

KAREN CAMACHO VERASTEGUI

Data Scientist | AI Developer | Data Analyst

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

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

RESUMEN PROFESIONAL

Data Scientist & AI Developer con sólida base cuantitativa y titulación de Máster en Data Science & IA (Nota final 9.4/10). Especializada en el ciclo de vida completo del dato (End-to-End): desde extracción y modelado relacional en SQL y pipelines analíticos con Python, hasta el entrenamiento de modelos predictivos supervisados (Scikit-Learn, calibración y explicabilidad con SHAP), desarrollo de asistentes de IA Generativa con arquitecturas RAG (LangGraph, ChromaDB, Gemini 2.5) y cuadros de mando operativos en Streamlit y Power BI. Trayectoria previa optimizando procesos organizacionales y métricas operativas y comerciales mediante toma de decisiones basada en evidencia analítica.

HABILIDADES TÉCNICAS

- Lenguajes & Core:** Python (Pandas, NumPy, Scikit-learn, SciPy, Matplotlib, Seaborn), SQL, DAX, Bash básico.
- Machine Learning:** Clasificación/Regresión supervisada, calibración (CalibratedClassifierCV), explicabilidad (SHAP), validación cruzada.
- IA Generativa & NLP:** RAG (Retrieval-Augmented Generation), ChromaDB, LangGraph, LangChain, Gemini API, Transformers, BERTopic, UMAP.
- Datos & Persistencia:** PostgreSQL, MySQL, SQLite, MariaDB, consultas SQL complejas (JOINS, CTEs, subconsultas), Apache Parquet, ETL.
- Visualización & BI:** Streamlit, Power BI (modelado relacional, DAX, dashboards interactivos), Plotly, Excel Avanzado (Power Query).
- Herramientas & Entornos:** Git, GitHub, entornos virtuales (venv), Docker (fundamentos de despliegue), VS Code, Jupyter, Linux.

PROYECTOS DESTACADOS DE DATA SCIENCE & IA

Beauty Girl Studio - Smart Data Suite | Machine Learning & Churn

[\[Ver Repositorio\]](#)

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

- **Predicción de abandono (Churn):** Modelo supervisado calibrado para predecir riesgo de no retorno a 90 días con explicabilidad SHAP.
- **Dashboard comercial interactivo:** Despliegue en Streamlit para priorizar clientas en riesgo, calcular impacto y gestionar reactivación.
- **Almacenamiento columnar:** Estructura de datos optimizada en Apache Parquet y persistencia relacional en SQLite.

AI Asistente de Extranjería | IA Generativa, RAG & LangGraph

[\[Ver Repositorio\]](#)

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

- **Recuperación semántica vectorial:** Arquitectura RAG sobre ChromaDB y Google Generative AI Embeddings para normativa migratoria.
- **Orquestación agéntica con memoria:** Flujos conversacionales multi-paso con LangGraph y memoria persistente multisesión.
- **Procesamiento multimodal y streaming:** Ingesta de PDFs y generación de respuestas en tiempo real token a token en Streamlit.

Análisis de Reviews con NLP | Topic Modeling & Sentimiento

[\[Ver Repositorio\]](#)

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

- **Minería de texto y modelado:** Agrupación semántica no supervisada con BERTopic, reducción UMAP y clustering HDBSCAN.
- **Extracción de insights:** Identificación automatizada de fricciones de clientes y generación de recomendaciones estratégicas.

Otros proyectos en GitHub | Data Engineering & Analítica

[\[Ver GitHub\]](#)

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

- **Data Warehouse SQL (COVID-19 vs PIB):** Pipeline ELT con capas staging, core y semántica, modelo dimensional, función, procedimiento almacenado y trigger sobre indicadores epidemiológicos y de PIB.
- **Modelo de Fidelización y Recompra:** Pipeline de predicción de recompra en retail con limpieza, análisis exploratorio, variables RFM y modelos Logistic Regression, Random Forest y XGBoost.
- **EDA Accidentalidad 2024:** Pipeline modular en Python (Pandas y Seaborn) para analizar más de 30.000 registros de accidentes de tráfico en Madrid.

EXPERIENCIA LABORAL

Air Traffic Agent | DELTACARGO

05/2026 - Presente | Santiago de Compostela

- **Control y análisis de datos logísticos:** Gestión y actualización de datos operativos en software y bases de datos del sector aéreo internacional.
- **Auditoría y consistencia documental:** Verificación de documentación aduanera y de transporte (AWB, manifiestos de carga), asegurando muy pocas incidencias.
- **Monitoreo métrico y resolución analítica:** Seguimiento en tiempo real de envíos y vuelos para anticipar disrupciones logísticas mediante análisis proactivo.
- **Optimización de procesos operativos:** Colaboración activa en la mejora de flujos y eficiencia en la gestión de expedientes del departamento.

- Analista de Compras & Operaciones de Datos | Hotel Congreso**

07/2024 - 05/2026 | Santiago de Compostela
- **Análisis cuantitativo de aprovisionamiento:** Modelos analíticos en Excel avanzado y SQL para monitorizar costes, reduciendo tiempos en un 50%.
 - **Gobernanza y depuración de datos:** Estandarización de registros de stock entre departamentos, incrementando la coherencia del dato en un 20%.
 - **Reporting estratégico:** Desarrollo de cuadros de mando e informes de gasto para respaldo directivo y control de márgenes.
- Analista de Bases de Datos (Contrato de Formación) | Clicktic**

12/2023 - 01/2024 | Santiago de Compostela
- **Optimización de bases de datos:** Gestión y mantenimiento de MySQL, PostgreSQL y MariaDB, reduciendo tiempos de consulta en un 25%.
 - **Desarrollo SQL avanzado:** Creación de consultas complejas, triggers y procedimientos almacenados para extracción y depuración ETL.
 - **Seguridad y consistencia:** Ejecución de políticas de respaldo periódico y verificación de integridad referencial.
- Encargada de Control de Procesos & KPIs | Integro Corp SpA**

03/2017 - 09/2021 | Santiago de Chile
- **Gestión métrica de rendimiento:** Seguimiento cuantitativo de KPIs de productividad, alcanzando mejoras de eficiencia de hasta un 38%.
 - **Informes ejecutivos basados en datos:** Elaboración de cuadros de mando para la gerencia basados en evidencia analítica.
 - **Liderazgo y reingeniería:** Detección de cuellos de botella y coordinación de equipos para la adopción de flujos estructurados.
- Coordinadora Operativa & Planificación | SunSol Hoteles**

08/2014 - 10/2016 | Isla de Margarita
- **Planificación analítica y logística:** Gestión de grandes volúmenes de información y demanda histórica, aumentando la satisfacción en un 15%.
 - **Gestión de partes interesadas:** Control presupuestario y negociación con clientes corporativos bajo entornos de alta exigencia.

EDUCACIÓN

- Máster en Data Science & Desarrollo de Inteligencia Artificial (480h) | Evolve**

01/2026 - 10/2026
- Nota final del programa: 9.4 / 10 (Sobresaliente)**
- Módulos 10/10: Asistente Experto con Gemini, RAG y Agentes; MLOps. Módulos 9.5/10: Machine Learning; Proyecto Final.
- Certificado de Profesionalidad: Administración de Bases de Datos (570h) | G12 Servicios**

01/2023 - 07/2023
- Especialización en SQL relacional, diseño de datos, seguridad y administración de servidores de bases de datos.
- Licenciatura en Idiomas Modernos (4 años) | Universidad de Margarita**

02/2009 - 07/2013
- Formación en análisis lingüístico estructurado, redacción técnica y comunicación internacional.

CERTIFICACIONES COMPLEMENTARIAS & IDIOMAS

- Programación en Python (40h) & Java (90h) - Centro de Novas Tecnoloxías de Galicia (CNTG) | Prevención Blanqueo (400h) - Euroinnova
- Español: Nativo | Inglés: Competencia Profesional Completa (C1 - Licenciatura en Idiomas Modernos, lectura técnica y papers IA)