Subject: Linear Models

Assignment 1

Name: Shruti Rane

PRN No. 23060641076

Problem Statement:

To develop a simple linear regression model and to predict the monthly revenue of a restaurant based on its marketing spend.

Objective:

To establish a reliable relationship between amount invested in marketing activities and the resulting revenue, enabling the restaurant to make informed decisions and optimize its marketing strategy for better financial outcomes. Here,

Y: Monthly Revenue (dependent Variable)

X: Marketing Spend (independent Variable)

DatasetLink:

https://drive.google.com/file/d/18y8jsq7MDivooVW4H_LqkBJ0YW0k4mR2/view?usp=sharing

The Restaurant Revenue Dataset is a comprehensive collection of simulated data designed for predicting monthly revenue for a set of restaurants.

Columns Included:

1. Number. of Customers
2. Menu Price
3. Marketing Spend
4. Cuisine Type
5. Average Customer Spending
6. Promotions
7. Reviews
8. Monthly Revenue

Pre-analysis using Excel:

https://docs.google.com/spreadsheets/d/1xyt7O9h1PFcno3z7zCrChijVTFHYe4_i/edit?usp=sharing&ouid=111232181404191323453&rtpof=true&sd=true

R code:

```
library(MASS)
data=read.csv("D:\\Restaurantrevenue_.csv")
y=data$Monthly_Revenue
x=data$Marketing_Spend
model=lm(y~x)
model
summary(model)
plot(model)
#training and testing model
install.packages("caret")
library(caret)
set.seed(123)
index <- sample(1:nrow(data), 0.8 * nrow(data))
train_data <- data[index, ]
test_data <- data[-index, ]
model1 <- lm(y ~ x, data = train_data)
predictions <- predict(model1, newdata = test_data)
predictions
mse <- mean((data$Monthly_Revenue-predictions)^2)
mse
```

Output:

```
lm(formula = y ~ x)
Residuals:
    Min      1Q  Median      3Q     Max
-274.186  -69.857    2.251   72.729  262.717
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 220.9067     6.2600  35.288   <2e-16 ***
x             4.8016     0.5422   8.856   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

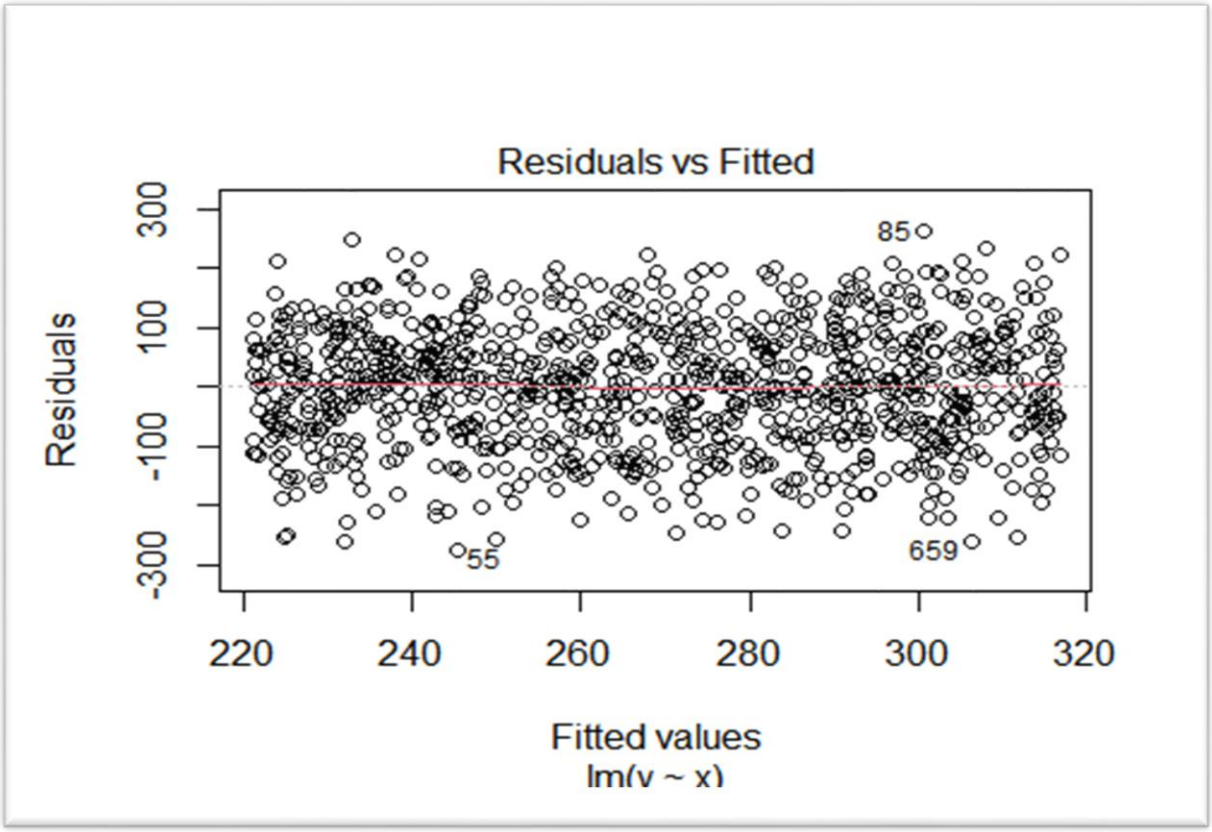
Residual standard error: 100.2 on 998 degrees of freedom

Multiple R-squared: 0.07286, Adjusted R-squared: 0.07193

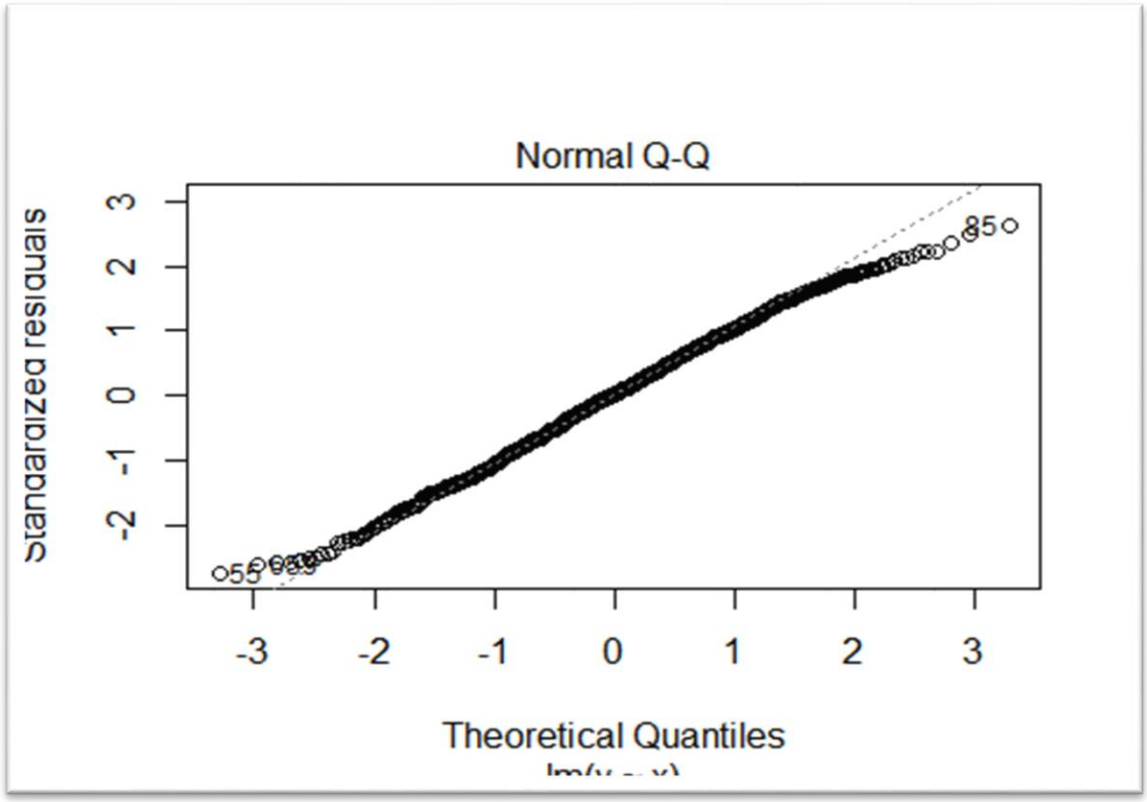
F-statistic: 78.43 on 1 and 998 DF, p-value: < 2.2e-16

Mean Squared Error: 10014.62

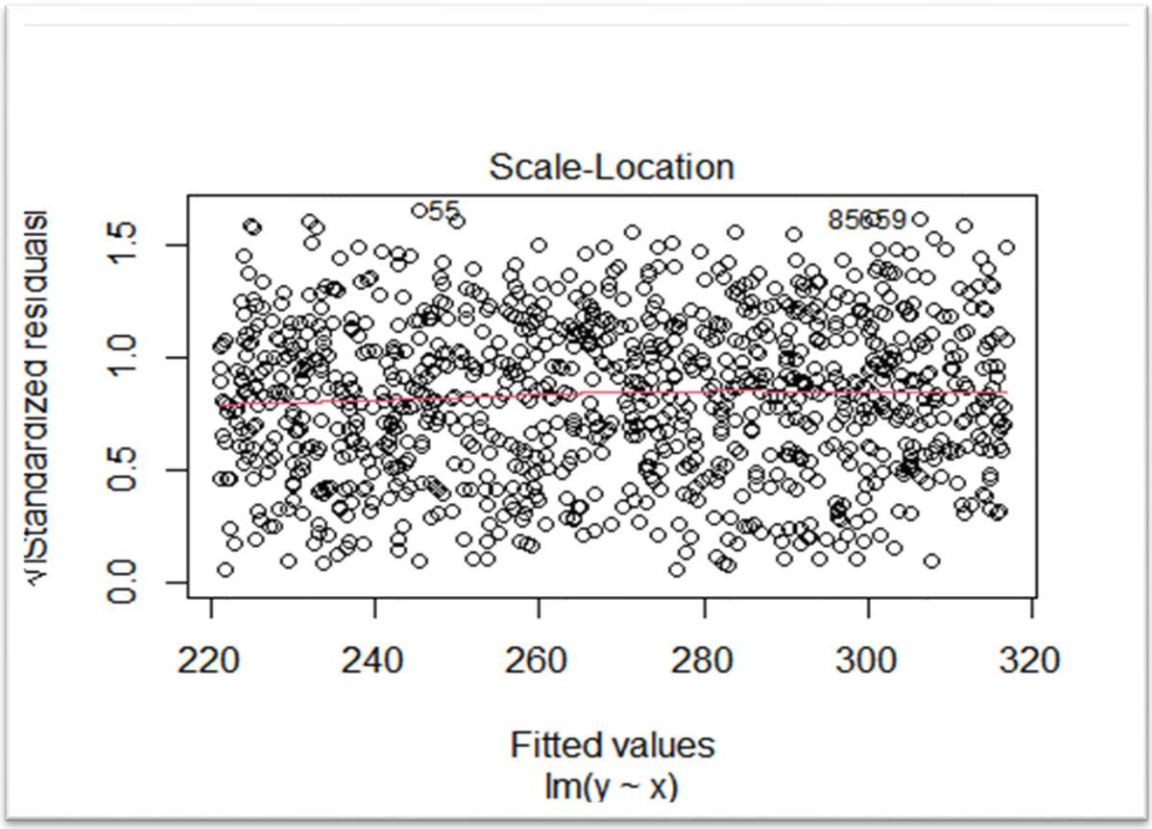
Plots:



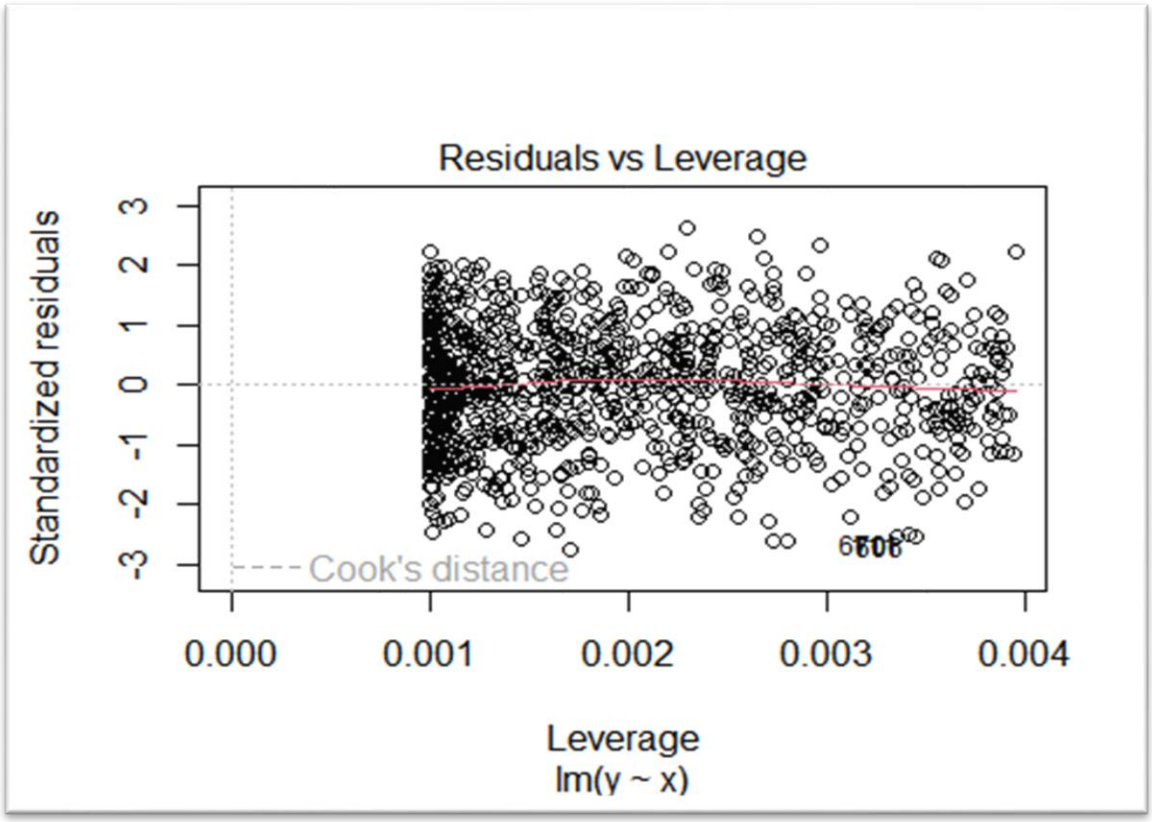
Interpretation: few outliers are present



Interpretation: Assumption of normality is satisfied



Interpretation: Heteroscedasticity is present.



Interpretation: There is no influence of outliers

Predicted Values:

Filelink:

https://drive.google.com/file/d/1oPi6EKPJ7OpBDmcJQxLJKn5W0AkccTAT/view?usp=sharing

Interpretation:

Mean Squared Error: 10014.62

Since the mean squared error is too large which suggests there is a large difference between actual and predicted value .

Also the R-square value= 0.07 which suggests that the model is not good fit.