



## A/B TESTING IN CRM:

A TECHNICAL METHODOLOGY TO  
ENHANCE CUSTOMER EXPERIENCE



---

How data, statistics, and automation  
can transform CRM

LEARN MORE

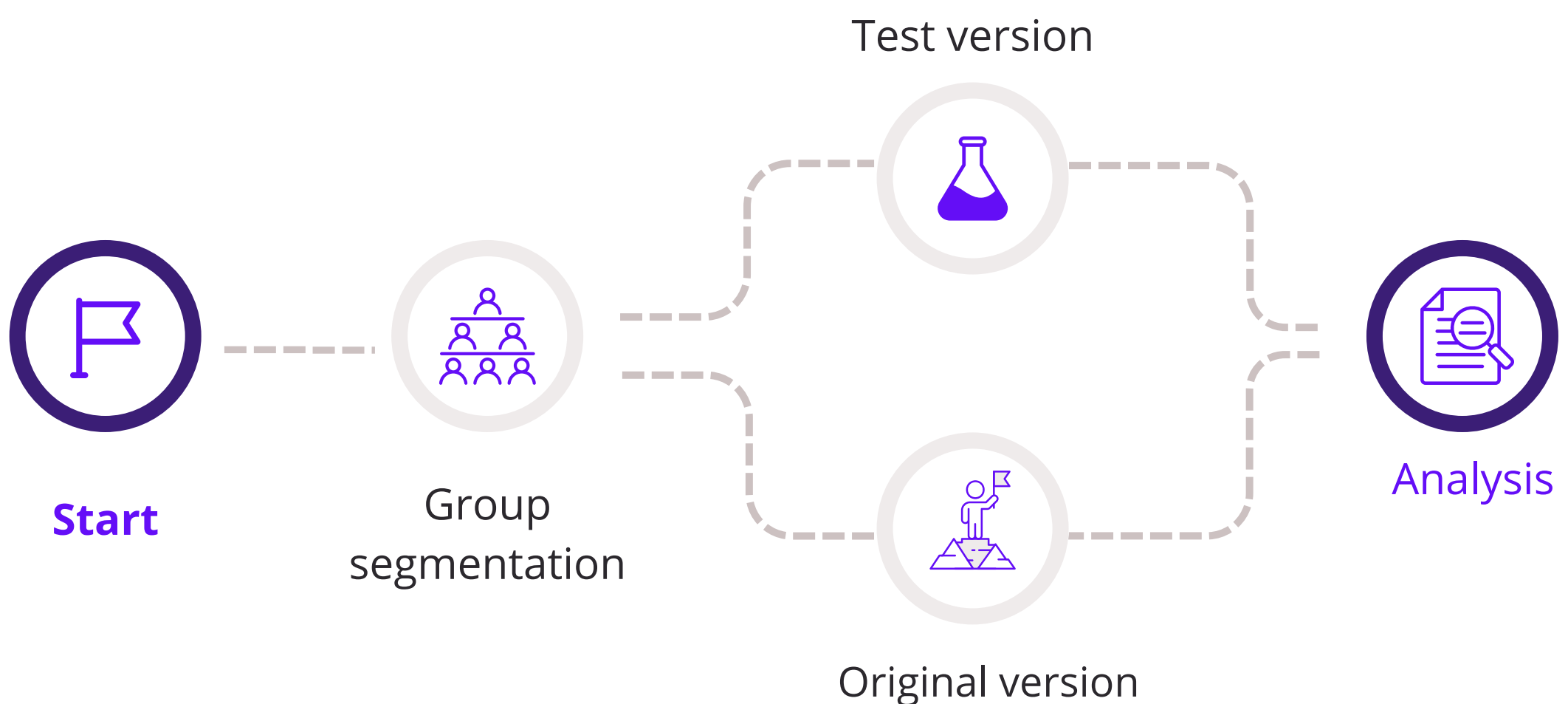


## IN TODAY'S CUSTOMER EXPERIENCE (CX) LANDSCAPE, INTUITION-BASED DECISIONS NO LONGER SUSTAIN CONSISTENT GROWTH.

To achieve scalable and replicable results, it is essential to adopt robust experimental methodologies—and A/B testing stands out as a cornerstone of data-driven decision-making in the context of CRM (Customer Relationship Management).

This article details the technical workings of A/B testing, from experimental design to data unification and statistical analysis, with a focus on reliability and result accuracy.

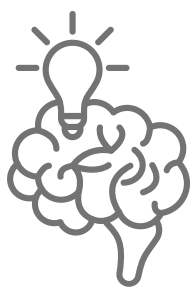
### What is A/B Testing?



A/B testing is a statistical methodology for controlled experimentation that compares the performance of two or more variations of an element under test, while keeping all other variables constant, in order to isolate the effect of the variable of interest.

In CRM, this may involve testing different email subjects, delivery channels, send times, messages, formats, or CTAs, always aiming to understand the direct impact on a business metric (such as open rates, conversion, retention, etc.).

## TECHNICAL FOUNDATIONS OF THE METHODOLOGY



### Hypothesis Definition

Establish the foundation for statistical comparison.



### Selection of the Primary Metric (KPI)

Measure what truly matters to the business.



### Sample Size and Statistical Power

Ensure reliability through mathematical rigor.

## 1. Hypothesis Definition

A/B testing begins with two hypotheses: the null ( $H_0$ ) – which asserts that there is no difference between the versions being tested – and the alternative ( $H_1$ ), which suggests a change (e.g., version B performs better than version A).

Example:

- $H_0$  (null hypothesis): There is no difference in proportion rate between tests A and B.
- $H_1$  (alternative hypothesis): Test B shows a higher proportion rate than test A.

## 2. Selection of the Primary Metric (KPI)

Choosing a single KPI as the primary metric in a test is key to ensuring clarity of objectives and reliable results. This maintains focus on business relevance and avoids confusing or conflicting interpretations of data. Common metrics include:

- **Open Rate:** Measures the proportion of recipients who opened the message (e.g., email), indicating initial interest.
- **CTR (Click-Through Rate):** Percentage of individuals who clicked on a link within the message, reflecting engagement.
- **Conversion Rate:** Assesses how many users completed the desired action (e.g., purchase, form fill). Essential to understand business impact.
- **Average Response Time:** Indicates the average time until the first user interaction.
- **Retention / Churn:** Respectively measure customer retention and loss over time—critical to understanding satisfaction and loyalty.

## 3. Sample Size and Statistical Power

Sample size should be calculated based on:

- **Significance level ( $\alpha$ ):** Determines tolerance for Type I error (false positive). Typically 5%, meaning up to a 5% chance of detecting a difference when none exists.
- **Statistical power ( $1 - \beta$ ):** Measures the ability to detect a true effect when one exists. Recommended at 80%+ to reduce risk of Type II error (false negative).
- **Expected effect size ( $\Delta$ ):** The minimum meaningful difference you wish to detect. Smaller effects require larger samples.
- **Historical data variability:** Refers to natural fluctuations in past data. High variability demands larger samples for reliable results.

To ensure trustworthy results, the sample size should be defined using appropriate statistical techniques and tools such as Optimizely or Evan Miller calculators.

---

# Sampling Strategies in CRM

How test groups are formed can affect result validity. Thus, defining the sampling process is key. Common approaches:

## A

### Pure Randomization

Customers are randomly distributed between control and test groups. This is simple and useful for large-scale campaigns, but may overlook behavioral nuances.

## B

### Stratified Randomization

Data is pre-segmented by key criteria (e.g., customer type, funnel stage, average ticket), followed by randomization within strata. Ensures balanced groups—ideal for CRM testing due to increased result comparability.

## C

### Clustering + Randomization

For large, heterogeneous datasets, clustering (e.g., K-means or DBSCAN) can group customers by behavior. Then randomization is applied within clusters for better control and balanced samples.

## CRM Sampling Strategy Summary Table:

Strategy	Description	When to Use	Advantages
<b>Pure Randomization</b>	Fully random group assignment	Homogeneous, large audiences	<b>Simple, reduces initial bias</b>
<b>Stratified Randomization</b>	Pre-segmentation with internal draw	Diverse profiles, need for balance	<b>Ensures comparability</b>
<b>Clustering + Randomization</b>	Grouping by behavior before randomization	Large, varied bases	<b>Controls variability, improves precision</b>

## 1. Group Definition

A structured A/B test begins with clearly defined groups to ensure statistical validity:

- **Control Group (CG):** Receives the standard experience or no communication. Serves as a baseline.
- **Group A:** Receives version 1, typically the current campaign.
- **Group B:** Receives version 2, the proposed new test.

For multivariate tests, additional groups (C, D, etc.) may be used—but **this requires a larger sample, as statistical power is diluted.**

## 2. Test Duration

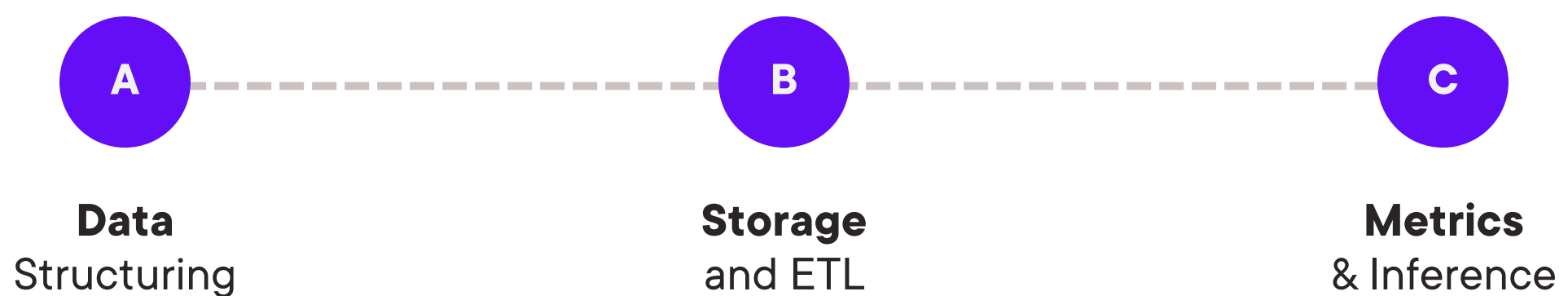
Test length must be sufficient to reach sample size goals without seasonal or behavioral distortions.

In CRM, duration varies:

- **Push Notifications:** Typically a few hours.
- **Retention or long-term behavior tests:** May require days or weeks.
-

**Important:** Do not stop a test early—even with promising results—as this compromises statistical significance and biases conclusions.

## DATA UNIFICATION AND MEASUREMENT



Reliable A/B test results depend on structured data collection and organization. Key data considerations:

### a) Data Structuring

Use unique customer IDs to track full journeys and avoid duplication.

- Capture all touchpoints: opens, clicks, replies, conversions.
- Standardize contextual variables like channel, time, device, and location to avoid biased comparisons.

### b) Storage and ETL

Store A/B data in robust platforms like data warehouses or lakes.

Use ETL/ELT pipelines for standardization and transformation.

#### Recommended tools:

- Storage: BigQuery, Snowflake, Redshift
- ETL/Orchestration: dbt, Apache Airflow

## c) Metrics & Inference

Use appropriate statistical methods based on data type:

- Z-Test / T-Test: Compare averages (e.g., average response time, purchase value).
- Chi-square or Proportion Test: For rates or proportions (clicks, conversions).
- Mann-Whitney U Test: For distribution comparisons when data isn't normally distributed.

Always calculate confidence intervals (usually 95%) to ensure robustness and reduce random error.

## HOW TO AVOID COMMON BIASES AND PITFALLS

Misinterpreting A/B test results can lead to poor strategic decisions, impacting campaigns, products, and investments. Avoid these risks:

- **Multiple variation testing:**  
Test only one variable at a time to isolate impact.
- **External interference:**  
Control test timing to avoid promotions, crises, or seasonal effects.
- **Group overlap:**  
Ensure a user is not in multiple concurrent tests with the same goal.
- **Cherry-picking:**  
Don't select only favorable data. Predefine analysis to ensure objectivity.

## MEASURING CUSTOMER EXPERIENCE TO GUIDE BUSINESS STRATEGIES

Adopting good statistical and methodological practices makes A/B testing a reliable, evidence-based tool for strategic decision-making.

Annex I contains a practical validation checklist to use before publishing results or acting on them.

---

## Dhauz Recommendations: Best Practices

For scalable and reliable A/B testing, structure the process around three pillars: automation, data centralization, and governance.



### AUTOMATE SEGMENTATION AND GROUP ALLOCATION

Use CRM and campaign orchestration tools for consistent and auditable segmentation.

**Examples:** Salesforce, RD Station, HubSpot, Oracle Eloqua.

## INTEGRATE DATA INTO A CENTRALIZED ENVIRONMENT

Unify user interaction data in a single data warehouse or data lake to improve analysis integrity.

**Create experiment governance:** Define rules for simultaneous tests, assign responsibilities, use standardized checklists, and maintain a historical test repository.

**Conclusion:** Data, Statistics, and Automation in Favor of CX.

Applying A/B testing in CRM with a technical approach fosters a culture of structured, statistically valid, and customer-centered experimentation.

**WITH THIS METHODOLOGY, YOUR COMPANY CAN MAKE FASTER, MORE ACCURATE, AND PERSONALIZED DECISIONS—ENHANCING CAMPAIGN PERFORMANCE AND ANALYTICAL MATURITY.**

## Annex 1

### Planning & Execution

- ✓ Planning & Execution
- ✓ Data Collection & Quality
- ✓ Statistical Analysis
- ✓ Anomalies & Bias Verification
- ✓ Interpretation & Action

To ensure the scalability and reliability of A/B testing, the process must be structured. There are three main pillars: automation, data centralization, and governance. Below are some recommendations to help advance the maturity of the operation:

#### 1. Planning & Execution

- Is the test objective clearly defined? (e.g., increase conversions, reduce churn)
- Was only one element tested at a time?
- Were groups randomly distributed with no overlap?
- Was the sample size statistically sufficient?
- Was the test duration fully respected?

#### 2. Data Collection & Quality

- Is the data centralized in a reliable repository?
- Were all customer journey touchpoints captured?
- Were contextual variables (e.g., time, channel, device) normalized or controlled?

### 3. Statistical Analysis

- Was the correct test applied (T/Z, chi-square, Mann-Whitney)?
- Are results statistically significant ( $p \leq 0.05$ )?
- Was the confidence interval calculated (typically 95%)?
- Was the analysis free of cherry-picking or post hoc filtering?

### 4. Anomalies & Bias Verification

- Were external interferences (e.g., promotions) recorded?
- Were group behaviors comparable before the test?
- Are the tested hypotheses and evaluation criteria clearly documented?

### 5. Interpretation & Action

- Are the results statistically and business-relevant?
- Is it clear which variation performed best and why?
- Has the responsible team validated findings before large-scale rollout?



**WANT TO STRUCTURE A CONTINUOUS,  
DATA-DRIVEN EXPERIMENTATION  
OPERATION?**

Dhauz can support you—from  
test modeling to analysis  
automation.

**Let's talk!**