```
> library(readr)
> original <- read_csv("D:/ad_campaigns_data.csv")
> View(original)
> summary(original)
       X1               TV              Radio          Newspaper          Sales      
 Min.   :  1.00   Min.   :  0.70   Min.   : 0.000   Min.   :  0.30   Min.   : 1.60  
 1st Qu.: 50.75   1st Qu.: 74.38   1st Qu.: 9.975   1st Qu.: 12.75   1st Qu.:10.38  
 Median :100.50   Median :149.75   Median :22.900   Median : 25.75   Median :12.90  
 Mean   :100.50   Mean   :147.04   Mean   :23.264   Mean   : 30.55   Mean   :14.02  
 3rd Qu.:150.25   3rd Qu.:218.82   3rd Qu.:36.525   3rd Qu.: 45.10   3rd Qu.:17.40  
 Max.   :200.00   Max.   :296.40   Max.   :49.600   Max.   :114.00   Max.   :27.00  


> library(dplyr)
> sample <- sample_frac(original, 1.5, replace = TRUE)
```

Fit an OLS model and answer the specific questions based on the data problem:

```
> ols <- lm(Sales ~ ., data = sample)
> summary(ols)

Call:
lm(formula = Sales ~ ., data = sample)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.7904 -0.8167  0.2583  1.1790  2.9614 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.833e+00  3.161e-01   8.962   <2e-16 ***
X1          -7.214e-05  1.698e-03  -0.042    0.966    
TV           4.540e-02  1.115e-03  40.723   <2e-16 ***
Radio        1.904e-01  7.041e-03  27.037   <2e-16 ***
Newspaper   -4.238e-04  4.942e-03  -0.086    0.932    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.668 on 295 degrees of freedom
Multiple R-squared:  0.8957,      Adjusted R-squared:  0.8943 
F-statistic: 633.1 on 4 and 295 DF,  p-value: < 2.2e-16
```

1. Is there a relationship between Advertisement Expenditure and Sales?

The relationship between Advertisement Expenditure and Sales can be seen from the F-test (which tests for a relationship between all X variables and the Y variable). The p-value for the F-test is `< 2.2e-16`. At a significance level of 0.05, the relationship between all types of advertisement expenditure and Sales is significant.

Oishi Bhattacharyya
PRN: 14060321050

2. How strong is that relationship between Advertisement Expenditure and Sales??

```
Adjusted R-squared:  0.8943
```

From this value, we can understand that the relationship (correlation) between Advertisement Expenditure and Sales is pretty high (1 being the highest). There is a 89.43% chance that Advertisement Expenditure and Sales are related.

3. Which Advertisement types contribute to sales?

This can be seen from the result of the t-tests for each X variable. At a significance level (alpha) of 0.05, `TV expenditures (<2e-16)` and `Radio expenditures (<2e-16)` have significant contribution since their p-values are less than alpha. However `Newspaper expenditures` with a p-value of `0.932` (much higher than alpha) makes no significant contributions.

4. What is the (Positive/Negative) effect of each Advertisement type of sales?

`TV` `4.540e-02` (for every 1 unit change in TV expenditure, there is approximately 4.540e-02 units change in Sales. Positive effect.)

`Radio` `1.904e-01` (for every 1 unit change in Radio expenditure, there is approximately 1.904e-01 units change in Sales. Positive effect.)

`Newspaper` `-4.238e-04` (for every 1 unit change in Newspaper expenditure, there is approximately -4.238e-04 units change in Sales. Negative effect.)

5. Using the Regression model you have fitted, given Advertisement spending in a particular market, can sales be predicted?

The general form of OLS states: y = XB + e. If the advertisement spending in a particular additional market is given, sales can be predicted by using the new X variable.

6. What will be the formula for predicting sales using Advertisement spending ?

We have 3 advertisement spending columns in this dataset. If we were have a 4th column, then the Ordinary Least Squares regression would look like: $\boldsymbol{y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4}$

In this particular case, $\boldsymbol{y = 4.540e\text{-}02X_1 + 1.904e\text{-}01X_2 - 4.238e\text{-}04X_3}$

7. Test if there is any multicollinearity among the independent variables

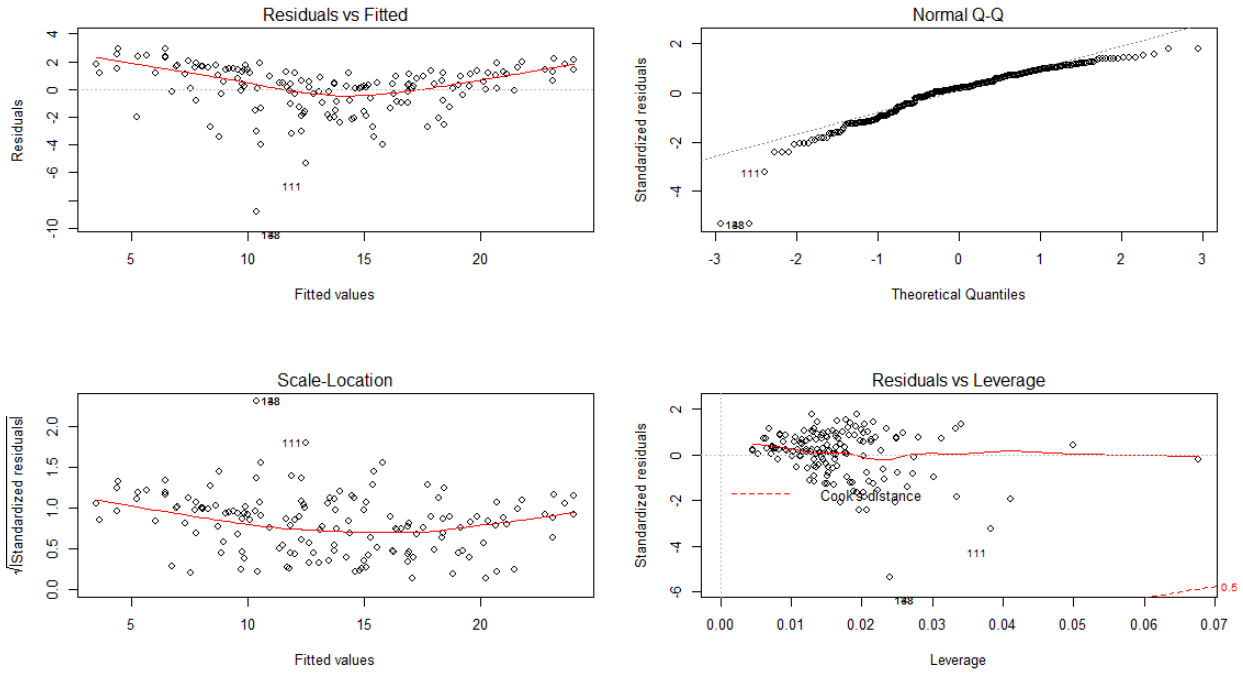
```
> library(car)
> vif(ols)
       X1        TV     Radio Newspaper
 1.029512  1.014044  1.161123  1.167636
```

The Variance Inflation Factors for all X variables are below 2 (low), therefore there is no multicollinearity among the independent variables.

Oishi Bhattacharyya
PRN: 14060321050

8. Look at the regression plots and give a comment on the model fit

```
par(mfrow=c(2,2))
plot(ols)
```



There are a few point falling beyond Cook's distance, which is not desirable. However the lines of best fit are fairly straight. In order for the model to fit properly, we need to take out the insignificant variable, which would further straighten out the line.

9. Check if any of the Assumptions of linear regression might have got violated based on the regression plots

In the top-left and bottom-left graphs, the line of best fit is fairly straight without any sharp dips, therefore the data is homoscedastic.

From the Q-Q Normal graph, we can see that the data is aligned to the Normal Distribution and doesn't deviate from it sharply. Therefore, the residuals follow a normal distribution.

Assumptions of linear regression are not violated.

Oishi Bhattacharyya
PRN: 14060321050