

# Housing Price

Atshaya Suresh

2024-01-30

```
options(repos = "http://cran.r-project.org")
```

```
hp.df <- read.csv("house price.csv")
head(hp.df)
```

```
##      X      Suburb      Address Rooms Type   Price Method SellerG      Date
## 1 1 Abbotsford    85 Turner St     2   h 1480000      S   Biggin 3/12/2016
## 2 2 Abbotsford   25 Bloomburg St    2   h 1035000      S   Biggin 4/02/2016
## 3 4 Abbotsford    5 Charles St     3   h 1465000     SP   Biggin 4/03/2017
## 4 5 Abbotsford  40 Federation La    3   h  850000     PI   Biggin 4/03/2017
## 5 6 Abbotsford    55a Park St     4   h 1600000     VB   Nelson 4/06/2016
## 6 10 Abbotsford  129 Charles St    2   h  941000      S   Jellis 7/05/2016
##      Distance Postcode Bedroom2 Bathroom Car Landsize BuildingArea YearBuilt
## 1         2.5      3067         2         1   1       202          NA        NA
## 2         2.5      3067         2         1   0       156          79       1900
## 3         2.5      3067         3         2   0       134         150       1900
## 4         2.5      3067         3         2   1        94          NA        NA
## 5         2.5      3067         3         1   2       120         142       2014
## 6         2.5      3067         2         1   0       181          NA        NA
##      CouncilArea Latitude Longitude      Regionname Propertycount
## 1         Yarra  -37.7996   144.9984 Northern Metropolitan      4019
## 2         Yarra  -37.8079   144.9934 Northern Metropolitan      4019
## 3         Yarra  -37.8093   144.9944 Northern Metropolitan      4019
## 4         Yarra  -37.7969   144.9969 Northern Metropolitan      4019
## 5         Yarra  -37.8072   144.9941 Northern Metropolitan      4019
## 6         Yarra  -37.8041   144.9953 Northern Metropolitan      4019
```

```
#removing missing values and replacing it with mean
```

```
mean_value <- mean(hp.df$Price, na.rm = TRUE)
```

```
# Impute missing values with the mean
```

```
hp.df$Price <- ifelse(is.na(hp.df$Price), mean_value, hp.df$Price)
```

```
install.packages("tidyverse")
```

```
## Installing package into 'C:/Users/Atshaya Suresh/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'tidyverse' successfully unpacked and MD5 sums checked
```

```

##
## The downloaded binary packages are in
## C:\Users\Atshaya Suresh\AppData\Local\Temp\RtmpsreNE1\downloaded_packages

install.packages("rpart")

## Installing package into 'C:/Users/Atshaya Suresh/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'rpart' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'rpart'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Atshaya
## Suresh\AppData\Local\R\win-library\4.3\00LOCK\rpart\libs\x64\rpart.dll to
## C:\Users\Atshaya
## Suresh\AppData\Local\R\win-library\4.3\rpart\libs\x64\rpart.dll: Permission
## denied

## Warning: restored 'rpart'

##
## The downloaded binary packages are in
## C:\Users\Atshaya Suresh\AppData\Local\Temp\RtmpsreNE1\downloaded_packages

install.packages("randomForest")

## Installing package into 'C:/Users/Atshaya Suresh/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'randomForest' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'randomForest'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Atshaya
## Suresh\AppData\Local\R\win-library\4.3\00LOCK\randomForest\libs\x64\randomForest.dll
## to C:\Users\Atshaya
## Suresh\AppData\Local\R\win-library\4.3\randomForest\libs\x64\randomForest.dll:
## Permission denied

## Warning: restored 'randomForest'

##
## The downloaded binary packages are in
## C:\Users\Atshaya Suresh\AppData\Local\Temp\RtmpsreNE1\downloaded_packages

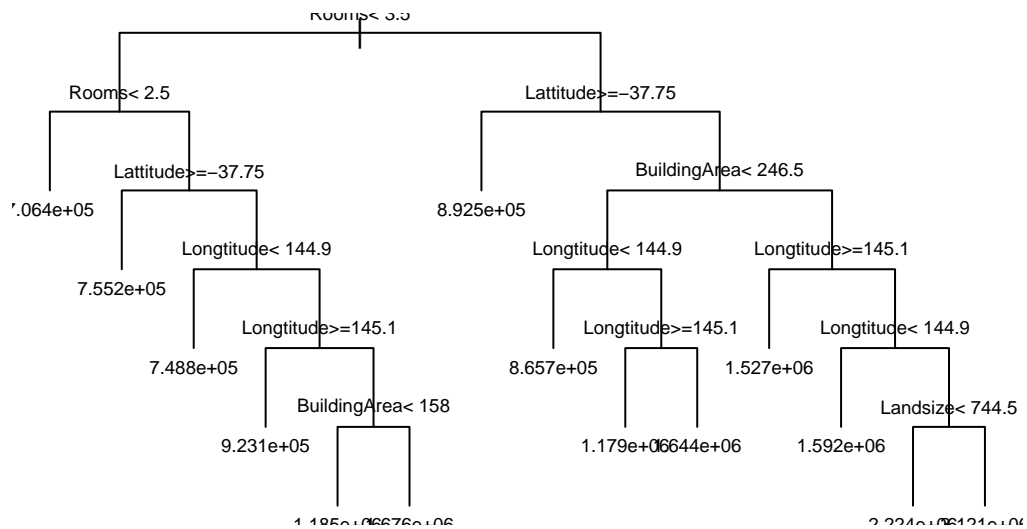
```

```
#Fitting a regression Tree to predict prices
library(rpart)
```

```
## Warning: package 'rpart' was built under R version 4.3.2
```

```
fit <- rpart(Price ~ Rooms + Bathroom + Landsize + BuildingArea + YearBuilt + Latitude + Longitude, data = hp)
```

```
plot(fit, uniform = TRUE)
text(fit, cex= 0.6)
```



```
head(hp.df)
```

```
##      X      Suburb      Address Rooms Type  Price Method SellerG      Date
## 1  1 Abbotsford  85 Turner St      2   h 1480000      S  Biggin 3/12/2016
## 2  2 Abbotsford 25 Bloomburg St      2   h 1035000      S  Biggin 4/02/2016
## 3  4 Abbotsford   5 Charles St      3   h 1465000     SP  Biggin 4/03/2017
## 4  5 Abbotsford 40 Federation La      3   h   850000     PI  Biggin 4/03/2017
## 5  6 Abbotsford   55a Park St      4   h 1600000     VB  Nelson 4/06/2016
## 6 10 Abbotsford 129 Charles St      2   h   941000      S  Jellis 7/05/2016
##      Distance Postcode Bedroom2 Bathroom Car Landsize BuildingArea YearBuilt
## 1         2.5      3067         2         1   1      202          NA         NA
## 2         2.5      3067         2         1   0      156           79        1900
## 3         2.5      3067         3         2   0      134          150        1900
```

```
## 4      2.5      3067      3      2      1      94      NA      NA
## 5      2.5      3067      3      1      2     120     142     2014
## 6      2.5      3067      2      1      0     181      NA      NA
##      CouncilArea Latitude Longitude      Regionname Propertycount
## 1      Yarra    -37.7996   144.9984 Northern Metropolitan      4019
## 2      Yarra    -37.8079   144.9934 Northern Metropolitan      4019
## 3      Yarra    -37.8093   144.9944 Northern Metropolitan      4019
## 4      Yarra    -37.7969   144.9969 Northern Metropolitan      4019
## 5      Yarra    -37.8072   144.9941 Northern Metropolitan      4019
## 6      Yarra    -37.8041   144.9953 Northern Metropolitan      4019
```

```
print(predict(fit, head(hp.df)))
```

```
##      1      2      3      4      5      6
## 706378.9 706378.9 1184840.0 1184840.0 1644287.4 706378.9
```

```
print(head(hp.df$Price))
```

```
## [1] 1480000 1035000 1465000 850000 1600000 941000
```

```
library(modelr)
mae(model = fit, data = hp.df)
```

```
## [1] 322001.3
```

```
splitData <- resample_partition(hp.df, c(test = 0.3, train= 0.7))
lapply(splitData, dim)
```

```
## $test
## [1] 5518 22
##
## $train
## [1] 12878 22
```

```
fit2 <- rpart(Price~ Rooms+ Bathroom+ Landsize+ BuildingArea+ YearBuilt+ Latitude+ Longitude, data =
mae(model = fit2, data = splitData$train)
```

```
## [1] 315033.5
```

```
get_mae <- function(maxdepth, target, predictors, training_data, testing_data)
{
  predictors <- paste(predictors, collapse="+")
  formula <- as.formula(paste(target, "~", predictors, sep = ""))
  model <- rpart(formula, data = training_data, control = rpart.control(maxdepth = maxdepth))
  mae <- mae(model, testing_data)
  return(mae)
}
```

```

target <- "Price"
predictors <- c("Rooms", "Bathroom", "Landsize", "BuildingArea", "YearBuilt", "Latitude", "Longitude")
for(i in 1:10)
{ mae <- get_mae(maxdept = i, target = target, predictors = predictors, training_data = splitData$train)
print(glue::glue("Maxdepth:" ,i, "\t MAE:",mae))}

```

```

## Maxdepth:1      MAE:416510.929518989
## Maxdepth:2      MAE:377530.880626343
## Maxdepth:3      MAE:349415.142668116
## Maxdepth:4      MAE:334611.963972307
## Maxdepth:5      MAE:331636.553502582
## Maxdepth:6      MAE:320681.392297585
## Maxdepth:7      MAE:320681.392297585
## Maxdepth:8      MAE:320681.392297585
## Maxdepth:9      MAE:320681.392297585
## Maxdepth:10     MAE:320681.392297585

```

The MAE drop when the Maxdepth of the Regression Tree is 6.