Modelling of Employee Skill Demand Evolution in the US using Dynamic Topic Modelling and Embedding Models

Rajireddy Gudipati
Marketing and Data Science
Regis University
Regis University , Denver, CO, USA
rgudipati@regis.edu

EXECUTIVE SUMMARY

The practicum project addresses the problem of the understanding of skill demands in the dynamic business environment of big economies. Understanding of the skill demand requires evaluation of the evolution of skill demand overtime and contextualizing this demand in terms of the present workforce and the national averages. Through the evaluation, a complete picture of the skill demand for big economies can be generated for use by policy makers and companies. Data from the current US Federal Job Listing, Combined Local and Federal Payroll, City Level Payroll and National Labor Statistics was collected and utilized for the modelling of the employee skill demand in US. The data preparation for the project included Natural Language Processing (NLP) for skill information extraction and text processing from the job listing data, and, data cleaning and data agglomeration across all the three datasets. Using the Dynamic Topic Modelling and Embedding Models, the employee skill demand evolution in was shown to be consistent with the trends in employment indicators, different across the national, local and city levels, and largely similar across the top paying jobs.

I. INTRODUCTION AND BACKGROUND

Human resource is not limited to availability and lack of availability, but instead extends to cover skill and demand. The analysis of human resource at a skill demand level provides for contextualized insight into human resource (Grama & Todericiu, 2025). This perspective to human resource is especially important when governance and organization is considered is big economies such as the US. Companies and policy makers would benefit from a good understanding of human resource in terms of the skills demand and how it evolves (Jotabá, 2022; Dhuniganh et al., 2024). Extending the analysis of human resource beyond availability and lack of availability can be completed through the modelling of the evolution of skill demand with dynamic topic modelling and embedding models allowing for insights to be gained on how skill demand evolves for the US.

A. *Problem Statement and Motivation*

The US represents a big economy where the large population and level of economic activities make management of human resource important. Magnitude, in terms of the size of the organization or economies, amplify the critical nature of the evolution in skill demands in human resource (Israel & Rutainurwa, 2025). The administration responsible for the governance of big economies need to have a complete picture of the nature of skill demand for the economy and how this skill demand evolves overtime so as to better manage the economy from the perspective of both the employees and employers. Skill may overlap different job types, which makes collection of data on skills more complex and different from collection of data on jobs available (Jotabá, 2022; Mishra et al., 2025); requiring extraction and processing of texts using Natural Language Processing (NLP) and analysis using Topic Modelling and Embedding Models.

II. Data and Methodology

A. Data

Data from the US Federal Government Job Postings API sourced from https://developer.usajobs.gov/api-reference/ will form the primary data source. Three other data sources will be utilized; US Federal Payroll Data (from https://www.census.gov/data/datasets/2024/econ/apes/annual-apes.html), New York City Payroll API (https://data.cityofnewyork.us/resource/k397-673e.json) and National Labor Statistics (https://data.bls.gov/oes/#/industry/000000). Data from the four sources will be cleaned and merged for evaluation of the evolution in dynamic topic modelling.

B. Data Description

The data used for the project was sourced from the following four sources, with the description of the variables provided below in Table 1; US Federal Government Job Postings API, US Federal Payroll Data, New York City Payroll API and National Labor Statistics.

Table 1

Source	**Data**	**Variable**
US Federal Government Job Postings API	US Federal Government Job Postings	Year
		State
		Job Descriptions
		Wages
National Labor Statistics	Employment Indicators Data	Year
		State
		Unemployment Rate
		Average Hourly Pay
US Federal Payroll Data	Combine Local and Federal Government Job Postings	Job Descriptions
New York City Payroll API	New York City Job Postings	Job Descriptions

C. Analysis Implementation

The methodology implemented for the investigation into employee skill demand evolution in US is presented below in Figure 1.

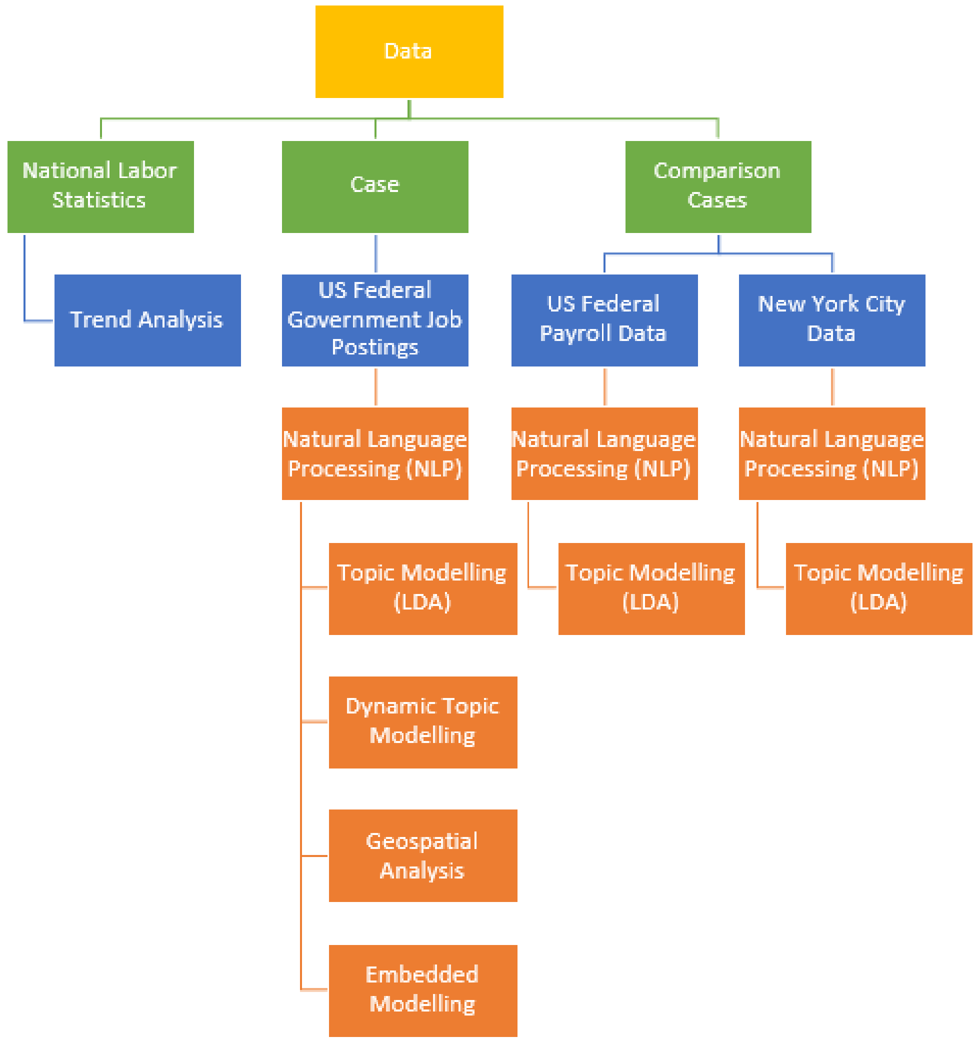


Figure 1

The summary in Figure 1 above shows that trend analysis and three cases for modelling were employed in the project. Trend Analysis provides an overview of time series data (Siswanto et al., 2025). In this case, the project obtained an overview of the time series for current and historic employment indicators; unemployment rate and average hourly pay. The three cases examined for the evolution of the skill demand in the US were the National/Federal level, Combine Local and Federal Government level and the City level for New York City. The case of the National/Federal level formed the primary case of focus for the project, since it presented a holistic view across the US. The Combine Local and Federal Government level, and the City level case for New York City formed comparison cases for which the National/Federal level case was compared against. For each of the three cases, job description data was first prepared through Natural Language Processing (NLP). Natural Language Processing (NLP) extracts key terms from textual data (O'Shaughnessy, 2026). For this project, the terms extracted from textual data formed skill signals. In terms of textual data, signals allow for derivations of implied meanings given the signals. For this case, the signals were useful for implying skills in the evaluation of the evolution of skill demand in the US. Following data preparation, the data for each case was then modelled through Topic Modelling. According to Hairani et al. (2024), Topic Modelling employs the Latent Dirichlet Allocation (LDA) to group terms into topics based on thematic analysis. From the skill signals, the LDA then grouped the implied skills into topics, based on thematic analysis, to create general impressions (skill demand set) on the evolution of skill demand. The topic modelling was completed for the primary case and the two comparison cases. The skill signals, implied skills and general impressions of the skill demand for the National/Federal level was then compared against the skill signals, implied skills and general impressions of the skill demand for the two comparison cases. In order to observe the evolution of the skill demand across the US, there is need to observe the skill signals, implied skills and general impressions of the skill demand over time. Dynamic Topic Modelling conducts LDA overtime to capture evolution of skill demand. In Dynamic Topic Modelling, each observation is allocated probabilities for each of the topics identified to determine which topic the observation is likely to fall under (Hairani et al., 2024). Observing how the probabilities for each of the topics changes over time constitutes the Dynamic Topic Modelling. For the skill signals and implied skills for the National/Federal level, the general impressions of the skill demand forming the topics, and hence skill demand sets, formed the target for the Dynamic Topic Modelling. Observing how the probabilities for each of the skill demand sets (topics) changes over time allowed for the determination of the evolution of the skill demand in the US at the National/Federal level; and comparison against trends in unemployment rate and average hourly pay. The computed probabilities for the implied skill demand sets (topics) from the Topic Modelling for each of the observations, allowed for the observations for the distribution of the skill set demands across the US. Geospatial analysis plots attributes based on spatial characteristics of a record into maps (Muhammad et al., 2025). In this case, Geospatial Analysis allowed for the mapping of topics from the LDA across the US. The demand for the different implied skill demand sets can then be observed across the US. The topics from the LDA can also be used to gauge the extent to which skill demands for various job positions are similar at the National/Federal level. Embedded Models evaluate the similarity and semantics in textual data (Ajallouda et al., 2025). Hence, an embedded model, the pre-trained all-MiniLM-L6-v2 embedding model, was employed for the skill signals, implied skills and skill demand sets to determine the similarity and uniqueness of the skill demands across the best paying jobs at the National/Federal level.

III. Results, Key Findings and Insights

A. *Trend Analysis*

Figure 2 below shows the line chart for the trend in the unemployment rate. We note that the unemployment rate has largely been stagnant, fluctuating over a range of values. The period between 2020 and 2021 represented an outlier where there was a very steep spike in unemployment followed by a steep then gentle fall in the unemployment rate; due to the COVID-19 pandemic.

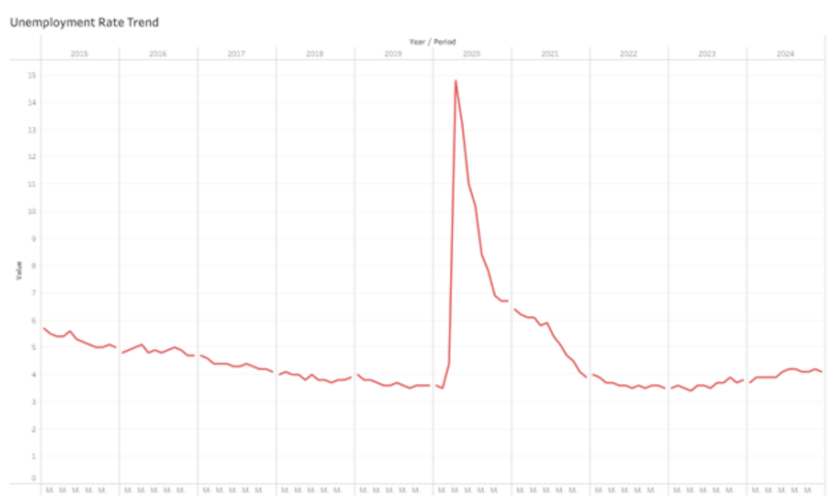


Figure 2

The results of the trend analysis for the average hourly pay in shown in the line chart in Figure 3 below. The plot shows a steady rising trend over time with a notable jump in the year 2020 followed by a drop; this can similarly be attributed to the COVID-19 pandemic.

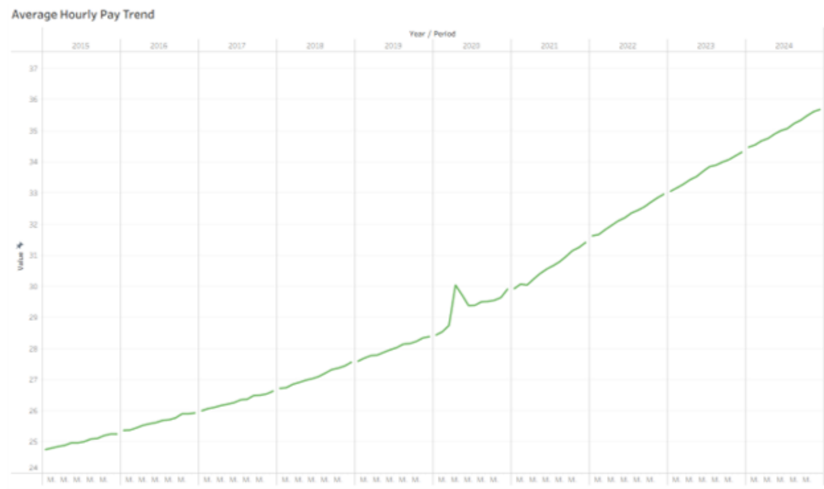


Figure 3

B. Primary Case: National/Federal Level Skill Demand

The results of the NLP and LDA for the National/Federal Level are shown below in Figure 4. The plot shows that the skill signals and implied skills are such that;

- From the first topic, the skill demand related to the aviation, safety, management and medical fields.
- From the second topic, the skill demand related to engineering and human resource.
- From the third topic, the skill demand related to the medical and engineering fields.

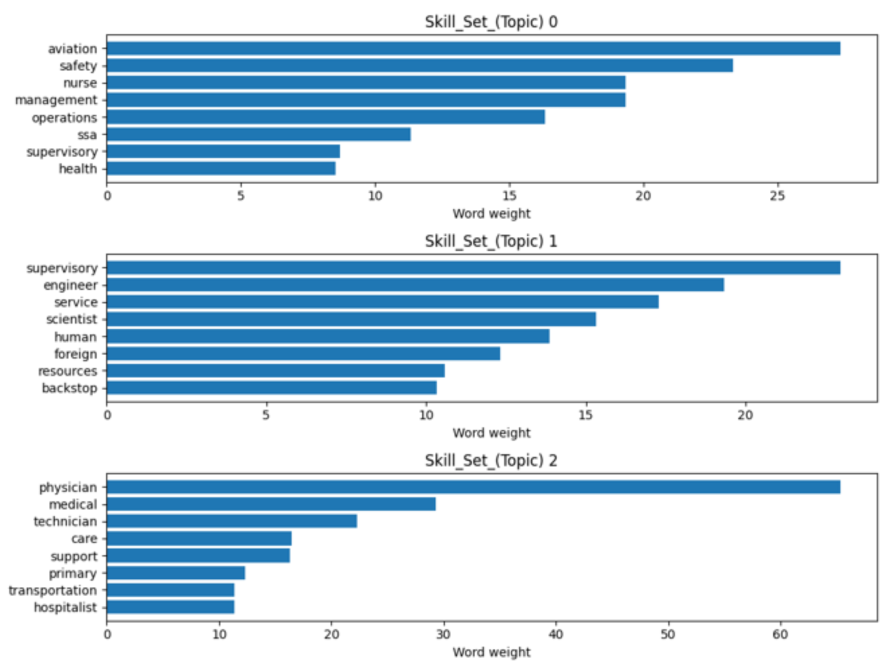


Figure 4

The line charts in Figure 5 present the Dynamic Topic Modelling for the evolution of the skill demand over time, from 2013 through to 2024. The demand for skill set 0 is initially high, then flattens between 2014 and 2018 before having a series of jumps and falls with a decline in recent time. The demand for skill set 1 is relatively flat from 2013 to 2023, but shows a rising trend in recent time. The demand for skill set 2 has three peaks showing possible seasonality (unequal season lengths) with the demand experiencing decline in recent time; a rise might characterize the next season.

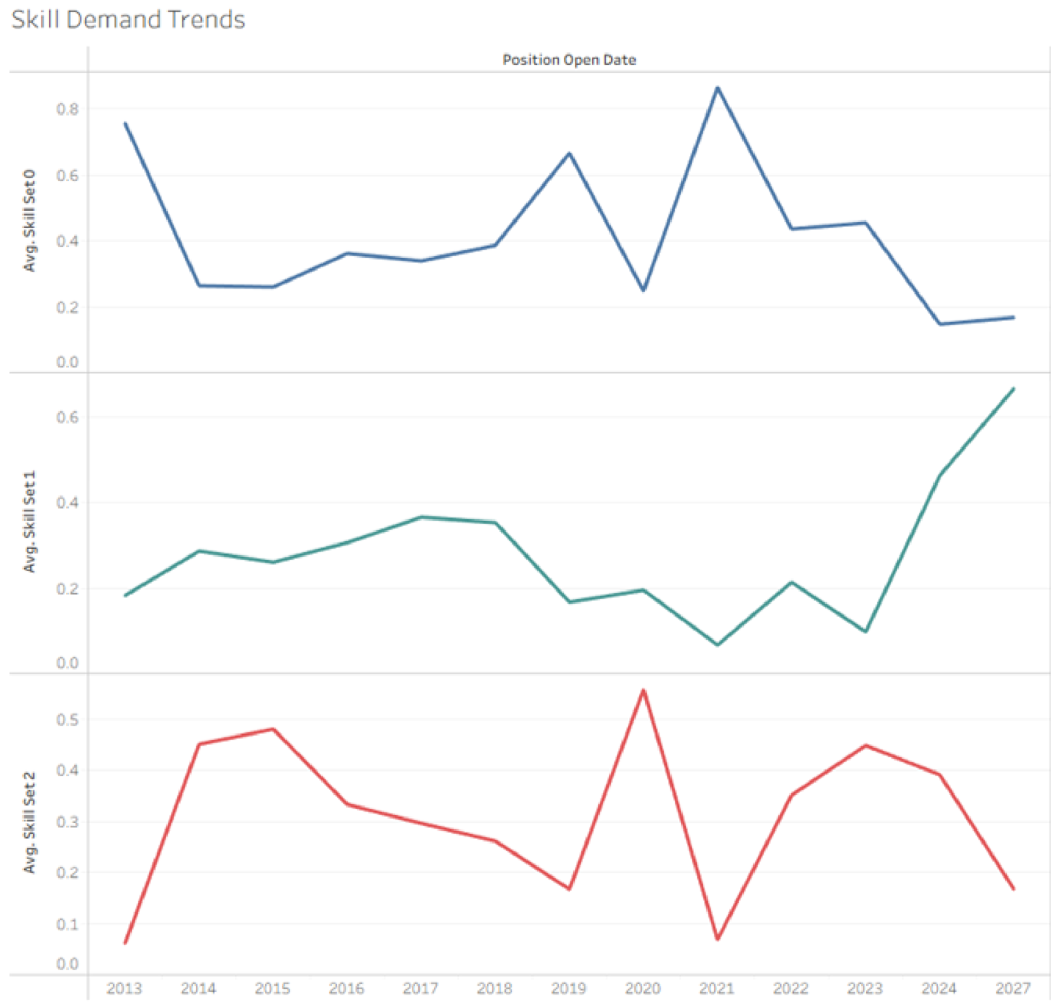


Figure 5

The results of the geospatial analysis for the distribution of the skills in skill set 0 are shown in the map in Figure 6 below. We note that the demand for the skills in skill set 0 were highest in the states of Montana, Maine, Connecticut, New Hampshire, Massachusetts and Kentucky, and lowest in the states of Nebraska and Alaska.

Skill Set 0

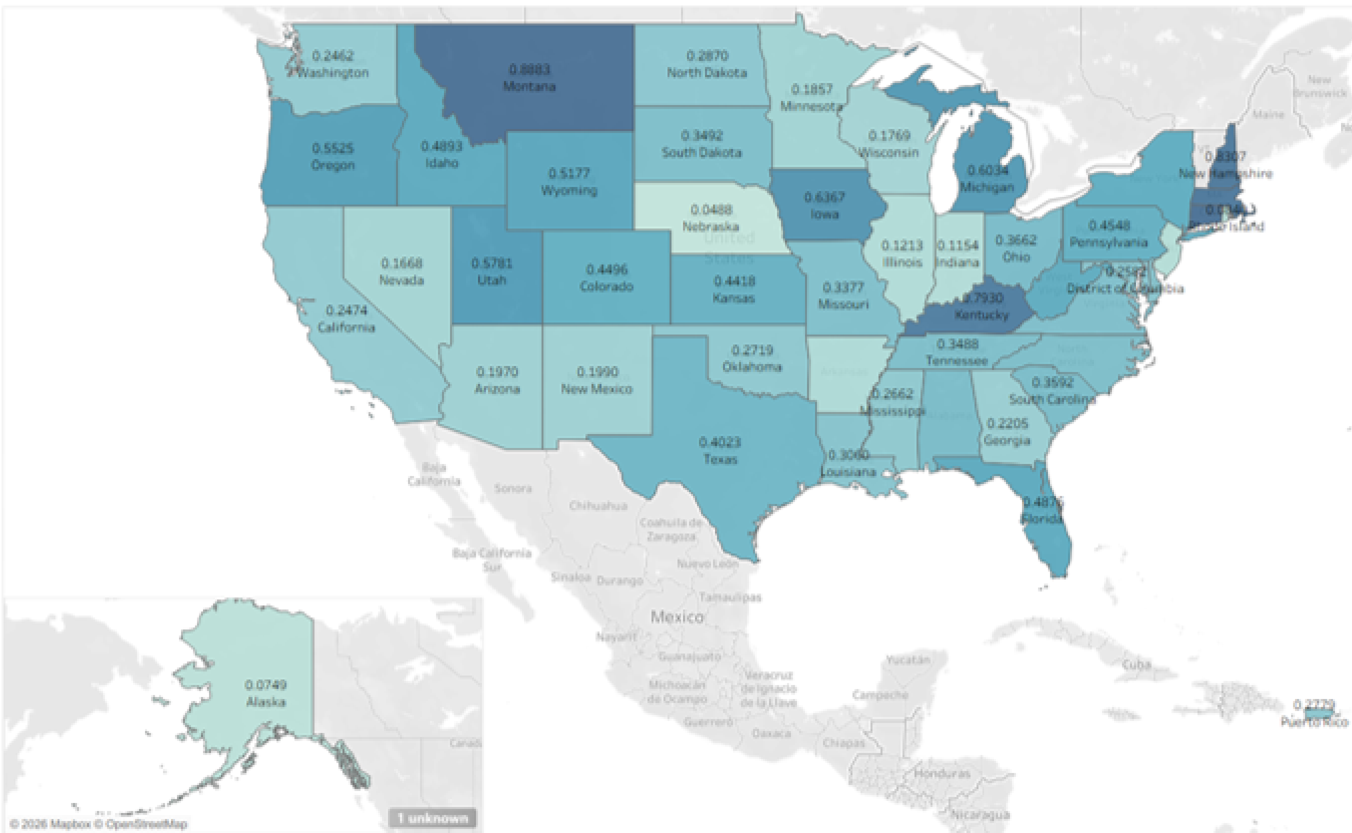


Figure 6

The results of the geospatial analysis for the distribution of the skills in skill set 1 are shown in the map in Figure 7 below. We note that the demand for the skills in skill set 1 were highest in the states of Nebraska, Nevada, and Mississippi, and lowest in the states of Montana, Kentucky, New Hampshire, Rhode Island and Massachusetts.

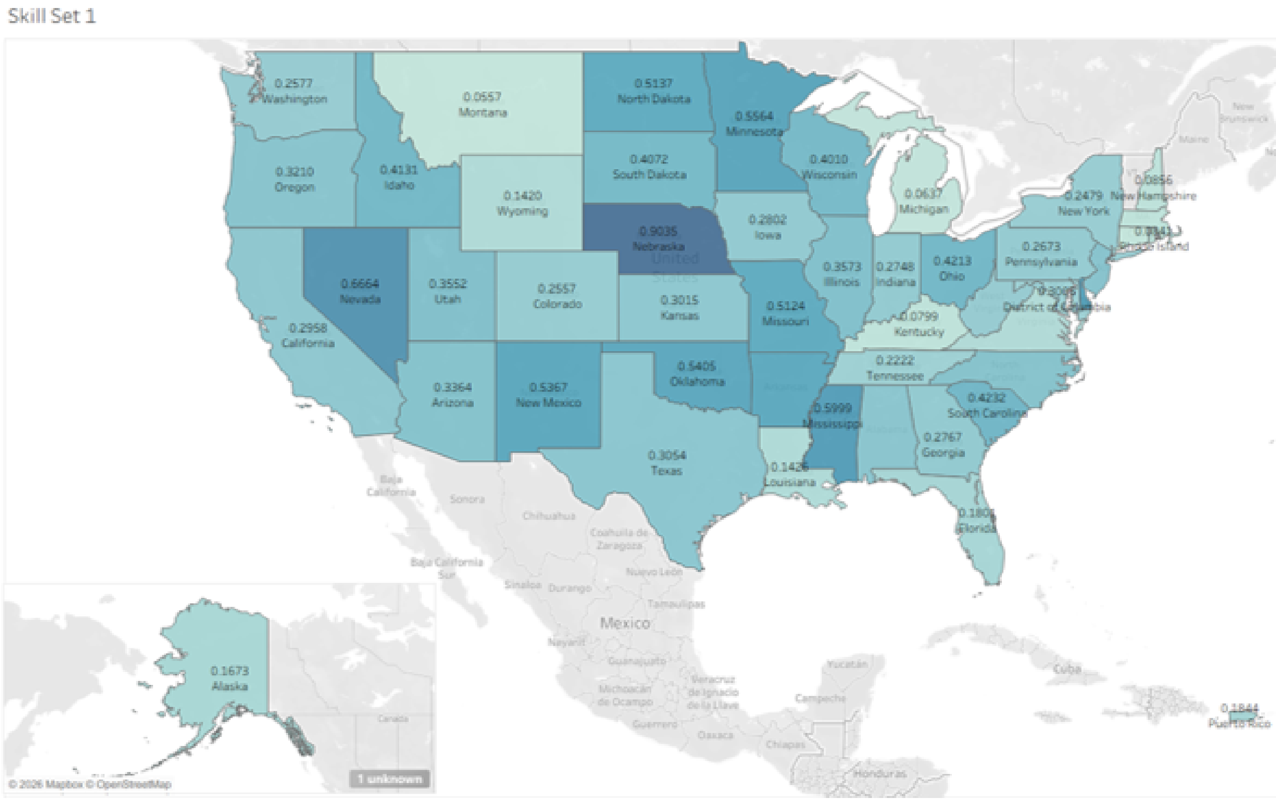


Figure 7

The results of the geospatial analysis for the distribution of the skills in skill set 2 are shown in the map in Figure 8 below. We note that the demand for the skills in skill set 2 were highest in the states of Rhode Island, Alaska and Indiana, and lowest in the states of Nebraska, Montana, Utah and Iowa.

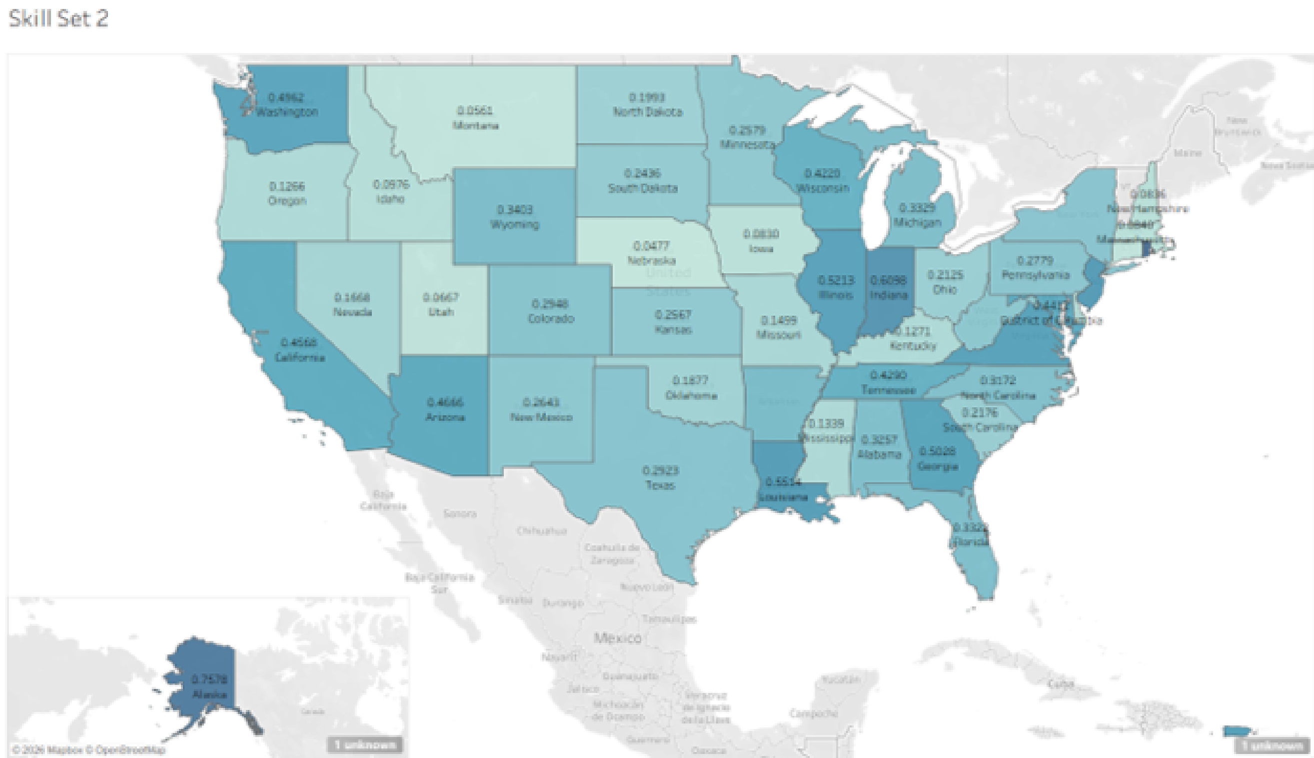
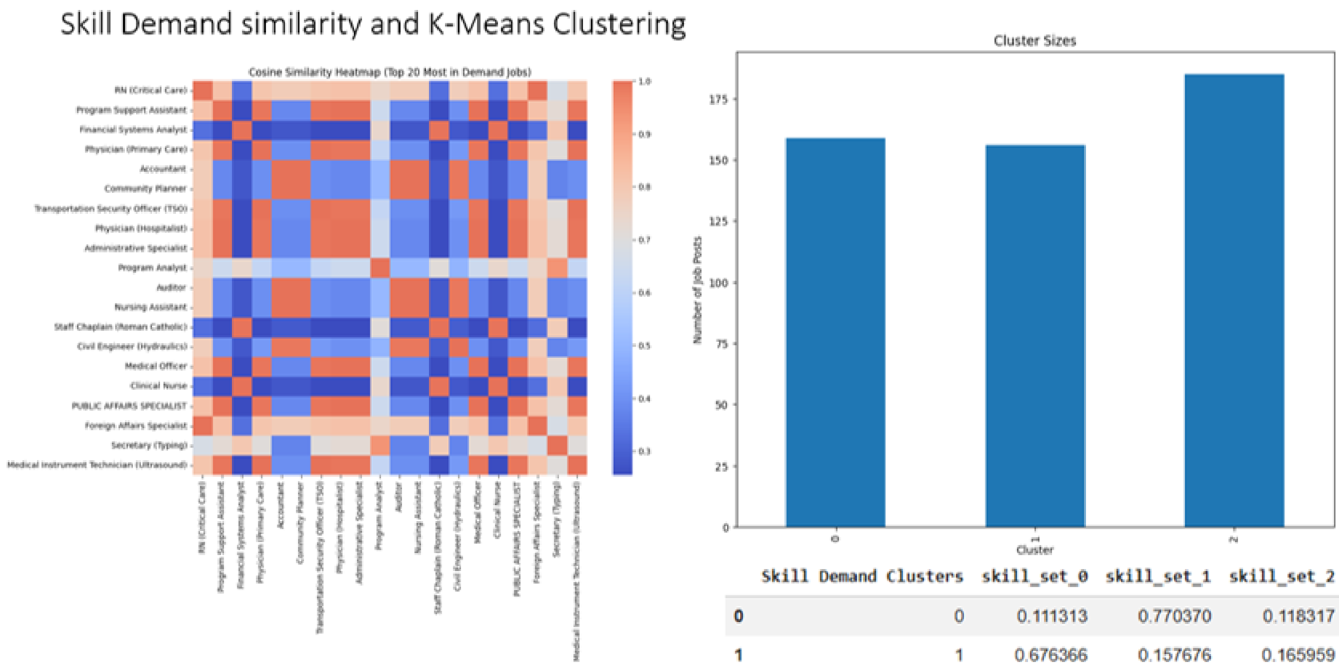


Figure 8

The results of the embedded models for the similarity of the implied skills from the NLP and LDA are shown below in Figure 9 in the form of a heatmap for the top 30 best paying jobs at the National/Federal level and bar plot for the clustering of the National/Federal level jobs by the similarity score. We observe from the heatmap that skills in Financial System Analyst, Staff Chaplain (Roman Catholic) and Clinical Nurse are the most unique skills among the top 20 most in demand positions. The bar plot showing the clustering of the top paying jobs by similarity showed that majority of the jobs fell under cluster 2, followed by cluster 0 then cluster 1. Additionally, on average, majority of jobs in skill demand cluster 0 had highest demand for skills in skill demand set 1, majority of jobs in skill demand cluster 1 had highest demand for skills in skill demand set 0 and majority of jobs in skill demand cluster 2 had highest demand for skills in skill demand set 2.



	Skill Demand Clusters	skill_set_0	skill_set_1	skill_set_2
0	0	0.111313	0.770370	0.118317
1	1	0.676366	0.157676	0.165959
2	2	0.110621	0.130044	0.759335

Figure 9

C. Comparison Cases: Combined Local and Federal Level and City Level

The results of the NLP and LDA for the Combined Local and Federal Level are shown below in Figure 10. The plot shows that the skill signals and implied skills are such that;

- From the first topic, skills in education were most predominant.
- From the second topic, skills in fire and safety management were most predominant.
- From the third topic, skills in governance and policing were most predominant.

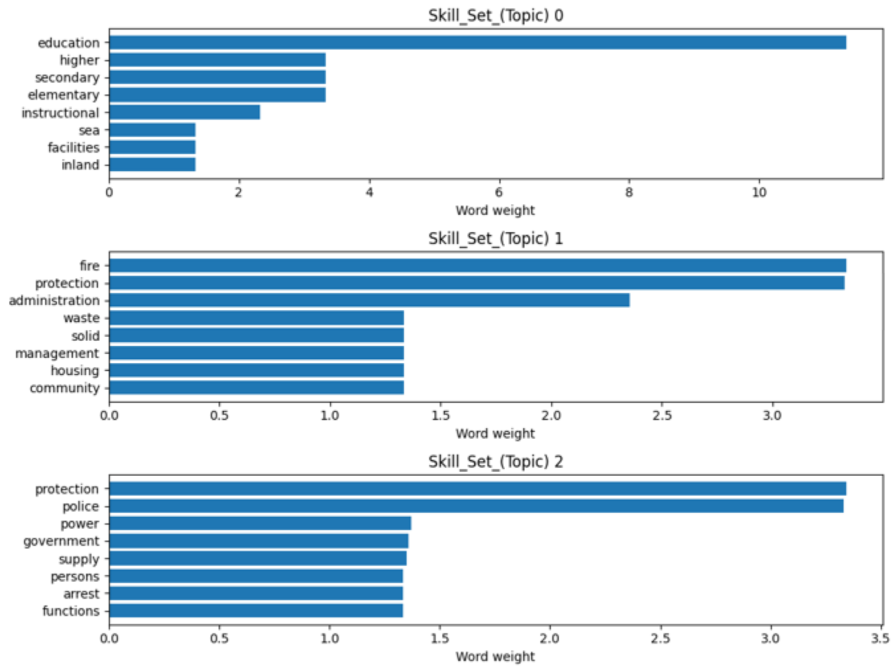


Figure 10

The results of the NLP and LDA for the City Level, in the case of New York City, are shown below in Figure 11. The plot shows that the skill signals and implied skills are such that;

- From the first topic, skills in administration and governance were most predominant.
- From the second topic, skills in social services and administration were most predominant.
- From the third topic, skills in social services and administration were most predominant.
-

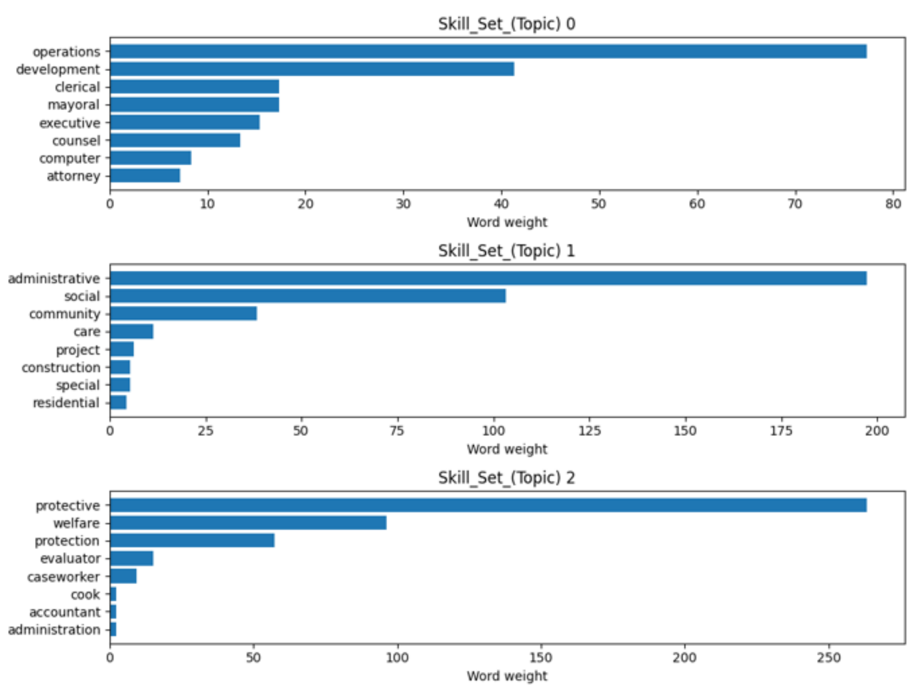


Figure 11

The comparison between the skill demand across the three skill demand sets for the primary case and the two comparison cases is shown below in Table 2.

Table 2

	Case	Comparison Cases	
	National Level - Federal Government Level (Longitudinal)	**Local Government and Federal Government (Cross-sectional)**	**City Level – New York City**
Skill Set 0	Aviation, Safety, Management and Medical fields	Education-Related skills	Administration and Governance related skills
Skill Set 1	Engineering and Human Resource	Fire and Safety Management skills	Administration and Social Service skills
Skill Set 2	Medical and Engineering skills	Governance and Policing skills	Administration and Social Service skills

IV. Conclusion

The results from the analysis show that consistency exist between skill demand trends and trends in both unemployment and wage rates. Also, the skill demand at national level, local government level and the business level largely differ, pointing a difference in skill needs across the three levels of employment. Further, the results show that the distribution of skills across the US states was such that the states with highest likelihood of skill demand for skills in skill set 0 are shown to have the lowest likelihood of skill demand for skills in skill set 1, and vice versa. Future research should provide the perspective of the business sector by using the evaluation of the evolution of skill demand in this project as a blue print for a wider investigation.

References

Ajallouda, L., Hassani Saissi, M., & Zellou, A. (2025). *Embedding models: A comprehensive review with task oriented assessment.* International Journal of Advanced Computer Science and Applications, 16(10), 539–550. http://dx.doi.org/10.14569/IJACSA.2025.0161056

Dhuniganh, R., Anjali, N., & Samarthy, J. (2024). *Development of managerial skills for success in a dynamic business environment.* Journal Development Manecos, 2(2), 46–56. https://doi.org/10.71435/604086

Grama, B., & Todericiu, R. (2025). *The evolution of skill dynamics in the context of the future of work.* Studies in Business and Economics, 20(2), 137–154. https://doi.org/10.2478/sbe-2025-0028

Hairani, H., Janhasmadja, M., Tholib, A., Guterres, J. X., & Ariyanto, Y. (2024). *Thesis topic modeling study: Latent Dirichlet Allocation (LDA) and machine learning approach.* International Journal of Engineering and Computer Science Applications, 3(2), 51–60. https://doi.org/10.30812/ijecsa.v3i2.4375

Israel, B., & Rutainurwa, V. (2025). *Dynamic skills for achieving profitability and long-term sustainability of start-up micro, small, and medium enterprises (MSMEs) in a competitive business environment.* Management Dynamics in the Knowledge Economy, 13(1), 85–104. https://doi.org/10.2478/mdke-2025-0006

Jotabá, M., Fernandes, C., Gunkel, M., & Kraus, S. (2022). *Innovation and human resource management: A systematic literature review.* European Journal of Innovation Management, 25(6), 1–18. https://doi.org/10.1108/EJIM-07-2021-0330

Mishra, C. E. B., Walters, D., Fraser, E. D. G., Gillis, D., & Jacobs, S. (2025). *Higher education fields of study and the use of transferable skills at work: An analysis using data from the Programme for the*

International Assessment of Adult Competencies (PIAAC) in Canada. Trends in Higher Education, 4(2), 19. https://doi.org/10.3390/higheredu4020019

Muhammad Hashim, Atta-ur Rahman, Muhammad Qasim, Muhammad Umar Farooq, Muhammad Dawood, Basit Nadeem, & Shazia Muneer. (2025). *Application of geospatial approaches for evaluation of urban growth pattern and trend prediction of Multan City, Pakistan.* International Journal of Innovations in Science & Technology, 7(9), 108–120. https://journal.50sea.com/index.php/IJIST/article/view/1512

O'Shaughnessy, D. (2026). *An overview of recent advances in natural language processing for information systems.* Applied Sciences, 16(2), 1122. https://doi.org/10.3390/app16021122

Siswanto, J., Goeltom, V. A. H., Pandawan, I. N. H., Lisangan, E. A., & Fitriani, A. (2025). *Market trend analysis and data-based decision making in increasing business competitiveness.* Sundara Advanced Research on Artificial Intelligence, 1(1), 1–8. https://journal.sundarapublishing.com/index.php/sundara/article/view/1