Jill Palma Garro

🌐 Portfolio | ✉ jpalmagarro@gmail.com | 📱 +34 634 609 089 | 💼 LinkedIn

SUMMARY

Computer Scientist specialized in building end-to-end data systems. Bridging the gap between **Engineering reliability** and **Data Science intelligence**. Experienced in orchestrating scalable ELT pipelines, designing real-time anomaly detection architectures, and applying rigorous statistical simulation to product analytics.

SKILLS

Hard Skills: Python · SQL · Apache Spark · Airflow · Snowflake · dbt Core · Docker · PyTorch · Power BI
Soft Skills: Analytical Thinking · Systems Design · Stakeholder Management · Agile Methodologies
Languages: Spanish (Native) · Catalan (Native) · English (Advanced/C1)

ENGINEERING PROJECTS

AnomalyDetection4Pharma (AD4P) | *PySpark, PyTorch, Docker*   *2025 - 2026*
- Achieved **100% detection** of Concept Drift scenarios, outperforming baseline models by **92.5%**.
- Reduced false positives by **40%**, achieving a **7x Signal-to-Noise Ratio** using physics-informed LSTM model.
- Optimized a Spark Medallion Architecture to process sensor data via distributed streaming ingestion.

Enterprise Data Quality Monitor | *Airflow, Snowflake, dbt, Elementary*   *2025*
- Engineered a Dual-DAG pipeline processing **55k+ records** with a **10%** custom chaos injection rate.
- Implemented automated observability with dbt, achieving **100% test coverage** across all models.
- Quantified "Revenue at Risk" in real-time by isolating data faults across **37k+** backfilled historical orders.

AB Decision Engine | *SciPy, Streamlit, Bayesian Stats*   *2025*
- Reduced experiment duration by **41.8%** (from 14 to **8.2 days**) using O'Brien-Fleming Sequential Testing.
- Validated **99% sensitivity** in detecting Sample Ratio Mismatch via **N=500** robust Monte Carlo simulations.
- Built a hybrid Frequentist/Bayesian engine identifying Simpson's Paradox across **10k+** user sessions.

WORK EXPERIENCE

SDG Group   **Barcelona, Spain**
Data Solutions Analyst   *Nov 2024 – May 2025*
- **Developed robust Python automation scripts** to replace manual data workflows, effectively reducing the **reporting cycle time by 30%** while eliminating critical human errors in production.
- **Optimized Snowflake SQL transformation logic** within distributed ETL pipelines, successfully reducing data processing latency by **20%** for Power BI and operational monitoring systems.
- **Orchestrated multi-source data integration** to analytical layers, ensuring **high reliability and data consistency** for 10+ cross-functional stakeholders while maintaining pipeline integrity.

Penguin Random House   **Barcelona, Spain**
Data Scientist Intern   *Jan 2024 – May 2024*
- **Developed scalable Python scripts** to process and validate **500k+ records** within AWS environments, identifying key inefficiencies that reduced manual editorial review time by **20%** globally.
- **Implemented rigorous Snowflake SQL data quality checks** and validation protocols, resulting in a measurable **25% improvement** in dataset consistency for downstream predictive models.
- **Engineered scalable image search model**, significantly improving the reliability of internal tools and increasing query performance by **30%** for editorial teams.

EDUCATION

Universitat Politècnica de Catalunya   **Barcelona, Spain**
Postgraduate in Artificial Intelligence with Deep Learning   *Exp. 2026*

Universitat Autònoma de Barcelona   **Barcelona, Spain**
Bachelor's Degree in Computer Science   *2025*