Richard Pallangyo

Seattle, WA — rapaugustino@gmail.com — (206) 665-6484 — linkedin.com/in/rapaugustino
Portfolio: datacamp.com/portfolio/rpallangyo — GitHub: github.com/rapaugustino

SKILLS

Technical Skills

- **Programming & Analytics:** Python (Pandas, NumPy), SQL, PostgreSQL, PySpark, R, Statistics (Regression, Hypothesis Testing), Machine Learning, Data Visualization (Plotly, Streamlit)
- **AI & Agents:** Retrieval-Augmented Generation (RAG), LangGraph, Google Gemini, LLMs, Prompt Engineering, Retrieval/Indexing, Human-in-the-Loop (HITL), OpenAI Batch API
- **Cloud & Data Engineering:** AWS (Glue, Athena, S3, Redshift, CloudFormation), Azure (Data Factory, Logic Apps, AI Search, APIM, Entra ID, App Service, ADLS), Redis, Docker
- **Web & Ops:** FastAPI, Next.js, React/TypeScript, REST APIs, GitHub Actions, CI/CD, Railway

Soft Skills

Leadership, Team Management, Technical Writing, Research Protocol Design, Public Speaking, Stakeholder Management

PROFESSIONAL EXPERIENCE

University of Washington (CBE-IT) — Seattle, WA
Student Consultant — Jan 2025 – Present

- **Provide** technical consultation and Tier 1 support for faculty, staff, and students within the College of Built Environments, troubleshooting hardware, software, and network issues across Windows and macOS environments
- **Manage** wide-format plotting operations and equipment circulation, ensuring high availability of critical design resources for architecture and planning departments
- **Assist** in IT infrastructure tasks including disk imaging, device deployment, and audiovisual setup for events, maintaining operational readiness of computer labs

SSA Marine — Procurement Department — Seattle, WA
Data Science Intern — Jun 2025 – Sep 2025

- **Authored** an internal **indexing guide** and prompt library to document best practices for data extraction and support future maintenance
- **Designed** the end-to-end architecture for a production-grade **AI Procurement Agent** (Python/FastAPI, LangGraph) focused on policy Q&A and document processing
- **Built** a **document-extraction batch pipeline** (OpenAI Batch) with job tracking and idempotent retries to process unstructured files at scale
- **Implemented** a retrieval & indexing pipeline in **Azure AI Search** with principled chunking and semantic/vector configs to improve answer relevance and grounding
- **Built** a human-in-the-loop approval system (Azure Logic Apps) for AI-drafted messages; added redacted logging and an auditable status trail to protect sensitive data
- **Deployed** the backend to **Azure App Service** behind **APIM** and **Entra ID SSO** to enforce enterprise identity and network boundaries
- **Delivered admin-grade observability**—health checks, metrics, job lists, and LangGraph node traces—to enable operational insight and auditability
- **Developed** a **React/TypeScript** admin UI and ops dashboard that surfaces metrics, batch queues, and tracking downloads (CSV/JSON)
- **Hardened** reliability with CI checks (lint, security audit), APIM documentation, and runtime data hygiene

LTIMindtree Limited   Seattle, WA
Senior Data Engineer   Jun 2022 – Aug 2024

- **Modernized** enterprise analytics stacks on **Azure** across ADF, Databricks, and warehouse patterns to accelerate BI and ML use cases
- **Led** a **30-person** Data & AI team and **reduced** client issue resolution time by **30%** through process rigor and coaching
- **Delivered** internal training on Azure Data & AI and **channeled** field feedback to Microsoft product teams for roadmap impact
- **Scaled** the practice from **0 → 30** engineers and **promoted 4** team members into leadership roles via targeted development plans
- **Resolved** mission-critical escalations in ADF, SSIS, and streaming pipelines to restore service reliability under tight SLAs
- **Streamlined** hiring and onboarding operations to **cut** ramp-up time by **20%** for new data engineers

---

LTIMindtree Limited   Seattle, WA
Junior Data Engineer   Oct 2021 – Jun 2022

- **Advised** Microsoft stakeholders on **cloud data architecture** best practices across ingestion, transformation, and governance
- **Diagnosed** and **resolved** production incidents in ADF, SSIS, and Azure Stream Analytics following defined incident protocols

EDUCATION

University of Washington (UW) — Master of Science in Data Science   Seattle, WA
Expected March 2026   **GPA: 3.90**

Seattle University — B.A. in Computer Science; Minors: Mathematics & Data Science   Seattle, WA
Jun 2019 – Jun 2021
Honors & Awards: Cum Laude; Dean's List (4x); Tri-Alpha First-Gen Honor Society; Alfie Scholar; Xavier Merit Award

Seattle Central College — A.S. in Computer Science   Seattle, WA
Jun 2017 – Mar 2019
Honors & Awards: Dean's List (5x); Phi Theta Kappa Honor Society

PROJECTS

PawMatch – AI-Powered Dog Breed Recommender   FastAPI, PostgreSQL, Gemini 2.0, Next.js
Award: Ranked 2nd out of 516 in DataCamp Competition
Live App — GitHub: Backend — GitHub: Frontend

- **Architected** a high-performance full-stack application separating concerns between a **Next.js** frontend and **FastAPI** backend, utilizing **PostgreSQL** for persistent data and **Redis** for session caching.
- **Engineered** a custom conversational agent with **Google Gemini 2.0 Flash** that parses natural language into structured query parameters to drive a weighted scoring algorithm (17 traits) for personalized recommendations.
- **Built** interactive discovery features including a "Browse" interface with faceted filtering, a "Compare" module with side-by-side trait visualization (Radar Charts), and a "Favorites" system for authenticated users.
- **Managed** a complex data pipeline using **SQLAlchemy** and **Pydantic** for robust data validation across 195 breeds, serving 6,800+ optimized assets via Google Cloud Storage.

CIT Impact Analysis: Police Use of Force   Python, Statsmodels, Plotly
Research Project: The Interventionist Effect — GitHub: rapaugustino/cit-impact

- **Analyzed** 104,000+ Seattle Police crisis contacts (2015-2025) to evaluate the efficacy of Crisis Intervention Team (CIT) training, cleaning complex administrative data to handle negative experience values and categorize risk types.

- **Modeled** predictors of force using **Binary Logistic Regression**, identifying that CIT officers were 13% more likely to use force (OR 1.13, p=0.027) when controlling for call severity and precinct, challenging the assumption that training reduces force.
- **Applied Multinomial Logistic Regression** to disposition outcomes, discovering an "Interventionist Effect": CIT officers were 27% more likely to refer subjects to services but also 45% more likely to arrest compared to taking "No Action."
- **Visualized** statistical findings using Plotly forest plots and grouped bar charts to communicate odds ratios and confidence intervals to non-technical stakeholders.
- **Addressed** ethical considerations regarding selection bias and population bias, ensuring nuanced interpretation of sensitive behavioral health data.

E-commerce Analytics Data Pipeline   AWS Glue, Athena, PySpark
GitHub: rapaugustino/ecommerce-analytics-pipeline

- **Designed** a production-grade Medallion Architecture (Bronze/Silver/Gold) to ingest and process 750,000 events every 5 minutes using AWS Glue and S3
- **Implemented** defensive data quality checks to filter out records with missing timestamps or malformed IDs, ensuring high data fidelity for downstream analysis
- **Optimized** query performance by converting raw JSONL to Snappy-compressed Parquet and implementing pre-aggregated Gold tables, achieving a **16.6x reduction in query latency**
- **Defined** infrastructure as code using CloudFormation to ensure reproducibility and version control

Fantasy Premier League (FPL) Decision Dashboard   Python, Streamlit, Plotly
Live App: fpl511.streamlit.app — GitHub: rrichards7/data511

- **Co-developed** an interactive dashboard for lineup planning and formation experimentation to support weekly decisions
- **Built** a player comparison module that visualizes trends in expected goals (xG), assists, minutes, and price changes
- **Engineered** modular data loaders and transformation scripts to keep visuals fast and maintainable
- **Deployed** the app on Streamlit Cloud with publicly accessible links for user feedback

World Cup 2026 Predictions   Python, Streamlit, Jupyter
Live App: 2026 World Cup Predictions — GitHub: KPfeil25/world-cup-26-predictions

- **Co-built** a Streamlit web app for interactive **team/player analytics** and **match-outcome predictions**
- **Integrated** multi-source data (Fjelstul World Cup DB, NOAA climate averages) to enrich model features
- **Implemented** a "what-if" page that lets users vary team, stadium, and temperature to see predicted outcomes

Safari King Africa Website   HTML, CSS, JavaScript
www.safarikingafrica.com

- **Delivered** a responsive, brand-aligned website from concept through deployment for an international safari operator
- **Designed** the UI system and **integrated** secure contact forms with automated acknowledgements