



## Sobre mí

Soy un profesional comprometido y proactivo, con gran interés en la Ciencia de Datos, la Inteligencia Artificial y el Desarrollo de Código. Mi objetivo es aplicar mis conocimientos técnicos en el desarrollo de soluciones innovadoras que respondan a problemas reales y generen un impacto positivo en la sociedad.

## Trayectoria Laboral

### Científico de Datos

UC3M-Santander Big Data Institute (IBiDat)

09/2024 - Actualidad

**Habilidades Clave:** Análisis y modelización de datos, Machine Learning (ML), Inteligencia Artificial Generativa (GenAI), Desarrollo de código y Automatización (Python, Git, APIs, Docker, MongoDB), Forecasting (Series Temporales), Documentación y Reporte Técnico.

### Proyectos Destacados:

- **Estimación de Demanda y Factores de Impacto en Parking Público (Grupo Ruiz):** Predicción a largo plazo de la ocupación de Parking de Atocha y estimación del impacto de construcción de un Área Comercial sobre el mismo. También se estudia el número de plazas óptimo para ser puestas en alquiler, así como el efecto de la apertura de un nuevo parking público en la Estación de Atocha.
- **Predicción de Demanda de Transporte Público Interurbano (Grupo Ruiz):** Predicción a largo plazo de la demanda de autobuses interurbanos en la Comunidad de Madrid utilizando Forecasting de Series Temporales y ML para generación de datos.
- **Asistente Virtual con IA Generativa (Universia - Banco Santander):** Diseño e implementación de un sistema GenAI (chat consultivo, chat evalaudor, generación de material de apoyo y de evaluación) para mejorar el aprendizaje con videos educativos. Puesta en producción con APIs, Docker y MongoDB.
- **Herramienta de Generación y Evaluación de Clips de Video (Universia - Banco Santander):** Desarrollo de una herramienta automatizada para generar reels/shorts educativos y un evaluador de calidad de los clips generados (valor educativo, cobertura de conceptos clave, relevancia y diversidad de contenido).
- **Investigación en LLMs y ASR:** Evaluación comparativa de sistemas de Transcripción Audiovisual (ASR) y análisis del impacto educativo de asistentes virtuales basados en Large Language Models (LLMs).

## Fabio Scielzo Ortiz

Científico de datos.

- ▶ Madrid, España.

## Contacto

✉ fabio.scielzoortiz@gmail.com

GitHub

## Idiomas

Español Native

Inglés B2

## Certificados

Grado en Estadística y Empresa (UC3M)

Máster en Big Data (UC3M)

Linguaskill English Test (Cambridge)

Desarrollo de la tesis doctoral "Métodos de aprendizaje estadístico basados en distancias y profundidades para datos masivos de estructura compleja" en el programa de doctorado en Estadística para la Ciencia de Datos en la Universidad Carlos III de Madrid, supervisado por la Catedrática Aurea Grané Chavez.

Mis principales contribuciones como doctorando han sido el desarrollo de distancias estadísticas robustas para datos multivariantes de tipo mixto, disponibles en el paquete de Python `robust-mixed-dist`, así como algoritmos de clustering robustos para grandes volúmenes de datos multivariantes de tipo mixto, disponibles en el paquete de Python `db-robust-clust`. Estos desarrollos han sido aplicados a dos casos de estudio prácticos, demostrando su potencial de aplicabilidad, como son la detección de perfiles de calidad de vida en la población europea de edad avanzada, así como la detección de grupos de opinión a nivel internacional sobre el conflicto en Gaza, utilizando datos de redes sociales.

#### **Artículos Académicos:**

- Grané, A., & Scielzo-Ortiz, F. (2025). On generalized gower distance for mixed-type data: Extensive simulation study and new software tools. *SORT-Statistics and Operations Research Transactions*, 49 (2), 65-96.
- Grané, A., & Scielzo-Ortiz, F. (2025). Fast k-medoids and q-fold fast k-medoids: New distance-based clustering algorithms for large mixed-type data. Retrieved from <https://hdl.handle.net/10016/46673> (Working Paper 2025-02, Statistics and Econometrics Series, University Carlos III of Madrid)
- Grané, A., & Scielzo-Ortiz, F., & Albarrán, I. (2026). Quality of life in older European population: Finding emotional well-being robust profiles from SHARE data. (*Work in progress*)
- Grané, A., & Scielzo-Ortiz, F., & Gorfinkel, M. (2026). Quantifying Global Polarization: Detection and analysis of opinion groups on the Gaza conflict at an international level based on data from the social network Reddit. (*Work in progress*)

#### **Presentaciones en Congresos:**

- Scielzo-Ortiz, F. (2024, November). New clustering algorithms for large mixedtype data. Oral dissertation. 6th International Workshop on Proximity Data, Multivariate Analysis and Classification, Girona, Spain.
- Scielzo-Ortiz, F. (2025, April). Robust clustering for large datasets of weighted and mixed type data. An application to mental well-being in older adults. Oral dissertation. 3th Conference of Statistic and Data Science Group of the Italian Statistical Society, Milan, Italy.
- Scielzo-Ortiz, F. (2025, June). Robust clustering for large datasets of weighted and mixed type data. An application to mental well-being in older adults. Oral dissertation. XLI National Congress in Statistics and Operation Research (SEIO), Lleida, Spain.

#### **Profesor de programación en Python para Data Science**

01/2025

Bootcamp, ANBAN

Encargado de diseño y docencia del módulo *Programación en Python para Ciencia de Datosramación en Python* del Bootcamp de Data Science de ANBAN.

#### **Profesor de Machine Learning**

04/2024

Saturdays AI, ANBAN

Profesor de la sesión "Machine Learning: clasificación supervisada" del curso de Data Science de las asociaciones Saturdays AI y ANBAN, impartido en la Universidad de Elche.

## Educación

### Máster en Big Data Analytics

09/2023 - 07/2024

Universidad Carlos III de Madrid

Máster universitario en Inglés que proporciona una excelente formación para realizar análisis estadísticos y computacionales complejos de grandes volúmenes de datos, con el objetivo de obtener resultados aplicables en identificación de patrones, predicción, previsión, simulación y optimización. Calificaciones a destacar (hasta ahora):

- Once (11) **sobresalientes**: Business applications of big data analytics (10), Master's Thesis (TFM) (10), Optimization for large scale data (10), High performance computing for Big Data: (9.6), Statistics for Data Analysis (9.5), Data Analytics for the Smart Society (9.3), Bayesian Learning (9.3), Machine Learning I (9.2), Web Analytics (9.2), Predictive modeling (9), Time Series Forecasting (9).
- Tres (3) **notables altos**: Social Networks and Data Viz (8.9), Back-end for Big Data (8.2), Technological fundamentals of Big Data (8).
- Nota media: 8.6

### Grado en Estadística y Empresa

09/2019 - 07/2023

Universidad Carlos III de Madrid

Este grado universitario está dividido en un bloque de **Ciencia de Datos** (70%) y otro de **Empresa** (30%). Este grado me ha otorgado conocimientos y habilidades relacionados con el mundo de la Ciencia de Datos, concretamente, he adquirido fuertes destrezas en Estadística, Machine Learning y programación. He sido el **mejor expediente del Grado en mi promoción (2019-2023)**, destacando por los siguientes méritos:

- Doce (12) **Matrículas de Honor**: Aprendizaje Automático, Aprendizaje Estadístico, Análisis de Muestras, Análisis Multivariante, Métodos de Regresión, Métodos Avanzados de Regresión, Métodos Bayesianos, Optimización, Probabilidad II, Técnicas de Inferencia Estadística I, Técnicas de Remuestreo, Trabajo Final de Grado.
- Ocho (8) **sobresalientes**: Cálculo I, Economía Financiera, Hojas de Cálculo, Prácticas, Programación I, Programación II, Principios de Economía, Visual Thinking.
- Cinco (5) **notables altos** en las siguientes asignaturas: Álgebra Lineal, Ampliación de Matemáticas, Bases de Datos, Marketing Digital, Métodos Econométricos, Procesos Estocásticos.
- Nota media del bloque de Ciencia de Datos de 9, y media total de 8.5.

Madrid, 30th January 2026

Fabio Scielzo Ortiz