# JOB MARKET ANALYSIS

*By: Ftalew Dagnaw, GRAPH Student*
Date: October 30, 2025
As part of the GRAPH WORKSHOP 3 WORKSHOP 3 | Analyzing Job Market Data with LLM APIs | AIW 2025 Q4

Brief Introduction

This dataset contains 400 job postings for data analyst positions, including information on median salary estimates, minimum years of experience required, and programming language

requirements (R, Python, both, or neither). This project analyzes how experience level and programming language skills relate to salary expectations in the data analyst job market.
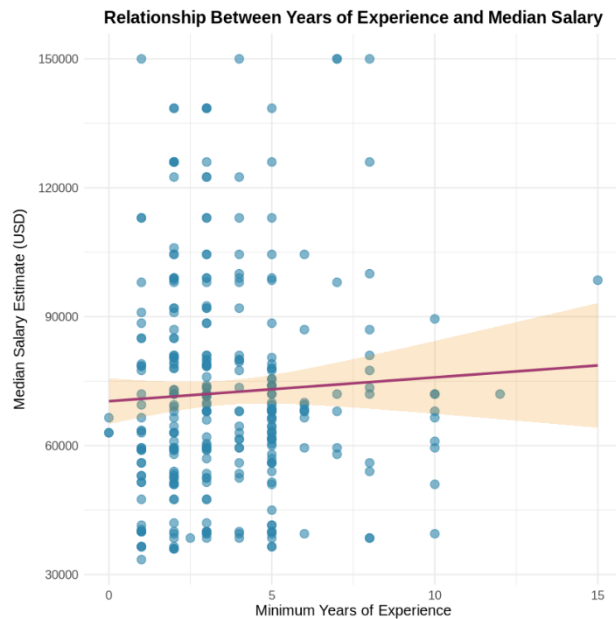
Interpretation: The scatter plot reveals a surprisingly weak relationship between years of experience and median salary (correlation = 0.05). While we might expect salaries to increase substantially with experience, the data shows considerable salary variation at every experience level. This suggests that factors other than years of experience—such as industry, location, company size, or specific technical skills—may play a more significant role in determining data analyst salaries. The relatively flat trend line indicates that entry-level positions and those requiring 10+ years of experience have similar salary ranges, which is noteworthy.

Median Salary Distribution by Programming Language
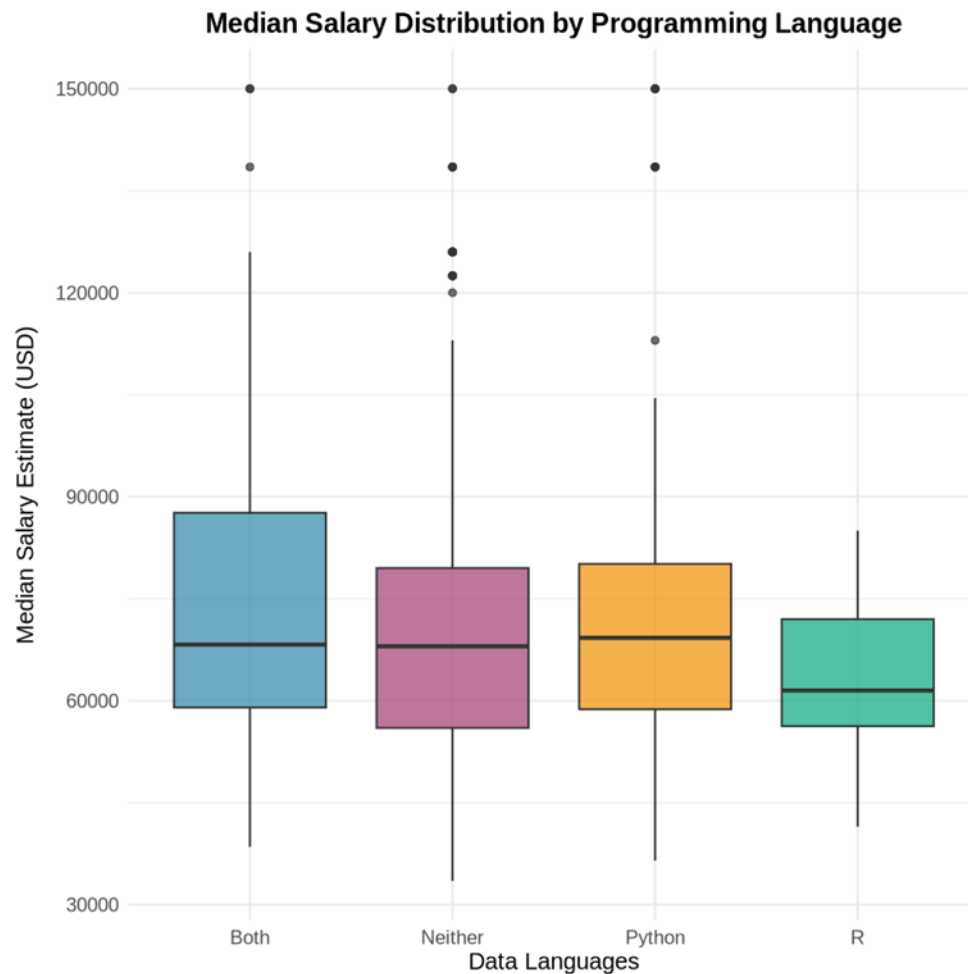
Interpretation: The box plot shows meaningful differences in salary distributions across programming language categories. Jobs requiring both R and Python have the highest mean salary ($75,471), followed closely by Python-only positions ($75,100). Interestingly, positions requiring neither programming language still command respectable salaries (mean: $70,433), while R-only positions show the lowest mean salary ($64,533) with the least variability. This suggests that Python skills, either alone or combined with R, are associated with higher-paying data analyst roles, while R-only positions may be more specialized or niche'

Reflection

I found the session very interesting and helpful in exploring how we can interact with LLM in one platform and conduct text analysis from large dataset in a couple of minutes that it would have been taken longer duration otherwise. I found the section straight forward in following the steps in the analysis. The challenging part is some how in choosing the right model and maybe in our case easily securing access in keys. I found this interesting how LLM is used to conduct job market data analysis and how much it gives huge opportunity to take decision making in both the government policy setting and academic institutions capacity building efforts..