

Examen 17 de diciembre 2025

Duración: 30 minutos

Instrucciones: Lee cuidadosamente cada escenario y resuelve los ejercicios utilizando operaciones de DataFrame de Spark.

Puntuación: El examen tiene una puntuación de 4 puntos

EJERCICIO 1 (0.5 puntos)

- Leer los csv de películas y ventas, indicando el esquema de forma explícita en ambos casos.
- Rellena los valores nulos de ventas con los valores por defecto:
 - cantidad: 2
 - metodo_pago: "Tarjeta"
- Rellena los valores nulos de películas con los valores por defecto:
 - duracion_min: 120.0
 - precio: 15,95
- Obtener todas las películas del género 'Drama' que sean posteriores al año 2000. Mostrar solo las columnas: id, titulo, genero, año (anyo) y ordenar los resultados por año (anyo) de forma ascendente
- Seleccionar de ventas las columnas: id_venta, id_pelicula, fecha, cantidad y crear una nueva columna llamada mes que extraiga el mes de la columna fecha. Mostrar solo los primeros 10 registros

EJERCICIO 2 (0.5 punto)

- Mostrar las 5 películas más caras, indicando las columnas: titulo, genero y precio
- Contar cuántas películas hay de cada género, mostrando dos columnas: género y el total de las películas. Ordenar los resultados de mayor a menor usando el total de las películas

EJERCICIO 3 (1 punto)

- Mostrar información de las ventas de enero 2025 junto con las películas que se han vendido. El resultado debe contener las columnas:
 - id_venta, fecha, cantidad, titulo, genero, precio
 - Ordenar por id_venta ascendente
 - Muestra solo los primeros 15 registros
- Calcular el total de unidades vendidas por película. Para cada película mostrar: id_pelicula, el total de unidades vendidas (suma de todas las cantidades). Ordena el resultado de mayor a menor usando el total de las unidades vendidas. Muestra solo las 5 películas más vendidas

EJERCICIO 4 (2 puntos)

- Calcular para cada género:



Ingeniería de Sistemas Intensivos en Datos
Curso 2025/2026

- Número total de películas
- Duración promedio (redondeado a 1 decimal)
- Precio promedio (redondeado a 2 decimales)
- El resultado debe estar ordenado por el número total de películas descendente

(nota: Se deben mostrar las columnas: genero, número total películas, duración promedio, precio promedio)

- Realizar un análisis financiero completo de ventas por género. Para ello obtén los ingresos por venta de películas que se calcula como (cantidad * precio). A continuación, agrupa por género de película y calcula:
 - Cantidad total de ventas realizadas
 - Cantidad total de unidades vendidas
 - Ingresos totales por ventas (redondeados a 2 decimales)
 - Ingresos promedios por venta (redondeado a 2 decimales)

(Nota: Mostrar género, cantidad de ventas, cantidad de unidades vendidas, ingresos totales por ventas e ingresos promedios por venta, además ordena el resultado por los ingresos totales de manera descendente)