Analyzing Job Market Data with LLMs in Sheets and AI Reporting Agents
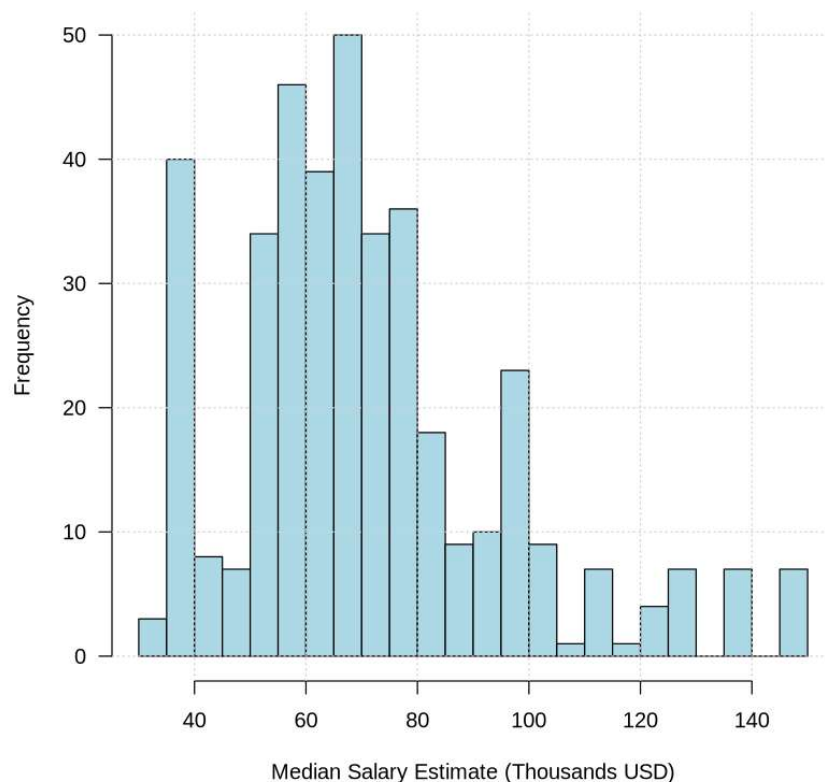
Kim Lindblade

In 2020, more than 8000 job postings in GlassDoor for data analysts, data scientists and data engineers were scraped into an online database. In this report, we analyze the association between the minimum years of experience and data language skills (R or Python) required for the position against the median salary estimates among 400 positions included in this assignment.

1. Salary ranges

The annual salary data ranged from $33,500 to $150,000, with a median of $68,000 and a mean of approximately $71,794 (Figure 1). The distribution of median salary was right skewed with a few jobs offering more than $120,000. No records were missing salary data.
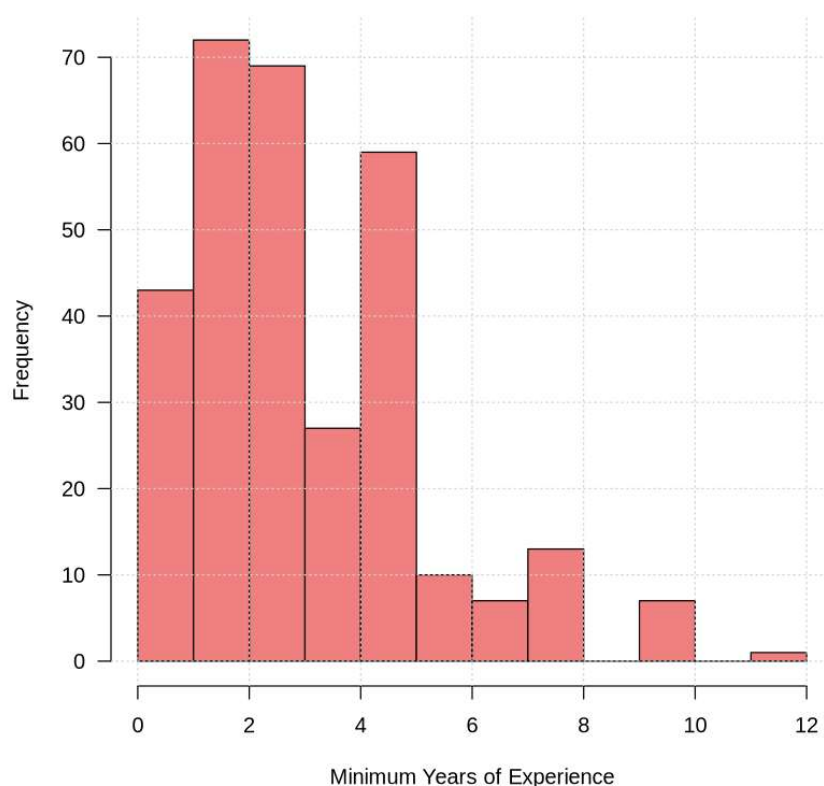

Figure 1: Median Salary Estimates

2. Minimum years experience

Many records (92, 23%) were missing information on minimum years of experience required. The experience requirements ranged from 0 to 12 years with a median of 3 and median of 3.5 (Figure 2).
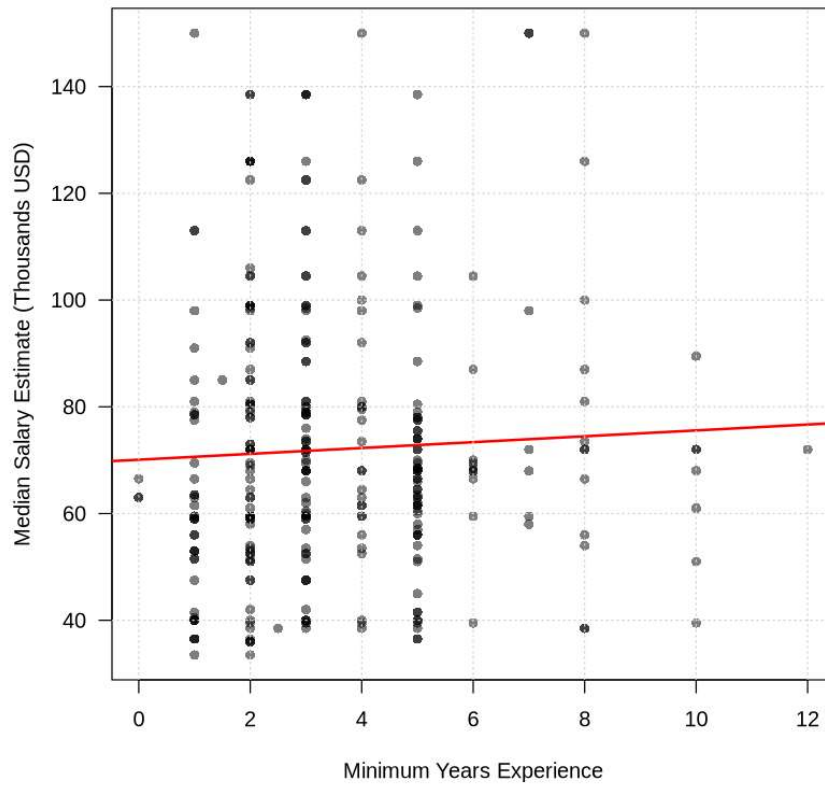
Figure 2: Minimum Years of Experience Required

3. Median salary estimates by minimum years of experience

The median salary estimate increased slightly with the minimum number of years experience required for the position (Figure 3). However, the 95% confidence interval for the Pearson correlation coefficient (r = 0.046, 95% CI, -0.067 - 0.157) suggested no statistically significant association. It is probable that other factors, including the location and level of the position, are stronger predictors of salary than years of experience. Additionally, the small sample size, which was further reduced by the number of records without minimum years of experience, may mask differences in salary by years of experience.

Figure 3: Relationship Between Experience and Salary
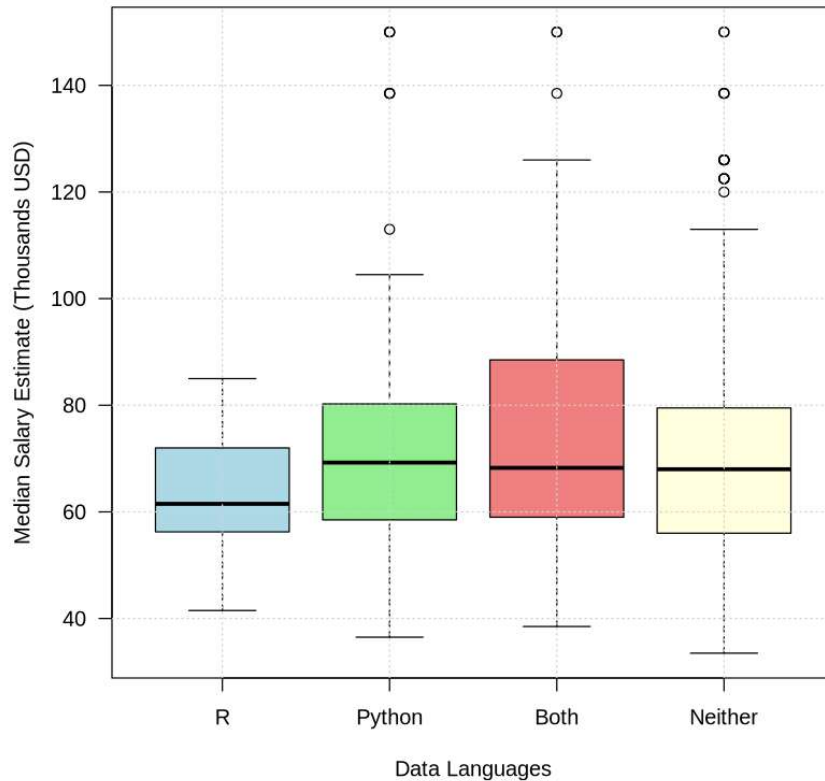
4. Programming language analysis

Among the 400 positions queried on data language, 3 returned an error. Most positions (252, 63.5%) did not mention a programming language requirement while 70 (17.6%) required both R and Python. Very few job postings mentioned only R (15, 3.8%).

Data Language	Number of Jobs	Percentage
R	15	3.8%
Python	60	15.1%
Both	70	17.6%
Neither	252	63.5%
Total	**397**	**100.0%**

5. Median salary estimates by programming language requirements

There was no strong association between the data language proficiency required for the position and median salary amounts (Figure 4).

Figure 4: Salary Distribution by Data Language Proficiency

6. Reflection

In this exercise, we learned how to use an application programming interface (API) to a large language model (LLM) embedded in a function in Google sheets to perform repeated actions. These actions could include internet searches or performing functions within a spreadsheet. The API provides a different interface for the LLM than the graphical user interface we were more accustomed to use and would be helpful when there are a number of similar actions to be conducted. I would be very interested in learning how to scrape data from websites. Given the current employment environment, being able to analyze job requirements in real time could be very helpful.