Email delivery experiment on funding rate improvement

## Background / Observation

Imagine you are working for a fintech company, such as Square, Webull, etc. The company’s goal is always to improve account link rate, funding rate, activities (trading or not) etc.

We see a common pattern, across different use cases, when it comes to encouraging users to take an action. The pattern is as follows - segment users, target them with messages and find an optimal message that works and productizes it.

We message our users to promote the action/feature. These messages could be sent via emails. In this experiment, we are focusing on the effect from emails running multiple A/B tests, which runs for 5 weeks.

Our goal is to discover the best Email to deliver and verify the effectiveness on various groups of users. To get there we’ll have to provide the following configuration for teams to adjust.

* Message(s).
* User-segments,
* Sequence and frequency.
* Metrics - target metrics and counter metrics.

## Hypothesis

Email is the best way to reach the approved and unfunded users. The category and inventory of messages are exhaustive and addresses the concerns that hold users back from funding the account.

We expect the impact to users behavior, including Email open rate, funding rate, comes from the following sources:

1. User segment
2. Email Template content, including subject title, GIF, text content
3. The order to send Email messages
4. The frequency of sending Emails

### Audience

* 480,000 unfunded users by 11/27/20, and we give 2+ weeks to these audiences to take action, if there is any.

### Goals

* Increase Email opening rate
* Increase Funding rate
* Confirm different funding and activation rates across unfunded segments
* Confirm different funding and activation rates across variance setup combinations to send out Emails.

### Metrics

* Email click through rates, app opens
* Funding Rate vs. Control
* Activation: % of users that links their bank account, funds, complete 1 trades, complete 2+ trades

## Experiment

### **General Approach**

1. Create a large pool of messages across different categories, 10 in total. (Done)
2. Send messages to understand relationship and engagement rate amongst message and user segment (12 user segments with 2 delivery frequencies, 24 treatments in total). (Done).

### **Users segmentation**

The segmentation to choose users samples is rule-based. In control and variants, we perform stratified sample of the user from each of the following segment:

1. Account approved within 6 months: True / False
2. Bank linked: True / False
3. If there is any activity in the past 20 days: True / False
4. If there is any trade activity in the past 5 days: True / False

Below is the summary table of user groups. There are 12 groups in total and we randomly sample 40,000 users from each group for our experiment. These groups are all treatment groups since they all receive Emails. The control groups are users with the same segment rules, but have not received any Emails.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **group\_id** | **approved\_within6M\_flag** | **link\_flag** | **recent\_activity\_flag(20days)** | **5day\_trade\_flag** | **num\_user** | **group\_name** |
| 0 | False | False | False | False | 4343086 | ML\_unfund\_exp\_control |
| 1 | False | False | True | False | 360430 | 20D\_Act |
| 2 | False | False | True | True | 374603 | 20D\_Act-5D\_Act |
| 3 | False | True | False | False | 921347 | link |
| 4 | False | True | True | False | 115686 | link-20D\_Act |
| 5 | False | True | True | True | 123536 | link-20D\_Act-5D\_Act |
| 6 | True | False | False | False | 845514 | 6M\_App |
| 7 | True | False | True | False | 227223 | 6M\_App-20D\_Act |
| 8 | True | False | True | True | 210603 | 6M\_App-20D\_Act-5D\_Act |
| 9 | True | True | False | False | 135009 | 6M\_App-link |
| 10 | True | True | True | False | 53211 | 6M\_App-link-20D\_Act |
| 11 | True | True | True | True | 65976 | 6M\_App-link-20D\_Act-5D\_Act |

### **Email Templates**

We prepared ten Emails for the experiments. The tops cover:

* Trust and safety
* Knowledge gaps
* Simplicity of use

You can use this hard coded list of Emails (‘PO\_number\_list’):

PO\_number\_list = ['ml\_funding\_enables\_investing','ml\_investing\_starts\_here','ml\_explore\_the\_app\_investing', 'ml\_funding\_faq','ml\_user\_clustering\_emails\_fracs','ml\_funding\_is\_safe','ml\_picking\_an\_investment', 'ml\_investing\_101','ml\_diversified\_portfolio','ml\_explore\_the\_app\_list']

### Email Delivery Strategy

We developed our own Email scheduling pipeline to deliver Emails for the experiment. The schedule has two frequencies, daily and twice a week. Each user segment is split equally further into two groups with different scheduling frequencies. Furthermore, to fully cover the space of Emails orders, each email is sent randomly (without replacement) for each Email on its scheduled day.

For this particular experiment, the starting date (day 0) is 11/30/20. The daily Emails are sent between Da\_0 and Day\_9 every day. The twice-a-week Emails are sent according to this schedule (Monday, Friday): Day\_0, Day\_4,Day\_7, Day\_11, Day\_14, Day\_18, Day\_21, Day\_25, Day\_28, Day\_32.

## Process details

We want to send out the emails and measure its impact. To cover the whole space of the audience and Email strategy, we went through the following steps:

* Produce Emails (10 total)
* Generate audience list (12 segments)
* Generate dates to send out Emails:   
  A. Split the audience groups into **24** groups to receive Emails at daily (‘\*\_D’) or twice a week (‘\*\_W’) frequency.
* Generate Email template IDs to be sent for each group at each day
* Send email out around 10 - 10:30 am.
* Track metrics for two plus weeks after the Emails are all sent.
* Synthesize results, reevaluate next steps.

## Timeline

As mentioned above, the experiment starts on 11/30/20 (Day\_0) and the last Email is sentay\_32. The waiting period to collect the final result is 2+ weeks.

## Description of Data files

Download link: <https://drive.google.com/drive/folders/1rq9B44LUACus-b3J0yF-_VRyOfhknzab?usp=sharing>

List of files:

1. ‘Sample\_segment\_groups.csv’
   1. Users segmentation rules as descriptive in the above users segmentation section.
   2. ‘group\_id’ is the ID for each user group: 0-11
   3. A note here: ‘user\_uuid’ is the count of user\_uuids, not the user\_uuids themselves.
2. ‘Sample\_uuid\_email\_order.csv'
   1. Email delivery schedule table for each user in the experiment.
   2. ‘user\_uuid’ is the unique ID for each user.
   3. ‘group\_id’ is the ID for each user groups the experiment: 0-23
   4. ‘group\_name’: actual user groups in the experiment as described in the process details section. There are 24 groups in total, because we deliver Email at two frequencies. The ‘\*\_D’ surfix stands for daily delivery. The ‘\*\_W’ surfix stands for twice a week delivery.
   5. ‘order\_\*’ columns hold the Email index to be delivered at each iteration.
      * For example, for ‘\*\_D’ user groups, ‘order\_0’ - ‘oder\_9’ means day\_0 to day\_9. For ‘\*\_W’ user groups, ‘order\_0’ and ‘order\_1’ means for the two deliveries in the first week.
      * The values in each column stands for the index (starting from 0) in this list. It represents the Emails to be sent on each day.  
        PO\_number\_list = ['ml\_funding\_enables\_investing','ml\_investing\_starts\_here','ml\_explore\_the\_app\_investing', 'ml\_funding\_faq','ml\_user\_clustering\_emails\_fracs','ml\_funding\_is\_safe','ml\_picking\_an\_investment', 'ml\_investing\_101','ml\_diversified\_portfolio','ml\_explore\_the\_app\_list']
3. ‘email\_events.csv'
   1. Table generated by our Email delivery system (Post Office) to track users' interaction with Emails.
   2. This table is optional since we provide you the user status summary table below.
4. ‘user\_events.csv'
   1. Current account status of users in the experiment.
   2. ‘ml\_\*’ columns holds the summarized Email interaction status of each user to each ‘ml\_\*’ Email. The status can be: delivered, opened, etc.
   3. ‘approved\_at’: account approved date
   4. ‘first\_funded\_at’: account first funded date
   5. ‘first\_linked\_bank\_account\_at’: first linked bank account date
   6. ‘5d\_trading\_avg\_event\_count’: count of average trading events in the last 5 days
   7. ‘2d\_non\_trading\_avg\_event\_count’: count of average none trading events in the last 2 days
   8. ‘20d\_trading\_avg\_event\_count’: count of average trading events in the last 20 days
   9. ‘8d\_non\_trading\_avg\_event\_count’: count of average none trading events in the last 8 days
   10. ‘1d\_trading\_avg\_event\_count’: count of average trading events in the last day
   11. ‘1d\_non\_trading\_avg\_event\_count’: count of average none trading events in the last day
5. ‘control\_groups\_rate.csv’:
   1. Summarized table on the latest status for control groups who have not received any Emails.
   2. ‘num\_users\_in\_control’: Total number of users in each control group.
   3. ‘num\_funded\_in\_control‘: number of funded users by now in each control group.
   4. ‘funding\_rate\_in\_control’: funding rate in each control group.
   5. ‘num\_link\_in\_control’: number of users linked their accounts by now in each control group.
   6. ‘link\_rate\_in\_control’: link rate in each control group.

## Requirements for this project

Do your analysis on the data and try to answer these questions at least.

1. Email Open rate
   1. What is the best Email to send?
   2. What are the important user segment factors that correlate to Email open rate?
   3. (Optional question) Is there any friction (e.g. ‘unsubscribed’, etc.) effect from Email delivery?
2. Link and Funding rate
   1. What are link and funding rates for the treatment groups?
3. A/B testing on treatment and control groups
   1. Does sending Emails actually cause a higher funding rate?
4. Time series analysis
   1. Does Day\_0 always have the highest Email open rate? If not, do you have some suggestions on the strategy to send Emails?
   2. Based on our current data, do you think the experiment has completely finished?
5. Funnel analysis (optional)
   1. What does the ‘Email open -> account link -> account fund’ funnel look like?