# Market-Basket Analysis of LinkedIn Job&Skills Dataset

Salima Tankibayeva

July 11th 2024

**Abstract**

This report provides a market-basket analysis of LinkedIn Job&Skills Dataset using the Apriori algorithm implementation. The main goal is to identify common sets of job skill elements that occur often together, which give insight into the skill requirements for the scope of jobs. The report describes the selected dataset, data organization, preprocessing methods used, algorithm implementation from scratch, and experimental results.

## 1 Introduction

Market-basket analysis is a data mining technique that is used to identify relationships between items in large data sets. For this project, this method was applied to the LinkedIn Job&Skills dataset in order to identify common job skills' combinations.

## 2 Data Collection

The dataset used in this analysis: "1.3M LinkedIn Jobs and Skills 2024" dataset from Kaggle. It includes detailed information on job descriptions and the required skills. The dataset consists of 3 files, for the project we were prescribed with "job_skills.csv" which contains job IDs and their associated skills. There are two columns:

- Job ID: Unique identifier for each job posting.
- Job skills: list of skills required for the job.

## 3 Preprocessing

First, I loaded data from Kaggle and used pandas dataframe. Next step was to check if there are missing values and solve an issue with them. The lists of job skills for each job posting were extracted from the DataFrame and stored in a list called transactions. This list is used as input for the frequent itemset generation algorithm. Then in order to ensure that each skill within a job posting is considered only once, we converted the list of skills for each job posting into a set. Removing duplicate skills within the same job posting, because otherwise it could skew the frequency counts in the frequent itemset generation algorithm. The transactions are used for the frequent itemset generation algorithm. The algorithm counts the frequency of each skill and then iteratively joins itemsets to find frequent itemsets of larger sizes.

## 4 Algorithm Implementation

- Frequent itemset Generation Algorithm

  Joining itemsets: itemsets of larger sizes are generated by joining smaller itemsets that meet the minimum support threshold. This process continues iteratively until no more frequent itemsets can be generated. em More space to store data and instructions closer to the CPU.

- Association Rule Generation Algorithm

  For each frequent itemset, the algorithm generates association rules by splitting the itemset into antecedent and consequent parts. The confidence of each rule is calculated, and rules that meet the minimum confidence threshold are retained.

## 5 Experimental results

The minimum support was set 0.01, (1%) to identify itemsets that appear in at least 1% of transactions. The minimum confidence was set to 0.001 (0.1%) for generating association rules.

According to the set conditions, frequent itemsets were identified, such as: Highly Frequent Skills:

Communication Skills:

- 'Communication' with a support of 0.27
- 'Teamwork' with a support of 0.17
- 'Communication skills' with a support of 0.08
- 'Problem Solving' with a support of 0.08
- 'Customer Service' with a support of 0.07
- 'Problem solving' with a support of 0.07
- 'Time management' with a support of 0.06
- 'Training' with a support of 0.06

- 'Attention to detail' with a support of 0.06

These results indicate that communication-related skills, teamwork, and problem-solving are highly valued across job postings.

Moderately Frequent Skills:

- 'Leadership' with a support of 0.12
- 'Collaboration' with a support of 0.07
- 'Scheduling' with a support of 0.05
- 'Customer service' with a support of 0.05
- 'Sales' with a support of 0.05
- 'Microsoft Office Suite' with a support of 0.05

Skills related to leadership, collaboration, and specific technical proficiencies like Microsoft Office are moderately frequent, reflecting their importance in various roles.

Less Frequent but Notable Skills:

- 'Attention to Detail' with a support of 0.04
- 'Adaptability' with a support of 0.04
- 'Documentation' with a support of 0.04
- 'Flexibility' with a support of 0.04
- 'Patient Care' with a support of 0.04

These skills, while less frequent than the top-tier skills, still play a crucial role in specific job functions, particularly in roles requiring precision, adaptability, and patient interaction. Rare Skills:

Skills like 'Cleaning', 'Sanitation', 'Food safety', 'Dental Insurance', 'Health Insurance', 'Vision Insurance', 'Integrity', 'PowerPoint', 'Word', 'Standing', 'Walking', 'Fastpaced environment', and 'High school diploma or equivalent' have a support of 0.01.

These skills appear infrequently across job postings, suggesting they are more niche or specific to certain industries or roles.

Despite identifying frequent itemsets, no association rules were generated with the specified confidence threshold. This suggests that while certain skills appear frequently, they do not strongly co-occur with other specific skills.

# 6 Discussion

The provided code and detailed description of preprocessing steps and parameter settings ensure that the experiments can be replicated. The dataset and transformations are explicitly defined to achieve consistent results. The algorithm's scalability is influenced by the min support parameter. Lowering the min support threshold increases the number of candidate itemsets, which can lead to higher computational costs. Adjusting min support and min confidence parameters can balance the number of itemsets/rules generated and computational efficiency. Experimenting with different minimum support and confidence thresholds could potentially uncover more association rules, especially for less frequent but important skills.

The high frequency of communication-related skills and teamwork underscores their universal importance in the workplace. Effective communication and collaboration are essential in almost all job roles, which is reflected in their high support values. The emphasis on problem-solving skills and leadership indicates a significant demand for individuals who can navigate challenges and lead teams. These skills are critical for roles that require strategic thinking and management capabilities. Skills with lower support values, such as those related to specific certifications or physical tasks, point to specialized requirements within certain industries. These skills are crucial for roles that demand specific technical knowledge or physical capabilities.

# 7 Conclusion

The comprehensive analysis of job postings has provided valuable insights into the prevalent skills demanded in the current job market. By employing frequent itemset mining, we identified key skills and competencies that appear with notable frequency, offering a clear picture of what employers prioritize across various industries.

- High-Demand Skills:

  Communication and Teamwork: The results highlight that skills related to communication and teamwork are the most frequently cited, emphasizing their critical importance in almost every job role. Effective communication, both verbal and written, alongside the ability to work well in teams, are foundational competencies that employers consistently seek.

- Problem Solving and Leadership:

  The prominence of problem-solving and leadership skills in the data indicates a strong demand for individuals who can navigate complex challenges and lead teams towards achieving organizational goals. These skills are essential for roles that require strategic thinking, decision-making, and management capabilities.

- Technical Proficiency:

Technical skills, particularly proficiency with Microsoft Office Suite and other administrative tools, are also frequently mentioned. This reflects the necessity for candidates to possess a solid foundation in basic technological and organizational tools, which are crucial for efficient job performance in many roles.

- Specialized and Niche Skills:

  Although less frequent, certain specialized skills, such as those related to specific certifications, physical tasks, or industry-specific knowledge, are critical for particular roles. These skills, while not universally required, are indispensable in their respective contexts and underscore the diversity of job requirements across different sectors.

This study successfully identifies the key skills that are currently in high demand across job postings, offering practical insights for various stakeholders. By focusing on these findings and implementing the recommended actions, job seekers can enhance their employability, educators can better prepare students for the workforce, and employers can optimize their recruitment and training strategies. The continuous evolution of the job market necessitates ongoing analysis to stay abreast of emerging trends and ensure alignment with the ever-changing skill requirements.

## 8 Implecations

This study's findings have several important implications for job seekers, educators, and employers:

- Job Seekers: Understanding the high-demand skills can guide job seekers in enhancing their resumes and developing competencies that are most valued by employers. Focusing on improving communication, teamwork, problem-solving, and technical skills can significantly increase employability.

- Educators: The insights from this analysis can inform curriculum development, ensuring that educational programs equip students with the skills that are in high demand. Emphasizing practical communication skills, teamwork exercises, problem-solving activities, and technical training can better prepare students for the job market.

- Employers: Employers can use these findings to refine their recruitment processes, ensuring they prioritize candidates with the most critical skills. Additionally, understanding the importance of these skills can help in designing effective training and development programs for current employees.

# 9 Recommendations

To build on the insights gained from this analysis, the following recommendations are proposed:

- Refinement of Analysis: Experimenting with different support and confidence thresholds could reveal more association rules, providing deeper insights into the co-occurrence of skills. This can help identify combinations of skills that are particularly valuable.
- Segmented Analysis: Conducting a segmented analysis by industry or job type can offer more granular insights into specific skill requirements. This approach can help uncover patterns that may not be evident in the overall dataset.
- Temporal Trends: Analyzing trends over time can reveal how the demand for certain skills is evolving, providing valuable information for future workforce planning and career development strategies.

# 10 Further implementation

Conducting a segmented analysis by industry or job type could provide more granular insights into the specific skill sets required for different roles. This could help identify patterns that are not evident in the overall dataset. Analyzing trends over time could reveal how the demand for certain skills is evolving. This would be particularly useful for understanding emerging skills that are becoming increasingly important in the job market.

# 11 Declaration

"I/We declare that this material, which I/We now submit for assessment, is entirely my/our own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my/our work, and including any code produced using generative AI systems. I/We understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. This assignment, or any part of it, has not been previously submitted by me/us or any other person for assessment on this or any other course of study."