Project

Proposal

Retaining Customers in the Telecommunications

Industry: A Data-Driven Approach

Team 4

Shahbaz Aslam

Haddyjatou Ndimbalan

Gauravi Bendre

Maria Molina

**Business Problem:**

What are the key drivers of customer churn in a telecommunications company in California, and how can the company leverage these insights to develop targeted retention strategies and reduce churn?

Customer churn, or the rate at which customers leave a company, is a major concern for businesses, especially in the telecommunications industry where customer acquisition can be expensive and competitive. Customer acquisition cost (CAC) for communication service providers (CSPs), known alternatively as Telecoms, is approximately $315 in the U.S. [[1]](#footnote-1)(Forbes, 2020). Therefore, CSPs aim to maximize the lifetime value of each subscriber, which represents the total amount of revenue that the customer is expected to generate over the entire duration of their relationship with the CSP. Reducing customer churn can help CSPs improve customer satisfaction, increase customer loyalty, reduce costs such as CAC, and increase profits.

Reducing customer churn is a complex challenge that can impact a wide variety of stakeholders. CSPs, themselves, have a personal stake in reducing churn since by increasing customer retention they can also improve the company’s reputation and increase revenue which can be allocated into new areas like the upgrade of their infrastructure, expanding customer base, investing in R&D, etc. Likewise, shareholders are also interested in this initiative as it can improve the company’s financial performance, leading to increased revenue and profitability. On the other hand, competitors have an interest in reducing customer churn to gain a competitive advantage and attract new customers. Finally, regulatory bodies may be concerned about this business problem because they must ensure that all customers have access to reliable and high-quality telecom services, they can do that by ensuring service quality improves and promoting fair competition between CSPs.

The primary audience for the results of the analysis is the telecommunications company in California that is seeking to reduce customer churn and retain its customer base, specifically, the company's management team. The management team will use the insights gained from the analysis to make data-driven decisions. That said, other internal stakeholders that may receive this information will be customer service representatives and marketing teams who will be able to create a better targeted and effective experience for the customer and marketing strategies.

**Data sources:**

The data used in this paper was obtained from the website Maven Analytics[[2]](#footnote-2), a data analytics and visualization platform. The source of the data is listed as IBM Cognos Analytics. The Telecom Customer Churn dataset has been designed to provide a realistic simulation of the key factors that drive customer churn in a telecommunications company in California.

**Variables in the Telecom Customer Churn dataset:**

|  |  |
| --- | --- |
| Field | Description |
| CustomerID | A unique ID that identifies each customer |
| Gender | The customer’s gender: Male, Female |
| Age | The customer’s current age, in years, at the time the fiscal quarter ended (Q2 2022) |
| Married | Indicates if the customer is married: Yes, No |
| Number of Dependents | Indicates the number of dependents that live with the customer (dependents could be children, parents, grandparents, etc.) |
| City | The city of the customer’s primary residence in California |
| Zip Code | The zip code of the customer’s primary residence |
| Latitude | The latitude of the customer’s primary residence |
| Longitude | The longitude of the customer’s primary residence |
| Number of Referrals | Indicates the number of times the customer has referred a friend or family member to this company to date |
| Tenure in Months | Indicates the total amount of months that the customer has been with the company by the end of the quarter specified above |
| Offer | Identifies the last marketing offer that the customer accepted: None, Offer A, Offer B, Offer C, Offer D, Offer E |
| Phone Service | Indicates if the customer subscribes to home phone service with the company: Yes, No |
| Avg Monthly Long Distance Charges | Indicates the customer’s average long distance charges, calculated to the end of the quarter specified above (if the customer is not subscribed to home phone service, this will be 0) |
| Multiple Lines | Indicates if the customer subscribes to multiple telephone lines with the company: Yes, No (if the customer is not subscribed to home phone service, this will be No) |
| Internet Service | Indicates if the customer subscribes to Internet service with the company: Yes, No |
| Internet Type | Indicates the customer's type of internet connection: DSL, Fiber Optic, Cable (if the customer is not subscribed to internet service, this will be None) |
| Avg Monthly GB Download | Indicates the customer’s average download volume in gigabytes, calculated to the end of the quarter specified above (if the customer is not subscribed to internet service, this will be 0) |
| Online Security | Indicates if the customer subscribes to an additional online security service provided by the company: Yes, No (if the customer is not subscribed to internet service, this will be No) |
| Online Backup | Indicates if the customer subscribes to an additional online backup service provided by the company: Yes, No (if the customer is not subscribed to internet service, this will be No) |
| Device Protection Plan | Indicates if the customer subscribes to an additional device protection plan for their Internet equipment provided by the company: Yes, No (if the customer is not subscribed to internet service, this will be No) |
| Premium Tech Support | Indicates if the customer subscribes to an additional technical support plan from the company with reduced wait times: Yes, No (if the customer is not subscribed to internet service, this will be No) |
| Streaming TV | Indicates if the customer uses their Internet service to stream television programing from a third party provider at no additional fee: Yes, No (if the customer is not subscribed to internet service, this will be No) |
| Streaming Movies | Indicates if the customer uses their Internet service to stream movies from a third party provider at no additional fee: Yes, No (if the customer is not subscribed to internet service, this will be No) |
| Streaming Music | Indicates if the customer uses their Internet service to stream music from a third party provider at no additional fee: Yes, No (if the customer is not subscribed to internet service, this will be No) |
| Unlimited Data | Indicates if the customer has paid an additional monthly fee to have unlimited data downloads/uploads: Yes, No (if the customer is not subscribed to internet service, this will be No) |
| Contract | Indicates the customer’s current contract type: Month-to-Month, One Year, Two Year |
| Paperless Billing | Indicates if the customer has chosen paperless billing: Yes, No |
| Payment Method | Indicates how the customer pays their bill: Bank Withdrawal, Credit Card, Mailed Check |
| Monthly Charge | Indicates the customer’s current total monthly charge for all their services from the company |
| Total Charges | Indicates the customer’s total charges, calculated to the end of the quarter specified above |
| Total Refunds | Indicates the customer’s total refunds, calculated to the end of the quarter specified above |
| Total Extra Data Charges | Indicates the customer’s total charges for extra data downloads above those specified in their plan, by the end of the quarter specified above |
| Total Long Distance Charges | Indicates the customer’s total charges for long distance above those specified in their plan, by the end of the quarter specified above |
| Total Revenue | Indicates the company's total revenue from this customer, calculated to the end of the quarter specified above (Total Charges - Total Refunds + Total Extra Data Charges + Total Lond Distance Charges) |
| Customer Status | Indicates the status of the customer at the end of the quarter: Churned, Stayed, or Joined |
| Churn Category | A high-level category for the customer’s reason for churning, which is asked when they leave the company: Attitude, Competitor, Dissatisfaction, Other, Price (directly related to Churn Reason) |
| Churn Reason | A customer’s specific reason for leaving the company, which is asked when they leave the company (directly related to Churn Category) |
| Zip Code | The zip code of the customer’s primary residence |
| Population | A current population estimate for the entire Zip Code area |

**Data quality concerns:**

Since the data is fictional, it is possible that the data was generated randomly or artificially, which could result in biases or inaccuracies in the data. There are several data quality concerns that should be taken into account when using it for analysis or modeling:

* Validity: While the fictional data is sourced from IBM, a reputable organization, it is still valid data which may contain inherent biases, as is common with all datasets.
* Completeness: It is possible that some variables or observations may be missing from the dataset, which could result in incomplete or inaccurate results from any analysis or modeling performed using the data.
* Consistency: There may be inconsistencies or discrepancies in the data that could affect the accuracy of any analysis or modeling performed using the data.
* Imbalanced Data: The dataset has around 30% of the records that are related to churn. This is good for the prediction model development in the churn prediction cases.

Note: CSPs lack the capability to measure the customers’ quality experience as it cannot detect the increasing frustration levels as subscribers reach their “churning” point. For example, in many cases customers change providers without showing previous a sign of dissatisfaction.

**Potential methods:**

This study aims to use several machine learning models, including but not limited to **logistic regression, decision tree and SVM** to perform predictive modeling techniques on the dataset. The goal is to identify customers who are most likely to churn based on key factors, including demographics, services, tenure, and location. The following step after gaining the data insights from the analysis, is to develop a model that can accurately predict which customers are most likely to churn, and thus, help the telecom operator in California to take proactive measures to retain their customers. Additionally, a comparison of **the performance of different machine learning models to determine which one provides the best prediction will be run**. Other than churn prediction, other potential insights include:

* Identifying the impact of tenure and the likelihood of churn
* How the services (like phone and internet services) which the customer has subscribed to impact their stay
* Identifying if there is any area (Zip code) where there is a particular high concentration of customer churn
* Customer clusters based on demographics

**Preliminary results:**

Not yet.

**Challenges:**

Not yet.

**Team members’ responsibilities:**

* Gauravi Bendre: Data Wrangling and Quality Assurance Lead - Responsible for overseeing the end-to-end process of data preparation, transformation, and validation to ensure high-quality data inputs for analysis and insights.
* Shahbaz Aslam: Machine Learning Architect - Leads the end-to-end lifecycle of designing, implementing, and deploying machine learning models to drive actionable insights and data-driven decision making.
* Haddyjatou Ndimbalan: Data Artisan- transform complex data analysis results (patterns and trends) into visually compelling and actionable insights in a clear and concise manner.
* Maria Fernanda Molina: Business strategy Architect– Spearhead the alignment of business goals with data-driven insights, and craft compelling narratives that drive informed decision-making through data storytelling.

1. Forbes. (2020, October 30). Acquiring Subscribers Is Only Half The Battle. <https://www.forbes.com/sites/forbestechcouncil/2020/10/30/acquiring-subscribers-is-only-half-the-battle/?sh=33fa49df1821> [↑](#footnote-ref-1)
2. Maven Analytics. (n.d.). Data Playground. Retrieved February 7, 2023, from <https://www.mavenanalytics.io/data-playground> [↑](#footnote-ref-2)