Email Marketing Analysis

Table of Contents

1. INTRODUCTION

The proper analysis of running campaigns for marketing is an important task to perform. It helps in decision-making and gives the direction of doing future campaigns. Generally, different companies use different approaches to get information about their previous campaigns. Our approach in this project will be to process, analyze, visualize, and plot insights of all the available email marketing interaction data to get suitable information about marketing campaigns. Furthermore, we will visualize different aspects or features of the datasets which will help us to run our future analysis and to make some decisions based on this visualization. Some basic queries will be analyzed and answered while doing the analysis.

The provided datasets comprise of huge data of up to 50 GB that contain about 50 CSV files, one GB each. Each file contains the following columns:

Campaign_id, campaign_content_link_url, interaction_contact, interaction_content_subject,

interaction_type, interaction_creation_utc_date_time, marketing_area.

All these columns play a key role in our analysis and it helps in visualization.

2. TOOLS USED

Several tools are available to deal with a large dataset. To deal with datasets like about 50 GB, GPU is preferable. To take access to GPU and analyze datasets, Kaggle is used in this project. It provides access to analyze datasets using Python and it is also good in terms of processing power. As mentioned, it provides GPU for computing which is far better than CPU. It was not possible to perform this analysis without GPU usage. On Kaggle, the CUDF library is used instead of Pandas because for GPUs, the Pandas is run on CUDA.

3. RESULTS AND DISCUSSION

In this section, all the questions related to email data analysis are answered using visualizations and suitable discussion about the query. The visualization is performed on all the available 50 files and the results are the combined influence of these files.

a. **What is the best time of day to send campaigns?**

The best time of the day to send campaigns will be the time when there will be a maximum chance of email open by the recipient. The best approach to answer this question is to check the best time when recipients open emails. The dataset is converted into the case where the recipient opens the email and combines the whole dataset. After analyzing and calculating the whole number of emails opened at each hour, we get the following analysis.

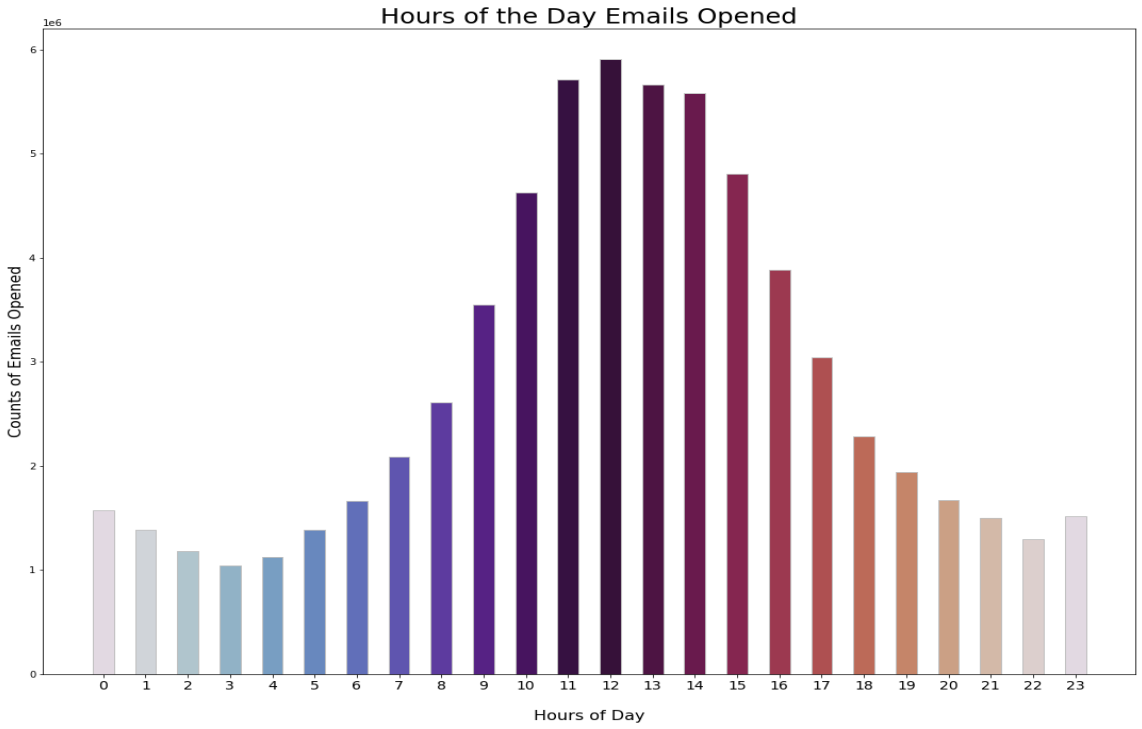

Figure 3.1 shows the time of the day when emails are opened by recipients.

The above figure shows the total number of emails opened by the recipient at different hours of the day. It also illustrates that around noon there is a peak of emails opened. The maximum number of emails opened at 1100, 1200, and 1300 hours. The least number of emails opened by the recipient is 300 at night. It is clear from the figure that the best time of the day to send the campaign is around noon. There is a maximum chance of the email being opened by the recipient at noon.

b. Can it be determined if certain/any of our list subsets are being hit too often?

To answer this question, the total number of emails sent to the recipients based on each marketing area is calculated. In this way, the proportion of emails being sent to each marketing area is visualized in the form of a pie chart. The pie chart is the best visualization here in this case because we get to know the proportion of lists subsets being hit. We get the following figure after analyzing the campaign datasets for emails being sent to each marketing area.

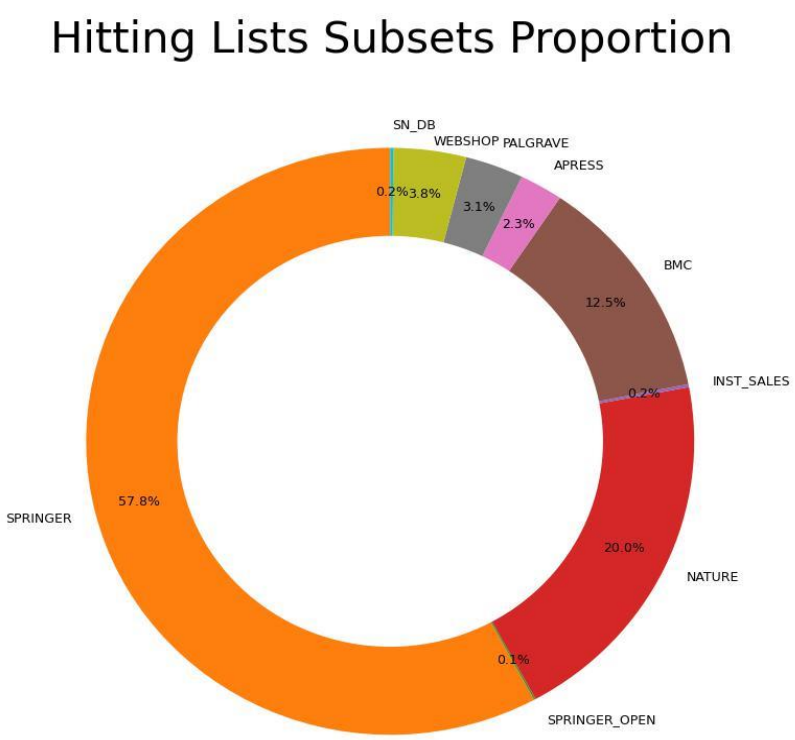

Figure 3.2 shows the proportion of emails being sent to recipients for each marketing area.

Figure 3.2 illustrates the number of emails sent to each marketing area in the form of percentages. It also depicts the highest number of emails sent to the recipients is of the marketing list named “Springer” (57.8%). The least area hit under this campaign is Springer Open (0.1%). It can be concluded that “Springer” is hit too often than all other marketing areas. Its percentage is even greater than the combined percentage of all other marketing areas. Contrary, “Springer Open” hit too less which can be considered negligible in front of “Springer” and even “NATURE”. “NATURE” is the second most hit area in campaigns.

c. How many emails on average does a recipient receive in a set period of time?

For calculating the average emails being sent to each recipient, the monthly time period is being selected. In this way, we get the different average values for each month in the available dataset. Firstly, the datasets were first converted to the email outbound status only. Because Out_bound shows the email being sent to the recipient. After that, the average email is sent by dividing the total number of emails by the total unique id in each month. We get the following figure.

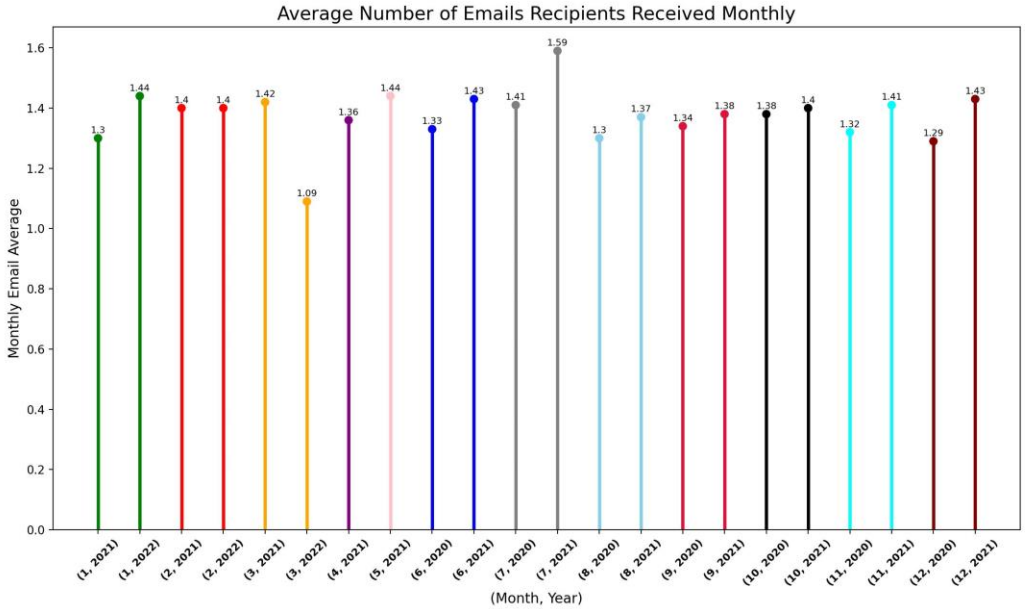

Figure 3.3 shows the average number of emails sent to each recipient each month.

Figures 3.3 depicts the average number of emails being sent to each recipient each month. The average number of emails sent to each recipient is 1.37. There is very little fluctuation of values in different months. It also illustrates that the average is almost uniform throughout the months. Furthermore, there is a little outlier at 07-2021 where the average value is 1.59. While the rest of the average for each month is not much different from others.

d. What is the frequency distribution?

To visualize the frequency distribution, the number of emails in each month is calculated and displayed. The time period set for this distribution is one month. Each month, the number of emails is calculated and it is different from others as shown in figure 3.4.

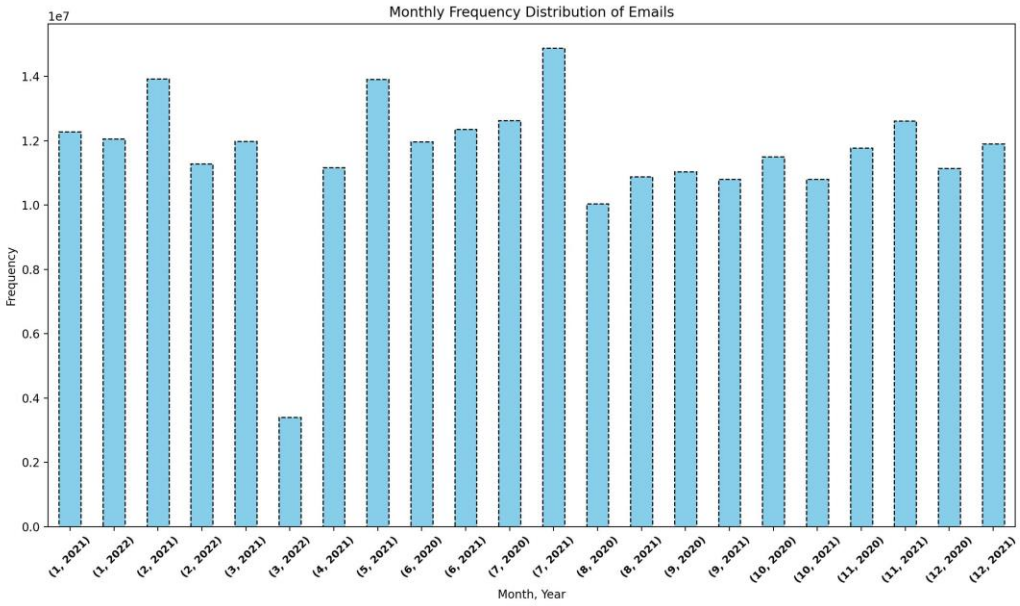

Figure 3.4 shows the frequency distribution based on monthly values.

Figure 3.4 illustrates the monthly frequency distribution. As mentioned earlier that the time period is one month and it shows different values after each month. The values are in the range of 10^7. The figure also shows that overall the distribution is uniform with the exception of the value in March 2022. The reason behind this is the less data availability in the dataset for March

2022. Apart from that, all the other values are nearly close to each other and show a normal distribution.

e. When are we bunching campaigns and therefore hitting lists too often?

All marketing areas are hit with emails while bunching campaigns. Some of the marketing areas might be hit too often than others. In this question, we have calculated the number of emails being sent to each marketing area on each day of the week. The total emails are then plotted against each day of the week as shown in figure 3.5. It will show on which day of the week, some areas are being hit too often than others.

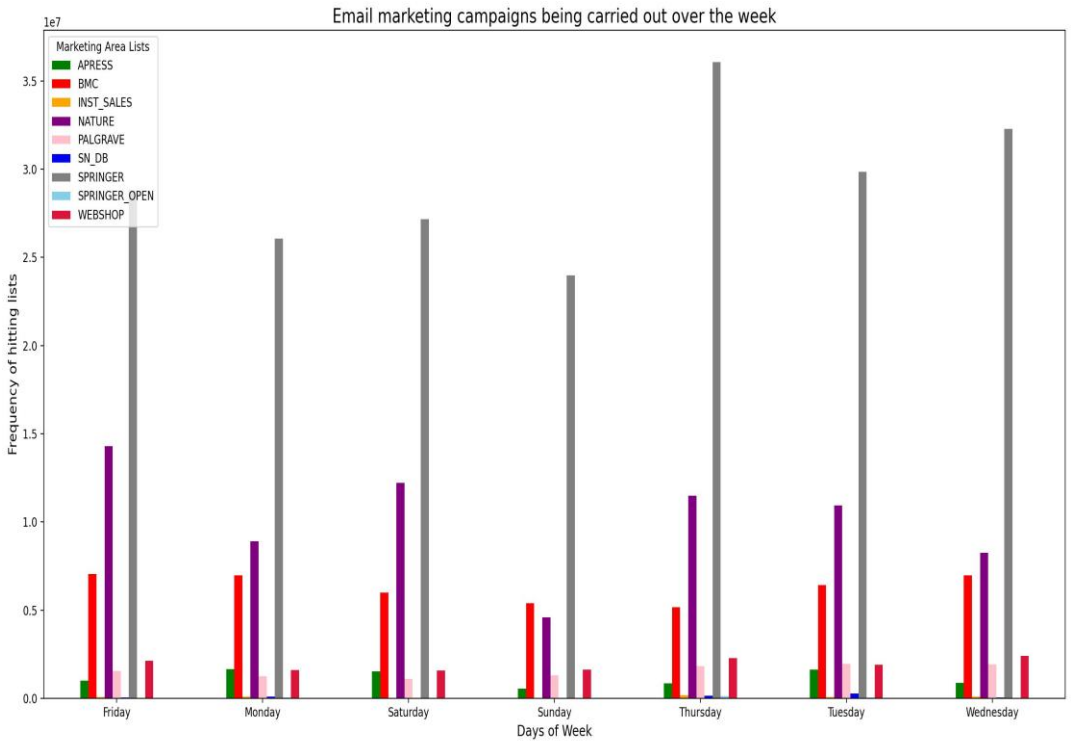

Figure 3.5 shows the marketing campaign of emails being carried out each day of the week.

The above figure shows the emails campaigns run each day of the week. It shows emails being sent to each marketing area on each day of the week. Different colors of the bars show different marketing areas. The figure also depicts that Springer hit too often than another marketing area. The Springer marketing area is being bunched a lot and therefore the recipient receives emails of this specific marketing area more frequently than others. It has also been discussed in the question above. Furthermore, the highest peak of emails is being recorded on Thursday. Some bars are missing which shows that the specific marketing areas were hit negligibly during the campaign. It can be concluded that when we were bunching campaigns on Thursday, "Springer" hit too often.

f. Are some people receiving many emails and others too few? If so, what is the driver behind this?

Some people might get more emails than others. It all depends on the several features of the datasets. The best approach to solve it is to check the total number of emails each individual's contact has received. We can see from the below table that around 40 percent of the recipients receive emails less than or equal to five. Whereas only 16.5 percent of the people receive emails greater than 15.

Percentage	Emails Received
39.64	<= 5
26.02	6 to 10
17.85	11 to 15
16.50	>15

g. **What is causing un-subscription from our email marketing lists?**

From figure 3.1 it is clear that the best time of the day for recipients to open their emails is around noon. It also illustrates that there is a maximum chance that the recipient will open the emails is noon but if the company runs a campaign or send an email at night or even early in the morning that there will be a maximum change of un-subscription. For instance, if a person gets an email at midnight he or she unsubscribes it.

The company takes care of special care about the timing of sending emails to each individual. If an individual is used to receiving an email at 1100 in the morning and the company sent an email at night then there will be a maximum chance that he un-subscribe the emails.

The un-subscription rate from the marketing campaigns is 1.95%.