouse database.

The data from the source files are in CSV (Comma Separated Values) format.

Week Decode table and Store tables were extracted from the manual given for Dominic's Finer Foods' retail operations. Data from both of the tables was stored in CSV files of the same name.

The data from all source files was extracted and loaded into staging database by the name "<603><Group-9>staging-area".

5.2.5 Determining Data Transformation and Cleansing Rules

Data extraction is followed closely by data transformation and loading processes. The input to next process i.e. transforming process is the output data of extraction process. Now, certain transformation rules will be enforced to make sure that resultant data of transformation phase is clean and consistent throughout. This will promote data integrity and will provide us reliable data on which a data warehouse could be built.

The transformation and cleaning rules applied are as given below:

1] Removal of NULL values

A large dataset will have a lot of null values. Furthermore, null values are also created during the data extraction process. Rows with such NULL values should be deleted as part of cleaning.

2] Removal of Dirty Data

Big data sets have the tendency of having dirty data. Dirty data specifically consists of invalid data like "negative" integers, ".", ",", "#", and other special characters. All data sheets such as CCOUNT, MOVEMENT (WANA), WEEK_DECODE, STORE etc. have such unwanted invalid values which can be considered as dirty data. All the dirty data will be deleted from corresponding tables in staging area.

3] Data Conversion

The data extracted from different flat files and excel sheets is stored as data with datatype varchar (string) in the staging area. The data has been then converted into other data types like int, float, double, or date. Our cleaning process shall ensure that data is consistent in nature and we should delete fields that don't conform to our desired datatype.

4] Creation of Surrogate Keys

Surrogate keys were used in all dimension and fact tables before data was loaded in the data warehouse.

5] Derived Attributes

Derived attributes exist in following dimensions, described as below:

Year and month have been derived from the functions YEAR (Date), MONTH (Date)

The 'Total_Sales' measure in the fact table 'Fact_Product_Sales' is obtained from the formula

Total_Sales = (Move * Price) / Qty.

The 'Bakery_Sales' and 'Cheese_Sales' measures in the fact table 'Fact_Ccount' are obtained by aggregating sales for a particular week by grouping the sales based on week number and store.

5.2.6 Plan for Aggregating Tables

Aggregating data properly will be of utmost importance since it will offer better performance results for our data warehouse. Aggregation plays a crucial role in servicing BI queries in a timely manner. Storing data at the most detailed and granular level will give

use the advantage of drilling down or rolling up as we want while catering to BI needs. However, this will consume lot of time and effort.

As such it depends upon us to choose proper approach while supporting BI queries. Consider fact table Fact_CCount for example. Total sale of cheese and bakery for each day has been saved as we need precise data like sales on a special day "Halloween", etc. But this data can be rolled up and can be compared on a monthly basis as well. Hence storing at a granular level and aggregating to a weekly or monthly level will cater to our demands in this situation. Similarly, in fact table Fact_Product_Sales, we need Total_Sales to be access performance of different products on a weekly basis so our level of aggregation will be at the week level. In conclusion, data warehousing analysts will have to consider business question at hand to figure out the level of aggregation that will be needed.

Data Staging Area Organization

During building of a warehouse, data flows from source files to staging area where it is cleaned, to warehousing area. We extracted data from source files like CCount, DONE-WANA, DONE-WBER, DONE-WCOO, UPCANA, UPCBER, UPCCOO into our staging area (<603><Group-9>staging-area) database. Data from pdf files was also extracted to create Store and Week columns in our staging database.

1. We created a time dimension by extracting data from Dominick's data manual pdf file weekly_decode.

2. UPC source files UPCANA, UPCBER, UPCCOO were used to extract data regarding UPC number, product descriptions and categories for each Analgesics, Beer, Cookies.

3. Move source files DONE-WANA, DONE-WBER, DONE-WCOO were used to extract data related to movement, rate, price, sale to calculate total sales of a product.

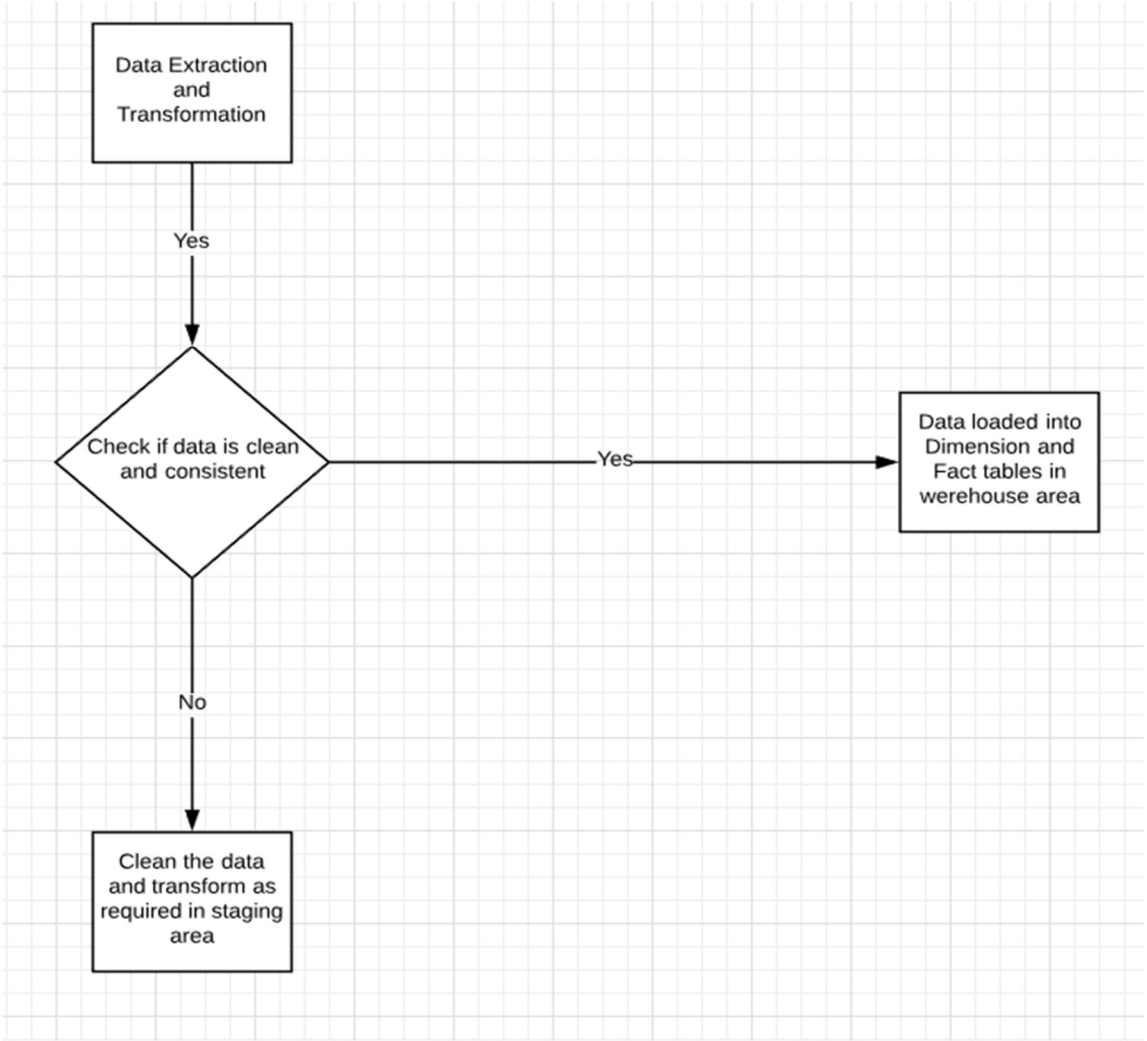
4. CCount source file was extracted to get data regarding Cheese and Bakery sales along with dates.

These tables were used to create dimension tables Dim_Product, Dim_Store, Dim_Time.

Write procedures for all data extractions and loadings

- Make sure the source files are of file types – CSV and Excel.
- Export Store and Week data from Dominik's data manual
- Do not load your staging database with all the data. Instead, use filters and select data that is pertinent to your business questions.
- Add Primary Keys and Foreign Keys if such data is missing to preserve referential integrity.

- Use SQL Server Integration's import-export wizard to move data from source files to staging database and then from staging database to warehousing database. Us the wizard to create tables by mapping fields correctly while importing data from source files.
- Make sure derived columns are added in the tables subsequently while the table is being created
- Check the data in staging area to see if it is fully clean before creating dimensions and fact tables in warehousing database.
- Verify if appropriate Lookups are done between the fact tables in the staging database and the dimension tables in the data warehouse.
- Create Lookup functions to map the data in the fact tables to the dimension tables
- Depending on the number of dimensions, verify the lookup connection created to link fact tables with the dimension tables.
- Extract the resultant data into fact tables



5.2.7 ETL for dimension tables

We have three dimension tables in our start schema design to go along with tow fact tables. Our dimension tables are Time (which is derived from week_decode), Store (which is derived from Store data from dominick's manual), and Product (which is derived from multiple product files like UPCANA, UPCBER, UPCCOO).

Product Dimension ETL Conversion

We need product information of three products primarily– Analgesics, Beer, Cookies. As such we extracted data from UPCANA, UPCBER, and UPCCOO and stored in the dbo.Product database in our staging area. We only need Universal Product Code (UPC) and Descrip (Description) of these products. As per our requirement, our design will append columns product category and category description in the Product table in staging database. We use the import export wizard to extract data from excel files to the Product table (OLE DB Destination). After that, we will again use the import export wizard to set up an OLE DB Destination to OLE DB Destination transaction to transfer data from staging database of product to Dimensional table of Product.

Time Dimension ETL Conversion:

We extracted the data related to weeks and special events from week_decode and stored it in our staging database table Week under the attributes Week #, Start, End, Special Events. We then extracted month by using the function MONTH(Date) and Year using YEAR(Date). We then loaded this data into dimension table Dim_Time with attributes Year, month, Week, Special_Event.
The Week table is loaded into the data warehouse database through SSIS dataflow components OLE DB source and OLE DB destination, and a new Dim_Time is created.

Store Dimension ETL Conversion:

Dominick's data manual provided use with store details like store number, zone, price tier, city, and address. The store details were first extracted into a csv file. The same csv file acted as a Source File when we used the import export wizard to extract store data to the staging area database of Store. Few irrelevant and null values were cleaned before the data was transferred from staging area to warehousing database by using import export wizard of SSIS. This time staging database acted as an OLE DB Source and warehousing database acted as an OLE DB Destination. As such, Dim_Store was created.

5.2.8 ETL for fact tables

Product Sales fact table:

Product sales fact table creation involved the measures from the new movement staging table along with three lookups – Store lookup, Time lookup and Product lookup to create the composite primary key for the fact table.

The final columns loaded into this fact table involved – Move, Price, Qty, Total_Sales (Derived column), Store Key, Time Key and Product Key.

Ccount Fact table:

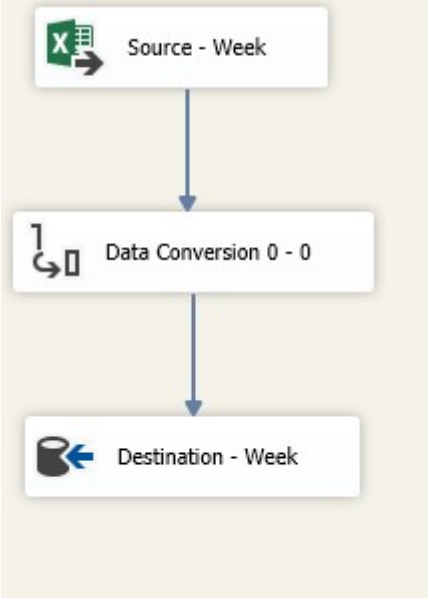
The creation of ccount fact table involved measures from the aggregated ccount fact table. The dimensions were connected to the fact table using Store lookup and Time lookup. The final columns loaded into the aggregated fact table are – Bakery Sales, Cheese Sales, Time Key, and Store Key.

5.3 ETL Implementation

All the extraction, cleansing and transforming process on the Dominicks finer foods data are performed using SQL server management studio and SSIS. The below screenshots explain the entire step by step process carried out from extraction till loading the dimension and fact tables in the data warehouse area.

5.3.1 Extraction of source files into the staging area

a. **Week Decode**



Snapshot of Week source table:

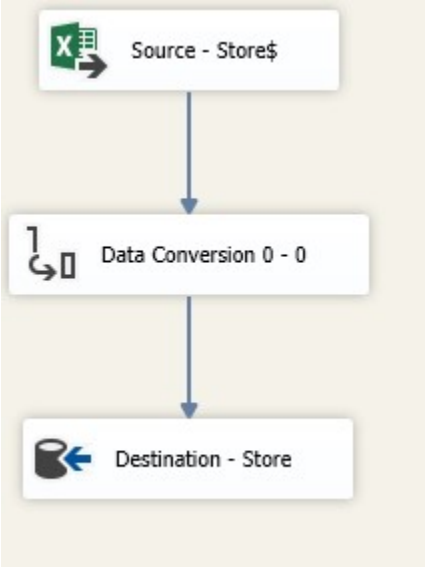
```
select * from week;
```

100 %

Results | Messages

	Week #	Start	End	Special Events
1	1	1989-09-14	1989-09-20	NULL
2	2	1989-09-21	1989-09-27	NULL
3	3	1989-09-28	1989-10-04	NULL
4	4	1989-10-05	1989-10-11	NULL
5	5	1989-10-12	1989-10-18	NULL
6	6	1989-10-19	1989-10-25	NULL
7	7	1989-10-26	1989-11-01	Halloween
8	8	1989-11-02	1989-11-08	NULL
9	9	1989-11-09	1989-11-15	NULL
10	10	1989-11-16	1989-11-22	NULL
11	11	1989-11-23	1989-11-29	Thanksgiving

b. Store



Snapshot of store source table:

```
select * from store;
```

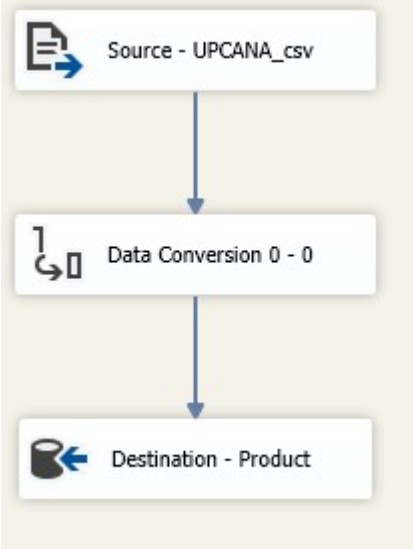
100 %

Results | Messages

	Store	City	Price Tier	Zone	Zip Code	Address
1	2	River Forest	High	1	60305	7501 W. North Ave.
2	4	Park Ridge	Medium	2	60068	Closed
3	5	Palatine	Medium	2	60067	223 Northwest HWY.
4	8	Oak Lawn	Low	5	60435	8700 S. Cicero Ave.
5	9	Morton Grove	Medium	2	60053	6931 Dempster
6	12	Chicago	High	7	60660	6009 N. Broadway Ave.
7	14	Glenview	High	1	60025	1020 Waukegan Rd.
8	18	River Grove	Low	5	60171	8355 W. Belmont Ave.
9	19	Glen Ellyn	NULL	NULL	60137	Closed
10	21	Hanover Park	CubFighter	6	60103	1440 Irving Park Rd.

c. Product

Loading products under Analgesics category:



Snapshot of loaded data:

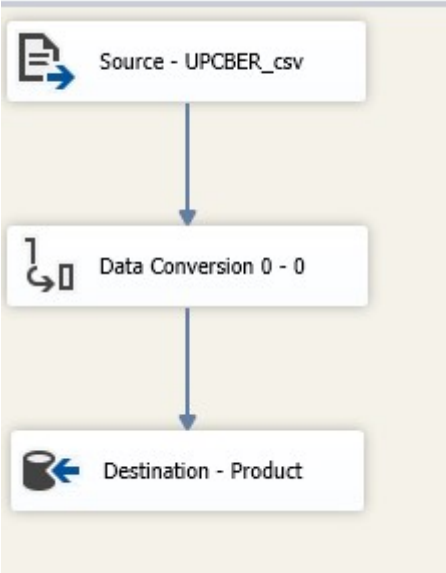
```
select * from product;
```

100 %

Results | Messages

	UPC	CATEGORY
1	1192603016	ANALGESICS
2	1192662108	ANALGESICS
3	1650001020	ANALGESICS
4	1650001022	ANALGESICS
5	1650004106	ANALGESICS
6	1650004108	ANALGESICS
7	1650004703	ANALGESICS
8	2140649030	ANALGESICS
9	2586600493	ANALGESICS
10	2586610493	ANALGESICS

Loading products under Beer category:



Snapshot of loaded data:

641	78765150...	ANALGESICS
642	294	BEER
643	307	BEER
644	710	BEER
645	711	BEER
646	712	BEER
647	720	BEER
648	721	BEER
649	723	BEER
650	731	BEER
651	732	BEER
652	735	BEER
653	740	BEER
654	750	BEER
655	757	BEER

Loading products under cookies category:

Snapshot of loaded data:

1...	79709637...	BEER
1...	79709638...	BEER
1...	1410007012	COOKIES
1...	1410007016	COOKIES
1...	1410007056	COOKIES
1...	1410007070	COOKIES
1...	1410007174	COOKIES
1...	1410007188	COOKIES
1...	1410007219	COOKIES
1...	1410007223	COOKIES
1...	1410007233	COOKIES
1...	1410007234	COOKIES
1...	1410007297	COOKIES
1...	1410007402	COOKIES

d. Movement

Snapshot of loaded data into Movement source table:

```
select * from dbo.Movement;
```

100 %

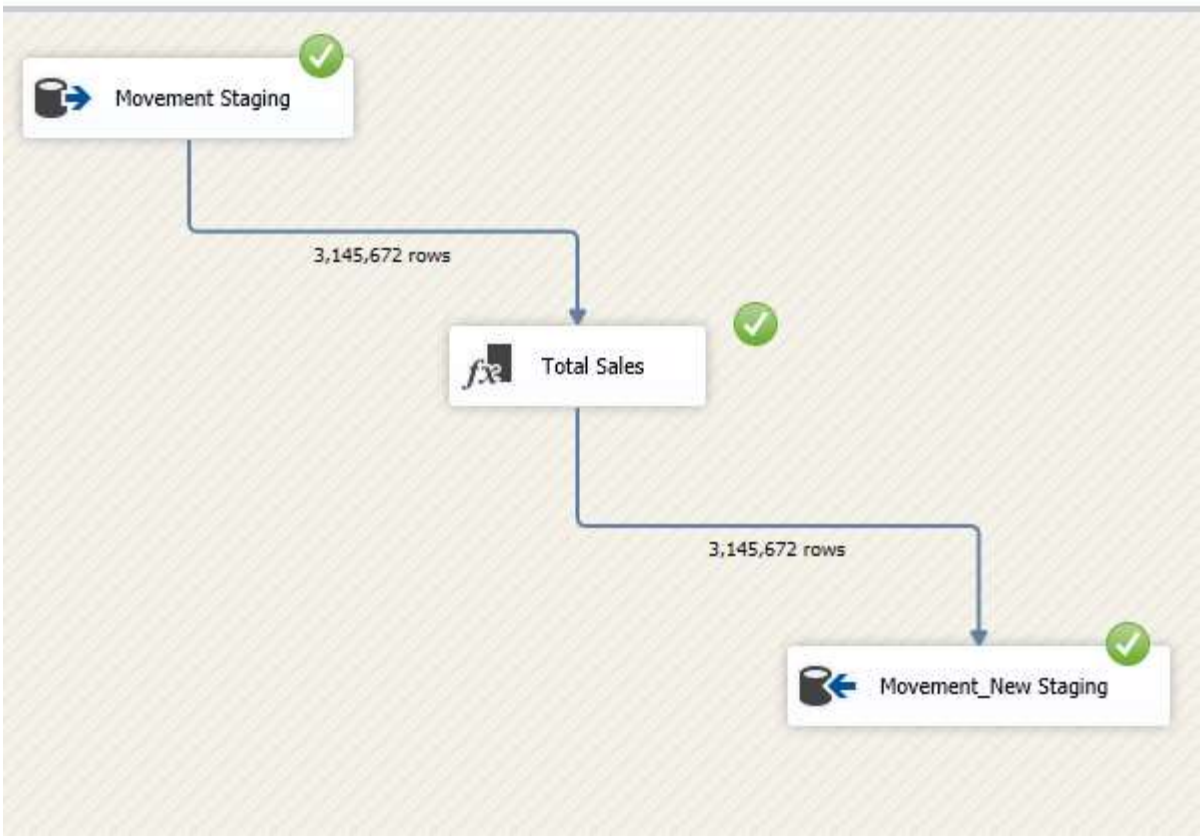
Results | Messages

	STORE	UPC	WEEK	MOVE	QTY	PRICE
1	12	2140649030	337	0	1	0
2	12	2140649030	338	0	1	0
3	12	2140649030	339	0	1	0
4	12	2140649030	340	0	1	0
5	12	2140649030	341	0	1	0
6	12	2140649030	342	0	1	0
7	12	2140649030	343	0	1	0
8	12	2140649030	344	0	1	0
9	12	2140649030	345	0	1	0
10	12	2140649030	346	0	1	0
11	12	2140649030	347	0	1	0
12	12	2140649030	348	0	1	0
13	12	2140649030	349	0	1	0
14	12	2140649030	350	1	1	8.69

Data flow for Movement_New_Staging table:

Derived column ‘Total Sales’ is added to the movement table. The new movement staging table is further used to load the measures into product sales fact table.

Total Sales = (Price*Move)/Qty



Snapshot of data:

```
select * from Movement_New_Staging;
```

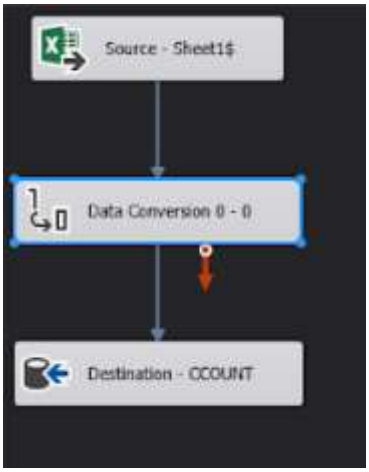
100 %

Results | Messages

	STORE	UPC	WEEK	MOVE	QTY	PRICE	Total_Sales
1	110	3828161041	153	2	1	5.38	10.76
2	110	3828161041	154	1	1	5.38	5.38
3	110	3828161041	155	9	1	5.38	48.42
4	110	3828161041	156	3	1	5.38	16.14
5	110	3828161041	157	1	1	5.38	5.38
6	110	3828161041	158	3	1	5.38	16.14
7	110	3828161041	159	6	1	5.38	32.28
8	110	3828161041	160	3	1	5.38	16.14
9	110	3828161041	161	2	1	5.38	10.76
10	110	3828161041	162	2	1	5.38	10.76
11	110	3828161041	163	3	1	5.38	16.14
12	110	3828161041	164	4	1	5.38	21.52

e. Ccount

Dataflow:



Snapshot of Ccount data:

```
select * from CCOUNT;
```

100 %

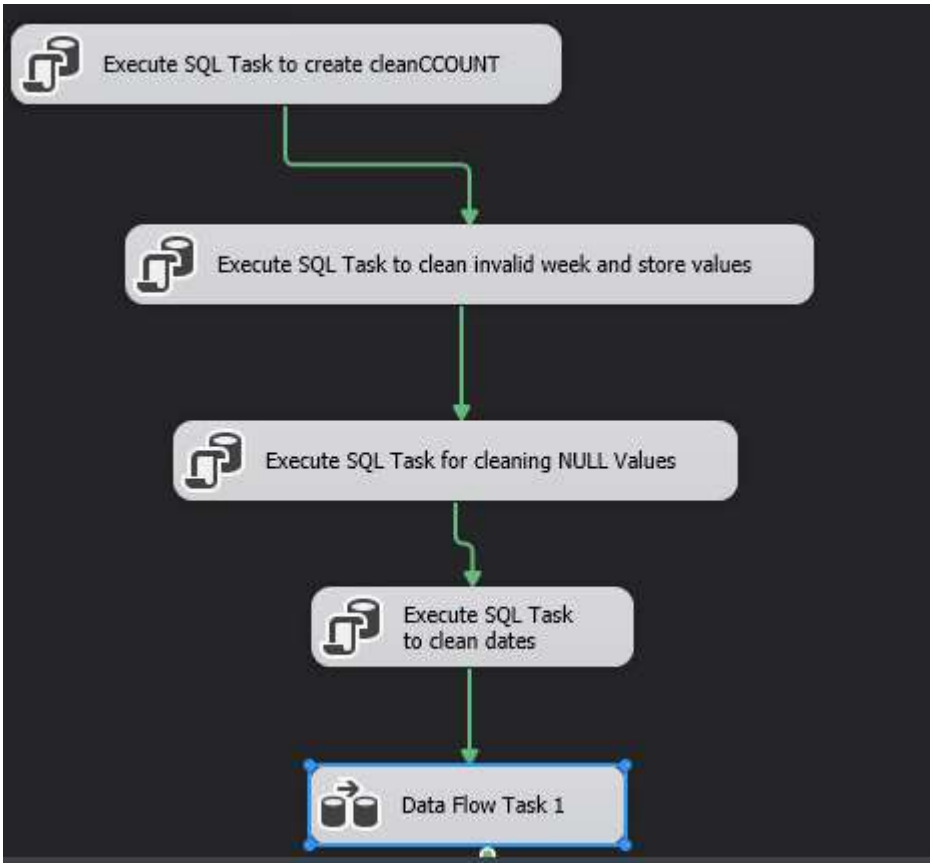
Results | Messages

	Store	Date	Bakery	Cheese	Week
1	33	1995-06-14 00:00:00.000	792.66	315.91	300
2	33	1995-06-15 00:00:00.000	987.31	345.37	301
3	33	1995-06-16 00:00:00.000	1195.46	447.08	301
4	33	1995-06-17 00:00:00.000	1448.41	430.93	301
5	33	1995-06-18 00:00:00.000	967.11	455.19	301
6	33	1995-06-19 00:00:00.000	901.76	322.94	301
7	33	1995-06-20 00:00:00.000	934.72	227.82	301
8	33	1995-06-21 00:00:00.000	802.9	242.42	301
9	33	1995-06-22 00:00:00.000	1008.71	232.57	302
10	33	1995-06-23 00:00:00.000	1064.54	422.97	302
11	33	1995-06-24 00:00:00.000	1211.17	535.68	302
12	33	1995-06-26 00:00:00.000	945.19	387.5	302
13	33	1995-06-27 00:00:00.000	878.11	379.86	302
14	33	1995-06-28 00:00:00.000	778.33	228.35	302
15	33	1995-06-29 00:00:00.000	998.19	298.35	303
16	33	1995-06-30 00:00:00.000	1303.17	382.69	303
17	33	1995-07-01 00:00:00.000	1397.51	533.41	303
18	33	1995-07-02 00:00:00.000	1089.57	425.88	303
19	33	1995-07-03 00:00:00.000	1353.2	474.07	303
20	33	1995-07-04 00:00:00.000	1152.41	416.83	303

CleanCcount

Execute SQL tasks in SSIS were used to cleanse the Ccount data. The clean data was then loaded into the cleanCcount table.

Data flow for CleanCcount:



SQL Queries used for cleanCcount:

```
/*Query for creating cleanCCOUNT:*/
SELECT [Store], [Date],[Cheese],[Bakery],[Week]
INTO cleanCCOUNT
FROM CCOUNT;

/*Query for Cleaning Negative week Values*/
DELETE FROM cleanCCOUNT
WHERE ([WEEK] < 0);

/*Query for Cleaning Invalid Store Values:*/
DELETE FROM cleanCCOUNT
WHERE ([Store] NOT IN
(SELECT Store_Number
FROM Dim_Store));

/*Query for Cleaning Null Values*/
DELETE FROM cleanCCOUNT
```

```
WHERE ([Week] IS NULL) OR
([Store] IS NULL) OR
([Date] IS NULL) OR
([Cheese] IS NULL) OR
([Bakery] IS NULL);

/*Query for Cleaning Date Values*/
UPDATE cleanCCOUNT
SET [Date] = REPLACE([Date], "", '');
```

Snapshot of data from cleanCcount:

```
SELECT *
  FROM [<603><Group-9>staging-area].[dbo].[cleanCCOUNT]
```

100 %

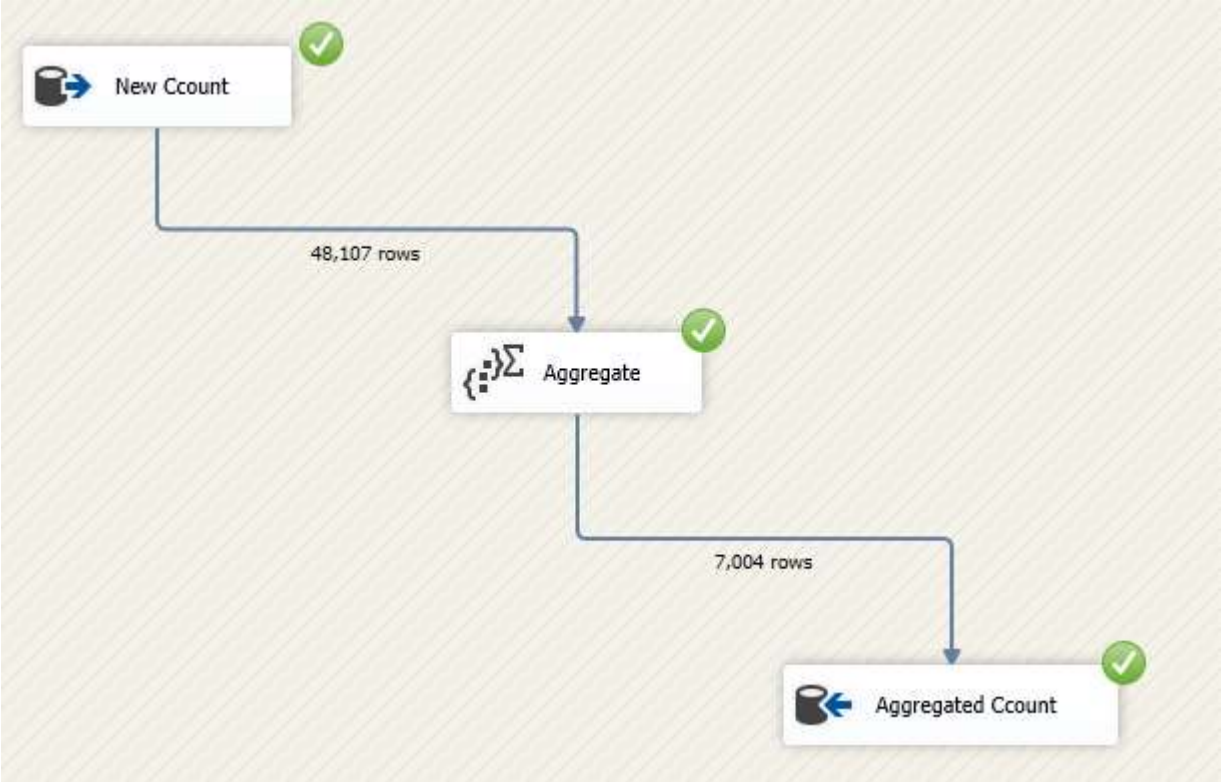
Results | Messages

	Store	Date	Bakery	Cheese	Week
1	33	1995-06-14 00:00:00.000	792.66	315.91	300
2	33	1995-06-15 00:00:00.000	987.31	345.37	301
3	33	1995-06-16 00:00:00.000	1195.46	447.08	301
4	33	1995-06-17 00:00:00.000	1448.41	430.93	301
5	33	1995-06-18 00:00:00.000	967.11	455.19	301
6	33	1995-06-19 00:00:00.000	901.76	322.94	301
7	33	1995-06-20 00:00:00.000	934.72	227.82	301
8	33	1995-06-21 00:00:00.000	802.9	242.42	301
9	33	1995-06-22 00:00:00.000	1008.71	232.57	302
10	33	1995-06-23 00:00:00.000	1064.54	422.97	302
11	33	1995-06-24 00:00:00.000	1211.17	535.68	302
12	33	1995-06-26 00:00:00.000	945.19	387.5	302
13	33	1995-06-27 00:00:00.000	878.11	379.86	302
14	33	1995-06-28 00:00:00.000	778.33	228.35	302
15	33	1995-06-29 00:00:00.000	998.19	298.35	303

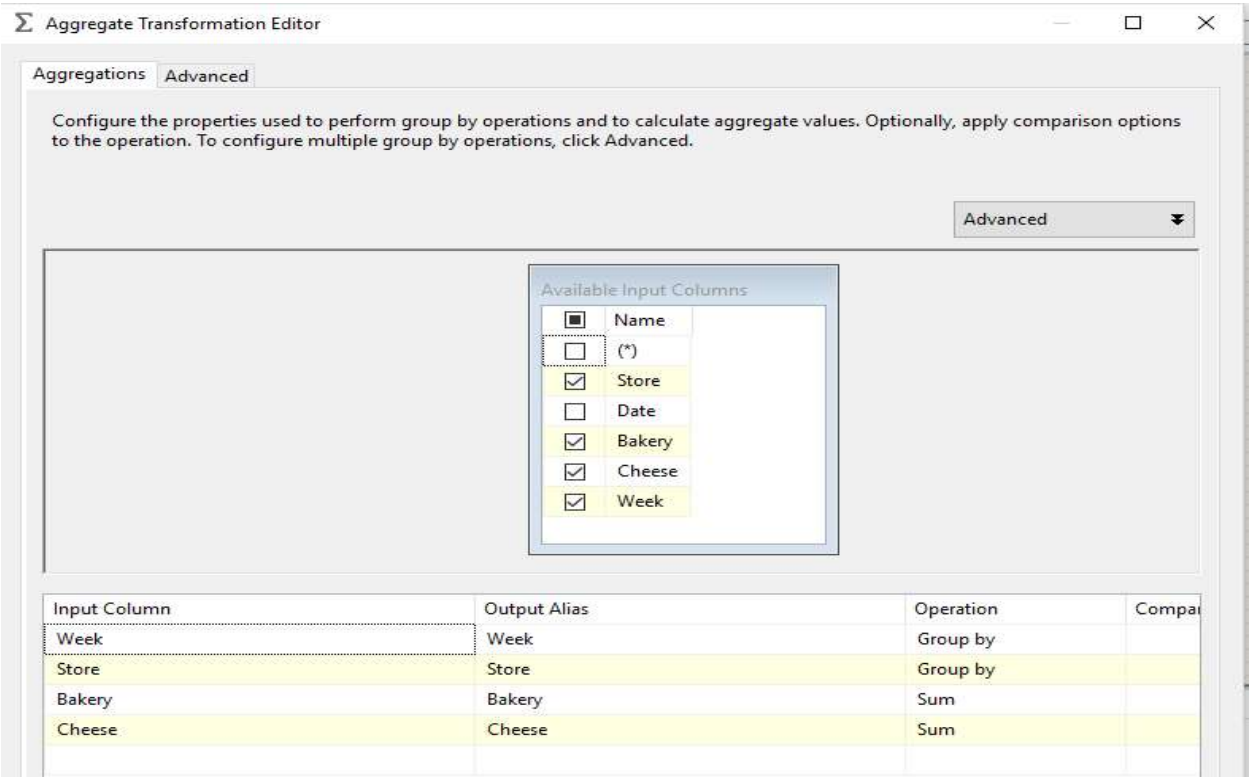
AggCcount:

The data in the cleanCcount table was day-wise for every store. This data was aggregate to display the week wise sales data for every store. The aggregated data was stored in the AggCcount table. AggCcount was then further used to load the measure into the Ccount fact table.

Data flow:



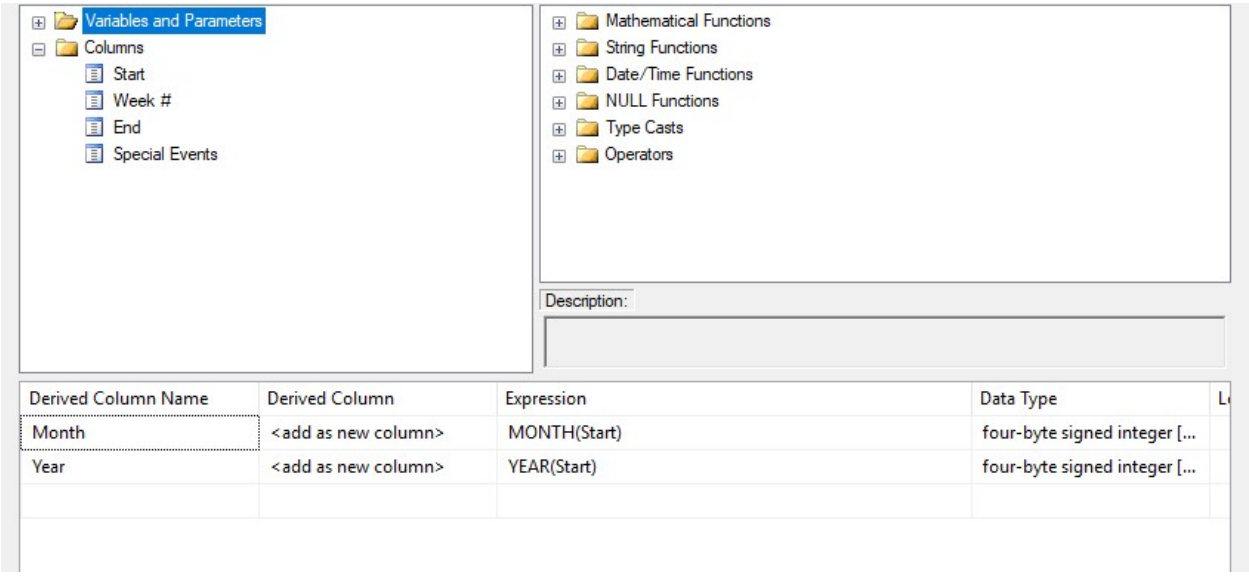
Aggregate function was used to group the sales data by week and store.



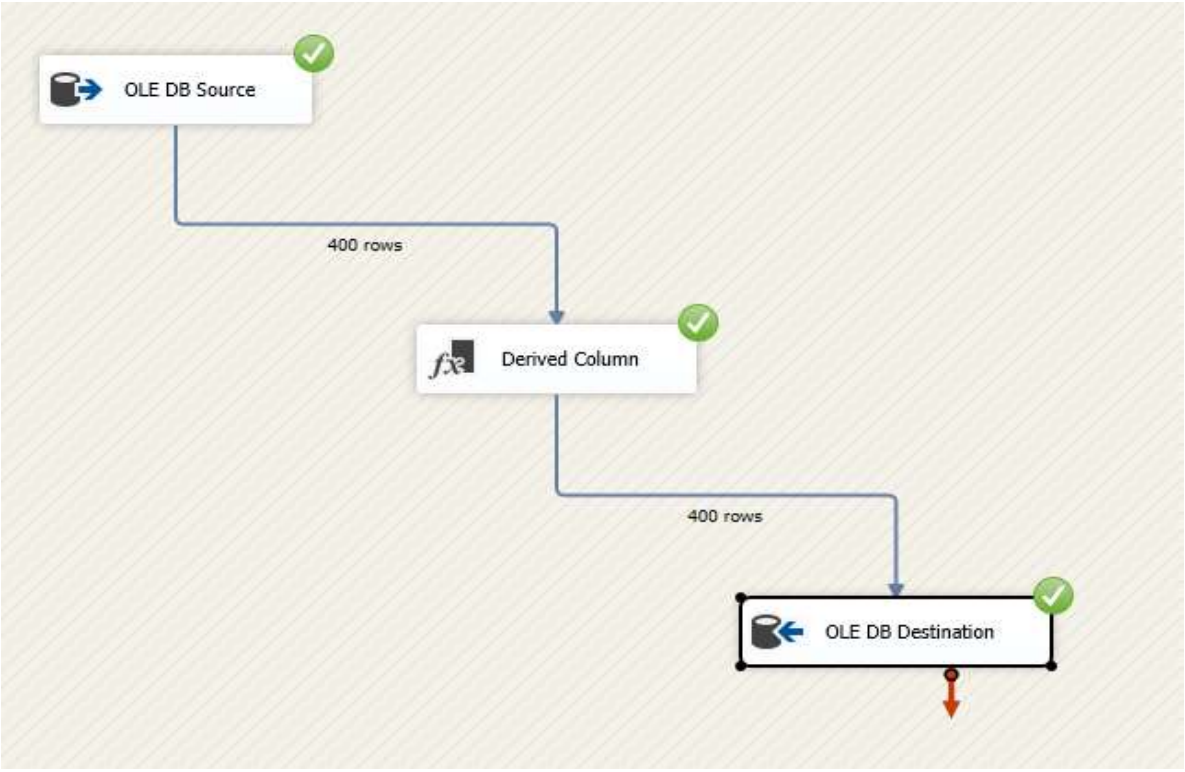
5.3.2 Creation of dimension tables in staging area

a) Time Dimension

Columns ‘Month’ and ‘Year’ were derived from the start date of every row in week source table.



Data flow



Snapshot of Time dimension table

```
select * from dbo.Dim_Time;
```

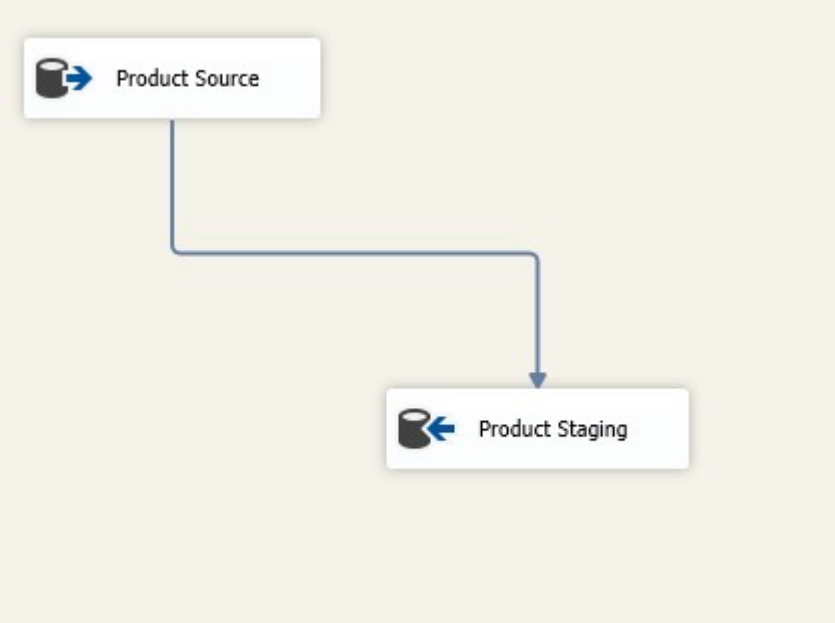
100 %

Results | Messages

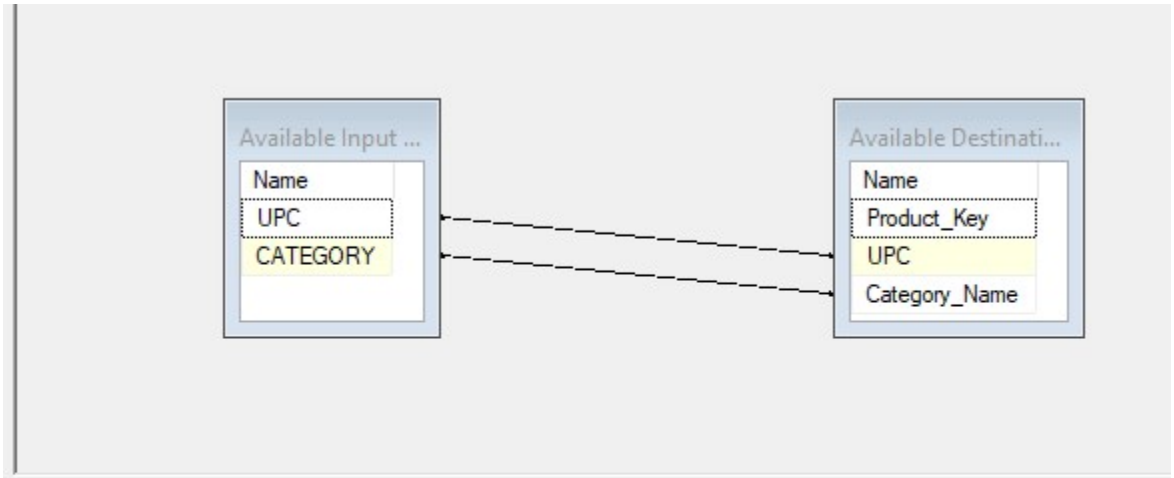
	Time_Key	Year	Month	Week_Number	Special_Event
1	1	1989	9	1	NULL
2	2	1989	9	2	NULL
3	3	1989	9	3	NULL
4	4	1989	10	4	NULL
5	5	1989	10	5	NULL
6	6	1989	10	6	NULL
7	7	1989	10	7	Halloween
8	8	1989	11	8	NULL
9	9	1989	11	9	NULL
10	10	1989	11	10	NULL
11	11	1989	11	11	Thanksgiving
12	12	1989	11	12	NULL
13	13	1989	12	13	NULL
14	14	1989	12	14	NULL

b) Product Dimension

Data flow



Column mapping from product source to product dimension table



Input Column	Destination Column
<ignore>	Product_Key
UPC	UPC
CATEGORY	Category_Name

Snapshot of loaded Product dimension

```
select * from Dim_Product;
```

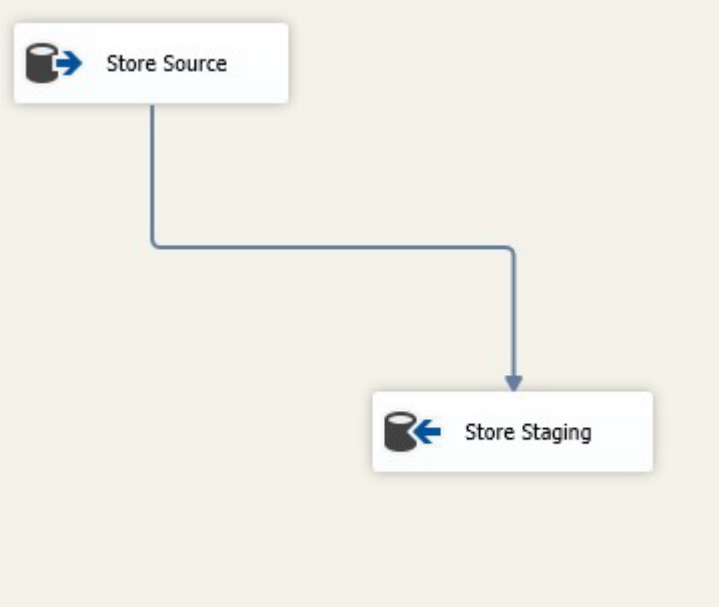
100 %

Results | Messages

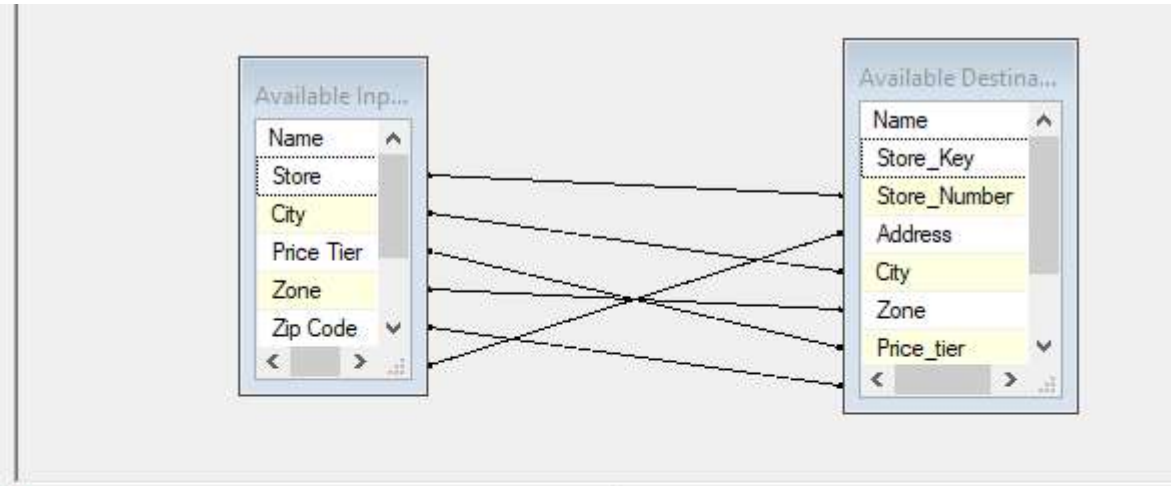
	Product_Key	UPC	Category_Name
1	1	1192603016	ANALGESICS
2	2	1192662108	ANALGESICS
3	3	1650001020	ANALGESICS
4	4	1650001022	ANALGESICS
5	5	1650004106	ANALGESICS
6	6	1650004108	ANALGESICS
7	7	1650004703	ANALGESICS
8	8	2140649030	ANALGESICS
9	9	2586600493	ANALGESICS
10	10	2586610493	ANALGESICS
11	11	2586610501	ANALGESICS
12	12	2586610502	ANALGESICS

c) Store Dimension

Data flow



Column mapping from store source table to store dimension



Input Column	Destination Column
<ignore>	Store_Key
Store	Store_Number
Address	Address
City	City
Zone	Zone
Price Tier	Price_tier
Zip Code	Zip_Code

Snapshot of loaded dimension table

```
select * from Dim_Store;
```

100 %

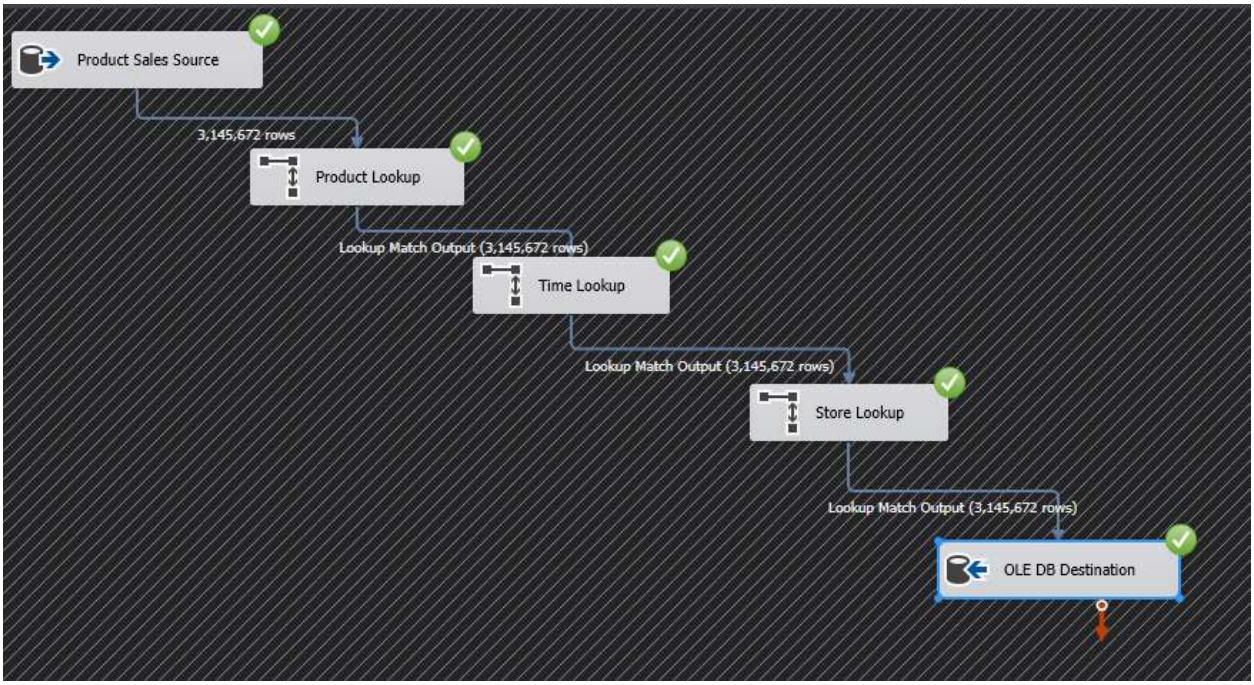
Results | Messages

	Store_Key	Store_Number	Address	City	Zone	Price_tier	Zip_Code
1	1	2	7501 W. North Ave.	River Forest	1	High	60305
2	2	4	Closed	Park Ridge	2	Medium	60068
3	3	5	223 Northwest HWY.	Palatine	2	Medium	60067
4	4	8	8700 S. Cicero Ave.	Oak Lawn	5	Low	60435
5	5	9	6931 Dempster	Morton Grove	2	Medium	60053
6	6	12	6009 N. Broadway Ave.	Chicago	7	High	60660
7	7	14	1020 Waukegan Rd.	Glenview	1	High	60025
8	8	18	8355 W. Belmont Ave.	River Grove	5	Low	60171
9	9	19	Closed	Glen Ellyn	NULL	NULL	60137
10	10	21	1440 Irving Park Rd.	Hanover Park	6	CubFighter	60103
11	11	25	Closed	Chicago	NULL	NULL	60639
12	12	28	1145-55 Mt Prospect Pz.	Mt. Prospect	2	Medium	60054
13	13	32	1900 S. Cumberland Ave.	Park Ridge	1	High	60068
14	14	33	3012 N. Broadway Ave.	Chicago	7	High	60657

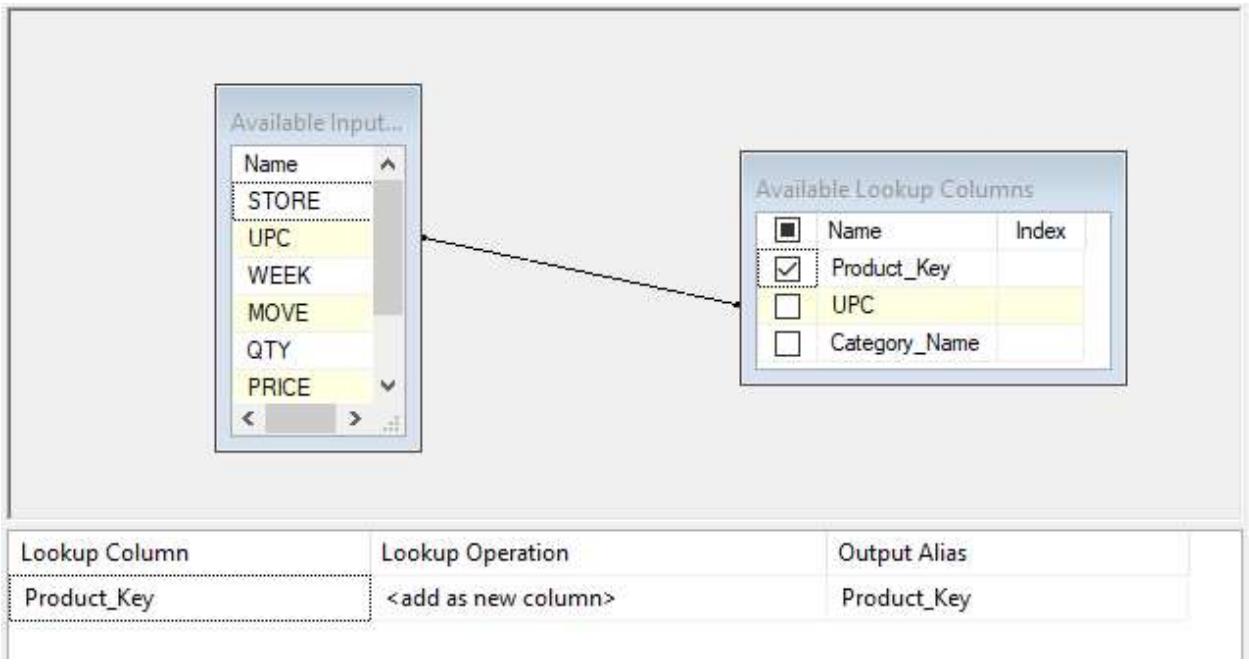
5.3.3 Creation of fact tables in staging area

A. Product Sales Fact table

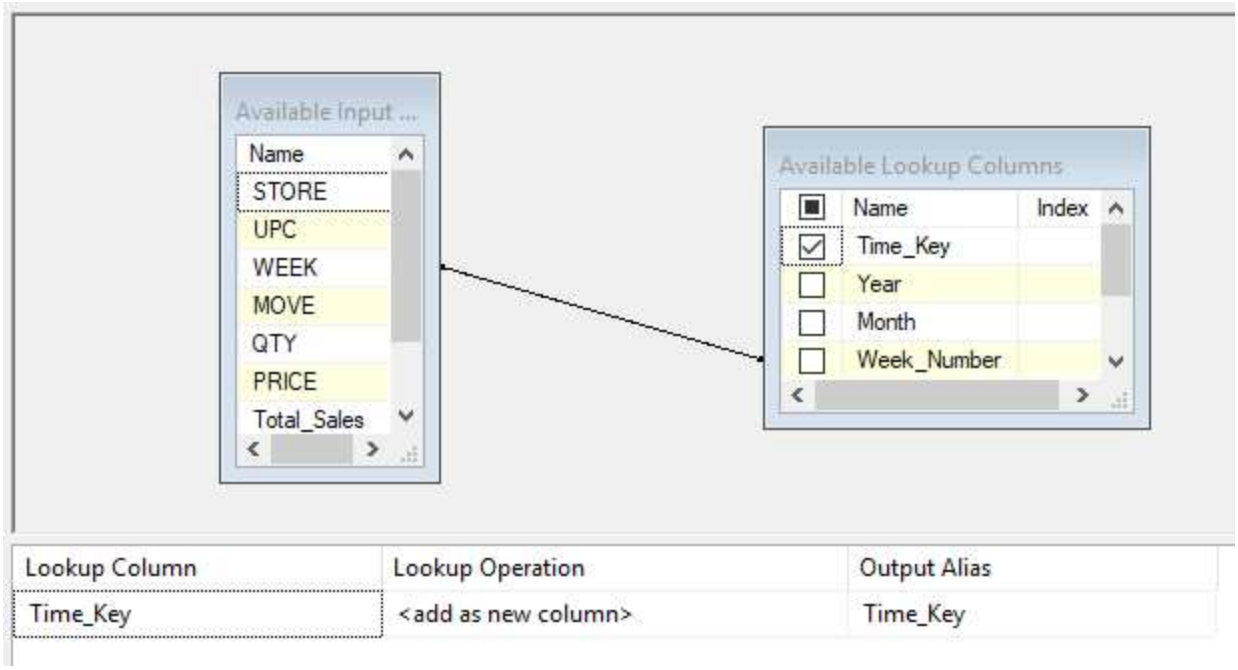
Data flow



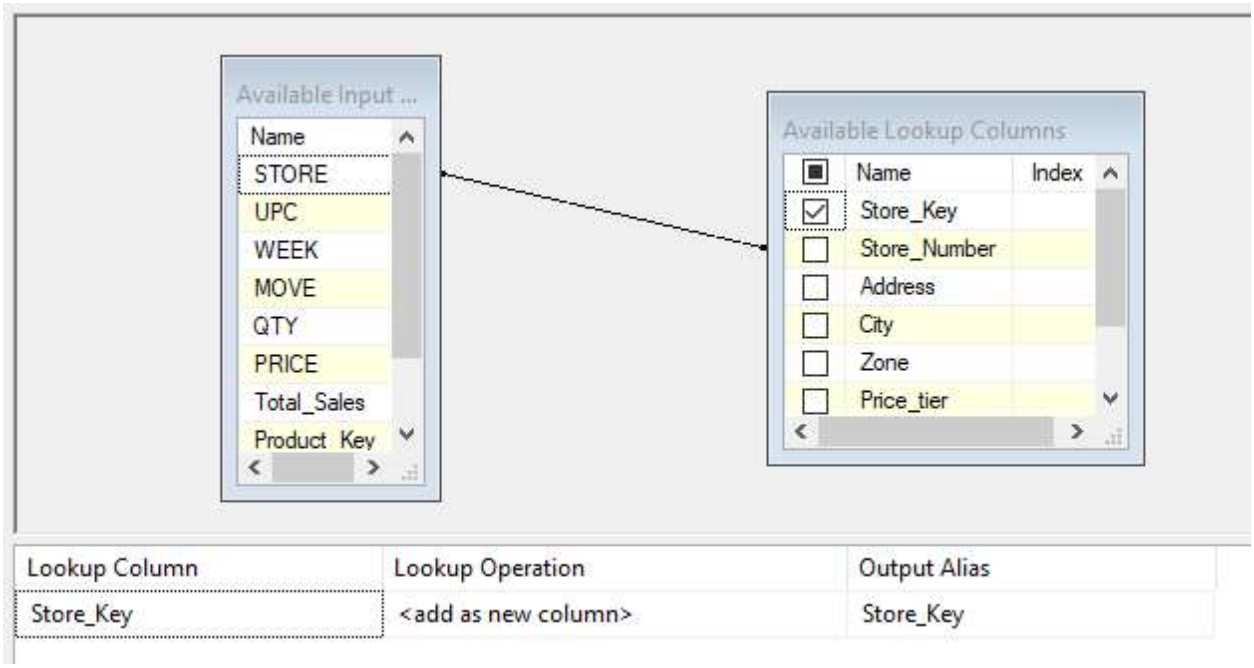
Product lookup



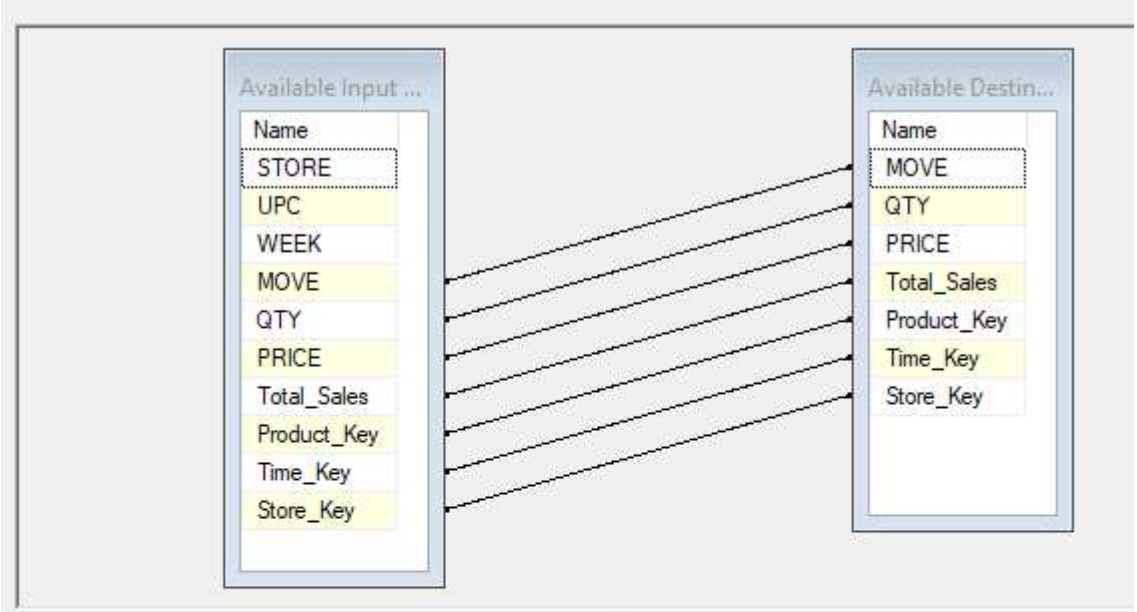
Time lookup



Store lookup



Column mapping for destination fact table



Input Column	Destination Column
MOVE	MOVE
QTY	QTY
PRICE	PRICE
Total_Sales	Total_Sales
Product_Key	Product_Key
Time_Key	Time_Key
Store_Key	Store_Key

Snapshot of product sales fact table from staging area

```
select * from [dbo].[Fact_Product_Sales_Staging];
```

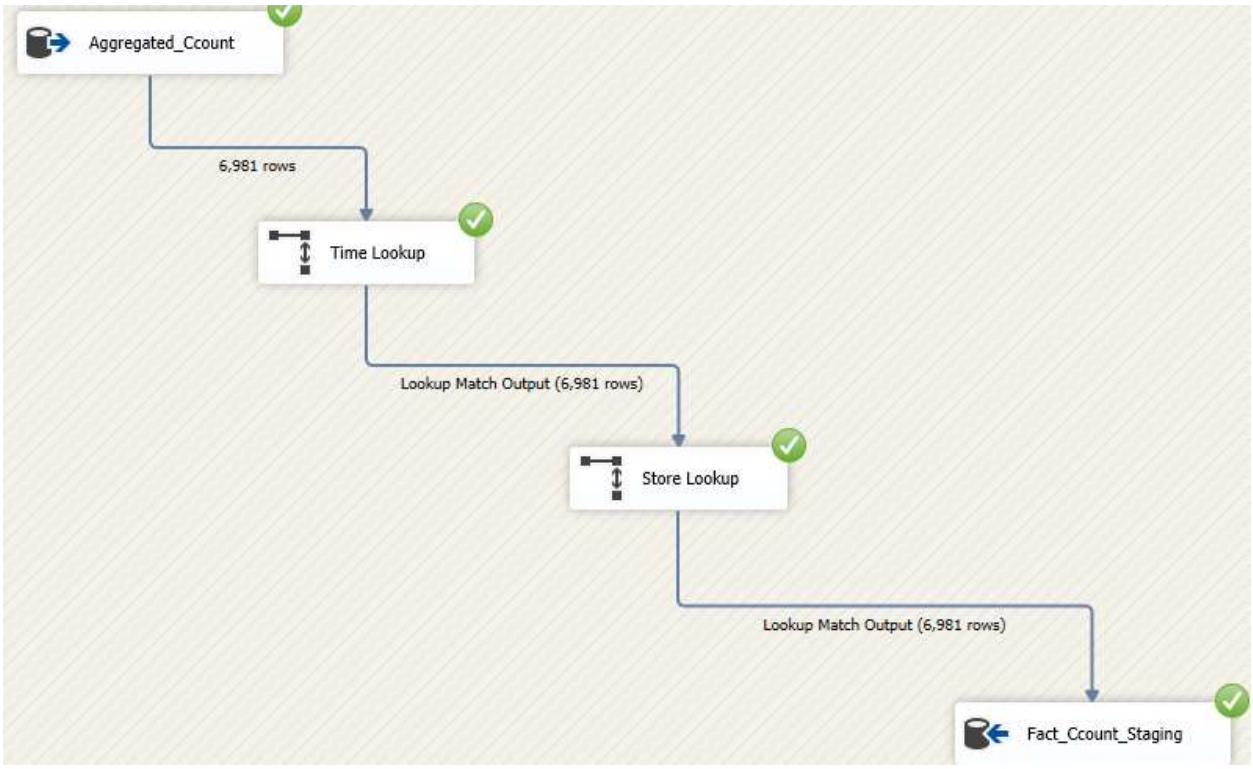
100 %

Results | Messages

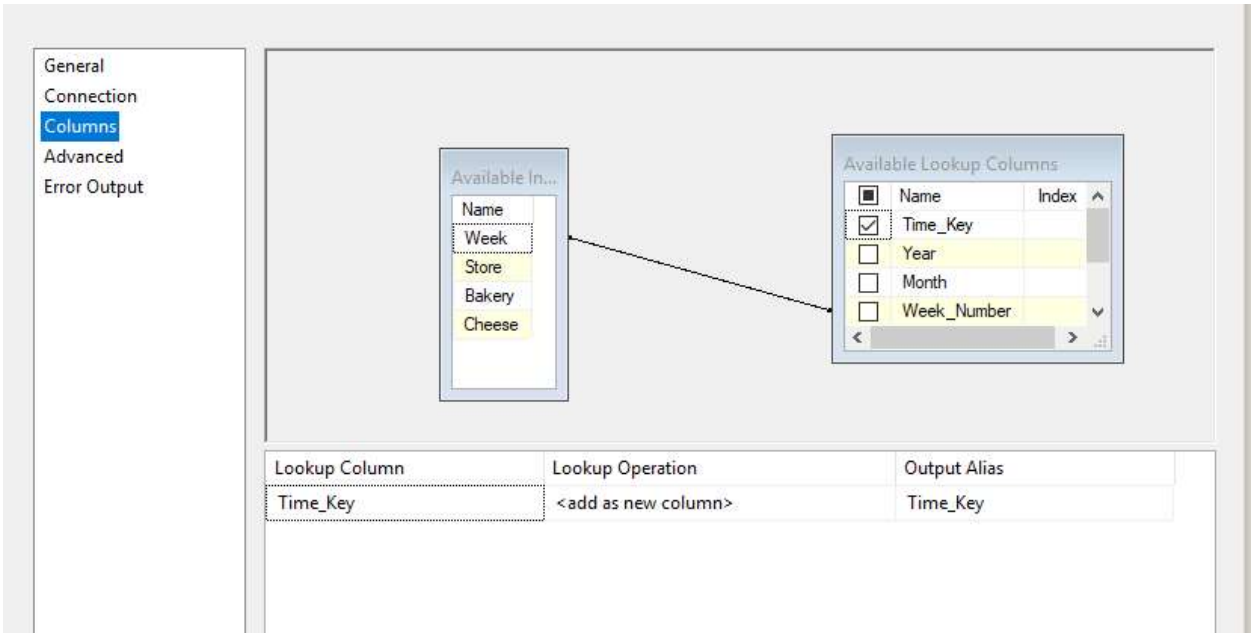
	MOVE	QTY	PRICE	Total_Sales	Product_Key	Time_Key	Store_Key
1	0	1	0	0	1	306	44
2	1	1	2.99	2.99	1	307	44
3	0	1	0	0	1	308	44
4	0	1	0	0	1	309	44
5	0	1	0	0	1	310	44
6	0	1	0	0	1	311	44
7	0	1	0	0	1	312	44
8	0	1	0	0	1	313	44
9	0	1	0	0	1	314	44
10	0	1	0	0	1	315	44
11	0	1	0	0	1	316	44
12	0	1	0	0	1	317	44

B. Ccount Fact table

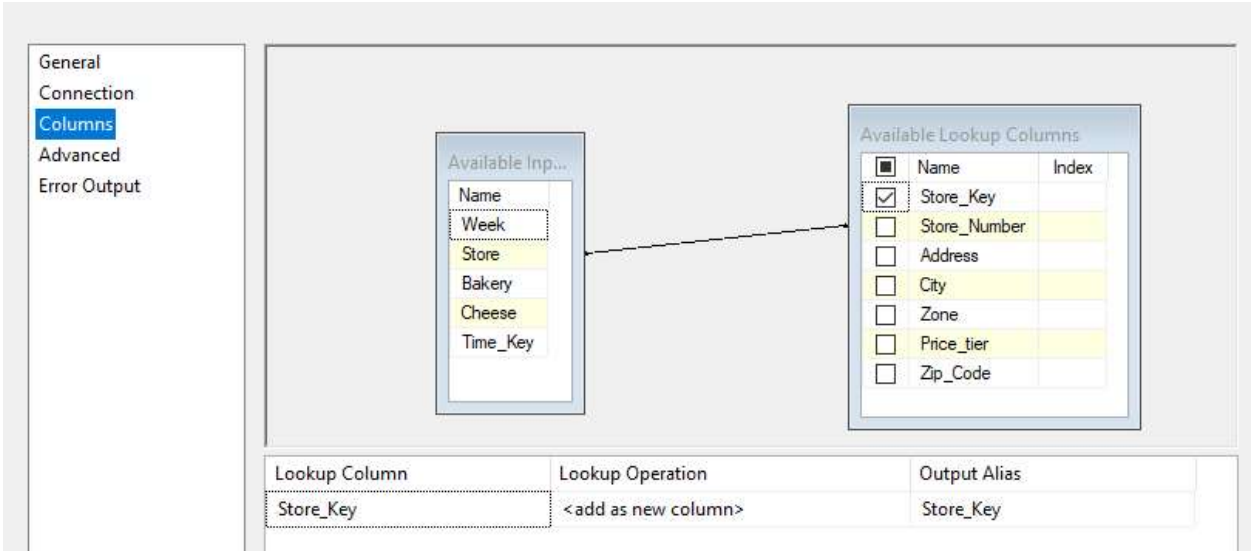
Data flow



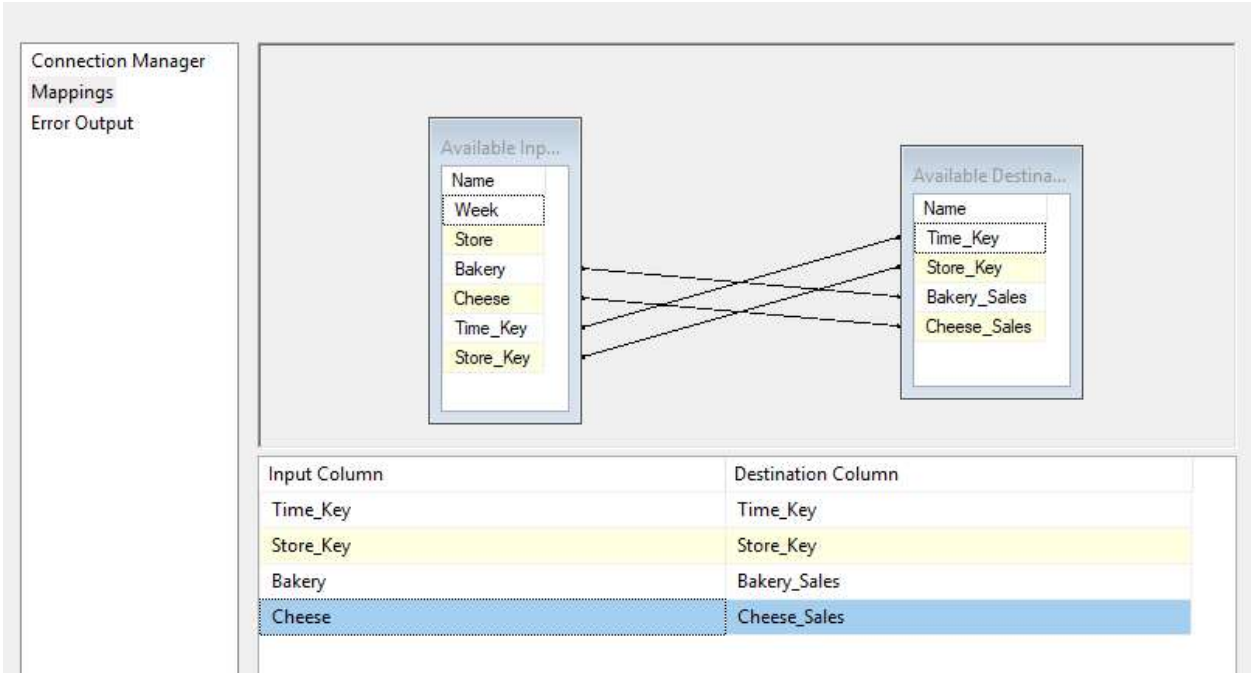
Time Lookup



Store lookup



Column mapping for destination Ccount fact table



Snapshot of loaded data in Fact Ccount staging table

```
select * from fact_ccount_staging;
```

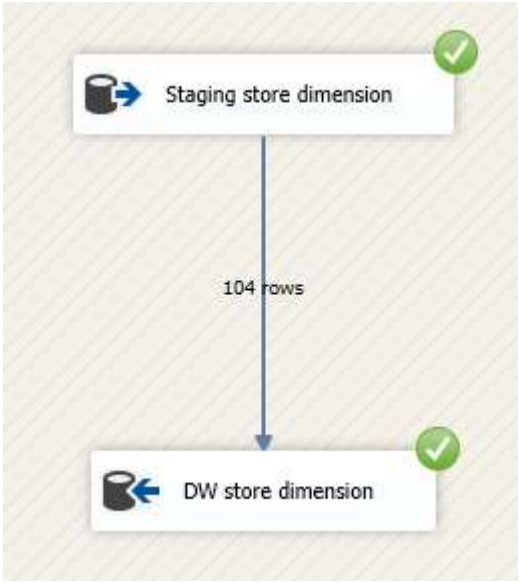
100 %

Results | Messages

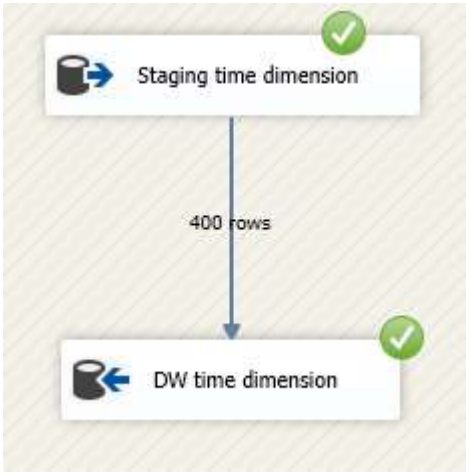
	Bakery_Sales	Cheese_Sales	Time_Key	Store_Key
1	11660.7	981.14	1	1
2	7943.28	666.68	1	2
3	11895.54	3007.53	1	3
4	13221.39	1305.69	1	4
5	12215.39	1112.23	1	5
6	7459.01	2457.46	1	6
7	11430.82	1305.08	1	7
8	15610.09	900.75	1	8
9	1903.75	309.92	1	9
10	7542.47	468.11	1	10
11	3943.52	167.94	1	11
12	7630.52	533.95	1	12
13	16710.37	2670.92	1	13
14	6695.31	1550.55	1	14
15	2060.17	276.52	1	15

5.4 Loading fact and dimension tables from staging to DW area

5.4.1 Store dimension



5.4.2 Time Dimension



5.4.3 Product dimension

5.4.4 Product Sales fact

Snapshot of loaded data in DW area

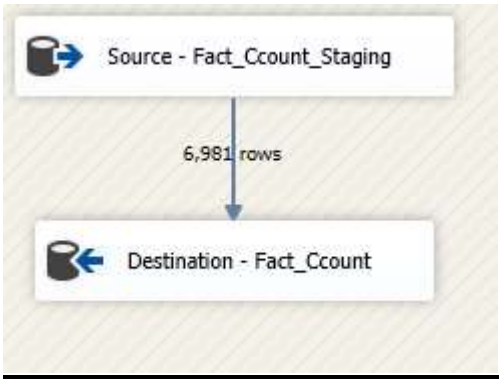
```
select * FROM [<603><Group-9>-dw-area].[dbo].[Fact_Product_Sales];
```

100 %

Results | Messages

	MOVE	QTY	PRICE	Total_Sales	Product_Key	Time_Key	Store_Key
1	0	1	0	0	1	306	44
2	1	1	2.99	2.99	1	307	44
3	0	1	0	0	1	308	44
4	0	1	0	0	1	309	44
5	0	1	0	0	1	310	44
6	0	1	0	0	1	311	44
7	0	1	0	0	1	312	44
8	0	1	0	0	1	313	44
9	0	1	0	0	1	314	44
10	0	1	0	0	1	315	44
11	0	1	0	0	1	316	44

5.4.5 Ccount Fact



Snapshot of loaded data in DW area

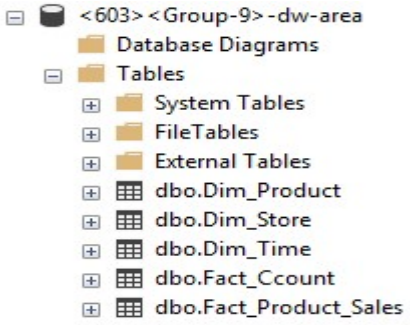
```
select * from Fact_Ccount;
```

100 %

Results | Messages

	Bakery_Sales	Cheese_Sales	Time_Key	Store_Key
1	11660.7	981.14	1	1
2	7943.28	666.68	1	2
3	11895.54	3007.53	1	3
4	13221.39	1305.69	1	4
5	12215.39	1112.23	1	5
6	7459.01	2457.46	1	6
7	11430.82	1305.08	1	7
8	15610.09	900.75	1	8
9	1903.75	309.92	1	9
10	7542.47	468.11	1	10
11	3943.52	167.94	1	11
12	7630.52	533.95	1	12
13	16710.37	2670.92	1	13
14	6695.31	1550.55	1	14
15	2060.17	276.52	1	15

Snapshot of the Data warehouse area



5.4.6 SQL Queries used to create the fact and dimension tables

Time Dimension

```
create table Dim_Time(
Time_Key int identity(1,1) primary key not null,
Year int not null,
Month int not null,
Week_Number int,
Special_Event nvarchar(255)
)
```

Product Dimension

```
create table Dim_Product(
Product_Key int identity(1,1) primary key not null,
UPC bigint not null,
Category_Name varchar(50) not null
)
```

Store Dimension

```
create table Dim_Store(
Store_Key int identity(1,1) primary key not null,
Store_Number int not null,
Address nvarchar(255),
City nvarchar(255),
Zone int,
```

Product Sales Fact table

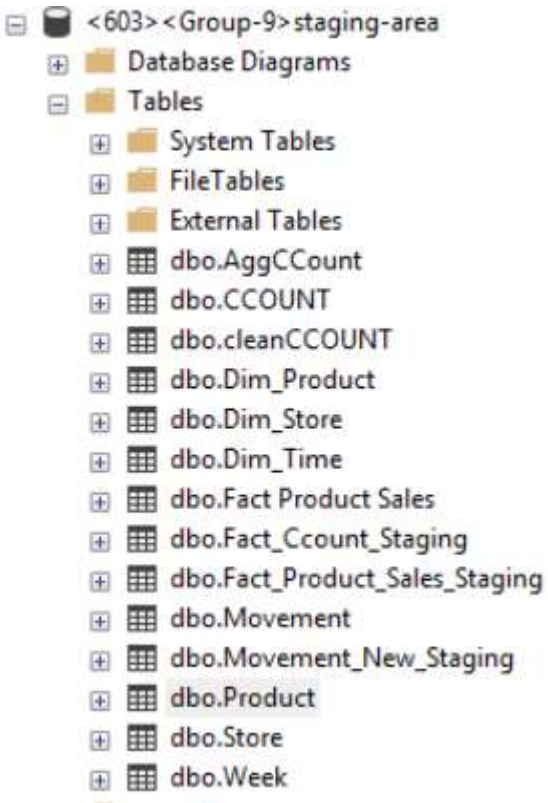
```
CREATE TABLE [dbo].[Fact_Product_Sales] (
[MOVE] int,
[QTY] int,
[PRICE] float,
[Total_Sales] float,
[Product_Key] int NOT NULL,
[Time_Key] int NOT NULL,
[Store_Key] int NOT NULL,
PRIMARY KEY ([Product_Key],[Time_Key],[Store_Key]),
FOREIGN KEY (Product_Key) REFERENCES Dim_Product(Product_Key),
FOREIGN KEY (Time_Key) REFERENCES Dim_Time(Time_Key),
FOREIGN KEY (Store_Key) REFERENCES Dim_Store(Store_Key)
)
```

Ccount Fact table

```
CREATE TABLE [dbo].[Fact_Ccount] (
[Bakery_Sales] float,
[Cheese_Sales] float,
[Time_Key] int NOT NULL,
[Store_Key] int NOT NULL
PRIMARY KEY ([Time_Key], [Store_Key]),
FOREIGN KEY (Time_Key) REFERENCES Dim_Time(Time_Key),
FOREIGN KEY (Store_Key) REFERENCES Dim_Store(Store_Key)
)
```

5.4.7 Removal of temporary tables created in the staging area

The temporary tables that were created during the cleansing and transformation process have been removed from the staging area. The below snapshot shows the tables that were retained in the staging area.



6. BI Reporting

6.1 Reporting Plan

Business Intelligence refers to applications, tools, practices that help compile and analyze data sets for providing actionable insights to business executives to bolster the process of strategic decision making. Furthermore, the different tools, applications and methodologies of BI enable collecting of data from external as well as internal data sources, preparing it for analysis, developing and running queries on the dataset, creating reports, and enabling creation of visualizations and dashboards to help corporate executives as well as employees of a firm.

In our analysis of data, we will be using SSAS, SSRS, SSAS+SSRS, and Report Builder 3.0 in this project with the following breakdown of each reports:

Reporting Tool	Question Number
SSRS	Question 1
SSRS	Question 2
SSAS	Question 3
SSRS+SSAS	Question 4
Report Builder 3.0	Question 5

6.1.1 Determination of all target reports that satisfy business questions of the project

Question 1: What has been the trend in sales of cookies over the years during Thanksgiving?

Tool Used: SSRS alone.

To answer this business question, we used the category cookies from product dimension Dim_Product, Week_Number and Special Event as "Thanksgiving" from the dimension table Dim_Time and we mapped them with Fact_Product_Sales. We then used a SQL query along with SSRS for generating a report to showcase trend in cookie sales during Thanksgiving.

The report uses a line chart to showcase a trajectory that represents a trend in cookie sales during Thanksgiving weeks over the years.

Question 2: Which zone had the maximum number of Beer sales during New Year week of 1997?

Tool Used: SSRS

To answer this question, we used the category of Beer in product dimension Dim_Product with time dimension Dim_Time providing us the special event of "New Year" and week number.We used store dimension Dim_Store to get stores in different zones who sold beer. We derived beer sales from the fact table Fact_Product_Sales. The data derived, represented sale of Beer during New Year week of 1997. We then ran a query along with SSRS to generate a visual representation of beer sales during New Year week of 1997 across different zones. The report comprises of a bar chart to give sales in Zone A to Zone F during New Year week of 1997.

Question 3: How will the cheese sales be over the next three years based on the data available from past three years?

Tool Used: SSAS

To answer this question, we used the Customer Count Data Mart wherein we had stored cheese sales in fact table Fact_Ccount. The cheese sales were mapped across time with the help of Dim_Time as we derived year from it. A cube with Fact_Ccount and Dim_Time were deployed to get a data set. We used the "Forecast" function in "Analyze" section in excel to forecast cheese sales over the next three years. The report comprises of a bar chart that shows cheese sales for years 1993,1994,1995 along with forecast for cheese sales in 1996,1997,and 1998.

Questions 4: Compare the bakery sales during Easter and 4th of July 1996 across all the stores?

Tools Used: SSAS+SSRS

To answer this business question, we derived Bakery sales from fact table Fact_Ccount along with store information from Dim_Store and time information(Special Event and Week Number) from Dim_Time. We created a cube comprising of Fact_Ccount, Dim_Store, and Dim_Time and deployed it on the server. After successful deployment, we used SSRS on the derived data set to generate a visual report depicting bakery sales during Easter and 4th of July. The report comprises of a line chart that plots two lines showcasing the trend of bakery sales with one line representing Easter and the other representing 4th of July.

Question 5: How is Analgesics performing in different Price Tiers and Zones?

Tools Used: Report Builder 3.0

To answer this question, we used the Product Sales Data Mart. We used Category_Name from Dim_Product to map it across price tiers and zones from Dim_Store. We got the total sales for Analgesics from our fact table Fact_Product_Sales. We built the required design query in Report Builder and arranged fields as per our convenience to derive a visual representation of Analgesics' performance across different zones and price tiers. The report generated was a bar chart showing performance of Analgesics in Zone A, B, and C. After drilling down, we could then see performance across various Price Tiers as well.

Mappings from independent Data Marts

Question 1: What has been the trend in sales of cookies over the years during Thanksgiving?

Source	Attributes
Dim_Time	Time_Key Week_Number Special_Event
Dim_Product	Product_Key Category_Name

Dim_Store	Store_Key
Fact_Product_Sales	Total_Sales

Report	Category_Name
	Total_Sales

Question 2: Which zone had the maximum number of Beer sales during New Year week of 1997?

Source	Attributes
Dim_Time	Time_Key Special_Event Year
Dim_Product	Product_Key Category_Name
Dim_Store	Store_Key Zone
Fact_Product_Sales	Total_Sales

Report	Zone
	Total_Sales

Question 3: How will the cheese sales be over the next three years based on the data available from past three years?

Source	Attributes
Dim_Time	Time_Key Year
Dim_Store	Store_Key
Fact_Ccount	Cheese_Sales

Report	Year
	Cheese_Sales

Questions 4: Compare the bakery sales during Easter and 4th of July 1996 across all the stores?

Source	Attributes
Dim_Time	Time_Key Year Special_Event Week_Number
Dim_Store	Store_Key Store_Number
Fact_Ccount	Bakery_Sales

Report	Year
	Bakery_Sales

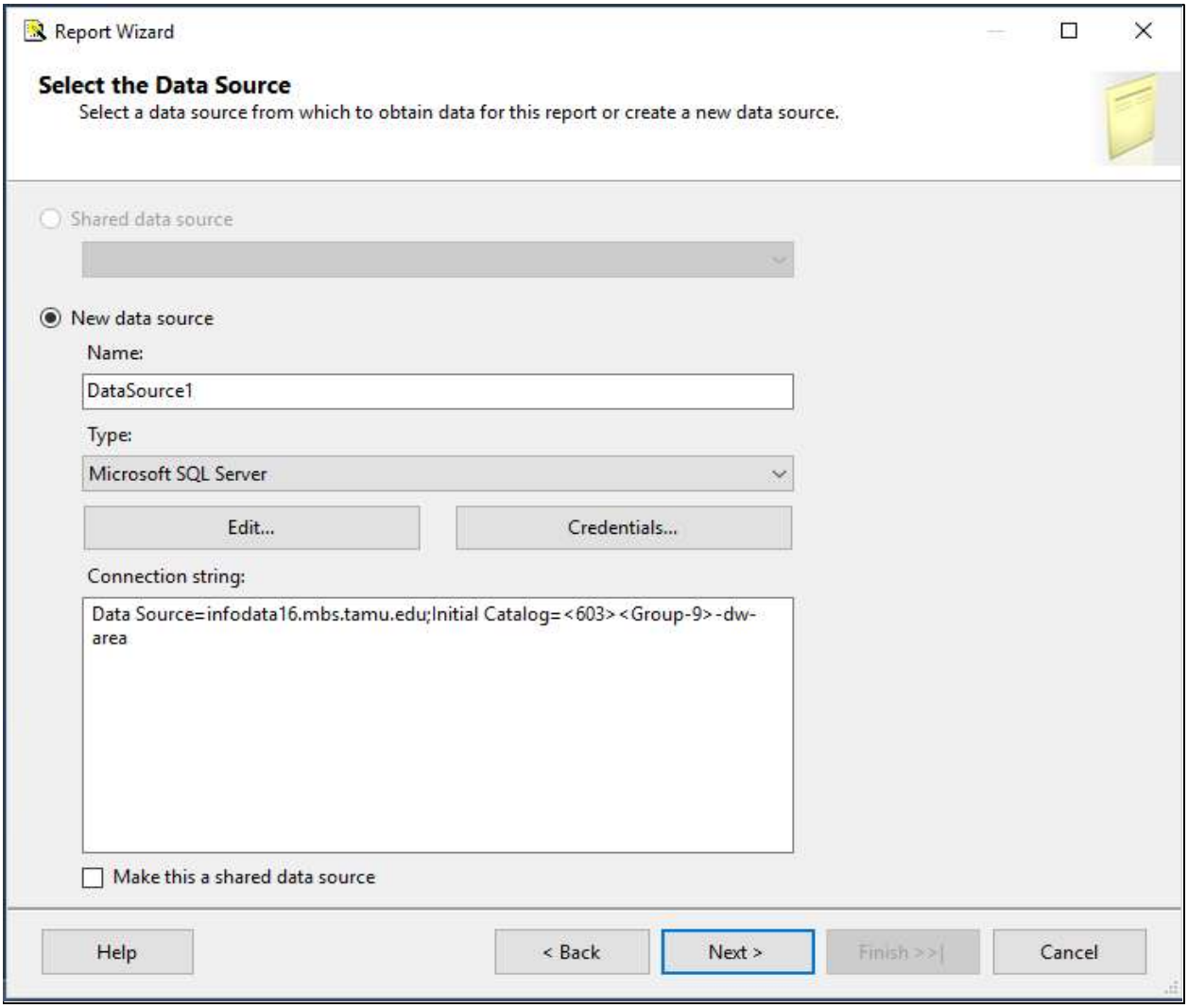
Question 5: How is Analgesics performing in different Price Tiers and Zones?

Source	Attributes
Dim_Time	Time_Key Year
Dim_Product	Product_Key Category_Name
Dim_Store	Store_Key Store_Number Price_Tier Zone
Fact_Product_Sales	Total_Sales

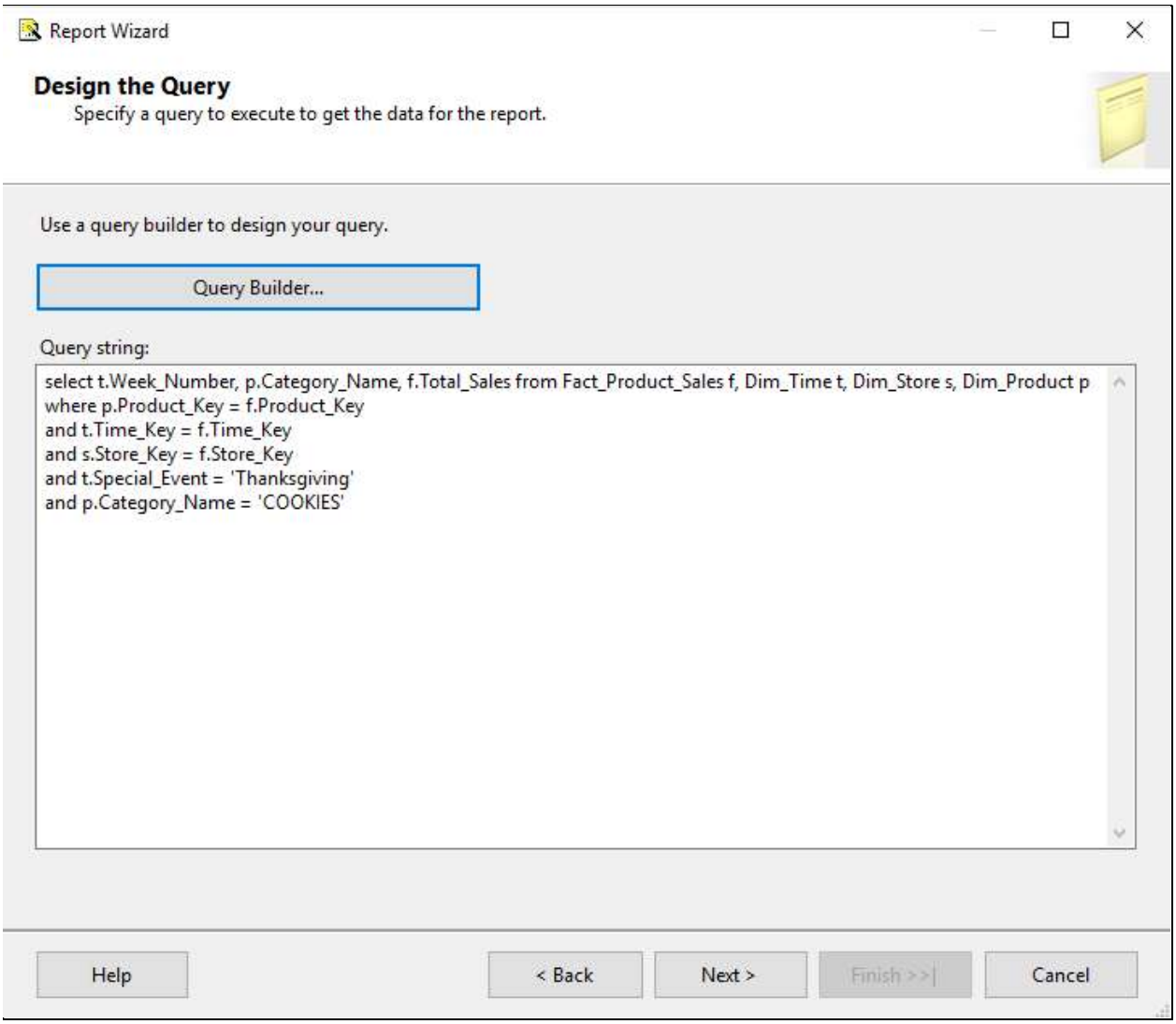
Report	
	Zone
	Price_Tier
	Store_Number
	Total_Sales

6.2 Report built from individual data marts using SSRS for Question 1

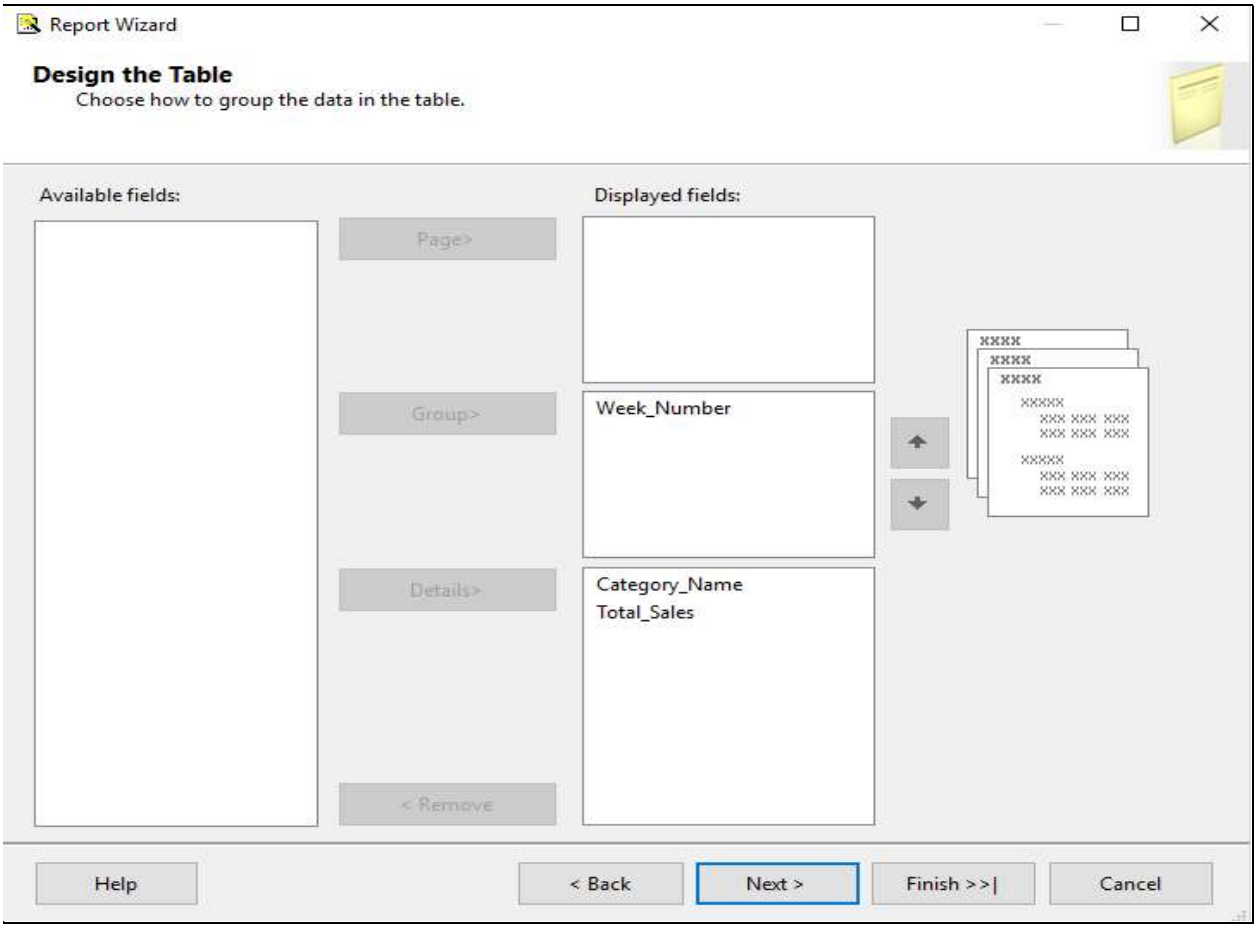
Data source:



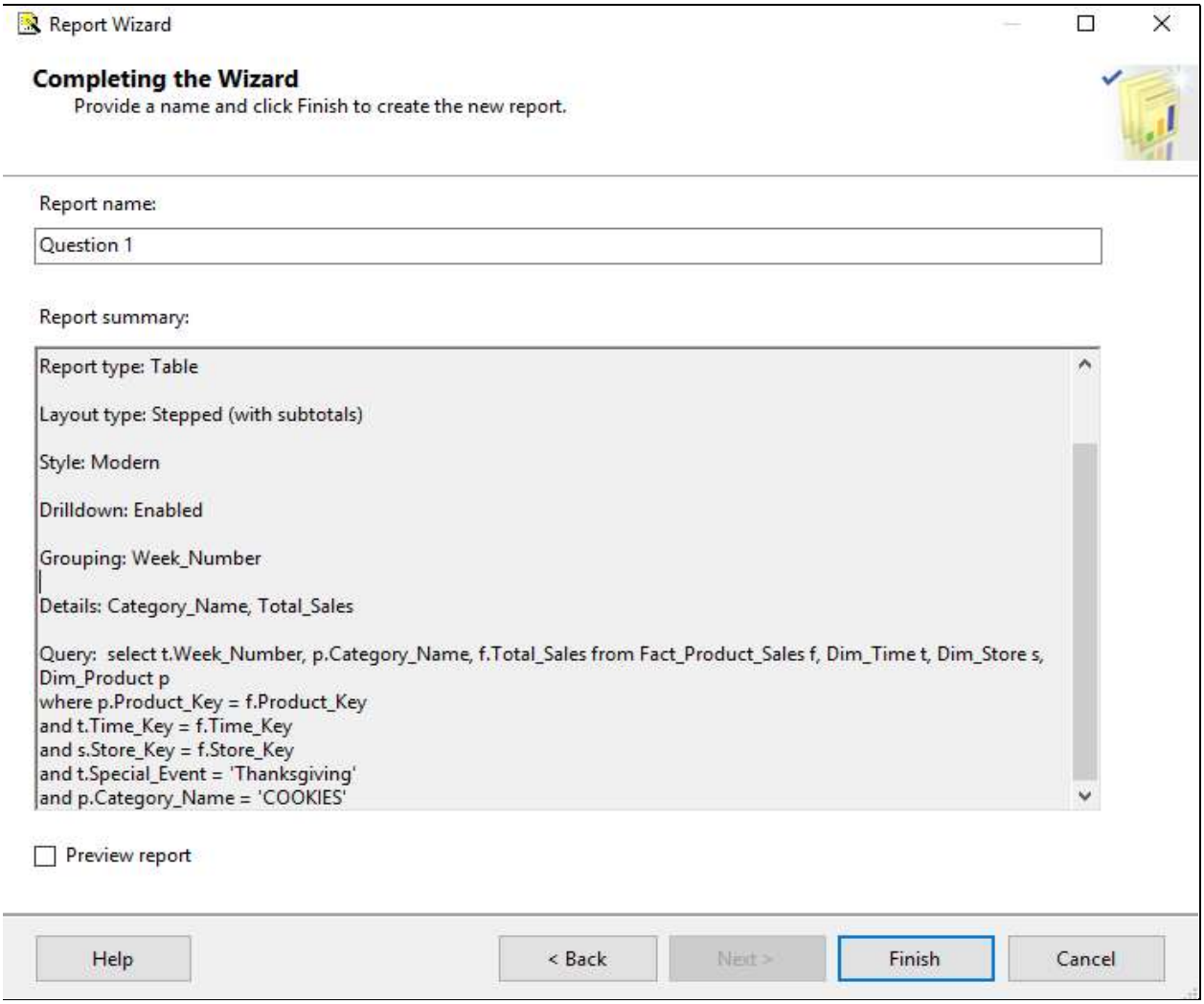
Query used in report:


Report Wizard
Design the Query
Specify a query to execute to get the data for the report.
Use a query builder to design your query.
Query Builder...
Query string:
select t.Week_Number, p.Category_Name, f.Total_Sales from Fact_Product_Sales f, Dim_Time t, Dim_Store s, Dim_Product p
where p.Product_Key = f.Product_Key
and t.Time_Key = f.Time_Key
and s.Store_Key = f.Store_Key
and t.Special_Event = 'Thanksgiving'
and p.Category_Name = 'COOKIES'
Help
< Back
Next >
Finish >>|
Cancel

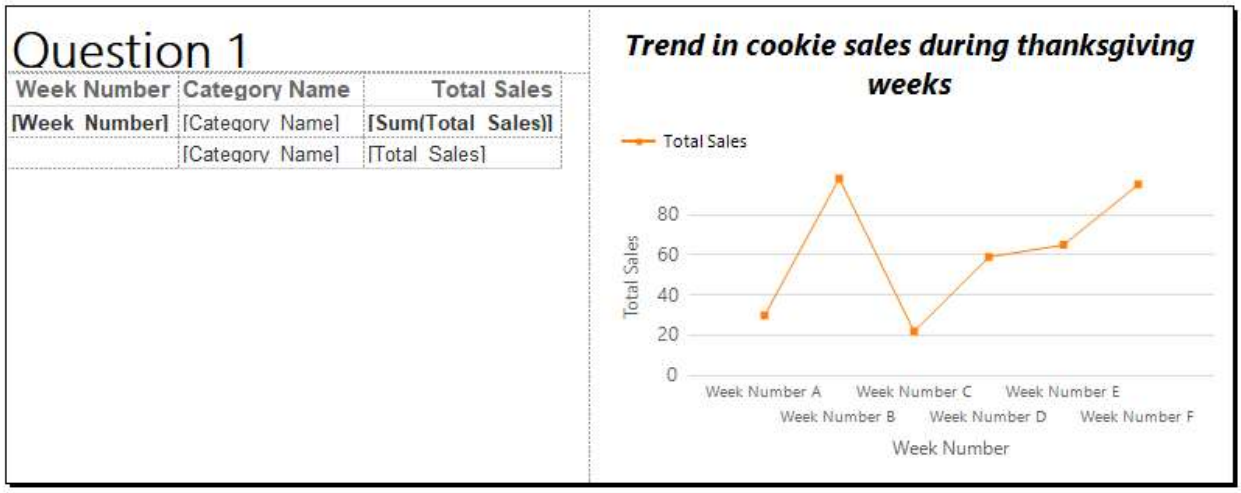
Grouping the total sales data according to Week number:



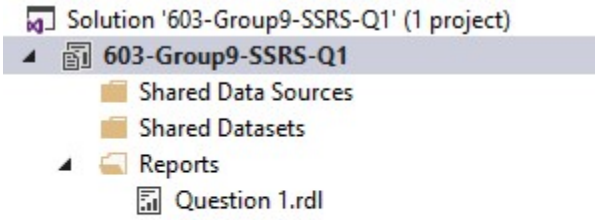
Report Summary:



Report design:



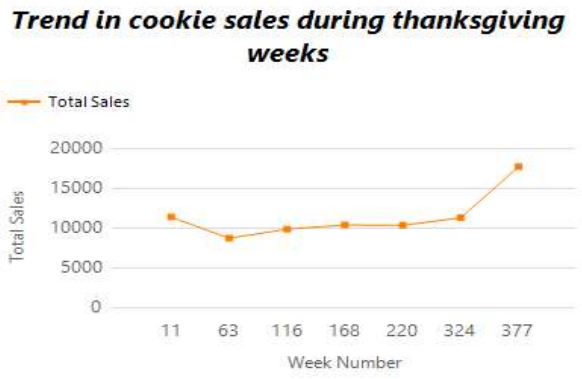
Final Report Deployment:



Deployed report before drilldown:

Question 1

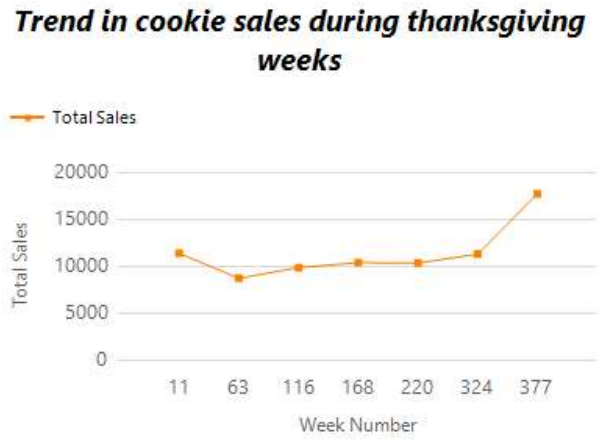
Week Number	Category Name	Total Sales
⊞ 11	COOKIES	11343.76
⊞ 63	COOKIES	8692.67000000006
⊞ 116	COOKIES	9838.4199999999
⊞ 168	COOKIES	10356.27
⊞ 220	COOKIES	10318.92
⊞ 324	COOKIES	11264.3799999999
⊞ 377	COOKIES	17704.8399999999



Deployed report after drilldown:

Question 1

Week Number	Category Name	Total Sales
⊞ 11	COOKIES	11343.76
⊞ 63	COOKIES	8692.67000000006
⊟ 116	COOKIES	9838.4199999999
	COOKIES	5.98
	COOKIES	2.99
	COOKIES	5.98
	COOKIES	8.97
	COOKIES	2.99
	COOKIES	8.97
	COOKIES	2.99
	COOKIES	5.98
	COOKIES	2.99
	COOKIES	20.93
	COOKIES	5.98
	COOKIES	5.98
	COOKIES	8.97
	COOKIES	2.99
	COOKIES	0
	COOKIES	0
	COOKIES	2.99
	COOKIES	2.99
	COOKIES	2.99
	COOKIES	2.99
	COOKIES	14.95
	COOKIES	8.97
	COOKIES	8.97
	COOKIES	8.97
	COOKIES	5.98
	COOKIES	8.97

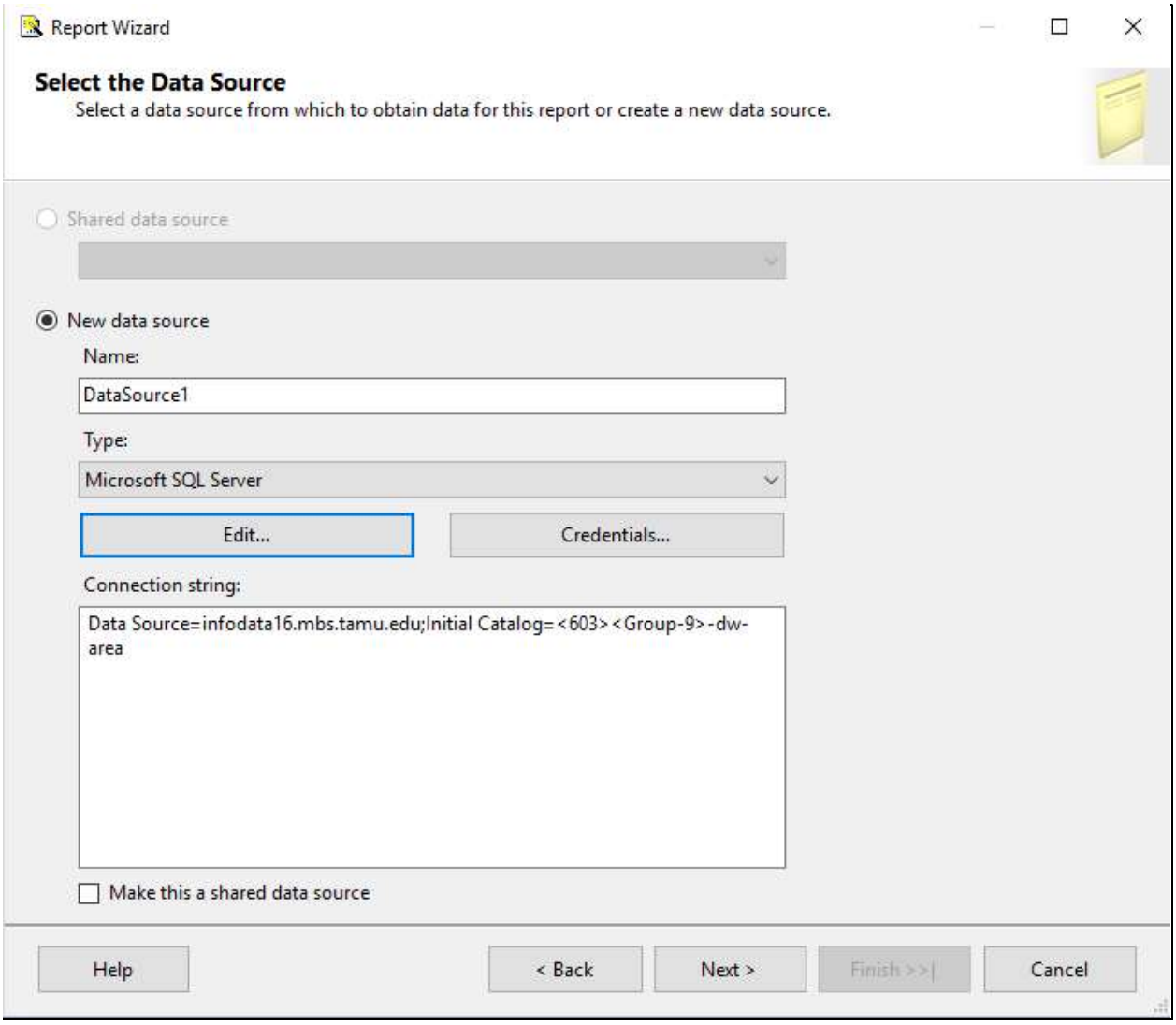


Conclusion:

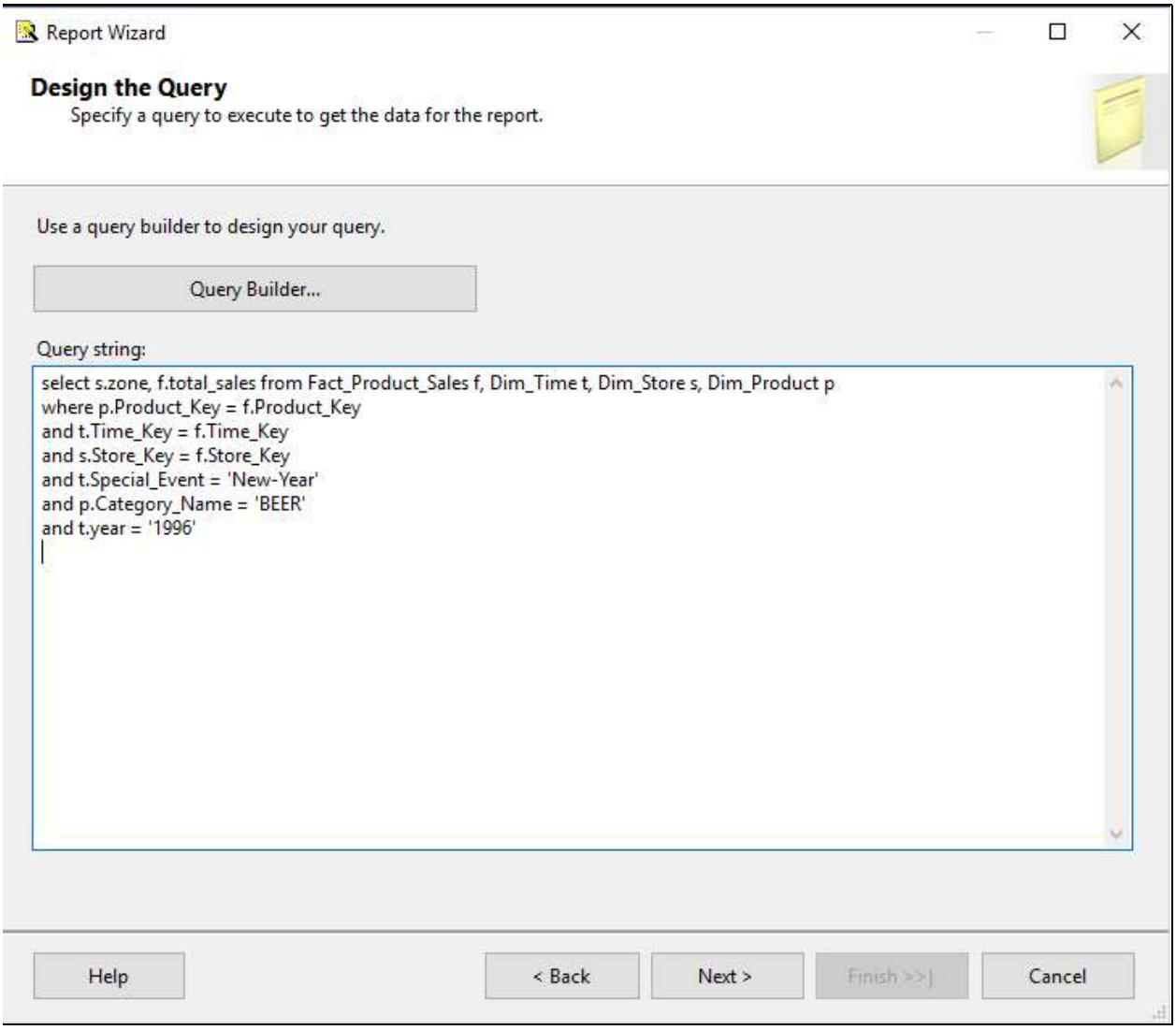
From the report, we can clearly see that the sales of cookies during thanksgiving has been seeing an increasing trend. DFF can make use of the same strategy (be it marketing or pricing) for other products that they wish performed well during this holiday season. Using, SSRS and data warehouse, we were able to clearly answer this business question.

6.3 Report built using SSRS for Question 2

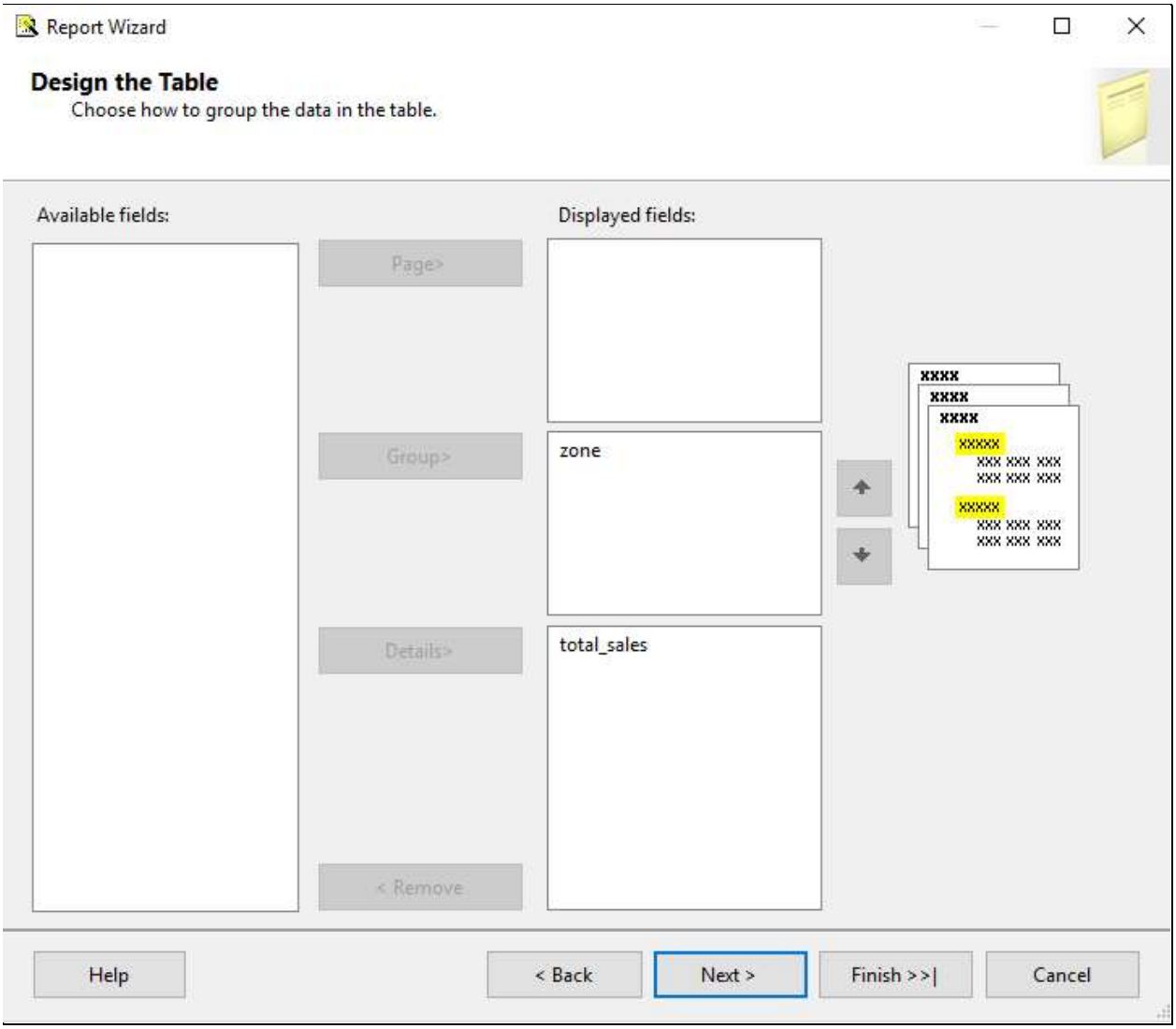
Data Source Selection:



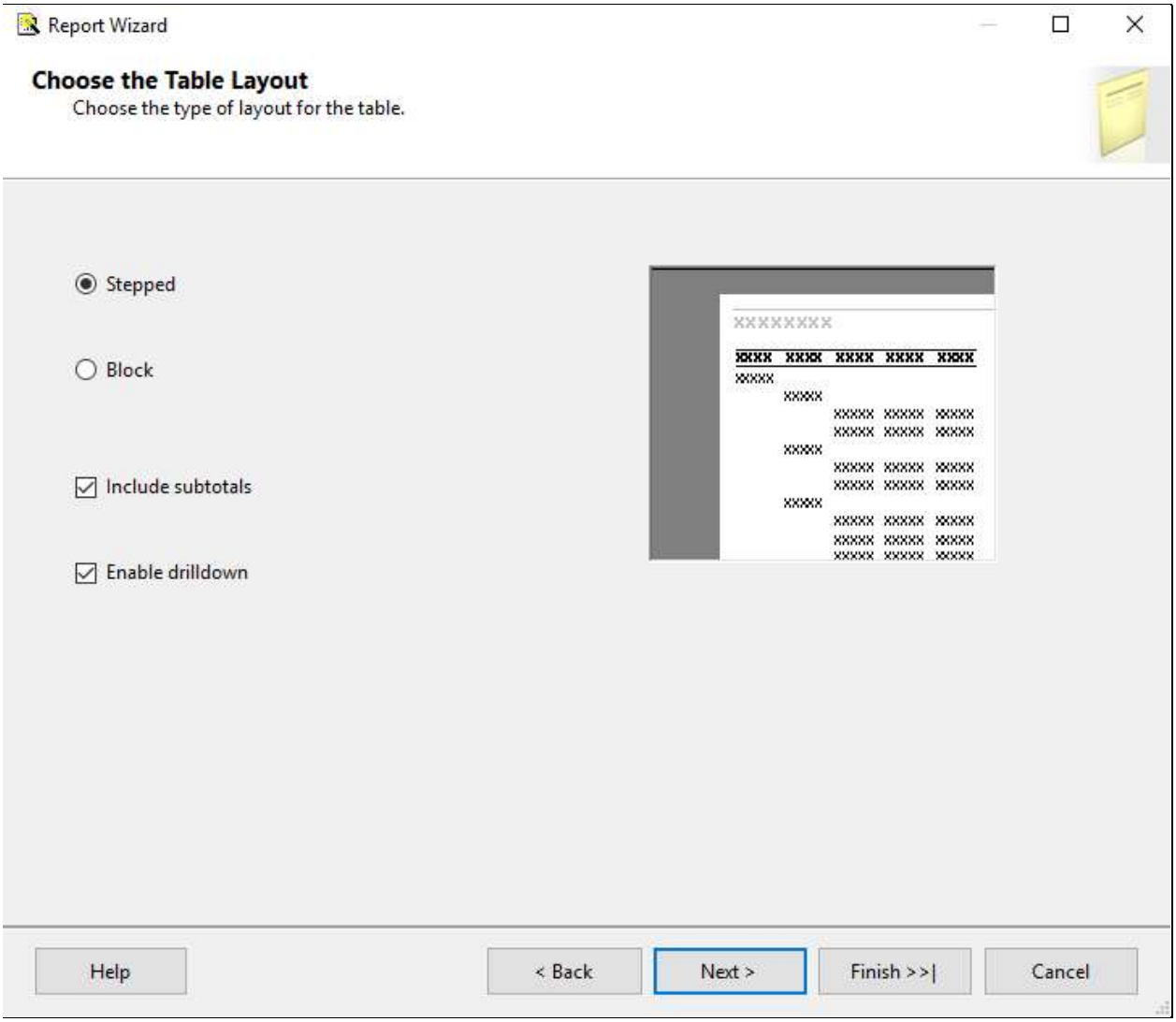
Query used for report:



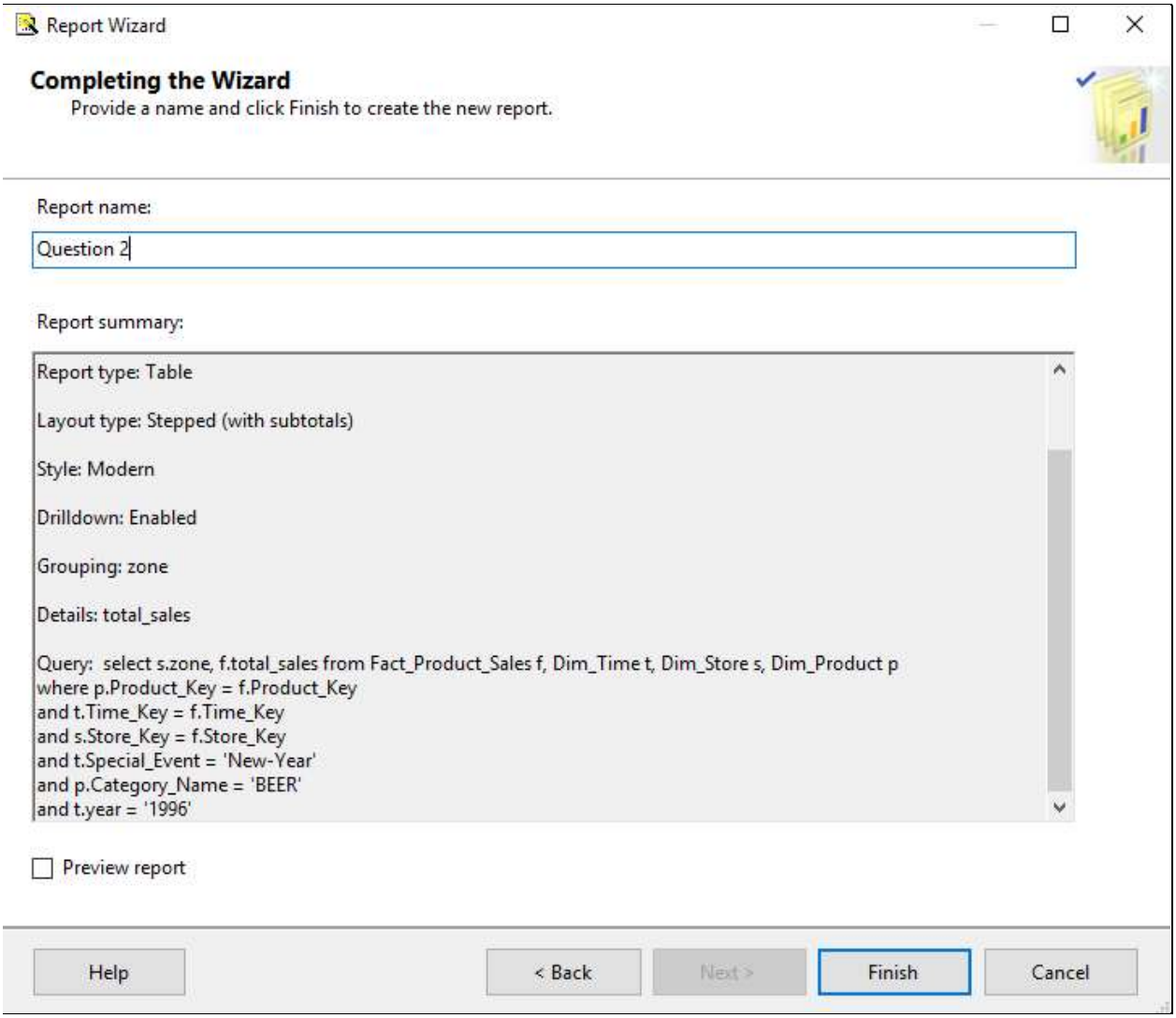
Grouping of total sales based on zone:



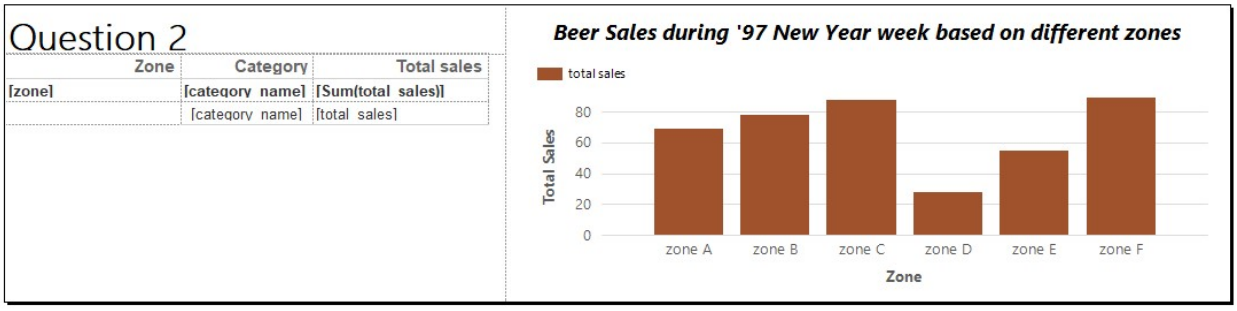
Enable drilldown and subtotals:



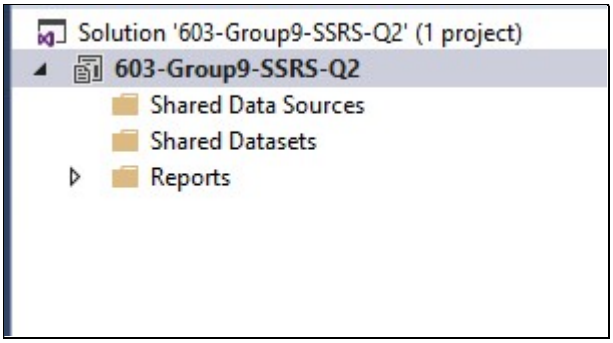
Report Summary:



Report design:



Final Report deployed:



infodata16.mbs.tamu.edu/ReportServer - /603-Group9-SSRS-Q2

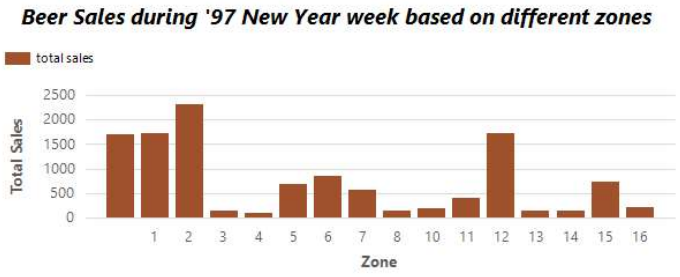
[To Parent Directory]
Friday, April 19, 2019 5:08 PM 37546 Question 2

Microsoft SQL Server Reporting Services Version 13.0.5081.1

Deployed report before drilldown:

Question 2

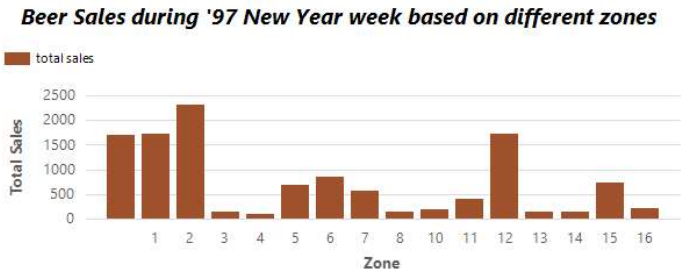
	Zone	Category	Total sales
⊞		BEER	1699.14
⊞1		BEER	1725.96
⊞2		BEER	2311.51
⊞3		BEER	136.66
⊞4		BEER	100.74
⊞5		BEER	684.66
⊞6		BEER	857.57
⊞7		BEER	562.11
⊞8		BEER	145.02
⊞10		BEER	190.14
⊞11		BEER	405.66
⊞12		BEER	1708.97
⊞13		BEER	133.16
⊞14		BEER	144.65
⊞15		BEER	724.08
⊞16		BEER	214



Deployed report after drilldown:

Question 2

Zone	Category	Total sales
⊞	BEER	1699.14
⊞1	BEER	1725.96
⊞2	BEER	2311.51
⊟3	BEER	136.66
	BEER	15.96
	BEER	48.86
	BEER	13.98
	BEER	0
	BEER	0
	BEER	0
	BEER	5.98
	BEER	11.96
	BEER	39.92
⊟4	BEER	100.74
	BEER	3.99
	BEER	41.88
	BEER	27.96
	BEER	0
	BEER	26.91
⊞5	BEER	684.66
⊞6	BEER	857.57
⊞7	BEER	562.11
⊞8	BEER	145.02
⊞10	BEER	190.14
⊞11	BEER	405.66
⊞12	BEER	1708.97
⊞13	BEER	133.16
⊞14	BEER	144.65
⊞15	BEER	724.08
⊞16	BEER	214

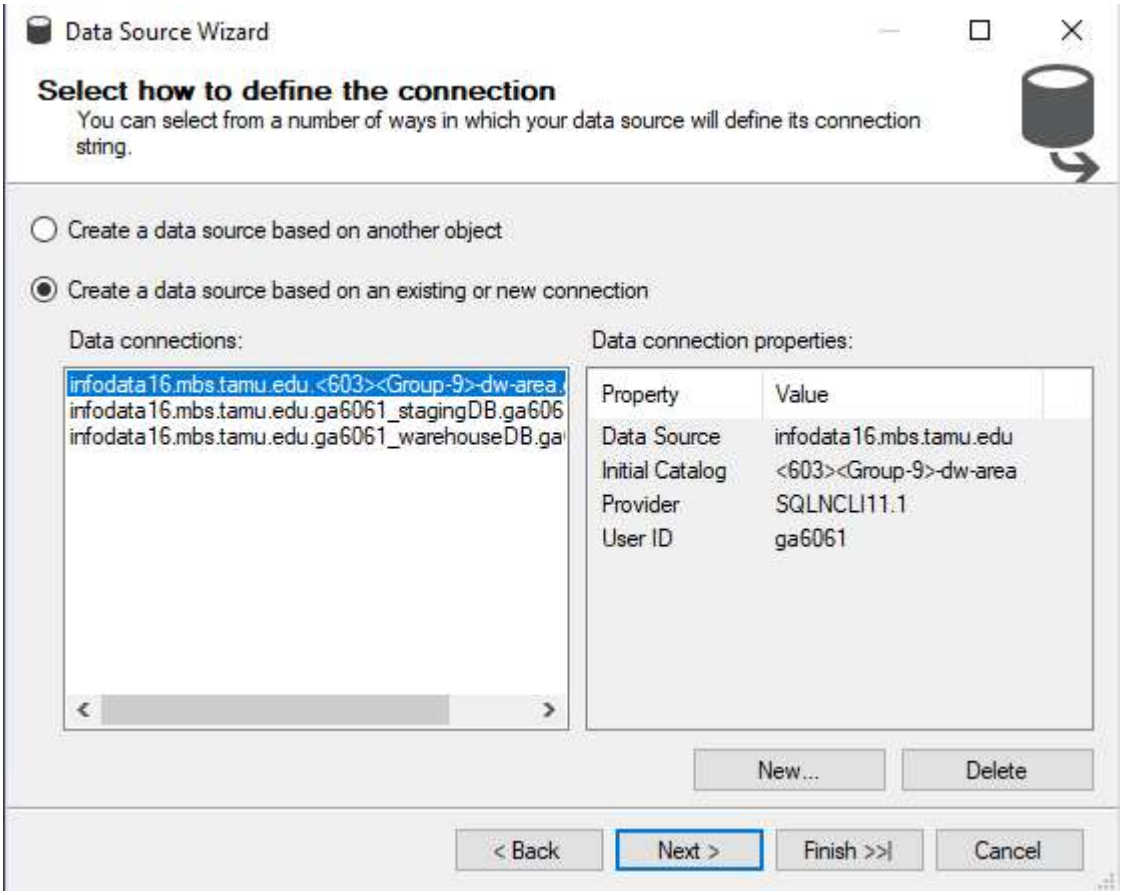


Conclusion:

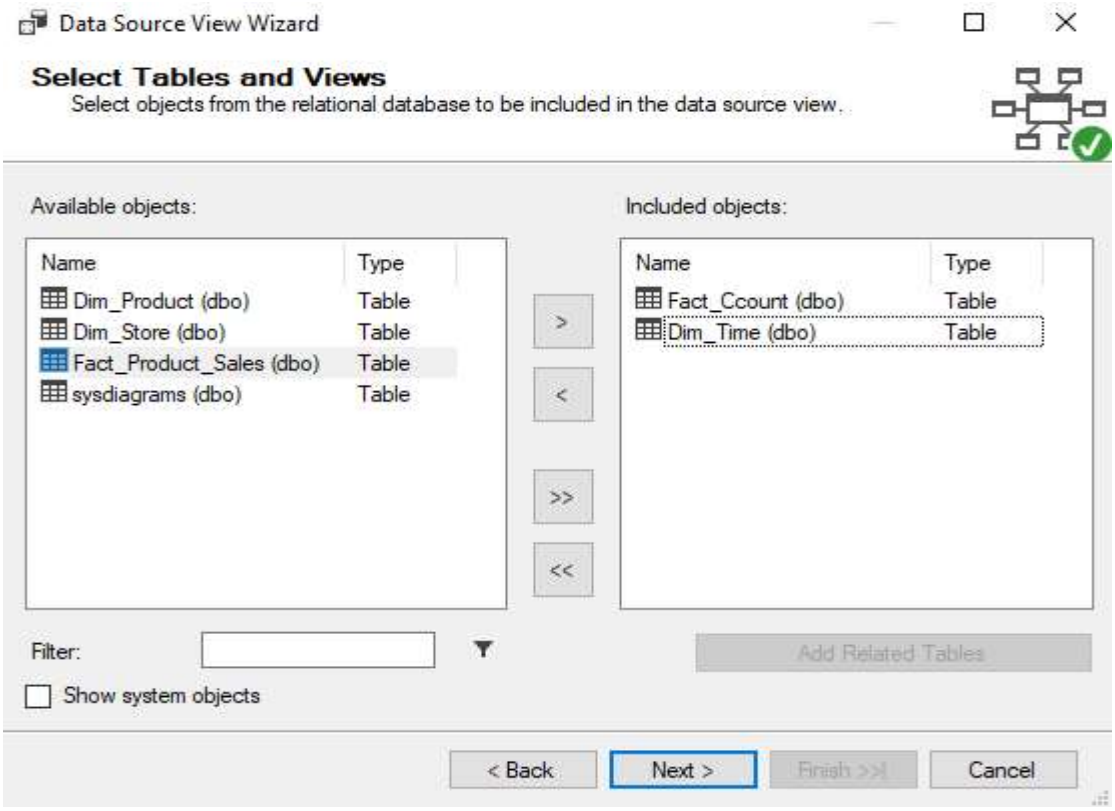
From this SSRS report, we see two specific zones – zone 1 and zone 2 that has performed really well when it comes to beer sales during the 1997 New Year's week. Deep dive analysis can be performed by DFF to circle in on what works best in these zones. The same sort of strategic decisions can then be implemented in other zones as well to boos beer sales.

6.4 Cubes from SSAS for Question 3

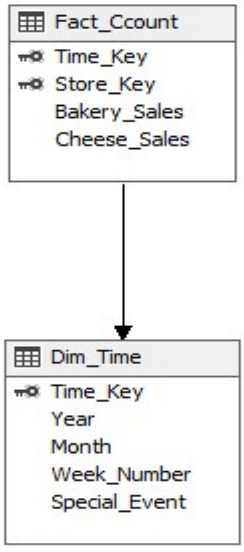
Data Source Connection:



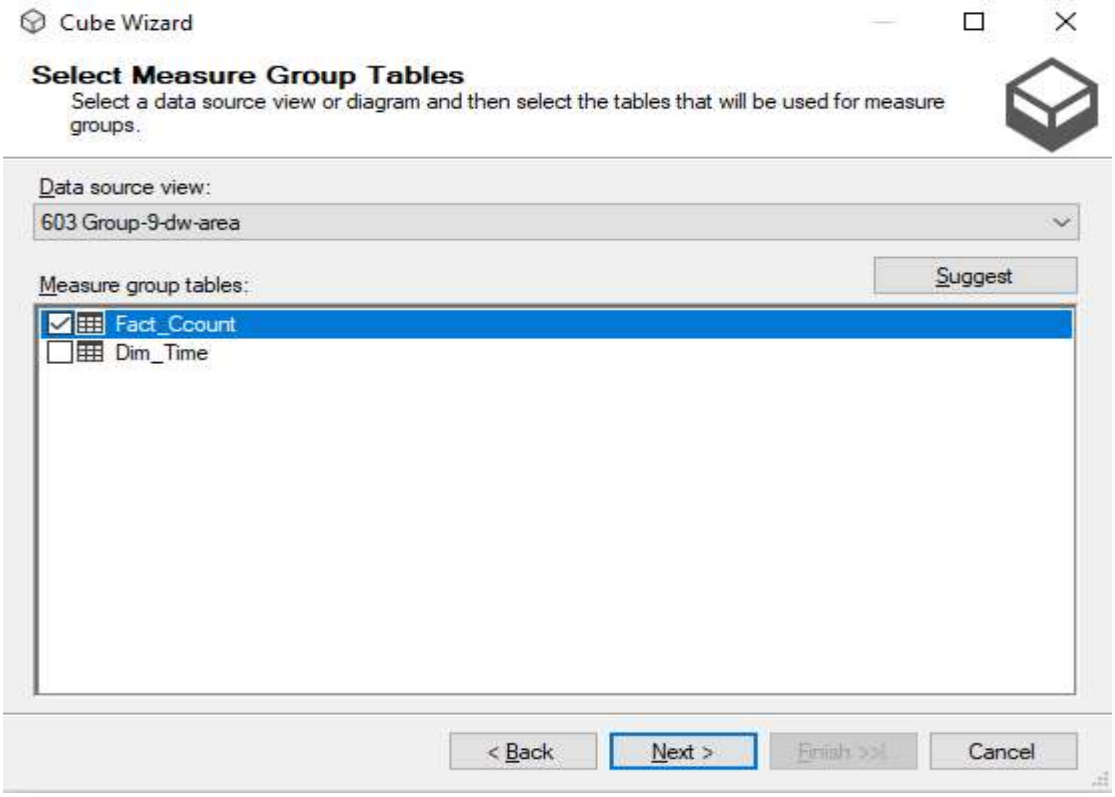
Creating a data source view by selecting the required fact and dimension tables:

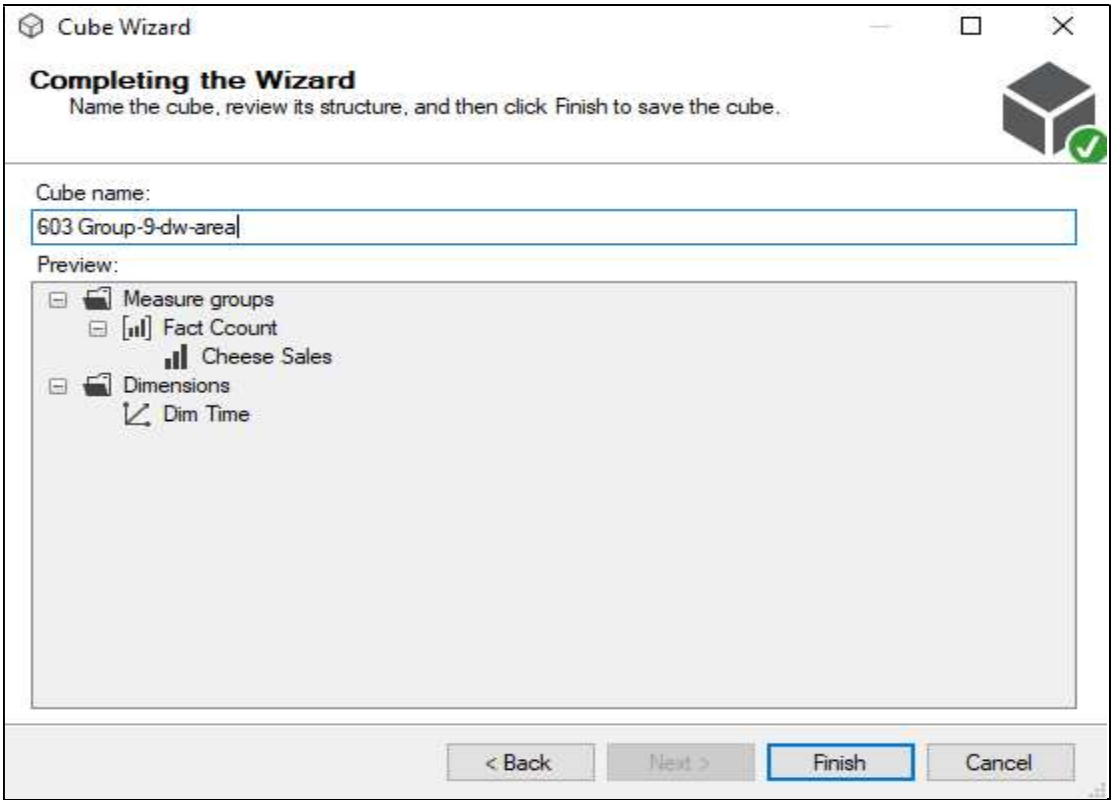


Data Source View:

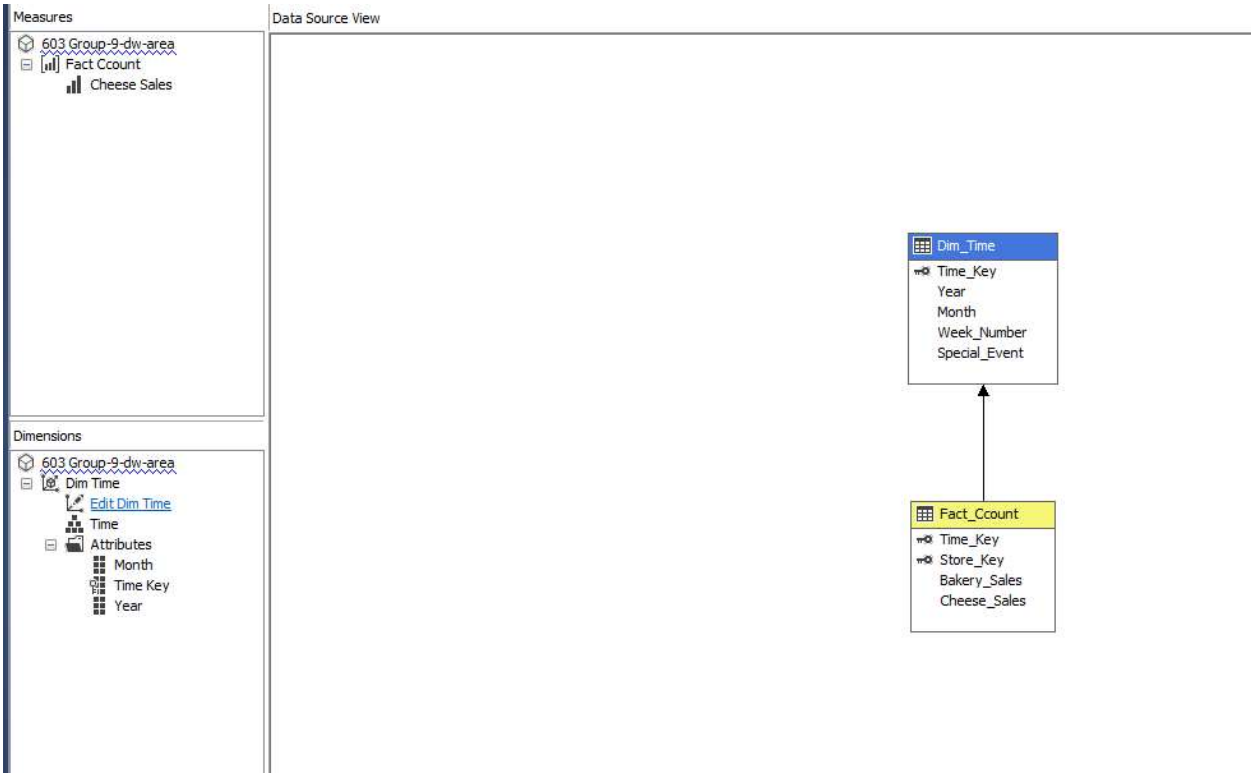


Selecting measure groups and dimensions to create the cube structure:

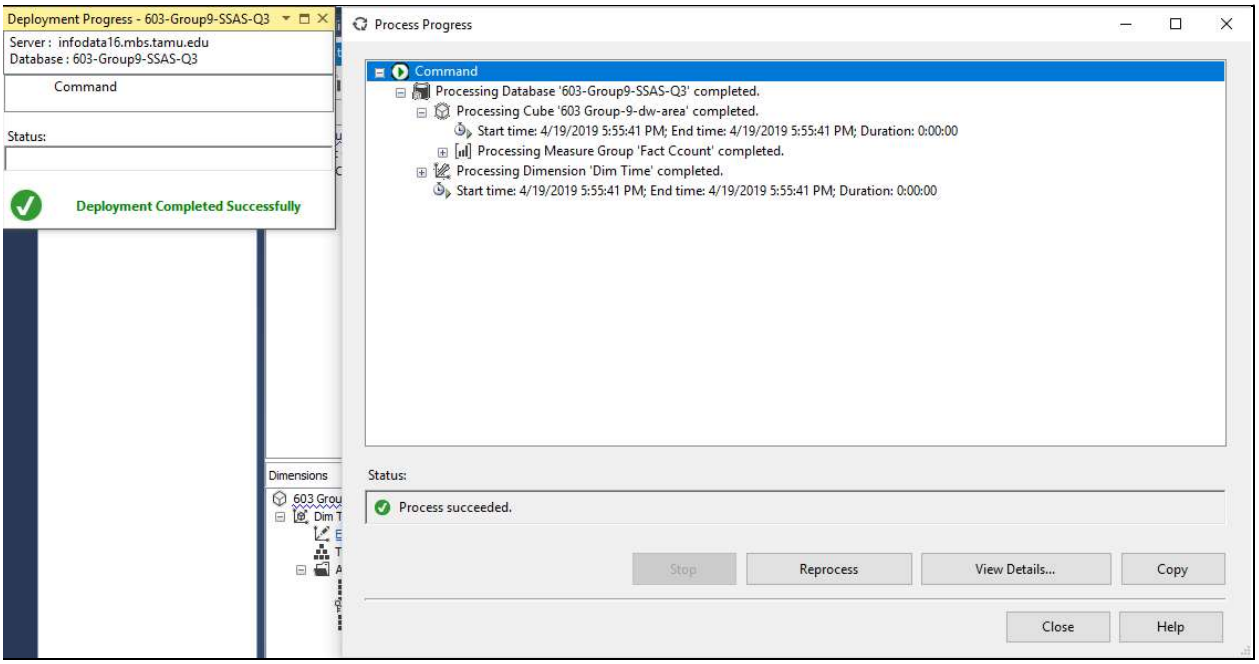


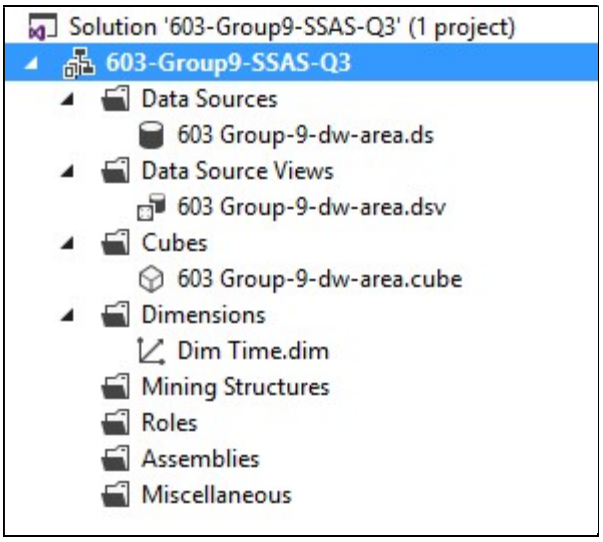


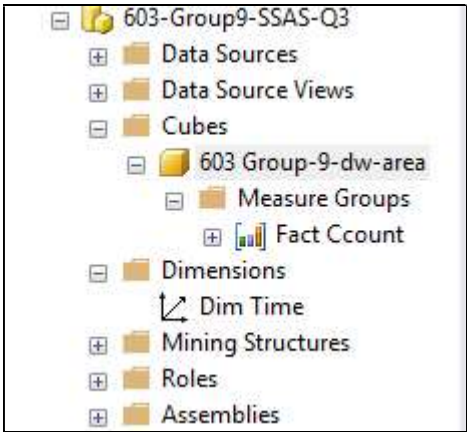
Cube structure:



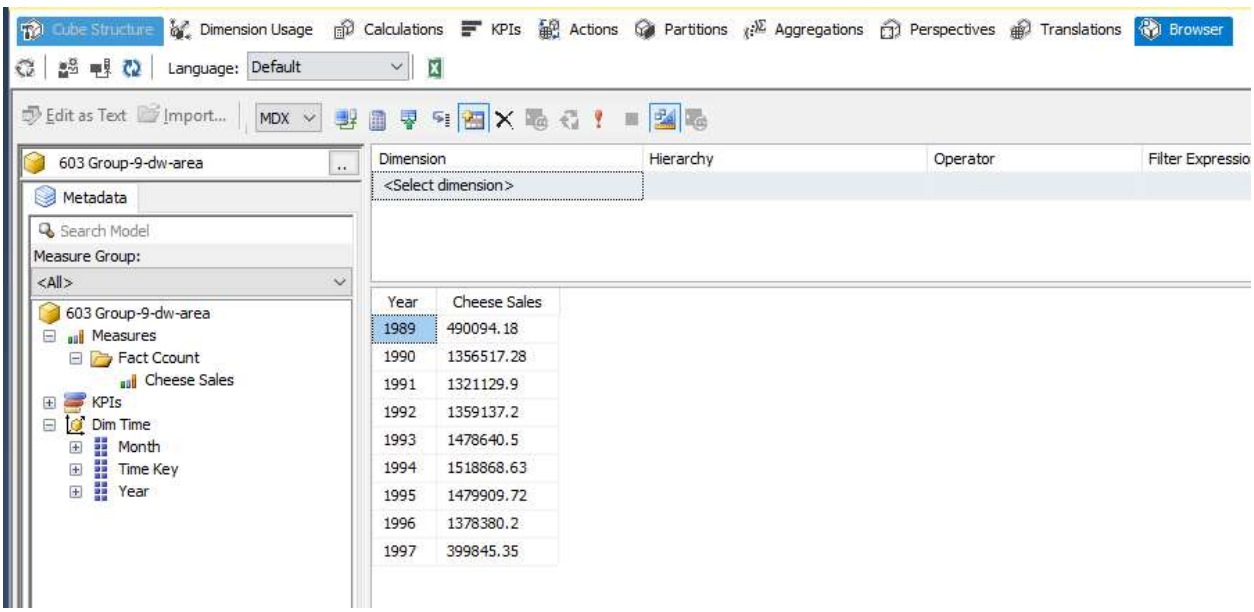
Successful process and deployment:





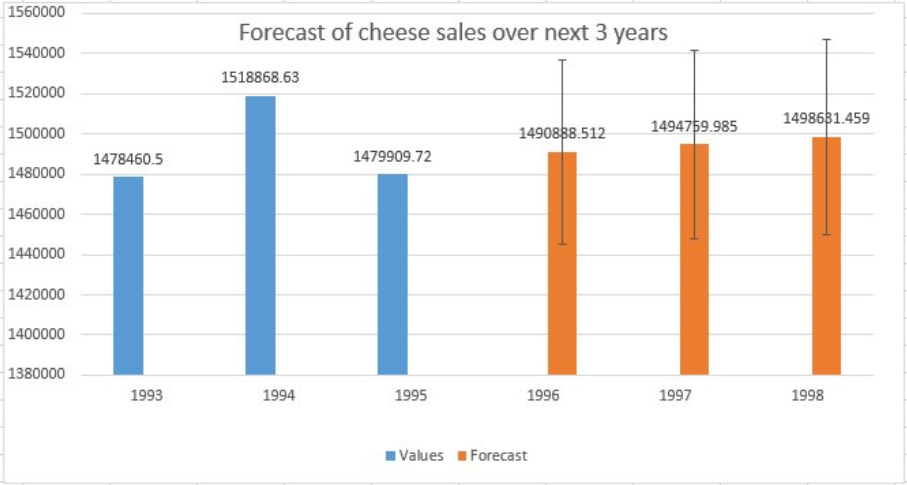


Cube browser:



Further data visualization and forecasting from SSAS cubes through Excel:

Row Labels	Cheese Sales
1989	490094.18
1990	1356517.28
1991	1321129.9
1992	1359137.2
1993	1478640.5
1994	1518868.63
1995	1479909.72
1996	1378380.2
1997	399845.35
Grand Total	10782522.96

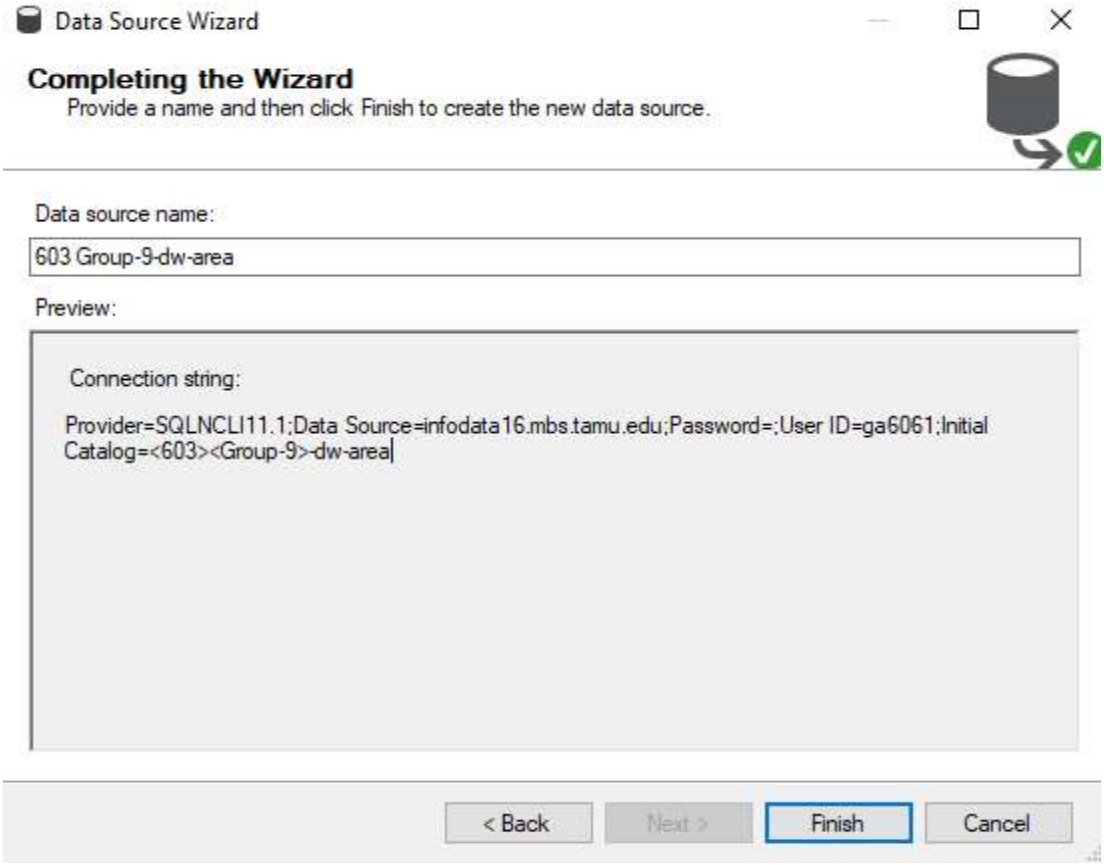


Conclusion:

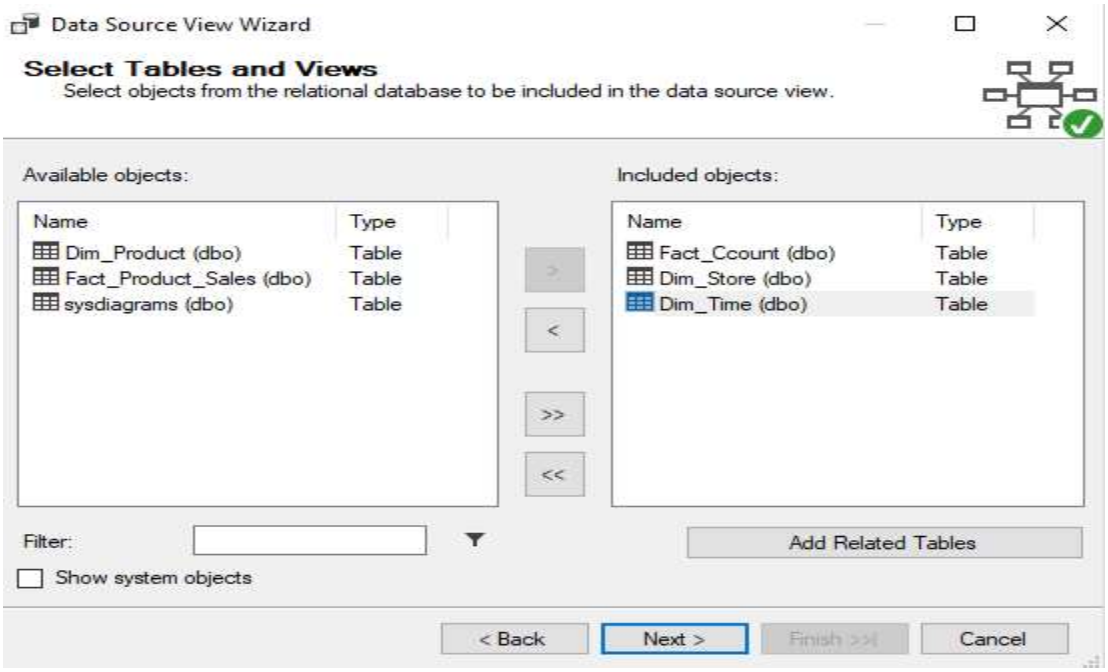
Cubes created using SSAS helps us in viewing the data from different perspectives. We can easily aggregate the cheese sales based on year to see the trend. Once the cubes were created, we used the 'Analyze in Excel' option in SSAS to further visualize the data as per requirement.

6.5 SSRS on top of SSAS for Question 4

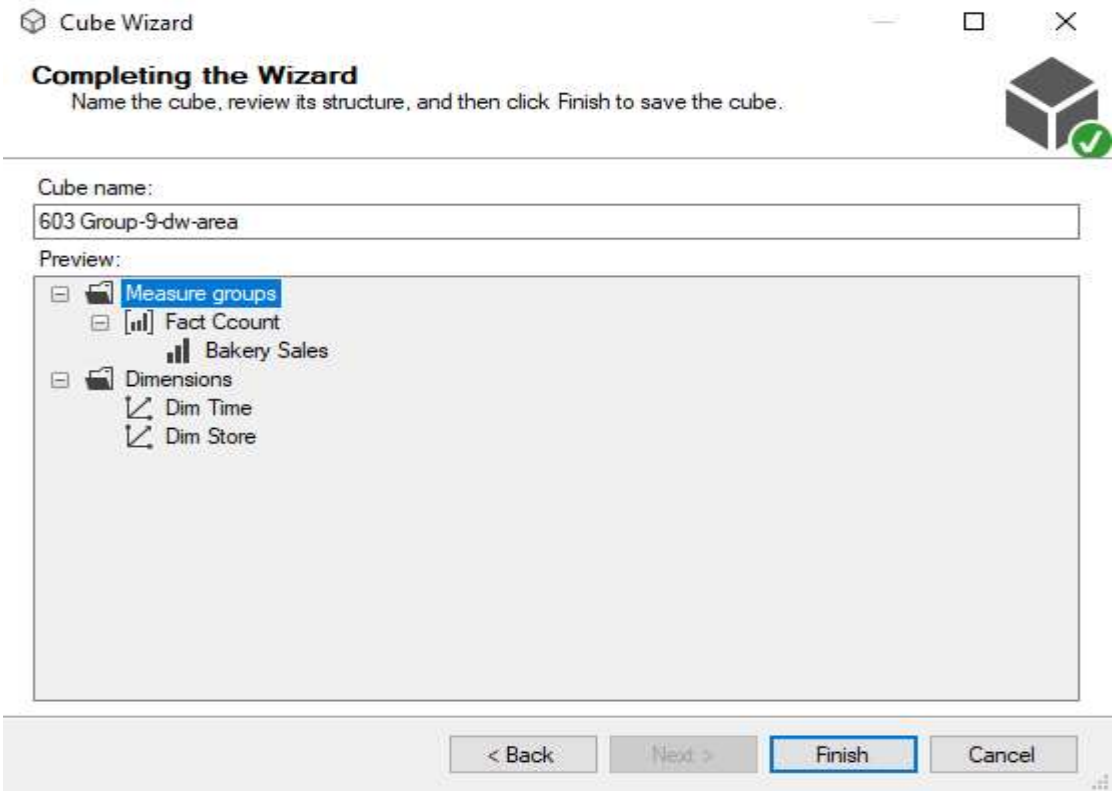
Data source selection:



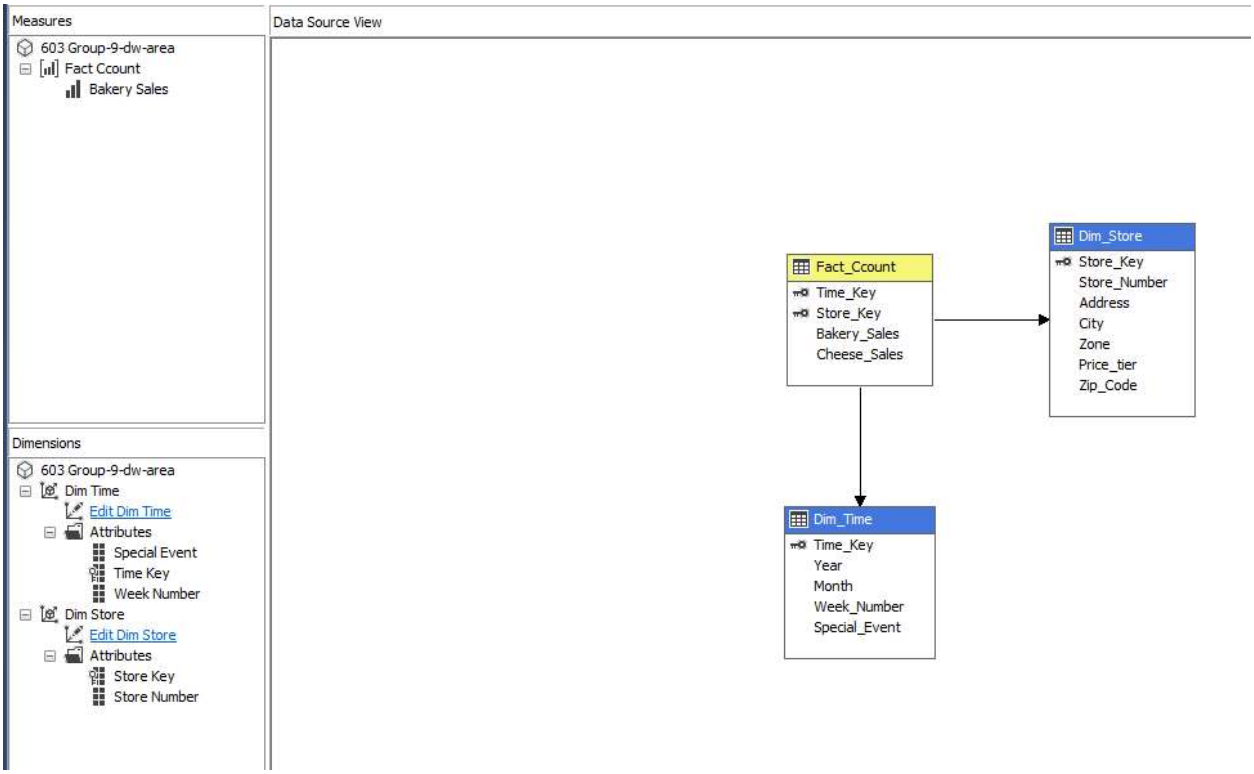
Creating data source view by selecting required fact and dimension tables:



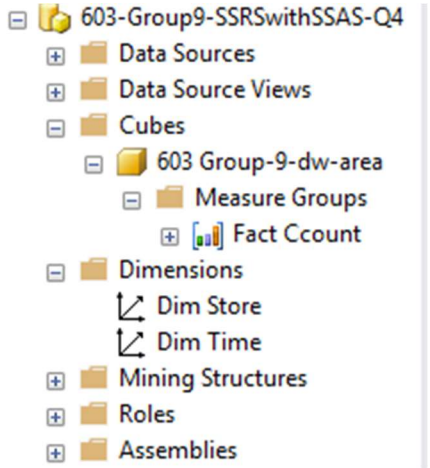
Creating cube structure by selecting measure groups and dimensions:



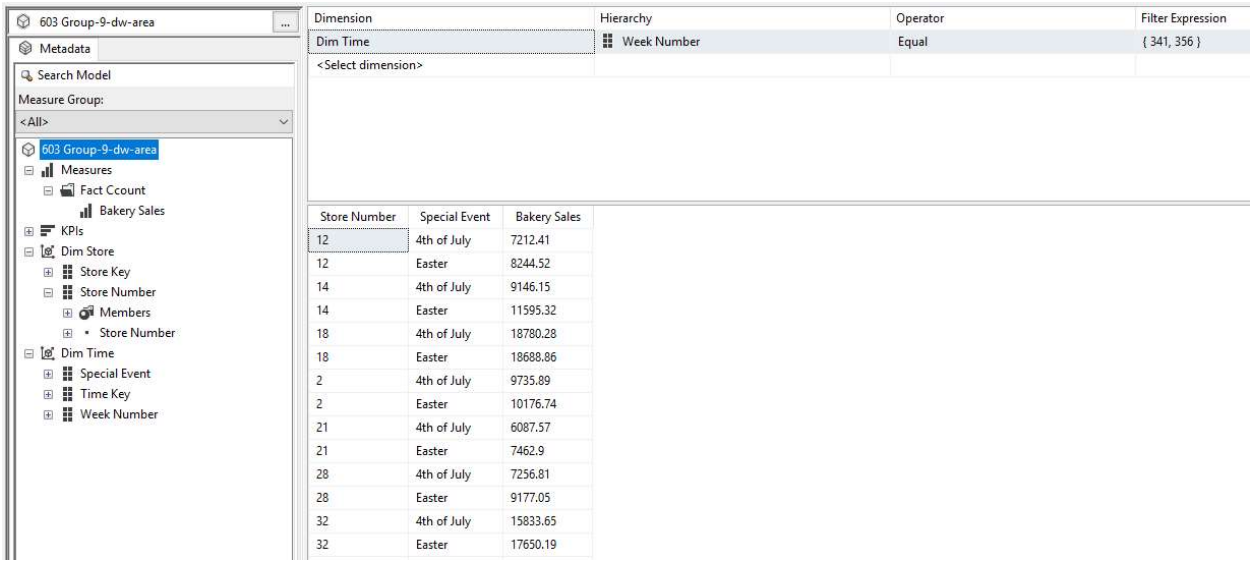
Cube structure:



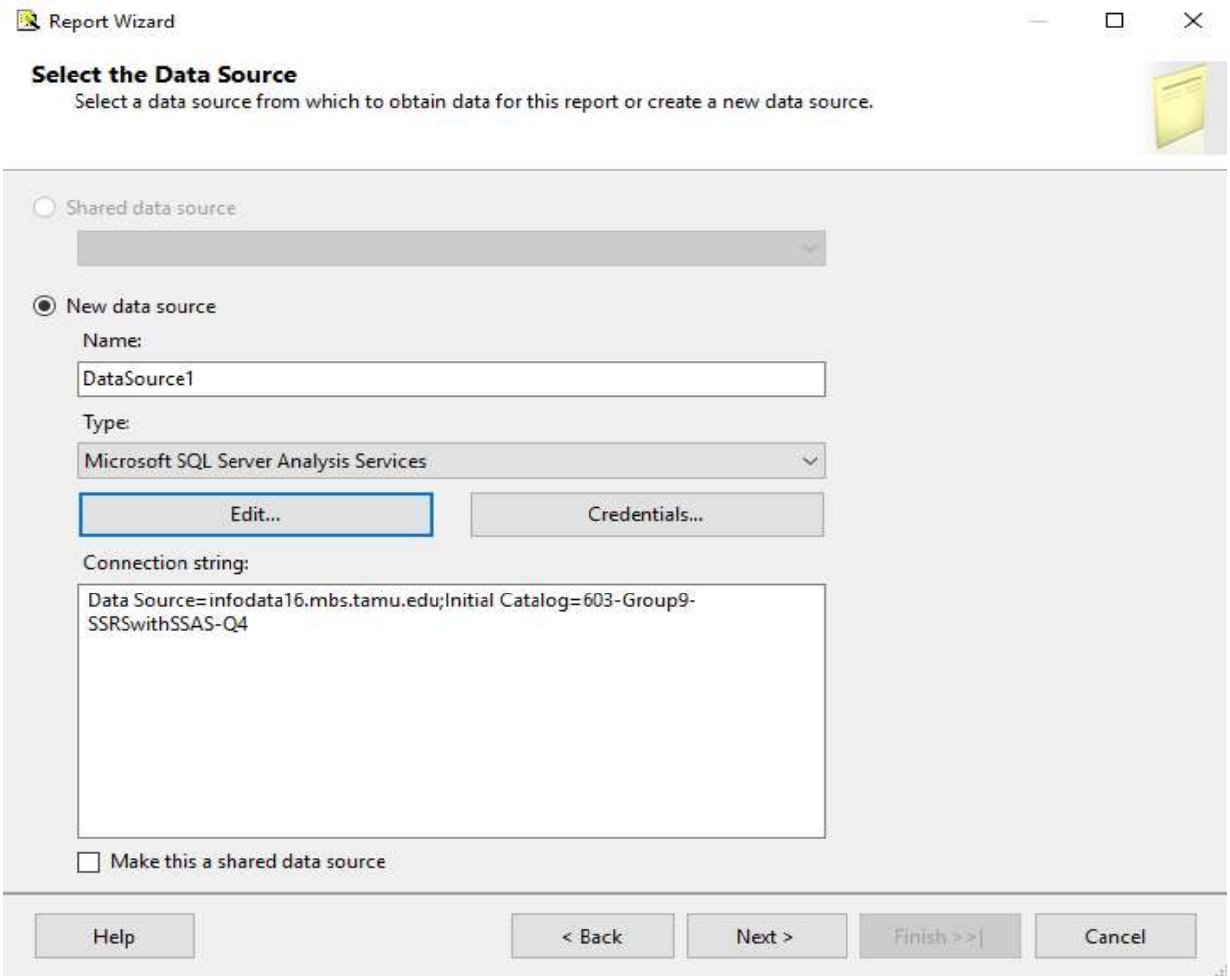
After successful deployment:



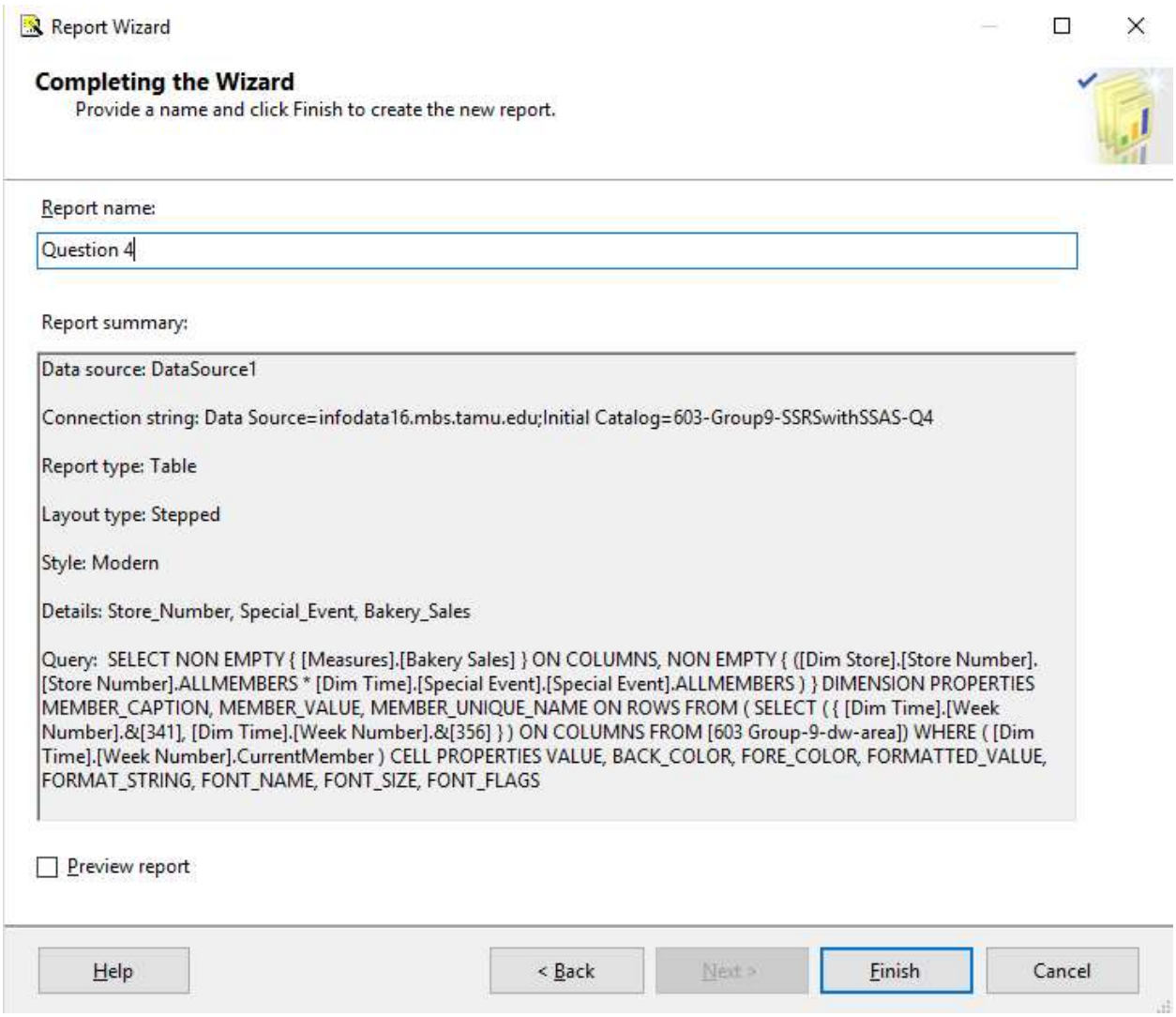
Cube browser:



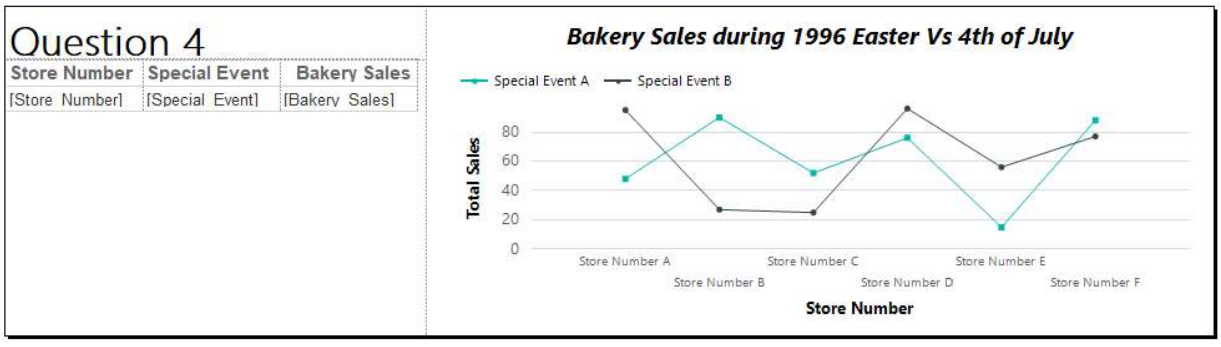
Selecting Analysis Services as data source in SSRS:



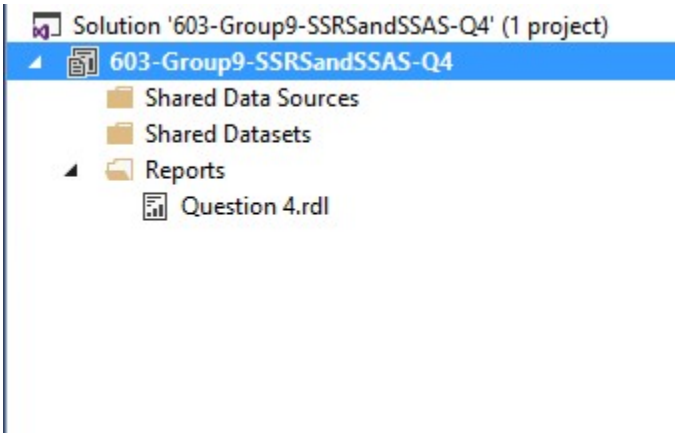
Report Summary:



Report Design:



Final report deployment:



infodata16.mbs.tamu.edu/ReportServer - /603-Group9-SSRSandSSAS-Q4

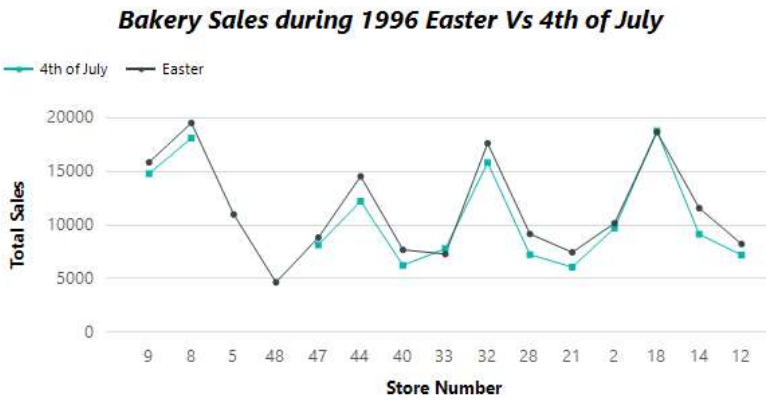
[To Parent Directory]
Friday, April 19, 2019 7:38 PM 36885 Question 4

Microsoft SQL Server Reporting Services Version 13.0.5081.1

Deployed report:

Question 4

Store Number	Special Event	Bakery Sales
12	4th of July	7212.41
12	Easter	8244.52
14	4th of July	9146.15
14	Easter	11595.32
18	4th of July	18780.28
18	Easter	18688.86
2	4th of July	9735.89
2	Easter	10176.74
21	4th of July	6087.57
21	Easter	7462.9
28	4th of July	7256.81
28	Easter	9177.05
32	4th of July	15833.65
32	Easter	17650.19
33	4th of July	7808.85
33	Easter	7302.14
40	4th of July	6252.45
40	Easter	7697.79
44	4th of July	12236.77
44	Easter	14555.44
47	4th of July	8157.82
47	Easter	8844.09
48	Easter	4662.44
5	Easter	11001.41
8	4th of July	18137.65
8	Easter	19546.77
9	4th of July	14791.91

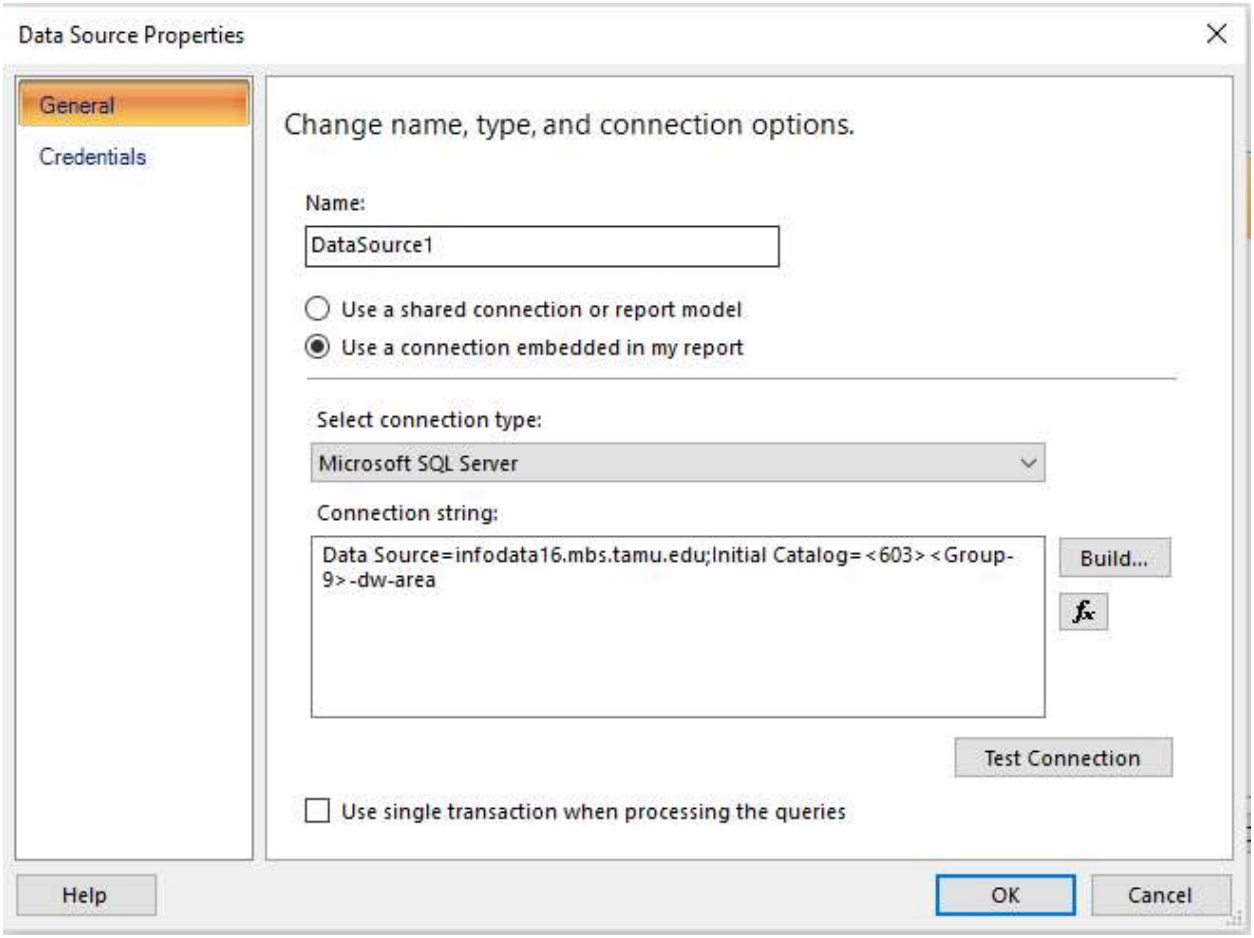


Conclusion:

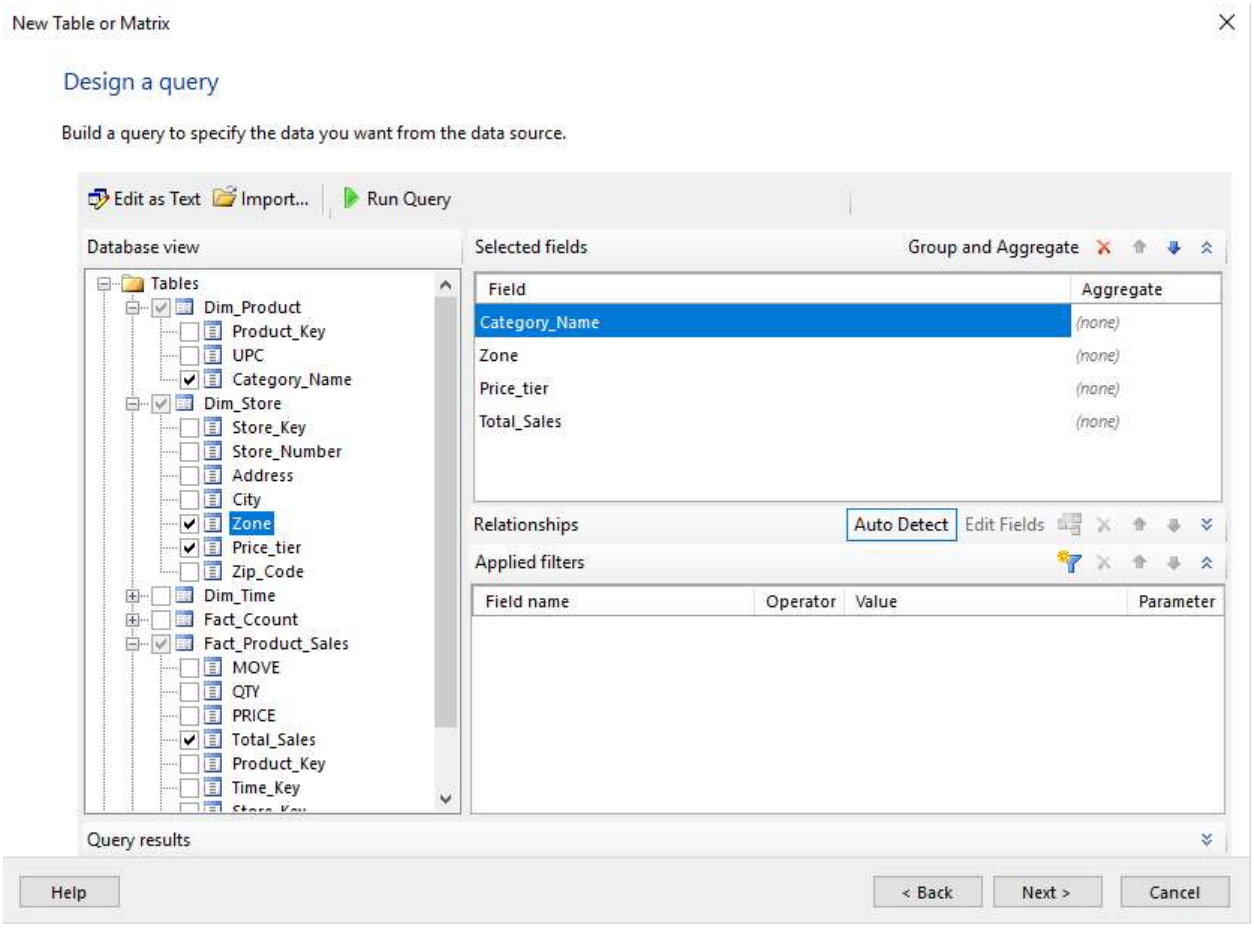
For this business question, we have created the cubes using SSAS and reports were further created using SSRS on top of it. The SSRS report built on top of the cube clearly conveys and compares the bakery sales between Easter and 4th of July 1996. We can see a very close comparison but bakery sales during Easter seems to have been slightly greater across every store. These results can be further used by DFF to make strategic decisions for 4th of July.

6.6 Report using Report Builder 3.0 for Question 5

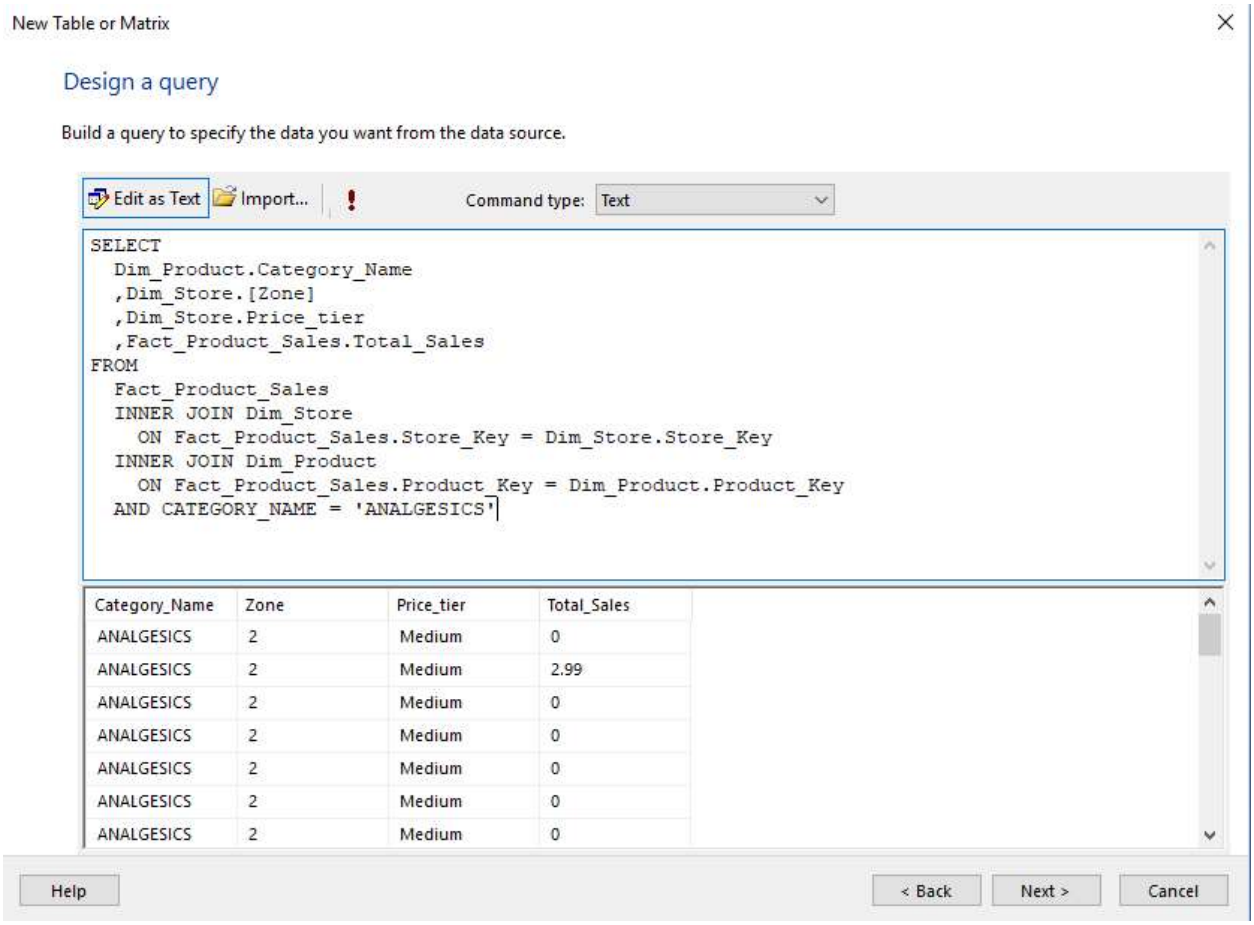
Data Source selection:



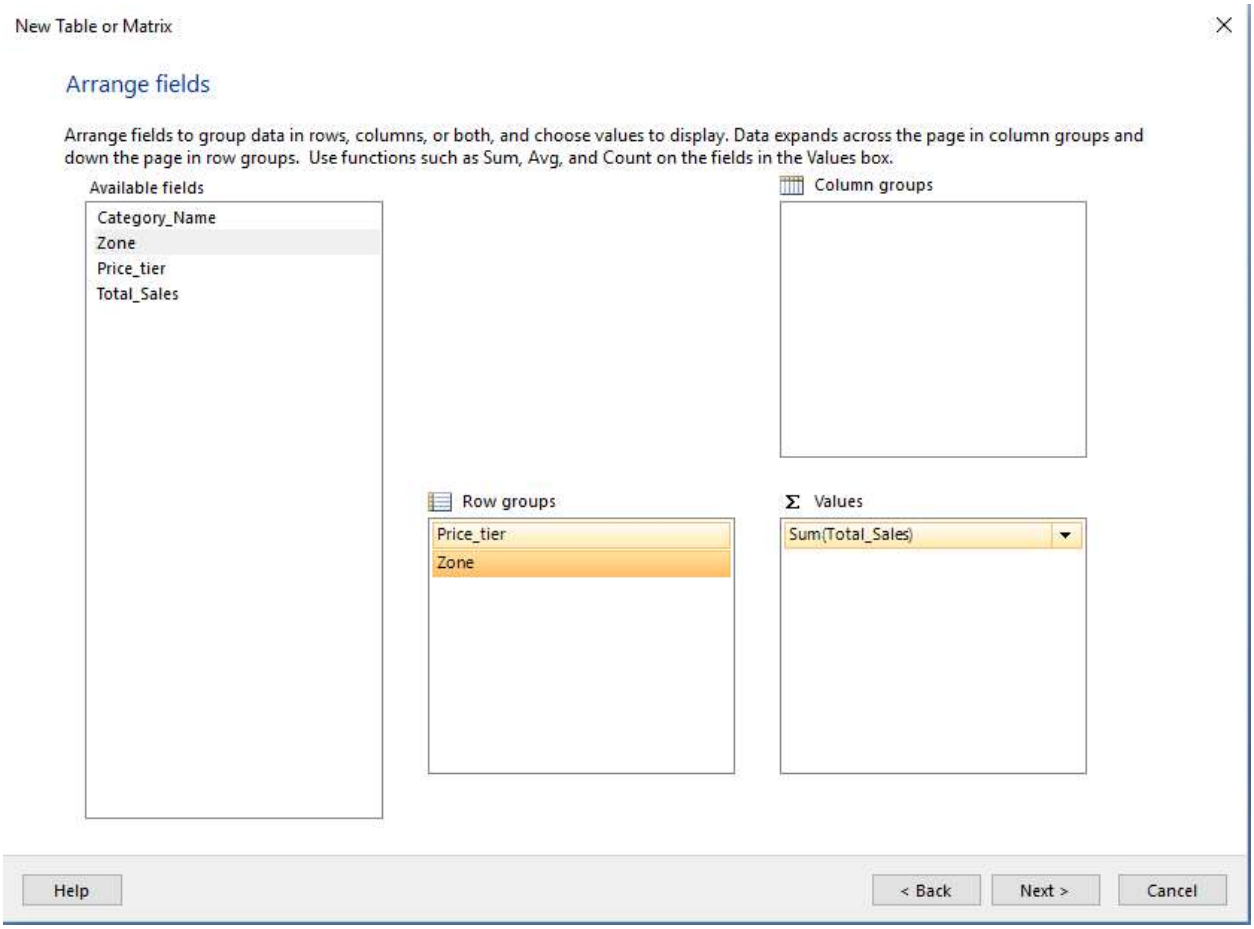
Selecting the required fields for the report:



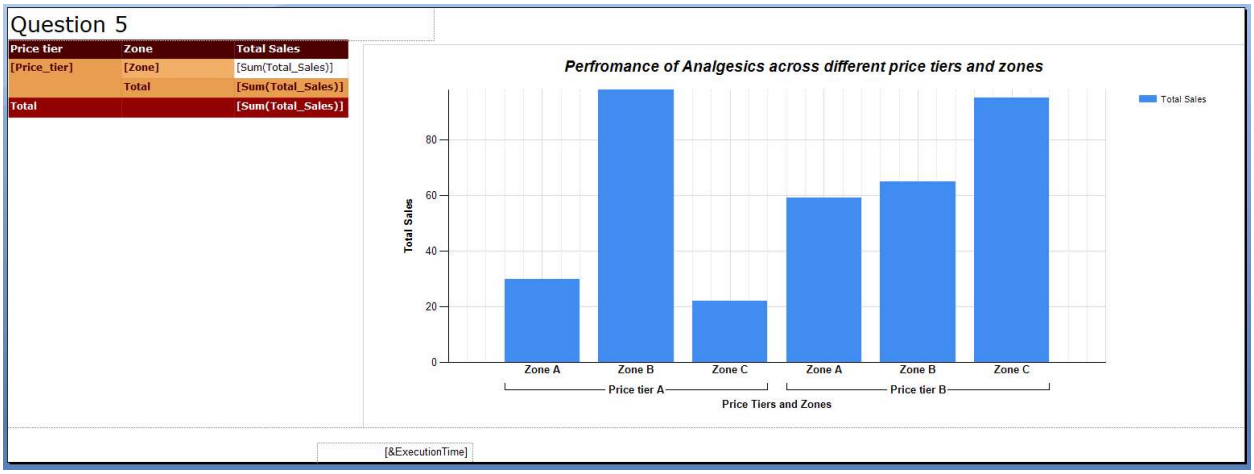
Query Used:


New Table or Matrix
Design a query
Build a query to specify the data you want from the data source.
Edit as Text
Import...
Command type:
Text
SELECT
Dim_Product.Category_Name
,Dim_Store.[Zone]
,Dim_Store.Price_tier
,Fact_Product_Sales.Total_Sales
FROM
Fact_Product_Sales
INNER JOIN Dim_Store
ON Fact_Product_Sales.Store_Key = Dim_Store.Store_Key
INNER JOIN Dim_Product
ON Fact_Product_Sales.Product_Key = Dim_Product.Product_Key
AND CATEGORY_NAME = 'ANALGESICS'
Category_Name Zone Price_tier Total_Sales
ANALGESICS 2 Medium 0
ANALGESICS 2 Medium 2.99
ANALGESICS 2 Medium 0
ANALGESICS 2 Medium 0
ANALGESICS 2 Medium 0
ANALGESICS 2 Medium 0
ANALGESICS 2 Medium 0
Help
< Back
Next >
Cancel

Grouping total sales based on price tier and zone:



Report design:

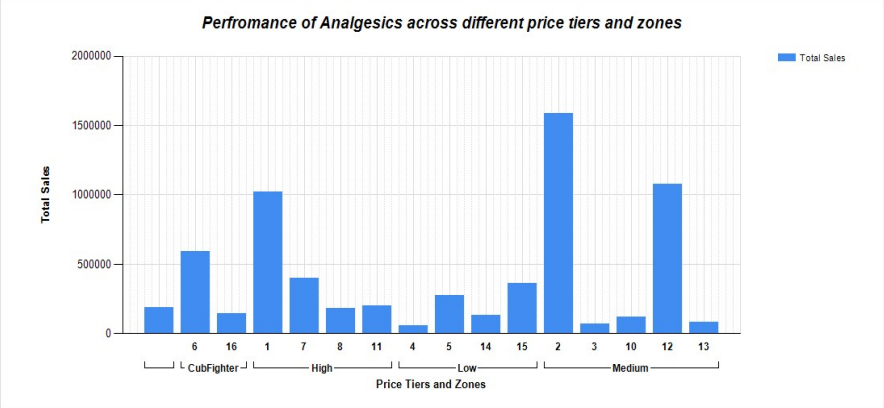


Final report:

Before drilldown:

Question 5

Price tier	Zone	Total Sales
⊞	Total	186247.249999998
⊞ CubFighter	Total	740163.789999851
⊞ High	Total	1806590.7999993
⊞ Low	Total	821628.01999981
⊞ Medium	Total	2944996.4800035
Total		6499626.34002014

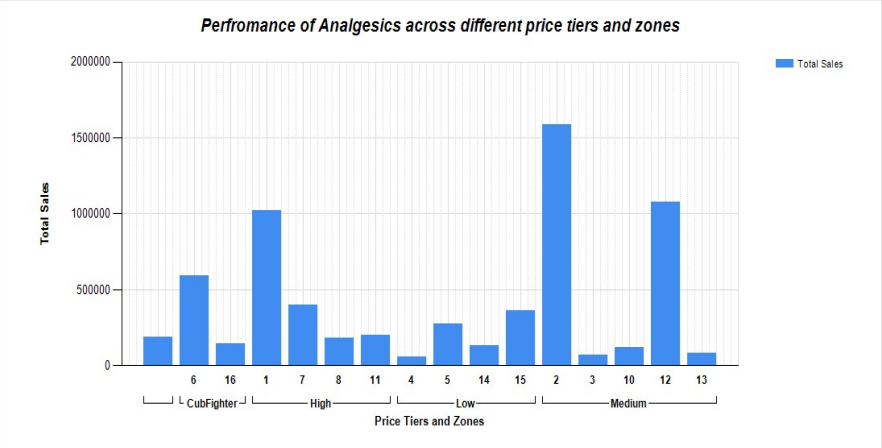


4/19/2019 4:58:12 PM

After drilldown:

Question 5

Price tier	Zone	Total Sales
⊞	Total	186247.249999998
⊟ CubFighter	6	592921.349999914
	16	147242.44
	Total	740163.789999851
⊟ High	1	1020722.15999972
	7	400841.699999967
	8	181637.299999999
	11	203389.639999999
	Total	1806590.7999993
⊟ Low	4	55378.5700000002
	5	273903.259999998
	14	132425.26
	15	359920.929999977
	Total	821628.01999981
⊟ Medium	2	1587576.30999952
	3	71130.1199999999
	10	123200.26
	12	1079739.7199997
	13	83350.070000001
	Total	2944996.4800035
Total		6499626.34002014



4/19/2019 4:58:12 PM

Conclusion:

The report built using Report Builder 3.0 facilitates data drilldown and subtotal calculations among other features. From the chart, we can see that Medium price tier has been performing the best by a large margin in the Analgesics category. Further analysis can be carried out to emulate the strategy used here on other price tier stores. Hence, report builder gave a clear insight on business question 5.

Team Task Sheet

Task	Completed by	Time taken (in days)
Reporting Plan	Pooja Chourey, Sai Shiva Ganesan, Sushant Patil	1
Reports from SSAS, SSRS, Report Builder and SSAS+SSRS	Pooja Chourey, Sai Shiva Ganesan, Sushant Patil	1
Integration of all reports	Pooja Chourey, Sai Shiva Ganesan, Sushant Patil	1
PowerPoint Presentation	Pooja Chourey, Sai Shiva Ganesan, Sushant Patil	1

Date: 4/25/2019

Pooja Chourey

Sai Shiva Ganesan

Sushant Patil