

# Examen de Admisión – Ingeniero de Datos

## Objetivo

Evaluar las competencias del candidato en modelado de datos, integración de información desde un Data Lake, construcción de reportes en Power BI, aplicación de técnicas de machine learning y aprovechamiento de modelos de lenguaje (LLM) para análisis de datos.

## Instrucciones Generales

1. El examen es individual.
2. Tiempo máximo sugerido: 4 días.
3. Puedes usar cualquier recurso público (documentación, foros, librerías open-source).
4. Deberás entregar:
  - Archivo .pbix con el modelo y reportes de Power BI.
  - Scripts/notebooks utilizados para el modelo de Machine Learning y transformación de datos.
  - Documentación breve (README o PDF) explicando tus pasos, supuestos y decisiones.

## Parte 1 – Modelado e Informe en Power BI

Archivos proporcionados:

- Tablas de dimensiones (dimCities, dimLocations, dimSublocations, dimDevices, dimSensors, dimUnidades).
- Carpeta 'Lecturas\_Enero\_2025' con subcarpetas (una por día) con lecturas de sensores almacenadas en archivos parquet (un archivo parquet cada 5 minutos).

Hint: Crear un script en Python donde concatene en una sola tabla (factLecturas) todos los archivos parquet del mes de Enero. Esa será su tabla de hechos en Power BI.

## Parte 2 – Machine Learning

1. Seleccionar y justificar un modelo de machine learning que consideres útil e interesante aplicar.
2. Explicar brevemente la lógica del modelo elegido y por qué aporta valor.
3. Entrenar el modelo con los datos proporcionados (Python, R u otra herramienta).
4. Entregar el notebook o script con los resultados y comentarios.

## 💡 Parte 3 – Bonus Extra (Inteligencia Artificial Generativa)

De manera opcional:

- 1. Construir un asistente de IA que, utilizando un modelo LLM gratuito (ej. Ollama con Mistral o LLaMA 2), responda preguntas en lenguaje natural sobre los datos.
  - Ejemplo de preguntas:
  - "¿Cuál fue el promedio de temperatura del cuarto de congelación en la última semana?"
  - "Dame el sensor de corriente con mayor valor registrado en Enero."
  - "Resume las métricas clave por ubicación."
- 2. El asistente debe conectarse al dataset para responder con base en datos reales (no inventados).
- 3. Documentar brevemente cómo se implementó y cómo interactuar con él.

### 📖 Contexto de los datos

Los datos provienen de un sistema de monitoreo instalado en varias sucursales de restaurantes de comida rápida. El sistema cuenta con diferentes dispositivos, cada uno de los cuales puede tener asociados uno o varios sensores.

Ejemplos:

- Una máquina de nieve puede tener tres sensores de corriente (uno por cada fase eléctrica).
- Un cuarto frío puede contar con sensores de temperatura y humedad.
- Un área de servicio puede tener sensores de energía y switches de control.

La tabla de hechos (factLecturas) contiene los registros de lecturas de cada sensor en distintos momentos del tiempo.

La tabla de dimensiones dimSensors describe las características de cada sensor (nombre, dispositivo asociado, modelo, fechas de alta/modificación, etc.).

## Indicaciones para el Dashboard

1. No es obligatorio utilizar todos los sensores en el informe. Puedes seleccionar los más relevantes.
2. El dashboard puede incluir varias pestañas con una distribución adecuada del informe.
3. Se valorará la capacidad para:
  - Relacionar correctamente las dimensiones y las lecturas.
  - Presentar métricas útiles y comprensibles para un usuario de negocio.
  - Diseñar un informe claro y organizado que permita analizar los datos de forma intuitiva.

## Criterios de Evaluación

- Parte 1: Correcto modelado de relaciones, calidad del dashboard, claridad en la visualización.
- Parte 2: Coherencia en la elección del modelo ML, justificación, implementación y análisis de resultados.
- Parte 3 (Bonus): Creatividad, innovación y capacidad de integrar un LLM con fuentes de datos.