# npillana_Assignment 1

## Loading Libraries

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

## Data Preparation

##Loading DataSet

```
SafeBabies <- Carseats %>% select("Sales", "Price", "ShelveLoc")
summary(SafeBabies)

##      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
```

### Clean the data. Check for any missing vales

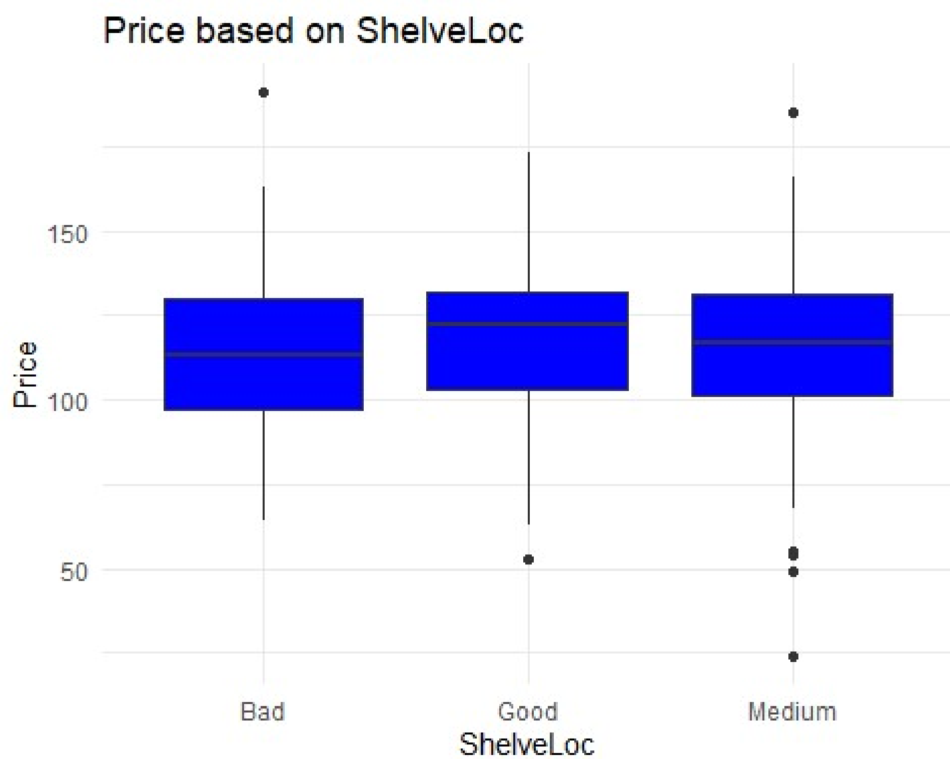
```
table(is.na(SafeBabies))

##
## FALSE
##  1200

sum(is.na(SafeBabies))

## [1] 0
```

## Divide data based on shelveLoc i.e, Good, Bad and Medium

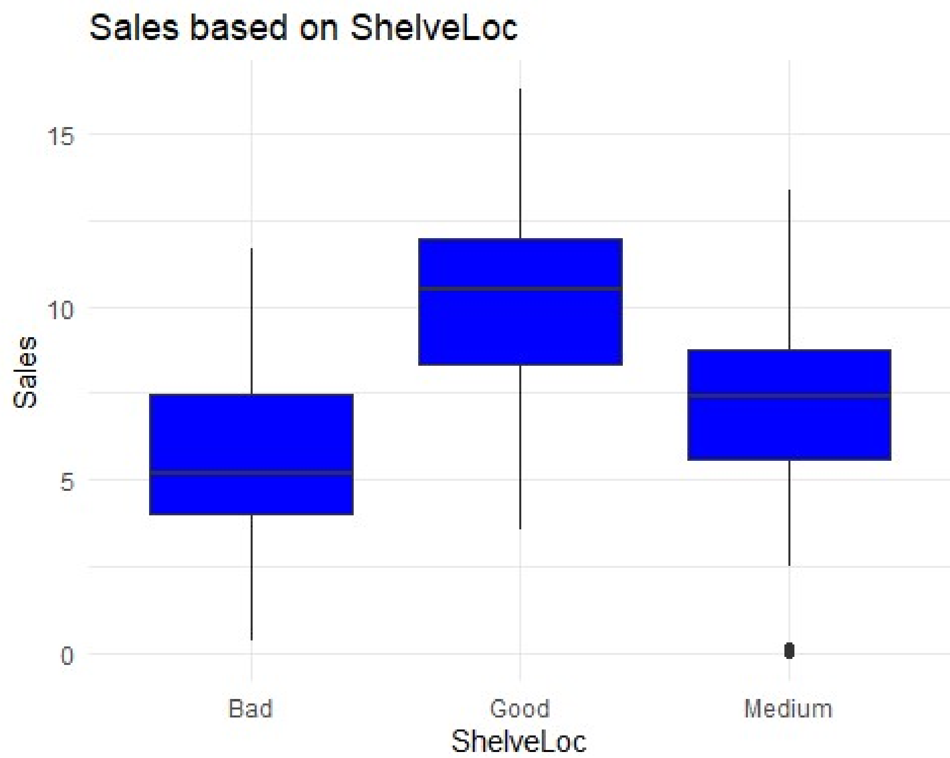
```
Good <- filter(SafeBabies, ShelveLoc == "Good")
Bad <- filter(SafeBabies, ShelveLoc == "Bad")
Medium <- filter(SafeBabies, ShelveLoc == "Medium")
```

## Visualizing the Price and Sales based on Shelve location

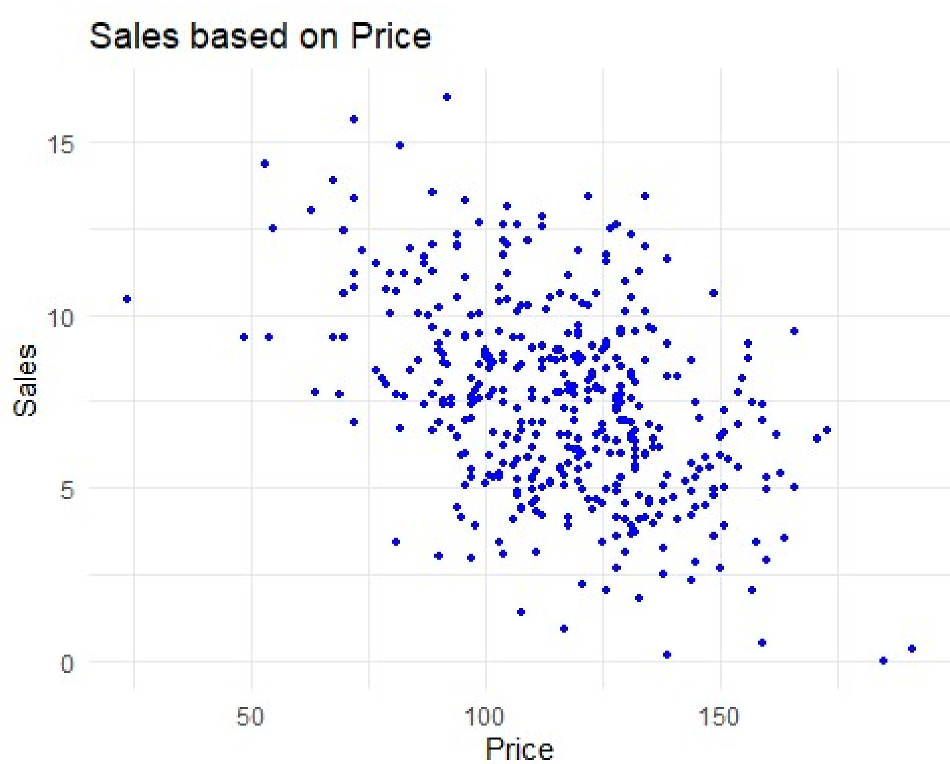
```
ggplot(data=SafeBabies) + aes(x = ShelveLoc, y = Price) + geom_boxplot(fill =
"blue") + labs(x = "ShelveLoc", title = "Price based on ShelveLoc") +
theme_minimal()
```




```
ggplot(data=SafeBabies) + aes(x = ShelveLoc, y = Sales) + geom_boxplot(fill =
"blue") + labs(x = "ShelveLoc", title = "Sales based on ShelveLoc") +
theme_minimal()
```




```r
ggplot(data=SafeBabies) + aes(x = Price, y = Sales) + geom_point(size = 1L,
colour = "blue") + labs(title = "Sales based on Price") + theme_minimal()
```



Based on above visuals the ShelveLoc with Good has the highest sales when compared to others and the price for Good ShelveLoc is slightly higher than Bad and Medium

# Build a Linear Model

### Lets build a linear model to predict the sales for all ShelveLoc

```
Good_Shelve <- lm(Sales ~ Price, data = Good)
Bad_Shelve <- lm(Sales ~ Price, data = Bad)
summary(Good_Shelve)
```

```
##
## Call:
## lm(formula = Sales ~ Price, data = Good)
##
## 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
```

```
summary(Bad_Shelve)
```

```
##
## Call:
## lm(formula = Sales ~ Price, data = Bad)
##
## 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
```

The adjusted R2 tells you the percentage of variation explained by only the independent variables that actually affect the dependent variable. Adjusted R-squared:0.43 for Good ShelveLoc and Adjusted R-squared:0.3032 for Bad ShelveLoc, which explains variability in the car price for Car Seats.

## Formulating the Problem

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

Now we have to find the optimal value, set the derivative to Zero Total Profit derivative = 2b1*price + b0-b1*cost

Then Price is; Price = (-b0)-(b1*cost))/2b1

## Solving the Problem

```
#Lets calculate optimal price for car seat for both Good and Bad Shelve at
cost $55

OptimalPrice <- function(prod_cost, b_0, b_1) {return(((-1 * b_0) + (b_1 *
prod_cost)) /(2 * b_1))}
Good_OptimalPrice<- OptimalPrice(55, Good_Shelve$coefficients[1],
Good_Shelve$coefficients[2])
Good_OptimalPrice

## (Intercept)
##    164.0731

paste("optimal price for car seats with good shelve",Good_OptimalPrice)

## [1] "optimal price for car seats with good shelve 164.07312564386"

Bad_OptimalPrice<- OptimalPrice(55, Bad_Shelve$coefficients[1],
Bad_Shelve$coefficients[2])
Bad_OptimalPrice

## (Intercept)
##    134.6435

paste("optimal price for car seats with Bad shelve",Bad_OptimalPrice)

## [1] "optimal price for car seats with Bad shelve 134.643464696399"
```

### Calculate the optimal prive when the cost varies from $40 to $80 with $5 increments

```
Varied_Cost <- seq(40,85, by=5)

for( i in Varied_Cost){
```

```
  OptimalPrice_Good <- OptimalPrice(Varied_Cost,
Good_Shelve$coefficients[1],Good_Shelve$coefficients[2])
  OptimalPrice_Bad <- OptimalPrice(Varied_Cost,
Bad_Shelve$coefficients[1],Bad_Shelve$coefficients[2])
}
```

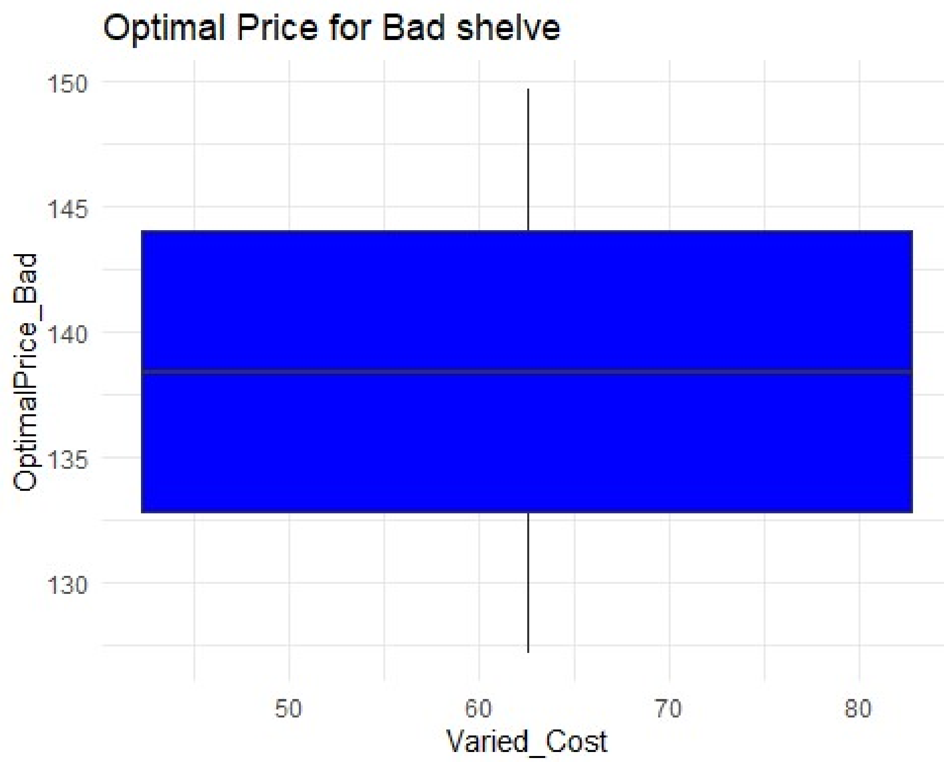
## Now, we will combine production costs and optimal prices for good and bad shelve location

```
Combine_cost <- data.frame(cbind(Varied_Cost,OptimalPrice_Good,
OptimalPrice_Bad))
Combine_cost

##    Varied_Cost OptimalPrice_Good OptimalPrice_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
```

## Visualizing the optimal cost with varying with Production costs from $40 to $85 for Good Shelve and Bad Shelve

```
ggplot(data=Combine_cost) + aes(x = Varied_Cost, y = OptimalPrice_Good) +
geom_boxplot(fill = "blue") + labs(x = "Varied_Cost", title = "Optimal Price
for Good shelve") + theme_minimal()

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```


```
ggplot(data=Combine_cost) + aes(x = Varied_Cost, y = OptimalPrice_Bad) +
geom_boxplot(fill = "blue") + labs(x = "Varied_Cost", title = "Optimal Price
for Bad shelve") + theme_minimal()

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```




**The above Visuals show that the Optimal Price for the Car seats at Good Shelve is higher than the Optimal Price for the Car seats at Bad Shelve.**