Email Marketing Campaign

Goal

Optimizing marketing campaigns is one of the most common data science tasks. Among the many marketing tools available, emails stand out as particularly efficient.

Emails are great because they are free, scalable, and can be easily personalized. Email optimization involves personalizing the content and/or the subject line, selecting the recipients, and determining the timing of the sends, among other factors. Machine Learning excels at this.

Case Description

The marketing team of an e-commerce site has launched an email campaign. This site has email addresses from all the users who created an account in the past.

They have chosen a random sample of users and emailed them. The email lets the user know about a new feature implemented on the site. From the marketing team perspective, success is if the user clicks on the link inside of the email. This link takes the user to the company site.

You are in charge of figuring out how the email campaign performed and were asked the following questions:

- What percentage of users opened the email and what percentage clicked on the link within the email?

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

Data

We have 3 tables in this case study.Available here

The 3 tables are:

email_table - info about each email that was sent

Columns:

- **email_id** : the Id of the email that was sent. It is unique by email.
- **email_text** : there are two versions of the email: one has "long text" (i.e. has 4 paragraphs) and one has "short text" (just two paragraphs)
- **email_version** : some emails were "personalized" (i.e. they had the name of the user receiving the email in the incipit, such as "Hi John"), while some emails were "generic" (the incipit was just "Hi,").
- **hour** : the local time on which the email was sent.

- **weekday** : the day on which the email was sent.
- **user_country** : the country where the user receiving the email was based. It comes from the user ip address when she created the account.
- **user_past_purchases** : how many items in the past were bought by the user receiving the email

---

email_opened_table - the id of the emails that were opened at least once.

Columns:

- **email_id** : the id of the emails that were opened, i.e. the user clicked on the email and, supposedly, read it.

---

link_clicked_table - the id of the emails whose link inside was clicked at least once.

Columns:

- **email_id** : if the user clicked on the link within the email, then the id of the email shows up on this table.