**data** panel\_dr;

infile 'H:\Final Project\laundet\_PANEL\_DR\_1114\_1165.csv' firstobs = **2** dsd missover;

INPUT PANID WEEK UNITS OUTLET $ DOLLARS IRI\_KEY COLUPC;

**run**;

/\* PROC PRINT data = panel\_dr (obs = 10); RUN; \*/

**data** panel\_gr;

infile 'H:\Final Project\laundet\_PANEL\_GR\_1114\_1165.csv' firstobs = **2** dsd missover;

INPUT PANID WEEK UNITS OUTLET $ DOLLARS IRI\_KEY COLUPC;

**run**;

/\* PROC PRINT data = panel\_gr (obs = 10); RUN; \*/

**data** panel\_ma;

infile 'H:\Final Project\laundet\_PANEL\_MA\_1114\_1165.csv' firstobs = **2** dsd missover;

INPUT PANID WEEK UNITS OUTLET $ DOLLARS IRI\_KEY COLUPC;

**run**;

/\* PROC PRINT data = panel\_ma (obs = 10); RUN; \*/

**data** panel\_sales;

set panel\_dr panel\_ma panel\_gr;

**run**;

/\* proc print data = panel\_sales (obs = 10); run; \*/

/\* proc print data = panel\_sales (obs = 10); run; \*/

**data** brands(keep= L2 L4 L5 PRODUCT\_TYPE PACKAGE FORM FLAVOR\_SCENT COLUPC VOL\_EQ\_NUM);

infile 'H:\Final Project\prod\_laundet.csv' DLM = ',' firstobs = **2** dsd missover;

length L1 $**100** L2 $**100** L3 $**100** L4 $**100** L5 $**100** L9 $**100** Level UPC $**100** SY GE VEND ITEM STUB\_SPEC\_1860RC\_00004 $**100** VOL\_EQ PRODUCT\_TYPE $**100** PACKAGE $**100** FORM $**100** FLAVOR\_SCENT $**100** CONCENTRATION\_LEVEL $**100** ADDITIVES $**100** TYPE\_OF\_FORMULA $**100**;

INPUT L1 $ L2 $ L3 $ L4 $ L5 $ L9 $ Level UPC $ SY GE VEND ITEM STUB\_SPEC\_1860RC\_00004 $ VOL\_EQ PRODUCT\_TYPE $ PACKAGE $ FORM $ FLAVOR\_SCENT $ CONCENTRATION\_LEVEL $ ADDITIVES $ TYPE\_OF\_FORMULA $;

L1 = strip(L1);

L2 = strip(L2);

L3 = strip(L3);

L4 = strip(L4);

L5 = strip(L5);

L9 = strip(L9);

UPC = strip(UPC);

STUB\_SPEC\_1860RC\_00004 = strip(STUB\_SPEC\_1860RC\_00004);

PRODUCT\_TYPE = strip(PRODUCT\_TYPE);

PACKAGE= strip(PACKAGE);

FORM= strip(FORM);

FLAVOR\_SCENT = strip(FLAVOR\_SCENT);

CONCENTRATION\_LEVEL = strip(CONCENTRATION\_LEVEL);

ADDITIVES = strip(ADDITIVES);

TYPE\_OF\_FORMULA = strip(TYPE\_OF\_FORMULA);

format COLUPC\_char $13. COLUPC **13.**;

len\_item = length(ITEM);

len\_vend = length(VEND);

if len\_vend < **5** then VEND = trim(repeat('0',**5**-len\_vend-**1**)||VEND);

if len\_item < **5** then ITEM = trim(repeat('0',**5**-len\_item-**1**)||ITEM);

COLUPC\_char = compress(CATT(trim(SY),trim(GE),trim(VEND),trim(ITEM)));

COLUPC = COLUPC\_char\***1**;

VOL\_EQ\_NUM = VOL\_EQ\***1**;

**run**;

**proc** **sql**;

create table brand\_details as

select

a.\*,

b.L5,

case

when b.L5 in ('ALL TIDE PRODUCTS' , 'TIDE', 'TIDE KICK' , 'TIDE SIMPLE PLEASURES') then 'TIDE'

when b.L5 in ('ALL' ,'ALL BABY CARE PLUS' , 'ALL SMALL & MIGHTY' , 'ALL SURF') then 'ALL'

when b.L5 in ('BABY SOFT PUREX' , 'PUREX', 'PUREX ADVANCED', 'PUREX BABY') then 'PUREX'

when b.L5 in ('WISK' , 'WISK POWER SCOOP') then 'WISK'

else 'OTHER' end as brand

from panel\_sales a

inner join brands b on a.COLUPC = b.COLUPC

where b.L2 = 'LIQUID LAUNDRY DETERGENT' and b.PRODUCT\_TYPE = 'LAUNDRY DETERGENT' and FORM = 'LIQUID';

**quit**;

**proc** **print** data = brand\_details (obs = **10**); **run**;

**proc** **sql**;

create table pre\_loyal as

SELECT brand\_details.panid, (counts/count(brand\_details.panid)) as loyality

FROM brand\_details, (SELECT brand\_details.panid, count(brand\_details.panid) as counts FROM brand\_details WHERE L5 = 'ALL' GROUP BY panid) as X

WHERE brand\_details.panid = X.panid GROUP BY brand\_details.panid;

**quit**;

**proc** **print** data = pre\_loyal (obs = **25**); **run**;

/\* ------------------------------------------------------------------- \*/

/\* Segmentation using only RFM Variables - Liquid + Laundary Products \*/

/\* ------------------------------------------------------------------- \*/

**proc** **sql**;

create table pre\_rfm as

Select panid

,mean(dollars) as monetary

,count(week) as frequency

,min(**1165**-week) as recency

from panel\_sales

group by panid;

**quit**;

**PROC** **PRINT** data = pre\_rfm (obs = **10**); **run**;

**proc** **corr** data=pre\_rfm;

var monetary frequency recency;

**run**;

/\*PROC SQL;

create table pre\_rfml as

SELECT \* FROM pre\_rfm a, pre\_loyal b WHERE a.PANID = b.PANID;

run;

proc print data = pre\_rfml (obs = 10); run;

proc corr data=pre\_rfml;

var monetary frequency recency loyality;

run;\*/

/\* Standardize the data to z-score \*/

**PROC** **STANDARD** data = pre\_rfm mean = **0** std = **1** out = stand;

VAR monetary frequency recency; **RUN**;

**proc** **print** data = stand(obs = **10**); **run**;

**proc** **sql**;

create table pre\_loyal\_tide as

SELECT brand\_details.panid, (counts/count(brand\_details.panid)) as loyality\_tide

FROM brand\_details, (SELECT brand\_details.panid, count(brand\_details.panid) as counts FROM brand\_details WHERE L5 = 'TIDE'

GROUP BY panid) as X

WHERE brand\_details.panid = X.panid GROUP BY brand\_details.panid;

**quit**;

**proc** **sql**;

create table pre\_loyal as

SELECT brand\_details.panid, (counts/count(brand\_details.panid)) as loyality\_All

FROM brand\_details, (SELECT brand\_details.panid, count(brand\_details.panid) as counts FROM brand\_details WHERE L5 = 'ALL'

GROUP BY panid) as X

WHERE brand\_details.panid = X.panid GROUP BY brand\_details.panid;

**quit**;

**PROC** **SQL**;

create table temp1 as

SELECT \* FROM pre\_rfm a LEFT OUTER JOIN pre\_loyal\_tide b ON a.PANID = b.PANID;

**run**;

**PROC** **SQL**;

create table temp2 as

SELECT \* FROM temp1 a LEFT OUTER JOIN pre\_loyal b ON a.PANID = b.PANID;

**run**;

**proc** **print** data = temp2(obs = **50**); **run**;

**data** pre\_rfm;

set temp2;

if loyality\_tide = "" then loyality\_tide = **0**;

if loyality\_All = "" then loyality\_all = **0**;

**run**;

**proc** **print** data = pre\_rfm(obs = **50**); **run**;

/\* Create Cluster \*/

**PROC** **fastclus** data=pre\_rfm maxclusters = **10** OUT=clust;

var monetary recency frequency loyality\_tide loyality\_All; **run**;

**proc** **print** data = clust(obs = **10**); **run**;

/\* Sort data \*/

**PROC** **SORT** data = clust out = clust\_sort (keep = PANID Cluster);

by PANID; **RUN**;

**PROC** **SORT** data = pre\_rfm out = pre\_rfm;

by PANID; **RUN**;

/\* Joining Clusters \*/

**PROC** **SQL**;

create table RFM as

SELECT \* FROM pre\_rfm a, clust b WHERE a.PANID = b.PANID;

**run**;

**proc** **print** data = RFM (obs = **10**); **run**;

/\* Descriptive Analysis \*/

title 'ALL THE BRANDS AND LIQUID CATEGORY';

**PROC** **MEANS** data = RFM;

class cluster;

VAR recency frequency monetary loyality\_tide loyality\_All;

OUTPUT out = temp(drop = \_Type\_);

**RUN**;

**proc** **export** data=RFM

outfile='H:\Final Project\10Clusters.csv'

dbms=csv

replace;

**run**;

/\* ---------------------------------------------------------------------- \*/

/\* Segmentation using only RFM Variables for LAUNDRY + LIQUID + ALL Brand \*/

/\* ---------------------------------------------------------------------- \*/

**proc** **sql**;

create table brand\_details\_ALL as

select

a.\* FROM brand\_details a where a.L5 = 'ALL'; **quit**;

**proc** **sql**;

create table pre\_rfm\_all as

Select panid

,mean(dollars) as monetary

,count(week) as frequency

,min(**1165**-week) as recency

from brand\_details\_ALL

group by **1**;

**quit**;

**PROC** **SQL**;

create table pre\_rfml\_all as

SELECT \* FROM pre\_rfm\_all a, pre\_loyal b WHERE a.PANID = b.PANID;

**run**;

**proc** **print** data = pre\_rfml\_all (obs = **10**); **run**;

**proc** **corr** data=pre\_rfml\_all;

var monetary frequency recency loyality;

**run**;

**proc** **print** data = pre\_rfml\_all (obs = **10**); **run**;

/\* Standardize the data to z-score \*/

**PROC** **STANDARD** data = pre\_rfml\_all mean = **0** std = **1** out = stand2;

VAR loyality monetary frequency recency; **RUN**;

**proc** **print** data = stand2(obs = **10**); **run**;

/\* Create Cluster \*/

**PROC** **fastclus** radius = **0** replace = FULL data=stand2 maxclusters = **4** maxiter = **20** OUT=clust2 list distance;

var loyality monetary frequency recency; **run**;

/\* Sort data \*/

**PROC** **SORT** data = clust2 out = clust\_sort2 (keep = PANID Cluster);

by PANID; **RUN**;

**PROC** **SORT** data = pre\_rfml\_all out = pre\_rfm2;

by PANID; **RUN**;

/\* Joining Clusters \*/

**PROC** **SQL**;

create table RFM2 as

SELECT \* FROM pre\_rfml\_all a, clust2 b WHERE a.PANID = b.PANID;

**run**;

**proc** **print** data = RFM2 (obs = **10**); **run**;

/\* Descriptive Analysis \*/

title 'LAUNDRY + LIQUID + ALL Brand';

**PROC** **MEANS** data = RFM2;

class cluster;

VAR loyality recency frequency monetary ;

OUTPUT out = temp(drop = \_Type\_);

**RUN**;

/\* ---------------------------------------------------------------------- \*/

/\* Segmentation using only RFM Variables for LAUNDRY + LIQUID + TIDE Brand \*/

/\* ---------------------------------------------------------------------- \*/

**proc** **sql**;

create table brand\_details\_TIDE as

select

a.\* FROM brand\_details a where a.L5 = 'TIDE'; **quit**;

**proc** **sql**;

create table pre\_loyal\_tide as

SELECT brand\_details.panid, (counts/count(brand\_details.panid)) as loyality

FROM brand\_details, (SELECT brand\_details.panid, count(brand\_details.panid) as counts FROM brand\_details WHERE L5 = 'TIDE'

GROUP BY panid) as X

WHERE brand\_details.panid = X.panid GROUP BY brand\_details.panid;

**quit**;

**proc** **sql**;

create table pre\_rfm\_tide as

Select panid

,mean(dollars) as monetary

,count(week) as frequency

,min(**1165**-week) as recency

from brand\_details\_TIDE

group by **1**;

**quit**;

**PROC** **SQL**;

create table pre\_rfml\_tide as

SELECT \* FROM pre\_rfm\_tide a, pre\_loyal\_tide b WHERE a.PANID = b.PANID;

**run**;

**proc** **print** data = pre\_rfml\_tide (obs = **10**); **run**;

**proc** **corr** data=pre\_rfml\_tide;

var monetary frequency recency loyality;

**run**;

/\* Standardize the data to z-score \*/

**PROC** **STANDARD** data = pre\_rfml\_tide mean = **0** std = **1** out = stand3;

VAR loyality monetary frequency recency; **RUN**;

**proc** **print** data = stand3(obs = **10**); **run**;

/\* Create Cluster \*/

**PROC** **fastclus** radius = **0** replace = FULL data=stand3 maxclusters = **4** maxiter = **20** OUT=clust3 list distance;

var loyality monetary frequency recency; **run**;

/\* Sort data \*/

**PROC** **SORT** data = clust3 out = clust\_sort3 (keep = PANID Cluster);

by PANID; **RUN**;

**PROC** **SORT** data = pre\_rfml\_tide out = pre\_rfm3;

by PANID; **RUN**;

/\* Joining Clusters \*/

**PROC** **SQL**;

create table RFM3 as

SELECT \* FROM pre\_rfml\_tide a, clust3 b WHERE a.PANID = b.PANID;

**run**;

**proc** **print** data = RFM3 (obs = **10**); **run**;

/\* Descriptive Analysis \*/

title 'LAUNDRY + LIQUID + TIDE Brand';

**PROC** **MEANS** data = RFM3;

class cluster;

VAR loyality monetary frequency recency;

OUTPUT out = temp(drop = \_Type\_);

**RUN**;

/\* ------------------------------------------------------------------------------------------------ \*/

/\* Segmentation using only RFM Variables for LAUNDRY + LIQUID + ALL Brand + DEMOGRAPHIC INFORMATION \*/

/\* ------------------------------------------------------------------------------------------------ \*/

**proc** **sql**;

create table brand\_details\_ALL as

select

a.\* FROM brand\_details a where a.L5 = 'ALL'; **quit**;

**data** adsdemo;

infile 'H:\Final Project\ads demo1.csv' DLM = ',' firstobs = **2** dsd missover;

INPUT PANID PANTYPE PreTaxInc FamSize HH\_Race Type\_Res\_Poss County HH\_Age HH\_Edu HH\_Occ

Male\_Age Male\_Edu Male\_Occ Male\_Whc Male\_Smoke Female\_Age Female\_Edu Female\_Occ Female\_Whc Female\_Smoke Dogs

Cats Child\_GC Marital Language TV\_used TV\_hooked Year HISP\_Flag HISP\_Cat Race2 Race3

MW Zipcode Fipscode Market\_zipcode IRI\_Geonum Ext\_fact;

DROP County HH\_Age HH\_Edu HH\_Occ Male\_Smoke Female\_Smoke Language HISP\_Flag HISP\_Cat Race2 Race3 MW Zipcode Fipscode

Market\_zipcode IRI\_Geonum Ext\_fact; **RUN**;

**PROC** **PRINT** DATA = adsdemo (obs = **12**);**RUN**;

/\* Demographic Feature Selection - Liquid Category \*/

**PROC** **SQL**;

create table adsdemo\_rfml as

SELECT \* FROM adsdemo a, RFM b WHERE a.PANID = RFM.PANID;

**run**;

**proc** **print** data=adsdemo\_rfml(obs = **10**);

**run**;

**proc** **sort** data = adsdemo\_rfml; by cluster; **run**;

**proc** **export** data=adsdemo\_rfml

outfile='H:\Final Project\demo\_everything.csv'

dbms=csv

replace;

**run**;

/\* Demographic Feature Selection - Liquid Category - All Brand \*/

**PROC** **SQL**;

create table adsdemo\_rfml\_all as

SELECT \* FROM adsdemo a, RFM2 b WHERE a.PANID = RFM2.PANID;

**run**;

**proc** **print** data=adsdemo\_rfml\_all(obs = **10**);

**run**;

**proc** **sort** data = adsdemo\_rfml\_all; by cluster; **run**;

**proc** **export** data=adsdemo\_rfml\_all

outfile='H:\Final Project\demo\_tide.csv'

dbms=csv

replace;

**run**;

/\* Demographic Feature Selection - Liquid Category - Tide Brand \*/

**PROC** **SQL**;

create table adsdemo\_rfml\_tide as

SELECT \* FROM adsdemo a, RFM3 b WHERE a.PANID = RFM3.PANID;

**run**;

**proc** **print** data=adsdemo\_rfml\_tide (obs = **10**);

**run**;

**proc** **sort** data = adsdemo\_rfml\_tide; by cluster; **run**;

**proc** **export** data=adsdemo\_rfml\_tide

outfile='H:\Final Project\demo\_all.csv'

dbms=csv

replace;

**run**;

/\* Selecting Important Variable \*/

**proc** **sql**;

create table cust\_desc as

select

b.PANID,

case

when b.FamSize in (**4**,**5**,**6**) then 'large'

when b.FamSize in (**1**,**2**,**3**) then 'regular'

else 'other' end as fam\_size,

case

when b.PreTaxInc in (**1**,**2**,**3**,**4**) then 'low'

when b.PreTaxInc in (**5**,**6**,**7**,**8**) then 'medium'

when b.PreTaxInc in (**9**,**10**,**11**) then 'high'

when b.PreTaxInc in (**12**) then 'very\_high'

else 'other' end as fam\_income,

case

when b.Male\_Age in (**1**) then 'young'

when b.Male\_Age in (**2**,**3**,**4**) then 'mid\_career'

when b.Male\_Age in (**5**,**6**) then 'elder'

else 'other' end as age\_male,

case

when b.Female\_Age in (**1**) then 'young'

when b.Female\_Age in (**2**,**3**,**4**) then 'mid\_career'

when b.Female\_Age in (**5**,**6**) then 'elder'

else 'other' end as age\_female,

case

when b.Male\_Edu in (**1**,**2**,**3**) then 'school'

when b.Male\_Edu in (**4**,**5**,**6**) then 'college'

when b.Male\_Edu in (**7**,**8**) then 'graduate'

else 'other' end as educ\_male,

case

when b.Female\_Edu in (**1**,**2**,**3**) then 'school'

when b.Female\_Edu in (**4**,**5**,**6**) then 'college'

when b.Female\_Edu in (**7**,**8**) then 'graduate'

else 'other' end as educ\_female,

case

when b.Child\_GC in (**1**,**2**,**3**) then **1**

when b.Child\_GC in (**4**,**5**,**6**) then **2**

when b.Child\_GC in (**7**) then **3**

else **0** end as child\_num,

case

when b.Male\_Occ in (**1**,**2**,**3**) then 'white\_high'

when b.Male\_Occ in (**4**,**5**,**6**) then 'white\_low'

when b.Male\_Occ in (**7**,**8**,**9**) then 'blue'

when b.Male\_Occ in (**10**,**11**,**12**) then 'no\_occup'

else 'other' end as occu\_male,

case

when b.Female\_Occ in (**1**,**2**,**3**) then 'white\_high'

when b.Female\_Occ in (**4**,**5**,**6**) then 'white\_low'

when b.Female\_Occ in (**7**,**8**,**9**) then 'blue'

when b.Female\_Occ in (**10**,**11**,**12**) then 'no\_occup'

else 'other' end as occu\_female,

Cats+Dogs as pets\_total

from adsdemo b

order by b.panid;

**quit**;

**proc** **print** data = cust\_desc (obs = **10**); **run**;

/\* Variables for Demographic Study \*/

**data** cust\_desc2;

set cust\_desc;

IF fam\_size='large' THEN fam\_size\_L=**1** ; ELSE fam\_size\_L=**0**;

IF fam\_size='regular' THEN fam\_size\_R=**1** ; ELSE fam\_size\_R=**0**;

IF fam\_size='other' THEN fam\_size\_O=**1** ; ELSE fam\_size\_O=**0**;

IF fam\_income="low" THEN fam\_income\_L=**1** ; ELSE fam\_income\_L=**0**;

IF fam\_income="medium" THEN fam\_income\_M=**1** ; ELSE fam\_income\_M=**0**;

IF fam\_income="high" THEN fam\_income\_H=**1** ; ELSE fam\_income\_H=**0**;

IF fam\_income="very\_high" THEN fam\_income\_VH=**1** ; ELSE fam\_income\_VH=**0**;

IF fam\_income="other" THEN fam\_income\_O=**1** ; ELSE fam\_income\_O=**0**;

IF age\_male="young" THEN age\_mY=**1** ; ELSE age\_mY=**0**;

IF age\_male="mid\_career" THEN age\_mM=**1** ; ELSE age\_mM=**0**;

IF age\_male="elder" THEN age\_mE=**1** ; ELSE age\_mE=**0**;

IF age\_male="other" THEN age\_mO=**1** ; ELSE age\_mO=**0**;

IF age\_female="young" THEN age\_fY=**1** ; ELSE age\_fY=**0**;

IF age\_female="mid\_career" THEN age\_fM=**1** ; ELSE age\_fM=**0**;

IF age\_female="elder" THEN age\_fE=**1** ; ELSE age\_fE=**0**;

IF age\_female="other" THEN age\_fO=**1** ; ELSE age\_fO=**0**;

IF educ\_male="school" THEN educ\_mS=**1** ; ELSE educ\_mS=**0**;

IF educ\_male="college" THEN educ\_mC=**1** ; ELSE educ\_mC=**0**;

IF educ\_male="other" THEN educ\_mO=**1** ; ELSE educ\_mO=**0**;

IF educ\_male="graduate" THEN educ\_mG=**1** ; ELSE educ\_mG=**0**;

IF educ\_female="school" THEN educ\_fS=**1** ; ELSE educ\_fS=**0**;

IF educ\_female="college" THEN educ\_fC=**1** ; ELSE educ\_fC=**0**;

IF educ\_female="graduate" THEN educ\_mG=**1** ; ELSE educ\_mG=**0**;

IF educ\_female="other" THEN educ\_fO=**1** ; ELSE educ\_fO=**0**;

IF occu\_male="white\_high" THEN occ\_mWH=**1**; ELSE occ\_mWH=**0**;

IF occu\_male="white\_low" THEN occ\_mWL=**1**; ELSE occ\_mWL=**0**;

IF occu\_male="blue" THEN occ\_mB=**1**; ELSE occ\_mB=**0**;

IF occu\_male="no\_occup" THEN occ\_mNO=**1**; ELSE occ\_mNO=**0**;

IF occu\_female="white\_high" THEN occ\_fWH=**1**; ELSE occ\_fWH=**0**;

IF occu\_female="white\_low" THEN occ\_fWL=**1**; ELSE occ\_fWL=**0**;

IF occu\_female="blue" THEN occ\_fB=**1**; ELSE occ\_fB=**0**;

IF occu\_female="no\_occup" THEN occ\_fNO=**1**; ELSE occ\_fNO=**0**;

IF child\_num=**1** THEN one\_child=**1**; ELSE one\_child=**0**;

IF child\_num=**2** THEN two\_child=**1**; ELSE two\_child=**0**;

IF child\_num=**3** THEN three\_child=**1**; ELSE three\_child=**0**;

IF child\_num=**0** THEN zero\_child=**1**; ELSE zero\_child=**0**;

**RUN**;

/\* ADDING DEMOGRAPHIC VARIABLES AFTER CLUSTERING \*/

**proc** **sql**;

create table rfm\_all\_demo as

select a.\*, b.\* from RFM a, cust\_desc2 b

where a.PANID = b.PANID;

**proc** **print** data = rfm\_all\_demo1(obs = **10**); **run**;

**data** rfm\_all\_demo1; set rfm\_all\_demo;

drop PANTYPE PreTaxInc FamSize HH\_Race Type\_Res\_Poss County HH\_Age HH\_Edu HH\_Occ

Male\_Age Male\_Edu Male\_Occ Male\_Whc Male\_Smoke Female\_Age Female\_Edu Female\_Occ Female\_Whc Female\_Smoke Dogs

Cats Child\_GC Marital Language TV\_used TV\_hooked Year HISP\_Flag HISP\_Cat Race2 Race3 MW Zipcode Fipscode Market\_zipcode IRI\_Geonum Ext\_fact fam\_size fam\_income age\_male age\_female educ\_male educ\_female occu\_male occu\_female ; **run**;

/\* ------ Data Preparation : Brand Choice Model ------- \*/

**proc** **print** data = brand\_details (obs = **10**); **run**;

**proc** **sql**;

create table choice\_model\_1 as

select a.\*, b.\* from brand\_details a join rfm\_all\_demo1 b

on a.PANID = b.PANID;

**data** choice\_model; set choice\_model\_1; if nmiss(of \_numeric\_) + cmiss(of \_character\_) > **0** then delete; **run**;

**proc** **print** data = choice\_model (obs = **10**); **run**;

**proc** **export** data=choice\_model

outfile='H:\Final Project\choice\_model.csv'

dbms=csv

replace;

**run**;

**proc** **print** data = choice\_model(obs = **10**); **run**;

/\* Subsetting the Data \*/

**proc** **sql**;

create table choice\_model\_1 as

select

a.\* FROM choice\_model a where a.cluster = **4**; **quit**;

**proc** **sql**;

create table choice\_model\_2 as

select

a.\* FROM choice\_model a where a.cluster = **5**; **quit**;

**proc** **sql**;

create table choice\_model\_3 as

select

a.\* FROM choice\_model a where a.cluster = **6**; **quit**;

**proc** **sql**;

create table choice\_model\_4 as

select

a.\* FROM choice\_model a where a.cluster = **7**; **quit**;

**proc** **sql**;

create table choice\_model\_5 as

select

a.\* FROM choice\_model a where a.cluster = **8**; **quit**;

/\* Survival \*/

**PROC** **IMPORT** OUT= WORK.S1

DATAFILE= "H:\Final Project\survival\_esh.csv"

DBMS=CSV REPLACE;

GETNAMES=YES;

DATAROW=**2**;

**RUN**;

**proc** **sql**;

create table s2 as

select

a.\*,

case

when a.L5 in ('ALL TIDE PRODUCTS' , 'TIDE', 'TIDE KICK' , 'TIDE SIMPLE PLEASURES') then 'TIDE'

when a.L5 in ('ALL' ,'ALL BABY CARE PLUS' , 'ALL SMALL & MIGHTY' , 'ALL SURF') then 'ALL'

when a.L5 in ('BABY SOFT PUREX' , 'PUREX', 'PUREX ADVANCED', 'PUREX BABY') then 'PUREX'

when a.L5 in ('WISK' , 'WISK POWER SCOOP') then 'WISK'

else 'OTHER' end as brand

from s1 a;

**quit**;

**proc** **sql**;

create table WORK.s4 as

select duration, dollars, Number\_of\_Cats, Panelist\_ID,Vol\_Eq, Number\_of\_Dogs, Units, Occupation\_Code\_of\_Male\_HH, Family\_Size,brand,

case when concentration\_level = 'ULTRA CONCENTRATED' then **1** else **0** end as cl\_uc,

case when concentration\_level = 'CLASSIC' then **1** else **0** end as cl\_c,

case when concentration\_level = 'MISSING' then **1** else **0** end as cl\_m,

case when concentration\_level = 'CONCENTRATED' then **1** else **0** end as cl\_c,

case when concentration\_level = 'REGULAR CONCENTRATE' then **1** else **0** end as cl\_rc,

case when Children\_Group\_Code = **1** then **1** else **0** end as cgc1,

case when Children\_Group\_Code = **2** then **1** else **0** end as cgc2,

case when Children\_Group\_Code = **3** then **1** else **0** end as cgc3,

case when Children\_Group\_Code = **4** then **1** else **0** end as cgc4,

case when Children\_Group\_Code = **5** then **1** else **0** end as cgc5,

case when Children\_Group\_Code = **6** then **1** else **0** end as cgc6,

case when Children\_Group\_Code = **7** then **1** else **0** end as cgc7,

case when Children\_Group\_Code = **8** then **1** else **0** end as cgc8,

case when Combined\_Pre\_Tax\_Income\_of\_HH in (**1**,**2**,**3**,**4**) then **1** else **0** end as inc\_low,

case when Combined\_Pre\_Tax\_Income\_of\_HH in (**5**,**6**,**7**,**8**) then **1** else **0** end as inc\_med,

case when Combined\_Pre\_Tax\_Income\_of\_HH in (**9**,**10**,**11**,**12**) then **1** else **0** end as inc\_high,

case when Education\_Level\_Reached\_by\_Femal in (**1**,**2**,**3**) then **1** else **0** end as edu\_f\_low,

case when Education\_Level\_Reached\_by\_Femal in (**4**,**5**,**6**) then **1** else **0** end as edu\_f\_med,

case when Education\_Level\_Reached\_by\_Femal in (**7**,**8**) then **1** else **0** end as edu\_f\_high,

case when Education\_Level\_Reached\_by\_Male in (**1**,**2**,**3**) then **1** else **0** end as edu\_m\_low,

case when Education\_Level\_Reached\_by\_Male in (**4**,**5**,**6**) then **1** else **0** end as edu\_m\_med,

case when Education\_Level\_Reached\_by\_Male in (**7**,**8**) then **1** else **0** end as edu\_m\_high,

case when Hh\_Race = **3** then **1** else **0** end as hisp\_flag,

case when Marital\_Status = **2** then **1** else **0** end as married,

case when Outlet = 'GR' then **1** else **0** end as outlet\_gr,

case when Outlet = 'DR' then **1** else **0** end as outlet\_dr,

case when Outlet = 'KK' or Outlet = 'MA' then **1** else **0** end as outlet\_oth,

case when Type\_of\_Residential\_Possession = **1** then **1** else **0** end as resi\_rent,

case when Type\_of\_Residential\_Possession = **2** then **1** else **0** end as resi\_own

from WORK.s2

where Combined\_Pre\_Tax\_Income\_of\_HH > **0**

and Education\_Level\_Reached\_by\_Femal not in (**0**,**9**)

and Education\_Level\_Reached\_by\_Male not in (**0**,**9**)

and Marital\_Status > **0**

and Type\_of\_Residential\_Possession > **0**;

**quit**;

**proc** **print** data = s4 (obs = **10**);

**run**;

**proc** **contents** data = s4; **run**;

/\* Creating data sets for each clusters \*/

**proc** **sql**;

create table survival\_1 as

select

a.\*, b.cluster, b.monetary, b.recency, b.frequency, b.loyality\_all, b.loyality\_tide FROM s4 a JOIN choice\_model\_1 b on s4.Panelist\_ID = choice\_model\_1.panid; **quit**;

**proc** **sql**;

create table survival\_2 as

select

a.\*, b.cluster, b.monetary, b.recency, b.frequency, b.loyality\_all, b.loyality\_tide FROM s4 a JOIN choice\_model\_2 b on s4.Panelist\_ID = choice\_model\_2.panid; **quit**;

**proc** **sql**;

create table survival\_3 as

select

a.\*, b.cluster, b.monetary, b.recency, b.frequency, b.loyality\_all, b.loyality\_tide FROM s4 a JOIN choice\_model\_3 b on s4.Panelist\_ID = choice\_model\_3.panid; **quit**;

**proc** **sql**;

create table survival\_4 as

select

a.\*, b.cluster, b.monetary, b.recency, b.frequency, b.loyality\_all, b.loyality\_tide FROM s4 a JOIN choice\_model\_4 b on s4.Panelist\_ID = choice\_model\_4.panid; **quit**;

**proc** **sql**;

create table survival\_5 as

select

a.\*, b.cluster, b.monetary, b.recency, b.frequency, b.loyality\_all, b.loyality\_tide FROM s4 a JOIN choice\_model\_5 b on s4.Panelist\_ID = choice\_model\_5.panid; **quit**;

**proc** **sql**;

create table survival\_0 as

select

a.\*, b.cluster, b.monetary, b.recency, b.frequency, b.loyality\_all, b.loyality\_tide FROM s4 a JOIN choice\_model b on s4.Panelist\_ID = choice\_model.panid; **quit**;

/\* Survival Analysis \*/

**PROC** **LIFEREG** data=survival\_1;

model duration = Dollars Number\_of\_Cats Vol\_Eq Number\_of\_Dogs Units Family\_Size cl\_uc cl\_c cl\_m cgc2 cgc3 cgc4 cgc5 cgc6 cgc7 cgc8 inc\_med inc\_high edu\_f\_med edu\_f\_high edu\_m\_med edu\_m\_high hisp\_flag married outlet\_gr outlet\_oth resi\_own

monetary recency frequency loyality\_all loyality\_tide / distribution = weibull;

**run**;

**PROC** **LIFEREG** data=survival\_2;

model duration = Dollars Number\_of\_Cats Vol\_Eq Number\_of\_Dogs Units Family\_Size cl\_uc cl\_c cl\_m cgc2 cgc3 cgc4 cgc5 cgc6 cgc7 cgc8 inc\_med inc\_high edu\_f\_med edu\_f\_high edu\_m\_med edu\_m\_high hisp\_flag married outlet\_gr outlet\_oth resi\_own

monetary recency frequency loyality\_all loyality\_tide / distribution = weibull;

**run**;

**PROC** **LIFEREG** data=survival\_3;

model duration = Dollars Number\_of\_Cats Vol\_Eq Number\_of\_Dogs Units Family\_Size cl\_uc cl\_c cl\_m cgc2 cgc3 cgc4 cgc5 cgc6 cgc7 cgc8 inc\_med inc\_high edu\_f\_med edu\_f\_high edu\_m\_med edu\_m\_high hisp\_flag married outlet\_gr outlet\_oth resi\_own

monetary recency frequency loyality\_all loyality\_tide / distribution = weibull;

**run**;

**PROC** **LIFEREG** data=survival\_4;

model duration = Dollars Number\_of\_Cats Vol\_Eq Number\_of\_Dogs Units Family\_Size cl\_uc cl\_c cl\_m cgc2 cgc3 cgc4 cgc5 cgc6 cgc7 cgc8 inc\_med inc\_high edu\_f\_med edu\_f\_high edu\_m\_med edu\_m\_high hisp\_flag married outlet\_gr outlet\_oth resi\_own

monetary recency frequency loyality\_all loyality\_tide / distribution = weibull;

**run**;

**PROC** **LIFEREG** data=survival\_5;

model duration = Dollars Number\_of\_Cats Vol\_Eq Number\_of\_Dogs Units Family\_Size cl\_uc cl\_c cl\_m cgc2 cgc3 cgc4 cgc5 cgc6 cgc7 cgc8 inc\_med inc\_high edu\_f\_med edu\_f\_high edu\_m\_med edu\_m\_high hisp\_flag married outlet\_gr outlet\_oth resi\_own

monetary recency frequency loyality\_all loyality\_tide / distribution = weibull;

**run**;

**PROC** **LIFEREG** data=survival\_0;

model duration = Dollars Number\_of\_Cats Vol\_Eq Number\_of\_Dogs Units Family\_Size cl\_uc cl\_c cl\_m cgc2 cgc3 cgc4 cgc5 cgc6 cgc7 cgc8 inc\_med inc\_high edu\_f\_med edu\_f\_high edu\_m\_med edu\_m\_high hisp\_flag married outlet\_gr outlet\_oth resi\_own

monetary recency frequency loyality\_all loyality\_tide / distribution = weibull;

**run**;