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.

CTR = No. of users who clicked on at least one of the CTA / No. of emails delivered

CTR depends on multiple factors like design, content, personalization, etc.

- How do you design the email content effectively?
- What should your subject line look like?
- What should be the length of the email?
- Do you need images in your email template?

Table Content

1. **Importing the Necessary Libraries:-** Firstly I have import all the libraries which are use in throughout model building.

2. **Data Loading/Inspection:-** Here I have read the both dataset using pandas then done some inspection about data like what is the shape of dataset, what type of columns it contain.

3. **Data Preprocessing/Feature Engineering:-** Here I have drop some of columns which are not relevant for prediction and convert the categorical features into numerical features.

4. **Exploratory Data Analysis:-** Here I done some kind of exploratory like what types of feature we have , how feature are correlated to each other or with label, which feature is more affective to label data.

5. **Data Cleaning:-** Here Firstly I have see is dataset contain null values or missing value or not. Then I have remove outliers.

6. **Model Building (Approach used to solve the problem):-** For getting better prediction result I have use four regression algorithms which are as follow:-
    1. Linear Regression
    2. XGBoost Regression
    3. Gradient Boosting Regression
    4. Random Forest Regression

Before start training model I have firstly split train dataset into train and valid where train for training the model and valid for testing the model to get the best possible result. I have split it with .9:.1 whereas .1 means 10% of train dataset for testing the model and .9 means 90% of train dataset for training the model.

Then for better result I have use hyperparameter tunning. For hyperparameter tunning I have use Random Search CV to find the best parameter for ML algorithms.

Then after finding all the best parameter I have select the model which will give the highest R2 score.

7. **Prediction of Test Dataset Click Rate:-** After creating model I have select the model which give the highest R2 score and predict the Click rate of test dataset with that model and store the value in pandas dataframe.

8. **Save DataFame in CSV File:-** Here I have save the pandas dataframe which contain test dataset campaign_id and predicted values of click_rate into CSV file.