

Case Study

Somyanath Mohanty

7th October 2018

Marketing Campaigns

Introduction

Problem Statement

The problem statement here is to provide detailed insights into the different marketing campaigns in one market. The insights to be presented are:

- a. Overview of the development of the entire market and the different campaigns.
- b. Assess the development of the quality of traffic. How is the overall development and how does each campaign evolve?
- c. A business developer wants to spend an additional 250€ per week from week 31 onwards.
 - Advise the business developer in which campaign to invest and why?
 - How does this impact the overall performance in the market from week 31 onwards?

Data

****First of all, on checking the data it is found that there is a redundant data entry for the week-30 Cottington campaign which should be deleted before starting the analysis.**

Given below is a sample of the dataset that has been used to perform the analysis on:

Week	Campaign	Visits	Revenue	Cost
1	Aldebaran	27	2.270	3.764
2	Aldebaran	64	15.323	15.323
3	Aldebaran	80	10.754	10.754
4	Aldebaran	93	16.906	16.906
5	Aldebaran	120	21.447	21.447

We have the following 5 variables:

1. **Week:** the number of the week
2. **Campaign:** the name of the campaign
3. **Visits:** number of visitors to a campaign in a week
4. **Revenue:** revenue of a campaign in a week
5. **Cost:** the investment in a particular campaign in a week

A new variable is also created in order to garner more insight from the data given to us:

- **Profit:** Profit (Revenue-Cost) generated by a particular campaign in a week

Data Analysis

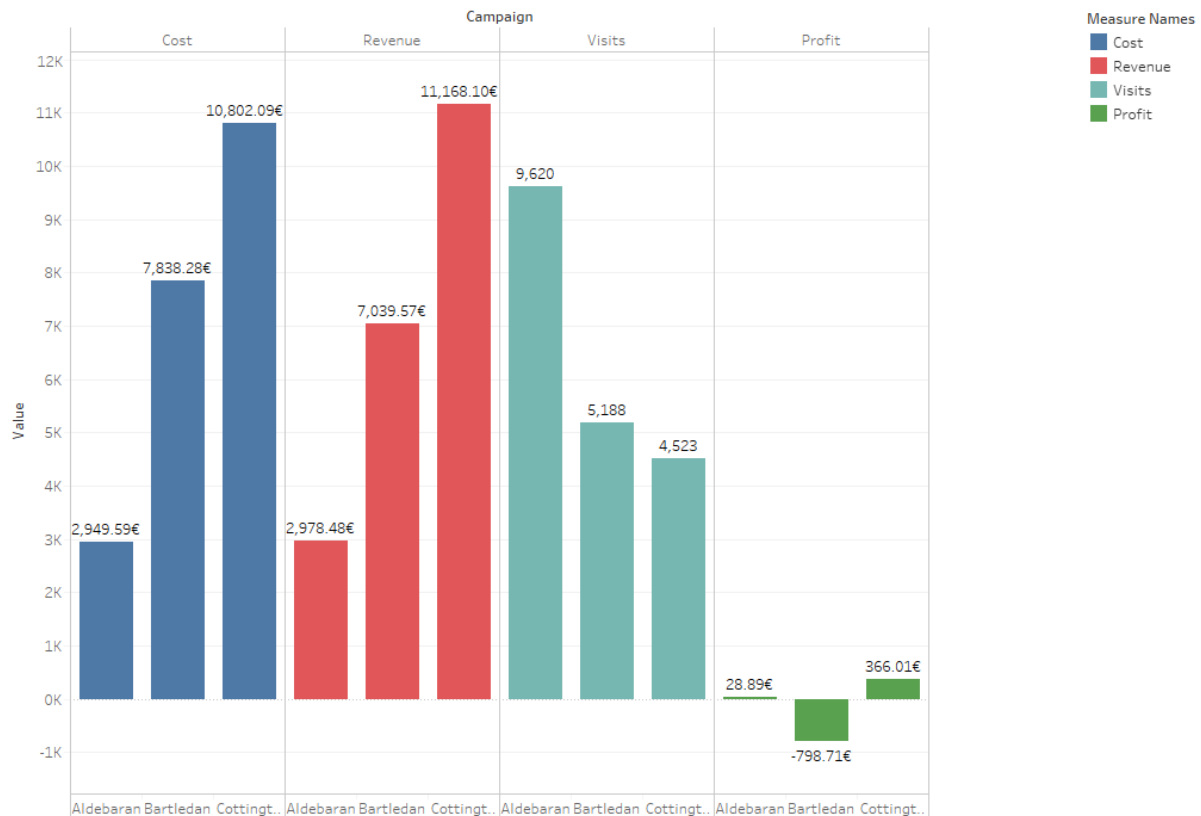
Some basic info about the content of the data:

1. The marketing campaigns data is of 30 weeks
2. There are 3 campaigns: **Aldebaran**, **Bartledan**, **Cottington**

Problem Statement a:

Overview of the development of the entire market and the different campaigns.

Firstly, we should study the total statistics of the campaigns over the whole course of 30 weeks.



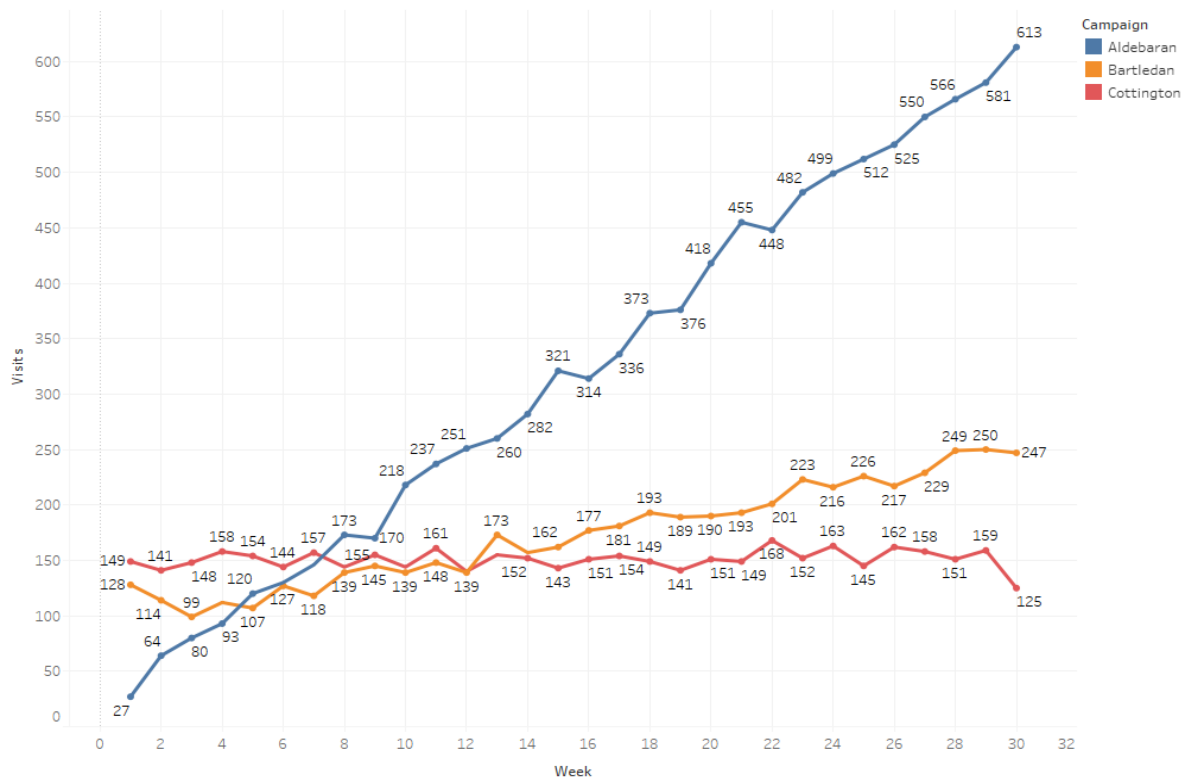
From the above plot, we observe the following about the campaigns:

- Aldebaran generated the highest number of visitors followed by Bartledan, and then Cottington
- Cottington generated the highest revenue followed by Bartledan, and then Aldebaran
- Cottington incurred the highest spending (cost) followed by Bartledan, and then Aldebaran
- Cottington generated the highest profit followed by Aldebaran, and then Bartledan

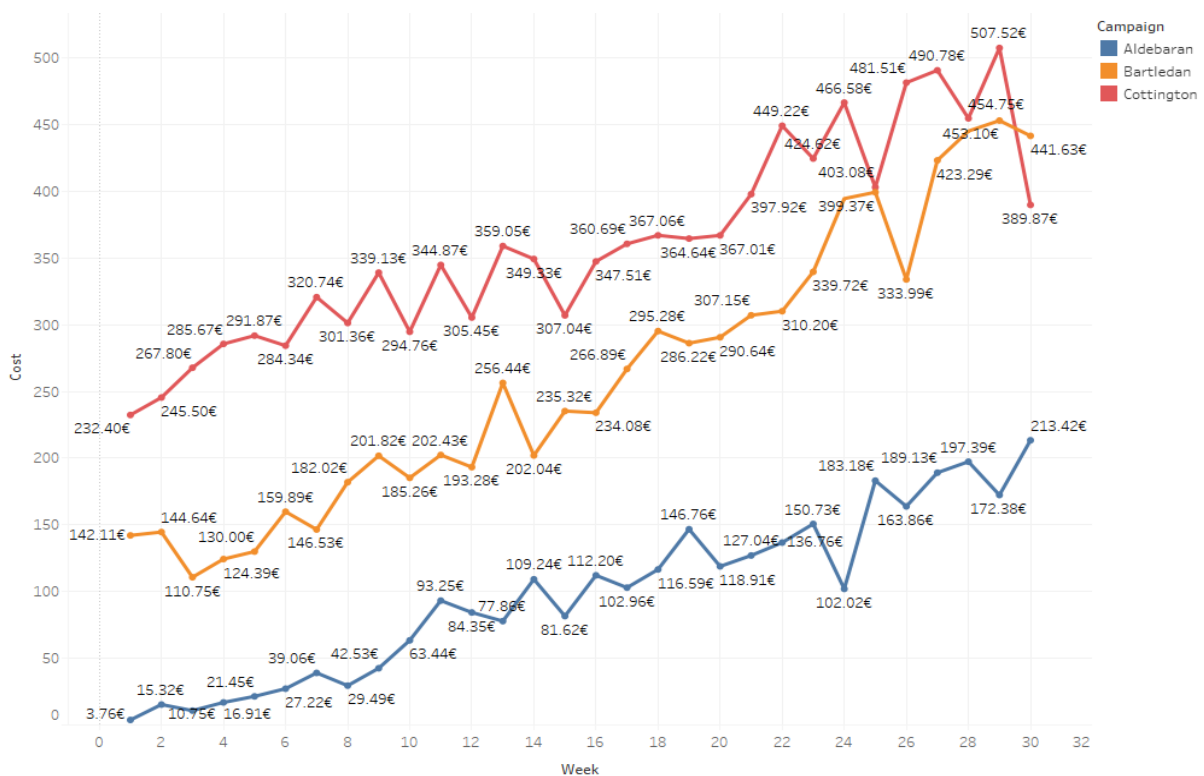
The important thing to observe in the above plot is that the profit generated by Bartledan is -798.71€ which means that over the course of whole 30 weeks the marketing campaign Bartledan is in a loss.

Now to present a deeper analysis of the development of the marketing campaigns we should study the trends on a week wise basis.

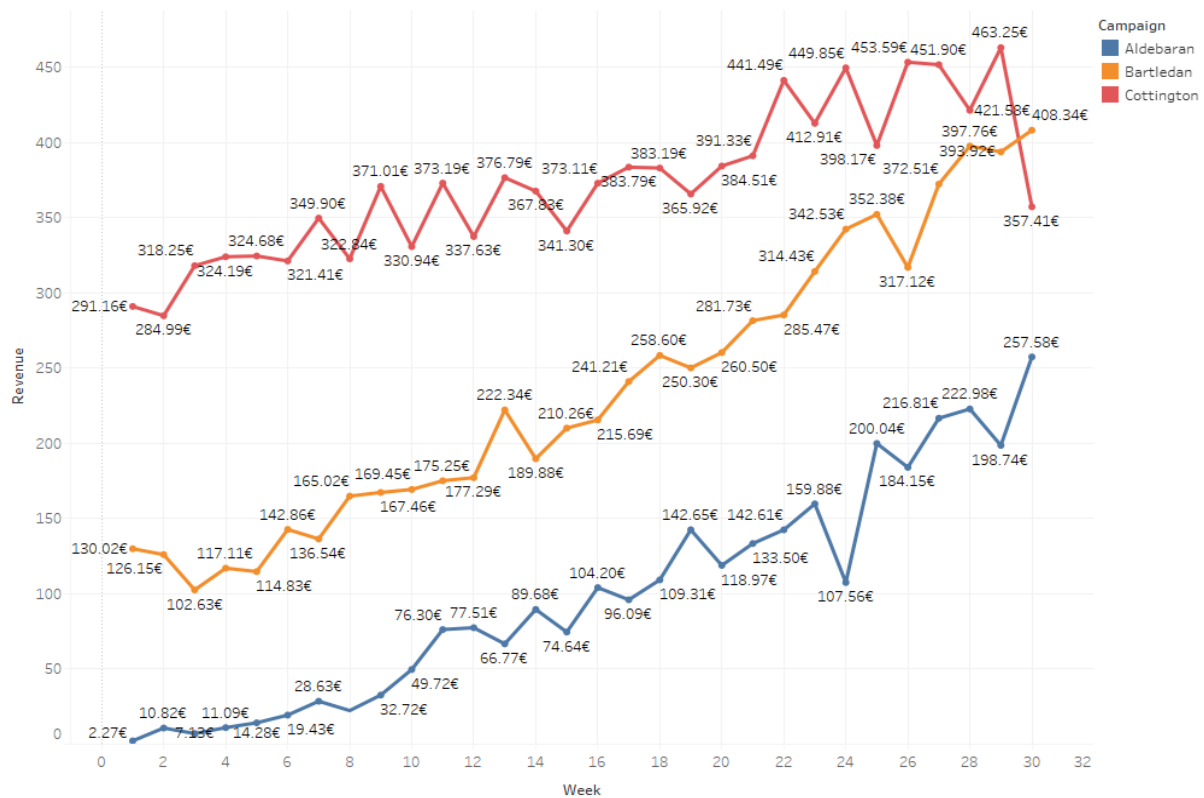
Let's study the Cost incurred by the 3 campaigns individually per week.



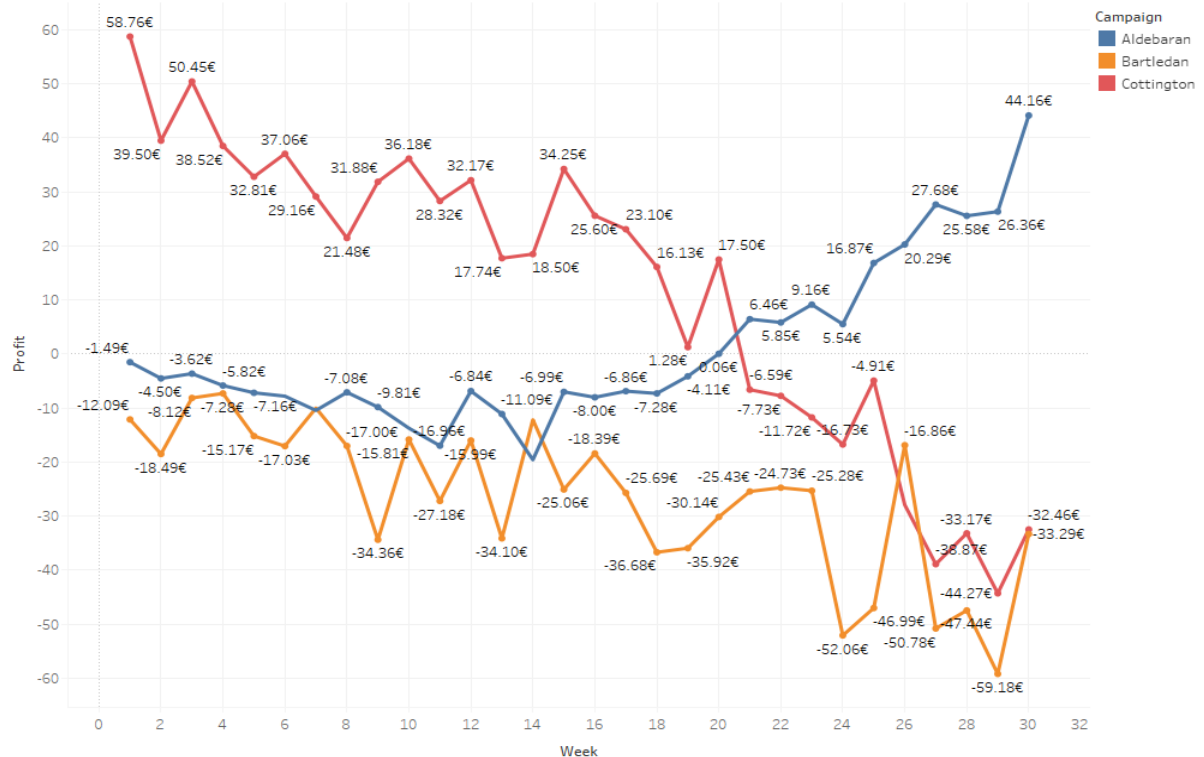
The trend in the above plot of visitors per week shows that the number of visitors for the Aldebaran campaign has increased at a very high rate as compared to that of the Bartledan and Cottington. At week 30, Aldebaran garnered 613 visitors whereas Bartledan and Cottington generated only 247 and 125 visitors respectively.



From the above plot of the variation of Cost, we observed that in any week except week 30, the spending on Cottington is always highest followed by Bartledan, and then Aldebaran. There is a linear increase in spending for all the three campaigns except in week 30 for Cottington campaign, there is a drastic decrease of about 120€ of the previous week's value.



From the above plot of the variation of revenue, we can say for all the weeks except week 30, the Cottington campaign is generating the highest revenue followed by Bartledan and Aldebaran. The above plot is identical to that of the variation of cost. This can be attributed to the fact that the revenue of any campaign is highly dependant on the spend in the campaign.



The above plot of profit per week shows a completely different picture. At week 30, even if the revenue generated by Aldebaran is least among the three campaigns but the profit generated is highest at 44.16€. The more surprising finding is that the profit generated by Bartledan and Cottington is

negative at -33€ and -32€ respectively, which means they are in loss considering the development of the campaigns on a weekly basis.

The summary of the few important findings presented above are as follows:

- Considering the overall sum of the measures like Revenue, Cost, and Profit, Cottingham is highest among all the three campaigns whereas considering the Visits, Aldebaran leads the chart.
- In week wise distributions, Aldebaran is more effective than Cottingham in terms of generating more visitors.
- There is a very strong relationship between the spend (Cost) and the revenue as both the plots of cost and revenue display nearly identical trends, *higher the spend higher the revenue*.
- In terms of profits generated per week, Aldebaran is most beneficial as it has generated more profits than Cottingham and Bartledan.

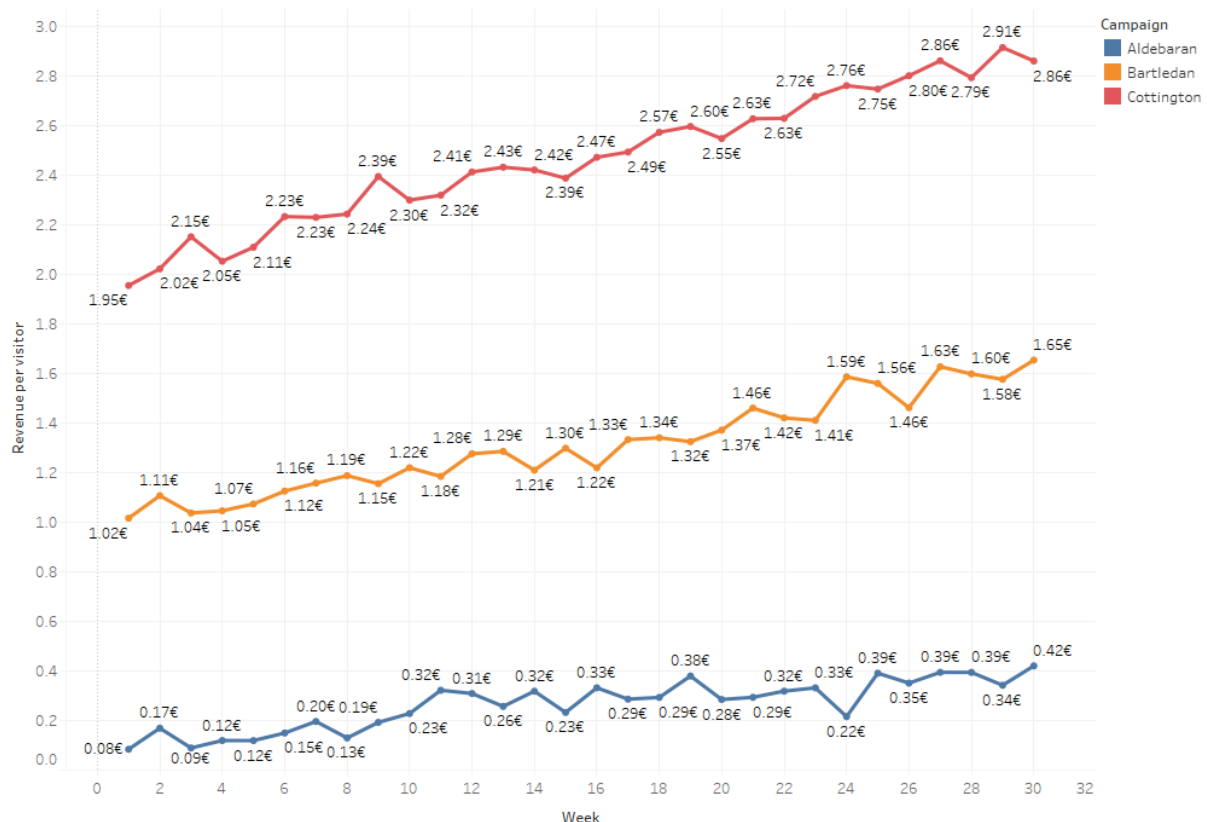
Problem Statement b:

Assess the development of the quality of traffic. How is the overall development and how does each campaign evolve?

In order to assess the development of quality of traffic we have to create few more metrics:

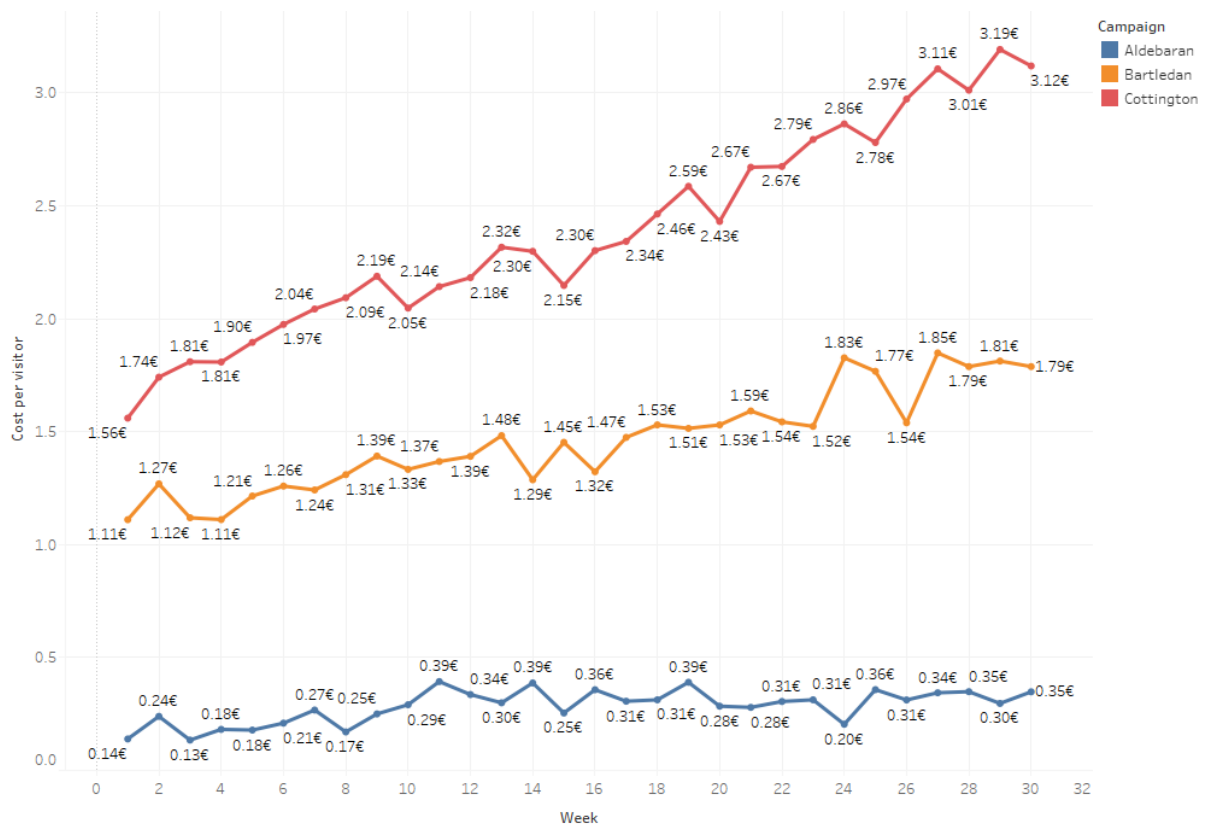
- **Profit per visitor:** Profit per visitor of a campaign in a week
- **Revenue per visitor:** Revenue generated in a week by one visitor of a campaign

Let's see the variation in revenue per visitor

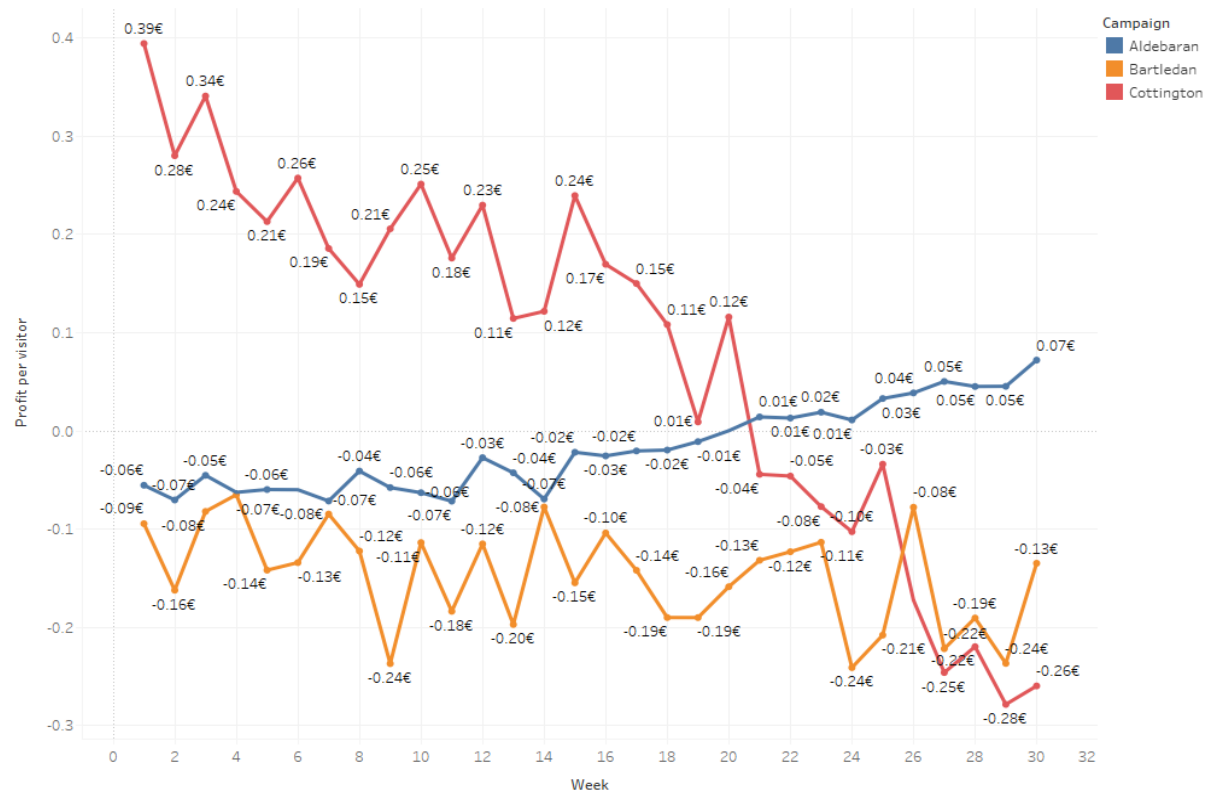


For all the three campaigns, there is a steady increase in the revenue every week. Cottingham stays at highest for all the weeks followed Bartledan and Aldebaran. The above plot suggests us that the revenue generated per visitor is highest for the Cottingham campaign at 2.86€ followed by Bartledan and Aldebaran at 1.65€ and 0.42€ respectively.

Actually, the revenue plot does not give us a clear idea about which campaign contributes the most in development in terms of quality of traffic, as the cost element could make the campaign with the most revenue as the one with the least profit due to high costs.



The above plot clearly proves our doubt of having very high costs can render one campaign into a loss. According to the above plot, the spending per visitor is very high for Cottingham (3.119€) followed by Bartledan (1.79€) and then Aldebaran (0.35€). This metric basically means: the expense done to attract one visitor. This should be as low as possible in order to generate more profit per visitor.



As suggested, the reality about the development in terms of quality of traffic is totally different from what the revenue plot shows us. The above plot clearly substantiates our above finding that there are no profits generated by Cottington and Bartledan, instead, they are in loss. The above plot of profit generated per visitor shows that at week 30 Aldebaran campaign is generating the highest profit per visitor at 0.07€ which is far better than Bartledan and Cottington campaigns which are at a loss of 0.13€ per visitor and 0.26€ per visitor respectively.

The overall summary of the development and evolution of a particular campaign:

- **Aldebaran:**
 1. The increase in revenue per visitor is steady from week 1 to week 30.
 2. The revenue generated per visitor increases by 2.5 folds from week 1 to week 30.
 3. It started in the loss and went on to become the campaign with the best quality of traffic. The profit generated is nearly doubled from week 1 to week 30.
 4. Week 20 was when it made its transition into a profit-generating campaign.
- **Bartledan:**
 1. The increase in revenue per visitor is steady from week 1 to week 30.
 2. The revenue generated per visitor increases by nearly 1.6 times from week 1 to week 30.
 3. It started as the campaign with the highest loss and continued to stay the same until week 26 but even after that, it was never able to generate profits.
 4. It was never able to generate profits but the least loss it incurred was on week 4 of - 7.28€.
- **Cottington:**
 1. The increase in revenue per visitor is steady from week 1 to week 29 except at week 30 where there is an unusual increase in the revenue due to the unusual increase in the spent.
 2. The revenue generated per visitor increases by nearly 1.5 times from week 1 to week 30.
 3. It started as the campaign generating the highest profit, in fact, it was the only campaign generating profits from the start of week 1 but as the campaign continued it gradually fell off and became the campaign generating the most loss per week.
 4. After week 20 it was never able to generate profits.

Problem Statement c:

A business developer wants to spend an additional 250€ per week from week 31 onwards.

- *Advice the business developer in which campaign to invest and why?*
- *How does this impact the overall performance in the market from week 31 onwards?*

In order to suggest the business developer in which campaign to invest let's create some more metrics in order to assess the development of business:

- Profit %
- Visits per euro spent

First, let's dive into the plot of Profit %.

Analysing the below *fig: Profit %*, we can clearly substantiate our previous findings of Aldebaran as the only campaign generating profit while the other two are in loss. Aldebaran is raking in about 21% of profit which is huge considering it started as the one with the highest loss.

Based on our findings we should advise the business developer to spend in Aldebaran campaign because if we are going to increase the spending in the campaigns the trend of Aldebaran generating profits while the other two campaigns in loss will continue more drastically. This statement can be better understood if we analyse the variation of profit w.r.t. cost.

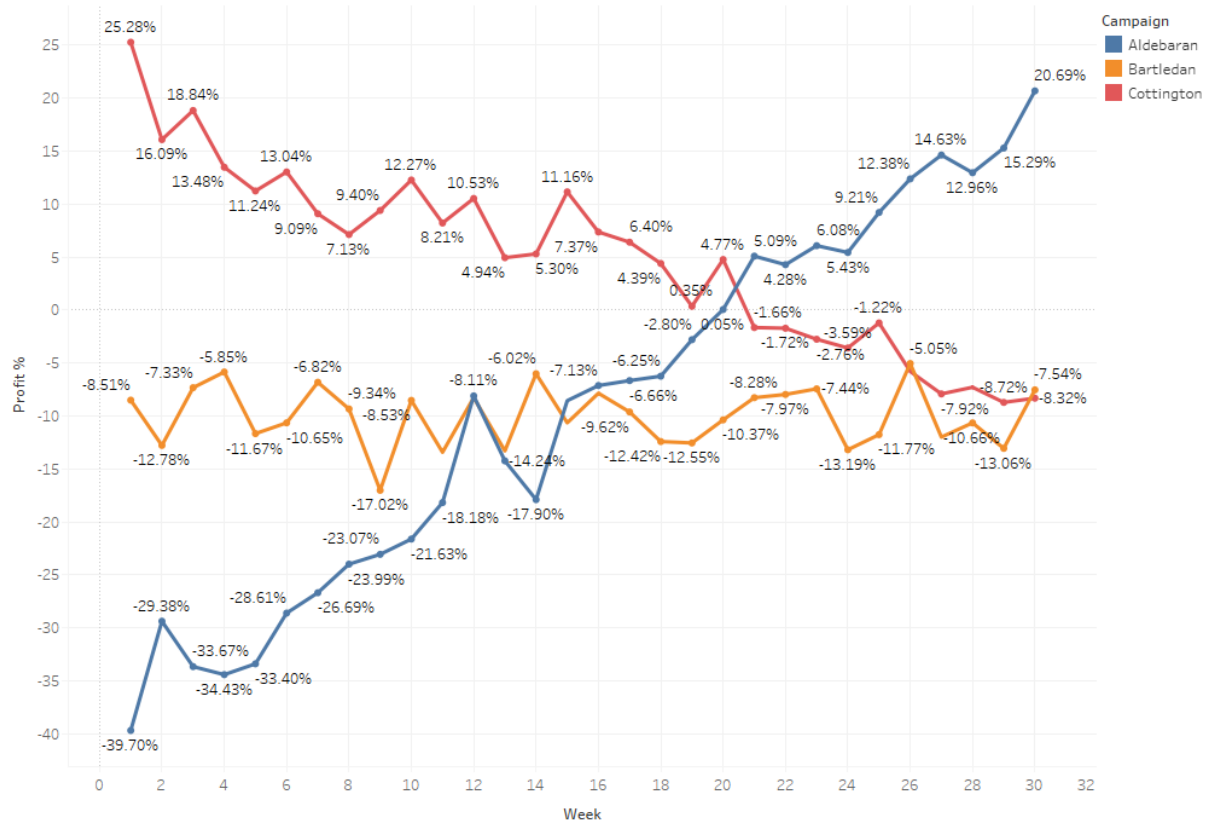


Fig: Profit %

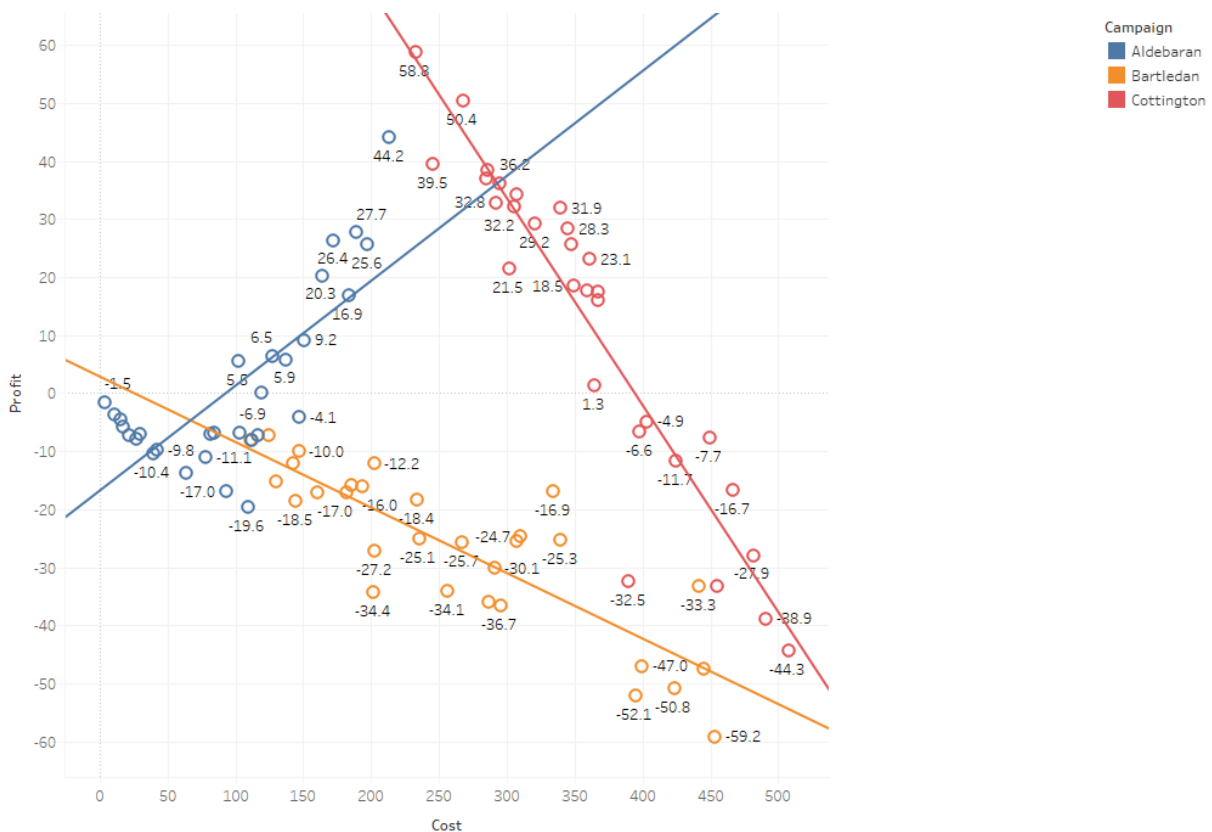


Fig: Profit as a function of Cost

The above plot clearly shows the trend and proves that the more we increase the spending the more loss we are going to incur from the Bartledan and Cottingham campaigns whereas increasing the spending will create more profits for Aldebaran campaign.

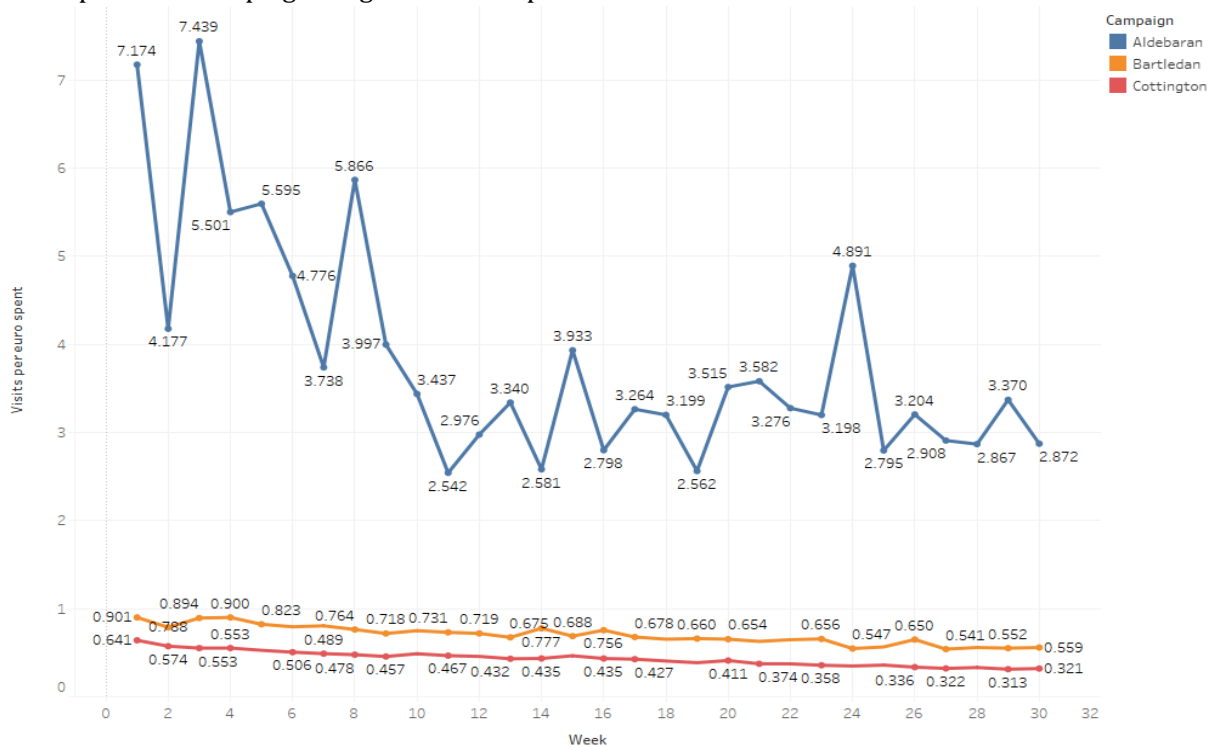
The above explanations justify our suggestion of Aldebaran as the one campaign to invest more in.

Now coming over to the impact of more spending on Aldebaran the overall performance can be summarised considering many metrics.

Let's first analyse the impact on the visitor count.

As the spend in Aldebaran is increased we expect the overall performance of the Aldebaran campaign in the market is going to decrease in terms of the number of visitors per euro spent on a campaign.

This might come as a surprise but on analysing the trend in the number of visitors garnered for one euro spent on a campaign we get the below picture.



Following are the findings from the above plot:

- For all the three campaigns as the spending is increased, the number of visitors per euro spent is decreased.
- Aldebaran follows a high-cost-high-volume model. This means that this is a type of campaign which can generate more profits if there are a high number of visitors which is possible by high spending. It is not going to prosper if there are fewer visitors.
- Cottingham, on the other hand, follows a low-cost-low-volume model. This means that this is a type of campaign which can even generate profits with very less number of visitors. But if the spending is increased to increase the visitors in expectation of making more profits, it backfires and results in a loss. So, this is a type of campaign which can only prosper with low spending.
- Bartledan follows the same model as that of Cottingham.

Now, the impact of more spending on revenue and profit is already explained while advising for the best campaign to do more investment in. But we can take it a step further to predict profit for the increased spending.

On analysing the profit trend line above we can find the variation equations:

- Aldebaran:

$$\text{Profit} = 0.180761 * \text{Cost} - 16.8093; \text{ At Cost} = 463.42, \text{ Profit} = 66.96$$

This further substantiates our suggestion that the Aldebaran campaign will be generating profits.

Session Data

Introduction

Problem Statement

Analyse the session data and find out if there is any relation between booking data and any other given information. Feel free to formulate hypothesis based on the findings.

Data

On observing the data it is found that the important variables for the analysis are:

- Booking
- Clickouts
- Session.Start.Time
- Session.End.Time

The “Session” variable only contains the session id which of no use in formulating relation between booking and other variables.

Now, let's modify the data a bit so that analysis can be performed on the data more efficiently. Subtracting the Session.Start.Time from the Session.End.Time the duration of a session is found. The duration is then converted into Minutes and rounded off and store as Session.Duratin.Minutes in order to avoid huge variation in data.

On checking the Session.Duration.Minutes it can be observed that it has negative values. On examining those values it is found that they are just interpreted in a wrong way by the R compiler. If we convert the negative values of minutes into seconds and add it to 86400 seconds(number of seconds in 24 hours), we will get the seconds which are actually the correct value to replaced and then replace them.

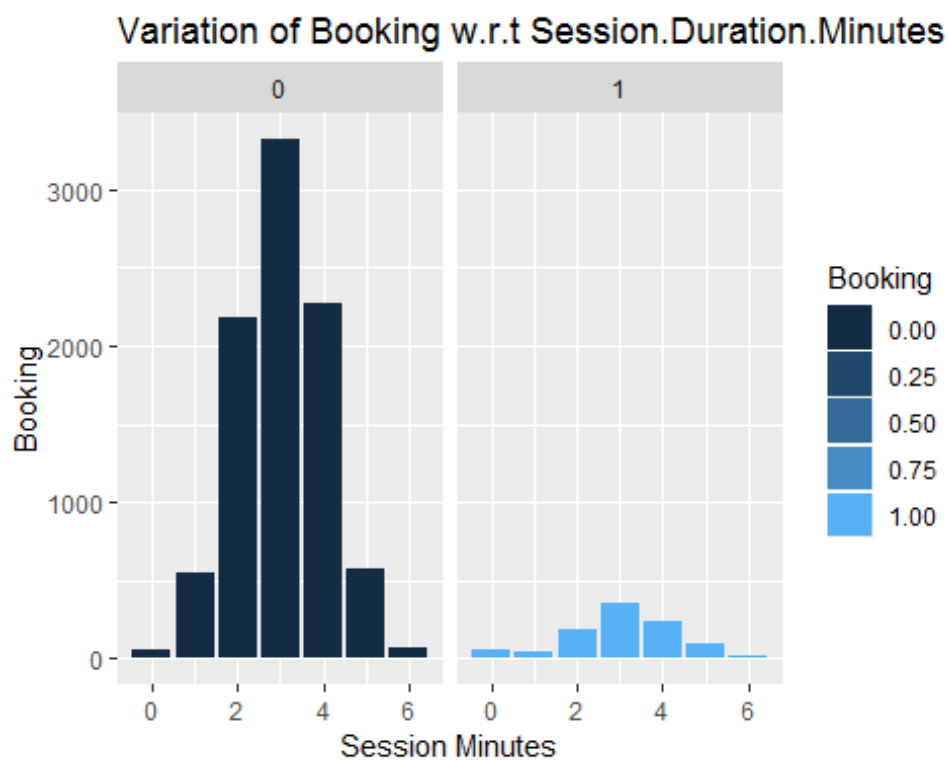
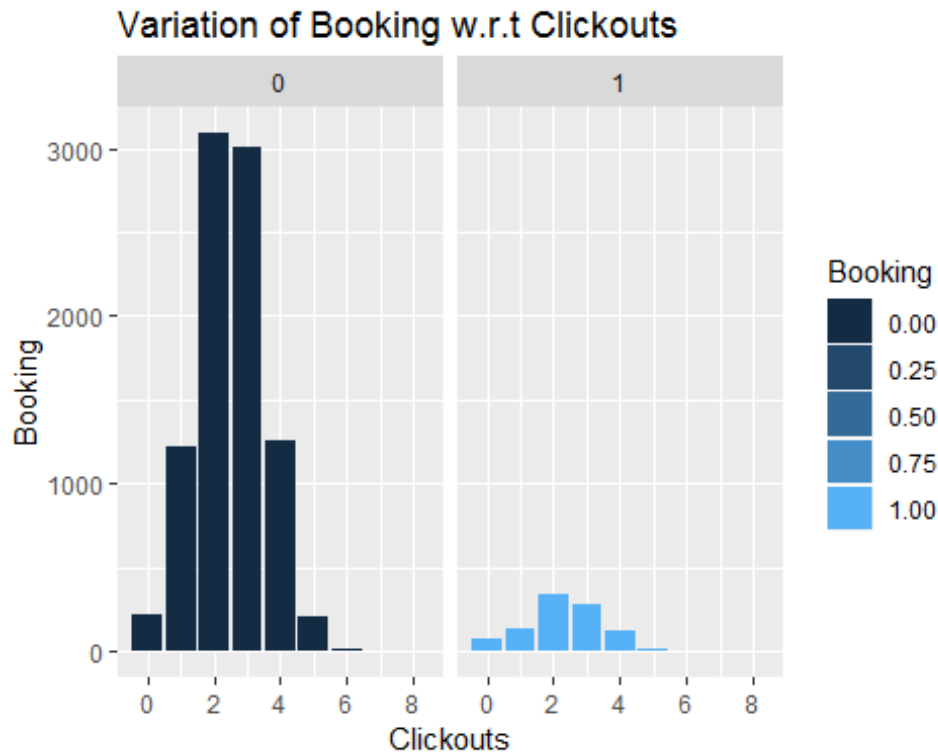
For example, $-1438 * 60 = -86280$ $86400 - 86280 = 120$ seconds (2 minutes)

Data Analysis

We should start with checking the distribution of the booking data with the other two variables, Clickouts and Session.Duration.Minutes.

It can see from both the plots above, that the Booking data is **normally distributed** and **follows a pattern in its variation with respect to Clickouts and Session.Duration.Minutes**.

The plots clearly show that **most booking is done when the session minutes is 3 minutes**. The above plot also shows that **most of the bookings are done when the clickouts are 2**.



This finding suggests that there could be a relation between Booking and Session.Duration.Minutes/Clickouts. Let's conduct a Hypothesis testing to find out if there is any relation between Booking and the other two variables.

According to hypothesis testing our null hypothesis(H_0) and alternate hypothesis(H_1) would be:

H_0 : There is no relation between Booking and (Session.Duration.Minutes/Clickouts).

H_1 : We fail to reject the assumption that there is a relation between Booking and (Session.Duration.Minutes/Clickouts).

The most important thing to keep in mind is that Booking is the dependent variable (the variable whose value depends on other variables) and it should be treated as Categorical variable as it is stated in the problem statement that it is binary.

Here we'll perform logistic regression as our dependent variable should be treated as categorical and the independent variables have to be treated as numerical.

The results of the test are:

```
## Call:
## glm(formula = Booking ~ Session.Duration.Minutes + Clickouts,
##      family = binomial(link = "logit"), data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5741  -0.4651  -0.4395  -0.4137   2.3807
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.97323    0.12382  -15.936  < 2e-16 ***
## Session.Duration.Minutes  0.04230    0.03138   1.348    0.178
## Clickouts        -0.16154    0.03214  -5.027 4.99e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 6355.4  on 9999  degrees of freedom
## Residual deviance: 6328.7  on 9997  degrees of freedom
## AIC: 6334.7
##
## Number of Fisher Scoring iterations: 5
```

For Session.Duration.Minutes, the p-value is 0.178. In statistics the most suitable value of alpha is assumed to be 0.05, then the p-value is greater than alpha, so we fail to reject the null hypothesis. That is, we do not have sufficient evidence to say that there is a relation between Booking and Session.Duration.Minutes.

For Clickouts, the p-value is 4.99×10^{-7} . As pointed out earlier, our alpha is 0.05, then the p-value is smaller than alpha, so we reject the null hypothesis. That is, we have sufficient evidence to say that there seems to be a relation between Booking and Clickouts.

So, our tests formulate that there is a relation between Booking and Clickouts data.

Note: All the modified .csv files and the code for the analysis is provided separately and a compiled report of the code is also provided as a HTML file.