

# Career Navigation with LinkedIn Job Posting Insights

What skills, careers, and cities should you be working towards?

**Alex Archer**

Applied Computer Science  
University of Colorado,  
Boulder  
Boulder, Colorado, USA  
alar2300@colorado.edu

## Project Statement

In today's rapidly evolving job market, finding the right career path can be like searching for a needle in a haystack. With industries constantly transforming and demanding new skills at record speeds, job seekers and professionals are often left wondering, 'Where do I fit in?' Which my project, "Career Navigation with LinkedIn Job Posting Insights," I intend to help answer this question. With the help of the massive public dataset "1.3 M LinkedIn Jobs & Skills (2024)," I aim to provide clarity about the current job market landscape, turning this massive data set into relevant and up-to-date insights.

This dataset will be a source of useful insights into current opportunities, with 1.3 million entries spanning across numerous industries. Each job listing provides abundant information, detailing not only the role, but the skills it demands, which company is offering, and the location of the posting. The attributes in the first of the two tables being examined include the link to the posting as the primary key, job title, company, job location, and search city, which is the criterion for collecting location relevant postings. The second table holds the list of relevant skills for each posting, as well as the postings link as the primary key.

My goal is twofold. First, I will discover where jobs in specific industries are booming, so that job seekers know where to find opportunities best suited to their skillset. Second, I will uncover the most sought after job titles in addition to the top skills they require by region, thus providing a roadmap for skill

development and career planning. I hope to reveal the current state of the 2024 job market, helping to answer questions such as, 'in what cities are my skills in demand?', 'What job titles match my current skillset?', 'what skills do I need to develop for my desired job title?'

What I find most interesting, and will try to reveal the patterns of, is the complex interlacing between job titles and skills within industries. What are the niche skills that can help broaden the scope of your job search? Are there emerging or established hotspots for specific industries? The knowledge mined from this dataset is not only academic, but also immensely practical. It could be useful for guiding educational institutions in relevant curriculum development, and also helping job seekers choose in-demand skills to attain.

In essence, my project seeks to make it easy to make informed choices about your career, powered by data-driven insights from the heart of the professional world: LinkedIn.

## Literature survey

The first piece of literature which I surveyed in preparation for this project is entitled, "Salience and Market-aware Skill Extraction for Job Targeting," which is a research paper directly from engineers at LinkedIn which elaborates on systems utilized by LinkedIn for skill extraction. This LinkedIn system addresses the limitations of traditional text-based approaches which often overlook skill salience (importance) and market dynamics. with the integration of market supply and demand factors, the

method utilized by significantly enhances job recommendation and skill suggestion accuracy, as evidenced by improved application rates (+1.92% job application rate) and reduced suggestion rejection rates (-37% rejection rate). Through detailed case studies, this paper offers essential insights into how these techniques vary across different industries, countries, and skill levels, demonstrating the practical benefits of their market-aware skill extraction system. These systems inform the inherent value of my own skill extraction and the accuracy of the insights I intend to attain.

The second piece of literature that I analyzed in preparation for my project is an open source Jupyter Notebook entitled, "Analysis of 1.3 Million Jobs on LinkedIn," authored by Galal Qassas. This notebook provides a detailed exploration into the skill demands reflected in LinkedIn the 1.3 million job postings held within the dataset. By employing Python libraries including numpy, pandas, matplotlib, and seaborn, the notebook systematically parses and analyzes the dataset to identify prevalent skills across job descriptions. Notably, the analysis highlights the top 15 skills in demand, including communication, teamwork, and customer service, and uncovers that an average of 19 specific skills is typically required to apply for jobs on LinkedIn. This analysis is highly relevant to my project given how it offers a direct glimpse into the skills landscape, aiding in my foundational understanding of LinkedIn skill trends as well as the most common demands within the job market as a whole, thus aligning closely with my objective of mapping career paths and skill requirements by industry.

Lastly, the open-source Jupyter Notebook entitled, 'EDA on LinkedIn Jobs and Skills' conducts a comprehensive exploratory data analysis using Python and libraries such as numpy, pandas, matplotlib, and seaborn. It first confirms the dataset's integrity by ensuring that there are no duplicate entries, and that all columns are non-null. The analysis further explores the dataset to discover some surface level insights including top job titles, leading companies by

job postings, and primary job locations, highlighting cities and regions with the most postings. Additionally, it counts the number of jobs associate v.s. mid level and senior positions, and position settings including remote, hybrid, or onsite positions which gives a detailed overview of the dataset as a whole.

## Proposed Work

To grant clarity in the 2024 job market landscape and provide valuable insights, my project begins in the data collection phase which will leverage the "1.3 M LinkedIn Jobs & Skills (2024)" dataset. The initial step in this process involves data cleaning to remove listings either with incomplete information or those lacking in job summaries or skill details. Following this initial step, I will implement a validation process that ensures that the integrity of a selected subset of job postings through manual checks.

My data preprocessing stage will integrate job postings with their corresponding skills given that the source data formats this information in separate tables, transforming this data into a structured format most convenient to my analysis. Job titles will then be then categorized into broad industry sectors, with a new 'industry' column created for better classification. This structured data will form the foundation for a two sided approach.

First, I will conduct a time-series analysis to identify and visualize the job market trends for 2024, with a focus on regional industry booms and the popularity of specific job titles within those regions and industries. This analysis aims to pinpoint where specific industries are thriving geographically, answering the question of where job opportunities are emerging most heavily.

Second, I will utilize clustering techniques, with k-means as the prime candidate, to group similar job titles and associated skills within industry sectors. This step is designed to uncover the patterns in skill demand, providing insights into the most sought after skills by region and job category. This analysis hopes to reveal skills that can broaden job search scopes and

identify both emerging and established industry hotspots.

My project sets itself apart from others both on this subject and on this dataset by not only mapping the current state of the job market extensively, but also providing a forward thinking perspective on skill demand trends and job opportunities. My current approach aims to arm job seekers as well as educational institutions and industry professionals with the knowledge to make more informed decisions, helping to bridge the gap between skill development and market needs.

## Data Set

The dataset in question is made up of two primary CSV files which are entitled `job_skills.csv` and `linkedin_job_postings.csv`, comprised of 1,348,435 unique posings. Luckily, these entries are entirely non-null in all columns with the exception of only 19 entries throughout the whole dataset. The `job_skills.csv` file contains a comma separated list of skills, as well as a URL foreign key for joining to the other table:

1. **job\_link**: String (Nominal) - URL, Foreign Key
2. **job\_skills**: String(Nominal) - comma separated list of skills required for that job

The second table, `linkedin_job_postings.csv` holds the comprehensive details about each posting:

1. **job\_link**: String (Nominal) - URL, Primary Key
2. **last\_processed\_time**: DateTime (Interval) - When posting was last processed
3. **got\_summary**: Boolean (Binary) - Indicates if posting summary was successfully extracted
4. **got\_ner**: Boolean (Binary) - Whether Named Entity Recognition was performed
5. **is\_being\_worked**: Boolean (Binary) - Status of posting
6. **job\_title**: String (Nominal) - Title of posting
7. **company**: String (Nominal) - Company that is posting

8. **job\_location**: String (Nominal) - Where the posting is located
9. **first\_seen**: DateTime (Interval) - Date posting was made
10. **search\_city**: String (Nominal) - City used to find posting

This massive dataset that is for the most part devoid of null values will set a strong foundation for impactful analysis and insights into current job market dynamics.

## Evaluation Methods

I intend to compare my foundational findings to the existing job market analytics which were discovered in the 'EDA on LinkedIn jobs and skills', and 'Analysis of 1.3 Million Jobs on linked-in' Jupyter Notebooks. By utilizing their findings as a benchmark I will assess the uniqueness and depth of my own insights. I also intend to quantify the coherence and distinctiveness of clusters within my analysis, adopting a threshold such as a silhouette score of at least 0.5 (a metric for measuring cluster quality) that ensures clusters are both internally consistent and distinctly separate from each other. This approach will ensure the robustness of the clustering methodology. Additionally, I plan on leveraging the Apriori algorithm, I will discover all of the frequent itemsets among skills and job titles across industry clusters, thus further informing this projects insights into the 2024 job market's structure and dynamics.

## Tools

I plan to utilize Python as well as it's flagship data analysis libraries including Numpy, Pandas, Matplotlib, and Seaborn. I intend to utilize Git and Github for version control and regular project updates. For any formulas and techniques, I intend to use 'Data Mining Concepts and Techniques - Third Edition.'

## Milestones

I plan to complete work on this project on regular intervals in accordance with class deadlines:

- **03/18 - 03/25** - Begin data collection and initiate data cleaning to remove sub-par positions.
- **03/26 - 04/02** - Conduct validation on subsets of postings and integrate postings & skills.
- **04/03 - 04/10** - Classify job titles and generate structured format suitable for analysis.
- **04/11 - 04/18** - Conduct time-series analysis, identifying trends in regional industry booms and job title popularity.
- **04/19 - 04/26** - Employ clustering to group similar job titles and associated skills.
- **04/27 - 05/01** - Complete analysis and compile findings.

## References

- [1] Baoxu Shi, Jaewon Yang, Feng Guo, and Qi He. 2020b. Salience and market-aware skill extraction for job targeting. (May 2020). Retrieved March 16, 2024 from <https://arxiv.org/abs/2005.13094>
- [2] Galal Qassas. 2024. Analysis of 1.3 million jobs on linked-in. (February 2024). Retrieved March 16, 2024 from <https://www.kaggle.com/code/galalqassas/analysis-of-1-3-million-jobs-on-linked-in/notebook>
- [3] Varnika Bhardwaj. 2024. EDA on linkedin jobs and Skills. (February 2024). Retrieved March 16, 2024 from <https://www.kaggle.com/code/varnikabhardwaj/eda-on-linkedin-jobs-and-skills>