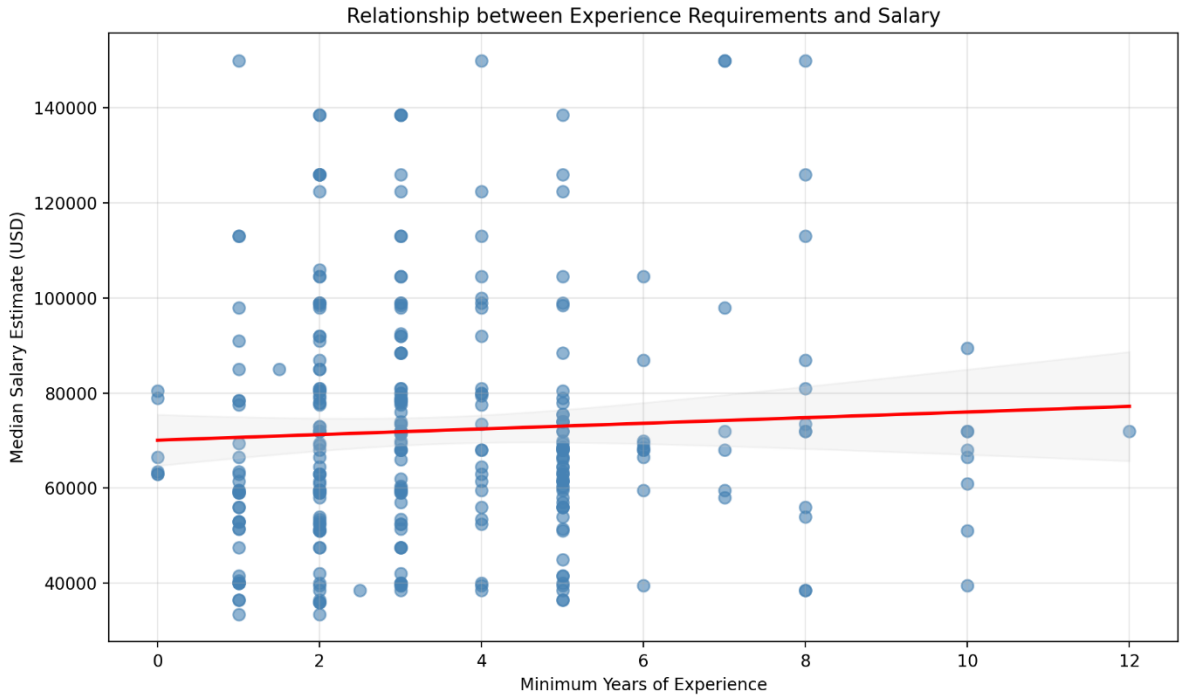
Analysis of the Data Analyst Job Market

Author Ester Nepolo – Generative AI for Work and Research
25 May 2025

1. Introduction

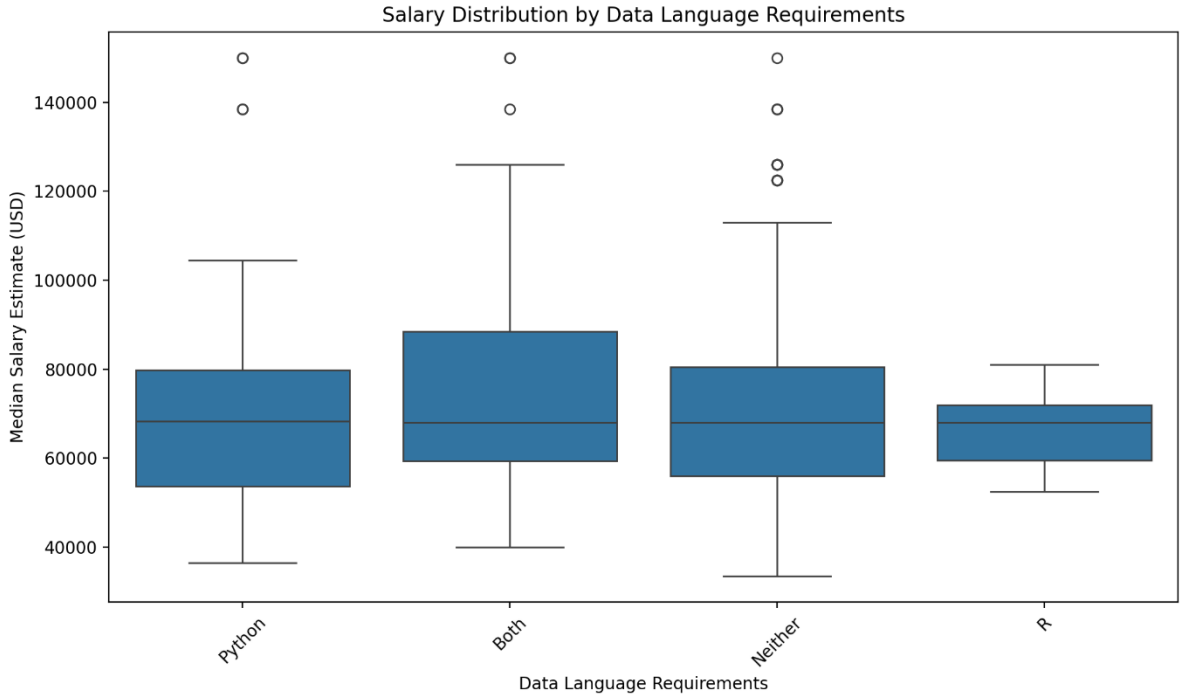
This exercise aims to extract specific information from hundreds of data analysis job postings for Data Analysts. Using LLMs and Google sheets, information about the number of years of required experience and data languages requirements were extracted and analyzed into the following graphs.

2. Experience vs. Salary Relationship



The plot shows a weak positive correlation between minimum years of experience required for job postings and the median salary offered for those positions, indicating that jobs requiring more experience tend to offer slightly higher salaries. However, the confidence interval indicates substantial variation in the data, suggesting that experience requirements alone are not a strong predictor of salary levels in data analyst positions."

3. Salary Distribution by Programming Language



The box plot reveals that data language requirements have minimal impact on salary levels, with all categories showing nearly identical median salaries around $68,000. Surprisingly, jobs requiring both R and Python skills do not offer higher median salaries compared to other categories - in fact, Python-only positions show the highest median salary at $68,250. This suggests that, proficiency in multiple data languages does not translate to significantly higher compensation, and other factors likely play more important roles in determining salary levels.

4. Reflections

It was surprising to see the model extracting information from the dataset effortlessly. Clear instruction makes the interaction easy and focused. There were a number of errors with dragging of the formula in the google sheets that I managed to spot and rectify manually. It will be interesting to know how such errors can be handled within big or large data sets.