

Job Market Analysis: PostgreSQL Deep Dive with Python Visualizations

Omar Ayman Tawfiq Bakr  

Abstract—This project demonstrates PostgreSQL skills through comprehensive job market analysis, combining SQL queries with Python visualizations. The analysis explores hiring patterns, salary trends, and skill demands using database techniques. The implementation showcases practical database expertise while uncovering interesting insights about the tech job landscape.

Keywords—PostgreSQL, job market analysis, SQL, Python visualization, database analysis, tech industry trends

1. Introduction

The tech job market is a fascinating ecosystem that generates massive amounts of data daily. This project tackles that data using PostgreSQL as the analytical engine, demonstrating how SQL techniques can extract meaningful patterns from datasets. The analysis reveals patterns in the job market structure, from geographic hiring patterns to the premium companies pay for specialized skills.

The dataset contains hundreds of thousands of job postings across multiple countries and platforms, creating an excellent foundation for database analysis. Each analysis reveals part of a larger story about how the tech industry operates and what skills are in demand.

2. Database Architecture and Data Exploration

The analysis begins with understanding the dataset structure. The database follows a dimensional model with a central fact table surrounded by dimension tables - a classic data warehouse pattern that enables efficient analytical queries.

The geographic analysis in Figure 1 immediately reveals something interesting - this isn't just US-centric data. The substantial representation from multiple countries means our salary and hiring analyses will reflect genuine global patterns, not just Silicon Valley trends.

The pie chart in Figure 2 shows that more than 99% of the data was collected in 2023, indicating the data is fairly recent and relevant for current market analysis.

Figure 3 shows the data aggregates multiple major platforms, eliminating the risk of single-source bias that would skew our analysis. Among these sources, **LinkedIn**, the established market leader, and **BeBee**, a professional networking site that also aggregates jobs from other major platforms, are taking the lead. This diversity is essential for drawing valid conclusions about the broader market.

3. Job Market Volume and Salary Analysis

The volume analysis in Figure 4 shows that traditional data analyst and software engineer roles dominate the market. Machine learning positions demonstrate substantial presence in the industry.

The salary analysis in Figure 5 demonstrates clear compensation patterns within the tech industry. The most paying jobs are data-related positions, followed by machine learning engineering, then software engineers.

The comprehensive dashboard in Figure 6 provides an integrated view of job market dynamics across volume, salaries, benefits, and work arrangements.

4. Benefits and Requirements Deep Dive

Figure 7 reveals important patterns in job requirements and benefits. Most jobs require a degree, with data jobs showing the highest educational requirements at over 80%. Remote work opportunities

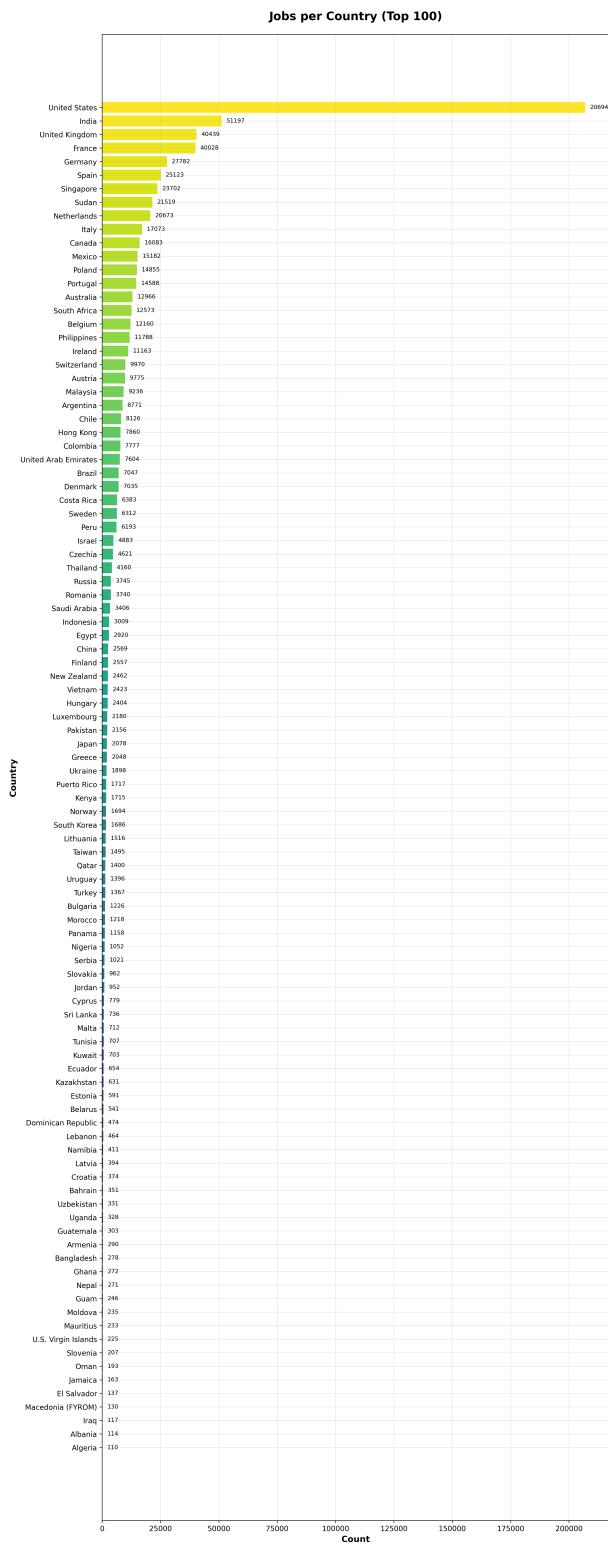


Figure 1. Geographic distribution reveals strong representation from major tech markets, with the US dominating but significant presence from European markets

vary dramatically by role - data analysts receive remote options more frequently than other positions.

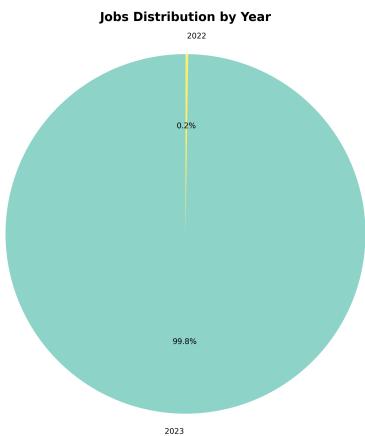


Figure 2. Temporal distribution shows that more than 99% of the data was collected in 2023

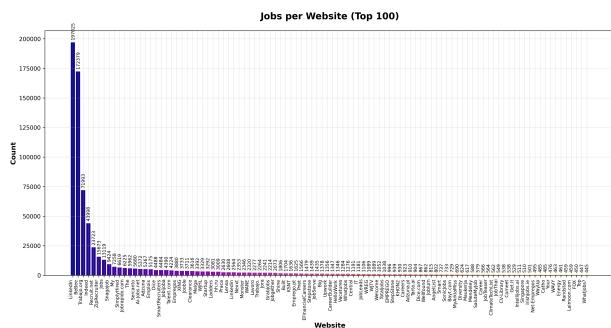


Figure 3. Platform diversity ensures comprehensive market coverage beyond any single job site's bias

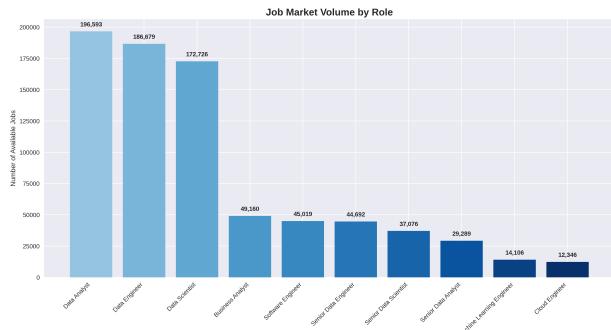


Figure 4. Job volume analysis reveals clear market leaders and emerging specializations

The education requirements analysis shows that data roles have the highest degree requirements, while other positions show more flexibility in educational backgrounds.

5. Skills Demand Analysis

The skills analysis in Figure 8 shows that programming languages dominate most jobs. However, analyst tools specifically dominate all data-focused positions, while cloud technologies hold the third position across all job types, which is interesting given the industry's shift toward cloud-first architectures.

Figure 9 confirms that while programming languages maintain overall dominance, the distribution varies significantly by role specialization, with data roles showing distinct preferences for analytical toolsets.

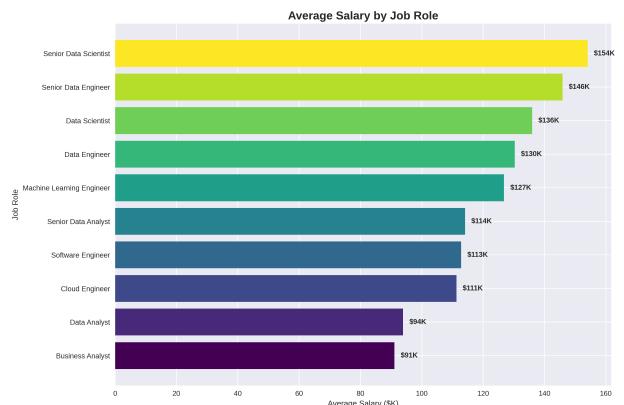


Figure 5. Salary analysis reveals distinct compensation tiers with machine learning roles commanding significant premiums

6. Company Hiring Patterns

I chose to analyze companies hiring for machine learning specifically because it's my interest, and we can see that Emprego is on the top, which is interesting, and Apple is the 4th, which is odd! It's important to note that the percentage of machine learning jobs in this dataset is considered marginal compared to the data-related jobs, which is why we are not seeing the FAANG companies up in the top.

Figure 11 provides deeper insight into machine learning hiring patterns, showing how this specialized market segment operates and the distribution of opportunities across different companies.

The comprehensive company analysis in Figure 12 demonstrates the broader hiring landscape across all technical roles, showing more distributed patterns than the specialized machine learning segment.

7. Key Market Insights

Figure 13 shows several important market characteristics:

- The dominant jobs in this data are data jobs (either analysis, engineer, or scientist)
- Most of the jobs are within the \$100-140k range
- Remote jobs are only around 9%
- Average health coverage is only 11%
- The number of jobs that require a degree is 64%

Figure 14 shows the percentage breakdown per each job category:

- Health care coverage rates by position type
- Remote work availability percentages
- Degree requirement distributions
- Average salary ranges
- Jobs available counts

You will find some interesting results in these breakdowns that reveal how different roles offer varying benefit packages and requirements.

8. Technical Implementation

This project demonstrates database analysis techniques applied to real-world job market data. The implementation covers several key areas:

Multi-Table Analysis: The analysis joins multiple tables to create comprehensive views of job market data, connecting job postings with company information, skills requirements, and geographic data.

Data Aggregation: Strategic use of grouping and aggregation functions to summarize data across different dimensions, providing insights into salary ranges, skill distributions, and company hiring patterns.

Analytical Queries: Implementation of ranking and statistical functions to identify top companies, most demanded skills, and compensation trends across different job categories.

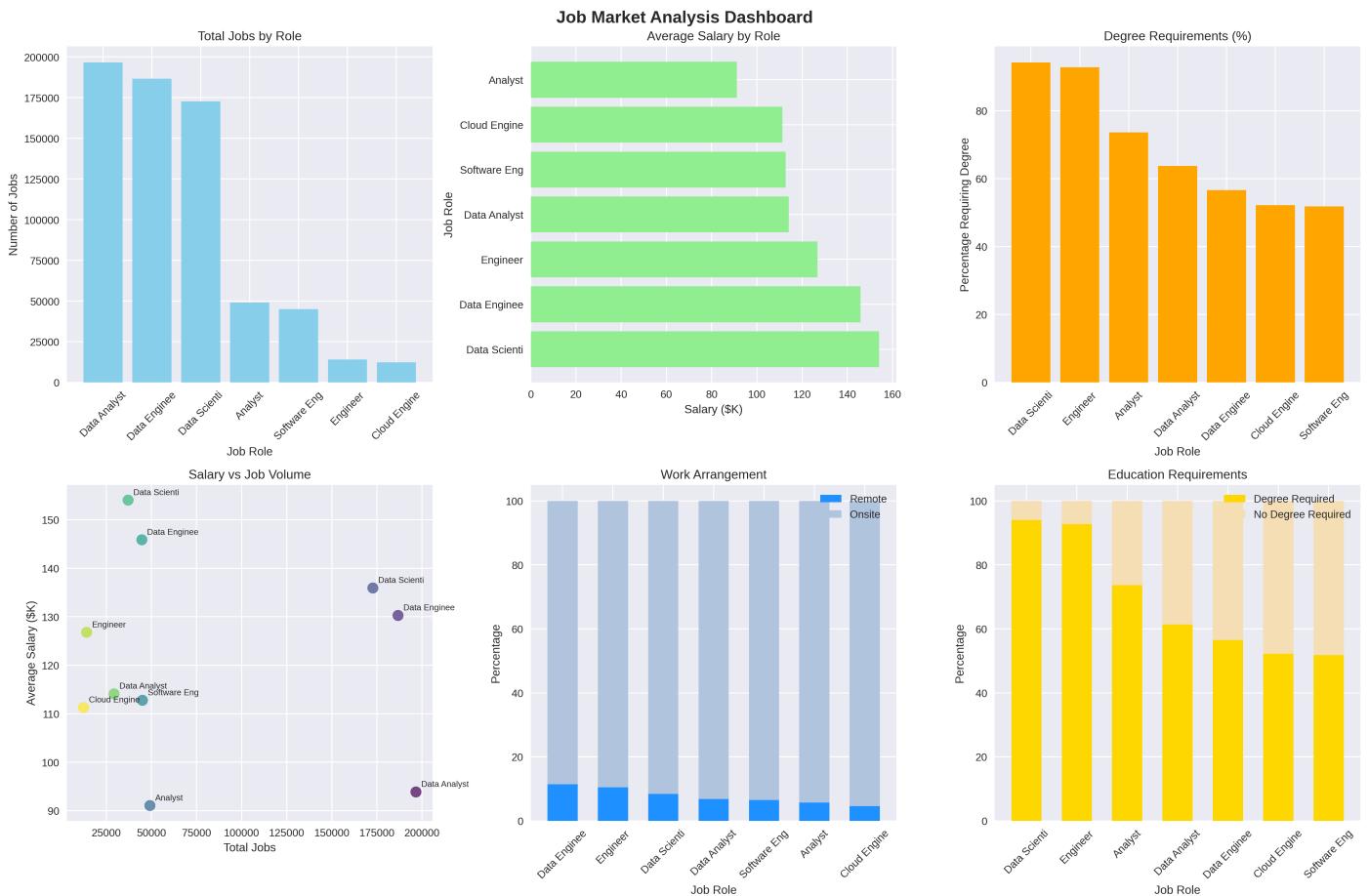


Figure 6. Comprehensive dashboard combining multiple analytical views of the job market

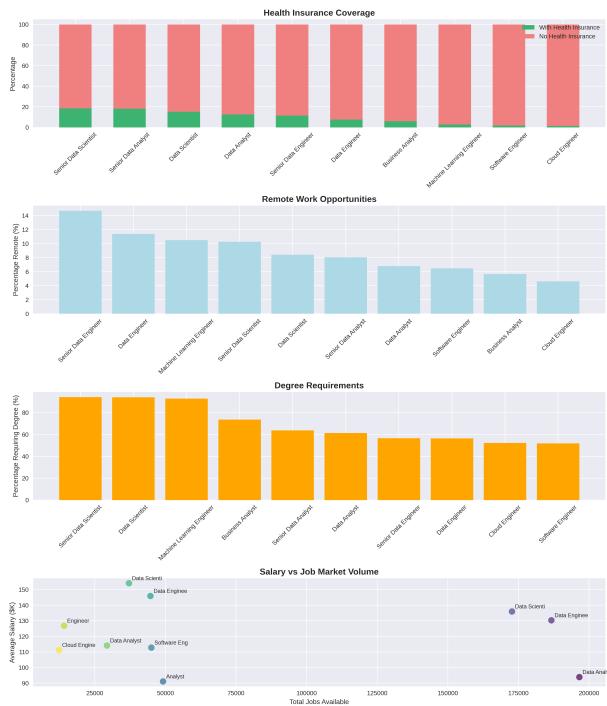


Figure 7. Multi-dimensional analysis of health insurance, remote work, and education requirements across roles

9. Python Visualization Integration

The database analysis connects with Python visualization tools to create comprehensive visual representations of the findings:

Database Connectivity: Direct connection to PostgreSQL with efficient data retrieval and processing using pandas for data manipulation and analysis.

Statistical Processing: Data processing and statistical analysis to identify patterns, trends, and correlations within the job market data.

Professional Visualization: Creation of publication-quality visualizations using matplotlib and plotly to present findings in clear, accessible formats.

10. Project Insights and Learnings

The analysis reveals several important patterns about the tech job market:

Data-related roles consistently command the highest salaries, with machine learning engineering positions showing particularly strong compensation packages. This trend reflects the growing importance of data science and machine learning capabilities in modern businesses.

Educational requirements vary significantly by role type, with data positions requiring degrees more frequently than other technical roles. This suggests that formal education remains important for data-focused careers.

Remote work opportunities remain limited across most technical positions, representing less than 10% of available roles. This indicates that most tech companies still prefer in-person or hybrid work arrangements.

Geographic distribution shows continued concentration in traditional tech markets, though the presence of international opportunities provides options for global talent mobility.

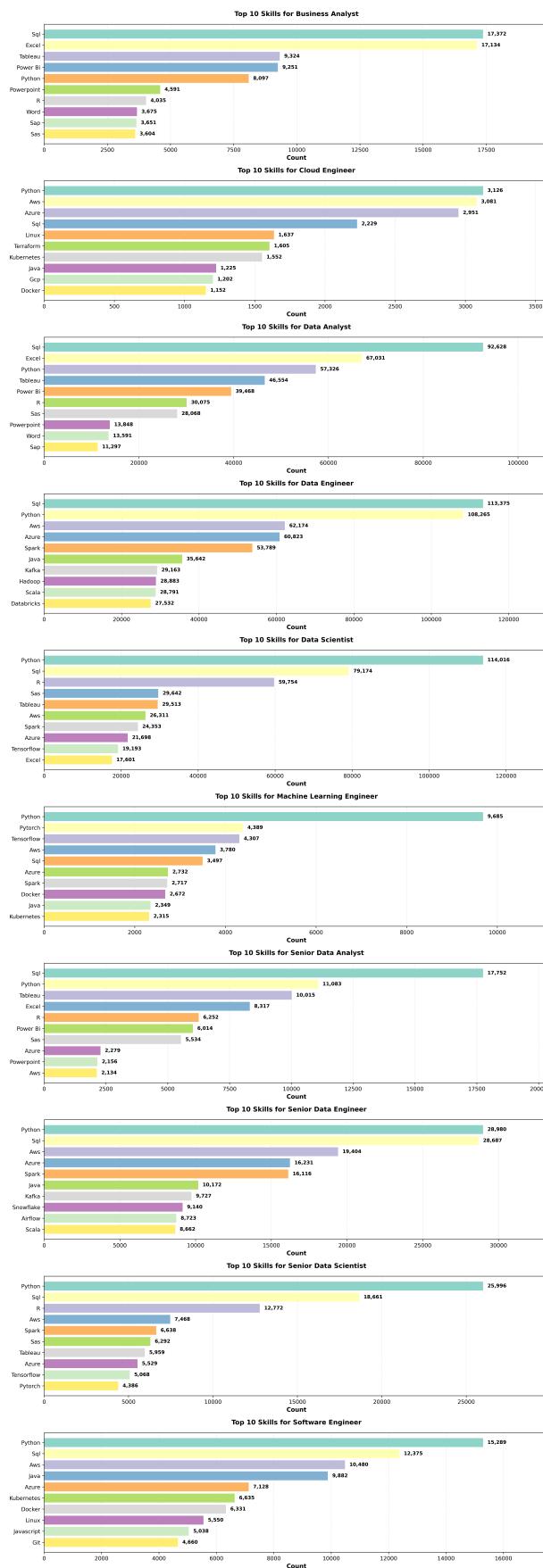


Figure 8. Individual skills demand reveals the technical competencies driving hiring decisions

11. Conclusion

This project shows good intermediate coverage of PostgreSQL. We were able to find some good insights from the data and analyze the

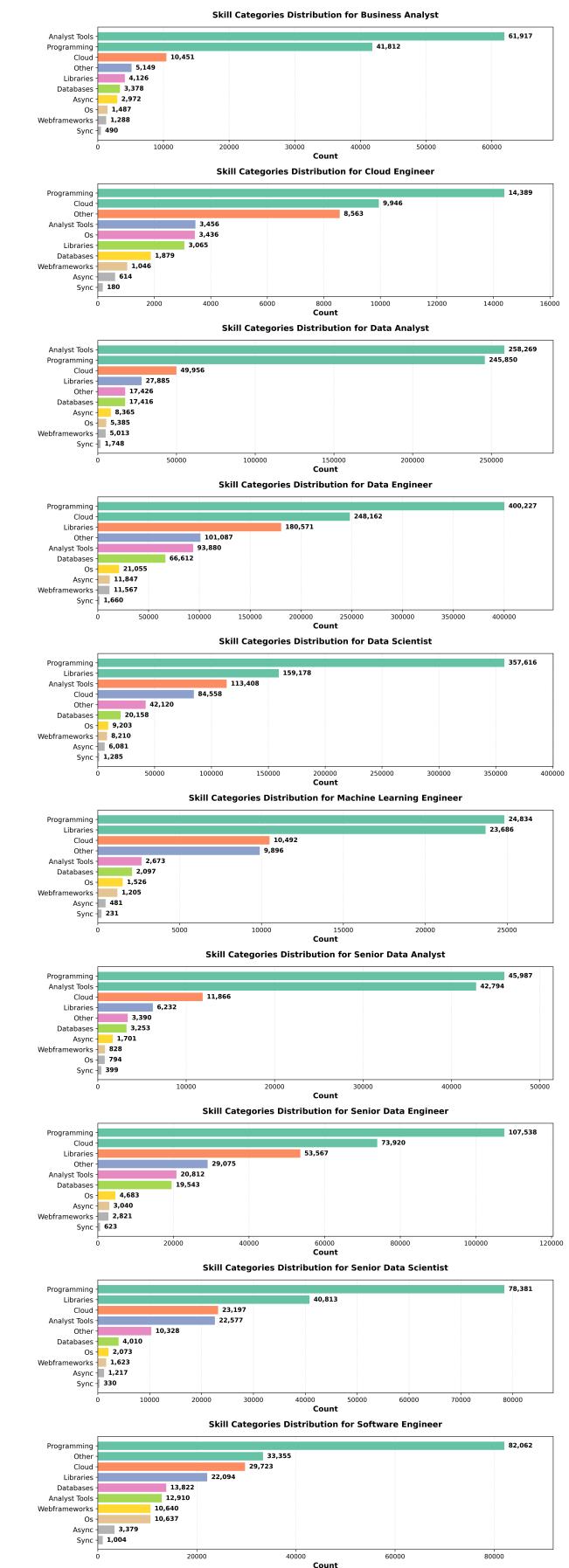


Figure 9. Skill category analysis showing the balance between different technical competency areas

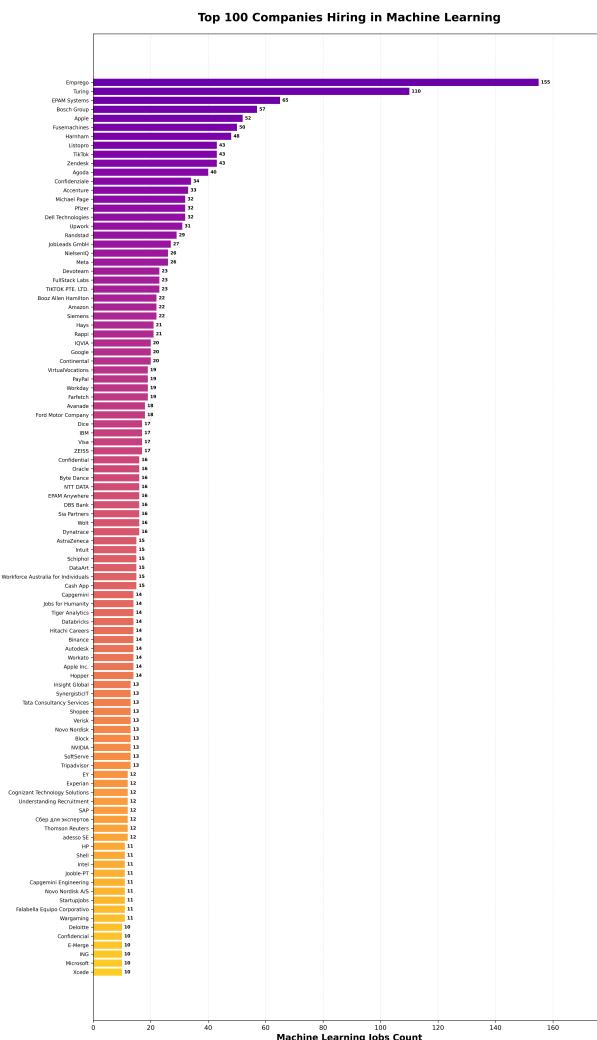


Figure 10. Machine learning hiring leaders demonstrate interesting market patterns

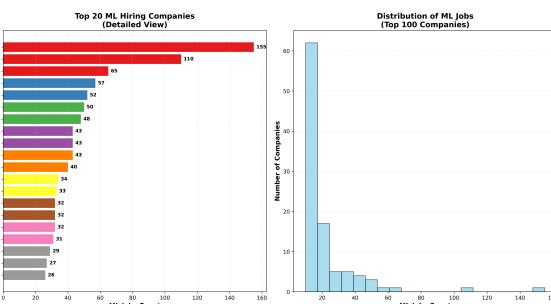


Figure 11. Detailed hiring pattern analysis reveals distribution characteristics and market concentration

patterns within the data and jobs. The analysis successfully identified key trends in compensation, skill requirements, and hiring patterns across different technical specializations.

The findings provide practical value for understanding current job market conditions and can inform decision-making for both job seekers and employers. The analytical approach demonstrates how database skills can be applied to extract meaningful business intelligence from real-world datasets.

The project establishes a solid foundation for ongoing market research and could be extended to examine additional aspects of employment trends and industry dynamics.

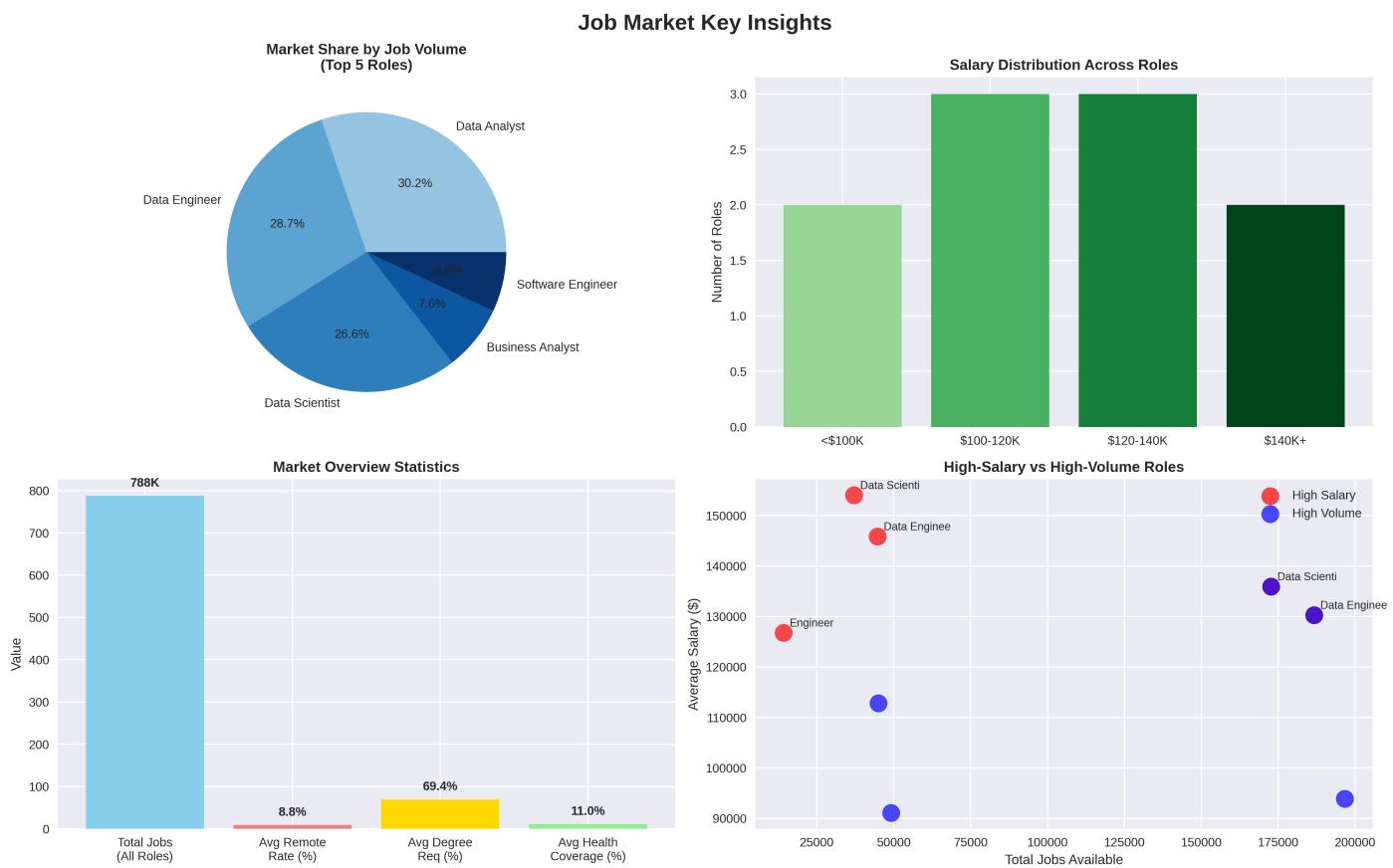


Figure 13. Executive insights summary highlighting critical market patterns

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 14. Comprehensive breakdown showing key metrics per job category