# Análisis del Mercado Laboral de Roles de Datos en el Reino Unido

## 📦 Entregable 1

### 📌 Descripción del Caso de Negocio

En los últimos años, los roles relacionados con datos han experimentado un notable crecimiento en el mercado laboral. Empresas de todos los sectores buscan cada vez más profesionales con competencias en análisis de datos, estadística y herramientas de visualización. Sin embargo, tanto los candidatos como los empleadores enfrentan dificultades para comprender cuáles son las habilidades más demandadas, qué rangos salariales se ofrecen y cómo varía la disponibilidad de empleo según la región o la industria.  
  
Este proyecto se plantea como una solución de inteligencia de mercado laboral, ofreciendo un análisis profundo de las ofertas de trabajo relacionadas con ciencia y análisis de datos publicadas en plataformas del Reino Unido. Su objetivo es proporcionar información clara y actualizada que sirva de apoyo a la toma de decisiones para:  
  
- Profesionales que desean orientar o redirigir su carrera.  
- Equipos de selección que necesitan entender tendencias salariales y tecnológicas.  
- Instituciones educativas que buscan adaptar su oferta formativa a la demanda real del mercado.

### 🎯 Objetivos del Proyecto e Impacto Esperado

Objetivos:  
1. Extraer y procesar datos de vacantes laborales relacionadas con datos (ciencia de datos, analista de datos y especialista en datos espaciales) en el Reino Unido.  
2. Limpiar, transformar y normalizar dichos datos para análisis significativos.  
3. Identificar patrones clave: tecnologías más solicitadas, soft skills, distribución geográfica, rangos salariales y tipos de contrato.  
4. Crear paneles interactivos que permitan explorar las tendencias del mercado.

5. Crear un chatbot interactivo para extraer información de la base de datos usando lenguaje humano.  
  
Impacto Esperado:  
- Ofrecer una visión clara sobre la evolución y demanda de perfiles en análisis de datos.  
- Apoyar el diseño de programas de formación basados en datos reales.  
- Ayudar a los profesionales a prepararse mejor para los puestos con mayor potencial.  
- Brindar a las agencias de reclutamiento herramientas para optimizar sus estrategias de contratación.

### 🛠️ Tecnologías y Herramientas Utilizadas

- Python: ETL, predicción de salarios, categorización de empleos.  
- PostgreSQL: Almacenamiento estructurado de los datos.  
- Power BI / Streamlit: Visualización interactiva de los resultados.  
- Git / GitHub: Control de versiones y colaboración.

- Chatbot usando modelo LLM local (Mistral 7B)

### 🔗 Fuentes de Datos

Se utilizaron APIs públicas de plataformas como Adzuna, Reed, DevITJobs, entre otras. Estas fuentes fueron elegidas por su frecuencia de actualización, cobertura geográfica y riqueza semántica de los datos.

**✅ Entregable 2: ETL**

Este entregable documenta el proceso completo de ETL (Extracción, Transformación y Carga) aplicado a los datos de empleos obtenidos desde APIs públicas como Adzuna y Reed. El objetivo principal es consolidar, enriquecer y estructurar la información de vacantes para su posterior análisis exploratorio y visualización.

**📁 Esquema de Base de Datos**

La base de datos PostgreSQL se ha diseñado con un enfoque relacional y contiene las siguientes tablas principales:  
  
- `companies`: Almacena los nombres de las empresas que publican ofertas.  
- `locations`: Registra información geográfica como ciudad, latitud y longitud.  
- `jobs`: Contiene los detalles de cada vacante, incluyendo salario, descripción, fecha y referencias a empresa, ubicación y nivel del puesto.  
- `job\_levels`: Define la jerarquía o seniority de los empleos (ej. Junior, Senior).  
- `job\_metadata`: Guarda información relacionada con la búsqueda, como palabras clave y fecha de descarga.  
  
Se han implementado validaciones, claves foráneas y múltiples índices para optimizar la eficiencia de consultas y garantizar la integridad referencial.

**🔄 Proceso de Transformación**

Durante la transformación se realizaron las siguientes tareas:  
  
- Limpieza y estandarización de datos.  
- Detección y eliminación de duplicados.  
- Predicción de salarios cuando no estaban disponibles, utilizando modelos de machine learning entrenados con descripciones y títulos.  
- Enriquecimiento de ubicaciones mediante una API geográfica.  
- Categorización de empleos basada en palabras clave extraídas de las descripciones.

**✅ Entregable 3: Análisis Exploratorio de Datos (EDA)**

El análisis exploratorio de datos (EDA) se realizó sobre la base de datos resultante del proceso ETL. Su objetivo fue identificar tendencias, patrones y relaciones clave en el mercado laboral de roles de datos en el Reino Unido.

**📊 Principales Áreas Analizadas**

- Distribución Salarial

- Palabras clave

-Tipos y niveles de trabajo

Gestión de duplicados

-Cálculo de salarios nulos

Gráfico, Histograma

El contenido generado por IA puede ser incorrecto.  
Este análisis sirvió como base para el diseño de gráficos y para alimentar modelos predictivos en futuras fases del proyecto.

## ✅ Entregable 4: Dashboard Power BI

### Estructura del Dashboard

El informe contiene cuatro pestañas principales:  
1. Análisis Salarial: Visualizaciones de salarios mínimos, máximos y predichos.  
2. Job Titles: Puestos más comunes y tendencias de demanda.  
3. Ubicación de Empleos: Distribución geográfica con mapas interactivos.  
4. Empresas: Empresas con mayor número de ofertas.

Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

## 🤖 Entregable Opcional: Chatbot Inteligente (JobBot)

Se ha implementado un chatbot local denominado JobBot, que responde en lenguaje natural a preguntas sobre la base de datos de empleos.  
  
Características:  
- Interfaz conversacional vía Streamlit.  
- Traducción de preguntas en lenguaje natural a SQL mediante un modelo LLM local (Mistral 7B).  
- Integración con la base de datos PostgreSQL.  
- En caso de error, responde con frases irónicas tipo “Skynet” en español para añadir humor al proceso.

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.

## 🧠 Próximos Pasos

A continuación se detallan algunas mejoras previstas para futuras versiones del proyecto:  
  
- Resolver problemas que he tapado y solo yo sé que están.  
- Agrupar las keywords en menos categorías.  
- Hacer que el código acepte múltiples tipos de trabajos.  
- Borrar empleos antiguos de la base de datos.  
- Mejorar la precisión y contextualización del chatbot.  
- Dockerizar el proyecto para facilitar su despliegue.  
- Desplegar una versión de la aplicación en la nube.