**Customer Segmentation: RFM Model & K-means Clustering Using SAS**

/\*Create your own library in SAS like here it is libref and mention the path \*/

libname libref “/home/aro1260/deep”;

/\*Importing customer\_seg dataset \*/

PROC IMPORT DATAFILE= “/home/aro1260/data/customer\_seg.csv”

DBMS=CSV Replace

OUT=libref.synthetic\_RFM;

GETNAMES=YES;

RUN;

/\*To check the contents of the data \*/

PROC CONTENTS DATA=libref.synthetic\_RFM;

RUN;

/\*Descriptive statistics of the data\*/

proc means data = libref.synthetic\_RFM;

vars Amount;

class Product\_Category;

run;

/\*Applying proc univariate for detailed summary statistics \*/

proc univariate data= libref.synthetic\_RFM;

var Amount;

run;

/\*Printing top 10 observations of the data \*/

PROC PRINT Data =libref.synthetic\_RFM(OBS=10);

Title 'Total Sales by Customer';

ID Customer\_ID;

Var Invoice\_No Invoice\_Date l\_Date Amount;

Run;

/\*Creating the Customer\_summary table by applying proc sql query \*/

PROC SQL;

Create table libref.Customer\_summary as

select distinct Customer\_ID,

max(Invoice\_Date) as Recency format = date9.,

(l\_Date - max(Invoice\_Date)) as Days\_since\_recent,

count(Invoice\_Date) as Frequency,

Sum(Amount) as Monetary format=dollar15.2

from libref.synthetic\_RFM

group by Customer\_ID;

quit;

/\*Printing top 10 observations of the data \*/

PROC PRINT Data =libref.Customer\_summary(OBS=10);

Run;

/\* creating rank \*/

PROC RANK DATA = libref.Customer\_summary OUT = libref.RFM ties=high group=5;

Var Days\_since\_recent Frequency Monetary;

Ranks R\_score F\_score M\_score ;

RUN;

/\* combining RFM \*/

DATA libref.Final\_RFM;

Set libref.RFM;

R\_score+1;

F\_score+1;

M\_score+1;

RFM = Cats(R\_score,F\_score,M\_score);

Total\_RFM = SUM(R\_score,F\_score,M\_score);

RUN;

/\*Printing top 10 observations of the data \*/

PROC PRINT Data =libref.Final\_RFM(OBS=10);

Run;

/\*Applying PROC FASTCLUS \*/

PROC FASTCLUS data= libref.Final\_RFM maxclusters=5 maxiter=100 converge=0 random=121 replace=random out=libref.Clust;

Var Days\_since\_recent Frequency Monetary;

Title 'FASTCLUS ANALYSIS';

run;

/\*CANDISC procedure to compute canonical variables for plotting the clusters \*/

proc candisc data = libref.Clust out=libref.Can noprint;

var Days\_since\_recent Frequency Monetary;

class cluster;

Run;

/\*Plots the two canonical variables generated from PROC CANDISC, can1 and can2 \*/

proc sgplot data= libref.Can;

scatter y=can2 x=can1 / group=cluster;

run;