Analysis of the Reactions/Comments to a LinkedIn Post
CS7DS1, Trinity College Dublin

Abhinaw Tiwari
Trinity College Dublin
atiwari@tcd.ie

Amogh Anil Rao
Trinity College Dublin
raoam@tcd.ie

Asish Prasad Rath
Trinity College Dublin
arath@tcd.ie

Haneesh Solasa
Trinity College Dublin
solasah@tcd.ie

Harshal Patil
Trinity College Dublin
patilh@tcd.ie

Karan Dua
Trinity College Dublin
duak@tcd.ie

Ketan Anil Patil
Trinity College Dublin
patilk@tcd.ie

Vaibhav Srivastava
Trinity College Dublin
vsrivast@tcd.ie

Abstract

A LinkedIn post featured an anecdotal story regarding interactions between a boss and an HR manager. In response, a comment expressed skepticism about the story's validity and discussed the role of HR managers in prioritizing support for management over employees, even in cases where the employees have a valid argument. This project aims to analyze the responses to this post, identify recurring themes and differing opinions, and understand the professional community's overall sentiment regarding the role of HR management in assisting employees and their relationship with management.

1 Introduction

Human resource management is an essential component of any organization, responsible for managing employee-related matters and ensuring a healthy and positive work environment. The balance between employee advocacy and management support in HR management is a constant topic of discussion in the professional community. This paper aims to analyze the responses to a professional's comment on LinkedIn, expressing skepticism about the validity of an anecdotal story and discussing the role of HR managers in prioritizing support for management over employees.

The paper follows a structured approach with a clear methodology. Section 2 outlines the data scraping journey and the sentiment analysis performed on the scraped data. Section 3 and 4 describe the explanatory data analysis, including the use of different visualization tools to extract valuable insights and trends to understand the underlying data. In the discussion section, we present our interpretation of the analysis conducted in sections 3 and 4. Finally, we discuss future work that could be carried out in this area.

2 Methodology

This section discusses the methodologies used for the analysis of the comments and reactions.

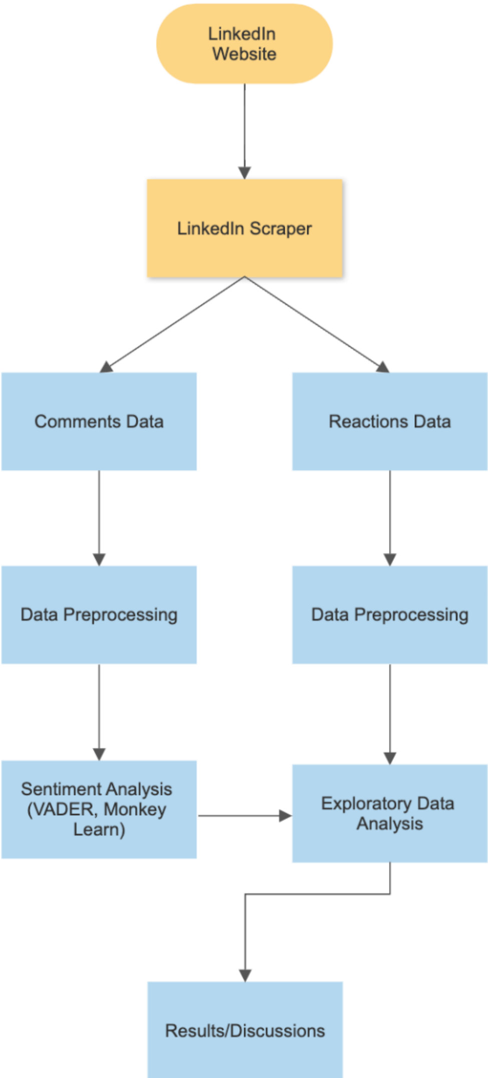

Figure 1: Project Flowchart

2.1 Data Scraping using Selenium:

The first step was to scrape the replies to professional's comment and the corresponding URL to the profile. Selenium and Beautiful Soup were used to get the information. Selenium WebDriver library automates the web browser actions while the BeautifulSoup is used to parse the HTML contents of the web page.

Navigate to the post:

Once all the required libraries are imported, the Selenium WebDriver is used to navigate to the URL of the post.

Load Comments and replies:

In its original state, LinkedIn displays only a limited number of comments and their replies, with additional comments and replies being hidden and requiring user action such as clicking a "Show Previous Replies" button. To do this, the button is located using its class name and waiting for a few seconds to allow the replies to load. This process is repeated until all the replies are loaded.

LinkedIn typically displays only a truncated version of a comment by default and the users need to click on a "See More" button to expand and view the entire comment. So, this button is located using its class name and waiting for a few seconds to allow the comments to load. This process is repeated until all the comments have been loaded.

Extract Comments and replies:

Once everything is loaded, BeautifulSoup was used to parse the HTML content of the LinkedIn post webpage. All the comment objects were located using their HTML tags and appended to a list. Similarly, all the replies to the comment are also appended to a list. The replies list was looped through to find all the replies to objects and from each object, Profile URL, Username and Comment Text were extracted.

Now that the scraped data contains the Profile URL, the next step was to get the information such as About, Experience, Education, Interest etc. about the profile. There are a few LinkedIn official APIs with which this data can be scraped. If there is no developer access, which is difficult to get, there is a chance that the account can be blocked by LinkedIn. Therefore, it was decided not to proceed with these APIs.

2.2 Data Scraping using linkedin_scraper

Linkedin_scraper is a python library that can be used to scrape LinkedIn pages. It provides an interface to extract LinkedIn profiles and posts through various filtering parameters. The "Person" class is used to extract information from a LinkedIn profile page, and it allows users to specify which data they want to retrieve by passing parameters such as About, Experience etc.

The output from the scraper was found to be incorrect because the LinkedIn profiles accessed were not within the user's contacts. As a result, the scraper retrieved information about random individuals from user's contacts who were not part of the intended group, which led to inaccurate output. Because of this, it was decided not to proceed with linkedin_scraper.

2.3 Data Scraping using Phantom Buster

Phantom Buster is a cloud automation platform which provides a quick and effortless way to get data from several web sources, including LinkedIn. With its LinkedIn Profile Scraper tool, the URLs of the profile which were extracted during [2.1] were given as an input. Relevant information such as job titles, education, skills, and more were extracted using this and got into an excel file. The limitations of the service were that it was not a free service and only a maximum of 80 comments could be retrieved per day. As a result, the data collection took a few days to complete and some of the redundant comments had to be removed.

To consolidate data extracted from Selenium and Phantom buster, we used VLOOKUP in Excel to combine both the files. This allowed us to match the data from each file based on a unique identifier Profile URL. A more comprehensive and accurate picture of the information about the target profile was obtained after combining both the files.

2.4 Sentiment Analysis

For getting the sentiments of the comments, they were fed as an input to VADER which is a rule-based sentiment analysis tool. It is specifically designed for social media text.

Monkey Learn: The platform provides an interface to perform sentiment analysis on CSV files. Once the file is uploaded, a few comments were to be manually specified if they were neutral, positive, or negative and the service provides the sentiments of the other comments based on the inputs that were given.

Vader is used for general sentiment analysis, but we want to perform sentiment analysis according to the comment to check whether the reply is in favor of the comment or against it or neutral. Due to this issue, we chose monkey learn in which we manually classified some of the comments and it uses those inputs to train the model which provides us with more accurate analysis.

The final sentiments for the replies for the professional's comments are 63 negative, 14 positive and 11 neutral.

2.5 Scraping & Analyzing Reactions Data

Once the comments were all analyzed, the data about the reactions was scraped. This was done to find out if there was a correlation between the comments and reactions. Also, the correlation between the people who reacted and those who commented on the post could be analyzed. The html code for the comment is copied to a file and created a reactions.html file from which the data is scraped.

BeautifulSoup library is used to parse the HTML file and all the <li> tags with the class for the reactions are found and they are stored. Information such as name, job title, reaction and profile URL are extracted by looping over all the <li> tags and stored inside a data frame.

The obtained reactions data is merged with the final comments data using an inner join to determine how many people who reacted also commented. The results of them are below:

- No one who reacted Funny or Love to the comment, replied to it.
- No one who reacted Curious or Insightful to the comment, replied to it.
- Only one person who reacted Celebrate to the comment replied to it.
- 15 people who liked the comment replied to it.

3. Exploratory Data Analysis

Tableau/Power BI, Matplotlib library in Python are interactive platforms for creating plots, reports and dashboards using diverse sources like Excel, SQL etc. They simplify the transformation of raw data into meaningful information to make informed business decisions.

These tools were used to perform Exploratory Data Analysis and get relevant visualizations. These visualizations in turn helped to extract valuable information and insights for discovering underlying patterns and trends.

Figure 2: Most common words in comments

From figure 2, we can see the most common words in the comments. This word map includes the professional's name which shows that most of the comments are addressed to him and not a general discussion.

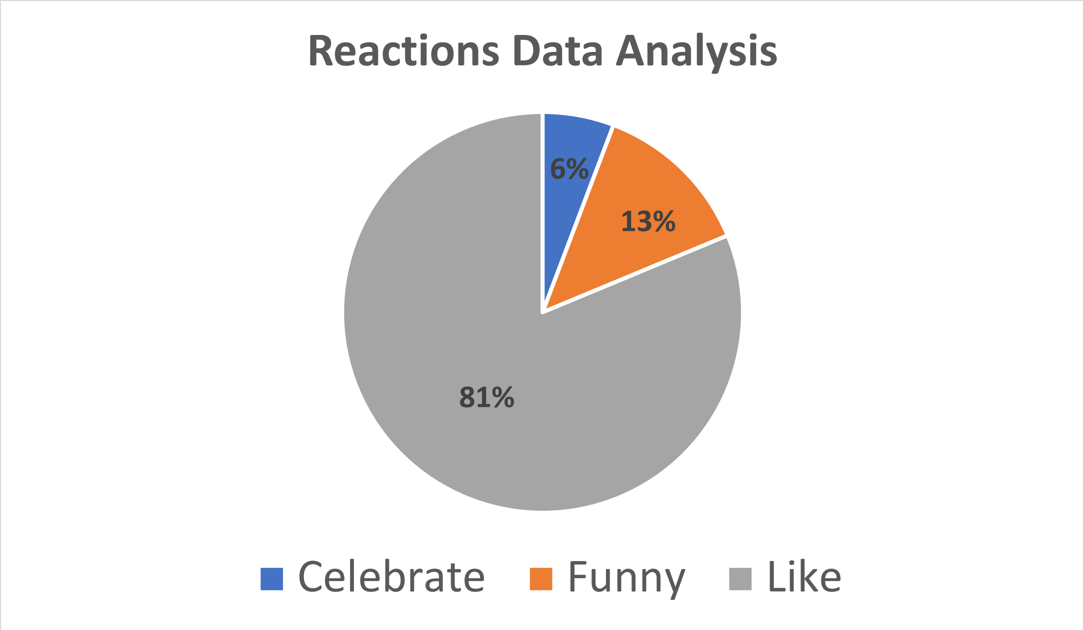

Figure 3

Figure 3 shows the classification of the reactions. For the reactions "Support", "Curious", "Insightful" and "Love" have less than 1%.

4. Results

After completing Exploratory Data Analysis, the different visualizations obtained are as follows.

4.1 Comments by Geographic Region:

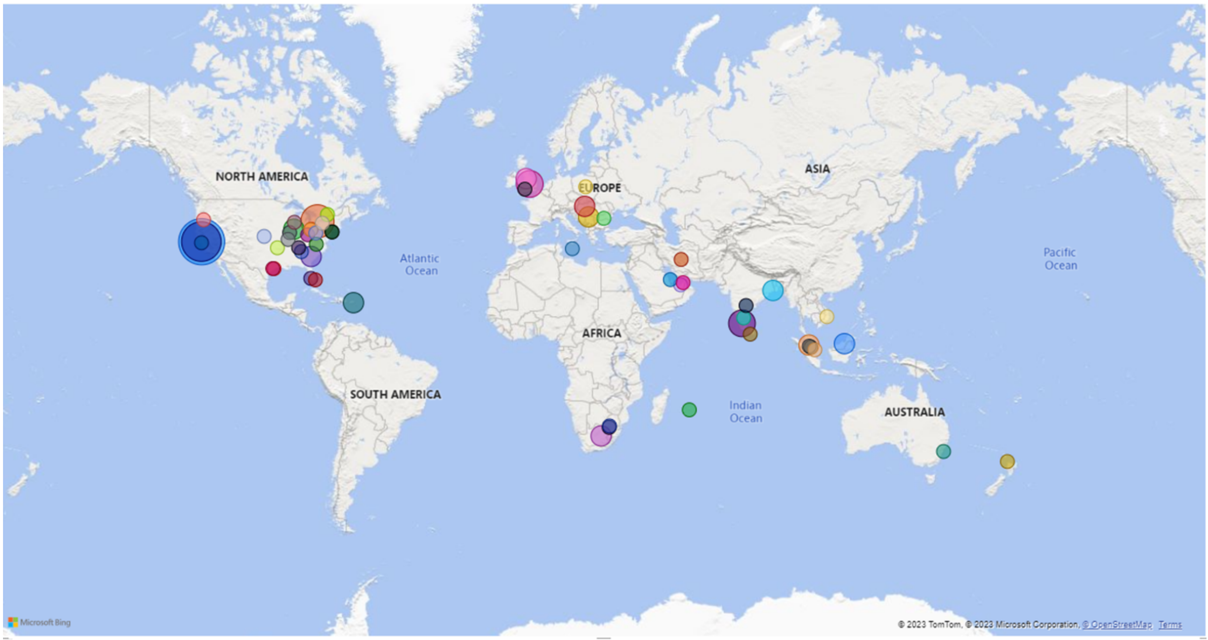

Figure 4: Comments by Geographic Region

This plot illustrates the geographic locations of individuals who responded to the comment on the problem statement. The size of the bubbles represents the number of respondents, with larger bubbles indicating a higher number of responses and smaller bubbles signifying fewer responses. As observed, most respondents are from the USA, which is likely because the original post was primarily shared by a person from the USA. Figure 4 shows the distribution of comments across different geographic regions.

This analysis was further refined to identify the specific regions from which the comments are concentrated in. Figure 5 gives us the regions with the highest comment count.

Figure 5: Regions with Highest comment count

4.2 Department-wise Employee Distribution:

This pie chart displays the percentage distribution of individuals from different departments and seniority levels. The chart reveals that approximately 43% of the respondents hold senior-level positions, providing a clear visualization of the representation of each department and seniority level.

As shown in Figure 6, management and senior-level employees contributed the highest number of comments on the post, followed by the education department.

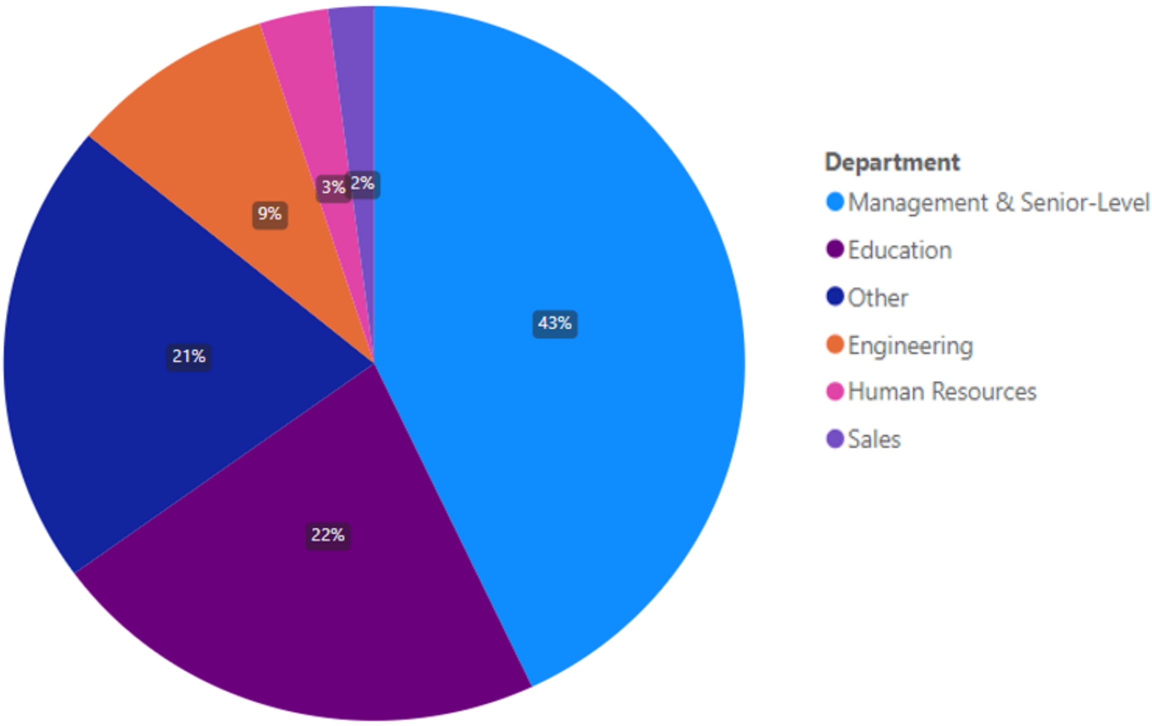

Figure 6: Department-wise Employee Distribution

The job titles of the commenters is analysed using a word map. Larger words indicate a more frequent occurrence which enables for a quick understanding of the common job titles among the commenters. Fig. 7 gives a word map in which we can see manager and engineer are the most common job titles across the commenters which are followed by directors and business representatives.

Figure 7: Word Map of the Job Titles

Integrating the analyses, we arrive at Figure 8, which illustrates the job titles of the commenters originating from a particular region. It can be observed from the figure that there is not one job which is common across different regions.

Figure 8: Regional Breakdown of Commenters by Job Title

4.3 Sentiments of Comments by Region

Figure 9 depicts the sentiments of different commentors across different regions in the world. Most of the negative comments are from the United States of America while there is no one region concentrated for positive and negative comments.

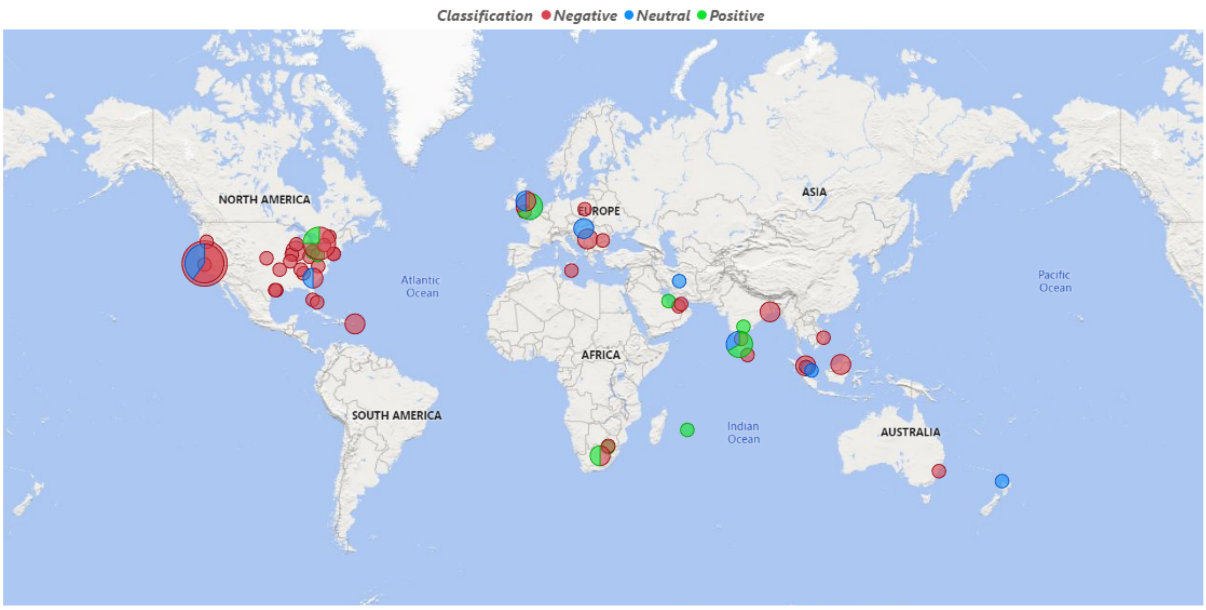

Figure 9: Sentiments of Comments by Region

4.4 Sentiment Analysis by Department

Upon the sentiment analysis's completion, the comments are classified into negative, positive, and neutral groups. To gain a comprehensive understanding of the responses from each department, these sentiment categories are further subdivided by department. We used job title of the respondents to classify them into different departments.

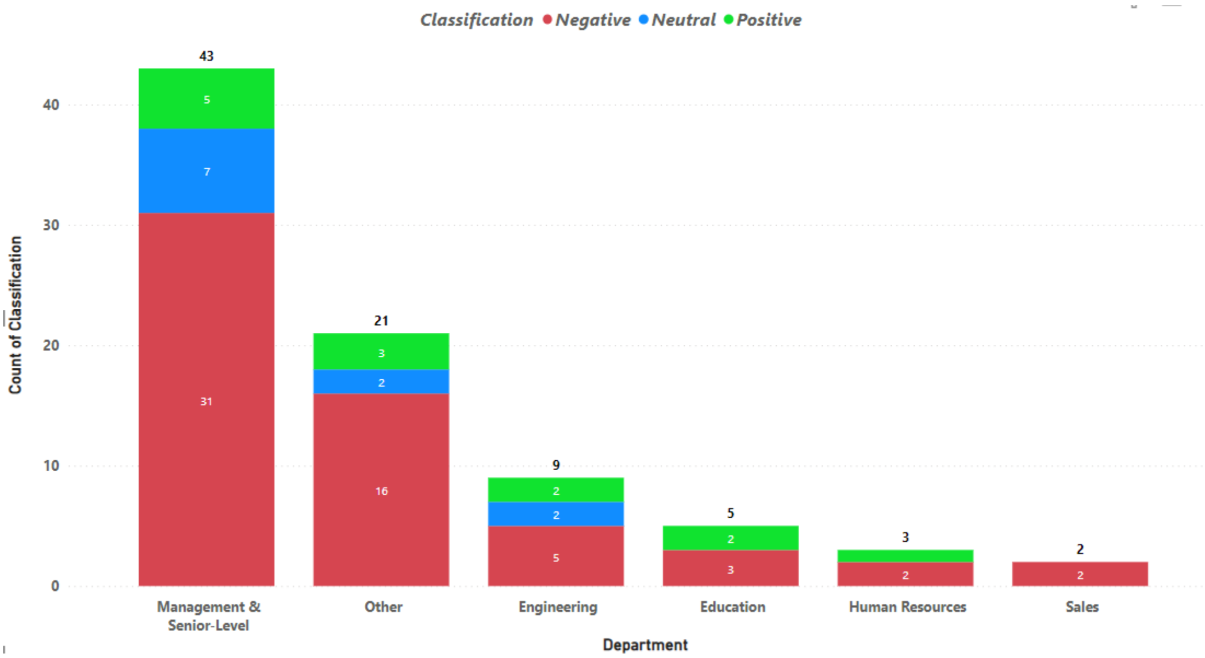

Figure 10: Classification of Sentiments over different Departments

Figure 10 depicts the number of people from various departments who have commented on the LinkedIn post and comment had an overall negative, positive, and neutral sentiments after the sentiment analysis.

It is clear that there were more negative sentiment comments from individuals in the management and education departments compared to other departments. However, in departments such as Engineering and Education, there were roughly equal numbers of non-negative and negative comments.

4.5 Reactions wise analysis by Department

Analyzing the distribution of the reactions: 'Like', 'Celebrate', 'Support', 'Love', 'Insightful' and 'Curious' can offer valuable insights into the engagement patterns of professionals from various fields. This helps to explore the relation between departments and types of reactions received.

The counts for "Curious", "Love" and "Support", "Insightful" reactions are low, with 1,3, 4 and 2 instances, respectively. As a result, visualizing these reactions in a graph may not provide substantial information. Therefore, focusing on the rest of the reactions for the analysis.

Figure 11 depicts the number of people from various departments that have reacted to the LinkedIn post with Funny Reaction. It is evident that the individuals from the engineering department exhibit highest number of funny reactions.

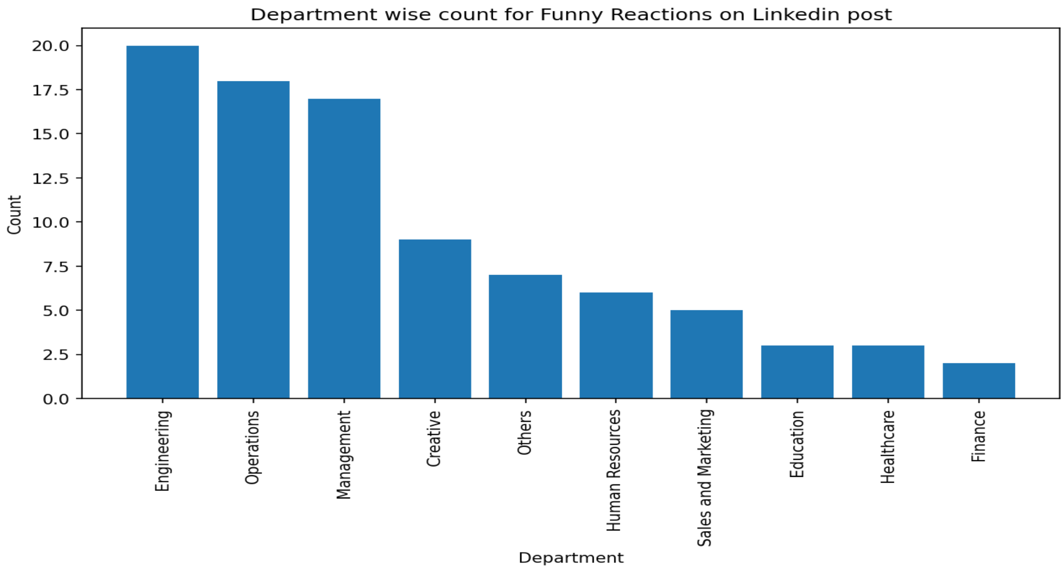

Figure 11: Funny Reactions by Department

Figure 12 depicts the number of people from various departments that have reacted to the LinkedIn post with Celebrate Reaction. It is evident that the individuals from the engineering department exhibit highest number of celebrate reactions.

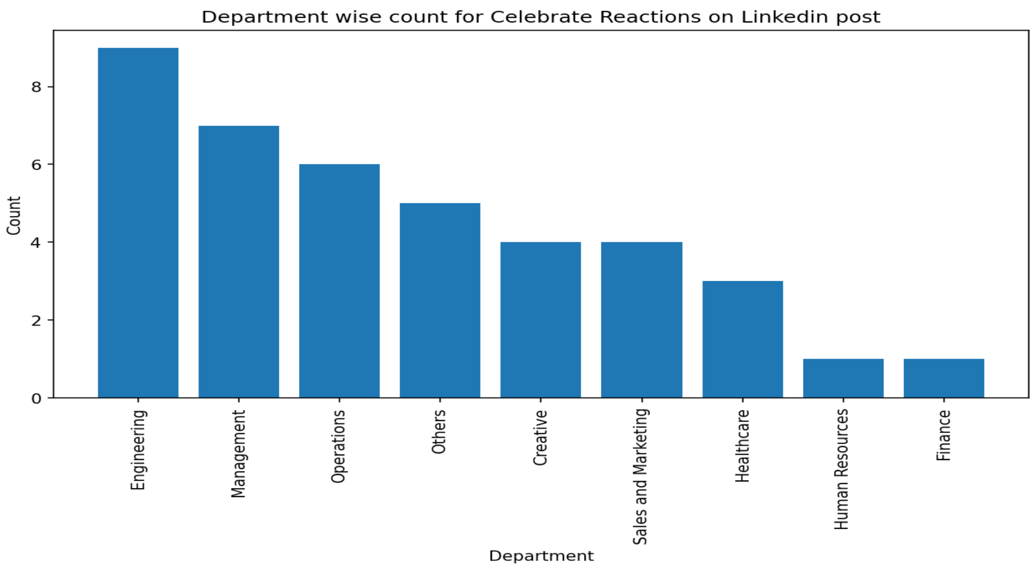

Figure 12: Celebrate Reactions by Department

Figure 13 depicts the number of people from various departments that have reacted to the LinkedIn post with Like reaction. It is evident that individuals from the engineering department highest number of Like reactions.

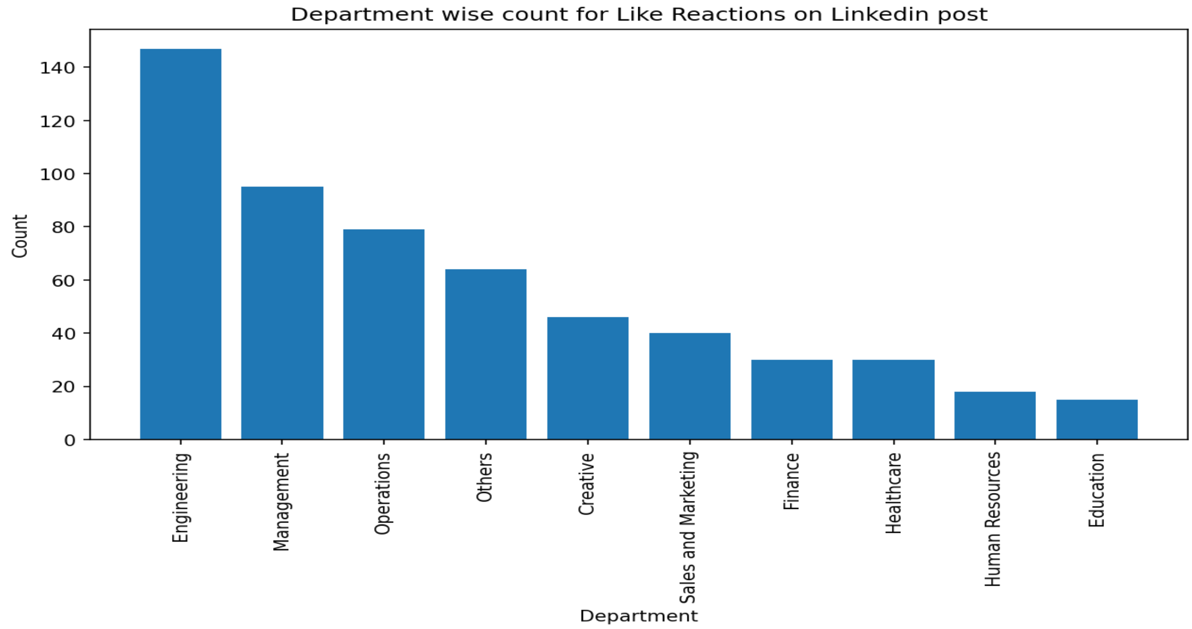

Figure 13: Like Reactions by Department

5. Discussions

Based on the above analysis, it can be inferred that the management department appeared to disagree with the professional's views, while the engineering department seemed to support him. The main purpose of conducting this analysis is to examine the reactions received and evaluate the sentiment expressed by professionals from different departments.

Our findings reveal that the reactions from Engineering and Education Departments professionals have as many as comments with non-negative sentiments as the negative ones. Conversely, the Management & Senior Level professionals reacted with the highest number of comments expressing negative sentiment. This observation is coherent to the professional's comment as the comment challenges the perspective of an HR manager while advocating for engineers.

The analysis of the comment reactions indicates that the Engineering department had the highest number of reactions among all departments. The high number of "celebrate" reactions suggests that the Engineering department strongly agrees with the professionals' comment, which supports their work and is critical of HR management. This is followed by the management profession because there might have been a logical argument which they accepted as well. The significant number of "celebrate" and humorous reactions to the professional's comment indicate that the Engineering and Operations departments strongly supports the professional's criticism of HR management and find it amusing. Since the “like” reason is a generic one, it can be used for a variety of reasons. It is difficult to determine whether the person truly liked the comment or was simply using it as a gesture. Therefore, it may not be appropriate to draw any compulsion solely based on the number of those reactions.

6. Further Work

There were limited instances of individuals who both commented and reacted. A larger sample of individuals who did both would enable further analysis to examine any potential correlation between the reaction and the comment. Furthermore, to ensure the validity of the analysis and mitigate potential bias, future work may include initial comment from a diverse group of professionals. This would allow for an exploration of potential correlations between the profession of the initial commenter and the sentiment of the replies.

7. References

[1] Honari, B. (2023, Feb). Comment sharing insight. LinkedIn. https://tinyurl.com/ycy7zy9d
[2] MonkeyLearn, "monkeylearn-python: Official Python client for the MonkeyLearn API," GitHub, 2023. [Online]. Available: https://github.com/monkeylearn/monkeylearn-python.
[3] Phantombuster, "Phantombuster: Data Extraction & Automation Platform," 2023. [Online]. Available: https://phantombuster.com/?deal=leesa57.
[4] JustAnotherArchivist, "snscrape: A Social Networking Service scraper in Python," GitHub, 2023. [Online]. Available: https://github.com/JustAnotherArchivist/snscrape.
[5] B. Liu, "Sentiment Analysis and Opinion Mining," University of Illinois at Chicago, Tech. Rep., 2012. [Online]. Available: https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf.