Forecasting and Simulation of Market Share for a CPG Company

by

Shreya Nair

and

Boyue Adam Kang

Submitted on May 8, 2026, in Partial Fulfillment of the

Requirements for the Degree of Master of Applied Science in Supply Chain Management

at the Massachusetts Institute of Technology

ABSTRACT

This capstone develops a data-driven forecasting and simulation framework to improve market share prediction for a leading consumer packaged goods (CPG) company in the U.S. cookies category. The current approach relies on historical trends and internal assumptions, limiting its ability to capture competitive dynamics and evolving consumer behavior. To address this gap, the project integrates SKU-level sales data, demographic information, macroeconomic indicators, and consumer search trends to identify key drivers of market share. A causal modeling framework is applied to estimate the impact of pricing, promotions, and distribution while accounting for endogeneity, using Double Machine Learning (DML) and Causal Forest models to capture both average and heterogeneous effects. The estimated relationships are translated into a scenario-based simulation tool that enables decision-makers to evaluate alternative strategies and assess their projected impact on future market share, supporting more informed and forward-looking business decisions.

Capstone Advisor: Thomas Koch

Title: Postdoctoral Associate (MIT CTL)

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Motivation

The sponsor company is a global leader in the confectionery and packaged food sector, offering products such as biscuits, chocolate, candy, and gum, while expanding into several health-focused categories. Headquartered in the United States, the company operates a diversified business structure with manufacturing and commercial activities distributed across multiple regions worldwide. In recent years, it has pursued strategic growth through portfolio modernization, digital transformation of its supply chain, and expansion into fast-growing markets in Asia-Pacific and Latin America.

Looking ahead, the company anticipates continued category expansion driven by evolving consumer preferences and increasing demand for convenient, impulse-driven food products. To support strategic planning, the company develops annual business plans based on projections of category growth and expected market share performance. These planning processes require an understanding of how both internal actions and external market conditions influence market share outcomes.

Understanding the drivers of market share is a critical component of strategic decision-making in the consumer packaged goods (CPG) industry. For the sponsor company, insights into how pricing, promotions, and product strategies affect market share are essential for optimizing commercial performance.

The company's current planning approach relies primarily on historical sales data and internally generated assumptions, such as expected seasonality patterns, planned marketing activities, and baseline category growth. While this approach provides a structured foundation, it does not fully incorporate external competitive dynamics or evolving consumer behavior.

In practice, market share outcomes are influenced by a complex set of interrelated factors. Competitors' actions such as temporary price reductions or intensified promotions can shift demand across brands, while changes in demographics, lifestyle preferences, and economic conditions can alter consumer purchasing behavior at the category level. As a result, relying solely on historical trends may lead to incomplete or biased assessments of market dynamics.

These limitations highlight the need for a more efficient analytical framework that can isolate the impact of key business actions while accounting for external demand conditions. This project addresses this need by developing a causal modeling and simulation framework that links market share outcomes to measurable drivers. By estimating the impact of pricing, promotions, retail environment mix, and macroeconomic factors, the framework enables scenario-based analysis that supports more informed and responsive decision-making.

1.2 Problem Statement and Key Questions

The sponsor company seeks to improve its ability to understand the drivers of market share dynamics in the U.S. cookies category and to support more informed commercial decision-making. This project develops a data-driven modeling framework that links observed changes in market share to measurable drivers, including pricing, promotions, retail environment mix and macroeconomic conditions.

A key challenge in this setting is that many business decisions such as pricing and promotional intensity—are not made independently, but are influenced by underlying demand conditions and competitive dynamics. These factors act as confounding influences, meaning they affect both the company's actions (e.g., lowering price or increasing promotions) and the resulting market share outcomes. For example, firms may increase promotions during periods of weakening demand or heightened competition, making it difficult to distinguish whether changes in market share are driven by the promotion itself or by the underlying market conditions. As a result, simple correlation-based analyses may produce biased estimates of the true impact of these actions. To address this, the project adopts a causal modeling approach that estimates the effect of company-controlled actions while explicitly accounting for external factors and confounding influences. The final output of the project is a simulation tool that enables the sponsor company to evaluate "what-if" scenarios and assess the expected impact of alternative strategic decisions on market share.

To guide the development of the modeling framework and ensure alignment with the sponsor company's decision-making needs, the following key questions are considered:

1. Which observable factors—such as pricing, promotions, retail environment mix and macroeconomic variables—most strongly drive variation in the sponsor company's market share in the U.S. cookies category?

2. What is the causal impact of company-controlled actions, particularly pricing and promotions, on market share, and how do these effects vary across different market conditions and time periods?

3. How can causal machine learning methods, such as Double Machine Learning and Causal Forests, be used to estimate these effects in a high-dimensional observational setting?

4. How can the estimated causal relationships be translated into a simulation tool that supports scenario analysis and strategic decision-making?

1.3 Project Goals and Expected Outcomes

The goal of this project is to provide the sponsor company with a data-driven causal modeling and decision-support framework that enhances its ability to understand and quantify the drivers of category-level market share.

The project delivers two primary outcomes. First, it develops a causal estimation framework that quantifies the impact of key commercial levers such as pricing and promotions while accounting for external demand conditions, including retail environment dynamics and macroeconomic factors. By explicitly addressing confounding influences, this approach enables more credible identification of the true drivers of market share beyond simple correlations. Second, the project translates these estimated relationships into an interactive simulation tool that allows business teams to evaluate "what-if" scenarios and assess the expected impact of alternative strategies on market share outcomes.

Once validated, the framework can be integrated into the company's category planning and strategic decision-making processes and extended to additional product categories and markets. By linking market outcomes to measurable and controllable drivers, this approach supports more effective resource allocation, improves responsiveness to changing market conditions, and enables more informed, data-driven decision-making.

2. STATE OF THE PRACTICE

The objective of this section is to review the current state of practice in modeling market share dynamics in consumer packaged goods (CPG) categories. Prior research highlights that market share is driven by a complex set of interrelated factors and interactions, rather than isolated individual variables. In addition, recent advances in causal inference and machine learning have provided more rigorous approaches for estimating these relationships in the presence of endogeneity and complex data structures. This section synthesizes insights from three areas: market share drivers, consumer behavior, and modern causal modeling approaches.

2.1 Market Share Drivers

Understanding the drivers of market share in consumer packaged goods (CPG) categories has been a longstanding focus in retail analytics. Prior research shows that market share is shaped by multiple factors, including pricing, promotion intensity, product innovation, distribution breadth, retail environment formats, macroeconomic conditions, and seasonality. The analytical challenge arises from the strong multicollinearity among these variables, the simultaneous relationship between price and demand, and the endogeneity associated with promotional activities. In this context, endogeneity refers to situations in which promotional decisions are correlated with unobserved demand shocks, leading to biased and inconsistent estimates of causal effects. These challenges highlight the need for more advanced modeling approaches to identify true causal drivers rather than mere correlations (Hanssens et al., 2001).

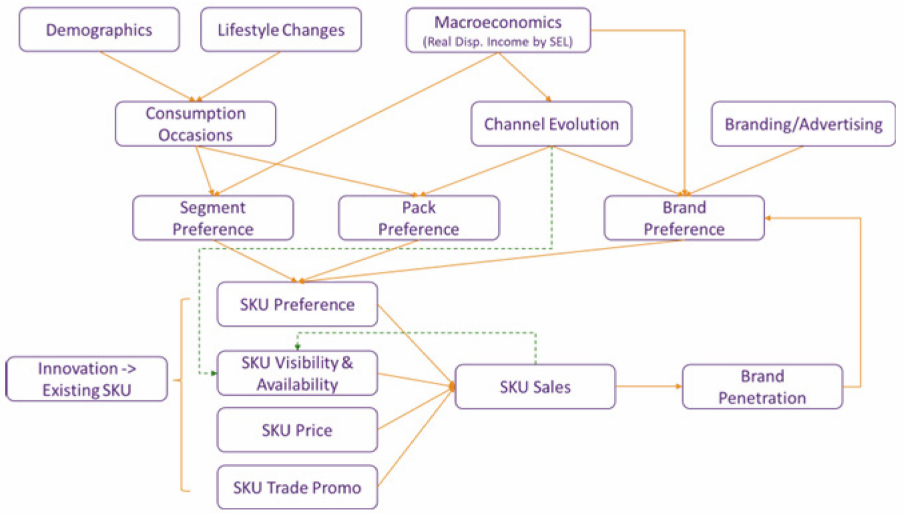

Figure 1. Sponsor's Company Framework for Market-Share Driver Analysis (Sponsor Company, 2025)

Figure 1 presents the sponsor company's internal view of the key drivers of market share, illustrating how macroeconomic conditions, consumer behavior, and commercial levers jointly influence category outcomes. In this study, the framework is used to inform variable selection and to ensure that the modeling approach reflects the interconnected nature of these drivers.

Market share has traditionally been modeled as a function of price, promotions, distribution breadth, product attributes, and competitive activity. Foundational studies such as Cooper and Nakanishi (1988) introduced the Multiplicative Competitive Interaction (MCI) model to explain how share responds to marketing inputs, while Tellis (1988) showed that promotions can generate significant short-term gains but have limited persistent effects on long-term share.

Building on this literature, several consistent insights emerge. First, seasonality and macroeconomic conditions play a significant role in shaping category dynamics, with market share often moving cyclically with income levels, unemployment, and commodity inflation (Dekimpe & Hanssens, 1995). Second, pricing and promotions are inherently endogenous, as firms and retailers adjust promotional intensity based on expected demand, making price non-exogenous (Nijs et al., 2001). Third, shifts in size tiers and price tiers often reflect consumer responses to economic conditions, with consumers trading down to smaller pack sizes or lower price tiers during periods of economic stress (Silverstein & Fiske, 2003). Finally, retail environments exhibit distinct behavioral patterns; for example, discount retailers tend to have higher price elasticity, while drug and pharmacy channels skew toward higher-tier, convenience-driven purchases (Bell & Lattin, 1998).

Overall, the literature suggests that accurate market share modeling must account for multiple interrelated drivers and their interactions. Simple regression approaches are often insufficient, particularly in the presence of endogeneity and multicollinearity. These findings motivate the use of more advanced causal modeling frameworks to better capture the underlying drivers of market share dynamics.

2.2 Consumer Behavior

Consumer behavior plays an important role in shaping category demand and market share. Extensive research in marketing science and behavioral economics shows that household demographics, consumption occasions, preferences, and psychological drivers all interact with marketing activities to shape purchase decisions.

Demographic factors exert strong influence. Income levels and household economic conditions affect price sensitivity, with lower-income households exhibiting higher price elasticity and stronger responses to promotions (Blattberg & Neslin, 1990). Age and life-stage dynamics also shape demand patterns; younger consumers, for example, tend to prefer novelty, variety, and innovation, leading to higher adoption rates for new SKUs (Rogers, 2003).

Behavioral economics further highlights systematic differences in consumer responses. Reference price theory shows that consumers form expectations about "fair prices," affecting their reactions to discounts and price increases (Kalyanaram & Winer, 1995). Prospect theory suggests that consumers are more sensitive to losses than gains, which amplifies the impact of price hikes during periods of economic stress (Kahneman & Tversky, 1979). These behavioral effects interact with marketing levers and can shift category-level share outcomes.

Consumption occasions also play a significant role. Research shows that many CPG categories exhibit strong seasonal or event-driven patterns, such as holidays, back-to-school periods, and weather-related shifts (Hoch et al. 1995). These changes affect both category volume and the relative share of brands positioned around specific occasions.

Consumer preferences regarding health, indulgence, sustainability, or dietary trends likewise influence category share. Studies show that changes in nutritional attitudes or wellness trends can reallocate share across segments or size tiers (Chandon & Wansink, 2007).

Information-search behavior has also become an important indicator of consumer demand. Google Trends and other digital search measures have been found to be predictive of short-term demand fluctuations and product adoption, especially for innovation and seasonal products (Choi & Varian, 2012).

These findings indicate that market share dynamics cannot be fully understood without explicitly accounting for underlying consumer behavior. In particular, variations in demographic composition shape differences in price sensitivity, promotion responsiveness, and preferences for innovation. In addition, seasonal patterns and consumption occasions introduce predictable but non-linear fluctuations in category demand. Moreover, measures of innovation-related interest, such as novelty-seeking tendencies and responsiveness to new product launches, help explain heterogeneity in "trial and repeat" behavior across consumer segments. Finally, search-based indicators such as Google Trends capture shifts in consumer attention and consideration that often precede changes in purchase behavior. Taken together, incorporating these behavioral signals into a causal modeling framework enables a more accurate representation of demand formation and improves the ability to detect emerging shifts in category-level share.

2.3 Causal Inference and Causal Prediction in Marketing

Identifying the causal drivers of sales and market share has been a longstanding challenge in marketing science due to the strategic and interdependent nature of firm and consumer behavior. Unlike traditional predictive models, causal inference focuses on isolating the true effect of marketing actions from confounding forces, enabling researchers to distinguish genuine causal mechanisms from spurious correlations (Hanssens et al., 2001). This distinction is critical in retail settings where prices, promotions, and distribution decisions are often endogenous, meaning that they are influenced by expected demand and competitive conditions. For example, firms may increase promotions in periods when they anticipate weaker demand or stronger competitive pressure. As a result, these marketing actions are correlated with underlying demand conditions, making it difficult to isolate their true causal impact on market share.

Addressing these endogeneity challenges requires a clear articulation of underlying causal assumptions and a disciplined approach to empirical estimation. Recent advances in causal inference provide structured frameworks for both identifying valid adjustment strategies and translating these assumptions into implementable analytical procedures. The following discussion outlines how these frameworks support more reliable causal estimation in marketing settings.

2.4 Machine Learning Approaches for Causal Prediction

While traditional econometric models remain foundational in marketing science, recent advances in machine learning have expanded the methodological toolkit for causal analysis in high-dimensional retail settings. Capturing nonlinear interactions and heterogeneous effects across multiple drivers often exceeds the capacity of standard linear models.

Double Machine Learning (DML) provides a structured approach to causal estimation in settings with many covariates (Chernozhukov et al., 2018). In this context, the treatment refers to company-controlled actions such as pricing, promotions, or distribution changes, while the outcome of interest is category-level market share. By leveraging flexible machine learning models to control for confounding factors, DML isolates the causal effect of these actions from underlying demand conditions. This approach reduces estimation bias while preserving interpretability, making it well suited for market share modeling where correlated predictors and endogeneity are common.

Tree-based methods such as Causal Forests (Athey & Wager, 2019) extend this framework by estimating heterogeneous treatment effects. A Causal Forest is a machine learning method that builds on the Random Forest algorithm to estimate how the effect of a treatment varies across observations. Rather than producing a single average effect, it partitions the data into groups with similar characteristics and estimates treatment effects within each group, which are then aggregated to produce stable and data-driven estimates. This allows the model to capture variation in the impact of pricing, promotions, or distribution across different market conditions, such as seasonality, channel mix, and product characteristics.

3. METHODOLOGY

In this section, we outline the methodology adopted in this project, the datasets used, and how these components come together to address our core objective: to estimate and predict category-level market share on a monthly basis using a causal framework that integrates key commercial drivers, including pricing, promotions, and distribution, alongside broader market conditions.

3.1 Methodology Selection

Based on the methods discussed in Section 2, we adopt a causal inference framework to quantify the impact of key commercial levers (price, promotion, and distribution) on market share. Traditional predictive models, such as standard regression or machine learning approaches, are primarily designed to capture correlations in the data and optimize forecast accuracy. However, they do not explicitly account for these factors or distinguish between correlation and causation, which limits their usefulness for decision-making in a business context. Since the objective of this analysis is to evaluate what-if scenarios and simulate the effect of strategic interventions, a causal methodology is required.

To address this, we employ Double Machine Learning (DML) approaches, including Linear DML and Causal Forest DML. These methods are grounded in the Double Machine Learning framework, which separates the estimation of nuisance components (such as the relationship between covariates and treatment or outcome) from the estimation of the causal effect. By leveraging flexible machine learning models (e.g., Random Forests) to control for high-dimensional confounders, DML reduces bias and enables more reliable estimation of treatment effects under the assumption of conditional independence.

Among the models considered, Linear DML provides a globally interpretable estimate of average treatment effects, offering a clear understanding of the overall direction and magnitude of each lever. However, it assumes a linear and homogeneous relationship across all observations. In contrast, Causal Forest DML relaxes these assumptions by allowing for nonlinearity and heterogeneity in treatment effects. It constructs an ensemble of decision trees that partitions the data into subgroups and estimates localized causal effects, thereby capturing variation in responsiveness across different market conditions, channels, or product segments.

The use of both models enables a balanced approach: Linear DML offers interpretability and stability, while Causal Forest DML captures more complex, real-world dynamics. By comparing results across these methods, we improve robustness and gain deeper insights into how different levers interact with underlying market characteristics. This combined approach ensures that the resulting simulation framework is both theoretically grounded and practically relevant for scenario analysis and strategic decision-making.

3.2 Data Collection and Preparation

The project focuses on the sponsor company's sales in the U.S., specifically within the cookies category. The sponsor company relies on data and analytics providers such as Nielsen to measure category performance across competitors and to analyze consumer behavior by demographic segments.

We analyze two primary datasets. The first dataset provided by the sponsor company contains SKU-level sales by volume and value for both the sponsor company and its competitors from January 2023 to December 2025. The second dataset comes from the Federal Reserve Economic Data (FRED), which is maintained by the Federal Reserve Bank of St. Louis.

In addition to the Nielsen retail data, this study incorporates macroeconomic indicators from the FRED database to capture broader economic conditions that may influence consumer demand. These variables are included as control factors to account for external demand shifts that are not directly driven by pricing, promotion, or distribution decisions. By controlling for these broader economic conditions, the model aims to isolate the causal impact of company-controlled levers on market share and reduce bias from underlying economic fluctuations.

These macroeconomic variables are included as control variables to account for exogenous demand shifts that are not directly driven by pricing, promotion, or distribution decisions. For instance, during periods of high unemployment or rising inflation, consumers may become more price-sensitive and reduce discretionary spending, while changes in industrial production may reflect broader economic cycles affecting demand. By incorporating these macroeconomic signals into the model, we aim to isolate the true causal impact of commercial levers from broader economic fluctuations.

The FRED data is aggregated at a monthly level and aligned temporally with the Nielsen dataset to ensure consistency in model inputs. This integration enhances the reliability of the causal analysis by controlling for external factors that could otherwise bias treatment effect estimates.

Dataset 1: Key Variables in the U.S. Cookies Market Dataset

This dataset, sourced from Nielsen retail measurement data, provides SKU-level information on sales value and volume for both the sponsor company and its competitors within the U.S. cookies category. It serves as the primary dataset for constructing market share and forms the empirical foundation of the analysis.

Relevant features of the dataset:

Variable Name	**Description**
global_manufacturer	The parent company or corporate group that owns and produces the brand.
global_sub_segment	A finer product classification within a category based on functional or product-type characteristics.
global_category	The broad product category to which the item belongs.
global_segment	A higher-level grouping of related sub-segments within a category, often reflecting consumer usage or need states.
global_brand	The commercial brand name under which the product is marketed.
retail_environment	The type of retail channel or store format where the product is sold (e.g., grocery, mass merchandiser, convenience).
Calmonth	The calendar month of observation, representing the time period of the sales record.
innovation_type	A categorical variable indicating the type of product innovation.
price_tier_bucket	A discrete classification of products into price tiers (e.g., value, mid-tier, premium) based on relative price levels.
size_tier	A categorical grouping of products based on package size.
price_corridor_band	A pricing band that places products within a competitive price corridor relative to similar items in the market.
Vol	Total unit sales volume during the given period, typically measured in physical units or standardized volume.
Valusd	Total sales value in U.S. dollars generated during the given period.
Promoanyvalusd	Total promotional sales value in U.S. dollars, capturing revenue generated from items sold under promotion.
Promoanyvol	Total promotional sales volume, capturing the number of units sold under promotional conditions.

Table 1. Key Variables in the U.S. Cookies Market Dataset

These variables shown above capture key aspects of pricing, sales performance, and promotional activity, providing the core inputs for the subsequent analysis of market share dynamics.

Dataset 2: Federal Reserve Economic Data (FRED)

This dataset incorporates macroeconomic indicators from the Federal Reserve Economic Data (FRED) to capture external economic conditions that may influence consumer demand. Specifically, the unemployment rate reflects labor market health, the Consumer Price Index (CPI) captures inflation and changes in purchasing power, and industrial production serves as a proxy for overall economic activity. These variables are included as control variables to account for broader economic fluctuations, ensuring that the estimated effects of pricing, promotion, and distribution are not confounded by macroeconomic trends.

The key macroeconomic variables used in this study are summarized in Table 2.

Variable Name	Description
Unemployment Rate	% of unemployed workforce; proxy for economic conditions
Consumer Price Index	Measures inflation; reflects price level changes
Industrial Production	Indicates economic activity and production levels

Table 2: Federal Reserve Economic Data (FRED)

3.3 Application of Methodology

Our analytical process follows a four-step structured framework to estimate the causal impact of pricing, promotions, and distribution on category-level market share, while controlling for demand-side factors and macroeconomic conditions.

Step 1: Data Integration and Feature Construction

We begin by integrating multiple datasets into a unified analytical dataset. The primary dataset consists of SKU-level Nielsen retail data, including market share, pricing, promotion intensity, and distribution metrics. This is complemented by external macroeconomic indicators from the Federal Reserve Economic Data (FRED), including the unemployment rate, Consumer Price Index (CPI), and industrial production, which capture broader economic conditions.

All datasets are aligned to a monthly time frequency. We then construct additional features to improve model performance and capture underlying behavioral dynamics. These include lagged market share variables to capture persistence effects; interaction terms between pricing, promotion, and distribution to model joint effects; and seasonality controls implemented using sinusoidal transformations.

These transformations ensure that the model captures both temporal dynamics and non-linear interactions while avoiding data leakage.

Step 2: Causal Identification Strategy

To estimate causal effects, we adopt a structural approach based on the potential outcomes framework. In this context, pricing, promotion intensity, and distribution are treated as continuous treatment variables, as they represent company-controlled levers that can influence market share, while market share serves as the outcome variable.

We assume that, conditional on observed covariates, the variation in these treatment variables is independent of unobserved factors affecting the outcome. The covariates include lagged performance, category characteristics, and macroeconomic conditions, which help capture underlying demand trends and external influences. Under this assumption, the model can identify causal effects by controlling for factors that simultaneously influence both the treatments and market share, such as past demand patterns, seasonality, and broader economic conditions.

By explicitly incorporating these controls, the model aims to isolate the impact of commercial decisions from underlying demand fluctuations and historical momentum effects.

Step 3: Causal Modeling Approaches

We estimate treatment effects using Double Machine Learning (DML), which enables reliable causal inference in high-dimensional settings. In this framework, machine learning models are first used to capture the relationships between observable factors, business actions such as pricing and promotions, and market share. The remaining variation, after accounting for these factors, is then used to estimate the causal impact of each action.

Specifically, we implement:

- Linear DML to capture average treatment effects under a parametric structure
- Causal Forest DML to model non-linear and heterogeneous treatment effects

Both models rely on flexible machine learning methods, such as Random Forests, to model underlying patterns in the data and account for factors that influence both business decisions and market share. This approach helps reduce bias and improves the credibility of the estimated causal effects.

This approach allows us to:

- Correct for endogeneity arising from correlated predictors
- Capture complex, non-linear relationships
- Estimate heterogeneous effects across different market conditions

Step 4: Scenario Simulation

To translate model outputs into actionable insights, we build a scenario simulation framework. Using the estimated causal models, we simulate counterfactual scenarios by applying controlled changes to key decision variables, including pricing, promotion, and distribution.

The model computes the resulting change in market share by comparing baseline and counterfactual outcomes. A logit transformation is applied to ensure predicted market share remains within valid bounds.

This simulation enables decision-makers to evaluate the expected impact of alternative commercial strategies in a controlled and interpretable manner.

4. RESULTS AND DISCUSSIONS

This section presents the estimated causal effects of key commercial levers on category-level market share and interprets their implications for decision-making. Results are derived from two complementary models, Linear DML and Causal Forest DML, and are further validated through scenario-based simulation.

4.1 Results

The analysis evaluates the impact of pricing, promotion, and distribution under five representative scenarios. Specifically, Price Cut refers to a reduction in product price while holding other factors constant; Promo Boost captures an increase in promotional intensity; Distribution Boost reflects an expansion in product availability or retail coverage; Full Mix combines simultaneous changes in price, promotion, and distribution; and Aggressive strategy represents a coordinated, large-scale adjustment across all three levers. The estimated treatment effects show clear directional patterns across models, although the magnitude and consistency vary.

The Linear DML model produces positive and economically meaningful effects across most scenarios. Price reductions lead to a substantial increase in market share, indicating strong price sensitivity within the category. Promotional increases also generate positive but smaller gains, suggesting that promotions are effective but less impactful than price changes. In contrast, distribution improvements show negligible effects, implying that incremental expansion in distribution may not translate directly into short-term share gains. When combined in the Full Mix and Aggressive scenarios, the effects are amplified, with the Aggressive strategy yielding the largest estimated increase in market share.

In contrast, the Causal Forest model produces more conservative, and in some cases negative, estimates, particularly for the Price Cut and Aggressive scenarios. While promotional increases continue to show small positive effects, most other scenarios result in minimal or slightly negative impacts. This divergence reflects the model's ability to capture heterogeneity and non-linear relationships, suggesting that the effectiveness of these levers may vary across market conditions and segments. These patterns are further illustrated in Figure 2.

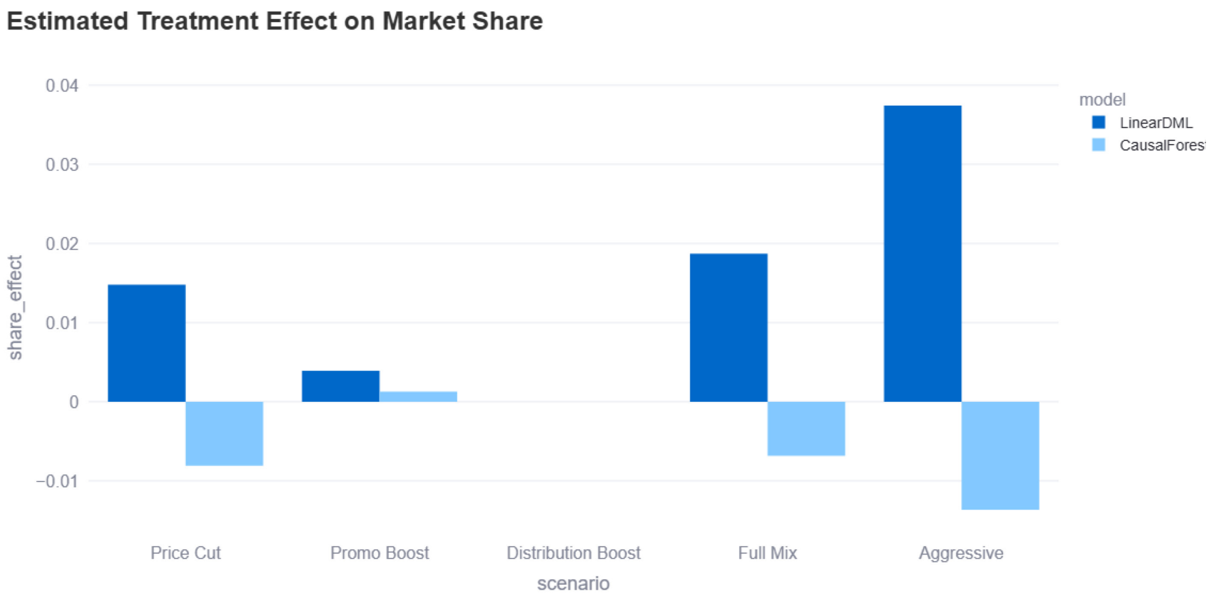

Figure 2: The Estimated Treatment Effects across Different Scenarios for Both Linear DML and Causal Forest Models

4.2 Findings and Implications

Several key insights emerge from the results.

First, pricing is the most influential lever driving market share. The consistent positive impact observed in the Linear DML model indicates that lowering relative price can significantly increase competitiveness. However, the weaker and sometimes negative effects in the Causal Forest model suggest that aggressive price reductions may not universally translate into gains and could depend on context such as competitive response or consumer perception.

Second, promotions have a positive but moderate impact. Across both models, promotional increases consistently generate small improvements in market share, indicating that promotions are a reliable but secondary growth lever compared to pricing. This suggests that promotions should be used to complement, rather than substitute for, pricing strategies.

Third, distribution expansion shows limited marginal impact in the current setting. The near-zero effects across both models imply that the category may already be close to distribution saturation, or that incremental distribution gains are less effective without supporting pricing or promotional strategies.

Fourth, combined strategies deliver stronger outcomes but also greater uncertainty. The Full Mix and Aggressive scenarios produce the largest gains under the Linear DML model, highlighting the potential benefits of coordinated commercial actions. However, the Causal Forest results indicate that such strategies may not always perform consistently, reflecting underlying heterogeneity in market response.

Finally, the scenario simulator demonstrates the practical value of the modeling framework. By translating estimated causal effects into predicted market share changes, the simulator allows decision-makers to evaluate alternative strategies in a forward-looking and interpretable way. This bridges the gap between statistical estimation and business application. Simulator results are represented in Figure 3 below:

	scenario	model	share_effect	new_share
0	Price Cut	LinearDML	0.0148	0.4132
2	Promo Boost	LinearDML	0.0039	0.4105
4	Distribution Boost	LinearDML	-0.000005	0.4096
6	Full Mix	LinearDML	0.0187	0.4141
8	Aggressive	LinearDML	0.0374	0.4187
1	Price Cut	CausalForest	-0.0081	0.4076
3	Promo Boost	CausalForest	0.0013	0.4099
5	Distribution Boost	CausalForest	-0.00001	0.4096
7	Full Mix	CausalForest	-0.0068	0.4079
9	Aggressive	CausalForest	-0.0137	0.4063

Figure 3. Causal Market Share Decision Simulator used to evaluate counterfactual pricing, promotion, and distribution strategies based on estimated treatment effects

Overall, the results suggest that while pricing remains the primary driver of market share, its effectiveness is context-dependent. Promotions provide stable but incremental gains, and distribution plays a more limited role. A balanced strategy that considers both magnitude and uncertainty of effects is therefore critical for optimizing performance.

4.3 Limitations

Despite these insights, several limitations should be acknowledged. First, the analysis relies on observational data and assumes that all relevant factors influencing both business decisions and market outcomes are adequately accounted for. In practice, unobserved demand shocks or competitive actions may still influence both pricing and promotions, potentially biasing the estimated effects. As a result, the findings should be interpreted as directional and indicative rather than strictly causal in a fully controlled sense.

Second, the divergence between Linear DML and Causal Forest estimates highlights model sensitivity to underlying assumptions. While Linear DML provides stable and interpretable average effects, Causal Forest captures heterogeneity and context-specific variation, sometimes producing more conservative or negative estimates. This suggests that the true impact of commercial levers is unlikely to be uniform and should be considered within a range of possible outcomes rather than a single point estimate.

Third, the relatively aggregated nature of the data and the limited time horizon may constrain the ability to detect finer-grained patterns across segments, channels, or product tiers. Important sources of heterogeneity, such as retailer-specific dynamics or micro-level consumer behavior, may therefore be partially obscured.

Finally, the simulation framework assumes static competitive behavior and does not explicitly model dynamic responses from competitors. In reality, pricing or promotional actions may trigger competitive reactions that amplify or offset the estimated effects. Future work could address this by incorporating competitive response models or scenario-based stress testing.

5. RECOMMENDATIONS

Building on the causal analysis and scenario simulation results, this section translates key findings into actionable guidance for the sponsor company. The recommendations focus on how pricing, promotions, and distribution strategies can be optimized in a way that balances expected impact with uncertainty. Beyond immediate tactical decisions, these insights also highlight opportunities to improve planning processes, resource allocation, and the use of data-driven decision tools in commercial strategy.

5.1 Management Recommendations

The following recommendations translate the empirical findings into actionable guidance for the sponsor company's commercial strategy. They are derived from the estimated causal effects and scenario simulations, which highlight the relative importance, variability, and interactions of pricing, promotions, and distribution. Rather than prescribing one-size-fits-all actions, these recommendations emphasize context-aware decision-making and the need to balance expected impact with uncertainty. Together, they provide a structured approach to improving market share performance while supporting more informed and data-driven strategic planning.

- **Adopt a Targeted and Context-Aware Pricing Strategy**

Pricing emerges as the most influential lever, but also the most variable across contexts. While price reductions can significantly increase market share, the differences between model estimates indicate that their effectiveness depends on underlying market conditions such as competitive intensity and demand sensitivity. As a result, price changes should not be applied uniformly. Instead, the company should adopt a more targeted approach by tailoring pricing actions to specific market conditions and competitive environments. Broad, aggressive price cuts may increase share in some cases but risk eroding margins without guaranteed returns in others. Leveraging the simulation tool to test alternative pricing scenarios before implementation can help balance upside potential with risk and improve decision quality.

- **Use Promotions as a Reliable Short-Term Growth Lever**

Promotions consistently generate positive but moderate gains in market share, indicating that they are a stable mechanism for driving incremental demand. However, their impact is smaller than that of pricing and appears less sensitive to context. This suggests that promotions are most effective when used strategically to support pricing actions rather than as a primary growth driver. In practice, this means focusing on targeted promotional campaigns aligned with short-term sales objectives, while prioritizing efficiency and return on investment over scale. Treating promotions as a complementary lever can help sustain consistent performance without over-reliance on discounting.

- **Reassess Incremental Investment in Distribution Expansion**

The results indicate that distribution expansion has a limited marginal impact in the current setting, suggesting that the category may already be approaching saturation. This implies that additional investment in broad distribution growth may yield diminishing returns. Instead, distribution strategies should be more selective and focused on underpenetrated channels, regions, or retailer formats where incremental gains are still achievable. At the same time, reallocating resources toward pricing and promotional optimization may generate higher returns. A more disciplined, ROI-driven approach to distribution investment can improve overall resource efficiency.

- **Prioritize Integrated Commercial Strategies**

The analysis shows that coordinated changes across pricing, promotions, and distribution generate the largest potential gains, but also introduce greater variability and uncertainty. This highlights the importance of adopting integrated commercial strategies rather than optimizing each lever in isolation. By combining actions in a structured and data-informed way, the company can capture synergies while managing risk. The simulation framework developed in this project provides a practical tool to evaluate such combined strategies and supports more informed, forward-looking decision-making.

- **Institutionalize the Causal Decision Simulator**

The developed simulator represents a valuable strategic asset for the sponsor company. By translating causal estimates into a practical decision-support tool, it enables forward-looking evaluation of alternative commercial strategies and reduces reliance on purely judgment-based planning. To fully realize its value, the simulator should be integrated into regular planning and forecasting processes, allowing business teams to systematically evaluate “what-if” scenarios before committing resources. This can improve decision quality by making trade-offs between pricing, promotions, and distribution more transparent and data-driven. In addition, maintaining the relevance and accuracy of the simulator requires continuous updates with new data and periodic model refinement to reflect evolving market conditions. Over time, embedding the simulator into decision-making workflows can support a more adaptive and evidence-based approach to commercial strategy, strengthening the company’s ability to respond to market changes and optimize performance.

5.2 Future Work

While the current framework provides valuable insights into the causal drivers of market share, several extensions could further enhance both the robustness of the analysis and its applicability to real-world decision-making contexts.

One important direction is to incorporate competitive dynamics more explicitly into the modeling framework. The current approach treats pricing, promotion, and distribution decisions as independent of competitors’ strategic responses, whereas in practice these actions often trigger reactions such as price matching, promotional escalation, or shifts in assortment and visibility. Future work could extend the framework to capture such interactions, for example by introducing competitor response variables, lagged competitive indicators, or interaction terms that reflect strategic interdependence. More advanced approaches, such as game-theoretic modeling or multi-agent simulation, may also provide a richer representation of competitive behavior. Incorporating these dynamics would allow the model to better approximate real market conditions and improve the credibility of scenario-based predictions.

Another area for improvement is the ability to capture finer-grained heterogeneity in consumer and channel responses. The current analysis is conducted at an aggregated category level, which may mask meaningful variation across customer segments, retail formats, geographic regions, or product subcategories. Future extensions could leverage more granular data, such as segment-level or SKU-level observations, to estimate heterogeneous effects with greater precision. This would enable a more detailed understanding of how different groups respond to pricing and promotional changes, and could support more targeted and differentiated strategy design. In addition, incorporating interaction structures across segments or channels may reveal cross-effects that are not observable in an aggregated framework.

Expanding the set of external demand signals also represents a promising avenue for improving model performance. While the current framework incorporates key macroeconomic indicators, additional data sources could provide more timely and behaviorally relevant signals of demand fluctuations. For instance, consumer sentiment indices, digital engagement metrics such as search trends or online activity, and seasonal or weather-related variables may capture shifts in attention, preferences, and purchasing intent. Integrating these signals could enhance the model's ability to detect short-term changes in demand conditions and reduce reliance on lagged variables. This, in turn, may improve both the responsiveness and predictive accuracy of the estimated effects.

A further extension involves refining the temporal structure of the model. The current approach primarily captures contemporaneous relationships and limited lagged effects, but consumer response and competitive dynamics often unfold over time. Future work could incorporate richer dynamic specifications, such as distributed lag models, state-space representations, or time-varying parameter approaches, to better capture persistence, delayed effects, and structural changes in the market. This would allow for a more nuanced understanding of how the impact of pricing and promotions evolves over different time horizons.

Finally, additional development could focus on enhancing the simulator as both an analytical and operational tool. While the current version enables scenario-based evaluation, future iterations could incorporate more advanced functionality, such as automated scenario generation, optimization routines, or probabilistic outcome ranges that reflect uncertainty in model estimates. Integration with internal data systems and pipelines may also support more frequent updates and facilitate real-time or near-real-time analysis. Over time, these enhancements could enable the simulator to evolve into a continuously learning platform that adapts to changing market conditions and supports ongoing analytical refinement.

Taken together, these extensions would contribute to a more comprehensive and flexible modeling framework, better aligned with the complexity of real-world market dynamics. They may also strengthen the connection between causal inference and practical decision support, providing a foundation for future research and continued improvement of data-driven commercial strategy.

References

Ailawadi, K. L., Gedenk, K., Lutzky, C., & Neslin, S. A. (2006). Consumer response to promotions. Marketing Science, 25(6), 625–643.

Athey, S., & Wager, S. (2019). Estimating treatment effects with causal forests. Journal of the American Statistical Association, 114(526), 607–619.

Bell, D. R., & Lattin, J. M. (1998). Shopping behavior and consumer preference for store price format: Why “large basket” shoppers prefer EDLP. Marketing Science, 17(1), 66–88.

Blattberg, R. C., & Neslin, S. A. (1990). Sales promotion: Concepts, methods, and strategies. Prentice Hall.

Chandon, P., & Wansink, B. (2007). Is food marketing making us fat? A multi-disciplinary review. Journal of Marketing, 71(1), 1–25.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and causal parameters. American Economic Review, 108(8), 2018–2054.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. Economic Record, 88(s1), 2–9.

Cooper, L. G., & Nakanishi, M. (1988). Market share analysis: Evaluating competitive marketing effectiveness. Kluwer Academic Publishers.

Dekimpe, M. G., & Hanssens, D. M. (1995). The persistence of marketing effects on sales. Marketing Science, 14(1), 1–21.

Hanssens, D. M., Parsons, L. J., & Schultz, R. L. (2001). Market response models: Econometric and time series analysis. Kluwer Academic Publishers.

Hoch, S. J., Kim, B.-D., Montgomery, A. L., & Rossi, P. E. (1995). Determinants of store-level price elasticity. Journal of Marketing Research, 32(1), 17–29.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263–291.

Kalayanaram, G., & Winer, R. S. (1995). Empirical generalizations from reference price research. Marketing Science, 14(3), G161–G169.

Nijs, V. R., Dekimpe, M. G., Steenkamp, J.-B. E. M., & Hanssens, D. M. (2001). The category-demand effects of price promotions. Marketing Science, 20(1), 1–22.

Pearl, J. (2009). Causality: Models, reasoning, and inference (2nd ed.). Cambridge University Press.

Rogers, E. M. (2003). Diffusion of innovations (5th ed.). Free Press.

Sharma, A., & Kiciman, E. (2020). DoWhy: A Python package for causal inference. Journal of Machine Learning Research, 21(1), 1–6.

Silverstein, M. J., & Fiske, N. (2003). Trading up: The new American luxury. Harvard Business Review, 81(2), 41–53.