How Do We Increase Positive Responses To Emails?

https://github.com/DLeenheer/CSPB_4502_project

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

ABSTRACT

This project investigates various ways of increasing positive responses to emails in business-to-business (B2B) 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 investigates internal and external data to see what trends are present and explore ways to optimize campaigns for future success. The specific focus of this project is a qualitative review of job titles within specific industries and locations to see what trends and information is present. This review took place with Excel software, largely utilizing pivot tables and summary functions. The data cleaning took place prior to the data set compilation and utilized various Excel and Google Sheet formulas.

Knowing which job titles tend to be most frequent within an industry will help the optimization of the contact sourcing process.

PROBLEM STATEMENT/MOTIVATION

Helping clients improve the effectiveness of their prospecting and general time management is the focus of Sailes. Freeing up the salesperson's time by handling much of their prospecting and assisting with market research is what Sailes offers and is accomplished through campaigns run by sailebots. Each sailebot is tailored to the personality and style of the salesperson they are assisting. This project aims to provide information to assist with the optimization of the campaigns, to ultimately increase the reach of the campaign that the salesperson would have had to do manually. Each sailebot does the work of 4 human salespeople [8]. The sales campaign's success is largely judged on the number and quality of positive responses to the emails sent out. The inputs provided to the sailebot drive the success, one of these inputs are the job titles entered.

A key focus of this project will be identifying the most common job titles for different criteria such as industry, job seniority level, and region/location. Given that the previously successful responses are heavily skewed by the focus of individual campaigns, target markets, target regions, and product or service sold by the client the previously successful email responses will be not reviewed in this project. Aligning a future campaign from a different company with the specifications of a previous company would not be successful or the most ideal strategy. Sometimes the campaign is focused on the senior level of the company, other times the focus would be on specific departments within the company based on the product or service sold. Given this the variety of specific job titles responding positively to emails would differ based on the targets sought after.

This investigation will try to answer questions like, what is the most common job title? Do job titles vary by region? Do jobs in the United States have different job title density than other countries? Do different industries have different varieties of job titles? What job titles are most frequent within different industries?

Are general job titles more popular than department specific titles (CEO vs. Director of Operations vs. Director of Engineering)?

INTRODUCTION

Data cleaning and preprocessing are the key first steps for my project. There are two main data sets for my project, one coming from Amazon Athena featuring the job titles, state, and country of a small sample grouping of contacts in the Sailes database and another set of data comprised from a set of CSV files.

The main external 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.

A second internal data set of 50k rows was exported from the Sailes database via Amazon Athena, using SQL. The attributes selected were title, state, and country. This data set was exported to provide a comparison to the external data set.

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). Placing the external compiled data set in Excel gave me the opportunity to utilize functions, find summary statistics like the average number of employees at a company for the data set overall, utilize pivot tables to drill down into multiple different aspects of the data. Pivot table filtering was an ideal efficient way to review the data.

A qualitative review of the job titles is the focus of this investigation. The previous successful email responses are useful information. However, discovering the trends within key industries of focus provides Sailes with valuable information to consider as part of the overall strategy of sales campaigns. The qualitative review is comprised of different drilled down views of the external dataset focusing on metrics of interest including location/region, industry, job seniority level, job title, and the number of employees at the company.

LITERATURE SURVEY (PREVIOUS WORK)

The literature review for this project focuses on a couple different areas including, sales prospecting strategies for B2B sales and artificial intelligence, the impact of email subject lines, job title assignment, job title lists for key industries, job title variance, and organizational structure (what job levels are present at a company, how many levels are there).

B2B prospecting is a key part of the sales strategy of any business selling a good or service to other businesses. There are different strategies for this type of sales compared to business to consumer selling. According to Zapier any B2B marketer needs purpose built B2B email marketing strategy to be successful [2]. The goal is to gain potential customers, build brand awareness, and generate sales [2]. Zapier gives 6 key strategies for email marketing 1) profile your target customers 2) think like a buyer 3) use email segmentation 4) plan your campaign in stages 5) use templates 6) save time by automating [2].

Sales emails are an important part of the sales cycle. HubSpot offers 150 different sales email subject lines with the goal of increased positive interaction with emails. The subject lines include separate sections for introduction emails and follow up emails [3]. With this much focus on the subject line alone of an email it's clear that trying to get the email to the correct person is key to success of the email campaign.

The next area that this literature review will touch on is the selection of job titles. Making sure that the email campaign is targeting the correct titles is a key component of a successful campaign and helps ensure that the ideal people are being reached. Indeed has an article that lists 230 job titles in 17 different industries [4]. The job titles vary within each industry which is key to consider when targeting an email campaign for a specific industry or group of industries. Interestingly, some of the industries listed could also be considered as departments. Business, customer service, engineering, finance, human resources, information

technology, leadership, marketing, operations, and sales are listed as separate industries when these categories could also be departments within the same company [4]. This is one example of why using keywords in addition to job titles might be a better strategy as even the definition of what an industry is can vary from time to time. Using keywords gives you a better chance to be inclusive of the ideal group of companies given variance in structure and assignment of job titles.

The variance in job titles and typical hierarchy is explained in the "Job Titles: The Definitive Guide" article on the Ongig website. It talks about the 6 major tiers of c-suite, vice president, director, manager, individual contributor, and entry level [5]. These tiers can be much more detailed and expanded for large companies. The U.S. Federal Government has 15 job grades with 10 steps each [5].

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 [1]. This review did not focus on the job title of the person, only what the result of the email correspondence was, positive or negative. Given that job title selection and entry is a key component of a successful sales email campaign this project focuses on the ways to optimize these inputs.

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 was compared against the internal data to account for this and to help identify areas of possible optimization.

DATA SET ETHICS

Catherine Cote defines data ethics as "the moral obligations of gathering, protecting, and using personally identifiable information and how it affects individuals" [6]. Cote also provides 5 principles of data ethics, ownership, transparency, privacy, intention, and outcomes [6]. A key step of this project was to de-identify [6] the data set. First name, last name, company name, company ID number, and any phone numbers were deleted from the individual 50k csv files prior to the data being added to the main external data set.

Commercial email falls under the CAN-SPAM Act [7] and B2B emails are commercial. Sailes emails follow the requirements of this law. The header information contains the first and either last name or last initial of the real person and company that the email is sent on behalf of, the subject lines clearly reflect the content of the message, the emails contain the physical address of the company that the email is sent on behalf of, the unsubscribe links are clearly shown, the unsubscribe links are processes through the API and recorded blocking future emails to the person contacted, and Sailes also conducts additional manual unsubscribe screening and processing [7].

DATA SETS

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 was sourced from Amazon Athena.

The first data set is a collection of 45 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 assisted in the compiling of the files into one main file. The selected columns or attributes of these files are:

- Title
- # of Employees (this is the number of employees at a respective contact's company)
- Seniority (this is the seniority level of the job title, ex. Senior, Vice President, Director, etc.)
- Industry
- State (of the employee)
- Country (of the employee)
- Company State
- Company Country
- Annual Revenue

- Apollo Contact Id (this attribute was kept while compiling the files, was used to find duplicate contacts/lines of data. After the duplicate rows were deleted this column, the contact ID column, was deleted)

The second data set was sourced from Amazon Athena. It is 50k lines of data with the following attributes.

- Title
- State (of the contact)
- Country (of the contact)

This second data set was created to compare with the external data set to see what similarities and differences there are between the two data sets. Both data sets will be kept on my personal computer.

TOOLS

I utilized Excel, Amazon Athena, and Google Sheets. The summary functions combined with my domain knowledge helped me find some interesting information and conclusions. Given the qualitative focus of this investigation the main tool used was Excel for the external data and Google Sheets for the internal data.

MAIN TECHNIQUES APPLIED

- Data cleaning – an initial cleaning of the individual external job title files took place in Google Sheets. This initial cleaning included deleting the columns of the files that go beyond the scope of this project or/and make the dataset harder to sort because the additional cells used for the data show down the computing speed of the Sheets and Excel software. The removed columns included data such as contact's first name, last name, company name, various phone numbers, and the company ID number.
- Data integration - The individual job titles documents have been combined into one main document in Excel. This process included the following steps, changing the file type from a CSV to a Google sheet file, removing the columns that go beyond the scope of the project, deleting all rows that feature an Employee number count of 0 (per my domain knowledge, these rows of data have been commonly inaccurate), changing the file type back to a CSV, emailing the CSV to my school email address, downloading the CSV on my school laptop, importing the CSV to a temporary Excel sheet, copying this data into the main dataset for the project.
- Data cleaning - Duplicate rows have been deleted from the external data set. There was a unique contact ID number column that each of the precompiled job titles had. The Excel 'Remove Duplicates' function was run on the entire dataset with the contact ID number field used as the comparison field. Once this operation was completed the contact ID number field was deleted to help reduce the total file size of the remaining dataset.
- Data cleaning – further cleaning of the job titles external data set took place in Excel. Some of the job titles include phrases like 'Operations Director at ABC Company'. I removed the 'at ABC Company' portion of the title so that the job titles could be grouped correctly when viewed in a filtered pivot table.
- A literature review of past work related to this project was conducted over the past week. This review was purposely inclusive of the different aspects of this project. The topics searched for included research done on B2B sales in general, B2B sales and artificial intelligence, job title creation, job title importance, data privacy issues, and business size classification.
- Data selection – 50000 contacts were pulled from the Sailes database with a SQL query using Amazon Athena, the fields exported were title, state, and country were selected and exported. Cleaning of this file then occurred along the same lines as the cleaning completed for the external data. The number of rows of this data set was selected to get a relative sample of the internal data at Sailes while also making sure that the file size was small enough to complete steps and functions in Googles

Sheets. I discovered early in the data cleaning process that Google Sheets was effective with smaller files, but when the data set increased to 75-100k lines of data the processing speed slowed down significantly and made task completion slow to a halt.

- Data transformation/mining – A pivot table was run on the external and internal data sets to evaluate multiple different aspects of the data. The table and the filtering ability was utilized in a variety of ways to uncover different drill down views of the data, discover trends, and make comparisons.

RESULTS

The first thing I explored once the external data set was fully compiled and cleaned was the average number of employees at the company of the contact and the industry layout of the data. By selecting the # Employees column I was able to find the overall average as it's given in the summarized data in the bottom right hand of the sheet. The average was 2367 people per company. This is in the ideal target range for many of the companies that Sailes works with. Figure 1 was created after using the pivot table functionality of Excel to list the industries in descending order. This breakdown informed the rest of the investigation. The top 5 most frequent industries in the job title dataset are information technology & services, hospital & healthcare, construction, banking, and insurance. Information technology & services is the most common industry in the external data set with 209579 people. Given that most of the industries in this top 10 group are also commonly requested for targeting by Sailes clients I decided to focus much of the investigation on this data set.

Before drilling more into the industry data, I decided to view the country of residence data of the contacts in the external data set. This was accomplished through a pivot table in Excel and is shown in Figure 2. My initial thought was that the contacts would be mostly based in the United States and while this ended up being the case the United States was in the majority with a much smaller margin than I anticipated. The companies in the external dataset are located in 210 countries. The top 5 countries are the United States, the United Kingdom, Canada, France, and Germany. The United States has the most companies in the dataset with 387,364 companies.

After the country detail I next drilled down into the job seniority data of the external data set. I again used a pivot table to isolate the job seniority data and sorted it in descending order. This data is found in Figure 3. The top 3 most common job seniority levels in the external data set are director, vice president, and head (ex. Head of Engineering). I then drilled down further by filtering to just the director seniority level, within the pivot table sheet. This data is found in Figure 4. Not surprisingly the top 3 most common industries were information technology, hospital & health care, and construction.

From here I decided to drill down into the job title data. First up was a look at the most common job titles overall in the external data set. This data is shown in Figure 5. The top five most common job titles across all industries in the data set are President, Vice President, Director, Managing Director, and CEO. My hypothesis is that they are general titles due to the variety of job titles found within specific industries.

I tested this hypothesis by drilling down into individual industry job title data with the pivot table. I looked at the most common job titles within the information technology industry, the hospital & health care, and the construction industries. The resulting top 20 data are shown in Tables 2, 3, and 4. Table 1 also provides a look at the top 20 most common job titles by the top 3 most common industries overall in the external data set. This view was created with a pivot table with additional filters.

I then decided to compare the top 10 most common job titles for contacts based in the United States of both the internal data set and external data set. Surprisingly the two groups did match. These data are found in Figure 6 and Figure 7. Vice President and Owner were the only two matches between the two groups.

The number of employees data grew in proportion to the number of rows of data for each job level of the external data set, larger companies had larger numbers

of data rows in the overall data set. This increase was also relatively even so further investigation of the number of employees metric was not conducted.

CONCLUSION/ APPLICATIONS

This project yielded useful information that will help inform the strategy for future client sales campaigns at Sailes. The top 10 most common industries of the data set are relevant to Sailes as they are commonly targeted by clients. Tables 2, 3, and 4 offer useful job title sets that can be considered for sailebot campaigns targeting the information technology, hospital & health care, or construction industries.

It is noteworthy that general job titles are common. Prior to this project I didn't realize that job titles like Vice President, Director, and Managing Director as stand-alone titles are common. I assumed that these titles always had department descriptors with them, ex. Vice President of Operations or Director of Marketing. I also didn't realize that a Chief Operating Officer is potentially more common than Chief Operations Officer, per Figure 5. Chief Operations Officer wasn't in the top 20 most common group overall or in the information technology or hospital & health care industries (per Table 1, Table 2, and Table 3). Operations is a common area of interest for many clients of Sailes so something as similar as including Chief Operating Officer vs Chief Operations Officer, or being sure to include both versions is a quick optimization that could have a big impact.

Given the job title information discovered in this project I recommend a strategy that includes both general job titles like Director and Vice President alongside department specific titles like Vice President Operations, Director of Nursing, or Chief Technology Officer.

REFERENCES

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

[2] Amanda Pell. 2022. B2B Email Marketing: Proven Strategies and Examples. DOI: B2B email marketing: Proven strategies + examples | Zapier https://zapier.com/blog/b2b-email-marketing/.

[3] Emma Brudner. 2023. 150 Sales Email Subject Lines That Get Opened, Read, and Responded To. DOI: https://blog.hubspot.com/sales/sales-email-subject-lines-that-get-prospects-to-open-read-and-respond.

[4] Indeed Editorial Team. 2023. 230 Job Titles in 17 Industries To Include on Your Resume. DOI: https://www.indeed.com/career-advice/resumes-cover-letters/job-title. Austin, TX.

[5] Job Titles: The Definitive Guide. DOI: https://www.ongig.com/job-titles#/. Oakland, CA.

[6] Catherine Cote. 2021. 5 Principles of Data Ethics For Business. DOI: https://online.hbs.edu/blog/post/data-ethics. Boston, MA.

[7] CAN-SPAM Act: A Compliance Guide for Business. DOI: https://www.ftc.gov/business-guidance/resources/can-spam-act-compliance-guide-business.

[8] Sailes. DOI: https://sailes.com/. Kansas City, MO.

Appendix A: Figures

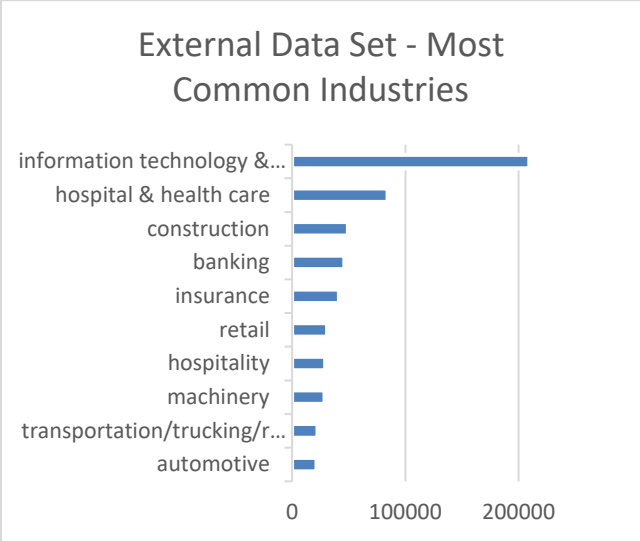

Figure 1: Bar chart of the most common industries of the data set.

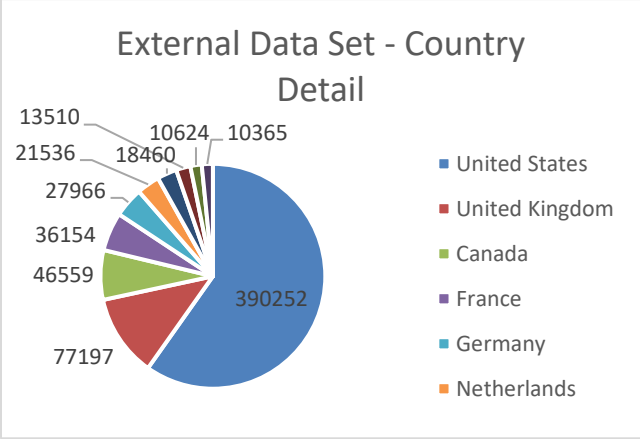

Figure 2: Pie chart showing the country of residence for the contacts in the external data set.

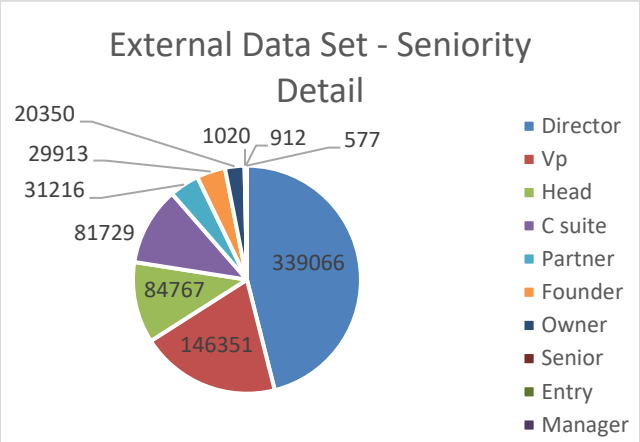

Figure 3: Pie chart showing the breakdown of job seniority level of the contacts in the external data set.

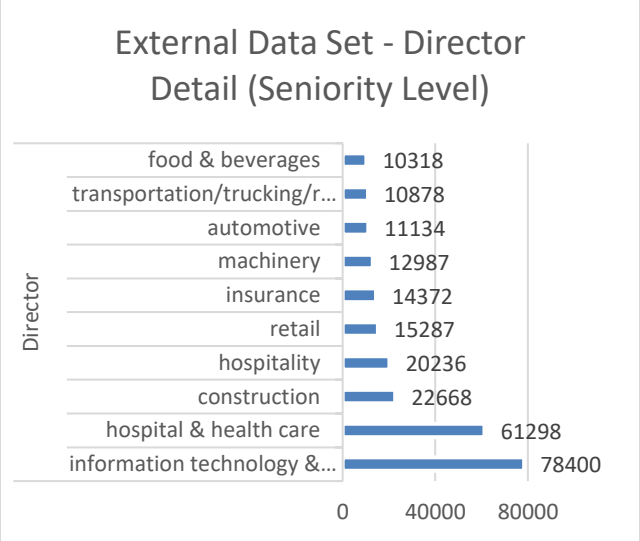

Figure 4: Bar chart showing the breakdown by industry of the contacts in the external data set with a job seniority level of Director.

Figure 5: Bar chart of the top 20 most common job titles overall of the external data set.

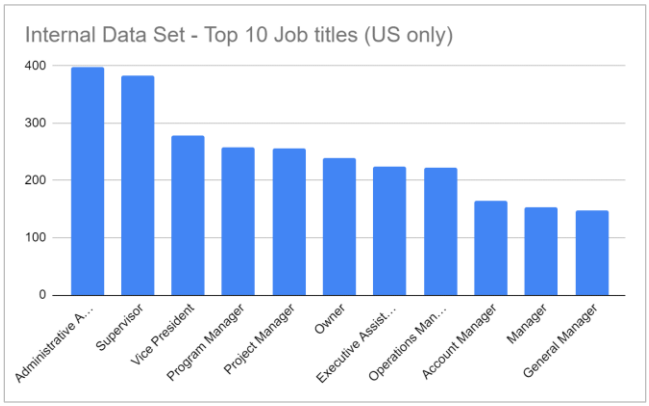

Figure 6: Bar chart of the top 10 most common job titles on the United States based contacts in the internal data set.

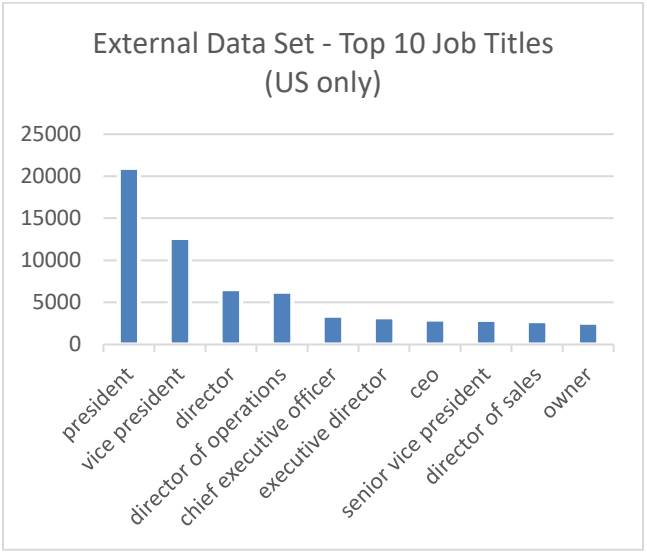

Figure 7: Bar chart of the top 10 most common job titles of the external data set.

Appendix B: Tables

Table 1: Table of the top 20 most common job titles with their quantities listed by industry of the external data set. The top 3 industries are shown in the table.

	construction	hospital & health care	information technology & services	total
president	6291	838	2972	**13470**
director	1918	1951	2989	**8146**
ceo	637	566	5343	**7782**
vice president	3151	489	1323	**7249**
managing director	1367	421	3024	**6841**
chief executive officer	443	831	3142	**5088**
director of operations	836	1044	1199	**3668**
chief operating officer	261	482	1906	**3023**
partner	607	239	1448	**2894**
owner	1021	147	1130	**2887**
executive director	162	2164	380	**2874**
sales director	254	172	1767	**2845**
chief technology officer	35	174	2428	**2814**
co-founder	67	130	2461	**2767**
product owner	20	57	2141	**2396**
chief financial officer	388	449	1163	**2334**
director of business development	347	382	835	**1772**
marketing director	307	342	671	**1585**
senior vice president	228	134	280	**1451**
director of sales	145	108	578	**1075**
Grand Total	18485	11120	37180	**82961**

Table 2: Table of the top 20 most common job titles within the hospital & health care industry in the external data set.

Hospital & Health Care	Total
Executive Director	2164
Director	1951
Director of Nursing	1288
Director of Operations	1044
Medical Director	966
President	838
Chief Executive Officer	831
Program Director	739
Clinical Director	574
CEO	566
Vice President	489
Chief Operating Officer	482
Director of Rehabilitation	450
Chief Financial Officer	449
Managing Director	421
Director of Pharmacy	411
Director of Business Development	382
Director of Finance	382
Assistant Director	360
Director of Clinical Services	352
Grand Total	15139

Table 3: Table of the top 20 most common job titles within the information technology & services industry in the external data set.

Information Technology & Services	Total
CEO	5343
Chief Executive Officer	3142
Managing Director	3024
Director	2989
President	2972
Co-Founder	2461
Chief Technology Officer	2428
Product Owner	2141
Chief Operating Officer	1906
Sales Director	1767
CTO	1517
Partner	1448
Founder	1422
Vice President	1323
Founder & CEO	1296
Director of Operations	1199
Chief Financial Officer	1163
Head of Marketing	1140
Owner	1130
Managing Partner	1108
Grand Total	40919

Table 4: Table of the top 20 most common job titles within the construction industry in the external data set.

Construction	Total
President	6291
Vice President	3151
Director	1918
Managing Director	1367
Owner	1021
Director of Operations	836
Project Director	708
CEO	637
Partner	607
Associate Director	590
Chief Executive Officer	443
Directeur	391
Chief Financial Officer	388
Executive Vice President	388
Vice President Operations	385
VP	369
Director of Business Development	347
Commercial Director	318
Marketing Director	307
Operations Director	307
Grand Total	20769