/\*Setting data set to a new data\*/

**data** LCloanstatus2011\_1;

set LCloanstatus2011;

**run**;

**proc** **contents** data = LCloanstatus2011\_1 order=varnum; **run**;

/\*exploratory analysis of numeric variables\*/

**proc** **means** data= LCloanstatus2011\_1;

var \_numeric\_;

**run**;

/\* To find out duplicate values \*/

**proc** **sql**;

title 'Lenging club data analysis';

select \*, count(\*) as Count

from WORK.LCLOANSTATUS2011\_1

group by id

having count(\*) > **1**;

**run**;

/\* Covering seven loan categories to 2 categories \*/

**PROC** **SQL**;

CREATE TABLE WORK.QUERY\_FOR\_LCLOANSTATUS2011\_1(label="QUERY\_FOR\_LCLOANSTATUS\_1") AS

SELECT t1.\*,

/\* Loan staus (N) \*/

(CASE

WHEN 'Charged Off' = t1.loan\_status THEN 'Default'

WHEN 'Current' = t1.loan\_status THEN 'No Default'

WHEN 'Default' = t1.loan\_status THEN 'Default'

WHEN 'Fully Paid' = t1.loan\_status THEN 'No Default'

WHEN 'In Grace Period' = t1.loan\_status THEN 'No Default'

WHEN 'Late (16-30 days)' = t1.loan\_status THEN 'No Default'

WHEN 'Late (31-120 days)' = t1.loan\_status THEN 'No Default'

END) FORMAT=$CHAR12. AS Loan\_status1

FROM WORK.LCLOANSTATUS2011\_1 t1;

**QUIT**;

/\* To get a sense of missing values in every row using CMISS\*/

**data** test1;

set WORK.QUERY\_FOR\_LCLOANSTATUS2011\_1;

miss\_n = cmiss(of id -- application\_type);

**run**;

**proc** **means** data= test1 n mean max min ;

var miss\_n;

**run**;

/\*exploratory analysis of categorical variables\*/

**proc** **freq** data=work.query\_for\_lcloanstatus2011\_1;

tables \_character\_;

**run**;

/\*what numeric varables impact loan status \*/

**proc** **means** data=work.query\_for\_lcloanstatus2011\_1 ;

class Loan\_status1;

var installment funded\_amnt funded\_amnt\_inv loan\_amnt annual\_inc dti revol\_util delinq\_2yrs inq\_last\_6mths pub\_rec int\_rate;

**run**;

ods graphics on;

/\* Frequency distribution by Loan Purpose, Term and Loan status \*/

**proc** **freq** data=WORK.QUERY\_FOR\_LCLOANSTATUS2011\_1 order=freq;

tables purpose term grade loan\_status1 ;

**run**;

/\* Frequency distribution Loan status by Purpose \*/

**proc** **freq** data=WORK.QUERY\_FOR\_LCLOANSTATUS2011\_1 order=freq;

tables purpose \*Loan\_status1 grade\*Loan\_status1 /nocol nofreq nopercent;

/\* suppresses row %, frequency values and cumulative % \*/

**run**;

/\* To find corelation between related numeric variable \*/

**proc** **corr** data= work.query\_for\_lcloanstatus2011\_1;

var funded\_amnt funded\_amnt\_inv loan\_amnt;

**run**;

ods graphics off;

ods graphics on;

/\* To find corelation between related numeric value \*/

**proc** **corr** data= work.query\_for\_lcloanstatus2011\_1;

var revol\_util delinq\_2yrs inq\_last\_6mths pub\_rec ;

**run**;

ods graphics off;

/\* Frequency distribution Loan status by grade \*/

**proc** **freq** data=WORK.QUERY\_FOR\_LCLOANSTATUS2011\_1 order=freq;

tables grade\* Loan\_status1 /norow nofreq nopercent;

/\* suppresses row %, frequency values and cumulative % \*/

**run**;

/\* does grade subgrade pupose term home\_ownership emp\_length affect status\*/

**proc** **freq** data= work.query\_for\_lcloanstatus2011\_1;

table grade\*Loan\_status1 / nocol nofreq nocum nopercent chisq;

table sub\_grade\*Loan\_status1 / nocol nofreq nocum nopercent chisq;

table purpose\*Loan\_status1 / nocol nofreq nocum nopercent chisq;

table term\*Loan\_status1 / nocol nofreq nocum nopercent chisq;

table home\_ownership\*Loan\_status1/nocol nofreq nocum nopercent chisq;

table emp\_length\* Loan\_status1/ nocol nofreq nocum nopercent chisq;

**run**;

**proc** **SGSCATTER** DAta=WORK.QUERY\_FOR\_LCLOANSTATUS2011\_1;

MATRIX annual\_inc dti int\_rate delinq\_2yrs

/

DIAGONAL=(HISTOGRAM )

START=TOPLEFT

ELLIPSE=(ALPHA=**0.05** TYPE=PREDICTED)

NOLEGEND;

**RUN**;

ods graphics off;

/\* Plot a bar chart of grades \*/

**proc** **gchart** data=WORK.QUERY\_FOR\_LCLOANSTATUS2011\_1;

vbar grade /discrete percent;

**run**;

/\* Plot a bar chart of loan status \*/

**proc** **gchart** data=WORK.QUERY\_FOR\_LCLOANSTATUS2011\_1;

vbar loan\_status1 /discrete percent;

**run**;

ods graphics on;

/\* Split the data set test and training\*/

**proc** **surveyselect** data=work.query\_for\_lcloanstatus2011\_1 method=srs seed=**1234** outall samprate=**0.75** out=LC2011\_subsets; **run**;

**data** LC2011\_train;

set LC2011\_subsets;

if selected =**1**;

**run**;

**data** LC2011\_test;

set LC2011\_subsets;

if selected =**0**;

**run**;

/\* Add key numeric Borrower variables dti \*/

ods graphics on;

**proc** **logistic** data=LC2011\_train;

class home\_ownership(ref=" OWN")/ param=ref;

model loan\_status1 (event =" Default")= dti annual\_inc home\_ownership/ outroc=troc;

score data=LC2011\_test outroc=vroc fitstat;

roccontrast ;

roc 'dti annual\_inc home\_ownership' dti annual\_inc home\_ownership;

**run**;

/\* Add key numeric credit variables dti \*/

ods graphics on;

**proc** **logistic** data=LC2011\_train;

class home\_ownership(ref=" OWN")/ param=ref;

model loan\_status1 (event ="Default")= dti annual\_inc home\_ownership revol\_util delinq\_2yrs inq\_last\_6mths/ outroc=troc;

score data=LC2011\_test outroc=vroc fitstat;

roccontrast ;

roc 'dti annual\_inc home\_ownership revol\_util delinq\_2yrs inq\_last\_6mths' dti annual\_inc home\_ownership revol\_util delinq\_2yrs inq\_last\_6mths;

**run**;

/\* Add key numeric variables loan and grade characteristics \*/

ods graphics on;

**proc** **logistic** data=LC2011\_train;

class 'purpose'n(ref="credit\_card")'grade'n(ref="A") / param=ref;

class home\_ownership(ref=" OWN")/ param=ref;

model loan\_status1 (event =" Default")= dti annual\_inc home\_ownership revol\_util delinq\_2yrs inq\_last\_6mths 'purpose'n'grade'n / outroc=troc;

score data= LC2011\_test outroc=vroc fitstat;

roccontrast ;

roc 'dti annual\_inc home\_ownership' dti annual\_inc home\_ownership;

roc 'dti annual\_inc home\_ownership revol\_util delinq\_2yrs inq\_last\_6mths' dti annual\_inc home\_ownership revol\_util delinq\_2yrs inq\_last\_6mths;

roc 'dti annual\_inc home\_ownership revol\_util delinq\_2yrs inq\_last\_6mths purpose grade ' dti annual\_inc home\_ownership revol\_util delinq\_2yrs inq\_last\_6mths purpose grade ;

**run**;

/\*

proc logistic data=LC2011\_train;

class'purpose'n(ref="credit\_card")'grade'n(ref="A")/ param=ref;

model loan\_status1 (event="Default")=dti funded\_amnt revol\_util delinq\_2yrs inq\_last\_6mths pub\_rec'purpose'n'grade'n/ outroc=troc;

score data=LC2011\_test outroc=vroc fitstat;

roccontrast;

roc 'dti' dti ;

roc 'dti, funded\_amnt' dti funded\_amnt;

roc 'dti, funded\_amnt term ' dti funded\_amnt term ;

roc 'dti, funded\_amnt term grade ' dti funded\_amnt term grade ;

run; \*/