

Ernesto Diaz

347-757-2325 | ernestoabel.diaz@gmail.com | [LinkedIn](#) | [Portfolio](#)

EDUCATION

BS, Business Analytics and Information Systems

Aug 2022 – May 2026

University of South Florida, Tampa, FL

- Study Abroad in Finance, Florence University of the Arts
- Bright Futures Academic Scholarship

EXPERIENCE

AI & Machine Learning Research Assistant | *Blue Cross Blue Shield*

Oct 2025 – Present

- Fine-tuned Med42-8B LLM using a **QLoRA** pipeline in **Python**, reducing training loss from **2.49** → **0.09** across **15,000+** clinical prompt-response pairs to improve model performance for oncology decision support.
- Engineered training dataset of **5,200+** synthetic TCGA-LUAD-style patient records using GDC/TCGA data standards to address the absence of public lung cancer training data, covering NCCN guideline adherence and mutation-driven reasoning.
- Identified and resolved a data quality gap in JSONL generation, redesigning pipeline logic to increase clinical citation coverage to **75%+** and significantly improve model response reliability.
- Deployed and evaluated **Med42 LLM**, achieving **72%** accuracy on USMLE benchmarks, outperforming GPT-3.5 by **~11 points**.
- Built **SQL**-based data pipelines to analyze **10,000+** lung cancer EHR records to surface patterns in patient survival.

Data Analytics Tutor | *Knack*

May 2024 – Present

- Created individualized learning plans for **20+** students with data-driven insights, resulting in a **25% average improvement** in scores.
- Earned consistent **5-star ratings** across sessions, simplifying complex concepts in Excel and Statistics into accessible outlines.
- Leveraged **Tableau** and **Excel** to provide visual, data-driven feedback that accelerated student comprehension and performance.

Data Analytics Intern | *Colegio Yannette*

Apr 2020 – Jun 2020

- Managed the registration and updates of critical data for **200+ student records** in a remote database system, ensuring data integrity.
- Reviewed invoices and payments, identifying opportunities for process improvements and enhancing data clarity and efficiency.
- Organized student record management through systematic **Excel** data entry, reducing errors and improving operational reporting.

PROJECTS

Oncology AI Clinical Decision Support System | Python, PyTorch, QLoRA/LoRA, Hugging Face, Hex, GCS | [Site](#)

- Fine-tuned Med42-8B on a synthetic TCGA-LUAD lung cancer dataset using **QLoRA** and 4-bit quantization, reducing training loss from **2.49** → **0.09** over **15,000+** clinical prompt/response pairs spanning treatment recommendations and mutation reasoning.
- Engineered a synthetic lung cancer dataset of **5,200+** patient records modeled after GDC/TCGA standards, generating **30,000+** instruction-tuned training pairs on treatment recommendations, mutation-driven reasoning, and NCCN guideline adherence.
- Built end-to-end QLoRA fine-tuning pipeline using bitsandbytes, PEFT, and TRL with LoRA adapter training and versioning.
- Designed a data analysis pipeline in Hex: Kaplan-Meier curves, cohort distribution dashboards, and data quality audits.

S&P 500 Market Regime & Sector Rotation Analysis | Python, Pandas, NumPy, Matplotlib, Seaborn, yfinance | [Site](#)

- Engineered **15+** technical features across **2.9M+** daily prices, enabling multi-horizon return forecasting and sector analysis.
- Designed a market regime classification system to produce **9** distinct market states, surfacing outperforming sectors historically.
- Built market-level indicators, including % of stocks above SMA-50/200 & advance/decline ratios, providing macro-level insights.

Hospital Operations & Patient Flow Analytics | PostgreSQL, Python, Tableau, Google Cloud (GCP) | [Link](#)

- Analyzed **1000+** appointment records to identify high-traffic departments handling **~30%** more appointments than average.
- Designed a 7-table relational database with patients, doctors, departments, and appointments using **SQL** and **Google Cloud**.
- Developed SQL-driven **KPIs** and **Tableau** visualizations aimed at reducing wait times and improving hospital management.

Mobile Expense Tracker Application | React Native, SQLite, JavaScript, Git | [Site](#)

- Built a mobile expense tracking app to manage **100+** records with full **CRUD** functionality using **React Native** and **SQLite**.
- Designed **SQL** queries to aggregate daily spending totals and power a time-series chart visualization across selectable ranges.

SKILLS

- **Languages:** Python, SQL (PostgreSQL, MySQL, Oracle), Spanish (fluent)
- **Data & ML:** pandas, NumPy, Matplotlib, Seaborn, scikit-learn, PyTorch
- **Cloud Platforms:** AWS, Azure, Google Cloud Platform (GCP), Snowflake, MongoDB
- **Analytics & BI Tools:** Tableau, Power BI, Excel, Hex, dbt, Git, Shell

INVOLVEMENT & VOLUNTEERING

Member | *Association of Latino Professionals for America (ALPFA)*

Volunteer | *WEllBuilt Tampa Bay*

CERTIFICATIONS

- DataCamp SQL Associate, DataCamp Data Analyst Associate, Citizen Data Science Certificate (Tableau)