

## Angel Alberto Merino Cedeño

Hermosillo, Sonora, México | [linkedin.com/in/anmerino-pnd](https://www.linkedin.com/in/anmerino-pnd) | +526623351465 |  
[acedeno00@gmail.com](mailto:acedeno00@gmail.com) | [anmerino-pnd.github.io/projects-hub/](https://anmerino-pnd.github.io/projects-hub/) |  
<https://medium.com/@anmerino-pnd>

### EDUCATION

---

Universidad de Sonora

Hermosillo

*Master's Student in Data Science*

*Aug 2023 - Present*

- Specialization in Natural Language Processing (NLP), Large Language Models (LLMs), Neural Networks, and Time Series Analysis.
- Proficient in **Python, SQL, and NoSQL databases**.
- Experience with **ETL pipelines, data visualization, and data wrangling** using Pandas, NumPy, Seaborn, and Matplotlib.
- Skilled in **version control** with **Git and GitHub**, and project collaboration using **DagsHub**.

Universidad de Sonora

Hermosillo

*Bachelor's in Chemical Engineering*

*Aug 2018 – June 2023*

- Skilled in ASPEN HYSYS and MATLAB for process simulation and optimization.

### WORK EXPERIENCE

---

CT Internacional

Hermosillo

*Data Scientist*

*Feb 2025 – Present*

- **Engineered and deployed a production-grade multi-agent chatbot** using LangGraph and VertexAI (Gemini 3.0), transitioning from previous OpenAI models and LangChain framework. This empowered users and the sales team across all branches to **independently resolve complex product and warranty queries in under 20 seconds**, eliminating the hour-long wait for IT support.
- **Designed a multi-tool agent architecture** with a dynamic RAG pipeline processing structured (SQL, XML) and semi-structured (PDF) data, integrating product search, real-time promotion validation, order status tracking, branch directory lookups, and policy/warranty retrieval — enabling end-to-end sales and customer support through a single conversational interface.
- **Designed an internal monitoring stack** combining a RAGAS-inspired LLM-as-judge evaluation system (faithfulness, answer relevancy, context precision and recall) with a BI reporting layer tracking assistant usage, user feedback, and cost metrics — **enabling data-driven model optimization** and early hallucination detection in production.
- **Built a CI/CD pipeline with GitHub Actions and Podman** that automatically runs the test suite and produces a reproducible deployment artifact for the LangGraph agent runtime on every push to main, eliminating manual deployment steps and catching integration bugs before they reach the live chatbot. **Cut deploy verification from ~30 minutes of manual smoke-tests to a single automated pipeline run.**
- **Agentic product recommendation chatbot:** a personalized approach ([https://anmerino-pnd.github.io/projects-hub/2\\_ct\\_chatbot/0\\_home.html](https://anmerino-pnd.github.io/projects-hub/2_ct_chatbot/0_home.html)).
- **Built and deployed a modular customer analytics pipeline** that automates RFM segmentation (K-Means k=5), market basket analysis (FP-Growth across 5 customer segments producing 3,800+ rules), and temporal pattern detection over 780K historical orders.
- **Reduced exploratory notebooks to a single 45-second daily ETL job** scheduled via cron, with frozen scikit-learn model artifacts, quality checks, and incremental MongoDB ingest — eliminating manual analysis cycles for the marketing team.
- **Developed an internal analytics dashboard** (FastAPI + DuckDB + Plotly) exposing 7 interactive views over segmentation outputs — including a Market Basket Opportunity Map, at-risk customer alerts, and per-segment cross-sell

recommendations — giving marketing self-service access to insights that previously required ad-hoc analyst requests.

- **Deployed the analytics platform to an on-premise server** with systemd service supervision (auto-restart, boot-resilient), firewall configuration, scheduled cron-based pipeline runs, and full reproducibility via uv and version-pinned dependencies — making the system operational without ongoing developer intervention.
- Pulse: a production customer analytics platform ([https://anmerino-pnd.github.io/projects-hub/5\\_ct\\_dashboard/0\\_home.html](https://anmerino-pnd.github.io/projects-hub/5_ct_dashboard/0_home.html)).

## Centro Nacional de Control de Energía (CENACE)

Freelance Data Scientist

Hermosillo

May 2024 – Feb 2026

- **Designed a secure, on-premise RAG system** (FastAPI, Ollama, FAISS) to optimize the IT help desk's incident resolution workflows across 2,800 historical tickets and dense manuals, achieving an average query response time of 24.4 seconds and delivering a solution rated 4.4/5 in technical precision by engineering experts.
- **Reduced manual search times and improved retrieval accuracy** by designing an automated feedback loop mechanism that re-ingests high-rated bot responses into the vector store.
- **Streamlined triage workflows** by building an ML-based incident classification system to automatically suggest accurate ticket labels.
- **Managed and optimized end-to-end Python ETL pipelines** to extract and transform structured (SQL) and unstructured (PDFs, MongoDB) data, successfully supporting RAG model training and real-time inference.
- Development of a Help Desk based on a large language model ([https://anmerino-pnd.github.io/projects-hub/1\\_cenace\\_helpdesk/0\\_home.html](https://anmerino-pnd.github.io/projects-hub/1_cenace_helpdesk/0_home.html)).

## DARIG – Consultores

Freelance Data Scientist

Hermosillo

March 2025 – Feb 2026

Client: Salsas Castillo

- **Automated the generation of reports and data visualizations** for the executive board by developing an internal AI assistant with OpenAI function-calling and SQL toolkits, **enabling faster, data-driven decisions via natural language**, reducing manual data extraction time from hours to real-time minutes for the executive board.
- **Eliminated the need for manual reporting** by securely integrating the assistant with Telegram, allowing executives to asynchronously query business KPIs on the go.
- Led the architectural transition from a standard RAG prototype to a **Model Context Protocol (MCP)** framework, enabling the assistant to autonomously execute SQL queries to extract real-time insights, significantly reducing analysis time and improving long-term maintainability.
- GlorIA: Business analysis from an AI-driven approach ([https://anmerino-pnd.github.io/projects-hub/3\\_salsas\\_chatbot/0\\_home.html](https://anmerino-pnd.github.io/projects-hub/3_salsas_chatbot/0_home.html)).

Freelance Data Engineer

Client: Frutal

October 2025 – Feb 2026

- **Resolved critical Stack Smashing memory errors** caused by unstable legacy ODBC drivers by designing and containerizing a cross-platform ETL pipeline (**Podman**). Successfully extracted and transformed 6 production tables *from a legacy HFSQL database to Parquet format*, loading them into **Google Cloud Platform (GCP)** for executive dashboard consumption.
- **Ensured consistent pipeline execution** across Windows development and headless Linux production environments by leveraging containerization to resolve complex library conflicts (iODBC vs unixODBC).
- Data engineer to create a data lake from HFSQL database ([https://anmerino-pnd.github.io/projects-hub/4\\_frutal/0\\_home.html](https://anmerino-pnd.github.io/projects-hub/4_frutal/0_home.html)).

## **PUBLICATIONS**

---

**Medium Publications:** Authored guides focusing on Data Engineering, LLMs, and API integrations:

- [I Built a Telegram Bot That Tells Me When My Stocks Are “Cheap”](#)
- [Turned Telegram Into a Local AI Voice Studio](#)
- [Ditch the Cloud: Build a Free, Local AI Coding Agent with llama.cpp](#)
- [Build a Telegram-to-Notion Expense Tracker with Python and LLMs](#)
- [Extracting Enterprise Data: From HFSQL to Parquet with Python on Windows](#)
- [Pocket AI: Building a Telegram Bot with Python & Ollama](#)
- [The Missing Manual: Connecting Python to HFSQL on Linux \(Podman Edition\)](#)
- [From Data to Beats: An Easy Guide to Spotify’s API](#)

## **SKILLS & INTERESTS**

---

**Skills:** Proficient in programming with Python, SQL, and NoSQL. Specialized in GenAI (Vertex AI, Gemini, LangChain/LangGraph, RAG, Agents, and Model Context Protocol (MCP) systems), with foundational knowledge of Google Cloud Platform (GCP). Certified in Data Analysis, Machine Learning Scientist in Python and Data Science with Python by DataCamp. Experienced using Microsoft Office. Fluent in English, Portuguese, and French; intermediate proficiency in Japanese and German, and a native Spanish speaker.

**Interests:** Enthusiastic about learning new languages and exploring diverse cultures. I enjoy dancing and travelling. Interested in data-driven problem solving and the application of technology to innovative projects.