Data Warehousing Group Project

Report - 4

ISYS 637 – Section 602

Group 5

Vamsi Kompella

Abhishek Pillai

Siddharth Sonti

Texas A&M University

May 2017

Contents

1. **Introduction** 4
2. **Metadata** 5
3. **Entity Relationship Diagram** 8
4. **Domain Understanding** 11
5. **Business Questions and Rationalizations** 13
6. **Independent Data Marts design using Kimball's approach** 26
   - a. **Dimension Tables** 26
     - i. **Product Dimension** 26
     - ii. **Store Dimension** 27
     - iii. **Demographics Dimension** 27
     - iv. **Time Dimension** 28
   - b. **Fact Tables** 29
     - **i. Sales Fact Table** 29
     - **ii. Promotion Sales Fact Table** 30
     - **iii. Demographics Fact Table** 30
   - c. **Define Dimension Matrix for data marts** 31
   - d. **Data Marts using STAR Schema** 32
     - **i. Marketing STAR Schema** 32
     - **ii. Strategy STAR Schema** 33
     - **iii. Overall Data Warehouse Schema** 34
7. **Mapping Tables** 35
   - a. **Mapping tables for Marketing Data Mart** 35
   - b. **Mapping tables for Strategy Data Mart** 37
8. **Justification of business questions corresponding to data marts** 39
9. **Development of ETL Plan** 40
   - a. **Determine all the target data needed in the data warehouse** 40
   - b. **Determining all the Data Sources** 40
   - c. **Data Mapping from Data Sources to Staging Area** 41
   - d. **Data Mapping from Staging Area to Data Warehouse** 42
   - e. **Establishing comprehensive data extraction rules** 43
   - f. **Determining data transformation and cleansing rules** 43
10. **Extraction and Transformation Implementation** 44

a. **WeekDecode Extraction and Transformation** 44
b. **Demographics Extraction and Transformation** 46
c. **Movement Extraction and Transformation** 48
d. **UPC Extraction and Transformation** 50
e. **CCOUNT Extraction and Transformation** 52
11. **Loading dimension and fact tables from Staging area to Data Warehouse** 55
12. **Overview of the data warehouse** 62
a. **Overview of all the Dimension tables in the data warehouse** 64
b. **Overview of all the Fact tables in the data warehouse** 68
13. **SQL statements to create Staging Area** 71
14. **SQL statements to create Data Warehouse** 73
15. **Removal of temporary tables that exists in the data staging area** 75
16. **Business Intelligence reporting with SSRS, SSAS and Report Builder** 76
a. **Reporting Plan** 76
b. **Determination of target reports that satisfy business questions** 76
c. **Mappings of the attributes from the data marts to the report** 78
17. **Report Building from Individual Data Marts** 83
18. **Task Sheet** 109

1. Introduction

Dominick's Finer Foods Inc. (DFF), referred to commonly as Dominick's, is a grocery store chain based in the Chicago area. They have their distribution center located in Northlake, Illinois, while their management offices are located in Oak Brook, Illinois. The chain was founded in 1918 by Dominick Dimatteo.

Throughout the years, DFF has been constantly striving to keep up with changing times. Just like any other business, there is always a need to adapt to the growing demands as well as dynamic needs of the customers. In this quest, DFF faced certain challenges. Although they had recorded the data of customers and stores over the past few years, they did not have the analytical processing resources to develop key insights into customer demographics, store sales statistics and data relating to the various products in the store. DFF had no way to cater to the needs of the customers without cars and were possibly losing out on customer retention. There is no accurate information available to support decisions like inventory management in stores with low sales with respect to products like jewelry and managing the supply of perishable products such as bakery items and meat over holiday seasons. In addition to all this, DFF may also be losing out on targeted marketing by not observing the buying trends of customer demographics. The stores still cater to a shrinking married household population as opposed to addressing the preferences of an ever growing single household population. Proper collection and sophisticated analysis of data can help a retail chain like Dominick's understand the needs of their customers and stay ahead of competitors.

To understand the data, we downloaded customer count sheet, demographic sheet, some of the UPC and movement files for commodities like grooming products, beer and cereals. The demographic sheet contains information for a particular store including attributes such as address and population demographic that visit the store where store number attribute can be considered as the primary key. The customer count table consists of data for the store sales for particular day which references store number attribute of the demographic sheet. Date and Store number form the composite primary key for this table. UPC sheets contains information about the products in stock including attributes such as the category, description and size. The movement table works as an associative entity between UPC and demographic table. It contains information about the sales, movement and profits across stores for a week.

We combined UPC data with the movement data using the VLookUp function in Excel in order to understand the sales of the commodities better and figure out the sales trend and profit margins for the products involved. We also combined this data with the demographic sheet to find out which areas sold most of the products and area wise profit margin. We used pivot tables on customer count sheet in order to figure out sales of products across the years by eliminating rows which had no valid dates.

2. Metadata

Variable	Description	Type	Length
DATE	Date of the Observation	Character	6
Week	Week Number	Numeric	8
Store	Store Code	Numeric	8
BAKCOUP	Bakery Coupons Redeemed	Numeric	8
BAKERY	Bakery Sales in Dollars	Numeric	8
BEER	Beer Sales in Dollars	Numeric	8
BOTTLE	Bottle Sales in Dollars	Numeric	8
BULK	Bulk Sales in Dollars	Numeric	8
BULKCOUP	Bulk Coupons Redeemed	Numeric	8
CAMERA	Camera Sales in Dollars	Numeric	8
CHEESE	Cheese Sales in Dollars	Numeric	8
CONVFOOD	Conventional Foods Sales in Dollars	Numeric	8
COSMCOUP	Cosmetics Coupons Redeemed	Numeric	8
COSMETIC	Cosmetics Sales in Dollars	Numeric	8
CUSTCOUN	Customer Count	Numeric	8
DAIRCOUP	Dairy Coupons Redeemed	Numeric	8
DAIRY	Dairy Sales in Dollars	Numeric	8
DELI	Deli Sales in Dollars	Numeric	8
DELICOUP	Deli Coupons Redeemed	Numeric	8
DELIEXPR	Deli Express Sales in Dollars	Numeric	8
DELISELF	Deli Self Service Sales in Dollars	Numeric	8
FISH	Fish Sales in Dollars	Numeric	8
FISHCOUP	Fish Coupons Redeemed	Numeric	8
FLORAL	Floral Sales in Dollars	Numeric	8
FLORCOUP	Floral Coupons Redeemed	Numeric	8
FROZCOUP	Frozen Items Coupons Redeemed	Numeric	8
FROZEN	Frozen Items Sales	Numeric	8
FTGCCOUP	Food-to-Go Coupons Redeemed	Numeric	8
FTGCHIN	Food-to-Go Chinese Sales in Dollars	Numeric	8
FTGICOUP	Food-to-Go Coupons Redeemed	Numeric	8
FTGITAL	Food-to-Go Italian Sales in Dollars	Numeric	8
GM	General Merchandise Sales in Dollars	Numeric	8
GMCOUP	General Coupons Redeemed	Numeric	8
GROCCOUP	Grocery Coupons Redeemed	Numeric	8
GROCERY	Grocery Sales in Dollars	Numeric	8
HABA	Health and Beauty Aids Sales in Dollars	Numeric	8
HABACOUP	Health and Beauty Aids Coupons Redeemed	Numeric	8
JEWELRY	Jewelry Sales in Dollars	Numeric	8
LIQCOUP	Liquor Coupons Redeemed	Numeric	8
MANCOUP	Manufacturer Coupons Redeemed	Numeric	8
MEAT	Meat Sales in Dollars	Numeric	8
MEATCOUP	Meat Coupons Redeemed	Numeric	8

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

Customer count table

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

Movement table

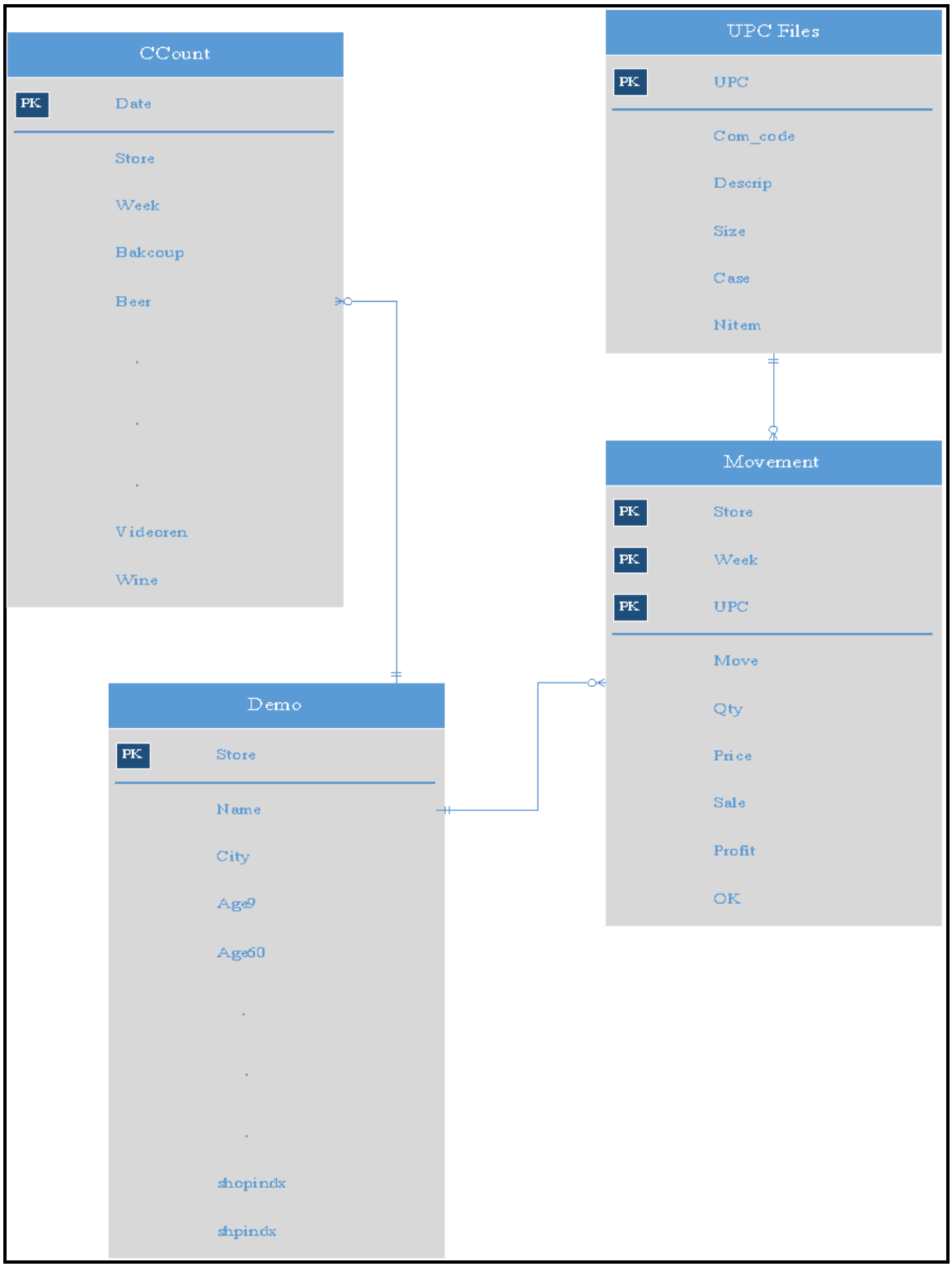
Variable	Description	Type	Length
upc	UPC Number	Numeric	8
com_code	Dominick's Commodity Code	Numeric	8
nitem	Dominick's item code	Numeric	8
descrip	Product Name	Character	20
size	Product Size	Character	6
case	Number of items in a case	Numeric	8

UPC table

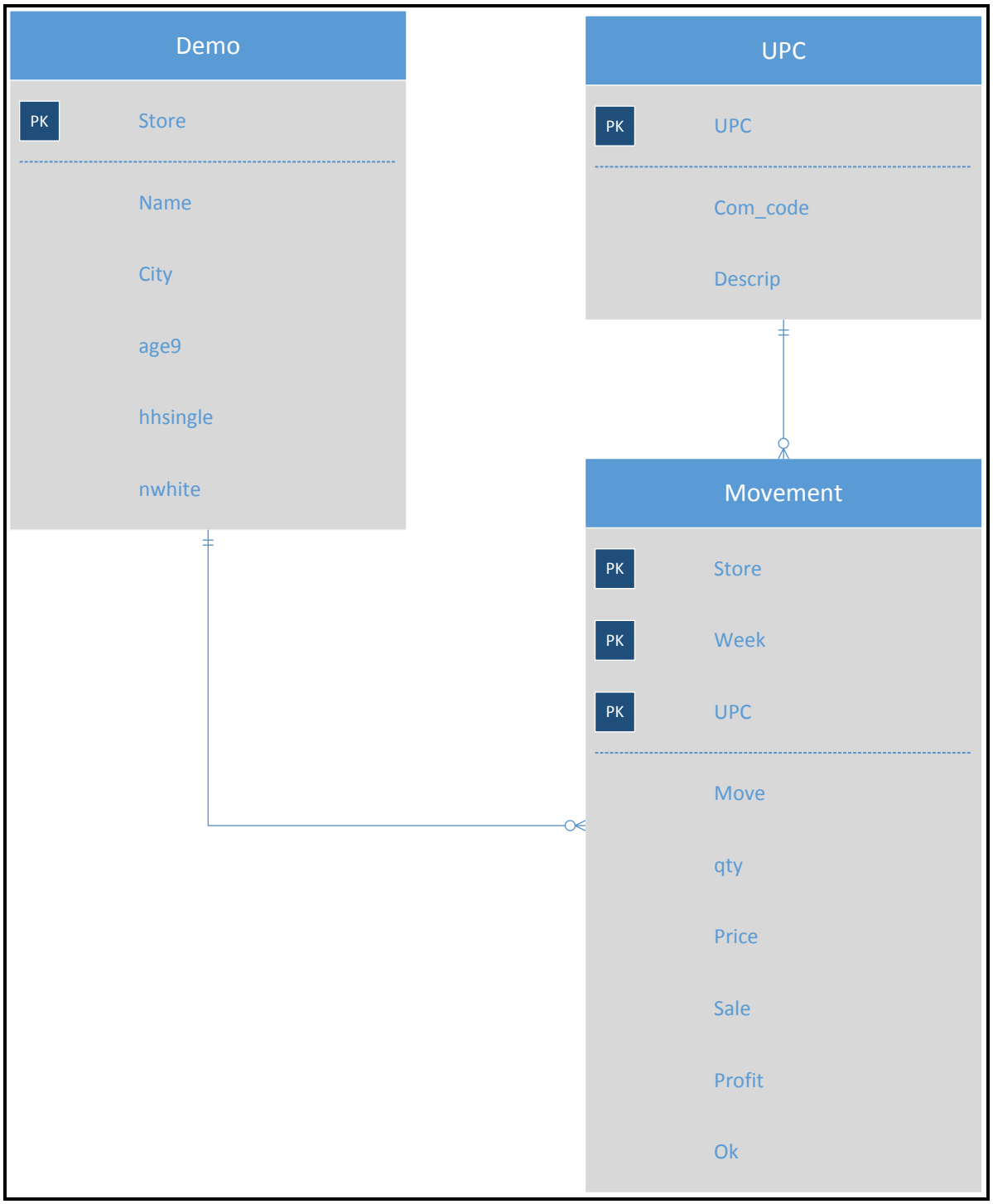
Variable Name	Description
age9	% Population under age 9
age60	% Population over age 60
ethnic	% Blacks & Hispanics
educ	% College Graduates
nocar	% With No Vehicles
income	Log of Median Income
incsigma	Std dev of Income Distribution (Approximated)
hsizeavg	Average Household Size
hsize1	% of households with 1 person
hsize2	% of households with 2 persons
hsize34	% of households with 3 or 4 persons
hsize567	% of households with 5 or more persons
hh3plus	% of households with 3 or more persons
hh4plus	% of households with 4 or more persons
hhsingle	% of households with 1 person
hhlarge	% of households with 5 or more persons
workwom	% Working Women with full-time jobs
sinhouse	% Detached Houses
density	Trading Area in Sq Miles per Capita
hval150	% of Households with Value over $150,000
hval200	% of Households with Value over $200,000
hvalmean	Mean Household Value (Approximated)
single	% of Singles
retired	% of Retired
unemp	% of Unemployed
wrkch5	% of working women with children under 5
wrkch17	% of working women with children 6 - 17
nwrkch5	% of non-working women with children under 5
nwrkch17	% of non-working women with children 6 - 17
wrkch	% of working women with children
nwrkch	% of non-working women with children
wrkwch	% of working women with children under 5

Demographic table

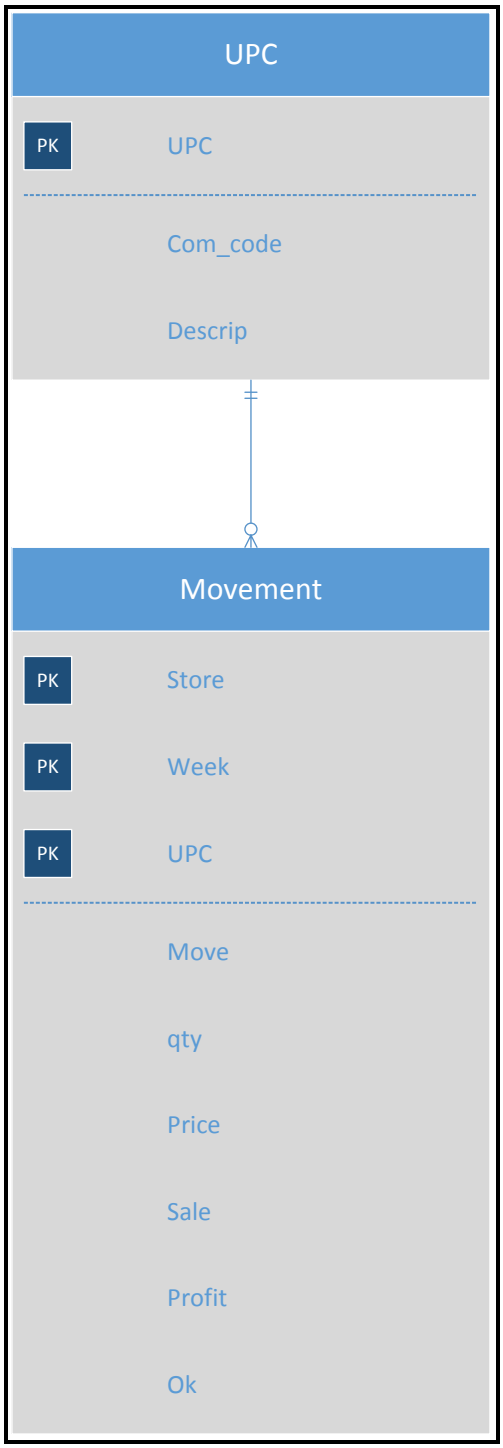
3. Entity Relationship Diagram



Entity-Relationship Diagram



Entity-Relationship Diagram for business questions 4, 6, 7 and 9



Entity-Relationship Diagram for business question 8

4. Domain Understanding

John Stanton mentions in his article [1] that America now has a majority of households that consists of single individuals as compared to married households. In fact, married households account has declined from 24 percent to 21 percent in 2010. This result provides a great deal of information to retail chains and food stores as single household consumers are valuable target customers. Single households are tipped to be the high consumers of resources and energy as they consist of mostly young professionals who can afford housing. An Hodgson [2] states that single person households are less interested in combo packs or large packs that have undesired items. Instead, the retail chain can focus on items that sell in small portions as it appeals to them more.

According to Dan Pahomi and Peter Leith of JustEnough [5], a demand management solutions provider for retail, wholesale and direct-to-customer businesses, analyzing the data on promotional sales provides a company with valuable insights that help in ultimately increasing the sales and the products purchased in a single transaction. This data analysis, according to them, helps evaluate the effectiveness of promotional coupons and discounts in the past to prepare strategies that target channels other than print media and TV advertisements. Also, the article by Analytic Partners [6] mentions that identifying the actual ROI and profitability of such a promotional campaign can help answer another important question as to whether the promotional frequency should be increased or not.

Daphne Howland explains in her article [7] that over 61% of retailers across the globe have some sort of "click and collect" in place in an effort to reduce the effort required by customers to place orders and complete the shopping trip. A hurried shopper is an easy to lose customer and something as trivial as the store being on the wrong side of the road could lead to a diminishing base. Empirically, it has been found that giving customers multiple options in ways to finish their order could lead to increased customer retention. A pilot program by CVS Pharmacy in San Francisco, Atlanta and Charlotte, NC has revealed that the curbside pickup solution has resulted in a 56% increase in repeat shoppers [8]. Other instances of similar services have all resulted in an increase in the customer loyalty to the business. The same approach could be tried by Dominick's in a phased approach to see if it makes business sense.

Maria Luisa Alaniz mentions in her paper [3] how neighborhood characteristics deeply affect alcohol consumption in the region. She provides examples of how neighborhoods consisting mostly of non-white population advertise alcohol and smoking that would increase the sales and profits if the popular brands are known. The Beer Industry Electronic Commerce Coalition, in their journal [4] provides the advantage of analyzing the demand in the market, it would lead to cost reduction in the distributor and supplier network. The journal talks about how historical data can be used in forecasting demand in different areas or neighborhoods. In fact, it is essential for an analyst to predict the market conditions using historical data and use these established trends in order to meet the supply and demand chain. The same principle can be applied to different commodities that include cereal boxes, front-end candies and grooming products for a target audience based on research.

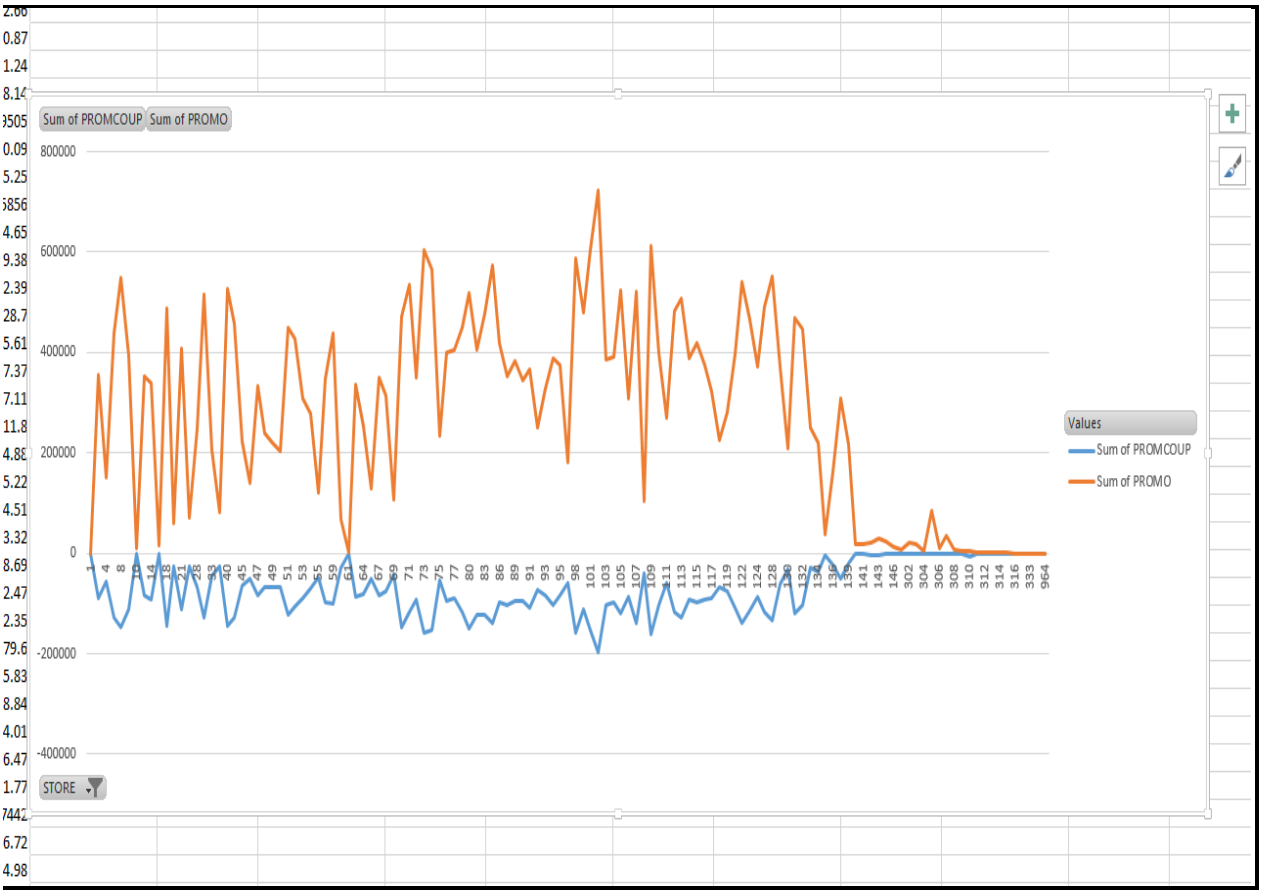
References:

1. *Editor, J. S. (n.d.). A Closer Look at the Single Household. Retrieved February 28, 2017, from http://www.foodprocessing.com/articles/2013/market-view-single-household/*
2. *One person households: Opportunities for consumer goods companies. (2016, December 28) from http://blog.euromonitor.com/2007/09/one-person-households-opportunities-for-consumer-goods-companies.html*
3. *Alaniz, M. L. (n.d.). Alcohol Availability and Targeted Advertising in Racial/Ethnic Minority Communities.*
4. *B. (n.d.). Hitting the Mark: Accuracy in Beer Sales Forecasting. Sales forecasting.*
5. *Pahomi, D., & Leith, P. (n.d.). Measuring the Effectiveness of Promotions.*
6. *Promotional Effectiveness. (n.d.). Retrieved February 28, 2017, from https://analyticpartners.com/promotional-effectiveness/*
7. *Howland, D. (2016, May 18). How retailers can use curbside pickup to build customer loyalty. Retrieved March 01, 2017, from http://www.retaildive.com/news/how-retailers-can-use-curbside-pickup-to-build-customer-loyalty/418801/*
8. *MMR, O. 3. (2016, September 26). CVS curbside goes nationwide. Retrieved March 01, 2017, from http://www.massmarketretailers.com/cvs-curbside-goes-nationwide/*

5. Business Questions and Rationalizations

This section explains the various lines of business ideas that were identified from DFF's available data. They provide some thinking guidelines on how the existing business could be improved by providing statistical trends extracted from the data over the past few years. The questions in this section are prioritized in the decreasing order of business importance and value. For example, it is of higher business value to cater to people who have no access to a vehicle than to optimize the inventory holding of jewelry as the latter would play a major role in increasing DFF's market share and raising the number of repeat customers. Each question is accompanied by the supporting data and rationalization as to why it is important for DFF.

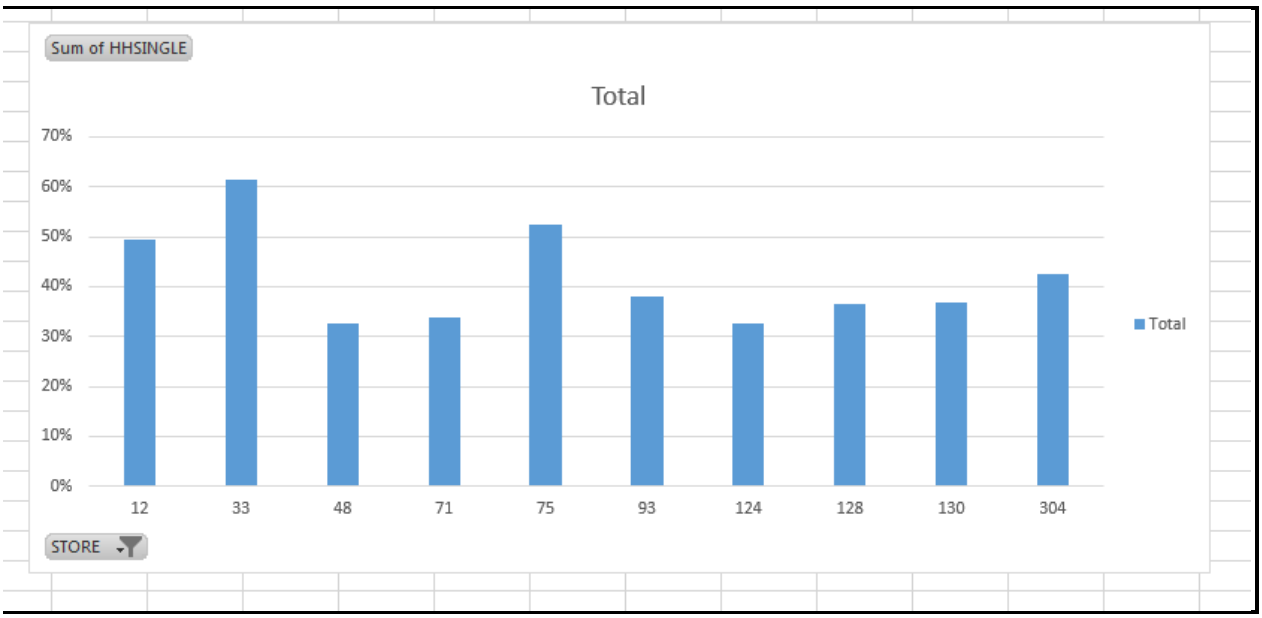
1. How have the promo sales been affected across the stores on the basis of coupons?



Coupon sales across stores

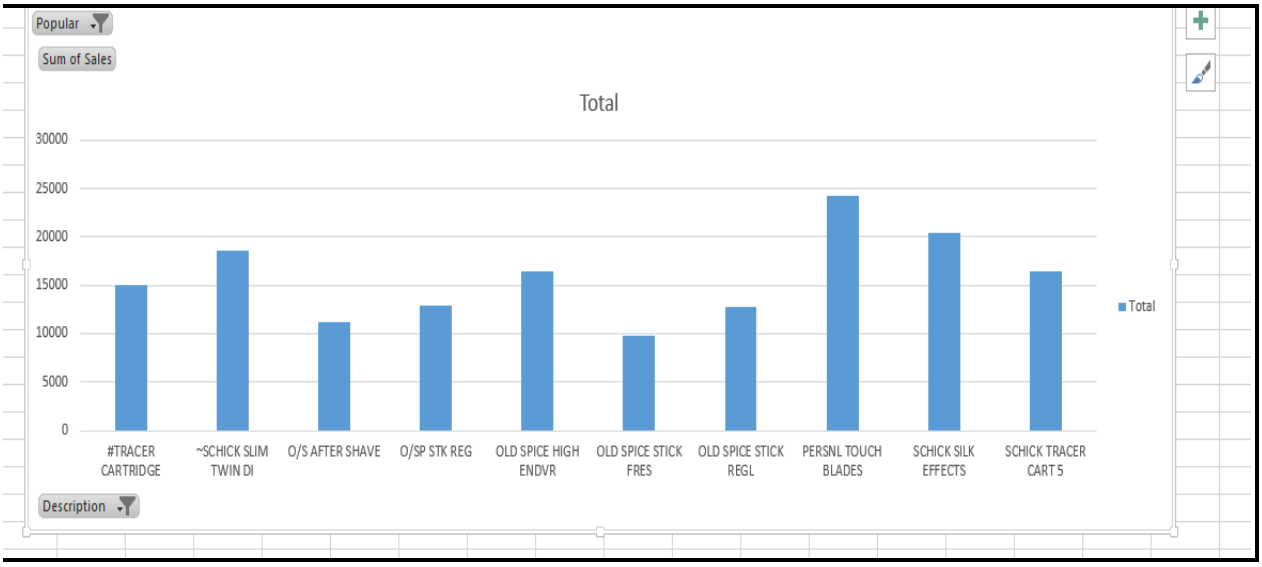
<u>Rationalization:</u> The pivot chart shows promo sales in orange and promo coupons use in blue for all the years plotted against the stores. As per the graph, promo coupon usage is proportional to sales leading to the conclusion that stores that have more coupons redeemed have more sales. This information can be used to increase sales across stores.

2. **Which stores are in areas of high single households and what are the commodities/products that sell more and which sell least?**



Stores in areas with high single household population

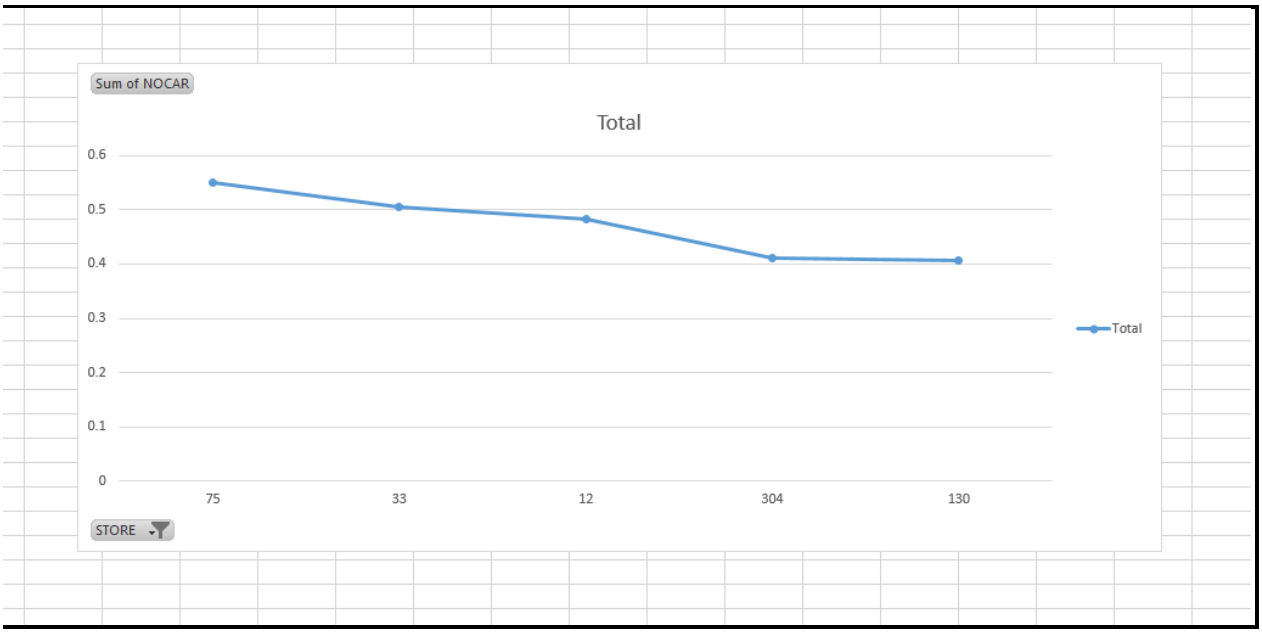
Rationalization: Single households constitute about half the customer base in a store. Therefore, it is essential to consider the products that are popular among single households. Single households often take up more resources and energy and will be the fastest rising household group, which will be very influential in shaping future consumer lifestyles. Since we cannot combine all the movement sheets, for this example we considered grooming products which are highly popular with single households.



Top products in areas with high single household population

3. **Which stores or store areas have a high percentage of customers with no cars in order to set up a home delivery service?**

Let's have a look at the stores that have more than 40% population having no cars

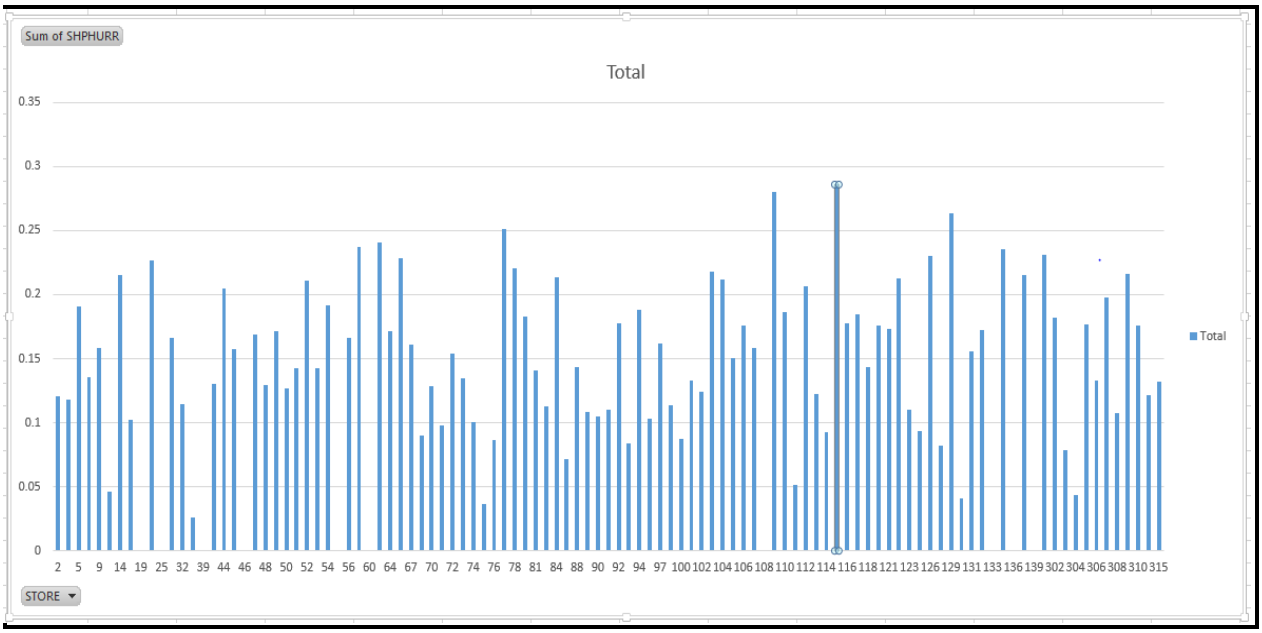


Stores in areas with more than 40% population with no cars

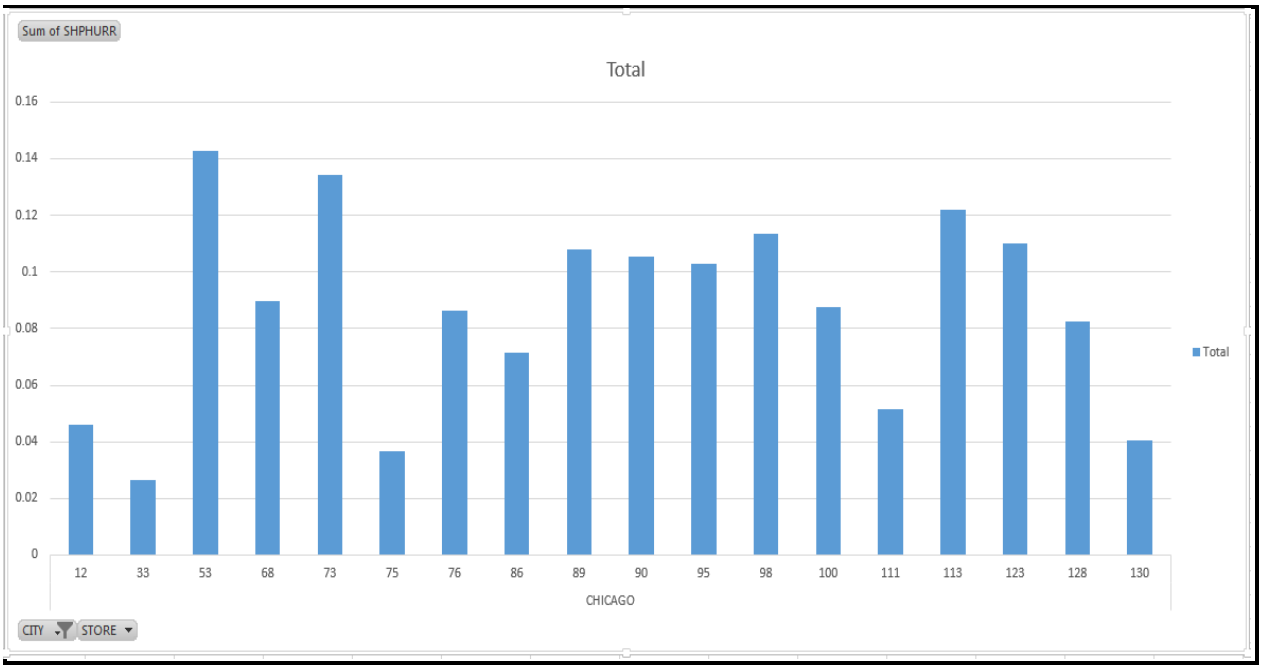
Rationalization: A store in an area having a high percentage of people having no cars can set up a home delivery service that will boost the sales. People having no cars would not regularly visit the store and would often buy a lot of things from nearby stores. Many stores have a home delivery service, however placing those services in the right stores would help in improving customer relations as well as drive up the profits.

4. **Which stores have a large number of hurried shoppers and would benefit from setting up a curbside pickup?**

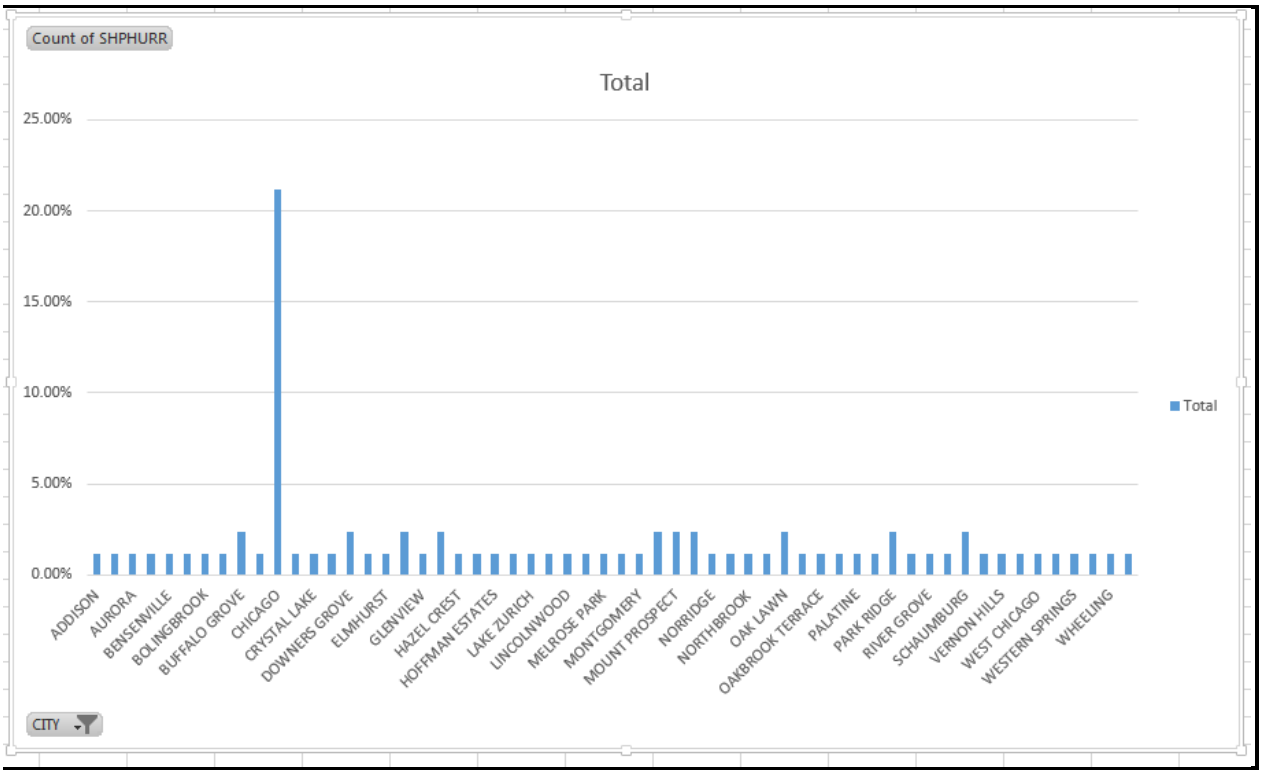
Rationalization: Curbside pickup is a new and coming business idea that could help stores in areas that have a high percentage of hurried customers save their time by being ready with their order curbside at a time slot convenient to the buyer. This has numerous benefits to both the store as well as the customer. first, for the customers, they don't have to search for a parking, walk all the way to various, spread out aisles, wait in line for checkout and walk all the way back. They could just place the order online or by phone and their order would be ready at a time suitable for them. Second, this is also beneficial to the store as it would result in freeing up store space for other customers in the event of high demand. The store could also try charging a premium to the curbside service, which is a small price to pay for a highly convenient service.



Percentage of hurried shoppers spread out across the stores



Percentage of hurried shoppers within Chicago area. Notice that over 50% have greater than 10% hurried shoppers.

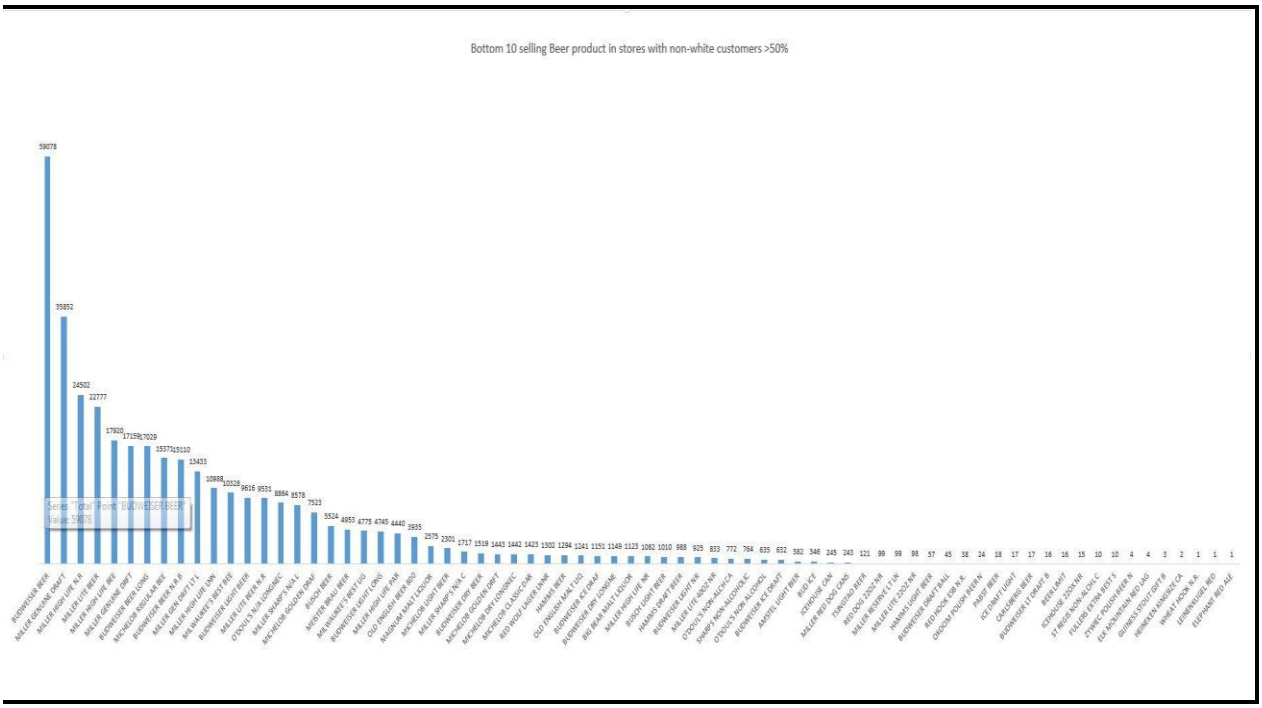


Count of stores that have a significant number of hurried shoppers

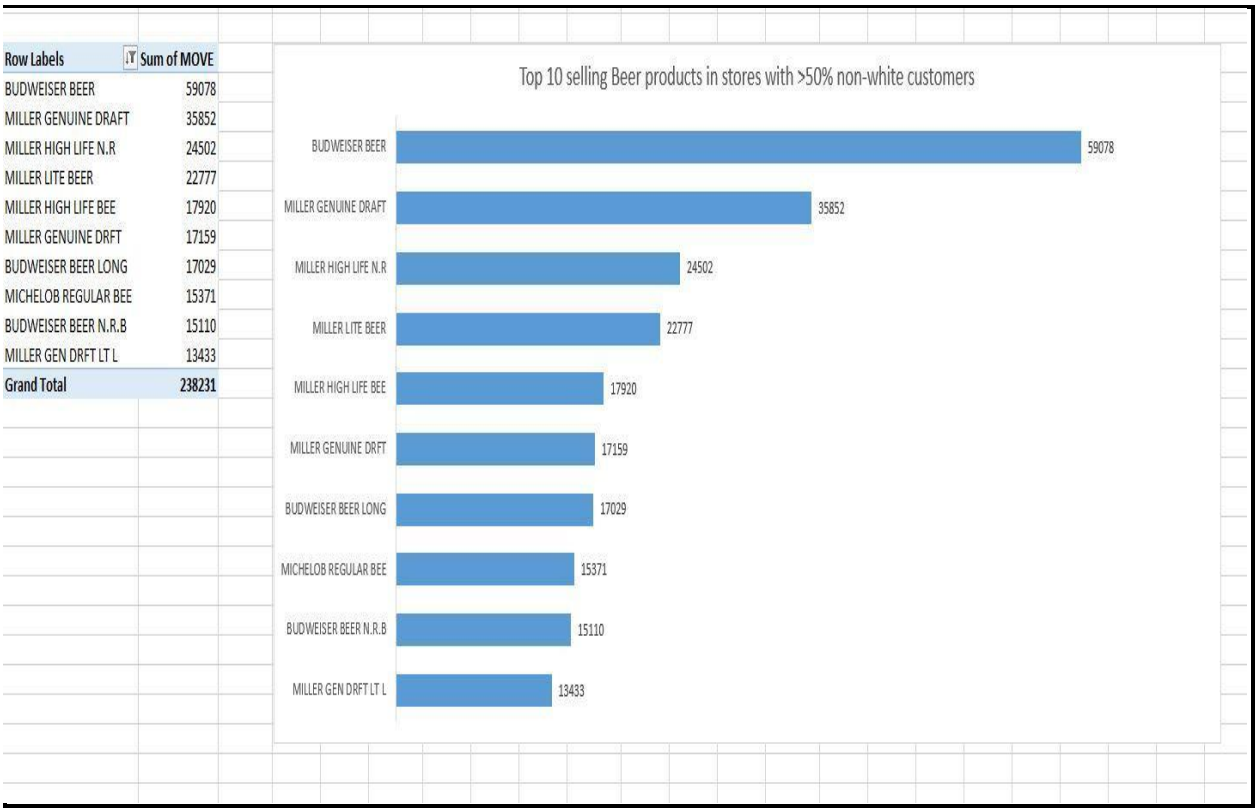
From the pivot chart images, we could see that a large number of stores have over 20% of its customers doing hurried shopping. Of these, a vast majority are from Chicago. This trend is understandable as Chicago area is one of the most densely populated areas in the US. Rolling out the service in a few stores in many cities is a bad business decision as we do not know the probability of success. It would be safe to assume that the places with highest percentage of hurried customers would see the most success for curbside pickup. We find that Chicago area has the most number or stores and almost all stores have an average of 20% hurried customers. Also, there are about 18 stores in this area and it makes a lot of business sense to roll out the service in this area to see its results.

5. **How is the sales performance of all the beer brands across the stores in areas with more than 50% non-white population in the last 6 years (1991 - 1997)?**

Rationalization: For any retail organization, analyzing customer demographic data provides an understanding of customer behavior and buying patterns. Identifying these patterns and correlations helps in making better business decisions through targeted marketing, improved management of inventory and pricing strategies. The current business question provides DFF with insights into the popularity of all brands of Beer in their inventory over the years across the stores in areas consisting of a majority of non-white population.

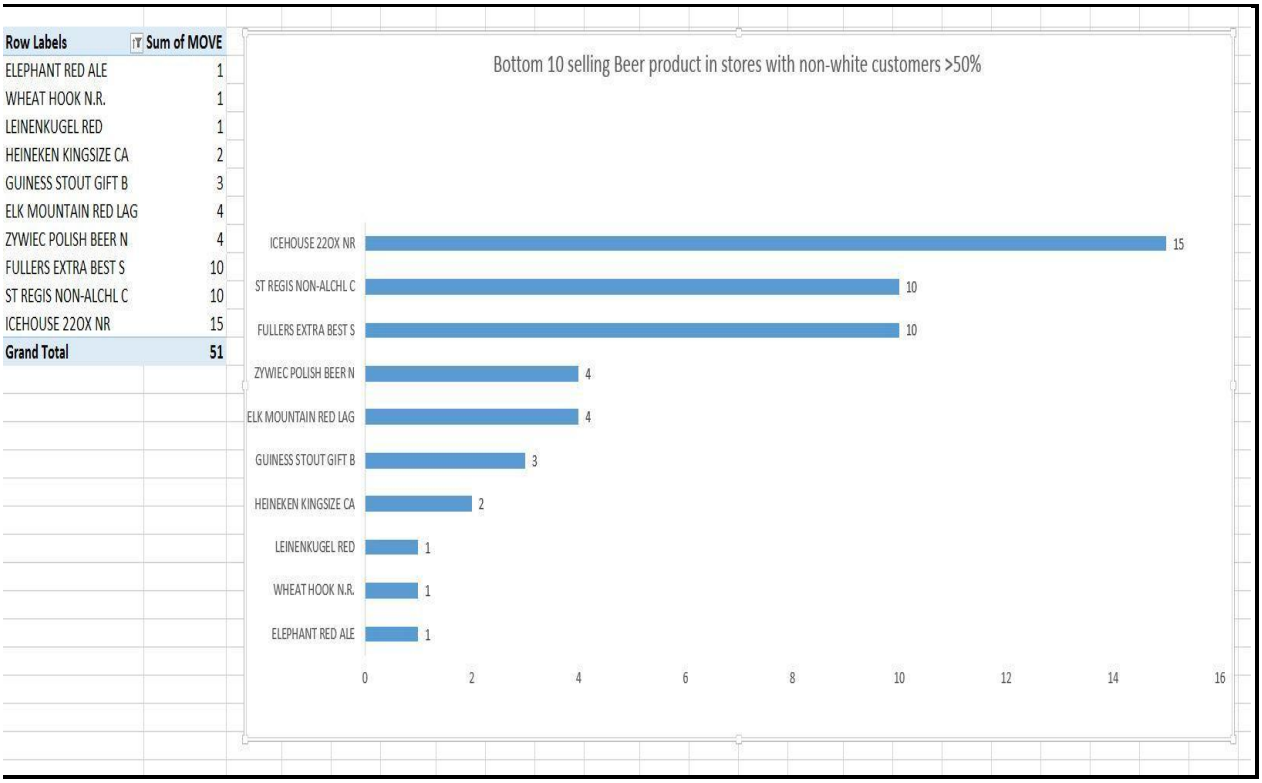


Sales trends of all beer products in areas with high non-white population



Row Labels	Sum of MOVE
BUDWEISER BEER	59078
MILLER GENUINE DRAFT	35852
MILLER HIGH LIFE N.R	24502
MILLER LITE BEER	22777
MILLER HIGH LIFE BEE	17920
MILLER GENUINE DRFT	17159
BUDWEISER BEER LONG	17029
MICHELOB REGULAR BEE	15371
BUDWEISER BEER N.R.B	15110
MILLER GEN DRFT LT L	13433
Grand Total	**238231**

Top 10 selling beers



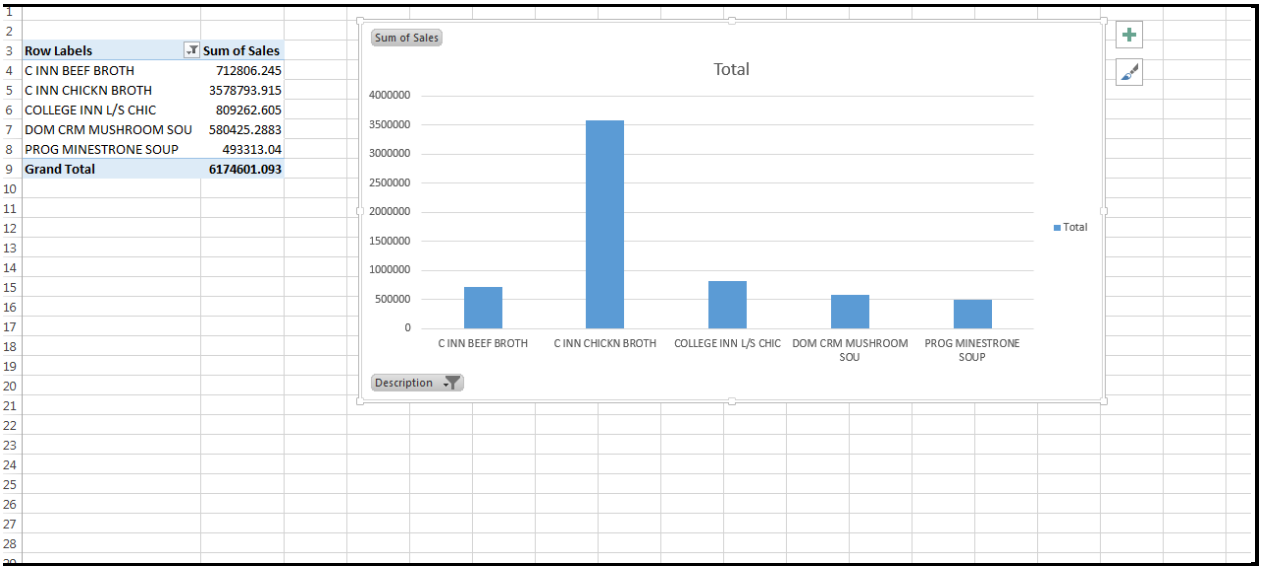
Row Labels	Sum of MOVE
ELEPHANT RED ALE	1
WHEAT HOOK N.R.	1
LEINENKUGEL RED	1
HEINEKEN KINGSIZE CA	2
GUINESS STOUT GIFT B	3
ELK MOUNTAIN RED LAG	4
ZYWIEC POLISH BEER N	4
FULLERS EXTRA BEST S	10
ST REGIS NON-ALCHL C	10
ICEHOUSE 22OX NR	15
Grand Total	**51**

Bottom 10 selling beers

From the above charts, we get the information about the selling trends of all the beer products sold in the DFF stores over the years. DFF can use this information to stock up their inventory intelligently by increasing the storage of the consistently popular beer products and cutting down on the consistently underselling products. They can sell the popular products in bulk with attractive discounts and also, employ promotional strategies to increase the sales of any particular product that is lagging currently.

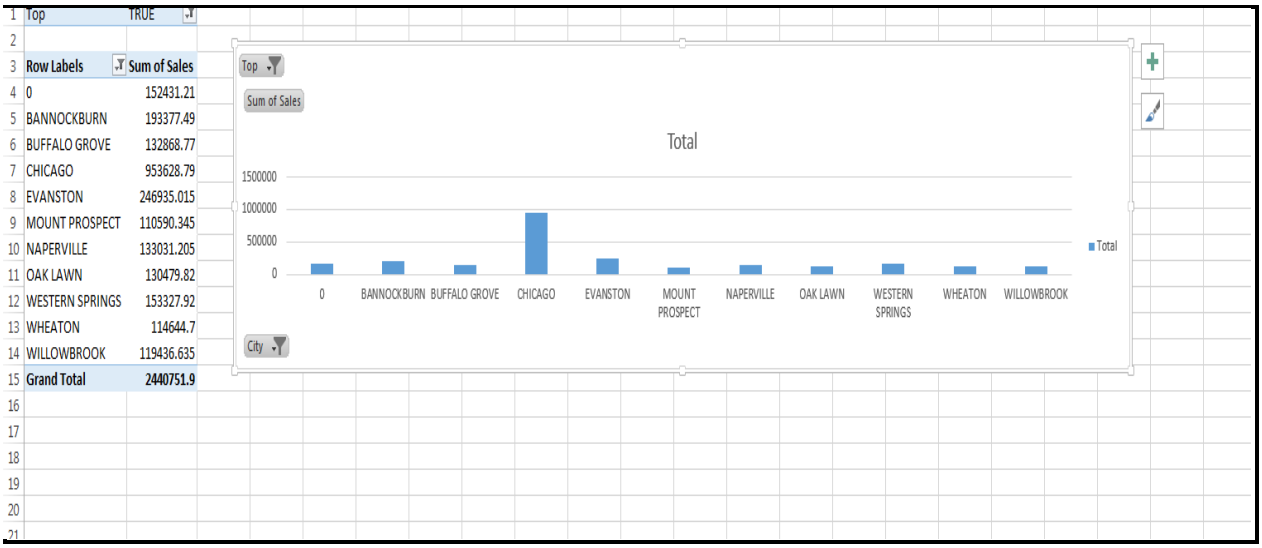
6. What are the top 5 sold commodities and which areas are they sold the most?

Rationalization: It is considered to be one of the key business interests to know which products in a retail company's offering have consistently been popular over the years. Perhaps, we could use that information to bundle some less known products with these and get the inventory moving. Having the information about the areas where they are sold will also help us understand the demographics of that area better and do targeted marketing in the best interest of the company.



Row Labels	Sum of Sales
C INN BEEF BROTH	712806.245
C INN CHICKN BROTH	3578793.915
COLLEGE INN L/S CHIC	809262.605
DOM CRM MUSHROOM SOU	580425.2883
PROG MINESTRONE SOUP	493313.04
Grand Total	**6174601.093**

Top 5 commodities by sales value sold across all stores of Dominic's Finer Food

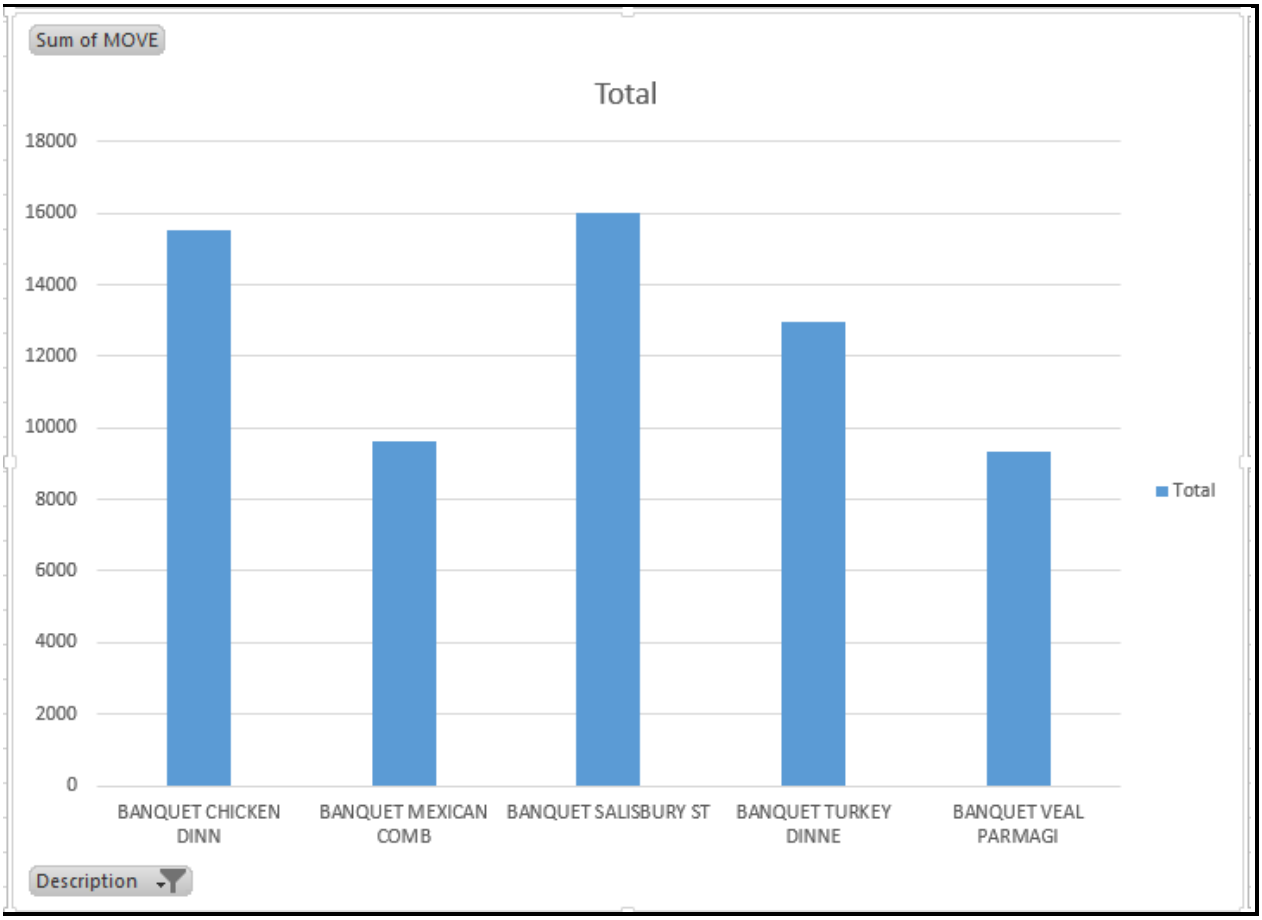


Top	TRUE
Row Labels	**Sum of Sales**
0	152431.21
BANNOCKBURN	193377.49
BUFFALO GROVE	132868.77
CHICAGO	953628.79
EVANSTON	246935.015
MOUNT PROSPECT	110590.345
NAPERVILLE	133031.205
OAK LAWN	130479.82
WESTERN SPRINGS	153327.92
WHEATON	114644.7
WILLOWBROOK	119436.635
Grand Total	**2440751.9**

Top 10 cities and the combined sales value of the top 5 commodities sold by Dominic's Finer Food

7. What are the top 5 frozen dinner products sold during each Thanksgiving week?

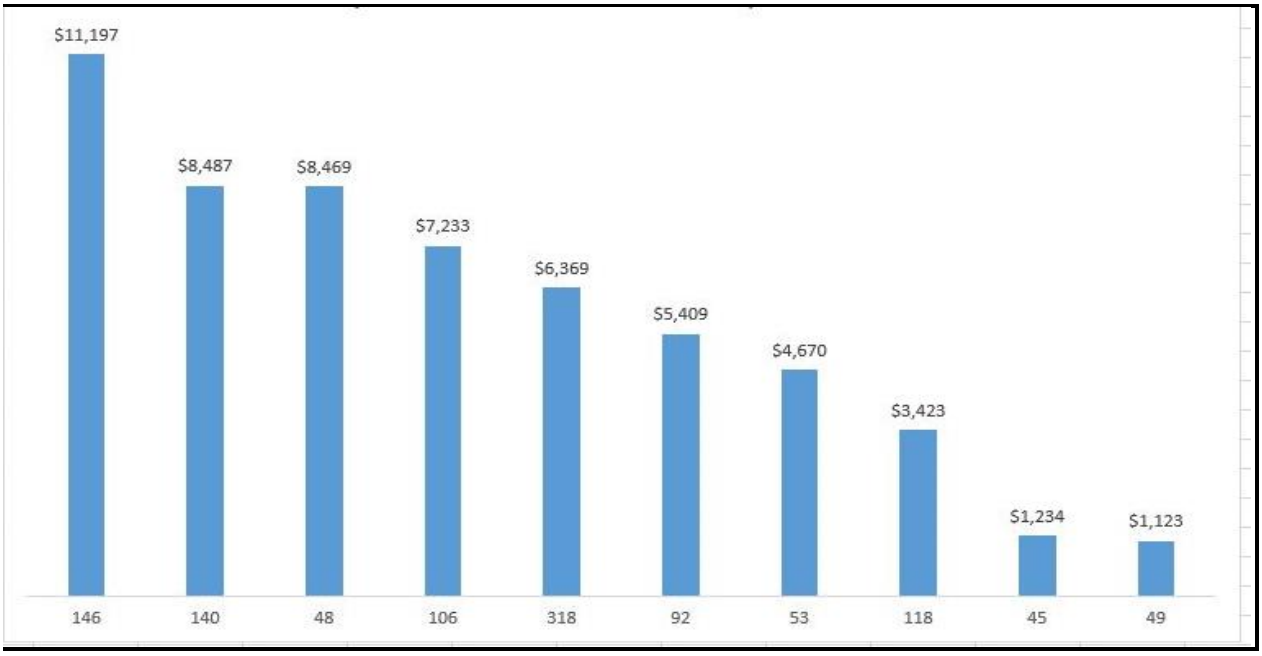
Rationalization: Thanksgiving is an important day in our lives where we get together with family and make some fond memories. An important part of thanksgiving is the food and we can expect to have a lot of increase in sales of meat during the period. It makes business sense to track the products that have been selling the most during thanksgiving, so we can estimate the amount we are expecting to sell in the future as well as be prepared to have the items in stock.



Top 5 units moved during the past five Thanksgiving weeks.

8. What are the bottom 10 stores with the least bakery sales during the years 95 - 97?

Rationalization: Bakery products have shorter lives resulting in loss of inventory in case products are not sold soon. Analyzing the bakery sales trends across the stores during the last three years (1995-97) provides DFF with a chance to reduce the inventory in the stores that have recorded consistently poor performance with respect to these products. Also, these existing unsold products in the underperforming stores can be clubbed with other product categories as an add-on which might help in improving the sales. The below figure shows the bottom 10 stores with the least bakery sales in the last three years.



Bottom 10 stores by jewelry sales.

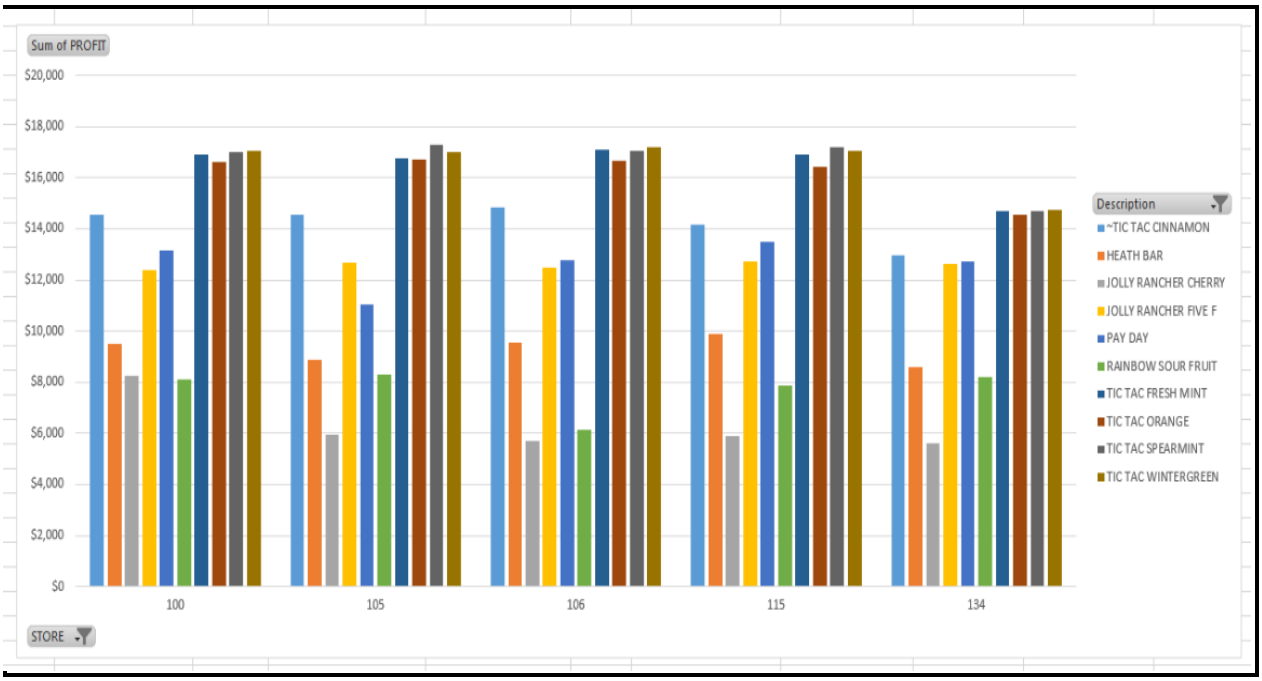
9. **Which are the stores in areas with high concentrations of children below the age of 9 that sell front end candies and which types of candies sell the most and which sells the least?**

We got a list of top 10 stores that have high customers below the age of 9.

Row Labels	Sum of AGE9
65	0.190055414
100	0.184360976
103	0.185393984
106	0.187403158
115	0.185493511
134	0.192817818
301	0.179988291
303	0.188213598
307	0.181203999
309	0.200155835
Grand Total	**1.875086582**

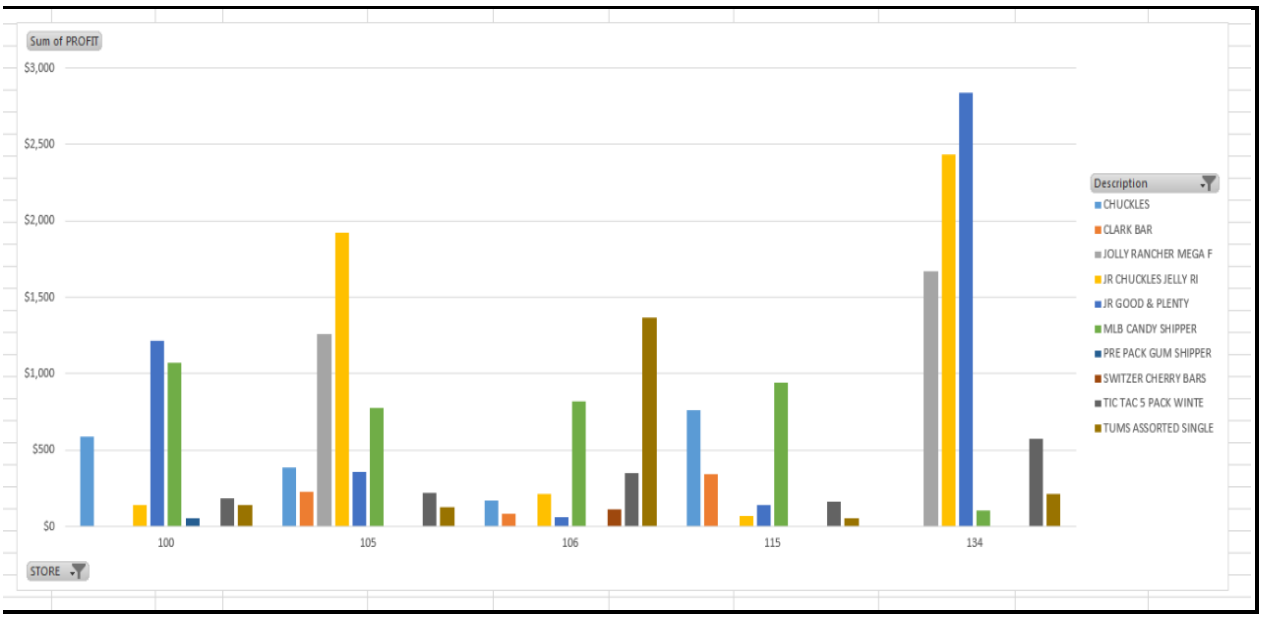
Top 10 stores in areas with high percentage of population under the age of 9

Rationalization: Front end candies are products that are more popular among children, specifically population under the age of 9. By advertising the right products and eliminating products that do not contribute to sales can be helpful to improve DFF profits and sales. Based on the information we have, we can find out which candies are popular among kids.



Top 10 selling candies

We can also consider the candies that are least popular.

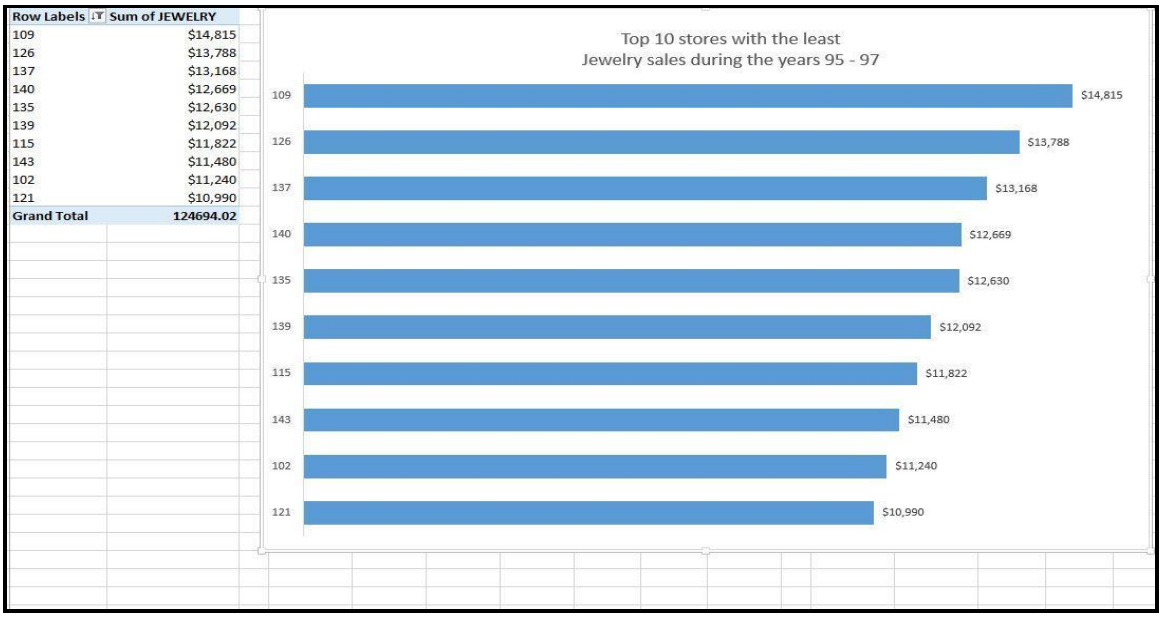


Bottom 10 selling candies

Based on this information, DFF can decide what candies to stock up more and which candies to eliminate.

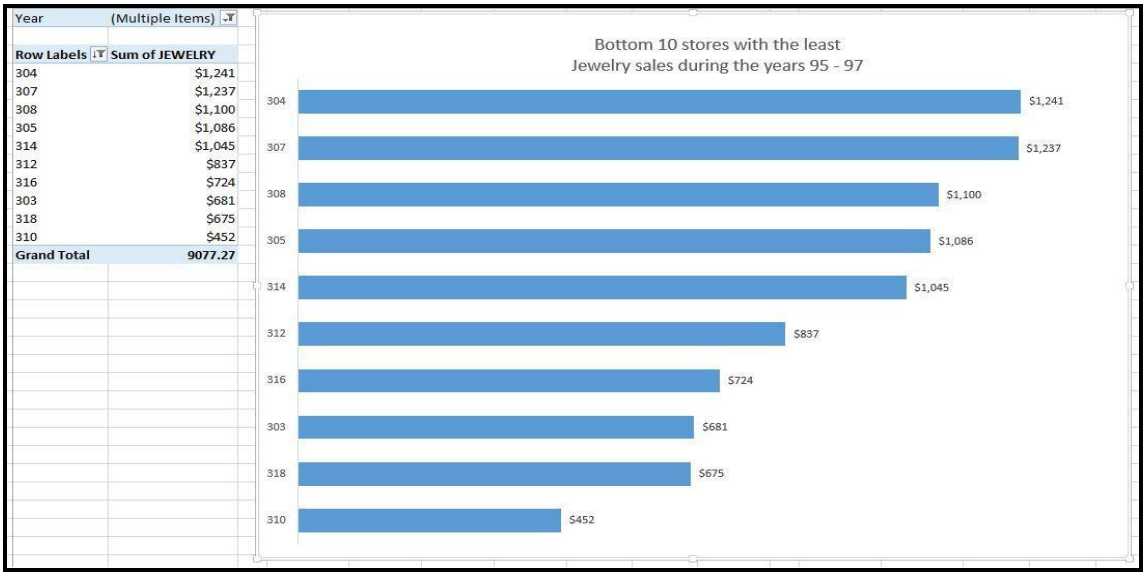
10. What are the top 10 stores with the most jewelry sales during the years 95 - 97? What are the bottom 10 stores with the least jewelry sales during the same period?

Rationalization: Jewelry sales by DFF are not very many in volume and it doesn't make sense to stock inventory in stores with less probability of sales. Not only does it increase the inventory costs, it also poses a certain amount of risk in case of theft. It makes business sense to track the top 10 stores with highest value of sales and focus the inventory and security aspects in those stores. It also makes sense to identify the bottom 10 stores and stop stocking inventory there altogether.



Row Labels	Sum of JEWELRY
109	$14,815
126	$13,788
137	$13,168
140	$12,669
135	$12,630
139	$12,092
115	$11,822
143	$11,480
102	$11,240
121	$10,990
Grand Total	124694.02

Top 10 stores by jewelry sales



Year	(Multiple Items)
Row Labels	Sum of JEWELRY
304	$1,241
307	$1,237
308	$1,100
305	$1,086
314	$1,045
312	$837
316	$724
303	$681
318	$675
310	$452
Grand Total	9077.27

Bottom 10 stores by jewelry sales.

6. Independent Data Marts design using Kimball's approach

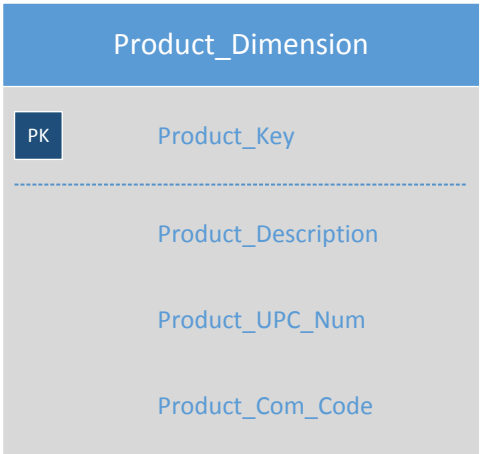
The STAR schemas to answer our business questions for Dominick's finer foods Inc. have been developed using the dimensional modeling technique, following Kimball's approach. The following sections provide the details regarding dimension tables, fact tables and the resulting Data Marts along with the Dimension Matrix. The justification as to how data mart design corresponds to the business questions is given in the next section.

a. Dimension Tables

There are a total of four dimensions, namely Product dimension, Store dimension, Demographics dimension and Time dimension. They are as follows

i. Product Dimension

The product dimension table contains the details of various products sold by Dominick's Finer Foods, pertaining to the business questions. It contains the description, UPC number, COM code for various products in addition to the surrogate product key.



PRODUCT_DIMENSION

Following is the description of the attributes of the product_dimension table –

- Product_Key – Surrogate key used to uniquely identify each of the rows
- Product_Description – Product name
- Product_UPC_Num – Number to identify the product and manufacturer
- Product_Com_Code – Code to identify the product category

ii. Store Dimension

The store dimension table contains the details of various stores operated by Dominick's Finer Foods, pertaining to the business questions. It contains the Store ID, Name, City, Zip code of the stores in addition to the surrogate store key.

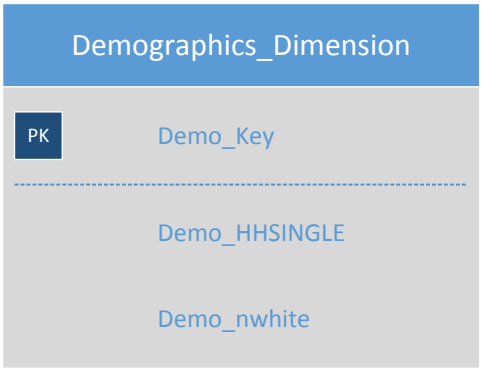
STORE_DIMENSION

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

- Store_Key – Surrogate key used to uniquely identify each of the rows
- Store_ID – Number to label the store
- Store_Name – Name of the store
- Store_City – City where the store is located
- Store_Zip – Postal zip code of the store

iii. Demographics Dimension

The Demographics dimension table contains the relevant details of various demographics that frequently visit the stores operated by Dominick's Finer Foods, pertaining to the business questions. It contains the percentage of households with one person and the percentage of nonwhite population corresponding to each store location in addition to the surrogate demo key.



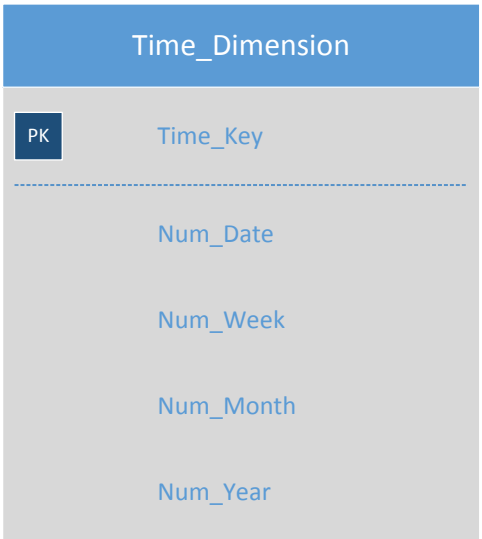
DEMOGRAPHICS_DIMENSION

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

- Demo_Key – Surrogate key used to uniquely identify each of the rows
- Demo_HHSINGLE – Percentage of Households with one person
- Demo_nwhite – Percentage of population that is non white

iv. Time Dimension

The Time dimension table contains all the time related measures for the stores operated by Dominick's Finer Foods, pertaining to the business questions. It contains the date, week, month and year by DFF in addition to the surrogate time key.



TIME_DIMENSION

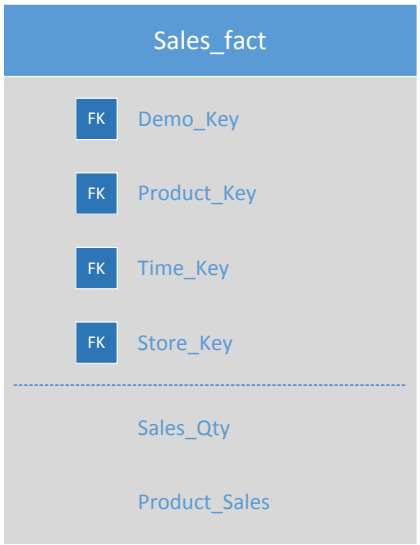
Following is the description of the attributes of the Time_Dimension dimension table –

- Time_Key – Surrogate key used to uniquely identify each of the rows
- Num_Date – Date of the observation
- Num_Week – Week of the observation
- Num_Month – Month of the observation
- Num_Year – Year of the observation

b. Fact Tables

i. Sales Fact Table

The sales fact table contains the details of various sales made by Dominick's Finer Foods, pertaining to the business questions. It contains the sales quantity and revenue from product sales for various products in addition to the foreign keys that refer to the surrogate keys of dimension tables, which are product_key, store_key, demo_key and time_key.



SALES_FACT

Following is the description of the attributes of the Sales fact table –

Fact Table Keys:

- Product_Key – Foreign key that refers to the surrogate key of Product dimension table
- Store_Key - Foreign key that refers to the surrogate key of Time dimension table
- Demo_Key - Foreign key that refers to the surrogate key of Demographics dimension table
- Time_Key - Foreign key that refers to the surrogate key of Time dimension table

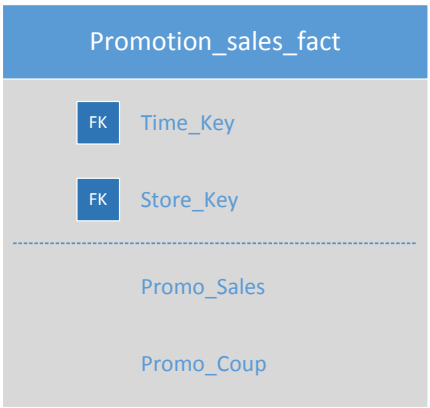
Fact Table Measures:

- Sales_Qty – Quantity of products sold

- Product_Sales – Revenue generated from product sales

ii. Promotion Sales Fact Table

The Promotion sales fact table contains the details of various promotional sales made by Dominick's Finer Foods, pertaining to the business questions. It contains the Dollar amount of promotional sales and the corresponding promotional coupons redeemed in addition to the foreign keys that refer to the surrogate keys of dimension tables, which are store_key and time_key.



PROMOTION_SALES_FACT

Following is the description of the attributes of the Promotion Sales fact table –

Fact Table Keys:

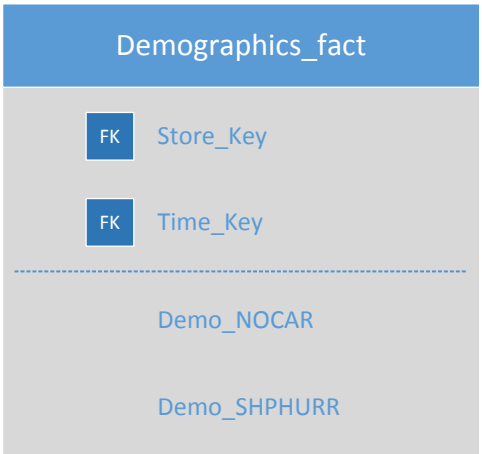
- Time_Key – Foreign key that refers to the surrogate key of Time dimension table
- Store_Key – Foreign key that refers to the surrogate key of Store dimension table

Fact Table Measures:

- Promo_Sales – Promotional sales in dollars
- Promo_Coup – Promotional coupons redeemed

iii. Demographics Fact Table

The Demographics fact table contains the demographic facts of the population that visit Dominick's Finer Foods, pertaining to the business questions. It contains the data about the percentage of customers with no vehicles and the percentage of shoppers that shop in a hurry, in addition to the foreign keys that refer to the surrogate keys of dimension tables, which are store_key and time_key.



DEMOGRAPHICS_FACT

Fact Table Keys:

- Time_Key – Foreign key that refers to the surrogate key of Time dimension table
- Store_Key – Foreign key that refers to the surrogate key of Store dimension table

Fact Table Measures:

- Demo_NOCAR – Percentage with no vehicles
- Demo_SHPHURR – Percentage of hurried shoppers

c. Define Dimension Matrix for data marts

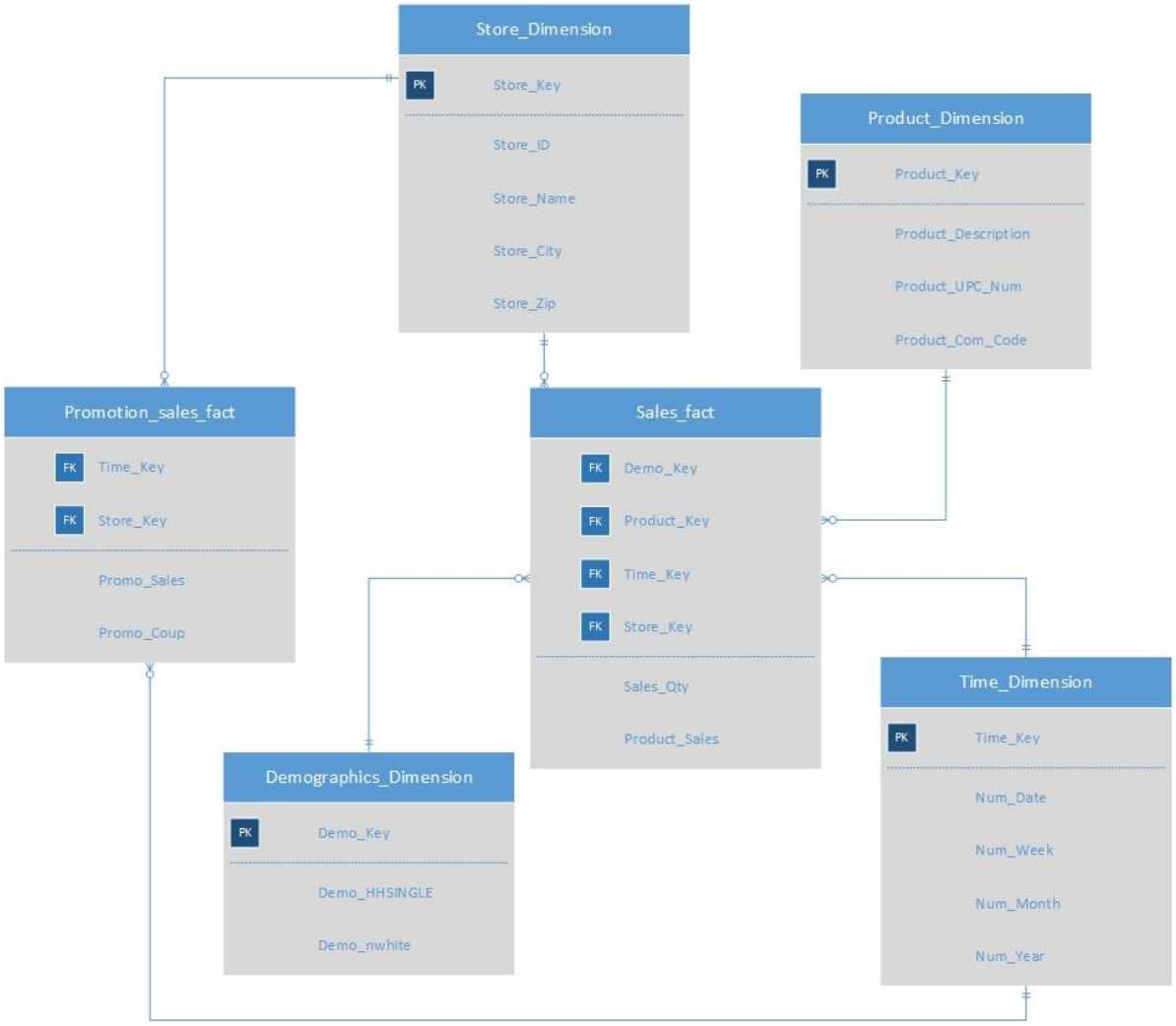
Dimensions / Facts	Store	Product	Time	Demographic
Sales_fact	X	X	X	X
Promo_sales_fact	X		X	
Demographic_fact	X		X	

d. Data Marts using STAR Schema

The business questions identified for Dominick's Finer Foods Inc. can be answered using the integrated STAR schemas developed using the above dimension and fact tables. These schemas make up the Marketing data mart and Strategy data mart.

i. Marketing STAR Schema

The marketing STAR schema comprises of the Store dimension, Product dimension, Time dimension, Demographics dimension, Promotion Sales fact and Sales fact tables. This data mart aims to answer the business questions posed by the marketing team.



MARKETING STAR SCHEMA

In the context of this report, the questions answered are as follows:

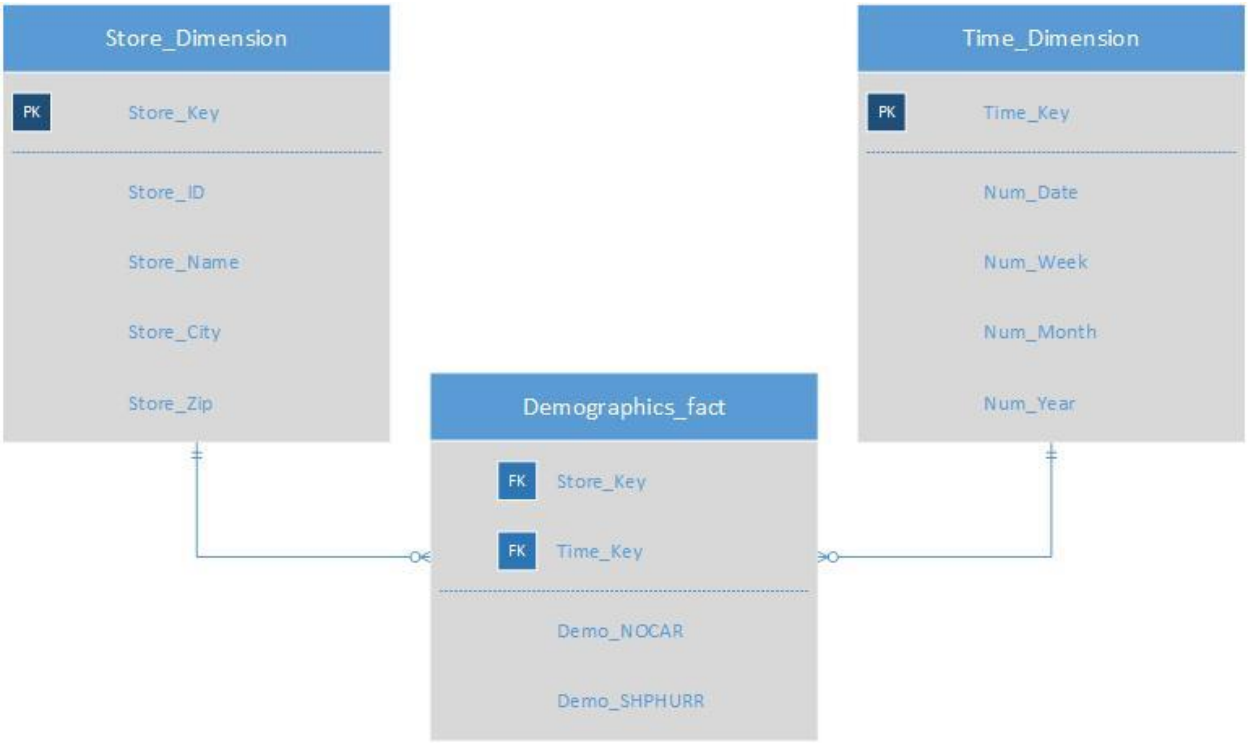
- How have the promo sales been affected across the stores on the basis of coupons?
- Which stores are in areas of high single households and what are the commodities/products

that sell more and which sell least?

- How is the sales performance of all the beer brands across the stores in areas with more than 50% non-white population in the last 6 years (1991 - 1997)?

ii. Strategy STAR Schema

The Strategy STAR schema comprises of the Store dimension, Time dimension and Demographics fact tables. This data mart aims to answer the business questions posed by the Strategy team.



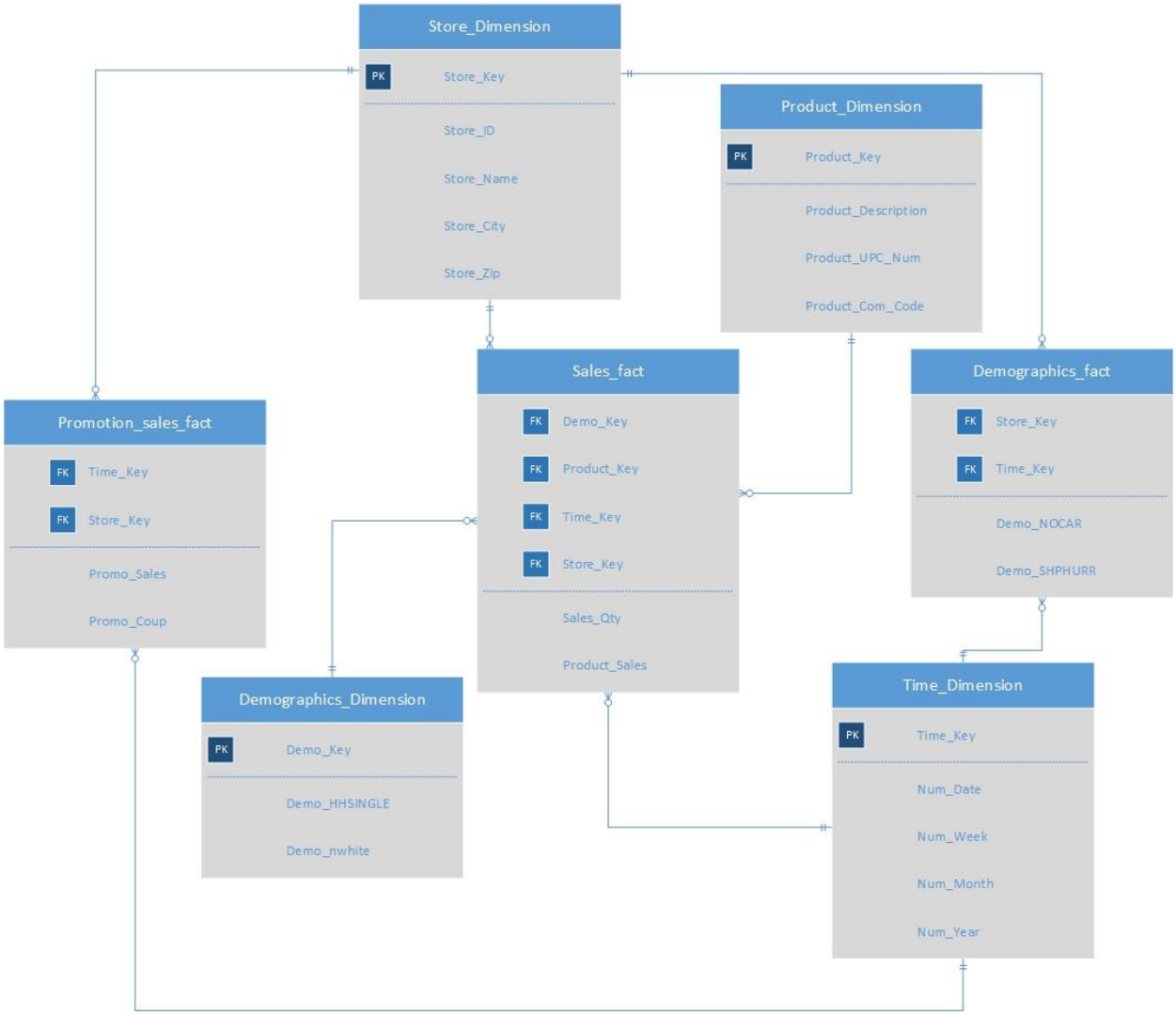
STRATEGY STAR SCHEMA

In the context of this report, the questions answered are as follows:

- Which stores or store areas have a high percentage of customers with no cars in order to set up a home delivery service?
- Which stores have a large number of hurried shoppers and would benefit from setting up a curbside pickup?

iii. Overall Data Warehouse Schema

The above data marts collectively comprise the overall data warehouse. It illustrates the various fact tables, dimension tables along with the conformed dimensions and the relations between them.



DATA WAREHOUSE SCHEMA

7. Mapping Tables

a. Mapping tables for Marketing Data Mart

DW Dimension Table	DW Dimension Attribute	Source Table	Source Table Attribute	Mapping Functions
Product		UPC table		
	Product_Key (Row number)			Create
	Product_Description		Description	Copy
	Product_UPC_Num		UPC	
	Product_Com_Code		Com_code	

DW Dimension Table	DW Dimension Attribute	Source Table	Source Table Attribute	Mapping Functions
Demographics		Demo		
	Demo_Key (Row number)			Create
	Demo_HHSINGLE		Hhsingle	Copy
	Demo_nwhite		Nwhite	

DW Dimension Table	DW Dimension Attribute	Source Table	Source Table Attribute	Mapping Functions
Store		Demo		
	Store_Key (Row number)			Create
	Store_ID		Store	Copy
	Store_Name		Name	
	Store_City		City	
	Store_Zip		Zip	

DW Dimension Table	DW Dimension Attribute	Source Table	Source Table Attribute	Mapping Functions
Time		CCount		
	Time_Key (Row number)			
	Num_Date		Date	Date of starting day of week
	Num_Week		Week	
	Num_Month		Date	Month of starting day of week
	Num_Year		Date	Year of starting day of week

DW Dimension Table	DW Dimension Attribute	Source Table	Source Table Attribute	Mapping Functions
Promotion_Sales_Fact		CCount		
	Time_Key (Row number)			Primary Key of Time dimension
	Store_Key			Primary Key of Store dimension
	Promo_Sales		Promo	Sum of Promo aggregated across a week
	Promo_Coup		Promo_coup	Sum of Promo_coup aggregated across a week

DW Dimension Table	DW Dimension Attribute	Source Table	Source Table Attribute	Mapping Functions
Sales_Fact		Movement		
	Demo_Key			Primary Key of Demo Dimension
	Product_Key			Primary Key of Product Dimension
	Time_Key			Primary Key of the Time Dimension
	Store_Key			Primary key of the Store Dimension
	Sales_Qty		Move	
	Product_Sales		Sale	

b. Mapping tables for Strategy Data Mart

DW Dimension Table	DW Dimension Attribute	Source Table	Source Table Attribute	Mapping Functions
Store		Demo		
	Store_Key (Row number)			Create
	Store_ID		Store	Copy
	Store_Name		Name	
	Store_City		City	
	Store_Zip		Zip	

DW Dimension Table	DW Dimension Attribute	Source Table	Source Table Attribute	Mapping Functions
Time		WeekDecode		
	Time_Key (Row number)			
	Num_Date		Date	Date of starting day of week
	Num_Week		Week	
	Num_Month		Date	Month of starting day of week
	Num_Year		Date	Year of starting day of week

DW Dimension Table	DW Dimension Attribute	Source Table	Source Table Attribute	Mapping Functions
Demographics _Fact		Demo		
	Store_Key			Primary Key of Store Dimension
	Time_Key			Primary Key of Time Dimension
	Demo_NOCAR		Nocar	
	Demo_SHPHURR		Shophurr	

8. Justification of business questions corresponding to data marts

a. **How have the promo sales been affected across the stores on the basis of coupons?**
This business question is answered by the Marketing data mart. The store details such as Store ID are taken from the store dimension, the time attributes spanning the desired duration for analysis are taken from the time dimension, which together enable us to fetch appropriate metrics from the Promotion Sales fact table, such as promotional sales and coupons, which can be used to answer the business question.

b. **Which stores are in areas of high single households and what are the commodities/products that sell more and which sell least?**
This business question is answered by the Marketing data mart. The store details such as Store ID are taken from the store dimension, the time attributes spanning the desired duration for analysis are taken from the time dimension, the demographics details such as the distribution of single households from the demographics dimension, the product details such as the product name, category which together enable us to fetch appropriate metrics from the Sales fact table, such as the quantity of sales and products sold, which can be used to answer the business question.

c. **Which stores or store areas have a high percentage of customers with no cars in order to set up a home delivery service?**
This business question is answered by Strategy data mart. The store details such as Store ID, city are taken from the store dimension, the time attributes spanning the desired duration for analysis are taken from the time dimension, which together enable us to fetch appropriate metrics from the Demographics fact table, such as the percentage of customers with no access to vehicles, to answer the business question.

d. **Which stores have a large number of hurried shoppers and would benefit from setting up a curbside pickup?**
This business question is answered by Strategy data mart. The store details such as Store ID are taken from the store dimension, the time attributes spanning the desired duration for analysis are taken from the time dimension, which together enable us to fetch appropriate metrics from the Demographics fact table, such as the percentage of hurried shoppers, to answer the business question.

e. **How is the sales performance of all the beer brands across the stores in areas with more than 50% non-white population in the last 6 years (1991 - 1997)?**
This business question is answered by the Marketing data mart. The store details such as Store ID are taken from the store dimension, the time attributes spanning the desired duration for analysis are taken from the time dimension, the demographics details such as the distribution of nonwhite population from the demographics dimension, the product details such as the product name, manufacturer which together enable us to fetch appropriate metrics from the Sales fact table, such as the quantity of sales and products sold, which can be used to answer the business question.

9. Development of ETL Plan

ETL is an acronym used for extract, transform and load. These three operations are crucial database functions that are together used to pull data from a database containing dirty, unorganized data, clean the extracted data and place it into another database in a much more convenient format. Extract involves reading data from a database. Transform involves converting the extracted data into a more usable form. Load operation involves placing the data into the target database.

The subsequent steps demonstrate the construction of a data warehouse for Dominick's Finer Foods Inc. to answer the business questions established in the previous section.

a. Determine all the target data needed in the data warehouse

Data Source	**Data Extraction**	**Data Transformation**	**Data in Data Warehouse**
UPC Files	UPC	UPC	Product_Dimension
Movement Files	Movement	Movement	Sales_Fact
DEMO.csv	Demographics	Demographics	Demo_Dimension, Store_Dimension, Demographics_Fact
CCOUNT.csv	CCOUNT	CCOUNT, CCOUNT_Agg	Promotion_Sales_Fact
WEEK_DECODE.csv	Week_Decode	Week_Decode_Final	Time_Dimension

b. Determining all the Data Sources

All the data used in building the data warehouse for Dominick's Finer Foods Inc. is obtained from Dominick's database and manual made available by the University of Chicago Booth School of Business. The following files are required to answer our business questions.

DATA SOURCE	**SOURCE FILES**
Product Details	UPCXXX (XXX is a product specific abbreviation)
Product Movement Details	WXXX (XXX is a product specific abbreviation)
Demographic Information	DEMO
Week Information	WEEK_DECODE
Customer Data	CCOUNT

c. Data Mapping from Data Sources to Staging Area

Source File Name	File Attributes	Staging Area Table Name	Staging Area Attributes
Week_Decode.csv	Week Number	WeekDecode	Week #
	Start Date		Start
	End Date		End
	Special Events		Special Events
CCOUNT.csv	Store	CCOUNT	Store_ID
	Promo		Promo_Sales
	Promo_Coup		Promo_Coup
	Date		Num_Date
	Week		Num_Week
	Store	CCOUNT_Agg	Store_ID
	Sum of (Promo)		Sum_Promo_Sales
	Sum of (Promo_Coup)		Sum_Promo_Coup
	Week		Num_Week
MOVEMENT Files	STORE	Movement	STORE_ID
	UPC		UPC_Number
	WEEK		Num_Week
	QTY		Sales_Qty
	SALE		Product_Sales
	MOVE		Product_Movement
	PRICE		Product_Price
	OK		OK
DEMO.csv	STORE	Demographics	STORE_ID
	NAME		STORE_Name
	CITY		STORE_City
	ZIP		STORE_Zip
	HHSINGLE		Demo_HHSINGLE
	NWHITE		Demo_NWHITE
	NOCAR		Demo_NOCAR
	SHPHURR		Demo_SHPHURR
UPC Files	UPC	UPC	Product_UPC_Num
	DESCRIPTION		Product_Description
	COM_CODE		Product_Com_Code

d. Data Mapping from Staging Area to Data Warehouse

Staging Area Table Name	Staging Area Attributes	Data Warehouse Table Name	Data Warehouse Attributes
UPC	Product_UPC_Num	Product_Dimension	Product_UPC_Num
	Product_Description		Product_Description
	Product_Com_Code		Product_Com_Code
WeekDecode	Week #	Time_Dimension	Num_Week
	Start		Num_Date , Num_Month, Num_Year
	End		
	Special Events		
Movement	STORE_ID	Sales_Fact	
	UPC_Number		
	Num_Week		
	Sales_Qty		Sale_Qty
	Product_Sales		Product_Sales
	Product_Movement		
	Product_Price		
	OK		
Demographics	STORE_ID	Store_Dimension	STORE_ID
	STORE_Name		STORE_Name
	STORE_City		STORE_City
	STORE_Zip		STORE_Zip
	STORE_ID	Demo_Dimension	STORE_ID
	Demo_HHSINGLE		Demo_HHSINGLE
	Demo_NWHITE		Demo_NWHITE
	Demo_NOCAR	Demographics_Fact	Demo_NOCAR
	Demo_SHPHURR		Demo_SHPHURR
CCOUNT_Agg	Store_ID	Promotional_Sales_Fact	
	Sum_Promo_Sales		Promo_Sales
	Sum_Promo_Coup		Promo_Coup
	Num_Week		

e. Establishing comprehensive data extraction rules

Data extraction from the source files is the initial step in the ETL process of designing and implementing a data warehouse. Data extraction has to be carried out with a lot of care as the amount of data can get overwhelming if performed without a plan of action. The data is extracted to be in the same format (Tables for MS SQL Server) that can be later used for transform and load operations. The data sources can be in a variety of formats but in this case, all our sources are in the Comma Separated Value (.csv) format.

f. Determining data transformation and cleansing rules

Data transformation and cleaning process is the next step after Data Extraction process. The general rules that apply to transformation and cleaning of the data are as follows:

i. **Removal of NULL values**: Null values that were existing, in addition to empty rows, in the data that affect referential integrity and those that were created while extracting data into tables will be deleted.
ii. **Removal of Garbage values:** Some of the values do not mean anything at all. Such attributes as well as attributes that are not part of the answers to the business questions will be removed. Blank records, rows with just a '.' (Dot) and other non-meaningful values such as negatives in week or sales will be removed. The records where the value for the sales of various departments are '.' will be removed. All these values are being removed as they are not relevant to answer our business questions.
iii. **Data conversion:** The data which was extracted from the source files is initially stored as nvarchar (string) in the staging area. These will be converted to appropriate data types such as int, float, date etc.
iv. **Creation of surrogate keys:** Surrogate keys have been added for all the dimension tables and fact tables to aid with associations before the loading operations.
v. **Derived attributes:** Derived attributes exist in two dimensions, described as below:
    a. Year and month in the Time dimension, obtained from the functions YEAR (date) and MONTH (date) on the respective columns of the WeekDecode table.
    b. The 'Product_sales' measure in the fact table 'Sales_Fact' is obtained by the formula 'Product_sales = (Unit_price * Number_of_units_sold) / Quantity'.
    c. The 'Promo_Sales' measure in the fact table 'Promotional_Sales_Fact' is obtained by aggregating sales for a particular week by grouping the sales based on week number.

SSIS Functions

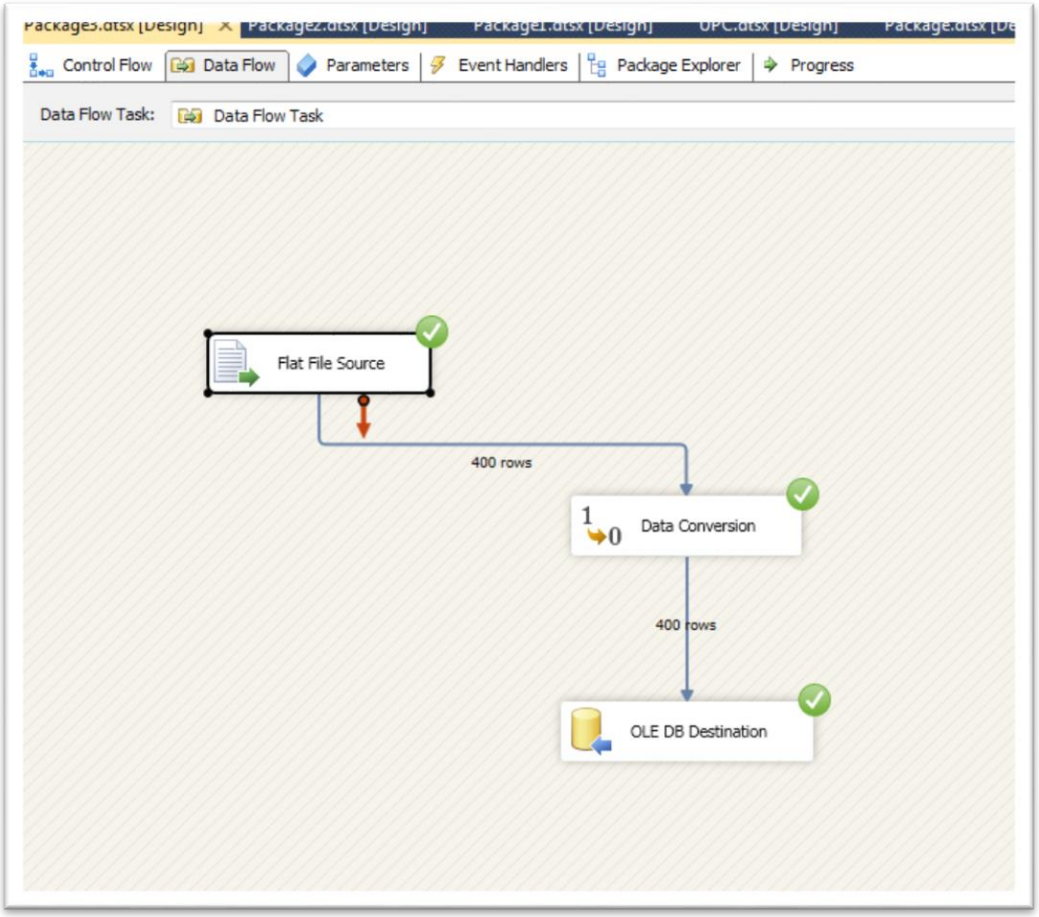
The following functions were used during during the ETL process:

- LOOKUP: Joins additional columns to the data flow by looking up values in a table
- DATA CONVERSION: Converts data from one data type to another
- DERIVED COLUMNS: Creates new column values by applying expressions to input columns
- EXECUTE SQL: Executes SQL queries in the control flow
- FOR EACH LOOP CONTAINER: Executes an operation for all files in a directory

10. Extraction and Transformation Implementation

The ETL process for the data warehouse being created was implemented using SQL Server Integration Services (SSIS) and SQL Server Management Studio. The following screenshots give a detailed Idea of the steps involved in the process.

a. WeekDecode Extraction and Transformation



EXTRACTING WEEKDECODE SOURCE FILE INTO THE STAGING AREA

SQLQuery41.sql - in...aster (so7339 (133)) | SQLQuery40.sql - in...aster (so7339

```sql
/****** Script for SelectTopNRows command from SSMS  ***
SELECT TOP 1000 [Week #]
      ,[Start]
      ,[End]
      ,[Special Events]
  FROM [602group5-staging-area].[dbo].[WeekDecode]
```

100 %

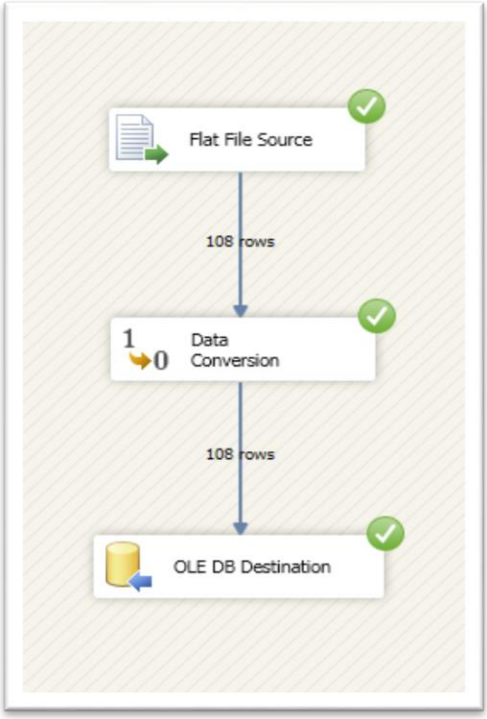
Results | Messages

	Week #	Start	End	Special Events
1	1	1989-09-14	Sep 20 1989 12:00AM	
2	2	1989-09-21	Sep 27 1989 12:00AM	
3	3	1989-09-28	Oct 4 1989 12:00AM	
4	4	1989-10-05	Oct 11 1989 12:00AM	
5	5	1989-10-12	Oct 18 1989 12:00AM	
6	6	1989-10-19	Oct 25 1989 12:00AM	
7	7	1989-10-26	Nov 1 1989 12:00AM	Halloween
8	8	1989-11-02	Nov 8 1989 12:00AM	
9	9	1989-11-09	Nov 15 1989 12:00AM	
10	10	1989-11-16	Nov 22 1989 12:00AM	
11	11	1989-11-23	Nov 29 1989 12:00AM	Thanksgiving
12	12	1989-11-30	Dec 6 1989 12:00AM	
13	13	1989-12-07	Dec 13 1989 12:00AM	
14	14	1989-12-14	Dec 20 1989 12:00AM	
15	15	1989-12-21	Dec 27 1989 12:00AM	Christmas
16	16	1989-12-28	Jan 3 1990 12:00AM	New-Year
17	17	1990-01-04	Jan 10 1990 12:00AM	
18	18	1990-01-11	Jan 17 1990 12:00AM	
19	19	1990-01-18	Jan 24 1990 12:00AM	
20	20	1990-01-25	Jan 31 1990 12:00AM	
21	21	1990-02-01	Feb 7 1990 12:00AM	
22	22	1990-02-08	Feb 14 1990 12:00AM	
23	23	1990-02-15	Feb 21 1990 12:00AM	Presidents Day
24	24	1990-02-22	Feb 28 1990 12:00AM	
25	25	1990-03-01	Mar 7 1990 12:00AM	
26	26	1990-03-08	Mar 14 1990 12:00AM	
27	27	1990-03-15	Mar 21 1990 12:00AM	
28	28	1990-03-22	Mar 28 1990 12:00AM	Easter
29	29	1990-03-29	Apr 4 1990 12:00AM	
30	30	1990-04-05	Apr 11 1990 12:00AM	
31	31	1990-04-12	Apr 18 1990 12:00AM	
32	32	1990-04-19	Apr 25 1990 12:00AM	
33	33	1990-04-26	May 2 1990 12:00AM	
34	34	1990-05-03	May 9 1990 12:00AM	
35	35	1990-05-10	May 16 1990 12:00AM	
36	36	1990-05-17	May 23 1990 12:00AM	

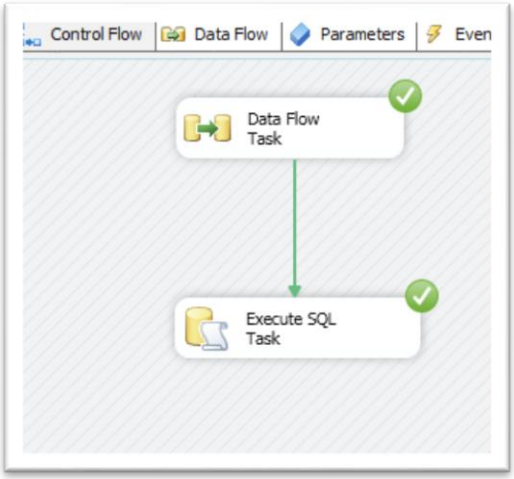
Query executed successfully.

CLEANED WEEKDECODE TABLE IN STAGING AREA

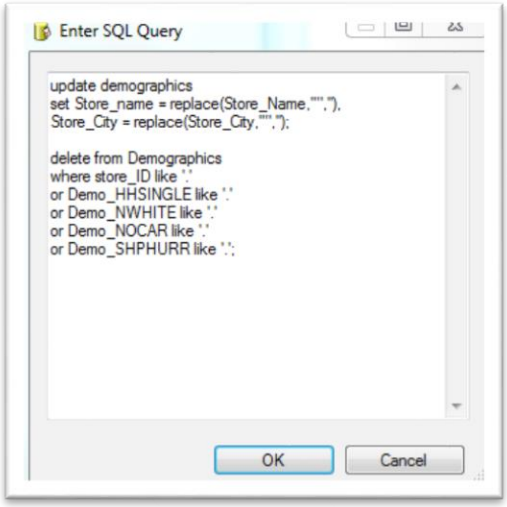
b. Demographics Extraction and Transformation



EXTRACTING DEMOGRAPHICS DATA TO THE STAGING AREA



SQL TASK TO CLEAN THE DEMOGRAPHICS STAGING AREA



SQL QUERY TO CLEAN DEMOGRAPHICS STAGING AREA

SQLQuery3.sql - inf...-area (so7339 (181))* SQLQuery2.sql - inf...aster (so7339 (214)) SQLQuery1.sql - inf...-area (so7339 (188))*

```
select * from Demographics
```

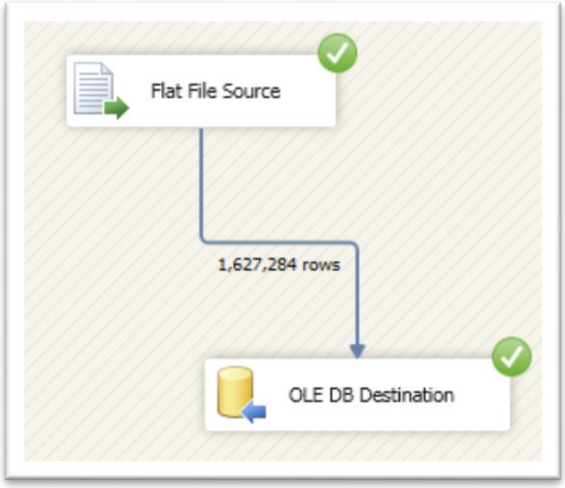
100 %

Results | Messages

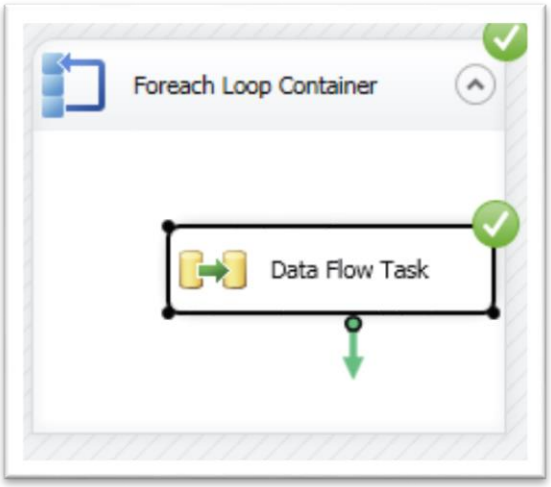
	Store_ID	Store_Name	Store_City	Store_Zip	Demo_HHSINGLE	Demo_NWHITE	Demo_NOCAR	Demo_SHPHURR
1	2	DOMINICKS 2	RIVER FOREST	60305	0.2820331804	0.1382256003	0.1246028945	0.1210730674
2	4	DOMINICKS 4	PARK RIDGE	60068	0.2694420964	0.1516669404	0.0555672935	0.1183181914
3	5	DOMINICKS 5	PALATINE	60067	0.2188516969	0.0911701989	0.0255695026	0.1906090191
4	8	DOMINICKS 8	OAK LAWN	60453	0.2108215111	0.0459968603	0.0751127241	0.1351737791
5	9	DOMINICKS 9	MORTON GROVE	60053	0.2115440535	0.1299819089	0.0401279442	0.1586216924
6	12	DOMINICKS 12	CHICAGO	60660	0.4925948468	0.4953317689	0.4835175981	0.0459478649
7	14	DOMINICKS 14	GLENVIEW	60025	0.1863322673	0.1203502415	0.0265858994	0.215415685
8	18	DOMINICKS 18	RIVER GROVE	60171	0.2688054199	0.08956747	0.1419746936	0.1023112174
9	21	DOMINICKS 21	HANOVER PARK	60103	0.1388341312	0.1633634178	0.0175981979	0.226978316
10	28	DOMINICKS 28	MOUNT PROSPECT	60056	0.2106000467	0.0965281431	0.0548552754	0.165810369
11	32	DOMINICKS 32	PARK RIDGE	60068	0.2903059648	0.0526046114	0.0717008344	0.1147743664
12	33	DOMINICKS 33	CHICAGO	60657	0.6140095339	0.1704708161	0.5062235169	0.0263506356
13	40	DOMINICKS 40	BRIDGEVIEW	60455	0.2293672186	0.056533999	0.0463295688	0.130373412
14	44	DOMINICKS 44	WESTERN SPRINGS	60558	0.173053404	0.0534809669	0.0407664085	0.2046056654
15	45	DOMINICKS 45	WHEELING	60090	0.2625675946	0.1296547325	0.0202323718	0.1574519231
16	47	DOMINICKS 47	ADDISON	60101	0.1756202383	0.1659792443	0.0212973818	0.1686205549
17	48	DOMINICKS 48	SCHAUMBURG	60193	0.3249071946	0.1744655404	0.0212089077	0.129397207
18	49	DOMINICKS 49	DOWNERS GROVE	60515	0.255505618	0.0991848992	0.0543820225	0.1719101124
19	50	DOMINICKS 50	HICKORY HILLS	60457	0.2334478809	0.0899879787	0.0364344638	0.1264604811
20	51	DOMINICKS 51	PALOS HEIGHTS	60463	0.1996674979	0.0425719319	0.025436409	0.1429758936
21	52	DOMINICKS 52	NORTHBROOK	60062	0.2059581782	0.1375667022	0.0148997135	0.2107449857
22	53	DOMINICKS 53	CHICAGO	60662	0.2793791574	0.2030500077	0.1453629032	0.1429435484
23	54	DOMINICKS 54	NAPERVILLE	60540	0.2672399537	0.1008991009	0.0208402249	0.1914999173
24	56	DOMINICKS 56	COUNTRYSIDE	60525	0.2280229033	0.0763397693	0.0319973055	0.1663859885
25	59	DOMINICKS 59	CRYSTAL LAKE	60014	0.1542296308	0.0357650029	0.0313181677	0.2375389408
26	62	DOMINICKS 62	NORTHFIELD	60093	0.1826497942	0.0789710177	0.0265284342	0.2404329928
27	64	DOMINICKS 64	VILLA PARK	60181	0.187960688	0.1075805073	0.043620725	0.17161581
28	65	DOMINICKS 65	HANOVER PARK	60103	0.1415006948	0.1713344316	0.0152848541	0.2283464567
29	67	DOMINICKS 67	OAKBROOK TERR...	60521	0.2549435028	0.1170322492	0.0450291365	0.1608403955
30	68	DOMINICKS 68	CHICAGO	60625	0.3160502566	0.3792154164	0.305255707	0.0897027601
31	70	DOMINICKS 70	JOLIET	60435	0.2578110725	0.1751592357	0.0811255253	0.1284370147
32	71	DOMINICKS 71	NORTH RIVERSIDE	60546	0.3384953322	0.1007916313	0.1534321801	0.0977592267
33	72	DOMINICKS 72	LINCOLNWOOD	60646	0.211928934	0.204995157	0.0676214648	0.1539437897
34	73	DOMINICKS 73	CHICAGO	60629	0.2289969996	0.1195237725	0.133319044	0.1343907405
35	74	DOMINICKS 74	NORRIDGE	60634	0.2654993383	0.0615682166	0.1436452799	0.1007941356
36	75	DOMINICKS 75	CHICAGO	60640	0.5228839261	0.5623832399	0.5506547209	0.0368022054
37	76	DOMINICKS 76	CHICAGO	60618	0.2697022767	0.4636889332	0.3479640981	0.0862521891
38	77	DOMINICKS 77	VERNON HILLS	60061	0.1803126451	0.1164416978	0.0168678428	0.2510058805
39	78	DOMINICKS 78	DOWNERS GROVE	60516	0.2026257122	0.1356198162	0.013869969	0.2203368838

CLEANED DEMOGRAPHICS TABLE IN STAGING AREA

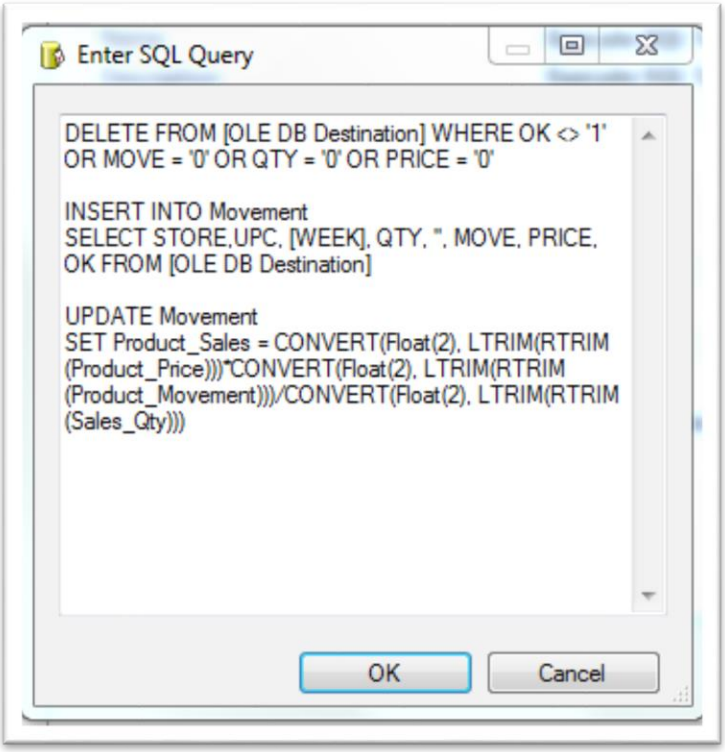
c. Movement Extraction and Transformation



EXTRACTING MOVEMENT DATA TO THE STAGING AREA



SQL TASK TO CLEAN THE MOVEMENT STAGING AREA



SQL QUERY TO CLEAN MOVEMENT STAGING AREA

SQLQuery4.sql - inf...-area (so7339 (324))* | SQLQuery3.sql - inf...-area (so7339 (198))* | SQLQuery2.sql - inf...-area (so7339 (201))* | SQLQ

```
SELECT TOP 100 * FROM MOVEMENT
```

100 %

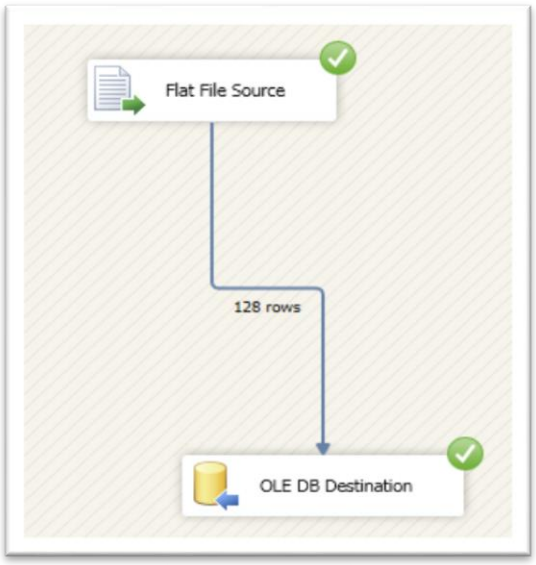
Results | Messages

	Store_ID	UPC_Number	Num_Week	Sales_Qty	Product_Sales	Product_Movement	Product_Price	OK
1	123	2080000054	45	1	17.49	11	1.59	1
2	123	2080000054	46	1	7.95	5	1.59	1
3	123	2080000054	47	1	14.31	9	1.59	1
4	123	2080000054	48	1	33.06	19	1.74	1
5	123	2080000054	49	1	39.75	25	1.59	1
6	123	2080000054	51	1	4.77	3	1.59	1
7	123	2080000054	52	1	5.22	3	1.74	1
8	123	2080000054	53	1	13.92	8	1.74	1
9	123	2080000054	54	1	282.17	203	1.39	1
10	123	2080000054	55	1	130.66	94	1.39	1
11	123	2080000054	56	1	31.32	18	1.74	1
12	123	2080000054	57	1	5.22	3	1.74	1
13	123	2080000054	58	1	1.85	1	1.85	1
14	123	2080000054	59	1	14.8	8	1.85	1
15	123	2080000054	60	1	5.55	3	1.85	1
16	123	2080000054	61	1	3.7	2	1.85	1
17	123	2080000054	63	1	20.35	11	1.85	1
18	123	2080000054	64	1	22.1	13	1.7	1
19	123	2080000054	65	1	37.4	22	1.7	1
20	123	2080000054	66	1	30.6	18	1.7	1
21	123	2080000054	67	1	25.5	15	1.7	1
22	123	2080000054	68	1	39.1	23	1.7	1
23	123	2080000054	69	1	283.8	220	1.29	1
24	123	2080000054	70	1	59.2	32	1.85	1
25	123	2080000054	71	1	49.95	27	1.85	1
26	123	2080000054	72	1	13.51	7	1.93	1
27	123	2080000054	73	1	11.58	6	1.93	1
28	123	2080000054	74	1	9.65	5	1.93	1
29	123	2080000054	75	1	17.37	9	1.93	1
30	123	2080000054	76	1	7.72	4	1.93	1
31	123	2080000054	77	1	17.82	9	1.98	1
32	123	2080000054	78	1	5.07	3	1.69	1
33	123	2080000054	79	1	23.66	14	1.69	1
34	123	2080000054	80	1	10.14	6	1.69	1
35	123	2080000054	81	1	3.38	2	1.69	1
36	123	2080000054	82	1	7.92	4	1.98	1
37	123	2080000054	83	1	5.94	3	1.98	1
38	123	2080000054	84	1	7.92	4	1.98	1
39	123	2080000054	85	1	3.96	2	1.98	1
40	123	2080000054	86	1	2.96	2	1.98	1

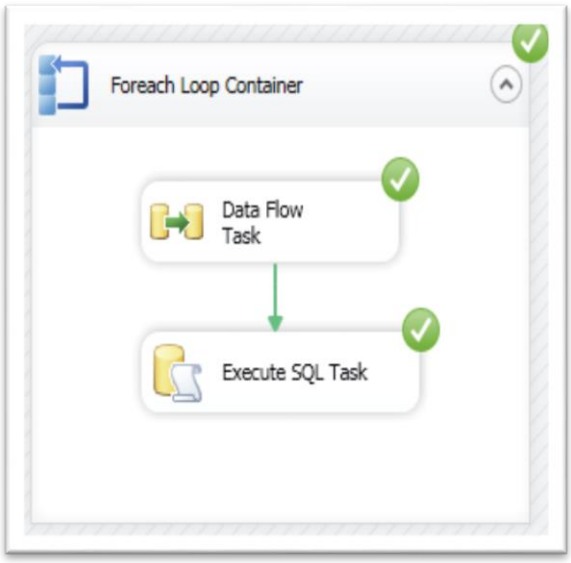
Query executed successfully.

CLEANED MOVEMENT TABLE IN STAGING AREA

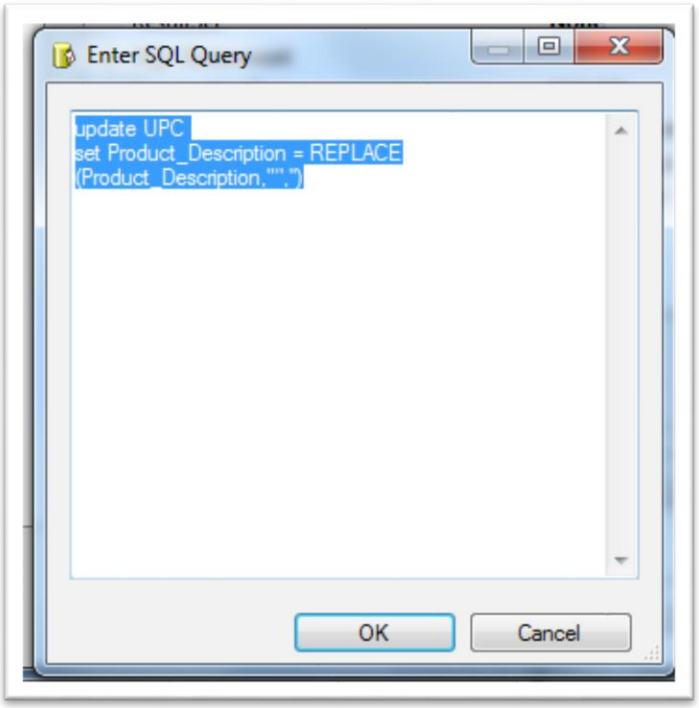
d. UPC Extraction and Transformation



Extracting UPC data to the staging area



SQL Task to clean the UPC Staging Area



SQL Query to Clean UPC Staging Area

SQLQuery11.sql - in...area (so7339 (436))* SQLQuery10.sql - in...area (so7339 (303))* SQLQuery8.sql - inf...-area (so7339 (175))*

```
select * from UPC
```

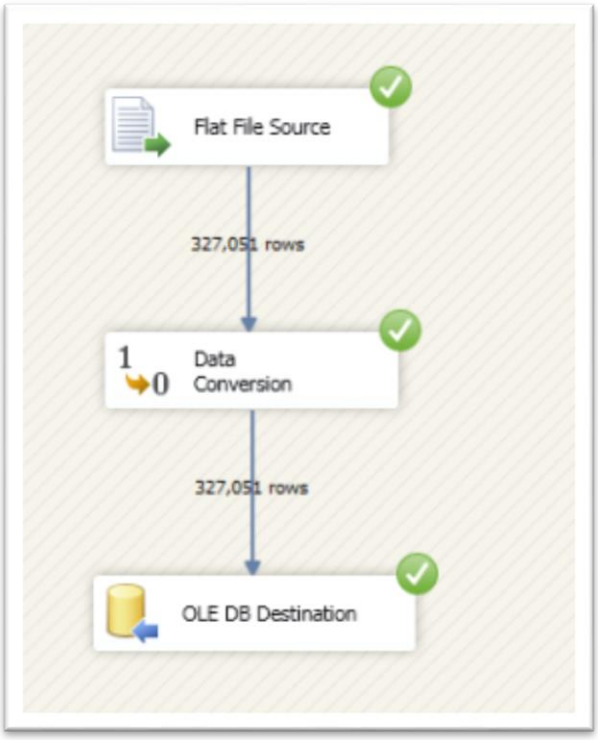
100 %

Results | Messages

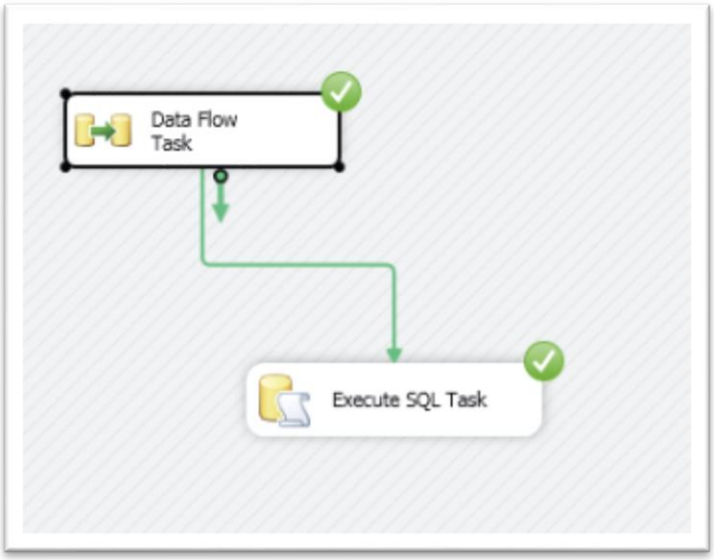
	Product_UPC_Num	Product_Description	Product_Com_Code
1	7430000438	BONINE	953
2	7430000590	RHEABAN TABS#	953
3	7430000606	UNISOM SLEEP AID	953
4	7430000607	UNISOM SLEEP AID TAB	953
5	7430000609	UNISOM SLEEP AID	953
6	7430000611	UNISOM SLEEP GELS	953
7	7498009600	YL PHARMACIST FORMUL	953
8	7498009700	YL PHARMACIST FORM.	953
9	7498009800	YL PHARMST FORMULA A	953
10	7782399152	VITALAN IBUPROFEN	953
11	7782399353	ACETAMINOPHEN 500 MG	953
12	30009340201	HALTRAN#	953
13	30009340202	HALTRAN TABLETS	953
14	30009346301	TRL-SZ MOTRINIB DIS	953
15	30009346302	MOTRIN IN TABLET	953
16	30009346303	MOTRIN IB TABLETS	953
17	30009346304	MOTRIN IB TABLETS	953
18	30009346311	~MOTRIN IB TABLETS	953
19	30009346318	MOTRIN IB TABLETS TW	953
20	30009346323	MOTRIN IB FASTCAP TA	953
21	30009346324	MOTRIN IB TAB 50+15	953
22	30009346330	~MOTRIN IB W/FREE 1	953
23	30009346332	MOTRIN BOGO TABLETS	953
24	30009348101	MOTRIN IB CAPLETS	953
25	30009348102	MOTRIN IB CPLT	953
26	30009348103	MOTRIN IB CAPLETS	953
27	30009348105	TRL-SZ MOTRINIB DIS	953
28	30009348109	~MOTRIN IB CAPLETS	953
29	30009348111	~MOTRIN IB CAPLETS C	953
30	30009348118	MOTRIN IB	953
31	30009348123	MOTRIN IB FASTCAP CA	953
32	30009348124	MOTRIN IB CPLT	953
33	30009348130	~MOTRIN IB W/FREE 1	953
34	30009348134	MOTRIN BOGO CAPLETS	953
35	30009364201	DRAMAMINE	953
36	30009364401	DRAMAMINE CHILD LIQ	953
37	30009364402	DRAMAMINE CHILD LIQ	953
38	30009364801	DRAMAMINE II TABLETS	953
39	30009373501	~MOTRIN IB SINUS TAB	953

CLEANED UPC TABLE IN STAGING AREA

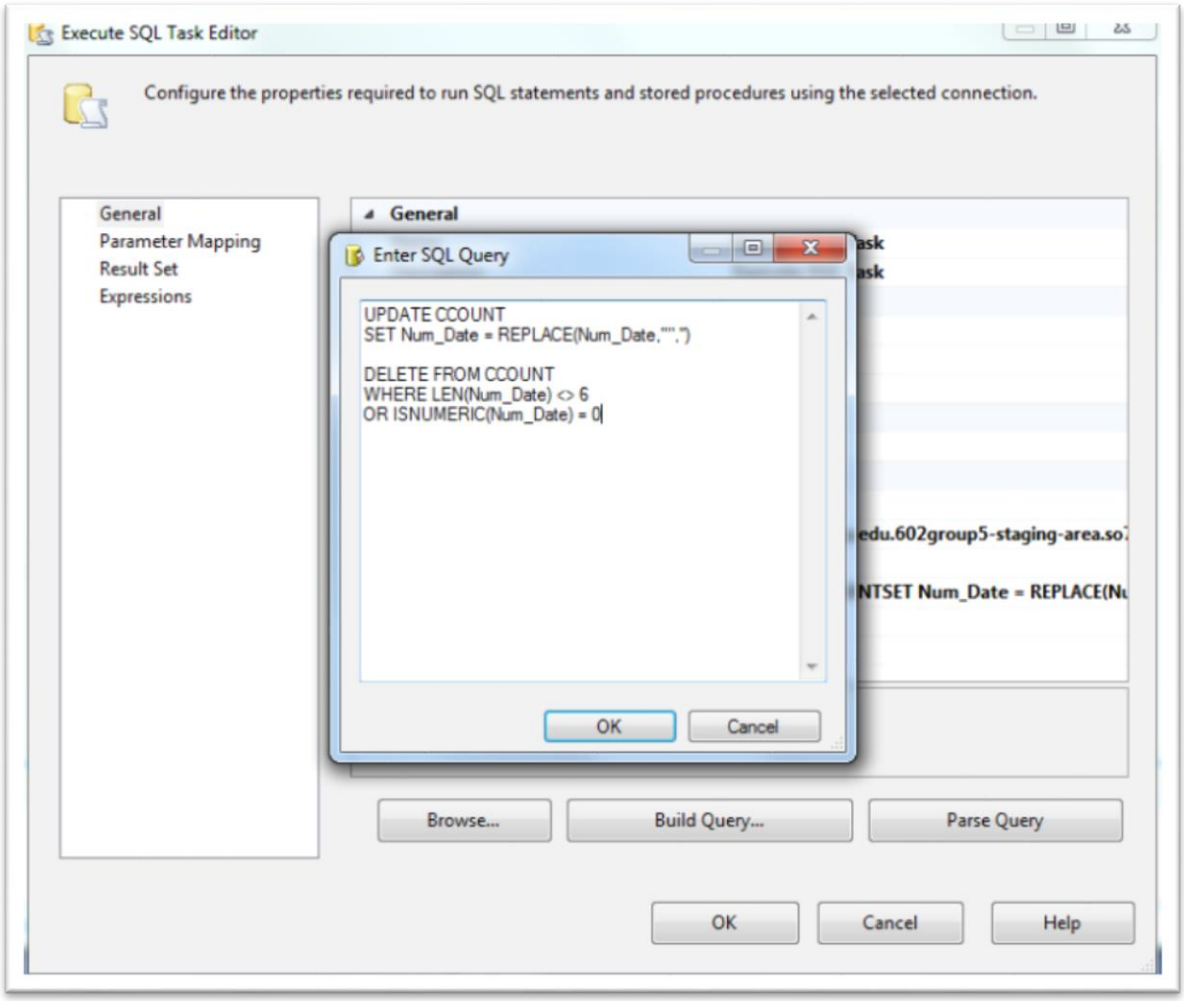
e. CCOUNT Extraction and Transformation



EXTRACTING CCOUNT DATA TO THE STAGING AREA



SQL TASK TO CLEAN THE CCOUNT STAGING AREA



SQL Query to Clean CCOUNT Staging Area

SQLQuery3.sql - inf...-area (so7339 (198))* SQLQuery2.sql - inf...-area (so7339 (201))* SQLQuery1.sql - inf.

```
SELECT * FROM CCOUNT
```

100 %

Results | Messages

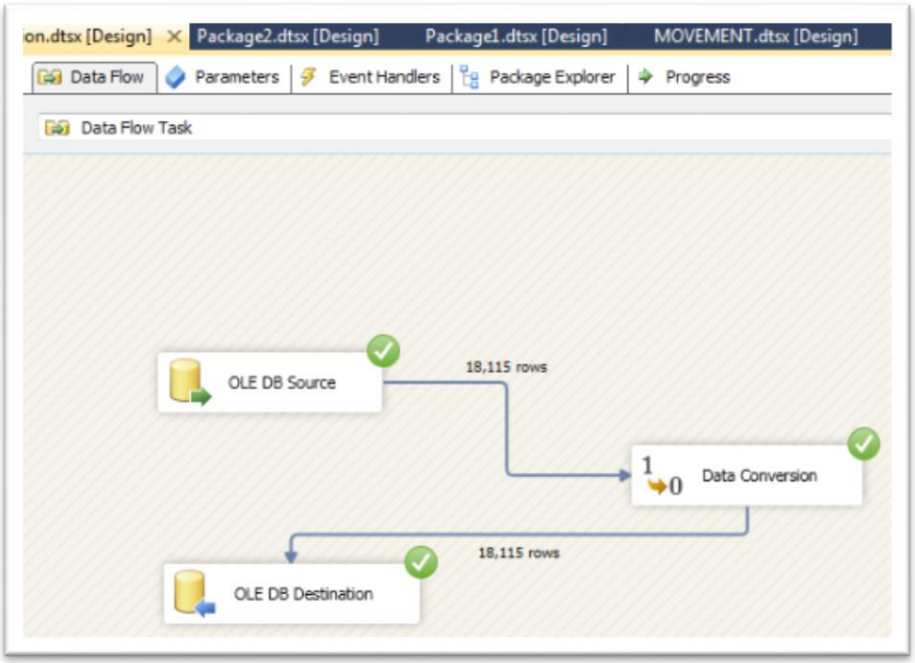
	Store_ID	Promo_Sales	Promo_Coup	Num_Date	Num_Week
1	114	46.56	0	950925	315
2	114	45.92	0	950926	315
3	114	48.34	0	950927	315
4	114	115.94	0	950928	316
5	114	74.88	0	950929	316
6	114	45.95	0	950930	316
7	114	112.93	0	951001	316
8	114	37.95	0	951002	316
9	114	35.96	0	951003	316
10	114	58.97	0	951004	316
11	114	27.78	0	951005	317
12	114	-7	0	951006	317
13	114	99.41	0	951007	317
14	114	144.94	0	951008	317
15	114	37.95	0	951009	317
16	114	116.38	0	951010	317
17	114	27.98	0	951011	317
18	114	31.46	0	951012	318
19	114	120.9	-11.13	951013	318
20	114	69.95	0	951014	318
21	114	35.97	0	951015	318
22	114	33.75	0	951016	318
23	114	138.93	0	951017	318
24	114	0	0	951018	318
25	114	82.94	0	951019	319
26	114	84.52	0	951020	319
27	114	56.97	0	951021	319
28	114	60.97	0	951022	319
29	114	4.99	0	951023	319
30	114	0	0	951024	319
31	114	90.44	0	951025	319
32	114	88.95	0	951026	320
33	114	34.98	0	951027	320
34	114	67.96	0	951028	320
35	114	4.99	-1.25	951029	320
36	114	13.36	0	951030	320
37	114	0	0	951031	320
38	114	23.57	0	951101	320
39	114	96.9	0	951102	321
40	114	47.93	0	951103	321

Query executed successfully.

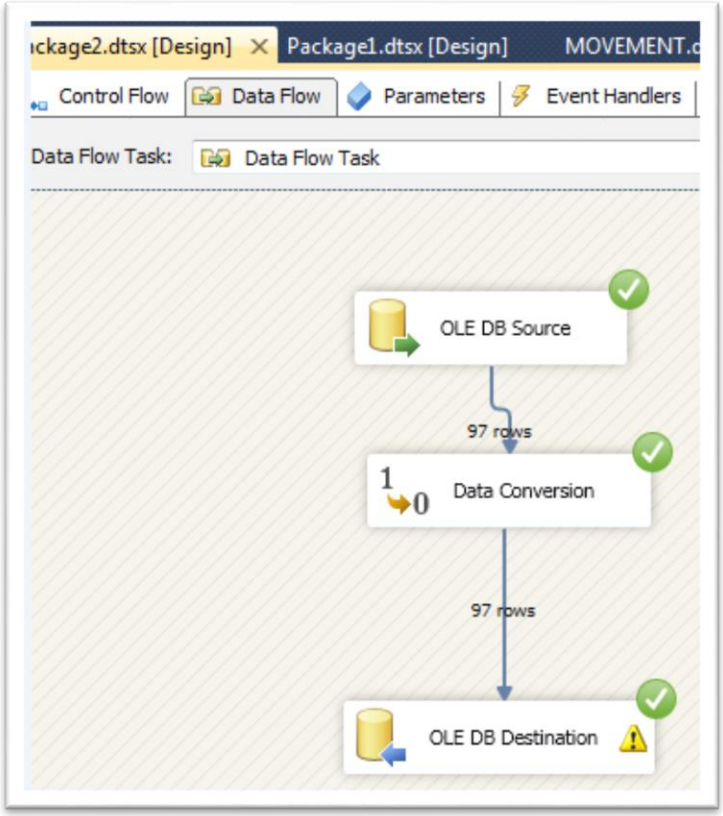
CLEANED CCOUNT TABLE IN STAGING AREA

11. Loading dimension and fact tables from Staging area to Data Warehouse

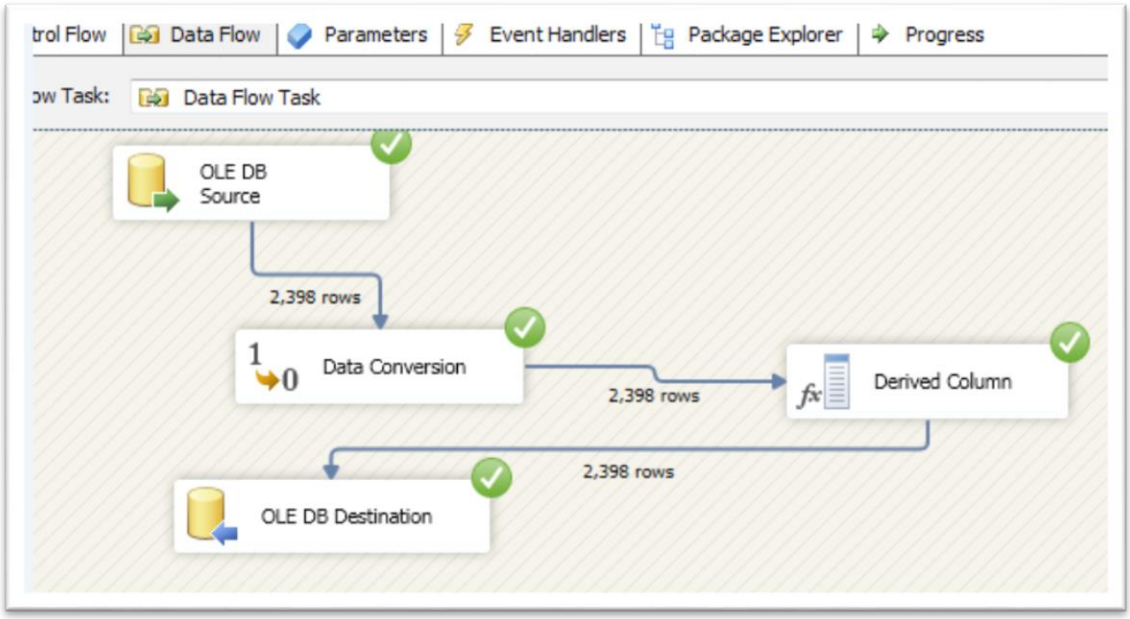
LOADING OF THE DEMO DIMENSION FROM STAGING AREA



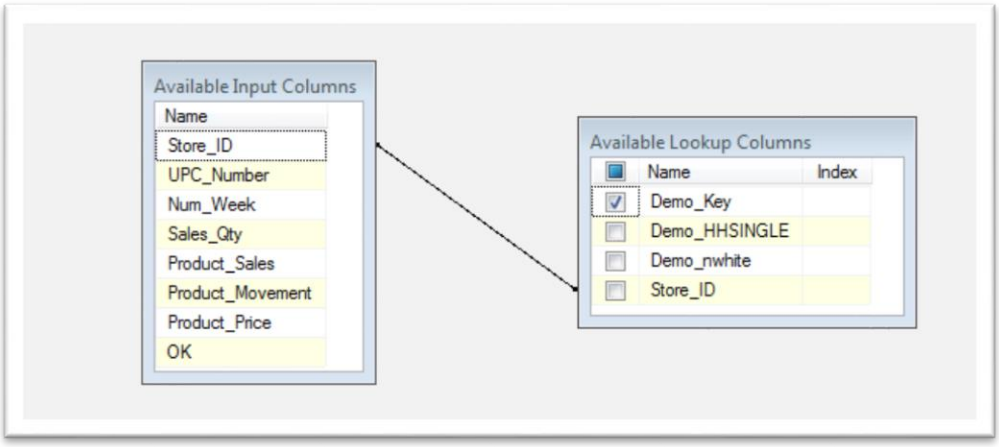
LOADING OF THE PRODUCT DIMENSION FROM STAGING AREA



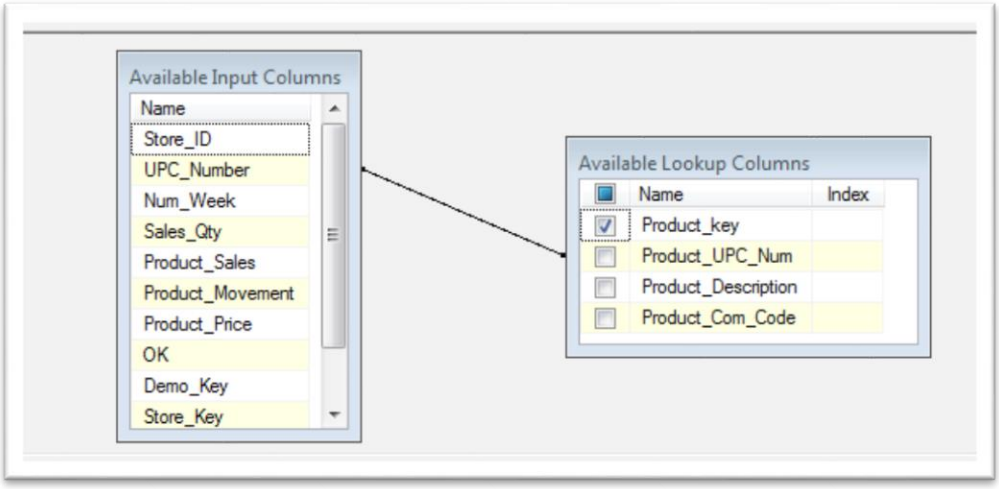
LOADING OF THE STORE DIMENSION FROM STAGING AREA



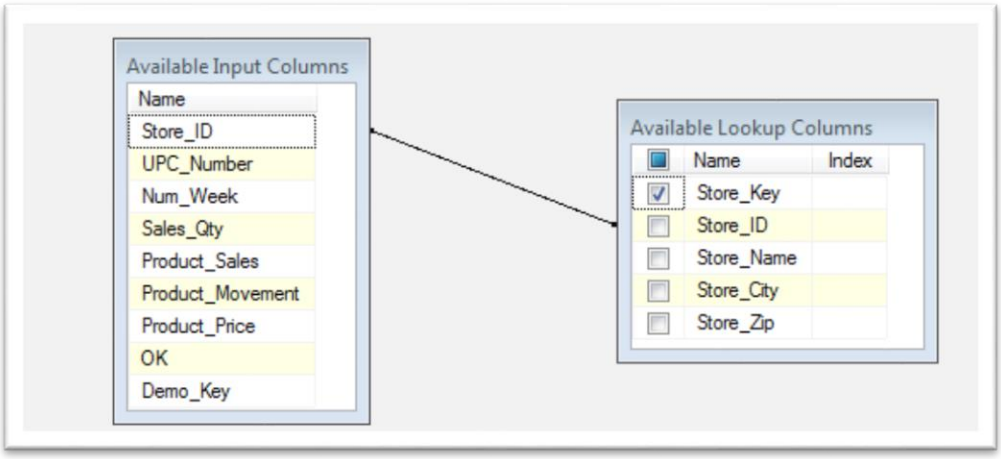
LOADING OF THE TIME DIMENSION FROM STAGING AREA



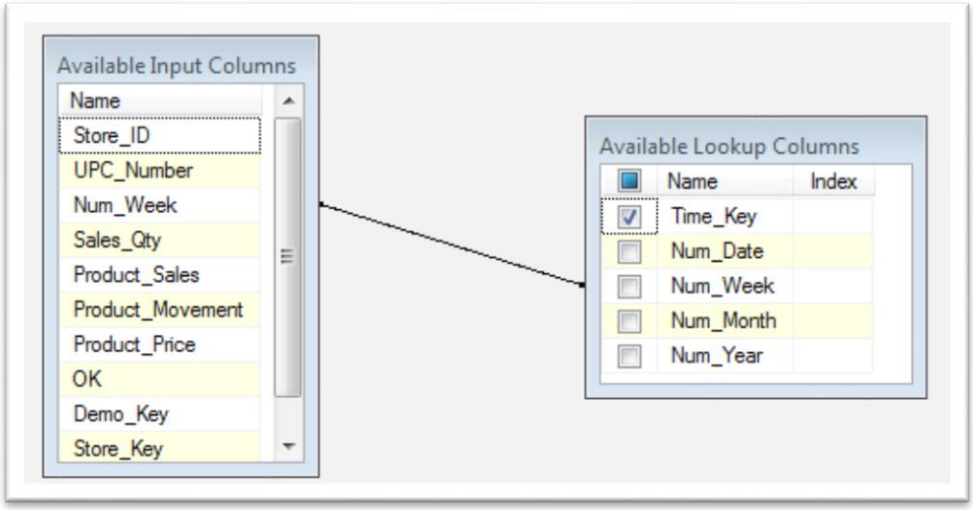
DEMO KEY LOOKUP FOR SALES FACT TABLE



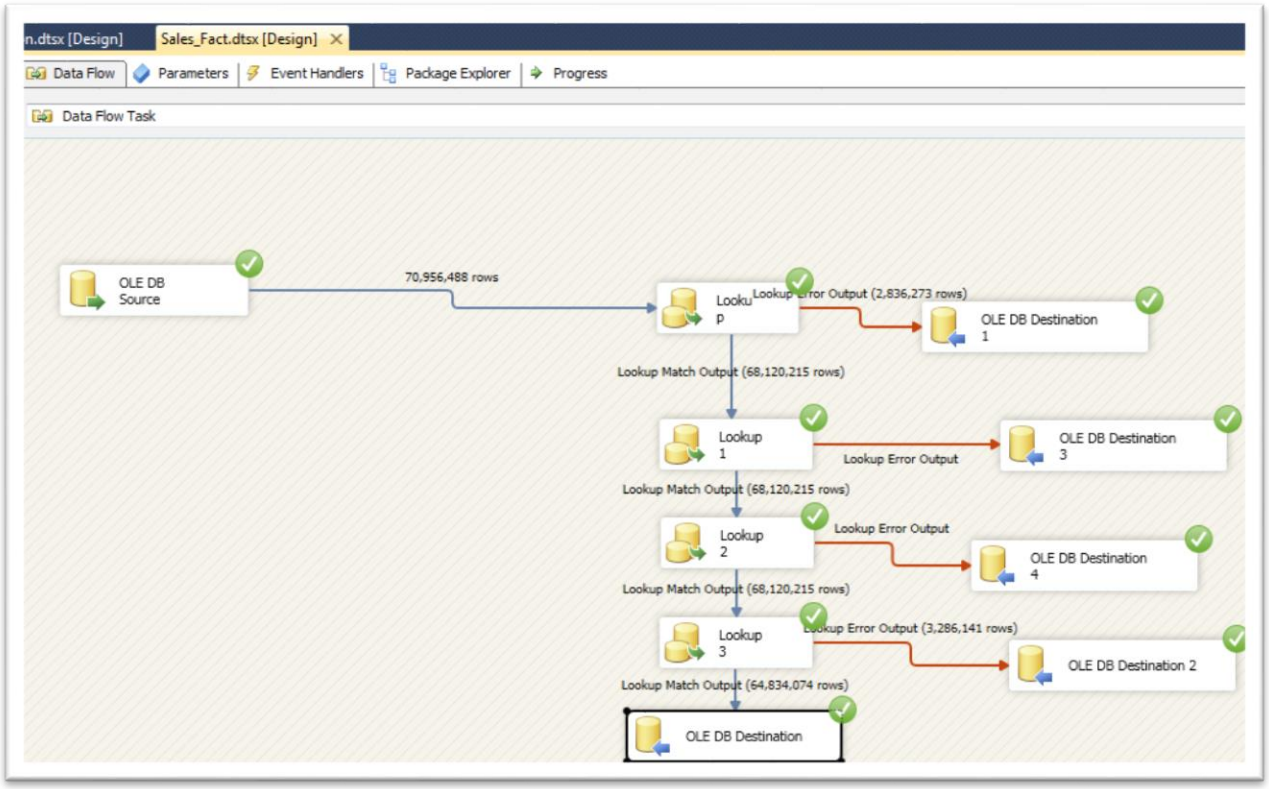
PRODUCT KEY LOOKUP FOR SALES FACT TABLE



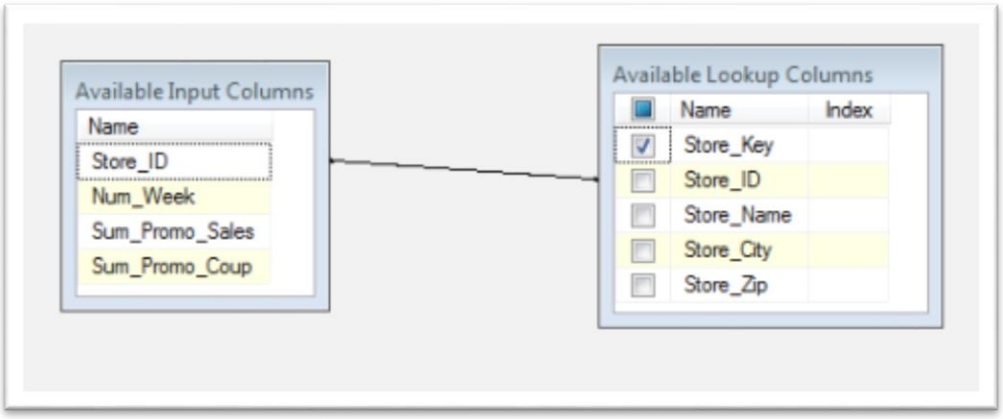
STORE KEY LOOKUP FOR SALES FACT TABLE



TIME KEY LOOKUP FOR SALES FACT TABLE

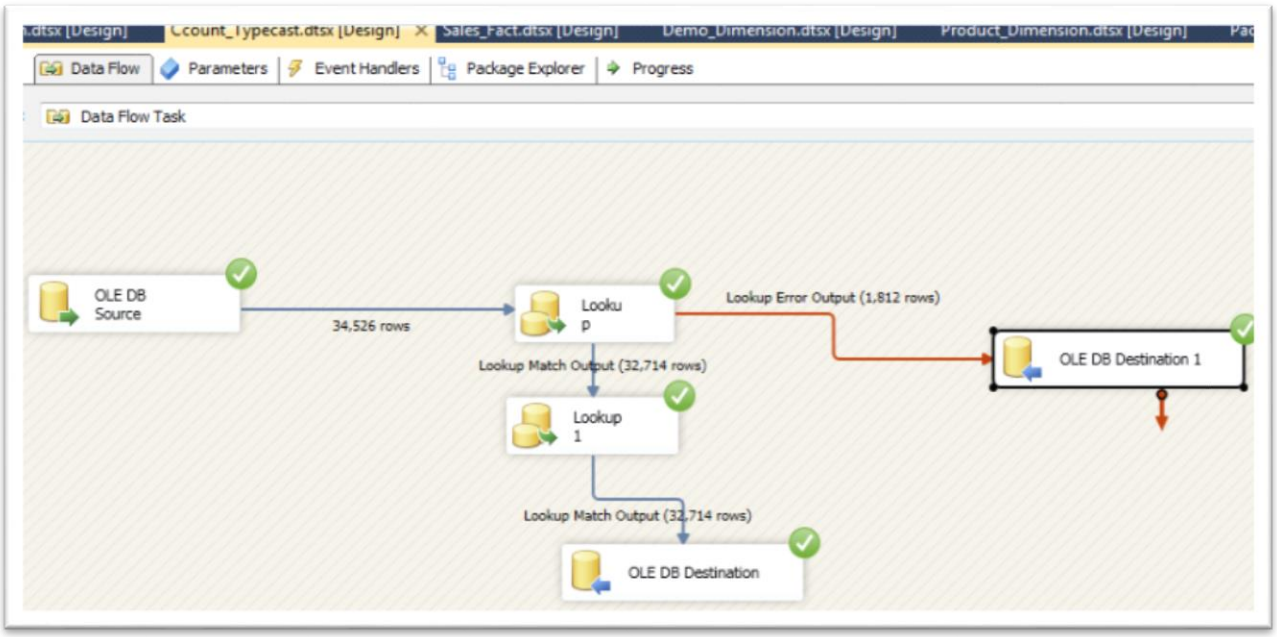


LOADING OF THE SALES FACT TABLE FROM STAGING AREA

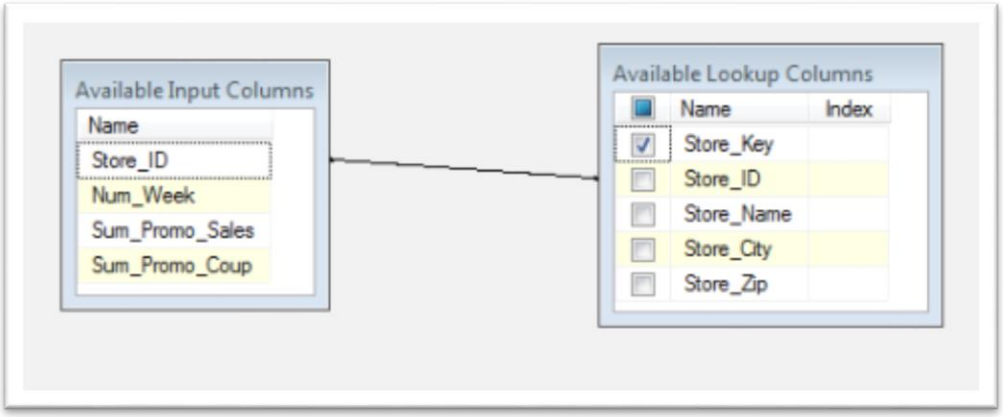


STORE KEY LOOKUP FOR PROMOTION SALES FACT TABLE

TIME KEY LOOKUP FOR PROMOTION SALES FACT TABLE

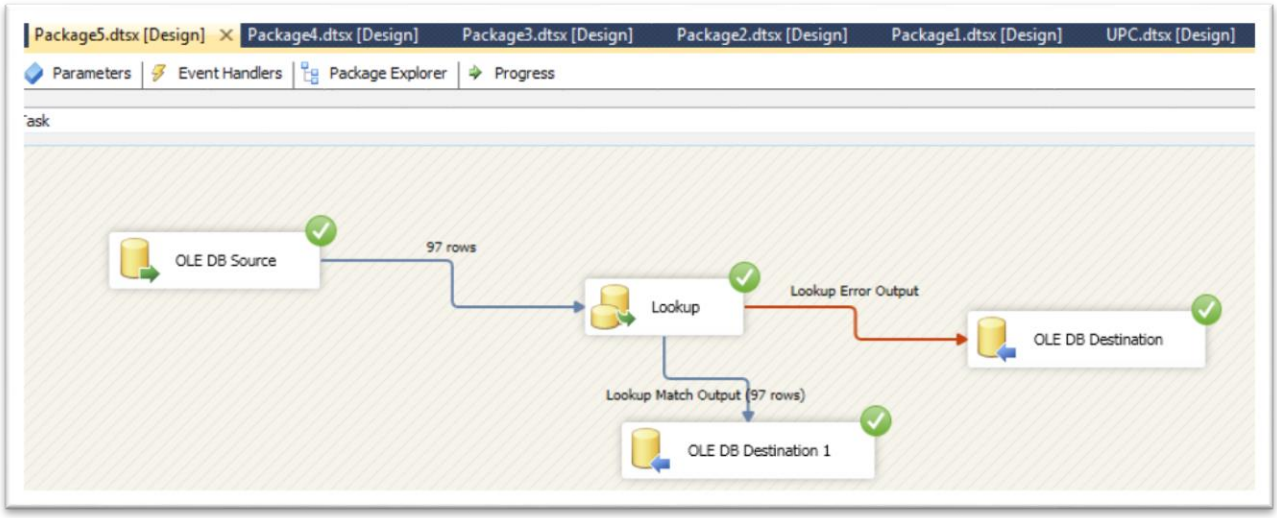


LOADING OF THE PROMOTION SALES FACT TABLE FROM STAGING AREA



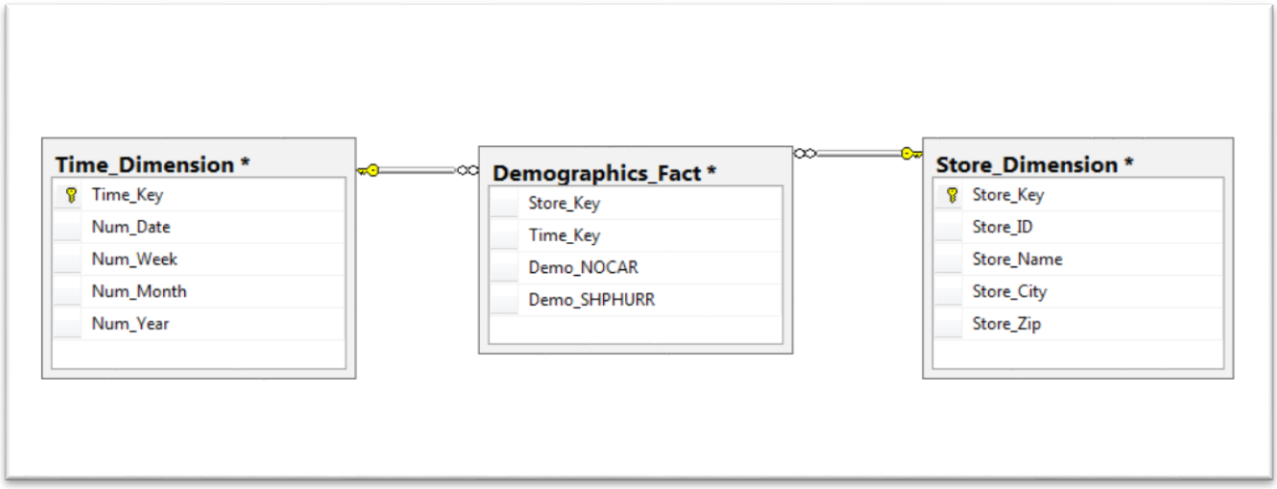
STORE KEY LOOKUP FOR DEMOGRAPHICS FACT TABLE

TIME KEY LOOKUP FOR DEMOGRAPHICS FACT TABLE

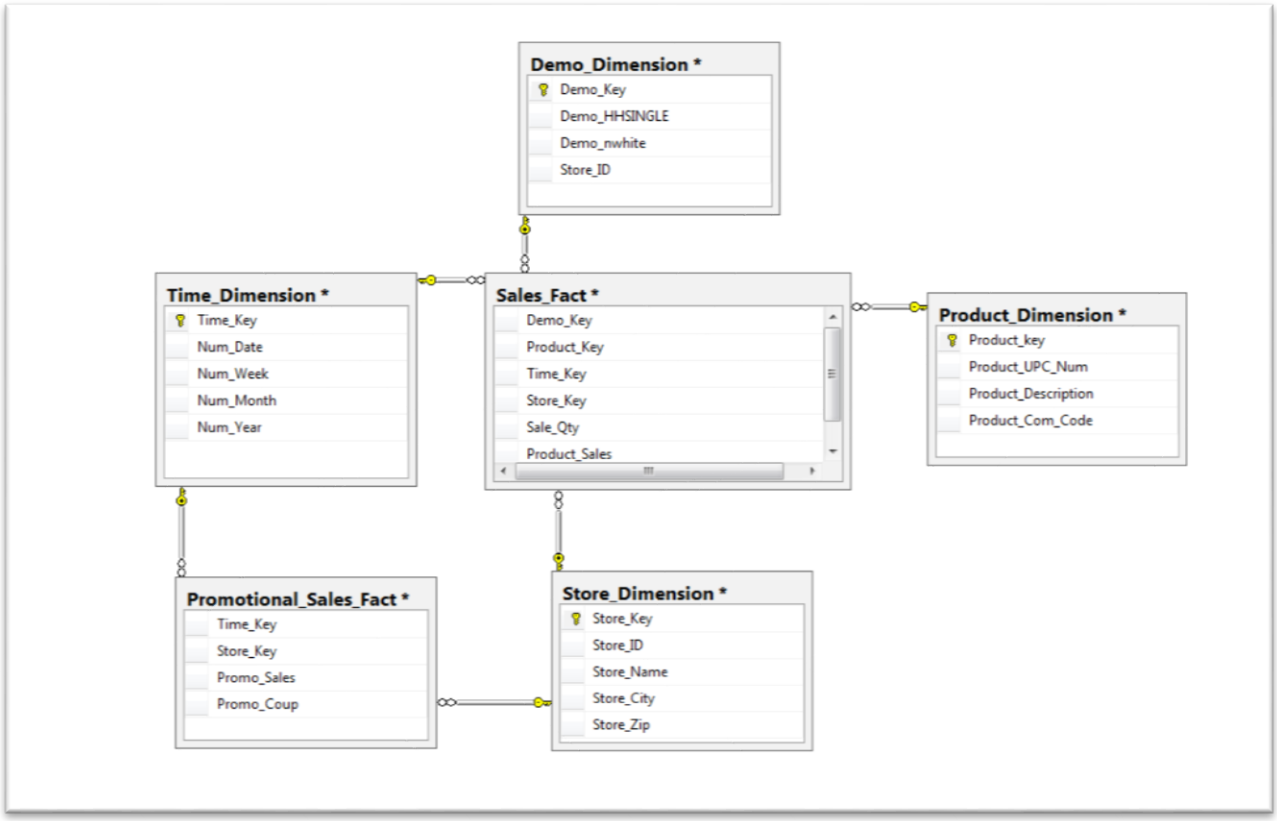


LOADING OF THE DEMOGRAPHICS FACT TABLE FROM STAGING AREA

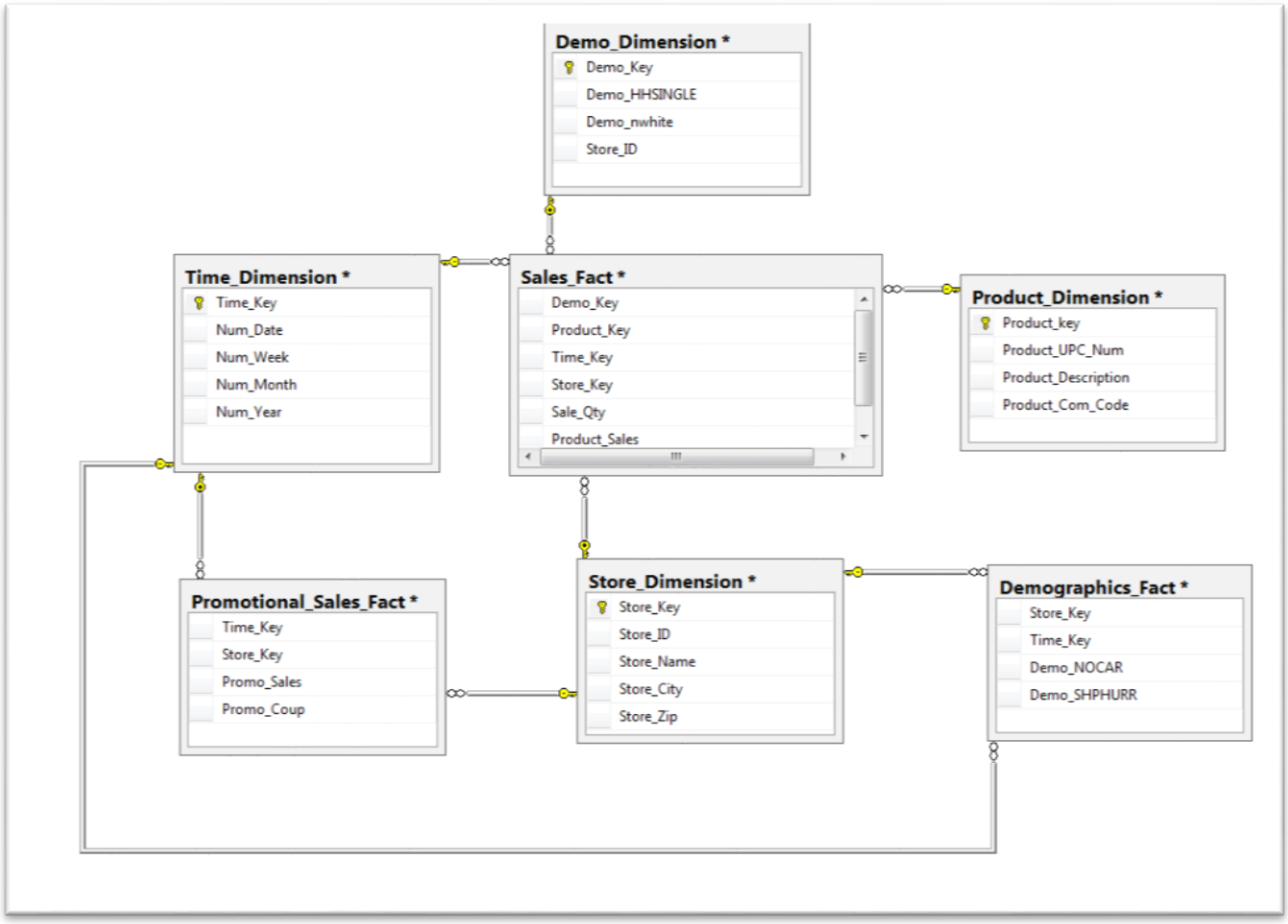
12. Overview of the data warehouse



DATABASE DIAGRAM OF THE STRATEGY DATA MART



DATABASE DIAGRAM OF THE MARKETING DATA MART



DATABASE DIAGRAM OF THE DATA WAREHOUSE

a. Overview of all the Dimension tables in the data warehouse

SQLQuery33.sql - in...aster (so7339 (145)) | SQLQuery32.sql - in...aster (so7339 (127))

```sql
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP 1000 [Product_key]
      ,[Product_UPC_Num]
      ,[Product_Description]
      ,[Product_Com_Code]
  FROM [602group5-dw-area].[dbo].[Product_Dimension]
```

100 %

Results | Messages

	Product_key	Product_UPC_Num	Product_Description	Product_Com_Code
1	1	1192603016	CAFFEDRINE CAPLETS 1	953
2	2	1192662108	SLEEPINAL SOFTGEL	953
3	3	1650001020	NERVINE TABS	953
4	4	1650001022	NERVINE SLEEP AID	953
5	5	1650004106	ALKA-SELTZER GOLD	953
6	6	1650004108	ALKA-SELTZER GOLD	953
7	7	1650004703	ALKA MINTS	953
8	8	2140649030	LEGATRIN PM	953
9	9	2586600493	PERCOGESIC A/F ANALG	953
10	10	2586610493	PERCOGESIC A/F ANALG	953
11	11	2586610501	ALEVE TABLETS	953
12	12	2586610502	ALEVE CAPLETS	953
13	13	2586610503	ALEVE TABLETS	953
14	14	2586610504	ALEVE CAPLETS	953
15	15	2586610505	ALEVE TABLETS	953
16	16	2586610506	ALEVE CAPLETS	953
17	17	3225259620	SUNBEAM HEAT WRAP MS	953
18	18	3680012732	TC MOTION SICKNESS T	953
19	19	3680012740	VALUE TIME ASPIRIN	953
20	20	3680012742	VALUE TIME ACETA	953
21	21	3680012888	TC IBUROFEN TABLETS$	953
22	22	3680012890	TC X/STR PAIN RLF TA	953
23	23	3680012892	TC X/STR N/A PAIN RE	953
24	24	3680012904	TOPCARE IBUPROFEN TA	953
25	25	3680014746	TOPCARE IBUPROFEN TA	953
26	26	3680019234	TC SLEEPAID TABLETS	953
27	27	3680019972	TC X/STR N/A GEL CAP	953
28	28	3680024827	TC X/STR PAIN RLF TA	953
29	29	3680029694	$TC PAIN REL INFANT	953
30	30	3680030527	TC STAY AWAKE TABLET	953
31	31	3680051752	~TC X/STR BACKACHE R	953
32	32	3680051754	~TC PMS MENSTRUAL CA	953
33	33	3680059914	TC NON-ASP CHILD COL	953
34	34	3680059916	TC X/STR N/A GEL TAB	953
35	35	3680059917	TC X/STR N/A GEL TAB	953

Query executed successfully.

SNAPSHOT OF THE PRODUCT DIMENSION

SQLQuery36.sql - in...aster (so7339 (123)) SQLQuery35.sql - in...

```
/****** Script for SelectTopNRows command from
SELECT TOP 1000 [Demo_Key]
      ,[Demo_HHSINGLE]
      ,[Demo_nwhite]
      ,[Store_ID]
  FROM [602group5-dw-area].[dbo].[Demo_Dimensi
```

100 %

Results | Messages

	Demo_Key	Demo_HHSINGLE	Demo_nwhite	Store_ID
1	1	0.28	0.13	2
2	2	0.26	0.15	4
3	3	0.21	0.09	5
4	4	0.21	0.04	8
5	5	0.21	0.12	9
6	6	0.49	0.49	12
7	7	0.18	0.12	14
8	8	0.26	0.08	18
9	9	0.13	0.16	21
10	10	0.21	0.09	28
11	11	0.29	0.05	32
12	12	0.61	0.17	33
13	13	0.22	0.05	40
14	14	0.17	0.05	44
15	15	0.26	0.12	45
16	16	0.17	0.16	47
17	17	0.32	0.17	48
18	18	0.25	0.09	49
19	19	0.23	0.08	50
20	20	0.19	0.04	51
21	21	0.20	0.13	52
22	22	0.27	0.20	53
23	23	0.26	0.10	54
24	24	0.22	0.07	56
25	25	0.15	0.03	59
26	26	0.18	0.07	62
27	27	0.18	0.10	64
28	28	0.14	0.17	65
29	29	0.25	0.11	67
30	30	0.31	0.37	68
31	31	0.25	0.17	70
32	32	0.33	0.10	71
33	33	0.21	0.20	72
34	34	0.22	0.11	73
35	35	0.26	0.06	74
36	36	0.52	0.56	75

Query executed successfully.

SNAPSHOT OF THE DEMO DIMENSION

SQLQuery36.sql - in...aster (so7339 (123)) | SQLQuery35.sql - in...aster (so7339 (115

```sql
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP 1000 [Store_Key]
      ,[Store_ID]
      ,[Store_Name]
      ,[Store_City]
      ,[Store_Zip]
  FROM [602group5-dw-area].[dbo].[Store_Dimension]
```

100 %

Results | Messages

	Store_Key	Store_ID	Store_Name	Store_City	Store_Zip
1	98	2	DOMINICKS 2	RIVER FOREST	60305
2	99	4	DOMINICKS 4	PARK RIDGE	60068
3	100	5	DOMINICKS 5	PALATINE	60067
4	101	8	DOMINICKS 8	OAK LAWN	60453
5	102	9	DOMINICKS 9	MORTON GROVE	60053
6	103	12	DOMINICKS 12	CHICAGO	60660
7	104	14	DOMINICKS 14	GLENVIEW	60025
8	105	18	DOMINICKS 18	RIVER GROVE	60171
9	106	21	DOMINICKS 21	HANOVER PARK	60103
10	107	28	DOMINICKS 28	MOUNT PROSPECT	60056
11	108	32	DOMINICKS 32	PARK RIDGE	60068
12	109	33	DOMINICKS 33	CHICAGO	60657
13	110	40	DOMINICKS 40	BRIDGEVIEW	60455
14	111	44	DOMINICKS 44	WESTERN SPRINGS	60558
15	112	45	DOMINICKS 45	WHEELING	60090
16	113	47	DOMINICKS 47	ADDISON	60101
17	114	48	DOMINICKS 48	SCHAUMBURG	60193
18	115	49	DOMINICKS 49	DOWNERS GROVE	60515
19	116	50	DOMINICKS 50	HICKORY HILLS	60457
20	117	51	DOMINICKS 51	PALOS HEIGHTS	60463
21	118	52	DOMINICKS 52	NORTHBROOK	60062
22	119	53	DOMINICKS 53	CHICAGO	60662
23	120	54	DOMINICKS 54	NAPERVILLE	60540
24	121	56	DOMINICKS 56	COUNTRYSIDE	60525
25	122	59	DOMINICKS 59	CRYSTAL LAKE	60014
26	123	62	DOMINICKS 62	NORTHFIELD	60093
27	124	64	DOMINICKS 64	VILLA PARK	60181
28	125	65	DOMINICKS 65	HANOVER PARK	60103
29	126	67	DOMINICKS 67	OAKBROOK TERR...	60521
30	127	68	DOMINICKS 68	CHICAGO	60625
31	128	70	DOMINICKS 70	JOLIET	60435
32	129	71	DOMINICKS 71	NORTH RIVERSIDE	60546
33	130	72	DOMINICKS 72	LINCOLNWOOD	60646
34	131	73	DOMINICKS 73	CHICAGO	60629
35	132	74	DOMINICKS 74	NORRIDGE	60634

Query executed successfully.

SNAPSHOT OF THE STORE DIMENSION

SQLQuery36.sql - in...aster (so7339 (123)) | SQLQuery35.sql - in...aste

```sql
/****** Script for SelectTopNRows command from SSI
SELECT TOP 1000 [Time_Key]
      ,[Num_Date]
      ,[Num_Week]
      ,[Num_Month]
      ,[Num_Year]
  FROM [602group5-dw-area].[dbo].[Time_Dimension]
```

100 %

Results | Messages

	Time_Key	Num_Date	Num_Week	Num_Month	Num_Year
1	2399	14	1	9	1989
2	2400	21	2	9	1989
3	2401	28	3	9	1989
4	2402	5	4	10	1989
5	2403	12	5	10	1989
6	2404	19	6	10	1989
7	2405	26	7	10	1989
8	2406	2	8	11	1989
9	2407	9	9	11	1989
10	2408	16	10	11	1989
11	2409	23	11	11	1989
12	2410	30	12	11	1989
13	2411	7	13	12	1989
14	2412	14	14	12	1989
15	2413	21	15	12	1989
16	2414	28	16	12	1989
17	2415	4	17	1	1990
18	2416	11	18	1	1990
19	2417	18	19	1	1990
20	2418	25	20	1	1990
21	2419	1	21	2	1990
22	2420	8	22	2	1990
23	2421	15	23	2	1990
24	2422	22	24	2	1990
25	2423	1	25	3	1990
26	2424	8	26	3	1990
27	2425	15	27	3	1990
28	2426	22	28	3	1990
29	2427	29	29	3	1990
30	2428	5	30	4	1990
31	2429	12	31	4	1990
32	2430	19	32	4	1990
33	2431	26	33	4	1990
34	2432	3	34	5	1990
35	2433	10	35	5	1990
36	2434	17	36	5	1990

Query executed successfully.

SNAPSHOT OF THE TIME DIMENSION

b. Overview of all the Fact tables in the data warehouse

SQLQuery39.sql - in...aster (so7339 (170)) | SQLQuery38.sql - in...aster (so7339 (112))

```sql
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP 1000 [Demo_Key]
      ,[Product_Key]
      ,[Time_Key]
      ,[Store_Key]
      ,[Sale_Qty]
      ,[Product_Sales]
  FROM [602group5-dw-area].[dbo].[Sales_Fact]
```

100 %

Results | Messages

	Demo_Key	Product_Key	Time_Key	Store_Key	Sale_Qty	Product_Sales
1	1	3393	2658	98	10	25
2	1	3393	2659	98	3	9
3	1	3393	2660	98	4	12
4	1	3393	2661	98	7	19
5	1	3393	2662	98	12	30
6	1	3393	2663	98	5	14
7	1	3393	2664	98	5	15
8	1	3393	2665	98	6	18
9	1	3393	2666	98	6	18
10	1	3393	2667	98	9	27
11	1	3393	2668	98	3	9
12	1	3393	2669	98	6	18
13	1	3393	2670	98	9	27
14	1	3393	2671	98	9	27
15	1	3393	2672	98	6	18
16	1	3393	2674	98	3	9
17	1	3393	2675	98	8	24
18	1	3393	2676	98	8	24
19	1	3393	2677	98	2	6
20	1	3393	2678	98	28	69
21	1	3393	2679	98	5	15
22	1	3393	2680	98	2	6
23	1	3393	2681	98	8	24
24	1	3393	2682	98	7	21
25	1	3393	2683	98	5	15
26	1	3393	2684	98	8	24
27	1	3393	2685	98	3	9
28	1	3393	2686	98	9	27
29	1	3393	2687	98	4	12
30	1	3393	2688	98	5	14
31	1	3393	2689	98	2	5
32	1	3393	2690	98	7	19
33	1	3393	2691	98	5	14
34	1	3393	2694	98	10	26
35	1	3393	2695	98	6	16

Query executed successfully.

SNAPSHOT OF THE SALES FACT

SQLQuery39.sql - in...aster (so7339 (170)) | SQLQuery38.sql

```sql
/****** Script for SelectTopNRows command
SELECT TOP 1000 [Time_Key]
      ,[Store_Key]
      ,[Promo_Sales]
      ,[Promo_Coup]
  FROM [602group5-dw-area].[dbo].[Promotio
```

100 %

Results | Messages

	Time_Key	Store_Key	Promo_Sales	Promo_Coup
1	105	157	216.67	0.00
2	45	112	274.96	-21.00
3	75	133	234.44	-6.00
4	129	177	735.50	-152.00
5	2	98	1428.97	-136.00
6	47	113	517.90	0.00
7	5	100	561.57	-89.00
8	72	130	662.57	-11.00
9	75	133	675.88	-238.00
10	44	111	674.06	0.00
11	48	114	253.73	0.00
12	77	135	2367.92	-24.00
13	134	181	308.05	0.00
14	2	98	375.52	-282.92
15	111	162	17.99	0.00
16	134	181	403.80	0.00
17	44	111	587.95	-292.32
18	45	112	165.90	0.00
19	53	119	25.00	0.00
20	77	135	626.57	-160.00
21	107	159	1818.34	-0.30
22	45	112	118.41	-3.00
23	309	191	0.00	-5.00
24	32	108	910.64	0.00
25	48	114	427.76	-2.00
26	111	162	1328.95	-126.00
27	106	158	165.24	-14.00
28	33	109	137.40	0.00
29	68	127	427.55	-3.00
30	74	132	683.51	0.00
31	105	157	1391.05	-819.94
32	104	156	1509.09	-268.00
33	56	121	690.61	-20.00
34	12	103	589.36	0.00
35	73	131	3817.46	-1630.42

Query executed successfully.

SNAPSHOT OF THE PROMOTIONAL SALES FACT

SQLQuery39.sql - in...aster (so7339 (170)) SQLQuery38.sql - in...

```sql
/****** Script for SelectTopNRows command from
SELECT TOP 1000 [Store_Key]
      ,[Time_Key]
      ,[Demo_NOCAR]
      ,[Demo_SHPHURR]
  FROM [602group5-dw-area].[dbo].[Demographics_
```

100 %

Results | Messages

	Store_Key	Time_Key	Demo_NOCAR	Demo_SHPHURR
1	98	60	0.12	0.12
2	99	60	0.06	0.12
3	100	60	0.03	0.19
4	101	60	0.08	0.14
5	102	60	0.04	0.16
6	103	60	0.48	0.05
7	104	60	0.03	0.22
8	105	60	0.14	0.10
9	106	60	0.02	0.23
10	107	60	0.05	0.17
11	108	60	0.07	0.11
12	109	60	0.51	0.03
13	110	60	0.05	0.13
14	111	60	0.04	0.20
15	112	60	0.02	0.16
16	113	60	0.02	0.17
17	114	60	0.02	0.13
18	115	60	0.05	0.17
19	116	60	0.04	0.13
20	117	60	0.03	0.14
21	118	60	0.01	0.21
22	119	60	0.15	0.14
23	120	60	0.02	0.19
24	121	60	0.03	0.17
25	122	60	0.03	0.24
26	123	60	0.03	0.24
27	124	60	0.04	0.17
28	125	60	0.02	0.23
29	126	60	0.05	0.16
30	127	60	0.31	0.09
31	128	60	0.08	0.13
32	129	60	0.15	0.10
33	130	60	0.07	0.15
34	131	60	0.13	0.13
35	132	60	0.14	0.10
36	133	60	0.55	0.04

Query executed successfully.

SNAPSHOT OF THE DEMOGRAPHICS FACT

13. SQL statements to create Staging Area

Staging Database Creation

```
CREATE DATABASE [602group5-staging-area]
```

WeekDecode Staging Table

```
CREATE TABLE [dbo].[WeekDecode](
	[Week #] [nvarchar](100) NULL,
	[Start] [nvarchar](100) NULL,
	[End] [nvarchar](100) NULL,
	[Special Events] [nvarchar](100) NULL
)
```

UPC Staging Table

```
CREATE TABLE [dbo].[UPC](
	[Product_Description] [nvarchar](100) NULL,
	[Product_UPC_Num] [nvarchar](100) NULL,
	[Product_Com_Code] [nvarchar](100) NULL
)
```

Movement Staging Table

```
CREATE TABLE [dbo].[Movement](
	[Store_ID] [int] NULL,
	[UPC_Number] [numeric](18, 0) NULL,
	[Num_Week] [int] NULL,
	[Sales_Qty] [int] NULL,
	[Product_Sales] [decimal](18, 2) NULL,
	[Product_Movement] [decimal](18, 2) NULL,
	[Product_Price] [decimal](18, 2) NULL,
	[OK] [nvarchar](100) NULL
)
```

Demographics Staging Table

```
CREATE TABLE [dbo].[Demographics](
	[Store_ID] [nvarchar](100) NULL,
	[Store_Name] [nvarchar](100) NULL,
	[Store_City] [nvarchar](100) NULL,
	[Store_Zip] [nvarchar](100) NULL,
	[Demo_HHSINGLE] [nvarchar](100) NULL,
	[Demo_NWHITE] [nvarchar](100) NULL,
	[Demo_NOCAR] [nvarchar](100) NULL,
	[Demo_SHPHURR] [nvarchar](100) NULL
)
```

CCOUNT Staging Table

```
CREATE TABLE [dbo].[CCOUNT](
	[Store_ID] [int] NULL,
	[Promo_Sales] [numeric](18, 2) NULL,
	[Promo_Coup] [numeric](18, 2) NULL,
	[Num_Date] [date] NULL,
	[Num_Week] [int] NULL
)
```

CCOUNT Aggregate Staging Table

```
CREATE TABLE [dbo].[CCOUNT_Agg](
	[Store_ID] [int] NULL,
	[Sum_Promo_Sales] [float] NULL,
	[Sum_Promo_Coup] [float] NULL,
	[Num_Week] [int] NULL
)
```

14. SQL statements to create Data Warehouse

Data Warehouse Creation

```
CREATE DATABASE [602group5-dw-area]
```

Time_Dimension Creation

```
CREATE TABLE [dbo].[Time_Dimension](
	[Time_Key] [int] IDENTITY(1,1) NOT NULL,
	[Num_Date] [int] NULL,
	[Num_Week] [int] NULL,
	[Num_Month] [int] NULL,
	[Num_Year] [int] NULL,
PRIMARY KEY CLUSTERED
(
	[Time_Key] ASC
))
```

Store_Dimension Creation

```
CREATE TABLE [dbo].[Store_Dimension](
	[Store_Key] [int] IDENTITY(1,1) NOT NULL,
	[Store_ID] [int] NULL,
	[Store_Name] [nvarchar](50) NULL,
	[Store_City] [nvarchar](50) NULL,
	[Store_Zip] [int] NULL,
PRIMARY KEY CLUSTERED
(
	[Store_Key] ASC
))
```

Demo_Dimension Creation

```
CREATE TABLE [dbo].[Demo_Dimension](
	[Demo_Key] [int] IDENTITY(1,1) NOT NULL,
	[Demo_HHSINGLE] [numeric](18, 2) NULL,
	[Demo_nwhite] [numeric](18, 2) NULL,
	[Store_ID] [int] NULL,
PRIMARY KEY CLUSTERED
(
	[Demo_Key] ASC
))
```

Product_Dimension Creation

```
CREATE TABLE [dbo].[Product_Dimension](
	[Product_key] [int] IDENTITY(1,1) NOT NULL,
	[Product_UPC_Num] [decimal](18, 0) NULL,
	[Product_Description] [nvarchar](100) NULL,
	[Product_Com_Code] [decimal](18, 0) NULL,
PRIMARY KEY CLUSTERED
(
	[Product_key] ASC
))
```

Sales_Fact Creation

```
CREATE TABLE [dbo].[Sales_Fact](
	[Demo_Key] [int] NULL,
	[Product_Key] [int] NULL,
	[Time_Key] [int] NULL,
	[Store_Key] [int] NULL,
	[Sale_Qty] [decimal](18, 0) NULL,
	[Product_Sales] [decimal](18, 0) NULL
)
```

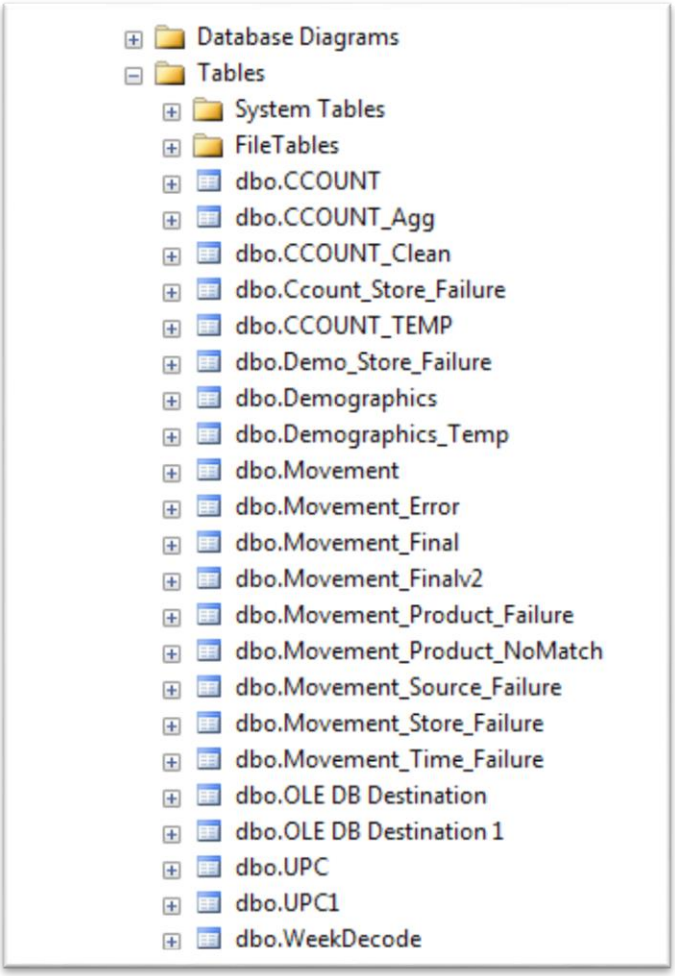
Promotional_Sales_Fact Creation

```
CREATE TABLE [dbo].[Promotional_Sales_Fact](
	[Time_Key] [int] NULL,
	[Store_Key] [int] NULL,
	[Promo_Sales] [decimal](18, 2) NULL,
	[Promo_Coup] [decimal](18, 2) NULL
)
```

Demographics_Fact Creation

```
CREATE TABLE [dbo].[Demographics_Fact](
	[Store_Key] [int] NOT NULL,
	[Time_Key] [int] NULL,
	[Demo_NOCAR] [numeric](18, 2) NULL,
	[Demo_SHPHURR] [numeric](18, 2) NULL,
PRIMARY KEY CLUSTERED
(
	[Store_Key] ASC
))
```

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



STAGING AREA BEFORE THE REMOVAL OF TEMPORARY TABLES

There have been several tables in the staging area that were created for having a temporary table to simplify cleaning or contained dirty data rejected from the initial versions of the tables before transformation. Such tables were dropped from the database. The tables existing before data transformation process which contained dirty data were removed. The above screen shot lists all the tables before this drop process was initiated.

16. Business Intelligence reporting with SSRS, SSAS and Report Builder

a. Reporting Plan

Business intelligence (BI) is the process in which data is analyzed and presented to corporate executives and business managers to make informed decisions. Business Intelligence spans a lot of tools applications and methodologies that facilitate organizations in data collection from internal systems and external data sources, transform them for analysis, develop and run queries against the data, and generate reports, dashboards and data visualizations to make the analytical results available to corporate decision makers as well as operational workers.

We are using SSRS, SSAS, and a combination of SSAS, SSRS and ReportBuilder3.0 in our report building with the following breakdown for each of the reports:

Question Number	Tool
1	Report Builder 3.0
2	Report Builder 3.0
3	Excel on SSAS
4	SSRS on SSAS
5	SSRS

b. Determination of target reports that satisfy business questions

Question 1: How have the promo sales been affected across the stores on the basis of coupons?

Analysis of the cube created from Report Builder 3.0 only

We used Report Builder 3.0 to answer this business question. We used store dimension and promotional sales fact table to answer this business question and build our report. We used the area chart to visualize the effect of coupons on the promotional sales.

Question 2: Which stores are in areas of high single households and what are the commodities/products that sell more and which sell least?

Analysis of the cube created from Report Builder 3.0 only

We used Report Builder 3.0 to answer this business question. We used store dimension, demo dimension, product dimension and time dimension and sales fact table to answer this business question and build our report. We used the bar charts to visualize the sales statistics of products across categories in areas of high single households.

Question 3: Which stores or store areas have a high percentage of customers with no cars in order to set up a home delivery service?

Analysis of the cube created from Excel on SSAS.

We used Excel on SSAS to answer this business question. We used store dimension and demographic fact table to answer this business question and build our report. We used the excel pivot table and chart to visualize the store areas that have a high percentage of customers with no cars in order to set up a home delivery service.

Question 4: Which stores have a large number of hurried shoppers and would benefit from setting up a curbside pickup?

Analysis of the cube created from SSRS on SSAS.

We used SSRS on SSAS to answer this business question. We used store dimension and demographic fact table to answer this business question and build our report. We used the SSRS report builder tool to visualize the store areas that have a large number of hurried shoppers in order to set up a curb side pick-up service.

Question 5: How is the sales performance of all the beer brands across the stores in areas with more than 50% non-white population in the last 6 years (1991 - 1997)?

Analysis of the cube created from SSRS only

We used SSRS to answer this business question. We used store dimension, demo dimension, product dimension and time dimension and sales fact table to answer this business question and build our report. We used the SSRS report builder tool to visualize the sales performance of all the beer brands across the stores in areas with more than 50% non-white population in the last 6 years.

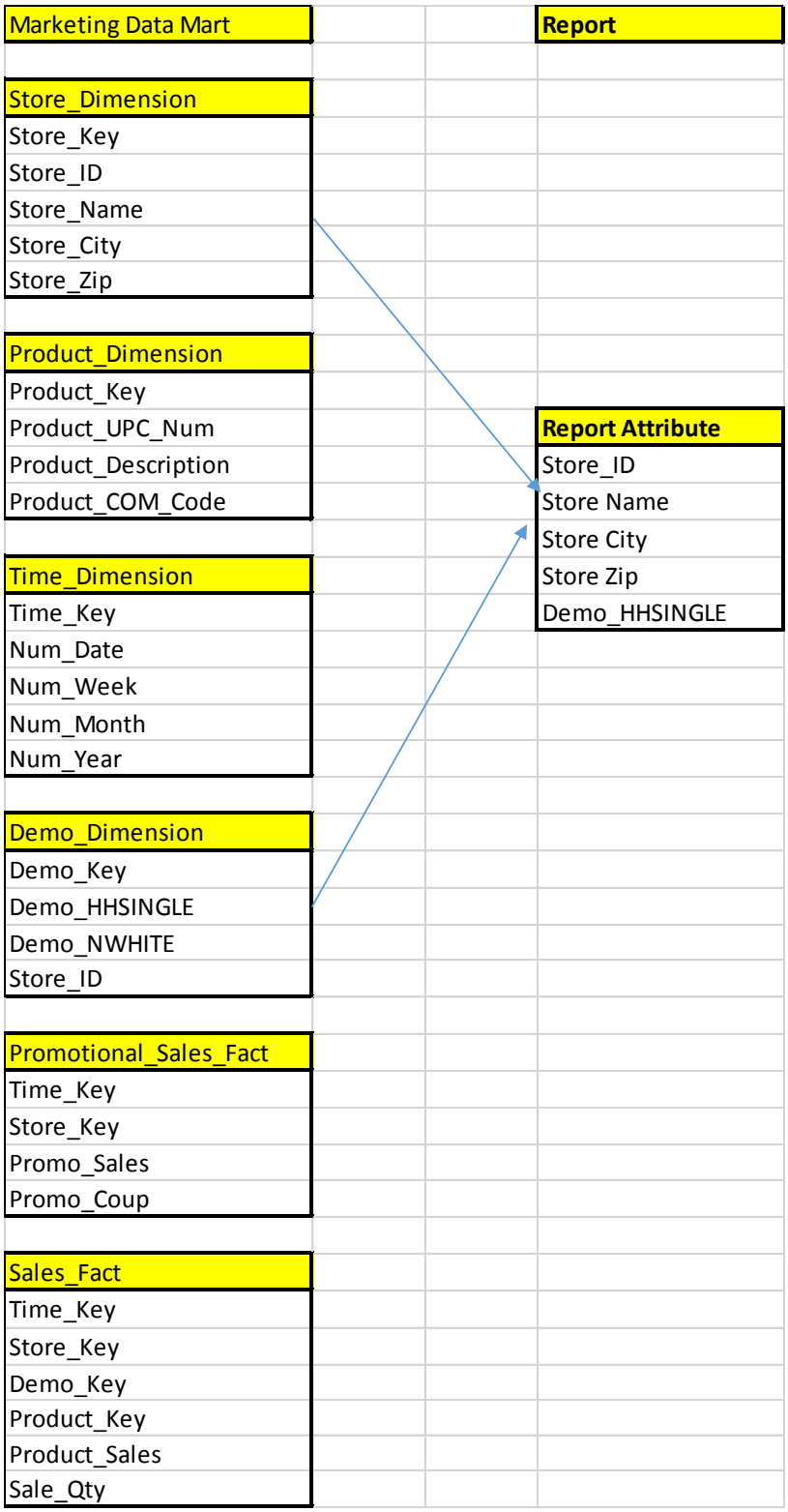
c. Mappings of the attributes from the data marts to the report

Question 1: How have the promo sales been affected across the stores on the basis of coupons?

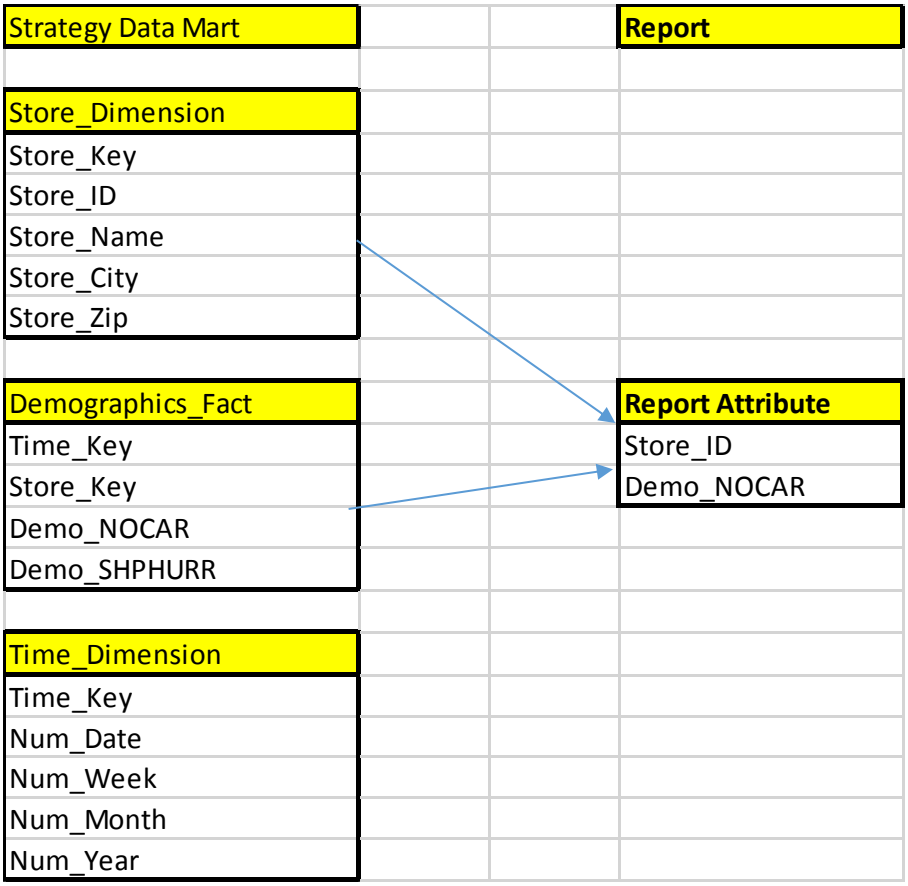
Marketing Data Mart
Store_Dimension
Store_Key
Store_ID
Store_Name
Store_City
Store_Zip
Product_Dimension
Product_Key
Product_UPC_Num
Product_Description
Product_COM_Code
Time_Dimension
Time_Key
Num_Date
Num_Week
Num_Month
Num_Year
Demo_Dimension
Demo_Key
Demo_HHSINGLE
Demo_NWHITE
Store_ID
Promotional_Sales_Fact
Time_Key
Store_Key
Promo_Sales
Promo_Coup
Sales_Fact
Time_Key
Store_Key
Demo_Key
Product_Key
Product_Sales
Sale_Qty

Report
Report Attribute
Store_ID
Store Name
Store City
Store Zip
Promo_Sales
Promo_Coup

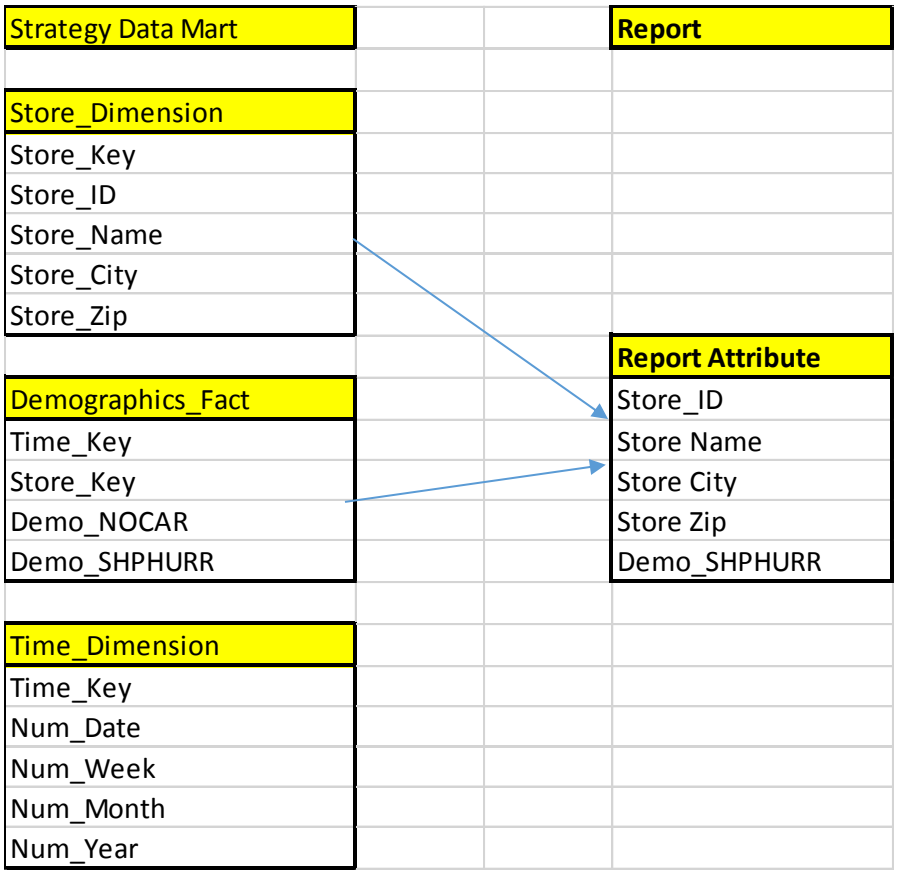
Question 2: Which stores are in areas of high single households and what are the commodities/products that sell more and which sell least?



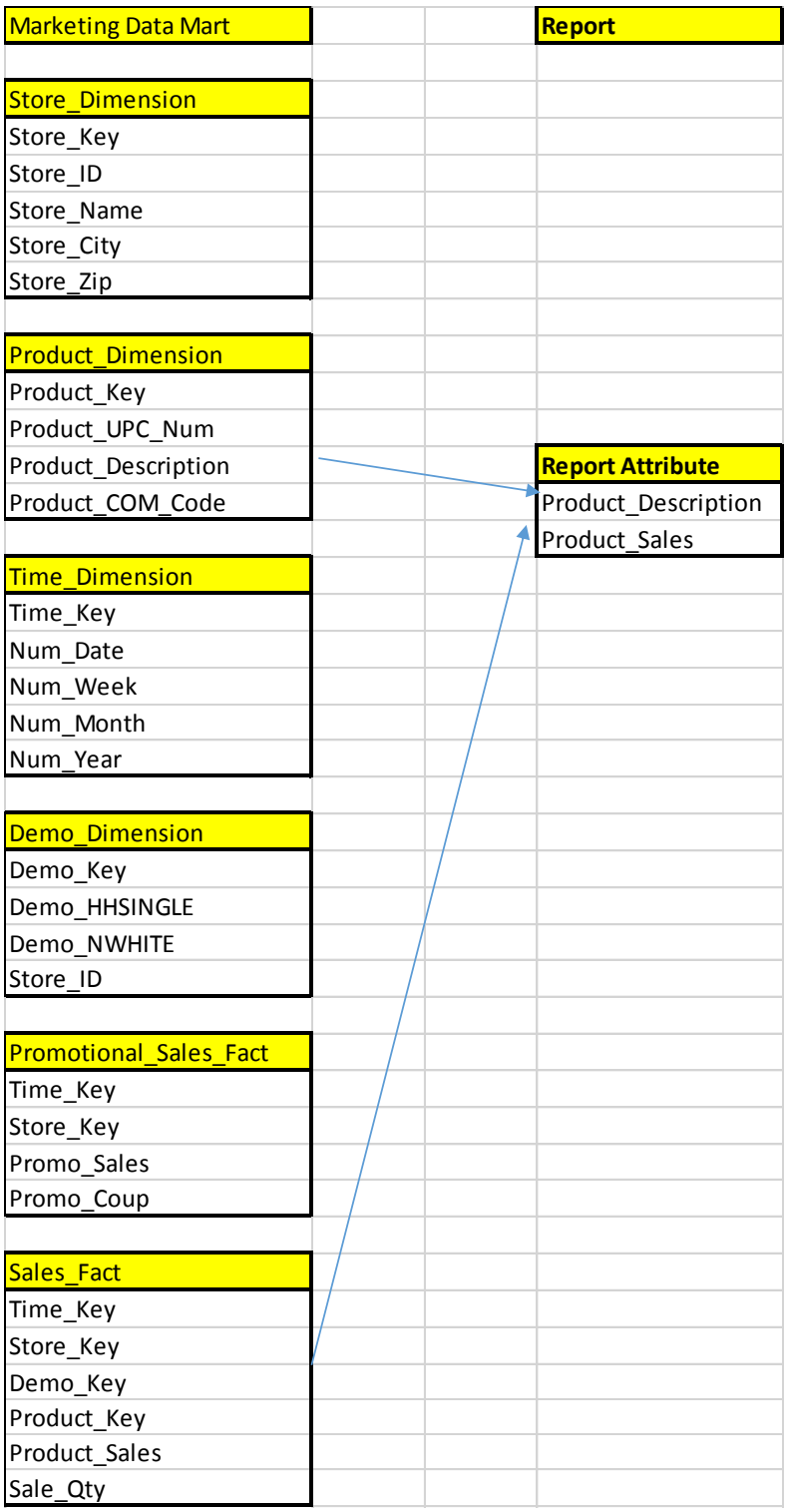
Question 3: Which stores or store areas have a high percentage of customers with no cars in order to set up a home delivery service?



Question 4: Which stores have a large number of hurried shoppers and would benefit from setting up a curbside pickup?



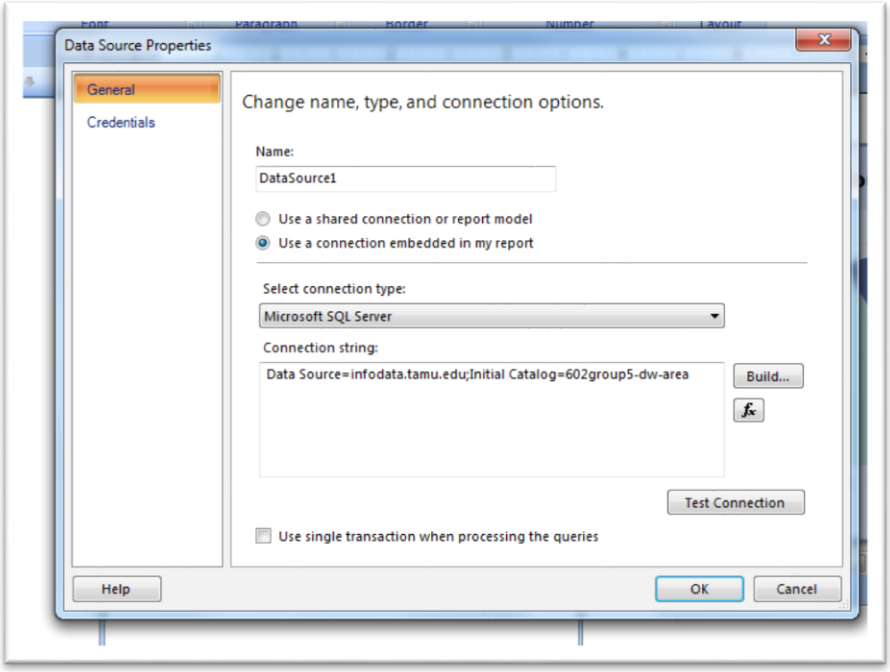
Question 5: How is the sales performance of all the beer brands across the stores in areas with more than 50% non-white population in the last 6 years (1991 - 1997)?



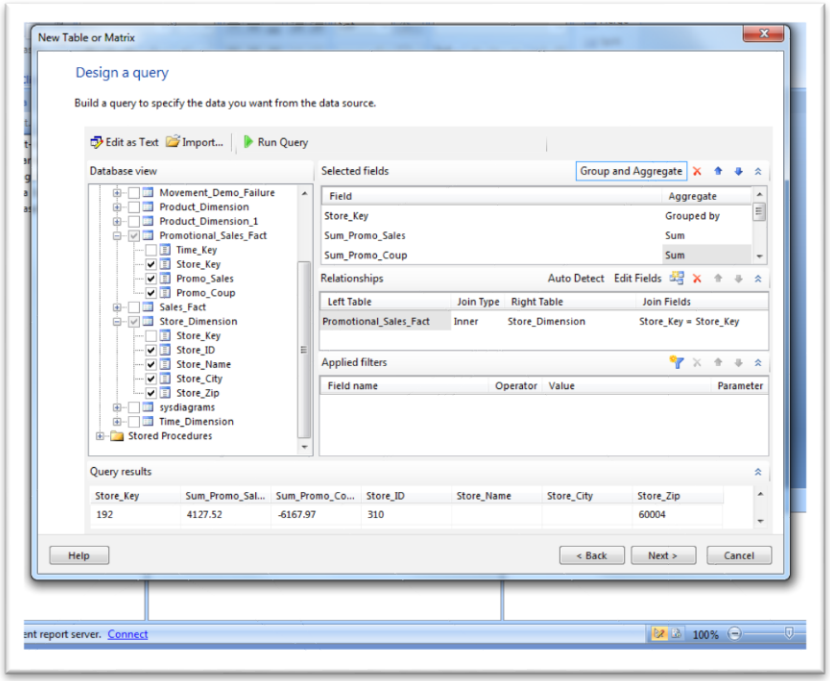
17. Report Building from Individual Data Marts

a. Question 1: How have the promo sales been affected across the stores on the basis of coupons?

Tools Used: *Report Builder 3.0*



DATA SOURCE CREATION

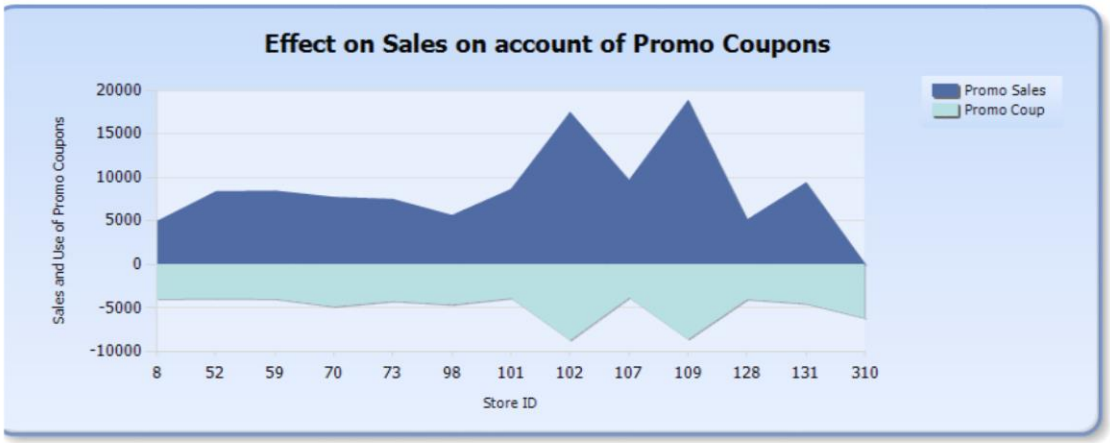


QUERY DESIGNER

/iews | Zoom | Navigation | Print | Export

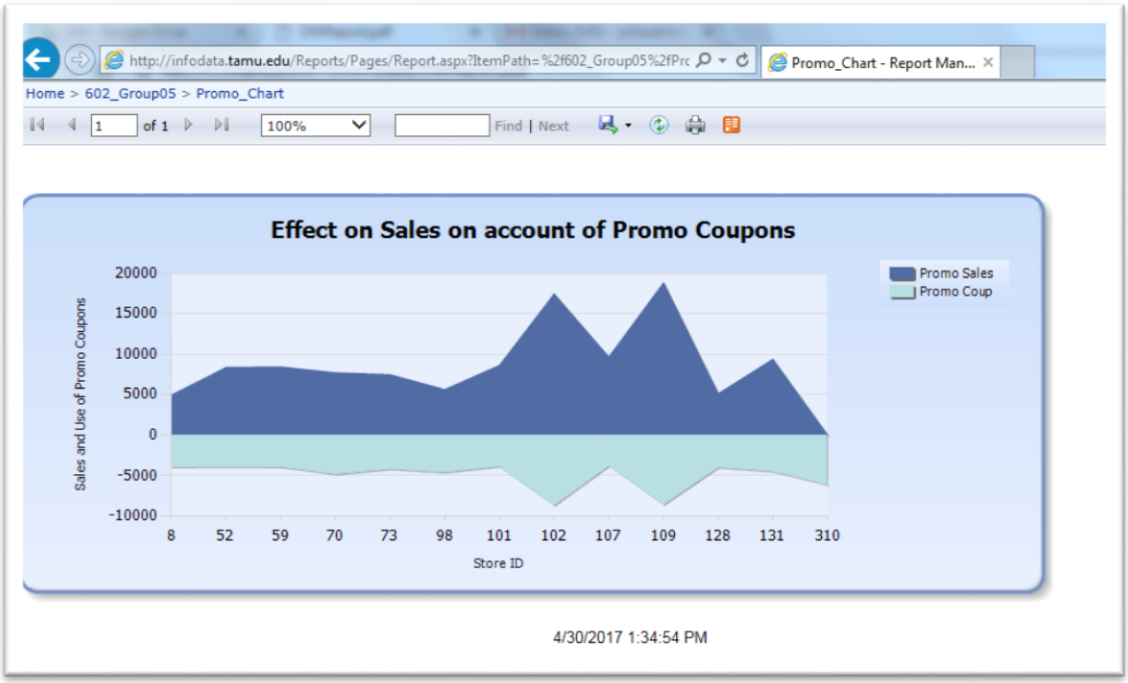
ore ID	Store Name	Store City	Store Zip	Sum Promo Coup	Sum Promo Sales
2	⊟ DOMINICKS 2	⊟ RIVER FOREST	60305	-52326.74	276310.46
			Total	-52326.74	276310.46
		Total		-52326.74	276310.46
	Total			-52326.74	276310.46
4	⊟ DOMINICKS 4	⊟ PARK RIDGE	60068	-29202.21	102554.44
			Total	-29202.21	102554.44
		Total		-29202.21	102554.44
	Total			-29202.21	102554.44
5	⊟ DOMINICKS 5	⊟ PALATINE	60067	-73566.09	333681.02
			Total	-73566.09	333681.02
		Total		-73566.09	333681.02
	Total			-73566.09	333681.02
8	Total			-88521.77	449427.27
9	Total			-68491.11	321876.98
12	Total			-48921.82	288967.04
14	Total			-55083.58	269525.29

REPORT DESIGN



4/23/2017 4:15:57 PM

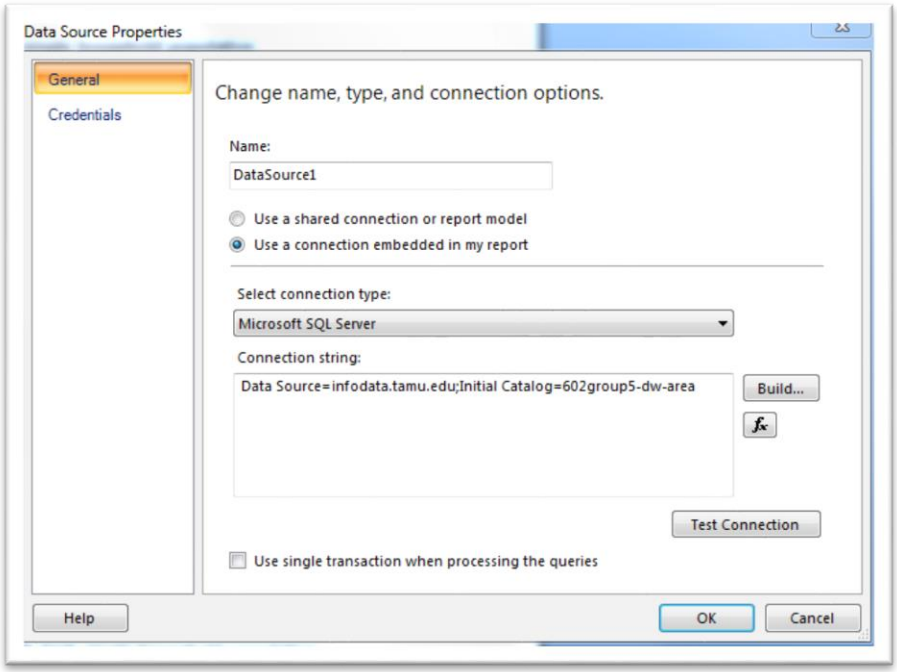
AREA CHART



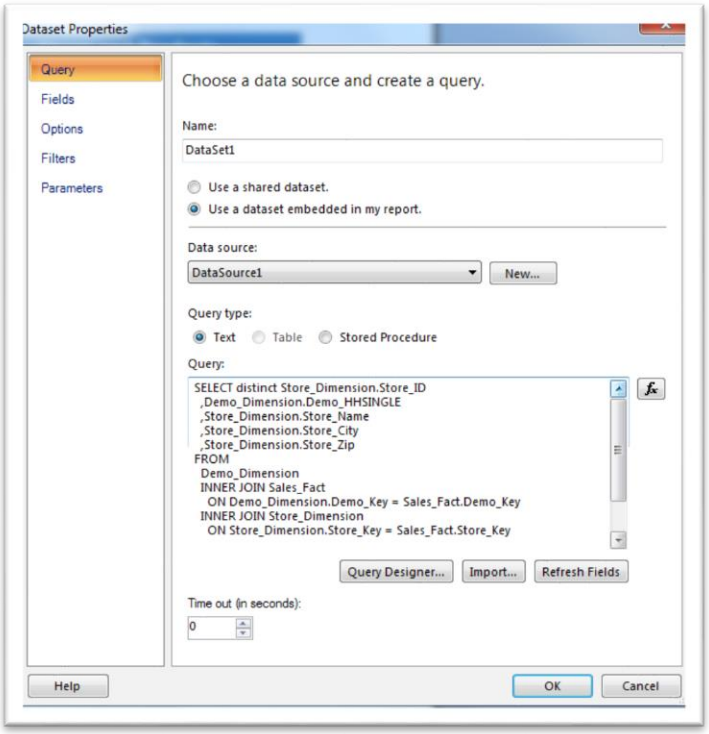
DEPLOYMENT OF REPORT ON INFODATA.TAMU.EDU

Conclusion: We find that the sales are positively affected by promotional coupons from the graphs above.

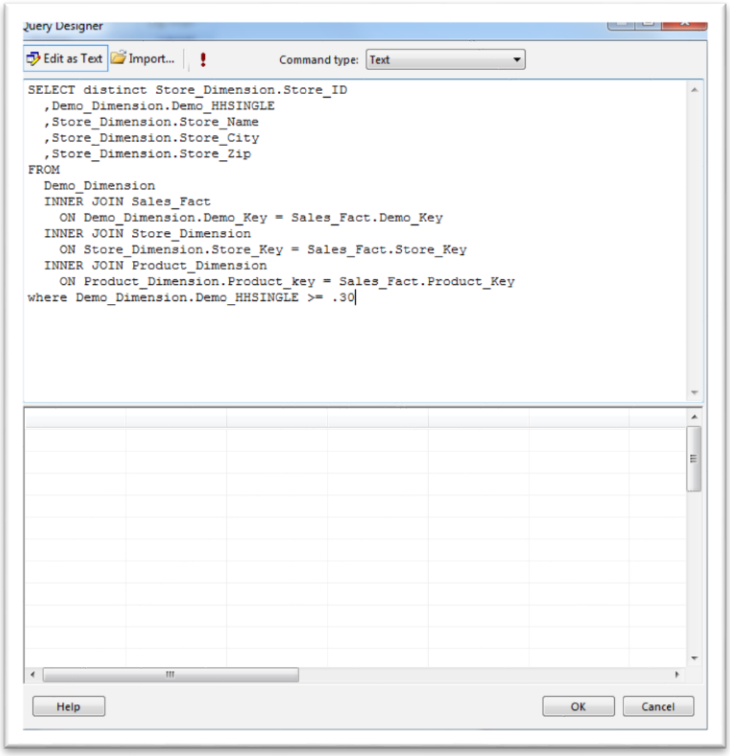
b. **Question 2: Which stores are in areas of high single households and what are the commodities/products that sell more and which sell least?**
Tools Used: Report Builder 3.0



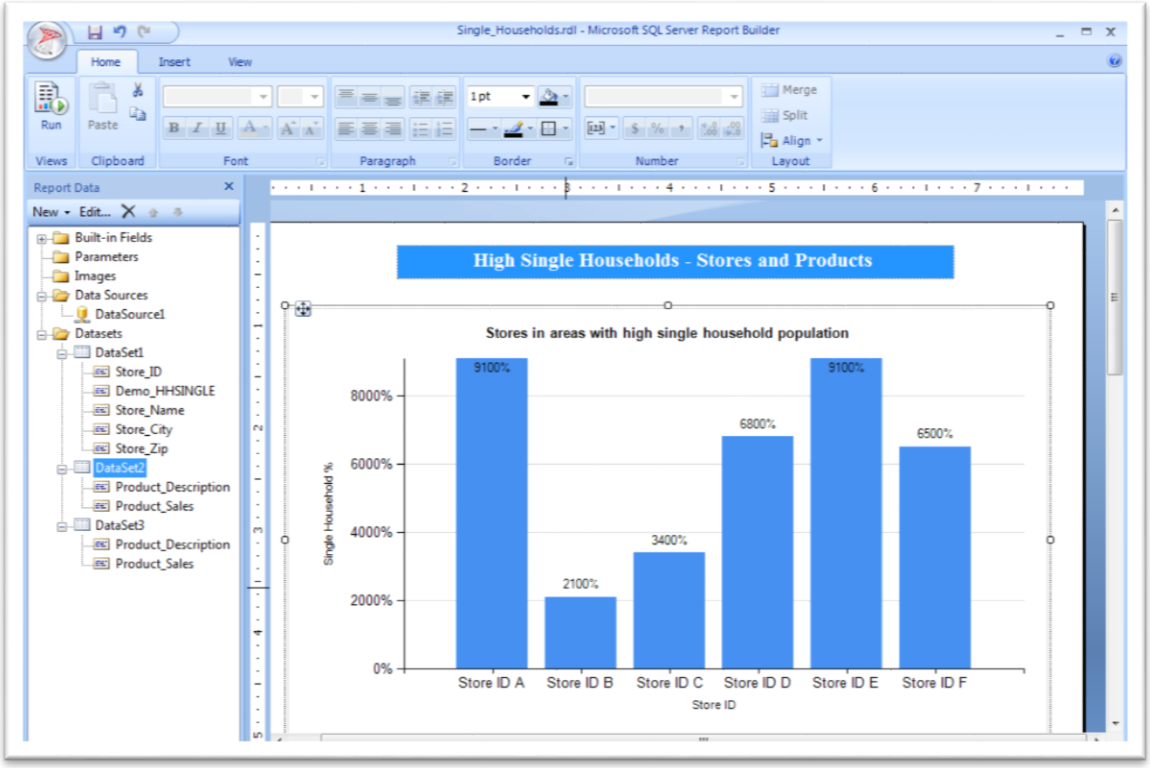
DATA SOURCE CREATION



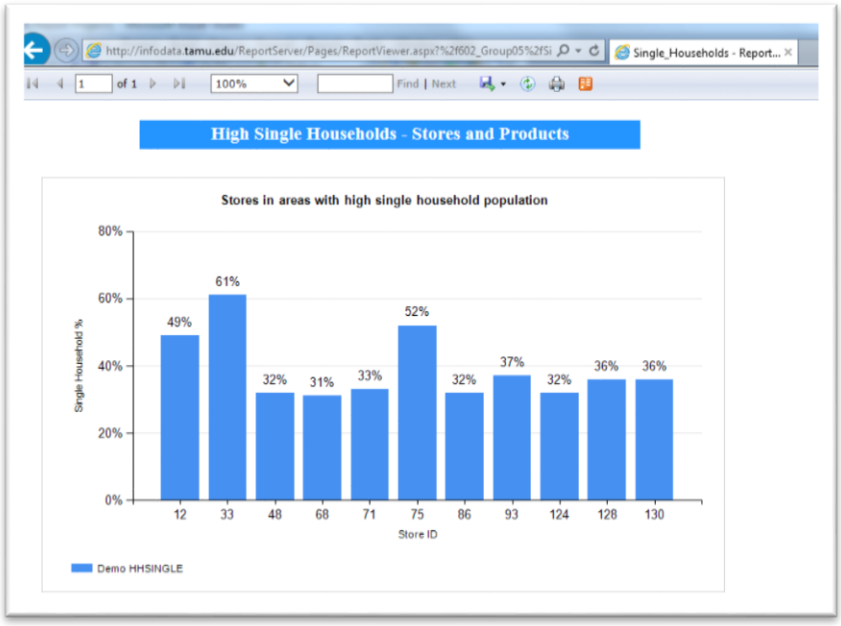
DATA SET CREATION



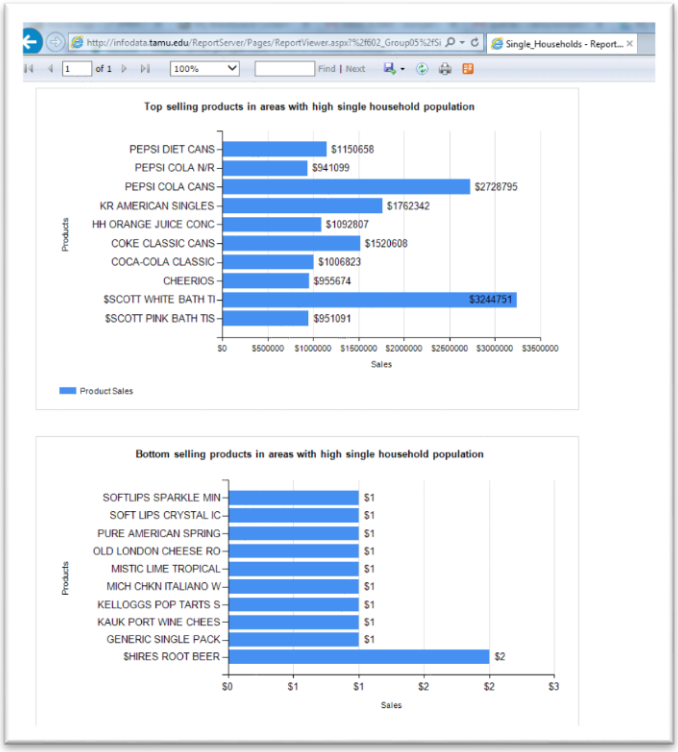
QUERY DESIGNER



DESIGNING THE REPORT TABLE



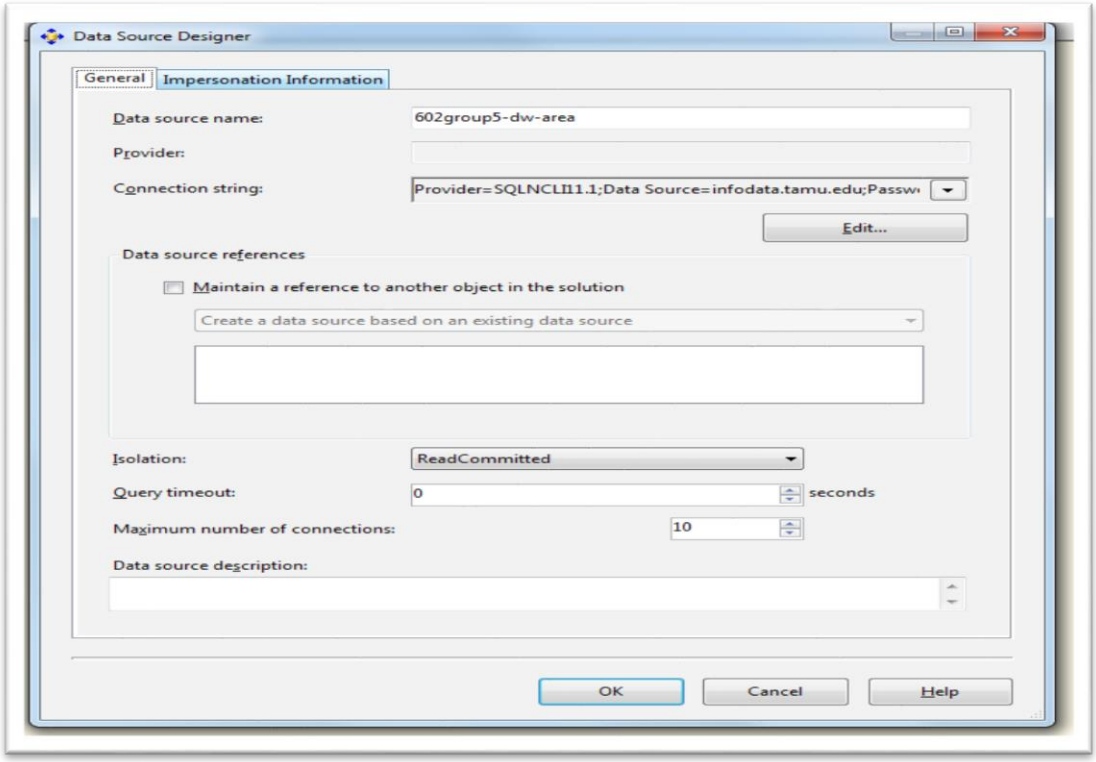
BAR CHART DEPLOYED REPORT FOR STORES IN HIGH (>30%) SINGLE HOUSEHOLD POPULATION AREAS



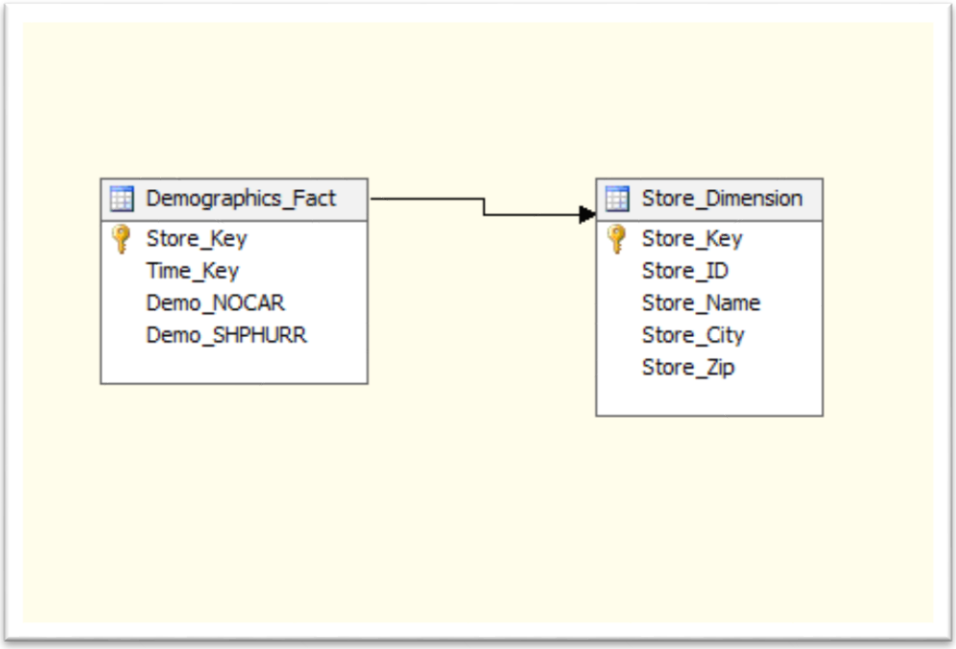
BAR CHART DEPLOYED REPORT FOR TOP AND BOTTOM SELLING PRODUCTS

Conclusion: The reports generated using Report builder give us a good perspective of the stores with high single household population and top and bottom most selling products across categories in those stores. This information enables us to make an informed decisions related to targeted marketing.

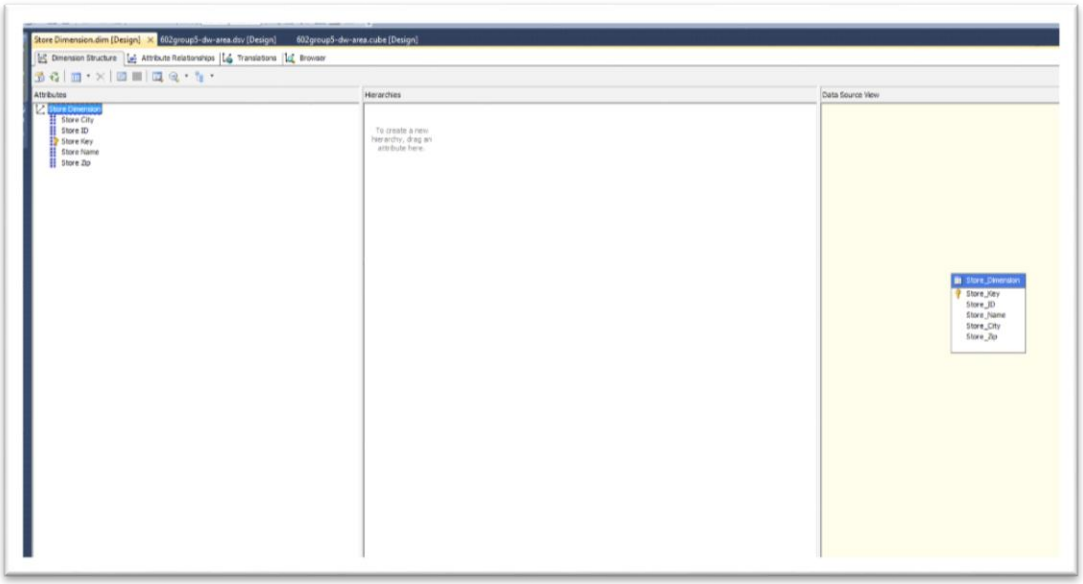
C. **Question 3: Which stores or store areas have a high percentage of customers with no cars in order to set up a home delivery service?**
Tools Used: Excel on SSAS



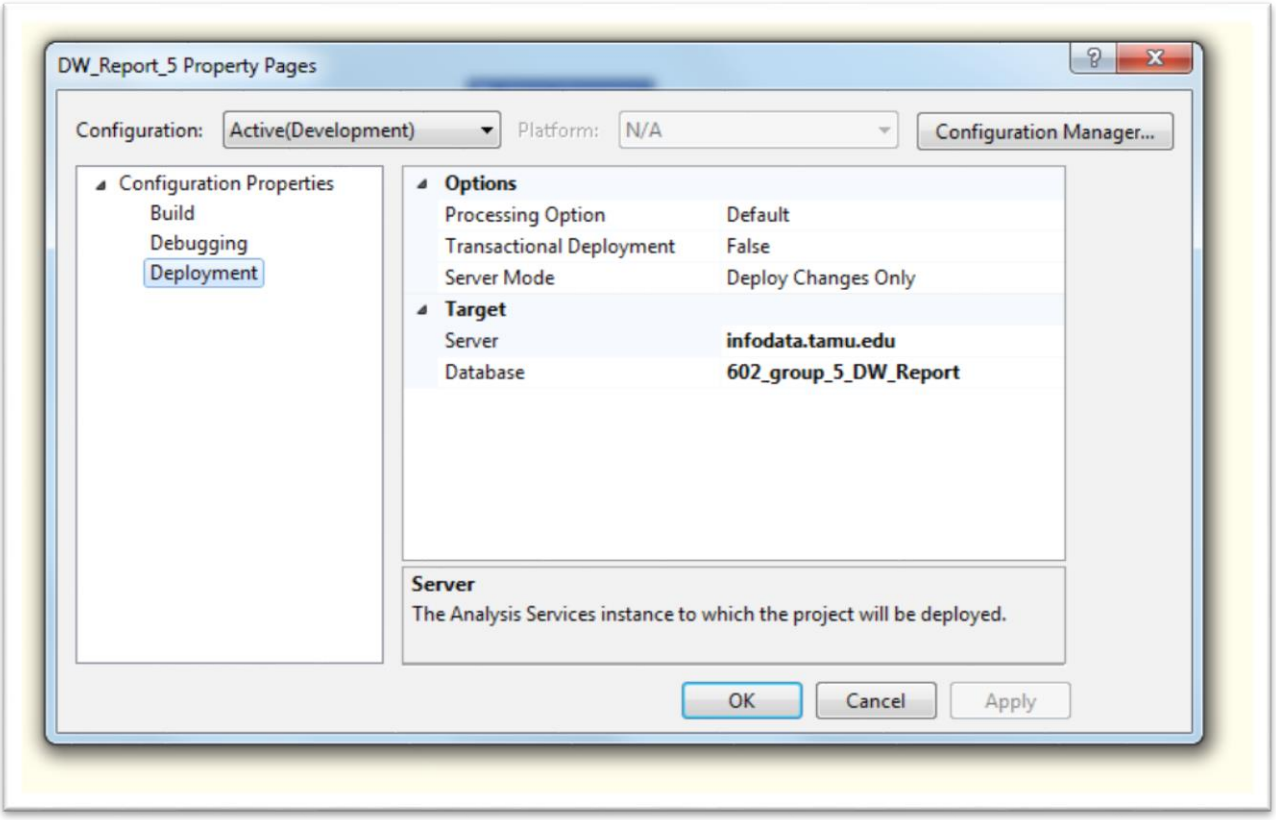
DATA SOURCE CREATION



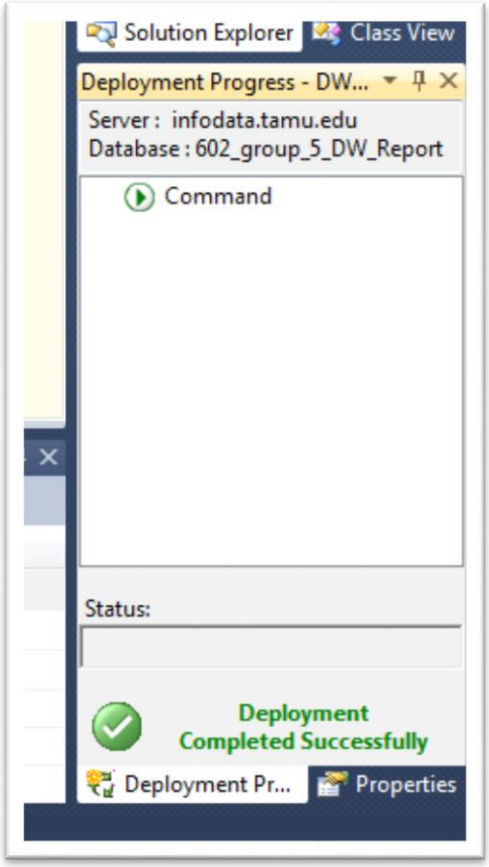
CUBE CREATION



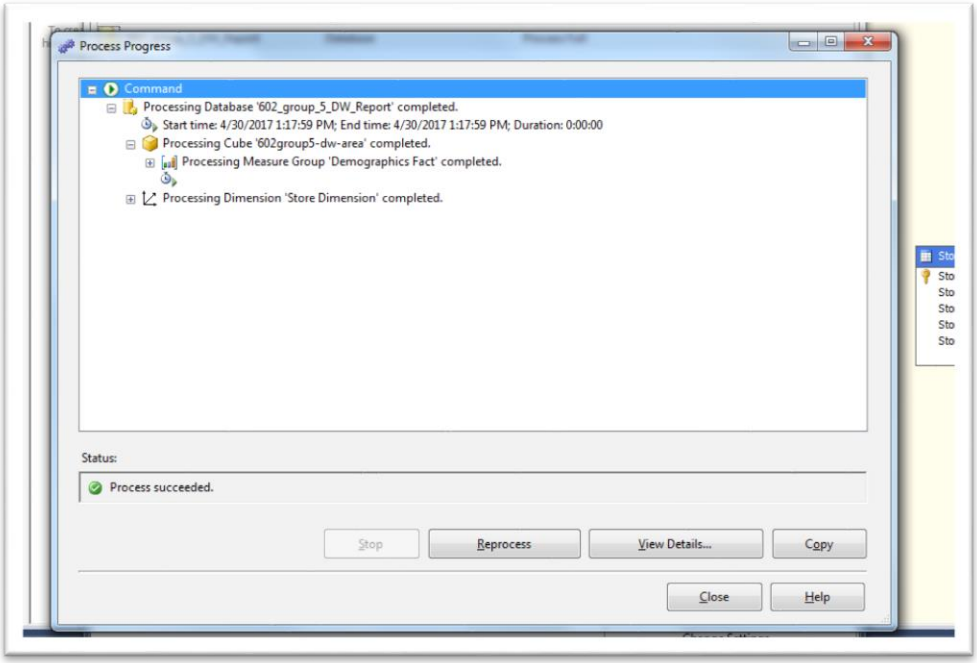
STORE DIMENSION



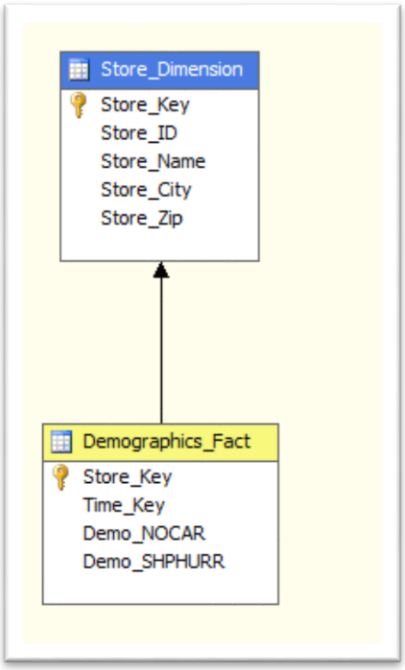
MULTIDIMENSIONAL DATABASE PROPERTIES



SUCCESSFUL DEPLOYMENT



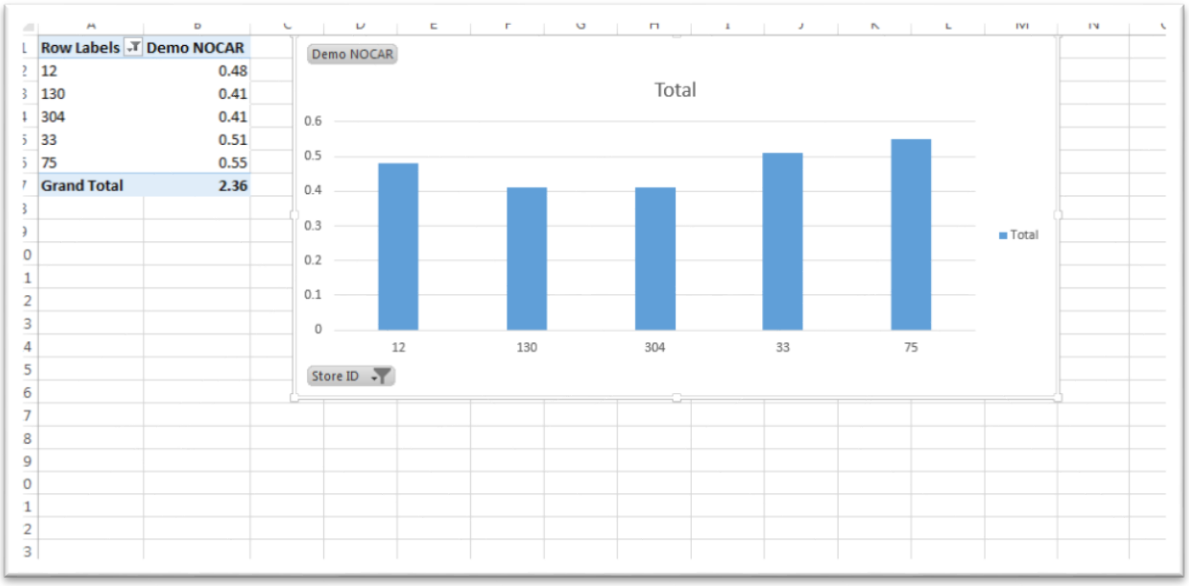
SUCCESSFUL PROCESS COMPLETION



CUBE STRUCTURE

Store ID	Store City	Store Name	Demo NOCAR	Demo SHPHURR
100	CHICAGO	DOMINICK...	0.37	0.09
101	DES PLAI...	DOMINICK...	0.06	0.13
102	MERRIO...	DOMINICK...	0.11	0.12
103	BOLINGB...	DOMINICK...	0.02	0.22
104	ST CHAR...	DOMINICK...	0.03	0.21
105	MELROS...	DOMINICK...	0.09	0.15
106	MONTGO...	DOMINICK...	0.06	0.18
107	WESTCH...	DOMINICK...	0.04	0.16
109	BANNOC...	DOMINICK...	0.03	0.28
110	EAST DU...	DOMINICK...	0.06	0.19
111	CHICAGO	DOMINICK...	0.33	0.05
112	BUFFALO...	DOMINICK...	0.01	0.21
113	CHICAGO	DOMINICK...	0.14	0.12
114	CALUME...	DOMINICK...	0.09	0.09
115	NAPERVI...	DOMINICK...	0.02	0.29
116	ELMHURST	DOMINICK...	0.06	0.18
117	SCHAUM...	DOMINICK...	0.02	0.18
118	MORTON...	DOMINICK...	0.08	0.14
119	BUFFALO...	DOMINICK...	0.02	0.18
12	CHICAGO	DOMINICK...	0.48	0.05
121	WILLOW...	DOMINICK...	0.02	0.17
122	HOFFMA...	DOMINICK...	0.02	0.21
123	CHICAGO	DOMINICK...	0.24	0.11

CUBE BROWSER

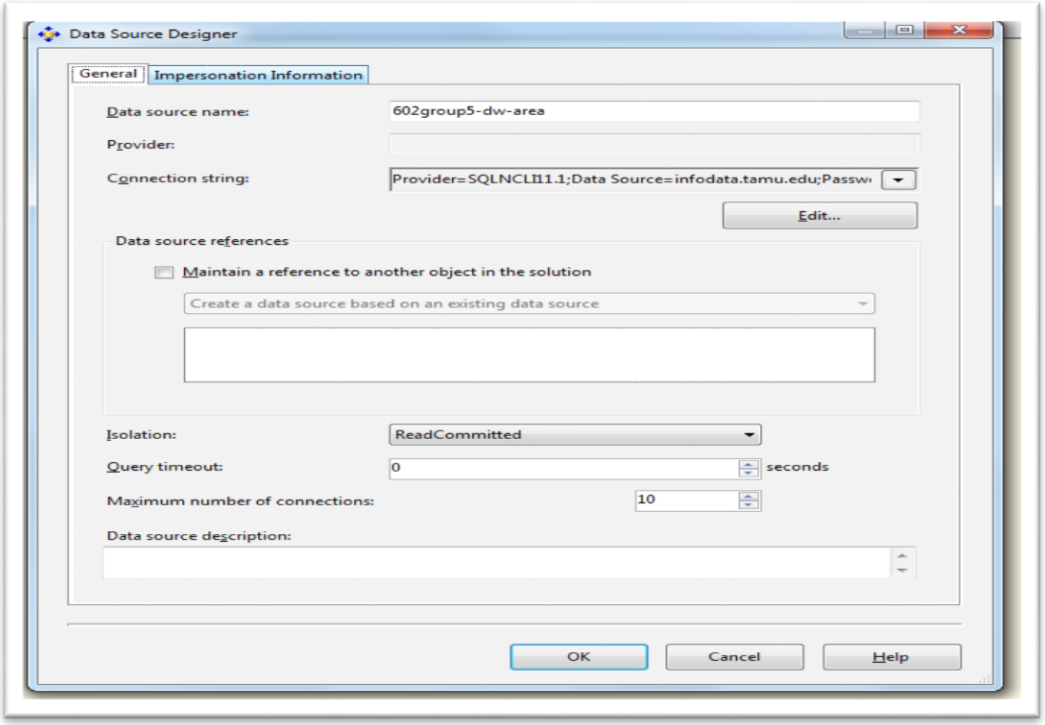


EXCEL: ANALYSIS USING PIVOT CHART

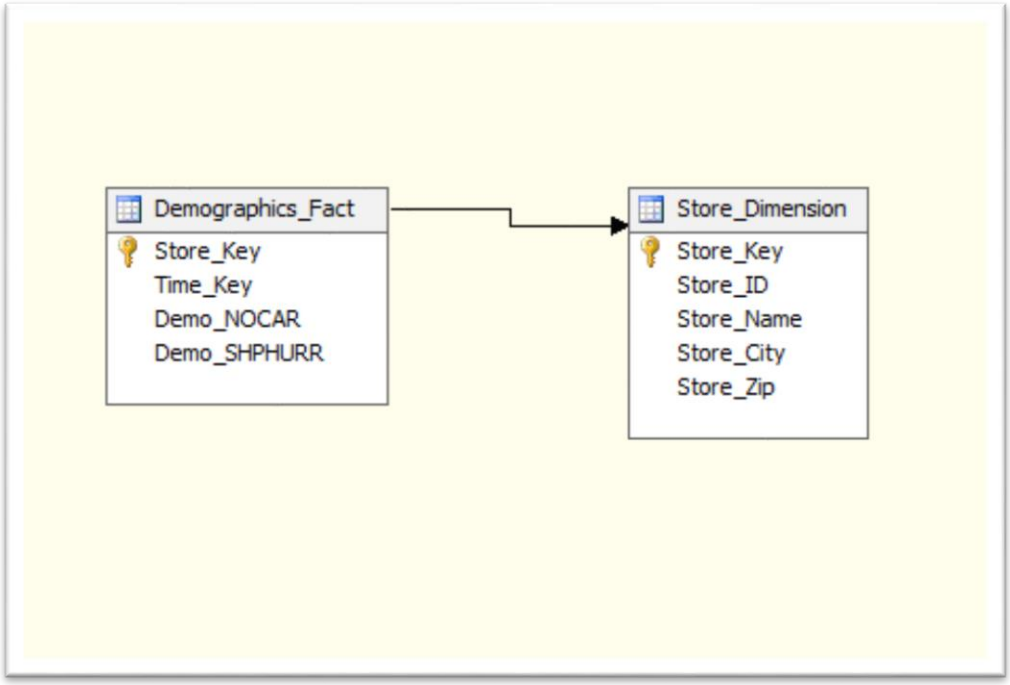
Conclusion: The pivot chart enables us to isolate stores with high percentage of people that come in without cars. This information makes us think about the possibility of setting up a home delivery service to better cater to the customers.

d. **Question 4: Which stores have a large number of hurried shoppers and would benefit from setting up a curbside pickup?**

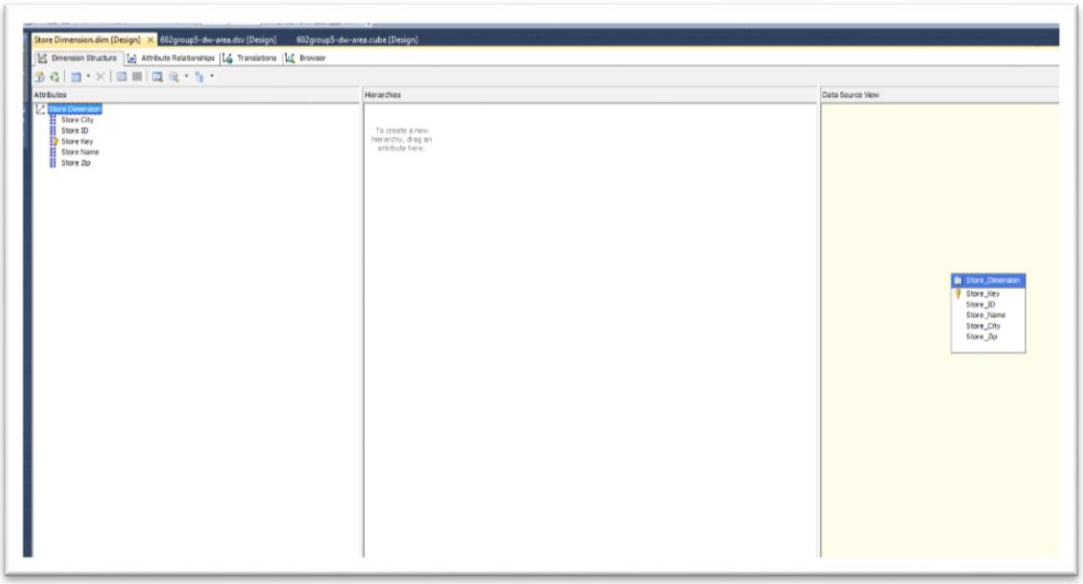
Tools Used:



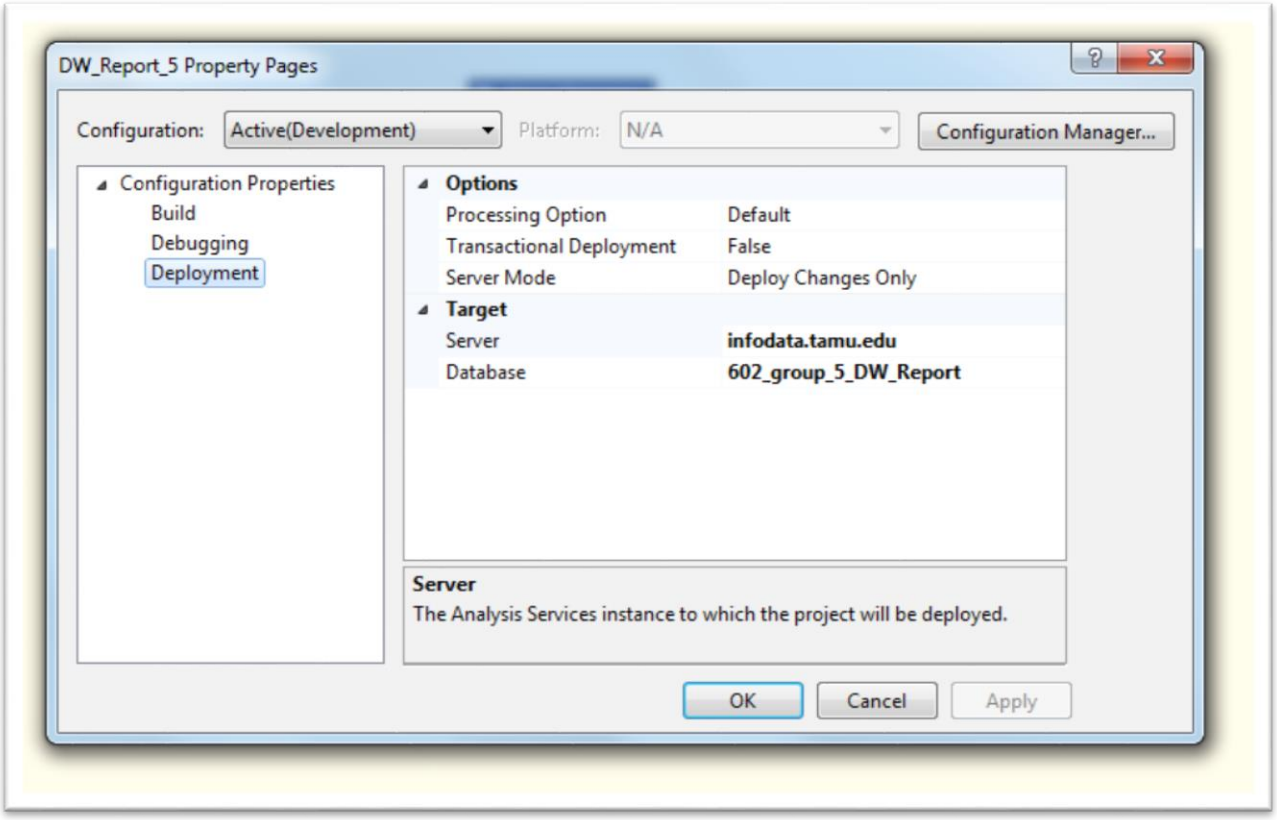
DATA SOURCE CREATION



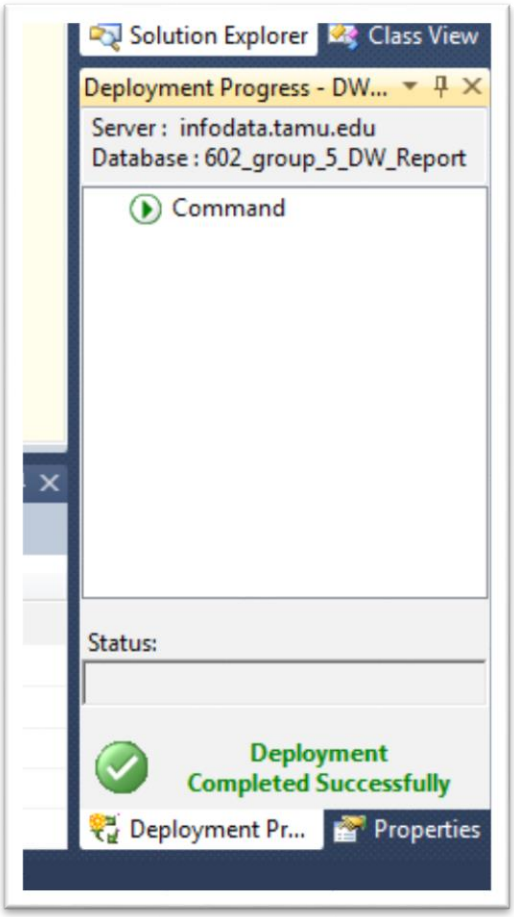
CUBE CREATION



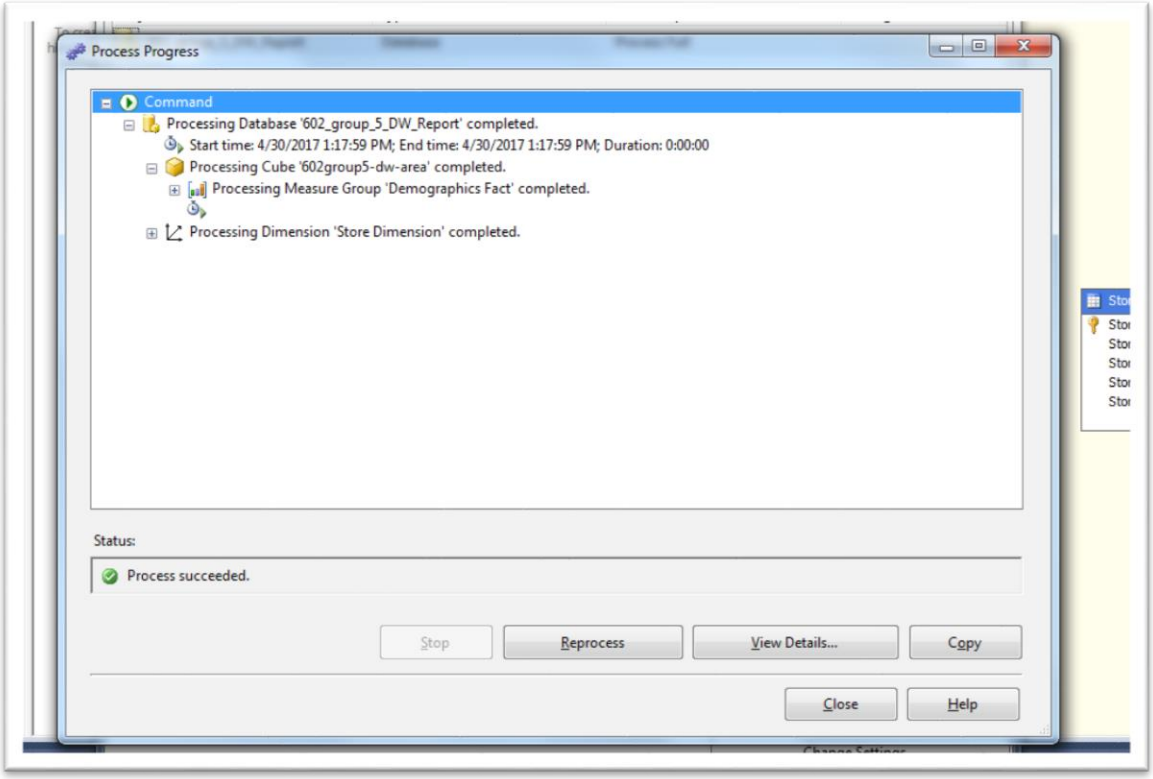
STORE DIMENSION



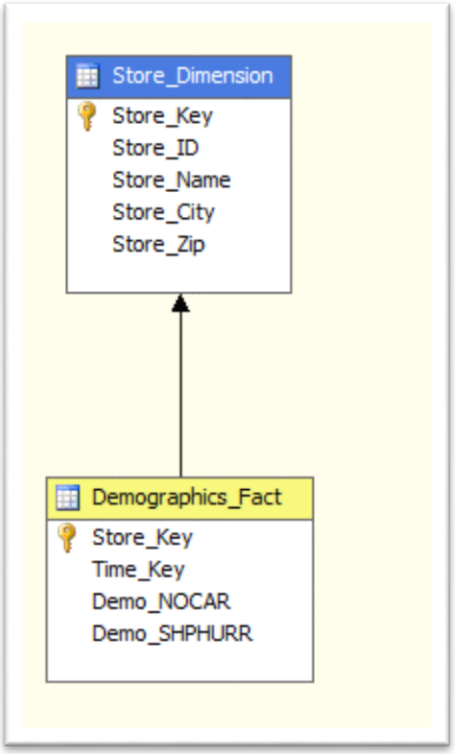
MULTIDIMENSIONAL DATABASE PROPERTY



SUCCESSFUL DEPLOYMENT



SUCCESSFUL PROCESS

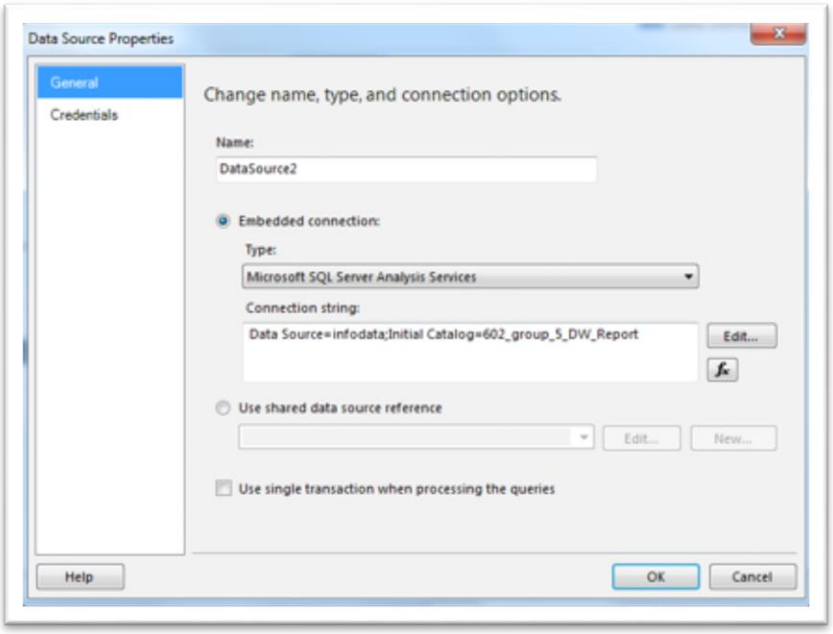


CUBE STRUCTURE

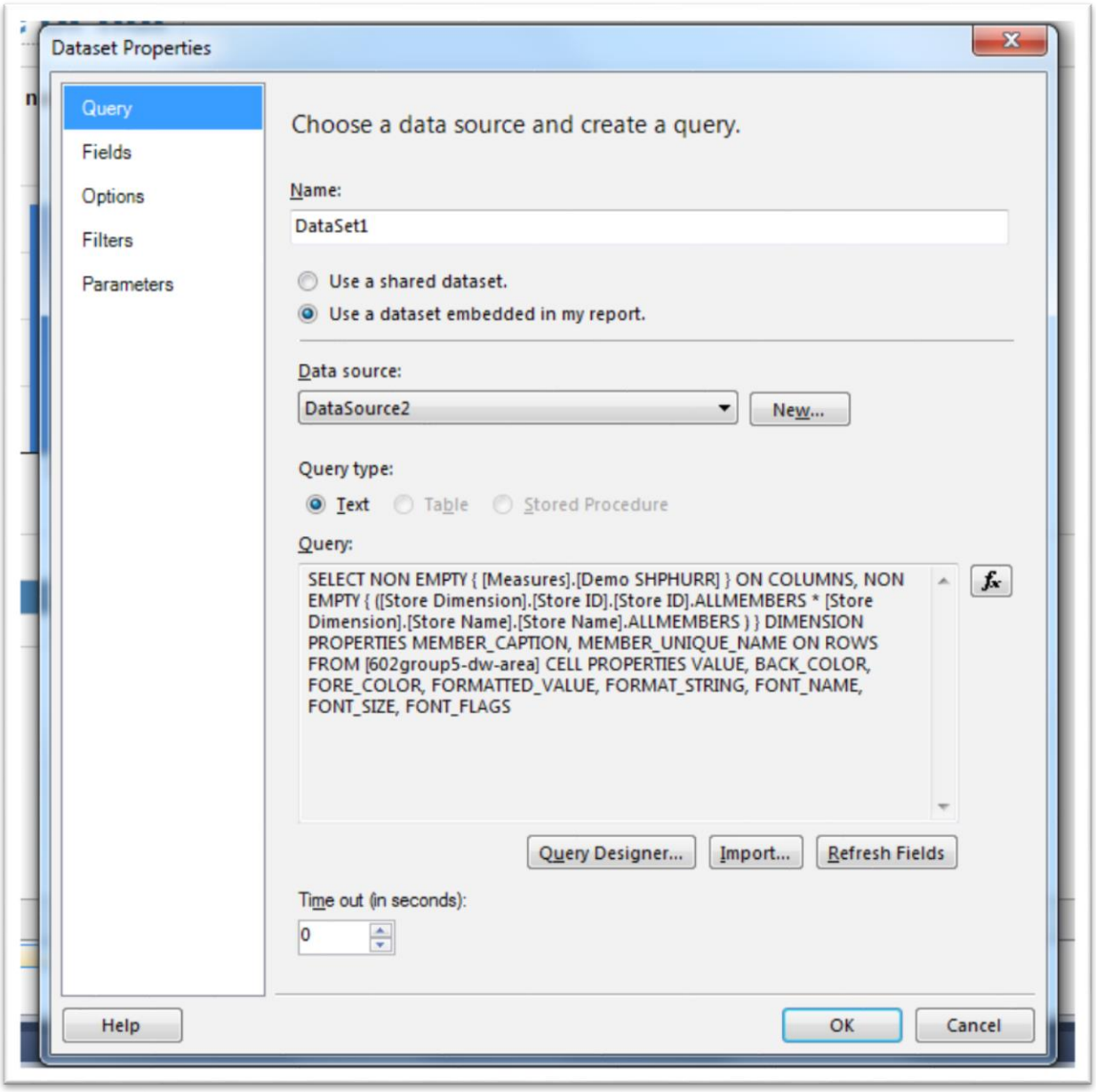
Store ID	Store City	Store Name	Demo NOCAR	Demo SHPHURR
100	CHICAGO	DOMINICK...	0.37	0.09
101	DES PLAI...	DOMINICK...	0.06	0.13
102	MERRIO...	DOMINICK...	0.11	0.12
103	BOLINGB...	DOMINICK...	0.02	0.22
104	ST CHAR...	DOMINICK...	0.03	0.21
105	MELROS...	DOMINICK...	0.09	0.15
106	MONTGO...	DOMINICK...	0.06	0.18
107	WESTCH...	DOMINICK...	0.04	0.16
109	BANNOC...	DOMINICK...	0.03	0.28
110	EAST DU...	DOMINICK...	0.06	0.19
111	CHICAGO	DOMINICK...	0.33	0.05
112	BUFFALO...	DOMINICK...	0.01	0.21
113	CHICAGO	DOMINICK...	0.14	0.12
114	CALUME...	DOMINICK...	0.09	0.09
115	NAPERVI...	DOMINICK...	0.02	0.29
116	ELMHURST	DOMINICK...	0.06	0.18
117	SCHAUM...	DOMINICK...	0.02	0.18
118	MORTON...	DOMINICK...	0.08	0.14
119	BUFFALO...	DOMINICK...	0.02	0.18
12	CHICAGO	DOMINICK...	0.48	0.05
121	WILLOW...	DOMINICK...	0.02	0.17
122	HOFFMA...	DOMINICK...	0.02	0.21
123	CHICAGO	DOMINICK...	0.24	0.11
124	OAK PARK	DOMINICK...	0.29	0.09

CUBE BROWSER

SSRS Report Creation



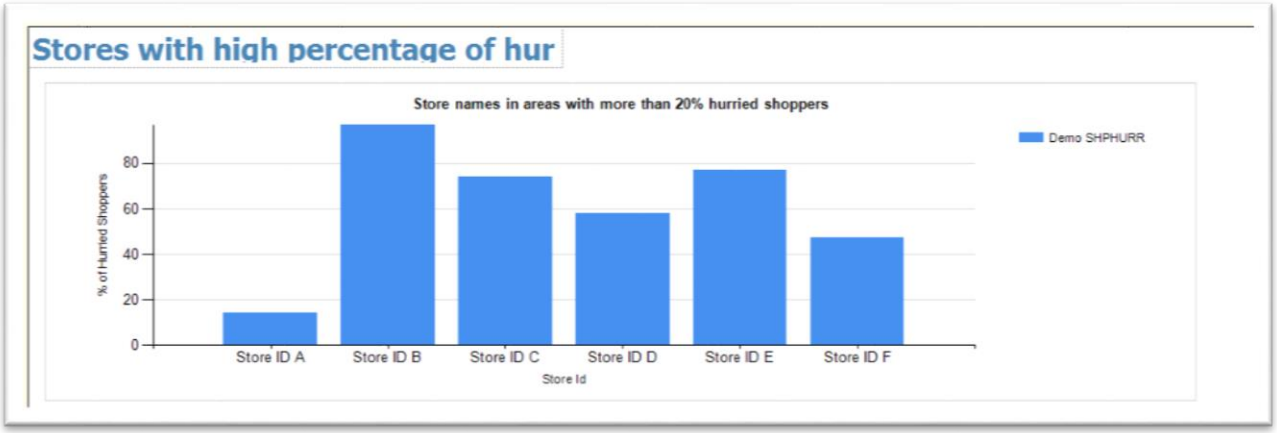
SSRS DATA SOURCE



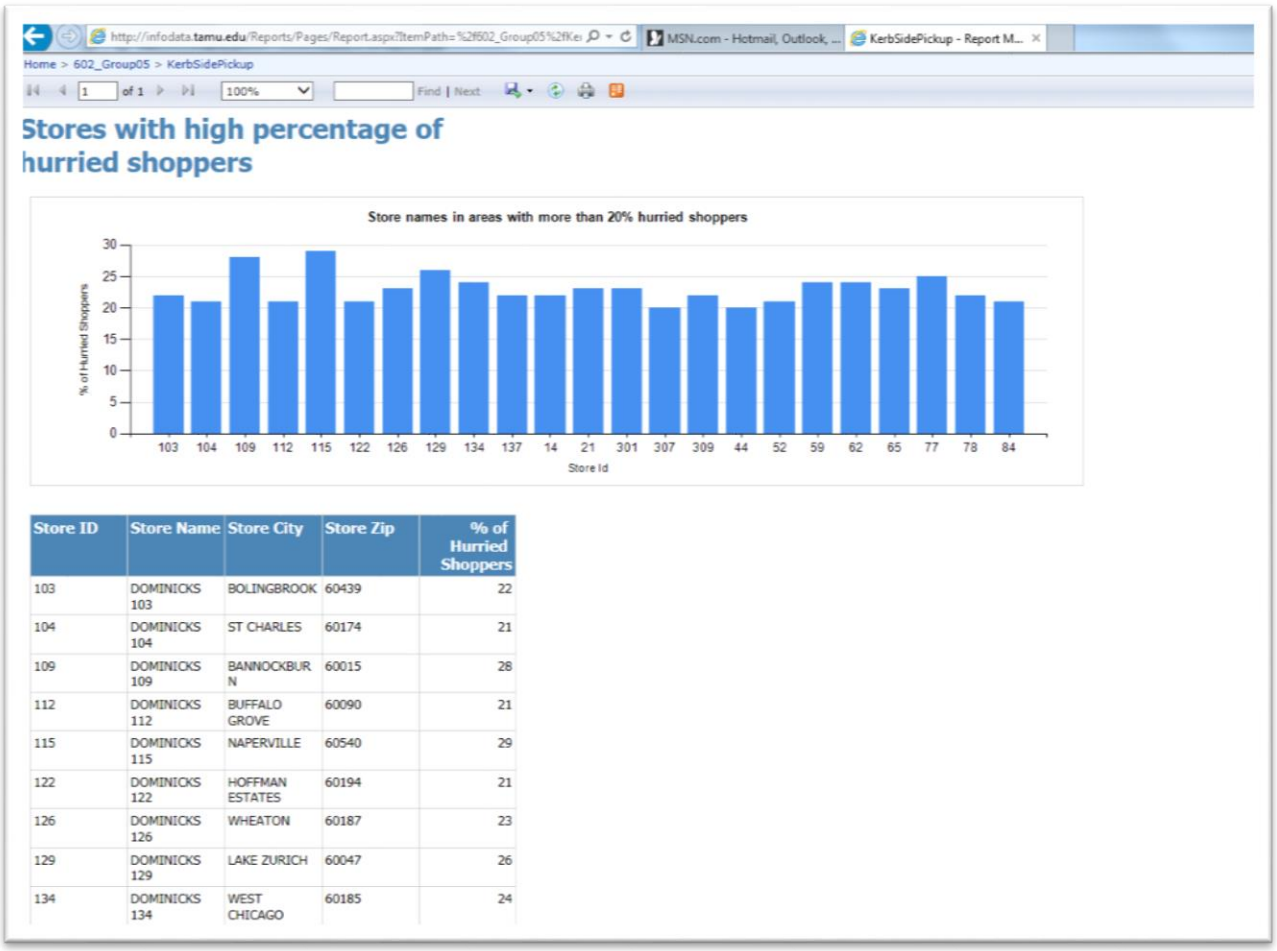
SSRS QUERY

Store ID	Store Name	Store City	Store Zip	% of Hurrie
[Store_ID]	[Store_Name]	[Store_City]	[Store_Zip]	«Expr»

TABLE DESIGN



BAR CHART DESIGN

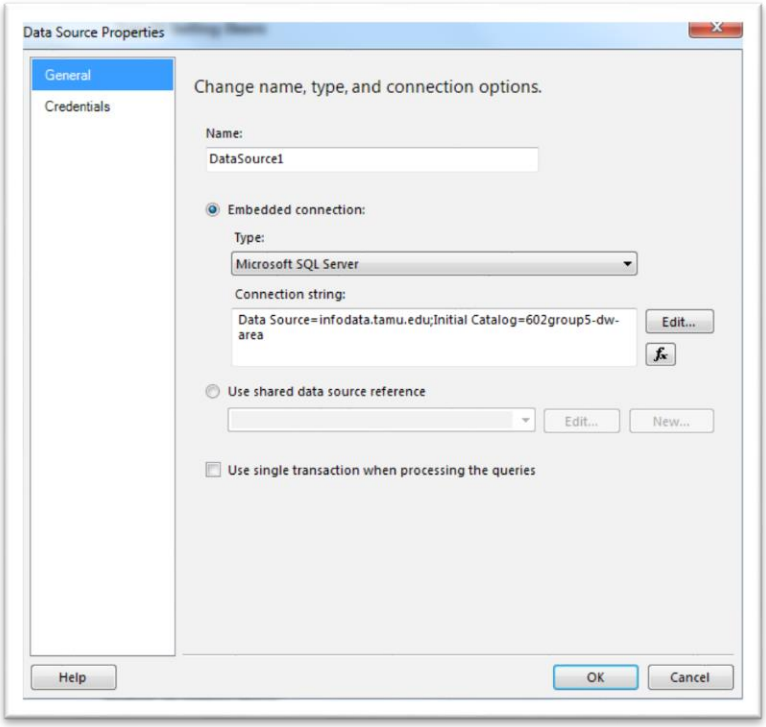


Store ID	Store Name	Store City	Store Zip	% of Hurried Shoppers
103	DOMINICKS 103	BOLINGBROOK	60439	22
104	DOMINICKS 104	ST CHARLES	60174	21
109	DOMINICKS 109	BANNOCKBURN	60015	28
112	DOMINICKS 112	BUFFALO GROVE	60090	21
115	DOMINICKS 115	NAPERVILLE	60540	29
122	DOMINICKS 122	HOFFMAN ESTATES	60194	21
126	DOMINICKS 126	WHEATON	60187	23
129	DOMINICKS 129	LAKE ZURICH	60047	26
134	DOMINICKS 134	WEST CHICAGO	60185	24

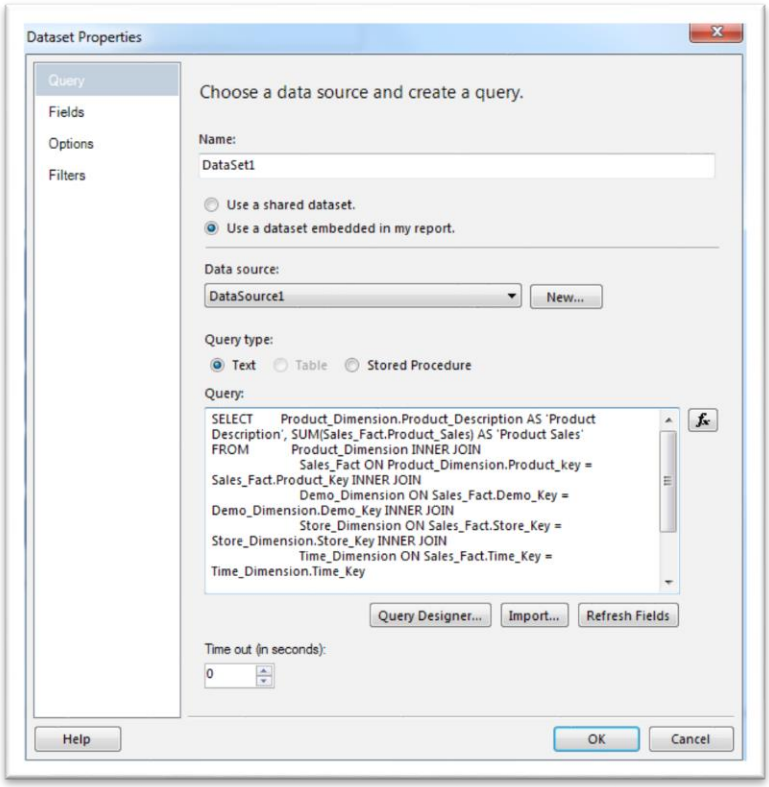
DEPLOYMENT ON IE

Conclusion: The reports generated on SSRS give us a good perspective of the number of hurried shoppers in certain stores. This information enables us to make an informed decision on whether implementing curb side pickup is worth it or not.

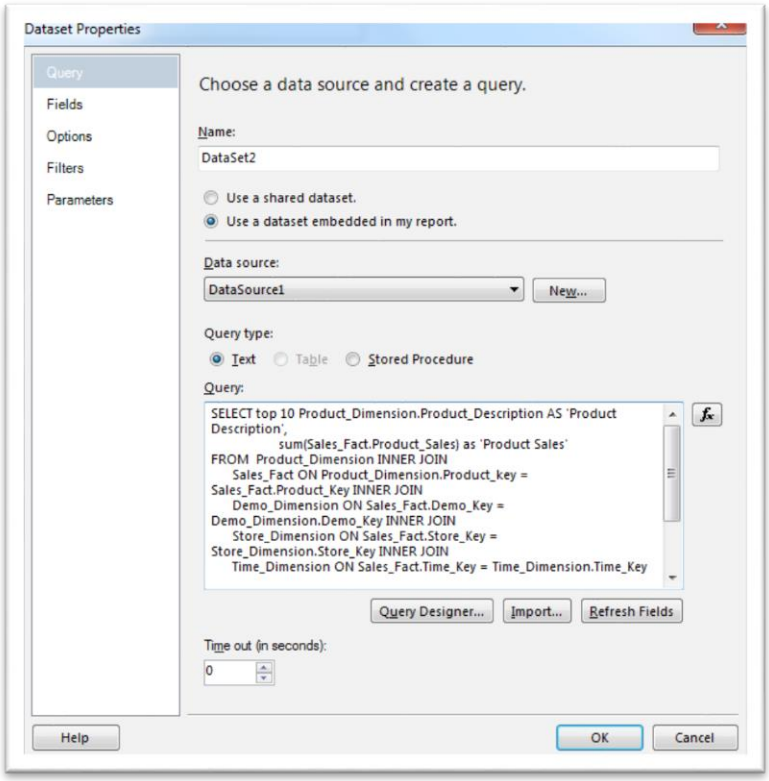
e. **Question 5: How is the sales performance of all the beer brands across the stores in areas with more than 50% non-white population in the last 6 years (1991 - 1997)?**
Tools Used: SSRS Only



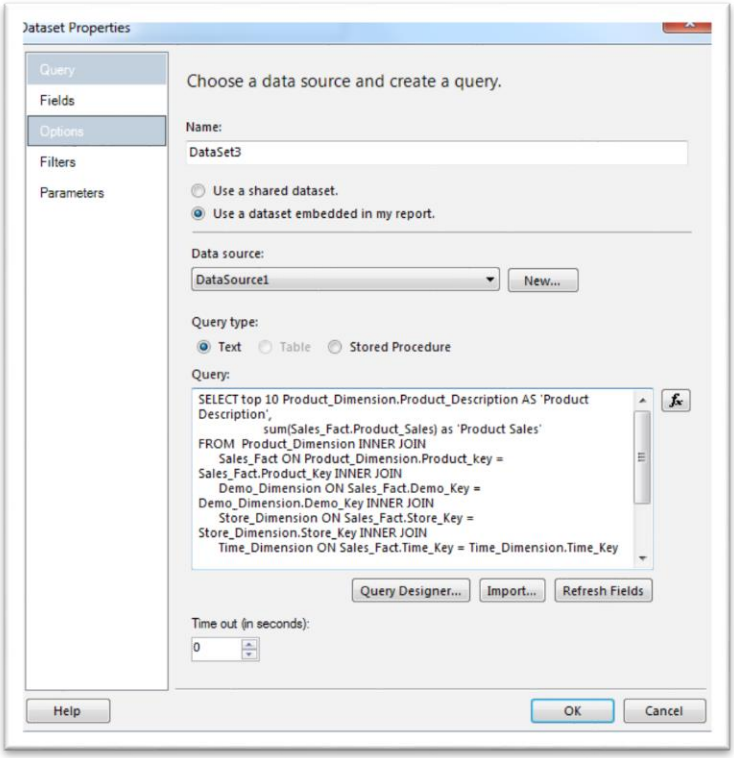
DATA SOURCE CREATION



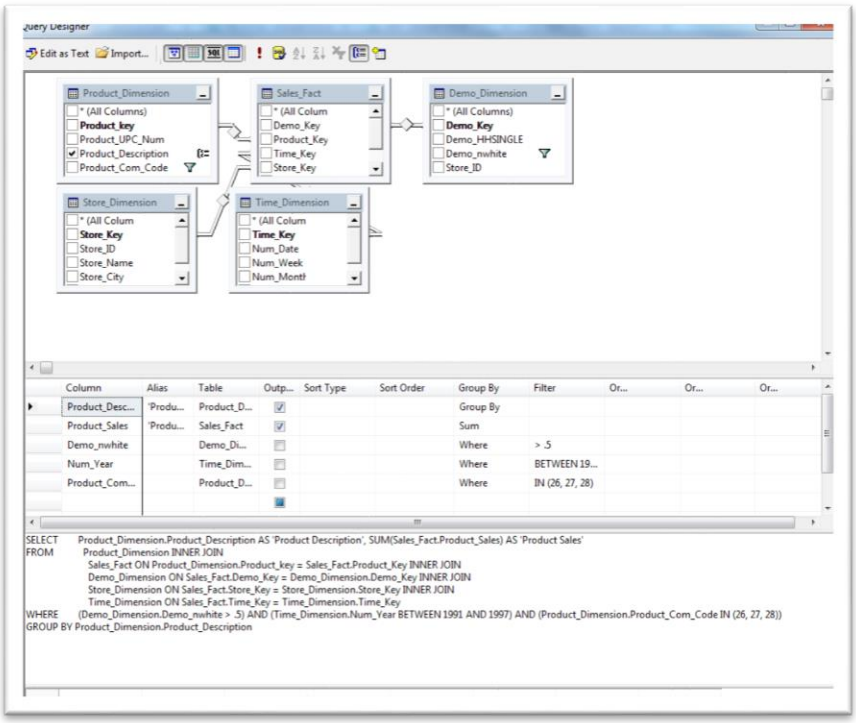
DATA SET FOR TOTAL BEER SALES CREATION



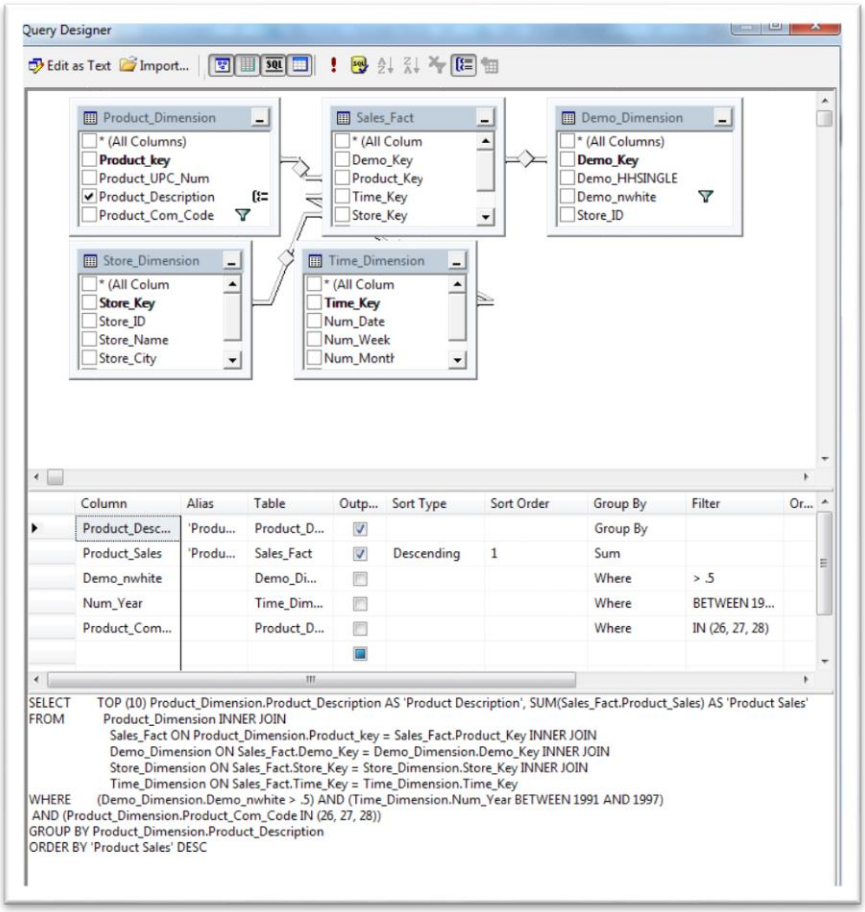
DATA SET FOR SALES OF TOP 10 BEER BRANDS CREATION



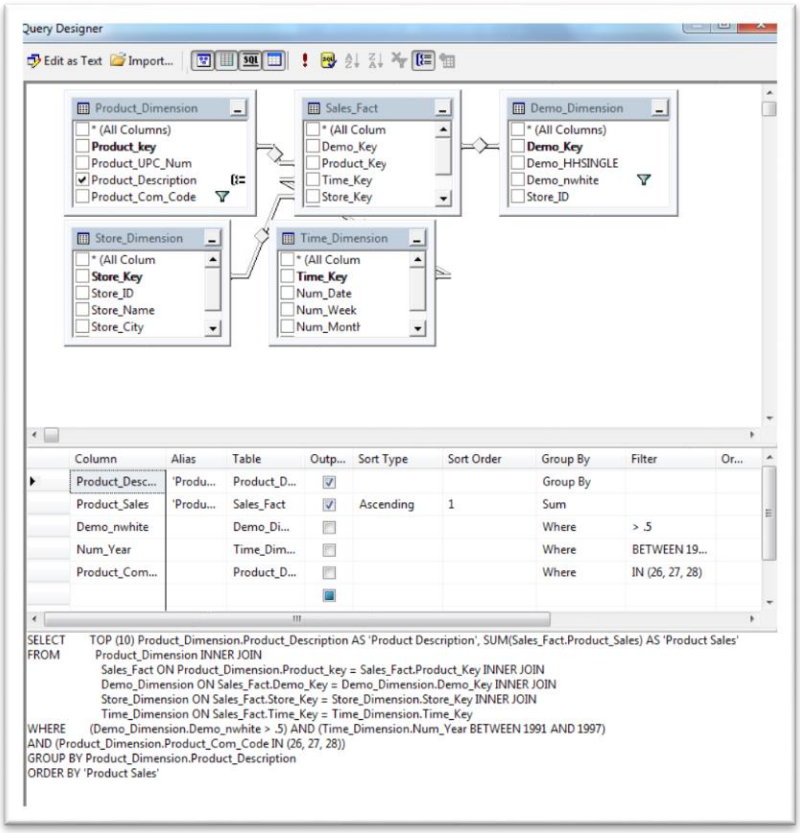
DATA SET FOR SALES OF BOTTOM 10 BEER BRANDS CREATION



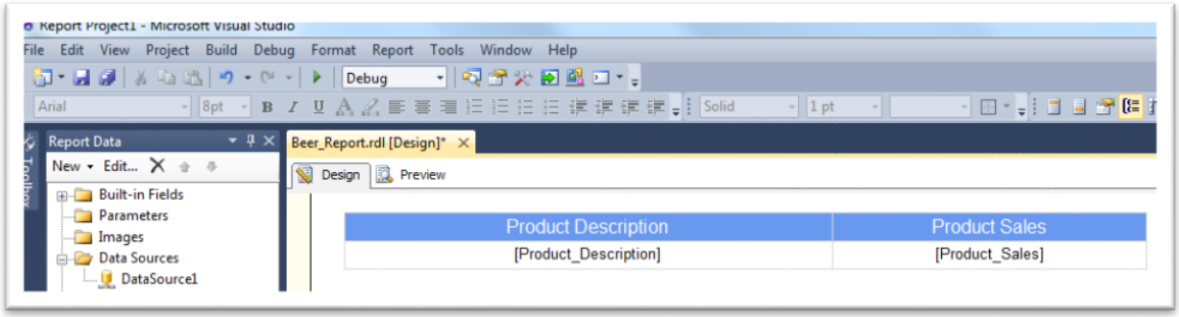
QUERY DESIGNER FOR TOTAL BEER SALES CREATION



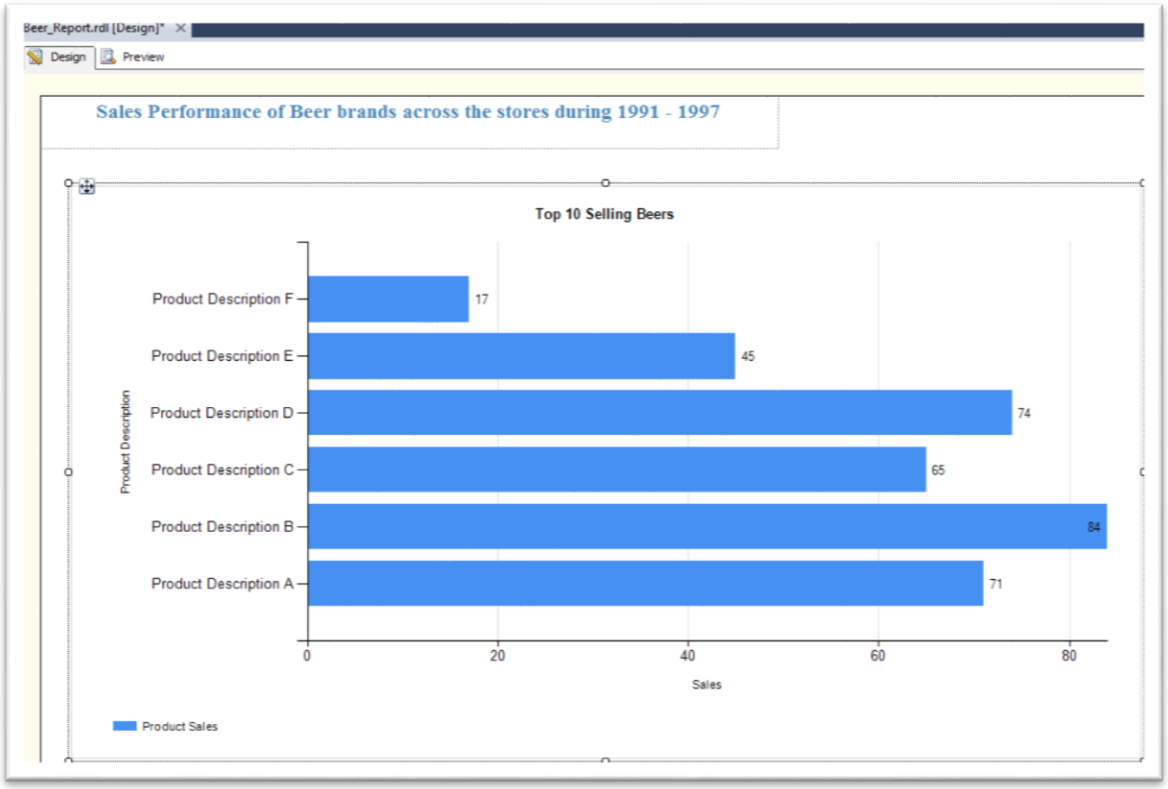
QUERY DESIGNER FOR SALES OF TOP 10 BEER BRANDS CREATION



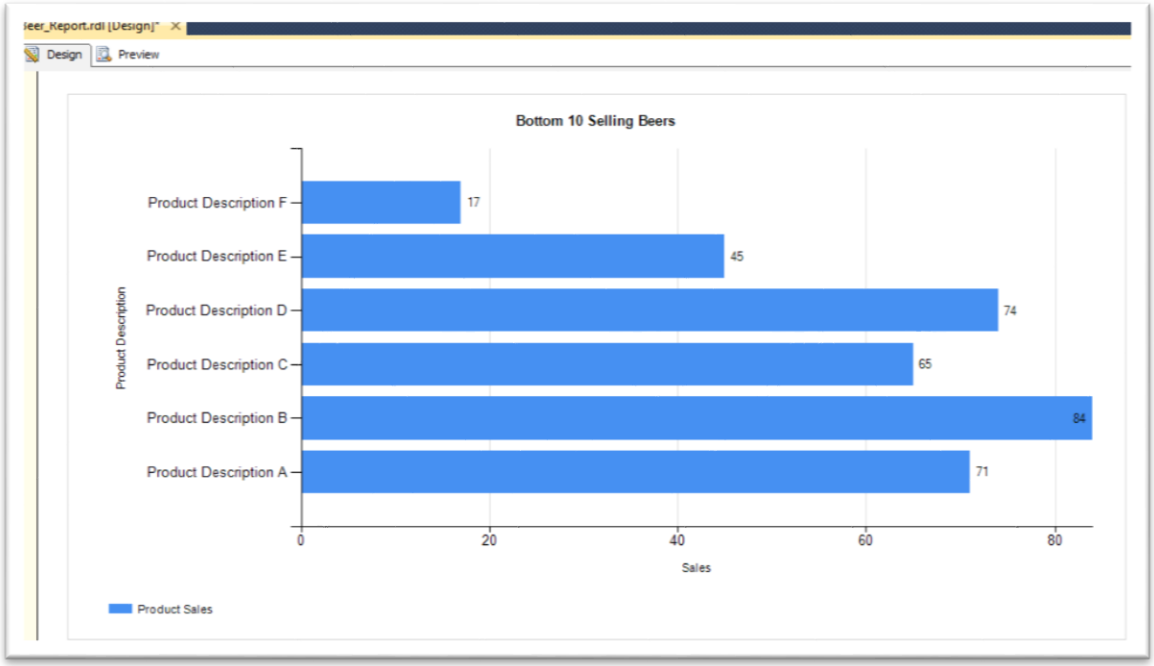
QUERY DESIGNER FOR SALES OF BOTTOM 10 BEER BRANDS CREATION



REPORT DESIGN FOR TOTAL SALES



REPORT DESIGN FOR SALES OF TOP 10 BEER BRANDS



REPORT DESIGN FOR SALES OF BOTTOM 10 BEER BRANDS

```
Output
Show output from: Build
------ Build started: Project: Report Project1, Configuration: Debug ------
Skipping 'Beer_Report.rdl'. Item is up to date.
Build complete -- 0 errors, 0 warnings
------ Deploy started: Project: Report Project1, Configuration: Debug ------
Deploying to http://infodata.tamu.edu/reportserver
Deploying report '/602_Group05/Beer_Report'.
Deploy complete -- 0 errors, 0 warnings
========== Build: 1 succeeded or up-to-date, 0 failed, 0 skipped ==========
========== Deploy: 1 succeeded, 0 failed, 0 skipped ==========
```

DEPLOYMENT SUCCESS

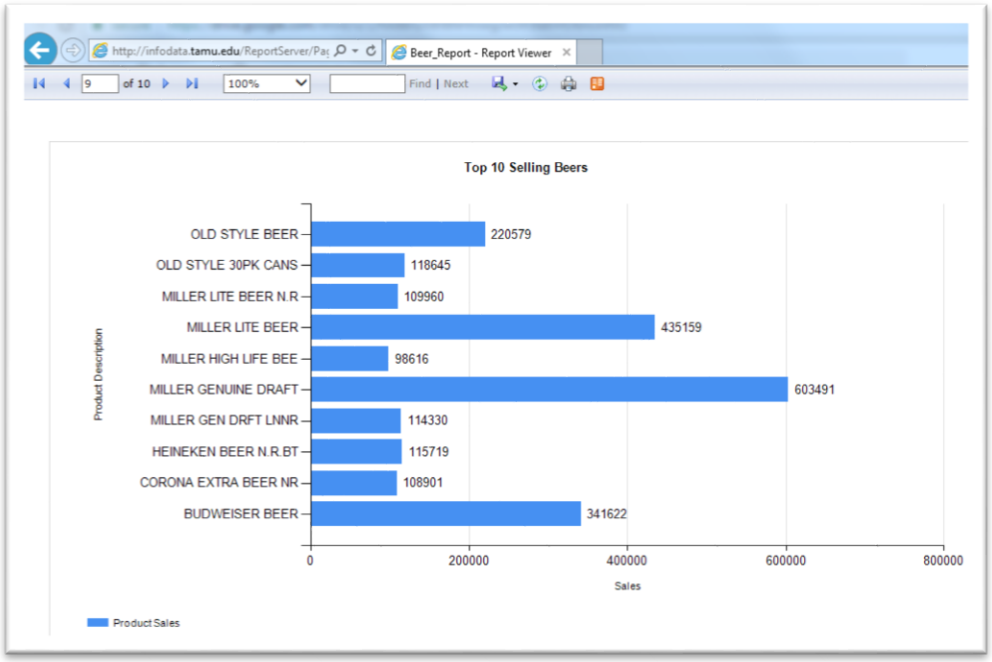
http://infodata.tamu.edu/ReportServer/Pages/ReportViewer.aspx?%2f602_Group05%2fB Beer_Report - Report Viewer

1 of 2 ? 100% Find | Next

Sales Performance of Beer brands across the stores during 1991 - 1997

Product Description	Product Sales
$ARIEL NA CHARDONNAY	77
$ARIEL NA WHITE ZINF	36
AMSTEL LIGHT BEER NR	17170
AMSTEL LIGHT BIER	683
ARIEL CABERNET SAUVI	164
ARIEL WHITE NON ALCH	5
ASAHI BEER	140
ASAHI SUPER DRY BIG	101
AUGSBURGER BOCK	1239
AUGSBURGER BOCK BEER	200
AUGSBURGER DARK BEER	7093
AUGSBURGER GOLDEN	12559
AUGSBURGER LIGHT BEE	426
AUGSBURGER OKTOBERFE	163
AUGSBURGER REG BEER	7251
BADERBRAU BEER NR BT	2758
BADERBRAU BOCK	403
BASS ALE KING SIZE B	1808
BASS ALE NR BTLS	10960
BECKS DARK	228
BECK'S DARK BEER BAM	1200
BECK'S DARK BEER NR'	17037
BECK'S LIGHT NR BTLS	5823

TOTAL SALES TABLE DEPLOYMENT ON IE



BAR DIAGRAM DEPLOYMENT FOR SALES OF TOP 10 BEER BRANDS

BAR DIAGRAM DEPLOYMENT FOR SALES OF BOTTOM 10 BEER BRANDS

Conclusion: The reports generated on SSRS give us a good perspective of the top and bottom most selling brands in certain stores. This information enables us to make an informed decision on stocking up inventory and reducing costs.

18. Task Sheet

Task	Members	Duration (Hours)
Project Planning	Vamsi Kompella Abhishek Pillai Siddharth Sonti	1
Background survey of DFF	Vamsi Kompella Abhishek Pillai Siddharth Sonti	1
Understanding Dominick's data	Vamsi Kompella Abhishek Pillai Siddharth Sonti	4
Description of metadata	Vamsi Kompella Abhishek Pillai Siddharth Sonti	2
Entity Relationship Diagrams in Visio	Vamsi Kompella Abhishek Pillai Siddharth Sonti	4
Domain understanding	Vamsi Kompella Abhishek Pillai Siddharth Sonti	8
Figuring out trends in data	Vamsi Kompella Abhishek Pillai Siddharth Sonti	4
Pivot tables and charts in Excel	Vamsi Kompella Abhishek Pillai Siddharth Sonti	6
Formulating business questions	Vamsi Kompella Abhishek Pillai Siddharth Sonti	12
Prioritization and rationalization	Vamsi Kompella Abhishek Pillai Siddharth Sonti	4
Documentation	Vamsi Kompella Abhishek Pillai Siddharth Sonti	5
Planning	Vamsi Kompella Abhishek Pillai Siddharth Sonti	1
Research and analysis of dimensional models	Vamsi Kompella Abhishek Pillai Siddharth Sonti	2
Description of dimensions and facts	Vamsi Kompella Abhishek Pillai Siddharth Sonti	1

Dimensional modeling for data marts in Visio	Vamsi Kompella Abhishek Pillai Siddharth Sonti	3
Justification of data marts for selected questions	Vamsi Kompella Abhishek Pillai Siddharth Sonti	2
Overall data mart	Vamsi Kompella Abhishek Pillai Siddharth Sonti	1
Mapping tables	Vamsi Kompella Abhishek Pillai Siddharth Sonti	2
Staging Area Creation	Vamsi Kompella Abhishek Pillai Siddharth Sonti	5
Data scrubbing and normalization	Vamsi Kompella Abhishek Pillai Siddharth Sonti	3
Dimensions Creation	Vamsi Kompella Abhishek Pillai Siddharth Sonti	4
Facts Mapping	Vamsi Kompella Abhishek Pillai Siddharth Sonti	4
Verification of data	Vamsi Kompella Abhishek Pillai Siddharth Sonti	2
Reporting Plan	Vamsi Kompella Abhishek Pillai Siddharth Sonti	2
Determination of target reports	Vamsi Kompella Abhishek Pillai Siddharth Sonti	2
Attribute mapping from data mart to report	Vamsi Kompella Abhishek Pillai Siddharth Sonti	2
Report Building	Vamsi Kompella Abhishek Pillai Siddharth Sonti	10
Documentation	Vamsi Kompella Abhishek Pillai Siddharth Sonti	10

By signing below, each member agrees to the above task – responsibility – duration table

1. Vamsi Kompella _____
2. Abhishek Pillai _____
3. Siddharth Sonti _____