**/\*IMPORTING DATA\*/**

PROC IMPORT OUT=mydata

DATAFILE="C:\Users\user\Desktop\Bank.xlsx"

DBMS=EXCEL REPLACE;

GETNAMES=YES;

RUN;

**/\*SAMPLING DATA\*/**

proc surveyselect data = work.Mydata

out= work.samp

method = srs

sampsize = 2500

seed = 3653317;

run;

proc print data = hsbs2 noobs;

run;

data Project9.sample;

set work.samp;

run;

**/\*Label statement to link variable names to descriptions\*/**

DATA samp; SET work.samp;

LABEL adbdda='Average daily balance in checking account'

atres='Lentgh of time at current residence'

income='Monthly income'

atmct='Number of ATM transactions per month'

savbal='Saving account balance'

ddatot='Total amount of checking transactions'

ddadep='Total amount of checking deposits'

acquire='The indicator of product acquisition(0=No, 1=Yes)'

invest='Amount of investment';

run;

proc means data=samp;

run;

**/\*Format statement for categorical variables\*/**

PROC FORMAT;

VALUE acquire 0='N0'

1='YES';

PROC PRINT DATA=work.SAMP;

VAR acquire;

FORMAT acquire;

run;

**/\*Checking distribution of data -e.g for income\*/**

PROC MEANS DATA=SAMP;

RUN;

PROC UNIVARIATE DATA=SAMP;

VAR income;

HISTOGRAM income;

RUN;

**/\*Try to transform non-normal variables\*/**

DATA work.SAMP;

SET work.samp;

new\_income=exp(income);

run;

**/\*QQ-NORMAL PLOTS FOR VARIABLES\*/**

PROC UNIVARIATE DATA=work.SAMP;

VAR income atmct ddadep savbal;

qqplot income atmct ddadep savbal;

RUN;

**/\*SORTING THE DATA\*/**

PROC SORT DATA=work.SAMP;

BY atmct ddadep savbal income adbdda atres ddatot ddadep invest;

run;

**Frequency tables**

PROC FREQ DATA=SAMP;

TABLES acquire;

TITLE 'The frequency table for product acquisition';

RUN;

**/\*grouping invest\*/**

**PROC** **FORMAT** library=fmt;

VALUE $fmtinvest **0**='no investment'

**0**-**10000**='small investo'

**10001**-**100000**='medium investor'

**100000**-high='big investor' ;

**RUN**;

ods html;

**PROC** **FREQ** order=formatted DATA=work.samp;

TABLES invest;

format invest $fmtinvest.;

**RUN**;

ods html close;

**/\*Frequency table for acquire by invest\*/**

PROC FREQ data=work.Sorted\_samp;

Tables acquire\*invest;

Run;

**/\*SUMMARY STATS \*/**

PROC MEANS DATA=WORK.SAMP;

RUN;

**IDENTIFY MISSING VALUES**

**proc** **means** data=work.samp NMISS N;

**run**;

**/\*Descriptive statistics for the two acquire groups\*/**

ODS RTF;

proc means data=work.samp NMISS N mean std min max clm;

by acquire;

run;

ODS RTF CLOSE;

PROC SORT DATA=work.SAMP;

BY acquire;

run;

ods graphics on;

proc corr data=work.samp nomiss plots=matrix(histogram);

var acquire invest income;

run;

ods graphics off;

**/\* Group invest\*/**

PROC FORMAT;

VALUE invest

0 = '0'

0-10000 = '0-10,000'

10000-100000 = '10,001-100,000'

100000-high = '100,000+'

;

RUN;

proc freq data=work.samp order=formatted;

tables invest;

format invest invest.;

run;

**/\*Amount of investment by acquire\*/**

ODS RTF;

proc freq data=work.samp order=formatted;

tables invest\*ACQUIRE;

format invest invest.;

run;

ODS RTF CLOSE;

proc means data=work.samp mean std min max clm;

by invest;

format invest invest.;

run;

ods rtf;

proc freq data=work.samp order=formatted;

tables invest\*acquire/ nocol norow;

format invest invest.;

run;

ods rtf close;