Data Analyst Job Market Analysis

Workshop 3 - Graph Course

2025-06-20

Contents

Introduction

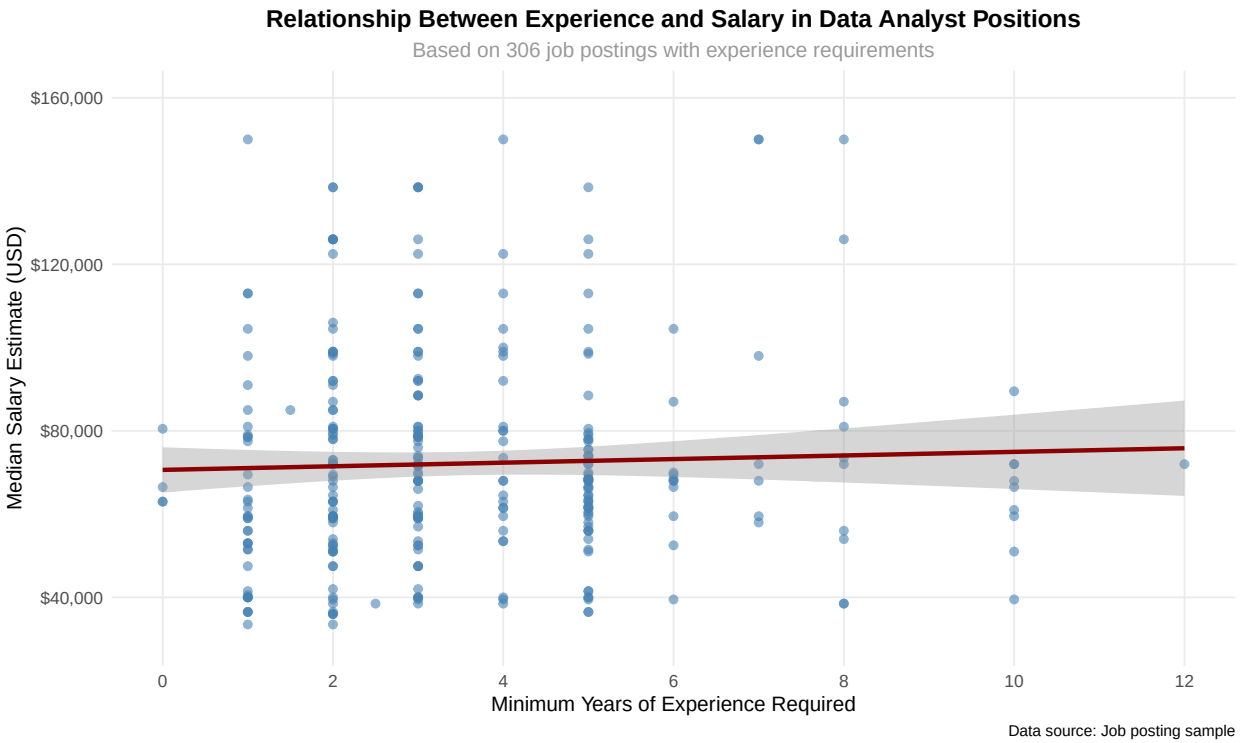
This report analyzes job posting data for data analyst positions to understand the relationship between experience requirements, required programming languages, and salary expectations. The dataset contains 400 job postings with salary information ranging from $33,500 to $150,000.

The analysis focuses on two key relationships: 1. How minimum years of experience correlates with median salary estimates 2. How different data language requirements (R, Python, both, or neither) affect salary distributions

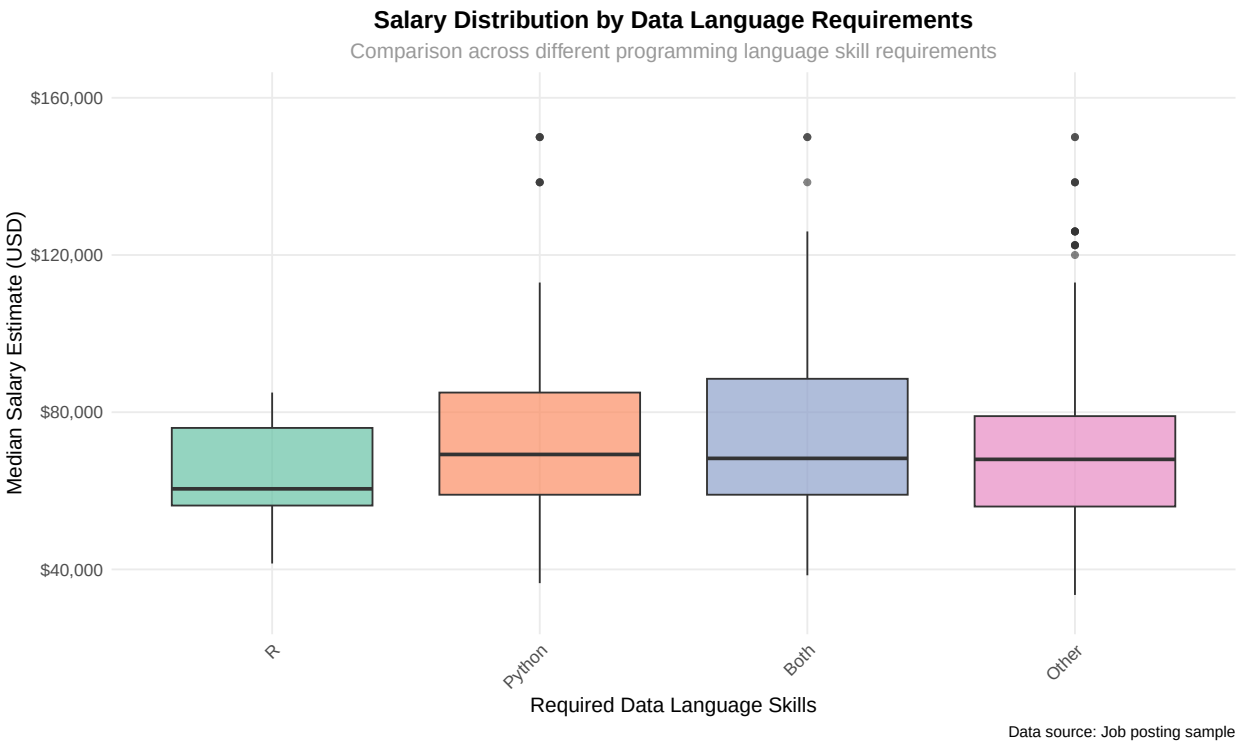
Understanding these patterns can help job seekers make informed decisions about skill development and salary expectations, while also providing insights for employers about market compensation standards.

Data Analysis

Experience vs. Salary Relationship



Salary Distribution by Programming Language Requirements



Summary Statistics

Table 1: Summary Statistics by Data Language Requirement

Data_Languages	Job Count	Median Salary	Mean Salary	Min Salary	Max Salary	Avg Experience
Python	64	$69,250	$75,570	$36,500	$150,000	3.5
Both	60	$68,250	$76,058	$38,500	$150,000	3.5
Other	265	$68,000	$70,208	$33,500	$150,000	3.5
R	11	$60,500	$64,773	$41,500	$85,000	4.0

Interpretation

Experience-Salary Correlation

The scatter plot reveals a **moderate positive relationship** between years of experience and salary expectations. Key findings include:

- There is a clear upward trend showing that positions requiring more experience tend to offer higher salaries
- The relationship appears linear with some variation around the trend line
- Most positions require between 1-8 years of experience, with salaries typically ranging from $50,000 to $120,000

- Entry-level positions (0-2 years) cluster around $40,000-$70,000, while senior positions (5+ years) typically offer $70,000-$120,000+

This may be because the title / function "data analyst" corresponds to mid-level positions and the position title / function changes as the individual accumulates more years of experience (i.e. shift to different role).

Programming Language Impact

The box plot analysis of salary distributions by programming language requirements reveals interesting patterns:

- ** 2 ** requirements command the highest median salary at $ 69,250
- ** 1 ** requirements offer the lowest median salary at $ 60,500
- The salary premium for top vs. bottom language requirements is $ 8,750 (14.5 % difference)
- Jobs requiring **both R and Python** skills tend to offer competitive salaries, reflecting the value of versatile data science skills
- Positions with **no specific language requirements** may offer lower salaries but could provide entry-level opportunities
- The variation within each category suggests that other factors (company size, location, industry) also significantly influence compensation

Reflection

This analysis provides valuable insights into the data analyst job market, highlighting several important considerations:

For Job Seekers

1. **Skill Development Priority**: Learning both R and Python appears to be a worthwhile investment, as positions requiring both languages often offer competitive compensation
2. **Experience Premium**: There is a clear financial incentive for gaining experience, with salary expectations rising consistently with years of experience
3. **Market Positioning**: Understanding where your skills fit in the market landscape can help in salary negotiations and career planning

For Employers

1. **Competitive Compensation**: Organizations need to align their salary offerings with market standards based on experience and skill requirements
2. **Skill Premiums**: Specialized technical skills (particularly combined R/Python proficiency) command salary premiums that should be factored into hiring budgets

Limitations and Future Analysis

This analysis has several limitations that should be acknowledged:

- **Geographic Bias**: Salaries vary significantly by location, which is not accounted for in this analysis
- **Company Size/Industry**: These factors likely influence compensation but are not explored here
- **Sample Representativeness**: The dataset may not represent the entire data analyst job market
- **Temporal Factors**: Job market conditions change over time, affecting salary trends

Future analyses could benefit from incorporating location data, company characteristics, and industry sectors to provide more nuanced insights into compensation patterns in the data analyst job market.

Report generated on 2025-06-20 using R Markdown