

BHARGAV PEDDI

Data Scientist | Machine Learning | Data Analytics

Jersey City, NJ | (551) 655-6725 | bhargavpeddi25@gmail.com | linkedin.com/in/bhargavpeddi

PROFESSIONAL SUMMARY

Data Scientist and Data Analyst with 4+ years of experience across healthcare analytics, predictive modeling, and machine learning. Proven track record building and deploying ML models (classification, regression, clustering) using Python and Scikit-learn, engineering production-ready data pipelines, and translating complex healthcare datasets into measurable business outcomes, including a 12% lift in retention-targeting accuracy and 85%+ predictive model accuracy on capacity-planning forecasts. Skilled in SQL, statistical analysis, feature engineering, and BI dashboarding (Power BI, Tableau), with hands-on experience integrating Generative AI and LLM-based workflows (RAG, LangChain, OpenAI/Anthropic APIs). Currently completing an MS in Information Systems at NJIT with a concentration in Data Science and AI-driven analytics.

CORE TECHNICAL SKILLS

Languages: Python, SQL, R, Java, C++, C, SAS

Machine Learning: Scikit-learn, TensorFlow, XGBoost, Classification, Regression, Clustering, Feature Engineering, Hyperparameter Tuning, Cross-Validation, SHAP / Explainable AI

Generative AI: OpenAI & Anthropic (Claude) APIs, Retrieval-Augmented Generation (RAG), LangChain, Prompt Engineering, n8n Workflow Automation

Data & Visualization: Power BI, Tableau, Advanced Excel, EDA, Statistical Analysis, Forecasting, Data Wrangling & Cleaning

Databases & Cloud: MySQL, PostgreSQL, MongoDB, SQL Server, Oracle, AWS, ETL Pipelines, Data Warehousing, SSIS/SSRS

Domain & Tools: Healthcare Claims & ADT Data, HIPAA, ICD-10, Git/GitHub, Jira, Agile/Scrum

PROFESSIONAL EXPERIENCE

Data Scientist / AI-ML Analyst | UnitedHealthcare, Remote

Jan 2026 – Present

- Analyzed 500K+ member engagement records (claims, appointments, portal usage) using SQL and Python, uncovering behavioral trends that informed outreach strategy across 3 regional campaigns.
- Built machine learning models in Python and Scikit-learn to predict customer engagement and appointment adherence, improving retention-targeting accuracy by 12%.
- Delivered 6+ interactive Power BI dashboards and automated reporting workflows to monitor engagement KPIs, cutting manual reporting time by 40%.
- Improved reporting accuracy by 20% across recurring analytics cycles through structured data cleaning, preprocessing, and validation.
- Partnered with a cross-functional team of 8+ analysts and business stakeholders to translate model outputs into actionable engagement strategies.

Operations Data Analyst | NJIT, Newark, NJ

Dec 2025 – Present

- Engineered advanced SQL queries (CTEs, window functions, joins, subqueries) across 1M+ operational records, improving data extraction efficiency by 30%.
- Built automated Python/Pandas ETL pipelines to ingest, cleanse, and reconcile data from SQL, Excel, and CSV sources, cutting manual prep time by 40% and improving accuracy by 25%.
- Developed Power BI dashboards using DAX and Power Query to track 20+ operational KPIs, reducing recurring reporting time by 35%.
- Applied Python and Scikit-learn to build forecasting and predictive analytics models for demand and resource planning, achieving 85%+ predictive accuracy.
- Automated data-quality validation and exception checks, eliminating 30+ hours of monthly manual effort and improving reporting turnaround by 45%.

Aug 2023 - Dec 2024

- Data Analyst | Optum, Hyderabad, India**

Aug 2022 – Jul 2023

- ## PROJECTS

- Built an XGBoost classification pipeline on 7,000+ telecom records achieving 80% accuracy and 0.85 AUC-ROC.
- Applied SMOTE oversampling to correct class imbalance, raising recall on the minority churn class from 52% to 61%.
- Performed feature engineering and selection across usage, billing, and contract variables to improve model signal.
- Used SHAP value analysis to explain top churn drivers (contract type, tenure, internet type) to non-technical stakeholders.
- Tuned hyperparameters via cross-validation to optimize model generalization and reduce overfitting.

- Analyzed 70K+ hospital discharge records using SQL-based EDA to identify factors driving 30-day readmissions.
- Engineered patient-level risk features (diagnosis codes, length of stay, prior visit frequency) to improve model input quality.
- Built an interactive Power BI dashboard tracking readmission rates and department-level risk trends for clinical stakeholders.
- Performed data cleaning and validation across multiple hospital data sources to ensure reporting consistency.
- Segmented patients by risk category to support targeted follow-up and care-coordination planning.

- Designed an end-to-end ETL pipeline transforming raw, multi-source datasets into analytics-ready data.
- Automated data ingestion and transformation workflows using Python and SQL, reducing manual processing effort.
- Implemented automated data validation and anomaly detection at each pipeline stage to catch missing/out-of-range values.
- Built BI dashboards in Power BI and Tableau tracking key business metrics with built-in data quality checks.
- Ensured pipeline reliability and reporting accuracy through scheduled validation and exception handling.

EDUCATION

Jan 2025 – Dec 2026

B.Tech, Information Technology: Anurag University, Hyderabad, India

Aug 2021 - May 2024

CERTIFICATIONS

- IBM Data Science Professional Certificate: Python, Machine Learning, SQL, Data Analysis
- Google Data Analytics Professional Certificate: Data Analysis, SQL, Tableau, R
- Tableau Certified Data Analyst: Data Visualization, Dashboard Development
- Oracle Certified Database Programmer with SQL: Query Optimization, Relational Databases
- AWS Academy Cloud Foundations: Cloud Computing, AWS
- Alteryx Designer Core Certification: Data Preparation, Data Blending, Automated Workflow Design