ADM Assignment 2

Hannah Cronin

2023-04-09

```
library(ISLR)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##      filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-7
```

```
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```
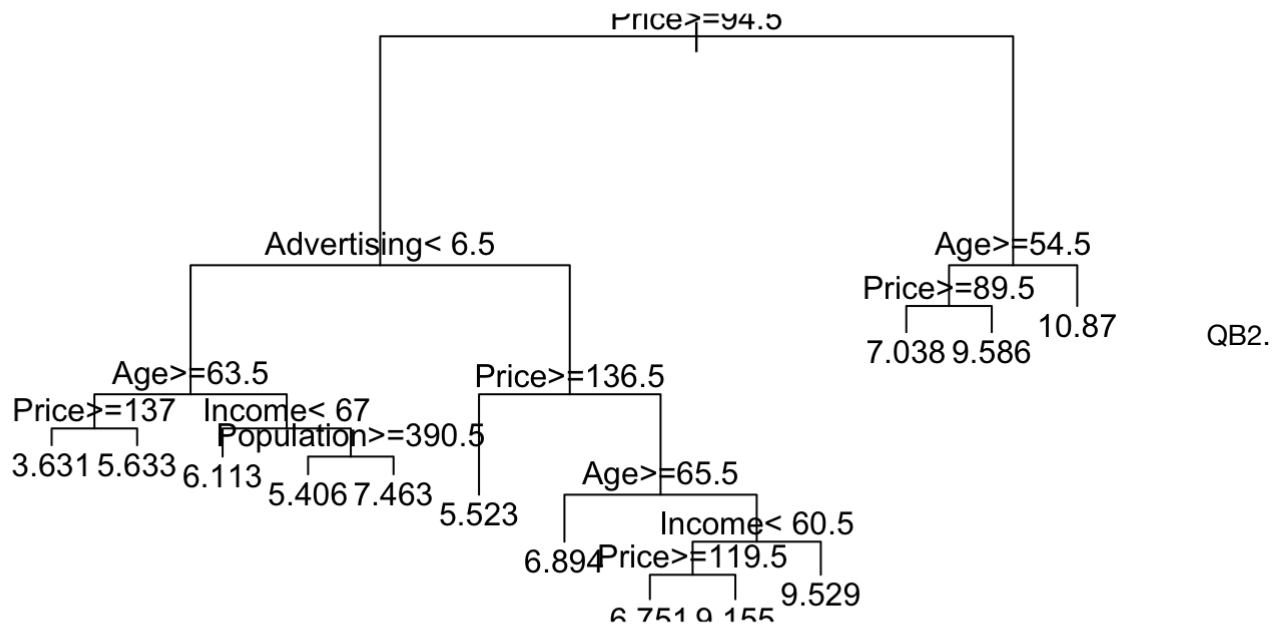
```
library(rpart)
library(rpart.plot)
library(rsample)
Carseats_Filtered <- Carseats %>% select("Sales", "Price",
"Advertising","Population","Age","Income","Education")
```

QB1.

```
set.seed(100)
Model_1 = rpart (Sales~., data = Carseats_Filtered, method = 'anova')
plot(Model_1)
text(Model_1)
```
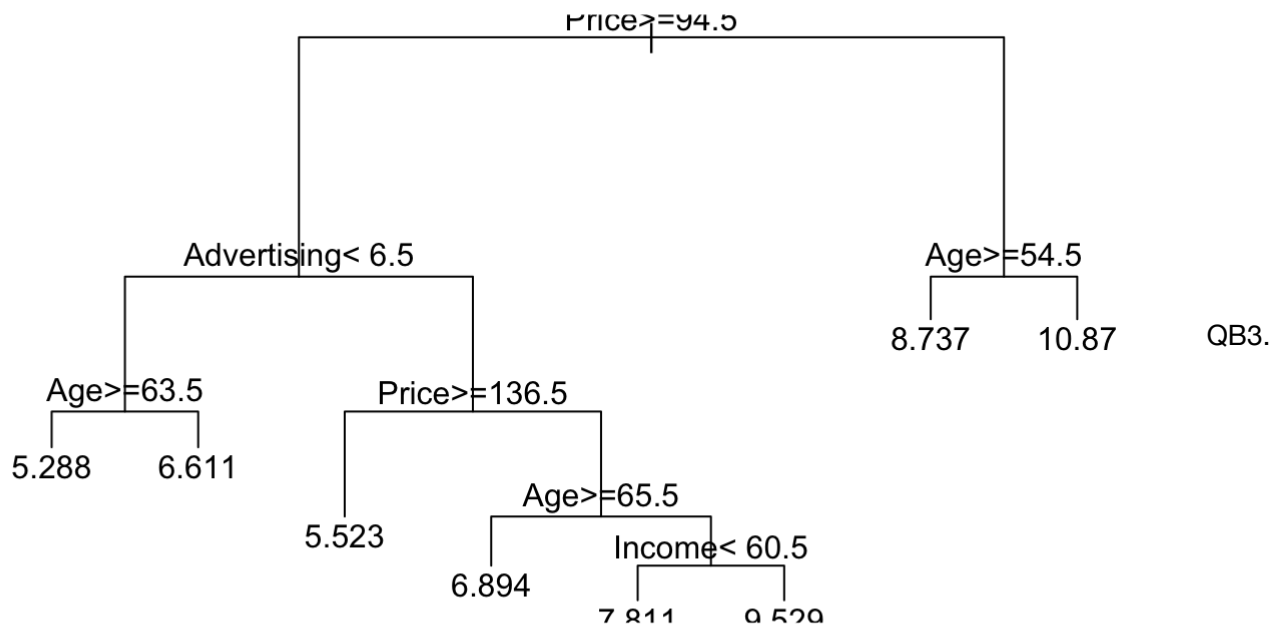
Price>=94.5

Advertising< 6.5

Age>=54.5

Price>=89.5

10.87

7.038 9.586

QB2.

Age>=63.5

Price>=137   Income< 67

Price>=136.5

3.631 5.633 6.113

Population>=390.5

5.406 7.463

5.523

Age>=65.5

Income< 60.5

6.894 Price>=119.5

9.529

6.751 9.155

```
Model_2 = rpart (Sales~., data = Carseats_Filtered, method = 'anova', control = rpart.co
ntrol(minsplit=60)) #making model less complex
plot(Model_2)
text(Model_2)
```

Price>=94.5

Advertising< 6.5   Age>=54.5

                                                   8.737   10.87   QB3.

Age>=63.5   Price>=136.5

5.288   6.611

                5.523   Age>=65.5

                        6.894   Income< 60.5

                                7.811   9.529

```
set.seed(123)
split = initial_split(Carseats_Filtered, prop = .7)
train = training(split)
test = testing(split)
Model_Caret_1 = train(Sales~., data = train, method = 'rf', trControl = trainControl(met
hod = "oob"))
print(Model_Caret_1)
```

```
## Random Forest
##
## 280 samples
##   6 predictor
##
## No pre-processing
## Resampling results across tuning parameters:
##
##   mtry  RMSE       Rsquared
##   2     2.414864   0.2668719
##   4     2.436562   0.2536383
##   6     2.450852   0.2448577
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 2.
```

QB4.

```
set.seed(123)
Model_Caret_2 = train(Sales~., data = train, method = 'rf', trControl = trainControl(met
hod = "repeatedcv", number = 5, repeats = 3), tuneGrid = expand.grid(mtry = 2:5))
print(Model_Caret_2)
```

```
## Random Forest
##
## 280 samples
##   6 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold, repeated 3 times)
## Summary of sample sizes: 224, 224, 224, 224, 224, 224, ...
## Resampling results across tuning parameters:
##
##   mtry  RMSE      Rsquared   MAE
##   2     2.428404  0.2676561  1.952151
##   3     2.427821  0.2686007  1.944892
##   4     2.435497  0.2662608  1.951003
##   5     2.447503  0.2614657  1.958052
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 3.
```