



## TESTE A/B NO CRM:

METODOLOGIA TÉCNICA PARA  
IMPULSIONAR A EXPERIÊNCIA DO CLIENTE



---

Como dados, estatística e automação  
**podem transformar o CRM**

SABER MAIS

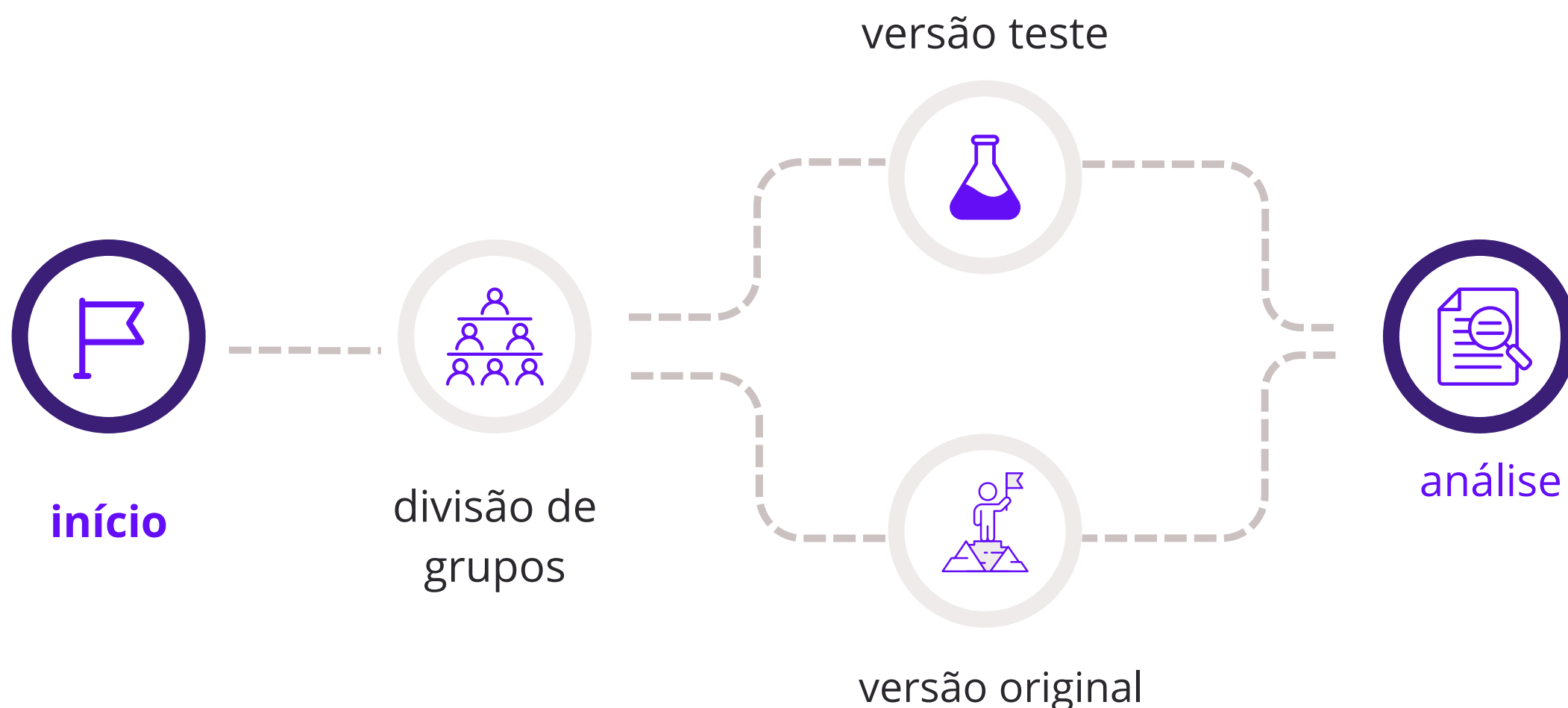


## NO CENÁRIO ATUAL DE EXPERIÊNCIA DO CLIENTE (CX), DECISÕES BASEADAS EM INTUIÇÃO JÁ NÃO SUSTENTAM CRESCIMENTO CONSISTENTE.

Para gerar resultados escaláveis e replicáveis, é necessário adotar metodologias experimentais robustas e, o teste A/B se destaca como um dos pilares da tomada de decisão baseada em dados no contexto de CRM (Customer Relationship Management).

Neste artigo, detalhamos o funcionamento técnico do teste A/B, desde o desenho experimental até a unificação dos dados e análise estatística, com foco em confiabilidade e precisão dos resultados.

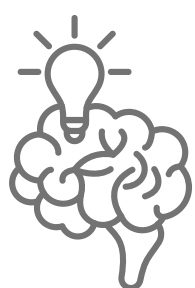
### O que é teste A/B?



O teste A/B é uma metodologia estatística de experimentação controlada, que compara o desempenho de duas ou mais variações de um elemento sob teste, mantendo todas as demais condições constantes, para isolar o efeito da variável de interesse.

No contexto de CRM, isso pode significar **testar diferentes assuntos** de e-mail, canais de disparo, tempos de envio, mensagens, formatos ou CTAs, sempre com o objetivo de entender o impacto direto sobre uma métrica de negócio (abertura, conversão, retenção, etc.).

## FUNDAMENTOS TÉCNICOS DA METODOLOGIA



### Definição de Hipóteses

Estabeleça a base da comparação estatística.



### Escolha da Métrica (KPI Primário)

Meça o que realmente importa para o negócio.



### Tamanho da Amostra e Poder Estatístico

Garanta confiabilidade com base matemática.

## 1. Definição de hipóteses

A experimentação em testes A/B, inicia-se com duas hipóteses: a nula ( $H_0$ ) – postula que não existe diferença entre as versões testadas, e a alternativa ( $H_1$ ) – que sugere uma mudança. (a versão B é superior à versão A)

### Por exemplo:

- $H_0$  (hipótese nula): não há diferença na taxa de proporção entre os testes A e B.
- $H_1$  (hipótese alternativa): o teste B apresenta uma taxa de proporção superior ao teste A.

## 2. Escolha da métrica (KPI primário)

Escolher um único KPI como métrica principal em um teste é fundamental para garantir clareza nos objetivos e resultados confiáveis. Assim, mantém-se o foco no que é relevante para o negócio e evita interpretações confusas ou conflitantes dos dados, assegurando que as decisões confiáveis. **Dentre as métricas mais utilizadas temos:**

- Taxa de abertura (Open Rate): computa a proporção de destinatários que abriam a mensagem (ex: e-mail), servindo como um indicador do interesse inicial na comunicação.
- CTR (Click-Through Rate): representa a porcentagem de indivíduos que clicaram em um link contido na mensagem (e-mail/anúncio), evidenciando o interesse despertado pelo conteúdo.
- Taxa de conversão (ex: vendas ou preenchimento de formulário): avalia quantas pessoas realizaram o que era esperado (exemplo: comprar um produto ou preencher um formulário). Essa métrica é essencial para entender o impacto nos resultados de negócio.
- Tempo médio de resposta: indica o tempo médio até a primeira resposta, revelando o grau de interesse do usuário.
- Retenção / churn: medem, respectivamente, a manutenção e a perda de clientes ao longo do tempo, essencial para compreender a sua satisfação e a lealdade.

## 3. Tamanho da amostra e poder estatístico

O tamanho da amostra é calculado com base em:

- Nível de significância ( $\alpha$ ): define a tolerância ao erro do tipo I (falso positivo). Normalmente 5%, o que significa aceitar até 5% de chance que há uma diferença quando, na verdade, não há.
- Poder estatístico ( $1 - \beta$ ): mede a capacidade do teste de detectar uma diferença real quando ela existe. Recomenda-se 80% ou mais, o que reduz o risco de erro do tipo II (falso negativo).

- Tamanho do efeito esperado ( $\Delta$ ): é a menor diferença entre as versões que se deseja identificar como relevante do ponto de vista do negócio. Quanto menor esse valor, maior precisa ser a amostra.
- Variabilidade dos dados históricos: refere-se à flutuação natural dos dados com base em testes ou campanhas anteriores. Alta variabilidade exige amostras maiores para garantir resultados confiáveis.

Para ter resultados confiáveis no teste A/B, **é preciso definir o tamanho da amostra seguindo o rigor estatístico**. O que pode ser feito através de técnicas estatísticas adequadas e ferramentas especializadas como, Optimizely ou Evan Miller,

---

## Estratégias de amostragem no CRM

A forma como os grupos de teste é formado pode impactar a validade dos resultados. Por isso, a definição do processo amostral é uma etapa importante em experimentos envolvendo clientes. Três abordagens comuns e suas aplicações são apresentadas a seguir:

### A

#### Randomização pura

Os clientes são distribuídos de forma aleatória entre os grupos de controle e teste. Essa técnica é simples e útil para campanhas em larga escala, pois reduz o risco de viés, mas pode ignorar particularidades comportamentais.

### B

#### Randomização estratificada

O conjunto de dados é segmentado previamente por critérios-chave (ex: tipo de cliente, estágio no funil, ticket médio), em seguida é realizado o sorteio aleatório dentro dos estratos. Dessa forma, garante balanceamento entre grupos com perfis semelhantes, ideal para testes em CRM, pois aumenta a comparabilidade dos resultados.

**C**

## Clusterização + randomização

Em bases grandes e heterogêneas, aplicar um modelo de clusterização (como: K-means ou DBSCAN) pode ajudar a agrupar clientes com comportamentos semelhantes. A randomização é feita dentro dos clusters, promovendo equilíbrio e maior controle sobre a heterogeneidade da amostra.

### Quadro resumo da estratégia de amostragem no CRM:

Estratégia	Descrição	Quando Usar	Vantagens
<b>Randomização pura</b>	Distribuição totalmente aleatória entre os grupos.	Campanhas amplas com público homogêneo.	<b>Simples, reduz viés inicial.</b>
<b>Randomização estratificada</b>	Segmentação prévia por critérios (ex: tipo de cliente), com sorteio interno.	Bases com perfis variados e foco em equilíbrio entre grupos.	<b>Garante comparabilidade entre perfis.</b>
<b>Clusterização + randomização</b>	Agrupamento por comportamento (ex: via K-means) seguido de randomização.	Bases grandes e heterogêneas, com muitos padrões distintos de clientes.	<b>Controla variabilidade e melhora a precisão dos grupos.</b>

## 1. Definição dos grupos

A etapa inicial para executar um teste A/B de forma estruturada é a definição clara dos grupos de teste. Essa segmentação é importante para garantir comparabilidade e validade estatística. Os grupos recomendados são:

- **Grupo Controle (GC):** recebe a experiência padrão ou nenhuma comunicação. Esse grupo serve de baseline, e permite comparar os resultados das variações testadas com o comportamento natural dos clientes.
- **Grupo A:** recebe a variação 1, normalmente representa a campanha em uso até o momento. Ou seja, serve de comparação com a nova proposta.
- **Grupo B:** recebe a variação 2, que representa a nova proposta/teste). Isto é, o elemento que se deseja testar, (uma nova mensagem, layout etc.)

Caso o teste envolva mais de duas variações (testes multivariados) pode haver mais grupos (C, D, etc.). No entanto, vale destacar que **quanto mais grupos forem adicionados, maior será a necessidade de uma amostra robusta**, pois o poder estatístico se dilui com a divisão do público(amostra).

## 2. Duração do teste

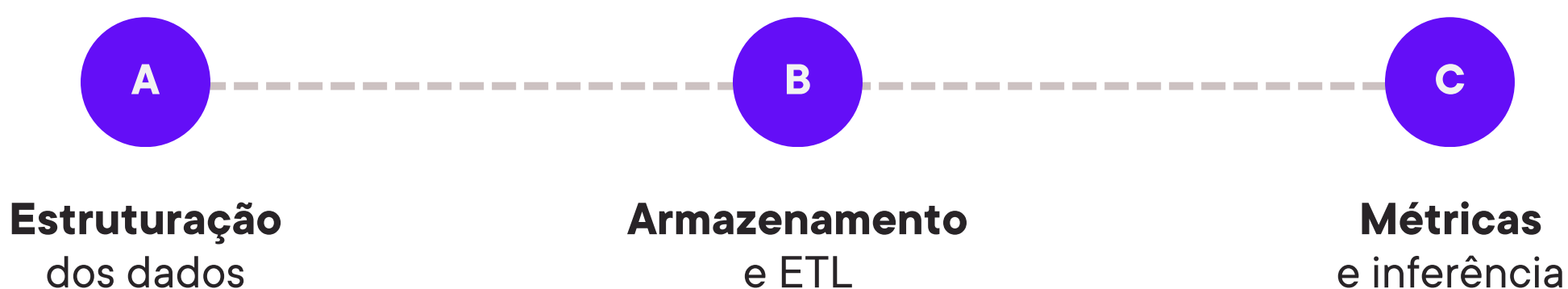
A extensão do teste deve ser suficiente para atingir o tamanho de amostra necessário sem interferência de ruídos sazonais ou comportamentos atípicos. No CRM, o tempo pode variar de horas (para campanhas push) a semanas (para retenção).

No contexto de CRM, a duração pode variar conforme o tipo de ação testada:

- **Campanhas de Notificação Push** normalmente requerem algumas horas para obter resultados significativos.
- **Testes focados na retenção ou comportamento** de longo prazo podem necessitar dias ou semanas de acompanhamento.

**Importante:** evite interromper um teste antes do período previamente estabelecido, mesmo que um resultado pareça promissor. A **interrupção antecipada** prejudica a significância estatística, tornando as conclusões inválidas ou enviesadas.

## UNIFICAÇÃO DE DADOS E MENSURAÇÃO



A confiabilidade dos resultados em um teste A/B está diretamente ligada à maneira como os dados são coletados e organizados. Para realizar o experimento de forma correta e conclusiva é importante estruturar os dados de forma consistente. A seguir, alguns cuidados com os dados antes de iniciar o estudo:

### a) Estruturação dos dados

- Ter identificadores únicos de clientes: atribuir um identificador único para cada indivíduo, permite acompanhar toda a sua jornada, evitando, assim, duplicações. Ou seja, garante que cada indivíduo esteja em apenas um grupo, possibilitando comparações justas entre as variações testadas.
- Capturar todos os pontos de contato da jornada: registre cada interação durante a jornada, tais como abertura de e-mails, cliques, respostas, conversões ou compras. Dessa forma, é possível compreender em que etapa da jornada cada versão teve maior ou menor impacto.

- Padronizar variáveis contextuais: elementos como canal de comunicação, horário de envio, dispositivo utilizado e localização podem afetar os resultados. Assim, padronize ou controle essas variáveis para evitar comparações enviesadas entre os grupos.

## b) Armazenamento e ETL

É importante **centralizar os dados coletados durante o teste A/B**, em uma estrutura de armazenamento robusta, como um data warehouse ou data lake.

- Use pipelines ETL/ELT para padronizar, transformar e organizar os dados para análise.
- Dessa forma é possível consolidar informações dos diferentes grupos (controle e variações) e etapas da jornada do usuário com segurança e escalabilidade.

**Ferramentas recomendadas para armazenamento:** BigQuery, Snowflake, Redshift. Já para transformações e orquestração: dbt (data build tool) e Apache Airflow.

## c) Métricas e inferência

A análise deve ser fundamentada em métodos estatísticos adequados ao tipo de dado e ao objetivo da comparação. **Escolher o teste corretamente é garante a significância e validade** dos resultados. Utilize testes estatísticos adequados ao tipo de dado. Exemplos:

- **Teste Z ou Teste T:** use para comparar médias, tais como o tempo médio de resposta ou o preço médio de uma compra.
- **Teste qui-quadrado ou teste de proporções:** indicado para taxas ou proporções, como taxas de cliques ou conversões.
- **Teste U de Mann-Whitney:** utilizado para comparar distribuições entre duas ou mais amostras independentes. Utilizado, quando os dados (amostras) não seguem a distribuição normal.

Adicionalmente, **sempre calcule o intervalo de confiança** (normalmente 95%) para validar a robustez do resultado e minimizar riscos de decisões baseadas em flutuações aleatórias.

## COMO EVITAR VIESES E ARMADILHAS COMUNS

A interpretação incorreta dos resultados em um teste A/B pode conduzir às decisões estratégicas equivocadas. Neste contexto, as campanhas, produtos e investimentos podem ser impactados. **Armadilhas metodológicas e vieses estatísticos comprometem a validade do teste, tornando as conclusões pouco confiáveis** — mesmo que os resultados pareçam promissores. Para garantir rigor e confiança nos resultados, fique atento aos principais riscos:

### **Variação múltipla:**

Teste apenas um elemento por vez (ex: assunto do e-mail, layout da página, benefício da oferta) para garantir isolamento da variável. Incluir múltiplas mudanças, simultaneamente, dificulta identificar qual fator foi responsável pela diferença nos resultados, comprometendo o aprendizado.

### **Interferência externa:**

Eventos promocionais, crises ou sazonalidades podem enviesar os resultados. Controle o período do teste, evite datas atípicas e registre anomalias que possam explicar flutuações inesperadas.

### **Overlap entre grupos:**

Garanta que o mesmo cliente não esteja em dois testes simultaneamente com o mesmo objetivo. A sobreposição pode contaminar os dados e dificultar a atribuição correta dos efeitos observados.

### **Cherry-picking:**

Evite selecionar dados que favorecem uma hipótese desejada. A análise deve ser completa, objetiva e pré-definida, sem aplicar filtros retroativos que distorçam a realidade. Essa prática compromete a credibilidade dos resultados e das decisões baseadas neles.

## MEDIR A EXPERIÊNCIA DO CLIENTE PARA DIRECIONAR ESTRATÉGIAS DE NEGÓCIOS

A adoção de boas práticas estatísticas e metodológicas é o que transforma o teste A/B em uma ferramenta confiável e estratégica, capaz de guiar decisões com base em evidências e não em suposições.

*O Anexo I contém um checklist prático de validação que pode ser usado antes de divulgar os resultados ou tomar ações com base no experimento.*

### dhauz recomenda (boas práticas):

Para garantir a escalabilidade e confiabilidade dos testes A/B, é necessário estruturar o processo. Existem três pilares principais: automação, centralização de dados e governança. A seguir, é proposto algumas recomendações para evoluir a maturidade da operação:



### AUTOMATIZE A SEGMENTAÇÃO E DIVISÃO DOS GRUPOS

Utilize ferramentas de CRM e orquestração de campanhas para segmentar os dados de forma aleatória, consistente e auditável. Dessa forma, reduz erros manuais e facilita a repetição do processo.

**Exemplos de ferramentas:** Salesforce, RD Station, HubSpot, Oracle Eloqua.

## INTEGRE DADOS EM UM AMBIENTE CENTRALIZADO PARA ANÁLISES MAIS CONFIÁVEIS

Unifique as fontes de dados em um data warehouse ou data lake para garantir que todas as interações dos usuários estejam disponíveis em um único local. Isso melhora a integridade da análise e reduz riscos de inconsistência entre sistemas.

**Crie uma governança para os experimentos:** Estabeleça processos e regras para o controle de testes simultâneos, evitando sobreposição de públicos e garantindo que os resultados possam ser replicados. Boas práticas incluem a definição de responsáveis, o uso de checklists padronizados e a criação de um repositório com o histórico de testes realizados.

**Conclusão:** dados, estatística e automação a favor da experiência do cliente.

Aplicar testes A/B no CRM de forma técnica possibilita criar uma cultura de experimentação estruturada, estatisticamente válida e centrada no cliente.

**COM ESSA ABORDAGEM, SUA EMPRESA PODE TOMAR DECISÕES MAIS RÁPIDAS, PRECISAS E PERSONALIZADAS, ELEVANDO A PERFORMANCE DE CAMPANHAS E A MATURIDADE ANALÍTICA COMO UM TODO.**

## Anexo 1

### Checklist de Validação para Testes A/B

- ✓ Planejamento e Execução
- ✓ Coleta e Qualidade dos Dados
- ✓ Análise Estatística
- ✓ Verificação de Anomalias e Viéses
- ✓ Interpretação e Ação

Para garantir a escalabilidade e confiabilidade dos testes A/B, é necessário estruturar o processo. Existem três pilares principais: automação, centralização de dados e governança. A seguir, é proposto algumas recomendações para evoluir a maturidade da operação:

#### 1. Planejamento e Execução

- O objetivo do teste está claramente definido (ex: aumentar conversão, reduzir churn)?
- Foi testado apenas um elemento por vez?
- Os grupos (controle e variações) foram distribuídos aleatoriamente e sem sobreposição?
- O tamanho da amostra foi suficiente para garantir significância estatística?
- A duração do teste foi respeitada, sem interrupções antecipadas?

#### 2. Coleta e Qualidade dos Dados

- Os dados estão centralizados e consolidados em um único repositório confiável?
- Todos os pontos de contato da jornada foram capturados (ex: abertura, clique, compra)?
- As variáveis que poderiam influenciar os resultados (ex: horário, canal, dispositivo) foram **controladas ou normalizadas?**

### 3. Análise Estatística

- Foi utilizado o **teste estatístico correto** para o tipo de métrica (T/Z, qui-quadrado, Mann-Whitney)?
- O resultado apresenta **significância estatística (p-valor  $\leq 0,05$ )**?
- Foi calculado o **intervalo de confiança (geralmente 95%)**?
- A análise não foi influenciada por cherry-picking ou filtragens retroativas?

### 4. Verificação de Anomalias e Viéses

- Foram registradas possíveis **interferências externas** (ex: datas promocionais, eventos inesperados)?
- O comportamento dos grupos foi **comparável** antes do início do teste?
- Existe documentação clara sobre **quais hipóteses foram testadas e com quais critérios**?

### 5. Interpretação e Ação

- Os resultados são estatisticamente válidos e relevantes do ponto de vista de negócio?
- Há clareza sobre qual variação venceu e por quê?
- A equipe responsável validou os achados antes de aplicar mudanças em larga escala?



**QUER ESTRUTURAR UMA OPERAÇÃO  
DE EXPERIMENTAÇÃO CONTÍNUA  
E DATA-DRIVEN?**

A dhauz pode te apoiar desde a modelagem dos testes até a automação da análise.

**Vem conversar com a gente!**