Section 3 - Experimental Design

By- Sudarshan Bahurupe

Context:

The original version of an ecommerce website is quite basic, featuring only text on the site. They plan to add some product images, and intend to run an A/B test on their website to see if it helps sales. You are the analyst for this test.

a. What would your hypothesis be for running this test?

Adding product images to the ecommerce website will lead to an increase in sales. The hypothesis could be stated as ***"The inclusion of product images on the website will result in a higher conversion rate and ultimately lead to increased sales compared to the text-only version."***

b. What would be your primary metric to measure the success and why?

The primary metric to measure the success of the A/B test would be **the conversion rate**. Conversion rate is the percentage of website visitors who take the desired action, which in this case is making a purchase. By **comparing the conversion rates of the two versions** (text-only vs. text with product images), we can determine which version is more effective at driving sales.

The reason for choosing **conversion rate as the primary metric** is that it directly **reflects the impact on sales and revenue**, which is the ultimate goal of the ecommerce website. It provides a clear and actionable insight into whether the addition of product images leads to better performance in terms of driving user engagement and sales.

c. Tell us what secondary metrics you might look at to help you make a call on if a test is performing correctly and why?

<u>Average Order Value (AOV):</u> By analysing AOV for both versions of the website, we can determine if the inclusion of product images influences customers to purchase more or higher-priced items. A **higher AOV indicates increased customer spending**, which can be beneficial for the business.

<u>Bounce Rate:</u> Bounce rate measures the percentage of visitors who leave the website without interacting further after landing on the homepage. A high bounce rate may indicate that the **changes made in the A/B test are not resonating with visitors**, leading them to leave quickly.

<u>Time on Page:</u> Longer average time on page can indicate increased **user engagement** and interest in the products, suggesting that the inclusion of images is capturing user attention and encouraging them to explore the products further.

<u>Click-through Rate (CTR):</u> It measures the percentage of users who click on these elements to proceed complete the desired action. A higher CTR would indicate that the changes made in the A/B test are **encouraging users to take the desired actions** more frequently.

<u>Cart Abandonment Rate:</u> This metric measures the percentage of users who add products to their carts but do not complete the checkout process. If the inclusion of product images leads to a higher cart abandonment rate, it may suggest that there are **issues with the checkout process** or that the **product images** are **not convincing** enough for users to make a purchase.

d. You are asked to filter down the results further to see if the primary metric is performing well on different breakdowns. What potential problems could occur with your analysis if you layer different filters on top of the other?

When layering different filters on top of each other to analyse the A/B test results, several potential problems may arise:

<u>Sample Size Issues:</u> As you apply multiple filters, you might end up with smaller and smaller subgroups, reducing the sample size for each segment. Smaller sample sizes can lead to less reliable and statistically

significant results. It may be challenging to draw meaningful conclusions from small subgroups with limited data.
Simpson's Paradox: Layering filters can lead to the Simpson's Paradox phenomenon, where an observed trend in the aggregated data reverses or disappears when the data is divided into subgroups. This can cause misleading interpretations and incorrect conclusions if not considered carefully.
Interaction Effects: When analysing different breakdowns, interaction effects between filters may occur. These interaction effects can significantly influence the results and might lead to misinterpretations if not carefully considered.
Misleading Conclusions: Depending on the filters applied, some subgroups might have significant differences in the primary metric, while others might not show any significant impact. In such cases, it might be tempting to pick results that support a particular narrative, leading to biased or misleading conclusions.
Data Sparsity: Some subgroups resulting from the filters might have very few or no data points, leading to sparse data. Sparse data can be challenging to analyze and may not provide reliable insights.

e. Below are the results for an A/B test that you have analysed. Based off the results displayed below, what would be your recommendation to your product manager? Why did you make this recommendation?

Based on the results, I highly recommend implementing Variant 1, the A/B test group. Variant 1 exhibited a significantly superior conversion rate of 37.5% (from 3000/8000), outperforming the Control group's rate of approximately 34.7% (From 25000/72000). Despite receiving only 10% of the traffic compared to the Control group's 90%, Variant 1 still demonstrated better performance, indicating the positive impact of the new feature on user conversions.
Given the impressive performance of Variant 1 and the potential enhancements it offers to the conversion rate, it is beneficial to swiftly implement the new feature for the entire user base.

f. In what situations would you choose not to run an experiment, favouring rolling out a new feature immediately? When this occurs, how should you aim to handle this as an Analyst?

As an analyst, there are situations where it might be more appropriate to pause running an experiment and instead roll out a new feature immediately. Here are some scenarios where we can opt for immediate rollout:

High Business Impact: Rolling out new changes for potentially **high business impact** features with low risk is recommended, especially for time-sensitive features aligning with marketing plans or seasonal events.

Overwhelmingly Positive Results: If a small test shows a **big and meaningful improvement** without any problems, it might be okay to introduce the new feature to more people right away.

Clear User Benefit: If users find the new feature very **useful** and are **satisfied** with it, and their feedback is **positive**, introducing it right away can lead to **better user experiences** and higher **user engagement.**

Resource Constraints: When doing a full A/B test is difficult and takes a **lot of resources and time**, it's better to prioritize introducing the new feature based on strong evidence from smaller tests.

Competitor Response: To keep up with competitors in fast-changing markets, it's important to introduce similar features quickly in order to stay competitive.

To handle any of such situations we can opt for following steps:

Comprehensive Analysis: Even if a formal experiment is skipped, it's crucial to **carefully analyze** the small pilot or test, comparing the new **feature's performance** to the existing one, and considering important metrics.

Report and Insights: Prepare a **detailed report summarizing the results** of the limited testing, including any statistical analyses performed and key findings.

Stakeholder Communication: **Communicate results** of the analysis, the **recommendation** for immediate rollout, and the **reasoning behind this decision** with stakeholders, including product managers, engineers, and decision-makers.

Monitor Post-Rollout Performance: After the feature is rolled out, **continue monitoring its performance** closely to validate the positive results observed during the limited testing.