

Ejercicio 1

Crear un flujo de trabajo que debe realizar las siguientes operaciones:

- Leer el archivo data1.txt con la columna "ranking" como String y denominada "marks";
- Eliminar los comentarios iniciales de los datos leídos del archivo;
- Eliminar la columna "class"
- Escribir los datos finales en el archivo en formato CSV (por ejemplo, con el nombre "data1_new.csv") utilizando el carácter ";" como separador

Escribe una breve descripción de todos los nodos del flujo de trabajo.

Guarda y ejecuta el workflow. La ejecución debe realizarse sin errores (luces verdes para todos los nodos).

Ejercicio 2

Partiendo del fichero CSV resultado del ejercicio anterior. Ahora crea un workflow que haga lo siguiente:

- Leer el archivo CSV y cambiar el nombre de la columna "marks" a "ranking"
- Filtrar las filas con valor 'average' en la columna 'comments'
- Excluir las columnas con datos del tipo Integer
- Escribir los datos finales en el archivo en modo "Append" y con tab como carácter de separación
- Cambia el nombre de todos los nodos cuando sea necesario. Guarda y ejecuta el workflow.

Ejercicio 3

A partir del dataset del Titanic, se pide realizar un workflow que efectúe los siguientes pasos:

- Cargar el fichero CSV desde una ruta relativa al *workflow*, en su subcarpeta *data*.
- Convertir la columna *Survived* a *string*
- Convertir la columna *Sex* a tipo entero (*Category To Number*), sobrescribiendo la propia columna
- Eliminar la columna *Cabin*, porque tiene demasiados nulos
- Reemplazar las edades nulas por la media de edad y los *Embarked* nulos por el valor más frecuente de esa columna (puedes hacerlo todo junto con un solo *Missing Value*)
- Guarda el resultado procesado con un *CSV Writer* en un fichero en la misma carpeta que el original

- Paralelamente, desde el penúltimo nodo, filtra los supervivientes, y quédate con la columna *Pclass*. Convierte este dato a *string* y saca un gráfico de barras de cuántos supervivientes hubo por clase.

Ejercicio 4

Vamos a realizar un proyecto más completo, con más fases. Recopilaremos datos de diferentes fuentes, los limpiaremos y combinaremos para producir un resultado final. Usaremos el fichero del dataset para el ejercicio 4, en él veremos varios archivos que son los que nos interesan:

- Product Info: información de los productos en venta de una compañía (bebidas energéticas), se indican precios de venta y coste de fabricación en cada país donde se venden.
- Sales Rep: información sobre las diferentes regiones de venta, los productos que se venden en cada región y quién es el responsable o jefe de ventas de ese producto en esa región.
- Finalmente, un conjunto de archivos CSV con las ventas realizadas en diversos meses. Se indican en cada registro su ID, la región donde se vendió, el nombre de producto que se vendió, país y estado, cantidad, fecha y precio que se pagó.

Crea un workflow group donde añadiremos 3 workflows diferentes. El objetivo es conseguir en cada uno lo siguiente:

- UnificacionTransacciones:
 - Usa *List Files/Folders* y especifica la carpeta donde están los ficheros. En Filter options configurar archivos con extensión csv y que empiecen por Transactions.
 - Itera sobre esos ficheros usando Table Row to Variable Loop Start
 - Luego usaremos un CSV Reader conectado a la salida del nodo anterior mediante *Show Flow Variable Ports* No obstante, le diremos que lea la variable Path que proporciona el nodo anterior.
 - Finaliza con un Loop End y luego un CSV Writer. Pídele que sobrescriba el fichero si ya existía.
- ProcesarSalesRep:
 - Lee el fichero Sales Rep.xlsx.
 - Lee el fichero RegionDictionary.xlsx.
 - Usa Value Lookup y conecta las salidas de los 2 nodos anteriores a este. Configuración: Lookup column: (Sales Area) - Key column: (Misspelled) Lookup column output: (Replace) - Replacement column: (Regions)
 - Escribe el fichero excel resultante, incluyendo sus cabeceras y eligiendo sobrescribir si ya existe.
- CombinacionDatos:
 - Carga el CSV generado en el 1er workflow.

- Cambiar “Price Paid” a numérica, tendrás que reemplazar las comas decimales por puntos para poder hacer la conversión. Usa un String Replacer y luego String To Number.
- Lee el excel generado en el 2º workflow.
- Añade un “Joiner” con esta configuración:
 - Left Outer Join
 - Sales Area con Sales Area
 - Product Name con Product Name
 - Merge join columns
- Lee el fichero Product Info.xlsx.
- Reemplaza los valores vacíos de la columna Country con el valor de la fila anterior.
- Añade otro joiner para unificar la salida del 1er joiner con la salida del procesamiento de los productos, con esta configuración:
 - Left Outer Join
 - Country con Country
 - Product Name con Product Name
 - Merge join columns
- Limpieza de datos:
 - Eliminar filas duplicadas
 - Eliminar filas con “Quantity” 0 o negativo mediante Row Filter.
 - Eliminar filas con Product Name a #NV
 - Calcular beneficio por fila usando Math Formula. La operación es: precio pagado menos coste de producción unitario, todo ello multiplicado por la cantidad. Indicar que el cálculo se almacene en una columna llamada “Revenue”.
 - Guardar el resultado final en un fichero CSV.
 - En paralelo antes de escribir el CSV, obtén un gráfico circular de los beneficios acumulados por país.