

Problem Statement: what makes a customer valuable?

Introduction:

In today's world, everything is digital and data reigns supreme. For businesses to flourish, they must keep up with the ever-evolving landscape of e-commerce.

[Drinks@home.uk](https://drinksathome.co.uk) is part of this chase for relevance and success, specializing in the direct-to-consumer business of both alcoholic and non-alcoholic beverages from all around the world to people in Great Britain. There is an urgent need to capitalize on consumer data, delve into factors that excite customers about the products, and figure out what qualities make a customer valuable.

Every purchase and every interaction on the website is a hint to solving this puzzle.

Key Objectives for this analysis are:

1) Customer Spending Habits

Analysing factors that contribute to both positive and negative effects on consumer spending. This involves explorative data analysis, deciphering correlations within the data and identifying key factors for increased revenue generation. while making a purchase.

2) Recommendation for Marketing Strategy

The marketing team has come up with three strategies for boosting sales:

1. Run an advertisement targeting customers who are older than 45 years old as they are likely to spend more money.
2. Provide a voucher for 20GBP on their next orders.
3. Spend more money on advertising with an influencer.

Based on the insights gained from the customer spending analysis, a recommendation is to be made for increasing revenue on the website.

Data Understanding and Preparation:

Data on 400 customers has been made available which contains information as shown in figure 1.

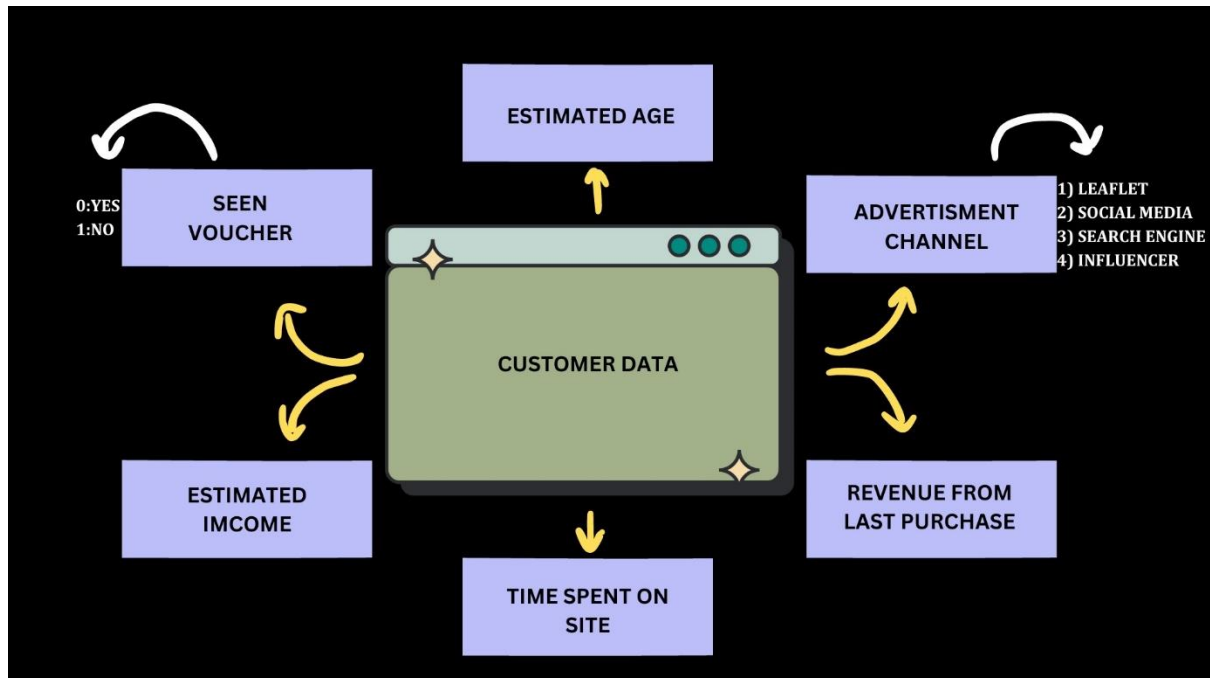


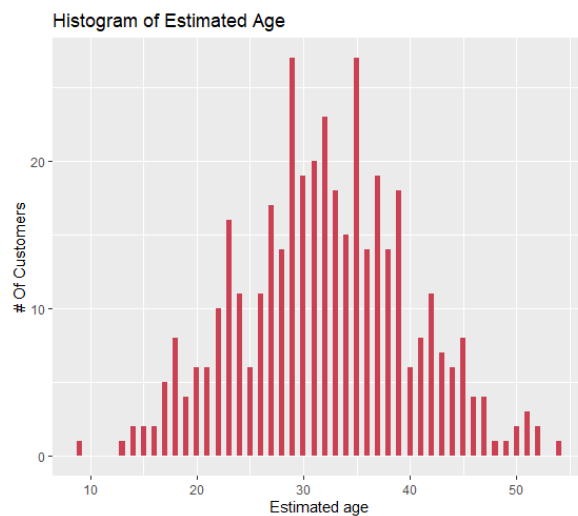
Figure 1: Features of Customer Dataset

- **Revenue (GBP)** – Revenue from the latest order by the customer.
- **Advertisement Channel** – Advertisement medium that brought them to the website:
 - 1: Leaflet
 - 2: Social Media
 - 3: Search Engine
 - 4: Influencer
- **Estimated Age** – Estimated age of the customer.
- **Estimated Income (GBP)** - Estimated income of the customer.
- **Time on website per week (Seconds)** – The estimated average time that the customer spends on the website per week.
- **Seen Voucher** – If the customer has seen any discount voucher popup.

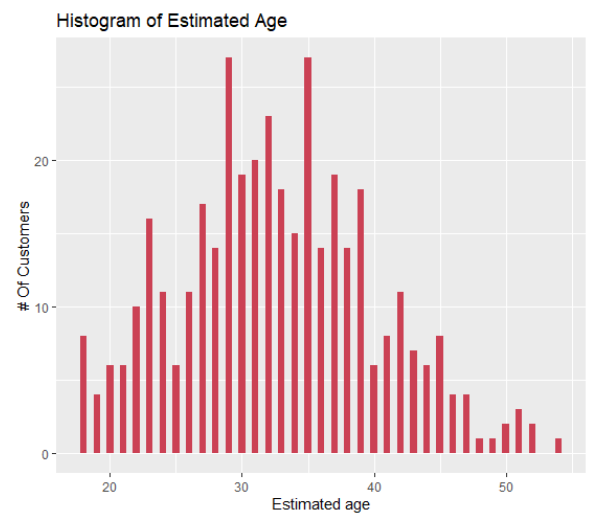
There are **no missing or special placeholder values** in the dataset, The unit of measurement is consistent across the dataset.

Explorative Data Analysis is done to identify general patterns in the dataset.

1) Age



(a) Age distribution of All 400 Customers



(b) Age distribution of all legally aged customers

Figure 2: Age Distribution of Customers (in years)

From Figure 2, It can be observed that the age of customers is between 9-54 years, and the average age is around 32.

Since the legal age for someone who can make an online purchase is 18 years, the removal of underaged customers is necessary for a fair analysis. **Hence, thirteen rows are removed.**

2) Seen Voucher:

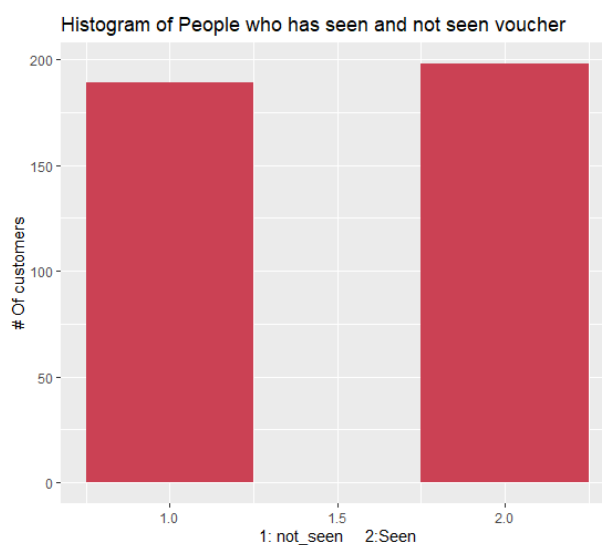
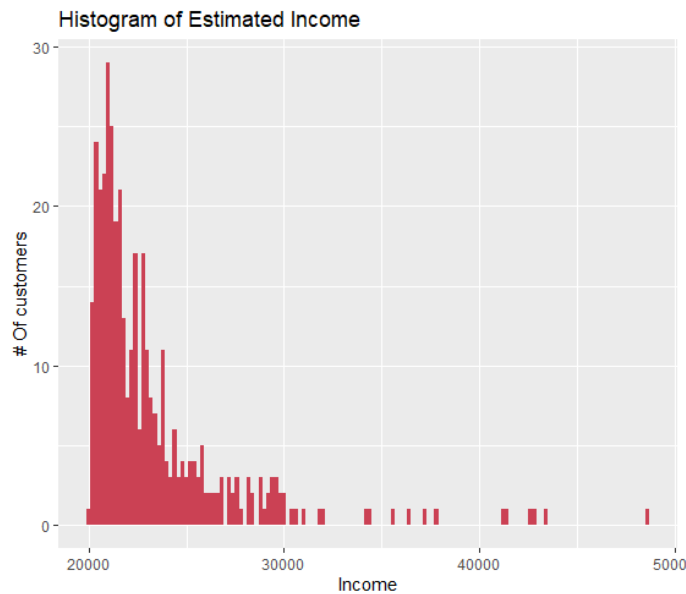


Figure 3: Voucher Pop up distribution

From Figure 3, it can be observed that 189 people have seen vouchers while 198 didn't.

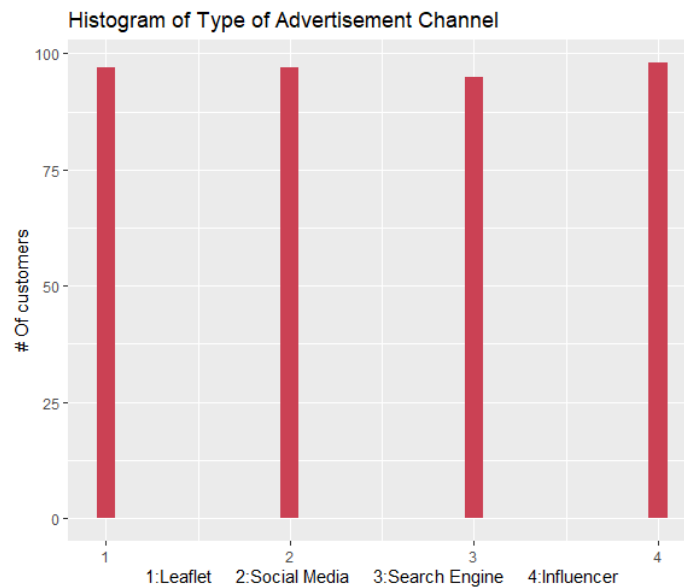
3) Estimated Income



From Figure 4, it can be observed that the Customer's Income lies between 20000-49000 GBP where the mean and median income is 23000 and 21000 GBP and most of the customers lie in the 20000-30000 bracket.

Figure 4: Income Distribution of Customers

4) Advertisement Channel



From Figure 5. Leaflet, social media, Search Engine, and Influencer brought 97,97,95,98 customers respectively to the website.

Figure 5: Advertisement channel Spread

5)Revenue

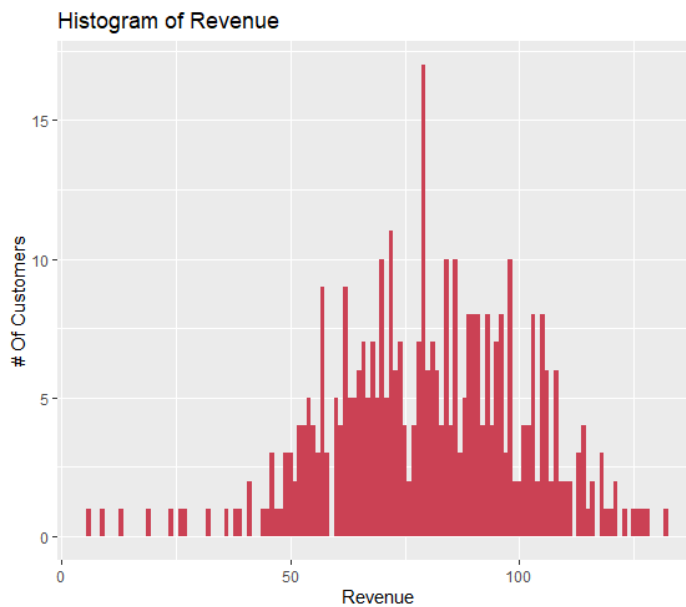


Figure 6: Revenue Distribution

From Figure 6, it can be observed that customers have spent 6-132 GBP on the website and the mean revenue generated per customer is around 80 GBP.

6) Time Spent By customer on the website:

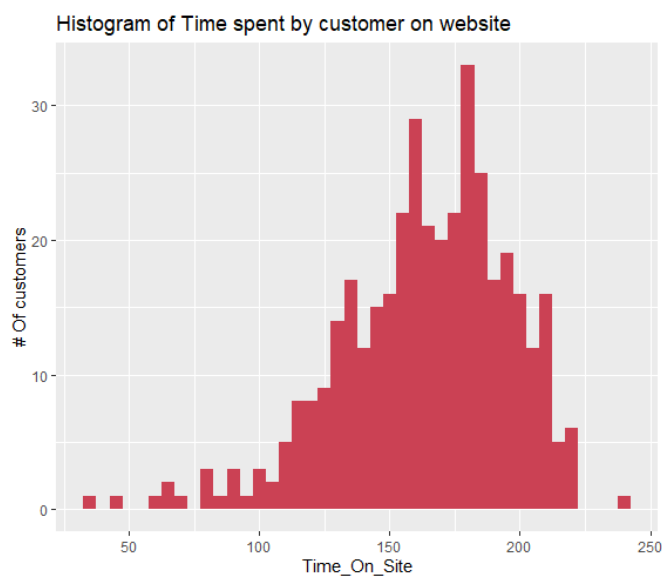


Figure 7: Distribution of Time spent by customers

Most of the customers spent 100-220 seconds on the website.



In Figure 8, Dark shades indicate a strong correlation, blue indicates positive and red indicates negative between two Features.

Strong positive correlations between revenue and estimated income and voucher, a weak positive correlation between revenue and estimated age and a weak negative correlation between revenue and time spent on a website.

Figure 8: Correlation between Features

*As the purpose of this analysis is to measure the impact of all these variables on revenue, **Linear regression can be a powerful tool for the same.***

Modelling

Linear Regression can be used to identify what variables influence revenue more strongly and in what direction (positively or negatively) and which do not influence it at all.

Components of linear Regression Model:

$$y = \alpha + \beta x + \epsilon$$

y = Dependent Variable
 α = Intercept / constant
 β = Slope / coefficient of X
 x = Independent Variable
 ϵ = Noise / error

Figure 9: Linear Regression Equation

The Dependent variable **Y** is **Revenue**, and the independent variable **X** is **Age, Income, Time spent on the website, Advertisement channel, and if the Voucher is seen.**

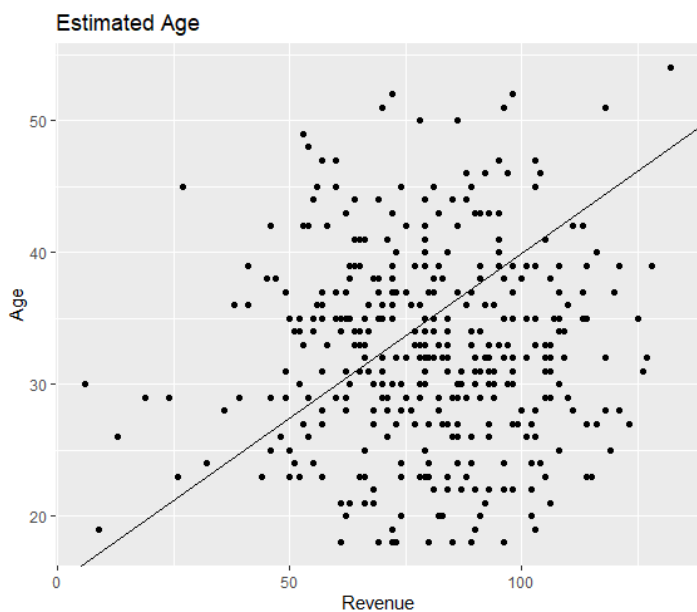
With linear regression, we try to fit a line on the data by estimating alpha and beta that minimises the distance of data points from the line and that's how a relation between the two variables X and Y is obtained.

To implement linear regression, certain conditions are required to be satisfied:

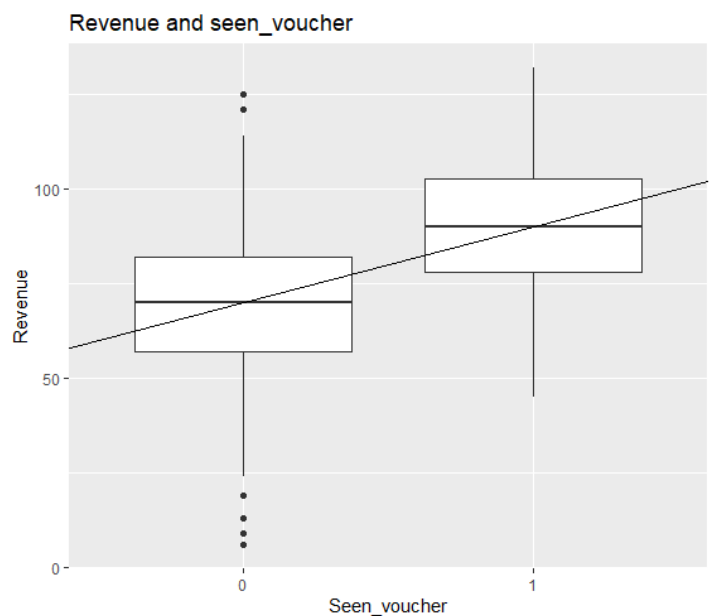
Condition 1: The expected value of the dependent variable(Y) has a linear Relation with the exploratory variable (X).

Condition 2: The variance of the dependent variable should be the same for all the values of the exploratory variable.

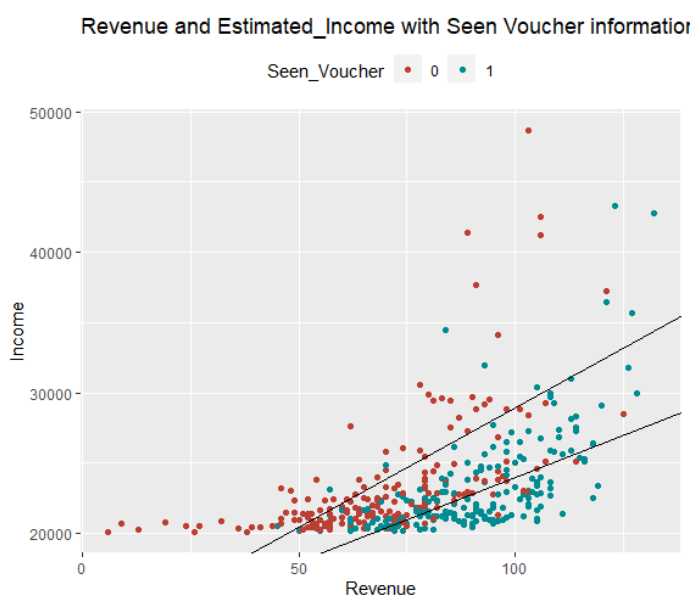
Condition 3: The distance of each point should be independent of the error of the points.



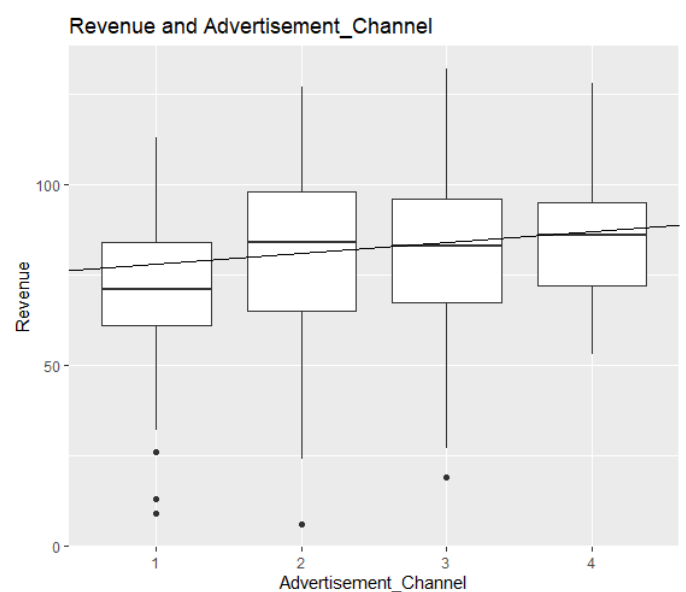
(a)



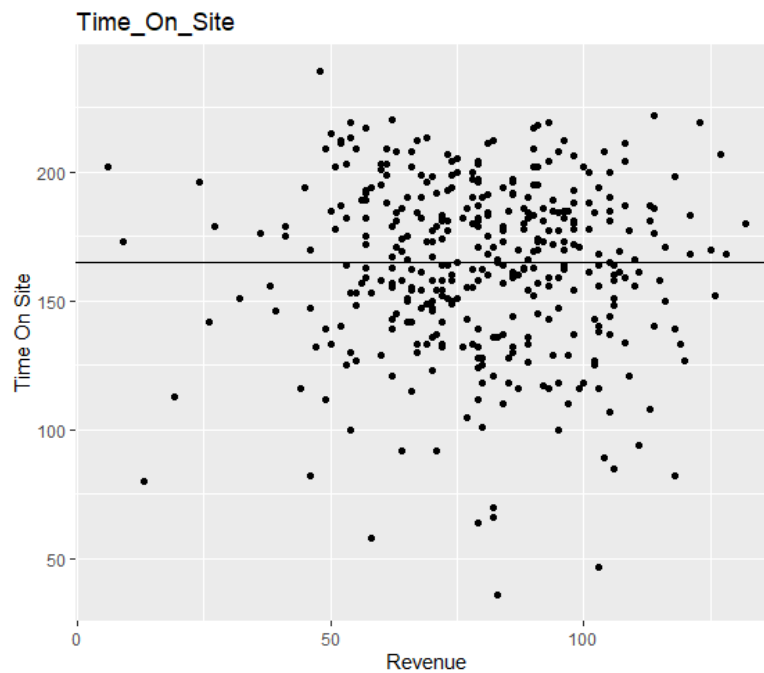
(b)



(c)



(d)



(e)

Figure 10: Conditions check for all independent variables:

All the features satisfy all conditions for linear regression as shown in Figure 10; Hence Linear Regression can be used to find the dependence of revenue on variables.

Modelling linear regression on the dataset we observe, only seen voucher, income and advertisement channels are statistically significant and can influence revenue generation on the website.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -8.3224892   4.7499497  -1.752   0.0806 .
Seen_Voucher    19.7143725   1.4436639  13.656 < 2e-16 ***
Estimated_Income  0.0029163   0.0001843  15.820 < 2e-16 ***
Advertisement_Channel 4.1071266   0.6436522   6.381 5.08e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.19 on 383 degrees of freedom
Multiple R-squared:  0.5505,    Adjusted R-squared:  0.547

```

Figure 11: Summary of Linear Regression Model with its parameters and model performance statistics

We obtain **t and p-values** of the variable from which it can be inferred that the **income of the customer is the most influential factor**, followed by voucher and advertisement channel.

T-test is used to check if coefficients are statistically equivalent to zero or not, if equivalent to zero, it means the exploratory variable does not influence the dependent variable and should be removed from the model and

The P-value is 1- confidence associated with a particular t-test, therefore $p\text{-values} < 0.05$ means the coefficient is statistically relevant, otherwise can be removed.

T-test value of age and time spent on site are beyond the threshold of the t-test i.e.

$$|t\text{-value}| < 1.96$$

They do not influence revenue generation on the website.

Objective 2:

The three marketing strategies proposed by marketing teams are:

1. *Run an advertisement targeting customers who are older than 45 years old as they are likely to spend more money.*

According to this strategy, the company should focus on customers older than 45 years because they are more likely to spend more money. T-test reveals that age is not a statistically relevant factor for predicting revenue on the website. Even so, the data available for analysis contains only 18 customers older than 45. So, the dataset is skewed towards younger audiences, hence generalising based on this dataset may not be wise, hence this strategy is not endorsed.

2. *Provide a voucher for 20 GBP on their next orders.*

If we compare customers who have seen the voucher and those who haven't, there is a clear distinction, Customers who have seen vouchers have spent more money with the mean and median revenue of **90 pounds as compared to 70 pounds** by customers who haven't seen the voucher.

So, providing a voucher does make sense and can be a reliable strategy to increase sales on the website. But a voucher of value 20 GBP might be on the higher end as this is 20 pounds out of profit.

The data available doesn't state the voucher value or if the voucher was used or not, so providing a voucher to attract customers can be a good strategy but voucher value should be considered again.

3. *Spend more money on advertising with an influencer.*

If we compare the advertisement channels, customers attracted by influencers are generating max revenue with mean and median revenue values of **85 and 86 pounds** but both social media and search engines are bringing customers with a mean revenue of **82 pounds** and median revenue of **84 and 83** pounds respectively. It's just the leaflet channel that is performing weak. Though influencers are bringing the maximum revenue out of all channels, the difference is 4 pounds between influencers and the other two channels. **Therefore, spending more money on advertising with an influencer might help but any other channel other than leaflet may produce the same impact.**

Conclusion:

After studying past consumer behaviours on our website, it can be concluded that out of all the features, **visible vouchers, the income of the customer and the advertising channels** are the most relevant factors for generating more sales. To boost sales, the best strategy could be to provide users with vouchers as they entice them to buy more and lead to the generation of more revenue on the website.

