



# FINAL PRACTICE

## The job of the data scientist

### Brief description

This project is a case study of a Business Intelligence (BI) solution for a telecommunications multinational. It involves designing the data analysis system architecture, modeling the database, loading extracted data, and performing data exploitation to support decision-making.

Sergio Hervás Aragón  
Sergiohervas9@gmail.com

Our client is a department of a multinational company in the telecommunications sector that manages requests and incidents occurring in a range of services and environments provided by the company to its employees and customers.

Each time there is an access request or a functionality issue, users open a ticket in the ticketing tool to report it and initiate its resolution. The client has provided us with data export from 2020 until the end of August 2021 to analyze specific questions they want to address through graphical reports or dashboards.

The client provides us with the following technological architecture:

- MySQL database
- ETL Pentaho Data Integration
- Analytical frontend: Microsoft Power BI or Tableau Public

In summary, our client wants to know:

- The evolution of the total ticket volume, differentiated by incidents and requests.
- Whether the priority of open tickets has changed in recent months (% distribution by priority).
- Which services accumulate the most tickets and which of them fail to meet SLAs the most.
- Which services have the highest backlog (currently open tickets).
- Which services resolve tickets the fastest and the slowest.
- Ticket volume in production environments versus non-production environments.
- The ability to filter by priority, ticket type, and service.

Our client has requested the following project documentation:

- Database schema
- Documentation of the data loading processes
- Accuracy checks of the data, including:
  - Expected values based on the loading dataset
  - SQL checks (via direct queries)
  - Comparison with the dashboard (MySQL data vs. results in Power BI/Tableau)
- Functionalities and visualizations developed in Power BI or Tableau
- Design decisions made at different stages

## Variable Information

Each dataset must include the following columns:

- **Fecha de creación:** Specifies the time when the ticket is created. It will be used to perform the temporal analysis.
- **Numero de incidente:** Unique ticket identifier.
- **Descripción:** Brief description of the request or incident.
- **Servicio:** The different services that the company provides to its users and clients.
- **Tipo de servicio:** Specify whether it is a request or an incident:
  - **Incidencias:** Restoration of infrastructure or restoration of service to user.
  - **Peticiones:** Service request by user
- **Prioridad:** Specifies the priority with which the ticket should be treated.
  - There are four priorities, ordered from most important to least: Critical, High, Medium, and Low.
  - Only incidents can have critical priority and should only be considered in production environments.
- **Estado:** Indicates the status of the ticket.
  - A ticket is considered finished if it is in Closed or Resolved status.
  - A ticket is considered open if it is pending or assigned.
  - Cancelled tickets do not count.
- **Torre:** It is a functional grouping of services. Each service belongs to a management tower.
- **Entorno:** Indicates the environment that the ticket affects. Production environments are all those that contain a PRO.
- **Estado cumplimiento:** Indicates whether the ticket has been resolved within the time required by the client or not.
  - Within the service objective: ticket resolved appropriately
  - Service objectives not met: ticket resolved late
- **Duración días:** Number of days it took to resolve the ticket.

## **OBJECTIVES**

With the given datasets and variables given, you are required to:

1. Analysis of initial data and proposed data model
2. Model development in MySQL
3. Developing loads with Pentaho Data Integration
4. Reports made with PowerBI/Tableau