



# BOOST THE CLICK EMAIL A/B OPTIMIZATION

# TABLE OF CONTENTS



Introduction



Data Description



Data Processing



Comparing Campaigns  
performance



Hypothesis Testing



Conclusion



# INTRODUCTION

This project focuses on conducting an **A/B Testing experiment in an email marketing campaign** to evaluate the effectiveness of two different email subject lines in attracting user engagement.

The campaign was split into two groups:

- **Subject Line A:** "Exclusive offer just for you today!"
- **Subject Line B:** "We have a special gift for you 📁"



# DATA DESCRIPTION

daocuong.data@gmail.com

daocuong.data@gmail.com

- 1. Email ID:** Unique identifier for each email sent.
- 2.Campaign Name:** Experimental group (A or B) used for A/B testing.
- 3.Subject Line:** The subject line of the email (used to test audience engagement.).
- 4.Sent Time:** The timestamp when the email was sent.
- 5.Delivered:** Indicates whether the email was successfully delivered to the recipient.
- 6.Opened:** Indicates whether the recipient opened the email.
- 7.Clicked:** Indicates whether the recipient clicked on any link inside the email.
- 8.Bounced:** Indicates whether the email failed to be delivered (hard or soft bounce).
- 9.Marked Spam:** Indicates whether the recipient marked the email as spam.
- 10.Unsubscribed:** Indicates whether the recipient unsubscribed from future emails.
- 11.Replied:** Indicates whether the recipient replied to the email.
- 12.Converted:** Indicates whether the recipient performed the desired conversion action (e.g., purchase, signup).
- 13.Conversion Target:** The specific goal of the campaign - e.g., Purchase, Signup, Download, etc.
- 14.Device:** The device type used to open the email.
- 15.Open Time:** Timestamp of when the email was opened.
- 16.Location:** The city or region inferred from IP or user profile.



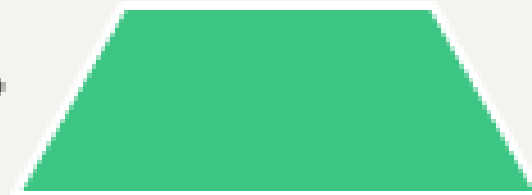
### Integer Data

Columns with 0/1 values



### Conversion Process

Transforming True/False to 0/1



### Boolean Data

Columns with True/False values



# DATA PROCESSING

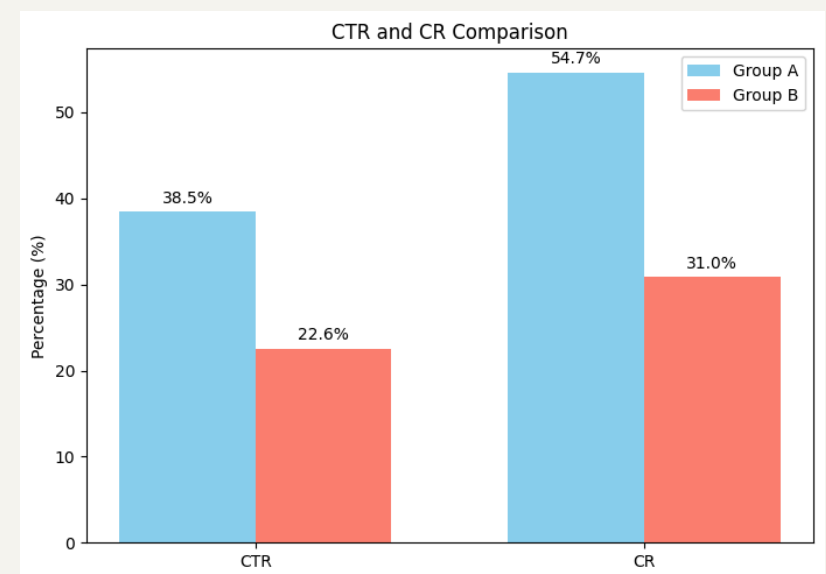
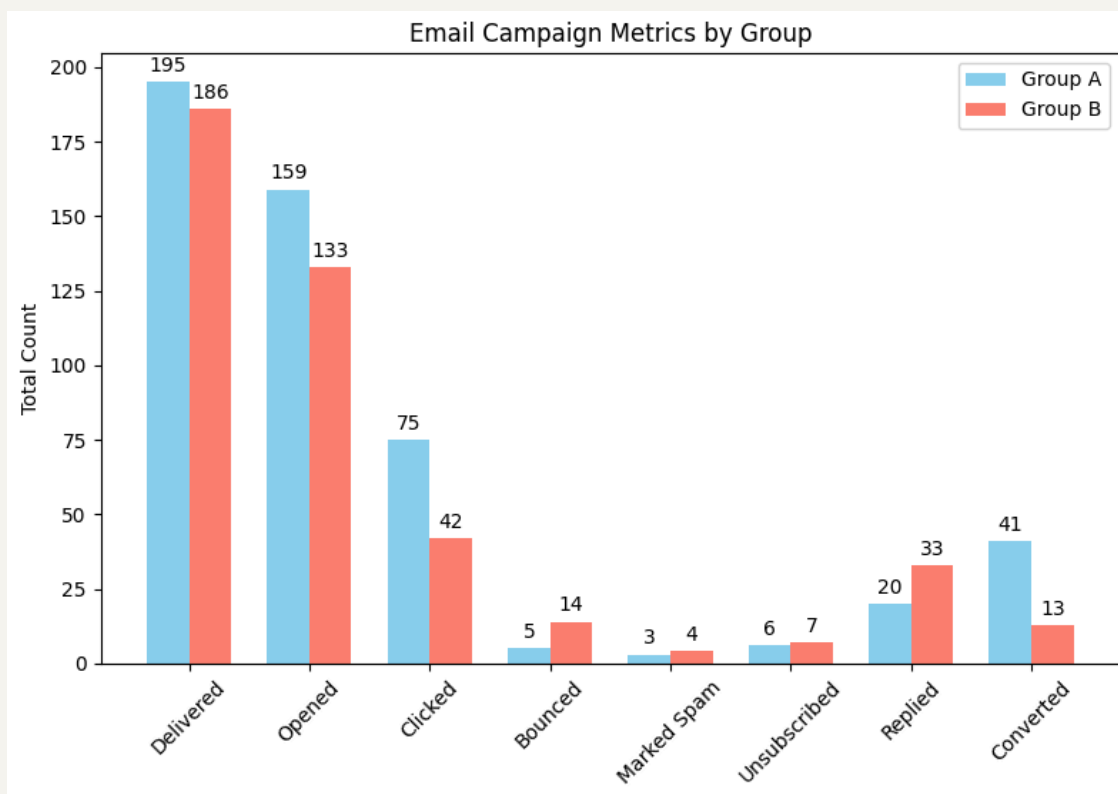
# COMPARING CAMPAIGNS PERFORMANCE

Group A outperforms Group B in most important metrics:

- Open Rate
- Click-Through Rate (CTR)
- Conversion Rate (CR)



This shows that Subject Line A ("Exclusive offer just for you today!") can be **more attractive and convincing** to users.



## Click-Through Rate (CTR)

The percentage of delivered emails that resulted in a click on a link within the email.

## Conversion Rate (CR)

The percentage of recipients who clicked a link and then completed a desired action (e.g., purchase, signup).



# HYPOTHESIS TESTING

## Testing Objective

Is Subject Line A **more engaging, attractive and effective** than Subject Line B?

## Hypothesis

- Null Hypothesis ( $H_0$ ):

There is no difference, or Subject Line A is not more attractive than Subject Line B

$$H_0: \mu_a \leq \mu_b$$

- Alternative Hypothesis ( $H_1$ ):

Subject Line A is more engaging than Subject Line B

$$H_1: \mu_a > \mu_b$$



- $\mu_a$ : Effectiveness metric (Open Rate / CTR / CR) of Group A
- $\mu_b$ : Effectiveness metric of Group B



### Shapiro-Wilk Test for its residuals:

- OpenRate - Group A:  $W = 0.4718$ ,  $p = 0.0000 \rightarrow \text{X}$  Not normal
- OpenRate - Group B:  $W = 0.5653$ ,  $p = 0.0000 \rightarrow \text{X}$  Not normal
- CR - Group A:  $W = 0.6333$ ,  $p = 0.0000 \rightarrow \text{X}$  Not normal
- CR - Group B:  $W = 0.5822$ ,  $p = 0.0000 \rightarrow \text{X}$  Not normal
- CTR - Group A:  $W = 0.6168$ ,  $p = 0.0000 \rightarrow \text{X}$  Not normal
- CTR - Group B:  $W = 0.5163$ ,  $p = 0.0000 \rightarrow \text{X}$  Not normal

Both Open rate, CR and CTR are **not normally distributed**, so you did the right thing by using a non-parametric test (**Mann-Whitney U test**) instead of a T-test.

# RESIDUALS NORMALITY CHECK

“Is the data of each group normally distributed?”



# MANN WHITNEY U-TEST

**OpenRate:**  $U = 19954.50$ ,  $p = 0.0104$

✓ Open Rate of **Group A** is significantly greater than **Group B**

**CR:**  $U = 1948.5$ ,  $p = 0.0071$

✓ CR of **Group A** is significantly greater than **Group B**.

**CTR:**  $U = 21015.0$ ,  $p = 0.0004$

✓ CTR of **Group A** is significantly greater than **Group B**.

Based on the statistical test results  
( $p\text{-value} < 0.05$ )

**Group A was more effective at attracting attention, generating interest, and driving user actions compared to Group B.**

# CONCLUSIONS AND RECOMMENDATIONS

A/B testing is an effective method for evaluating different versions of email elements, such as subject lines, designs, and calls-to-action. Key performance metrics like Open Rate, Click-Through Rate (CTR), and Conversion Rate (CR) offer valuable insights into recipient behavior and engagement levels. To determine whether observed differences are statistically significant and not due to random variation, statistical tests such as the Mann-Whitney U Test can be applied. Additionally, conducting thorough data cleaning and ensuring balanced group distribution in the test enhances the reliability and validity of the results.

