Ajay K Pai

Data Analyst

Mangalore, Karnataka, India | ajaykpai895@gmail.com | +91 79755 45779
linkedin.com/in/ajay-k-pai-815a10365 | github.com/ajaykpai895

Professional Summary

Entry-level Data Analyst with a fintech internship at Bluestock and three projects using Python, SQL, and Power BI. Built a customer churn model with 84% ROC-AUC that flagged over $1.2M in yearly revenue at risk, and a fraud model that cut test-set fraud losses by 65.6%.

Experience

Data Analyst Intern April 2026 to May 2026
Bluestock Fintech *India*

- Analyzed financial and stock market data to support business decisions. Cleaned the data with Python and Excel and built dashboards to track key metrics.
- Built automated pipelines to bring in financial data in real time, which improved data accuracy.
- Worked with product managers to set KPIs and plan A/B tests for new features.
- Analyzed historical stock data to forecast market trends.

Skills

Programming & Analytics: Python (Pandas, NumPy, Scikit-learn), SQL (MySQL, SQLite), C Programming (Basics), Data Cleaning, EDA, KPI Definition

Data Visualization & BI: Tableau, Power BI, Matplotlib, Seaborn, Microsoft Excel (Pivot Tables, VLOOKUP, Power Query, Macros)

Machine Learning Basics: Scikit-learn, Logistic Regression, Model Evaluation

Projects

TelcoChurn360: Churn Prediction & Retention Dashboard
Python, SQL, Scikit-learn, Power BI github.com/ajaykpai895/TelcoChurn360-End-to-End-Churn-Prediction-Retention-Dashboard

- Analyzed 7,043 customer records (21 features) with SQL and Python to find why customers leave.
- Built a logistic regression model with an ROC-AUC of 84%, 6 points better than a SQL rule-based baseline.
- Estimated over $1.2M in yearly revenue at risk and built a Power BI dashboard for retention decisions.

Credit Card Fraud Detection
Python, Pandas, Scikit-learn, XGBoost, Streamlit github.com/ajaykpai895/fraud-detection

- Built a fraud detection model on 284,807 credit card transactions (0.17% fraud) and compared Logistic Regression, Random Forest, and XGBoost.
- Split the data by time (train on older transactions, test on newer ones) and judged models by money lost to fraud, not accuracy.
- Tuned the alert cutoff to cut test-set fraud losses by 65.6% ($6,168.88 to $2,123.10) versus flagging nothing.
- Built a Streamlit dashboard to review flagged transactions and wrote 27 unit tests.

Heart Disease Risk Scoring
Python, Pandas, Scikit-learn, XGBoost, Streamlit github.com/ajaykpai895/heart-disease-risk-scoring

- Built a heart disease risk model on 303 patient records. Logistic Regression and XGBoost scored almost the same (ROC-AUC 0.906 vs. 0.894).
- Lowered the risk cutoff so no patient with heart disease was missed in the test set (0 missed vs. 7 at the default cutoff).
- Checked results for male and female patients. There were few female cases, so those results are less reliable.
- Built a Streamlit app and wrote 54 automated tests.

Education

Bachelor of Engineering (B.E.) in Computer Science & Design
Canara Engineering College, Mangalore, Karnataka, India
Expected Graduation: 2027 | CGPA: 7.9/10 (current)

Certifications

Data Science Essentials with Python, Cisco Networking Academy (2026)
Data Analytics Essentials, Cisco Networking Academy (2026)
Go Beyond the Numbers: Translate Data into Insights, Google on Coursera (2026)