```
/* Generated Code (IMPORT) */
/* Source File: Final Assessment.xls */
/* Source Path: /home/u37548343/my_courses/MKT809 - DG */
/* Code generated on: 6/22/19, 10:15 AM */

FILENAME REFFILE '/home/u37548343/my_courses/MKT809 - DG/final_data.csv';

PROC IMPORT DATAFILE=REFFILE
    DBMS=CSV REPLACE
    OUT=final_data;
    GETNAMES=YES;

RUN;

/* we want to remove the unneeded variables */
data mkt_data;
set final_data(drop=zipcode);
run;

/* QUESTION #1 */
/* 1    While designing a new campaign, can we model the previous campaign's customer behavior to  */
/*  analyze what combination of parameters make a customer more likely to  */
/*  accept a personal loan? */

/* we will factor the data of columns B -> I */
/* according to the output there are only 2 factors with eigen values */
/* greater than, so will use those. It also appears that variables age, education, and mortgage */
/* are deemed  */
proc factor data=mkt_data n=2 rotate=varimax out=factored_data;
var Age Experience Income Family CCAvg Education Mortgage;
run;

/* we will rassess the factor analysis after removing the unimport models */
proc factor data=mkt_data n=2 rotate=varimax out=factored_data;
var Experience Income Family CCAvg;
run;

/* run model with vanilla and factored data and compare Somers' D and c vaues */

/* run a plain vanilla model (multiple logistic regression)*/
/* when running this stepwise model we see similar behavior to the factor analysis */
/* the variables age and mortgage have been excluded from the logistic model */
/* note that stepwise kept the education variable where as we removed it based on  */
/* our personal threshold during factor analysis */

/* somers' d = 0.891; c = 0.945; */
proc logistic data=mkt_data;
model PersonalLoan = Age Experience Income Family CCAvg Education Mortgage / stb selection=stepwise;
run;

/* somers' d = 0.803; c = 0.901; */
proc logistic data=factored_data;
model PersonalLoan = Factor1 Factor2 / stb selection=stepwise;
run;

/* both the vanilla model and the factor model are great predictors for PersonalLoan */
/* based on the somers' d and c values */

/* QUESTION #2 */
/* 2    While designing a new campaign, can we model the previous campaign's customer behavior to  */
/*  analyze what combination of parameters make a customer more likely to  */
/*  accept a securities account? */

/* run model with vanilla and factored data and compare Somers' D and c vaues */
```

```
/* somers' d = null; c = null; */
proc logistic data=mkt_data;
model SecuritiesAccount = Age Experience Income Family CCAvg Education Mortgage / stb selection=stepwise;
run;

/* somers' d = null; c = null; */
proc logistic data=factored_data;
model SecuritiesAccount = Factor1 Factor2 / stb selection=stepwise;
run;

/* run one more time to see variable significance */
proc logistic data=mkt_data;
model SecuritiesAccount = Age Experience Income Family CCAvg Education Mortgage;
run;

/* both the vanilla model and the factor model are terrible predictors for SecuritiesAccount */
/* based on the somers' d and c values */

/* Question #3 */
/* 3    Should the bank use Market Segmentation. If Yes, pose your recommendation    */

/********Cluster Analysis**********/
/* run a proc cluster function on the factored data */
/* analyze the 3 plots and look for the inflection point to determine */
/* the optimal number of clusters, we will simulate 10 clusters for analysis */
ods graphics on;
proc cluster data=factored_data method=ward print=10 ccc pseudo;
var Factor1 Factor2;
copy Age Experience Income Family CCAvg Education Mortgage;
run;

/* displays a tree to show the groups of observations */
/* not printing as we only want the merging of cluster information to */
/* the origially factored dataset, we find 3 to be the optimal cluster count*/
proc tree noprint ncl=3 out=out;
copy Age Experience Income Family CCAvg Education Mortgage Factor1 Factor2;
run;

/* analyze the statistics to describe each cluster by demographic information */
data cl1;
set out;
where cluster=1;
run;

proc univariate data=cl1;
run;

data cl2;
set out;
where cluster=2;
run;

proc univariate data=cl2;
run;

data cl3;
set out;
where cluster=3;
run;

proc univariate data=cl3;
run;

/********Canonical Discriminant Analysis**********/
/* creates composite variables so we can see the groupinngs */
proc candisc data=out(drop=_name_) out = can  distance anova;
```

```
class cluster;
var factor1 factor2;
run;

proc sgplot data = can;
title "Cluster Analysis for Income Datasets";
scatter y = can2 x = can1 / group=cluster;
run;

/* yes the bank should utilize segmentation techniques. according to our cluster analysis */
/* we can see that there are 3 distinct clusters with the following characteristics: */
/*  */
/* CL1 variable_mean; std=1 (68% data) */
/*  age=55;        48-62 yrs */
/*  experience=30; 23-37 yrs */
/*  income=$64k;   $25k-$101k */
/*  */
/* CL2 variable_mean; std=1 */
/*  age=42;        32-52 yrs */
/*  experience=16; 6-26 yrs */
/*  income=$140k;  $104k-$176k */
/*  */
/* CL3 variable_mean; std=1 */
/*  age=40;        30-50 yrs */
/*  experience=15; 5-15 yrs */
/*  income=$54k;   $26k-$82k */
/*   */
/* because we know the characteristics and more importantly he differences among the  */
/* clusters, we can provide more targeting approaches to cross selling bank products. */
/* one of the key differences to capitalize on is the income. these customers can be targeted */
/* for desposits into interest bearing accounts with varying return rates based on the sum of deposits. */
```