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

**SOLUTION:**

After many iteration obtained an optimal logistic regression model(CHURN MODEL)

And on the basis of this model found that there are many things given in market survey work or influence the churn behavior:

We can better understand through churn Model that is given below:

CHURN MODEL:   
  
Top Five factor to drive churn at MOBICOM:

1. optimal\_rate\_plan
2. asl\_flag\_N
3. eqpdays900
4. hnd\_webcap\_wc
5. comp\_plcd\_call\_low

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Maximum Likelihood Estimates** | | | | | |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| Intercept | 1 | -1.4440 | 0.1450 | 99.1127 | <.0001 |
| refurb\_new\_N | 1 | -0.2253 | 0.0371 | 36.9318 | <.0001 |
| hnd\_webcap\_wc | 1 | 0.2682 | 0.0379 | 50.0123 | <.0001 |
| asl\_flag\_N | 1 | 0.4666 | 0.0423 | 121.8527 | <.0001 |
| age1\_25 | 1 | 0.1532 | 0.0351 | 19.0840 | <.0001 |
| age2\_25 | 1 | 0.2175 | 0.0430 | 25.6379 | <.0001 |
| age2\_50 | 1 | 0.1439 | 0.0419 | 11.8069 | 0.0006 |
| prizm\_social\_one\_R | 1 | 0.1658 | 0.0612 | 7.3349 | 0.0068 |
| prizm\_social\_one\_T | 1 | 0.1384 | 0.0365 | 14.3812 | 0.0001 |
| area\_central | 1 | -0.1408 | 0.0657 | 4.5952 | 0.0321 |
| area\_midwestarea | 1 | -0.1433 | 0.0546 | 6.9034 | 0.0086 |
| area\_southflorida | 1 | 0.1919 | 0.0720 | 7.1135 | 0.0077 |
| area\_tennessee | 1 | -0.1839 | 0.0824 | 4.9830 | 0.0256 |
| months15 | 1 | 0.1849 | 0.0332 | 31.0753 | <.0001 |
| eqpdays900 | 1 | 0.3043 | 0.0352 | 74.5369 | <.0001 |
| income1\_7 | 1 | 0.1004 | 0.0404 | 6.1695 | 0.0130 |
| optimal\_rate\_plan1 | 1 | -0.5128 | 0.1289 | 15.8315 | <.0001 |
| comp\_plcd\_call\_low | 1 | 0.2282 | 0.0283 | 64.8321 | <.0001 |
| mou\_Mean400 | 1 | 0.1278 | 0.0275 | 21.5090 | <.0001 |

After Different Iteration we found few variable which having significant probability.

We can see that there are 19 variables that significantly influence the churn behavior.

refurb\_new = Handset: refurbished or new

hnd\_webcap\_wc= Handset web capability

asl\_flag\_N= Account spending limit

age1= Age of first household member

age2= Age of second household member

prizm\_social\_one= Social group letter only

area= Geographic area

months= Total number of months in service

income= Estimated income

optimal\_rate\_plan=low value of (average revenue/total revenue)\*100

non\_optimalrate\_plan= high value of (average revenue/total revenue)\*100 as rate plan not given in data set so derived from average revenue and total revenue variable.

mou\_Mean400=monthly usage below 400.

comp\_plcd\_call\_low=completed call/attempted call

**Recommendation:**

1.variables like refurb\_new\_N , area\_central area\_midwestarea, area\_tennessee having negative

influence on churn behavior, means if unit increase in these variable will impact respective amount of decrease in churn behavior.

So it seems people having new handset and people residing in central area, midwestarea ,Tennessee are loyal to Mobicom.

2.variables like optimal\_rate\_plan, non\_optimalrate\_plan, mou\_Mean400, hnd\_webcap\_wc, comp\_plcd\_call\_low ,asl\_flag\_N, age1\_25, age2\_25, prizm\_social\_one\_R, prizm\_social\_one\_T, area\_southflorida, months15, eqpdays900, income1\_7 having positive impact on churn behavior and these variables abbreviate the churn behavior.

These variable are most likelihood to churn at mobicom:

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

**SOLUTION:**

1. After many iteration obtained an optimal logistic regression model(CHURN MODEL).

And on the basis of this model found that there are many things given in market survey work or influence the churn behavior of customer’s.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Maximum Likelihood Estimates** | | | | | |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| mou\_Mean400 | 1 | 0.0562 | 0.0276 | 4.1586 | 0.0414 |
| comp\_plcd\_call\_low | 1 | 0.2205 | 0.0284 | 60.4394 | <.0001 |

Here we can see that monthly usage below 400 impact the churn at mobicom. As decrease the MOU increase the churn so both go hand in hand. Through the churn model “ cost and billing “do not impact on churn but “ monthly usage “does and it may be possible because mobicom is an established company In US costing and billing is more important for new player to be established in market.

Network and service having their importance in telecom market as we can see that derived variable comp\_plcd\_call\_low (completed call/total call) having significant influence on churn .

If completed and placed call rate is less this means network is not working as per requirement of the customers. Customers having less ratio of completed to placed call are most prone to churn.

So, Mobicom must work on network quality as it is important factor to churn.

As per service, it doesn’t have significant probability so it is not included in model like customer care mean value, waiting \_time, RETDATE variables.

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Maximum Likelihood Estimates** | | | | | |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| optimal\_rate\_plan | 1 | -0.5128 | 0.1289 | 15.8315 | <.0001 |

**Recommendation:**

As per churn Model I will recommend rate plan migration retention

Strategy because in the model optimal\_rate\_plan drives churn with negative impact

.It means as optimal\_rate\_plan will increase the churn at mobicom will decrease.

So I will recommend mobicom to rate plan migration strategy.

As non\_optimalrate\_plan and optimal rate plan is not given in dataset so a proxy for non\_optimalrtae\_plan is higher value of ( overrev\_mean/total revenue)\*100.

And optimal\_rate\_plan is lower value of ( overrev\_mean/total revenue)\*100.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

Recommendation:

In proactive retention strategy there are given three area to work on.

1.Increase the minute usage for data and voice.

2.Rate plan migration means to say switch non-optimal rate plan to optimal rate plan.

3.Bundling(package deal)

So when we look upon Churn Model the same things comes out .

1. mou\_Mean400 variables having positive impact on churn rate .It shows customer those are using less than 400 min of usage are most prone to churn so Mobicom should work on this segment of customers.
2. Rate plan migration means switching from non-optimal rate plan to optimal rate plan,

Generally optimal rate plan customers are postpaid they have fixed minutes of call for a month if usage increases maximum allowance than company charged at a premium per minutes.

And it is observed and shown in this model is optimal rate plans have significantly less odd to churn.

As optimal rate plan will increase churn behavior will decrease.

Contrary **non\_optimal rate** plan customer may be spinner they could switch network for taking advantage of offers.

1. Service and network: we can see in churn model through derived variable

**comp\_plcd\_call\_low** if completed call % is low churn at mobicom is increased and it proves customer are more concerned for “service and network”.

Apart from these factors many more factors influence the churn at mobicom

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Maximum Likelihood Estimates** | | | | | |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| Intercept | 1 | -1.4440 | 0.1450 | 99.1127 | <.0001 |
| refurb\_new\_N | 1 | -0.2253 | 0.0371 | 36.9318 | <.0001 |
| hnd\_webcap\_wc | 1 | 0.2682 | 0.0379 | 50.0123 | <.0001 |
| asl\_flag\_N | 1 | 0.4666 | 0.0423 | 121.8527 | <.0001 |
| age1\_25 | 1 | 0.1532 | 0.0351 | 19.0840 | <.0001 |
| age2\_25 | 1 | 0.2175 | 0.0430 | 25.6379 | <.0001 |
| age2\_50 | 1 | 0.1439 | 0.0419 | 11.8069 | 0.0006 |
| prizm\_social\_one\_R | 1 | 0.1658 | 0.0612 | 7.3349 | 0.0068 |
| prizm\_social\_one\_T | 1 | 0.1384 | 0.0365 | 14.3812 | 0.0001 |
| area\_central | 1 | -0.1408 | 0.0657 | 4.5952 | 0.0321 |
| area\_midwestarea | 1 | -0.1433 | 0.0546 | 6.9034 | 0.0086 |
| area\_southflorida | 1 | 0.1919 | 0.0720 | 7.1135 | 0.0077 |
| area\_tennessee | 1 | -0.1839 | 0.0824 | 4.9830 | 0.0256 |
| months15 | 1 | 0.1849 | 0.0332 | 31.0753 | <.0001 |
| eqpdays900 | 1 | 0.3043 | 0.0352 | 74.5369 | <.0001 |
| income1\_7 | 1 | 0.1004 | 0.0404 | 6.1695 | 0.0130 |
| optimal\_rate\_plan1 | 1 | -0.5128 | 0.1289 | 15.8315 | <.0001 |
| comp\_plcd\_call\_low | 1 | 0.2282 | 0.0283 | 64.8321 | <.0001 |
| mou\_Mean400 | 1 | 0.1278 | 0.0275 | 21.5090 | <.0001 |

1. Customer handset web capacity and account spending limits increase the churn at mobicom.age of first householder below 25 and age of 2nd householder age below 25 and below 50 are most prone to churn.

Social group letter only also influence the churn.

People from **central, Midwest** area and Tennessee are seems loyal customer but people of **southflorida** having churn behavior..

People of income group 7, equipments of 900days and 15 months of job are also increase the churn behavior.

So mobicom can use the churn model to priorities the customer on basis of given variables in model for their proactive retention campaign in future.

And the churn model and proactive retention strategy are justified to each other therefore Mpbicom can use both for their respective purpose.

**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.**

SOLUTION: we will use “ cluster analysis” to prioritisation of customer if “revenue saves” is also a priority besides the controlling churn.

k-means cluster analysis:

If mobicom primary priority is controlling the churn and secondary is ARPU

(average revenue per unit ).

Then weightage the variable as follows:

churn = 3\*churn;

avgrev= 2\*avgrev;

Coding :

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

out=telecomfinal dbms=csv replace;

**run**;

/\*Data exploration\*/

**proc** **contents** data=univariate;

**run**;

**proc** **univariate** data=telecomfinal;

**run**;

**proc** **means** data=telecomfinal;

**run**;

**proc** **freq** data=telecomfinal;

table numbcars occu1 refurb\_new retdays proptype prizm\_social\_one mailordr mailresp marital income hnd\_webcap ethnic dwlltype dwllsize div\_type

csa crclscod children cartype car\_buy asl\_flag area;

**run**;

**proc** **means** nmiss mean data= telecomfinal;

**data**;

/\*data prepartion\*/

/\*Outlier:

1.totmrc\_mean have negative values that shows Monthly Recurring Charge is the base cost of the calling plan regardless of actual minutes used.

but it cannot possible in telecom industry as company always charge some base cost so it is outlier .

2.Change\_Mou has some negative values as it depicts Percentage change in monthly minutes of use vs previous three month average that can be negative as some customer

might be decreased their use of minutes.

3.eqpdays has negative values as it depicts Number of days (age) of current equipment that can’t be possible so these are outlier and will replace with ZERO

4.rev\_mean has some negative values as it shows Mean monthly revenue (charge amount) and it can be possible as some customer do not generate but they are become bad debt for company there for it is not outlier.\*/

**data** telecomfinal1;

set telecomfinal;

if eqpdays=-5 then eqpdays=0;

if eqpdays =-4 then eqpdays=0;

if eqpdays=-3 then eqpdays=0;

if eqpdays=-2 then eqpdays=0;

if eqpdays=-1 then eqpdays=0;

if totmrc\_mean=-26.915 then totmrc\_mean=47.0073631;

if totmrc\_mean=-8.71 then totmrc\_mean=47.0073631;

if totmrc\_mean=-6.1675 then totmrc\_mean=47.0073631;

if totmrc\_mean =-6.05 then totmrc\_mean=47.0073631;

if totmrc\_mean=-5.33 then totmrc\_mean= 47.0073631;

if totmrc\_mean=-2.2150000 then totmrc\_mean=47.0073631;

if totmrc\_mean=-1.6133333 then totmrc\_mean=47.0073631;

if totmrc\_mean=-1.61333333 then totmrc\_mean=47.0073631;

if totmrc\_mean=-1.613333333 then totmrc\_mean=47.0073631;

if totmrc\_mean= -1.4225 then totmrc\_mean=47.0073631;

if totmrc\_mean= -0.27 then totmrc\_mean= 47.0073631;

if totmrc\_mean = -0.0825 then totmrc\_mean=47.0073631;

if totmrc\_mean = -0.08 then totmrc\_mean=47.0073631;

if totmrc\_mean =0 then totmrc\_mean=47.0073631;

**run**;

/\*Missing values ,cutoff-15%\*/

**data** telecomfinal2;

set telecomfinal1;

if mou\_Mean=. then mou\_Mean=529.42;

if totmrc\_Mean=. then totmrc\_Mean=47.01;

if rev\_Range=. then rev\_Range=44.10;

if mou\_Range=. then mou\_Range=376.51;

if change\_mou=. then change\_mou=-9.19;

if ovrrev\_Mean=. then ovrrev\_Mean=13.2153508;

if rev\_Mean=. then rev\_Mean=59.0793053;

if ovrmou\_Mean=. then ovrmou\_Mean=40.1849949;

if avg6mou=. then avg6mou=521.3638642;

if avg6qty=. then avg6qty=182.2250432;

if age1=. then age1=31.3909126;

if age2=. then age2= 21.1353135 ;

if models=. then models=2;

if hnd\_price=. then hnd\_price=105.1650564;

if forgntvl =. then forgntvl=0.0583928;

if mtrcycle=. then mtrcycle=0.0134316;

if truck=. then truck=0.1897613;

if roam\_Mean=. then roam\_Mean=1.2614875;

if da\_Mean=. then da\_Mean=0.9042070;

if da\_Range=. then da\_Range=1.6450836;

if datovr\_Mean=. then datovr\_Mean=0.2540895;

if datovr\_Range=. then datovr\_Range=0.7277170;

if eqpdays=. then eqpdays=377.1488778;

**run**;

**data** telecomfinal1;

set telecomfinal1;

if refurb\_new="NA" then refurb\_new="N";

if prizm\_social\_one="NA" then prizm\_social\_one="S";

if marital="NA" then marital="U";

if hnd\_webcap="NA" then hnd\_webcap="WCMB";

if ethnic="NA" then ethnic="N" ;

if car\_buy="NA" then car\_buy="UNKNOWN";

if area="NA" then area="NEW YORK CITY AREA";

**run**;

/\*Dummy variable: There are some continues and categorical variables therefore its converted into dummy variable\*/

**data** telecomfinal2(drop=numbcars occu1 retdays mailresp income ethnic dwlltype dwllsize div\_type children cartype );

set telecomfinal1;

if refurb\_new="N" then refurb\_new\_N=1;

else refurb\_new\_N=0;

if refurb\_new="R" then refurb\_new\_R=1;

else refurb\_new\_R=0;

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;

if hnd\_webcap="UNKW" then hnd\_webcap\_unkw=1;

else hnd\_webcap\_unkw=0;

if hnd\_webcap="WC" then hnd\_webcap\_wc=1;

else hnd\_webcap\_wc=0;

if hnd\_webcap="WCMB" then hnd\_webcap\_wcmb=1;

else hnd\_webcap\_wcmb=0;

if asl\_flag="N" then asl\_flag\_N=1;

else asl\_flag\_N=0;

if asl\_flag="Y" then asl\_flag\_Y=1;

else asl\_flag\_Y=0;

if age1 <25 then age1\_25=1;

else age1\_25=0;

if 25 =< age1 < 50 then age1\_50=1;

else age1\_50=0;

if 50=< age1 < 75 then age1\_75=1;

else age1\_75=0;

if 75 =< age1 =< 99 then age1\_99 =1;

else age1\_99=0;

if prizm\_social\_one = "C" then prizm\_social\_one\_C=1;

else prizm\_social\_one\_C=0;

if prizm\_social\_one = "R" then prizm\_social\_one\_R=1;

else prizm\_social\_one\_R=0;

if prizm\_social\_one = "S" then prizm\_social\_one\_S=1;

else prizm\_social\_one\_S=0;

if prizm\_social\_one = "T" then prizm\_social\_one\_T=1;

else prizm\_social\_one\_T=0;

if prizm\_social\_one = "U" then prizm\_social\_one\_U=1;

else prizm\_social\_one\_U=0;

if area="ATLANTIC SOUTH AREA" then area\_atlantic=1;

else area\_atlantic=0;

if area="CALIFORNIA NORTH AREA" then area\_california=1;

else area\_california=0;

if area="CENTRAL/SOUTH TEXAS AREA" then area\_central=1;

else area\_central=0;

if area="CHICAGO AREA" then area\_chicago=1;

else area\_chicago=0;

if area="DALLAS AREA" then area\_dallas=1;

else area\_dallas=0;

if area="DC/MARYLAND/VIRGINIA AREA" then area\_dc=1;

else area\_dc=0;

if area="GREAT LAKES AREA" then area\_greatlakes=1;

else area\_greatlakes=0;

if area="HOUSTON AREA" then area\_houston=1;

else area\_houston=0;

if area="LOS ANGELES AREA" then area\_losangles=1;

else area\_losangles=0;

if area="MIDWEST AREA" then area\_midwestarea=1;

else area\_midwestarea=0;

if area="NEW ENGLAND AREA" then area\_newengland=1;

else area\_newengland=0;

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

else area\_newyorkcity =0;

if area="NORTH FLORIDA AREA" then area\_northflorida=1;

else area\_northflorida=0;

if area="NORTHWEST/ROCKY MOUNTAIN AREA" then area\_northwest=1;

else area\_northwest=0;

if area="OHIO AREA" then area\_ohio=1;

else area\_ohio=0;

if area="PHILADELPHIA AREA" then area\_philadelphia=1;

else area\_philadelphia=0;

if area="SOUTH FLORIDA AREA" then area\_southflorida=1;

else area\_southflorida=0;

if area="SOUTHWEST AREA" then area\_southwest=1;

else area\_southwest=0;

if area="TENNESSEE AREA" then area\_tennessee=1;

else area\_tennessee=0;

if eqpdays < 450 then eqpdays450=1;

else eqpdays450=0;

if 450 =< eqpdays < 900then eqpdays900=1;

else eqpdays900=0;

if 900 =< eqpdays < 1350 then eqpdays1350=1;

else eqpdays1350=0;

if 1350 =< eqpdays =< 1823 then eqpdays1823=1;

else eqpdays1823=0;

if hnd\_price < 150 then hnd\_price150=1;

else hnd\_price150=0;

if 150 =< hnd\_price < 300 then hnd\_price300=1;

else hnd\_price300=0;

if 300 =< hnd\_price < 500 then hnd\_price500=1;

else hnd\_price500=0;

if months < 15 then months15=1;

else months15=0;

if 15 =< months < 30 then months30=1;

else months30=0;

if 30 =< months <45 then months45=1;

else months45=0;

if 45 =< months =< 60 then months60=1;

else months60=0;

if age2 <25 then age2\_25=1;

else age2\_25=0;

if 25 =< age2 < 50 then age2\_50=1;

else age2\_50=0;

if 50=< age2 < 75 then age2\_75=1;

else age2\_75=0;

if 75 =< age2 =< 99 then age2\_99 =1;

else age2\_99=0;

**run**;

/\*Scaling\*/

**proc** **standard** data =telecomfinal2(keep= Customer\_ID churn refurb\_new\_N marital\_U hnd\_webcap\_unkw hnd\_webcap\_wc asl\_flag\_N

age1\_25 age2\_25 age2\_50 prizm\_social\_one\_R prizm\_social\_one\_T area\_central

area\_midwestarea area\_southflorida

area\_tennessee months15 eqpdays900 hnd\_price150 hnd\_price300 avgrev )

mean=0 std= 1 out=clustering\_data;

var churn refurb\_new\_N marital\_U hnd\_webcap\_unkw hnd\_webcap\_wc asl\_flag\_N

age1\_25 age2\_25 age2\_50 prizm\_social\_one\_R prizm\_social\_one\_T area\_central

area\_midwestarea area\_southflorida

area\_tennessee months15 eqpdays900 hnd\_price150 hnd\_price300 ;

**run**;

/\*Weighning\*/

**data** clustering\_data;

set clustering\_data;

churn = 3\*churn;

avgrev= 2\*avgrev;

**run**;

/\*Cluster formation\*/

**proc** **fastclus** data=clustering\_data maxclusters=6 converge=0 maxiter=120

out =cluster\_output

outstat=cluster\_stat;

var churn refurb\_new\_N marital\_U hnd\_webcap\_unkw hnd\_webcap\_wc asl\_flag\_N

age1\_25 age2\_25 age2\_50 prizm\_social\_one\_R prizm\_social\_one\_T area\_central

area\_midwestarea area\_southflorida

area\_tennessee months15 eqpdays900 hnd\_price150 hnd\_price300 avgrev ;

**run**;

/\*Sorting and adding the cluster to orginal file \*/

**data** cluster\_output1(keep=Customer\_ID cluster);

set cluster\_output;

**run**;

**proc** **sort** data=cluster\_output1;

by Customer\_ID;

**run**;

**proc** **sort** data=telecomfinal2;

by Customer\_ID;

**run**;

**data** telecomfinal\_cluster;

merge telecomfinal2(in=a ) cluster\_output1(in=b);

by Customer\_ID;

**run**;

/\*Profiling\*/

**proc** **sort** data=telecomfinal\_cluster;

by cluster;

**run**;

**proc** **means** mean data=telecomfinal\_cluster;

var churn refurb\_new\_N marital\_U hnd\_webcap\_unkw hnd\_webcap\_wc asl\_flag\_N

age1\_25 age2\_25 age2\_50 prizm\_social\_one\_R prizm\_social\_one\_T area\_central

area\_midwestarea area\_southflorida

area\_tennessee months15 eqpdays900 hnd\_price150 hnd\_price300 avgrev ;

By cluster ;

**run**;

**proc** **means** mean std data=telecomfinal\_cluster;

var churn refurb\_new\_N marital\_U hnd\_webcap\_unkw hnd\_webcap\_wc asl\_flag\_N

age1\_25 age2\_25 age2\_50 prizm\_social\_one\_R prizm\_social\_one\_T area\_central

area\_midwestarea area\_southflorida

area\_tennessee months15 eqpdays900 hnd\_price150 hnd\_price300 avgrev ;

**run**;

/\*Formation of target clusters \*/

**DATA** telecomfinal\_1234 telecomfinal\_56;

set telecomfinal\_cluster;

if cluster=1 then output telecomfinal\_1234;

if cluster=2 then output telecomfinal\_1234;

if cluster=3 then output telecomfinal\_1234;

if cluster=4 then output telecomfinal\_1234;

if cluster=5 then output telecomfinal\_56;

if cluster=6 then output telecomfinal\_56;

**run**;

Cluster telecomfinal\_1234 contain about 25%observations and it could be appropriate target customer for Mobicom.

Deciding the number of cluster:

NUMBER OF CLUSTER:5

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster Summary** | | | | | | |
| **Cluster** | **Frequency** | **RMS Std Deviation** | **Maximum Distance from Seed to Observation** | **Distance Between Cluster Centroids** | **ratio(strenghth metric)** | **proportion of clusters** |
| 1 | 10623 | 3.1044 | 32.52 | 35.0181 | 11.2801508 | 16.0233495 |
| 2 | 399 | 19.3882 | 616.9 | 133 | 6.85984258 | 0.60183719 |
| 3 | 2980 | 5.5861 | 66.7245 | 59.5055 | 10.652423 | 4.49492436 |
| 4 | 23155 | 2.1489 | 24.4193 | 25.2621 | 11.7558286 | 34.9261656 |
| 5 | 29140 | 2.2359 | 32.3681 | 25.2621 | 11.2984033 | 43.9537234 |
|  | 66297 |  |  |  | 10.3693296 |  |

NUMBER OF CLUSTER:6

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster Summary** | | | | | | |
| **Cluster** | **Frequency** | **RMS Std Deviation** | **Maximum Distance from Seed to Observation** | **Distance Between Cluster Centroids** | **ratio(strenghth metric)** | **proportion of clusters** |
| 1 | 11004 | 2.482 | 27.5559 | 28.9834 | 11.6774376 | 16.5980361 |
| 2 | 119 | 23.7943 | 521.3 | 182.2 | 7.65729608 | 0.1794953 |
| 3 | 1169 | 6.974 | 90.8859 | 73.0695 | 10.4774161 | 1.76327737 |
| 4 | 4260 | 3.656 | 39.0685 | 41.5547 | 11.3661652 | 6.42563012 |
| 5 | 27432 | 2.1843 | 31.6385 | 23.3007 | 10.6673534 | 41.3774379 |
| 6 | 22313 | 1.9619 | 24.2474 | 23.3007 | 11.8765992 | 33.6561232 |
|  | 66297 |  |  |  | 10.6203779 |  |

NUMBER OF CLUSTER:7

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster Summary** | | | | | | |
| **Cluster** | **Frequency** | **RMS Std Deviation** | **Maximum Distance from Seed to Observation** | **Distance Between Cluster Centroids** | **ratio(strenghth metric)** | **proportion of clusters** |
| 1 | 25723 | 2.1413 | 30.924 | 21.353 | 9.97197964 | 38.799644 |
| 2 | 2126 | 4.2168 | 43.7313 | 48.0092 | 11.385221 | 3.2067816 |
| 3 | 46 | 26.7342 | 427.2 | 233.5 | 8.73413081 | 0.06938474 |
| 4 | 553 | 8.7266 | 115.8 | 86.5987 | 9.92353265 | 0.83412522 |
| 5 | 11178 | 2.092 | 24.1825 | 23.6409 | 11.3006214 | 16.8604914 |
| 6 | 21050 | 1.8158 | 23.4843 | 21.353 | 11.759555 | 31.7510596 |
| 7 | 5621 | 2.7147 | 27.8602 | 30.8948 | 11.3805577 | 8.47851336 |
|  |  |  |  |  | 10.636514 |  |

NUMBER OF CLUSTER:8

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster Summary** | | | | | | |
| **Cluster** | **Frequency** | **RMS Std Deviation** | **Maximum Distance from Seed to Observation** | **Distance Between Cluster Centroids** | **ratio(strenghth metric)** | **proportion of clusters** |
| 1 | 24288 | 2.113 | 30.3359 | 19.8853 | 9.41093232 | 36.6351419 |
| 2 | 20003 | 1.7364 | 22.3923 | 19.8853 | 11.4520272 | 30.1718026 |
| 3 | 6461 | 2.3834 | 27.0858 | 27.2174 | 11.4195687 | 9.74553901 |
| 4 | 2754 | 3.4267 | 34.8928 | 39.6212 | 11.5624945 | 4.15403412 |
| 5 | 11726 | 1.9477 | 23.9214 | 20.8279 | 10.6935873 | 17.6870748 |
| 6 | 894 | 5.6101 | 59.4371 | 64.34 | 11.4686013 | 1.34847731 |
| 7 | 153 | 10.6393 | 134.4 | 118.3 | 11.1191526 | 0.23077967 |
| 8 | 18 | 25.9523 | 312.5 | 269.1 | 10.3690232 | 0.02715055 |
|  | 66297 |  |  |  | 11.018052 |  |

On the basis of above table k=6 having fairly good strength and proportion.

Profiling of clusters:

As churn is primary and ARPU is secondary concerned area at mobicom so mainly we look upon these two variable.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster=1** | | | | | | |
| **Variable** | **Label** | **Mean** | **Mean** | **Std Dev** | **DIFF** | **Z-SCORE** |
| **churn** | **churn** | 0.23164 | 0.23921 | 0.42661 | -0.007568 | -0.01774099 |
| **refurb\_new\_N** |  | 0.82406 | 0.86173 | 0.34519 | -0.037664 | -0.10911255 |
| **marital\_U** |  | 0.45883 | 0.38447 | 0.48647 | 0.0743663 | 0.152868393 |
| **hnd\_webcap\_unkw** |  | 0.00309 | 0.00264 | 0.05131 | 0.0004502 | 0.008774135 |
| **hnd\_webcap\_wc** |  | 0.08688 | 0.12931 | 0.33555 | -0.042435 | -0.12646347 |
| **asl\_flag\_N** |  | 0.78481 | 0.84941 | 0.35766 | -0.064599 | -0.18061888 |
| **age1\_25** |  | 0.40622 | 0.3224 | 0.4674 | 0.0838182 | 0.179329471 |
| **age2\_25** |  | 0.64876 | 0.56114 | 0.49625 | 0.0876226 | 0.176568973 |
| **age2\_50** |  | 0.23846 | 0.27864 | 0.44833 | -0.040181 | -0.08962393 |
| **prizm\_social\_one\_R** |  | 0.04117 | 0.04703 | 0.21171 | -0.005864 | -0.02769875 |
| **prizm\_social\_one\_T** |  | 0.12859 | 0.14882 | 0.35591 | -0.020226 | -0.05682803 |
| **area\_central** |  | 0.04762 | 0.04328 | 0.20348 | 0.004344 | 0.021348861 |
| **area\_midwestarea** |  | 0.0588 | 0.06595 | 0.24819 | -0.007149 | -0.02880426 |
| **area\_southflorida** |  | 0.03508 | 0.03222 | 0.17658 | 0.0028595 | 0.016193637 |
| **area\_tennessee** |  | 0.02526 | 0.02759 | 0.16379 | -0.002325 | -0.01419192 |
| **months15** |  | 0.43175 | 0.42985 | 0.49506 | 0.0018986 | 0.0038351 |
| **eqpdays900** |  | 0.18929 | 0.26959 | 0.44375 | -0.080295 | -0.18094666 |
| **hnd\_price150** |  | 0.83851 | 0.88075 | 0.32409 | -0.042236 | -0.13032173 |
| **hnd\_price300** |  | 0.15667 | 0.1162 | 0.32047 | 0.0404659 | 0.126269611 |
| **avgrev** | **avgrev** | 84.2682 | 58.1085 | 36.6506 | 26.159682 |  |

1. Aveg churn of 1st cluster and population aveg churn is about the same and avegrev of cluster is higher than population avgrev.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster=2** | | | | | | |
| **Variable** | **Label** | **Mean** | **Mean** | **Std Dev** | **DIFF** | **Z-SCORE** |
| **churn** | **churn** | 0.2437 | 0.23921 | 0.42661 | 0.004486 | 0.01051581 |
| **refurb\_new\_N** |  | 0.84874 | 0.86173 | 0.34519 | -0.012989 | -0.03762823 |
| **marital\_U** |  | 0.4874 | 0.38447 | 0.48647 | 0.102928 | 0.21158042 |
| **hnd\_webcap\_unkw** |  | 0 | 0.00264 | 0.05131 | -0.00264 | -0.05144426 |
| **hnd\_webcap\_wc** |  | 0.10084 | 0.12931 | 0.33555 | -0.028472 | -0.08485147 |
| **asl\_flag\_N** |  | 0.85714 | 0.84941 | 0.35766 | 0.007738 | 0.02163501 |
| **age1\_25** |  | 0.42857 | 0.3224 | 0.4674 | 0.106174 | 0.22715918 |
| **age2\_25** |  | 0.7395 | 0.56114 | 0.49625 | 0.178354 | 0.35940312 |
| **age2\_50** |  | 0.18487 | 0.27864 | 0.44833 | -0.093766 | -0.20914391 |
| **prizm\_social\_one\_R** |  | 0.03361 | 0.04703 | 0.21171 | -0.013417 | -0.06337742 |
| **prizm\_social\_one\_T** |  | 0.09244 | 0.14882 | 0.35591 | -0.056378 | -0.15840627 |
| **area\_central** |  | 0.05042 | 0.04328 | 0.20348 | 0.007145 | 0.03511553 |
| **area\_midwestarea** |  | 0.06723 | 0.06595 | 0.24819 | 0.001281 | 0.00516219 |
| **area\_southflorida** |  | 0.03361 | 0.03222 | 0.17658 | 0.001395 | 0.00789833 |
| **area\_tennessee** |  | 0.0084 | 0.02759 | 0.16379 | -0.019185 | -0.11712896 |
| **months15** |  | 0.51261 | 0.42985 | 0.49506 | 0.082752 | 0.16715489 |
| **eqpdays900** |  | 0.08403 | 0.26959 | 0.44375 | -0.185556 | -0.41815495 |
| **hnd\_price150** |  | 0.7395 | 0.88075 | 0.32409 | -0.141253 | -0.43584982 |
| **hnd\_price300** |  | 0.2521 | 0.1162 | 0.32047 | 0.135896 | 0.42405051 |
| **avgrev** | **avgrev** | 381.072 | 58.1085 | 36.6506 | 322.9639 | 8.81195913 |

1. Cluster churn rate is higher than population churn and avgrev of cluster is much higher than population avgrev so this could be mobicom target customer.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster=3** | | | | | | |
| **Variable** | **Label** | **CMean** | **PMean** | **Std Dev** | **DIFF** | **Z-SCORE** |
| **churn** | **churn** | 0.23353 | 0.23921 | 0.42661 | -0.005679 | -0.01331 |
| **refurb\_new\_N** |  | 0.8195 | 0.86173 | 0.34519 | -0.042225 | -0.12232 |
| **marital\_U** |  | 0.48161 | 0.38447 | 0.48647 | 0.0971413 | 0.199685 |
| **hnd\_webcap\_unkw** |  | 0.00599 | 0.00264 | 0.05131 | 0.0033484 | 0.065258 |
| **hnd\_webcap\_wc** |  | 0.05988 | 0.12931 | 0.33555 | -0.069432 | -0.20692 |
| **asl\_flag\_N** |  | 0.78272 | 0.84941 | 0.35766 | -0.066685 | -0.18645 |
| **age1\_25** |  | 0.42002 | 0.3224 | 0.4674 | 0.0976194 | 0.208857 |
| **age2\_25** |  | 0.67151 | 0.56114 | 0.49625 | 0.1103726 | 0.222413 |
| **age2\_50** |  | 0.2361 | 0.27864 | 0.44833 | -0.042541 | -0.09489 |
| **prizm\_social\_one\_R** |  | 0.05047 | 0.04703 | 0.21171 | 0.0034397 | 0.016248 |
| **prizm\_social\_one\_T** |  | 0.11206 | 0.14882 | 0.35591 | -0.036754 | -0.10327 |
| **area\_central** |  | 0.05475 | 0.04328 | 0.20348 | 0.0114726 | 0.056383 |
| **area\_midwestarea** |  | 0.05218 | 0.06595 | 0.24819 | -0.013764 | -0.05546 |
| **area\_southflorida** |  | 0.03849 | 0.03222 | 0.17658 | 0.0062757 | 0.03554 |
| **area\_tennessee** |  | 0.02908 | 0.02759 | 0.16379 | 0.0014967 | 0.009138 |
| **months15** |  | 0.4722 | 0.42985 | 0.49506 | 0.042345 | 0.085535 |
| **eqpdays900** |  | 0.11206 | 0.26959 | 0.44375 | -0.157528 | -0.35499 |
| **hnd\_price150** |  | 0.76903 | 0.88075 | 0.32409 | -0.111715 | -0.34471 |
| **hnd\_price300** |  | 0.21215 | 0.1162 | 0.32047 | 0.0959427 | 0.299379 |
| **avgrev** | **avgrev** | 198.891 | 58.1085 | 36.6506 | 140.78284 | 3.841211 |

1. cluster 3churn rate and population churn rate is about the same and avgrev of cluster is higher than population mean. This could be also target customer of mobicom.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster=4** | | | | | | |
| **Variable** | **Label** | **Mean** | **Mean** | **Std Dev** | **DIFF** | **Z-SCORE** |
| **churn** | **churn** | 0.22559 | 0.23921 | 0.42661 | -0.01362 | -0.03194 |
| **refurb\_new\_N** |  | 0.8108 | 0.86173 | 0.34519 | -0.05093 | -0.14754 |
| **marital\_U** |  | 0.4831 | 0.38447 | 0.48647 | 0.098632 | 0.202749 |
| **hnd\_webcap\_unkw** |  | 0.00399 | 0.00264 | 0.05131 | 0.001351 | 0.02633 |
| **hnd\_webcap\_wc** |  | 0.06643 | 0.12931 | 0.33555 | -0.06288 | -0.1874 |
| **asl\_flag\_N** |  | 0.76033 | 0.84941 | 0.35766 | -0.08908 | -0.24906 |
| **age1\_25** |  | 0.42817 | 0.3224 | 0.4674 | 0.105771 | 0.226298 |
| **age2\_25** |  | 0.66784 | 0.56114 | 0.49625 | 0.106699 | 0.21501 |
| **age2\_50** |  | 0.23239 | 0.27864 | 0.44833 | -0.04625 | -0.10315 |
| **prizm\_social\_one\_R** |  | 0.04648 | 0.04703 | 0.21171 | -0.00055 | -0.00261 |
| **prizm\_social\_one\_T** |  | 0.13075 | 0.14882 | 0.35591 | -0.01806 | -0.05075 |
| **area\_central** |  | 0.04272 | 0.04328 | 0.20348 | -0.00055 | -0.00271 |
| **area\_midwestarea** |  | 0.05258 | 0.06595 | 0.24819 | -0.01336 | -0.05384 |
| **area\_southflorida** |  | 0.03897 | 0.03222 | 0.17658 | 0.006748 | 0.038217 |
| **area\_tennessee** |  | 0.02582 | 0.02759 | 0.16379 | -0.00177 | -0.01078 |
| **months15** |  | 0.46127 | 0.42985 | 0.49506 | 0.031414 | 0.063455 |
| **eqpdays900** |  | 0.13615 | 0.26959 | 0.44375 | -0.13344 | -0.30071 |
| **hnd\_price150** |  | 0.80376 | 0.88075 | 0.32409 | -0.07699 | -0.23757 |
| **hnd\_price300** |  | 0.18662 | 0.1162 | 0.32047 | 0.070415 | 0.219724 |
| **avgrev** | **avgrev** | 125.822 | 58.1085 | 36.6506 | 67.71365 | 1.847544 |

1. Cluster 4th churn rate is a bit a lower than population churn rate and avgrev of cluster is much higher than population avgrev.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster=5** | | | | | | |
| **Variable** | **Label** | **Mean** | **Mean** | **Std Dev** | **DIFF** | **Z-SCORE** |
| **churn** | **churn** | 0.2379 | 0.23921 | 0.42661 | -0.001314 | -0.00308036 |
| **refurb\_new\_N** |  | 0.88871 | 0.86173 | 0.34519 | 0.0269783 | 0.078155473 |
| **marital\_U** |  | 0.3139 | 0.38447 | 0.48647 | -0.070563 | -0.1450511 |
| **hnd\_webcap\_unkw** |  | 0.00204 | 0.00264 | 0.05131 | -0.000598 | -0.01165857 |
| **hnd\_webcap\_wc** |  | 0.1706 | 0.12931 | 0.33555 | 0.0412917 | 0.123057689 |
| **asl\_flag\_N** |  | 0.91645 | 0.84941 | 0.35766 | 0.0670429 | 0.187450581 |
| **age1\_25** |  | 0.24067 | 0.3224 | 0.4674 | -0.08173 | -0.17486154 |
| **age2\_25** |  | 0.4751 | 0.56114 | 0.49625 | -0.086039 | -0.17337865 |
| **age2\_50** |  | 0.31343 | 0.27864 | 0.44833 | 0.0347895 | 0.077597386 |
| **prizm\_social\_one\_R** |  | 0.05089 | 0.04703 | 0.21171 | 0.0038587 | 0.018226666 |
| **prizm\_social\_one\_T** |  | 0.16597 | 0.14882 | 0.35591 | 0.0171588 | 0.048211214 |
| **area\_central** |  | 0.03926 | 0.04328 | 0.20348 | -0.004014 | -0.01972853 |
| **area\_midwestarea** |  | 0.07291 | 0.06595 | 0.24819 | 0.0069619 | 0.0280508 |
| **area\_southflorida** |  | 0.03084 | 0.03222 | 0.17658 | -0.001379 | -0.00780828 |
| **area\_tennessee** |  | 0.02993 | 0.02759 | 0.16379 | 0.0023406 | 0.014290215 |
| **months15** |  | 0.41805 | 0.42985 | 0.49506 | -0.011802 | -0.02383878 |
| **eqpdays900** |  | 0.33938 | 0.26959 | 0.44375 | 0.0697948 | 0.157284021 |
| **hnd\_price150** |  | 0.91754 | 0.88075 | 0.32409 | 0.0367928 | 0.113527751 |
| **hnd\_price300** |  | 0.08133 | 0.1162 | 0.32047 | -0.034876 | -0.10882691 |
| **avgrev** | **avgrev** | 31.994 | 58.1085 | 36.6506 | -26.11456 | -0.71252663 |

1. cluster 5th churn rate is less than population mean and avgrate as well.

So mobicom must exclude these population from their subscriber pool.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Cluster=6** | | | | | | |
| **Variable** | **Label** | **Mean** | **Mean** | **Std Dev** | **DIFF** | **Z-SCORE** |
| **churn** | **churn** | 0.24743 | 0.23921 | 0.42661 | 0.008223 | 0.01927496 |
| **refurb\_new\_N** |  | 0.85914 | 0.86173 | 0.34519 | -0.002588 | -0.00749708 |
| **marital\_U** |  | 0.41007 | 0.38447 | 0.48647 | 0.025608 | 0.05263995 |
| **hnd\_webcap\_unkw** |  | 0.00273 | 0.00264 | 0.05131 | 9.42E-05 | 0.0018359 |
| **hnd\_webcap\_wc** |  | 0.11527 | 0.12931 | 0.33555 | -0.014043 | -0.04185071 |
| **asl\_flag\_N** |  | 0.8193 | 0.84941 | 0.35766 | -0.030107 | -0.084178 |
| **age1\_25** |  | 0.35567 | 0.3224 | 0.4674 | 0.033269 | 0.07118005 |
| **age2\_25** |  | 0.5966 | 0.56114 | 0.49625 | 0.035461 | 0.07145854 |
| **age2\_50** |  | 0.26724 | 0.27864 | 0.44833 | -0.011397 | -0.02542037 |
| **prizm\_social\_one\_R** |  | 0.04518 | 0.04703 | 0.21171 | -0.001855 | -0.00876356 |
| **prizm\_social\_one\_T** |  | 0.14337 | 0.14882 | 0.35591 | -0.005446 | -0.01530139 |
| **area\_central** |  | 0.04553 | 0.04328 | 0.20348 | 0.002259 | 0.011102 |
| **area\_midwestarea** |  | 0.06418 | 0.06595 | 0.24819 | -0.001768 | -0.0071232 |
| **area\_southflorida** |  | 0.03088 | 0.03222 | 0.17658 | -0.00134 | -0.00758742 |
| **area\_tennessee** |  | 0.02622 | 0.02759 | 0.16379 | -0.00137 | -0.00836496 |
| **months15** |  | 0.43477 | 0.42985 | 0.49506 | 0.004916 | 0.00992912 |
| **eqpdays900** |  | 0.2581 | 0.26959 | 0.44375 | -0.011489 | -0.02589115 |
| **hnd\_price150** |  | 0.87765 | 0.88075 | 0.32409 | -0.003099 | -0.00956226 |
| **hnd\_price300** |  | 0.11993 | 0.1162 | 0.32047 | 0.003726 | 0.01162566 |
| **avgrev** | **avgrev** | 55.2871 | 58.1085 | 36.6506 | -2.821434 | -0.07698186 |

1. Cluster 6th churn rate more than population churn rate but avgrev is less than population therefore It should also excluded from subscriber pool as well.

As per above observation and financial constraint at mobicom cluster 1,2,3,4 should be target customer for mobicom.

These cluster also collectively satisfied the criteria of 20% of subscriber pool as well

That’s how mobicom has their target customer to implement their proactive retention campaign.