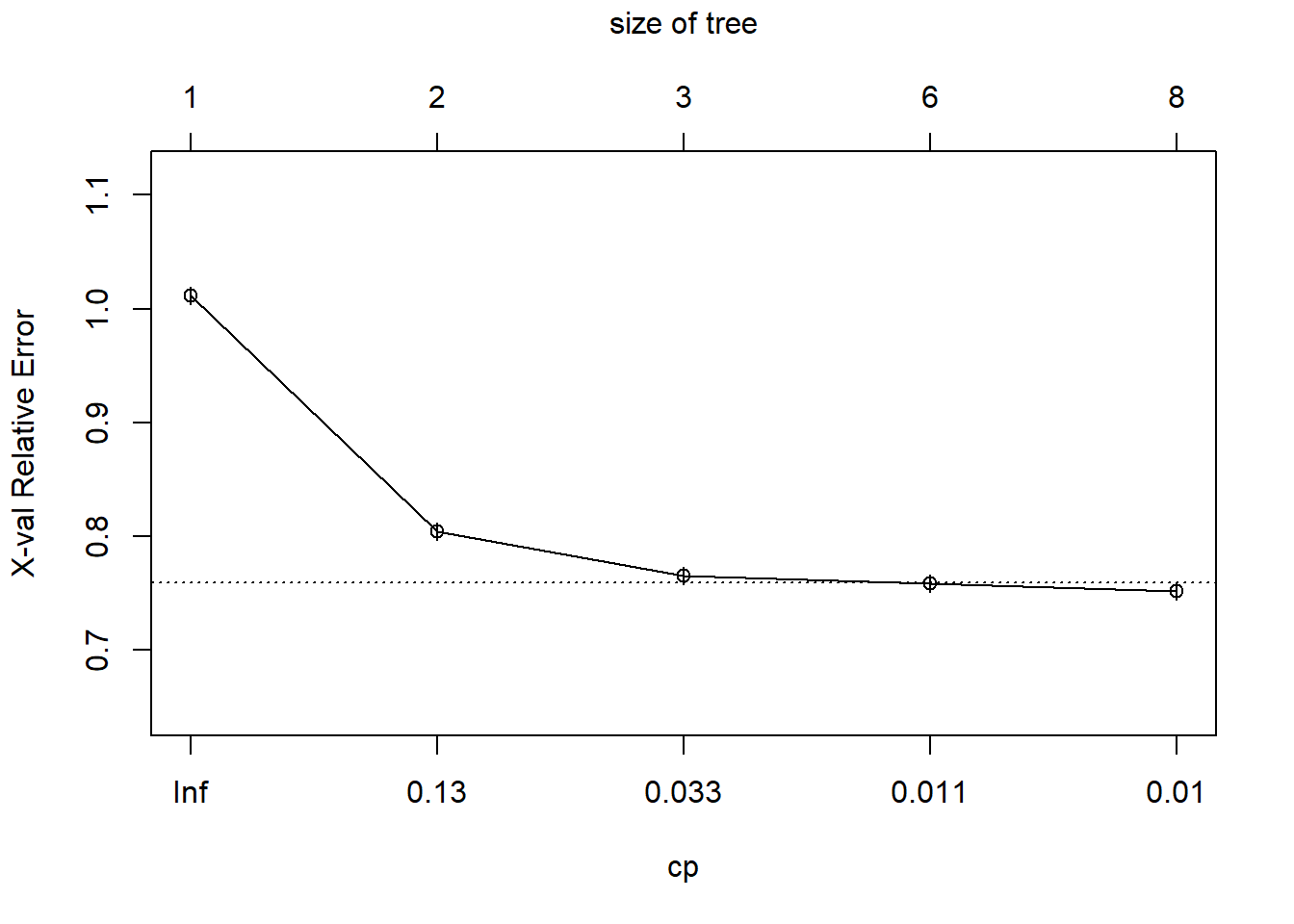
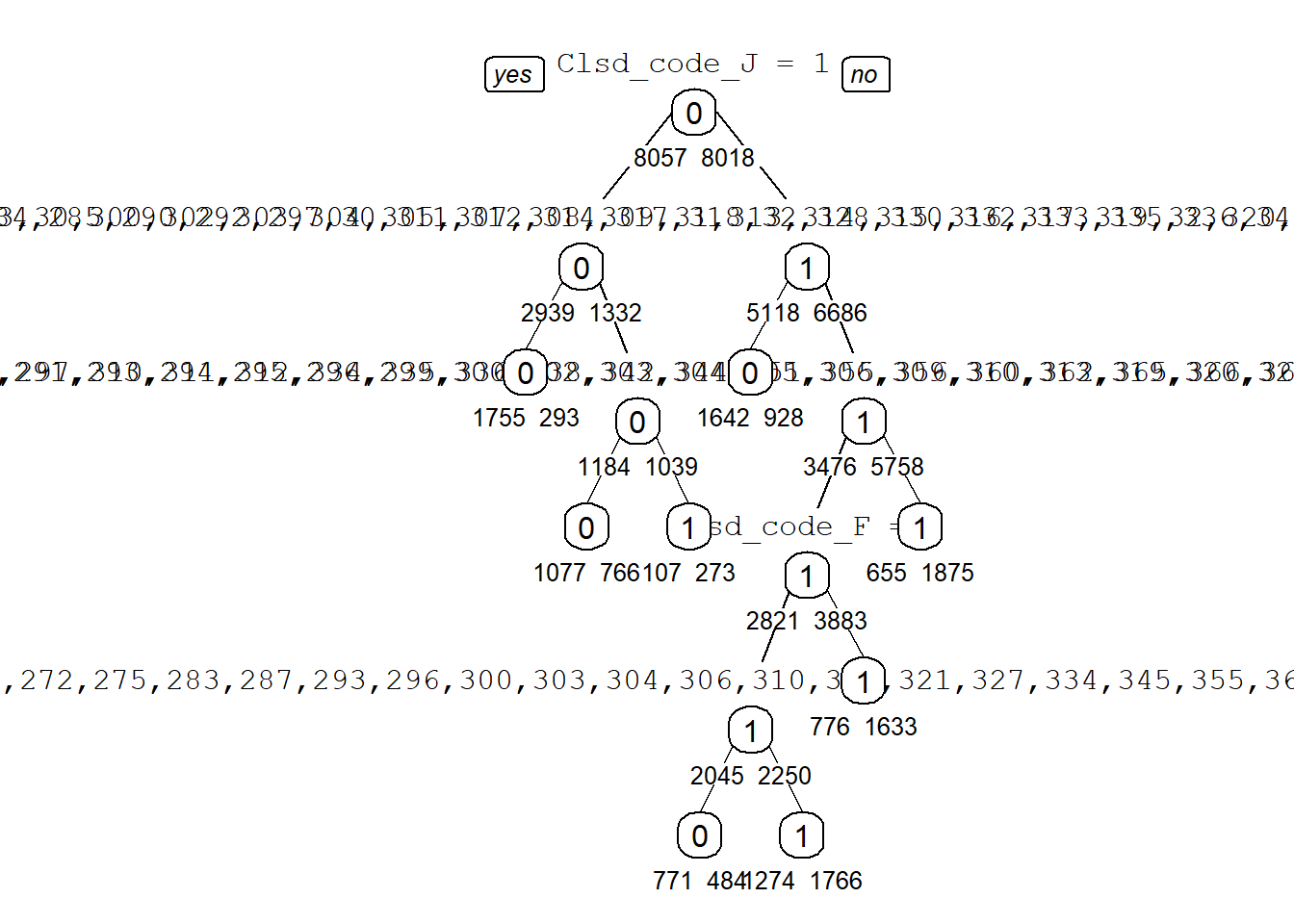
**The predictive analytics:**





(The picture is blurry as there are many factors involved the summary gives a better idea. They are in next sheets)

Decision tree outputs:

For training data:

## Confusion Matrix and Statistics

##

##

## pred 0 1

## 0 5245 2471

## 1 2812 5547

##

## Accuracy : 0.6714

## 95% CI : (0.664, 0.6786)

## No Information Rate : 0.5012

## P-Value [Acc > NIR] : < 0.00000000000000022

##

## Kappa : 0.3428

##

## Mcnemar's Test P-Value : 0.0000029

##

## Sensitivity : 0.6510

## Specificity : 0.6918

## Pos Pred Value : 0.6798

## Neg Pred Value : 0.6636

## Prevalence : 0.5012

## Detection Rate : 0.3263

## Detection Prevalence : 0.4800

## Balanced Accuracy : 0.6714

##

## 'Positive' Class : 0

##

For validation data after oversampling:

Confusion Matrix and Statistics

##

##

## pred 0 1

## 0 15274 8596

## 1 8962 15643

##

## Accuracy : 0.6378

## 95% CI : (0.6335, 0.6421)

## No Information Rate : 0.5

## P-Value [Acc > NIR] : < 0.00000000000000022

##

## Kappa : 0.2756

##

## Mcnemar's Test P-Value : 0.005877

##

## Sensitivity : 0.6302

## Specificity : 0.6454

## Pos Pred Value : 0.6399

## Neg Pred Value : 0.6358

## Prevalence : 0.5000

## Detection Rate : 0.3151

## Detection Prevalence : 0.4924

## Balanced Accuracy : 0.6378

##

## 'Positive' Class : 0

**The Logistic regression predicted that the areas in red are the most problematic areas.**

A picture containing text, map

Description automatically generated

Crime in terms of hours: (Based on model importance level)

**Areas of crucial importance: (In order)**

1.Multifamily homes

2.Garage/sheds

3.School

4.Church

5.Hospitals

6.financial Institution

7.Entertainment centers

8.Departmental stores

9.Yard

10. Parking lots

11.Playground

12.Public transit vehicle.