Predicting Member Churn

For Fitness Center

Scientia Consulting

Agenda

01

Business Problem and Overview

Brief overview of the business problem at hand

02

Data Exploration

Preliminary insights into provided dataset

03

Model Proposal and Early Insights

Assessing member retention and predicting member churn

04

Projected Plan and Requests

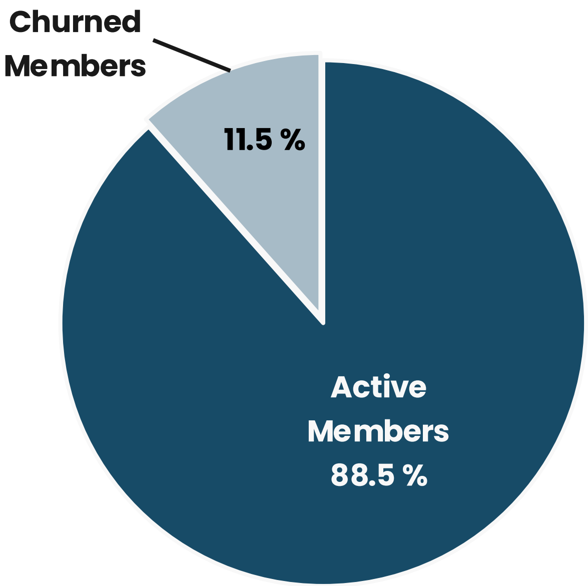
Estimated timelines for deliverable and anticipated cost

01

Business Problem and Overview

Brief overview of the business problem at hand

Business Problem



- The Fitness Center has noticed a **decline in membership retention** after the initial sign-up period and wants to understand the economic benefit of reducing the amount of churn.

We found that the Fitness Center has experienced only **11.5% in member churn** in the past 3 years (Avg. churn for fitness centers, 30-60%).

Therefore, we suggest a **modified** a approach to **understanding** the Fitness Center's **member retention** and **predicting potential churn.**

Why Scientia Consulting?

Domain Expertise

Our consulting group brings specialized knowledge in the wellness industry which enables us to understand the unique challenges and nuances that clients face. We develop solutions that are not only data-driven but also grounded in industry insights, ensuring relevance and impact for our clients.

Proven Track Record

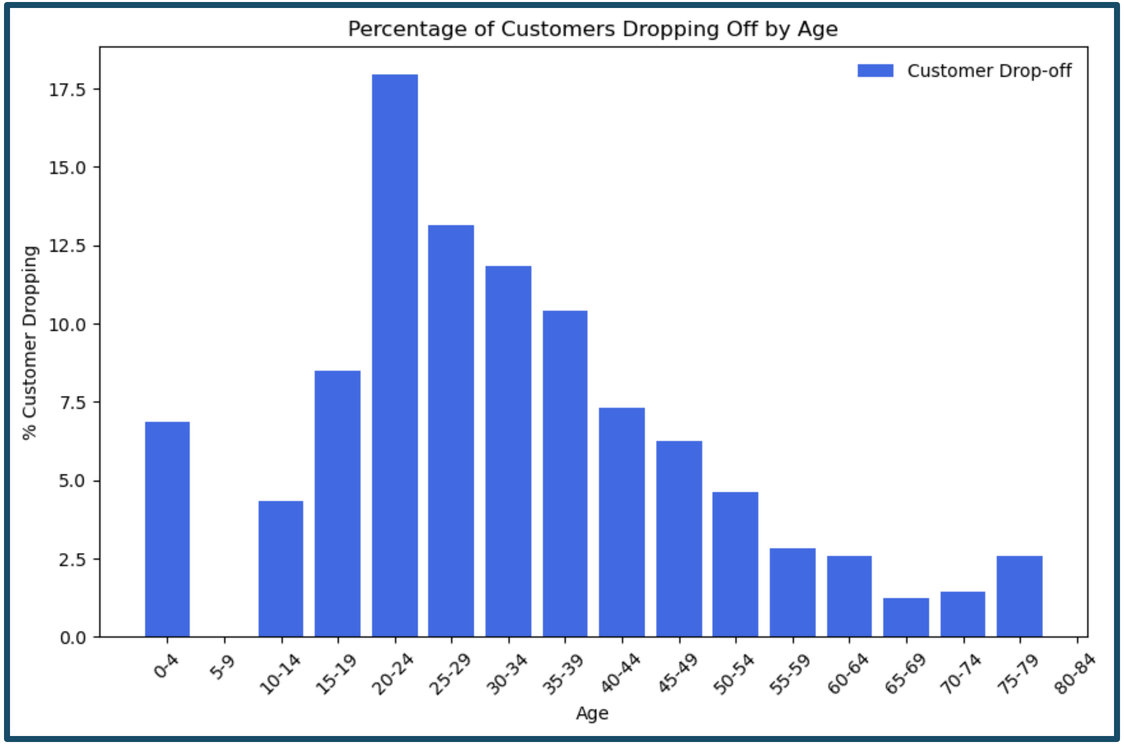
We pride ourselves on over two decades of successful engagements, demonstrated through measurable outcomes and over 100 satisfied clients. Our team's past projects highlight our ability to drive results, from increasing operational efficiency to boosting profitability.

02

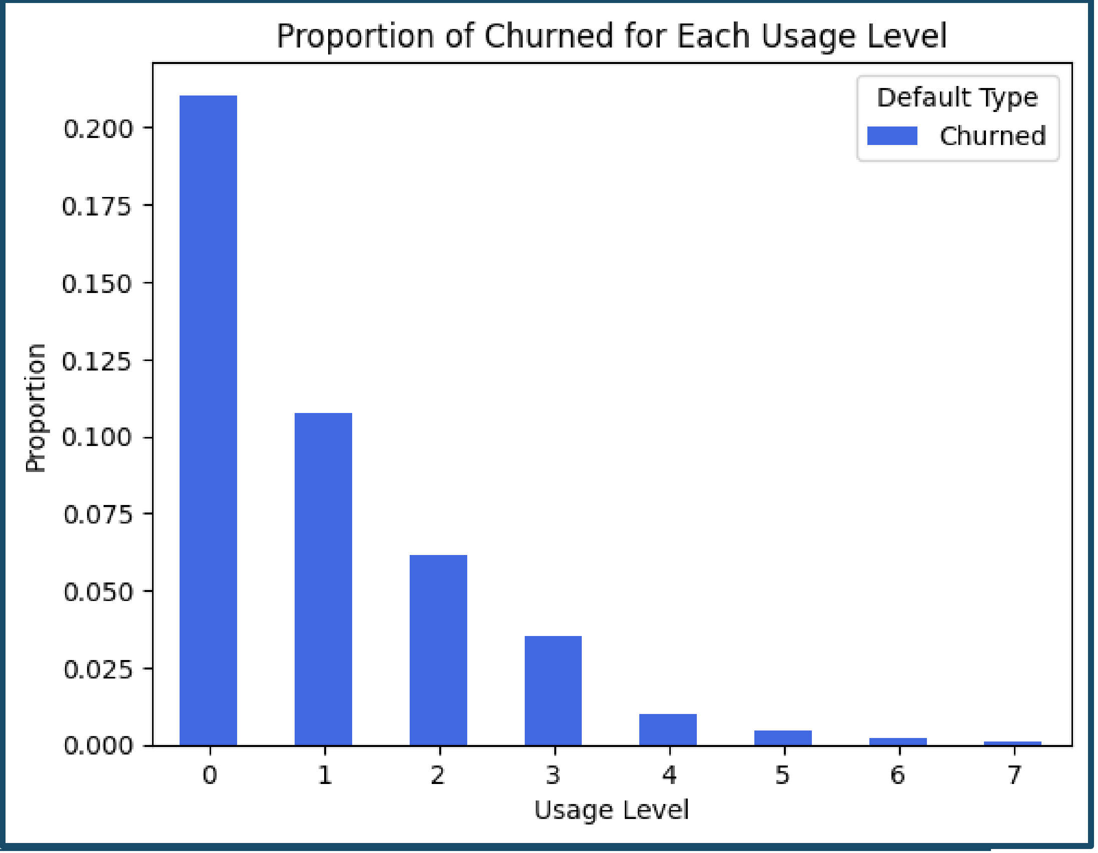
Data Exploration

Preliminary insights into the provided dataset

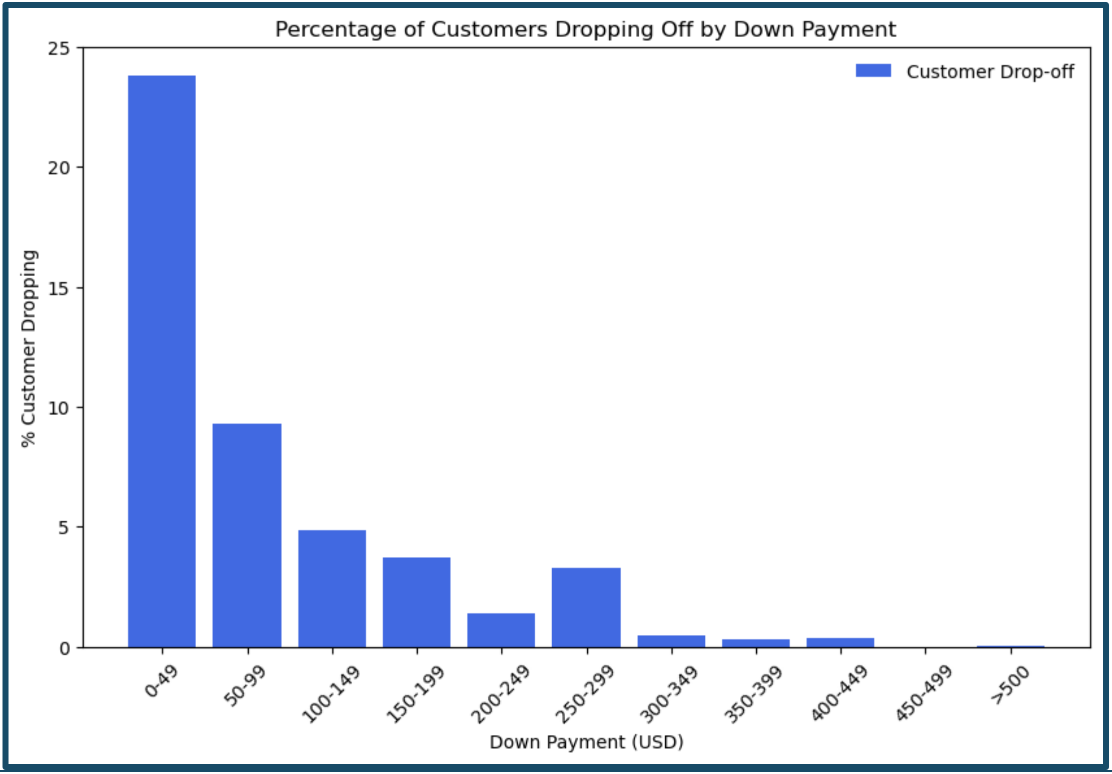
Younger age groups have higher churning than older age groups.



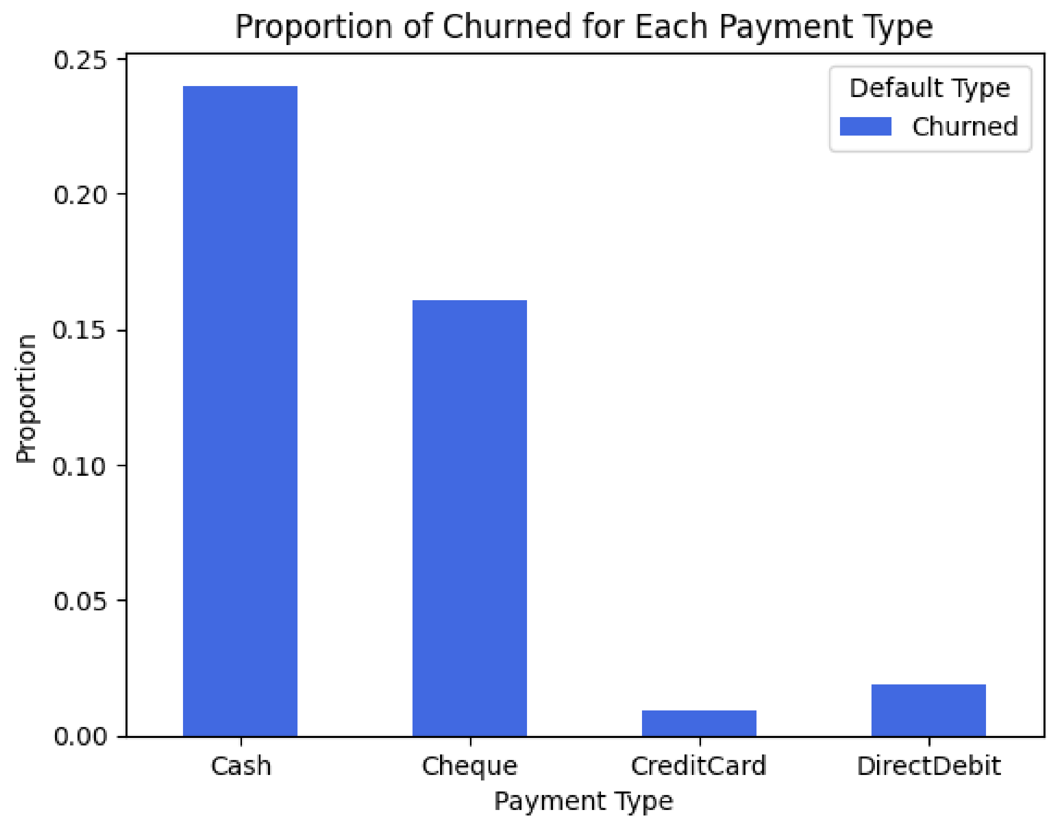
Members that use the Fitness Center less frequently have a higher proportion of churning.



Customers with lower down payment have a higher churning.



Members that use cash and checks for payments have a higher proportion of churning.



03

Model Proposal and Early Insights

Assessing member retention and predicting member churn

Implementing a Logistic Regression Model

Why this is a good model:

- Customer churn is represented as a **binary outcome** (1 for churned customers, 0 for retained customers).
- Provides **probability scores,** which help to understand how likely a customer is to churn.
- Provide a solid starting point to benchmark against more advanced models, like Random Forest, Gradient Boosting, XGBoost.

Variables to use:

- Age
- Payment Type
- Gender
- Usage Level
- Down payment

Early Insights (Logistic Regression)

The model reveals important insights into the predictors of churn, such as age, gender, usage level, and payment type

The probability of churning:

- Decreases with an increase in age, usage level, or downpayment
- Decreases when going from cash payment to check, credit card, or direct debit

Correctly identified most of the non-churned members (6416 out of 6646)

Due to class imbalance, failed to identify some of the churned members (339 out of 807), predicting they would stay.

Implementing an XGBoost Model

XGBoost is an advanced tool for making precise predictions by learning from previous patterns and adjusting based on past errors. Think of it as a team working together, with each member improving on the last, to reach the most accurate result.

Why XGBoost is right for this problem –

- **Handles Imbalance:** We have fewer churn cases (11.5%) than non-churn cases (88.5%).
- **Handles Large Datasets:** XGBoost can run cheaper and faster with large datasets compared to other advanced models.
- **Feature Importance:** XGBoost shows which factors are most important in predicting churn, helping us understand why members might leave.
- **Captures Complexity:** XGBoost can capture complex patterns in the data, which is helpful since the reasons behind churn can be complicated.

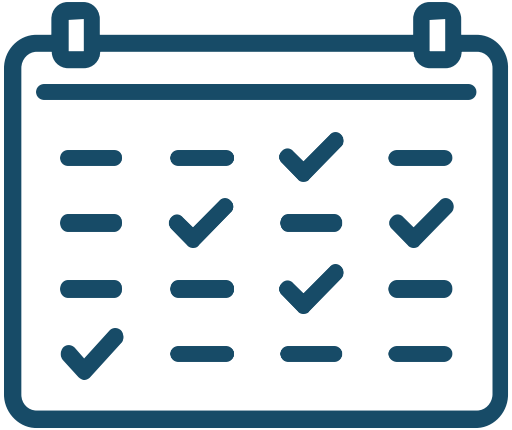
Early Insights – XGBoost

From the preliminary XGBoost model, we observe that **usage**, **downpayment** amount, and **cash** payment have the most impact on churn. We observe a model accuracy of 90%.

However, **data cleaning** based on data collection methodology and **additional data** such as cancellation date, customer ID, length of membership, cost of membership etc would be needed for best model performance and interpretability.

04

Projected Plan and Requests



Estimated timelines for deliverable and anticipated cost

Projected Timeline

Task/ Deliverable		Nov. 8 - 15					Nov. 16- 22				
		M	T	W	Th	F	M	T	W	Th	F
1	Complete Data Cleaning	■	■								
2	Conduct in-depth analysis of churning trends			■	■	■					
3	Implement models						■	■	■		
4	Draft recommendations									■	
5	Present final report and findings										■

Projected Cost of *Scientia Consulting* Services

25k – 30k

Includes a 2-week turnaround, data cleaning, model development and deployment, and formal recommendations.

Formal Data Requests

- **Length of membership**. This would help us understand how long do members stay before they churn.
- **Fixed and Variable costs** such as G&A costs, equipment costs and other expenses to understand the cost pattern of the Fitness Center and whether these expenses are contributing to reducing the customer churn rate.
- **Price of membership plans** corresponding to the length of membership. This will help us understand which plans have higher churn rates and find incentives to reduce those churn rates.

Key Takeaways

Strong Retention with Room for Improvement

The Fitness Center's 11.5% churn rate is a sign of strong member retention, but there's an opportunity to optimize this even further by proactively addressing at-risk members and enhancing engagement strategies..

Data-Driven Approach to Churn Reduction

We recommend implementing predictive analytics to identify at-risk members. This will allow the company to act before members leave, reducing churn and maximizing member value.

Early Insights Indicate Targetable Segments

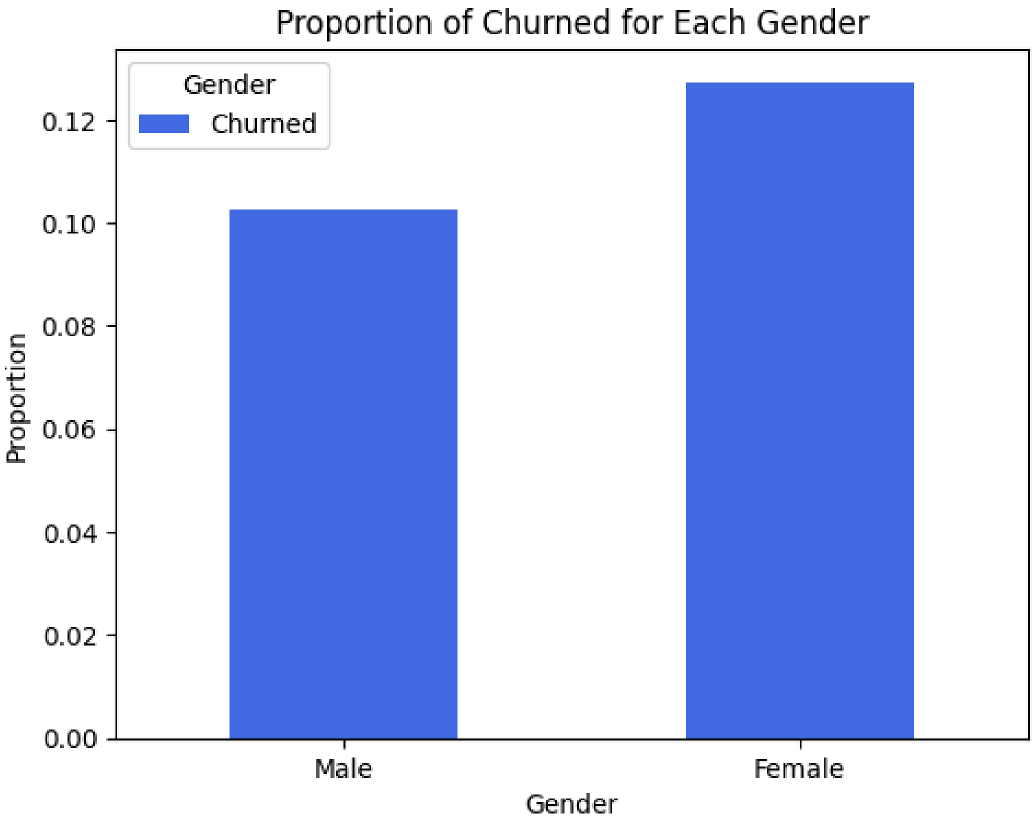
Preliminary data analysis shows that certain member characteristics are more likely to churn. Indicating that targeted retention efforts for these groups could improve outcomes.

Fast Execution of Clear Deliverables

In just two weeks, we will develop predictive models to identify churn risk and provide actionable insights and retention strategies based on model results.

Appendix

Female members have a higher proportion of churning than male members.



Model Comparison

Criterion	Logistic Regression	Random Forest	Gradient Boosting
Model Purpose	Binary Classification	Binary Classification with non-linear relationships	Binary Classification with complex patterns
Cost	Low	Moderate	High
Handling on non-linearity	Assumes linearity between predictors	Captures non-linear patterns in data	Captures complex, non-linear relationships
Sensitivity of Outlier	Can be affected by extreme values	Robust to outlier due to averaging over tree	Robust but sensitive in overfitting

Early Insights (Logistic Regression)

The model reveals important insights into the predictors of churn, such as age, gender, usage level, and payment type

Variable	Coefficient	Standard Error
Intercept	0.8841	0.094
age	-0.0055	0.003
gender	0.2898	0.058
use	-0.6890	0.030
downpayment	-0.0138	0.001
Payment Type: Check	-0.9085	0.063
Payment Type: Credit Card	-3.9294	0.165
Payment Type: Direct Debit	-2.9006	0.123

Early Insights (Logistic Regression)

The model reveals important insights into the predictors of churn, such as age, gender, usage level, and payment type

```
Logistic Regression Model Results:
                           Logit Regression Results
==============================================================================
Dep. Variable:                default   No. Observations:                17390
Model:                          Logit   Df Residuals:                    17382
Method:                           MLE   Df Model:                            7
Date:                Wed, 06 Nov 2024   Pseudo R-squ.:                  0.4074
Time:                        12:23:00   Log-Likelihood:                -3738.8
converged:                       True   LL-Null:                       -6309.1
Covariance Type:            nonrobust   LLR p-value:                     0.000
=========================================================================================
                          coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
Intercept               0.8841      0.094      9.412      0.000       0.700       1.068
age                    -0.0055      0.003     -2.150      0.032      -0.011      -0.000
gender                  0.2898      0.058      4.955      0.000       0.175       0.404
use                    -0.6890      0.030    -22.969      0.000      -0.748      -0.630
downpmt                -0.0138      0.001    -24.130      0.000      -0.015      -0.013
pmttype_Cheque         -0.9085      0.063    -14.345      0.000      -1.033      -0.784
pmttype_CreditCard     -3.9294      0.165    -23.827      0.000      -4.253      -3.606
pmttype_DirectDebit    -2.9006      0.123    -23.670      0.000      -3.141      -2.660
=========================================================================================
```

Early Insights (Logistic Regression)

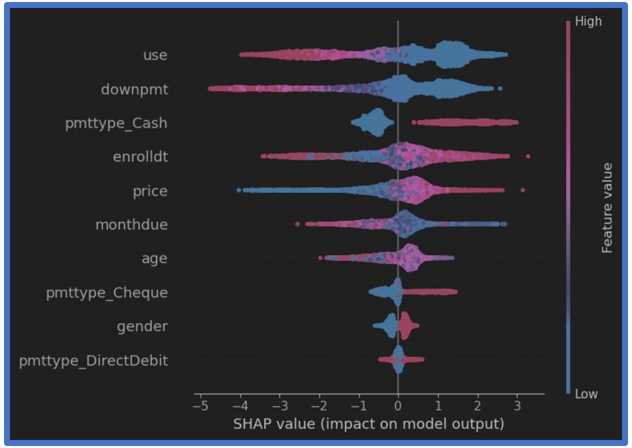
```
Accuracy: 0.9063

Confusion Matrix:
[[6416  230]
 [ 468  339]]

Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.97      0.95      6646
           1       0.60      0.42      0.49       807

    accuracy                           0.91      7453
   macro avg       0.76      0.69      0.72      7453
weighted avg       0.90      0.91      0.90      7453
```

Correctly identified all of the non-churned members (6416 out of 6646)

Due to class imbalance, failed to identify all of the churned members (339 out of 807), predicting they would stay.

XGBoost preliminary model insights

Model accuracy: 90%



	precision	recall
0	0.98	0.91
1	0.55	0.85
accuracy		
macro avg	0.76	0.88
weighted avg	0.93	0.90

Feature	Gain Importance
pmttype_Cash	25%
pmttype_Cheque	22%
use	19%
downpmt	13%
enrolldt	5%
pmttype_DirectDebit	4%
price	4%
monthdue	4%
gender	3%
age	3%

Detailed Projected Cost of *Scientia Consulting* Services

Service	Data Cleaning	Churn Analysis	Model Development	Recommendations & Financial Projections	Final Report
Price	5K	5K	5-10K	5K	5K
Description	• Ensure data quality for reliable analysis • Includes standardization and initial exploratory work.	• Provide a nuanced understanding of churn drivers. • Identifies high-risk groups.	• Develop and validate the predictive model. • Ensures accuracy and robustness for actionable insights.	• Targeted retention strategies. • Focuses on financial benefits and operational considerations.	• Professional, client-focused report and presentation.

End of Deck