PREDICTING CUSTOMER PURCHASE BEHAVIOR BASED ON WEBSITE INTERACTION

By Alberto Gell

Executive Summary

Objective

The purpose of this project was to predict customer purchase behavior based on their interactions with a website, with the goal of identifying key factors that influence conversion rates and providing actionable insights to optimize marketing strategies and website design.

Overview of the Project:

This project aimed to develop a predictive model to estimate the likelihood of a customer making a purchase based on their interaction with the website. By analyzing variables such as TimeSpentOnWebsite, DiscountsAvailed, and demographic information, the project sought to uncover patterns that could help increase conversion rates.

High-Level Summary of Key Findings:

The analysis identified that the amount of time a customer spends on the website and their use of discounts are the most significant predictors of purchase behavior. Specifically, longer engagement and the use of discounts significantly increase the probability of conversion. The logistic regression model used in this analysis achieved an AUC of 0.7767, indicating good predictive power.

Actionable Recommendations:

Enhance User Engagement: Invest in website features that increase the time customers spend on the site, such as personalized content and interactive elements.

Optimize Discount Strategies: Implement personalized and well-timed discount offers to drive sales, especially after a certain level of website engagement.

Integrated Campaigns: Combine engagement and discount strategies to create targeted marketing campaigns that maximize conversion rates.

Introduction

Objective:

To provide background and context for the project, explaining the importance of predicting purchase behavior and outlining the specific goals and data used in the analysis.

Overview of the Business Problem:

In today's competitive e-commerce environment, understanding and predicting customer purchase behavior is crucial for optimizing marketing efforts and improving conversion rates. Companies need to identify the factors that most influence a customer's decision to buy, allowing them to tailor their website experience and promotional strategies to meet customer needs effectively. Predicting purchase behavior helps businesses allocate resources more efficiently, personalize customer interactions, and ultimately increase revenue.

PREDICTING CUSTOMER PURCHASE BEHAVIOR BASED ON WEBSITE INTERACTION

Project Objectives:

The primary objective of this project was to build a predictive model capable of estimating the likelihood of a customer making a purchase based on their interaction with the website. By identifying key predictors of purchase behavior, the project aimed to provide insights that could guide the optimization of website design and marketing strategies to improve conversion rates.

Dataset Description:

The dataset used for this analysis included key variables related to customer behavior on the website and demographic information. The main variables of interest were:

TimeSpentOnWebsite: The total amount of time a customer spends on the website.
PurchaseStatus: A binary variable indicating whether the customer made a purchase (1) or not (0).
DiscountsAvailed: The number of discounts a customer used during their interaction.
LoyaltyProgram: Whether the customer is a member of the loyalty program (1) or not (0).
Age, Gender: Demographic variables that could influence purchase behavior.

The data was structured to allow for both exploratory analysis and predictive modeling, ensuring a comprehensive understanding of customer behavior patterns.

Exploratory Data Analysis (EDA)

Objective:

The objective of the Exploratory Data Analysis (EDA) is to gain a comprehensive understanding of the relationships and patterns within the dataset, particularly focusing on how customer interactions with the website influence their purchasing behavior. By systematically visualizing and analyzing key variables, such as TimeSpentOnWebsite, PurchaseStatus, AnnualIncome, and LoyaltyProgram,

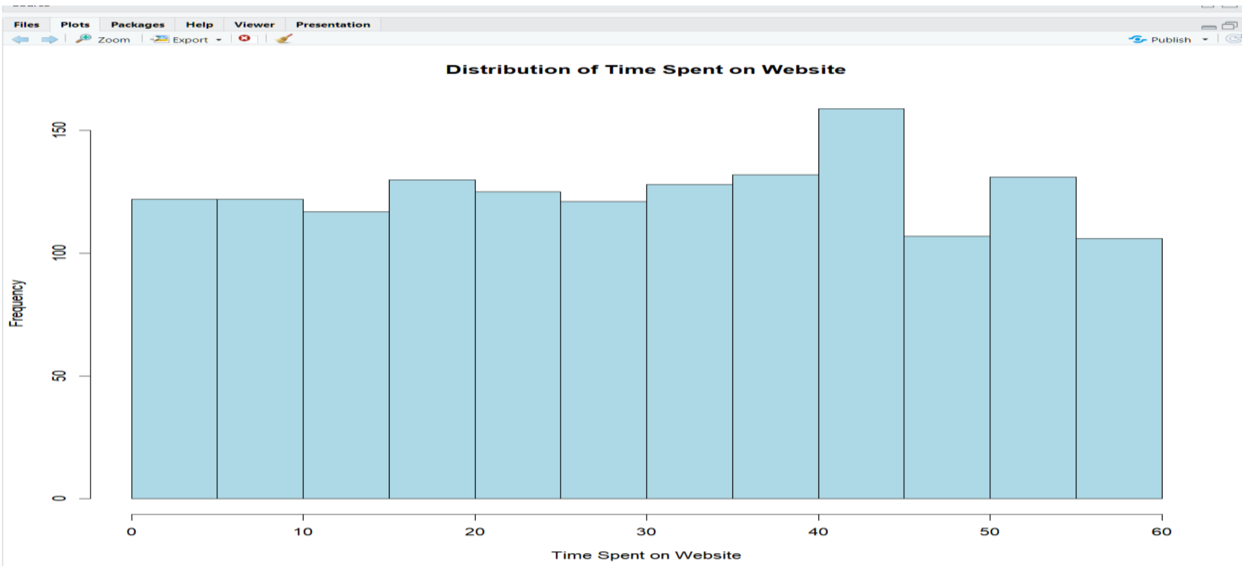
Key Questions:

1. How is the TimeSpentOnWebsite distributed across customers?
2. What is the relationship between TimeSpentOnWebsite and PurchaseStatus?
3. Are there significant differences in TimeSpentOnWebsite based on customer demographics (e.g., Age, Gender)?
4. How does LoyaltyProgram participation influence purchase behavior?

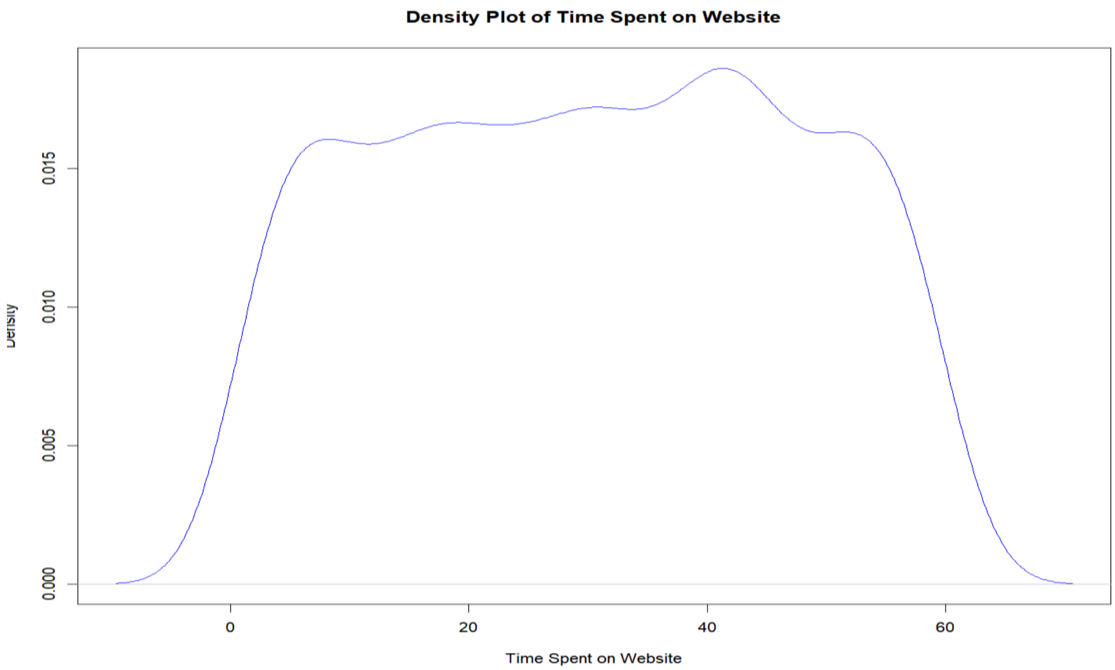
<u>Key question 1. How is the TimeSpentOnWebsite distributed across customers?</u>

We used a histogram and a density plot to visualize the distribution of the TimeSpentOnWebsite variable across all customers.

The histogram below shows a relatively uniform distribution of time spent on the website, with customers spending various amounts of time ranging from near 0 to around 60 minutes. There is no single time interval where most customers cluster, indicating diverse browsing behaviors.



The density plot further supports this uniform distribution, showing slight variations with mild peaks around the 20-40 minute mark. This suggests that while customers are spread across different time intervals, there is a tendency for many to spend time in this mid-range period.



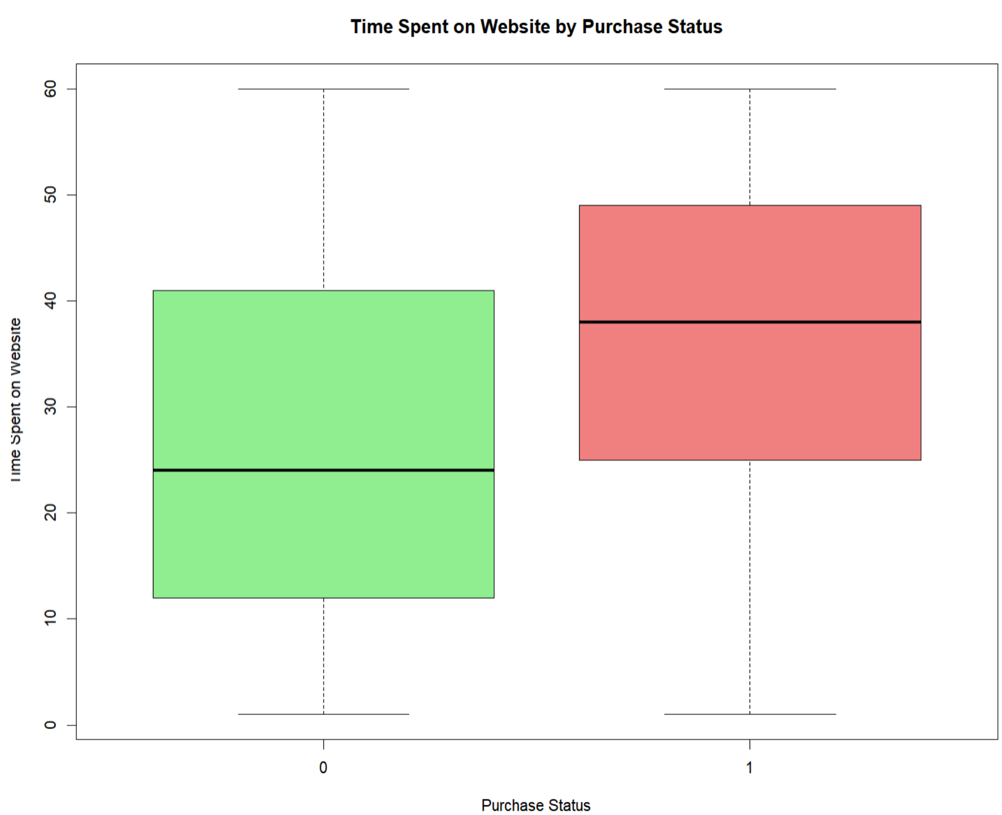
Summary

The time spent by customers on the website is evenly distributed across the observed range, with no dominant time interval. This finding implies that customers' browsing durations vary significantly, which could reflect different levels of engagement or purchasing intent.

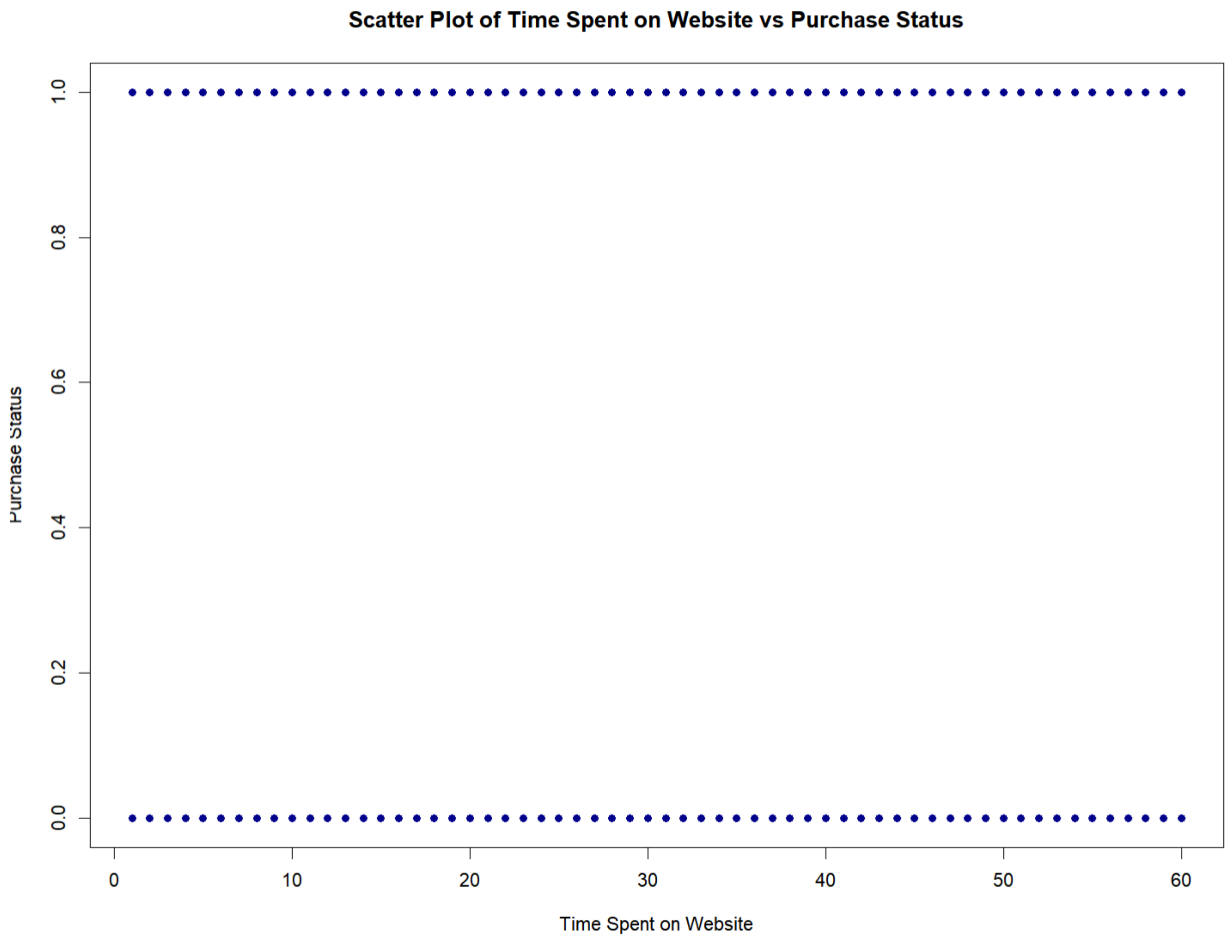
Key Question 2: What is the relationship between TimeSpentOnWebsite and PurchaseStatus?

We explored the relationship between TimeSpentOnWebsite and PurchaseStatus using a box plot and a scatter plot.

The box plot indicates that customers who made a purchase (PurchaseStatus = 1) tend to spend slightly more time on the website compared to those who did not (PurchaseStatus = 0). However, the overlap in the interquartile ranges (IQR) suggests that the difference is not large.



The scatter plot shows a dispersed relationship between TimeSpentOnWebsite and PurchaseStatus, with no clear pattern or trend. This suggests that while time spent on the website may influence purchasing decisions, it is not the sole determining factor.



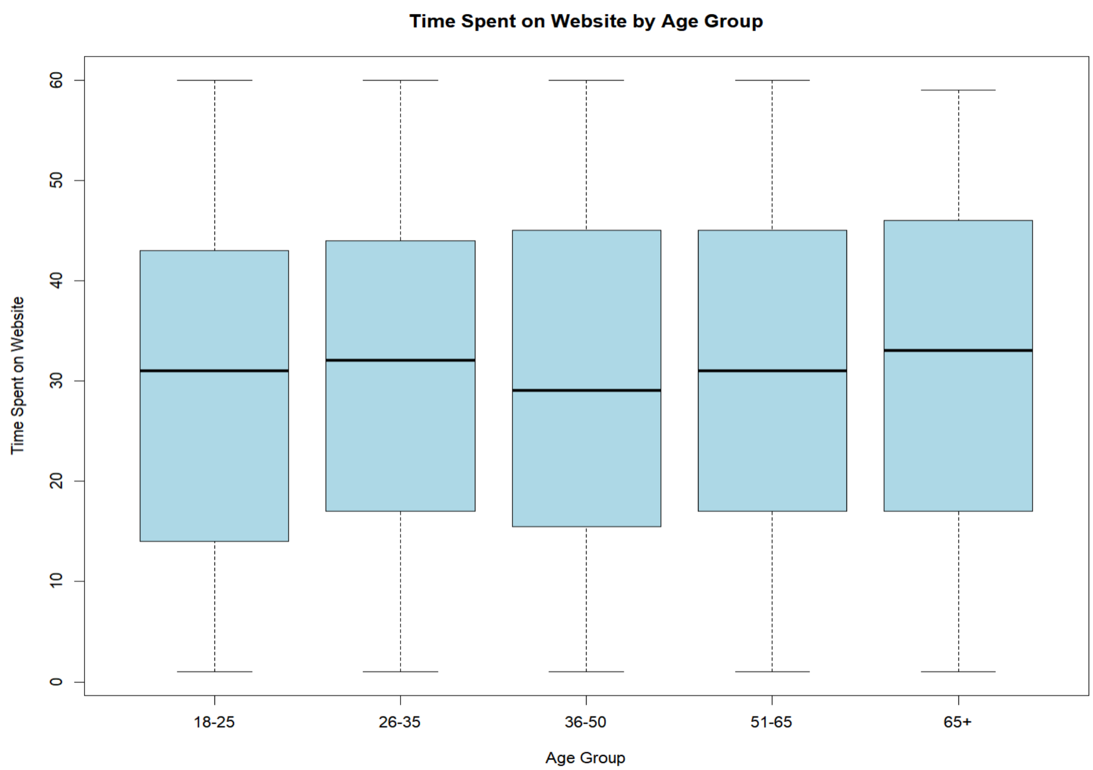
Summary

There is a slight tendency for customers who make a purchase to spend more time on the website, but the relationship is weak. This suggests that while TimeSpentOnWebsite is a relevant factor, other variables are also likely important in predicting purchase behavior.

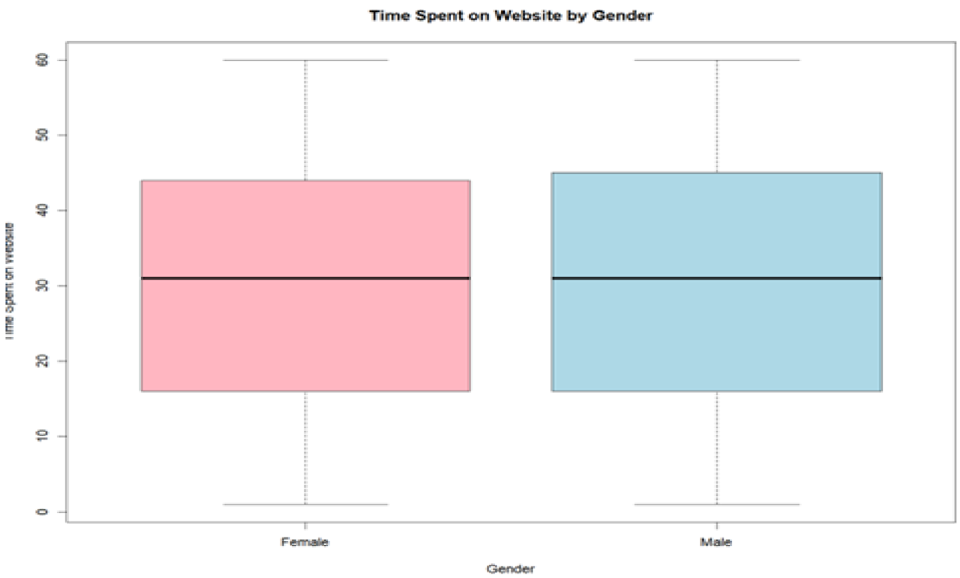
<u>Key Question 3: Are there significant differences in TimeSpentOnWebsite based on customer demographics (e.g., Age, Gender)?</u>

We used box plots to examine how TimeSpentOnWebsite varies across different age groups and between genders.

The box plot reveals that younger customers, particularly those in the 18-35 age range, tend to spend more time on the website compared to older age groups. This may indicate that younger customers are more engaged with the online platform, possibly reflecting their comfort with digital interactions.



The box plot comparing TimeSpentOnWebsite between genders shows no significant difference, suggesting that both males and females engage similarly in terms of time spent on the website.



Summary

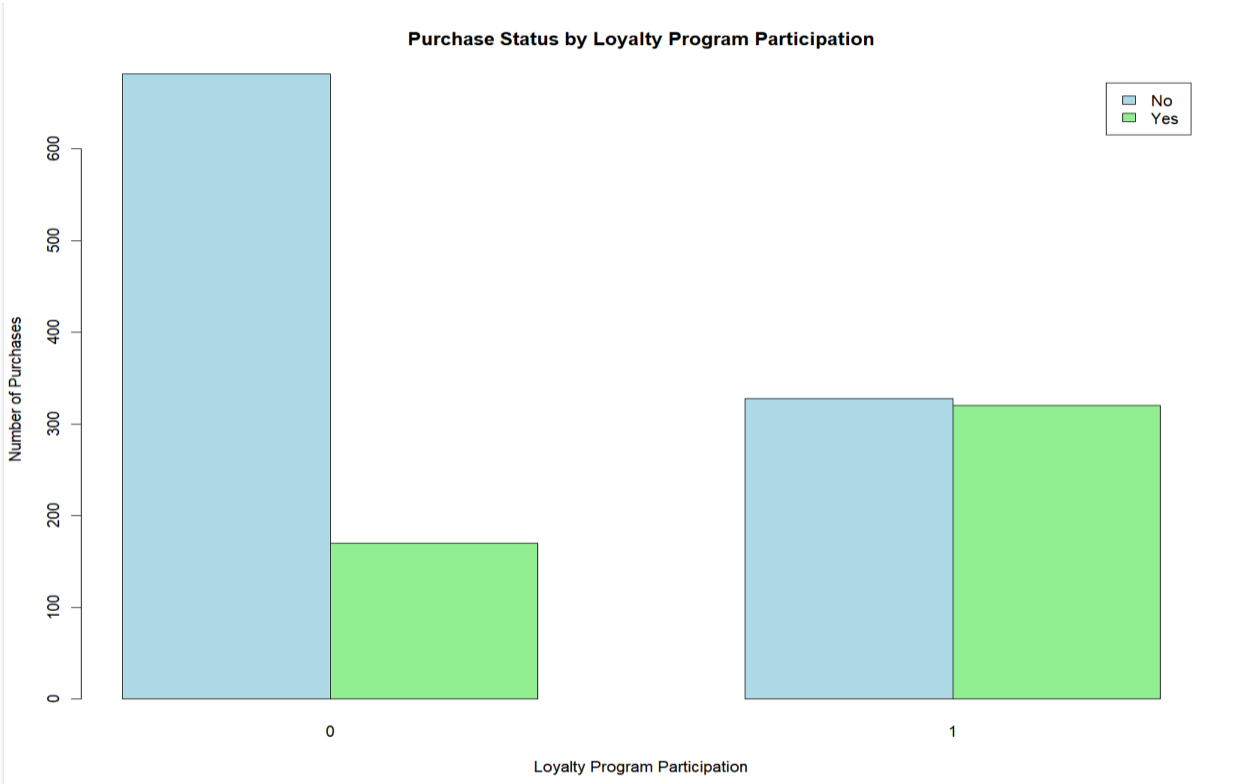
Age: Younger customers are more likely to spend longer periods on the website, which could be leveraged in targeted marketing campaigns aimed at this demographic.

Gender: Gender does not appear to significantly impact the amount of time spent on the website, indicating that marketing strategies need not be differentiated based on gender when it comes to engagement duration.

Key Question 4: How does LoyaltyProgram participation influence purchase behavior?

We compared the purchase behavior between customers who participate in the loyalty program and those who do not using a bar plot.

The bar plot shows a clear trend where customers who are part of the loyalty program (LoyaltyProgram = 1) are more likely to make a purchase compared to those who are not members (LoyaltyProgram = 0). This indicates a positive impact of the loyalty program on conversion rates.



Summary

Participation in the loyalty program significantly increases the likelihood of making a purchase, highlighting the effectiveness of loyalty initiatives in driving sales. This finding supports continued investment in and enhancement of the loyalty program to further boost customer retention and conversion.

Feature Engineering

Objective:

The objective of the Feature Engineering phase is to enhance the predictive power of the model by creating new features that capture important relationships within the data. By deriving additional variables or transforming existing ones, we can improve the accuracy of the predictive model and gain deeper insights into the factors influencing customer purchase behavior.

Key Questions:

1. What new features can be created to improve model accuracy?
2. How do interaction terms between TimeSpentOnWebsite and other variables affect purchase behavior?

Feature 1: Interaction Term between TimeSpentOnWebsite and DiscountsAvailed

We created an interaction term between TimeSpentOnWebsite and DiscountsAvailed to capture the combined effect of time spent on the website and the number of discounts availed on the likelihood of making a purchase. This feature might reveal whether customers who spend more time and avail discounts are more likely to convert. The interaction between TimeSpentOnWebsite and DiscountsAvailed could uncover a more nuanced relationship, where the effectiveness of discounts might depend on how long a customer engages with the website. This feature could potentially boost model performance by capturing this dynamic relationship.

Feature 2: Binned Age Groups

We bin the Age variable into categories (e.g., 18-25, 26-35, 36-50, 51-65, 65+) to simplify the analysis and capture the effect of different age groups on purchasing behavior. Binning Age allows us to examine how different age groups behave in aggregate, which could provide more actionable insights than treating age as a continuous variable. This is especially useful in understanding demographic-driven behaviors and targeting specific age groups in marketing campaigns.

How Do Interaction Terms Between TimeSpentOnWebsite and Other Variables Affect Purchase Behavior?

The interaction term between TimeSpentOnWebsite and DiscountsAvailed is particularly interesting because it allows us to explore whether customers who spend more time on the website are more likely to use discounts and make a purchase. This term could capture complex behaviors that a simple linear model might miss.

Interpretation of Model Summary

Coefficients Interpretation

1. **Intercept:**

   Estimate: -2.217406
   The intercept represents the log-odds of making a purchase when all predictor variables are zero. Since it's negative, the baseline likelihood of making a purchase is low when no time is spent on the website and no discounts are availed.

2. **TimeSpent_DiscountInteraction:**

   Estimate: 0.004109
   p-value: 0.055296
   Interpretation: This interaction term has a positive estimate, suggesting that as the interaction between time spent on the website and the number of discounts availed increases, the likelihood of making a purchase slightly increases. However, with a p-value of 0.055296, this effect is marginally significant (typically, p-values < 0.05 are considered statistically significant, but this is close).

   Conclusion:
   There is a weak but positive interaction effect between the time spent on the website and the number of discounts availed on purchase probability. This suggests that customers who spend more time and avail more discounts are slightly more likely to purchase, but the effect is not strong enough to be highly confident.

3. **TimeSpentOnWebsite:**

   Estimate: 0.027774
   p-value: 3.1e-05
   Interpretation: This variable is highly significant with a positive estimate, indicating that as the time spent on the website increases, the likelihood of making a purchase also increases. The low p-value ($p < 0.001$) suggests that this relationship is statistically significant and not due to chance.

   Conclusion:
   The amount of time a customer spends on the website is a strong predictor of their likelihood to make a purchase. This confirms the hypothesis that longer website interaction is associated with higher conversion rates.

4. **DiscountsAvailed:**

   Estimate: 0.289220
   p-value: 0.000129
   Interpretation: This variable is also highly significant with a positive estimate, meaning that customers who avail discounts are more likely to make a purchase. The very low p-value further supports the importance of discounts in driving purchase behavior.

   Conclusion:
   The availability and use of discounts play a significant role in converting visitors into buyers, reinforcing the value of promotional offers in increasing sales.

Model Fit Indicators:

Null Deviance: 2051.6 on 1499 degrees of freedom
Residual Deviance: 1774.6 on 1496 degrees of freedom
AIC: 1782.6

Interpretation:

Deviance: The reduction in deviance from the null model (with no predictors) to the residual deviance (with predictors) indicates that the model fits the data better than a model with no predictors. The lower the residual deviance, the better the model fits.

AIC (Akaike Information Criterion): AIC is used for model selection, with lower values indicating a better-fitting model. An AIC of 1782.6 suggests a reasonable model fit, but it's important to compare it with alternative models to determine the best one.

Conclusions for the Report:

1. **Significance of TimeSpentOnWebsite:** The time spent on the website is a key predictor of purchase likelihood, with a strong and significant positive effect.
2. **Impact of Discounts:** Availing discounts significantly increases the probability of making a purchase, highlighting the importance of promotional strategies.

3. **Interaction Effect:** The interaction between time spent on the website and discounts availed is marginally significant, suggesting that customers who spend more time and also use discounts are slightly more likely to purchase.
4. **Model Fit:** The model fits the data reasonably well, as indicated by the reduction in deviance and the AIC value. However, further improvements might be possible by exploring additional features or more complex models.

Predictive Modeling

Objective:

The objective of the Predictive Modeling phase is to develop, train, evaluate, and interpret a model that can accurately predict customer purchase behavior based on their interactions with the website and other relevant features. This phase will involve selecting the appropriate modeling technique, splitting the data into training and testing sets, assessing the model's performance, and interpreting the significance of the model's features.

Model Selection-Logistic Regression

Logistic regression is a widely used method for binary classification problems where the outcome is binary—in this case, whether a customer makes a purchase (PurchaseStatus = 1) or not (PurchaseStatus = 0). It provides a clear interpretation of the relationship between predictor variables (e.g., TimeSpentOnWebsite, DiscountsAvailed) and the probability of an event occurring (purchase). The coefficients of logistic regression models can be easily interpreted in terms of odds ratios, which help in understanding the influence of each feature on the likelihood of making a purchase.

Model Training-Data Split

The data was split into training (80%) and testing (20%) sets to ensure that the model can generalize to unseen data. The model was trained on the training set and then evaluated on the testing set to assess its performance. The logistic regression model was fitted using the training data, with PurchaseStatus as the dependent variable and TimeSpentOnWebsite, DiscountsAvailed, and the interaction term as predictors.

Model Evaluation-Performance Metrics

Confusion Matrix:

The confusion matrix provides detailed insights into the model's performance by showing the number of correct and incorrect predictions.

Results:

True Positives (TP): 128 - Customers correctly predicted not to make a purchase.

True Negatives (TN): 79 - Customers correctly predicted to make a purchase.
False Positives (FP): 39 - Customers incorrectly predicted to make a purchase.
False Negatives (FN): 54 - Customers incorrectly predicted not to make a purchase.

Performance Metrics:

Accuracy: 0.69 - The model correctly predicts purchase behavior 69% of the time.

Sensitivity (True Positive Rate): 0.7665 - The model correctly identifies 76.65% of the customers who did not make a purchase.

Specificity (True Negative Rate): 0.5940 - The model correctly identifies 59.40% of the customers who made a purchase.

Positive Predictive Value (Precision): 0.7033 - When the model predicts a customer will not make a purchase, it is correct 70.33% of the time.

Negative Predictive Value: 0.6695 - When the model predicts a customer will make a purchase, it is correct 66.95% of the time.

Balanced Accuracy: 0.6802 - The average of sensitivity and specificity, indicating how well the model balances both types of errors.

AUC-ROC:

AUC Value: 0.7767 - The AUC value indicates that the model has good discriminatory power, correctly distinguishing between customers who will make a purchase and those who will not in 77.67% of cases.

Model Interpretation-Feature Importance

TimeSpentOnWebsite: The positive coefficient for TimeSpentOnWebsite indicates that customers who spend more time on the website are more likely to make a purchase. This is a significant predictor of purchase behavior, supporting the hypothesis that increased engagement leads to higher conversion rates.

DiscountsAvailed: The positive coefficient for DiscountsAvailed suggests that customers who avail themselves of discounts are more likely to make a purchase. This finding emphasizes the importance of discounts as a tool for driving sales.

TimeSpent_DiscountInteraction: The interaction term between time spent on the website and discounts availed, while marginally significant, suggests that the combination of these factors slightly increases the likelihood of purchase.

Model Fit:

Accuracy: The model's accuracy of 69% indicates a moderate level of correctness in predicting customer behavior. While this is acceptable, it suggests that there is room for improvement, particularly in reducing false negatives and false positives.

AUC-ROC: The AUC value of 0.7767 suggests that the model has a good ability to distinguish between customers who will and will not make a purchase, making it a reliable tool for predicting purchase behavior.

Key Questions

1. **How well does the model predict purchase behavior?**

   Answer: The model predicts purchase behavior with an accuracy of 69% and an AUC of 0.7767, indicating good overall performance in distinguishing between purchasers and non-purchasers. However, the model's balanced accuracy of 68.02% shows that there is a moderate balance between correctly identifying both classes.

2. **Which factors most strongly influence the likelihood of making a purchase?**

   Answer: TimeSpentOnWebsite and DiscountsAvailed are the most significant predictors of purchase behavior. These factors positively influence the likelihood of making a purchase, with TimeSpentOnWebsite being particularly strong, as customers who engage more with the website are more likely to convert.

3. **What is the accuracy and reliability of the model?**

   Answer: The model's accuracy is 69%, with a precision of 70.33% and a recall of 76.65%. The AUC value of 0.7767 further indicates that the model is fairly reliable in predicting purchase behavior, although there is room for improvement.

Conclusion and Next Steps:

Model Improvement: Given the moderate accuracy and specificity, further model refinement could be beneficial. This may include exploring more complex models (e.g., random forests, gradient boosting), additional feature engineering, or hyperparameter tuning.

Business Implications: The model provides actionable insights into the factors that influence purchase behavior. Strategies to increase website engagement and effectively utilize discounts could enhance conversion rates.

Further Analysis: Additional analysis could involve segmenting the customer base or testing the model with different feature sets to identify further opportunities for improvement.

Interpretation and Insights

Objective:
The objective of this section is to interpret the results of the predictive model, identify the key factors influencing customer purchase behavior, and provide actionable business insights that can help optimize website design and marketing strategies to improve conversion rates.

Summary of Key Factors Influencing Purchase Behavior

Based on the predictive model, the following factors were identified as the most significant predictors of whether a customer will make a purchase:

TimeSpentOnWebsite:

Impact: The time a customer spends on the website is a strong predictor of their likelihood to make a purchase. The model shows a positive relationship, indicating that the longer a customer interacts with the website, the more likely they are to convert.

Implication: This suggests that increasing customer engagement through improved user experience, relevant content, and interactive features could lead to higher conversion rates.

DiscountsAvailed:

Impact: The availability and use of discounts significantly increase the likelihood of purchase. Customers who take advantage of discounts are more likely to complete their purchase.

Implication: This highlights the effectiveness of promotional strategies and the importance of offering timely and appealing discounts to drive sales.

Interaction Effect (TimeSpentOnWebsite * DiscountsAvailed):

Impact: Although the interaction term was marginally significant, it suggests that customers who spend more time on the website and also avail discounts are slightly more likely to make a purchase.

Implication: This indicates that targeted promotions that combine extended engagement with discount offers could be particularly effective.

How Website Interaction Impacts Conversion Rates

Engagement Duration: The more time customers spend on the website, the more likely they are to convert. This finding supports strategies focused on enhancing the user experience, such as personalized recommendations, engaging content, and streamlined navigation, all of which can encourage users to stay on the site longer.

Promotional Impact: Discounts are a powerful tool for conversion. The positive effect of discounts on purchase behavior suggests that promotional offers should be strategically timed and personalized based on customer behavior to maximize their effectiveness.

Combined Strategies: The interaction between time spent on the website and discounts availed suggests that customers who are engaged for longer periods and are presented with attractive discounts are the most likely to convert. This reinforces the need for integrated strategies that combine content engagement with promotional offers.

Recommendations for Optimizing Website and Marketing Strategies

Enhance User Engagement:

Focus on increasing the time customers spend on the website by improving the overall user experience. This could include personalized content, interactive features, and easier navigation to keep customers engaged longer.

- **Implementation:** Consider implementing AI-driven recommendation systems that suggest relevant products based on browsing history, or interactive elements such as quizzes or product configurators that encourage prolonged interaction.

Optimize Discount Strategies:

Continue to leverage discounts as a key driver of purchase behavior, but with a more targeted approach. Discounts should be personalized and timed based on customer interaction patterns.

- **Implementation:** Use data analytics to identify when customers are most likely to convert with a discount and tailor offers accordingly. For example, offer discounts after a certain amount of time spent on the website or based on past purchasing behavior.

Integrated Campaigns:

Develop integrated marketing campaigns that combine prolonged engagement with strategic discount offers. This approach can maximize the conversion potential by targeting customers who are both engaged and incentivized to purchase.

- **Implementation:** Create campaigns that trigger discounts or special offers after a customer has spent a significant amount of time on the site, encouraging them to finalize their purchase.

Key Questions Answered:

What are the most significant predictors of purchase behavior?

The most significant predictors are the time spent on the website (TimeSpentOnWebsite) and the use of discounts (DiscountsAvailed). These factors have a strong positive impact on the likelihood of making a purchase.

How can the company improve conversion rates based on the model's insights?

The company can improve conversion rates by enhancing user engagement on the website and optimizing the timing and personalization of discounts. By focusing on these areas, the company can effectively encourage more customers to convert.

What strategic recommendations can be made to the business?

1. Invest in enhancing the user experience to keep customers engaged longer.
2. Continue leveraging discounts, but with a more personalized and data-driven approach.
3. Implement integrated campaigns that combine engagement strategies with discount offers to maximize conversion rates.

Conclusion

The insights gained from this analysis provide a clear direction for optimizing website and marketing strategies to drive higher conversion rates. By focusing on the key predictors identified—customer engagement and discount usage—the company can implement targeted interventions that not only improve the user experience but also increase sales and customer loyalty. The recommendations outlined here offer a roadmap for achieving these goals and ensuring continued business growth.