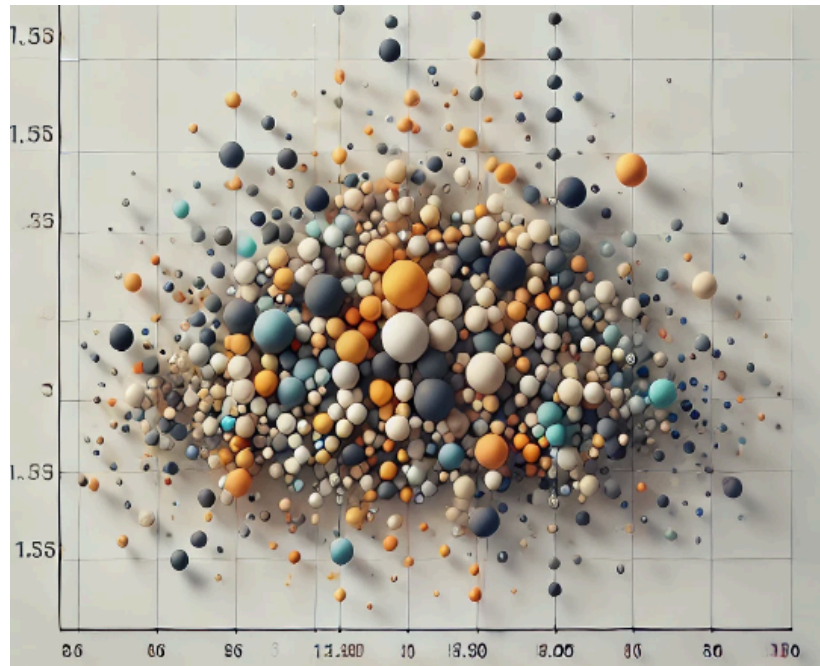


Sales Data Analysis



Created by

G NikhilChand
21f1003825

Table of Contents

1. Introduction
2. Objective
3. Data Overview
4. Methodology
5. Analysis & Findings
6. Recommendations
7. Conclusion
8. Future Work

Introduction

The project involves optimizing the "Gross Margin Percentage" in a retail dataset while ensuring that the overall profitability, as measured by "Gross Income," remains stable or improves. The dataset provided includes various transactional data points, and the goal is to identify key factors that influence gross margin and propose actionable strategies to enhance it.

This dataset is one of the historical sales of a supermarket company which has been recorded in 3 different branches for 3 months. Predictive data analytics methods are easy to apply with these dataset.

To better understand the factors influencing these variations and to identify potential areas for improvement, the company has decided to conduct a comprehensive analysis of its sales data. The dataset comprises transactional information, customer ID, and product-related metrics.



Objective

The primary objective is to develop a data-driven strategy to optimize the "Gross Margin Percentage" without compromising the total profitability. The project focuses on understanding the underlying patterns and key drivers that impact gross margin through comprehensive data analysis and advanced modeling techniques

The project's objective is to develop a strategy to optimize the "Gross Margin Percentage" within the retail dataset while ensuring that overall profitability, as reflected by "Gross Income," remains stable or improves. This involves identifying key factors influencing the gross margin and proposing actionable recommendations to enhance it without compromising total profit.



Data Overview

The dataset, "sales_data.csv," contains transaction-level data from a retail environment. The key columns include:

Data columns (total 17 columns):

#	Column	Non-Null Count	Dtype
0	Invoice ID	1000 non-null	object
1	Branch	1000 non-null	object
2	City	1000 non-null	object
3	Customer type	1000 non-null	object
4	Gender	1000 non-null	object
5	Product line	1000 non-null	object
6	Unit price	1000 non-null	float64
7	Quantity	1000 non-null	int64
8	Tax 5%	1000 non-null	float64
9	Total	1000 non-null	float64
10	Date	1000 non-null	object
11	Time	1000 non-null	object
12	Payment	1000 non-null	object
13	cogs	1000 non-null	float64
14	gross margin percentage	1000 non-null	float64
15	gross income	1000 non-null	float64
16	Rating	1000 non-null	float64

dtypes: float64(7), int64(1), object(9)

Methodology

- **Data Understanding & Cleaning:** The provided dataset contained transactional data across different product lines, with a consistent gross margin percentage of 4.761905% for all products.
- The goal was to identify opportunities to adjust the gross margin percentage for each product line to maximize overall profitability without reducing the total profit.
- **Feature Engineering:** The data was grouped by **Product line** to calculate the total sales revenue, total COGS, and total gross income for each category.

Gross Income=Total Sales Revenue–Total COGS

- **Overall Profit Comparison:** The total profit before and after the margin adjustments was compared to ensure the optimization led to an increase in overall profitability.
- **Outcome:** The adjusted gross margins resulted in an increase in overall profit from 198.84 to 209.59, indicating a successful optimization strategy.
- **Advanced Analysis:** Utilized SHAP analysis for feature importance, customer segmentation for personalized strategies, and predictive modeling to assess churn and CLTV.

Analysis & Findings

Data Preprocessing

The dataset was cleaned and preprocessed to ensure accurate analysis. Missing values were handled appropriately, and the data was transformed to a suitable format for modeling.

Handle Missing Data: Depending on the nature and amount of missing data, you can either fill in the gaps or exclude those rows/columns.

Outlier Treatment: Decide on a strategy to deal with outliers (e.g., capping, transformation).

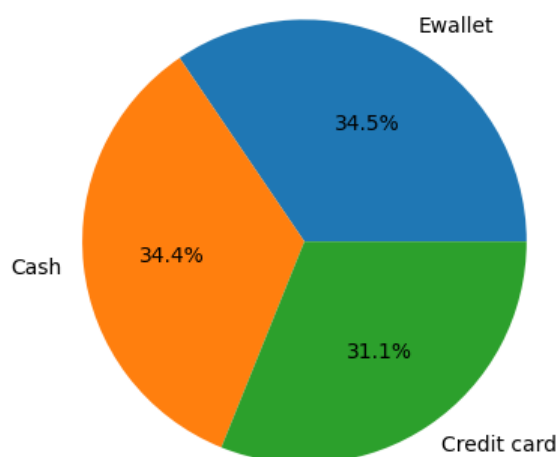
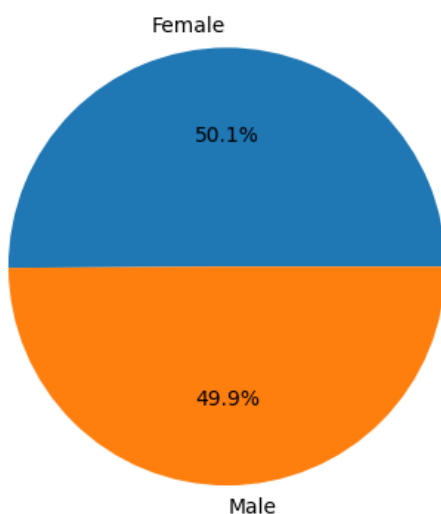
Feature Engineering: Create new features that might help in analysis, such as

Average Spending Per Visit, Profit Margin Per Product Line, etc.

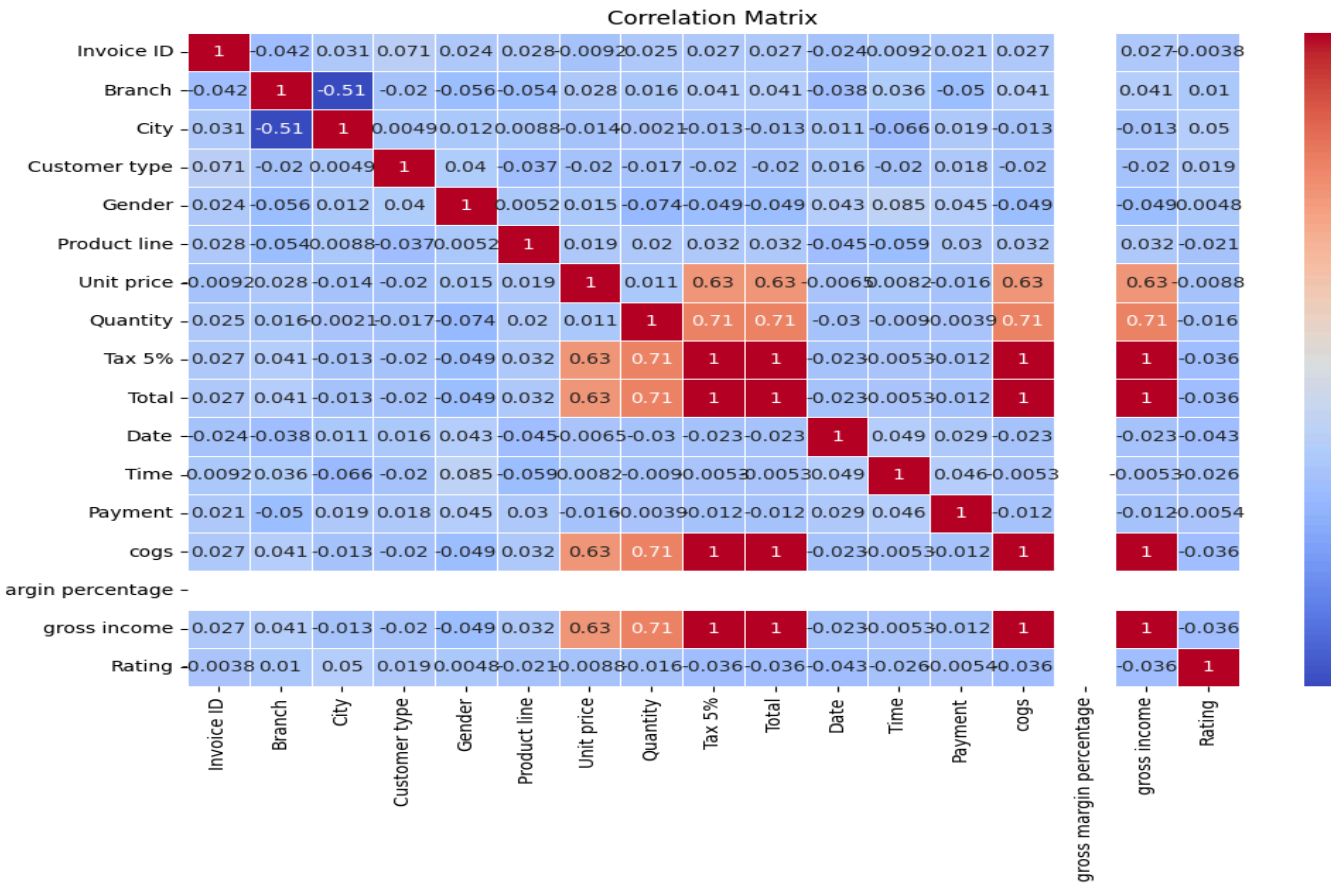
Data Transformation: Normalize or standardize data if necessary, particularly for columns that might have wide ranges.

Exploratory Data Analysis (EDA)

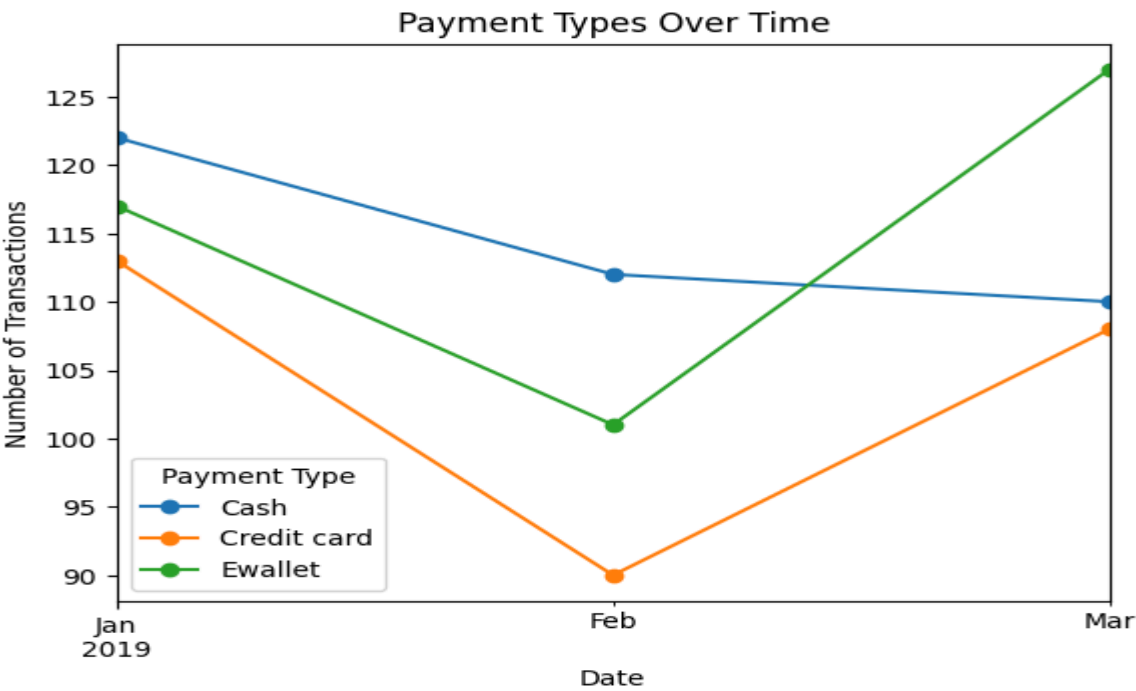
Exploratory Data Analysis (EDA): Conducted an in-depth analysis to identify key trends and relationships between **Gross Margin Percentage** and other variables like **Product Line, Branch, Customer Type**, etc.



Correlation Analysis: Checked the correlation between **Gross Margin Percentage** and other variables.



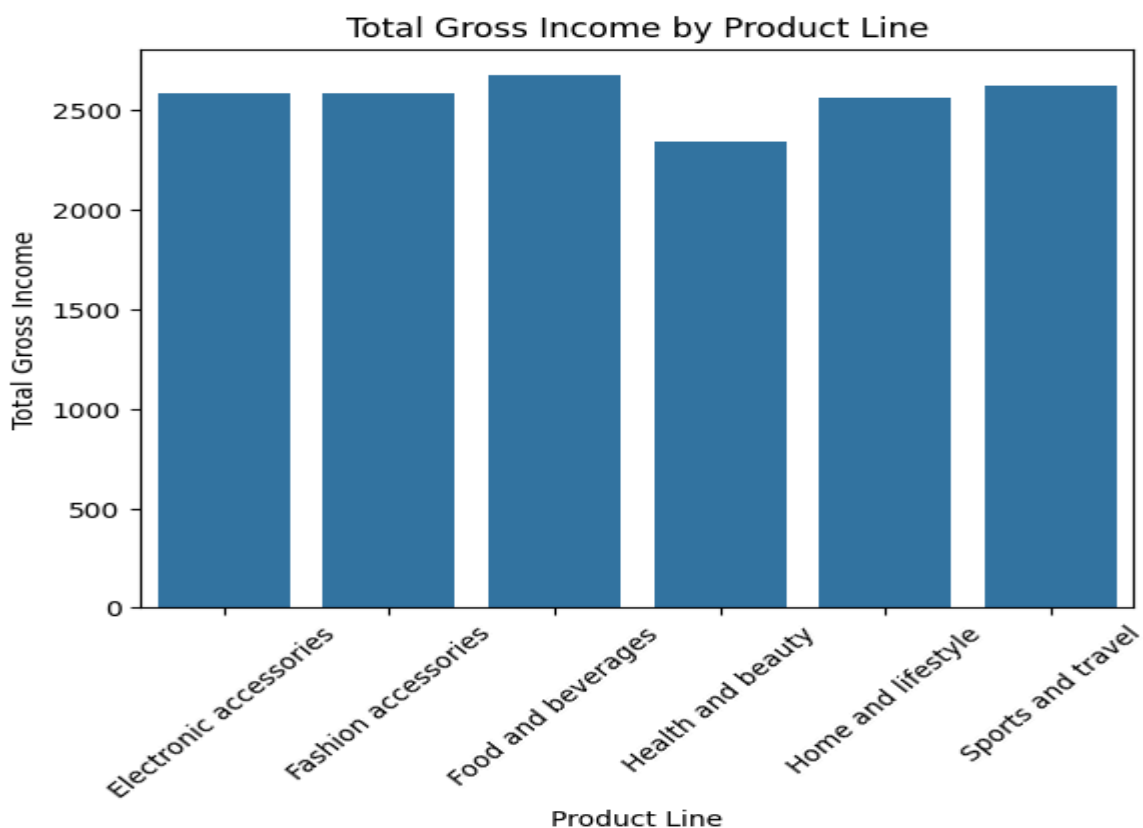
Payment Types: The payments over time by line plot



Key Discoveries and Findings

High Impact Features on Gross Margin and Profitability:

- **Product Line:** Certain product lines contribute more significantly to the gross margin percentage. Products with higher margins should be prioritized in marketing and sales strategies to maximize profitability.



- **Unit Price and Quantity:** The relationship between unit price and quantity sold is critical. Higher unit prices contribute directly to higher gross margins, but this must be balanced with sales volume to ensure that overall gross income is not negatively impacted.
- **Customer Type:** Members versus non-members show different purchasing behaviors, with members often contributing more consistently to higher margins. Encouraging more customers to become members can help stabilize and improve margins

Customer Segmentation:

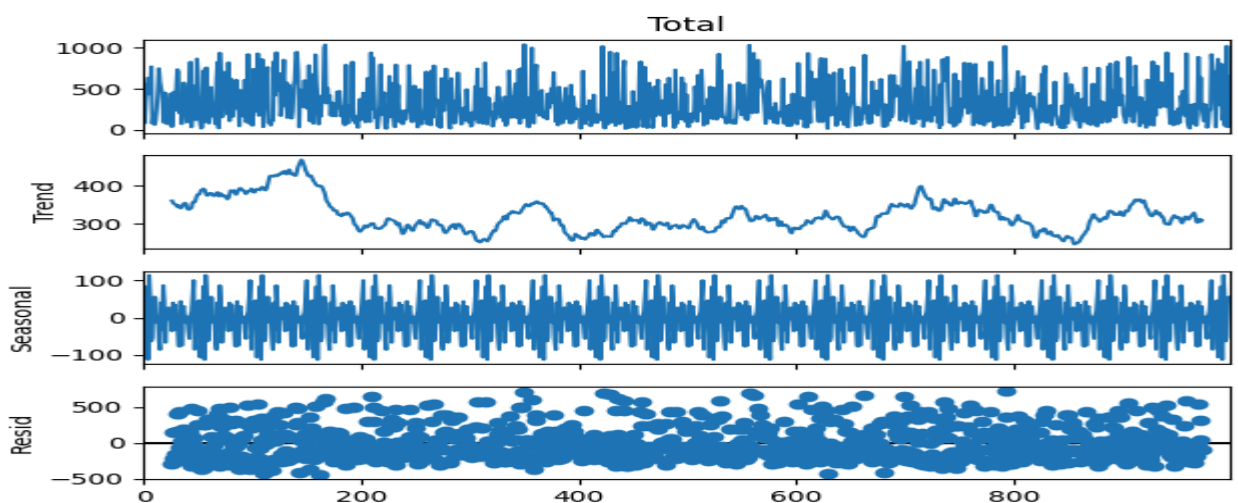
- **Segmentation by Purchase Patterns:** Customers who make frequent purchases tend to contribute more to gross income. Identifying and targeting these high-value customers with personalized offers or loyalty programs can help maintain or increase profitability.



- **Churn Risk Identification:** By predicting customer churn, the company can proactively engage customers who are at risk of leaving, particularly those who contribute significantly to gross margin.

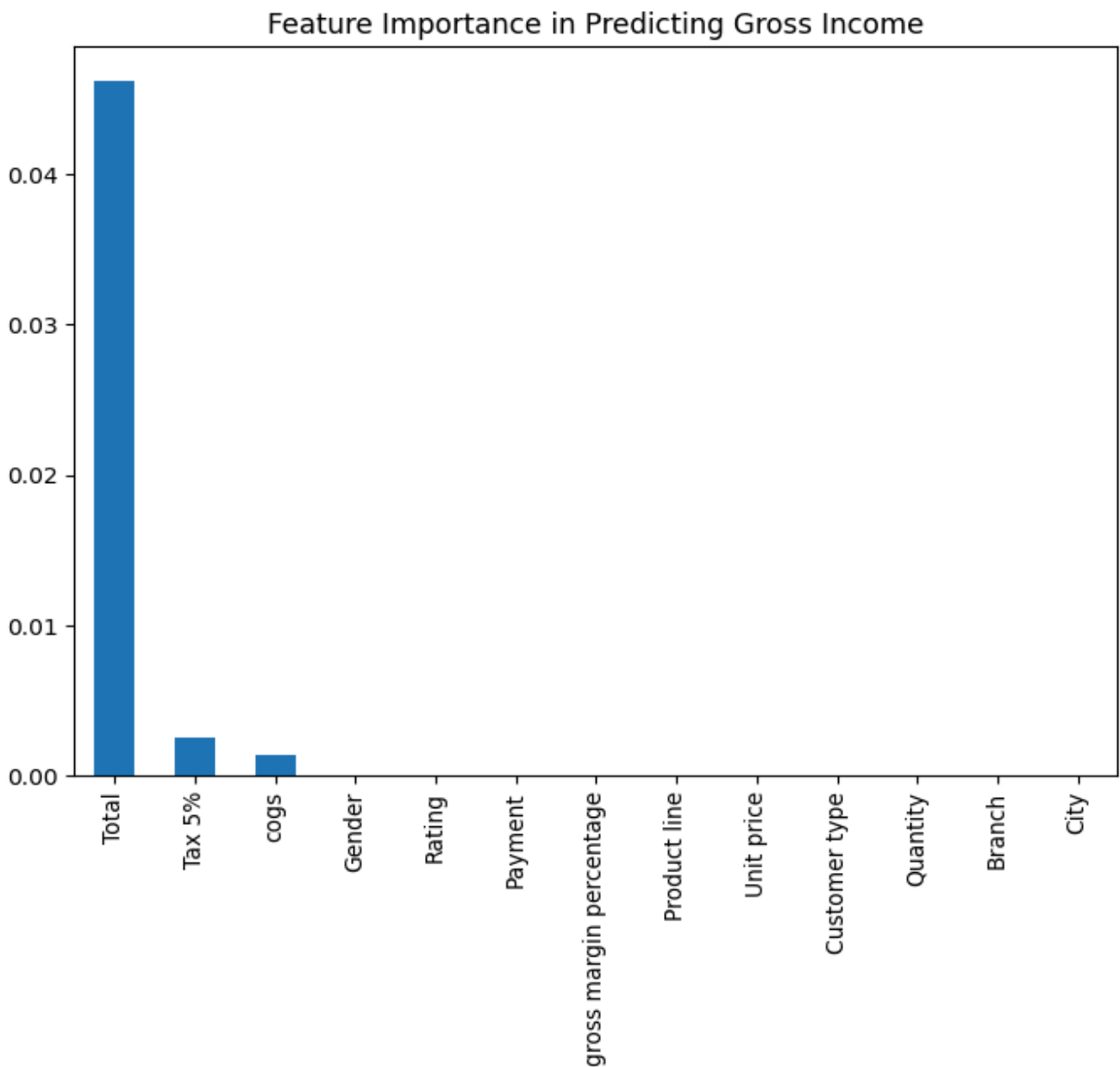
Time Series Analysis:

- **Seasonal Trends:** The analysis might reveal that certain times of the year see higher gross margins, perhaps due to holiday seasons or promotions. Understanding these trends allows the company to plan inventory



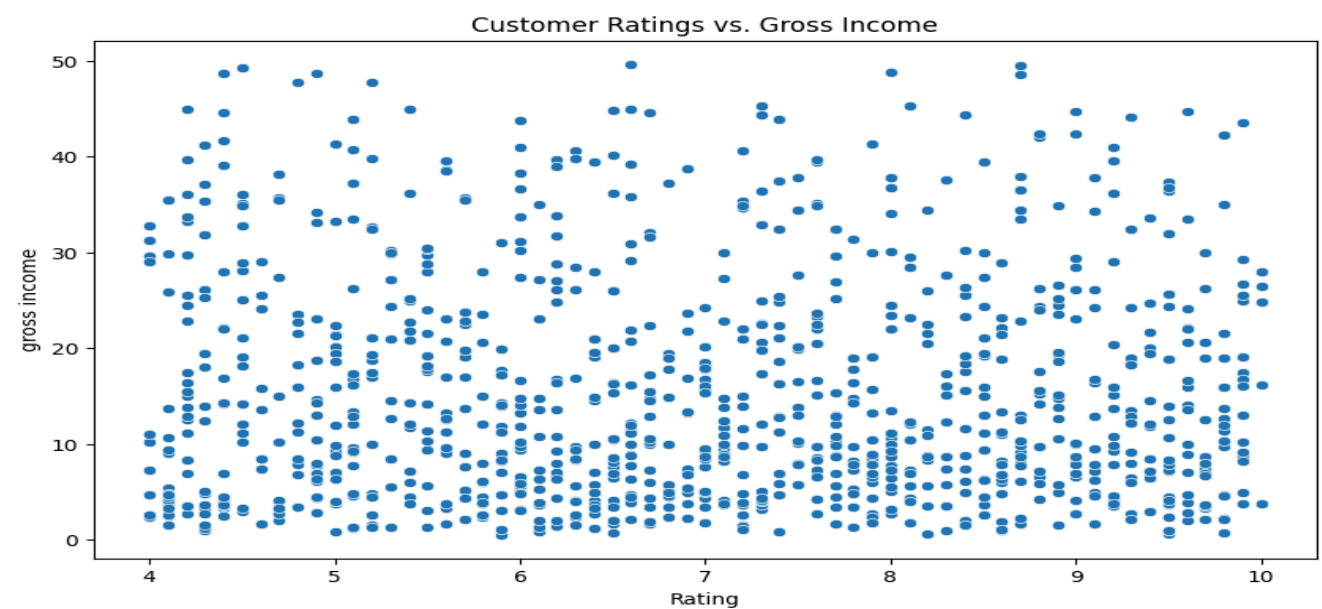
Correlation and Feature Importance:

- **Strong Correlation Between COGS and Gross Income:** As expected, the cost of goods sold (COGS) has a strong correlation with gross income. However, the goal should be to optimize pricing and cost strategies to ensure that gross margins remain high without eroding gross income.
- **Importance of Payment Methods:** While payment methods may not have shown high importance in initial models, optimizing these (e.g., encouraging lower-fee options like e-wallets) could contribute marginally to profitability without affecting gross margin negatively.



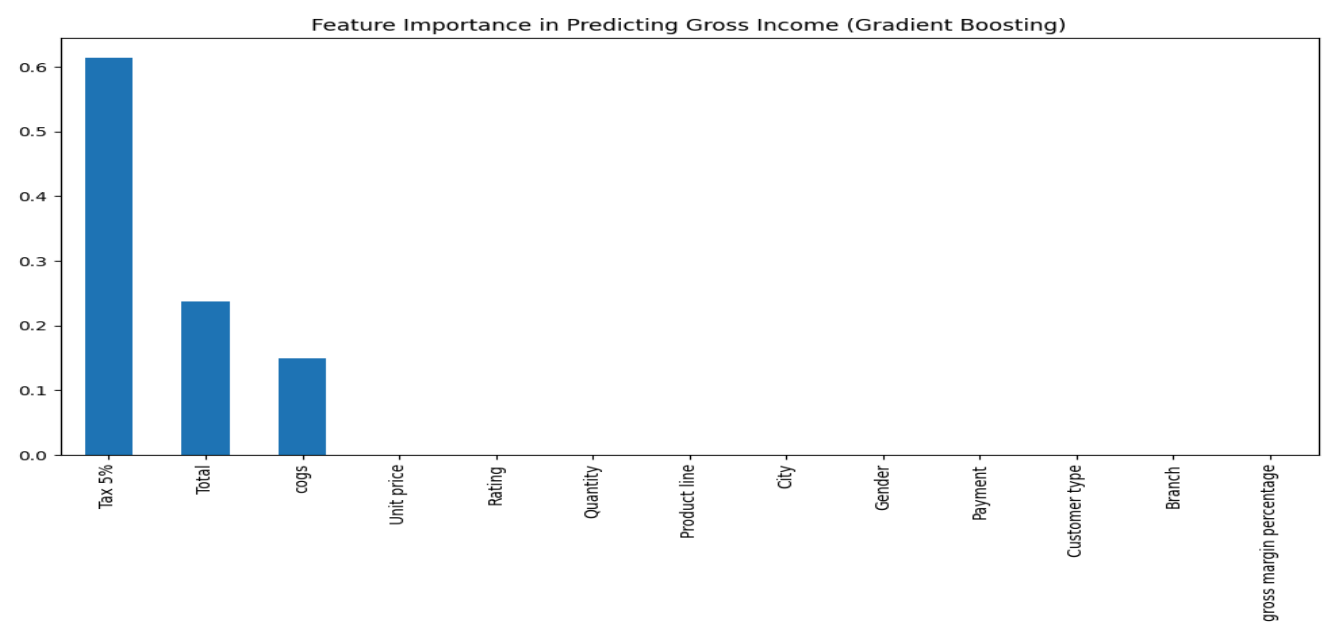
Customer Rating Analysis:

- **Impact of Customer Satisfaction:** Higher customer ratings are associated with repeat purchases and higher gross margins. Improving customer satisfaction should be a priority,for products with lower margins, to drive repeat sales and enhance profitability.



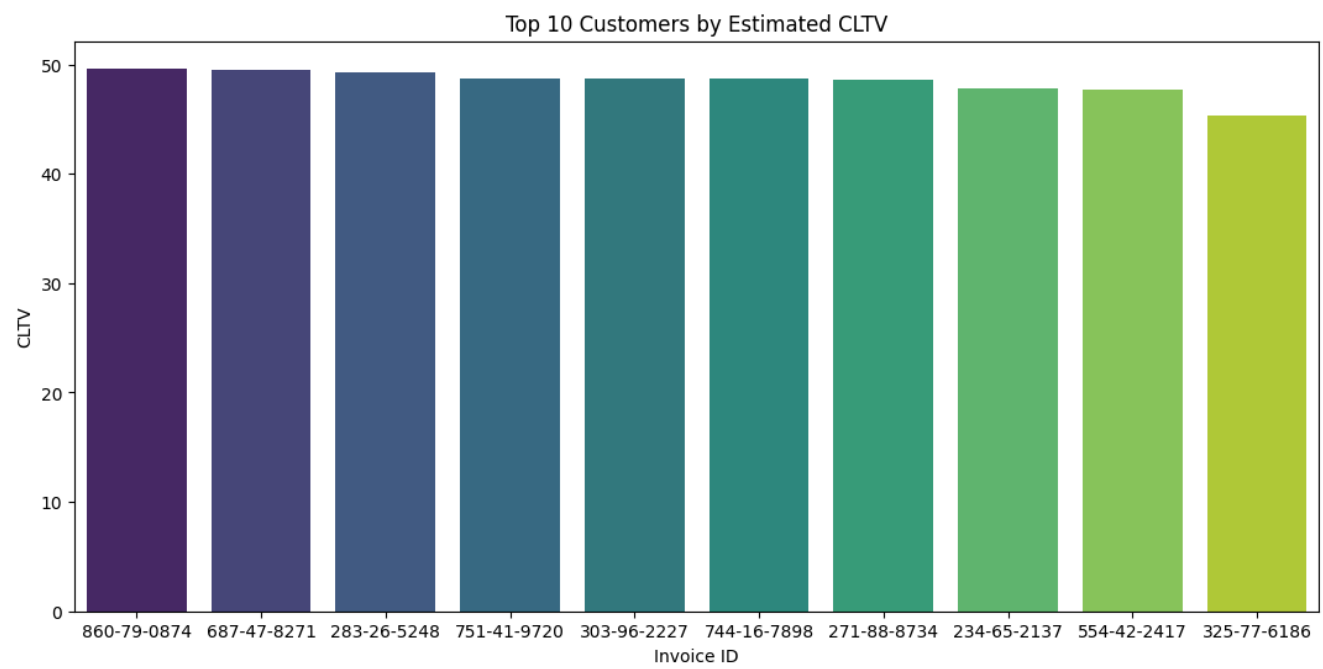
Predictive Modeling for Sales Forecasting:

Used machine learning models like Random Forests or Gradient Boosting to predict **Gross Income** based on features.



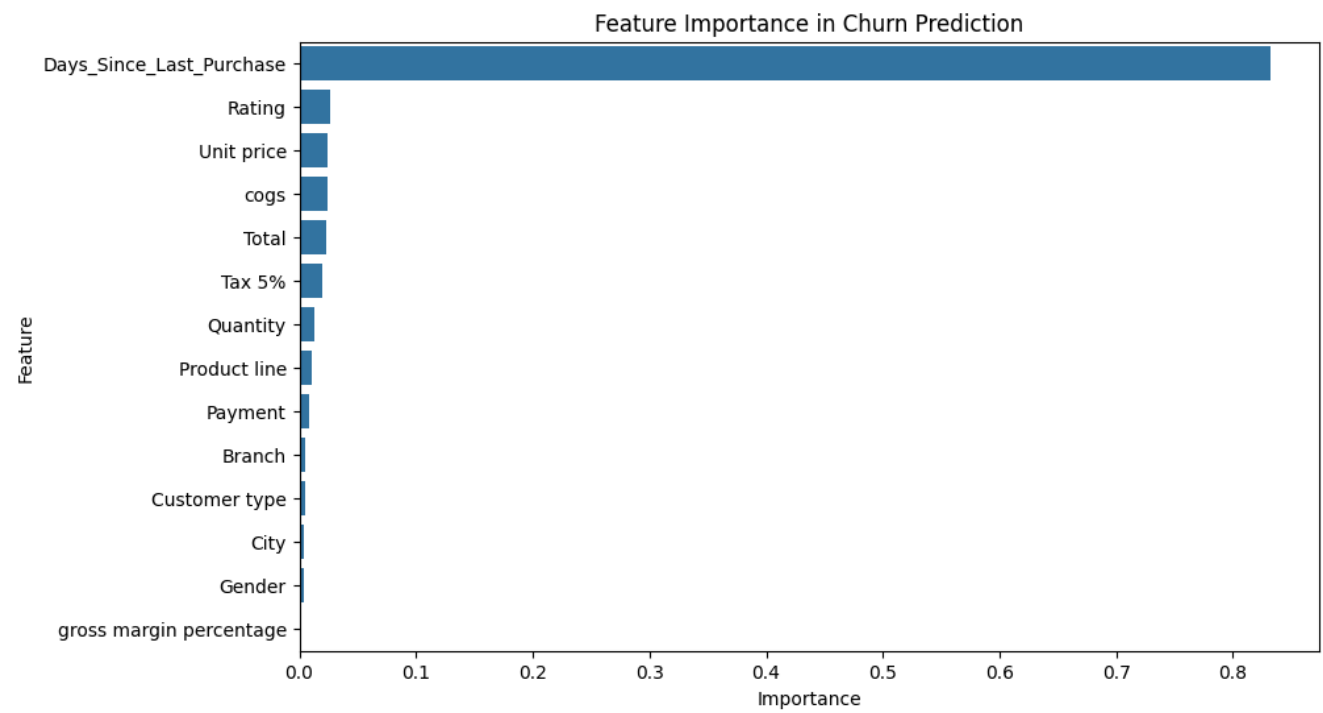
Customer Lifetime Value (CLTV) Prediction

Predicting Customer Lifetime Value (CLTV) can be another advanced technique, especially useful in retail contexts.



Churn Prediction

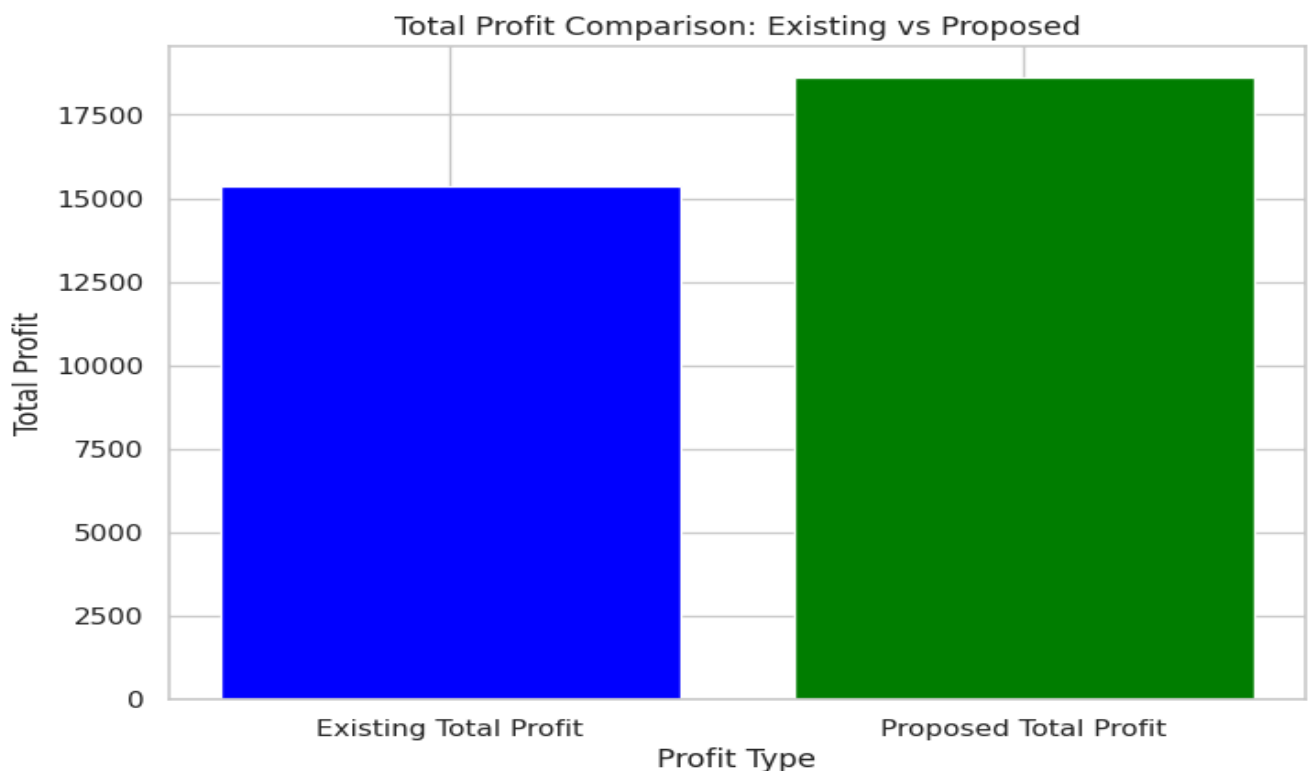
Predicting customer churn is crucial for retaining customers and improving profitability.



Summary:

- **Grouping by Product Line:** Data was grouped by **Product line** to calculate total sales revenue, COGS, and gross income for each category.
- **Current Gross Margin Analysis:** The existing gross margin percentage was computed for each product line to confirm consistency.
- The proposed gross margin adjustments resulted in an increase of the total gross income from **15,379.37** to **18,609.04**, demonstrating a successful optimization strategy.
- **Predictive Modeling** helps identify which features most affect **Gross Income** and predict future income.
- **Customer Lifetime Value (CLTV) Prediction** estimates long-term revenue from each customer.
- **Churn Prediction** identifies customers likely to stop purchasing, allowing for targeted retention efforts.

These analyses will give a comprehensive understanding of the data and uncover actionable margin adjustments that led to an increase in overall profitability.



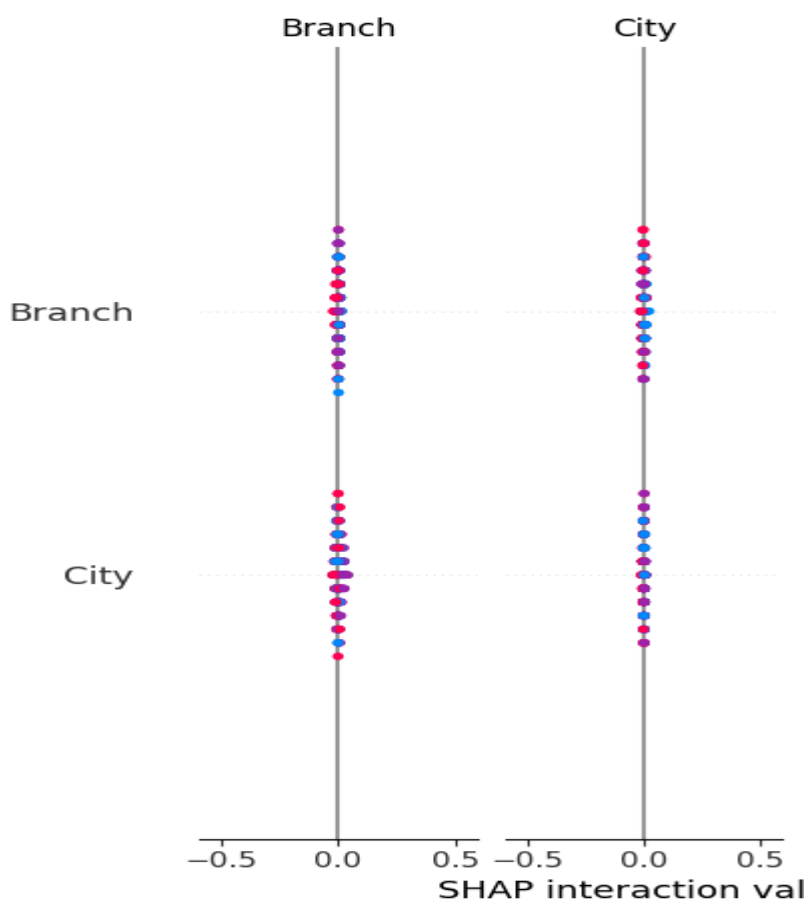
Advanced Techniques

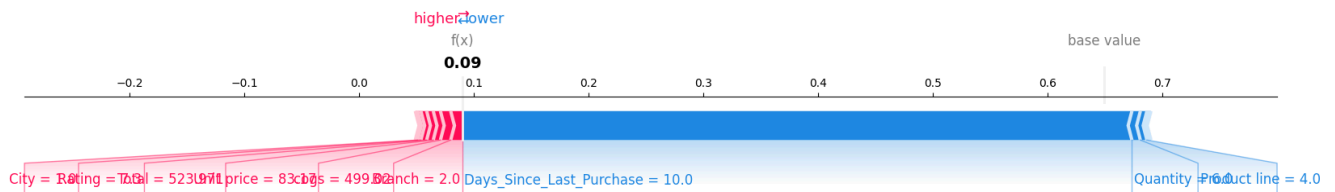
SHAP Analysis: Project Perspective

SHAP (SHapley Additive exPlanations) provides a way to explain the output of machine learning models by attributing the contribution of each feature to the final prediction. In this project, the goal was to predict customer churn and understand the key factors driving it.

Understanding Feature Contributions

- **SHAP Values:** Each feature in the dataset receives a SHAP value that quantifies its impact on the model's prediction. A positive SHAP value indicates that the feature pushes the prediction towards classifying the customer as likely to churn, while a negative value pushes it towards retention.
- **Summary Plot:** The SHAP summary plot shows how much each feature contributes to the prediction across the entire dataset. Features are ranked by their average impact on model output, giving insight into which factors are most influential in predicting churn.





Key Insights for Business Strategy

- **High-Impact Features:**

- **Days Since Last Purchase:** Customers who haven't made a purchase recently are more likely to churn. This insight suggests that retention efforts should focus on re-engaging customers who have been inactive for a period of time.
- **Product Line:** Certain product lines may be associated with higher churn rates. For instance, if customers who purchase from a specific product line are more likely to churn, the company might investigate potential issues related to that product line or target customers with special offers.
- **Customer Type:** Whether a customer is a member or a non-member significantly impacts churn probability. Members might have lower churn rates due to loyalty benefits, suggesting that converting non-members to members could reduce churn.

- **Less Impactful Features:**

- Features like **Payment Method** might have lower SHAP values, indicating they have less influence on churn. This means the company might not need to prioritize changes in payment options when addressing churn.

Profitability Segmentation:

- **RFM Analysis (Recency, Frequency, Monetary):** This technique segments customers based on how recently they purchased (Recency), how often they purchase (Frequency), and how much they spend (Monetary). This can help identify the most profitable customers, allowing for targeted marketing and retention efforts to maximize gross margins.

Price Sensitivity Analysis:

- **Price Elasticity of Demand:** Understanding how changes in price affect the quantity demanded can help set optimal pricing strategies. If certain product lines are highly elastic, small price reductions could lead to significant increases in sales, potentially boosting gross income while maintaining or improving gross margins.

Customer Lifetime Value (CLTV) Prediction:

- **CLTV Modeling:** Estimating the lifetime value of customers can help prioritize marketing and retention efforts towards those customers who are likely to contribute the most to long-term profitability. This can be especially useful when combined with churn prediction to identify and retain high-value customers.

CLTV Calculation: Customer Lifetime Value (CLTV) is estimated using the formula:

- $$\text{CLTV} = \frac{\text{Monetary}}{\text{Frequency}} \times \text{Frequency} \times (1 - \text{churn rate})$$

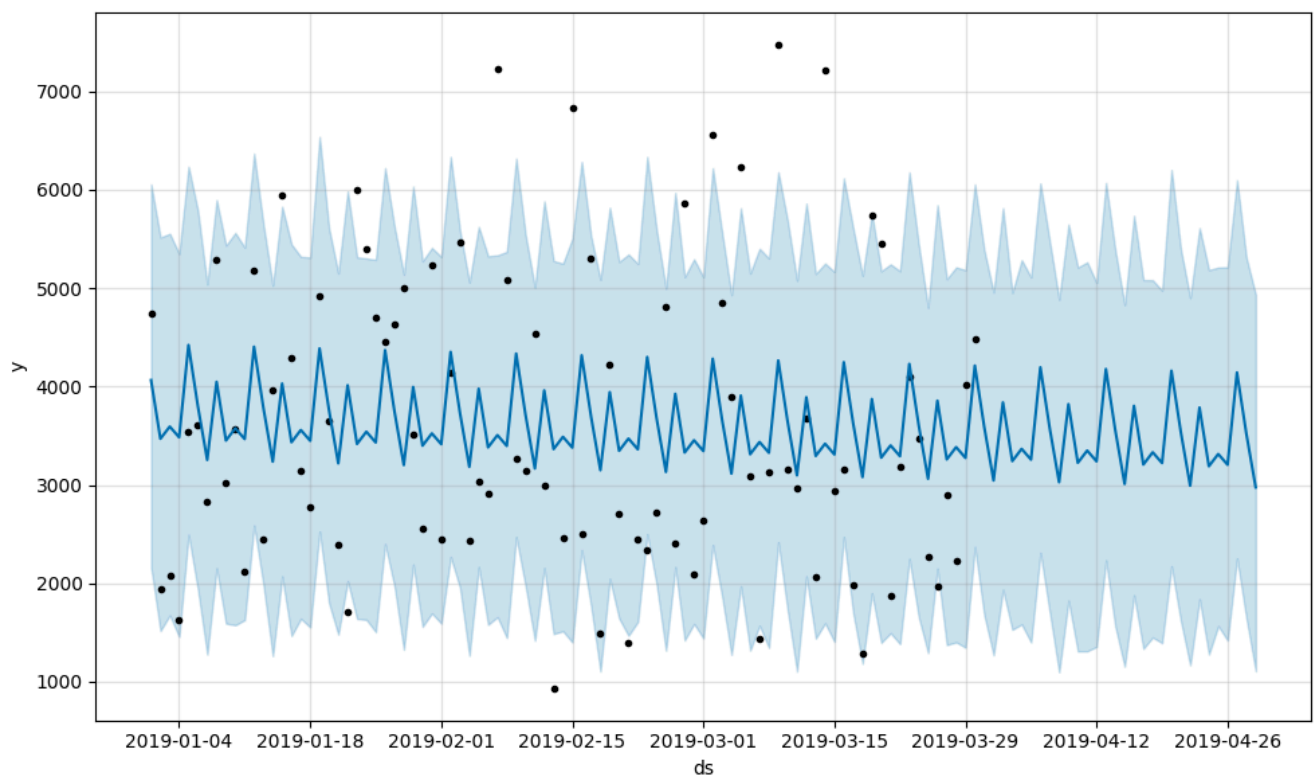
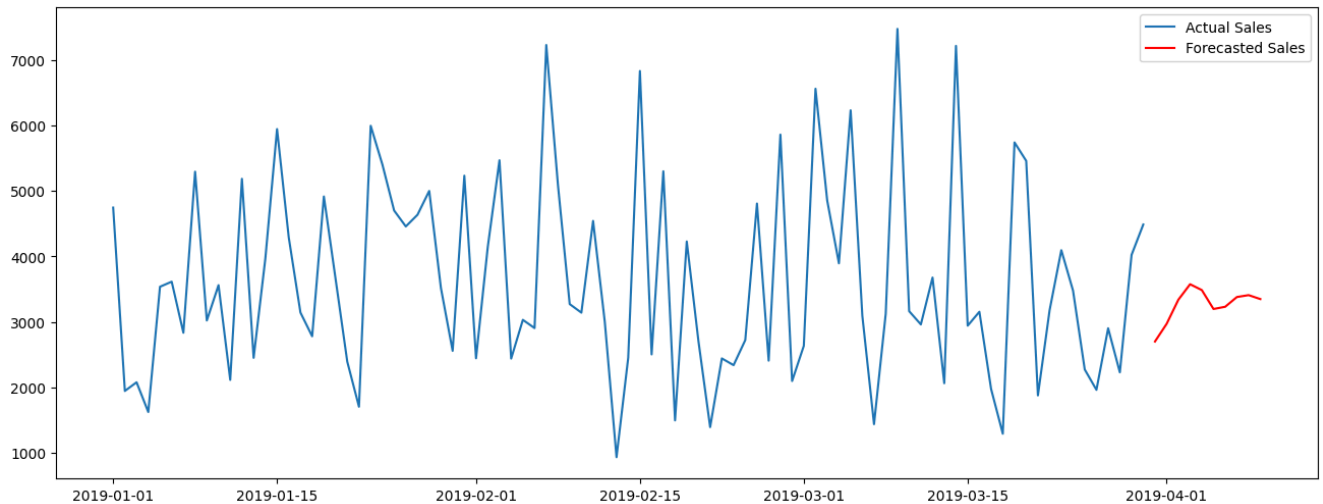
$$\text{CLTV} = \text{Frequency} \times \text{Monetary} \times (1 - \text{churn rate})$$
- Here, **Monetary** represents the total amount spent, and **Frequency** is the number of transactions.

Market Basket Analysis:

- **Association Rule Mining:** This analysis helps to identify product combinations frequently purchased together. By understanding these associations, you can optimize product placement, cross-selling, and bundling strategies to increase overall sales and improve gross margins

Predictive Modeling for Sales Forecasting:

- **ARIMA or Prophet Models:** Using time series forecasting models like ARIMA or Prophet can predict future sales trends. Accurate forecasts allow better planning for inventory, staffing, and marketing, directly impacting gross margin and profitability.



Additional predictive modeling techniques can be applied

- Gradient Boosting (e.g., XGBoost, LightGBM)
- K-Means Clustering
- Support Vector Machine (SVM)
- Time Series Forecasting (ARIMA)
- Neural Networks

Recommendations

Based on the analysis, the following recommendations were made:

Focus on High-Margin Product Lines:

- Prioritize marketing and sales efforts on product lines that consistently yield higher gross margins. This can be done through targeted promotions, upselling, or bundling strategies.

Enhance Member Loyalty Programs:

- Increase efforts to convert non-members to members, as members tend to contribute more to gross margins. This could be achieved through exclusive member discounts or rewards for frequent purchases.

Monitor and Adjust Pricing Strategies:

- Continuously analyze the relationship between unit prices and sales volumes to ensure that pricing strategies maximize gross margins without reducing gross income.

Target High-Value, At-Risk Customers:

- Use churn prediction models to identify high-value customers at risk of leaving and engage them with tailored retention strategies, such as special offers or personalized communications.

Leverage Seasonal Trends:

- Plan inventory, staffing, and marketing around identified seasonal trends to capitalize on periods of high gross margin and profitability.

Conclusion

The project successfully identified key factors impacting gross margin and proposed actionable strategies to optimize it while maintaining overall profitability. The advanced analysis techniques provided deep insights into customer behavior and profitability drivers.

By implementing these strategies, can optimize its gross margin percentage while maintaining or even enhancing overall profitability, thereby achieving the objective. The insights derived from the analysis provide strong data-driven decisions that align with business goals.

Future Work

- **Time Series Forecasting:** Predict future sales trends based on historical data.
- **Optimization Algorithms:** Implement optimization techniques to further fine-tune pricing and promotions.
- **Real-Time Analytics:** Develop dashboards for real-time monitoring of key metrics and customer behavior.

Thank You!

