Project 4: AI vs Non-AI Careers

About The Team

Connor Coulter

Wei Wang

Balqis Bevi Abdul Hannan Kanaga

October 8, 2025

**Team:** Connor Coulter, Wei Wang, Balqis Bevi Abdul Hannan Kanaga

**Topic:** AI vs. Non-AI Job Growth — Is AI taking over or creating more jobs?

**Course:** AD 688 – Cloud Analytics for Business

This site hosts our research rationale, intro, and literature review for Project Selection I, II and III.

title: “Data Cleaning & Exploration” subtitle: “Preparing and Preprocessing the Job Market Dataset” author: - name: Connor Coulter affiliations: - id: bu name: Boston University city: Boston state: MA - name: Wei Wang affiliations: - ref: bu - name: Balqis Bevi Abdul Hannan Kanaga affiliations: - ref: bu bibliography: references.bib csl: csl/econometrica.csl format: html: embed-resources: true toc: true number-sections: true df-print: paged docx: default execute: echo: false eval: true freeze: auto jupyter: python3 —

# Import Data

Loaded dataset: (72498, 131)

# Data Cleaning & Preprocessing

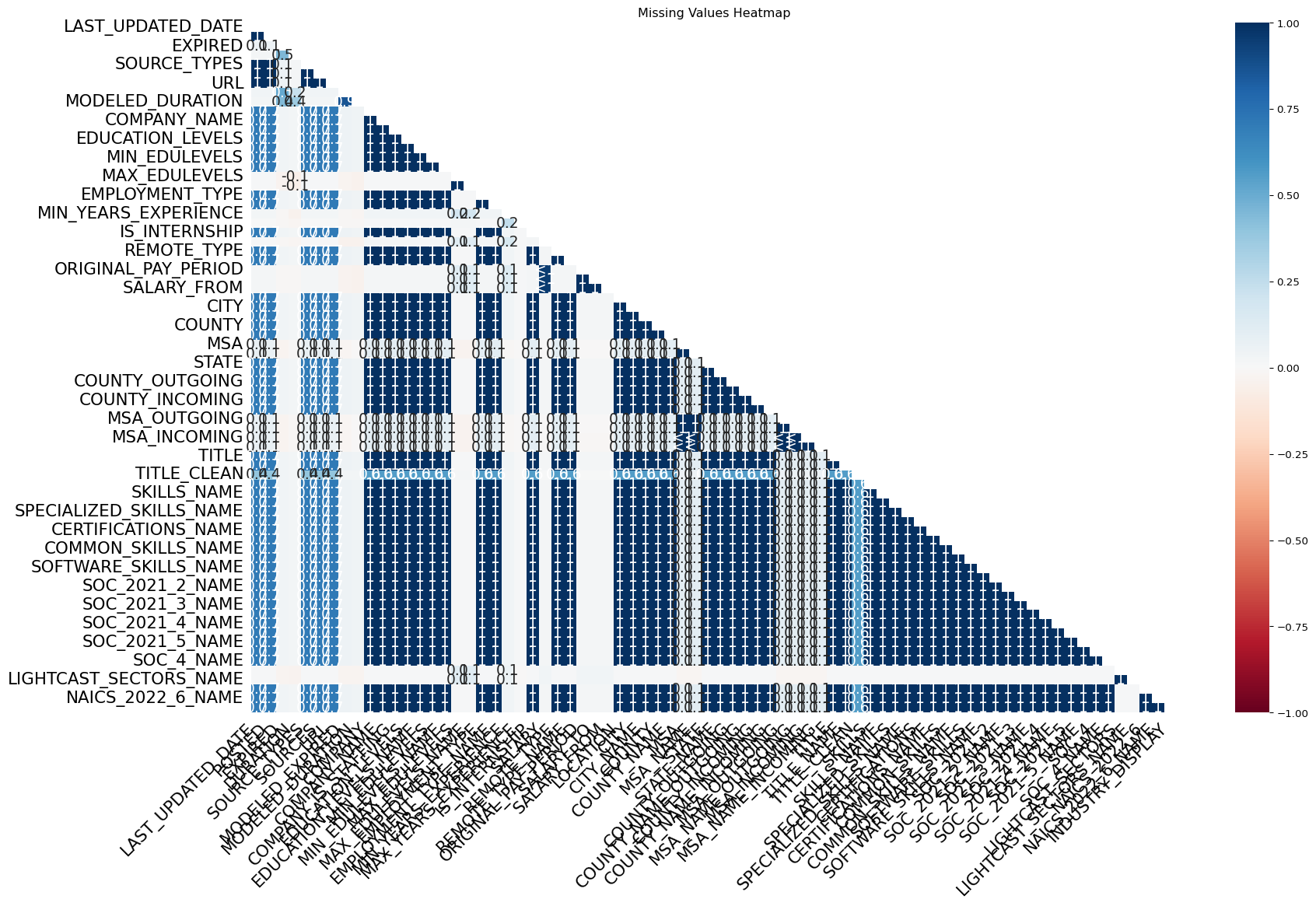
## Drop Unnecessary Columns

Derived non-null: {'INDUSTRY\_DISPLAY': np.int64(72454), 'SALARY\_DISPLAY': np.int64(72498)}

## Drop Unnecessary Columns

Remaining columns (first 30): ['LAST\_UPDATED\_DATE', 'POSTED', 'EXPIRED', 'DURATION', 'SOURCE\_TYPES', 'SOURCES', 'URL', 'MODELED\_EXPIRED', 'MODELED\_DURATION', 'COMPANY', 'COMPANY\_NAME', 'COMPANY\_IS\_STAFFING', 'EDUCATION\_LEVELS', 'EDUCATION\_LEVELS\_NAME', 'MIN\_EDULEVELS', 'MIN\_EDULEVELS\_NAME', 'MAX\_EDULEVELS', 'MAX\_EDULEVELS\_NAME', 'EMPLOYMENT\_TYPE', 'EMPLOYMENT\_TYPE\_NAME', 'MIN\_YEARS\_EXPERIENCE', 'MAX\_YEARS\_EXPERIENCE', 'IS\_INTERNSHIP', 'SALARY', 'REMOTE\_TYPE', 'REMOTE\_TYPE\_NAME', 'ORIGINAL\_PAY\_PERIOD', 'SALARY\_TO', 'SALARY\_FROM', 'LOCATION']

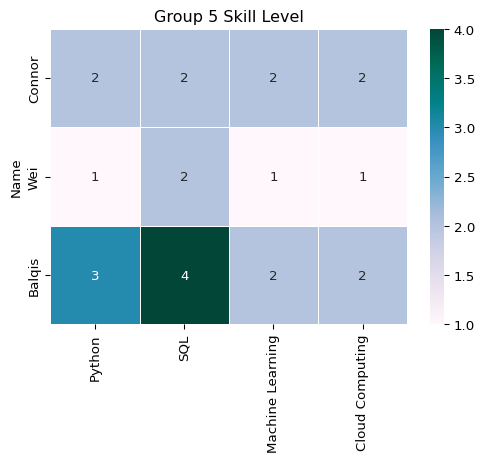
## Handle Missing Values



## Remove Duplicates

Removed 3300 duplicates using ['TITLE', 'COMPANY\_NAME', 'LOCATION', 'POSTED']

# Group 5 skill level



# Compare our group’s skills against job market demand

|  | Python | SQL | Machine Learning | Cloud Computing | Docker | AWS |
| --- | --- | --- | --- | --- | --- | --- |
| Name |  |  |  |  |  |  |
| Connor | 2 | 2 | 2 | 2 | 0 | 0 |
| Wei | 1 | 2 | 1 | 1 | 0 | 0 |
| Balqis | 3 | 4 | 2 | 2 | 0 | 0 |

# Improvement Plan

* **Balqis**: Her Machine Learning and Cloud Computing are at a basic level, leaving room to grow. With a career in data analysis and visualization, Machine Learning isn’t her top priority, but Cloud Computing is worth developing further. Strengthening Python would also be valuable, as it’s essential for data analysts. A good approach is to sharpen her skills through small personal projects and apply what she learns at work. If her fundamentals feel solid, she can move towards certifications.
* **Wei**: Her Python and Machine Learning are at a basic level, so she has the option to develop them further depending on how relevant they are to her career path. Since her SQL is already stronger, focusing on Python would be the most practical next step if she chooses to continue building technical skills. A good approach is to take it gradually through small projects and applied practice, and then expand into more advanced areas only if it fits her goals.
* **Connor**: His skills are fairly even across all areas, at a basic stage, which gives him room to build depth. Bumping Python up to a stronger level would give him the most flexibility, while also continuing to grow in Cloud Computing to keep pace with current tools and workflows. A steady way forward is to practice Python through hands-on work and then bring in cloud tools as he becomes more confident.

# Introduction

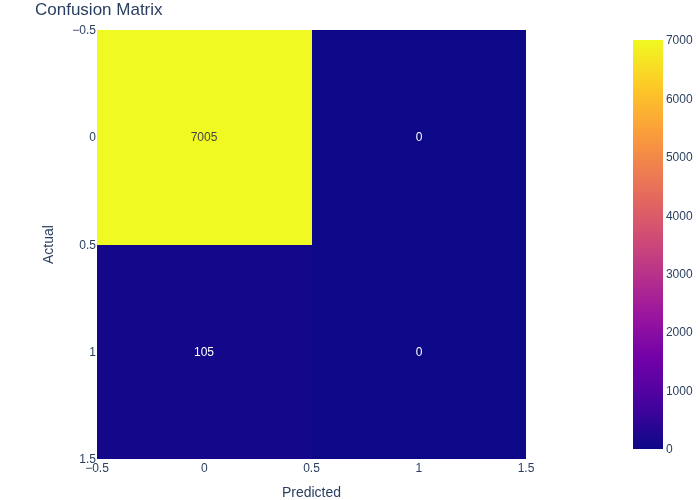
This project explores **AI vs Non-AI careers** using the lightcast\_job\_postings.csv dataset.  
We apply **clustering, regression, and classification** to evaluate trends in job markets, with a focus on **salary, experience, and employability**.  
The goal is to help job seekers understand how AI is shaping opportunities in 2024.

# Analysis

## Load dataset

Original dataset: 72498 rows  
After removing missing salary and years\_experience: 23697 rows

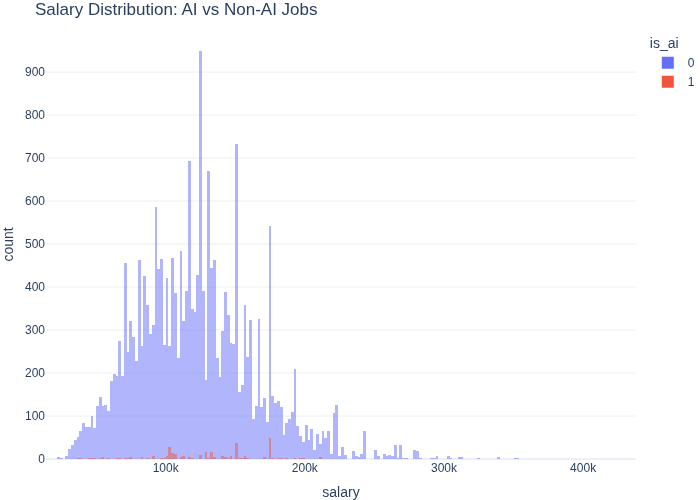
## Classification: AI vs Non-AI Jobs



## Clustering: Job Segmentation



## Visualizations



# Insights for Job Seekers

* **AI roles often cluster at higher salaries** compared to non-AI roles.
* **Experience remains critical** — higher years of experience align with higher pay clusters.
* **Industries with strong AI adoption** (e.g., tech, finance) show clearer salary advantages.

## Takeaways:

* Highlight **AI-related skills** to access higher-paying roles.
* Leverage **industry trends** to target fields with high AI adoption.
* Use clustering insights to understand **where your profile fits** (AI-heavy vs. traditional roles).