Samatrix.io

Website A/B Testing for Conversion Optimization

Problem:

An e-commerce startup wants to know if a new checkout funnel (B) outperforms the old one (A).

Steps:

1. Simulate click-through data (visits vs. purchases) for A and B.
2. Calculate conversion rates and 95% confidence intervals.
3. Plot conversion rates with error bars.
4. Perform a two-proportion z-test to see if B beats A.
5. Real-Time Monitoring & Sequential Testing

Key Learnings:

1. **Streaming data handling** How to update metrics as new visitors arrive, rather than in one big batch.
2. **Sequential testing pitfalls** The p-value can fluctuate wildly early on. Peeking too often increases false positives — a real industry concern.
3. **Monitoring dashboards** The dual plots let students visualize both the **p-value trajectory** and the **observed lift**, mirroring real A/B dashboards.
4. **Practical inference** Learn to interpret "not yet significant" vs. "now significant" in a dynamic context and appreciate the need for **pre-defined stopping rules**.
5. **Streaming data handling** How to update metrics as new visitors arrive, rather than in one big batch.
6. **Sequential testing pitfalls** The p-value can fluctuate wildly early on. Peeking too often increases false positives — a real industry concern.
7. **Monitoring dashboards** The dual plots let students visualize both the p-value trajectory and the observed lift, mirroring real A/B dashboards.
8. **Practical inference** How to interpret "not yet significant" vs. "now significant" in a dynamic context and appreciate the need for pre-defined stopping rules.

Gain a more holistic understanding of how A/B tests are deployed and monitored in production environments.

What to Submit:

Submission Type: Individual

Each student must submit the following:

1. **Jupyter Notebook (.ipynb file) or python filr (.py file)**
    a. Filename: YourFullName_ABTest.ipynb (e.g., AnanyaKumar_ABTest.ipynb)
    b. Your notebook must follow the steps and structure discussed in class following the instructions in the submission guideline
2. **Word or PDF file**
    a. Answers "Questions for Report" in separate file

Required Sections and Instructions:

1. Setting Up Your A/B Testing Project in Python

Objective:
Get your Python environment ready for analyzing A/B test data. In this step, you will **load essential tools** that help with calculations, data handling, statistical testing, and visualizations.

Step 1: Understand the Goal
Before analyzing any A/B test results, you need tools that will:
- Organize and store your data (like a spreadsheet)
- Help you run math and statistics easily
- Let you create charts and graphs to visualize findings
- Perform the actual **A/B test comparison** to check if group differences are real

Step 2: Open Your Python Notebook
Use any one of the following:
- **Jupyter Notebook**
- **Google Colab** (runs in your browser)
- **VS Code with Python plugin**

Create a new notebook or script file to begin.

Step 3: Add Tools (Called Libraries)
These are special add-ons in Python that make your job easier.
1. **NumPy** – helps with numbers and percentages
2. **Pandas** – helps manage data like Excel
3. **SciPy** – helps you do statistics (like checking if group A is really better than B)
4. **Matplotlib** – helps you draw graphs and charts
5. **Statsmodels** – helps run special statistical tests like Z-tests for proportions

Step 4: What Each Tool Will Do

Tool	What You Will Use It For
NumPy	Calculate percentages, differences, and averages
Pandas	Read data from a file or table and work with groups A and B
SciPy	Do hypothesis testing (like calculating p-values)
Matplotlib	Draw bar charts to compare A and B visually
Statsmodels (Z-test)	Check if A and B conversion rates are **statistically different**

Step 5: What to Do Next
Once this setup is complete, you'll move on to:
- Loading the A/B test data
- Finding conversion rates for both groups
- Running the Z-test to compare them

- Drawing a graph to show your result
- Making a decision: “Did group B actually perform better?”

Learning Outcome from This Step
By completing this setup, students will:
- Be able to use Python for real-world statistical analysis
- Understand which tool does what in an A/B test
- Get ready to move on to hypothesis testing and result interpretation

Questions for Report

1. What is the purpose of each library (NumPy, Pandas, Matplotlib, SciPy, and Statsmodels) that you imported in this project?
Your answer should briefly describe how each tool helps in analyzing or visualizing data.

2. Why do we need to import tools before analyzing A/B test data in Python? Can we do it without them?
Think about whether these tools are optional or essential — and how they help simplify the work.

3. Which library will help you create visual charts in this project? Why are visualizations helpful when analyzing test results?
Hint: Think about how charts help us understand data better than just reading numbers.

4. What kind of real-world situations might benefit from using an A/B test and tools like the ones you've set up?
Try to give one or two simple examples — like a company testing a new website button.

5. Which tool or step in this setup did you find most interesting or new to you, and why?
This is your personal reflection — there’s no wrong answer!

2. Simulating A/B Test Results

Objective:
You will **simulate the outcome of an A/B test** experiment. In this simulation:
- Group A (the original version) has a 10% conversion rate
- Group B (the new version) has a 12% conversion rate
- You will “pretend” to run the test on 10,000 people in each group
- You'll use a random process to see how many people actually converted (took the desired action)

Step 1: Understand the Scenario
Imagine you are a product manager. You are testing two versions of a webpage:
- **Group A** sees the current version (this is the control group)
- **Group B** sees a new version with some improvements (this is the test group)

You want to know: **Is Version B actually better?**

Step 2: Decide How Many Users You Will Simulate
In a real experiment, thousands of users visit each version of the webpage.
In this step, we'll **pretend** that:

- 10,000 users saw **Version A**
- 10,000 users saw **Version B**

This is called your **sample size**.

Step 3: Set the True Conversion Rates
Now assume that:

- 10% of users in Group A will convert (success rate = 0.10)
- 12% of users in Group B will convert (success rate = 0.12)

These percentages are hidden in real life — you don't know them in advance. But for simulation purposes, we're defining them now.

Step 4: Simulate the Results
You will now **randomly generate** the number of users who convert in each group. Think of it like flipping a coin for each user:

- For Group A, each of the 10,000 users has a 10% chance of converting
- For Group B, each of the 10,000 users has a 12% chance of converting

At the end, you'll get:

- A number like **984** conversions for Group A
- And maybe **1,205** for Group B

Every time you run the simulation, the numbers may change slightly — just like in real life!

Step 5: Why This Step Is Important

What You're Doing	Why It Matters
Simulating user behavior	Helps you learn without real data
Creating fake but realistic numbers	Prepares you to test differences
Getting "conversion counts"	You'll use these numbers in the next step to run a statistical test

What to Do Next
In the next lab step, you'll:

- Compare how many people converted in A vs. B
- Use a test to check if the difference is **big enough** to be considered "real" (and not just luck)

Questions for Report

1. What do the numbers 10,000 and the conversion rates 10% and 12% represent in the context of this simulation?
Explain what these numbers mean and why they are used.

2. Why do we use random simulation to generate conversion outcomes instead of entering fixed numbers?

Think about how real-world user behavior involves chance and uncertainty.

3. In your simulation, how many users converted in Group A and how many in Group B? What does this suggest about their performance?
Write the actual numbers you got and compare them.

4. If you run the simulation again, will you get the exact same results? Why or why not?
Consider how randomness affects simulations and what the random seed does.

5. Do the results of your simulation suggest that Version B is better than Version A? What other information might you need to confirm this?
Think about the difference between observed results and statistical significance.

3. Calculate Conversion Rate & Confidence Interval for A/B Testing

Objective:
In this step, you will calculate:

- How many people converted in each group (A and B)
- What the **conversion rate** is for each group
- What the **confidence interval** is — a range that tells us how accurate the conversion rate estimate is

This will help you understand:
Is the new version (B) really performing better than the old one (A)?

Step 1: Understand What You're Comparing
You ran a test on two groups:

Variant	Total Visitors	Converted Users
A	10,000	Example: 973
B	10,000	Example: 1,134

You want to know:

- What percent of users converted in each group?
- Is that number reliable, or could it just be from chance?

Step 2: Calculate the Conversion Rate (CR)
To get the conversion rate:

- Divide the number of converted users by the total number of visitors.
- Then multiply by 100 to get a percentage.

Example:

- Variant A: 973 / 10,000 = 9.73%
- Variant B: 1,134 / 10,000 = 11.34%

This shows you how successful each group was.

Step 3: Calculate the 95% Confidence Interval
The **confidence interval** tells you:
"How sure are we that this result reflects the truth, and not just random luck?"

You will calculate a **range** around the conversion rate where the true conversion rate likely falls.
Example:

- Group A might be 9.15% to 10.31%
- Group B might be 10.72% to 11.96%

If these ranges don't overlap, it's a good sign that one version is truly better than the other.

Step 4: Create a Table with All Results
Once you calculate everything, you will organize your results like this:

Variant	Visitors	Conversions	Conversion Rate	Confidence Interval (95%)
A	10,000	973	9.73%	9.15% – 10.31%
B	10,000	1,134	11.34%	10.72% – 11.96%

This table will help you clearly compare both groups.

Step 5: Interpret the Results
Ask yourself:

- Which group has a higher conversion rate?
- Are the results different enough to say it's not random?
- Do the confidence intervals overlap?

If **Group B has a higher rate** and the intervals **don't overlap**, you can conclude:
"Group B performed significantly better, and we can be confident about it."

What You Learn from This Step

Concept	What You're Practicing
Conversion rate	Understanding user behavior data
Confidence interval	Measuring accuracy of results
Data comparison	Making decisions based on numbers

Questions for Report

1. What was the conversion rate (CR) for Variant A and Variant B in your experiment? What does this number represent?
Tip: Look at how many people converted out of the total and what that tells you.

2. What do the confidence intervals for each variant tell you about the reliability of the conversion rate?
Explain what the range (from lower bound to upper bound) means and how it helps you make decisions.

3. Do the confidence intervals for Variant A and B overlap? If not, what does that suggest about the performance of the two variants?
Think about whether one group clearly outperformed the other or not.

4. Which variant would you recommend based on the results, and why?
Write a short decision statement like a product manager: which version is better and how confident are you?

5. If you were presenting this to your team, how would you explain the meaning of "confidence interval" in simple terms?
Try to describe it in your own words, like explaining to a friend.

4. Visualizing A/B Test Results with Confidence Intervals

Objective:
In this step, you will create a **bar chart** that compares the **conversion rates** of Variant A and Variant B. You will also show the **confidence intervals** as error bars to help you understand how reliable the results are.
This chart makes it easier to answer the key question:
"Is Variant B really better than A — or is it just random chance?"

Step 1: Know What You're Comparing
Before drawing the chart, make sure you have calculated the following for both A and B:

Variant	Visitors	Conversions	Conversion Rate	Confidence Interval
A	10,000	973	9.73%	9.15% to 10.31%
B	10,000	1,134	11.34%	10.72% to 11.96%

You'll use this data to build your chart.

Step 2: Set Up the Chart
You will now draw a **bar for each group**:
- One bar for **Variant A**
- One bar for **Variant B**

Each bar's **height** represents its conversion rate.

Step 3: Add Error Bars for Confidence
Above each bar, draw a **short vertical line** that shows the **range of values** from the confidence interval. This tells the viewer:
"We are 95% confident the real conversion rate falls in this range."
Example:
- For A, show from 9.15% to 10.31%
- For B, show from 10.72% to 11.96%

This is called a **confidence interval** — and it helps you understand how accurate the conversion rate is.

Step 4: Label Your Chart
Make sure your chart has:
- A **title** (e.g., "A/B Test: Conversion Rate with 95% Confidence Interval")
- The **Y-axis label** (e.g., "Conversion Rate")
- Labels for each bar: "A" and "B"
- Use different **colors** for the bars to clearly separate them (like blue and pink)

Step 5: Interpret What You See
Once your chart is ready, answer:
- Is the bar for B **taller** than A? (Higher conversion rate?)
- Do the **error bars overlap** or are they separate?

- Is there **enough difference** to believe B is truly better?

What Students Learn from This Step

Skill Learned	Why It Matters
Visualizing data	Helps communicate results clearly
Reading error bars	Understands statistical confidence
Comparing outcomes	Learns to make data-based decisions

Questions for Report

1. Which variant had the higher conversion rate in your chart? How can you tell just by looking at the bars?
Hint: Look at the height of each bar. Which bar was taller?

2. What do the small vertical lines (error bars) above the bars represent? Why are they important?
Explain in your own words what confidence intervals show and why we use them.

3. Do the confidence intervals for Variant A and Variant B overlap in your chart? If not, what does this tell you about the results?
Try to describe whether the result is clear and reliable, or if it might just be due to chance.

4. If this test was for a real website or app, which version would you recommend using and why?
Support your decision using the chart and what you learned from the data.

5. How does seeing the results in a chart help you understand the data better than just looking at numbers in a table?
Reflect on how visuals help communicate insights.

5. Testing If Variant B is Statistically Better (Z-Test for Proportions)

Objective:
In this step, you will use a **Z-test** to check whether the difference in conversion rates between Variant A and Variant B is **real** or just due to **random chance**. This helps you answer:
"Is Variant B actually better, or did we just get lucky?"

Step 1: Get Your Two Key Inputs
Make sure you already know:

- How many users **converted** in each group
- How many **total visitors** were in each group

You should have something like this:

Variant	Total Visitors	Conversions
A	10,000	973

B	10,000	1,134

Step 2: Understand the Goal of the Z-Test
You're going to run a statistical test to **compare two percentages**:

- Group A: 973 / 10,000 = 9.73%
- Group B: 1,134 / 10,000 = 11.34%

The Z-test checks:
Is that 1.61% difference big enough to be trusted? Or could it be random?

Step 3: Run the Z-Test for Proportions
The test gives you two important values:

Term	Meaning
Z-statistic	How far apart A and B are, measured statistically
P-value	Probability that the difference is due to chance

Step 4: Interpret the Results
Now use these **rules**:
If p-value < 0.05:

- The result is **statistically significant**
- You can say with confidence: **Variant B is better**
- You **reject the null hypothesis** (H_0)

If p-value ≥ 0.05:

- The difference might just be random
- You **fail to reject the null hypothesis**
- You **don't have enough proof** to say B is better

Step 5: Record Your Decision
Based on your p-value, write a conclusion. For example:
"Because our p-value was 0.000, which is less than 0.05, we reject the null hypothesis. We are confident that Variant B performs significantly better than Variant A."

What Students Learn from This Step

Concept	What It Teaches
Hypothesis testing	How to make decisions based on data
Statistical thinking	Not all differences are meaningful
Real-world reasoning	When to trust that one version is better

Questions for Report

1. What is the purpose of running a Z-test in an A/B test project?
Explain why we don't just compare percentages, and why we need a statistical test.

2. What were your Z-statistic and p-value? What do these values tell you?
Write the exact numbers from your output and briefly say what they mean.

3. Did you reject or fail to reject the null hypothesis (H_0)? What does that decision mean in simple words?

State your decision clearly and explain whether it means B is better or not.

4. If Variant B had a slightly higher conversion rate than A, but your p-value was above 0.05, what would you conclude?
Think about whether a small improvement is always enough to make a change.

5. How would you explain the meaning of the p-value to a non-technical friend?
Use your own words — imagine explaining it over a coffee.

6. Preparing to Simulate A/B Test Results Over Time

Objective:
In this step, you will **prepare to simulate live A/B testing data** — as if users are arriving at your website every minute. You'll set up the test environment and get ready to track key metrics like:

- How many people visit each version
- How many of them convert
- How your confidence in the result changes over time

Step 1: Understand What You're Simulating
In real-world A/B tests:

- Users don't come all at once — they arrive over time.
- We want to observe: "How does the performance of Variant A vs. Variant B look as data flows in batch by batch?"

This setup mimics that by adding users in **small time steps**, such as 100 visitors at a time.

Step 2: Define the True Performance of Each Variant
Imagine we know this in advance (just for simulation purposes):

- **Variant A** has a true conversion rate of **10%**
- **Variant B** has a better conversion rate of **12%**

These percentages are used to randomly simulate how users behave on each version.

Step 3: Decide the Batch Size
A **batch** is a group of visitors added at one time. For example:

- Every minute, 100 new users arrive and are split between Group A and Group B.

This allows us to **track how results evolve over time**, just like in real life.

Step 4: Set Up Counters to Track Progress
You'll keep track of:

- How many users have visited each version so far
- How many have converted
- What the conversion rate is
- How much better one version is over the other
- How confident we are in that difference

These values will be updated automatically as each new batch arrives.

Step 5: Prepare to Store and Visualize Results
Before we start adding users, we also prepare **empty containers** (called lists) that will later store:

- **Batch number** (time step)
- **Conversion lift** — how much better B is than A
- **p-value** — a measure of how confident we are in the result

This helps you later **draw a graph** that shows how your decision becomes stronger over time.

What Students Learn from This Step

Concept	What It Builds
Live testing mindset	Data arrives over time, not all at once
Real-world thinking	Confidence takes time to build with data
Statistical awareness	Introduces the idea of tracking conversion lift and p-value dynamically

Questions for Report

1. What is the purpose of running an A/B test in batches instead of using all the data at once?
Think about how websites work in real life — do all users come at the same time or gradually?

2. What do the values 0.10 and 0.12 represent in your simulation setup? Why are they important?
These values describe the behavior of your variants — explain what they mean.

3. How does tracking visitors and conversions in small batches help us better understand A/B test results?
Consider how this can show trends or build confidence over time.

4. Why is it useful to store data like conversion lift and p-values during each time step of the simulation?
Think about what we learn from seeing those values change instead of just one final result.

5. If your simulation showed that results fluctuated a lot in early batches, what would that tell you about making early decisions in A/B testing?
Reflect on the importance of waiting for enough data before making decisions.

7. Simulating a Live A/B Test and Watching Results Over Time

Goal of This Step
You will now simulate a **live A/B test** — watching what happens **as visitors arrive batch-by-batch**. For each batch, you'll update:

- How many people visited
- How many converted

- The difference between A and B (called **lift**)
- Whether the difference is statistically meaningful (**p-value**)

You'll also **see graphs update live** as each new batch of visitors arrives.

Step 1: Simulate Visitors for One Time Step

- A batch of visitors arrives (e.g., 100 people for A and 100 for B).
- Each visitor either **converts (buys/signs up)** or doesn't, based on the group's conversion rate.
- This step **simulates that behavior randomly**, as it would happen in real life.

Step 2: Update Totals for Each Variant

- Keep track of how many visitors have seen A and B so far.
- Count how many total conversions have happened in each group.
- **Why?** This lets you keep score over time — like tracking goals in a football match.

Step 3: Calculate Conversion Rates and Lift

- Calculate the **conversion rate** for A and B (conversions ÷ total visitors).
- Find the **lift**: how much better B is doing than A.
- Example: If A is at 10% and B is at 12%, the lift is 2%.

Step 4: Run a Z-Test to Check for Significance

- This step answers: *"Is the difference real, or just luck?"*
- A **Z-test** checks if B is doing significantly better than A.
- It returns a **p-value**:
    - If the p-value is **less than 0.05**, the result is considered **statistically significant**.
    - If it's higher, we can't be sure yet.

Step 5: Save Results for Plotting

- Each batch's **p-value and lift** is saved in a list.
- This lets us draw charts that **update in real time**, showing how confidence grows.

Step 6: Display Live Output

You'll see printed updates like:
yaml
CopyEdit
Batch 12/60
Variant A: 1200 visits, 120 buys → CR = 10.0%
Variant B: 1200 visits, 144 buys → CR = 12.0%
Observed lift: 2.0%
p-value = 0.0341
→ Significant lift detected (p<0.05)
This helps you practice **reading A/B test results** like a real data scientist.

Step 7: View Real-Time Charts

Two charts update as each batch arrives:

1. **p-value chart**: shows your statistical confidence growing over time

2. **Lift chart**: shows how much better B is doing compared to A

You'll see if:

- The p-value **drops below 0.05** (a good sign!)
- The lift stays **above 0**, meaning B is better than A

Step 8: Wait and Repeat

The loop **pauses briefly** before moving to the next batch — just like real traffic flowing into a site.

What You Learn in This Step

Concept	Why It Matters
A/B test simulation	Mimics how real-time experiments work
Conversion lift	Measures the advantage of one variant
p-value tracking	Shows when results become statistically reliable
Live data interpretation	Builds real-world decision-making skills

Questions for Report

1. How did the conversion rates of Variant A and Variant B change over time as more data came in?

Describe any patterns you noticed in the conversion rates — were they steady, fluctuating, or improving?

2. What does the "lift" represent in this simulation? Did you observe the lift increasing, decreasing, or remaining stable over time?

Explain what lift tells you and how it behaved across the 60 batches.

3. At which batch (or around which batch number) did the p-value first drop below 0.05? What does that mean in practical terms?

Think of the moment when you had enough evidence that Variant B was better.

4. Why is it important to monitor A/B test results over time instead of deciding early based on just a few visitors?

Reflect on how randomness or early variation can lead to incorrect decisions if you act too soon.

5. If you were running this test in the real world, how would you decide when to stop the test and choose a winner? What would you look for?

Use the concepts of p-value, lift, and conversion stability in your explanation.