PROJECT REPORT

On

Trends and Skill Analysis in Data Science Job Posting

Submitted To
Genniesphere

Submitted By
Name: Sakshi Gupta

1. Abstract

This project focuses on analysing trends in the data science job market using a cleaned dataset of job postings from Glassdoor. By leveraging data analysis techniques, we aim to extract insights into the most in-demand skills, high-growth job titles, company hiring patterns, and geographic job hubs. The analysis uses Python and visualization libraries to explore job descriptions, titles, company details, and skill requirements. The outcome helps stakeholders—including job seekers, educational institutions, and employers—understand the dynamic landscape of data roles.

2. Introduction to the Project

2.1 Background

In recent years, data science has emerged as one of the most sought-after fields across multiple industries. Companies are increasingly relying on data professionals to generate insights, improve decision-making, and develop AI-powered solutions. However, the skills required, job roles offered, and industry preferences often vary significantly. Analyzing job market trends helps stakeholders—such as job seekers, educators, and HR professionals—adapt to the evolving landscape by focusing on relevant skills and opportunities.

2.2 Problem Statement

The project aims to solve the following questions:

- What are the most in-demand technical and soft skills across industries?
- Which job titles are growing rapidly in the data domain?
- What regions in the U.S. are emerging as major hubs for data jobs?
- How do skill requirements differ by job level and industry?

This project provides a data-driven solution to identify patterns and trends from real job postings using natural language processing and data visualization.

3. Software and Hardware Requirement Specification

3.1 Methods

The methodology includes:

- Data loading and cleaning
- NLP-based text extraction of skills
- Categorization of job titles and career levels
- Visualization of trends using bar charts, pie charts, and 3D plots
- Grouping and aggregation based on sector, company, and locations

4. Overview

This report summarizes final insights derived from job posting data related to data roles. The analysis highlights current market demands, growth areas, and location-based trends. The insights are designed to support strategic hiring and workforce planning.

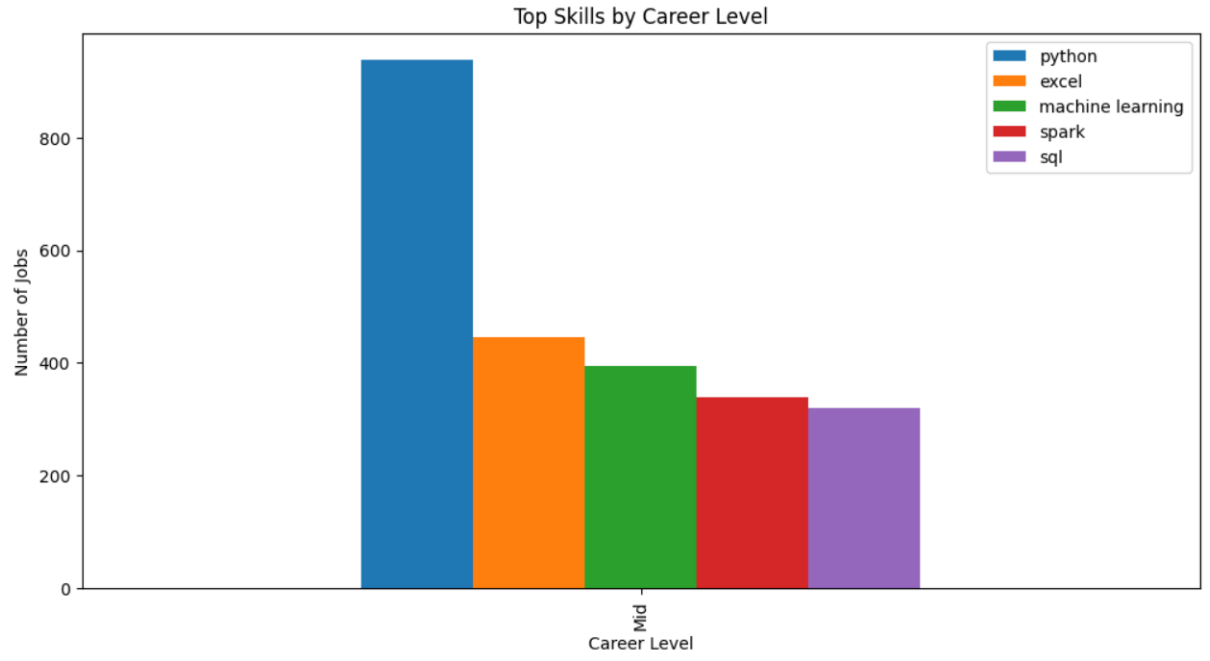
5. Key Insights

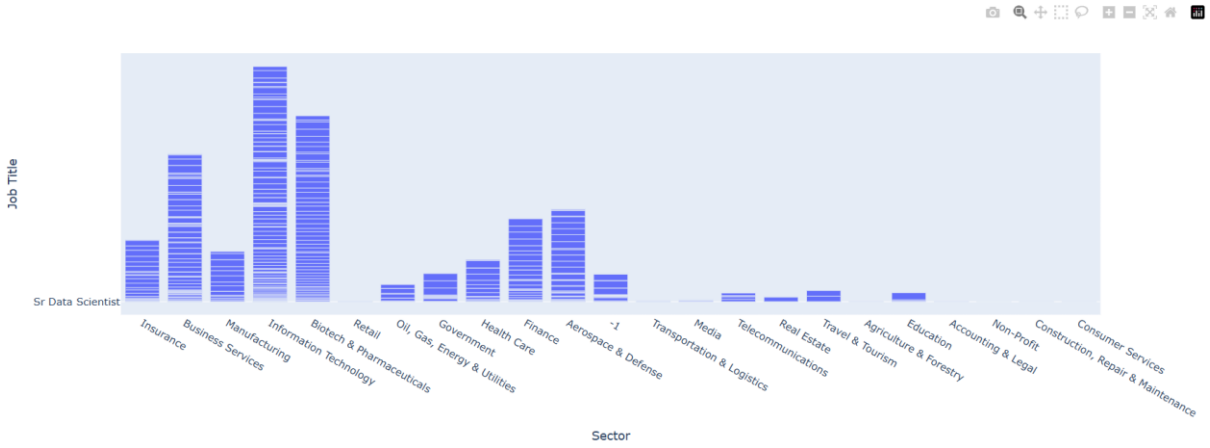
5.1 Technical Skills

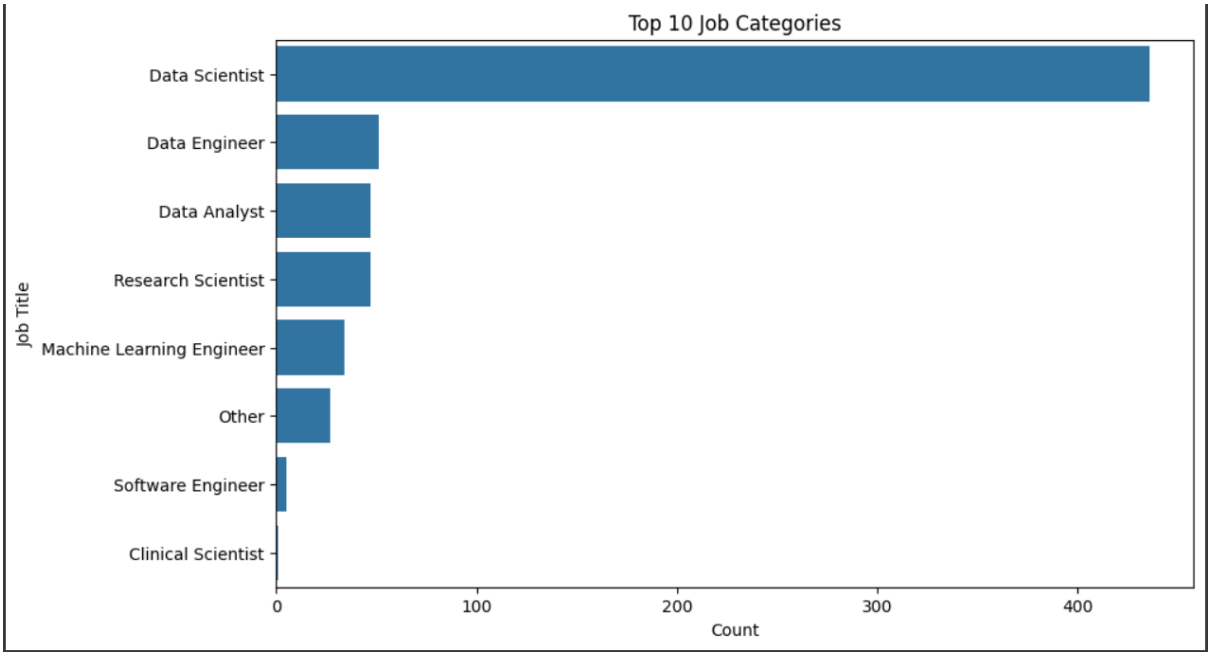
- Python: Very High - Used across all data roles; foundational for analytics and ML
- SQL: Very High - Core for data extraction, manipulation, and reporting
- Excel: High - Preferred for analysts, business intelligence, and quick reporting
- AWS: High - Needed in mid-to-senior roles for cloud computing
- Tableau: High - In demand for data visualization, often in analyst roles
- Spark & Hadoop: Moderate - More common in engineering and big data roles
- Power BI: Moderate - Common in business and healthcare analytics

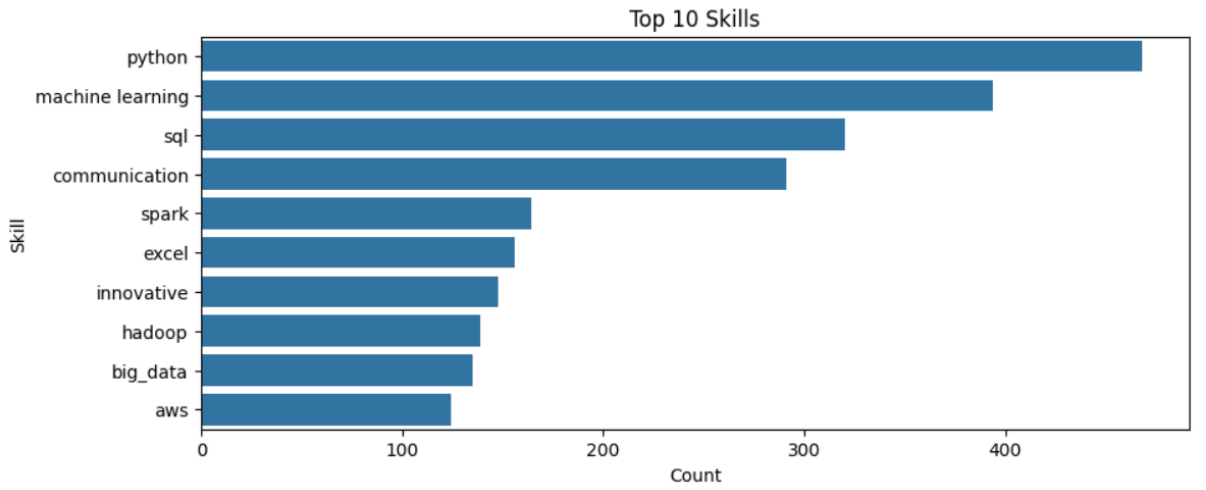
5.2 Soft Skills

- Communication
- Problem Solving
- Teamwork
- Adaptability
- Innovation


Top Skills by Career Level
python
excel
machine learning
spark
sql
Number of Jobs
0
200
400
600
800
Mid
Career Level


Job Title
Sr Data Scientist
Insurance
Business Services
Manufacturing
Information Technology
Biotech & Pharmaceuticals
Retail
Oil, Gas, Energy & Utilities
Government
Health Care
Finance
Aerospace & Defense
-1
Transportation & Logistics
Media
Telecommunications
Real Estate
Travel & Tourism
Agriculture & Forestry
Education
Accounting & Legal
Non-Profit
Construction, Repair & Maintenance
Consumer Services
Sector


Top 10 Job Categories
Data Scientist
Data Engineer
Data Analyst
Research Scientist
Machine Learning Engineer
Other
Software Engineer
Clinical Scientist
Job Title
0
100
200
300
400
Count



6. High- Growth Job Titles

- Data Scientist: Most dominant role, includes Senior and Lead variants
- Data Analyst: High demand, especially in entry and mid-level positions
- Machine Learning Engineer: Fast-growing with demand in AI-driven organizations
- Data Engineer: Key role in data infrastructure and pipeline development
- Business Analyst: Often hired in healthcare, consulting, and finance

7. Regional Hubs for Hiring

- California: Silicon Valley leads in hiring for data and ML professionals
- New York: High activity in finance, healthcare, and consulting sectors
- Texas: Strong presence in Austin, Dallas for tech and health analytics
- Illinois: Particularly Chicago; hub for finance and business analytics
- Massachusetts: Boston leads in education, biotech, and health tech roles