

TANNU JHA

Delhi 110088 | jhatannu828@gmail.com | +91 8447156251 | linkedin.com/in/tannu-jha-5989a9392 | github.com/Tannujha828

SUMMARY

Data Science trainee with hands-on experience building and evaluating machine learning models across regression, classification, and clustering. Completed a Master Diploma in Data Science & AI from Boston Institute of Analytics and interned at Spinnaker Analytics, delivering 5 end-to-end ML projects using Python, SQL, and Power BI. Seeking an entry-level Data Analyst / Data Science role to turn data into decisions.

SKILLS

Data Analysis & BI: Advanced Excel (Pivot Tables, VLOOKUP/XLOOKUP), SQL (MySQL), Power BI, Tableau, Google BigQuery, Data Cleaning, Report Preparation

Programming & Libraries: Python, Pandas, NumPy, Scikit-learn, Jupyter Notebook

Statistics & Machine Learning: Statistics & Math, Feature Engineering, Machine Learning, Model Evaluation, Deep Learning (Basics)

Data Visualization: Matplotlib, Seaborn, Data Visualization

AI/NLP: NLP (Basics), Gen AI

PROJECTS

NYC Salary Range Prediction | *Python, Scikit-learn, XGBoost, Pandas*

- Built a multi-output regression pipeline on 5,120+ NYC job postings, comparing Decision Tree, Random Forest, and XGBoost; achieved best $R^2 \approx 0.757$ (MAE $\approx 10,892$) after 5-fold cross-validation and GridSearchCV tuning.

Predicting Used Car Prices | *Python, Scikit-learn, XGBoost, Pandas, Matplotlib*

- Compared Decision Tree, Random Forest, and XGBoost (R^2 : 0.36 $\rightarrow$ 0.56) on a used car dataset; improved accuracy through outlier removal and feature engineering, and built a Power BI dashboard to present results.

Classifying Shooting Incident Fatality | *Python, Logistic Regression, Random Forest, XGBoost, SMOTE, Power BI*

- Built a classification pipeline on the NYPD Shooting Incident dataset to predict fatality; handled class imbalance with SMOTE, evaluated with F1-score/ROC-AUC, and visualized results in a Power BI dashboard.

Hotel Reservation Cancellation Prediction | *Python, Logistic Regression, Random Forest, Pandas, Seaborn*

- Analyzed hotel booking data to predict cancellations; performed EDA and feature engineering, then compared classification models using accuracy, precision, recall, and F1-score.

Product Cluster Analysis (K-Means) | *Python, Scikit-learn, K-Means, Matplotlib, Seaborn*

- Applied K-Means clustering on warehouse and retail sales data to segment products into four business-relevant groups; presented findings via a Word report and PowerPoint deck.

EXPERIENCE

Data Science Intern — Spinnaker Analytics (Boston Institute of Analytics Program)

03/2026 – 09/2026

- Built and evaluated 5 end-to-end machine learning projects across regression, classification, and clustering using Python, Scikit-learn, and XGBoost, achieving $R^2 \approx 0.76$ through feature engineering and hyperparameter tuning.
- Created Power BI dashboards and Word/PowerPoint reports to communicate model findings to stakeholders.

Office Assistant — Apeejay Education, New Delhi

01/2025 – Current

- Led a digitization initiative converting physical files into organized digital databases, improving accessibility; prepared Excel-based reports supporting internal data analysis.

Junior Accountant — Ravi Chaudhary & Associates, Delhi

06/2024 – 12/2024

- Managed financial transactions, reconciliation, and ledger maintenance; tracked cash flow and prepared TDS return sheets and ITR filings to ensure regulatory compliance.

EDUCATION

Bachelor of Commerce — Lakshmibai College

2022 – 2025

CERTIFICATIONS

- Master Diploma in Data Science & AI — Boston Institute of Analytics, Connaught Place Campus

LANGUAGES

Hindi | English