

Safe Babies Car Seat Price Optimization

Mark Bruner

21 March 2021

```
library(tidyverse)
library(glmnet)
library(corrplot)
library(ISLR)

safe_babies <- Carseats %>% select("Sales", "Price", "ShelveLoc")

# Creating profit column for future visualizations. Equation for profit: (price
# - cost) * (quantity of sales)
safe_babies$profit <- (safe_babies$Price - 55) * safe_babies$Sales
```

Part 1: Exploratory Data Analysis

Univariate Analysis

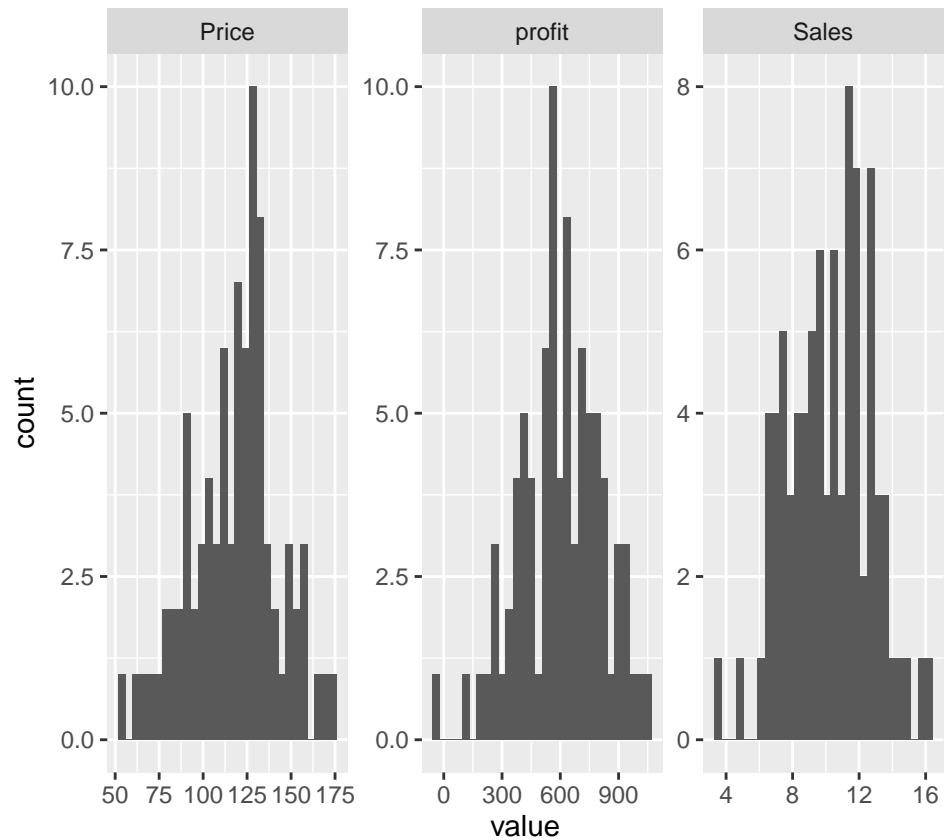
Measures of Central Tendency and Dispersion

All columns have means and medians fairly close together, which means that they are more normally distributed. The 1st and 3rd quartiles are mostly symmetrical around the mean/median. The “ShelveLoc” column’s category “Medium” is where most of the carseats are located.

| ## | Sales | Price | ShelveLoc | profit |
|----|----------------|----------------|-------------|----------------|
| ## | Min. : 0.000 | Min. : 24.0 | Bad : 96 | Min. : -323.3 |
| ## | 1st Qu.: 5.390 | 1st Qu.: 100.0 | Good : 85 | 1st Qu.: 287.1 |
| ## | Median : 7.490 | Median : 117.0 | Medium: 219 | Median : 399.9 |
| ## | Mean : 7.496 | Mean : 115.8 | | Mean : 426.1 |
| ## | 3rd Qu.: 9.320 | 3rd Qu.: 131.0 | | 3rd Qu.: 549.1 |
| ## | Max. : 16.270 | Max. : 191.0 | | Max. : 1057.8 |

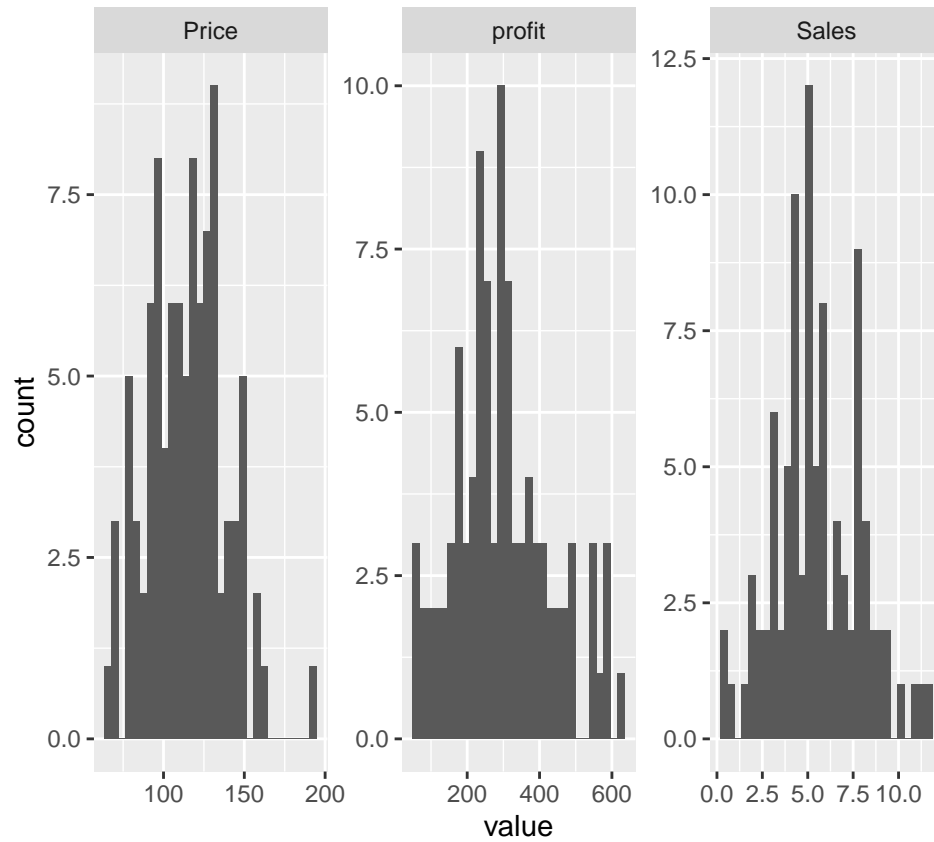
Histograms of Carseats in Good Shelf locations

From these plots, it confirms that the columns of profit and price are normally distributed. The sales column is more skewed to the right. I don't see any outliers that are mistakes. The prices are all positive as well as the sales which is what you would expect to see. The profit has a negative value but profit can be negative if the price of carseats is less than the cost of production so that doesn't seem unrealistic.



Histograms of Carseats in Bad Shelf locations

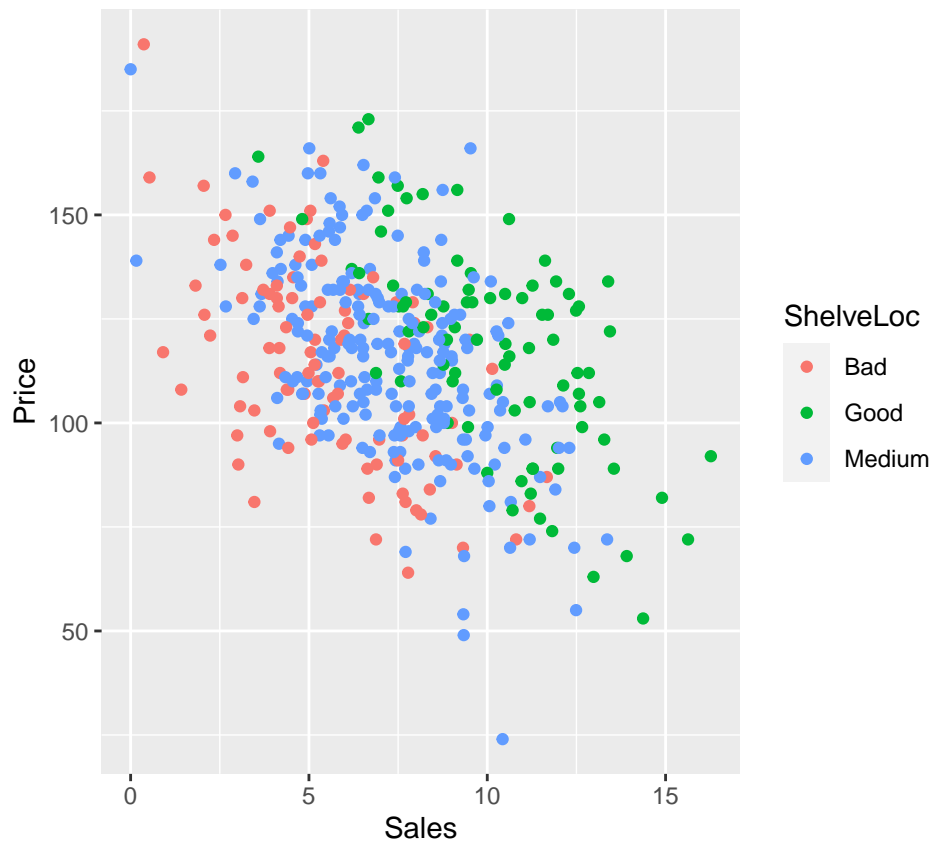
From these plots, it confirms that the columns are more normally distributed. The price column seems to have an outlier that is more extreme but since the dataset is fairly small, 1 price that is outside the norm is worth looking into but it doesn't seem too extreme.



Bivariate Analysis

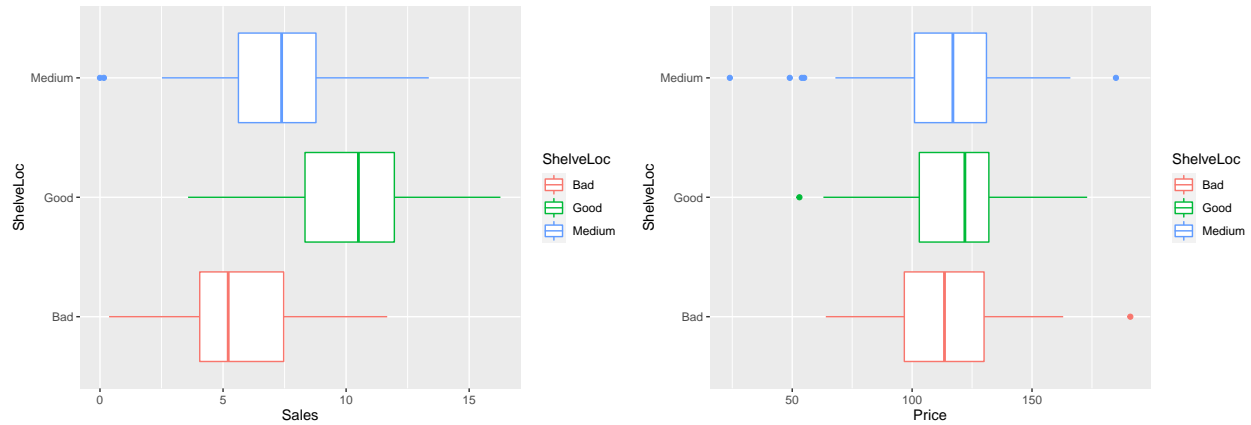
Relationship Between Sales & Price

There is a negative relationship between sales and price, which is what you would expect. As price increases, sales decrease and visa versa. You would also expect carsseats placed in a good shelf location would have, on average, higher sales. The lower the shelf location the lower the sales. You can see that the green dots (representing good shelf location), have mostly higher sales compared to the other locations.



Boxplot Analysis by Sales and Price

Most of the outliers were in the medium shelf location and the better the shelf location the better the sales, on average. The other interesting relationship was that the price spread for both medium and good shelf locations were almost identical.



Outlier Table Analysis

The negative profit is associated with those with prices below the \$55 cost which is to be expected. The others with high prices have low sales which again makes sense. The outliers seem to be more informative than mistakes. Also, most of them are in the “Medium” location and we will be looking at Bad and Good shelf locations so I don’t believe any further action is necessary for the outliers.

| ## | Sales | Price | ShelveLoc | profit |
|--------|-------|-------|-----------|---------|
| ## 43 | 10.43 | 24 | Medium | -323.33 |
| ## 126 | 9.34 | 49 | Medium | -56.04 |
| ## 368 | 14.37 | 53 | Good | -28.74 |
| ## 314 | 9.33 | 54 | Medium | -9.33 |
| ## 172 | 12.49 | 55 | Medium | 0.00 |
| ## 107 | 0.16 | 139 | Medium | 13.44 |
| ## 175 | 0.00 | 185 | Medium | 0.00 |
| ## 166 | 0.37 | 191 | Bad | 50.32 |

Multivariate Analysis

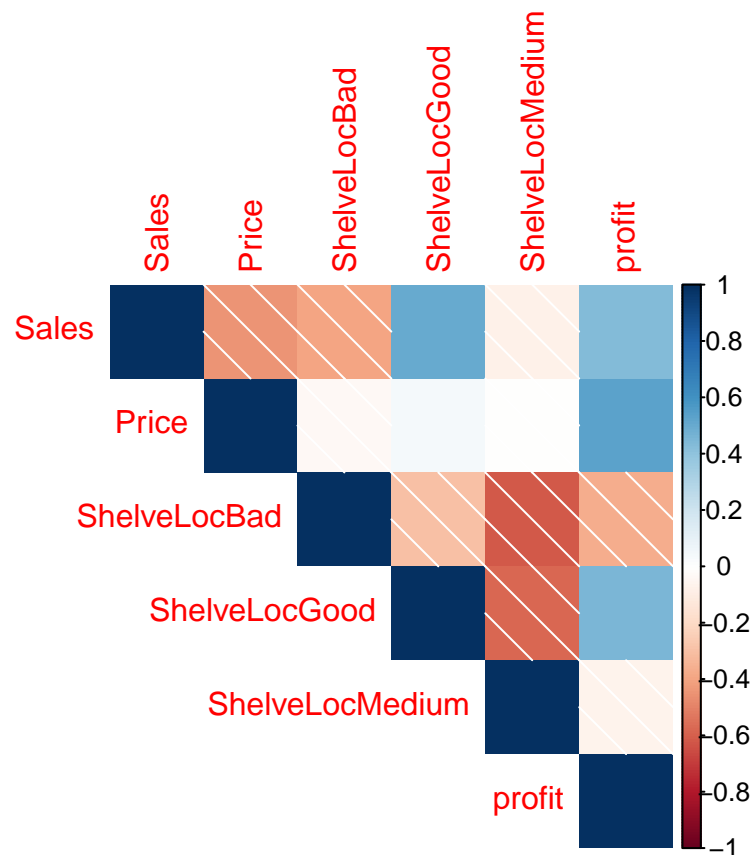
The price, good and bad shelf locations are all fairly uncorrelated. The “Good” and “Bad” shelf locations have a slight negative correlation of .29 but shouldn’t be too much of a problem. “Medium” is highly negatively correlated with “Good” and “Bad”.

```
# One hot encoding the shelf location to keep them in the glm.
safe_babies_mx <- model.matrix(~. - 1, safe_babies[, 1:4])

# Separating the sales variable from the other variables. Will check to see if
# the predictors are independent of one another.
x = safe_babies_mx[, 2:5]
y = safe_babies_mx[, 1]

safebabies_cor <- cor(safe_babies_mx)

corrplot(safebabies_cor, method = "shade", type = "upper")
```



Part 2: Model Building

Optimal GLM model for Sales & Price Based on Location

```
# Using 20 fold cross-validation to find the optimal glm model for our dataset.
for (a in seq(0, 1, by = 0.1)) {
  cv.fit <- cv.glmnet(x, y, alpha = a, nfolds = 20)
  print(paste0("alpha is ", a, " and best MSE is ", min(cv.fit$cvm)))
}
```

```
## [1] "alpha is 0 and best MSE is 3.71719206021491"
## [1] "alpha is 0.1 and best MSE is 3.71889469811023"
## [1] "alpha is 0.2 and best MSE is 3.71853137933202"
## [1] "alpha is 0.3 and best MSE is 3.72547962279592"
## [1] "alpha is 0.4 and best MSE is 3.72271773633929"
## [1] "alpha is 0.5 and best MSE is 3.72942339515315"
## [1] "alpha is 0.6 and best MSE is 3.68781166253832"
## [1] "alpha is 0.7 and best MSE is 3.70040428793626"
## [1] "alpha is 0.8 and best MSE is 3.70530892490352"
## [1] "alpha is 0.9 and best MSE is 3.69900292488701"
## [1] "alpha is 1 and best MSE is 3.73205685504604"
```

It looks like the alpha value at .9 is optimal.

```
# The Best Fit Model for This Dataset
best_fit <- coef(cv.fit, s = "lambda.min")
best_fit
```

```
## 5 x 1 sparse Matrix of class "dgCMatrix"
##              1
## (Intercept) 13.81940138
## Price      -0.05631621
## ShelvelocBad -1.84605181
## ShelvelocGood 3.01699343
## ShelvelocMedium .
```

Using the best fitted model from above to model the relationship between sales and price + shelf location. You will notice that Medium shelf location was eliminated from this model.

Part 3: Optimal Price for \$55 Cost

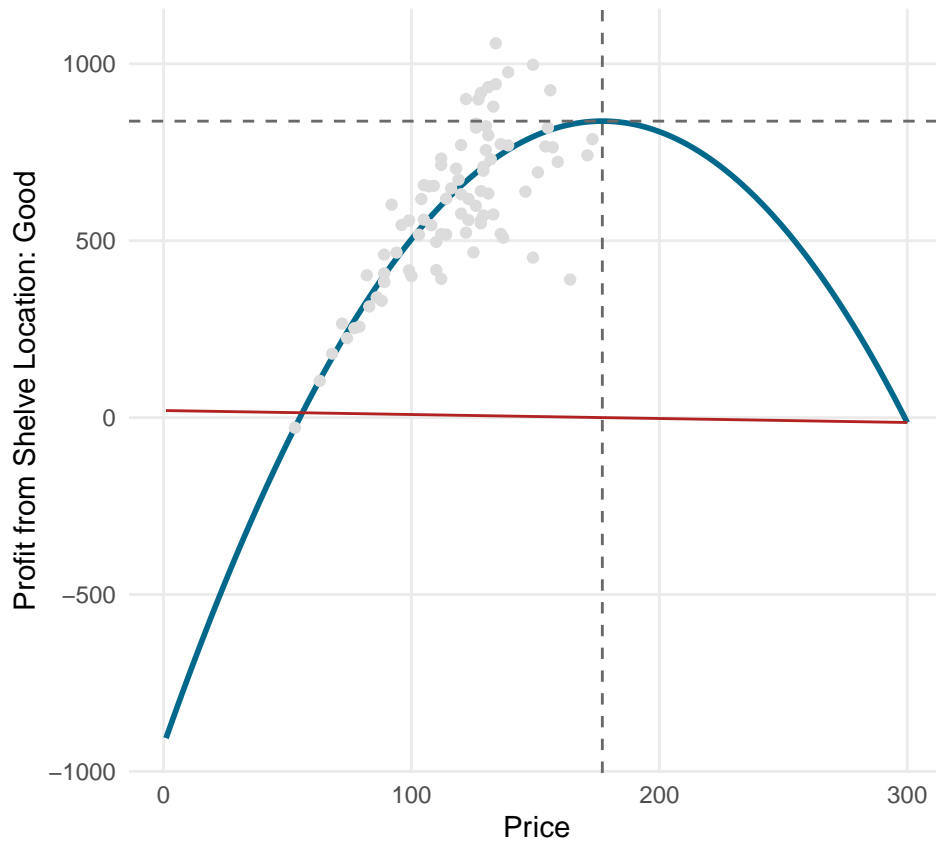
```
# Created a function to find the optimal value based on shelf location. cost is  
# the cost of carseat production, good and bad represent shelf location and they  
# are binary.  
optimal <- function(cost, good, bad) {  
  # Derivative of the first term in the quadratic profit equation  
  x1 <- 2 * best_fit[2]  
  
  # Derivative of the second term in the profit equation  
  x2 <- -1 * best_fit[1] + -1 * best_fit[3] * bad + -1 * best_fit[4] * good + (best_fit[2] *  
    cost)  
  
  # Solving the derivative for max price unknown  
  optimal_price <- x2/x1  
  
  optimal_sales <- best_fit[1] + best_fit[3] * bad + best_fit[4] * good + best_fit[2] *  
    optimal_price  
  
  optimal_profit <- (optimal_price - cost) * optimal_sales  
  return(as.data.frame(cbind(price = round(optimal_price, 2), sales = round(optimal_sales,  
    2), profit = round(optimal_profit, 2))))  
}
```

Optimal Price, Sales, and Profit for Good Shelf Location

```
optimal(55, 1, 0)
```

```
##      price sales profit  
## 1 176.98  6.87 837.95
```

Optimal price for a \$55 cost is \$176.98 or \$177. Optimal sales is 6.87 thousand carseats which will provide \$837.95 thousand dollar profit.



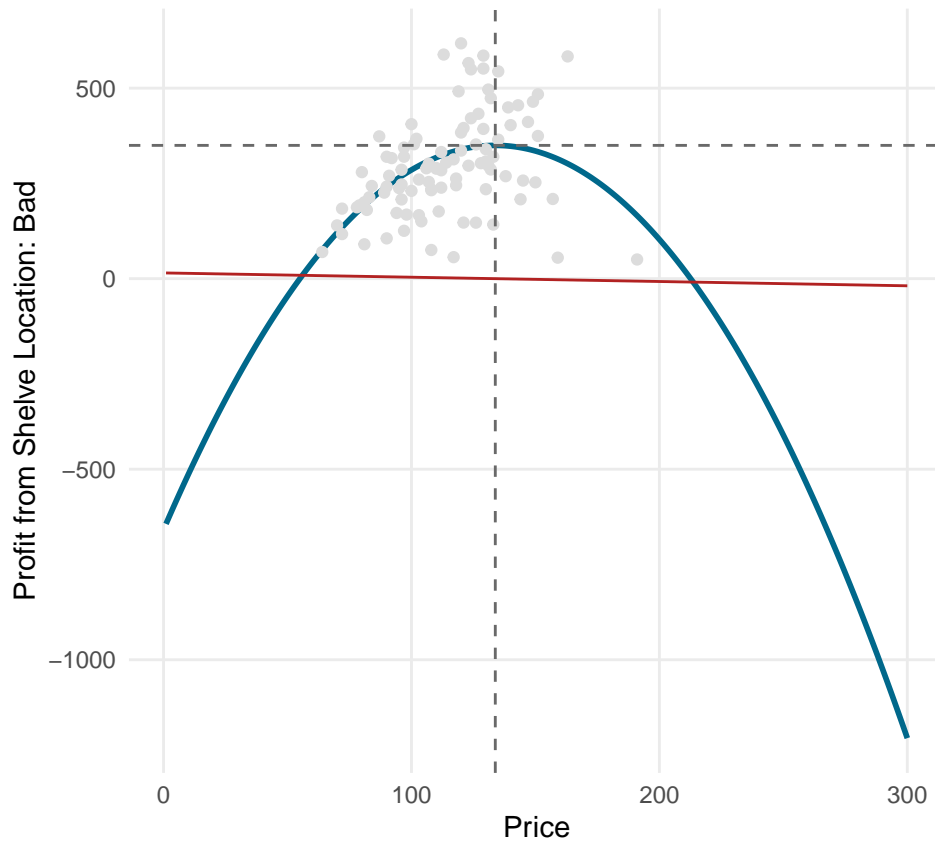
The plot above shows the fit of the profit quadratic equation to the good shelf data and shows that the derivative intercepts the x-axis at the optimal price of \$177.

Optimal Price, Sales, and Profit for Bad Shelf Location

```
optimal(55, 0, 1)
```

```
## price sales profit
## 1 133.8 4.44 349.73
```

Optimal price for bad shelf location at a 55 cost is \$133.8 or about \$134. Optimal sales is 4.44 thousand carseats which will provide \$349.73 thousand dollar profit.



The plot above shows the fit of the profit quadratic equation to the good shelf data and shows that the derivative intercepts the x-axis at the optimal price of \$177.

```
# Creating
costs <- as.data.frame(seq(40, 85, by = 5))
costs <- costs %>% rename(costs = 'seq(40, 85, by = 5)')

for (i in costs) {
  # i-th element of 'costs' used to find good shelf optimal sales, price, and
  # profit
  good_results <- as.data.frame(optimal(i, 1, 0))
  good_results <- (cbind(costs, good_results))
}

for(i in costs) {
  # i-th element of 'costs' used to find bad shelf optimal sales, price, and profit
  bad_results <- as.data.frame(optimal(i, 0, 1))
  bad_results <- (cbind(costs, bad_results))
}
```

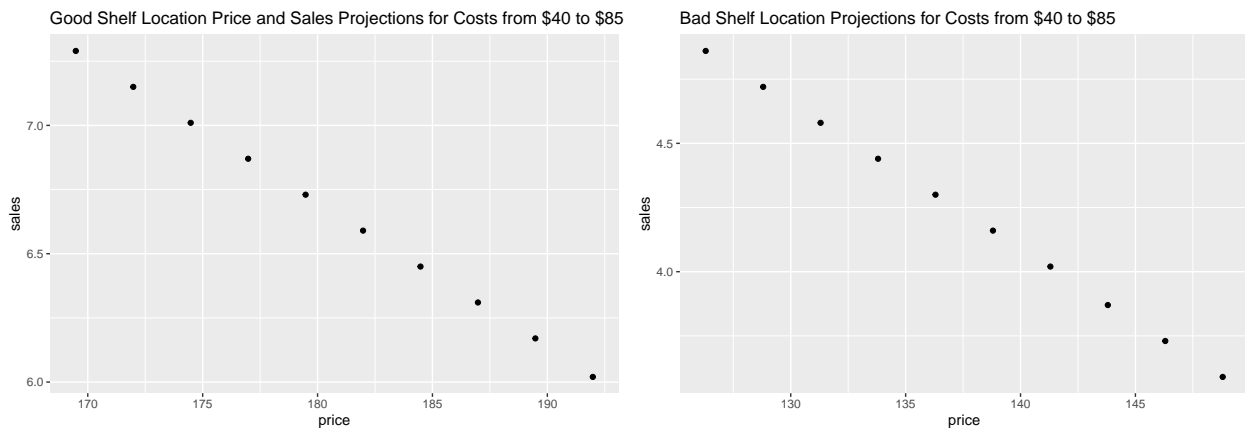
```
good_results %>% mutate(profit_margin_percent = round((price - costs)/price * 100))
```

| ## | costs | price | sales | profit | increase_profit | profit_margin_percent |
|-------|-------|--------|-------|--------|-----------------|-----------------------|
| ## 1 | 40 | 169.48 | 7.29 | 944.16 | 36.11 | 76 |
| ## 2 | 45 | 171.98 | 7.15 | 908.05 | 35.40 | 74 |
| ## 3 | 50 | 174.48 | 7.01 | 872.65 | 34.70 | 71 |
| ## 4 | 55 | 176.98 | 6.87 | 837.95 | 34.00 | 69 |
| ## 5 | 60 | 179.48 | 6.73 | 803.95 | 33.29 | 67 |
| ## 6 | 65 | 181.98 | 6.59 | 770.66 | 32.59 | 64 |
| ## 7 | 70 | 184.48 | 6.45 | 738.07 | 31.88 | 62 |
| ## 8 | 75 | 186.98 | 6.31 | 706.19 | 31.18 | 60 |
| ## 9 | 80 | 189.48 | 6.17 | 675.01 | 30.48 | 58 |
| ## 10 | 85 | 191.98 | 6.02 | 644.53 | 0.00 | 56 |

```
bad_results %>% mutate(profit_margin_percent = round((price - costs)/price * 100))
```

| ## | costs | price | sales | profit | increase_profit | profit_margin_percent |
|-------|-------|-------|-------|--------|-----------------|-----------------------|
| ## 1 | 40 | 126.3 | 4.86 | 419.47 | 23.95 | 68 |
| ## 2 | 45 | 128.8 | 4.72 | 395.52 | 23.24 | 65 |
| ## 3 | 50 | 131.3 | 4.58 | 372.28 | 22.55 | 62 |
| ## 4 | 55 | 133.8 | 4.44 | 349.73 | 21.83 | 59 |
| ## 5 | 60 | 136.3 | 4.30 | 327.90 | 21.14 | 56 |
| ## 6 | 65 | 138.8 | 4.16 | 306.76 | 20.43 | 53 |
| ## 7 | 70 | 141.3 | 4.02 | 286.33 | 19.72 | 50 |
| ## 8 | 75 | 143.8 | 3.87 | 266.61 | 19.03 | 48 |
| ## 9 | 80 | 146.3 | 3.73 | 247.58 | 18.31 | 45 |
| ## 10 | 85 | 148.8 | 3.59 | 229.27 | 0.00 | 43 |

Lowering the costs by \$15 a unit would increase our profits by about \$106,000 for good shelf location and \$68,000 for bad shelf locations. Increasing the profit by \$174,000. Also for good shelf location, the highest profit margin can be achieved at a cost of \$40 which is a margin increase of 7% from our current \$55 cost. For bad shelf location, the highest profit margin can be achieved at a cost of \$40 which is a margin increase of 9% from our current \$55 cost.



The optimal price range for costs between 40 to 85 for good shelf location is about \$170 to \$192 and bad shelf location of \$126 to \$149.