Design change on the Udacity website:
Testing if the new feature improves user experience and elevates retention

BACKGROUND

BUSINESS CONTEXT

- Udacity is an educational organization offering massive open online courses.
- It has invested heavily in the course content, leaving the design team with sparse resources for quantitative data-driven decisions

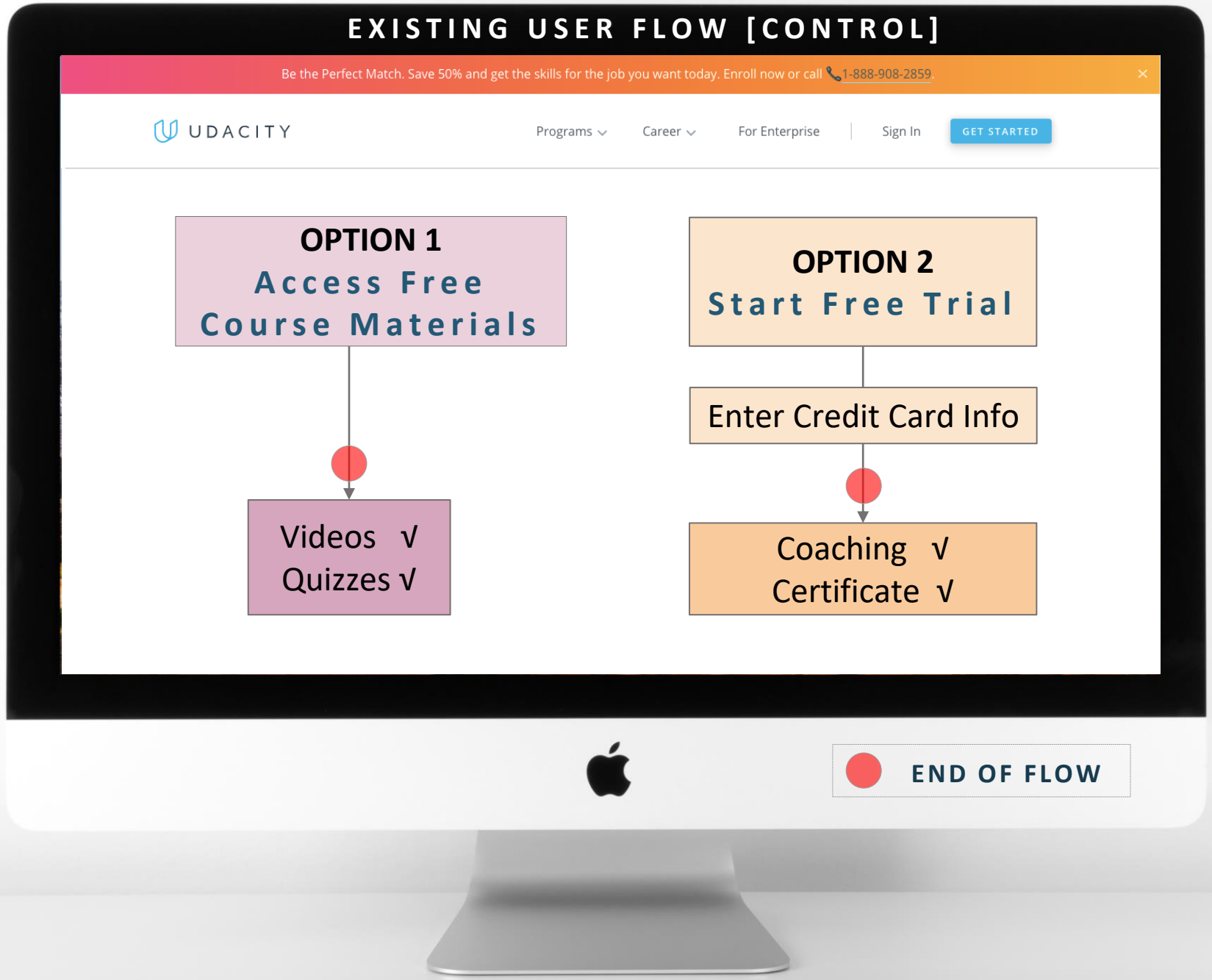
CAUSAL QUESTION

The design team wants to assess if a proposed new feature improves user experience and elevates retention

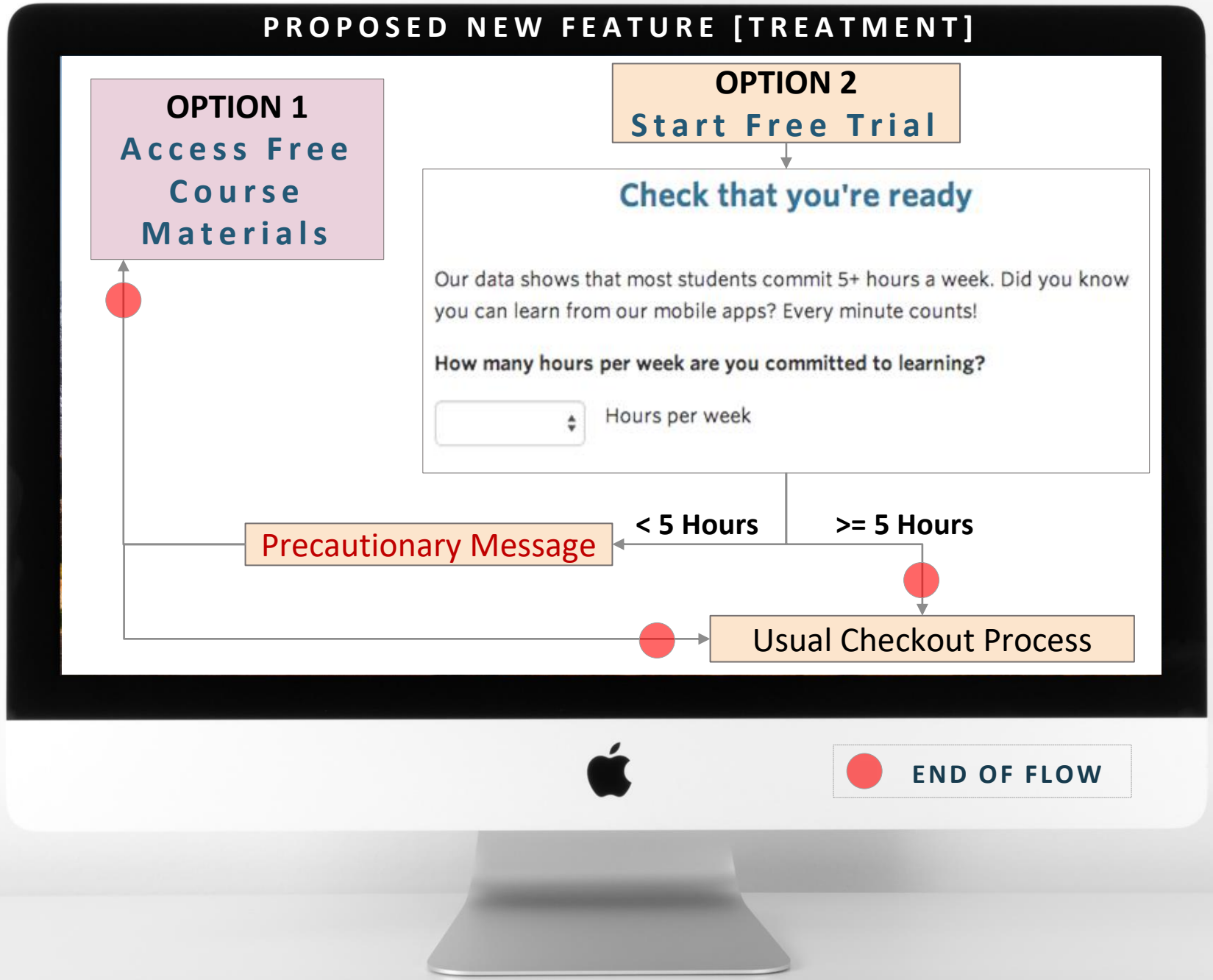
IMPORTANCE

It is useful to understand the causal impact of this design feature, because this design feature is intended to increase positive customer experience and bring in more revenue per user who enrols for free trial

EXISTING USER FLOW [CONTROL]



PROPOSED NEW FEATURE [TREATMENT]



Threats to Causal Inference

The profitability is determined by initial investment in purchasing properties, as well as the subsequent returns from renting that property

1 NO BASELINE FOR BENCHMARKING	Since we do not have data of any other online teaching platform as our baseline, it's possible that we overestimate the influence of this experiment.
2 INTERFERENCE	Even though the experiments are randomized, the users who are a part of the experiment might know each other. The behavior of people towards the experiments might deviate from natural free choice in that case
3 MEASUREMENT ERROR	Some users may pay for the courses after free trial if they forget to cancel the enrollment. This might lead to measurement error of one of our independent variables, which is enrollment.
4 OMITTED VARIABLE	There might be other unaccounted variables which influence the three performance evaluation metrics, for example inequitable exposure of test and control populations to some advertising campaigns run by Udacity

2

Analysis of Experiment

DATASET/ MEASURES

METHODS: CAUSAL ANALYSIS

RESULTS

IMPLICATIONS

Considering the limited budget, the Facebook's Atlas platform only gives visibility to five attributes, for measuring the efficacy of our experiment

	DESCRIPTION	TYPE
DATE	Date the data is recorded	Date
PAGE VIEWS	Number of unique users to view the course overview page that day	Numeric
CLICKS	Number of unique users to click the course overview page that day	Numeric
ENROLLMENTS	Number of user-ids to enroll in the free trial that day	Numeric
PAYMENTS	Number of user-ids who enrolled on that day to remain enrolled for 14 days and thus make a payment	Numeric

We are considering three evaluation metrics to get a holistic read on the effects of the experiment

	DESCRIPTION
GROSS CONVERSION	$\frac{Enrollments}{Clicks}$
NET CONVERSION	$\frac{Payments}{Clicks}$
RETENTION	$\frac{Payments}{Enrollments}$

Methods employed in the quantitative analysis

1

RANDOMIZATION CHECK

Is the treatment and control population truly randomized for the experiment? (This task was performed by a black-box model in Atlas platform)

2

POWER TEST

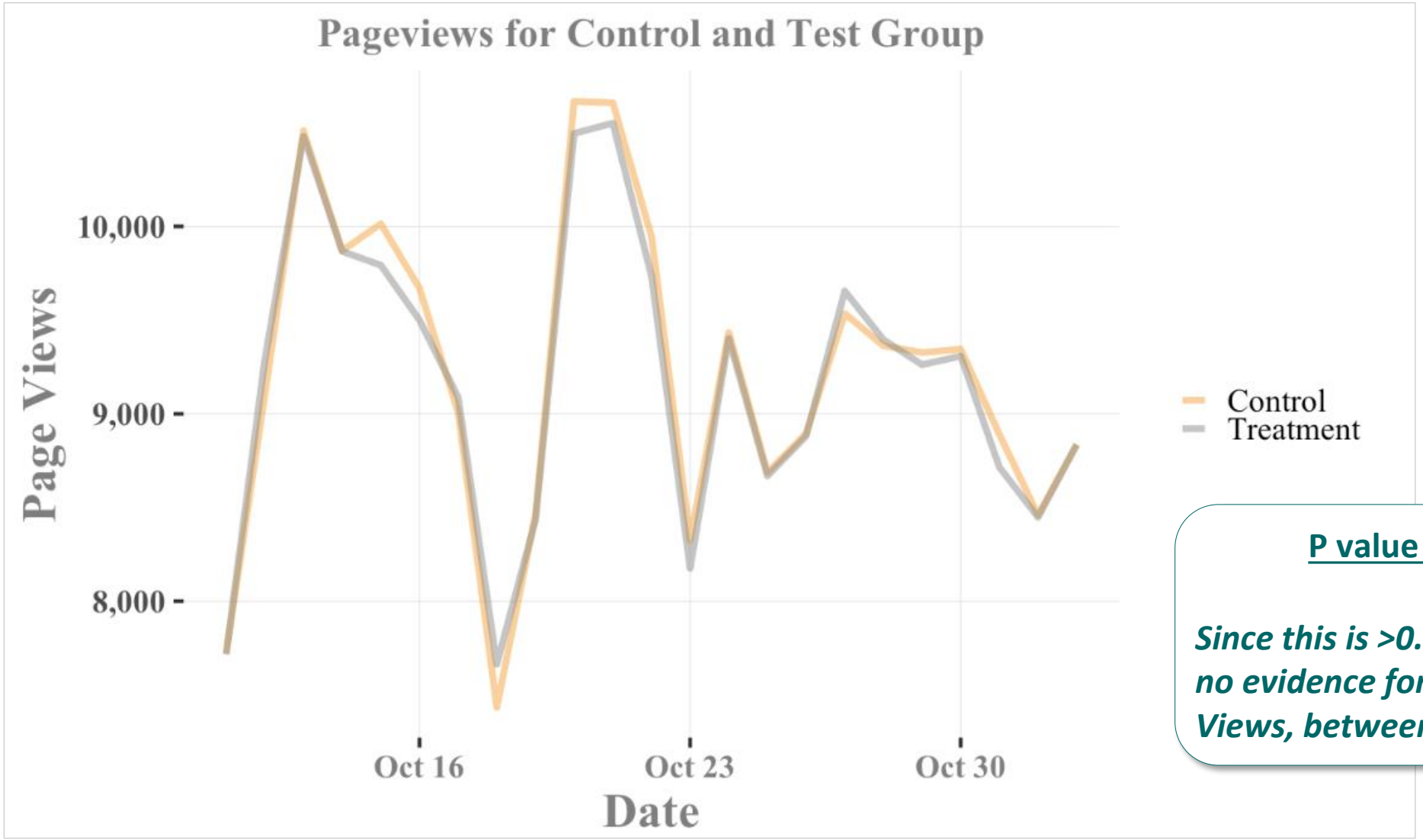
How much data is required?

3

TEST FOR CAUSALITY

Does the new feature significantly affect our evaluation metrics (metrics where we expect change)?

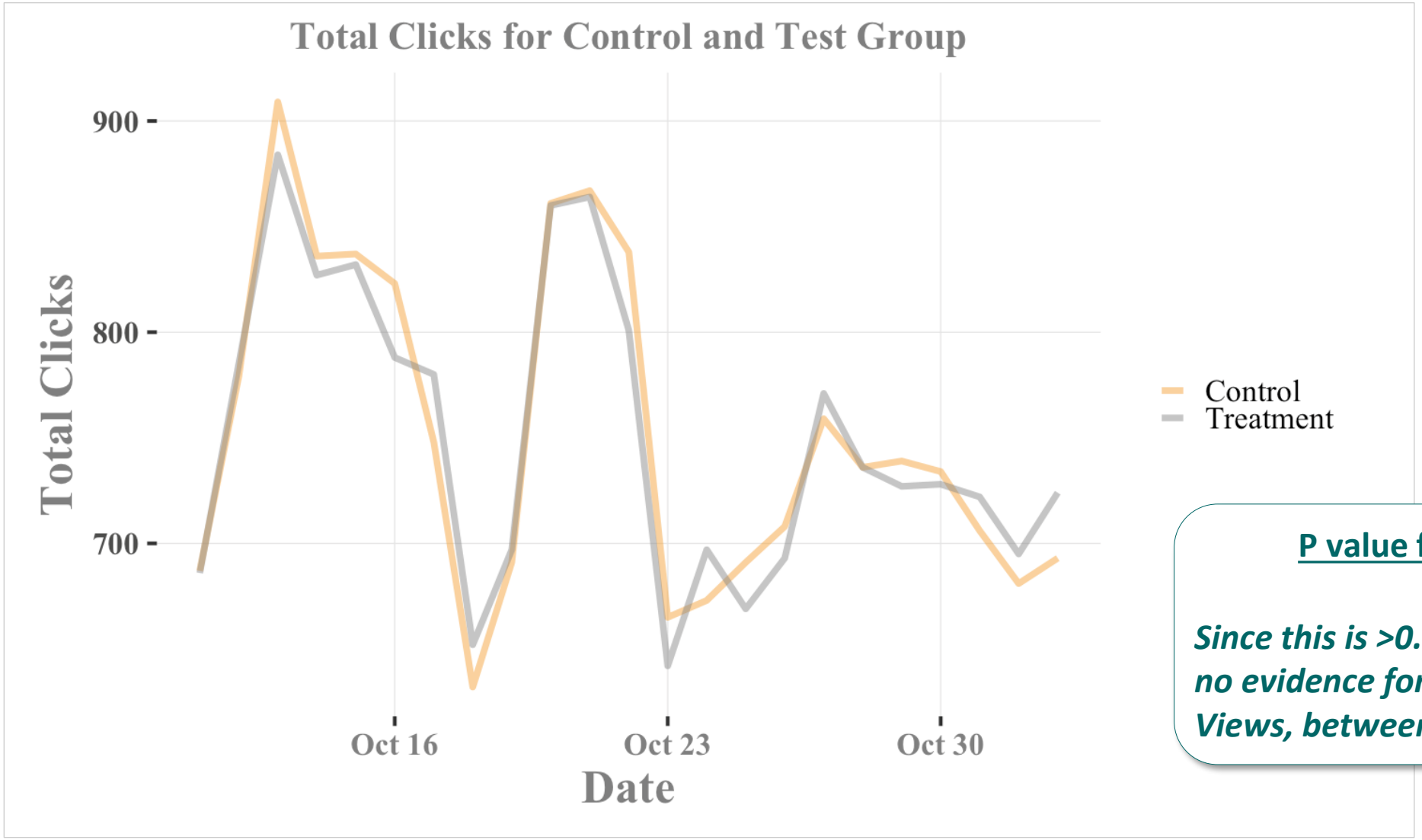
Randomization Check to assess if the Daily Page Views on Udacity website are invariant between Test and Control Group



P value from T-Test: 0.887

Since this is >0.05, it implies that there is no evidence for difference in Page Views, between Test and Control groups

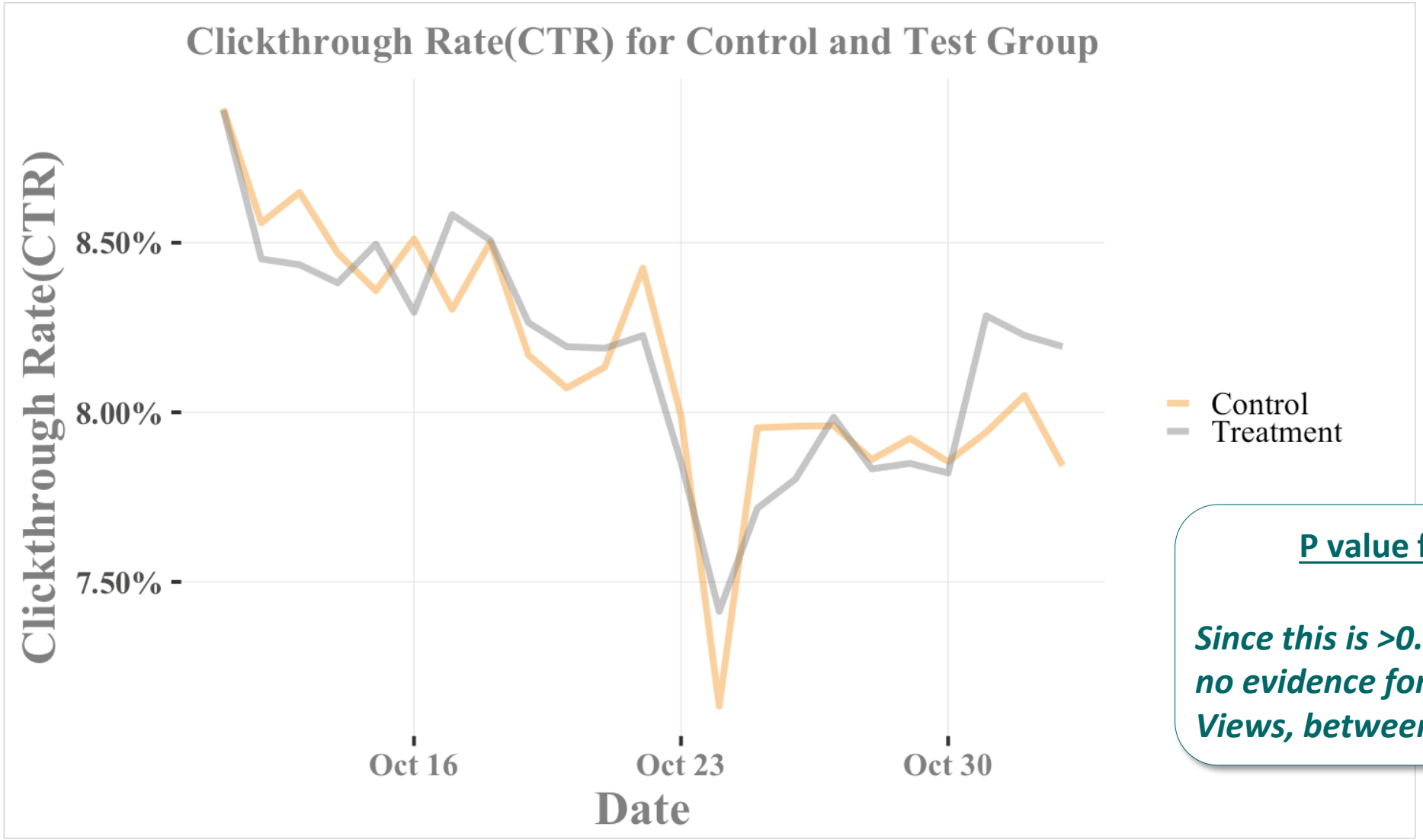
Randomization Check to assess if the Daily User Clicks on Udacity website are invariant between Test and Control Group



P value from T-Test: 0.9477

Since this is >0.05, it implies that there is no evidence for difference in Page Views, between Test and Control groups

Randomization Check to assess if the Daily Click-through Rate (CTR) on Udacity website are invariant between Test and Control Group



P value from T-Test: 0.8753

Since this is >0.05, it implies that there is no evidence for difference in Page Views, between Test and Control groups

What does this mean?

We have successfully performed the sanity checks, to assess if the test and control groups are not inherently biased in terms of randomization

Randomization Check to assess if the Daily Click-through Rate (CTR) on Udacity website are invariant between Test and Control Group

	STATISTICALLY SIGNIFICANT DIFFERENCE BETWEEN TEST AND CONTROL POPULATION?
GROSS CONVERSION	YES
NET CONVERSION	NO
RETENTION	NO

What does this mean?

- ***A statistically and practically significant decrease in Gross Conversion was observed but with no significant differences in Net Conversion.***

- ***This translates to a decrease in enrollment not coupled to an increase in students staying for the requisite 14 days to trigger payment***

Deeper Dive into how the Gross conversion metric is being affected by the experiment

- Gross Conversion

$$\frac{Enrollments}{Clicks}$$

VARIABLE	COEFFICIENT	SIGNIFICANCE
Treatment	- 0.019	Significant
Pageviews	0.000084	Very Significant
Clicks	- 0.0011	Very Significant
Weekend	- 0.0031	Not significant

Interpretation: The treatment group tends to have 0.019 decrease in gross conversion compared with control group.

Power Test for sample size

Metrics	Baseline	Required Detectable Change	Required Data Size
Gross Conversion	0.20625	1% change	6278 days
Net Conversion	0.53	1% change	5540 days
Retention	0.109313	0.75% change	17201 days

Reality - *Only have 23-day data!*

3

Limitations and Recommendations

The experiment has multiple limitations, which can be addressed to some degree with a higher budget allocation for tracking clickstream data

1 DURATION OF EXPERIMENT	• Data is available for a month. It may be insufficient to establish a measure of success (causal relationship) with strong certainty
2 SECOND ORDER EFFECTS	• We are missing customer attribute information. As such, even if the treatment yields an overall positive effect, we do not know if it has a negative outcome on a certain section of users. • In other words, there is no way to breakdown the analysis by customer segments and observe the variation in each segment
3 ETHICAL CONSIDERATIONS	• By showing a calculated warning to a section of users in the test group, and not to the control group, we are revealing more information to them in order to decide. • This is putting the users in the control group at a disadvantage. Informed consent has also not been taken from users.

Recommendations to the executive team

1

DEEP DIVE INTO USER JOURNEY DURING FREE TRIAL

- A successful experiment would be one in which there is a signficant decrease in Gross Conversion coupled to a significant increase in Net Conversion.
- Additional deep dive analysis into user journey during free trial, may give insights into required invention for improving retention

2

INVEST MORE RESOURCES FOR USER LEVEL ATTRIBUTES

- Facebook's ATLAS platform can provide access to additional user level information, with a premium access
- With more resources, the experiment should be run for a longer duration to pick up stabilized effect of the new feature
- Access to new variables can also allow study into the people who entered <5 hours Vs. >=5 hrs