**Bank Telemarketing Case Study**

**Background**: A European bank wishes to understand which of its customers are likely to respond to a phone marketing offer to set up a long term deposit. A success is considered when a customer agrees to open a term deposit with the bank.

**Problem Definition**: How can the bank identify which customers are likely to respond to a direct marketing offer.

**Solution Design**: After preliminary data exploration and data preparation (preparing data for modelling by removing outliers etc.), Binary logistic regression method is used to model the data to obtain the response score for each customer and then rank order customer in descending order of response score. A cut-off is used to identify the potential customers to be called for direct telemarketing offer. Business insights are generated from the finalized model i.e. which attributes are important in determining response. Recommendations are also generated based on the model outcome.

--------------------------------------------------------------SAS Code ---------------------------------------------------------------

\* Simple approach - start with proc import \*;

**proc** **import** datafile = "Z:\Class Codes and Data\Topic 5 - An Introduction to SAS Language\Bank\_data.csv"

out = bank dbms = csv replace;

**run**;

\* Check contents to see if any import code needs modification\*;

/\*proc contents data = bank;

run;\*/

\* The contact information is truncated \*\*;

\* Can re-run import code with a longer length specified for Contact \*;

**data** bank;

infile 'Z:\Class Codes and Data\Topic 5 - An Introduction to SAS Language\Bank\_data.csv' delimiter=',' MISSOVER DSD firstobs=2 LRECL=32760;

informat age BEST32.;

informat job $13.;

informat marital $9.;

informat education $10.;

informat default $4.;

informat balance BEST32.;

informat housing $3.;

informat loan $4.;

informat contact $15.;

informat day BEST32.;

informat month $3.;

informat duration BEST32.;

informat campaign BEST32.;

informat pdays BEST32.;

informat previous BEST32.;

informat poutcome $7.;

informat y $5.;

input age

job $

marital $

education $

default $

balance

housing $

loan $

contact $

day

month $

duration

campaign

pdays

previous

poutcome $

y $ ;

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*DATA EXPLORATION \*\*\*\*\*\*\*;

**PROC** **CONTENTS** DATA = BANK;

**RUN**;

**proc** **freq** data = bank;

tables contact

default

education

housing

job

loan

marital

month

poutcome

y

;**run**;

**proc** **means** min max mean nmiss median stddev data = bank;

**run**;

**proc** **univariate** data = bank;

var balance;

**run**;

**proc** **univariate** data = bank;

var duration;

**run**;

**proc** **univariate** data = bank;

var campaign;

**run**;

**proc** **univariate** data = bank;

var previous;

**run**;

**proc** **univariate** data = bank;

var pdays;

**run**;

\*\* Should remove -1 from pdays \*\*\*;

**data** bank\_p;

set bank;

if pdays ne -1;

**run**;

**proc** **univariate** data = bank\_p;

var pdays;

**run**;\*/

\* Data Preparation \*;

\*\* For possible outlier deletions, check # of obs impacted \*\*;

/\*

data bank\_1;

set bank;\*/

\*if balance > 20000;

\*if duration > 2000;

/\*if pdays ne -1;

if pdays > 550;

\*/

\*if campaign > 20;

/\*if previous > 20;

proc freq;

tables y /norow nocol;

run;\*/

\*\* Replacing missing values in education \*\*;

**data** check2;

set bank;

**proc** **freq**;

tables job\*education /nocol nopct;

**run**;

**data** prep;

set bank;

if y = "no" then ny = 0;

else ny =1;

\* dropping previous outcomes \*\*;

drop previous;

\* imputing missing educ value using job type \*\*;

if job = "admin." and education = "unknown" then education = "secondary";

if job = "blue-collar" and education = "unknown" then education = "secondary";

if job = "management" and education = "unknown" then education = "tertiary";

if job = "services" and education = "unknown" then education = "secondary";

if job = "student" and education = "unknown" then education = "secondary";

if job = "technician" and education = "unknown" then education = "secondary";

length btype $20.;

if balance < 0 then btype = "OD";

else if 0 le balance lt 500 then btype = "<500";

else if 500 le balance lt 3500 then btype = "500-3500";

else if 3500 le balance lt 13500 then btype = "3500-13500";

else btype = ">13500";

length dtype $20.;

if duration < 10 then dtype = "< 10s";

else if 10 le duration lt 60 then dtype = "10-60s";

else if 60 le duration le 300 then dtype = "60-300s";

else if 300 le duration le 1200 then dtype = "300-1200s";

else dtype = ">1200s";

if contact = " " then cell = .;

if contact = "cellular" then cell = 1;

else if contact = "telephone" then cell = 0;

if default = "no" then def = 1;

else def = 0;

if education = "primary" then pri\_ed = 1;

else pri\_ed = 0;

if education = "secondary" then sec\_ed = 1;

else sec\_ed = 0;

if education = "tertiary" then ter\_ed = 1;

else ter\_ed = 0;

if education = "unknown" then un\_ed = 1;

else un\_ed = 0;

if education = "primary" then ed = 1;

else if education = "secondary" then ed = 2;

else if education = "tertiary" then ed = 3;

else if education = "unknown" then ed = .;

if job = "blue-collar" then bl\_jb = 1;

else bl\_jb = 0;

if job = "management" or job = "entreprenuer" or job = "self employed" then mg\_jb = 1;

else mg\_jb =0;

if job = "services" or job = "housemaid" then sr\_jb = 1;

else sr\_jb = 0;

if job = "technician" then tc\_jb = 1;

else tc\_jb = 0;

if job = "admin." then ad\_jb = 1;

else ad\_jb = 0;

if job = "retired" or job = "unemployed" or job = "student" then no\_jb = 1;

else no\_jb = 0;

if job = "blue-collar" then jb = 1;

else if job = "management" or job = "entreprenuer" or job = "self employed" then jb = 2;

else if job = "services" or job = "housemaid" then jb = 3;

else if job = "technician" then jb = 4;

else if job = "admin." then jb = 5;

else if job = "retired" or job = "unemployed" or job = "student" then jb = 6;

if loan = "no" then ln = 0;

else ln = 1;

if marital = "married" then married = 1;

else married = 0;

if age < 30 then ag30 = 1;

else ag30 = 0;

if 30 le age lt 45 then ag45 = 1;

else ag45 = 0;

if 45 le age lt 60 then ag60 = 1;

else ag60 = 0;

if age ge 60 then ag61 = 1;

else ag61 = 0;

if age < 30 then ag = 1;

if 30 le age lt 45 then ag = 2;

if 45 le age lt 60 then ag = 3;

if age ge 60 then ag = 4;

if pdays = -1 then pdays = 2000;

if month = "jan" then jan = 1;

else jan = 0;

if month = "feb" then feb = 1;

else feb = 0;

if month = "mar" then mar = 1;

else mar = 0;

if month = "apr" then apr = 1;

else apr = 0;

if month = "may" then may = 1;

else may = 0;

if month = "jun" then jun = 1;

else jun = 0;

if month = "jul" then jul = 1;

else jul = 0;

if month = "aug" then aug = 1;

else aug = 0;

if month = "sep" then sep = 1;

else sep = 0;

if month = "oct" then oct = 1;

else oct = 0;

if month = "nov" then nov = 1;

else nov = 0;

if month = "dec" then dec = 1;

else dec = 0;

if day <= 15 then first15 = 1;

else first15 = 0;

if balance < 0 then od = 1;

else od = 0;

**run**;

\*\*\* Logistic Regression Model \*\*\*\*\*\*\*\*\*\*\*\*;

/\*Split data into training and validation\*;\*/

**proc** **freq** data = prep;

tables ny;

**run**;

**proc** **surveyselect** data = prep outall out = sample

method = srs samprate = 0.5;

**run**;

**data** training validation;

set sample;

if selected = 0 then output training;

else output validation;

**run**;

**proc** **freq** data = training;

tables ny;

**run**;

**proc** **freq** data = validation;

tables ny;

**run**;

**proc** **logistic** data= training descending outest = modelx outmodel = respm;

model ny = ag45 ag60 ag61 balance duration sec\_ed ter\_ed

ad\_jb no\_jb bl\_jb ln married first15

feb mar apr may aug sep oct nov dec

campaign pdays/outroc=roc1 ;

score out=out1;

**run**;

**proc** **logistic** data= validation descending;

model ny = ag45 ag60 ag61 balance duration sec\_ed ter\_ed

ad\_jb no\_jb bl\_jb ln married first15

feb mar apr may aug sep oct nov dec

campaign pdays/outroc=roc2 ;

score out=out2;

**run**;

\*\* ROC curve \*\*;

symbol1 i=join v=none c=blue;

**proc** **gplot** data=roc1;

title 'ROC plot';

plot \_sensit\_\*\_1mspec\_=1 / cframe=ligr ;

**run**;

\*\* Gain Charts \*\*;

**proc** **rank** data= out2 out = gain groups=10 ties=mean;

var P\_1;

ranks decile;

**run**;

**proc** **sort** data = gain;

by descending P\_1;

**run**;

**proc** **means** n data = gain;

class decile;

var decile;

**run**;

**proc** **means** n data = gain;

class decile;

var ny;

where ny = 1;

**run**;

**proc** **export** data = gain (keep = ny decile) outfile = "D:\Users\Jig14765\Programes\Jig14765\running\_case\_gainschart.csv"

dbms = csv replace;

**run**;