Background:
A/B Testing

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 appear indicating that Udacity courses usually require a greater time commitment for successful completion, and suggesting that the student might like to access the course materials for free.

Initial Unit of Diversion: Cookie

Experiment Design
Metric Choice
Invariant Metrics: Number of Cookies, Number of Clicks
Evaluation Metrics: Gross Conversion, Retention, Net Conversion

Number of Cookies:
It is the number of unique cookies to visit the course overview page.
> Since the Unit of Diversion is cookies and the number of cookies is not going to be affected by the change that udacity is launching at the time of enrolment . Therefore, number of cookies is well suited for invariant matric because there is no impact on the number of cookies

Number of Clicks:
Number of unique cookies to click the start free trial button.
> Since, the page asking for the number of hours the student can devote to the course appeared after clicking the " Start free Trial " button, the course overview page remains the same for both the control and experiment group.

Number of user-ids:
That is, number of users who enroll in the free trial.
> As per the experiment , the new pop up message is likely to affect the total number of user ids who enrol in the program. For this reason , this matric can not be used an invariant matric because it is more likely to differ in control and experiment group.

Click through probability :
That is, number of unique cookies to click the "Start free trial" button divided by number of unique cookies to view the course overview page.
> Since, the page asking for the number of hours the student can devote to the course appeared after clicking the " Start free Trial " button, the click through probability should remain the same for both the control and experiment group. Therefore, it can be chosen as invariant matric.

Gross Conversion:
Number of users who enrolled in the free trial / Number of users who clicked the Start N Free Trial Button
> After clicking the Start Free Trial button, a popup appears for the Experiment group users asking for the amount of time that the Student can devote to the course. Based on

the user choice it then makes a suggestion whether the student should enroll for the course or continue with the free course material.In the experiment group the user can make a decision based on the pop message and choose to continue exploring the material only. But for the control group no popup page appears, thus the user enrols for the course anyways. Hence, the gross conversion can be different in both control and experiment group hence can be used as an evaluation matric.

Retention:
Number of user − ids to remain enrolled for 14 days trial period and make their first payment / Number of users who enrolled in the free trial.

Retention ratio can also be a good evaluation matric because the retention ratio in experiment group is expected to be higher because of low enrolment , if experiment meets the assumption. After seeing the message , there might be fewer users who would enrol and hence retention would be higher because it should filter out those users who can leave the course as frustrated user. Therefore this ratio can be used as evaluation matric because ratio in control group and experiment group would be different and for the same reason it cannot be chosen as invariant matric.

Net Conversion:
Number of user − ids to remain enrolled for the 14 days trial period and make their first payment /Number of users who clicked the Start Free Trial button

As per the intention and assumption of the experiment , experiment group users are made aware that the course require some minimum of hours each week by showing the pop message at the time of enrolment. This message should filter out those users who cannot devote the required hours and are prone to be frustrated later on. This ratio should be different among control group users and experiment group users, if experiment's assumptions holds true.Hence , it can be used as evaluation matric and for the same reason it cannot be chosen as invariant matric.

Gross Conversion, will show us whether we lower our costs by introducing the new pop up. Net conversion will show how the change affects our revenues. After the experiment, we expect that, Gross conversion should have a practically significant decrease, and Net conversion should have a statistically significant increase.

Measuring Standard Deviation

Evaluation matric	Standard Deviation
Gross Conversion	0.0202
Retention	0.0549
Net Conversion	0.0156

The Unit of Diversion is equal to the Unit of Analysis for Gross Conversion and Net Conversion. Thus the analytical estimate would be comparable to the empirical variability.

For Retention the Unit of Analysis is " Number of users who enrolled in the courseware " which is not equal to the Unit of Diversion. Hence the empirical variability may be different from the Analytical estimate.

Sizing

Number of Samples vs. Power

We will not be using Bonferroni correction during analysis phase.
Using the [online](#) calculator , we calculated number of samples required as following:

Probability of enrolling, given click:
20.625% base conversion rate, 1% min d.
Samples needed: 25,835

Probability of payment, given enroll:
53% base conversion rate, 1% min d.
Samples needed: 39,115

Probability of payment, given click:
10.93125% base conversion rate, 0.75% min d.
Samples needed: 27,413

In the first iteration , we used the max sample size , which is 39115 to calculate the page view but later on (explained in next section) , it was observed that this was leading to very high duration to complete the experiment and hence we took the Retention Matric out and carried out our experiment with Gross Conversion and Net Conversion matric and since the base line probability values are not in 1:1 ratio , we had to convert our sample size to get the correct page views. Our correct page view came out to be **685325** ((27413/0.08)*2)
So at alpha = 0.05 and beta = 0.2 , we would need at least 685325 page views to get enough power for our experiment.

Duration vs. Exposure

I would divert the 50 % of the traffic of Udacity for this experiment even though this experiment is not going to overlap with any other experiment but I want to minimize the end user experience impact during my first phase.I am trying to be little safe in my first attempt in the experiment. If , I get expected results then in order to further gain confidence I may choose to divert entire traffic to the experiment.

Number of days needed to perform the experiment:
685325 page views / 20,000 pageviews per day = 35 days approx.

We can see that 35 days are a long period of time , so we may have to rethink about our decision, which is very subjective and can depend on various factors. In the any given situation , this decision is not in the hands of experiment designer only , there should be other groups involved and should take this decision on a mutual consent .This decision should align with the business objectives. For now, we can assume that we can conduct this experiment for 35 days.

Experiment Analysis
Sanity Checks
1.Number of Cookies/Pageviews:

Control Group	Experiment Group
345,543	344,660

Total Number of Pageviews (Control + Experiment) = 690203
Probability of a cookie = 0.3 (approximately)
Standard Error (SE) = sqrt(0.3 * (1 − 0.7) * (1/345543 + 1/344660))= 0.00055
Margin of error (m) = SE * 1.96 = 0.00235
Confidence Interval = [0.5 m, 0.5 + m] = [0.4989, 0.5011]
Observed Value = 345543 / 690,203 = 0.5006 , since observed value is within the expected range it **passes** the Sanity Check.

2. Number of Clicks:

Control Group	Experiment Group
28378	28325

Total Number of Clicks = 56703
Probability of a click = 0.5
Standard Error (SE) = sqrt(0.5 * (1 − 0.5) * (1/28378 + 1/28325)) = 0.00209
Margin of error (m) = SE * 1.96 = 0.00411
Confidence Interval = [0.5 m,0.5 + m] = [0.4959, 0.5041]
Observed Value = 28378 / 56703 = 0.50046, since observed value is within the expected range it **passes** the Sanity Check.

Click Through Probability:

Number of	Control Group	Experiment Group
Page Views	34543	344660
Clicks	28378	28325

Click through probability = 0.082125814
Standard Error = sqrt(0.082125814 * (1 − 0..082125814) * (1/345543)) = 0.000467053
Margin of error (m) = SE * 1.96 = 0.000915424
Confidence Interval = [0.082125814-m, 0.082125814 + m] = [0.0812, 0.0830]
Observed Value = 28325/ 344660= 0.0822, since observed value is within the expected range it **passes** the Sanity Check.

Result Analysis
Effect Size Tests
Gross Conversion :
d_min : minimum practical significance = +/0.01

Probability of Gross Conversion-Cont	3785/17293=.21887

Probability of Gross Conversion-Exp	3423/17260=.19831
Diff	-0.02055
SE	=.0044
ME	SE* 1.96(z score for 95% confidence interval)=.00856
Lower C.I	Diff-ME = -.02911
Upper C.I	Diff+ME = -.0120
Statistically significant	Yes, since Confidence interval does not contain zero
Practically significant	Yes, since confidence interval does not contain d_min value.

NetConversion :
d_min : minimum practical significance = +/0.0075

Probability of Net Conversion-Cont	2033/17293 = .11756
Probability of Net Conversion-Exp	1954/17260 = .11269
Diff	-.0049
SE	=.00343
ME	SE* 1.96(z score for 95% confidence interval)=.0067
Lower C.I	Diff-ME = -.0116
Upper C.I	Diff+ME = .0018
Statistically significant	No, since Confidence interval does contain zero
Practically significant	No, since confidence interval does contain d_min value.

Sign Tests

I used online sign test calculator to calculate p-values. Following are the results.
Gross Conversion:
alpha_individual = 0.05

Number of Success- Where exp Gross Conversion is greater than control group's Gross conversion	4
Number of Failure- Where exp Gross Conversion is less than control group's Gross conversion	19
Number of days	23
Probability	0.5
Two tailed p-value	0.0026

We can see that two tailed p-value is less than our alpha value hence results are statistically significant.
Net Conversion:
alpha_individual = 0.025

Number of Success- Where exp net conversion is greater than control group's Gross conversion	10
Number of Failure- Where exp net conversion is less than control group's Gross conversion	13
Number of days	23
Probability	0.5
Two tailed p-value	0.6776

The result is not statistically significant , which is inline with our expectations.

Summary
State whether you used the Bonferroni correction, and explain why or why not. If there are any discrepancies between the effect size hypothesis tests and the sign tests, describe the discrepancy and why you think it arose.

I did not use Bonferroni correction because in this case , we need both Gross Conversion and Net Retention matric to be significant (statistically and practically) and if results show that any one of the matric is not significant , launch is not recommended. Therefore, Bonferroni correction in this case is not necessary. Also ,in this experiment , we are not considering FWER , family wise error rate, and bonferroni correction can actually lead to false negatives.

Recommendation
I would not recommend the change. Following is my reasoning:

From our results , we see that results for Gross Conversion evaluation matrix are both statistically and practically significant, which means that the change(pop up) has positive impact on experiment user group i.e reducing the number of users enrolling after viewing the message. As expected, it should reduce number of frustrated users , who leaves the course in the middle. This should also allow the coaches to spend more time helping those students, who are really likely to complete the course. However, Net conversion evaluation matric results were neither statistically significant nor practically significant. This means that there is a risk that the introduction of the trial screener may lead to a decrease in revenue and udacity should consider test other designs before deciding whether to release the feature, or abandon the idea entirely.
        I think udacity should not launch this change even though Udacity could improve the overall student experience and improve coaches' capacity to support students who are likely to complete the course .This decision is in line with our hypothesis that the results might set clearer expectations for students upfront but can impact revenues(net conversion matric). From the experiment results , we can see that second part of hypothesis "**without significantly reducing the number of students to continue past the free trial and eventually complete the course"** , which pertain to Net conversion matric does not hold true. Hence , it is not recommended to take the risk and launch the change.

Follow-Up Experiment

In order to help udacity to reduce number of frustrated users , Udacity can perform following experiment. (Very similar to previous experiment but with different strategy)

When student click "start free trial", they should be shown a pop up with total number of Courses ,# of Pre-requisite courses, # of chapters , # of In chapter-exercises ,# of quizzes and # of projects that student would have to complete in order to finish the course successfully. Pop should have option to continue or Cancel (use the free course material). If student click on continue then they should be taken to next screen for enrolment else not.

Hypothesis : The hypothesis is that this might set clearer expectations for students upfront, thus reducing the number of frustrated students who left the free trial because they didn't expect the course to be so much intensive—without significantly reducing the number of students to continue past the free trial and eventually complete the course. If this hypothesis held true, Udacity could improve the overall student experience and improve coaches' capacity to support students who are likely to complete the course without any impact on revenues.

Unit of Diversion :
The unit of diversion is a cookie, students are are tracked by user-id after enrolment. The same user-id cannot enroll in the free trial twice.

Since this experiment is very similar to previous experiment , the choice of matrices is also going to be very similar.
Metric Choice
Invariant Metrics:
Number of Cookies:
It is the number of unique cookies to visit the course overview page.
> Since the Unit of Diversion is cookie and the number of cookies is not going to be affected by the change that udacity is experimenting at the time of enrolment . Therefore, number of cookies is well suited for invariant matric because there is no impact on the number of cookies.

Number of Clicks:
Number of unique cookies to click the start free trial button.
> Since, the page asking for the number of hours the student can devote to the course appeared after clicking the " Start free Trial " button, the course overview page remains the same for both the control and experiment group.

Evaluation Metrics:
Gross Conversion:
Number of users who enrolled in the free trial / Number of users who clicked the Start N Free Trial Button
> After clicking the Start Free Trial button, a popup appears for the Experiment group users asking for the amount of time that the Student can devote to the course. Based on the user choice it then makes a suggestion whether the student should enroll for the course or continue with the free course material.In the experiment group the user can make a decision based on the pop message and choose to continue exploring the material only. But for the control group no popup page appears, thus the user enrols for the course anyways. Hence, the gross conversion can be different in both control and experiment group hence can be used as an evaluation matric.

Net Conversion:
Number of user − ids to remain enrolled for the 14 days trial period and make their first payment /Number of users who clicked the Start Free Trial button

As per the intention and assumption of the experiment , experiment group users are made aware that the course require some minimum of hours each week by showing the pop message at the time of enrolment. This message should filter out those users who cannot devote the required hours and are prone to be frustrated later on. This ratio should be different among control group users and experiment group users, if experiment's assumptions holds true.Hence, it can be used as evaluation matric and for the same reason it cannot be chosen as invariant matric.

Gross Conversion, will show us whether we lower our costs by introducing the new pop up. Net conversion will show how the change affects our revenues. After the experiment, we expect that, Gross conversion should have a practically significant decrease, and Net conversion should have a statistically significant increase.