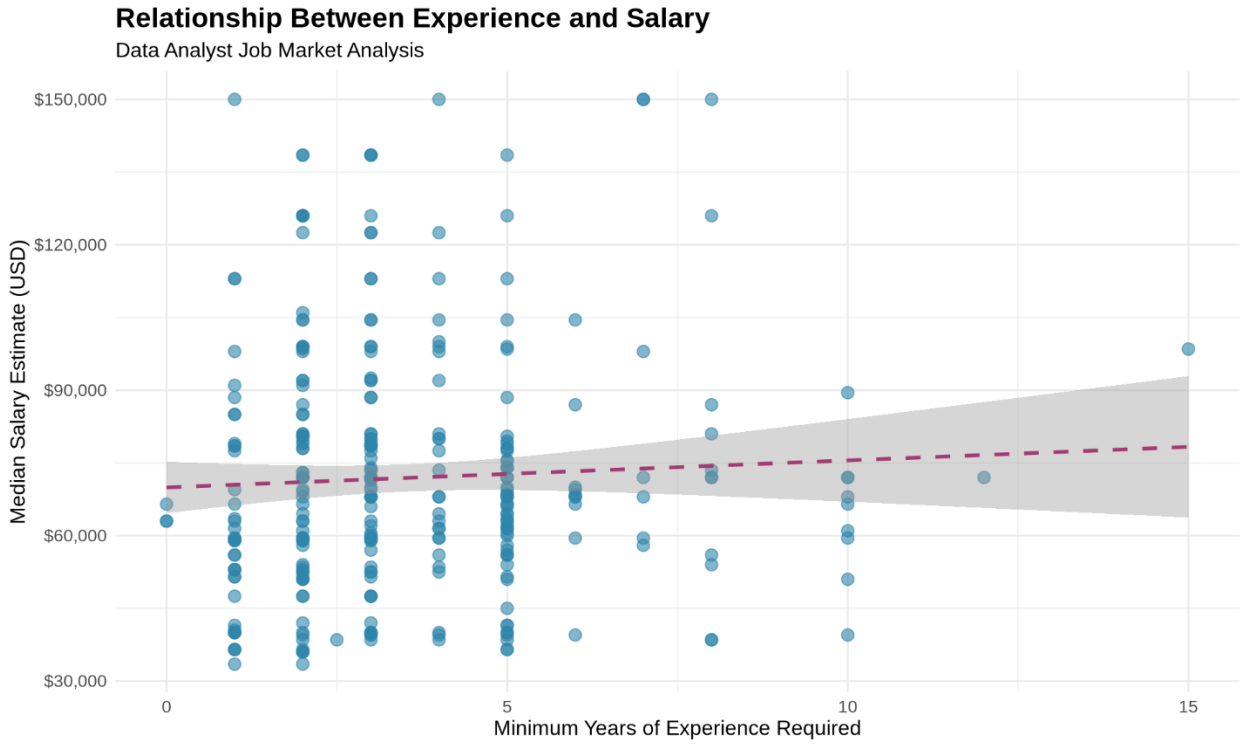
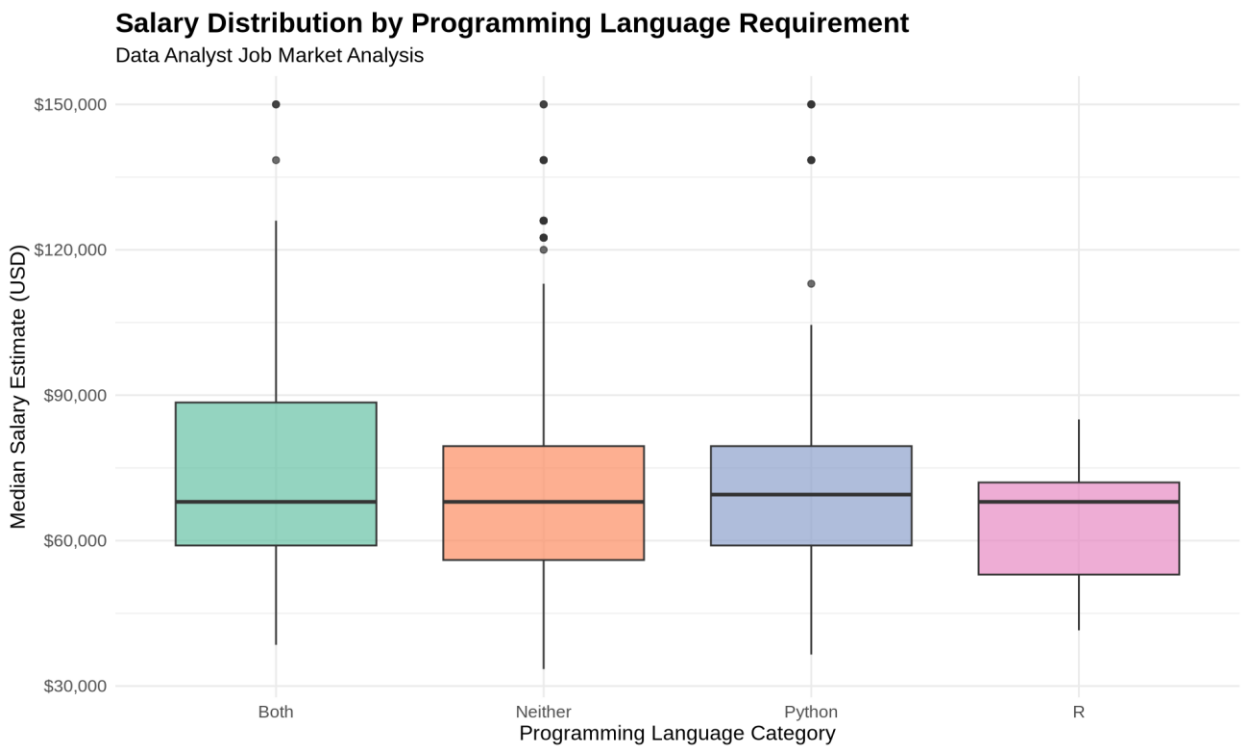
Data Analyst Job Market Analysis Report - Ignacio Esteban

1) **Brief introduction:** This report analyzes a dataset of 400 data analyst job postings to explore the relationship between experience requirements, programming language skills, and salary expectations. The analysis focuses on understanding how minimum years of experience and programming language proficiency (Python, R, both, or neither) influence median salary estimates in the data analyst job market.
2) **Analysis of "Years of Experience" :**



Interpretation: the scatter plot reveals a **surprisingly weak relationship** between years of experience and salary (correlation coefficient: 0.05). While one might expect higher experience requirements to correlate with higher salaries, the data shows considerable salary variation at every experience level. This suggests that factors beyond experience—such as industry, location, company size, or specific technical skills—play a more significant role in determining data analyst salaries. The nearly flat trend line indicates that entry-level positions (0-2 years) can offer salaries comparable to those requiring 5-10 years of experience, highlighting the diverse nature of data analyst roles across different sectors.

3) **Analysis of "Programming Language Requirements" :**



Summary Statistics:

Language Category	Mean Salary	Median Salary
Both (Python & R)	$75,458	$68,000
Python Only	$75,127	$69,500
Neither	$70,356	$68,000
R Only	$65,077	$68,000

Interpretation: The box plot demonstrates that **programming language requirements have a modest impact on salary distributions**. Jobs requiring both Python and R skills show the highest mean salary ($75,458), followed closely by Python-only positions ($75,127). Interestingly, the majority of positions (257 out of 400) require neither Python nor R, suggesting many data analyst roles focus on other tools like SQL, Excel, or business intelligence platforms. The similar median salaries across all categories ($68,000-$69,500) indicate that while programming skills may provide a slight salary advantage, they are not the primary salary determinant. The relatively small sample size for R-only positions (13 jobs) makes it difficult to draw strong conclusions about this category.

4) **Reflections:**

It was pretty straightforward, with very well-explained instructions with feasible objectives. I'm happy to be using Julius again to solidify what we saw last week.
The most surprising part was the results, more specifically, the lack of a strong positive correlation between the variables of interest. Also that most of the positions didn't require either R or Python. However, results and data interpretations are limited by the small sample size and missing values in the experience field (n=90, 22.5% of the dataset).
While we had used Julius before, it was my first time and probably not the last using my GPT function in Google Sheets, which seems to be incredibly helpful, time-saving, and interesting.