# Internship Project Report

Project Title  
Data Science Job Salaries Analysis & Modeling

Name:

Kavitha Shree H.R

Course:  
B.Sc. Data Science

Internship Domain  
Finance Analyst / Data Analytics

Date  
June 2025

## Table of Contents

1. 1. Abstract
2. 2. Company & Context
3. 3. Objectives & Project Definition
4. 4. Methodology & Design
5. 5. Implementation & Results
6. 6. Challenges & Solutions
7. 7. Learnings & Skill Growth
8. 8. Future Recommendations
9. 9. Conclusion
10. 10. Tech Stack
11. 11. Appendices

## 1. Abstract

This project analyzed global data science job salaries with a focus on finance analyst applications. Through data cleaning, EDA, and machine learning modeling, the project provided insights into how factors such as experience level, job type, and company size impact compensation. A linear regression model was implemented, achieving an MAE of ~$14,000. This project showcased how data science can inform compensation and workforce planning strategies.

## 2. Company & Context

This internship project was designed to simulate a real-world financial analysis scenario using data from AI-jobs.net. It aligns with financial modeling, helping HR or finance teams understand cost implications of hiring in various job markets. The project was conducted as part of a data analytics internship.

## 3. Objectives & Project Definition

- Analyze salary trends based on categorical and numerical job-related factors  
- Build regression models to predict salary  
- Identify highest paying roles and locations  
- Support HR decision-making with insights into pay structures

## 4. Methodology & Design

- Data Cleaning: Removed duplicates, standardized category labels  
- Feature Engineering: Mapped categorical levels, converted currencies, created job type variable  
- EDA: Used bar plots, violin plots, and correlation analysis to understand salary distribution  
- Modeling: Trained linear regression using experience level, job type, company size

## 5. Implementation & Results

- Final MAE: ~$14,000  
- Remote workers had the highest mean salaries (~$120K)  
- Executive experience level yielded ~$199K, Entry ~$61K  
- Top-paying roles: Principal Data Engineer, Financial Data Analyst

## 6. Challenges & Solutions

- Dataset imbalance between high and low salary roles  
- Solution: Used data binning and filtering to focus on meaningful categories  
- Complexity in encoding categorical features  
- Solution: Used label encoding and one-hot encoding as needed

## 7. Learnings & Skill Growth

- Gained strong understanding of real-world salary factors  
- Improved regression modeling, encoding, and evaluation skills  
- Mastered visualization tools (Seaborn, Matplotlib)  
- Learned Streamlit for interactive dashboard development

## 8. Future Recommendations

- Use advanced models like Random Forest or XGBoost  
- Incorporate job title embedding with NLP techniques  
- Add more granular time-series salary analysis  
- Include industry-specific salary segmentation

## 9. Conclusion

This project effectively demonstrated the power of data analytics and ML in understanding salary patterns. It provided insights that are relevant to finance analysts, HR teams, and job seekers. With more data and improved modeling, the predictive power and decision support potential of such analysis can be greatly enhanced.

## 10. Tech Stack

- Language: Python  
- Libraries: Pandas, Scikit-learn, Matplotlib, Seaborn, Plotly, country\_converter  
- Tools: Jupyter Notebook, Google Colab, Streamlit

## 11. Appendices

Key Visuals & Code Snippets  
- EDA Charts (salary distribution, box plots)  
- Feature Encoding Examples  
- Regression Model Training & Evaluation Code