Advance Business Analytics and Visualization

ASIA PACIFIC UNIVERSITY OF TECHNOLOGY & INNOVATION (APU)
SCHOOL OF COMPUTING AND TECHNOLOGY

NAME: Ameer Ali khoso

STUDENT ID: TP043357

INTAKE: UCMF1610DSBA

Table of Contents

1. Introduction

The organic food market has become one of the most rapidly growing sectors in developed economies around the world, especially in the European Union (Chen, 2007). In 2010, this market reached 18.1 billion euros in sales, compared to 10.0 billion euros in 2004. The growth in demand for organic food in the last two decades is partly due to food scandals that have heightened consumer awareness about natural, healthy, safe, and quality food (Miles and Frewer, 2001; Onyango et al., 2007; Schmid et al., 2007; Kalogeras et al., 2009). However this study is about to predict the purchasing organic food that which age group and city has to more purchase the organic food. Furthermore, the study has to predict the Southwell and Scottish city.

2. Problem Statement

Statistics have shown that there is decreased in percentage of organic food market from year to year. Therefore it is extremely vital to use predictive models to determine the which cities and age groups will bought more organic food in future.

3. Aim

The aim of this study is to investigate the potential of data mining techniques as a tool for predicting the organic food market

4. Objectives

- The objective of this study is to predict the organic food in Southwell and Scottish.
- To generate graphs using several attributes for effective prediction using SAS visual analytics.

5. METHODOLOGY

5.1 CRISP-DM

The Cross Industry Standard Process for Data-Mining – CRISP-DM is a process which follows a structured method when planning a data mining project that is used to solve problems. The 6 phases are described below

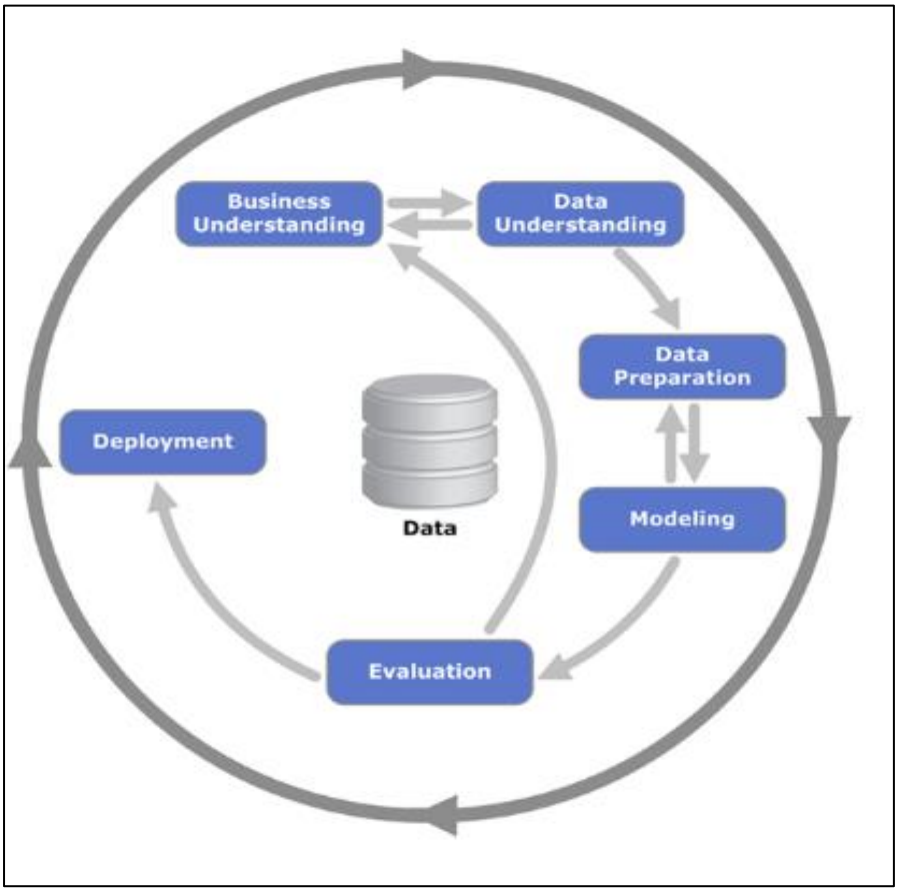


Figure 1: CRISP-DM Lifecycle

Source: Villena, 2016

5.1.1 Business Understanding – to understand the objectives of the project.

5.1.2 Understanding Data – to collect and describe data.

5.1.3 Data Preparation – activities needed to construct the final dataset

5.1.4 Modelling – to select the modelling technique to be used.

5.1.5 Evaluation – to evaluate the process to see if the technique solves the problem of modelling and creation of rules.

5.1.6 Deployment – Applying the model in real time environment.

Business Understanding

The first phase of the CRISP-DM is the Business Understanding, where an understanding of the objectives is aimed at. For the sake of this study, the proposed objective is to predict the organic food of two cities

Data Understanding

The second phase of the CRISP-DM is the Data Understanding. The initial data should be collected where for this study the existing dataset from SAS was used and a description of this was produced. This is where the organic food bought history of the dataset is synthesized, with the required attributes such as organic purchase age group gender geographic loyalty card cities and so on.

Data Preparation

This phase involves the activities needed to construct the final dataset. In this phase the relevant attributes that will be used for the modelling and analysis of the project was done. The selection of the relevant attributes selected included the driver gender as the response variable and cities as the predictor variable.

Modelling

This phase is where the relevant modelling technique on the data prepared in the Data preparation phase is selected. The selected modelling technique used for this study is the decision tree and logistic regression.

Evaluation

In this phase a checking procedure is performed to assess whether we have use the best tool for data mining and verifies that the data is really portraying the reality understood in the Business Understanding phase. If more processes are to be modelled, the process returns to the Business Understanding phase and reiterates the whole process.

Deployment

In this phase the implementation of the model is done, where the results of the model are either presented or recorded.

6 DATA ANALYTIC TECHNIQUES

6.1 Decision Tree

A decision tree is a classification tree structure that consists of a root node, branches and leaf nodes that utilizes a divide and conquer algorithm. Each internal node represents the tests done, while the branches represent the outcome of the test and the leaf nodes represent class labels. The highest node in the tree is known as the root node. A classification tree is a tree-like graph that represents decisions and their possible results. For each inner node, a test is carried out on attribute or input variable. The branches that connect the nodes lead to leaf nodes that represent test results. Classification trees are applied when response variables are quantitative, discrete and qualitative. The main advantage of using decision trees is that it is comprehendible and explainable. However, the downside to it is that a general design cannot be utilized from one context to the other (Mines.humanoriented.com, 2013).

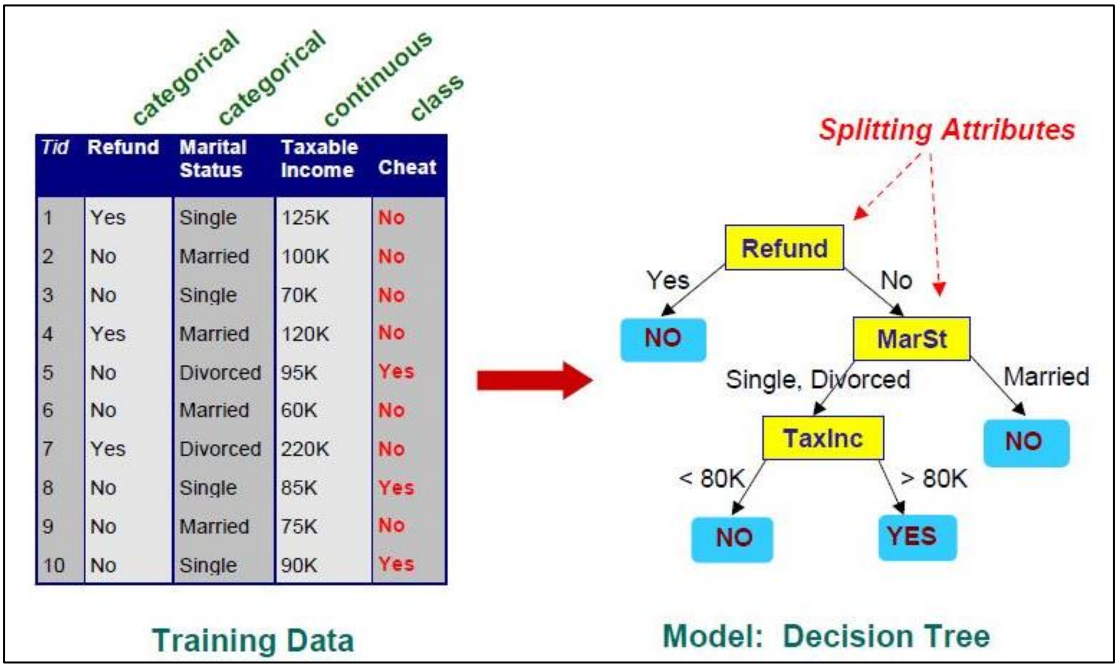
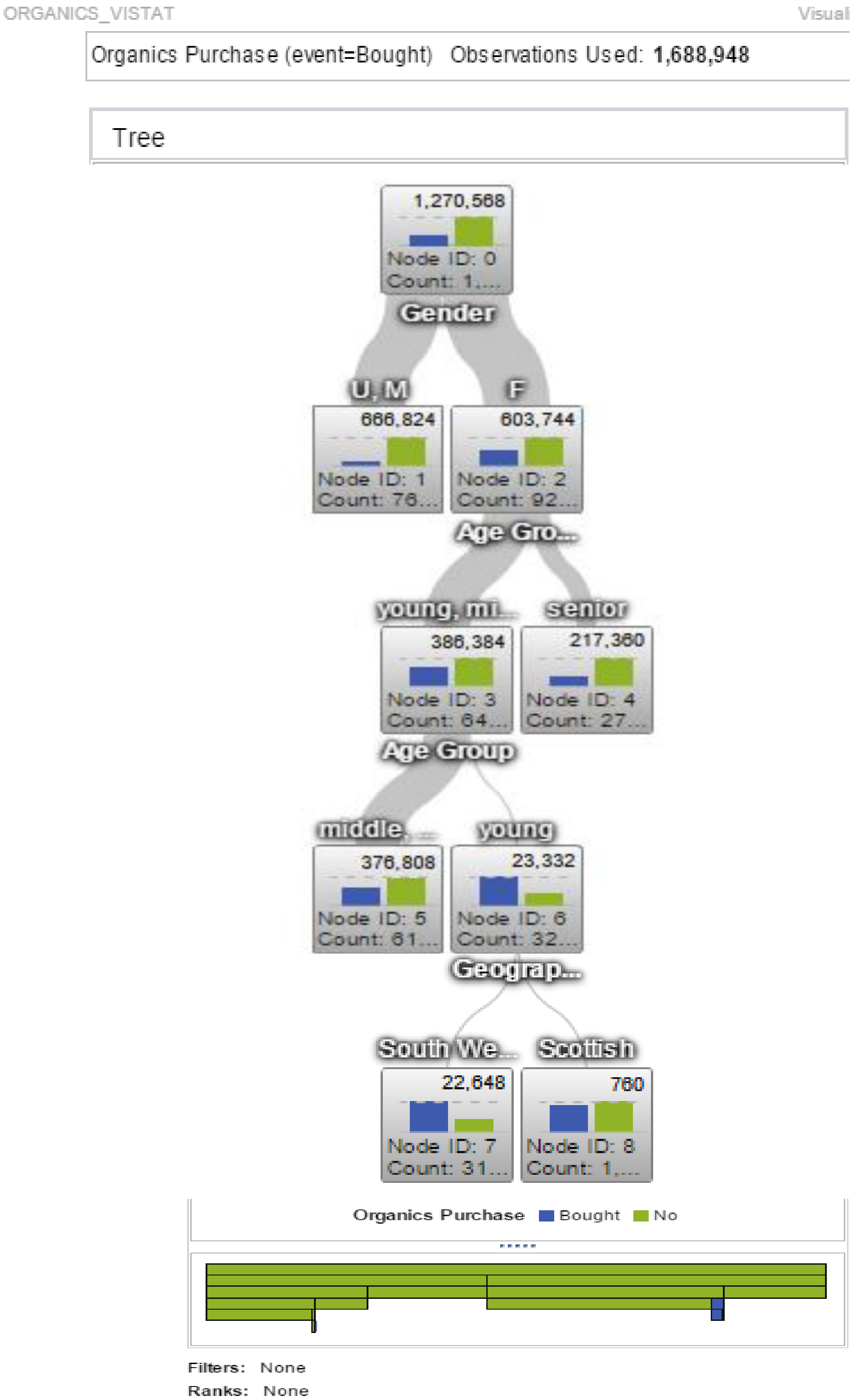


Figure 2: Example of a Decision Tree

Source: Mines.humanoriented.com, 2013

6.1.1 Application of Decision Tree

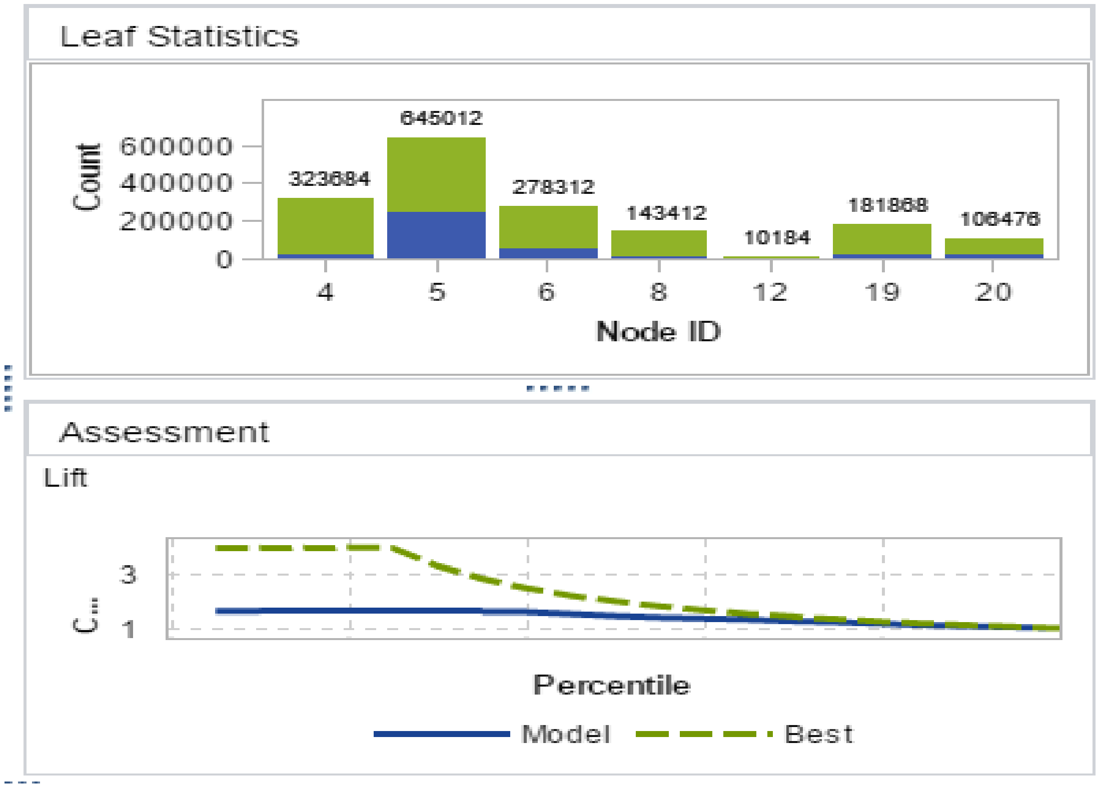
The figure below shows the output results of the decision tree which was implemented using SAS visual analytics. The decision tree was used for the prediction of the age group, which more bought the organic food in south wells and Scottish.



The above results shows the prediction of bought organic products and did not purchase the products in two cities known as southwells and Scottish. The blue nodes represents the bought products and green shows did not purchase the organic products. On the disregarded node it predicts a higher number of organic products purchased. However the female are younger to age are more likely to purchase organic products. Meanwhile in geographical location, southwells city has more number of young female who have to purchase products with the number of 22,648, followed by Scottish city which has higher number of female which do not buy the organic products with number of 760.

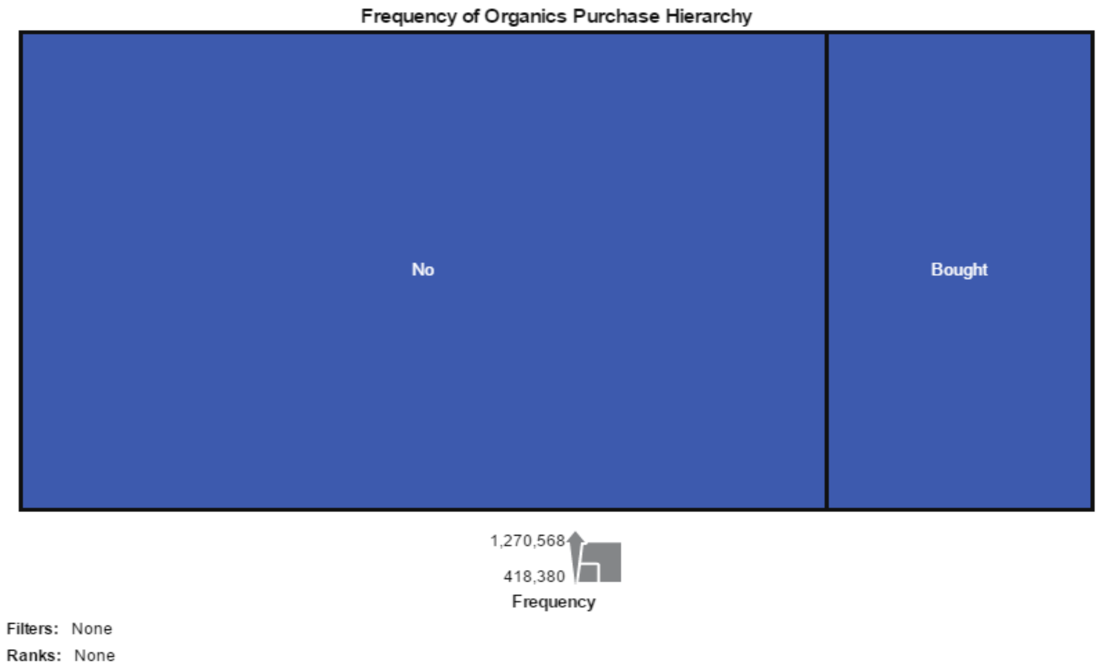
6.1.2 STATISTICAL GRAPH

6.1.2.1 Leaf Statistics and Assessment



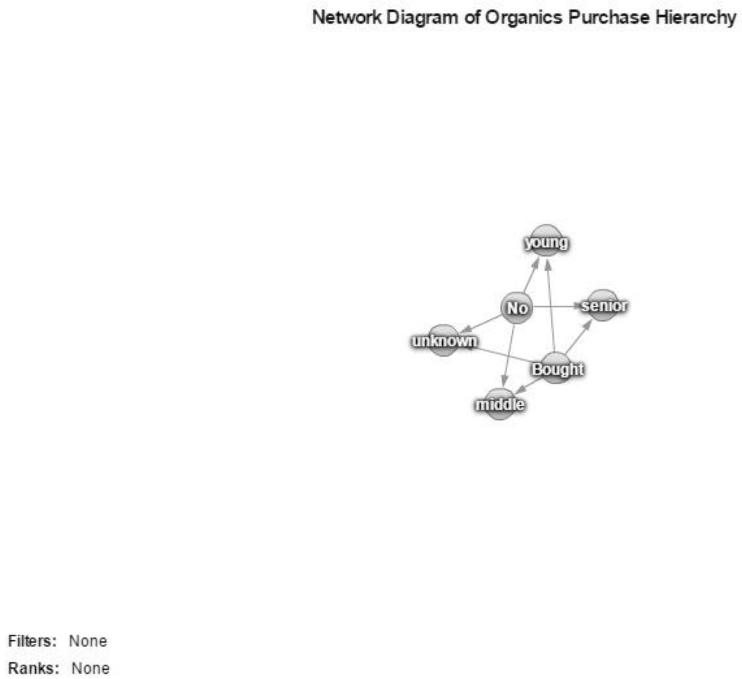
The above results show the statistics of organic products. It shows that Node ID 5 has higher number of people which buying the products followed by Node ID 4 with count of 645012 and 323684 respectively. However the Node 12 has lower number of people to buy the products.

6.1.2.2 Fequency of Organic Purchase Hierarchy



The above graph shows the frequency of purchase hierarachy. It tells the 1,270,568 observation have not purchased the products followed by bought with count 418,380 which is more slightly lower.

6.1.2.3 Network Diagram of organic purchase Hierarchy



6.2 Logistic Regression

Logistic regression is a statistical method for analysing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes).

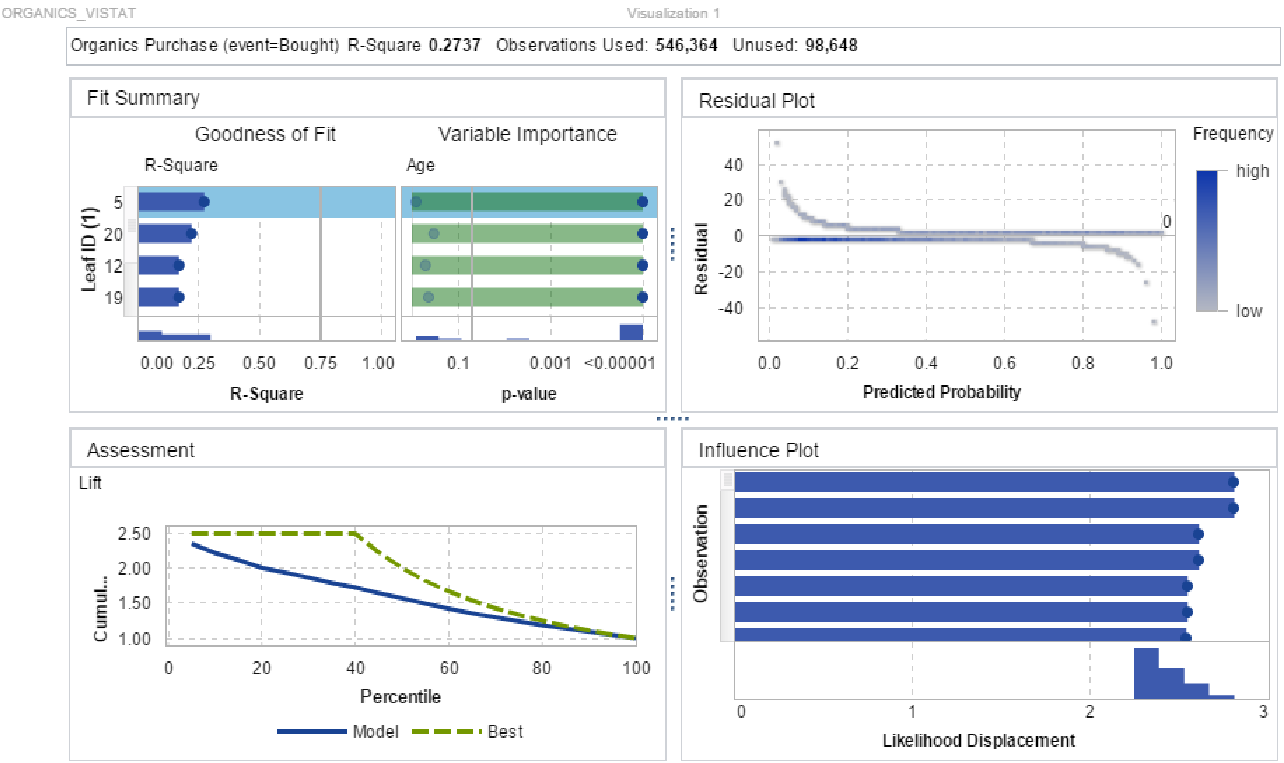
In logistic regression, the dependent variable is binary or dichotomous, i.e. it only contains data coded as 1 (TRUE, success, pregnant, etc.) or 0 (FALSE, failure, non-pregnant, etc.).

The goal of logistic regression is to find the best fitting (yet biologically reasonable) model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression generates the coefficients (and its standard errors and significance levels) of a formula to predict a logit transformation of the probability of presence of the characteristic of interest:

$$logit(p) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \ldots + b_k X_k$$

6.2.1 Application of Logistic Regression

The figure below shows the output results of the logistic regression which is implemented using SAS visual analytics. The logistic regression has used for the prediction of the age group, which more bought the organic food.



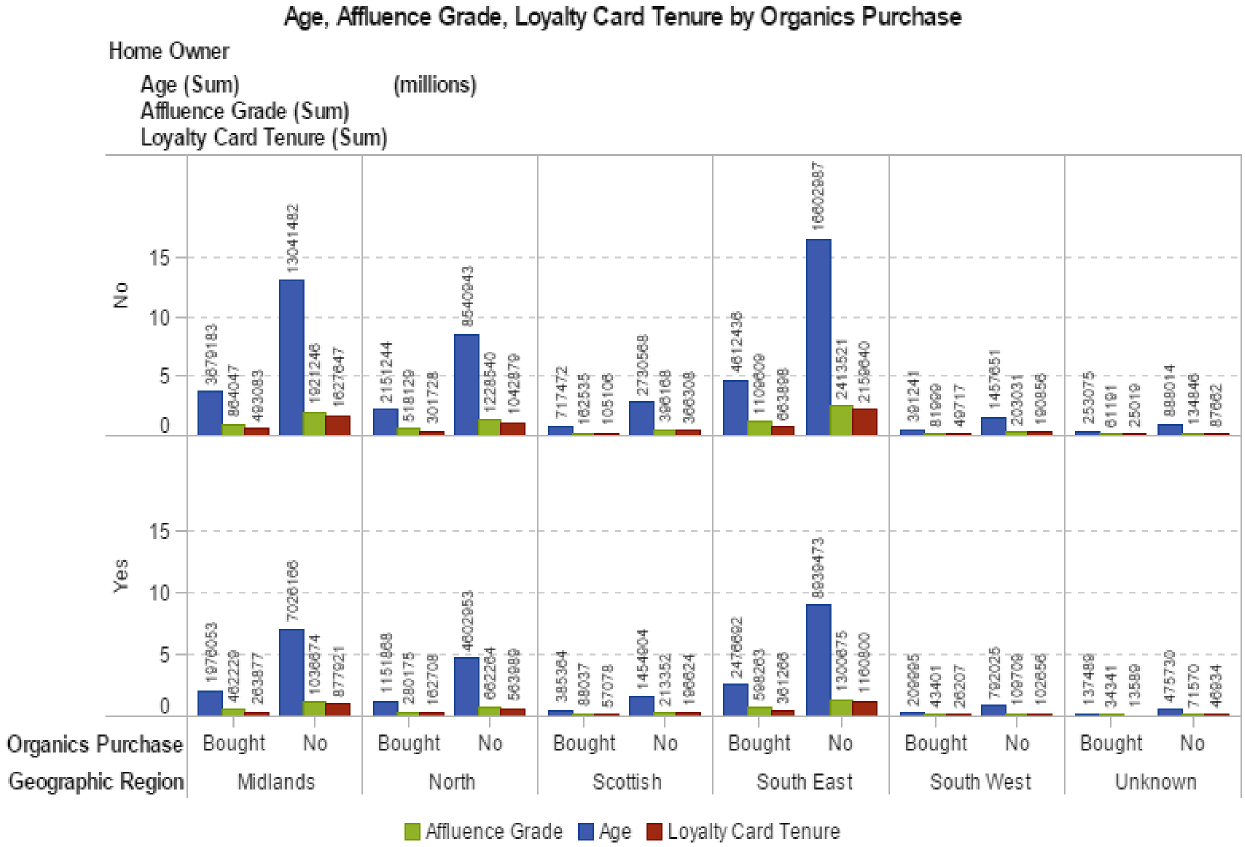
Filters: None

Ranks: None

The above logistic regression graph shows the predictable result with age group we have built 4 simultaneously model. Each of the nodes represents the model which creates in decision tree. However age did play very important role in all 3 models which shown in logistic graph but they did not play important role in model 4. Meanwhile the p-vale of age variable is less then level of significance thus we reject the null hypothesis which means age play important roles in above model.

STATISTICAL GRAPH

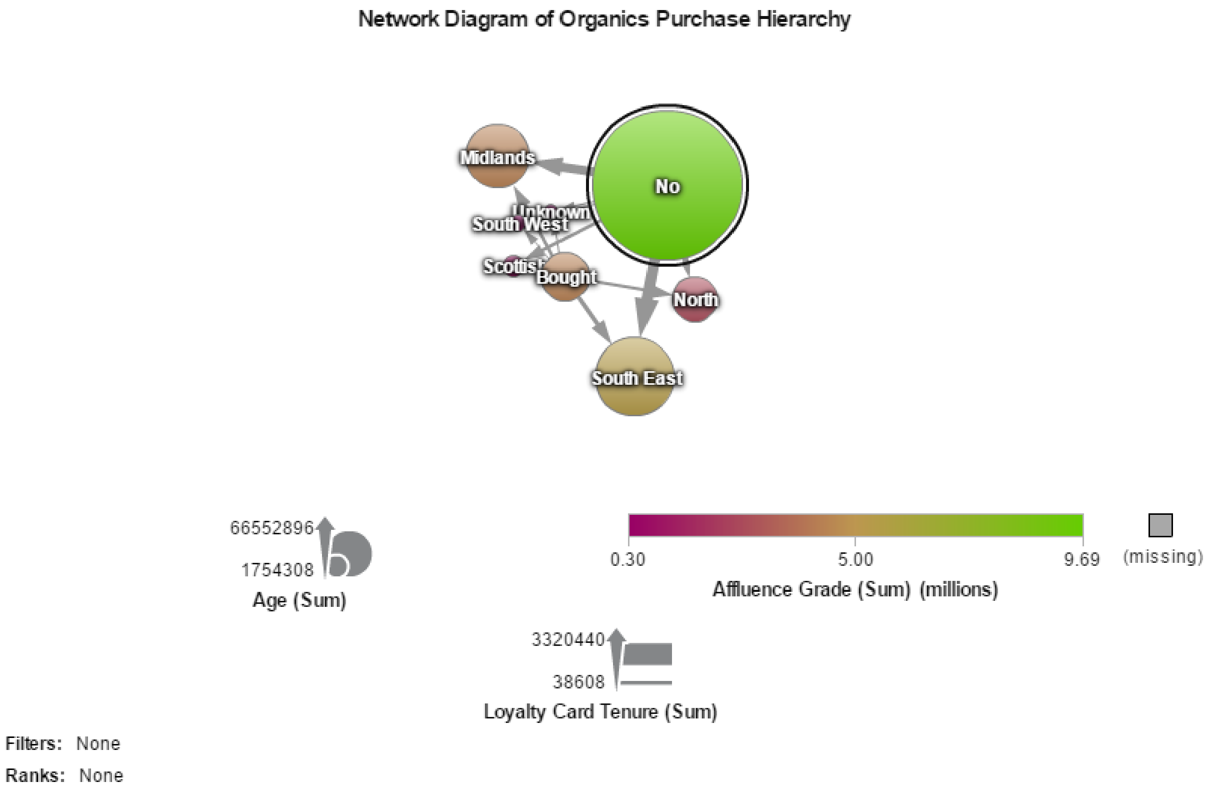
6.2.2.1 Loyalty Card Tenure by organic purchase



Filters: None

Ranks: None

6.2.2.2 Network Diagram



7 Conclusion:

In this study, the two (2) techniques has been used such as decision tree and logistic regression in SAS visual analytics software. The organic product data set used to predict the purchase behaviour of 2 cities. Like Southwell and Scottish. The result shown that young females are likely to buy organic products in both cities. However we have used open data sets which is available in SAS visual analytics software.

8 References

* Villena, J. (2016). *s|ngular - CRISP-DM: The methodology to put some order into Data Science projects*. [online] Data.sngular.team. Available at: https://data.sngular.team/en/art/40/crisp-dm-the-methodology-to-put-some-order-into-data-science-projects.

* Moro, S., Laureano, R. and Cortez, P., 2011. Using data mining for bank direct marketing: An application of the crisp-dm methodology. In*Proceedings of European Simulation and Modelling Conference-ESM'2011*(pp. 117-121). Eurosis.

* Mines.humanoriented.com. (2013). *Decision Tree Classifier*. [online] Available at: http://mines.humanoriented.com/classes/2010/fall/csci568/portfolio_exports/lguo/decisionTree.html.

* Smart Vision - Europe. (2016). *What is the CRISP-DM methodology?*. [online] Available at: http://www.sv-europe.com/crisp-dm-methodology/#businessunderstanding

APPENDIX

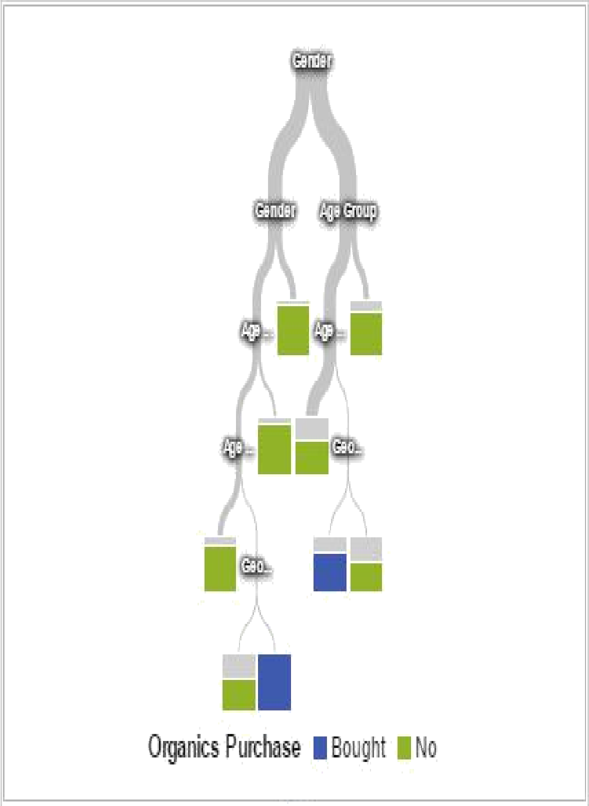
Exploration 1

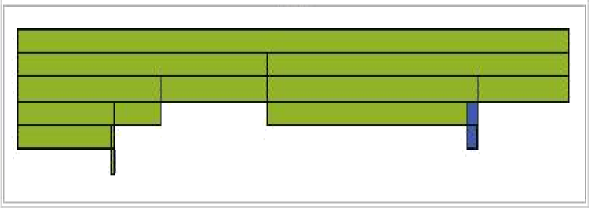
Creation date: Sunday, June 18, 2017

Author: Ameer Ali Khoso

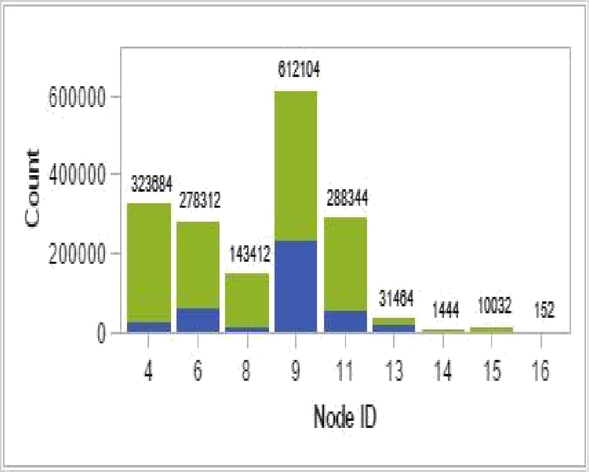
Organics Purchase (event=Bought) Observations Used: **1,688,948**

Tree



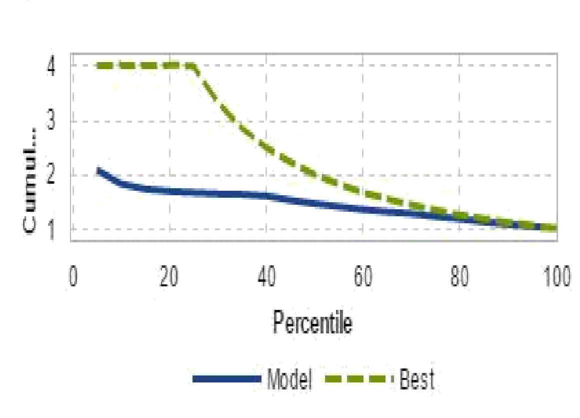


Leaf Statistics



Assessment

Lift



Filters: None

Ranks: None

Node ID	Depth	Parent ID	N Children	Type	Observations
0	0	-1	2	Class	1688948
1	1	0	2	Class	765624
2	1	0	2	Class	923324
3	2	1	2	Class	441940
4	2	1	0	Leaf	323684
5	2	2	2	Class	645012
6	2	2	0	Leaf	278312
7	3	3	2	Class	298528
8	3	3	0	Leaf	143412
9	3	5	0	Leaf	612104
10	3	5	2	Class	32908
11	4	7	0	Leaf	288344
12	4	7	2	Class	10184
13	4	10	0	Leaf	31464
14	4	10	0	Leaf	1444
15	5	12	0	Leaf	10032
16	5	12	0	Leaf	152

% Observations	Percent of Parent	Gain	Predicted Value	Bought	No
100.00%	.	0.0474	No	418380 (24.77%)	1270568 (75.23%)
45.33%	45.33%	0.0127	No	98800 (12.90%)	666824 (87.10%)
54.67%	54.67%	0.0233	No	319580 (34.61%)	603744 (65.39%)
26.17%	57.72%	0.0109	No	73416 (16.61%)	368524 (83.39%)
19.16%	42.28%	0.0065	No	25384 (7.84%)	298300 (92.16%)
38.19%	69.86%	0.0151	No	258628 (40.10%)	386384 (59.90%)
16.48%	30.14%	0.0002	No	60952 (21.90%)	217360 (78.10%)
17.68%	67.55%	0.0078	No	58748 (19.68%)	239780 (80.32%)
8.49%	32.45%	0.0020	No	14668 (10.23%)	128744 (89.77%)
36.24%	94.90%	0.0054	No	235296 (38.44%)	376808 (61.56%)
1.95%	5.10%	0.0080	Bought	23332 (70.90%)	9576 (29.10%)
17.07%	96.59%	0.0038	No	54264 (18.82%)	234080 (81.18%)
0.60%	3.41%	0.0179	No	4484 (44.03%)	5700 (55.97%)
1.86%	95.61%	0.0027	Bought	22648 (71.98%)	8816 (28.02%)
0.09%	4.39%	0.0506	No	684 (47.37%)	760 (52.63%)
0.59%	98.51%	0.0125	No	4332 (43.18%)	5700 (56.82%)
0.01%	1.49%	0.0000	Bought	152 (100.00%)	

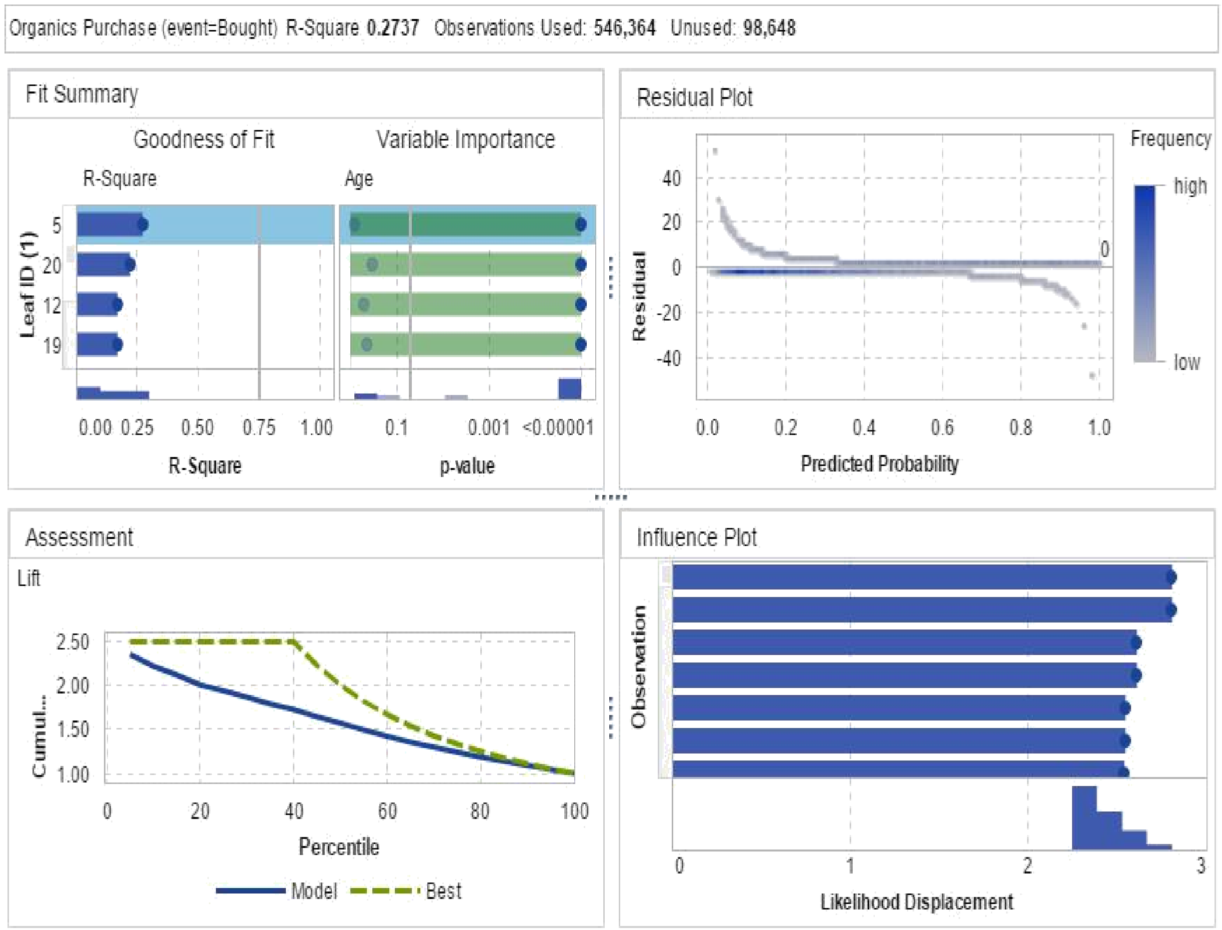
% Observations	Percent of Parent	Gain	Predicted Value	Bought	No
100.00%	.	0.0474	No	418380 (24.77%)	1270568 (75.23%)
45.33%	45.33%	0.0127	No	98800 (12.90%)	666824 (87.10%)
54.67%	54.67%	0.0233	No	319580 (34.61%)	603744 (65.39%)
26.17%	57.72%	0.0109	No	73416 (16.61%)	368524 (83.39%)
19.16%	42.28%	0.0065	No	25384 (7.84%)	298300 (92.16%)
38.19%	69.86%	0.0151	No	258628 (40.10%)	386384 (59.90%)
16.48%	30.14%	0.0002	No	60952 (21.90%)	217360 (78.10%)
17.68%	67.55%	0.0078	No	58748 (19.68%)	239780 (80.32%)
8.49%	32.45%	0.0020	No	14668 (10.23%)	128744 (89.77%)
36.24%	94.90%	0.0054	No	235296 (38.44%)	376808 (61.56%)
1.95%	5.10%	0.0080	Bought	23332 (70.90%)	9576 (29.10%)
17.07%	96.59%	0.0038	No	54264 (18.82%)	234080 (81.18%)
0.60%	3.41%	0.0179	No	4484 (44.03%)	5700 (55.97%)
1.86%	95.61%	0.0027	Bought	22648 (71.98%)	8816 (28.02%)
0.09%	4.39%	0.0506	No	684 (47.37%)	760 (52.63%)
0.59%	98.51%	0.0125	No	4332 (43.18%)	5700 (56.82%)
0.01%	1.49%	0.0000	Bought	152 (100.00%)	

logistic regression

Creation date: Sunday, June 18, 2017

Author: Ameer Ali Khoso

File: /User Folders/tp043357@mail.apu.edu.my/My Folder/decission tree



Filters: None

Ranks: None

Description	Value
Number of Model Effects	6
Number of Classification Effects	2
Number of Columns in X	12
Rank of Cross-product Matrix	10
Number of Observations Read	645,012
Number of Observations Used	546,364

ORGANICS_VISTAT

Visualization 1: Fit Statistics

Statistic	Value
-2 Log Likelihood	561501.9
AIC	561521.9
AICC	561521.9
BIC	561634
R-Square	0.273716
Max-rescaled R-Square	0.369828

ORGANICS_VISTAT

Visualization 1: Parameter Estimates

Parameter		Estimate	Standard Error	z Value	Pr > |z|
Intercept		1.214445	0.030168	40.25633	<0.0001
Age		-0.09742	0.000387	-251.528	<0.0001
Affluence Grade		0.276644	0.001081	255.8889	<0.0001
Loyalty Card Tenure		-0.00442	0.00098	-4.51348	<0.0001
Geographic Region	Midlands	0.247732	0.02442	10.14442	<0.0001
Geographic Region	North	0.079773	0.024812	3.215119	0.0013
Geographic Region	Scottish	0.23956	0.027263	8.787103	<0.0001
Geographic Region	South East	0.171983	0.024257	7.089979	<0.0001
Geographic Region	South West	0.427103	0.031866	13.40303	<0.0001
Geographic Region	Unknown	0	.	.	.
Home Owner	No	-0.00165	0.006865	-0.24085	0.8097
Home Owner	Yes	0	.	.	.

Code

```
options VALIDMEMNAME=EXTEND VALIDVARNAME=ANY;

   /**** User ID: tp043357@mail.apu.edu.my ****/

   /**** Model Name: Visualization 1 ****/
   /*------------------------------------------
   SAS Code Generated by LASR Analytic Server
     Date               : 17Jun2017:18:43:36
     Locale             : en_US
     Model Type         : Logistic Regression
     Group-By variable : _va_800_C1000000
     Class variable     : DemReg
     Class variable     : DemHomeowner
     Response variable : _va_response_2
     Distribution       : Binary
     Link Function      : Logit
     ------------------------------------------*/

   /*-------------------------------------------*/
   /*Temporary computed columns                */
if
(('DemGender'n = 'M') AND ('DemAgeGroup'n IN ('middle','unknown'))
AND ('PromClass'n IN ('Silver','Gold','Platinum')))
then
'C1000000'n
=
19.0;
else
if
(('DemGender'n = 'M') AND ('DemAgeGroup'n IN ('middle','unknown'))
AND ('PromClass'n = 'Tin'))
then
'C1000000'n
=
20.0;
else
if
(('DemGender'n = 'M') AND ('DemAgeGroup'n = 'young'))
then
'C1000000'n
=
12.0;
else
if
(('DemGender'n = 'M') AND ('DemAgeGroup'n = 'senior'))
then
'C1000000'n
=
8.0;
else
if
('DemGender'n = 'U')
```

```
then
'C1000000'n
=
4.0;
else
if
(('DemGender'n = 'F') AND ('DemAgeGroup'n IN
('young','middle','unknown')))
then
'C1000000'n
=
5.0;
else
if
(('DemGender'n = 'F') AND ('DemAgeGroup'n = 'senior'))
then
'C1000000'n
=
6.0;
else
'C1000000'n
=
.;
length
_va_response_2
$10;
'_va_response_2'n
=
'TargetBuy'n;
'_va_800_C1000000'n
=
round('C1000000'n,0.01);

   ;
   /*------------------------------------------*/

   /*------------------------------------------*/
   /*Defining temporary arrays and variables   */
   array _xrow_2_{12} _temporary_;
   array _beta_2_{12} _temporary_;

   drop _badval_ _linp_ _temp_ _i_;
   _badval_ = 0;
   _linp_   = 0;
   _temp_   = 0;
   _i_      = 0;

   /*Formatted values of GROUPBY variables*/
   length __va_800_C1000000_ $32; drop __va_800_C1000000_;
   __va_800_C1000000_ = left(trim(put(_va_800_C1000000,BEST32.)));

   /*Formatted values of CLASS variables*/
   length _DemReg_ $10; drop _DemReg_;
   _DemReg_ = left(trim(put(DemReg,$10.)));
```

```
length _DemHomeowner_ $3; drop _DemHomeowner_;
_DemHomeowner_ = left(trim(put(DemHomeowner,$3.)));
length __va_response_2_ $10; drop __va_response_2_;
__va_response_2_ = left(trim(put(_va_response_2,$10.)));
/*------------------------------------------*/

/*------------------------------------------*/
/*Predicted values are computed for records */
/*that match one of the response values used*/
/*in modeling the data.                     */
select (__va_response_2_);
   when ('Bought') _badval_ = 0;
   when ('No') _badval_ = 0;
   otherwise do; _badval_ = 1; goto skip_2_0; end;
end;
/*------------------------------------------*/

/*------------------------------------------*/
/*Missing values in model variables result  */
/*in missing values for the prediction.     */
if missing(DemAge)
   or missing(DemAffl)
   or missing(PromTime)
   then do;
      _badval_ = 1;
      goto skip_2_0;
end;
/*------------------------------------------*/

select (__va_800_C1000000_);
   when ('4') do;
      /*Initialize temporary array for x row*/
      do _i_=1 to 12; _xrow_2_{_i_} = 0; end;

      /*------------------------------------------*/
      /*Fill in the X row for the observation     */
      /*based on converting the values of model   */
      /*variables to effect positions.            */
      /*Effect: Intercept*/
      _xrow_2_[1] = 1;
      /*Effect: Age*/
      _xrow_2_[2] = DemAge;
      /*Effect: 'Affluence Grade'n*/
      _xrow_2_[3] = DemAffl;
      /*Effect: 'Loyalty Card Tenure'n*/
      _xrow_2_[4] = PromTime;
      /*Effect: 'Geographic Region'n*/
      _temp_ = 1;
      select (_DemReg_);
         when ('Midlands') _xrow_2_[5] = _temp_;
         when ('North') _xrow_2_[6] = _temp_;
         when ('Scottish') _xrow_2_[7] = _temp_;
         when ('South East') _xrow_2_[8] = _temp_;
         when ('South West') _xrow_2_[9] = _temp_;
```

```
      when ('Unknown') _xrow_2_[10] = _temp_;
      otherwise do; _badval_ = 1; goto skip_2_0; end;
   end;

   /*Effect: 'Home Owner'n*/
   _temp_ = 1;
   select (_DemHomeowner_);
      when ('No') _xrow_2_[11] = _temp_;
      when ('Yes') _xrow_2_[12] = _temp_;
      otherwise do; _badval_ = 1; goto skip_2_0; end;
   end;

   /*------------------------------------------*/

   /*Initialize beta*/
   _beta_2_{1} = -2.45271414260925;
   _beta_2_{2} = -0.05262666446671;
   _beta_2_{3} = 0.29045969217497;
   _beta_2_{4} = 0.02743604695601;
   _beta_2_{5} = -0.12574286388897;
   _beta_2_{6} = -0.22390064536448;
   _beta_2_{7} = -0.40010549897153;
   _beta_2_{8} = -0.23707804152471;
   _beta_2_{9} = 0.05481505185494;
   _beta_2_{10} = 0;
   _beta_2_{11} = -0.00557002770619;
   _beta_2_{12} = 0;


   /*------------------------------------------*/
   /*Compute the linear predictor              */
   do _i_=1 to 12;
      _linp_ + _xrow_2_{_i_} * _beta_2_{_i_};
   end;
   /*------------------------------------------*/
end;

when ('5') do;
   /*Initialize temporary array for x row*/
   do _i_=1 to 12; _xrow_2_{_i_} = 0; end;

   /*------------------------------------------*/
   /*Fill in the X row for the observation     */
   /*based on converting the values of model   */
   /*variables to effect positions.            */
   /*Effect: Intercept*/
   _xrow_2_[1] = 1;
   /*Effect: Age*/
   _xrow_2_[2] = DemAge;
   /*Effect: 'Affluence Grade'n*/
   _xrow_2_[3] = DemAffl;
   /*Effect: 'Loyalty Card Tenure'n*/
   _xrow_2_[4] = PromTime;
   /*Effect: 'Geographic Region'n*/
```

```
    _temp_ = 1;
    select (_DemReg_);
       when ('Midlands') _xrow_2_[5] = _temp_;
       when ('North') _xrow_2_[6] = _temp_;
       when ('Scottish') _xrow_2_[7] = _temp_;
       when ('South East') _xrow_2_[8] = _temp_;
       when ('South West') _xrow_2_[9] = _temp_;
       when ('Unknown') _xrow_2_[10] = _temp_;
       otherwise do; _badval_ = 1; goto skip_2_0; end;
    end;

    /*Effect: 'Home Owner'n*/
    _temp_ = 1;
    select (_DemHomeowner_);
       when ('No') _xrow_2_[11] = _temp_;
       when ('Yes') _xrow_2_[12] = _temp_;
       otherwise do; _badval_ = 1; goto skip_2_0; end;
    end;

    /*------------------------------------------*/

    /*Initialize beta*/
    _beta_2_{1} = 1.21444459282263;
    _beta_2_{2} = -0.09742309716793;
    _beta_2_{3} = 0.27664357746948;
    _beta_2_{4} = -0.00442376220383;
    _beta_2_{5} = 0.24773157984164;
    _beta_2_{6} = 0.07977280042884;
    _beta_2_{7} = 0.23955990700338;
    _beta_2_{8} = 0.17198343983921;
    _beta_2_{9} = 0.42710328423198;
    _beta_2_{10} = 0;
    _beta_2_{11} = -0.00165338612423;
    _beta_2_{12} = 0;


    /*------------------------------------------*/
    /*Compute the linear predictor             */
    do _i_=1 to 12;
       _linp_ + _xrow_2_{_i_} * _beta_2_{_i_};
    end;
    /*------------------------------------------*/
 end;

 when ('6') do;
    /*Initialize temporary array for x row*/
    do _i_=1 to 12; _xrow_2_{_i_} = 0; end;

    /*------------------------------------------*/
    /*Fill in the X row for the observation     */
    /*based on converting the values of model   */
    /*variables to effect positions.            */
    /*Effect: Intercept*/
    _xrow_2_[1] = 1;
```

```
   /*Effect: Age*/
   _xrow_2_[2] = DemAge;
   /*Effect: 'Affluence Grade'n*/
   _xrow_2_[3] = DemAffl;
   /*Effect: 'Loyalty Card Tenure'n*/
   _xrow_2_[4] = PromTime;
   /*Effect: 'Geographic Region'n*/
   _temp_ = 1;
   select (_DemReg_);
      when ('Midlands') _xrow_2_[5] = _temp_;
      when ('North') _xrow_2_[6] = _temp_;
      when ('Scottish') _xrow_2_[7] = _temp_;
      when ('South East') _xrow_2_[8] = _temp_;
      when ('South West') _xrow_2_[9] = _temp_;
      when ('Unknown') _xrow_2_[10] = _temp_;
      otherwise do; _badval_ = 1; goto skip_2_0; end;
   end;

   /*Effect: 'Home Owner'n*/
   _temp_ = 1;
   select (_DemHomeowner_);
      when ('No') _xrow_2_[11] = _temp_;
      when ('Yes') _xrow_2_[12] = _temp_;
      otherwise do; _badval_ = 1; goto skip_2_0; end;
   end;

   /*------------------------------------------*/

   /*Initialize beta*/
   _beta_2_{1} = -3.14465158326961;
   _beta_2_{2} = 0.00083648427367;
   _beta_2_{3} = 0.21451685132757;
   _beta_2_{4} = -0.01032344902381;
   _beta_2_{5} = -0.0590322129303;
   _beta_2_{6} = -0.10403816936141;
   _beta_2_{7} = -0.12986101138161;
   _beta_2_{8} = 0.01137499954899;
   _beta_2_{9} = -0.22138913463814;
   _beta_2_{10} = 0;
   _beta_2_{11} = 0.01025958406289;
   _beta_2_{12} = 0;


   /*------------------------------------------*/
   /*Compute the linear predictor              */
   do _i_=1 to 12;
      _linp_ + _xrow_2_{_i_} * _beta_2_{_i_};
   end;
   /*------------------------------------------*/
end;

when ('8') do;
   /*Initialize temporary array for x row*/
   do _i_=1 to 12; _xrow_2_{_i_} = 0; end;
```

```
/*------------------------------------------*/
/*Fill in the X row for the observation     */
/*based on converting the values of model   */
/*variables to effect positions.            */
/*Effect: Intercept*/
_xrow_2_[1] = 1;
/*Effect: Age*/
_xrow_2_[2] = DemAge;
/*Effect: 'Affluence Grade'n*/
_xrow_2_[3] = DemAffl;
/*Effect: 'Loyalty Card Tenure'n*/
_xrow_2_[4] = PromTime;
/*Effect: 'Geographic Region'n*/
_temp_ = 1;
select (_DemReg_);
   when ('Midlands') _xrow_2_[5] = _temp_;
   when ('North') _xrow_2_[6] = _temp_;
   when ('Scottish') _xrow_2_[7] = _temp_;
   when ('South East') _xrow_2_[8] = _temp_;
   when ('South West') _xrow_2_[9] = _temp_;
   when ('Unknown') _xrow_2_[10] = _temp_;
   otherwise do; _badval_ = 1; goto skip_2_0; end;
end;

/*Effect: 'Home Owner'n*/
_temp_ = 1;
select (_DemHomeowner_);
   when ('No') _xrow_2_[11] = _temp_;
   when ('Yes') _xrow_2_[12] = _temp_;
   otherwise do; _badval_ = 1; goto skip_2_0; end;
end;

/*------------------------------------------*/

/*Initialize beta*/
_beta_2_{1} = -7.1672754186378;
_beta_2_{2} = 0.03946642814959;
_beta_2_{3} = 0.25911924885958;
_beta_2_{4} = 0.00238558676971;
_beta_2_{5} = -0.08157863380409;
_beta_2_{6} = -0.08523497266295;
_beta_2_{7} = -0.00533290277378;
_beta_2_{8} = -0.2105615020503;
_beta_2_{9} = 0.4138037582053;
_beta_2_{10} = 0;
_beta_2_{11} = 0.05470444324966;
_beta_2_{12} = 0;


/*------------------------------------------*/
/*Compute the linear predictor              */
do _i_=1 to 12;
   _linp_ + _xrow_2_{_i_} * _beta_2_{_i_};
```

```
   end;
   /*------------------------------------------*/
end;

when ('12') do;
   /*Initialize temporary array for x row*/
   do _i_=1 to 12; _xrow_2_{_i_} = 0; end;

   /*------------------------------------------*/
   /*Fill in the X row for the observation     */
   /*based on converting the values of model   */
   /*variables to effect positions.            */
   /*Effect: Intercept*/
   _xrow_2_[1] = 1;
   /*Effect: Age*/
   _xrow_2_[2] = DemAge;
   /*Effect: 'Affluence Grade'n*/
   _xrow_2_[3] = DemAffl;
   /*Effect: 'Loyalty Card Tenure'n*/
   _xrow_2_[4] = PromTime;
   /*Effect: 'Geographic Region'n*/
   _temp_ = 1;
   select (_DemReg_);
      when ('Midlands') _xrow_2_[5] = _temp_;
      when ('North') _xrow_2_[6] = _temp_;
      when ('Scottish') _xrow_2_[7] = _temp_;
      when ('South East') _xrow_2_[8] = _temp_;
      when ('South West') _xrow_2_[9] = _temp_;
      when ('Unknown') _xrow_2_[10] = _temp_;
      otherwise do; _badval_ = 1; goto skip_2_0; end;
   end;

   /*Effect: 'Home Owner'n*/
   _temp_ = 1;
   select (_DemHomeowner_);
      when ('No') _xrow_2_[11] = _temp_;
      when ('Yes') _xrow_2_[12] = _temp_;
      otherwise do; _badval_ = 1; goto skip_2_0; end;
   end;

   /*------------------------------------------*/

   /*Initialize beta*/
   _beta_2_{1} = -14.2145396772944;
   _beta_2_{2} = -0.05319711248371;
   _beta_2_{3} = 0.240490295592;
   _beta_2_{4} = 0.21682703060606;
   _beta_2_{5} = 12.1873792682116;
   _beta_2_{6} = 13.0305296189784;
   _beta_2_{7} = 12.3665771769083;
   _beta_2_{8} = 12.1485960554157;
   _beta_2_{9} = 25.0667987805237;
   _beta_2_{10} = 0;
   _beta_2_{11} = -0.03205850841643;
```

```
    _beta_2_{12} = 0;


    /*------------------------------------------*/
    /*Compute the linear predictor              */
    do _i_=1 to 12;
       _linp_ + _xrow_2_{_i_} * _beta_2_{_i_};
    end;
    /*------------------------------------------*/
 end;

 when ('19') do;
    /*Initialize temporary array for x row*/
    do _i_=1 to 12; _xrow_2_{_i_} = 0; end;

    /*------------------------------------------*/
    /*Fill in the X row for the observation     */
    /*based on converting the values of model   */
    /*variables to effect positions.            */
    /*Effect: Intercept*/
    _xrow_2_[1] = 1;
    /*Effect: Age*/
    _xrow_2_[2] = DemAge;
    /*Effect: 'Affluence Grade'n*/
    _xrow_2_[3] = DemAffl;
    /*Effect: 'Loyalty Card Tenure'n*/
    _xrow_2_[4] = PromTime;
    /*Effect: 'Geographic Region'n*/
    _temp_ = 1;
    select (_DemReg_);
       when ('Midlands') _xrow_2_[5] = _temp_;
       when ('North') _xrow_2_[6] = _temp_;
       when ('Scottish') _xrow_2_[7] = _temp_;
       when ('South East') _xrow_2_[8] = _temp_;
       when ('South West') _xrow_2_[9] = _temp_;
       when ('Unknown') _xrow_2_[10] = _temp_;
       otherwise do; _badval_ = 1; goto skip_2_0; end;
    end;

    /*Effect: 'Home Owner'n*/
    _temp_ = 1;
    select (_DemHomeowner_);
       when ('No') _xrow_2_[11] = _temp_;
       when ('Yes') _xrow_2_[12] = _temp_;
       otherwise do; _badval_ = 1; goto skip_2_0; end;
    end;

    /*------------------------------------------*/

    /*Initialize beta*/
    _beta_2_{1} = 1.50027774356977;
    _beta_2_{2} = -0.11448485619938;
    _beta_2_{3} = 0.24741210197092;
    _beta_2_{4} = 0.01578320503247;
```

```
   _beta_2_{5} = -0.09778772434504;
   _beta_2_{6} = -0.03362613418269;
   _beta_2_{7} = -0.22874852430825;
   _beta_2_{8} = -0.12658668600807;
   _beta_2_{9} = -0.22471650494822;
   _beta_2_{10} = 0;
   _beta_2_{11} = 0.01253634101801;
   _beta_2_{12} = 0;


   /*------------------------------------------*/
   /*Compute the linear predictor              */
   do _i_=1 to 12;
      _linp_ + _xrow_2_{_i_} * _beta_2_{_i_};
   end;
   /*------------------------------------------*/
end;

when ('20') do;
   /*Initialize temporary array for x row*/
   do _i_=1 to 12; _xrow_2_{_i_} = 0; end;

   /*------------------------------------------*/
   /*Fill in the X row for the observation     */
   /*based on converting the values of model   */
   /*variables to effect positions.            */
   /*Effect: Intercept*/
   _xrow_2_[1] = 1;
   /*Effect: Age*/
   _xrow_2_[2] = DemAge;
   /*Effect: 'Affluence Grade'n*/
   _xrow_2_[3] = DemAffl;
   /*Effect: 'Loyalty Card Tenure'n*/
   _xrow_2_[4] = PromTime;
   /*Effect: 'Geographic Region'n*/
   _temp_ = 1;
   select (_DemReg_);
      when ('Midlands') _xrow_2_[5] = _temp_;
      when ('North') _xrow_2_[6] = _temp_;
      when ('Scottish') _xrow_2_[7] = _temp_;
      when ('South East') _xrow_2_[8] = _temp_;
      when ('South West') _xrow_2_[9] = _temp_;
      when ('Unknown') _xrow_2_[10] = _temp_;
      otherwise do; _badval_ = 1; goto skip_2_0; end;
   end;

   /*Effect: 'Home Owner'n*/
   _temp_ = 1;
   select (_DemHomeowner_);
      when ('No') _xrow_2_[11] = _temp_;
      when ('Yes') _xrow_2_[12] = _temp_;
      otherwise do; _badval_ = 1; goto skip_2_0; end;
   end;
```

```
         /*------------------------------------------*/

         /*Initialize beta*/
         _beta_2_{1} = 1.48927053549323;
         _beta_2_{2} = -0.11324360466851;
         _beta_2_{3} = 0.29535217180033;
         _beta_2_{4} = 0.05040546065422;
         _beta_2_{5} = -0.90866569305208;
         _beta_2_{6} = -0.8723864711379;
         _beta_2_{7} = -1.26499296063709;
         _beta_2_{8} = -0.85372383102273;
         _beta_2_{9} = -0.90184072351173;
         _beta_2_{10} = 0;
         _beta_2_{11} = -0.01882111719423;
         _beta_2_{12} = 0;


         /*------------------------------------------*/
         /*Compute the linear predictor              */
         do _i_=1 to 12;
            _linp_ + _xrow_2_{_i_} * _beta_2_{_i_};
         end;
         /*------------------------------------------*/
      end;

      otherwise do; _badval_ = 1; goto skip_2_0; end;
   end;

   /*------------------------------------------*/
   /*Compute predicted mean value from the     */
   /*linear predictor, taking into account the */
   /*link function.                            */
   skip_2_0:
   if (_badval_ eq 0) and not missing(_linp_) then do;
      if (_linp_ > 0) then do;
         P__va_response_2 = 1 / (1+exp(-_linp_));
      end; else do;
         P__va_response_2 = exp(_linp_) / (1+exp(_linp_));
      end;
   end; else do;
      _linp_ = .;
      P__va_response_2 = .;
   end;
   /*------------------------------------------*/
   drop _va_:;
   'P_TargetBuy'n='P__va_response_2'n;

   drop 'C1000000'n;
   drop 'P__va_response_2'n;
/*------------------------------------------*/
```