Data Analysis Report

Tech Job Market Analysis: PostgreSQL and Python Integration

Omar Ayman Tawfiq Bakr

Abstract—This project analyzes 787,686 tech job postings using PostgreSQL and Python, revealing critical insights about compensation, skills demand, and hiring patterns. The analysis uncovers that data-related positions dominate the market (73%), remote opportunities remain limited (9%), and significant compensation gaps exist between specialized roles. Implementation demonstrates advanced SQL techniques including CTEs, window functions, and complex joins, integrated with Python visualizations.

Keywords—*PostgreSQL, job market analysis, SQL, Python, data visualization, tech industry*

1. Introduction

This analysis examines 787,686 tech job postings collected in 2023 from multiple platforms including LinkedIn and BeBee. The dataset spans global markets with strong representation from the US, UK, France, Germany, and India, providing comprehensive insights into current tech employment trends.

The project demonstrates PostgreSQL proficiency through complex analytical queries while uncovering actionable insights about salary distributions, skill requirements, and market dynamics.

2. Database Architecture

The database follows a dimensional model with a central fact table (`job_postings_fact`) connected to dimension tables for companies, skills, and geographic data. This structure enables efficient analytical queries across multiple dimensions.

Key tables include:

- `job_postings_fact`: 787,686 job records
- `company_dim`: Company information
- `skills_job_dim`: Skill-job relationships
- `skills_dim`: Skill definitions

3. Market Overview

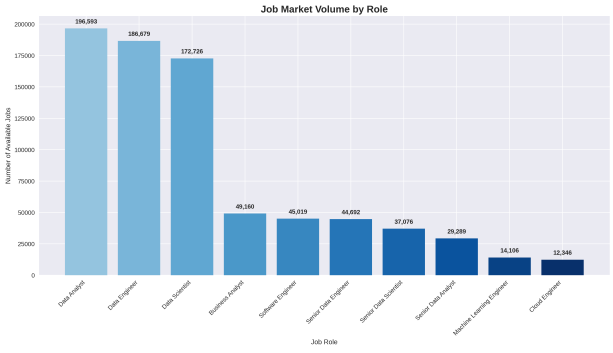

Figure 1. Job distribution showing data analyst dominance at 196,593 positions

The market analysis reveals striking concentration in data-related roles:

- Data Analysts: 196,593 jobs (25%)
- Data Engineers: 113,220 jobs (14%)
- Data Scientists: 106,563 jobs (13.5%)
- Software Engineers: 99,077 jobs (12.6%)
- Machine Learning Engineers: 14,106 jobs (0.01%)

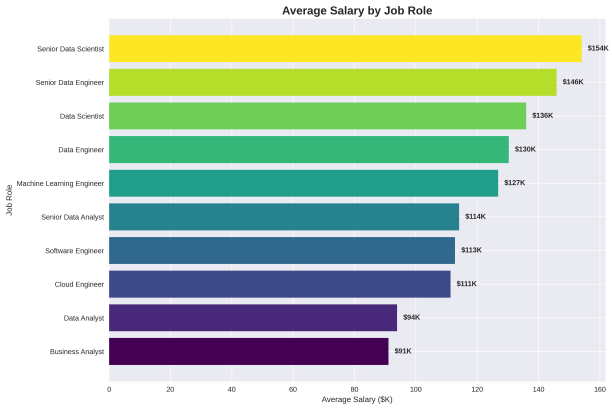

Figure 2. Average salaries by role showing premium for specialized positions

4. Compensation Analysis

Salary analysis reveals distinct compensation tiers:

Premium Tier ($140k+):

- Senior Data Scientists: $157
- Senior Data Engineers: $146

Standard Tier ($100-140k):

- Data Scientists: $136
- Data Engineers: $130
- ML Engineers: $127
- Data Analyst Engineers: $114
- Software Engineers: $113
- Cloud Engineers: $111

The 61% premium for senior data scientists over standard analysts highlights the value of specialization and experience.

5. Benefits and Requirements

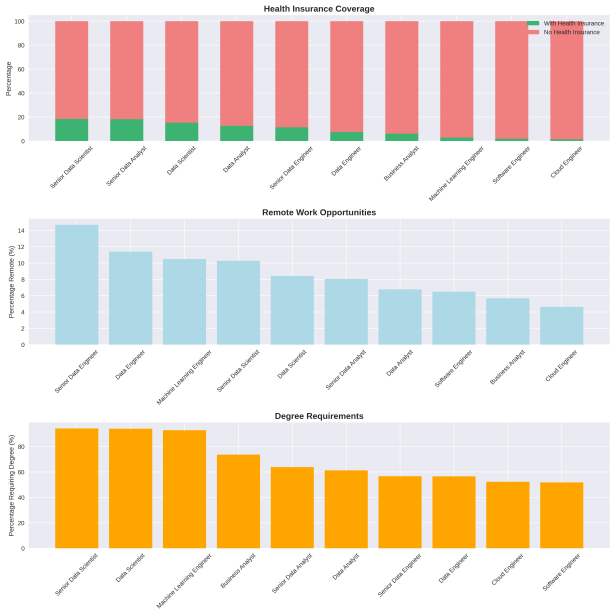

Figure 3. Benefits and requirements breakdown by position

Critical findings on benefits:

- Health insurance coverage: Only 11% overall

- Remote work: 9% of positions (highest in data analyst roles at 15%)
- Degree requirements: 64% overall (80%+ for data roles)

The low health insurance coverage (11%) likely reflects global data mixing US and international markets where healthcare systems differ.

6. Skills Demand

Top skills analysis reveals:

1. SQL (34% of jobs)
2. Python (31% of jobs)
3. Cloud platforms (AWS, Azure, GCP)
4. Tableau/Power BI for visualization
5. Spark for big data processing

The skill type distribution shows that while programming languages dominate overall, each role has distinct technical requirements, with data roles emphasizing analytical tools alongside programming capabilities.

7. Company Hiring Patterns

Hiring patterns reveal:

- Staffing agencies lead volume (Emprego, Robert Half)
- ML positions remain niche (14,000 jobs total)
- Data roles dominate across all company types

Note: ML company rankings may not reflect typical expectations due to the small sample size (14,000 positions) compared to data roles.

8. Key Insights

The analysis reveals several market realities:

1. Data Dominance: 73% of positions are data-related, confirming the shift toward data-driven decision making across industries.

2. Compensation Gaps: 50%+ salary premiums for senior and specialized roles justify investment in skill development.

3. Limited Remote Work: Despite post-pandemic expectations, only 9% of positions offer remote work, suggesting return-to-office trends.

4. Education Barriers: 80%+ degree requirements for data roles may limit talent pool accessibility.

5. Benefits Scarcity: Low health insurance offerings (11%) reflect global market mix and varying healthcare systems.

9. Technical Implementation

The project demonstrates advanced PostgreSQL techniques:

Complex Joins: Multi-table joins connecting jobs, companies, skills, and locations for comprehensive analysis.

CTEs (Common Table Expressions): Modular query construction for readability and performance optimization.

Aggregation Patterns: Strategic grouping for multi-dimensional analysis across time, geography, and job categories.

Python integration includes:

- **SQLTools plugin** for database connectivity
- pandas for data manipulation
- matplotlib/plotly for visualizations

10. Limitations and Future Work

Current limitations include:

- 2023 data snapshot may not reflect current conditions
- Salary data available for only 0.02% of postings
- Geographic bias toward English-speaking markets

Future enhancements could include:

- Time-series analysis for trend prediction
- Skills co-occurrence network analysis
- Salary prediction modeling

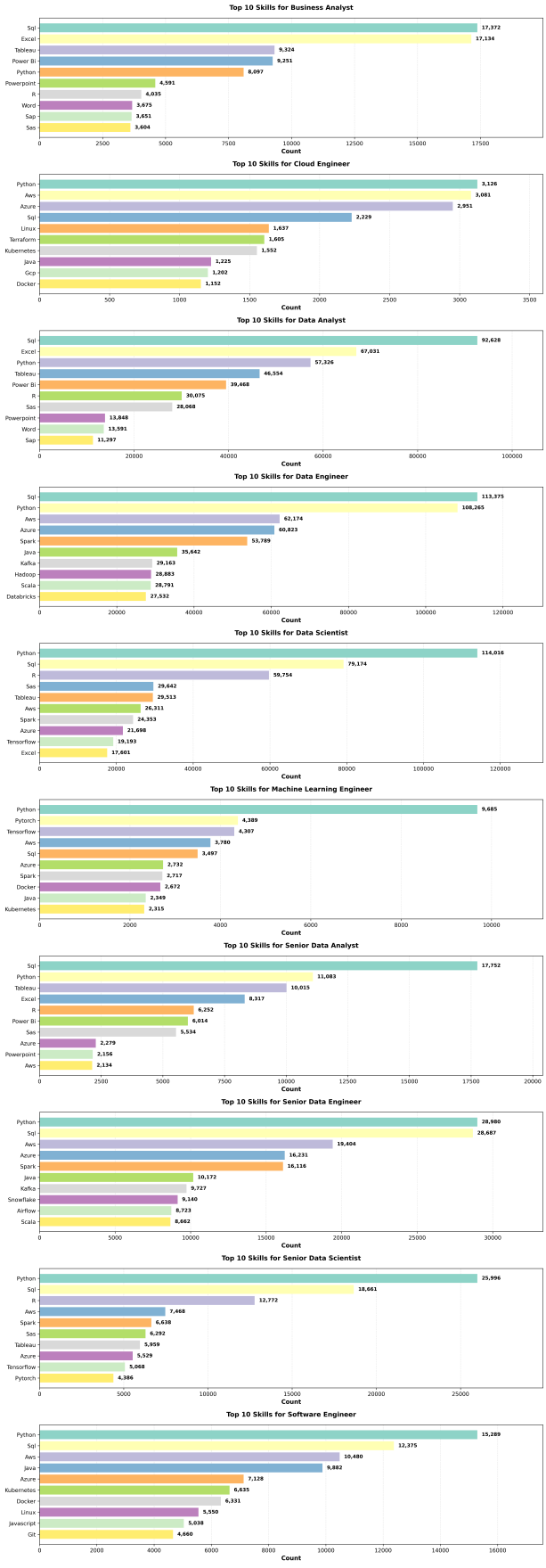

Figure 4. Most demanded technical skills across all positions

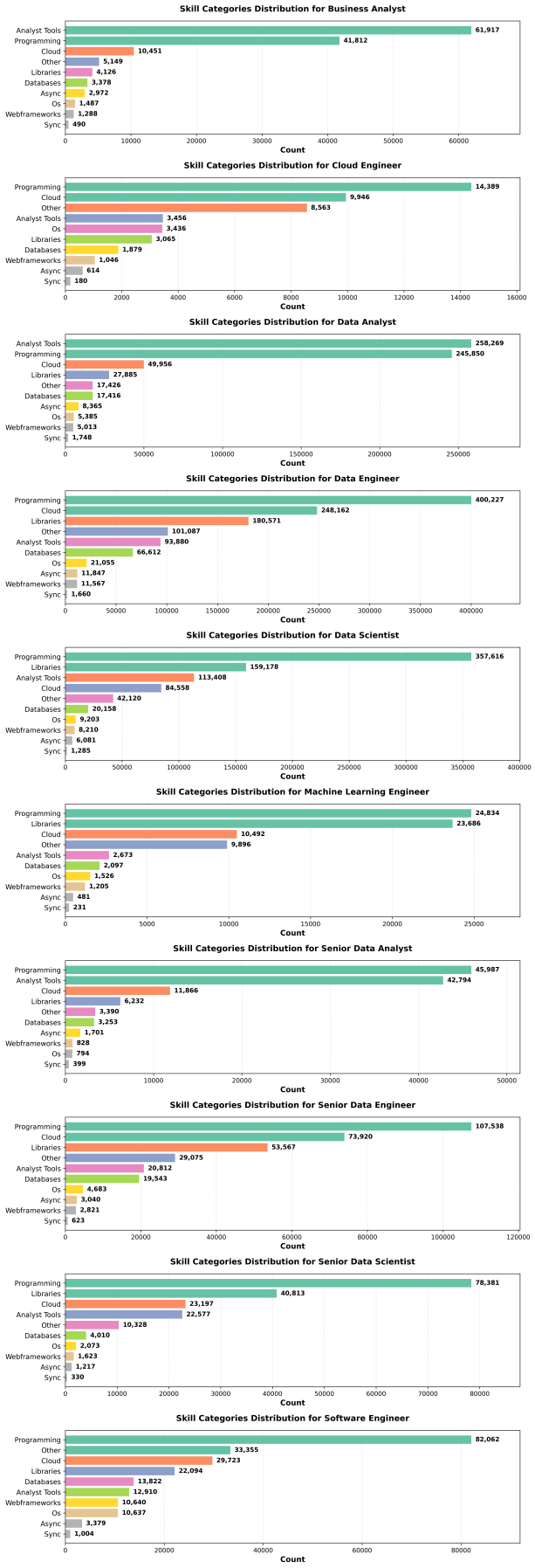

Figure 5. Skills distribution by role type showing varying technical requirements

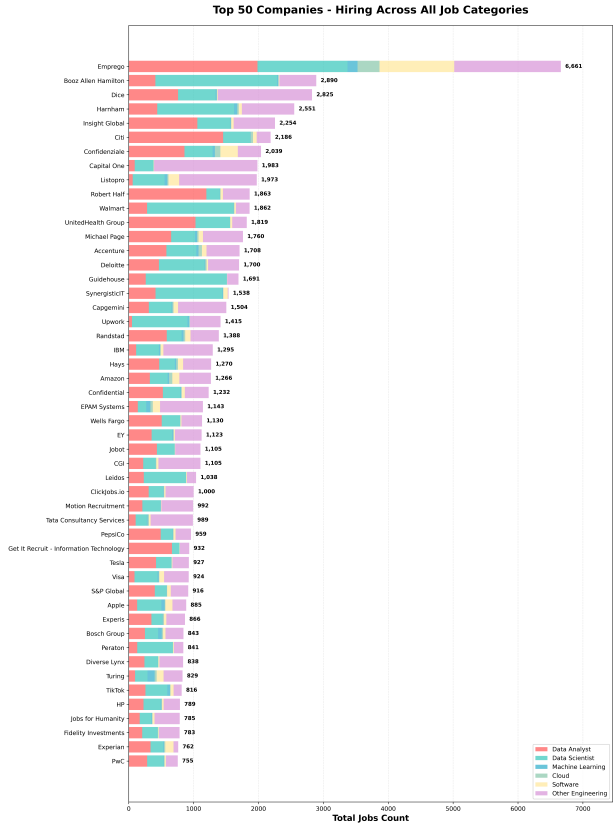

Figure 6. Top 50 companies' hiring patterns across ML, cloud, and data-related roles

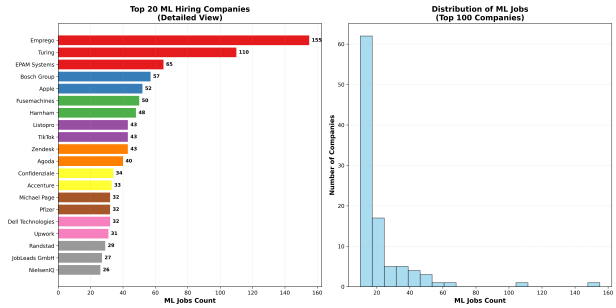

Figure 7. Machine learning hiring concentration analysis (based on 14,000 ML positions)

11. Conclusion

This analysis successfully demonstrates The usage of effective PostgreSQL Queries while uncovering valuable job market insights. The dominance of data-related positions (73%), significant compensation premiums for specialization (50%+), and limited remote opportunities (9%) provide actionable intelligence for career planning.

The technical implementation showcases database analysis capabilities applicable to business intelligence challenges. The integration of SQL analytics with Python visualization creates a comprehensive analytical framework suitable for ongoing market research.

For job seekers, the findings emphasize the value of data skills, particularly SQL and Python, while highlighting opportunities in underserved specializations like machine learning engineering. For employers, the analysis reveals competitive dynamics in technical hiring and potential talent acquisition strategies.

Job Market Executive Summary
(Ranked by Average Salary)

Role	Jobs Available	Avg Salary	Degree Req	Remote Rate	Health Coverage
Senior Data Scientist	37,076	$154,050	94%	10.3%	18.4%
Senior Data Engineer	44,692	$145,867	57%	14.7%	11.4%
Data Scientist	172,726	$135,929	94%	8.4%	15.2%
Data Engineer	186,679	$130,267	56%	11.4%	7.5%
Machine Learning Engineer	14,106	$126,786	93%	10.5%	2.8%
Senior Data Analyst	29,289	$114,104	64%	8.0%	18.1%
Software Engineer	45,019	$112,778	52%	6.5%	1.9%
Cloud Engineer	12,346	$111,268	52%	4.6%	1.3%
Data Analyst	196,593	$93,876	61%	6.8%	12.6%
Business Analyst	49,160	$91,071	74%	5.7%	5.9%

Figure 8. Executive summary showing metrics breakdown by job category