Project Brief

Optimise Request for Feedback Schedules to Maximise Response Rates and Quality at SoPost

James Burton
SoPost
jburton@sopost.com

July 18, 2023

1 Background and Motivation

Over SoPost's 10 years of operation it has gathered a plethora of data regarding the performance of various sampling campaigns. For context, a typical campaign follows three stages:

1. Claim: Samples are being claimed by consumers.
2. Feedback: SoPost is gathering feedback from consumers regarding the sample.
3. Purchase: Emails are sent to drive the consumers to purchase a full product.

One of the key factors that sets SoPost apart from its competitors is the aforementioned wealth of information, knowledge and experience that allows SoPost to position itself, not just as a sampling platform, but as a sampling consultant.

For the purposes of this project, we are interested in data gathered surrounding the "Feedback" stage of a campaign. When comparing the performance of campaigns, one metric that gets used frequently is the Request for Feedback (RFF) response rate. For example if RFFs are sent to 1000 consumers and we receive 70 responses then the response rate would be 0.07 or 7%.

SoPost is interested in using its historic data to determine how factors such as day of the week affect the response rate of RFF emails. As an example, one might imagine that Mondays are a bad day to send emails since most people are busy dealing with tasks at the start of a new week or handling issues that arose over the weekend. Having insights such as this would allow SoPost to advise brands on the best options when configuring a campaign and further cement SoPost as "the sampling experts".

The Head of Innovation at SoPost has expressed particular interest in the outcome of this project and will serve as a key stakeholder throughout its duration.

2 Project Objectives

In this section we shall outline what we hope to achieve with this project by using the OKR framework and defining our objectives as measured by various key results.

O1: Determine the best day of the week for sending RFFs to maximise response rates.

- KR1: Identify relevant data from SoPost and any necessary external sources.
- KR2: Produce a dashboard in Looker Studio showing average response rates per day of the week and providing various filters.
- KR3: Train a predictive model which takes into account many factors at once such time between sample claim and RFF send.
- KR4: Produce a report for key stakeholders.

O2 (Stretch): Gain similar insights optimising for metrics other than response rate.

- KR4: Repeat analysis optimising for review sentiment.
- KR5a: Produce an appropriate method for measuring review quality.
- KR5b: Repeat analysis optimising for review quality.

3 The Data

This section defines the data requirements of this project. We have divided the discussion further into two, firstly discussing the internal data that we shall be using from SoPost and secondly any external data we shall require.

3.1 Internal Data

SoPost uses Google BigQuery as a data warehouse, this shall be our internal data source for this project. One significant advantage in taking our data from BigQuery is that we do not store any personal information in there meaning that GDPR concerns have already been handled for us.

There are two tables we shall be primarily focusing on. The first of theses tables is `scheduled_notifications` which contains the scheduled send times for both RFF and Request for Purchase (RFP) emails. The second table is the `review` table which which can be used to join on other tables to obtain various details about the product type. The schemas for these tables can be found in Figures 1 and 2.

Field	Type	Mode
count	INTEGER	NULLABLE
campaign_id	STRING	NULLABLE
activity_id	STRING	NULLABLE
event_type	STRING	NULLABLE
event_date	TIMESTAMP	NULLABLE

Figure 1: Schema of the `scheduled_notifications` table.

Field	Type	Mode
id	STRING	REQUIRED
order_id	STRING	REQUIRED
product_id	STRING	REQUIRED
activity_id	STRING	REQUIRED
campaign_id	STRING	NULLABLE
questions	JSON	NULLABLE
created	TIMESTAMP	NULLABLE
cluster	STRING	NULLABLE

Figure 2: Schema of the `review` table.

3.2 External Data

While SoPost can provide the majority of the data we will need for our analysis there are two issues which require the use of external data to solve. The first is that countries have public holidays which may affect response rates on the given days. Secondly, during 2020 and 2021 various countries experienced varying levels of lockdown in response to the COVID-19 pandemic.

We shall decide what to do about these dates later in the project but we must first obtain data to actually identify when those affected dates were. With respect to public holidays, Microsoft provides an open dataset [1]. As for lockdown data, a quick search shows that Kaggle provides the following dataset [2].

References

[1] Microsoft. Public holidays - azure open datasets — microsoft learn, 2022.

[2] user "jcyzag", K. Covid-19 lockdown dates by country — kaggle, 2020.