2019

Product Case Study

TESTING A CHANGE TO THE EVENT CATEGORIES ON THE VIAGOGO HOMEPAGE

AKSEL KOHEN

Table of Contents

THE EXPERIMENT

The current viagogo homepage displays the ten most popular categories by sales from customers. Clicking on a category takes the user to a new page with a list of events in that category. We'd like to investigate if changing this homepage set-up would boost its performance, and thus, the company revenue.

The new modified homepage would list the 10 categories closest to the user's location rather than the ten most popular by sales. We believe this proposal would significantly increase homepage performance and sales revenue. To test this hypothesis, we conduct an A/B test in which we show half our visitors the original version (control) and the other half the new version (variant) and compare the performance of these groups. The main metric we are interested in is the conversion rate, as defined below:

$$Conversion\ Rate = \frac{\#\ visitors\ to\ the\ homepage\ that\ subsequently\ make\ a\ purchase}{\#\ total\ visitors\ to\ the\ homepage^{1}}$$

The conversion rate is the proportion of visitors to the homepage who subsequently buy a ticket. It's the primary metric because it's directly tied to revenue. Another useful metric is the bounce rate, which is the proportion of visitors who navigate away (bounce) from the homepage after browsing only one page.

$$Bounce\ Rate = \frac{\#\ visitors\ that\ bounce\ from\ the\ home\ page}{\#\ total\ visitors\ that\ land\ on\ the\ homepage}$$

In other terms, the bounce rate is an indicator for the proportion of users that decide to stay on the website after just seeing the home page. It may not be as powerful as the conversion rate because it's not directly tied to sales. However, it's still quite important as it could give us important clues about the user-friendliness of the homepage, something we may not be able to directly infer from the conversion rate.

The data we have available for this experiment is the user visit dates, the channels which brought them, whether they visited before, whether they landed directly on the homepage, whether they bounced and whether they made a purchase.[2]

[1] It's important to note that by "total visitors" we mean the number of *unique individuals* who visited the homepage over the test period, not "total visits". This is important as users may convert (purchase tickets) on their first *as well as* subsequent visits. Using "total visits" would ignore sales cycles from previous visits as it would treat every visit independently. The result would be an underestimation of the conversion rate. The number of unique visitors is the number of new users since every user (returning or not) was at one point a new user.

[2] Some of these categories can potentially influence the decision to convert or bounce (like channel, user type). This experiment is to test changes resulting specifically from the homepage modification. Therefore, to avoid any potential bias from these other factors and better isolate the impact of the homepage modification, both the variant and control contain approximately equal amounts of users across these other categories

RESULTS

Aggregate Data Analysis

We begin with a collective analysis of all users, without segmenting them according to channels and user types. This form of preliminary analysis is useful in giving us a quick sense of what changes we might expect to see in our Key Performance Indicators (KPIs) if we were to change the homepage *across all users.*

The results at this stage certainly do not favor the variant. Over the whole test period, the variant conversion rate was about 5.5% less than the control's, meaning that far fewer people made a purchase in the variant group compared to the control after seeing the modified homepage (Figure 1)

The bounce rate paints a similar picture (Figure 2). Users who were shown the new homepage bounced about 4% more than those who were shown the original.

Figure 1: The Conversion Rate

Control Conversion Rate	Variant Conversion Rate	Agg. Relative Difference (%)	Statistical Significance (95% level)
0.094	0.088	-5.548	True

Calculations in Appendix A

Figure 2: The Bounce Rate

Control Bounce Rate	Variant Bounce Rate	Agg. Relative Difference (%)	Statistical Significance (95% level)
0.397	0.413	4.037	True

Calculations in Appendix B

While these numbers may seem like a lot of evidence against the variant, they are not conclusive enough. It's imperative that we also test for "statistical significance", i.e. if the difference between the variant and control is due to an actual statistical relationship. Even without any changes to the homepage, we would always expect some natural variation in the data for randomly taken visitor samples. After all, visitors that make up different samples are never identical to one another – they have different budgets, tastes of entertainment, preferences etc. Hence, the conversion and bounce rates across different samples will always vary to some degree, with or without a test variable.

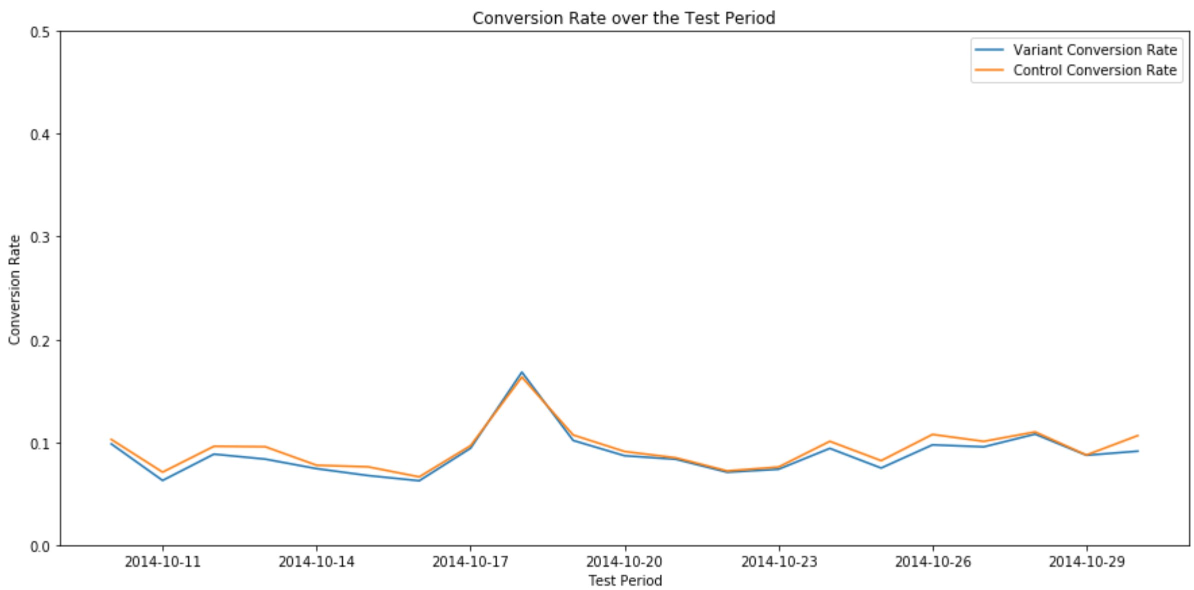
We want to make sure that the difference we observe between the variant and control is not due to this natural randomness but due to the homepage modification we are testing. Hence, we employ a hypothesis testing method called the " Paired T-test". [3] On an aggregate level

[3] In simple terms, the T-test estimates the probability that we would observe the sample data we were given under the assumption that the homepage modification has no impact on the KPI. If this value is below a pre-set threshold (0.05 is usually standard), we conclude that the homepage modification had a statistically meaningful influence over the KPI (and the differences between the variant and control are not due to natural variation). Paired T-test is a special form of T-test where every observation in one sample has a counterpart in the other sample

(considering all users), this test has revealed the variant to be statistically different from the control group for both the conversion and bounce rate, i.e. the poor performance across variant users is not due to chance but due to the new homepage. Hence, our preliminary finding is that changing the homepage *across all users* will probably damage the homepage performance and reduce revenue.

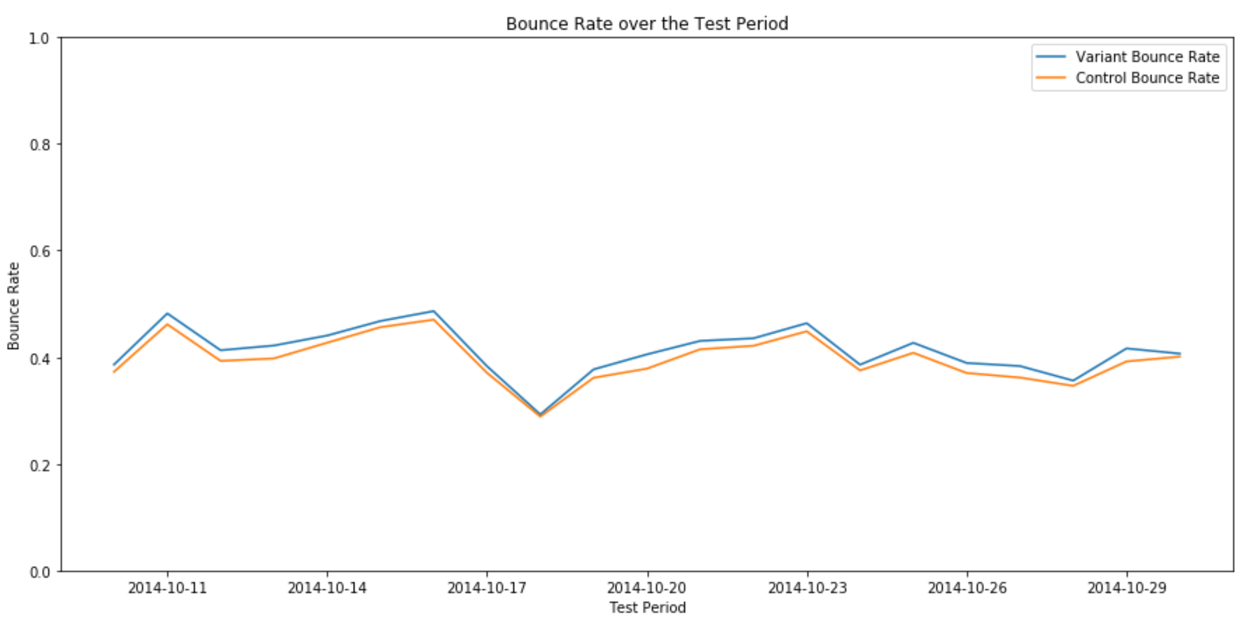
The calculations thus far have given us valuable insights into the overall performance of the variant and control, but they do not do much in terms of showing trends – which can tell us more about the consistency of any particular relationship. The charts below depict the trends in the conversion and bounce rate for the variant and control over the test period.

Figure 3: Daily Conversion Rate



Calculations in Appendix A

Figure 4: Daily Bounce Rate



Calculations in Appendix B

Looking at daily conversion rates in Figure 3, we see that both the variant and control follow the same steady pattern, with small rises and drops which roughly correspond to weekends (plus Friday) and weekdays respectively. On weekends, people have more leisure time so it makes sense they would be spending more time booking tickets for entertainment events – hence the slightly higher conversation rate (for both groups). One unusually strong day was October 18, 2014 (the peak day, a Saturday). My guess is that this was the day of an important event (say, Taylor Swift concert) and many users were surfing the viagogo website to buy last-minute tickets. (hence the high conversion rate)

More importantly, the control conversion rate was consistently higher than the variant conversion rate over the test period, which supports our initial finding that, when all users are considered collectively, the variant fares worse than the control. There was one notable exception, however – again October 18, 2014. Although small, this is the only day the variant had a higher conversion rate than the control. This makes intuitive sense; assuming there was a very important event on this day, those looking to buy last minute tickets would be those living close enough to the event to be able to attend. Therefore, the homepage modification – sorting categories according to user location – must have helped these users find available tickets, hence improving the conversion rate.

Compared to one another, the control and variant bounce rates follow quite similar trends as well. As we would expect, the rises in the conversion rate closely mirror the drops in the bounce rate (and vice versa) – more people converting means less people bouncing since they would have to visit more than just the homepage to make a purchase. October 18, 2014 is also significant here for the same reason discussed above. This is the day when the difference between the control and variant bounce rate was lowest - again attributable to the last-minute ticket seekers living close enough to a very important event.

Results section continued on next page

Subgroup Analysis

While the data as a whole may seem to be favoring the control (the original homepage) over the variant (the new homepage), this may not be true for subgroups present within the data. For example, a returning user who regularly comes from social media may have different expectations from the homepage than a user coming for the first time from direct marketing. Therefore, we need to segment the data into relevant categories and compare the control against the variant for each category. This approach will help us see which homepage set-up is more effective for which sub-group.

Let's start by segmenting by the user type – new user or returning user.

Figure 5: Conversion Rate by User Type

	Control Conversion Rate	Variant Conversion Rate	Agg. Relative Difference (%)	Statistical Significance (95% level)
New User	0.053	0.051	-3.134	True
Returning User	0.060	0.056	-6.272	True

Calculations in Appendix C

Figure 6: Bounce Rate by User Type

	Control Bounce Rate	Variant Bounce Rate	Agg. Relative Difference (%)	Statistical Significance (95% level)
New User	0.399	0.415	3.871	True
Returning User	0.393	0.410	4.259	True

Calculations in Appendix D

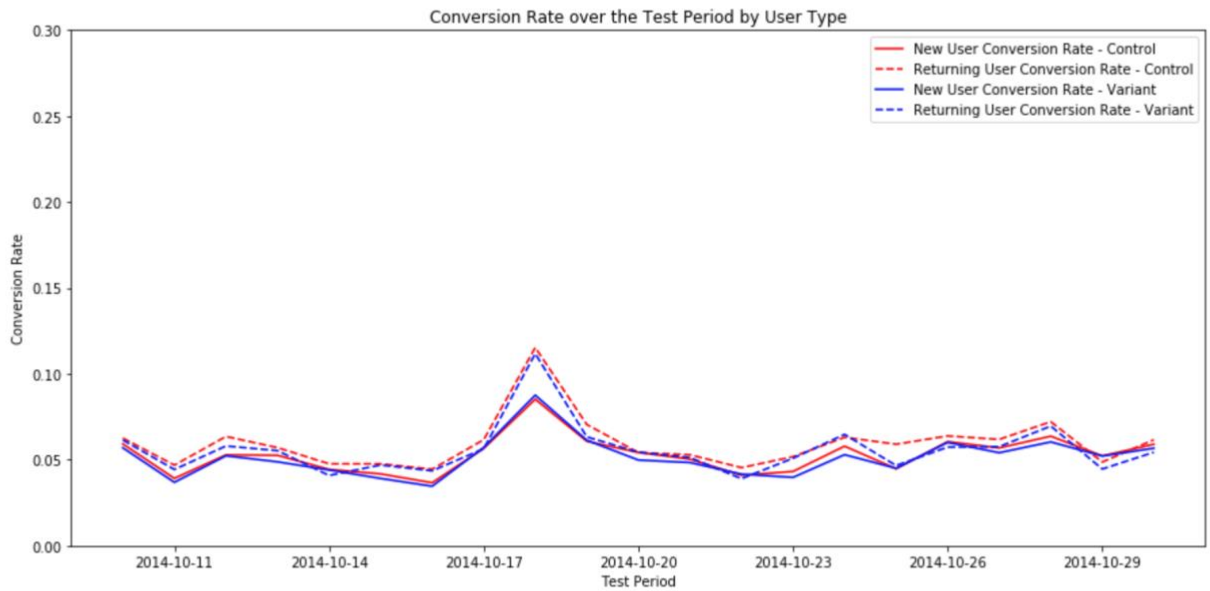
Returning users overall seem to convert more and bounce less (i.e. when compared within the same test group) – many of them have probably compared prices across different platforms already and are coming back with more readiness to buy.

The findings by user type mirror the aggregate-level findings as well as give us some additional interesting insights. Overall, the control performed better than the variant like we observed before. However, compared to new users, returning uses were much less likely to convert after visiting the new homepage (-6.3% versus -3.1%) and are more likely to bounce because of it (4.3% versus 3.9%). While both returning and new users clearly dislike the new homepage (the variant), returning users seem to dislike even more. We also have statistical significance for all sub-tests which tells us these findings must be linked to the new homepage design and cannot be random.

Let's consider the conversion and bounce rate by user type over the test period to get a better sense of trends.

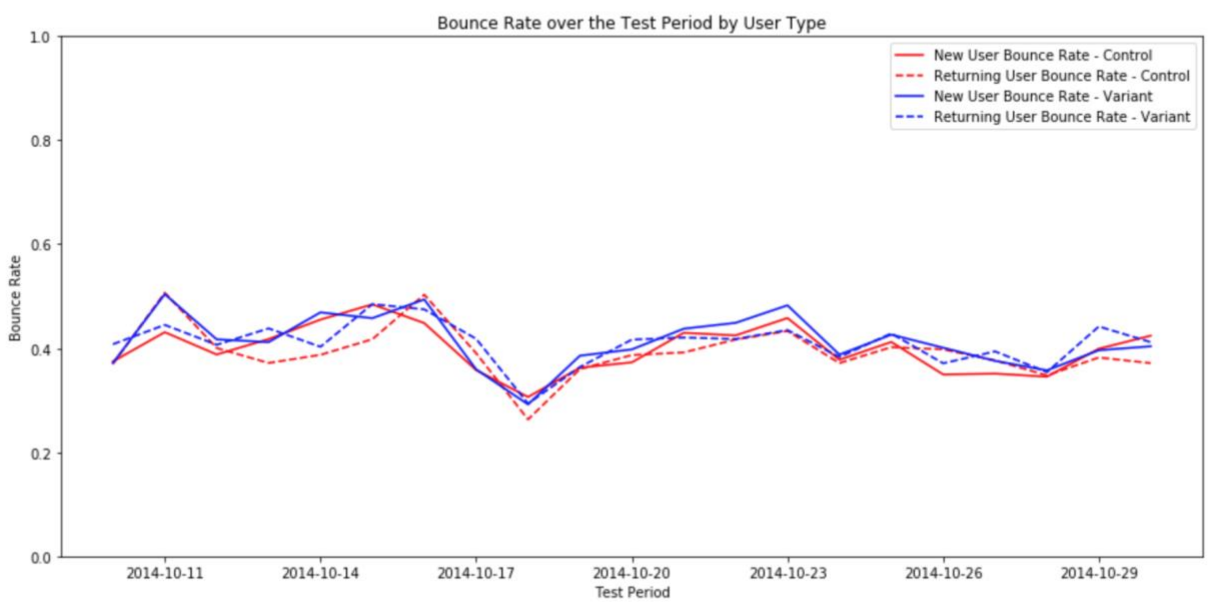
Charts on next page

Figure 8: Daily Conversion Rate by User Type



Calculations in Appendix C

Figure 9: Daily Bounce Rate by User Type



Calculations in Appendix D

Trends in the Daily Conversion Rate by User Type (Figure 8) supports an important finding we had made from the Daily Conversion Rate chart (Figure 3): that the control performed consistently better than the variant. We see this to be true with respect to both new and returning users. In addition, like we observed in Tables 5 and 6, returning users seem to have disliked the new homepage consistently more than new users - the control line for returning users is noticeably over the variant line for returning users, more than how much the control line for new users is over the variant line for new users. Furthermore, October 18, 2014 appears to be significant here also. It seems like returning users have converted much more compared to new users on this special day across both the variant and control samples (interesting fact but irrelevant to the homepage performance)

The trends we observe for the daily bounce rate by user type are much fuzzier than those for the conversion rate, with several trend lines repeatedly crossing over each other. We know (from Table 6) that on average a returning user bounced less frequently than a new one but the extent of it seems to vary significantly according to Figure 9 . This tells us that, as a metric, the bounce rate has more natural variation across this subgroup (new vs. returning user), making it difficult for us to isolate the impact of the homepage modification in this subgroup.

Now, let's apply a further level segmentation to test the homepage performance in more specific subgroups. Here, we consider every combination of user type and channel and compare the variant against the control within these subgroups.

Figure 9: Conversion Rate by User Type and Channel

User Type	Channel	Control Conversion Rate	Variant Conversion Rate	Agg. Relative Difference (%)	Statistical Significance (95% level)
New User	Affiliate	0.053	0.054	2.173	False
	Direct	0.053	0.050	-5.774	True
	Email	0.053	0.050	-5.647	True
	Paid Search	0.052	0.052	0.507	False
	SEO	0.052	0.050	-4.139	False
	Social Media	0.055	0.055	0.027	False
Returning User	Affiliate	0.060	0.057	-5.084	False
	Direct	0.057	0.054	-6.743	True
	Email	0.067	0.063	-6.376	False
	Paid Search	0.065	0.058	-10.817	True
	SEO	0.056	0.052	-6.291	False
	Social Media	0.075	0.075	0.782	False

Calculations in Appendix E

This table supports previous findings that overall the original homepage has performed better than the new one (the aggregate relative differences are mostly negative). It also provides us with additional insights and some notable exceptions. Specifically, for certain subgroups the variant was much worse than it was for others (e.g. (New User, Email), (Returning User, Paid Search)). For others, the variant was either negligibly worse or even resulted in an improvement (e.g. (New User, Affiliate),(New User, Paid Search)). Having "False" statistical significance for these subgroups may mean that the homepage modification perhaps wasn't "that bad" for the conversion rate in these subgroups. Yet it may also be that we simply do not have enough data to establish statistical significance in these small sub-groups – not that it doesn't exist. This uncertainty makes it difficult to make any conclusive remarks beyond what we already discussed.

EXPERIMENT LIMITATIONS

Before making any recommendations, we should discuss the limitations of the data we used to generate our findings. First of all, the data dates back to 2014; there is a chance that user and marketing demographics have changed since then. Hence the insights we gathered for various subgroups may no longer be true today. The data used also did not contain a unique ID for each individual visitor. This made it impossible to identify returning users who visited before the test period[4] and those who may have purchased tickets multiple times over the test period, adding a significant level of bias to our calculation of conversion rates.

User location would have also been very important in this experiment but was ignored due to lack of data. Presumably, individuals in different locations would react very differently to the new homepage over the current one. Those in rural areas or in cities with long distances and poor public transport may benefit from event categories sorted by proximity. On the other hand, those in densely populated areas may prefer event categories sorted by sales. Due to the absence of data, we were not able to test and control for this important factor.

RECOMMENDATIONS

The new homepage did not produce a statistically significant improvement over the original homepage. In fact, on the aggregate level and in many subgroups, it caused varying degrees of deterioration in the conversion and bounce rate – most of which were found to be statistically significant. Due to this overwhelming evidence, I recommend not deploying the variant and maintaining the current version of the website. Nevertheless, there is some evidence that in some niche cases the new homepage set-up may potentially be useful (days with very important events, the affiliate and new user subgroup). However we simply did not have enough data to establish statistical significance for these cases in our experiment. I recommend collecting more data on these cases and repeating the experiment.

As a next step, I also recommend undertaking the new experiment ideas outlined in the next section.

[4] total (unique) visitors were assumed to be new users when calculating the conversion rate so this group was essentially ignored due to lack of data

FURtTHER RESEARCH

A/B testing is a useful method in testing the effectiveness of potential improvements. I would like to propose five new improvements to the homepage which I believe are worth testing for:

1) *Show the most popular event categories in the user's city rather those most popular overall*

   Location is a big consideration for many event-goers. Users are more inclined to buy tickets for events they can travel to relatively easily. However, these need not be the events closest to them, so a city filter can be helpful for increasing the conversion rate

2) *Create and sort homepage categories based on past browsing and purchase history on the viagogo website (similar to an Instagram feed).*

   Personalizing homepage categories based on past browsing and purchase history may increase the chances a user will find events matching his interests and preferences. This may increase the conversion rate.

3) *Add homepage alerts for popular events that are selling out fast*

   This may help create faster conversions as many users will fear missing out on a great event and will want to buy sooner than later.

4) *Replace category photos with videos*

   Some users may prefer videos over photos as they might feel it gives them a better taste of what the actual event will be like.

5) *Repeat the experiment discussed in this paper (sorting categories by proximity to the user) for rural and urban areas as well as for cities with different population densities.*

   Like discussed earlier, users in certain areas may potentially benefit from categories sorted by proximity. Cities and rural areas where events are few and travel is difficult can be compared against urban areas and well-connected cities through a control and variant set-up similar to the experiment detailed in this written report.