

Data visualization

```
Proc contents data= xyz.train;
```

```
Run;
```

Sorting the data based on Cust_id and also removing duplicates

```
Proc sort data=xyz.train nodupkey;  
by cust_id;  
run;
```

Removing observations with missing Customer ID

```
data xyz.train;  
set xyz.train;  
if cust_id=. then delete;  
run;
```

Univariate analysis

```
proc means data=xyz.train1 n nmiss min max median mean ;  
run;
```

Outlier estimation (Percentile distribution and Extreme values list) for Numeric Variables

```
proc univariate data=xyz.train1 nextrobs=100;  
run;
```

(For Character variable)

```
proc freq data=xyz.train1;  
table occupation_partner occupation_type debt_burden_ratio__monthly_emi_  
gender income_group  
location marital_status months_on_book no__of_times_30_dpd_in_last_6_m  
no__of_times_60_dpd_in_last_6_m ;  
run;
```

Removing Missing observations

```
proc sql;  
create table xyz.train2 as select * from xyz.train1 where  
no__of_times_30_dpd_in_last_6_m  
is not null and no__of_times_60_dpd_in_last_6_m is not null;  
quit;
```

Bivariate Analysis (Fine Classing)

```
data xyz.train2;  
rename Debt_Burden_Ratio__Monthly_EMI__=DBR  
Education_Qualification__1__no=Educational_qualification  
No__of_times_30_dpd_in_last_6_m= DPD_30_6M  
No__of_times_60_dpd_in_last_6_m= DPD_60_6M;  
set xyz.train2;  
run;
```

```

%macro bivariate(var);
proc sql;
create table bi_&var as
select &var , sum(Default_Flag) as default,
count (*)as total, sum(Default_Flag)/count (*) as default_rate
from xyz.train2
group by &var;
quit;

proc print data = bi_&var;
run;

%mend;

ods html file='C:\Documents and Settings\Administrator\Desktop\ODS1.xls' ;

%bivariate (Age);
%bivariate (Assets);
%bivariate (DBR);
%bivariate (Educational_qualification);
%bivariate (Employed since yrs);
%bivariate (Existing Customer);
%bivariate (Dependents);
%bivariate (Gender);
%bivariate (Housing__1__Own__2__Rent__);
%bivariate (Income group);
%bivariate (Job Seniority);
%bivariate (Marital Status);
%bivariate (Months on book);
%bivariate (DPD_30_6M);
%bivariate (DPD_60_6M);
%bivariate (Occupation Partner);
%bivariate (Occupation_type);
%bivariate (Savings);
%bivariate (Years_at__current_address);
ods html close;

```

Coarse classing

```

data xyz.train3;
set xyz.train2;

if Employed since yrs <=7 then employed_band=1;
else if 7 < Employed_since_yrs < 12 then employed_band=2;
else employed_band=3;

if age <=27 then age_band=1;
else if 27 < age < 43 then age_band=2;
else age_band=3;

if Marital Status= "Married" then Marital_band=1;
else Marital_band=2;

if Occupation_Partner="fixed" then Occupartner_band=1;

else if Occupation_Partner="freelance" then Occupartner_band=2;
else Occupartner_band=3;

```

```

if Occupation_type="Govt" then Occutype_band=1;
else if Occupation_type="Private" then Occutype_band=2;
else Occutype_band=3;

if DPD_30_6M=0 then DPD306M=1;
else DPD306M=2;

if DPD_60_6M=0 then DPD606M=1;
else DPD606M=2;

if Income_group "< = 15000" then Income_band=1;
else if Income_group= "< = 25000" then Income_band=1;
else if Income_group="< = 35000" then Income_band=2;
else if Income_group="< = 50000" then Income_band=2;
else if Income_group="<=100000" then Income_band=2;
else Income_band =3;
run;

%macro bivariate(var);
proc sql;
create table bi_&var as
select &var , sum(Default_Flag) as default,
count (*)as total, sum(Default_Flag)/count (*) as default_rate
from xyz.train3
group by &var;
quit;

proc print data = bi_&var;
run;

%mend;

ods html file='C:\Documents and Settings\Administrator\Desktop\ODS2.xls' ;

%bivariate (age_band);
%bivariate (Marital band);
%bivariate (Occupartner band);
%bivariate (Occutype band);
%bivariate (DPD306M);
%bivariate (DPD606M);
%bivariate (Income band);
%bivariate (employed_band);
ods html close;

```

Creating WOE variable

```

data xyz.train4;
set xyz.train3;

if Income_band =1 then income_woe=0.244591;
else if Income_band=2 then income_woe=0.088829;
else income_woe= -0.39331;

if Occutype_band= 1 then Occutype_woe= 0.170621;
else if Occutype_band=2 then Occutype_woe= -0.2689;
else Occutype_woe= 0.621877;

if Occupartner_band=1 then Occupartner_woe = 0.181674;
else if Occupartner_band=2 then Occupartner_woe= -0.10572;
else Occupartner_woe= -0.4476;

```

```

if Marital_band=1 then Marital_woe= -0.10049;
else Marital_woe=0.092989;

if age_band=1 then age_woe= 0.345728;
else if age_band=2 then age_woe=0.091898;
else age_woe=-0.21301;

if Housing__1__Own__2__Rent__ =1 then housing_woe= -0.20757;
else housing_woe= 0.137949;

if Employed_since_yrs =1 then employed_woe=0.15991;
else if Employed_since_yrs =2 then employed_woe= 0.085352112;
else employed_woe= -0.334822545;
run;

```

Multicollinearity check

```

proc reg data= xyz.train4;
model Default_Flag= income_woe housing_woe age_woe Marital_woe
Occupartner_woe Occutype_woe employed_woe /vif;
run;

```

Logistic Regression Analysis

```

ods rtf;
proc logistic data = xyz.train4 descending ;
model
Default_Flag = income_woe housing_woe age_woe Marital_woe
Occupartner_woe Occutype_woe /*employed_woe*/ ;
output out = xyz.train5 p = phat;
run;
ods rtf close;

```

Creating Data for KS and Gini Coefficient (discriminatory test)

```

proc sort data = xyz.train5 ; by descending phat;
run;

/* divide data into 10 equal observation bins */
ods html file='C:\Documents and Settings\Administrator\Desktop\ODS3.xls';
%let Noofrecs = 18248;
%let Noofbins = 10;

data xyz.pred_default;
set xyz.train5;
retain cumm_count;
count = 1;
cumm count = sum(cumm count,count);
bin = round(cumm_count/(&Noofrecs/&Noofbins)-0.5)+1;
if bin GT &Noofbins then Bin = &Noofbins;
run;
proc sql;
create table xyz.gains_dev as
select bin, count(*) as freq, sum(Default_Flag) as default,
mean(phat) as exp_bad
from xyz.pred_default
group by bin;

```

```
run;
proc print data = xyz.gains_dev;
run;
ods html close;
```

Validation

```
proc means data=xyz.test n nmiss max min median mean p1 p5 p95 p99;
run;
```

Creating woe variables for validation sample

```
data xyz.test1;
set xyz.test;

if age <=27 then age_band=1;
else if 27 < age < 43 then age_band=2;
else age_band=3;

if Income_group = "< = 15000" then Income_band=1;
else if Income_group= "< = 25000" then Income_band=1;
else if Income_group="< = 35000" then Income_band=2;
else if Income_group="< = 50000" then Income_band=2;
else if Income_group="<=100000" then Income_band=2;
else Income_band =3;

if Occupation_Partner="fixed" then Occupartner_band=1;
else if Occupation_Partner="freelance" then Occupartner_band=2;
else Occupartner_band=3;

if Occupation_type="Govt" then Occutype_band=1;
else if Occupation_type="Private" then Occutype_band=2;
else Occutype_band=3;

run;

data xyz.test2;
set xyz.test1;

if Income_band =1 then income_woe=0.244591;
else if Income_band=2 then income_woe=0.088829;
else income_woe= -0.39331;

if Occutype_band= 1 then Occutype_woe= 0.170621;
else if Occutype_band=2 then Occutype_woe= -0.2689;
else Occutype_woe= 0.621877;

if Occupartner_band=1 then Occupartner_woe = 0.181674;
else if Occupartner_band=2 then Occupartner_woe= -0.10572;
else Occupartner_woe= -0.4476;

/*if Marital_band=1 then Marital_woe= -0.10049;
else Marital_woe=0.092989; */

if age_band=1 then age_woe= 0.345728;
else if age_band=2 then age_woe=0.091898;
else age_woe=-0.21301;
```

```

if Housing__1__Own__2__Rent_ =1 then housing_woe= -0.20757;
else housing_woe= 0.137949;

logit=
- 1.6879+income_woe*1.6038+housing_woe*2.2502+age_woe*1.0382+Occupartner_
r_woe*2.3120+Occutype_woe*2.2050;
phat = 1 / (1 + exp(-logit));

run;

```

- The final logistic equation is $\log(P/1-P) = -$
 $1.6879 + \text{income_woe} * 1.6038 + \text{housing_woe} * 2.2502 + \text{age_woe} * 1.0382 + \text{Occupartner_w}$
 $\text{oe} * 2.3120 + \text{Occutype_woe} * 2.2050;$

Creating data for KS and Gini

```

proc sort data = xyz.test2; by descending phat;
run;
ods html file= 'C:\Documents and Settings\Administrator\Desktop\ODS4.xls';
%let Noofrecs = 4567;
%let Noofbins = 10;
data xyz.test3;
set xyz.test2;
retain cumm_count;
count = 1;
cumm_count = sum(cumm_count,count);
bin = round(cumm_count/(&Noofrecs/&Noofbins)-0.5)+1;
if bin GT &Noofbins then Bin = &Noofbins;
run;
proc sql;
create table xyz.gains_val as
select bin, count(*) as freq, sum(Default_Flag) as default
from xyz.test3
group by bin;
run;
proc print data = xyz.gains_val;
run;
ods html close;

```