# Problem Statement

Most organizations today rely on email campaigns for effective communication with users. Email communication is one of the popular ways to pitch products to users and build trustworthy relationships with them.

Email campaigns contain different types of CTA (Call To Action). The ultimate goal of email campaigns is to maximize the Click Through Rate (CTR).

CTR is a measure of success for email campaigns. The higher the click rate, the better your email marketing campaign is. CTR is calculated by the no. of users who clicked on at least one of the CTA divided by the total no. of users the email was delivered to.

As a part of the Data Science team, in this hackathon, I need to build a smart system to predict the CTR for email campaigns and therefore identify the critical factors that will help the marketing team to maximize the CTR.

# Objective:

**My task at hand is to build a machine learning-based approach to predict the CTR of an email campaign.**

# About the Dataset :

You are provided with the information of past email campaigns containing the email attributes like subject and body length, no. of CTA, date and time of an email, type of the audience, whether it’s a personalized email or not, etc and the target variable indicating the CTR of the email campaign.

# Procedure

# **Table of Content:**

1. Step 1: Importing the Relevant Libraries
2. Step 2: Data Loading and Inspection
3. Step 3: Data Cleaning
4. Step 4: Exploratory Data Analysis
5. Step 5: Building Model

# **Approach For Better Prediction:**

For the better prediction of CTR I have used following machine learning algorithms:

1. Linear Regression
2. XGBoost Regression
3. Gradient Boosting Regression
4. Random Forest Regression

At the starting point of the model building, I have split my data set into two parts:

1. Train set: For Training the model which occupies 90% of data points of entire dataset
2. Validation set: For verifying the accuracy of our model which occupies the 10% of data point of entire dataset

In next step I tried to figure out best parameters for my models for the better result.

For this I have used RandomSearchCV library for figuring out best parameters for the respective model based on data set.