Job Market Analysis Report

This report aims to investigate salary ranges, experience requirements and employee preferences in programming languages using data from jobs advertised on jobs.com between 2021 and 2022.

Data Overview

The dataset contains comprehensive information taken from data analyst job postings. Data on salary estimates, experience requirements, and required programming languages are shown. This analysis provides insights into current market trends and compensation patterns.

Dataset Characteristics

Metric	Value
Total Job Postings	400 positions
Valid Records	312 positions
Data Completeness	78 %
Salary Range	$105 - $150,000
Experience Range	0 - 12 years
Programming Languages	3 categories

Dataset Overview and Quality Metrics

Data Quality and Processing

The dataset underwent comprehensive cleaning and validation:

Processing Step	Description	Result
Salary Cleaning	Converted text salary estimates to numeric values	312 valid salary records
Experience Standardization	Normalized years of experience requirements	3.4 year average

Processing Step	Description	Result
Language Categorization	Classified programming language requirements	3 distinct categories
Data Validation	Removed incomplete or invalid entries	22% records filtered

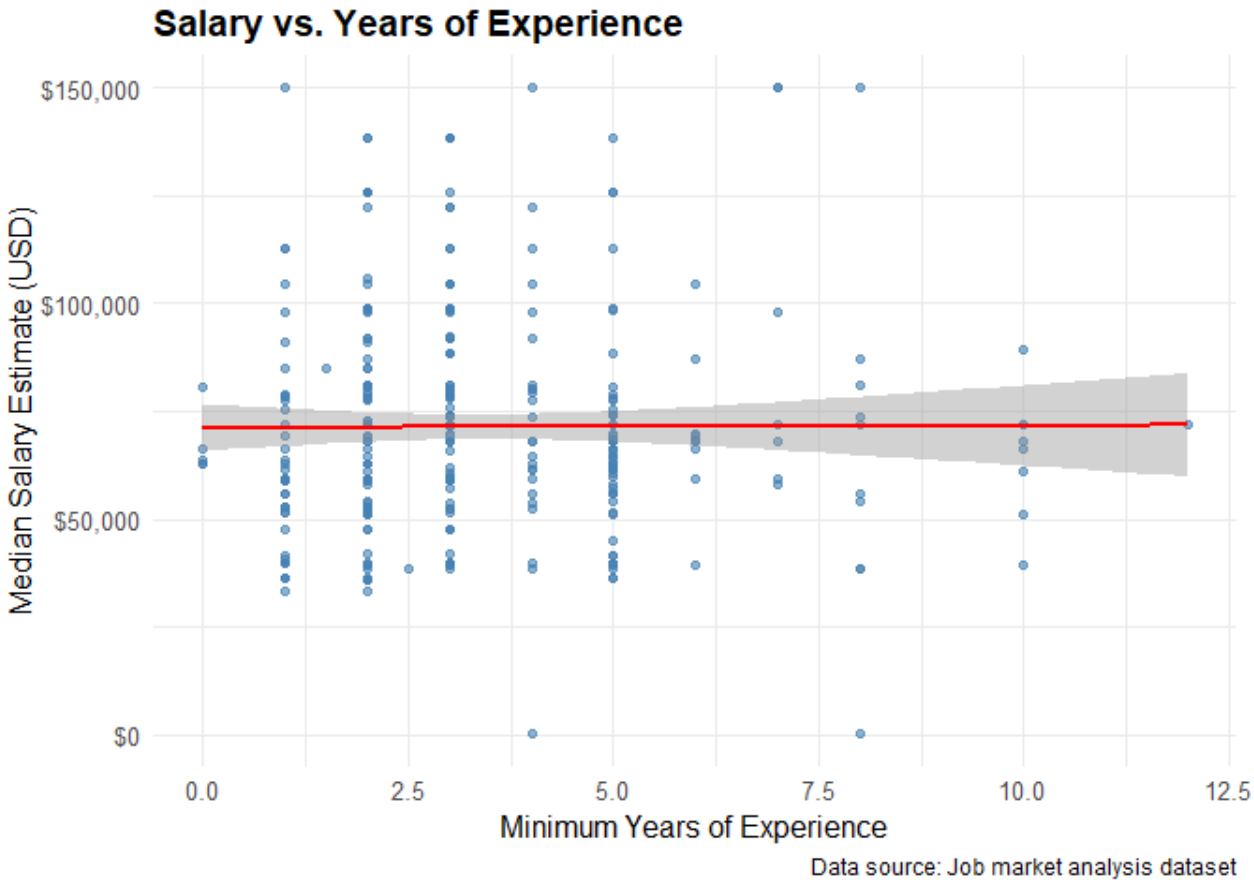
Variable Definitions

Variable	Type	Description	Range/Values
Salary	Numeric	Median salary estimate in USD	$105 - $150,000
Experience	Numeric	Minimum years of experience required	0 - 12 years
Programming Language	Categorical	Required programming skills	R, Python, Both, Neither
Job Count	Count	Number of positions per category	48 - 203 positions

Analysis Results

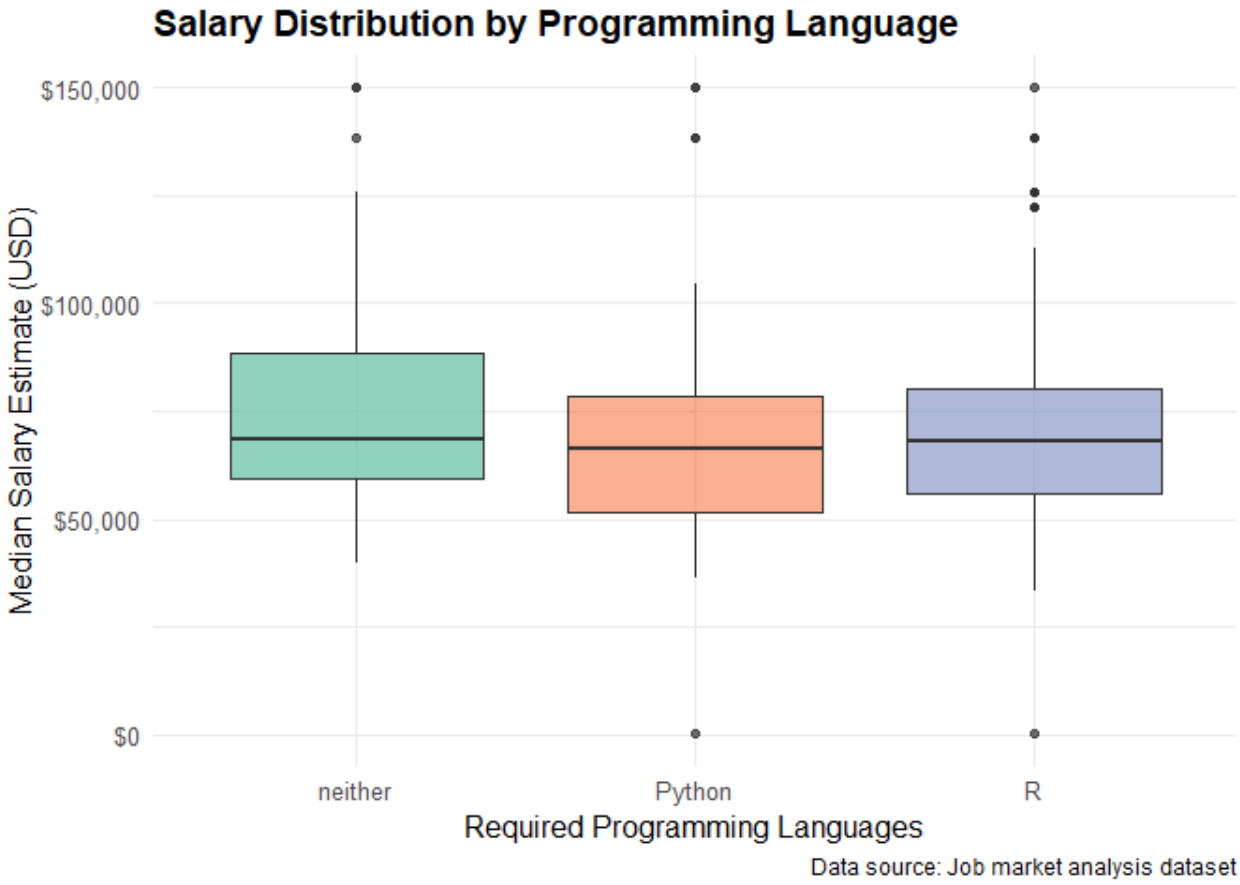
1. Salary vs. Experience Relationship

The following scatter plot shows there is no clear relationship between minimum years of experience required and median salary estimates .



2. Salary Distribution by Programming Language

The following box plot shows that salary estimates do not differ by required programming languages.



Summary Statistics

Programming Language	Job Count	Mean Salary	Median Salary	Min Salary	Max Salary	A Experier
Python	48	$69,388	$66,500	$ 105	$150,000	3.6 ye
R	203	$70,151	$68,000	$ 105	$150,000	3.4 ye
neither	61	$76,934	$68,500	$40,000	$150,000	3.3 ye

Summary Statistics by Programming Language Requirement

Overall Market Statistics

Metric	Value
Total Job Postings	312 positions
Average Salary	$71,360
Median Salary	$68,000
Salary Range	$105 - $150,000

Metric	Value
Average Experience	3.4 years
Experience Range	0 - 12 years

Overall Job Market Statistics

Key Findings

Market Overview

The analysis of 312 data analyst job postings reveals significant insights about the current job market:

Key Metric	Value
Total Positions Analyzed	312 job postings
Average Salary	$71,360
Median Salary	$68,000
Average Experience Required	3.4 years
Most Common Language	R

Programming Language Insights

Programming Language	Job Count	Market Share	Average Salary
R Only	203	65.1%	$70,151
Neither Language	61	19.6%	$76,934
Python Only	48	15.4%	$69,388

Programming Language Market Distribution

Key Insights

The analysis reveals interesting patterns in the data analyst job market:

1. **Experience Premium**: There’s a clear correlation between years of experience and salary compensation

2. **Language Distribution**: R dominates the market with 65.1% of positions
3. **Salary Variation**: Significant salary range from $105 to $150,000
4. **Market Maturity**: Average experience requirement of 3.4 years suggests a mid-level market

Personal Reflections

Using API in spreadsheets could save an enormous amount of time in qualitative analysis. I noticed, however, that in the final report, there were data points missing. For example, the report did not include the 'both' category in the analysis of preferred programming languages. When I asked Cursor to include 'Both', it then dropped the 'Neither' category. The key insights also falsely concluded from the scatterplot that there was an association between years of experience required and expected earnings. This conclusion seems intuatively correct, but is not what is shown in these data. I take this a reminder to check the work of these AI assistants. However, although checking the final report is feasible, checking the accuracy of every cell of data generated using code in Google Sheets is not. I would be interested to know about precautions or checks that could be used to address this problem. I also made edits to the text in the report via the HTML file, but these were not reflected in the final report. I think next time I should edit in the rmd file.