



Promotion Bump Assignment

Hasan Enes Güray

Table of Contents

1	Introduction.....	3
2	Dataset Overview.....	3
3	Objective	3
4	Methodology.....	4
4.1	Clustering of Products and Stores	4
4.1.1	Calculating Key Metrics	4
4.1.2	Quantile-Based Separation.....	4
4.1.3	Justification of the Method.....	4
4.2	Model to Measure Promotion Impact.....	5
4.2.1	Model Overview.....	5
4.2.2	Steps for Model Implementation.....	5
4.2.3	Rationale for Model	6
4.2.4	Interpretation of Results	6
4.3	Forecasting and Evaluating Future Promotions	6
4.3.1	Prediction	7
5	Results.....	8
5.1	The Effect of Promotion with Exploratory Data Analysis.....	8
5.2	The Effect of Promotion with Model	10
5.2.1	Product	10
5.2.2	Store	11
5.3	Promotion Analysis with Clustering	12
5.3.1	Effectiveness of Promotions	12
5.4	Test Model.....	14
5.5	Return Analysis	17
6	Recommendations	18
6.1	Tailored Promotion Strategies for Fast vs. Slow Items	18
6.2	Store-Specific Promotion Optimization	18
6.3	Strategic Timing and Duration of Promotions	18
6.4	Improvement of Predictive Models for Future Promotions	19
6.5	Exploration of Return Rates Post-Promotion.....	19
6.6	Further Data Collection.....	19
7	Conclusion.....	20
8	References	21

1 Introduction

The purpose of this analysis is to evaluate the effect of promotional campaigns on sales performance across different stores and products for a retail company. The main goal is to model the impact of promotions and offer actionable insights to the marketing department for future campaigns. For this assignment, we utilized sales data for a selected group of products, focusing on four major promotional campaigns that took place in 2015.

2 Dataset Overview

Three key datasets were provided for this analysis:

1. Sales Data (Assignment4.1a.csv and Assignment4.1b.csv):
 - The first file contains sales records from July 1, 2015, to August 1, 2015, with information on daily sales quantities for various products across multiple stores.
 - The second file extends the sales data from August 1, 2015, to January 1, 2016, which enables us to evaluate the predictive performance of the model developed in the initial period.
 - Each record in these files represents the sales (or return) of a specific product in a store on a particular date. The dataset includes 340 distinct stores and 317 distinct products.
2. Product Group Data (Assignment4.1c.csv):
 - This dataset categorizes each product into two levels of grouping: “ProductGroup” (10 distinct categories) and “ProductGroup2” (30 distinct subcategories). These classifications help in segmenting products for more granular analysis of promotion impacts.
3. Promotion Schedule (Promotiondates.csv):
 - The promotion file outlines six major promotional periods in 2015, with the first four campaigns running between February and June. The promotions vary in duration, and their impact on sales performance will be measured.

3 Objective

The analysis aims to answer several critical questions:

- What is the effect of promotions on different product groups and stores?
- Which products and stores benefit the most from promotions?
- Are there significant differences in promotion responses between fast-selling and slow-selling products and stores?
- How well can the model developed during the initial analysis predict future sales behavior during promotions?

This report outlines the methodology used for clustering products and stores into fast, medium, and slow categories based on their sales performance, the statistical models employed to analyze

promotional impacts, and a validation of the model's predictive power using the extended dataset from August to January.

Finally, recommendations will be made for future promotion strategies based on the findings, accompanied by a discussion of potential areas for improvement in the modeling process.

4 Methodology

4.1 Clustering of Products and Stores

To effectively analyze the impact of promotions on different products and stores, it was necessary to categorize them into distinct groups based on their sales performance during non-promotion periods. This clustering allowed us to differentiate between stores and products that exhibit varying sales velocities. The categories are defined as "Fast," "Medium," and "Slow" based on average weekly sales, which helps in isolating how each group responds to promotions. Here's the detailed process for clustering both products and stores:

4.1.1 Calculating Key Metrics

For each store and product, I calculated the following:

- `No_Promotion_Avg_Weekly_Sales`: The average weekly sales during non-promotion periods.

These metrics form the foundation of our clustering approach, as they reflect the typical sales behavior of products and stores across different periods.

4.1.2 Quantile-Based Separation

I used a quantile-based method (specifically `pd.qcut` in Python) to divide the products and stores into three clusters:

- Fast items/stores: The top 33rd percentile of the `No_Promotion_Avg_Weekly_Sales` distribution.
- Medium items/stores: The middle 33rd percentile of the `No_Promotion_Avg_Weekly_Sales` distribution.
- Slow items/stores: The bottom 33rd percentile of the `No_Promotion_Avg_Weekly_Sales` distribution.

This approach was chosen because it ensures that the dataset is split into relatively balanced groups, with an equal proportion of items and stores in each category. In scenarios where external benchmarks or thresholds are unavailable, using quantiles provides a logical and data-driven method to classify products and stores.

4.1.3 Justification of the Method

Using quantiles ensures that each category contains roughly the same number of products and stores, which provides a well-balanced representation of the entire dataset. This method is particularly useful in scenarios where there are no clear external standards for defining fast- or slow-moving products.

- Fast items/stores: These products and stores exhibit consistently high sales even during non-promotion periods, indicating strong baseline demand. Understanding how promotions affect this category is critical, as these items may show a lower relative lift from promotions.
- Slow items/stores: These have lower average sales in non-promotion periods, and promotions are often used to drive additional demand for these items. Analyzing this group's promotion sensitivity can reveal whether they benefit more from promotional efforts.

By clustering both products and stores in this way, I was able to compare the promotion effects across different categories, allowing for a more granular analysis and targeted recommendations for future campaigns.

4.2 Model to Measure Promotion Impact

To assess the impact of promotions on product sales and store performance, I utilized a regression-based approach. Two separate models were developed: one to analyze the impact of promotions on different products and another to analyze the impact on stores. The objective was to quantify how promotions affect sales, and whether this effect is statistically significant.

4.2.1 Model Overview

The models for both products and stores are built using Ordinary Least Squares (OLS) regression, where the target variable is the total SalesQuantity for each week, and the key explanatory variable is whether a promotion was active or not. By controlling for product and store differences using categorical variables (one-hot encoded), the model isolates the promotion effect.

- Dependent variable (y): The total SalesQuantity (sum of sales for each product or store per week).
- Independent variables (X):
 - Promotion: A binary indicator (1 for promotion periods, 0 for non-promotion periods).
 - ProductCode or StoreCode: One-hot encoded categorical variables to account for individual product/store differences.

4.2.2 Steps for Model Implementation

1. Week-Based Aggregation:
 - The dataset was aggregated at the weekly level to align sales data across products and stores. This allows for smoother trends and mitigates the variability seen in daily sales.
 - I created a new “Week” column that calculates the number of weeks since the custom start date of December 30, 2014.
2. Categorical Variables (One-Hot Encoding):
 - For both product and store models, categorical variables such as “ProductCode” and “StoreCode” were converted into binary columns using one-hot encoding. This enables the model to account for product-specific and store-specific effects without treating these codes as numerical values.

3. Handling Missing Columns:
 - As different products and stores might not appear in every week, I ensured that all possible product or store columns were included, assigning a default value of zero when necessary.
4. Feature Matrix and Target Variable:
 - I created the feature matrix (“X”) consisting of the Promotion variable and all one-hot encoded product/store variables. The target variable (“y”) is the total SalesQuantity for each week.
5. Model Training:
 - The data was split into training and testing sets using an 80-20 split. The OLS regression model was then trained on the training data.
6. Model Evaluation:
 - The model's performance was evaluated using the R-squared value, which measures the proportion of variance in the sales data explained by the model.
 - The p-value for the promotion coefficient (“ $P > |t|$ ”) was checked to determine whether the effect of promotion on sales is statistically significant.

4.2.3 Rationale for Model

- OLS Regression: Ordinary Least Squares is a standard method to quantify relationships between dependent and independent variables. It provides easy interpretability in terms of how much sales increase can be attributed to promotions, while controlling for differences between products and stores.
- Significance Testing: The p-value for the Promotion variable helps in determining if the promotion effect is statistically significant. This is key for making recommendations to the marketing team—if the effect of promotion is not significant, it may not be worth continuing the same strategy.
- R-Squared: The R-squared value provides an indication of how well the model fits the data, offering insight into the explanatory power of the model.

4.2.4 Interpretation of Results

- A statistically significant p-value (typically less than 0.05) for the promotion coefficient suggests that the promotion has a meaningful impact on sales.
- A higher R-squared value indicates that the model successfully explains the variation in sales due to promotions and other factors. However, lower values might indicate the need for more sophisticated models or inclusion of additional variables (such as external factors like holidays).

By using these models, we can quantify the promotion’s impact on different products and stores, allowing for a data-driven approach to optimizing future marketing efforts.

4.3 Forecasting and Evaluating Future Promotions

To evaluate how well the models predict the impact of future promotions, I used the data from Assignment4.1b.csv, which contains sales data from August 1, 2015, to January 1, 2016. This data allows us to assess the predictive performance of the models trained using the initial data (July 1 to August 1, 2015). The steps involved in the forecasting and evaluation process are described below.

1. Preparing the Test Data
 - The first step is to prepare the test dataset in a manner consistent with the training data. This ensures that the model can accurately predict sales quantities based on the learned patterns.
2. One-Hot Encoding:
 - Similar to the training data, I converted the “ProductCode” and “StoreCode” in the test dataset into categorical variables using one-hot encoding. This allows the model to interpret the test data correctly based on the encoded product and store identifiers.
3. Handling Missing Columns:
 - Since not all products or stores may appear in the test set, I ensured that all possible product and store columns were included, filling in any missing columns with zeros. This ensures consistency between the training and test data.
4. Feature Matrix for Testing:
 - The feature matrix (“X_test_new”) was constructed by including the promotion variable and the encoded product/store columns. A constant was also added for the model intercept, just as in the training phase.

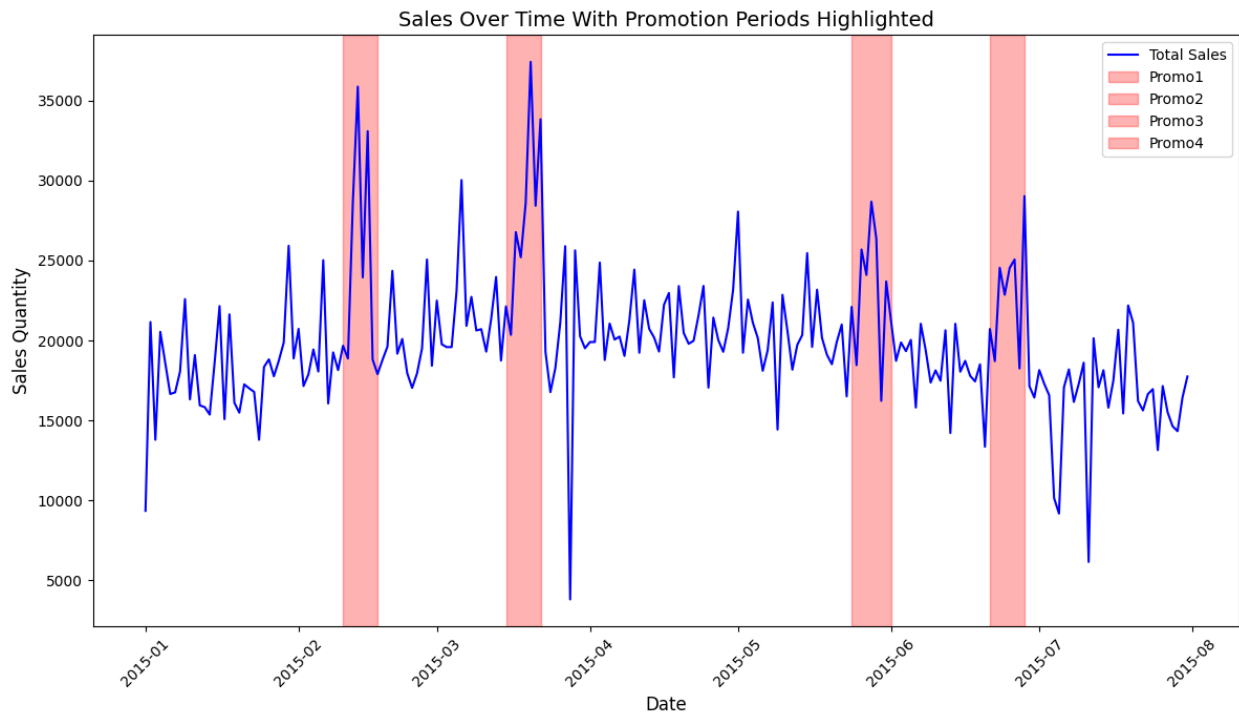
4.3.1 Prediction

With the prepared test data, I used the trained models to predict sales quantities during the future promotion periods.

1. Product Model Predictions:
 - The product model, which was previously trained, was used to predict sales quantities for the new data points in the test set.
2. Store Model Predictions:
 - Similarly, the store model was used to forecast the sales quantities based on store-level data in the test set.
3. Evaluation Metrics
 - To evaluate how well the models performed in predicting future sales during the promotion periods, I used two key metrics: Root Mean Squared Error (RMSE) and R-squared.
4. Root Mean Squared Error (RMSE):
 - RMSE measures the average magnitude of the prediction error. It indicates how closely the predicted sales quantities match the actual observed values. A lower RMSE value indicates better model performance.
5. R-squared:
 - The R-squared value provides insight into how much of the variance in the observed sales is explained by the model. A higher R-squared value indicates that the model is better at explaining the sales trends based on promotions.
6. Model Evaluation:
 - For the product model, the RMSE and R-squared values were calculated to assess how well the model predicted future sales during Promotion 5 for different products.
 - For the store model, similar metrics were used to evaluate the model’s performance at the store level.

5 Results

5.1 The Effect of Promotion with Exploratory Data Analysis

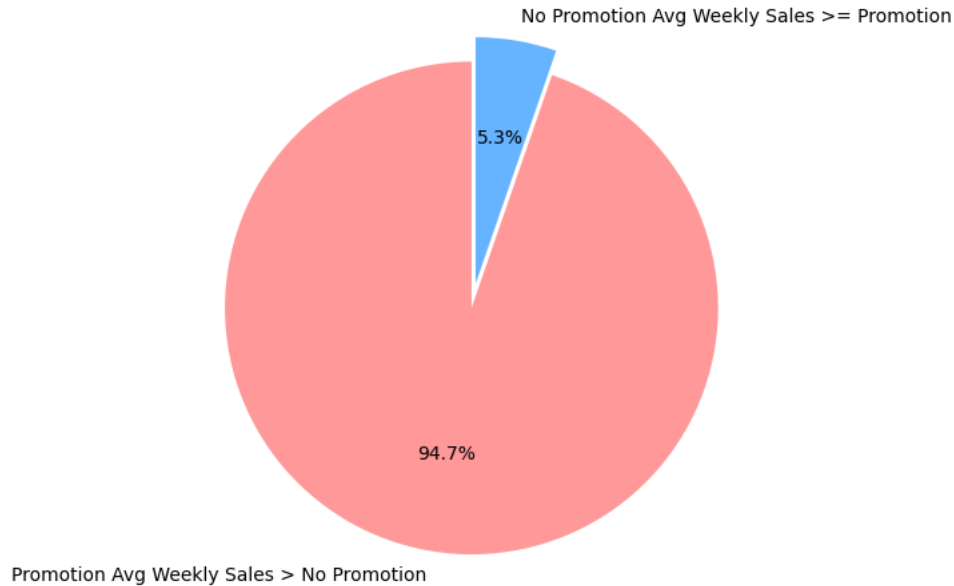


The graph depicts Sales Over Time with Promotion Periods Highlighted in red. The sales data, shown as a blue line, tracks the total sales quantity from the start of 2015 through August 2015. Significant peaks in sales are visible during the promotion periods, indicating a clear positive impact of promotions on overall sales volumes.

The first promotion (Promo1), which occurred in February 2015, shows a sharp spike in sales. Promo2 in March also led to a sales increase, with quantities reaching their highest point during the observed period. Promo3 and Promo4 in May and June show similar patterns, with substantial increases in sales quantities during the promotional periods.

This visualization provides a clear indication that promotions have a tangible effect on sales, causing temporary spikes in demand. The consistency of sales increases during promotion periods suggests that the marketing campaigns are effective in driving customer purchasing behavior. However, post-promotion sales tend to decline or stabilize back to pre-promotion levels, highlighting the short-term nature of the promotion effect.

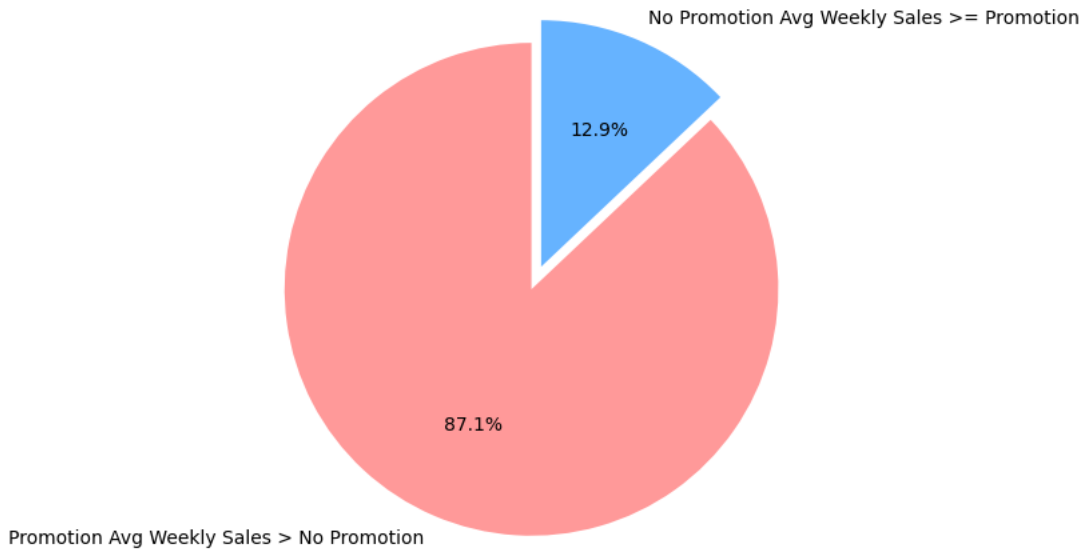
Stores: Promotion vs No Promotion Avg Weekly Sales



The pie chart compares Promotion vs No Promotion Average Weekly Sales for stores, showing the percentage of stores where promotional sales exceed non-promotional sales. The chart reveals that 94.7% of stores experienced higher average weekly sales during promotional periods compared to non-promotion periods, while only 5.3% of stores had similar or higher sales during non-promotion periods.

This result underscores the overall effectiveness of promotions in increasing store sales, as the vast majority of stores saw a significant sales uplift during promotional campaigns. The small percentage of stores where non-promotion sales were higher or similar suggests that these stores either have consistent demand for certain products regardless of promotions or may not be as responsive to promotion-driven strategies.

Product: Promotion vs No Promotion Avg Weekly Sales



The pie chart illustrates the comparison between Promotion vs No Promotion Average Weekly Sales for products. The data shows that 87.1% of products experienced higher average weekly sales during promotion periods compared to non-promotion periods, while 12.9% of products had equal or higher sales during non-promotion periods.

This indicates that the majority of products benefit from promotional activities, resulting in increased sales. However, there is a notable subset of products (12.9%) where sales were not significantly impacted by promotions, possibly due to these products being consistent sellers without the need for additional marketing efforts, or due to limited responsiveness to promotions.

5.2 The Effect of Promotion with Model

5.2.1 Product

OLS Regression Results						
Dep. Variable:	SalesQuantity	R-squared:	0.812			
Model:	OLS	Adj. R-squared:	0.805			
Method:	Least Squares	F-statistic:	110.0			
Date:	Tue, 24 Sep 2024	Prob (F-statistic):	0.00			
Time:	20:01:52	Log-Likelihood:	-63026.			
No. Observations:	8384	AIC:	1.267e+05			
Df Residuals:	8066	BIC:	1.289e+05			
Df Model:	317					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	329.9782	85.848	3.844	0.000	161.695	498.262
Promotion	-58.4238	12.136	-4.814	0.000	-82.214	-34.634
ProductCode_2	-88.4373	115.834	-0.763	0.445	-315.502	138.627
ProductCode_3	-132.5105	117.459	-1.128	0.259	-362.760	97.738
ProductCode_4	-105.2510	118.342	-0.889	0.374	-337.231	126.729

The OLS Regression Results provide a detailed statistical summary of the model used to measure the effect of promotions on sales quantities across products. Key findings from the model are as follows:

- R-squared: The model has an R-squared value of 0.812, meaning that 81.2% of the variation in sales quantities is explained by the model, which indicates a strong fit. The adjusted R-squared of 0.805 confirms this, even after accounting for the number of predictors in the model.
- Promotion Coefficient: The p-value is less than 0.05, which suggests that the effect of promotions on sales quantities is statistically significant. However, the negative coefficient indicates that the overall sales quantity might decrease during the promotion period. This could be due to the constant value being too large.
- F-statistic: The F-statistic of 110.0 with a p-value of 0.00 confirms that the model is statistically significant overall.

In summary, the results indicate that promotions have varied effects across different products, with some benefiting greatly and others showing diminished sales. This suggests that future promotional efforts should be tailored to target specific products that respond more positively to marketing campaigns.

5.2.2 Store

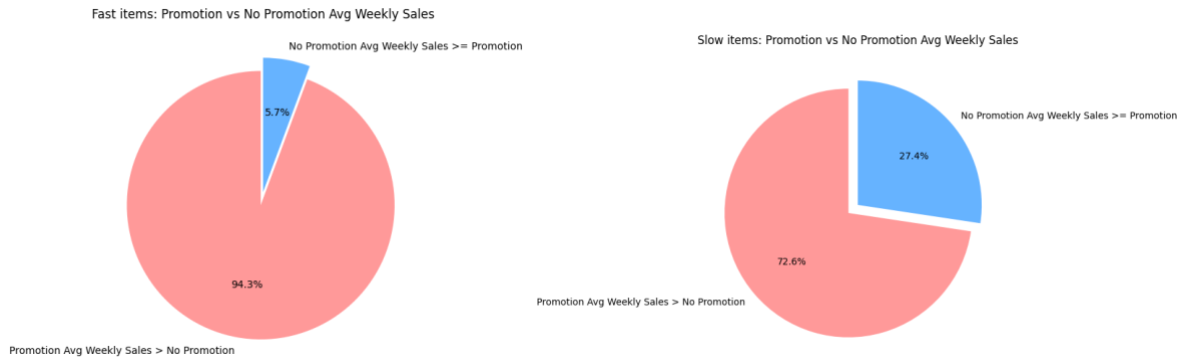
OLS Regression Results						
=====						
Dep. Variable:	SalesQuantity	R-squared:	0.599			
Model:	OLS	Adj. R-squared:	0.584			
Method:	Least Squares	F-statistic:	41.58			
Date:	Tue, 24 Sep 2024	Prob (F-statistic):	0.00			
Time:	20:01:54	Log-Likelihood:	-64444.			
No. Observations:	9791	AIC:	1.296e+05			
Df Residuals:	9451	BIC:	1.320e+05			
Df Model:	339					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	349.6888	31.947	10.946	0.000	287.065	412.312
Promotion	-47.7255	4.379	-10.899	0.000	-56.309	-39.142
StoreCode_2	845.4105	45.165	18.718	0.000	756.878	933.943
StoreCode_3	-103.0359	46.807	-2.201	0.028	-194.788	-11.284
StoreCode_4	32.8988	44.811	0.734	0.463	-54.940	120.738

- R-squared = 0.599: This indicates that approximately 59.9% of the variation in sales quantities is explained by the store codes and promotions in the model.
- Adj. R-squared = 0.584: This adjusts the R-squared for the number of predictors and shows a slightly lower explanatory power due to the high number of variables.
- Key Coefficients and Effects:
 - Constant (Intercept): The constant is 349.6888, meaning that in the absence of any promotion or effect from a store code, the average sales quantity is approximately 350 units.
 - Promotion: This effect is statistically significant ($p < 0.001$), implying a significant impact of promotions in this dataset.

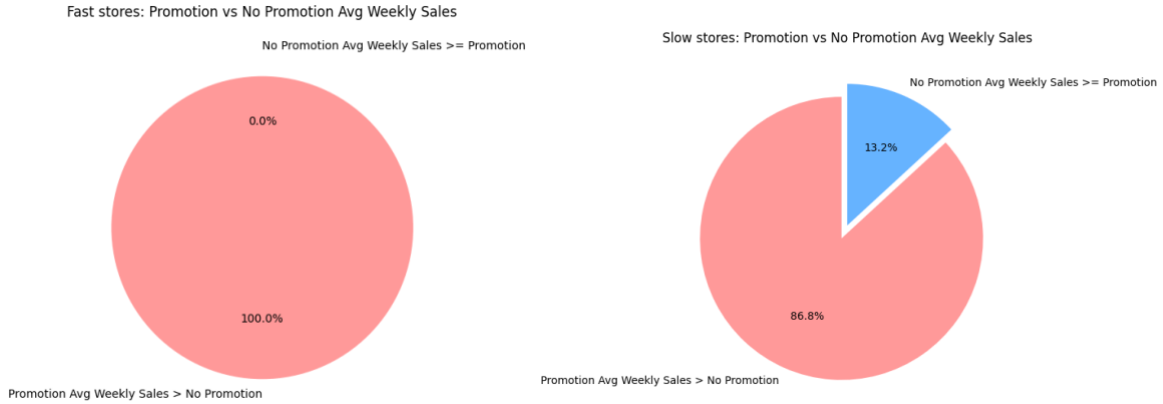
- Model Performance:
 - F-statistic = 41.58, $p < 0.001$: The model as a whole is statistically significant, meaning that the independent variables (store codes and promotion) together have a significant relationship with sales quantities.

5.3 Promotion Analysis with Clustering



5.3.1 Effectiveness of Promotions

- Promotions are more effective for fast-moving items compared to slow-moving ones. Fast items see a much greater boost from promotions, with over 94% of cases showing that promotions lead to higher sales.
- For slow-moving items, promotions are still beneficial, but they have a smaller effect. About 27% of slow items see similar or better sales without promotions, compared to only 5.7% of fast items.
- For fast-moving items, continuous or more frequent promotions could maximize sales due to the clear advantage promotions bring.
- For slow-moving items, it might be worth analyzing individual product performance without promotions, as some may not benefit significantly from promotional efforts, saving promotional costs in those cases.
- Strategic Implications:
 - Fast items: Maintain aggressive promotional strategies to capitalize on the high effectiveness of promotions.
 - Slow items: Consider selective promotional strategies. Focus on products that show higher sensitivity to promotions while potentially minimizing promotions for items that perform equally well without them.



- **Promotion Dependence:** Fast stores show a complete reliance on promotions to drive sales, as no store had equal or better performance without promotion. In contrast, slow stores show a small but significant percentage where promotions don't necessarily lead to better results.
- **Effectiveness of Promotions:** The effectiveness of promotions is universal in fast stores, while in slow stores, it is dominant but not absolute.
- **Key Similarities:**
 - Both fast and slow stores largely benefit from promotions. In both cases, the majority of sales increases occur during promotion periods, making promotions a key strategy for boosting sales across all store types.

This suggests a tailored promotional strategy might be effective: aggressive promotions for fast stores to sustain high performance, while more nuanced strategies for slow stores, perhaps exploring what factors contribute to sales when promotions aren't running, would be useful.

Promotion	StoreCode	0	1	No_Promotion_Avg_Weekly_Sales	Promotion_Avg_Weekly_Sales	Avg_Weekly_Sales	Store_Cluster	Promotion_Effect
91	92	2662.0	1114.0	103.522222	236.303030	124.093897	Slow stores	1.282631
163	164	2705.0	854.0	105.194444	181.151515	116.962441	Slow stores	0.722064
144	145	11872.0	3724.0	461.688889	789.939394	512.544601	Fast stores	0.710978
182	183	7593.0	2323.0	295.283333	492.757576	325.877934	Medium stores	0.668762
188	189	6670.0	2014.0	259.388889	427.212121	285.389671	Slow stores	0.646995

- The **Promotion_Effect** column shows how strongly a store's average weekly sales respond to promotions, calculated by comparing Promotion Avg Weekly Sales to No Promotion Avg Weekly Sales.
- Store 92 has the highest promotion effect with a value of 1.282631. This means its sales increased by 128.3% during the promotion period compared to non-promotion periods, indicating it is the most responsive store to promotions.

Promotion	ProductCode	0	1	No_Promotion_Avg_Weekly_Sales	Promotion_Avg_Weekly_Sales	Avg_Weekly_Sales	Product_Cluster	Promotion_Effect
290	291	3.0	3.0	0.116667	0.636364	0.197183	Slow items	4.454545
12	13	6.0	3.0	0.233333	0.636364	0.295775	Slow items	1.727273
228	229	17.0	7.0	0.661111	1.484848	0.788732	Slow items	1.245989
121	122	98.0	34.0	3.811111	7.212121	4.338028	Slow items	0.892393
65	66	241.0	82.0	9.372222	17.393939	10.615023	Slow items	0.855903

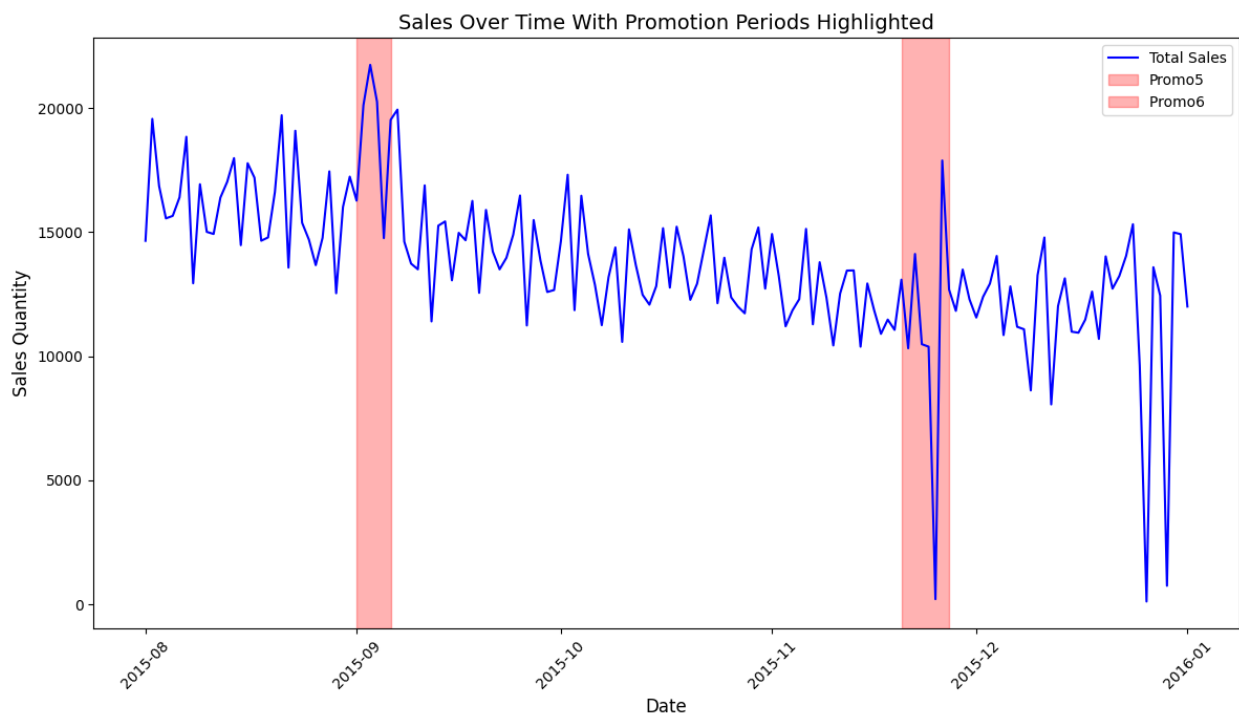
- Product 291 has the highest promotion effect with a value of 4.454545, indicating that its sales increased by more than 4.45 times during the promotion period compared to non-promotion periods. This product is the most responsive to promotions.

The biggest effect explaining sales change during promotions is primarily tied to the promotion itself. Looking at both stores and products, we see that promotion has a significant impact on sales. However, beyond the promotion itself, factors like store cluster and product cluster also play important roles in moderating the impact of promotions.

There is a significant difference between the promotion impacts on Fast versus Slow items. Fast items show more impact from promotions compared to Slow items. These items are more sensitive to promotions, as their sales are often more significantly boosted during promotion periods. The clustering of fast vs slow items also indicates that fast items are more likely to respond positively to promotions.

There is a significant difference between the promotion impacts on Fast versus Slow stores. Fast stores appear to rely heavily on promotions, with 100% of their sales being higher during promotional periods. There are no instances where sales in fast stores were better without a promotion. This suggests that slow stores are somewhat less sensitive to promotions compared to fast stores, but still experience substantial positive effects from promotional activities. The biggest factor driving sales changes during promotions is the presence of the promotion itself, especially for fast items and stores. Fast stores show a stronger dependency on promotions to boost sales compared to slow stores, which still benefit but are less reliant on promotions for their overall performance.

5.4 Test Model

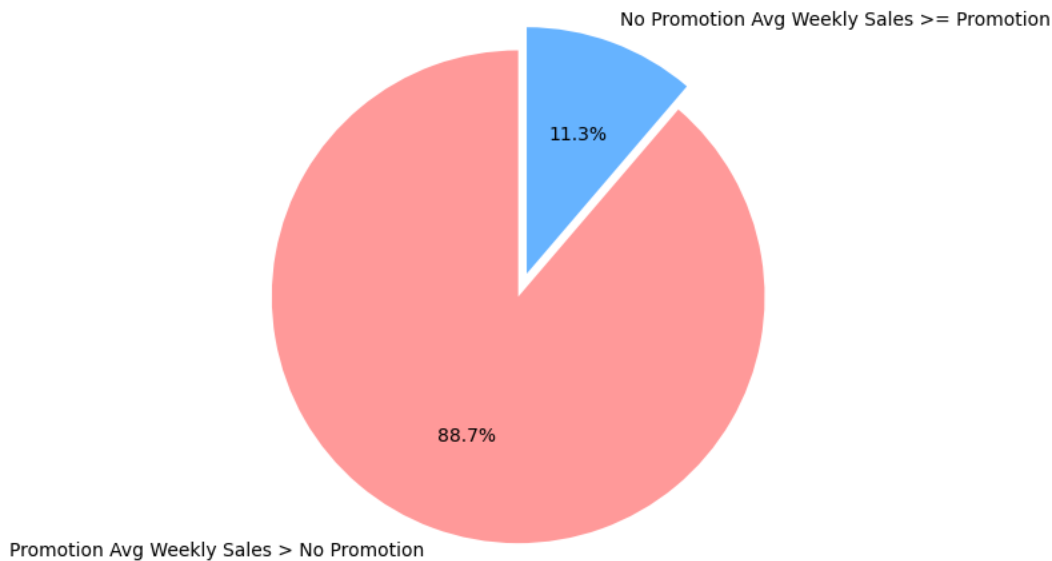


Sales quantity in this period shows a more stable pattern with fewer extreme peaks compared to the earlier period. Promo5 and Promo6 have moderate impacts on sales compared to Promo1 through Promo4. While promotions still lead to an increase in sales, the overall magnitude of change seems less pronounced in the latter half of the year. Overall, the first period (January to August 2015) demonstrates larger and more volatile sales increases during promotions, while the second period (August 2015 to January 2016) has relatively stable sales with promotions having a more tempered effect.



In the second period, the impact of promotions was slightly less significant, with a higher percentage of stores showing comparable or better sales during non-promotion periods. However, the majority of stores still demonstrated a clear sales boost during promotions across both periods, with the first period showing slightly more effectiveness in the promotional impact. This suggests that, while promotions were effective in both time frames, the latter half of the year saw a small decline in the promotional influence on store sales.

Test Product: Promotion vs No Promotion Avg Weekly Sales



The promotion's effect on products remained consistent across both periods, with a slightly higher percentage of products benefiting from promotions during the second period (August 2015 - January 2016).

From the results:

- Product-level model:
 - R^2 : 0.46, meaning around 46% of the variance in sales quantity is explained by the model. This is moderate but not excellent.
 - RMSE: 491.22, which indicates an average error of approximately 491 units in the predicted sales quantity. This can be considered a relatively high error depending on the scale of the sales data.
- Store-level model:
 - R^2 : 0.21, meaning only 21% of the variance in sales quantity is explained by the model. This is quite low, indicating the model struggles to predict sales quantities at the store level.
 - RMSE: 172.88, which indicates an average error of approximately 173 units in the predicted sales quantity.

Overall, the product-level model performs better than the store-level model based on both R^2 and RMSE. However, both models exhibit a certain degree of misfit, particularly at the store level.

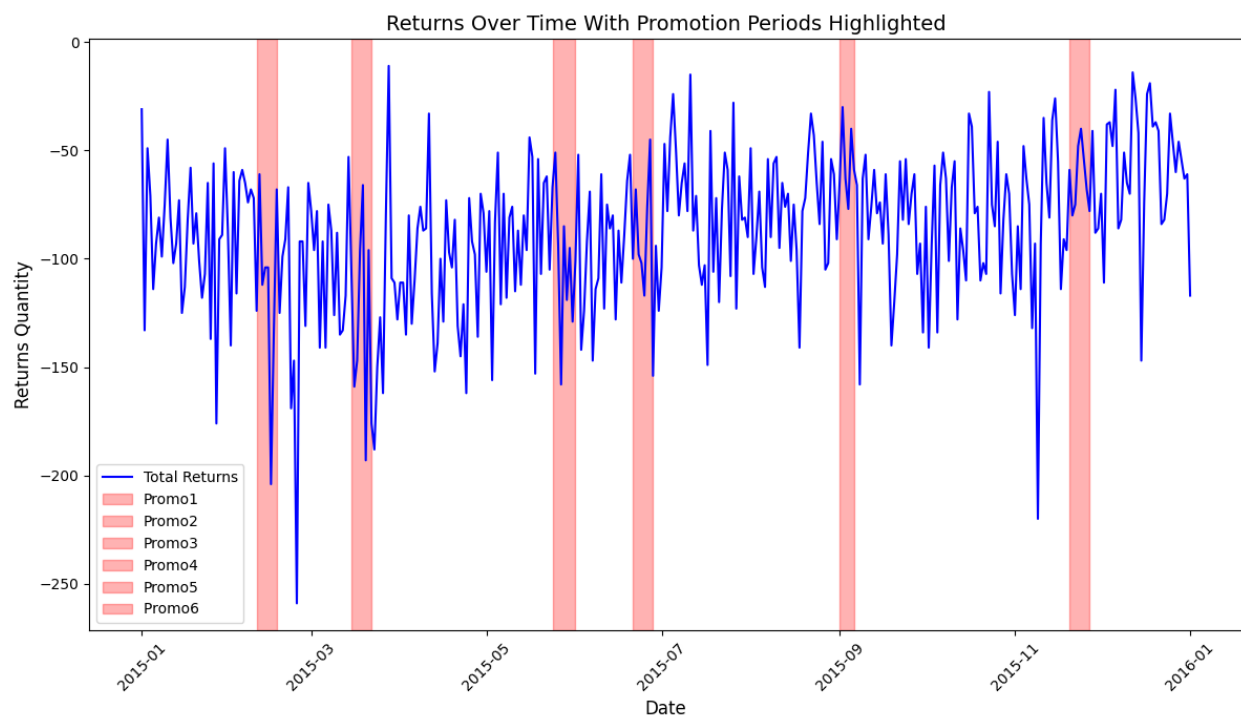
The main problem points causing bad fits in the current model can be attributed to several factors. First, multicollinearity may be present between different product codes or store codes, which can negatively impact the model's predictive power. When product or store codes are highly correlated, the model may struggle to assign the correct weights to each feature, leading to confusion. Second, the models may be missing relevant features that influence sales, such as seasonality, holidays, discounts, customer trends, and macroeconomic conditions. These omitted factors might explain some of the unexplained variance in the predictions.

Another issue is the possibility of overfitting to the training dataset, particularly when dealing with categorical variables like product or store codes. This can lead to poor generalization to unseen test data, resulting in subpar performance on new promotional data. Additionally, the explanatory power of promotions may be limited for certain products or stores, particularly those that are slow-moving or less responsive to promotions, which could diminish the model's accuracy. Lastly, heterogeneity in how different products and stores respond to promotions may also contribute to prediction errors. Some fast-moving stores or items might benefit disproportionately from promotions, while slower-moving ones may show little to no impact, further complicating the model's ability to predict accurately.

To address these issues, several adjustments can be made to improve the model. Introducing interaction terms between promotion and product/store features can help capture the varying responses to promotions across different products and stores. Additionally, adding time-based features such as month, quarter, or week of the year will allow the model to account for seasonal effects that significantly drive sales in retail data. Using lagged sales quantities as features can also improve predictions by capturing the temporal dependencies often present in sales data.

Regularization techniques such as Ridge or Lasso regression should be applied to reduce overfitting, especially in models with numerous categorical variables. Furthermore, more complex models like Random Forest or Gradient Boosting can be considered to better capture the non-linear relationships between promotions and sales. Lastly, handling outliers in the sales data, particularly extreme values during promotions, can help prevent skewed results and enhance the model's performance.

5.5 Return Analysis



To evaluate whether there is a significant difference in item return rates after promotions, we can begin by analyzing the visual representation of returns over time, with the promotion periods

clearly marked. The graph illustrates fluctuations in total returns, with frequent spikes, particularly during and after the promotional periods. Notably, these periods, highlighted in red, seem to coincide with an increase in return volatility, with sharper downward movements indicating higher returns. This suggests that promotions might have an impact on return rates.

To explore this relationship further, it would be useful to calculate the average return rates for periods before, during, and after promotions. A simple comparison of these averages could reveal whether return rates increase or decrease in response to promotions. However, to draw more robust conclusions, a statistical test, such as a paired t-test or a non-parametric equivalent, would be appropriate. These tests would help determine if the observed changes in return rates during promotion periods are statistically significant.

6 Recommendations

6.1 Tailored Promotion Strategies for Fast vs. Slow Items

- **Fast-Moving Items:** Promotions have shown a significant positive impact on fast-moving items, with over 94% of these products experiencing a substantial sales lift during promotional periods. To maximize sales, the marketing team should consider more frequent promotions for these items, as they are likely to generate high returns. Additionally, bundling fast-moving products with slower-moving items could help boost sales for the latter without saturating demand for the former.
- **Slow-Moving Items:** While promotions benefit slow-moving items, the effect is less pronounced, and some products do not exhibit significant sales increases during promotion periods. A selective approach should be considered for promoting these items. Promotions could be targeted only at items that show a greater sensitivity to promotional efforts, which could help reduce promotional costs and prevent unnecessary discounting of products that are unlikely to benefit from such efforts.

6.2 Store-Specific Promotion Optimization

- **Fast Stores:** Fast stores exhibit a heavy reliance on promotions, with all fast stores showing higher sales during promotional periods. To maintain this momentum, the company should continue aggressive promotional strategies in these stores. Further, testing multiple types of promotions (e.g., discounts, limited-time offers) could help identify which formats drive the greatest sales uplift in these high-performing locations.
- **Slow Stores:** While slow stores also benefit from promotions, 13.2% of them perform equally well or even better during non-promotional periods. For these stores, it may be worth exploring alternative sales strategies, such as loyalty programs, customer engagement campaigns, or product assortments tailored to local preferences. Reducing the frequency or intensity of promotions in slow stores that don't react as strongly could also save on marketing expenses.

6.3 Strategic Timing and Duration of Promotions

The analysis suggests that promotions drive short-term sales spikes, but post-promotion sales often return to pre-promotion levels. This indicates that promotions are effective in generating immediate demand but may not have long-term effects on customer behavior. To mitigate this,

promotions could be designed to encourage repeat purchases, such as offering incentives for future purchases (e.g., vouchers or discounts on the next purchase). Additionally, the company should consider testing different promotion lengths to determine the optimal duration that balances cost and sales impact.

6.4 Improvement of Predictive Models for Future Promotions

The current models explain a significant portion of the sales variation due to promotions, but there are areas where performance can be improved. Introducing interaction terms between promotion effects and product/store clusters could capture more nuanced responses. Furthermore, incorporating additional features such as seasonality, holidays, or external factors (e.g., economic conditions) may improve model accuracy. Advanced modeling techniques such as Random Forest or Gradient Boosting could also be explored to better capture complex relationships in the data.

6.5 Exploration of Return Rates Post-Promotion

The analysis indicates a potential increase in return rates during and after promotional periods. Further analysis should be conducted to identify the specific products or stores with higher return rates, allowing the company to adjust its promotional strategies accordingly. For example, stricter return policies during promotions or better product descriptions could help mitigate the higher returns.

6.6 Further Data Collection

To further enhance the analysis and provide more actionable insights, several additional datasets would be valuable. These datasets would allow a deeper understanding of the promotional impacts, not just on sales volume but also on profitability, market positioning, and competition.

1. Pricing Data

- Rationale: While this assignment focuses on sales quantities, it's important to understand how promotions affect profitability. Sales volume alone doesn't provide a complete picture of success, profit margins may vary across products, and promotions often involve price discounts. By incorporating product pricing data before, during, and after promotions, we can calculate the true impact of promotions on profits.
- How to Obtain: Pricing data can typically be extracted from the company's internal sales and financial systems, where historical prices for each product are stored. If dynamic pricing strategies are employed, it would be important to include those variations in the analysis.

2. Cost Data

- Rationale: To assess the true profitability of promotions, understanding the cost associated with each product is essential. Costs include production, storage, and marketing expenses during promotions. With both pricing and cost data, we could calculate the profit margin and evaluate whether promotions lead to profitable sales or just increased volume at reduced margins.
- How to Obtain: Cost data can usually be sourced from the company's supply chain, procurement, and financial departments. This data would need to include both fixed and variable costs for producing and marketing each product.

3. Competitor and Market Data

- Rationale: Understanding the competitive landscape would provide critical context for the company's performance during promotions. Competitor data could reveal whether sales changes are due to internal promotions or external factors, such as competitors' marketing campaigns or pricing adjustments. Market trends and sector-specific data would also help benchmark performance against industry standards.
- How to Obtain: Competitor data can be gathered through market research, industry reports, or third-party data providers. Tracking competitor pricing and promotions using web scraping tools or competitive intelligence software can provide real-time insights into the market environment.

4. Product Category and Sector Data

- Rationale: Understanding which sector the company operates in and the characteristics of the products being sold is vital for designing tailored promotion strategies. Certain sectors (e.g., electronics, fashion) respond differently to promotions compared to others (e.g., FMCG). Additionally, customer preferences, product lifecycle stages, and seasonal trends in different sectors can significantly affect promotion outcomes.
- How to Obtain: This information can typically be sourced from internal product databases, market analysis reports, or customer segmentation data. Insights from the marketing and product teams would help in classifying products more effectively and understanding sector-specific dynamics.

5. Customer Demographics and Behavioral Data

- Rationale: Customer preferences and demographics play a significant role in how promotions perform. Certain customer segments may be more responsive to price reductions or specific types of promotions (e.g., loyalty programs, flash sales). Demographic data (e.g., age, location, income levels) and customer purchasing behaviors can provide insights into which customer groups are driving promotional success.
- How to Obtain: This data can be obtained from customer loyalty programs, CRM systems, or third-party market research agencies. Analyzing transaction history can also reveal valuable patterns in customer behavior during promotional periods.

7 Conclusion

This analysis demonstrates the significant impact of promotions on sales performance across various product and store categories for the retail company. By clustering products and stores into Fast, Medium, and Slow categories, we were able to provide a more granular understanding of how promotions influence different segments. The findings show that fast-moving products and stores benefit most from promotions, with consistent sales spikes during promotional periods, while slow-moving items and stores exhibit more varied responses.

The OLS regression models provided a statistical basis for quantifying the effect of promotions, revealing that while promotions are generally effective, their impact can vary significantly based on the product or store. Fast items and stores showed stronger promotional effects, making them prime targets for future campaigns. However, a more nuanced approach may be needed for slow-

moving items, where selective promotion strategies could optimize results without overspending on marketing efforts.

The evaluation of the model's performance on future data, while moderately successful, highlighted areas for improvement. Enhancements such as including interaction terms, seasonality, and more complex machine learning models could provide better accuracy in predicting future sales behavior. Moreover, the analysis of return rates suggested a potential link between promotions and increased product returns, warranting further investigation into the costs and benefits of promotional strategies.

Ultimately, the key takeaway from this analysis is the need for tailored promotion strategies. Promotions should be designed to maximize the sales uplift in fast-moving products and stores, while slower-moving segments might benefit from a more measured and selective approach. By refining promotional campaigns based on data-driven insights, the company can improve both sales performance and profitability, ensuring a more effective allocation of marketing resources.

8 References

assignment4.1a.csv , provided by Invent Analytics

assignment4.1b.csv , provided by Invent Analytics

assignment4.1c.csv , provided by Invent Analytics

PromotionDates.csv , provided by Invent Analytics