

COSC6323 - Exercise 7

Sachin Shubham

3/27/2021

1 Task We are running a small sales company and we want to estimate the marketing budget for the next year. Use marketing data set which contains the impact of the amount of money spent on three advertising medias (YouTube, Facebook and newspaper) on sales. Install datarium package using devtools :: install_github("kassmbara/datarium") then load and inspect the marketing data. Build a model for estimating sales based on the advertising budget invested in youtube, facebook and newspaper. Use the following outline for your report: 0. Build a multiple linear regression model. 1. Create the necessary plots. 2. Examine the F-statistic and the associated p-value, interpret it. 3. Examine which predictor variables are significant. 4. Obtain the regression equation. 5. What are your recommendations for the budget planning based on the reviewed data?

```
data("marketing", package = "datarium")
input <- marketing[,c("youtube", "facebook", "newspaper", "sales")]
print(head(input))
```

```
##  youtube facebook newspaper sales
## 1  276.12    45.36     83.04 26.52
## 2   53.40    47.16     54.12 12.48
## 3   20.64    55.08     83.16 11.16
## 4  181.80    49.56     70.20 22.20
## 5  216.96    12.96     70.08 15.48
## 6   10.44    58.68     90.00  8.64
```

```
library(GGally)
```

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

```
## Registered S3 method overwritten by 'GGally':
```

```
##   method from
```

```
##   +.gg    ggplot2
```

#0. Build a multiple linear regression model

```
model <- lm(sales ~ youtube + facebook + newspaper, data = input)
summary(model)
```

```
##
```

```
## Call:
```

```
## lm(formula = sales ~ youtube + facebook + newspaper, data = input)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -10.5932  -1.0690   0.2902   1.4272   3.3951
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.526667   0.374290   9.422  <2e-16 ***
## youtube      0.045765   0.001395  32.809  <2e-16 ***
## facebook     0.188530   0.008611  21.893  <2e-16 ***
## newspaper    -0.001037   0.005871  -0.177    0.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.023 on 196 degrees of freedom
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8956
## F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16

AIC(model)

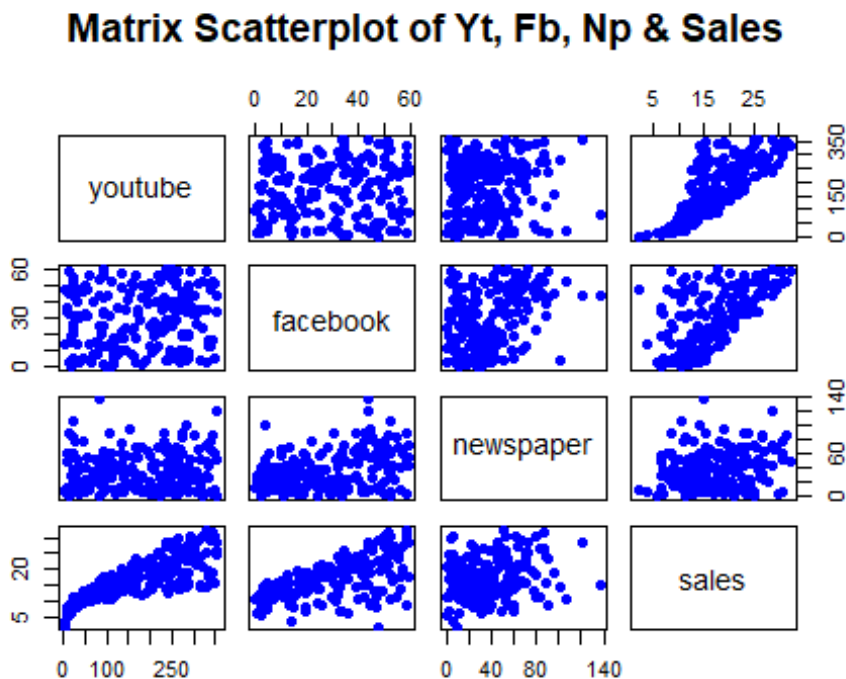
## [1] 855.2909

BIC(model)

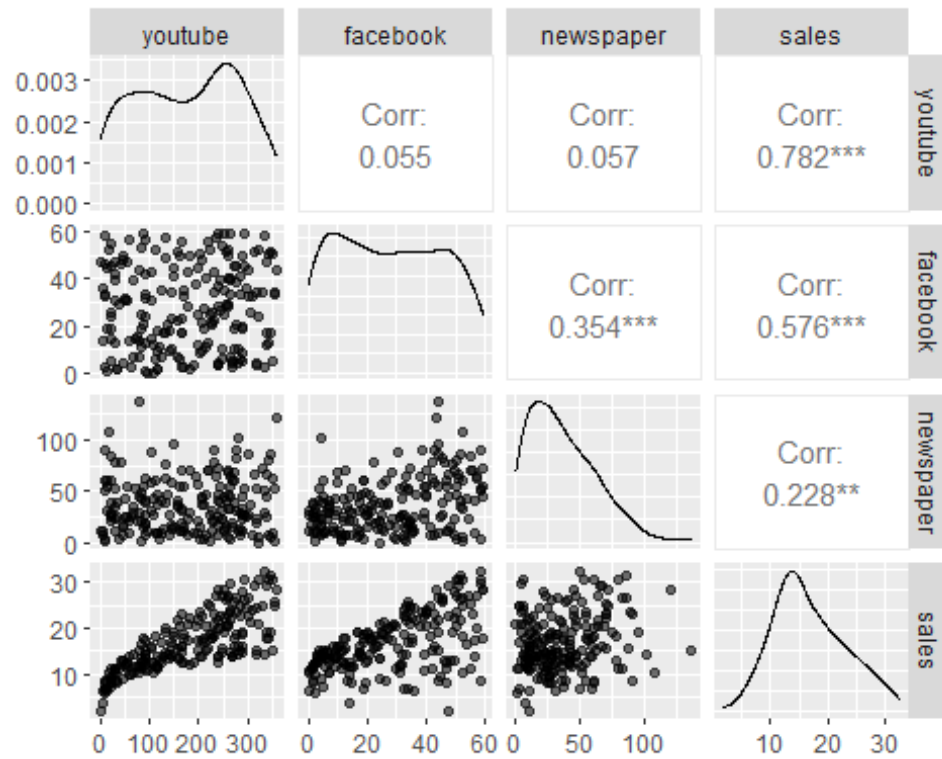
## [1] 871.7824
```

#1. Create the necessary plots.

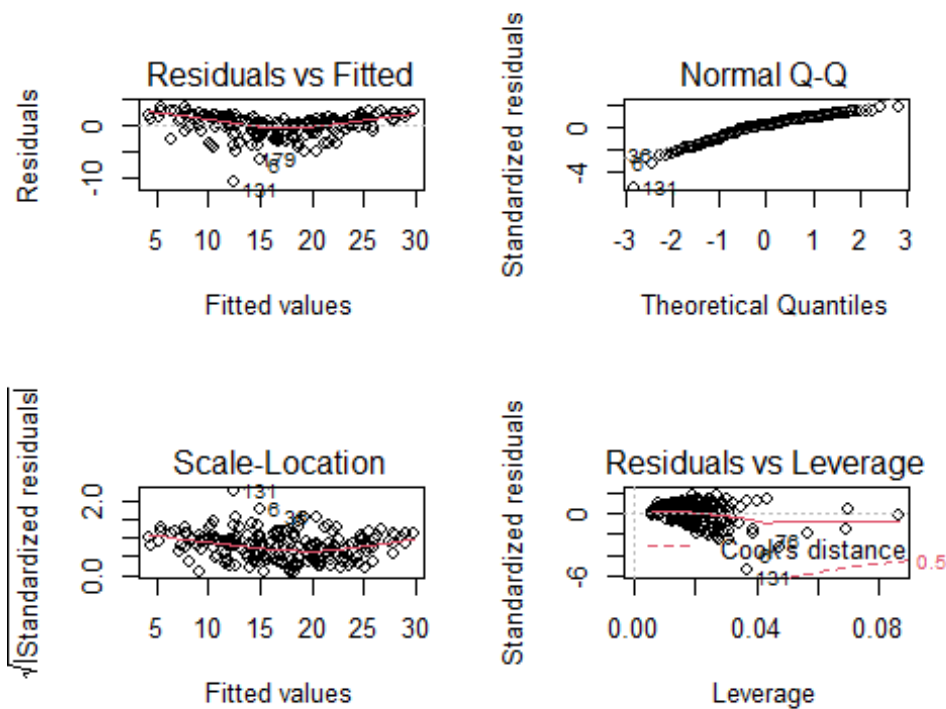
```
# Matrix scatterplot:
plot(input, pch=16, col="blue",
      main="Matrix Scatterplot of Yt, Fb, Np & Sales")
```



```
ggpairs(marketing,aes(alpha=0.4))
```



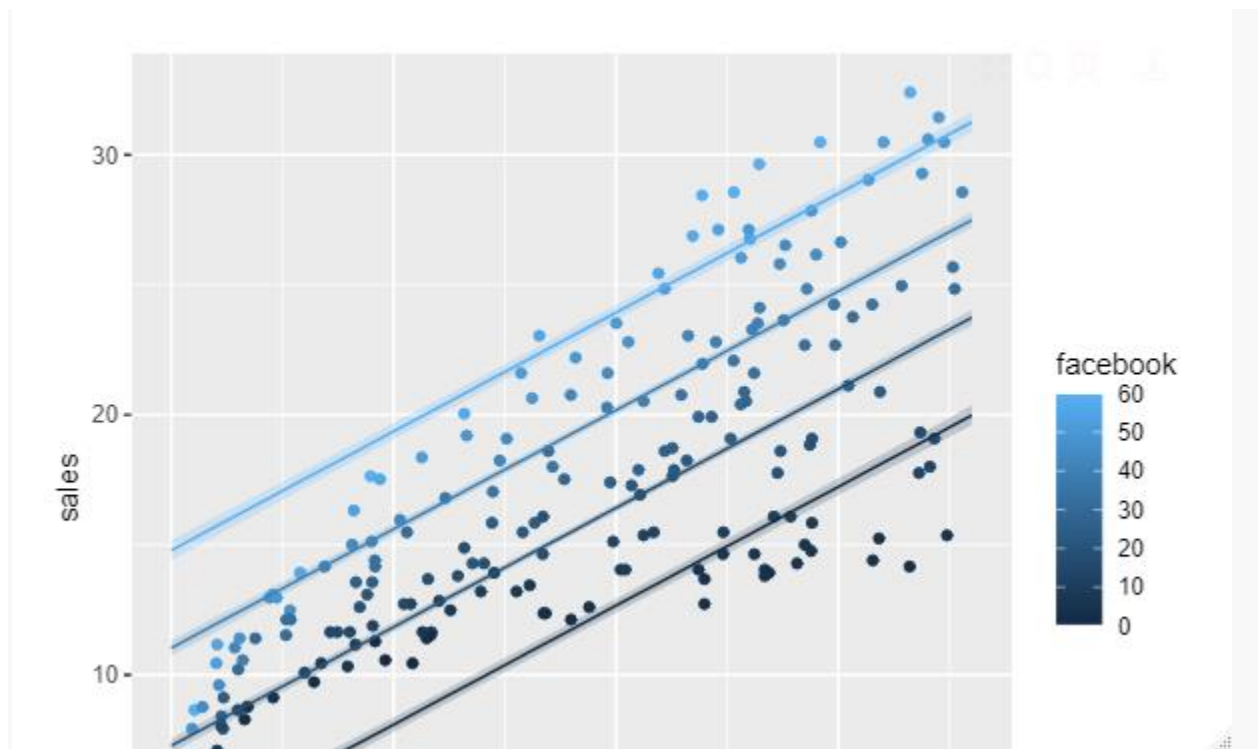
```
# Diagnostic plots for the model
par(mfrow=c(2, 2))
plot(model)
```



```
model2 <- lm(sales~ youtube + facebook, data = input)
summary(model2)
```

```
ggPredict(model2,se=TRUE,interactive=TRUE)
```

```
##
## Call:
## lm(formula = sales ~ youtube + facebook, data = input)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5572  -1.0502   0.2906   1.4049   3.3994
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.50532    0.35339   9.919  <2e-16 ***
## youtube       0.04575    0.00139  32.909  <2e-16 ***
## facebook      0.18799    0.00804  23.382  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.018 on 197 degrees of freedom
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8962
## F-statistic: 859.6 on 2 and 197 DF,  p-value: < 2.2e-16
```



#2. Examine the F-statistic and the associated p-value, interpret it.

As per our study F-statistic: 570.3 on 3 and 196 DF, p-value: $< 2.2e-16$, i.e. very less than 0.05 and extremely significant.

#3. Examine which predictor variables are significant.

```
anova(model)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: sales
```

```
##          Df Sum Sq Mean Sq    F value    Pr(>F)
## youtube    1 4773.1  4773.1 1166.7308 <2e-16 ***
## facebook   1 2225.7  2225.7  544.0501 <2e-16 ***
## newspaper   1    0.1    0.1    0.0312  0.8599
## Residuals 196   801.8    4.1
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The facebook and youtube ads will increase the sales with extremely significant p-values, whereas the newspaper ads appears that it will not increase the sales and has a very high p-value of 0.86 meaning we can keep the null hypothesis that newspaper ads don't contribute to sales. High sum squares for youtube and facebook ads while very low sum squares for newspaper ads. To conclude the newspapers ads don't contribute to sales.

#4. Obtain the regression equation.

```

Interc <- coef(model)[1]
Xyoutube <- coef(model)[2]
Xfacebook <- coef(model)[3]
Xnewspaper <- coef(model)[4]

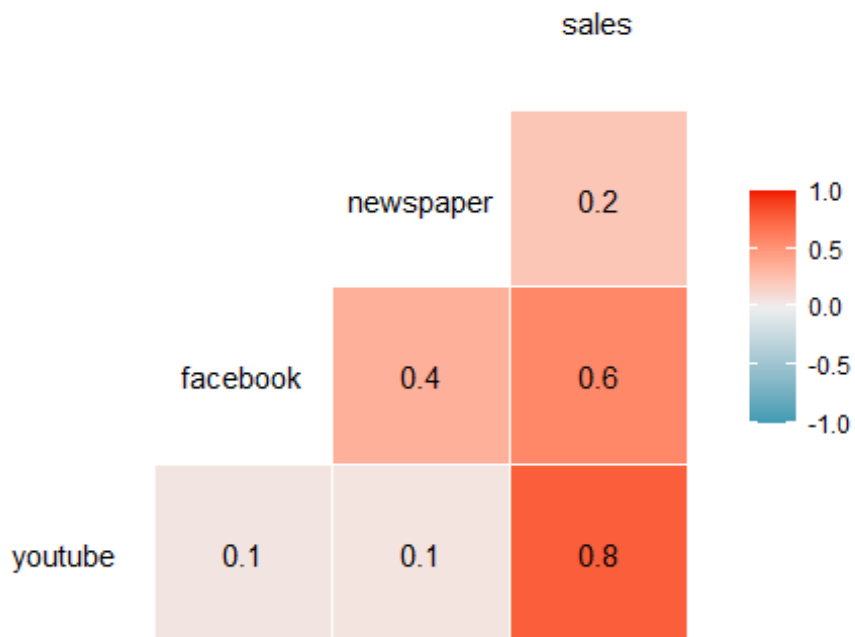
x1 = median(input$youtube)
x2 = median(input$facebook)
x3 = median(input$newspaper)

Y = Interc + Xyoutube*x1 + Xfacebook*x2 + Xnewspaper*x3

```

#5. What are your recommendations for the budget planning based on the reviewed data?

```
ggcorr(marketing, palette = "RdBu", label = TRUE)
```



Measuring correlation of predictor variables, it is found that Youtube ads contribute most to increased sales followed by facebook ads. Newspaper ads do not appear to have a relationship with increased sales so it is recommended to budget just for facebook and youtube ads only.