{Data Mining Assignment}

Assignment 1

Hadeer Elmahdy

20 March 2021

Contents

I. The Problem

SafeBabies is a large company who is producing car seats for babies and toddlers. They sell their products all over the US and abroad. The management team has hired you as a Business Analytics consultant to help them maximize their profit.

You have been told that the cost of producing each car seat is $55.0

Your primary task is to determine:

1- The optimal price for selling the car seats at those stores where the shelve location is "good" (i.e., the product is highly visible) 2- The optimal price for selling the car seats at those stores where the shelve location is "bad" (i.e., the product is poorly visible)

3- Plot the optimal price for selling the car seats at those stores where the shelve location is "good" and separately for those stores where the shelve location is "bad" when varying the production costs from $40 to $85 (in $5 increments).

```r
#Loading Liberaries
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
```

```r
library(ISLR)
library(ggplot2)
```

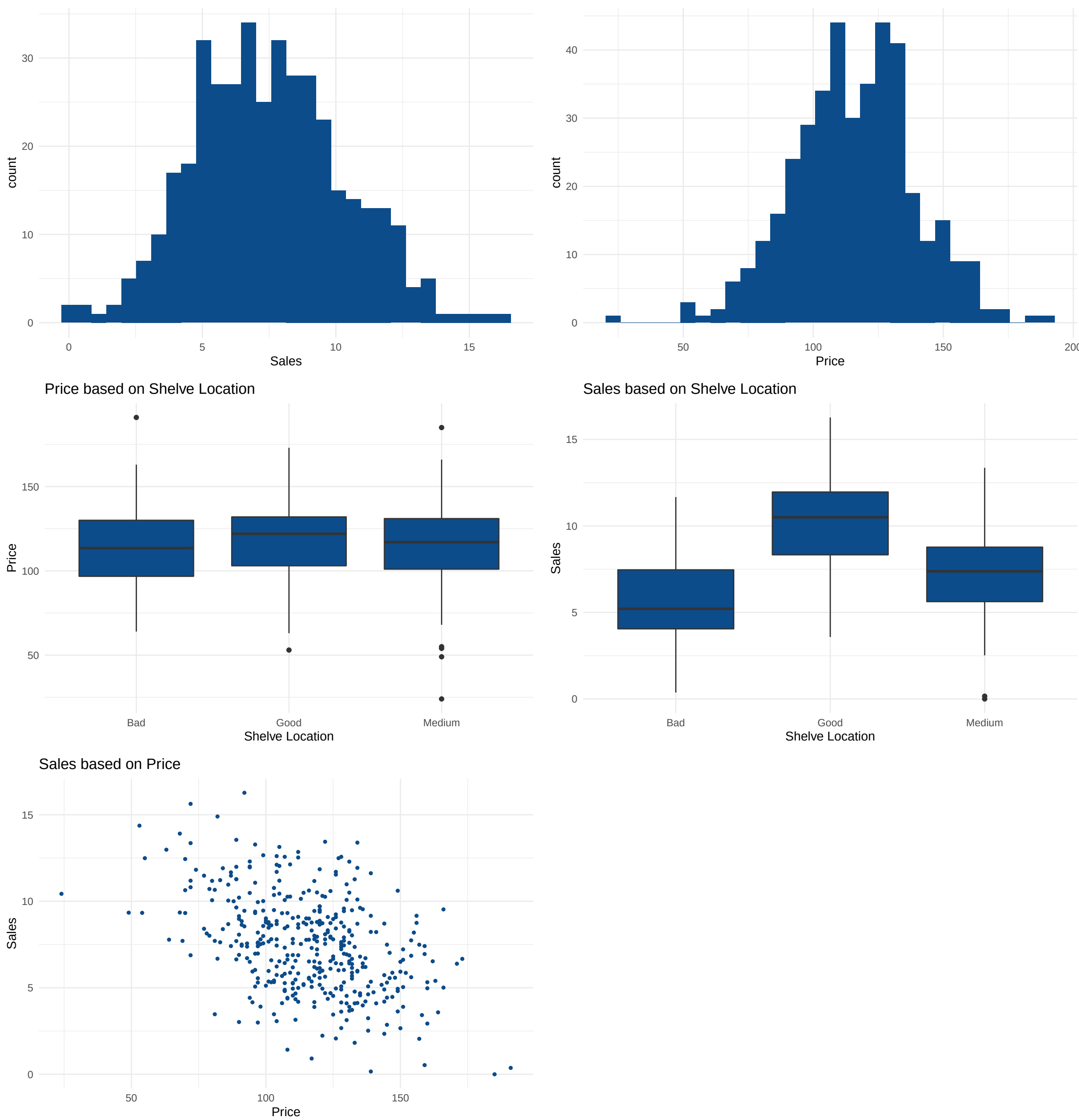
II. Data Prepration & Exploration

```
##      Sales            Price        ShelveLoc
##  Min.   : 0.000   Min.   : 24.0   Bad   : 96
##  1st Qu.: 5.390   1st Qu.:100.0   Good  : 85
##  Median : 7.490   Median :117.0   Medium:219
##  Mean   : 7.496   Mean   :115.8
##  3rd Qu.: 9.320   3rd Qu.:131.0
##  Max.   :16.270   Max.   :191.0

##   Sales Price ShelveLoc
## 1  9.50   120       Bad
## 2 11.22    83      Good
## 3 10.06    80    Medium
## 4  7.40    97    Medium
## 5  4.15   128       Bad
## 6 10.81    72       Bad
```

```
##
## FALSE
##  1200

## [1] 0
```



#Summary: Price is slightly higher for car seats with good shelve location. #Intuitively, Sales are higher for car seats with good shelve location. #There is a negative relationship between Sales and Price.

III. Building the Model

Building linear model to predict sales for car seats with good and bad shelve location separately

```
Model_GoodShelve <- lm(Sales ~ Price, data = Good_Shelve)

summary(Model_GoodShelve)
```

```
##
## Call:
## lm(formula = Sales ~ Price, data = Good_Shelve)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.721 -1.351 -0.098  1.483  4.353
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.968864   0.988008  18.187  < 2e-16 ***
## Price       -0.065785   0.008199  -8.023 5.85e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.888 on 83 degrees of freedom
## Multiple R-squared:  0.4368, Adjusted R-squared:   0.43
## F-statistic: 64.37 on 1 and 83 DF,  p-value: 5.848e-12
```

```
Model_BadShelve <- lm(Sales ~ Price, data = Bad_Shelve)

summary(Model_BadShelve)
```

```
##
## Call:
## lm(formula = Sales ~ Price, data = Bad_Shelve)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4622 -1.0617 -0.2014  1.2050  4.6412
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.832984   0.990317  11.949  < 2e-16 ***
## Price       -0.055220   0.008486  -6.507  3.7e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.967 on 94 degrees of freedom
## Multiple R-squared:  0.3105, Adjusted R-squared:  0.3032
## F-statistic: 42.34 on 1 and 94 DF,  p-value: 3.702e-09
```

```
#Both models confirm the negative relationship between Sales and Price. As price increases, sales decre
#R squared for the good-shelve model is 43%, which means that the model explains 43% of the variability
```

IV. Problem Formulation:

Total Profit= Sales * Unit Profit Unit Profit= Unit price – Unit cost
Sales= b1*price + b0 Total Profit= (Unit price – Unit cost)* (b1*price+b0) = b1*price^2+(b0 -cost*b1)*price -cost*b0

Now we will set the derivative to Zero to find optimal price

Total Profit derivative = 2b1*price + b0-b1*Cost

Then we solve for the price:

Price = (-b0)-(b1*Cost)) / 2b1

V. Problem Solving:

```
#create function to calculate optimal price for car seats given fixed productions cost of $55
Optimal_Price <- function(cost, b0, b1) {
    return(((-b0) + (b1 * cost)) / (2 * b1))
}


#Optimal price for car seats with good shelve location given that the cost = $55 is $164.0731
OptimalPrice_Good <- Optimal_Price(55, Model_GoodShelve$coefficients[1], Model_GoodShelve$coefficients[
OptimalPrice_Good
```

```
## (Intercept)
##    164.0731
```

```
#Optimal price for car seats with bad shelve location given that the cost = $55 is $ 134.6435
OptimalPrice_bad<- Optimal_Price(55, Model_BadShelve$coefficients[1], Model_BadShelve$coefficients[2])
OptimalPrice_bad
```

```
## (Intercept)
##    134.6435
```

Plot the optimal price when cost varies between $40 and $85

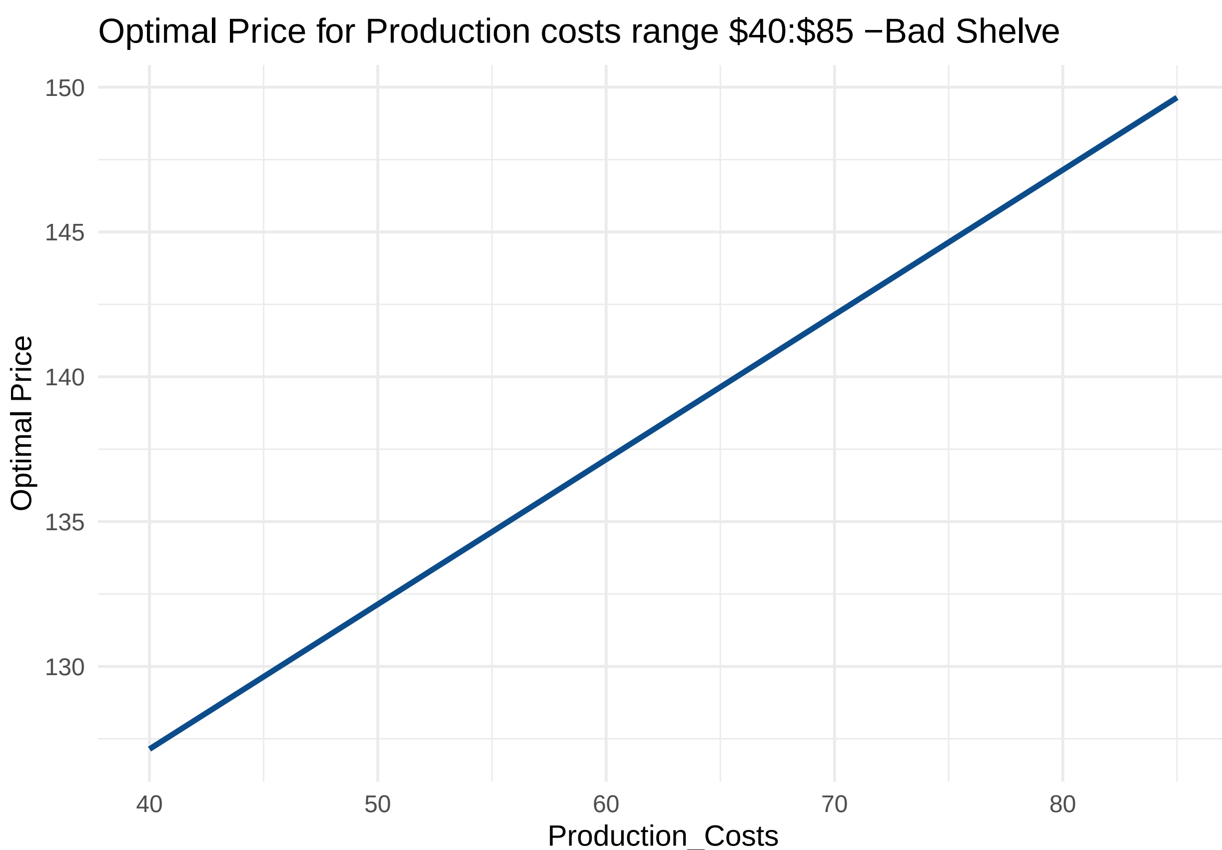
```
Production_Costs <- seq(40,85, by =5) #production costs from $40 to $85 (in $5 increments)


for (i in Production_Costs) {
  Optimal_good <- Optimal_Price(Production_Costs, Model_GoodShelve$coefficients[1], Model_GoodShelve$co
  Optimal_bad <- Optimal_Price(Production_Costs, Model_BadShelve$coefficients[1], Model_BadShelve$coeffi
}


#combining production costs and optimal prices for good and bad shelve location in a data frame
df <- as.data.frame (cbind(Production_Costs, Optimal_good, Optimal_bad))
df
```

```
##    Production_Costs Optimal_good Optimal_bad
## 1                40     156.5731    127.1435
## 2                45     159.0731    129.6435
## 3                50     161.5731    132.1435
## 4                55     164.0731    134.6435
## 5                60     166.5731    137.1435
## 6                65     169.0731    139.6435
## 7                70     171.5731    142.1435
## 8                75     174.0731    144.6435
## 9                80     176.5731    147.1435
## 10               85     179.0731    149.6435
```

```r
#Plotting Optimal Prices for Production costs range $40:$85 -Bad Shelve location
ggplot(df) +
 aes(x = Production_Costs, y = Optimal_bad) +
 geom_line(size = 1L, colour = "#0c4c8a") +
 labs(y = "Optimal Price", title = "Optimal Price for Production costs range $40:$85 -Bad Shelve ") +
 theme_minimal()
```




```r
#Plotting Optimal Prices for Production costs range $40:$85 -Good Shelve location
ggplot(df) +
 aes(x = Production_Costs, y = Optimal_good) +
 geom_line(size = 1L, colour = "#0c4c8a") +
 labs(y = "Optimal Price", title = "Optimal Price for Production costs range $40:$85 -Good Shelve") +
 theme_minimal()
```


Optimal Price for Production costs range $40:$85 −Good Shelve
Optimal Price
180
175
170
165
160
40
50
60
70
80
Production_Costs