A/B testing :

Project Overview

In this experiment udacity tested a change where if the student clicked ‘start free trial’,they were asked how much time they had available to devote to the course.If the student indicated 5 or more hours per week they would be taken through the checkout process as usual .If they indicated fewer than 5 hours per week a message would pop up indicating that Udacity courses usually requires a greater time commitment for successful completion and suggesting that the student might like to access the course material for free.The unit of diversion is a cookie, although if a student enroll in free trial they are tracked by user id from that point.

Experiment Design

Metric Choice

- Invariant metrics: number of cookie, number of clicks, click-through -probability

- Evaluation metrics: Gross-conversion,Retention ,Net conversion

1.**Number of cookie**: Since the unit of diversion is cookie that is not going to be affected by the change that udacity is launching at the time of enrollment.Hence it can be used as a invariant metric.It is expected to remain same across control and experiment group.It cant be used as evaluation metric since it remains unaffected by the change udacity is about ot launch .Hence it is invariant in nature.

2.**Number of clicks**: It is defined as the number of unique cookie to click the start free trial button.It can be used as a invariant metric because whatever changes are being launched by the udacity appears after clicking’Start free trial ‘ button.Hence it is expected to remain same across both control and experiment group.We expect our evluation metric to vary across the groups so it is not a suitable choice for evaluation metric.

3.**Click-through -probability**:It should not change across control and experiment group as it is the ratio of two unchanging metrics namely number of unique cookie who click and total number of cookie.Hence it can be used as invariant metric.Again for the above mentioned reasons it cannot be used as evaluation metric.

4.**Gross-conversion**: it is the ratio of number of users who enrolled in free trial and number of unique cookie to click 'start free trial'.This metric can vary depending upon how users respond to the pop-up message displayed in the experiment group.If they are willing to make time commitment only then they will enroll. Hence ,in our experiment group this number should be less as compared to control group.Since it is expected to vary between control and experiment group , it is suitable choice for evaluation metric but not as invariant metric which is used for doing sanity check.It can help us to check whether we reduced our cost by introsucing the pop-up message.

5.**Retention**: It is the ratio of number of users who made their first payment and the number of users who enrolled for the free trial.It is expected to be higher in experiment group as compared to control group.Again retention is expected to vary between control and experiment group ,hence it can be a choice for evaluation metric.It can't be used for doing sanity checks as it is expected to remain higher in experiment group. So, it cant be used for sanity checking.

6.**Net-conversion**: It is the ratio of number of users to remain enrolled for the 14 day trial period and make their first payment and number of users who clicked 'start free trial' button. As per the assumption of the experiment , the pop-up message displaying the required number of hours to be devoted in order to complete the course , is expected to weed out frustrated users who cant devote required time.As per asumption it is expected to increase significantly in our experiment group.It will show us how the change affects the overall revenue of udacity. This metric is expected to vary as the number of user-id to remain enrolled and make their first payment is not expected to remain same in both group.Hence it can be our evaluation metric.Since net conversion metric is expected to vary across groups it won't be a suitable candidate to be used as invariant metric(sanity checking).

7.**Number of User-id**: the experiment under consideration is expected to affect the total number of students getting enrolled after seeing the pop-up message .The metric can't be chosen as invariant metric as it is expected vary significantly between groups. Again this metric is just raw numbers. Usually we prefer a ratio or rate to raw numbers as they are more robust. Hence this metric would not be an ideal choice for evaluation metric.

Measuring Standard Deviation

Evaluation metric	Standard Deviation
1.Gross conversion	0.0202
2.Retention	0.0549
3.Net conversion	0.0156

Unit of analysis is equal to unit of diversion in case of gross -conversion and net-conversion but not retention.Empirical variability may vary from analytical variability when our evaluation metric is retention.

Sizing

Number of Samples vs. Power

1.Gross conversion:
Baseline conversion rate = 20.635%
Minimum detectable effect=1%
alpha=5%
beta=20%
Sample size=25835
2.Retention:
Baseline conversion rate = 53%
Minimum detectable effect=1%
alpha=5%
beta=20%
Sample size=39115
3.Net conversion:
Baseline conversion rate = 10.93%
Minimum detectable effect=0.75%
alpha=5%
beta=20%
Sample size=27414

If sample size of 39115 is chosen ,we would need 12223437 total pageviews which would require 611 days if we chose to divert 50 % of our traffic(20000 pageviews). Hence i chose to discard this metric and chose the sample size of 27414 which is for net conversion metric and total pageviews required would be 685325 at alpha=0.05 and beta=0.2.

Duration vs. Exposure

I chose to divert 50% of udacity traffic to this experiment to minimize end-user related impacts on our experiment results and there are always a chance that data collected over short period of time might be affected by specific events, as could be the case if i chose to divert 100% traffic.if this traffic diversion gives expected result we may try other combinations also to gain more confidence.At this rate the number of days needed would be 685325/20000=35 days approximately. 35 days might be a long time to run such experiments but in any given situation this decision is not in the hands of experiment designer only. There should be other groups

involved and should take the decision on mutual consent. such decision must align with our business objective in real world scenario. For now we can assume that it would run for 35 days.

Experiment Analysis

Sanity Checks

1.Number of cookie

Control group=345543
Experiment group=344660
Total number of cookie=690203
Probability of cookie=0.5
Standard error= sqrt(0.5 *(1-0.5)*(1/345543+1/344660))=0.001203
Margin of error=1.96*SE
Confidence interval=[0.4988 ,0.5011]
Observed value=(345543/690203)=0.50053

Since the observed value is within the confidence interval ,it will pass the sanity check.

2.Number of clicks

Control group=28378
Experiment group=28325
Total number of clicks=56703
Probability of click=0.5
Standard error=sqrt(0.5*0.5*(1/28378+1/28325))=0.00209
Confidence interval=[0.4959,0.5041]
Observed value=(28378/56703)=0.50046

Since the observed value is within the range ,it will pass the sanity check.

3.Click through probability

Total clicks=56703
Total pageviews=690203
click-through-probability=0.082125
Standard error=sqrt(0.082125*(1-0.082125)*(1/345543+1/344660))=0.00066
Confidence interval=[0.0812,0.0830]

Since our click through probability falls within the range, it passes the sanity check.

Result Analysis

Effect Size Tests

1.Gross conversion

d_min=$^{+}_{-}$0.01
probability$_{(\text{gross conversion})}$(control)=3785/17293=0.2180
probability$_{(\text{gross conversion})}$(experiment)=3423/17260=0.1983
Difference (d)=-0.02055
Standard Error=0.0044
Margin of Error=0.008624
Confidence interval=[-0.02911, -0.0120]

Since confidence interval does not contain zero and d_min also does not lie in confidence interval, it is statistically as well as practically significant.

2.Net Conversion

d_min=$^{+}_{-}$0.0075
probability$_{(\text{net conversion})}$(control)=2033/17293=0.11756
probability$_{(\text{net conversion})}$(experiment)=1954/17260=0.11269
Difference(d)=-0.0049
Standard error=0.00343
Margin of error=0.0067
Confidence interval=[-0.0116, 0.0018]

Since zero lies between the confidence interval ,it is statistically not significant.The confidence interval also contains the d_min value , hence it is also practically not significant.

Sign Tests

Gross conversion:
alpha=0.05

No. of success(where gross conversion is higher in exp. than cont.)= 4
No. of failure(where gross conversion is less in exp. than cont.)=19
Probability =0.5
P- value= 0.0026
Since p - value is less than alpha *it is statistically significant.*

Net-conversion:
alpha=0.05
No of success(where net conversion is higher in exp. than cont.)=10
No of failure(where net conversion is lower in exp. than cont.) =13
probability=0.5
p-value=0.677
Since p- value is greater than the alpha level ,*it is statistically not significant.*

Summary

If we need just some of our metrics to be statistically significant to be able to make decision about the launch of the experiment then it is a good idea to use Bonferroni correction.I prefer to not use Bonferroni correction as in this case it may lead to false negative.
We need our both the evaluation metric to be significant in order to recommend for launch Bonferroni correction could have given us the false result if used.

Recommendation

I would not recommend the launch of the experiment. For the gross conversion evaluation metric our result is practically and statistically significant indicating that the change under consideration have positive impact on experiment user group by reducing the number of students enrolling by viewing the pop-up message and finally reducing the number of frustrated students but considering the net conversion metric ,whose results are neither statistically nor practically significant indicating that there is risk involved with experiment launch which might reduce the revenue,which is not inline with the second part of hypothesis .Hence it would be advised to not take risk and launch the change as it might affect the financial aspect of the company as overall number of students making payment might reduce.

Follow-Up Experiment

The following follow-up experiment can be proposed:

Students should be allowed to register for the course as usual by clicking on “start free trial” with with a pop message asking them how many hours they can devote and also asking them what are their course expectations or outcome and career goals . Depending upon this, udacity can decide upon their requirement and they might be directed to either “free access of material ” or paid version of it where they can be asked for payment details depending upon whether the student is directed to paid version . If they are directed to free version of the course they should not be charged .After seeing the pop-up and getting the details like time student can devote and what are their course expectations, udacity can recommend suitable career path.This can help us to weed out those students who cant devote required time and those also who are not committed enough for the paid version of the course.In this case udacity can improve their coaching facility depending upon the number of users being enrolled for paid version as they would be committed enough to complete the course and would require coaching and feedback facilities.

Hypothesis:THe hypothesis is that asking students about what they are expecting from the course and how much time they can devote and depending upon this allocating appropriate course track can help students to remain engaged as per their requirement and help udacity to focus on those students who have shared outcome goals with udacity .

Unit of diversion: unit of diversion is a cookie and once users enrolled in the paid version of course they can be tracked by user-id.

Metrics:

Invariant metrics:

1.Number of cookies: It is the number of cookies to visit the course overview page. Since change under consideration appears after the home page load , it is expected to remain same and that's why it would be a suitable choice for invariant metrics.

2. Number of clicks :It the the number of unique cookie who clicks. This number is also not expected to vary as the change appears after the click on the “Start free trial “ button. It can also be a suitable choice for invariant metric.

Evaluation metrics:

1. Gross conversion: it can be defined as the ratio of the number of users who enrolled in the paid version of the course depending upon their career goal and course outcome and the number of unique cookie who clicks the “start free trial “ button.It is expected to vary between control and experiment group as more people is expected be directed to free course track in experiment group depending upon their goals and expected outcome .

2. Net conversion: It can be defined as ratio of the the number of users who enrolled and made their first payment after enrolling the paid version of the course and the number of unique cookie who clicks, This is also expected to vary as many people might not be able to adapt to the course or due to their personal reasons , ultimately drop out despite opting for the paid version of the course before their trial expires.