**Marketing Email Campaign**

**Goal**

Optimizing marketing campaigns is one of the most common data science tasks.

Among the many possible marketing tools, one of the most efficient is using emails.

Emails are great because they are free and can be easily personalized. Email optimization

involves personalizing the text and/or the subject, who should receive it, when should be

sent, etc. Machine Learning excels at this.

**Challenge 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 let the user know

about a new product on the site. From the marketing team perspective, a success

is if the user clicks on the link inside of the email. This link takes the user to the company site where user can buy the product.

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 head of marketing thinks that it is stupid to send emails to a random subset and 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 email campaigns 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 received 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 downloadable by clicking [here](https://docs.google.com/spreadsheets/d/1ziKRqkNPDuCFUrw-3NuYT4ZUoR5F1cOAgk6eiTw6rlg/edit#gid=0).

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 6 paragraphs) and one has "short text" (just 2 paragraphs)

email\_subject: there are 2 versions of the email subject: one with product name and one generic text.

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 user local time when the email was sent.

weekday : the day when 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

opened\_flag: the user clicked or not on the email and, supposedly, read it.

**"link\_clicked\_table"** - the id of the emails whose link inside was clicked at least once. This user was then brought to the site.

Columns:

email\_id : if the user clicked on the link within the email, then the id of the email shows

up on this table.

unsubscribe\_link\_click: Has the user clicked on unsubscribed link

CTA\_link\_click: Has the user clicked on Call to Action link

**Example**

Let's check one email that was sent

**email\_table**

|  |  |  |
| --- | --- | --- |
| Col Name | Value | Description |
| email\_id | 1234 | Id of the email |
| email\_text | short\_email | Was a short email |
| email\_subject | generic | Generic text in email subject |
| email\_version | personalized | Personalized with the user name in text |
| hour | 6 | 6 AM user local time |
| weekday | Thursday | Sent on Thursday |
| user\_country | US | User based in the US |
| user\_past\_purchases | 20 | 20 purchases on the site |

**email\_opened\_table**

|  |  |  |
| --- | --- | --- |
| Col Name | Value | Description |
| email\_id | 1234 | Email id |
| opened\_flag | Yes | User opened the email |

**link\_clicked\_table**

|  |  |  |
| --- | --- | --- |
| Col Name | Value | Description |
| email\_id | 1234 | Record gets created when emails opened |
| unsubscribe\_link\_click | No | User didn’t click on unsub link |
| CTA\_link\_click | Yes | User clicked on CTA link |