**CASE STUDY**

1. *What percentage of users opened the email and what percentage clicked on the link within the email?*  
   After analyzing the 3 given tables and evaluating the metrics for calculating the percentage of users who opened the email where around 10.35 %.   
     
   As for the percentage of users who clicked on the link within the email was found out to be 2.12 %.
2. *The VP of marketing thinks that it is stupid to send emails in a random way. Based on all the information you have about the emails that were sent, can you build a model to optimize in future how to send emails to maximize the probability of users clicking on the link inside the email?*  
   According to the given question, I have built multiple machine learning models to predict the likelihood of users clicking on the email link using features from the email and user behavior data such as Simple Logistic Regression, Decision Tree and Random Forest Classifier.

I have also done an ROC AUC (Receiver Operating Characteristic Curve – Area Under the Curve) Score Evaluation too because it gives a robust, threshold-independent measure of how well our model can separate users who are likely to click and those who aren’t – which is needed to optimize email targeting among the list of users. (likelihood to click is seen)

1. *By how much do you think your model would improve click through rate (defined as # of users who click on the link/total users who receive the email). How would you test that?*

In the original email campaign, emails were sent randomly. The baseline CTR is calculated as around 2.119 %.

But after using model based CTR Approach, we get around 6.17 % (like targeting the top 30 % with the highest predicted click probability).

With a baseline CTR of 2.119%, our model can improve click performance to 6.17% when targeting the top 30% most likely users — a 191.17% boost in engagement.  
  
This testing proves that targeted email strategies can deliver dramatically better results than random sends.

1. *Did you find any interesting pattern on how the email campaign performed for different segments of users? Explain.*

According to the exploratory data analysis, I found insights such as-  
  
-> Email Text: long\_email performs better than short\_email in terms of CTR.

-> Email Version: personalized emails lead to higher clicks than generic ones.

-> Hour of Day: There are noticeable peaks in certain hours, suggesting time-of-day optimization can help.

-> Weekday: Some weekdays (like Wednesday or Thursday) show higher engagement than others like Sunday.

After the segment analysis using grouping methodology and deriving visualizations for each of them we found out that-  
  
Personalized emails outperformed generic versions by 3–5% in CTR. Short text emails had slightly higher CTR than long versions. Best send hours: 2 PM – 5 PM (local time). Best days: Tuesday to Thursday. High-value customers (≥5 past purchases) clicked at almost double the rate. Some countries (e.g., US, Canada) had significantly higher engagement.

These patterns can inform targeted content creation and scheduling in future campaigns.

*References-*  
1. To understand about Email Campaign Data, I utilized the help of-  
  
<https://www.kaggle.com/datasets/loveall/email-campaign-management-for-sme/>

<https://www.analyticsvidhya.com/blog/2022/05/solving-business-case-study-assignments-for-data-scientists/>

2. How to optimize Click-Through Rate-  
  
<https://www.kaggle.com/datasets/swekerr/click-through-rate-prediction/>