Design and Implementation of a Data Warehouse for A Retail Store with Store-level Data

Jha, Pragya

Contents

Requirements Gathering to Create Business Questions

Section 1: Introduction

Dominick's was a retail store chain based out of Chicago area which was closed in 2013 due to poor sales and low performance, not only owing to fierce competition but rather because it made blunders in its good days. For years, Dominick's dominated the Chicago metropolitan area and had a slightly upscale feel to it. Later, Dominick's was acquired by Safeway and it made major changes that altered the future of business. Store layouts were altered, few stores that were underperforming were closed and replaced by newer stores without understanding the demographics of those areas. "According to a grocery business consultant, "Dominick's focused on purchasing produce and meat on quality first, price second. Safeway did just the opposite." Dominick's lost market share and profits following the Safeway takeover."[1]

We have been provided data, which was received as a part of partnership between Chicago Booth and Dominick's Finer Foods for conducting store level research for shelf management and pricing. There are three types of information used in this paper, each coming from different sources. As a part of the project, our team is supposed to build a data warehouse for better analysis of this data that could have helped Dominick to survive.

Challenges faced during the various phases of the group project:

1) Data provided has many file formats. (CSV, TXT, HTML) and it was quite burdensome to study them as they all had to be converted to Excel format initially.
2) It was quite a task to understand the relationships between the different sheets and understand the huge data.
3) Brainstorming business questions was equally tough and were decided after reading a few research papers and after reading about the retailing business.
4) Time management was also an issue as all three of us had different work schedules and there was a lot of chaos in the beginning.
5) Entity Relationship diagram also took a lot of time as it had to be perfect for further work.
6) Prioritizing business questions was the toughest of all as this was going to be the deciding factor for our future.
7) Owing to the huge size of data, finding out inconsistencies in data was a herculean task and it shall remain an ongoing process throughout the project.
8) Not having the right skills for ETL can be a hindrance later, so improving on our skills is another challenge for our team.

Section 2: Details about the data

2.1 Understanding of the data

The data contains information about Dominick's Finer Foods for a period of nine years and has store level information for more than 3500 UPCs in this 100-store chain.

Customer Count Files: It contains information on store traffic and coupon usage by store. That is, to say, it has information on a daily basis about the purchases made and the count of customers who visit the store. It also holds information regarding the coupons redeemed. These figures are compiled daily from the register/ scanner receipts.

Store Level Demographics file: It contains information for each of the DFF stores. The data originally came from US government census data (1990) for the Chicago metropolitan area. Later, Market metrics processed this data. It contains information about the age groups of customers, the household income, the dependent members etc.

UPC Files: These files contain description of each UPC in a category. It is historical data and contains information regarding the Product name, size, description and UPC number of a product.

Movement Files: They contain sale information at the store level for each UPC in a category. In addition, this information is stored on a weekly basis. It contains information like gross margin of UPC product.

Apart from the files, there is also information on store addresses and the list of holidays (the exact dates and weeks) in the pdf provided.

2.2. Metadata for all the OLTP source files

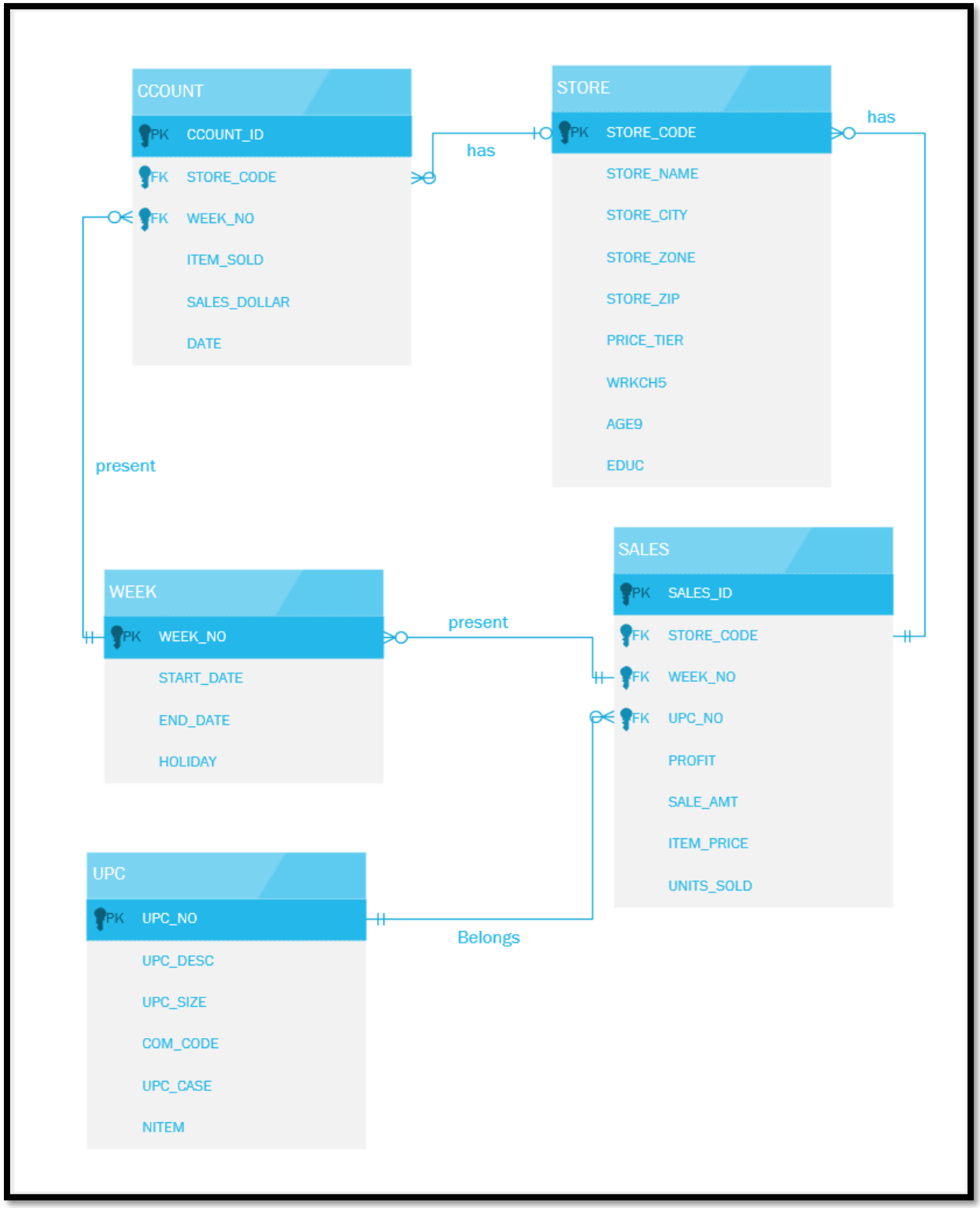
Customer Count Files: It contains store, date, sales on different types of items, coupons redeemed and the corresponding weeks.

Store level Demographics File: It contains information regarding the age of customers, the household income conditions, the availability of cars, the singles and retired, the household with working women and the type of shoppers etc.

UPC Files: These files contain the description, the size, the number of items in the case, the item code and the commodity code.

Movement Files: These files contain the sales, profit margins, the number of units sold and the promotions offered when sold or not and they are collected on a weekly basis.

2.3. Entity-Relationship Diagram



Section 3: Domain Understanding

We have studied five research papers to get a good understanding of the Retail domain. Below are the papers and the brief summary of our understanding from them.

From the paper, **"Why Do Some Prices in the Retail Sector Drop When Demand Rises? Evidence from the Chilean Case," Revista ABANTE 10 (Octubre 2007): 151-168 by Lira Loreto**, we conclude that the procyclical demand elasticity model of Warner and Barsky (1995) which focuses on the prevalence of economies of scale is the best imperfection competition that explains the results of the study. According to this theory, retailers tend to offer prominent discounts on particular products to attract customers to the store to gain profits. Once inside the store, it is easier for customers to purchase more than one product since they are utilizing the economies of scale in their purchase. [2]

From the paper, **Georg Muller, Mark Bergen, Shantanu Dutta, and Daniel Levy, "Private Label Price Rigidity During Holiday Periods," Applied Economics Letters, 13 (2006): 57-62**, we found that the reason for the difference in the prices cannot be determined by changes in holiday promotional practices as private label promotions tend to decrease at least as much as the national brands. The authors propose that the variations may be related to differences in the extent of rigidity in their wholesale prices during the holiday and non-holiday periods. However in conclusion, they suggest that the higher private label price rigidity might be due to the increased emphasis on social consumption during holiday periods, raising the customer's value of nationally branded products relative to the private labels. [3]

From the paper, **Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data, By JUDITH A. CHEVALIER, ANIL K KASHYAP, AND PETER E. ROSSI**, we concluded that generally it is expected that prices either stay the same or increase when the demand is high but retails stores do not follow this standard model. They exploit the fact that demands are more during holiday seasons and do the research accordingly. Moreover, they find that the falling prices are due to falling markups. A few of the conclusions of the paper: Retail margins fall for foods at their peak demands, during Thanksgiving and Christmas, prices tend to increase and that there was no truth in the statement that elasticity of demand for product increases at their seasonal high demands.[4]

From the paper, **Who Benefits from Price Promotions? Price cuts boost sales and traffic. So where are the long-term gains? - By Shuba Srinivasan, Koen Pauwels, Dominique Honssens, and Marnik Dekimpe**, we concluded that retailers and manufacturers often employ different price promotion tactics to attract customers into their stores and boost their sales. Researches have shown that after these promotions run their course, customers quickly go back to their regular buying habits and sales drop back to normal. The conclusion was that promotions are tactical, not strategic, and should be managed as such. Only then will both retailers and manufacturers be able to reap the benefits of promotions while limiting the potential downside.[5]

From the paper, **The Value of Purchase History Data in Target Marketing - By Peter E. Rossi, Robert E. McCulloch and Greg M. Allenby. Source: Marketing Science, Vol. 15, No. 4, (1996), pp. 321-340,** we concluded that based on the premise that there exist distinct segments of homogeneous consumers who can be identified by readily available demographics, the targeting of specific consumer groups for differential promotional activities has become an important aspect of marketing. The possibility of this direct marketing strategy has been further strengthened by the widespread use of optical scanning and other automatic data collection methods and the increased availability of individual consumer panel data. [6]

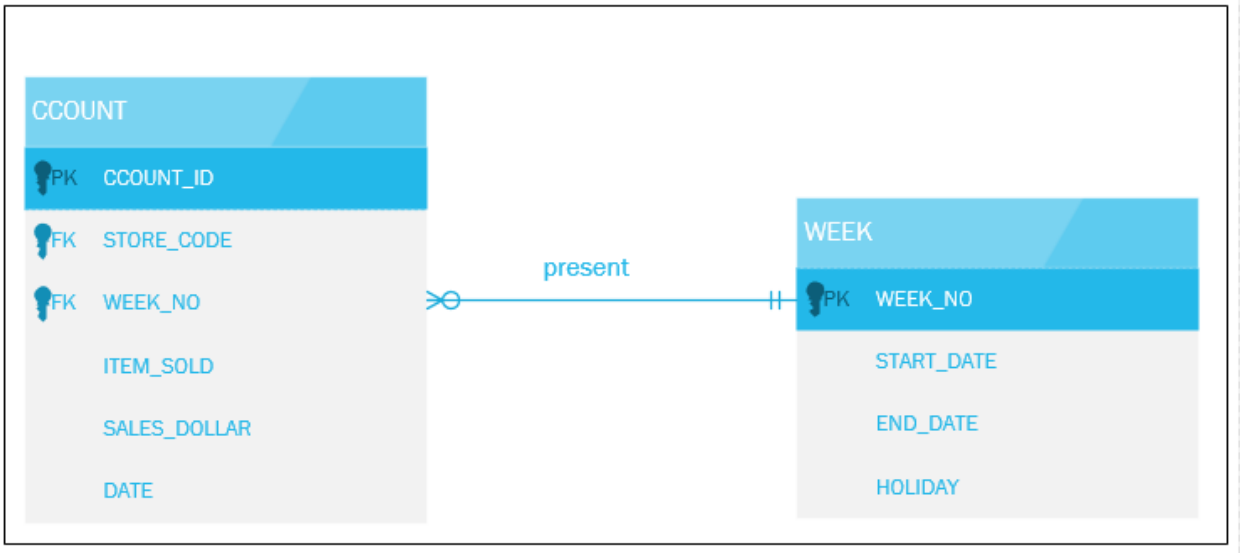
Section 4: Business Questions

From our subject area research and analysis of Dominick's data and metadata, we came up with ten business questions aimed at providing deeper insights into DFF's retail operations. The business questions have been prioritized and classified as either High, Medium or Low based on their strategic relation and importance to business goals and also how much they can support management decisions to boost sales and increase revenue for DFF.

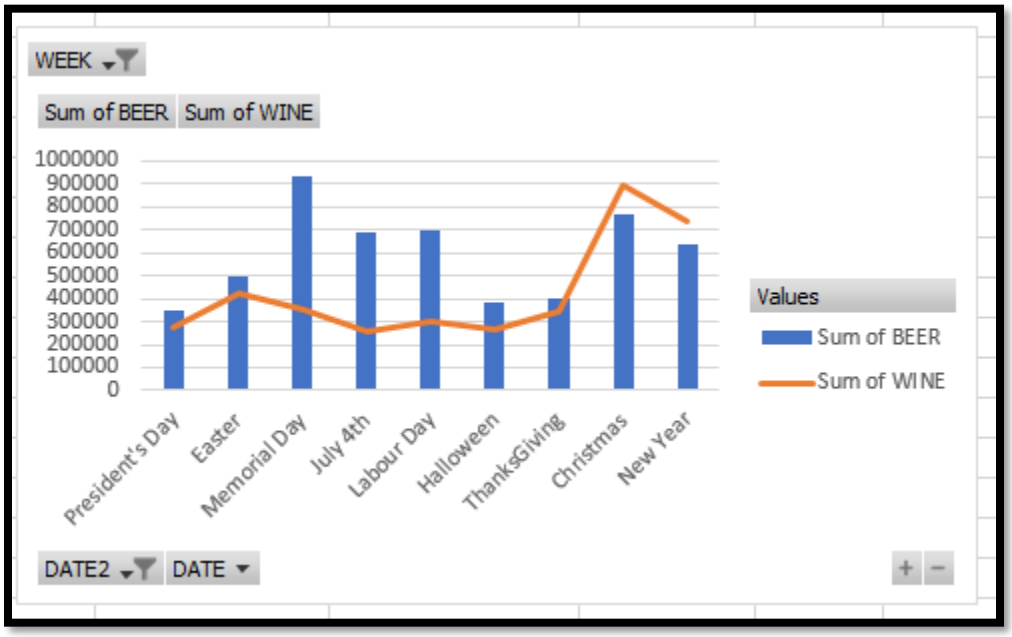
Question 1 - What impact do holidays have over sale of beer and wine and which holidays have the highest demands?

Priority – HIGH

Entity-Relationship Diagram:



Row Labels	Sum of BEER	Sum of WINE
President's Day	346912.51	271454.19
Easter	500252.69	420812.06
Memorial Day	933612.98	355208.11
July 4th	690520.75	258058.87
Labour Day	695052.37	303508.53
Halloween	383228.38	266221.18
ThanksGiving	404957.35	347778.3
Christmas	765450.54	894678.26
New Year	632680.93	740853.11
Grand Total	5352668.5	3858572.61



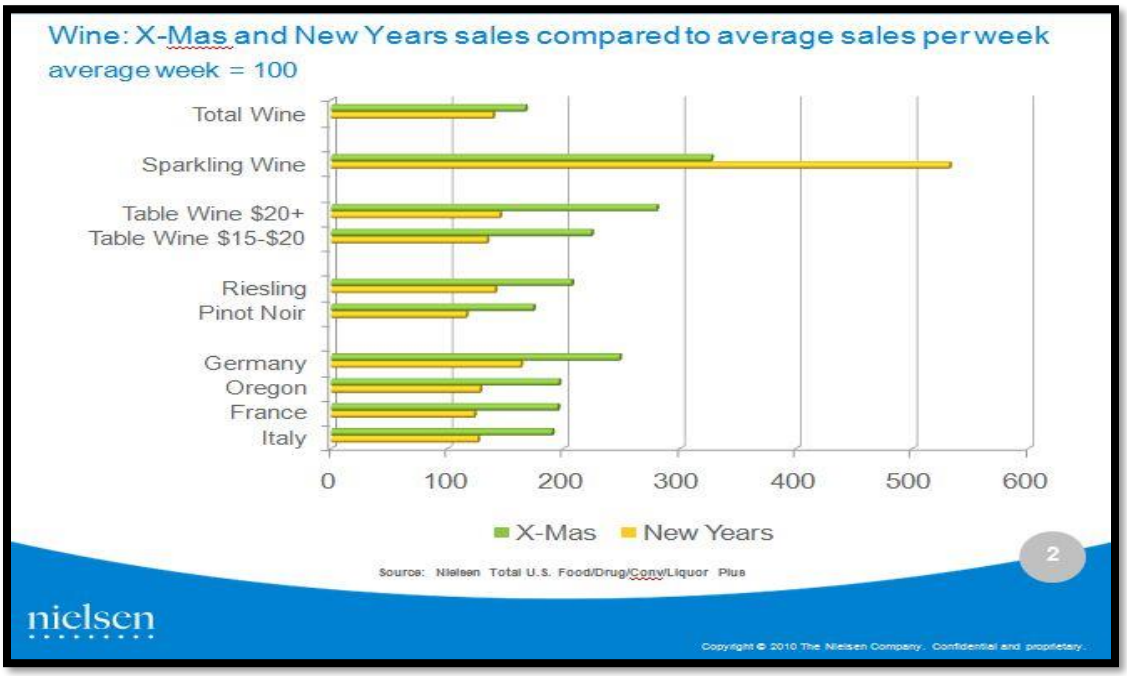
Justification

During holidays, sale of beer and wine increases dramatically. We compared the sales on all the leading holiday times in United States and from our exposure to the limited data, we found that beer sales are maximum on Memorial day and wine sales are maximum on Christmas. This could have helped Dominick's to deal with holidays in a better manner. They could have stocked up beer and wine during those particular holidays and also give discounts to maximize sales.

"Although it seems almost natural to crack open a cold beer on Memorial Day or Fourth of July, these aren't the only big beer-buying holidays. Nielsen has compiled the Top 7 beer holidays in the United States in 2010, based on the sale of cases of beer two weeks before the holiday.

7. Super Bowl, 49.3 million
6. Halloween, 51.3 million
5. Thanksgiving, 51.8 million
4. July Fourth, 52.7 million

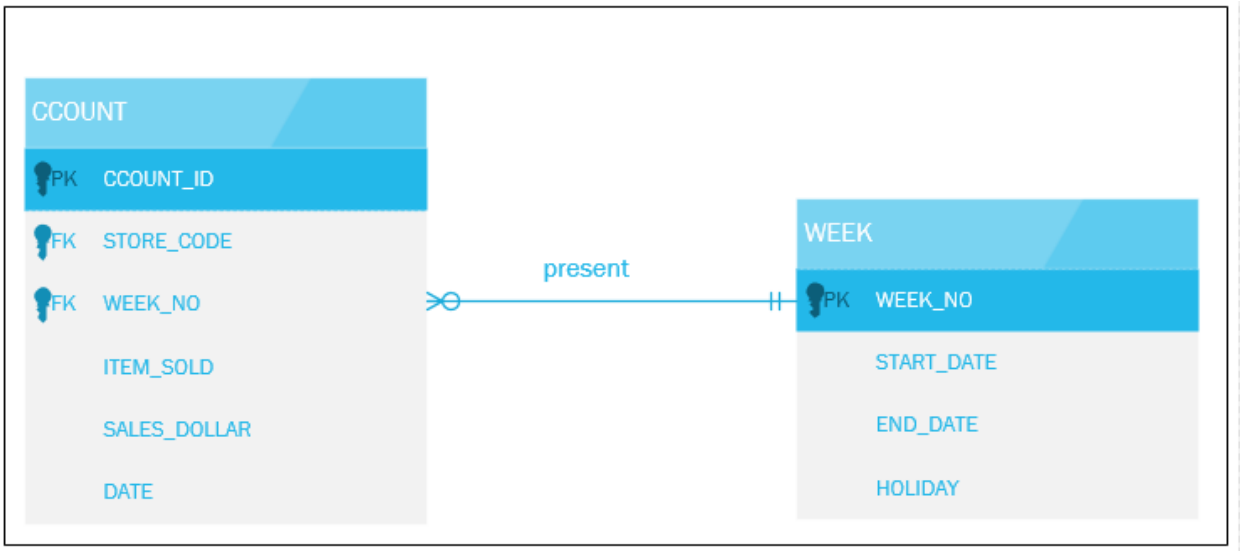
3. Christmas, 54.3 million
2. Memorial Day, 57.8 million
1. Labor Day, 60.1 million" [7]



[8]

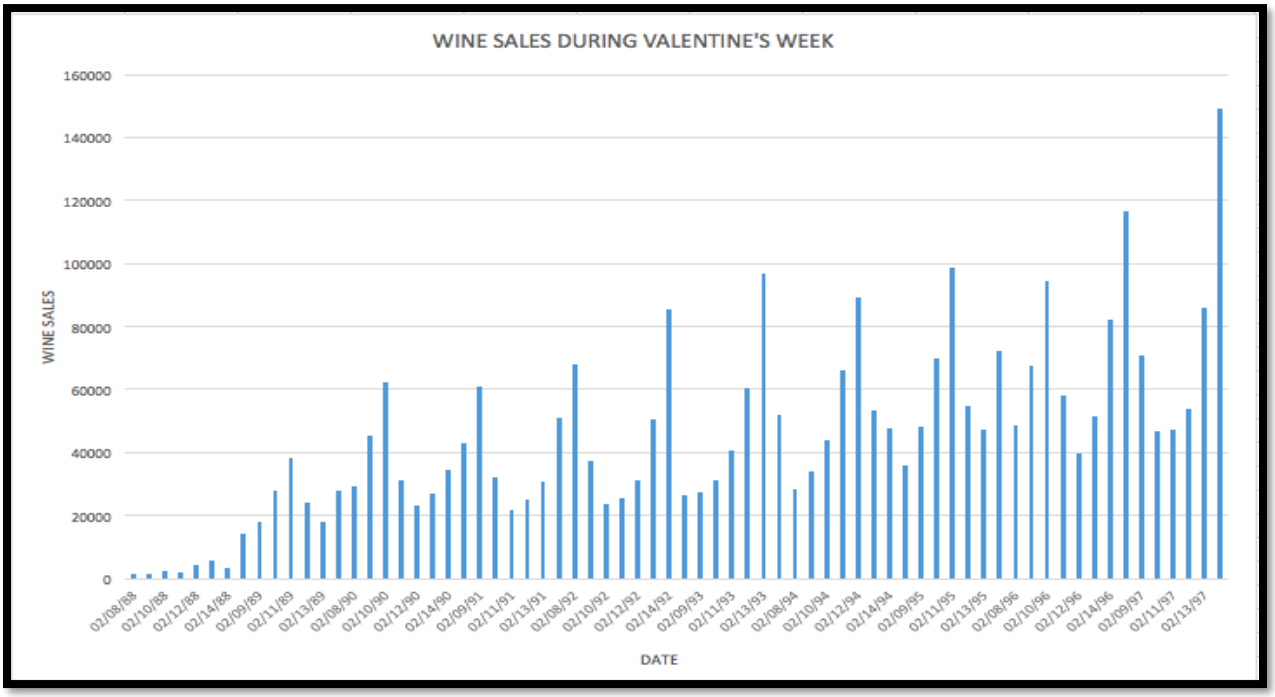
Question 2 - What is the trend in wine sales over the years during Valentine's week?
Priority – HIGH

Entity-Relationship Diagram:



Row Labels	Sum of WINE
02/14/1988	3244
02/14/1989	27911
02/14/1990	34401
02/14/1991	51169
02/14/1992	85489
02/14/1993	51969
02/14/1994	47580
02/14/1995	72450
02/14/1996	82285
02/14/1997	149406
Grand Total	**605903**


Wine Sales on Valentines Day
Wine Sales
700000
600000
500000
400000
300000
200000
100000
0
02/14/1988
02/14/1989
02/14/1990
02/14/1991
02/14/1992
02/14/1993
02/14/1994
02/14/1995
02/14/1996
02/14/1997
Grand Total
Date


WINE SALES DURING VALENTINE'S WEEK
WINE SALES
160000
140000
120000
100000
80000
60000
40000
20000
0
DATE

Row Labels	Sum of WINE
02/08/88	1413
02/09/88	1470
02/10/88	2285
02/11/88	2115
02/12/88	4223
02/13/88	5696
02/14/88	3244
02/08/89	13999
02/09/89	18171
02/10/89	27714
02/11/89	38268
02/12/89	24241
02/13/89	18010
02/14/89	27911
02/08/90	29469
02/09/90	45494
02/10/90	62443
02/11/90	31393
02/12/90	23074
02/13/90	27046
02/14/90	34401
02/08/91	42858
02/09/91	60740
02/10/91	32236
02/11/91	21923
02/12/91	24971
02/13/91	30610
02/14/91	51169
02/08/92	67833
02/09/92	37213
02/10/92	23460
02/11/92	25555
02/12/92	30967
02/13/92	50633

Justification

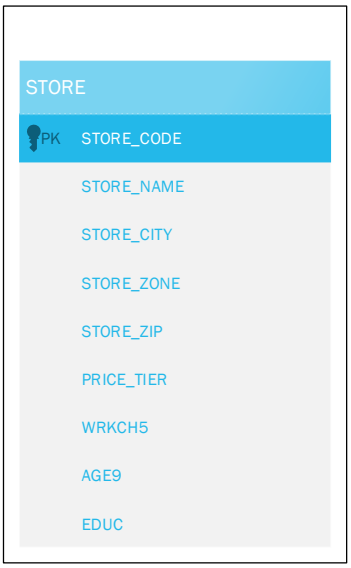
Most of the retailers in US depend on strong holiday sales and nearly 74% of retailers who have been surveyed say it's the time of the year where they get 20% or more of their annual sales. [9]. We wanted to analyze the sales pattern of one such product, Wine, which makes most of the sales during the Valentine's week.

From graph 1, it is evident that Wine sales are high during the Valentine's week, especially on the Valentine's day. While graph 2, emphasizes on the increase in the sales of Wine during Valentine's Week over the years. For example, sales in the year 1997 is almost 150 times more than the sales in 1988 and twice as much as the previous year 1996. This analysis can be used by retail managers to plan the amount of inventory for the upcoming year and avoid the loss of sales due to shortage of goods. In addition to this, retail managers can plan promotions and offers to attract the customers and increase the sales of Wine on the days when sales are not as high as on the Valentine's week

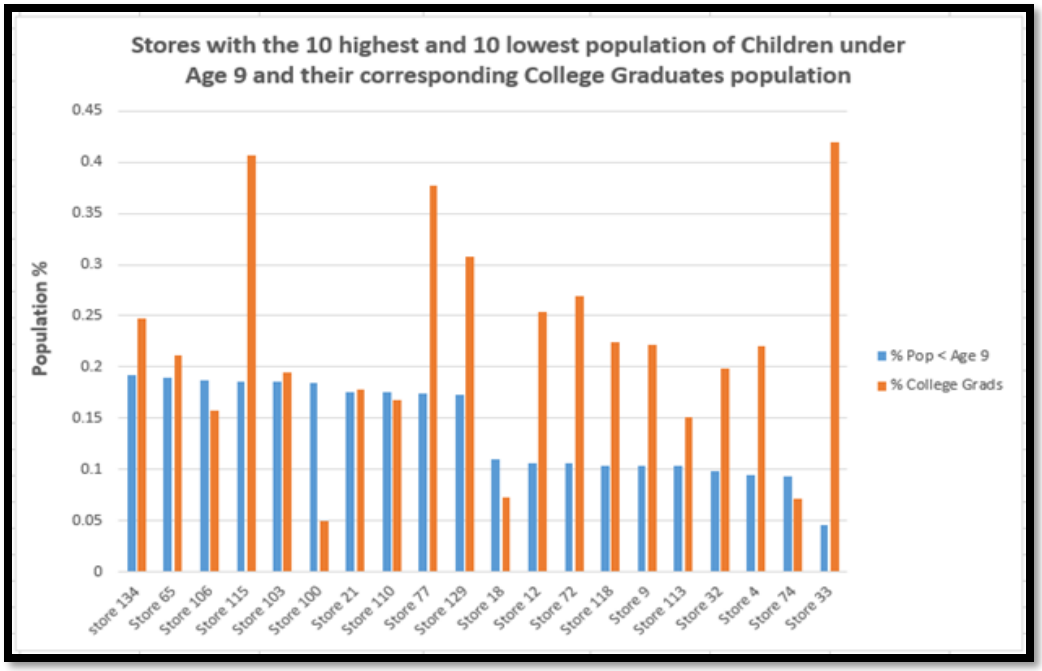
Question 3 - Which stores account for the 10 highest and the 10 lowest percent population of children under the age of nine (9) and what is the corresponding percent population of college graduates for the identified stores?

Priority – HIGH

Entity-Relationship Diagram:



STORE	% Pop < Age 9	% College Grads
store 134	0.192817818	0.247949139
Store 65	0.190055414	0.210995164
Store 106	0.187403158	0.157939308
Store 115	0.185493511	0.406312482
Store 103	0.185393984	0.194621102
Store 100	0.184360976	0.049550286
Store 21	0.175926346	0.17750345
Store 110	0.175594056	0.167553192
Store 77	0.174716756	0.376871097
Store 129	0.172501467	0.307842548
Store 18	0.110094984	0.072246456
Store 12	0.10569674	0.253412969
Store 72	0.105576311	0.268724553
Store 118	0.103791822	0.224725883
Store 9	0.103503097	0.222172318
Store 113	0.103084931	0.151592429
Store 32	0.099060632	0.198259861
Store 4	0.095089506	0.220789415
Store 74	0.093843178	0.07119776
Store 33	0.046070917	0.419688004
Grand Total	**2.790075602**	**4.399947416**



Justification

Youth marketing is a marketing strategy increasingly being employed by marketers and advertisers today. It is targeted at youth in every effort to capture their attention and inspire lifelong customer loyalty. Marketers have come to realize the criticality of this demographic, not only in their buying power and influence on family spending, but also in their ability to set trends for other demographics. Wikipedia states in its article that, 'The influence that youth have on purchases made in a household are extremely high, even on high-end items such as what vehicle the family decides to purchase. For example, one study estimated that children influenced $9 billion worth of car sales in 1994. One car dealer explains: "Sometimes, the child literally is our customer. I have watched the child pick out the car."'[10]

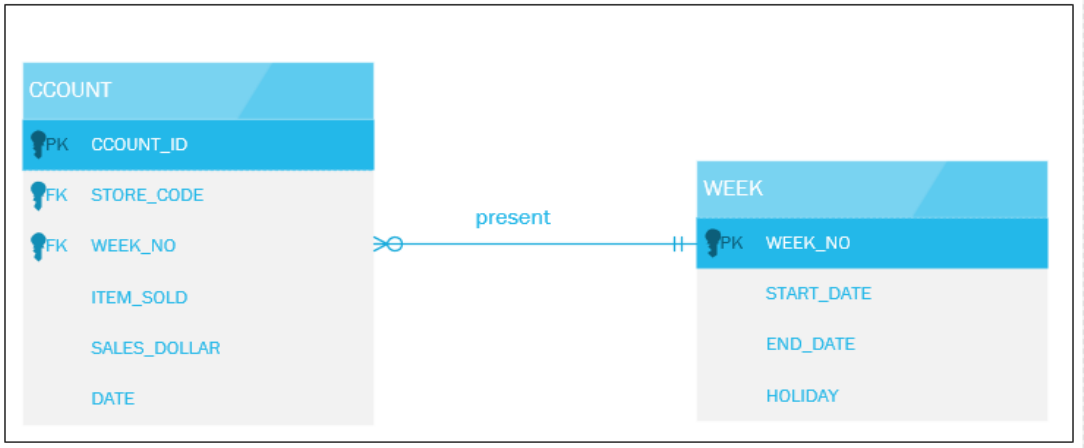
Another highly influential demographic is that of children. Statistics from various research have gone on to analyze and demonstrate their impact on the consumer market. One of such resources had stated that, "With children either spending or influencing $500 billion worth of purchases, marketing techniques have been turned upside down. In the past, the most effective way to sell children's products was through mom and dad. Now the opposite is true, children are the focal point for intense advertising pressure seeking to influence billions of dollars of family spending."[11]

With the insight of these facts and numbers, we wanted to find out the percentage of these target demographics for DFF stores and how it could be leveraged by DFF managers to employ more direct marketing strategies and boost sales. We examined the store-specific demographics data for DFF and found that for over 50% of all the stores captured, the % population of children under age 9 was significantly lower than the % population of college graduates for the same stores. We further drilled this down to stores that account for the 10 highest and 10 lowest % population of children under age 9, and found that the corresponding % of college graduates was still higher for 15 (75%) of those stores.

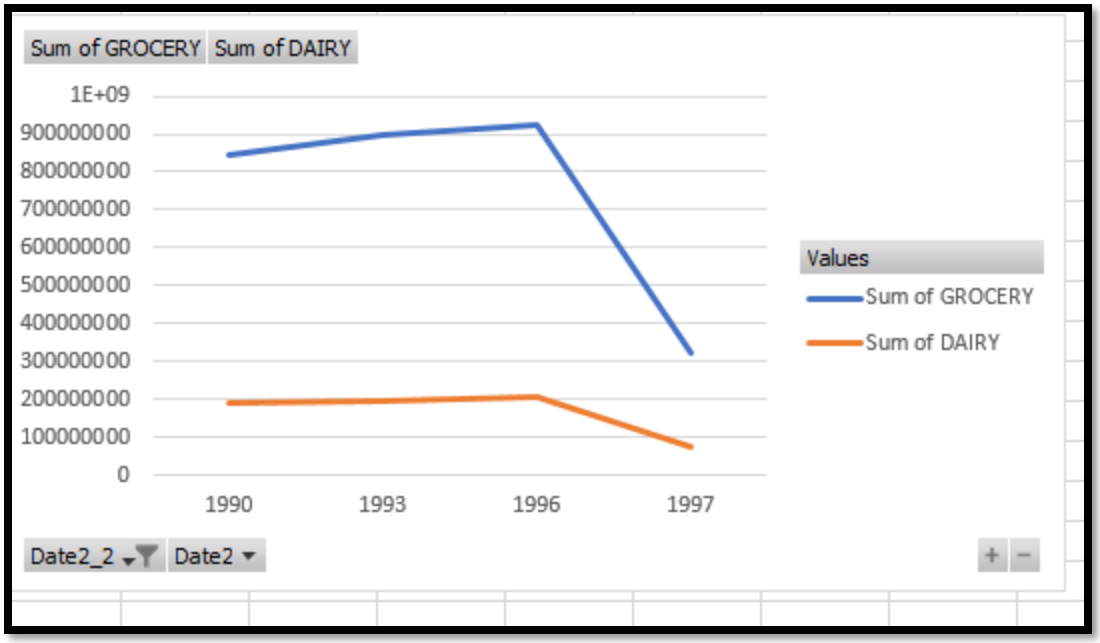
Question 4 - How do the total sales of Grocery and Dairy vary over years? What happened in 1997?

Priority – HIGH

Entity-Relationship Diagram:



Row Labels	Sum of GROCERY	Sum of DAIRY
⊞1990	842037487.2	191878415.3
⊞1993	897102757.6	196827141.3
⊞1996	922328335.9	207521630.5
⊞1997	324162494.9	76671371.1
Grand Total	2985631076	672898558.1



Justification

We compared the sales of Grocery and Dairy over a range of few years and with our limited data, we found that the growth was stagnant for a while and then in 1997, the sales went down. As it is evident from the graph that Dominick's did not pay much attention to increase sales for all the years and had become complacent which ultimately led to its closure. Walmart opened its stores in 1997 and that affected Dominick's Finer Foods business badly. Dominick's should have kept up with the new market trends earlier than it did.

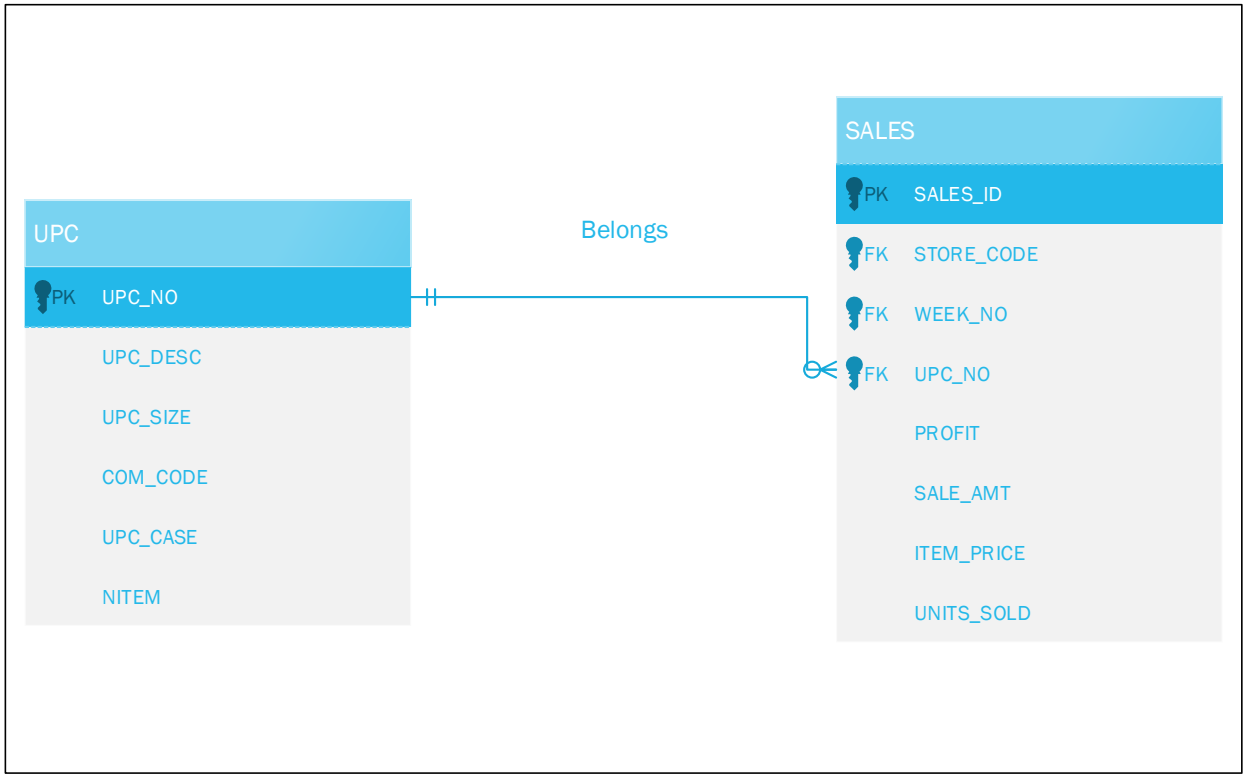
"Between 1997 and 2002, sales of grocery products in traditional grocery stores fell by approximately 2% in real terms, while sales of grocery products in "general merchandise" stores, which include Wal-Mart, grew by 48% in real terms (and by 57% in the narrower category of warehouse clubs and superstores). [12]

Around 1997, when Safeway(Dominick's Finer Foods was acquired by Safeway) began making prominent changes to the pricing structure with personalized deals via a loyalty program, opening new stores with a fresh market focus and implementing competitive pricing, it was too late. [13]

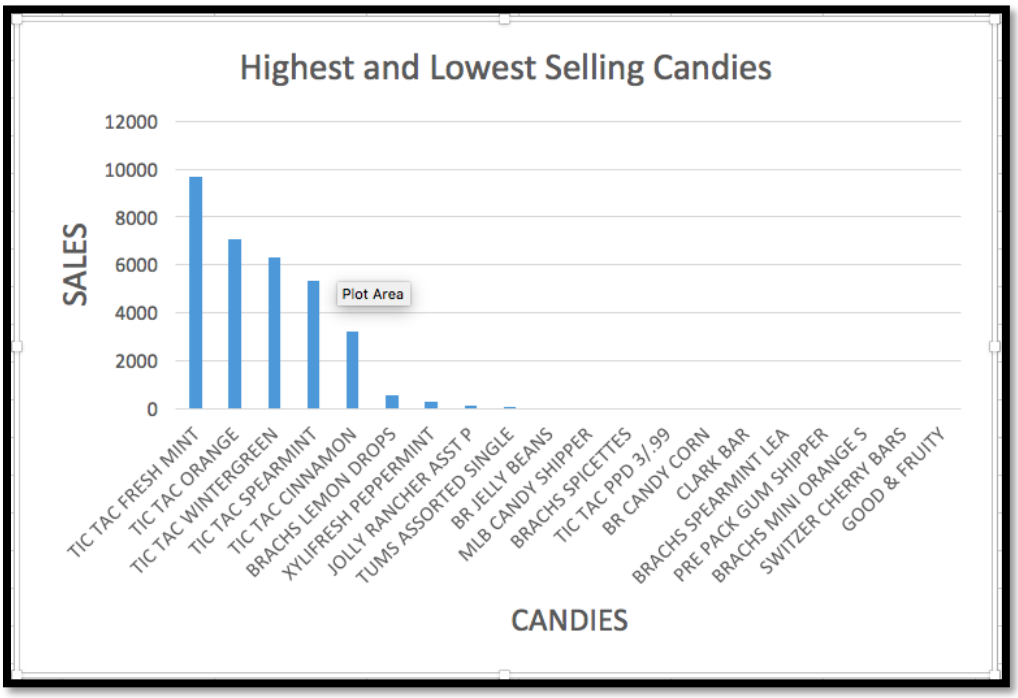
Question 5 - Which candies made the highest and lowest number of sales during Halloween week over the years?

Priority – MEDIUM

Entity-Relationship Diagram:



Row Labels	Sum of SALE
TIC TAC FRESH MINT	9657
TIC TAC ORANGE	7078
TIC TAC WINTERGREEN	6308
TIC TAC SPEARMINT	5321
TIC TAC CINNAMON	3244
BRACHS LEMON DROPS	541
XYLIFRESH PEPPERMINT	297
JOLLY RANCHER ASST P	147
TUMS ASSORTED SINGLE	54
BR JELLY BEANS	0
MLB CANDY SHIPPER	0
BRACHS SPICETTES	0
TIC TAC PPD 3/.99	0
BR CANDY CORN	0
CLARK BAR	0
BRACHS SPEARMINT LEA	0
PRE PACK GUM SHIPPER	0
BRACHS MINI ORANGE S	0
SWITZER CHERRY BARS	0
GOOD & FRUITY	0



Justification

Americans consume more than 7 billion pounds of candy per year. Not surprisingly, a large amount of that candy consumption happens during Halloween. It's the number one holiday for candy sales. The week of Halloween alone, Americans buy 90 million pounds of chocolate.On that basis, we wanted to analyse data of Dominick's candy sales data during Halloween and aim to determine the top selling candies and candies with least number of sales. The purpose behind doing so is to compare the sales trend between the candies and to find the trend of candy sales during the Halloween weeks over the years.

The data analysis shows that sum of sales of Tic Tac candies in all the dominick's stores is highest. While in the above graph and snippet of Pivot table details of the candies that have made no sales during Halloween have been emphasized. This information would enable retail managers to plan strategies to increase the sales of the products poor sales so as to benefit the business. Alternatively, they can avoid ordering the candies that have made poor sales for the upcoming year and try ordering more candies that have made highest sales. This will not only save Dominick's from facing losses due to unsold candy but only an increase in gross margin can be experienced. [14]

Question 6 – Compare meat sales during Christmas for the different store tiers - High, Medium, Low and Cub Fighter

Priority – MEDIUM

Entity-Relationship Diagram:

Christmas Weeks	(Multiple Items)
Row Labels	Average of MEAT
⊞ High Tier	6689.15314
⊞ Medium Tier	6049.940559
⊞ Low Tier	7935.051423
⊞ Cub Fighter	7287.834935
Grand Total	6584.164729

Justification

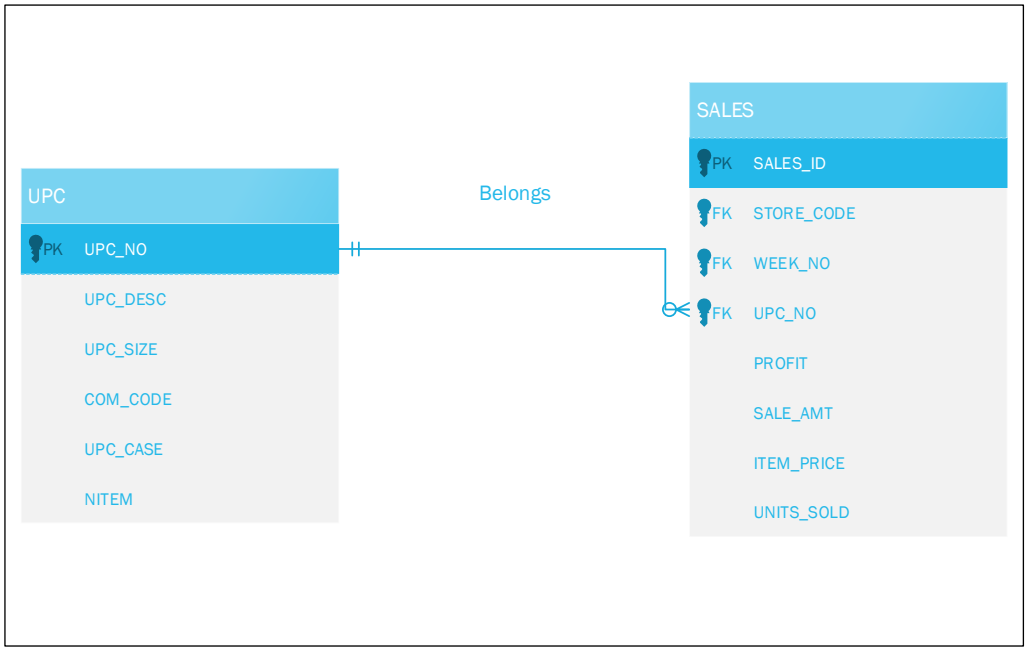
DFF priced products by 16 zones. And each zone has four price tiers: Cub Fighter, Low, Medium and High. So knowing the zones better will help distribute the resources in a better manner. We checked the meat sales in 4 tiers and found that surprisingly, Low Tier had maximum sales during Christmas. So Dominick's should have done more marketing for meat sales in other tiers during Christmas and also made sure that Low Tier had enough supply of meat and enough discounts to maximize the profits. There is quite a bit of disparity in the four tiers according to the demographics of the area. [15]

"Different price tier stores likely to have different assortments, clientele demographics (e.g., Hoch et al. 1995), pass-through rates (e.g., Besanko et al. 2005) and potentially different pricing rules."[16]

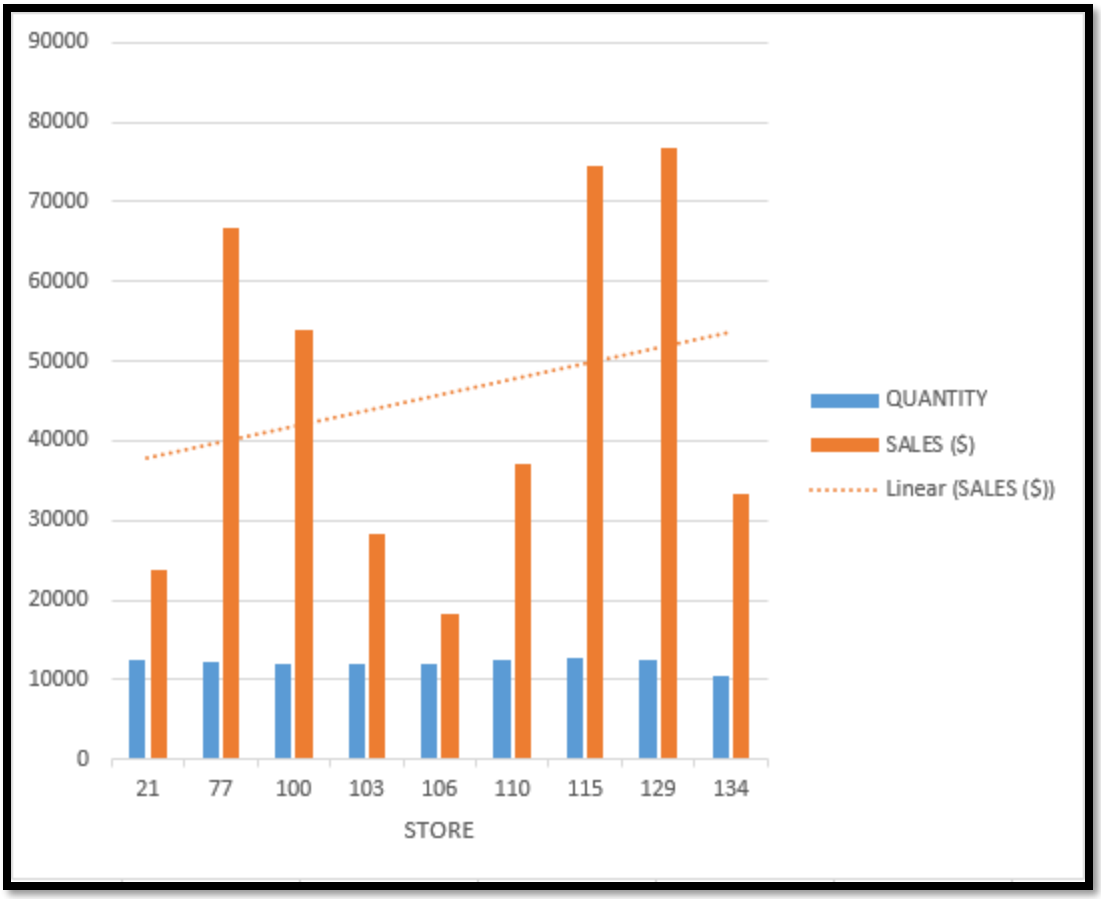
Question 7 - What is the trend in the sale of cookies (from week 1 – 399) in stores with the highest % population of kids under age 9?

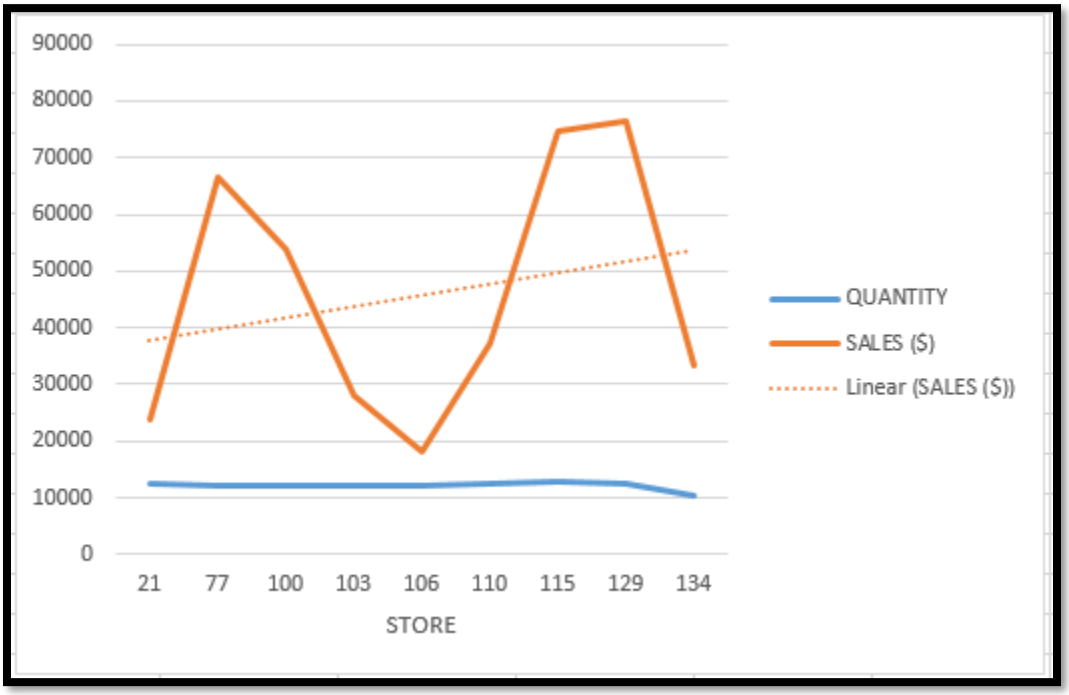
Priority – MEDIUM

Entity-Relationship Diagram:



WEEK	(All)	
STORE	**QUANTITY**	**SALES ($)**
21	12455	$23,741.11
77	12236	$66,672.46
100	12092	$53,878.04
103	12109	$28,184.83
106	12118	$18,219.68
110	12458	$37,071.93
115	12831	$74,507.21
129	12500	$76,620.67
134	10501	$33,319.25
Grand Total	**109300**	**$412,215.18**


90000
80000
70000
60000
50000
40000
30000
20000
10000
0
21
77
100
103
106
110
115
129
134
STORE
QUANTITY
SALES ($)
Linear (SALES ($))


90000
80000
70000
60000
50000
40000
30000
20000
10000
0
21
77
100
103
106
110
115
129
134
STORE
QUANTITY
SALES ($)
Linear (SALES ($))

Justification

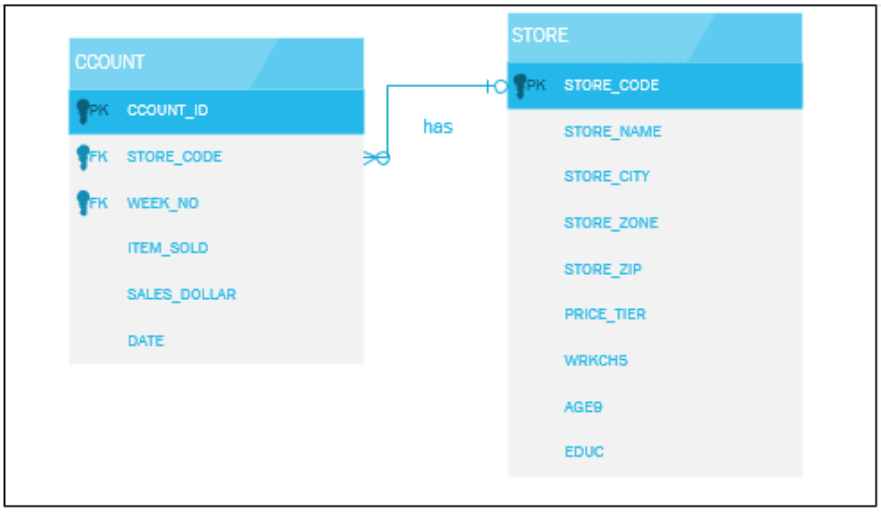
With the concept of youth marketing and/or targeted marketing strategy in mind, we wanted to analyze the sales trend for these targeted demographics.

For children under age 9, we ran the analysis using cookies which is a product that is very popular with this demographic. We further drilled down to examine only the stores with the highest population of children under age 9. We found that linear sales across these stores showed an upward going (increasing) trend, from about $38,000 to just below $55,000. This can be observed in the charts below. DFF can take advantage of this positive trend to introduce more brands and varieties of cookies in these stores in order to increase revenue and profit margins.

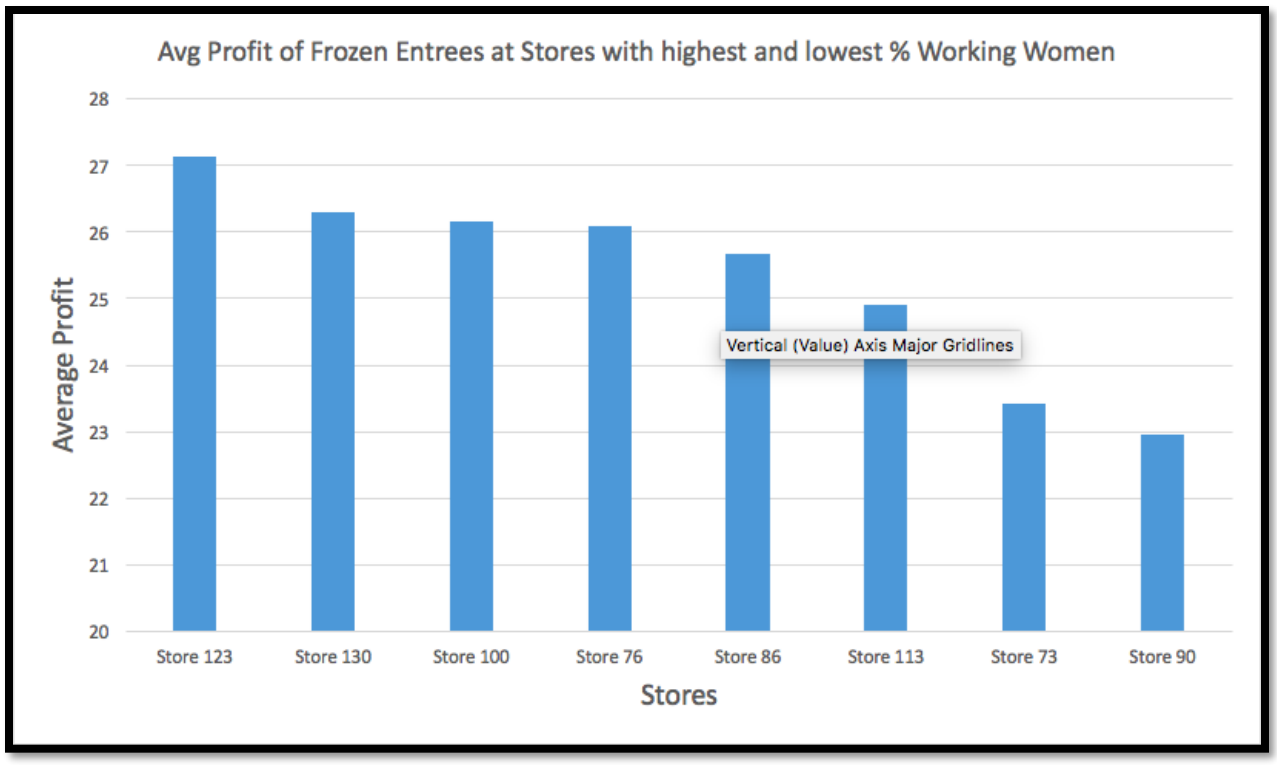
Question 8 - What is the profitability of Frozen entrees in stores with the highest % of working women with children and in stores with the lowest % of working women with children under 5 in the city of Chicago?

Priority – LOW

Entity-Relationship Diagram:



Row Labels	Average of PROFIT
Store 123	27.1403256
Store 130	26.28788249
Store 100	26.16837803
Store 76	26.09466037
Store 86	25.67508193
Store 113	24.90859526
Store 73	23.41431825
Store 90	22.95674972



Row Labels	Average of WRKCH5
Store130	0.085318923
Store 76	0.069974047
Store 86	0.067163417
Store 100	0.065638729
Store 123	0.064942947
Store 111	0.061604445
Store 75	0.056855662
Store 89	0.056479162
Store 128	0.052722186
Store 68	0.051282051
Store 12	0.050668216
Store 98	0.043757346
Store 53	0.043748919
Store 73	0.038221815
Store 33	0.030111206
Store 113	0.025834329

Justification

In cities like Chicago, working woman will lead a different lifestyle to a non working woman. Working women will tend to spend money on lunch snacks from shops and eat out in restaurants

more than unemployed as working women avoid cooking and this is one reason for eating out.[17] In general, there has been an increase in the number of consumers looking for frozen/prepared foods.[18]

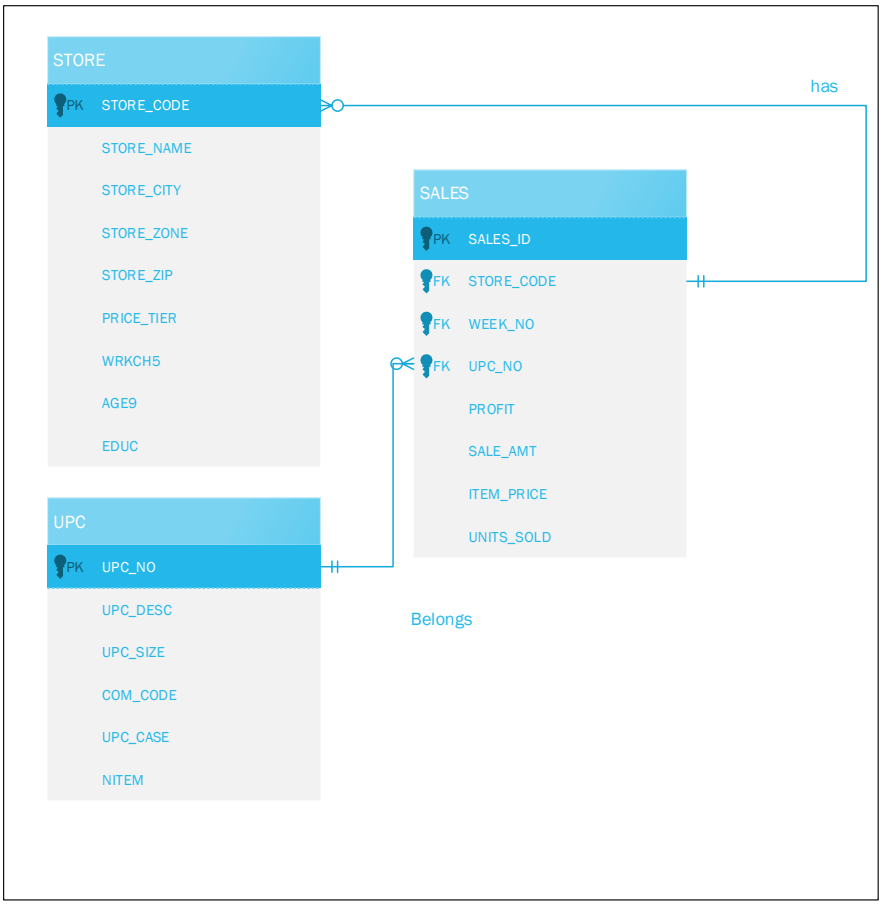
Based on this research results, we wanted to analyse the data about sales of Frozen in the stores with highest % of working women with children under five years. The results of the analysis shows that the average profit on frozen foods in stores with highest % of working women is much higher when compared to the profits acquired on frozen foods. These results will help DFF to gain more profits on the frozen food section. They can also introduce wide range of frozen food items and ready to go meals in the stores with higher % of working women.

A similar combination of product and demography, for example, jewelry with working women could be used to benefit DFF in the longer run. Such analysis could also help other stakeholders like manufacturers, employees to promote or demote any particular product in a particular area for future benefits.

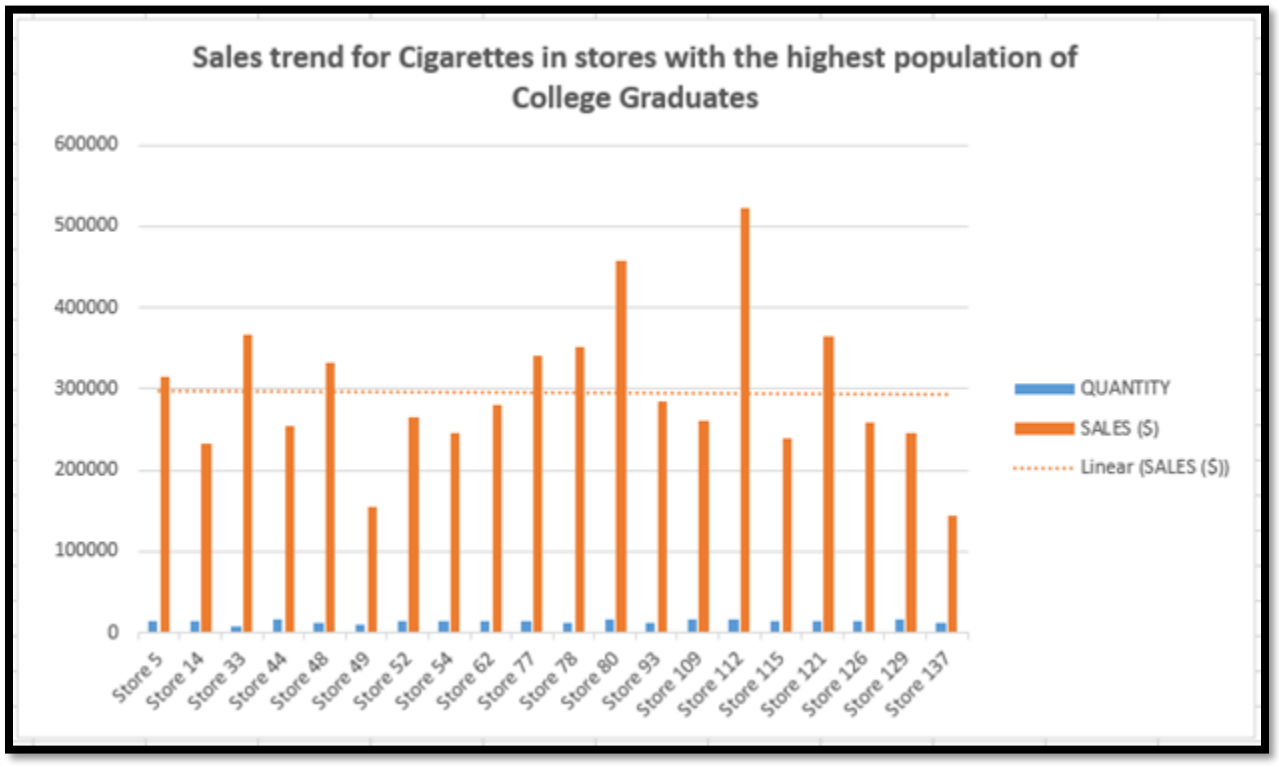
Question 9 - What is the trend in the sale of Cigarettes in stores with the highest population of College Graduates?

Priority – LOW

Entity-Relationship Diagram:



STORE	QUANTITY	SALES ($)
Store 5	14560	314225.29
Store 14	13800	232151.12
Store 33	7690	367613.9
Store 44	16242	254144.63
Store 48	12778	331682.11
Store 49	10532	155267.57
Store 52	13606	266264.88
Store 54	14202	244924.12
Store 62	14279	279929.75
Store 77	14607	341751.49
Store 78	12600	351526.29
Store 80	16641	457209.69
Store 93	11251	285303.99
Store 109	15568	260257.34
Store 112	15875	522089.11
Store 115	13904	239052.16
Store 121	15003	364015.71
Store 126	13571	258097.9
Store 129	16078	245392.67
Store 137	11688	144211.72
Grand Total	**274475**	**5915111.44**



Justification

Still analyzing the possibilities of the youth marketing strategy, we decided to study the trend for the college graduates' demographic.

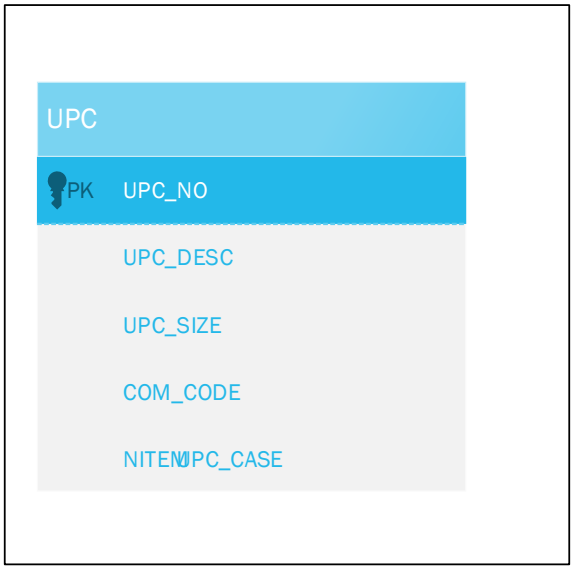
However, this time we ran the analysis for cigarettes which is a product that has had conflicting reports over the years, because of its adverse effects on health. We wanted to be able to tell whether linear sales were rising or declining and whether it was still a viable product for this demographic.

We focused on stores with the highest population of college graduates and results from DFF's data showed that the linear sales trend for cigarettes remained almost constant at about $30,000 over the years. With this information, DFF managers can better plan on profitable direct marketing strategies which they can employ to boost sales for this product.

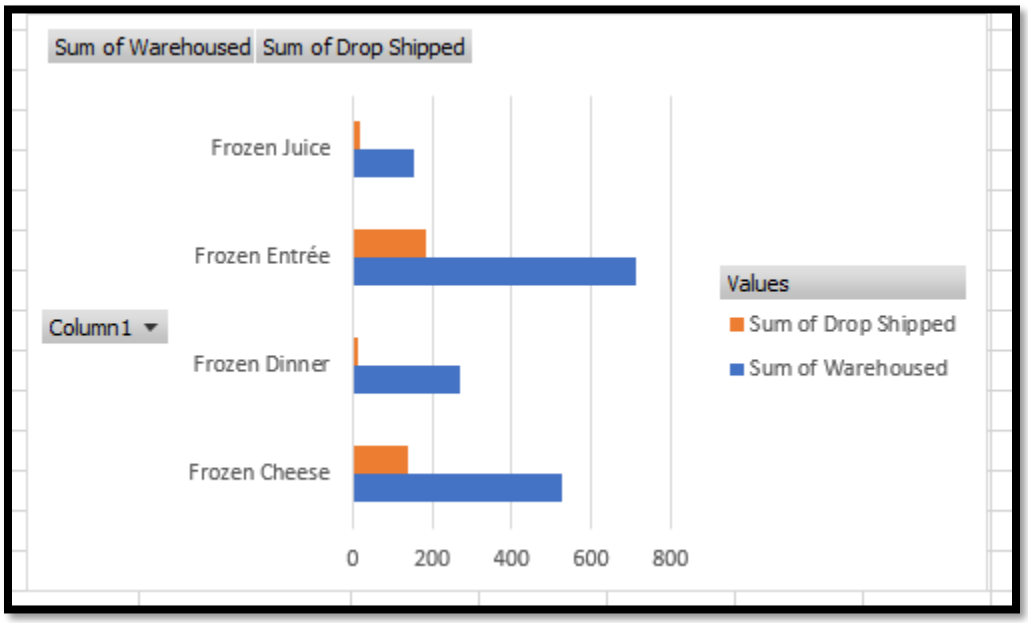
Question 10 - How many of perishable items were drop-shipped vs warehoused?

Priority – LOW

Entity-Relationship Diagram:



Row Labels	Sum of Warehoused	Sum of Drop Shipped
Frozen Cheese	526	141
Frozen Dinner	271	12
Frozen Entrée	714	186
Frozen Juice	155	19
Grand Total	**1666**	**358**



Justification

We analyzed the count of perishable items that were warehoused or drop shipped. With our limited data on Frozen Juice, Frozen entree, Frozen Dinner and cheese, we found that all of them were warehoused more often than drop shipped. And we concluded that maybe Dominick's could have drop shipped such items more often than they did and hence prevented the losses to such perishable items.

There are formidable investment challenges for storing frozen items. A major hurdle is the cost of cold storage facility. This can be avoided completely by drop shipping frozen items.[19]

Logical Design of the Data Warehouse Schema

Section 1: Data Marts and Dimension Matrix

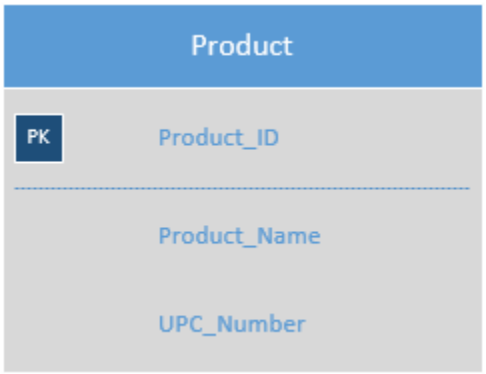
In this project, Kimball's Dimensional Modeling technique is used to create STAR schema and HOLAP style in the design of data warehouse for Dominic's Finer Foods retail store chain. This section of the reports has the details of the different dimensions of Dominic's Data Warehouse and fact tables which would help answer the business questions we have designed earlier. This report also provides a clear justification of business questions corresponding to independent Data Marts.

1.1. Dimension Tables

There are four dimensions created namely, Product dimension, Store dimension, Category dimension and Time dimension. Each dimension table and its attributes are described below.

- DProduct

The details of the various products sold by DFF retail store chain are defined in the Product Dimension table (DProduct). The attributes in the Product Dimension table helps define the facts or measures that are stored in the Product Sales fact table.



DProduct

The attributes of the DProduct dimension table are as follows:

DProduct - Dimension table defining the products sold in the DFF retail store chain.	
Product_ID	A surrogate key designated as the unique identifier for the product dimension table
Product_Name	Name of a particular product
UPC_Number	UPC number of the product

- DStore

The details of the various stores in the DFF retail chain are defined in the Store Dimension table (DStore). The attributes in the Store Dimension table help define the facts or measures that are stored in Product Sales fact table as well as Category Sales fact table.

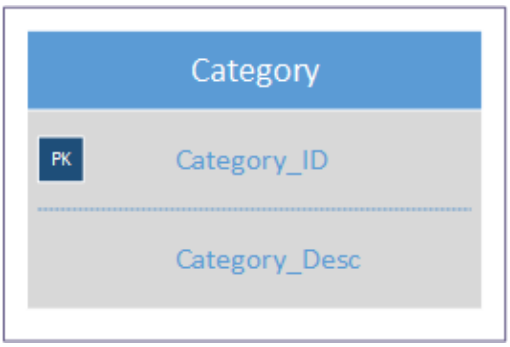
DStore

Following is the description of the attributes of the DStore dimension table –

DStore - Dimension table defining the stores in the DFF retail store chain.	
Store_ID	Unique identifier of Store dimension table; A surrogate key
Store_Number	Store number
Store_Name	Name of the store
Store_City	City of the store
Age_Nine	% Population under age 9
EDUC	% College Graduates

- DCategory

The details of the categories of DFF store chain are defined in the Category Dimension table (DCategory). The attributes in the Category Dimension table helps define the facts or measures that are stored in Category Sales fact table.



DCategory

The attributes of the DCategory dimension table are as follows:

DCategory - Dimension table defining the categories of DFF retail chain store.	
Category_ID	Unique identifier of category dimension table; A surrogate key
Category_Desc	Description of the category

- DTime

The Time dimension table (DTime) gives the details of the sales periods for DFF retail store chain, including special events (e.g. Valentine's) or holiday periods. The attributes in the Time Dimension table help define the facts or measures stored in both the Product Sales fact table and the Category Sales fact table.

DTime

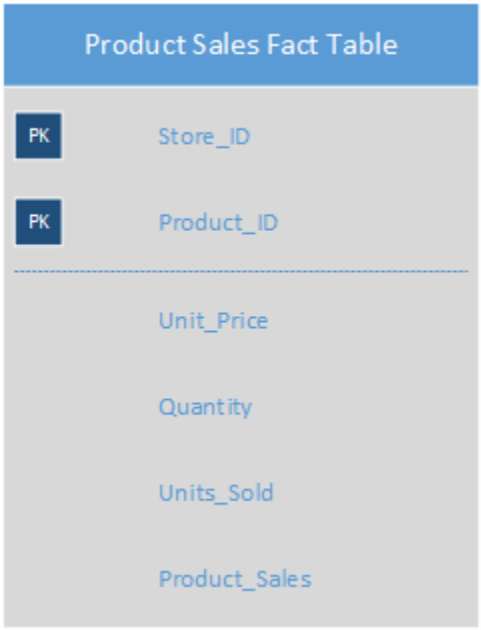
The attributes of the DTime dimension table are as follows:

DTime - Dimension table defining sales periods for DFF retail chain store.	
Time_ID	Unique identifier of the time dimension table; a surrogate key.
Year	The given year
Week_Number	The specific week number. The calendar weeks have been converted to database weeks with start and end dates)
Special_Event	Special events and/or holidays

1.2. Fact Tables

- Product Sales fact table

This table is located at the center of the Product Sales STAR Schema. It has the information of the total sales (in dollars) pertaining to different products of the Dominic's Finer Foods retail store chain.



ProductSalesFact

Fact Table Keys:
Store_ID – A unique identifier for stores which has been generated as a surrogate key
Product_ID - A unique identifier for products which has been generated as a surrogate key

Fact Table Measures:
Unit_Price – Price of a bundle of the product
Quantity – Size of a bundle
Units_Sold – Number of units of a product sold

Product_Sales – This is a derived attribute. It indicates the dollar value of the product sales across various dimensions. Product_sales = (Unit_price * Number_of_units_sold) / Quantity

- Category Sales fact table

This table is located at the center of the Category Sales STAR Schema. It has the information of the total sales (in dollars) pertaining to different categories of the Dominic's Finer Foods retail store chain.

CategorySalesFact

Fact Table Keys:
Time_ID - A unique identifier for time dimension that has been generated as a surrogate key.
Category_ID - A unique identifier for category dimension that has been generated as a surrogate key.
Fact Table Measures:
Category_sales – This attribute is the total number of sales (in dollars) across different dimensions.

1.3. Dimension Matrix for Data Marts

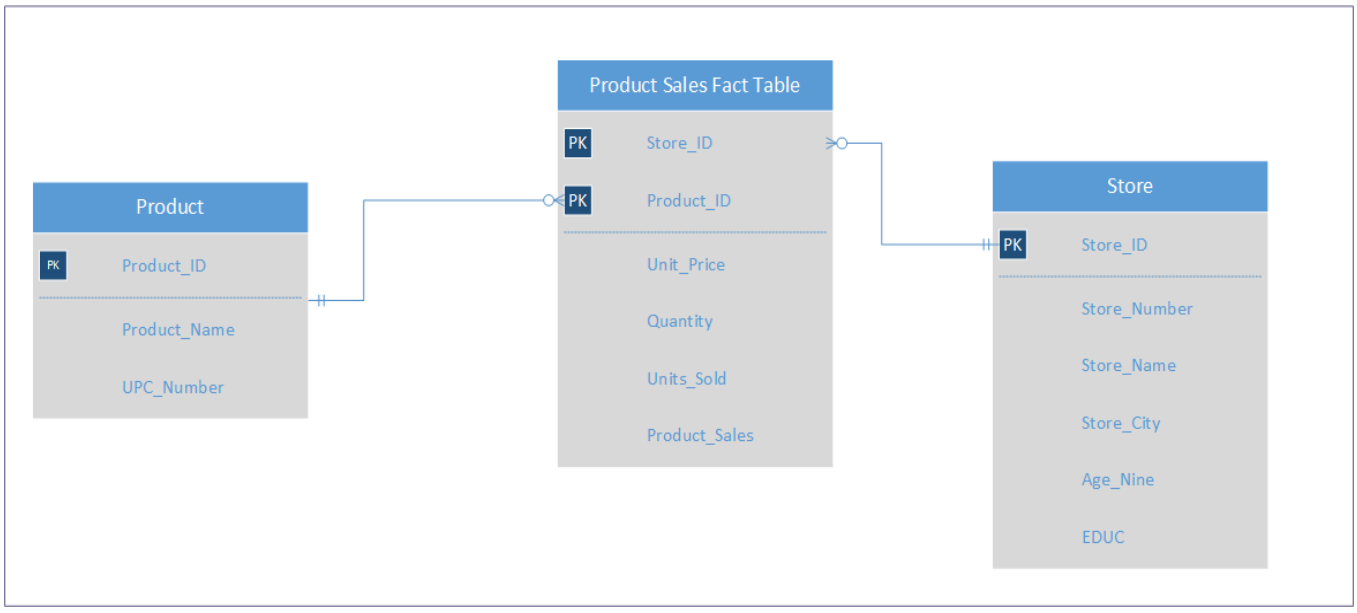
Dimension / Data Mart	DProduct	DStore	DCategory	DTime
Product Sales	✓	✓		
Category Sales			✓	✓

1.4. Star Schemas

The fact tables and dimension tables have been integrated to create two STAR schemas for the data marts of DFF's data warehouse. Those are Product Sales and Category Sales. The data marts would help answer the business questions we have designed and have been addressed in Section 3 of this report.

Star Schema for Product Sales data mart

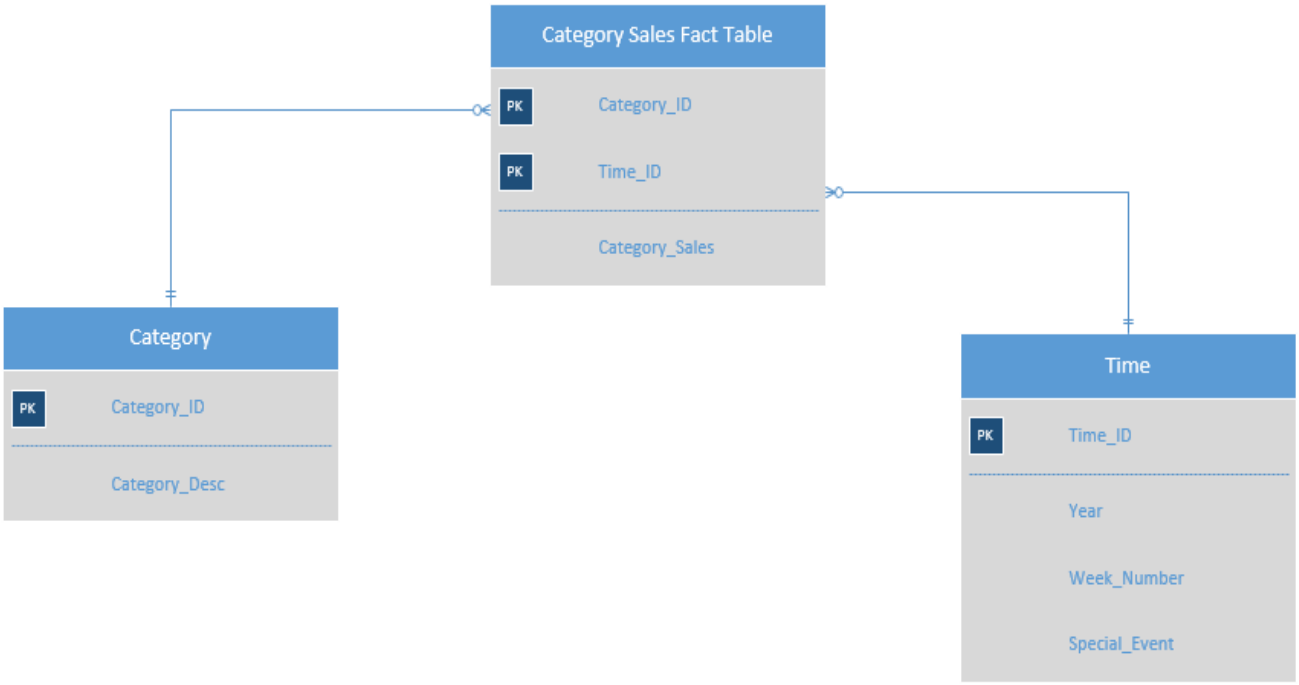
Following is the STAR schema for the Product Sales data mart. It consists of Product Sales fact table i.e. ProductSalesFact and two dimension tables, DProduct and DStore. We have used this data mart schema to address the business questions 4 and 5 and the details are provided in Section 3 of this report.



Product Sales STAR Schema

Star Schema for Category Sales data mart

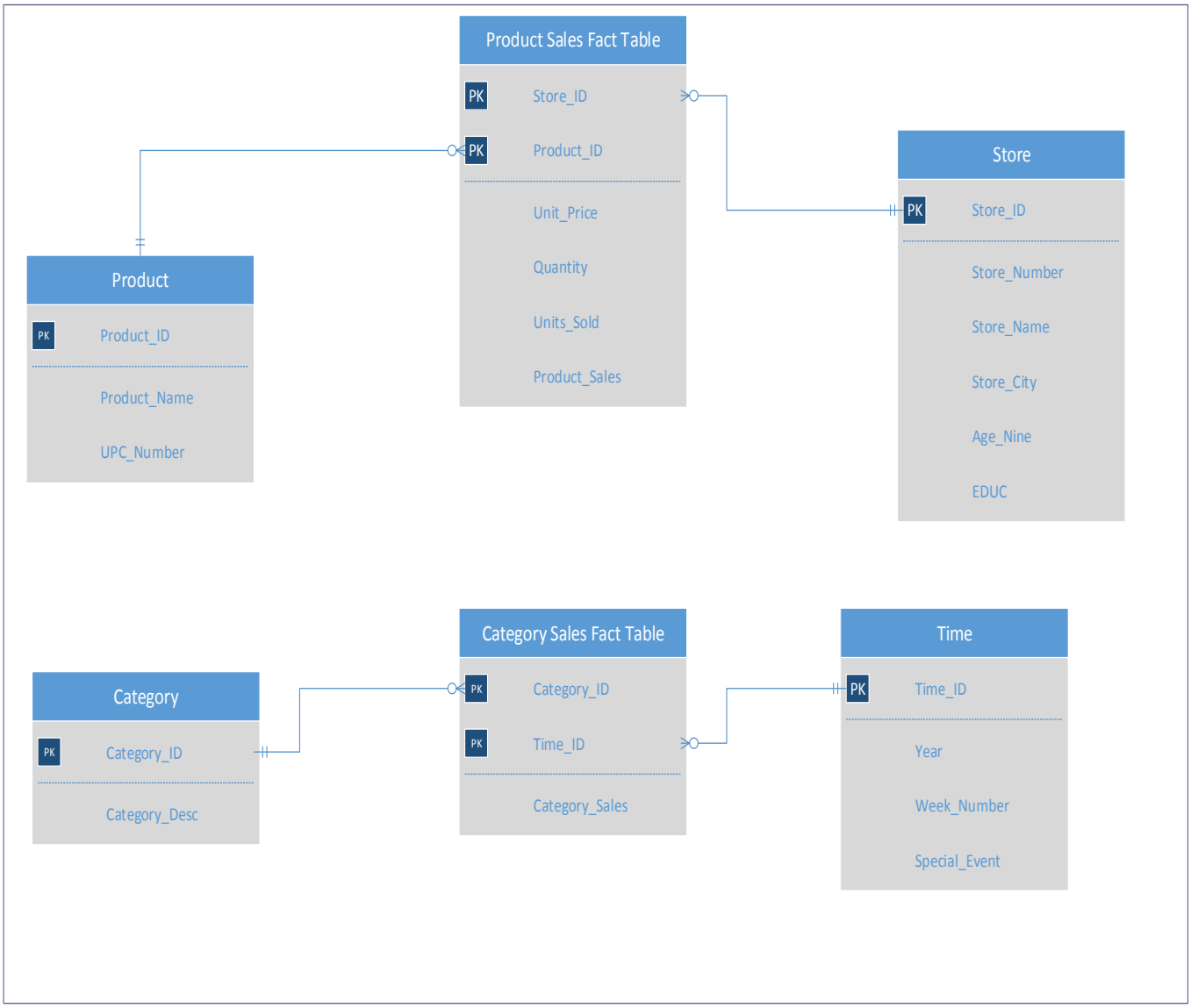
The STAR schema for the Category Sales data mart consists of the Category Sales fact table i.e. CategorySalesFact and two dimension tables, DCategory and DTime. We have used this data mart schema to address the business questions 1, 2 and 3. The details are provided in Section 3 of this report.



Category Sales STAR Schema

Section 2: Data Warehouse Schema and Mapping tables

2.1. Data Warehouse Schema



Data Warehouse Schema

2.2. Mapping Tables for Product Sales Data Mart

Dimension Table	Dimension Attribute	Source Table	Source Table Attribute	Mapping Function
Product		UPC		
	Product_ID (Row Number)			
	Product_Name		Description	
	UPC_Number		UPC	

Dimension Table	Dimension Attribute	Source Table	Source Table Attribute	Mapping Function
Store		DFF's Store Specific Demographics		
	Store_ID (Row number)			
	Store_Number		Store	
	Store_Name		Name	
	Store_City		City	
	Age_Nine		Age9	
	EDUC		EDUC	

Fact Table	Fact Table Attribute	Source Table	Source Table Attribute	Mapping Function
Product Sales Fact Table		Movement Table		
	Store_ID			Unique identifier (Primary key) of the Store dimension
	Product_ID			Unique identifier (Primary key) of the Product dimension
	Unit_Price		price	
	Quantity		qty	
	Units_Sold		move	
	Product_Sales			Derivation: Sales = (price * move)/qty

2.3. Mapping Tables for Category Sales Data Mart

Dimension Table	Dimension Attribute	Source Table	Source Table Attribute	Mapping Function
Time		Week Decode Table		
	Time_ID (Row Number)			
	Year			Can be obtained from Start and End dates in Week Decode Table
	Week_Number		Week	Week number
	Special_Event		Special Events	

Fact Table	Fact Table Attribute	Source Table	Source Table Attribute	Mapping Function
Category Sales Fact Table		CCOUNT		
	Category_ID			Unique Identifier (Primary Key of the Category Dimension)
	Time_ID			Unique Identifier (Primary Key of the Time Dimension)
	Category_Sales			Sale of categories like beer, wine, grocery etc

Dimension Table	Dimension Attribute	Source Table	Source Table Attribute	Mapping Function
Category		CCOUNT		
	Category_ID (Row Number)			
	Category_Desc			Category descriptions from Ccount file like beer, wine, grocery etc

Section 3: Business Questions Corresponding to Independent Data Marts

Question 1 - What impact do holidays have over sale of beer and wine and which holidays have the highest demands?

This business question is addressed by Category Sales data mart. The special_event and week attribute from the Time dimension table DTime is used to get the sales data by holidays. The Category_Sales measure from Category Sales fact table gives the required information about the sales of the categories Beer and Wine. So, using the week dimension attribute and the Category_Sales measure, we can analyze the sales of beer and wine during the holidays.

Question 2 - What is the trend in wine sales over the years during Valentine's week?

This business question is addressed by the Category Sales data mart. The sales data during Valentine's week is gotten from the Special_Event and Week attributes of the Time dimension table (DTime) for the wine category. Then obtaining the Category_Sales measure from the Category Sale fact table, we can thus analyze the trend in the sale of wine over the years during Valentine's week.

Question 3 - How do the total sales of Grocery and Dairy vary over years? What happened in 1997?

This business question is addressed by Category Sales data mart. The year attribute from the Time dimension table DTime is used to get the sales data by years (1990, 1993, 1996, 1997). The Category_Sales measure from Category Sales fact table gives the required information about the sales of the categories Grocery and Dairy. So, using the year dimension attribute and the Category_Sales measure, we can analyze the sales of grocery and dairy during the selected years. And we check the external factors to help understand the dip in 1997.

Question 4 - What is the trend in the sale of cookies (from week 1 – 399) in stores with the highest % population of kids under age 9?

This business question is addressed by Product Sales Store data mart. The age_nine attribute from the Time dimension table DTime is used to get the data about store with highest number of children under age 9. The Product_Sales measure from Product Sales fact table gives the required information about the sales of the cookies and hence we can use this information to find the trend of the cookie sales in stores with highest population of kids under age 9.

Question 5 - What is the trend in the sale of Cigarettes in stores with the highest population of College Graduates?

This business question is addressed by Product Sales Store data mart. The educ attribute from the Time dimension table DTime is used to get the data about store with highest number of college graduates. The Product_Sales measure from Product Sales fact table gives the required information about the sales of the cigarettes and hence we can use this information to find the trend of the cigarette sales in stores with highest population of college graduates.

ETL Design using SSIS

Section 1: Data Quality issues in the DFF data sets

Accurate and high quality data is very critical to the building of a comprehensive data warehouse that can support business analytics and provide business intelligence to the users of DFF data. Amongst other things, it reduces or eliminates the risks of strategic business decisions based on wrong or incorrect data and facts. During the data integration process, we discovered many data quality issues with our source DFF data sets, e.g. null and inconsistent data values, incorrect and non-meaningful values etc.

The following table categories the types of data quality issues considered and gives details on the specific issues encountered with the DFF data sets.

Category	Quality	Issues Considered	Data Quality Problems with DFF Data Sets
Relation to other data	Referential integrity	Do records exist where expected? Do they contain unnecessary or inactive data? Are reference files/tables complete?	In DFF's data, all records did not exist in the expected places. There were a lot of null and non-meaningful values in the source files. There were also some referential integrity issues, for instance, some products in the CCOUNT source file had product sales values for stores with Store Code as '0', even though in the data manual, no stores had Store Code = 0
	Cardinality	Is the structure of relationships among entities and attributes maintained consistently?	There were many inconsistencies in the structure of relationships among entities and attributes.
Structure of fields	Format	Do values follow consistent formatting standards?	There were inconsistences in some of the data values e.g. Some UPC numbers had 10-digit formats (xxxxxxxxxx) while others had 3, 8 and 9-digit formats etc.
	Standard	Are data elements consistently defined and understood?	Information about the data elements were well defined in the data manual but not easy to understand when looking at the data sets because of the various inconsistencies e.g. in CCOUNT file, the negative values in

			dollar sales was difficult to decipher whether they were dirty data or meant loss in sales.
	Consistent	Do values represent the same meaning across systems and files?	Though data was well defined in the manual to represent the same meaning across systems and files, inconsistencies in data values that were present in multiple places (e.g. UPC number) gave misrepresentations of the data.
Content within data values	Complete	Is all necessary data present?	Some of the necessary data were not present and had to be derived. For examples, the product sales values for the Cookies and Cigarettes products (Done-WCOO and Done-WCIG) had to be derived in order to perform the sales trend analysis.
	Accurate	Does the data accurately represent reality or a verifiable source?	DFF data sets represents real data collected from the actual DFF stores, however, there were many inaccuracies in the data too. For instance, the null values, the negative sales values, the meaningless values etc.
	Valid	Do data values fall within acceptable ranges defined by the business?	Not all data was valid because they fall outside the acceptable ranges defined by the business, e.g. sales values for stores with Store Code = 0, because 0 is an invalid store code.
	Fit for purpose	Is the information valuable to the business? Does the data convey information that can intelligently be consumed by the business?	The DFF data sets is nevertheless fit for purpose because it contains a lot of information on sales and customer data which are extremely valuable to the business.

Section 2: ETL Plan

The plan that we shall follow is to get data out of the source and load it into the staging tables, make a copy of those tables and then extract, transform and load the data needed as per our business questions in the staging table. For data warehouse loading, it shall simply be a process of copying data from one database to other. The DFF data sets was fit for purpose because it contained a lot of information on sales and customer data which are extremely valuable to the business.

"Data extraction is where data is extracted from homogeneous or heterogeneous data sources; data transformation where the data is transformed for storing in the proper format or structure for the purposes of querying and analysis; data loading where the data is loaded into the final target database, more specifically, an operational data store, data mart, or data warehouse."

Following are steps involved in the ETL plan for data warehouse implementation-

2.1. Determining all the target data needed in the data warehouse

DATA SOURCE	DATA EXTRACTION	DATA TRANSFORMATION	DATA IN DATA WAREHOUSE
CCOUNT	CCOUNT	CCOUNT_COL	DCatgeory, Category Sales Fact
WEEK	Week_Decode	Week_Decode	DTime
STORE	Store	STORE_COL	DStore
MOVEMENT	DONE-WCIG, DONE-WCOO	Merged	Product Sales Fact
UPC	UPCCIG, UPCCOO	PROD_MID	DProduct

2.2. Determining all the data sources

University of Chicago Booth School of Business has provided the Dominick's database as the data sources for the data warehouse implementation. The files that are required according to our business questions are listed as below:

DATA	SOURCE FILES
CUSTOMER COUNT	CCOUNT
WEEK	Week_Decode (Created)
STORE	STORE (Created)
MOVEMENT	DONE-WCIG, DONE-WCOO
UPC	UPCCIG, UPCCOO

2.3. Preparing data mappings for data elements from sources in CSV to staging and then data mapping from staging to data warehouse (include all transformations)

SOURCE FILE NAME	SOURCE FILE ATTRIBUTES	INITIAL STAGING AREA TABLE NAME	INITIAL STAGING AREA TABLE ATTRIBUTES
CCount.csv	Store	CCOUNT_COL	Store
	Date		Date
	Week		Week
	Beer		Beer
	Wine		Wine
	Grocery		Grocery
	Dairy		Dairy
Store.csv	store number	STORE_COL	Store
	city		City
Week_decode.csv	week number	WEEKDECODE	Week_NO
	start date		Start_Date
	end date		End_Date
	special events		Special_Events
upccig.csv, upccoo.csv	UPC	UPCCIG_COL, UPCCOO_COL	UPC
	Descrip		Descrip
DONE-WCIG.csv DONE-WCOO.csv	Price	WCIG_COL WCOO_COL	Price
	Qty		Qty
	Move		Move
	Store		Store
	UPC		UPC
	Sale		Sale

INITIAL STAGING AREA TABLE NAME	INITIAL STAGING AREA ATTRIBUTE	FINAL STAGING AREA TABLE NAME	FINAL STAGING AREA ATTRIBUTE	MAPPING FUNCTION
CCOUNT_COL		DCATEGORY	Category_ID	Used UNPIVOT
			Catgeory_Name	
STORE_COL, DEMO_COL		DSTORE	Store_ID	Used Inner Join
			Store_Number	
			Store_Name	
			Age_Nine	
			EDUC	
			Store_City	
WCIG_COL WCOO_COL →PRODFACT_MID-→Prodfa3	Used Sort and Merge to get PRODFACT_MID to get combined data from the Cigarettes and Cookies UPC Files. And then Used Lookup on Store and Product to make Prodfa3	PRODUCTSALESFACT	Store_ID	Surrogate Key
			Product_ID	Surrogate Key
			Unit_Price	
			Quantity	
			Units_Sold	
			Product_Sales	Derived from the formula Unit_Price * Units_Sold / Quantity
CCOUNT_sample, DCategory, DTime	Used Lookup for Category and Time on corrected CCOUNT file	CATEGORYSALES_FACT	Category_ID	Surrogate Key
			Time_ID	Surrogate Key
			Category_Sales	Weekly Sales of particular categories
UPCCIG_COL, UPCCOO_COL ,PROD_MID	Used Derived Column to get PROD_MID to get Classification from the Cigarettes and Cookies UPC Files.	DPRODUCT	Product_ID	Using Row_Number
			Product_Name	
			UPC_Number	
			Classification	Used Derived Column on UPCCIG_COL, UPCCOO_COL

FINAL STAGING AREA TABLE NAME	FINAL STAGING AREA ATTRIBUTE	DATAWAREHOUSE TABLE	DATAWAREHOUSE ATTRIBUTES	MAPPING FUNCTION
WEEKDECODE_COL		DTime	Time_ID	Surrogate key
	Week_NO		Week_NO	Direct Mapping
	Start_Date		Year	RIGHT(Start_Date,4)
	End_Date			
	Special_Events		Special Events	Direct Mapping
DCATEGORY	Category_ID	DCategory	Category_ID	Direct Mapping
	Catgeory_Name		Category_Desc	
CATEGORYSALES_FACT		Category_Sales_Fact	FactCategorySales_ID	Surrogate Key
	Category_ID		Category_ID	Direct Mapping
	Time_ID		Time_ID	
	Category_Sales		Category_Sales	
DSTORE	Store_ID	DStore	Store_ID	Direct Mapping
	Store_Number		Store_Number	
	Store_Name		Store_Name	
	Age_Nine		Age_Nine	
	EDUC		EDUC	
	Store_City		Store_City	
DPRODUCT	Product_ID	DProduct	Product_ID	Direct Mapping
	Product_Name		Product_Name	
	UPC_Number		UPC_Number	
	Classification		Classification	
PRODUCTSALESFACT		Product_Sales_Fact	FactProductSales_ID	Surrogate Key
	Store_ID		Store_ID	Direct Mapping
	Product_ID		Product_ID	
	Unit_Price		Unit_Price	
	Quantity		Quantity	
	Units_Sold		Units_Sold	
	Product_Sales		Product_Sales	

2.4. Establishing comprehensive data extraction rules

Before extracting data from the source systems, we took care of very detailed information such as the database size and planned the time of extraction when there were relatively few users. The data from the source files were loaded into the data staging area 601-GROUP14-STAGING. Comma separated Value (csv) files were created for Week and Store details using the DFF's data manual provided. The data from sources were CSV files. Using Microsoft SQL Server Studio, the diverse types of data (different UPCs) were combined and converted into tables. Later, the tables were transformed further as per the business questions and finally loaded into the data warehouse.

2.5. Determining data transformation and cleansing rules

Accurate and high quality data is very critical to the building of a comprehensive data warehouse that can support business analytics and provide business intelligence to the users of DFF data. Amongst other things, it reduces or eliminates the risks of strategic business decisions based on wrong or incorrect data and facts. During the data integration process, we discovered many data quality issues with our source DFF data sets, e.g. null and inconsistent data values, incorrect and non-meaningful values etc.

In DFF's data, all records did not exist in the expected places. There were a lot of null and non-meaningful values in the source files. There were also some referential integrity issues, for instance, some products in the CCOUNT source file had product sales values for stores with Store Code as '0', even though in the data manual, no stores had Store Code = 0. There were many inconsistencies in the structure of relationships among entities and attributes. There were inconsistencies in some of the data values e.g. Some UPC numbers had 10-digit formats (xxxxxxxxxx) while others had 3, 8 and 9-digit formats etc.

Information about the data elements were well defined in the data manual but not easy to understand when looking at the data sets because of the various inconsistencies e.g. in CCOUNT file, the negative values in dollar sales was difficult to decipher whether they were dirty data or meant loss in sales. Some of the necessary data were not present and had to be derived. For examples, the product sales values for the Cookies and Cigarettes products (Done-WCOO and Done-WCIG) had to be derived in order to perform the sales trend analysis.

After Data extraction, transformation and loading is done to ensure the clean and consistent data as per the business questions. The clean data is finally loaded into the DW area so as to create the data marts and get the appropriate results.

The general transformation and cleaning rules that were applied to the data are as follows:

i. **Surrogate keys creation**
   Before loading the data in the two warehouses, Category Sales and Product Sales, surrogate keys were created for all the dimension tables and fact tables.

ii. **Format Revisions**
   The data types were changed when loading into the data warehouse as per the requirements.
   The date has been split using functionalities in SSIS and then stored as Year and Month attributes.

iii. NULL values removal

Null values present in the data and that were created while loading data into staging tables were deleted. Also, the blank records, rows with just a '.' (Dot) and other non-meaningful values such as unnecessary negatives were removed.

iv. Retention of Meaningful data

Attributes that were not relevant to the business questions will be removed. For instance, attributes other than Grocery, Beer, Wine, Diary and Week in the Customer Count files were removed.

v. Derived attributes

The 'Category_Sales' measure in the fact table 'Category_Sales_Fact' is obtained from the

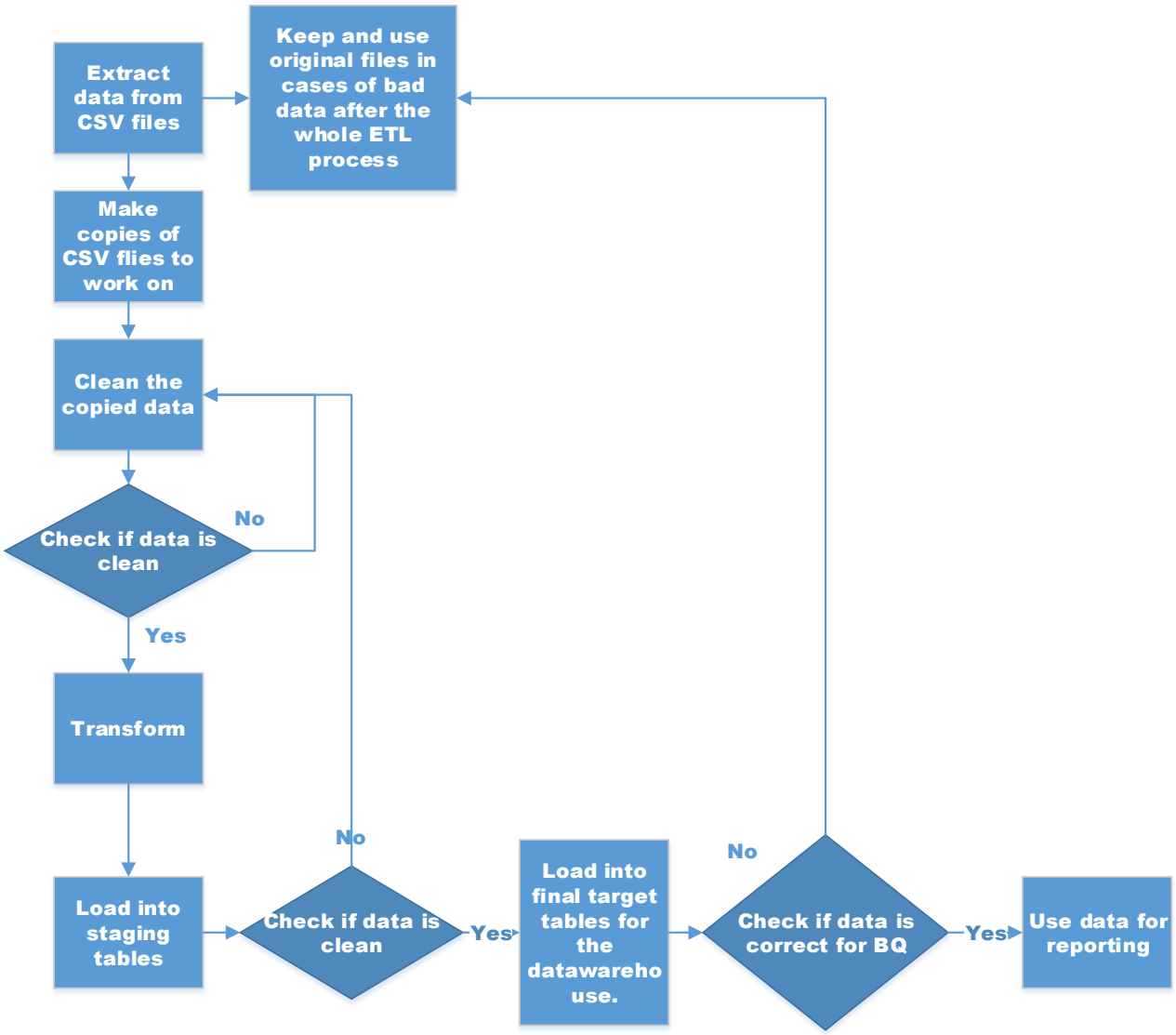
Category sales of a particular week. And 'Product_Sales = (Unit_price * Number_of_units_sold) / Quantity'. And, Year in the Time dimension, obtained from the functions YEAR (date)

SSIS Functions

The SSIS functions that were used during the transformation and cleaning process include:

1. **UNPIVOT:** It performs almost the reverse operation of PIVOT, by rotating columns into rows and was used to get beer and wine sales.
2. **LOOKUP:** For the final fact tables, lookup function was used to join additional columns to the data flow by looking up values in the dimension tables.
3. **DERIVED COLUMNS:** New column values were created as per the business question needs by applying expressions to input columns.
4. **SORT AND MERGE:** Using, sort and merge, movement files were merged.

vi. Procedure for all Data Extraction and Loading



Section 3: ETL implementation for dimension tables and fact tables

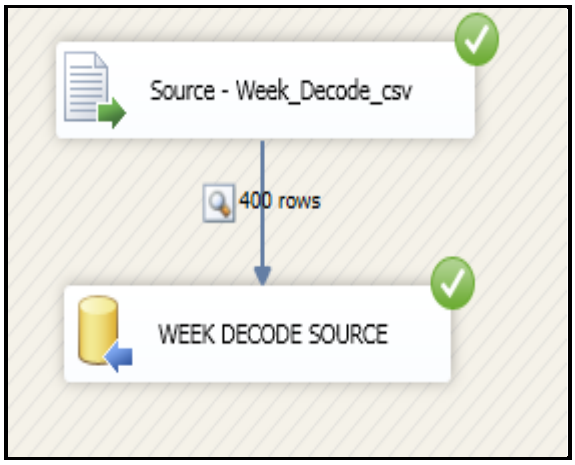
Using the SQL Server Management Studio and SSIS (Microsoft Visual studio – SQL Server Data Tools), the ETL process for the data warehouse was implemented. The steps involved are provided as below, along with the SSIS tool screen shots explaining the entire process step-by-step:

3.1. Extraction and Transformation of Source data into Dimensions and Fact Tables

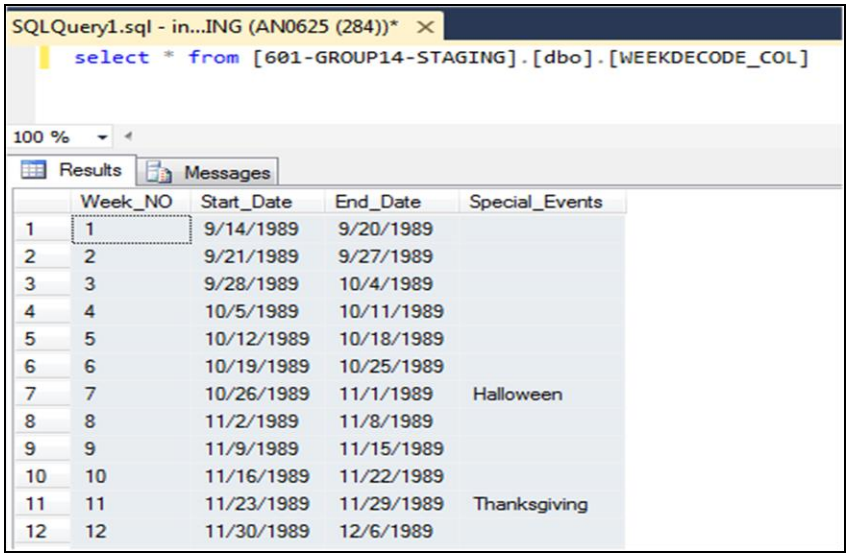
SQL Queries to create tables in DW area

3.1.1. Time Dimension Creation

Extracting Week decode source file into the staging area

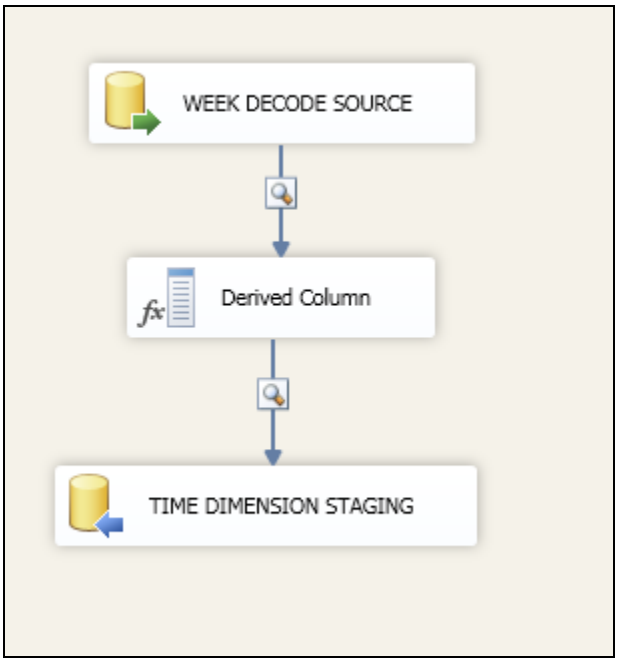


Snapshot of Week Decode table in the Staging Area



	Week_NO	Start_Date	End_Date	Special_Events
1	1	9/14/1989	9/20/1989	
2	2	9/21/1989	9/27/1989	
3	3	9/28/1989	10/4/1989	
4	4	10/5/1989	10/11/1989	
5	5	10/12/1989	10/18/1989	
6	6	10/19/1989	10/25/1989	
7	7	10/26/1989	11/1/1989	Halloween
8	8	11/2/1989	11/8/1989	
9	9	11/9/1989	11/15/1989	
10	10	11/16/1989	11/22/1989	
11	11	11/23/1989	11/29/1989	Thanksgiving
12	12	11/30/1989	12/6/1989	

Transforming Week decode to Time Dimension



Snapshot of Time Dimension table

SQLQuery1.sql - in...ING (AN0625 (284))*

```
select * from [601-GROUP14-STAGING].[dbo].[DTIME]
```

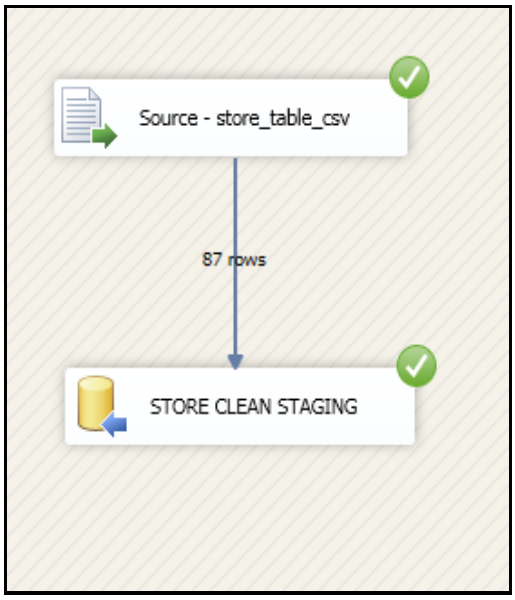
100 %

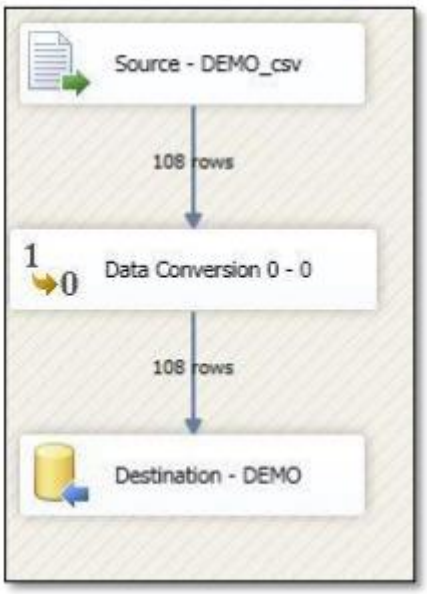
Results | Messages

	Time_ID	Week_NO	Year	Special_Events
1	1	1	1989	
2	2	2	1989	
3	3	3	1989	
4	4	4	1989	
5	5	5	1989	
6	6	6	1989	
7	7	7	1989	Halloween
8	8	8	1989	
9	9	9	1989	
10	10	10	1989	
11	11	11	1989	Thanksgiving
12	12	12	1989	
13	13	13	1989	
14	14	14	1989	

3.1.2. Store Dimension Creation

Extracting the Store and Demo source file into the staging area





SQL Queries Used to clean Demographics data

```
delete from [601-GROUP14-STAGING].[dbo].[DEMO_COL]
where NAME = 'NULL' OR NAME = ''
or CITY = 'NULL' OR CITY = ''
or STORE = '.'
or AGE9 = 'NULL' OR AGE9 = '.'
or EDUC = 'NULL' OR EDUC = '.'
```

Snapshot of Demo Table

SQLQuery1.sql - in...ING (AN0625 (284))*

```
select * from [601-GROUP14-STAGING].[dbo].[DEMO_COL]
```

100 %

Results | Messages

	NAME	CITY	STORE	AGE9	EDUC
1	DOMINICKS 2	RIVER FOREST	2	0.117508576	0.2489349342
2	DOMINICKS 4	PARK RIDGE	4	0.0950895057	0.2207894147
3	DOMINICKS 5	PALATINE	5	0.1414334827	0.3212257298
4	DOMINICKS 8	OAK LAWN	8	0.123155416	0.0951732743
5	DOMINICKS 9	MORTON GROVE	9	0.1035030974	0.2221723183
6	DOMINICKS 12	CHICAGO	12	0.1056967397	0.2534129693
7	DOMINICKS 14	GLENVIEW	14	0.129589372	0.3482930237
8	DOMINICKS 18	RIVER GROVE	18	0.1100949839	0.0722464558
9	DOMINICKS 21	HANOVER PARK	21	0.1759263459	0.1775034504
10	DOMINICKS 28	MOUNT PROSPECT	28	0.1288795371	0.233162564
11	DOMINICKS 32	PARK RIDGE	32	0.0990606319	0.1982598608
12	DOMINICKS 33	CHICAGO	33	0.0460709172	0.4196880043
13	DOMINICKS 40	BRIDGEVIEW	40	0.1336846485	0.0721286047
14	DOMINICKS 44	WESTERN SPRINGS	44	0.1448834853	0.3297383876

Snapshot of Store Table

SQLQuery3.sql - inf...ster (AN0625 (205))* | SQLQuery2.sql - inf...ster (AN0625 (346))

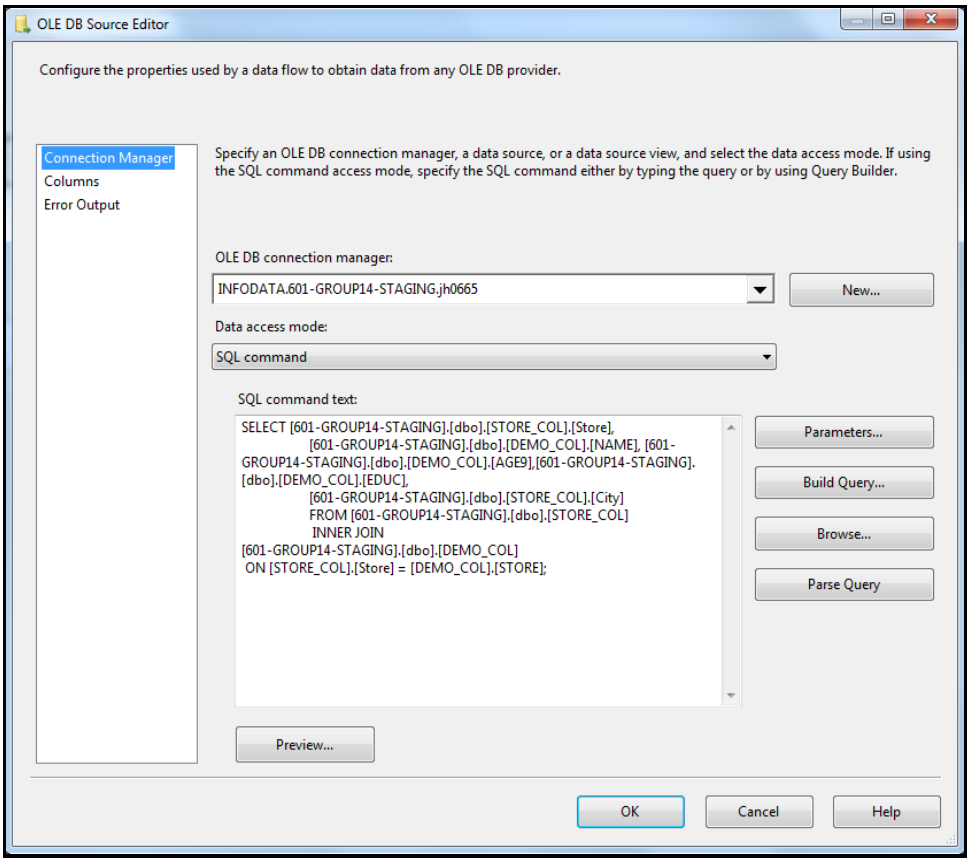
```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT *
  FROM [601-GROUP14-STAGING].[dbo].[STORE_COL]
```

100 %

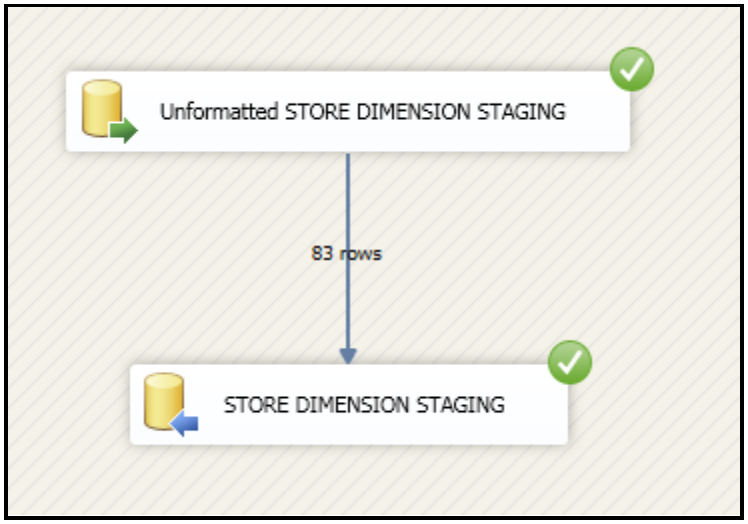
Results | Messages

	Store	City
1	2	River Forest
2	5	Palatine
3	8	Oak Lawn
4	9	Morton Grove
5	12	Chicago
6	14	Glenview
7	18	River Grove
8	21	Hanover Park
9	28	Mt. Prospect
10	32	Park Ridge
11	33	Chicago
12	39	Waukegan
13	40	Bridgeview
14	44	Western Spring

Inner join Demographics data into Store Table



Transformation of Store data into Store Dimension



Snapshot of Store Dimension

SQLQuery1.sql - in...ING (AN0625 (284))*

```
select * from [601-GROUP14-STAGING].[dbo].[DSTORE]
```

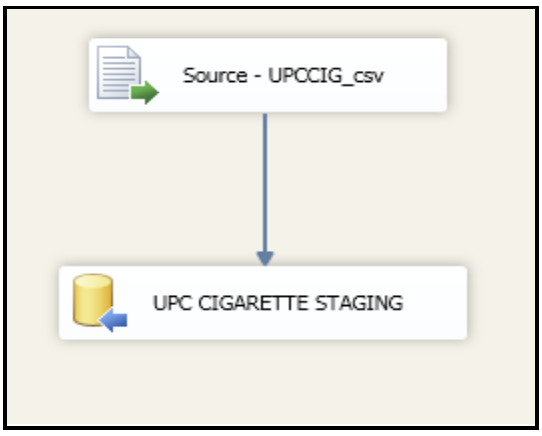
100 %

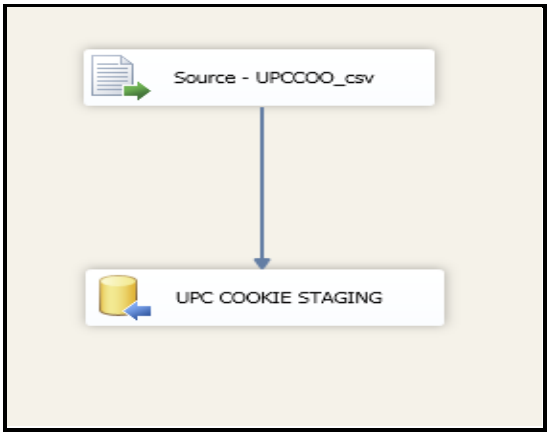
Results | Messages

	STORE_ID	Store_Number	Store_Name	Age_Nine	EDUC	Store_City
1	1	2	DOMINICKS 2	0.117508576	0.2489349342	River Forest
2	2	5	DOMINICKS 5	0.1414334827	0.3212257298	Palatine
3	3	8	DOMINICKS 8	0.123155416	0.0951732743	Oak Lawn
4	4	9	DOMINICKS 9	0.1035030974	0.2221723183	Morton Grove
5	5	12	DOMINICKS 12	0.1056967397	0.2534129693	Chicago
6	6	14	DOMINICKS 14	0.129589372	0.3482930237	Glenview
7	7	18	DOMINICKS 18	0.1100949839	0.0722464558	River Grove
8	8	21	DOMINICKS 21	0.1759263459	0.1775034504	Hanover Park
9	9	28	DOMINICKS 28	0.1288795371	0.233162564	Mt. Prospect
10	10	32	DOMINICKS 32	0.0990606319	0.1982598608	Park Ridge
11	11	33	DOMINICKS 33	0.0460709172	0.4196880043	Chicago
12	12	40	DOMINICKS 40	0.1336846485	0.0721286047	Bridgeview
13	13	44	DOMINICKS 44	0.1448834853	0.3297383876	Western Spring
14	14	45	DOMINICKS 45	0.1467187625	0.2801501642	Wheeling

3.1.3. Product Dimension Creation

Extracting UPC source file into the staging area*





**Using only Cigarette and Cookies since those are the target files*

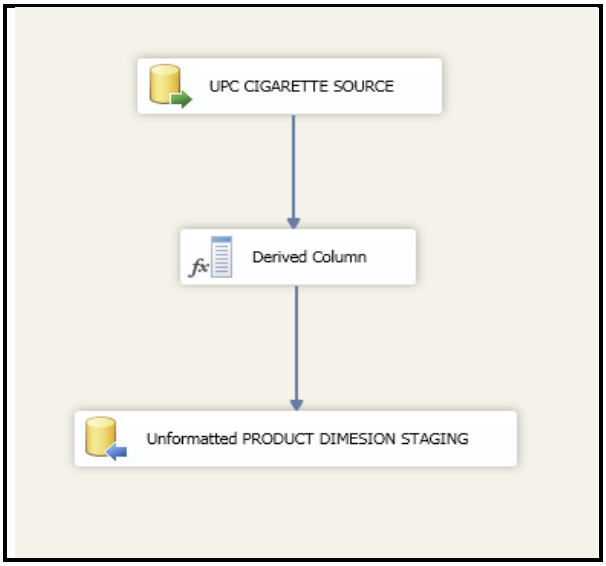
SQL Query Used to clean Cigarette UPC data

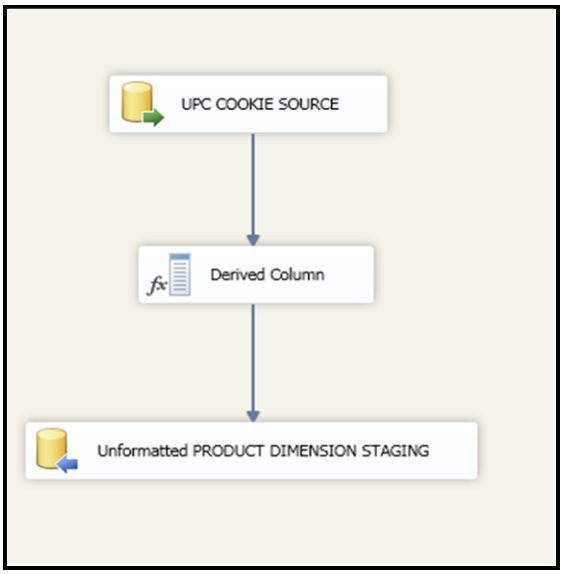
```
delete from [601-GROUP14-STAGING].[dbo].[UPCCIG_COL]
where LEN([UPC]) <> 10
```

SQL Query Used to clean Cookie UPC data

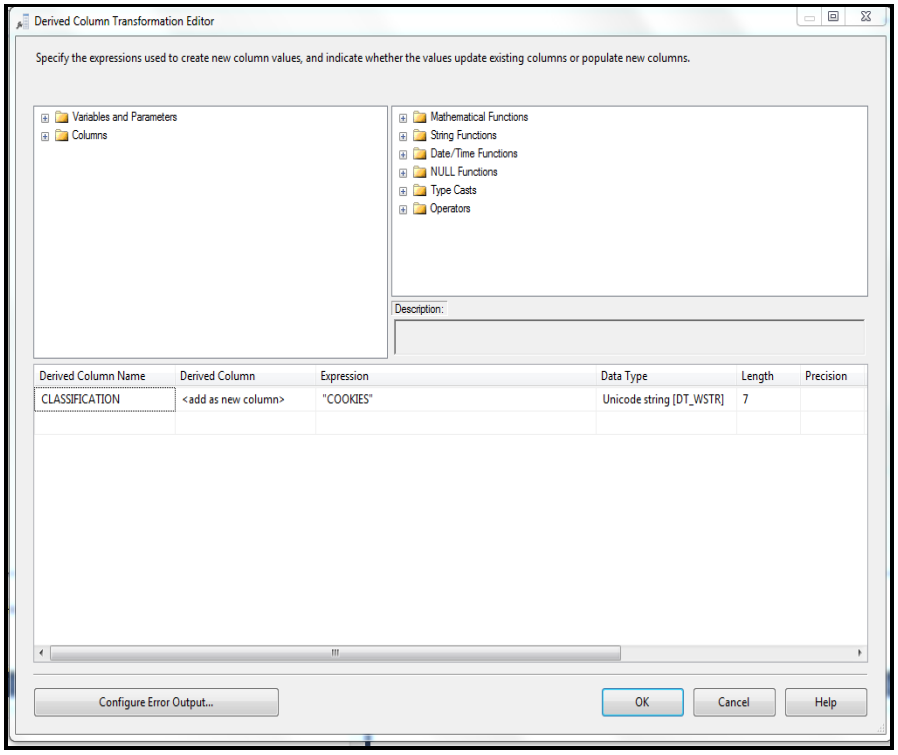
```
delete from [601-GROUP14-STAGING].[dbo].[UPCCOO_COL]
where LEN([UPC]) <> '10'
```

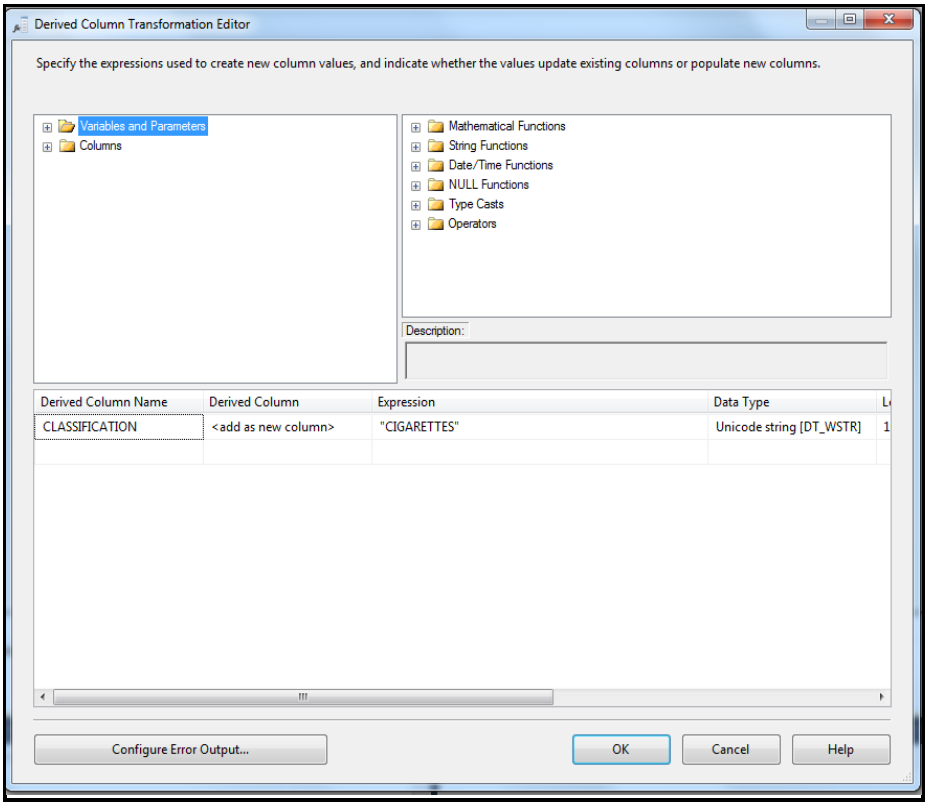
Transforming Source Data into Product Dimension





Derived Column for Product Dimension in Cigarette and Cookies





Snapshot of Data in Product Dimension

SQLQuery3.sql - inf...ster (AN0625 (205))* | SQLQuery2.sql - inf...ster (AN0625 (346))

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT *
  FROM [601-GROUP14-STAGING].[dbo].[DPRODUCT]
```

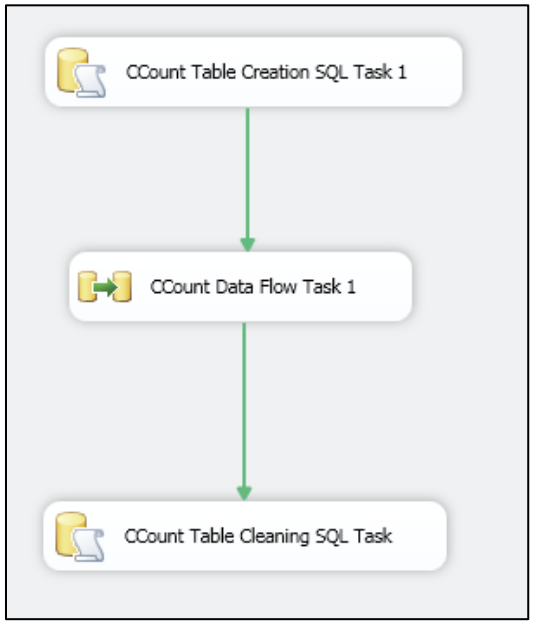
100 %

Results | Messages

	UPC_Number	Product_Name	CLASSIFICATION	PRODUCT_ID
446	1560001610	LU BUTTER TWIST	COOKIES	446
447	1560001670	LU MILK LUNCH	COOKIES	447
448	1560002322	LU LE PATISSIRE	COOKIES	448
449	1560008353	LU CREME WAFER	COOKIES	449
450	1560008358	LU SCHOOLBOY MILK	COOKIES	450
451	1560008359	LU PECAN DELLE	COOKIES	451
452	1560008361	LU JAVA CHOCOLATIE...	COOKIES	452
453	1560008362	LU SPICE CRISPS	COOKIES	453
454	1560008363	MINI MARIE LU	COOKIES	454
455	1560008370	$LU PIMS RASPBERRY	COOKIES	455
456	1560008372	LU LETRUFFE-PRALINE	COOKIES	456
457	1560008973	SAL LU CHIPS CHOC	COOKIES	457
458	1899600005	CADBURY MKCHOC C...	COOKIES	458
459	1899600006	CADBURY DKCHOC C...	COOKIES	459
460	2040000000	SUB GENERIC SINGLE	CIGARETTES	460
461	2040000001	SUB GENERIC SINGLE	CIGARETTES	461
462	2040000002	SUB GENERIC SINGLE P	CIGARETTES	462
463	2040000003	SUB GENERIC SINGLE P	CIGARETTES	463
464	2040000006	SUB GENERIC SINGLE P	CIGARETTES	464
465	2040000007	SUB GENERIC SINGLE P	CIGARETTES	465
466	2040000009	CIGARETTES (SINGLE P	CIGARETTES	466

3.1.4. Category Dimension Creation

Cleaning CCOUNT files using SQL queries



Snapshot of CCOUNT

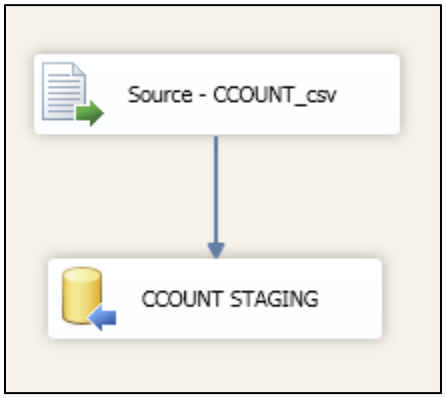
```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT * FROM [601-GROUP14-STAGING].[dbo].[CCOUNT]
```

100 %

Results | Messages

	STORE	DATE	GROCERY	DAIRY	FROZEN	BOTTLE	MVPCLUB	GROCCOUP	MEAT	MEATFROZ	MEATCOUP	FISH	FISHCOUP
1	97	920812	8999.82	2368.19	1640.23	-2.4	12.78	-202.57	1935.41	238.9	-4.38	96.92	0
2	97	920813	12183.04	2336.87	2075.06	0	13.38	-168.2	2295.98	204.27	-7.76	170.47	0
3	97	920814	14315.94	2642.23	2387.58	0	33.75	-122.15	3026.44	243.29	-4.69	348.9	0
4	97	920815	17030.72	3247.85	3120.37	-4	38.94	-226.72	4014.5	292.36	-3	240.51	0
5	97	920816	13723.61	2806.75	2301.07	0	13.98	-319.51	2736.33	247.36	-3.38	205.47	0
6	97	920817	11292.22	2435.78	1930.01	0	0	-173.64	2181.4	213.25	-4.69	141.77	0
7	97	920818	10702.26	2226.37	1749.06	-4.8	14.58	-135.05	2043.12	135.53	-2.69	95.85	0
8	97	920819	10114.11	2348.1	1779.43	-0.7	34.55	-115.01	2000.23	90.64	-5.38	168.75	0
9	97	920820	11739.15	2426.24	3391.28	-1.6	20.87	-133.04	2674.76	220.91	-14.01	104.3	0
10	97	920821	12623.61	2467.94	3061.38	-0.8	30.47	-133.31	2899.16	211.59	0	224.24	0
11	97	920822	16352.14	3390.67	3695.02	0	84.65	-107.75	4571.05	236.27	-0.75	153.94	0
12	97	920823	13186.51	2974.03	2869.96	0	43.43	-49.84	2919.39	230.3	-9.57	117.7	0
13	97	920824	10858.32	2419.76	2511.23	0	42.14	-33.62	2093.62	222.58	-2.19	121.66	0
14	97	920825	10587.97	2330.63	2600.4	-5.6	19.77	-30.45	2278.6	272.09	-2.19	201.94	0
15	97	920826	9152.44	2092.15	2497.78	-0.8	30.95	-40.44	2313.51	228.34	0	121.77	0
16	97	920827	11537.44	2559.55	1820.22	-4	23.96	-373.5	2698.32	326.03	-4.38	171.76	0
17	97	920828	13182.36	2923.37	2103.2	0	41.33	-430.14	3078.52	317.26	-4.38	347.49	0

Extraction of CCOUNT files into Staging area



SQL Queries Used to clean CCOUNT

UPDATE [601-GROUP14-STAGING].[dbo].[CCOUNT_COL]
set [store] = 'NULL'
where [STORE] LIKE '%[^0-9]%'

DELETE from [601-GROUP14-STAGING].[dbo].[CCOUNT_COL]
where [STORE] = 'NULL'

DELETE from [601-GROUP14-STAGING].[dbo].[CCOUNT_COL]
where [DATE] = 'NULL'

DELETE from [601-GROUP14-STAGING].[dbo].[CCOUNT_COL] where [WEEK] LIKE '%[^0-9]%' OR [WEEK] = '.'

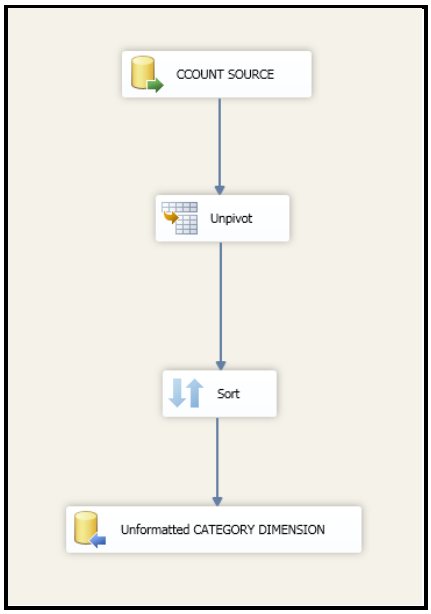
DELETE FROM [601-GROUP14-STAGING].[dbo].[CCOUNT_COL] where TRY_CAST([DAIRY] as decimal(6,0)) < 0

DELETE FROM [601-GROUP14-STAGING].[dbo].[CCOUNT_COL] where TRY_CAST([GROCERY] as decimal(6,0)) < 0

DELETE FROM [601-GROUP14-STAGING].[dbo].[CCOUNT_COL] where TRY_CAST([BEER] as decimal(6,0)) < 0

DELETE FROM [601-GROUP14-STAGING].[dbo].[CCOUNT_COL] where TRY_CAST([WINE] as decimal(6,0)) < 0

Extracting CCOUNT files in the staging area



Snapshot of CCOUNT

SQLQuery3.sql - inf...ster (AN0625 (205))* | SQLQuery2.sql - inf...ster (AN0625 (346))

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT *
  FROM [601-GROUP14-STAGING].[dbo].[CCOUNT_COL]
```

100 %

Results | Messages

	STORE	DATE	WEEK	BEER	WINE	GROCERY	DAIRY
1	97	920812	152	354.22	213.03	8999.82	2368.19
2	97	920813	153	514.94	256.07	12183.04	2336.87
3	97	920814	153	847.33	355.42	14315.94	2642.23
4	97	920815	153	694.94	415.9	17030.72	3247.85
5	97	920816	153	591.31	168.3	13723.61	2806.75
6	97	920817	153	328.16	180.88	11292.22	2435.78
7	97	920818	153	527.69	207.71	10702.26	2226.37
8	97	920819	153	310.61	217.93	10114.11	2348.1
9	97	920820	154	472.74	236.06	11739.15	2426.24
10	97	920821	154	779.67	349	12623.61	2467.94
11	97	920822	154	976.68	273.57	16352.14	3390.67
12	97	920823	154	473.08	162.54	13186.51	2974.03
13	97	920824	154	451.31	155.73	10858.32	2419.76
14	97	920825	154	429.12	221.11	10587.97	2330.63

Transforming CCount to Category Dimension

SQLQuery3.sql - inf...ster (AN0625 (205))* | SQLQuery2.sql - inf...ster (AN0625 (346))

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT *
  FROM [601-GROUP14-STAGING].[dbo].[DCATEGORY]
```

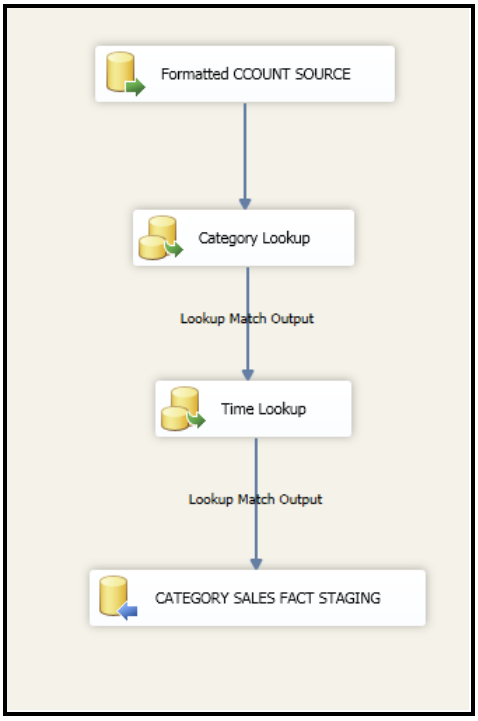
100 %

Results | Messages

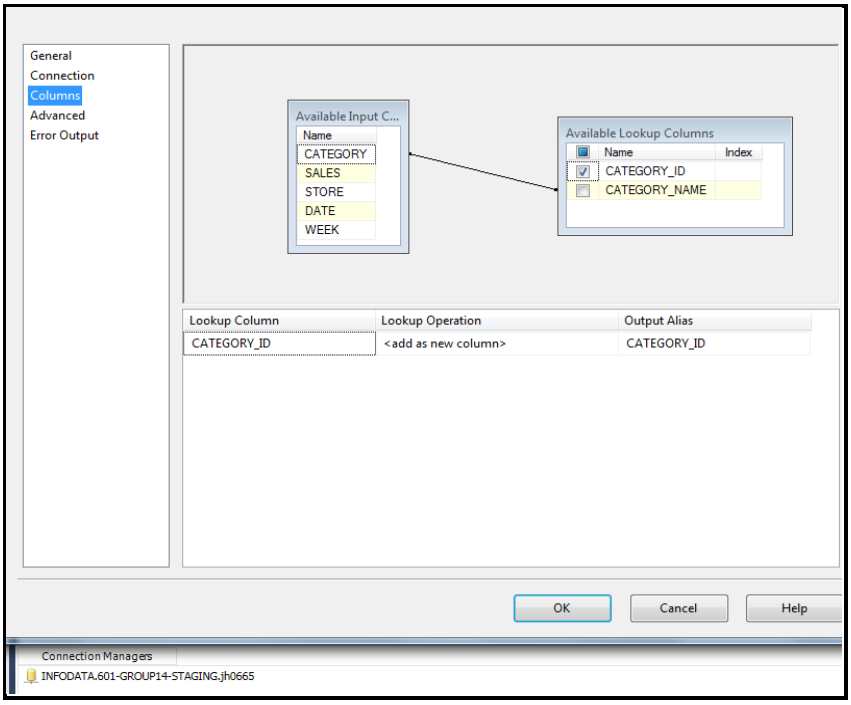
	CATEGORY_ID	CATEGORY_NAME
1	1	BEER
2	2	DAIRY
3	3	GROCERY
4	4	WINE

3.1.5. Category Fact Table Creation

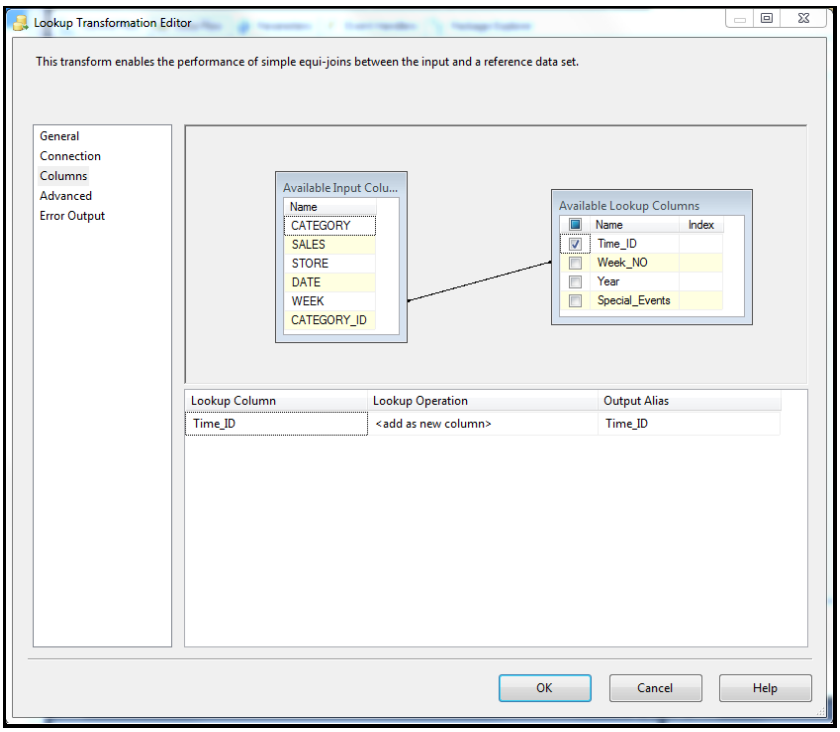
Snapshot of Category Fact Table creation from CCOUNT Source



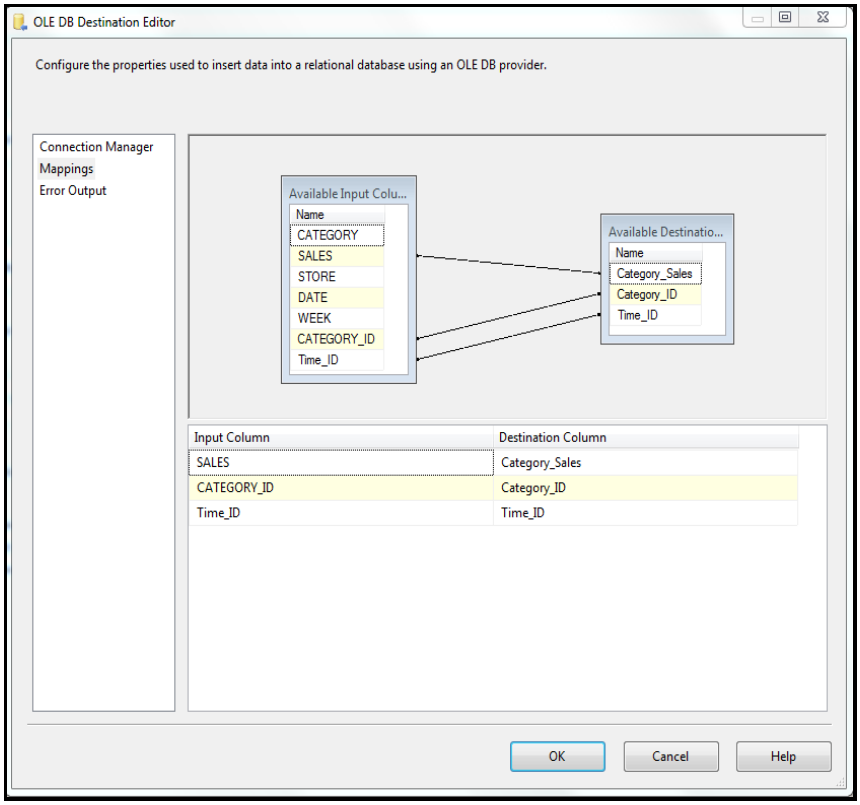
Snapshot of Category Lookup



Snapshot of Time Lookup



Screen shot of Final Mapping in to Desitnation Table



Snapshot of Data in Category Fact table

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT * FROM [601-GROUP14-STAGING].[dbo].[CATEGORYSALES_FACT]
```

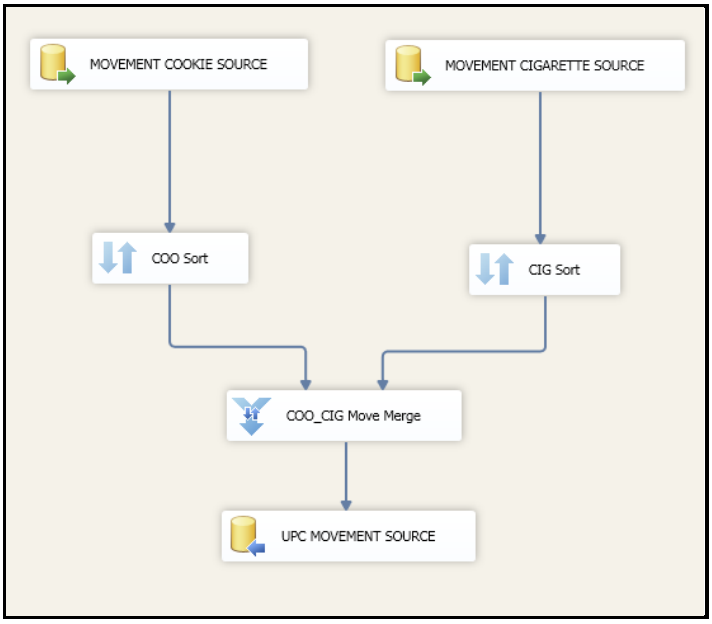
00 %

Results | Messages

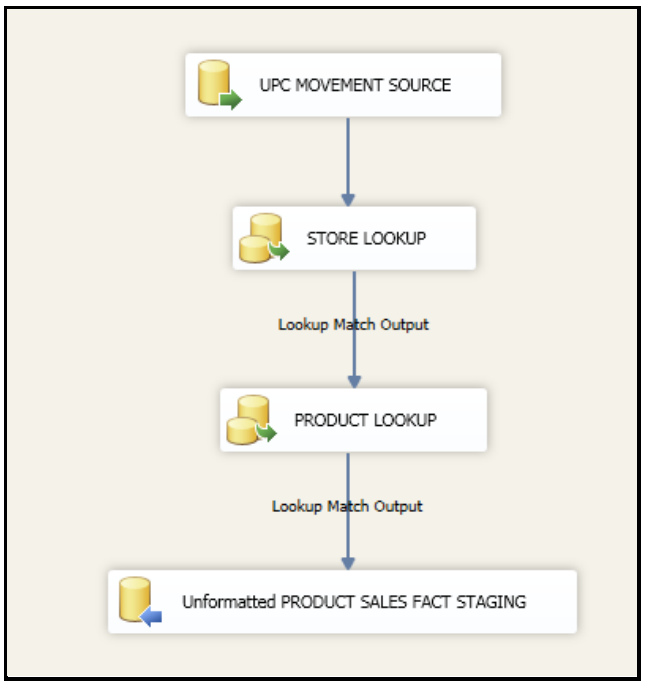
	Category_Sales	Category_ID	Time_ID
1	19138.86	3	141
2	836.35	4	141
3	825.46	1	142
4	4646.98	2	142
5	20479.9	3	142
6	1053.95	4	142
7	1031.36	1	142
8	5057.5	2	142
9	22361.35	3	142
10	1449.54	4	142
11	1250.43	1	142
12	6524.87	2	142
13	27018.04	3	142
14	1415.73	4	142

3.1.6. Product Fact Table Creation

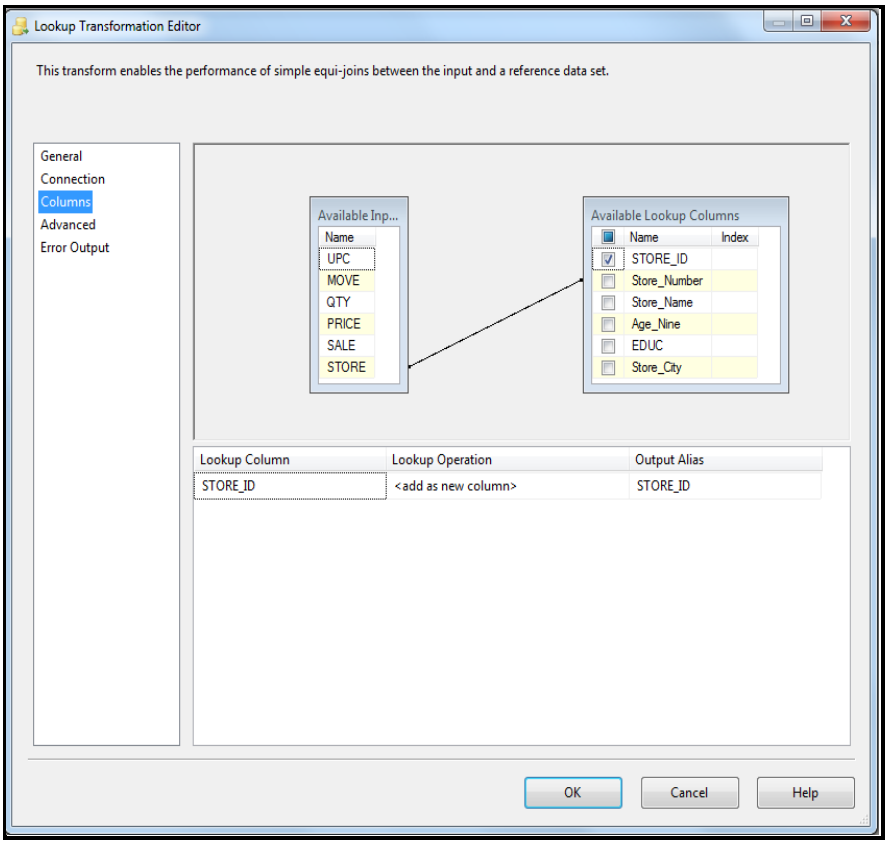
Merge of Movement Source Files



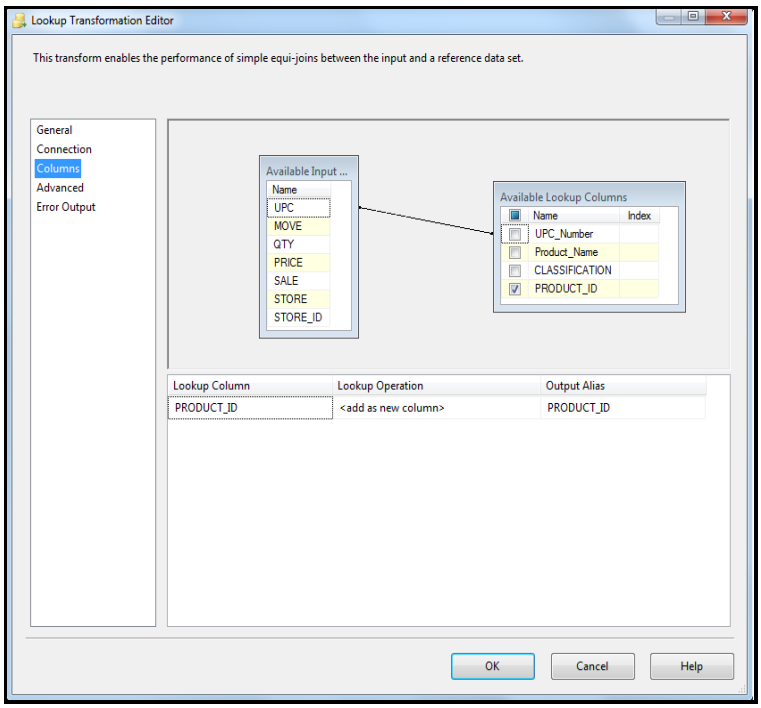
Snapshot of Product Fact Table Creation from Movement Source Files



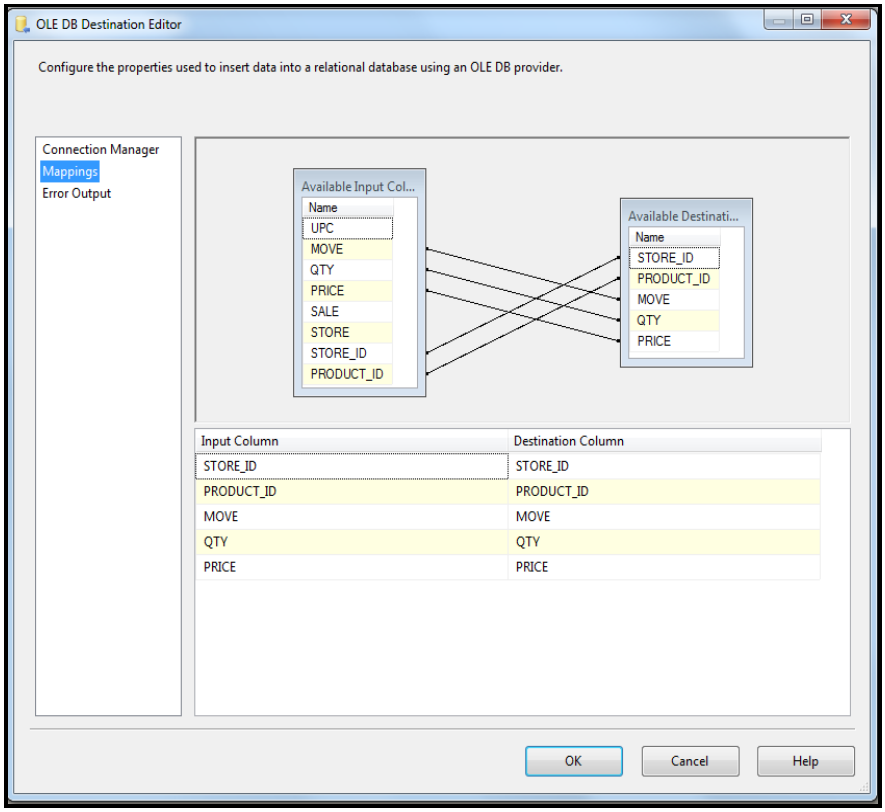
Snapshot of Store lookup



Snapshot of Product Lookup



Snapshot of Final Mapping



Section 4: Loading the dimension and fact tables from Staging area to DW area

4.1. SQL Queries to create tables in DW area

Following are the SQL statements to create Staging Area and Data Warehouse Staging Database

Category Dimension

```
CREATE TABLE [601-Group14-CategorySales-DWArea].[dbo].[DCategory] (
   [Category_ID] bigint NOT NULL,
   [Category_Desc] nvarchar(255) NULL,
) ;
```

Product Dimension

```
CREATE TABLE [601-Group14_ProductSales-DWArea].[dbo].[DProduct] (
(
  [Product_ID] bigint NOT NULL,
  [Product_Name] varchar(50) NULL,
  [UPC_Number] varchar(50) NULL,
  [Classification] nvarchar(10) NULL,
  PRIMARY KEY ([Product_ID])
);
```

Store Dimension

```
CREATE TABLE [601-Group14_ProductSales-DWArea].[dbo].[DStore]
(
[Store_ID] bigint       NOT NULL,
        [Store_Number] varchar(50) NULL,
        [Store_Name] varchar(50) NULL,
        [Store_City] varchar(50) NULL,
        [Age_Nine] varchar(50) NULL,
        [EDUC] varchar(50) NULL,
 PRIMARY KEY ([Store_ID])
);
```

Time Dimension

```
CREATE TABLE [601-Group14-CategorySales-DWArea].[dbo].[DTime](
[Time_ID] bigint NOT NULL,
[Year] nvarchar(4) NULL,
[Week_Number] varchar(50) NULL,
[Special_Event] varchar(50) NULL,
PRIMARY KEY ([Time_ID] )
) ;
```

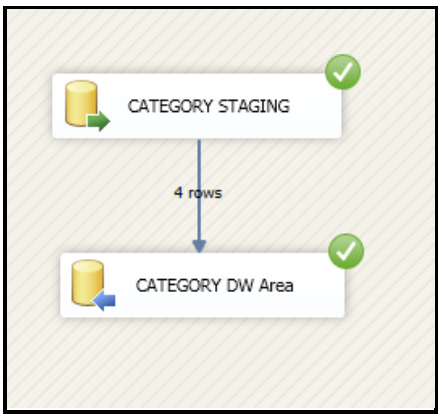
Category Sales Fact table Source

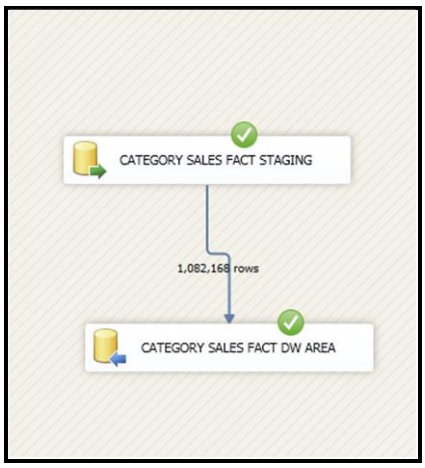
```
CREATE TABLE [601-Group14-CategorySales-DWArea].[dbo].[Category_Sales_Fact]
(
  [FactCategorySales_ID] int IDENTITY(1,1) NOT NULL,
  [Category_ID] bigint NOT NULL,
  [Time_ID] bigint NOT NULL,
  [Category_Sales] varchar(50) NULL,
  PRIMARY KEY ([FactCategorySales_ID]),
  FOREIGN KEY([Category_ID]) REFERENCES DCategory([Category_ID]),
  FOREIGN KEY([Time_ID]) REFERENCES DTime([Time_ID])
);
```

Product Sales Fact table Source

```
CREATE TABLE [601-Group14_ProductSales-DWArea].[dbo].[Product_Sales_Fact]
(
[FactProductSales_ID] int IDENTITY(1,1) NOT NULL,
[Store_ID] bigint NOT NULL,
        [Product_ID] bigint NOT NULL,
        [Unit_Price] float NULL,
        [Quantity] float NULL,
        [Units_Sold] float NULL,
        [Product_Sales] float NULL,
 PRIMARY KEY ([FactProductSales_ID]),
FOREIGN KEY([Store_ID]) REFERENCES DStore([Store_ID]),
FOREIGN KEY([Product_ID]) REFERENCES DProduct([Product_ID])
);
```

Loading of Category Sales Fact Data Mart from Staging area





Loading of Category Sales Fact Data Mart from Staging area

Table structure for Category Sales Data mart

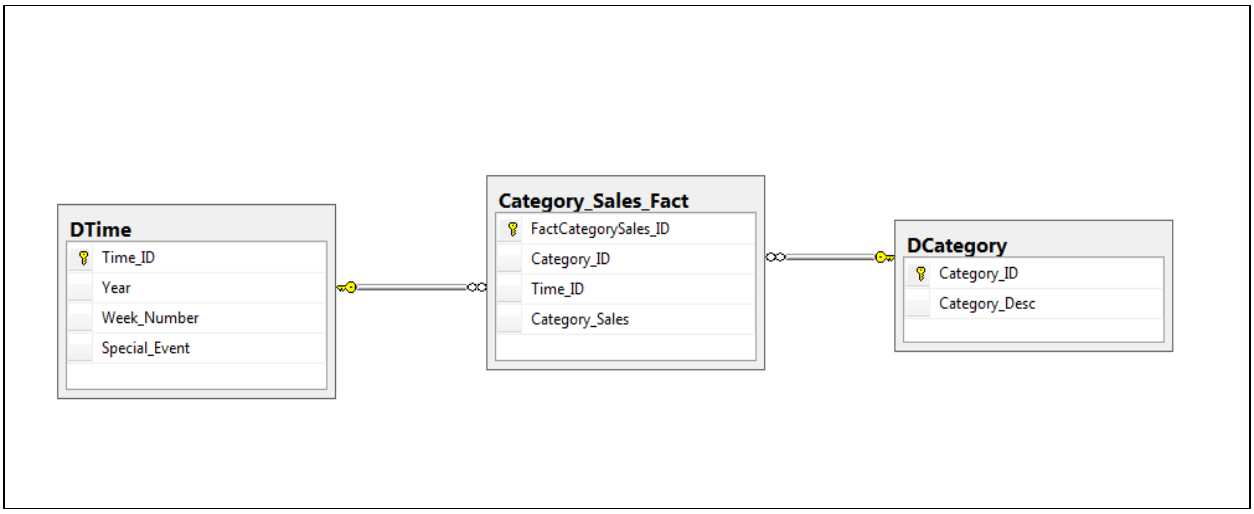
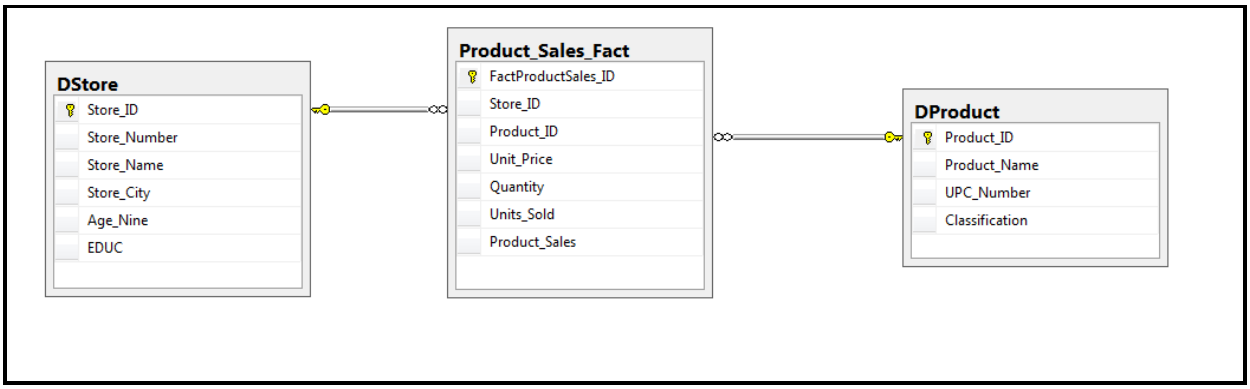
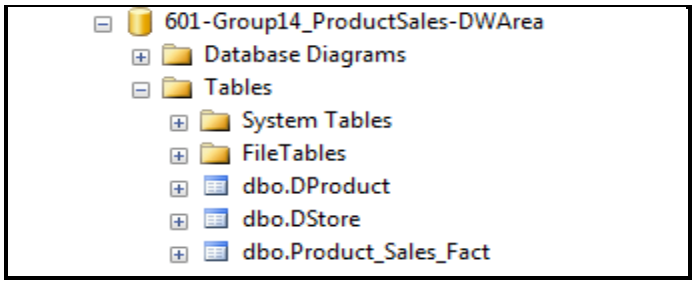


Table structure for Category Sales Data mart



4.2. Snapshot of the Product Sales Data Warehouse area



Snapshots of the data in the dimension tables of the 601Group14_ProductSales_DW_area

```
SELECT * FROM [601-Group14_ProductSales-DWArea].[dbo].[DProduct]
```

100 %

Results | Messages

	Product_ID	Product_Name	UPC_Number	Classification
453	453	LU SPICE CRISPS	1560008362	COOKIES
454	454	MINI MARIE LU	1560008363	COOKIES
455	455	$LU PIMS RASPBERRY	1560008370	COOKIES
456	456	LU LETRUFFE-PRALINE	1560008372	COOKIES
457	457	SAL LU CHIPS CHOC	1560008973	COOKIES
458	458	CADBURY MKCHOC COV F	1899600005	COOKIES
459	459	CADBURY DKCHOC COV F	1899600006	COOKIES
460	460	SUB GENERIC SINGLE	2040000000	CIGARETTES
461	461	SUB GENERIC SINGLE	2040000001	CIGARETTES
462	462	SUB GENERIC SINGLE P	2040000002	CIGARETTES
463	463	SUB GENERIC SINGLE P	2040000003	CIGARETTES
464	464	SUB GENERIC SINGLE P	2040000006	CIGARETTES
465	465	SUB GENERIC SINGLE P	2040000007	CIGARETTES
466	466	CIGARETTES (SINGLE P	2040000009	CIGARETTES
467	467	SUB GENERIC SINGLE P	2040000022	CIGARETTES
468	468	SUB GENERIC SINGLE P	2040000023	CIGARETTES
469	469	SUB GENERIC SINGLE P	2040000034	CIGARETTES
470	470	SUB GENERIC 100'S CA	2040000035	CIGARETTES

```
SELECT * FROM [601-Group14_ProductSales-DWArea].[dbo].[DStore]
```

100 %

Results | Messages

	Store_ID	Store_Number	Store_Name	Store_City	Age_Nine	EDUC
1	1	2	DOMINICKS 2	River Forest	0.117508576	0.2489349342
2	2	5	DOMINICKS 5	Palatine	0.1414334827	0.3212257298
3	3	8	DOMINICKS 8	Oak Lawn	0.123155416	0.0951732743
4	4	9	DOMINICKS 9	Morton Grove	0.1035030974	0.2221723183
5	5	12	DOMINICKS 12	Chicago	0.1056967397	0.2534129693
6	6	14	DOMINICKS 14	Glenview	0.129589372	0.3482930237
7	7	18	DOMINICKS 18	River Grove	0.1100949839	0.0722464558
8	8	21	DOMINICKS 21	Hanover Park	0.1759263459	0.1775034504
9	9	28	DOMINICKS 28	Mt. Prospect	0.1288795371	0.233162564
10	10	32	DOMINICKS 32	Park Ridge	0.0990606319	0.1982598608
11	11	33	DOMINICKS 33	Chicago	0.0460709172	0.4196880043
12	12	40	DOMINICKS 40	Bridgeview	0.1336846485	0.0721286047
13	13	44	DOMINICKS 44	Western Spring	0.1448834853	0.3297383876
14	14	45	DOMINICKS 45	Wheeling	0.1467187625	0.2801501642

```
SELECT * FROM [601-Group14_ProductSales-DWArea].[dbo].[Product_Sales_Fact]
```

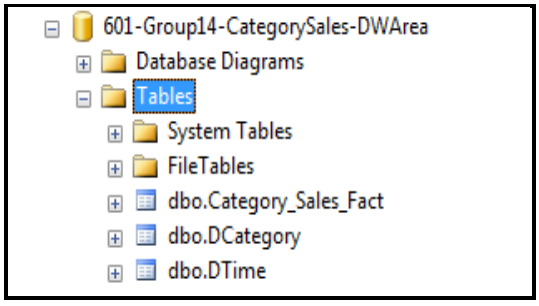
100 %

Results | Messages

	FactProductSales_ID	Store_ID	Product_ID	Unit_Price	Quantity	Units_Sold	Product_Sales
8...	8494	17	419	0	1	0	0
8...	8495	54	419	2.29	1	5	11.45
8...	8496	68	419	2.29	1	2	4.58
8...	8497	46	419	2.03	1	8	16.24
8...	8498	36	419	1.99	1	6	11.94
8...	8499	17	419	0	1	0	0
8...	8500	54	419	2.29	1	4	9.16
8...	8501	68	419	2.29	1	4	9.16
8...	8502	46	419	1.79	1	10	17.9
8...	8503	36	419	2.09	1	6	12.54
8...	8504	17	419	0	1	0	0
8...	8505	54	419	2.09	1	7	14.63
8...	8506	68	419	0	1	0	0
8...	8507	46	419	2.03	1	16	32.48
8...	8508	36	419	1.97	1	3	5.91
8...	8509	17	419	0	1	0	0
8...	8510	54	419	2.29	1	9	20.61
8...	8511	68	419	2.29	1	8	18.32

4.3. Snapshot of the Category Sales Data Warehouse area

Snapshots of the dimension tables of the 601Group14_CategorySales_DW_area



Snapshots of the data in the dimension tables of the 601Group14_CategorySales_DW_area

```
select * from [601-Group14-CategorySales-DWArea].[dbo].[DCategory]
```

100 %

Results | Messages

	Category_ID	Category_Desc
1	1	BEER
2	2	DAIRY
3	3	GROCERY
4	4	WINE

```
select * from [601-Group14-CategorySales-DWArea].[dbo].[DTime]
```

100 %

Results | Messages

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

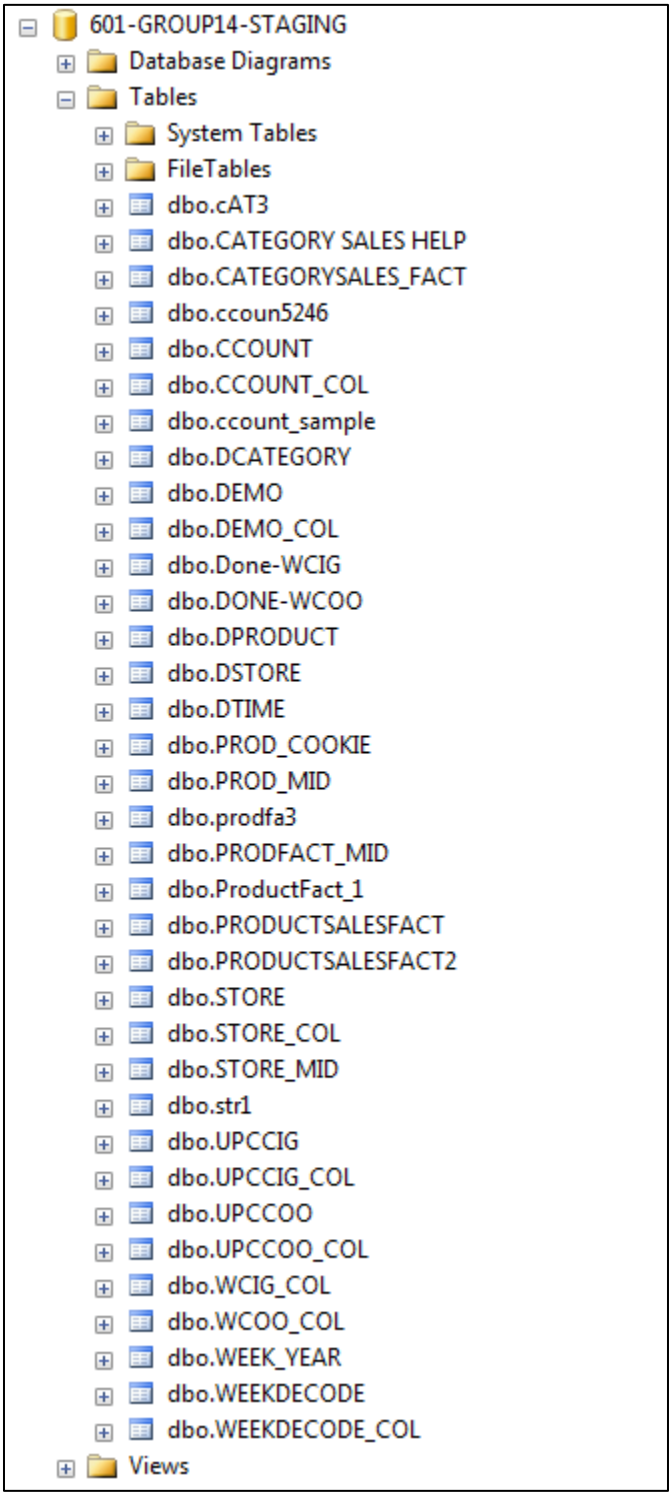
```
select * from [601-Group14-CategorySales-DWArea].[dbo].[Category_Sales_Fact]
```

100 %

Results | Messages

	FactCategorySales_ID	Category_ID	Time_ID	Category_Sales
1	1	3	141	19138.86
2	2	4	141	836.35
3	3	1	142	825.46
4	4	2	142	4646.98
5	5	3	142	20479.9
6	6	4	142	1053.95
7	7	1	142	1031.36
8	8	2	142	5057.5
9	9	3	142	22361.35
10	10	4	142	1449.54
11	11	1	142	1250.43
12	12	2	142	6524.87
13	13	3	142	27018.04
14	14	4	142	1415.73

4.4. Removal of temporary tables that exists in the data staging area



The tables existing before data transformation process which contained dirty data were removed. The above snapshot lists the tables which were retained in the staging area after the transformation process and required for the loading process.

BI Reports Design and Implementation

Section 1: BI Reporting Using SSAS, SSRS and ReportBuilder 3.0

1.1. Reporting plan

After ETL, it is absolutely necessary to implement reporting, as ultimately it is the visualizations that will be useful and helpful to the users. Thus, in order to make sense of the data and to implement strategic business decisions accordingly, reporting is done.

In designing and implementing our Business Intelligence (BI) reports, we used **SQL Server Analysis Services (SSAS), SQL Reporting Services (SSRS)** and **ReportBuilder 3.0** as end user tools. These were implemented in the following way: SSRS, SSAS, SSRS with ReportBuilder and SSRS on top of SSAS. A breakdown of our reporting plan is shown below:

Reporting Tool	Business Question Number
SSRS	1
SSRS with ReportBuilder	2, 5
SSAS	3
SSRS on top of SSAS	4

1.2 Determining the target reports that satisfy business questions.

Question 1 - What impact do holidays have over sale of beer and wine and which holidays

Report generated from SSRS alone

To solve this business question, we used the category name as BEER and WINE from the Category dimension, week number in ascending order and all the special events from the Time dimension and mapped them with the Category Sales fact table to get the Category sales. The sales from the fact table will be the trend of Beer and Wine sales during the holidays for the entire duration. We have used SQL query and SSRS to generate the report for this business question. We have made use of bar chart to give a better visualization to answer this question.

Question 2 - What is the trend in wine sales over the years during Valentine's week?

Report generated using SSRS and ReportBuilder

In order to analyze the trend in wine sales during Valentine's week over the years, we selected the category WINE from the Category dimension and the week numbers corresponding to Valentine's week over the different years from the Time dimension and then mapped these two to the Category Sales fact table. The total sales over the years provided by the Category Sales fact table will give insights to the sales trend over the specified period. We used SSRS to implement the detailed report and we also implemented with ReportBuilder for better visualization, so as to readily and easily spot the sales trend.

Question 3 - How do the total sales of Grocery and Dairy vary over years? What happened in 1997?

Report generated from Cube Analysis created from SSAS only

We used SSAS only to answer this business question. We deployed a cube which consists of the Category sales fact table, Store dimension, Category dimension and Time dimension tables. The cube groups the Category sales by the Category name “Grocery” and “Dairy” the SQL query gives the results of the sales of grocery and dairy trending over the period. In order to analyze the results, we used the export to Excel tool to create a pivot table and implemented a pivot chart (line chart) for the visualization of the trend.

Question 4 - What is the trend in the sale of cookies (from week 1 – 399) in stores with the highest % population of kids under age 9?

Report generated from SSRS on top of SSAS

In order to answer this business question, we created a cube in the Analytics Services of the SQL Server data tools. We used the Product Sales Data mart with Store dimension and Product Dimension along with the Product Sales fact table. We then created a SQL query to filter out the stores with highest % population of kids under age 9 from the Store dimension, Product name “Cookie” from the Product dimension and Product Sales from the Product Sales fact table. After successful deployment of the cube on server, we used SSRS on top of SSAS cube to generate a report. We used a table to visualize the sales in the selected stores.

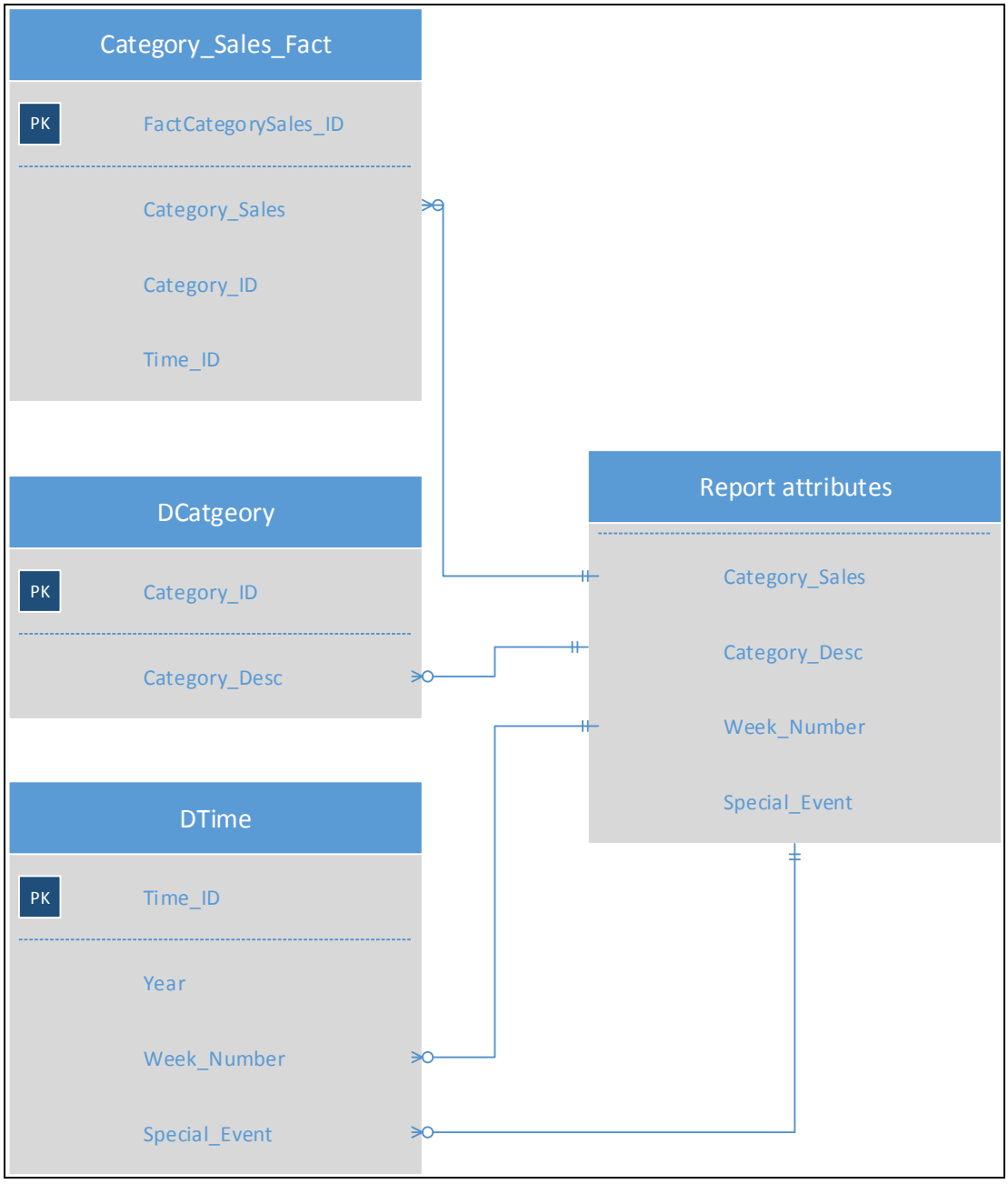
Question 5 - What is the trend in the sale of Cigarettes in stores with the highest population of College Graduates?

Report generated using SSRS and ReportBuilder

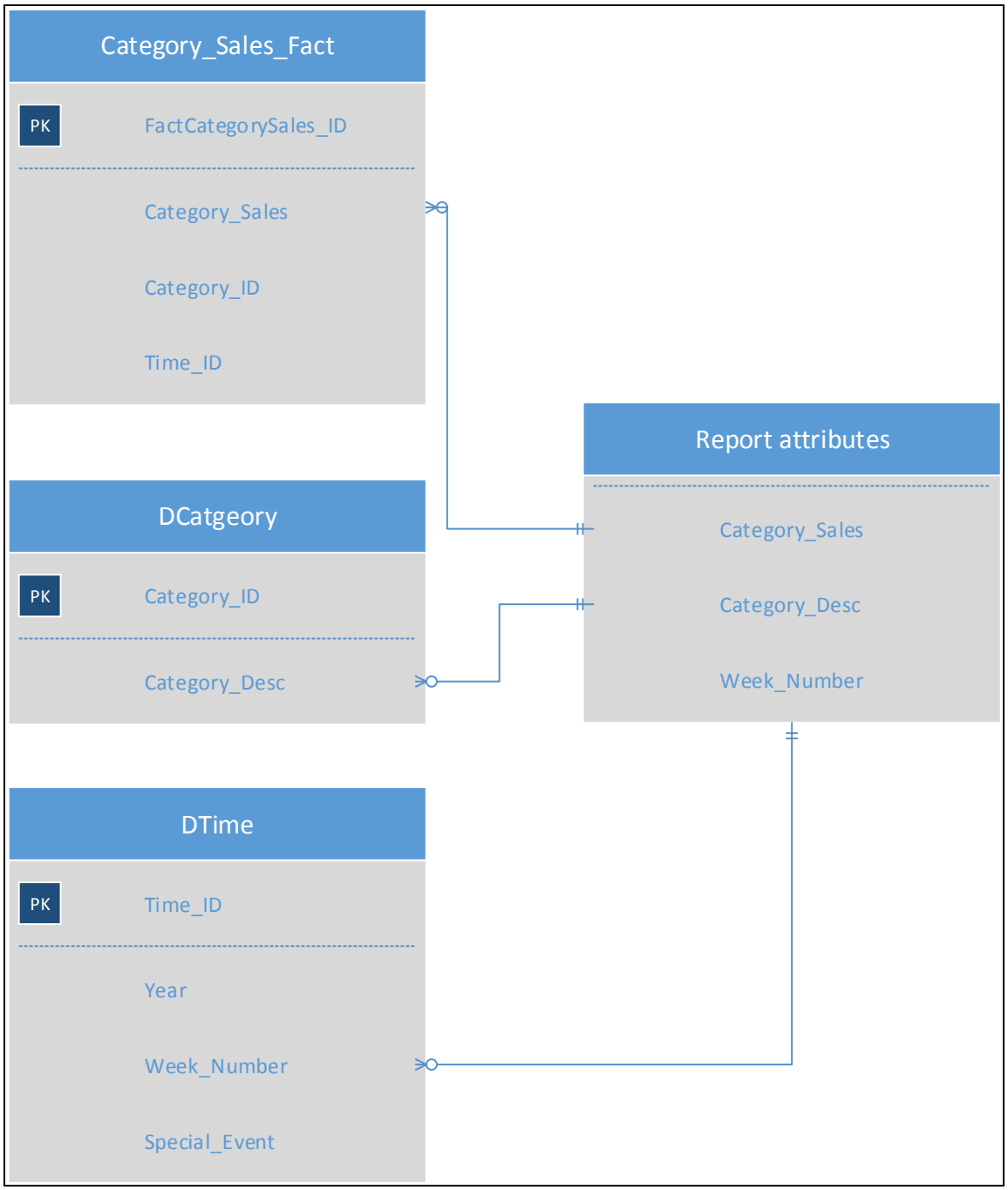
In analyzing the trend in cigarette sales in stores with the highest population of college graduates, we selected store number from the Store dimension and CIGARETTES from the Product dimension and then mapped these to the Product Sales fact table. However, in order to determine the sales for those stores with the highest population of college graduates, we applied filters to first order the results by 'EDUC' (i.e. the description for college graduates in the Store dimension) and then selected the top ten (10) stores corresponding to the 10 highest EDUC population. The total sales from the Product Sales fact table provided the relevant details as to the sales figures which we also implemented with ReportBuilder for a better visualization of the sales trend.

Section 2: Mappings from the Tables in the Data Marts to the Attributes in the Reports

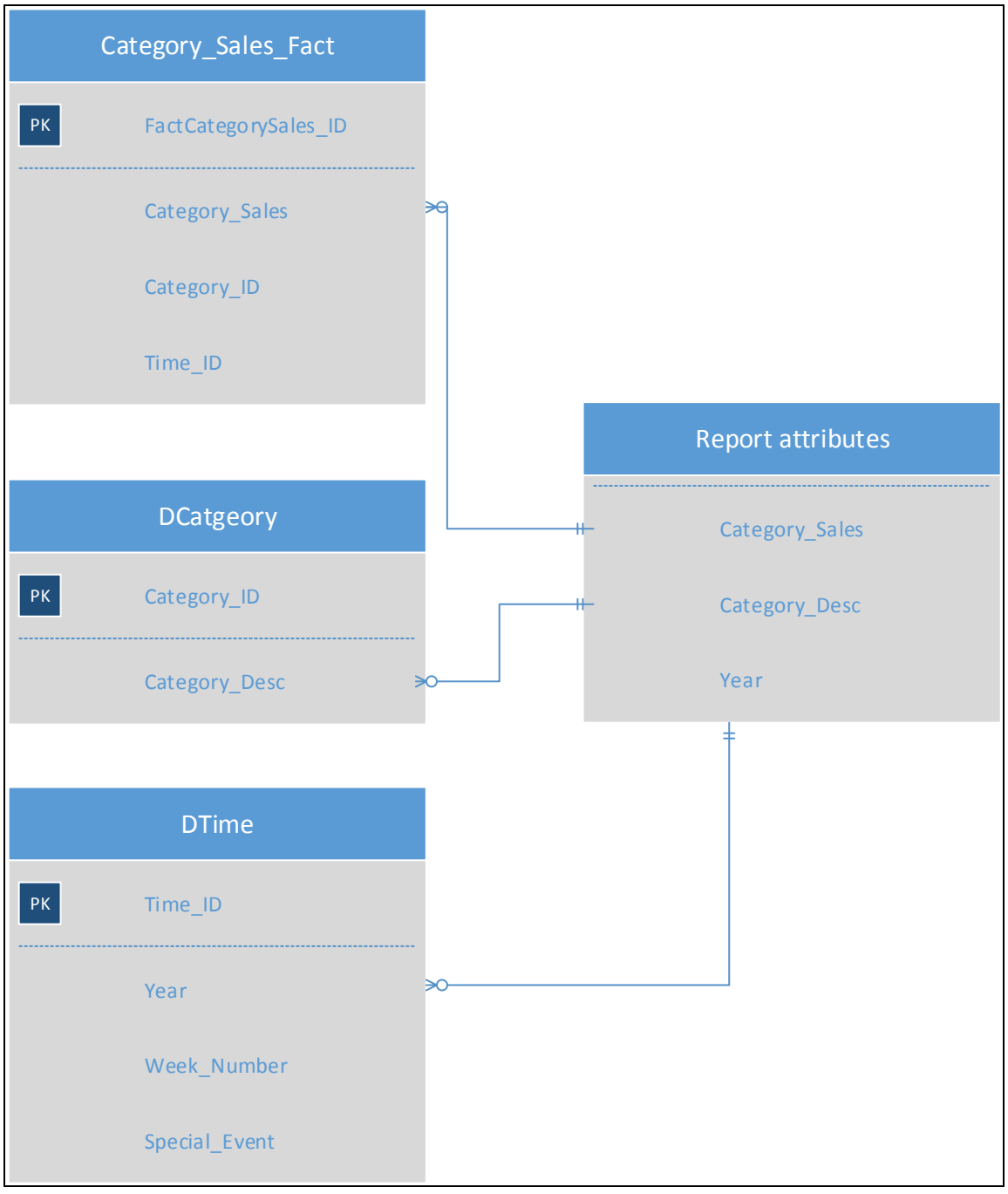
Question 1 - What impact do holidays have over sale of beer and wine and which holidays have the highest demands?



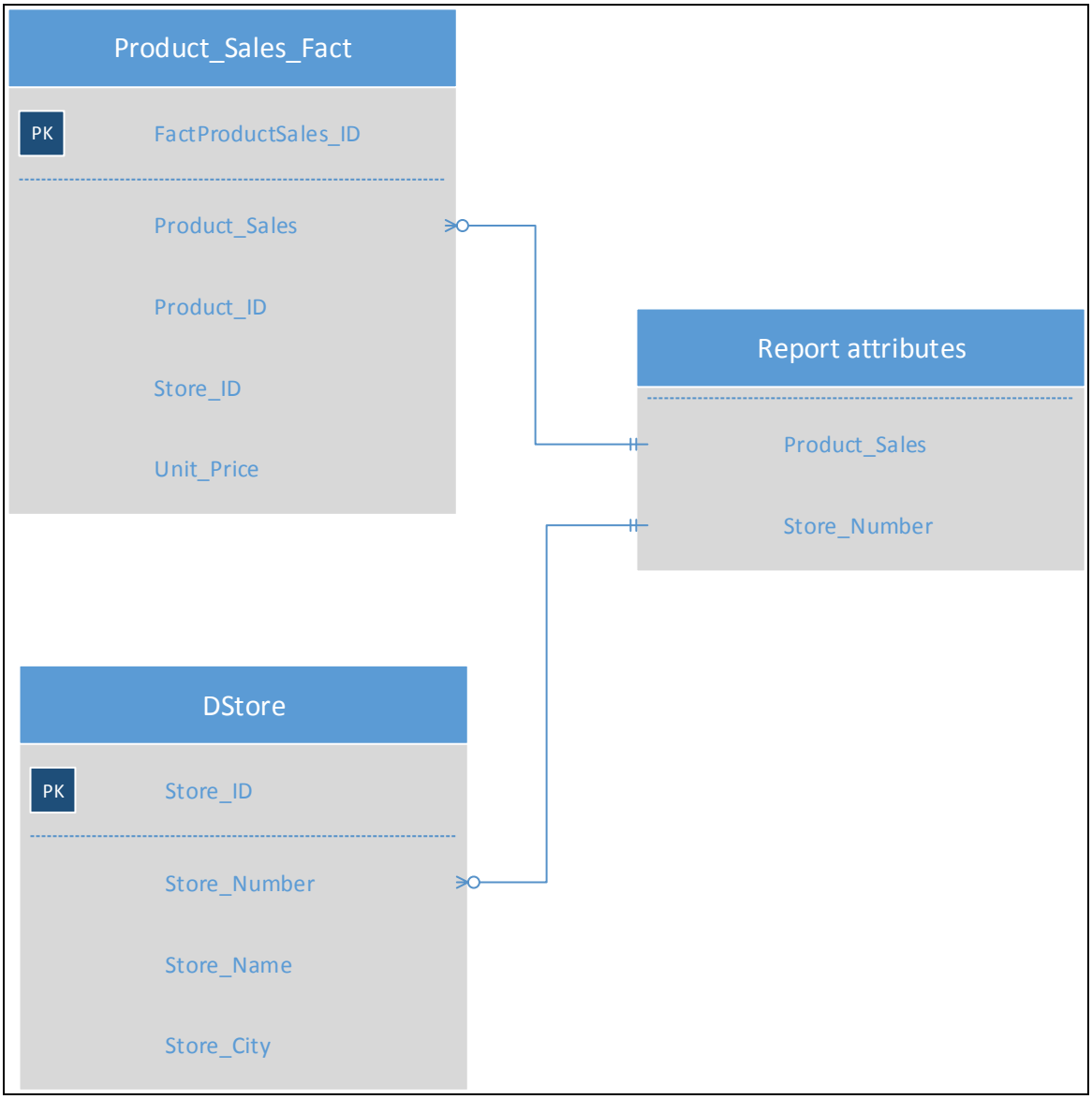
Question 2 - What is the trend in wine sales over the years during Valentine's week?



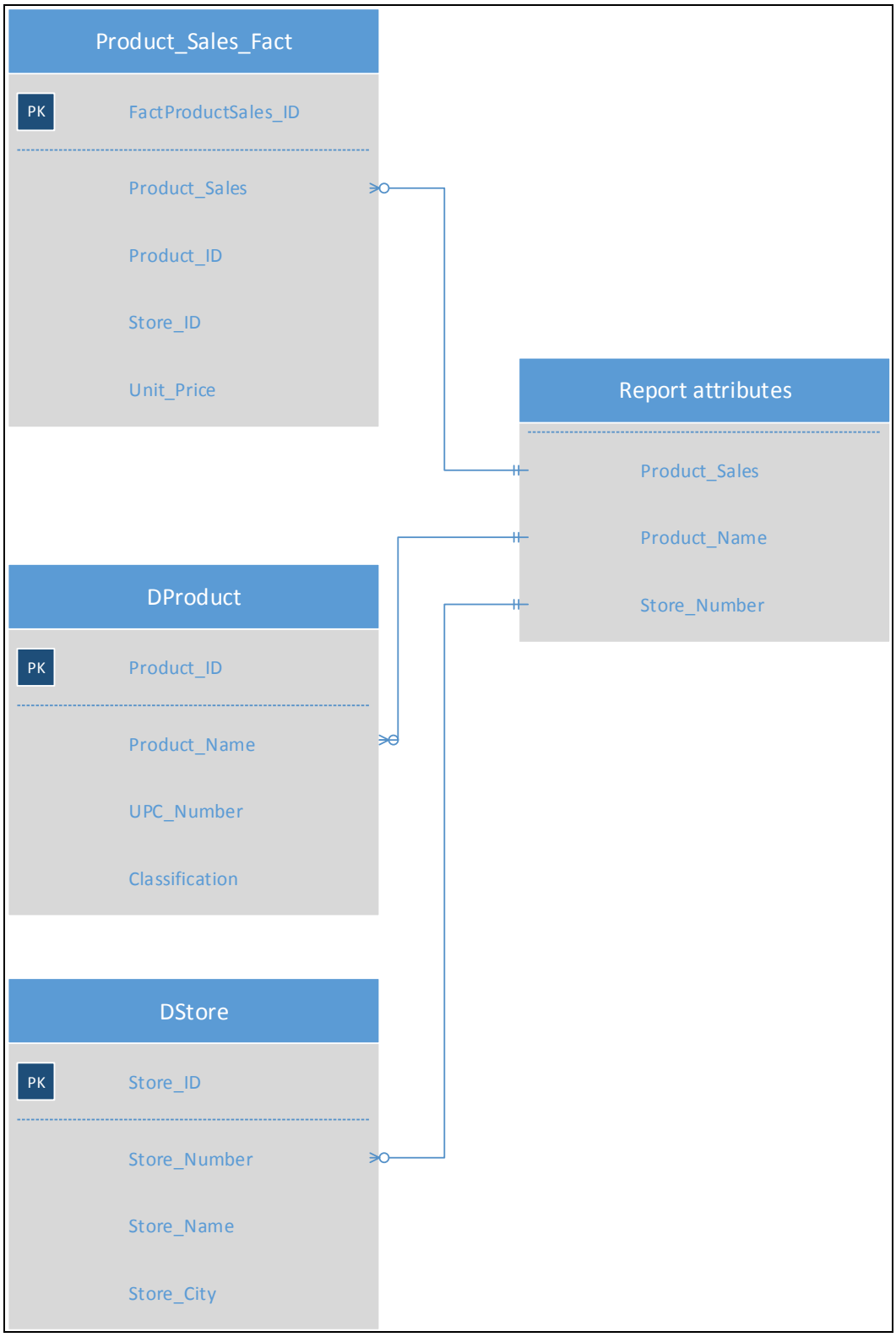
Question 3 - How do the total sales of Grocery and Dairy vary over years? What happened in 1997?



Question 4 - What is the trend in the sale of cookies (from week 1 – 399) in stores with the highest % population of kids under age 9?



Question 5 - What is the trend in the sale of Cigarettes in stores with the highest population of College Graduates?

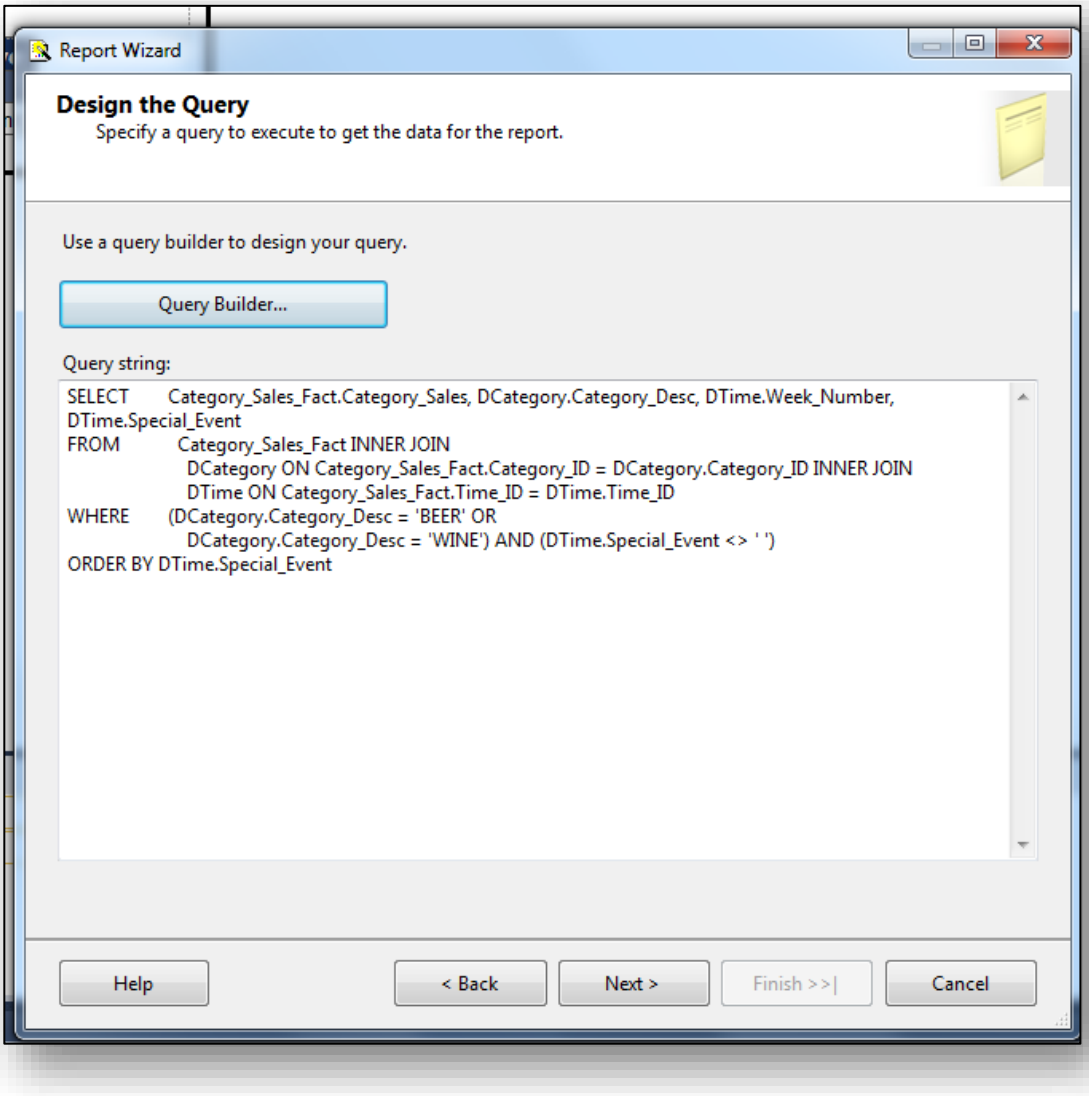


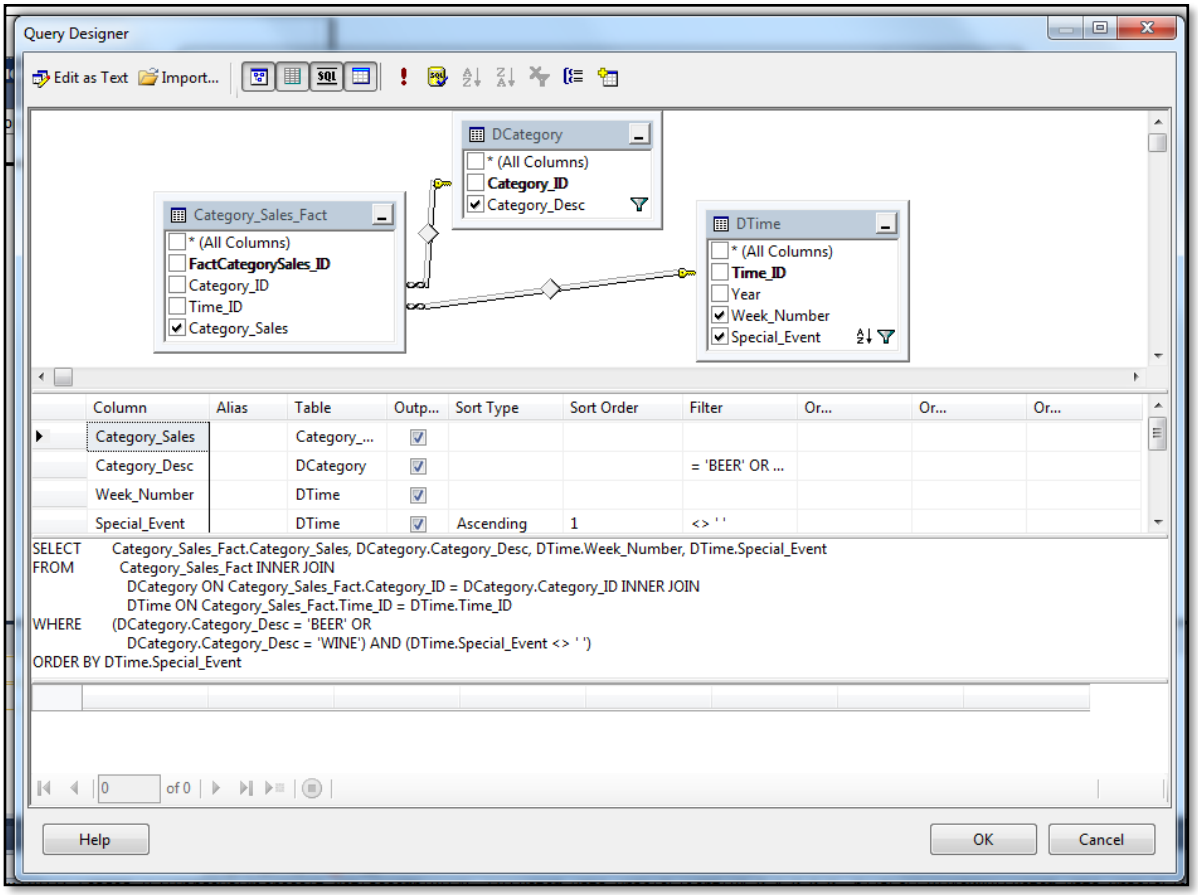
Section 3: Report Building from the Individual Data Marts

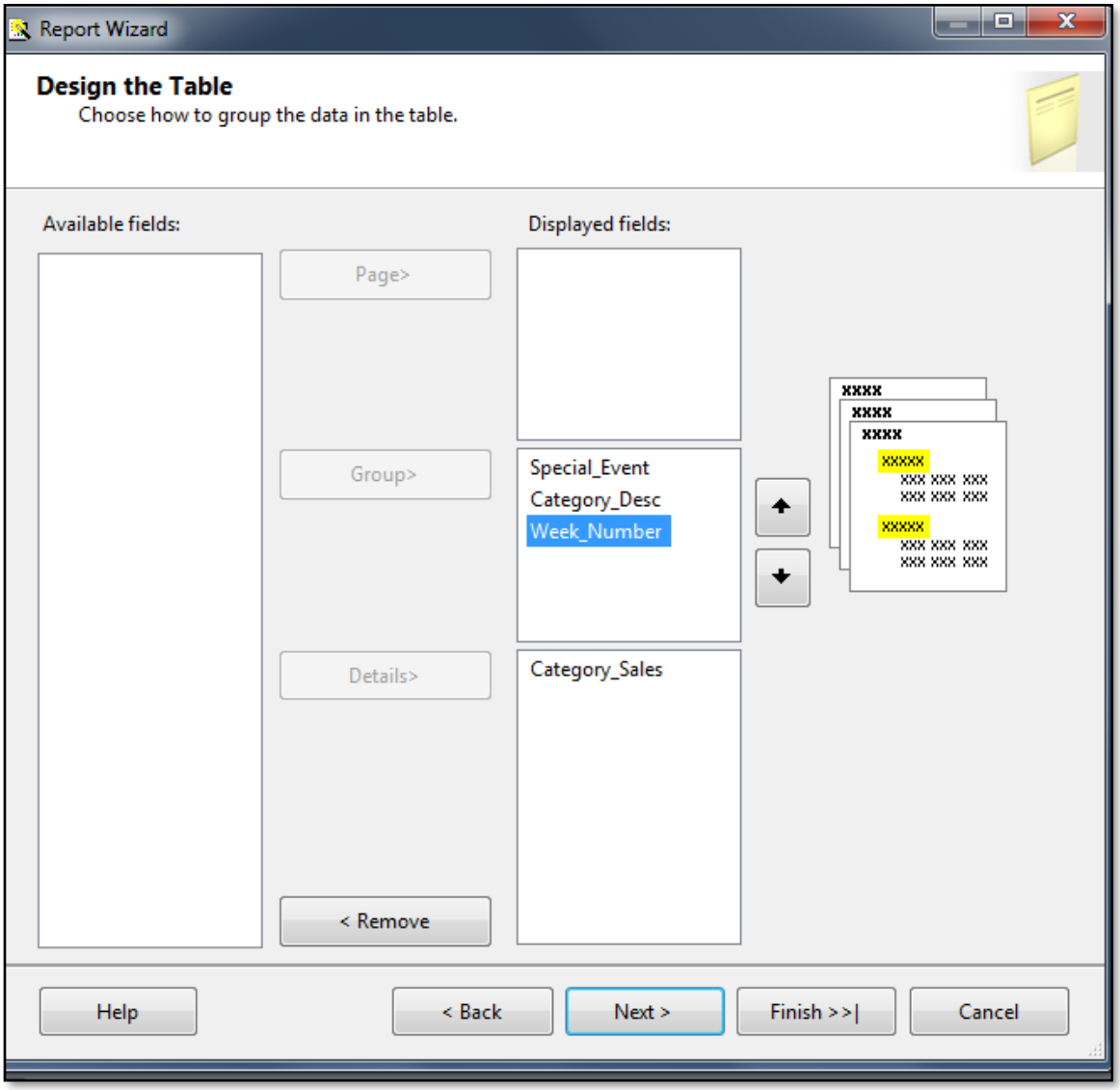
3.1 Report Building Using SSRS

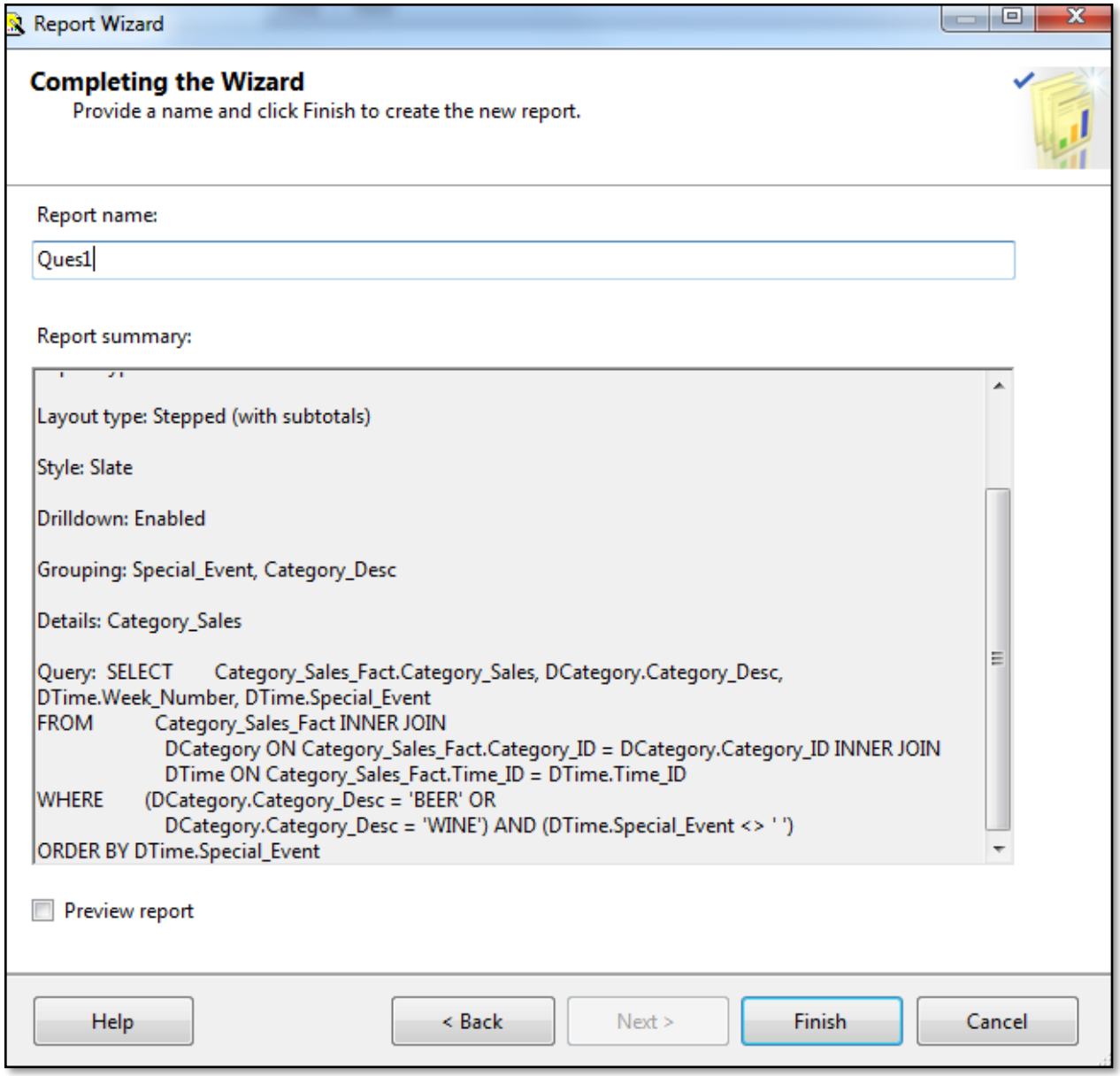
Based on the reporting plan discussed in the report earlier, we are using report building from individual data mart using SSRS tool to create visualizations for Question 1.

Question 1 - What impact do holidays have over sale of beer and wine and which holidays have the highest demands?




Query Designer
Edit as Text
Import...
DCategory
* (All Columns)
Category_ID
Category_Desc
Category_Sales_Fact
* (All Columns)
FactCategorySales_ID
Category_ID
Time_ID
Category_Sales
DTime
* (All Columns)
Time_ID
Year
Week_Number
Special_Event
Column
Alias
Table
Outp...
Sort Type
Sort Order
Filter
Or...
Or...
Or...
Category_Sales
Category_...
Category_Desc
DCategory
= 'BEER' OR ...
Week_Number
DTime
Special_Event
DTime
Ascending
1
<> ' '
SELECT Category_Sales_Fact.Category_Sales, DCategory.Category_Desc, DTime.Week_Number, DTime.Special_Event
FROM Category_Sales_Fact INNER JOIN
DCategory ON Category_Sales_Fact.Category_ID = DCategory.Category_ID INNER JOIN
DTime ON Category_Sales_Fact.Time_ID = DTime.Time_ID
WHERE (DCategory.Category_Desc = 'BEER' OR
DCategory.Category_Desc = 'WINE') AND (DTime.Special_Event <> ' ')
ORDER BY DTime.Special_Event
0
of 0
Help
OK
Cancel


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


Report Wizard
Completing the Wizard
Provide a name and click Finish to create the new report.
Report name:
Ques1
Report summary:
Layout type: Stepped (with subtotals)
Style: Slate
Drilldown: Enabled
Grouping: Special_Event, Category_Desc
Details: Category_Sales
Query: SELECT Category_Sales_Fact.Category_Sales, DCategory.Category_Desc,
DTime.Week_Number, DTime.Special_Event
FROM Category_Sales_Fact INNER JOIN
DCategory ON Category_Sales_Fact.Category_ID = DCategory.Category_ID INNER JOIN
DTime ON Category_Sales_Fact.Time_ID = DTime.Time_ID
WHERE (DCategory.Category_Desc = 'BEER' OR
DCategory.Category_Desc = 'WINE') AND (DTime.Special_Event <> ' ')
ORDER BY DTime.Special_Event
Preview report
Help
< Back
Next >
Finish
Cancel

Question 1

Special Event	Category Desc	Week Number	Category Sales
[Special_Event]			
	[Category_Desc]		
		[Week_Number]	
			[Category_Sales]
			[Sum(Category_Sale:

Question 1

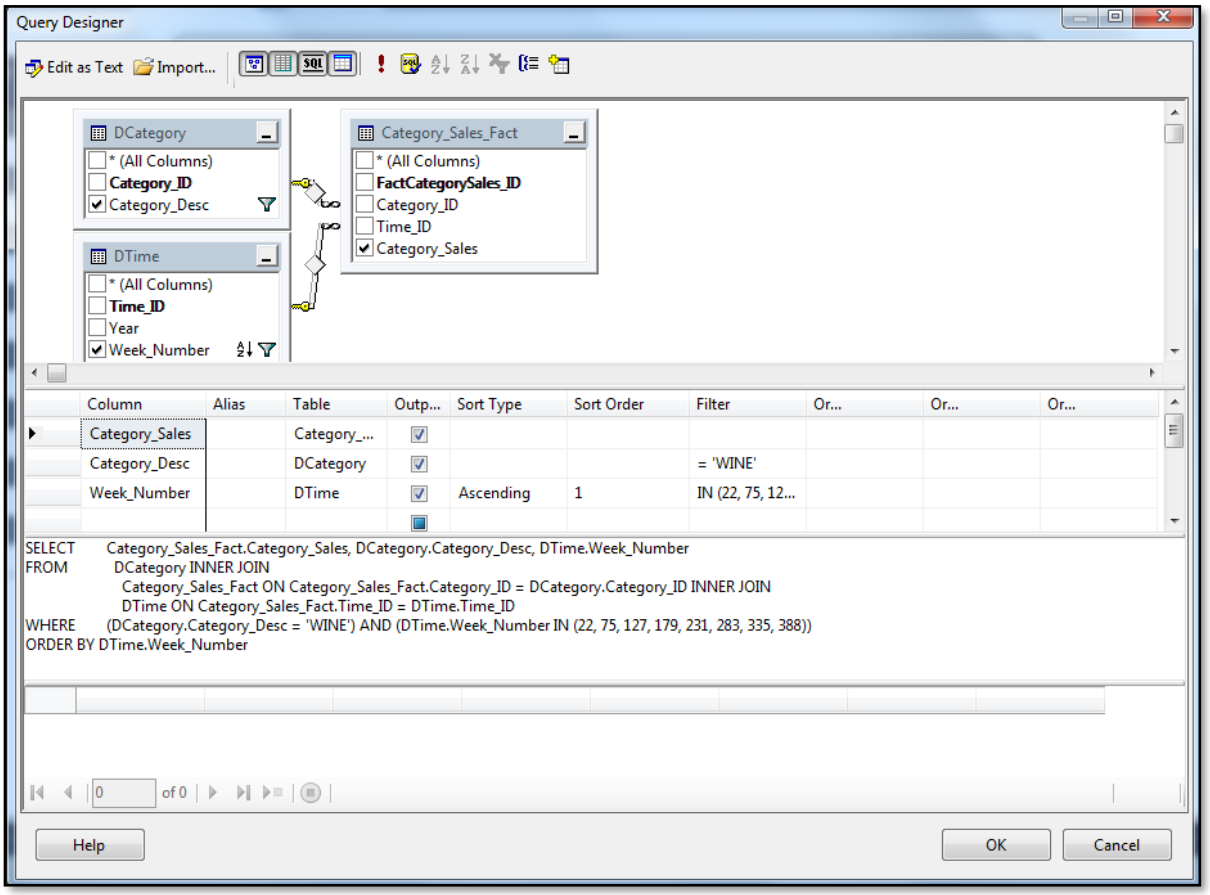
Special Event	Category Desc	Week Number	Category Sales
⊟ 4th of July			
	⊞ BEER		
	⊞ WINE		
⊟ Christmas			
	⊟ BEER		
		⊞ 119	
			765452
		⊞ 15	
			586102
		⊞ 172	
			628766
		⊞ 224	
			657391
		⊞ 276	
			802544
		⊞ 328	
			856666
		⊞ 380	
			1102815
		⊞ 67	
			856197
	⊞ WINE		
⊞ Easter			
⊞ Halloween			
⊞ Labor Day			
⊞ Memorial Day			
⊞ New-Year			
⊞ Presidents Day			
⊞ Thanksgiving			

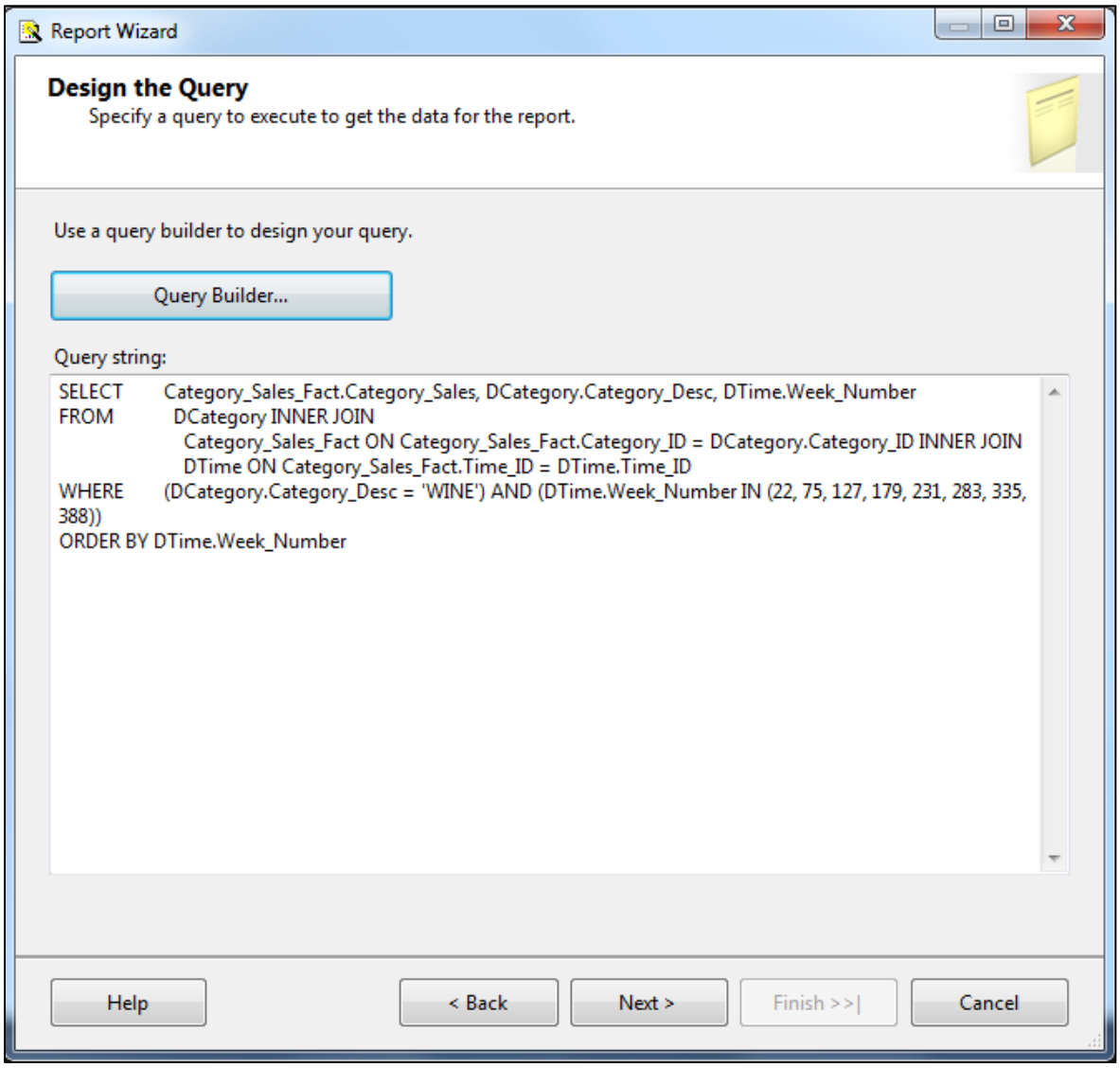
Analysis and Conclusion -

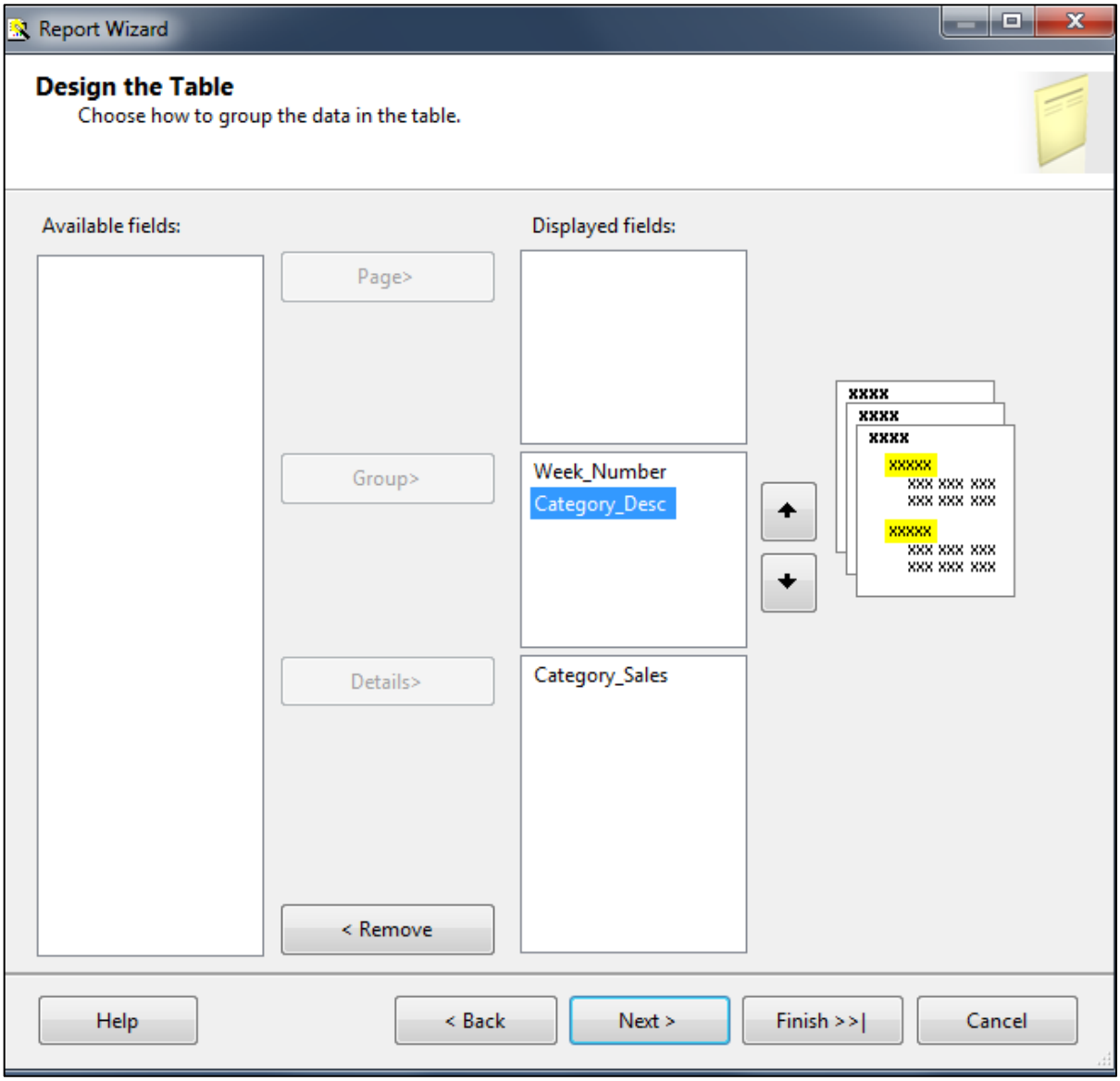
From reports generated using SSRS (as shown above), we clearly see that during holidays, the sales of wine and beer increases dramatically. We compared the sales on all the recognized holidays in the US, and from the exposure of our limited data, we found that beer sales are maximum on Memorial Day while wine sales are maximum on Christmas Day. This insight could have helped Dominick's to employ strategies such as higher stock-ups on products, promotions and discounts, so as to better deal with the high demands on these holidays and to maximize sales.

3.2 Report Building Using SSRS with ReportBuilder

Question 2 - What is the trend in wine sales over the years during Valentine's week?




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 Category_Sales_Fact.Category_Sales, DCategory.Category_Desc, DTime.Week_Number
FROM DCategory INNER JOIN
Category_Sales_Fact ON Category_Sales_Fact.Category_ID = DCategory.Category_ID INNER JOIN
DTime ON Category_Sales_Fact.Time_ID = DTime.Time_ID
WHERE (DCategory.Category_Desc = 'WINE') AND (DTime.Week_Number IN (22, 75, 127, 179, 231, 283, 335, 388))
ORDER BY DTime.Week_Number
Help
< Back
Next >
Finish >>|
Cancel


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

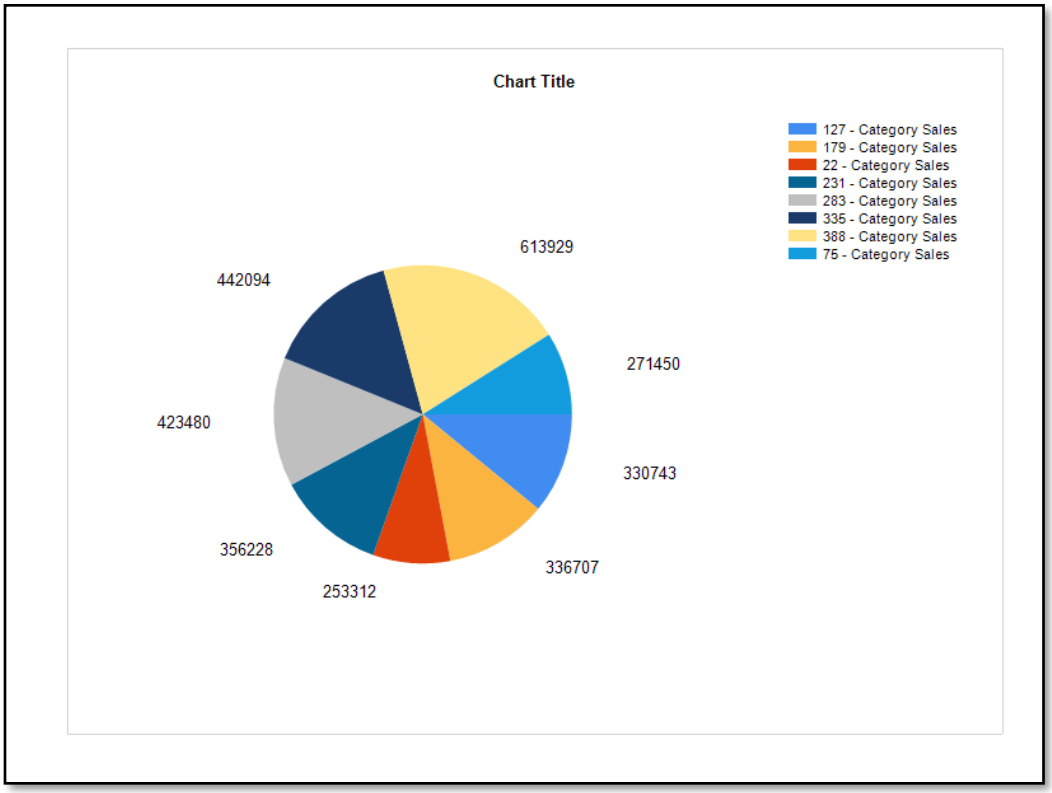


Chart Title
127 - Category Sales
179 - Category Sales
22 - Category Sales
231 - Category Sales
283 - Category Sales
335 - Category Sales
388 - Category Sales
75 - Category Sales
613929
442094
271450
423480
330743
356228
336707
253312

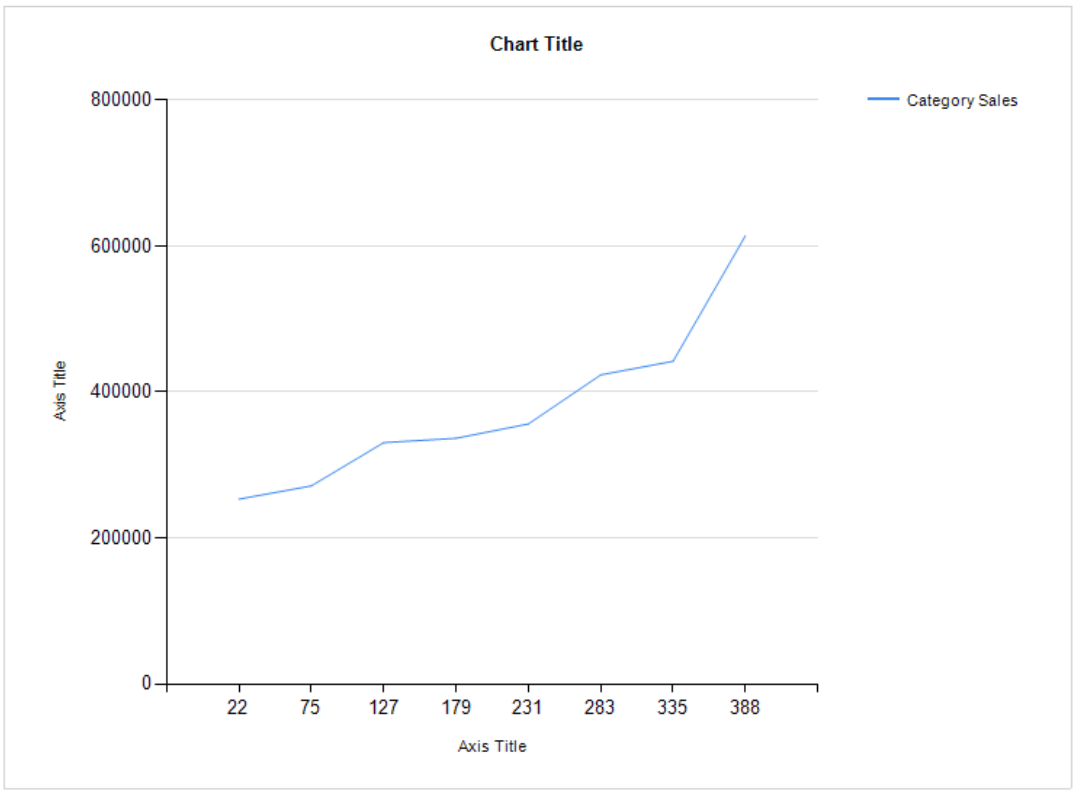


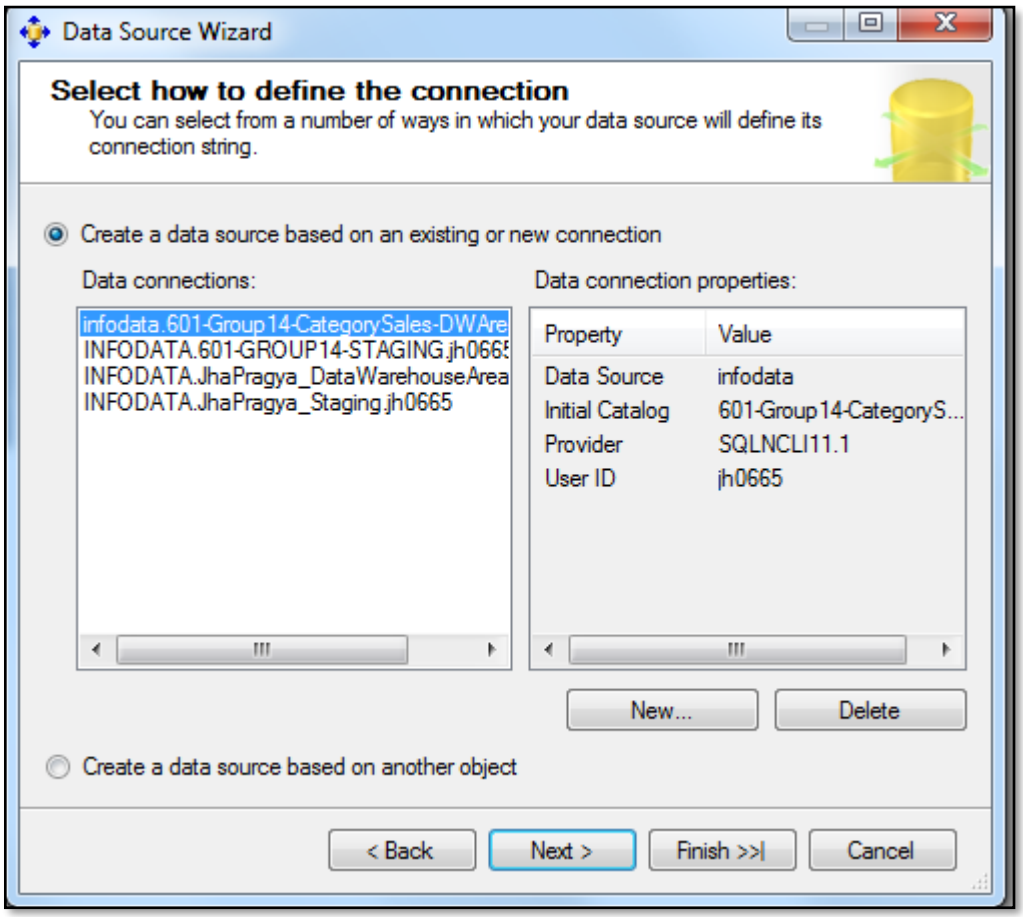
Chart Title
Category Sales
800000
600000
400000
200000
0
Axis Title
22
75
127
179
231
283
335
388
Axis Title

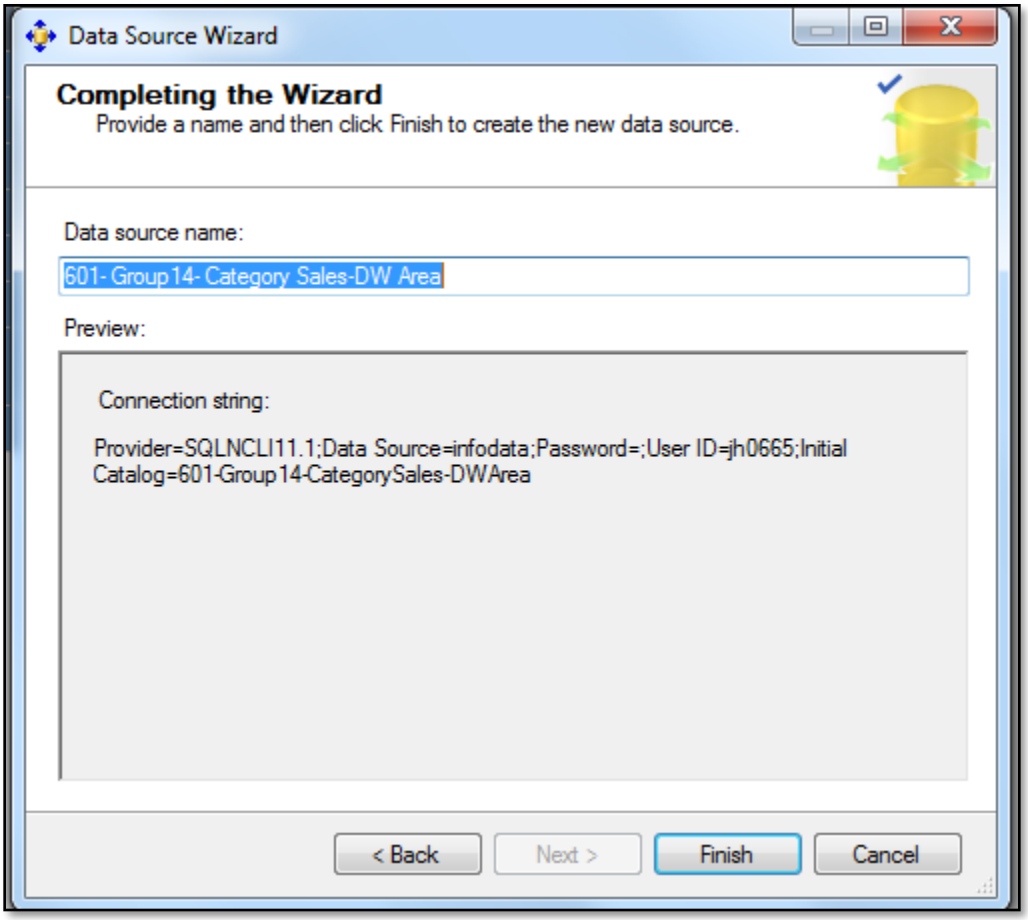
Analysis and Conclusion -

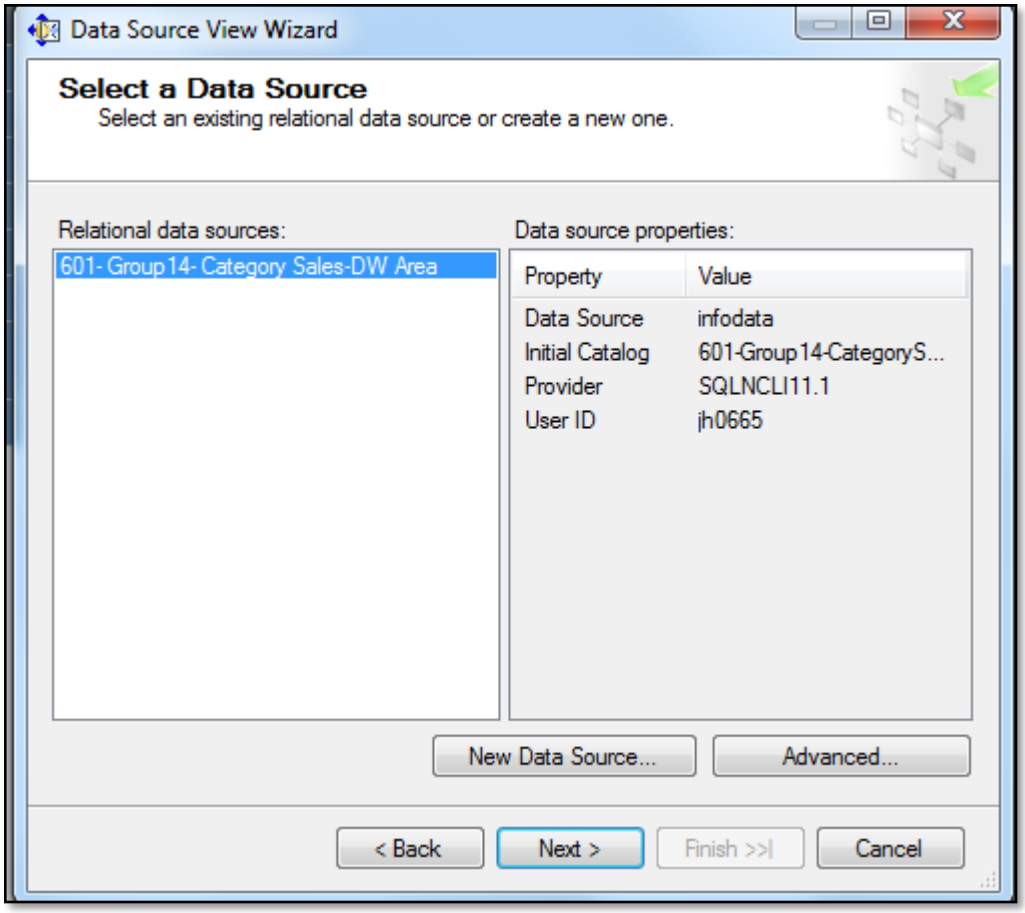
From the ReportBuilder chart and graph, we clearly see an increasing trend in wine sales during Valentine's week over the time period. This information could be used to compare the wine sales during other weeks of the year and determine which periods have the lowest sales and declining trends. DFF could then leverage this business insight to find appropriate strategies that can be implemented so as to equally increase the wine sales all year round.

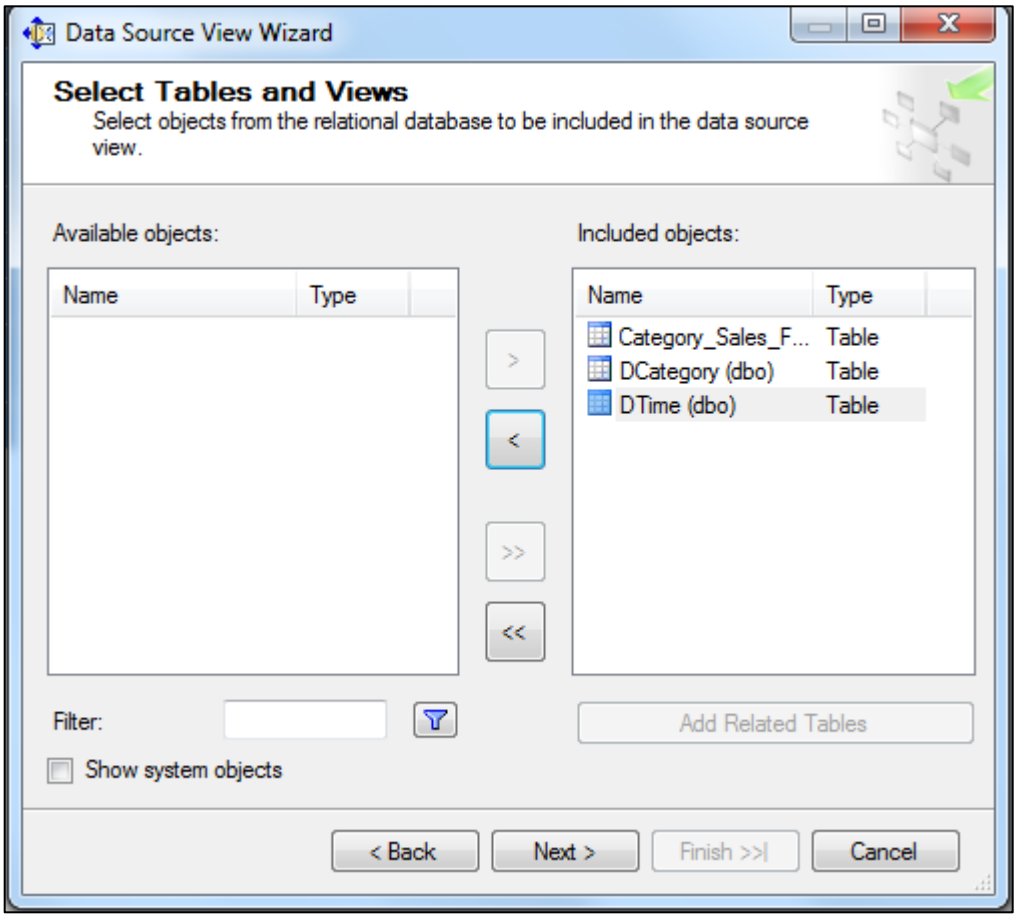
3.2 Report Building Using SSAS

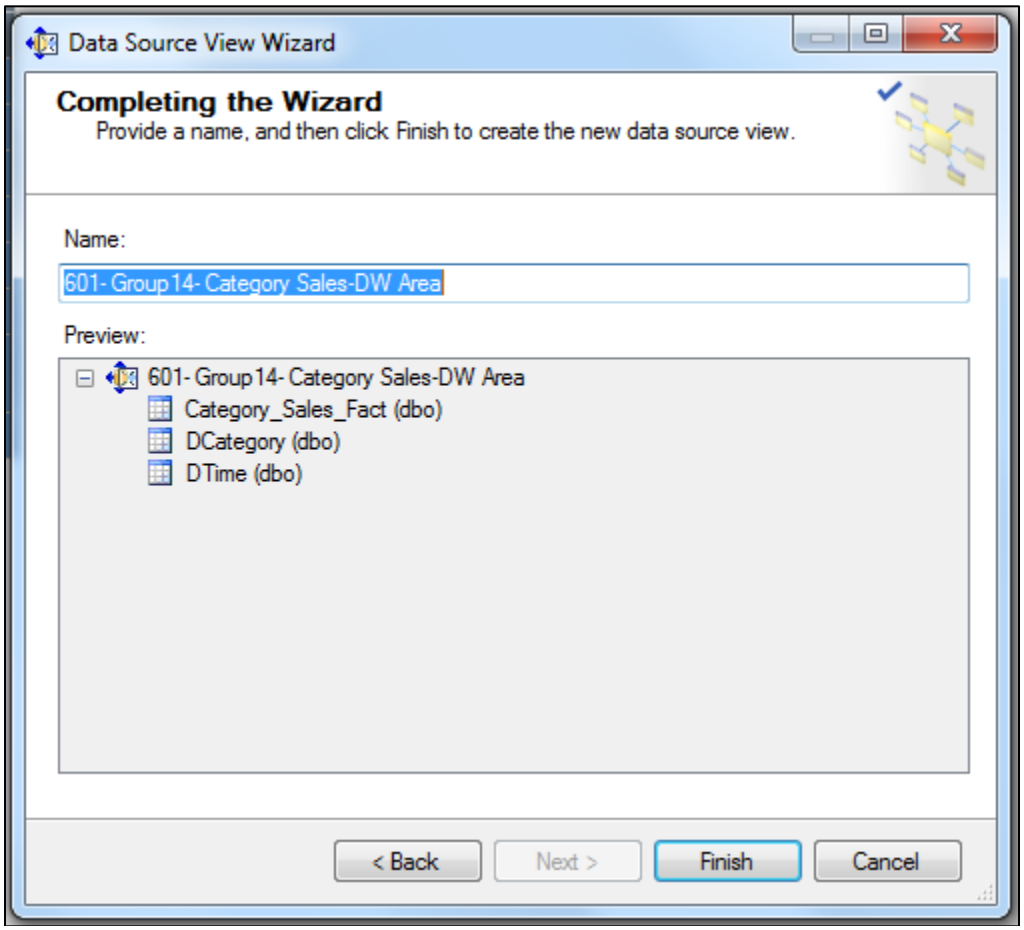
Question 3 - How do the total sales of Grocery and Dairy vary over years? What happened in 1997?

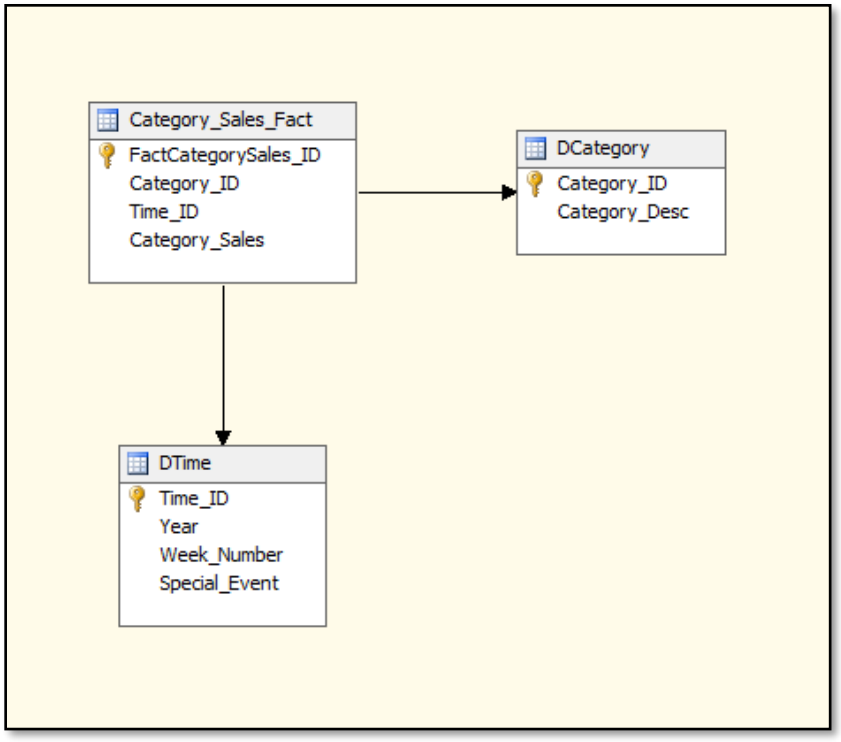


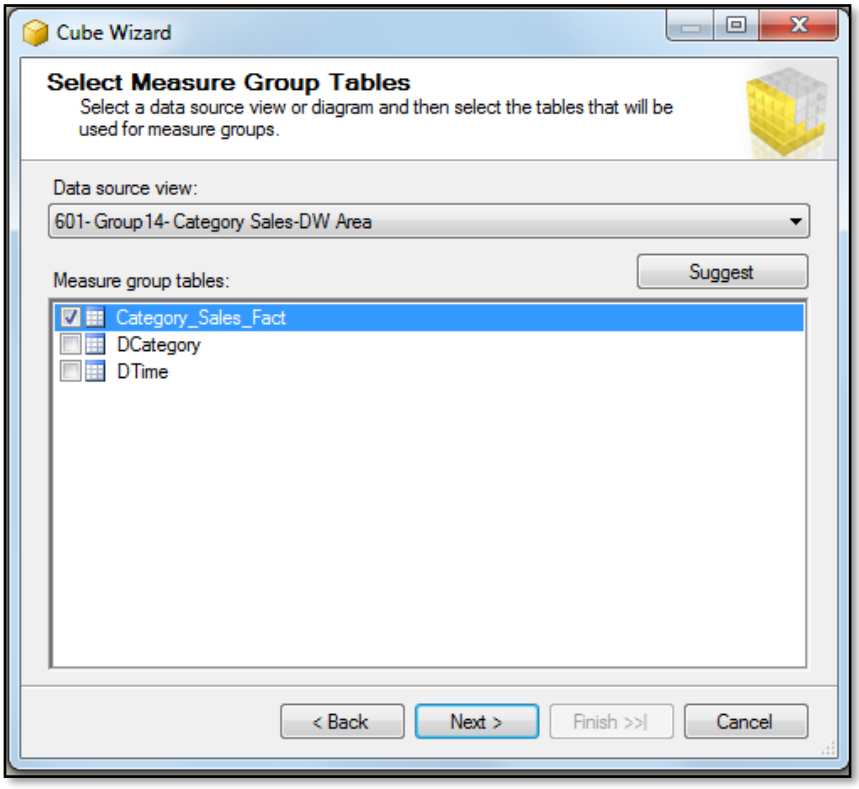

Data Source Wizard
Completing the Wizard
Provide a name and then click Finish to create the new data source.
Data source name:
601- Group14- Category Sales-DW Area
Preview:
Connection string:
Provider=SQLNCLI11.1;Data Source=infodata;Password=;User ID=jh0665;Initial Catalog=601-Group14-CategorySales-DWArea
< Back
Next >
Finish
Cancel

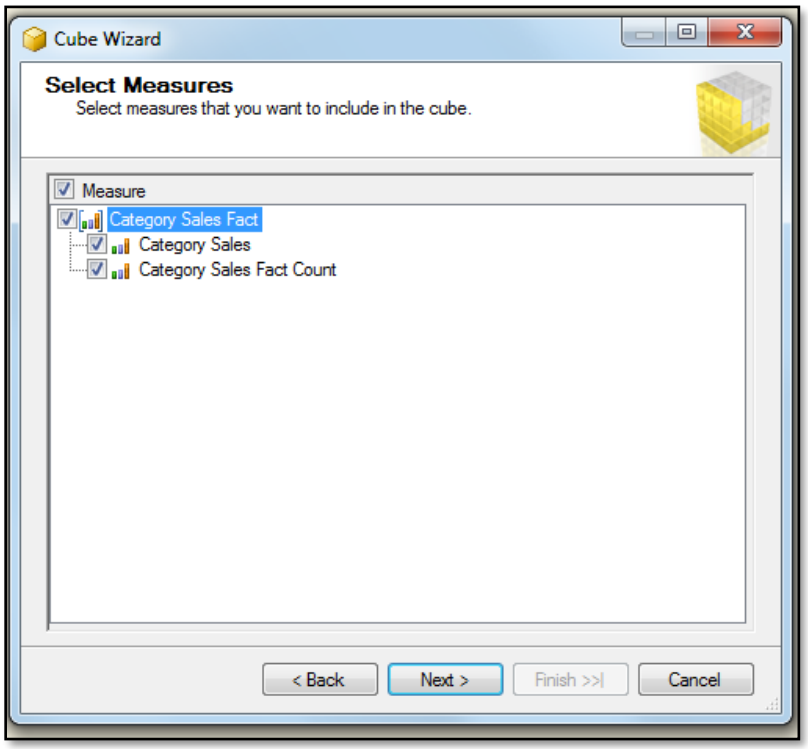

Data Source View Wizard
Select a Data Source
Select an existing relational data source or create a new one.
Relational data sources:
601- Group14- Category Sales-DW Area
Data source properties:
Property
Value
Data Source
infodata
Initial Catalog
601-Group14-CategoryS...
Provider
SQLNCLI11.1
User ID
jh0665
New Data Source...
Advanced...
< Back
Next >
Finish >>|
Cancel

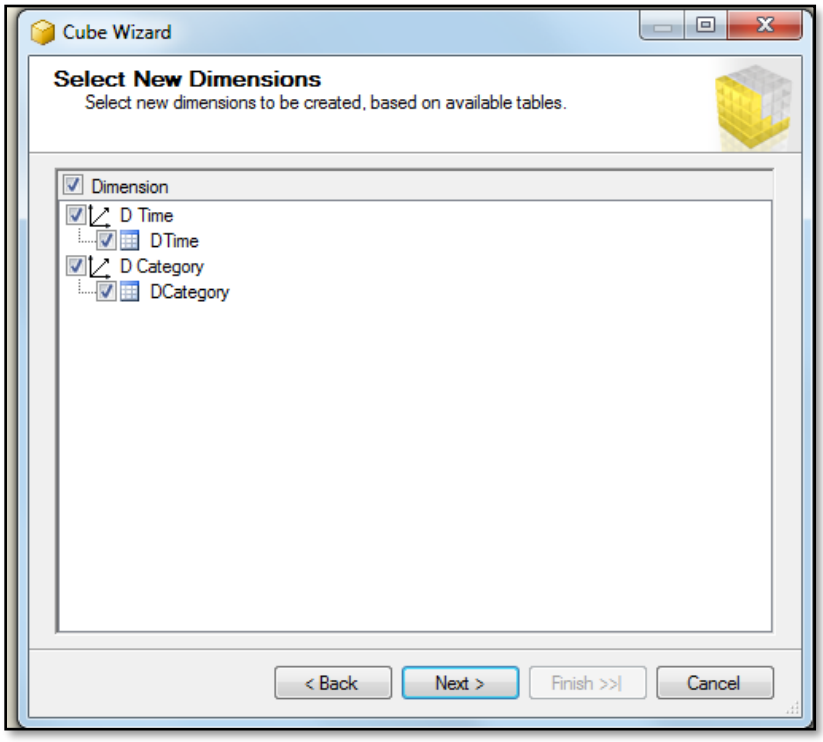

Data Source View Wizard
Select Tables and Views
Select objects from the relational database to be included in the data source view.
Available objects:
Name
Type
Included objects:
Name
Type
Category_Sales_F...
Table
DCategory (dbo)
Table
DTime (dbo)
Table
>
<
>>
<<
Filter:
Add Related Tables
Show system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

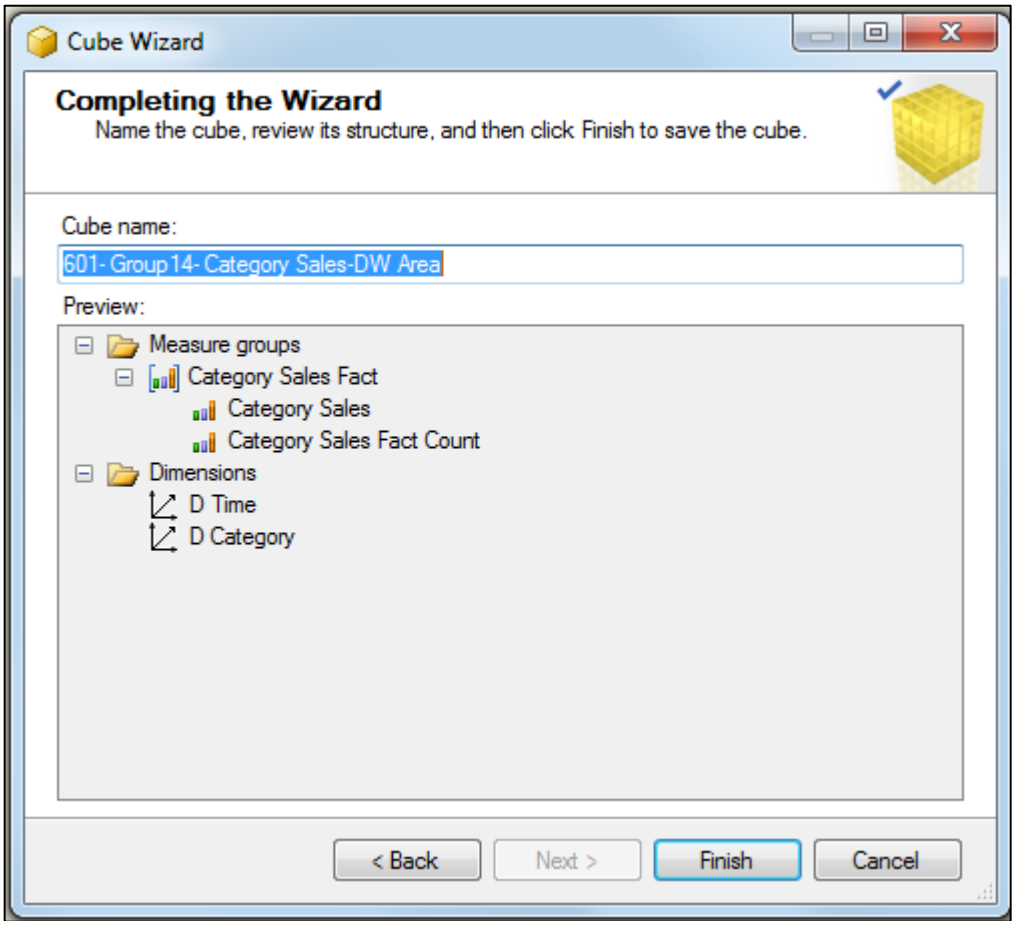

Data Source View Wizard
Completing the Wizard
Provide a name, and then click Finish to create the new data source view.
Name:
601- Group14- Category Sales-DW Area
Preview:
601- Group14- Category Sales-DW Area
Category_Sales_Fact (dbo)
DCategory (dbo)
DTime (dbo)
< Back
Next >
Finish
Cancel

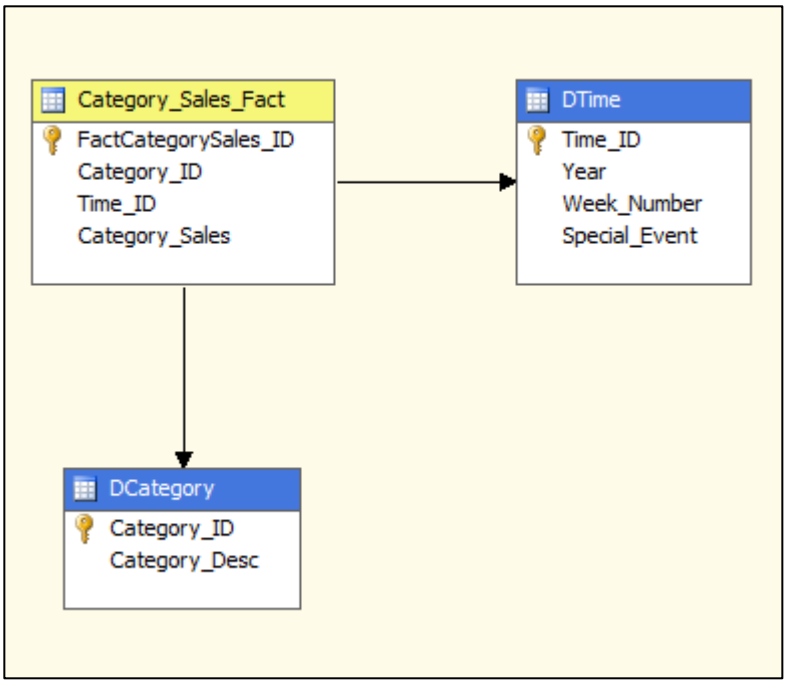

Category_Sales_Fact
FactCategorySales_ID
Category_ID
Time_ID
Category_Sales
DCategory
Category_ID
Category_Desc
DTime
Time_ID
Year
Week_Number
Special_Event

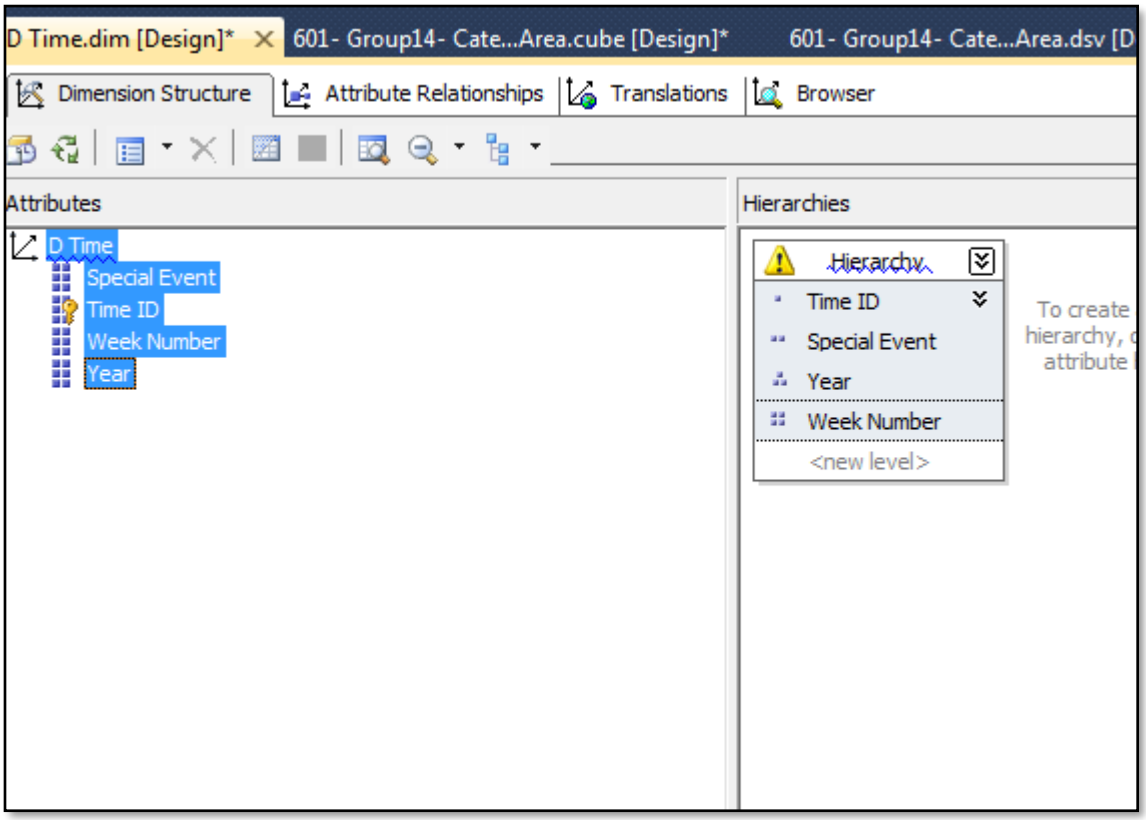

Cube Wizard
Select Measure Group Tables
Select a data source view or diagram and then select the tables that will be used for measure groups.
Data source view:
601- Group14- Category Sales-DW Area
Measure group tables:
Suggest
Category_Sales_Fact
DCategory
DTime
< Back
Next >
Finish >>|
Cancel

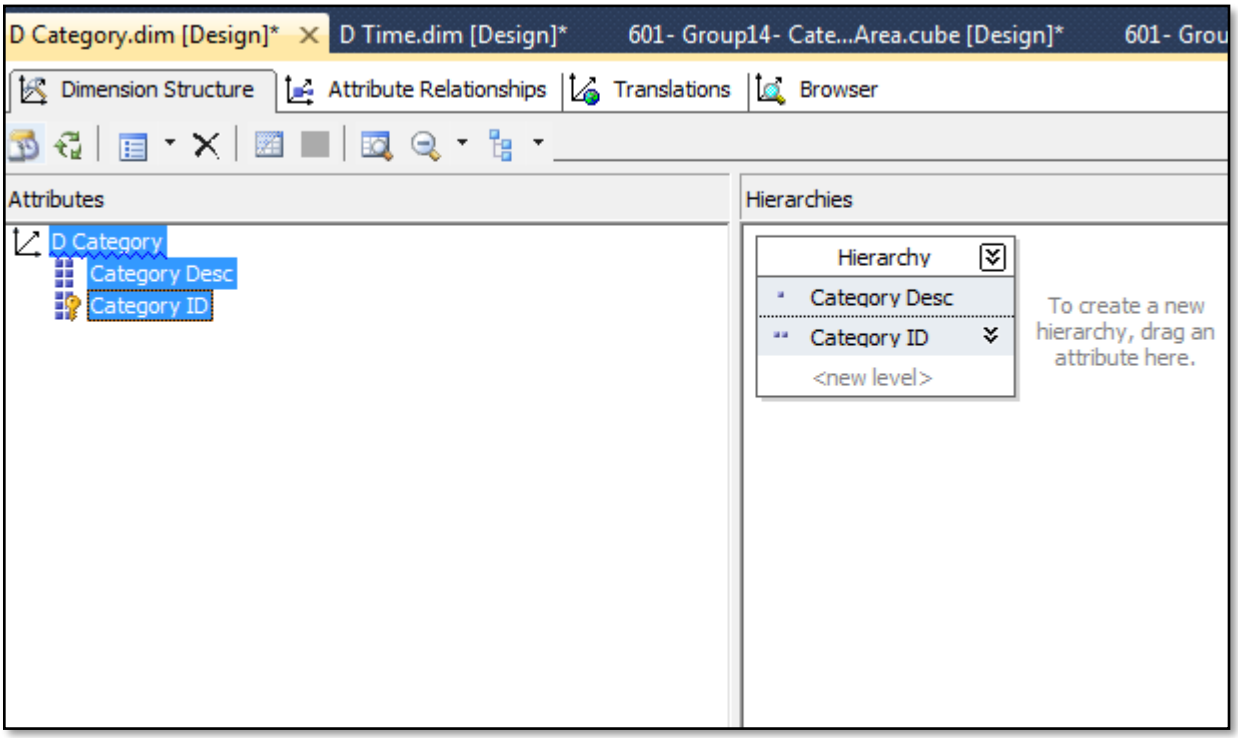

Cube Wizard
Select Measures
Select measures that you want to include in the cube.
Measure
Category Sales Fact
Category Sales
Category Sales Fact Count
< Back
Next >
Finish >>|
Cancel

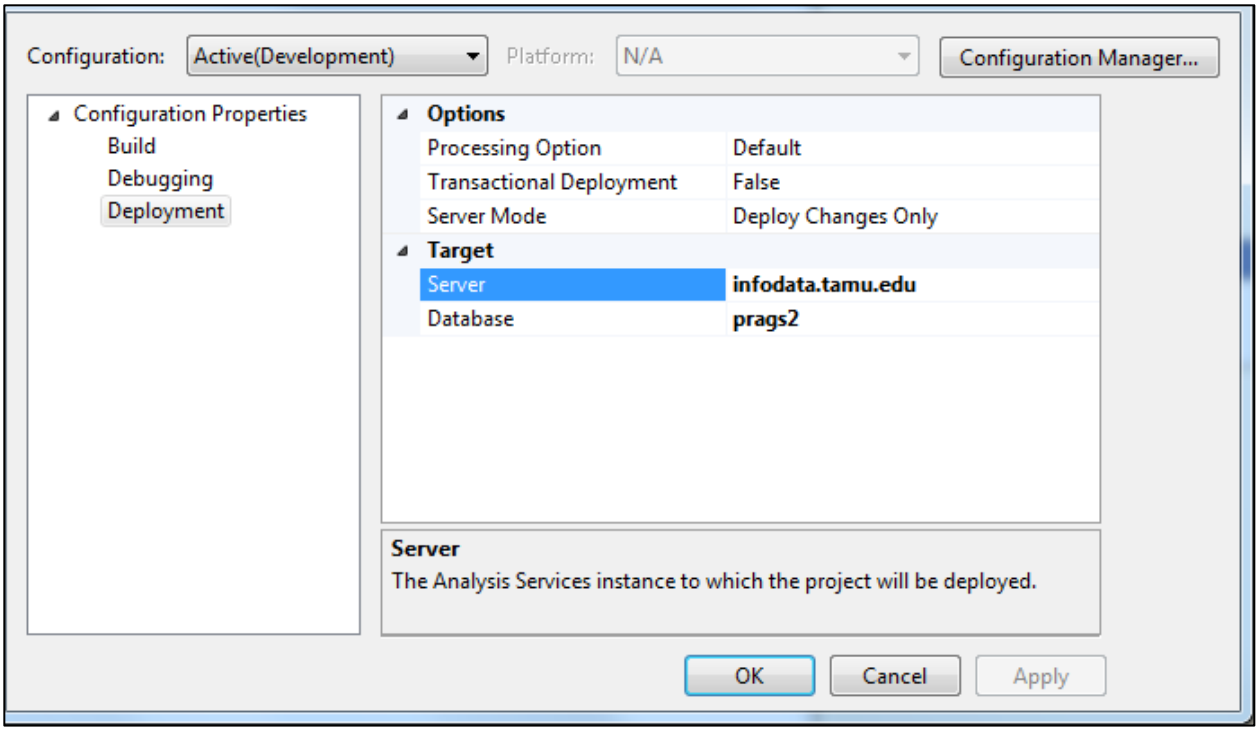

Cube Wizard
Select New Dimensions
Select new dimensions to be created, based on available tables.
Dimension
D Time
DTime
D Category
DCategory
< Back
Next >
Finish >>|
Cancel

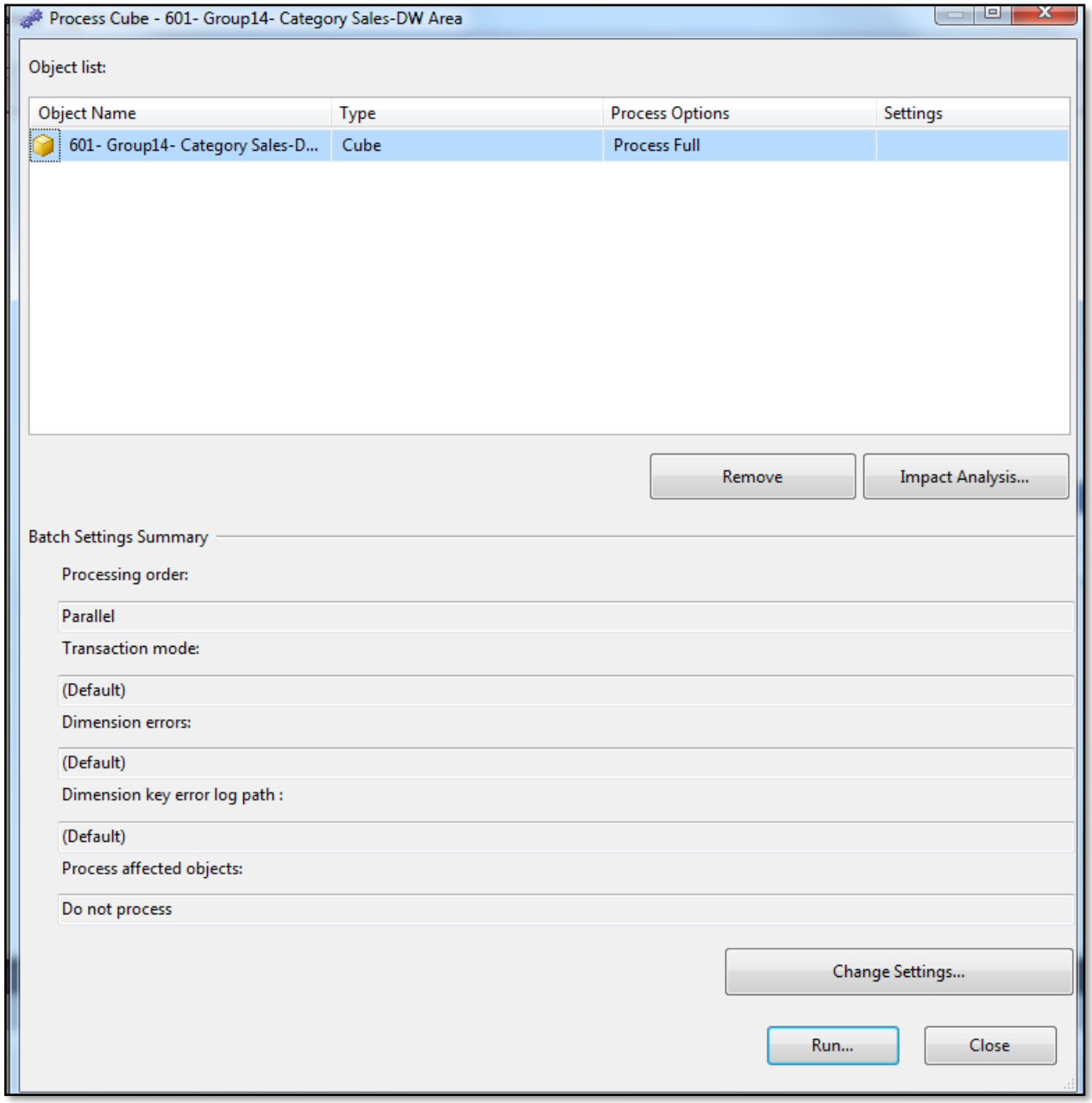

Cube Wizard
Completing the Wizard
Name the cube, review its structure, and then click Finish to save the cube.
Cube name:
601- Group14- Category Sales-DW Area
Preview:
Measure groups
Category Sales Fact
Category Sales
Category Sales Fact Count
Dimensions
D Time
D Category
< Back
Next >
Finish
Cancel

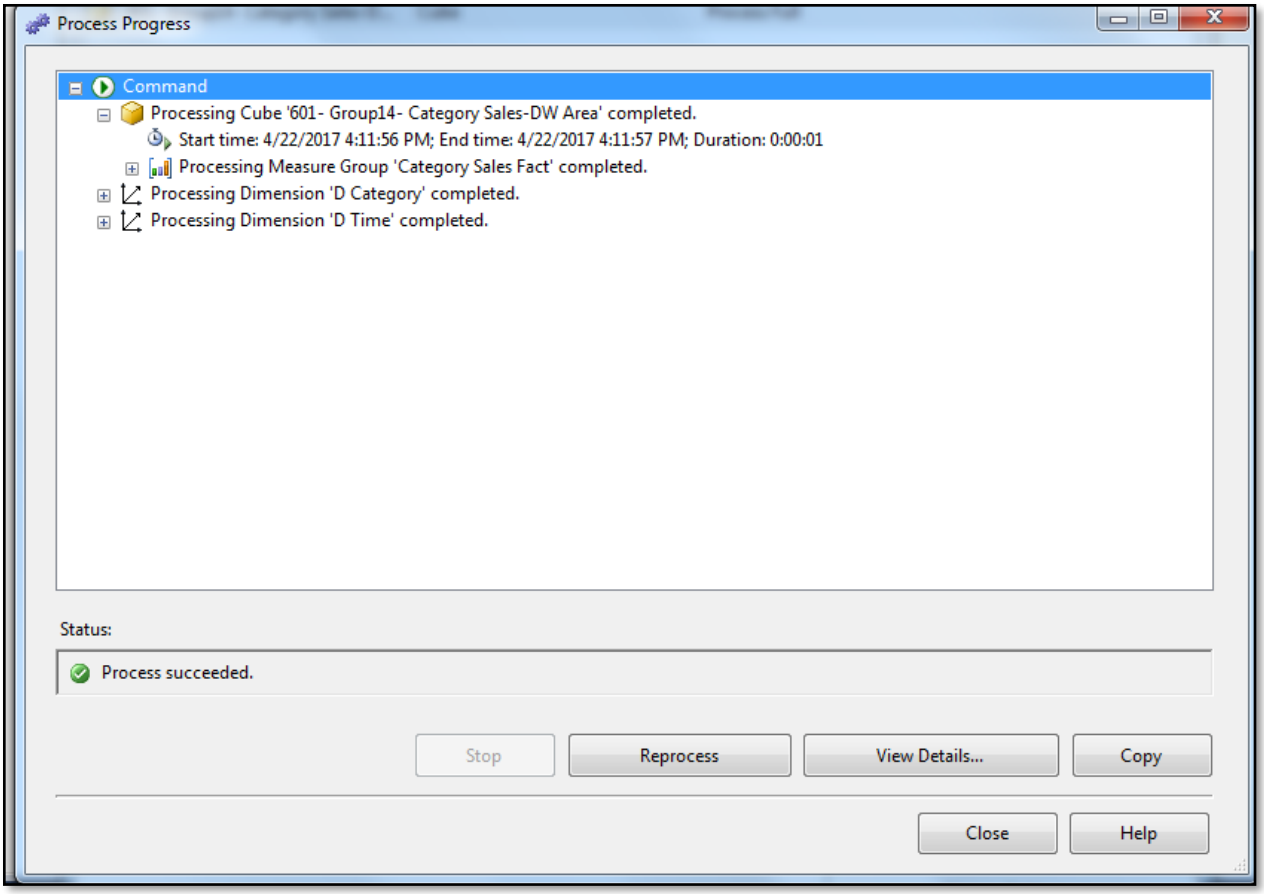

Category_Sales_Fact
FactCategorySales_ID
Category_ID
Time_ID
Category_Sales
DTime
Time_ID
Year
Week_Number
Special_Event
DCategory
Category_ID
Category_Desc

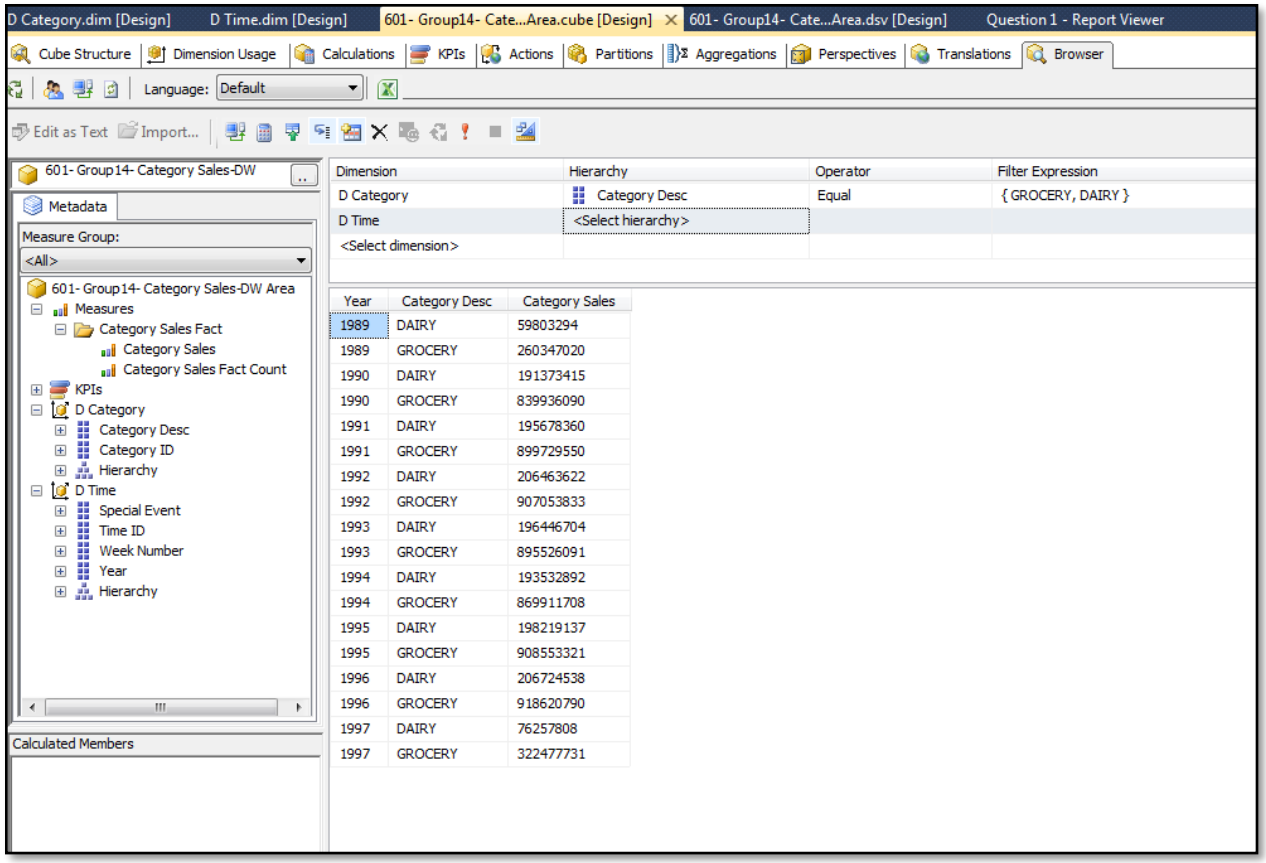

D Time.dim [Design]*
601- Group14- Cate...Area.cube [Design]*
601- Group14- Cate...Area.dsv [D
Dimension Structure
Attribute Relationships
Translations
Browser
Attributes
Hierarchies
D Time
Special Event
Time ID
Week Number
Year
Hierarchy
Time ID
Special Event
Year
Week Number
<new level>
To create
hierarchy,
attribute


D Category.dim [Design]*
D Time.dim [Design]*
601- Group14- Cate...Area.cube [Design]*
601- Grou
Dimension Structure
Attribute Relationships
Translations
Browser
Attributes
Hierarchies
D Category
Category Desc
Category ID
Hierarchy
Category Desc
Category ID
<new level>
To create a new hierarchy, drag an attribute here.

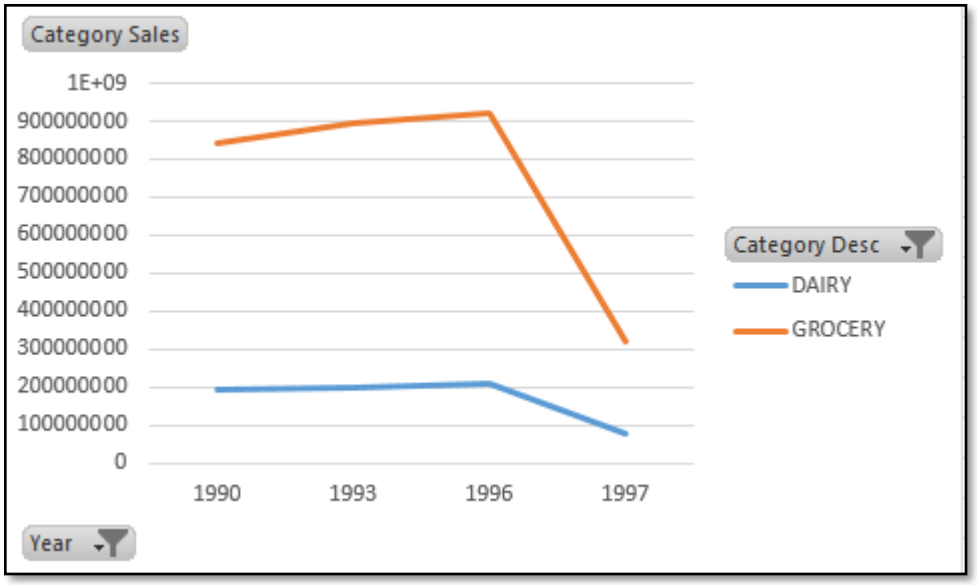

Configuration:
Active(Development)
Platform:
N/A
Configuration Manager...
Configuration Properties
Build
Debugging
Deployment
Options
Processing Option
Default
Transactional Deployment
False
Server Mode
Deploy Changes Only
Target
Server
infodata.tamu.edu
Database
prags2
Server
The Analysis Services instance to which the project will be deployed.
OK
Cancel
Apply


Process Cube - 601- Group14- Category Sales-DW Area
Object list:
Object Name
Type
Process Options
Settings
601- Group14- Category Sales-D...
Cube
Process Full
Remove
Impact Analysis...
Batch Settings Summary
Processing order:
Parallel
Transaction mode:
(Default)
Dimension errors:
(Default)
Dimension key error log path :
(Default)
Process affected objects:
Do not process
Change Settings...
Run...
Close


Process Progress
Command
Processing Cube '601- Group14- Category Sales-DW Area' completed.
Start time: 4/22/2017 4:11:56 PM; End time: 4/22/2017 4:11:57 PM; Duration: 0:00:01
Processing Measure Group 'Category Sales Fact' completed.
Processing Dimension 'D Category' completed.
Processing Dimension 'D Time' completed.
Status:
Process succeeded.
Stop
Reprocess
View Details...
Copy
Close
Help



Category Sales	Category Desc	
Year	DAIRY	GROCERY
1990	191373415	839936090
1993	196446704	895526091
1996	206724538	918620790
1997	76257808	322477731

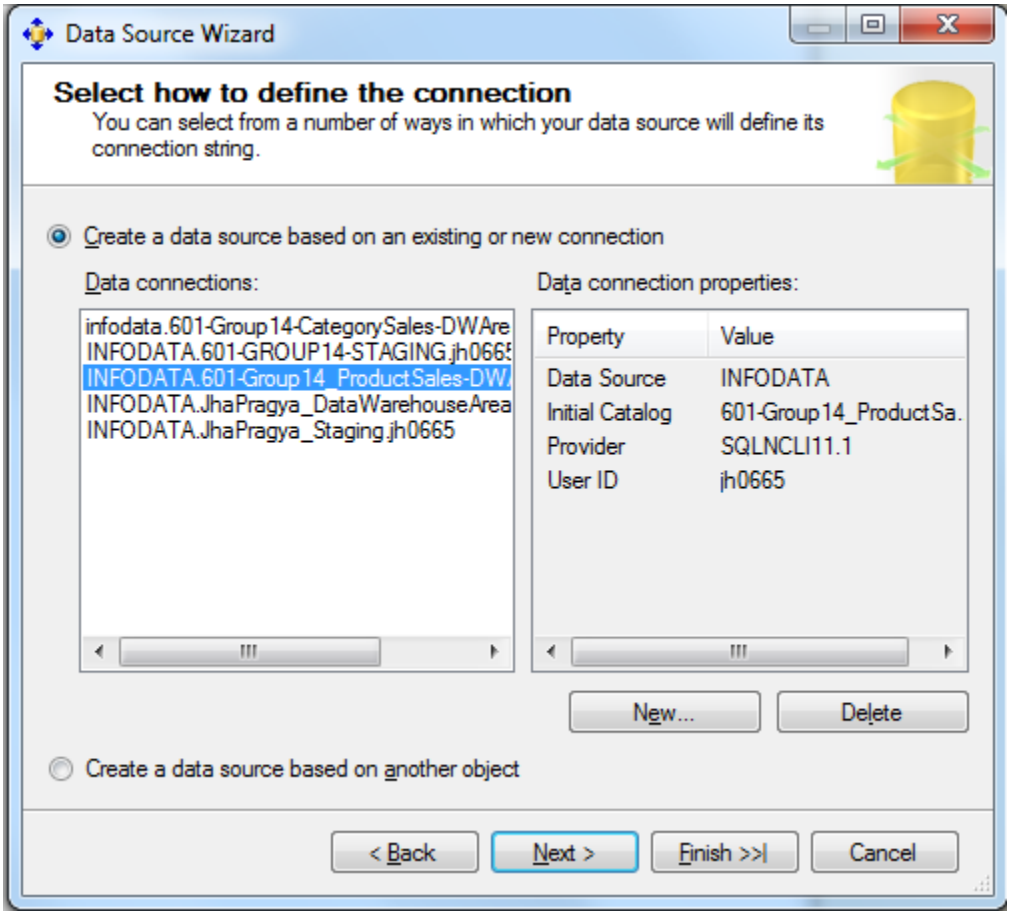


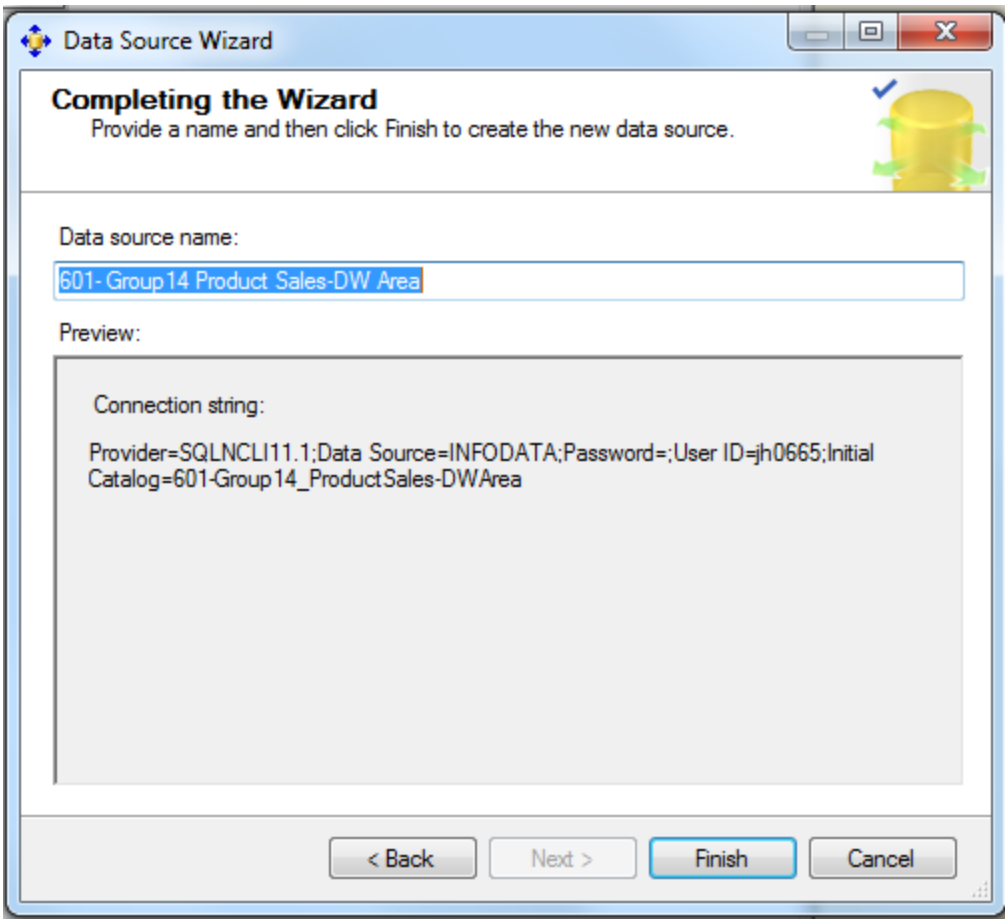
Analysis and Conclusion -

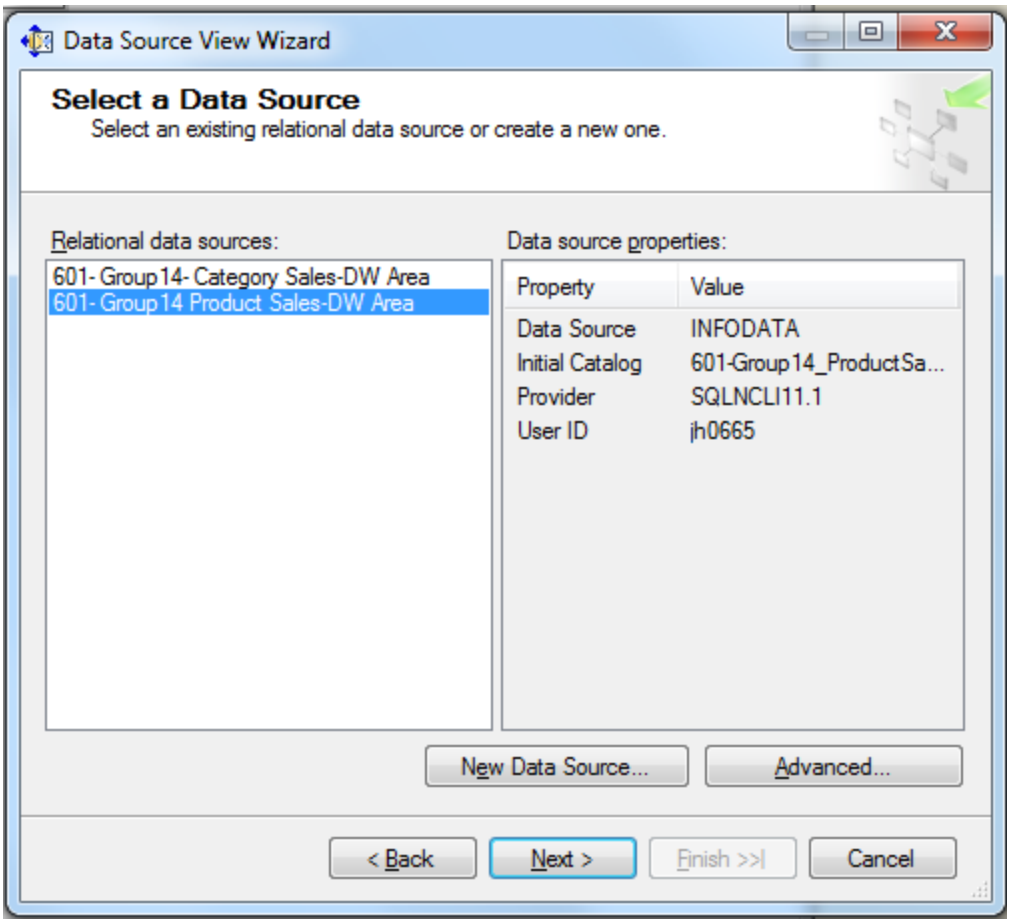
"What happened in 1997?" - this is an important business question which our report helps to answer. After creating a cube in SSAS, we selected the product category, product sales and time to build a report. In order to better visualize the results, we used the Export to Excel tool to view the pivot table and chart. From the pivot chart above, we can see a slowly but steadily increasing trend in sales from 1990 to 1996, after which there is a very sharp decline in 1997. The insight provided by this information is definitely one that will be of a major concern to DFF and thus, inspire them to urgently find and implement strategies to remedy their rapidly declining sales.

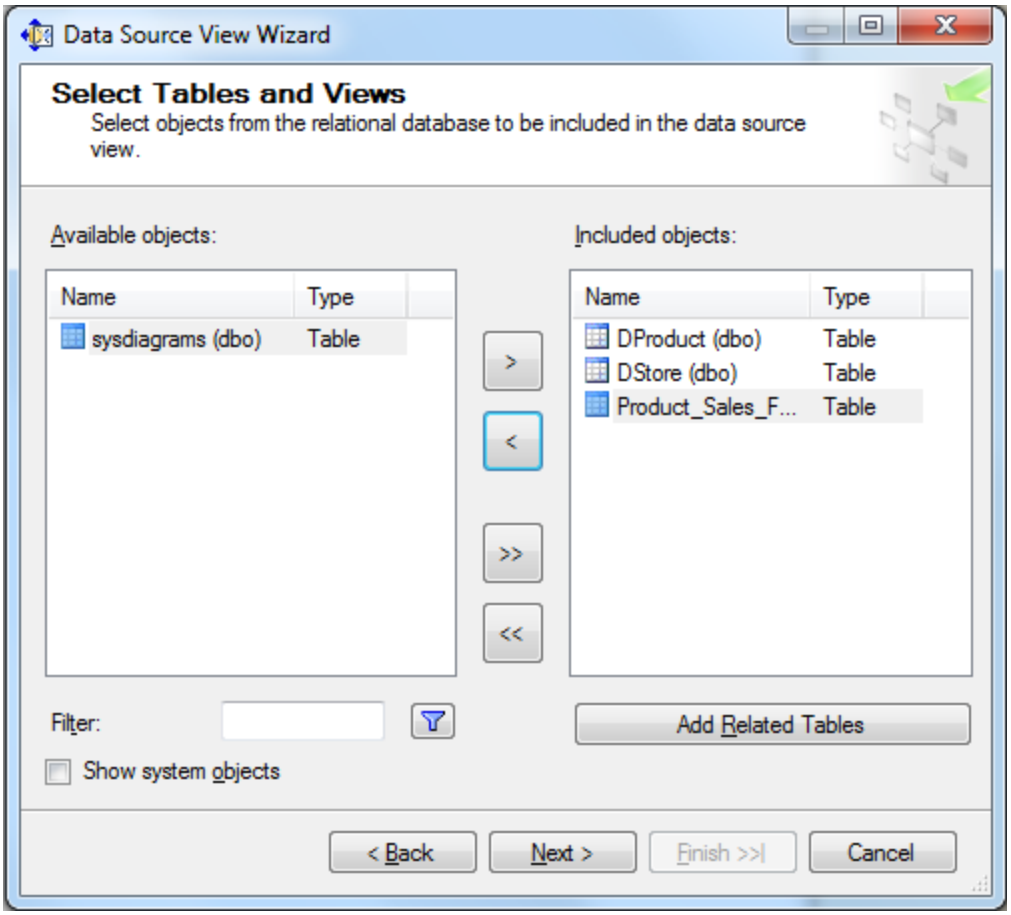
3.4 Report Building Using SSRS on top of SSAS

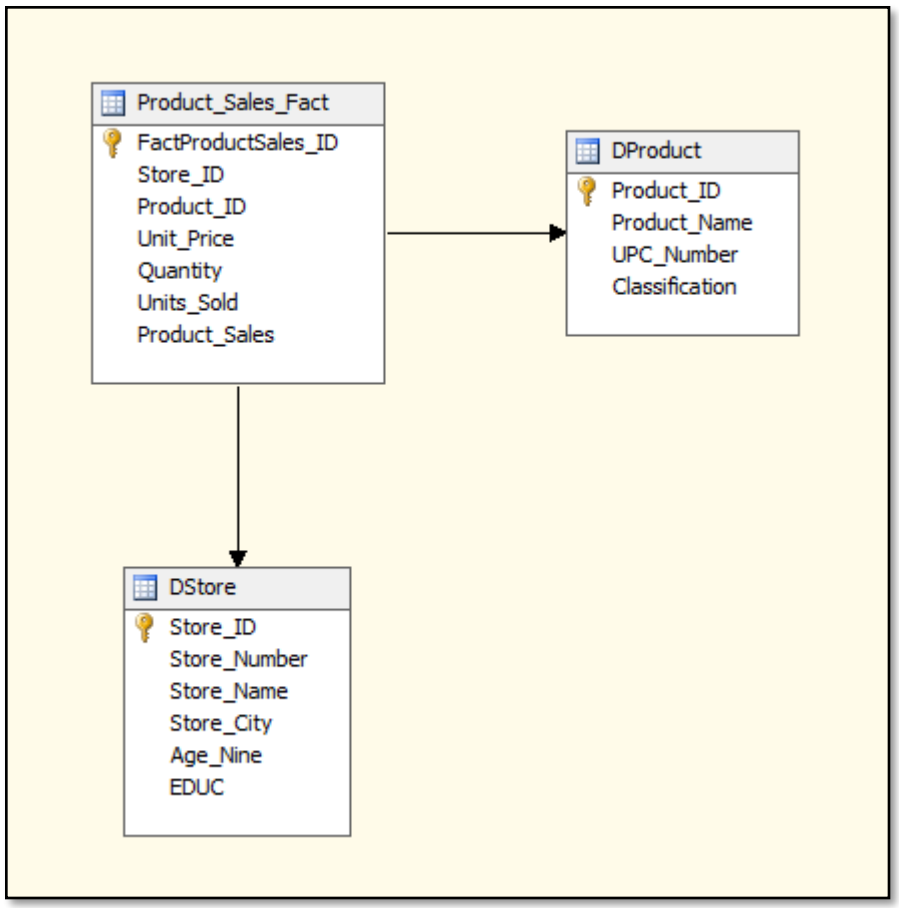
Question 4 - What is the trend in the sale of cookies (from week 1 – 399) in stores with the highest % population of kids under age 9?

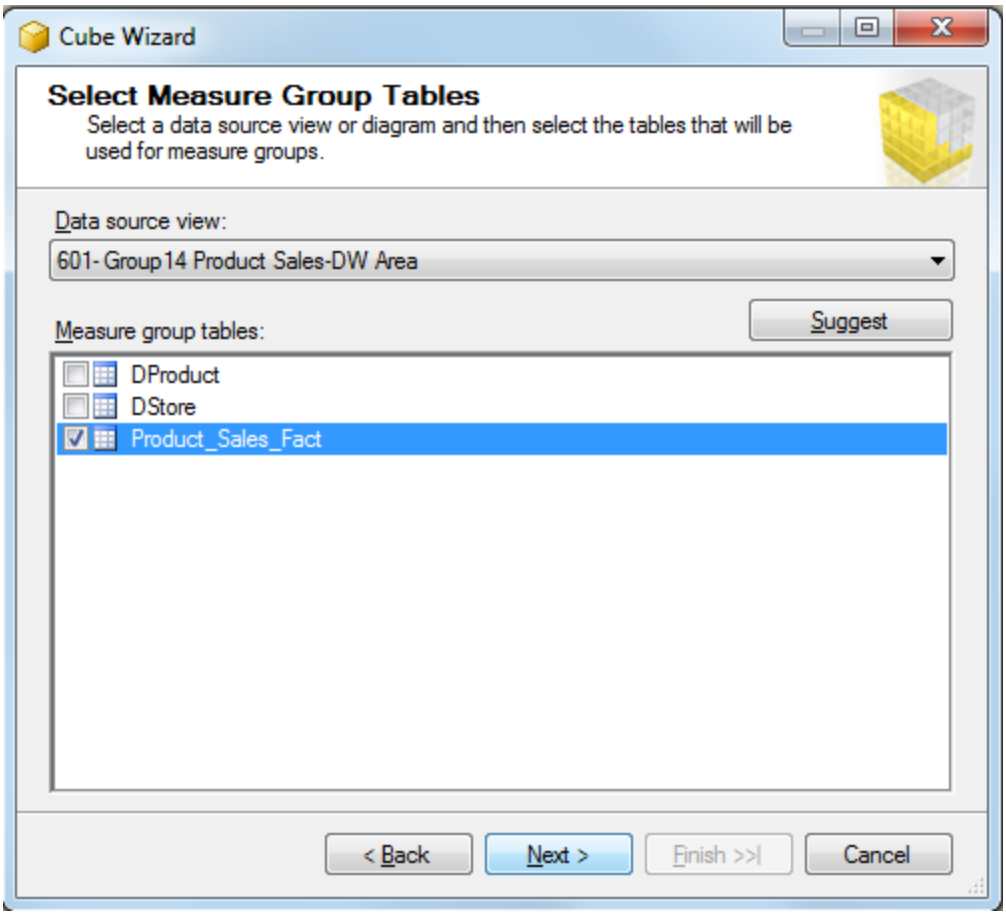


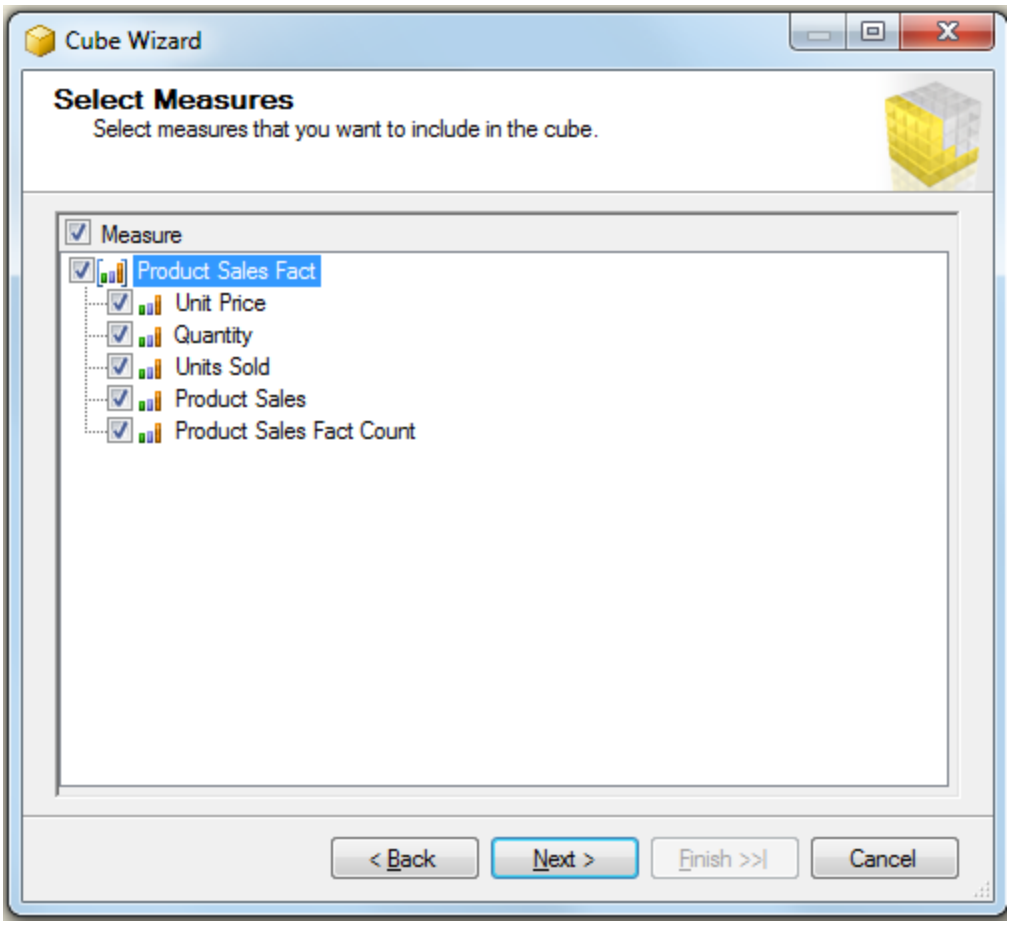

Data Source Wizard
Completing the Wizard
Provide a name and then click Finish to create the new data source.
Data source name:
601- Group14 Product Sales-DW Area
Preview:
Connection string:
Provider=SQLNCLI11.1;Data Source=INFODATA;Password=;User ID=jh0665;Initial Catalog=601-Group14_ProductSales-DWArea
< Back
Next >
Finish
Cancel

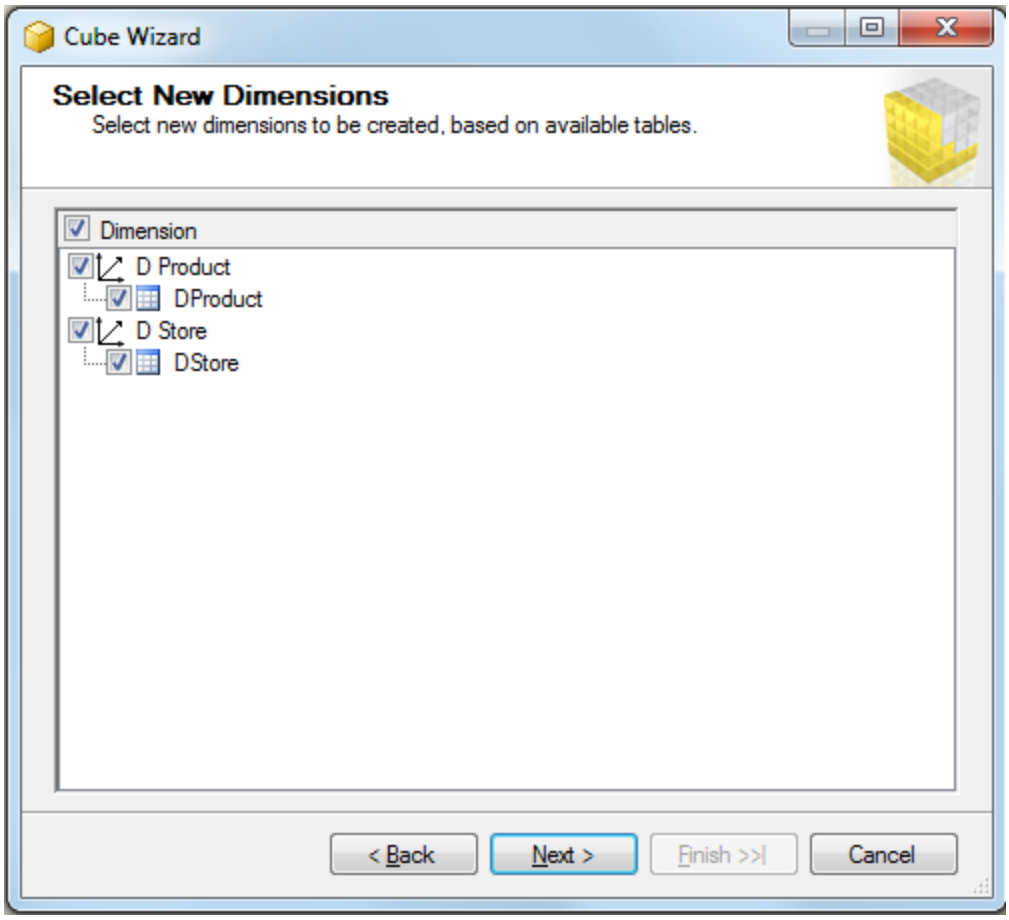

Data Source View Wizard
Select a Data Source
Select an existing relational data source or create a new one.
Relational data sources:
601- Group14- Category Sales-DW Area
601- Group14 Product Sales-DW Area
Data source properties:
Property
Value
Data Source
INFODATA
Initial Catalog
601-Group14_ProductSa...
Provider
SQLNCLI11.1
User ID
jh0665
New Data Source...
Advanced...
< Back
Next >
Finish >>|
Cancel

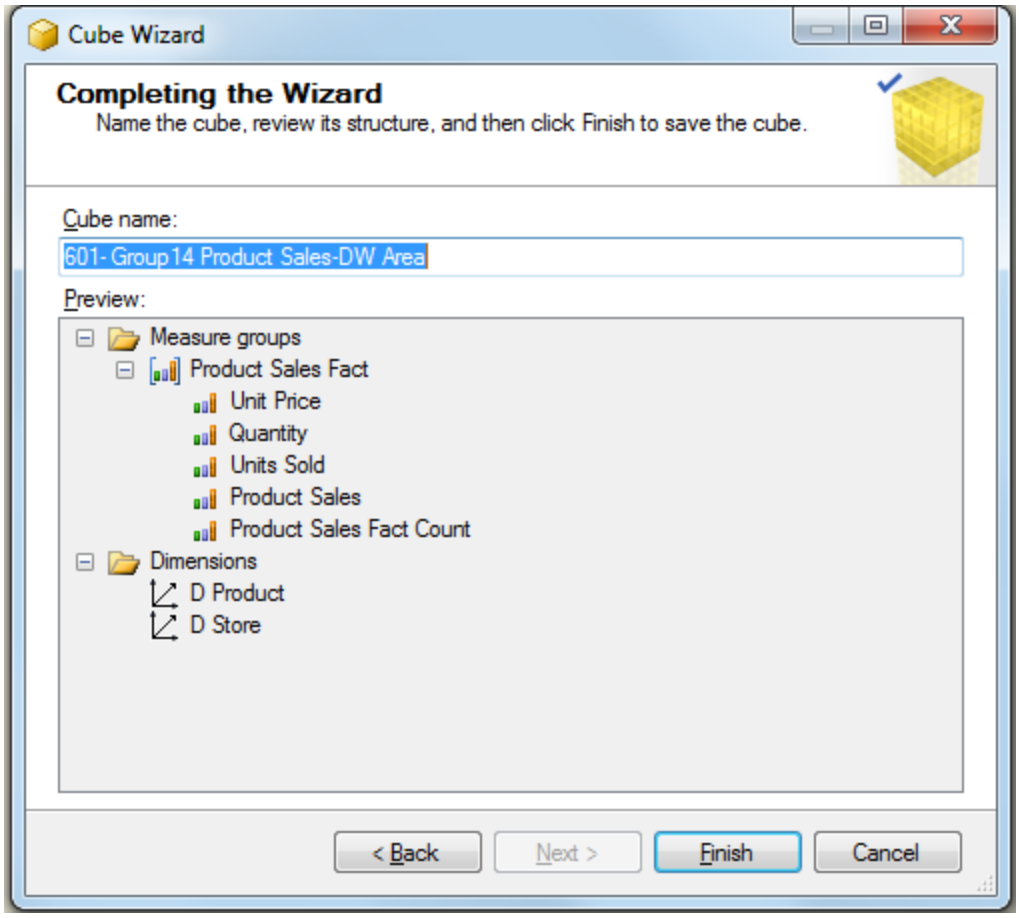

Data Source View Wizard
Select Tables and Views
Select objects from the relational database to be included in the data source view.
Available objects:
Name
Type
sysdiagrams (dbo)
Table
Included objects:
Name
Type
DProduct (dbo)
Table
DStore (dbo)
Table
Product_Sales_F...
Table
>
<
>>
<<
Filter:
Add Related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

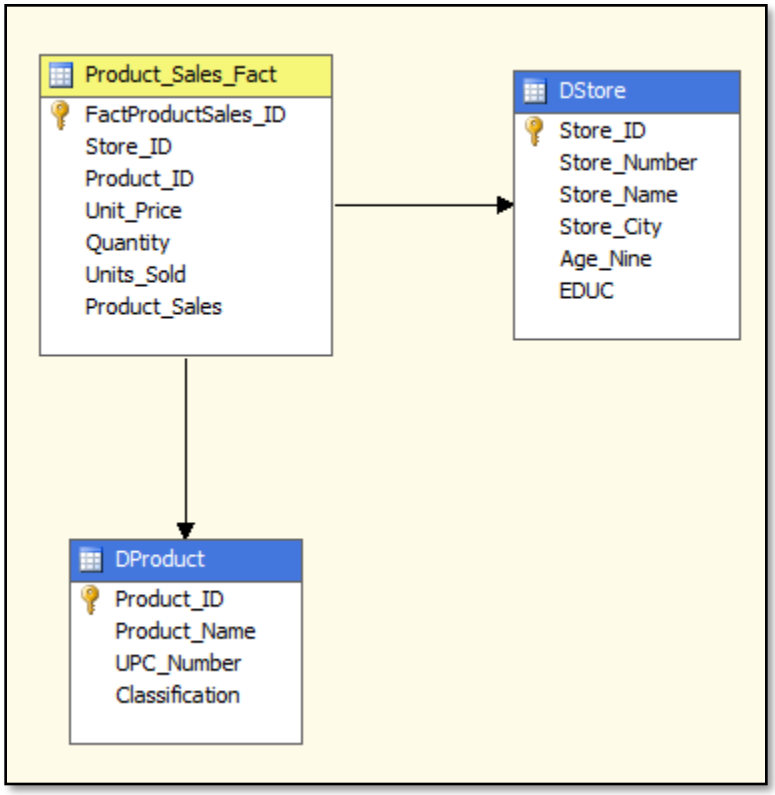

Product_Sales_Fact
FactProductSales_ID
Store_ID
Product_ID
Unit_Price
Quantity
Units_Sold
Product_Sales
DProduct
Product_ID
Product_Name
UPC_Number
Classification
DStore
Store_ID
Store_Number
Store_Name
Store_City
Age_Nine
EDUC

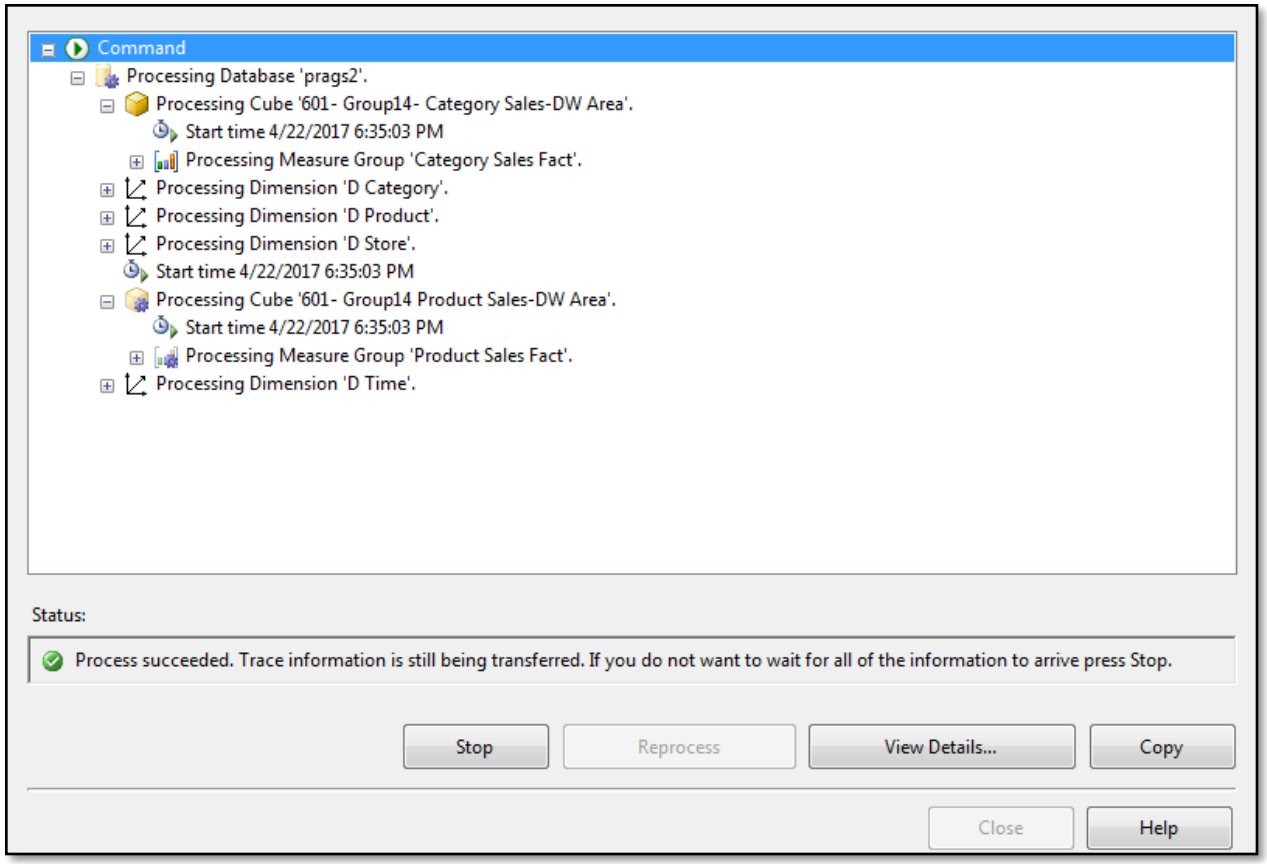

Cube Wizard
Select Measure Group Tables
Select a data source view or diagram and then select the tables that will be used for measure groups.
Data source view:
601- Group14 Product Sales-DW Area
Measure group tables:
Suggest
DProduct
DStore
Product_Sales_Fact
< Back
Next >
Finish >>|
Cancel

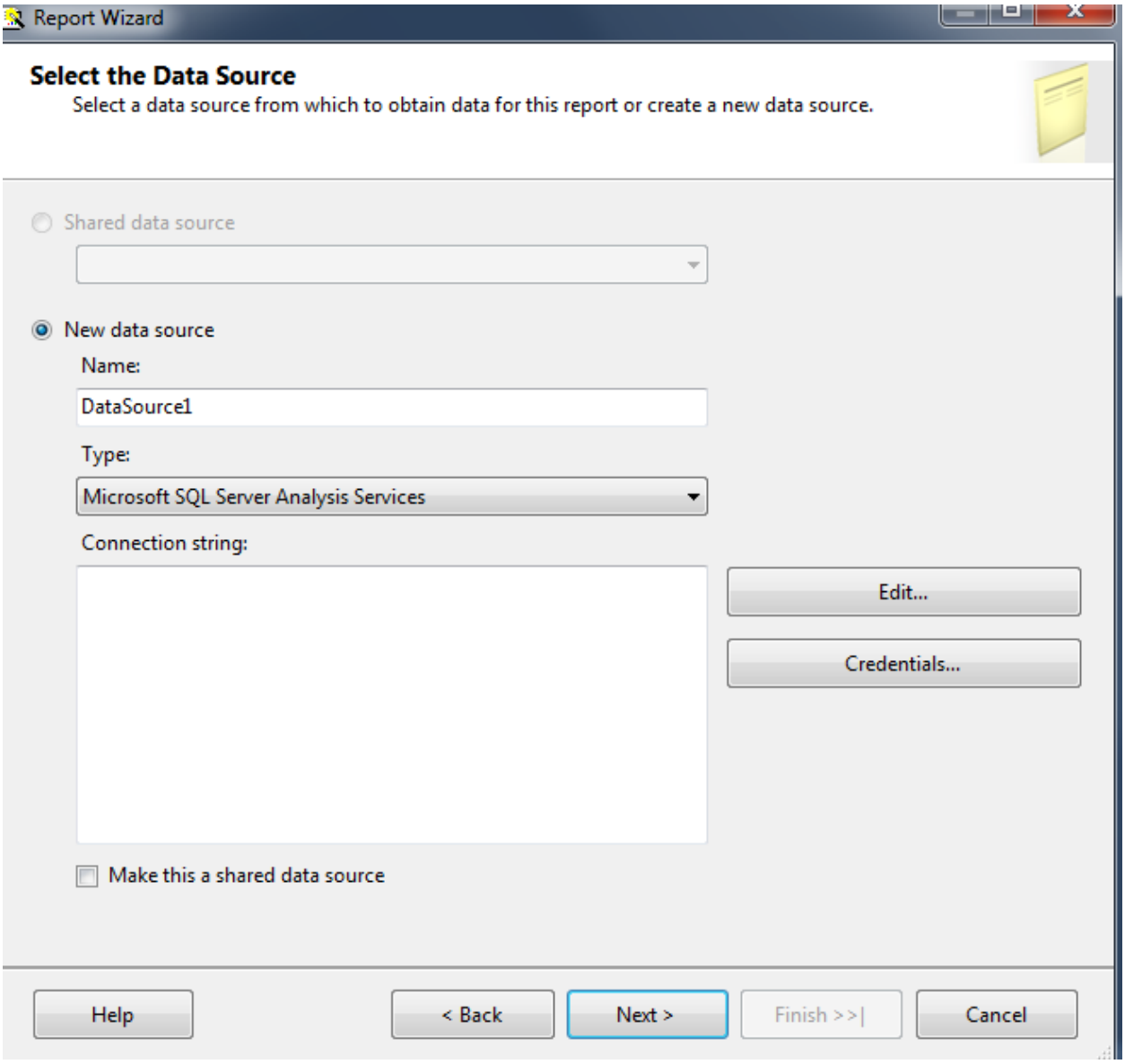

Cube Wizard
Select Measures
Select measures that you want to include in the cube.
Measure
Product Sales Fact
Unit Price
Quantity
Units Sold
Product Sales
Product Sales Fact Count
< Back
Next >
Finish >>|
Cancel

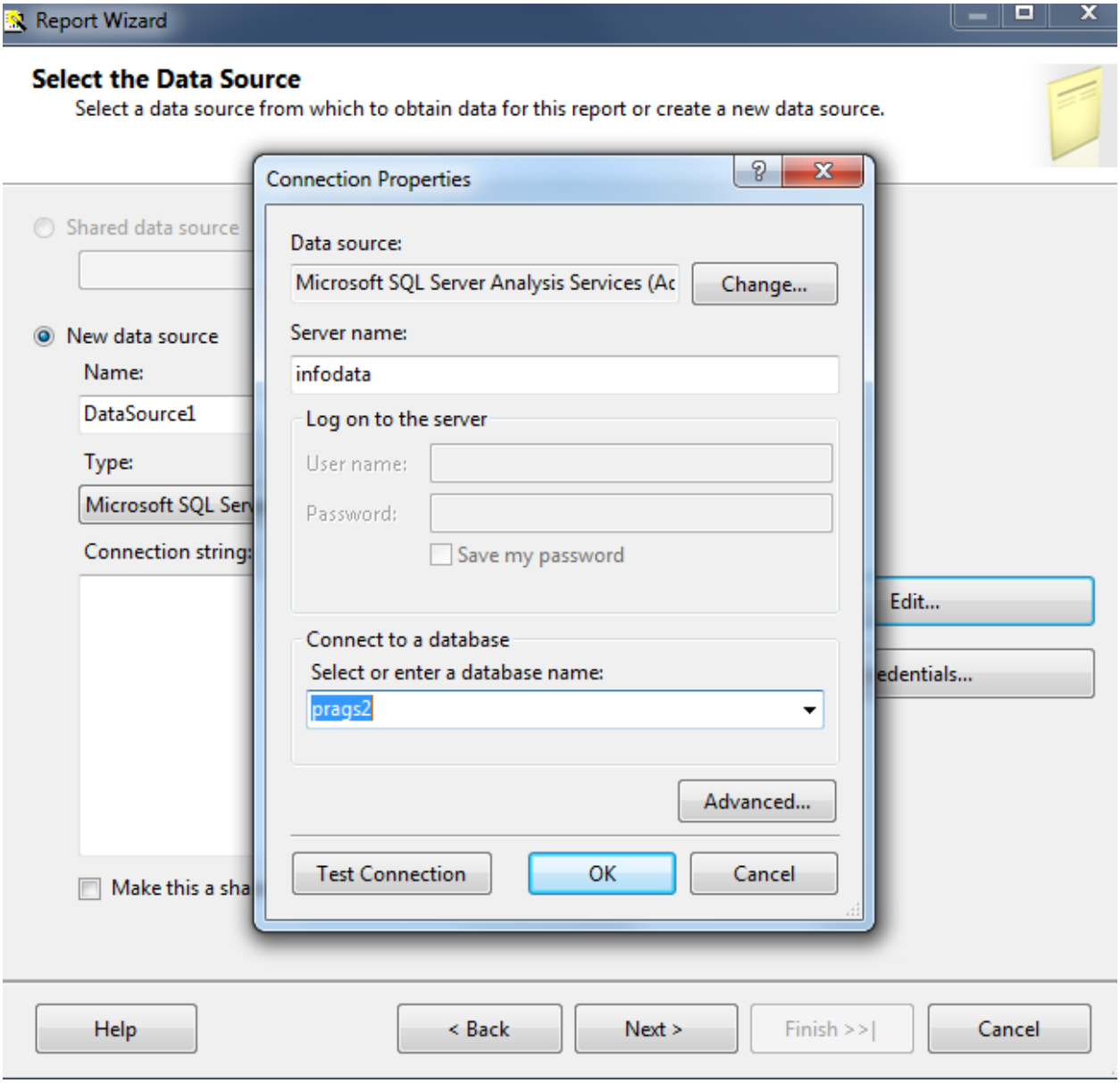

Cube Wizard
Select New Dimensions
Select new dimensions to be created, based on available tables.
Dimension
D Product
DProduct
D Store
DStore
< Back
Next >
Finish >>|
Cancel

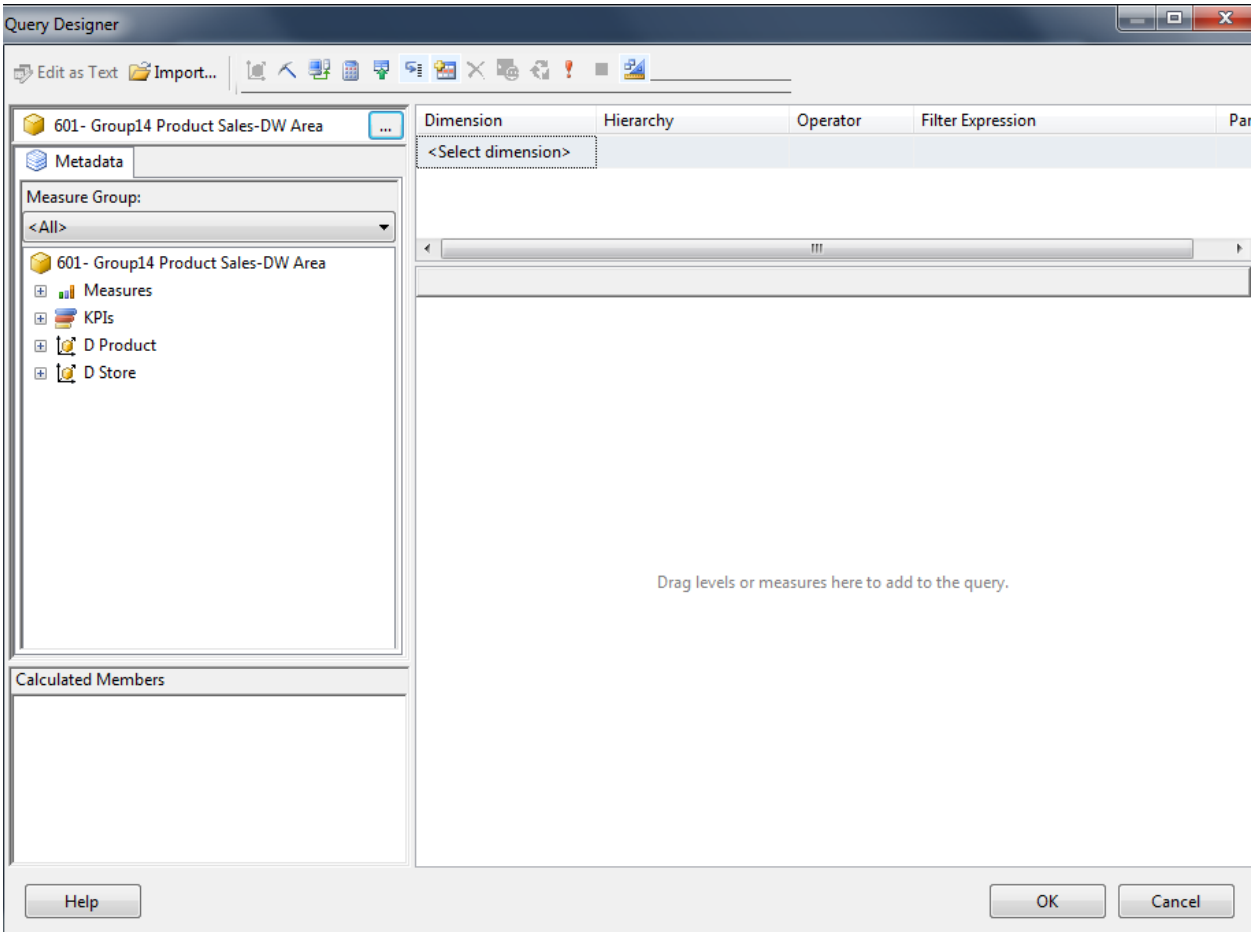

Cube Wizard
Completing the Wizard
Name the cube, review its structure, and then click Finish to save the cube.
Cube name:
601- Group14 Product Sales-DW Area
Preview:
Measure groups
Product Sales Fact
Unit Price
Quantity
Units Sold
Product Sales
Product Sales Fact Count
Dimensions
D Product
D Store
< Back
Next >
Finish
Cancel

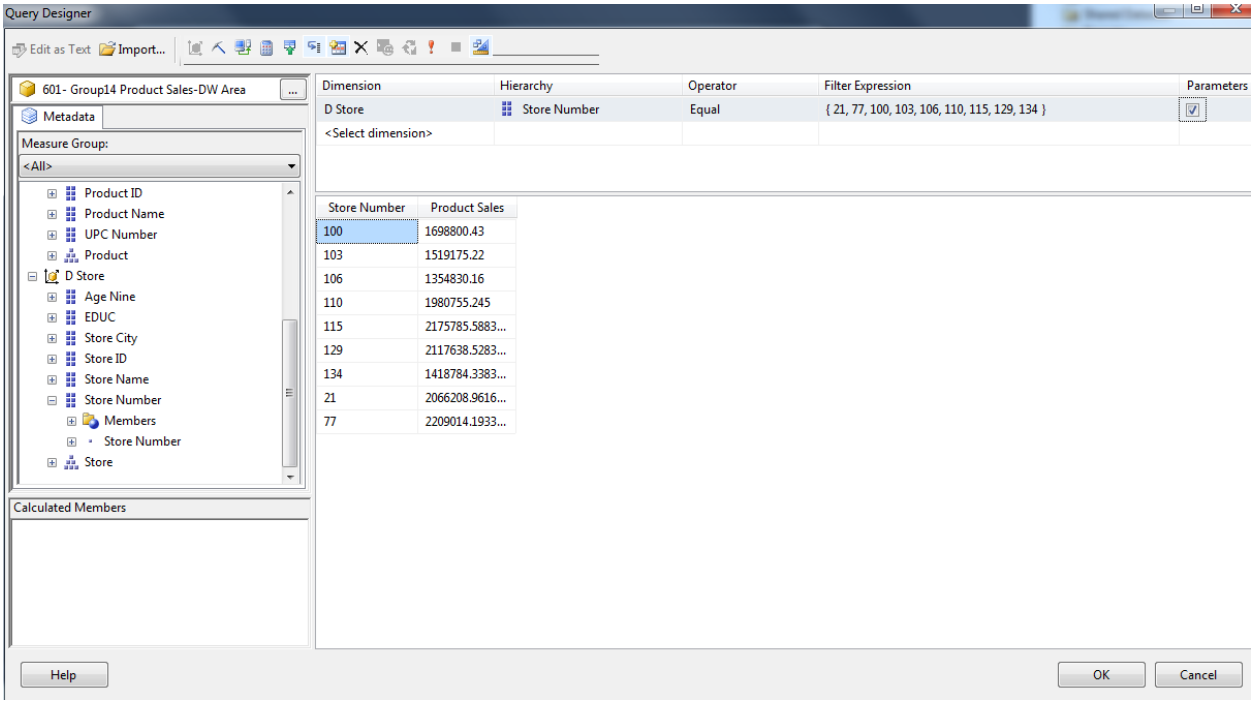

Product_Sales_Fact
FactProductSales_ID
Store_ID
Product_ID
Unit_Price
Quantity
Units_Sold
Product_Sales
DStore
Store_ID
Store_Number
Store_Name
Store_City
Age_Nine
EDUC
DProduct
Product_ID
Product_Name
UPC_Number
Classification

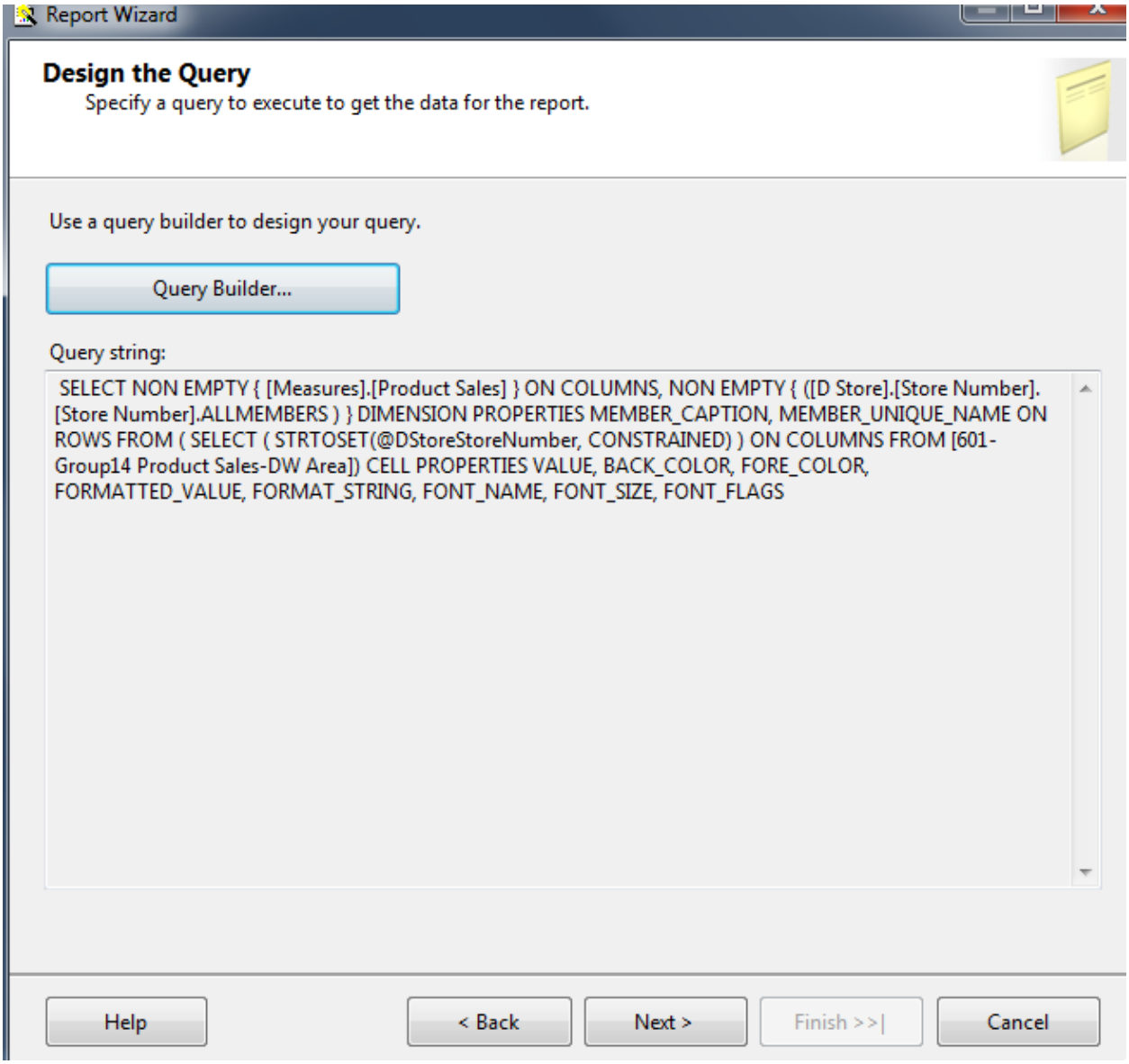

Command
Processing Database 'prags2'.
Processing Cube '601- Group14- Category Sales-DW Area'.
Start time 4/22/2017 6:35:03 PM
Processing Measure Group 'Category Sales Fact'.
Processing Dimension 'D Category'.
Processing Dimension 'D Product'.
Processing Dimension 'D Store'.
Start time 4/22/2017 6:35:03 PM
Processing Cube '601- Group14 Product Sales-DW Area'.
Start time 4/22/2017 6:35:03 PM
Processing Measure Group 'Product Sales Fact'.
Processing Dimension 'D Time'.
Status:
Process succeeded. Trace information is still being transferred. If you do not want to wait for all of the information to arrive press Stop.
Stop
Reprocess
View Details...
Copy
Close
Help

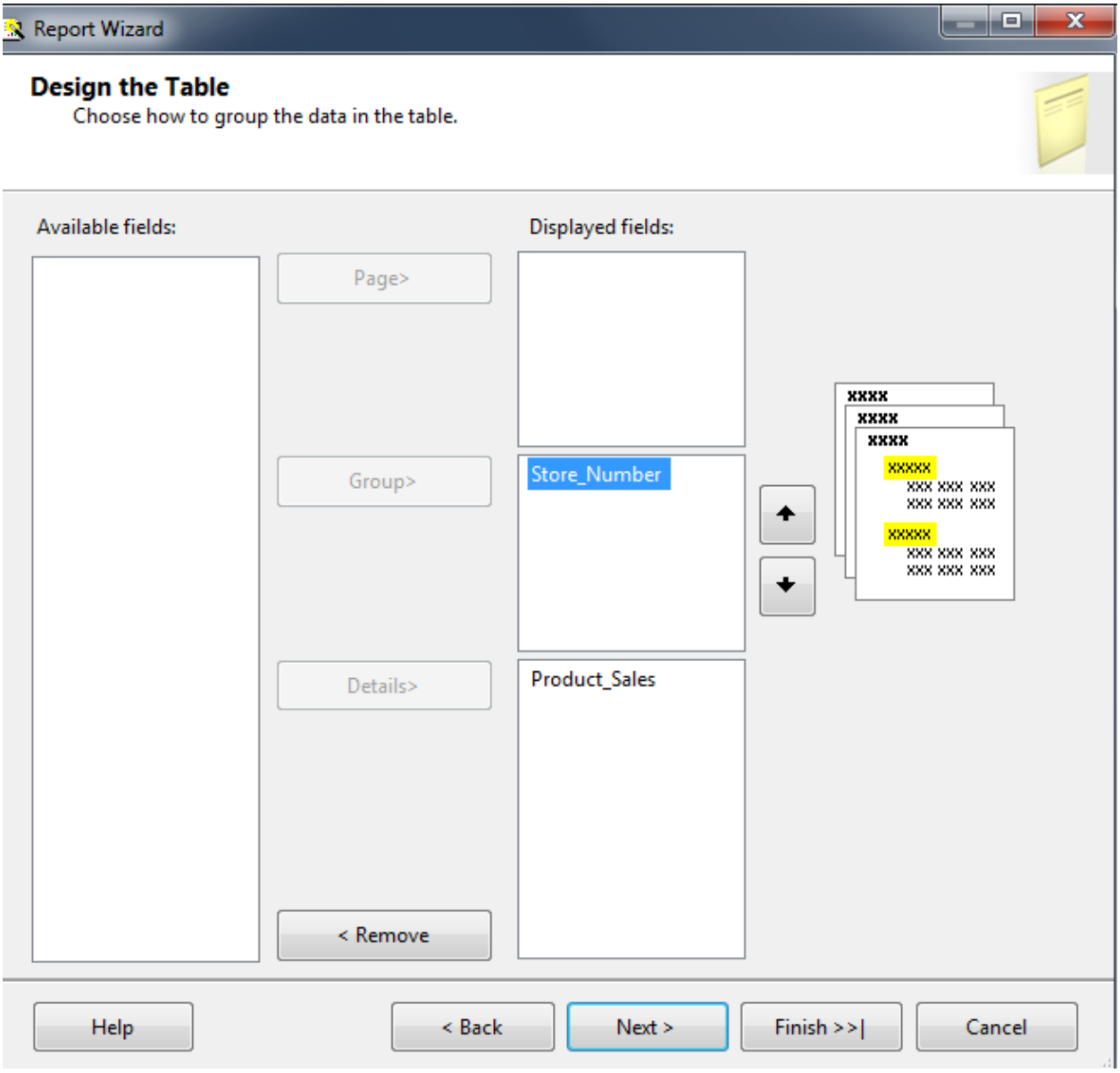

Report Wizard
Select the Data Source
Select a data source from which to obtain data for this report or create a new data source.
Shared data source
New data source
Name:
DataSource1
Type:
Microsoft SQL Server Analysis Services
Connection string:
Edit...
Credentials...
Make this a shared data source
Help
< Back
Next >
Finish >>|
Cancel

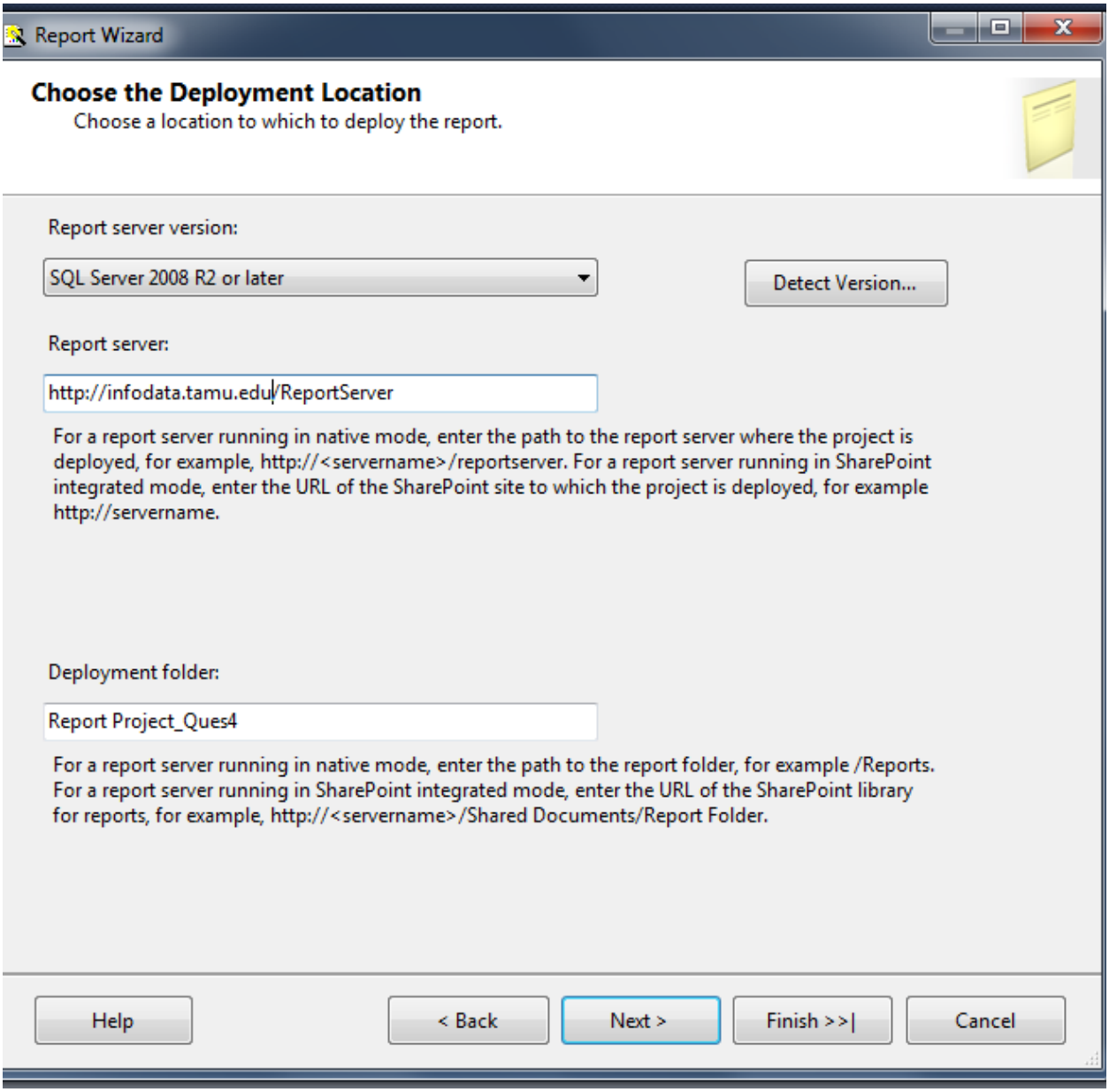

Report Wizard
Select the Data Source
Select a data source from which to obtain data for this report or create a new data source.
Shared data source
New data source
Name:
DataSource1
Type:
Microsoft SQL Serv
Connection string:
Make this a sha
Edit...
edentials...
Help
< Back
Next >
Finish >>|
Cancel
Connection Properties
Data source:
Microsoft SQL Server Analysis Services (Ac
Change...
Server name:
infodata
Log on to the server
User name:
Password:
Save my password
Connect to a database
Select or enter a database name:
prags2
Advanced...
Test Connection
OK
Cancel

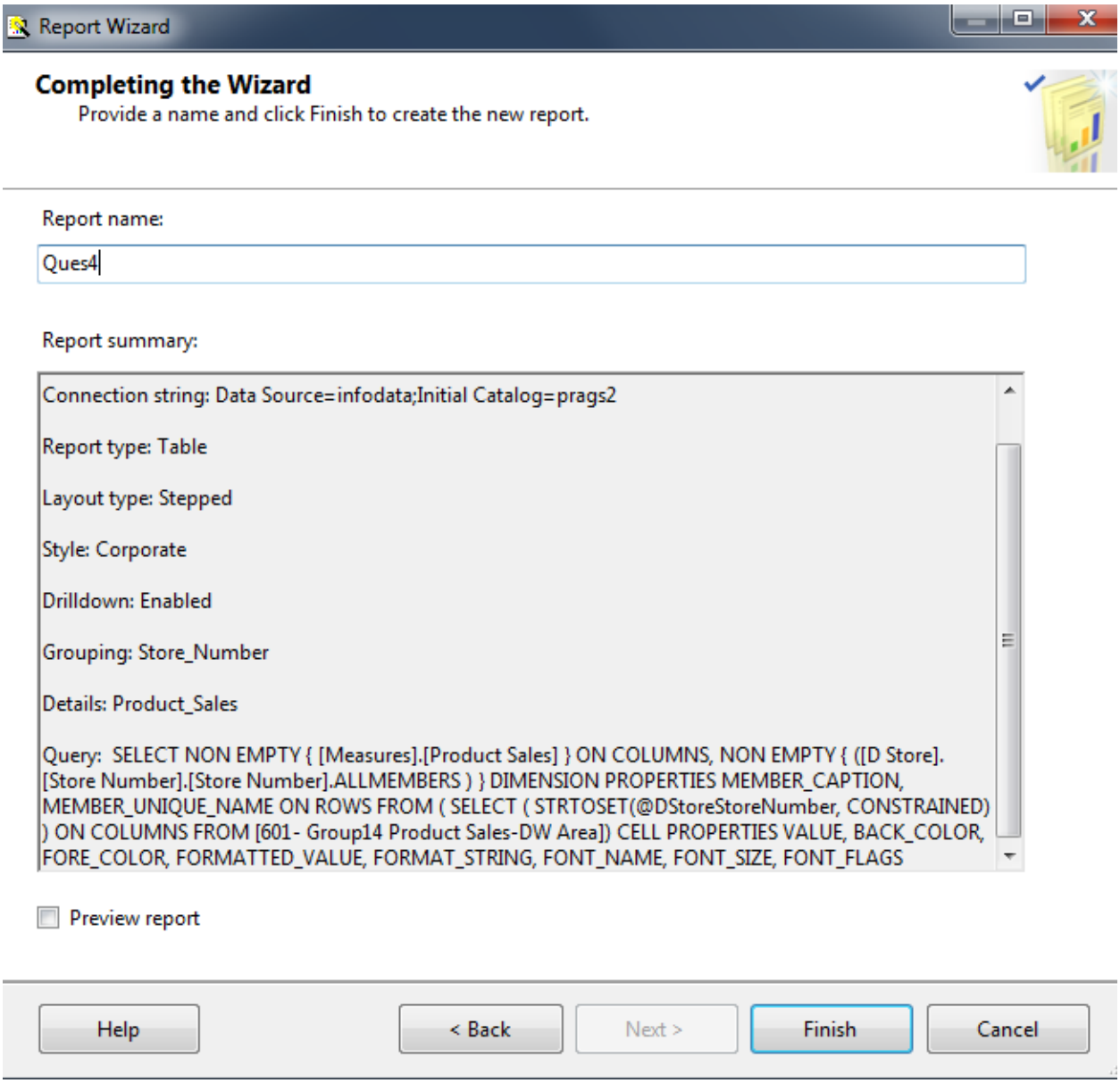

Query Designer
Edit as Text
Import...
601- Group14 Product Sales-DW Area
Metadata
Measure Group:
<All>
601- Group14 Product Sales-DW Area
Measures
KPIs
D Product
D Store
Calculated Members
Dimension
Hierarchy
Operator
Filter Expression
Par
<Select dimension>
Drag levels or measures here to add to the query.
Help
OK
Cancel


Query Designer
Edit as Text
Import...
601- Group14 Product Sales-DW Area
Metadata
Measure Group:
<All>
Product ID
Product Name
UPC Number
Product
D Store
Age Nine
EDUC
Store City
Store ID
Store Name
Store Number
Members
Store Number
Store
Calculated Members
Dimension
Hierarchy
Operator
Filter Expression
Parameters
D Store
Store Number
Equal
{ 21, 77, 100, 103, 106, 110, 115, 129, 134 }
<Select dimension>
Store Number
Product Sales
100
1698800.43
103
1519175.22
106
1354830.16
110
1980755.245
115
2175785.5883...
129
2117638.5283...
134
1418784.3383...
21
2066208.9616...
77
2209014.1933...
Help
OK
Cancel


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 NON EMPTY { [Measures].[Product Sales] } ON COLUMNS, NON EMPTY { ([D Store].[Store Number].[Store Number].ALLMEMBERS) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM (SELECT (STRTOSET(@DStoreStoreNumber, CONSTRAINED)) ON COLUMNS FROM [601-Group14 Product Sales-DW Area]) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
Help
< Back
Next >
Finish >>|
Cancel


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


Report Wizard
Choose the Deployment Location
Choose a location to which to deploy the report.
Report server version:
SQL Server 2008 R2 or later
Detect Version...
Report server:
http://infodata.tamu.edu/ReportServer
For a report server running in native mode, enter the path to the report server where the project is deployed, for example, http://<servername>/reportserver. For a report server running in SharePoint integrated mode, enter the URL of the SharePoint site to which the project is deployed, for example http://servername.
Deployment folder:
Report Project_Ques4
For a report server running in native mode, enter the path to the report folder, for example /Reports. For a report server running in SharePoint integrated mode, enter the URL of the SharePoint library for reports, for example, http://<servername>/Shared Documents/Report Folder.
Help
< Back
Next >
Finish >>|
Cancel



Ques4

Store Number	Product Sales
[Store_Number]	
	[Product_Sales]
	[Sum(Product_Sales)]

Ques4

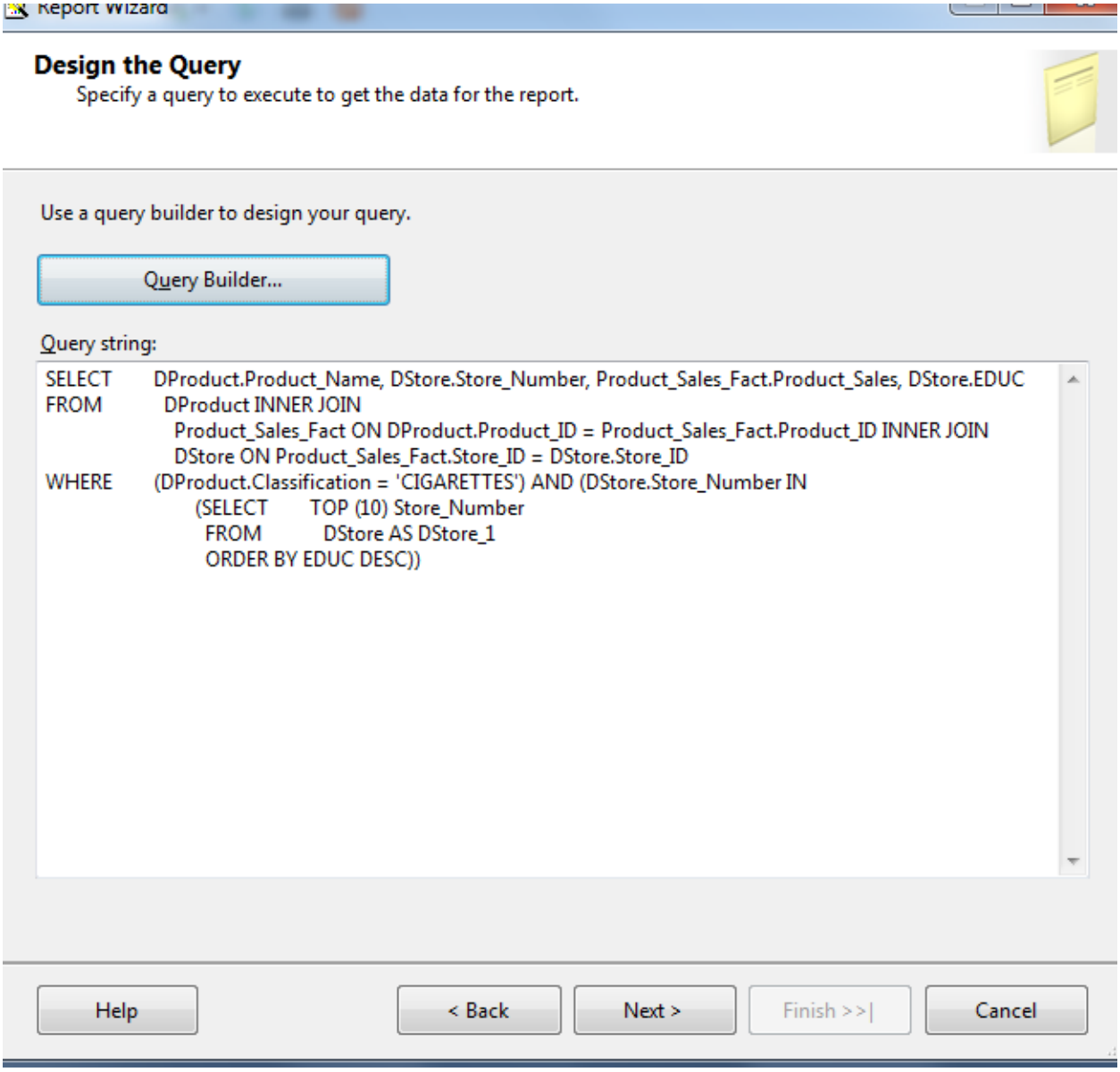
Store Number	Product Sales
⊞ 100	
	1698800.43
⊞ 103	
	1519175.22
⊞ 106	
	1354830.16
⊞ 110	
	1980755.245
⊞ 115	
	2175785.58833334
⊞ 129	
	2117638.52833333
⊞ 134	
	1418784.33833333
⊞ 21	
	2066208.96166666
⊞ 77	
	2209014.19333333

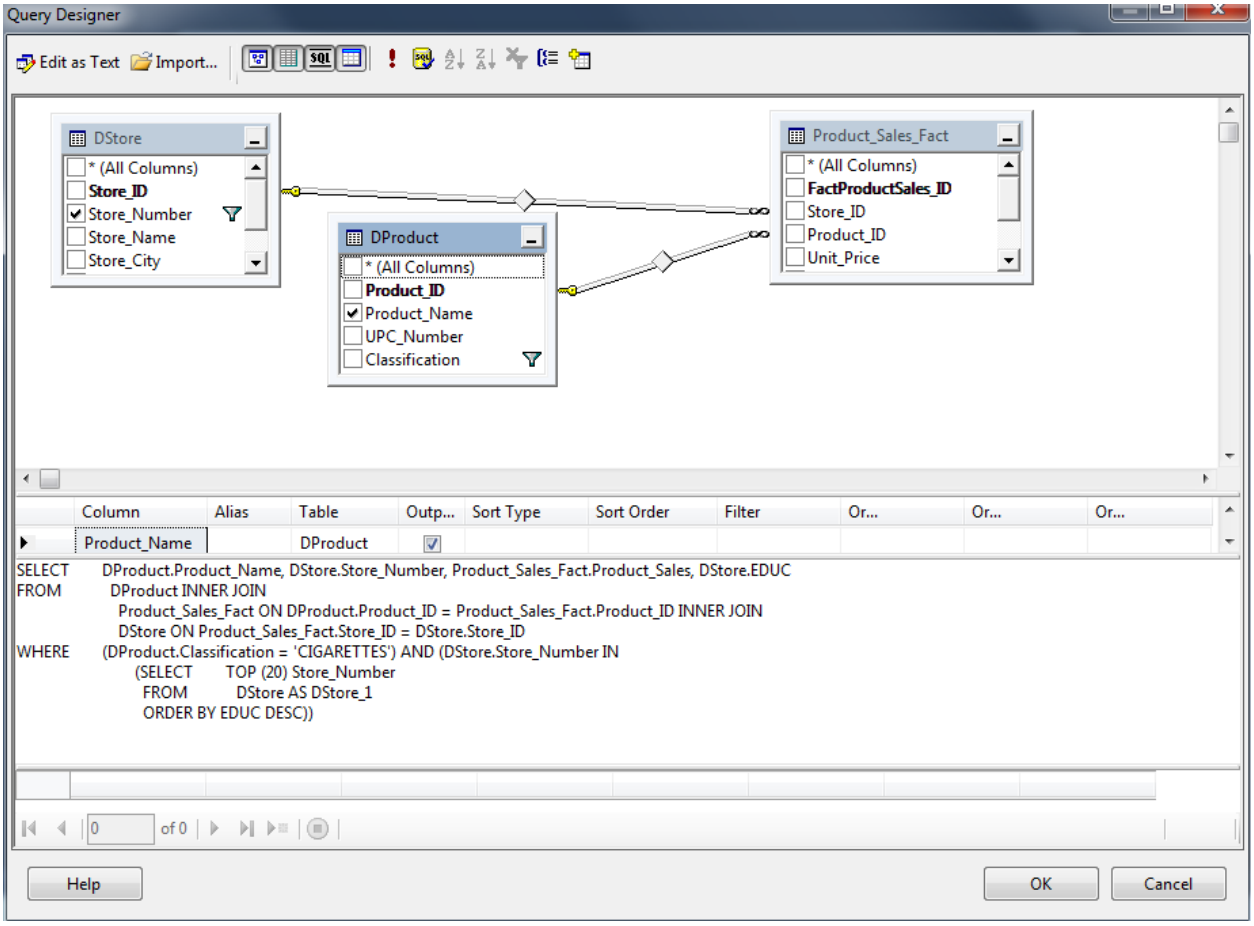
Analysis and Conclusion -

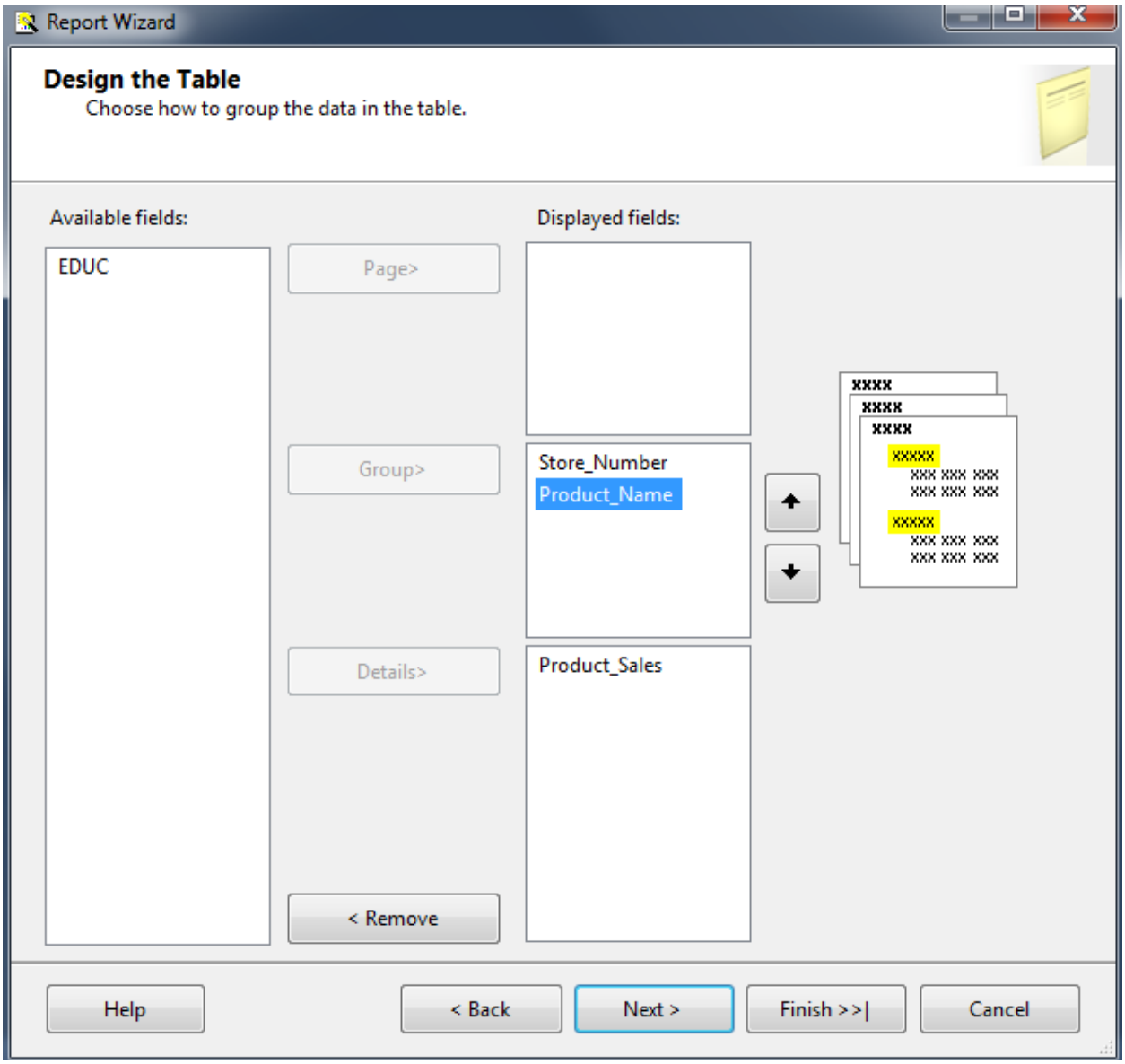
To determine the trend in the sale of cookies (from Week 1-399) in stores with the highest population of kids under age 9, we generated an SSRS report on top of SSAS, after creating a cube in the analytics services of SQL Server data tools. With the concept of youth marketing in mind and/or targeted marketing strategy, we wanted to analyze the sales trend for this targeted demographic, so we conducted the analysis using cookies which is a very popular product with this demographic. We further drilled down to examine only the stores with the highest population of children under age 9 and found that linear sales across these stores showed an upward going (increasing), thus confirming our earlier predictions.

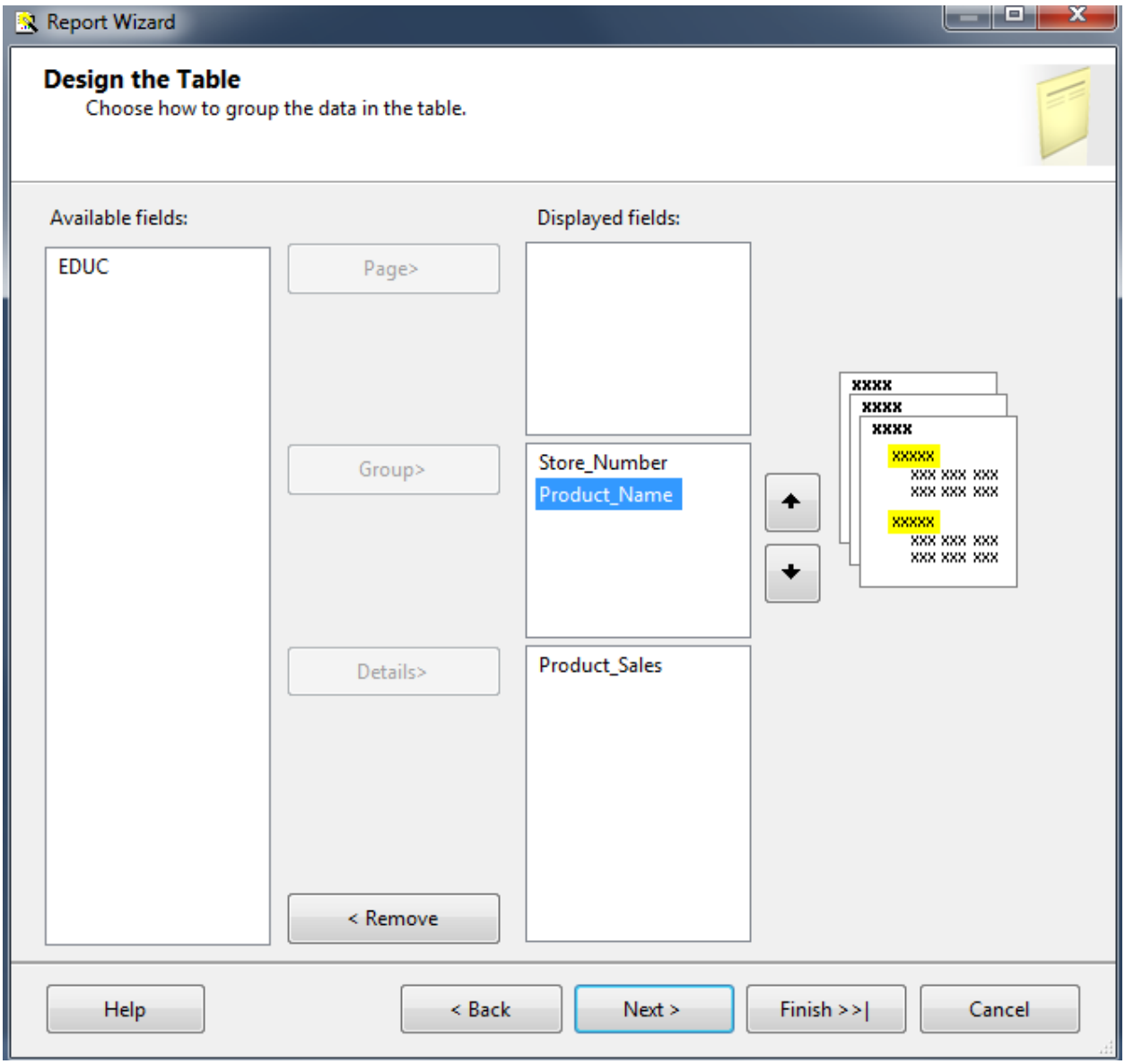
3.5 Report Building Using SSRS with ReportBuilder

Question 5 - What is the trend in the sale of Cigarettes in stores with the highest population of College Graduates?




Query Designer
Edit as Text
Import...
DStore
* (All Columns)
Store_ID
Store_Number
Store_Name
Store_City
DProduct
* (All Columns)
Product_ID
Product_Name
UPC_Number
Classification
Product_Sales_Fact
* (All Columns)
FactProductSales_ID
Store_ID
Product_ID
Unit_Price
Column
Alias
Table
Outp...
Sort Type
Sort Order
Filter
Or...
Or...
Or...
Product_Name
DProduct
SELECT DProduct.Product_Name, DStore.Store_Number, Product_Sales_Fact.Product_Sales, DStore.EDUC
FROM DProduct INNER JOIN
Product_Sales_Fact ON DProduct.Product_ID = Product_Sales_Fact.Product_ID INNER JOIN
DStore ON Product_Sales_Fact.Store_ID = DStore.Store_ID
WHERE (DProduct.Classification = 'CIGARETTES') AND (DStore.Store_Number IN
(SELECT TOP (20) Store_Number
FROM DStore AS DStore_1
ORDER BY EDUC DESC))
0
of 0
Help
OK
Cancel


Report Wizard
Design the Table
Choose how to group the data in the table.
Available fields:
EDUC
Page>
Group>
Details>
< Remove
Displayed fields:
Store_Number
Product_Name
Product_Sales
xxxx
xxxx
xxxx
xxxxx
xxx xxx xxx
xxx xxx xxx
xxxxx
xxx xxx xxx
xxx xxx xxx
Help
< Back
Next >
Finish >>|
Cancel

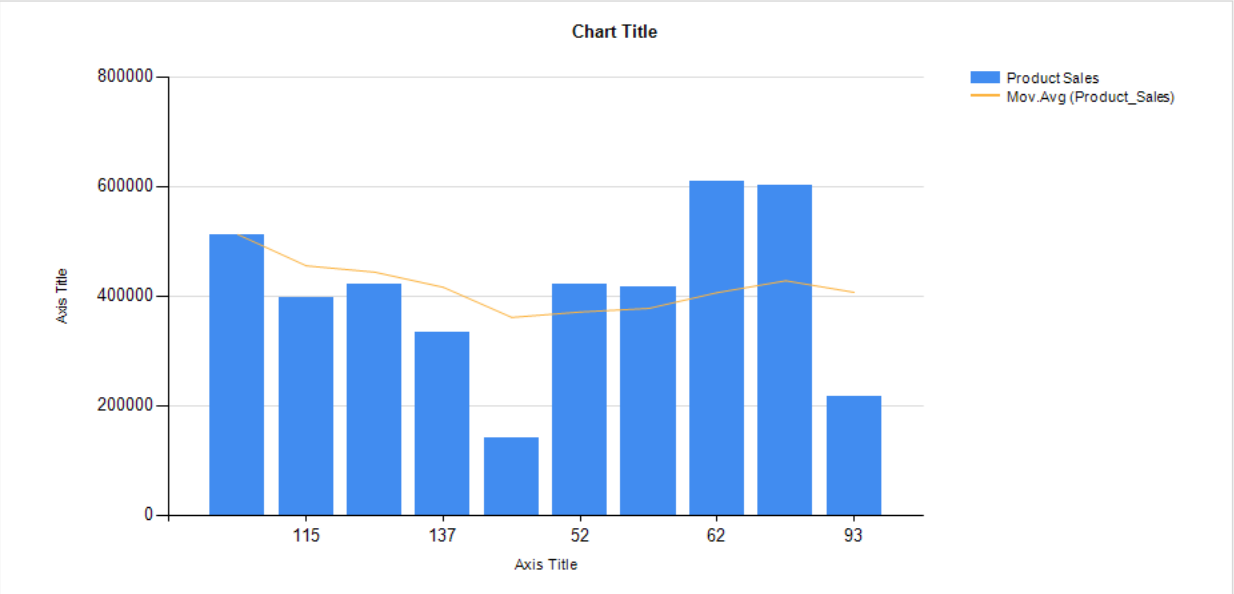


Question5

Store Number	Product Name	Product Sales
[Store_Number]		
	[Product_Name]	
		[Product_Sales]
		[Sum(Product_Sales)]

Question5

Store Number	Product Name	Product Sales
⊞ 109		
⊞ 115		
⊞ 126		
⊞ 137		
⊞ 33		
⊞ 52		
⊟ 54		
	⊞ 100'S CARTON	
		274.57
	⊞ 100'S CARTON CIGARET	
		17285.05
	⊞ 100'S CIGARETTE CART	
		90123.7200000002
	⊞ 25CT SINGLE PACK	
		41.6
	⊞ B & H SPECIAL $4 OFF	
		0
	⊞ B & H SPECIAL SNGLE	
		0
	⊞ BASIC DISCOUNT CRTN	



Analysis and Conclusion -

The ReportBuilder chart above shows the product sales for cigarettes for the top ten(10) stores with the highest population of college graduates. A quick analysis tells us that while Store 62 records the highest sales at a little over $600,000; Stores 93 and 33 record contrastingly lower sales at about $200,000 and $130,000 respectively. The insight provided by this information will typically give rise to such strategic questions as: How can product sales be increased in Stores 93 and 33? What business strategies can we implement to increase and maintain the average sales for all these stores at a minimum of, say $500,000? Thus, DFF can make informed business decisions on what strategies they can employ to increase their sales and profit margin.

Make sure to tell where the DW, staging, reports, and analysis services, etc. are stored. PLEASE PROVIDE THE USERID TO LOGIN AND THE PASSWORD so that I can test. This should be listed separately in a page next to the Title page.

References

1. https://en.wikipedia.org/wiki/Dominick's
2. Lira, Loreto, "Why Do Some Prices in the Retail Sector Drop When Demand Rises? Evidence from the Chilean Case," *Revista ABANTE* 10 (Octubre 2007): 151-168.
3. Georg Muller, Mark Bergen, Shantanu Dutta, and Daniel Levy, "Private Label Price Rigidity During Holiday Periods," *Applied Economics Letters*, 13 (2006): 57-62.
4. Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data By JUDITH A. CHEVALIER, ANIL K KASHYAP, AND PETER E. ROSSI*
5. Who Benefits from Price Promotions? Price cuts boost sales and traffic. So where are the long-term gains? - *By Shuba Srinivasan, Koen Pauwels, Dominique Honssens, and Marnik Dekimpe*
6. The Value of Purchase History Data in Target Marketing - *By Peter E. Rossi, Robert E. McCulloch and Greg M. Allenby. Source: Marketing Science, Vol. 15, No. 4, (1996), pp. 321-340*
7. http://www.huffingtonpost.com/2011/06/14/top-7-beer-holidays_n_876548.html
8. https://wineeconomist.files.wordpress.com/2011/12/nielsen.jpg
9. https://blog.adroll.com/trends/holiday-retail-sales-growth-in-the-us-explained-in-5-graphs
10. https://en.wikipedia.org/wiki/Youth_marketing#cite_note-51
11. http://www.kidclubmarketing.com/fastfacts.aspx
    http://www.bizjournals.com/bizjournals/how-to/marketing/2015/04/stats-to-help-marketers-reach-young-adults.html
12. https://economics.missouri.edu/working-papers/2007/wp0712_basker.pdf
13. https://www.forbes.com/sites/lauraheller/2014/01/31/this-is-what-a-failed-supermarket-looks-like/#7148d44c4866
14. http://www.livestrong.com/blog/scary-truth-about-halloween-infographic/#ixzz4a7L2Shrr
15. www.stat.cmu.edu/~kass/montgomery/written/sec4.html
16. http://essayonrequest.com/an-analysis-of-the-working-women-and-the-buying-of/
17. http://www.refrigeratedfrozenfood.com/articles/88837-frozen-and-prepared-foods-trends-in-2015
18. http://www.us.jll.com/united-states/en-us/Research/cold-storage-investment-perspectives-jll-nov2014.pdf