# **Identifying Key Characteristics of Potential Customers Likely to Sign Up for a New Credit Card**

Maxim Bilenkin

Department of Science and Technology, Bellevue University - Bellevue

DSC 630: Predictive Analytics

Professor Toni Farley

August 2, 2025

# **IDENTIFYING KEY CUSTOMER CHARACTERISTICS**

**Introduction of Topic/Problem**

In the 21st century, the banking industry became so saturated and competitive that it is difficult to acquire new customers. In most cases, if a person opens a new bank account or owns one or few credit cards, then it is hesitant to switch one for a significant period. According to bankrate.com (see last page for reference) many consumers stick with the same bank account for decades, if the bank continues to deliver convenience. Frequently, in most cases if a customer opens bank accounts, either checking or saving, then it is most likely to acquire credit card from the same bank and hesitant to switch or acquire a new one. However, loyalty to credit cards might be more volatile. If there are lower rates, better reward programs and personalized experiences offered, then many customers are willing to switch to another credit card issuer.

Nevertheless, today banks have become very competitive and upgraded their services to offer great services and experiences to their clients, so great services most likely are not a decisive and the most effective key factor that will make customers sign up for a new credit card.

The goal of this project is to identify the key characteristics of potential customers that are likely to sign up for a new credit card. Banks and financial companies use various marketing strategies to acquire new credit card customers. To name a few, direct mail, cold phone calls, internet advertising, and television advertising. But the cost of marketing can be very expensive and might not be justified to pursue for economic reasons. According to [Bankrate (2019)](https://www.bankrate.com/credit-cards/news/one-third-american-credit-cardholders-never-changed-cards-ccdc/), millions of Americans have never changed their primary credit card. According to [Bankrate (2019)](https://www.bankrate.com/credit-cards/news/one-third-american-credit-cardholders-never-changed-cards-ccdc/), 30% of card holders have never switched their primary cards as of March 2019.

**IDENTIFYING KEY CUSTOMER CHARACTERISTICS**

However, 35% have switched their most used cards in the past three years. As indicated by the research, of those customers who hold their cards for a long time, 14 % switched during the last three and five years ago. 12% switched during the last five and ten years ago. 9% switched more than ten years ago, and 30% had never switched their credit cards.

Looking at these statistics, we can quickly see that it will not be an easy task for financial institutions to acquire new credit card customers. However, there might be vast amount of people who never had credit cards in their life and want to acquire one or perhaps the one who are willing to switch one, who are those people and what kind of key characters they have. If financial institutions can answer all these questions, then they can use customized marketing strategies specifically target groups of customers that are more likely to sign up for new credit cards. By using this strategy, companies can lower expenses, increase revenue, and improve their earnings.

**Overview of Data used**

For this specific project, the best dataset that would address the goal of this model would be to use credit card marketing campaign data that contain demographic variables, such as age, marital

# **IDENTIFYING KEY CUSTOMER CHARACTERISTICS**

status, job, defaulted or not in the past, final result (yes) or (no) subscribed and many other variables that would give description of clients.

Since it was not possible to find the perfect dataset. It was decided to use data from direct marketing phone call campaign conducted by Portuguese banking institution. The classification goal of this marketing campaign was to predict if a client will subscribe to the bank term deposit, which was designated with variable (y = yes). Although, credit card and term deposit are completely different and opposite financial instruments. Nevertheless, the data contains many attributes that are needed for this project, making it a good fit. Also, companies often use data from different products for their model to forecast response for different financial products because if customers response was positive for one product, then it might be the same response for other products. From the economic point view, the term deposit data might be a good fit for this project because signing up for a new credit card can give customers the ability to postpone immediate cash payment until the end of a month and at the same time cash can be deposited in bank deposits earning interest. This justifies the use of the term deposit data.

In many cases, multiple calls needed to be made to the same client to assess if the product (bank-term deposit) will be (yes = 1) or (no = 0) subscribed. The data contains variables like age, job, marital status, education, default, balance, housing, loan, and target variable “y” with result of

# **IDENTIFYING KEY CUSTOMER CHARACTERISTICS**

campaign that has binary outcome (y) or (n) subscribed, which stands for yes or no respectively. The dataset contains 45,211 rows and 17 columns. The 7 columns are numerical and contain variables like age, balance, duration. The remaining 10 columns are categorical like job, education, and contact. There were no null or missing values. However, several fields contained the value “unknown”. Specifically, the following four variables like “job” ('admin.', 'blue-collar',

'entrepreneur', 'unemployed) had 288 “unknown” count, “education” ('basic.4y', 'basic.6y', 'basic.9y', 'high school', 'university degree') had 1,857, “contact” (contact communication type like 'cellular', 'telephone') had 13,020, and “poutcome” (outcome of previous marketing campaign: 'failure', 'nonexistent', 'success') 36,959. For this project, the value “unknown” was treated as separated group because it would be impossible to identify what was the value for the variable.

**Methods of Analysis**

**Data Preprocessing & Cleaning**

The data for this project had no missing value. However, as mentioned previously, there were value “unknown”, and it was decided to keep it as separate group since it would be impossible to identify what should be the correct value instead. One column named with feature “duration”

# **IDENTIFYING KEY CUSTOMER CHARACTERISTICS**

was dropped to avoid data leakage. The variable “duration” described in seconds how long the phone call with the customer lasted in seconds and had value outcome either “0” or “1” for target variable “y”. If duration was “0”, then it means unsuccessful, and client didn’t sign up for bank term deposit. If it was “1”, then it means success, which means client sign up for term deposit. Obviously, such information would be known after the marketing campaign but not before. Thus, it was dropped to avoid data leakage and to have realistic predictive model. Also, this data didn’t have duplicates.

Out of 17 columns, 7 columns are numerical like age, account balance, duration during last call, and 10 are categorical describing job, education level, contact by cellular or telephone. The numerical features were standardized using StandardScaler to improve the performance of machine learning algorithms. Scaling was applied after the train/test split to avoid data leakage.

**Train-Test Split**

For this project, the stratified sampling method was used to maintain the same proportion of positive cases in both sets because the data is highly imbalance, there are substantially more no (not subscribed) than yes (subscribed) for target variable “y”. The data was split into two parts, a training set comprising 80% and a test set comprising 20%.

# **IDENTIFYING KEY CUSTOMER CHARACTERISTICS**

**Feature Selection**

All categorical features were kept because they described an individual and can provide key factors about an individual that is more likely to sign up for a credit card. For example, job, marital, education, default, housing, loan, contact, month, and poutcome are all important features that can help to identify key features of an individual who are more likely to sign up for a new credit card. From beginning many assumptions can be made like a person with low pay job, single, no college degree, had default in the past on credit loan, rents instead of owning, has

some other loan to pay, would probably more likely to sign up for a new credit card because one doesn’t have good paying job, alone without partner, not earning enough because no college degree, and had defaulted in the past, which most likely indicate the person needs money and credit card could provide support. However, it is important to note that common assumptions can often be misleading. A data-driven approach is essential to uncover patterns that may not align with initial intuition. The goal of this project is to find those key factors that have influence on one to sign up for credit card backed by findings. As previously mentioned, “duration” is the only categorical feature that was dropped to avoid data leakage.

# **IDENTIFYING KEY CUSTOMER CHARACTERISTICS**

**Modeling Technique**

Logistic Regression model was used for this project as primary classification model. This is a commonly used model when dealing with binary classification tasks like this one, where the goal is to predict one of two outcomes, whether a customer will subscribe to term deposit or not. Also, the Logistic Regression model was chosen due to its simplicity, interpretability, and ease of implementation. However, because the target variable is imbalanced – with only 11% positive responses – it is expected it might struggle to precisely detect the minority class without further tuning. Additional adjustments like splitting data into two parts, training and testing sets using stratified sampling, numerical features were standardized using StandardScaler.

**Evaluation Metrics**

The model was trained on the training set and evaluated on the test set and metrics like accuracy, precision, recall, F1 score, and ROC AUC were used for evaluation. The use of recall and AUC were specifically important to be used for evaluation because recall would show how the model captured as many likely subscribers as possible and ROC AUC is a robust metric to evaluate performance of our binary classification model, when dealing with imbalance dataset.

# **IDENTIFYING KEY CUSTOMER CHARACTERISTICS**

**Visualization**

Various types of plots were used to evaluate model performance. These include a confusion matrix and a ROC AUC curve for logistic regression classifier. In addition, exploratory data analysis (EDA) was conducted by inspecting the first few rows of the dataset and using bar charts to understand feature distributions and relationships.

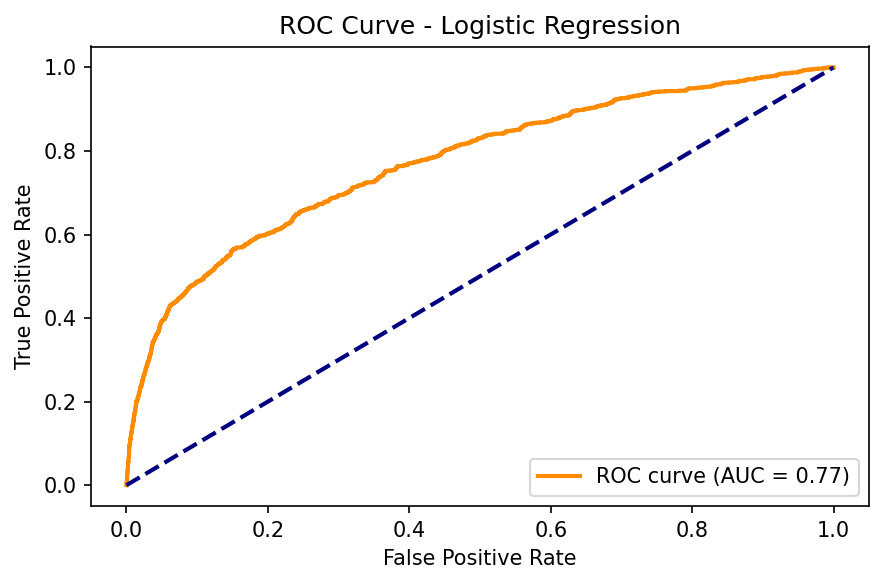
A blue squares with white text

AI-generated content may be incorrect.

***Figure 1***

Confusion matrix for the logistic regression classifier.

# **IDENTIFYING KEY CUSTOMER CHARACTERISTICS**



***Figure 2***

Receiver operating characteristic (ROC) curve showing discrimination performance.

# **IDENTIFYING KEY CUSTOMER CHARACTERISTICS**

**A graph of a number of people

AI-generated content may be incorrect.**

***Figure 3***

Distribution of job by target variable showing higher subscription rates among students.

**Results & Findings**

This Logistic Regression model was trained to predict if after the marketing campaign a customer will subscribe to term deposit based on various demographic and behavioral features. For this project dataset from Portuguese Banking marketing campaign for term deposit product was used, which is also serves as a proxy for signing up for a new credit card because it’s possible that customers who signed up for term deposit might sign up for new credit card. These two different products might be complementary indeed because a person who decides to sign up for a term loan must give away for some time a certain amount of money and by signing up for a new credit card one can alleviate the shortage of cash that it might need in case of emergency spending.

# **IDENTIFYING KEY CUSTOMER CHARACTERISTICS**

The model achieved the following results on the test set:

Accuracy: 75.53%

Precision: 26.70%

Recall: 62.57%

F1 Score: 37.43%

ROC AUC: 77.22%

These results indicate that while model is moderately good at distinguishing between customers who are likely to subscribe and those who are not, as shown by the ROC AUC score of 77.22%,

its precision is not impressive and relatively low, which means high number of false positives that are predicted as “yes”, were “no”. Nevertheless, the model achieved a relatively strong recall of 62.6%, suggesting the model is capturing a significant portion of the actual “yes” customers correctly. The confusion matrix confirms this trade-off. The model correctly predicted 662 subscribers, but also falsely predicted 1,817 non-subscribers as subscribers. At the same time, it correctly identified 6,168 non-subscribers.

**Recommendations/Conclusion**

Although this model didn’t have impressive performance, it can be still used for broad targeting campaigns. Given its high recall, the model can be useful in broad campaigns where identifying as many potential customers as possible is more important than minimizing false positives like sending promotional emails, ads, or trial offers. The model could be paired with cost-benefit analysis. Since false positives can incur significant costs like contacting uninterested customers, marketing teams should always consider from an economic standpoint where the campaign cost against the potential revenue per converted customer is below and worth pursuing. Other feature engineering or new interaction terms may also improve model performance. Marketing

# **IDENTIFYING KEY CUSTOMER CHARACTERISTICS**

team should consider adjusting the classification threshold to improve precision, depending on the marketing goal, like targeting fewer but more likely customers who will sign up for a credit card.

An important observation from the data shows that students had a subscription success rate of approximately 50% - the highest among all job categories. Although students made up only about 1,000 out of 45,211 total records, this group responded very well to marketing efforts. This suggests that future credit card marketing campaigns may benefit from specifically targeting student populations, such as through university partnerships, student credit-building programs, or online campaigns aimed at young adults entering financial independence. Targeting this niche group could yield high conversion rates with relatively low effort and cost.

**Reference**

*Bankrate*. (2019, March 10). Survey: Many consumers stick with same bank accounts for decades, cite convenience as a factor.<https://www.bankrate.com/banking/checking-fees-survey/>

*Bankrate*. (2019, March 9). Millions of Americans have never changed their primary credit card.<https://www.bankrate.com/credit-cards/news/one-third-american-credit-cardholders-never-changed-cards-ccdc/>