

Práctica 10: Spark

EJERCICIO 1: CONFIGURACIÓN Y LECTURA DE DATOS

Crea una sesión de Spark y carga el archivo empleados.csv definiendo un esquema explícito.

EJERCICIO 2: INSPECCIÓN DE DATOS

Examina la estructura del DataFrame:

- Muestra el esquema
- Visualiza las primeras 10 filas
- Obtén estadísticas básicas.

EJERCICIO 3: SELECCIÓN Y PROYECCIÓN

Selecciona las las siguientes columnas de empleados: nombre, apellido y departamento. Luego crea una selección que incluya solo empleados del departamento de "Ventas".

EJERCICIO 4: TRANSFORMACIÓN DE COLUMNAS

Crea una nueva columna que combine nombre y apellido, por ejempl: "Pepe García", y otra columna que calcule un bono del 5% sobre el salario. Renombra la columna "puesto" a "cargo".

EJERCICIO 5: FILTRADO BÁSICO

- Filtra los empleados que tienen un salario mayor a 45000 y muestra su nombre, apellido, departamento y salario.
- Filtra aquellos empleados que trabajan en "Madrid" y muestra su nombre, apellido, ciudad y puesto.
- Filtra los empleados que están activos en la empresa y muestra su nombre, apellido y activo

EJERCICIO 6: AGRUPACIONES SIMPLES

Agrupar los empleados por departamento y calcula: el número de empleados, el salario promedio y el salario máximo por departamento.

EJERCICIO 7: ORDENAMIENTO

Ordena los empleados por salario de forma descendente y luego por nombre de forma ascendente. Muestra solo los top 10 empleados con mayor salario.

EJERCICIO 8: FUNCIONES DE STRING

Utiliza funciones de string para añadir al dataframe las siguientes columnas:

- convertir los nombres a mayúsculas
- crear emails corporativos con el formato:
nombre.apellido@empresa.com
- extraer las primeras tres letras del apellido.

EJERCICIO 9: FUNCIONES DE FECHA

Calcula la antigüedad de cada empleado en días y añádela como una columna nueva.

EJERCICIO 10: CONSULTAS SQL

Crea una vista temporal y ejecuta una consulta SQL que muestre nombre, departamento y salario de empleados con salario mayor a 40000, ordenados por salario descendente.

EJERCICIO 11: ESCRITURA

Guarda en un CSV el DataFrame obtenido en el ejercicio 8