Exploratory Data Analysis of LinkedIn Job Postings and Stock Price Trends

Carnegie Mellon University
Data Visualization
Course ID: 36-613

Authors:
Perry Lin
Yahan Yang
Karan Karthik

Table of Contents

Abstract

This study explores the relationship between industry demand, company popularity, and stock performance, focusing on how these factors influence job posting behavior. Using LinkedIn job postings and daily stock price data, we first analyzed trends across industries and regions to identify sectors with the highest demand for talent. Next, we evaluated the stock performance of top-hiring companies by examining their cumulative returns and year-over-year (YoY) growth. Finally, we investigated the impact of company popularity, measured through employee count and social media engagement, on job posting trends using linear regression.

Our analysis revealed that sectors like Hospitals and Health Care have distributed hiring patterns, while industries such as Industrial Machinery Manufacturing are dominated by a few key employers. Stock performance analysis showed companies like Amazon and Meta demonstrated strong cumulative returns. Additionally, the regression analysis found that employee count is positively correlated with social media engagement, though it explains only a small portion of job posting activity, suggesting other factors, such as industry type and hiring strategies, play significant roles.

These findings highlight the complex interplay between company size, industry characteristics, and recruitment strategies. Future work could incorporate variables like salary data and economic trends to further understand how these factors shape the job market.

Introduction

The relationship between the job market and stock market performance is a key area of interest for economists, investors, and business leaders. Changes in employment trends—such as job demand, hiring practices, and salary data—often reflect broader economic conditions and can significantly impact financial markets. Understanding these dynamics is crucial for corporate and individual investors, as well as decision-makers involved in hiring and growth strategies.

In this study, we aim to assess the current state of the job market by identifying the industries with the highest demand for job postings and the companies that dominate hiring within these sectors. By examining job postings at both industry and company levels, we will highlight the leading companies and their share of job postings, providing a clearer understanding of the sectors driving job market demand.

Next, we will analyze the stock market performance of these top-hiring companies, focusing on their cumulative returns and year-over-year (YoY) growth. This will allow us to assess how these leading employers compare to one another in terms of financial growth, offering insights into their market success.

Finally, we will explore how company popularity—measured through factors like employee count and social media engagement—affects job posting behavior. Using regression analysis, we aim to determine whether companies with larger workforces and greater online visibility post more job openings, providing insights into how company popularity influences recruitment strategies.

Data

LinkedIn Dataset

Data for this analysis report can be found in a comprehensive record of job listings from 2023 to 2024. Each job posting contains various attributes such as title, job description, salary, location, application URL, and work type (remote, contract, etc.). Additional datasets include company-specific details like benefits, skills, and industries associated with each posting. Other files include the mapping between company id, company name, industry id, industry name, salaries, skills, etc.

For the purposes of our analysis, we first prepare the input data by merging the raw data on respective ids to produce two combined files. In the following analysis, we mainly focus on using two merged data sets: company information data set and job information data set.

Company information data: Variables include company id, company name, company size, state, country, city, zip code, address, url, speciality, industry, employee count, follower count, and time recorded. After merging and cleaning, we have a total of 17781 observations in this data set.

Job information data: Variables include job id, skill name, benefit type, salary id, max salary, med salary, min salary, pay period, currency type, industry id, and industry name. We have 11644 observations.

Stock Price Dataset

We focused on publicly traded companies identified as the top 3 performers in terms of job postings across various industries. These industries included Financial Services, IT Services and Consulting, Hospitals and Health Care, Retail, Industrial Machinery Manufacturing, and Software Development. The selected companies were:

- Financial Services: Wells Fargo (WFC), Capital One (COF)
- IT Services and Consulting: Verizon (VZ)
- Retail: Macy's (M), Lowe's Companies Inc. (LOW)
- Industrial Machinery Manufacturing: Johnson Controls (JCI), SKF Group (SKFRY), Ingersoll Rand (IR)
- Software Development: Amazon (AMZN), Google (GOOG), Meta (META)

These companies were chosen based on their job postings within each industry, and we ensured that only publicly traded companies were included, as some top employers were privately held.

Methods

Data cleaning and preparation

The original data consists of many '.csv' files, which makes it hard to perform data analysis. For the purpose of creating graphs to investigate the dataset, we started cleaning the raw data we collected before EDA.

LinkedIn Dataset

First, we combined the job skills and benefits data. Job skills were merged with relevant data to provide descriptive skill names, and these were grouped by job, consolidating all skills into a single field for each posting. Similarly, we grouped the benefits data by job, summarizing the types of benefits associated with each job into one field. These datasets were then merged to provide a complete view of the skills and benefits for each job.

Next, we added salary and industry data to the job dataset. This ensured that each job posting was matched with the corresponding salary information and industry classification. Industry mappings were also included to provide more descriptive industry names. This resulted in a dataset with comprehensive job-level information, including skills, benefits, salaries, and industry details.

In parallel, we processed the company-related data. We grouped company specialties and industries, connecting all specialties and industries for each company into single fields. We also filtered the employee count data to retain only the most recent entries for each company, ensuring we captured the latest company size. This created a complete dataset containing each company's specialties, industries, and employee count.

After processing the company data, we merged it with the job postings using company_id and company_name. This step allowed us to link job postings with their corresponding company attributes, combining detailed job data with company-level information, such as industry and size. This enriched our dataset for a more in-depth analysis. The processed data was used for regression analysis in trying to answer the third question. We counted the number of job postings of each listed company and conducted linear regression to capture the correlation effect.

To clean the data further, we excluded certain companies, such as Talentify.io, Dice, and myGwork. These companies are recruiting platforms with large numbers of job postings in the software category, but they are not actual software companies. Including them could have skewed our analysis, so we removed them to focus on real software employers. Additionally, we filtered out job postings from industries considered less relevant, such as unskilled labor sectors like Staffing and Recruiting, Construction, Hospitality, and Retail. This allowed us to narrow our analysis to more skilled industries.

Stock Price Dataset

We used the ‘tidyquant’ package in R to retrieve historical stock price data for the selected companies. The data was collected for the period from January 1, 2016, to August 1, 2023. The stock tickers for the companies were defined in advance, and the tq_get() function was used to extract stock prices.

Additionally, we mapped each stock ticker to the full company name for easier reference in the analysis. A distinct color palette was applied to each company to visually differentiate them in future visualizations, ensuring clarity in the data presentation.

This method allowed us to integrate stock performance data into our analysis alongside the job posting data, enabling a comprehensive view of company trends over time.

Linear Regression

Some variables in our dataset are worth investigating as they may involve correlations. In our analysis, we focus on job posting count as the dependent variable, with employee count as the independent variable. We want to explore the relationship between company size and job posting activity. Specifically, we used the job posting count as the dependent variable and employee count as the independent variable. This model was chosen to investigate whether the size and scope of a company, measured by its employee count, has a significant influence on its hiring activity. By analyzing this relationship, we aim to determine whether larger companies, which typically have more resources, post more job opportunities on platforms like LinkedIn. This approach allows us to quantify the impact of employee count on job posting behavior and assess its statistical significance.

Results

Industry Demand and Top Hiring Companies

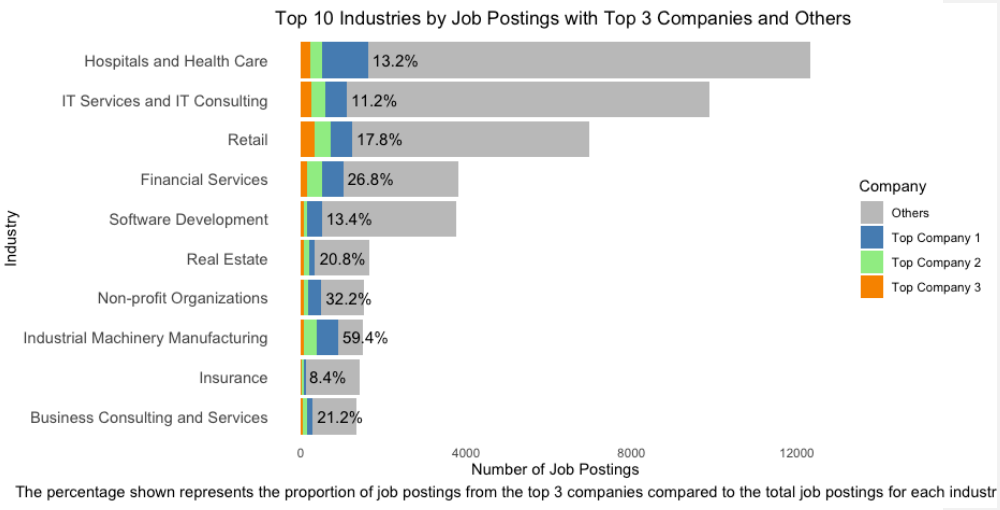

Figure 1. Job Postings and Industries (top 10)

To create this plot, we first prepared the data by grouping it by industry and calculating the total number of job postings for the top 3 companies within each industry. This allowed us to compute the proportion of job postings from the top 3 companies relative to the total postings for each industry. We then filtered the data to focus on the top 10 industries with the highest number of job postings, which made it easier to highlight the concentration of job postings among the top companies in those industries.

For the visualization, we used a stacked bar plot where each industry's total job postings were divided into segments representing the top 3 companies and the remaining companies, grouped as "Others." We flipped the bars horizontally to improve readability and applied distinct colors to differentiate the company segments. Additionally, we added percentage labels above the "Others" section of each bar to show the proportion of job postings from the top 3 companies. Finally, we applied a minimalist theme by removing grid lines and adjusting text sizes to ensure clarity and enhance visual appeal.

The bar graph (Figure 1) displays the top 10 industries by the number of job postings, highlighting the contributions from the top 3 companies in each industry compared to other companies. This analysis reveals several key trends regarding the distribution of job postings across industries and the dominance of certain companies in each sector.

Among the top 10 industries, Hospitals and Health Care recorded the highest number of job postings. The top 3 companies accounted for just 13.2% of the total job postings, highlighting the intense competition among companies for hiring in this sector.
Industrial Machinery Manufacturing stood out as an exceptional case. The top 3 companies in this sector contributed to a remarkable 59.4% of all job postings, making it the most concentrated industry, where hiring is dominated by a few large companies. This suggests that a small number of major players are driving the majority of hiring in this industry, making it unique compared to other sectors.

Given these observations, we decided to focus on the first five industries (Hospitals and Health Care, IT Services and Consulting, Retail, Financial Services, and Software Development) as well as Industrial Machinery Manufacturing due to the latter's unique concentration of hiring within the top 3 companies. These industries were selected because they either had significant overall hiring demand or exhibited notable patterns in the distribution of job postings among companies.

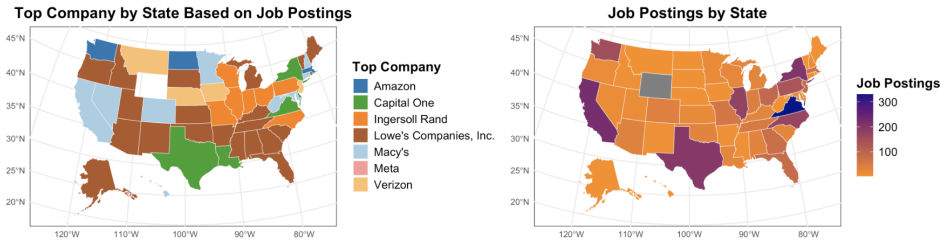

Figure 2. Aerial Job Posting distribution across the United States

Figure 2 illustrates two aspects of job postings. Only the top companies from the Fig 2 were selected for the purposes of this plot. The left map highlights the top companies based on the number of job postings by state. Companies like Lowe's Companies Inc., Macy's, and Ingersoll Rand seem to be hiring the most people across the states. The right map displays the distribution of job postings by state, where darker shades indicate a higher number of postings. The states with the most job postings are predominantly concentrated in regions with major urban centers, with states like California, Texas, and Virginia standing out. Together, these maps provide a comprehensive overview of the geographical distribution of job opportunities and the key employers driving hiring in each state.

Stock Performance of Leading Employers

Following the analysis above, we focus on the top 3 companies of each industry ranked in the top 10. The companies are listed as follows:

The companies were selected based on a previous bar graph analysis, where we identified key sectors, and within those sectors, we shortlisted the top companies. From that list, we chose publicly traded companies for analysis, making this a focused study on the most influential players in their respective industries.

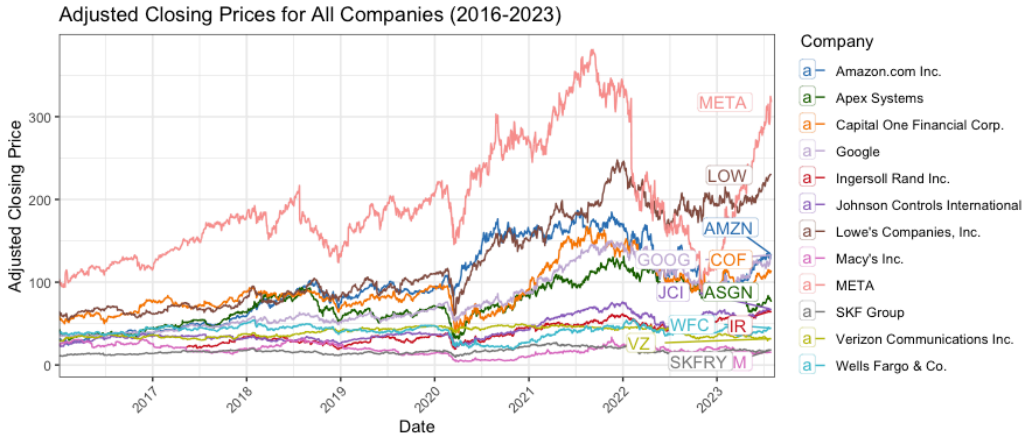

Figure 3. Adjusted Closing Prices Across Top Companies

The first graph (Figure 3), "Adjusted Closing Prices for All Companies (2016-2023)," provides a historical view of the stock price trends for all the selected companies. Notably, META (Meta Platforms, Inc.) and Amazon (AMZN) exhibited a consistent upward trend in adjusted closing prices over time, despite occasional fluctuations. META especially stands out, with a sharp price increase starting around 2020, possibly influenced by its shift to focus on the metaverse and other digital services. Google (GOOG) and Ingersoll Rand (IR) also demonstrate steady growth, although at a slower pace compared to META and Amazon. On the lower end, Macy's (M) and Verizon (VZ) show relatively flat performance, indicating either limited growth or challenges in maintaining stock price momentum during the observed period.

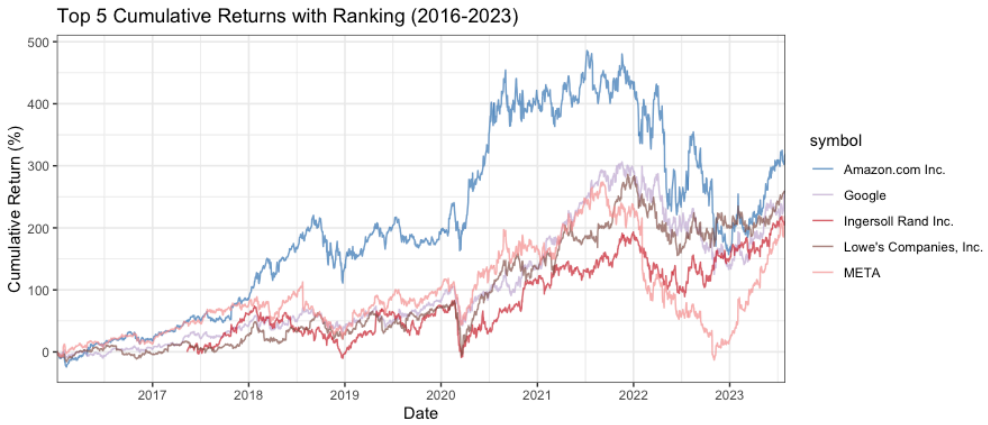

Figure 4. Cumulative Returns Across Top Companies

In the second graph (Figure 4), "Top 5 Cumulative Returns with Ranking (2016-2023)," we see a comparison of cumulative returns for the top 5 companies. Amazon leads with the highest cumulative return, reaching over 400% at its peak. This highlights Amazon's strong growth trajectory during the period, benefiting from its e-commerce dominance and cloud computing ventures. META and Google follow, with META showing significant volatility but still maintaining a strong upward trend. Both Ingersoll Rand and Lowe's (LOW) show more consistent growth but with smaller cumulative returns, indicating their relatively stable but slower growth compared to the tech giants.

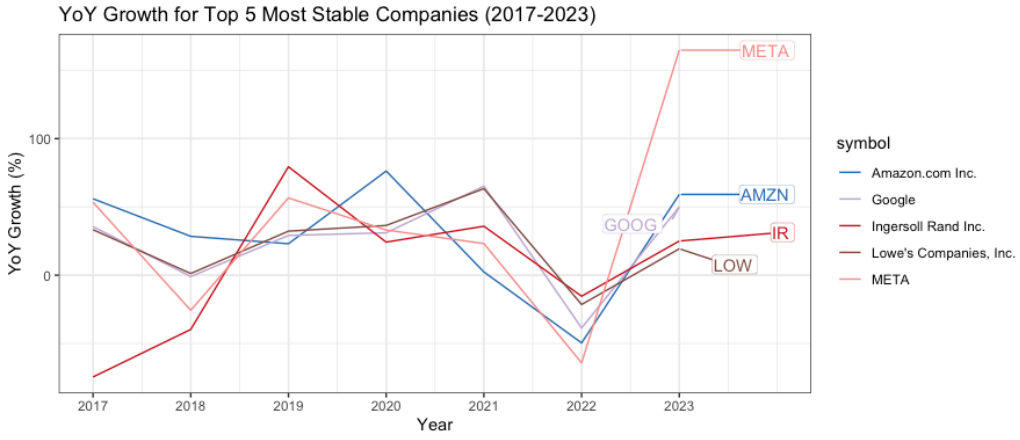

Figure 5. Year-over-year growth Across Top Companies

The third graph (Figure 5), "YoY Growth for Top 5 Most Stable Companies (2017-2023)," focuses on year-over-year growth rates for the top 5 most stable companies. A key finding here is the sharp decline in YoY growth across all companies during 2020 and 2021, likely due to the global economic impact of the COVID-19 pandemic. However, META experienced an extraordinary rebound in 2022 and 2023, posting the highest YoY growth of all companies, potentially driven by post-pandemic recovery and its investments in digital services. Amazon and Google also recovered, showing strong growth, while Ingersoll Rand and Lowe's displayed more moderate, consistent YoY growth, indicating their stability as compared to the more volatile tech companies.

Together, these three figures illustrate the varied growth patterns, stability, and volatility of major companies across sectors, highlighting differences in long-term growth, returns, and resilience in the face of external shocks like the pandemic.

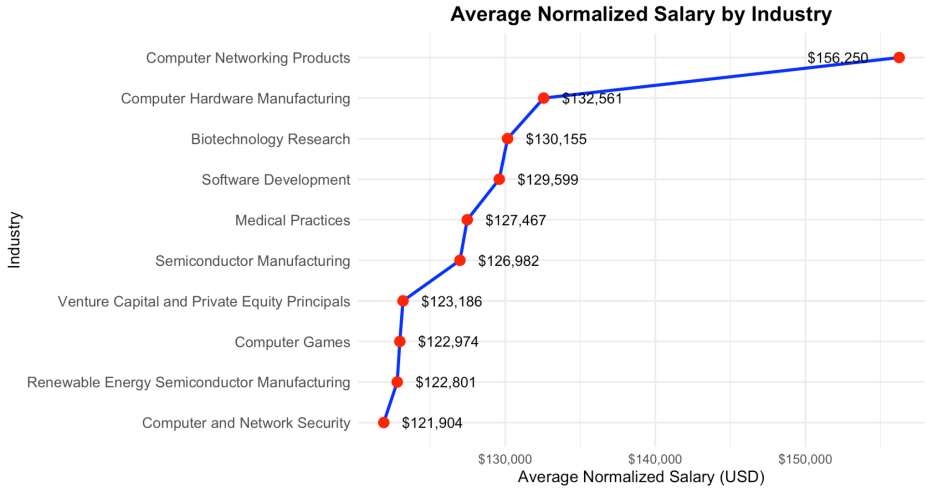

Figure 6. Average Normalized Salary by Industries (Top 10)

Figure 6 displays the average normalized salaries across various industries highlighting how technology-related sectors dominate the top ranks. The only non-tech industry is medical practices which is intuitive due to the high salaries of doctors in the US. The domination of tech- industries in the top salaries indicates demand for professionals who are highly skilled in tech and engineering-related areas.

The Effect of Company Popularity on Job Posting Trends

So far, our analyses have shown that the number of job postings varies across different industries. Additionally, we have examined the target companies in relation to their stock prices over recent years. With the variables we have for now, we want to delve deeper into what factors influence job postings.

Figure 7. Correlation Table Between Numeric Variables

We begin by plotting a correlation matrix using the available numerical data. Figure 7 shows several relationships between the number of job postings and other organizational metrics. First, the number of employees has a relatively large positive correlation with the number of followers (0.74), suggesting that larger companies tend to have more social media engagement, which could influence their visibility in job postings. However, the number of employees shows weak or negligible correlations with the number of views (-0.03) and number of applications (-0.04), indicating that company size does not directly affect the number of people viewing or applying for jobs. Finally, experience shows the strongest correlation with salary (0.41), which implies that companies offering higher salaries may attract more experienced candidates, which could indirectly affect the job posting dynamics.

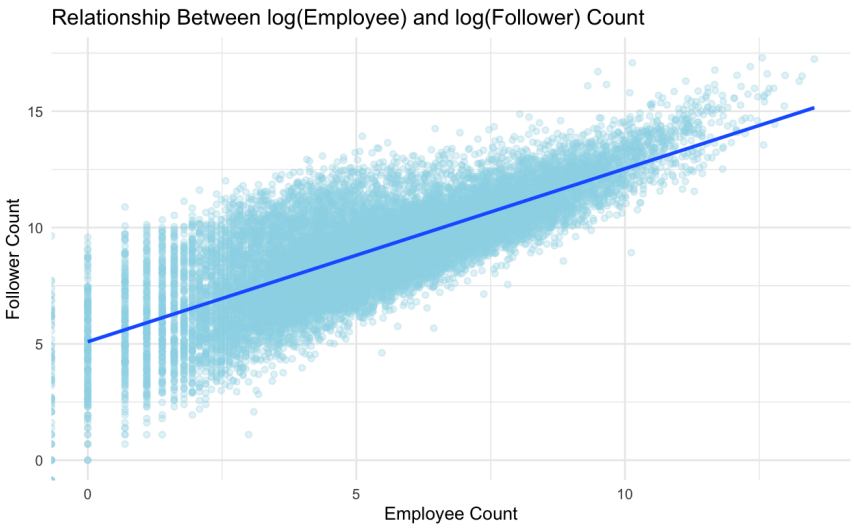

Figure 8. Relationship between Employee count and Follower count (logarithm) with regression line

Term	Estimate	Std.Error	T value	Pr(>|t|)
(Intercept)	1.347e+03	8.816e+01	15.28	<2e-16***
No. of Followers	1.736e-02	1.317e-04	131.77	<2e-16***

Commented [1]: table 1 or fig 9 need label

Table 1. Regression Results for Employee and Follower Count

The regression analysis between employee count and follower count, as illustrated in Figure 8 and Table 1, demonstrates a significant positive relationship. The scatter plot in Figure 8 shows that as the number of employees increases, so does the follower count, with a clear positive trend in the log-transformed data. This finding aligns with the results from Figure 7, which shows a strong positive correlation (0.74) between employee count and follower count. Larger companies tend to have greater social media engagement, which could increase their visibility in the job market.

The regression results in Table 1 indicate that employee count is a strong predictor of follower count, with a highly significant coefficient ($p < 2e\text{-}16$) and an estimate of 1.736e-02. This suggests that as companies expand their workforce, their social media following grows proportionally, reinforcing the idea that company size contributes to greater online visibility. While larger companies may benefit from this increased social media engagement to attract more attention to their job postings, employee count itself shows weak correlations with job views and applications.

Discussion

Our analysis of Industry Demand and Top Hiring Companies revealed that sectors like Hospitals and Health Care exhibit more distributed hiring patterns, while industries such as Industrial Machinery Manufacturing are dominated by a few large employers. This suggests that certain industries concentrate on hiring within a small number of companies, possibly due to the need for specialized skills or market dominance. Additionally, we observed regional concentration in states like California, Texas, and Virginia, which likely reflects the strength of their economic centers. However, our focus on specific industries may have excluded unadvertised or temporary positions, limiting the scope of the findings. Future studies could explore more sectors and investigate how regional and economic factors drive hiring patterns.

In evaluating the stock performance of leading employers, companies like Amazon and Meta demonstrated substantial cumulative returns and year-over-year growth. However, we found no direct correlation between hiring activity and stock performance. Future research could explore this relationship further by analyzing how changes in stock market performance influence hiring and salary decisions through a time series approach.

In exploring the Effect of Company Popularity on Job Posting Trends, we found that employee count strongly predicts social media follower count, suggesting that larger companies are more visible online. However, employee count alone explains only 6.4% of the variance in job posting activity, indicating that other factors—such as industry type, company reputation, and hiring strategies—play more significant roles. Further research could investigate how external factors, such as economic shifts or technological advances, impact job posting behavior and whether company visibility directly influences recruitment strategies.