Local v Popular Home Screen: AB Test Pilot Results

Viagogo Case Study

Alokik Mishra

Columbia Univeristy
New York City, NY

01 10 2019

Main Report

Key Findings

Aggregate Conversion Rate

Variant has lower conversion rate.

Aggregate Bounce Rate

Variant has higher bounce rate.

Conversion Rate - Time Trends

Conversion rate differentials mostly persist over time

Bounce Rate - Time Trends

Bounce rate differentials heterogenous over time.

Experiment Info

- Experiment run from Oct 10th to Oct 30th 2014
- Control - Home page shows popular events
- Variant - Home page shows location based events
- Control visitors - 2996337, Variant visitors - 3045853
- Both split between returning and new users
- Both split between different channels

Conversion Rate

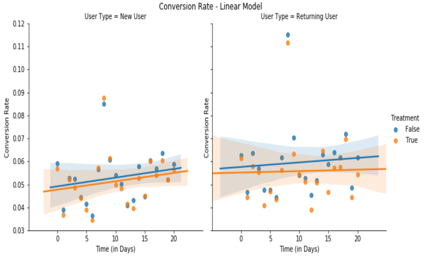
- The conversion rate for the variant was 5.31% compared to the controls 5.56%.
- The aggregate relative difference was -4.5%
- This difference is statistically significant at the 99% level

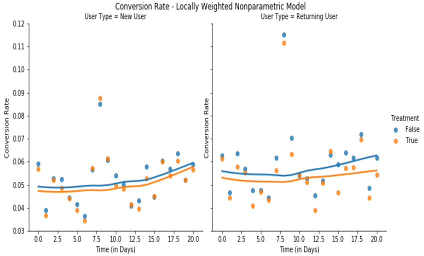
Bounce Rate

- The bounce rate for the variant was 41.27% compared to the controls 39.67%
- The aggregate relative difference was +4.0%
- This difference is statistically significant at the 99% level

Time trends - Conversion Rate 1

Conversion rates across time, conditioned on the type of user can be seen on the right. The top panel shows a linear OLS fit along with 95% confidence intervals. The bottom panel shows locally wieghted nonparatmetric trends which deal better with the outlier observation (Oct 18th). In the next page I highlight some important insights.

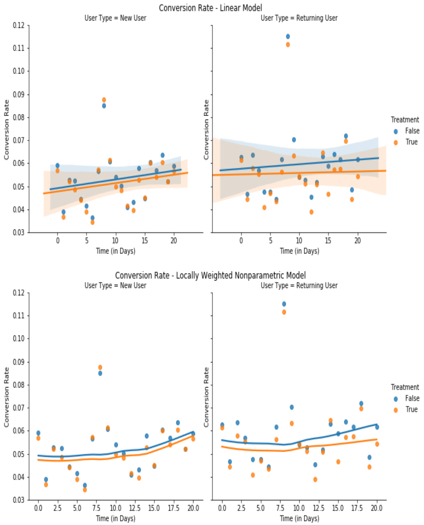




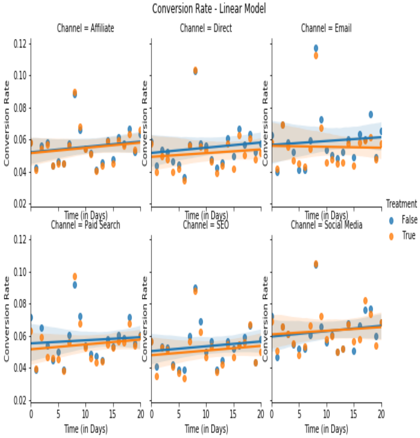
Time trends - Conversion Rate 2

1. Conversion rates rise across the board.
2. The difference between treatment and control stays constant for new users.
3. The difference between treatment and control increases for returning users.

Points 2 and 3 show that the damage to conversion rate by the variant may increase over time. Both due to new users $\rightarrow$ returning users, and the trends for returning users.



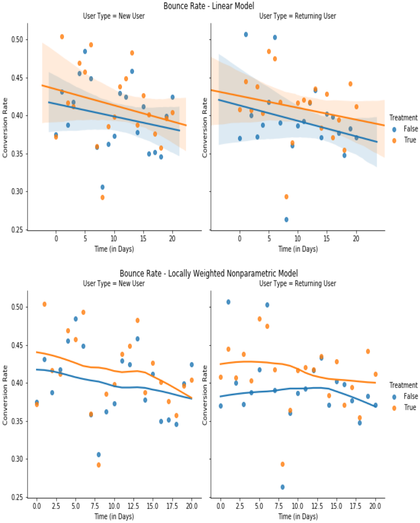
We turn to time trends conditioned on channel. There is not much that is surprising here. Across almost all the channels, the control outperforms over time. The one exception being social media (more on that later). The nonlinear fit isn't presented, but the results are similar.



Time trends - Bounce Rate 1

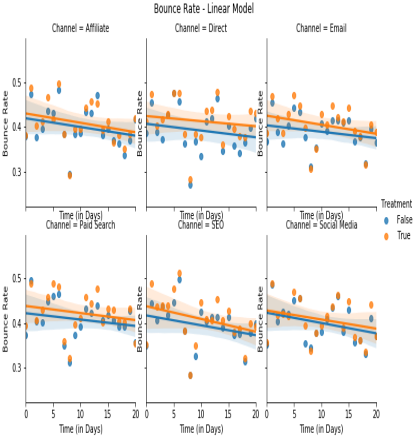
Key points to notice:

1. Bounce rates fall across the board.
2. The difference between treatment and control decreases slightly over time for new users.
3. The difference between treatment and control increases for returning users.



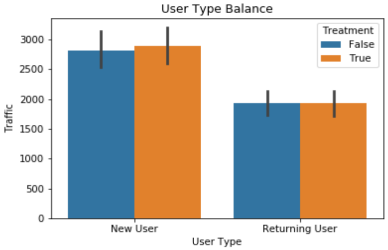
Time trends - Bounce Rate 2

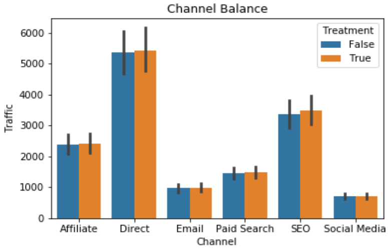
The bounce trends in do not seem to vary much across channel.



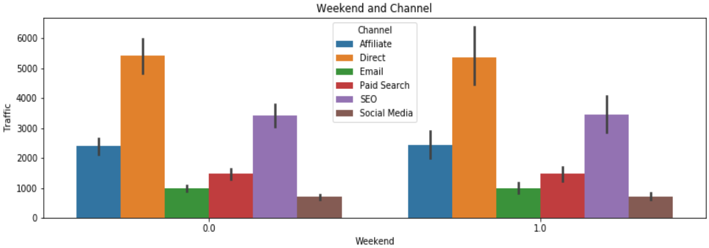
Data Exploration 1

Before further analysis, let's examine the data a little. To the right (top) we see the distribution of traffic across user type and see that new users dominate the market on average. We also see (bottom) that "direct" and "SEO" are the main methods through which users arrive at the page. Both user type and channel are balanced across treatment and control.





Below we see dsitribution acrosss different channels on weekdays (left) vs weekends (right). The distributions looks almost the same, suggesting that the weeknd traffic is not much different with regards to the distribution across different channels. The data overall is fairly balanced.



Next I look at an individual level analysis, where each observation is a person and we study what correlates to thier decision to make a purchase. Below I show the outcome of a logistic regression. The outcome variable is purchase. As expected we see a negative coefficient on the Treatment variable, showing that the local events home page negatively **affects** someone's liklihood or making a purchase. The data has been resampled to balance it out (see appendix for details.)

Variable	Coefficient	Std Error	P-value
Treatment	-0.0276	0.002	0.000

Table: Average Treatment Effect

In the next page we dig a little further.

Here we looks at how the treatment interacts differentially based on whether the person came from a social media channel and whether it was a weekend.

Variable	Coefficient	Std Error	P-value
Treatment	**-0.0972**	0.004	0.000
Weekend	0.1474	0.007	0.000
Social	**0.1180**	0.018	0.000
T x Social	**0.1081**	0.026	0.000
T x Weekend	0.0468	0.010	0.000
Weekend x Social	-0.0261	0.034	0.444
T x Weekend x Social	-0.0535	0.048	0.266

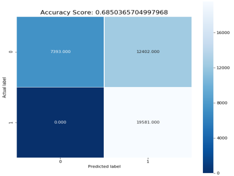
Table: Weekend and Social Media interactions

The results from the previous slide suggest that it is possible the variant page works slightly better on those users directed from social media, but **only** on the weekdays. This may be possible considering that if someone just wants to hang out with thier friends on a weekday (directed from social media), they might look for local (and possibly cheaper) events.

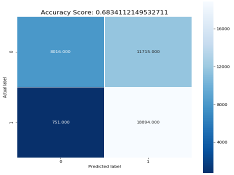
Predictions (optional)

Next I look at how well we can predict whether someone will buy based on just this data. The confusion matrices below.

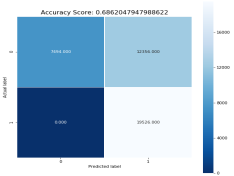
Logistic regression



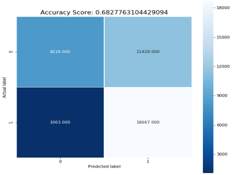
Random Forest



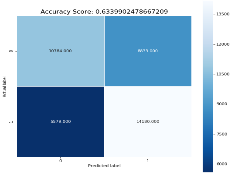
Neural Net



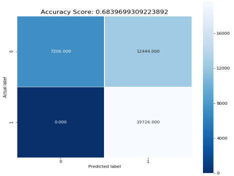
Decision Tree



K-nearest neighbor



Naive Bayes



Predictions (optional)

The frame above shows the limitations of the data with respect to a predictive model. All the models max out at around an accuracy rate of 68% on a testing set. This is expected as the data itself does not tell us much about the individuals themselves. For additional information on methodology, check appendix.

Conclusions 1

Based on the results of the test, my conclusion would be to revert to the control (popularity based) home page. The control does better in both the conversion rate and the bounce rate across most configurations. It is also important to note that the control does **even better** on the returning users. If market share is projected to grow, this is the pool that we would expect to increase over the long run, relative to the new user pool.

Conclusions 2

I am extremely confident in the performance of the control relative to the variant/treatment, however something I would like to know is the whether the types of tickets we sell are different between control and variant. For example if the types of tickets sold in the variant (local) are higher margin (fees) on average, it may make sense to take a lower conversion and higher bounce rate for a larger margin per conversion.

Other Improvements, Data, and Experiments

Other Improvements

Some possible improvements to explore are:

1. A simple filter bar below the search box
2. Switching static images to videos of trailers or similar events
3. Adding a 'suggested' event at the top for returning users
4. Adding a bar for each event corresponding to the percent of tickets sold
5. A button on each item that allows the user to return to it more easily later

Other Improvements - Data Requirements

I would need details from the engineering teams on the figures on both the fixed costs (developing the page), and the variable costs (supporting the bandwidth for x number of users) for each of the ideas. I would then want to use these figures when doing power and sample size analysis on the implementation of a 5x5 treatment cell. Not all cells need to be of equal size, and those with lower cost would be larger. Following the initial set of results, if the less expensive ideas clearly worked, I would implement those. If the more expensive ones seemed like they *could* work (desired direction and magnitude on the coefficients but high standard errors) , but were underpowered, I would then retest those.

Other Improvements - Evaluating Success

In order to evaluate the success I would focus on three metrics:

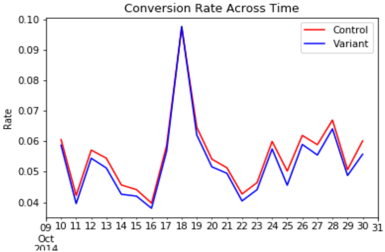
1. Conversion rate. Relevant to all.
2. Average margin per conversion. For example if users are buying more tickets but of lower margin (due maybe to a price filter), that may not be net postitive.
3. New to returning conversion. How many users who 'were' new have returned within a small-medium time period. This may be particularily relevant for item 5.

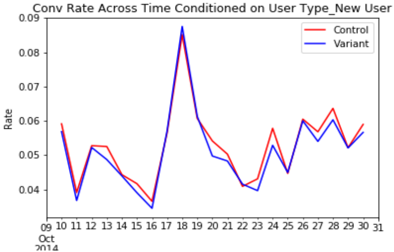
Thanks for your time and consideration!

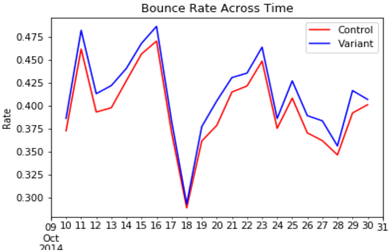
Appendix

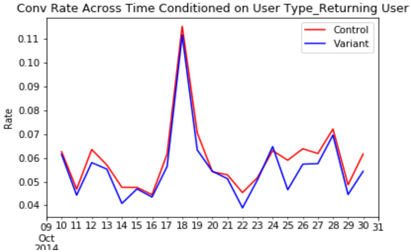
Raw trends

Time trends (raw values)



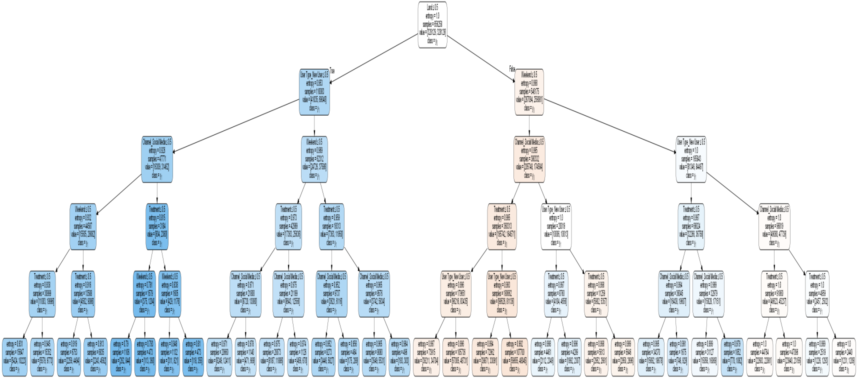






Raw trends

Decision Tree



Individual Analysis and Modelling Methodology

For the indvidual analysis I expanded the data based on the visitor number for each scenario to mimic a dataset with users as the observational unit. I added some features including a binary for the weekend, and a binary for the outlier day (Oct 18th). I then resampled the data as the amount with purchase equal to yes was about 5%, making most algorithms overweight the prediction of y = 0 (no purchase). For the resampling I simply resampled used the number of observations with a label of 1 (purchase), and sampled that number, with replacement, from the obervations with the label 0 (no purchase). To make sure the sampling didn't affect results, I used some bootstrapped samples, where I resampled multiple times.
The predictive modelling with then done using a 80% sample for traning and a 20% sample for testing, both model with and without interactions were tested.
All the analysis was done in python.