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

Comprehensive Report

[redacted]

Satvir Singh Grewal

Rakesh Kumar

Venkata Pavan Tej Deevi

Texas A&M University

May 2, 2017

Table of Contents

1 Credentials for DW, Reporting and Analysis Services 4

2 Introduction 5

2.1 Details about the Data 6

2.1.1 Understanding of the Data 6

2.1.2 Metadata for all OLTP Files 7

2.2 Entity – Relationship Diagram 11

2.3 Domain Understanding 12

3 Independent Data Marts design using Kimball's approach 13

3.1 Dimension Tables 13

3.1.1 DimProduct 13

3.1.2 DimStore 14

3.1.3 DimTime 15

3.1.4 DimProductCategory 16

3.2 Fact Tables 17

3.2.1 FactProductSales 17

3.2.2 FactCategorySales 18

3.3 Star Schemas 20

3.3.1 Product Sales Data Mart 20

3.3.2 Category Sales Data Mart 21

3.4 Mapping of Product Sales Data Mart 22

3.5 Mapping of Category Sales Data Mart 24

3.6 Justification of Business Questions 26

4 Data Integration 27

4.1 Data Quality issues in DFF dataset 27

4.2 ETL Plan 28

4.2.1 Determining all the target data needed in the data warehouse: 29

4.2.2 Determining all the Data Sources: 29

4.2.3 Data mappings for data elements from sources in CSV to staging and then from staging to data warehouse (includes all transformations): 29

4.2.4 Establishing comprehensive data extraction rules: 32

4.2.5 Determining data transformation and cleansing rules: 32

4.2.6 SSIS Functions Used: 33

4.3 ETL Implementation with Screenshots and descriptions: 34
4.3.1 Data Extraction: 34
4.4 Transformation and Cleansing of Data 36
4.4.1 Data Cleansing 36
4.4.2 Transformation of Source data into Dimensions Tables 38
4.4.3 Transformation of Source data into Fact Tables 40
4.4.4 Data Snapshots in the DW Area: 42
5 BI Reporting (using SSRS, SSAS and Report Builder 2012) 43
5.1 Reporting Plan 43
5.2 Determining target reports that satisfies business questions 44
5.3 Mappings from all Independent Data Marts to the Report Attributes 46
5.4 Report Building using Report Builder 3.0 for Question 1 51
5.5 Report Building using SSRS for Question 2 52
5.6 Report Building using SSAS for Question 3 53
5.7 Report Building using Report Builder 3.0 for Question 4 54
5.8 Report Building using SSRS on top of SSAS for Question 5 55
6 References 57

1 Credentials for DW, Reporting and Analysis Services

Credentials of SQL server staging area and data marts

Server: [redacted]

Login: [redacted]

Password: [redacted]

Staging area DB (contains staging tables): 602_Group1-staging-areas

Data mart DB (contains fact and dimension tables): 602_Group1-dw-area

Credentials of SSAS cubes

Server: [redacted]

Question 3 cube: 602 Group1-SSASQ3

Question 4 cube: 602 Group1-SSRSonSSAS_Q5

Report deployment details

Server: [redacted]

Report Builder 3.0, Question 1 –

Folder: SSRSQ2_Attempt2, **Report name:** 602 Group1 – Q1

SSRS Report, Question 2 –

Folder: SSRSQ2_Attempt2, **Report name:** Report1

Report Builder 3.0, Question 4 –

Folder: SSRSQ2_Attempt2, **Report name:** 602 Group1 – Q4

SSRS on SSAS, Question 5 –

Folder: SSRSonSSAS_Q5, **Report name:** Comparison of Sales when different coupons are applied

2 Introduction

Dominick's Finer Foods is a Chicago-area grocery store chain. It was founded in 1918 and by 1998 the chain has established 116 stores. Dominick's management head offices are located in Oak Brook, Illinois whereas the main distribution center is located in Northlake, Illinois. The chain was selling a variety of products and services such as grocery, pharmacy, fresh prepared foods, in-store restaurants/cafés, Starbucks cafés video departments, one-hour photo, bulk foods, etc. So, the chain was a giant in terms of the number and variety of customers it is serving and the diversity in the services and products it provides.

The retail industry is a very dynamic and complex industry as the business success depends on multiple factors such as offerings, pricing, customer preferences, marketing, branding, competitors etc. So, making informed strategic decision is a tough choice and data from past can help a lot. This brings in the need of building a capability to analyze the various historical information present and to learn from it. Data warehouse is precisely what the company needs.

So, using the data about the products and services offered, demographics of the stores, pricing and promotional activities, the group is trying to come up with findings and numbers that can serve as the lighthouse for the company in making their decisions. We have tried to briefly cover some of the key factors which can directly impact the success of the stores. Also, the sample questions provide an idea about the possibilities made possible by the warehouse. The key areas touched upon in the report are:

a) **Deciphering the business numbers:** In a vast chain as DFF, keeping a clear track of the profits, losses, margins at the granular level is very important. The overall numbers can be misleading and inappropriate in judging the health of the chain. We have tried to understand how the different stores from different tiers are doing overall, how have the stores performed over time.
b) **Stocking:** A smooth management of warehousing and stocking of products ensure that the stores are adequately stocked with the goods. Under stocked stores turn away the customer and the brand loses reputation. Similarly, over stuffed stores mean the produce may expire on the shelves, or we have to drastically reduce our prices to sold the excessive stock. We have tried to understand how the DFF stores perform in different product categories, how different time of the year impact the consumption of the products etc.
c) **Pricing and promotion strategies:** A successful store needs to price its products and develop promotional activities smartly. Understanding of how the different stores react to price changes and promotional activities is instrumental I achieving the desired profits even by keeping the prices low.
d) **Significance of demographics:** For most of the businesses the success of the business depends on satisfying the customer needs. With a retail store like DFFs the task of attracting and retaining the customers becomes more difficult because of the diversity of the customer base. We have tried to utilize the demographics data and tried to understand how the stats affect the sales of various stores and how to tailor make the strategies at store level.

The above-mentioned categories serve as the guideline in an attempt to view the data from different angles. The team has tried to touch upon different facets of the business and tried to deliver facts and numbers to be used for better business strategies.

2.1 Details about the Data

2.1.1 Understanding of the Data

Dominick's data contains store-level scanner data spanning over 7 years and more than 3500 UPC's. This sales data belongs to 100 stores of DFF all across the United States. The total volume of this data is around 4.76 Gigabytes. The entire research data can be divided into *General Files* and *Category Specific Files.* The general files contain data related to the different categories in the project. The general files are further subdivided into *Customer-Centric* and *Demographic* data. The category files are further subdivided into *UPC files* and *Movement* data.

Some of the characteristics of the data can be elucidated as follows:

- The data is collected through daily transactions across multiple stores of DFF retail chain and describes the purchase trend of the customers.
- A good section of the data provided is dirty in nature and requires manipulation and cleaning before being analyzed.
- The customer data has been divided into multiple customer categories to facilitate identification of customer sections based on different demographics.

2.1.2 Metadata for all OLTP Files

1) Customer Count Files:

The customer count file contains information regarding in-store traffic. The data contained in these files on a daily basis is store specific. This data refers to the number of customers visiting the store and purchasing something. This file also contains the total dollar sales and total coupons redeemed details, by DFF defined department. These figures are compiled daily from the register/scanner receipts. The details of the CCOUNT file variables are described below:

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

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

2) Store Specific Demographics:

The demographics file consists of store-specific demographic data. The data originally comes from U.S. government (1990) census data for the Chicago metropolitan area. Market Metrics processed this data to generate demographic profiles for each of the DFF stores. The table below gives the descriptions for all the files in the demographics database. The demographics file contains the following variables:

Demographics	
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
wrkwnch	% of working women with no children
telephn	% of households with telephones
mortgage	% of households with mortgages
nwhite	% of population that is non-white
poverty	% of population with income under $15,000
shopcons	% of Constrained Shoppers
shophurr	% of Hurried Shoppers
shopavid	% of Avid Shoppers
shopstr	% of Shopping Stranges
shopunft	% of Unfettered Shoppers
shopbird	% of Shopper Birds
shopindx	Ability to Shop (Car and Single Family House)
shpindx	Ability to Shop (Car and Single Family House)

3) Category Descriptions:

Category descriptions include the list of all UPC's and their description and size for a particular category. For example, under the analgesics category we can have pain relievers can be aspirin, ibuprofen, or acetaminophen base.

4) Week's Decode Table:

This data is part of Dominick's manual and Codebook document. These files consist of information on the codes that correspond to the week when sales data was recorded. This table is particularly useful in comparison of sales trends between holiday and non-holiday weeks. The data is this table could also be of essence in evaluating monthly, quarterly or semiannual sales reports.

5) UPC Files:

The UPC files contains information of UPC belonging to each category. The files are named upcxxx, where xxx is the three-letter acronym for the category. Information extracted can be Product Name, Size, UPC Number etc. The variables in this file are described below:

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

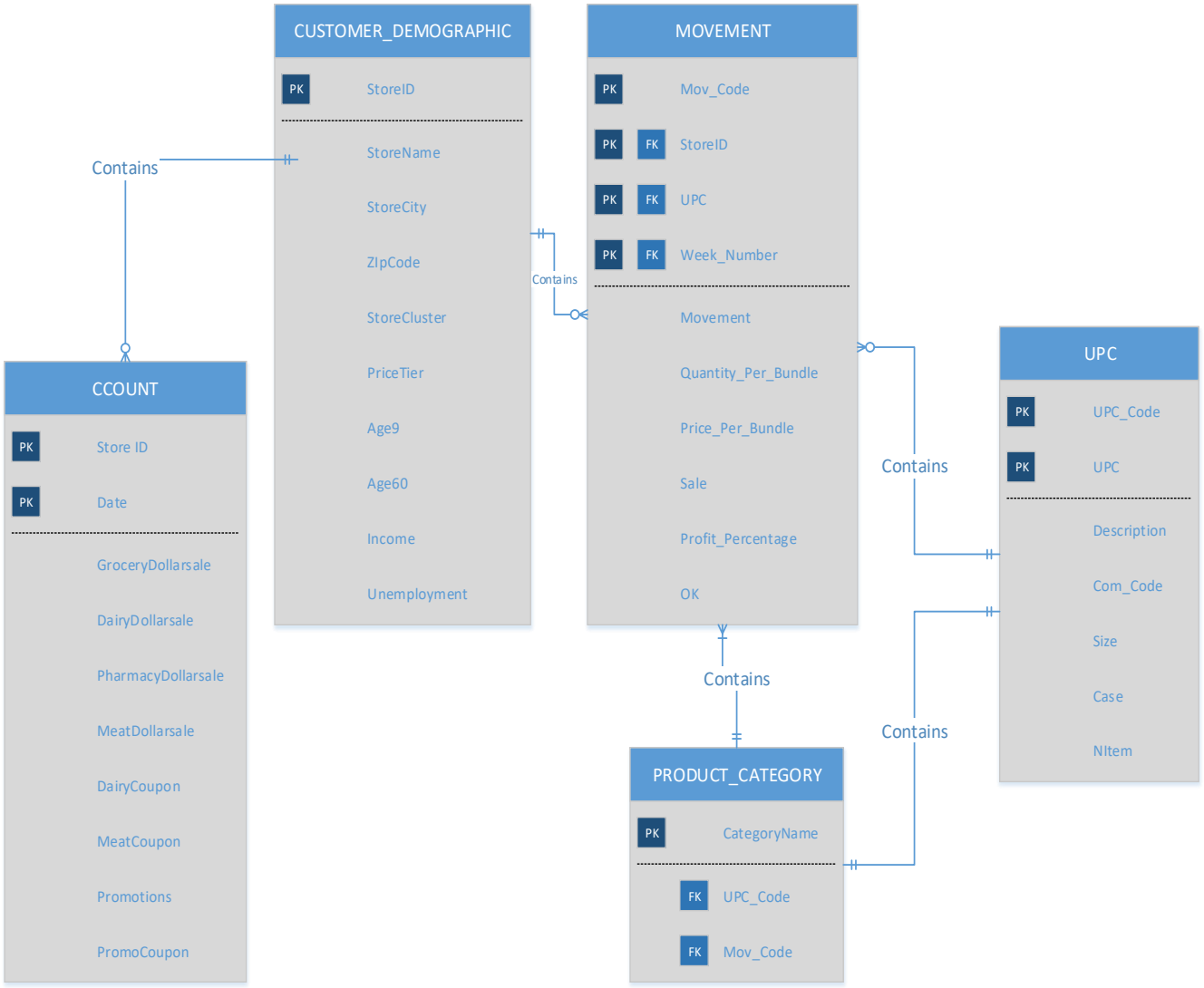
6) Movement Data by UPC:

These files contain the information that are crucial in answering business questions pertaining to holiday sales, demographic-level sales etc. These files contain weekly sales data of the products belonging to different categories. The files are named wxxx where xxx is the three- letter acronym for the category. The variable details for this file are as follows:

Movement of Data by UPC			
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

2.2 Entity – Relationship Diagram



2.3 Domain Understanding

Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data:

Micro-marketing refers to the customization of marketing mix variables to the store-level. This paper shows how prices can be profitably customized at the store-level, rather than adopting a uniform pricing policy across all stores (Montgomery, 1997). We have justified some of the business questions based on this understanding. In the business question related to price tiers, we propose micro-marketing pricing strategies to improve sales and overall profits for underperforming stores in the *high* price tier.

Chain-wide and store-level analysis for cross-category management

The main purpose of this study is to present a relatively simple, feasible and easy-to-implement approach for chain wide, store-level cross-category analysis. This analysis is intended to help retail managers make both chain-wide and store-specific decisions (Wagner A. Kamakura, 2007). The paper elaborates on the cross-category store level impact of price promotions. This theory can be applied to Dominick retail chain. This paper also discusses the effect of discount on one brand to the sale of other brands in the same category. For example, a discount on say a soft drink like coca cola in DFF can have an impact on the sales of Pepsi and other brands under the same category across stores. This paper provides an insight into the effect of price reduction or discounts across different brands under different categories.

How markets and revenues react to price promotions

The study also tries to understand how different promotional offers such as price promotions, coupons, two-for-one offers and temporary price cuts. Every store employ one or other form of such marketing strategies. Generally the offers are expected to help the businesses grow but the study 'Who benefits from price promotions' by Srinivasan, Shuba, et al, 2002 show a different trend altogether. They build on top of the previous work that price promotions have little long-term effect on sales volume. They further established that growth in revenues and margins also doesn't sustain for very long. This theory can be applied to DFF and we have tried to establish the impact of price reduction has on net profits. We have tried to verify that reducing prices beyond a certain price leads to slump in profit earnings. The price reduction alone cannot help companies build profits. Also, the type of promotion has different responses in the way they increase the sales. The paper also tries to touch upon that.

Impact of holidays or seasonal demand of commodities

The holiday periods and festivals generally bring along with themselves a changed shopping pattern of the patrons. There is a temporary spike in a particular set of products over a particular period of time. Such irregularities call for making decisions accordingly. Nevo, Aviv and Hatzitaskos, Konstantinos in their paper "Why Does the Average Price

of Tuna Fall during Lent?" (August 2005) tried to understand how the holiday period impacts various numbers pertaining to the category. This report takes a cue from the paper and tries to understand how the special occasions impact the sales of the products to better understand how to manage the stocking and promotional activities around the products. Also, the report tries to understand the pattern of sales of different products over a period of time to understand what those seasonal hikes and the timings are.

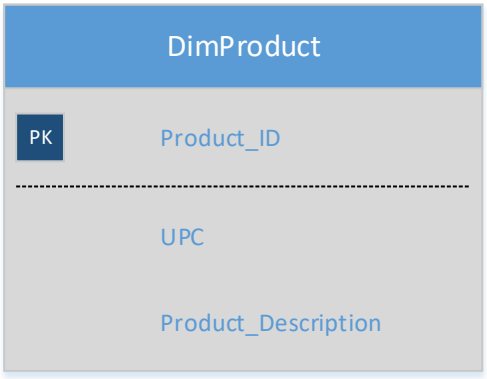
Demographic variables impacting the business

Understanding the demographics of a locality before opening a new store or devising strategies of an already open store is critical in the success of a business. The article 'The Demographic Variables That Affect a Business' by Rick Suttle has tried to identify some of the reasons that affect the revenues and sales of a particular business. Understanding the demographic impact on the business is critical in smartly targeting the consumers. This report tries to establish some trends in sales and profits with respect to certain demographic attributes such as income, age tec.

3 Independent Data Marts design using Kimball's approach

3.1 Dimension Tables

3.1.1 DimProduct



The description of the attributes of the PRODUCT dimension table are summarized below:

Product_ID – A unique identifier of the product dimension table which has been generated as a surrogate key.

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

Product_Description – Description of the product available in the DFF retail store.

3.1.2 DimStore

The description of the attributes of the STORE dimension table are summarized below:

Store_ID – A unique identifier of the store dimension table which has been generated as a surrogate key.

Store_Number – The assigned number to a particular store.

Store_Name – The name of a particular store in the DFF retail store chain.

Store_City – The city where the store is located.

Store_Zip – The zip where the store is located.

Store_Zone – The identifier of the zone to which a store belongs.

3.1.3 DimTime

The description of the attributes of the TIME dimension table are summarized below:

Time_ID – A unique identifier of the time dimension table which has been generated as a surrogate key.

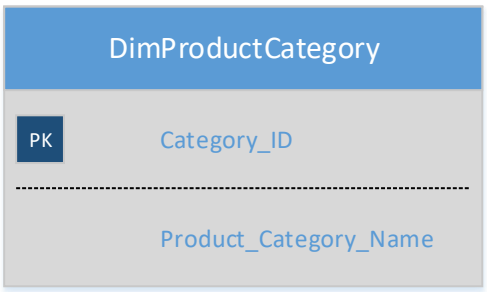
Week_Number – The identifier for the week in the year when sales have occurred.

Month – The identifier for the month when sales have occurred.

Year – The identifier for the year when sales have occurred.

Special_Event_Name – The identifier to indicate the occurrence of a special event in the year.

3.1.4 DimProductCategory



The description of the attributes of the PRODUCT_CATEGORY dimension table are summarized below:

Category_ID – A unique identifier of the product_category dimension table, which has been generated as a surrogate key.

Product_Category_Name – The name of a product category defined in the DFF retail store chain.

3.2 Fact Tables

3.2.1 FactProductSales

The description of the attributes of the PRODUCT SALES fact table are summarized below:

Fact Table Keys:

Store_ID – A unique identifier for stores, which has been generated as a surrogate key and helps in dividing data by store.

Time_ID – A unique identifier for time, which helps in dividing data by time dimensions like year, month, and week number.

Product_ID – A unique identifier for products, which has been generated as a surrogate key.

Fact Table Measures:

Sale_Type – This attribute denotes the sale code (B, C, S) for a product. This is an indicator if a product was sold on promotion.

Store_Price_Tier – This attribute denotes the price tier for a store.

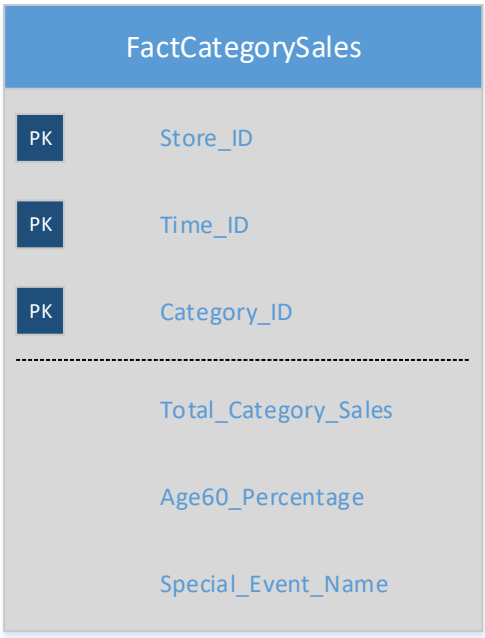
Quantity_Sold – This represents the number of units of product sold for every store.

Total_Sale – This is a derived attribute. It indicates the dollar value of the product sales across various dimensions.

Total_Profit – This attribute denotes the total profit made by DFF for every store.

Profit_Percentage – This attribute represents the profit percentage for each store.

3.2.2 FactCategorySales



The description of the attributes of the CATEGORY SALES fact table are summarized below:

Fact Table Keys:

Store_ID – A unique identifier for stores, which has been generated as a surrogate key and helps in dividing data by store.

Time_ID – A unique identifier for time, which helps in dividing data by time dimensions like year, month, and week number.

Category_ID – A unique identifier for category, which has been generated as a surrogate key.

Fact Table Measures:

Total_Category_Sales – This attribute indicates the dollar value of the category sales across various dimensions.

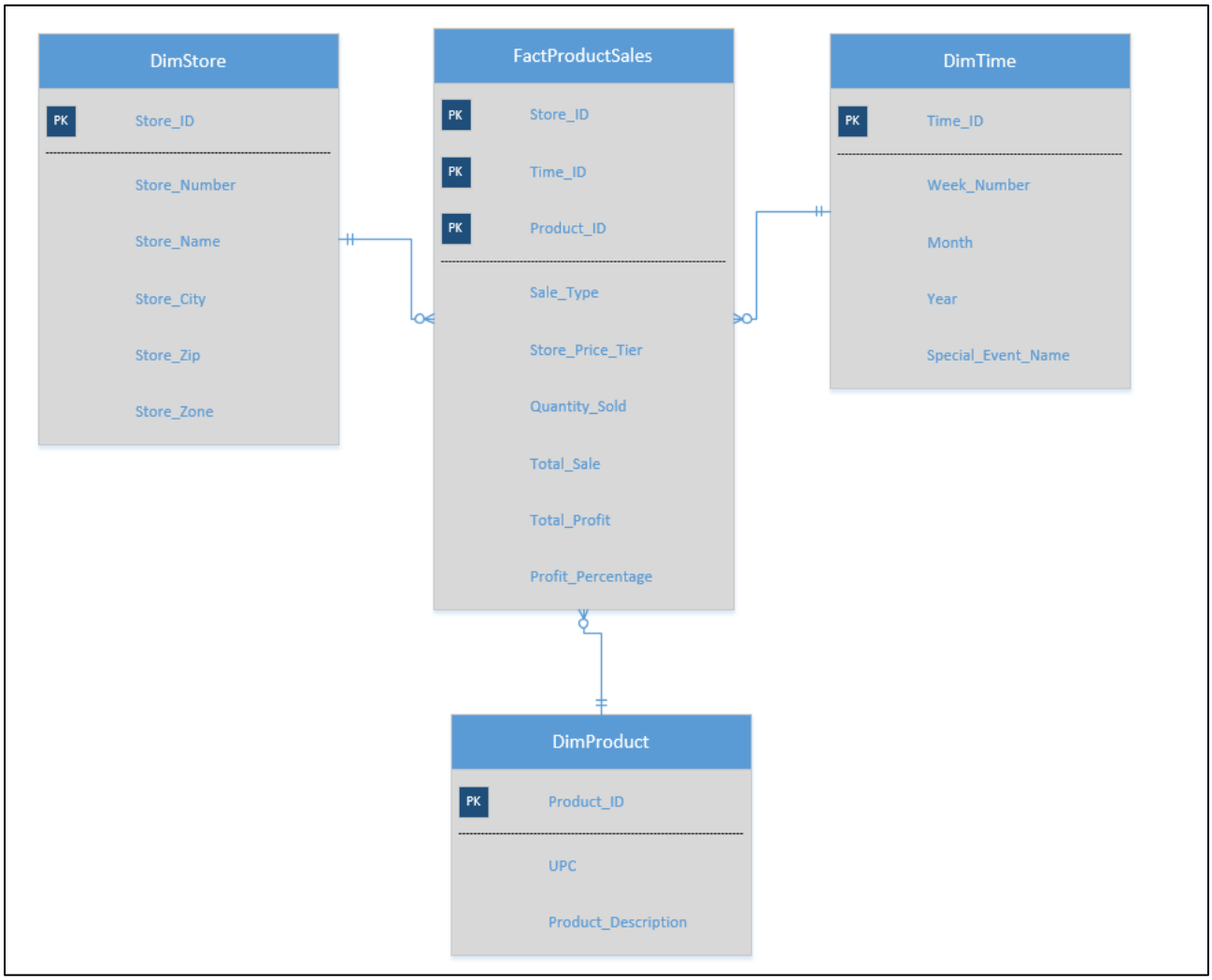
Age60_Percentage – This attribute indicates the percentage of customers above the age of 60.

Special_Event_Name – The identifier to indicate the occurrence of a special event in the year.

3.3 Star Schemas

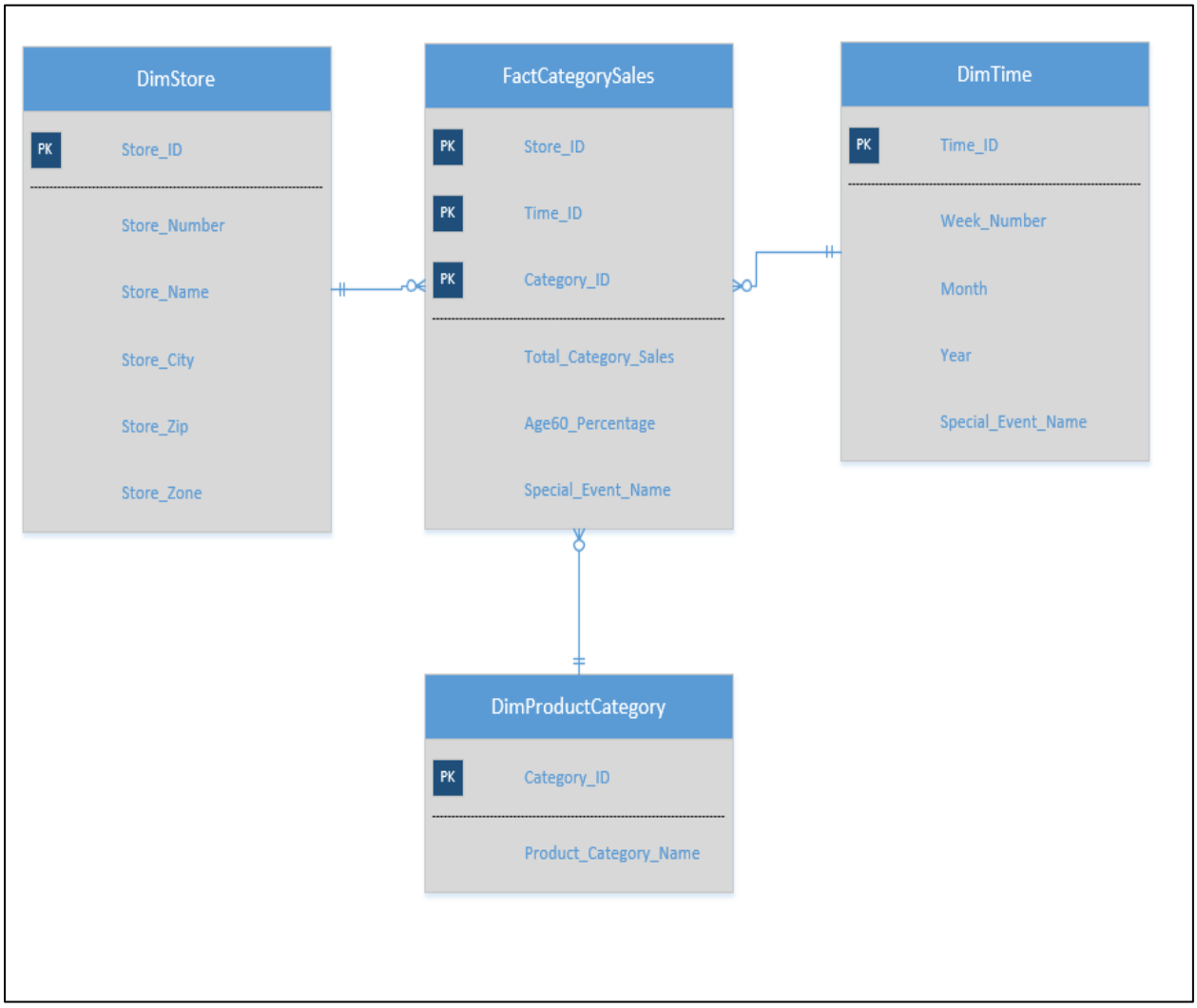
3.3.1 Product Sales Data Mart

We have integrated the above fact and dimension tables to create two Star Schemas. The data marts created using these tables are sufficient to answer business questions. The *Product Sales* data mart is shown below which comprises of Product Sales fact table and three dimension tables namely Store, Time and Product. The data mart shown below helps answer the business questions 1, 2 and 5.



3.3.2 Category Sales Data Mart

The *Category Sales* data mart is shown below which comprises of Category Sales fact table and three dimension tables namely Store, Time and Product_Category. The data mart shown below helps answer the business questions 3 and 4.



3.4 Mapping of Product Sales Data Mart

Dimension Table	Dimension Attribute	Source Table	Source Table Attribute	Mapping Function
DimProduct	Product_ID	UPC Table		
	UPC		UPC	
	Product_Description		Description	

Dimension Table	Dimension Attribute	Source Table	Source Table Attribute	Mapping Function
DimStore	Store_ID	Dominick's Store and Store Specific Demographics		A surrogate generated to use as the primary key
	Store_Number		StoreID	
	Store_Name		StoreName	
	Store_City		StoreCity	
	Store_ZIP		ZipCode	
	Store_Zone		PriceTier	

Dimension Table	Dimension Attribute	Source Table	Source Table Attribute	Mapping Function
DimTime	Time_ID	Week Decode		
	Week_Number		Week_Number	
	Month			This field is calculated using the week numbers
	Year			This field is calculated using the week numbers
	Special_Event_Name		Special_Events	

Fact Table	Dimension Attribute	Source Table	Source Table Attribute	Mapping Function
FactProductSales	Store_ID	Movement	StoreID	This ID is part of the primary key.
	Time_ID			The attribute serves as the primary key for the DimTime.
	Product_ID	Movement	UPC	The attribute serves as the primary key for the DimProduct.
	Sale_Type		Sale	
	Store_Price_Tier			
	Quantity_Sold		Movement	
	Total_Sale		Sale	
	Total_Profit			This is calculated using the total sales and the profit percentage
	Profit_Percentage	Movement Table	Profit_Percentage	

3.5 Mapping of Category Sales Data Mart

Dimension Table	Dimension Attribute	Source Table	Source Table Attribute	Mapping Function
DimStore	Store_ID	Dominick's Store and Store Specific Demographics		A surrogate generated to use as the primary key
	Store_Number		StoreID	
	Store_Name		StoreName	
	Store_City		StoreCity	
	Store_ZIP		ZipCode	
	Store_Zone		PriceTier	

Dimension Table	Dimension Attribute	Source Table	Source Table Attribute	Mapping Function
DimProductCategory	Category_ID	CCOUNT		A surrogate generated to use as the primary key
	Product_Category_Name		Various Columns	The columns in the CCOUNT table are mapped to get the various values to be put in the catgory names.

Dimension Table	Dimension Attribute	Source Table	Source Table Attribute	Mapping Function
DimTime	Time_ID	Week Decode Table		
	Week_Number		Week_Number	
	Month			This field is calculated using the week numbers
	Year			This field is calculated using the week numbers
	Special_Event_Name		Special_Events	

Fact Table	Dimension Attribute	Source Table	Source Table Attribute	Mapping Function
FactCategorySales	Store_ID	Movement	StoreID	The attribute serves as the primary key and come from DimStore.
	Time_ID			The attribute serves as the primary key and come from DimTime.
	Category_ID	Movement	UPC	The attribute serves as the primary key and come from DimCategory.
	Total_Category_Sales		Sale	
	Age60_Percentage	Customer_Demograhics		
	Special_Event_Name	Week_Decode	Special_Events	

3.6 Justification of Business Questions

1) **Identify the store price tier that contributes most to overall profit of DFF.**

This question is answered by the Product Sales Data Mart. The attributes Store_ID and Store_Zone are used from DimStore dimension table. These attributes help us locate the stores in different price tier. Then the fact table FactProductSales is used to find the sales made by the stores across all products.

2) **Analyze the store wise revenue contribution in the last 3 years.**

The dimension table DimStore and FactProductSales fact table come together to allow us analyze the sales made by different stores. The fact table FactProductSales can be rolled up or drilled down to a grain of a week to see the trend in sales. The Store_ID and Total_Sales attributes are used for answering the questions.

3) **Identify the sales trend in DFF for different product categories over time.**

The business question is addressed by the Category Sales data mart. Category_Id field from the DimProductCategory table is connected to the FactCategorySales fact table to get sales made in different categories. The DimTime dimension table is used to analyze the sales over different weeks of the year. Special_Event_Name field in the DimTime dimension table helps us analyze the data over holiday or non-holiday weeks.

4) **What is the effect of age demographics on the sale of pharmaceutical products?**

This question is answered by the Category Sales data mart. The fact Age60_Percentage in table FactCategorySales gives us the age demographics. The fact table is then linked to the dimension table DimProductCategory to get the sales of pharmaceutical products.

5) **Comparison of sale of products when there is a bonus buy offered, coupon discount offered, simple price reduction, and when no discount is offered.**

This business question is answered by the Product Sales data mart. The Sale_Type fact in the FactProductSales fact table classifies the sales made into different categories based on the offer given. The analysis helps us understand how the sales are impacted by different promotional offers.

4 Data Integration

4.1 Data Quality issues in DFF dataset

Parameter	Data Quality Problem in Dominic Finer Foods
Referential Integrity	Referential integrity was difficult to figure out as there were values that were null even in keys which could be assumed to be primary key. e.g.:- Store number was repeated and had inconsistent values.
Format	Values did not follow consistent formatting standards. E.g.: Date attribute had a format as XXXXXX instead of XX/XX/XX
Data Standard	Though data elements were explained well in DFF catalog, looking at the data it was difficult to understand what it represented. E.g.: Coupons had decimal values, so shall a user assume it is coupon's dollar value or it is dirty data.
Consistency	Data was explained well in the catalog and hence the data had same meanings across the systems. However, amongst files there was an overlap with the data values which created misrepresentation of data.
Completion	After formulating the business question, it was found that a lot of data was not present. For example, the movement files were not available for all the UPCs.
Accuracy	Accuracy was questionable as data had null values where value was expected, negative values for Sales data and decimal value for coupon.
Validity	Data values did not fall within acceptable ranges defined by the business. Store number have out of range values and hence have invalid values.
Fit for purpose	The information contains a lot of sales data and customer data and hence is valuable to the business.

4.2 ETL Plan

ETL stands for Extract-Transform-Load. ETL covers a process of how the data is loaded from the source system to the data warehouse.

Extract

The Extract step covers the data extraction from the source system and makes it accessible for further processing. The main objective of the extract step is to retrieve all the required data from the source system with as little resources as possible. The extract step should be designed in a way that it does not negatively affect the source system in terms or performance, response time or any kind of locking.

Clean

The cleaning step is one of the most important as it ensures the quality of the data in the data warehouse. Cleaning should perform basic data unification rules, such as making identifiers unique, Convert null values into standardized not available/not provided value, convert phone numbers to a standardized form and validate address fields against each other (State/Country, City/State, City/ZIP code, City/Street).

Transform

This involves a set of rules to transform data from source to the target. This involves converting the data into conformed dimensions using the same units so that they can later be joined. This also involves joining the data from numerous sources, generating aggregates, surrogate keys, sorting and deriving calculated values.

Load

It is imperative that the load is performed correctly and with as little resources as possible. In order to make the load process efficient, it is helpful to disable any constraints and indexes before the load and enable them back only after the load completes. The referential integrity needs to be maintained by ETL tool to ensure consistency.

4.2.1 Determining all the target data needed in the data warehouse:

Data from Sources	Data in staging database (Table names)	Data in Data Warehouse (Table names)
Dominick Stores details	Dominick_Stores	DimStore
Time details	Week_Decode	DimTime FactCategorySales
Product (UPC) details	UPC	DimProduct
Movement details	Sales_Product	FactProductSales
Demographic	Demographic	FactCategorySales
CCount	Category_Count	DimProductCategory, FactCategorySales

4.2.2 Determining all the Data Sources:

Data needed	Data Source
Dominick Stores details	Dominick_Store.xls (created using Dominick's Data Manual.pdf)
Time details	Week_Decode.xls (created using Dominick's Data Manual.pdf)
Product (UPC) details	UPC<productAcronym>.csv
Movement details	W<productAcronym>.csv
Demographic	DEMO.csv
CCount	CCOUNT.csv

4.2.3 Data mappings for data elements from sources in CSV to staging and then from staging to data warehouse (includes all transformations):

SOURCE FILE NAME	SOURCE FILE ATTRIBUTES	STAGING AREA TABLE NAME	STAGING AREA TABLE ATTRIBUTES
Week_Decode.csv	Week #	Week_Decode	Week_Number
	Start		Start_Date
	End		End_Date
	Special Events		Special_Event
Dominick_Stores.csv	Store	Dominick_Stores	Store_Number
	City		Store_City
	Price Tier		Price_Tier
	Zone		Store_Zone
	Zip Code		Store_Zip
CCount.csv	STORE	Category_Count	Store_Number
	DATE		Date
	GROCERY		Grocery

	DAIRY		Dairy
	FROZEN		Frozen
	BOTTLE		Bottle
	MEAT		Meat
	MEATFROZ		FrozenMeat
	FISH		Fish
	PRODUCE		Produce
	SALADBAR		Salad_Bar
	FLORAL		Floral
	DELI		Deli
	CONVFOOD		Conv_Food
	CHEESE		Cheese
	BAKERY		Bakery
	PHARMACY		Pharmacy
	GM		General_Merchandise
	JEWELRY		Jewelry
	COSMETICS		Cosmetics
	HABA		Haba
	CAMERA		Camera
	PHOTOFIN		Photo
	VIDEO		Video
	VIDEOREN		Video_Rental
	BEER		Beer
	WINE		Wine
	SPIRITS		Spirits
	FTGCHIN		Foodtogo_Chinese
	FTGITAL		Foodtogo_Italian
	WEEK		Week
UPC<productAcronym.csv	UPC	Upc	Upc
	DESCRIP		Product_Description
W<productAcronym>.csv	UPC	Sales_Product	Upc
	WEEK		Week
	STORE		Store_Number
	QTY		Quantity
	PRICE		Price
	PROFIT		Profit_Percentage
	SALE		Sale_Type
	OK		Ok
DEMO.csv	STORE	Demographics	Store_ Number
	AGE60		Age_60
	NAME		Store_Name

DATAWAREHOUSE TABLE	DATAWAREHOUSE TABLE ATTRIBUTES	MAPPING FUNCTION	STAGING AREA TABLE NAME	STAGING AREA ATTRIBUTE
DimStore	Store_ID	Surrogate Key		
	Store_Number		Dominick_Stores	Store_Number
	Store_Name		Demographics	Store_Name
	Store_City		Dominick_Stores	Store_City
	Store_Zip			Store_Zip
	Store_Zone			Store_Zone
DimProduct	Product_ID	Surrogate Key		
	Upc		Upc	Upc
	Product_Description			Product_Description
DimProduct Category	Category_Id	Surrogate Key		
	Product_Category_Name		Category_Count	Derived from the column headings
DimTime	Time_ID	Surrogate Key		
	Week_Number		Week_Decode	Week_Number
	Month	Derived		
	Year	Derived		
	Special_Event_Name		Week_Decode	Special_Event
FactProductSales	Store_Id		DimStore	Store_Id
	Product_Id		DimProduct	Product_Id
	Time_Id		DimTime	Time_Id
	Sale_Type		Sales_Product	Sale_Type
	Store_Price_Tier		Dominick_Stores	Price_Tier
	Quantity_Sold		Sales_Product	Quantity
	Total_Sale	Derived		
	Total_Profit	Derived		
	Profit_Percentage		Sales_Product	Profit_Percentage
FactCategorySales	Store_Id		DimStore	Store_Id
	Time_Id		DimTime	Time_Id
	Category_Id		DimProductCategory	Category_Id
	Total_Category_Sales	Derived		
	Age60_Percentage		Demographics	Age_60
	Special_Event_Name		DimTime	Special_Event_Name

4.2.4 Establishing comprehensive data extraction rules:

The initial step involved in the ETL process is Data Extraction. Data Extraction for Dominick's data has been performed meticulously in order to efficiently build the data warehouse.

Data Extraction involved getting all the data sources to a single format i.e. tables in Microsoft SQL Server Studio, that can be later used for transformation and loading.

The data from different sources are in the Comma Separated Value (.csv) formats.

In addition to the data provided, we have also extracted week and store details data from the data manual for Dominick.

The data from the source files has been loaded into the data staging area '*602Group1-staging-area*'.

4.2.5 Determining data transformation and cleansing rules:

The data imported into our database from the csv files has some data quality issues. The data must be transformed and cleaned before it can be used to make correct business decisions. The extracted data serve as the input to the transformation and cleaning stage. Some of the transformation and cleaning rules that we applied to the extracted data are as follows:

a) **Handling NULL Values**: Some of the fields contain blanks, where as some other fields put a default value such as '.'(dot) Depicting a blank field. We adopted a uniform convention where fields with no data were set NULL throughout the data marts. In some cases where the column value was necessary to answer the business questions but was not available, the records were ignored during the look up.

b) **Removing Dirty Data**: The attributes with dirty data or the data that doesn't make business sense was removed. There were some records which are non-meaningful e.g. negative value for Week column, one UPC code having 2 product descriptions etc. were removed.

c) **Data Conversion**: Most of the data in the source files i.e. .csv files was of data type varchar. The data was simply imported into the staging area tables. To perform meaningful computations such as aggregation, joins, the data should be transformed into different data types such as int, float and date.

d) **Surrogate Keys**: Surrogate keys were generated for all the dimension tables. Surrogate key ensures that we don't use any intelligent keys as primary and foreign keys. As the attributes which signify any information can be changed by the suppliers and that can lead to breaking of our data warehouse. Surrogate keys keep us safe from change in data values down the period of time.

e) **Derived Attributes**: The business questions chosen by us require some values which need to be calculated based on the given data. The derived fields are:

- The ‘Total _Sale’ field in the fact table FactProductSales is obtained by calculating ‘Total _Sale’ = (Movement*Price)/Quantity.
- The ‘Total_Profit’ field in the fact table FactProductSales is obtained by calculating ‘Total_Profit’ = Total_Sale*Profit_Percentage.
- The Start_Date column in Week_Decode table was used to derive month and year attributes.

4.2.6 SSIS Functions Used:

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

1) DERIVED COLUMNS:
This transformation creates new column values by applying expressions to transformation input columns. An expression can contain any combination of variables, functions, operators, and columns from the transformation input. The result can be added as a new column or inserted into an existing column as a replacement value. We used the derived column transformation for calculating *Total_Sales* and *Total_Profit* for the *FactProductSales* fact table.

2) AGGREGATE:
The aggregate transformation applies aggregate functions, such as sum, count, average to column values and copies the results to the transformation output. It also provides the GROUP BY clause, which you can use to specify groups to aggregate across. We implemented the aggregation transformation to evaluate *Total_Category_Sales* in the *FactCategorySales* fact table.

3) UNPIVOT:
The unpivot transformation makes an unnormalized dataset into a more normalized version by expanding values from multiple columns in a single record into multiple records with the same values in a single column. We made use of this transformation in the *DimProductCategory* dimension table and the *FactCategorySales* fact table.

4) LOOKUP:
This transformation performs lookups by joining data in input columns with columns in a reference dataset. We made use of lookups to access additional information in related tables that is based on values in common columns.

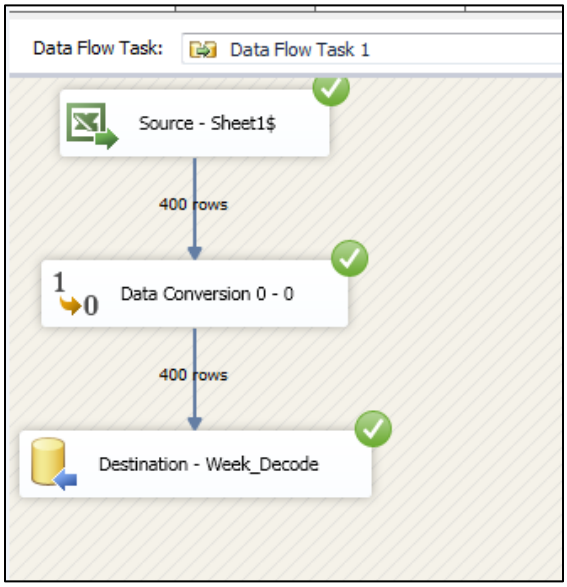
5) DATA CONVERSION:
The data conversion transformation converts the data in an input column to a different data type and then copies it to a new output column. For example, a package can extract data from multiple sources, and then use this transformation to convert columns to the data type required by the destination data store.

4.3 ETL Implementation with Screenshots and descriptions:

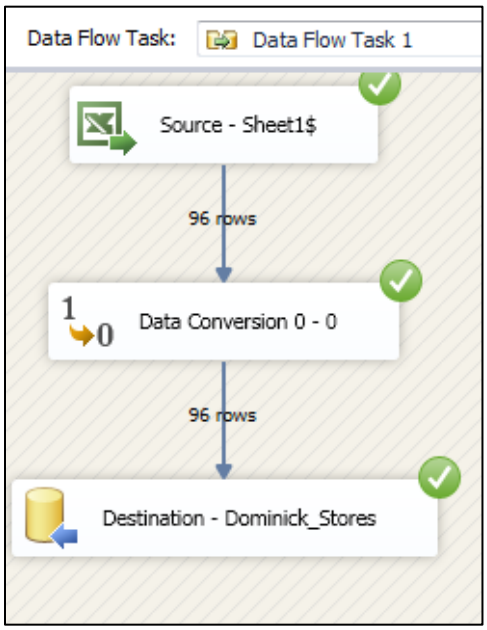
The ETL process for the data warehouse was implemented using the SQL Server Management Studio and SSIS (Microsoft Visual studio – SQL Server Data Tools). The steps involved in this process are elaborated below, along with the SSIS tool screen shots explaining the entire process step-by-step:

4.3.1 Data Extraction:

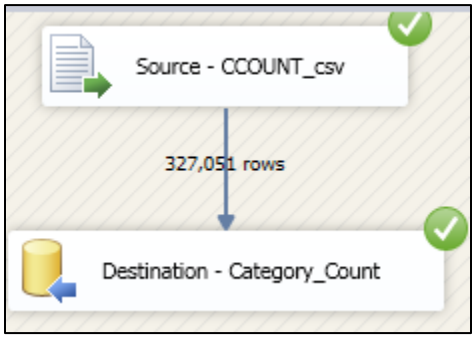
Extraction of Week_Decode.xls



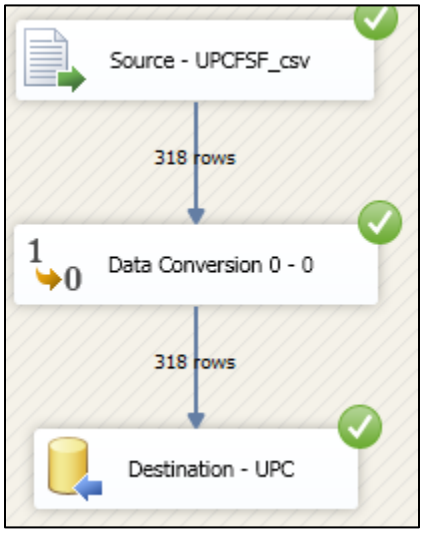
Extraction of Dominick_Stores.xls



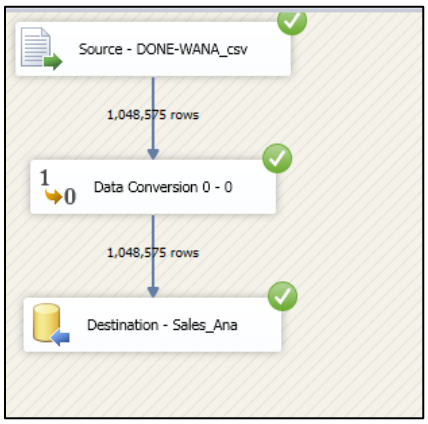
Extraction of CCount.csv



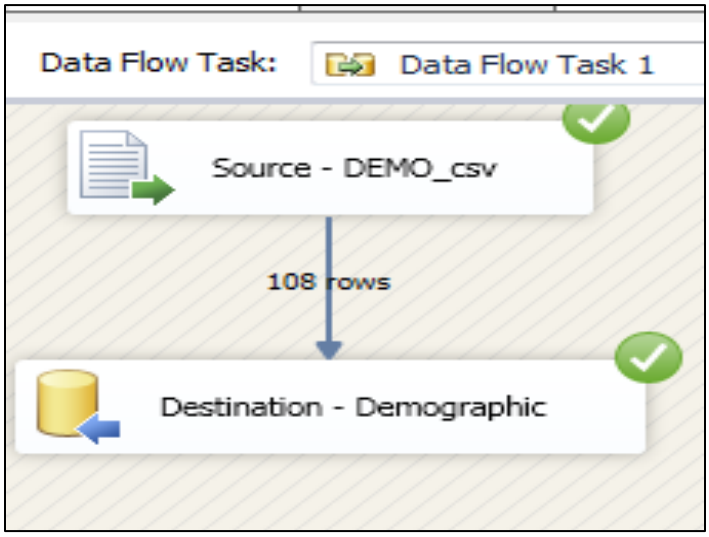
Extraction of UPC<product_acronym>.csv



Extraction of W<product_acronym>.csv



Extraction of DEMO.csv



4.4 Transformation and Cleansing of Data

4.4.1 Data Cleansing

We used 'SSIS SQL task' functionality to clean data in staging area before performing transformation operations. Below are SQL statements and snap shots of SQL tasks created.

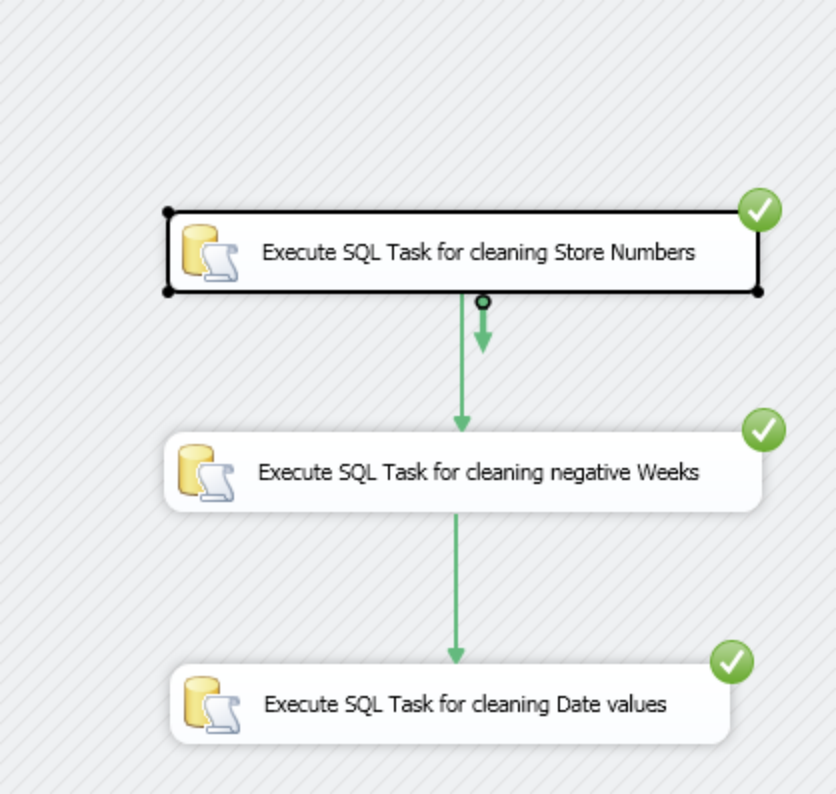
1) **Category_Count:**

Below statements are used in SSIS SQL task to clean Category_Count table for removing invalid store number rows, weeks with negative values and invalid dates.

DELETE from [dbo].[Category_Count] where Store_Number not like '%[0-9]%'

DELETE from [dbo].[Category_Count] where Week like '-%'

DELETE from [dbo].[Category_Count] where Date not like '%[0-9]%'



2) **Demographics:**

Below SQL statement is used in SSIS SQL task to clean Demographics table for removing rows with invalid store numbers, Age_60 values and no store names.

DELETE from [602_Group1-staging-area].[dbo].[Demographic] where Store_Number not in (select Store_Number from [602_Group1-staging-area].[dbo].[Demographic] where Store_Number like '%[0-9]%' and Age_60 <> '.' and Store_Name <> '"')

3) **UPC:**

Below SQL statements are used in SSIS SQL task to clean UPC table to update product description column where there are invalid characters.

UPDATE dbo.UPC SET[Product_Description]= REPLACE([Product_Description], '"', '') WHERE CHARINDEX('"', [Product_Description]) <> 0

UPDATE [602_Group1-staging-area].[dbo].[UPC] set Product_Description = substring(Product_Description,2,len(Product_Description)) where Product_Description like '%' or Product_Description like '~%'*
or Product_Description like '$%'
or Product_Description like '#%'

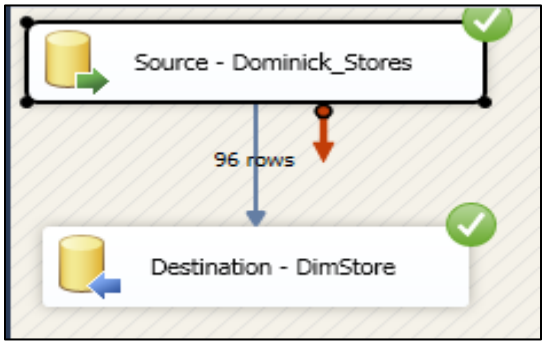
4) **Sales_Product:**

Below SQL statements are used in SSIS SQL task to clean Sales_Product table to update Sale_Type column with valid and business defined values.

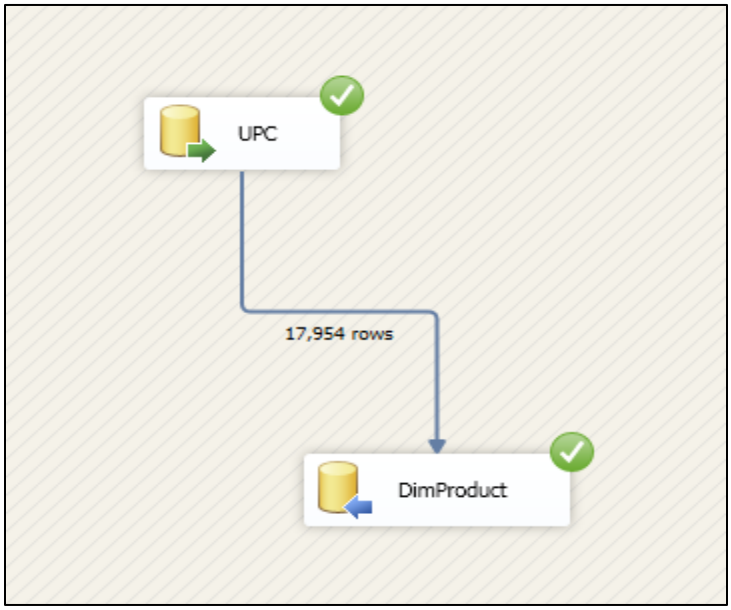
UPDATE FactProductSales set Sale_Type = NULL where (Sale_Type = '' OR Sale_Type = '''')
UPDATE FactProductSales set Sale_Type = 'B' where Sale_Type = '''B'''
UPDATE FactProductSales set Sale_Type = 'C' where Sale_Type = '''C'''
UPDATE FactProductSales set Sale_Type = 'S' where Sale_Type = '''S'''

4.4.2 Transformation of Source data into Dimensions Tables

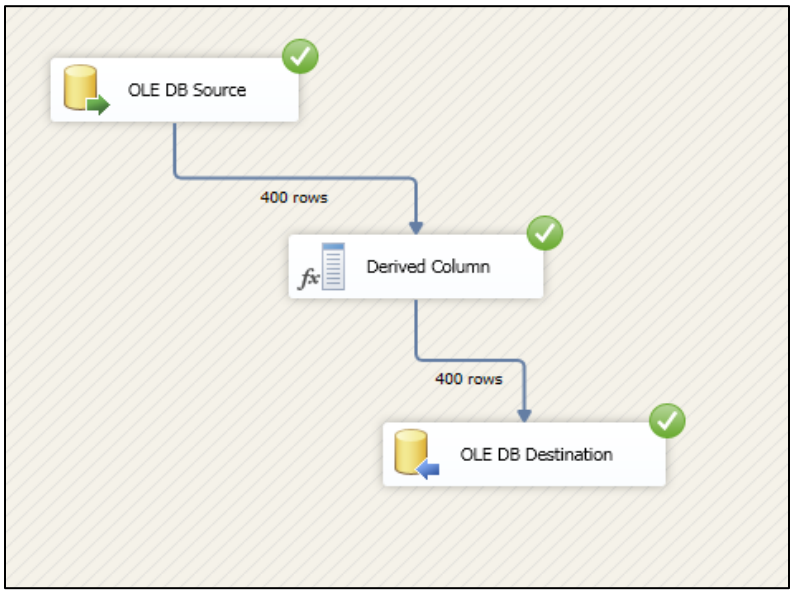
DimStore Dimension Creation



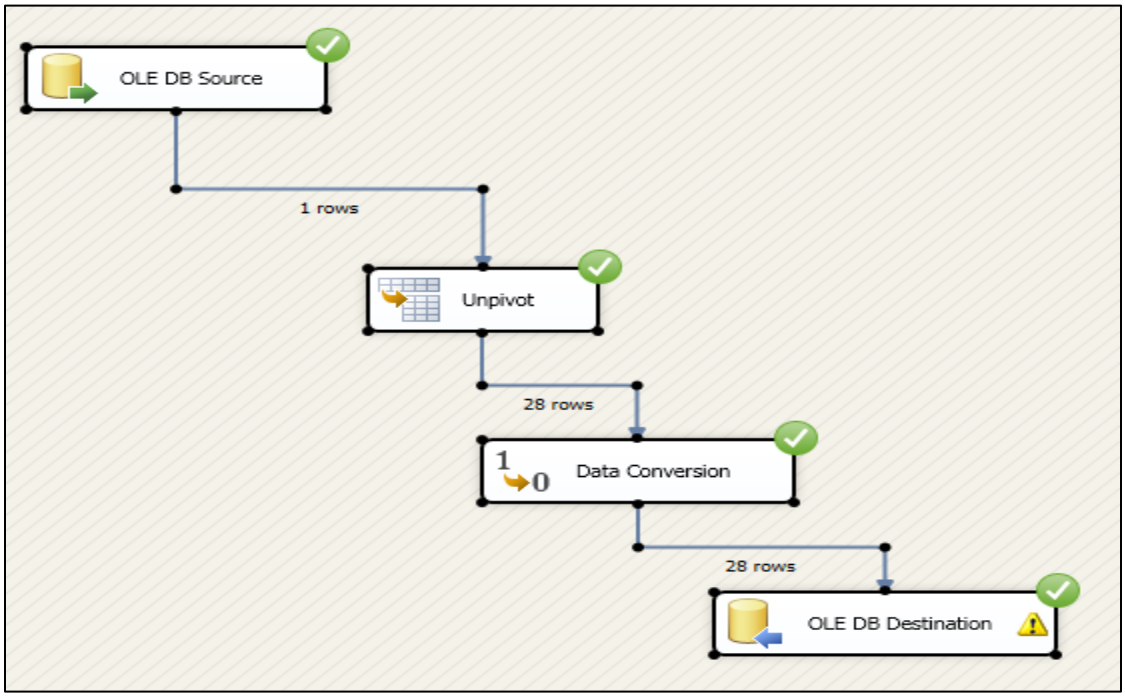
DimProduct Dimension Creation



DimTime Dimension Creation

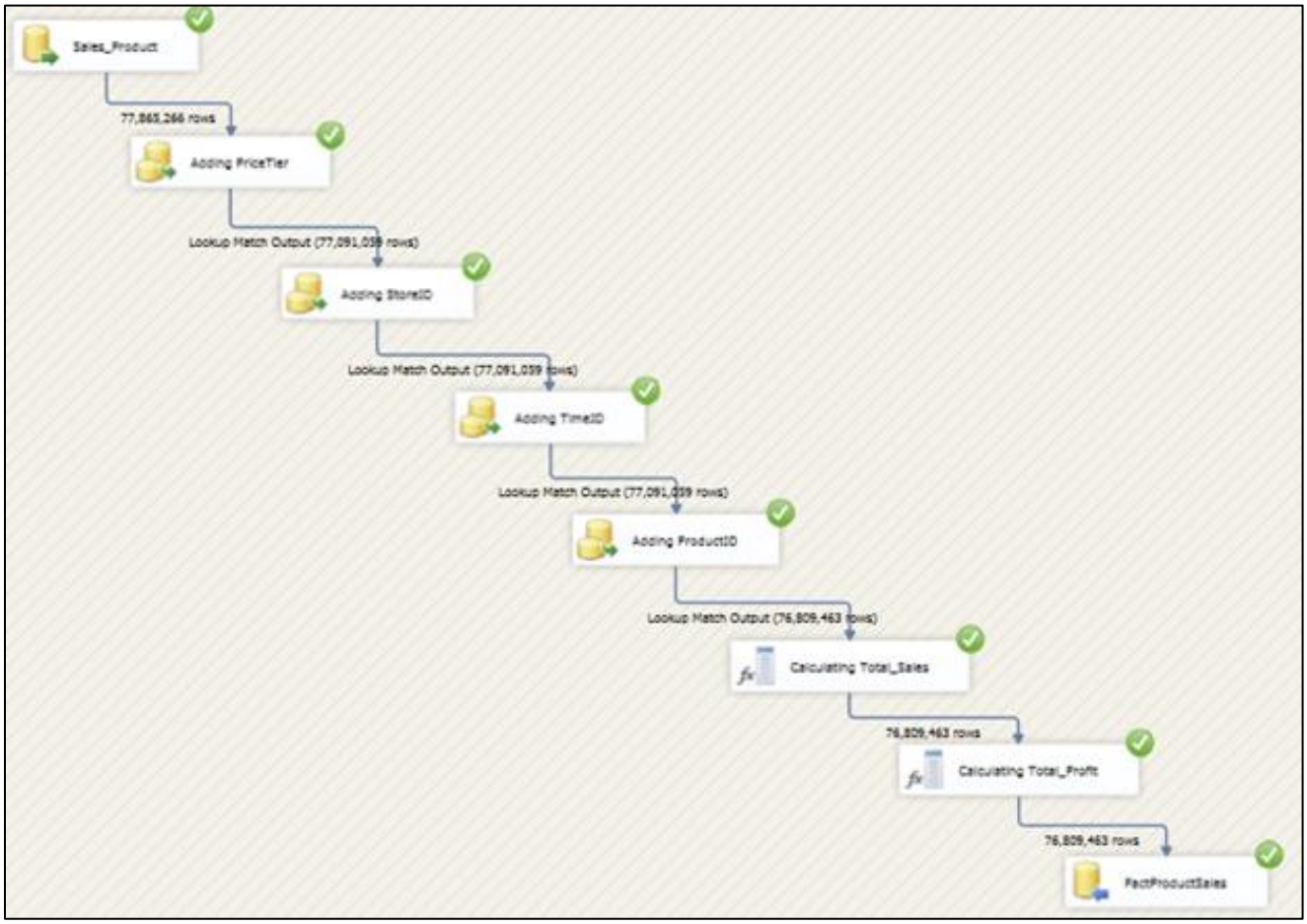


DimProductCategory Dimension Creation

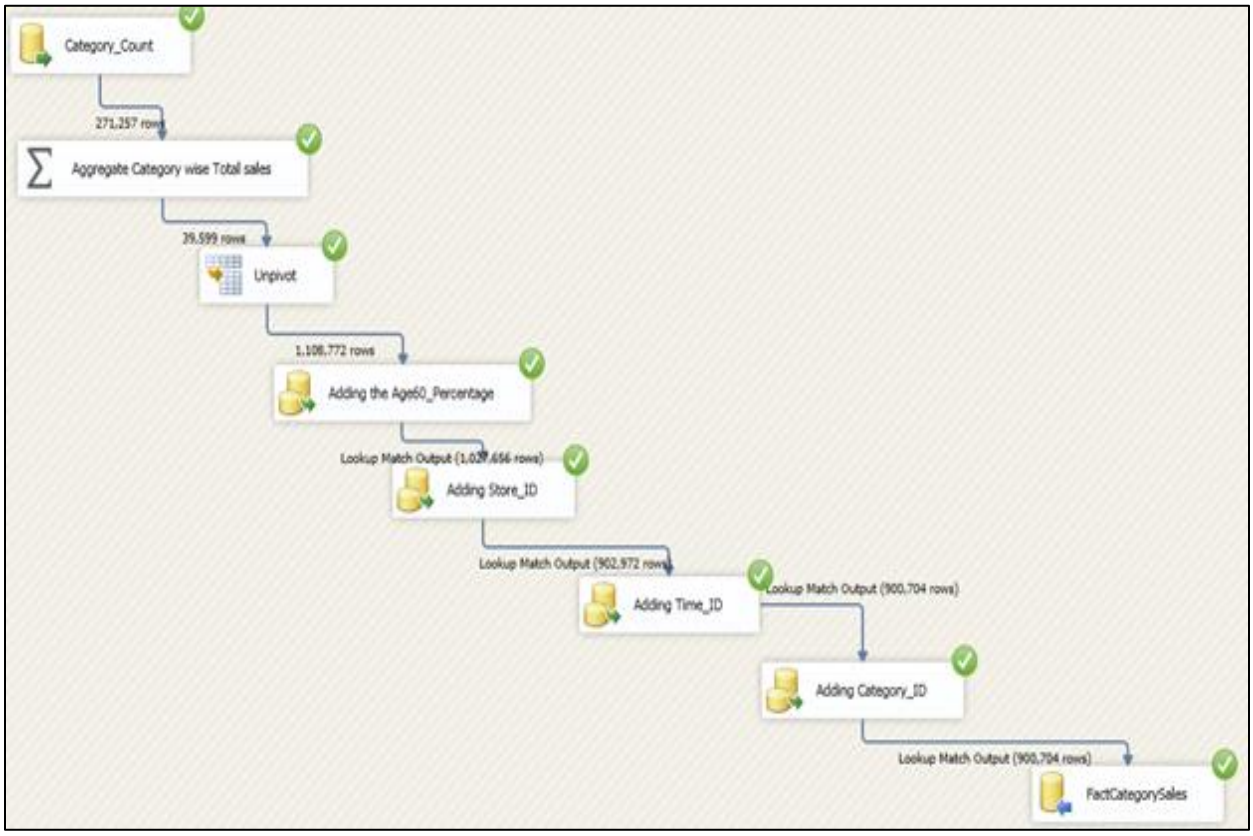


4.4.3 Transformation of Source data into Fact Tables

4.4.3.1 FactProductSales Fact Creation



4.4.3.2 FactCategorySales Fact Table Creation



4.4.4 Data Snapshots in the DW Area:

Table structure (database diagram) for the Category Sales Data Mart

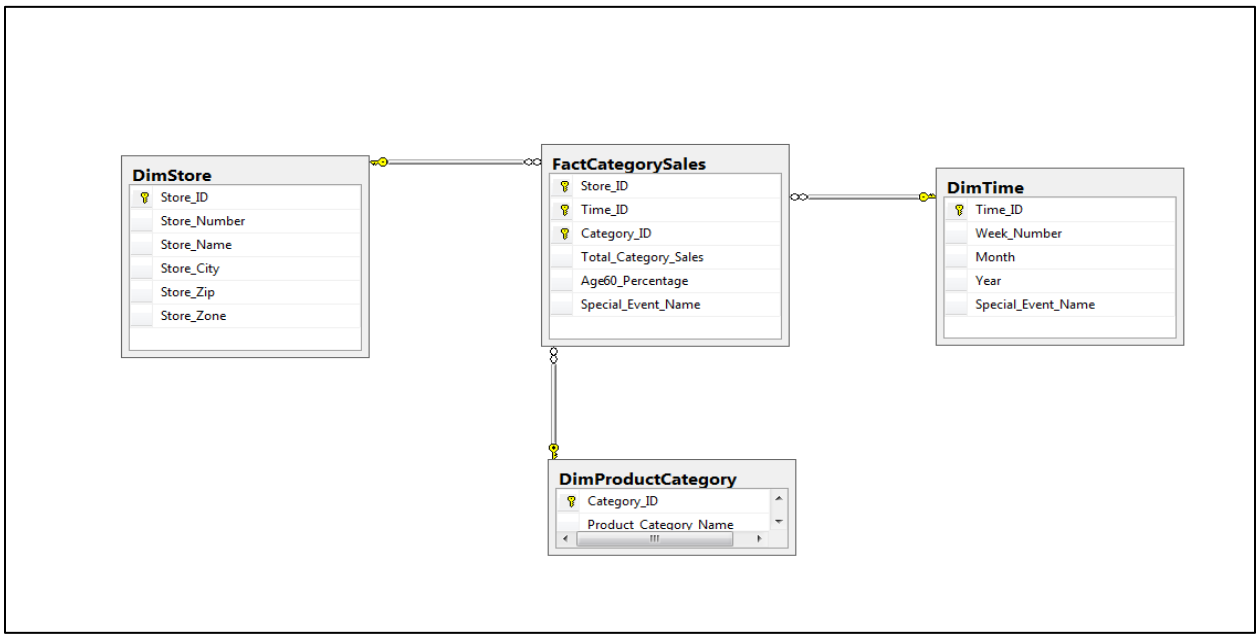
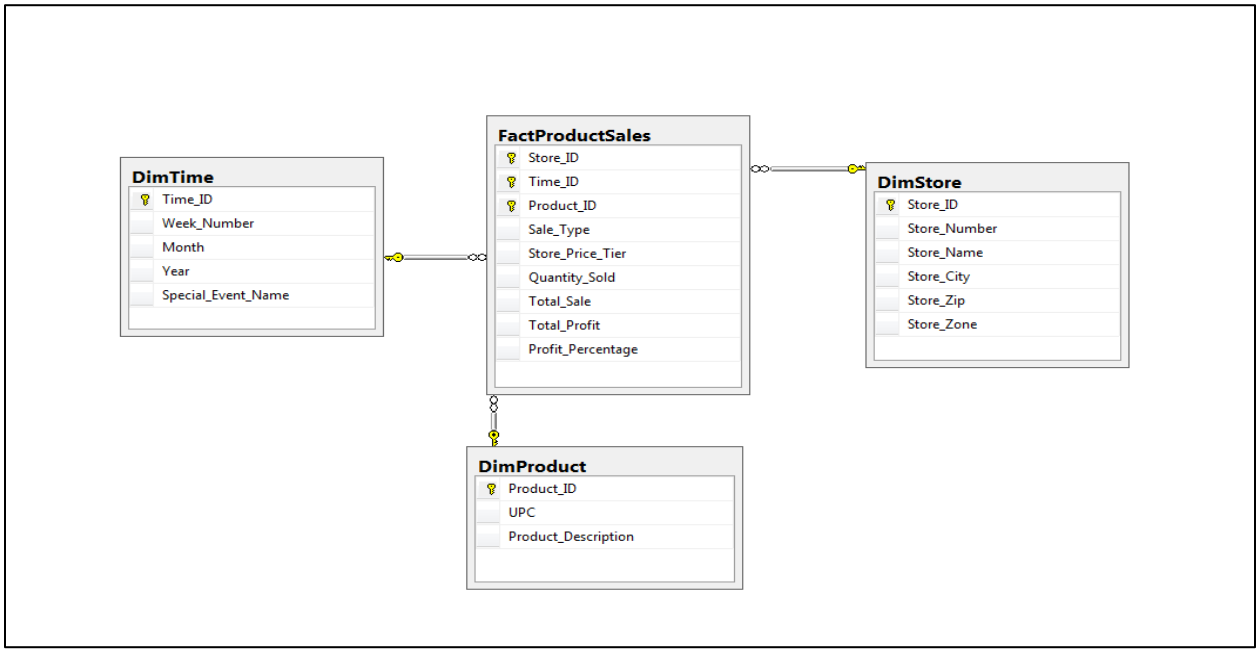


Table structure (database diagram) for the Product Sales Data Mart



5 BI Reporting (using SSRS, SSAS and Report Builder 2012)

5.1 Reporting Plan

Business intelligence reporting (BI reporting) is referred to the process of receiving /providing information or reports to end -users, organizations or applications through a BI software solution. It is typically part of BI software for delivering summarized and structured reports for the analysis or operations performed one or more set of data. The results of BI reporting are generally in the form of actionable results that help the organization or individual in short term, long term tactical and/or strategic decision making. It is also integrated with other application that takes the results to perform any further operation.

Reporting in BI is achieved through the use of myriads of tools and applications that allow organizations to collect internal and external data. This data can then be further analyzed, queried and interactive reports, data visualizations and dashboards can be generated using them. Programming frameworks like the popular D3.js can also be integrated with the data to create interactive real time reports.

Benefits of Reporting

Reporting is a prerequisite of analysis; as such, it should be viewed in light of the goal of making data understandable and ready for easy, efficient and accurate analysis.

- Collecting and presenting data ready to be analyzed, including historical data that can be tracked over time
- Empowering end-users with the knowledge to become experts in their area of business
- Having the underlying figures to back up actions and explain decisions

In our project we make use of SSRS, SSAS, SSAS+SSRS and Report Builder 3.0 to answer business questions. The table below shows the reporting tool used for corresponding business question:

Question Number	Reporting Tool Used
2	SSRS
3	SSAS
5	SSRS plus SSAS
1 & 4	Report Builder 3.0

5.2 Determining target reports that satisfies business questions

Question1: Identify the store price tier that contributes most to overall profit of DFF.

This report is build using Report Builder 3.0. The Product Sales data mart is used to answer this question. Store_price_tier attribute from FactProductSales table is used to group stores. Total_Profit attribute is summed for all the stored grouped based on Store_price_tier. Profit percentage field is calculated based on (Total_Profit for that tier/Sum of Total_Profit for all the stores in DFF). Bar graph is plotted with Store_Price_Tier on x-axis and Total_Profit on y1-axis. Line graph is plotted with Store_Price_Tier on x-axis and Profit percentage on y2-axis.

Question2: Analyze the store wise revenue contribution in the last 3 years.

This report is created using SSRS tool. The product sales data mart is used as primary data source. FactProductSales and DimTime tables are used to derive report attributes. Total_Sale attribute in table FactProductSales is summed grouped by field's store_id and time_id. Corresponding store number and year are derived from dimension tables. Line graph is plotted with store number on x-axis and total_sale for that store on y-axis. Results are filtered to show the data for years 1995, 1996, 1997 only on separate graphs.

Question3: Identify the sales trend in DFF for different product categories over time.

For answering this question, we chose SSAS tool clubbed with Excel pivot table and charts. Category sales data mart is used to answer this question. Tables FactCategorySales, DimTime and DimProductCategory are used as data sources. From FactCategorySales, attribute Total_Category_Sales is used as a measure in the cube. It is measured along dimensions Product category name and week. Finally, using Excel pivot charts, line graph is plotted with week number as x-axis and total_category_sale on y-axis. Results are filtered out for Bakery, Beer and Floral product categories.

Question4: What is the effect of age demographics on the sale of pharmaceutical products?

Report Builder 3.0 is used to answer this question. Category sales data mart is used as the primary source of data. FactCategorySales and DimProductCategory tables are used for data. Attribute Total_Category_Sales is summed in table FactCategorySales with data grouped using Age60_Percentage and CategoryID. This sum is divided by the corresponding store total sale to derive sale percentage. Results are filtered out to derive data for Product category Pharmaceutical only. Bar graph is plotted with Age60_Percentage on x-axis and sale of pharma as percentage of total sale on y-axis.

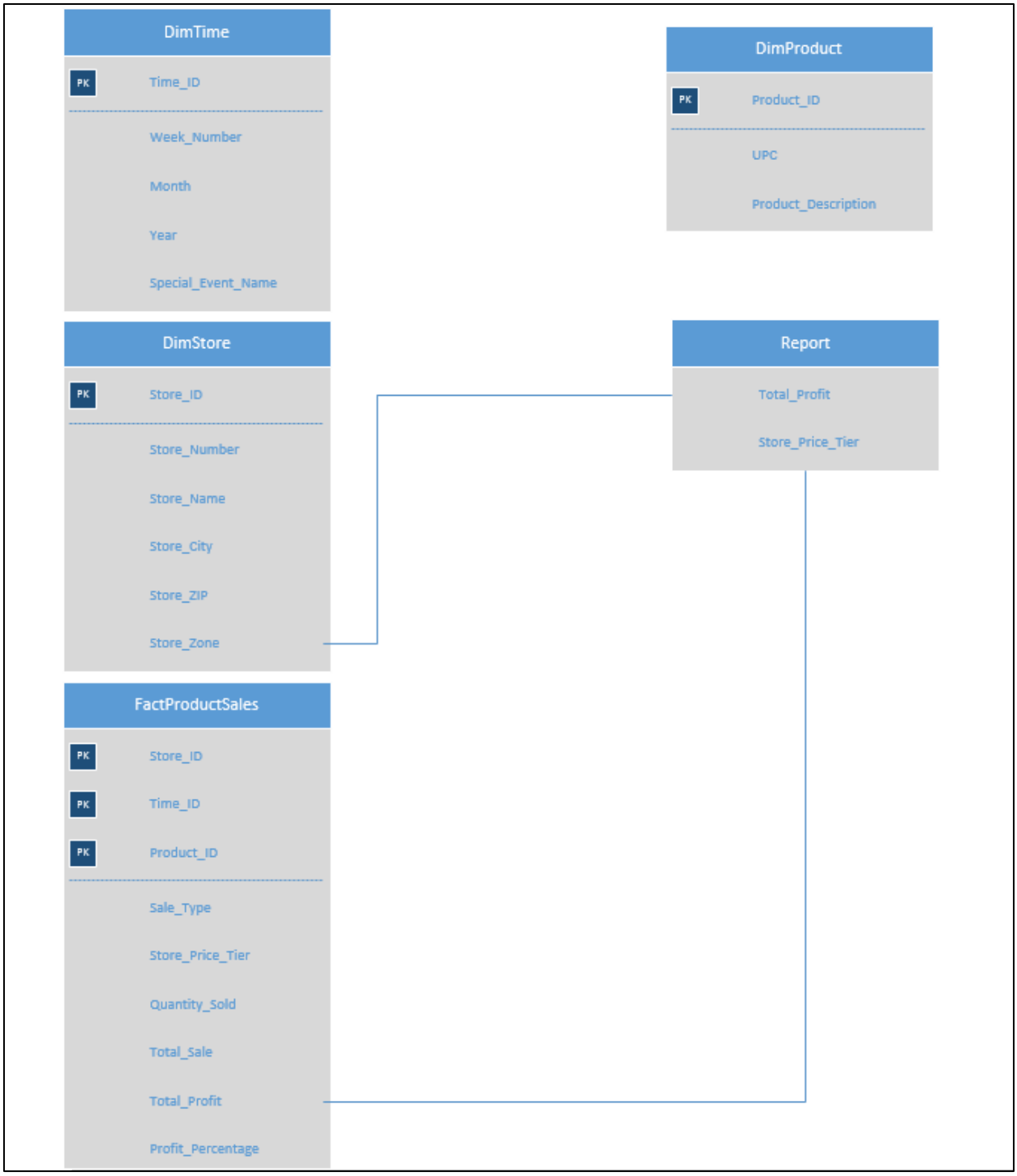
Question5: Comparison of sale of products when there is a bonus buy offered, coupon discount offered, simple price reduction, and when no discount is offered.

An SSAS cube is created and using SSRS, report is designed with SSAS cube as data source. Product sale data mart is used as primary data source. Tables FactProductSales, DimProduct and DimTime are used. SSAS cube is built with Quantity_sold attribute in table FactProductSales as measure with dimensions as Sale_type and Time_ID. Obtained cube is used as data source to design SSRS report. A bar chart is plotted with sale_type on x-axis and average quantity_sold per week on y-axis (quantity_sold / no. of weeks).

5.3 Mappings from all Independent Data Marts to the Report Attributes

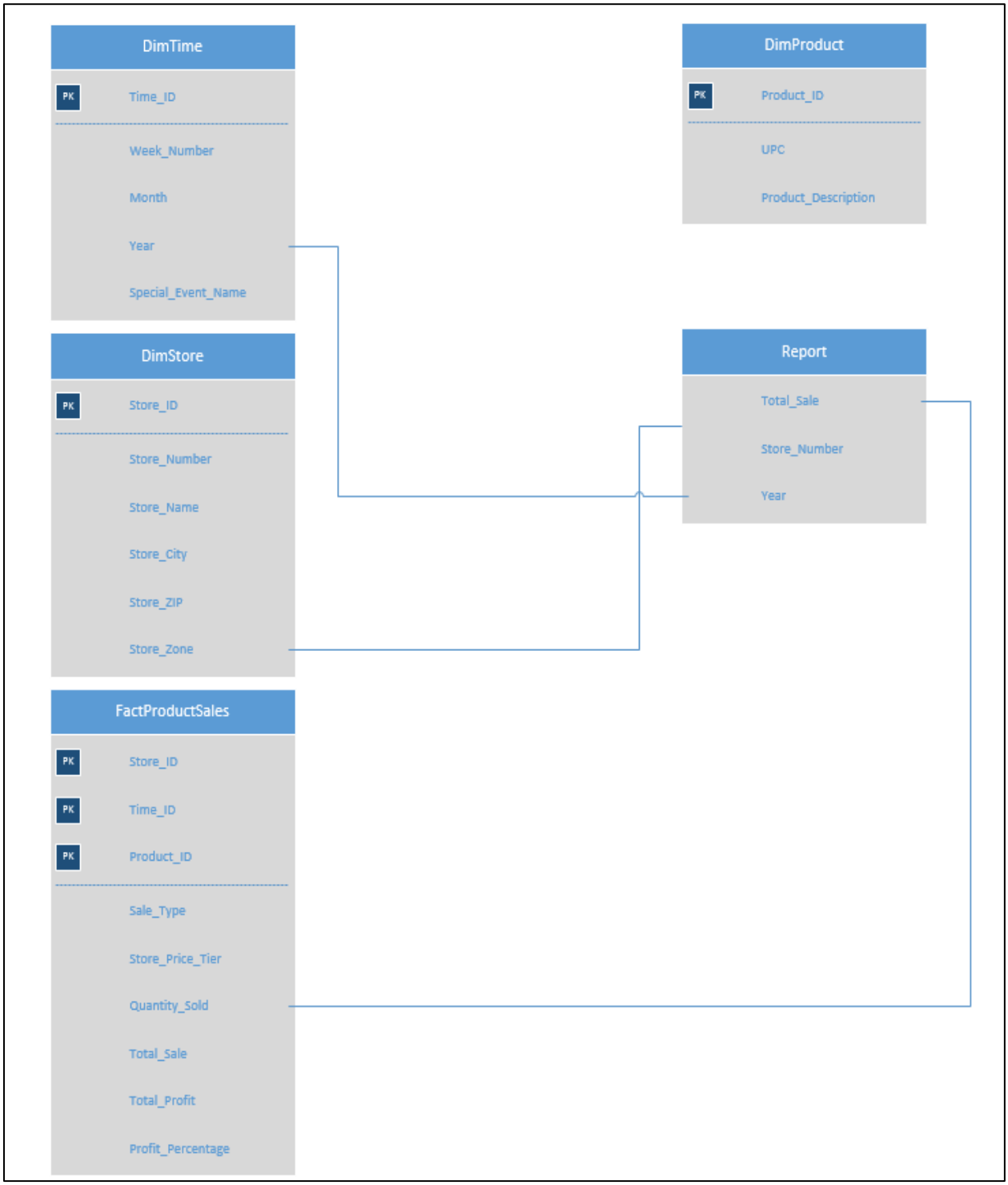
Question 1

Identify the store price tier that contributes most to overall profit of DFF.



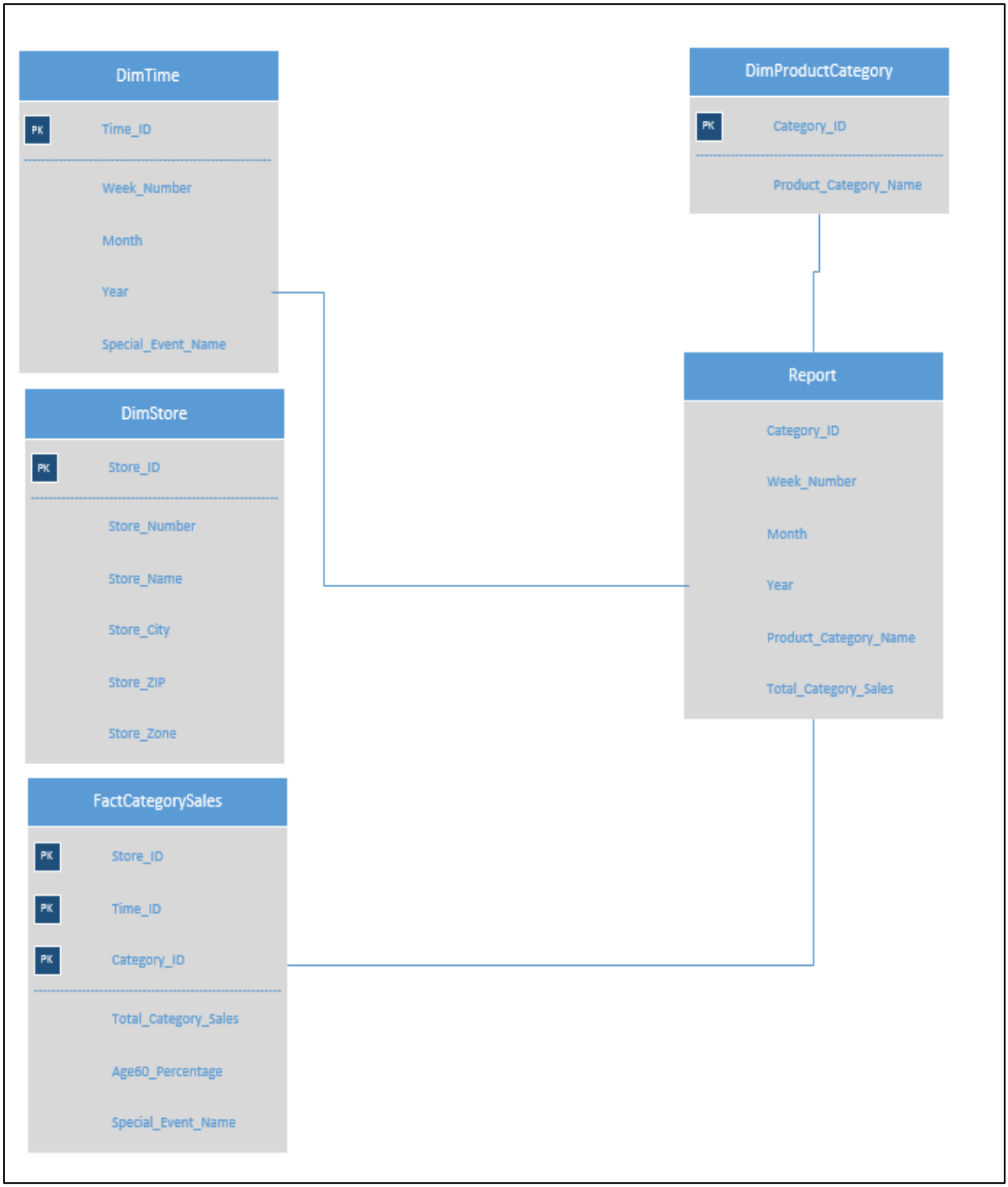
Question 2

Analyze the store wise revenue contribution in the last 3 years.



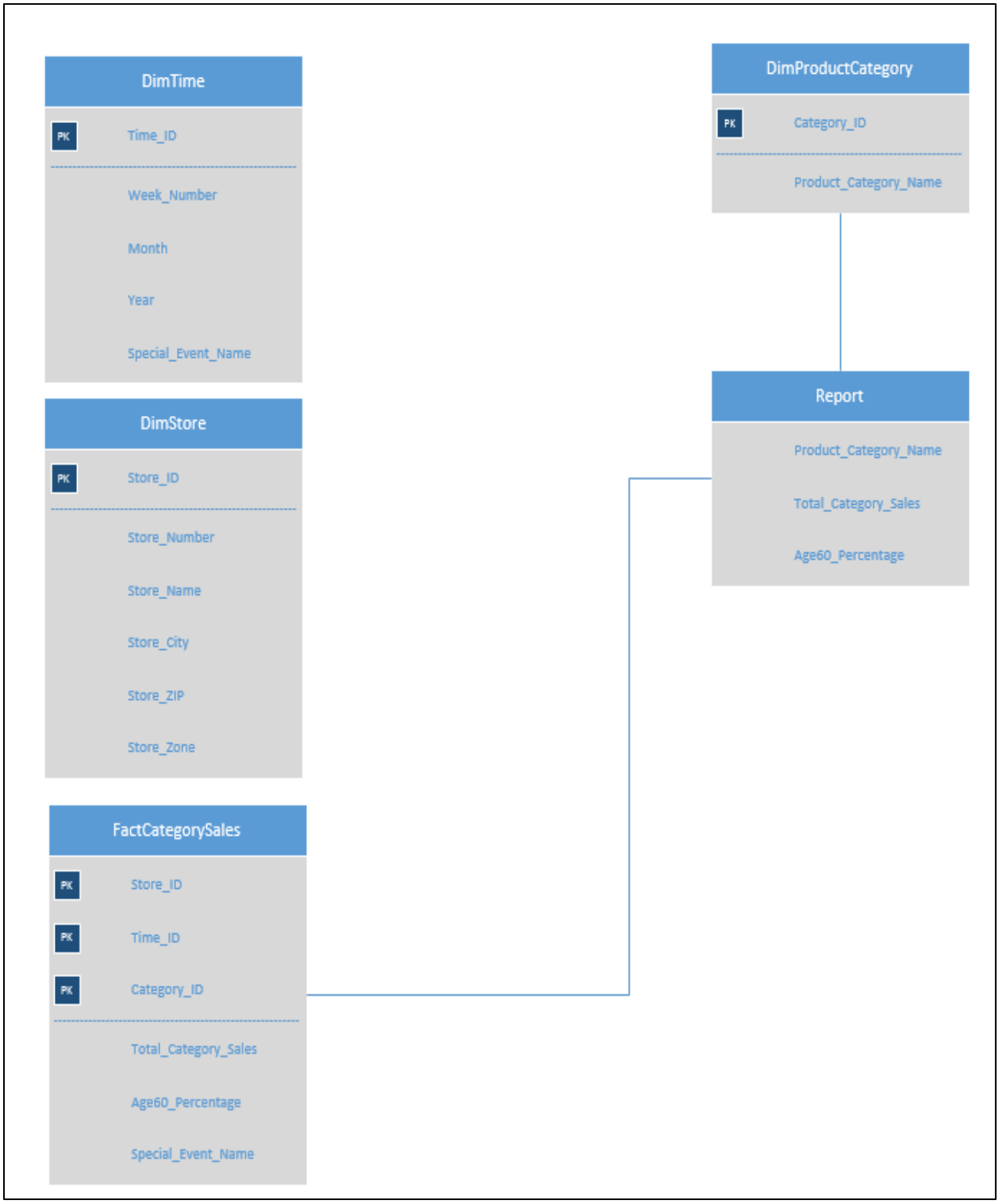
Question 3

Identify the sales trend in DFF for different product categories over time.



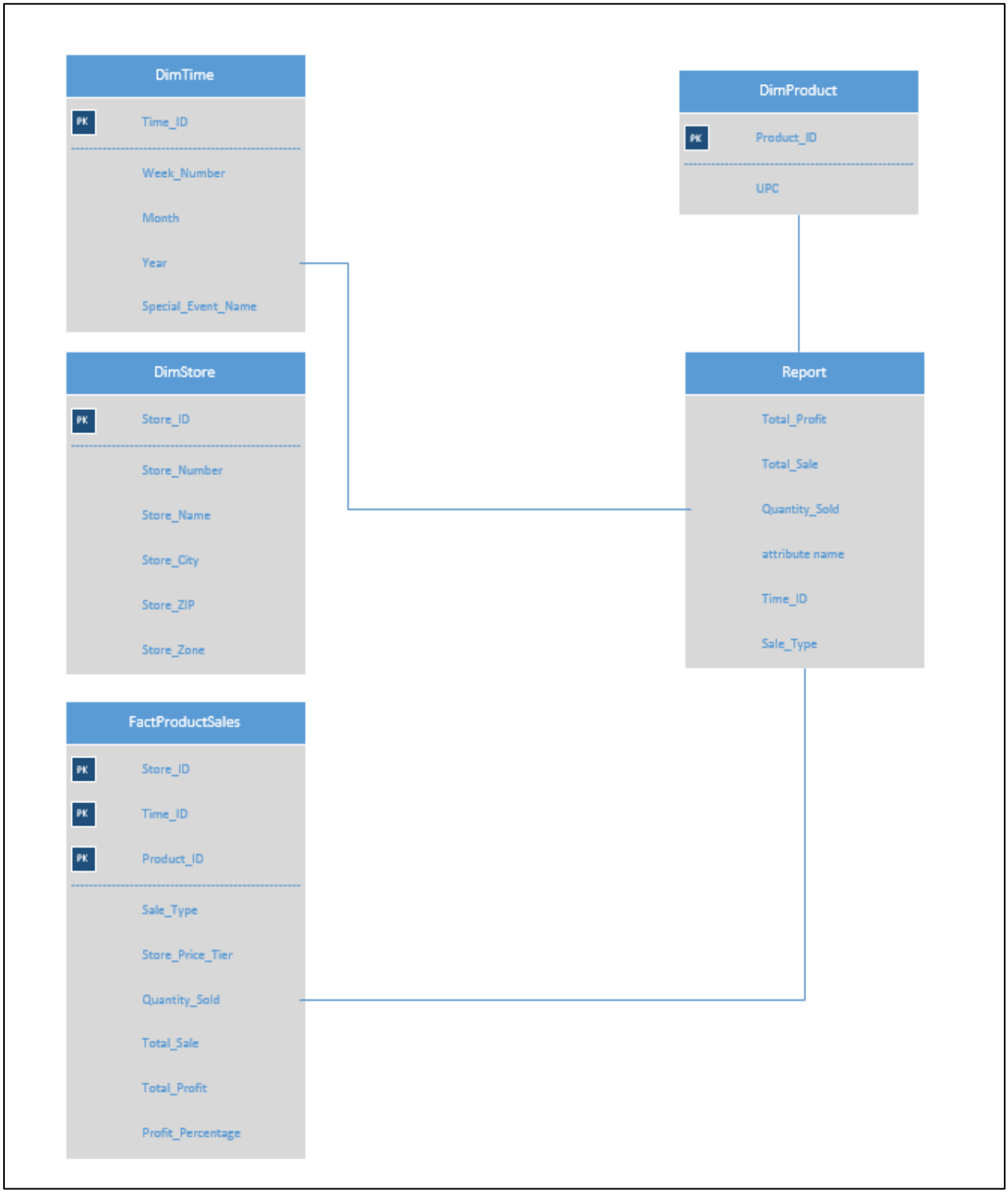
Question 4

What is the effect of age demographics on the sale of pharmaceutical products?



Question 5

Comparison of sale of products when there is a bonus buy offered, coupon discount offered, simple price reduction, and when no discount is offered.



5.4 Report Building using Report Builder 3.0 for Question 1

Question 1

Identify the store price tier that contributes most to overall profit of DFF.

Final Deployed Report

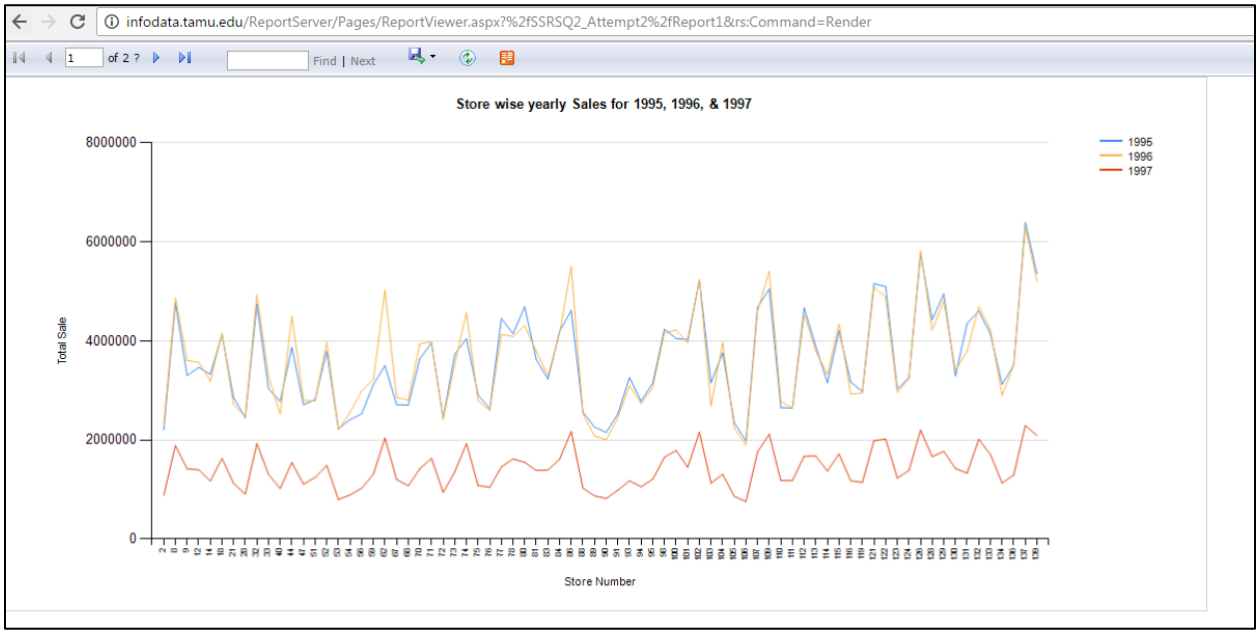
Conclusion:

We made use of Report Builder 3.0 in order to answer this question. In this business question, we are identifying the price tier that contributes most to the overall profit of Dominick. The above report shows the overall profit contributed by individual price tiers. The primary x-axis shows the four price tiers while the primary y-axes show the net profit both as absolute and percentage of total profit. On similar terms, other relevant questions related to price tier can be answered and their relative performance can be gauged.

5.5 Report Building using SSRS for Question 2

Question 2
Analyze the store wise revenue contribution in the last 3 years.

Final Deployed Report



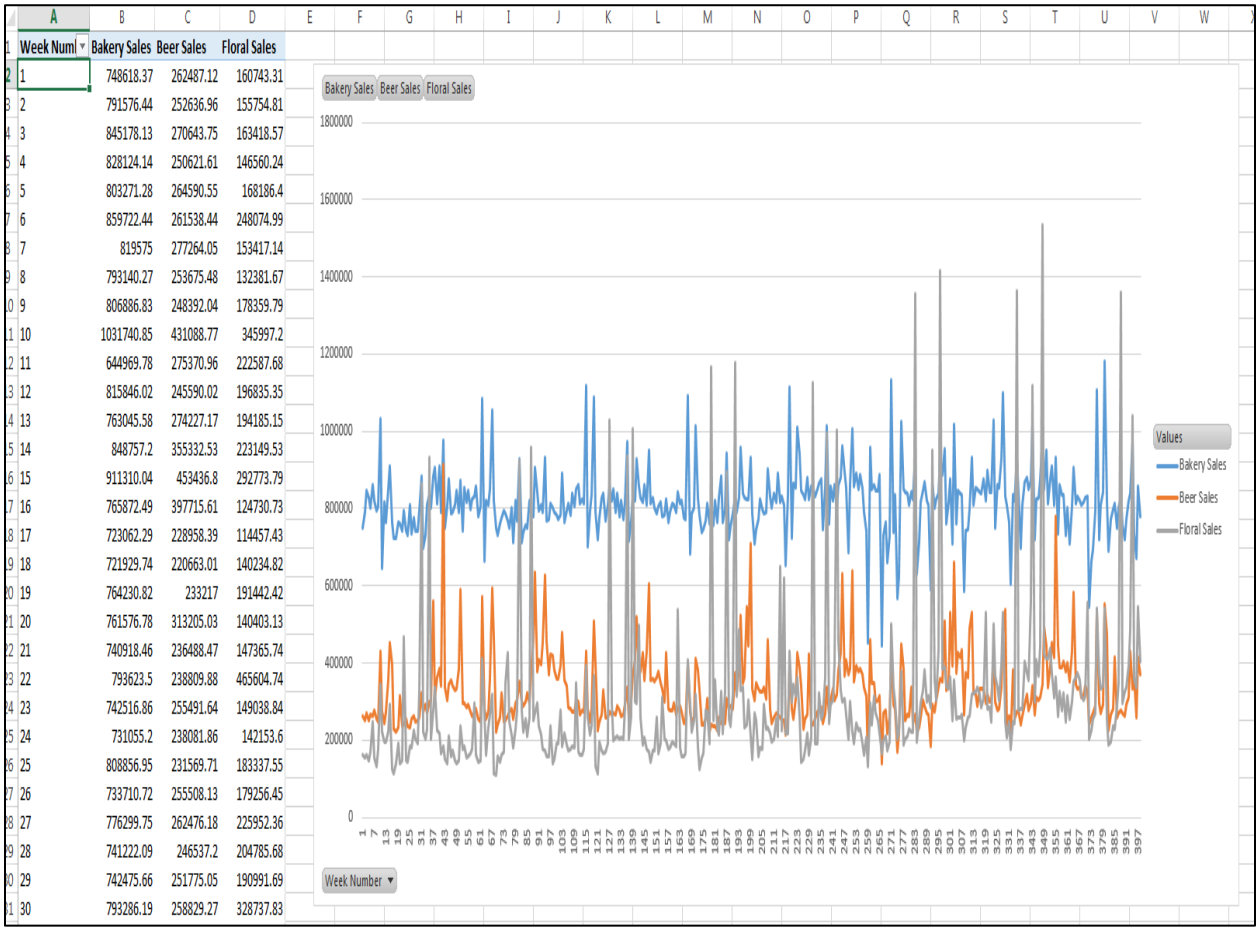
Conclusion:

Question 2 has been answer with the help of SSRS. The final report above shows the total sales on primary y-axis plotted against store number for three years. The sales in stores over a period of time is clearly shown and the performance of any specific store can be analyzed. Hence, the main idea of this question to identify stores with stagnant and slow revenue growth is fulfilled.
The business purpose of this question to narrow down on some of the reasons why some stores are under performing continuously over a period of time can be identified with the help of the report. Thorough analysis of this store wise report, can help DFF drill down to some of the inherent or external reasons for underperforming stores and brainstorm possible solutions to improve revenue.

5.6 Report Building using SSAS for Question 3

Question 3
Identify the sales trend in DFF for different product categories over time.

Final Analysis using Pivot Chart



Week Num	Bakery Sales	Beer Sales	Floral Sales
1	748618.37	262487.12	160743.31
2	791576.44	252636.96	155754.81
3	845178.13	270643.75	163418.57
4	828124.14	250621.61	146560.24
5	803271.28	264590.55	168186.4
6	859722.44	261538.44	248074.99
7	819575	277264.05	153417.14
8	793140.27	253675.48	132381.67
9	806886.83	248392.04	178359.79
10	1031740.85	431088.77	345997.2
11	644969.78	275370.96	222587.68
12	815846.02	245590.02	196835.35
13	763045.58	274227.17	194185.15
14	848757.2	355332.53	223149.53
15	911310.04	453436.8	292773.79
16	765872.49	397715.61	124730.73
17	723062.29	228958.39	114457.43
18	721929.74	220663.01	140234.82
19	764230.82	233217	191442.42
20	761576.78	313205.03	140403.13
21	740918.46	236488.47	147365.74
22	793623.5	238809.88	465604.74
23	742516.86	255491.64	149038.84
24	731055.2	238081.86	142153.6
25	808856.95	231569.71	183337.55
26	733710.72	255508.13	179256.45
27	776299.75	262476.18	225952.36
28	741222.09	246537.2	204785.68
29	742475.66	251775.05	190991.69
30	793286.19	258829.27	328737.83

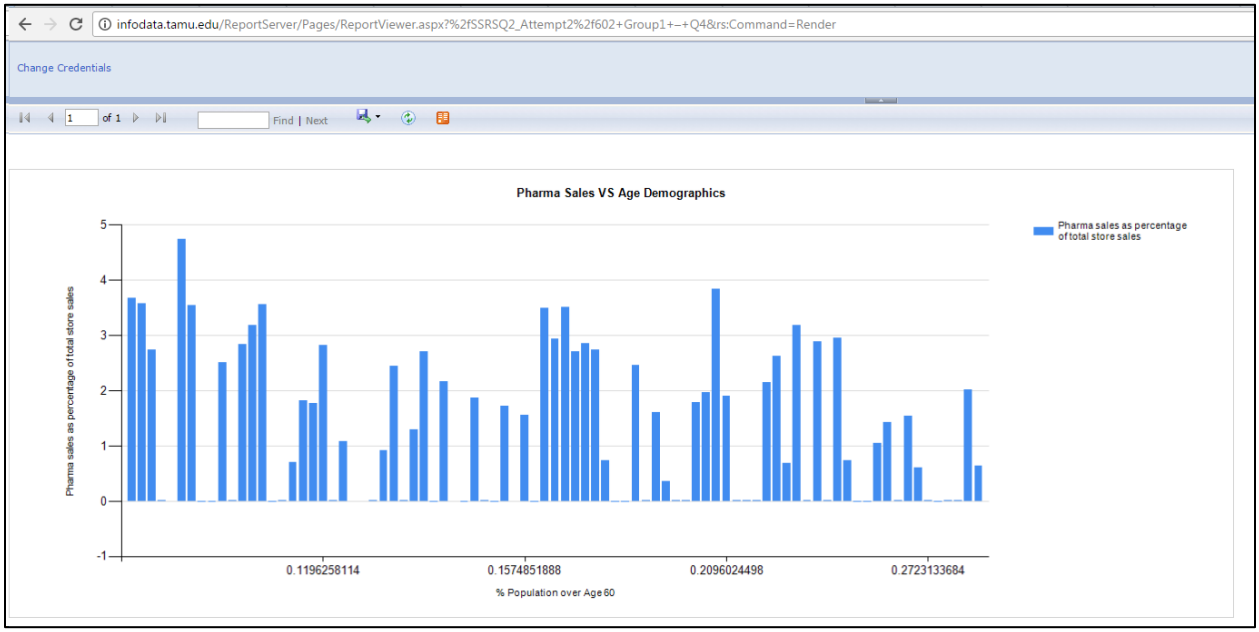
Conclusion:

We have answered question 3 using SSAS and Excel. The above chart shows the sales data for three categories namely beer, floral and bakery over a period of time (in weeks). The sales are represented on the y-axis while progressive time in weeks in shown on the x-axis. The report can be effectively used to identify sales trends for various categories over time. Likewise, sales peaks for other categories of products can also be evaluated and analyzed to answer other relevant and specific business questions. Based on this trend analysis of sales data of DFF, timely decisions like price cuts and stock increase for a particular category of items can be implemented.

5.7 Report Building using Report Builder 3.0 for Question 4

Question 4
What is the effect of age demographics on the sale of pharmaceutical products?

Final Deployed Chart



Conclusion:

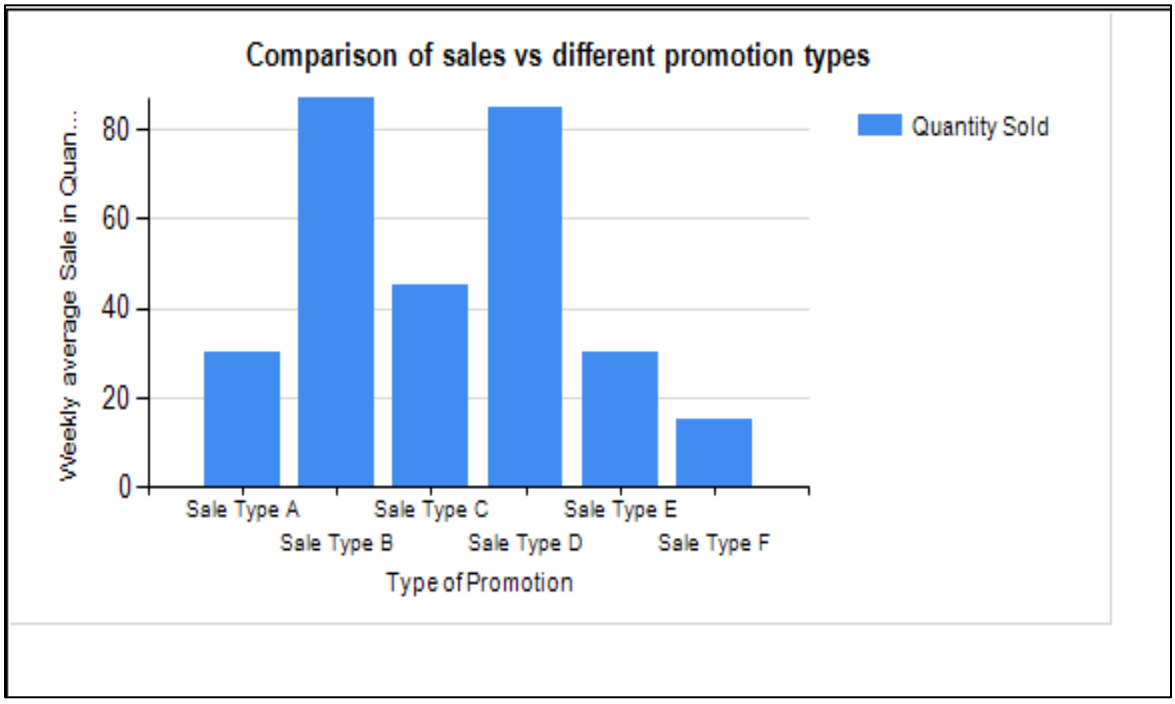
We have answered this question using Report Builder 3.0. The final report above shows the pharma sales as percentage of total sales against the percentage of population over the age of 60. It can be seen that the pharma sales have random spikes and have no specific correlation with the percentage of population above 60. As you can see, the contribution of pharmacy sales in store having 0.058 % demographic is greater that the sales in store having 0.28% demographic. Thus, we can conclude from the sample data that there is no direct relationship between pharmacy sale and Age60 demographic. Other factors weigh in when we consider the sale contribution of analgesics.
This question can help DFF in carefully considering the decision of expansion of analgesic warehouses or marketing campaigns, by not basing solely on demographic data, but also to investigate and discover other significant factors in pharmacy product sale. Such a careful analysis can avoid investment in wrong areas by throwing light on apt markets.

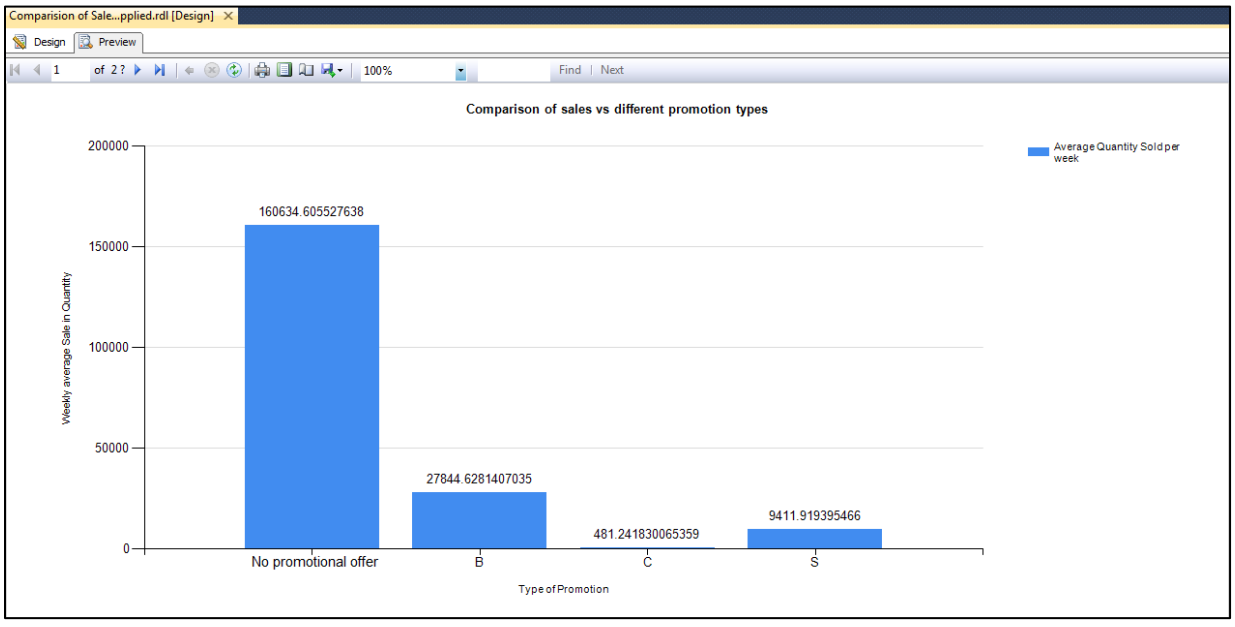
5.8 Report Building using SSRS on top of SSAS for Question 5

Question 5

Comparison of sale of products when there is a bonus buy offered, coupon discount offered, simple price reduction, and when no discount is offered.

Final Deployed Report





Conclusion:

We make use of SSRS on top of SSAS to solve this business question. The report shows the movement of Canned Diet Pepsi in store 5 when there is any type of sale offered on it and compares it to sales during no – promotion periods. The x-axis shows the type of sale offered – B, C, S or no promotions and on y-axis is the average weekly movement quantity. The report clearly shows a sharp rise in movement during sale periods than during the no sale periods. This analysis can help business in planning the sale periods on a particular product. In addition, depending on the amount of movement rise trend during sale and no sale periods, one can plan the stocking, distribution and strategic marketing of products.

6 References

- Montgomery, A. L. (1997). Creating Micro-Marketing Pricing Strategies. *Marketing Science*, 4-6.
- Wagner A. Kamakura, W. K. (2007). Chain-wide and store-level analysis for. *Journal of Retailing*, 169.
- The Demographic Variables That Affect a Business by Rick Suttle (http://smallbusiness.chron.com/demographic-variables-affect-business-24344.html)
- Nevo, Aviv and Hatzitaskos, Konstantinos, "Why Does the Average Price of Tuna Fall During Lent?" (August 2005). NBER Working Paper No. W11572.
- Srinivasan, Shuba, Koen Pauwels, Dominique Hanssens, and Marnik Dekimpe, "Who Benefits from Price Promotions?" *Harvard Business Review*, 80 (September 2002): 22-23.
- https://www.techopedia.com/definition/30217/business-intelligence-reporting-bi-reporting