Data Science Salary Analysis

An in-depth analysis of data science job salaries to uncover trends, identify salary drivers, and predict compensation based on job-related factors using a dataset of 607 entries across various roles, experience levels, and company types.

Dataset Overview

Dataset Size

607 entries with 11 features before cleaning

565 entries after removing duplicates

Key Features

- Work year (2020-2022)
- Experience level
- Employment type
- Job title
- Salary (in USD)

Additional Factors

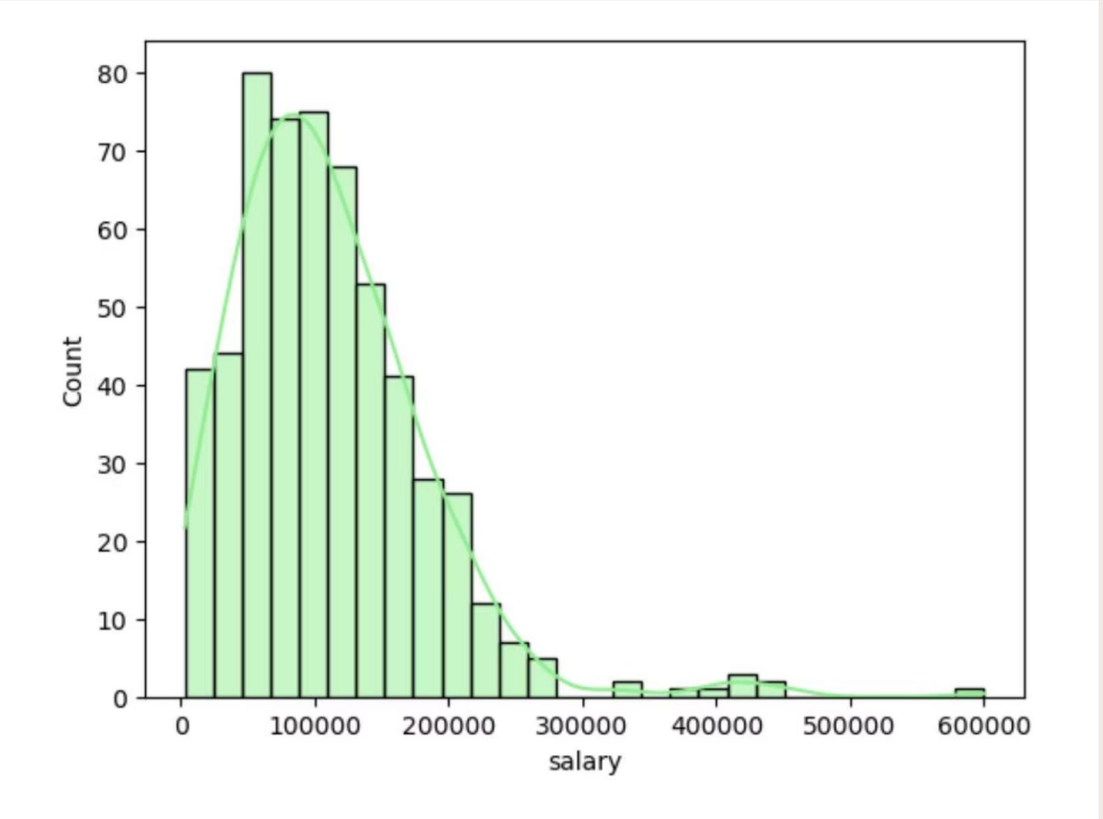
- Remote ratio (0%, 50%, 100%)
- Company size (S, M, L)
- Company location
- Employee residence

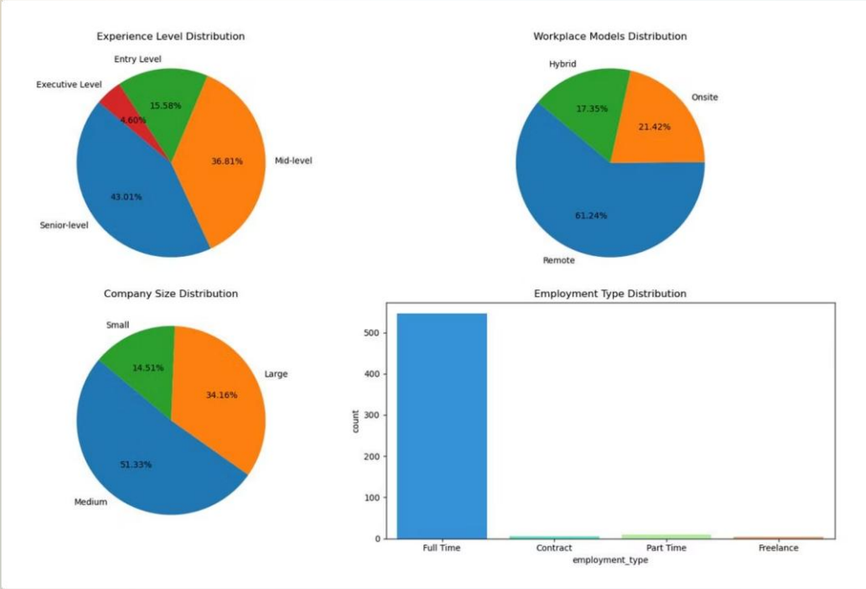
The dataset was cleaned by removing duplicates and unnecessary columns, with categorical variables transformed for analysis.

Salary Distribution

The salary distribution exhibits positive skewness (1.73), indicating that most salaries are concentrated at the lower end while a few high salaries pull the mean upward.

This pattern is typical in compensation data, where a small number of executive or specialized roles command significantly higher salaries than the majority of positions.





Workforce Composition

Experience Level

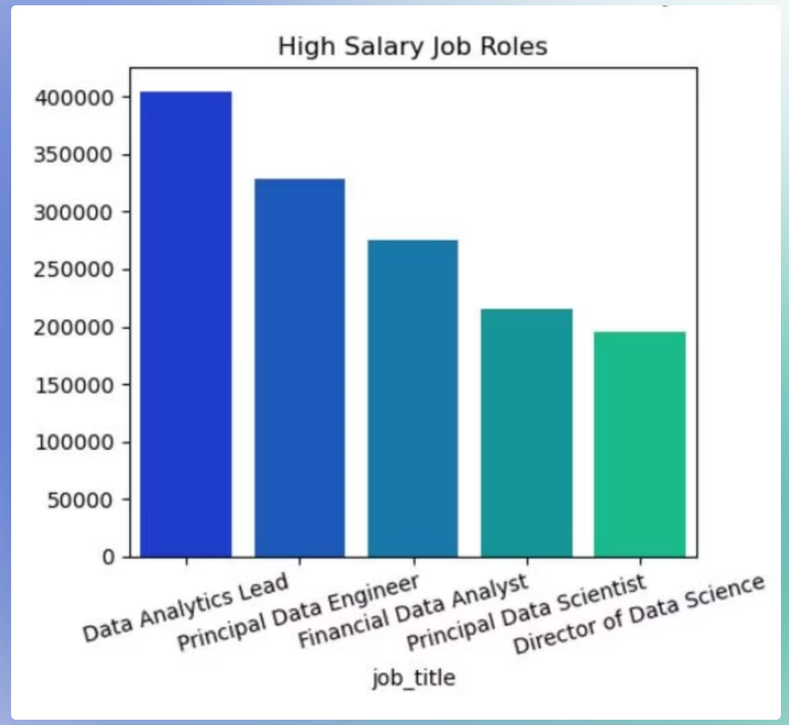
Senior-level professionals dominate the dataset, representing the largest segment of the data science workforce.

Work Models

Remote work is the predominant mode, highlighting the industry's shift toward flexible work environments.

Company Size

Medium-sized companies constitute the majority, suggesting a balanced organizational scale in the dataset.



Salary by Job Role

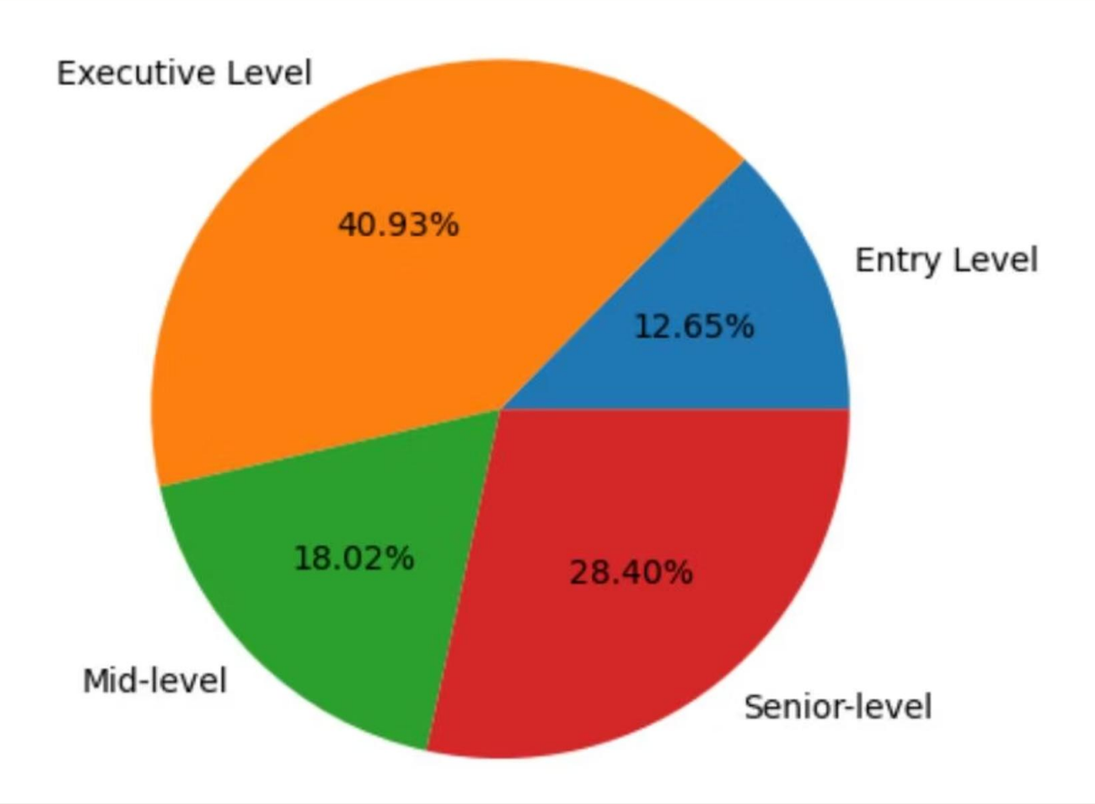
Highest Paid Roles

- Data Analytics Lead
- Principal Data Engineer
- Financial Data Analyst
- Director of Data Science
- Data Science Tech Lead

Lowest Paid Roles

- 3D Computer Vision Researcher
- BI Data Analyst
- Computer Vision Engineer
- NLP Engineer
- AI Scientist

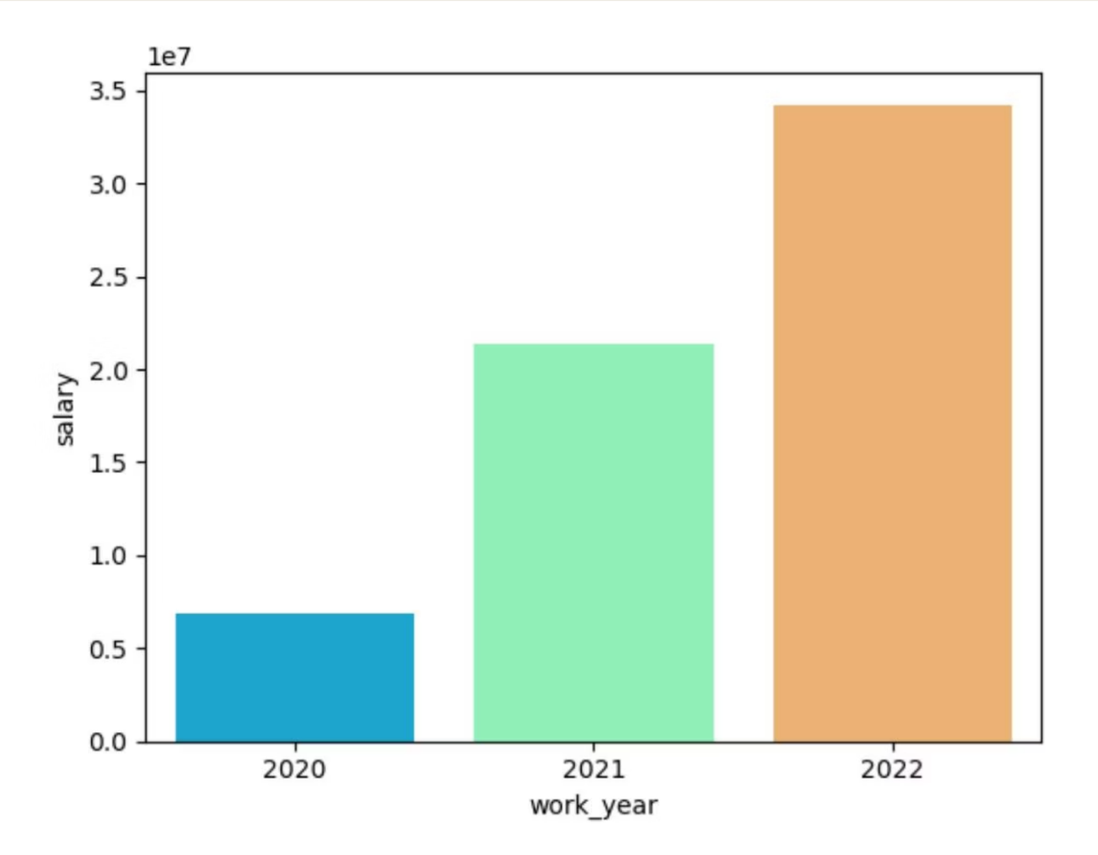
Salary by Experience Level



Key Findings

- Executive Level positions command the highest salaries, as expected
- Senior-level positions earn significantly more than mid-level roles
- Entry-level positions have the lowest compensation
- The progression from entry to executive level shows a clear salary ladder in data science careers

Salary Trends Over Time

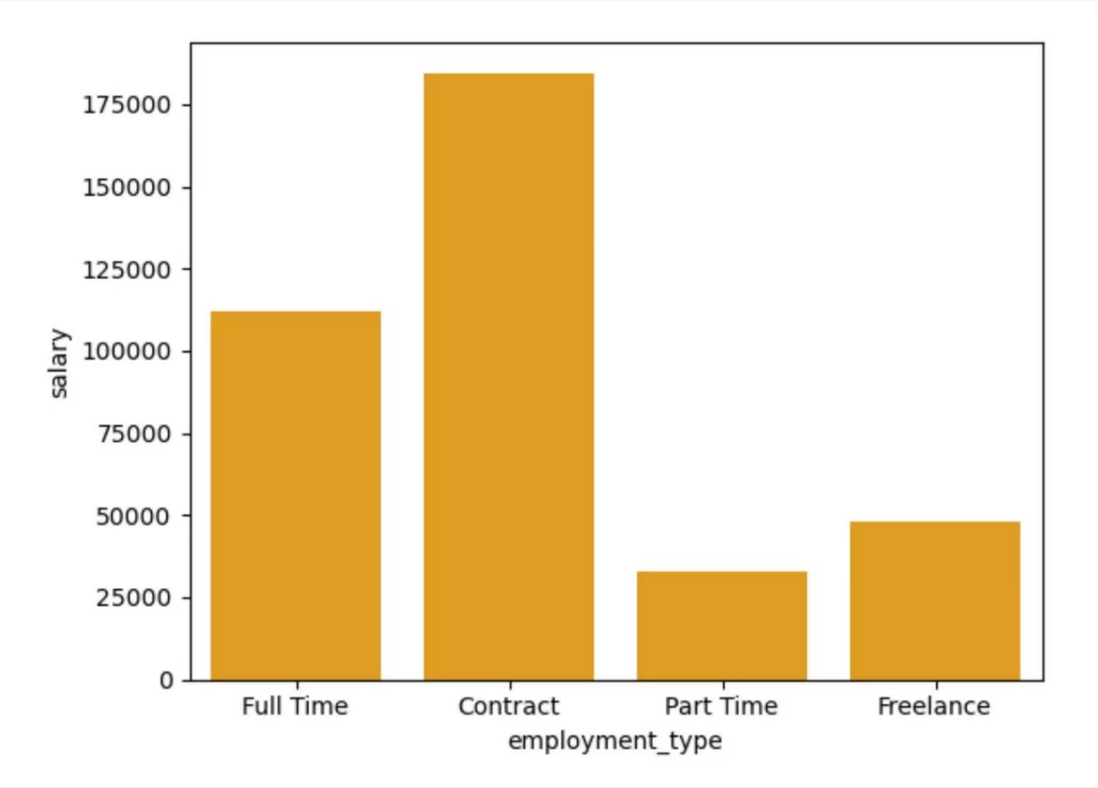


Year-over-Year Growth

The data shows a clear upward trend in data science salaries from 2020 to 2022, indicating growing demand for these skills in the market.

This trend aligns with the industry's expansion and the increasing importance of data-driven decision making across sectors.

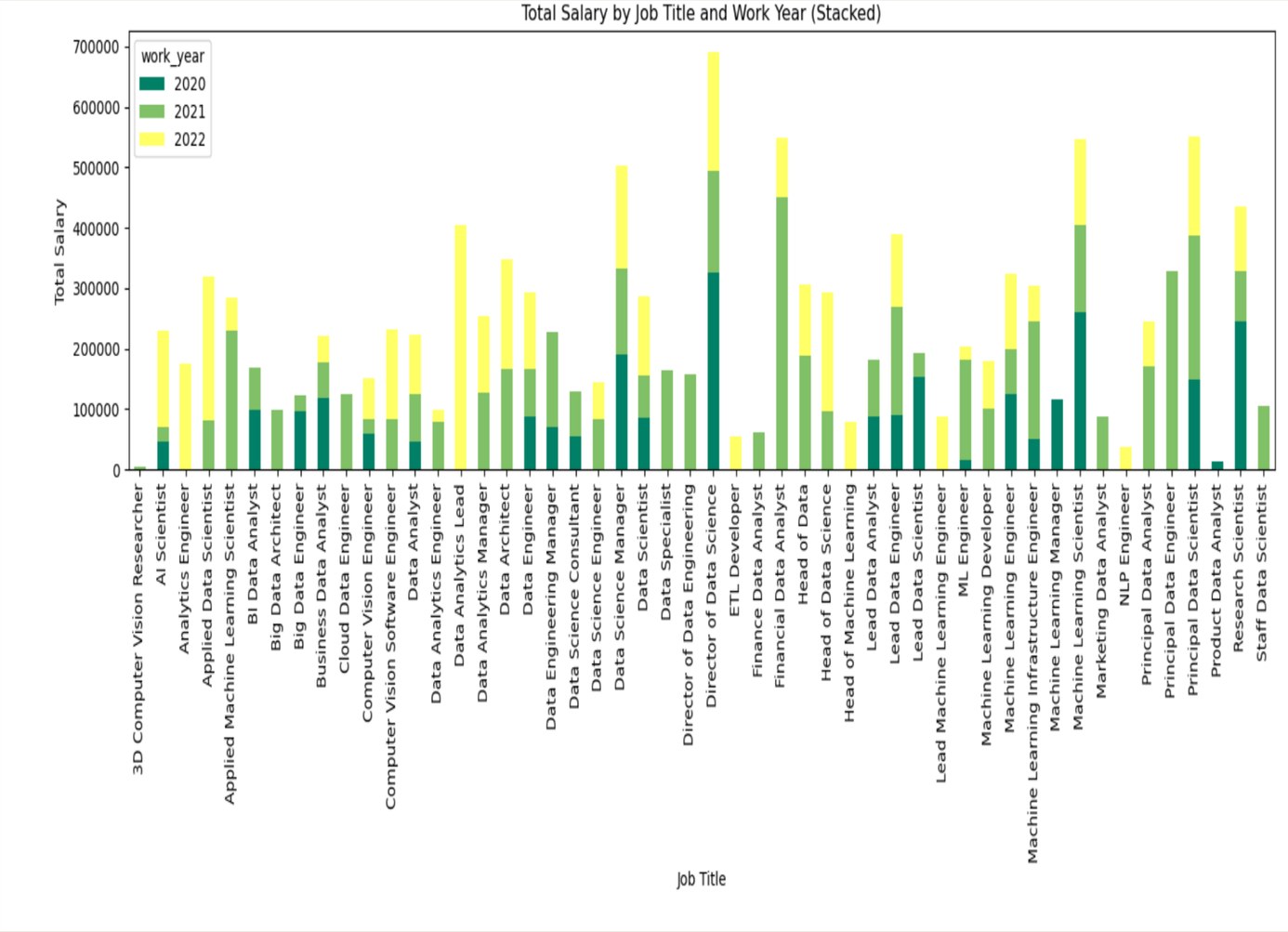
Salary by Employment Type



Contract Workers Earn More

Interestingly, contract employees earn higher salaries compared to full-time, part-time, and freelance workers.

This may reflect the premium paid for specialized, temporary expertise and the lack of benefits typically provided to permanent employees.



Job Title Salary Trends

The stacked bar chart reveals that Director of Data Science positions maintain consistently high compensation across all years, while 3D Computer Vision Researcher roles reflect the lowest compensation levels. This visualization also shows how certain roles have seen significant salary growth over the three-year period.

Predictive Modeling Results

26.2%

Random Forest R^2 Score

Best performing model with MSE of 3,844,751,527

23.0%

Linear Regression R^2 Score

MSE of 4,009,784,855

23.2%

KNN Regressor R^2 Score

MSE of 3,999,564,008

After hyperparameter tuning with Optuna, the Random Forest model achieved the best performance. However, the relatively low R^2 scores suggest that additional features or more complex models might be needed to better predict data science salaries.