Influenceable vs non-influenceable channels

## 

## Introduction:

In the online marketing analytics field, marketing managers steer their budget in two types of channels. Channels in which they can directly influence with investing more money and channels in which it is not possible to affect the performance directly. For example Paid Search advertising can be influenced yz increasing the bids of the ads, i.e investing more money, thus reaching higher positions in google/bing search ads and therefore reaching a broader audience, which at the end can lead to more people adding proudcts to their carts. On other hand, non-influenceable channels cannot be affected by investing more money, for example organic results. On example is Search organic results, which are based on the search engine’s algorithm to order the results.

In this study, two groups of channels are analyzed: Influenceable channels and non-influenceable channels. The objective is to determine if one of the groups has statistically higher cart additions than the other one. In other words, if one group leads to more people adding products to the cart.

Marketing managers are interested in having a higher amount of cart additions since this shows an intention of purchase. Therefore high amount of cart additions indicate a good performance of the channel.

## Methodology:

The data used in this study is based on a multinational european fast-moving consumer goods dataset. The same as the previous research question. For privacy reasons the data is completely anonymized and only 35k rows of data will be used in this study, the original dataset is 2 million.

Each row represents a different customer/user. Cart additions is the number of products added into the basked in the e-commerce website.

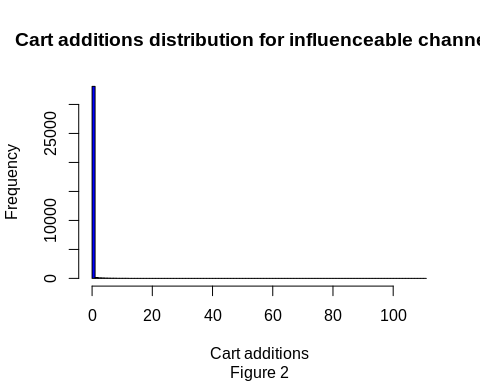
| customer\_id | cart\_ad\_infl | cart\_ad\_non.infl |
| --- | --- | --- |
| 0 | 0 | 0 |
| 1 | 0 | 0 |
| 2 | 0 | 0 |
| 3 | 0 | 0 |

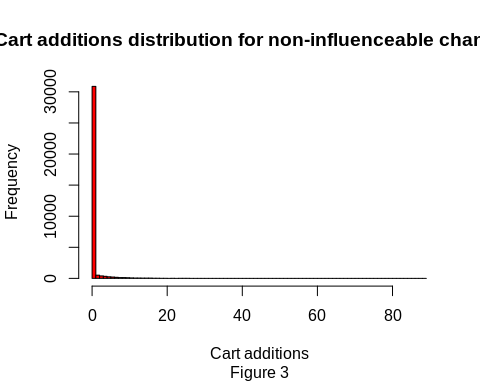
The average amount of cart additions is 0.10 for influencebable channels vs 0.6 for non-influenceable channels and the median is 0 for both. In fact, as can be seen in Figure 1 and Figure 2, the distribution is very skewed to the right. Most of the users haven’t made a cart addition. We can say with security that this data does not follow a normal distribution. Despite not ideal, this is the reality. Most of the customers or better said potential customers do not end up adding produts to the cart, and despite high investment in marketing the conversion rate at the end is very low (maximum 5%).

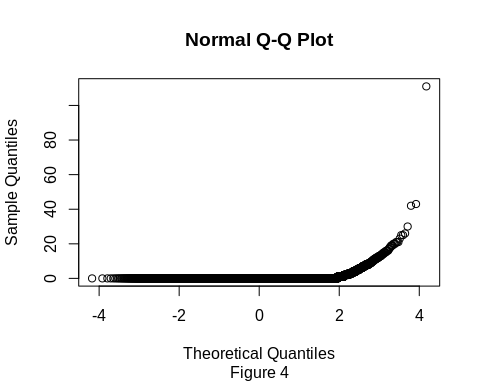
Figure 1

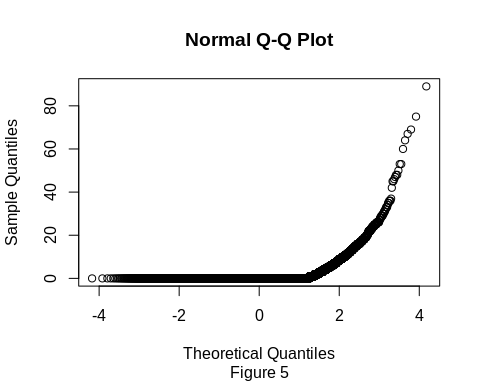
| customer\_id | cart\_ad\_infl | cart\_ad\_non.infl |
| --- | --- | --- |
| Min. : 0 | Min. : 0.0000 | Min. : 0.0000 |
| 1st Qu.: 8421 | 1st Qu.: 0.0000 | 1st Qu.: 0.0000 |
| Median :16842 | Median : 0.0000 | Median : 0.0000 |
| Mean :16842 | Mean : 0.1081 | Mean : 0.6163 |
| 3rd Qu.:25263 | 3rd Qu.: 0.0000 | 3rd Qu.: 0.0000 |
| Max. :33684 | Max. :111.0000 | Max. :89.0000 |

Summary table. Main descriptive statistics

 Figure 1 shows the distribution of the cart additions. The majority of the cart additions are 0. The distribution is positively skewed.



 In Figure 3, we can observe a Q-Q plot for the influenceable group. We can see that there is not a straight line, meaning the data is not approximately normal. In the x-axis we plot the theoretical quantiles (average = 0, standard deviation = 1) and on the y-axis we plot the ordered values for the cart additions.

 The same can be said for Figure 5, which represents a Q-Q plot for the non-influenceable channels.

Since we are dealing with non-normally distributed data I will go directly with the Wilcox rank-sum test (unpaired test). The non-parametric test that does not require normally distributed data. It only requires that the two populations are independent of each other and that the have equal variances. The independence assumption is not violated in with this dataset, however the variances of both groups seem to be very different, not ideal for this test ( 1.31 vs 7.6 for influenceable and non-influenceable channels respectively ).

We just need to remember that this test does not check whether the means of the two populations are the same, but whether the medians are equal.

## [1] 1.314518

## [1] 7.660838

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: df$cart\_ad\_infl and df$cart\_ad\_non.infl  
## W = 521793152, p-value < 2.2e-16  
## alternative hypothesis: true location shift is less than 0

Figure 6

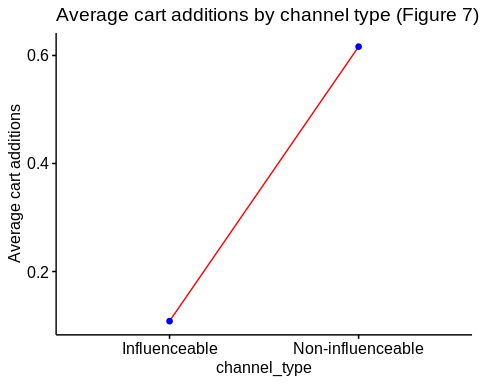
| statistic | p.value | method | alternative |
| --- | --- | --- | --- |
| 521793152.500 | 0.000 | Wilcoxon rank sum test with continuity correction | less |

Results of the wilcox sum-rank with alternative =“less”

We can reject the null hypothesis that both medians of the two groups are the same and that the non-influenceable group has a higher location shift than the influenceable.

The statistic is very high: 521793152, while the p-value is extremely low, way below the significance level of 5%.

With a 95% confidence level we can say that the non-influenciable group has a higher median than the influenceable group. The degrees of freedom are n-1 = 33683

 Figure 7 represents the means of the cart additions by channel type, showing a clear difference between both groups.

## Results

The objective of this study was to find if there is a real difference between two groups of channels: Influenceable channels and non-influenceable channels. The former can be affected by marketing spent while the second cannot. After checking for normality in the data, we can conclude the data does not follow a normal distribution and it is positively skewed.

In order to check if there is a significant difference between the two groups, a non-parametric test (wilcox rank-sum test) was implemented. The result of the test shows a clear interpretation. The medians of both groups are different, being the influenceable median below the non-influenceable median in terms of cart additions.

A message to the marketing managers would be to review which exact channels are causing this and why. The budget distribution might not be the most appropriate one since channels where capital investment is not necessary are brinting a higher amount of cart additions than more expensive channels.