A/B Testing

Quan Zhang, Ph.D.

What is A/B Testing?

- If Amazon wants to change the UI (user interface) of their App or webpage, what do they expect? What if they want to change their recommending systems?

What is A/B Testing?

- If Amazon wants to change the UI (user interface) of their App or webpage, what do they expect? What if they want to change their recommending systems?
- A/B testing is a process of showing two variants of the same product (webpage, App, recommending systems…) to different segments of users at the same time and comparing which variant is better

What is A/B Testing?

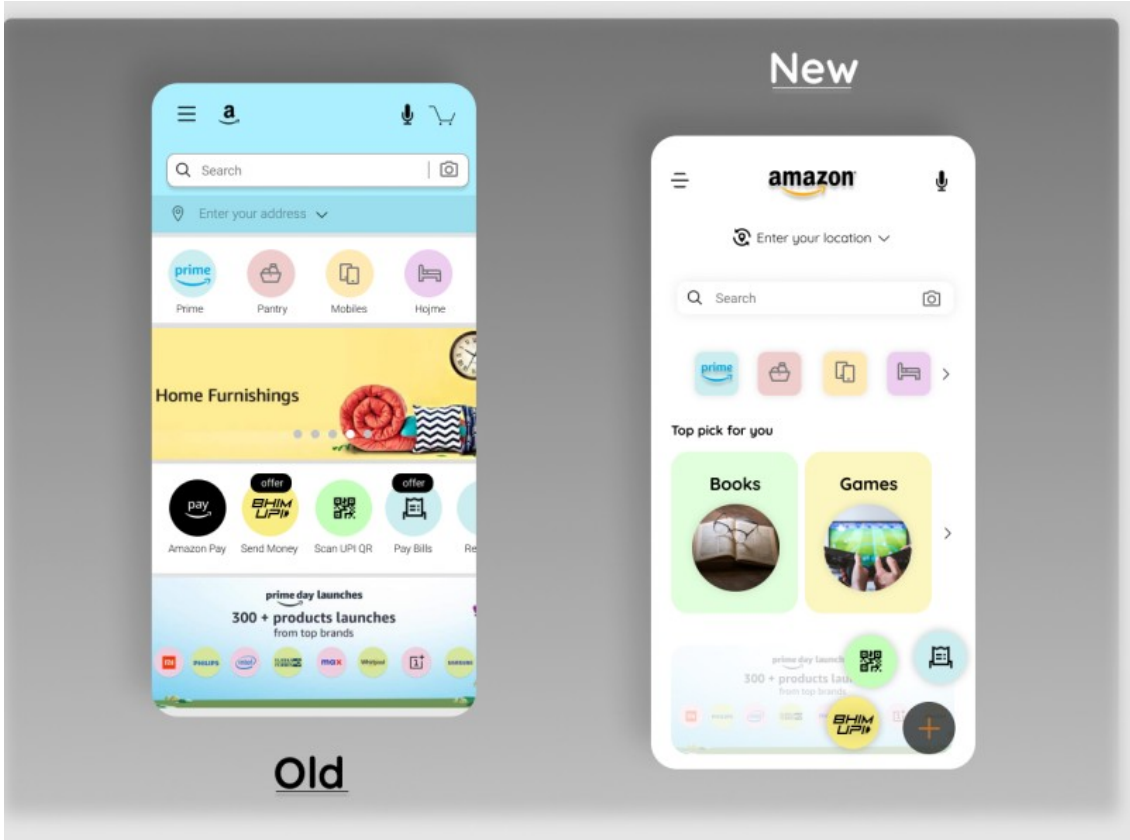
- If Amazon wants to change the UI (user interface) of their App or webpage, what do they expect? What if they want to change their recommending systems?
- A/B testing is a process of showing **two variants of the same product** (webpage, App, recommending systems…) to different segments of users at the same time and comparing which variant is better
- It enables you to determine which one of them performs better (such as generating better conversion rates). It is one of the easiest ways to analyze an application or a web page to create a new version that is more effective

Do you think these are good changes?

What Google see:

What I see:

Which one is better?



Metrics to use in A/B testing

- Typically in A/B testing, the variant that gives higher **conversions** is the winning one, and that variant can help you optimize your site for better results
- The metrics for conversion are unique to each website

Metrics to use in A/B testing

- Typically in A/B testing, the variant that gives higher **conversions** is the winning one, and that variant can help you optimize your site for better results
- The metrics for conversion are unique to each website
  - For eCommerce (B2C), it may be the sale of the products
  - For B2B, it may be the generation of qualified leads (A lead is a person who indicates interest in a company's product or service)
  - For media and publishing houses, it may be viewer engagement (time spent on the products)

Metrics to use in A/B testing

- Typically in A/B testing, the variant that gives higher **conversions** is the winning one, and that variant can help you optimize your site for better results
- The metrics for conversion are unique to each website
  - For eCommerce (B2C), it may be the sale of the products
  - For B2B, it may be the generation of qualified leads (A lead is a person who indicates interest in a company's product or service)
  - For media and publishing houses, it may be viewer engagement (time spent on the products)
- By A/B testing, you can gather quantitative user insights and use them to understand your potential customers and to optimize your metrics (conversion, leads, engagement...) based on data.

Why A/B Test?

- Struggling with a high cart abandonment rate
- Unhappy with all the unqualified leads they get per month
- Dealing with low viewer engagement

Why A/B Test?

- Struggling with a high cart abandonment rate
- Unhappy with all the unqualified leads they get per month
- Dealing with low viewer engagement
- These core conversion metrics are affected by some common problems like leaks in the conversion funnel, drop-offs on the payment page, etc

Why A/B Test?

- Struggling with a high cart abandonment rate
- Unhappy with all the unqualified leads they get per month
- Dealing with low viewer engagement
- These core conversion metrics are affected by some common problems like leaks in the conversion funnel, drop-offs on the payment page, etc
- A/B testing to deal with all these problems:

Solve visitor pain points

- Visitors have goals in their mind
  - To understand more about your product or service
  - To buy a product
  - To read/learn more about a particular topic
  - Simply to browse
- Common pain points while achieving their goal:
  - It can be hard to find a call-to-action button like buy now, request a demo, etc
  - Bad user experience will increase friction and eventually impacts conversion rates
- Get data through visitor behavior analysis tools
  - How long does it take a visitor to find the “Buy Now” button

Get better ROI from existing traffic

- The cost of acquiring any quality traffic can be huge
- You want to make the most out of your existing traffic and increase conversion without having to spend on acquiring new traffic.
- Sometimes, even the most minor changes can result in a significant increase in conversions.
- A/B testing can tell you if a change is profitable or harmful

Reduce bounce rate

- A ***bounce*** is a single-page session on a website
- A high bounce rate means many visitors browsed one webpage then left (that’s too bad!)
- Reasons for your website’s high bounce rate: too many options, expectations mismatch and so on.
- One way to reduce the bounce rate is through A/B testing.
- With A/B testing, you can test multiple variations of an element of your website till you find the best possible version.
- This improves your user experience and makes visitors spend more time on your site.

Make low-risk modifications

- Make minor, incremental changes to your web page instead of getting the entire page redesigned.
- This can reduce the risk of jeopardizing your current conversion rate.
- A/B testing lets you target your resources for maximum output with minimal modifications, resulting in increased ROI.
- Example:
  - Product descriptions changes: You can perform an A/B test when you plan to remove or update your product descriptions.
  - Introduction of a new feature change: Before introducing a new feature, launching that new feature as an A/B test.
- Testing and then making changes can make the outcome certain.

Achieve statistically significant improvements

- A/B testing is completely data-driven with no room for guesswork, gut feelings, or instincts
- You can easily determine a “winner” and a “loser” based on statistically significant improvements on metrics like time spent on the page, number of demo requests, cart abandonment rate, click-through rate and so on.

Profitably redesign your website

- The decision to implement one version or the other should always be based on data-driven A/B testing.
- Keep improving your website: As the new version goes live, test other elements of your webpage to make sure that the most engaging version is being served.

How to perform an A/B test

- A/B testing in marketing allows you to make the most out of your existing traffic.
- Should always be done through a well-defined **conversion rate optimization** (CRO) process.
- Includes the following steps:
  - Research
  - Formulate hypothesis
  - Create Variations
  - Run test and analyze the result

How to perform an A/B test

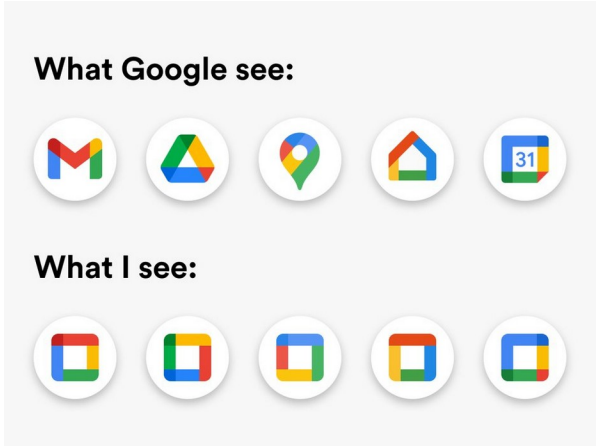
Step 1. Conduct research on how the website is currently performing

- You may want to find
  - how many users are coming onto the site
  - which pages drive the most traffic
  - what are the various conversion goals of different pages
- Then you know
  - your most visited pages
  - pages with most time spent
  - pages with the highest bounce rate.
- For example, you may want to start by shortlisting pages which have the highest revenue potential or the highest daily traffic. Following this, you dive deeper

How to perform an A/B test

Step 2. Formulate Hypothesis based on the research

- Creating data-backed hypotheses aimed at increasing conversions
  - Can be done by analyzing visitor behavior data qualitatively and quantitatively → EDA
- Example
  - A high bounce rate (visited a single page then left) → poor recommending systems
  - Abandoned shopping cart → difficulty in checking out
  - Sudden low usage of the product → change of icons



How to perform an A/B test

Step 3. Create variations based on your hypothesis

- A variation is another version of your current version with changes that you want to test
- You can test multiple variations against the existing version (control) to see which one works best
- Create a variation based on your hypothesis of what might work from a user-experience perspective
- For example, many people not filling forms
  - Does your form have too many fields? Try a variation with a shorter form
  - Does it ask for personal information? Omit fields that ask for personal information

How to perform an A/B test

Step 4. Run the test: **Randomly** assign users to different groups

- A/B testing: two variants
- Multivariate testing
  - Changes are made to multiple sections of a webpage, and variations are created for all the possible combinations
  - You can test all the combinations within a single test (more difficult than an A/B testing)
  - The multivariate test helps you figure out which element on a web page makes the most impact on its conversion rate
  - Example: You decide to test two versions of the **cover image** and the **Add To Cart** button color on a webpage. Then you test the following four versions

Image 1 + Red button (10% users)	Image 2 + Red button (9% users)	Image 1 + Blue button (11% users)	Image 2 + Blue button (8% users)

Step 4. Run the test

Two different statistical approaches to testing: Frequentist and Bayesian

- Example: Conversion rate comparison between App version 1 and version 2
  - N1 converted and M1 not converted for version 1
  - N2 converted and M2 not converted for version 2

Step 4. Run the test

Two different statistical approaches to testing: Frequentist and Bayesian

- Example: Conversion rate comparison between App version 1 and version 2
  - N1 converted and M1 not converted for version 1
  - N2 converted and M2 not converted for version 2
  - Frequentist approach: Hypothesis test (proportion test)
    - H0: Conversion rates are equal for both versions

Step 4. Run the test

Two different statistical approaches to testing: Frequentist and Bayesian

- Example: Conversion rate comparison between App version 1 and version 2
  - N1 converted and M1 not converted for version 1
  - N2 converted and M2 not converted for version 2
  - Frequentist approach: Hypothesis test (proportion test)
    - H0: Conversion rates are equal for both versions
  - Bayesian approach:
    - Priors for the conversion rates: *Beta(a1, b1)* for version 1, *Beta(a2,b2)* for version 2

Step 4. Run the test

Two different statistical approaches to testing: Frequentist and Bayesian

- Example: Conversion rate comparison between App version 1 and version 2
  - N1 converted and M1 not converted for version 1
  - N2 converted and M2 not converted for version 2
  - Frequentist approach: Hypothesis test (proportion test)
    - H0: Conversion rates are equal for both versions
  - Bayesian approach:
    - Priors for the conversion rates: *Beta(a1, b1)* for version 1, *Beta(a2,b2)* for version 2
    - Posteriors: *Beta(a1+N1, b1+M1)* for version 1, *Beta(a2+N2, b2+M2)* for version 2

    - What's the prob of version 1 giving a higher conversion rate?

Step 4. Run the test

Two different statistical approaches to testing: Frequentist and Bayesian

- Example: Conversion rate comparison between App version 1 and version 2
  - N1 converted and M1 not converted for version 1
  - N2 converted and M2 not converted for version 2
  - Frequentist approach: Hypothesis test (proportion test)
    - H0: Conversion rates are equal for both versions
  - Bayesian approach:
    - Priors for the conversion rates: *Beta(a1, b1)* for version 1, *Beta(a2,b2)* for version 2
    - Posteriors: *Beta(a1+N1, b1+M1)* for version 1, *Beta(a2+N2, b2+M2)* for version 2

    - What's the prob of version 1 giving a higher conversion rate?

      c1=rbeta(10000, a1+N1, b1+M1); c2=rbeta(10000, a2+N2, b2+M2);

Step 4. Run the test

Two different statistical approaches to testing: Frequentist and Bayesian

- Example: Conversion rate comparison between App version 1 and version 2
  - N1 converted and M1 not converted for version 1
  - N2 converted and M2 not converted for version 2
  - Frequentist approach: Hypothesis test (proportion test)
    - H0: Conversion rates are equal for both versions
  - Bayesian approach:
    - Priors for the conversion rates: *Beta(a1, b1)* for version 1, *Beta(a2,b2)* for version 2
    - Posteriors: *Beta(a1+N1, b1+M1)* for version 1, *Beta(a2+N2, b2+M2)* for version 2

    - What's the prob of version 1 giving a higher conversion rate?

      c1=rbeta(10000, a1+N1, b1+M1); c2=rbeta(10000, a2+N2, b2+M2);
      print(mean(c1>c2))

Step 4. Run the test: Frequentist v.s. Bayesian

- Frequentist approach
  - You only use data from your current experiment. The frequentist solution is to conduct tests and draw conclusions.
  - Requires the test to run for a set period of time to get correct data from it but can’t figure out how close or how far A and B actually are. It fails to tell you the probability of A beating B.
  - Often requires more observations and longer duration
- Bayesian approach
  - You can use prior knowledge from the previous experiments and try to incorporate that information into your current data; you use existing data to draw conclusions.
  - Takes into account the possibility of A beating B and also calculates how much improvement you can expect
  - Requires less data even if you don’t have previous experiments --- interim reports in the middle of the experiments; end the experiment early

Challenges of A/B Testing

1. **Deciding what to test**

- Not all small changes can significantly improve conversion
- The same goes for complex tests.
- EDA can help you overcome the challenge of “not knowing what to test”
- A possible starting point:

    On pages/features with the highest traffic, try to find the elements which may have the most impact on your conversion rates

Challenges of A/B Testing

2. Formulating hypotheses

- With the help of data gathered in the first step (i.e., research) of A/B testing, you need to discover where the problems lie with your product (site, App…) and come up with a hypothesis.
- Example problems leading to low conversion rates:
  - High bounce rate (users spending too much time on a single item)
  - Abandoned shopping cart
  - Spending too much time in payment sessions (promotion codes, credit card info, shipping address...)

Challenges of A/B Testing

3. **Appropriate sample size**

- Avoid the mistake of calling conclusive results too quickly
- As a statistician or data scientist/engineer, we need to learn about sample sizes, in particular, how large should our testing sample size be based on our web page’s traffic
- The larger sample size, the better?

Challenges of A/B Testing

4. Analyzing test results

- Successful campaigns: Interpreting test results after a statistical conclusion
  - Important to understand why the test succeeded
    - Why did users react a certain way with one version and not with the other versions?
    - What insights did you get?
    - Not only make sense of the current test, but also provide inputs for future tests
- Failed campaigns:
  - Failed test is useful if you draw learnings from them
  - The data gathered during the entire A/B testing process contains valuable information and insights
  - Find potential mistakes; use the data to provide prior belief for the next testing

Challenges of A/B Testing

5. **Changing Experiment Settings in the Middle of an A/B Test**

- Once the testing starts, everything had better to be fixed
- Try **not** to change your experiment settings, edit or omit your test goals, or play with the design of the control or the variation while the test is running.
- Try **not** to change the traffic allocations to variations Failed campaigns

Conclusion

- A/B testing is valuable when it is used to improving conversion rates.
- If done with complete dedication, and with the knowledge you now have, A/B testing can reduce a lot of risks involved when undertaking an optimization program.
- It will also help you significantly improve your products’ user experience by eliminating all distracting features and finding the most optimized version.
- Like EAD, A/B testing is Art more than Science.

Steps to perform an A/B test

- Research
- Formulate hypothesis
- Create Variations
- Run test and analyze the result

Which are the most important steps?

Steps to perform an A/B test

- **Research: Data engineering**
- **Formulate hypothesis: Data science + product management**
- Create Variations: software engineering
- Run test and analyze the result: Data science

Steps to perform an A/B test

- Research
- Formulate hypothesis
- Create Variations
- Run test and **analyze the result**

  This may be the only thing you can learn in class

In class practice

- An e-commerce website with localized versions in different countries
- A data scientist noticed that Spain-based users have a much higher conversion rate than any other Spanish-speaking country. She therefore talked to the international team to see why that was happening.
- The manager suggested that one reason could be translation. All Spanish-speaking countries had the same translation of the site. They agreed to try a test where each country would have its one translation written by a local. That is, Argentinian users would see a translation written by an Argentinian, Mexican users by a Mexican and so on. Nothing would change for users from Spain.
- After they run the test however, they are really surprised as the test is negative, i.e., it appears that the non-localized translation was doing better!

In class practice

You are asked to:

- Confirm that the test is actually negative. That is, it appears that the old version of the site with just one translation across Spain and LatAm performs better
- Explain why that might be happening. Are the localized translations really worse?
- If you identified what was wrong, design an algorithm that would decide if the same problem is happening in the future or everything is good and the results can be trusted

In class practice

Two data sets:

test_table.csv

Column Name	Value	Description
user_id	315281	this is id of the user
date	2015-12-03	he came to the site on Dec, 3 for the first time since the test started
source	Direct	his marketing channel was direct. No SEO or Ads.
device	Web	he visited the site using "web" (i.e. laptop/desktop, but not mobile)
browser_language	ES	his browser language settings are Spanish
ads_channel	NA	didn't come via an ad, so this has to be NA
browser	IE	he used Internet Explorer!
conversion	1	he converted
test	0	he was in control. That is, he saw the old translation written by a Spaniard

In class practice

Two data sets:

user_table.csv

Column Name	Value	Description
user_id	315281	same id as in the example above
sex	M	he is a Male
age	32	he is 32 y/o
country	Spain	he is based in Spain. So, in his case, he could have not been in the test no matter what.