

# Machine Learning Is Important, But Programming Skills Rank Higher in Data Science Job Demand

Group 4: Rachel Kang, **Morgan Simmons**, Rowan Rosenblum

## Outline:

Motivation & research Question -> Data Acquisition -> Analysis -> Ranking Results -> Skill Intensity -> Conclusion and & Implications



# Overview

- **Motivation:** With rise to AI/ML programs, we wanted to see if this translated to skills needed for jobs in this market especially since we are college students potentially looking for jobs in Data Science within the next couple of years.
- **Hypothesis:** *The top three requested skills for data scientist job postings in the United States will be machine learning or AI (ML/AI) oriented.*
- **Research Question:** Which skills are most frequently requested in recent data science job postings, and do machine learning/AI skills rank highest among them?
- **Modeling Approach:** Analyzed 2025 data science job postings, extracted and standardized key technical skills, identified AI-related skills using rule-based methods, computed AI skill prevalence across roles, built a fully reproducible R workflow



# Data Acquisition

- **Initial challenge/decision:** Not being able to use INDEED data
- **Source:** Provided 2025 data science job postings dataset
- **Format:** Structured tabular data (CSV file)
- **Unit of Observation:** One row per job posting
- **Key Variables:** Job title, company, location,, skill fields
- **Size:** 944 postings
- **Acquisition:** Secondary dataset (no scraping performed)
- **Ethics & Licensing:** Public job posting data; no personal or sensitive information



# Analysis Plan + Justification

## Workflow

Dataset → Cleaning → Skill Extraction → AI /ML Classification → Results

## Challenges

- Limiting amount of data to work with
- Inconsistent formatting

## Validation

- Manual spot-checks
- Frequency diagnostics
- Fully reproducible R script

# Results & Conclusion

Rank	skill	Frequency	Percentage_of_Postings
1	python	375	50.47
2	r	306	41.18
3	machine learning	219	29.48
4	sql	202	27.19
5	aws	198	26.65
6	pytorch	148	19.92
7	spark	147	19.78
8	pandas	76	10.23
9	numpy	60	8.08
10	git	56	7.54

AI_Skill_Count	Number_of_Postings	Percentage_of_Postings
0	345	46.43
1	315	42.40
2	77	10.36
3	6	0.81



# Next Steps

- Expand analysis to compare AI/ML skill demand by specific role, location, and/or company
- Improve skill extraction with more advanced NLP methods
- Refine skill dictionary to capture emerging technologies/methods



# References

D. A. Southwick, Z. V. Liu, C. Baldwin, A. L. Quirk, L. H. Ungar, C.-J. Tsay, and A. L. Duckworth, “The trouble with talent: Semantic ambiguity in the workplace,” *Organizational Behavior and Human Decision Processes*, vol. 174, Art. 104223, 2023, doi: 10.1016/j.obhdp.2022.104223.

J. Brasse, M. Förster, P. Hünn, J. Klier, M. Klier, and L. Moestue, “Preparing for the future of work: A novel data-driven approach for the identification of future skills,” *Journal of Business Economics*, vol. 94, no. 3, pp. 467–500, 2023, doi: 10.1007/s11573-023-01169-1.

Nalisha, *Data Science Careers & Salaries 2025*, Kaggle, 2025. [Online]. Available: <https://www.kaggle.com/datasets/nalisha/data-science-careers-and-salaries-2025> (accessed Feb. 4, 2026).

OpenAI, *ChatGPT*, [Online]. Available: <https://chat.openai.com/> (accessed Feb. 2, 2026).

U.S. Bureau of Labor Statistics, U.S. Department of Labor, “Data Scientists,” Occupational Outlook Handbook, [Online]. Available: <https://www.bls.gov/ooh/math/data-scientists.htm> [Accessed: Jan. 30, 2026].

<https://github.com/yxs4yt/MRR--Project-1>



### Distribution of ML/AI Skill Ratio in Data Science Job Postings

