IBM Telco

Predictive Customer Churn strategies

Christopher Hughes
Data Scientist

Agenda

Background

Key findings

Additional steps

Summary

Next Steps

Questions

You wanted to identify customers who would churn to add in a save strategy

This project was to assess the important features that management should consider in developing customer retention programme.

We focussed on two main questions:

a) Can we accurately predict which customers will churn?
b) Which factors heavily relate churn and can be used in a save strategy?

With sub questions of:

a) How the model performs
b) The financial savings from its implementation

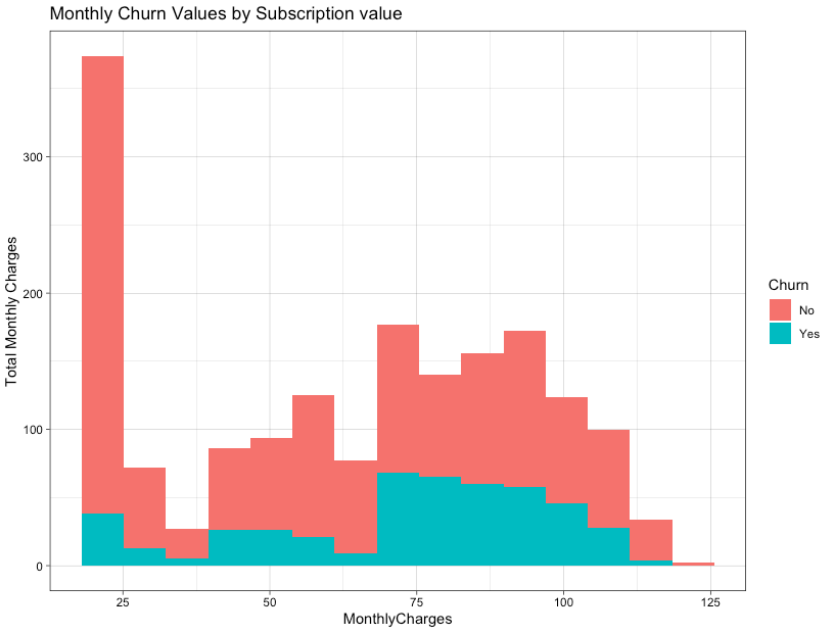
We determined that the approach could save £4,000 per month per 1,000 customers

We believe this because…

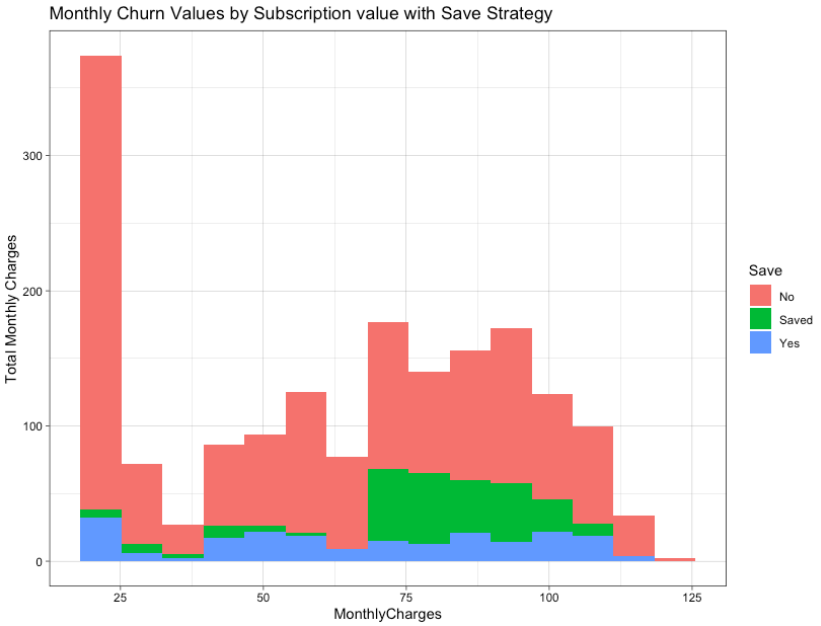
Key Findings

- £4000 in savings over first month per 1000 customers. Based on successfully 252 of 467.
- Model can identify 8 out of 10 customers' behave correct (stay or churn)
- Key features highlighted for customer retention programme are:
  - Contract Type
  - Payment Method

Without Retention Strategy



With Retention Strategy



Actual vs predicted savings performance for cohort of 1760 (example test set)

Summary

You asked:	Answer:
Can we predict churn customers?	Yes – Model can accurately identify potential churn customers.
Which churn factors can be used for retention?	They are – contract type and payment type
How model performs?	Well – Model can identify the behaviour of 8 out of 10 customers correctly (stay or churn)
Financial savings?	Yes – £4,000 per month per 1,000 customers.

Next steps

We can use this model to create a retention programme

Key Influencer	Situation and Result	Suggested Use
Contract Type	• The contract types are: Month-to-month, One year, Two year • **Month-to-month** has the highest churn	Offer discounted average monthly rates for 2 year and 1 year contracts to monthly and new customers.
Payment Type	• The payment methods are: Bank transfer (automatic), Credit card (automatic), Electronic check, Mailed check • **Electronic Check** Customers have a higher churn rate	Incentive **Electronic Check** customers to move Credit Card (automatic) with reward or discount

Potential next steps

Productionising the planned project

- The clear next step is to have the model in production so as to be ready for business use
- Establishing a stable production infrastructure to enable this can be as challenging as the creation of model code

Create and run Customer retention strategy

- Work with marketing to create a customer “save strategy” driven by model features and additional customer segmentation
- Design live test for churn model and save strategy.
- Implement strategy in live test

Other projects

- Customer segmentation
- Product recommendation and upselling

Questions?

Appendix

To assess the value of using the model we used a set of assumptions

Assumptions: based on values supplied by organisation and applied uniformly

Cost retention

- £20 per customer

Cost of Churn

- £100 per customer

Retention success

- 100% success rate

Definition: we define the value as below

Saved monthly revenue - (Unidentified Churn x Cost of Churn + Identified Churn x retention cost)

=

Actual net saving

EDA and Modelling Process (CRISP-DM)

1. Data Preparation

a) Missing data for TotalMonthly with tenure 0 imputed to 0.
b) All categorical variables
c) Difference between Tenure x MonthlyCharge and TotalCharge – highly correlated but not exact*
d) Gender removed as low influence variable.

2. Model Selection

a) Binary Classification problem: churn (1) or retain (0)
b) Explainable model required e.g. Decision Tree, Logistic Regression

3. Evaluation and Tuning

a) Logistic Regression selected as based on metric performance**
b) Modelled validated and tuned with Cross-Validation and hyper parameter optimisation, respectively

4. Deployment

a) Model written in package and version controlled
b) Unit tests and continuous integration added for quality monitoring

*Is there missing field? Monthly increases, Customers fines, etc.
** Precision, recall, accuracy

Model and Performance

a) Logistic regression model selected
b) Using feature set, excluding gender
c) All Categorical implemented as dummy variables

Feature Scores

Feature	Estimate	Std..Error	z.value	Pr...z..
tenure	-0.0621024	0.00737227	-8.4237763	3.65E-17
ContractTwo year	-1.446768	0.20337668	-7.113736	1.13E-12
ContractOne year	-0.6457614	0.12295241	-5.2521249	1.50E-07
TotalCharges	3.45E-04	8.31E-05	4.15425374	3.26E-05
PaymentMethodElectronic check	0.35423998	0.11068447	3.20044886	0.00137214
PaperlessBillingYes	0.24943427	0.08671088	2.87662005	0.00401959
DependentsYes	-0.2981046	0.10475951	-2.8456094	0.00443265
MultipleLinesYes	0.42276868	0.2037378	2.07506252	0.03798074
SeniorCitizen1	0.15097005	0.09748243	1.54868994	0.12145627
TechSupportYes	-0.3156324	0.20884249	-1.511342	0.13070133

Performance Metrics

Metrics	Estimate
accuracy	0.79431818
kap	0.44670142
sens	0.8863109
spec	0.53961456
precision	0.84202792
recall	0.8863109