

Laboratorio II

[Repositorio](#)

El trabajo se realizó sobre un [conjunto de datos](#) proveniente de Kaggle que contiene información de una empresa con miles de empleados distribuidos en más de veinticinco departamentos y múltiples países. Este dataset incluye datos demográficos, información departamental, ubicaciones geográficas, historiales de promociones, salarios anuales y las relaciones jerárquicas entre gerentes y empleados. Dado que el dataset original se encontraba en un estado limpio y bien estructurado, las transformaciones necesarias fueron mínimas. Durante la fase de transformación se estandarizaron los tipos de datos para cada columna según su naturaleza, se crearon nuevas medidas calculadas para facilitar el análisis de indicadores clave de desempeño (KPIs) como totales salariales, promedios ponderados y métricas de compensación ajustadas por costo de vida, y se incorporó una tabla calendario vinculada a la fecha de contratación de los empleados, permitiendo así análisis temporales sofisticados con capacidad de filtrado por año. Cabe mencionar que se prescindió de utilizar la tabla organizacional (“org_edges”) que relacionaba gerentes con empleados, reservando su implementación para futuros desarrollos que requieran representar la estructura jerárquica completa de la organización.

El modelo de datos resultante se presenta en un esquema de estrella donde la tabla “employees” funciona como tabla de hechos central, conectándose con cinco tablas dimensionales. Tres de estas relaciones siguen una cardinalidad de muchos a uno: “employees” se vincula con “departments” mediante “department_id”, con “locations” a través de “location_id”, y con “calendar” usando “hire_date”, lo que permite que múltiples empleados compartan departamento, ubicación o fecha de contratación, pero cada empleado pertenece a valores únicos en estas dimensiones. Las tablas “promotions” y “salaries_annual” presentan relaciones de uno a muchos desde “employees” mediante “employee_id”, permitiendo que cada empleado tenga múltiples registros históricos de promociones y salarios anuales que documentan su trayectoria laboral. Las relaciones, cardinalidades y direcciones de filtros cruzados detectadas automáticamente fueron validadas y resultaron correctas, habilitando el desarrollo de un dashboard interactivo que permite consultas multidimensionales sobre distribución geográfica de empleados, análisis de compensación por nivel de seniority, evolución salarial histórica, y comparaciones departamentales de estructura organizacional, todo esto con capacidad de filtrado dinámico por país, año y departamento.