Unit 1 Boot camp: Capstone project

Research proposal

**Problem/Hypothesis**

We want to increase sales in the company by increasing the signups for monthly plan in a music subscription company. The sales team have come up with a promotion to be rolled out in emails which when viewed will attract customers to signup for the monthly plan.

The problem here is that we have two versions of the promotion, i.e. Version A and Version B. Which among the two versions of the email needs to be sent to prospective customers is the question.

The hypothesis or question here is: Which version of the promotion is more effective in leading to signups for the monthly music subscription plan?

**Rollout plan**

The main factor here is how many people need to be sent out the email and how long do we test or experiment.

The number of people or emails ids we have is 120,000. We are going to do the finalized version of email with 100,000 prospective customers. From these 120,000 emails ids we are going to randomly take a subset of population and test the two versions of emails/promotions. We will be testing with 20,000 people. i.e. 10,000 emails with the test version A and 10,000 emails with test version B.

We can do the rollout in two phases, initial phase using 20% rollout (total 4000 emails) and then getting a pilot experiment results and doing the full phase (extended phase) of remaining 80% rollout ( total 16,000 emails).

The initial phase will be carried out for two weeks. Where emails are sent every day for two weeks to avoid the selection bias.

Based on the initial phase I test results we get the idea of the standard deviation of the sample population. This is can be used to plug into a power calculation and see if our test size is enough to find how much of a difference i.e. 25%, 50% or 75% or 100% difference with version A and version B. <https://www.stat.ubc.ca/~rollin/stats/ssize/n2.html>

Once we have the power calculation results, we can adjust the sample size if possible or needed to find certain difference in two versions and execute the second phase of A/B testing.

The final phase will be carried out for more two weeks. The data collected will then evaluated to see if which version of A or B is giving more sign ups.

**Evaluation plan**

Success in this experiment here is the signups or the conversions. The signup rate of Version A email is compared to the signup rate of version B email. The time curve will also be plotted to see how the signup rate varies with treatment date.

The t-statistic is calculated with signup rates of version A and version B and with the help of p-value (p < or > 0.05) at that t-Value we can conclude if Version A or B is statistically significant over other. The significance level planned to use here is 5%. i.e. we will be 95% confident that version A or B will be higher than the other in our results.

The data set selected here for testing experimentation if needed

<https://raw.githubusercontent.com/Thinkful-Ed/data-201-resources/master/1.4.11_experimentation_guided_example/experiment_data_2_wks.csv>

**Secondary metrics**: We can also measure here the read receipt of email versions. This will be useful to troubleshoot the reason for very low signup for either version A or version B. Is the failure due to not opening the emails or not subscribing to the plan ever after opening the email? can be answered here.

The experiment should be planned such that there is no other event either in time i.e. no major holidays or celebrations and also there are no other website versions changes which may have an effect on the sign up rate.