



# Gustavo João Gonçalves Caldas

**Fecha de nacimiento:** 30/07/2000 | **Nacionalidad:** Portuguesa | **Número de teléfono:** (+34) 614146384 (Móvil) | **Dirección de correo electrónico:** [gj.goncalvescaldas@gmail.com](mailto:gj.goncalvescaldas@gmail.com) | **Sitio web:** <https://gj-goncalvescaldas.github.io/> | **LinkedIn:** [Gustavo Caldas](#) | **Dirección:** Arroyomolinos, Madrid, España (Domicilio)

## ● SOBRE MI

### Intereses

Desde mis años en la universidad, he desarrollado un fuerte interés por el mundo de los datos, orientando mi aprendizaje y experiencia hacia la **Ingeniería de Datos**, el **Machine Learning** y la **Visualización**. Me apasiona construir bases de datos sólidas, diseñar modelos de predicción y extraer información valiosa que aporte un impacto significativo a las empresas. Mis principales intereses giran en torno a transformar datos en conocimiento accionable, impulsando la toma de decisiones estratégicas y el éxito de los proyectos en los que participo.

Valoro trabajar en entornos multiculturales, donde se fomente la diversidad de ideas y perspectivas. Además, aprecio la flexibilidad horaria, ya que me permite optimizar mi tiempo y evitar el tráfico. Dado que resido en el Sur de Madrid, el teletrabajo también es un aspecto fundamental para mantener un equilibrio entre la vida profesional y personal.

En 2024 enfrenté un desafío personal significativo, lo que llevó a pausar temporalmente mi trayectoria profesional.

## ● EXPERIENCIA LABORAL

01/09/2022 – 01/11/2023 Porto, Portugal  
**DATA MINING RESEARCHER IOTECH**

- Colaboré con **IOTECH** en el proyecto **ioCity**, financiado por la Unión Europea y centrado en transformar la movilidad urbana mediante soluciones inteligentes e innovadoras. Este proyecto buscó aumentar la eficiencia del transporte público y reducir la contaminación, promoviendo el uso de transportes colectivos a través de modelos de gamificación y sistemas predictivos.
- Realicé la preparación, limpieza e integración de datos utilizando la metodología **CRISP-DM**. Comprendí las necesidades del negocio, cuyo objetivo era prever la **tasa de ocupación** de los transportes públicos en la ciudad de **Porto** en cada parada. Analicé los datos disponibles, eliminé variables irrelevantes para la predicción, identifiqué problemas y transformé los datos para garantizar modelos predictivos más eficientes.
- Desarrollé un **sistema en tiempo real** para predecir la disponibilidad de asientos en los transportes públicos de Porto, mejorando significativamente la experiencia del usuario mediante el uso de **Machine Learning**. Implementé y evalué modelos de regresión, incluidos **Random Forest Regression**, **LightGBM**, **CatBoost**, **Gradient Boosting** y **XGBoost**.  
Tecnologías utilizadas: **Python**, **pandas**, **numpy**, **scikit-learn**, **matplotlib**, **Excel**.

01/08/2023 – 30/11/2023 Madrid, España  
**DATA ENGINEER NATIONALE-NEDERLANDEN**

- Especialista en Ingeniería de Datos con una sólida base en ingeniería de software.
- Construcción de un **Data Mart** desde cero en **esquema estrella**, incluyendo diseño e implementación de tablas de dimensiones y hechos. Procesamiento y carga diaria de datos desde Excel mediante **Azure Data Factory** y **Databricks**, asegurando una manipulación y carga correcta utilizando herramientas del entorno **Microsoft Azure**.
- Creación de **informes y plantillas en PowerBI** para presentar la información del Data Mart, implementando funciones avanzadas en **DAX** para análisis dinámico y visualización de datos.
- Participación en proyectos complejos, corrigiendo errores e integrando nuevas variables de datos según los requerimientos, como promociones específicas según el tipo de cliente.  
Tecnologías utilizadas: **Databricks**, **Azure Data Factory**, **SQL Server**, **SQL Management**, **Excel**, **Python** y **diversas bibliotecas**, **PowerBI**, **Microsoft Azure**, **DAX**, **Copilot**.

## ● EDUCACIÓN Y FORMACIÓN

10/09/2018 – 04/07/2022 Braga, Portugal  
**LICENCIATURA EN INGENIERÍA DE SISTEMAS DE INFORMACIÓN** Universidade do Minho

Web [uminho.pt](http://uminho.pt) | Nivel en el MEC Nivel 6 EQF-MEC

10/09/2021 – 20/06/2024 Braga, Portugal  
**MASTER EN INGENIERÍA DE SISTEMAS DE INFORMACIÓN** Universidade do Minho

Web [uminho.pt](http://uminho.pt) | Nivel en el MEC Nivel 7 EQF-MEC

Web [www.unir.net](http://www.unir.net) | Nivel en el MEC Nivel 5 EQF-MEC

COMPETENCIAS DE IDIOMAS

Lengua(s) materna(s): **PORTUGUÉS**  
Otro(s) idioma(s):

	COMPRENSIÓN		EXPRESIÓN ORAL		EXPRESIÓN ESCRITA
	Comprensión auditiva	Comprensión lectora	Producción oral	Interacción oral	
ESPAÑOL	C1	C1	C1	C1	C1
INGLÉS	B2	B2	B2	B2	B2

Niveles: A1 y A2 (usuario básico), B1 y B2 (usuario independiente), C1 y C2 (usuario competente)

COMPETENCIAS DIGITALES

Data Gathering, Data Processing, Data Visualization, Data Analysis | Experiencia en lenguajes de programación como Java, Python y C | Python(Pandas, Numpy, APIs, etc) | Pyspark | Extracción, Transformación y Carga de datos (ETL) | Machine Learning | PostgreSQL + SQLite + SQLServer + MySQL | Power Bi & Power Query | Microsoft Azure | Azure Databricks | Dominio microsoft office incluyendo excel | Data Warehouse

PUBLICACIONES

2022  
[Dashboards de Control Inteligentes para Monitorear los Sucesos en Ciudades Inteligentes: Un Estudio de Caso Portugués](#)

Este artículo aborda la mejora de las solicitudes de intervención del Regimiento de Bomberos Profesionales de Lisboa a través de dashboard de control inteligentes. El proyecto busca analizar indicadores clave para mejorar la toma de decisiones en emergencias. Entre los resultados, se destaca que el **58% de las falsas alarmas se cancelan después de activar los servicios de emergencia**, y que el **domingo es el día con más registros** de sucesos, especialmente a las **9 am y 8 pm. El otoño también tiene el mayor número de incidentes**. Estos datos ayudan a anticipar escenarios y mejorar la toma de decisiones en situaciones de emergencia.

PROYECTOS

01/01/2025 – 10/02/2025  
[Pipeline de Datos en Tiempo Real para Análisis de Clima y COVID-19 Usando Azure Data Factory, Databricks y Power BI](#)

Este proyecto demuestra la creación de dos flujos de trabajo **ETL (Extracción, Transformación y Carga)** completos utilizando **Azure Data Factory (ADF), Databricks, Delta Lake y Power BI**. La primera parte se centra en un **pipeline en tiempo real para datos meteorológicos**, donde los datos se obtienen a través de una API, se transforman con **PySpark en Databricks**, se cargan en **Delta Lake** y se visualizan en **Power BI**. La segunda parte procesa **datos históricos de COVID-19 del ECDC**, siguiendo pasos similares para la ingesta, transformación y carga en **Azure SQL** para su posterior visualización. El proyecto destaca el uso de **triggers automáticos en Data Factory** para actualizaciones programadas, así como la **automatización en Power BI**. Muestra el flujo completo de datos, desde la ingesta hasta la visualización, proporcionando una solución escalable para el procesamiento y análisis eficiente de grandes volúmenes de datos.

Enlace <https://github.com/gj-goncalvescaldas/Azure-Data-Factory>