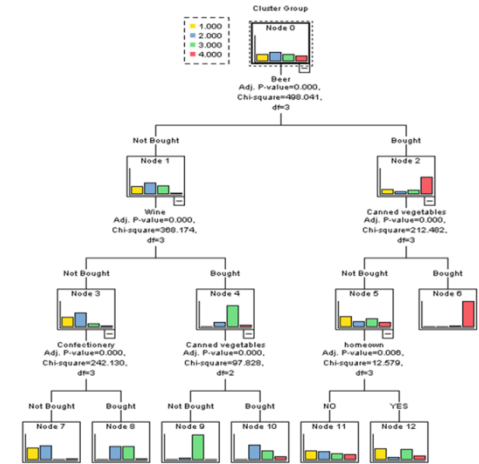


Introduction to Decision Trees

Jarlath Quinn

www.sv-europe.com

A SELECT INTERNATIONAL COMPANY



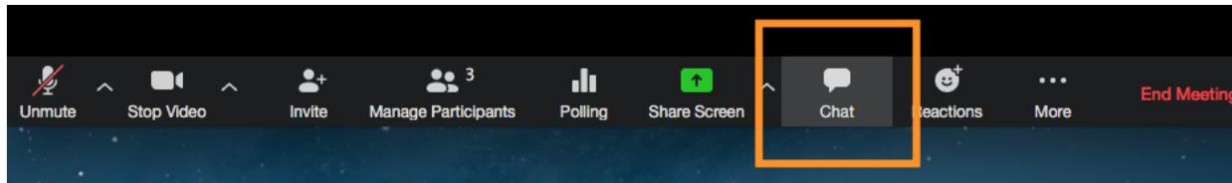
Introduction to Decision Trees

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

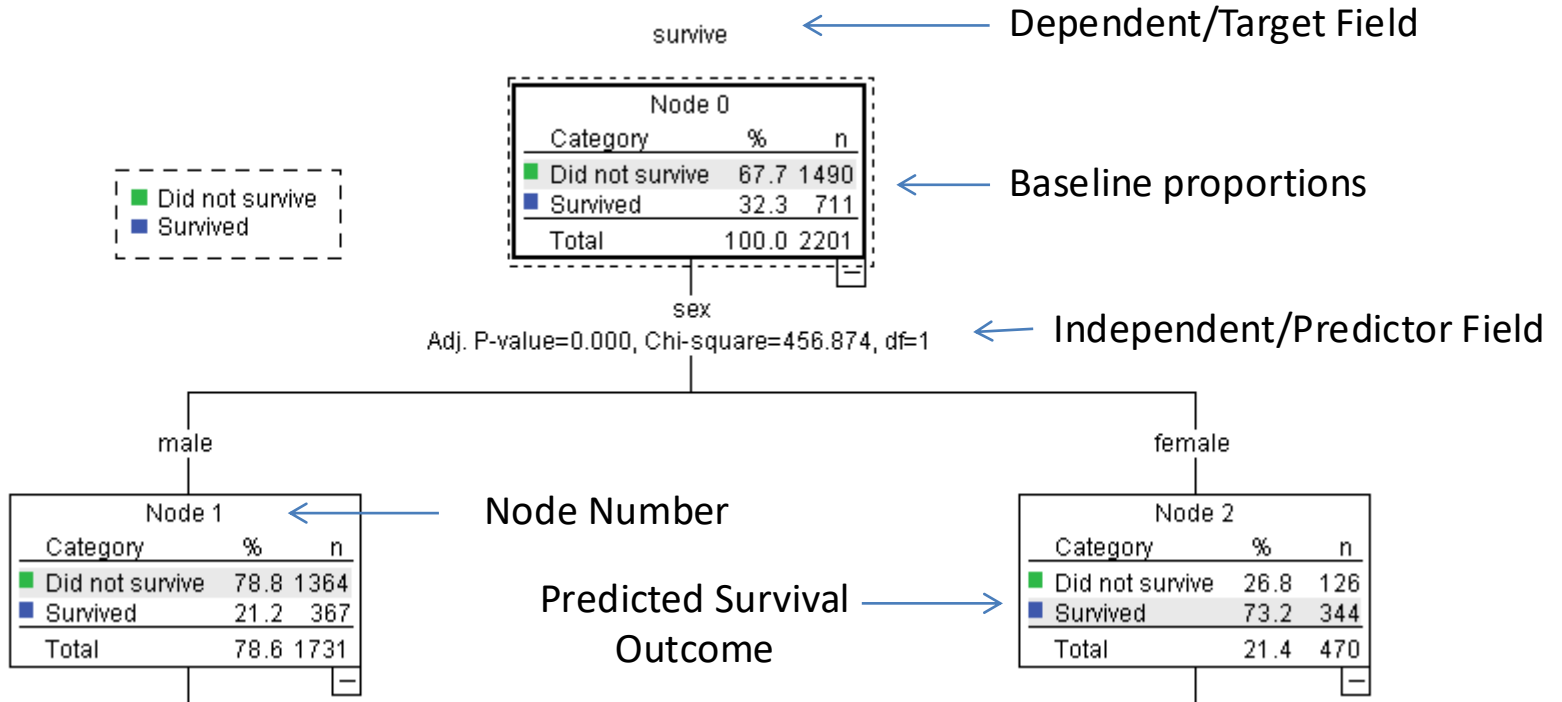
		class							
		1st		2nd		3rd		crew	
		Count	Column Percent %	Count	Column Percent %	Count	Column Percent %	Count	Column Percent %
survive	Did not survive	122	37.5%	167	58.6%	528	74.8%	673	76.0%
	Survived	203	62.5%	118	41.4%	178	25.2%	212	24.0%
	Total	325	100.0%	285	100.0%	706	100.0%	885	100.0%

Pearson Chi-Square Tests

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

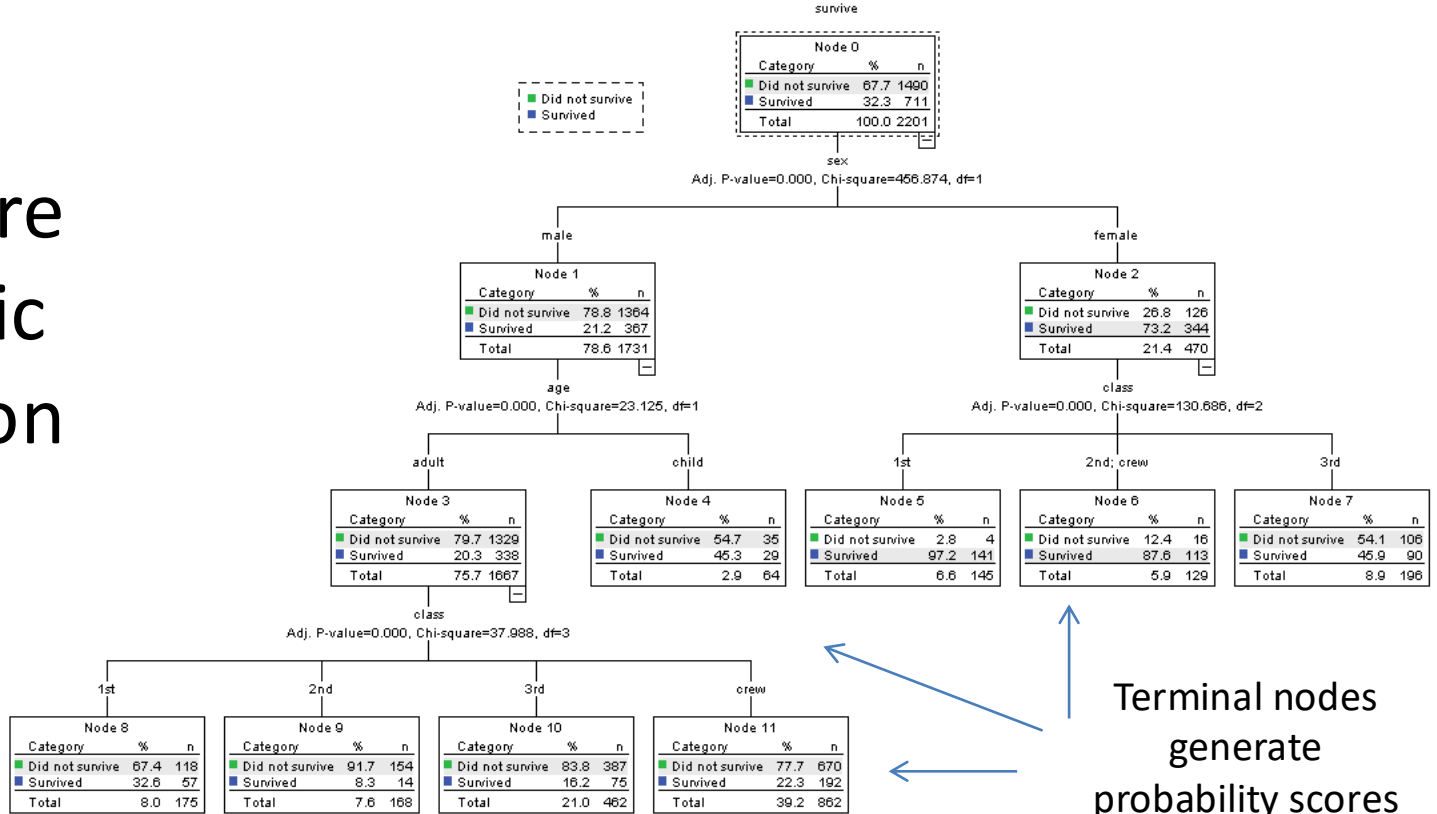
Gender is most important

...and a CHAID Decision tree will reflect this....



Full CHAID Decision Tree

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

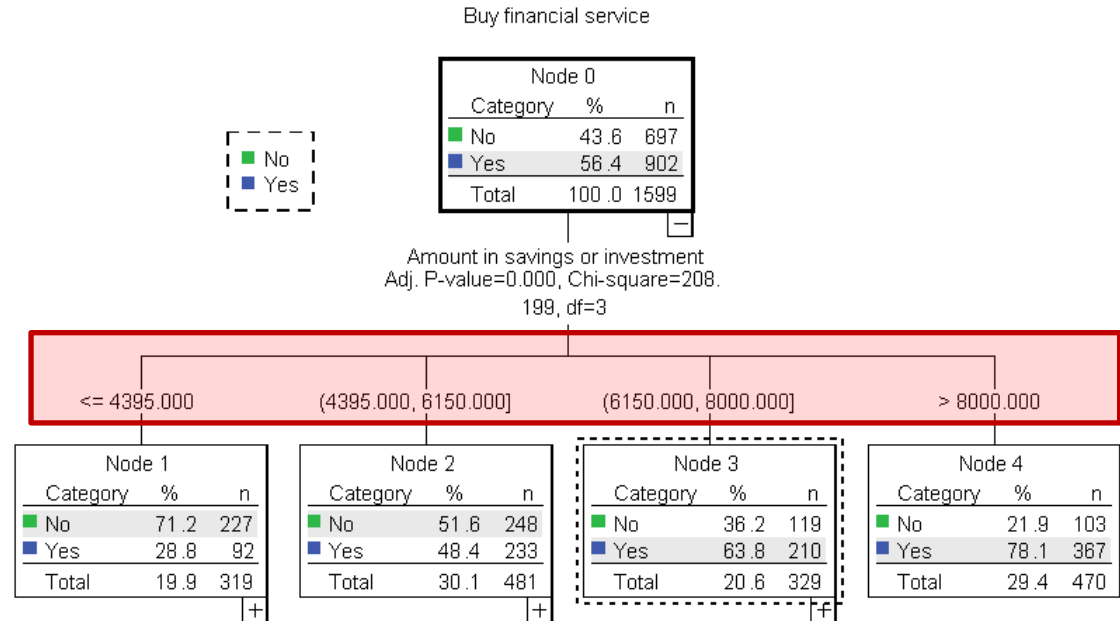


Terminal nodes
 generate
 probability scores

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



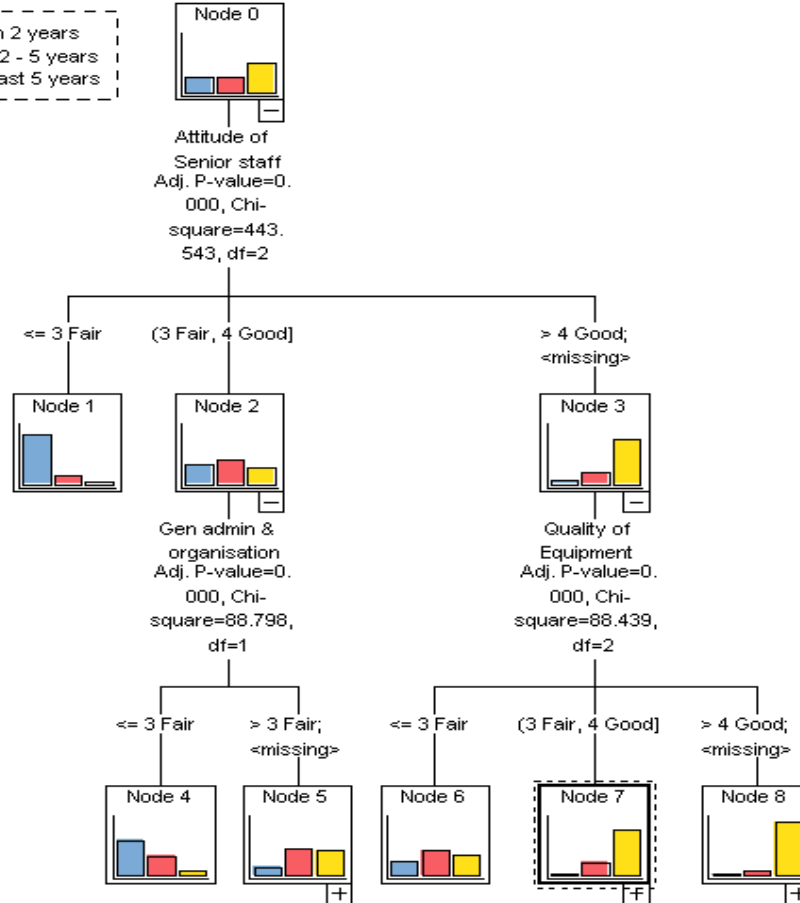


Let's see a demonstration...

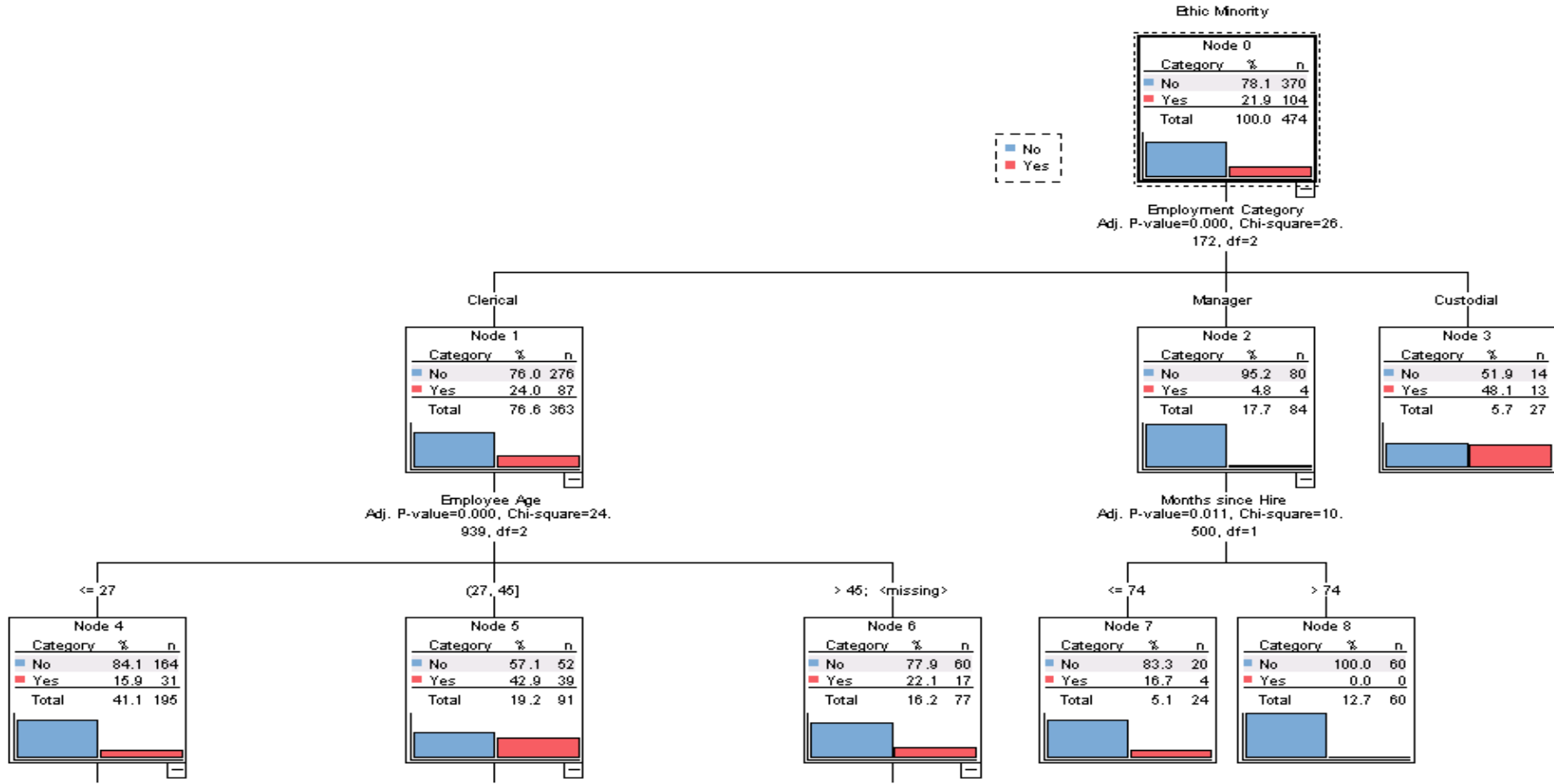
Drivers of Satisfaction

When will
employee leave

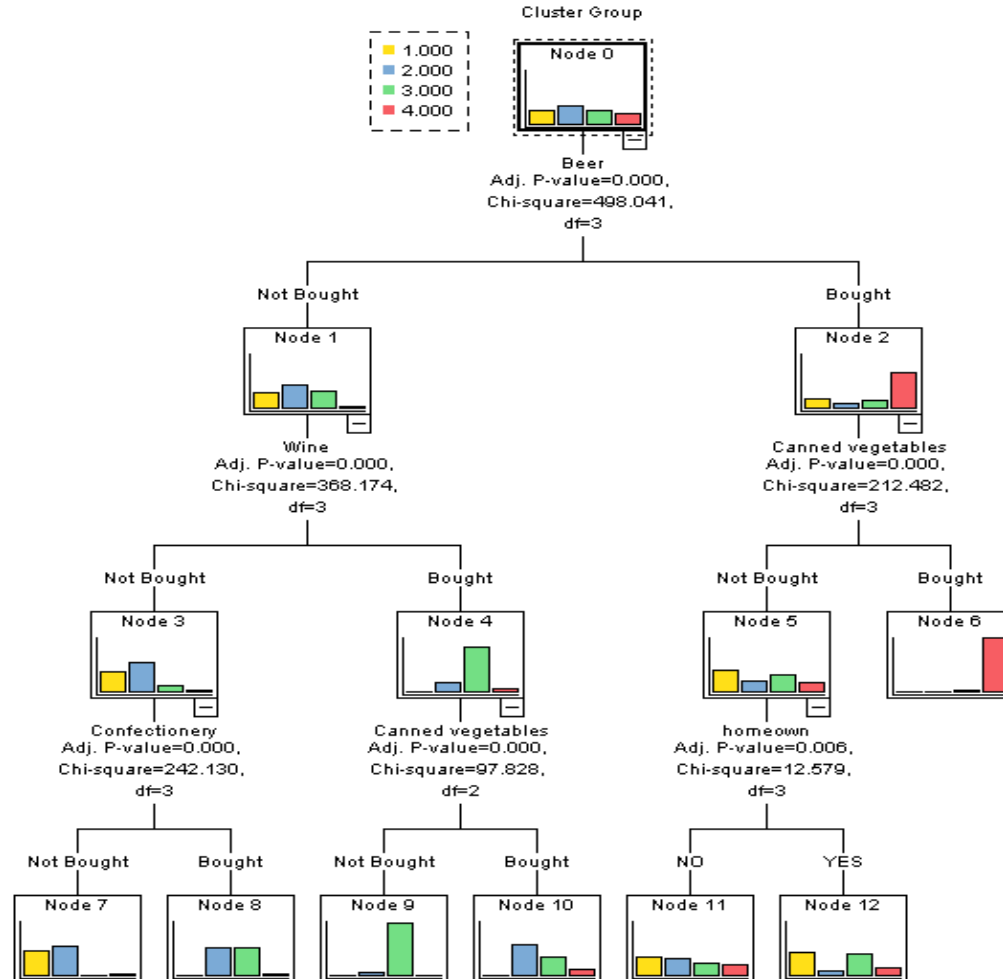
- 1 Leave within 2 years
- 2 Leave after 2 - 5 years
- 3 Remain at least 5 years



Demographic Profiling



Interpreting Clusters





Before we go



ChatSPSS is a free generative AI chatbot developed by Smart Vision Europe Ltd based on official IBM documentation of SPSS procedures

The screenshot shows the ChatSPSS website interface. At the top, there's a navigation bar with 'NEW CONVERSATION', 'EXPORT PDF', 'LOG OUT', 'PRIVACY', and 'SMARTVISION'. The main content area is divided into two columns. The left column shows a list of conversations, with the first one selected: 'How can I recode a string variable called region to create a numeric equivalent?'. Below this is a 'DOCUMENT REFERENCES' section with two links: 'IBM_SPSS_Statistics_Algorithms.pdf' and 'IBM_SPSS_Statistics_Command_Syntax_Re...'. The right column shows the chat response, which includes a blue header with the question, a paragraph explaining the RECODE command, a section titled 'Steps to Recode the Variable' with a bulleted list of instructions, a section titled 'Example Syntax' with a code snippet and a bulleted list of notes, and a section titled 'Alternative Method' with a text input field and a submit button.

Chat SPSS

https://chatspss.com

NEW CONVERSATION EXPORT PDF LOG OUT PRIVACY SMARTVISION

CONVERSATIONS

How can I recode a string variable called region to create a numeric equivalent?

How can I recode a string variable called region to create a numeric equivalent?

How can I recode a string variable called region to create a numeric equivalent?

To recode a string variable called region to create a numeric equivalent, you can use the RECODE command with the INTO keyword.

How can I recode a string variable called region to create a numeric equivalent?

Steps to Recode the Variable

- Use the RECODE command to specify the original values and the new values.
- Use the INTO keyword to specify the target numeric variable.
- Ensure that the target variable is declared before using it.

Example Syntax

```
RECODE region ('North'=1) ('South'=2) ('East'=3) ('West'=4) INTO region_num.
```

- In this example, the string variable region is recoded into the numeric variable region_num.
- The original values 'North', 'South', 'East', and 'West' are replaced with the numeric values 1, 2, 3, and 4, respectively.

Alternative Method

Alternatively, you riable.

```
AUTORECODE VAI  
/INTO region_num  
/PRINT.
```

Working with Smart Vision Europe

- **Sourcing Software**
 - You can buy your analytical software from us often with discounts
 - <http://www.sv-europe.com/buy-spss-online/>
 - We can also help with the development of customised analytical tools
- **Training**
 - Formal classroom/virtual training
 - Online self-paced training resources
- **Advice and Support**
 - ‘No strings attached’ technical and business advice relating to analytics
 - Tracked technical support services around the IBM SPSS product line

Smart Vision Europe: Services and Expertise

We have decades of experience providing guidance, training and consultancy in the delivery of effective data science initiatives.



Contact us:

+44 (0)207 786 3568

info@sv-europe.com

[Follow us on Linked In](#) 

[Sign up for our Newsletter](#)

Thank you