



Understanding Characteristics of Caravan Insurance Policy Holders

BUDT 733
May 10, 2007

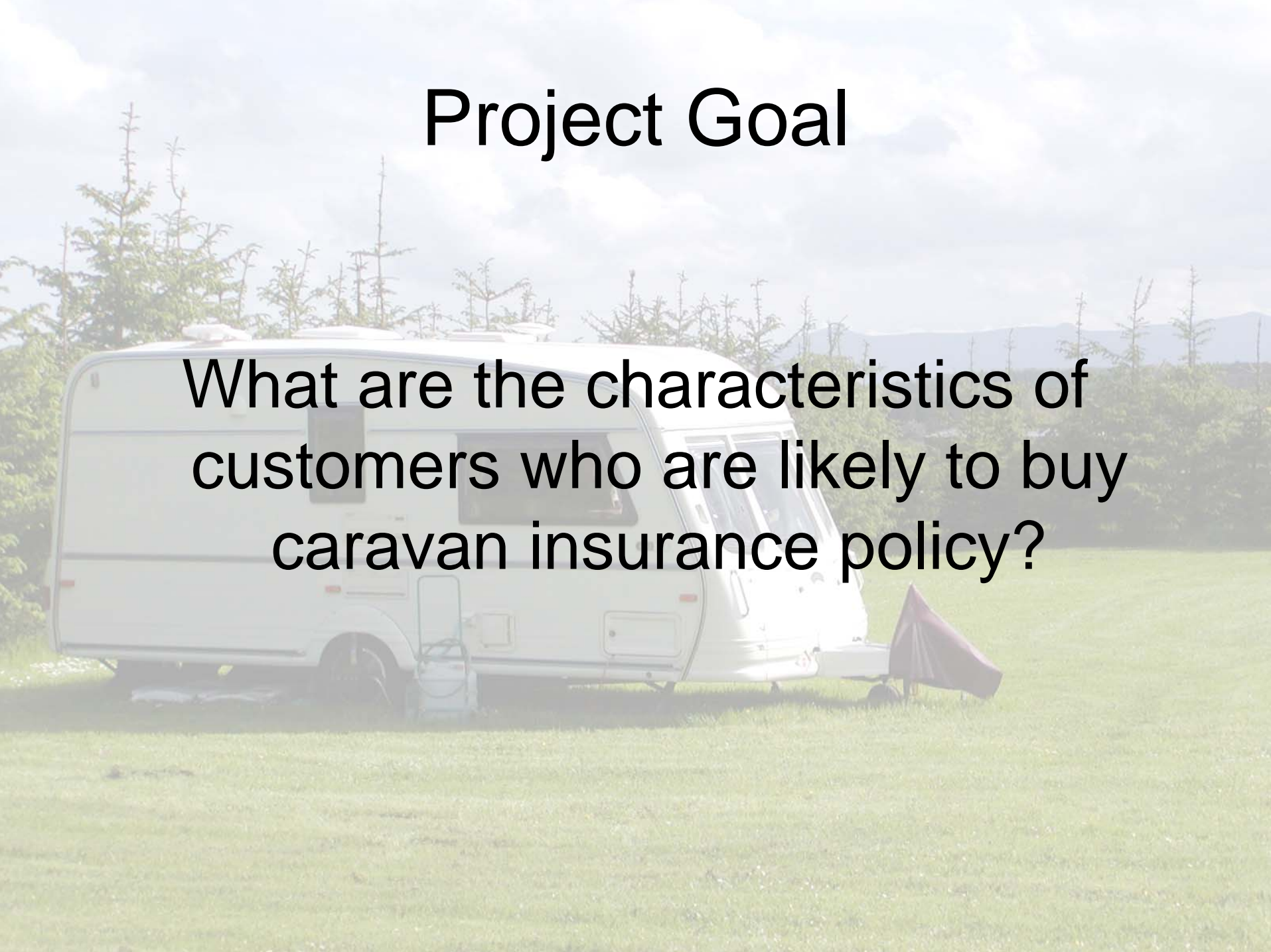
Group 5
Chih-Hau Huang
Masami Mabuchi
Muthita Songchitruksa
Nopakoon Visitrattakul

Agenda

- ④ Project Goal
- ④ Introduction
- ④ Data Source and Descriptions
- ④ Data Visualization
- ④ Data Pre-Processing
- ④ Models and Analysis
- ④ Findings
- ④ Recommendations
- ④ Q & A

Project Goal

What are the characteristics of customers who are likely to buy caravan insurance policy?



Introduction

@ What is a caravan?

🏠 A small trailer in which people can live and travel



Data Source & Description

- Supplied by the Dutch data mining company Sentient Machine Research

<http://kdd.ics.uci.edu/databases/tic/tic.html>

- 5,822 customer records

- Buyer: 5.98% (Success Class)
- Non-buyer: 94.02%

- 86 attributes

- 43 socio-demographic variables

- Derived from zip codes; customers in the same zip code have the same socio-demographic attributes

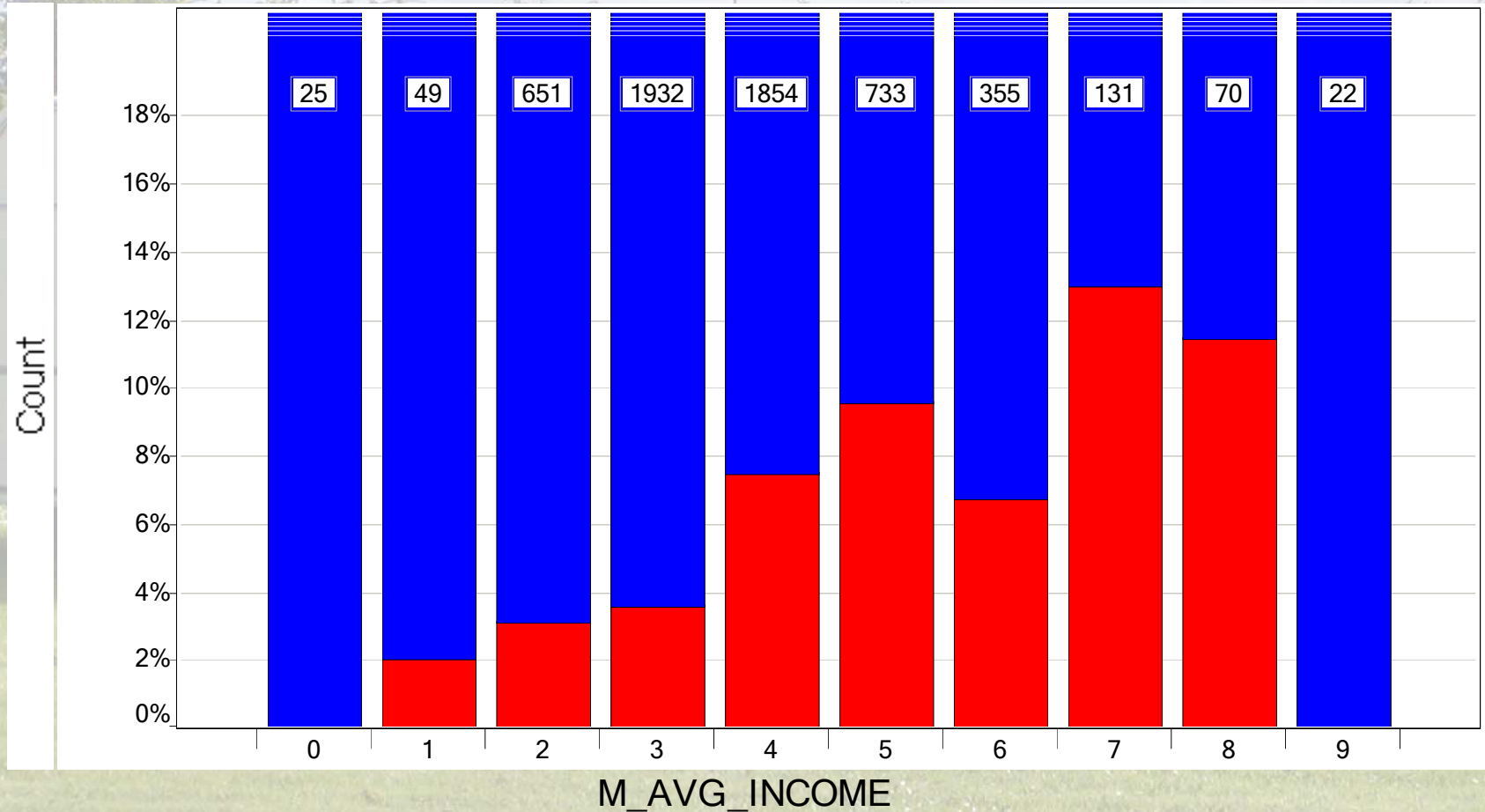
- 43 insurance-product-ownership variables

- # of specific insurance policies
- Contributions on corresponding policies

- Categorical (binned) and Numerical

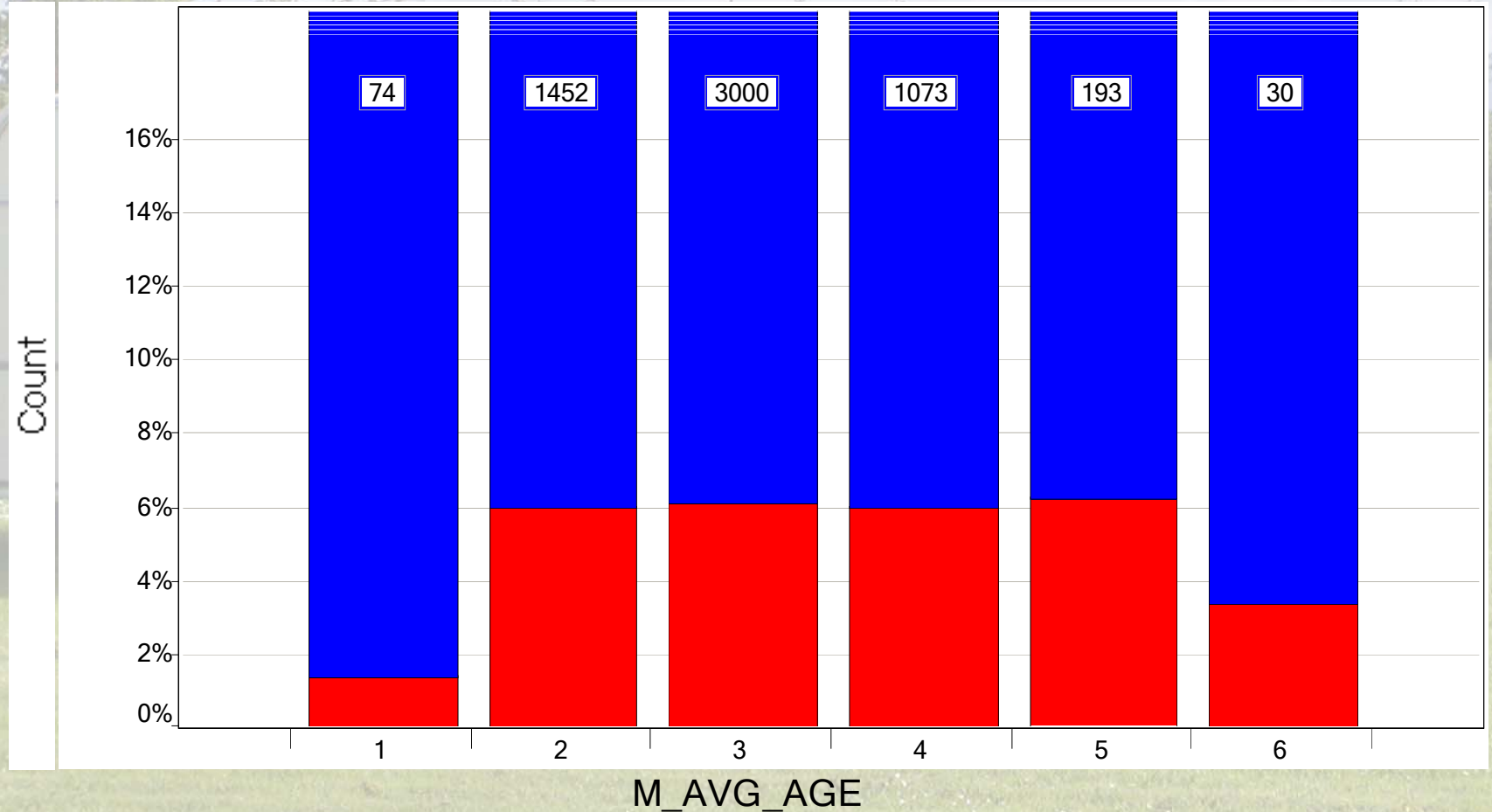
Data Visualization

Bar Chart



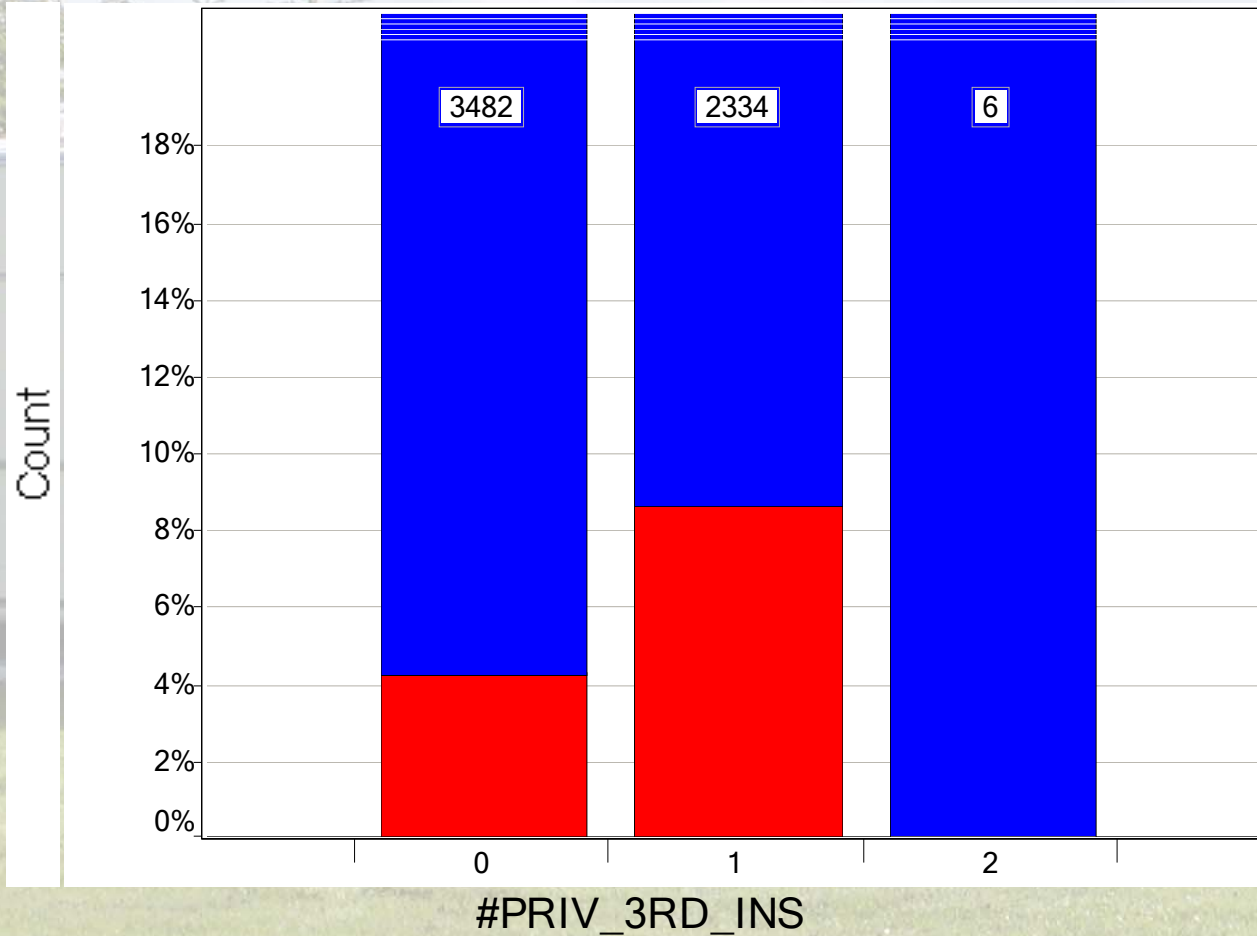
Data Visualization

Bar Chart



Data Visualization

Bar Chart



Data Pre-processing

- ④ Over-Sampling Training Set
- ④ Remove Variables
- ④ Create Derived Variables
- ④ Un-binning Data
- ④ Partition Data



Logistic Regression Model

Logistic Regression Model

Input variables	Coefficient	Std. Error	p-value	Odds
Constant term	-2.05754209	0.31352952	0	*
M_EDU_HIGH	0.17362288	0.05885206	0.00317612	1.1896069
M_AVG_INCOME	0.17481972	0.07740599	0.02391586	1.19103146
PRIV_3RD_INS	0.4663696	0.17033134	0.00618115	1.59419608
CAR_INS	1.32716846	0.18058152	0	3.77035236

Training Set

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buyer	Non-Buyer
Buyer	248	100
Non-Buyer	131	217

Validation Set

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buyer	Non-Buyer
Buyer	155	83
Non-Buyer	1455	2307

Error Report			
Class	# Cases	# Errors	% Error
Buyer	348	100	28.74
Non-Buyer	348	131	37.64
Overall	696	231	33.19

Error Report			
Class	# Cases	# Errors	% Error
Buyer	238	83	34.87
Non-Buyer	3762	1455	38.68
Overall	4000	1538	38.45

Findings

📍 Four Predictors:

📍 Car Insurance (most significant!)

📍 Private 3rd Party Insurance

📍 High Education

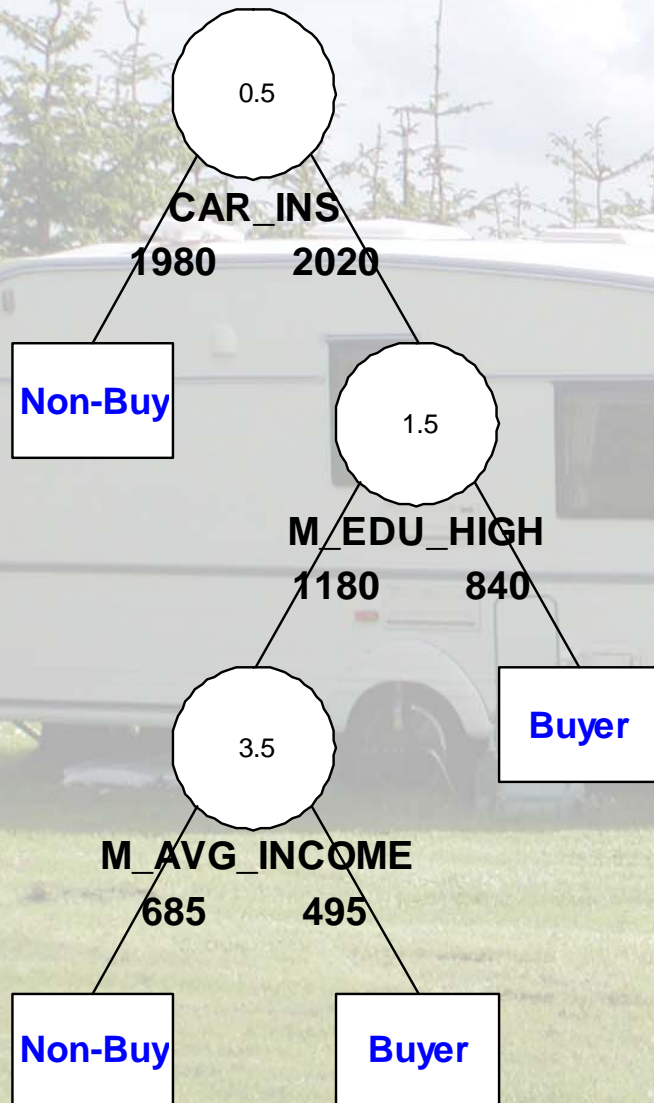
📍 Average Income

📍 Two Key Factors:

📍 Risk Aversion

📍 Wealth

Classification Tree (Best Pruned Tree)



- Input same set of predictors
- The result:
 - CAR_INS (car insurance or not)
 - M_EDU_HIGH (area with high education)
 - M_AVG_INCOME (area with average income)
- CAR_INS has the most explanatory power.
- These predictors were also chosen in Logistic Regression. !!!

Recommendations

- ④ Area-Focused Marketing
- ④ Advertising Campaign
- ④ Bundled Products
- ④ Joint Marketing with Caravan Makers

