JIGSAW CAPSTONE PROJECT - R

TELECOM CHURN ANALYSIS

ANSWERING BUSINESS QUESTION

1. What are the top five factors driving likelihood of churn at Mobicom?

- The model results show that the top 5 factors affecting churn are:
- a. hnd_price_249.98 with beta coefficient of 1.1908253
- b. ethnic_C with beta coefficient of 0.8290688
- c. uniq_9 with beta coefficient of 0.8184267
- d. uniq_7 with beta coefficient of 0.7289988
- e. retdays_1 with beta coefficient of 0.6746204

Continued.......

- The 1st factor explains, with a unit increase in level of variable hnd_price_249.98 there is 1.1908253unit increase in churn.
- The 2nd Factor explains, with a unit increase in variable ethnic_C, there is 0.8290688 unit increase in churn.
- Same explanation applies to the next 3 variables.
- The fifth factor explains unit increase in retdays_1 variable there is 0.6746204 unit increase in churn.

Continued.....

- Thus family bundles should be rolled out for families with 9 and 7unique subscribers.
- Special offers should be given to customers who makes retention calls, at the earliest as per their grieviances.
- Special plans should be rolled out for people with current hand set price of level 249.98 customers having ethnicity_C category level.

Continued......

2. **Validation of survey findings.**

a) Whether "cost and billing" and "network and service quality" are important factors influencing churn behaviour.

- The following variables explain "cost and billing" and "network and service quality"
- Variables totmrc_Mean i.e. 'base plan charge' representing cost to customer,
- Var rev_Range i.e. 'Range of Revenue(charge amount)' representing billing amount,
- Var ovrrev_Mean = DATOVR_MEAN + VCEOVR_MEAN i.e. 'Mean overage revenue' (It is the sum of data and voice overage revenues) representing the overage revenue earned from customers after billing the same to them. and var totrev i.e. 'Total revenue' representing total revenue earned from customers.

Continued.....

Var totmrc_Mean has beta coefficient value of -0.00541871 meaning a unit increase in this variable is causing decrease in churn by 0.00541871/unit.

Var rev_Range has beta coefficient value of 0.00204299 meaning a unit increase in this variable is causing increase in churn by 0.00204299/unit

Var ovrrev_Mean has beta coefficient value of 0.00709310 meaning a unit increase in this variable is causing increase in churn by 0.00709310/unit

Continued……

Var totrev has beta coefficient value of 0.00025578 meaning a unit increase in this variable is causing increase in churn by 0.00025578/unit.

Having said that, if we notice above mentioned beta values, a unit increase in them is having almost 0% impact on churn. So it seems cost and billing is not very important factors here influencing churn behaviour at Mobicom.

The following variables explain "network and service quality"

Variable	Beta Value	Variable	Beta value
mou_Range	0.00031484	mou_opkv_Range	-0.001080128
change_mou	-0.00064874	iwylis_vce_Mean	-0.01525907
drop_blk_Mean	0.00785318	avgqty	0.00119470
drop_vce_Range	0.01864740	avg6mou	-0.00025388
		retdays_1	0.67462037
		Complete_Mean	-0.00173140

Continued.....

- From the above statistics, data explains the following.
- Variables mou_Range 1.e. with a unit increase in 'Range of number of minutes of use', there is increase in Churn by 0.00031484 units.
- Var change_mou i.e. with a unit increase in 'Percentage change in monthly minutes of use vs previous three month average, there is decrease in Churn by -0.00064874 units.
- Var drop_blk_Mean i.e. with unit increase in 'Mean number of dropped or blocked calls', there is an increase in churn by 0.00785318 units

Continued.....

- Var drop_vce_Range i.e. with a unit increase in 'Range of number of dropped (failed) voice calls', there is an increase in Churn by 0.01864740 units.
- Var mou_opkv_Range i.e. with a unit increase in 'Range of unrounded minutes of use of off-peak voice calls, there is a decrease in Churn by -0.001080128 units.
- Var iwylis_vce_Mean i.e. with a unit increase in 'Mean number of inbound wireless to wireless voice calls', there is a decrease in churn by -0.01525907 units.

Continued......

- Var avgqty i.e. with a unit increase in 'Average monthly number of calls over the life of the customer', there is an increase in Churn by 0.00119470 units.
- Var avg6mou i.e. with unit increase in 'Average monthly minutes of use over the previous six months', there is a decrease in Churn by -0.00025388 units.
- Var retdays_1 representing values captured in the variable retdays i.e. with a unit increase in 'Number of days since last retention call', there is an increase in Churn by 0.67462037 units. This variable is probably explaining the service quality of the company.
- Var complete_Mean i.e. with unit increase in 'Mean number of completed voice and data calls' there is a decrease in Churn by -0.00173140 units

Continued....

- Of the above variables, the beta coefficient of variable retdays_1 is expressing a very important factor influencing Churn behaviour. That is with the increase in the number of days since a customer .makes a retention call, the customer's chances of churning is very high.
- This could probably be because their grieviances are not being catered to properly. These customers should be paid more attention to and special offers should be made to them depending upon their grieviances.

Continued.....

2b) Are data usage connectivity issues turning out to be costly? In other words, is it leading to churn?

- comp_dat_Mean - Mean no. of completed data calls.
- plcd_dat_Mean - Mean number of attempted data calls placed
- opk_dat_Mean - Mean number of off-peak data calls
- blck_dat_Mean - Mean no. of blocked / failed data calls
- datovr_Mean - Mean revenue of data overage.
- datovr_Range - Range of revenue of data overage
- drop_dat_Mean - Mean no. of dropped / failed data calls

Continued...

- The Data Quality Report for all the above variables show that only 10% to 15% customers are actualy making data calls or using the internet.
- This could be a matter of concern since the global market survey report shows "Subscribers who have switched operators in recent months reported two key information sources in their decision: the Internet and recommendation of family and friends. In this case it seems customers are not really using the internet.
- So it would be good to work towards attaining more customers to use data and also towards proving quality network connectivity and service to provide maximum customer satisfaction and reduce Churn.
- Since there is not enough usable data for the above variables they are not showing any influence on the Churn Behaviour at Mobicom.

Continued....

3. Would you recommend rate plan migration as a proactive retention strategy?

- Variable ovrrev_Mean has beta coefficient of 0.00709310.
- var ovrrev_Mean = DATOVR_MEAN + VCEOVR_MEAN i.e. 'Mean overage revenue' It is the sum of data and voice overage revenues representing the overage revenue earned from customers after billing the same to them.
- The Beta coefficient is not showing a strong impact of overage billing as an influencer of churn behaviour.
- Though this might be a matter of concern for few individual customers and they could be catered to on case to case basis. But overall rate plan migration as a proactive retention strategy might not help much at Mobicom.

Continued.....

4. What would be your recommendation on how to use this churn model for prioritisation of customers for a proactive retention campaigns in the future?

- Gains Chart
- library(gains)
- (Running R codes)
- gains(test$churn,predict(mod6,type="response",newdata=test),groups = 10)
- The Gains Chart shows that the top 20% of the probabilities contain 29.2% customers that are highly likely to churn.

Continued…

- Selecting Customers with high churn rate (Running R codes)
- test$prob<-predict(mod6,type="response",newdata=test)
- quantile(test$prob,prob=c(0.10,0.20,0.30,0.40,0.50,0.60,0.70,0.80,0.90,1))
- Top 20% of the probabilities lie between 0.3481952 and 0.7391842.
- Applying cutoff value to predict customers who Will Churn
- pred4<-predict(mod6, type="response", newdata=test)
- pred4<-ifelse(pred4>=0.3481952 , 1, 0)
- table(pred4,test$churn)

Continued.......

(the following lines are R codes)

- `Targeted<-test[test$prob>0.3481952 & test$prob<=0.7391842 & test$churn=="1","Customer_ID"]`
- `Targeted<-as.data.frame(Targeted)`
- `nrow(Targeted)`
- `write.csv(Targeted,"Target_Customers.csv",row.names = F)`
- Thus Using the model can be used to predict customers with high probability of Churn and extract the target list using their "Customer ID".

Continued....

5. What would be the target segments for proactive retention campaigns? Falling ARPU forecast is also a concern and therefore, Mobicom would like to save their high revenue customers besides managing churn. Given a budget constraint of a contact list of 20% of the subscriber pool, which subscribers should prioritized if "revenue saves" is also a priority besides controlling churn. In other words, controlling churn is the primary objective and revenue saves is the secondary objective.

Continued…….

- (Running R codes)
- # Solution:
- pred5<-predict(mod6, type="response", newdata=test)
- test$prob<-predict(mod6,type="response",newdata=test)
- quantile(test$prob,prob=c(0.10,0.20,0.30,0.40,0.50,0.60,0.70,0.80,0.90,1))
- pred6<-ifelse(pred5<0.20,"Low_Score", ifelse(pred5>=0.20 & pred5<0.30,"Medium_Score","High_Score"))
- table(pred6,test$churn)

Continued……

- str(test$totrev)
- quantile(test$totrev,prob=c(0.10,0.20,0.30,0.40,0.50,0.60,0.70,0.80,0.90,1))
- Revenue_Levels<-ifelse(test$totrev<670.660,"Low_Revenue",ifelse(test$totrev>=670.660 & test$totrev<1034.281,"Medium_Revenue","High_Revenue"))
- table(Revenue_Levels)
- table(pred6,Revenue_Levels)

Table of Revenue_ Levels

HIGH_ REVENUE	LOW_ REVENUE	MEDIUM_ REVENUE
5726	7766	5924

Table of customers who are to be targeted

	High_ Revenue	Low_ Revenue	Medium _ Revenue
High_ score	1687	1078	1355
Low_ score	1740	3034	2022
Medium _ score	2299	3654	2547

Continued.....

- Thus this table can be used to select the levels of customers are to be targeted and the Target list can be extracted as follows:
- test$prob_levels<-ifelse(pred5<0.20,"Low_Score",ifelse(pred5>=0.20 & pred5<0.30,"Medium_Score","High_Score"))
- test$Revenue_Levels<-ifelse(test$totrev<670.660,"Low_Revenue",ifelse(test$totrev>=670.660 & test$totrev<1034.281,"Medium_Revenue","High_Revenue"))

Continued....

- Targeted1<-test[test$prob_levels=="High_Score" & test$Revenue_Levels=="High_Revenue","Customer_ID"]
- Targeted1<-as.data.frame(Targeted1)
- nrow(Targeted1)
- **write.csv(Targeted1,"High_Revenue_Target_Customers.csv",row.names = F)**

*****

THANK YOU