How Do We Increase Positive Responses To Emails?

Dana Leenheer
Applied Computer Science
University of Colorado
Boulder, Colorado, United States
dana.leenheer@colorado.edu

Abstract

This project investigates the various ways of increasing positive responses to emails in B-to-B sales. A positive response includes the contact asking for a meeting, referring the email to a coworker, or a response asking for a follow up on a future date.

Sailes.com is an AI startup company that uses AI to automate the sales process so that enterprise sales teams can focus their energy on high value tasks. This investigation will investigate current data of positive responses to see what trends are present and explore ways to optimize campaigns for future success.

Problem Statement/Motivation

The current client onboarding process at Sailes includes utilizing ideal job titles either provided by the client, internal customer success team or operations team. This investigation aims to review the current success rates of job titles by various metrics including location, size of company, and day/week/month of the year. This job title and company data will be compared with external data focused on various target countries to determine if the current strategy is effectively covering the intended target population or if further refinements are needed (such as using general data on prevalence and density of job titles by region to better inform campaigns).

This investigation will answer questions like, what is the most common job title? Do job titles vary by region or company size? Do different industries have different varieties of job titles? What job title(s) respond positively to emails most often? What is the response rate between industries? What is the response rate between countries? Does email content or persona affect responses? Is there a variance in the day of the week or month for responses? What metrics affect positive responses the most?

Literature Survey (Previous Work)

Little research has been done on my topic, which partially led to the creation of this project. This project will help support efforts by providing past success and overall market information, going beyond the typical campaign by campaign approach. A campaign for the purpose of this investigation is a sales or business development campaign.

An overall review of positive response trends was completed by Clive Cadogan, Sailes CTO in 2022. He reviewed the overall trends of positive responses via various monthly metrics across all clients for Sailes [Cadogan 2022].

My role as a Data Operations Specialist at Sailes also lends insights to this project and helps direct the dataset selection. My understanding of the current internal data confirms that it is focused on the specific requirements of the client specifications for campaigns. Additional external data will be compared against the internal data to account for this.

Proposed Work

Data cleaning and preprocessing are the key first steps for my project. There are two main data sets for my project, one coming from QuickSight featuring data that has already gone through one round of cleaning prior to getting utilized by the API and another set of data that is currently a set of CSV files that I will be compiling.

The data in QuickSight will be pulled from the data lake that Sailes maintains using SQL and currently resides in 3 different internal tables (this method and the 3 associated tables were identified after meeting with one of the software engineers at Saile). These tables also include columns that go outside the scope of this project so they will be excluded from the data set created. This is partially why a custom query will be created, so that only the specified data columns will be added to the data set for this project vs. the entire table. I will review the data once it is pulled from the data lake to confirm whether cleaning is needed. I will also utilize Quicksight to mine the positive response rate per title and/or location. The details of this data set will be further discussed in my data set section.

The other main data set is a collection of CSV files from Apollo.io featuring 50k rows each of contact data. These files were exported from this software in August 2023. These files feature the same column order and format which simplifies the data cleaning and processing steps needed to assemble this project. The first step will be to combine each of the CSV files together, confirm that the columns do line up exactly, trim the white space, and remove the duplicate rows, both processes are common tools I routinely utilize in Google Sheets.

My approach to this project will focus on the characteristics of each contact, without focusing on the campaign they are associated with (beyond the general category of industry). Once the two data sets are ready for mining, I will review the applicable data mining algorithms and utilize the corresponding ones. One initial algorithm is the InfoGain algorithm with positive response, yes or no being the class under consideration. The CSV file data set lends itself to clustering to help form conclusions about the job title information.

Data Set
There are two main data sets for this investigation. The first data set is data that has not been uploaded to a campaign the second data set is being sourced from a data lake via QuickSight.

The first data set is a collection of 49 files of contact data sourced from Apollo.io that contain at most 50k lines of data in each file. This data was sourced in August 2023. Each of the files has identical column headers which will assist in the compiling of the files into one main file. The columns or attributes of these files are:

- Title
- # Employees (this is the number of employees at a respective contact's company)
- Email Domain (this is the email domain of the contact)
- Website (this is the website associated with the company of the contact)
- Industry
- City
- State
- Country
- Company City
- Company State
- Company Country
- Annual Revenue
- Apollo Contact Id (this attribute will only be kept while compiling the files and will be used to remove duplicate contacts/lines of data)

The second data set will be sourced from QuickSight and will also feature the above attributes (minus the Apollo Contact ID) but will also have:

- Contact ID
- Is_deliverable
- Email_response
- # of Email_response
- Date email was sent

The second dataset timeframe will be either the past 6 months or from January 2023 to October 31st of this year. This is due to staff and API structure being consistent and improved from pre-2023. Data sourced from 2021-2022 could still be valuable but 2023 will be the focus of this investigation. Both data sets will be kept on my personal computer.

Evaluation Methods

I will utilize the graphs and charts available in QuickSight to help gain understanding of the trends with the data. The timeline information will be displayed with a line chart to show positive response trends over time. An initial thought is to break this out by location as well. Athena also has a querying function that I will use to test my query on a sample set prior to querying the data lake. This gives me the ability to focus on the SQL query components first and then apply the specific data lake references for tables second.

I will utilize a word cloud to gain understanding of the most common job title and most common keywords used among the titles. My initial thought was to create my own word cloud with Python for the keywords but if that is outside the scope of this class my next option would be to find a word cloud website online and copy the words in.

Amazon QuickSight also gives me the ability to create various charts and graphs with the data so a separate python library or framework will not be needed for this project. Any fields I need to calculate (like the positive responses per location or job title) will be created with the custom calculated field option within QuickSight. I will also utilize the available functions within Excel as needed (ex. Count).

Tools

I will utilize Excel, Amazon Athena, Amazon QuickSight, VS Code (if needed), and Google Sheets. The data cleaning will take place within Excel and Google Sheets and be uploaded via CSV to QuickSight for evaluation.

Milestones

The benchmarks here are all listed with the intention of being completed prior to the time frame given.

- Data cleaning – complete within the 1-2 weeks and wrapping up over the fall break if needed
- Data integration – the csv files have been copied into a separate folder, compiling the files into one main file will be completing this week

- Data selection – the data in the data lake will be selected and organized this week
- Data transformation will take place shortly after the completion of the selection and integration, both QuickSight and Google Sheets have summary formulas that are user friendly.
- Data mining will take place shortly after the data transformation is completed.
- Pattern evaluation will take place shortly after the mining is completed, further mining may take place after the initial mining is completed.
- There will be a final report and presentation of my findings and recommendations.
- My GitHub account link will be added to my next project submission.

REFERENCES

[1] Clive Cadogan. 2022. ML to Targeting: Optimizing Positive Engagement. Kansas City, MO. (Confluence document).