JOSHUA PIÑA

Atlanta, GA | Columbus, GA | 954-470-1449
joshuadariuspina@gmail.com | linkedin.com/in/joshuadpina
github.com/joshuapina

Professional Summary

Data Scientist with Secret Clearance and 8+ years of leadership in military healthcare and government program management. Proven expertise across the machine learning lifecycle, from data engineering and feature development through model training (CNN, U-Net, ensemble methods) to production deployment (FastAPI, Docker, MLOps). Successfully managed data systems supporting military logistics and enhancing program efficacy through data-driven insights. Recent B.S. in Data Science; currently exploring graduate opportunities at Georgia State and Georgia Tech.

Technical Skills

Programming: Python (Pandas, NumPy, SciPy), R, SQL, JavaScript
ML Frameworks: PyTorch, scikit-learn, XGBoost, CatBoost, Transformers, WandB, CNN, U-Net
Data Engineering/Geo: PostgreSQL, Supabase, MongoDB, PostGIS, Geopandas, Vector Embeddings
Deployment/Tools: Docker, FastAPI, Streamlit, Git, GitHub Projects, Cookiecutter, RLS
Cloud/Storage: AWS (S3, EC2), Azure, OAuth

Projects

Spatiotemporal Forecasting of Burglary Risk in Atlanta   Aug – Dec 2025

- Led a 4-person data science capstone team through the entire data science lifecycle, utilizing GitHub Projects for coordination with robust data lineage and Cookiecutter Data Science structure for project reproducibility
- Developed a predictive modeling pipeline by integrating Atlanta Police Department data with external sources, including Census data, Open-Meteo weather, and Atlanta/Georgia shapefiles
- Performed extensive data engineering, creating distance-based features and utilizing Weights & Biases (WandB) for comprehensive experiment tracking
- Containerized the ML model and dashboard with Docker for reproducible deployment

The Living Library: Full-Stack Research App   Aug – Dec 2025

- Built a comprehensive study tool leveraging FastAPI, PostgreSQL (Supabase), and vector embeddings for semantic search across 55+ textbooks
- Architected database with OAuth and Row Level Security (RLS) to manage complex user authentication and secure multi-user data access

Machine Learning Study on Head & Neck Squamous Cell Carcinoma   Jun – Aug 2025

- Led a multi-modal machine learning capstone utilizing logistic regression, CNN, and U-Net for classification/segmentation of unlabeled axial CT scans
- Implemented runtime optimization, reducing the dataset from 6GB to < 500MB (90%+ compression), significantly improving training efficiency
- Preprocessed and augmented 1,500+ CT scan images using data normalization, windowing techniques, and geometric transformations while addressing class imbalance through weighted loss functions to improve model generalization

Ticket Heroes: Price Optimization Analysis   Aug – Dec 2025

- Engineered a comprehensive dataset through web scraping, aggregation, and advanced data augmentation to overcome data sparsity
- Designed a FastAPI backend with a Streamlit dashboard to analyze optimal ticket purchasing timing across event categories

Everything is a Graph: Edmonds-Karp Segmentation   Aug – Nov 2025

- Implemented graph-based image segmentation through Edmonds-Karp and network flow
- Added exponential decay to create a use-case to compare against deep learning techniques

Mining Flight Delays for Tracking and Prediction   Aug – Nov 2025

- Executed an end-to-end machine learning project predicting domestic flight delays, utilizing feature engineering and ensemble classification methods

Experience

Program Manager, Public Contracting Institute   May 2022 – Present

- Analyzed attendance, registration, and post-training survey data from 200+ webinars to identify top-performing programs and inform future offerings
- Designed and maintained structured workflows to support large-scale training operations
- Cleaned, organized, and validated participant data to ensure accuracy across certification and reporting
- Improved website UX/UI and SEO to increase engagement and training enrollment

Healthcare Specialist, United States Army   Jan 2017 – Apr 2022

- Managed medical records for 5,000+ personnel in Army EHR systems
- Supported command-level logistics and operations through accurate reporting and coordination
- Oversaw $100K+ in medical supplies and equipment

Education

Bachelor of Science in Data Science, Georgia State University   December 2025
Relevant Coursework: Machine Learning, Data Mining, Big Data Programming

AWS Solutions Architect – Associate   Expected January 2026

IBM Data Analyst Professional Certificate   August 2025

CodePath Intermediate Technical Interview Prep (Advanced Track)   July 2025

Leadership & Professional Development

CodePath – Technical Mentor   Aug – Dec 2025

- Mentored two students in data structures and technical problem-solving through weekly sessions

FourBlock Professional Development Cohort   Sept – Nov 2025

- Completed FourBlock's competitive leadership and career-readiness program