Unlocking Opportunities: A comprehensive Analysis of LinkedIn Job Postings

Aiswarya Sen
Master's in Information systems
George Mason University, Fairfax, Virginia, USA
asen3@gmu.edu

Narasimha Krishna Amruth Vemuganti
Master's in Data Analytics Engineering
George Mason University, Fairfax, Virginia, USA
nvemugan@gmu.edu

Monali Sheetal Kathepaga
Master's in Data Analytics Engineering
George Mason University, Fairfax, Virginia, USA
mkathepa@gmu.edu

AIT 664: Represent, Process & Visualize Applied Information Technology
Prof. Ebrima N Ceesay

Abstract: Data-driven insights play a critical role in creating strategies for talent acquisition, workforce planning, and career decisions in the present dynamic and growing labor market scenario. This research presents a comprehensive analysis of the LinkedIn dataset for the year 2023, aiming to uncover emerging trends and valuable insights within the professional networking platform. The dataset, comprising diverse user profiles and engagement metrics, is meticulously examined using advanced data analytics and visualization techniques. The issue statement emphasizes the significance of answering important concerns regarding the employment environment in 2023 and giving governments, companies, and job searchers vital information. The project seeks to provide decision-making tools to individuals, organizations, and governments by navigating various factors, so promoting a flexible and resilient labor market. This project aims to investigate the sectors, jobs from various industries that are in high demand, which industries are seeing the fastest job growth, how employment possibilities vary by location, and more. Our goal in doing this research is to improve knowledge about the state of the labor market today and how it affects companies and job seekers alike. For people navigating the labor market, choosing careers, and creating data-driven workforce and recruiting plans going forward, the resultant study will be an invaluable resource. In summary, by offering practical insights, this initiative seeks to close the knowledge gap between developments in the labor market and career choices.

Keywords—Data Analytics, Visualization, Labor Market, Talent Acquisition, Decision-Making Tools, Employment.

I. INTRODUCTION

In today's dynamic and ever-evolving job market, data-driven insights play a pivotal role in shaping talent acquisition strategies, workforce planning, and career decisions. Social network is a web-based service which allows users to create a public/semi-public or professional profile within a domain such that they can connect with other individuals within the network.[11] Recruitment and Job seeking are two major factors that are directly proportional to each other. Due to the competitive nature of the present world, the process of acquiring the best resource effectively and efficiently has become a challenging aspect for the companies. As a result, modern job portals have become increasingly popular to address the challenges identified in the early recruitment and job search process.[19] Facebook, Twitter, LinkedIn, Google+ etc. are very popular social websites. Facebook has more than 1000 million active users. These users are increasing about 85% annually since 2008.[12] This incredibly large and fast-growing demand has enticed software giants such as Google and Facebook to enter this space, which was previously dominated by companies such as LinkedIn, indeed, Dice and CareerBuilder.[2]

LinkedIn is a social networking site focused on professional and business relationships.[14] It is the world's largest professional networking platform, continuously collects and updates job postings from millions of companies across various industries and regions. [10] The LinkedIn Job Postings - 2023 dataset is a snapshot of this sizable database that reflects the situation of the labor market as of 2023. It includes a wide range of details about the jobs available, such as job names, company information, job descriptions, necessary skills, experience levels, and locations.

The LinkedIn employment Postings dataset is a priceless tool that offers a thorough overview of the worldwide employment market and is a goldmine of data for researchers, companies, policymakers, and job seekers alike. Every day, thousands of businesses and people use LinkedIn to find talents. New skills and knowledge are needed for jobs in the future due in part to the rapid development of workplace technology such as artificial intelligence and internet of things. Jobs in the future will likely require skills that are not taught in schools nor in standard training programs. Instead, job seekers will have to either upskill as they move to new jobs within the same industry

or reskill themselves through the lifelong learning process to move to another industry. Therefore, studying how users acquire skills in their lifelong learning process and how those skills affect their future career is of crucial importance to the world economy. Fortunately, the big data revolution has created an opportunity for researchers to collect and analyze large-scale data to understand the evolving labor market landscape and the upskilling and reskilling processes. [17]

This project's objective is to examine several topics, including the most in-demand talents, the sectors with the fastest rate of job growth, geographical differences in employment possibilities, and more. By doing this, we hope to help people better understand the state of the employment market now and how it will affect both employers and job seekers in future. This report will be an invaluable tool for anyone trying to successfully navigate the job market in future, make wise career decisions, or create data-driven recruitment and workforce initiatives.

II. RESEARCH AND PROBLEM STATEMENT

This analysis aims to answer key questions about the state of employment in 2023 and provide valuable information to job seekers, employers, and policymakers by identifying market trends, job demands, and improving the overall job satisfaction experience.

A.Why is the problem important?

The year 2023 is anticipated to be a pivotal moment in the labor market, marked by a nuanced equilibrium between obstacles and prospects. The economic recession and other recent turbulent events, such as mass layoffs, have had a significant impact on people's lives. Rapid job growth was a feature of the economy throughout 2021 and 2022. The U.S. Bureau of Labor Statistics reported that for all of 2022, non-farm payrolls grew by an average of 399,000 jobs per month. This represented a decline from 2021's rapid job growth, which topped 605,000 per month. In 2023, the pace of new job growth is even slower, though it continues to demonstrate stability. For the ten months ending in October, the monthly average dropped to 238,800, roughly 20,000 fewer jobs added per month than the year-to-date average through September. [18]

These upheavals have shaken the foundations of financial security in addition to changing career paths. Some people are bravely pursuing new career paths in the face of uncertainty during these changes, while others are taking advantage of newly presented chances for development and adaptation. This complex environment is a sobering reminder of how important it is for people and societies to keep funding programs, training, and policies that promote everyone's economic security and general well-being.

B. The 2023 Labor Market: An Intricate Tapestry

1. The Winds of Change: The winds of change starkly contrast in the 2023 labor market. Recent economic turmoil, marked by layoffs and a recession, has upended life and career. Many individuals now grapple with uncertainty as they navigate the maze of career transitions. This highlights people's resilience in the face of hardship—an evident display of unwavering determination to turn difficulties into spurs for advancement and adaptation.

2. Investing in Stability and Well-Being: In the face of current challenges, individuals and societies must steadfastly commit to funding policies, advancing education, and providing training for the enhancement of both economic stability and well-being. These investments form the cornerstone, enabling people to seize new opportunities and navigate the labor market's volatility. Long-term economic stability hinges on prioritizing expenditure on education and training to broaden horizons, acquire new skills, and adapt to evolving industry trends.

3. The Analytical Environment: Comprehending the Labor Market Dynamics of 2023 A thorough analysis is necessary considering this complex labor market situation. The analysis is designed to provide important insights into the employment situation in 2023. Its main objective is to give policymakers, employers, and job seekers insightful information. It aims to improve the overall experience of job satisfaction by analyzing market trends and job demands.

4. Job Market Insights: Analyzing the LinkedIn dataset is a key aspect of our effort to gain profound insights into the current job market. This analysis allows us to unveil real-time trends across various industries, providing valuable information on sectors experiencing growth and those encountering challenges. Scrutinizing the data empowers us to assess the overall health of different industries, distinguishing those flourishing from those requiring transformation.

In this dynamic landscape, job seekers benefit significantly from access to these insights. Understanding which industries are thriving enables individuals to make informed career decisions. They can align their aspirations and educational pursuits with sectors offering stability and growth opportunities. Moreover, insight into industry trends enables organizations to adapt business strategies, staying competitive in flourishing markets and identifying opportunities for innovation in those facing challenges.

5. Demand for Talents Matching Skills with Market Needs: One of the key areas of exploration in this analysis is the determination of the most sought-after talents in the labor market. For job seekers, having a clear understanding of the talents and skills that are currently in high demand is a notable change. It empowers them to choose career pathways that are not only personally fulfilling but also economically viable. By

aligning their skill development with market demands, they position themselves for better career prospects and financial stability.

6. Emerging Responsibilities Adapting to a Changing Landscape: New job roles and responsibilities are always emerging in a world of technological advancements and industry changes. These new duties frequently mirror the changing nature of the workplace and the expectations of the digital era. For those looking to stay ahead of the curve in the employment market, it is imperative that they recognize and comprehend these new roles. People can actively get ready for the future by identifying these new job opportunities.

7. Geographic Trends Regional Opportunities and Discrepancies: Another important factor that needs to be looked at is the geographic distribution of employment opportunities. Certain employment trends and disparities can be identified through a regional examination of job postings. Governments looking to address regional employment imbalances as well as individuals thinking about moving for career opportunities will find this information to be extremely helpful. Individuals can choose where to look for work opportunities by having a thorough understanding of geographic trends.

8. Salary Trends: Examining compensation ranges in job postings more closely reveals the intricate web of differences in pay between industries and regions. An understanding of these salary trends can be advantageous to both employers and job seekers. Employers who want to attract and retain top talent can use this insight to set competitive pay rates. Equipped with knowledge regarding current salary ranges, individuals can more effectively guarantee they obtain just and equitable recompense for their abilities and background. Armed with this knowledge, people can speak up for their own interests—financially and professionally.

C. Motivation: The motivation behind selecting the LinkedIn Job Postings - 2023 dataset lies in the recognition of the paramount importance of comprehending the labor market's intricate landscape in the pivotal year of 2023. This dataset provides a unique window into the ever-evolving world of work, offering real-time insights into job market trends. By scrutinizing aspects such as job market insights, demand for talent, emerging responsibilities, geographic trends, and salary trends, individuals, organizations, and policymakers are armed with the knowledge necessary to make informed decisions. In the context of rapidly changing economic, technological, and societal landscapes, these insights empower job seekers to make well-informed career choices, aligning their aspirations and skill development with market demands. Educational institutions can adapt their curricula to ensure that graduates are job market-ready, and employers can fine-tune their recruitment strategies to attract top talent. Policymakers gain the tools to address regional disparities and foster economic stability.

This dataset's analysis contributes to shaping a labor market that is not only robust but also adaptable, ensuring that both individuals and organizations can thrive amidst the complexities of the modern employment landscape.

These crucial areas—job market insights, talent demand, emerging responsibilities, geographic trends, and salary trends—are examined in depth in the labor market analysis of 2023. By managing these factors, people, businesses, and legislators can make well-informed choices that influence the paths taken by professions, sectors, and geographical areas, resulting in a labor market that is more resilient and flexible. Those with these insights will be in a better position to adapt and prosper in the ever-changing modern employment landscape as the labor market continues to change. We can analyze our own LinkedIn dataset to understand the different insights. [31] This paper centers on investigating "The relationship between job listing attributes and application success rates is being analyzed."

The following inquiries are addressed by this study's examination and analysis of a dataset:

1. Do companies with a higher number of job postings tend to have a more substantial market influence or financial success?
2. Do specific industries or sectors dominate in cities or states with a higher number of job postings?
3. Which industries dominate the job market, and what factors contribute to the sectors exhibiting the highest number of job postings?
4. What are the most common job titles among the job postings in the LinkedIn dataset, and how does the prevalence of these titles vary across different industries, regions, and company sizes?
5. How is the distribution of work types represented in the LinkedIn dataset, and what insights can be gained from visualizing this information?
6. How does the salary distribution vary across different experience levels, and what insights can this offer into the compensation structure based on professional experience?
7. Does the visibility of job postings, as measured by the number of job views on LinkedIn, have a significant impact on the level of interest and engagement from potential candidates, as reflected in the number of job applications received?
8. How do the explicit mentions of specific skills in job descriptions correlate with the perceived importance or priority of those skills in various industries, and how does this correlation vary across job roles and levels?

This research project spans twelve weeks. To prepare for analysis, we will first select a dataset, then proceed to clean, and preprocess it. After collecting and cleansing the data, we will create visualizations to display the researched information.

III. LITERATURE REVIEW

The research being discussed sheds important light on the prevalent practice of online job postings and draws attention to its subtleties, particularly with relation to the educational prerequisites of job searchers. One such analysis done on the online job postings is researched and summarized here.

1. ***"The Online College Labor Market: Where the Jobs Are."***

The recruiting environment has changed significantly in the competitive job market of today. Employers are increasingly using digital media to advertise job opportunities, especially their own websites and online job boards. This transition is supported by the internet's widespread use and ability to provide a quicker, more effective, and more affordable employment procedure.[6]

When compared to the size of job opportunities in the larger economy, the size of online job adverts is astounding. The study shows that there are 3.7 million job opportunities in the United States each month, whereas only about 2.7 million job adverts are published online. This statistical study highlights how heavily companies rely on internet channels for their employment requirements. This discovery is made more fascinating, though, by the apparent gap in employment postings according to educational requirements.

According to the survey, more than 80% of job vacancies geared at candidates with a bachelor's or more are posted online, however this percentage reduces to less than 50% for positions requiring less education. This discovery raises important concerns about the efficiency of internet job postings in reaching various labor groups.

The reliability of online job ad data, particularly for positions needing a bachelor's degree or above, is one of the study's significant conclusions. This pattern highlights the interests and level of digital literacy of recent college graduates, who are more likely to interact with online job postings when doing a job search.

The study incorporates data from Burning Glass Technologies, a well-known provider of information on internet job adverts, to corroborate these conclusions. These data show that every quarter, about 2 million new job postings online with a bachelor's degree or above are made. [15] This statistic shows how frequently businesses utilize online job boards to find highly qualified and educated candidates.

In conclusion, the study emphasizes the paradigm shift in how firms promote job opportunities and the critical importance of online job postings. This change demonstrates nuanced differences in the distribution of online job adverts based on educational qualifications in addition to reflecting the changing dynamics of job recruiting. Additionally, the inclusion of data from Burning Glass Technologies supports the research's assertions by offering concrete proof of the considerable number of online job postings targeted at those with higher educational backgrounds. These results help us comprehend the changing labor market and how digital technology affects the job search process.

Another research done on a revolutionary change in thinking in comprehending the changing labor markets and skill demands has developed with the use of big data in labor market analysis. This chapter explores the significant shift that the enormous amount of data gleaned from millions of online job ads brought about, making it a potent instrument for monitoring labor market developments.

2. ***"Online job postings as a data source to analyze the impact of digitalization on labor markets."***

This literature review explains how natural language processing (NLP) techniques were used to analyze data from online job listings, as well as the benefits and drawbacks of using big data for labor market intelligence in comparison to traditional statistics. Big Data's Benefits for Labor Market Analysis: Insights into labor markets may be gained using big data in a remarkable number of ways, as the research highlights. Big data usage, particularly information derived from the enormous database of internet job advertisements, provides an unmatched size and immediateness. Big data offers almost instantaneous access to current labor market trends, in contrast to conventional techniques of collecting data that sometimes include time-consuming surveys and lag real-time developments.[7]

Additionally, the study highlights how big data may be used to address labor market sectors. To provide a more specialized and nuanced perspective, it enables customized research of job ads based on educational backgrounds, industrial sectors, and geographic regions. This can help close skill gaps and match educational and training programs to the demands of the labor market.

Metrics and Natural Language Processing: The study presents the use of NLP methods to examine data from online job postings. Using NLP, it is possible to extract important data from unstructured text, like necessary skills, educational attainment, and market trends. [16] The study shows the metrics generated by NLP analysis, which might include current skill demand updates, location-specific job trends, and industry-specific insights.

The revolutionary potential of big data, particularly from online job listings, is highlighted in this chapter's conclusion to comprehend the growth of the labor market and skill requirements. Big data is not immune to problems like data bias and quality, despite providing immeasurable benefits in size and speed. The use of NLP methods offers a viable method for deriving essential information from unstructured text data, improving our comprehension of labor market dynamics.

In conclusion, this research establishes the groundwork for the efficient use of big data to monitor labor market changes and match workforce and educational plans with the needs of the always shifting labor market.
Two specific research papers are discussed: "Use of LinkedIn Data and Machine Learning to Analyze Gender Differences in Construction Career Paths" [8] and "Analyzing Career Paths and STEM Professional Development Using LinkedIn Datasets." [9]

The first research examines the gender differences in career routes within the architecture, engineering, and construction (AEC) industry, highlighting the value of diversity while utilizing sophisticated machine learning techniques to predict gender with astounding precision. The second research, which focuses on STEM workers, emphasizes the importance of LinkedIn data for figuring out career paths, job changes, cross-disciplinary talents, and academic accomplishments.

IV. INTRODUCTION TO DATASET

The dataset encompasses over 15,000 job postings from LinkedIn over a two-day period, featuring vital details like job titles, descriptions, salary ranges (maximum, median, minimum), pay periods, work types (full-time, part-time, contract), locations, application numbers, posting times, and remote work options. This dataset is a valuable resource for deep analysis, enabling exploration of high-paying roles, investigation of salary and benefit trends using NLP techniques, and cross-industry and company comparisons for internships and benefits.

Additionally, it comprises supplementary files. 'job_details/benefits.csv' reveals benefit types for each job posting, whether explicitly tagged or inferred. 'company_details/companies.csv' provides company-specific data including names, descriptions, employee count categories, and headquarters details (country, state, city, ZIP code, address), along with LinkedIn page links. 'company_details/employee_counts.csv' contains employee and follower counts for each company, recorded at specific Unix timestamps.

This dataset is a comprehensive resource for analyzing the job market, company profiles, and employment benefits. The dataset we selected “LinkedIn job postings:2023” is retrieved from Kaggle.[1] The dataset can be accessed from LinkedIn API, Web Scraping techniques and Third-Party Data Providers. [3] [4] [5].

V. ARCHITECTURE

Python and Google Collab are the two technologies used to find the answers to research questions. This architecture encapsulates the workflow and tools utilized in exploring the LinkedIn Job Postings - 2023 dataset, conducting in-depth analysis, deriving meaningful insights, and paving the way for future investigations. The use of Python, ML, NLP, and data visualization techniques forms the backbone of this research endeavor.

Python: Python is a widely recognized and versatile high-level programming language used across diverse domains. [21] Python offers several plotting libraries, namely Matplotlib, Seaborn and many other such data visualization packages with different features for creating informative, customized, and appealing plots to present data in the most simple and effective way. [22] Matplotlib provides extensive control over visual figure creations, enabling precise customization for various needs. Seaborn emphasizes enhancing visual aesthetics and simplifying complex plot generation, ideal for statistical data representation. NumPy provides several techniques for data visualization like line plots, scatter plots, bar graphs, and histograms.[23]

Python's strength in data visualization lies in the adaptability and user-friendly nature of these libraries, enabling users to effortlessly generate rich, insightful visual representations. Its broad acceptance and the wealth of libraries contribute to its status as the preferred choice for crafting engaging and informative visualizations, catering to various visualization needs with ease and efficiency.

The following steps are used to acquire the results by setting up the project and preprocessing the selected dataset.

- **Importing necessary libraries:** The code starts by importing the necessary libraries.
- **Loading the dataset:** The next step is to load the dataset from a CSV file using the pandas read_csv() function. The dataset is stored in a pandas data frame named data.
- **Displaying few records:** The head () function is used to display the first five records of the dataset.
- **Handling missing values:** The isnull().sum() function is used to check for any missing values in the dataset, and the fillna() function is used to fill in any missing values with 0.
- **Removing duplicates:** The drop_duplicates() function is used to remove any duplicate records from the dataset.

Google Colab: Google Colab (Colaboratory) is a cloud-based platform provided by Google that allows you to write and execute Python code in a collaborative environment. It's particularly popular for machine learning and data analysis tasks.
Google Colab makes data science, deep learning, neural network, and machine learning accessible to individual researchers who cannot afford costly computational infrastructure. [24] It is a web-based platform designed for

Python coding, predominantly favored within the realms of data science and machine learning. It stands out for its provision of free access to Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs), which significantly benefit tasks involving large datasets and the training of machine learning models. Its foundation lies in Jupyter Notebooks, a versatile tool that allows users to create documents containing executable code, visualizations, explanatory text, equations, and more in a cohesive manner. The best part is, it's completely free and can be accessed from anywhere with an internet connection.[25]

One of its primary attractions is the accessibility and ease of initiating coding endeavors. With pre-installed libraries like Pandas, NumPy, Matplotlib, and TensorFlow, users can quickly jump into coding without the hassle of manual installations. This convenience makes it a favored starting point for beginners in the field. Beyond its functional attributes, Google Collab fosters collaborative work akin to Google Docs. It supports real-time sharing and simultaneous editing, enabling seamless collaboration among team members. Furthermore, its integration with Google Drive streamlines data access and storage, allowing users to import datasets, save work directly to Drive, and access files effortlessly.

Additionally, Google Collab offers hardware acceleration through GPUs and TPUs, allowing for enhanced computational speed in tasks like training complex machine learning models. This feature significantly reduces the burden on local machines and facilitates the handling of resource-intensive computations, thereby widening its utility and appeal within the data science and machine learning communities.

Google Colab is an exceptional tool for data scientists and machine learning engineers. It offers a free and convenient platform to work on your projects, all you need is a browser and internet connection. With the increasing popularity of data science, machine learning and artificial intelligence in recent times, Google Colab has gained more relevance. [26]

VI. PROPOSED METHOD FOR EVALUATION

Harnessing the power of data analytics, data visualization strives to unearth important insights on job market trends, industry-specific demands, and the changing employment landscape. We will diligently examine job posts while upholding data privacy and legal compliance, giving a brief yet comprehensive analysis of the characteristics of the employment market. This method aims to provide useful information to policy makers, employers, and job searchers by anticipating future trends in addition to comprehending the current employment market.

1. Data Collection and preparation: Using legal compliance and allowed means (such the LinkedIn API or web scraping), obtain a dataset of LinkedIn job postings. By correcting duplicates, missing values, and inconsistencies, we preprocess the data. Real world information is very liable to outliers normally referred to as information noise. Data preprocessing deals with data preparation and transformation and seeks to enhance the method of knowledge mining and at the same time make the method of information discovery a lot of economical. [20]

2. Exploration of the Data and Initial Analysis: As the source of the dataset contains more than one csv file, we try to merge the files according to the job_id and company_id as standards and make two comprehensive datasets for further analysis.

3. Visualizing data: To emphasize categorical distributions, temporal trends, and probable outliers in the data, create key data visualizations such as histograms, bar charts, and scatter plots.

4. Recommendations and Reporting: Sum up all the findings in a detailed report or presentation for applicants, recruiters, and businesses, emphasizing key perspectives, concrete suggestions, and data-driven guidance.

VII. PRELIMINARY ANALYSIS

A LinkedIn job posting dataset analysis can reveal important patterns in the employment market, industry-specific needs, and more. Here is a broad strategy we used on the dataset:

1. Data collection: Company details, job details, and job postings are the three folders of datasets we have available.

1. Company Details Folder: companies.csv, company_industries.csv, company_specialities.csv, and employee_counts.csv
2. Job Details Folder: benefits.csv, job_industries.csv, and job_Skills.csv
3. Job Postings Folder: job_postings.csv

We have aggregated the CSV files found in each folder for analysis purposes and have consolidated the required datasets into a single dataset, on which we will base our further data analysis.

- The files job_postings.csv, benefits.csv, and job_skills.csv are combined to obtain more aggregated information about job details.
- The files in the Company Details folder are combined. We have aggregated the same company industries, same company specialties, and in employee_counts.csv file to get the latest data based on the 'time_recorded' column.
- By combining company details and job posting data, we can obtain information on companies that use LinkedIn to post jobs.

2. Data exploration: We looked at the datasets, which contain details about the structure of the data, including column names, data types, and the number of rows and columns.

3. Data cleaning:

The following measures were taken:

- Deal with Missing Values: Recognized and addressed missing data. This involved adding missing values, eliminating rows with missing values, or imputation missing data using statistical methods.
- Eliminate Duplicate Records: The dataset had duplicate records, which were eliminated.
- Conversion of Data Types: Choose the proper data types for each column.
- Managing outliers: Outliers are significant or unexpected data points that could skew the study. We could identify them and choose how to handle them.

Cleaned dataset consists of 15888 rows and 41 columns.

4. EDA (exploratory data analysis):

We have made use of the data visualization method known as World Cloud, which shows the words that appear the most frequently in a dataset. Word clouds have emerged as a straightforward and visually appealing visualization method for text. [13]

Fig .1. Job Title World Cloud

It is a method of visualizing text data in which words are presented in diverse sizes, with the size of each word signifying its frequency or significance in the dataset. Word clouds are frequently employed to quickly absorb the most important terms contained in a body of text or to pinpoint related topics.

VIII. RESULTS

1. Do companies with a higher number of job postings tend to have a more substantial market influence or financial success?

Analyzing this research question provides valuable insights into the dynamics between a company's hiring practices and its market influence or financial success, contributing to informed decision-making and strategic planning. Recruitment is a key role for human resource professionals, as because new talent is essential for an organization to meet its goals and to succeed in a rapidly changing marketplace.[27] A higher number of job postings can indicate a company's growth and expansion. If a company actively hires, it suggests that they are in a phase of development, which may positively correlate with market influence and financial success. Companies invest significant resources in hiring talent.

Analyzing the relationship between job postings and market influence/financial success can reveal the extent to which companies prioritize human capital as a factor in their overall success. The number of job postings can be a competitive metric. Companies vying for top talent might increase job postings to attract skilled professionals, and this competition may positively impact their market influence and financial success. However, it may not directly indicate these firms' market impact or financial success. Several factors, like development, expansion, turnover, or hiring tactics, might affect job listings.

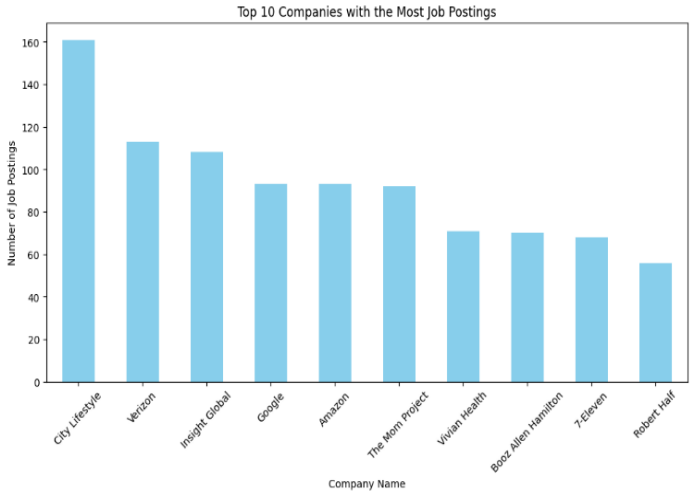

Fig .2. Top 10 companies with the most job postings

The above plot provides a snapshot of the top 10 companies that exhibit the highest volume of job postings, extracted from the LinkedIn dataset. Each company accompanies the respective count of job postings, offering insights into their active recruitment efforts on the platform.

Employers aiming to attract top talent can learn about the kinds of positions, competencies, and credentials in demand by examining job advertisements from prosperous businesses. This information assists companies in creating their talent acquisition strategy by pinpointing the industry's primary areas of concentration and specialty.

Analyzing job listings from various organizations allows for the identification of industry trends, such as newly created jobs, in-demand talents, and changing job responsibilities. Professionals attempting to align their skill set with market expectations and academic institutions adjusting their curricula to suit industry needs find this information helpful.

In summary, knowing which 10 organizations have the most job posts benefits businesses, professionals, and stakeholders alike. It offers an all-encompassing perspective of the labor market, facilitating strategic decision-making and well-informed actions in the areas of industry positioning, hiring, and general company planning.

2. Do specific industries or sectors dominate in cities or states with a higher number of job postings?

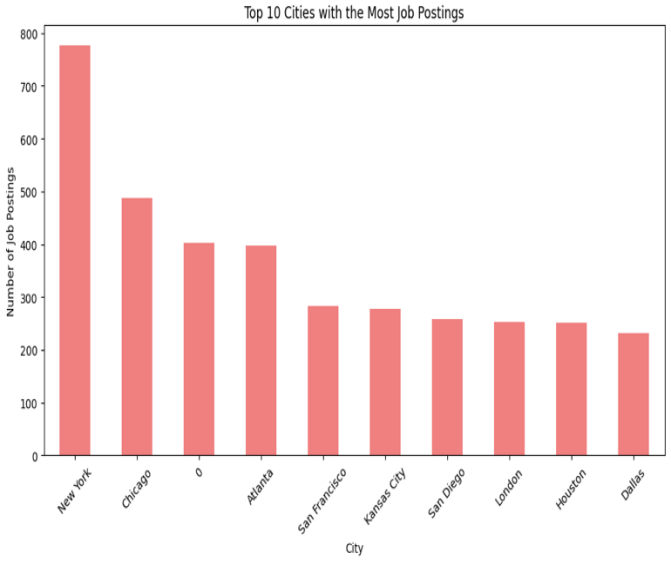

Fig .3. Top 10 cities with the most job postings

Analyzing job postings across cities and states in terms of specific industries or sectors can yield valuable insights into regional employment trends and economic activity. The results reveal:

Industry centers: Cities like New York City, San Francisco, and Los Angeles abound with job advertisements due to their status as significant industry centers. These cities frequently accommodate a wide variety of businesses and sectors.
Tech Dominance: States like California and locations like San Francisco prominently feature job ads, indicating their dominance in the tech sector. This trend is partly driven by the popularity of occupations connected to technology.

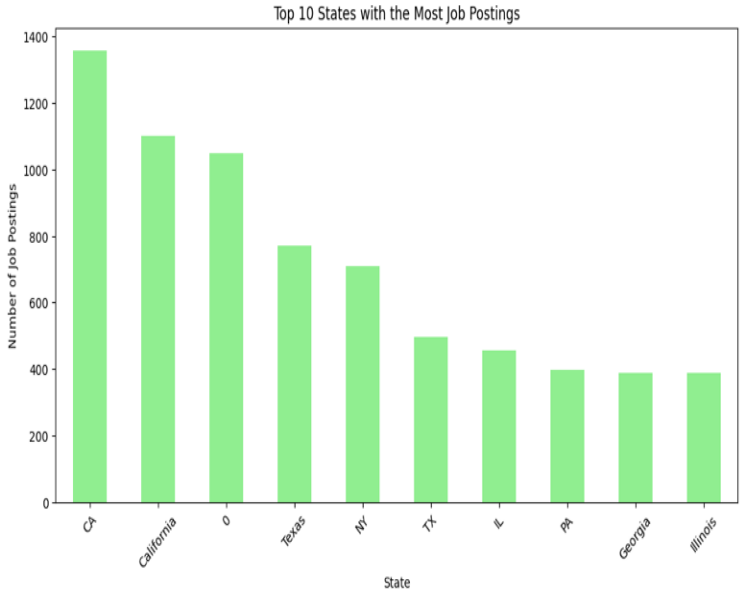

Fig .5. Top 10 states with the most job postings

Economic Centers: Places like New York City, considered economic centers, offer numerous work possibilities in various industries such as journalism, healthcare, and finance.

Regional Variances: The distribution of job posts among states indicates regional economic differences. States with vast and diversified economies, like Texas and California, naturally offer more work opportunities.

Fig .4. Tree Map

The above chart is a Tree map which are used in visualizations for hierarchical data. They are made of a series of nested rectangles of sizes proportional to the corresponding data value. A large rectangle represents a branch of a data tree, and it is subdivided into smaller rectangles that represent the size of each node within that branch. [28]

3. Which industries dominate the job market, and what factors contribute to the sectors exhibiting the highest number of job postings?

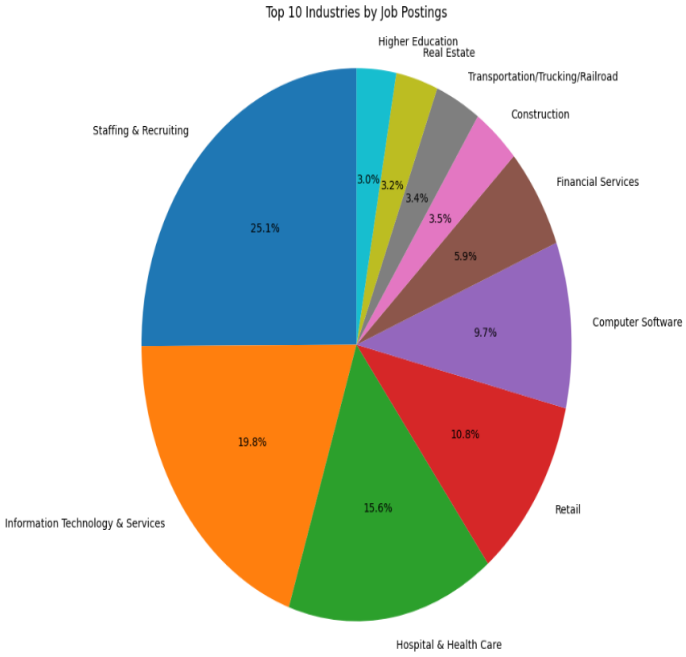

Fig .6. Top 10 industries by job postings

The research question tackles critical aspects of the job market, providing valuable information for various stakeholders. Understanding the dominant industries and the factors influencing job postings contributes to informed decision-making at both individual and organizational levels. It plays a significant role in shaping workforce strategies, economic policies, and career planning.

The primary goal is to identify the industries that have the highest number of job postings. This information helps in understanding which sectors actively seek new talent. Analyzing the top industries by job postings provides insights into current job market trends. It helps recognize which sectors experience growth and demand for skilled professionals. Industries with a high number of job postings often indicate economic vitality and growth in those sectors. Staffing and recruiting, at the top with 25%, suggest a high demand for workforce acquisition and human resources services. IT closely follows at 19.8%, indicating the continued growth and demand for professionals in the information technology sector. Health, at 15%, suggests a substantial number of job opportunities in the healthcare industry.

Employers, job seekers, and policymakers can use this information for strategic workforce planning. Companies can align their hiring strategies with the demands of dominant industries, and job seekers can focus on industries with abundant job opportunities. Dependence on a single industry can pose risks, and a diverse job market contributes to economic resilience. For individuals planning their careers, knowing which industries are dominant in terms of job opportunities allows for informed decision-making. It helps align career choices with market demands. The research question indirectly emphasizes the need for adaptability and continuous skill development. As industries evolve, staying competitive in the job market may require acquiring new skills aligned with the demands of dominant sectors.

4.What are the most common job titles among the job postings in the LinkedIn dataset, and how does the prevalence of these titles vary across different industries, regions, and company sizes?

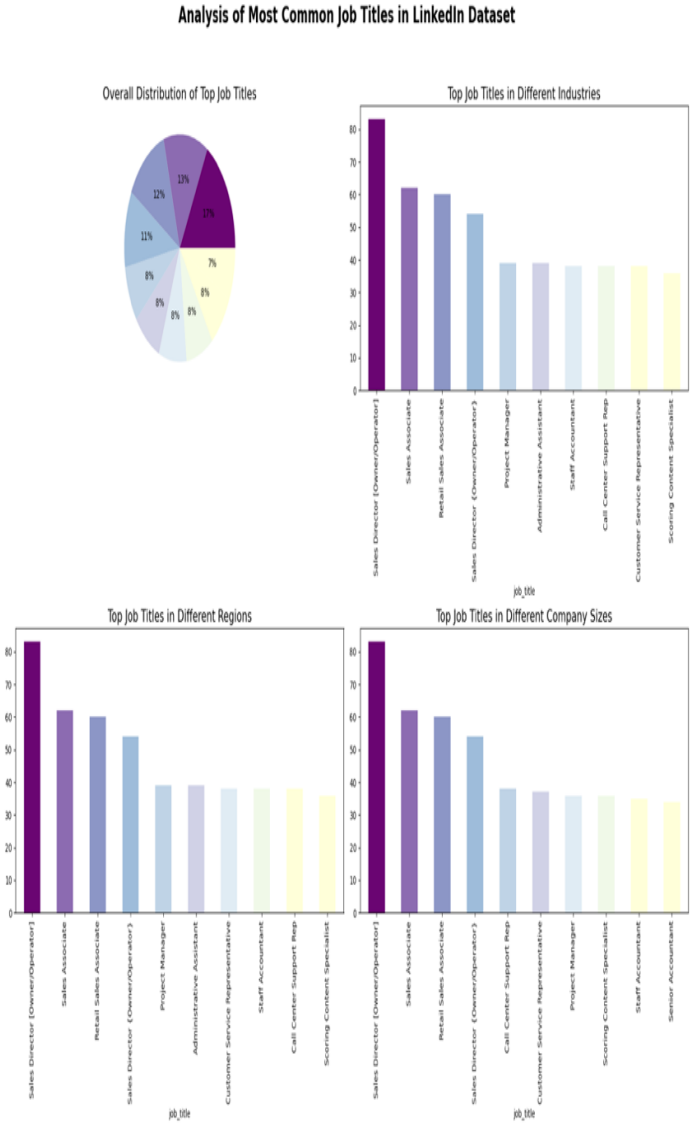

Fig .8. Job Titles

The research question drives the determination of the most prevalent job titles overall and the understanding of the most

popular positions in the dataset. We can analyze changes in job titles unique to a sector, identifying distinct job titles or in-demand roles in specific industries. We can also investigate regional variations in the demand for specific job titles.

Job seekers benefit from understanding the kinds of job titles that organizations of various sizes prefer. This sheds light on businesses' personnel requirements based on their size. We can determine the job titles in great demand both generally and within certain industries or locations. This allows us to identify the most common job titles in the field and area, enabling the modification of job advertisements to attract the right applicants.

The graph illustrates that various firms commonly choose job titles such as "Sales Director," "Sales Associate," and "Project Manager." This directs our attention to the work abilities required to fulfill these responsibilities. The graph's visualizations and analysis provide a thorough overview of job title distributions, empowering stakeholders to make informed decisions based on factors including industry, location, and firm size.

5. How is the distribution of work types represented in the LinkedIn dataset, and what insights can be gained from visualizing this information?

To answer this research question, we first examined the overall distribution of work types (such as full-time, temporary, and internship) across the entire LinkedIn dataset. We created a pie chart using the graph below to illustrate the percentage of each work category on LinkedIn. The percentage labels will display the relative distribution, and each slice of the pie will represent a distinct work category. These visualizations can offer HR professionals, job seekers, and researcher's valuable insights into the typical work patterns of the LinkedIn community.

Full-time employment comprises a significant 80.84% of the information, making it the most common work type on LinkedIn. This suggests that a sizable percentage of LinkedIn users engage in full-time jobs. Temporary roles constitute 0.76 percent of the jobs in LinkedIn's dataset, while internships make up 0.70 percent. This could imply that, although these work arrangements exist, they are less frequent than full-time, steady jobs. Contract work and part-time employment show moderate representation, with percentages of 10.95% and 6.36%, respectively. This indicates that the LinkedIn community offers a healthy blend of contract and part-time job options. Volunteer work and other non-essential labor kinds contribute very little to the distribution, with percentages of 0.05% and 0.33%, respectively.

This demonstrates how these kinds of jobs are either less widespread or less specifically mentioned on LinkedIn profiles. These insights provide a thorough summary of the distribution of work types on LinkedIn, offering helpful information for researchers, employers, and job seekers seeking to comprehend the dynamics of the professional environment on LinkedIn. The prevalence of full-time jobs indicates a healthy labor market for professionals looking for long-term positions, while individuals seeking internship or temporary work may face greater competition.

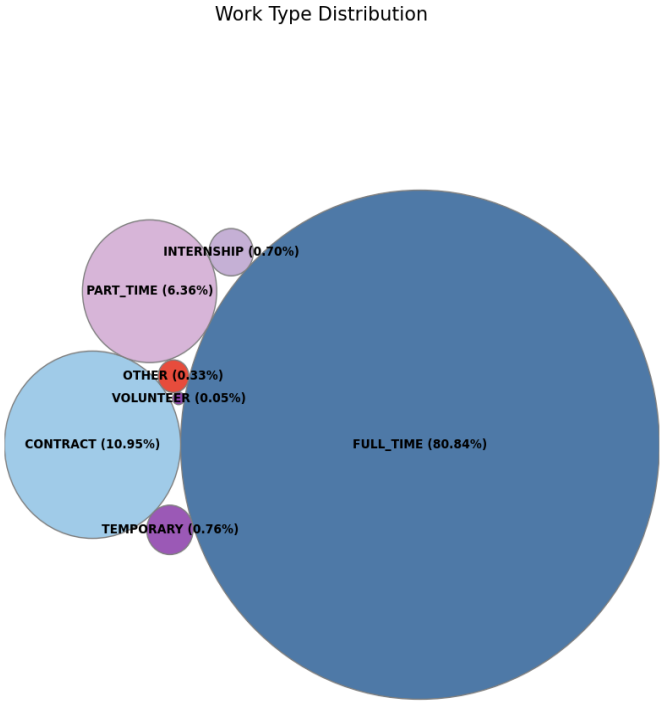

Fig .9. Job Work Types

6.How does the salary distribution vary across different experience levels, and what insights can this offer into the compensation structure based on professional experience?

Salary Distribution by Experience Level: Salaries were plotted against the required experience levels for jobs. Understanding salary distribution across different experience levels is crucial for comprehending how compensation aligns with professional growth. However, the analysis shows an absence of specific information regarding remote job salaries. This absence suggests that the dataset may not capture the influence of work nature (remote, hybrid, or in-person) on salary distribution.

From the available data, it's evident that salary variations might not directly correlate with the nature of work but rather with other factors like experience level, industry type, company size, or financial performance. This observation underscores the complexity of determining how compensation structures align with professional experience. Exploring this relationship

allows for insights into whether salary increments follow a linear trajectory with increasing experience or if there are industry-specific spikes at certain experience levels. By acknowledging that work nature might not directly impact compensation, professionals and organizations can focus on other critical factors influencing salary growth and ensure more informed compensation negotiations and career development strategies.

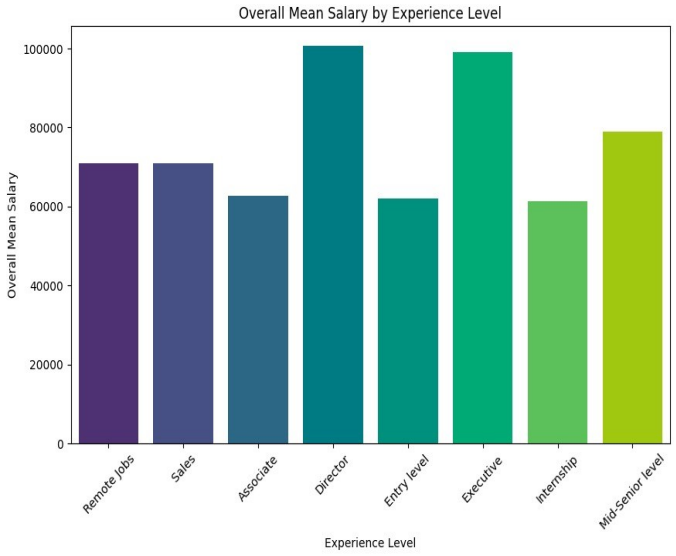

Fig.10.Overall Mean Salary by Experience Level

The salary distribution by experience level shows a gradual increase from the entry-level to the executive level. This is in line with the common understanding that experienced professionals tend to earn more than their less experienced counterparts.

Here's a detailed breakdown of the salary distribution by experience level:

Entry Level: The starting point for a professional career. The salaries for these jobs are typically lower than for jobs at higher experience levels.

Mid-Senior Level: The mid-point of the salary distribution. Salaries for these jobs are generally higher than those for entry-level positions, but lower than those for executive-level jobs.

Executive Level: The topmost point of the salary distribution. Salaries for these jobs are usually the highest, indicating that the highest level of expertise is rewarded with the highest salaries.

This analysis offers valuable insights into the compensation structure based on professional experience. It highlights that experience is an important factor in determining an individual's salary, with higher levels of experience corresponding to higher salaries.

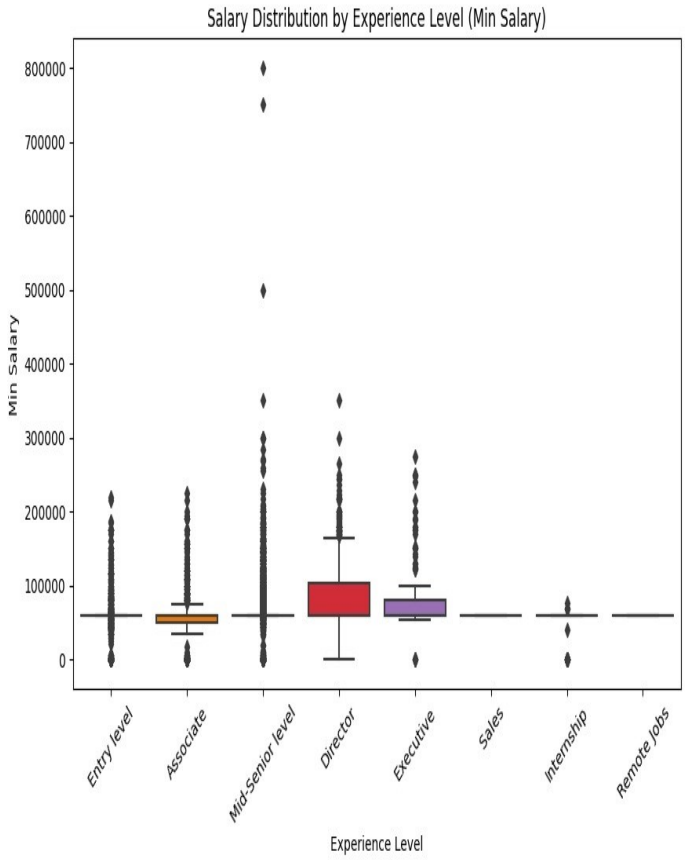

Fig.11.Salary Distribution by Experience level (Min Salary)

The analysis of Salary Distribution by Experience Level provides a comprehensive overview of salary ranges across various professional tiers, each delineated with a specific salary range indicative of the level's typical roles and responsibilities.

The graphical representation through a box plot is a powerful tool portraying the distribution of salary data for each experience level. It encapsulates vital statistics like minimum, maximum, first quartile (Q1), median (Q2), and third quartile (Q3). The median salary serves as a pivotal point, dividing the dataset into two halves—indicating that half of the professionals within that experience level earn more, and the other half earns less than the median.

The shape of the box plot reflects the data's dispersion. A shorter, wider box implies more uniformity in salaries around the median, while a longer, narrower box suggests skewness or potential outliers within the dataset.

Examining this salary distribution offers valuable insights into compensation structures based on professional experience. For instance, the delineation of salary ranges across various levels—ranging from entry-level positions to executive roles—clearly indicates the salary progression with career advancement. Higher-level positions, such as directors and executives, command substantially higher salaries compared to lower-tier roles like entry-level or associate positions.

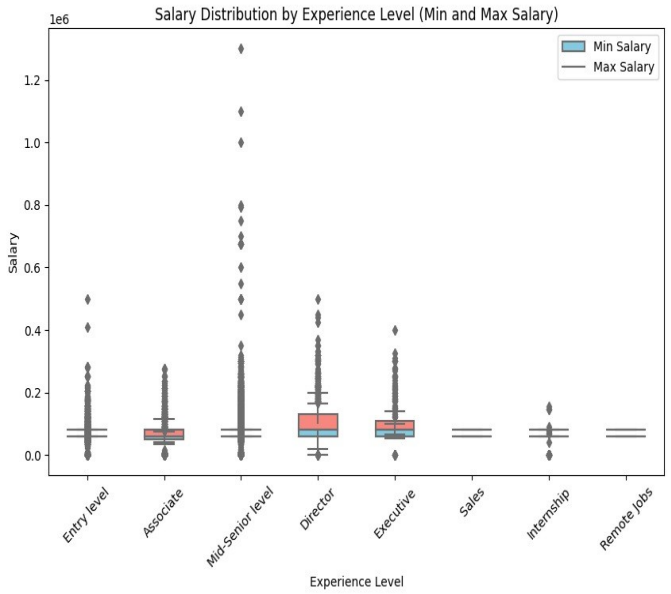

Fig.12.Salary Distribution by Experience level (Min and Max Salary)

The analysis of Salary Distribution by Experience Level, portrayed through a box plot, highlights a consistent trend: as experience levels rise, so does the average salary. The box plot illustration provides a comprehensive overview of the salary data's spread and distribution. It employs quartiles—Q1 and Q3—and a median line to delineate the salary range.

The key insights derived from this analysis are multifold. Firstly, the similarity in salary trends across varying experience levels indicates a clear association between salary and the candidate's years of experience. Secondly, the portrayal of wider salary spreads at higher experience levels suggests increased variability in compensation among these roles. Moreover, the identification of outliers within specific experience levels is significant. Outliers could signify individuals earning substantially above or below the median salary for that experience level, implying unique circumstances or skill sets influencing compensation.

In summary, this analysis offers valuable insights into the compensation structure concerning professional experience. It notably demonstrates an escalating median salary with increasing experience, suggesting a positive correlation between the two. However, the notable variability in salary ranges among higher experience tiers underscores the diversity and potential complexity in compensation for these roles.

7. Does the visibility of job postings, as measured by the number of job views on LinkedIn, have a significant impact on the level of interest and engagement from potential candidates, as reflected in the number of job applications received?

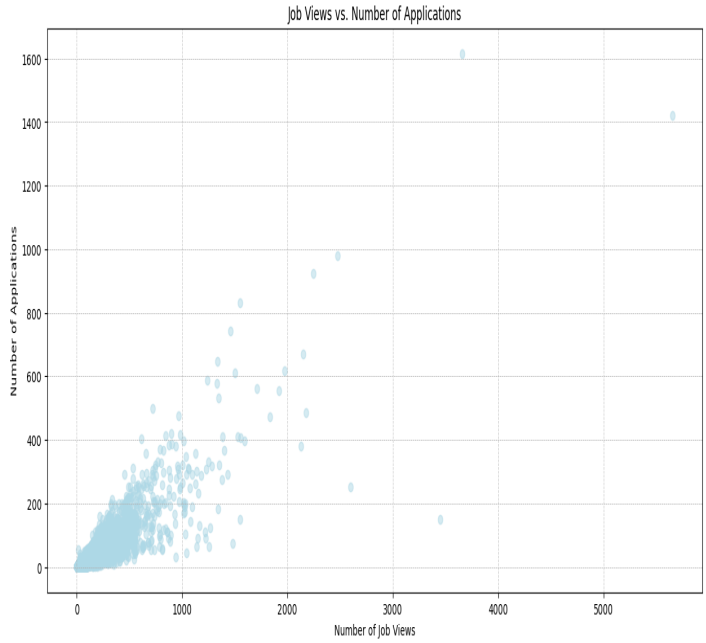

Fig.15. Job Views vs Number of job Applications

The scatter plot and the correlation value both offer insights into the relationship between the number of job views and the number of job applications. The scatter plot visually illustrates the data points for each job posting, with the x-axis denoting the number of job views and the y-axis representing the number of job applications. The general trend in the scatter plot indicates that as the number of job views increases, there is a tendency for the number of job applications to also increase. Each point on the plot corresponds to a specific job posting, suggesting a positive correlation.

The obtained correlation value is 0.847. The correlation coefficient quantifies the strength and direction of a linear relationship between two variables. In this case, the correlation value of approximately 0.847 indicates a strong positive correlation. A correlation coefficient close to 1 suggests that as one variable (job views) increases, the other variable (job applications) tends to increase as well, and vice versa. The positive sign of the correlation indicates that both variables move in the same direction: as job views increase, job applications also increase.

The strong positive correlation suggests a consistent and robust relationship between the visibility of a job posting (measured by the number of views) and the level of interest it generates (measured by the number of applications). Jobs that attract more views are more likely to receive a higher number of applications, implying a close link between visibility and interest. Employers and recruiters can find this information valuable, as it suggests that efforts to increase the visibility of job postings can positively impact the number of applications received.

In summary, both the scatter plot and correlation value offer a comprehensive understanding of the positive relationship between job views and job applications. This insight can guide recruitment strategies, enabling organizations to focus on enhancing the visibility of jobs to attract a larger pool of applicants.

8. How do the explicit mentions of specific skills in job descriptions correlate with the perceived importance or priority of those skills in various industries, and how does this correlation vary across job roles and levels?

For this research question analysis, we have considered taking job_postings.csv and job_skills.csv files. The side-by-side comparison of the distribution of text length for job descriptions before and after cleaning is demonstrated in the chart below. The histograms visually assess whether the text cleaning process had a significant impact on the distribution of text lengths. The descriptive statistics tables provide a numerical summary of the central tendency and spread for both the original and cleaned lengths, aiding in understanding the changes. The code explores and presents the differences in the distribution of text lengths before and after the cleaning process, offering insights into the impact of the cleaning on the job description data. The nature of the underlying content can be inferred from text lengths. For example, in a dataset of job descriptions, longer descriptions might convey more detail or specialization, while shorter ones may be more concise.

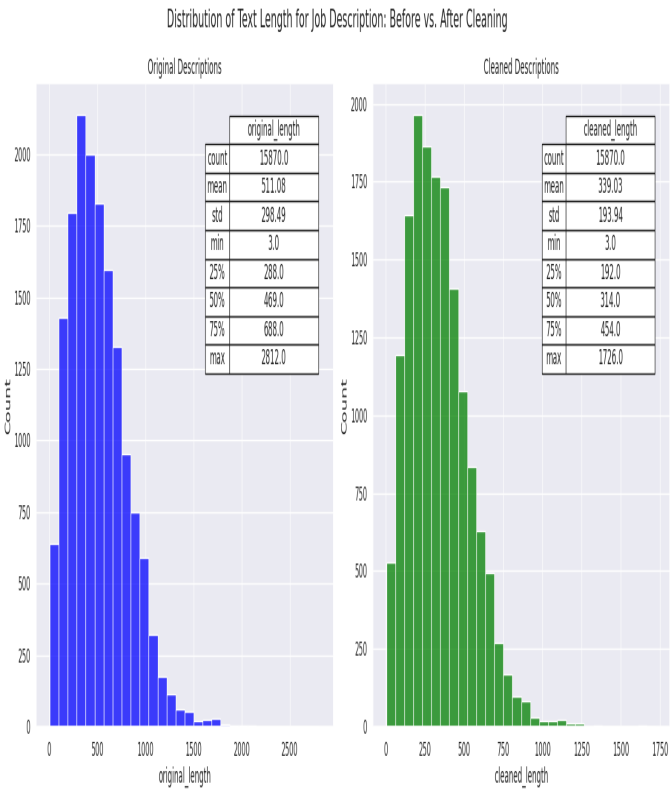

Fig.16. Distribution of Text Length for Job Description

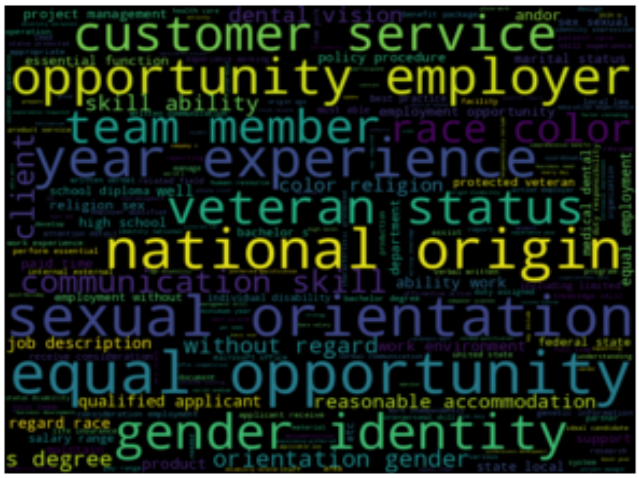

Fig.17. Job Description Word Cloud

We have implemented scikit-learn's Count Vectorizer to perform n-gram analysis on cleaned job descriptions, calculating and plotting the count of unigrams, bigrams, and trigrams. The blue line plot and markers represent the counts of n-grams, with the y-axis indicating the count and the x-axis representing different types of n-grams (unigram, bigram, trigram).

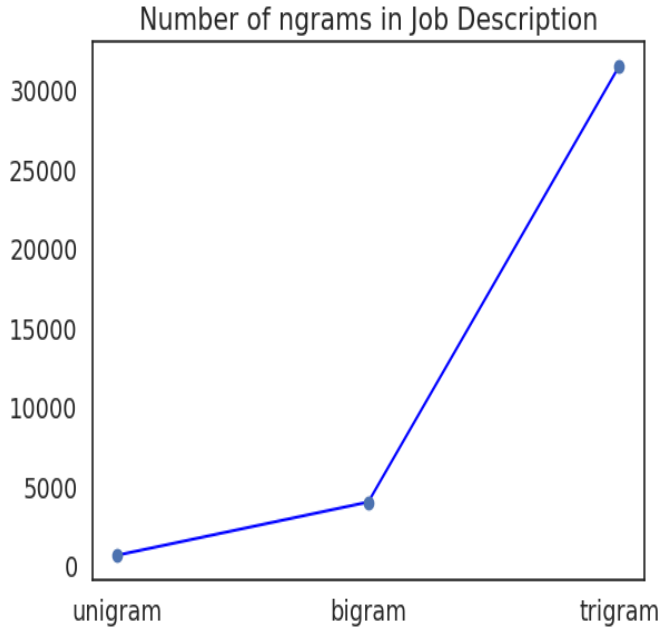

Fig.18. N-grams

Observing the plot allows us to assess how the count of n-grams changes as we transition from unigrams to bigrams and trigrams. An increase in counts from unigrams to bigrams and trigrams may indicate more diverse and complex language structures. Higher counts of bigrams and trigrams compared to unigrams could suggest the presence of multi-word phrases and expressions in the job descriptions. This information is valuable

for understanding the level of detail and complexity in the language used.

The Count Vectorizer is configured with tokenizer=lambda doc: doc and lowercase=False, indicating that it does not perform tokenization or lowercasing. Each entire document is treated as a single token. This choice of tokenization affects the size and nature of the resulting n-grams.

N-grams are continuous sequences of words or symbols, or tokens in a document. In technical terms, they can be defined as the neighboring sequences of items in a document. [29] The n-gram analysis provides insights into the structure and complexity of language in the job descriptions, helping in tasks such as feature engineering, model selection, and improving the overall quality of natural language processing applications. In the modeling section, a text classification model is implemented to predict job skills from job descriptions.

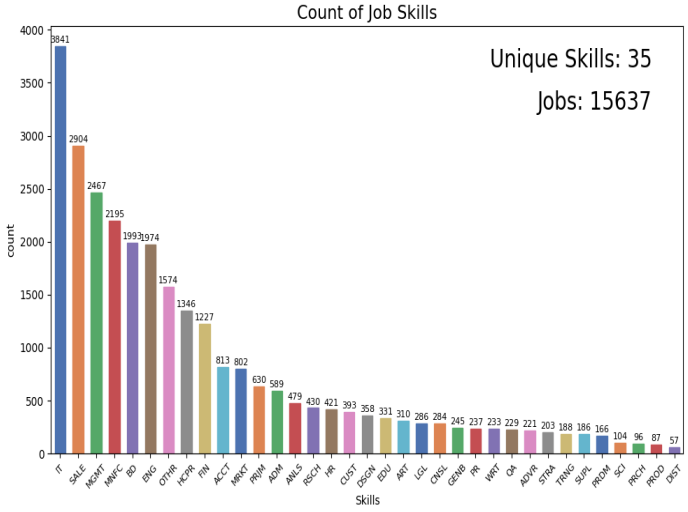

Fig.19. Job skills

The graph illustrates the distribution of job skills across different categories, with the top three categories being IT, Sales, and Management. IT boasts the highest count with 3841 occurrences, followed by Sales with 2904, and Management with 2467. This representation provides valuable insights into the prevalence of skills within distinct professional domains, aiding in the assessment of industry emphasis and workforce composition.

The dataset is split into training and testing sets, and text features are processed using both Count Vectorization (Bag of Words) and TF-IDF methods. Logistic Regression is employed as the classifier, and performance metrics such as accuracy, precision, recall, and F1-score are calculated for each skill category. The results show that the Count Vectorization approach achieves an accuracy of 63.90%, while the TF-IDF method yields an accuracy of 50.87%. The classification report provides a detailed breakdown of precision, recall, and F1-score for each skill category. It is observed that some skills have higher predictive accuracy than others, suggesting variations in the frequency and uniqueness of terms within the job descriptions. Hyperparameter tuning is performed for the Naive Bayes classifier using the TF-IDF method. The code utilizes a grid search with cross-validation to find the optimal value for the smoothing parameter (alpha) in the Multinomial Naive Bayes model. The best alpha is found to be 0.1, and a plot is generated to visualize the relationship between different alpha values and mean test scores. [30]

The varying accuracy across skills suggests that certain skills are more easily distinguishable from the job descriptions, while others might share common terms, making them harder to predict accurately. The TF-IDF method, despite its widely acknowledged effectiveness, appears to perform less well in this specific case compared to the simpler Count Vectorization approach. It's also worth noting that tuning hyperparameters, as seen with Naive Bayes, can improve model performance.

These insights can guide improvements in the model, such as refining text preprocessing steps, experimenting with different vectorization methods, or exploring more advanced models. Understanding the strengths and weaknesses of the model is crucial for its application in real-world scenarios, especially in the context of matching job descriptions to specific skill requirements.

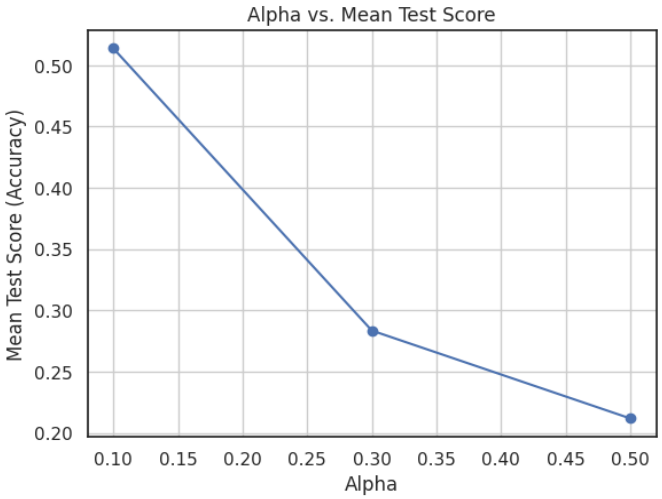

Fig.20. Alpha vs Mean Test Score

IX. CONCLUSION

Finally, our study offers a thorough analysis of the labor market for 2023 utilizing the large dataset of over 15,000 job listings from LinkedIn. Our approach uses ML algorithms, data visualization, NLP, and powerful data analytics to provide essential information about industry dynamics, job market trends, and the changing nature of work. The review technique that has been suggested is comprehensive and encompasses several critical areas, including preparation and collection of

data, exploratory data analysis, visualization, text mining based on natural language processing, geospatial analytics, sophisticated predictive modeling, and recommendations and reporting at the conclusion.

Our investigation has yielded significant insights into the correlation between an organization's recruiting procedures, market share, and financial prosperity. Additionally, by looking at job posts in different states and localities, we have investigated regional employment trends and provided information on economic activity and industrial domination. Additionally, our analysis of the top sectors based on job listings covers important facets of the labor market and provides data that is essential for making decisions at the individual and organizational levels. In addition, our examination of innovative technologies and abilities in job advertisements considers the changing demands of the labor market.
Given the circumstances, this study initiative is positioned to provide insightful information to a wide range of stakeholders, including employers, job seekers, and policymakers. The research seeks to provide decision-makers with practical information by forecasting future trends and offering a deeper view of the existing labor market. This study's examination of geographical differences, industry dominance, and the elements influencing job listings makes it an essential tool for economic policy, workforce planning, and well-informed career decision-making.

As the global economy undergoes rapid transformations, the findings from this initiative are poised to guide policymakers in formulating adaptive strategies. By aligning workforce policies with the identified trends, governments can better support economic growth, foster innovation, and enhance overall labor market resilience.

X. FUTURE WORKS

In our pursuit of a thorough and informative research project, our focus has been methodically dedicated on uncovering the intricate dynamics present in the LinkedIn Job Postings - 2023 dataset. The purpose of this preliminary investigation is to better understand the complex role that LinkedIn job postings and how they play a role in professional settings and to identify the numerous ways in which these algorithms contribute to hiring practices, job market trends, and the efficient matching of candidate skill.

As we envision the future direction of our study, we are interested in exploring the wider world of business networking sites, such as Indeed, Glassdoor, and other platforms. The purpose of our upcoming study phase is to examine how these social media platforms affect work-life balance and professional progress. We will undertake an in-depth investigation to assess whether increased activity on social networks may inhibit career advancement, giving subtle insights into the developing environment of modern professional interactions.

Furthermore, we intend to investigate the efficacy of digital well-being solutions that phone manufacturers have offered to control social media usage. This study will evaluate how well these tools work to assist users in finding a balance between their online and offline lives, providing insightful information for anybody looking to have a better connection with digital platforms. We will collect data on layoffs, identifying the job roles most affected to enhance the study's depth and relevance in understanding the broader impact on employment dynamics.

Additionally, a crucial component of our upcoming projects is a thorough examination of the function that social media algorithms perform in the context of employment-related platforms. We will investigate how these algorithms affect recruitment tactics, user behavior, and the whole recruiting process for employers and prospects.

In conclusion, our future research is well-positioned to traverse the varied terrain of social media platforms focused on jobs, investigating their effects on work and personal life, hiring procedures, and methods for engaging candidates. This comprehensive investigation seeks to give a thorough grasp of the changing dynamics in the digital recruiting market, providing insightful information to people, businesses, and governments alike.

XI. PROJECT WEBSITE

The website project was created utilizing the resources of George Mason University in accordance with the provided guidelines. [32] [33]

Project website: https://mason.gmu.edu/~asen3/

XII. AUTHOR INFORMATION

Aiswarya Sen, currently pursuing, a Master's degree in Information Systems at George Mason University, brings expertise in Data Analysis and a passion for research in machine learning, data visualization, and NLP.

Narasimha Krishna Amruth Vemuganti, currently pursuing a Master’s in Data Analytics and Engineering at George Mason University. Deeply interested in machine learning and actively exploring the field of prompt engineering in research endeavors.

Monali Sheetal Kathepaga, currently pursuing a Master's degree in Data Analytics at George Mason University, has a keen interest in the advancements occurring in the field of data analysis.

XIII. REFERENCES

[1] A. Kon, “LinkedIn job postings - 2023,” Kaggle, https://www.kaggle.com/datasets/arshkon/linkedin-job-postings (accessed Nov. 11, 2023).

[2] “Data-Driven Job Search Engine using skills and company attribute filters,” *IEEE Conference Publication | IEEE Xplore*, Nov. 01, 2017. https://ieeexplore.ieee.org/document/8215664.

[3] G. Karatas, “LinkedIn Datasets: Types, applications and providers in 2023,” *AIMultiple*, Sep. 08, 2023. https://research.aimultiple.com/linkedin-dataset/

[4] “LinkedIn Data: Best Datasets & Databases 2023 | Datarade.” https://datarade.ai/data-categories/linkedin-data

[5] Bright Data, “LinkedIn datasets,” *Bright Data*, Nov. 06, 2023. https://rb.gy/d53moo

[6] “The online college labor Market: Where the jobs are - CEW Georgetown,” *CEW Georgetown*, May 07, 2020. https://cew.georgetown.edu/cew-reports/the-online-college-labor-market/

[7] “Home.” https://www.oecd-ilibrary.org/sites/7d99dfbe-en/index.html?itemId=/content/component/7d99dfbe-en

[8] P. J. Hickey, A. Erfani, and Q. Cui, “Use of LinkedIn data and machine learning to analyze gender differences in construction career paths,” *Journal of Management in Engineering*, vol. 38, no. 6, Nov. 2022, doi: 10.1061/(asce)me.1943-5479.0001087.

[9] B. Data, “STEM Jobs Analysis with LinkedIn Datasets | Analytics Vidhya,” *Medium*, Aug. 23, 2022. [Online]. Available: https://medium.com/analytics-vidhya/analyzing-career-paths-and-stem-professional-development-using-linkedin-datasets-4c47e87678ba

[10] “LinkedIn Careers: Find your opportunity here,” *LinkedIn Careers: Find Your Opportunity Here*. https://careers.linkedin.com/.

[11] M.A. Olowe, M.M. Gaber and F.Stahl, “A survey of Data Mining techniques for social media analysis”, Journal of Data Mining and Digital Humanities, 2014.

[12] E. Ferrara, D.P. Meo, G. Fiumara and R. Baumagartner, “Web data extraction, applications and techniques: A survey”, Knowledge based systems, Vol. 70, pp.301-323, 2014.

[13] F. Heimerl, S. Lohmann, S. Lange and T. Ertl, "Word Cloud Explorer: Text Analytics Based on Word Clouds," 2014 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 2014, pp. 1833-1842, doi: 10.1109/HICSS.2014.231.

[14] A. T. Dang, “Mining data on LinkedIn - towards Data Science,” *Medium*, Dec. 14, 2021. [Online]. Available: https://towardsdatascience.com/mining-data-on-linkedin-9b70681b1467

[15] A. P. Carnevale, “The Online College Labor Market: Where the Jobs Are,” *DigitalGeorgetown*, Apr. 01, 2014. https://repository.library.georgetown.edu/handle/10822/1050294

[16] K. Tran, “Explore and Visualize your LinkedIn Network with Python and Sentiment Analysis,” *Medium*, Dec. 15, 2021. [Online]. Available: https://towardsdatascience.com/sentiment-analysis-of-linkedin-messages-3bb152307f84

[17] A. Ghosh, B. Woolf, S. Zilberstein and A. Lan, "Skill-based Career Path Modeling and Recommendation," 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020, pp. 1156-1165, doi: 10.1109/BigData50022.2020.9377992.

[18] “The effect of the job market on the economy | U.S. Bank,” Nov. 09, 2023. https://www.usbank.com/investing/financial-perspectives/market-news/effect-of-job-market-on-the-economy.html

[19] K. Dissanayake, S. Mendis, R. Subasinghe, D. Geethanjana, D. Lunugalage and D. Kasthurirathna, "Career Aura – Smart Resume and Employment Recommender," 2021 3rd International Conference on Advancements in Computing (ICAC), Colombo, Sri Lanka, 2021, pp. 353-358, doi: 10.1109/ICAC54203.2021.9671212.

[20] S. Gharatkar, A. Ingle, T. Naik and A. Save, "Review preprocessing using data cleaning and stemming technique," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2017, pp. 1-4, doi: 10.1109/ICIIECS.2017.8276011.

[21] *Welcome to Python.org*. (2023, November 23).Python.org. https://www.python.org/

[22] S, R. A. (2023, January 27). *An interesting guide to visualize data using Python Seaborn*. Simplilearn.com. https://www.simplilearn.com/tutorials/python-tutorial/python-seaborn

[23] *NumPY Data Visualization (With Examples)*. (n.d.). https://www.programiz.com/python-programming/numpy/data-visualization#:~:text=NumPy%20provides%20several%20techniques%20for,of%20data%20quickly%20and%20efficiently.

[24] Das, T. (2023, September 20). *Google Colab: Everything you Need to Know*. Geekflare. https://geekflare.com/google-colab/

[25] Mukhtiar, M. (2023, June 23). Effortlessly Scale Your Data Science Projects with Google Colab. *Medium*. https://medium.com/@rao.mohsin.54/effortlessly-scale-your-data-science-projects-with-google-colab-11117f591353

[26] Michael, & Michael. (2023, May 20). *Google CoLAB: Comprehensive Guide for Data Science and ML | Mike Urmeneta, Ed.D.* Mike Urmeneta, Ed.D. | I Help Organizations "move the Needle" With an Agile and Collaborative Approach to Research and Analysis. https://allthingy.com/google-colab-comprehensive-guide-for-data-science-and-ml/

[27] *JOSSO 2 by Atricore*. (n.d.). https://www.shrm.org/resourcesandtools/tools-and-samples/toolkits/pages/recruitinginternallyandexternally.aspx

[28] *TreeMaps: Data visualization of complex hierarchies*. (n.d.). Nielsen Norman Group. https://www.nngroup.com/articles/treemaps/

[29] V, N. (2023, September 14). *What are N-Grams and how to implement them in Python?* Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/09/what-are-n-grams-and-how-to-implement-them-in-python/

[30] Sklearn.feature_*extraction.text.TfidfVectorizer*. (n.d.). Scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

[31] Suwandi, R. C. (2022, May 7). Analyzing my own LinkedIn data - Richard Cornelius Suwandi - Medium. *Medium*. https://richardcsuwandi.medium.com/analyzing-my-own-linkedin-data-24df63a9dc28

[32] George Mason University. (2021, March 22). *How to Set Up a Personal Site - Information Technology Services*. Information Technology Services. https://its.gmu.edu/knowledge-base/how-to-set-up-a-personal-site/

[33] George Mason University. (2023, May 31). *How to connect to the Mason Cluster via terminal (Mac) - Information Technology Services*. Information Technology Services. https://its.gmu.edu/knowledge-base/how-to-connect-to-the-mason-cluster-via-terminal-mac/