Introduction to Decision Tree

- Also called objective segmentation
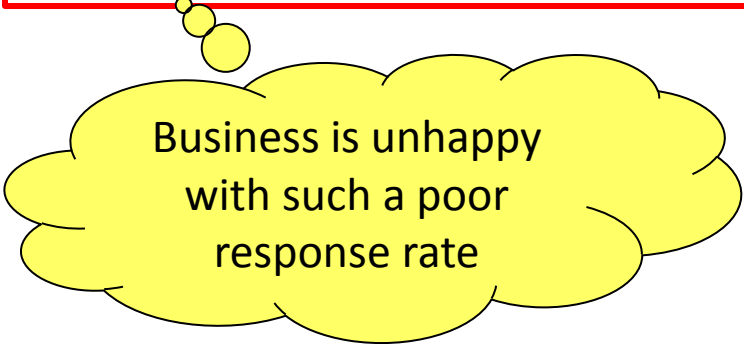
Agenda – Section 01

- Business scenario for categorical outcome
- See a sample decision tree – output
- Understand terms associated with the decision tree
- Understand the gains obtained from the decision tree
- Understand how it is different from logistic regression based scoring
- Understand what are the advantages of decision tree approach

Business Scenario – Need of a model

Business Scenario – need of a model?

- Say 100,000 prospect
- Say 1,000 takes up the product

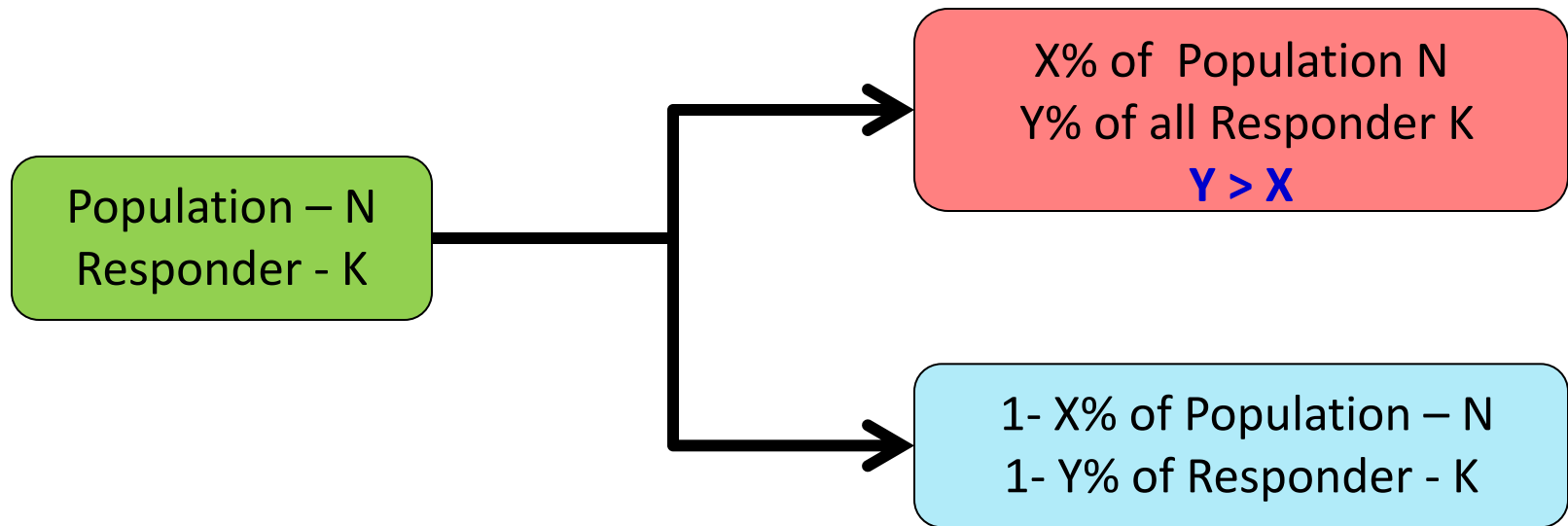*Business is unhappy with such a poor response rate*

- Think of – if $2 is the cost of mailer then one has spend $200 per new customer acquisition, right?
- Can we find a base where by working on less number of prospect, we can still get almost all the responder

- Say by working on 20000 prospect
- Can we get 900 responder

- Note – no possibility of exact match in real life scenarios
- Also very rare possibility of getting all the responder by working on part of population
- Target is to get **almost all the responder** by working on only **small portion of the population**

So the target is .....

- Target is to get almost all the responder by working on only part of the population

```
                                    ┌─────────────────────────────┐
                                    │  X% of  Population N         │
                                    │  Y% of all Responder K       │
                                    │  Y > X                       │
                                    └─────────────────────────────┘
┌─────────────────────┐
│ Population – N       │
│ Responder - K        │
└─────────────────────┘
                                    ┌─────────────────────────────┐
                                    │  1- X% of Population – N      │
                                    │  1- Y% of Responder - K       │
                                    └─────────────────────────────┘
```

- Note RGB concept –
  - ✓ Green the bench mark response rate
  - ✓ more response rate – **red**
  - ✓ Less response rate - **blue**
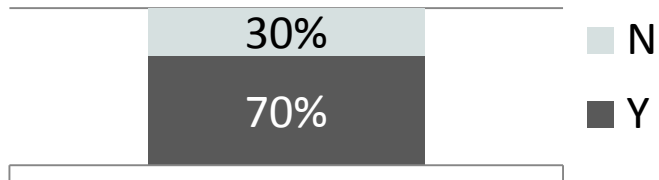- Work on red / blue– higher response / lower response rate section

Decision Tree Example– Understand the anatomy

Decision Tree Example

- Send files to bureau for credit worthiness of existing customers
- 70% gets good rating, 30% bad rating

30%	☐ N
70%	◼ Y

Credit Rating Y: Good, N: Bad

- Say $5 is the cost of sending each record for check to bureau
- Can we send records selectively to only those base where we have doubts
- Because ultimately, we want to stop loss and want to know, who will get bad rating hence risky

Decision Tree Example

- Can we forecast, among current population, who will Have good credit rating
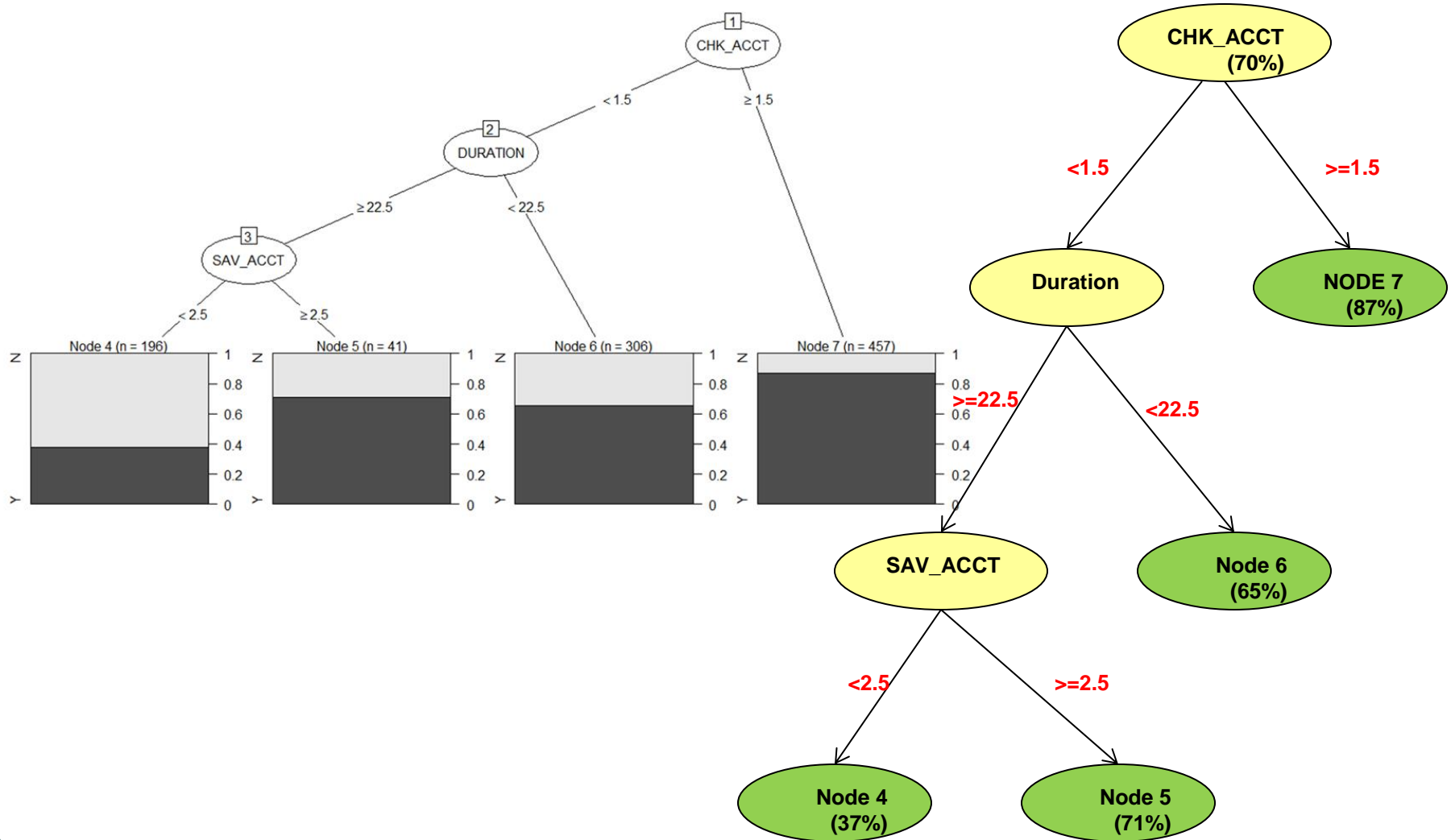- Decision tree improves the accuracy of decisioning

Credit Rating Y: Good, N: Bad

Root Note

CHK_ACCT < 1.5 and Duration >= 22.5 and SAV_ACCT < 2.5
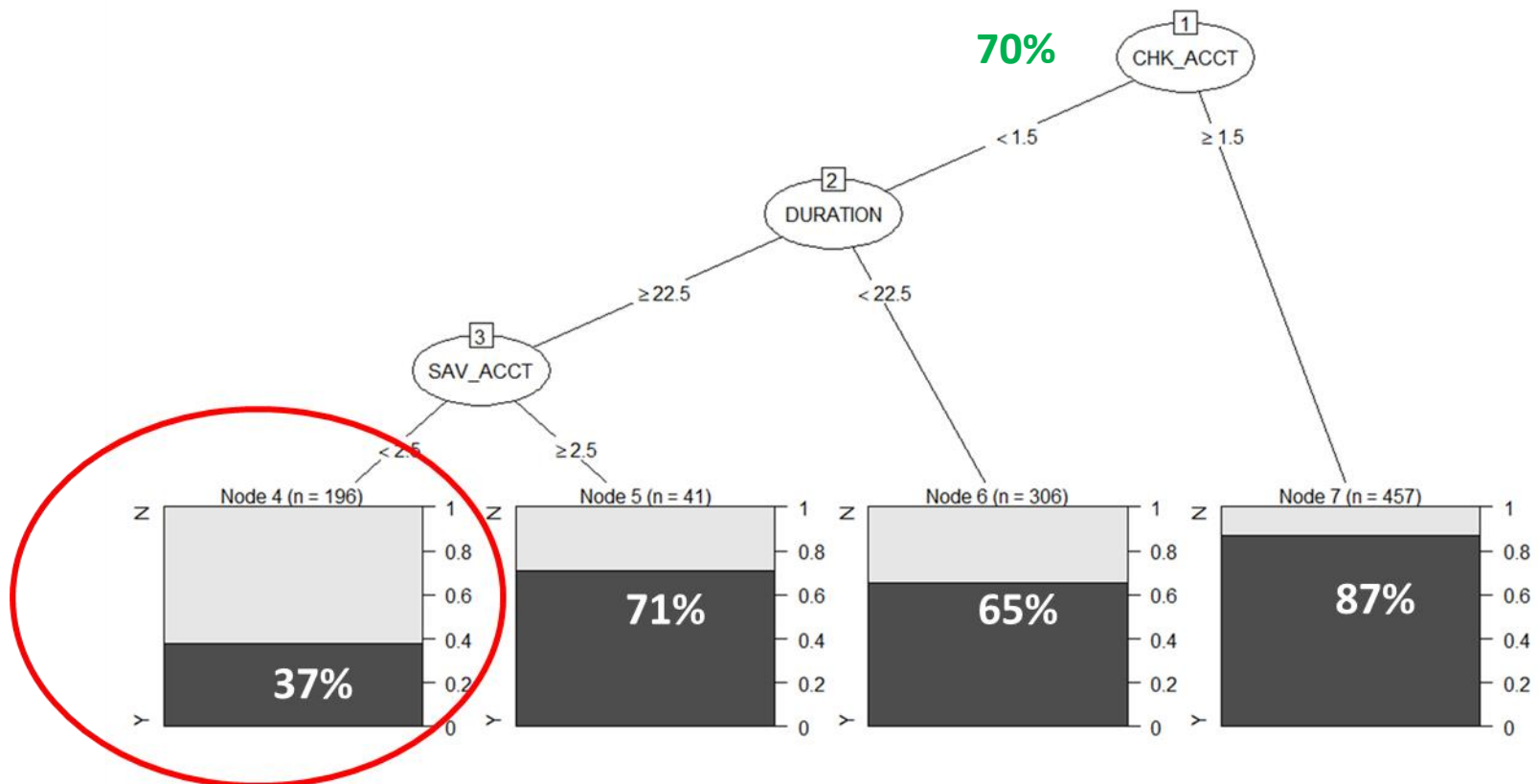
✓ Node Size
✓ Depth

Leaf Node
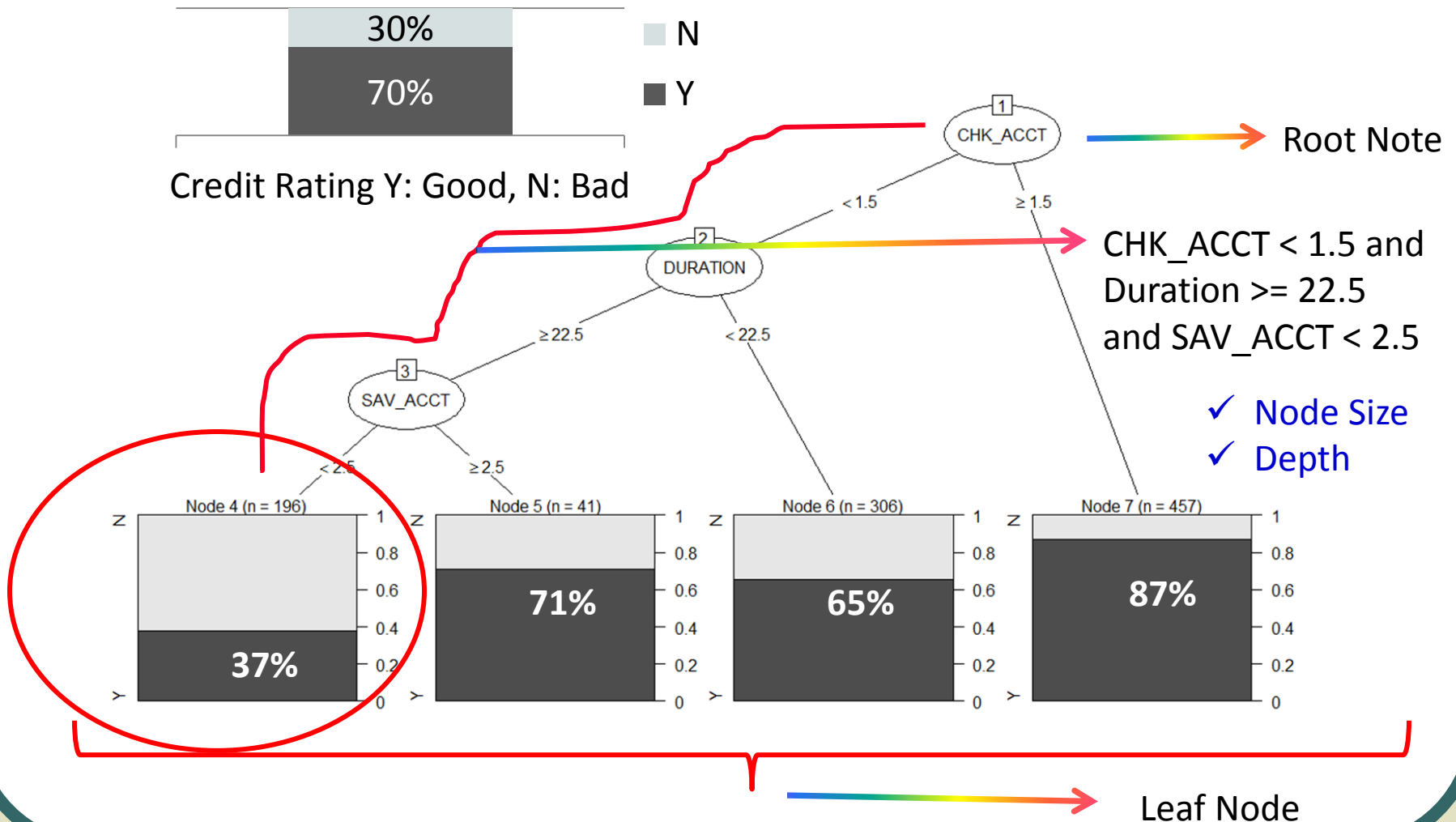
Decision Tree Example–
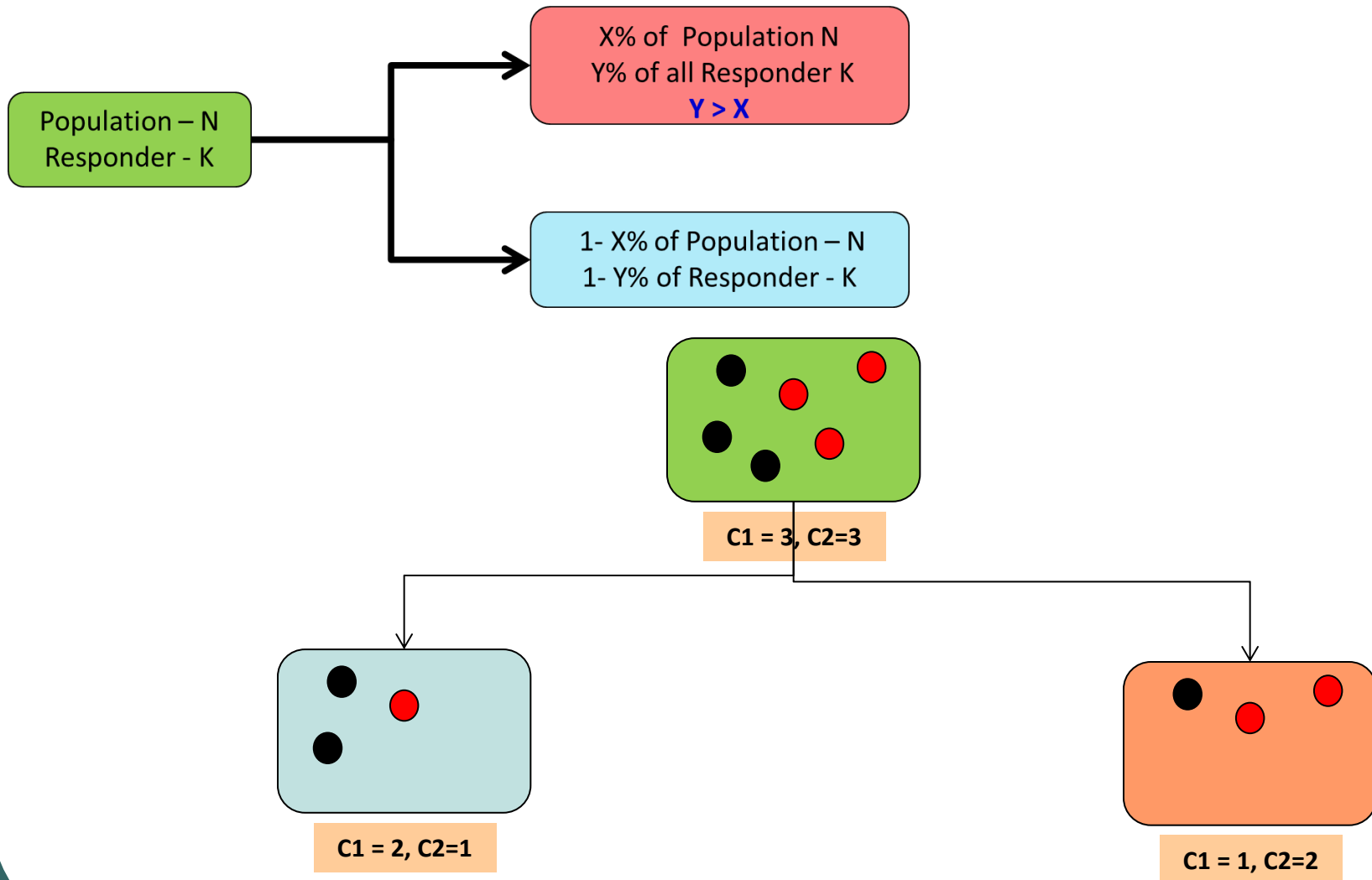**Understand the Gain from Decision Tree**

Decision Tree Example

Decision Tree Example

- ✓ Understand gain by working on different nodes
- ✓ Now we can keep a documentation cell to demand more document from a subset of population and then send them to bureau after receipt of documents.
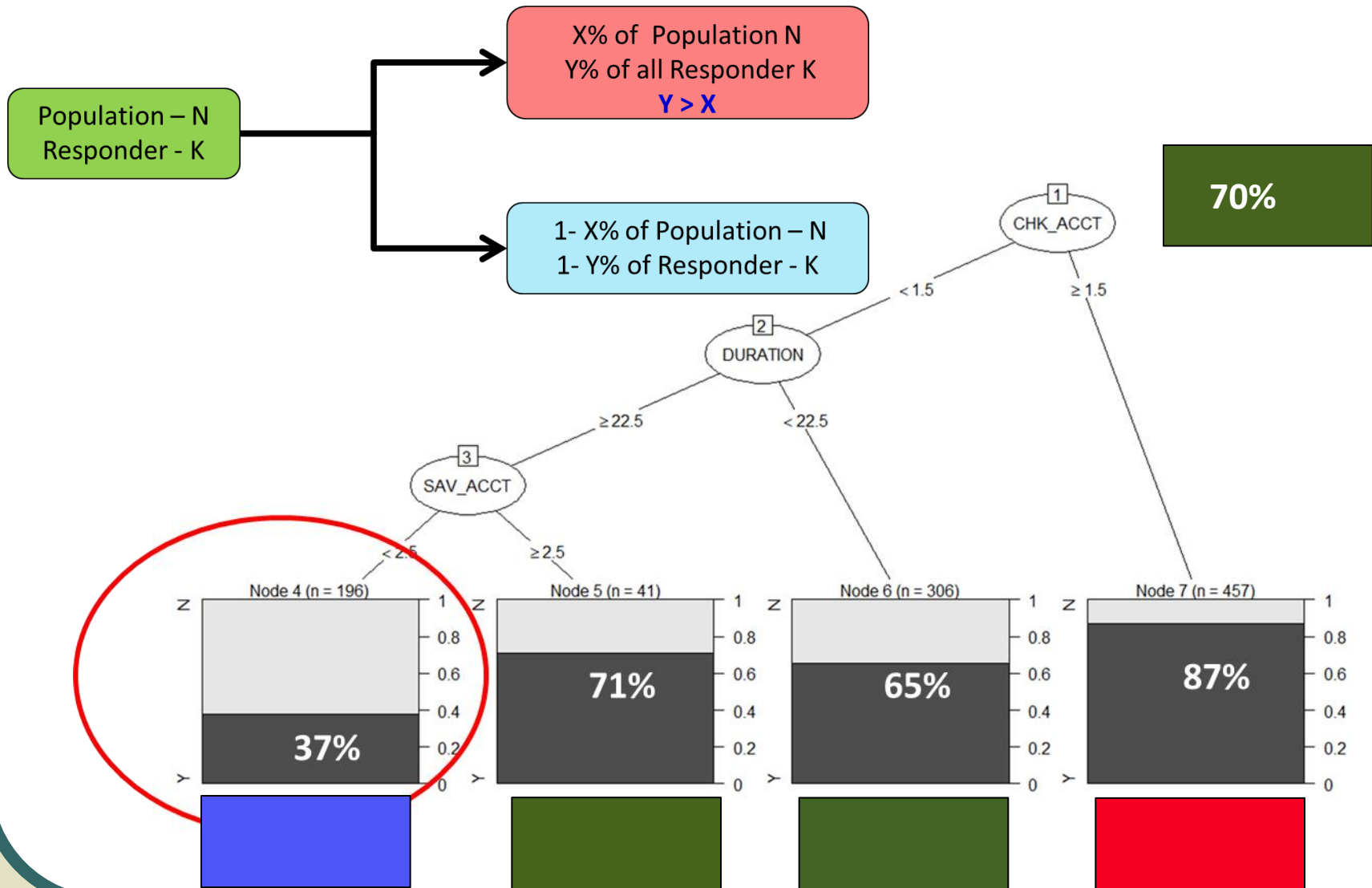
Decision Tree Example

✓ RGB Concepts

Population – N
Responder - K

X% of Population N
Y% of all Responder K
Y > X

1- X% of Population – N
1- Y% of Responder - K

C1 = 3, C2=3

C1 = 2, C2=1

C1 = 1, C2=2

Decision Tree Example

✓ RGB Concepts

Population – N
Responder - K

→ X% of Population N
Y% of all Responder K
Y > X

→ 1- X% of Population – N
1- Y% of Responder - K

1
CHK_ACCT

70%

< 1.5   ≥ 1.5

2
DURATION

≥ 22.5   < 22.5

3
SAV_ACCT

< 2.5   ≥ 2.5

Node 4 (n = 196)   Node 5 (n = 41)   Node 6 (n = 306)   Node 7 (n = 457)

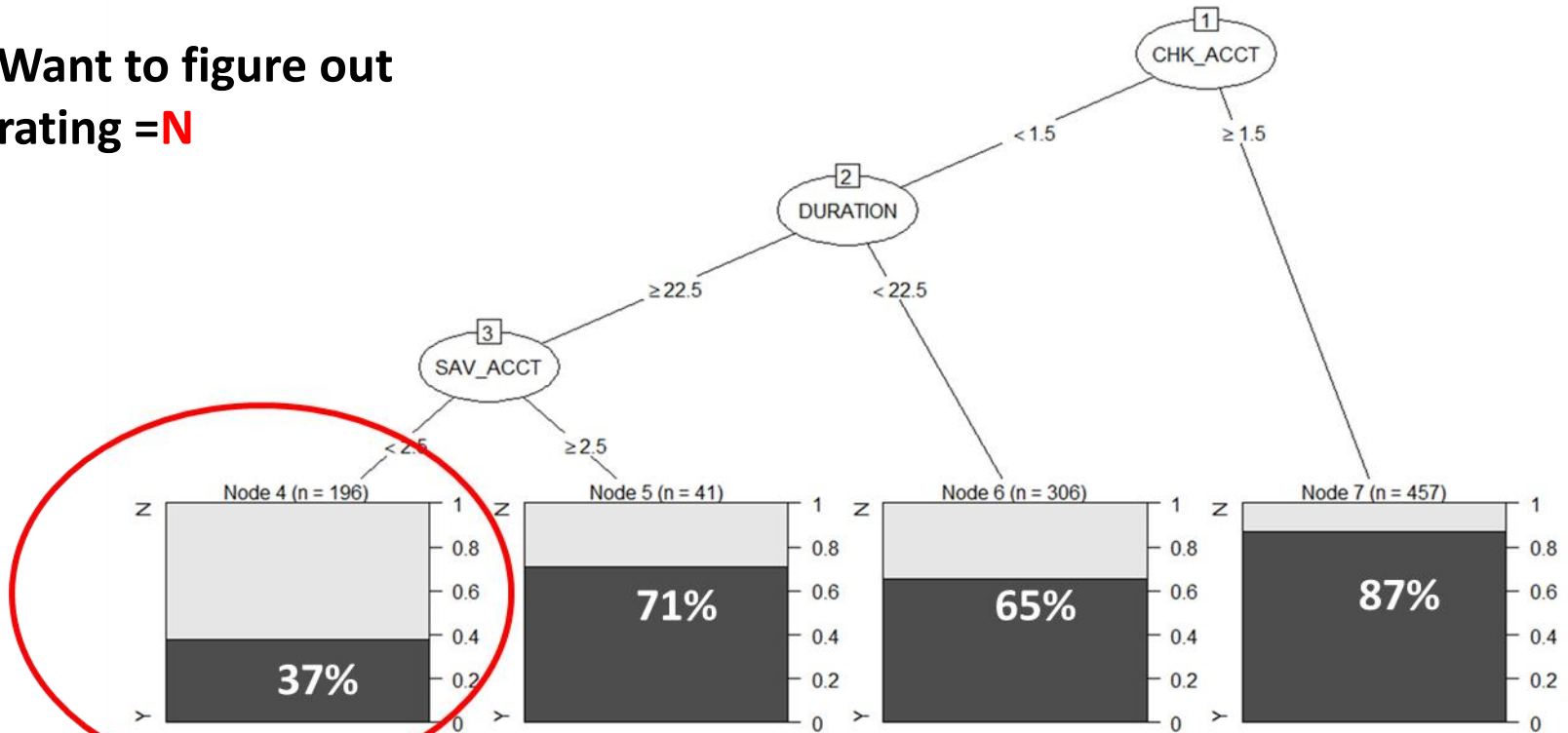37%   71%   65%   87%

KS of a decision Tree

KS of a decision tree

Want to figure out rating =N

| Segment	Response Rate (%)	Sample size	# of Responder	Cumulative Responders	Cumulative Size	Cumulative non responder	Cumulative non res %	Cumulative responder %	KS
4	63	196	123	123	196	73	10.4%	40.9%	30.5%
6	35	306	107	231	502	271	38.9%	76.4%	37.5%
5	29	41	12	242	543	301	43.0%	80.3%	37.3%
7	13	457	59	302	1,000	698	100.0%	100.0%	0.0%

Business Applications of a decision tree
– Use of a model

Business Examples – Scenario and Advantage

- Among prospect, Find who will default vs. non defaulter.
  - ✓ So by not giving loan to set of prospect, you avoid lots of bad loan.
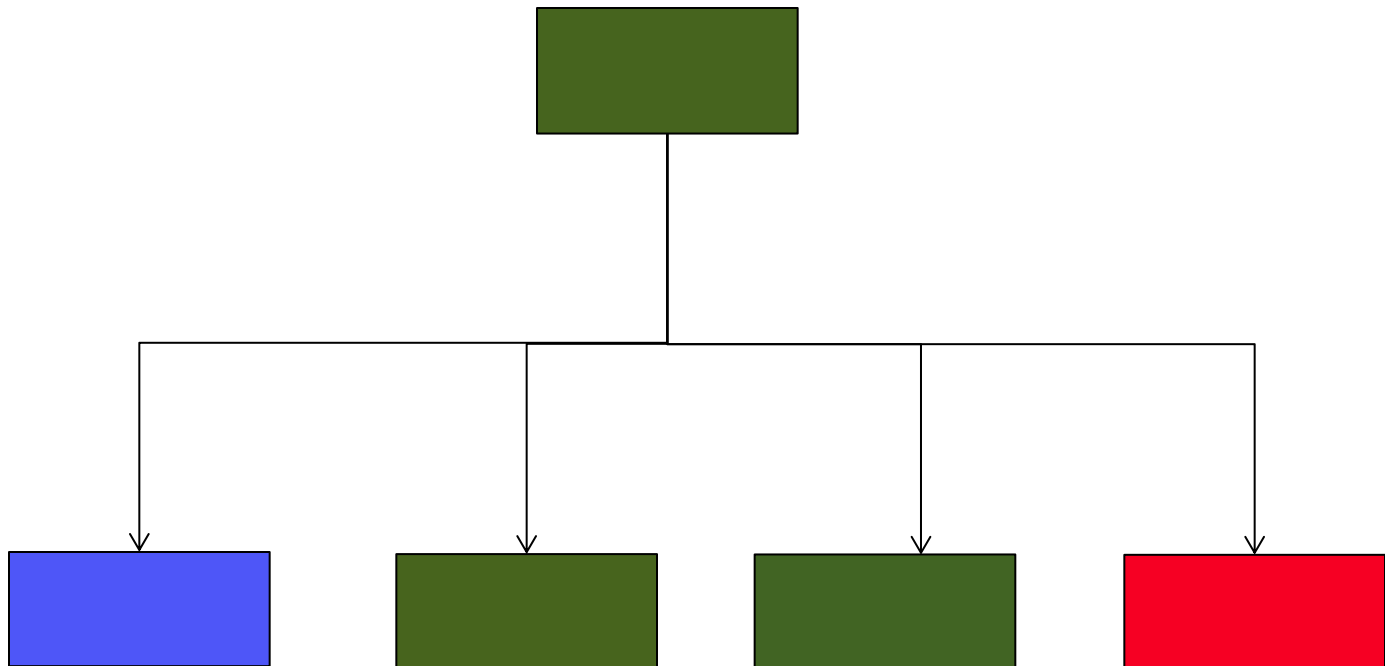
Business Examples – Scenario and Advantage

- Among patients profile, who will respond better with such treatment.
  - ✓ So by putting rest of them into another kind of treatment.

- Among customers, Find profile of those who will attrite vs. those will stay with the business.
  - ✓ So by targeting such customer you can reduce attrition?

- Among applicants, Find which are the applicants, who can be fraud (such as cases of account take over).
  - ✓ So by working on few selected applications you can avoid lots of account take over fraud cases.

- Among prospect of home loan pool, Find who are the prospects customer, who will switch over their home loan.
  - ✓ So by not working on few prospect, bank can quickly grow their portfolio by taking over existing home loans.

- Find who among current base will move into delinquency
  - ✓ So that their credit limit can be reduced to reduce exposure and losses.

Decision Tree related Definitions–

Definitions

- Objective segmentation -

  - **Homogenous within**: Groups, which has members as similar to each other as possible In terms of value of Dependent variable (Objective)

  - **Heterogeneous Across**: Members of one group are as dissimilar to member of others segments In terms of value of Dependent variable (Objective)

  - Popularly called **Objective** segmentation

  - The segmentation develops on how **independent variables** can predict the **dependent variable**.

  - There are several popular technique for the same. We will go through and understand following terms and methods

    - ✓ CHAID
    - ✓ Gini Index of Node and Split
    - ✓ Entropy
    - ✓ CART – CA for categorical outcome and RT for Numeric outcome

Decision Tree vs. Logistic Regression
And Key Decision Tree model development Features

Logistic regression Score vs. Decision Tree
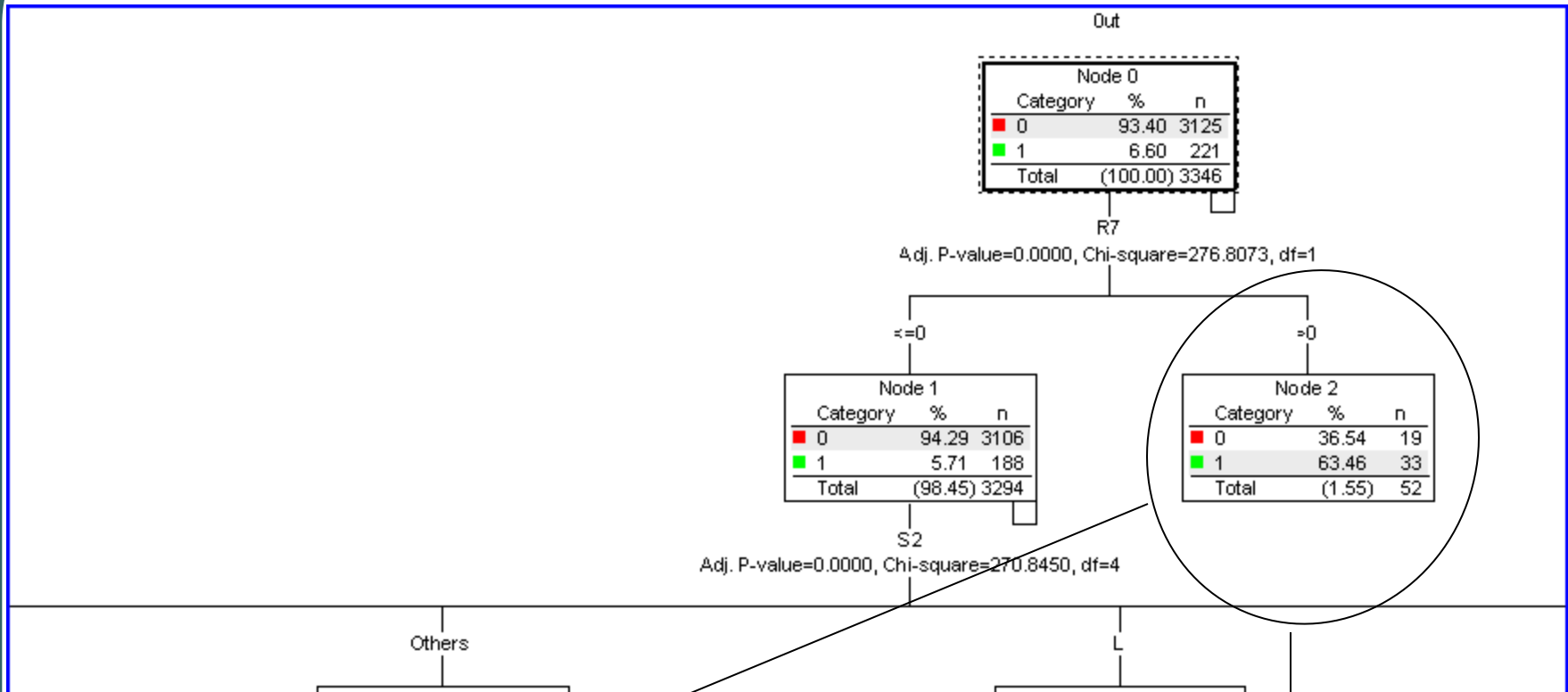
In segmentation

All 52 of one node are considered the same

Scoring methodology

1) Each account is given a score

2) Ranking among these, so more granular

3) Helps to make better selection

Key Decision Tree features

- **Automated field selection**
  - handles any number of fields
    - automatically selects relevant fields
- **No data preprocessing needed**
    - Does not require any kind of variable transforms
    - Impervious to outliers
- **Missing value tolerant**
    - Moderate loss of accuracy due to missing values
- **Quick development and validation**

Thanks