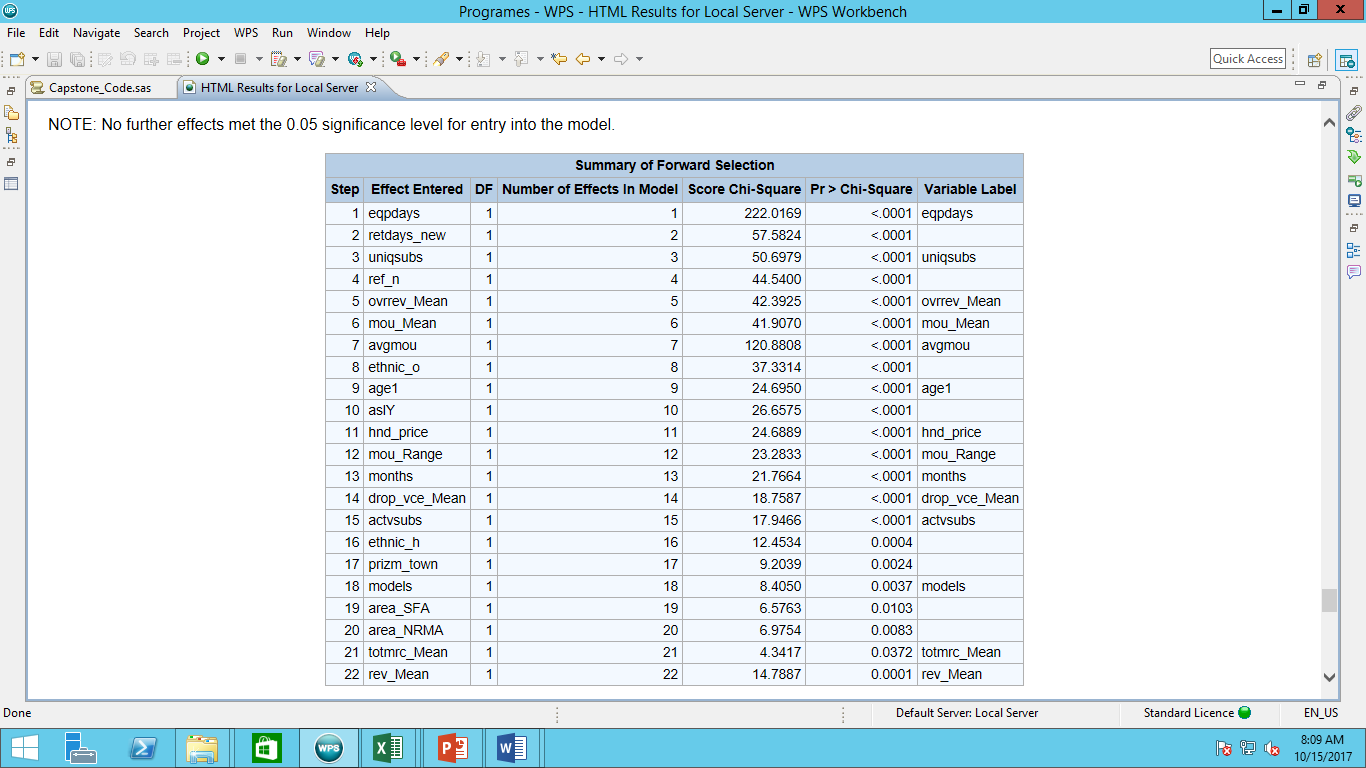
**Background**: Mobicon is a telecom company which is concerned of their higher churn rate. Churn is when a customer with an existing operator switch to a different operator. This project is aimed at developing a model which would answer a few key questions of interest to roll out targeted proactive retention programs, which would help in reducing the churn rate.

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



The top five factors driving likelihood of churn at mobicon are:

* Eqpdays - Number of days (age) of current equipment
* Retdays - Number of days since last retention call (Retention calls include any calls from the customer regarding loyalty or retention, e.g. contract renewal, relating competitor's offer, etc.)
* Uniqsubs - Number of unique subscribers in the household
* Refurb\_new - Handset: refurbished or new
* Ovrrev\_mean - Mean overage revenue (Mean overage revenue is the sum of data and voice overage revenues.)

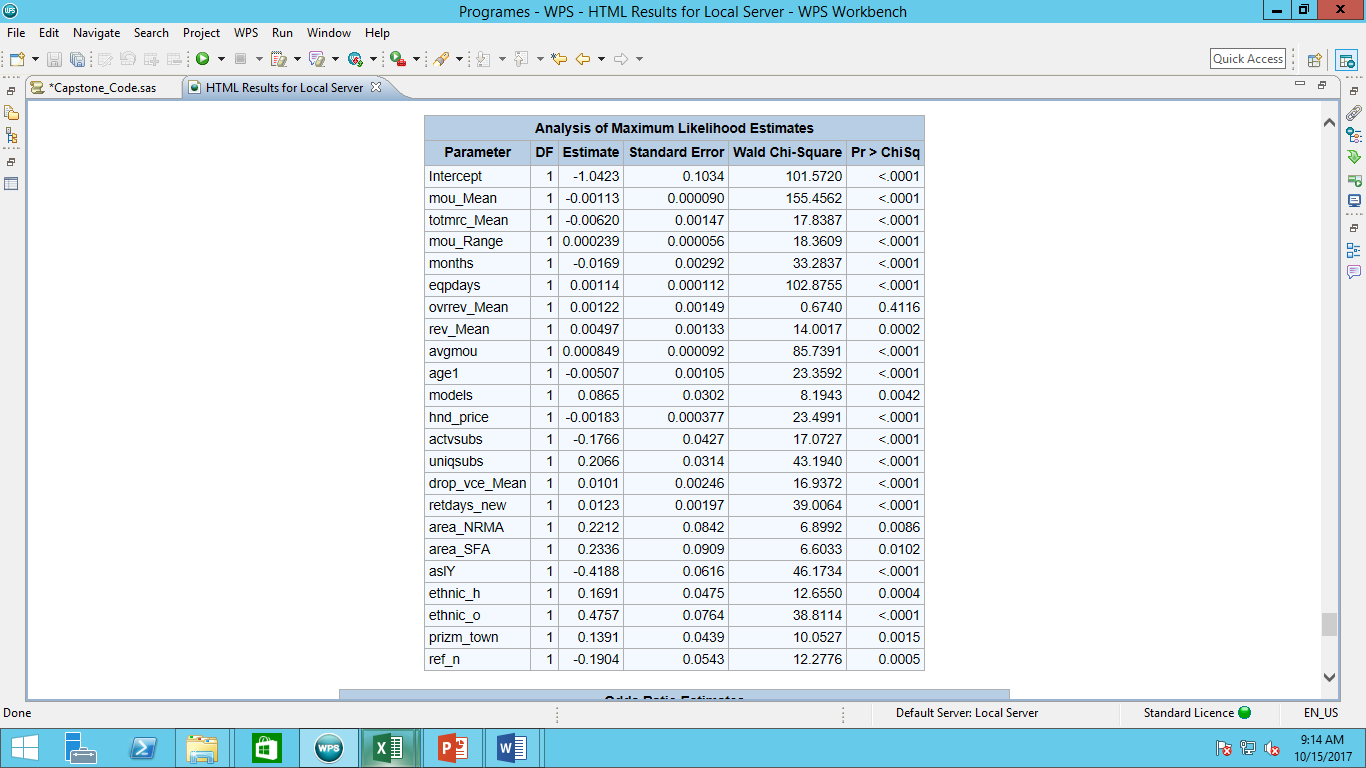
**Validation of survey findings. a) Whether “cost and billing” and “network and service quality” are important factors influencing churn behaviour. b) Are data usage connectivity issues turning out to be costly? In other words, is it leading to churn?**

1. Mean total monthly recurring charge(totmrc\_mean), Mean monthly revenue (rev\_mean) and mean overage revenue (ovrrev\_mean) are predictor variables in the following model. Therefore, cost and billing are important factors for churn behaviour.
2. Mean number of dropped (failed) voice calls (drop\_vce\_mean) is a predictor variables with p-value less than 0.0001. Therefore, network and sevice quality are important factors for churn behaviour.
3. Since Mean number of dropped data calls or Mean number of completed data calls are not predictor variables in the model, hence we can probably conclude that data usage connectivity issue is not influencing the churn behaviour.

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

Mean total monthly recurring charge(totmrc\_mean), Mean number of monthly minutes of use(Mou\_mean), Average monthly minutes of use over the life of the customer(avgmou) are predictor variables with significant affect on churn and the customer can be advised accordingly on the rate plan. Customer’s rate plan suitability is determined based on these parameters. Since mean overage revenue (ovrrev\_Mean) is also a major churn influencer, only after detailed analysis of customer’s rate plan suitability can rate plan migration be recommended to subscribers as they often switch to non-optimal rate plans to save overage costs.

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



From the analysis of predictor variables we can infer that:

* Mou\_Mean, mou\_Range and avgmou can be used to determine a customer’s rate plan suitability. Using revenue information like totmrc\_Mean, ovrrev\_Mean, rev\_mean and account spending limit a suitable plan can be constructed and recommended for proactive retention campaigns in the future.
* Since mean number of dropped calls is a predictor variable, thus ensuring better network facilities can help in retention of customers.
* Since increase in the total number of months in service reduces the probability of churning, targeted plans can be provided to those who have stayed loyal for a long time. This will ensure in decreasing the churn rate.
* Since with increase in active subscribers in a family the probability of churn rate decreases, offer bundling for family can be used as a means to increase active subscribers in a household which brings down churn rate. This will prove beneficial as referrals from family members are a deciding factor for switching carrier.

**proc** **import** datafile = "Z:\Assignments\Graded Assignment\Topic 13 - Final Case Study Implementation\telecomfinal.csv"

out=telecom dbms=csv replace;

**run**;

/\*------------------------ DATA EXPLORATION ---------------\*/

**proc** **contents** data = telecom;

**run**;

**proc** **means** n nmiss mean stddev min max data = telecom;

**run**;

**proc** **freq** data = telecom;

tables churn;

**run**;

**proc** **freq** data=telecom;

table area asl\_flag car\_buy cartype children crclscod csa div\_type

dwllsize dwlltype ethnic hnd\_webcap income mailordr mailresp marital numbcars occu1

prizm\_social\_one proptype refurb\_new retdays solflag wrkwoman;

**run**;

/\*--------------------- DATA PREPARATION ------------------------\*/

**data** telecom\_prepd(drop = car\_buy cartype children crclscod csa div\_type dwllsize dwlltype mailordr mailresp numbcars occu1 proptype wrkwoman solflag);

set telecom;

retdays\_new = *input*(retdays,8.);

**run**;

**data** telecom\_prepd(drop = retdays);

set telecom\_prepd;

if retdays\_new = . then retdays\_new = 0;

**run**;

**data** telecom\_prepd;

set telecom\_prepd;

where mou\_Mean<10000 and rev\_Range<5000 and mou\_Range<20000

and change\_mou<5000 and custcare\_Mean<150 and iwylis\_vce\_Mean<400

and ccrndmou\_Range<1000 and adjqty<90000 and rev\_Mean<1500 and ovrmou\_Mean<2500

and avg3mou<7000 and avgmou<5500 and hnd\_price<200 and opk\_dat\_Mean<200

and roam\_Mean<400 and recv\_sms\_Mean<100 and blck\_dat\_Mean<100

and mou\_pead\_Mean<300 and da\_Mean<100 and datovr\_Mean<120

and drop\_dat\_Mean<60 and adjmou<180000 and totrev<16000 and adjrev<15000

and avgrev<800 and mtrcycle != . and truck != . and forgntvl != .

and area ne "NA" and hnd\_webcap ne "UNKW" and ethnic ne "U"

and income ne "NA" and prizm\_social\_one ne "NA";

**run**;

**proc** **means** n nmiss mean stddev min max data = telecom\_prepd;

**run**;

**proc** **freq** data=telecom\_prepd;

table area asl\_flag ethnic hnd\_webcap income marital

prizm\_social\_one refurb\_new;

**run**;

**data** telecom\_prepd;

set telecom\_prepd;

if mou\_Mean=. then mou\_Mean=486.55;

if totmrc\_Mean=. then totmrc\_Mean=45.74;

if rev\_Range=. then rev\_Range=39.68;

if mou\_Range=. then mou\_Range=342.61;

if change\_mou=. then change\_mou=-8.16;

if ovrrev\_Mean=. then ovrrev\_Mean=12.13;

if rev\_Mean=. then rev\_Mean=56.52;

if ovrmou\_Mean=. then ovrmou\_Mean=36.95;

if avg6mou=. then avg6mou=478.68;

if avg6qty=. then avg6qty=170.24;

if age1=. then age1=40.57;

if age2=. then age2=27.86;

if hnd\_price=. then hnd\_price=101;

if roam\_Mean=. then roam\_Mean=1.05;

if da\_Mean=. then da\_Mean=0.84;

if da\_Range=. then da\_Range=1.54;

if datovr\_Mean=. then datovr\_Mean=0.21;

if datovr\_Range=. then datovr\_Range=0.62;

**run**;

/\*data check;

set telecom;

where truck != .;

run;

proc means n nmiss data = check;

var mou\_Mean;

run;

proc freq data = telecom\_prepd;

tables churn;

run;\*/

/\*----- Creating dummy variables ------- \*/

**data** telecom\_prepd;

set telecom\_prepd;

if area="ATLANTIC SOUTH AREA" then area\_ASA=1; else area\_ASA = 0;

if area="CALIFORNIA NORTH AREA" then area\_CNA=1; else area\_CNA = 0;

if area="CENTRAL/SOUTH TEXAS AREA" then area\_CSTA=1; else area\_CSTA = 0;

if area="CHICAGO AREA" then area\_CA=1; else area\_CA = 0;

if area="HOUSTON AREA" then area\_HA=1; else area\_HA = 0;

if area="LOS ANGELES AREA" then area\_LAA=1; else area\_LAA = 0;

if area="MIDWEST AREA" then area\_MA=1; else area\_MA = 0;

if area="NEW ENGLAND AREA" then area\_NEA=1; area\_NEA = 0;

if area="NEW YORK CITY AREA" then area\_NYCA=1; area\_NYCA = 0;

if area="NORTH FLORIDA AREA" then area\_NFA=1; else area\_NFA = 0;

if area="NORTHWEST/ROCKY MOUNTAIN AREA" then area\_NRMA=1; else area\_NRMA = 0;

if area="OHIO AREA" then area\_OA=1; else area\_OA = 0;

if area="PHILADELPHIA AREA" then area\_PA=1; else area\_PA = 0;

if area="SOUTH FLORIDA AREA" then area\_SFA=1; else area\_SFA = 0;

if area="SOUTHWEST AREA" then area\_SA=1; else area\_SA = 0;

if area="TENNESSEE AREA" then area\_TA=1; else area\_TA = 0;

if area="DALLAS AREA" then area\_DA=1; else area\_DA = 0;

if area="DC/MARYLAND/VIRGINIA AREA" then area\_DMVA=1; else area\_DMVA = 0;

if area="GREAT LAKES AREA" then area\_GLA=1; else area\_GLA = 0;

**run**;

**data** telecom\_prepd;

set telecom\_prepd;

if asl\_flag='Y' then aslY=1;

else aslY=0;

**run**;

**data** telecom\_prepd;

set telecom\_prepd;

if ethnic='B' then ethnic\_b=1; else ethnic\_b=0;

if ethnic='C' then ethnic\_c=1; else ethnic\_c=0;

if ethnic='D' then ethnic\_d=1; else ethnic\_d=0;

if ethnic='F' then ethnic\_f=1; else ethnic\_f=0;

if ethnic='G' then ethnic\_g=1; else ethnic\_g=0;

if ethnic='H' then ethnic\_h=1; else ethnic\_h=0;

if ethnic='I' then ethnic\_i=1; else ethnic\_i=0;

if ethnic='J' then ethnic\_j=1; else ethnic\_j=0;

if ethnic='M' then ethnic\_m=1; else ethnic\_m=0;

if ethnic='N' then ethnic\_n=1; else ethnic\_n=0;

if ethnic='O' then ethnic\_o=1; else ethnic\_o=0;

if ethnic='P' then ethnic\_p=1; else ethnic\_p=0;

if ethnic='R' then ethnic\_r=1; else ethnic\_r=0;

if ethnic='S' then ethnic\_s=1; else ethnic\_s=0;

if ethnic='X' then ethnic\_x=1; else ethnic\_x=0;

if ethnic='Z' then ethnic\_z=1; else ethnic\_z=0;

**run**;

**data** telecom\_prepd;

set telecom\_prepd;

if hnd\_webcap='NA' then hndna=1;

else hndna=0;

if hnd\_webcap='WC' then hndwc=1;

else hndwc=0;

if hnd\_webcap='WCMB' then hndwcmb=1;

else hndwcmb=0;

**run**;

**data** telecom\_prepd;

set telecom\_prepd;

if income='1' then income1=1;

else income1=0;

if income='2' then income2=1;

else income2=0;

if income='3' then income3=1;

else income3=0;

if income='4' then income4=1;

else income4=0;

if income='5' then income5=1;

else income5=0;

if income='6' then income6=1;

else income6=0;

if income='7' then income7=1;

else income7=0;

if income='8' then income8=1;

else income8=0;

if income='9' then income9=1;

else income9=0;

**run**;

**data** telecom\_prepd;

set telecom\_prepd;

if marital='A' then marital\_a=1;

else marital\_a=0;

if marital='B' then marital\_b=1;

else marital\_b=0;

if marital='M' then marital\_m=1;

else marital\_m=0;

if marital='S' then marital\_s=1;

else marital\_s=0;

if marital='U' then marital\_u=1;

else marital\_u=0;

**run**;

**data** telecom\_prepd;

set telecom\_prepd;

if prizm\_social\_one='C' then prizm\_city=1;

else prizm\_city=0;

if prizm\_social\_one='R' then prizm\_rural=1;

else prizm\_rural=0;

if prizm\_social\_one='S' then prizm\_sub=1;

else prizm\_sub=0;

if prizm\_social\_one='T' then prizm\_town=1;

else prizm\_town=0;

if prizm\_social\_one='U' then prizm\_urban=1;

else prizm\_urban=0;

**run**;

**data** telecom\_prepd;

set telecom\_prepd;

if refurb\_new='N' then ref\_n=1;

else ref\_n=0;

if refurb\_new='R' then ref\_r=1;

else ref\_r=0;

**run**;

/\*----------------- LOGISTIC REGRESSION ------------------\*/

**proc** **surveyselect** data = telecom\_prepd

method = srs out = telecom\_samp samprate = 0.5 outall;

**run**;

**data** telecom\_train telecom\_validation;

set telecom\_samp;

if selected = 0 then output telecom\_train;

else output telecom\_validation;

**run**;

**proc** **logistic** data = telecom\_train descending;

model churn = mou\_Mean totmrc\_Mean /\*rev\_Range\*/ mou\_Range /\*change\_mou

drop\_blk\_Mean drop\_vce\_Range owylis\_vce\_Range mou\_opkv\_Range\*/ months

/\*totcalls\*/ eqpdays /\*custcare\_Mean callwait\_Mean iwylis\_vce\_Mean

callwait\_Range ccrndmou\_Range adjqty\*/ ovrrev\_Mean rev\_Mean /\*ovrmou\_Mean

comp\_vce\_Mean plcd\_vce\_Mean avg3mou\*/ avgmou /\*avg3qty avgqty

avg6mou avg6qty\*/ age1 /\*age2\*/ models hnd\_price actvsubs uniqsubs /\*forgntvl

opk\_dat\_Mean mtrcycle\*/ truck /\*roam\_Mean recv\_sms\_Mean blck\_dat\_Mean

mou\_pead\_Mean da\_Mean da\_Range datovr\_Mean datovr\_Range drop\_dat\_Mean\*/

drop\_vce\_Mean /\*adjmou totrev adjrev avgrev\*/ retdays\_new

area\_ASA area\_CNA /\*area\_CSTA area\_CA area\_HA area\_LAA area\_MA area\_NEA

area\_NYCA area\_NFA\*/ area\_NRMA /\*area\_OA area\_PA\*/ area\_SFA /\*area\_SA area\_DA

area\_DMVA area\_GLA\*/ aslY /\*ethnic\_b\*/ ethnic\_c /\*ethnic\_d ethnic\_f ethnic\_g\*/

ethnic\_h /\*ethnic\_i ethnic\_j ethnic\_m ethnic\_n\*/ ethnic\_o /\*ethnic\_p ethnic\_r

ethnic\_s ethnic\_x\*/ /\*hndna hndwc income1 income2 income3\*/ income4 /\*income5

income6 income7 income8 marital\_a\*/ marital\_b /\*marital\_m marital\_s prizm\_city

prizm\_rural prizm\_sub\*/ prizm\_town ref\_n;

score out = telecom\_train\_score;

**run**;

**proc** **logistic** data = telecom\_validation descending;

model churn = mou\_Mean totmrc\_Mean /\*rev\_Range\*/ mou\_Range change\_mou

/\*drop\_blk\_Mean drop\_vce\_Range owylis\_vce\_Range mou\_opkv\_Range\*/ months

/\*totcalls\*/ eqpdays /\*custcare\_Mean callwait\_Mean iwylis\_vce\_Mean

callwait\_Range ccrndmou\_Range adjqty\*/ ovrrev\_Mean rev\_Mean /\*ovrmou\_Mean

comp\_vce\_Mean plcd\_vce\_Mean avg3mou\*/ avgmou /\*avg3qty avgqty

avg6mou avg6qty\*/ age1 /\*age2\*/ models hnd\_price actvsubs uniqsubs /\*forgntvl

opk\_dat\_Mean mtrcycle\*/ truck /\*roam\_Mean recv\_sms\_Mean blck\_dat\_Mean

mou\_pead\_Mean da\_Mean da\_Range datovr\_Mean datovr\_Range drop\_dat\_Mean\*/

drop\_vce\_Mean /\*adjmou totrev adjrev avgrev\*/ retdays\_new

area\_ASA area\_CNA /\*area\_CSTA area\_CA area\_HA area\_LAA area\_MA area\_NEA

area\_NYCA area\_NFA\*/ area\_NRMA /\*area\_OA area\_PA\*/ area\_SFA /\*area\_SA area\_DA

area\_DMVA area\_GLA\*/ aslY /\*ethnic\_b\*/ ethnic\_c /\*ethnic\_d ethnic\_f ethnic\_g\*/

ethnic\_h /\*ethnic\_i ethnic\_j ethnic\_m ethnic\_n\*/ ethnic\_o /\*ethnic\_p ethnic\_r

ethnic\_s ethnic\_x\*/ /\*hndna hndwc income1 income2 income3\*/ income4 /\*income5

income6 income7 income8 marital\_a\*/ marital\_b /\*marital\_m marital\_s prizm\_city

prizm\_rural prizm\_sub\*/ prizm\_town ref\_n/selection = forward;

score out = telecom\_validation\_score outroc = roc;

**run**;

**proc** **gplot** data = roc;

plot \_sensit\_\*\_1MSPEC\_;

**run**;

**proc** **rank** data = telecom\_validation\_score out = decile groups = 10 ties = mean;

var P\_1;

ranks decile;

**run**;

**proc** **sort** data = decile;

by descending P\_1;

**run**;

**proc** **means** n data = decile;

class decile;

var decile;

**run**;

**proc** **means** n data = decile;

class decile;

var churn;

where churn = 1;

**run**;

**proc** **export** data = decile (keep = churn P\_1 decile) outfile = "D:\Users\Jig14765\Programes\Jig14765\Capstone\_gainschart.csv"

dbms = csv replace;

**run**;