

KAREN CAMACHO VERASTEGUI

Data Scientist | AI Developer | Data Analyst

Valencia, Spain | +34 618 668 343 | karenverastegui22@outlook.com

Portfolio: karen-jcv.github.io | LinkedIn: in/karen-camacho-verastegui | GitHub: github.com/Karen-JCV

PROFESSIONAL SUMMARY

Data Scientist & AI Developer with a solid quantitative background and a Master's degree in Data Science & AI (Final grade 9.4/10). Specialised in the full data life cycle (End-to-End): from extraction and relational modelling in SQL and analytical pipelines with Python, to training supervised predictive models (Scikit-Learn, calibration and explainability with SHAP), developing Generative AI assistants with RAG architectures (LangGraph, ChromaDB, Gemini 2.5) and operational dashboards in Streamlit and Power BI. Prior experience optimising organisational processes and operational and commercial metrics through decision-making based on analytical evidence.

TECHNICAL SKILLS

Languages & Core: Python (Pandas, NumPy, Scikit-learn, SciPy, Matplotlib, Seaborn), SQL, DAX, basic Bash.
Machine Learning: Supervised classification/regression, calibration (CalibratedClassifierCV), explainability (SHAP), cross-validation.
Generative AI & NLP: RAG (Retrieval-Augmented Generation), ChromaDB, LangGraph, LangChain, Gemini API, Transformers, BERTopic, UMAP.
Data & Persistence: PostgreSQL, MySQL, SQLite, MariaDB, complex SQL queries (JOINS, CTEs, subqueries), Apache Parquet, ETL.
Visualisation & BI: Streamlit, Power BI (relational modelling, DAX, interactive dashboards), Plotly, Advanced Excel (Power Query).
Tools & Environments: Git, GitHub, virtual environments (venv), Docker (deployment fundamentals), VS Code, Jupyter, Linux.

FEATURED DATA SCIENCE & AI PROJECTS

Beauty Girl Studio - Smart Data Suite | [Machine Learning & Churn](#) [\[View Repository\]](#)

Technologies: Python, Scikit-learn, SHAP, SQLite, Apache Parquet, Streamlit, Plotly

- **Churn prediction:** Calibrated supervised model to predict 90-day non-return risk with SHAP explainability.
- **Interactive commercial dashboard:** Streamlit deployment to prioritise at-risk customers, estimate impact and manage reactivation.
- **Columnar storage:** Data structure optimised with Apache Parquet and relational persistence in SQLite.

AI Immigration Assistant (AI Asistente de Extranjería) | [Generative AI, RAG & LangGraph](#) [\[View Repository\]](#)

Technologies: Python, Google Gemini 2.5 Flash, LangGraph, LangChain, ChromaDB, Streamlit, pypdf

- **Semantic vector retrieval:** RAG architecture over ChromaDB and Google Generative AI Embeddings for immigration regulations.
- **Agentic orchestration with memory:** Multi-step conversational flows with LangGraph and persistent multi-session memory.
- **Multimodal processing and streaming:** PDF ingestion and real-time token-by-token response generation in Streamlit.

NLP Review Analysis | [Topic Modeling & Sentiment](#) [\[View Repository\]](#)

Technologies: Python, Transformers, BERTopic, Hugging Face, UMAP, HDBSCAN, Scikit-learn, Seaborn

- **Text mining and modelling:** Unsupervised semantic grouping with BERTopic, UMAP reduction and HDBSCAN clustering.
- **Insight extraction:** Automated identification of customer friction points and generation of strategic recommendations.

Other GitHub projects | [Data Engineering & Analytics](#) [\[View GitHub\]](#)

Technologies: SQL (MySQL), Python, Pandas, Scikit-learn, XGBoost, Seaborn

- **SQL Data Warehouse (COVID-19 vs GDP):** ELT pipeline with staging, core and semantic layers, dimensional model, function, stored procedure and trigger over epidemiological and GDP indicators.
- **Loyalty and Repurchase Model:** Retail repurchase prediction pipeline with cleaning, exploratory analysis, RFM features and Logistic Regression, Random Forest and XGBoost models.
- **Road Accident EDA 2024:** Modular Python pipeline (Pandas and Seaborn) analysing more than 30,000 road accident records in Madrid.

WORK EXPERIENCE

Air Traffic Agent | [DELTACARGO](#) 05/2026 - Present | Santiago de Compostela

- **Logistics data control and analysis:** Management and updating of operational data in software and databases of the international air sector.
- **Documentary audit and consistency:** Verification of customs and transport documentation (AWB, cargo manifests), ensuring very few incidents.
- **Metric monitoring and analytical resolution:** Real-time tracking of shipments and flights to anticipate logistics disruptions through proactive analysis.
- **Operational process optimisation:** Active collaboration in improving workflows and efficiency in the department's file management.

Purchasing & Data Operations Analyst | [Hotel Congreso](#) 07/2024 - 05/2026 | Santiago de Compostela

- **Quantitative procurement analysis:** Analytical models in advanced Excel and SQL to monitor costs, reducing time by 50%.
- **Data governance and cleansing:** Standardisation of stock records across departments, improving data consistency by 20%.
- **Strategic reporting:** Development of dashboards and spend reports to support management and margin control.

- Database Analyst (Training Contract) | [Clictic](#)**

12/2023 - 01/2024 | Santiago de Compostela

 - **Database optimisation:** Management and maintenance of MySQL, PostgreSQL and MariaDB, reducing query times by 25%.
 - **Advanced SQL development:** Creation of complex queries, triggers and stored procedures for ETL extraction and cleansing.
 - **Security and consistency:** Execution of periodic backup policies and verification of referential integrity.
- Process Control & KPI Manager | [Integro Corp SpA](#)**

03/2017 - 09/2021 | Santiago, Chile

 - **Performance metrics management:** Quantitative tracking of productivity KPIs, achieving efficiency improvements of up to 38%.
 - **Data-driven executive reports:** Development of management dashboards based on analytical evidence.
 - **Leadership and re-engineering:** Bottleneck detection and team coordination for the adoption of structured workflows.
- Operations & Planning Coordinator | [SunSol Hoteles](#)**

08/2014 - 10/2016 | Margarita Island

 - **Analytical and logistics planning:** Management of large volumes of information and historical demand, increasing satisfaction by 15%.
 - **Stakeholder management:** Budget control and negotiation with corporate clients in high-demand environments.

EDUCATION

- Master's in Data Science & Artificial Intelligence Development (480h) | [Evolve](#)**

01/2026 - 10/2026

Final programme grade: 9.4 / 10 (Outstanding)

Modules 10/10: Expert Assistant with Gemini, RAG and Agents; MLOps. Modules 9.5/10: Machine Learning; Final Project.
- Professional Certificate: Database Administration (570h) | [G12 Servicios](#)**

01/2023 - 07/2023

Specialisation in relational SQL, data design, security and database server administration.
- Bachelor's Degree in Modern Languages (4 years) | [Universidad de Margarita](#)**

02/2009 - 07/2013

Training in structured linguistic analysis, technical writing and international communication.

ADDITIONAL CERTIFICATIONS & LANGUAGES

- Python Programming (40h) & Java (90h) - Centro de Novas Tecnoloxías de Galicia (CNTG) | Anti-Money Laundering Prevention (400h) - Euroinnova

Spanish: Native | English: Full Professional Proficiency (C1 - Bachelor's in Modern Languages, technical reading and AI papers)