



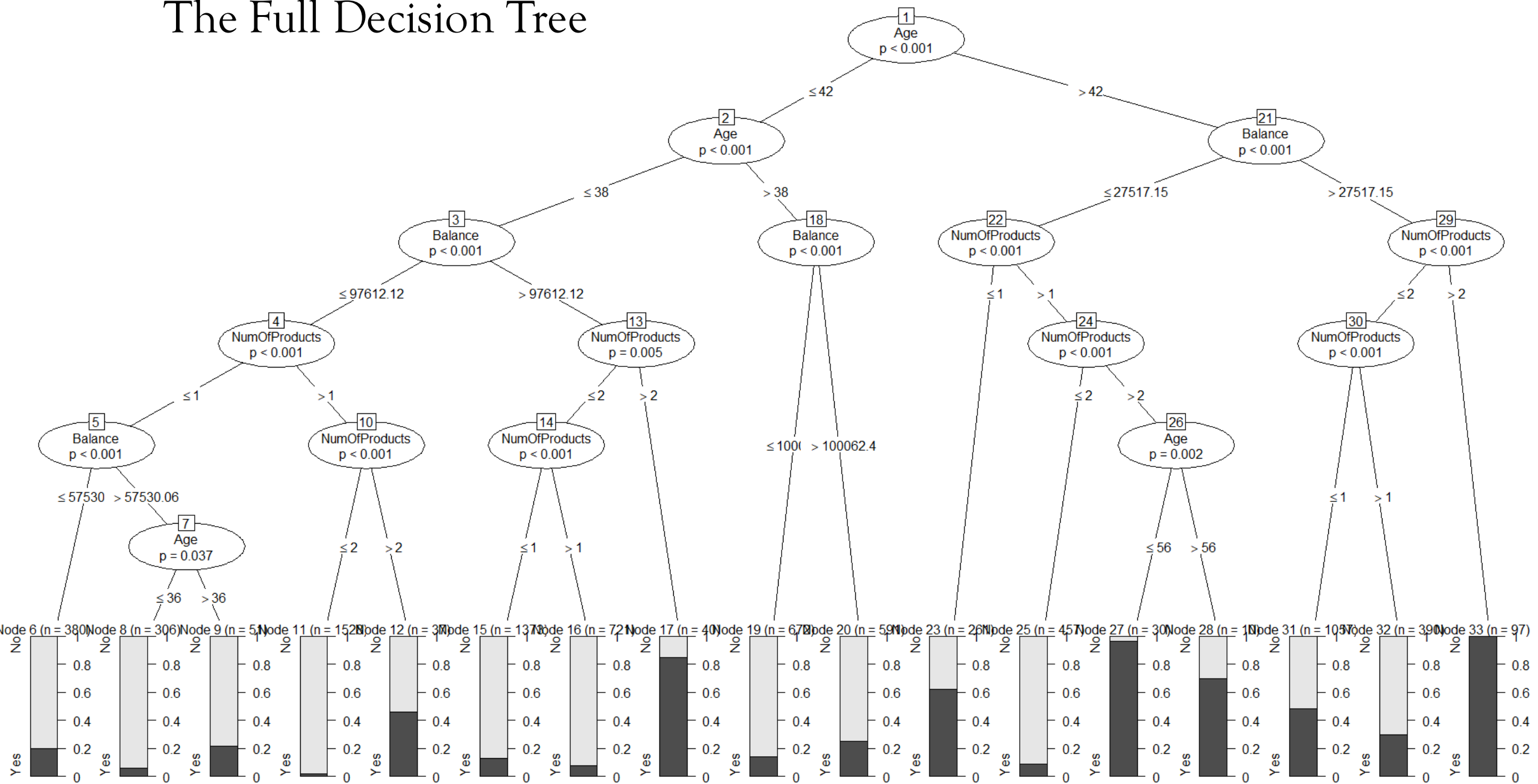
Bank Churn Analysis

Calvin Gerard

Business Question & Process

In this project I used [this](#) dataset from Kaggle to determine what factors can predict a bank's customer churn. In other words, the goal is to discover insights in order to advise the bank on how they can retain more customers. To answer this business question, I imported the dataset into R and checked for important factors and null values (for more details check the R Markdown document in the Github repository for this project [here](#)). Then I created decision trees based off of the factors that showed the highest impact on customer churn, which were the age of the customer, the customer's balance, and the number of financial products the customer purchased through the bank. Here is the decision tree using those factors:

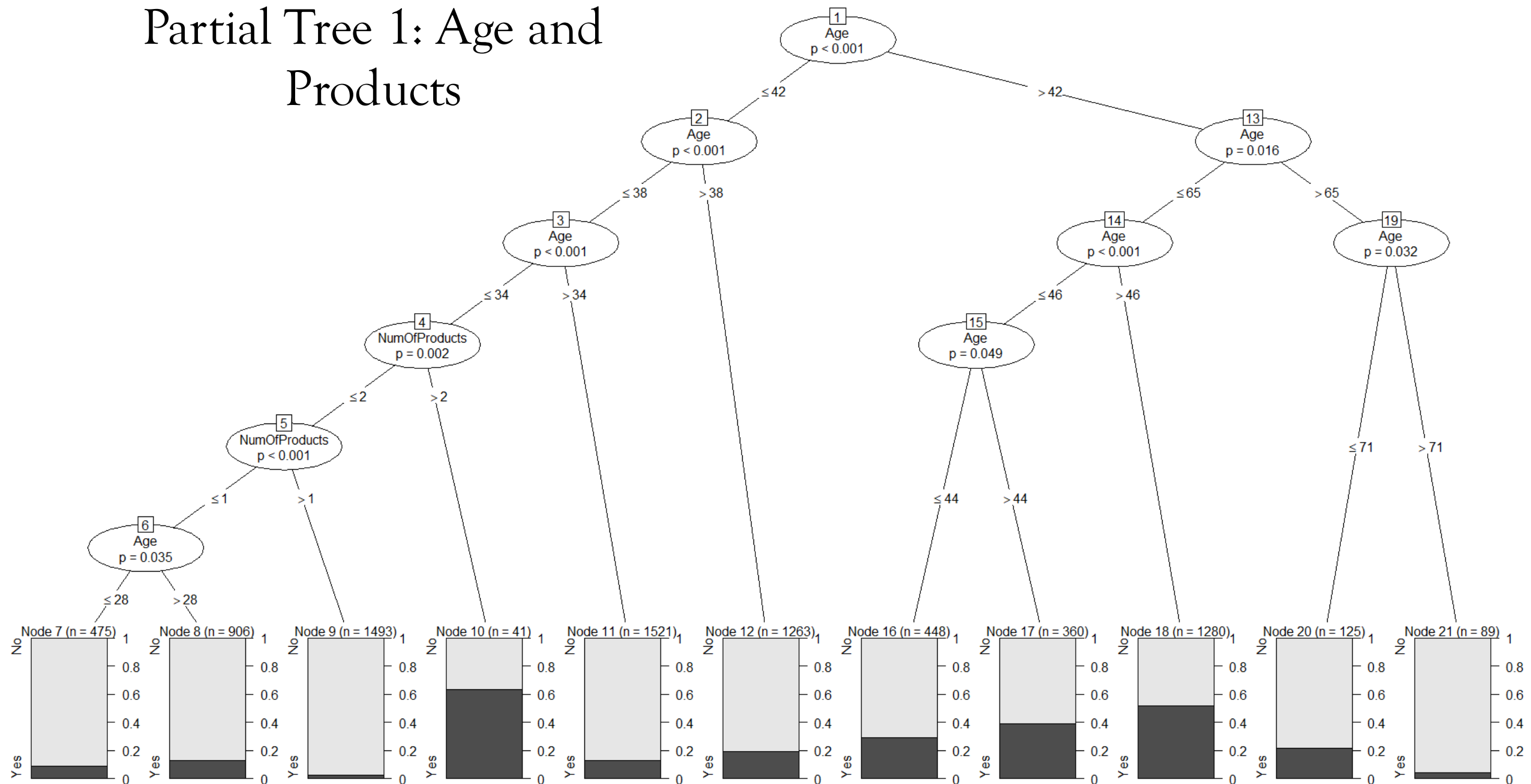
The Full Decision Tree



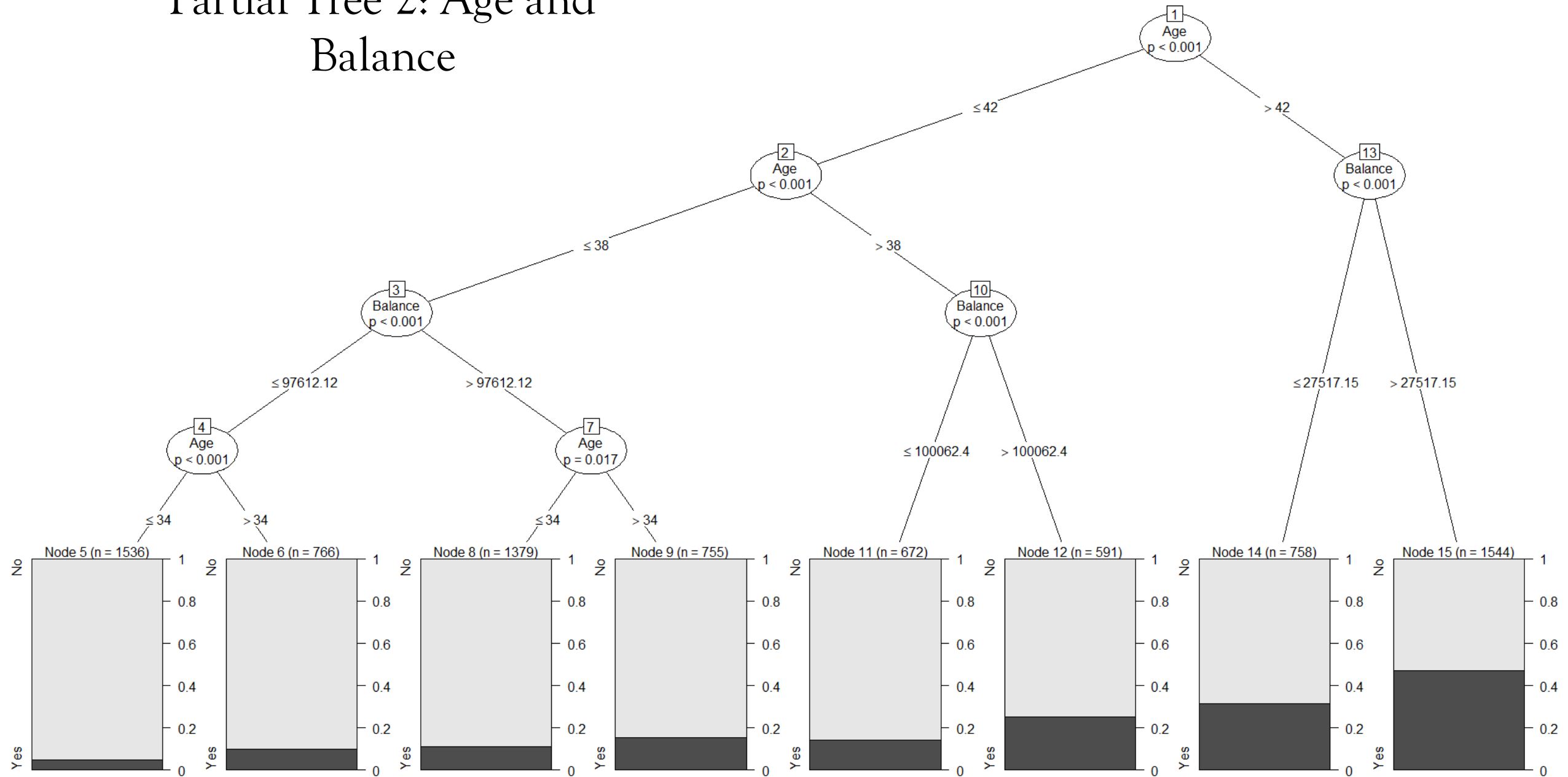
Trimming the Tree

That tree includes all of the most relevant factors that break down the different conditions that lead to high amounts of customer churn. The decision tree model shown on the last slide produced accuracy results of about 83% when running the prediction testing that's included in the R Markdown document. This result is good for our purposes, but the tree itself is slightly crowded and less readable as a result. To remedy this, I made two additional trees that include only two factors each, one tree for age and number of products bought and one tree for age and balance of the account.

Partial Tree 1: Age and Products



Partial Tree 2: Age and Balance



Trimming the Tree (cont.)

The two partial trees have accuracy results of about 79% and 80% respectively. Thus the bank could use either of these trees and produce predictive results that are nearly as accurate as the first, but their main goal is just to serve as more easily readable visualizations. The full tree is the most accurate predictive model of the three, but reading the nodes and gaining insights at a glance is not as easy as with the partial trees. Nevertheless, the recommendations I will be making will be based off the full tree since the other two are meant for the bank to present internally.

Findings

The decision tress show us that the top 5 categories of customers most likely to leave the bank are as follows:

- People with balances between ~27k and ~97k that have purchased more than two products
- People with balances less than ~27k age 56 years or younger that have purchased more than two products
- People with balances more than ~97k that have purchased more than two products
- People with balances less than ~27k age more than 56 years that have purchased more than two products
- People with balances less than ~27k that have purchased 1 or less products

Recommendations and Considerations

Based on my findings throughout the analysis, I would advise the bank to investigate why customers that buy more products are more likely to leave the bank, since that seems to be the most obvious contributing factor regardless of age or account balance. Some reasons for this could be that the products they're offering aren't producing enough value for the customers or another bank is outcompeting them. Additionally, a large contingent of customers that churn are those with account balances less than ~27k, so the bank should consider offering more products and/or benefits for customers in that range.

Improvements

In future churn analysis projects, depending on the dataset I would like to include more factors such as gender and geography that were included within this set. However, for this particular project I wanted to showcase decision trees within R and those qualitative factors don't work within that model. It would also be nice to have additional information that was not provided in this dataset, such as what the card type and points earned specifically mean for the customers' usage.