# Assessment 2C – Week 4 Workbook Exercise

## Logistic Regression

|  |  |  |
| --- | --- | --- |
| Question | Answer | SAS VA visualisation reference |
| 1. How many rows of data does BIOORGANICS contain? | 111,115 |  |
| 1. How many missing values are there for *Organics Purchase Indicator*? | None |  |
| 1. How many customers purchased organic products? | 27,525 | Cross-tab 1  (not exported) |
| 1. Which gender buys more organic products? | Females | Visualisation 1 |
| 1. Which statement most correctly describes the conclusion you can make about those people who buy organic products? | 1. People who are younger and more affluent tend to buy more organic products. | Visualisation 2 |
| 1. How many observations were used to build the model? | 82,040 | Logistic 1 |
| 1. Which variables are not considered important to the model? | Total Spend, Geographic Region and Neighbourhood Cluster-7 level | Logistic 1 – Fit Summary |
| 1. What is the R square for the new model? | 0.2258 | Logistic – with IM & VarSel |
| 1. What is the K-S statistic for the new model? | 0.4706 | Logistic – with IM & VarSel Assessment |
| 1. Which model is selected when the fit statistic is Misclassification? | Logistic – with IM & VarSel |  |
| 1. Which model is selected when the prediction cutoff is set to 0.25 and the FDR statistic is chosen? | Logistic 1 |  |