Jakob Kuemmerle, Lowan Li, Linda Liang, Yidan Wang

Final Report

MLDS 400

Dec 8, 2023

Enhancing Retail Profitability through Data-Driven Decision Making: A Case Study on Price Elasticity and Discount Strategy Optimization at Dillard's

Executive Summary

This project is designed to transform Dillard's discounting strategies using a data-driven approach, focusing on the 4.5% of SKUs that significantly contribute to revenue. We conducted thorough data cleaning and feature engineering, incorporating seasonality and demographic variables. The Apriori Market Basket Algorithm, augmented with PrefixSpan, was employed to analyze key items, and Random Forest models were used to predict sales for each SKU. Our research indicates that customized discount strategies can greatly enhance profits, but they need strategic focus and must be validated through field experiments. Immediate actions include prolonging the duration of discounts and adjusting prices, while our long-term strategy involves improving data collection and extending our models to all product lines. This initiative is vital for profit growth and marks a significant shift in Dillard's business approach.

Introduction

In this project, we are redefining the way Dillard's approaches its discount strategies. Given the variety of Dillard's product offerings and their diverse pricing strategies, there lies a significant opportunity for optimization. A key challenge with the current system is its insufficient consideration of customer behavior, which often results in lost revenue opportunities. By concentrating on the 4.5% of stock-keeping units (SKUs) that account for half of the company's revenue and adopting a data-driven approach to discounting, our aim is to tailor discounts to align with customer preferences and buying patterns. This initiative is dedicated to refining pricing strategies, boosting revenue, and improving customer satisfaction by offering a shopping experience that is customized to meet individual customer needs.

Methodology

EDA and Feature Engineering

Our analysis began with a comprehensive exploratory data analysis (EDA) of extensive transactional datasets. To facilitate this process, we imported all five tables – SKSTINFO, STRINFO, SKUINFO, DEPTINFO, and TRNSACT – into a PostgreSQL database, given the challenges posed by the large size of datasets. We also verified relevant columns within Dillard's transaction data, rectifying erroneous entries such as transactions with incorrect pricing, and addressing gaps in the dataset. Where original prices were missing, we assigned the known price of identical SKUs, and in cases where no such data existed, we imputed these values using the mean retail price. This stage extended to unveil purchase behavior trends, seasonality effects, and price sensitivity while exploring sales data from various dimensions. Noteworthy discoveries included an increase in sales leading up to Christmas and peaks throughout the year, presumably associated with holidays. Additionally, we observed a clear weekly pattern indicating higher sales during weekends. These insights allowed us to identify anomalies and critical performance indicators, providing actionable information on customer behavior and sales performance.

Feature engineering, a pivotal step that followed, involved transforming raw data into informative predictors that could significantly improve the predictive power of our models. By creating new variables that captured the core aspects of the patterns in the transactional data, we managed to represent complicated connections within the data in a clear and concise manner.

This included introducing variables that capture the seasonality in sales (*Figure 1*), such as 'month'(January=0) and 'day of week'(Monday=0, Sunday=6), and discount (*Figure 2*) calculated as (original_price – amount_paid)/ original_price. We also added a binary weekend indicator, marking Friday to Sunday as 1, to distinguish weekend sales trends. Additionally, we identified customer 'baskets' (*Figure 3*) within the transaction data with a subset of primary keys, setting the stage for association rule mining. We also took into account socio-economic factors by integrating data on median earnings and population statistics specific to the zip codes of each Dillard's location (*Figure 4*). To obtain this valuable information, we employed web scraping techniques to gather relevant data from online sources (uspopulation.org). This approach helped us understand the economic context of our customer base.

Market Basket Algorithm

Employing the Apriori Market Basket Algorithm, augmented with the PrefixSpan technique, our analysis encompassed 100,000 baskets that included items critical to revenue generation. We specifically targeted products that account for 50% of our revenue, items with multiple price points, and those with a known launch date. In our analytical process, we set a lift threshold exceeding 30% and established a support level at 0.001, aiming to guarantee that the insights derived were both significant and applicable on a larger scale.

Modeling and Simulation

In the modeling and simulation phase, we adopted random forest regression models to predict sales for each stock-keeping unit (SKU). This decision was based on the superior performance of the Random Forest algorithm compared to the Gradient Boosting Regressor (GBR). Notably, the Random Forest models demonstrated a mean squared error (MSE) of 16.88, significantly lower than the GBR model's MSE of 21.96. Additionally, the Random Forest approach yielded a higher R-squared value of 53%, in contrast to the GBR model's 38%, indicating a more accurate and reliable predictive capability. The models were refined using key features, including population and income data obtained through web scraping by zipcode, the time elapsed since product launch, day of the week, month, and the average discount level. For items identified via the Association Rule Analysis, we conducted a Price Demand Model analysis on an individual basis. To gauge the demand response, we simulated a range of discounts from 0% to 80% in 5% increments. For each proposed discount level, we projected the demand, computed the profit (by deducting cost from the selling price), and subsequently determined the most profit-maximizing discount strategy for each SKU.

Findings

Our findings indicate that optimal discount strategies for specific SKUs can significantly boost profits. For example (*Figure 5*), setting a 10% discount on SKU 6696135 and a 5% discount on SKU 6656135 can potentially increase profits by over $20,000. However, a limitation of our current approach is the independent consideration of each item, which may overlook broader sales dynamics. To address this, we propose conducting field experiments to explore discount strategies based on association rules, aiming to identify which discounts on particular items can drive overall sales increases.

The findings indicate that while strategic discounting can boost profits, it's important to note that discounts are just one of many factors influencing consumer demand. Factors like market trends, economic conditions, and special events, which our model does not account for, are also vital. Thus, while our results are encouraging, they should be implemented with caution and a focus on adaptable pricing and ongoing market analysis. Moreover, modeling each item individually, though effective, is computationally demanding, necessitating a selective approach for practicality. It's crucial to validate these models with real-world experiments to confirm their alignment with actual market trends and customer behaviors.

Recommendations

To improve Dillard's discount strategy, we initially decided to extend discount periods. This extension will yield more comprehensive data on customer reactions, enhancing our analysis of sales trends and discount impact. Alongside this, we recommend targeted price adjustments for specific SKUs to test our model's accuracy and gather immediate insights.

For our long-term strategy, we plan to broaden our approach by collecting more extensive customer purchase history data. This will help us better understand consumer buying patterns and uncover new insights. We also aim to improve our price elasticity models by adding advanced features, making them more accurate in predicting the impact of price changes on consumer demand. An important step will be to apply these enhanced models across all Dillard's product lines. This widespread application will maximize the benefits of our data-driven strategy, applying our insights across the full product range to drive sales growth and make our discount strategies more effective and efficient.

Conclusions

This project presents a pivotal opportunity for Dillard's to enhance its profitability through a data-driven discount strategy. By harnessing detailed transaction data and advanced analytical models, we can optimize pricing, align discounts with consumer behavior, and unlock significant revenue potential. The need for immediate and strategic action is clear. This initiative, with its potential for scalability across various product lines, represents not just an improvement in sales tactics but a transformative step in Dillard's overall business strategy.

Appendix

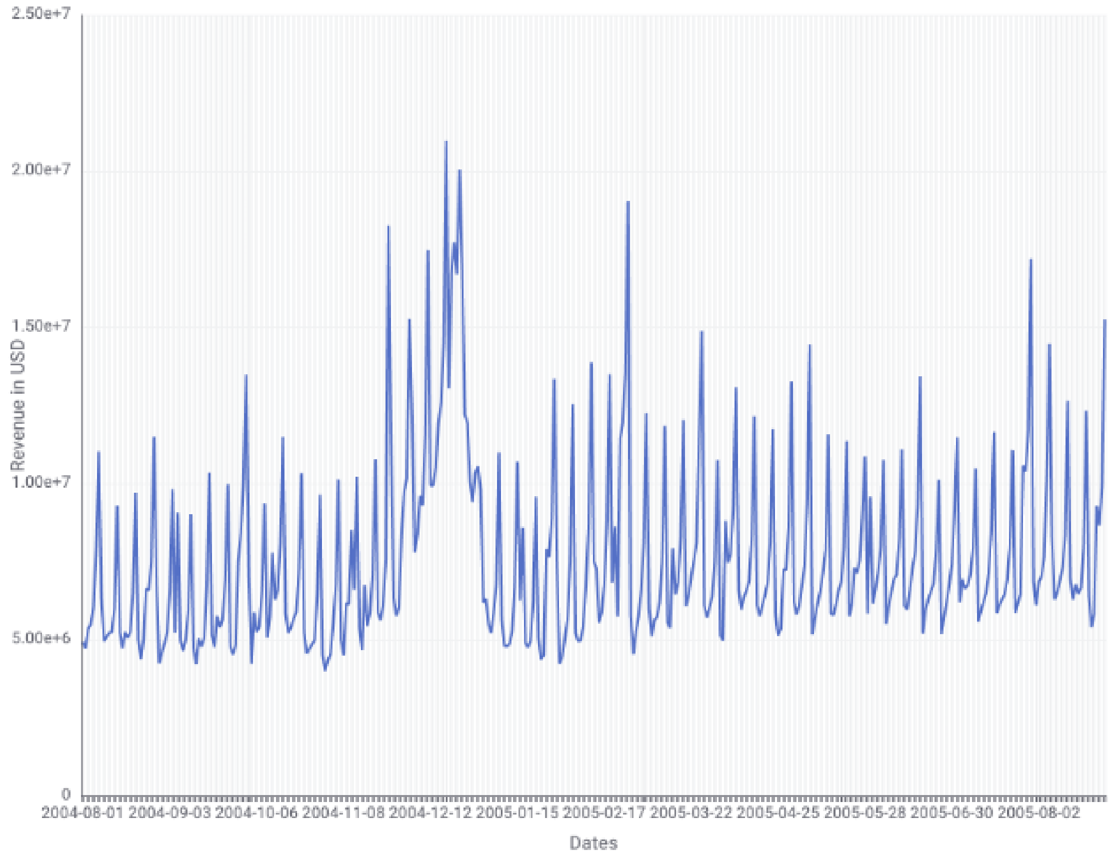

Figure 1 Seasonality of sales

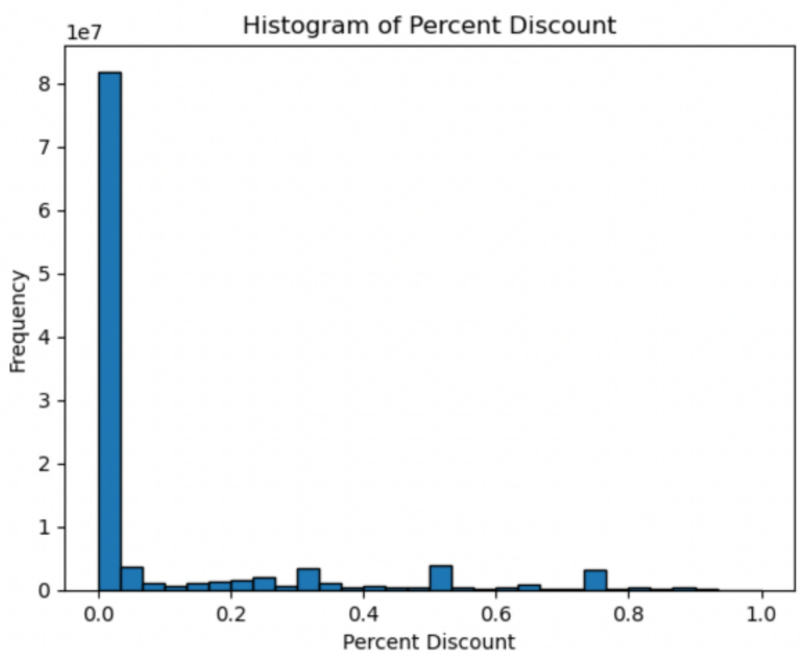

Figure 2 Discount distribution

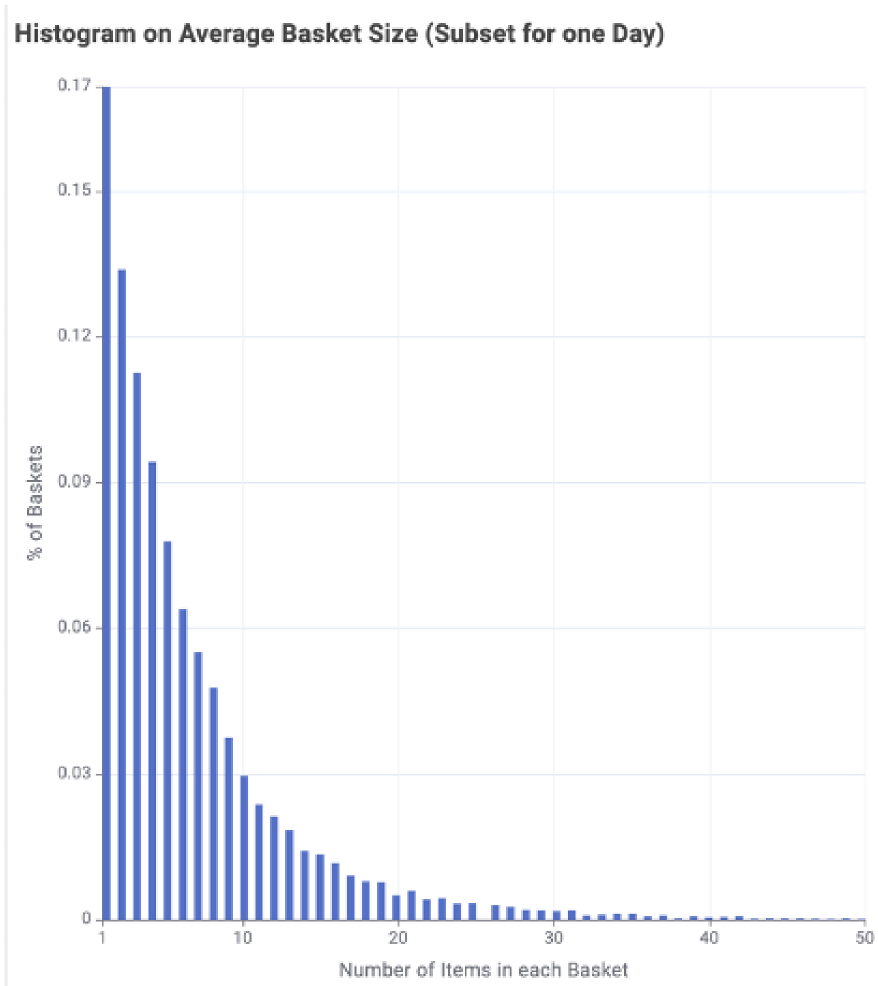

Figure 3 Basket size distribution

	storeid	state	zip	population	median_earning
0	2	FL	33710	34462	41198
1	3	MO	63126	15362	55585
2	4	AR	72201	875	63063
3	7	TX	76137	59273	42936
4	9	AZ	85281	66878	30311
..	...	...	...	...	...
448	9808	AZ	85233	37651	53126
449	9812	LA	70006	15455	35704
450	9900	AR	72201	875	63063
451	9906	AR	72201	875	63063
452	9909	WY	82009	34509	53831

Figure 4 Population and median income of each zipcode