Good evening, everyone. My name is Veena, and I am joined by my teammates—Gowtami, Likhita, Sushma, and Amith. We are excited to be here today for this case competition. Together, we will be presenting insights and recommendations for this case competition.

Marketing campaigns are a critical tool for businesses to engage customers and drive sales. However, not all customer segments respond equally to promotional efforts. Gourmet Haven is facing ongoing challenges, and refining its marketing strategies is key to boosting the effectiveness of its campaigns.

To achieve this, it is important to understand the factors that influence customer engagement and spending behavior. Our report seeks to identify key customer trends, improve targeting efficiency, and provide actionable recommendations to boost campaign effectiveness.

We look forward to sharing our insights and strategies with you.

Model accuracy

I WILL NOW HAND IT OVER TO AMITH...

### **1. Customer Response Rate by Household Composition-AMITH**

* Households **without children** have the **highest response rate (~27%)**.
* Response rates **decline** as the number of children increases.
* Households with **three or more children** have the **lowest response rate (~4.9%)**.

**Marketing recommendations:**

* Focus marketing efforts on **child-free households**, as they are more likely to engage.
* For households with children, consider offering **family-oriented promotions** and discounts tailored to their needs.

Add more columns like where this campaign took place how did the customer respond (email,store,)

Add a collum in discounts were there if they responded.

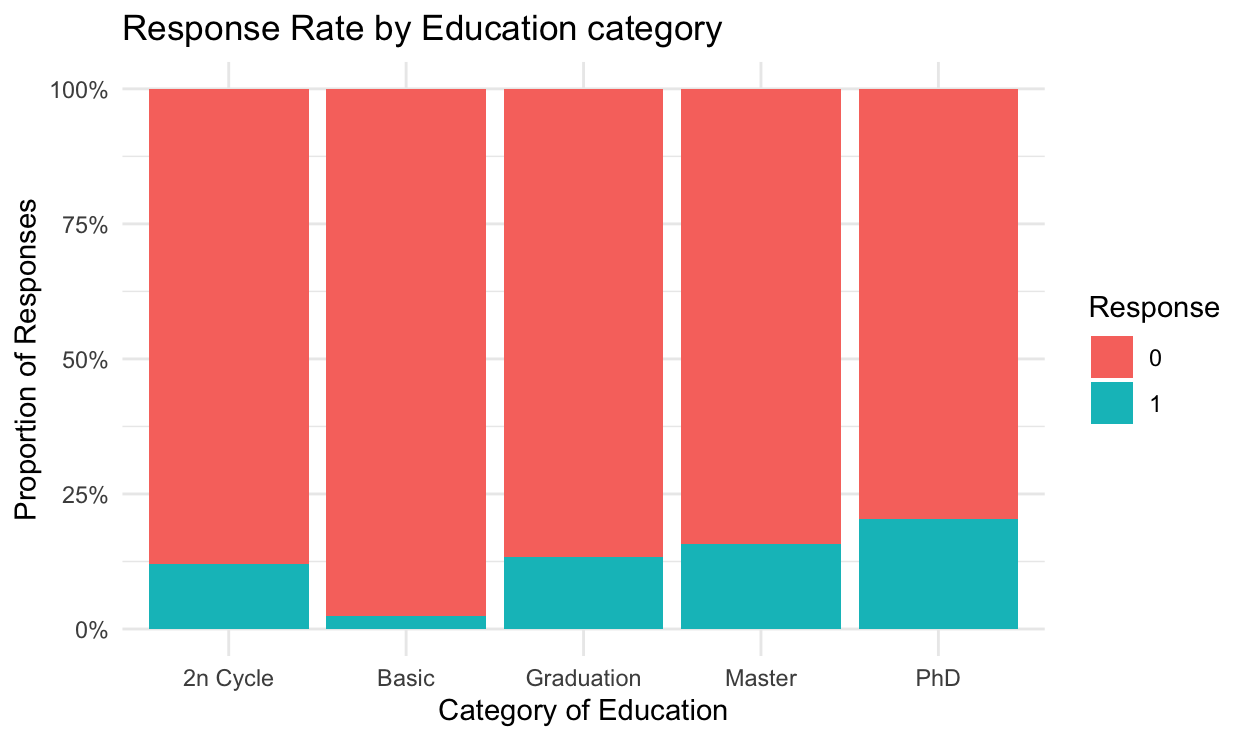
A graph with blue squares

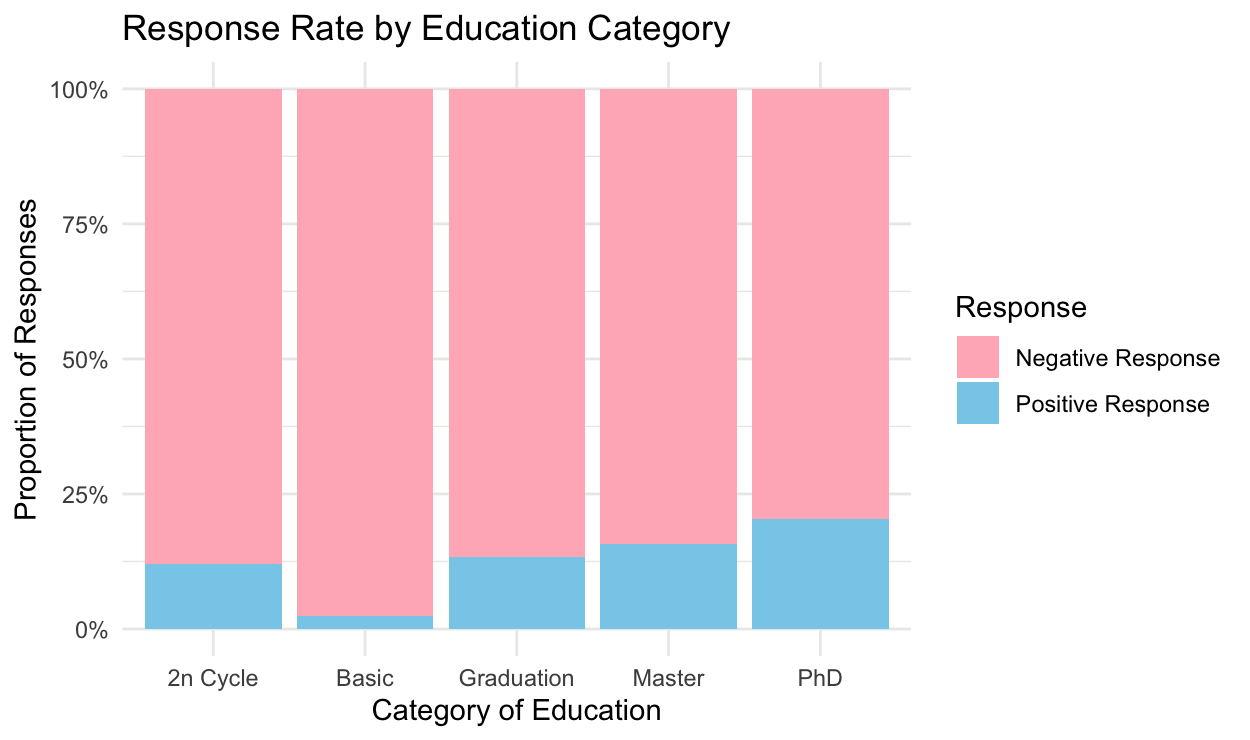
Description automatically generated

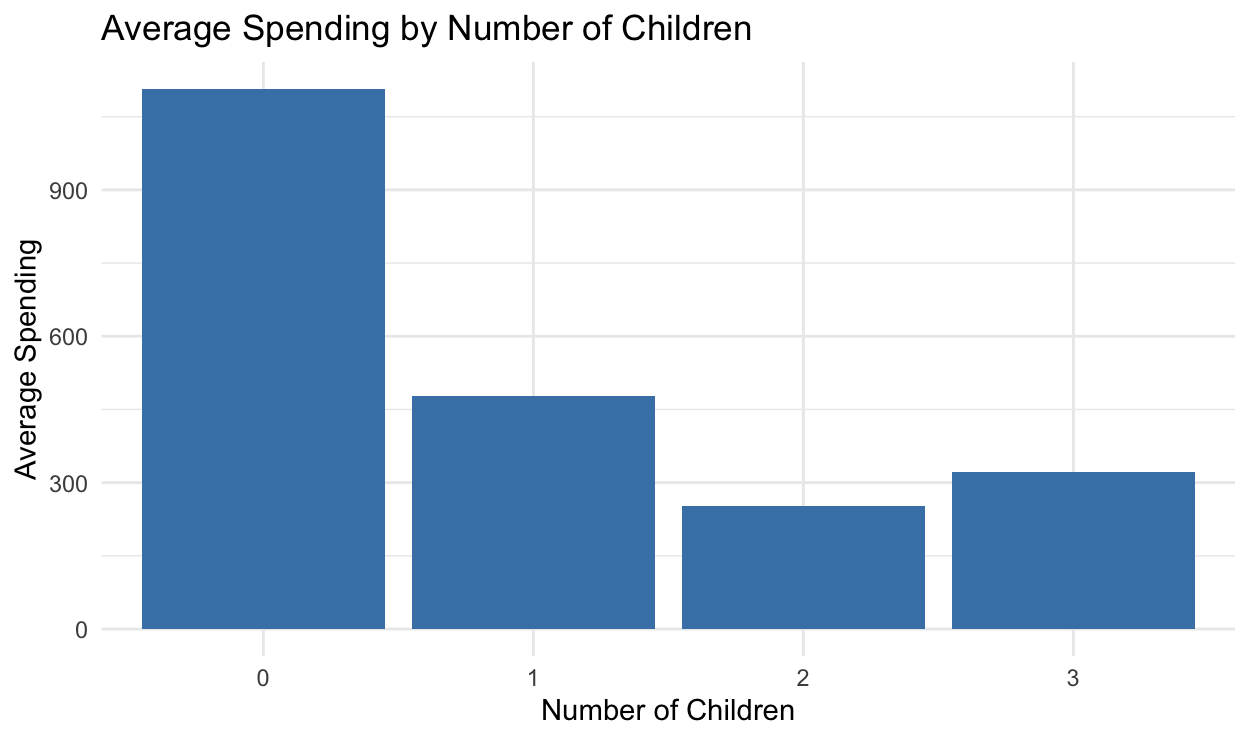
**2. Customer Response Rate by Education**

* \**07)**.
* Spending **significantly decreases** as the number of children increases.

**Marketing Implications:**

* Promote **premium and high-value products** to child-free households.
* Introduce **affordable pricing strategies, installment plans, or special discounts** to encourage spending among families with children.
* 





1. Households without children have higher response rates and higher spending.

2. Wine purchases strongly correlate with meat and other luxury products.

**Best Model: Lasso Regression**

Based on the metrics table, **Lasso Regression** is the best model as it has the highest **Accuracy (91.52%)**, **AUC (0.7129)**, **Precision (69.69%)**, and **F1-Score (54.76%)**.

**Predictions that Can Be Made Using Lasso Regression**

Since the model is trained to predict the **Response** variable (whether a customer will respond to a marketing campaign), we can use it to:

* **Predict customer response probability** for future campaigns.
* **Identify high-value customers** who are most likely to engage.
* **Optimize marketing strategy** by targeting customers with higher predicted probabilities.
* **Analyze feature importance** to determine which factors (e.g., total spending, number of web purchases) influence customer response.

**Evaluation Metric Used**

The model performance is assessed using:

* **Accuracy (91.52%)** – Measures the proportion of correct predictions.
* **AUC (0.7129)** – Measures the area under the ROC curve, showing how well the model distinguishes between positive and negative classes.
* **Precision (69.69%)** – Measures how many predicted positives were actually positive.
* **Recall (45.09%)** – Measures how many actual positives were correctly identified.
* **F1-Score (54.76%)** – Balances precision and recall.

**Generalization Approach**

To ensure that Lasso Regression generalizes well to unseen data, the model applies:

* Cross-validation (cv.glmnet) to select the best λ (penalty term) that balances bias and variance.
* Feature selection to reduce noise by eliminating irrelevant predictors.
* Normalization of numerical features to maintain consistent weight distribution.
* Train-test split to prevent overfitting and measure real-world performance.

**Bias-Variance Tradeoff in Lasso Regression**

* **Bias:** Lasso introduces some bias by shrinking coefficients, which reduces variance but increases error slightly.
* **Variance:** By eliminating unnecessary features, Lasso reduces model complexity and variance, making it more robust to new data.
* **Overall Tradeoff:** Lasso balances bias and variance well, making it ideal when there are many correlated features, as it avoids overfitting while preserving interpretability.

Thus, Lasso Regression is the best choice for **predicting customer response efficiently while maintaining simplicity and avoiding overfitting**.

## **Final Recommendations**

### **Marketing Strategy improvement:**

* **Target child-free households** with luxury and premium product promotions.
* **Leverage bundle discounts** (e.g., wine and meat pairings) to encourage cross-selling.
* **Implement personalized promotions** for families to boost engagement and spending.
* **Utilize predictive analytics** to refine and enhance marketing outreach efforts.

## **Conclusion and Next Steps**

### **Key Takeaways:**

* **Household composition significantly impacts response rates and spending habits.**
* **Product purchase correlations provide opportunities for targeted bundling and cross-selling.**
* **Lasso Regression is the best predictive model for customer targeting.**

### **Next Steps for Implementation:**

1. **Adjust marketing campaign focus** based on response rate and spending insights.
2. **Test bundled promotions and personalized offers** to evaluate effectiveness.
3. **Continuously track model performance and customer engagement trends** to refine strategies.

By integrating these insights into marketing strategies, businesses can enhance engagement, improve conversion rates, and maximize overall sales performance.

**End of Report**

**Model metrics:**

**Model Accuracy AUC Precision Recall F1\_Score**

**1 Logistic Regression 0.9107143 0.7103768 0.6571429 0.4509804 0.5348837**

**2 Random Forest 0.8950893 0.5990270 0.6111111 0.2156863 0.3188406**

**3 XGBoost 0.9107143 0.6676545 0.7200000 0.3529412 0.4736842**

**4 Lasso Regression 0.9151786 0.7128957 0.6969697 0.4509804 0.5476190**

**5 Decision Tree 0.8950893 0.6332049 0.5769231 0.2941176 0.3896104**

Sure! Here’s how you can expand your **data cleaning** and **feature engineering** process to include **SMOTE** (for handling class imbalance) and **SHAP** (for feature importance analysis):

### **1. Data Cleaning**

✔ **Handling Missing Values**

* Imputed missing values using the **mean** for numerical variables and the **mode** for categorical variables. This ensures no missing data disrupts model training.

✔ **Encoding Categorical Variables**

* Applied **one-hot encoding** to categorical features like Education, Marital\_Status, Complain, and Response to convert them into numerical format, which is suitable for model input.

✔ **Creating New Meaningful Features**

* Derived interaction terms and aggregated purchase behavior, such as Total\_Spent (sum of individual spending categories) and Total\_Purchases (sum of web, catalog, and store purchases), to provide more insightful predictors.

### **2. Feature Engineering**

✔ **Customer Segmentation Features**

* Created **age groups**, **income bands**, and **spending categories** to better capture purchasing behavior, grouping customers into meaningful segments for targeted marketing.

✔ **Behavioral Features**

* Engineered variables like **average purchase frequency**, **total spending**, and **recency of last purchase** to quantify customer activity, helping models identify patterns that could predict future purchases.

✔ **Derived Ratio Features**

* Created **spending-to-income ratio** to identify high-value customers who are more likely to engage in future purchasing behavior, enhancing the model’s ability to spot profitable customers.

### **3. Class Imbalance Handling**

✔ **SMOTE (Synthetic Minority Over-sampling Technique)**

* Applied **SMOTE** to balance the dataset by generating synthetic samples for the minority class in the Responsevariable. This helps to address class imbalance and prevent the model from being biased towards the majority class.
* ✅Highest AUC (0.7129) → Indicates best ability to separate classes.
* ✅ Better Generalization → Lasso regression reduces overfitting by selecting important features and shrinking irrelevant ones to zero.
* ✅ Interpretable Model → Feature selection makes it easier to understand the model's decisions compared to complex models like Random Forest or XGBoost.

**4. Model Interpretation & Feature Importance**

✔ **SHAP (SHapley Additive exPlanations)**

* Used **SHAP** values to interpret the model’s predictions and understand feature importance. SHAP values provide a way to explain how each feature contributes to the final prediction, helping to identify key factors driving customer behavior.

This step helps you interpret which features (e.g., Total\_Spent, Age, Income) have the most impact on predicting customer response. By analyzing SHAP values, you can better understand and explain how your model is making decisions.

### **1. Data Cleaning**

✔ **Handling Missing Values:**

✔ **Encoding Categorical Variables:** Education, Marital\_Status, Complain, and Response to convert them into numerical format

✔ **Creating New Meaningful Features:** Total\_Spent (sum of individual spending categories) and Total\_Purchases (sum of web, catalog, and store purchases)

### **2. Feature Engineering**

✔ **Customer Segmentation Features**

* Created **age groups**, **income bands**, and **spending categories** to better capture purchasing behavior, grouping customers into meaningful segments for targeted marketing.

✔ **Behavioral Features**

* Engineered variables like **average purchase frequency**, **total spending**, and **recency of last purchase** to quantify customer activity, helping models identify patterns that could predict future purchases.

### **3. Class Imbalance Handling**

✔ **SMOTE (Synthetic Minority Over-sampling Technique)**

* SMOTE was used to **oversample the minority class** and create a balanced dataset.
* This helped improve the model's ability to predict the minority class (e.g., Response = 1).

**4. Model Interpretation & Feature Importance**

✔ **SHAP (SHapley Additive exPlanations)**

* Used **SHAP** values to interpret the model’s predictions and understand feature importance.