5/1/2018

Consulting Project: Report 4

Design and Implementation of a Data Warehouse for a Retail Store

Group# 6
Anushree Tak, Pranjal Gururani, Sumeer Angra
Primary Email: anushree_tak@tamu.edu
TEXAS A&M UNIVERSITY AT COLLEGE STATION

Credentials for access to Data warehouse, Reporting and Analysis Services and Report Server

SQL Server Authentication details are as below:

Server name:	**infodata16.mbs.tamu.edu**
User id:	**gu5127**
Password:	**Mays5127**

The above credentials will provide a user access to the following:

1) Database Engine:

Staging area:	**602Group6_stagingDB**
Data warehouse area:	**602Group6_DW_Area**

2) SSAS:

Cubes deployed for SSAS:	
Question 2	**602Group6_HyperCube_Q2**
Question 5	**602Group6_HyperCube_Q5**

3) Report Server: http://infodata16.mbs.tamu.edu/reportserver

Project folder:	**602_Group6_ProjectReport**

Contents

Table of Tables

Table of Figures

1 Introduction

Data warehouse is a collection of subject-specific and time-variant data that can be used for making important business decisions. It aims at providing data for reporting and analysis purposes and helps with planning, problem solving, and decision support. Data warehouse gets its data from OLTP databases. It contains data from disparate sources which is both current and historic in nature and is used for strategic decision making. The stored data reveals multi-dimensional view of business activities.

Data warehousing is the process of designing data warehouse. It includes fetching data from different sources and storing it in the Staging area. Data in the staging area is cleaned using various methods to remove the data inconsistencies. Clean data is then loaded into the data warehouse and can be then used for reporting and analysis.

The data warehouse being designed for this project is for Dominick's Finer Food (DFF). DFF is a retail business in Chicago area which deals with the sale of grocery. At one point, DFF had a chain of 99 stores spread across Chicago and suburbs, northwest Indiana and Wisconsin. It was first set up in 1918 and had to shut down by 2014 due to continuous fall in revenue. Strategic analysis is important to understand the lack of sales for DFF and the overall poor performance.

As a part of this project, a data warehouse will be designed for DFF to help integrate data across stores and improve its sales, revenue, profit and customer base. The data being used for this project was generated by partnership between Chicago Booth and DFF for a store-level research. The data will be carefully studied, cleaned and stored in the data warehouse to perform analytical reports and solve business questions that could help DFF improve sales, efficiency and customer base.

1.1 Challenges Faced

Every project has to deal with a lot of challenges during its lifetime. The challenges faced during the implementation of this project are:

1. The data was huge, so team had to spend a lot of time in understanding the data and the relationship between various data files.
2. The data was dirty and had a lot of missing values and inconsistencies making it difficult to analyze and draw conclusions. For example, in CCount data files many values are negative for Week column.
3. The file format for data was not uniform. The data files are in csv, txt or HTML format so it took some time to convert data in Excel format for analysis.
4. Deciding on the Business Questions and prioritizing them was challenging because it required in depth understanding of the business and its problems.

1.2 Understanding of the Data

The partnership between Chicago Booth and DFF for research between 1989 and 1994, resulted in store-level scanner data. Spread over 100 stores, the data takes into account 3500 products (UPCs) which belong to 29 different categories.
The Dominick's database contains 2 types of files: files containing details of a particular category, and files containing general details. Details about the data files are as follows:

a. Customer Count Files
   - The customer count files consist of customer counts for each day for every store. It has the data for customer visits and purchase.
   - This file contains sales data for different products and also coupon redemption per day per store.

b. Store-specific Demographics
   - The Demographic file contains demographic data related to each store. The U.S. government census data for Chicago area was processed by Market Metrics to create demographic data specific to DFF store.
   - This data provides information about percentage of people in various age groups, ethnicity, sex, shopping style, employment, income, household size, etc. It will help in understanding the customer base and deciding the target customer.

c. UPC Files
   - This data file has information regarding the UPC associated with different categories.
   - This contains data about Product Name, size, number of items in a case.

d. Movement Files
   - These are category-specific files which contain sales data for each UPC for every store
   - The granularity is week level. It contains store number, week number, items sold that week, price, gross margin, etc.

e. Week's decode table
   - It is available in Dominick's Manual and Codebook pdf. It gives the decode description of week number by including week's start date and end date and also contains information about special events. It could help in understanding the sales patterns during special events or particular part of the year.

f. Dominick Stores
   - It is available in Dominick's Manual and Codebook pdf. It gives the details corresponding to each store number like city, address, zone, zip code and price tier.

1.3 Metadata

1.3.1 Sales Data

Variable	Description	Type	Length
STORE	Store Code	Numeric	3
DATE	Date of the Observation	Character	6

WEEK	Week Number	Numeric	6
MEAT	Meat Sales in Dollars	Numeric	8
BEER	Beer Sales in Dollars	Numeric	8
GROCERY	Grocery Sales in Dollars	Numeric	8
DELI	Deli Sales in Dollars	Numeric	8
DELI_COUP	Deli Coupons Redeemed	Numeric	8
FROZEN	Frozen Sales in Dollars	Numeric	8
DAIRY	Dairy Sales in Dollars	Numeric	8

Table 1.3.1: Sales Data

1.3.2 Demographics

Variable	**Description**	**Type**	**Length**
MMID	Mobile Money Identification Number	Numeric	8
NAME	Name of Store	Character	16
CITY	Name of City	Character	16
ZIP	Zip Code of Area	Character	8
INCOME	Log of Median Income	Numeric	8
STORE	Store Number	Numeric	3
AGE60	% of population over 60	Numeric	8
POVERTY	% of population n with income under $15,000	Numeric	8
AGE9	% of population under 9	Numeric	8

Table 1.3.2: Demographics

1.3.3 UPC

Variable	Description	Type	Length
UPC	UPC Number	Numeric	8
COM_CODE	Dominick's commodity code	Numeric	8
NITEM	Dominick's item code	Numeric	8
DESCRIP	Product Name	Character	20
SIZE	Product Size	Character	6
CASE	Number of items in a case	Numeric	8

Table 1.3.3: UPC

1.3.4 Movement

Variable	Description	Type	Length
UPC	UPC Number	Numeric	8
STORE	Store Number	Numeric	3
WEEK	Week Number	Numeric	3
MOVE	Number of Units Sold	Numeric	8
PRICE	Retail Price	Numeric	8
QTY	Number of items bundled together	Numeric	3
PROFIT	Gross Margin	Numeric	8
SALE	Sale Code (B, C, S)	Numeric	8
OK	1 for valid data, 0 for trash	Numeric	3

Table 1.3.4: Movement

1.3.5 Movement Files to Category Mapping

Category	UPC	Movement
Analgesics	upcana	wana
Bath Soap	upcbat	wbat
Beer	upcber	wber
Bottled Juice	upcbjc	wbjc
Canned Soup	upccso	wcso
Canned Tuna	upctna	wtna
Cereals	upccer	wcer
Cigarettes	upccig	wcig
Cheese	upcche	wche
Cookies	upccoo	wcoo
Crackers	upccra	wcra
Dish Detergent	upcdid	wdid
Fabric Softener	upcfsf	wfsf
Front-End-Candies	upcfec	wfec
Frozen Dinners	upcfrd	wfrd
Frozen Entrees	upcfre	wfre
Frozen Juices	upcfrj	wfrj
Grooming Products	upcgro	wgro
Laundry Detergents	upclnd	wlnd

Paper Towels	upcptw	wptw
Refrigerated Juices	upcrfj	wrfj
Soap	upcsoa	wsoa
Soft Drinks	upcsdr	wsdr
Shampoos	upcsha	wsha
Snack Crackers	upcsna	wsna
Toothbrushes	upctbr	wtbr
Toothpastes	upctpa	wtpa

Table 1.3.5: Movement Files to Category Mapping

1.3.6 Store Data

Store	City	Price Tier	Zone	Zip Code	Address
2	River Forest	High	1	60305	7501 W. North Ave.
4	Park Ridge	Medium	2	60068	Closed
5	Palatine	Medium	2	60067	223 Northwest HWY.
8	Oak Lawn	Low	5	60435	8700 S. Cicero Ave.
9	Morton Grove	Medium	2	60053	6931 Dempster
12	Chicago	High	7	60660	6009 N. Broadway Ave.

Table 1.3.6: Store Data

1.3.7 Week Decode

Week	Start	End	Special
1	09/14/89	09/20/8	
2	09/21/89	09/27/8	
3	09/28/89	10/04/8	
4	10/05/89	10/11/8	
5	10/12/89	10/18/8	
6	10/19/89	10/25/8	
7	10/26/89	11/01/8	Halloween
8	11/02/89	11/08/8	
9	11/09/89	11/15/8	
10	11/16/89	11/22/8	
11	11/23/89	11/29/8	Thanksgiving
12	11/30/89	12/06/8	
13	12/07/89	12/13/8	
14	12/14/89	12/20/8	
15	12/21/89	12/27/8	Christmas
16	12/28/89	01/03/9	New-Year
17	01/04/90	01/10/9	
18	01/11/90	01/17/9	
19	01/18/90	01/24/9	
20	01/25/90	01/31/9	
21	02/01/90	02/07/9	
22	02/08/90	02/14/9	
23	02/15/90	02/21/9	Presidents
24	02/22/90	02/28/9	

Table 1.3.7: Week Decode

1.4 Domain Understanding

a) **Hamister, J. W., & Suresh, N. C. (2008). The impact of pricing policy on sales variability in a supermarket retail context. International Journal of Production Economics, 111(2), 441-455.doi:10.1016/j.ijpe.2007.01.011**

This paper focuses on how the pricing policy impacts the sales in retail. This study states that using constant price and not dynamic would reduce profit and increase demand volatility. This research uses the data from Dominick to test the hypothesis.
Two important hypothesis that are tested and supported in this study are:

- Demand variability is lower with dynamic pricing
- Profitability is higher with dynamic pricing than it is with static

Learning
We will use this study to check and understand how the sales and profit margin of each cheese category has been affected as per their price during the period 1991-1995. These supported hypotheses indicate that dynamic pricing should be followed by retailers to increase profit and reduce demand variability. Under situation of serial correlation, constant pricing may reduce the profit which will be a significant reduction for retailers in highly competitive market. We will incorporate this finding while analyzing cheese sales vs price fluctuation across cheese products.

b) **Kamakura, Wagner A., and Wooseong Kang. "Chain-Wide and store-Level analysis for cross-Category management." Journal of Retailing, vol. 83, no. 2, 2007, pp. 159–170., doi:10.1016/j.jretai.2006.02.006.**

This paper helps the retail managers to make decisions which are chain-wide and store-specific. It discusses the impact of price promotions of specific products across the organization. The paper analyzes the effect of discounts at chain level across global organization. It discusses how the price discounts have store-level and cross-category impacts.

Learning
We will use the parameters mentioned in this case study to check for profit fluctuation in cheese. We will see if a low priced cheese product sells in high volume or a higher one. Hence we will be able to understand optimal level of pricing for maximizing profit. This study analyzes Dominick data using the sales and prices of nine brands of toothpaste and 8 brands of toothbrush in 66 stores to investigate store-level and cross- category promotion. These two categories are chosen because of their close connection in consumption. The results indicate that price promotions in toothpaste can have positive and negative effects on various toothpaste brands. This study helps in realizing the effects price promotion has on the sale of brands across stores and categories.

c) **Nevo, Aviv, and Catherine Wolfram. "Why Do Manufacturers Issue Coupons? An Empirical Analysis of Breakfast Cereals." The RAND Journal of Economics, vol. 33, no. 2, 2002, p. 319., doi:10.2307/3087436.**

This paper analyzes the impact of coupons on sales. Manufacturers issue coupons to lure the low-valuation customers who tend to postpone their purchases. New buyers are attracted by issuing coupons and are tested if they would repurchase the product at full price. The cheapest way for sellers to increase sales, for example at the end of company fiscal year to reach the target, is to provide discounts to everyone but higher price promotions by providing coupons to a few. This increases the sales without much decrease in the profit ratio.

Learning

We will check for the high sales periods for Beer, Grocery, Bakery and video. This will help us figure out when the sales of each of these categories is the highest. This will then be correlated with the pricing of these products during those periods. For Dominick data, on comparing the sales of products during promotional and non-promotional weeks it was found that coupons have different impacts on sale for different products. Offers during some holiday like President's Day and Labor Day did not have much impact on the sales as compared to other special events.

2 Business Questions

2.1 Business Questions List

The business questions listed in order of priority:

a) How are the store sales affected by the poverty level in the regions?
b) How are the sales of Bakery products affected during Christmas period over the period 1989 -1996?
c) Are there any periods during the years 1989-1995 where the video sales are higher or lower?
d) How are jewelry sales affected by the store price tier?
e) From year 1991 to 1995, which cigarette brand has the highest profit?
f) Based on different zones, how much is consumption of non-vegan products like fish, meat and dairy across Chicago?
g) What is the profit trend of cheese during 1991 to 1995?
h) How the trend of grocery sales varies across different stores in Chicago?
i) What are the top cookies that have highest profit in all stores combined?
j) Detect trends in Beer Sales encompassing the entire duration that look for high period of sales. Detect peaks not only during designated holidays but also on events that are not in the Week_Decode table.

2.2 Business Questions Substantiation and Explanation

Q1. How are the store sales affected by the poverty level in the regions?

Data Processing

- All Sales in Ccount table were summed up for a store to get the total sales
- Poverty percentage from Demographics table is matched to Ccount table using the STORE attribute
- Data grouped by percentage intervals(each interval of 3%) for better visualization

Resulting Pivot

Poverty	Sum of Sales
0.03-0.06	$7,066,977,014
0-0.03	$6,260,286,283
0.06-0.09	$3,867,227,647
0.15-0.18	$1,723,796,450
0.09-0.12	$698,426,296
0.12-0.15	$422,924,831
0.21-0.24	$169,266,615
0.18-0.21	$163,159,911
Grand Total	**$20,372,065,047**

Figure 2.2-1: Pivot Table

Resulting Graph

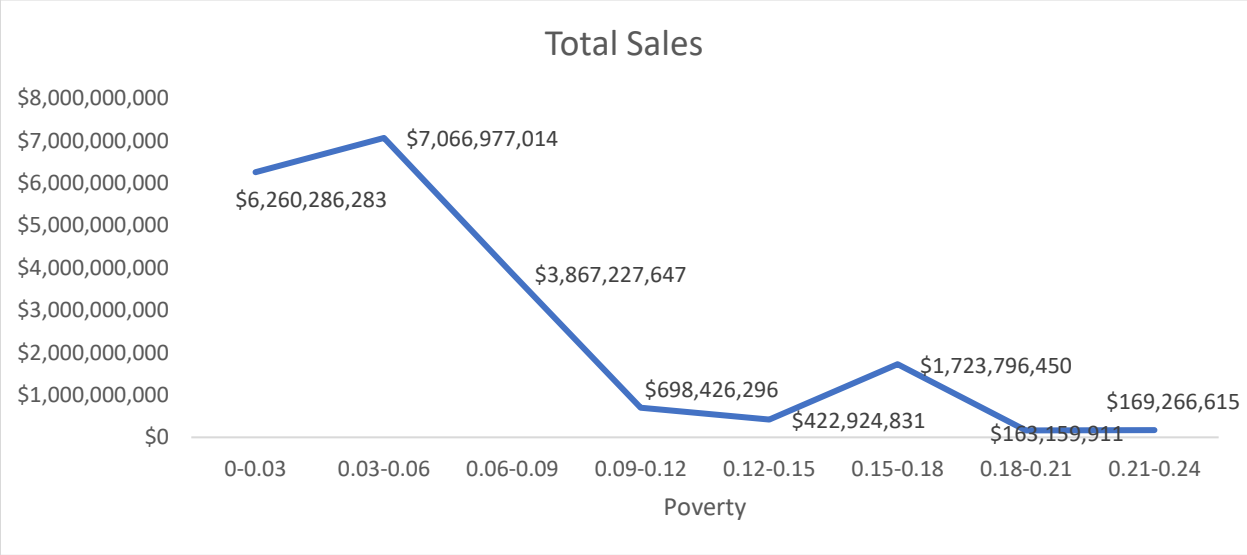

Figure 2.2-2: Pivot Chart

ERD

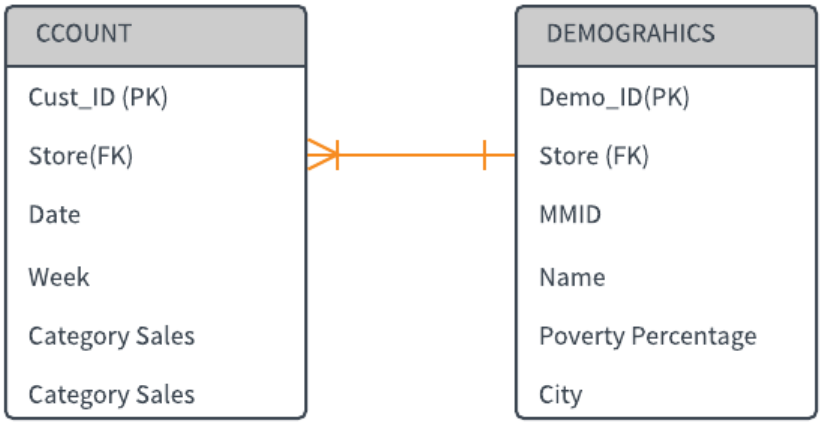

Figure 2.2-3: ERD

Justification

This data can be used to understand how opening a store in a poverty stricken area will impact the business revenue growth. Discovering this trend through current data will make it possible to estimate increase in topline of the business when a store in opened in a new area. Business heads can align opening of new stores as per their commercial targets. For example, if the new store needs to generate high revenue it must be opened in a region with low poverty percentage. Hence an area can be mapped to desired levels of sales estimations. Another application of this pattern is to check if all stores meet a minimum level of sales generation in order to meet the operational cost required to make them profitable. Any stores below this minimum levels must be closed.

Q2. How are the sales of Bakery products affected during Christmas period over the period 1989 -1996?

Data Processing

- Christmas week for each of the year(‘89-‘96) was identified from the Week_Decode table
- The week was then mapped to the bakery product in the CCount table
- Sales variance is then calculated for each of the day of the Christmas week

Resulting Pivot

Bakery Sales		
WEEK	DATE	Total
15	891221	$139,171
	891222	$198,816
	891223	$289,826
	891224	$224,360
	891225	$0
	891226	$79,063
	891227	$85,595
67	901220	$162,995
	901221	$184,009
	901222	$231,678
	901223	$227,401
	901224	$280,865
	901225	$18,382
	901226	$85,078
119	911219	$155,673
	911220	$154,954
	911221	$221,315
	911222	$174,649
	911223	$236,052
	911224	$297,513
	911225	$24,511

Figure 2.2-4 Pivot Table

Resulting Graph

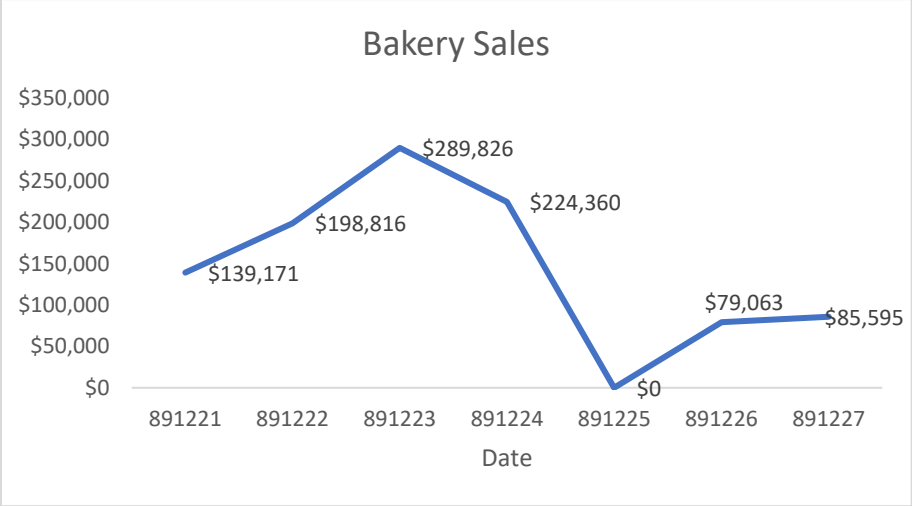

Figure 2.2-5: Pivot Chart

ERD

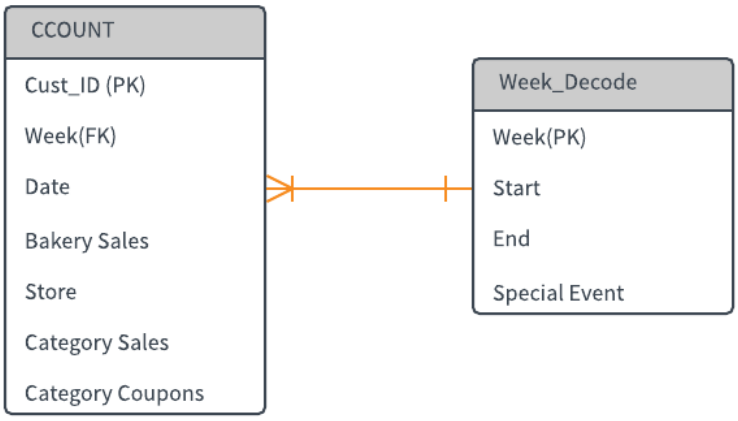

Figure 2.2-6: ERD

Justification

It is critical for the Retail chain to exactly know the period when demand relating to Christmas activities is at its peak. Christmas is the time when consumers are on a spending spree. For example, there is no point in procuring more inventory after Peak Sales Period has passed. This question answers what "Peak Sales Period" is. It will allow store managers to predict when they will experience peak demand. They can hence prepare their inventories to meet this demand. Replenishing inventory always involves a certain amount of lead time after an order has been placed to when it is delivered to store. Hence store manager can stock up their inventories using the pattern discovered in this question. For example – as per the graph, peak sales for bakery products happened two days before Christmas. No sales for Bakery took on Christmas day. Correspondingly, store managers need to place purchase orders to meet demand experienced two days before Christmas and no orders are needed for the Christmas day itself. This pattern if repeats every year forms a strong basis for ordering. Our question will answer this aspect. It is to notice that a short period of time before Christmas is only being considered as we are analyzing the sales of bakery products which are *perishable*. If the analysis is done for a *non –perishable* product the time period can range from 2-3 weeks prior to Christmas.

Q3. Are there any periods during the years 1989-1995 where the video sales are higher or lower?

Data Processing

- Video Sales are mapped against their weekly volumes in the Ccount table
- This data will be aggregated on monthly basis to check the trend.

Resulting Pivot

Week	Video Sales
1	102006.92
2	89368.81
3	284169.44
4	130738.06
5	221438.18
6	119813.59
7	93769.21
8	114456.83
9	209826.94
10	563262.31
11	314194.01
12	315516.72
13	108314.95
14	130643.33
15	151202.66

Figure 2.2-7: Pivot Table

Resulting Graph

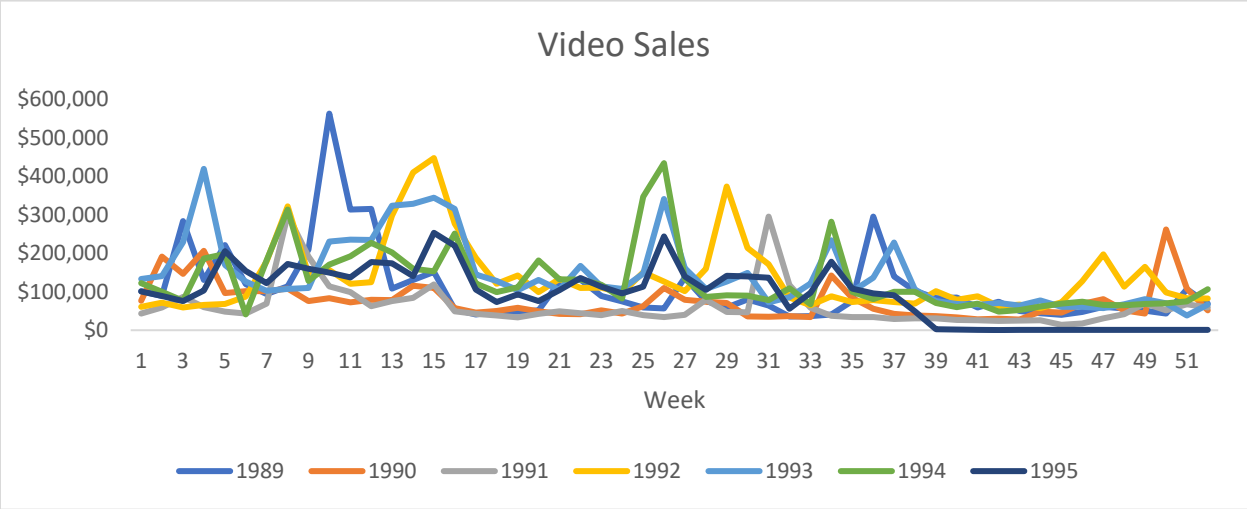

Figure 2.2-8: Pivot Chart

ERD

Figure 2.2-9: ERD

Justification

During the period 1989 to 1995 videos were a novel technology. Video watching was not a regular day to day activity as it is today. This questions analyzes when were these videos were being watched. The question aims to discover if there is any pattern in the way users were consuming videos in this era. This insight is useful to stock videos in a store during high demand period. In the low demand period the shelfs space used for videos can be reduced and be release for day to day items such as groceries, cookies, chocolates etc. Hence this question will help in store shelf management by only stocking videos only during the time of demand. Our preliminary analysis depicts period peaks and troughs. This will be further explored in the project.

Q4. How are jewelry sales affected by the store price tier?

Data Processing

- Store Price tier is selected from Store Table
- Price Tier is mapped to store in Ccount file

Resulting Pivot

Price Tier	Sum of JEWELRY
Medium	$1,384,166
Low	$682,406
High	$642,934
CubFighter	$302,270
Grand Total	**$3,011,777**

Figure 2.2-10: Pivot Table

Resulting Chart

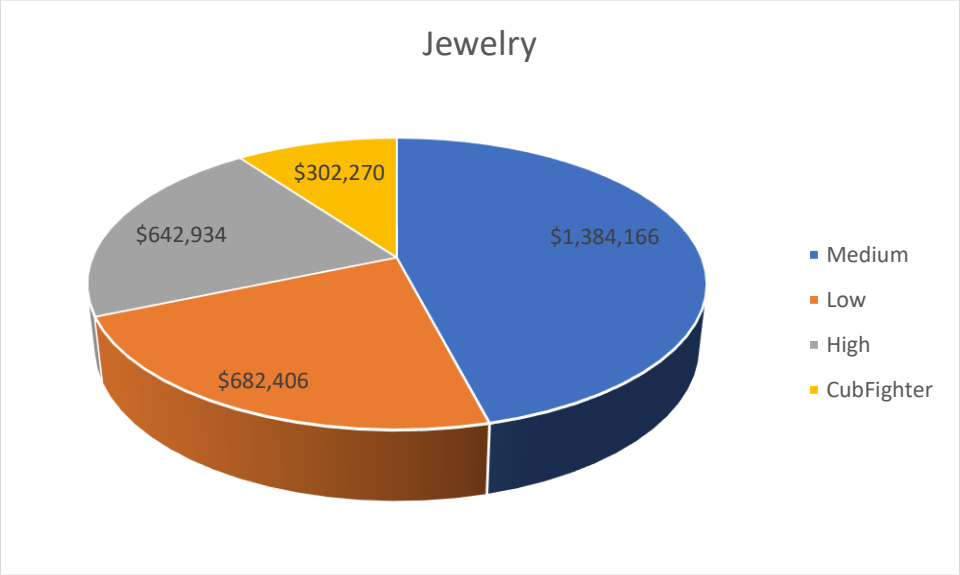

Figure 2.2-11: Pivot Chart

ERD

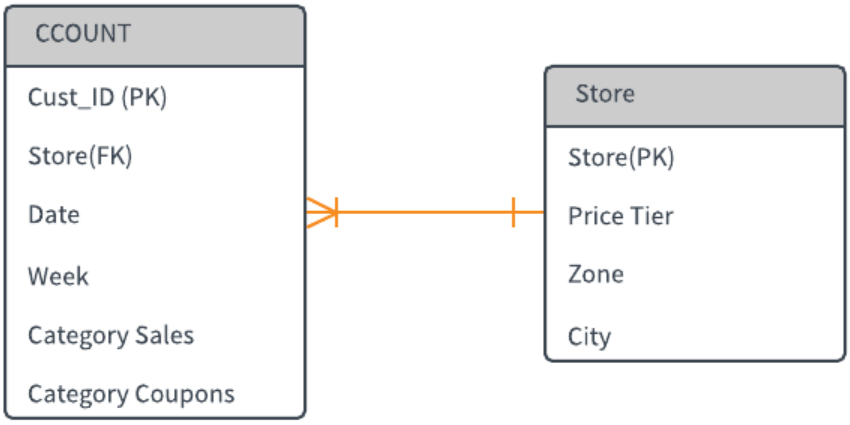

Figure 2.2-12: ERD

Justification

A close study is required for a luxury item such as jewelry because usually it is not the core competency of a retail store. One would intuitively believe that jewelry sales for the retail store would be the most in its highest price tier. But our preliminary analysis suggests that the medium price tier stores are selling more jewelry than the high price tier stores. This question would look at data and analyze which stores should DFF focus on for its jewelry business.

Q5.From year 1991 to 1995, which cigarette brand has the highest profit?

Data Processing

- Profit for each item in the movement file for cigarette was calculated using the formulae ((move*price/ qty)*profit/100)
- A pivot was then calculated to determine what are the most profitable items
- The names of these items were taken from UPC cigarette file through vlookup

Resulting Pivot

UPC	Sum of Final Profit	DESCRIP
197	$16,648,945	CIGARETTES SINGLE PA
1230000024	$389,207	CIGARETTES (SINGLE P
1230000010	$370,702	CIGARETTES (SINGLE P
1230000011	$305,693	CIGARETTES (SINGLE P
1230000025	$237,548	CIGARETTES (SINGLE P
1230000020	$235,239	CIGARETTES (SINGLE P
194	$217,040	GENERIC SINGLE
1230000013	$213,203	CIGARETTES (SINGLE P
1230000026	$210,678	CIGARETTES (SINGLE P
1230000014	$178,920	CIGARETTES (SINGLE P
1230000022	$161,957	CIGARETTES (SINGLE P
1230000012	$108,789	CIGARETTES (SINGLE P
1230000017	$89,858	CIGARETTES (SINGLE P
1230000018	$83,485	CIGARETTES (SINGLE P
1100015308	$76,399	KING CARTON
1230000007	$68,047	CIGARETTES SINGLE PA
1230000004	$64,659	CIGARETTES SINGLE PA
1230000001	$63,779	CIGARETTES SINGLE PA
193	$53,396	25 CT SINGLE PACK CI
1230000019	$48,472	CIGARETTES SINGLE PA

Figure 2.2-13: Pivot Table

Resulting Graph

This is a ranking exercise to select Top brands which will be visualized only through a pivot table. Hence no graph has been prepared

ERD

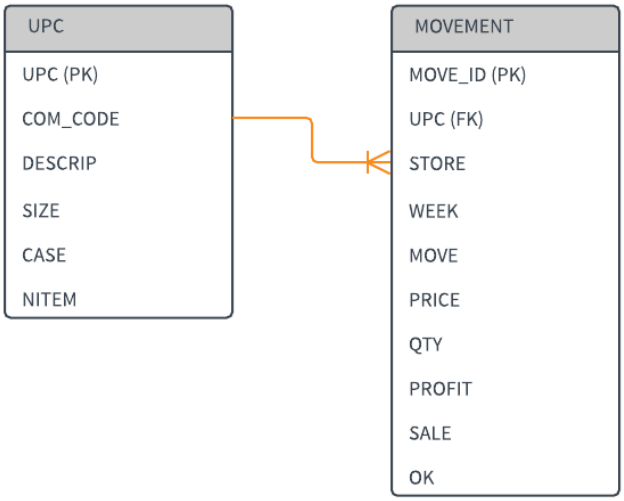

Figure 2.2-14: ERD

Justification

By analyzing profits of cigarettes over the five year period from 1991 to 1995, we can understand the trend in cigarettes sales. Moreover, we will able to recognize the brand of cigarette which has more margin than the other brands. This information will help a retailer to stock up the brand which has higher margin to cater the customer needs. The question can also find out which are low performing brands and will help in weeding out brands not meeting the minimum profit criteria. Such brands will be taken off from shelves. This information will assist in both shelf as well as inventory management.

Q6. Based on different zones, how much is consumption of non-vegan products like fish, meat and dairy across Chicago?

Data Processing

- City and Zone for a store we looked up from the Demographics table
- Sales dollars for Dairy, Fish and Meat were taken from Ccount table

Resulting Pivot

City	CHICAGO		
Zone	**Sum of FISH**	**Sum of MEAT**	**Sum of DAIRY**
1	$13,156,490	$88,448,796	$69,324,088
2	$8,874,569	$68,920,087	$57,837,695
5	$2,995,019	$26,070,552	$20,411,042
7	$8,714,674	$51,263,163	$55,261,456
10	$1,417,653	$13,878,544	$9,521,848
11	$7,228,776	$47,029,714	$39,654,619
12	$2,041,704	$23,095,245	$19,594,750
Grand Total	**$44,428,884**	**$318,706,101**	**$271,605,497**

Figure 2.2-15: Pivot Table

Resulting Graph

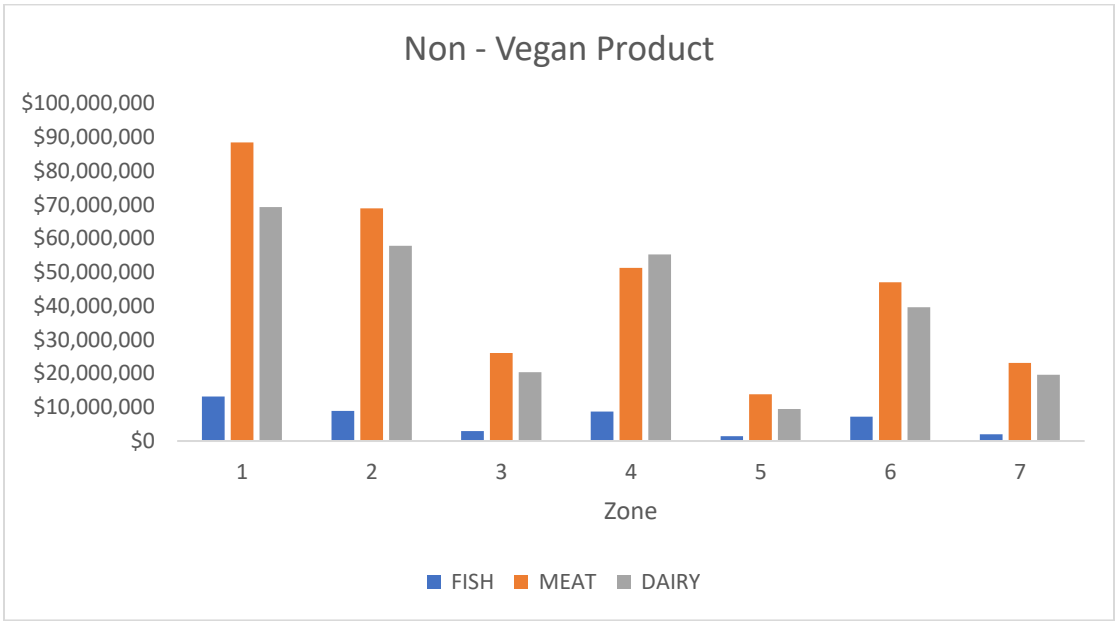

Figure 2.2-16: Pivot Chart

ERD

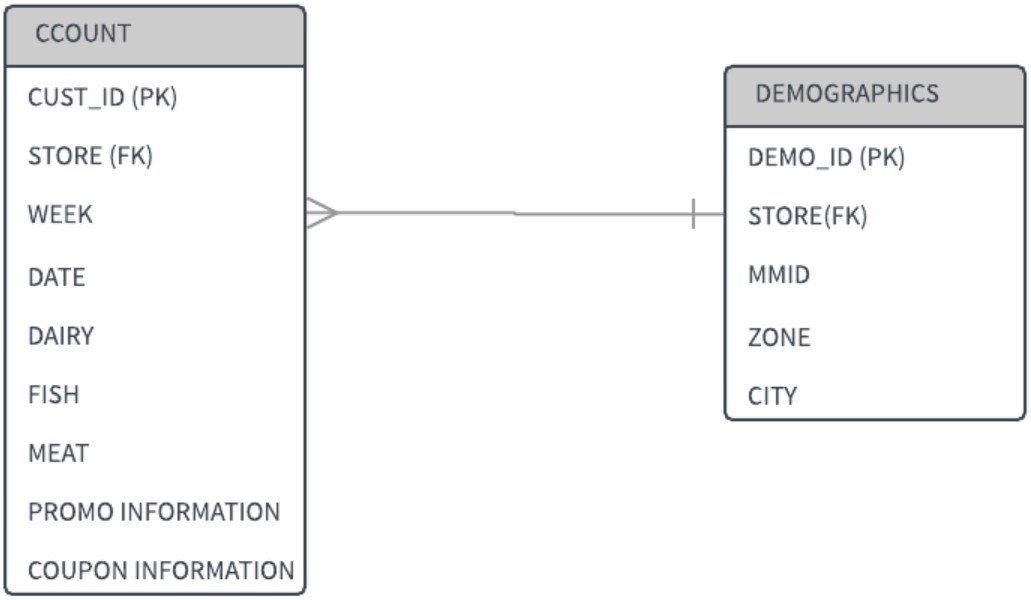

Figure 2.2-17: ERD

Justification

Going vegan has been a prevailing food trend. A vegan is a person who abstains from consuming animal products. Studying the sales of non-vegan products like fish, meat and dairy can help estimate the customer base of vegetarians and vegans in the zone. This will help DFF to manage the stock of vegan foods as per the sale trends. The store in the zone which has higher sale of non-vegan foods should stock up more fish, meat and dairy, whereas, the store in the zone with higher vegan food consumption should reduce the stock of fish, meat and dairy, in addition increase the stock of vegetables and other vegan food.

Q7. What is the profit trend of cheese during 1991 to 1995?

Data Processing

- Cheese profit for each UPC in Movement file is calculated using the formulae ((move*price/ qty)*profit/100)
- Data will be aggregated on annual basis

Resulting Pivot

Week Num	Sum of Final Profit
1	21163.71379
2	20959.02892
3	22631.23713
4	23267.28938
5	24747.56578
6	24273.32336
7	25643.08066
8	26948.80003
9	24777.82375
10	29169.12299
11	20552.58679
12	26643.48755
13	53990.0209
14	29692.48695
15	25836.26845
16	21661.88723
17	20997.37402
18	21623.2445
19	24095.24999
20	29065.50602

Figure 2.2-18: Pivot Chart

Resulting Graph

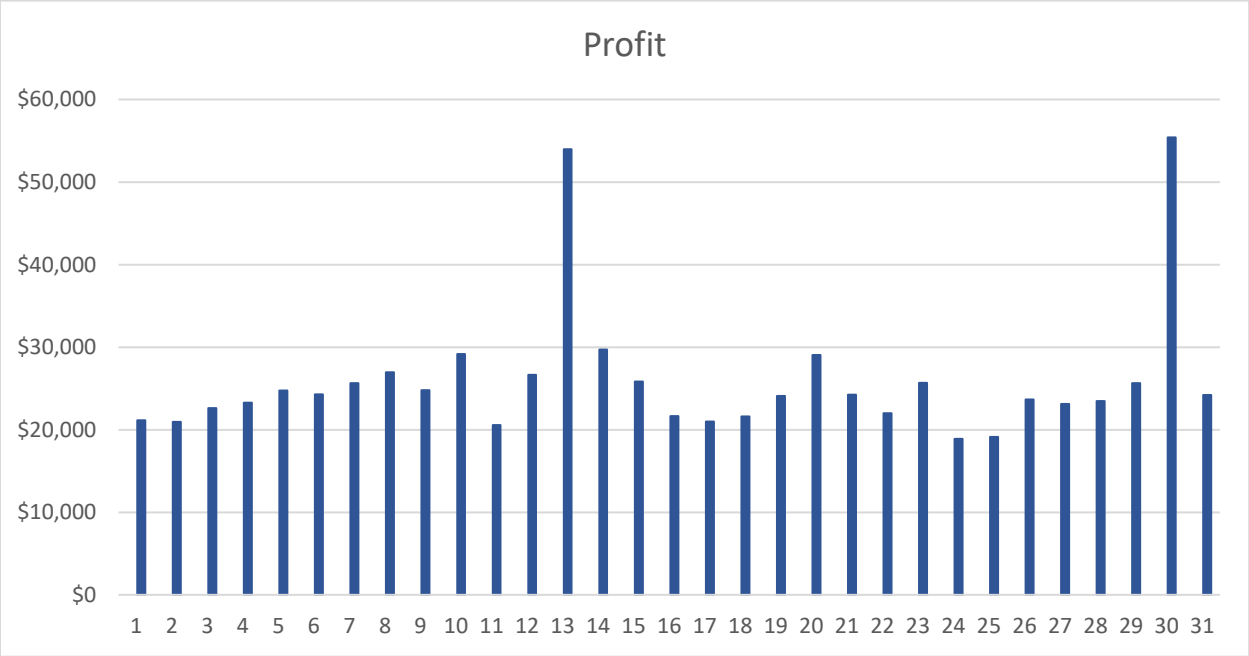

Figure 2.2-19: Pivot Chart

ERD

Figure 2.2-20: ERD

Justification

By analyzing profits of cheese over the five-year period from 1991 to 1995, we can understand the trend in cheese sales across years. This also will help retailers understand which cheese products generate high profit from their sale so that the production of those cheese products can be increased. It also helps in finding out the cheese products with low profit from cheese sale so that they can shift their focus from cheese sale because perhaps the demand of cheese is not high in that area or high profit margin cheese are not popular there. This information will assist in both shelf as well as inventory management.

Q8. How the trend of grocery sales varies across different stores in Chicago?

Data Processing:

- City for a store is looked up from Store table
- Grocery Sales for stores are picked up from Ccount table

Resulting Pivot

City	Chicago
Store	**GROCERY SALES**
12	$74,943,358
25	$11,479,544
33	$57,030,266
53	$58,973,692
68	$53,274,821
73	$101,435,589
75	$54,922,153
76	$65,495,733
86	$77,421,637
89	$51,988,043
90	$53,394,946
95	$61,128,336
98	$93,185,926
100	$87,738,612
111	$62,092,452
113	$83,278,312
123	$67,852,188
128	$86,124,054
130	$55,661,537
Grand Total	**$1,257,421,198**

Figure 2.2-21: Pivot Chart

Resulting Graph

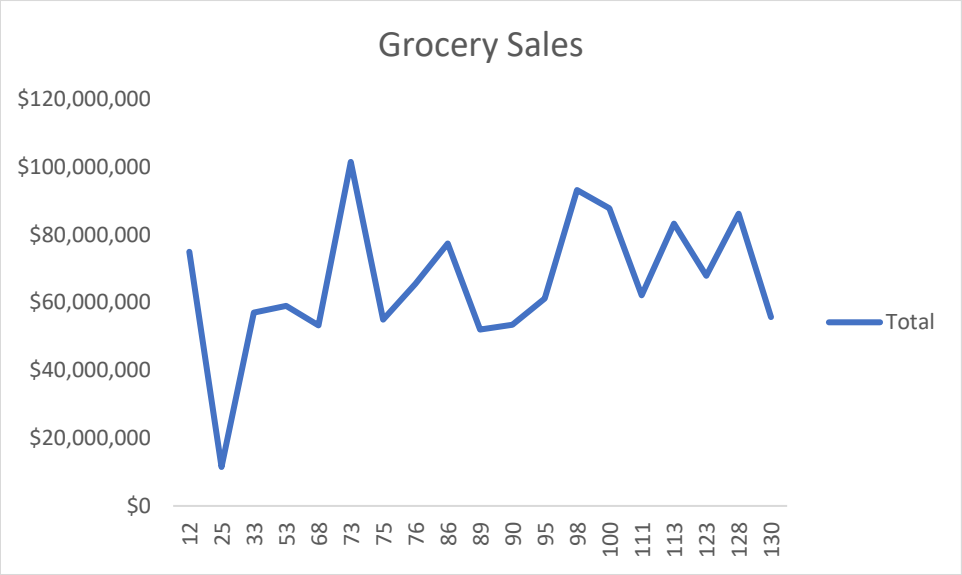

Figure 2.2-22: Pivot Chart

ERD

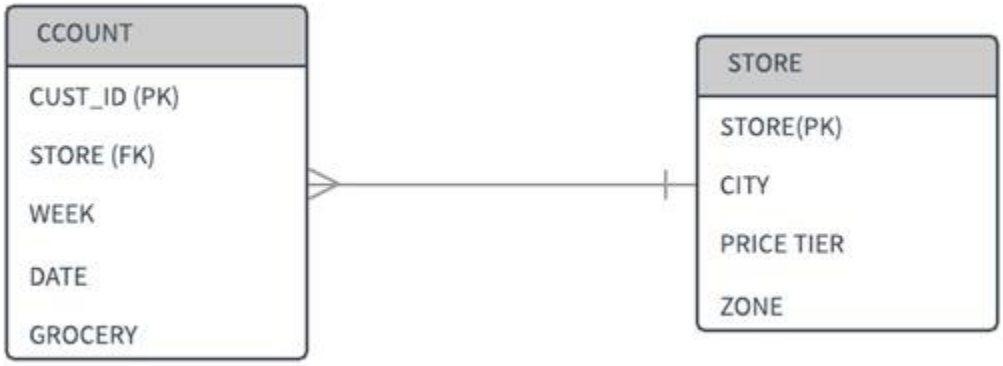

Figure 2.2-23: ERD

Justification

Grocery are brought routinely by everyone and is provided in the each store. Studying the sales of grocery across different stores in a city will help DFF in stock management. Moreover, this analysis will help identifying the stores where the grocery sales are higher and lower. Store with higher sales of grocery should stock up more groceries, whereas, for stores with lower sale, strategies can be planned to enhance sales or these stores can be shut if the losses are high.

Q9. What are the top cookies that have highest profit in all stores combined?

Data Processing

- Profit is calculated for each cookie as per the profit formulae((move*price/qty)*profit/100)
- UPCs with top profits are selected from the movement file using a pivot
- Cookie names are then looked up from UPC table

Resulting Pivot

Cookie Name	Sum of Profit
PPF PARTY FAVORITES	$90,782
P FARM DOUBLE CHOC M	$78,520
PPF DESSERT FAVORITE	$48,138
P FARM TAHOE AMER CO	$25,806
$PEPP FARM CAFE' FAV	$25,591
PEP FARM OATMEAL RAI	$25,445
PEP FRM CHERRY COBBL	$11,527
~PF	$5,121

Figure 2.2-24: Pivot Table

Resulting Graph

This is a ranking exercise to select Top brands which will be visualized only through a pivot table. Hence no graph has been prepared

ERD

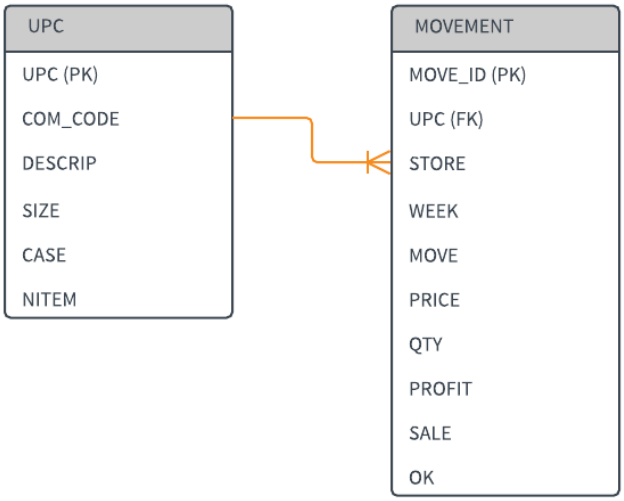

Figure 2.2-25: ERD

Justification

This analysis helps in identifying the top cookies with highest profit in all the stores combined which can help understand the trend in cookies sale. The top profitable brands should be focused on more and their stocks should be increased to gain high profits and meet customer needs. This analysis can also find out which are low profit brands and will help in taking the brands off the shelf which are not meeting the minimum profit criteria. This will help in better inventory management and hence improve the overall profits.

Q10. Detect trends in Beer Sales encompassing the entire duration that look for high period of sales. Detect peaks not only during designated holidays but also on events that are not in the Week_Decode table.

Data Processing

- Beer Sales are mapped against their weekly volumes in the Ccount table

Resulting Pivot

Week	Sum of BEER
1	$323,262
2	$310,749
3	$334,978
4	$310,008
5	$319,366
6	$326,395
7	$361,653
8	$329,395
9	$317,514
10	$543,976
11	$339,089
12	$312,602
13	$355,500
14	$475,092
15	$586,086
16	$507,795
17	$286,805
18	$280,094
19	$294,689
20	$406,287

Figure 2.2-26: Pivot Table

Resulting Graph

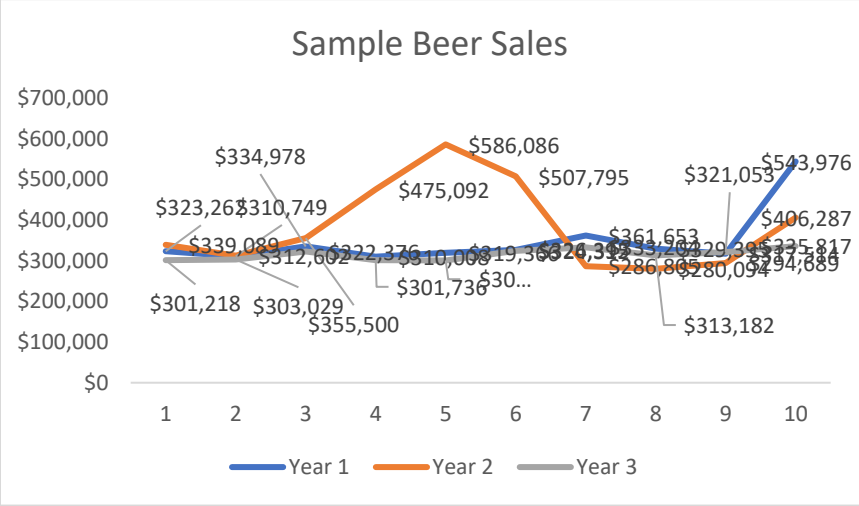

Figure 2.2-27: Pivot Chart

ERD

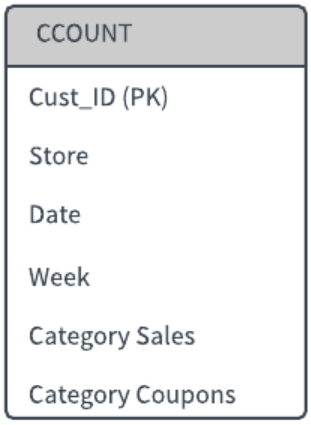

Figure 2.2-28: ERD

Justification

Most of us consider beer as a staple. Be it a college graduate, a working professional or a retired person, almost everybody consumes it. Also, beer is best served cold, it has to be refrigerated before it is to be. Also, since beer is a non-perishable product, it can be stocked up. By studying the sale of beer across multiple DFF stores, stores with higher and lower beer sales can be identified. This analysis will prove to be a valuable insight as it can help DLL in stock management and inventory management. Moreover, this will help management take decisions for the beer pricing. Management can use strategies like, stores with higher sales of beer and lower prices can use an idea to reduce the prices by little to increase beer sales and profit.

2.3 Final Selected Business Questions

The business question finally selected are as below -

Q1. How are the sales of Bakery products affected during Christmas period over the period 1989 -1996?

Q2. Are there any periods during the years 1989-1995 where the video sales are higher or lower?

Q3. What is the profit trend of cheese during 1991 to 1995?

Q4. How the trend of grocery sales varies across different stores in Chicago?

Q5. Detect trends in Beer Sales encompassing the entire duration that look for high period of sales. Detect peaks not only during designated holidays but also on events that are not in the Week_Decode table.

3 Independent Data Marts Design using Kimball's approach

3.1 Build Matrix

3.1.1 List Data Marts

We have created two data marts to model data required to fulfill business questions requirements. These data marts have been carefully crafted to ensure accurate and relevant information is stored and retrieved to support business questions. The two data marts are as below-

1. Fact Category Sales Data Mart
2. Fact Product Margin Data Mart

3.1.2 List Dimensions

We have used four dimensions to analyze the data. These are –

1. Category Dimension
2. Time Dimension
3. Store Dimension
4. Product Dimension

3.1.3 Mark the Intersection

The matrix indicates how dimension tables will be used in combination with fact tables to create required data marts. A * in row-column indicates the presence of dimension-fact table combination in the matrix.

	Dimension Table			
Fact table	**Category**	**Time**	**Store**	**Product**
Fact Category Sales	*	*	*	
Fact Product Margin		*		*

Table 3.1.1: Dimension Matrix

3.2 Design Each Fact Tables

3.2.1 Choose Data Mart

Data Mart 1 – Category Sales Data Mart
It will cater to business questions across Category, Time and Store dimensions with measures on each category. It will provide insights on questions 1 to 3 and 5.

Data Mart 2 – Product Margin Data Mart
It will cater to business questions across Time and Product dimensions with measures on each Product. It will provide insights on questions 4.

3.2.2 Declare Grain

Data Mart 1 – Category Sales Data Mart
Sales Data will be aggregated along Category, Time and Store Dimension. Time will aggregate data, at the lowest level, on weekly basis for each store and each category. Hence, the lowest level of detail i.e. data granularity of this data mart is at week level for each store and category.

Data Mart 2 – Product Margin Data Mart
Sales Data will be aggregated along Product and Time Dimension. Time dimension will aggregate data, at the lowest level, on weekly basis for each product. Hence, the lowest level of detail i.e. data granularity of this data mart is at week level for each product.

3.2.3 Choose Dimensions

The dimensions that will be used for Data Mart 1 – Category Sales Data Mart
1. Store
2. Category
3. Time

The dimensions that will be used for Data Mart 2 – Product Margin Data Mart
1. Product
2. Time

3.2.4 Include Derived Fact

Derived fact is included only in Data Mart 2 – Product Margin Data Mart. The profit margin for each record is calculated using the formulae ((move*price/ qty)*profit/100).

3.2.5 Design Fact Table

3.2.5.1 Fact Category Sales

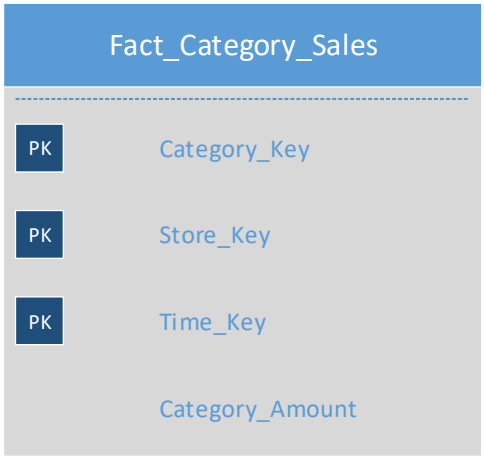

Figure 3.2-1: Fact Category Sales

Fact Table Attribute Description:

- **Category_ID** - This is a surrogate key which identifies the category of the product.
- **Store_ID** - Primary key which uniquely identifies each retail store of Dominick's Finer Foods (DFF). This is a surrogate key.
- **Time_ID** - This is a surrogate key uniquely identifies time of purchase.
- **Category_Amount** – Total sales amount for that category.

3.2.5.2 Fact Product Margin

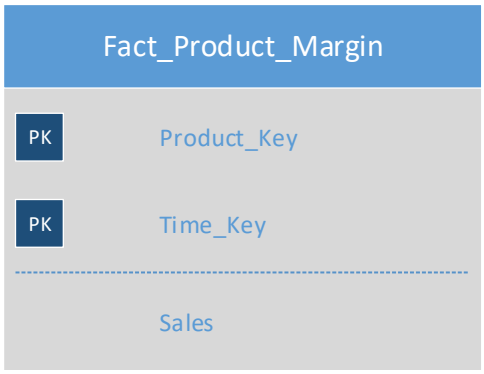

Figure 3.2-2: Fact Product Margin

Fact Table Attribute Description:

- **Product_ID** - Primary key which uniquely identifies each product in Dominick's Finer Foods (DFF) retail stores. This is a surrogate key.
- **Time_ID** - This is a surrogate key uniquely identifies time of purchase.
- **Product_Sales** – Identifies the total dollar sales.

3.3 Design Dimension **Table**

3.3.1.1 dimProduct

Dimensions with information on Product Names and UPC Numbers.

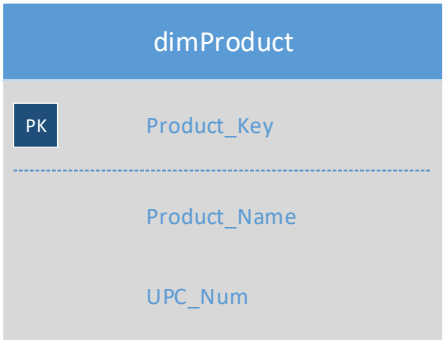

Figure 3.3-1: dimProduct Dimension

Dimension Table Attribute Description:

- **Product_ID** - Primary key for Product dimension table which uniquely identifies each product in Dominick's Finer Foods (DFF) retail stores. This is a surrogate key.
- **Product_Name** - Product name for each item.
- **UPC_Num** - Last five digit of the UPC number identify the product, the remaining digits identify the manufacturer.

3.3.1.2 dimStore

Dimensions with information on Store Number with corresponding city and price tiers.

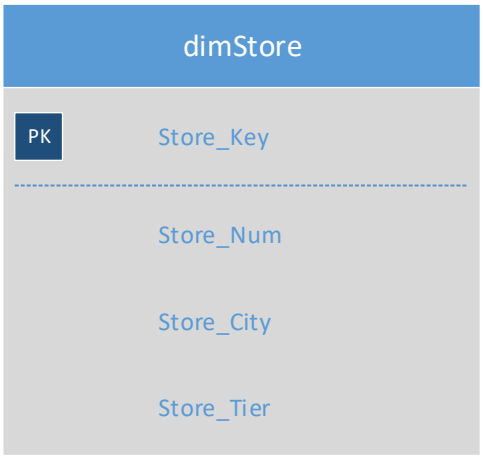

Figure 3.3-2: dimStore Dimension

Dimension Table Attribute Description:

- **Store_ID** - Primary key for Store dimension table which uniquely identifies each retail store of Dominick's Finer Foods (DFF). This is a surrogate key.
- **Store_Num** - Number assigned to each of DFF's stores.
- **Store_City** – City where the DFF store is located.

- **Store_Tier** – Price tier in which the store falls.

3.3.1.3 dimCategory

Dimension with category name information.

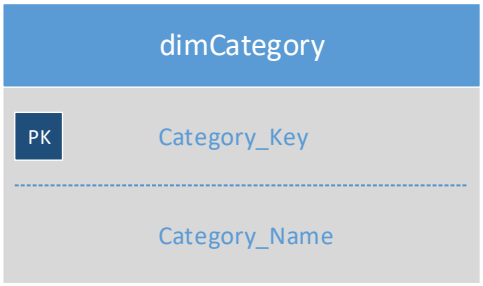

Figure 3.3-3: dimCategory Dimension

Dimension Table Attribute Description:

- **Category_ID** - This is a surrogate key which identifies the category of the product.
- **Category_Name** – Name of the category under which the product falls.

3.3.1.4 dimTime

Dimension with Time information on Year, Month and Week Number along with special events.

Figure 3.3-4: dimTime Dimension

Dimension Table Attribute Description:

- **Time_ID** - Surrogate key for time dimension table.
- **Year** - Identifier for the year to record the sales of products in DFF's stores
- **Month** - Identifier for the month to record the sales of products in DFF's stores.
- **Week_Num** - Identifier for the week number to record the sales of products in DFF's stores.

- **Week_Event** - Identifier for the special event during the week.

3.4 Design Star Schema for Each Data Mart

The selected dimension tables and fact tables along with dimensional matrix have been used to create and design following two data marts in STAR schema:

- Category Sales Data Mart
- Product Margin Data Mart

3.4.1.1 Fact Category Sales Data Mart

The Category Sales Data Mart encompasses following fact and dimension tables:
Fact table: Category Sales fact table
Dimension tables: Store dimension table, Category dimension table and Time dimension table

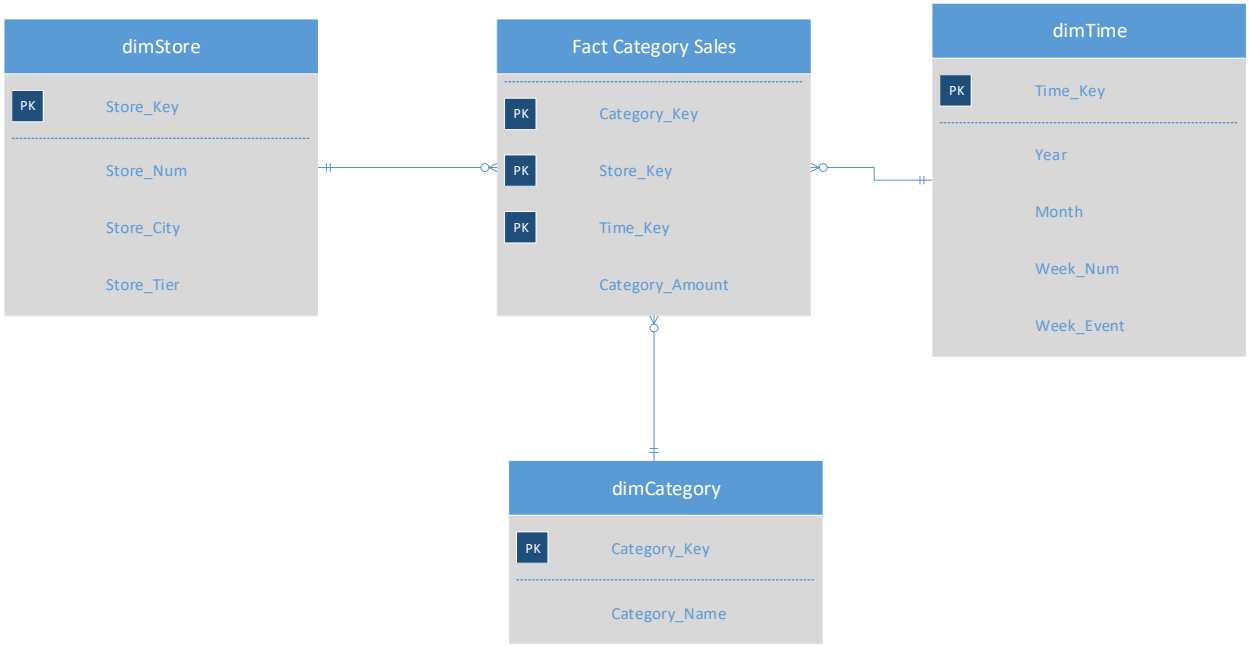

Figure 3.4-1: Fact Category Sales Data Mart

3.4.1.2 Fact Product Margin Data Mart

The Product Margin data mart encompasses following fact and dimension tables:
Fact table: Product Margin fact table
Dimension tables: Product dimension table and Time dimension table

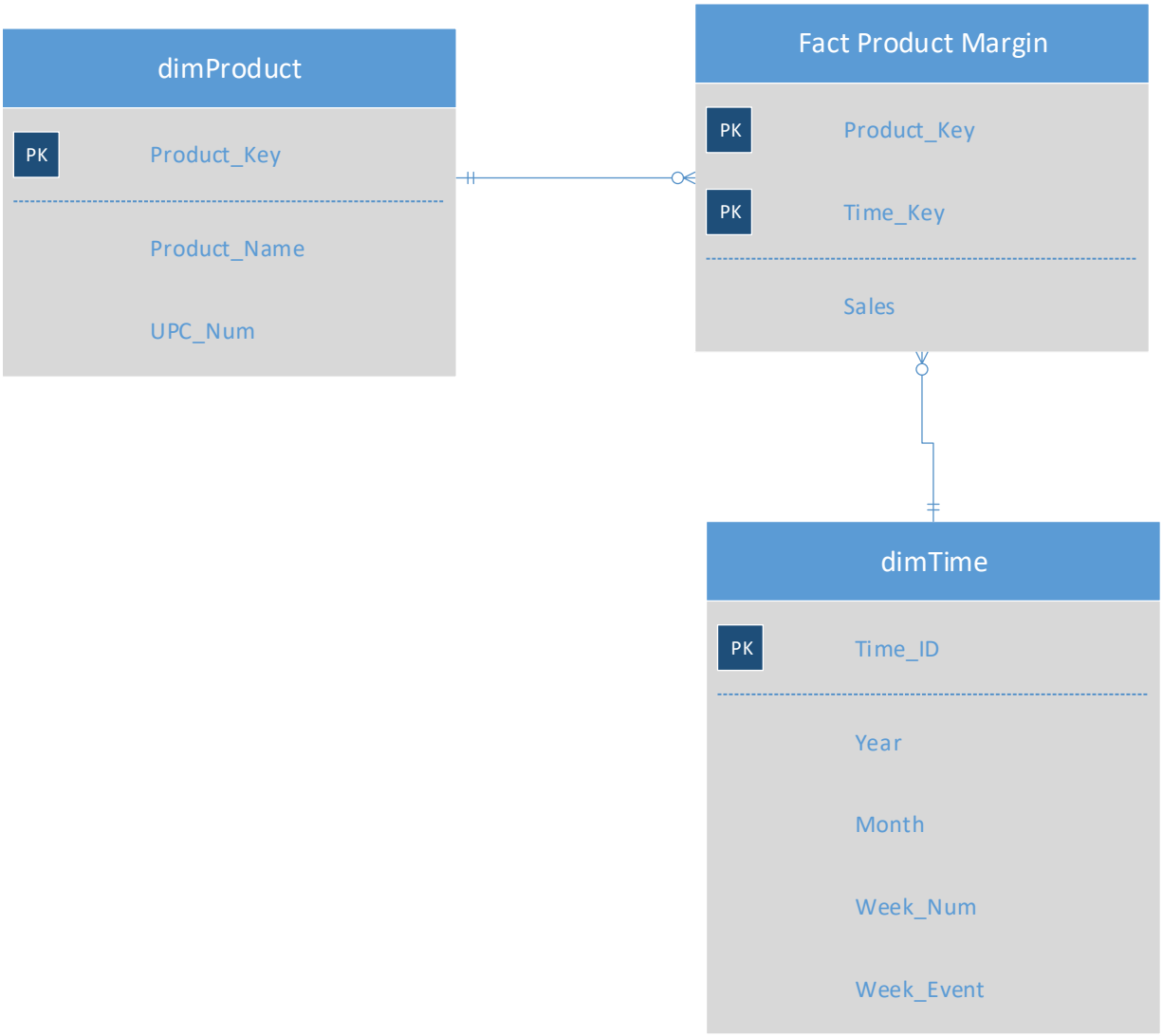

Figure 3.4-2: Fact Product Margin Data Mart

3.5 Mapping Tables

3.5.1.1 Category Sales Data Mart

DW Dimension Table	Source Table	Source Table Attribute	DW Dimension Table Attribute	Mapping Function
dimCategory			Category_Key	None
	Ccount		Category_Name	Column Names as Categories

Table 3.5.1: Category Dimension Table

DW Fact Table	Source Table	Source Table Attribute	DW Fact Table Attribute	Mapping Function
Fact Category Sales	Category Dimension		Category_Key	Primary key
	Store Dimension		Store_Key	Primary key
	Time Dimension		Time_Key	Primary key
	Ccount	None	Category_Amount	Sales of each categories

Table 3.5.2: Category Sales Fact Table

DW Dimension Table	Source Table	Source Table Attribute	DW Fact Table Attribute	Mapping Function
dimTime			Time_Key	None
	Week Decode	Week #	Week_Num	None
		Start Column - End Column	Year	Determine Year by Start & End Date
		Start Column - End Column	Month	Calculate month by Start & End Date
		Special Events	Week_Event	None

Table 3.5.3: Time Dimension Table

3.5.1.2 Product Margin Data Mart

DW Dimension Table	Source Table	Source Table Attribute	DW Dimension Table Attribute
dimProduct			Product_Key
	UPC Table	UPC	UPC_Num
		Descrip	Product_Name

Table 3.5.4: Product Dimension Table

DW Dimension Table	Source Table	Source Table Attribute	DW Dimension Table Attribute
dimStore			Store_Key
	Dominick's Stores	Store	Store_Num
		City	Store_City
		Price Tier	Store_Tier

Table 3.5.5: Store Dimension

DW Fact Table	Source Table	Source Table Attribute	DW Fact Table Attribute	Mapping Function
Fact Product Sales	Product Dimension		Product_Key	Primary key
	Store Dimension		Store_Key	Primary key
			Product_Sales	Product Sales = (Price*Move)/Qty

Table 3.5.6: Product Margin Fact Table

3.6 Data Design Justification

Q1 How are the sales of Bakery products affected during Christmas period over the period 1989 -1996?

Sales of Bakery product will be analyzed through the Category Sales Data Mart. Total sales amount data in a product category will be stored in the fact table Category Sales. Category will be selected using Category dimension. This data will be sliced and diced using Time dimension to analyze trend in sales of Bakery Product during period 1989 -1996.

Q2 Are there any periods during the years 1989-1995 where the video sales are higher or lower?

Sales of Videos will be analyzed through the Category Sales Data Mart. Total sales amount data for videos will be stored in the fact table Category Sales. Video Category will be selected using Category dimension. This data will be sliced and diced using Time dimension to examine trend in sales of videos during period 1989-1995.

Q3 What is the profit trend of cheese during 1991 to 1995?

Profit trend will be analyzed through the Product Margin Data Mart. Total Profit Sales Amount data for cheese will be derived and saved in the Product Margin Data Mart as per the mapping function. Each of the cheese product will be identified and studied using Product dimension. The profit will be sliced and diced using the Time dimension.

Q4 How the trend of grocery sales varies across different stores in Chicago?
Sales of Grocery will be analyzed through the Category Sales Data Mart. Total sales amount data for grocery will be stored in the fact table Category Sales. Grocery Category will be selected using Category dimension. This data will be sliced and diced using Store and Time dimension to examine trend in sales of grocery.

Q5 Detect trends in Beer Sales encompassing the entire duration that look for high period of sales. Detect peaks not only during designated holidays but also on events that are not in the Week_Decode table.
Sales of Beer will be analyzed through the Category Sales Data Mart. Total sales amount data for beer will be stored in the fact table Category Sales. Beer Category will be selected using Category dimension. This data will be sliced and diced using Time dimension to examine trend in sales of grocery.

4 Data Cleaning and Integration

4.1 Data Quality Issues

The source data used for analysis was polluted. It had various quality issues such as dummy values in fields, absence of data values, unofficial use of fields, violation of business rules, inconsistent and incorrect values etc. The data quality issues specific to the creation of our data mart have been provided in table

Group	Quality	Issues Considered	Example of Data Quality Problem
Relation to other data	Referential Integrity	Do records exist where expected? Do they contain unnecessary or inactive data? Are reference files/tables complete?	It was difficult to track referential integrity as there were a lot of null values for attributes which could be considered as primary keys. For example, Store number had repeated and inconsistent values.
	Cardinality	Is the structure of relationships among entities and attributes maintained consistently?	There is a high level of inconsistency in the structure of relationships among entities and attributes.

Structure of fields	Format	Do values follow consistent formatting standards?	In Week_Decode table, prior to the data conversion operation on this file, we were unable to use the dates. The format of date is different in different excel sheets.
	Standard	Are data elements consistently defined and understood?	Lack of standardization makes it difficult to understand what values the attributes hold. For example, Coupons had certain decimal values which makes it difficult to guess if it is a dollar value or dirty data.
	Consistent	Do values represent the same meaning across systems and files?	Certain attributes has inconsistency across files and hence led to data misinterpretation.
Content within data values	Complete	Is all necessary data present?	A lot of data entries were missing. There were null values for week in ccount, sales in movement file, etc.
	Accurate	Does the data accurately represent reality or a verifiable source?	The movement file we used was Done-CHE.csv. This table contained a lot of junk values. The records with value of "OK" field set as 0 were inaccurate.
	Valid	Do data values fall within acceptable ranges defined by the business?	The file used was Store.csv. The file still contained information on stores which were closed.
	Fit For Purpose	Is the information valuable to the business? Does the data convey information that can intelligently be consumed by the business?	The file used was UPCCHE.csv. The UPC number was in a format which was not suitable for lookup operations in the warehouse.

Table 4.1.1: Data Quality Issues

4.2 ETL Development Plan

An ETL Development Plan was designed to frame the roadmap for the data load process into the data warehouse. After determining the target data and source data, we mapped the tables for

staging and data mart loads by using comprehensive data extraction rules. We have attached the screen shots for the same displaying the data transformation and cleansing rules. The steps followed in ETL process are as below -

4.2.1 Target Data Required in Data Warehouse

We need six table in Data Warehouse area. They are categorized as below -
Four Dimension tables

1. dimTime
2. dimStore
3. dimProduct
4. dimCategory

Two Fact Tables

1. factProductMargin
2. factCategorySales

4.2.2 Source Data Required in Data Warehouse

We need the below data files

1. CCount.csv for category data
2. Done-CHE.csv for product transaction data
3. Store.csv for store data
4. Week_Decode.csv for time data
5. UPCCHE.csv for product information

4.2.3 ETL Specific Mapping Tables

These are required to map source data to target data. ETL specific mapping tables are extension of Mappings establish in section 3.2.7. Mapping tables provide in detail the flow of source data into staging area table which then flows into data warehouse tables.

Source to Staging Mapping Tables
This includes intermediate mapping from raw extracted file to intermediate tables with cleaned data. These intermediate tables are not temporary table. They are essential for storing and loading data to dimension and fact table.

Staging Table	Staging Attribute	Used	DataType	Cleaned Staging Table	Cleaned Staging Table Attribute	DataType	Comments
CCount	Store	Yes	varchar	**finalCCount**	Store	bigint	
CCount	Grocery	Yes	varchar	**finalCCount**	Grocery	numeric	
CCount	Bakery	Yes	varchar	**finalCCount**	Bakery	numeric	
CCount	Video	Yes	varchar	**finalCCount**	Video	numeric	
CCount	Beer	Yes	varchar	**finalCCount**	Beer	numeric	
CCount	Week	Yes	varchar	**finalCCount**	Week	bigint	
CCount	Dairy, Frozen, Bottle, Mvpclub, Groccoup, Meat, Meatfroz, Meatcoup, Fish, Fishcoup, Promo, Promcoup, Produce, Bulk, Saladbar, Prodcoup, Bulkcoup, Salcoup, Floral, Florcoup, Deli, Deliself, Deliexpr, Convfood, Cheese, Delicoup, , Pharmacy, Pharcoup, Gm, Jewelry, Cosmetic, Haba, Gmcoup, Camera, Photofin, , Videoren, Vidcoup, , Wine, Spirits, Miscscp, Mancoup, Custcoun, Ftgchin, Ftgccoup, Ftgital, Ftgicoup, Daircoup, Frozcoup, Habacoup, Photcoup, Cosmcoup, Ssdelicp, Bakcoup, Liqcoup	No					
finalCCount	Store	No					
finalCCount	Grocery	Yes	numeric	**tempCategory**	Category Name		
finalCCount	Bakery	Yes	numeric	**tempCategory**			
finalCCount	Video	Yes	numeric	**tempCategory**			
finalCCount	Beer	Yes	numeric	**tempCategory**			
finalCCount	Week	No					
UPCCHE	Com_Code	No					
UPCCHE	Upc	Yes	varchar	**UPC**	Upc	bigint	
UPCCHE	Descrip	Yes	varchar	**UPC**	Descrip	varchar	
UPCCHE	Size	No					
UPCCHE	Case	No					
UPCCHE	Nitem	No					
Done-WCHE	Store	Yes	varchar	**finalMovement**	Store	bigint	
Done-WCHE	UPC	Yes	varchar	**finalMovement**	UPC	bigint	
Done-WCHE	Week	Yes	varchar	**finalMovement**	Week	bigint	
Done-WCHE	Move	Yes	varchar	**finalMovement**	Move	bigint	
Done-WCHE	QTY	Yes	varchar	**finalMovement**	QTY	bigint	
Done-WCHE	Price	Yes	varchar	**finalMovement**	Price	real	
Done-WCHE	Sale	No					
Done-WCHE	Profit	Yes	varchar	**finalMovement**	Profit	real	
Done-WCHE	Ok	For Calculation		**finalMovement**	Ok	bigint	
				finalMovement	Product Sales	real	Derived Column

Table 4.2.1: Source to Staging Mapping Table

Staging to Data Warehouse Mapping Tables

Staging Table	Staging Attribute	Used	DataType	Warehouse Table	Warehouse Attribute	DataType	Comments
Week_Decode	Week#	Yes	bigint	**dimTime**	Week_Num	bigint	
Week_Decode	Start	For Derivation	date				
Week_Decode	End	No	date				
Week_Decode	Special Events	Yes	varchar	**dimTime**	Week_Event	varchar	
				dimTime	Year	int	Derived from Start
				dimTime	Month	int	Derived from Start
				dimTime	Time_ID	bigint	Surrogate Key
Store	Store	Yes	varchar	**dimStore**	Store_Num	bigint	
Store	City	Yes	varchar	**dimStore**	Store_City	varchar	
Store	Price Tier	Yes	varchar	**dimStore**	Store_Tier	varchar	
Store	Zone	No					
Store	Zip	No					
Store	Address	No					
				dimStore	Store_ID	bigint	Surrogate Key
UPC	Upc	Yes	bigint	**dimProduct**	UPC_Num	bigint	
UPC	Descrip	Yes	varchar	**dimProduct**	Product_Name	varchar	
				dimProduct	Product_key	bigint	Surrogate Key
tempCategory	Category Name	Yes	varchar	**dimCategory**	Category Name	varchar	
				dimCategory	Category_Key	bigint	Surrogate Key
				factCategorySales	Sales	real	Aggregated from finalCCount
				factProductMargin	Product Sales	real	Aggregated from finalMovement

Table 4.2.2: Staging to Data Warehouse Mapping Table

4.2.4 Data Extraction Rules

Data extraction involves mining data from various data sources. It is the first step in the ETL process. The simple rule that has been followed in this project is to dump all source data in staging area as it is and then clean it.

Data was extracted from source files mentioned in **Section 4.2.2** and was convert to tables in SQL Server. All the source files we used in data extractions were .csv files.

All the loaded was loaded into staging area 602Group6_stagingDB.

4.2.5 Data Transformation and Cleaning Rules

Data Quality issues have been discussed in **Section 4.1**. To remove these issues and prepare data to be fit for analysis data cleaning and transformation exercise was performed. Data cleaning and transformation follows immediately after Data Extraction.

Each of the data quality issue was handled as per below rules-

1. Dirty Data
Records with any dirty data were removed. Record with any field value equal to "." was removed. Records with Week as negative values were removed.

2. Null Values
Records with Null values or blank values in any of the fields were removed.
3. Data Conversion

All text attributes were converted to varchar format. Number attributes such as Week_Num, Year, Month were converted to int. Measures such as sales were converted to numeric format.

5. Creation of Derived Attributes
Product profit margins were derived from existing data fields. The profit margin for each record is calculated using the formulae ((move*price/ qty)*profit/100).

4. Surrogate Key
Surrogate keys were defined for all records. All surrogate keys were stored as bigint.

Functions that were used for Data Cleaning and Transformation are as below –

- Unpivot- It is used for making a dataset more normalized by expanding the single record into multiple records with same value in single column.
- Pivot- It is used for making a dataset less normalized by compacting multiple records into single record.
- Lookup- It performs lookups by joining data in input columns with columns in a reference dataset.
- Derived Column- It creates new column values by applying transformation to the input columns.
- Data Conversion- It converts the data type of a column to another desired data type.
- Aggregate- It applies aggregate functions like sum, average, group by to the input columns and produce transformed columns.

ETL Implementation in **Section 4.3** provides detailed execution of the plan.

4.2.6 Form Aggregate Tables

Two aggregate operations have been used to create the two fact tables.

- Category Sales Fact table has been grouped by Time, Store and Category key and sum operation is applied on sales
- Product Margin Fact table has been group by Time and Product key and sum operation is applied on profit

ETL Implementation of fact tables in **Section 4.3.1** provides detailed execution of the plan.

4.2.7 Load the data

After all data has been cleaned, the staging area tables acts as a feeder to data warehouse dimension and fact tables.

All data has been loaded into 602Group6_DW_Area.

Data loading has been covered for each table in detail in **Section 4.3.2**

4.3 ETL Implementation

4.3.1 Extraction and Transformation of Source Data in Staging Area

4.3.1.1 Extraction and Transformation for Store Dimension Creation

The first step in this process is to create the staging area 602Group6_stagingDB. This step is automatically completed when we import our first file Store.csv to the staging area through a SSIS package.

Data Extraction

The first step in this process is to create the staging area 602Group6_stagingDB. This step is automatically completed when we import our first file Store.csv to the staging area through a SSIS package.

Extract Store.csv file to Staging Area

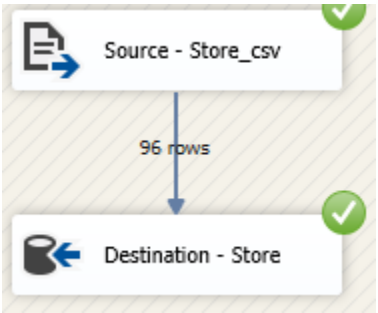

Figure 4.3-1: Extract Store.csv file to Staging Area

Snapshot of Store Table in Staging Area

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [Store]
      ,[City]
      ,[Price Tier]
      ,[Zone]
      ,[Zip]
      ,[Address]
  FROM [602Group6_stagingDB].[dbo].[Store]
```

100 %

Results | Messages

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

Figure 4.3-2: Snapshot of Store Table in Staging Area

Data Transformation

Transformation of Store Table to dimStore

1. Select only Store, City and Price Tier from Store table where address of store is not marked as Closed

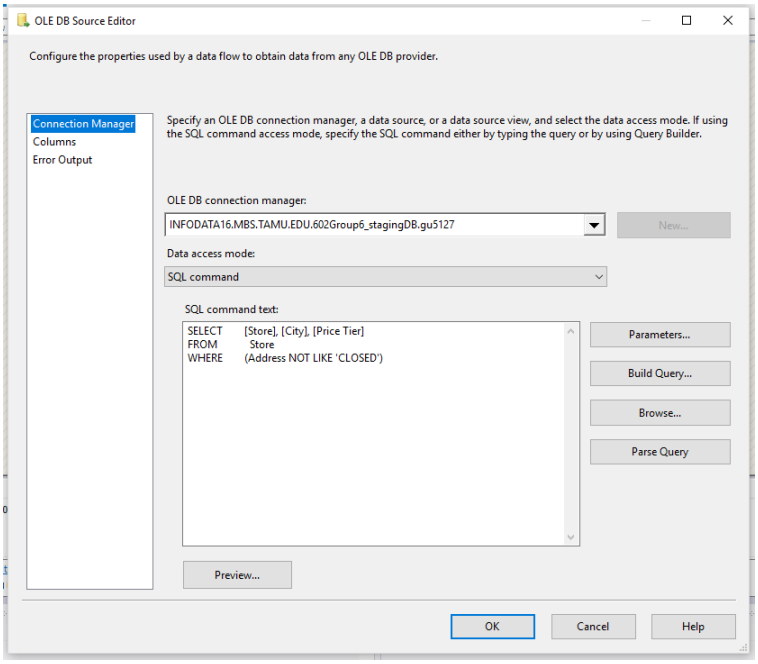

Figure 4.3-3: Selecting attributes from Store table

2. Convert Store from datatype varchar to a Converted Store with datatype bigint

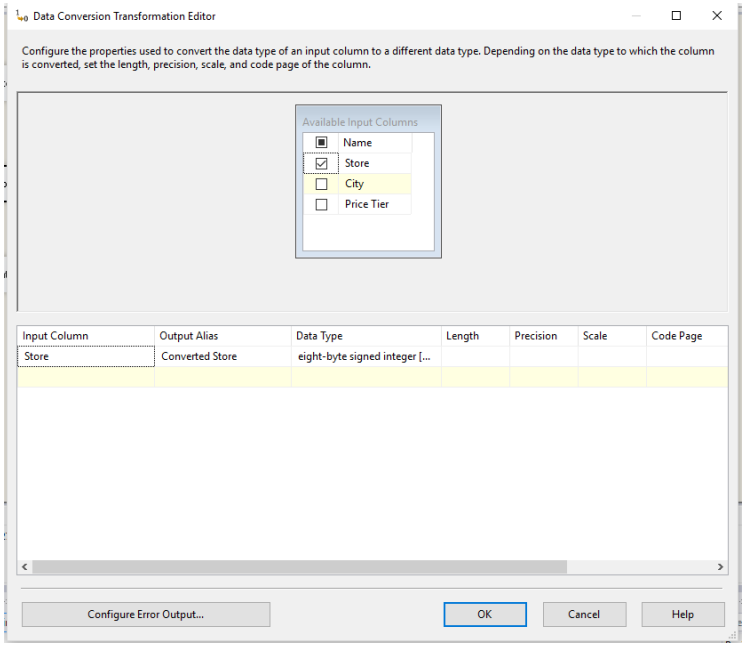

Figure 4.3-4: Converting Data Type of Store

4.3.1.2 Extraction and Transformation for Time Dimension

Data Extraction

Extract Week_Decode.csv file to Staging Area

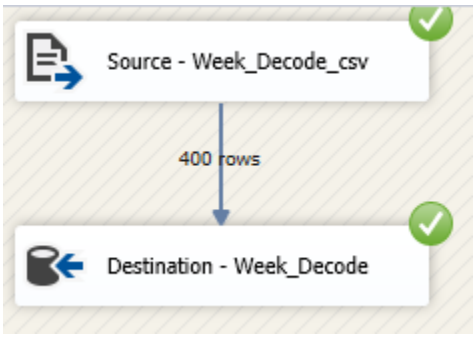

Figure 4.3-5: Extract Week_Decode.csv file to Staging Area

Snapshot of Week_Decode Table in Staging Area

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [ Week # ]
      ,[Start ]
      ,[End ]
      ,[Special Events ]
  FROM [602Group6_stagingDB].[dbo].[Week_Decode]
```

100 %

Results | Messages

	Week #	Start	End	Special Events
1	1	9/14/1989	9/20/1989	
2	2	9/21/1989	9/27/1989	
3	3	9/28/1989	10/4/1989	
4	4	10/5/1989	10/11/1989	
5	5	10/12/1989	10/18/1989	
6	6	10/19/1989	10/25/1989	
7	7	10/26/1989	11/1/1989	Halloween
8	8	11/2/1989	11/8/1989	
9	9	11/9/1989	11/15/1989	
10	10	11/16/1989	11/22/1989	

Figure 4.3-6: Snapshot of Week_Decode Table in Staging Area

Data Transformation

Week _Decode table to dimTime in Data Warehouse

1. Transform Start and End fields to date datatype from varchar. Also transform Week # to bigint datatype from varchar

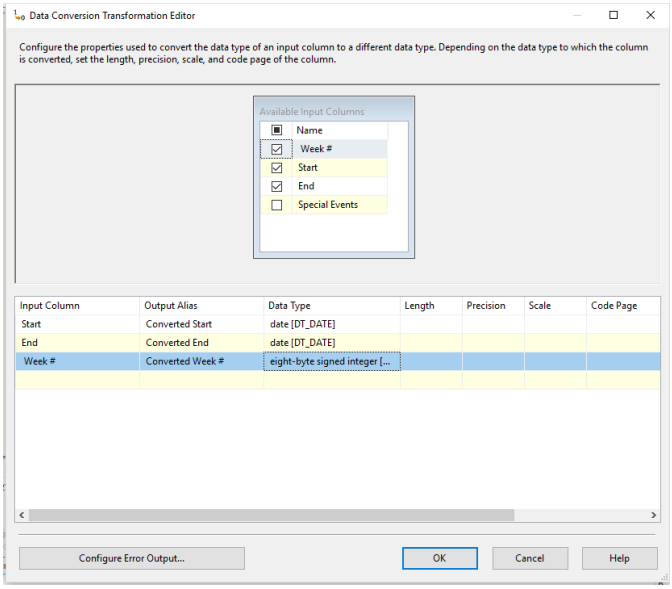

Figure 4.3-7: Transforming datatypes in Week_Decode

2. Derive Year and Month from Start field

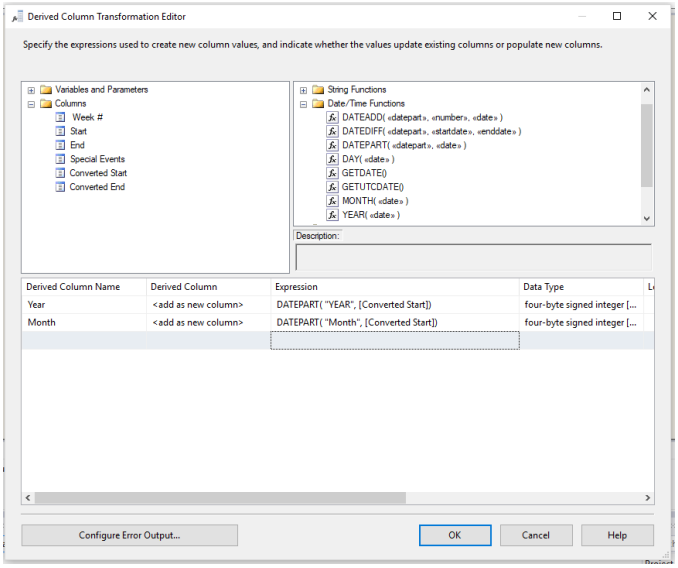

Figure 4.3-8: Deriving attributes in Week_Decode

4.3.1.3 Extraction and Transformation for Category Dimension

The csv file CCOUNT is used to create the category Dimension. This dimension will contain the sales amount for the categories Bakery, Beer, Grocery and Videos along with Date, Week and Store number.

Data Extraction

Extract CCount.csv file to Staging Area

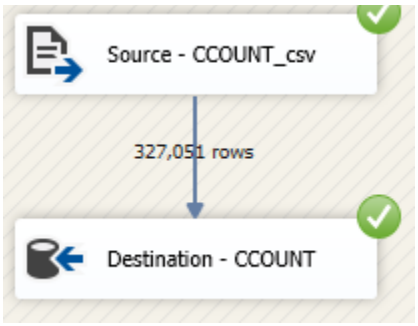

Figure 4.3-9: Extract CCount.csv to staging area

Snapshot of CCount Table in Staging Area

```
select * from CCOUNT
```

100 %

Results | Messages

	"STORE"	"DATE"	"GROCERY"	"DAIRY"	"FROZEN"	"BOTTLE"	"MVPCLUB"	"GROCCOUP"	"MEAT"	"MEATFROZ"	"MEATCOUP"	"FISH"	"FISHCOUP"	"PROMO"	"PROMCOUP"	"PRODUCE"	"BULK"	"SALADBAR"	"PRO
1	59	"950209"	21657.91	4678.37	4143.88	0	319.44	-779.49	5877.19	763.35	0	857.95	0	0	0	5382.21	695.45	157.18	-3.3
2	59	"950210"	23671.27	5105.01	4467.24	0	245.56	-720.64	6257.17	923.45	0	1200.22	0	0	0	6378.12	770.85	161.81	-3.2
3	59	"950211"	30531.72	6811.52	5963.35	0	341.92	-718.31	7308.45	1117.14	-2	1276.88	0	0	0	7543.23	786.6	206.07	-5.7
4	59	"950212"	23125.24	5057.43	4351.24	0	200.09	-624.23	5301.06	675.77	0	705.69	0	0	0	5509.92	535.71	158.3	-4.1
5	59	"950213"	17441.45	3779.4	3356.89	0	229.5	-550.92	4005	467.29	-3	630.51	0	0	0	4242.53	591.45	153.8	-2.95
6	59	"950214"	13883.89	3107.12	2786.31	0	202.45	-422.56	3079.86	442.65	0	958.87	0	0	0	3556.32	437.44	171.28	-5.4
7	59	"950215"	11446.47	2648.31	2287.15	0	180.38	-260.14	2259.17	380.66	0	273.14	0	11.98	0	2720.81	257.55	139.63	0
8	59	"950216"	19175.56	5082.77	6365.61	-1.6	139.66	-47.33	6660.68	415.37	-22	793.77	0	0	0	4794.53	479.04	150.7	-2
9	59	"950217"	21841.54	6069.9	7165.42	0	180.91	-40.62	7083.62	548.32	-14	886.02	0	0	0	5660.42	467.44	192.64	-4.5
10	59	"950218"	25644.11	6569.21	7246.73	0	197.03	-64.21	8649.32	713.1	-11	816.78	0	0	0	6715.37	600.45	184.54	-5.5

Figure 4.3-10: Snapshot of CCount Table in Staging Area

Data Transformation

1. Clean CCount file with the following transformations

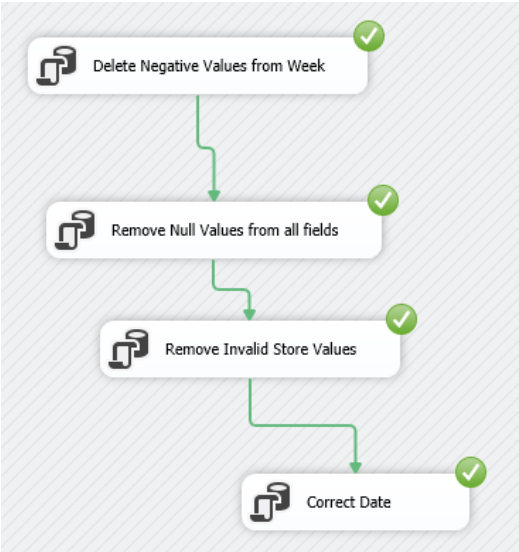

Figure 4.3-11: Clean CCount

2. Select required categories from CCount

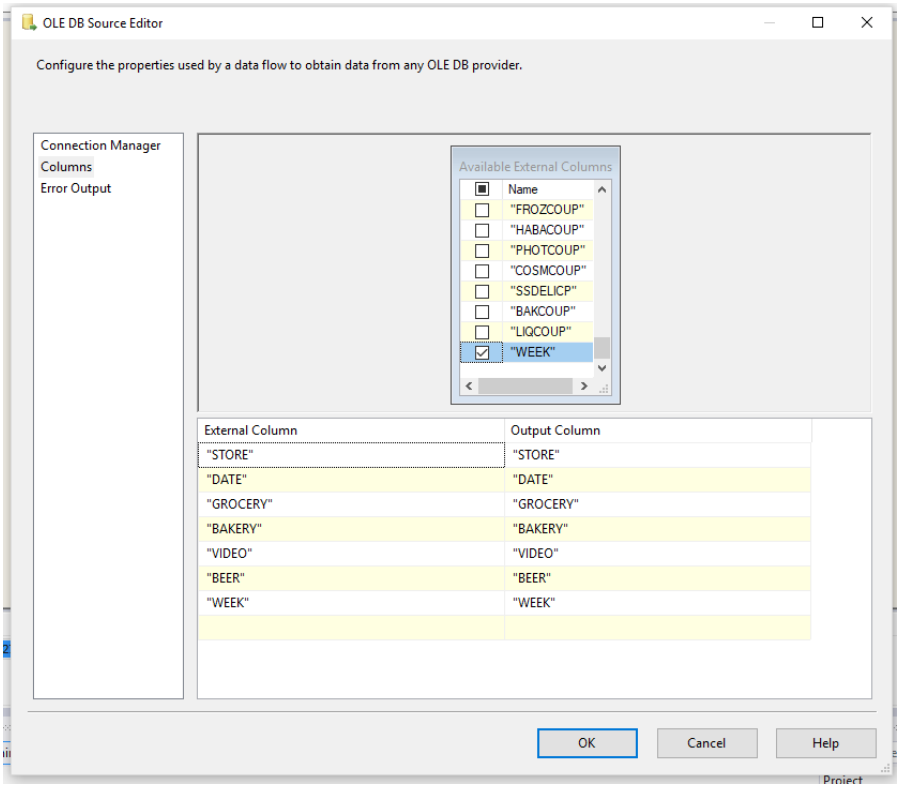

Figure 4.3-12: Selecting required categories from CCount

3. Create finalCCount file with required datatypes

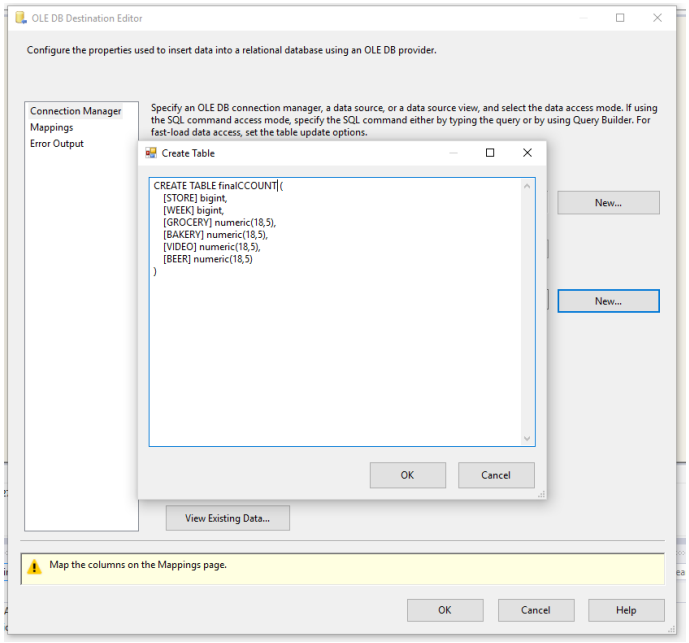

Figure 4.3-13: Creating finalCCount with required attributes

4. Map the fields to correct data types

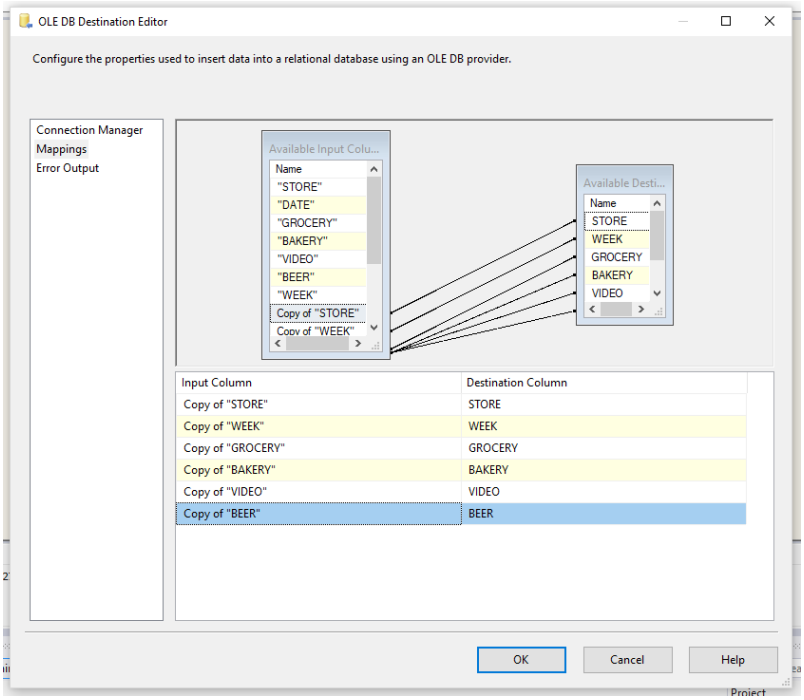

Figure 4.3-14: Mapping CCount fields to correct datatypes

5. Execute the package

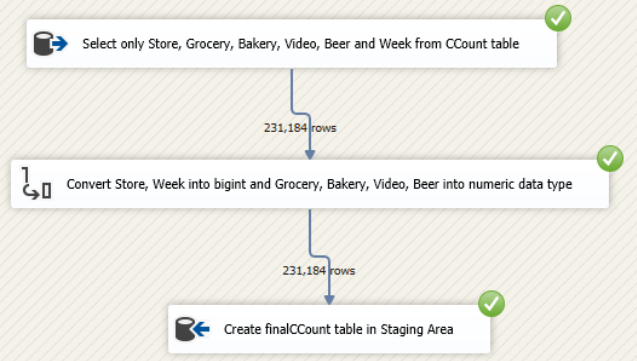

Figure 4.3-15: Executing finalCCount table

6. Create a temporary Category table in Staging Area to store the Category Names

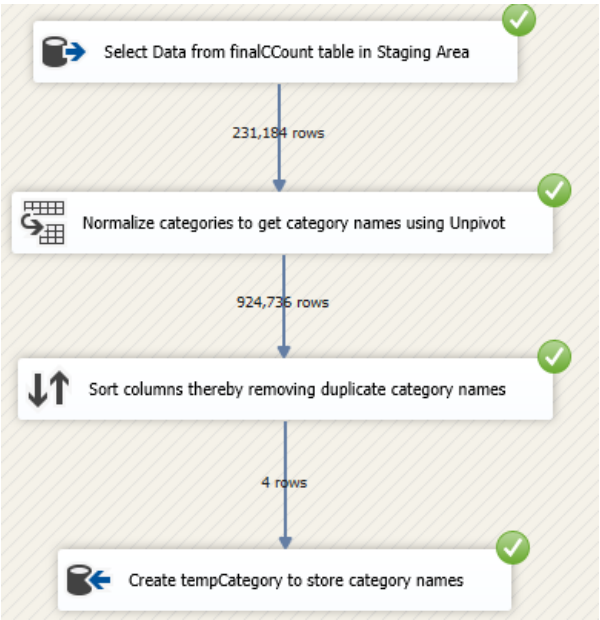

Figure 4.3-16: Creating temporary Category table in Staging Area

Snapshot of tempCategory Table in Staging Area

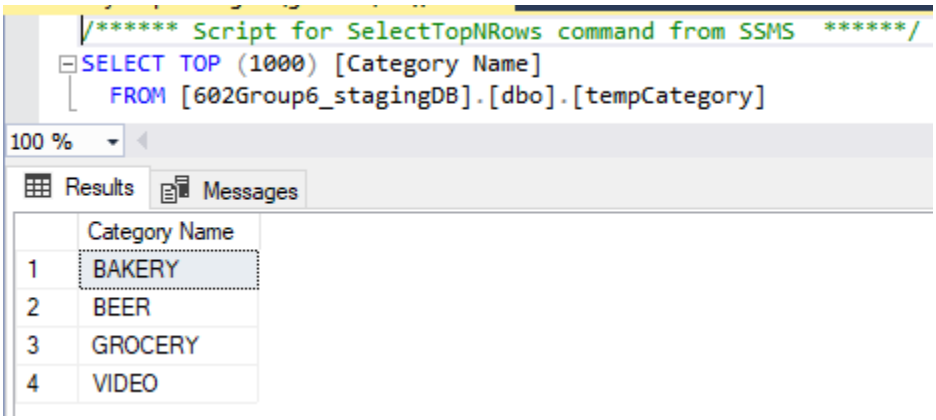

Figure 4.3-17: Snapshot of tempCategory Table in Staging Area

4.3.1.4 Extraction and Transformation for Product Dimension

Data Extraction

Extract UPCCHE to Staging Area

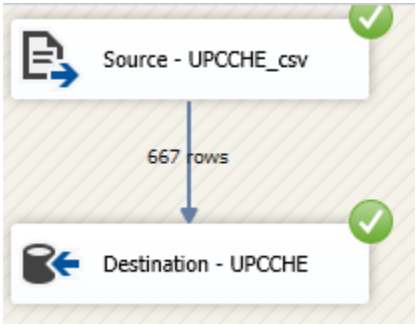

Figure 4.3-18: Extract UPCCHE to Staging Area

UPCCHE file in Staging Area

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [COM_CODE]
      ,[UPC]
      ,[DESCRIP]
      ,[SIZE]
      ,[CASE]
      ,[NITEM]
  FROM [602Group6_stagingDB].[dbo].[UPCCHE]
```

100 %

Results | Messages

	COM_CODE	UPC	DESCRIP	SIZE	CASE	NITEM
1	153	1570077445	CTY LN COLBY MILD	10 OZ	12	5390551
2	153	1570077446	CTY LN CHEDDAR MILD	10 OZ	12	5390571
3	153	1570077454	CTY LN CHEDDAR MED/S	8 OZ	12	5390591
4	153	1570077455	CTY LN CHEDDAR SHARP	8 OZ	12	5390511
5	153	1570077456	CTY LN SWISS OLD WOR	10 OZ	12	5390531
6	153	1570077458	COUNTY LINE/COLBY JA	10 OZ	12	5390621
7	153	1570077501	CTY LN MONTEREY JACK	10 OZ	12	5390491
8	153	1570077503	~COUNTY LINE MUENSTE	8 OZ	12	5390441
9	153	1570077507	~COUNTY LINE MOZZ CH	8 OZ	12	5390421
10	153	1570077508	COUNTY LINE EXTRA SH	8 OZ	12	5390401

Figure 4.3-19: Snapshot of UPCCHE file in Staging Area

Data Transformation

1. Convert data types as shown in figure

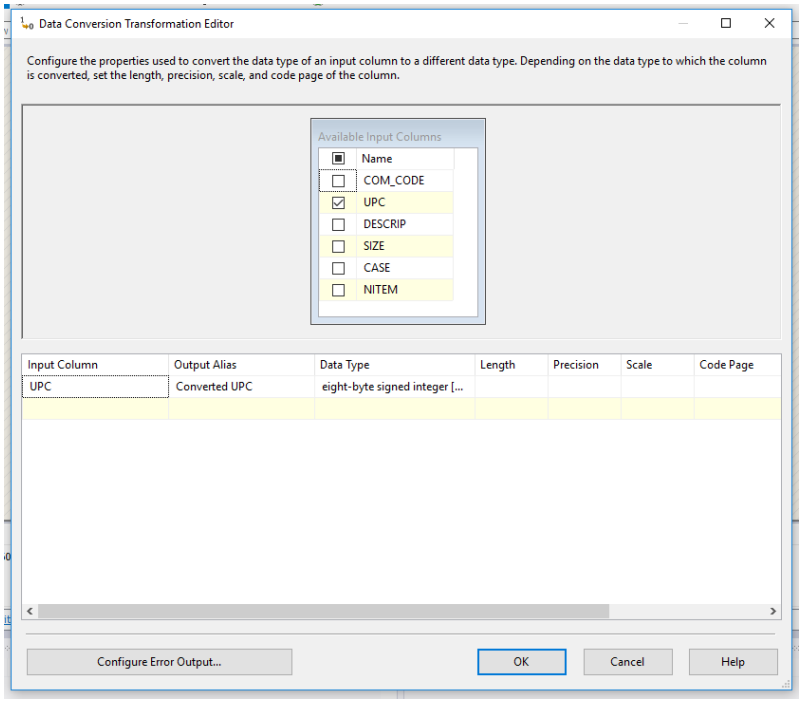

Figure 4.3-20: Converting data types for UPCCHE

2. Store the clean data in UPC table in staging area

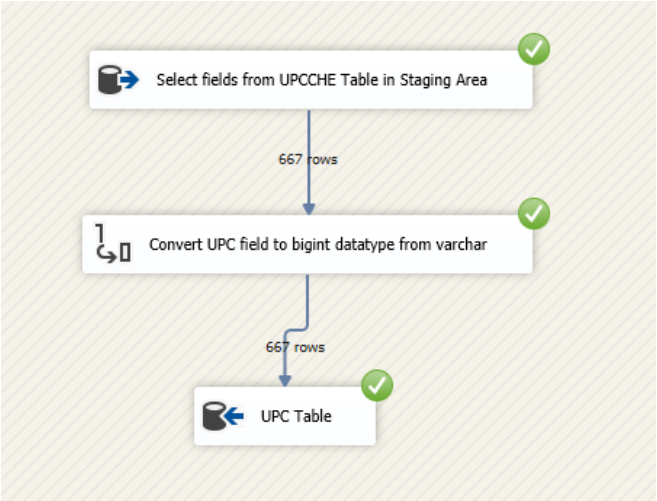

Figure 4.3-21: Store the clean data in UPC table in staging area

4.3.1.5 Extraction and Transformation for factProductMargin Fact Table

1. Select Fields from finalMovement table

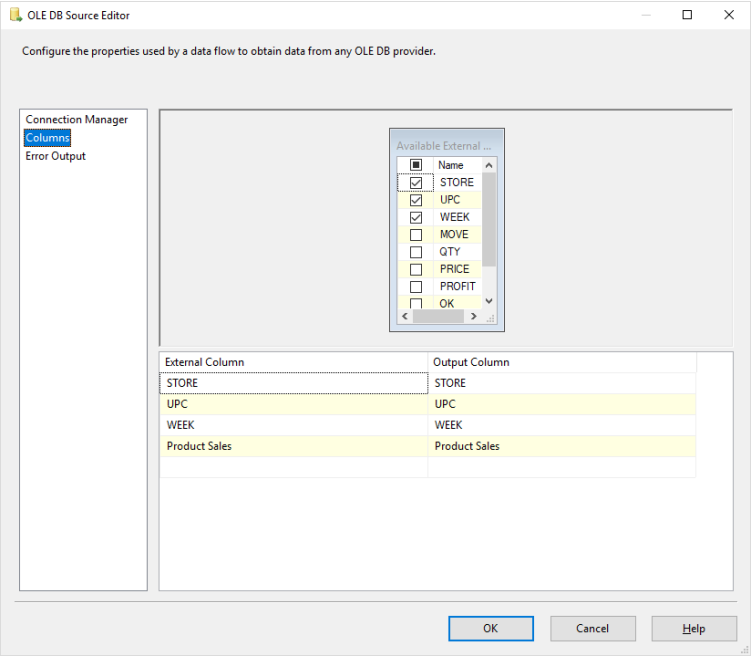

Figure 4.3-22: Select Fields from finalMovement table

3. Perform Product Lookup

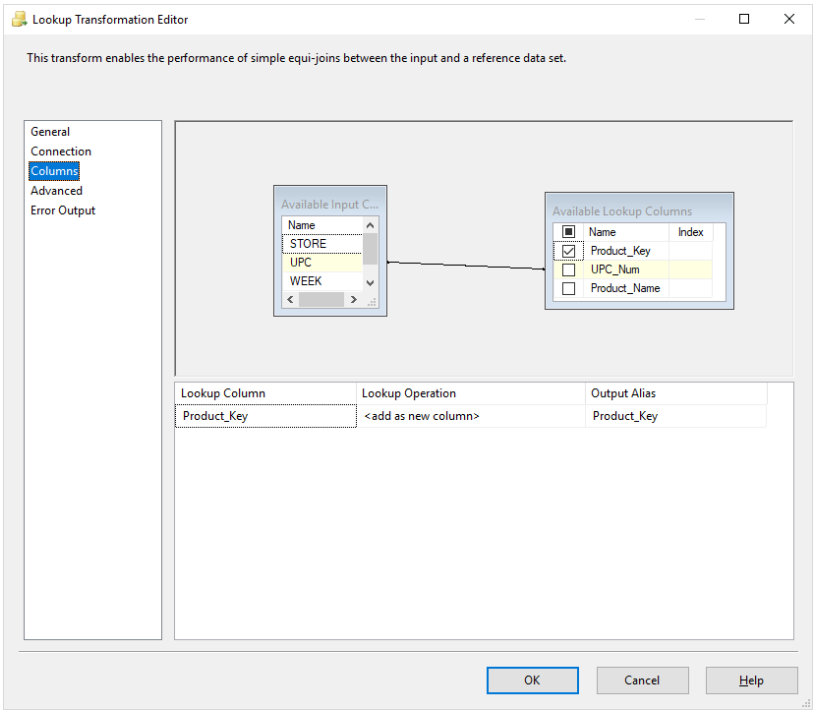

Figure 4.3-23: Perform Product Lookup

3. Perform Time Lookup

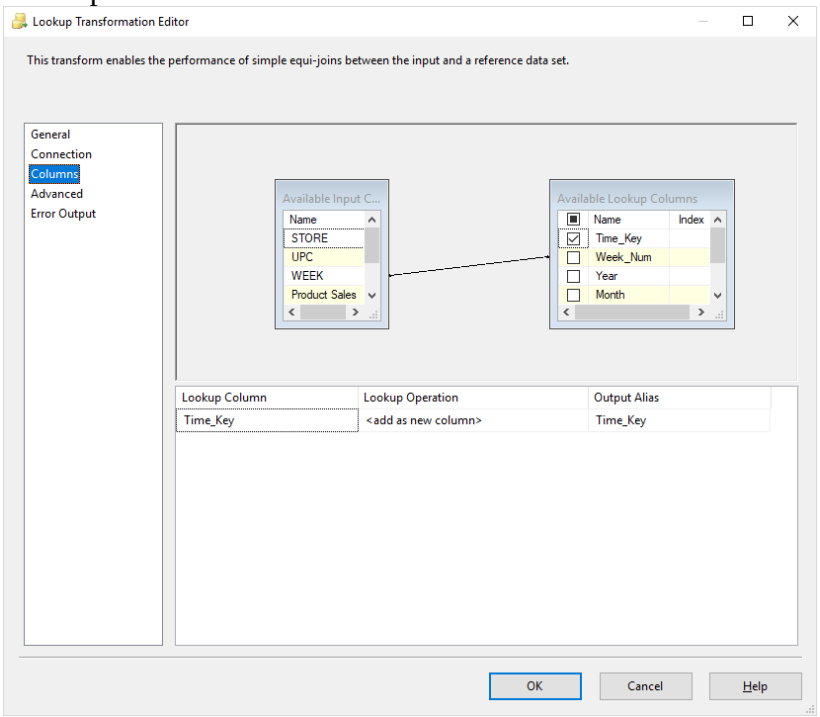

Figure 4.3-24: Perform Time Lookup

4. Aggregate Fields

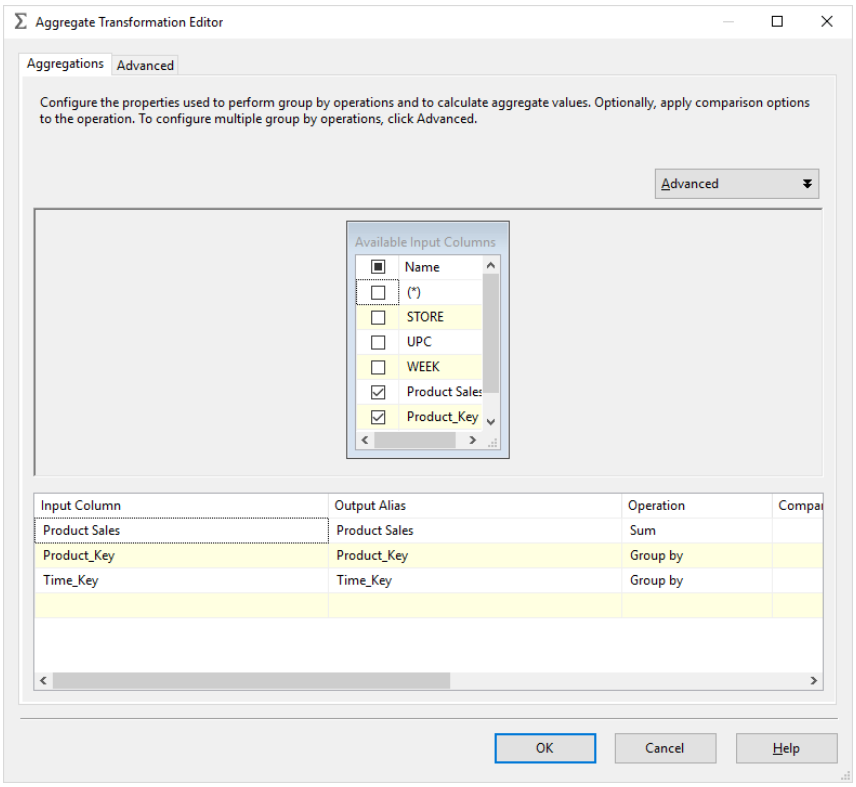

Figure 4.3-25: Aggregate Fields

4.3.1.6 Extraction and Transformation for factCategorySales Fact Table

1. Select all fields from finalCCount table

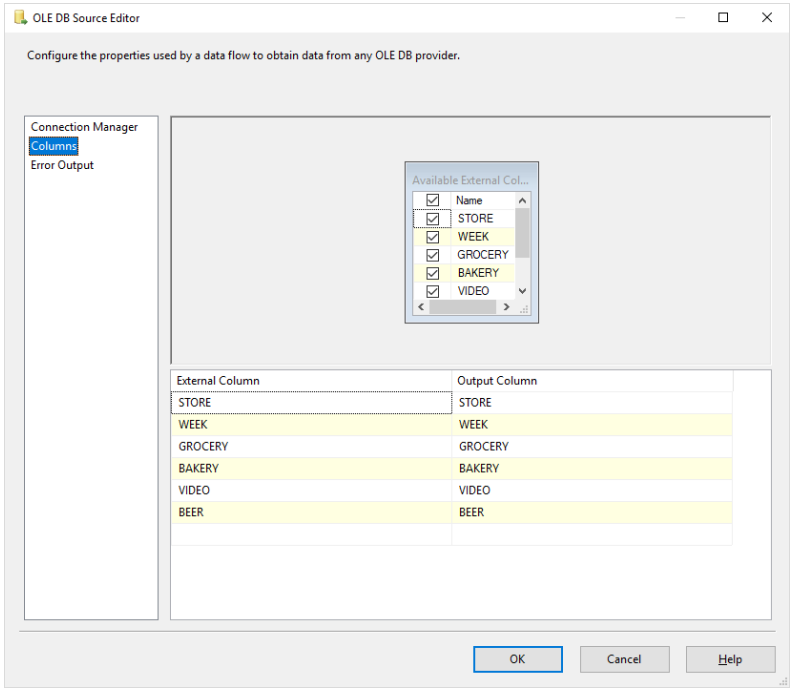

Figure 4.3-26: Select all fields from finalCCount table

2. Use Unpivot transformation to get Categories Name

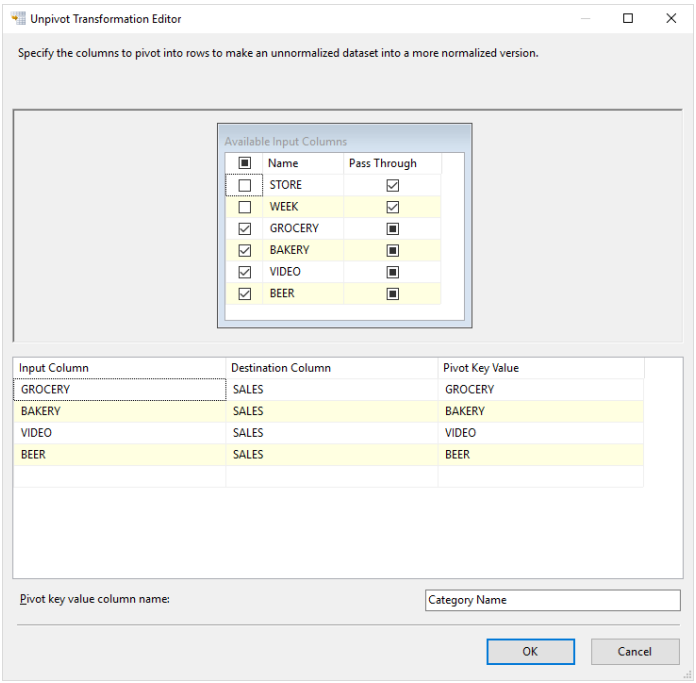

Figure 4.3-27: Use Unpivot transformation to get Categories Name

3. Perform Category Look Up

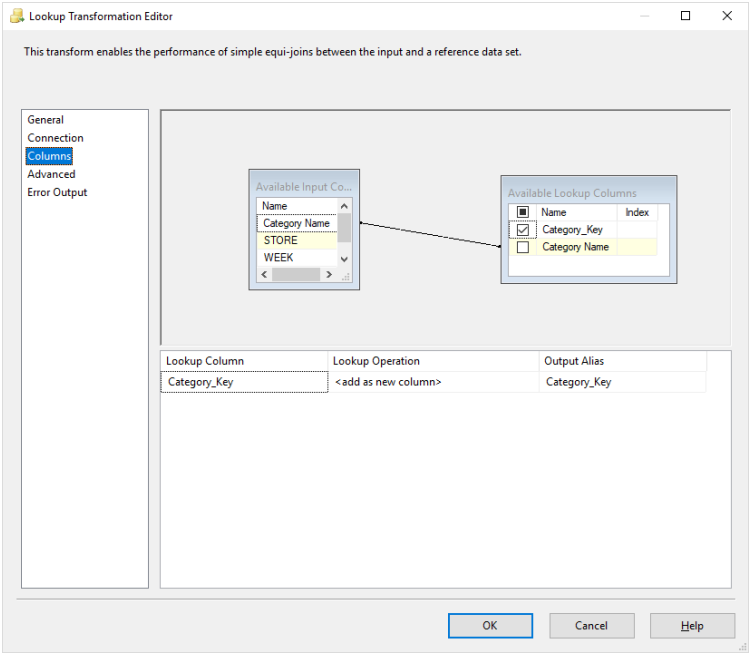

Figure 4.3-28: Perform Category Look Up

4. Perform Store Lookup

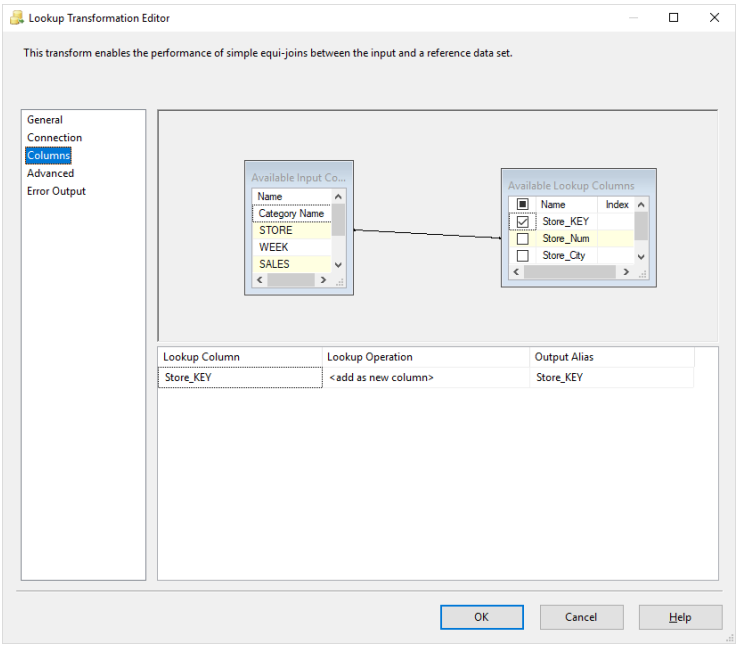

Figure 4.3-29: Perform Store Lookup

5. Perform Time Lookup

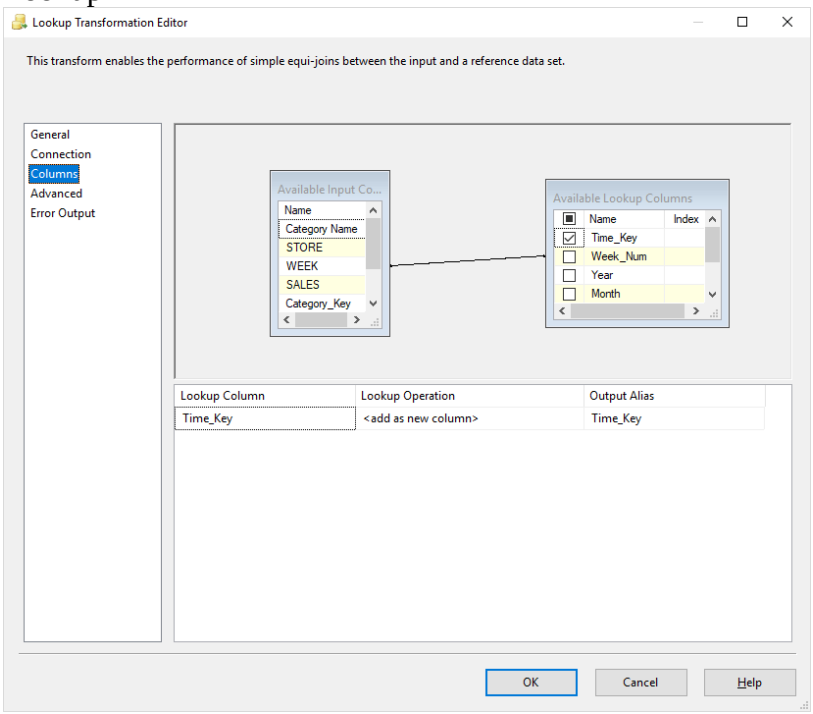

Figure 4.3-30: Perform Time Lookup

6. Aggregate Sales Data

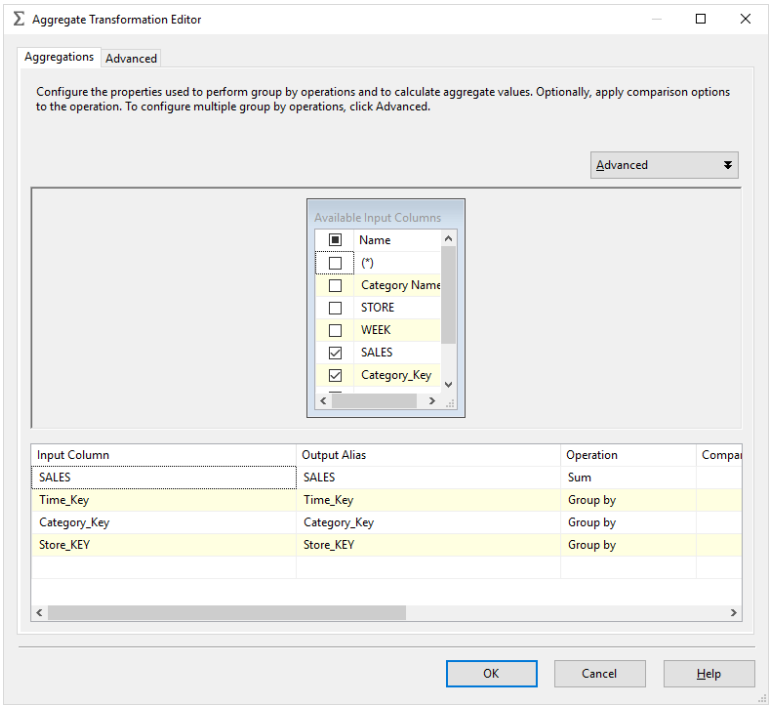

Figure 4.3-31: Aggregate Sales Data

4.3.2 Loading of Dimension and Fact Tables from Staging to DW Area

4.3.2.1 Data Loading

Data loading is the process of moving the data from staging area to the Data Warehouse. The phrases related to data loading are:

- Initial load: This is when the Data warehouse tables are loaded for the first time. This was the type of Data Loading we performed in our project.
- Incremental load: This is when any update, delete or insert occurs in the source database and those changes are applied to the data warehouse in a period manner. We did not perform this loading in our project.
- Full Refresh: This is when the entire data warehouse data is deleted and reloaded with fresh data. We did not require a full refresh anytime while data loading in this project.

The data loading process for our project has been depict in Figure 4.3-24.

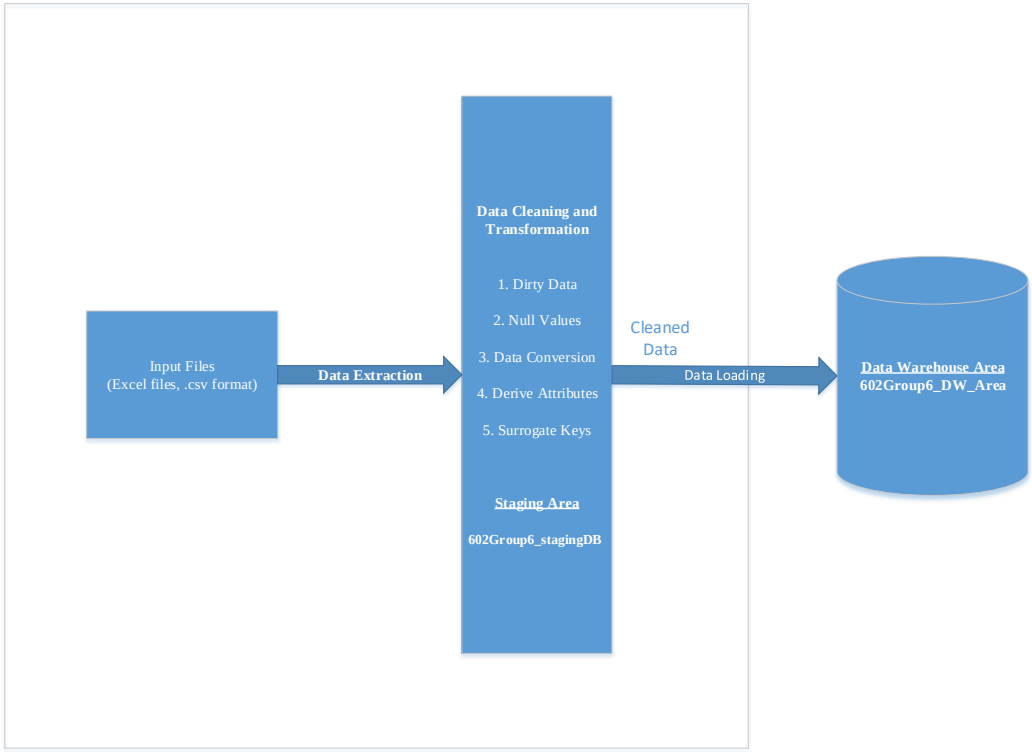

Figure 4.3-32: Data Loading Process

4.3.2.2 Data Loading for Store Dimension

1. Create Dimension table dimStore along with a surrogate key for each record in Data Warehouse

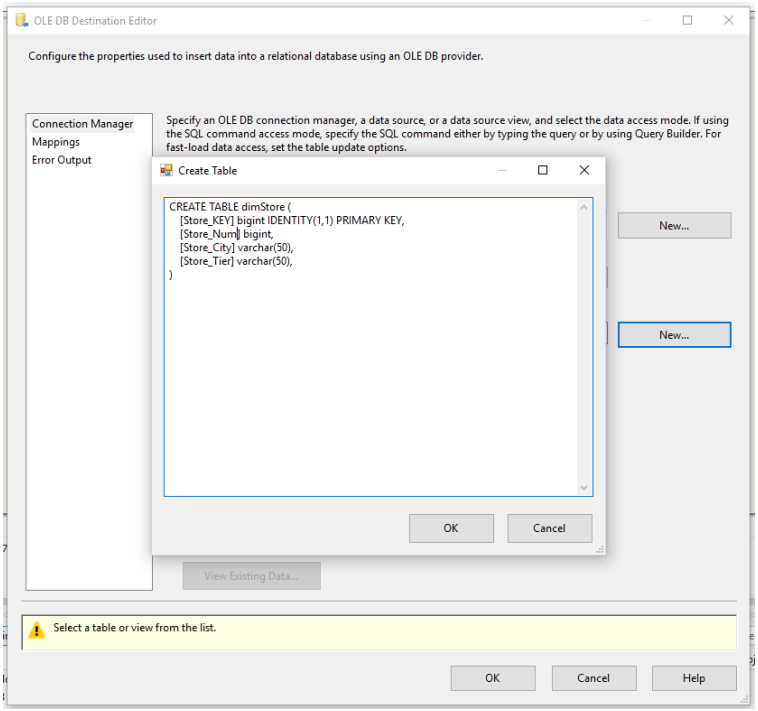

Figure 4.3-33: Creating dimStore

2. Map fields into dimStore

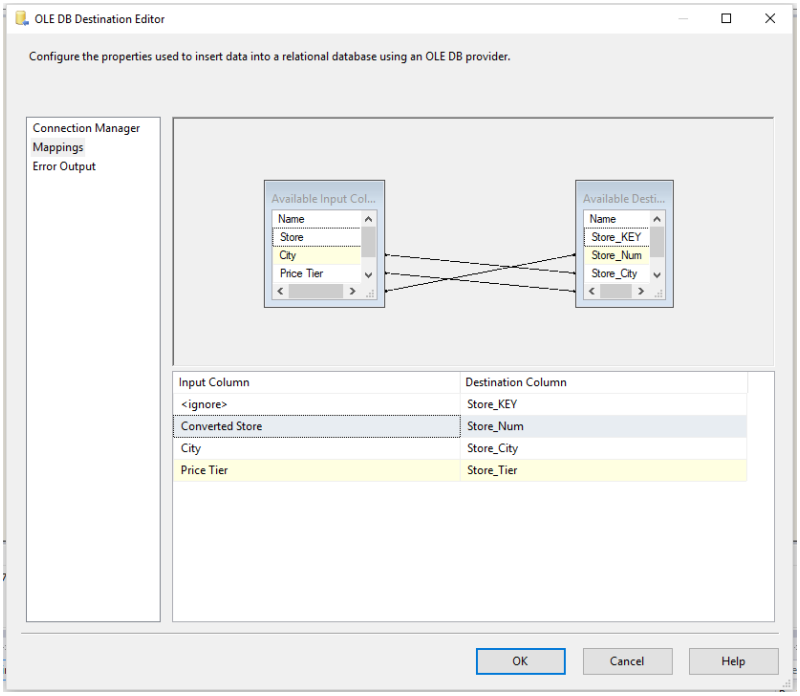

Figure 4.3-34: Mapping attributes from Staging area to dimStore

3. Execute the package

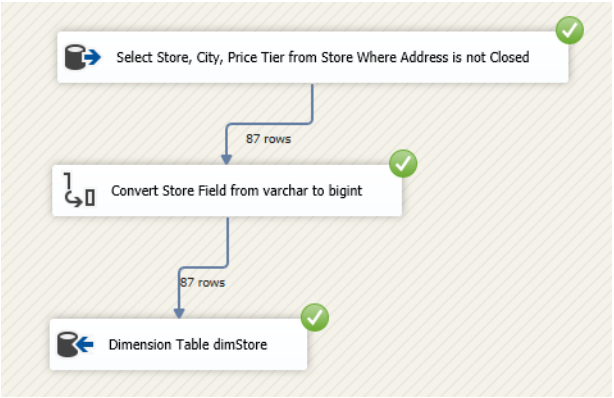

Figure 4.3-35: Execute dimStore Package

Snapshot of dimStore Dimension Table

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [Store_KEY]
      ,[Store_Num]
      ,[Store_City]
      ,[Store_Tier]
  FROM [602Group6_DW_Area].[dbo].[dimStore]
```

100 %

Results | Messages

	Store_KEY	Store_Num	Store_City	Store_Tier
1	175	2	River Forest	High
2	176	5	Palatine	Medium
3	177	8	Oak Lawn	Low
4	178	9	Morton Grove	Medium
5	179	12	Chicago	High
6	180	14	Glenview	High
7	181	18	River Grove	Low
8	182	21	Hanover Park	CubFighter
9	183	28	Mt. Prospect	Medium
10	184	32	Park Ridge	High

Figure 4.3-36: Snapshot of dimStore Dimension Table

4.3.2.3 Data Loading for Time Dimension

1. Create dimTime along with a surrogate key for each record

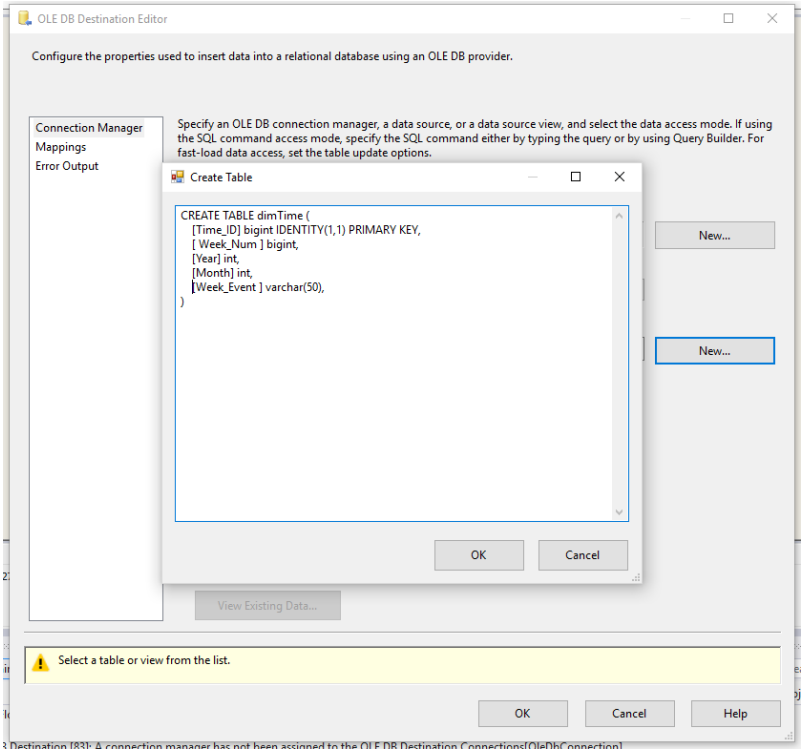

Figure 4.3-37: Creating dimTime with surrogate key

2. Map fields into dimTime

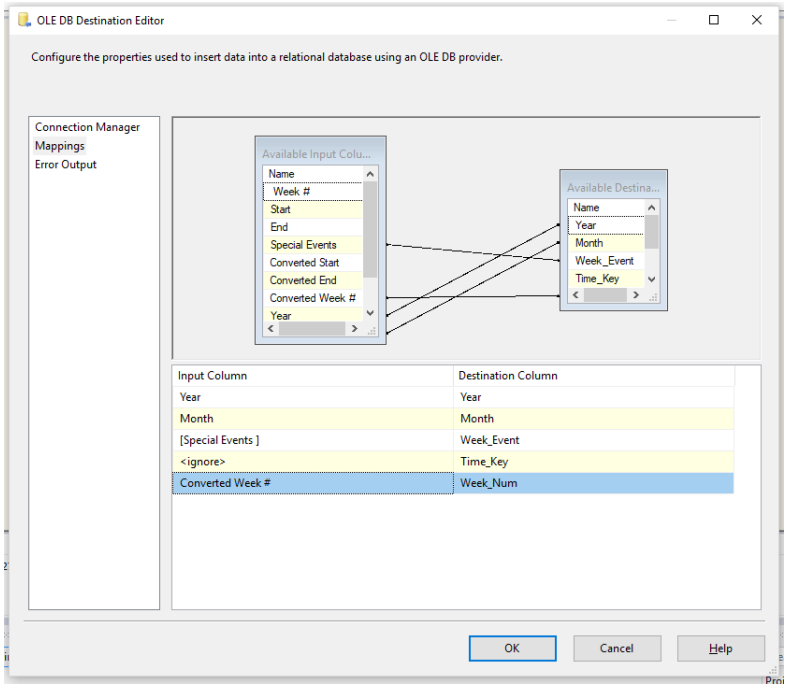

Figure 4.3-38: Mapping data from staging area to dimTime

3. Execute the package

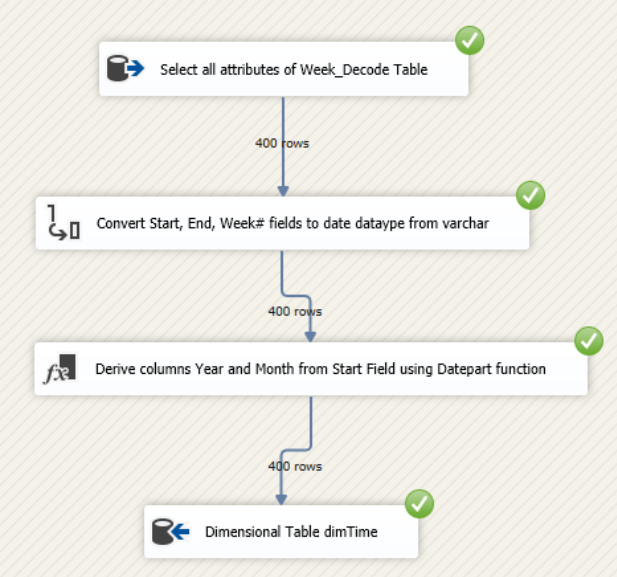

Figure 4.3-39: Executing dimTime package

Snapshot of dimTime Dimension Table

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [Time_Key]
      ,[Week_Num]
      ,[Year]
      ,[Month]
      ,[Week_Event ]
  FROM [602Group6_DW_Area].[dbo].[dimTime]
```

100 %

Results | Messages

	Time_Key	Week_Num	Year	Month	Week_Event
1	1601	1	1989	9	
2	1602	2	1989	9	
3	1603	3	1989	9	
4	1604	4	1989	10	
5	1605	5	1989	10	
6	1606	6	1989	10	
7	1607	7	1989	10	Halloween
8	1608	8	1989	11	
9	1609	9	1989	11	
10	1610	10	1989	11	

Figure 4.3-40: Snapshot of dimTime Dimension Table in Data Warehous Area

4.3.2.4 Data Loading for Category Dimension

Load dimCategory table data from tempCategory table in Staging Area

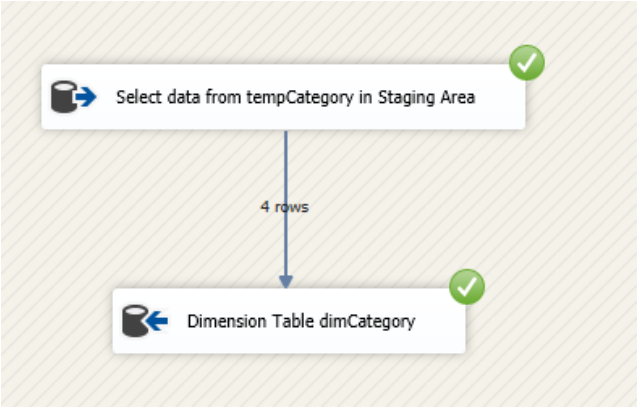

Figure 4.3-41: Load dimCategory table data from tempCategory table in Staging Area

Snapshot of dimCategory Dimension Table

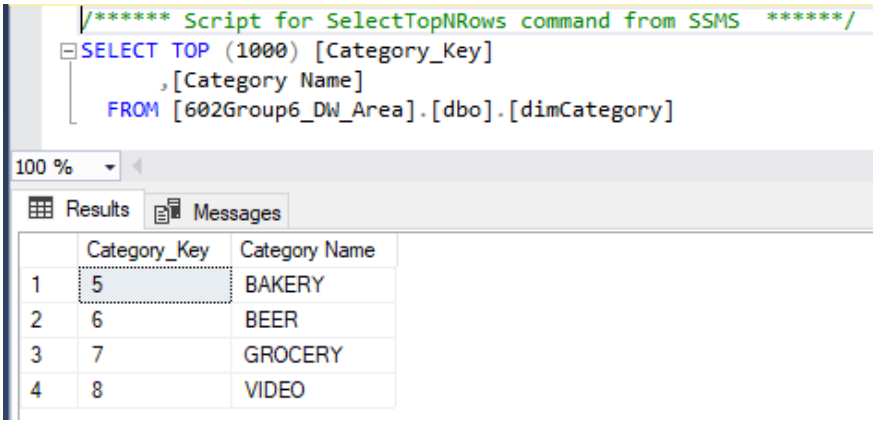

Figure 4.3-42: Snapshot of dimCategory Dimension Table

4.3.2.5 Data Loading for Product Dimension

1. Create dimProduct in DataWarehouse

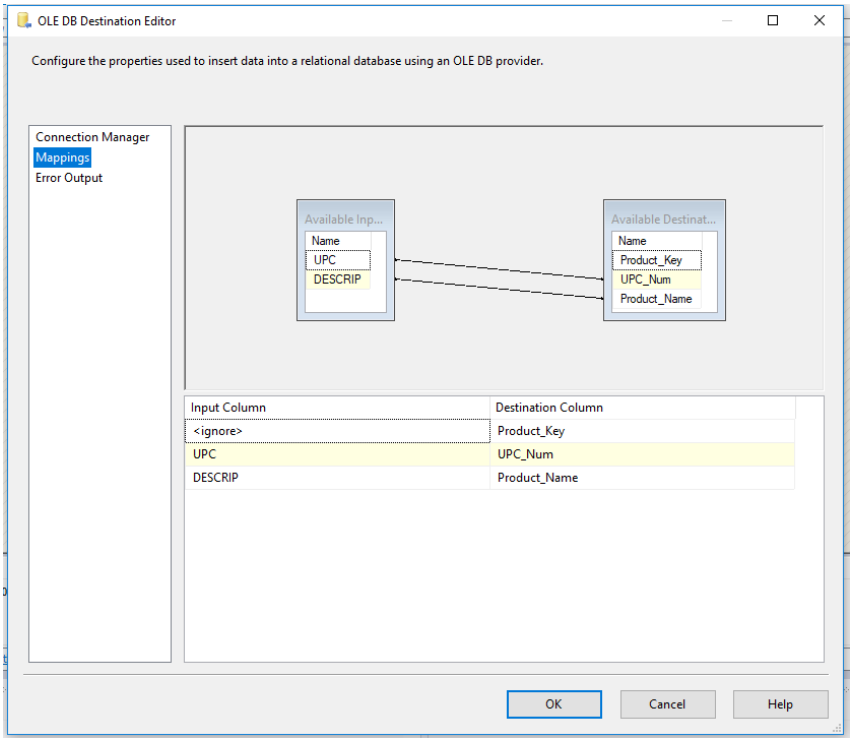

Figure 4.3-43: Mapping attributes from UPCCHE to dimProduct

2. Load the data from UPC table in Staging Area

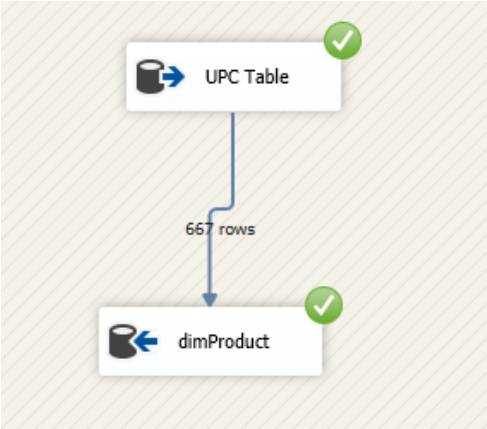

Figure 4.3-44: Loading data in dimProduct

Snapshot of dimProduct Dimension Table

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [Product_Key]
      ,[UPC_Num]
      ,[Product_Name]
  FROM [602Group6_DW_Area].[dbo].[dimProduct]
```

100 %

Results | Messages

	Product_Key	UPC_Num	Product_Name
1	1	1570077445	CTY LN COLBY MILD
2	2	1570077446	CTY LN CHEDDAR MILD
3	3	1570077454	CTY LN CHEDDAR MED/S
4	4	1570077455	CTY LN CHEDDAR SHARP
5	5	1570077456	CTY LN SWISS OLD WOR
6	6	1570077458	COUNTY LINE/COLBY JA
7	7	1570077501	CTY LN MONTEREY JACK
8	8	1570077503	~COUNTY LINE MUENSTE
9	9	1570077507	~COUNTY LINE MOZZ CH
10	10	1570077508	COUNTY LINE EXTRA SH

Figure 4.3-45: 3. Snapshot of dimProduct Dimension Table

4.3.2.6 Data Loading for factProductMargin Fact Table

Execute the package

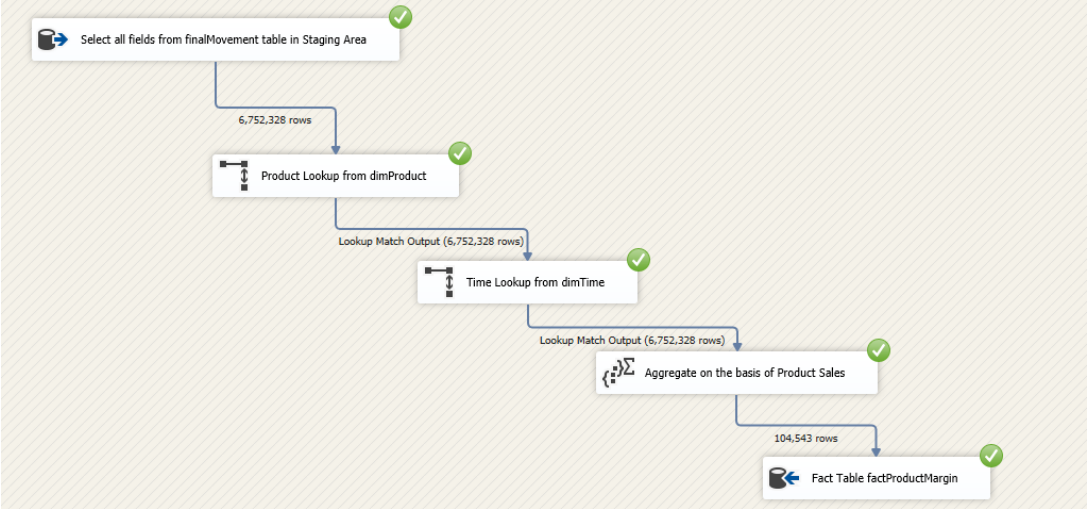

Figure 4.3-46: Product Margin Fact Table

Snapshot of factProductMargin Fact Table

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [Product Sales]
      ,[Product_Key]
      ,[Time_Key]
  FROM [602Group6_DW_Area].[dbo].[factProductMargin]
```

100 %

Results | Messages

	Product Sales	Product_Key	Time_Key
1	1091.73003935814	1	1601
2	1831.60006833076	2	1601
3	931.000011444092	3	1601
4	2668.25989532471	4	1601
5	2973.16015386581	5	1601
6	1064.80002403259	6	1601
7	1122.30004024506	7	1601
8	1197.66998291016	23	1601
9	7.44999980926514	24	1601
10	3.49000000953674	25	1601

Figure 4.3-47: Snapshot of factProductMargin Fact Table

4.3.2.7 Data Loading for factProductMargin Fact Table

Execute the package

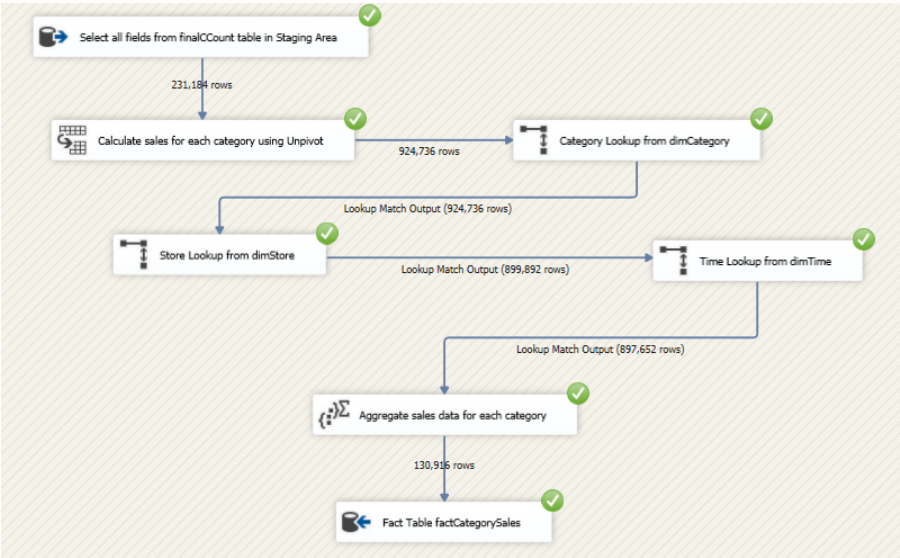

Figure 4.3-48: Category Sales fact table

Snapshot of factCategorySales Fact Table

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP (1000) [SALES]
      ,[Time_Key]
      ,[Category_Key]
      ,[Store_KEY]
  FROM [602Group6_DW_Area].[dbo].[factCategorySales]
```

100 %

Results | Messages

	SALES	Time_Key	Category_Key	Store_KEY
1	11660.70000	1601	5	175
2	11895.54000	1601	5	176
3	13221.39000	1601	5	177
4	12215.39000	1601	5	178
5	7459.01000	1601	5	179
6	11430.82000	1601	5	180
7	15610.09000	1601	5	181
8	7542.47000	1601	5	182
9	7630.52000	1601	5	183
10	16710.37000	1601	5	184

Figure 4.3-49: Snapshot of fact Category Sales

4.3.3 SQL Queries Used in SSIS

4.3.3.1 Table Creation Commands in SSIS

The following commands were used during creation OLE DB Destination through SSIS for Staging Area and Data Warehouse. These commands need to be entered when we click on New button in OLE DB Destination.

dimCategory

```
CREATE TABLE dimCategory(
        Category_Key bigint IDENTITY(1,1) PRIMARY KEY,
        Category Name varchar,
)
```

dimProduct

```
CREATE TABLE dimProduct(
        Product_Key bigint IDENTITY(1,1) PRIMARY KEY,
        UPC_Num bigint,
        Product_Name varchar,
)
```

dimStore

```
CREATE TABLE dimStore(
        [Store_KEY] bigint IDENTITY(1,1) PRIMARY KEY,
        [Store_Num] bigint,
        [Store_City] varchar(50),
        [Store_Tier] varchar(50),
)
```

dimTime

```
CREATE TABLE dimTime(
        [Time_ID] bigint IDENTITY(1,1) PRIMARY KEY,
        [Week_Num] bigint,
        [Year] int,
        [Month] int,
        [Week_Event] varchar(50),
)
```

finalCCOUNT

```
CREATE TABLE finalCCOUNT(
```

```
	[STORE] bigint,
	[WEEK] bigint,
	[GROCERY] numeric(18, 5),
	[BAKERY] numeric(18, 5),
	[VIDEO] numeric(18, 5),
	[BEER] numeric(18, 5)
```

finalMovement

```
CREATE TABLE finalMovement(
	[STORE] bigint,
	[UPC] bigint,
	[WEEK] bigint,
	[MOVE] bigint,
	[QTY] bigint,
	[PRICE real,
	[PROFIT] real,
	[OK] bigint,
	[Product Sales] real
)
```

tempCategory

```
CREATE TABLE tempCategory(
	[Category Name] varchar(255)
)
```

UPC

```
CREATE TABLE UPC(
	[COM_CODE] varchar(50),
	[UPC] bigint,
	[DESCRIP] varchar(50),
	[SIZE] varchar(50),
	[CASE] varchar(50)
)
```

4.3.3.2 Data Cleaning Commands in SSIS

Commands to clean CCount file

Command Location – Control Flow

Delete Negative Values from Week

```
DELETE FROM CCOUNT
WHERE	(["WEEK"] LIKE '-%') OR
```

(["WEEK"] LIKE '.')

Remove Null Values from All fields

```
DELETE
FROM CCOUNT
WHERE (["WEEK"] IS NULL) OR
(["VIDEO"] IS NULL) OR
(["GROCERY"] IS NULL) OR
(["BEER"] IS NULL) OR
(["DATE"] IS NULL) OR
(["STORE"] IS NULL) OR
(["BAKERY"] IS NULL)
```

Remove Invalid Store Values

```
DELETE FROM CCOUNT
WHERE       (NOT (["STORE"] IN
                  (SELECT       Store
                   FROM          Store)))
```

Correct Date

```
UPDATE CCOUNT
SET ["DATE"] = REPLACE(["DATE"], "'", '')
```

```
DELETE FROM CCOUNT
WHERE      ["DATE"] LIKE  'a'
```

Commands to clean Store file

Command Location – Data Flow

Select only stores which are not “Closed”

```
SELECT [Store], [City], [Price Tier]
FROM   Store
Where   (Address NOT LIKE 'CLOSED')
```

Commands to clean Done-CHE table which is used to create UPC file

Command Location – Control Flow

Remove Records with Invalid “OK”

```
DELETE
FROM [DONE-WCHE]
WHERE ["OK"] LIKE '0'
```

Remove Records with No Move

```
DELETE
FROM [DONE-WCHE]
WHERE ["MOVE"] LIKE '0'
```

4.3.4 Staging Area Tables after Removal of Temporary Tables

This is the final view of tables in staging area after removing all the temporary tables from staging area.

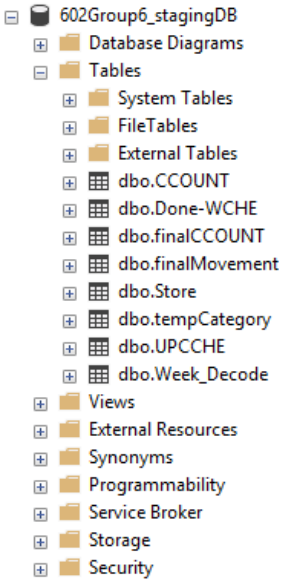

Figure 4.3-50: Staging area tables

4.3.5 Data Warehouse Area Tables

This is the final view of tables in data warehouse area after removing all the temporary tables from data warehouse area.

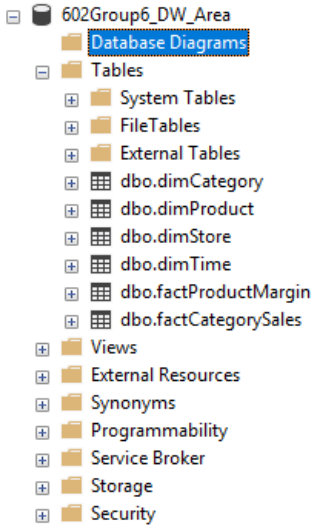

Figure 4.3-51: Data Warehouse Area tables

4.3.6 Table Structure for factCategorySales Fact Table

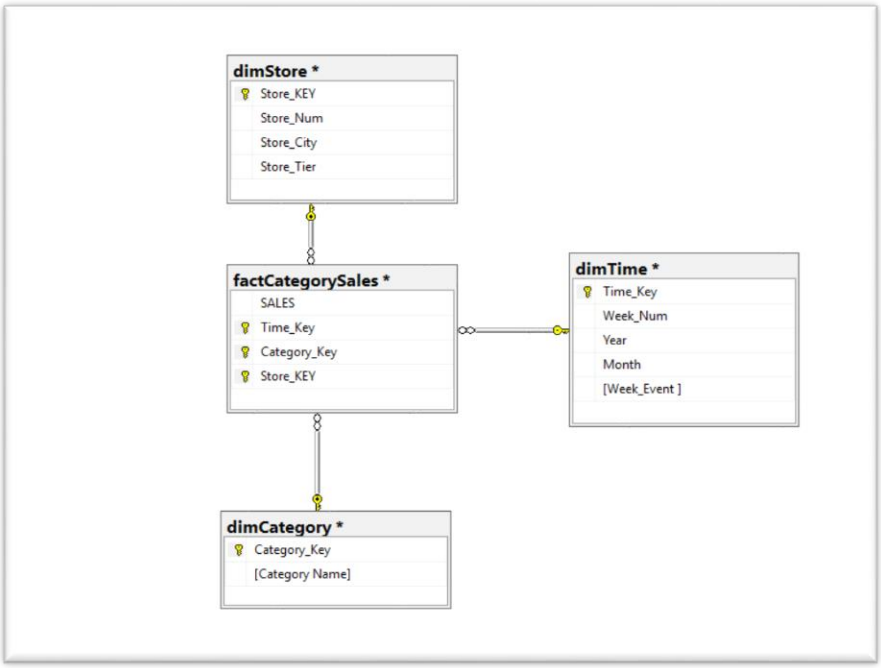

Figure 4.3-52: Table Structure for factCategorySales

4.3.7 Table Structure for factProductMargin Fact Table

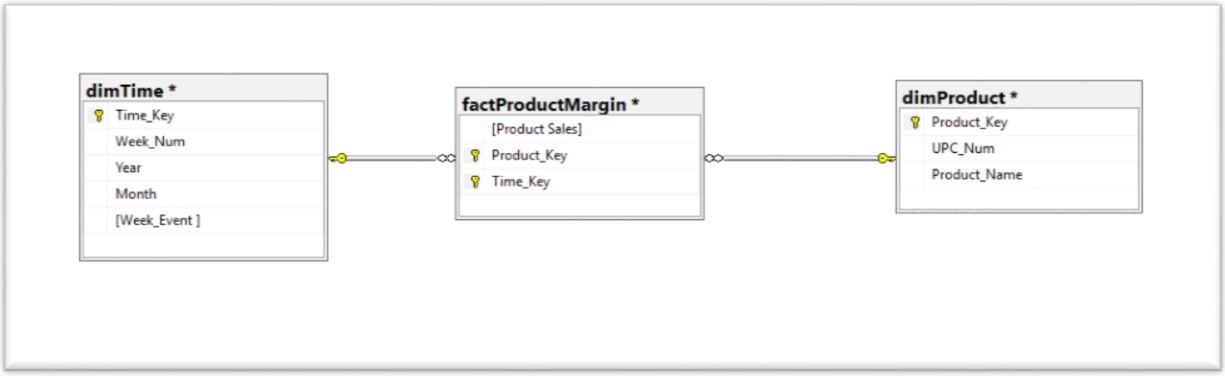

Figure 4.3-53: Table Structure for factProductMargin

5 BI reporting

5.1 Reporting Plan

This section lays out what reports will be generated for each of the business questions. We have decided to use Report Builder for Question 1 and Question 4. We will use SSRS only for Question

3 and SSAS only for Question 5. For SSAS only (Question 5) graph will be created through Report Builder. Report for Question 2 will be generated through SSRS on top on SSAS. The reporting plan is further explained in sections 5.1.1 to 5.1.4.

5.1.1 Target reports satisfying business questions

Q1. How are the sales of Bakery products affected during Christmas period over the period 1989 -1996?

Report generated from Report Builder 3.0

To solve this business question, we used the Category Name as BAKERY from the Category dimension, week number from Time dimension and week event as "Christmas" from the Time dimension and mapped them with the Category Sales fact table to get the Category sales. This will give the trend of Bakery sales during the Christmas week over the period 1989 -1996. We have made use of a bar chart to show the bakery trends.

Q2. Are there any periods during the years 1989-1995 where the video sales are higher or lower?

Report generated from SSRS on top of SSAS

This business question was answered by creating a cube in the Analytics Services of the SQL Server data tools. We have used the Category Sales fact table along with Time dimension table and Category dimension table. An SQL query was created to use Year and month from Time dimension, and Category Sales from the Category Sales fact table while putting the filter on Category name to get the values for Video category from category dimension. After successful deployment of the cube on server, we used SSRS on top of SSAS cube to generate a report. We used a horizontal bar chart to visualize higher or lower video sales during the years 1989-1995.

Q3. What is the profit trend of cheese during 1991 to 1995?

Report generated from SSRS alone

We used SSRS only to answer this business question. We used the product name from Product dimension, Years from 1991-1995 from the Time dimension and mapped them with the product Sales fact table to get the product sales. The sales from the fact table will be the trend of Cheese sales during 1991 to 1995. We have used SQL query and SSRS to generate the report for this business question. We have made use of bar chart to show the trend.

Q4. How the trend of grocery sales varies across different stores in Chicago?

Report generated from Report Builder 3.0
In order to answer this business question, we have used category name a GROCERY from category dimension, Story city as Chicago from Store dimension, Store number from Store dimension, year from time dimension and mapped them with the Category Sales fact table to get

the category sales. This will give the trend of grocery sales across different stores in Chicago. We have used two pie charts for this question- one showing the percentage of grocery sales across stores in Chicago and the other shoeing Dollar value of grocery sales across stores in Chicago.

Q5. Detect trends in Beer Sales encompassing the entire duration that look for high period of sales. Detect peaks not only during designated holidays but also on events that are not in the Week_Decode table.

Analysis of the cube created from SSAS only

We used SSAS only to answer this business question. We deployed a cube which consists of year and month from time dimension and mapped them with the Category Sales fact table to get the category sales. The Category Name was used as BEER from category dimension to group the category sales. We used SSRS to analyze the cube. We made use of two bar charts to show the Beer sale trends- one shows the monthly beer sale trends and the other shows the beer sales by year and month.

5.1.2 Mappings from the independent data marts to the report attributes

Q1 How are the sales of Bakery products affected during Christmas period over the period 1989 -1996?

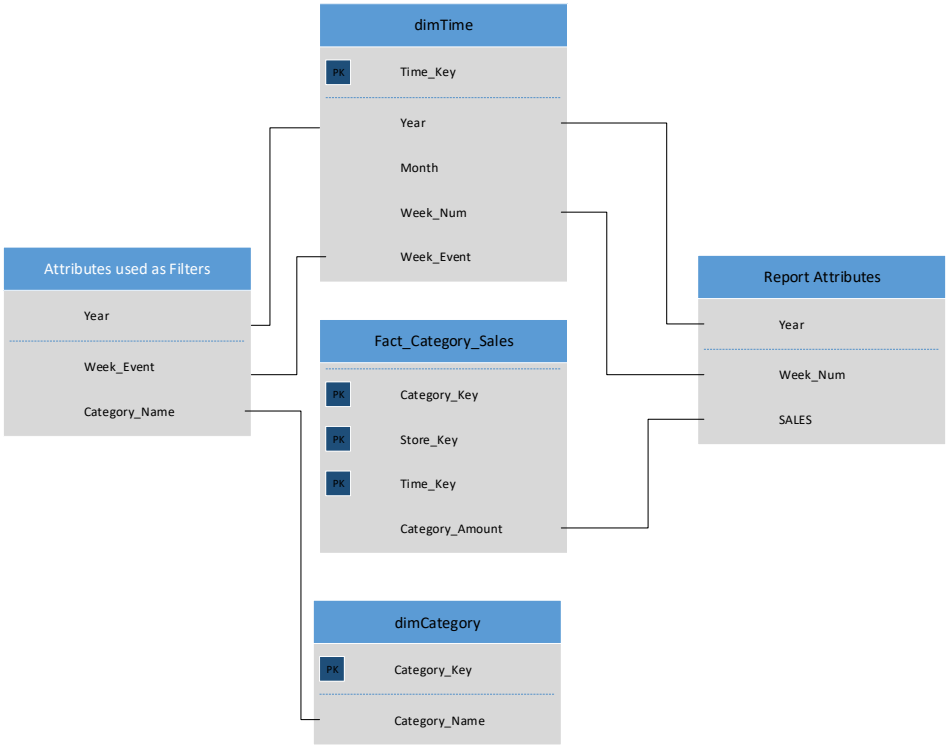

Figure 5.1-1: Data Mapping for Business Question 1

Q2 Are there any periods during the years 1989-1995 where the video sales are higher or lower?

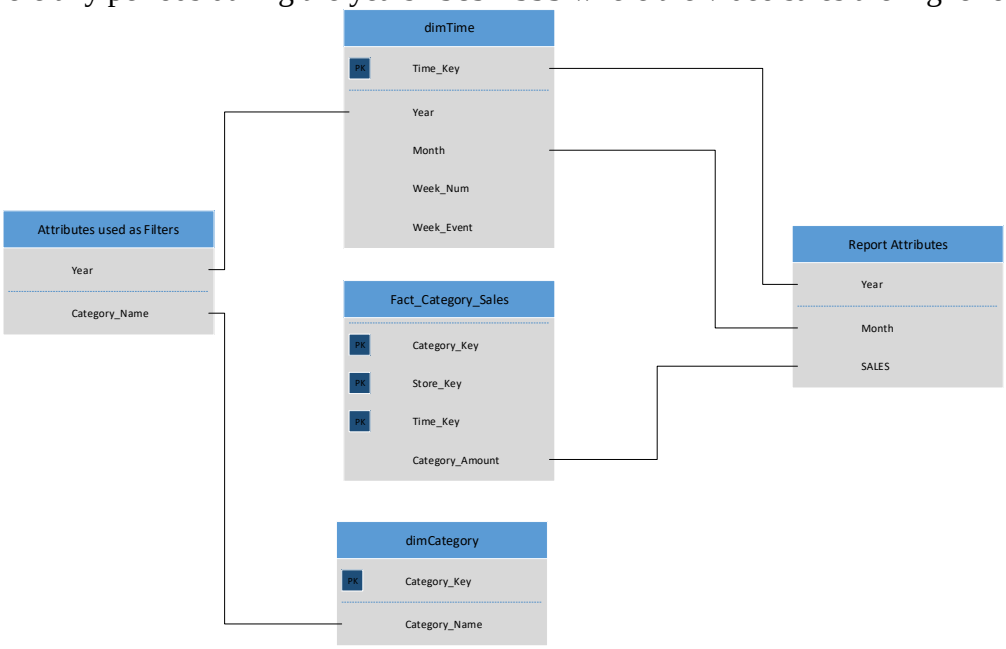

Figure 5.1-2: Data Mapping for Business Question

Q3 What is the profit trend of cheese during 1991 to 1995?

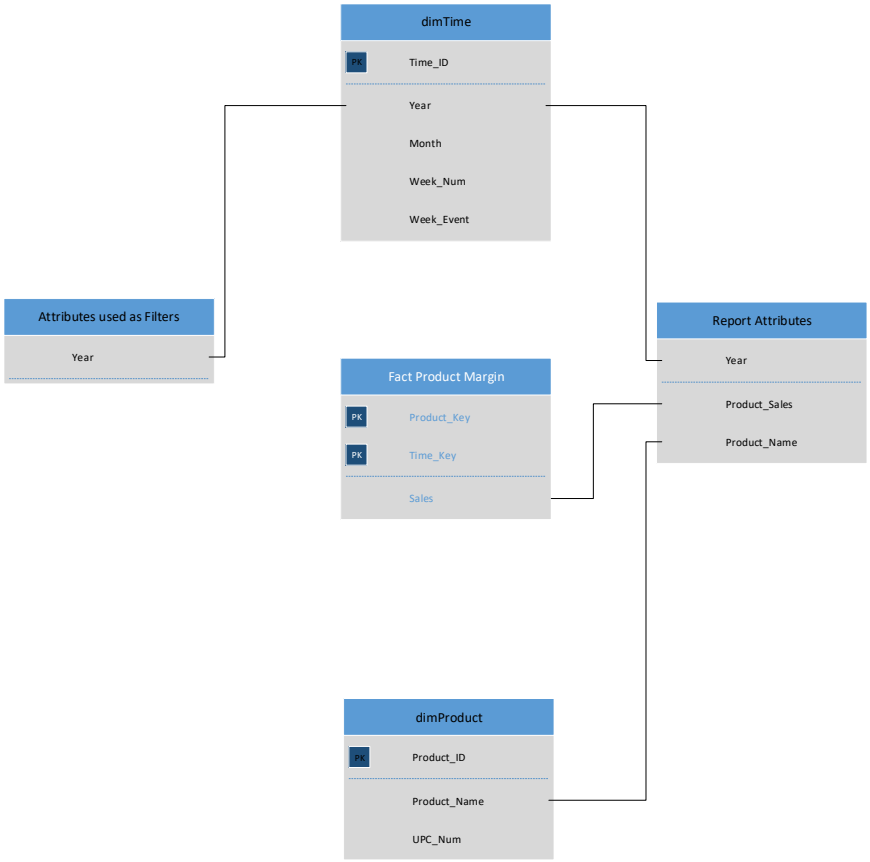

Figure 5.1-3: Data Mapping for Business Question 3

Q4 How the trend of grocery sales varies across different stores in Chicago?

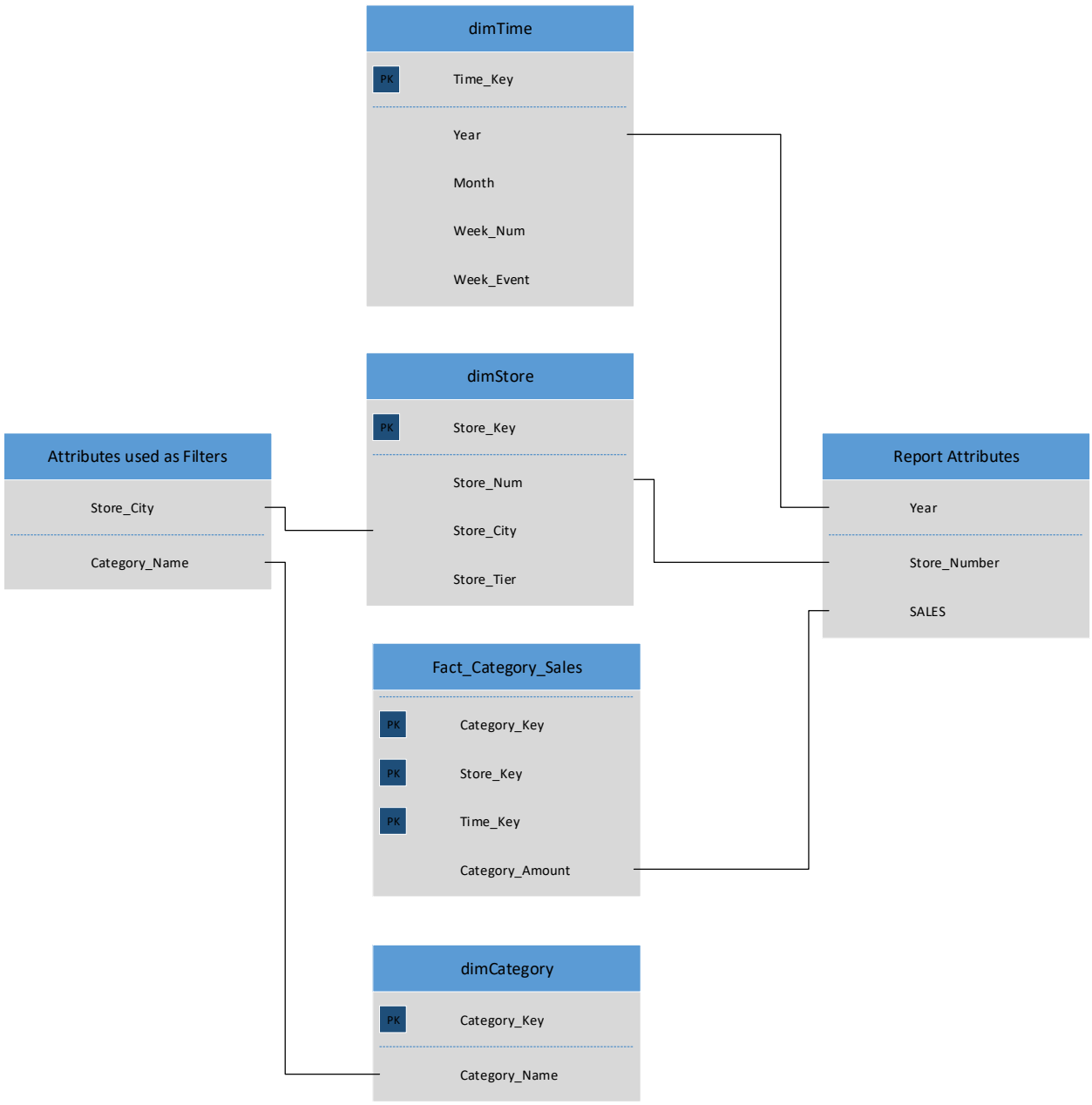

Figure 5.1-4: Data Mapping for Business Question 5

Q5 Detect trends in Beer Sales encompassing the entire duration that look for high period of sales. Detect peaks not only during designated holidays but also on events that are not in the Week_Decode table.

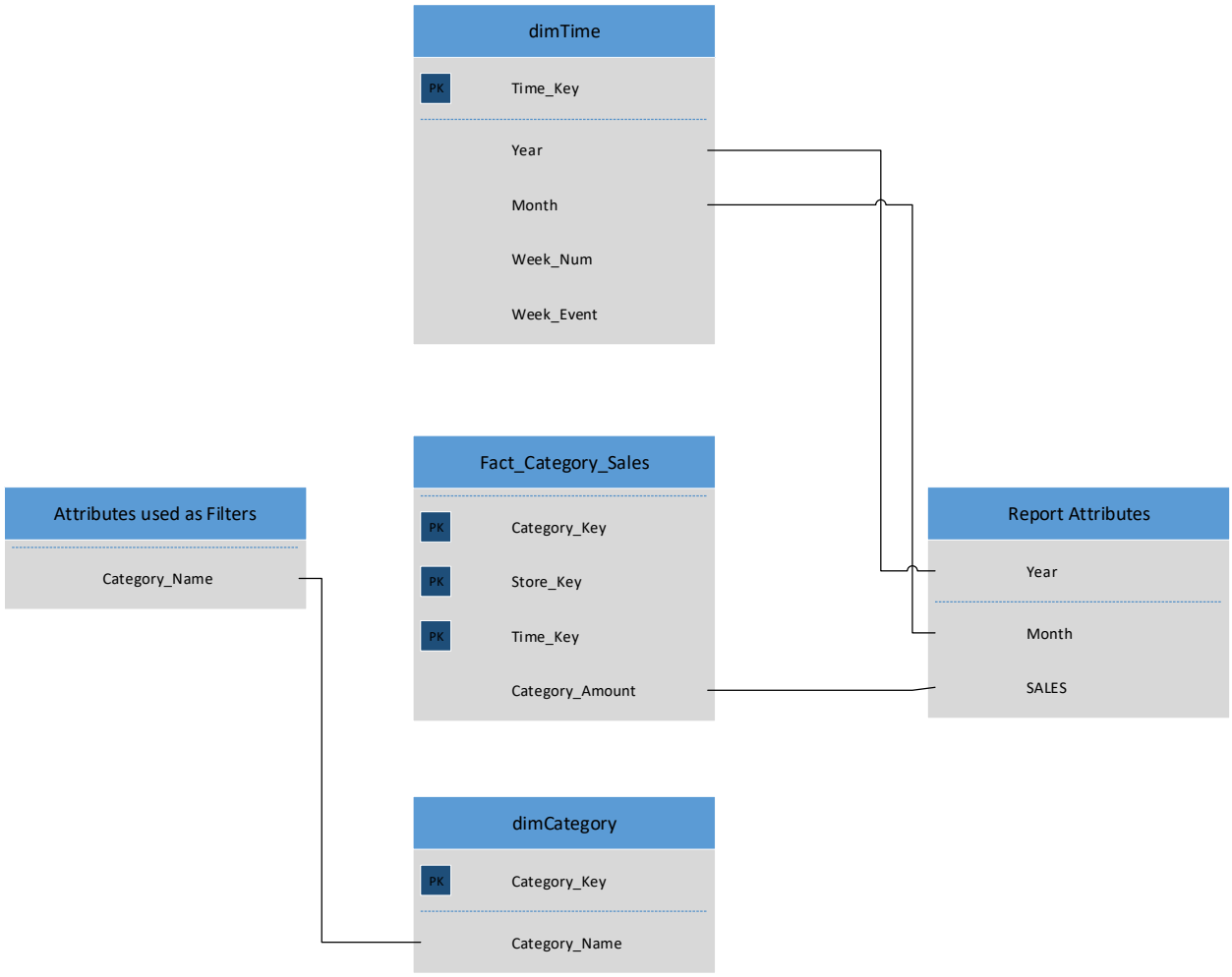

Figure 5.1-5: Data Mapping for Business Question 5

5.1.3 Report Templates

Q1. How are the sales of Bakery products affected during Christmas period over the period 1988-1996?
We will generate a horizontal bar chart for Bakery Sales during Christmas over the period 1989 – 1996. This report will be created in Report Builder.

Q2. Are there any periods during the years 1989-1995 where the video sales are higher or lower?
We will generate a vertical bar chart for Video Sales which will depict the breakdown of video sales on monthly basis during the period 1989-1995. This will be augmented by a report with drilldown available on year and month basis for actual sales. This report will be created through SSRS on top of SSAS.

Q3. What is the profit trend of cheese during 1991 to 1995?
A horizontal bar chart will be generated to understand cheese profit trend from 1991 to 1995. This chart will be augmented by a report will drilldown enabled on Product Name, Year for Product Profit. This report will be created through SSRS.

Q4. How the trend of grocery sales varies across different stores in Chicago?
Report builder will be used to generate two pie charts that will depict the share of each store in sales. One pie chart will depict the percentage sales of each store vs total sales and another pie chart will show amount of sales done by each store to understand its overall sales.

Q5. Detect trends in Beer Sales encompassing the entire duration that look for high period of sales. Detect peaks not only during designated holidays but also on events that are not in the Week_Decode table.
SSAS will used to generate trend for beer sales on monthly and yearly basis which will assist in pointing out all the peak periods for Beer Sales in the past.

5.1.4 Reporting Tools

In order to create reports for answering the business questions, we have employed reporting tools such as SSRS, SSAS, a combination of SSAS and SSRS as well as Report Builder 3.0 with the following breakdown for each of the business questions:

Business Question	*Reporting Tool Employed*
Question 1	**Report Builder 3.0**
Question 2	**SSRS on SSAS**
Question 3	**SSRS**
Question 4	**Report Builder 3.0**
Question 5	**SSAS**

Table 5.1.1 Reporting Plan

5.2 Report building using Reporting Tools

Report Server: http://infodata16.mbs.tamu.edu/reportserver
Report Folder: 602_Group6_ProjectReport

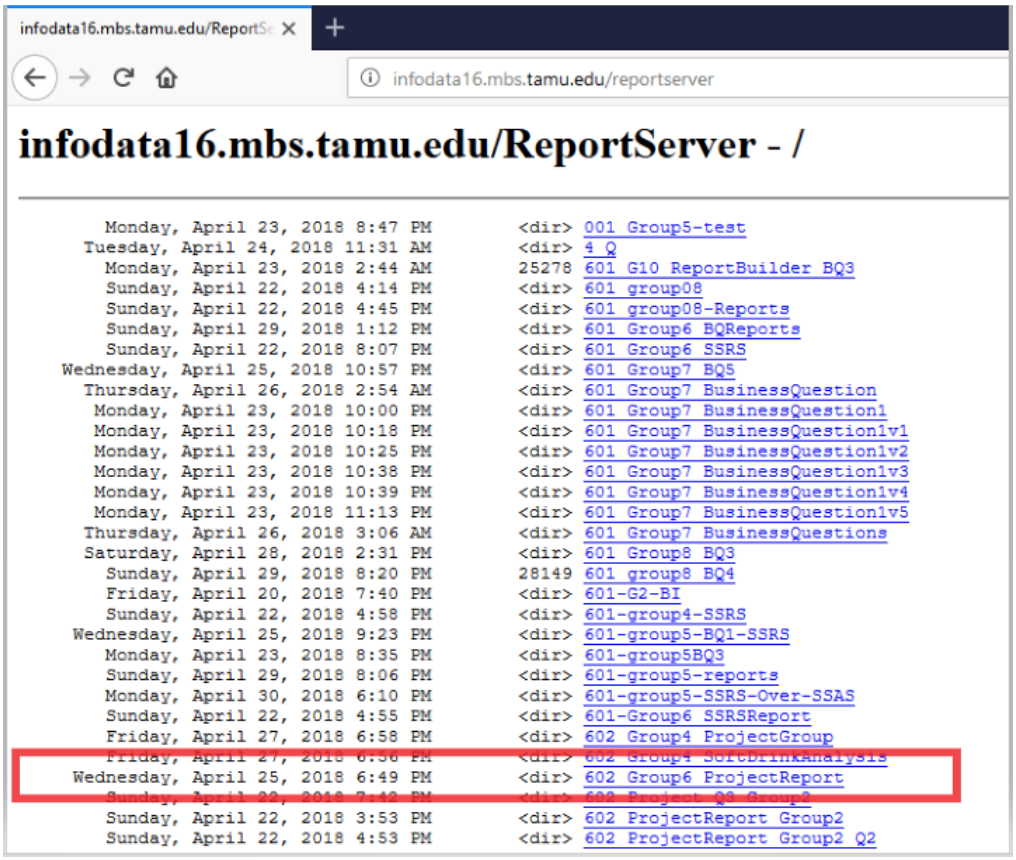

Figure 5.2-1: Report Folder deployed on Report Server

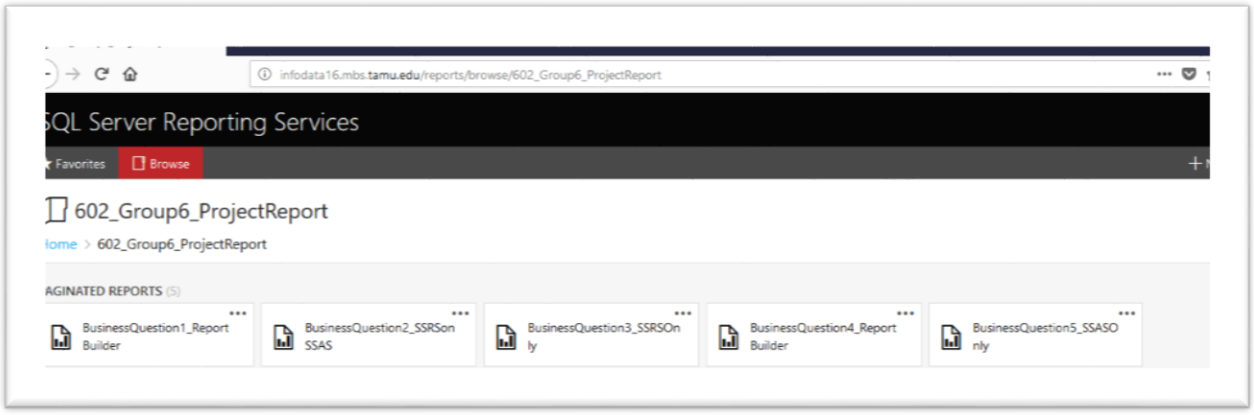

Figure 5.2-2: 5 Business Questions in Report Folder

5.2.1 Report using ReportBuilder: Business Question 1

How are the sales of Bakery products affected during Christmas period over the period 1989 - 1996?

Reporting tool employed: ReportBuilder 3.0

Screenshots of report building

1. Create Data Source Connection

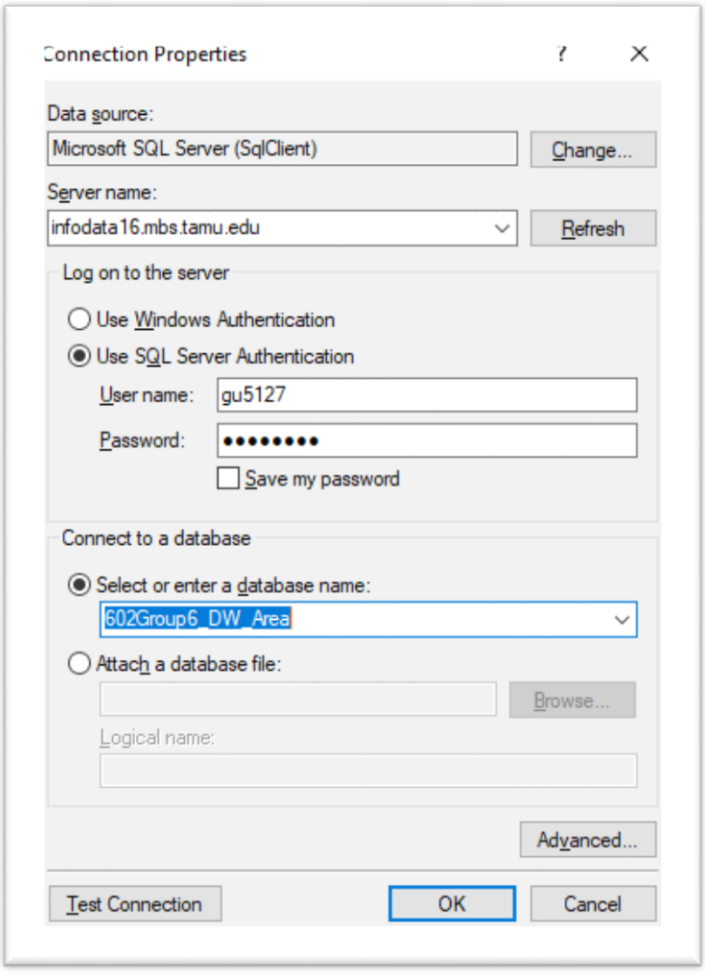

Figure 5.2-3 Connecting To Data Base

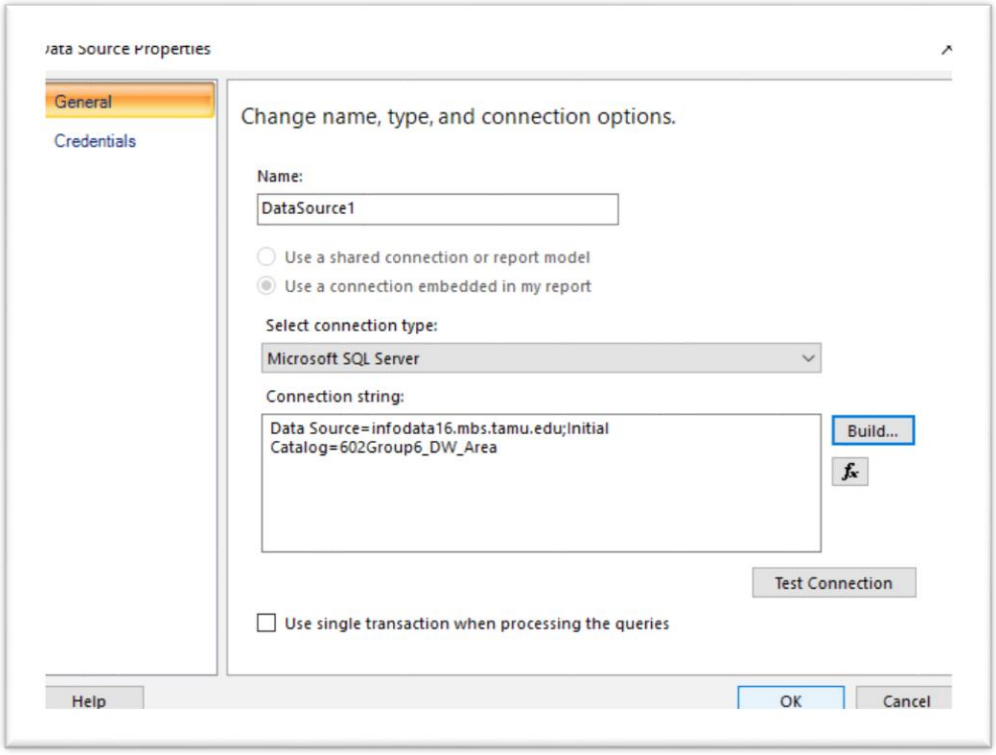

Figure 5.2-4: Configuring data source properties

2. Design Query

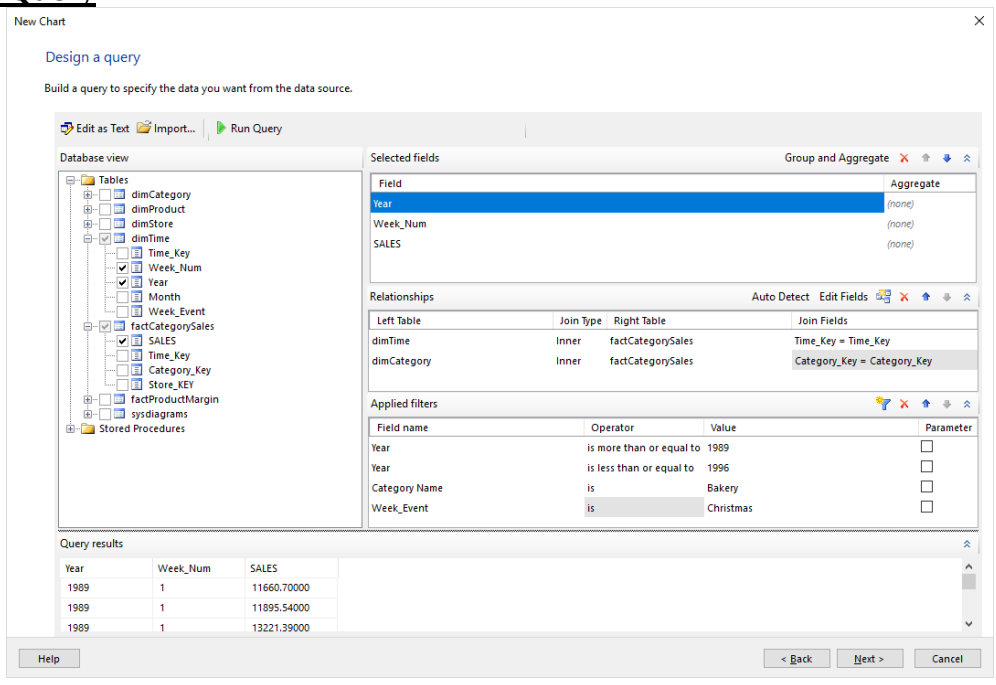

Figure 5.2-5: Designing Query

3. Arrange Chart Fields

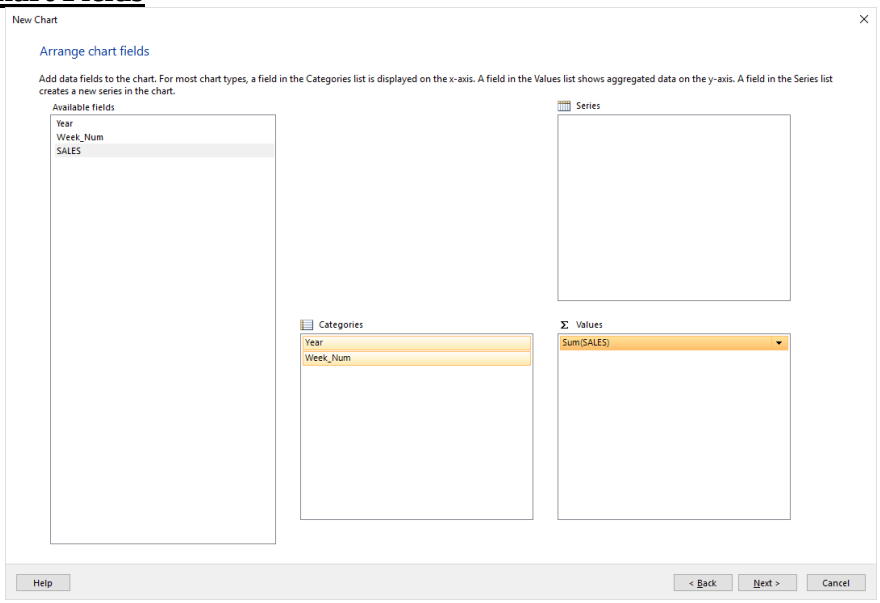

Figure 5.2-6: Arranging Chart values

4. Edit Formatting of Chart

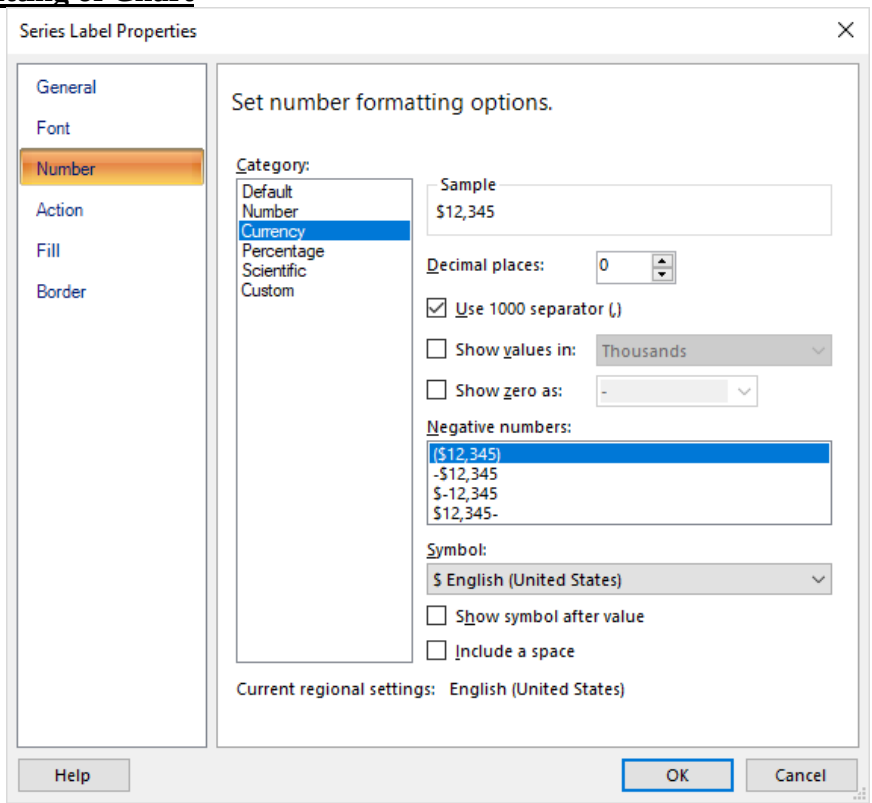

Figure 5.2-7: Edit Formatting of chart

3. **Final Bar Chart showing trend of Bakery Sales**

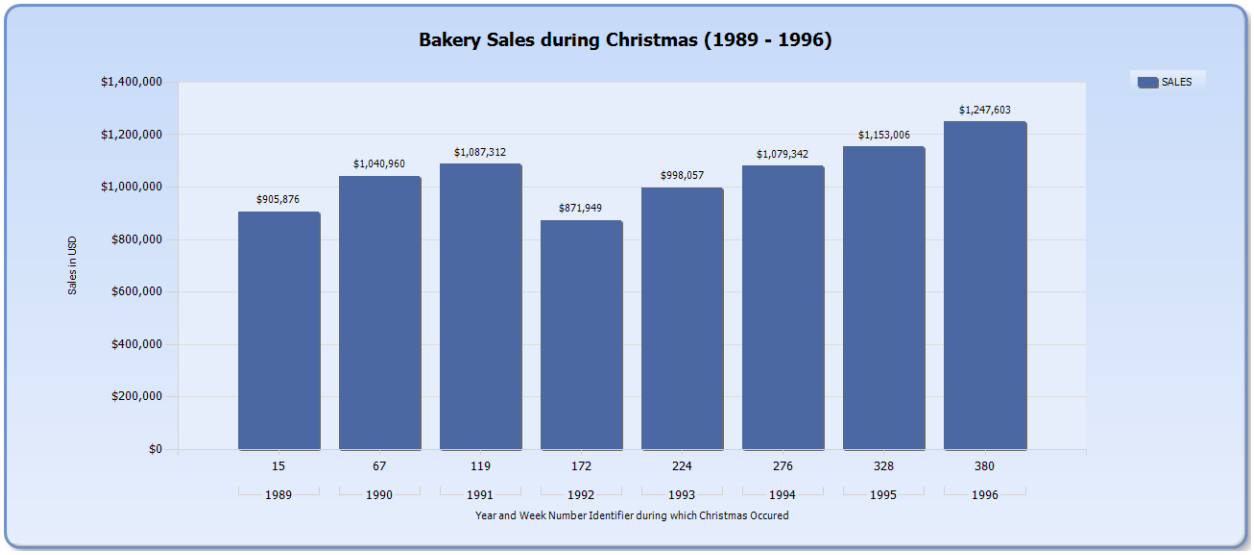

Figure 5.2-8: Final Bar Chart showing trend of Bakery Sales

6. Report Deployed on Report Server

Figure 5.2-9: Report Deployed on Report Server

Conclusion

We used Report Builder to answer Question 1. Our aim was to find if there was any trend in bakery sales during Christmas over the period 1989 to 1996. The graph for Bakery Sales during Christmas

shows that there is an increasing trend in bakery sales during Christmas period. Consumers are now purchasing more and more bakery products from these retail stores. This information is useful as managers can safely assume that the trend i.e. increase in bakery sales will continue in 1997 too, in absence of any isolated or unlikely event, as the sales have increased continuously for the past 4 years. Business managers can utilize this information to take decisions such as to introduce new bakery products, increase bakery production of existing products etc. Such a decision will help business managers to tap into growth of bakery sales by considering results of this business question. This graph also emphasizes the importance of stocking up on bakery inventory during the Christmas week. Seeing an increasing trend in the sales of bakery products, business managers might want to stock up more inventory of bakery products prior to Christmas week.

5.2.2 Report from Cubes using SSRS and SSAS: Business Question 2

Question: Are there any periods during the years 1989-1995 where the video sales are higher or lower?

Reporting tool employed: SSRS on SSAS

Screenshots of report building

Hyper Cube Creation through SSAS

1. Create Data Source

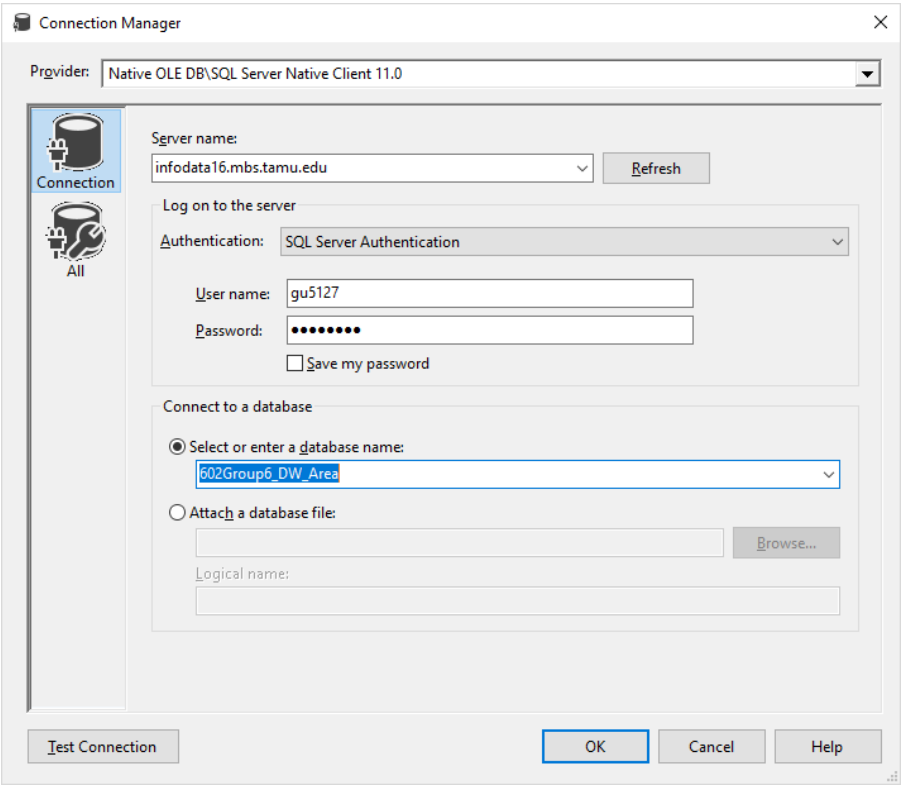

Figure 5.2-10: Create Data Source

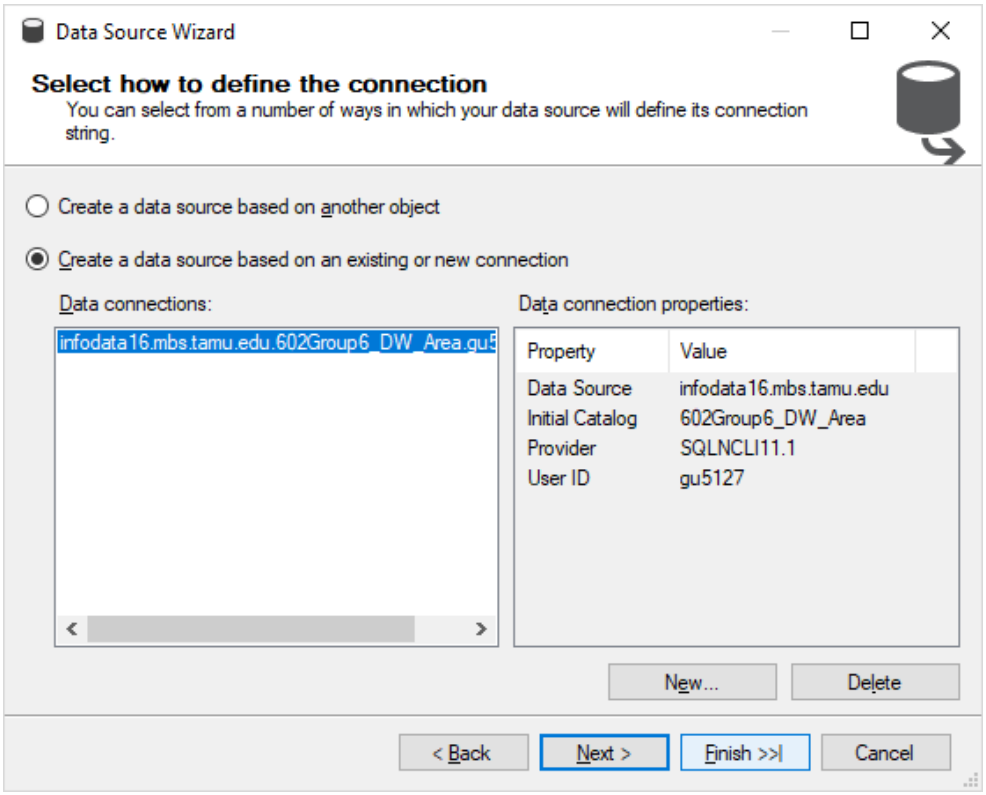

Figure 5.2-11: Data source connection

2. Select Service Account

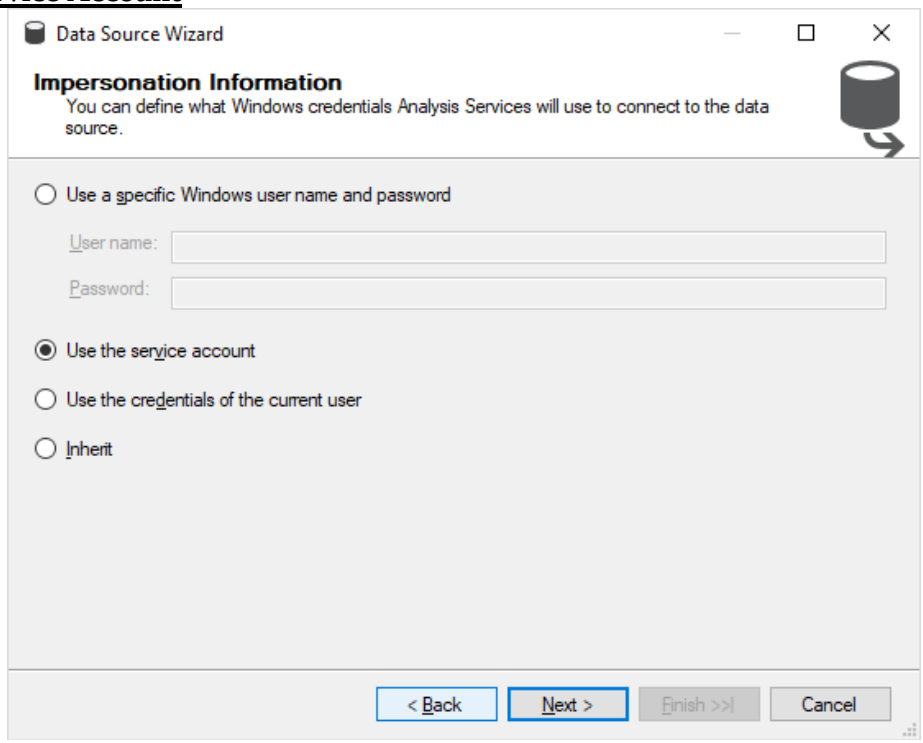

Figure 5.2-12: Select Service Account

3. Select Data Source Created in previous step to create a view

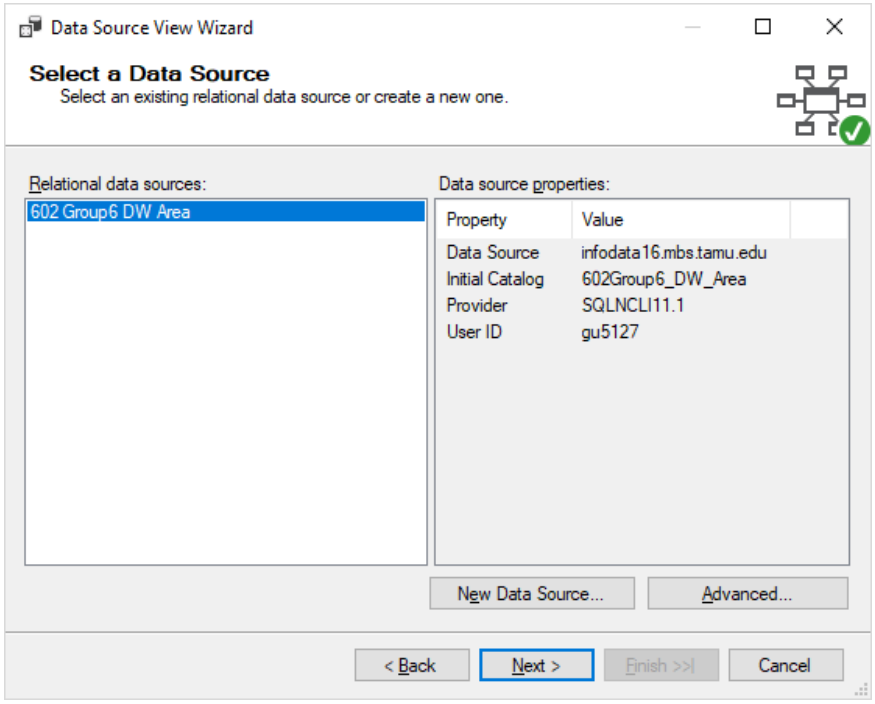

Figure 5.2-13: Select Data Source

4. Create logical relationships

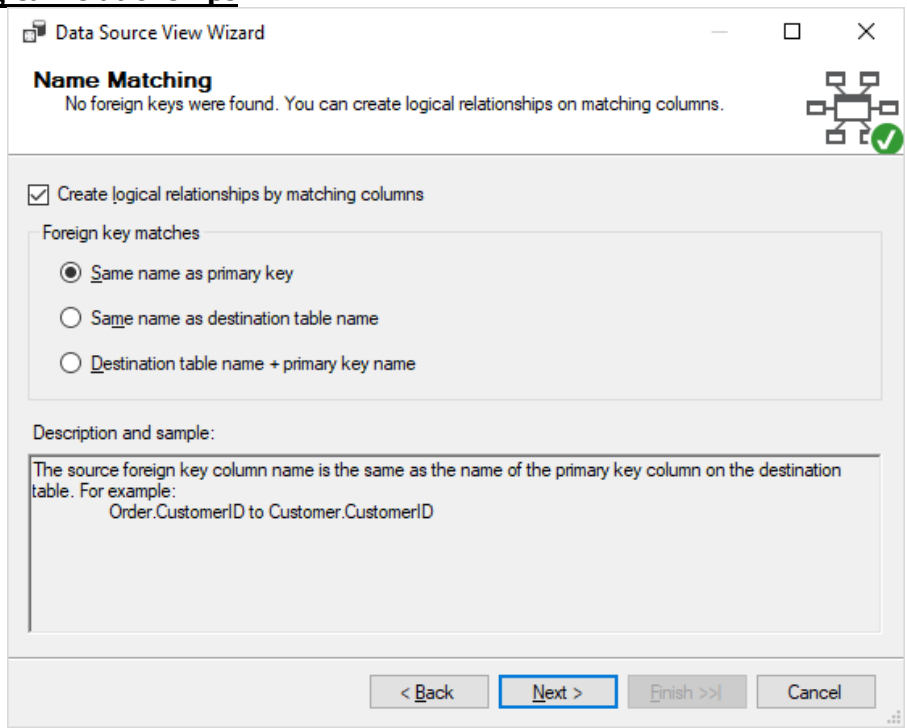

Figure 5.2-14: Create logical relationships

5. Select Dimensions and Fact tables for View

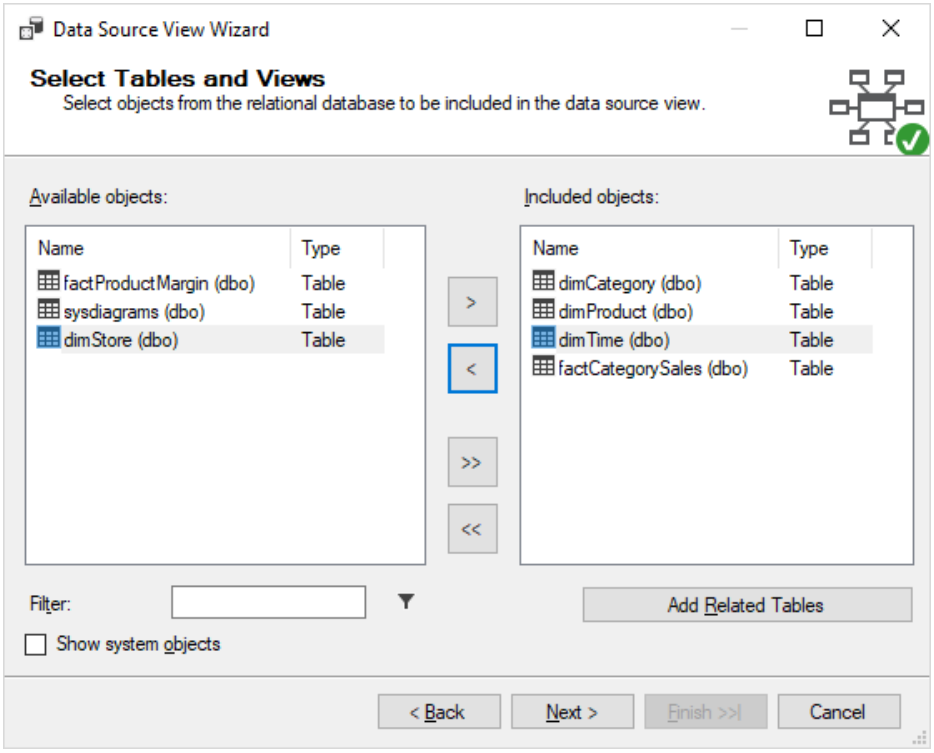

Figure 5.2-15: Select Dimensions and Fact tables for View

6. Select existing tables for creating a cube

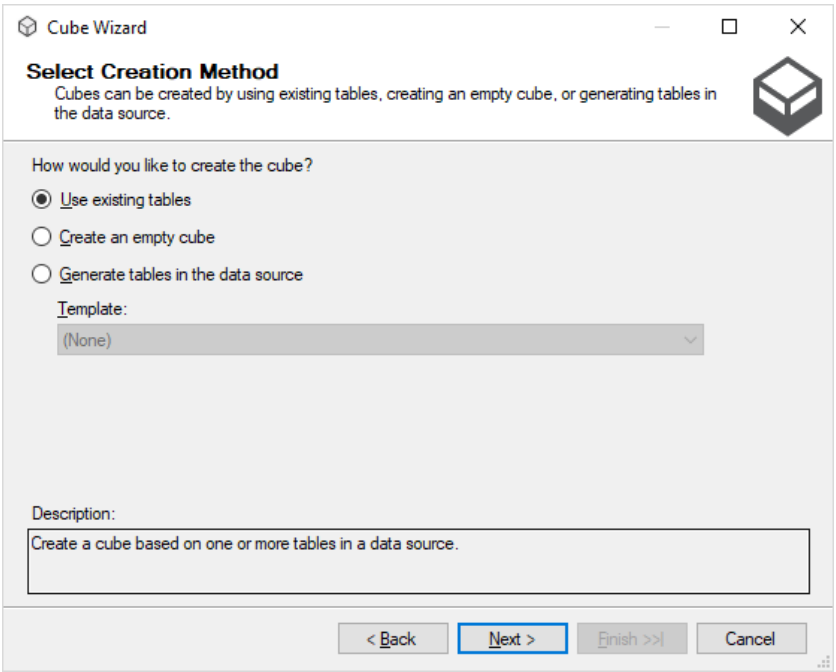

Figure 5.2-16: Selecting Cube Creation Method

7. Select Fact Tables

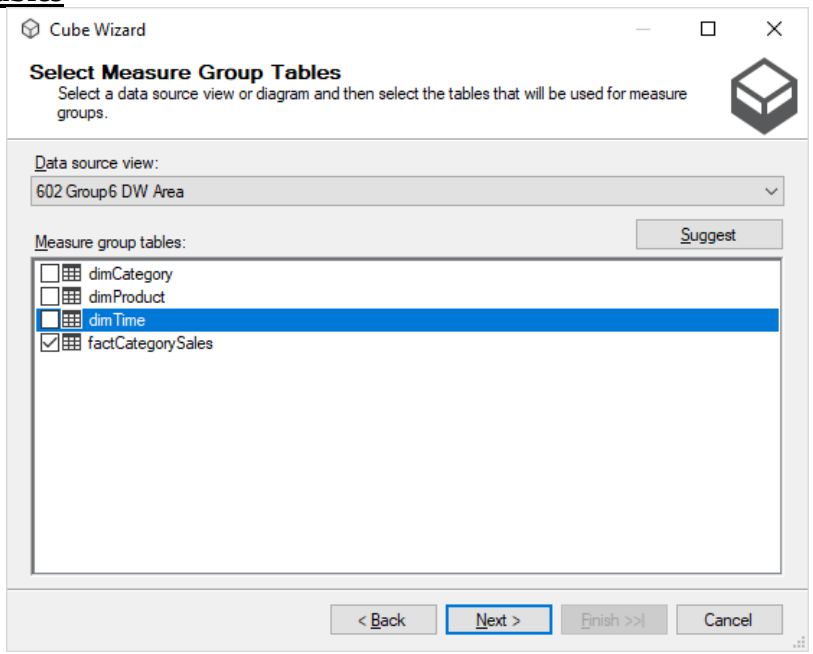

Figure 5.2-17: Selecting measure group tables

8. Select Dimension Tables

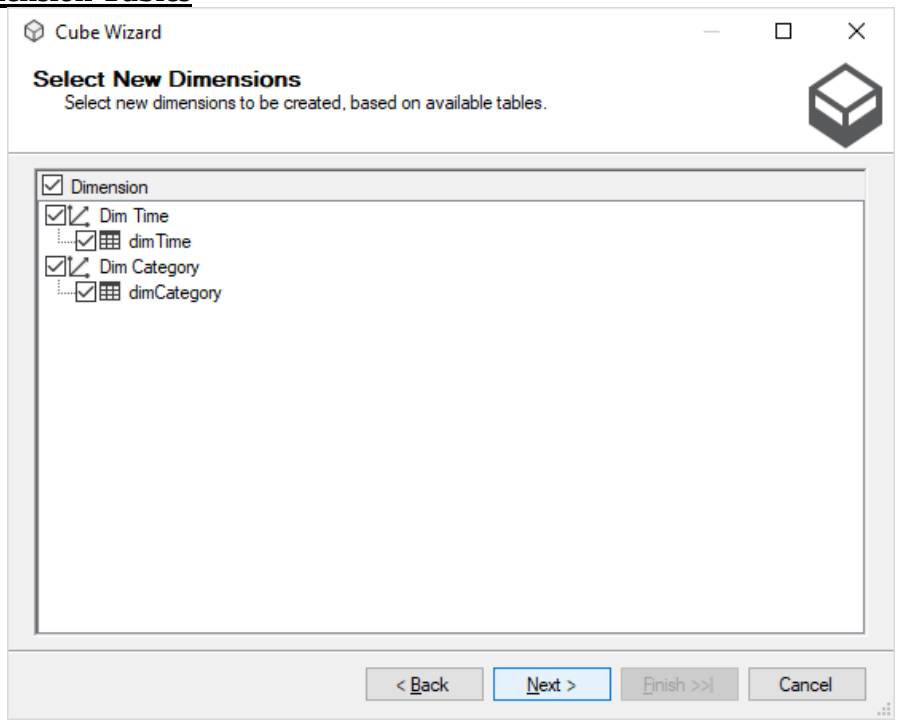

Figure 5.2-18: Selecting Dimensions

9. Cube Screenshot

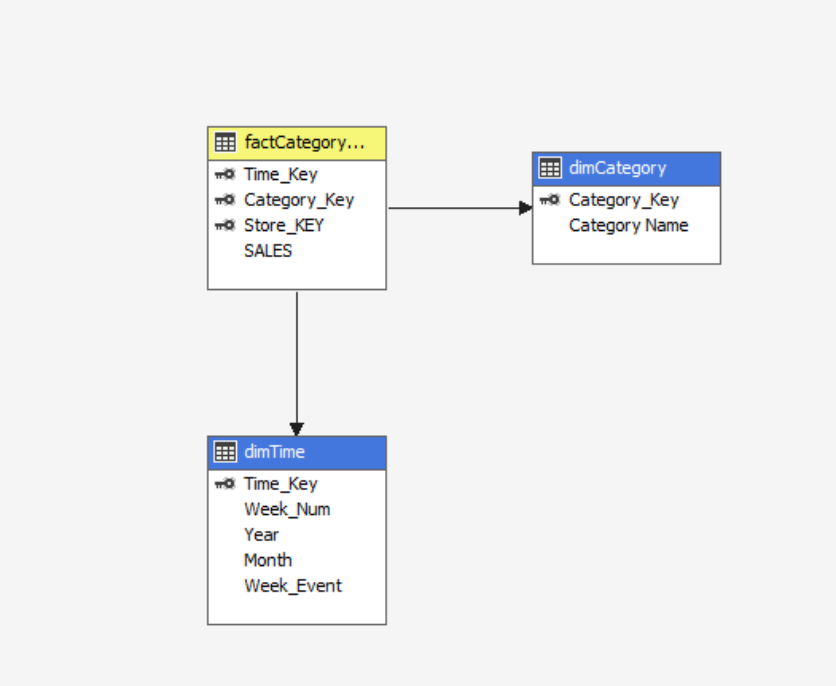

Figure 5.2-19: Cube

11. Create Time Hierarchy

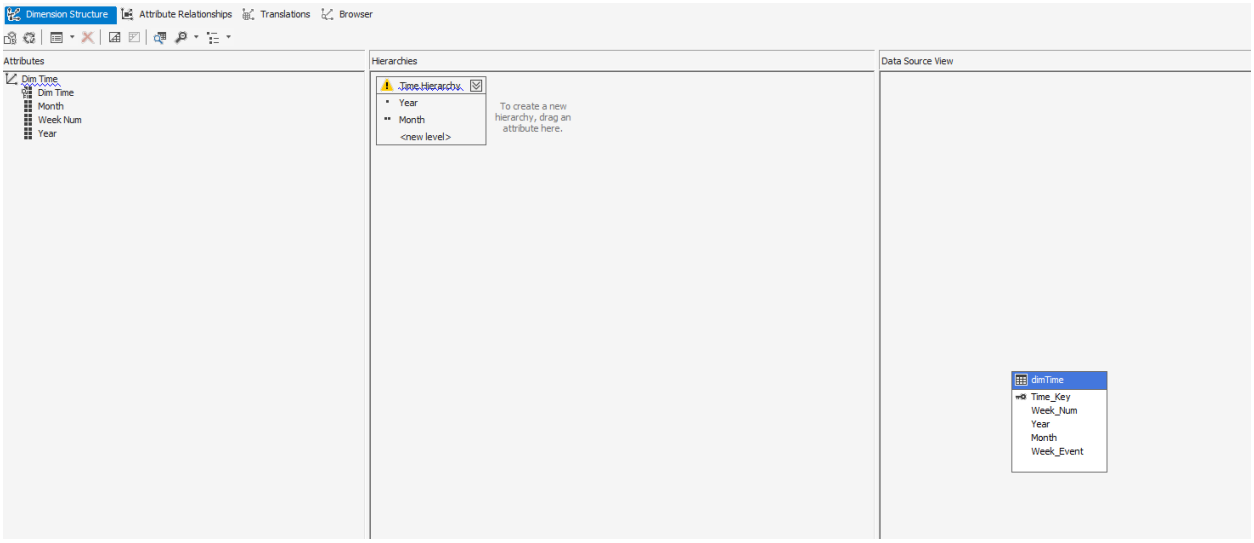

Figure 5.2-20: Time Dimension

12. Create Dim Category

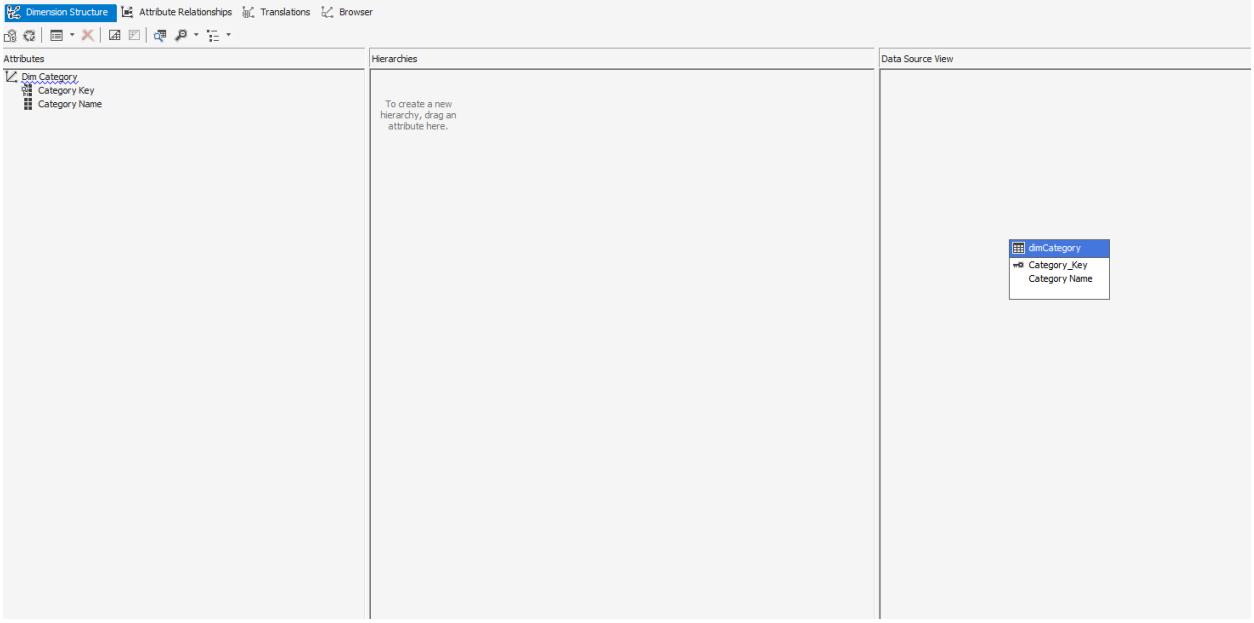

Figure 5.2-21: Category Dimension

13. Deploy Cube

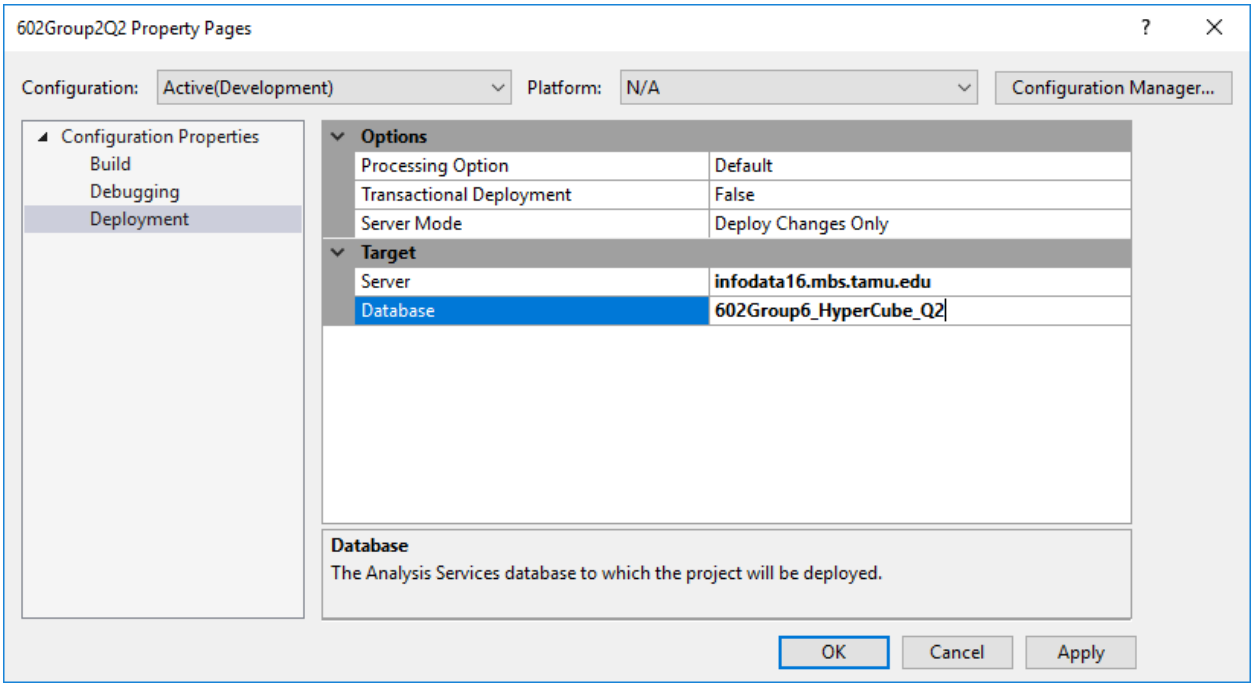

Figure 5.2-22: Setting deployment Target

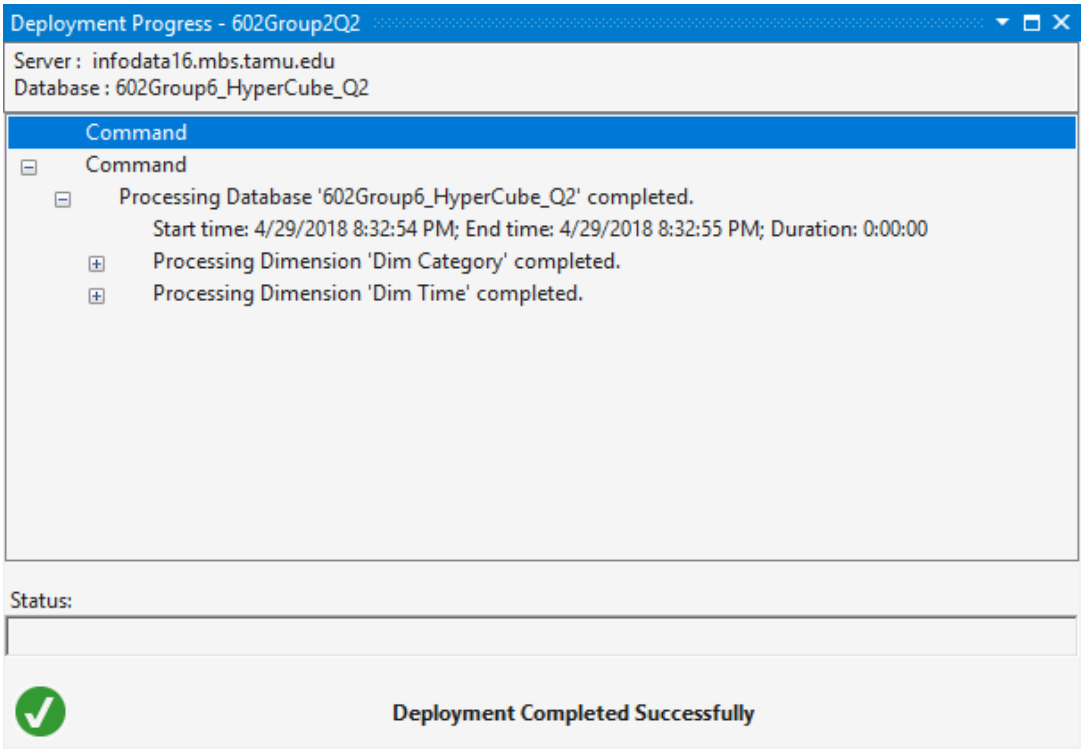

Figure 5.2-23: Successful Deployment

14. Report Preview in SSAS Browser

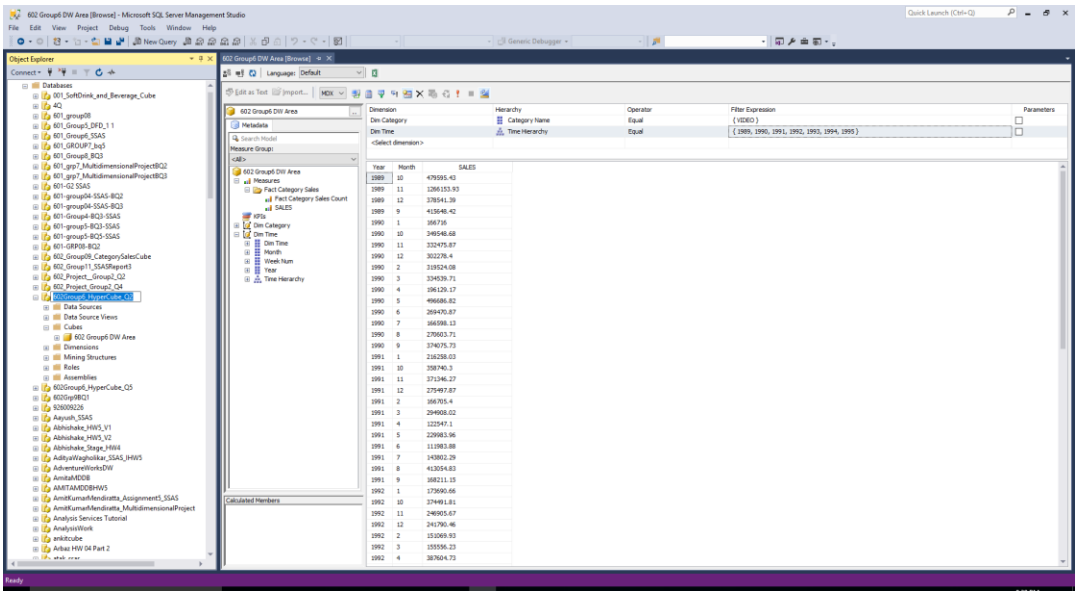

Figure 5.2-24: Data in the cube

Chart Creation through SSRS

1. Create Data Source Connection

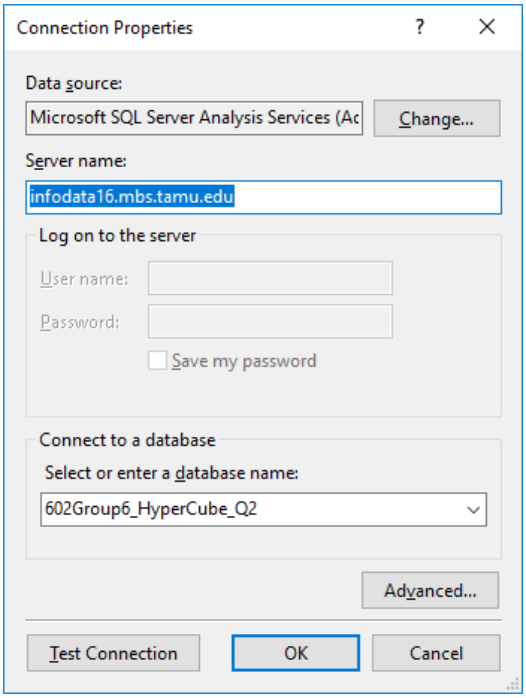

Figure 5.2-25: Connection to Cube

2. Select Data Source

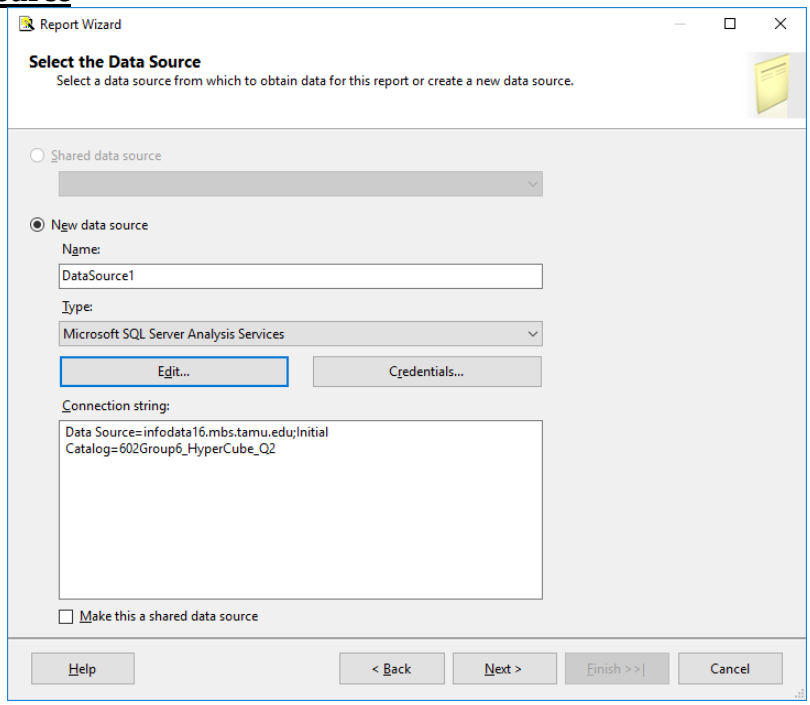

Figure 5.2-26: Selecting Data Source

3. Design Query

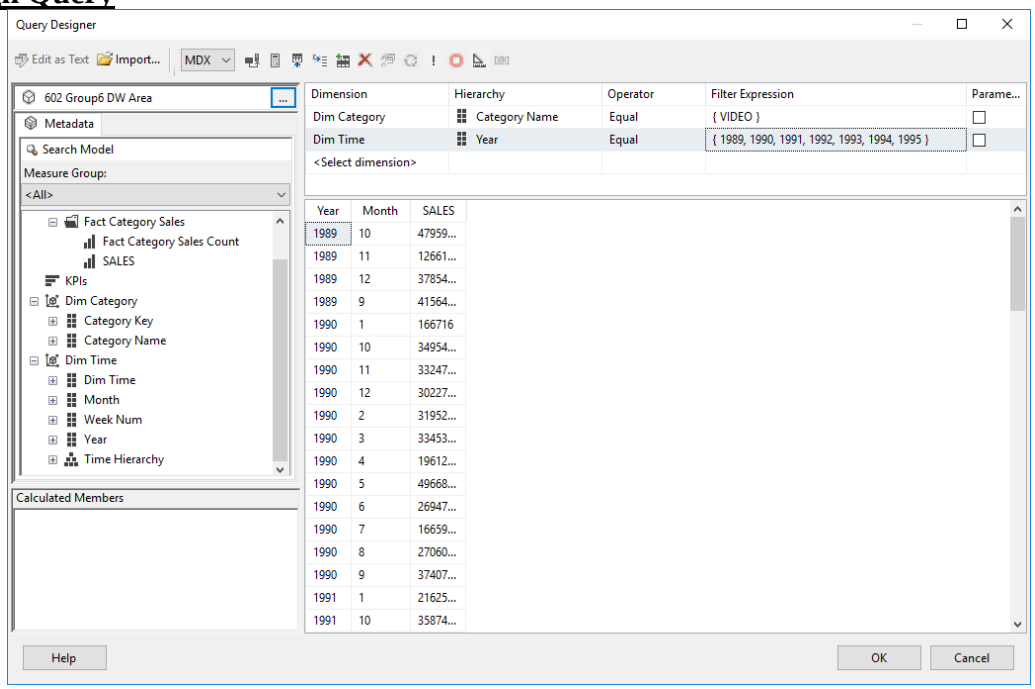

Figure 5.2-27: Designing Query

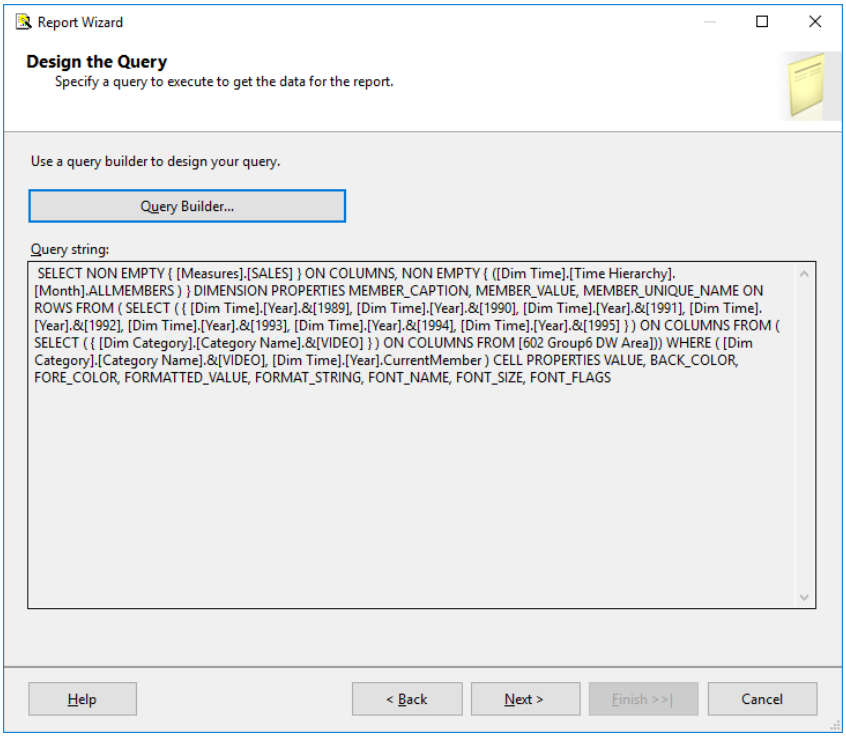

Figure 5.2-28: Query Builder

3. Arrange Report Fields

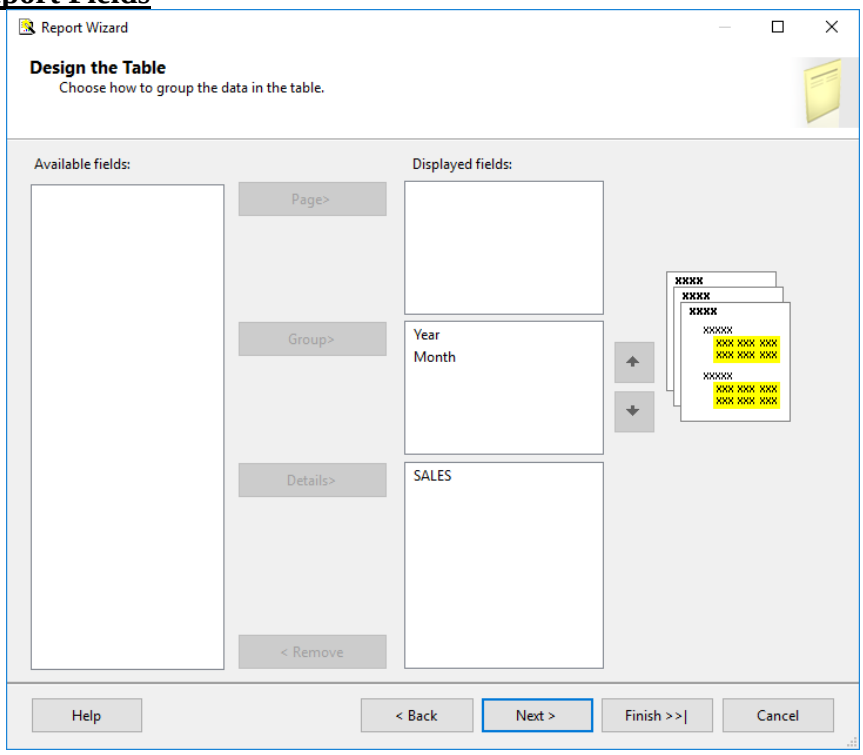

Figure 5.2-29: Designing Table

4. Enable Drilldown in Report

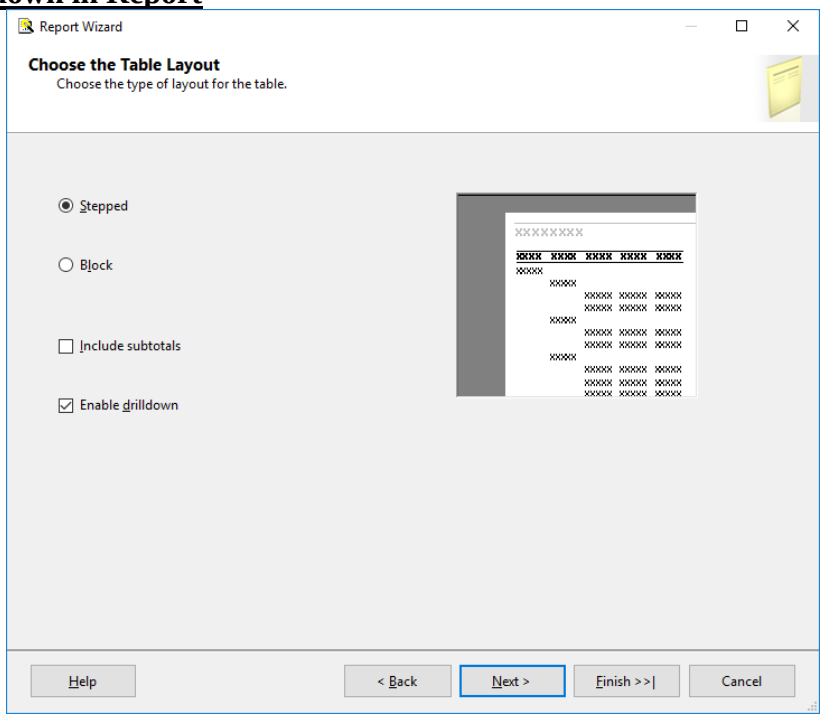

Figure 5.2-30: Table Layout

5. Edit Formatting of Chart

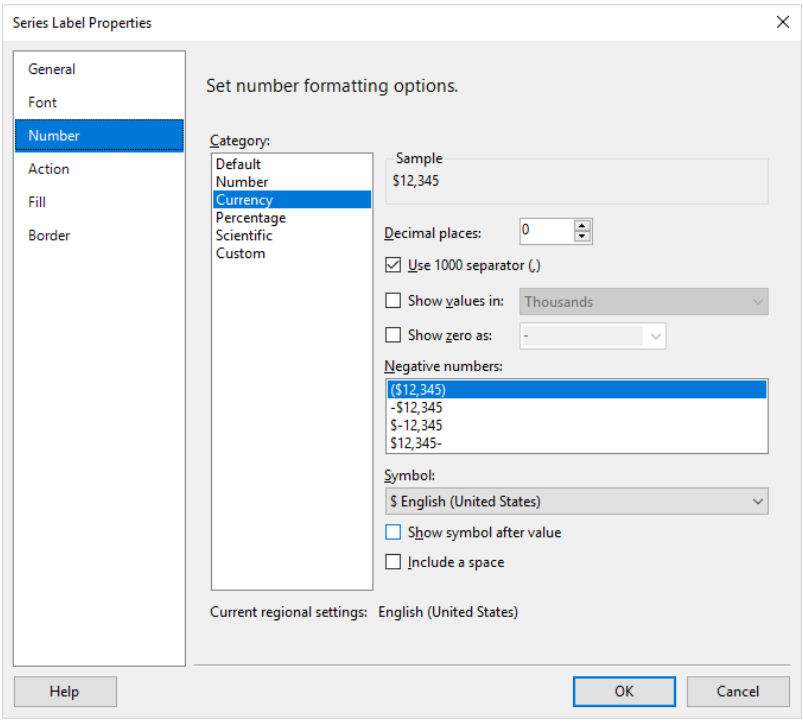

Figure 5.2-31: Edit Formatting of Chart

6. Report in Design Mode

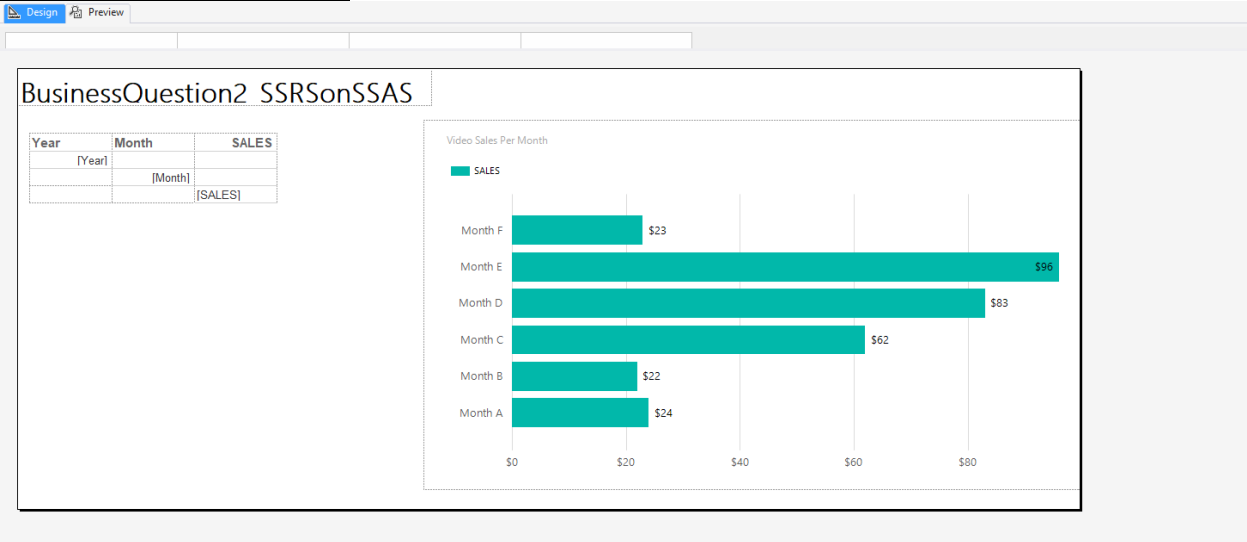

Figure 5.2-32: Report in Design Mode

7. Report in Preview Mode

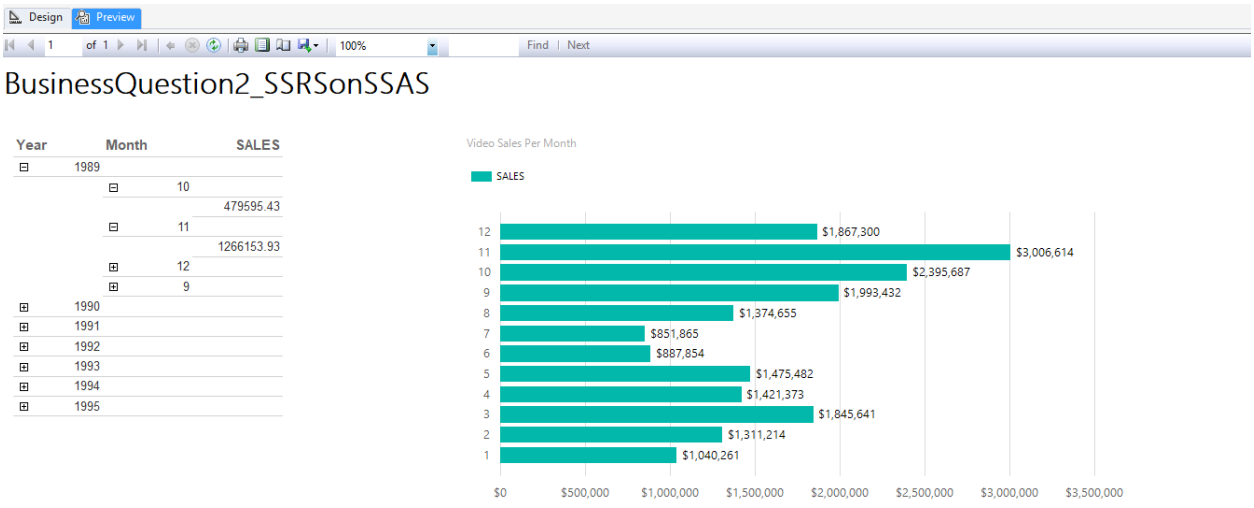

Figure 5.2-33: Report in Preview Mode

8. Deploy Report on Report Server

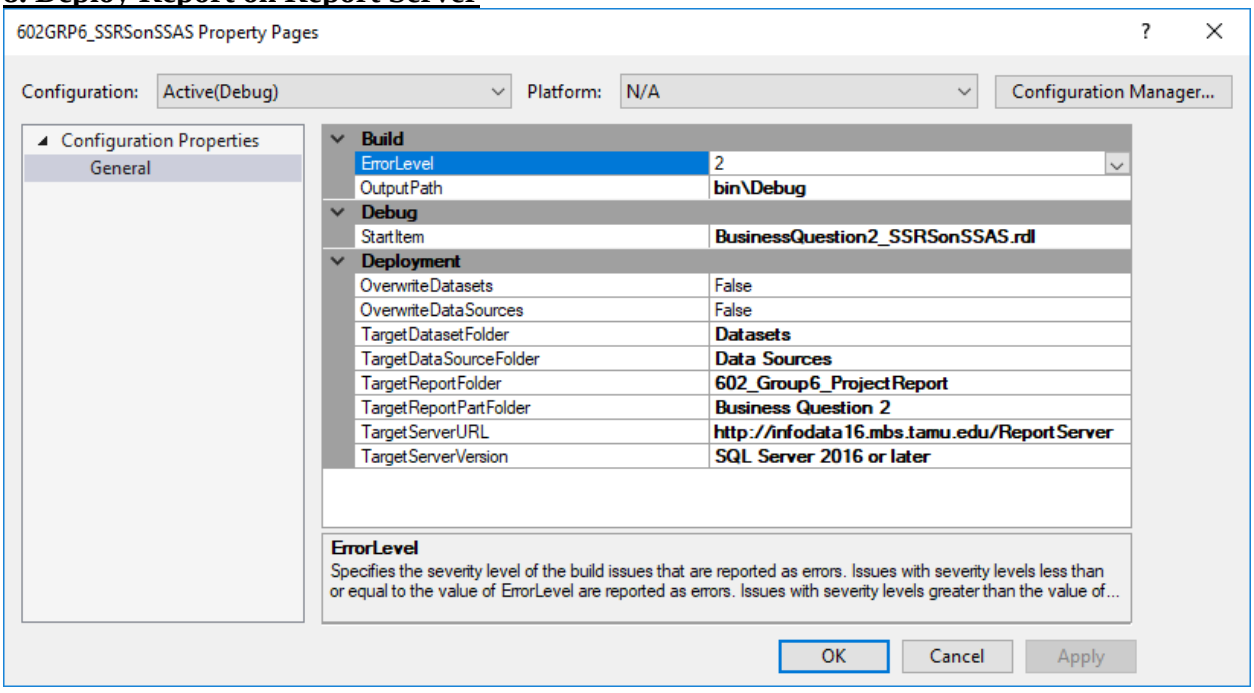

Figure 5.2-34: Deploy Report on Report Server

9. Deployed Report on Report Server

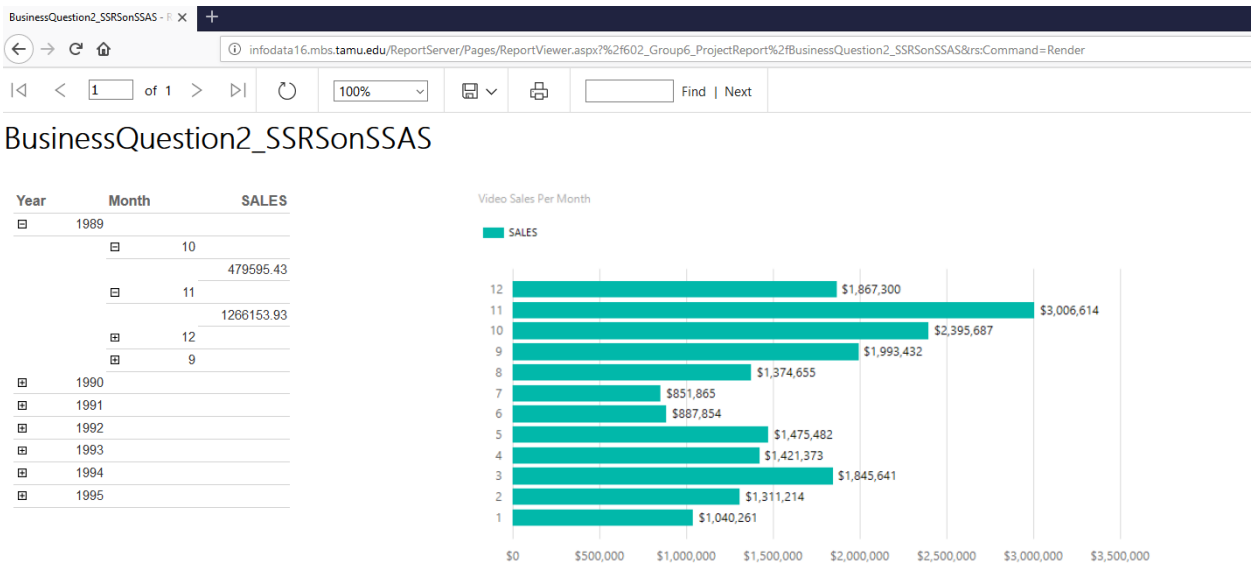

Figure 5.2-35: Deployed Report on Report Server

Conclusion

We used SSRS on top of SSAS to understand the trend for video sales. We constructed a hyper cube and then derived insights on monthly sales of videos. This analysis will help business manager understand what the periods are where customers purchase videos. Videos were a novel technology during the period 1989 -1995. The graph for video sales per month helps us understand that videos were being purchased a lot during the months of November, October and March where their sale was lowest during the months of June and July. Business Managers can look at this data to make a decision whether to stock up on videos or use the shelf space for other products. This analysis points out that videos must be stocked up during the months of November, October and March while during June and July shelf space allotted to them must be reduced.

5.2.3 Report from Independent Data Marts using SSRS: Business Question 3

What is the profit trend of cheese during 1991 to 1995?

Reporting tool employed: SSRS

Screenshots of report building

1. Create Data Source

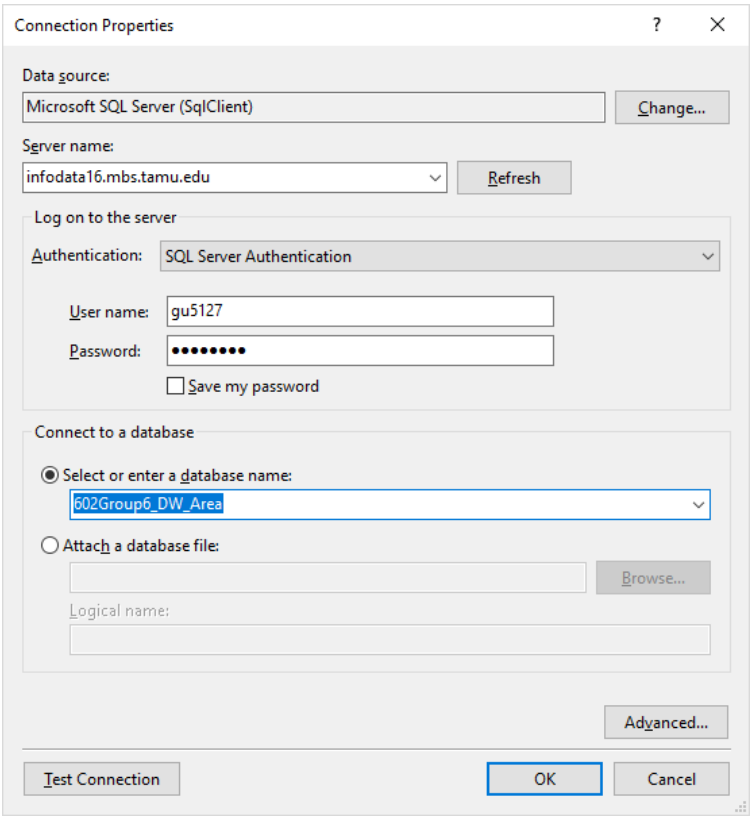

Figure 5.2-36: Data base connection

3. Select Dimension and Fact Tables to be Used

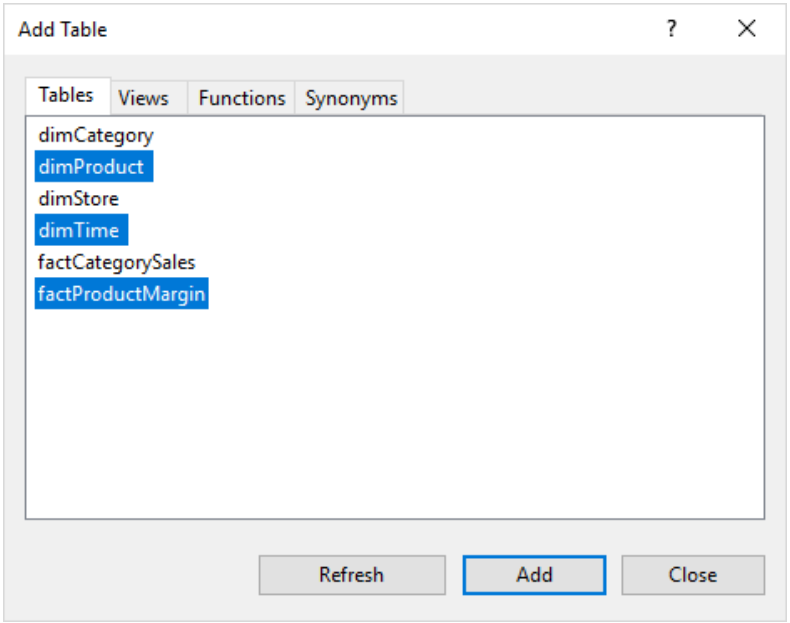

Figure 5.2-37: Adding Tables

3. Design Query

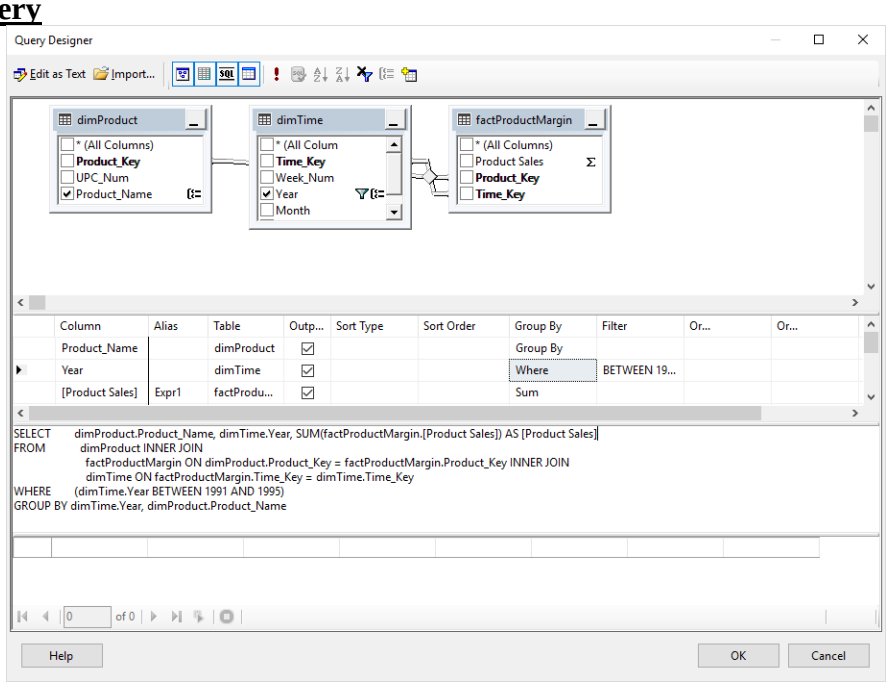

Figure 5.2-38: Query Designer

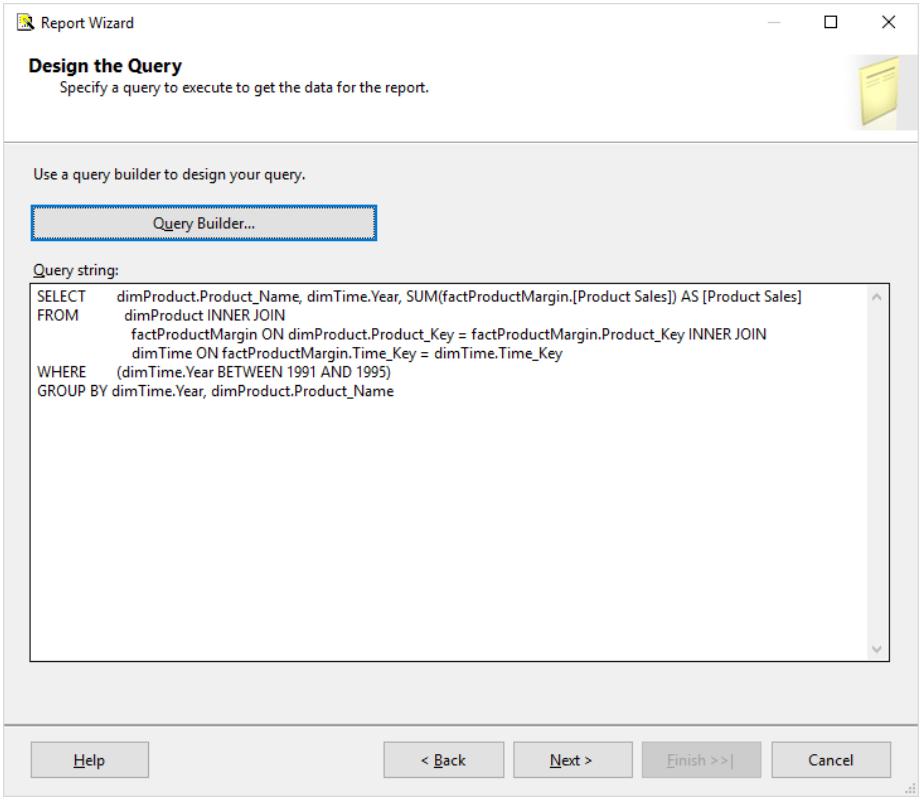

Figure 5.2-39: Query Builder

4. Arrange Report Fields

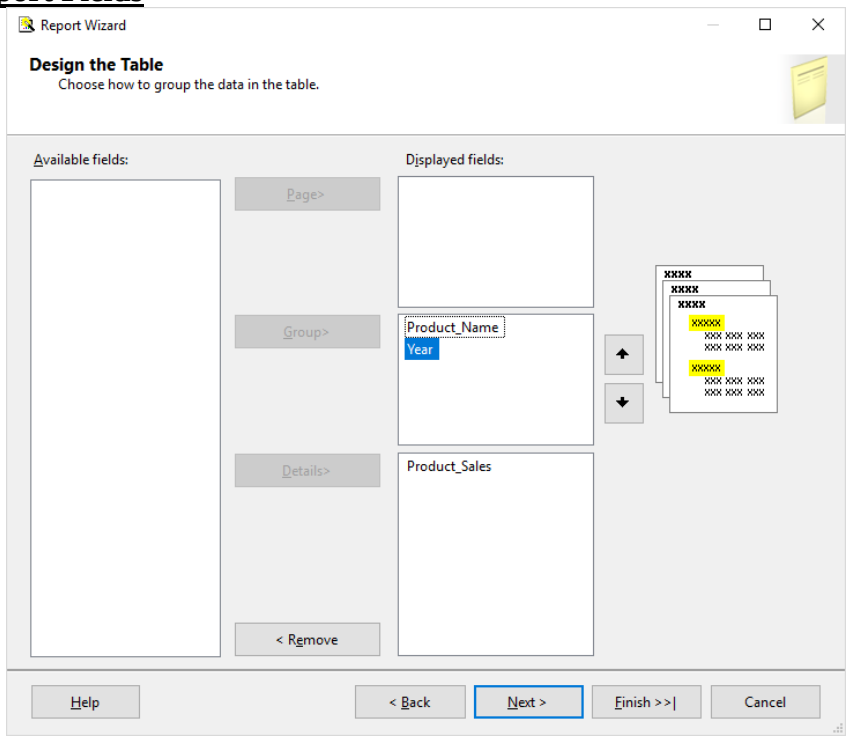

Figure 5.2-40: Designing Table

5. Enable Drilldown

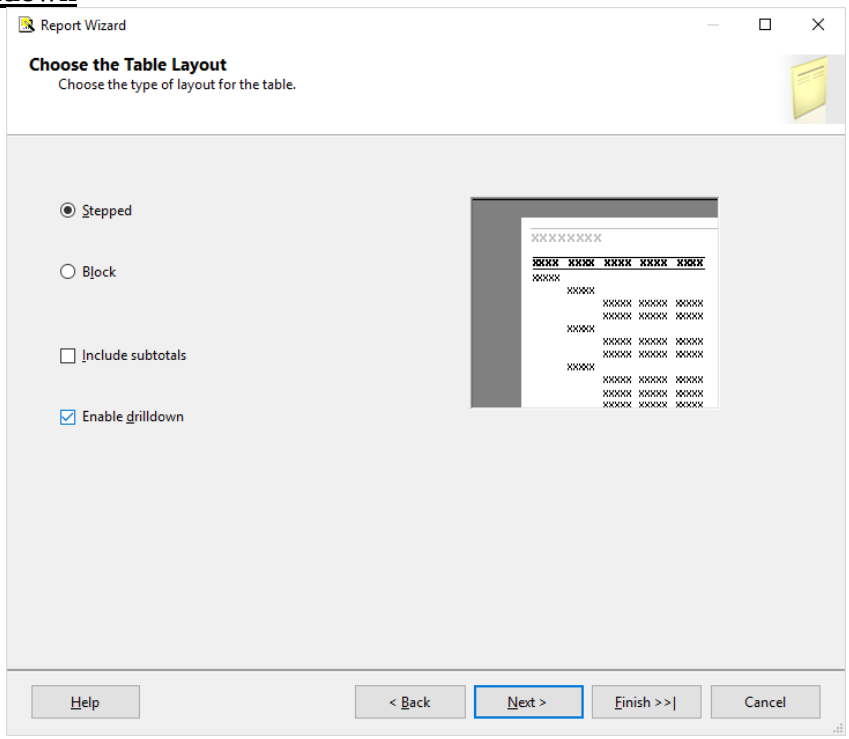

Figure 5.2-41: Table layout

6. Report in Design Mode

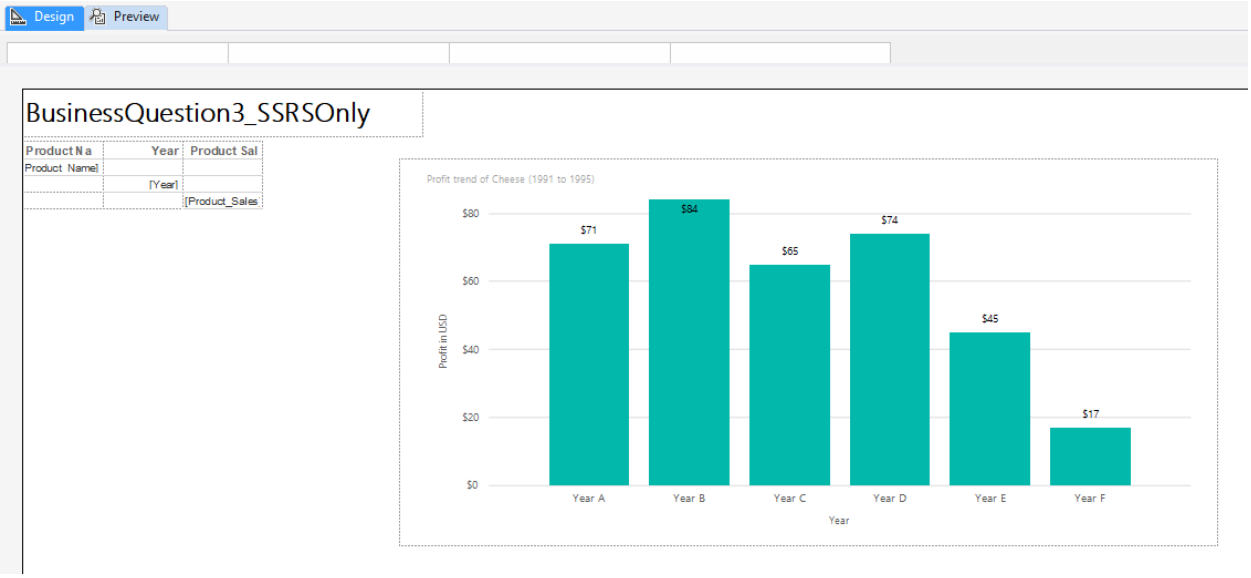

Figure 5.2-42: Report in Design Mode

6. Report in Preview Mode

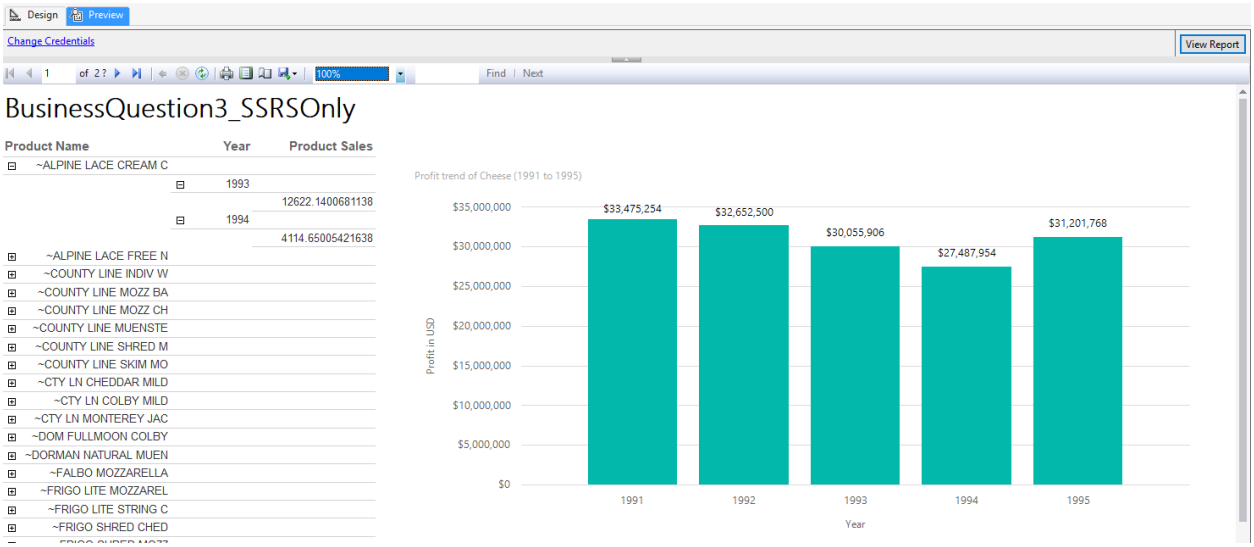

Figure 5.2-43: Report in Preview Mode

7. Deploy Report on Report Server

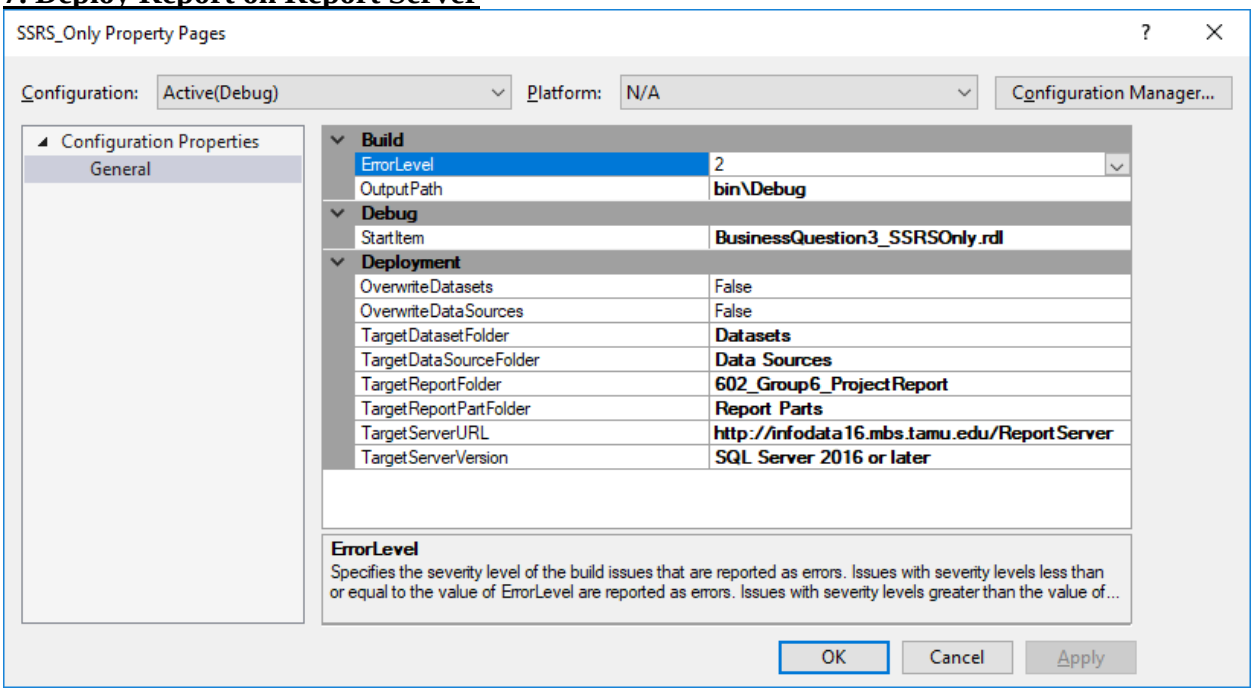

Figure 5.2-44: Deploy Report on Report Server

6. Report Deployed on Report Server

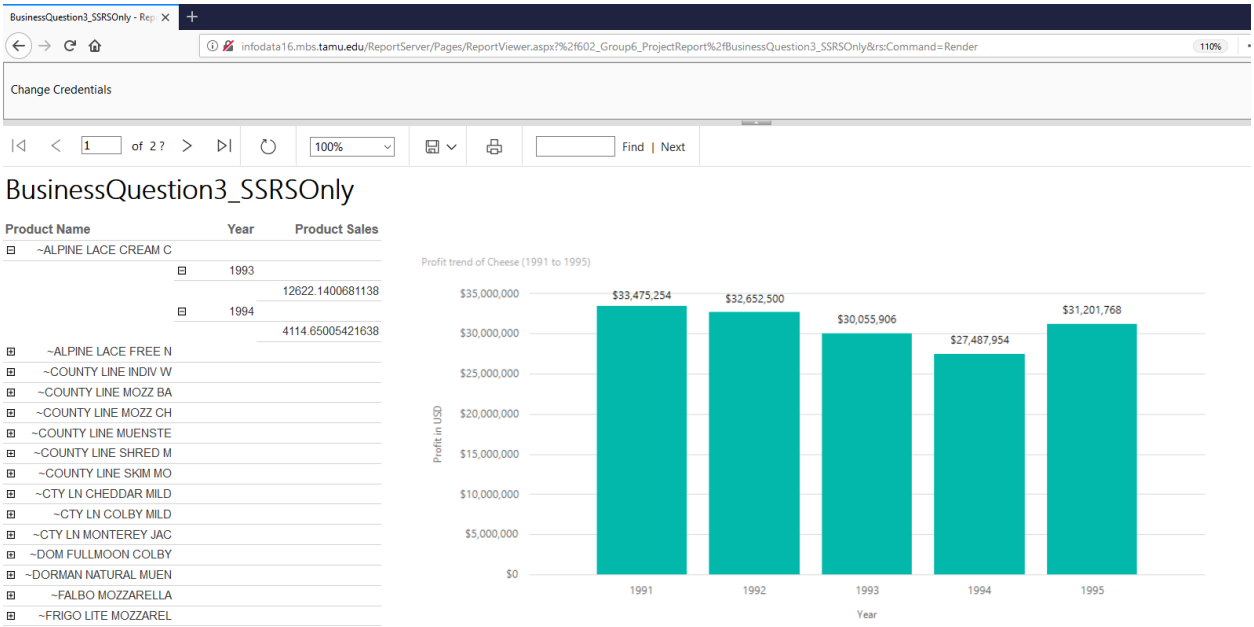

Figure 5.2-45: Report Deployed on Report Server

Conclusion

We used SSRS only to understand profit trend of cheese products from the period 1991 to 1995. We created a report with product names and their yearly profit and augmented it visually through a horizontal bar chart. From the report and corresponding chart Business Managers can understand that cheese sales have been dropping during the period 1991 to 1994. However, there is an increase in the profit margins during 1995. A business manager can utilize the report we have created to understand which cheese products profitable and selling and which ones are not. The ones which are generated high profits can be stocked more while the ones that are not profitable should be removed from the retail stores shelf.

5.2.4 Report using Report Builder: Business Question 4

How the trend of grocery sales varies across different stores in Chicago?

Reporting tool employed: Report Builder 3.0

Screenshots of report building

1. Create Data Source

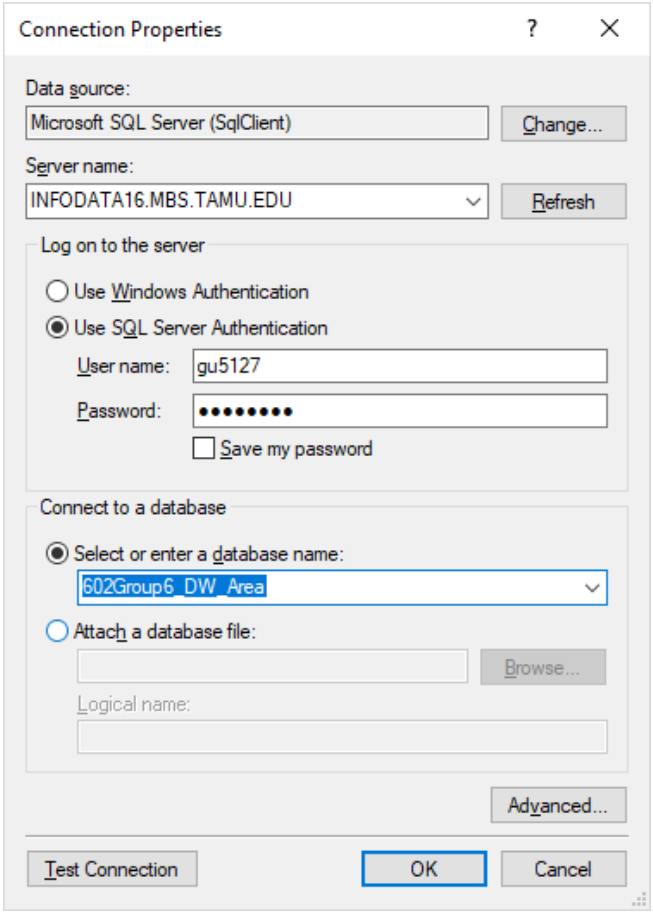

Figure 5.2-46: Connecting to database

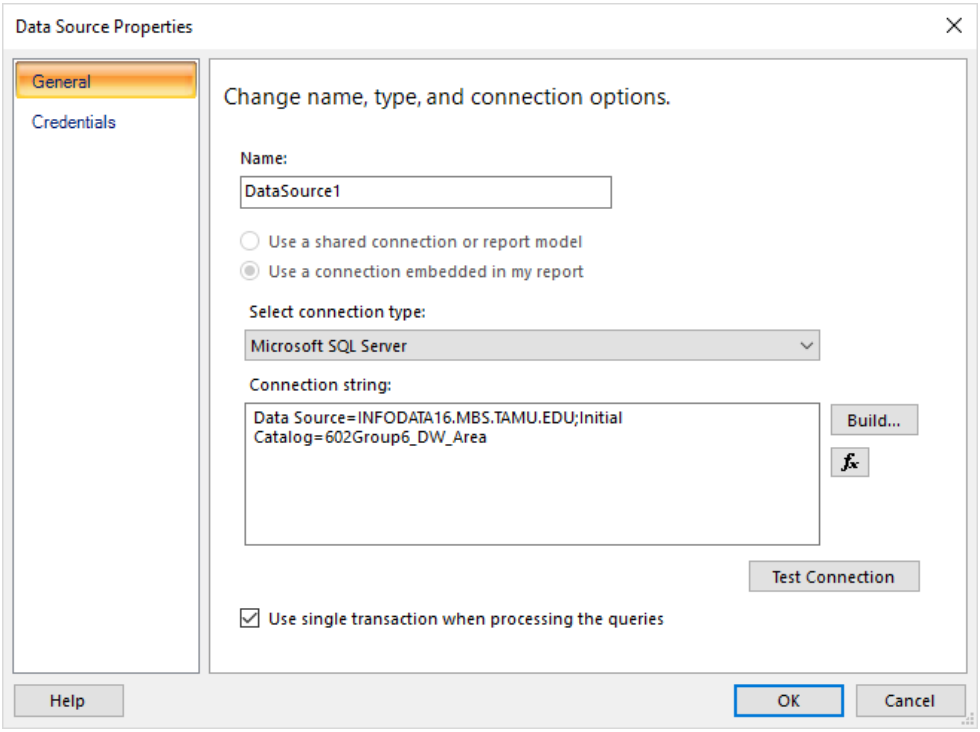

Figure 5.2-47: Data source properties

2. Design Query

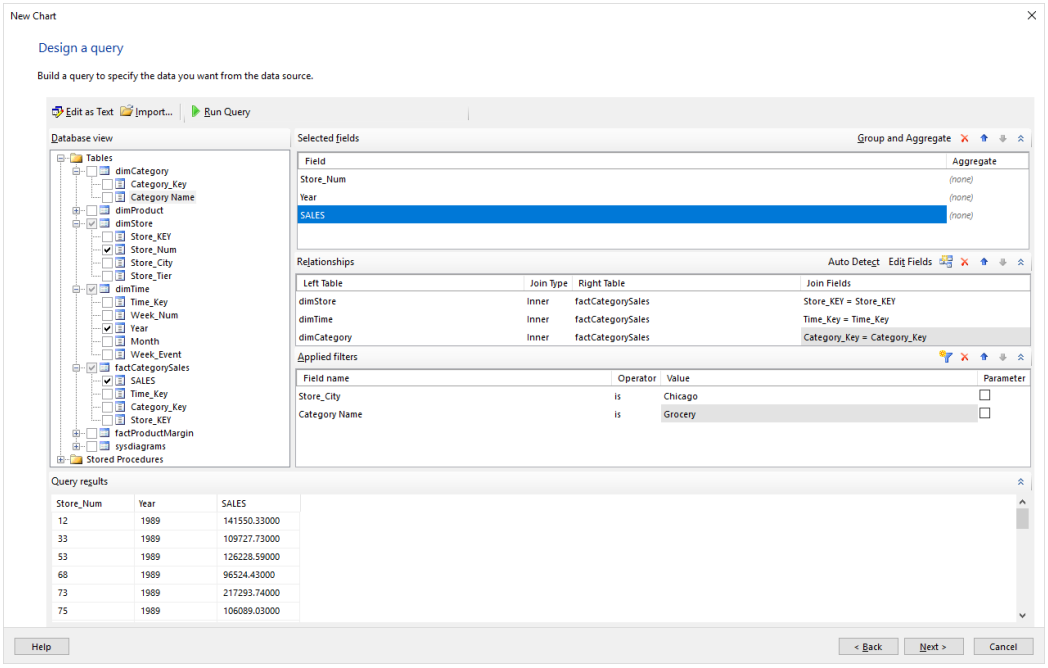

Figure 5.2-48: Designing Query

3. Arrange Report Fields

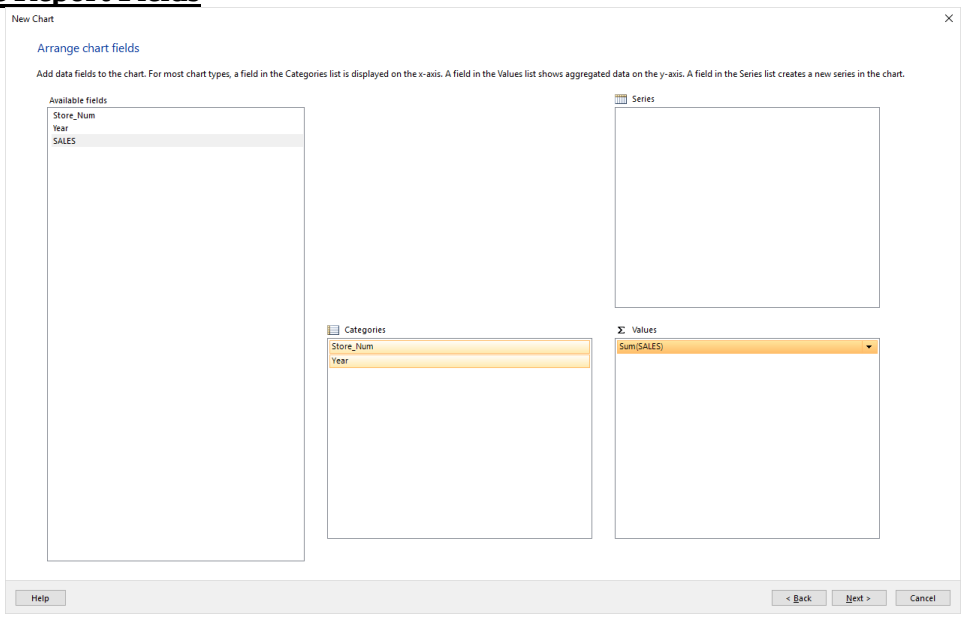

Figure 5.2-49: Arranging Chart Fields

4. Edit Formatting of Chart

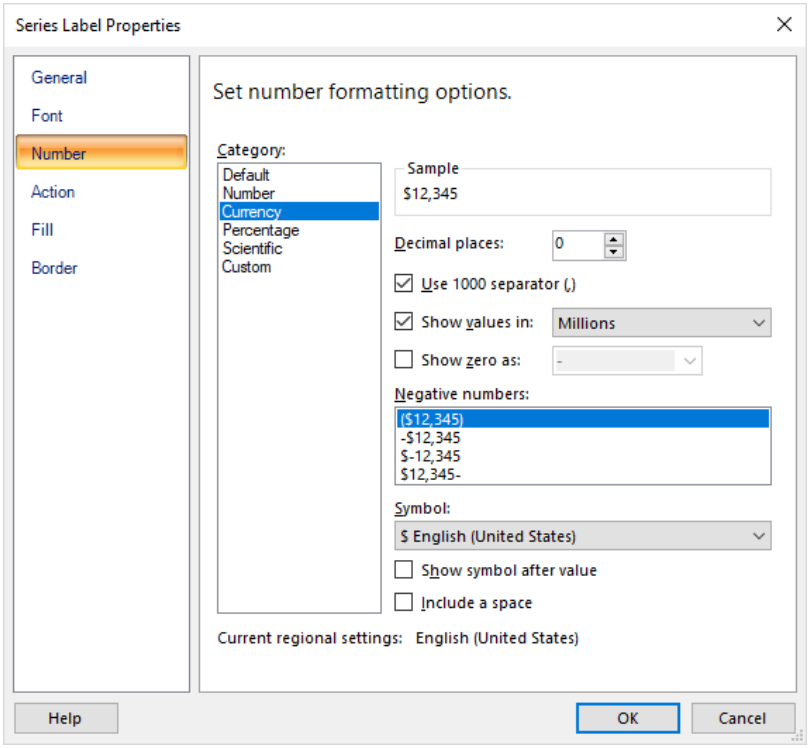

Figure 5.2-50: Edit Formatting of Chart

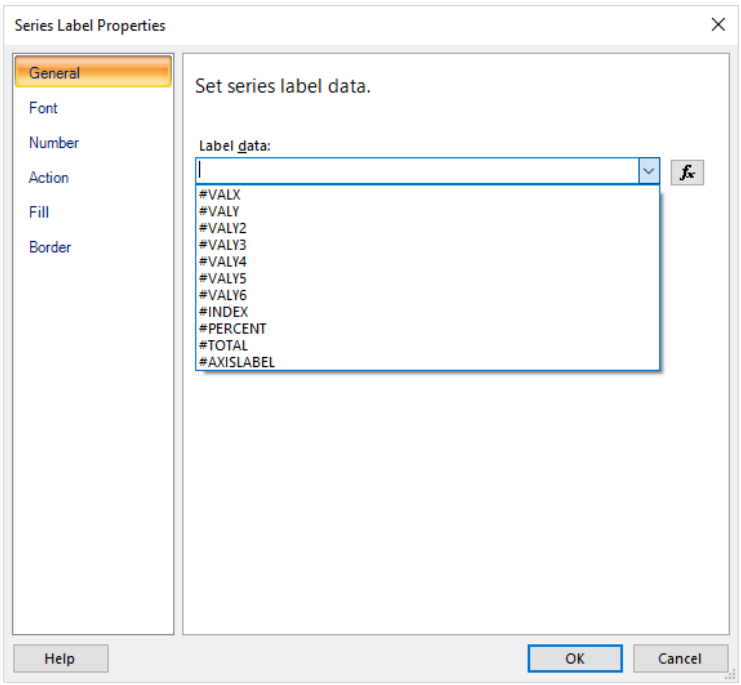

Figure 5.2-51: Series label properties

5. Report Preview in Report Builder

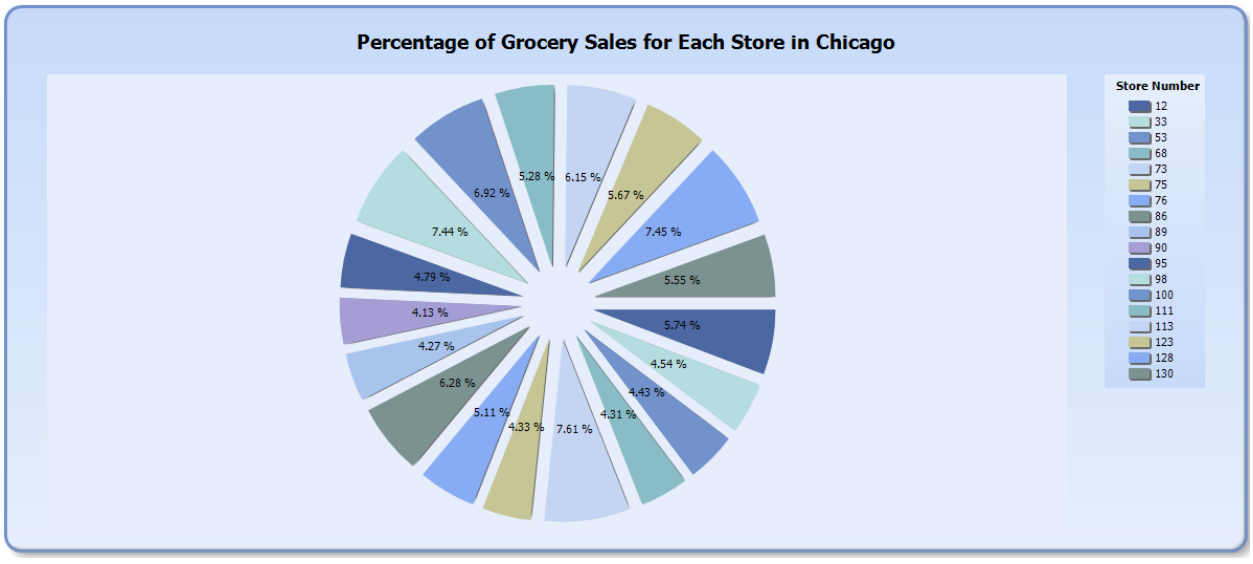

Figure 5.2-52: Percentage of grocery sales for each store in Chicago

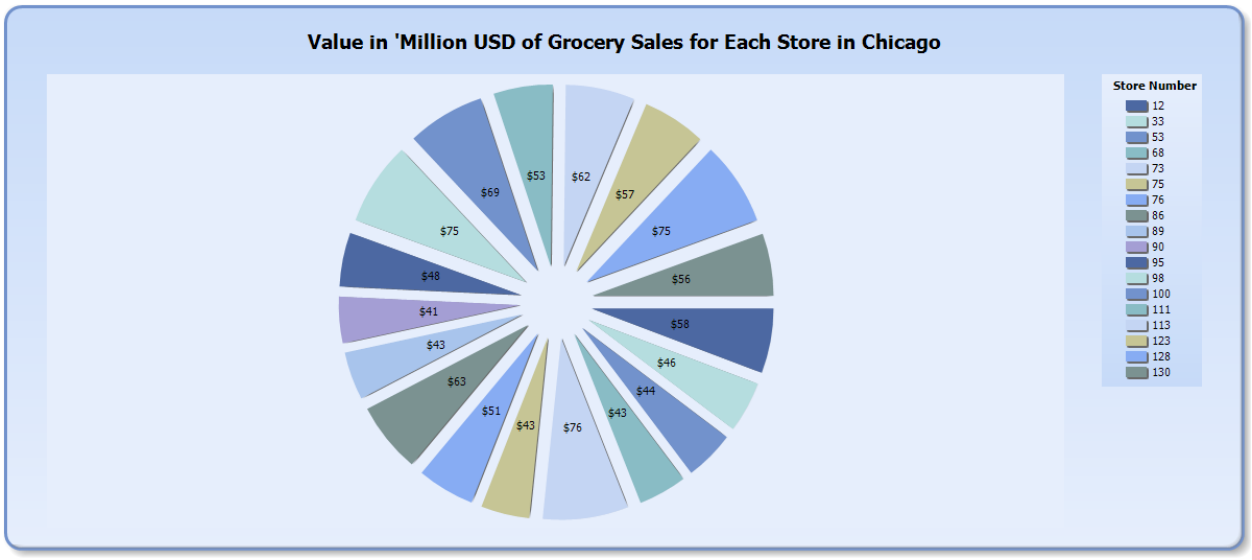

Figure 5.2-53: Dollar value of grocery sales for each store

6. Report Deployed on Report Server

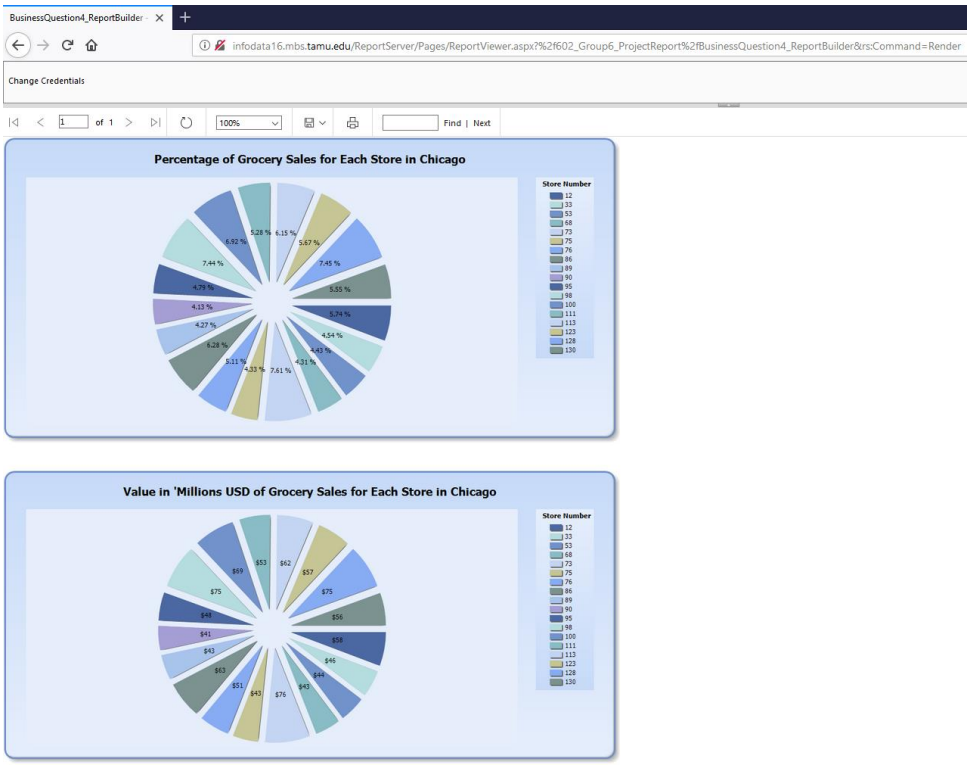

Figure 5.2-54: Report Deployed on Report Server

Conclusion

We used Report Builder to analyze the sales of grocery products across stores. We created two pie charts – one that shows percentage of total grocery sales of a store and another which depicts the exact amount of grocery sales per store. These two pie charts provide important insights to

Business Managers on how much is each store contributing to overall grocery sales and how much is the total amount of sales. These two measures can be used by business managers to stock up or remove stock from retail stores. For example, if a store contributes to less than 4.5% of overall sales and its total contribution is less than $42mn a manager can decide to stock less grocery inventory in that store. But if the percentage grocery sales is less than 4.5% but sales are higher than $42mn it would still be an important store for grocery inventory stocking. This analysis will assist a business manager in such kind of decisions.

5.2.5 Cube from SSAS: Business Question 5

Detect trends in Beer Sales encompassing the entire duration that look for high period of sales. Detect peaks not only during designated holidays but also on events that are not in the Week_Decode table.

Reporting tool employed: SSAS
Screenshots of report building

1. Create Data Source

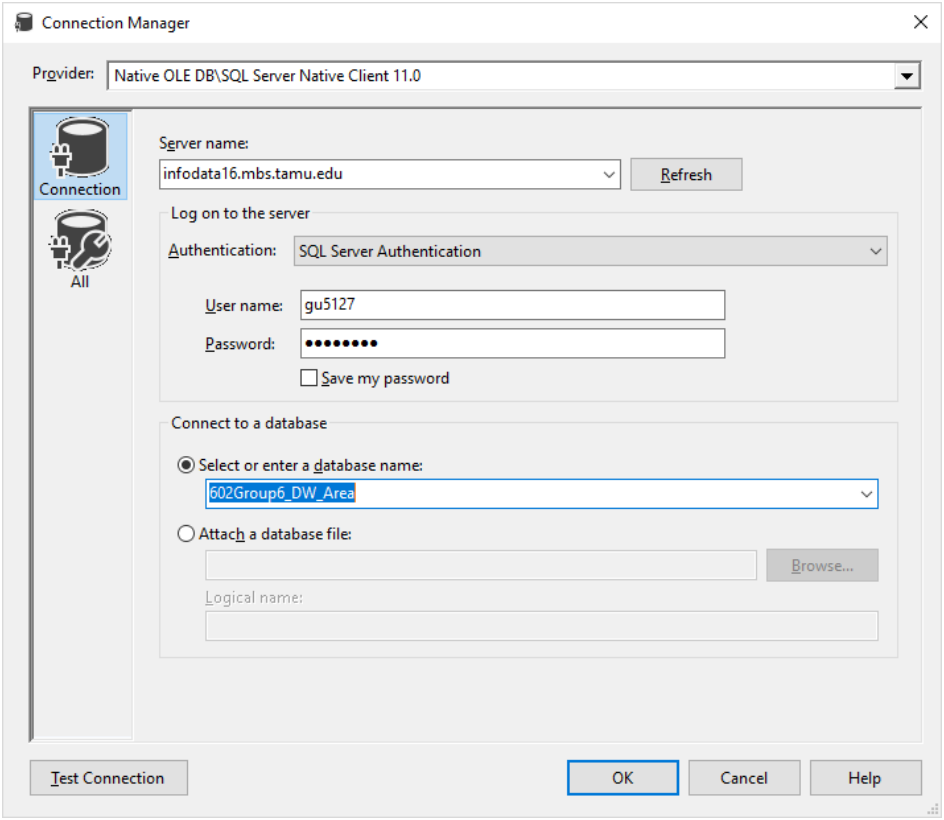

Figure 5.2-55: Data source connection

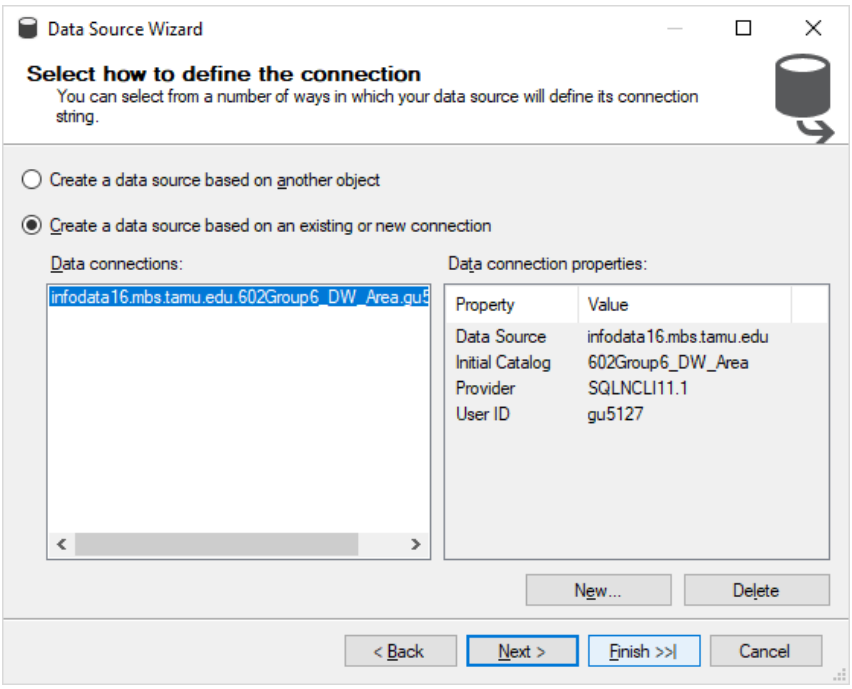

Figure 5.2-56: Data source connection defined

2. Select Service Account

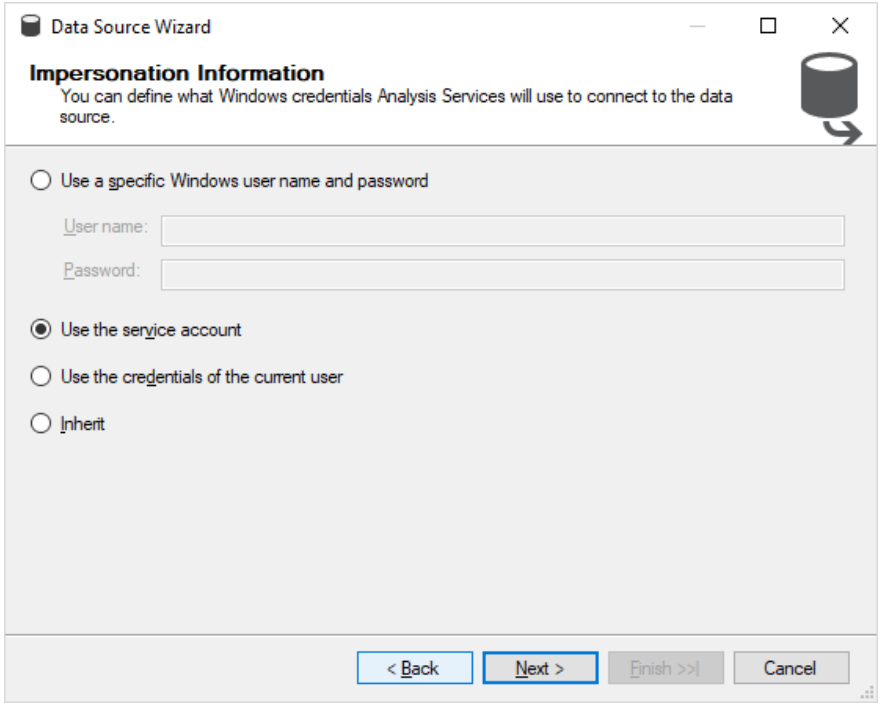

Figure 5.2-57: Impersonation information

3. Select Data Source Created in previous step to create a view

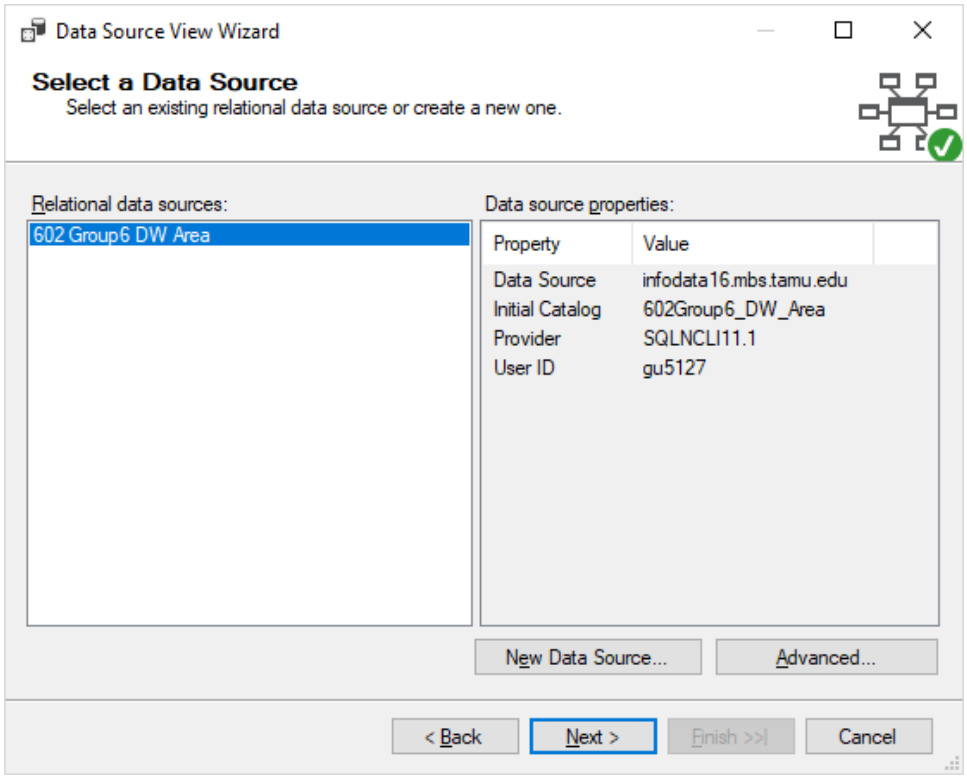

Figure 5.2-58: Selecting data source

4. Create logical relationships

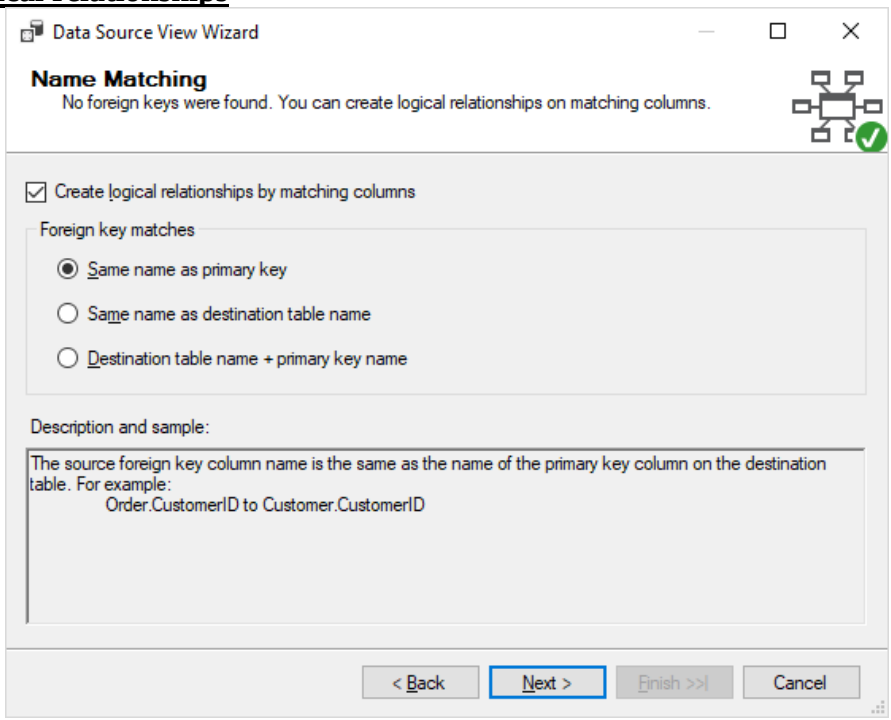

Figure 5.2-59: Name Matching

5. Select Dimensions and Fact tables for View

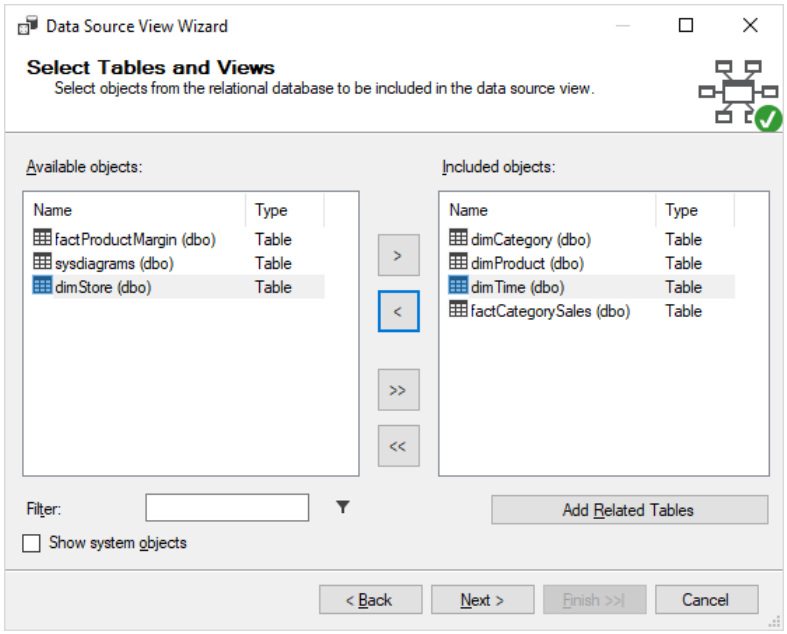

Figure 5.2-60: Selecting Tables and Views

6. Select existing tables for creating a cube

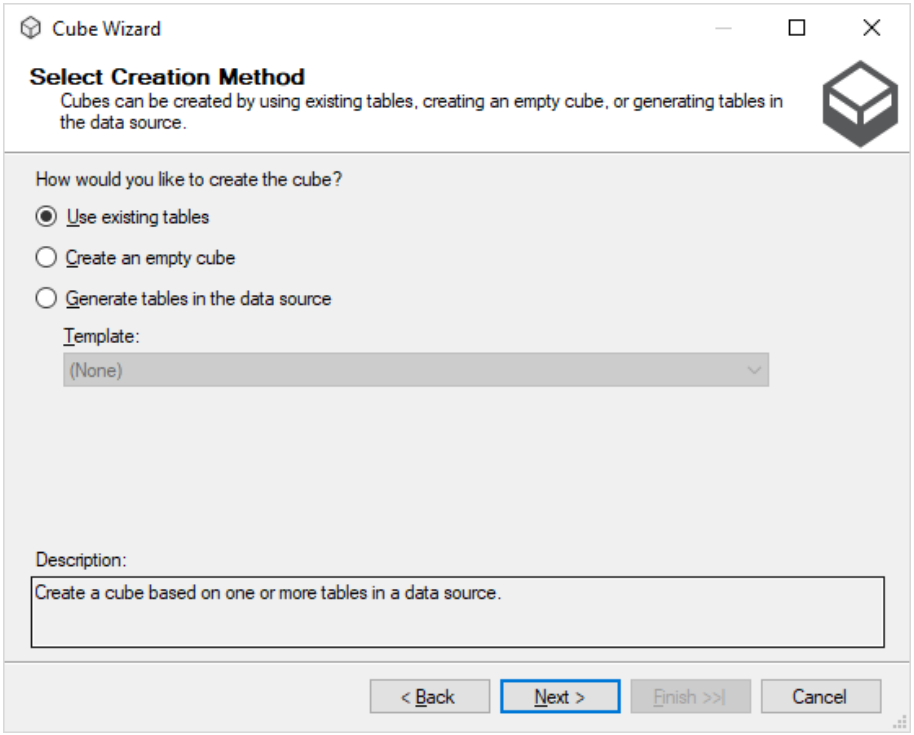

Figure 5.2-61: Selecting Cube Selection Method

7. Select Fact Tables

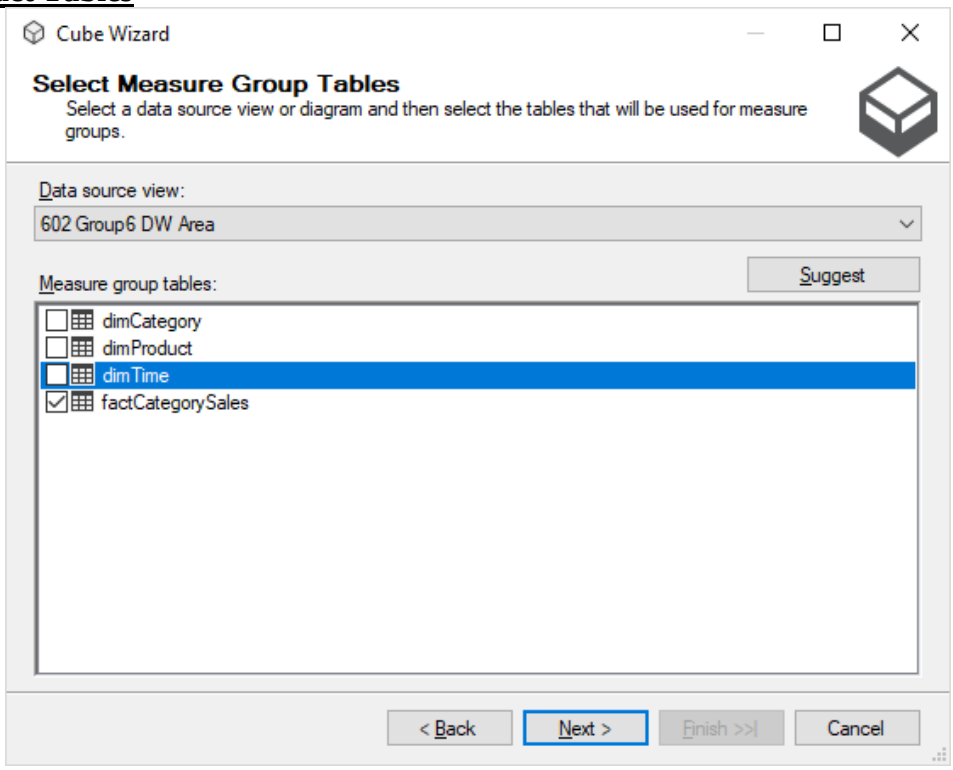

Figure 5.2-62: Selecting measure group tables

8. Select Dimension Tables

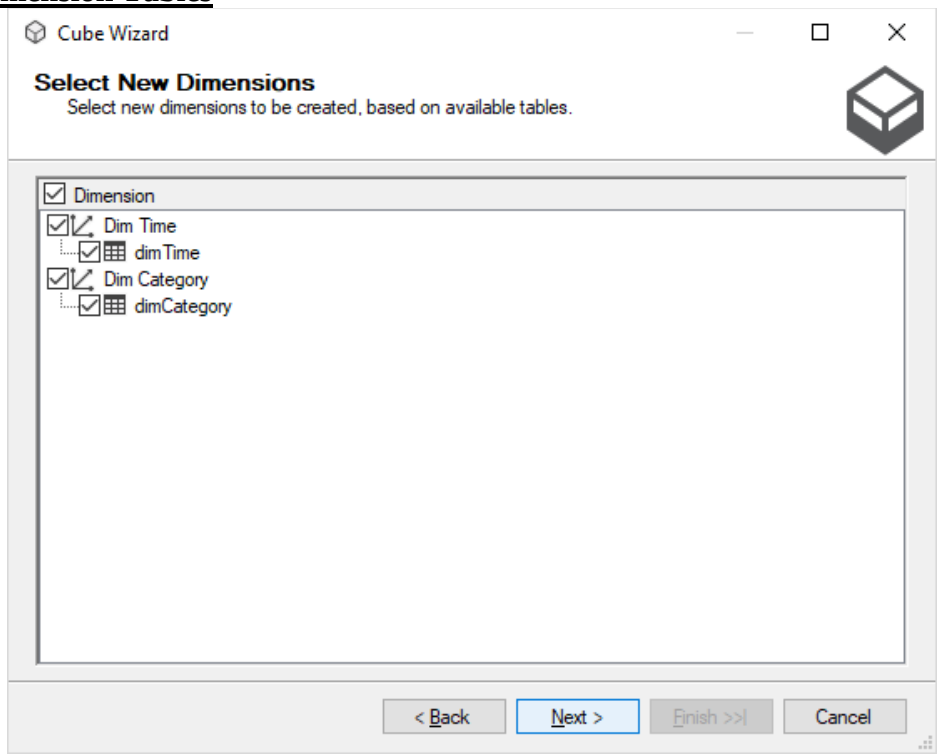

Figure 5.2-63: Selecting Dimensions

9. Cube Screenshot

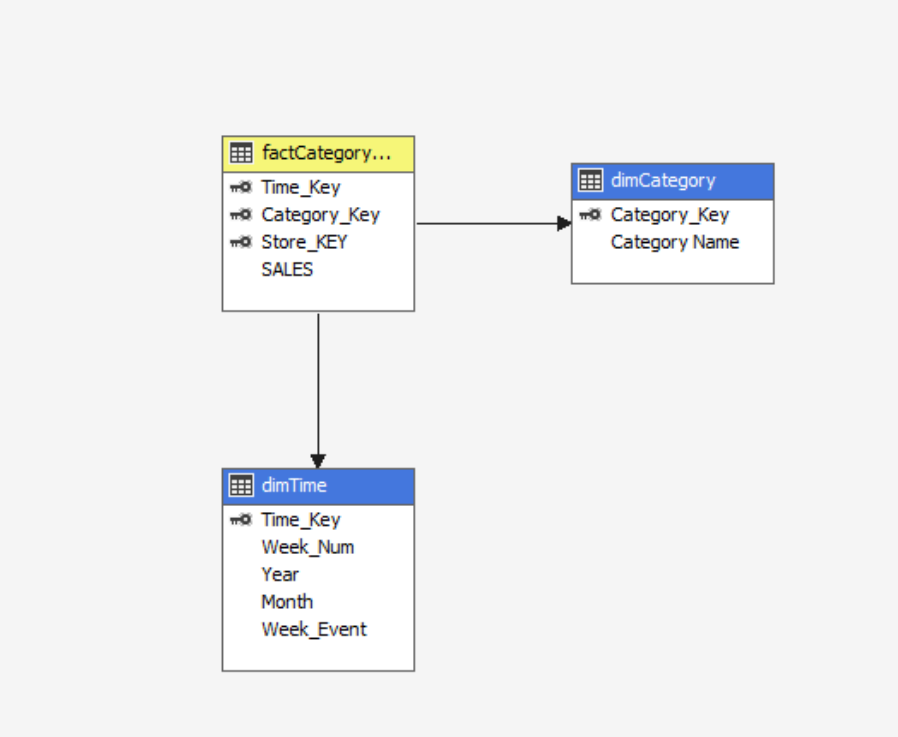

Figure 5.2-64: Cube

11. Create Time Hierarchy

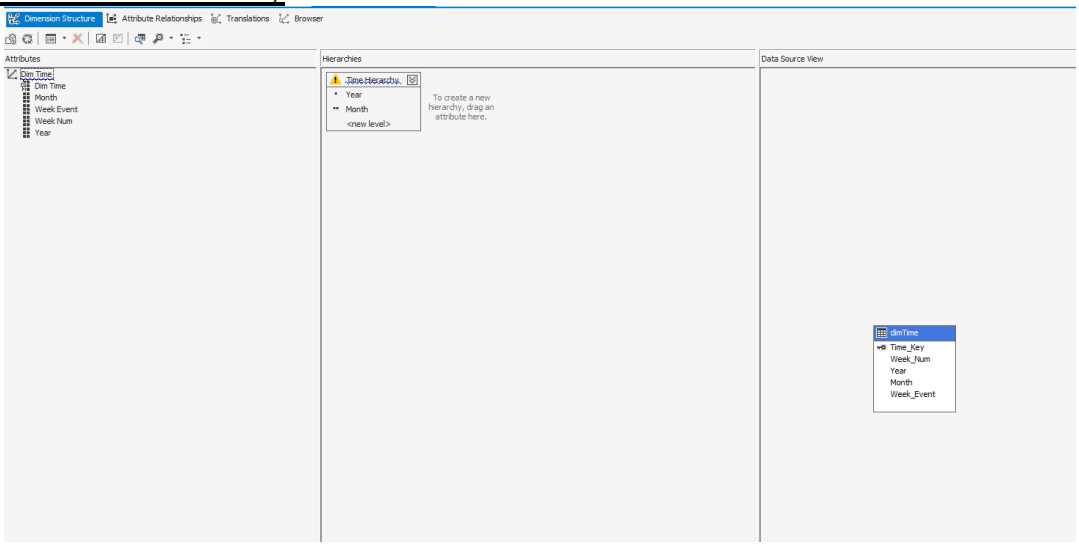

Figure 5.2-65: Time Dimension

12. Create Dim Category

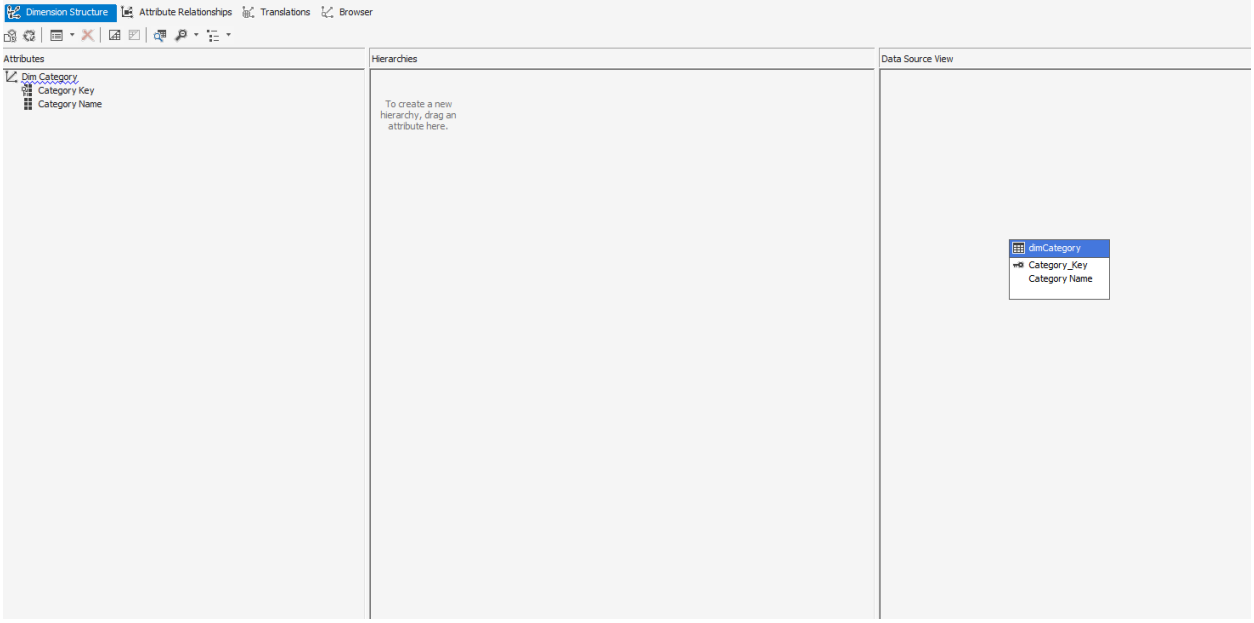

Figure 5.2-66: Category Dimension

13. Deploy Cube

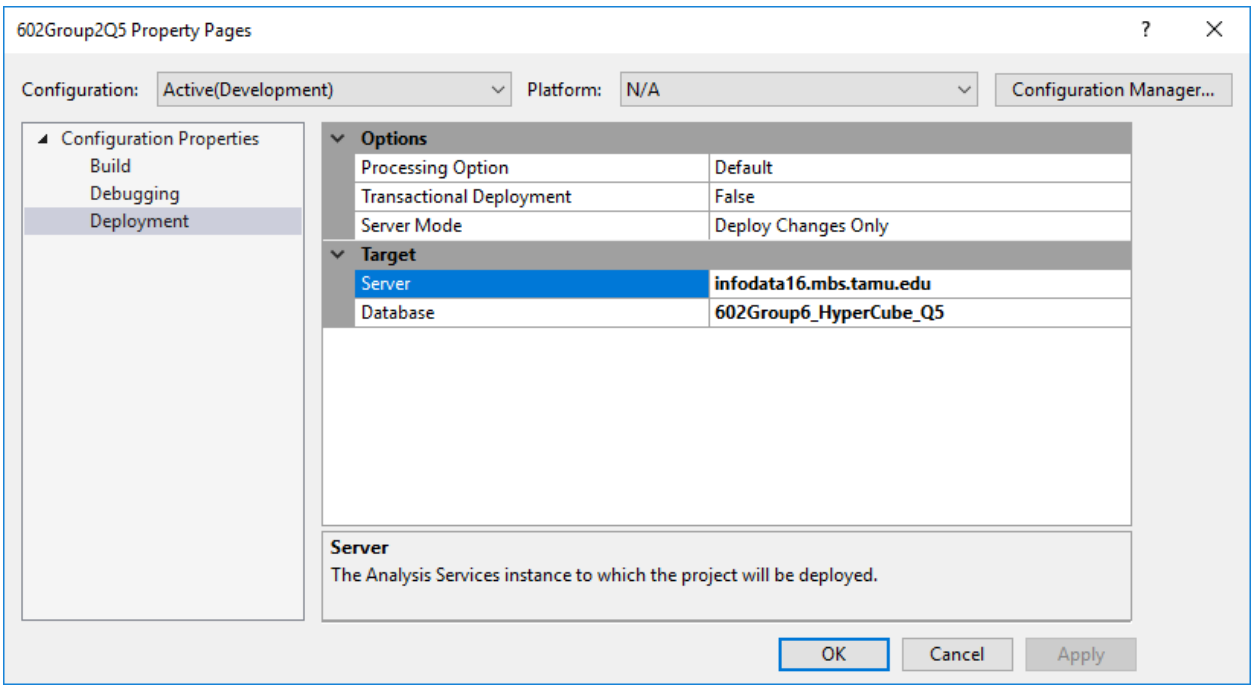

Figure 5.2-67: Selecting Target Server for cube deployment

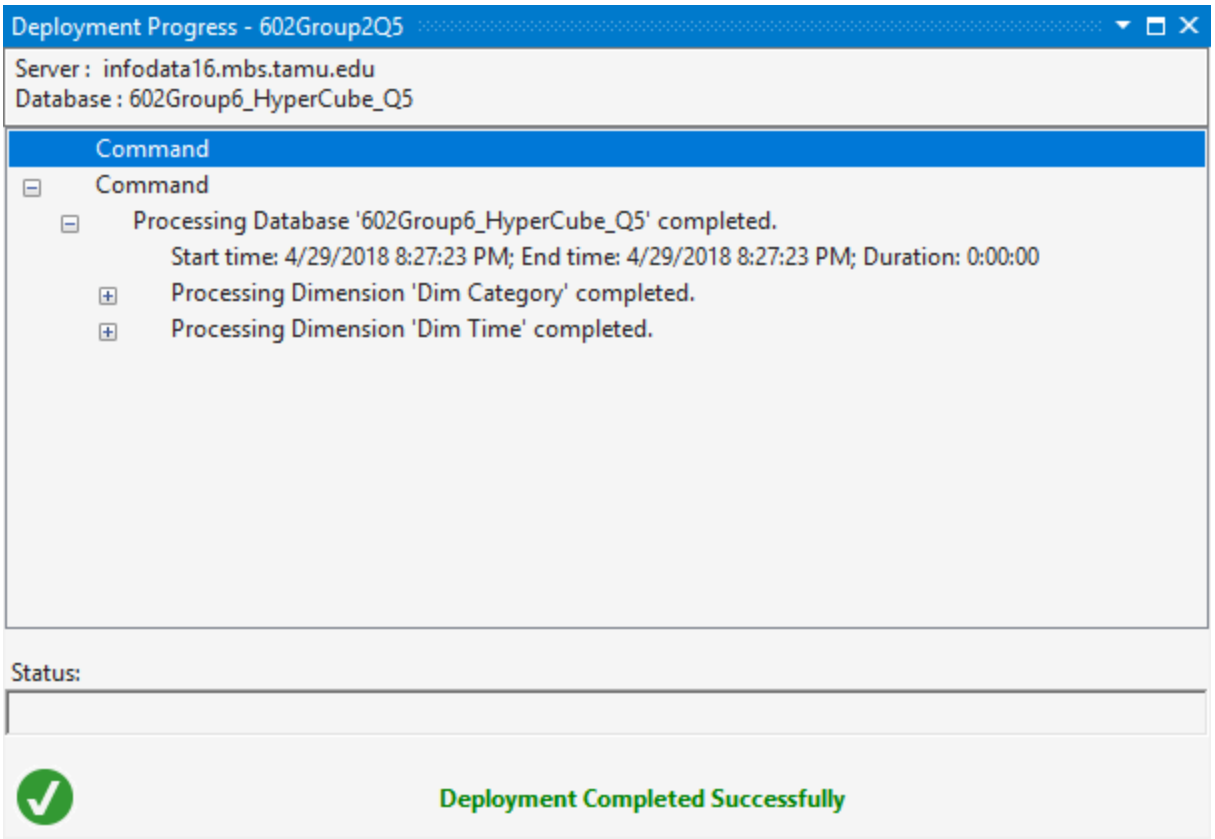

Figure 5.2-68: Cube deployment successful

14. Cube in SSAS

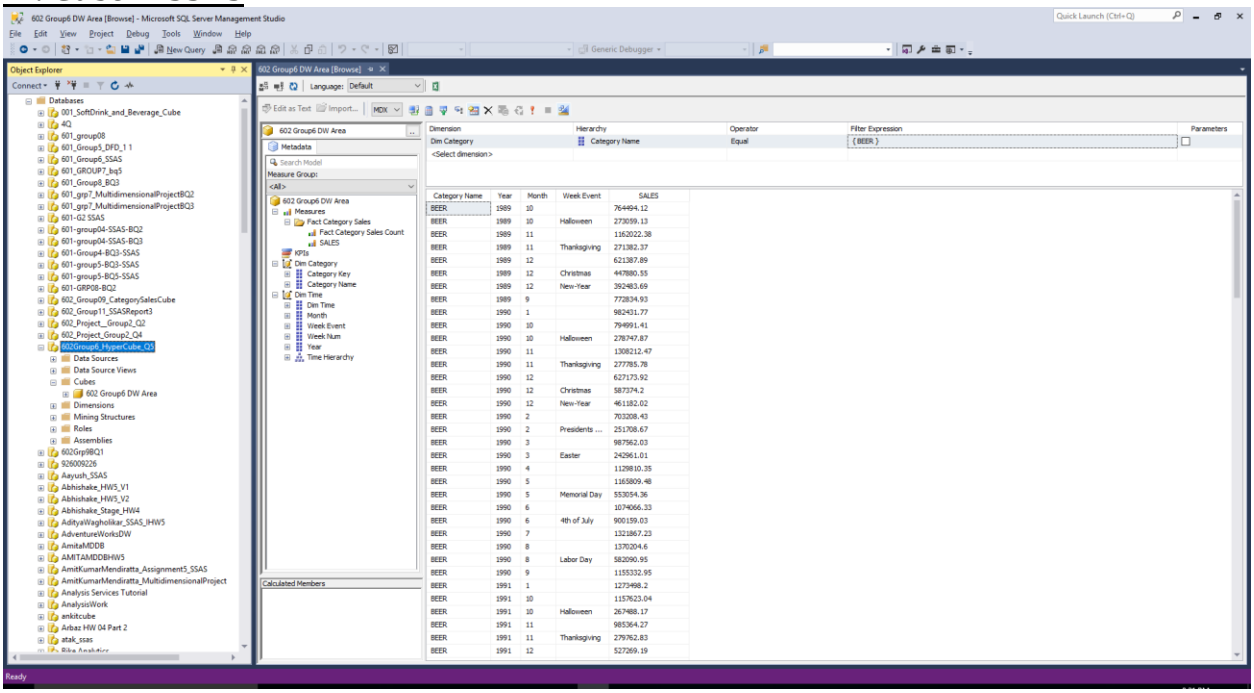

Figure 5.2-69: Cube Data in Solution Explorer

15. Create Data Source Connection to Cube from SSRS

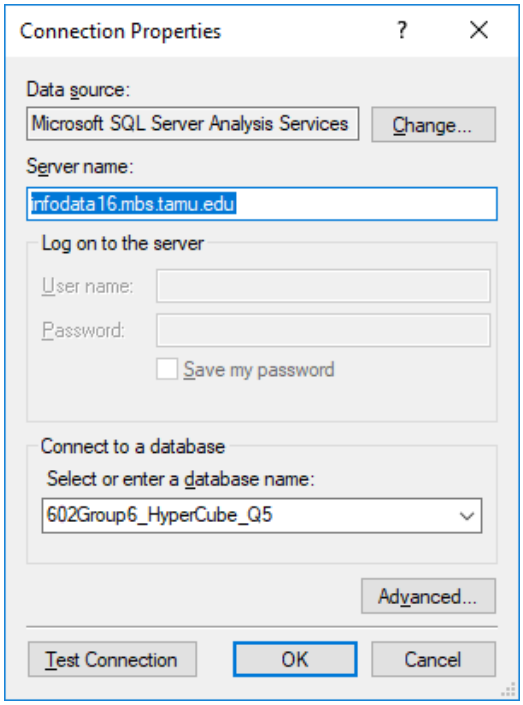

Figure 5.2-70: Data Source Connection

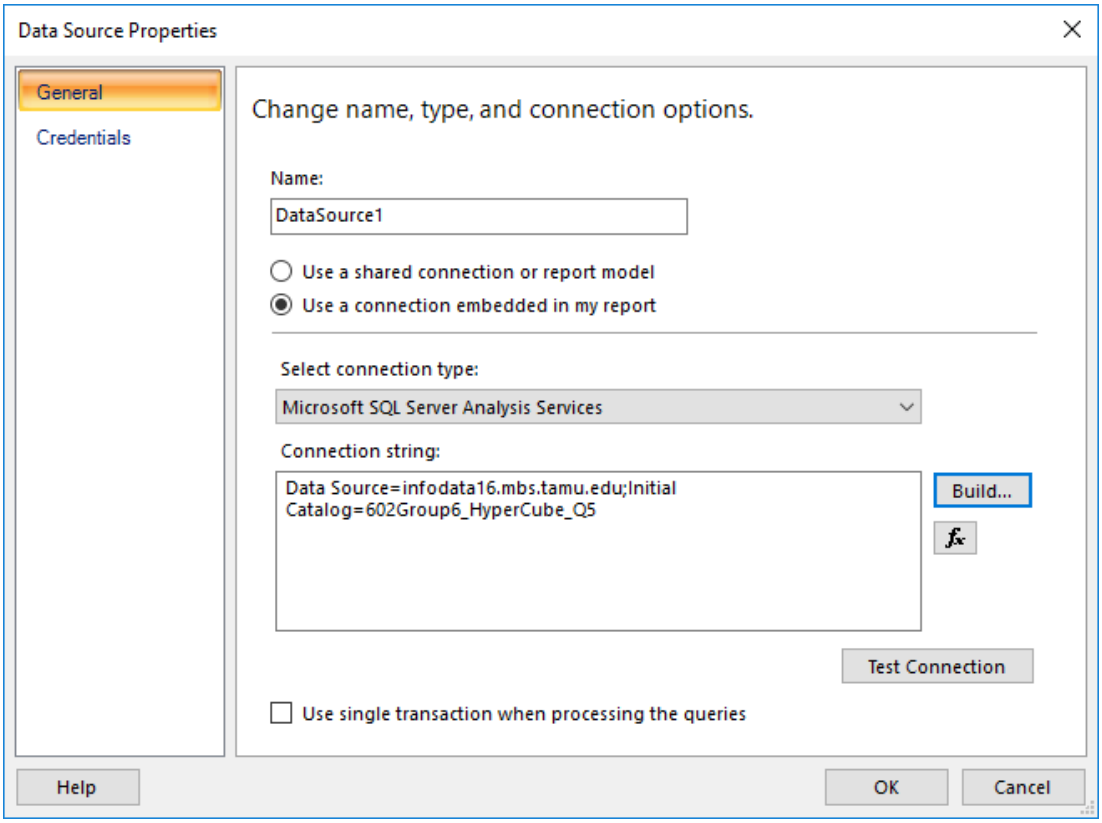

Figure 5.2-71: Data source Properties

16. Design Query

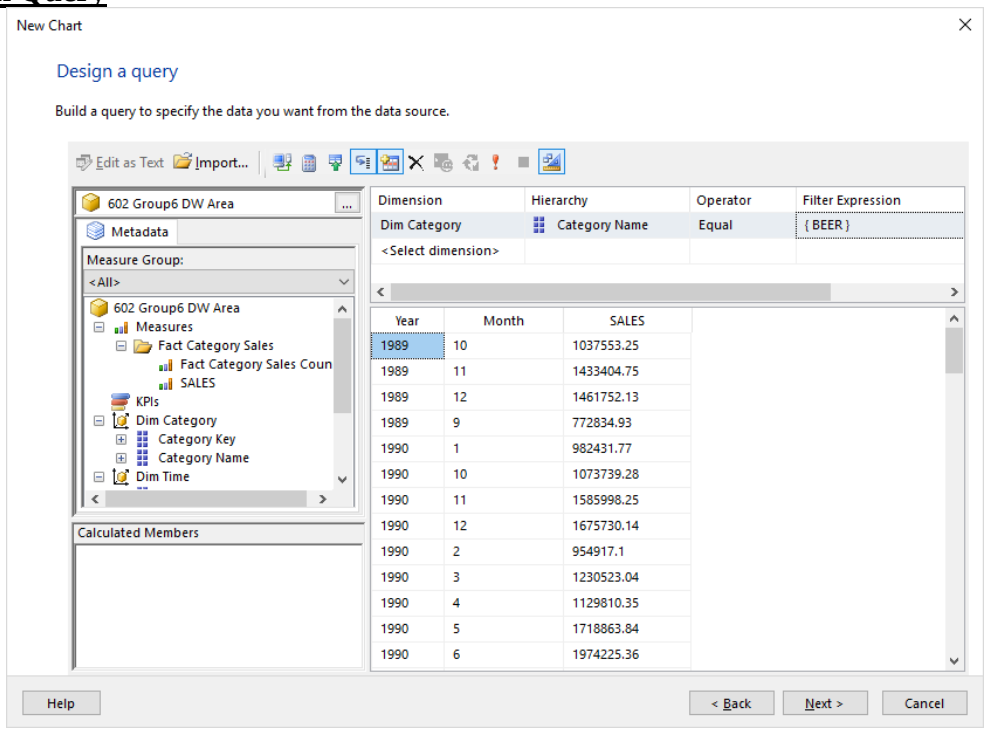

Figure 5.2-72: Design Query

17. Report preview in Design Mode in SSRS

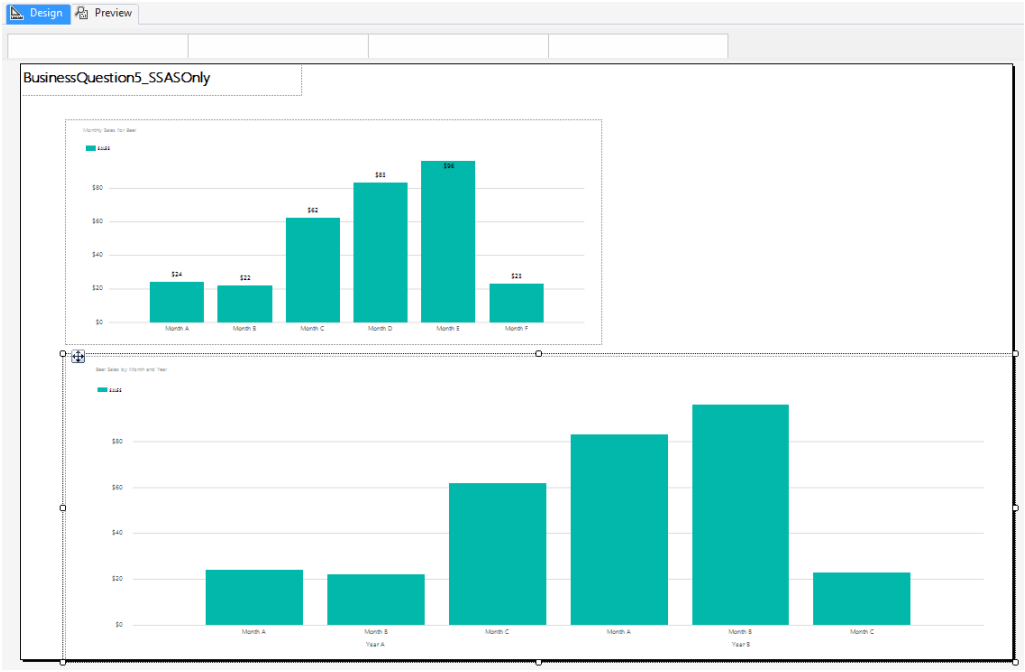

Figure 5.2-73: Report in Design Mode

18. Report preview in Preview Mode in SSRS

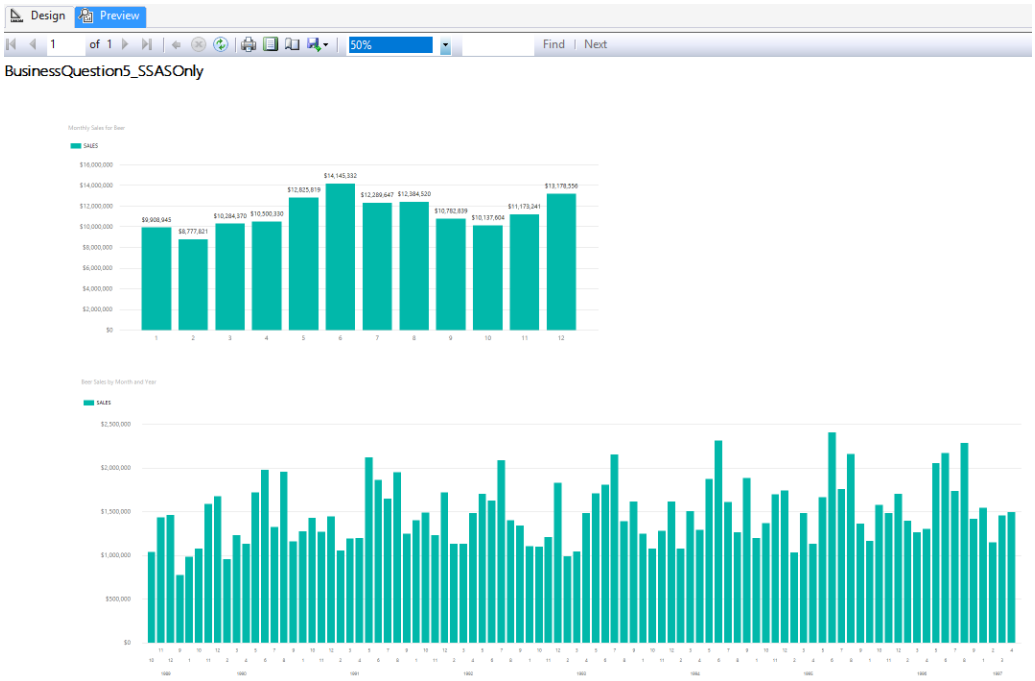

Figure 5.2-74: Report in preview Mode

19. Deploy the Report

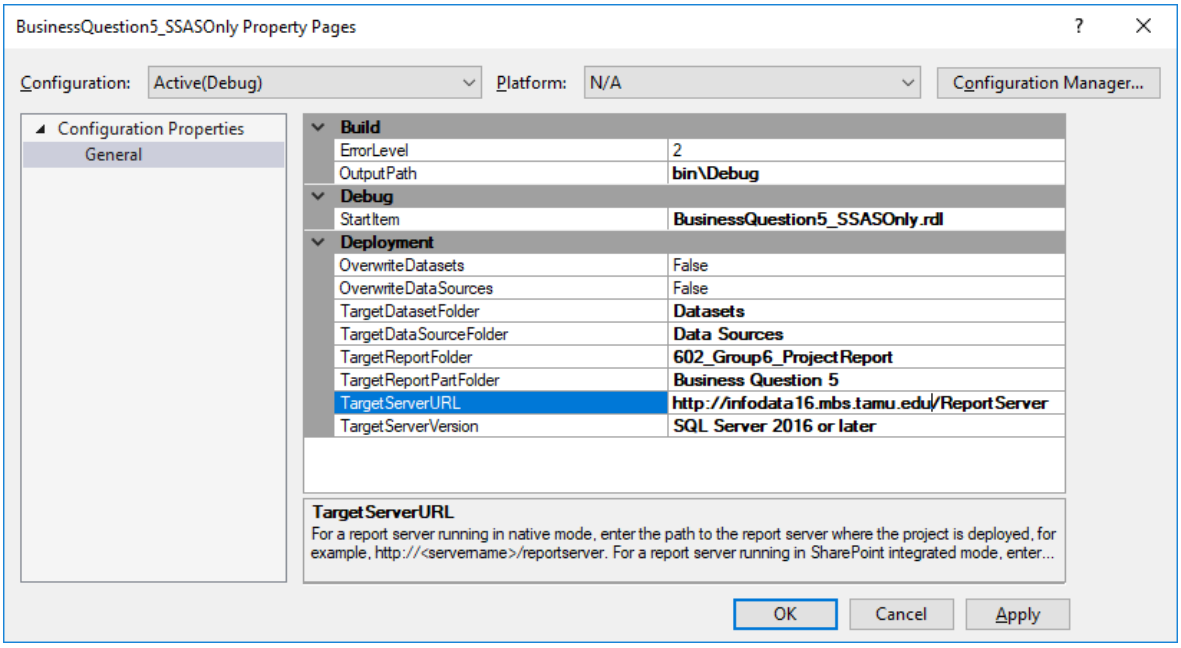

Figure 5.2-75: Deployment Target Source

20. Report Deployed on Report Server

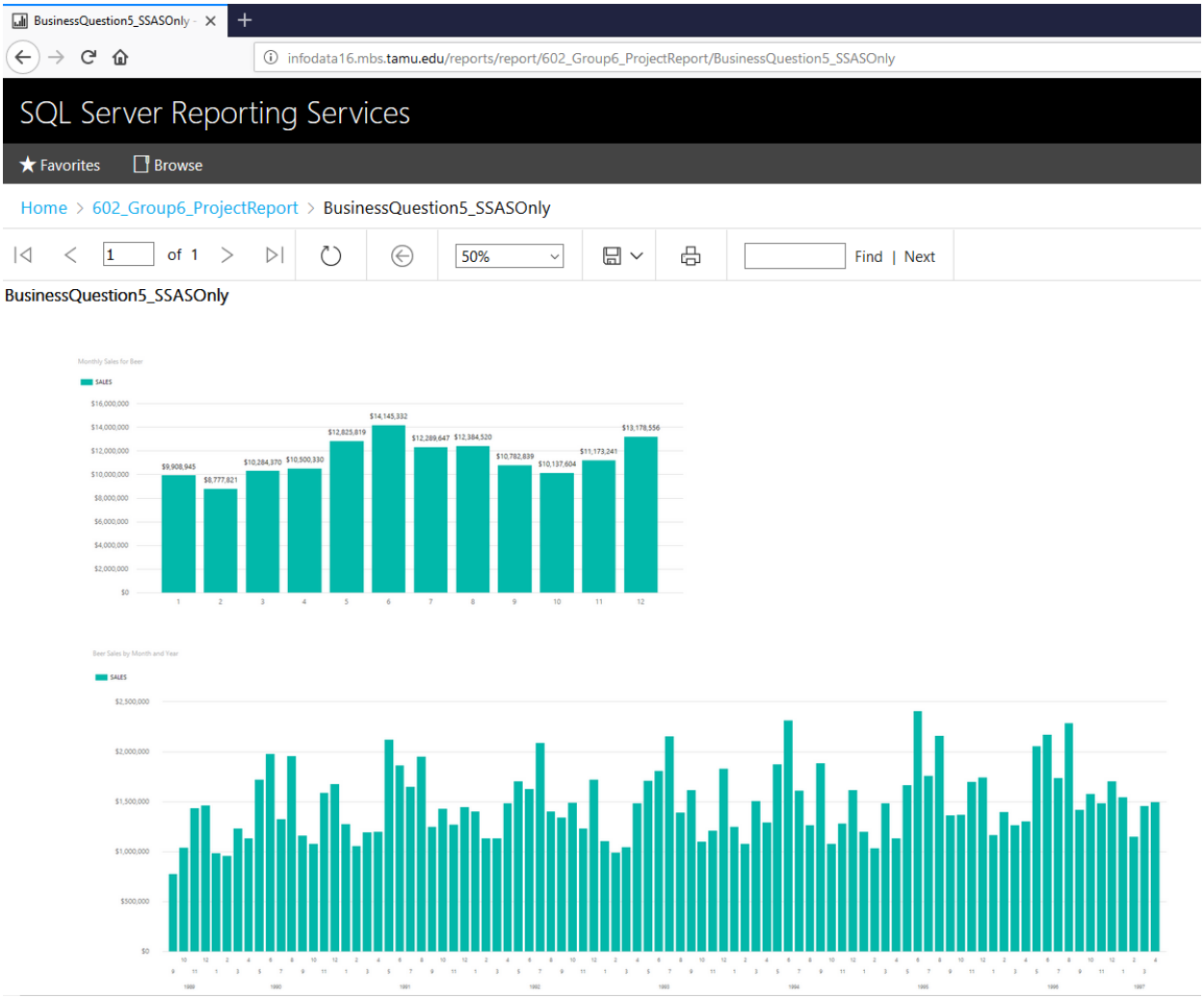

Figure 5.2-76: Report Deployed on Server

Conclusion

We used SSAS only to detect the trends in BEER sale over the special events throughout the year. We created two bar charts to show the Beer sale trends- one shows the monthly beer sale trends and the other shows the beer sales by year and month. These bar charts will help the Business managers to understand which special events cause a rise in the sale of beer. The graph of monthly beer sale indicates that beer sales are high during July and December. On further analysis of the week numbers, we figured that beer sales were high during 4th of July and Christmas week. The graph of beer sales by year and month shows that the trend remains pretty much same with higher sales during July and December. This analysis shows that beer must be stocked up during 4th of July and Christmas week.

6 References

1. Hamister, J. W., & Suresh, N. C. (2008). The impact of pricing policy on sales variability in a supermarket retail context. International Journal of Production Economics, 111(2), 441-455.doi:10.1016/j.ijpe.2007.01.011
2. Kamakura, Wagner A., and Wooseong Kang. “Chain-Wide and store-Level analysis for cross-Category management.” Journal of Retailing, vol. 83, no. 2, 2007, pp. 159–170., doi:10.1016/j.jretai.2006.02.006.
3. Nevo, Aviv, and Catherine Wolfram. “Why Do Manufacturers Issue Coupons? An Empirical Analysis of Breakfast Cereals.” The RAND Journal of Economics, vol. 33, no. 2, 2002, p. 319., doi:10.2307/3087436.
4. “Dominick's Database.” Dominick's Database | The University of Chicago Booth School of Business, research.chicagobooth.edu/kilts/marketing-databases/dominicks/.