

Research Proposal

1. Introduction

Marketing companies want to run successful campaigns but the market is complex and several options can work. So normally they run A/B tests, which is a randomized experimentation process wherein two or more versions of a variable (web page, page element, banner, etc.) are shown to different segments of people at the same time to determine which version leaves the maximum impact and drive business metrics.

The dataset I am interested in exploring is the results of an A/B test run by a company. The company has conducted A/B tests with the majority of the people exposed to ads and a small portion of people to a Public Service Announcement (PSA) (or nothing).

2. Data

The data comes from [Kaggle](#) and contains 588101 observations and 6 variables. There are no missing values. The 6 variables provided are:

- a. User id
- b. Test group - If "ad" the person saw the advertisement, if "psa" they only saw the public service announcement
- c. Converted - If a person bought the product then True, else is False
- d. Total ads - Amount of ads seen by person
- e. Amount of ads seen by person - Amount of ads seen by person
- f. most ads day: Day that the person saw the biggest amount of ads
- g. Most ads hour - Hour of day that the person saw the biggest amount of ads

3. Research design:

The business question that I am interested to answer are:

1. Does Number of ads people see contribute to conversion?

2. At what day, people watch the highest number of ads?

Hypothesis

1. Do people who converted have been shown a bigger number of ads on average than people who are not converted?

Steps to prove:

- a. The variable test group indicates "ad" if the person saw the advertisement, "psa" if they only saw the public service announcement. For testing the above hypothesis, we need to have a dataset that has only an "ad" Test group.
 - b. Next step is the variable 'Converted', which indicates whether a person bought the product or not. Split the data into two groups on this variable i.e 'converted_yes' and 'converted_no'.
 - c. Check if two groups are normally distributed. We'll do that both visually via histograms and through descriptive statistics using the measures of skewness and kurtosis.
 - d. If samples of total ads variable are normally distributed, proceed with t-test, to get t-statistic and p-value and see If the t-test shows a significant difference in the means of number of ads shown to two groups.
 - e. If samples of total ads variables are not normally distributed, use a non parametric test like Kruskal-Wallis Test to check if there is significant difference between two samples.
 - f. If there is no significant difference between two samples, we can say that the number of ads doesn't contribute to conversions.
 - g. If there is a significant difference in the number of ads shown in two samples, we can prove/reject the above hypothesis by calculating 95% confidence interval and also via visualization using boxplot or point plot.
2. Is there a significant difference in the number of ads seen on average from Monday to Friday? If yes, Which day do people see the biggest number of ads?

Steps to prove:

- a. The variable test group indicates "ad" if the person saw the advertisement, "psa" if they only saw the public service announcement. For testing the above hypothesis, we need to have a dataset that has only an "ad" Test group.
- b. Next, the variable 'most ads day' indicates the day that the person saw the biggest amount of ads. For this analysis we will be grouping our data based on the variable 'most ads day', which leaves us 7 different groups for our analysis.

- c. Check if all groups of 'total ads' variable are normally distributed. We'll do that both visually via histograms and through descriptive statistics using the measures of skewness and kurtosis.
- d. If seven independent samples of data are normally distributed, compare using one-way Analysis of Variance (ANOVA), identify whether any of the groups differs from the others., and use a set of pairwise t-tests comparing each group with the rest to check which one is odd one and visual inspection using a box plot for answering which day people see the biggest number of ads.
- e. If 7 independent samples of data are not normally distributed, use a non parametric test like Kruskal-Wallis Test to check if there is significant difference between samples.
- f. If there is no significant difference between two samples, we can say that people on average see the same number of ads from Monday through Sunday.
- g. If there is a significant difference in the number of ads shown in samples, we can prove the above hypothesis by calculating visualization using boxplot or point plot.

4. Audience

The above proposal benefits the company who ran A/B test to find key insights regarding impact of ads on conversions. With the above proposal, the company can understand whether the people are buying their products mostly because of their ads. Also, it gives them info regarding the day of the week people on average saw most of their ads. This understanding will also be helpful to the companies who run the similar campaign and driving future business decisions.