Report

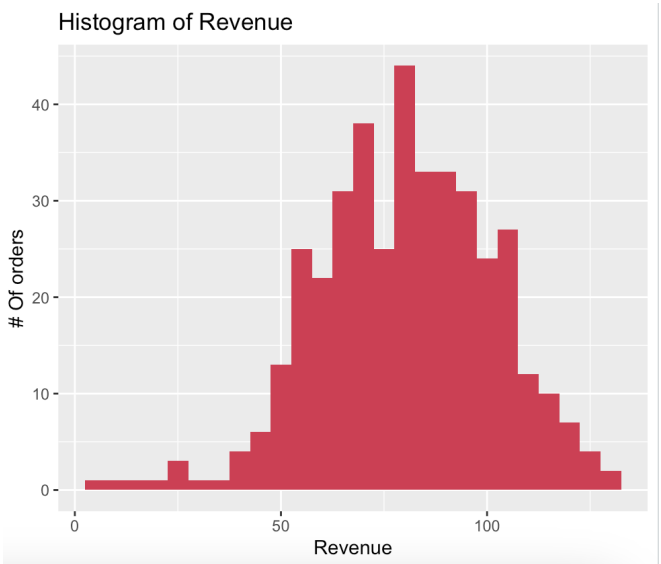
Drinks@home.uk is organizing a marketing campaign and would like to get a better understanding of the customers to increase the revenue from the sale of alcoholic and non-alcoholic drinks. Data on 400 customers is provided. The focus is to understand which of the factors positively or negatively impact the revenue generated from the orders. To understand this, the relationship between the factors i.e., the customer demographic and the revenue must be obtained. Also, a recommendation is required on which of the three proposed choices of marketing campaign would increase the profits of the company.

Task 1

The dataset provided has 400 observations of customer orders.

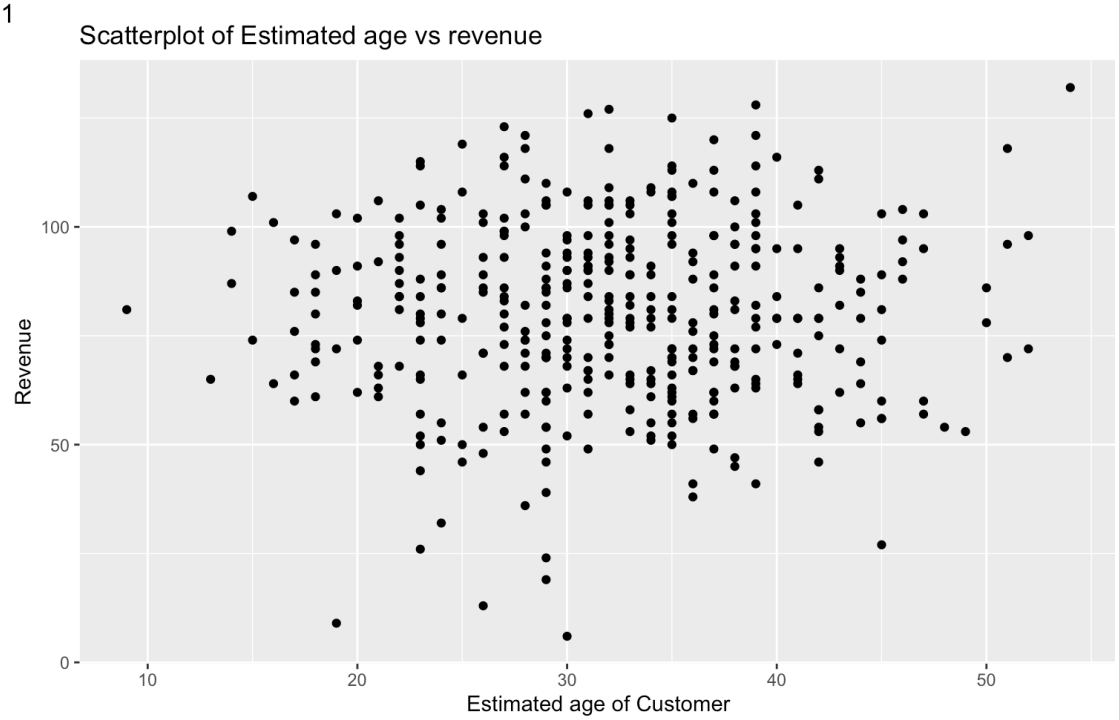
- The Revenue is the revenue in GBP from the most recent order made by the customer.
- Estimated_Age is the age of the customer estimated by the software.
- Time_On_Site is the time in seconds spent by the customer on the website per week.
- Seen_Voucher is a categorical variable that shows if the customer has come across a discount voucher pop-up. It has two values 0 and 1.
- Estimated_Income is the income of the customer in GBP estimated by the software.
- Advertisement_Channel is a categorical variable indicating the means of advertisement that led the customer to the website. (1-Leaflet, 2-Social media, 3-Search Engine, 4-Influencer).

As the objective is to find what factors positively or negatively impact the spending of customers, it is appropriate to apply multiple linear regression. Multiple linear regression allows an understanding of the influence of two or more independent variables on the dependent variable. This also enables to explain and identify which marketing campaign to recommend to ensure an increase in profits. The Revenue is the dependent variable as it is the variable of interest and the rest of the variables are considered independent variables.



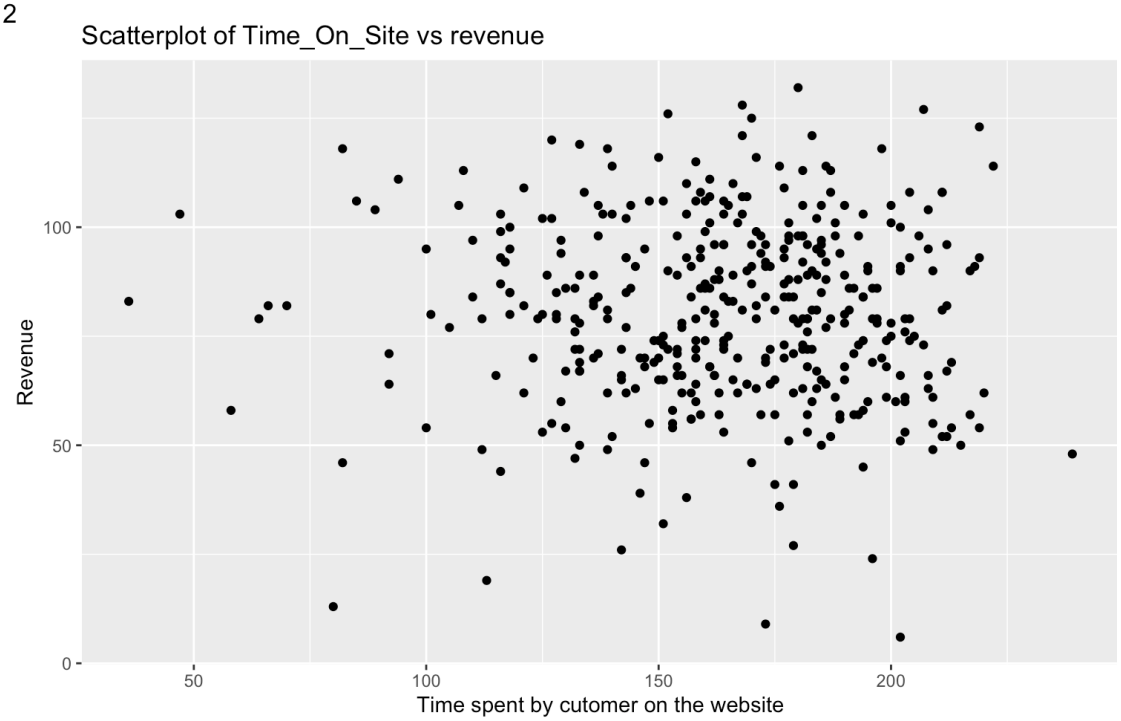
The histogram above gives the no. of orders distributed over the range of revenues. It is observed that the number of orders is larger for revenues between 50 to 100 GBP. The effect of each scalar independent variable on the dependent variable is visualized in the plots below.

Figure 1



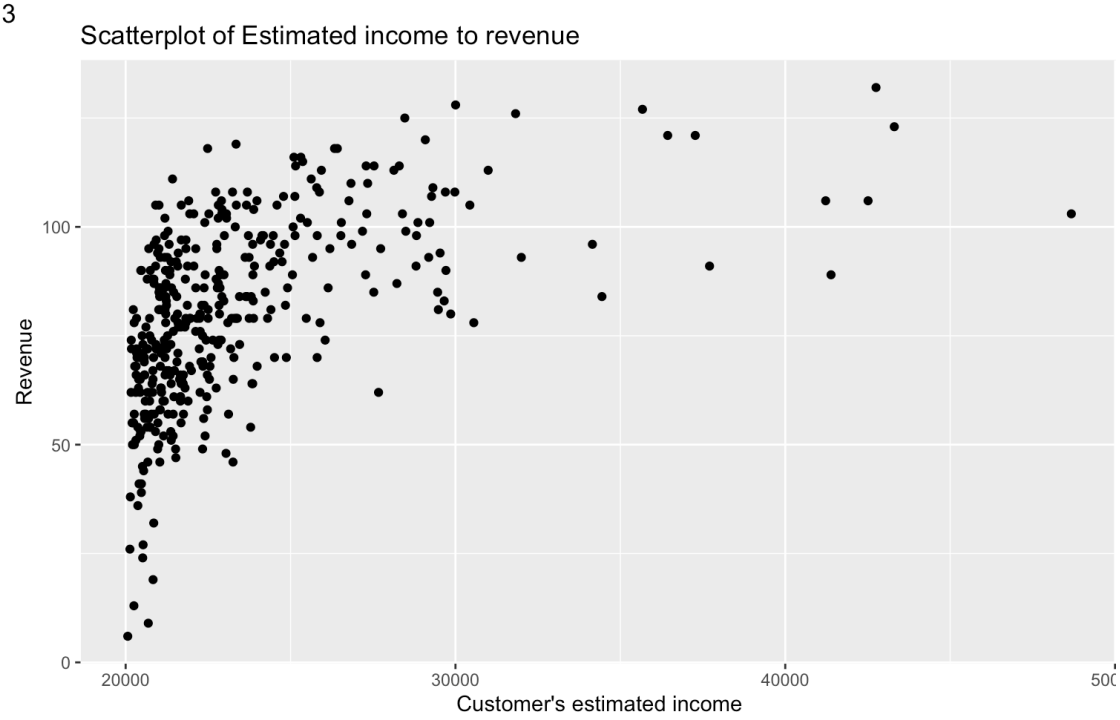
From Figure 1, the relationship between Age and Revenue is not very clear.

Figure 2



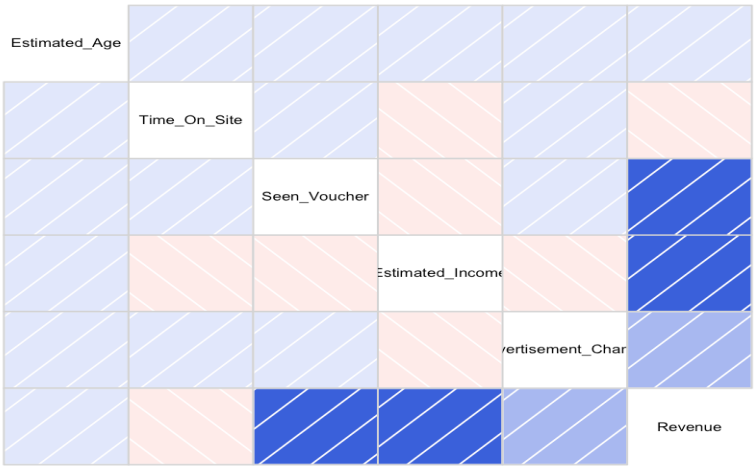
From Figure 2, the relationship between Time on site and Revenue is not evident.

Figure 3



From Figure 3, it is observed that increase in the income of the customer increases the revenue.

To better understand the relationship between the variables in the dataset proceed to check the correlation of the dataset. In the image below, shades of blue indicate a positive correlation and red indicates a negative correlation. Darker shades indicate a stronger correlation. It is evident that Seen_Voucher and Estimated income have a strong positive correlation of 0.466 and 0.532 respectively on the Revenue. The advertisement channel has a moderate positive correlation of 0.212 with revenue. Age and time spent on site have a very low correlation with the revenue. There is no high correlation among the independent variables, hence multicollinearity is avoided.



The values of correlation among the variables are as in the below image.

	Estimated_Age	Time_On_Site	Seen_Voucher	Estimated_Income	Advertisement_Channel	Revenue
Estimated_Age	1.00000000	0.02774864	0.011600588	0.014782713	0.09709081	0.02628325
Time_On_Site	0.02774864	1.00000000	0.050760998	-0.038676812	0.02294742	-0.02826230
Seen_Voucher	0.01160059	0.05076100	1.000000000	-0.009811076	0.02013695	0.46597667
Estimated_Income	0.01478271	-0.03867681	-0.009811076	1.000000000	-0.02015121	0.53165184
Advertisement_Channel	0.09709081	0.02294742	0.020136949	-0.020151214	1.00000000	0.21240069
Revenue	0.02628325	-0.02826230	0.465976667	0.531651842	0.21240069	1.00000000

With the above understanding, visualizing the effect of the categorical variables on the revenue since seen voucher and advertisement channel have a positive correlation. The voucher seen is split into categories of whether the voucher is seen by the customer or not and the advertisement channel is split into categories by the different channels through which the customer has reached the website. Below are the scatterplots showing the relationship of the explanatory variables with revenue.

Figure 4

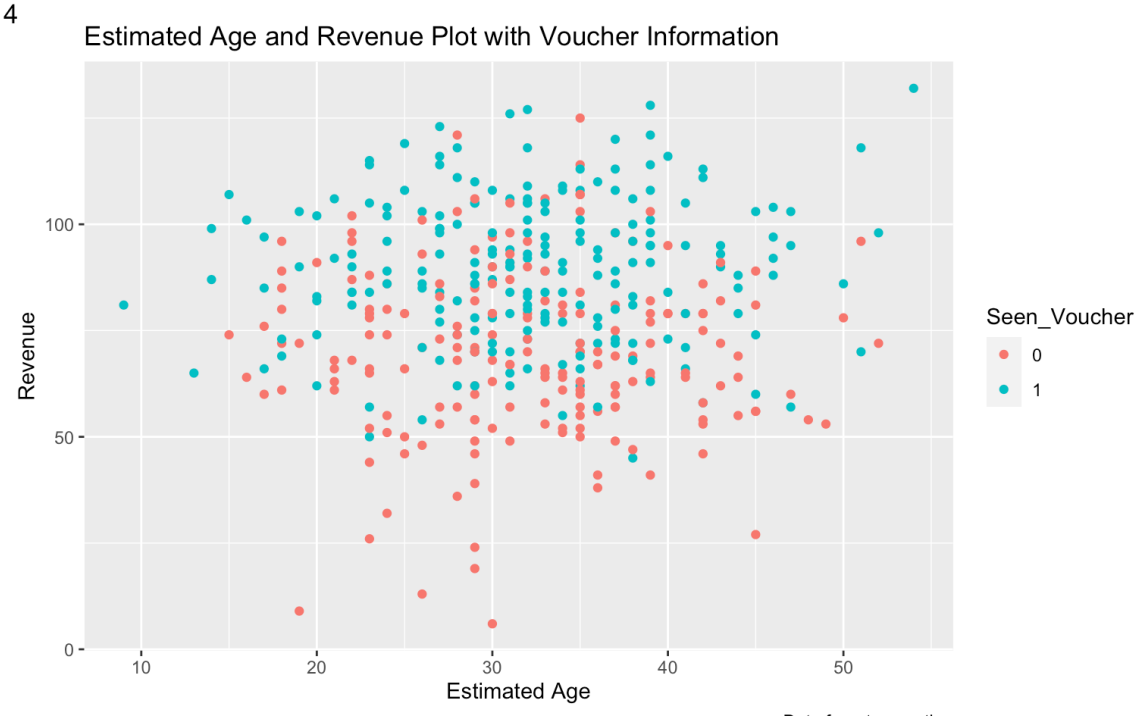


Figure 5

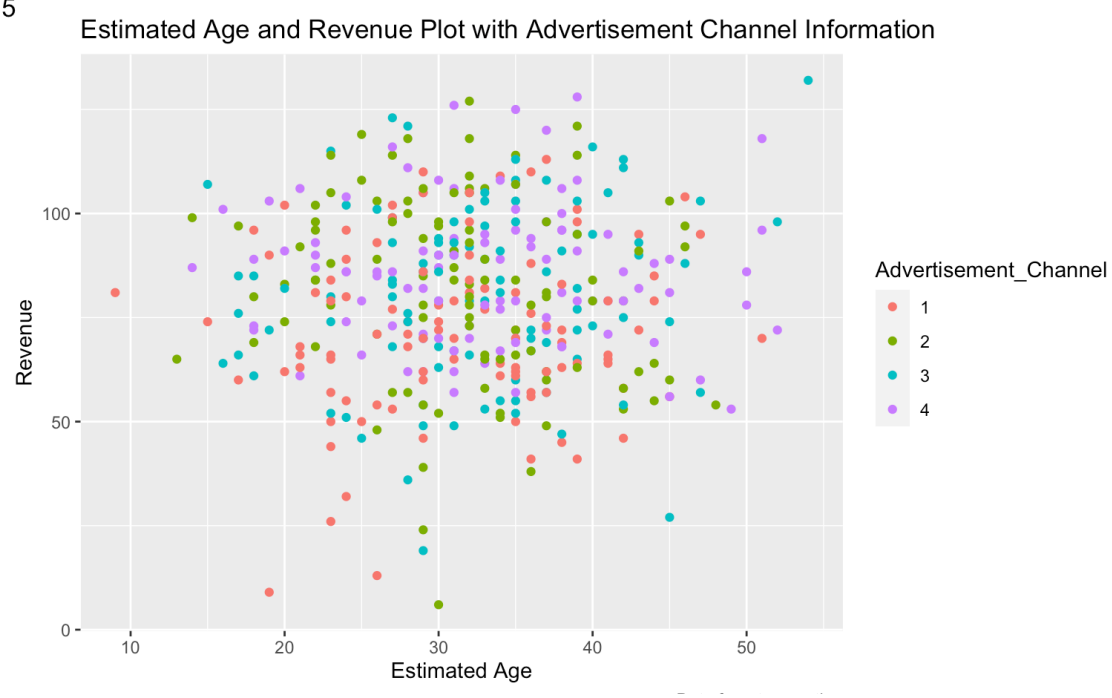


Figure 6

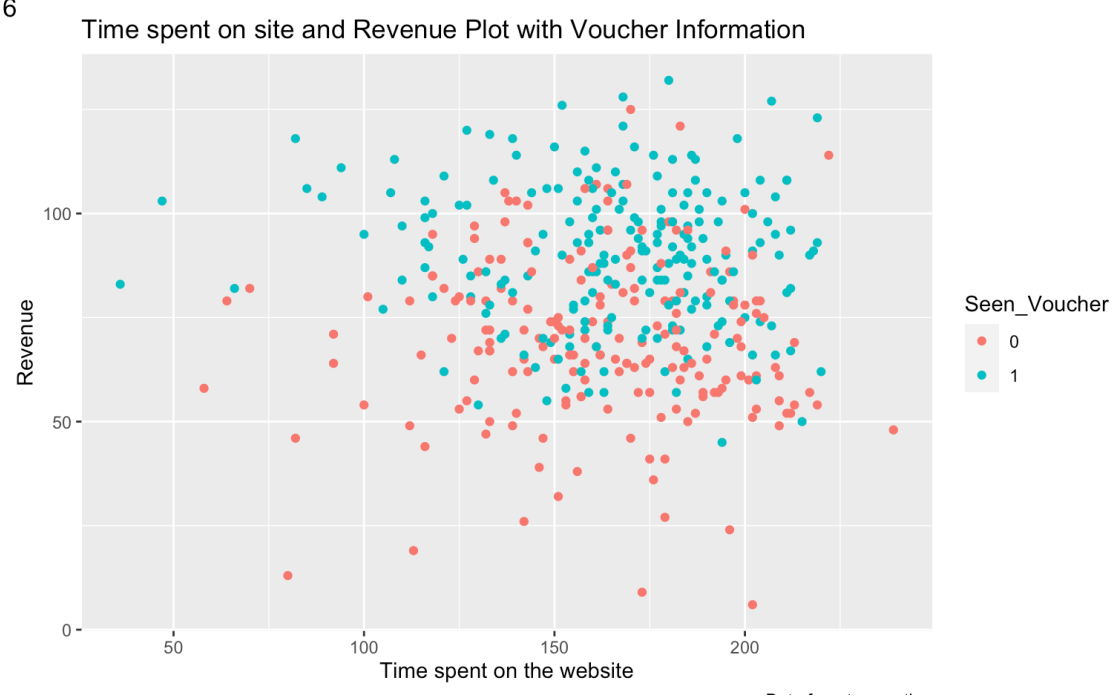


Figure 7

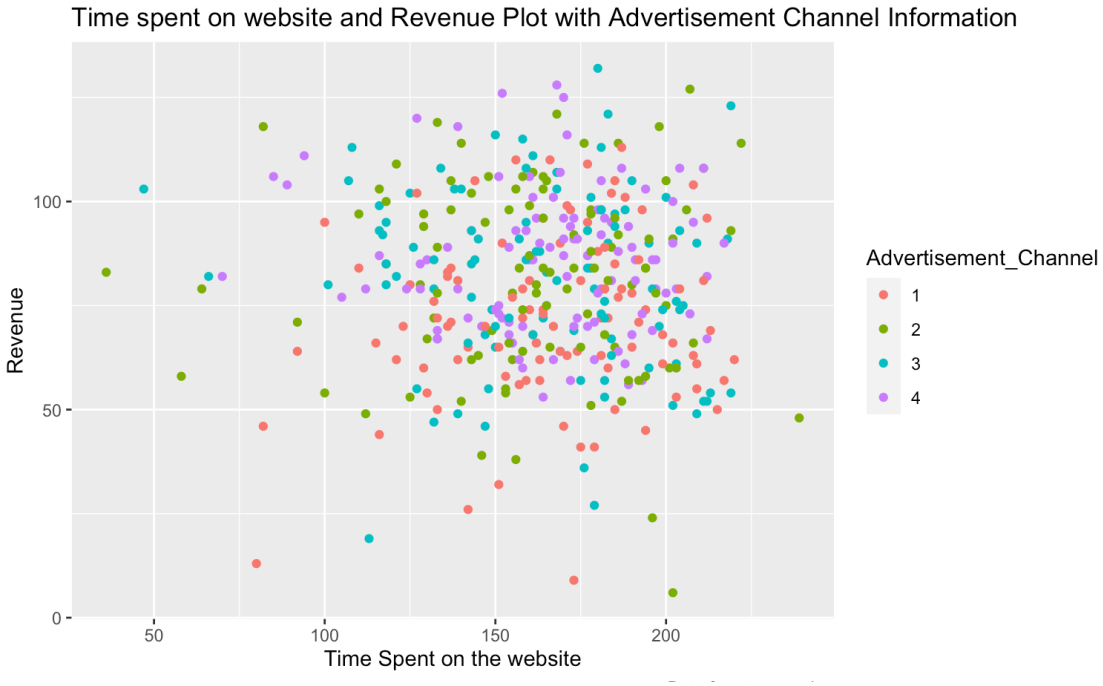


Figure 8

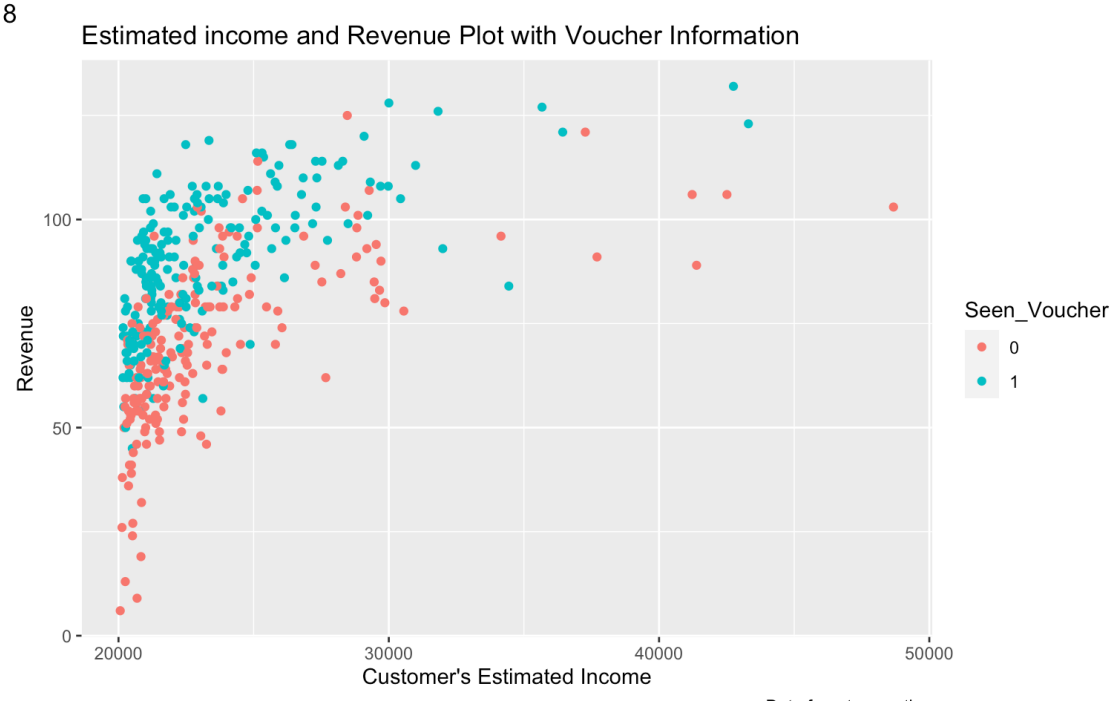
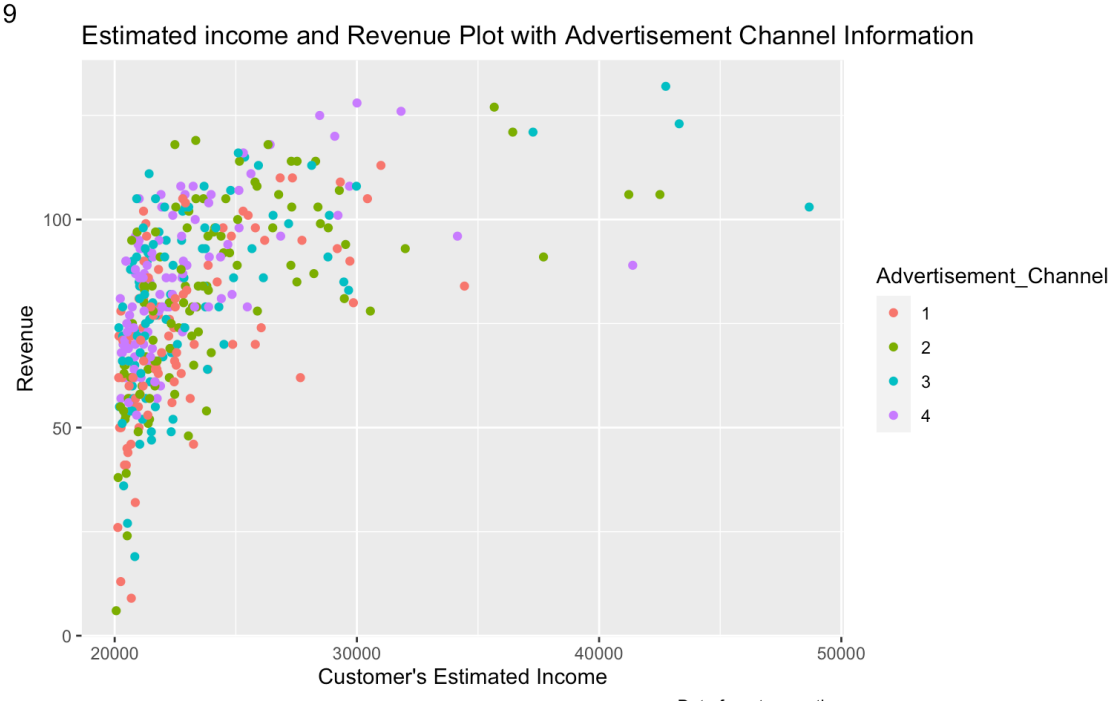
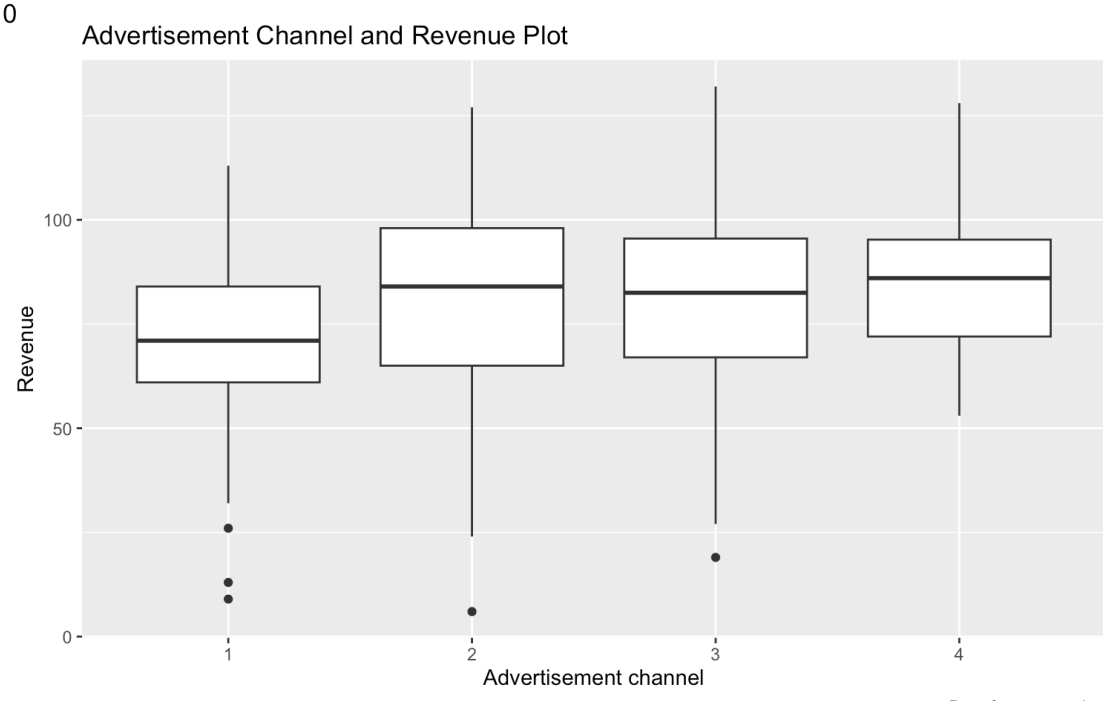


Figure 9



From the above plots Linearity is observed only between income and revenue, this verifies the value of correlation. The age and time on site do not show any linearity with the revenue. When SeenVoucher is 1, the revenue is higher than when it is 0, hence a positive relation can be observed. The impact of the advertisement channels is not very obvious from the above plots. Therefore, drawing a boxplot to check the relationship of advertisement_channel vs revenue. From Figure 10 it can be seen that the advertisement channel has a moderate positive relation with the revenue.

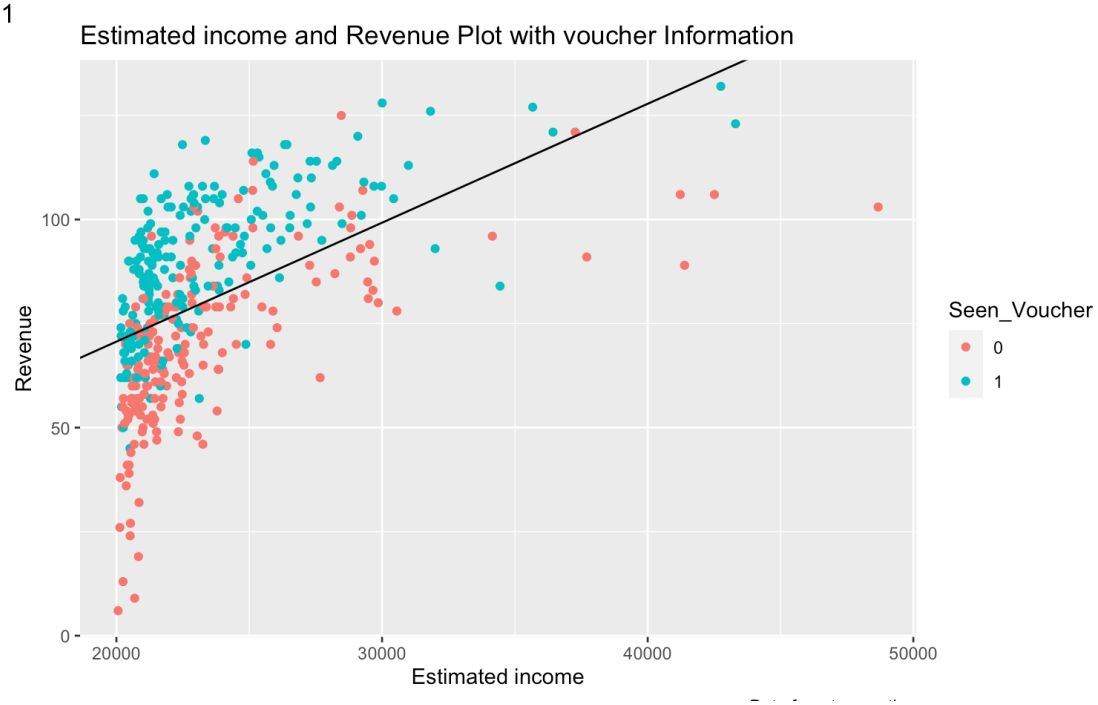
Figure 10



From the above, it can be concluded that income, voucher and advertisement channel satisfy linearity and the assumption that the variance of the dependent variable is independent of

the values of the explanatory variable which is very crucial while considering the variables for a linear regression. To check for the error of the points, visualize a line through the data points. From Figure 11 it is noticed that on average the error of the points is independent from each other.

Figure 11



Fitting the regression model for Income, Voucher Seen and Advertisement channel vs the revenue gives the following insights

```
Call:
lm(formula = Revenue ~ Estimated_Income + Seen_Voucher + Advertisement_Channel,
    data = dataplot)

Residuals:
    Min      1Q  Median      3Q     Max
-54.575  -7.738   1.029   8.861  38.841

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)            -3.8215992  4.4645012  -0.856 0.392520
Estimated_Income        0.0028656  0.0001834  15.622  < 2e-16 ***
Seen_Voucher1          19.6270625  1.4113549  13.907  < 2e-16 ***
Advertisement_Channel2  6.9039364  2.0132224   3.429 0.000669 ***
Advertisement_Channel3  8.1139387  1.9943405   4.068 5.72e-05 ***
Advertisement_Channel4 12.9103883  1.9901035   6.487 2.63e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.07 on 394 degrees of freedom
Multiple R-squared:  0.5534,    Adjusted R-squared:  0.5478
F-statistic: 97.66 on 5 and 394 DF,  p-value: < 2.2e-16
```

- Estimated income is statistically significantly related to the revenue (p-value<0.05) and it can be seen that an increase of 1 GBP in income increases the revenue by 0.002 GBP on average.
- Seen voucher is significantly related to the revenue with the revenue increasing by 19.627 GBP when the voucher is seen.

- Advertisement Channel 2 is significantly related to revenue and using advertisement channel 2 increases the revenue by 6.90 GBP on average with reference to using channel 1.
- Advertisement Channel 3 is significantly related to revenue and increases the revenue by 8.11 GBP on average with reference to channel 1.
- Advertisement Channel 4 is significantly related to revenue and increases the revenue by 12.91 GBP on average with reference to channel 1.
- Considering that channel 1 is reference channel with respect to the other channels, Impact of Advertisement channel 1 is inferred from the intercept and is not significantly related to the revenue.

From the above analysis, it can be concluded that estimated income, voucher seen and advertisement channels 2, 3 and 4 have a positive impact while advertisement channel 1, time on site and Estimated age of customer have no impact on the spending of the customer on the website.

Task 2

Drinks@home is looking to increase profits on the website and has three choices for the marketing campaign that are being considered.

1. Launch an advertisement targeting customer above the age of 45 considering the likelihood of them spending more.
   From the results of the analysis conducted it is inferred that the age of the customer has no impact on the revenue. Hence running a commercial targeting an age group would not aid in increasing profits.

2. Grant a voucher of 20 GBP off the next purchase to the customer.
   It is evident from the analysis and regression model summary that whether the customer has been presented with a voucher pop-up largely impacts the revenue. The revenue increases by 19.6 GBP depending on if the customer has seen the voucher. Hence providing vouchers to customers increases revenue.

3. Invest on influencer advertising
   The data provided has an equal distribution with 100 data points for each channel. Advertisement Channel 1 represents leaflets, 2 represents social media, 3 represents search engines and 4 represents influencers. From the analysis, we see that if the customer has reached the website through advertisement channel 4 which is influencer advertising, then the revenue increases by 12.91 GBP. Therefore, investing in Influencer Ads increases profits.

To conclude, the factors that positively impact the spending of customers are income, voucher information and advertisement channels 2, 3, 4. The estimated age and time spent on site do not have a significant influence on the revenue. given the choice of 3 campaigns, campaign 3 of investing in influencer advertising is is recommended. Although campaign 2 is a good option with a lower coupon amount, the current coupon offering of 20 GBP discount might result in a loss.