July 7th, 2025

Ryan Webber

011889933

Choice Mobile Churn Prediction Application

Part A: Letter of Transmittal 2

Part B: Project Proposal Plan 4

[Project Summary 4](#_Toc904507251)

[Data Summary 4](#_Toc1736393957)

Implementation 5

[Timeline 6](#_Toc1357178365)

[Evaluation Plan 7](#_Toc623361460)

[Resources and Costs 8](#_Toc1538507987)

[Part D: Post-implementation Report 10](#_Toc651895932)

[Solution Summary 10](#_Toc1134136520)

[Data Summary 10](#_Toc182221765)

[Machine Learning 12](#_Toc1505466430)

[Validation 15](#_Toc391434166)

[Visualizations 16](#_Toc201059345)

[User Guide 18](#_Toc1365484010)

# Part A: Letter of Transmittal

July 7th, 2025

Hank Mardukas

Choice Mobile

73 McCaul Street,

Toronto, Ontario Canada

M5T 2X2

Dear Mr. Mardukas,

This letter outlines a crucial project for Choice Mobile, aimed at addressing the significant challenge of customer churn, which directly impacts future revenue and growth of the company. The rate at which customers discontinue their subscriptions with Choice Mobile is a problem that can be substantially mitigated with our proposed solution.

The proposed solution centers around the development of a Customer Churn Prediction Application that will leverage a bespoke machine learning model to forecast customer churn by analyzing past churn behavior and associated customer features. The practical benefits to Choice Mobile are substantial. By proactively identifying at-risk customers, the company can intervene with potential offers or support, improving customer satisfaction and ultimately reducing customer turnover. This will allow Choice Mobile to retain valuable customers and safeguard its revenue streams.

The implementation plan involves a data-driven approach utilizing the Customer Churn Dataset which is comprised of historical information from Choice Mobile’s records. This dataset includes key information such as age, gender, tenure, usage frequency, support calls, payment delay, subscription type, contract length, total spend, and last interaction. Our development process includes data loading, understanding, preparation, model selection, building, assessment, and evaluation, culminating in a final application that presents insights and visualizations. The project is projected to be completed by August 10th, 2025.

The initial cost of this project will be $25,000. This budget includes the initial training of the model based on the internal dataset, as well as the building and deployment of the application. A significant portion of this cost is allocated to ongoing model retraining, which is crucial to maintaining and enhancing the model’s accuracy as new customer data becomes available and business patterns evolve. This will ensure the predictive power of the model on unseen data will allow Choice Mobile to optimize their customer retention efforts effectively. The remaining costs are attributed to the hosting and maintenance of the model and its associated application for the term of one year. After the initial year has elapsed the cost of maintenance, retraining, and hosting of the model and associated application will cost 12,000 per year with additional costs required for work outside of the scope maintenance.

As a dedicated developer and project manager, I bring a comprehensive skill set ideally suited to this initiative. My expertise encompasses the entire lifecycle of machine learning projects, from training robust models capable of identifying complex patterns, to designing and building user-friendly applications that put these insights into action. Furthermore, I have practical experience in implementing these applications within an organizational context, ensuring they effectively integrate with existing systems and address specific business problems. My background in developing and managing such projects positions me to oversee this proposed project successfully, ensuring the delivery of tangible benefits to Choice Mobile by proactively identifying at-risk customers and improving overall customer satisfaction.

Sincerely,

Ryan Webber

Ryan Webber,

Technical ML Project Manager

# Part B: Project Proposal Plan

## Project Summary

The problem this project addresses is customer churn within Choice Mobile. Customer churn, defined as the rate at which customers discontinue their subscription, directly impacts the company’s revenue and growth. The need is to develop a mechanism to proactively identify customers who are likely to churn.

The proposed solution involves creating a customer churn prediction model. The primary deliverables will be a trained machine learning model, an accessible application that utilizes the model, as well as a user guide that will explain how to use the application. This application will leverage machine learning to forecast customer churn by analyzing historical data and associated features. The solution will also deliver visualizations representing data trends that will allow for Choice Mobile to have a better understanding of the composition of their customer base.

This application will benefit Choice Mobile by allowing them to identify at-risk customers proactively. This early identification will enable the company to intervene with potential offers or targeted support, ultimately improving customer satisfaction and significantly reducing customer turnover.

## Data Summary

The source of the raw data for this project is the Customer Churn Dataset which is comprised of historical customer information from Choice Mobile’s customer base. This dataset features columns such as age, gender, tenure, usage frequency, support calls, payment delay, subscription type, contract length, total spend, and last interaction. The dataset will initially be hosted on GitHub before being imported for cleaning and ingestion.

Before beginning model development a data understanding phase will be conducted, involving the collection of initial data, examination of its format, record count, and field identities. This will always require an in-depth deploration through querying and visualization to identify relationships.

Following the data understanding phase, a thorough data preparation phase will be necessary. This can be the most time-consuming aspect of the project as the quality of the data will directly influence the quality of the model going forward. This phase will consist of data selection, cleaning (addressing outliers, incomplete data, and erroneous values), as well potentially constructing new attributes. Throughout development ethical considerations and legal concerns regarding data privacy and usage will be thoroughly addressed.

## Implementation

The project will follow an iterative and adaptive development methodology, drawing on principles from the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework which will provide a structured yet flexible approach. This will allow accommodation for potential changes in requirements and for continuous refinement. The implementation will proceed through distinct phases, often revisited in an iterative manner.

The initial phase will be one of business understanding, which will focus on clearly defining the project’s objectives from a business perspective. This will include churn reduction targets, assessing available resources, and risks. At this stage data mining goals will be determined, outlining what technical success for the model will look like.

The following phase will be data understanding. This involves collecting, describing, exploring, and verifying the quality of the necessary datasets. A more thorough explanation of this step is available above in the Data Summary.

The next crucial phase will be data preparation which will involve the selection, cleaning, construction, integration, and formatting of raw data into a final dataset suitable for modelling.

Once data preparation is complete it will then be possible to begin modelling. In this phase various machine learning algorithms will be explored and applied to the prepared dataset. Models will be trained and assessed against performance criteria. This phase is an iterative process as the goal is to find a solution that is functional and can be improved with further iterations.

Following the modelling, evaluation of the solution will be required. Beyond technical assessment this phase will evaluate whether the selected model meets the defined business success criteria that had previously been established. The entire development process will be reviewed to identify areas for improvement, and next steps will be determined. This will include determining whether further iteration is required or if deployment may begin.

In deployment, the developed model and its associated application will be transitioned into a production environment. This phase will involve careful planning for the rollout, user training, as well as the establishment of monitoring and maintenance protocols.

Post-deployment, the model will undergo continuous maintenance, including retraining with new data to ensure its accuracy and relevance. This task is critical to maintaining optimal performance.

The project will also incorporate modern software practices such as version control with Git as well as Continuous Integration/Continuous Deployment (CI/CD) pipelines to streamline development and deployment.

## Timeline

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Stage #** | **Milestone or deliverable** | **Project Dependencies** | **Resources** | **Start and End Date** | **Duration** |
| 1 | Business & Data Understanding | N/A | Project Manager | July 7th, 2025 – July 13th, 2025 | 6 Days |
| 2 | Data Preparation | 1 | Data Scientist | July 13th, 2025 – July 24th, 2025 | 11 Days |
| 3 | Modeling & Initial Evaluation | 1,2 | Data Scientist & Machine Learning Engineer | July 24th, 2025 – July 30th, 2025 | 6 Days |
| 4 | Final Evaluation & Deployment Preparation | 1,2,3 | Machine Learning Engineer & Software Engineer | July 30th, 2025 – August 8th, 2025 | 9 Days |
| 5 | Initial Deployment | 1,2,3,4 | Cloud Deployment Engineer | August 8th, 2025 - August 10th 2025 | 2 Days |
| 6 | Post-Deployment Maintenance & Retraining | ALL PREVIOUS | Project Manager, Machine Learning Engineer, & Software Developer | Starting: August 10th, 2025 | Ongoing |

## Evaluation Plan

Verification Methods

The development process will require regular methods of verification to ensure the application is being developed as intended. Regular check-ins with senior leadership and middle management stakeholders will verify that the current state of the application is successfully providing the intended utility.

Data quality checks, including identification and handling of outliers and missing values, will be continuously verified by the Data Scientists and will be reviewed by the Project Manager for completeness as well as accuracy.

The internal model performance will be verified using appropriate machine learning metrics. These metrics will most importantly include accuracy, precision, recall, and F1-score.

Validation Methods

The primary validation will assess whether the deployed model meets the predefined business success criteria, which is the ability to successfully predict customer churn allowing for Choice Mobile to intervene to increase retention. Ensuring that the delivered application meets the established business success criteria is the paramount focus of this validation process.

The validation process will also require user acceptance testing. This will require a group of end-users to validate the applications usability and the model’s practical utility in identifying at-risk customers to support retention efforts. Their feedback will ensure that the application effectively solves the proposed business problem.

Finally, a post implementation review will be executed. This comprehensive review will be conducted to assess the overall project execution against the plan, documenting lessons learned as well as identifying areas for future process improvement.

## Costs

The initial cost for this project will be $25,000 CAD, with $5,000 CAD due at the time of acceptance of this proposal and the remainder due at time of successful delivery of the completed application. This initial cost will include 12 months of hosting, maintenance of the model, and application. The model maintenance will include ongoing retraining of the model which is critical to maintain and enhance accuracy as new customer data emerges and market dynamics shift.

Following the initial 12 months of model and application maintenance, the annual cost of this service will be $12,000 CAD. This continual service is optional but will require a one-year contract.

From initial investment in the delivery of the application, the breakdown of costs are as follows:

* $2,000 CAD – Cost of hosting the machine learning application and its model on a cloud server (for one year) allowing for instant and reliable access to the model and its utility.
* $10,000 CAD – One year of maintenance of both the model and its application estimated to be 100 hours over the course of the year, at a rate of $100 CAD / Hour
* 13,000 CAD – Skilled labor for Data Scientists, Project Manager, Software Engineers, and Machine Learning Engineers. 130 total hours estimated at a rate of $100 CAD / Hour which will cover Stages 1-5 of the timeline above.
* The programming environment, development computers, as well as the operating systems are all either open source or are already included in the development rate above.

# 

# 

# Part D: Post-implementation Report

## Solution Summary

The core problem addressed by this project was the significant challenge of customer churn at Choice Mobile, which directly impacts the company’s future revenue and growth. Customers discontinuing their subscriptions presented a quantifiable risk to business stability.

The proposed solution was the creation of a customer churn prediction model, delivered via an accessible application. This application leverages machine learning to forecast customer churn by analyzing historical data and associated features. By proactively identifying at-risk customers, Choice Mobile can implement interventions such as targeted offers or more substantial technical support. These interventions will improve customer satisfaction and ultimately can reduce customer turnover. Along with the predictive application, the data analysis also delivered visualizations representing data trends, offering Choice Mobile a clearer understanding of their customer base composition. These visualizations are included below.

## Data Summary

The raw data for this project was derived from an internal dataset belonging to Choice Mobile (the dataset was found located on Kaggle at <https://www.kaggle.com/datasets/muhammadshahidazeem/customer-churn-dataset>). The data underwent a structured processing and management lifecycle, adhering to an iterative and adaptive methodology influenced by the CRISP-DM framework.

The process began with a data understanding phase, which involved collecting the initial data and examining its surface properties, such as format, record count, and field identities. This included an in-depth exploration through querying and visualization to identify relationships and verify data quality. For instance, upon checking the dataset for values of NaN (not a number) it was discovered that there was a record which entirely consisted of NaN values.

(Shown in the screenshot below)

A screenshot of a computer

AI-generated content may be incorrect.

Following this, a thorough data preparation phase was conducted, this was the most time-consuming aspect due to its direct influence on the final quality of the model. This included data cleaning including addressing issues with outliers and incomplete data. The record noted above consisting of NaN values was selected and then dropped from the dataset.

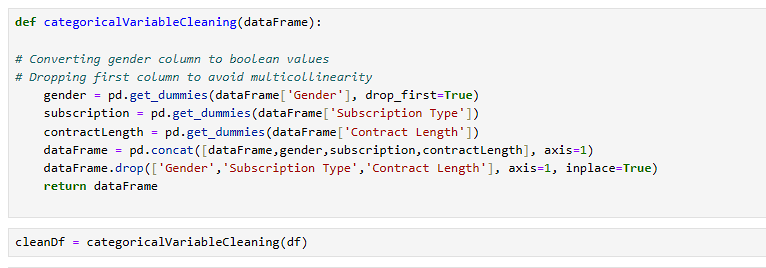
(Shown in the screenshot below)

A screenshot of a computer

AI-generated content may be incorrect.

The following step in data preparation was data construction. This primarily consisted of converting categorical columns that consist of multiple distinct values into columns consisting of Boolean values. In the dataset the “Gender”, “Subscription Type”, and “Contract Length” columns were converted into Boolean values using the Pandas method “get\_dummies”. In the case of the “Gender” column it was then necessary to drop one of the resulting columns in order to avoid multicollinearity.

(Shown in the screenshot below)



Finally, before beginning model training raw data was combined and reformatted to ensure that the dataset was suitable for modelling. Initially the dataset was split into both a training and testing set, but upon data exploration it was noticed that the set had been split in an unbalanced manner. The two datasets were combined and then re-split to train a more reflective model.

## 

## Machine Learning

The machine learning application employed predictive modeling using classification models to forecast customer churn. Specifically, the models developed and assessed were Logistic Regression and Random Forest Classifier. Ultimately only Random Forest Classifier was utilized for the delivered application.

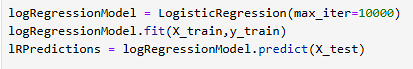
The development process followed the modeling phase of the CRISP-DM framework, where various algorithms suitable for classification were explored and applied to the prepared dataset. The data was split into training and testing sets to evaluate performance and generalization capabilities. This was achieved using sklearn’s train\_test\_split() method.

(Shown in the screenshot below)



The models were built using sklearn libraries, using LogisticRegression().fit() and RandomForestClassifier().fit() for training.

(Shown in the screenshot below)





These models once trained will take input consisting of customer data excluding a value for “Churn” and then predict a value for this missing column. It will either result in a 1 or a 0 being returned. The application allows for Churn Mobile to use an interface to submit customer information then upon a model prediction the application will return either “Predicted to Churn” or “Not Predicted to Churn”.

(Examples of both application results shown in the screenshots below)

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

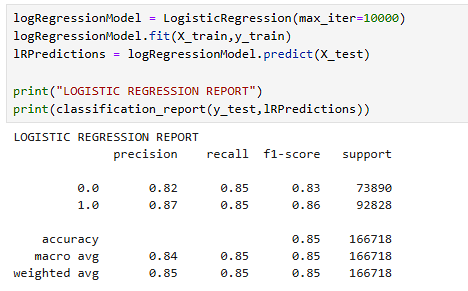
AI-generated content may be incorrect.

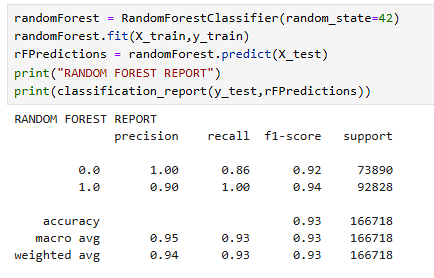
The selection of the classification models was justified by the nature of the churn prediction problem, which requires categorizing customers. Both Logistic Regression and Random Forest Classifier were explored for their suitability. The Random Forest Classifier was ultimately chosen as the primary model for its superior performance.

## Validation

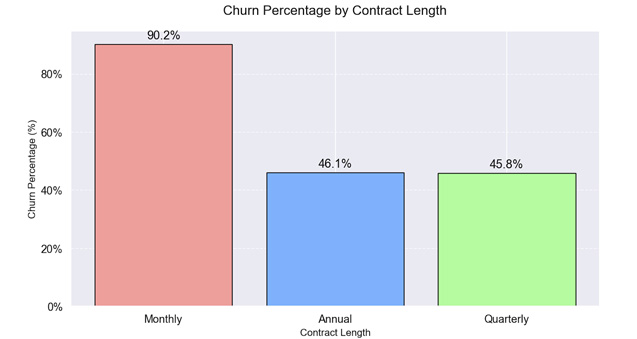
Both machine learning models explored for suitability for the application are categorized as supervised models. The validation method for both potential models was sklearn’s train\_test\_split() method. Both models were tested and then compared using sklearn’s classification\_report() method. While this report returns numerous metrics to consider when picking a model, it was determined that recall, specifically recall of positive results, was most important to avoid false negatives. Practically speaking, Choice Mobile wants to avoid missing retention efforts on at risk customers and there are not negative consequences to these efforts being applied to customers who potentially won’t churn.

(Training of both models as well as their classification reports are in the screenshots below)

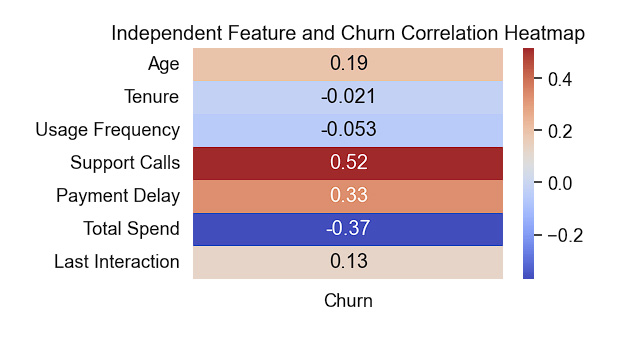




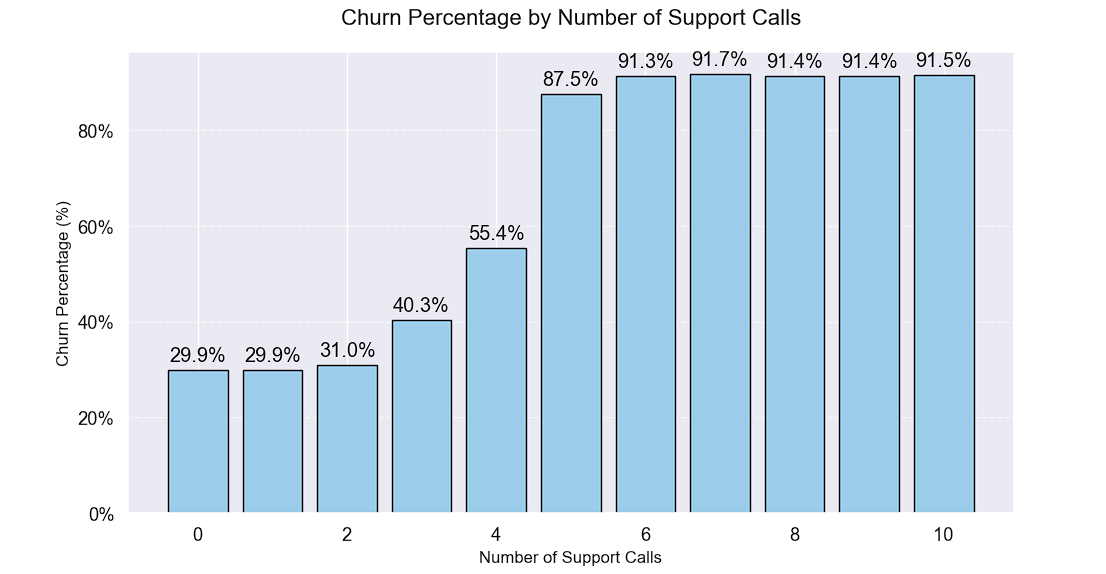
## Visualizations

“Churn Percentage by Contract Length” (Bar Plot) was created using matplotlib.pyplot and seaborn. It clearly illustrates the churn rate for different contract lengths. It reveals that customers on a monthly subscription exhibit a significantly higher churn percentage (90.2%) compared to annual (46.1%) or quarterly (45.8%) contracts.

“Independent Feature and Churn Correlation Heatmap” was also generated with seaborn. It displays the correlation between various independent numerical features (e.g., Age, Tenure, Usage Frequency, Support Calls, Payment Delay, Total Spend, Last Interaction) and the Churn variable. It effectively highlights which factors are strongly associated with Churn.



“Churn Percentage by Number of Support Calls” (Bar Chart) visualizes how the number of support calls correlates with the churn percentage, providing direct insight into the impact of customer support interactions on retention.



## User Guide

# Open Web Browser

# Navigate to <https://colab.research.google.com/drive/1_6m4g7ztkqV9AemxTlbMHBnqLojoBq2M?usp=sharing#scrollTo=4b0eea02-7d1f-485e-af92-0e1a08de7221>

# Once page has loaded click button that says, “Run All” (

# Once all cells have executed; scroll to bottom cell and enter client values in form and click “Submit” Button.