

Pratik Raj

+91-9341890052 · rajpratik139@gmail.com · linkedin.com/in/pratik-raj · github.com/PRATIK-4751 · pratik05.vercel.app
· Haldia, West Bengal

Summary

B.Tech Data Science student (2027 batch) who enjoys getting into messy datasets and finding something useful. I work primarily in Python and SQL, build dashboards people actually use, and have maintained a 261-day LeetCode streak. Currently seeking a Data Analyst internship where I can apply these skills to real business problems.

Technical Skills

Languages & Querying	Python, SQL (MySQL) — window functions, CTEs, joins, subqueries
Data Analysis & EDA	Pandas, NumPy, data cleaning, outlier detection (IQR, Z-score), imputation, statistical analysis
Visualization	Matplotlib, Seaborn, Plotly, Streamlit dashboards, KPI reporting, business storytelling
Machine Learning	Scikit-learn, regression, classification, model evaluation, feature engineering
AI & LLM Engineering	LangChain, Hugging Face, Ollama, RAG systems, vector databases, prompt engineering
Dev Tools	Git, GitHub, Docker, Jupyter Notebook, Flask, FastAPI, Supabase, Render, Streamlit

Projects

Retail Sales EDA & Business Intelligence Dashboard | *Python, Pandas, SQL, Seaborn, Streamlit* github.com

- Profiled a 10,000+ row superstore dataset end to end — treated missing values using median and mode imputation by column type, removed 37 duplicate order IDs, and fixed inconsistent date formats before any analysis began
- Wrote SQL queries using window functions (RANK, LAG, DENSE_RANK with CTEs) to extract month-over-month revenue changes, top-10 products by profit margin, and state-level sales rankings across all four regions
- IQR outlier check on the discount column revealed that every order with a discount above 40% produced a negative profit margin — held true across all product categories without a single exception
- Built a Streamlit dashboard with KPI cards for Revenue, Profit Margin, Average Order Value, and Top Category, all filterable by region, category, and date range so stakeholders can explore the data without writing SQL
- Delivered a structured insight report flagging Furniture as the weakest category by profit despite ranking 2nd in total sales; primary recommendation was a 30% discount cap estimated to recover the majority of margin losses

AI Financial Analyst | *Python, Streamlit, Pandas, OpenRouter API, Ollama* github.com

- Built a stock analysis platform covering 5+ years of price history across 50+ equities; computed rolling averages, volatility scores, and Pearson correlation matrices to surface relationships between asset classes that are not visible in raw price charts
- Performed exploratory data analysis on historical market data to identify trend patterns, seasonal behaviour, and anomalies using statistical methods including moving averages, standard deviation bands, and correlation heatmaps
- Layered an LLM backend using OpenRouter and Ollama that reads statistical outputs and generates plain-English summaries — making complex financial analysis accessible to users without a quantitative background
- Integrated live web scraping so users get current price data alongside historical context in one interface; all filters (ticker, date range, metric) update dynamically on the Streamlit front end without page reloads
- Designed a multi-model comparison feature allowing users to switch between different LLM backends and evaluate analytical perspectives side by side — reducing time to insight by eliminating manual interpretation

Achievements & Certifications

- **LeetCode** — 216 problems solved (144 Easy / 68 Medium / 4 Hard) with a 261-day maximum streak; consistent daily practice in Python and SQL | profile
- **Oracle Generative AI Professional Certificate** — GenAI architectures, retrieval-augmented generation, LLM evaluation, and prompt engineering patterns
- **Hugging Face NLP Specialization** — Transformer model architecture, fine-tuning pre-trained models, text classification, and sequence-to-sequence tasks
- **Hacktoberfest 2024** — Active open-source contributor with pull requests reviewed and merged across public repositories

Education

Haldia Institute of Technology <i>B.Tech in Data Science</i> Coursework: Machine Learning, Statistics & Probability, Database Management Systems, Data Structures & Algorithms, Linear Algebra, Python Programming	<i>Haldia, West Bengal</i> <i>Expected May 2027</i>
Bihar School Examination Board (BSEB) <i>Class XII — Science Stream 75.6%</i>	<i>Patna, Bihar</i> <i>2022</i>
Central Board of Secondary Education (CBSE) <i>Class X 88%</i>	<i>India</i> <i>2020</i>