



# INTRODUCCIÓN A LA CIENCIA DE DATOS

## CREACIÓN DE UN PERFIL PROFESIONAL. TRAYECTORIA DE UN CIENTÍFICO DE DATOS ACTIVIDAD GRUPAL

CIENCIA DE DATOS, UNIVERSIDAD EUROPEA

### ACTIVIDAD GRUPO 3

AUTORES: GEMA FERNÁNDEZ CAMPOS, LUNA ESTEBAN GARCÍA, M.<sup>a</sup> DEL CARMEN RODRÍGUEZ PÉREZ  
EDUARDO PÁEZ RODRÍGUEZ, JUAN MANUEL GRUESO RIVELLES



# Índice

<b>Informe Grupal – Caso Estudio</b> .....	2
<b>Líder de datos</b> .....	2
<i>Responsabilidades, metodologías e interacciones de equipo aplicadas al caso de estudio</i> .....	3
<b>Ingeniero de datos</b> .....	4
<i>Responsabilidades, metodologías e interacciones de equipo aplicadas al caso de estudio</i> .....	5
<b>Analista de datos</b> .....	6
<i>Rol y responsabilidades</i> .....	6
<i>Metodologías e interacción con otros roles</i> .....	6
<b>Especialista en visualización de datos</b> .....	8
<i>Responsabilidades</i> .....	8
<i>Interacción con otros roles</i> .....	8
<b>Administrador de datos</b> .....	9
<i>Rol y responsabilidades</i> .....	9
<i>Responsabilidades, metodologías e interacciones de equipo aplicadas al caso de estudio</i> .....	9
<b>Bibliografía</b> .....	12

# **Informe Grupal – Caso Estudio**

El objetivo principal de este proyecto es desarrollar una herramienta digital, junto con su base de datos, para el departamento de Recursos Humanos (RRHH) de una empresa mediana del sector industrial. La herramienta permitirá la gestión de contratos para empleados de la empresa en cuestión.

Hasta ahora esta tarea de gestión y listado de los contratos se realizaba utilizando Excel de forma muy básica.

Esta herramienta consistirá en un formulario con restricciones específicas para asegurar que la información contractual ingresada sea completa y correcta antes de almacenarse en la base de datos. Los campos de este formulario incluirán datos esenciales como el Id del empleado, edificio, área, país, educación, género, jornada, modalidad, fecha de contratación, estatus y fecha de baja. La implementación de esta herramienta facilitará la administración y actualización de datos, mejorando la precisión y eficiencia de la gestión de contratos en RRHH.

## **Líder de datos**

El rol principal de un líder de datos sería coordinar y guiar al equipo de ciencia de datos, asegurando que el proyecto cumpla con los objetivos de la empresa, se mantenga en el cronograma y cumpla los requerimientos específicos del cliente, en este caso, RRHH.

El líder también se encargaría de gestionar la colaboración entre los roles y facilitar la comunicación entre ellos para asegurar una ejecución fluida del proyecto.

En este caso, como se comenta dentro de las fases, será el encargado de recibir el feedback de cualquiera de los roles, escuchar y esforzarse en entender cual es el problema, para tomar decisiones, que pueden ir desde suministrar mas recursos, organizar una reunión no programada para tratarlo con el resto de roles, etc..., hasta reasignar tareas si es necesario. Siempre priorizando intentar cumplir los plazos dados al cliente, pero manteniendo la motivación de todo el equipo.

## *Responsabilidades, metodologías e interacciones de equipo aplicadas al caso de estudio.*

- **Fases iniciales: Definición del problema & recopilación de datos:**

Al inicio del proyecto, el Líder de datos debe encargarse de organizar un kick-off meeting (reunión inicial) con el cliente que ha solicitado el proyecto para su dpto. de RRHH para definir y entender lo que quiere y necesita. En esta reunión es imprescindible que acudan tanto los stakeholders del cliente, como los de nuestra empresa. Ya que en esa reunión inicial, el resto de roles, como expertos, pueden ayudar a definir mejor los KPI's que necesita la empresa, u ofrecer otras posibilidades (mejoras) que puede el cliente no hubiera pensado inicialmente.

Una vez esta bien definido el proyecto, y tenemos recopilados los datos, se realiza otra reunión inicial ya con el equipo, y mediante técnicas de toma de decisiones (brainstorming, isikawa, etc...), se definirán las fases.

El Líder debe dividir las tareas de cada una de esas fases, tanto para asignarlas a cada uno de los roles, como para poder crear un timing, donde se mostrará los plazos para cada una de las fases, así como quien es o son los responsables de cada una de esas fases. En ese timing también se programarán reuniones para revisar el status de cada una de esas fases, así como estudiar posibles problemas o incluso mejoras que se les haya ocurrido al resto de roles, para tratarlas con el equipo.

- **Fases intermedias: Revisión con el equipo del draft del proyecto, y posterior exposición al cliente para su validación.**

Puesto que habrá una reunión en la fase intermedia del proyecto, se mostrará un draft, donde tanto el Líder como el resto de roles, valorarán si correlaciona con la expectativa del cliente, y una vez revisado y validado por el equipo, se programará una reunión con el cliente, para enseñarle este draft. En esta reunión, el líder expondrá el proyecto al cliente, y en que fase se encuentra. Una vez validado por el cliente este draft, o teniendo en cuenta sus sugerencias en caso de que las haya. Se volverá a realizar otra reunión con el equipo, para dar el feedback del cliente, y proceder con la siguiente fase.

- **Fase final: Versión final de la aplicación y “entrega de llaves”**

En esta fase, ya debemos tener la herramienta depurada y finalizada. Antes de entregarla al cliente, todo el equipo debemos hacer una reunión final, donde cada Rol dará un feedback de los insights de cada una de las tareas del proyecto, de las cuales el equipo tomará las principales para exponérselas al cliente. Esto es una forma de mostrar la herramienta final al cliente, ensalzando el poder que tiene esta nueva herramienta para RRHH.

Además, se ofrece al cliente una formación inicial de uso de la herramienta para todos los usuarios.

Una vez entregada y validada al cliente la nueva herramienta para su Dpto de RRHH, comienzo el periodo contratado para mantenimiento de esta herramienta (acordado en el kick-off meeting), y se firman las cláusulas de este contrato de mantenimiento y actualización.

Al cliente se le suministra el contacto directo del Administrador de datos, para que funcione como “Helpdesk” para cualquier usuario de esta herramienta. Reduciendo así la burocracia y facilitando un contacto directo y rápido para el cliente.

## **Ingeniero de datos**

Como miembro del grupo de proceso de datos, el rol específico consiste en interpretar las necesidades del negocio planteadas por el departamento de RRHH, colaborar en la definición de los objetivos del análisis y asegurar que los resultados obtenidos sean relevantes y aplicables. En este contexto, debe garantizar que los sistemas diseñados cumplan con los requisitos de integración, seguridad y eficiencia necesarios para este proyecto.

Se encarga de diseñar y mantener las bases de datos que almacenan información clave sobre empleados, contratos, beneficios y métricas de desempeño. En este proyecto, la función principal es estructurar el sistema de datos que facilite el análisis realizado por el equipo de analistas y la automatización de procesos operativos, como la generación de informes y seguimiento de nóminas.

El proyecto se desarrolla utilizando un enfoque colaborativo. Como ingeniero de datos diseñar e implementar las bases de datos y flujos ETL necesarios. Junto al equipo validar los resultados con el departamento de RRHH y ajustar el sistema según las necesidades detectadas.

Respecto a las técnicas, experiencia en modelado de bases de datos relacionales (SQL Server) y no relacionales (MongoDB). Y también conocimiento de procesos ETL (Extracción, Transformación y Carga) para manejar datos provenientes de diversas fuentes como Excel o sistemas heredados. Es importante la capacidad para interpretar necesidades de negocio y traducirlas en modelos de datos. Además los conocimientos básicos de análisis predictivo y machine learning aplicado a datos de empleados (e.g., predicción de rotación). Otras herramientas transversales como la comunicación efectiva para colaborar con equipos de desarrollo, RRHH y dirección y la atención al detalle para garantizar la calidad y seguridad de los datos.

Estas habilidades permiten manejar grandes volúmenes de datos de empleados y transformar sistemas rudimentarios (como Excel) en soluciones robustas.

### *Responsabilidades, metodologías e interacciones de equipo aplicadas al caso de estudio.*

Como ingeniero de datos, el rol en el grupo es fundamental para garantizar la calidad y utilidad de la información procesada en el proyecto. Colaborar estrechamente con el resto del equipo para la extracción y limpieza de datos, asegurando de que sean relevantes y estén listos para su análisis. La labor incluye identificar patrones y tendencias clave que puedan influir en las decisiones estratégicas del departamento de RRHH, desarrollando visualizaciones efectivas para comunicar resultados y facilitar su interpretación.

Además, mantener una interacción constante con el equipo para coordinar esfuerzos y ajustar enfoques. Trabajar con el especialista en visualización para transformar los *insights* en *dashboards* claros y funcionales, y participar en reuniones con el líder de datos y el cliente para alinear los objetivos del proyecto con las necesidades específicas de RRHH. El enfoque debe ser colaborativo y garantizar que las recomendaciones estratégicas basadas en los datos sean aplicables y de alto impacto para la organización.

1. **Diseño y mantenimiento de bases de datos:** Crear estructuras optimizadas para almacenar datos sobre contratos, desempeño, compensaciones y beneficios.
2. **Integración de sistemas:** Facilitar la conexión entre herramientas de RRHH y la base de datos (e.g., software de nómina o gestión de talento).
3. **Soporte analítico:** Proveer datos limpios y procesados para análisis que mejoren la gestión de talento.
4. **Automatización:** Configurar flujos automatizados para actualizaciones de datos y generación de reportes.

Este proyecto con el rol de ingeniero de datos es un ejemplo de cómo la colaboración del equipo multidisciplinar de proceso de datos, permite alinear la tecnología con los objetivos estratégicos de los departamentos empresariales. Además, refuerza la importancia de la gestión basada en datos para la toma de decisiones informadas en recursos humanos.

## Analista de datos

El analista de datos tiene un papel crucial en el proyecto, sirviendo como puente entre los datos brutos generados y las necesidades del cliente (departamento de RRHH). Las responsabilidades específicas son:

### *Rol y responsabilidades*

- Calidad y validación de datos: Verificar que los datos sean completos, precisos y útiles.
- Interpretar los algoritmos: Identificar patrones o tendencias relevantes, como tipos de contratos más utilizados o áreas con mayor rotación de empleados.
- Definición de KPIs: Desarrollar indicadores para evaluar la gestión de contratos.

### *Metodologías e interacción con otros roles*

#### **Fase inicial: Definición del problema y recopilación de datos**

- Se sienta con el cliente (RRHH) y el equipo para entender qué datos son importantes y cómo se usarán.
- Revisa los datos que ya existen para asegurarse de que no estén incompletos o llenos de errores.

- Ayuda a decidir cómo organizar los datos en la base para que sean útiles y fáciles de analizar.
- **Interacciones**
  - Con el arquitecto e ingeniero de datos para armar bien la estructura.
  - Con el administrador para asegurarse de que los datos importantes estén bien protegidos.

### **Fase intermedia: Procesamiento, análisis y testeo**

- Examina los datos procesados para asegurarse de que todo esté bien ordenado y limpio.
- Busca patrones o tendencias, como por ejemplo, áreas con más rotación o qué tipo de contratos son más comunes.
- Crea reportes y gráficos para mostrar resultados iniciales y verifica con el cliente si está satisfecho.
- **Interacciones**
  - Con el especialista en visualización para diseñar dashboards claros y útiles.
  - Con el líder de datos y el cliente para ajustar cualquier análisis según los comentarios que den.

### **Fase final: Versión final y entrega**

- Confirma que todos los datos y reportes que salen de la herramienta estén bien y sean fáciles de entender.
- Prepara gráficos y reportes finales que resuman cosas clave, como rotación, duración de contratos, etc.
- Explica al cliente cómo usar los dashboards y sacarles el máximo provecho para tomar decisiones.
- **Interacciones**
  - Con el administrador para que todo esté bien integrado en la herramienta.



- Con el especialista en visualización para que los gráficos finales queden impecables.

## Especialista en visualización de datos

Un especialista en visualización de datos en este proyecto será responsable de transformar la información almacenada en la base de datos en representaciones visuales significativas, esto ayudará a RRHH y otros departamentos para analizar, entender y comunicar tendencias, métricas de contratación, índices de rotación de empleados, y distribución por áreas, educación, género, modalidad de trabajo, etc.

### Responsabilidades

- Creación de visualizaciones fáciles de comprender y dinámicas → generación de *dashboards* que permitan a los usuarios de RRHH monitorear la información recopilada de los contratos a tiempo real.
- Dar soporte a la hora de la interpretación de las visualizaciones, no técnico.
- Exploración de datos para descubrir patrones → analizar de manera visual los datos de los contratos para analizar tendencias y patrones relevantes, como estacionalidad en la contratación, duración promedio de los empleados, y factores asociados a bajas o rotación.

### Interacción con otros roles

- **Líder de equipo:** le dejará claro los objetivos del proyecto en relación con su puesto, facilitará la priorización de tareas y asegurará que las visualizaciones cumplan con las necesidades de negocio.
- **Analista de datos:** le proporcionará los datos ya analizados y los *insights* que deberán ser visualizados.
- **Ingeniero de datos:** se asegurará de que la data esté limpia y lista para ser estudiada y utilizada. El especialista trabajará en estrecha colaboración con el ingeniero para establecer y validar *pipelines* de datos que alimenten las visualizaciones en tiempo real o de manera periódica.
- **Administrador de datos:** dará soporte creando una herramienta para ayudar a la visualización y accesibilidad de los datos, y colaborará con el especialista en visualización para optimizar consultas y garantizar la compatibilidad de los datos con las herramientas de visualización.

# Administrador de datos

## *Rol y responsabilidades*

El administrador de datos tiene un perfil de gestión, mantenimiento y control, durante todo el proyecto, en las siguientes áreas:

- Gestión, mantenimiento y control de la base de datos.
- Gestión, mantenimiento y control de la seguridad de los datos.
- Cumplimiento de la legislación y protección de datos.
- Asegurar la calidad del producto y la resolución de incidencias.

## *Responsabilidades, metodologías e interacciones de equipo aplicadas al caso de estudio.*

### • **Fases iniciales: Definición del problema & recopilación de datos:**

Al inicio del proyecto lo primordial es entender el negocio, comprender los datos, y queremos saber que se espera de la aplicación y que requerimientos mínimos necesita. Para ello nos reuniremos con el cliente (parte interesada o "stakeholder"), en este caso el departamento de recursos humanos de la empresa en cuestión.

Particularmente, en esta fase inicial, las tareas del administrador de datos serán entender las necesidades con respecto a la seguridad, la normativa de protección de datos y la calidad mínima esperadas.

Por ejemplo, deberíamos aclarar cuales son las variables o datos más relevantes para ellos: el ID, el tipo de contrato... y que datos deberían estar encriptados según en qué nivel de seguridad, y cuántos niveles de seguridad quieren tener.

Además, fundamental será apoyar al ingeniero y al analista de datos en la definición, selección y depuración de los datos, con el fin de tener una BBDD y aplicación robustas y ágiles.

Con respecto a las herramientas, en esta fase inicial del proyecto principalmente utilizaremos el lenguaje SQL y bases de datos relacionales que trabajen con este lenguaje, como por ejemplo el SQL Server.

Las habilidades de comunicación, creatividad y generación de ideas (softskills) serán también fundamentales en esta fase inicial, de elevada interacción personal y directa con el cliente.

- **Fases intermedias: Procesamiento de datos, análisis , interpretación de los resultados (testeo e iteraciones):**

El proyecto avanzará, y eso significa que los datos están limpios y etiquetados, habremos empezado con los primeros modelados, análisis, testeos y presentaciones de resultados.

Con la primera versión de prueba, es muy importante recibir la primera impresión del cliente, saber si vamos por el buen camino y que hay que corregir y mejorar.

Para el administrador de datos esto significará que la aplicación funciona con al menos los mínimos de calidad y seguridad requeridos, pero aún queda mejorarla. El cliente nos puede indicar dónde y responderemos haciendo propuestas.

Con respecto a la BBDD haremos iteraciones y depuraciones de su funcionamiento junto con el ingeniero de datos.

A la vez contrastaremos con el analista de datos si los modelos funcionan, mostrando al cliente los resultados esperados. Por ejemplo, el analista podría hacer análisis sobre la duración media de los contratos, y buscar patrones no evidentes en los datos.

Además, será importante la presentación y visualización de los datos, la cual la revisaremos cualitativamente con el especialista en visualización, los cuales serán presentados en un dashboard de Power BI a nuestro cliente. Por ejemplo, tendremos visualizaciones atractivas y gráficos de las grandes cifras: número de empleados totales, empleados según en qué oficinas, tipos de contratos etc.

En paralelo revisaremos que se cumple las normativas, la protección de datos (GDPR-UE), estaremos alineados con respecto a la seguridad y los permisos de acceso, para lo cual podemos utilizar las funciones que ofrece el SQL Server para ello. Por ejemplo, definiríamos dos niveles de seguridad: el básico para todos los empleados y el avanzado, para RRHH y los managers o jefes de departamento.

Además, estaremos encriptando los datos utilizando las opciones que nos da SQL server para esto y haciendo copias de seguridad regularmente.

Sobre todo esto, el administrador de datos irá reportando de los avances e incidencias al líder de datos, con el objetivo de cumplir los hitos y fechas de entrega del mismo.

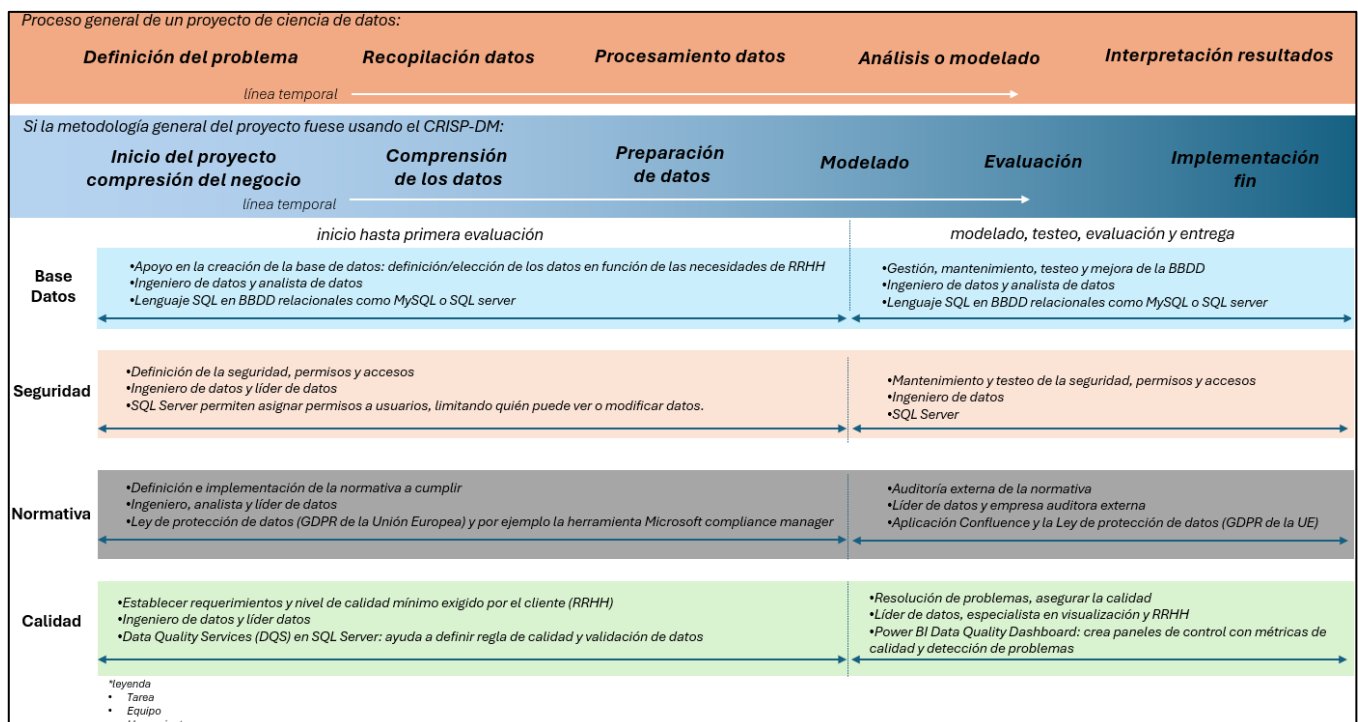
- **Fase final: Versión final de la aplicación y “entrega de llaves”**

Tras varias iteraciones, y si todo va correctamente, habremos obtenido una aplicación robusta, que funciona tal y como el cliente espera.

Como administrador de datos habremos ido solucionando incidencias y problemas, interactuando con el equipo según la naturaleza de la incidencia, usando las diferentes herramientas, desde las BBDD con SQL, las funciones de SQL server para permisos y accesos, así como herramientas específicas de calidad, como el Power BI Data Quality Dashboard, y validación de datos como el "Data Quality Services" (DQS), integrado en el SQL Server.

Finalmente, se le entrega la versión final de la aplicación al cliente, ofreciéndole un mantenimiento, actualización y mejora de esta, de al menos 2 años, pudiendo hacer prórrogas de esta.

*El esquema describe el rol del administrador de datos en un proyecto de CD siguiendo la metodología CRISP-DM, diferenciando responsabilidades desde las fases iniciales hasta el modelado y testeos, y las fases posteriores. Se enfatiza que las tareas pueden solaparse en el tiempo, sin un inicio o final exacto alineado con las fases del proyecto.*



## Bibliografía

- *Apuntes de clase, Universidad Europea, Asignatura: Introducción a la ciencia de datos.*
- SQL server: [https://es.wikipedia.org/wiki/Microsoft\\_SQL\\_Server](https://es.wikipedia.org/wiki/Microsoft_SQL_Server)
- DQS: <https://learn.microsoft.com/es-es/sql/data-quality-services/data-quality-services?view=sql-server-ver16>
- Power BI: [https://es.wikipedia.org/wiki/Power\\_BI](https://es.wikipedia.org/wiki/Power_BI)