

IBM SPSS Decision Trees

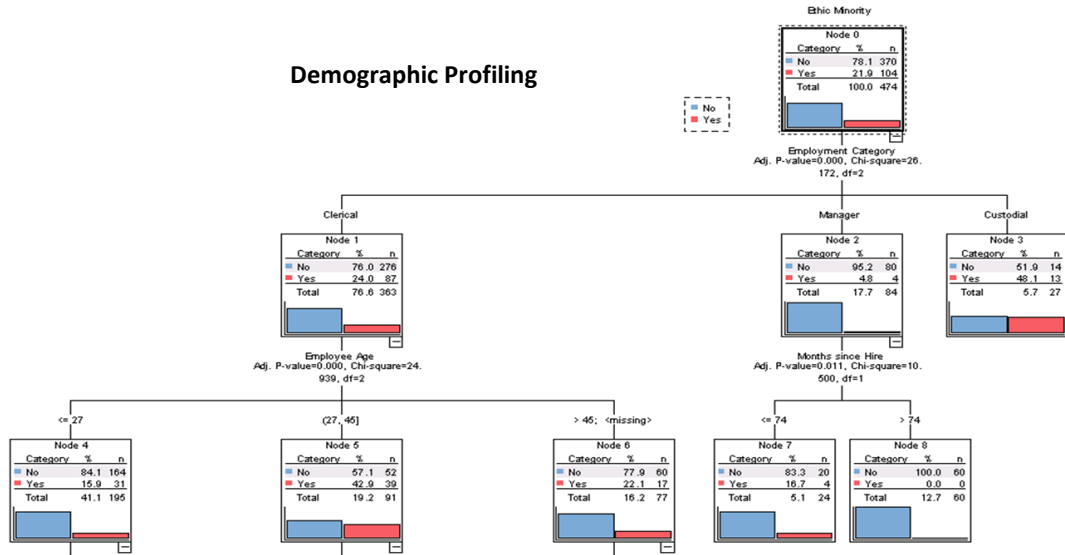
- The IBM SPSS Decision Trees procedure creates a tree-based classification model. Decision Trees can be used as predictive models to predict the values of a dependent (target) variable based on values of independent (predictor) variables. This approach is often used as an alternative to methods such as Logistic Regression.
- Because the Decision Trees module is frequently used to correctly categorise cases into a target group, it may be applied in *segmentation* and *profiling* applications where the analysts wish to describe customers who are more likely to be more dissatisfied than others. It can also be used to describe cluster membership where the target field is the resultant cluster variable of an SPSS cluster analysis.
- Decision Trees uses four algorithms as growing methods:
 - CHAID (Chi Square Automatic Interaction Detector)
 - Exhaustive CHAID
 - CRT (Classification and Regression Tree)
 - QUEST (Quick Unbiased Efficient Statistical Tree)
- The following table describes the various features of the different growing methods:

Feature	CHAID*	CRT	QUEST
Chi-square-based**	■		
Surrogate independent (predictor) variables		■	■
Tree pruning		■	■
Multiway node splitting	■		
Binary node splitting		■	■
Influence variables	■	■	
Prior probabilities		■	■
Fast calculation	■		■

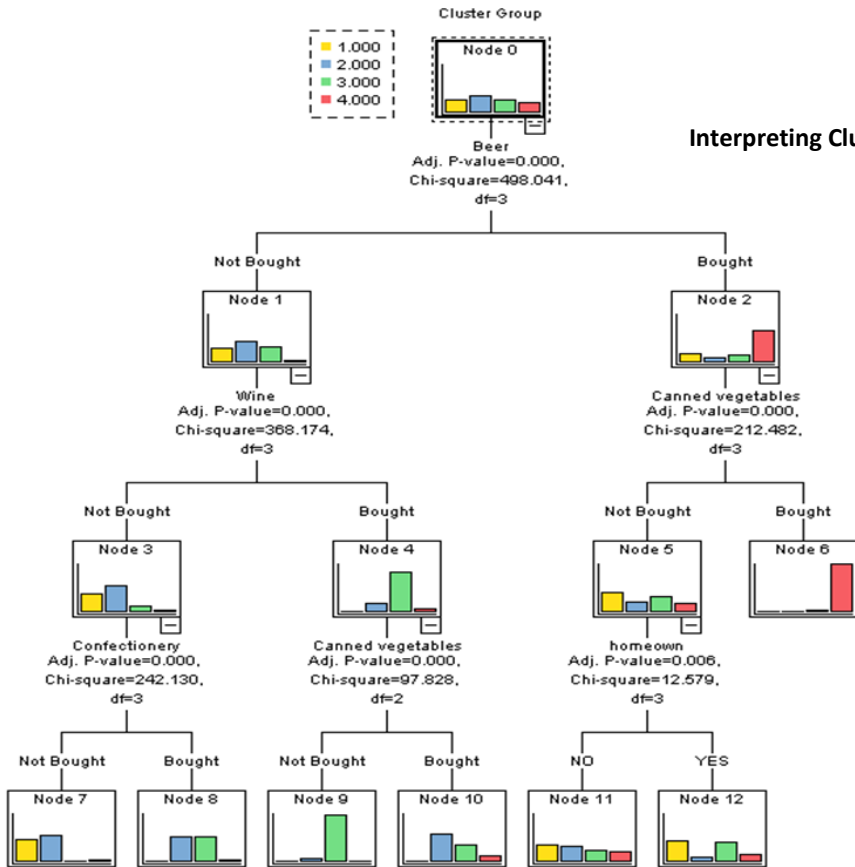
*Includes Exhaustive CHAID.

**QUEST also uses a chi-square measure for nominal independent variables.

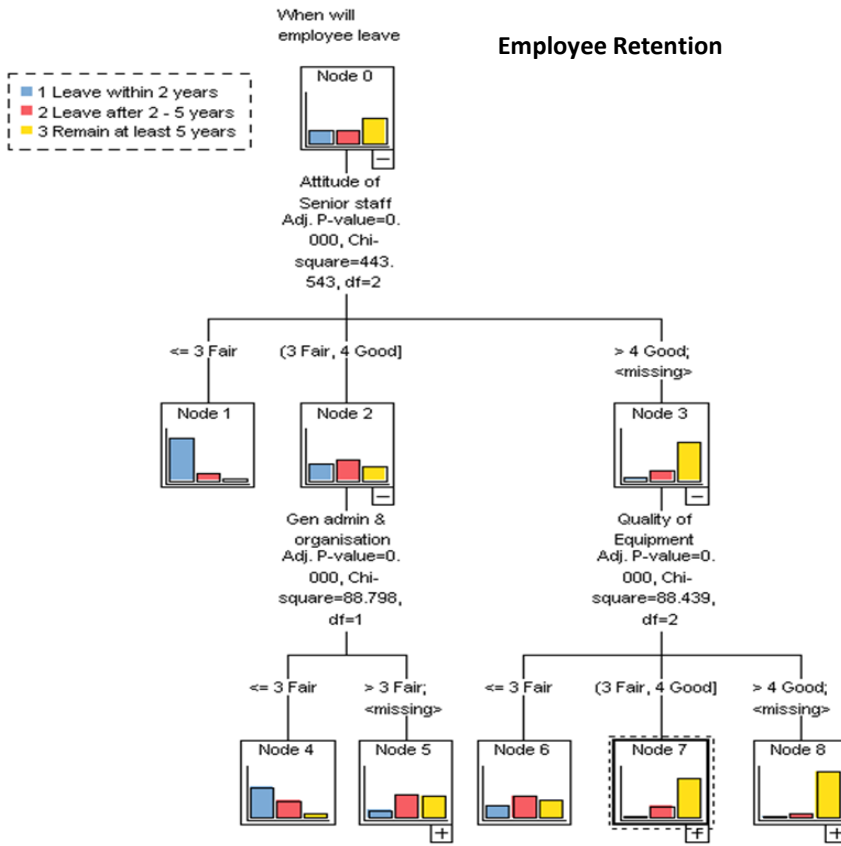
Demographic Profiling



Interpreting Cluster Membership



Employee Retention



Examining Passenger Survival from the RMS Titanic

