

PUJA PATIDAR

DATA ANALYST | DATA SCIENCE | MACHINE LEARNING

PYTHON | SQL | POWER BI | TABLEAU | EXCEL

Jersey City, NJ 07306 | pujapatidar46@gmail.com | (551) 260-1044 | linkedin.com/in/puja-patidar | github.com/patidarpuja

Authorized to Work in the U.S. | No Sponsorship Required

SUMMARY

Data analyst with an engineering background who turns large healthcare, pharma, and financial datasets into clear, data-driven decisions using Python, SQL, Excel, Tableau, and Power BI. Built XGBoost models on up to 1.85M records and a fully local RAG system that raised answer accuracy to 88% and cut latency 71%. Led 2 teams to 1st place in hackathons.

CORE SKILLS

Programming and SQL: Python, SQL, PostgreSQL, Pandas, NumPy, CTEs, Window Functions, Query Optimization

Reporting and BI: Excel, Tableau, Power BI, DAX, Power Query, KPI Dashboards, Data Modeling, Data Storytelling

Problem Solving and Data Quality: Root Cause Analysis, Data Profiling, Data Cleansing, Validation, Documentation

Machine Learning: Scikit-learn, XGBoost, Random Forest, Linear Regression, Feature Engineering, Model Evaluation, SMOTE

Analysis and Tools: Exploratory Data Analysis, Statistical Analysis, Cohort Analysis, Matplotlib, Seaborn, Git, GitHub, Jupyter

Artificial intelligence (AI): RAG, Mistral 7B, FAISS, BM25, Tesseract OCR, Prompt Engineering, Gradio, Claude Code, Codex

PROFESSIONAL EXPERIENCE

Data Science Extern | Remote, U.S. | July 2026 to Present

- Developed a fully local RAG system for pharma compliance documents using Tesseract OCR, LLM-based document splitting, BM25 retrieval benchmarked against FAISS, and Mistral 7B, with rule-based certificate expiry checks in Python.
- Ran controlled train/dev/test experiments (51 questions) on chunking, retrieval, and top_k, lifting accuracy from 82% to 88% and MRR from 0.845 to 0.917 while cutting latency 71% (6.1 s to 1.8 s).

Data Analyst, NumpyNinja | Remote, U.S. | July 2025 to Present

- Led a 5-member team to clean and model a clinical diabetes dataset in Power BI, designing a star schema for reporting.
- Transformed raw data in Power Query with 27 steps, resolving missing values, data types, and invalid entries.
- Created 20+ DAX measures for a KPI dashboard, revealing a 3x rise in kidney-marker elevation across visits.
- Built an interactive Tableau dashboard on 1.5M+ ICU patient-hour records, flagging SIRS risk across 40+ variables.
- Segmented patients by age and gender with heatmaps, uncovering high-risk biomarker trends in ICU data.
- Won 1st place in 2 hackathons as team lead with a Random Forest model (0.90 R²) and a 7-table PostgreSQL database.

Career Development and U.S. Relocation | August 2022 to May 2025

- Relocated to the U.S. and upskilled in SQL, Python, Tableau, and machine learning through applied projects.

Engineer, Engineering and Application Support, Sterling E-Mobility | January 2022 to July 2022

- Analyzed performance data from 100+ EV control boards, identifying quality patterns to support engineering decisions.
- Investigated 50+ field returns through root cause analysis, documenting failure patterns for corrective action.

Trainee, System Design, Philips India | September 2020 to July 2021

- Performed power-budget and derating analysis across 500+ components to support system feasibility and design reviews.

PROJECTS

Springboard Data Science Career Track | Completed April 2026

- Built a time-aware XGBoost sales-prediction model on 420K+ weekly records (20+ engineered features), reaching 0.88 R² validated with TimeSeries Split, to support inventory and demand-planning decisions.
- Developed an XGBoost credit card fraud detection model on 1.85M transactions (0.5% fraud rate) using stratified sampling and SMOTE to reach 0.784 PR-AUC, improving rare-event detection over baseline.

EDUCATION

Master of Technology in Power System | National Institute of Technology, Kurukshetra, India | 2019 to 2021

Bachelor of Engineering in Electrical Engineering | Government Rewa Engineering College (RGPV), India | 2014 to 2018

CERTIFICATIONS

- [Data Science Career Track Certificate, Springboard \(2026\)](#) • [Google Data Analytics Certificate, Google \(2024\)](#)