Email Campaign Analysis and Insights for Improved Engagement

Gaurav Singh

04.1.2023

Objective:

Evaluate the candidate's skills in data analysis, machine learning application, and effective communication of findings.

Part 1: Data Analysis

Data Overview

The dataset contains 154 entries with 7 columns. Here's a summary of each column:

Subject: 154 unique email subjects, indicating a variety of campaigns.

Body: 154 unique email bodies, corresponding to the subjects.

Opened: Binary variable (True/False), representing whether the email was opened.

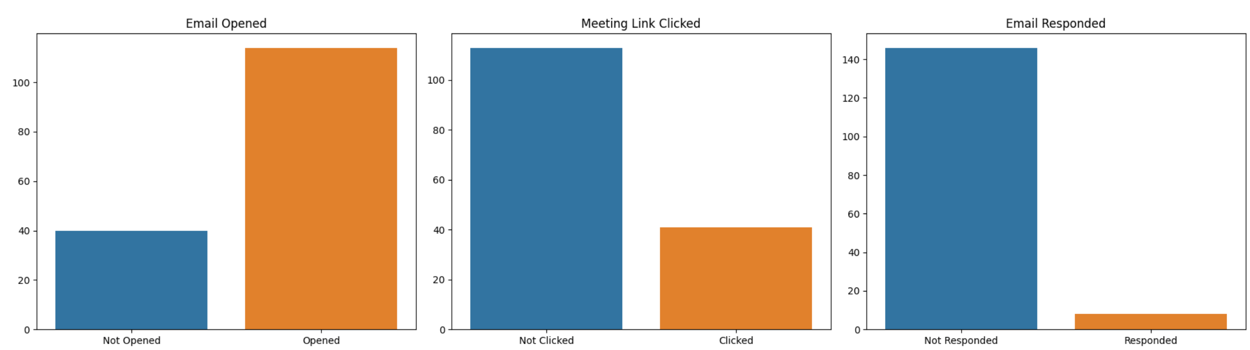
Meeting Link Clicked: Binary variable (True/False), but there are 16 missing values.

Responded: Binary variable (True/False), indicating if there was a response.

Campaign: There are 63 non-null entries with 7 unique campaign names. The rest are missing.

Meeting_Link_Clicked: Seems redundant with 'Meeting Link Clicked' and mostly contains missing values.

Exploratory Data Analysis :



Open Rates and Click-Through Rates (CTR):

The first chart shows that a significant number of emails are opened compared to those that are not. This suggests that the subject lines may be effective or the recipients are well-targeted.

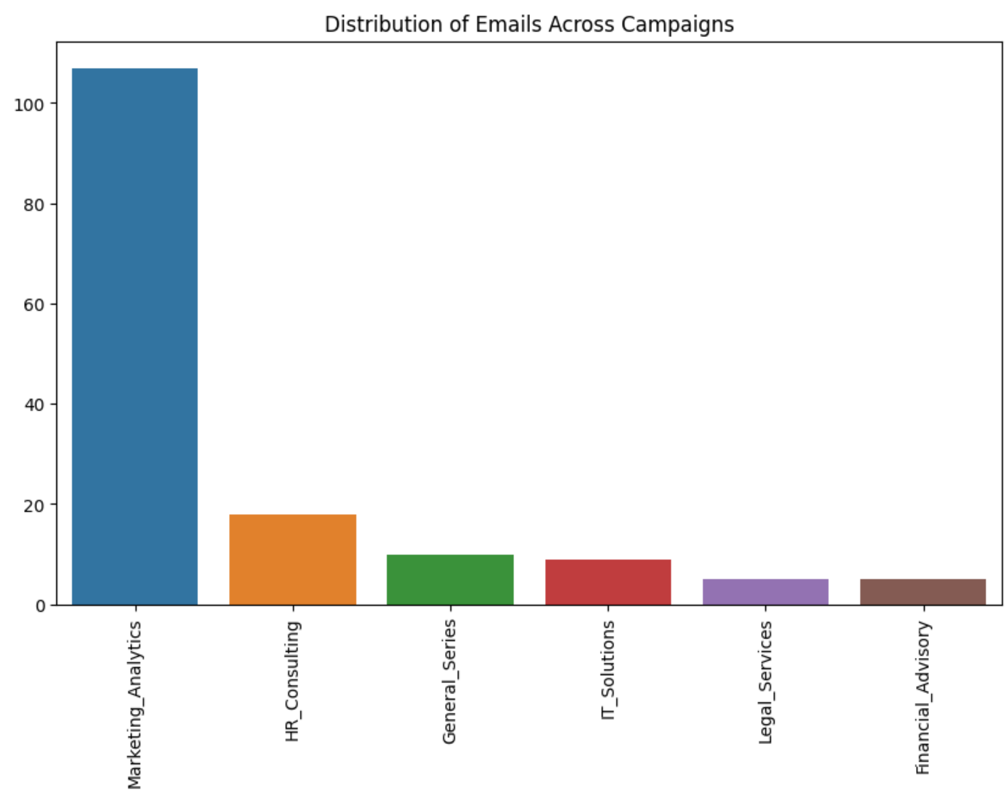
However, the CTR is lower as seen in the second chart, indicating that while the initial engagement is high, the content within the emails might not be compelling enough to drive recipients to click on the links provided.

Response Rates:

The third chart shows that the response rate is much lower than the open and click rates. This could indicate that the call-to-action within the emails is not strong or clear enough to elicit a response, or that the content is not resonating enough to prompt further action from the recipients.

Campaign Effectiveness:

The distribution of emails across campaigns indicates that 'Marketing_Analytics' is the most common category, suggesting a potential focus area or interest among the recipients.



The table down below showing the rates by campaign type suggests that 'IT_Solutions' has the highest open and CTR, indicating that the content in this category is particularly relevant or appealing to the recipients.

campaign	total_emails	opened_emails	clicked_emails	responded_emails	open_rate	click_through_rate	response_rate
Marketing_Analytics	107	79	28	5	0.738318	0.261682	0.046729
HR_Consulting	18	12	4	0	0.666667	0.222222	0.000000
General_Series	10	7	3	1	0.700000	0.300000	0.100000
IT_Solutions	9	8	3	2	0.888889	0.333333	0.222222
Financial_Advisory	5	4	1	0	0.800000	0.200000	0.000000
Legal_Services	5	4	2	0	0.800000	0.400000	0.000000

Insights for Improvement:

Content Relevance:

Work on tailoring the content within the emails to be more relevant to the recipient's interests or needs, especially for campaigns with lower engagement rates.

Call-to-Action (CTA):

Refine the CTAs to be more compelling and clear, encouraging recipients to take the desired action. A/B testing different CTAs could be beneficial to see which ones resonate more with the audience.

Personalization:

Increase the level of personalization in the emails. Personalized emails tend to have higher engagement rates. Use data analytics to understand the interests of the recipients and segment the audience accordingly.

Follow-up Strategy:

Implement a follow-up strategy for recipients who click but do not respond. The follow-up could provide additional information or incentives that might be needed to encourage a response.

Subject Line Testing:

Continue to test different subject lines to further increase open rates. The effectiveness of a subject line can diminish over time as recipients become accustomed to the sender's patterns.

Optimize Sending Time:

Analyze the best time to send emails based on when recipients are most likely to open and engage with them. This might vary by audience segment.

Mobile Optimization:

Ensure that emails are mobile-friendly. Many users read emails on their mobile devices, and poor mobile optimization can lead to lower engagement.

Engagement Tracking:

Develop a more granular tracking and analytics system to understand which parts of the email are engaging the users and at what point you are losing their interest.

Part 2: Model Development

In developing a machine learning model to predict email engagement, specifically open rates, we utilized a dataset containing various features such as the email subject, body, and engagement metrics (opened, clicked on meeting link, and responded). Here's an overview of the model development process and its inferences:

Data Preprocessing:

The 'campaign' category was numerically encoded to facilitate model processing, and the text content from 'subject' and 'body' was concatenated to form a single text feature.

Binary features like 'opened', 'meeting link clicked', and 'responded' were converted to numeric values for model compatibility.

Feature Engineering:

We employed BERT (Bidirectional Encoder Representations from Transformers) to extract sophisticated text features. The encoding process involved tokenization and attention mask application, ensuring that each text was adequately represented in a fixed-size embedding.

Non-text features were standardized to bring them onto a similar scale, and then combined with the BERT-derived text features, creating a comprehensive feature set.

Model Selection:

A simple neural network architecture with one hidden layer was chosen for its ability to learn non-linear patterns in the data.

Training and Validation:

- To address class imbalance, we applied SMOTE (Synthetic Minority Over-sampling Technique) to the training data.
- The model was trained for 10 epochs, showing a decrease in validation loss over time, which indicates learning and generalization capability.
- The validation process involved a qualitative assessment of model performance per epoch to ensure that the model was not overfitting and was generalizing well to unseen data.

Model Evaluation:

The classification report revealed an overall high accuracy of 88% on the test dataset, with a particularly high precision for the positive class (emails that were opened).

	precision	recall	f1-score	support
0.0	0.60	0.75	0.67	4
1.0	0.95	0.90	0.92	20
accuracy			0.88	24
macro avg	0.77	0.82	0.79	24
weighted avg	0.89	0.88	0.88	24

The model achieved a high f1-score for the positive class (0.92), suggesting a good balance between precision and recall.

Inferences:

- The model appears to be effective at predicting open rates, with strong performance metrics across precision, recall, and f1-score.
- The precision of 0.60 for the negative class (emails that were not opened) suggests the model is quite cautious, preferring to predict an email will be opened unless it's quite sure otherwise.

Insights for Improvement:

- To further improve engagement, consider refining the content within the emails. This could involve A/B testing different types of content and calls-to-action to see what resonates best with the audience.
- Analyze the subject lines of emails that were not opened to identify any commonalities that may be off-putting to recipients.
- Investigate the emails that were opened but did not result in a clicked link to understand if there's a disconnect between the email content and the linked content.
- Review the content of emails that led to clicks but not responses, as this could indicate a promising engagement that needs a final nudge, possibly through personalization or follow-up strategies.
- This analysis and the insights derived from it can guide strategic decisions to enhance the effectiveness of future email campaigns, driving up open rates and overall engagement.