Ejercicio 3 – Práctica entregable. Análisis de datos con KNIME.

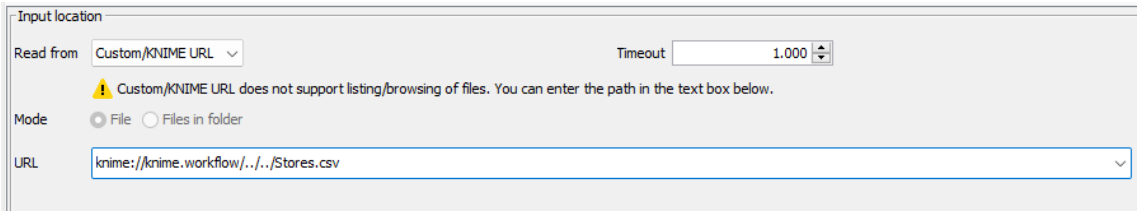
Realizar un workflow en KNIME que cubra los requerimientos que se exponen en los puntos siguientes.

Los **objetivos** de esta práctica son: acceder, limpiar, transformar, fusionar y agregar datos e informar posteriormente de los resultados del análisis realizado usando la plataforma de modelado KNIME.

Caso de uso: El equipo de ventas de una empresa recopila mensualmente datos de las transacciones de sus clientes en diferentes formatos, como por ejemplo el precio total por cliente y el tamaño de su cesta de la compra por tipo de tienda.
Dado que los datos se reciben de diferentes fuentes será necesario limpiarlos y transformarlos antes de poder realizar ningún análisis o informe.

Entrega de la práctica:

- Se debe entregar un único workflow llamado **UT6_EJ3_Apellidos_Nombre.knwf**
- Todas las rutas de acceso a datos, tanto las de los ficheros como la de la BD SQLite se deben configurar de forma que sean relativas al workflow y los ficheros estén ubicados en un nivel superior a la ubicación del workflow.



- Cada uno de los bloques/puntos se deben identificar claramente con anotaciones tal y como se indica en el Bloque 7
- En la mayoría de los puntos os sugiero algunos nodos para obtener los resultados esperados, pero podéis utilizar otros.

Bloque 1: Acceso a datos

Se reciben distintos ficheros y bases de datos que deben leerse.

- CustomerInfoSystem1.xlsx , que tiene datos de clientes.
- CustomerInfoSystem2.table, que tiene datos de clientes.
- Stores.csv, que tiene datos de tiendas.
- Sales.sqlite, que tiene datos de ventas.

Acciones a realizar:

- Lee los tres ficheros con Reader Node(s)
- Conéctate a la BD SQLite y con dos nodos Table Selector carga las tablas **Transactions** y **ProductNrAndPrice**

Observa la estructura de los datos cargados.

Bloque 2: Limpia los datos

- Limpia los datos del primer fichero excluyendo la primera fila y la primera columna con un nodo Table Cropper.
- Elimina las filas con CustomerID duplicado con el nodo Duplicate Row FIlter.
- Cuando no haya edad inserta la media de edad del dataset con un nodo Missing Value

Bloque 3: Transformación de datos (2 puntos)

Objetivo: estandarización y merge de datos de ficheros

- Crea una nueva columna **Age Group** con los valores Adolescente, Adulto, Adulto mayor según el valor de edad con el nodo **Rule Engine** (<18 adolescentes, 18-65 adultos, >65 adultos mayores)
- Reemplaza el carácter "_" con un espacio " " en la columna país **País** con el nodo **String Manipulation**.
- Divide la columna **CustomerID** en el carácter "_" con el nodo **Cell Splitter**

- Cambia el nombre de las columnas creadas por la división con el nodo **Column Renamer.**
- Fusione las columnas **Email** y **Corporate Email** con el nodo **Column Merger**. Nos quedaremos con el Email cuando tengamos ambos.
- Convierte la columna **Newsletter** con el nodo **Number to String.**

Se deben realizar las mismas transformaciones para ambos ficheros.

Bloque 4: Une los datos (2 puntos)

Objetivo: ampliar la información de la tabla de clientes con los datos de la tienda y el precio

- Concatena las dos tablas de clientes con el nodo **Concatenate**.
- Añade el tipo de tienda (StoreType de Stores.csv) y Price (de la tabla ProductNrAndPrice de la DB) a cada transacción (tabla Transactions DB) con dos nodos **Value Lookup**
- Une las tablas **Cliente** y **Transacction** con la columna **CustomerID** con el nodo Joiner.
  Nota: El primer puerto de salida del Joiner muestra las filas que coinciden en las dos tablas. Utiliza sólo estas filas en los siguientes pasos.

Bloque 5: Agregación de datos (2 puntos)

Objetivo: agregar los datos por cliente y tipo de tienda

- Suma el precio para cada **CustomerID** con un nodo **Row Agreggator**.
- Calcula el tamaño del carrito (número de productos en el pedido) con un nodo **GroupBy**
- Quédate con la información del **tipo de tienda**.
- Muestra el número de pedidos por tamaño de carrito en la tienda Online y en la tienda OnSite respectivamente con el nodo **Pivot**.
- Repite el punto en el que se suma el precio con un nodo GroupBy. Quédate con las siguientes columnas: CustomerID, CurtomerGroup, AgeGroup, Suma de Precio, Media de Precio, contados de productos únicos.

Bloque 6: Visualización de datos (2 puntos)

- Exporta a un fichero Excel la tabla que contiene los precios totales por cliente con un nodo **Excel Writer**

Genera un componente llamado **Visualizaciones** que realice las siguientes tareas:

- Convierte el tamaño del carrito con un nodo **Number to String** y muestra en un diagrama de barras la cantidad de pedidos por cada tamaño de carrito, tanto para tienda Online como OnSite
- Asigna un color a cada fila según la columna **CustomerGroup** con el nodo **Color Manager** y en un diagrama de dispersión, muestra la suma (precio total) y el recuento único (Nº de productos)
- Muestra el recuento de apariciones de **CustomerGroup** en un gráfico de barras.
- Muestra un gráfico de coordenadas paralelas con
    - Grupo de edad
    - Grupo de clientes
    - Suma(Price)
    - Recuento único (ProductNr)
    - Crear un título con el nodo Vista de texto
- Exporta los gráficos resultantes a un PDF

Bloque 7: Presentación y documentación (2 puntos)

- Crea anotaciones para cada uno de los bloques o apartados.
- Explica en ellos las operaciones que realizas
- Modifica la descripción de los nodos para que también sea explicativa.