/\* create a library "mylib" \*/

libname mylib '\\client\C$\Users\prady\_000\Documents\Vishwa\MSDA\practicum1\Exam1';

/\* Import "FA12\_Data" dataset containing 25564 records into a new dataset "bigrec" \*/

filename bigrec '\\client\C$\Users\prady\_000\Documents\Vishwa\MSDA\practicum1\Exam1\FA12\_Data.txt' lrecl =**65576**;

/\* create "mytemp", that copies data from bigrec but for only certain variables, selectively, by specifying the positions from where

the data should be extracted\*/

**data** mylib.mytemp;

infile bigrec;

input my\_id **1**-**7**

/\* NORMALLY COUNT CALORIES THE FOODS I EAT \*/

cntcal\_agree\_alot **4143**

cntcal\_agree\_little **4188**

cntcal\_neither **4278**

cntcal\_disagree\_little **4323**

cntcal\_disagree\_alot **4368**

/\* FATTENING FOOD MAKES ME FEEL GUILTY \*/

fatglty\_agree\_alot **4166**

fatglty\_agree\_little **4211**

fatglty\_neither **4301**

fatglty\_disagree\_little **4346**

fatglty\_disagree\_alot **4391**

/\* I OFTEN FEEL I OVER-EAT \*/

overeat\_agree\_alot **4156**

overeat\_agree\_little **4201**

overeat\_neither **4291**

overeat\_disagree\_little **4336**

overeat\_disagree\_alot **4381**

/\* MOST TIME TRYING LOSE WEIGHT BY DIETING \*/

diet\_agree\_alot **4147**

diet\_agree\_little **4192**

diet\_neither **4282**

diet\_disagree\_little **4327**

diet\_disagree\_alot **4372**

/\* PEOPLE ASK MY OPINION/BUYING NEW TCHNLGY \*/

opinion\_agree\_alot **6836**

opinion\_agree\_little **6853**

opinion\_neither **6887**

opinion\_disagree\_little **6904**

opinion\_disagree\_alot **6921**

/\* I TRY KEEP UP/DEVELOPMENTS IN TECHNOLOGY \*/

keepup\_agree\_alot **6833**

keepup\_agree\_little **6850**

keepup\_neither **6884**

keepup\_disagree\_little **6901**

keepup\_disagree\_alot **6918**

/\* BUY TECHNOLOGY/CONNECTS TO PRODS I HAVE \*/

buy\_agree\_alot **6837**

buy\_agree\_little **6854**

buy\_neither **6888**

buy\_disagree\_little **6905**

buy\_disagree\_alot **6922**

/\* I MAKE SURE I EXERCISE REGULARLY \*/

exreg\_agree\_alot **3925**

exreg\_agree\_little **3943**

exreg\_neither **3979**

exreg\_disagree\_little **3997**

exreg\_disagree\_alot **4015**

/\* XTRA FEATURES MRE IMPT THN TRAD FEATURES \*/

extfeat\_agree\_alot **3752**

extfeat\_agree\_little **3776**

extfeat\_neither **3824**

extfeat\_disagree\_little **3848**

extfeat\_disagree\_alot **3872**

/\*USE INFO FROM CELL TO DECIDE FREE TIME \*/

cellinfo\_agree\_alot **3759**

cellinfo\_agree\_little **3783**

cellinfo\_neither **3831**

cellinfo\_disagree\_little **3855**

cellinfo\_disagree\_alot **3879**

/\* LIKE IDEA OF CELL MOVING BYND VOICE/TEXT \*/

beyond\_agree\_alot **3767**

beyond\_agree\_little **3791**

beyond\_neither **3839**

beyond\_disagree\_little **3863**

beyond\_disagree\_alot **3887**

/\* MY CELL PHONE SHOULD HELP GET WORK DONE \*/

workdone\_agree\_alot **3773**

workdone\_agree\_little **3797**

workdone\_neither **3845**

workdone\_disagree\_little **3869**

workdone\_disagree\_alot **3893**

/\* DEMOGRAPHIC VARIABLES \*/

AGE\_18TO24 **2299**

AGE\_25TO49 **2305**

AGE\_50PLUS **2313**

GRADUATED\_HIGHSCHOOL **2335**

FEMALE **2282**

WEBAPP\_LAST12MON\_YES **9143**

FB\_LAST7DAYS **9313**

;

**run**;

/\* create another dataset "mycalcs" that copies mytemp \*/

**data** mycalcs;

set mylib.mytemp;

/\* create an 2-D array of size 12X5 for 12 questions and its 5 available options \*/

array missy(**12**,**5**)

/\* NORMALLY COUNT CALORIES THE FOODS I EAT \*/

cntcal\_agree\_alot

cntcal\_agree\_little

cntcal\_neither

cntcal\_disagree\_little

cntcal\_disagree\_alot

/\* FATTENING FOOD MAKES ME FEEL GUILTY \*/

fatglty\_agree\_alot

fatglty\_agree\_little

fatglty\_neither

fatglty\_disagree\_little

fatglty\_disagree\_alot

/\* I OFTEN FEEL I OVER-EAT \*/

overeat\_agree\_alot

overeat\_agree\_little

overeat\_neither

overeat\_disagree\_little

overeat\_disagree\_alot

/\* MOST TIME TRYING LOSE WEIGHT BY DIETING \*/

diet\_agree\_alot

diet\_agree\_little

diet\_neither

diet\_disagree\_little

diet\_disagree\_alot

/\* PEOPLE ASK MY OPINION/BUYING NEW TCHNLGY \*/

opinion\_agree\_alot

opinion\_agree\_little

opinion\_neither

opinion\_disagree\_little

opinion\_disagree\_alot

/\* I TRY KEEP UP/DEVELOPMENTS IN TECHNOLOGY \*/

keepup\_agree\_alot

keepup\_agree\_little

keepup\_neither

keepup\_disagree\_little

keepup\_disagree\_alot

/\* BUY TECHNOLOGY/CONNECTS TO PRODS I HAVE \*/

buy\_agree\_alot

buy\_agree\_little

buy\_neither

buy\_disagree\_little

buy\_disagree\_alot

/\* I MAKE SURE I EXERCISE REGULARLY \*/

exreg\_agree\_alot

exreg\_agree\_little

exreg\_neither

exreg\_disagree\_little

exreg\_disagree\_alot

/\* XTRA FEATURES MRE IMPT THN TRAD FEATURES \*/

extfeat\_agree\_alot

extfeat\_agree\_little

extfeat\_neither

extfeat\_disagree\_little

extfeat\_disagree\_alot

/\*USE INFO FROM CELL TO DECIDE FREE TIME \*/

cellinfo\_agree\_alot

cellinfo\_agree\_little

cellinfo\_neither

cellinfo\_disagree\_little

cellinfo\_disagree\_alot

/\* LIKE IDEA OF CELL MOVING BYND VOICE/TEXT \*/

beyond\_agree\_alot

beyond\_agree\_little

beyond\_neither

beyond\_disagree\_little

beyond\_disagree\_alot

/\* MY CELL PHONE SHOULD HELP GET WORK DONE \*/

workdone\_agree\_alot

workdone\_agree\_little

workdone\_neither

workdone\_disagree\_little

workdone\_disagree\_alot ;

/\* loop through the array to check for missing values in any of the 5 cells for each of 25564 records

and make it zero when found, else, do nothing\*/

do i = **1** to **12**;

do j = **1** to **5**;

if missy(i,j) = **.** then missy(i,j) = **0**;

end;

end;

/\* compute the sum of each record by adding all the columns and drop the looping variables \*/

array mysum(**12**);

do k = **1** to **12**;

mysum(k) = missy(k,**1**)+ missy(k,**2**)+missy(k,**3**) + missy(k,**4**) + missy(k,**5**);

end;

drop i; drop j;

/\* define a variable to capture the option selected by the person if mysum is 1 else mark it as missing \*/

array myvar(**12**);

do m = **1** to **12**;

if mysum(m) = **1** then

myvar(m) = (missy(m,**1**)\***5**) + (missy(m,**2**)\***4**)+ (missy(m,**3**)\***3**) + (missy(m,**4**))\***2** + (missy(m,**5**)\***1**);

else

myvar(m) = **.**;

end;

/\* myvar has the either the option selected for the question or missing field indicating nothing is selected \*/

/\* naming it relevant to the question of interest \*/

cntcal = myvar(**1**);

fatglty = myvar(**2**);

overeat = myvar(**3**);

diet = myvar(**4**);

opinion = myvar(**5**);

keepup = myvar(**6**);

buy = myvar(**7**);

exreg = myvar(**8**);

extfeat = myvar(**9**);

cellinfo = myvar(**10**);

beyond = myvar(**11**);

workdone = myvar(**12**);

/\* make sum variables available \*/

mysum1=mysum(**1**);

mysum2=mysum(**2**);

mysum3=mysum(**3**);

mysum4=mysum(**4**);

mysum5=mysum(**5**);

mysum6=mysum(**6**);

mysum7=mysum(**7**);

mysum8=mysum(**8**);

mysum9=mysum(**9**);

mysum10=mysum(**10**);

mysum11=mysum(**11**);

mysum12=mysum(**12**);

**run**;

**proc** **freq** data=mycalcs;

/\* check incidence of multiple check marks \*/

tables

mysum1

mysum2

mysum3

mysum4

mysum5

mysum6

mysum7

mysum8

mysum9

mysum10

mysum11

mysum12

;

**run**;

/\* cleaning the demographic variables \*/

**data** demo;

set mycalcs;

if AGE\_18TO24 = **1** then

AGE = **1**;

ELSE

IF AGE\_25TO49 = **1** THEN

AGE = **2**;

ELSE

IF AGE\_50PLUS = **1** THEN

AGE = **3**;

if female =**.** then female=**0**;

if WEBAPP\_LAST12MON\_YES = **.** then WEBAPP\_LAST12MON\_YES = **0**;

if FB\_LAST7DAYS = **.** then FB\_LAST7DAYS =**0**;

if graduated\_highschool=**.** then graduated\_highschool=**0**;

**RUN**;

/\* now sniff new variables and compare - should be approx same as cell counts \*/

**proc** **freq** data=demo;

tables

cntcal

fatglty

overeat

diet

opinion

keepup

buy

exreg

extfeat

cellinfo

beyond

workdone

female

WEBAPP\_LAST12MON\_YES

FB\_LAST7DAYS

GRADUATED\_HIGHSCHOOL

age;

**run**;

**data** mylib.myfactor;

set mycalcs;

**run**;

/\* do the factor analysis - set eigenval rotate=varimax scree PCA \*/

**proc** **factor** data = mylib.myfactor

maxiter=**100**

method=principal

mineigen=**1**

rotate=varimax

scree

score

print

nfactors=**2**

plots=scree

out=myscores;

var cntcal

fatglty

overeat

diet

cellinfo

extfeat

beyond

workdone ;

**run**;

/\* rename the factor variables \*/

**data** mylib.myscores1;

set myscores;

rename factor1 = cell;

rename factor2 = wtcon;

rename my\_id = resp\_id;

**run**;

/\* k-means Clustering \*/

**data** letscluster;

set mylib.myscores1;

**proc** **fastclus** data=letscluster maxclusters=**3** out=clus3;

var cell wtcon opinion keepup buy exreg ;

**run**;

**proc** **fastclus** data=letscluster maxclusters=**4** out=mylib.clus4;

var cell wtcon opinion keepup buy exreg;

**run**;

**proc** **fastclus** data=letscluster maxclusters=**5**;

var cell wtcon opinion keepup buy exreg;

**run**;

**proc** **fastclus** data=letscluster maxclusters=**6**;

var cell wtcon opinion keepup buy exreg;

**run**;

**proc** **fastclus** data=letscluster maxclusters=**7**;

var cell wtcon opinion keepup buy exreg ;

**run**;

/\* Gap analysis \*/

**proc** **hpclus** data=myscores1 maxclusters=**7**

noc=abc(b=**20** minclusters=**3** align=pca criterion=firstpeak);

input cell wtcon opinion keepup buy exreg / level=interval;

**run**;

/\* Choosing k-means with 4 clusters \*/

**proc** **fastclus** data=letscluster maxclusters=**4** out=mylib.clus4;

var cell wtcon opinion keepup buy exreg;

**run**;

/\* cluster visualization \*/

**proc** **candisc** data=mylib.clus4 out= can3 noprint ;

var cell wtcon opinion keepup buy exreg;

class cluster;

**proc** **sgplot** data=can3;

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

**run**;

/\* differentiate clusters based on demographic and media descriptors/non-drivers \*/

**PROC** **MEANS** DATA=demo

FW=**12**

PRINTALLTYPES

CHARTYPE

NWAY

VARDEF=DF

MEAN

STD

N ;

VAR FEMALE WEBAPP\_LAST12MON\_YES FB\_LAST7DAYS AGE graduated\_highschool;

CLASS CLUSTER / ORDER=UNFORMATTED ASCENDING;

**RUN**;