Consulting Project

Design and Implementation of a Data Warehouse

for a Retail Store with Store-level Data

Report 1: Requirements Gathering

Prepared By:
Sri Sai Kowshik Reddy Boyalla
Mithilesh Menakuru
Prajakta Ingle
Varshitha Ravikumar

Index

Introduction **2**

Data analysis **3**

Data Overview 3

Metadata Description 6

ERD Diagram 12

Industry Research **13**

Emerging trends in retail pricing practice: implications for research 13

An Examination of the extent, causes, and efforts to address retail out-of-stocks 13

Communication and Promotion Decisions in Retailing: A Review and Directions for Future Research 14

A latent class segmentation analysis of e-shoppers 16

Proposed Business Questions **17**

Prioritization of Business Questions: **27**

References **28**

Introduction

Dominick's Finer Foods (DFF) founded in 1918 by a Sicilian immigrant Dominick DiMatteo was a major grocery store chain in the Chicago area. The company evolved from a small deli to become Chicago's second-largest supermarket chain. Over the decades, DFF expanded significantly and they underwent several changes in ownership which impacted DFF's performance.

They were one of the first to feature in-store delicatessens and frozen food sections and pioneered online grocery ordering well before e-commerce became the norm. They partnered with Starbucks to install coffee bars in stores to enhance the customer's in-store experience. DFF's success was largely credited to its ability to stay ahead of the market trends. The company also employed the latest technologies to consistently improve efficiency and customer service, staying committed to meeting customers' needs and requirements.

However, the company faced significant challenges after it was acquired by Safeway Inc. as the changes implemented were unfamiliar to the long-time local customers and impacted their shopping experience. They replaced the familiar local products with Safeway's in-house brands disrupting the shopping experience of long-term customers. Additionally, the entry of competitors such as Target, Walmart, Kroger, etc. resulted in a drop in DFF's market share stressing the impact of integrating local stores into national operations [1].

The primary objective of this project is to design and implement a data warehouse for DFF using the store-level data collected from 1989-1994. This involves integrating the data from various sources ensuring data consistency to provide a consolidated view of the company's operations. The data warehouse will facilitate historical data analysis to understand the patterns in customer behavior, sales trends, and marketing strategies. This information is essential to understand the factors that contributed to DFF's success and downfall.

Furthermore, The data warehouse will enable the creation of reports tailored to the organization and its stakeholders. These reports will provide valuable insights into various aspects of the business enabling targeted marketing and strategic decision-making. Successful implementation of this project will allow us to understand and optimize the operations.

Data analysis

Data Overview

The DFF's historical data is divided into two categories - General Files and Category Specific Files.

A) General Files:

1) Customer Count File: This dataset provides daily and store-specific customer traffic information along with the sales data, and coupon usage (if any) data for different product categories at Dominick's Fine Foods. Each row or record in the file represents a single day's data related to the number of customers and a summary of their purchases for a specific store. In addition to customer count, the dataset tracks coupon usage and sales revenue for a wide range of product categories, from groceries to pharmacy, cosmetics, alcohol, bakery, frozen items, and more.

The data in the Customer count file is mainly related to 3 things -

1. **Customer Traffic:**
    - There is a column (field) called CUSTCOUNT available in the Customer count file. This field indicates the total number of customers who made purchases on a particular given day.
    - The other fields DATE and STORE represent the particular date on which sales happened and the store number (at which store sales occurred) respectively
2. **Sales Data:**
    - This file contains multiple fields or columns that capture sales in dollars for different departments such as bakery, beer, cosmetic, pharmacy frozen foods, etc. For instance, the columns in the data file are - BAKERY, BEER, COSMETIC, PHARMACY, and FROZEN which capture related sales information per day respectively.
    - Each department-specific column indicates the total sales in dollars for that day and specific store.
3. **Coupon Usage:**

- Not all departments are offering coupons but most of the departments are offering coupons for customers and in the dataset there are multiple fields to capture the coupon usage for various departments.
- For instance, columns such as BAKCOUP, COSMCOUP, LIQCOUP, and PHARCOUP indicate how many coupons were redeemed in those departments.
- MANCOUP tracks the number of manufacturer coupons redeemed.

This dataset contains sales and coupon data broken down into categories such as bakery, frozen items etc. Time-series data is recorded per store and day, allowing for temporal analysis.

2) Demographics file: Dominick's Fine Foods contains detailed store-specific demographic data derived from the **1990 U.S. Census** for the **Chicago metropolitan area**. This data provides insight into the socio-economic characteristics of the population in the trading area of each store, helping to understand the customer base and tailor store offerings accordingly.
The data includes:

- **Age distribution** (e.g., % population under 9 or over 60)
- **Ethnic composition** (e.g., % Blacks and Hispanics)
- **Education levels** (e.g., % College Graduates)
- **Income distribution** (e.g., Median income, income variance)
- **Household characteristics** (e.g., household size, % single-person households)
- **Employment data** (e.g., % working women, % unemployed)
- **Homeownership and home value** (e.g., % households with mortgages, mean household value)
- **Shopping behavior** (e.g., shopper types like "Avid Shoppers", "Hurried Shoppers")

This data was processed by **Market Metrics** to provide a demographic profile for each store's trading area, which can be leveraged for various analytical purposes, such as pricing strategies, shelf management, product placement, and understanding customer behavior.

B) Category Specific Files:

1) UPC Files: The UPC datasets contain detailed information related to the products sold at Dominick's Fine Foods, identified by their UPC codes. The data includes product-specific details such as category (commodity code), item code, product description, size and the number of items per case. This information helps to track individual products across the store's inventory system and facilitates better organization and sales tracking at the item level.

Some important columns in the dataset are:

1. **UPC Code:** The last 5 digits of the code identify the specific product, and the preceding digits represent the manufacturer. This allows for easy and fast identification of both the product and manufacturer within the dataset.
2. **Commodity Code (COM_CODE):** A classification system used by Dominick's to categorize products in different sections and departments. While multiple commodity codes may exist within a single dataset (file), a single commodity code will not be repeated across different files.
3. **Item Code (NITEM):** This is a product tracking number across UPCs. If 2 different UPCs have the same item code, it indicates the one is a newer version of the other (though this system is not foolproof).
4. **Description:** This product description field includes various special characters to denote specific attributes:
    - # - Indicates that the product is available in Combo stores (Stores with a pharmacy)
    - < - Denotes trial size products (though this is not always accurate)
    - ~ - Marks discontinued products
    - $ and * - Have no specific meaning
5. **Case:** The number of items in a case delivered by the manufacturer. This field is mainly used for inventory purposes but not visible to customers.

2) Understanding of the Movement Data by UPC:

The movement data files (named wxxx, where xxx is a category acronym) provide weekly sales data at the store level for each unique product (UPC) across various categories. This data includes sales, price, quantity, and other relevant metrics that can be used to analyze sales performance and retail strategy over time.

Key aspects to understand about this data include:

1. **UPC Code as a Key**: The **UPC** is a unique identifier for products, which is crucial when merging data with other files (e.g., product descriptions or categories).

2. **Price, Quantity, and Movement**:
   - Sometimes products are sold in bundles (e.g., "3 for $2"). The **qty** variable represents the bundle size, and the **move** variable represents the actual number of items sold (not the number of bundles).
   - To calculate total sales dollars, the formula is: Sales=Price×MoveQty\text{Sales} = \frac{\text{Price} \times \text{Move}}{\text{Qty}}Sales=QtyPrice×Move
3. **Profit**: This variable represents the gross margin for each product sold. It does not directly reflect replacement or last transaction costs but an average acquisition cost (AAC), which adjusts based on historical inventory and current purchases.
4. **Sales**: This is a promotional flag that tracks whether the product was sold with a promotion (Bonus Buy, Coupon, or Price Reduction). However, it's noted that this field may not always be set consistently.
5. **Data Validation**: The **OK** variable is a flag indicating if the data for a particular week is valid. If the flag is set to 0, that data should not be used in analysis.

Metadata Description

1. Customer Count File:

Column Name	Description	Type	Length
DATE	Date of the observation. The date on which sales occurred at DFF.	Character	6
WEEK	Week number of the observation.	Numeric	8

STORE	Unique store identifier.	Numeric	8
CUSTCOUN	Total number of customers on a specific day at a specific store.	Numeric	8
BAKERY, BEER, GROCERY, COSMETIC, PHARMACY, FROZEN, etc.	Sales in dollars for various specific product categories.	Numeric	8
BAKCOUP, COSMCOUP, GROCCOUP, PHARCOUP, FROZCOUP, etc.	Coupons redeemed for specific product categories at specific stores.	Numeric	8

2) Demographic File:

Column Name	Description	Data Type	Example Value
AGE9	% Population under age 9	Float	12.5
AGE60	% Population over age 60	Float	15.3
ETHNIC	% Blacks & Hispanics	Float	45.2
EDUC	% College Graduates	Float	25.1
NOCAR	% With No Vehicles	Float	18.9

INCOME	Log of Median Income	Float	10.5 (log scale)
INCSIGMA	Std dev of Income Distribution (Approximated)	Float	1.1
HSIZEAVG	Average Household Size	Float	3.2
HSIZE1	% of households with 1 person	Float	23.5
HSIZE2	% of households with 2 persons	Float	31.0
HSIZE34	% of households with 3 or 4 persons	Float	26.4
HSIZE567	% of households with 5 or more persons	Float	10.2
HH3PLUS	% of households with 3 or more persons	Float	36.6
HH4PLUS	% of households with 4 or more persons	Float	18.3
HHSINGLE	% of households with 1 person	Float	21.4
HHLARGE	% of households with 5 or more persons	Float	8.1
WORKWOM	% Working Women with full-time jobs	Float	47.5
SINHOUSE	% Detached Houses	Float	64.2
DENSITY	Trading Area in Sq Miles per Capita	Float	2.1
HVAL150	% of Households with Value over $150,000	Float	33.2

HVAL200	% of Households with Value over $200,000	Float	18.7
HVALMEAN	Mean Household Value (Approximated)	Float	195.6 (in $1,000s)
SINGLE	% of Singles	Float	31.3
RETIRED	% of Retired	Float	19.1
UNEMP	% of Unemployed	Float	5.4
WRKCH5	% of working women with children under 5	Float	6.3
WRKCH17	% of working women with children 6 - 17	Float	11.2
NWRKCH5	% of non-working women with children under 5	Float	4.5
NWRKCH17	% of non-working women with children 6 - 17	Float	7.6
WRKCH	% of working women with children	Float	17.5
NWRKCH	% of non-working women with children	Float	12.1
WRKWCH	% of working women with children under 5	Float	7.4
WRKWNCH	% of working women with no children	Float	19.6
TELEPHN	% of households with telephones	Float	95.5
MORTGAGE	% of households with mortgages	Float	69.2

NWHITE	% of population that is non-white	Float	27.8
POVERTY	% of population with income under $15,000	Float	15.9
SHOPCONS	% of Constrained Shoppers (low budget)	Float	22.4
SHOPHURR	% of Hurried Shoppers (time-sensitive)	Float	10.7
SHOPAVID	% of Avid Shoppers (frequent shoppers)	Float	34.5
SHOPSTR	% of Shopping Strangers (rare shoppers)	Float	6.8
SHOPUNFT	% of Unfettered Shoppers (shop without budget constraints)	Float	4.2
SHOPBIRD	% of Shopper Birds (window shoppers)	Float	8.9
SHOPINDX	Ability to Shop (Car and Single Family House)	Float	1.8
SHPINDX	Ability to Shop (Car and Single Family House)	Float	1.8

3. UPC Files:

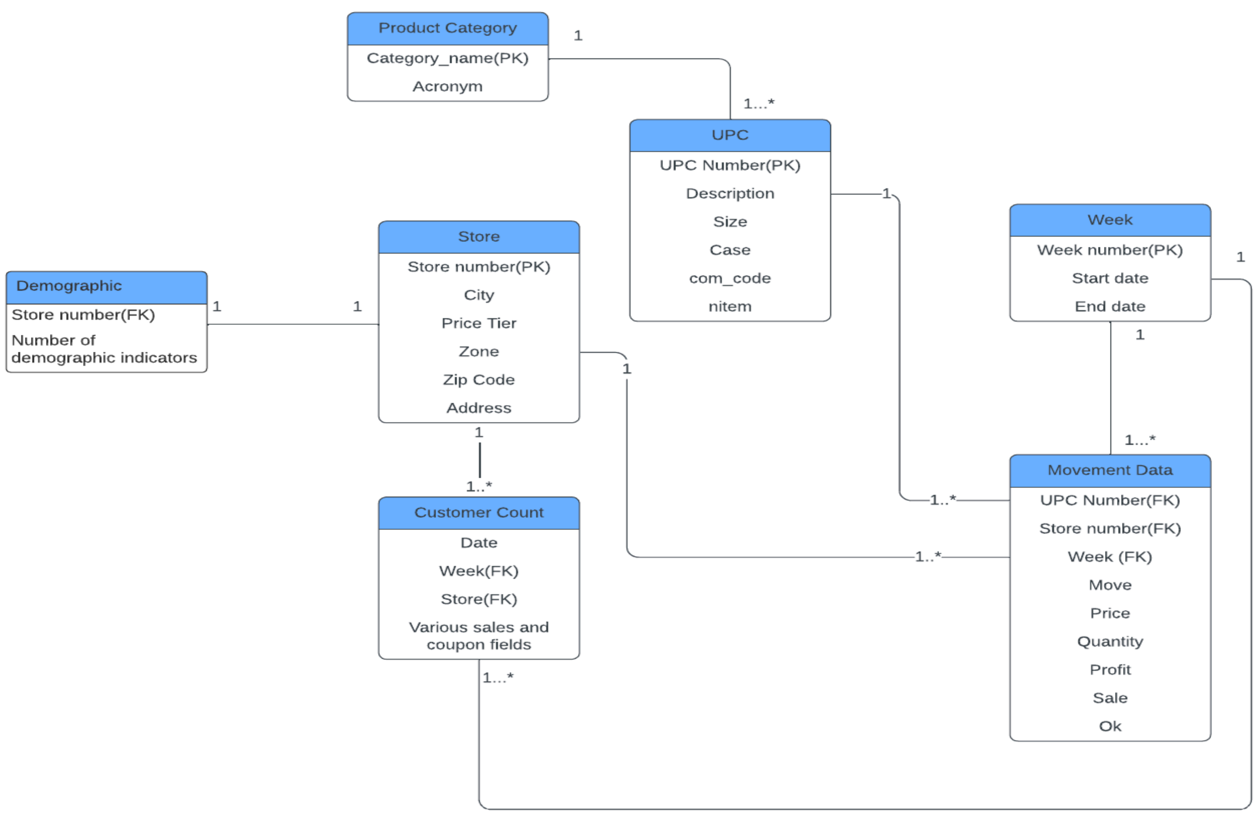
Column Name	Description	Type	Length
UPC	UPC number used to identify the product and the manufacturer.	Numeric	8
COM_CODE	Dominick's commodity code (category).	Numeric	8
NITEM	Item code used to track products.	Numeric	8
DESCRIP	Product Description (with special characters)	Character	20
SIZE	Product size (Quantity)	Character	6
CASE	Number of items in a case (for internal inventory use)	Numeric	8

4) Movement Files:

Column Name	Description	Type	Length
UPC	Unique Product Code (UPC) for each item	Numeric	8
STORE	Store identifier	Numeric	3
WEEK	Week number during which sales occurred	Numeric	3

MOVE	Number of units sold	Numeric	8
PRICE	Retail price for the product	Numeric	8
QTY	Number of items bundled together	Numeric	3
PROFIT	Gross profit margin (%)	Numeric	8
SALE	Sales promotion code ('B', 'C', 'S')	Character	8
OK	Data validity flag (1 = valid, 0 = invalid)	Numeric	3

ERD Diagram



Industry Research

Emerging trends in retail pricing practice: implications for research

The paper [2] discusses the comparison of retailers that have used a simple rule-based approach to defining prices rather than using data-driven decisions. This can also be applied to DFF's historical pricing model which might have led to suboptimal decisions and lost profits. Another issue that the paper brought to notice is the lack of consideration for the effect that cross-product sales have in the pricing model. Analyzing interrelated product categories, and calculating how the price of one product affects the demand for related products is important for increasing profitability.

It also shows that traditional pricing methods can make it difficult to adjust the prices to factors like inventory levels, competition, and demand. Organizations can benefit from adopting a dynamic pricing model. Profits can also be increased by implementing pricing strategies based on customer segments and locations. This customer segregation was highlighted as a missed opportunity by DFF. With a company like DFF that deals with perishable and non-perishable goods, inventory considerations for pricing decisions are crucial.

All these issues emphasize the complexity of retail pricing and the strategies that can be used for advanced analytics and price optimization.

An Examination of the extent, causes, and efforts to address retail out-of-stocks

The research paper [3] provides a comprehensive analysis of the retail out-of-stock (OOS) issue which is a critical issue in the retail industry that can significantly impact the profit and customer satisfaction. It is particularly important to the project as it highlights the importance of data-driven decisions.

The primary causes of the OSS are due to supply chain inefficiencies, demand forecasting errors, and poor store-level operations. These are interconnected involving multiple stakeholders such as manufacturers, distributors, retail stores, etc. A small disruption between any one of these stakeholders can result in OOS despite the demand for the product.

The paper emphasizes the price sensitivity and effectiveness of promotions across different product categories, analyzing this information will provide valuable insights on the effect that price changes and promotions will have on the sales of different products. As part of this project we have access to a vast dataset that can be leveraged to avoid potential OOS issues.

Likewise, inaccuracies in forecasting can lead to either overstocking or understocking. Forecasting errors are often the result of the inability to predict the changes in the market. In this case analyzing the historical data is essential to see past sales to accurately predict the future demands for the product.

Implementing a Data warehouse can help address these issues by centralizing the data from various sources to provide a comprehensive view of the retail operations. It enables trend analysis to accurately forecast future demands. It also enables better coordination between the stakeholders by providing a single source of truth for inventory and sales data. This can help reduce the inefficiencies and potential OOS.

The finding in the research paper further stresses the importance of implementing a data warehouse for DFF to address the issues proactively to enhance demand forecasting, optimize pricing, and improve customer satisfaction.

Communication and Promotion Decisions in Retailing: A Review and Directions for Future Research

The research paper [4] discusses the communication and promotion strategies in retailing, focusing on the roles of trade and consumer promotions, new media, and budget allocation. It highlights the shifting dynamics between manufacturers and retailers in the context of evolving consumer behavior and technology.

DFF faces significant challenges as it adapts to the rapidly evolving retail landscape, particularly in light of new media marketing. One of the most critical issues is adapting to new media marketing. As manufacturers increasingly allocate their promotional budgets towards digital platforms, DFF must reassess its marketing strategies, which may have relied heavily on traditional advertising. A failure to effectively embrace new media channels such as social media, email marketing, and targeted online promotions could result in missed opportunities for customer engagement and brand visibility. A robust online presence is essential for attracting a diverse customer base and fostering loyalty due to the competitive nature of the retail industry.

Coordination with manufacturers is a significant challenge for DFF as it implements new media promotions. Ensuring alignment with manufacturer promotional strategies is crucial to avoid inconsistent communication, which can confuse customers. For instance, if a manufacturer promotes a product on social media while DFF runs a different in-store campaign, it could disengage customers and lead to lost sales and brand trust.

Personalizing promotions offers both opportunities and challenges. With increasing consumer data availability, DFF can tailor promotions to individual preferences. However, failure to leverage this data effectively risks losing customers to competitors. Consumers now expect marketing that directly addresses their needs, so DFF's investment in building a data warehouse and using it for data analytics and customer insights is necessary.

Enhanced in-store promotions and product placement are necessary to convert foot traffic into sales. Understanding consumer price sensitivity is essential, especially since research indicates that online shoppers are more price-sensitive without non-price information. DFF must maintain competitive pricing and communicate product value to avoid losing customers to competitors. Additionally, effective budget allocation for promotions is crucial. DFF needs to balance its communication budget across channels to prevent resource waste and missed connections with consumers. Conducting thorough market research to grasp consumer behavior and preferences will be vital for success.

By embracing new media marketing, coordinating with manufacturers, personalizing promotions, optimizing the in-store experience, understanding price sensitivity, and effectively allocating budgets, DFF can enhance customer engagement and drive sales, positioning itself for future success.

A latent class segmentation analysis of e-shoppers

The research paper[5] explains that the primary focus is on exploring key retail management challenges, particularly in areas such as pricing strategies, shelf management, and inventory control. Dominick's Fine Foods (DFF) partnered with the University of Chicago Booth School to conduct randomized experiments from 1989 to 1994, aiming to understand consumer responses to various pricing and promotional strategies across their 100-store chain in the Chicago metropolitan area. These experiments were conducted in more than 25 product categories and generated data that is highly unique for its breadth and depth, including detailed information about retail margins, sales performance, and product movement at the store level. The study not only analyzed product pricing but also considered bundled promotions and coupon usage, which allowed DFF to test different sales techniques and their effects on consumer purchasing behavior.

These findings are crucial for informing the design and development of the data warehouse for DFF, as they highlight the importance of integrating comprehensive store-level sales data with demographic and promotional information. One of the key insights from the research is the impact of pricing on consumer behavior, especially in a competitive retail environment. By systematically collecting and analyzing this data, DFF can refine its pricing strategies, assess the success of various promotional campaigns, and make more informed decisions about future marketing tactics. Understanding which promotions work best in different regions or store branches allows for better-targeted sales efforts and helps DFF maximize profitability. For example, analyzing data about product bundling or coupon redemptions can provide insights into how DFF's consumers react to promotions, which can then be used to design more effective sales strategies.

Moreover, the paper emphasizes the role of shelf management and product placement, which directly impacts how products are sold. The research shows that effective shelf management strategies, such as product positioning and stocking, can significantly influence sales. This is a vital aspect of retail management that the data warehouse must capture, as the data can help optimize product assortment and shelf layout decisions based on store-specific sales trends and consumer preferences. The integration of sales data with demographic information, such as income levels and household sizes, will also provide insights into how regional variations affect product performance, allowing DFF to customize strategies for each location. In doing so, the data warehouse will empower DFF to manage its inventory more efficiently, ensuring that high-demand products are always available while minimizing the risk of overstocking low-demand items.

Ultimately, the insights from the research are integral to achieving the broader objective of the project: to create a data warehouse that enables data-driven decision-making across all stores in the DFF chain. By consolidating historical sales data with real-time store performance and demographic insights, the data warehouse will provide DFF with a powerful tool to analyze trends, predict consumer demand, and optimize retail operations. This will lead to better management of pricing, inventory, promotions, and store layouts, all of which are key to enhancing customer satisfaction and increasing overall sales performance. In summary, the research not only sheds light on consumer behavior and retail management challenges but also serves as a guide for the project's data warehouse, which will help DFF harness data for more efficient and profitable operations.

Proposed Business Questions

1. How does the profit margin of toothpaste vary by brand?

   Impact:

   Toothpaste is a staple product in our everyday life and has constant demand. Understanding the profit margin based on the brand can help DFF make informed decisions on pricing, promotion, and shelf space allocation. This will also provide them leverage to negotiate better terms with the manufacturers. It will also allow DFF to renegotiate with brands with lower profit margins or replace them with better-performing brands.

   Support:

   For doing preliminary analysis, we have used wtpa.csv which contains movement data for toothpaste. Then I created a pivot table with upc as rows and calculated the average profit margin and total sales = [MOVE*PRICE]/QTY. I then sorted it in descending order of profit margin to show the UPCs with the highest profit margin. In the later stages, we can join the data from the UTC file for toothpaste to then do further analysis.

Row Labels	Sum of TOTAL SALES	Average of PROFIT
1111341101	₹3,26,582.49	3882.98%
1111319120	₹1,08,967.34	2923.70%
1851527243	₹3,837.18	2833.27%
1111323101	₹2,25,230.06	2758.12%
2260064631	₹43.25	2727.93%
1111381660	₹11,388.63	2501.19%
1111374102	₹4,01,777.65	2335.77%
1111320202	₹1,18,305.18	2335.21%
1851527239	₹1,698.37	2319.96%
1111329780	₹73,106.72	2285.85%
1111381680	₹6,343.64	2234.82%
1111381650	₹1,438.23	2104.13%
1111315120	₹9,365.67	2088.49%
1111374801	₹1,19,397.43	2085.45%
1851527235	₹4,825.92	2067.52%
1111363761	₹2,09,512.87	2037.47%
3320018180	₹6,145.95	2022.24%
1111363961	₹2,27,068.84	1992.14%
1111363861	₹1,63,336.63	1984.50%
2260062031	₹1,486.71	1830.03%
1111363701	₹1,56,990.97	1826.63%
1111325381	₹47,113.64	1825.87%
1111374242	₹39,191.01	1789.12%
1111379161	₹1,91,531.29	1785.78%
3320018190	₹2,396.47	1782.39%
1111349510	₹25,957.74	1699.79%
1111363801	₹1,20,946.36	1696.76%
1111363901	₹1,23,975.72	1691.38%
1111379152	₹24,200.36	1640.34%

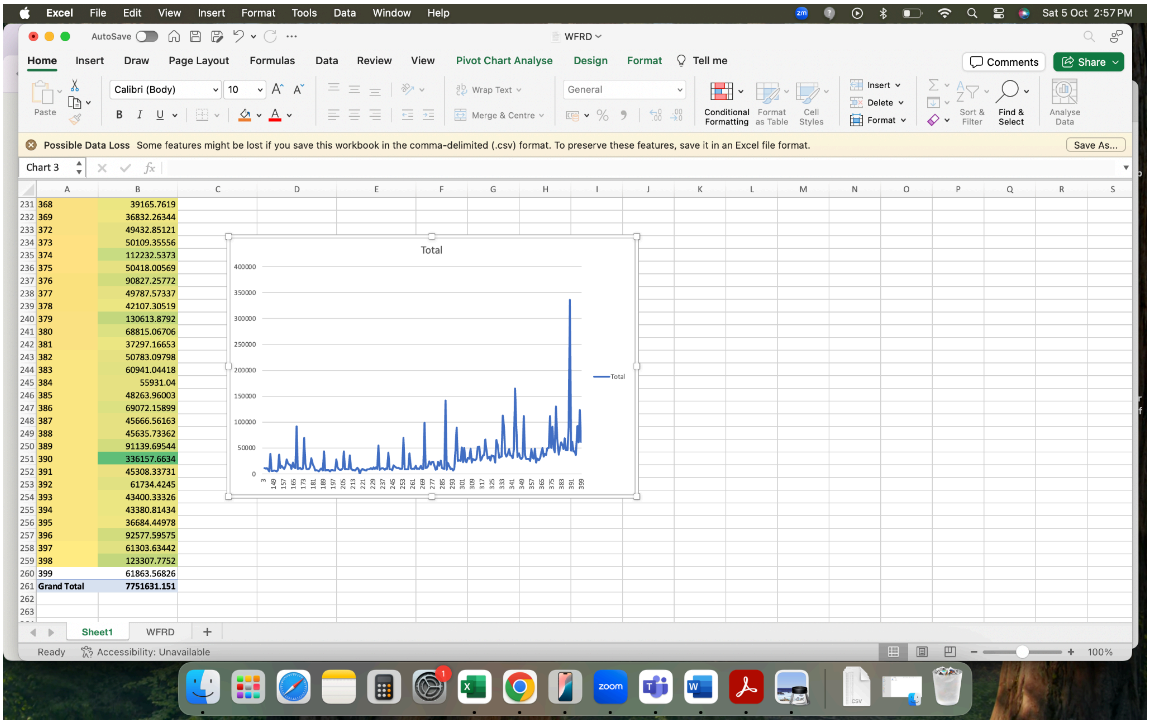
2. What is the trend in frozen dinner sales? How does it vary across different stores?

   Impact

   The frozen dinner market is growing as people are getting more and more busy and looking for convenient and hassle-free ways to get in their meals. Understanding the sales trend for this product category helps DFF to capitalize on the growing need for frozen dinners. This provides valuable insights that will help DFF manage its inventory, and plan product placement, promotions, and overall business strategy. It also helps to identify the regional preferences tailoring the product stocking specific to stores to increase the overall sales.

   Support:

   For this, I used WFRD.csv which has the data for frozen food. For preliminary analysis, I have created a pivot table with Week as rows and calculated the total sales = (MOVE * PRICE)/QTY. Further, I have added stores as filter to filter data as per stores for in-depth analysis. Finally we have highlighted the table using a heatmap to highlight the low and high sales and also displayed a pivot table to see the trend.



3. Which product is purchased most by college graduate students?

   Impact:

   Understanding what a key demographic is purchasing and their preference is invaluable. College graduates are an important set of customers even though they might not generate the most revenue, the insights from this analysis can help with targeted marketing and product selection in the stores near the universities. It can also be leveraged to capture a potential loyal customer base.

   Support:

   We need to join the movement data using the UPC to answer this question and get the product codes. Then we would join the store demographics to get the percentage of graduate students and then filter stores having above-average college students. Finally, we would aggregate the sales for these stores. Below is a preliminary analysis of the store ranking based on total sales.

4. What is the average price and sales volume of women's shampoo products in stores located in areas with above-average percentages of working women?

Row Labels	Sum of TOTAL SALES	Average of PROFIT
102	₹74,932.02	1629.52%
128	₹72,978.30	1795.47%
122	₹59,618.61	1649.19%
12	₹58,820.51	1686.35%
100	₹57,160.72	1640.56%
130	₹55,334.77	1506.06%
98	₹54,150.80	1582.68%
75	₹54,102.32	1597.44%
126	₹53,643.65	1590.31%
121	₹53,244.91	1652.95%
132	₹50,902.41	1551.68%
114	₹49,604.09	1523.60%
86	₹48,984.22	1607.93%
112	₹48,447.61	1643.20%
109	₹47,085.92	1586.51%
74	₹46,877.80	1515.14%
8	₹45,649.49	1620.81%
124	₹45,011.11	1592.22%
32	₹44,453.21	1531.34%
133	₹44,401.17	1437.33%
71	₹44,252.74	1549.73%
101	₹43,702.75	1525.23%
123	₹43,556.36	1557.78%
131	₹42,861.49	1553.04%
107	₹42,177.07	1589.55%
18	₹42,052.70	1475.25%
73	₹41,534.20	1492.59%
80	₹40,771.60	1594.16%
84	₹40.593.97	1556.07%

Impact:

This analysis combines the demographic data with the product performance providing DFF with valuable insights to place targeted marketing. Working women are a significant consumer group, and understanding their purchase patterns can help DFF tailor their offering and adjust its pricing strategy accordingly. It also creates an opportunity to provide value-oriented options in stores with more working women traffic which can potentially increase sales.

Support:

To answer this question, we need to join the movement, store, and store demographics data. Next, we need to calculate the average percentage of women across all the stores. Then find the stores with above average percentage of working women. Finally, calculate the average sales and sales volume.

5. What is the percentage contribution of cookies towards the total sales during the holiday season (i.e., Christmas and New Year) of the year 1994?

Impact:

The seasonal sales patterns are important to focus on as it's the time of the year when people buy products for hosting events and gifting. We have chosen cookies as its

Row Labels	Sum of Total sales
126	35262.41838
122	32635.56071
98	31665.41467
128	31574.0808
121	29785.93784
102	28105.60027
101	27476.1678
131	26799.7167
112	26112.51845
139	24802.71345
137	24686.91468
100	24521.06985
133	24435.27913
115	24410.4053
32	23831.94081
129	22978.93117
86	22450.12341
12	22147.58947
132	21026.24038
107	20559.47582
8	19655.63939
109	18687.77604
136	18291.89259
80	17482.26858
123	17117.62023
70	17044.50913
71	16989.53923
18	16947.65127

purchased to be consumed as well as gifts. Understanding a particular product's sales pattern can provide valuable insights into the overall holiday sales. DFF can leverage this information for planning holiday promotions, store layout planning, and inventory management during the holiday season.

Support:

To solve this question we need to join the movement and upc data for the weeks of holiday as per the manual. Calculate the total sales of cookies. Then calculate the total sums. Finally, calculate the percentage contribution.

Row Labels	Sum of total sales
276	3659.357126
Grand Total	**3659.357126**

6. Which store has the highest store traffic for the last quarter of 1994?

Impact

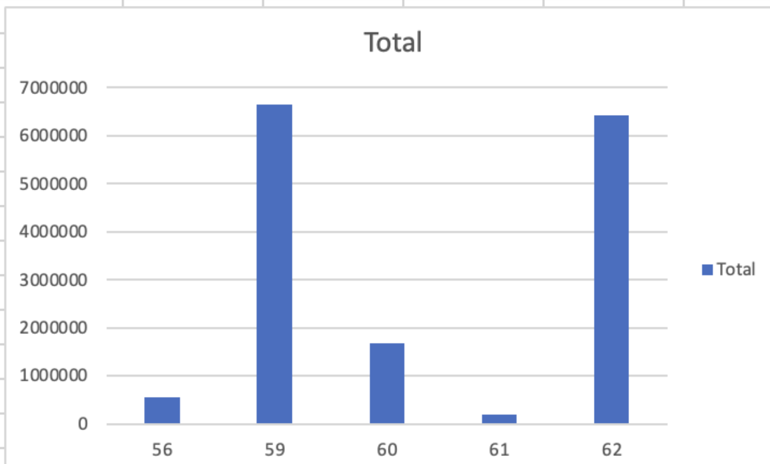
Store traffic is a key indicator of the overall business performance and sales. By identifying the highest-traffic store, DFF can analyze what contributed to its success and replicate the same across other locations. The focus on the last quarter of the year is important as it's the holiday season, and the insights from this can be used to improve the performance across other stores as well.

Support

For this, I used the CCOUNT.csv and filtered out the data for weeks 264-276 [The week code for the last quarter of 1994]. I then created pivot tables with stores as a row number and sum of CustCount to calculate the store traffic. I finally displayed the bar chart showing the result.

As per this, store number 59 which is Crystal Lake [As per the Dominicks Manual] store had the highest store traffic in the last quarter.

Row Labels	Sum of CUSTCOUN
56	560708
59	6636839.95
60	1681199.5
61	198928
62	6409721.82
Grand Total	**15487397.27**



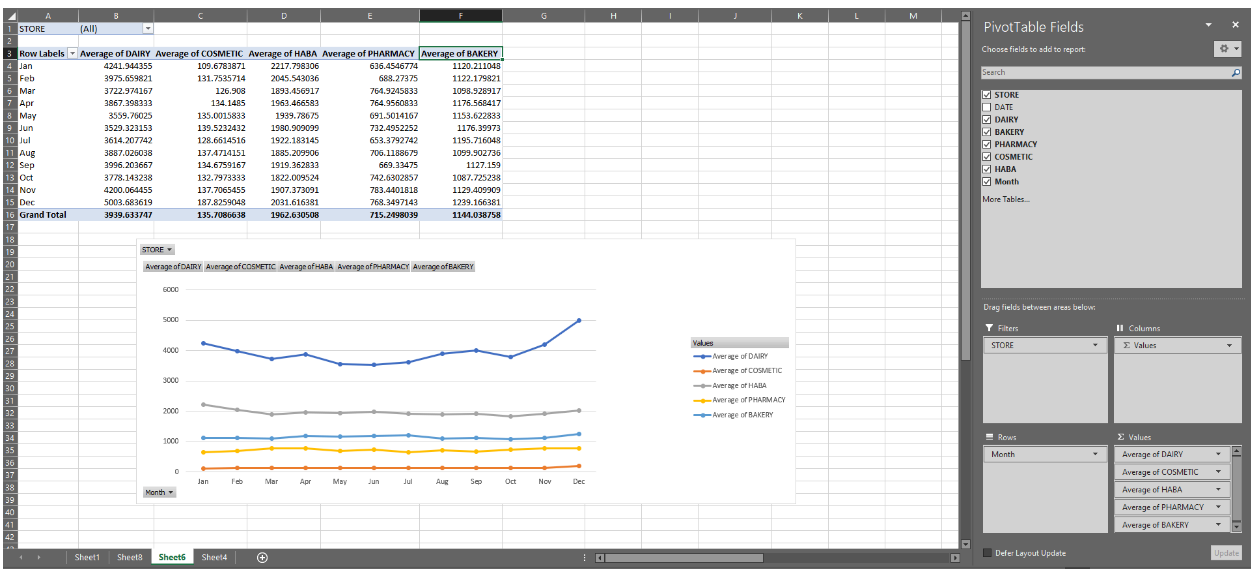
7. What are the monthly average sales amounts for the BAKERY, DAIRY, PHARMACY, COSMETIC, and HABA departments across Dominick's Fine Food stores located in Naperville and Schaumburg during the year 1994, and how do these averages compare among the different departments within these locations?

Impact:

Analyzing the monthly average sales amounts for the BAKERY, DIARY, PHARMACY, COSMETIC, and HABA departments across Dominick's Fine Food stores in Naperville and Schaumburg during 1994 is beneficial for DFF as it provides insights into consumer preferences and sales trends. Understanding which departments perform better can guide inventory management, marketing strategies, and resource allocation. This data-driven approach enables DFF to optimize promotions and improve customer satisfaction, ultimately enhancing profitability.

Support:

For this, I have used CCOUNT.csv file date. Naperville and Schaumburg store numbers are 54, 115 and 48, 117 respectively. As we are looking the data related to these stores for the year 1994, we initially fetched Store, Date, Diary, Bakery, Pharmacy, Cosmetic, and Hba categories data from CCOUNT.csv to different files. In the new file, we filtered records accordingly. Our new files consist of 4 stores of data related to those categories in the year 1994. We have added a new column called month which stores the month of each observation respectively. After that, we created a pivot table and a line chart as mentioned below.



STORE (All)

Row Labels	Average of DAIRY	Average of COSMETIC	Average of HABA	Average of PHARMACY	Average of BAKERY
Jan	4241.944355	109.6783871	2217.798306	636.4546774	1120.211048
Feb	3975.659821	131.7535714	2045.543036	688.27375	1122.179821
Mar	3722.974167	126.908	1893.456917	764.9245833	1098.928917
Apr	3867.398333	134.1485	1963.466583	764.9560833	1176.568417
May	3559.76025	135.0015833	1939.78675	691.5014167	1153.622833
Jun	3529.323153	139.5232432	1980.909099	732.4952252	1176.39973
Jul	3614.207742	128.6614516	1922.183145	653.3792742	1195.716048
Aug	3887.026038	137.4714151	1885.209906	706.1188679	1099.902736
Sep	3996.203667	134.6759167	1919.362833	669.33475	1127.159
Oct	3778.143238	132.7973333	1822.009524	742.6302857	1087.725238
Nov	4200.064455	137.7065455	1907.373091	783.4401818	1129.409909
Dec	5003.683619	187.8259048	2031.616381	768.3497143	1239.166381
Grand Total	**3939.633747**	**135.7086638**	**1962.630508**	**715.2498039**	**1144.038758**

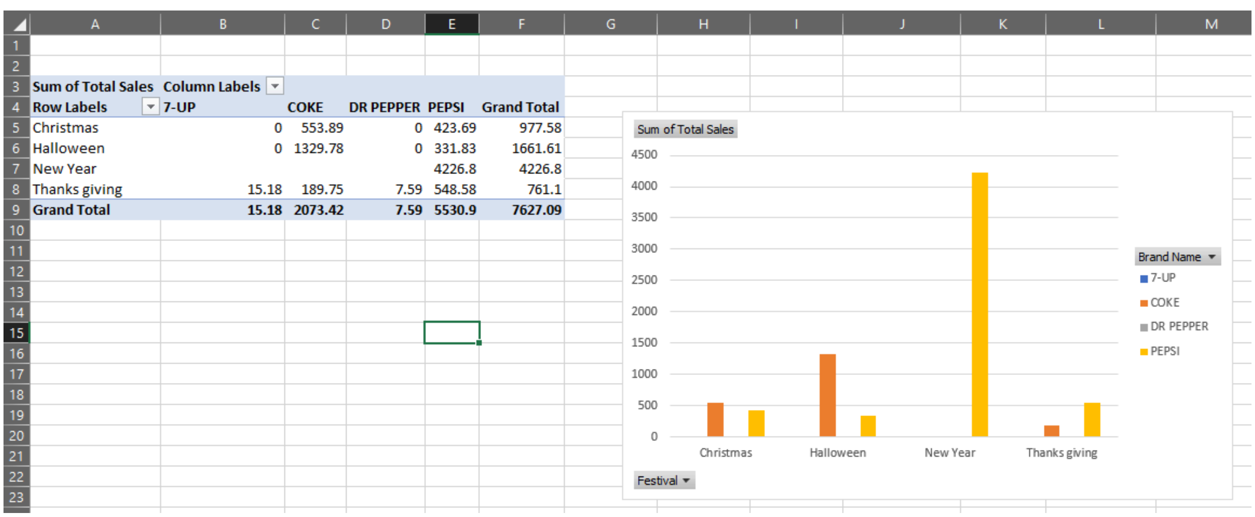
The line chart showcases different category's average sales amount monthly and how it varies month by month and how it is performing compared to the monthly average sales of other categories. We can drill down this further by using the store filter to know more about monthly average sales at each store level. From the chart we can see that, the Diary category is performing better than any other category as its average monthly sales amount is higher than others. We can see trends in the DIARY category month by month. In the month of May, the diary sales went down compared to all other months' performance of the diary category and its monthly sales were highest in Dec 1994.

8. Which soft drink brand (PEPSI, COKE, 7 UP, DR PEPPER) recorded the highest sales performance at the Orland Park store during the 1992 holiday season, specifically across Halloween, Thanksgiving, Christmas, and New Year? [6]

Impact:

Understanding which soft drink brand PEPSI, COKE, 7 UP, or DR PEPPER performed best during key holiday periods in 1992 at the Orland Park store provides valuable insights for Dominick's Fine Foods (DFF). This information helps in optimizing product placements, tailoring promotions, and aligning stock levels with consumer demand during high-traffic seasons. It also enables DFF to better negotiate with suppliers and brands based on performance metrics, ultimately improving profitability and customer satisfaction.

Support:



Sum of Total Sales	Column Labels				
Row Labels	7-UP	COKE	DR PEPPER	PEPSI	Grand Total
Christmas	0	553.89	0	423.69	977.58
Halloween	0	1329.78	0	331.83	1661.61
New Year				4226.8	4226.8
Thanks giving	15.18	189.75	7.59	548.58	761.1
Grand Total	15.18	2073.42	7.59	5530.9	7627.09

For this, I have used UPCSDR.csv and wsdr.csv files. First I have found out the PEPSI, COKE, 7-UP, and DR PEPPER product's UPC codes from the UPCSDR.csv file. After finding out the codes, I filtered the records in a wsdr.csv file. I selected store number 84 which is the store number of the Orland Park store. I fetched the sales of those brands exclusively during the holiday season. In 1992 holiday season weeks were - 120, 164, 168, and 172. Using the filters I have fetched sales of brands during the holiday season and stored the results in a separate sheet. From that data, I created the pivot table and the bar chart.

From the bar chart, we can understand that on New Year and Thanksgiving week - the PEPSI brand has more sales than all others, on Halloween and Christmas week, COKE has the highest sales recorded. In this way, we can analyze for any area or city or across all stores if needed.

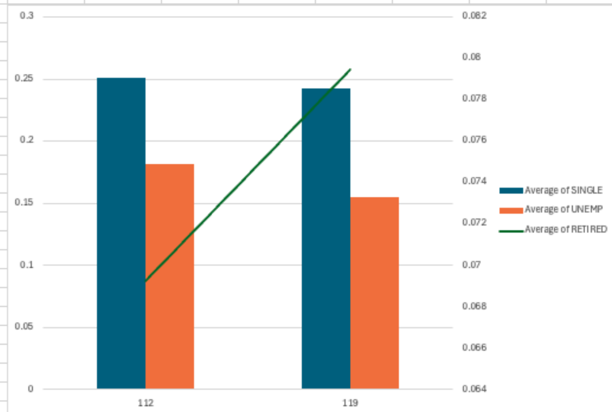
9. What were the highest sales contributions from single and retired individuals in the Buffalo Grove stores for BUDWEISER BEER N.R.B during the Thanksgiving week of 1993?

This analysis helps **Dominick's Fine Foods** by revealing how **single** and **retired** shoppers contribute to sales during Thanksgiving week. Understanding demographic-driven sales for high-demand products like **BUDWEISER BEER N.R.B** enables DFF to optimize promotions and inventory for these customer segments. It allows for more targeted marketing and better stocking strategies, helping the retailer maximize sales and meet customer demand during key holiday periods.

Support:

To analyze the sales data, I used the demographic CSV file to filter the relevant week of sales, focusing on specific customer segments, namely single, unemployed, and retired individuals. I separated the necessary columns and rows to isolate the required data. Next, I worked with the UPC and Movement CSV files. From the UPC file, I extracted the product UPC codes and sizes. Then, using the Movement CSV file, I filtered the data for stores located in Buffalo Grove, identifying the sales transactions for the target products during the specified week. After filtering the data, I created a pivot table to summarize the findings. Finally, I generated a pivot chart to visually depict the differences in sales during that particular week.

Row Labels	Average of SINGLE	Average of UNEMP	Average of RETIRED
112	0.250960531	0.181411077	0.06922866
119	0.242272811	0.15471004	0.079484945
Grand Total	**0.246616671**	**0.168060558**	**0.074356803**

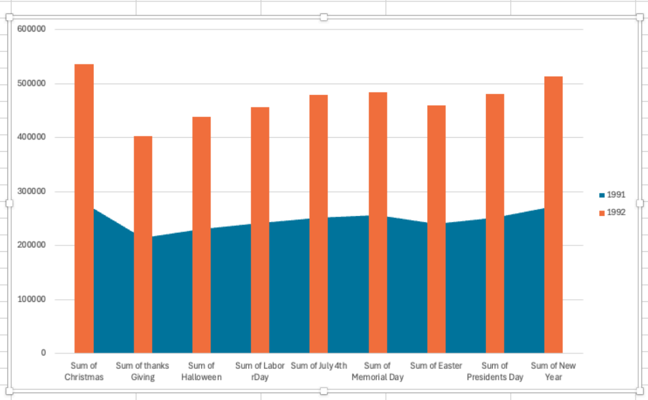


10. Which holiday week in 1991 and 1992 saw the highest grocery sales in the low-tier Buffalo Grove stores? [3]

    This analysis is valuable for **Dominick's Fine Foods (DFF)** as it helps identify peak grocery sales periods in low-tier stores during key holidays. By pinpointing which holiday weeks in 1991 and 1992 had the highest sales, DFF can optimize future inventory management, staffing, and promotions for low-tier locations. This allows DFF to better align supply with demand, minimizing stockouts while maximizing revenue during high-traffic holiday periods.

    Support:

	Column Labels		
Values	**1991**	**1992**	**Grand Total**
Sum of Christmas	279329.44	536413.4	815742.84
Sum of thanks Giving	213829.46	403076.49	616905.95
Sum of Halloween	231238.68	438970.39	670209.07
Sum of Labor rDay	241408.64	457395.02	698803.66
Sum of July 4th	251013.43	478740.61	729754.04
Sum of Memorial Day	255888.35	484881.8	740770.15
Sum of Easter	240882.34	459713.73	700596.07
Sum of Presidents Day	252316.04	481162.36	733478.4
Sum of New Year	273034.22	514303.27	787337.49



To perform this analysis, I utilized the account CSV file alongside the DFF manual PDF. I began by collecting the grocery purchase data for the entire year of 1991, followed by repeating the same process for 1992. After gathering the data, I filtered the weekly grocery purchases for both years individually. Once the data was cleaned, I created a pivot table to summarize the findings. Using this pivot table, I then generated a pivot chart to visualize the results. The chart clearly shows that sales during holiday weeks in 1992 were higher compared to the same period in 1991, highlighting a notable increase in holiday-related sales.

Prioritization of Business Questions:

1. Which store has the highest store traffic for the last quarter of 1994?
2. What is the trend in frozen dinner sales? How does it vary across different stores?
3. Which holiday week in 1991 and 1992 saw the highest grocery sales in the low-tier Buffalo Grove stores?
4. How does the profit margin of toothpaste vary by brand?
5. What are the monthly average sales amounts for the BAKERY, DAIRY, PHARMACY, COSMETIC, and HABA departments across Dominick's Fine Food stores located in Naperville and Schaumburg during the year 1994, and how do these averages compare among the different departments within these locations?
6. What is the percentage contribution of cookies towards the total sales during the holiday season (i.e., Christmas and New Year) of the year 1994?
7. What is the average price and sales volume of women's shampoo products in stores located in areas with above-average percentages of working women?
8. Which soft drink brand (PEPSI, COKE, 7 UP, DR PEPPER) recorded the highest sales performance at the Orland Park store during the 1992 holiday season, specifically across Halloween, Thanksgiving, Christmas, and New Year?
9. Which product is purchased most by college graduate students?
10. What were the highest sales contributions from single and retired individuals in the Buffalo Grove stores for CORONA EXTRA LIGHT N during the Thanksgiving week of 1993?

We decided on this prioritization because we believe focusing on business operations and strategic decision making can help DFF to analyze customer traffic and performance to find the most successful locations. They can try to replicate the model at all the other stores. Next, we focused on the sales of essential categories like toothpaste, and frozen dinner sales and their sale trends. This information can be used to manage store inventory and decide pricing strategies. In the last section, we focused on demographic-related questions to do targeted marketing to increase sales for that demographic during the festive season.

References

1. *https://www.company-histories.com/Dominicks-Finer-Foods-Inc-Company-History.html*
2. *Michael Levya, Dhruv Grewala, Praveen K. Kopalleb, James D. Hessc," Emerging trends in retail pricing practice: implications for research", Journal of Retailing, 80(3), xiii-xxi,2004*
3. *Daniel Corsten, Thomas Gruen, "Desperately seeking shelf availability: An Examination of the extent, causes, and efforts to address retail out-of-stocks", 2003*
4. *Kusum L. Ailawadi, J.P. Beauchamp, Naveen Donthu, Dinesh K. Gauri, Venkatesh Shankar "Communication and Promotion Decisions in Retailing: A Review and Directions for Future Research", Journal of Retailing 85 (1) 42–55, 2009*
5. *Amit Bhatnagar, Sanjoy Ghose, "A latent class segmentation analysis of e-shoppers", Journal of Business Research 57 758 – 767, 2004*
6. *https://www.company-histories.com/Dominicks-Finer-Foods-Inc-Company-History.html*
7. *https://www.chicagobooth.edu/research/kilts/research-data/dominicks*
8. *Dominick's Data Manual and Codebook - "Dominicks-Manual-and-Codebook_KiltsCenter2013"*