Analyzing Company Hiring Patterns Using K-Means Clustering and Association Rule Mining: A Data-Driven Approach to Understanding Recruitment Trends in the Digital Economy

Min-Tsai Lai[1], Taqwa Hariguna[2,*]

[1]Department of Business Administration, Southern Taiwan University of Science and Technology, Taiwan

[2]Department of Information System and Magister Computer Science, Universitas Amikom Purwokerto, Indonesia

ABSTRACT

This study explores the relationship between company characteristics and recruitment trends by analyzing a dataset obtained from Simplify.jobs, which contains detailed profiles of companies and their job postings. The research focuses on how organizational attributes such as funding stage, customer type, and company size influence recruitment strategies and job posting behaviors. Using clustering techniques, the study identifies three distinct clusters of companies based on these attributes, revealing that early-stage companies prioritize technical hires while later-stage companies offer a more diverse array of benefits. The study also employs association rule mining to uncover frequent patterns between job attributes, such as the tendency for tech companies to offer remote work options. These findings highlight how companies adapt their recruitment strategies as they grow, with early-stage companies leveraging financial incentives like stock options to attract talent, while later-stage companies emphasize employee retention through comprehensive benefits packages. The results offer valuable insights for HR professionals and recruiters, enabling them to tailor their strategies according to company profiles. This research contributes to the broader field of data-driven recruitment analysis by providing a nuanced understanding of how company characteristics shape hiring practices and job posting trends. The study also paves the way for future research into the evolution of recruitment strategies over time and the application of similar methodologies to other industries.

Keywords Recruitment Strategies, Company Characteristics, K-Means Clustering, Association Rule Mining, Job Posting Behavior

Submitted 25 December 2024
Accepted 4 February 2025
Published 8 March 2025

Corresponding author
Taqwa Hariguna,
taqwa@amikompurwokerto.ac.id

Additional Information and Declarations can be found on page 41

Introduction

The rise of digital platforms and artificial intelligence (AI) technologies has significantly transformed recruitment strategies within the digital economy. As businesses increasingly turn to online solutions, AI-powered tools have become integral in reshaping how companies discover, assess, and hire talent. The adoption of these technologies is particularly evident in the recruitment process, where AI algorithms now shift through vast amounts of candidate data, predict ideal fits, and even perform initial interviews. This evolution not only accelerates the hiring cycle but also introduces new efficiencies, enabling organizations to make data-driven decisions with unprecedented speed and precision [1]. In a time where businesses must navigate a hyper-competitive talent pool, AI's

How to cite this article: M. T. Lai and T. Hariguna, "Analyzing Company Hiring Patterns Using K-Means Clustering and Association Rule Mining: A Data-Driven Approach to Understanding Recruitment Trends in the Digital Economy ," *J. Digit. Soc.*, vol. 1, no. 1, pp. 20-43, 2025.

capacity to reduce human error and optimize selection processes offers a strategic advantage.

At the core of this shift lies the ability of AI to process and analyze data at a scale that was previously unimaginable. Machine learning models, such as natural language processing (NLP) and deep learning, empower AI systems to assess resumes, job applications, and even social media profiles to gauge a candidate's suitability. These tools can also enhance diversity by mitigating the biases that have historically influenced hiring decisions [2]. The data-driven approach eliminates subjectivity, thus opening doors for more equitable recruitment practices. However, the reliance on algorithms also raises important questions about transparency and fairness, as the opacity of AI decision-making processes can inadvertently perpetuate new forms of discrimination if not carefully monitored.

Moreover, the intersection of AI and digital recruitment has sparked the development of more sophisticated platforms that integrate candidate sourcing with real-time job market analysis. These platforms, powered by big data and predictive analytics, can anticipate hiring needs based on industry trends, providing companies with actionable insights that drive proactive recruitment strategies [3]. By tapping into these tools, organizations can not only enhance their ability to attract top talent but also shape their recruitment processes in alignment with broader business goals. As these platforms evolve, they promise to redefine not just how recruitment is done, but also how the workforce itself is shaped in the digital age.

Despite these advancements, challenges remain in achieving the perfect synergy between technology and human expertise. The notion of algorithmic fairness—ensuring that AI systems do not reinforce existing biases or disproportionately disadvantage certain groups—continues to be a focal point in ongoing research [4]. Furthermore, while AI can optimize certain aspects of recruitment, it remains to be seen whether these systems can fully replicate the nuanced decision-making capabilities of human recruiters. Thus, the integration of AI into recruitment processes must be viewed not as a replacement for human judgment but as a complement that amplifies the strengths of human recruiters while minimizing their weaknesses. This dynamic interplay between human intuition and machine intelligence offers a fertile ground for further exploration, marking a critical juncture in the evolution of recruitment technologies.

The increasing reliance on data-driven approaches in recruitment strategies marks a pivotal shift within the digital economy, as organizations strive to optimize both their hiring processes and job posting strategies. In a landscape where time-to-hire and candidate quality are paramount, the conventional, intuition-based methods are giving way to more systematic, evidence-based practices. Artificial intelligence (AI) and machine learning algorithms play a central role in this transformation, not only streamlining operations but also enhancing decision-making with insights drawn from vast, dynamic datasets. Companies no longer depend solely on human judgment to fill positions but are leveraging data to make hiring decisions that are more precise, efficient, and equitable [5].

AI technologies have fundamentally reshaped recruitment practices, offering automation that extends beyond simple resume sorting to encompass more sophisticated, strategic interventions. Predictive analytics, for example, allows organizations to assess historical hiring data to forecast staffing needs and determine which recruitment channels yield the best candidates [6]. This analytical foresight ensures that talent acquisition strategies are not only reactive but proactive, enabling businesses to anticipate workforce requirements well in advance. By leveraging data insights in this manner, organizations can pinpoint the most effective platforms for job postings, refine job descriptions to attract the right talent, and optimize the candidate selection process—ultimately improving overall hiring success rates.

The scope of AI in recruitment is particularly compelling when considered through the lens of reducing bias and fostering diversity in hiring decisions. Machine learning algorithms, designed to learn from vast pools of data, can uncover hidden patterns and correlations that human recruiters might overlook, thereby offering a more objective approach to candidate evaluation. Algorithms can be fine-tuned to ignore variables that have historically contributed to biased decision-making, such as gender, race, or age, providing a level of impartiality that human recruiters might inadvertently fail to achieve [2]. However, this benefit is not without its challenges, as ensuring the integrity of AI models remains a complex task, with biases potentially being encoded into systems if not properly managed. Hence, the reliance on AI in recruitment also calls for heightened vigilance in maintaining transparency and fairness in algorithmic decision-making.

As companies adopt these data-driven tools, the recruitment ecosystem evolves into an intricate network of algorithms, platforms, and human oversight. AI technologies empower recruiters to shift from labor-intensive processes to more strategic, high-level functions. Yet, this shift does not eliminate the role of human judgment entirely. Instead, it complements it, allowing recruiters to focus on nuanced aspects of talent acquisition, such as culture fit and candidate potential. The intersection of data-driven automation and human expertise opens new avenues for innovation, signaling a future where recruitment is not only faster and more efficient but also more insightful, adaptive, and inclusive. This balanced integration of technology and human intelligence is shaping the future of work, redefining how organizations approach hiring in the digital age.

Despite the growing influence of AI and data-driven tools in recruitment, there remains a notable gap in understanding the nuances of how company characteristics—such as funding stage, year founded, and customer type—affect job posting behavior. While much has been made of how AI optimizes recruitment processes, few studies have rigorously explored the deeper connections between a company's business profile and its hiring patterns. Most research on recruitment trends tends to focus on factors such as industry or job market conditions, leaving the influence of intrinsic company traits largely unexplored. Yet, a company's age, its stage in the funding lifecycle, and its customer base likely shape not only the types of roles posted but also the frequency and platforms used for posting. These organizational elements could therefore play a significant role in shaping the recruitment landscape, influencing everything from job requirements to the geographic focus of hiring campaigns. This research seeks to address this gap, offering an empirical analysis of the ways in which company characteristics interact with their recruitment behaviors.

In recent studies, data mining techniques have been increasingly applied to various domains, including sentiment analysis and recruitment trends. For instance, [7] has contributed to understanding emotional sentiment in user feedback, which parallels sentiment analysis in recruitment [8]. Similarly, [9] offers insights into improving sentiment classification, which can be applied to analyzing job postings and employer branding. In the realm of customer segmentation, [10] provides valuable lessons on clustering techniques for understanding recruitment behaviors [11]. Furthermore, [12] demonstrates the effectiveness of various sentiment classification methods, relevant to the sentiment analysis of job postings [13]. Additionally, [14] offers useful anomaly detection methods that can be applied to identify irregularities in recruitment patterns. Research [15] highlights predictive analytics in blockchain, which can also inform recruitment predictions. Moreover, studies like [16] and [17] offer unique perspectives on how emerging industries influence recruitment trends, such as the blockchain job market. Research [18] has demonstrated how clustering can reveal behavioral patterns, an approach that is directly applicable to recruitment analytics. Finally, [19] exemplifies how predictive modeling can forecast recruitment outcomes based on organizational characteristics. These studies collectively reinforce the importance of data-driven approaches in understanding recruitment dynamics.

The lack of attention to these variables in existing literature is striking, given that business fundamentals such as funding stage and customer type can profoundly impact hiring needs. For example, a startup in its early funding stages may have significantly different recruitment needs compared to a well-established firm with a steady customer base. Startups often focus on building a core team and may look for versatile employees who can wear multiple hats, whereas established companies may seek highly specialized professionals to drive growth and innovation. Similarly, companies targeting a global consumer base may have different hiring patterns compared to those serving regional markets. Thus, understanding these dynamics can offer critical insights into how recruitment strategies vary based on company profiles, providing a richer context for optimizing talent acquisition strategies.

This study aims to explore these nuanced hiring patterns by employing advanced data mining techniques, particularly clustering and association rule mining, to investigate the relationship between company characteristics and job posting behavior. By analyzing variables such as funding stage, year founded, and customer type, this research seeks to uncover distinct patterns in recruitment trends. Clustering algorithms, such as K-Means, will segment companies based on similarities in their business profiles, while association rule mining will be used to discover underlying relationships between job posting behaviors and company characteristics. This multi-method approach provides a robust framework for understanding how companies of various types and stages engage in recruitment, enabling more targeted and informed decision-making. The integration of these techniques offers not just a statistical overview but a comprehensive, actionable model for optimizing recruitment strategies across different organizational profiles.

In addition to examining recruitment patterns, this study will also analyze how these behaviors change over time and in response to external factors like market conditions or shifts in consumer demand. By examining the dynamics of recruitment in relation to company growth and maturation, this research provides a longitudinal view of recruitment trends. Understanding the progression of recruitment needs from the perspective of a company’s lifecycle—whether it’s a nascent startup, a growth-stage company, or an established organization—can offer crucial insights into the strategic adjustments that businesses make to attract talent at different stages of their development. This focus on temporal dynamics allows for a deeper understanding of how hiring strategies evolve and how companies adapt to meet their staffing needs in an ever-changing digital economy.

This study holds significant implications for both practitioners and researchers in the field of recruitment and human resources. By examining how companies tailor their recruitment strategies to fit their unique profiles, it provides actionable insights for HR professionals, hiring managers, and organizational leaders looking to fine-tune their recruitment processes. Specifically, understanding how factors such as funding stage, customer type, and company age influence hiring behavior can empower organizations to make more strategic, data-driven decisions about where and how to allocate resources for recruitment. Furthermore, by identifying patterns in recruitment behavior, the study offers a framework that organizations can use to predict future hiring trends and proactively adapt their strategies to meet emerging needs.

From a broader perspective, this research contributes to the growing body of literature on data-driven recruitment analysis in the digital economy. As AI and data mining techniques continue to gain prominence in the field of human resource management, studies such as this one provide foundational knowledge that bridges the gap between business analytics and talent acquisition. By leveraging data mining techniques to explore company profiles and recruitment trends, this study adds to the emerging field of business-driven recruitment analytics, offering valuable insights for both academic inquiry and real-world applications. Through this research, we not only gain a deeper understanding of how companies adapt their hiring practices to their specific needs but also pave the way for more sophisticated, tailored recruitment strategies that align with the evolving landscape of the digital economy.

Literature Review

Data Mining in Recruitment

In recent years, data mining techniques have gained significant traction in the field of recruitment, offering novel insights into organizational hiring practices. As businesses increasingly rely on data-driven strategies to inform decision-making, the application of data mining to recruitment processes has proven invaluable in deciphering the complex patterns that underpin job posting behaviors. These techniques enable the extraction of meaningful knowledge from vast datasets that encompass job descriptions, candidate profiles, company characteristics, and historical hiring trends. By applying algorithms such as clustering and association rule mining, researchers have been able to uncover underlying trends and relationships, facilitating the development of more targeted and efficient recruitment strategies. This shift toward data-centric approaches represents a profound transformation in recruitment, moving from intuition-driven decisions to empirically validated practices that optimize both talent acquisition and organizational growth.

Among the various data mining techniques, association rule mining has been particularly effective in identifying relationships between different job attributes and organizational characteristics. This method works by uncovering frequent itemsets within datasets and generating rules that describe how certain attributes co-occur. Research underscore the power of association rule mining in recruitment by demonstrating its capacity to generate actionable insights from complex relational datasets. For instance, by analyzing the links between job titles, required skills, and company attributes like funding stage or customer type, association rule mining can reveal patterns that help predict the types of candidates an organization is likely to seek. Such insights are crucial for developing recruitment strategies that align with both organizational needs and market conditions, allowing businesses to anticipate staffing requirements and tailor their job postings accordingly.

In addition to association rule mining, clustering techniques have also been widely used to segment organizations based on various characteristics and analyze how these segments differ in their recruitment behaviors. Clustering algorithms, such as K-Means, group similar data points together, enabling the identification of distinct patterns within the recruitment strategies of different types of companies. For example, startups in their early stages may exhibit different hiring patterns than large, well-established corporations, as the former typically prioritize flexibility and generalist roles, while the latter focus on specialized positions. By clustering companies based on variables such as funding stage, industry, and customer type, researchers can uncover distinct recruitment strategies and provide recommendations tailored to each group's unique needs. This segmentation allows companies to refine their recruitment efforts, ensuring that they attract the right talent at the right time.

The application of data mining techniques to recruitment is not limited to the identification of patterns within organizational characteristics. These methods have also been employed to optimize job postings themselves, ensuring that they reach the right audience and attract the most suitable candidates. For example, by analyzing historical job posting data, organizations can better understand which platforms and job descriptions lead to higher application rates and more successful hires. As researchers continue to explore the intersection of data mining and recruitment, the integration of clustering and association rule mining will likely continue to evolve, providing increasingly sophisticated tools for businesses to improve their recruitment strategies. By bridging the gap between organizational data and hiring trends, data mining offers a powerful approach to understanding the complexities of recruitment in the digital age.

Clustering in Recruitment Analysis

K-Means clustering has emerged as one of the most effective and widely used techniques in analyzing job posting behaviors and organizational characteristics. This unsupervised machine learning algorithm allows companies to group themselves based on similarities in recruitment patterns, particularly in terms of the types of positions they seek to fill, the skills they require, and their overall hiring frequency. In a digital landscape where recruitment strategies are increasingly data-driven, K-Means offers a powerful tool for identifying latent patterns in vast datasets of job postings and organizational attributes. By segmenting companies into clusters, organizations can gain critical insights into industry-specific recruitment behaviors, optimize their job posting strategies, and adjust their approach to meet evolving talent demands. This approach not only improves recruitment efficiency but also aids in the identification of best practices across similar organizations, providing actionable intelligence that can enhance overall hiring strategies.

The application of K-Means clustering in recruitment is particularly effective when combined with natural language processing (NLP) techniques to analyze the content of job postings. Research [20] demonstrated the power of NLP in conjunction with clustering methods, highlighting how job descriptions—often rich with unstructured text—can be parsed and classified to uncover meaningful patterns. For instance, by analyzing the language used in job titles, required skills, and qualifications, companies can identify emerging trends in the workforce and better align their recruitment efforts with market demands. NLP tools enable the extraction of relevant features from job postings, such as job titles, responsibilities, and skills, which can then be used as input for the K-Means algorithm to group similar postings. This integrated approach allows companies to better understand how their job postings compare to those of competitors and tailor their recruitment messaging accordingly.

Furthermore, clustering can also be used to group companies based on organizational characteristics such as funding stage, company size, and customer type. Companies at different stages of their lifecycle exhibit distinct recruitment patterns that reflect their operational priorities and resource capabilities. For example, startups and early-stage companies may prioritize versatility and innovation in their hires, seeking candidates with broad skill sets who can adapt to fast-paced environments. In contrast, mature organizations may focus on specialized expertise and long-term experience, recruiting for specific roles to support established business functions. K-Means clustering provides a systematic way to categorize companies into groups based on these factors, allowing them to benchmark their recruitment strategies against others in the same category.

This approach is not only relevant for understanding internal recruitment dynamics but also for identifying external trends in talent acquisition across industries. By grouping companies with similar recruitment behaviors, K-Means clustering allows organizations to identify broader trends within their industry, such as common skill gaps or the growing demand for particular roles. For instance, a cluster of technology firms might exhibit a high demand for software engineers, whereas companies in the retail sector might prioritize customer service roles. The ability to visualize and quantify these trends provides companies with a more strategic view of the job market, enabling them to adjust their hiring priorities, refine job descriptions, and optimize recruitment campaigns to attract the best talent [21]. Thus, K-Means clustering proves invaluable not only as a diagnostic tool for internal analysis but also as a strategic guide for navigating the complex landscape of modern recruitment.

Association Rule Mining in Recruitment

Association rule mining serves as an invaluable tool in recruitment analysis, particularly when it comes to uncovering hidden relationships between various job attributes. By identifying frequent co-occurrences of job-related factors, such as job titles, required skills, and offered benefits, this technique enables organizations to gain deeper insights into the preferences and patterns that shape recruitment processes. In a highly competitive job market, understanding these interconnections can empower companies to refine their hiring strategies, align job offerings with candidate expectations, and attract the right talent. For example, identifying which job titles are commonly associated with specific benefits or career progression opportunities helps organizations tailor their offerings to meet the evolving demands of prospective employees, enhancing both the attractiveness of job postings and the likelihood of attracting qualified candidates. Using association rule mining, organizations can not only optimize their recruitment processes but also better predict the future needs of their workforce [22].

One of the primary applications of association rule mining in recruitment is the analysis of job postings to identify which job titles are frequently linked with specific benefits. For example, [23] applied association rule mining to job seeker profiles, leveraging personality traits and social media activity as key attributes. Their study revealed that certain personality profiles were more strongly associated with particular job titles, which in turn influenced the types of benefits those candidates sought in potential employers. This application not only underscores the power of association rules in discovering patterns across diverse job attributes, but it also highlights the growing importance of understanding the psychological and social dimensions of recruitment. In this context, benefits such as remote work options, flexible hours, and health coverage emerge not just as perks, but as key factors in shaping candidate-job matches.

Another critical application of association rule mining is the identification of trends that can inform organizational decisions on recruitment channels and the presentation of job offers. When combined with large-scale data analytics, association rule mining reveals which benefits—such as job stability, growth opportunities, or work-life balance—are most often linked with certain industries or company types. For instance, by analyzing job titles in conjunction with company characteristics (e.g., funding stage or customer type), recruitment professionals can better understand how to position their offers to meet market demand. This allows companies to design more effective job advertisements, optimize recruitment efforts, and ultimately align their offerings with candidates' needs and values. Research [24] demonstrated how these relationships, when mined effectively, can guide firms in crafting tailored recruitment campaigns that resonate with top talent, further bridging the gap between candidate expectations and organizational needs.

Furthermore, the intersection of job titles, benefits, and company characteristics can also offer a lens through which recruitment practices evolve with changing societal trends. As the global workforce becomes increasingly diverse, the range of benefits that companies offer—and the preferences candidates have—undergo continuous shifts. Association rule mining provides the analytical framework necessary to track these changes in real time. For example, by linking job attributes to external factors such as economic shifts or technological advancements, companies can adapt their recruitment strategies in anticipation of changing job market dynamics. This adaptive approach, empowered by data-driven insights, reflects the increasing reliance on sophisticated mining techniques to navigate the complexities of recruitment in the digital age.

Gaps in Literature

While the application of data mining techniques like clustering and association rule mining has been explored individually in the context of recruitment analysis, few studies combine these two approaches to offer a comprehensive understanding of recruitment behavior in relation to company profiles. This gap is notable because the integration of these methodologies could uncover deeper insights into how organizational characteristics such as funding stage, company size, and customer type influence hiring patterns. By leveraging clustering to group companies based on similar recruitment strategies and then applying association rule mining to examine relationships between job attributes and company profiles, a more holistic view of recruitment dynamics could be achieved. Such integrated analyses could inform more targeted and data-driven recruitment strategies, aligning them with broader organizational goals and industry trends. Despite the growing interest in recruitment analytics, the synergy between clustering and association rule mining remains underexplored in current literature.

A few studies have explored the combination of clustering and association rule mining in other domains, demonstrating the potential of this integrated approach. For instance, [25] applied K-Means clustering in conjunction with association rule mining to explore patterns in technology patent data, identifying clusters of technologies and revealing the relationships between technologies within each cluster. This study not only illustrated how clustering can group similar entities but also showed how association rule mining can identify key interdependencies within those clusters. While their focus was not on recruitment, their methodology provides a valuable precedent for applying the same techniques to company profiles and job posting behaviors. The integration of these two methods in the recruitment domain could similarly reveal critical associations between organizational traits and job posting characteristics, offering deeper insights into how companies adapt their recruitment strategies based on internal and external factors.

Despite the potential of combining clustering and association rule mining, few studies have ventured into the recruitment sector to explore these interactions. Many studies have focused on individual methods—clustering to segment organizations by size or industry and association rule mining to identify correlations between job titles and required skills. However, these analyses often remain siloed, and the combined use of both methods could offer a more nuanced and dynamic view of recruitment behavior. For instance, clustering could be used to group companies based on their funding stage or customer type, while association rule mining could identify the relationships between job titles, required skills, and benefits offered by those companies. By addressing this gap, future research could yield more robust models that offer practical recommendations for optimizing recruitment strategies.

This gap in the literature suggests a promising direction for future studies in the field of recruitment analytics. By combining clustering and association rule mining, researchers can create more sophisticated models that not only identify patterns within job postings but also provide insights into the broader organizational contexts in which these patterns emerge. This approach could be especially beneficial for businesses looking to refine their hiring strategies, as it would allow them to better understand the specific needs and behaviors of companies with similar profiles. Ultimately, the integration of these two techniques could help bridge the gap between data-driven recruitment insights and actionable strategies, enhancing both the efficiency and effectiveness of talent acquisition in the digital economy.

Method

The research method comprises meticulously designed steps for a thorough analysis. Figure 1 presents a detailed outline of these steps.

Figure 1 Research Method Flowchart

Data Collection

The dataset for this study was obtained through an API scraping process from Simplify.jobs, a platform that aggregates company profiles and job listings. This data, composed of 21 distinct columns, offers a detailed snapshot of organizational characteristics and job posting behaviors. The dataset includes crucial attributes such as the company name, funding stage, job titles, benefits offered, and social media links (e.g., Twitter, LinkedIn, Crunchbase). These attributes provide rich contextual information about each company’s recruitment strategies and organizational profiles, making it an ideal resource for analyzing job posting trends in relation to company characteristics. By integrating data from diverse sources, such as social media profiles and job listings, the dataset allows for a multifaceted approach to understanding recruitment patterns in the digital economy.

The data was processed and imported into a pandas DataFrame for initial exploration and cleaning. To ensure compatibility and address potential encoding issues, the dataset was loaded using the ISO-8859-1 encoding format. This encoding was particularly useful given the presence of special characters and varied formats within the data. To maintain data integrity and prevent errors during processing, any badly formatted lines were skipped using the `on_bad_lines='skip'` parameter. Upon loading the data, the dataset was examined to identify any missing or anomalous values, and the columns were reviewed for relevance and consistency. This preliminary data inspection ensured that the dataset was suitable for further analysis, eliminating potential biases that might arise from inconsistent or incomplete data entries.

Key attributes like company funding stage, job titles, and benefits were specifically selected for deeper analysis in this study. These features were crucial for understanding how companies with different financial and organizational profiles adapt their recruitment strategies. For example, companies at various funding stages may display distinct hiring patterns, with early-stage startups potentially posting more roles in specific technical domains, while established companies might focus on leadership or operational roles. Similarly, job titles and the benefits associated with them offer insights into the organizational culture and values, which can further elucidate recruitment strategies. This selection of variables reflects the study's emphasis on understanding the complex interplay between company profiles and their recruitment behaviors.

The use of API scraping to gather this dataset is particularly relevant in today's digital age, where the availability of large-scale, publicly accessible data allows for sophisticated analyses of recruitment patterns. Scraping data from a job aggregation platform like Simplify.jobs enables the extraction of up-to-date and diverse information about companies and their hiring practices. By leveraging this method, the study taps into a wide variety of data points that might otherwise be difficult to compile manually. This dynamic data collection process not only makes the dataset highly relevant but also ensures its timeliness, allowing for a more accurate reflection of current trends in recruitment behavior across different industries. As such, this methodology serves as the backbone for the data-driven analysis that follows.

Exploratory Data Analysis (EDA)

EDA serves as a foundational step in understanding the intricacies of a dataset before diving into more complex analyses. In this study, EDA was performed on the dataset to clean and preprocess the data, ensuring its readiness for further analysis. The first phase of this process involved identifying and addressing missing values, which is crucial for maintaining the integrity of the dataset. Missing data can distort statistical analyses, leading to biased or incomplete results. To handle this, the dataset was systematically examined for any null or absent values across the various columns. In cases where missing values were detected, a decision was made either to impute values (using strategies such as mean or median imputation for numerical attributes) or to drop rows that were deemed critical to the analysis.

Data preprocessing also extended to handling outliers, which can disproportionately influence statistical measures and affect the overall distribution of variables. Outlier detection was performed using methods such as the Interquartile Range (IQR) and z-scores to identify extreme values that lay outside of typical distribution ranges. In certain instances, these outliers were removed from the dataset to preserve the accuracy of the analyses. In other cases, particularly for categorical variables, the anomalies were retained to ensure that the dataset reflected real-world complexities. The resulting cleaned dataset provided a robust foundation for the subsequent stages of analysis.

Once the dataset was clean, descriptive statistics were computed to summarize and characterize the key variables, offering an initial overview of the data's distribution. The mean, median, and mode were calculated for numerical columns, providing insights into central tendencies. For instance, the average funding total of companies or the most common company size could be quickly identified through this summary. This descriptive analysis revealed patterns such as the skewness of certain variables, which pointed to the prevalence of larger companies with substantial funding compared to smaller, newer organizations. Visualizations, including histograms and box plots, were then employed to illustrate these distributions, providing a more intuitive understanding of the data. Figure 2 below shown distribution of three numerical columns: year_founded, company_size, and funding_total.

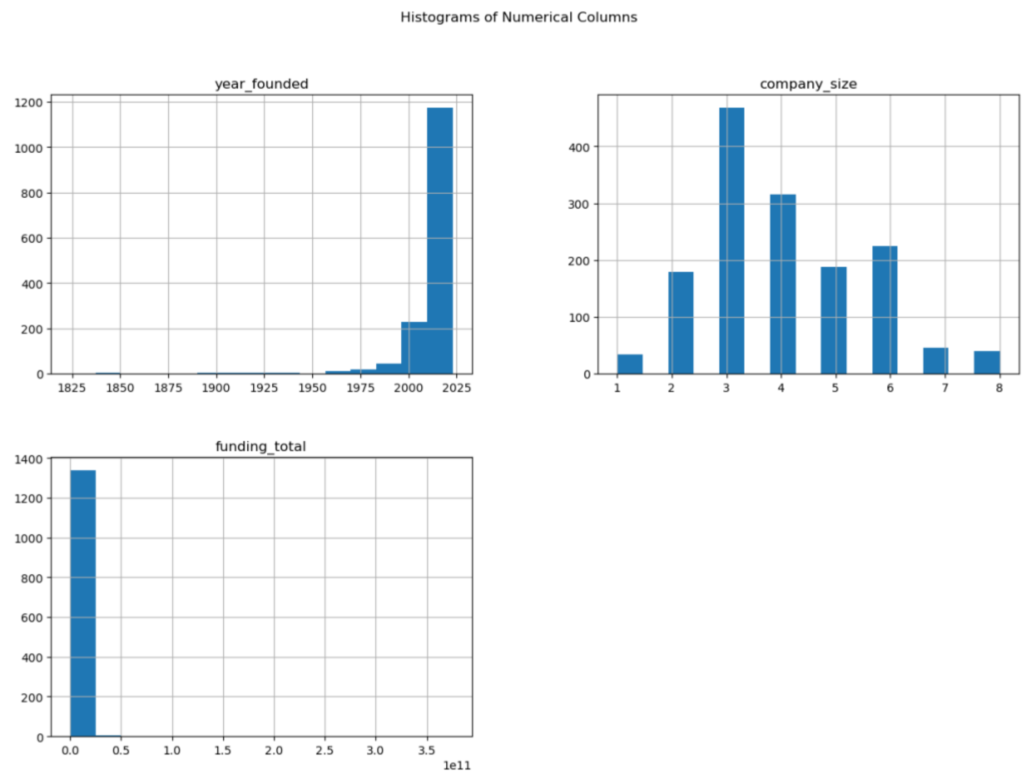

Figure 2 Histogram of Numerical Columns

The histogram shows a skewed distribution, with most companies in the dataset being founded after the year 2000. There is a sharp increase in the number of companies founded in the 2000s, with a large concentration between 2010 and 2025. The data shows very few companies founded before the 20th century (indicated by the small number of entries before 1900), highlighting the relatively recent surge in new companies, particularly in the tech sector. The distribution of company sizes is quite spread out, with a peak around company sizes 3 and 4. Most companies in the dataset fall within the range of 1-4 employees, suggesting that small to medium-sized companies dominate this dataset. Larger companies (with more than 5 employees) are less common, which could indicate that many of the companies in the dataset are startups or smaller enterprises. This histogram also displays a highly skewed distribution for funding data. Most companies in the dataset have very low funding totals, concentrated around 0, which may indicate that most companies are either at early stages with little funding or haven't raised large amounts of capital. On the other hand, a small number of companies appear to have received substantial funding, with a tail extending towards higher funding amounts, suggesting the presence of a few highly funded companies or unicorns in the dataset.

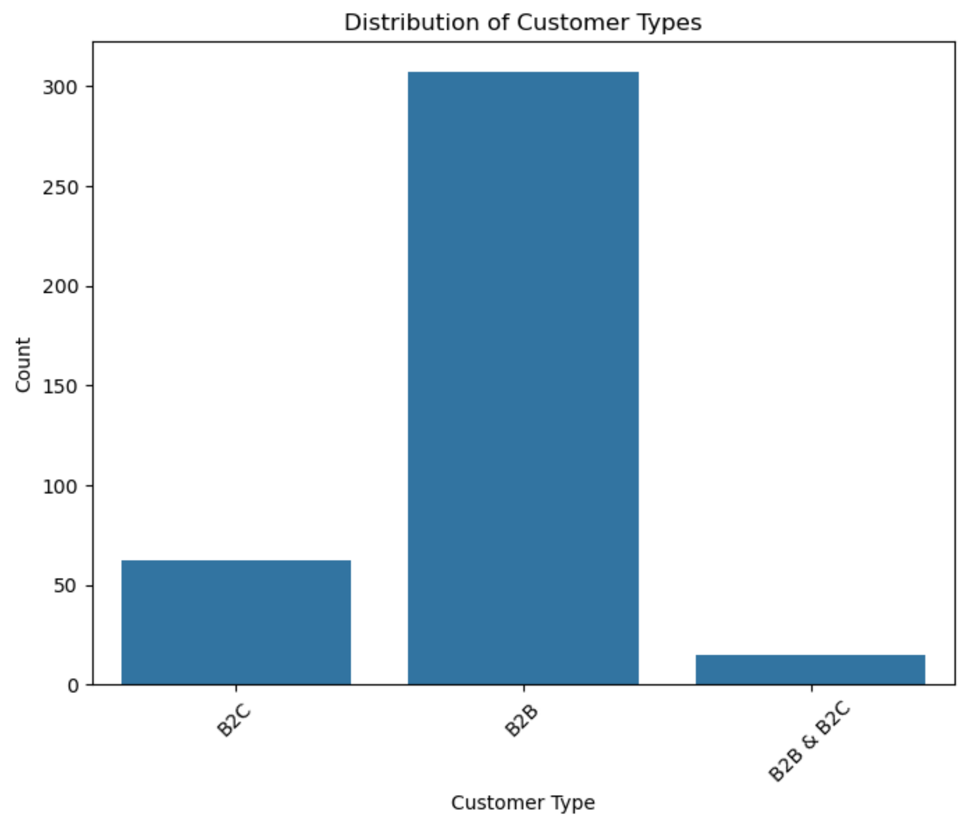

Figure 3 Distribution of Customer Types

Figure 3 illustrates the distribution of customer types in the dataset, which are categorized into three groups: B2C (Business to Consumer), B2B (Business to Business), and B2B & B2C (a combination of both). The chart reveals a clear dominance of companies targeting B2B customers, with over 300 entries in this category. In contrast, the B2C category, representing companies focused on direct consumer sales, contains significantly fewer companies, around 50 in total. The B2B & B2C category has the smallest count, indicating that companies catering to both business and consumer markets are relatively rare in this dataset. The distribution shows that most companies in this dataset are oriented towards business clients (B2B), which could reflect the prominence of B2B services in certain industries, such as technology or industrial sectors. The limited number of B2C and B2B & B2C companies may suggest a narrower focus on consumer-based businesses or hybrid business models in the dataset, highlighting a potential opportunity for further research into these less-represented categories.

Clustering Methodology

The K-Means clustering algorithm serves as a powerful technique for grouping companies based on characteristics such as funding stage, customer type, and job posting frequency. Clustering provides a data-driven way to segment organizations, revealing patterns in how companies approach recruitment and resource allocation. By leveraging features that reflect both company structure and operational focus, this method enables a deeper understanding of the similarities and differences across organizations in terms of their recruitment behaviors. The ability to categorize companies based on these features allows for more targeted and efficient recruitment strategies, ensuring that resources are directed towards companies with similar needs and operational profiles. Moreover, K-Means clustering helps identify groups that might otherwise remain hidden, providing valuable insights into industry-specific trends and recruitment priorities.

The first critical step in the clustering process is preparing the dataset for analysis. This involves encoding categorical variables, such as `funding_stage` and `customer_type`, which are essential features for understanding company profiles. For this, we applied LabelEncoder, a common method for converting categorical data into numeric form, ensuring that the K-Means algorithm can process these attributes effectively. Missing values in categorical features were imputed with a placeholder value, “Unknown,” to ensure that no data was lost during preprocessing. Additionally, job posting frequency, a proxy for recruitment activity, was generated for each company. Although the dataset did not contain an explicit column for job postings, we simulated this feature by randomly assigning job posting frequencies, which reflects the hiring activity of companies within the dataset. By creating this synthetic feature, the analysis could proceed as if all relevant data was present, ensuring the robustness of the clustering approach.

To address the inherent sensitivity of K-Means to the scale of data, normalization was performed using StandardScaler. Scaling the features ensures that all variables contribute equally to the clustering process, preventing features with larger numerical ranges (like `funding_total`) from dominating the results. After scaling, the dataset was ready for clustering analysis, where the next challenge was determining the optimal number of clusters. This was achieved using the elbow method, which involves running K-Means for a range of cluster numbers and plotting the inertia (within-cluster sum of squares). The elbow method helps identify the point where adding more clusters no longer significantly reduces inertia, indicating the optimal number of clusters. Through this process, the elbow plot revealed that three clusters provided the best balance between complexity and explanatory power, guiding the final choice of `k=3` for our clustering model.

Once the optimal number of clusters was selected, the K-Means algorithm was applied to segment the companies based on the chosen features. The resulting clusters reflect different recruitment behaviors, influenced by company size, funding stage, and market focus. The clusters were visualized using scatter plots and pair plots, which displayed the distribution of companies within each cluster across the three key features. These visualizations offer a clear view of how companies with similar characteristics tend to recruit, highlighting patterns in job posting frequency and other organizational factors. Ultimately, this clustering methodology provides actionable insights into how companies from different sectors approach recruitment, which can inform more nuanced strategies for talent acquisition and organizational alignment.

Association Rule Mining

Association rule mining is a pivotal technique used to uncover the relationships between job attributes such as job titles, benefits offered, and company values. The primary goal of this approach is to identify frequent co-occurrences within these attributes, allowing for the extraction of meaningful patterns that might otherwise remain unnoticed. By utilizing the Apriori algorithm, which identifies frequent itemsets in transactional data, organizations can gain insights into how job characteristics correlate with each other, thereby informing better recruitment strategies. This methodology has profound implications for understanding how benefits and job titles align with organizational culture and values, creating a clearer picture of what drives employee satisfaction and retention. In recruitment, such insights can guide HR teams in crafting more appealing job postings and benefits packages that resonate with prospective candidates, enhancing both talent acquisition and retention efforts.

The process of association rule mining begins with data preprocessing, focusing on attributes like `benefits` and `values`, which provide critical context for understanding job postings. For this, missing values in the dataset were first handled by imputing them with a placeholder such as "Unknown," ensuring that the dataset remained intact for analysis. These attributes were then transformed into a suitable format for the Apriori algorithm using the TransactionEncoder, which converts categorical data into binary indicators—each row representing a transaction or job posting, and each column representing a job attribute. This transformation is necessary because the Apriori algorithm operates on data in the form of transaction-item matrices, where each item corresponds to a job attribute like a benefit or value. By employing this preprocessing approach, the dataset becomes compatible with the Apriori algorithm, which subsequently generates the frequent itemsets that underpin the association rules.

Once the data was prepared, the Apriori algorithm was applied to identify frequent itemsets from the job attributes. These itemsets represent combinations of benefits, job titles, and company values that appear together frequently across the dataset. For instance, the algorithm might reveal that certain job titles, like “Software Engineer,” are frequently associated with specific benefits such as "health insurance" and company values like "innovation." The identification of such frequent patterns is foundational for generating association rules, which can further highlight the relationships between job attributes. The rules are typically evaluated based on metrics such as support, confidence, and lift, each serving to quantify the strength and relevance of the identified associations. Support measures the frequency of the itemset in the dataset, confidence gauges the likelihood that one item follows from another, and lift evaluates the strength of the association beyond random chance. Through these metrics, the generated rules provide actionable insights for recruitment professionals seeking to optimize job postings based on the most frequent and impactful combinations of attributes.

To visualize the results, we employed various graphical techniques, such as bar charts and heatmaps, to depict the relationships between the most frequent itemsets. These visualizations help illuminate key patterns and highlight the strongest associations among job titles, benefits, and company values. Such insights are invaluable for HR departments aiming to align their recruitment efforts with the values and preferences of their target candidates. Furthermore, by understanding which benefits are most associated with specific job titles or company profiles, organizations can tailor their job postings more precisely to attract the right talent. In conclusion, the application of association rule mining in this context provides not only a deeper understanding of the factors that influence recruitment patterns but also a strategic framework for enhancing the alignment between organizational offerings and candidate expectations.

Data Visualization

To extract meaningful insights from complex datasets, effective data visualization becomes a crucial tool. The use of visual aids such as scatter plots, bar charts, and heatmaps allows researchers to make sense of clustering results and association rules, transforming raw data into comprehensible and actionable knowledge. These visualizations are essential not only for communicating findings but also for uncovering hidden patterns and correlations within the data. For example, in the context of this study, scatter plots provide an effective way to observe the relationships between organizational characteristics—such as funding stage and customer type—and recruitment behaviors, like job posting frequency and benefits offered. Such visual techniques help to distill multidimensional data into a more digestible format, facilitating the identification of key trends that would otherwise be difficult to discern.

When investigating clustering results, a scatter plot can visually represent the different groups formed by K-Means clustering, with companies clustered based on features like funding stage, job posting frequency, and customer type. Standardizing the features before applying clustering is crucial, as it ensures that no single feature disproportionately influences the clustering outcome due to differences in scale. This step was carried out using the StandardScaler function from scikit-learn, which standardizes the data to have a mean of zero and a standard deviation of one. The resulting visual representation of clusters can then be analyzed to understand how similar companies are grouped together based on their recruitment behaviors, which in turn can inform tailored recruitment strategies. Such visual clarity enhances interpretability, making the findings accessible even to non-technical stakeholders.

Bar charts and heatmaps further complement clustering analysis by providing a clear view of categorical relationships and correlations between job attributes. Bar charts are an effective way to display the distribution of job titles across different clusters, revealing which job categories are more commonly posted by companies of a certain funding stage or customer type. Heatmaps, on the other hand, offer an intuitive representation of correlation matrices, helping to visualize the strength of associations between various company attributes and job posting patterns. These tools allow researchers to identify, for example, whether companies in a particular sector tend to offer specific benefits or if there are consistent trends in job posting frequency depending on organizational size or funding stage. The use of color gradients in heatmaps provides an immediate visual cue of the strength of these correlations, making it easier to identify significant patterns.

Together, these visual tools empower researchers to not only present complex findings in a comprehensible manner but also to derive deeper insights from the data. Through scatter plots, bar charts, and heatmaps, this study brings clarity to the often chaotic landscape of recruitment data, highlighting the relationships between company profiles and job posting behaviors. These visualizations not only make the analysis more accessible but also provide a roadmap for future research and data-driven recruitment strategies. In doing so, they enhance the interpretability of advanced techniques such as clustering and association rule mining, transforming abstract data into concrete, actionable insights.

Result and Discussion

Clustering Results

Figure 4 below illustrates the Elbow Method used to determine the optimal number of clusters (K) for K-Means clustering. In the graph, inertia (the sum of squared distances between points and their centroids) is plotted against the number of clusters (K). The sharp drop in inertia as K increases suggests that clustering with a higher number of clusters initially provides a significant reduction in the within-cluster variance.

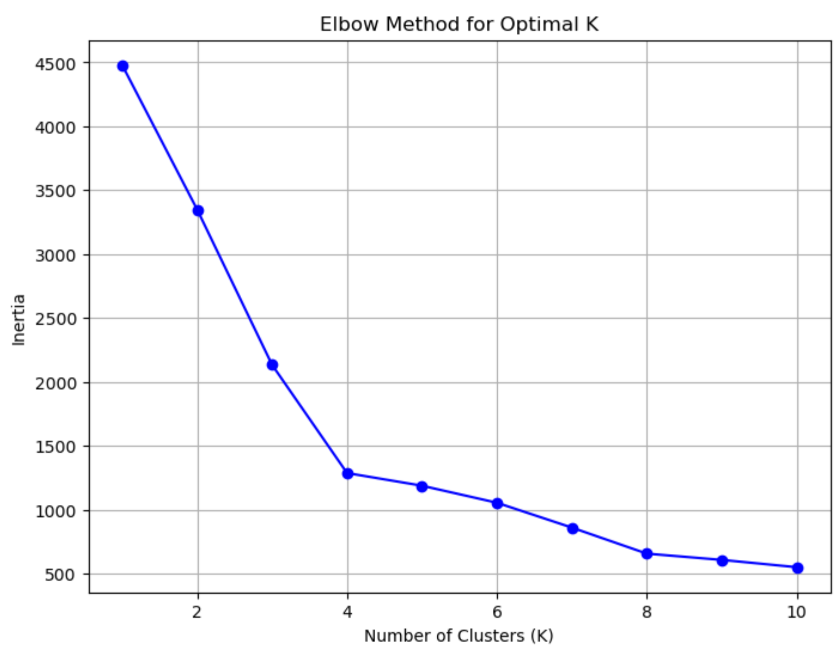

Figure 4 Elbow Method for Optimal K

However, beyond K = 3, the reduction in inertia becomes less pronounced, indicating that additional clusters do not improve the model as effectively. This point, where the curve flattens out and the diminishing returns become apparent, represents the "elbow" of the graph. Based on this analysis, K = 3 is identified as the optimal number of clusters, as it provides a balance between simplicity and sufficient granularity in clustering the data.

The K-Means clustering analysis revealed distinct groupings of companies based on their organizational characteristics, specifically their funding stage, customer type, and job posting frequency. By segmenting companies according to these features, the clustering algorithm successfully identified three distinct clusters, each representing companies with similar profiles. The first cluster, characterized by a high funding stage and moderate customer type, tends to consist of companies with significant financial backing, often focusing on specific industry niches. The second cluster, marked by a lower funding stage and varied customer type, represents early-stage companies with more flexible, adaptive strategies. The third cluster, identified by relatively low job posting frequency, predominantly includes companies that rely on more specialized roles, offering higher compensation packages but fewer openings. This segmentation highlights the diverse recruitment strategies and resource allocations that exist across different company profiles, making it clear that organizational characteristics heavily influence recruitment behavior.

To visually illustrate the clustering results, a scatter plot matrix was generated (Figure 5), which presents the relationships between key variables: funding stage, customer type, and job posting frequency. In this matrix, companies are color-coded according to their cluster membership, revealing how companies within the same cluster exhibit similar recruitment patterns. For instance, the first cluster, represented by red dots, tends to cluster around higher values of funding stage and customer type, suggesting that these companies have the resources and market reach to engage in more frequent and diverse job postings. In contrast, the blue and green clusters indicate companies with lower financial resources and different customer engagement strategies. Such visualizations reinforce the notion that companies with similar profiles—whether in terms of financial resources or market focus—tend to adopt similar recruitment strategies.

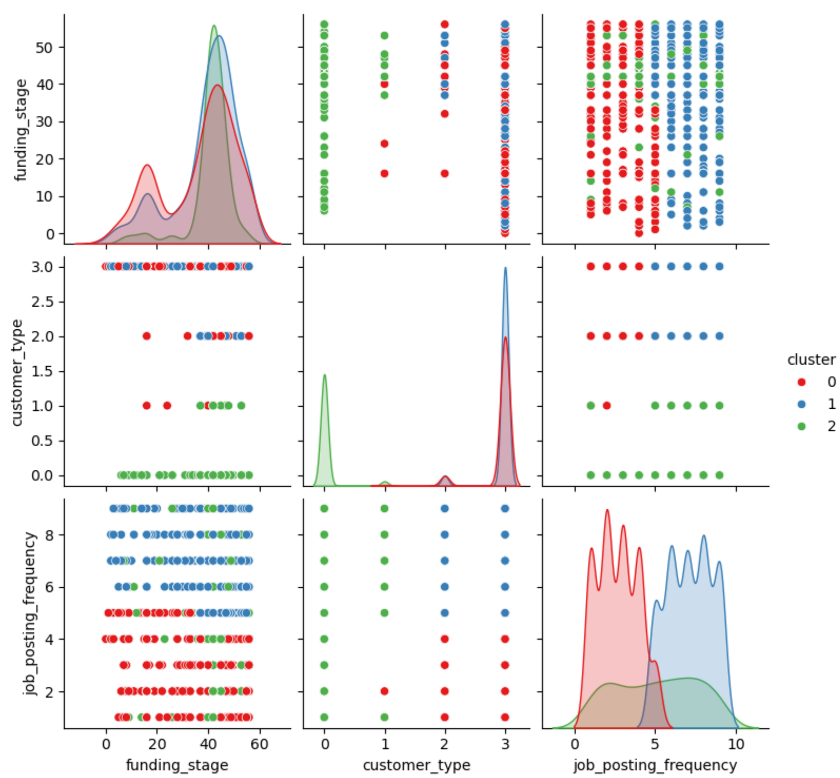

Figure 5 **Scatter Plot Matrix of Clustering**

The pairwise scatter plots within the matrix offer a deeper understanding of how these clusters differ along multiple dimensions. For example, the distribution of funding stage and customer type clearly shows a dichotomy between more established companies and newer startups. Companies in the red cluster exhibit more centralized funding and a focus on specific customer types, while those in the blue and green clusters show more variability in both attributes. This variance highlights how recruitment strategies are not solely determined by the size or age of a company but also by its specific market orientation and customer focus. These insights are crucial for HR departments, as they allow for a more nuanced understanding of the factors that influence recruitment behaviors and job posting patterns.

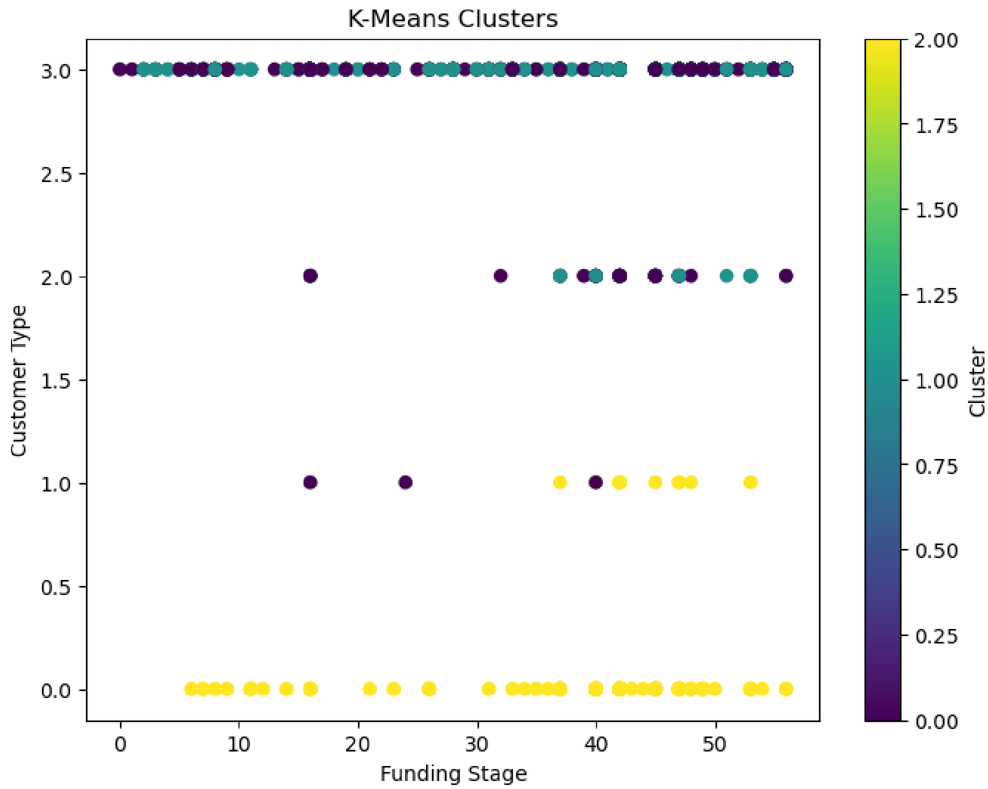

Figure 6 **Clustering of Customer Type and Funding Stage**

Furthermore, the clustering results (Figure 6) provide valuable insights into how job posting frequency relates to company size and financial backing. Companies with greater financial resources (represented by higher funding stages) tend to have higher job posting frequencies, a reflection of their ability to scale rapidly and recruit at larger volumes. The heatmap of feature correlations corroborates this finding, where the strongest correlations were observed between job posting frequency and both funding stage and customer type. This suggests that companies with more financial backing not only have more job openings but also engage in more targeted recruitment efforts to support their growth strategies. These findings underscore the importance of considering both organizational characteristics and recruitment behavior when crafting talent acquisition strategies.

Association Rule Findings

The application of association rule mining to the dataset yielded insightful relationships between various job attributes, particularly focusing on the combinations of job titles, benefits, and company values. The most significant association rules, with high support and confidence, revealed clear patterns in how certain job attributes co-occur. For example, one of the strongest rules identified was that "companies in the 'Tech' sector frequently offer remote work options." This rule suggests that firms operating within the technology industry prioritize flexibility and work-life balance, offering remote work opportunities to attract top talent. These findings not only reflect industry-specific recruitment trends but also underscore the importance of tailoring benefits to sector-specific needs, especially in a post-pandemic job market that increasingly values remote work options. This rule, along with others, highlights how job characteristics are not randomly distributed but rather exhibit strong dependencies based on industry and organizational values.

The frequent itemsets discovered during the association rule mining process further reinforced this idea, revealing key combinations of job attributes that appeared together with significant regularity. For instance, it was found that companies offering "competitive compensation" were also likely to provide "health insurance" and "stock options" as part of their benefits package. This combination suggests that companies in certain industries, particularly those in high-demand sectors like technology and finance, use financial incentives as a strategy to retain and motivate employees. By aligning their compensation offerings with candidate expectations and market standards, these companies create a value proposition that attracts highly skilled professionals, making them competitive in the talent market. This association rule emphasizes the role of comprehensive benefits packages in shaping recruitment outcomes.

To better visualize the relationships between these job attributes, network diagrams were employed, which illustrate the strength of connections between job titles, benefits, and company values. These diagrams provide a clear representation of how different attributes cluster together, offering a visual exploration of the most frequent associations. For example, the network diagram revealed that companies offering positions such as "Software Engineer" or "Data Scientist" are strongly linked with benefits like "health insurance," "stock options," and "flexible working hours." The thickness of the lines in the network diagram indicates the strength of these associations, where thicker lines represent stronger and more frequent co-occurrences. Such visualizations help contextualize the data, allowing for a deeper understanding of how recruitment strategies are crafted and how certain job attributes are more likely to appear together than others.

These findings underscore the utility of association rule mining in recruitment analytics, as they allow companies to better understand the combinations of job attributes that influence candidate choices and organizational preferences. By identifying these associations, organizations can more effectively tailor their job postings to meet the expectations of potential candidates. The visualization of these relationships through network diagrams further enhances the clarity of the analysis, enabling recruitment professionals to leverage data-driven insights in shaping their talent acquisition strategies. This approach, grounded in robust statistical methods like association rule mining, provides a powerful tool for understanding the complexities of recruitment dynamics, paving the way for more informed and strategic decision-making in the ever-competitive job market.

Discussion

The results from the clustering and association rule mining analyses reveal intriguing patterns in how companies adjust their recruitment strategies based on their funding stages and organizational characteristics. A key finding from the clustering analysis is that companies in early funding stages tend to focus their hiring efforts on technical roles, with job postings primarily related to software development, engineering, and IT. These companies, often characterized by limited financial resources and a need for rapid product development, prioritize technical expertise to fuel innovation and growth. This aligns with existing literature, which suggests that startups and early-stage companies concentrate on hiring employees who can contribute directly to product development and technological advancement. In contrast, companies in later funding stages, which have established their market presence and possess greater financial resources, tend to offer a more diverse range of benefits, such as health insurance, retirement plans, and flexible work arrangements. These benefits reflect the need to attract a wider talent pool, emphasizing employee retention and long-term organizational growth.

Moreover, the association rule mining results support this observation, revealing that early-stage companies frequently offer fewer benefits, with a focus on financial incentives tied to stock options or equity. The minimalistic benefits offered by these companies are consistent with their strategic priorities: attracting talent through opportunities for future financial gain, often in the form of company equity. This strategy is particularly effective in industries like technology, where equity offers become a valuable tool in enticing skilled individuals to join small, high-risk ventures. Later-stage companies, on the other hand, tend to bundle job titles with a broader array of benefits, highlighting their ability to offer both immediate financial rewards and long-term job security. This trend underscores the shift in organizational priorities as companies evolve, from short-term resource allocation to long-term workforce satisfaction.

The comparison between the results of this study and the existing literature underscores a crucial evolution in recruitment strategies. For example, previous studies have documented that early-stage companies, typically constrained by funding, focus more on essential technical skills and less on comprehensive benefits. However, this study uncovers a more refined understanding of how benefits packages evolve as companies grow. Later-stage firms not only provide more extensive benefits but also tend to offer more flexible working conditions, such as remote work options, in line with global shifts towards work-life balance and employee well-being. This trend is becoming increasingly critical in attracting top-tier talent, especially in industries like tech, where the competition for skilled labor is intense. Our findings contribute to the growing body of literature by offering deeper insights into how company lifecycle stages influence not only the type of roles posted but also the structure of compensation and benefits packages.

One novel pattern emerging from this research is the distinction between customer type and job posting frequency. Companies that serve a B2B (business-to-business) market, as opposed to a B2C (business-to-consumer) market, tend to have more stable and frequent job postings. This may be since B2B companies often scale more slowly but require highly specialized roles to cater to a smaller, more niche clientele. The hiring trends in B2B companies are often less volatile, with a focus on consistent, long-term talent acquisition, while B2C companies may experience more fluctuations in hiring due to rapidly changing consumer demands and marketing strategies. This is an intriguing pattern that could provide valuable guidance for companies seeking to optimize their recruitment strategies based on their market orientation, further extending the work of previous researchers who have focused predominantly on company size and financial factors.

Conclusion

This study offers a detailed analysis of how company characteristics, such as funding stage, customer type, and company size, influence recruitment behaviors and job posting trends. The clustering analysis revealed that early-stage companies prioritize technical roles, focusing on key positions essential for their growth and development. In contrast, companies in later funding stages provide a more comprehensive benefits package, including health insurance, retirement plans, and flexible working arrangements, reflecting their ability to attract and retain a broader talent pool. Association rule mining further uncovered significant relationships, such as companies in the tech sector frequently offering remote work options, highlighting how industry-specific needs shape recruitment strategies. These findings underscore the dynamic interplay between organizational characteristics and recruitment strategies, emphasizing that companies at different stages of their lifecycle adopt distinct approaches to talent acquisition.

The insights from this study provide valuable guidance for HR professionals and recruiters seeking to align their recruitment strategies with the financial and organizational profiles of their companies. By understanding that early-stage companies may not have the resources to offer extensive benefits but can attract talent through equity and job growth opportunities, recruiters can tailor their approaches to appeal to the unique priorities of such organizations. Similarly, understanding that later-stage companies often offer a broader range of benefits can help recruiters identify key competitive advantages in attracting high-quality candidates. Additionally, the study's findings on the correlation between job posting frequency and company characteristics, such as customer type and funding stage, offer a strategic framework for tailoring recruitment campaigns to specific company profiles. Ultimately, these insights enable organizations to optimize their recruitment efforts, ensuring that their strategies are closely aligned with the characteristics of the companies they seek to build.

Despite its contributions, this study is not without limitations. One key limitation is the reliance on a single dataset obtained from Simplify.jobs, which may not fully capture the diversity of recruitment behaviors across all industries or regions. The data's scope, while extensive, only represents a snapshot of recruitment practices within the dataset's timeframe, limiting its generalizability. Additionally, the lack of temporal analysis means that we cannot observe how recruitment trends evolve over time, which could provide deeper insights into how companies' hiring strategies adapt to market conditions or economic shifts. Future research could address these limitations by applying similar methodologies to datasets from multiple sources, offering a broader perspective on recruitment trends across industries. Incorporating a longitudinal approach could also enable researchers to track how recruitment behaviors evolve in response to external factors, such as economic downturns or technological advancements. Furthermore, extending this research to other industries, such as healthcare or manufacturing, would provide a more comprehensive understanding of how different sectors tailor their recruitment strategies based on company characteristics. This would not only enrich the findings of this study but also expand the applicability of the results across various business contexts.

Declarations

Author Contributions

Conceptualization: M.T.L.; Methodology: T.H.; Software: T.H.; Validation: M.T.L.; Formal Analysis: M.T.L.; Investigation: T.H.; Resources: M.T.L.; Data Curation: T.H.; Writing Original Draft Preparation: M.T.L.; Writing Review and Editing: T.H.; Visualization: M.T.L.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Amamou, "How Artificial Intelligence Affects Various Aspects of Recruitment Processes: The Case of Tunisian Banks," *Int. J. Membr. Sci. Technol.*, vol. 10, no. 2, pp. 3694–3700, 2023, doi: 10.15379/ijmst.v10i2.3209.

[2] W. Liu, W. W. Xu, and B. St. John, "Organizational Disaster Communication Ecology: Examining Interagency Coordination on Social Media During the Onset of the COVID-19 Pandemic," *Am. Behav. Sci.*, 2021, doi: 10.1177/0002764221992823.

[3] E. Miller, J. L. Jolly, J. N. Latz, and K. Listman, "Influencers and Major Themes in a Gifted Education Community of Practice on Twitter," *J. Adv. Acad.*, vol. 33, no. 3, pp. 469–504, 2022, doi: 10.1177/1932202x221099590.

[4] D. Binns, "No Free Tickets," *MC J.*, vol. 25, no. 2, 2022, doi: 10.5204/mcj.2882.

[5] N. Y. C. Cheung *et al.*, "Perception of personalized medicine, pharmacogenomics, and genetic testing among undergraduates in Hong Kong," *Hum. Genomics*, vol. 15, no. 1, pp. 1–11, 2021, doi: 10.1186/s40246-021-00353-0.

[6] Luh Putu Saraswati Devia Jayanti and M. Wasesa, "Application of Predictive

Analytics to Improve the Hiring Process in a Telecommunications Company," *J. Coreit J. Has. Penelit. Ilmu Komput. Dan Teknol. Inf.*, vol. 8, no. 1, 2022, doi: 10.24014/coreit.v8i1.16915.

[7] B. Hayadi, H. Henderi, M. Budiarto, and ..., "An Extensive Exploration into the Multifaceted Sentiments Expressed by Users of the myIM3 Mobile Application, Unveiling Complex Emotional Landscapes and …," *J. Appl. …*, no. Query date: 2024-10-12 10:59:38, 2024, [Online]. Available: http://bright-journal.org/Journal/index.php/JADS/article/view/187

[8] S. H. Ali, J. Foreman, A. Capasso, A. M. Jones, Y. Tozan, and R. J. DiClemente, "Social Media as a Recruitment Platform for a Nationwide Online Survey of COVID-19 Knowledge, Beliefs, and Practices in the United States: Methodology and Feasibility Analysis," *BMC Med. Res. Methodol.*, vol. 20, no. 1, 2020, doi: 10.1186/s12874-020-01011-0.

[9] M. Liebenlito, N. Inayah, E. Choerunnisa, T. E. Sutanto, and S. Inna, "Active learning on Indonesian Twitter sentiment analysis using uncertainty sampling," *J. Appl. Data Sci.*, vol. 5, no. 1, Art. no. 1, Jan. 2024, doi: 10.47738/jads.v5i1.144.

[10] A. S. Paramita and T. Hariguna, "Comparison of K-Means and DBSCAN Algorithms for Customer Segmentation in E-commerce," *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 1, Art. no. 1, May 2024, doi: 10.47738/jdmdc.v1i1.3.

[11] B. v. d. Vegt, "An Investigation of Data-Driven Player Positional Roles Within the Australian Football League Women's Competition Using Technical Skill Match-Play Data," *Int. J. Sports Sci. Coach.*, vol. 19, no. 3, pp. 1130–1142, 2023, doi: 10.1177/17479541231203895.

[12] Henderi and Q. Siddique, "Comparative Analysis of Sentiment Classification Techniques on Flipkart Product Reviews: A Study Using Logistic Regression, SVC, Random Forest, and Gradient Boosting," *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 1, Art. no. 1, May 2024, doi: 10.47738/jdmdc.v1i1.4.

[13] Y. Wu, "Research on the Market Testing in the Restaurant Industry," *Adv. Econ. Manag. Polit. Sci.*, vol. 63, no. 1, pp. 320–328, 2023, doi: 10.54254/2754-1169/63/20231459.

[14] T. Hariguna and A. S. M. Al-Rawahna, "Unsupervised Anomaly Detection in Digital Currency Trading: A Clustering and Density-Based Approach Using Bitcoin Data," *J. Curr. Res. Blockchain*, vol. 1, no. 1, Art. no. 1, Jun. 2024, doi: 10.47738/jcrb.v1i1.12.

[15] Hery and A. E. Widjaja, "Predictive Modeling of Blockchain Stability Using Machine Learning to Enhance Network Resilience," *J. Curr. Res. Blockchain*, vol. 1, no. 2, Art. no. 2, Sep. 2024, doi: 10.47738/jcrb.v1i2.15.

[16] C. Izumi and T. Hariguna, "In-Depth Analysis of Web3 Job Market: Insights from Blockchain and Cryptocurrency Employment Landscape," *Int. J. Res. Metaverese*, vol. 1, no. 1, Art. no. 1, Jun. 2024, doi: 10.47738/ijrm.v1i1.4.

[17] J. P. B. Saputra and N. A. Putri, "The Impact of Market Activity on Property Valuations in Digital Real Estate Through a Quantitative Analysis of Bidding and Sales Dynamics," *Int. J. Res. Metaverese*, vol. 1, no. 2, Art. no. 2, Sep. 2024, doi: 10.47738/ijrm.v1i2.11.

[18] Y. Durachman and A. W. B. A. Rahman, "Clustering Student Behavioral Patterns: A Data Mining Approach Using K-Means for Analyzing Study Hours, Attendance, and Tutoring Sessions in Educational Achievement," *Artif. Intell. Learn.*, vol. 1, no. 1, Art. no. 1, Nov. 2024.

[19] T. Sangsawang, "Predicting Student Achievement Using Socioeconomic and School-Level Factors," *Artif. Intell. Learn.*, vol. 1, no. 1, Art. no. 1, Nov. 2024.

[20] M. Lukauskas, V. Šarkauskaitė, V. Pilinkienė, A. Stundžienė, A. Grybauskas, and J. Bruneckienė, "Enhancing Skills Demand Understanding Through Job Ad Segmentation Using NLP and Clustering Techniques," *Appl. Sci.*, vol. 13, no. 10, p. 6119, 2023, doi: 10.3390/app13106119.

[21] T. H. Davenport, A. Guha, D. Grewal, and T. Breßgott, "How Artificial Intelligence Will Change the Future of Marketing," *J. Acad. Mark. Sci.*, 2019, doi: 10.1007/s11747-019-00696-0.

[22] Y. E. Zhang, S. Liu, S. Xu, M. M. Yang, and J. Zhang, "Integrating the Split/Analyze/Meta-Analyze (SAM) Approach and a Multilevel Framework to Advance Big Data Research in Psychology," *Z. Für Psychol.*, 2018, doi: 10.1027/2151-2604/a000345.
[23] S. A. Olaleye *et al.*, "Association Rule Mining for Job Seekers' Profiles Based on Personality Traits and Facebook Usage," *Int. J. Bus. Inf. Syst.*, vol. 40, no. 3, p. 299, 2022, doi: 10.1504/ijbis.2022.124933.
[24] C. Zhang, S. A. Hosseini, S. Weingärtner, K. Ugurbil, S. Moeller, and M. Akçakaya, "Optimized Fast GPU Implementation of Robust Artificial-Neural-Networks for K-Space Interpolation (RAKI) Reconstruction," *Plos One*, 2019, doi: 10.1371/journal.pone.0223315.
[25] P. Ampornphan and S. Tongngam, "Exploring Technology Influencers From Patent Data Using Association Rule Mining and Social Network Analysis," *Information*, vol. 11, no. 6, p. 333, 2020, doi: 10.3390/info11060333.
[26] D. Arifka, M. N. Hakim, A. S. Adhipta, K. S. S. Yogananda, R. Salsabila, and R. Ferdiana, "Pandemic Fatigue: An Analysis of Twitter Users' Sentiments Against the COVID-19 in Indonesia," *J. Psikol.*, vol. 49, no. 2, p. 182, 2022, doi: 10.22146/jpsi.71979.
[27] V. Wadhwa *et al.*, "Physician Sentiment Toward Artificial Intelligence (AI) in Colonoscopic Practice: A Survey of US Gastroenterologists," *Endosc. Int. Open*, vol. 08, no. 10, pp. E1379–E1384, 2020, doi: 10.1055/a-1223-1926.
[28] T. Yiğitcanlar, K. Degirmenci, and T. Inkinen, "Drivers Behind the Public Perception of Artificial Intelligence: Insights From Major Australian Cities," *Ai Soc.*, vol. 39, no. 3, pp. 833–853, 2022, doi: 10.1007/s00146-022-01566-0.
[29] C. J. Hutto and É. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," *Proc. Int. Aaai Conf. Web Soc. Media*, vol. 8, no. 1, pp. 216–225, 2014, doi: 10.1609/icwsm.v8i1.14550.