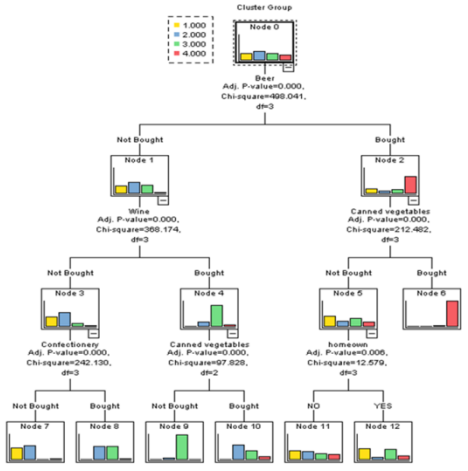
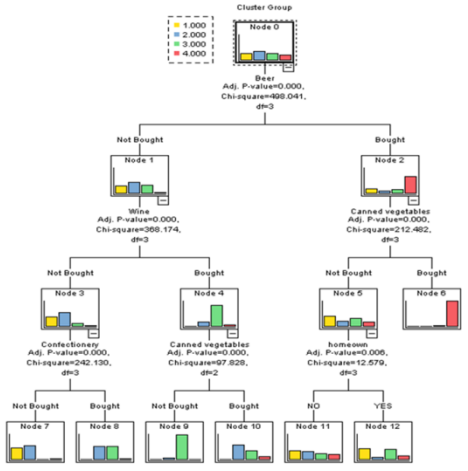
SMARTVISION Europe



Getting started with Decision Trees in SPSS Statistics

Jarlath Quinn

SMARTVISION Europe



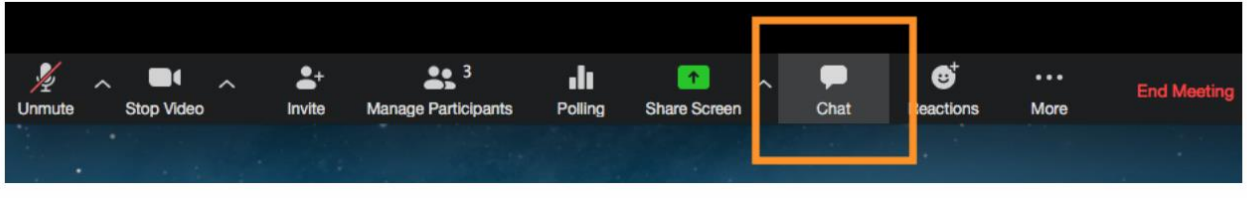
Introduction to Decision Trees

Jarlath Quinn

Just waiting for all attendees to join…

FAQ's

- Is this session being recorded? Yes
- Can I get a copy of the slides? Yes, we'll email links to download materials after the session has ended.
- Can we arrange a re-run for colleagues? Yes, just ask us.
- How can I ask questions? All lines are muted so please use the chat panel – if we run out of time we will follow up with you.



- Gold accredited partner to IBM, Predictive Solutions and DataRobot specialising in advanced analytics & big data technologies
- Work with open-source technologies (R, Python, Spark etc.)
- Team each have 15 to 30 years of experience working in the advanced and predictive analytics industry

- Deep experience of applied advanced analytics applications across sectors
  - Retail
  - Gaming
  - Utilities
  - Insurance
  - Telecommunications
  - Media
  - FMCG

Types of Predictive Analytics

- **Classification / Propensity**
  - Who is most likely to respond / upgrade/recommend/defect based on the historical behavioural data we have about them?
- **Clustering**
  - How can I divide my customers into meaningful and usable groups as a framework for marketing communications?
- **Association & Sequence**
  - What combinations of product purchases or events co-occur more often than normal? What natural affinities exist within the data?
- **Time Series**
  - What will product demand/revenue/website hits/visitor numbers look like in the next hour/day/month/quarter/ year?

Where do Decision Trees fit within Predictive Analytics?

- Decision trees are used *extensively and widely* within Predictive Analytics
- Decision trees can be used to
  - Build profiles of customers/employees/clients
  - Find key behavioural segments
  - Generate predictive models
- Decision Trees can be expressed as a series of hierarchical rules which means that they can be converted in languages like SQL for database scoring
- Decision Trees are especially popular because
  - they are fairly visual representations of models
  - relatively easy to understand

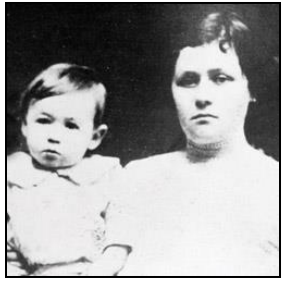
Understanding Decision Trees – a worked example

- What were the most important factors determining survival during the sinking of the RMS Titanic?

Survival on the RMS Titanic

		Count	Percent %
survive	Did not survive	1490	68%
	Survived	711	32%
	Total	2201	100%

Gender?



Age?

Class?

Statistical Tests Like Chi Square help to answer this

Survival on the RMS Titanic

		sex			
		female		male	
		Count	Column Percent %	Count	Column Percent %
survive	Did not survive	126	26.8%	1364	78.8%
	Survived	344	73.2%	367	21.2%
	Total	470	100.0%	1731	100.0%

Pearson Chi-Square Tests

		sex
survive	Chi-square	456.874
	df	1
	Sig.	.000*

Statistical Tests Like Chi Square help to answer this

Survival on the RMS Titanic

		age			
		adult		child	
		Count	Column Percent %	Count	Column Percent %
survive	Did not survive	1438	68.7%	52	47.7%
	Survived	654	31.3%	57	52.3%
	Total	2092	100.0%	109	100.0%

Pearson Chi-Square Tests

		age
survive	Chi-square	20.956
	df	1
	Sig.	.000*

Statistical Tests Like Chi Square help to answer this

Survival on the RMS Titanic

		class							
		1st		2nd		3rd		crew	
		Count	Column Percent %	Count	Column Percent %	Count	Column Percent %	Count	Column Percent %
survive	Did not survive	122	37.5%	167	58.6%	528	74.8%	673	76.0%
	Survived	203	62.5%	118	41.4%	178	25.2%	212	24.0%
	Total	325	100.0%	285	100.0%	706	100.0%	885	100.0%

Pearson Chi-Square Tests

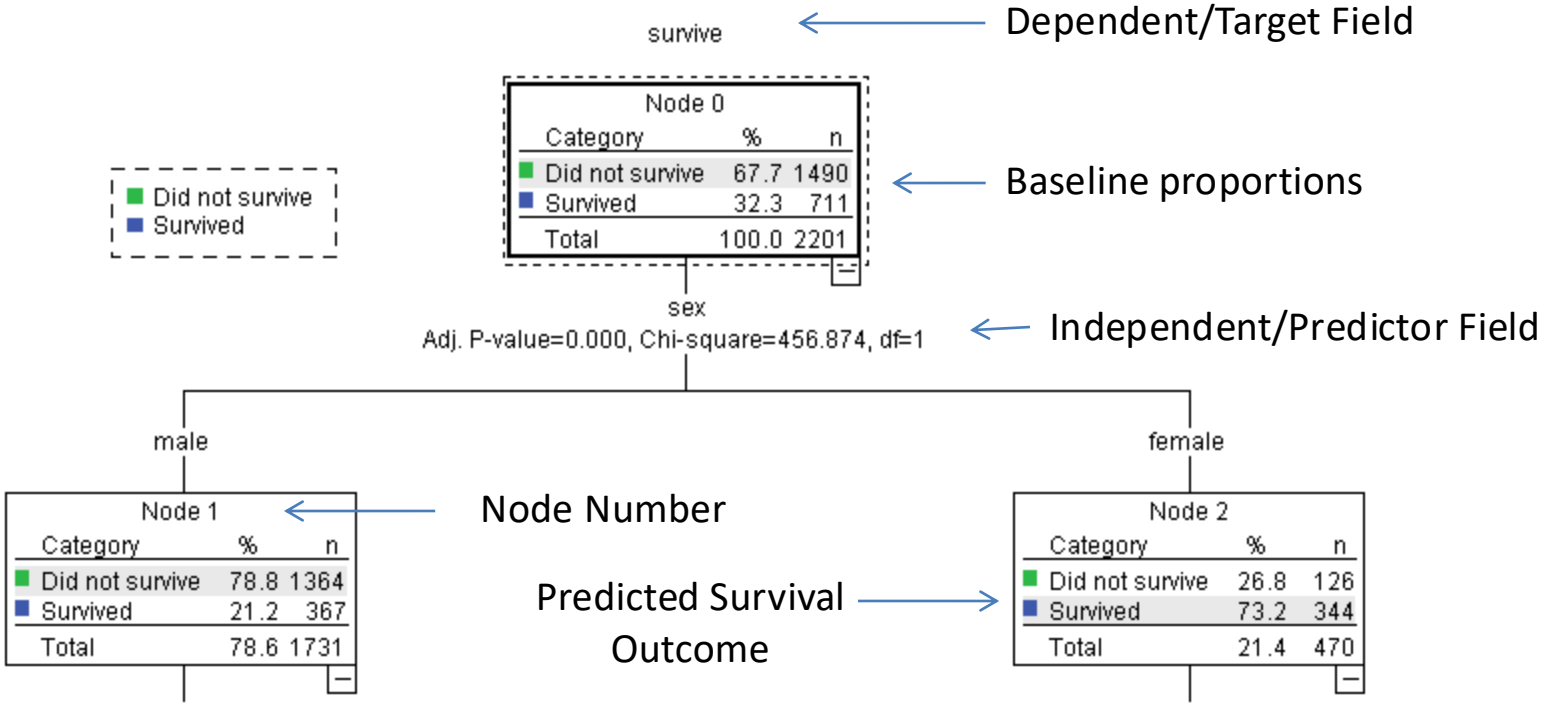
		class
survive	Chi-square	190.401
	df	3
	Sig.	.000*

SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

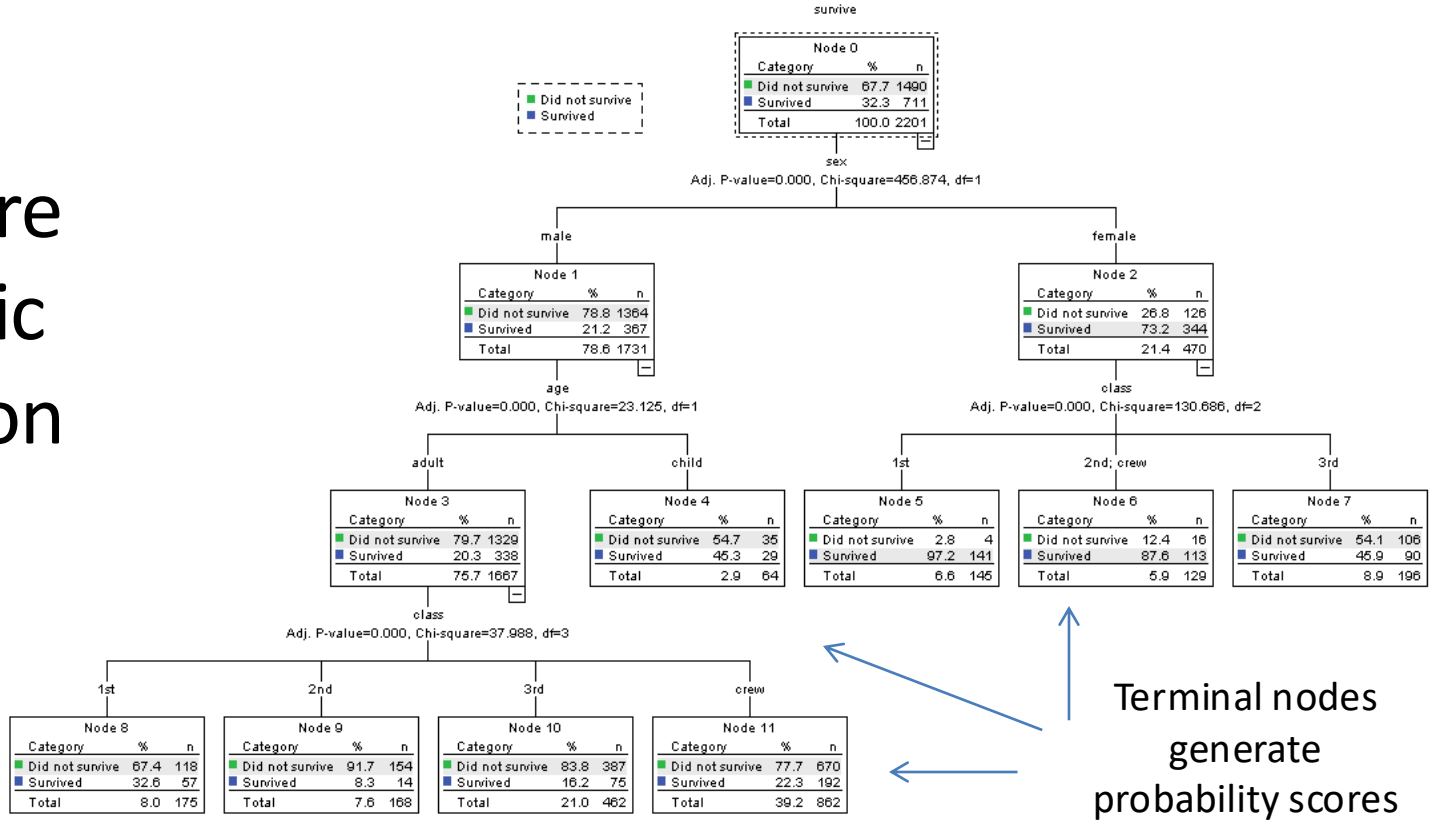
Gender is most important

…and a CHAID Decision tree will reflect this….



Full CHAID Decision Tree

C.H.A.I.D
Chi-Square
Automatic
Interaction
Detector



Terminal nodes generate probability scores

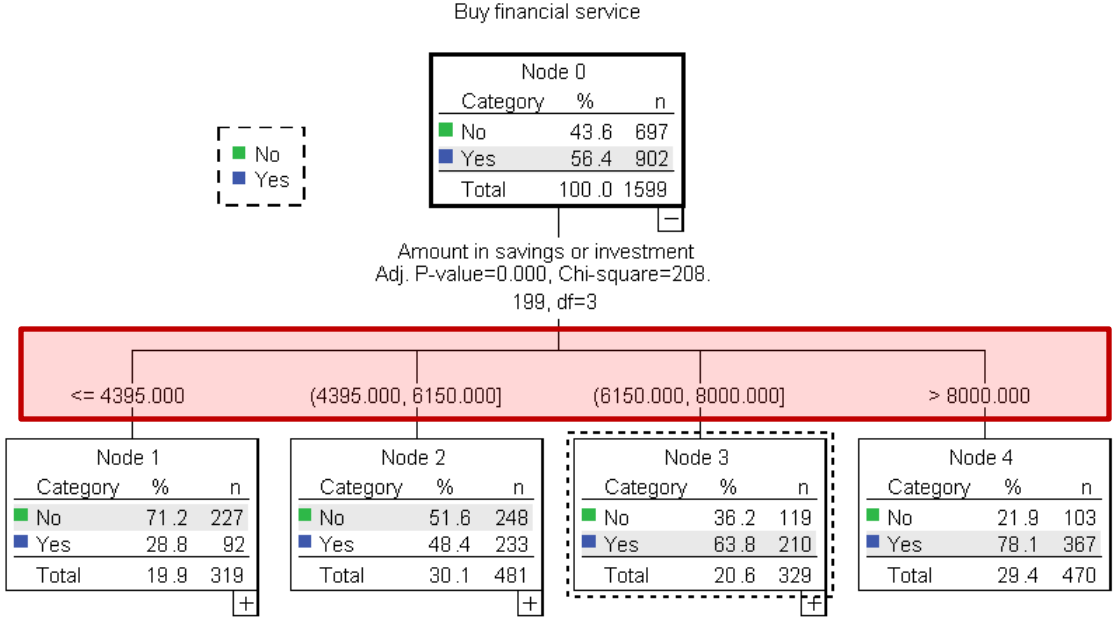
SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Merging/Splitting in CHAID Trees

Decision Trees can merge values of numeric *and* categorical predictors together

This makes the tree more efficient and easier to read



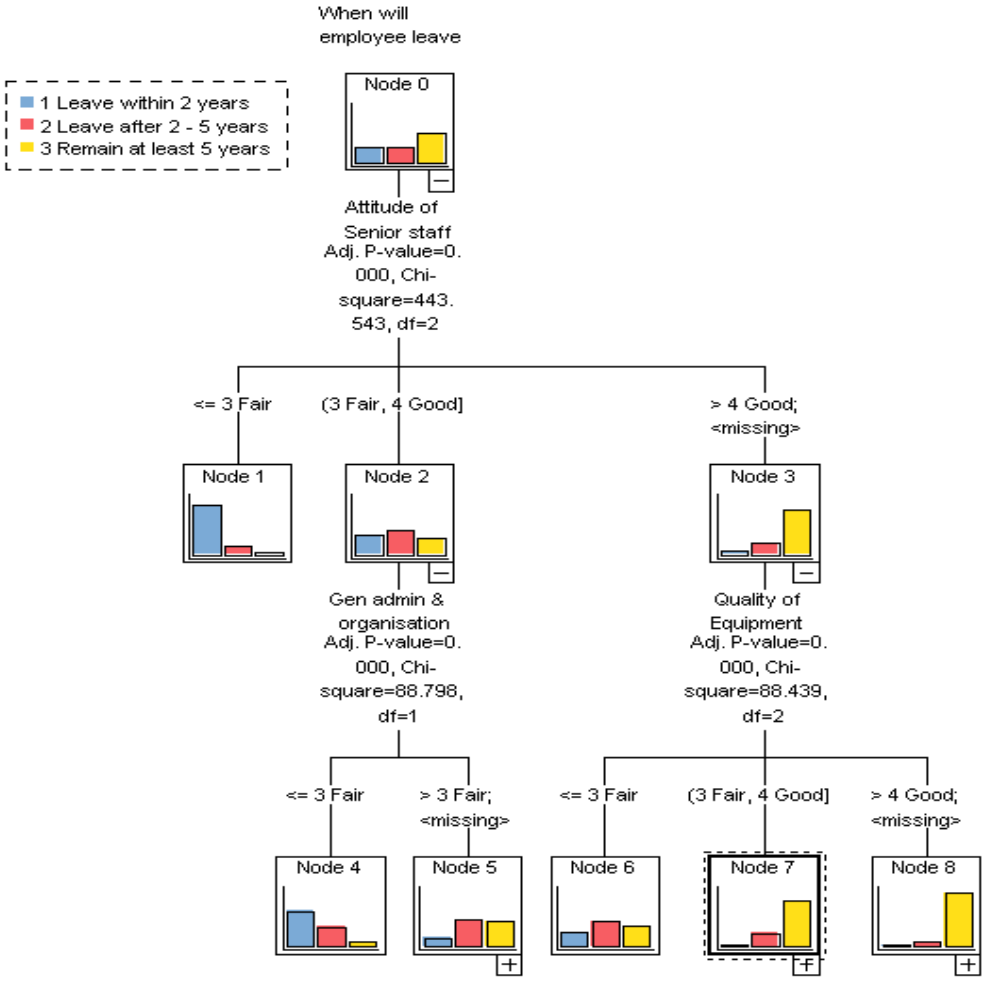
SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

SMARTVISION
Europe

Let’s see a demonstration…

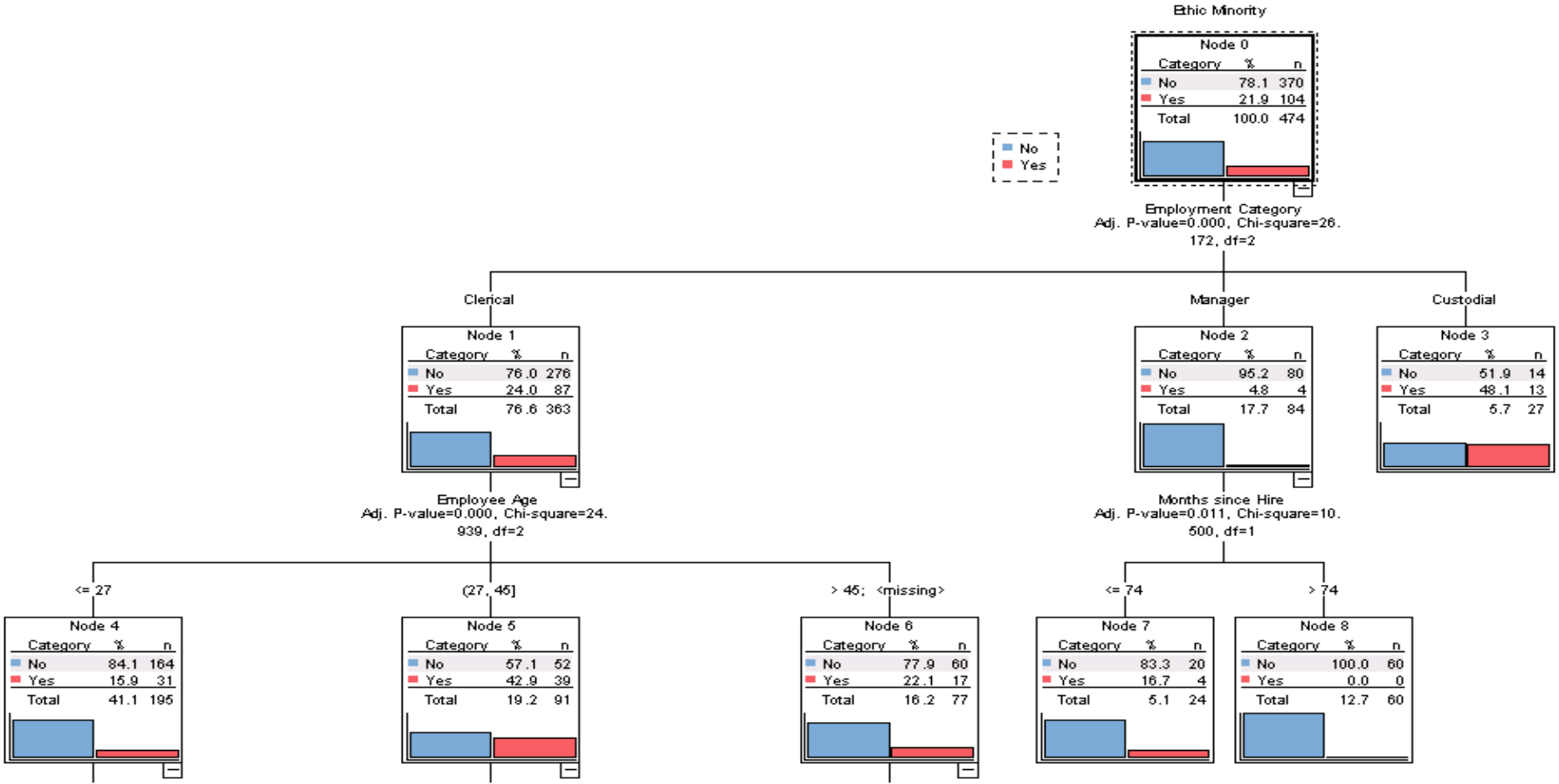
Drivers of Satisfaction



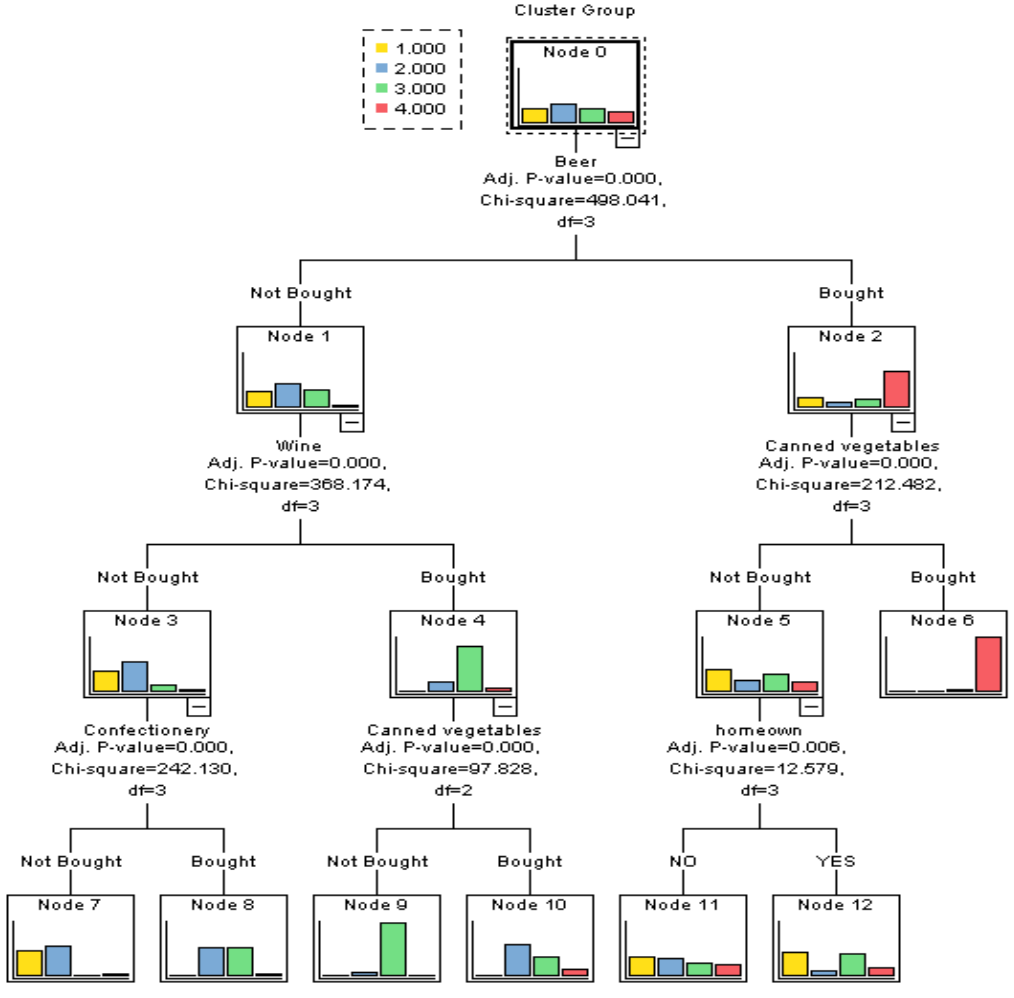
SMARTVIS Europe

SELECT INTERNATIONAL COMPANY

Demographic Profiling

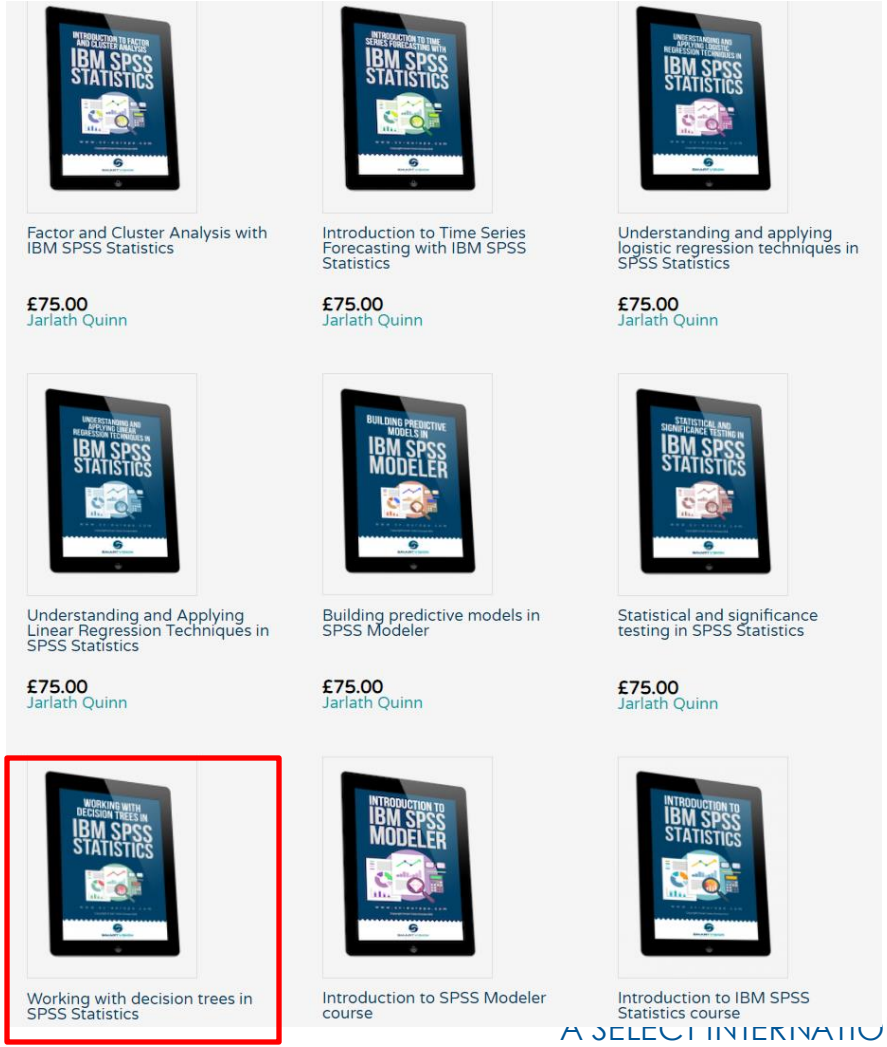


Interpreting Clusters



SMARTVISION Europe

Online training materials free to Smart Vision customers or available for purchase



SMARTVISION Europe

A SELECT INTERNATIONAL COMPANY

Working with Smart Vision Europe

We can help with processing and analysing your data.

- Self-paced, virtual and in-person training courses in how to use SPSS products and appropriate statistical techniques
- A mix of consultancy and training whereby we do the initial work and then teach you how to replicate it

Working with Smart Vision Europe

- **Sourcing Software**
  - You can buy your analytical software from us often with discounts
  - Assist with selection, pilot, implementation & support of analytical tools
  - http://www.sv-europe.com/buy-spss-online/
- **Training**
  - Formal classroom/virtual training
  - Custom course development
  - Informal 'bite-size' training split over time
- **Advice and Support**
  - 'No strings attached' technical and business advice relating to analytics
  - Tracked technical support services around the IBM SPSS product line

SMARTVISION Europe

Contact us:

+44 (0)207 786 3568
info@sv-europe.com
Twitter: @sveurope
Follow us on Linked In
Sign up for our Newsletter

Thank you