Email Analysis

Task

For the given set of Marketing Emails along with their engagement data, find patterns and insights into how we can increase customer engagement.

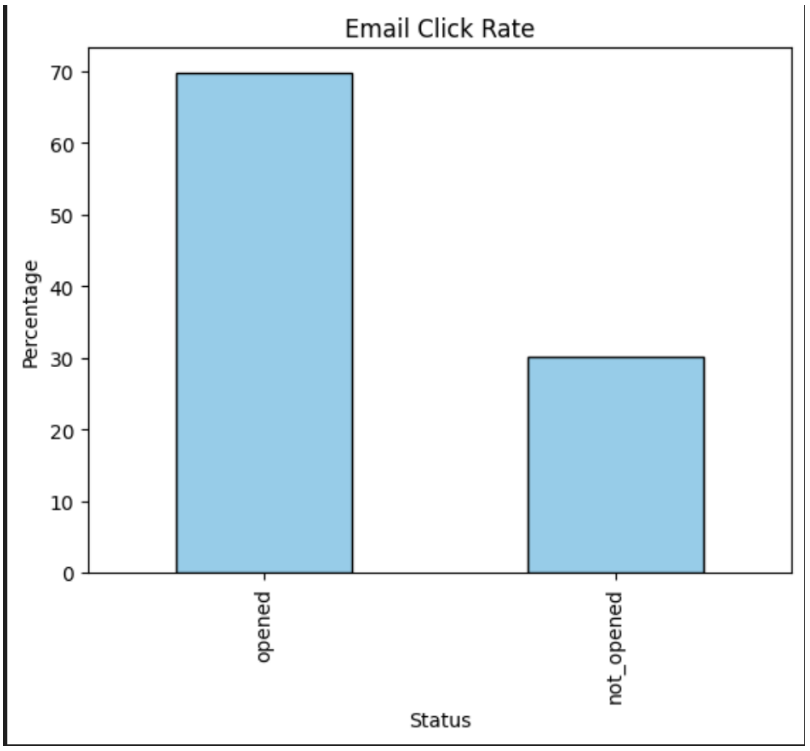
We have split it into two tasks

1. Develop a model to find out why a particular email will be opened by the customer.
2. Find out insights, on what leads to a customer engagement (responding/ scheduling a meeting) once he opens up the email.

SubTask1

We first looked at the percentage of how many first-time emails sent to a customer get opened.

We can see that only 70% of emails are viewed.



To Explore this a bit more we trained a model that would classify given a subject line whether an email will be opened or not.

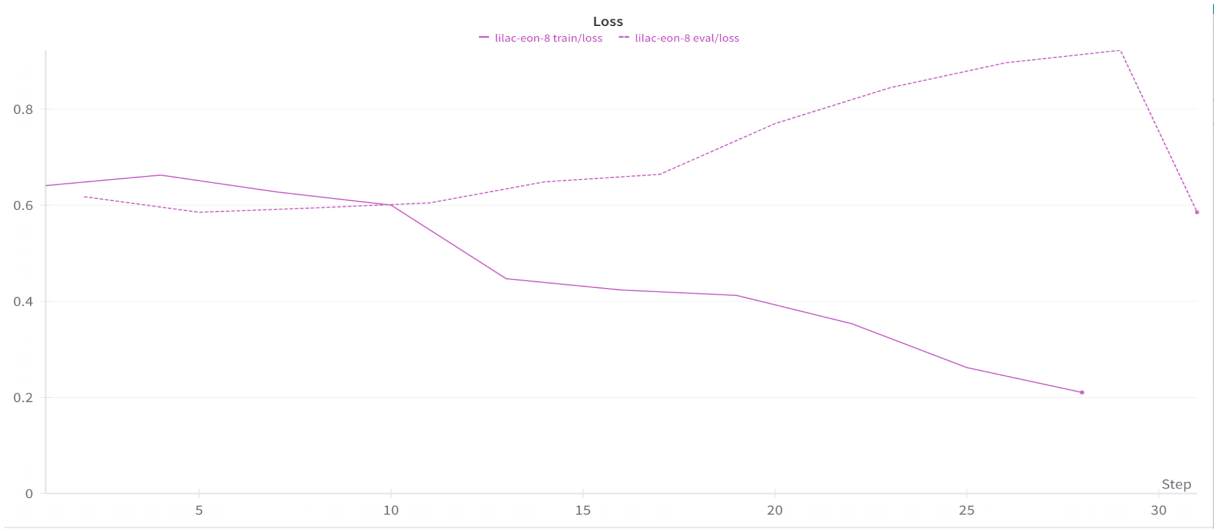
We used the subject lines from only the first emails sent to the customer.

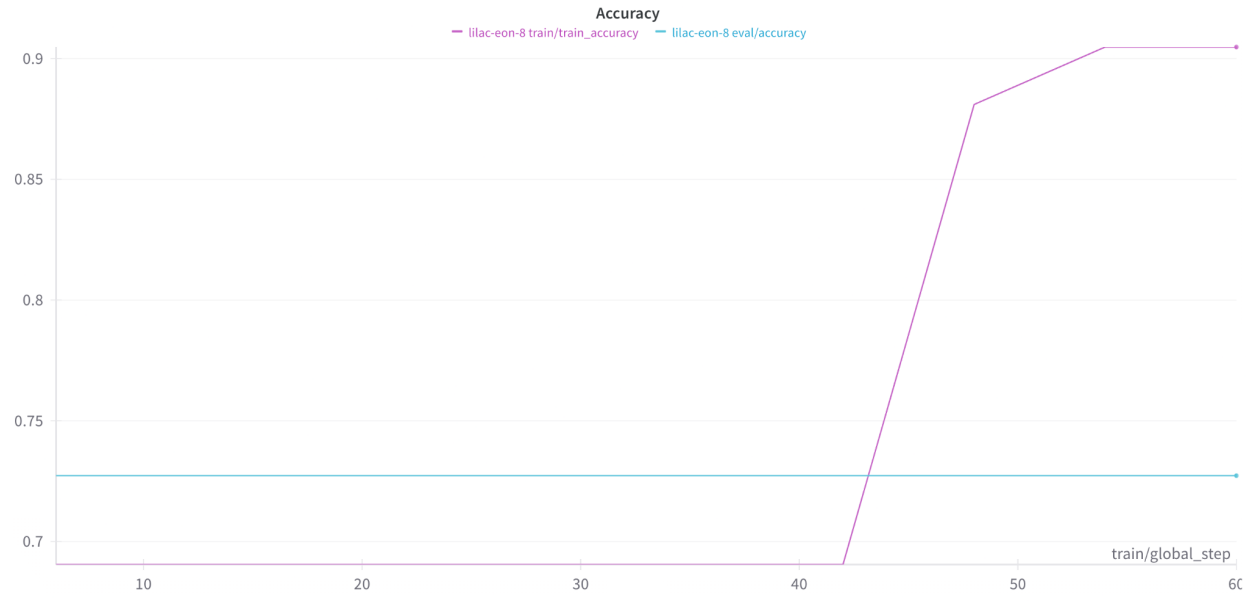
Preprocessing

We cleaned the subject line of any emoji and punctuations, and split the dataset keeping 80% for training and 20% for validation.

Model Development

We used a DistilBERT with a classification head on top, and trained for 10 epochs, with linear decay lr_scheduler.





We can see that the model is learning something, but our eval accuracy does not seem to be changing, one issue could be that our eval set is a bit too small, As we can see improvement in our train accuracy.

Model Interpretation

Using `transformers_interpret` library, we computer word attribution scores for all the words in our corpus, and saw which words were leading to lower engagement.

Following are the top 3 words that lead to people not opening the email

Word	Engagement Rate	Frequency	Example
Precision	33%	3	Unlock Growth Precision Marketing Analytics
Analytics	63%	27	Propel Marketing ROI Advanced Analytics!
Elevate	53%	13	Elevate Marketing Insights Next-Gen Analytics!

On further analysis, we found that for "Analytics" all the rejections were for Marketing Analytics related emails.

And we saw that as soon as we removed the Analytics keyword, the click-through rate for the Marketing emails increased from 71% (25/35) to 100%(7/7).

So to increase click-through rates for Marketing Emails, we should replace the Marketing Analytics keyword, with phrases like:
1. Marketing Actionable Insights
2. Marketing Data Insights

Example:
Propel Marketing ROI Advanced Analytics: 67% chance of getting opened
Propel Marketing Data-Driven Insights: 92% chance of getting opened

Similarly for **Precision** if we replace it with words like **Tailored**, the chances of the customer opening the email increases.

Subject	Body	Opened
Unlock Growth Precision Marketing Analytics	Hi [Recipient's Name] Do you feel your marketing campaigns could use a boost in efficacy? At [Your Company], we transform data into actionable insights that drive successful marketing strategies. Interested in a quick overview? Check out this link [meeting link] to lock in a time this week.	False
Unlock Growth Tailored Marketing Analytics	Hi [Recipient's Name] Do your marketing efforts feel like a shot in the dark? 💡 At [Your Company], we turn data into direction! Let's illuminate your strategy with actionable insights. How's your calendar for a quick intro call? [meeting link]	True

We can see that both the emails had the same content but when we replace Precision with Tailored it has a higher chance of getting opened.

Similarly for **Elevate** if we replace it with words like **Unlock** and **Empower** we see an increase in the click-through rate.

Sub-Task 2

Now we wanted to find out once a user has opened up the email what leads him to engage with it.
So one of the main ways for this is to use various persuasion techniques in the email, to convince the customer to engage in some way. As we did not have this data here I used GPT-4 as nowadays LLMs are being used extensively for annotating data. We defined a set of existing Persuasion Techniques and asked the LLM to annotate it.

Tag	Description
FOMO	Creating a sense of urgency or exclusivity to encourage immediate action.
SOC	Showcasing testimonials, reviews, or social media endorsements to build trust.
REC	Offering something valuable upfront to encourage recipients to reciprocate.
AUT	Establishing credibility by highlighting expertise, credentials, or endorsements.
PIA	Aligning the product or service with the recipient's self-image or aspirations.
EMO	Eliciting emotions, such as joy, excitement, or empathy, to create a connection.
VAL	Clearly communicating the unique value and benefits of a product or service.

We used the following system prompt to annotate the data
“””
You are a bot that finds persuasion techniques used in the given text. These are the following techniques you are going to look for:
SOC (Social Proof):Showcasing testimonials, reviews, or social media endorsements to build trust.
REC (Reciprocity):Offering something valuable upfront to encourage recipients to reciprocate.
AUT (Authority):Establishing credibility by highlighting expertise, credentials, or endorsements.
CTA (Call to Action):Clearly instructing recipients on the desired action, often using compelling language.
PIA (Personal Image and Aspirations):Aligning the product or service with the recipient's self-image or aspirations.
VAL (Value Proposition):Clearly communicating the unique value and benefits of a product or service.
EMO (Emotional Appeal):Eliciting emotions, such as joy, excitement, or empathy, to create a connection.
FOMO (Fear of Missing Out):Creating a sense of urgency or exclusivity to encourage immediate action.
Format your output as a list of dictionaries having the "persuation_technique" and "span" as keys. below is an example

```
[
  {   "span": <span of the text>
      "persuation_technique":"FOMO",
  },
  {   "span": <span of the text>
      "persuation_technique":"PIA",
  },
]
"""
```

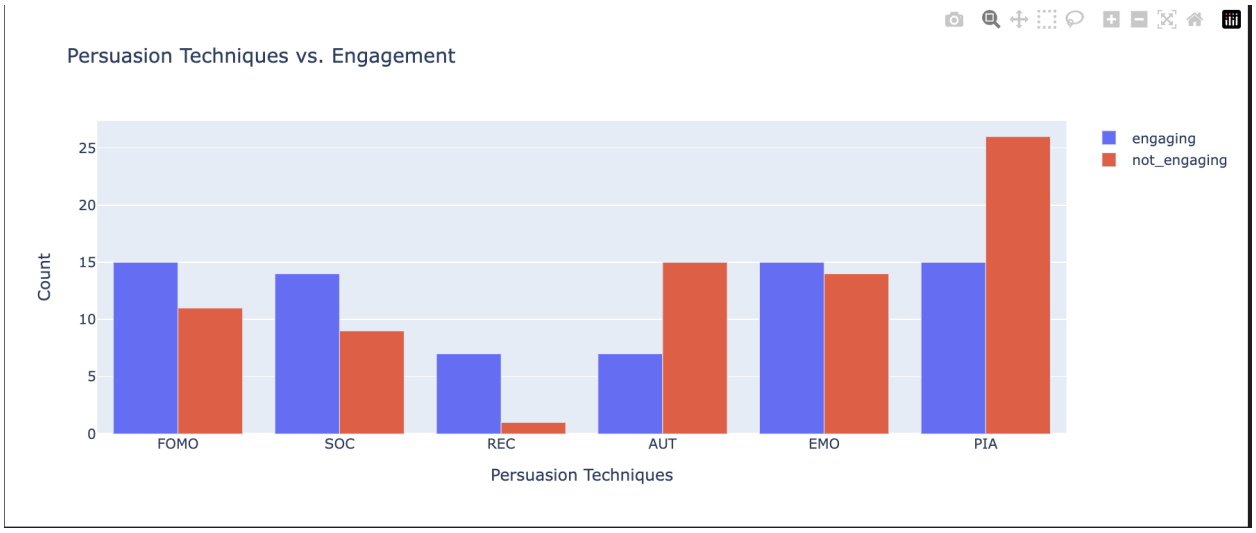
We tried with a variation of models gpt-3.5, gpt-3.5-turbo, gpt-4-1106-preview and found gpt-4-1106-preview to work the best.

Also tried a couple of different prompts but the one that extracts spans, works best and also gives a bit more interpretability about the annotations

As it could have some errors, I saved the output to a JSON, reiterated through it, and removed any wrong annotation to the best of my expertise.

I saw that there were a lot of false positives for **VAL** so I have dropped them as part of my analysis.

Results



So the three techniques that seem to add some value to the click-through rate are:
- REC (Reciprocity)
- SOC (Social Proof)
- FOMO (Fear of Missing Out)

The Technique that stands out the most here is Reciprocity, which means that if we provide something upfront to the customer lie a report or a free consultation, then the chances of engagement are very high.

And even though on the face of it it looks like PIA might lead to negative engagement on further analysis, we found that out of 26 negative samples, 25 of them contain none of the FOMO, SOC, and REC techniques, which has led to this non-engagement.