NPC Retention Case Study

1 - Problem Statement

A Newspaper Publishing Company (NPC) is facing increasing churn rates since years. This evolution has been jumpstarted by the rise of news- websites, and has continued at an accelerating rate due to the popularity of tablet computers and (social) news- aggregator applications.

Hence the NPC requires a predictive retention (or churn) model in order to predict which customers will not renew their newspaper subscriptions.

Customers are not allowed to cancel their subscriptions; hence a customer is defined as a churner if he or she does not renew the subscription once the current subscriptions end.

2 – Data Description

The following dictionaries provide names of tables, and meanings of each of their columns:

Customers

Variable	Description
CustomerID	Unique customer identifier
Gender	M=Male F=Female NA=Not Available
DoB	Date of Birth
District	Macro- geographical grouping: District 1-8
ZIP	Meso- geographical grouping
StreetID	Micro- geographical grouping: Street

Subscriptions

Variable	Definition
SubscriptionID	Unique subscription identifier
CustomerID	Customer identifier
ProductID	Product identifier
Pattern	Denotes delivery days. The position in the string equals the days in the week. {1=delivery,0=non-delivery} e.g., 1000000= delivery only on Monday 1111110= delivery on all days (except Sundays)
StartDate	The subscription's start date
EndDate	The subscription's end date
NbrNewspapers	Total number of copies, including NbrStart as determined at the start of the subscription.
NbrStart	Maximum number of copies before payment is received.
RenewalDate	Date that renewal is processed.
PaymentType	BT= Bank transfer DD= Direct debit
PaymentStatus	Paid | Not Paid
PaymentDate	Date of payment
FormulaID	Formula identifier.
GrossFormulaPrice	GrossFormulaPrice=NetFormulaPrice + TotalDiscount
NetFormulaPrice	NetFormulaPrice = GrossFormulaPrice – TotalDiscount
NetNewspaperPrice	Price of a single newspaper
ProductDiscount	Discount based on the product
FormulaDiscount	Discount based on the formula
TotalDiscount	TotalDiscount= ProductDiscount + FormulaDiscount
TotalPrice	TotalPrice = NbrNewspapers * NetNewspaperPrice TotalPrice=NetFormulaPrice + TotalCredit
TotalCredit	The customer's credit of previous transactions. Credit is negative, debit is positive.

Delivery

Variable	Definition
SubscriptionID	Subscription identifier
DeliveryType	DR=Delivery Rerouting MD=Main Delivery
	DI=Delivery Interruption
DeliveryClass	NOR= Normal ABN=Abnormal: every delivery that deviates from normal delivery (see DeliveryContext)
DeliveryContext	OTH=Other ACH=Address Change PCH=Product Change REN=Renewal NPA=Non- Payment VAC=Vacation
StartDate	Start date delivery
EndDate	End date delivery

Formula

Variable	Definition
FormulaID	Formula identifier
FormulaCode	Formula code
FormulaType	CAM=Campaign REG=Regular
Duration	Duration of the formula in months.

Credit

Variable	Definition
SubscriptionID	Subscription identifier
ActionType	PO=Payout NC=No consequence EN=Extra Newspapers CC=Create Credit
ProcessingDate	Date on which the credit is processed.
CreditSource	SUB=Subscription TRA=Transaction (e.g, Double payment) COM=Complaint
Amount	Credit Amount in $
NbrNewspapers	Number of newspapers as credit. Only relevant when ActionType=EN

Complaints

Variable	Definition
ComplaintID	Unique complaint identifier
CustomerID	Customer identifier
ProductID	Product Identifier
ComplaintDate	Date of complaint
ComplaintType	1=Non delivery 2=Late delivery 3=Wrong product 4=Messy delivery 5=No post delivery
	6=Delivery while nothing ordered 7=Damaged product 8=Wrong address 9=Other
SolutionType	1=Assign Credit 2=Post delivery 3=No solution needed 4=Take precautions
FeedbackType	1=Force Majeure 2=Weather conditions 3=Error postman 4=Late delivery to postal office 5=Postman confirmed delivery 6=Strike postal office

3. Instructions

Following the steps similar to the first project, use these datasets to create a Base-table after deciding on a Timeline for modeling. Then as usual, build multiple models and report your best one.