Case Study - Email Data at Greenly

It is essential to understand the analysis that I am working on is aiming for realising Greenly's goal and vision. Since Greenly has a broad scope of clients, segmentation can be crucial. This analysis could identify patterns/trends in engagement among different segments of Greenly's user base and that could inform personalised email campaigns to increase effectiveness or possibly the revenue. I want to know how these email campaigns cater to different segments of Greenly's customers. These different email campaigns can be anything related to Greenly's service, for instance, tips for reducing emissions, client success stories, or new features in the carbon footprint assessment tool. Also, I would like to learn more about these email campaigns, whether they are primarily used for customer acquisition, retention, or something else.

Here are the steps I would take to tackle this project.

Step 1 (2-3 days):

I will start by understanding the requirements entirely. That includes having a good understanding of the KPIs and the filtering criteria.

In this step, I am going to have a discussion with the CRM managers about the analysis they want to have.

- What are the KPIs and filtering criteria mean? For instance, how to clarify unique clicks and unique opens or if they want to have the ability to filter by date range etc.
- Comparison or trend analysis or any specific analysis? For instance, are they interested in comparing different campaign performances? Or keeping track of a single campaign performance over time? Or are they interested in going into the performance of certain emails that are in one single campaign?

I am also having a discussion with the Data team about the data acquired from customer.io.

- I will go through the documentation of customer.io site and have an overview of the data type, data structure, and what data are collected. For instance, what is the structure and format of the raw datasets on BigQuery? How are the compatibilities between that and customer.io export format?

Finally, a last discussion with the BI Metabase Team.

- What are the visualisation that are possible to be presented on the platform? How to implement the desired features that was expected from CRM managers?

Of course, constant communication is essential throughout the entire project but the goal of this first step is to make sure that there are no misunderstandings or gaps in the requirements.

Step 2 (1-2 days):

In this step, I want to get myself familiar with the raw dataset in BigQuery. For each features that we have, a good understanding of the data characteristic is necessary. That includes assessing the quality of the data we have, the distribution of the data, and the relationship of each data elements. Some questions should be answered, for instance, what are the differences between email sent and delivered? Or the meanings of specific values in certain columns of the data.

Then, once I am done with exploring the raw data, I am going to prepare the data so that we have a clean and complete data for further calculation. Also, I am going to identify what the important features are in the data that in further steps are required for setting up the filters and calculating the KPIs.

Step 3 (1 day):

This is the step where all calculations of KPIs are done. First, we start by creating tables for different KPI calculations.

- Campaign-level analysis table: `campaign_id`, `date_sent`, `num_emails_sent`, `deliverability_rate`, `bounce_rate`, `open_rate`, `click_rate`, `reactivity_rate`, `unsubscribe_rate`. This table comes from the existing deliveries table and metrics table. In this table, you can have an overview of the KPIs in respect of the campaigns.
- Conversion table: `campaign_id`, `date_sent`, `num_emails_sent`, `num_emails_clicked`, `num_converted`. This table is going to track conversions over time. One question to be clarified is whether we want to have visibility on the conversion over time to see trends and patterns of the overall email marketing performance, or the conversion rate per campaign to provide insights about the effectiveness of individual ones by comparison. In this table, you can have an overview of the conversion rate in respect of the campaigns.
- A/B test table: `campaign_id`, `action_id`, `date_sent`, `num_emails_sent`, `deliverability_rate`, `bounce_rate`, `open_rate`, `click_rate`, `reactivity_rate`, `unsubscribe_rate`, `num_converted`. This table is derived from the existing deliveries table and the metrics table. In this table, you can have an overview of the KPIs and conversion in respect of action_id within the same campaign. It is also important to ensure that the results of the A/B tests are statistically significant and not due to random chance with the help of t-test.

I would write SQL queries in BigQuery to perform these calculations and summarise the data at a campaign level. For t-test, a python script can be useful to calculate the p-value and determine whether the results are significantly different or not.

Step 4 (2-3 days):

Once I have all the required new analysis tables, I am going to then proceed to build a dashboard in Metabase. The dashboard includes all visualisations for each KPI and the required filters to allow the CRM manager to slice and dice the data. When building the dashboard, I am going to have a lot of discussions with the CRM manager/dashboard user/data team/BI team to make sure that the dashboard actually is reflecting what they want to see for business insights, and incorporate their feedback into my work.

In this step, since the dashboard is built, it is also useful to look into how certain KPIs are indicative of the email campaigns' performances and what actions can be helpful to improve them. For instance, a low open rate might suggest that the subject is not engaging enough, a low click rate after opening might indicate that the email content is not clear or compelling, a high unsubscribe rate indicates that emails irrelevant or too frequent. These can be useful feedback for growth/marketing team to increase the effectiveness of email campaigns

Step 5 (1-2 days):

Document the entire process. I would put down the document to explain everything related to this project. This includes the steps and technicalities to prepare the data, the KPI definitions and how they are calculated, the dashboard setup, and detailed instructions on how to use the dashboard with use cases included. The documentation will then be shared with the CRM manager and other users for reference.

Also, what can be useful in further development is to explore additional data to enrich the data analysis. For instance, more demographic data on the recipients of the email deliveries, or client/company background of specific email campaigns so that I can make a visualisation that shows the correlation between engagement with those email campaigns and subsequent carbon offsetting.

Step 6:

A presentation to different teams relevant to summarise and wrap up the project. It’s also important to change the communication style when facing different audience. For instance, when presenting to the technical team, I am going to dive into details, and explain all the technicalities of the dashboard, the analysis methods, and all the technical improvements that can be made. When presenting to the growth/marketing team, I will focus more on the result of different KPIs that could affect the marketing strategy, as well as the trends of different email campaigns.

Here are some of the SQL codes I made for creating new analysis tables.

Campaign level analysis table: (The data is not completed but I tried it out anyway)

```
SELECT
  d.campaign_id,
  SUBSTRING(d.created_at, 1, 10) AS date_sent,
```

```
  COUNT(DISTINCT CASE WHEN m.metric = 'sent' THEN d.delivery_id END) AS num_emails_sent,
  COUNT(DISTINCT CASE WHEN m.metric = 'delivered' THEN d.delivery_id END) /
    NULLIF(COUNT(DISTINCT CASE WHEN m.metric = 'sent' THEN d.delivery_id END), 0) AS deliverability_rate,
  COUNT(DISTINCT CASE WHEN m.metric IN ('bounced') THEN d.delivery_id END) /
    NULLIF(COUNT(DISTINCT CASE WHEN m.metric = 'sent' THEN d.delivery_id END), 0) AS bounce_rate,
  COUNT(DISTINCT CASE WHEN m.metric = 'opened' THEN d.delivery_id END) /
    NULLIF(COUNT(DISTINCT CASE WHEN m.metric = 'delivered' THEN d.delivery_id END), 0) AS open_rate,
  COUNT(DISTINCT CASE WHEN m.metric = 'clicked' THEN d.delivery_id END) /
    NULLIF(COUNT(DISTINCT CASE WHEN m.metric = 'delivered' THEN d.delivery_id END), 0) AS click_rate,
  COUNT(DISTINCT CASE WHEN m.metric = 'clicked' THEN d.delivery_id END) /
    NULLIF(COUNT(DISTINCT CASE WHEN m.metric = 'opened' THEN d.delivery_id END), 0) AS reactivity_rate,
  COUNT(DISTINCT CASE WHEN m.metric = 'unsubscribed' THEN d.delivery_id END) /
    NULLIF(COUNT(DISTINCT CASE WHEN m.metric = 'sent' THEN d.delivery_id END), 0) AS unsubscribe_rate
FROM
  deliveries d
JOIN
  metrics m ON d.delivery_id = m.delivery_id
GROUP BY
  d.campaign_id,
  date_sent;
```

Conversion table:

```
SELECT
  d.campaign_id,
  SUBSTRING(d.created_at, 1, 10) AS date_sent,
  COUNT(DISTINCT CASE WHEN m.metric = 'sent' THEN d.delivery_id END) AS num_emails_sent,
  COUNT(DISTINCT CASE WHEN m.metric = 'clicked' THEN d.delivery_id END) AS num_emails_clicked,
  COUNT(DISTINCT CASE WHEN m.metric = 'converted' THEN d.delivery_id END) AS num_converted
FROM
  deliveries d
JOIN
  metrics m ON d.delivery_id = m.delivery_id
GROUP BY
  d.campaign_id,
  date_sent;
```

A/B test table: (I showcased with number of emails here. If we want to get the rates, we can simply calculate them.)

```
SELECT
  d.campaign_id,
  d.action_id AS action_id,
  d.delivery_id,
  SUBSTRING(d.created_at, 1, 10) AS date_sent,
  COUNT(DISTINCT CASE WHEN m.metric = 'sent' THEN d.delivery_id END) AS num_emails_sent,
  COUNT(DISTINCT CASE WHEN m.metric = 'delivered' THEN d.delivery_id END) AS num_delivered,
  COUNT(DISTINCT CASE WHEN m.metric = 'opened' THEN d.delivery_id END) AS num_opened,
  COUNT(DISTINCT CASE WHEN m.metric = 'clicked' THEN d.delivery_id END) AS num_clicked,
```

```
  COUNT(DISTINCT CASE WHEN m.metric = 'unsubscribed' THEN d.delivery_id END) AS num_unsubscribed,
  COUNT(DISTINCT CASE WHEN m.metric = 'converted' THEN d.delivery_id END) AS num_converted
FROM
  deliveries d
JOIN
  metrics m ON d.delivery_id = m.delivery_id
WHERE
  d.campaign_id IN (SELECT campaign_id FROM deliveries GROUP BY campaign_id HAVING COUNT(delivery_id) > 1)
GROUP BY
  d.campaign_id, action_id, date_sent;
```