**Strategic Customer Targeting for Credit Product Line**

Written Supplement

Michael Charara

11.2023

**Video:** <https://youtu.be/KcR3jcBAssE>

**[A close-up of a blue and purple text

Description automatically generated](https://youtu.be/KcR3jcBAssE)**

**Executive Summary:**

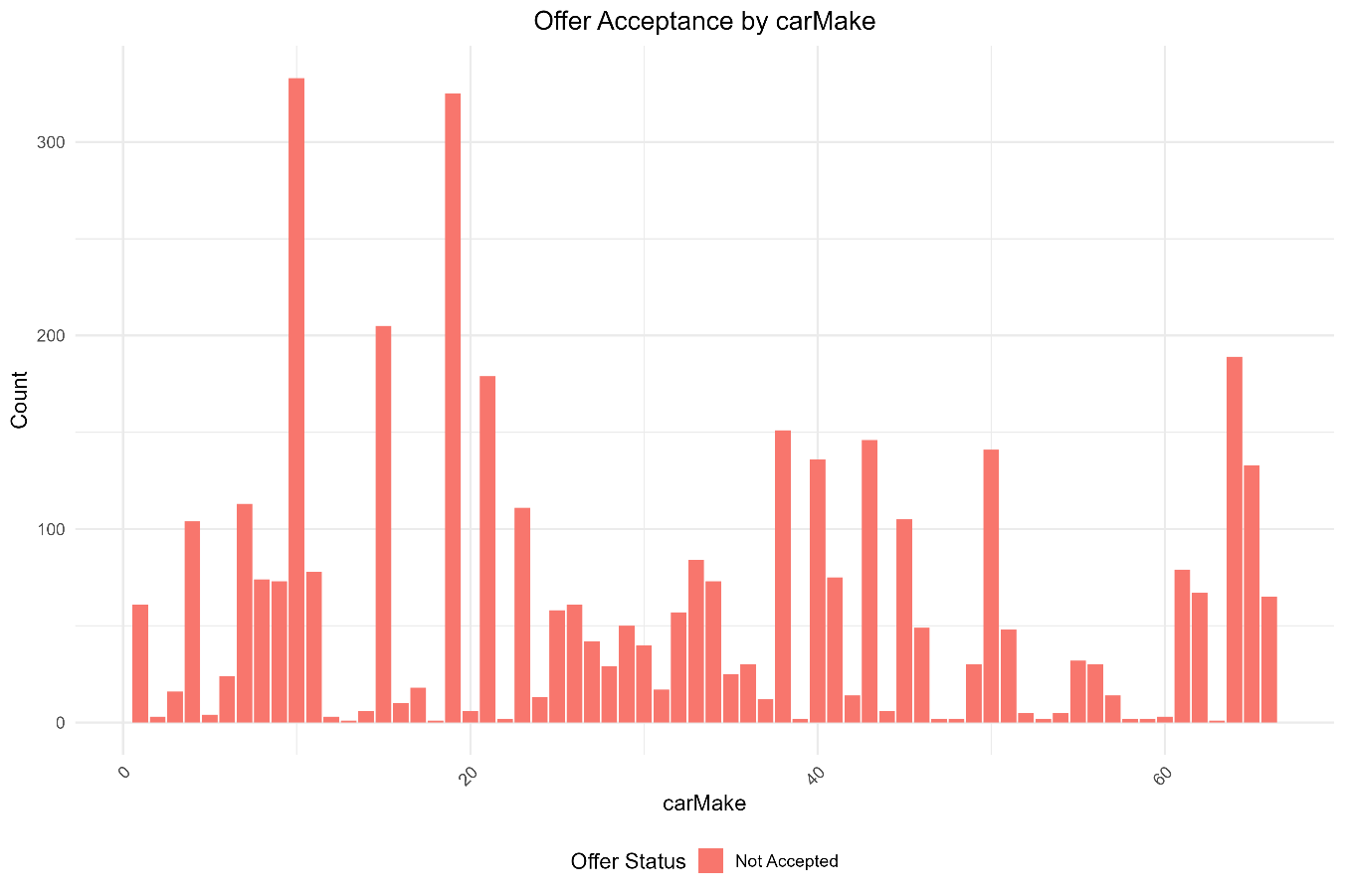
This report presents the efforts of National City Bank to strategically identify and engage potential customers for its new line of credit products. By applying advanced machine learning techniques, specifically Logistic Regression and XGBoost models, to historical marketing and customer data, we have determined a target list of the top 100 prospects with the highest likelihood of product uptake. The insights derived from this analysis will guide the bank's pilot marketing campaign, optimizing outreach efficiency and customer response rates.

**Methodology:**

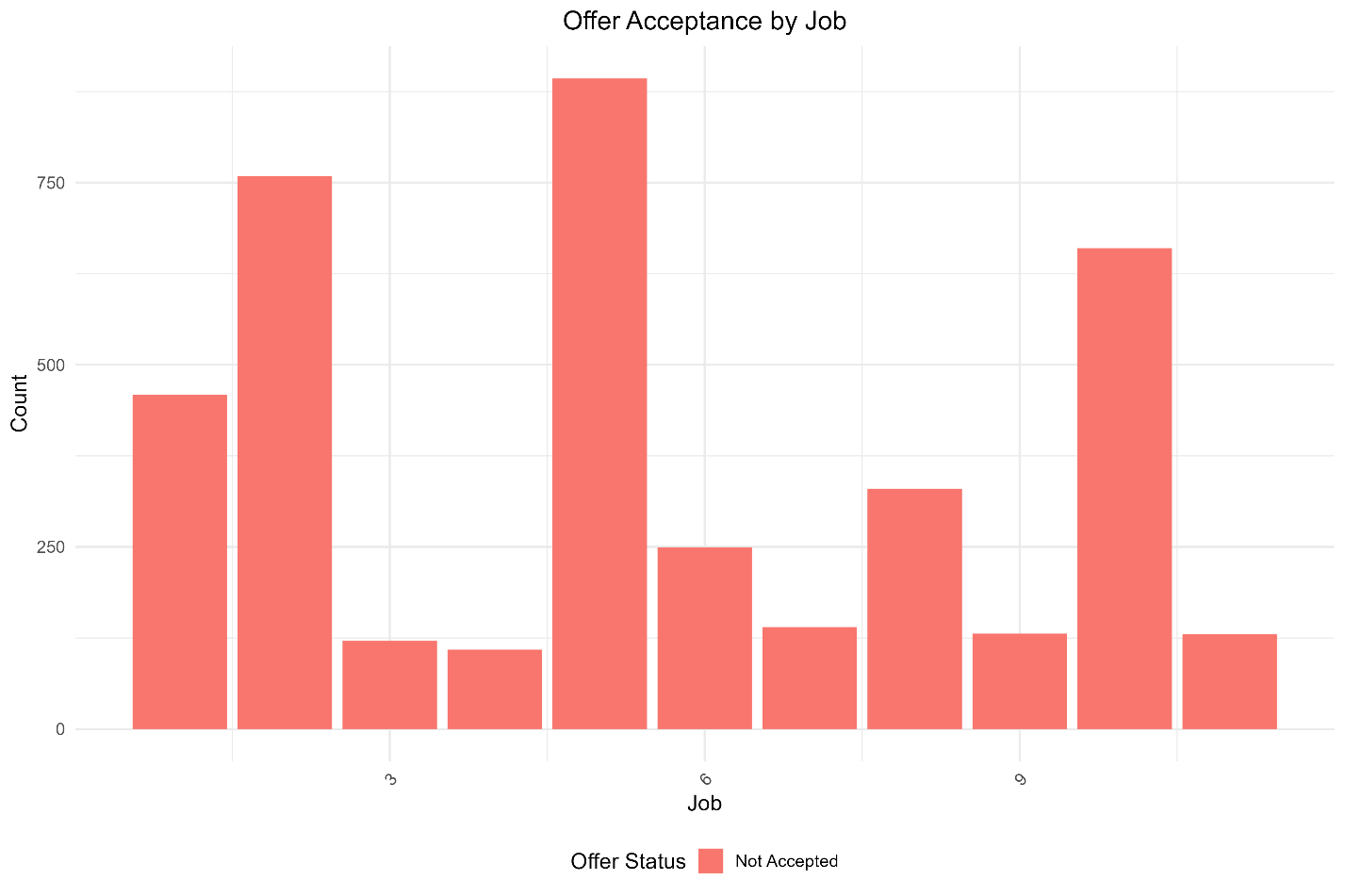
Our methodology involved a comprehensive data integration process, collating information from customer interactions and vehicle ownership records. After meticulous preprocessing, we engineered features that reflect customer profiles and past interactions. Two predictive models, Logistic Regression and XGBoost, were trained to estimate the propensity of customers to accept the new credit offer. The models were evaluated for accuracy and interpretability, with XGBoost providing a nuanced understanding of complex patterns and Logistic Regression offering clear insights into variable significance.

**Data Analysis and Exploratory Data Analysis (EDA):**

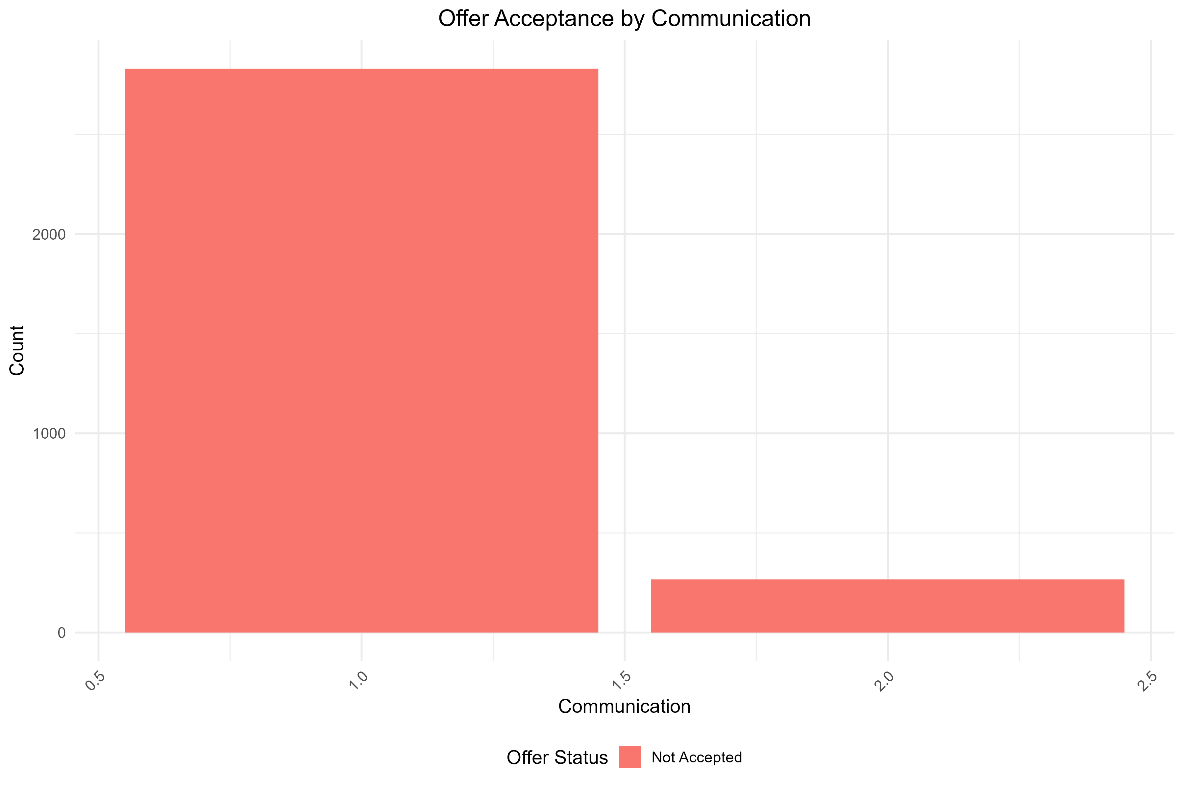
The EDA phase underscored the significance of variables like car make, job category, and communication frequency. Visualizations depicted a higher likelihood of offer acceptance among customers with specific educational backgrounds and marital statuses, guiding the feature engineering process.



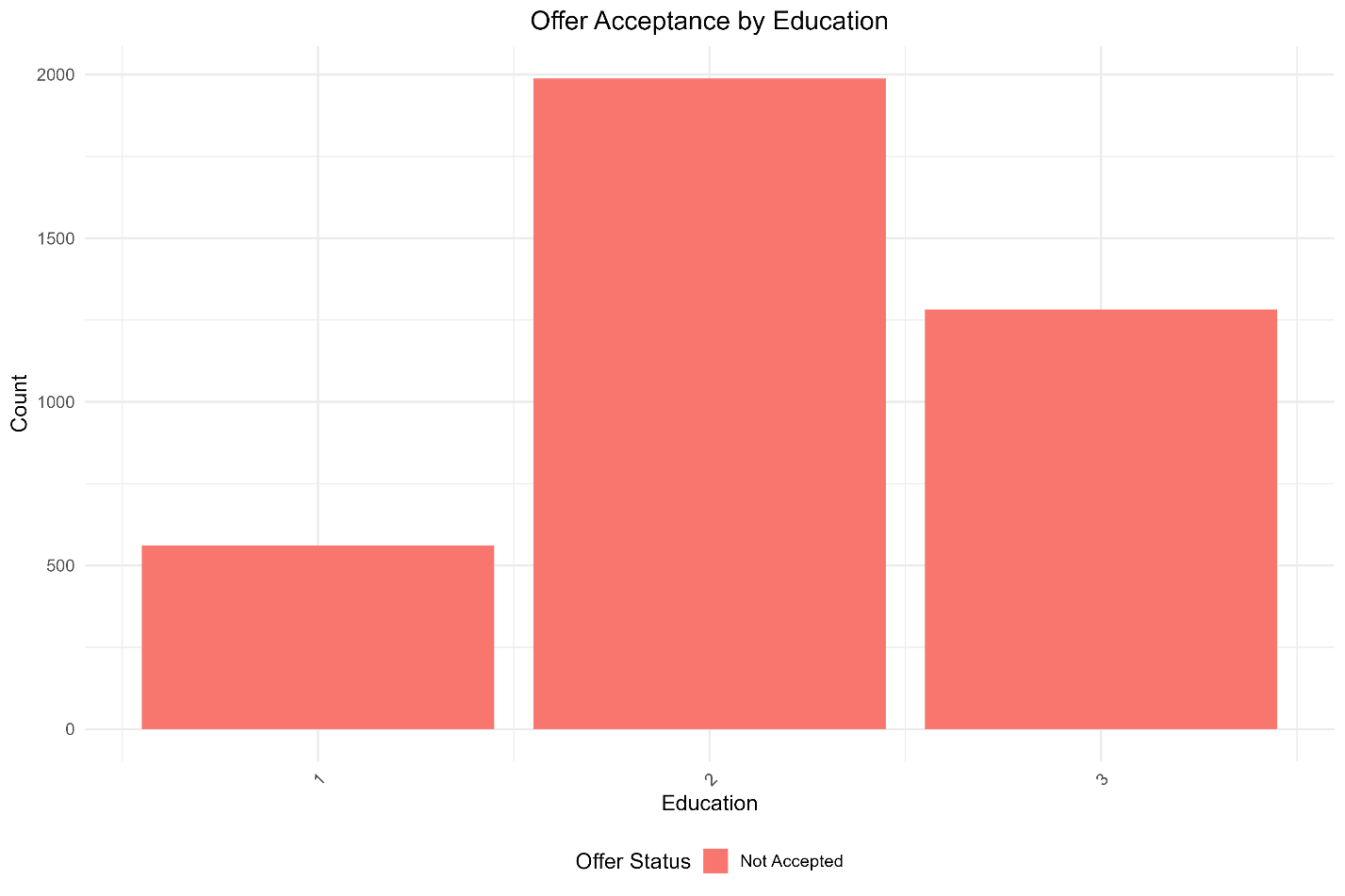
The bar chart for 'carMake' revealed varying frequencies of car brands within our customer base, indicating potential correlations with credit offer acceptance. A higher prevalence of specific car makes among those who accepted the offer suggests their relevance as predictors in our propensity model. This could be due to the market position of these brands, which may resonate with specific customer segments known for their openness to financial products. Additionally, the financial status of particular car brands could influence customers' decisions regarding new lines of credit, as ownership may reflect on their economic stability and purchasing power. Moreover, the perceived collateral value of different car makes, especially those that maintain their value over time, could impact the attractiveness of the credit offer.



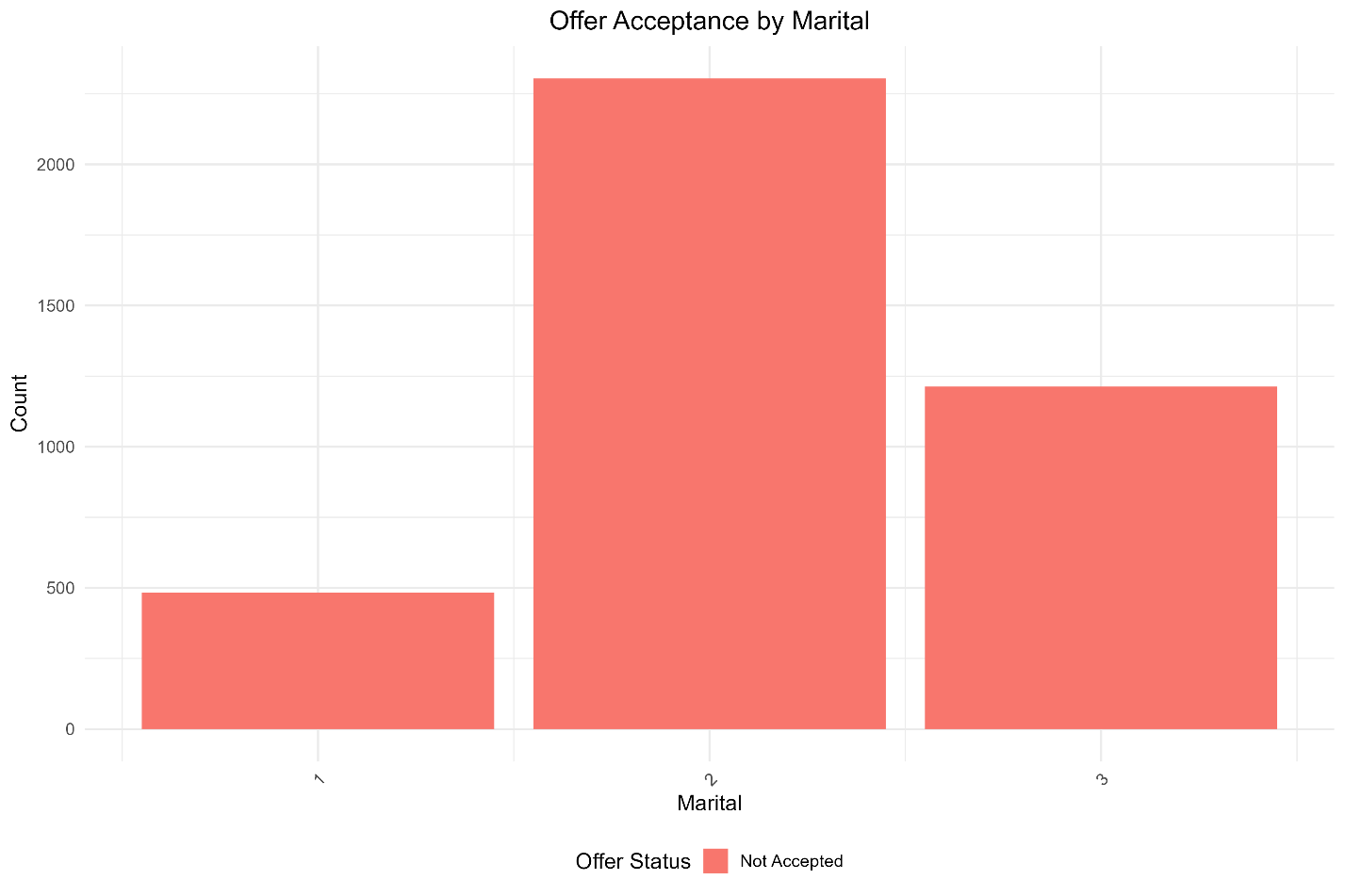
The distribution shown in the 'jobCategory' bar plot points to the importance of customers' professions in predicting offer acceptance. The analysis indicated that specific job categories are more likely to accept the credit offer, which could be attributed to the income levels associated with those professions. Higher-income jobs may increase the need for or the ability to manage additional credit. The stability of a customer’s profession also came into focus, as job security could enhance the confidence required to engage with new financial commitments. Lastly, the propensity to accept the credit offer might be influenced by industry-specific trends captured by the job categories to which customers belong.



The bar plot for 'Communication' outlined the frequency of interactions between the bank and its customers. This variable's significance is inferred from the pattern that customers in frequent contact with the bank tend to be more receptive to offers. Regular communication signifies higher customer engagement and ensures that customers are well-informed about the bank’s offerings. This constant information flow could significantly sway a customer’s decision-making process. Furthermore, the strength of the relationship built through ongoing communication may foster trust and influence customers' willingness to consider and accept new products.

****

The ‘Education’ bar chart likely showed the levels of education attained by customers who accepted the credit offer. If the data indicated that customers with specific educational backgrounds, such as a college degree or higher, were more likely to accept the offer, education level could be a strong predictor of offer acceptance. This may reflect the financial literacy or income potential associated with higher educational qualifications, which could influence an individual’s decision to purchase new financial products.

****

The bar plot for 'Marital Status' would have displayed the marital statuses of customers and how these statuses correlate with the acceptance of the credit offer. For example, a pattern where married customers show a higher likelihood of acceptance could imply that marital status is a significant variable. This could be due to various socioeconomic factors, such as combined household income or marriage stability, affecting the propensity to use credit products.

**Modeling Results:**

The results section details the outcomes from both the Logistic Regression and XGBoost models. While Logistic Regression facilitated an understanding of the influence of individual variables, the XGBoost model's ensemble approach captured intricate data relationships. A comparative analysis of the predictions from both models was conducted to consolidate the list of top prospects, ensuring robustness in our selection.

**Customer Profile Analysis:**

The combined analysis of both models paints a comprehensive picture of prospective customers. The report delves into the attributes that both models found predictive, such as communication preferences, household demographics, and vehicle details. This section discusses how the interplay of these factors was used to refine the model and enhance its predictive power.

**Business Implications and Recommendations:**

Building on the predictive insights, we recommend targeted marketing strategies that align with the identified influential factors. The XGBoost model highlights the complex interactions between customer attributes, suggesting a segmented approach to marketing communications. The bank can leverage these findings to craft personalized messaging that resonates with the identified customer segments.

**Conclusion:**

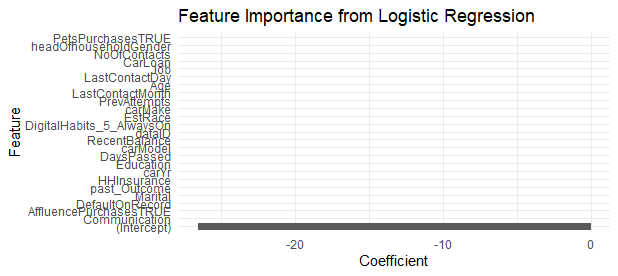
The propensity models have equipped National City Bank with a data-driven strategy for customer engagement. By utilizing Logistic Regression and XGBoost models, the bank can confidently approach its marketing campaign backed by a robust analytical foundation. The Logistic Regression model has illuminated the key factors that significantly influence a customer's likelihood to accept the line of credit offer. Features such as 'PetsPurchases', 'headOfHouseholdAge', and 'NoContacts' exhibit strong negative coefficients, suggesting that these factors may decrease the probability of offer acceptance. Conversely, features with positive coefficients were not prominent in this model, indicating that their influence might be more nuanced or require further investigation.

The prominence of 'PetsPurchases' as a strong negative predictor could indicate that customers who have recently made pet-related purchases might be less inclined to take on additional financial products. This could be due to the discretionary nature of such spending or an indication of different financial priorities. The age of the head of household appears to be another decisive factor, where certain age groups might not see the value in or need a line of credit against their vehicle.

These findings suggest a strategic pivot in the bank's customer segmentation and targeting approach. By understanding the traits that diminish the propensity to accept the offer, National City Bank can refine its marketing strategy to avoid fewer promising prospects and enhance the efficiency of its outreach.

Furthermore, the absence of strong positive influences in the feature importance graph suggests that the bank may benefit from exploring additional data sources or employing more complex models, like XGBoost, to capture other potential predictors of positive customer response. Therefore, while the Logistic Regression model provides clear insights into specific customer segments, it also underscores the need for a multifaceted analytical approach.

As National City Bank moves forward with its pilot marketing campaign, these data-driven insights will be crucial in shaping communication strategies and optimizing resource allocation, ensuring that the right customers are engaged with the right message at the right time.



**Code Explanation for National City Bank Project**

The National City Bank project analysis involved a series of R scripts, each designed to address specific aspects of the data analytics process. Central to this analysis were the scripts `ncb\_eda.r` for exploratory data analysis, `ncb\_logistic\_regression.r` and `ncb\_xgboost.r` for modeling. These scripts formed the backbone of our data preprocessing, exploratory analysis, and predictive modeling, ensuring a data-driven approach to identifying the top 100 prospects for the bank's new credit line product.

In the initial data preprocessing phase, essential libraries such as `xgboost`, `dplyr`, `readr`, and `ggplot2` were loaded. These libraries are essential for handling complex data operations, including data manipulation, machine learning, and visualization. The datasets loaded included `CurrentCustomerMktgResults.csv`, `householdAxiomData.csv`, `householdCreditData.csv`, `householdVehicleData.csv`, and `ProspectiveCustomers.csv`, each providing unique insights into the customers' profiles.

A pivotal step in the preprocessing was the merging of these datasets. This was achieved using the `HHuniqueID` field as a shared key, allowing a comprehensive view of each customer's information by combining data from multiple sources. This integrative approach allowed us to build a dataset that provided a complete view of the customers.

Feature engineering, a critical step in preparing the data for modeling, involved transforming categorical variables such as `Communication`, `LastContactMonth`, `past\_Outcome`, `headOfhouseholdGender`, `EstRace`, `Job`, `Marital`, `Education`, `carMake`, and `carModel` into numeric factors. This transformation was imperative for the application of machine learning algorithms. Additionally, the target variable `Y\_AcceptedOffer` was converted into a binary format to fit the needs of binary classification models.

The `ncb\_eda.r` script was instrumental in conducting exploratory data analysis. It created a series of bar plots for each categorical variable, visually depicting their frequency and relationship with the offer acceptance. This EDA process was essential in identifying key variables that significantly impact customer responses.

Regarding model training, the logistic regression and XGBoost scripts focused on predicting customer behavior. The XGBoost model was trained using parameters optimized for binary classification, and its performance was evaluated based on its accuracy in predicting customer responses.

One of the most critical outcomes of the XGBoost modeling script was the identification of the top 100 prospects. This was achieved by scoring each customer in the `ProspectiveCustomers.csv` dataset based on their likelihood of accepting the offer. The selection of the top 100 was thus data-driven, guided by the model's prediction scores, ensuring a focus on those most likely to respond positively.

Lastly, the variable importance analysis within the XGBoost script shed light on the most influential factors in the model's predictions. This insight was vital for understanding the driving forces behind customer decisions and refining the bank's marketing strategies.

In conclusion, the code used in this project was not just a series of scripts but a well-orchestrated analytical process. It involved meticulous data merging using `HHuniqueID`, careful feature engineering, insightful exploratory analysis, and strategic modeling, converging to identify the top 100 prospects for the bank's new credit product.

**References:**

Data Source: <https://www.kaggle.com/kondla/carinsurance>

Additional Resource: <https://www.kaggle.com/kondla/simple-random-forest-on-insurance-call-forecast/code>

R-Studio Copyright (C) 2022 by Posit Software, PBC  
R packages (xgboost, dplyr, readr, ggplot2)