Determining the Impact of
Booking Forms with Guest Ratings on Purchase Likelihood

An A/B test Statistical Analysis

By

Kinjal Majumdar

Contents

- Background

- Objective

- Proposed Methodology

- A/B Testing Analysis

- Results

- Recommendations

Background

Why are we considering changes to the Booking form?

When we put ourselves in a customer's shoes it becomes evident that like us, many users look for validation – often till the very last minute- that they have made the right purchase decision.

It is therefore critical to explore and ascertain the value of any product changes – i.e. changes to the booking form – in order to:

Assuage customer doubts, and ***boost purchase likelihood***.
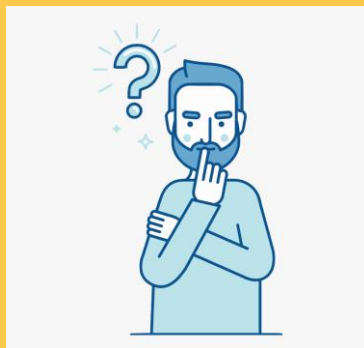
Objective

The product team wants to facilitate an increase in conversions

Expedia

In order to meet its revenue and engagement targets, the product team has proposed a change to the booking form that could increase a user's likelihood to purchase.

We want to focus on the question:

Will replacing star ratings with customer's guest ratings on the booking form, reinforce customer confidence in the choice they have selected?
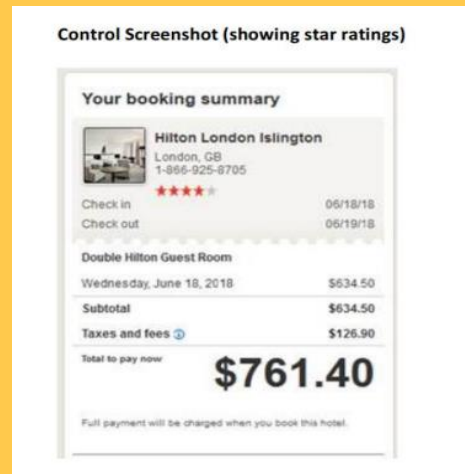
★★★★   vs.   4.5 out of 5
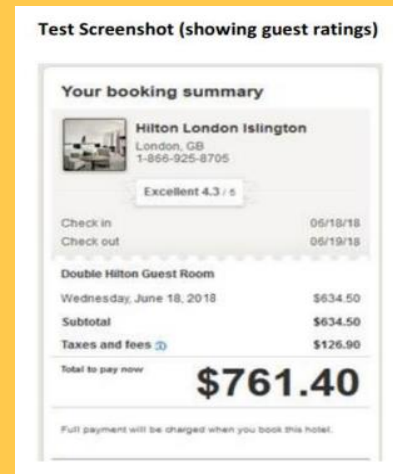
What do we expect to discover?

Expedia

Hypothesis:

Replacing star ratings with customer's guest ratings will have a significant impact on confidence that the user has made the right choice and therefore increase their likelihood to purchase.

Current   VS   **Challenger**

Proposed Methodology

An A/B test allows us to compare and evaluate changes to the Booking Form

- We test the hypothesis to evaluate changes in the product and observe if a change will lead to **improved performance in a specific KPI – conversions**.

- The goal of A/B testing is to track the primary metric during the test period and **find out whether there is a difference** in the performance of the product and if so, what type it is.

Methodology of an A/B test

1. Determine metric to track from hypothesis – did a customer purchase (convert)?

2. Clean the data. Treat nulls and duplicates

3. Allocate control group – users who are not shown the new variant.

4. Allocate variant group – users who are shown the new variant.

5. Determine a statistically significant sample size, to be able to generalize the findings to the overall customer base.

6. Compare the control and variant groups to ensure its an apples to apples comparison.

7. Determine null and alternate hypothesis.

8. Determine confidence level desired. An accepted convention is a confidence level of 95%.

9. Test the hypothesis using the normal approximation method or 2 sided z-test.

10. Validate or invalidate the null hypothesis.

11. Draw conclusions.

A/B Testing Analysis

Our results will help validate or invalidate the null hypothesis which is based on the original hypothesis

Null

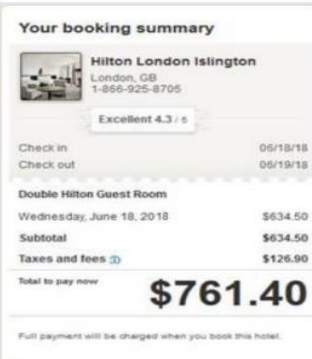The two variants shown on the price panel do not perform significantly different.

Current

Challenger

Alternate

The two variants shown on the price panel perform significantly different.

Current

Challenger

For this analysis we will use the normal approximation test (2 sample z-test) to test our hypotheses

Expedia

A z-test can be justified for the following reasons:
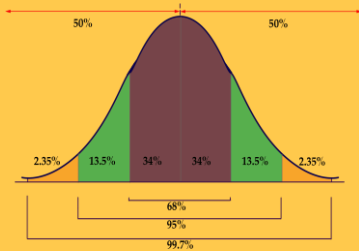
Both control and variant groups are normally distributed.
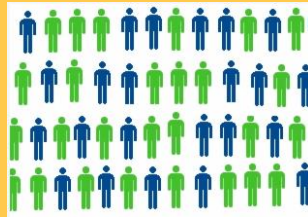
We have a large enough sample size.

When we have quantitative data points.
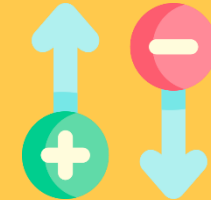
We're looking for a specific effect from our A/B test. i.e. variant performs better or worse than control.

Results

Displaying the customer's guest ratings does a better job of boosting customer confidence and purchase likelihood

Expedia

- The results of the z-test are statistically significant and since p value < α we can reject the null hypothesis that there is no difference between the star ratings and the customer's guest ratings.

- We can say that there is a difference in the likelihood to purchase between the two booking forms.

- Even though there is a **0.11% lift in purchase likelihood due to guest ratings**, the z-test determines that this is a statistically significant impact

We can conclude that the guest ratings booking form is the more optimal compared to the star ratings booking form as it facilitates more guest purchases.
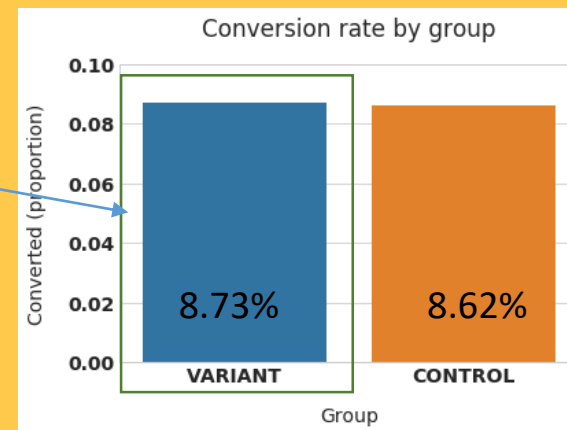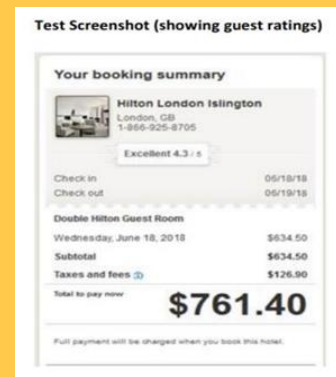
WINNER – Booking Form with Guest Ratings

Recommendations

We recommend the following actions on the website

- Run an **NPS survey** to further validate and understand customer sentiment around the booking form and guest ratings.
- Since the **New, Latin America, Web and SEO customers responded the most favorably to guest ratings** on the booking form, they could be the **low hanging fruit for further experimentation**.
- Similar to the above, we can start rolling out the guest ratings change to the **booking form in phases**.
    - Since New, Latin America, Web, and SEO customers had the highest lift in purchase likelihood, we can start rolling out the changes to these customers first, and observe the impact on conversion and revenue, before moving to more customer segments.
- Since guest ratings are shown to increase purchase likelihood, **incorporate a 'top review' preview which not only provides guest ratings but highlights the top guest review** to **enforce purchase confidence even more**.
- Incentivize guest ratings, through loyalty points or credit, so as to have guest ratings available for all purchase options.
- Pause experiment and then reset traffic allocation to ensure results are valid before ensuring wholesale migration to guest reviews on booking form.

Recommendations for further analysis

Expedia

- Although we took a brief segment wise impact of conversion amongst the control and treatment groups, I'd also a **segmentation exercise** to understand customer perspective with **segment-wise A/B testing** to see how each segment responds.

- Carry out A/B tests on a new/return visitor, platform, geography, and marketing channel level to see if any one performs better than the other. In addition, test on browser level to see if it has an impact.

- Ensure that experiments are conducted at different points in the year **to understand seasonality** and its impact better. The same applies with marketing and promotional campaigns.

- While they are not direct conversions **analyzing micro conversion data** will certainly help persuade prospects down the conversion funnel.

- Do a deep dive on **customer forums** to understand sentiment around the new booking form.

Thank You

Appendix

Benefits and Tradeoffs of A/B testing

Expedia

Benefits of A/B testing

- Determine what works and what doesn't.

- Enables real feedback.

- Unbiased results.

Demerits of A/B testing

- Presenting different content/price/features to different customers especially in the same geolocation might result in negative sentiment or Change Aversion.

- Requires a significant amount of investment from Product, Engineering, and Data Science.

- Incorrect conclusions possible if not conducted properly.

Across segments as well, guest ratings had a positive effect on purchase likelihood

% Lift in Conversions by New/Returning User
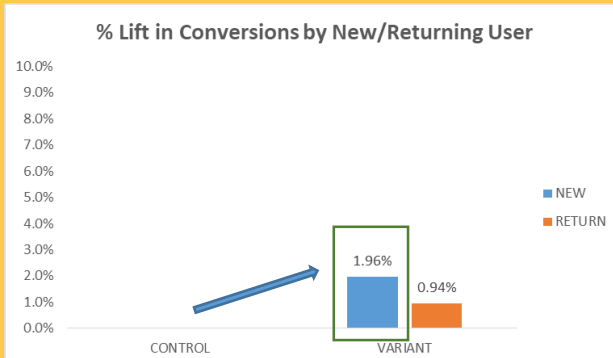
Purchase likelihood improved the most amongst new customers when shown guest ratings

% Lift in Conversions by Geography
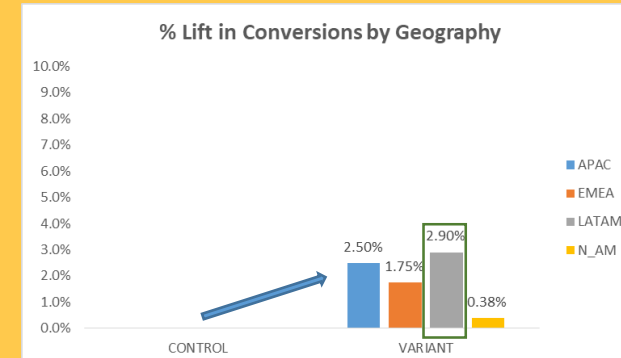
Purchase likelihood improved the most in Latin America customers when shown guest ratings

Note:
The y-axis in all charts captures the % difference in conversion between the control and variant groups. As shown, the variant group showed a positive lift in conversion across all segments

% Lift in Conversions by Platform
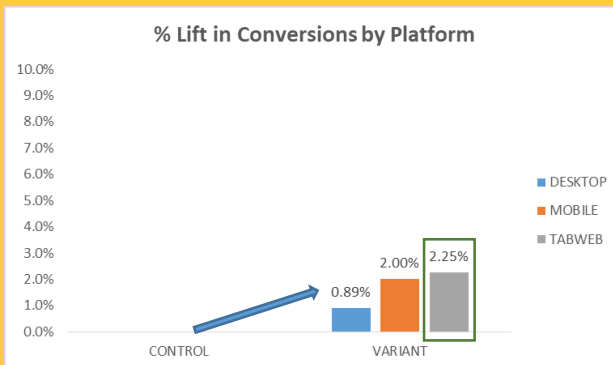
Purchase likelihood improved the most amongst Web customers when shown guest ratings

% Lift in Conversions by Marketing Channel
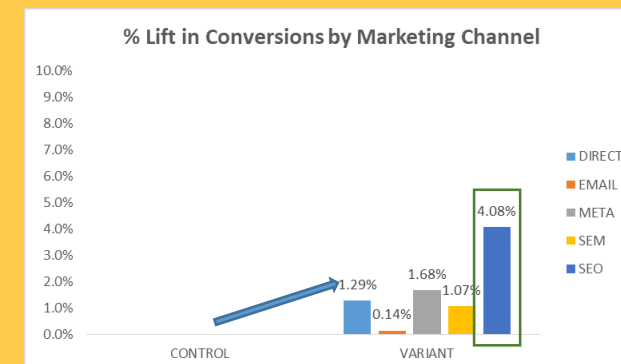
Purchase likelihood improved the most when the customer arrived at purchase through SEO when shown guest ratings

If we were to execute an A/B test once again from scratch we'd do the following in preparation…

- Determine baseline conversion rate, which **factors in the actual current state of affairs**.

- Determine A/B test duration. This is a critical parameter, as longer experiments require more resources, and insufficient durations lead to inaccurate results. **It is therefore very important to determine the optimal experiment length.**

- Determine power of the test, which is the **probability of making a correct decision** (to reject the null hypothesis) when the null hypothesis is false. This allows us to ascertain how accurate our results are.

- Establish significance level of the test which is the **probability of making a false discovery**.

- Determine **minimum desirable effect** which is the **minimum impact the business wants to see of the new version** to find this variant investment-worthy

- Determine sample size. How many users should we have in the control and treatment groups? **The sample size influences how accurate our results are and whether our conclusions from the test apply to the whole population.**

  The sample size we need is estimated through something called Power analysis, and it depends on a few factors:

  o Power of the test (1 — β) — This represents the probability of finding a statistical difference between the groups in our test when a difference is actually present. This is usually set at 0.8 by convention

  o Alpha value (α) — The critical value set earlier at 0.05.

  o Effect size — How big of a difference we expect there to be between the conversion rates

If we were to execute an A/B test once again from scratch we'd do the following in preparation...

Expedia

- Determine A/B test duration. This is a critical parameter, as longer experiments require more resources, and insufficient durations lead to inaccurate results. **It is therefore very important to determine the optimal experiment length.**

- Determine desired confidence level. By convention we assume a 95% confidence interval. This essentially means that whatever difference we observe, **we want to say with at least 95% statistical certainty that it is different from the control**.

- Not only are we looking for statistical significance, we're looking for **practical significance**. This will help us understand whether the detected difference in the performances of the two groups is large enough to justify the investment involved in changing the booking form.

- Analyze external and internal factors such as:
  - Seasonality or holiday period: Web traffic and sales are not stable throughout the year. For instance they might peak on Black Friday and Cyber Mondays. This could influence test results.
  - Marketing promotions and campaigns: Running a marketing campaign on the same site that an A/B test is running, may likely affect the test results.