A/B Test Analysis

Free Trial Screening Initiative

Team 8

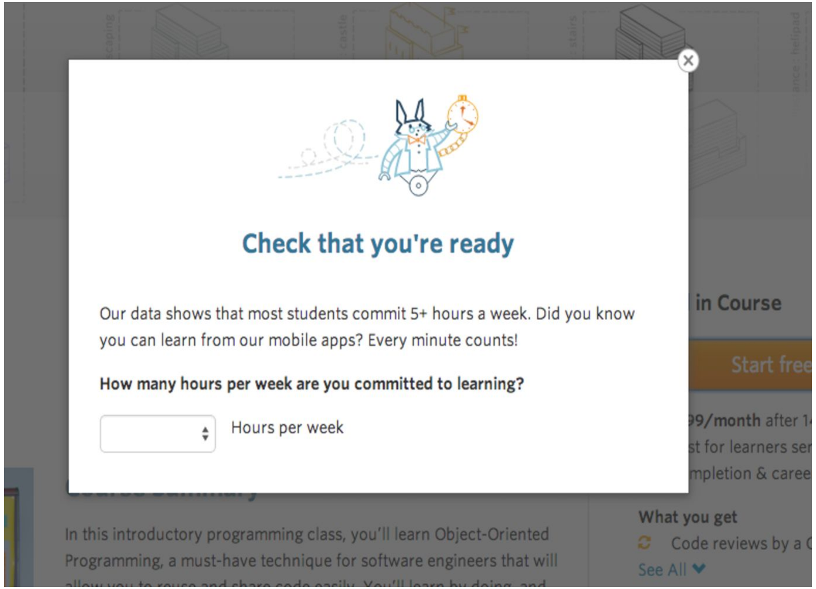
Joseph Miguel, Jeremy Grace, Jonathan Hilgart, & Victor Vulovic

Abstract

- Udacity noticed that there is a high level of frustration by users who don't have the recommended time and are billed.

- This experiment offered these users a free 14 day trial offer in an attempt to relieve those issues via a squeeze box.

- Results indicate a >96% confidence that the Retention Rate increased by 1%.



Introduction and Hypothesis

Problem: Number of 14-day free trial user dropout is high and it’s affecting our overall user experience

Objective: Introduce a new layer of expectations and gauge whether the feature has a positive effect on the experience and retention

Hypothesis: Users who can dedicate less than 5 hours per week will have a higher retention rate when offered a 14 day test drive trial period before being billed.

Ho : Retention rate change = 0	*The Retention rate would remain unaffected.*
H1 : Retention rate change >= 0.01	*Give users a free 14-day trial so they can test-drive our product before having to make a payment.*

Metric Choice from Baseline Data

Invariant metrics

Number of cookies & clicks

- # of unique cookies to click the "Start free trial"
- # of unique cookies to view the course [***Clicks***] overview page [***Pageviews***]
- Happens before the free trial screener is trigger

Click-through-probability (CTP)

- # of unique cookies to click the "Start free trial" button divided by number of unique cookies to view the course overview page.

Evaluation metrics

Retention rate:

$$\frac{\textit{Payments}}{\textit{Enrollments}}$$

Net conversion rate:

$$\frac{\textit{Payments}}{\textit{Clicks}}$$

Sample Size Needed to Conduct Test

Assumptions:
- Alpha = 0.05
- Beta = 0.2
- Retention dmin = 1%
- Net Conversion dmin = 0.75%

Retention Rate
- 39,051 enrollments ~ **2.37M pageviews**

Net Conversion Rate
- 27,978 clicks ~ **350k pageviews**

Choose execution strategy: Duration vs. Exposure

- Baseline pageviews/day = **40k**
- 60 day experiment window @ 100% user base participation

Goals: Determine minimum sample size of pageviews to power our experiment

```
import statsmodels.stats.api as sms
```

Sample size

```
In [21]: # Enrollment sample size for improving retention benchmark
es = sms.proportion_effectsize(0.53, 0.53+0.01)
sms.NormalIndPower().solve_power(es, power=0.8, alpha=0.05, ratio=1)

Out[21]: 39050.67796811014
```

```
In [22]: # Clicks sample size for improving net conversion benchmark
es = sms.proportion_effectsize(0.109313, 0.109313+0.0075)
sms.NormalIndPower().solve_power(es, power=0.8, alpha=0.05, ratio=1)

Out[22]: 27977.96232476509
```

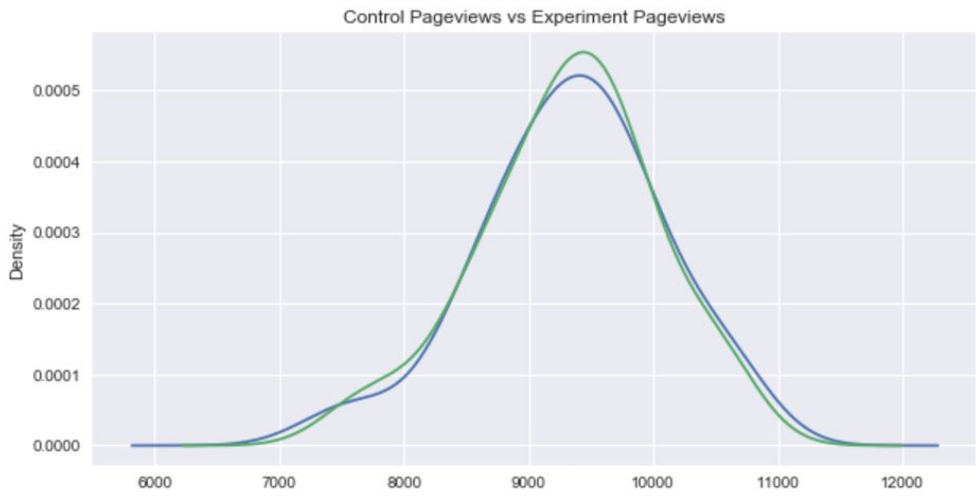
Analyze the Results

Sanity Checks

Invariant metrics (CTP, Pageviews, Clicks)

- CTP
  - The 1.96 z-score confidence interval is: -0.12% to 0.14%
- This is NOT significant (good!) our invariant metric holds

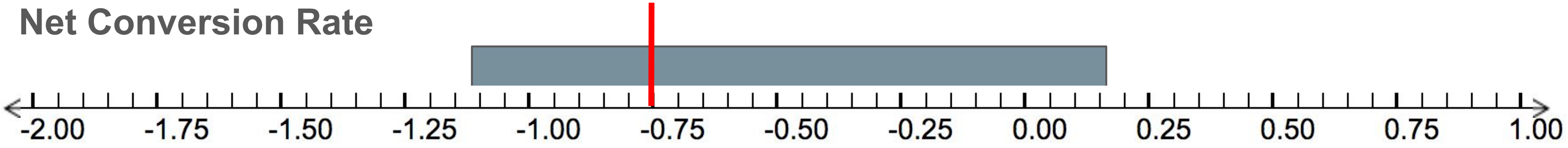
- Pageviews and Clicks
  - About 50/50 split of traffic



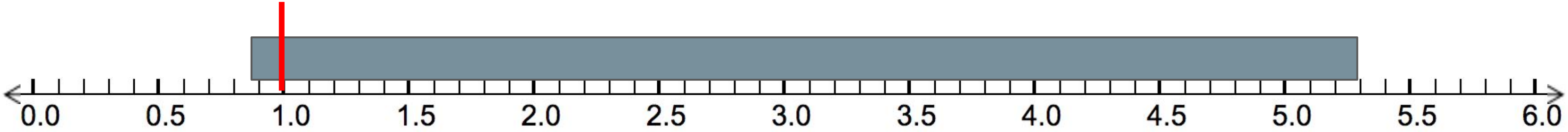
Frequentist Analysis

- Frequentist analysis can be used to evaluate the net conversion rate because we have an appropriate sample size. Can NOT use for Retention rate.

Net Conversion Rate



Retention Rate



** All values in percentages*

Bayesian Analysis

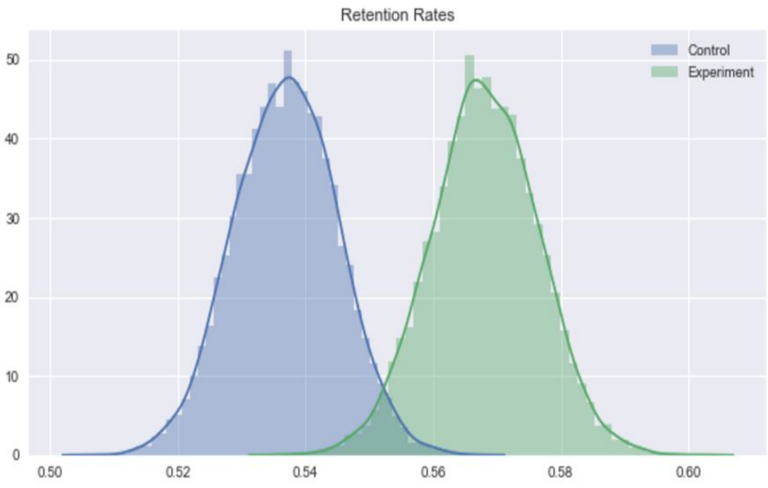
$$\Pr(p_B > p_A) = \sum_{i=0}^{\alpha_B - 1} \frac{B(\alpha_A + i, \beta_B + \beta_A)}{(\beta_B + i) B(1 + i, \beta_B) B(\alpha_A, \beta_A)}$$

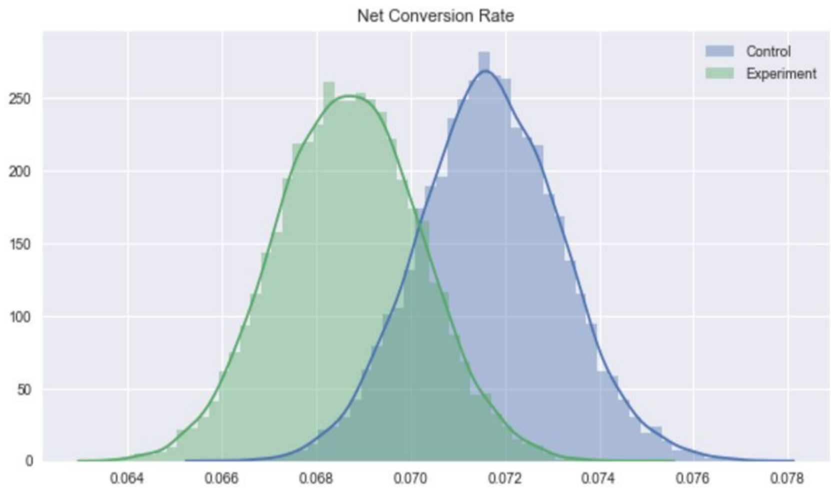
```
In [215]: with pm.Model() as retention: # model specifications in PyMC3 are wrapped in a with-statement
              # Define random variables
              control_prior = pm.Beta('control_retention', alpha=1 + control_number_of_payments,
                                      beta=1+control_number_of_enrollments - control_number_of_payments ) # uniform prior
              experiment_prior = pm.Beta('experiment_retention', alpha=1 + experiment_number_of_payments,
                                         beta=1+experiment_number_of_enrollments - experiment_number_of_payments) # uniform prior


              # Inference!
              start = pm.find_MAP() # Find good starting point
              step = pm.Slice() # Instantiate MCMC sampling algorithm
              trace = pm.sample(10000, step, start=start, progressbar=False) # draw posterior samples using slice sampling
```

- PyMc3 is a library to model stochastic processes
- Here, we are simulating the equation above to calculate the probability that our experiment performs better than our control. (posterior probability)

* http://www.evanmiller.org/bayesian-ab-testing.htm
* http://twiecki.github.io/bayesian_pymc3_europy_ab.slides.html#/l

Bayesian Markov Chain Monte Carlo Analysis





- We are 99.59% confident that our experiment significantly improved Retention Rate.

- We are 96.44% confident that our experiment significantly improved Retention Rate by over 1% (dmin)

- We are 91.59% confident that our experiment avoided making an improvement to Net Conversion Rate.

- We are 99.99% confident that our experiment avoided making an improvement to Net Conversion Rate by .75% (dmin)

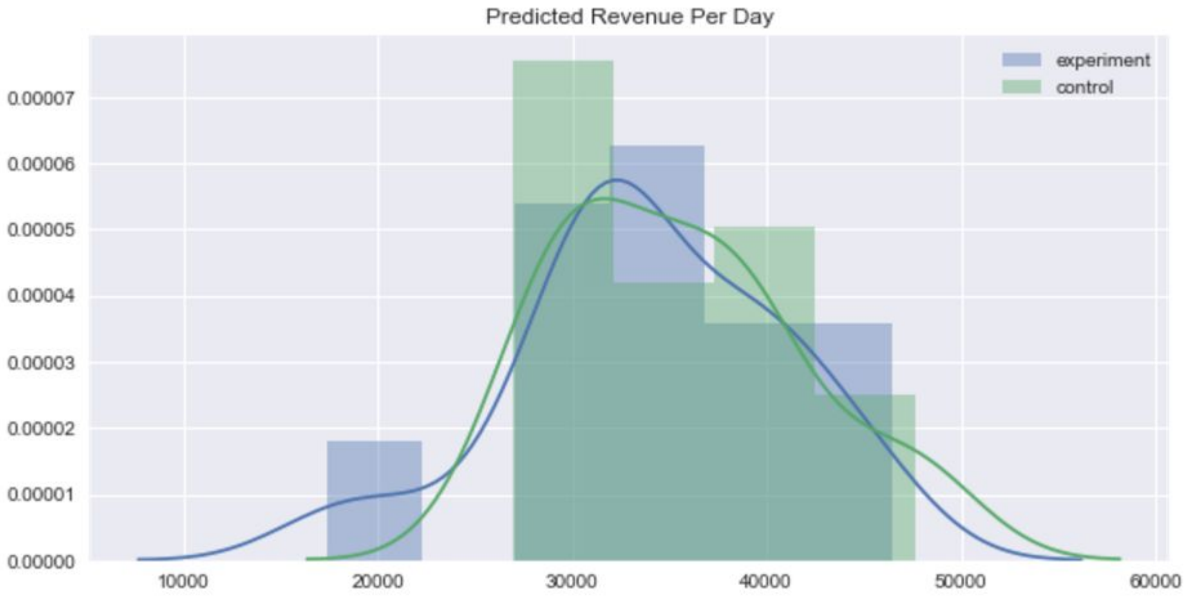
** Used 10,000 Monte Carlo samples. X-axis in percent*

Business Impact

Predicted Revenue Control :
- $817,332.00
- RMSE: $25.36

Predicted Revenue Experiment:
- $781,516.0
- RMSE: $19.21

Userbase Growth:
- Reduce churn & CAC
- Increase LTV



- *Assume each payment is a 'term' and each term costs $400*
- *Random Forest Regression model with 1,000 trees*

Recommendation

Summary of Analysis:

- Frequentist - Inconclusive that net conversion rate (NCR) is affected. Not enough info for retention rate (RR).
- Bayesian - >96% confidence that retention rate will increase; 99% confidence of negligible negative effect on NCR
- Risk - small, positive impact on short-term revenue and user growth. Potentially fewer users per course.

Conclusion:

- Given the projected low risk, start diverting additional traffic to this squeeze page and re-evaluate when the frequentist sample size needed is reached.

Next Steps

The following tests can help better determine the impact of this squeeze page.

1. Varying trial days:
   a. 7 day trial
   b. 30 day trial
2. Collect payment information at the beginning or the end of the trial offer.
3. Test different courses to receive this test
4. Gather additional data, such as revenue, to properly assess LTV (Lifetime Value)
5. Reassess data once sample size requirement is met for frequentist analysis

Questions?