```sas
libname churn "Y:\FINAL_CASE _STUDY_CHURN";


proc import datafile="Y:\FINAL_CASE _STUDY_CHURN\telecomfinal.csv"
out=churn.telecom dbms=csv replace;
run;


proc contents data=churn.telecom;
run;


proc means n nmiss mean data=churn.telecom;
run;
/* missing value for character variables

proc format;
value $missfmt ' ' = 'Missing' other = 'Not Missing';
value missfmt . = 'Missing' other = 'Not Missing';
run;

proc freq data= churn.telecom;
format income $missfmt.;
tables income / missing missprint nocum nopercent;
format _NUMERIC_ missfmt.;
tables _NUMERIC_ /missing missprint nocum nopercent;
run;
*/
proc freq data= churn.telecom;
tables income retdays 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;


/* deleting missing value */
data churn.telecom1;
set churn.telecom;
if mou_Mean = . or totmrc_Mean = . or rev_range = . or mou_Range = . or
change_mou = . or  ovrrev_Mean = . or
rev_Mean = . or ovrmou_Mean = . or avg6mou = . or avg6qty = . or age1 = . or
age2 = . or hnd_price = . or
forgntvl = . or mtrcycle = . or truck = . or roam_Mean = . or da_Mean = . or
da_Range = . or datovr_Mean = . or
datovr_Range = .
then delete ;
run;


data churn.telecom2;
set churn.telecom1;
if income=' ' or retdays=' '
then delete;
run;


PROC freq DATA= churn.telecom;
tables churn;
run;
```

```sas
proc means nmiss data=churn.telecom2;
run;
/*checking whether the missing values are deleted or not*/
proc means nmiss n mean data=churn.telecom1;
run;

/* IMPUTING MISSING VALUE USING PROC MI USING MLE(MAXIMUM LIKELIHOOD
ESTIMATION)
PROC MI DATA =churn.telecom1;
EM OUT= churn.telecom_MI;
VAR mou_Mean totmrc_Mean rev_range mou_Range change_mou ovrrev_Mean rev_Mean
ovrmou_Mean rev_Mean  avg6mou  avg6qty age1 age2
hnd_price forgntvl mtrcycle truck roam_Mean da_Mean da_Range datovr_Mean
datovr_range;
RUN;


PROC MEANS NMISS N MEAN DATA= churn.TELECOM_MI;
RUN;

DATA churn.telecom_1;
set churn.telecom1;
merge_1=1;
run;

data churn.TELECOM_MI;
set churn.TELECOM_MI;
merge_1=1;
run;

data churn.TELECOM2;
merge churn.telecom_1 churn.TELECOM_MI;
by merge_1;
run;
/* PROC FREQ
proc freq data=churn.telecom1;
tables income /nocum nopercent nocol;
run;


proc means n data=churn.telecom2;
run;

*/

/*IMPUTING MEANS MISSING VALUE FOR CHARACTER VARIABLE*/
data churn.TELECOM2;
set churn.TELECOM2;
income1=input(income,10.);
retdays1=input(retdays,5.);
run;

PROC MEANS NMISS DATA=churn.TELECOM2;
RUN;

data churn.TELECOM3(drop=income retdays);
set churn.TELECOM2;
```

```
run;



/* renaming variable as it was earlier*/
data churn.TELECOM4;
set churn.TELECOM3;
rename income1=income;
rename retdays1=retdays;
run;



proc means  data=churn.TELECOM4;
RUN;



/* calculating mean further to be stored as missing values for these
variables*/

proc means mean data=churn.TELECOM4;
var avg6mou avg6qty income retdays;
output out=churn.TELECOM4_MEAN
mean(avg6mou)= mean_avg6mou
mean(avg6qty)= mean_avg6qty
mean(income)= mean_income
mean(retdays)= mean_retdays;
run;

PROC MEANS DATA=churn.TELECOM4_MEAN;
run;



proc means nmiss data=churn.TELECOM4_MEAN;
RUN;



/* preparing datasets to be merged*/
data churn.TELECOM5;
set churn.TELECOM4;
merge1=1;
run;

data churn.TELECOM4_MEAN;
set churn.TELECOM4;
merge1=1;
run;



/*merging both dataset using one to one merge*/
data churn.TELECOM6;
merge churn.TELECOM4_MEAN churn.TELECOM5;
BY merge1;
run;
/*imputing mean values to all missing values*/
```

```sas
data churn.TELECOM7;
set churn.TELECOM6;
if avg6mou=. then avg6mou1=round(mean_avg6mou,0.01);
else avg6mou1=avgmou;
if avg6qty=. then avg6qty1=round(mean_avg6qty,0.01);
else avg6qty1=avg6qty;
if income=. then income1=round(mean_income,0.01);
else income1=income;
if retdays=. then retdays1=round(mean_retdays,0.01);
else retdays1=retdays;
run;


data churn.TELECOM8(drop=avg6mou avg6qty income retdays mean_avg6mou
mean_avg6qty
mean_income mean_retdays merge1);
set churn.TELECOM7;
run;

proc means nmiss data=churn.telecom7;
run;


data churn.TELECOM9;
rename avg6mou1=avg6mou avg6qty1=avg6qty income1=income retdays1=retdays;
set churn.TELECOM8;
run;

proc means nmiss mean data=churn.TELECOM9;
run;

proc freq data= churn.telecom;
tables  income retdays 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 churn.backup;
set churn.TELECOM9;
run;
/*------------------------------------------------ DATA PREPRATION----------
-------------------------------------------------------------------------------
-
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-
------------------------------------------------ CREATING DUMMY VARIABLES---
-------------------------------------------------------------------------------
*/




proc means data=churn.TELECOM9;
```

```sas
run;
/*1. asl_flag  */
data churn.telecom_dummy;
set churn.TELECOM9;
if asl_flag='Y' then asl_flag_y=1;
else asl_flag_y=0;
*run;

/*2.prizm_social_one*/

if prizm_social_one='C' then city_area=1;
else city_area=0;
if prizm_social_one='R' then rural_area=1;
else rural_area=0;
if prizm_social_one='S' then suburban_area=1;
else suburban_area=0;
if prizm_social_one='T' then town_area=1;
else town_area=0;
if prizm_social_one='U' then urban_area=1;
else urban_area=0;
if prizm_social_one='NA' then NA_area=1;
else NA_area=0;
*run;

*/3.refurb_new;
if refurb_new='N' then handset_new=1;
else handset_new=0;
/*if refurb_new='R' then handset_refurb=1;
else handset_refurb=0;*/
*run;


*/4.marital;
if marital='M' then marital_status=1;
else marital_status=0;
if marital='A' then marital_inferred=1;
else marital_inferred=0;
if marital='B' then marital_inferred_single=1;
else marital_inferred_single=0;
/*if marital='S' then marital_unmarried=1;
else marital_unmarried=0;*/
if marital='U' then marital_unknown=1;
else marital_unknown=0;
*run

*/5.hnd_webcap;
if hnd_webcap='WC' then hnd_webcapable=1;
else hnd_webcapable=0;
if hnd_webcap='UNKW' then hnd_unkw=1;
else hnd_unkw=0;
/*if hnd_webcap='WCMB' then hnd_webcapable_mini=1;
else hnd_webcapable_mini=0;*/
if hnd_webcap='NA' then hnd_na=1;
else hnd_na=0;
*run;
```

```
/*6.dwlltype */
if dwlltype='M' then dwll_multiple=1;
else dwll_multiple=0;
if dwlltype='S' then dwll_single=1;
else dwll_single=0;
/*if dwlltype='NA' then dwll_unknown=1;
else dwll_unknown=0;*/
*run;

/*7.dwllsize */
if index(dwllsize,'A') > 0 then dwllsize_1=1;
else dwllsize_1=0;

if index(dwllsize,'B') > 0 then dwllsize_2=1;
else dwllsize_2=0;

if index(dwllsize,'C') > 0 then dwllsize_3=1;
else dwllsize_3=0;

if index(dwllsize,'D') > 0 then dwllsize_4=1;
else dwllsize_4=0;

if index(dwllsize,'E') > 0 then dwllsize_5=1;
else dwllsize_5=0;

if index(dwllsize,'F') > 0 then dwllsize_6=1;
else dwllsize_6=0;

if index(dwllsize,'G') > 0 then dwllsize_7=1;
else dwllsize_7=0;

if index(dwllsize,'H') > 0 then dwllsize_8=1;
else dwllsize_8=0;

if index(dwllsize,'I') > 0 then dwllsize_9=1;
else dwllsize_9=0;

if index(dwllsize,'J') > 0 then dwllsize_J=1;
else dwllsize_J=0;

if index(dwllsize,'K') > 0 then dwllsize_K=1;
else dwllsize_K=0;

if index(dwllsize,'L') > 0 then dwllsize_L=1;
else dwllsize_L=0;

if index(dwllsize,'M') > 0 then dwllsize_M=1;
else dwllsize_M=0;

if index(dwllsize,'N') > 0 then dwllsize_N=1;
else dwllsize_N=0;

/*if index(dwllsize,'O') > 0 then dwllsize_O=1;
else dwllsize_O=0;*/

if index(dwllsize,'NA') > 0 then dwllsize_NA=1;
else dwllsize_NA=0;
```

```sas
*run;

/*8.occu1 */

if index(occu1,'1') > 0 then occu_tech=1;
else occu_tech=0;
if index(occu1,'2') > 0 then occu_admin=1;
else occu_admin=0;
if index(occu1,'3') > 0 then occu_sales=1;
else occu_sales=0;
if index(occu1,'4') > 0 then occu_clercial=1;
else occu_clercial=0;

if index(occu1,'5') > 0 then occu_craftsman=1;
else occu_craftsman=0;
if index(occu1,'6') > 0 then occu_student=1;
else occu_student=0;
if index(occu1,'7') > 0 then occu_homemaker=1;
else occu_homemaker=0;
if index(occu1,'8') > 0 then occu_retires=1;
else occu_retires=0;

if index(occu1,'9') > 0 then occu_farmer=1;
else occu_farmer=0;

if index(occu1,'A') > 0 then occu_military=1;
else occu_military=0;
if index(occu1,'B') > 0 then occu_religious=1;
else occu_religious=0;
if index(occu1,'C') > 0 or index(occu1,'D') > 0 or index(occu1,'E')>0
or index(occu1,'F') > 0 or index(occu1,'G') > 0 or   index(occu1,'H') > 0
or index(occu1,'I') > 0 or index(occu1,'J') > 0 or index(occu1,'K') > 0
or index(occu1,'L') > 0   then occu_self=1;
else occu_self=0;
*run;
/*9.numbcars */
if index(numbcars,'1')>0 then numbcars_1=1;
else numbcars_1=0;
if index(numbcars,'2')>0 then numbcars_2=1;
else numbcars_2=0;
/*if index(numbcars,'3')>0 then numbcars_3=1;
else numbcars_3=0;*/
if index(numbcars,'NA')>0 then numbcars_NA=1;
else numbcars_NA=0;
*RUN;

/*10.cartype */
if index(cartype,'A')>0 then cartype_luxury=1;
else cartype_luxury=0;
if index(cartype,'B')>0 then cartype_truck=1;
else cartype_truck=0;
if index(cartype,'C')>0 then cartype_suv=1;
else cartype_suv=0;
if index(cartype,'D')>0 then cartype_mini=1;
else cartype_mini=0;
if index(cartype,'E')>0 then cartype_regular=1;
else cartype_regular=0;
```

```sas
if index(cartype,'F')>0 then cartype_upper=1;
else cartype_upper=0;
/*if index(cartype,'G')>0 then cartype_basic=1;
else cartype_basic=0;
/*if index(cartype,'NA')>0 then cartype_NA=1;
else cartype_NA=0;*/

/*11.children */
if index(children,'Y')>0 then children_yes=1;
else children_yes=0;
/*if index(children,'N')>0 then children_no=1;
else children_no=0;*/
if index(children,'NA')>0 then children_na=1;
else children_na=0;
*run;

/*12. carbuy */
if car_buy="New" then car_new=1; else car_new=0;
/*if car_buy="UNKNOWN" then car_unk=1; esle car_unk=0; */

/*if index(car_buy,'New')>0 then car_new=1;
else car_new=0;
if index(car_buy,'NA')>0  then car_na=1; else car_na=0;
/*if index(car_buy,'UNKNOWN')>0 then car_unk=1;
else car_unk=0;*/
*run;

/*13. crclscod*/
if substr(crclscod,1,1)="A" or substr(crclscod,1,1)="B" or
substr(crclscod,1,1)="C" or substr(crclscod,1,1)="D" or
substr(crclscod,1,1)="E"
or substr(crclscod,1,1)="F"
then crclscod_good=1;
else crclscod_good=0;

 if substr(crclscod,1,1)="G" or substr(crclscod,1,1)="H" or
substr(crclscod,1,1)="I" or substr(crclscod,1,1)="J"
 or substr(crclscod,1,1)="K"
 then crclscod_satisfactory=1;
 else crclscod_satisfactory=0;

 if substr(crclscod,1,1)="L" or substr(crclscod,1,1)="M" or
substr(crclscod,1,1)="N" or substr(crclscod,1,1)="O"
 or substr(crclscod,1,1)="P"
 then crclscod_avg=1;
 else crclscod_avg=0;

 if substr(crclscod,1,1)="Q" or substr(crclscod,1,1)="R" or
substr(crclscod,1,1)="S" or substr(crclscod,1,1)="T"
 or substr(crclscod,1,1)="U" or substr(crclscod,1,1)="V" or
substr(crclscod,1,1)="W" or substr(crclscod,1,1)="X"
 or substr(crclscod,1,1)="Y" /*or substr(crclscod,1,1)="Z"*/
 then crclscod_bad=1;
 else crclscod_bad=0;
 *run;

 /*14 csa*/
```

```sas
    if substr(csa,1,1)="A" or substr(csa,1,1)="B" or substr(csa,1,1)="C"
    or substr(csa,1,1)="D" or substr(csa,1,1)="E"
    then csa_1=1;
    else csa_1=0;

    if substr(csa,1,1)="F" or substr(csa,1,1)="G" or substr(csa,1,1)="H"
    or substr(csa,1,1)="I" or substr(csa,1,1)="J"
    then csa_2=0;
    else csa_2=1;

    if substr(csa,1,1)="K" or substr(csa,1,1)="L" or substr(csa,1,1)="M"
    or substr(csa,1,1)="N" or substr(csa,1,1)="O"
    then csa_3=1;
    else csa_3=0;

    if substr(csa,1,1)="P" or substr(csa,1,1)="S" /*or substr(csa,1,1)="V"*/
    then csa_4=1;
    else csa_4=0;
    *run;




     /* 15 div_type*/
    if index(div_type,'BTH')>0 then div_long_local=1;
    else div_long_local=0;
    if index(div_type,'LDD')>0 then div_long=1;
    else div_long=0;
    if index(div_type,'LTD')>0 then div_local=1;
    else div_local=0;
    /*if index(div_type,'NA')>0 then div_NA=1;
    else div_NA=0;
    *run;




     /*16.ethnic*/

    if index(ethnic,'B')>0 then ethnic_asian_nonor=1;
    else ethnic_asian_nonor=0;
    if index(ethnic,'D')>0 then ethnic_south_european=1;
    else ethnic_south_european=0;
    if index(ethnic,'F')>0 then ethnic_french=1;
    else ethnic_french=0;
    if index(ethnic,'G')>0 then ethnic_german=1;
    else ethnic_german=0;
    if index(ethnic,'H')>0 then ethnic_hispanic=1;
    else ethnic_hispanic=0;
    if index(ethnic,'I')>0 then ethnic_italian=1;
    else ethnic_italian=0;
    if index(ethnic,'J')>0 then ethnic_jewish=1;
    else ethnic_jewish=0;
    if index(ethnic,'M')>0 then ethnic_misc=1;
```

```
else ethnic_misc=0;
if index(ethnic,'N')>0 then ethnic_North_european=1;
else ethnic_North_european=0;
if index(ethnic,'O')>0 then ethnic_asian=1;
else ethnic_asian=0;
if index(ethnic,'P')>0 then ethnic_polynesia=1;
else ethnic_polynesia=0;
if index(ethnic,'R')>0 then ethnic_arab=1;
else ethnic_arab=0;
if index(ethnic,'S')>0 then ethnic_scot=1;
else ethnic_scot=0;
if index(ethnic,'Z')>0 then ethnic_afro_american=1;
else ethnic_afro_american=0;
if index(ethnic,'NA')>0 or index(ethnic,'U')>0 then ethnic_unknown=1;
else ethnic_unknown=0;
*run;

/* 17mailordr.*/
if index(mailordr,'B')>0 then mailordr_buy=1;
else mailordr_buy=0;
/*if index(mailordr,'NA')>0 then mailordr_NA=1;
else mailordr_NA=0;
*run;

/*18 wrkwoman*/
if index(wrkwoman,'Y')>0 then wrkwoman_yes=1;
else wrkwoman_yes=0;
/*if index(wrkwoman,'NA')>0 then wrkwoman_na=1;
else wrkwoman_na=0;
*run;

/*19 proptype*/
if index(proptype,'A')>0 then proptype_sing_family=1;
else proptype_sing_family=0;
if index(proptype,'B')>0 then proptype_condominium=1;
else proptype_condominium=0;
if index(proptype,'D')>0 then proptype_duplex=1;
else proptype_duplex=0;
if index(proptype,'E')>0 then proptype_misc=1;
else proptype_misc=0;
if index(proptype,'G')>0 then proptype_apartment=1;
else proptype_apartment=0;
/*if index(proptype,'M')>0 then proptype_mobile=1;
else proptype_mobile=0;
*run;

/*20.Area*/
if index(area,'ATLANTIC SOUTH AREA')>0 then area_atlantic=1;
else area_atlantic=0;
if index(area,'CALIFORNIA NORTH AREA')>0 then area_CAL=1;
else area_CAL=0;
if index(area,'CENTRAL/SOUTH TEXAS AREA')>0 then area_texas=1;
else area_texas=0;
if index(area,'CHICAGO AREA')>0 then area_CHICAGO=1;
else area_CHICAGO=0;
if index(area,'DALLAS AREA')>0 then area_DALLAS=1;
else area_DALLAS=0;
```

```sas
if index(area,'DC/MARYLAND/VIRGINIA AREA')>0 then area_DC=1;
else area_DC=0;
if index(area,'GREAT LAKES AREA')>0 then area_GRTLAKES=1;
else area_GRTLAKES=0;
if index(area,'HOUSTAN AREA')>0 then area_HOUSTAN=1;
else area_HOUSTAN=0;
if index(area,'LOS ANGELES AREA')>0 then area_LOS=1;
else area_LOS=0;
if index(area,'MIDWEST AREA')>0 then area_MIDWEST=1;
else area_MIDWEST=0;
if index(area,'NEW ENGLAND AREA')>0 then area_ENGLAND=1;
else area_ENGLAND=0;
if index(area,'NEW YORK CITY AREA')>0 then area_NEWYORK=1;
else area_NEWYORK=0;
if index(area,'NORTH FLORIDA AREA')>0 then area_FLORIDA=1;
else area_FLORIDA=0;
if index(area,'NORTHWEST/ROCKY MOUNTAIN AREA')>0 then area_NORTHWEST=1;
else area_NORTHWEST=0;
if index(area,'OHIO AREA')>0 then area_OHIO=1;
else area_OHIO=0;
if index(area,'PHILADELPHIA AREA')>0 then area_phily=1;
else area_phily=0;
if index(area,'SOUTH FLORIDA AREA')>0 then area_SOUTHFLORIDA=1;
else area_SOUTHFLORIDA=0;
if index(area,'SOUTHWEST AREA')>0 then area_swest=1;
else area_swest=0;
if index(area,'TENNESSEE AREA')>0 then area_tenesse=1;
else area_tenesse=0;
*run;



/*21. mailresp*/
if index(mailresp,'R')>0 then mailresp_yes=1;
else mailresp_yes=0;
/*if index(mailresp,'NA')>0 then mailresp_NA=1;
else mailresp_NA=0;
*run;

/*22.solflag*/
if index(solflag,'Y')>0 then solflag_yes=1;
else solflag_yes=0;
/*if index(solflag,'N')>0 then solflag_no=1;
else solflag_no=0;*/
if index(solflag,'NA')>0 then solflag_na=1;
else solflag_na=0;
run;

/*-------------TAKING BACKUP-------------------------------------*/
data churn.telecom_dummy_bckup;
set churn.telecom_dummy;
run;

proc means data=churn.telecom_dummy_bckup;
run;
```

```sas
data churn.telecom_outlier(drop= income retdays 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);
set churn.telecom_dummy_bckup;
run;

proc means n nmiss data= churn.telecom_outlier;
run;



/*-------------------------------DETECTING OUTLIERS FOR VARIABLES---------
-------------------------------------------

actvsubs
adjqty
adjrev
age1
age2
avg3mou
avgmou
avg3qty
avg6mou
avg6qty
avgrev
callwait_Mean
callwait_Range
ccrndmou_Range
change_mou
comp_vce_Mean
custcare_Mean
da_Mean
da_Range
datovr_Mean
datovr_Range
drop_blk_Mean
drop_dat_Mean
drop_vce_Mean
drop_vce_Range
eqpdays
iwylis_vce_Mean
months
mou_mean
mou_opkv_Range
mou_pead_Mean
mou_Range
opk_dat_Mean
ovrmou_Mean
ovrrev_Mean
owylis_vce_Range
plcd_vce_Mean
recv_sms_Mean
rev_Mean
rev_Range
roam_Mean
```

```
    totcalls
    totmrc_mean
    totrev
    uniqsubs
    blck_dat_Mean
    */
proc means data=churn.telecom_outlier;
run;

proc univariate data=churn.telecom_outlier;
var actvsubs;
output out=percentile1 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if actvsubs>5 then delete;
run;

 proc univariate data=churn.telecom_outlier;
 var adjqty;
 output out=percentile2 pctlpts=99.9 pctlpre=percentile;
 run;


data churn.telecom_outlier;
set churn.telecom_outlier;
if adjqty>44228 then delete;
run;

proc univariate data=churn.telecom_outlier;
var adjrev;
output out=percentile3 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if adjrev>7982.28 then delete;
run;




proc univariate data=churn.telecom_outlier;
var age1;
output out=percentile4 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if age1>86 then delete;
run;


proc univariate data=churn.telecom_outlier;
var age2;
output out=percentile5 pctlpts=99.9 pctlpre=percentile;
```

```
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if age2>90 then delete;
run;


proc univariate data=churn.telecom_outlier;
var avg3mou;
output out=percentile6 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if avg3mou>3811 then delete;
run;


proc univariate data=churn.telecom_outlier;
var avg6qty;
output out=percentile7 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if avg6qty>1616 then delete;
run;


proc univariate data=churn.telecom_outlier;
var avgrev;
output out=percentile8 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if avgrev>326.02 then delete;
run;



proc univariate data=churn.telecom_outlier;
var callwait_Mean;
output out=percentile9 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if callwait_Mean>62 then delete;
run;



proc univariate data=churn.telecom_outlier;
var callwait_Range;
output out=percentile10 pctlpts=99.9 pctlpre=percentile;
```

```sas
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if callwait_Range>52 then delete;
run;



proc univariate data=churn.telecom_outlier;
var ccrndmou_Range;
output out=percentile12 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if ccrndmou>197 then delete;
run;

proc univariate data=churn.telecom_outlier;
var change_mou;
output out=percentile14 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if change_mou>1510.5 then delete;
run;

proc univariate data=churn.telecom_outlier;
var custcare_Mean;
output out=percentile15 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if custcare_Mean>49.33 then delete;
run;



proc univariate data=churn.telecom_outlier;
var datovr_Mean;
output out=percentile16 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if datovr_Mean>31.2 then delete;
run;

proc univariate data=churn.telecom_outlier;
var datovr_Range;
output out=percentile17 pctlpts=99.9 pctlpre=percentile;
run;
```

```
data churn.telecom_outlier;
set churn.telecom_outlier;
if datovr_Range>86.97 then delete;
run;


proc univariate data=churn.telecom_outlier;
var drop_blk_Mean;
output out=percentile18 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if drop_blk_mean>157 then delete;
run;


proc univariate data=churn.telecom_outlier;
var drop_dat_Mean;
output out=percentile19 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if drop_dat_Mean>6.33 then delete;
run;

proc univariate data=churn.telecom_outlier;
var drop_vce_Mean;
output out=percentile20 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if drop_vce_Mean>63 then delete;
run;


proc univariate data=churn.telecom_outlier;
var drop_vce_Range;
output out=percentile21 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if drop_vce_Range>69 then delete;
run;


proc univariate data=churn.telecom_outlier;
var eqpdays;
output out=percentile22 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if eqpdays>1520 then delete;
```

```
run;



proc univariate data=churn.telecom_outlier;
var iwylis_vce_Mean;
output out=percentile23 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if iwylis_vce_Mean>154 then delete;
run;



proc univariate data=churn.telecom_outlier;
var months;
output out=percentile24 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if months>56 then delete;
run;

proc univariate data=churn.telecom_outlier;
var mou_mean;
output out=percentile25 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if mou_mean>3593.5 then delete;
run;



proc univariate data=churn.telecom_outlier;
var mou_opkv_Range;
output out=percentile26 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if mou_opkv_range>1638.39 then delete;
run;



proc univariate data=churn.telecom_outlier;
var mou_pead_mean;
output out=percentile27 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if mou_pead_mean>104.16 then delete;
```

```sas
run;



proc univariate data=churn.telecom_outlier;
var mou_Range;
output out=percentile28 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if mou_range>3648 then delete;
run;




proc univariate data=churn.telecom_outlier;
var opk_dat_Mean;
output out=percentile29 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if opk_dat_mean>58.66 then delete;
run;




proc univariate data=churn.telecom_outlier;
var ovrmou_mean;
output out=percentile29 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if ovrmou_mean>1022 then delete;
run;


proc univariate data=churn.telecom_outlier;
var ovrrev_mean;
output out=percentile30 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if ovrrev_mean>306.05then delete;
run;


proc univariate data=churn.telecom_outlier;
var owylis_vce_range;
output out=percentile30 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
```

```sas
set churn.telecom_outlier;
if owylis_vce_range>219 then delete;
run;


proc univariate data=churn.telecom_outlier;
var plcd_vce_mean;
output out=percentile31 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if plcd_vce_mean>1198.33 then delete;
run;


proc univariate data=churn.telecom_outlier;
var recv_sms_mean;
output out=percentile32 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if recv_sms_mean>10.66then delete;
run;


proc univariate data=churn.telecom_outlier;
var rev_mean;
output out=percentile33 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if rev_mean>447.78then delete;
run;


proc univariate data=churn.telecom_outlier;
var rev_Range;
output out=percentile34 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if rev_Range>686.37 then delete;
run;



proc univariate data=churn.telecom_outlier;
var roam_mean;
output out=percentile35 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
```

```sas
if roam_mean>81.60 then delete;
run;




proc univariate data=churn.telecom_outlier;
var totcalls;
output out=percentile36 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if totcalls>41466 then delete;
run;




proc univariate data=churn.telecom_outlier;
var totmrc_mean;
output out=percentile37 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if totmrc_mean>199.99 then delete;
run;




proc univariate data=churn.telecom_outlier;
var totrev;
output out=percentile38 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if totrev>8025.11 then delete;
run;




proc univariate data=churn.telecom_outlier;
var uniqsubs;
output out=percentile39 pctlpts=99.9 pctlpre=percentile;
run;

data churn.telecom_outlier;
set churn.telecom_outlier;
if uniqsubs>7 then delete;
run;




proc univariate data=churn.telecom_outlier;
var blck_dat_mean;
output out=percentile40 pctlpts=99.9 pctlpre=percentile;
run;
```

```
data churn.telecom_outlier;
set churn.telecom_outlier;
if blck_dat_mean>4then delete;
run;


data churn.telecom_outlier_bckup;
set churn.telecom_outlier;
run;




/*------------------------------------------------------------END OF DATA
PREPRATION STEP-------------------------------------------------------
-----------------------------------------------------------------------
----------------------------------------------------------------------
-----------------------------------------------------------------------
----------------------------------------------------------------/*




/*---------------------------------------------------------- Modelling
starts-----------------------------------------------------------
-----------------------------------------------------------------------
--------------------------------------------------------------------
-----------------------------------------------------------------------
------------------------------------------------------------/*


/* dividing data into training and validation data set using proc
surveyselect*/

proc surveyselect data=churn.telecom_outlier
method=SRS out=SAMP1 samprate=0.5 outall;
run;

data churn.train churn.validate;
set SAMP1;
if selected=0 then output churn.train;
else if selected=1 then output churn.validate;
run;



proc contents data=churn.train;
run;
proc freq data=train;
tables churn;
run;
```

```sas
proc freq data=validate;
tables churn;
run;



/* 1st iteration including all variables*/

proc logistic data=churn.train descending;
model churn =actvsubs
adjmou
adjqty
adjrev
 age1
 age2
 area_CAL
 area_CHICAGO
 area_DALLAS
 area_DC
 area_ENGLAND
 area_FLORIDA
 area_GRTLAKES
 /*area_HOUSTAN*/
 area_LOS
 area_MIDWEST
 area_NEWYORK
 area_NORTHWEST
 area_OHIO
 area_SOUTHFLORIDA
 area_atlantic
 area_phily
 area_swest
 area_tenesse
 area_texas
 asl_flag_y
 avg3mou
 avg3qty
 /*avg6mou*/
 avg6qty
 avgmou
 avgqty
 avgrev
 blck_dat_Mean
 callwait_Mean
 callwait_Range
 car_new
 /*cartype_NA*/
 /*cartype_basic*/
 cartype_luxury
 cartype_mini
 cartype_regular
 cartype_suv
 cartype_truck
 cartype_upper
 change_mou
```

```
         children_na
         children_yes
         city_area
         comp_vce_Mean
         crclscod_avg
         crclscod_bad
         crclscod_good
         crclscod_satisfactory
         csa_1
         csa_2
         csa_3
         csa_4
         custcare_Mean
         da_Mean
         da_Range
         datovr_Mean
         datovr_Range
         div_local
         div_long
         div_long_local
drop_blk_Mean
         drop_dat_Mean
         drop_vce_Mean
drop_vce_Range
         dwll_multiple
         dwll_single
         dwllsize_1
         dwllsize_2
         dwllsize_3
         dwllsize_4
         dwllsize_5
         dwllsize_6
         dwllsize_7
         dwllsize_8
         dwllsize_9
         dwllsize_J
         dwllsize_K
         dwllsize_L
         dwllsize_M
         dwllsize_N
         dwllsize_NA
         eqpdays
         ethnic_North_european
         ethnic_afro_american
         ethnic_arab
         ethnic_asian
         ethnic_asian_nonor
         ethnic_french
         ethnic_german
         ethnic_hispanic
         ethnic_italian
         ethnic_jewish
         ethnic_misc
         ethnic_polynesia
         ethnic_scot
         ethnic_south_european
         ethnic_unknown
```

forgntvl
handset_new
hnd_na
hnd_price
hnd_webcapable
iwylis_vce_Mean
mailordr_buy
mailresp_yes
marital_inferred
marital_inferred_single
marital_status
marital_unknown
models
months
mou_Mean
mou_Range
mou_opkv_Range
mou_pead_Mean
mtrcycle
numbcars_1
numbcars_2
occu_admin
occu_clercial
occu_craftsman
occu_farmer
occu_homemaker
occu_military
occu_religious
occu_retires
occu_sales
occu_self
occu_student
occu_tech
opk_dat_Mean
ovrmou_Mean
ovrrev_Mean
owylis_vce_Range
plcd_vce_Mean
proptype_apartment
proptype_condominium
proptype_duplex
proptype_misc
proptype_sing_family
recv_sms_Mean
rev_Mean
rev_Range
roam_Mean
rural_area
solflag_na
solflag_yes
suburban_area
totcalls
totmrc_Mean
totrev
town_area
truck
uniqsubs

```sas
 urban_area
 wrkwoman_yes;
output out=model_1 predicted=pred_prob;
 run;
/* 2ND Iteration using selection= backward option*/

proc logistic data=churn.train descending;
model churn =actvsubs
adjmou
adjqty
adjrev
 age1
 age2
 area_CAL
 area_CHICAGO
 area_DALLAS
 area_DC
 area_ENGLAND
 area_FLORIDA
 area_GRTLAKES
 /*area_HOUSTAN*/
 area_LOS
 area_MIDWEST
 area_NEWYORK
 area_NORTHWEST
 area_OHIO
 area_SOUTHFLORIDA
 area_atlantic
 area_phily
 area_swest
 area_tenesse
 area_texas
 asl_flag_y
 avg3mou
 avg3qty
 avg6mou
 avg6qty
 /*avgmou*/
 avgqty
 avgrev
 blck_dat_Mean
 callwait_Mean
 callwait_Range
 car_new
 /*cartype_NA*/
 /*cartype_basic*/
 cartype_luxury
 cartype_mini
 cartype_regular
 cartype_suv
 cartype_truck
 cartype_upper
 change_mou
 children_na
 children_yes
 city_area
 comp_vce_Mean
```

```
crclscod_avg
crclscod_bad
crclscod_good
crclscod_satisfactory
csa_1
csa_2
csa_3
csa_4
custcare_Mean
da_Mean
da_Range
datovr_Mean
datovr_Range
div_local
div_long
div_long_local
drop_blk_Mean
drop_dat_Mean
drop_vce_Mean
drop_vce_Range
dwll_multiple
dwll_single
dwllsize_1
dwllsize_2
dwllsize_3
dwllsize_4
dwllsize_5
dwllsize_6
dwllsize_7
dwllsize_8
dwllsize_9
dwllsize_J
dwllsize_K
dwllsize_L
dwllsize_M
dwllsize_N
dwllsize_NA
eqpdays
ethnic_North_european
ethnic_afro_american
ethnic_arab
ethnic_asian
ethnic_asian_nonor
ethnic_french
ethnic_german
ethnic_hispanic
ethnic_italian
ethnic_jewish
ethnic_misc
ethnic_polynesia
ethnic_scot
ethnic_south_european
ethnic_unknown
forgntvl
handset_new
hnd_na
hnd_price
```

```
  hnd_webcapable
  iwylis_vce_Mean
  mailordr_buy
  mailresp_yes
  marital_inferred
  marital_inferred_single
  marital_status
  marital_unknown
  models
  months
mou_Mean
mou_Range
  mou_opkv_Range
  mou_pead_Mean
  mtrcycle
  numbcars_1
  numbcars_2
  occu_admin
  occu_clercial
occu_craftsman
  occu_farmer
  occu_homemaker
  occu_military
  occu_religious
  occu_retires
  occu_sales
  occu_self
  occu_student
  occu_tech
  opk_dat_Mean
ovrmou_Mean
  ovrrev_Mean
  owylis_vce_Range
  plcd_vce_Mean
  proptype_apartment
  proptype_condominium
  proptype_duplex
  proptype_misc
  proptype_sing_family
  recv_sms_Mean
  rev_Mean
  rev_Range
  roam_Mean
  rural_area
  solflag_na
  solflag_yes
  suburban_area
  totcalls
  totmrc_Mean
  totrev
  town_area
  truck
  uniqsubs
  urban_area
  wrkwoman_yes /selection=backward;
output out=model_2 predicted=prediction_model;
  run;
```

```sas
/* 3rd ITERATION using selection=forward; option*/

proc logistic data=churn.train descending;
model churn =actvsubs
adjmou
adjqty
adjrev
 age1
 age2
 area_CAL
 area_CHICAGO
 area_DALLAS
 area_DC
 area_ENGLAND
 area_FLORIDA
 area_GRTLAKES
 /*area_HOUSTAN*/
 area_LOS
 area_MIDWEST
 area_NEWYORK
 area_NORTHWEST
 area_OHIO
 area_SOUTHFLORIDA
 area_atlantic
 area_phily
 area_swest
 area_tenesse
 area_texas
 asl_flag_y
 avg3mou
 avg3qty
 avg6mou
 avg6qty
 /*avgmou*/
 avgqty
 avgrev
 blck_dat_Mean
 callwait_Mean
 callwait_Range
 car_new
 /*cartype_NA*/
 /*cartype_basic*/
 cartype_luxury
 cartype_mini
 cartype_regular
 cartype_suv
 cartype_truck
 cartype_upper
 change_mou
 children_na
```

children_yes
city_area
comp_vce_Mean
crclscod_avg
crclscod_bad
crclscod_good
crclscod_satisfactory
csa_1
csa_2
csa_3
csa_4
custcare_Mean
da_Mean
da_Range
datovr_Mean
datovr_Range
div_local
div_long
div_long_local
drop_blk_Mean
drop_dat_Mean
drop_vce_Mean
drop_vce_Range
dwll_multiple
dwll_single
dwllsize_1
dwllsize_2
dwllsize_3
dwllsize_4
dwllsize_5
dwllsize_6
dwllsize_7
dwllsize_8
dwllsize_9
dwllsize_J
dwllsize_K
dwllsize_L
dwllsize_M
dwllsize_N
dwllsize_NA
eqpdays
ethnic_North_european
ethnic_afro_american
ethnic_arab
ethnic_asian
ethnic_asian_nonor
ethnic_french
ethnic_german
ethnic_hispanic
ethnic_italian
ethnic_jewish
ethnic_misc
ethnic_polynesia
ethnic_scot
ethnic_south_european
ethnic_unknown
forgntvl

```
  handset_new
  hnd_na
  hnd_price
  hnd_webcapable
  iwylis_vce_Mean
  mailordr_buy
  mailresp_yes
  marital_inferred
  marital_inferred_single
  marital_status
  marital_unknown
  models
  months
mou_Mean
mou_Range
  mou_opkv_Range
  mou_pead_Mean
  mtrcycle
  numbcars_1
  numbcars_2
  occu_admin
  occu_clercial
occu_craftsman
  occu_farmer
  occu_homemaker
  occu_military
  occu_religious
  occu_retires
  occu_sales
  occu_self
  occu_student
  occu_tech
  opk_dat_Mean
ovrmou_Mean
  ovrrev_Mean
  owylis_vce_Range
  plcd_vce_Mean
  proptype_apartment
  proptype_condominium
  proptype_duplex
  proptype_misc
  proptype_sing_family
  recv_sms_Mean
  rev_Mean
  rev_Range
  roam_Mean
  rural_area
  solflag_na
  solflag_yes
  suburban_area
  totcalls
  totmrc_Mean
  totrev
  town_area
  truck
  uniqsubs
  urban_area
```

```sas
 wrkwoman_yes/ selection=forward;
output out=model_3 predicted=prediction_model;
 run;

 /*final iteration*/

 proc logistic data=churn.train descending;
model churn =
actvsubs
 adjrev
 age1
 area_NORTHWEST
 area_SOUTHFLORIDA
 asl_flag_y
 /*area_HOUSTAN*/
 avgqty
 avg3mou
 avg6mou
 avgrev
 cartype_luxury
 change_mou
 children_yes
 comp_vce_mean
 blck_dat_Mean
  city_area
  crclscod_good
 crclscod_satisfactory
 datovr_Mean
 drop_dat_Mean
 div_local
 div_long
 drop_blk_mean
 drop_vce_Range
 dwll_multiple
 dwllsize_8
 eqpdays
 ethnic_North_european
 ethnic_afro_american
 ethnic_arab
 ethnic_asian
 ethnic_asian_nonor
 ethnic_french
 ethnic_german
 ethnic_hispanic
 ethnic_italian
 ethnic_jewish
 ethnic_misc
 ethnic_polynesia
 ethnic_scot
 ethnic_south_european
 ethnic_unknown
 handset_new
 hnd_price


 mailordr_buy
 marital_inferred
```

```
   marital_inferred_single
   marital_status
   marital_unknown
   models
   months
   mou_range
   occu_self
ovrmou_Mean
   occu_sales
   owylis_vce_Range

   proptype_sing_family
   rev_Mean
   rev_range
   roam_mean
   rural_area
   suburban_area
   totcalls
   totmrc_Mean
   uniqsubs;
output out=model_4 predicted=prediction_model;
   run;




/* performing logistic regression on validation data set by further removing
all nonessential variables*/
proc logistic data=churn.validate descending;
model churn =
actvsubs
   adjrev
   age1
   area_NORTHWEST
   area_CHICAGO
   area_SOUTHFLORIDA
   asl_flag_y
   /*area_HOUSTAN*/
   avg3mou
   avg6mou
   /*avgmou*/
   change_mou
   children_yes
   children_na
   csa_1
   csa_2
   csa_3
   csa_4
   comp_vce_mean
   div_long
   drop_blk_mean
   drop_vce_mean
   dwll_multiple
   eqpdays
   ethnic_North_european
   ethnic_arab
```

```
 ethnic_asian
 ethnic_asian_nonor
 ethnic_french
 ethnic_german
 ethnic_hispanic
 ethnic_italian
 ethnic_jewish
 ethnic_scot
 ethnic_south_european
 ethnic_unknown
 handset_new
 hnd_price
 hnd_na
 hnd_webcapable
 mailresp_yes
 marital_unknown

 mou_opkv_range
 mailordr_buy
 models
 months
 mou_mean
 mou_range
 numbcars_2
 ovrrev_mean
ovrmou_Mean
 roam_mean
 rural_area
 town_area
 suburban_area
 totcalls
 totmrc_Mean
 uniqsubs/selection=backward;
output out=model_5 predicted=prediction_model;
 run;



proc logistic data=churn.validate descending;
model churn =
adjmou
adjrev
actvsubs

 age1
 area_NORTHWEST
 area_SOUTHFLORIDA
 asl_flag_y
 /*area_HOUSTAN*/
 avg3mou
 avg6mou
 change_mou
 children_yes
 children_na

 comp_vce_mean
 div_long
```

```sas
    drop_vce_mean

 eqpdays
 ethnic_North_european
 ethnic_arab
 ethnic_asian
 ethnic_asian_nonor
 ethnic_french
 ethnic_german
 ethnic_hispanic
 ethnic_italian
 ethnic_jewish
 ethnic_scot
 ethnic_south_european
 ethnic_unknown
 handset_new
 hnd_price

 mailresp_yes



 models
 months
 mou_range


ovrmou_Mean

 rural_area


 totmrc_Mean
 uniqsubs/selection=backward;
output out=model_5 predicted=prediction_model;
 run;



 /* final iteration on validation data set after removing non essential
varaibles found through validation dataset */
 proc logistic data=churn.train descending;
model churn =
 actvsubs
 age1
 area_NORTHWEST
 area_SOUTHFLORIDA
 asl_flag_y
 /*area_HOUSTAN*/
 avg3mou
 avg6mou
 change_mou
 children_yes

 comp_vce_mean
```

```sas
  div_long
  drop_vce_mean
  eqpdays
  ethnic_North_european
  ethnic_arab
  ethnic_asian
  ethnic_asian_nonor
  ethnic_french
  ethnic_german
  ethnic_hispanic
  ethnic_italian
  ethnic_jewish
  ethnic_scot
  ethnic_south_european
  ethnic_unknown
  handset_new
  hnd_price
  mailresp_yes
  models
  months
  mou_range
  ovrmou_Mean
  totmrc_Mean
uniqsubs/selection=backward;
output out=churn.final_model predicted=prediction_model;
 run;



 /*sorting final dataset*/
 proc sort data= churn.final_model out=churn.sorted_model;
 by descending prediction_model;
 run;

proc rank data=churn.sorted_model out=churn.model_decile
groups=10
ties=mean;
var prediction_model;
ranks decile;
run;


/*
 proc export data=churn.model_decile
 outfile="Y:\FINAL_CASE _STUDY_CHURN\liftchart.csv"
 dbms=csv replace;
 run;*/




/* scoring dataset creating confusion matrix*/

proc logistic data=churn.train descending;
model churn =
 actvsubs
 age1
```

```
 area_NORTHWEST
 area_SOUTHFLORIDA
 asl_flag_y
 /*area_HOUSTAN*/
 avg3mou
 avg6mou
 change_mou
 children_yes

 comp_vce_mean
 div_long
 drop_vce_mean
 eqpdays
 ethnic_North_european
 ethnic_arab
 ethnic_asian
 ethnic_asian_nonor
 ethnic_french
 ethnic_german
 ethnic_hispanic
 ethnic_italian
 ethnic_jewish
 ethnic_scot
 ethnic_south_european
 ethnic_unknown
 handset_new
 hnd_price
 mailresp_yes
 models
 months
 mou_range
 ovrmou_Mean
 totmrc_Mean
uniqsubs;
output out=churn.confusion_matrix predicted=prediction_model;
score data=churn.validate out=churn.score;
 run;



proc freq data=churn.score;
tables F_churn*I_churn/norow nocum nopercent nocol;
run;



/*-------------------------------------------------------------------------
---------------------------------------------------------------------
                                           VALIDATION OF MODEL
*/



proc logistic data=churn.validate descending;
model churn =
 actvsubs
 age1
 area_NORTHWEST
```

```sas
     area_SOUTHFLORIDA
     asl_flag_y
     /*area_HOUSTAN*/
     avg3mou
     avg6mou
     change_mou
     children_yes

     comp_vce_mean
     div_long
     drop_vce_mean
     eqpdays
     ethnic_North_european
     ethnic_arab
     ethnic_asian
     ethnic_asian_nonor
     ethnic_french
     ethnic_german
     ethnic_hispanic
     ethnic_italian
     ethnic_jewish
     ethnic_scot
     ethnic_south_european
     ethnic_unknown
     handset_new
     hnd_price
     mailresp_yes
     models
     months
     mou_range
     ovrmou_Mean
     totmrc_Mean
uniqsubs;
output out=churn.validate_model predicted=prediction_model;
run;

proc sort data=churn.validate_model out=churn.validate_sorted_model;
by descending prediction_model;
run;


proc rank data=churn.validate_sorted_model out=churn.validate_decile
groups=10
ties=mean;
var prediction_model;
ranks decile;
run;

/*proc export data=churn.validate_decile
outfile="Y:\FINAL_CASE _STUDY_CHURN\validation_liftchart.csv"
 dbms=csv replace;
 run;*/



proc logistic data=churn.validate descending;
model churn =
```

```
   actvsubs
   age1
   area_NORTHWEST
   area_SOUTHFLORIDA
   asl_flag_y
   /*area_HOUSTAN*/
   avg3mou
   avg6mou
   change_mou
   children_yes

   comp_vce_mean
   div_long
   drop_vce_mean
   eqpdays
   ethnic_North_european
   ethnic_arab
   ethnic_asian
   ethnic_asian_nonor
   ethnic_french
   ethnic_german
   ethnic_hispanic
   ethnic_italian
   ethnic_jewish
   ethnic_scot
   ethnic_south_european
   ethnic_unknown
   handset_new
   hnd_price
   mailresp_yes
   models
   months
   mou_range
   ovrmou_Mean
   totmrc_Mean
uniqsubs;
output out=churn.validate_confusion predicted=prediction_model;
score data=churn.train out= churn.score_valid;
run;




   proc freq data=churn.score_valid;
   tables F_churn*I_churn/nocum norow nopercent nocol;
   run;



   /*solution to qus no 5 AND FOR ALL STRATEGIES*/
   proc univariate data=churn.train;
   var avgrev;
   run;


   data churn.top_revenue_customers;
```

```sas
set churn.telecom_outlier;
if avgrev>67.86;
run;
proc print data= churn.top_revenue_customers;
run;

PROC FREQ DATA= churn.top_revenue_customers;
tables churn;
run;


data churn.low_revenue_customers;
set churn.telecom_outlier;
if avgrev<67.86;
run;

proc freq data=churn.low_revenue_customers;
tables churn;
run;

proc means  data=churn.top_revenue_customers;
var avgmou;
run;

proc means data= churn.low_revenue_customers;
var avgmou;
run;

proc means data=churn.top_revenue_customers;
var ovrmou_mean;
run;

proc means data=churn.top_revenue_customers;
var drop_vce_mean;
run;
proc means data=churn.low_revenue_customers;
var drop_vce_mean;
run;

proc freq data=churn.top_revenue_customers;
tables drop_vce_mean*area_texas/norow nocol nopercent;
run;

proc means data=churn.top_revenue_customers;
var age1 age2;
run;
proc means data=churn.low_revenue_customers;
var age1 age2;
run;

proc means data=churn.top_revenue_customers;
var eqpdays handset_new;
run;
```