**Predictive Analysis**

**of**

**Retail Sales and Promotions**

**By: Pranjali Rokade**

**Institution: Northeastern University**

**Date: 05.25.24**

Contents

[Introduction 3](#_Toc167733794)

[Background 3](#_Toc167733795)

[Objective 3](#_Toc167733796)

[Methodology 4](#_Toc167733797)

[Tools 4](#_Toc167733798)

[Data Source 4](#_Toc167733799)

[Exploratory Data Analysis 5](#_Toc167733800)

[Data Cleaning 5](#_Toc167733801)

[Descriptive Statistics 5](#_Toc167733802)

[Data Transformation and Feature Engineering 5](#_Toc167733803)

[Visualizations and Insights 7](#_Toc167733804)

[1. Histograms 7](#_Toc167733805)

[2. Units Uplift with Time 10](#_Toc167733806)

[3. ROI and Company 11](#_Toc167733807)

[4. Weighted ROI 12](#_Toc167733808)

[5. Dashboard 13](#_Toc167733809)

[Predictive Analysis 17](#_Toc167733810)

[Purpose: 17](#_Toc167733811)

[Inference: 17](#_Toc167733812)

[Conclusion 18](#_Toc167733813)

## Introduction

### Background

Retailers run promotions frequently to motivate sales. Sales volume typically expects an uplift at the cost of the reduced price. Retailers and CPG companies need to measure how effective retail promotions are in terms of impact on margin and sales. Understanding the dynamics of promotional strategies is crucial for making informed decisions that maximize revenue and profitability. By analyzing sales data and promotional activities, companies can gain insights into which promotions are driving the most significant increases in sales volume and which are merely eroding profit margins without yielding substantial returns.

### Objective

- Understand the impact of promotions on overall sales performance.

- Identify key factors that contribute to successful promotions and to forecast future sales trends based on historical promotional activities

- Use combination of Tableau and Python for data visualizations and exploratory data analysis respectively

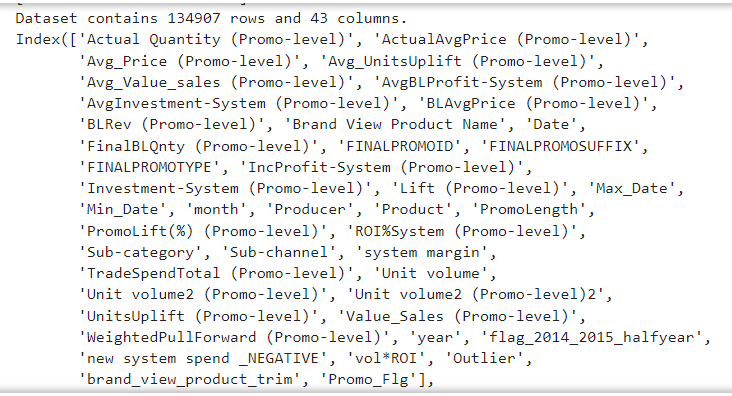
## Methodology

### Tools

Used combination of Tableau and Python for data visualizations and exploratory data analysis respectively.

### Data Source

The dataset used in this analysis was obtained from our Professor Wang, who provided us with a comprehensive set of retail sales and promotional data. This dataset is crucial for understanding the effectiveness of various promotional strategies on sales performance. It consists of 134,907 rows and 43 columns, capturing a wide array of information at both the system and promo levels.



## Exploratory Data Analysis

I have uploaded my Python notebook on GitHub, which details how I conducted exploratory data analysis. You can find all my projects there, including a folder named 'Retail\_Sales\_and\_Promotions'. Inside this folder, you'll find my Python notebook titled 'EDA Retail'.

Here is the link: [Pranju97/Data-Analyst-Pranjali (github.com)](https://github.com/Pranju97/Data-Analyst-Pranjali)

### Data Cleaning

I performed data cleaning code using Python to check any missing values.

In the above link, we see that there are no missing values in the dataset which says that the data is already cleaned.

### Descriptive Statistics

Descriptive statistics provide an overview of the main characteristics of the dataset by summarizing the central tendency, dispersion, and shape of the data's distribution. My python notebook will show you the descriptive statistics for the numerical features in our dataset, including measures such as mean, standard deviation, minimum, and maximum values.

### Data Transformation and Feature Engineering

Below transformations are done using Python:

1. Stripping Spaces from Column Names:

Transformation: Leading and trailing spaces are removed from column names.

Impact: Ensures consistency and prevents errors in column referencing.

2. Converting Columns to Numeric:

Transformation: Columns such as 'ActualAvgPrice (Promo-level)', 'Value\_Sales (Promo-level)', 'Avg\_UnitsUplift (Promo-level)', and 'AvgInvestment-System (Promo-level)' are converted to numeric types using pd.to\_numeric with errors='coerce'.

Impact: Ensures that non-numeric entries are converted to NaN, allowing for proper numerical analysis and calculations.

3. Handling Missing Values:

Transformation: Rows containing NaN values are dropped.

Impact: Removes incomplete data entries, ensuring that the dataset used for modeling is clean and complete.

4. Categorical Conversion:

Transformation: The 'FINALPROMOTYPE' column is converted to a categorical data type.

Impact: Facilitates the creation of dummy variables, enabling the model to handle categorical data.

5. Date Conversion and Feature Extraction:

Transformation: The 'Date' column is converted to a datetime format. Month and year are extracted into new columns 'Month' and 'Year'.

New Features: 'Month' and 'Year' columns.

Impact: Enables temporal analysis and allows the model to consider seasonal patterns and yearly trends.

6. Creating Dummy Variables:

Transformation: Dummy variables are created for the 'FINALPROMOTYPE' column using pd.get\_dummies.

New Features: Columns like 'FINALPROMOTYPE\_Type 1 Promo', 'FINALPROMOTYPE\_Type 2 Promo', etc.

Impact: Converts categorical promotion types into a binary format that the model can understand and use effectively.

7. Feature Selection:

Transformation: Selected features for the model include numerical columns, extracted date features, and dummy variables.

Features: 'ActualAvgPrice (Promo-level)', 'Avg\_UnitsUplift (Promo-level)', 'AvgInvestment-System (Promo-level)', 'Month', 'Year', and dummy variables for 'FINALPROMOTYPE'.

Impact: Ensures that relevant features are included for model training, which can significantly influence the model's predictive power.

9. Feature Importance Analysis:

Transformation: Using a trained Random Forest model to determine the importance of each feature.

Impact: Identifies the most significant features influencing the target variable, which can be used for feature selection and improving model performance.

## Visualizations and Insights

Before we start with our visualizations, there are many outliers that could skew our data analysis. Those outliers specifically filtered in the dataset using column “Outliers” with “N” in Tableau. Also, just for the sake of better and easy analysis, I have only considered promotions that are currently going by filtering column “Promo\_flg” with value 1. You can access the tableau file in my GitHub link posted in ‘Exploratory Data Analysis’ section.

### 1. Histograms

Purpose: The histograms for each column provide insights into the distribution of the data

Actual Quantity (Promo-level):

The distribution is right-skewed, indicating that most promotion-level quantities are on the lower end, with a few instances of very high quantities.

A white graph with black lines

Description automatically generated with medium confidence

Figure 1

Avg\_UnitsUplift (Promo-level):

The histogram shows a wide range of values, indicating high variability. There are both negative and positive units uplift, with a notable number of promotions having small uplifts or reductions.

A graph with a bar

Description automatically generated with medium confidence

Figure 2

Value\_Sales (Promo-level):

The distribution is right-skewed, like the quantity and price distributions. Most sales values during promotions are relatively low, with some promotions generating very high sales values.

A graph with a blue line

Description automatically generated with medium confidence

Figure 3

**Key Takeaways**

Skewed Distributions:

The data for quantities, prices, and sales during promotions are right-skewed, indicating that most promotions result in lower values with a few high-value outliers. This suggests that while many promotions may have moderate success, some are exceptionally effective.

Variability in Promotions:

The units uplift during promotions shows high variability, with both positive and negative impacts. This indicates that promotions can have different outcomes, sometimes even reducing the number of units sold.

Presence of Outliers:

Outliers are present in all the variables, highlighting that some promotions significantly deviate from the norm. These outliers could represent highly successful or unsuccessful promotions and warrant further investigation.

### 2. Units Uplift with Time

Purpose:

Firstly, we will examine how promotions affect the increase in quantities sold. This analysis is essential because the success of a promotion is measured not just by the average price, but also by the number of units sold.

Inference:

As shown in the graph (figure 4), different promotion types led to varying increases in quantities sold. Promotion type 2 resulted in a significantly higher increase compared to promotion type 1. The units uplift for promotion type 2 showed a notable rise from January to March 2014, a slight decline in April, and then a resurgence through June 2014. This indicates that promotion type 2 was highly effective, likely due to large sales volumes, appealing offers, better-targeted marketing, or higher consumer demand for the promoted products. In contrast, promotion type 1 saw minimal quantities sold, indicating a much lower impact.A graph with a line

Description automatically generated

Figure 4

### 3. ROI and Company

Figure 5 shows the return on investment (ROI) of five companies over time. ROI is a metric that compares the cost of an investment with the return on that investment. In this case, the investment is the cost of the promotion, and the return is the increase in sales.

The graph shows that Company III had a promotion that ran from 2012 to 2015. The ROI for this promotion was positive throughout the entire period, which means that the promotion was successful in generating a return on investment. The other four companies had promotions that ran from 2014 to 2015. The ROI for these promotions was also positive, but it was lower than the ROI for Company III's promotion.

There are a few reasons why this graph is important for retail sales and promotions. First, it shows that promotions can be an effective way to increase sales. Second, it shows that the timing of a promotion can be important. Company III's promotion was more successful than the other companies' promotions, possibly because it ran for a longer period.

By analyzing data on promotions, retailers can learn what types of promotions are most effective, when to run promotions, and how to target promotions to specific customers.

**A graph with lines and numbers

Description automatically generated with medium confidence**

Figure 5

### 4. Weighted ROI

Purpose:

* Illustrate the weighted ROI calculated at an aggregate level with sales volume as a weighting factor.
* Compare the weighted ROI of promotion types 1 and 2 to determine which generates a higher return on investment.
* Identify high-performing categories and sub-channels to inform future promotional strategies.

Inference:

* The ROI is influenced more by the sales volume of a particular product or category, with products having higher sales volume impacting the overall ROI calculation more significantly.
* Figure 6 shows that promotion type 2 has a higher weighted ROI, likely due to its larger sales volume, helping retailers decide which promotion type to prioritize in the future.
* The breakdown of ROI by category (1, 2, and 3) and sub-channel (A, B, C, and D) allows retailers to see which categories and sub-channels benefit most from promotions.
  + For example, Company V under category 3 and sub-channel D has the highest weighted ROI, indicating it might be a good candidate for further promotions or strategic pricing.
  + Conversely, Company I invested heavily in type 2 promotion under category 3 but failed to gain returns, suggesting this promotion type might not have been effective for Company I.
* The graph uses color to represent the average investment made in promotions, providing additional insights into the effectiveness of promotional investments.

A screenshot of a graph

Description automatically generated

Figure 6

This graph is a valuable tool for retailers because it helps them assess the effectiveness of different promotion types and identify which categories and sub-channels benefit most from them. This information can be used to optimize promotional strategies and maximize return on investment.

### 5. Dashboard

The dashboards collectively tell a compelling story of how Type 1 promotions have impacted incremental profit and ROI across different time periods: 2014 January-June to 2015 January-June, and a broader period from April 2012 to July 2015. By examining these time frames, we can understand the effectiveness of Type 1 promotions over time and across different market conditions

Case I: 2014 January-June

The first half of 2014 was a highly successful period for Type 1 promotions, driving significant incremental profits and high ROI. This suggests that the promotional strategies implemented during this time were well-targeted and resonated with consumers.

In the Promotional ROI Trends chart, Company IV in sub-channel D achieved a higher ROI of 352% with an investment of $10,570. The investment levels are visually represented by the thickness of the lines, where thicker lines indicate higher investments, and green signifies higher ROI. The chart shows that while some companies achieved significant returns even with smaller investments, others experienced higher losses due to type 1 promotions. Additionally, the chart reveals that companies focused their investments on only a few sub-channels; for instance, Company II never invested in type 1 promotion.

A screenshot of a computer screen

Description automatically generated

Figure 7

In the first half of 2014, Type 2 promotions showed moderate profits initially, spiking in March but declining significantly in May and June, resulting in negative profits. Company IV, in sub-channel D, invested $27,961 in Type 2, more than in Type 1, but incurred losses, indicating the promotion was unsuitable or poorly executed for this sub-channel. Conversely, other companies investing solely in Type 2 saw gains. We should check if Company IV continued investing in Type 1 promotions.

A screenshot of a computer screen

Description automatically generated

Figure 8

Case II: 2015 Jan – June

In the first half of 2015, incremental profits showed significant losses in January and February, with some recovery later. The baseline profit often exceeded actual profit, indicating Type 1 promotions were less effective or faced more competition. Company IV's reinvestment in Type 1 resulted in a heavy loss of -171%. Other companies also invested in Type 1, leading to increased competition and further losses.

A screenshot of a computer screen

Description automatically generated

Figure 9

Not just with promotion type 1, type 2 also faced same situation where their profits dropped till February. However, the companies started investing in sub-channel C and resulted in higher ROI even after keeping their investment almost constant. This overall result showed in incremental profit in March 2015. But again, the results dipped drastically after March.

A screenshot of a graph

Description automatically generated

Figure 10

## Predictive Analysis

I have done predictive analysis for the next 8 months starting after June 2015 which you will find out in my GitHub link in the above pages.

### Purpose

The purpose of conducting this predictive analysis is to forecast the sales performance based on different promotion types. By leveraging historical sales data and key features such as average price, units uplift, and investment, the analysis aims to provide insights into future sales trends. This allows for data-driven decision-making regarding which promotion strategies are likely to be more effective, ultimately helping in optimizing marketing efforts and improving sales outcomes.

### Inference

The predictive analysis indicates varying sales performance for the two promotion types over the forecasted period. The results show that Promotion Type 2 is generally predicted to yield higher sales compared to Promotion Type 1 in most months. However, it is essential to continuously monitor actual sales data against these predictions to adjust promotion strategies accordingly. The analysis underscores the importance of using historical data to inform future marketing strategies, thereby enhancing the effectiveness of promotional campaigns and maximizing return on investment.

## Conclusion

This project aimed to analyze and predict the impact of different promotional strategies on retail sales. By leveraging such comprehensive dataset, we conducted an in-depth analysis using a combination of Python for exploratory data analysis and Tableau for visualizations.

We observed that Promotion Type 2 generally predicts higher sales compared to Promotion Type 1 in most months, suggesting that Promotion Type 2 strategies might be more effective in driving sales. However, we cannot draw definitive conclusions solely based on these results. The increase in sales units was observed only for Promotion Type 2 and had there been a similar increase for Promotion Type 1, the outcome might have been different. Additionally, Promotion Type 1 proved effective for certain companies when examined in more detail, indicating that its impact may vary across different contexts.

Overall, this project underscores the importance of using historical data to inform future marketing strategies. By understanding the dynamics of promotional strategies, retailers can make data-driven decisions that maximize revenue and profitability. The combination of Python and Tableau provided a robust framework for data analysis and visualization, enabling us to draw meaningful insights and actionable conclusions.

In conclusion, predictive analysis serves as a valuable tool for forecasting sales trends and evaluating the effectiveness of promotional strategies. Retailers can leverage these insights to enhance the effectiveness of their promotional campaigns, maximize return on investment, and ultimately drive business growth.