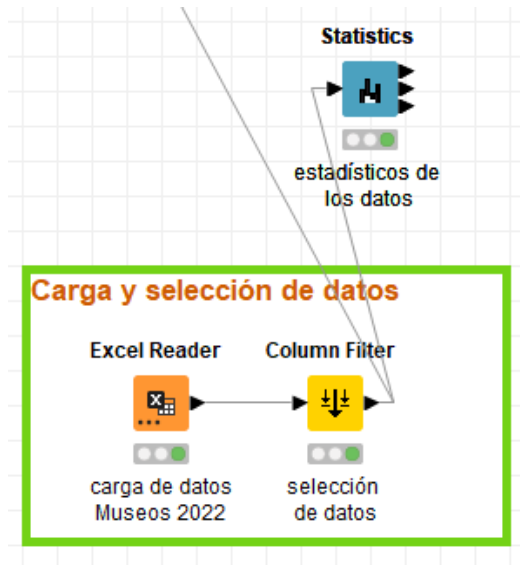


PRÁCTICA MANEJO Y TRATAMIENTO DE DATOS EN KNIME

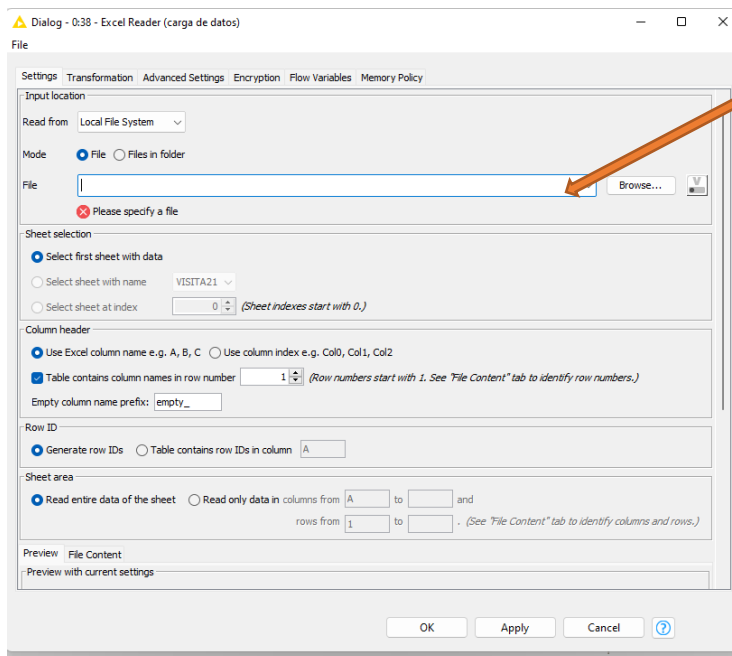
PANTALLAS DE CONFIGURACIÓN

CARGA Y SELECCIÓN DE DATOS



NODO EXCEL READER

Este nodo lee archivos de Excel (formatoxlsx, xlsx, xlsb y xls). Puede leer uno o varios archivos al mismo tiempo, sin embargo, solo lee una hoja por archivo.



Agregar la ruta y el nombre del archivo. En este caso: VISITA21.xlsx

Se especifica que los datos se encuentran en la primera hoja y que los nombres de las columnas están en el renglón 1.

Sheet selection

☒ Select first sheet with data (VISITA21)

☐ Select sheet with name VISITA21

☐ Select sheet at index 0 (Sheet indexes start with 0.)

Column header

☒ Use Excel column name e.g. A, B, C ☐ Use column index e.g. Col0, Col1, Col2

☒ Table contains column names in row number 1 (Row numbers start with 1. See "File Content" tab to identify row numbers.)

Empty column name prefix: empty_

Row ID

☒ Generate row IDs ☐ Table contains row IDs in column A

Sheet area

☒ Read entire data of the sheet ☐ Read only data in columns from A to and rows from 1 to . (See "File Content" tab to identify columns and rows.)

Preview File Content

Dialog - 0:38 - Excel Reader (carga de datos)

File

Settings Transformation Advanced Settings Encryption Flow Variables Memory Policy

☒ Table contains column names in row number 1 (Row numbers start with 1. See "File Content" tab to identify row numbers.)

Empty column name prefix: empty_

Row ID

☒ Generate row IDs ☐ Table contains row IDs in column A

Sheet area

☒ Read entire data of the sheet ☐ Read only data in columns from A to and rows from 1 to . (See "File Content" tab to identify columns and rows.)

Preview File Content

Preview with current settings

The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	[S] ANIO...	[S] ENT_R...	[I] MES_E...	[I] DIA_EN...	[S] SEXO	[I] EDAD	[S] ENT_R...	[S] MUN_R...	[S] PA
Row0	2021	01	7	1	2	57	02	001	000
Row1	2021	01	7	1	2	21	01	001	000
Row2	2021	01	7	1	2	19	32	010	000
Row3	2021	01	7	1	2	22	22	014	000
Row4	2021	01	7	1	1	22	01	010	000
Row5	2021	01	7	1	1	25	01	001	000
Row6	2021	01	7	2	1	35	01	001	000
Row7	2021	01	7	2	1	22	01	001	000
Row8	2021	01	7	2	1	18	01	001	000
Row9	2021	01	7	2	1	43	01	001	000
Row10	2021	01	7	1	2	47	01	001	000
Row11	2021	01	7	1	1	50	08	037	000
Row12	2021	01	7	1	1	51	33	000	221

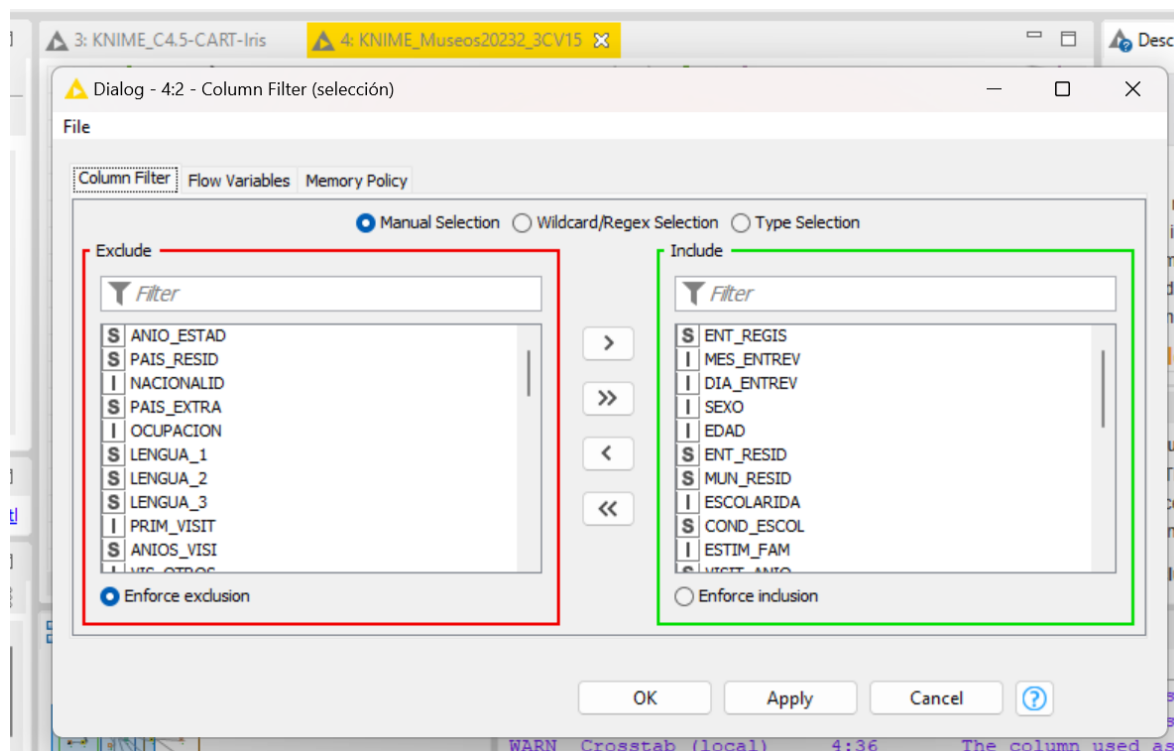
OK Apply Cancel ?

Si se desplaza a la parte inferior de la configuración del nodo, tendrá una vista previa del conjunto de datos que cargó.

NODO COLUMN FILTER

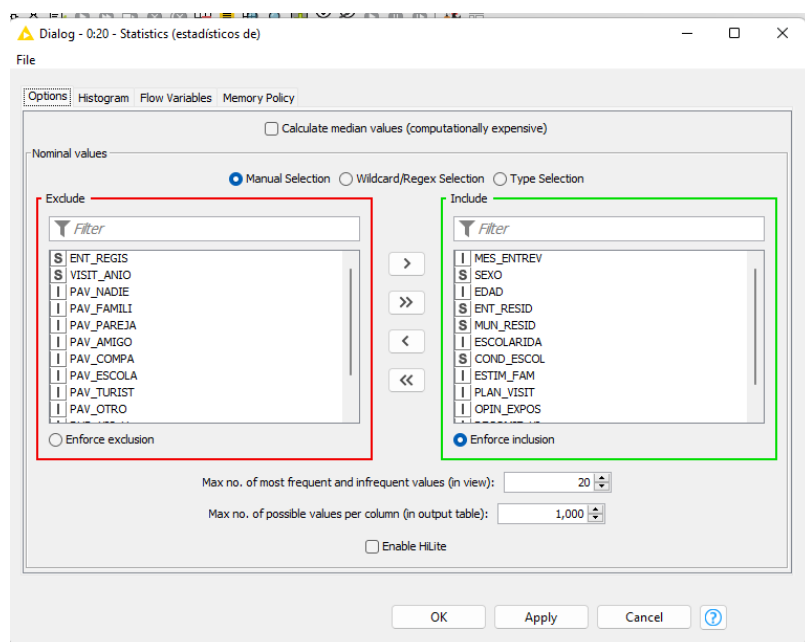
Este nodo permite filtrar las columnas de la tabla de entrada.

Seleccione los atributos indicados en la práctica.



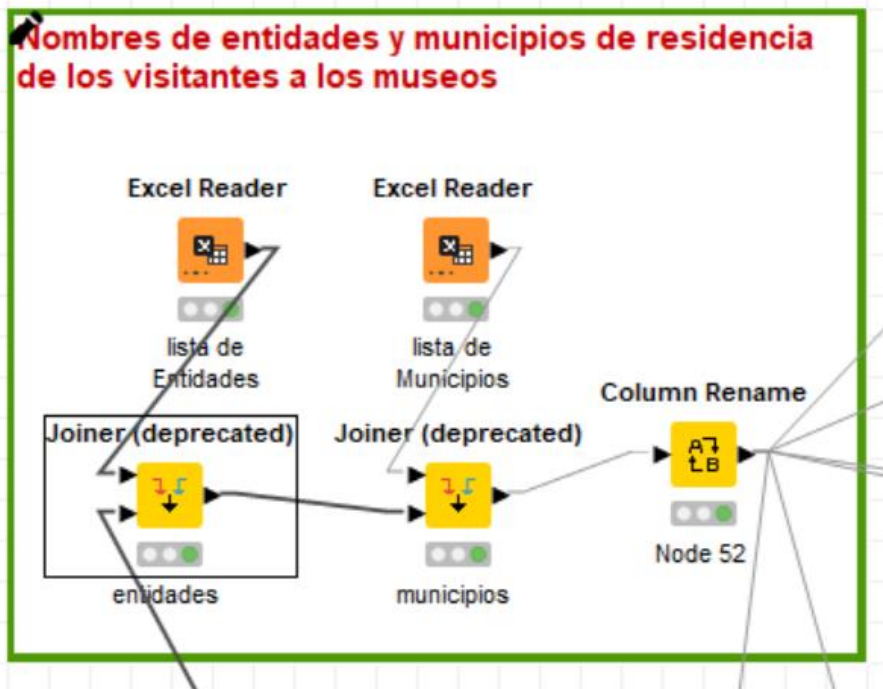
NODO STATISTICS

Este nodo calcula los estadísticos de datos numéricos y categóricos. Agregue los datos seleccionados para que analice su comportamiento.



NOMBRES DE ENTIDADES Y MUNICIPIOS

Esta sección le permitirá agregar los nombres de las entidades y municipios.



NODO EXCEL READER. LISTA DE ENTIDADES

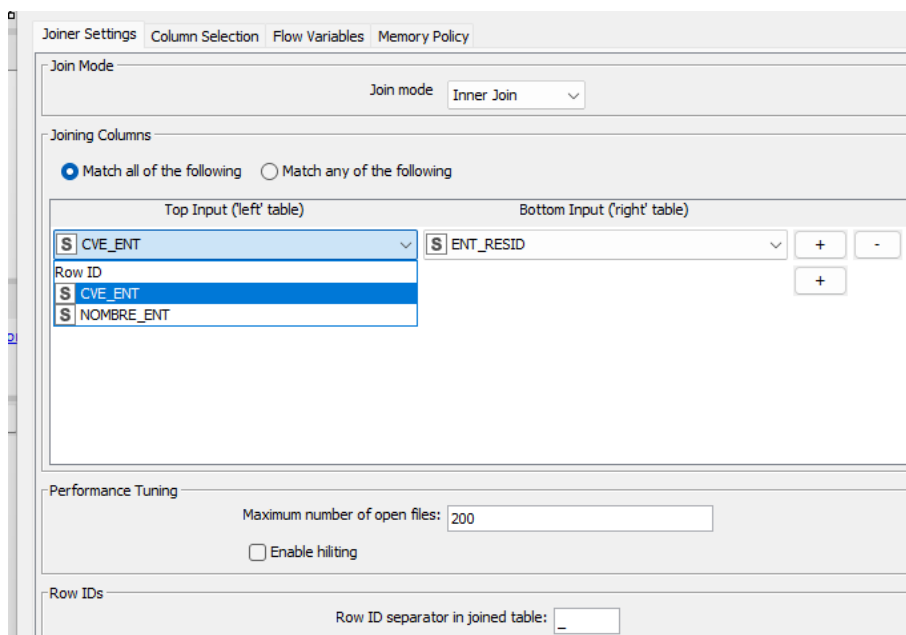
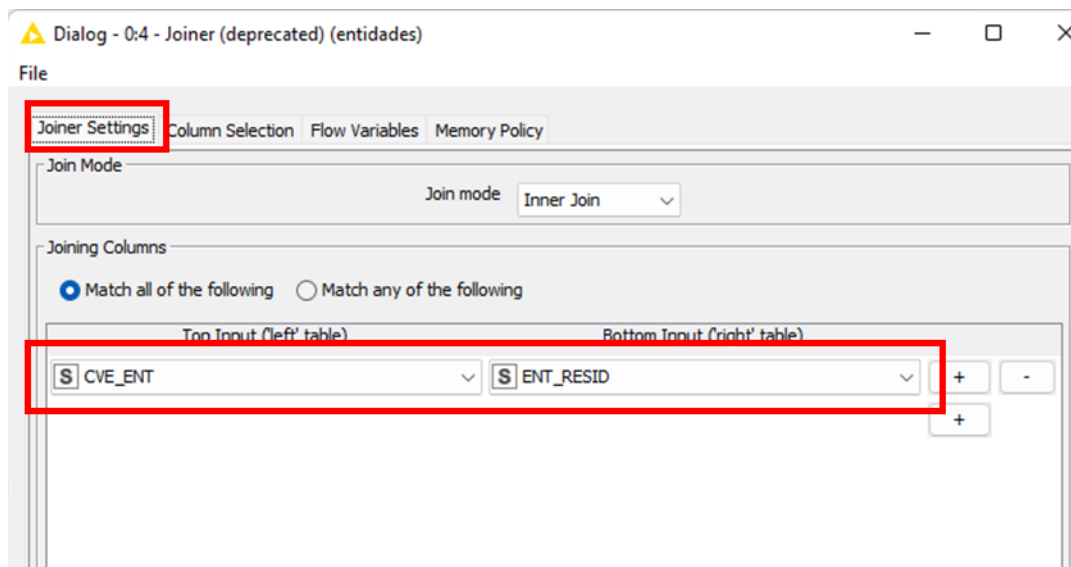
Con este nodo cargará el archivo llamado: CENTIDAD.xlsx y realizará la configuración similar a la del archivo de visitas.

NODO EXCEL READER. LISTA DE MUNICIPIOS

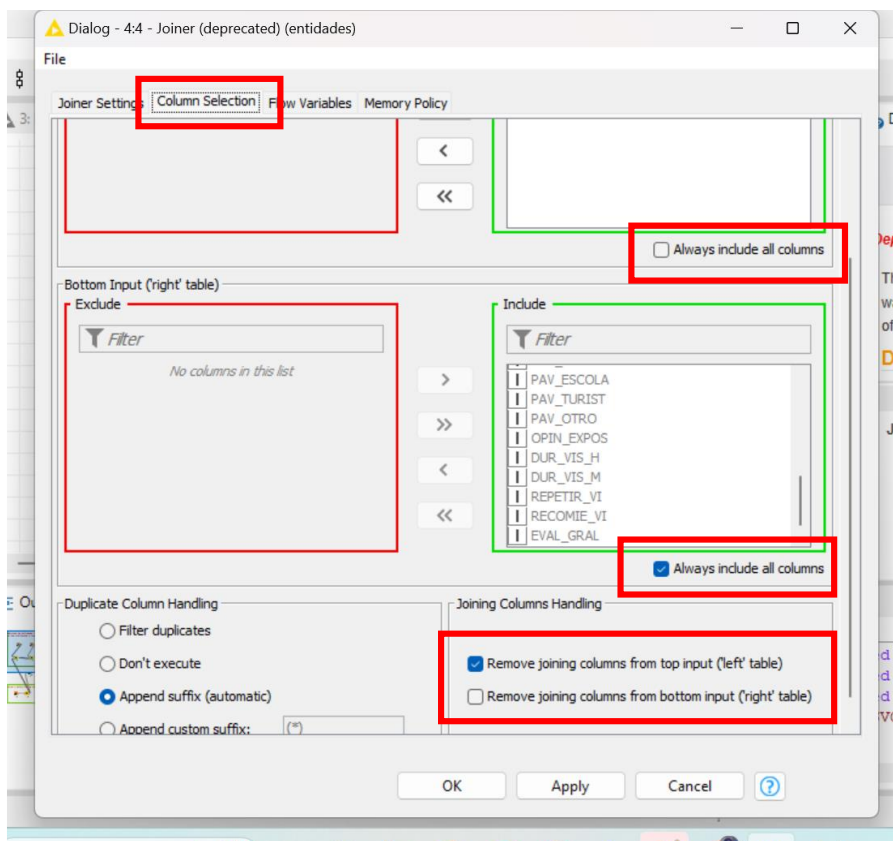
Con este nodo cargará el archivo llamado: MPIO2021.xlsx y realizará la configuración similar a la del archivo de visitas.

NODO JOINER (DEPRECATED). ENTIDADES

Se realizará un Join con las dos tablas que reúne: entidades y la resultante del Nodo Column Filter. En la pestaña Joiner Settings.

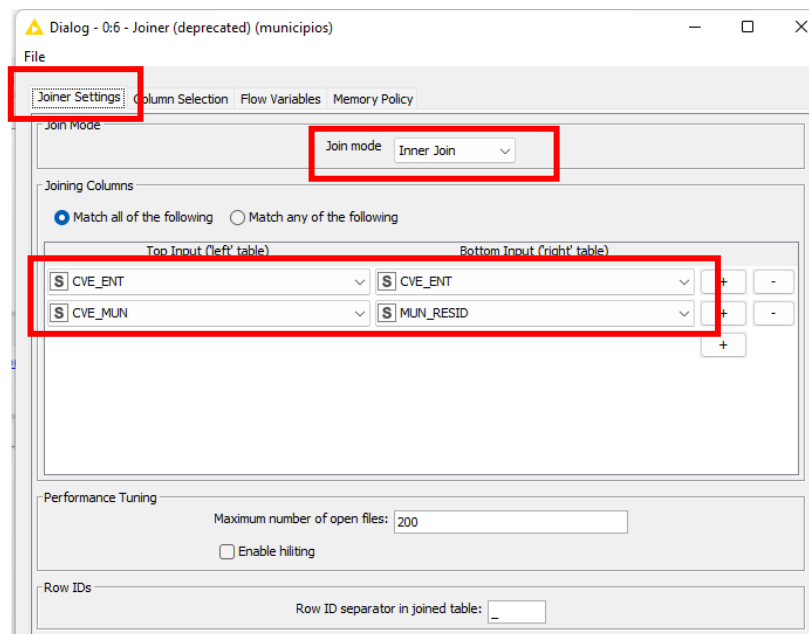


Debe seleccionar la opción de remover (de la pestaña Column Selection) los campos del archivo del catálogo de los nombres de entidades para que no estén repetidos, cuidando de mantener los nombres de la entidad de residencia (ENT_RESID) del archivo de datos de los visitantes a los museos. La pestaña Column Selection le permite elegir los atributos de salida de la operación de Joiner, los atributos o columnas de la tabla final.

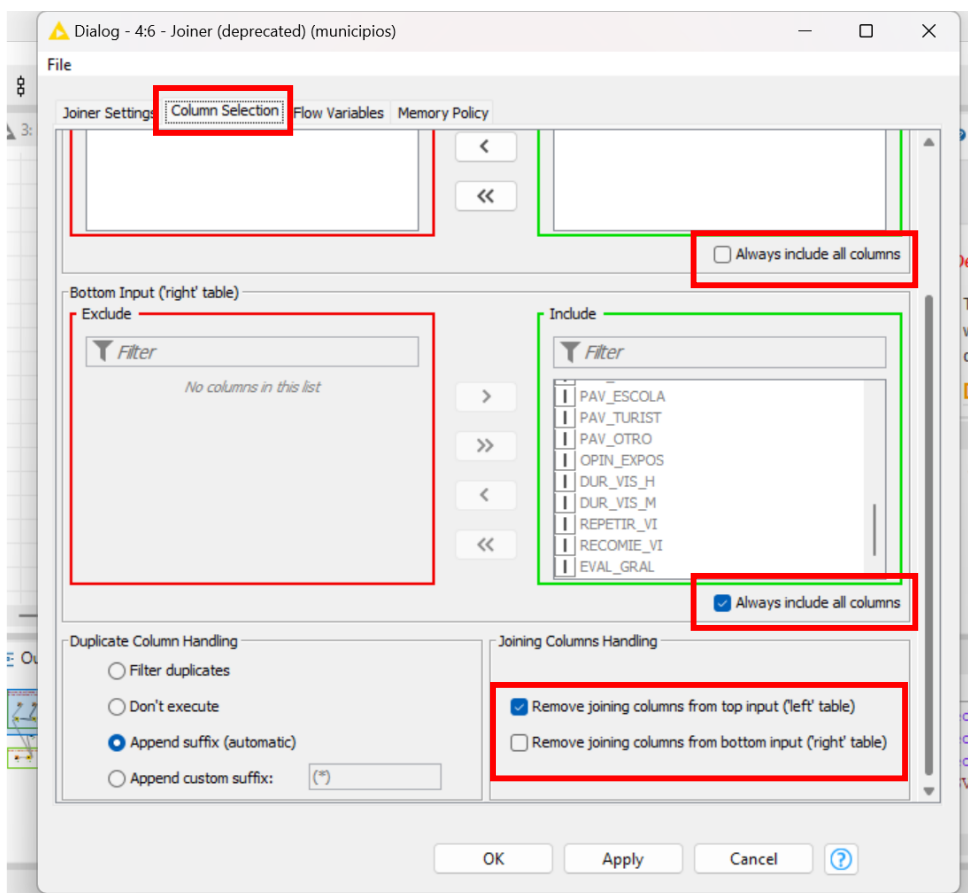


NODO JOINER (DEPRECATED). MUNICIPIOS

Se realizará un Join con las dos tablas que reúne: municipios y la resultante del Nodo Column Filter. Es un Join con dos atributos.

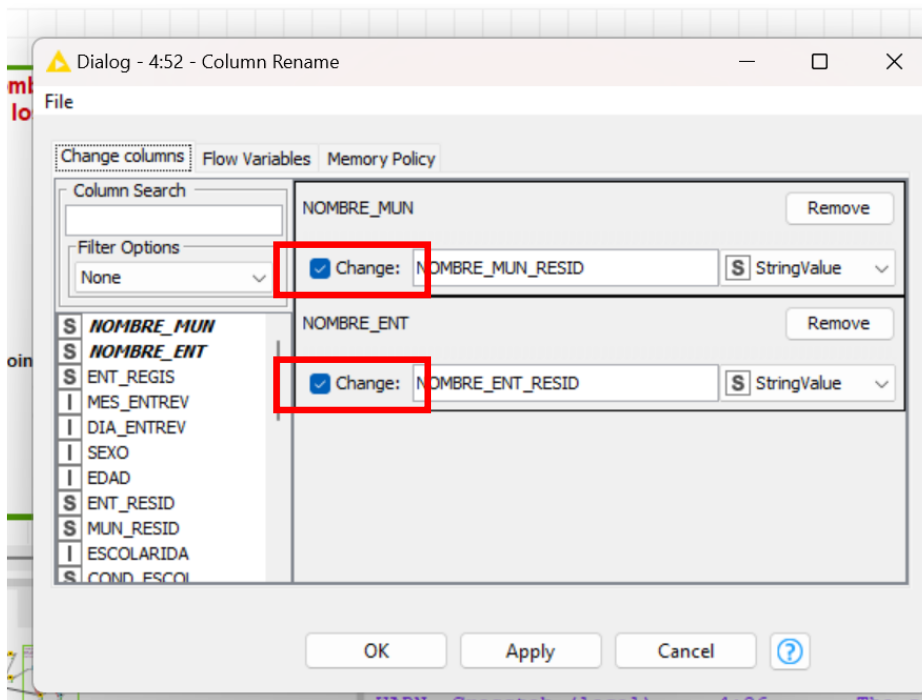


Debe seleccionar la opción de remover (de la pestaña Column Selection) los campos del archivo del catálogo de los nombres de municipios para que no estén repetidos, cuidando de mantener los nombres de los municipios de residencia (MUN_RESID) del archivo de datos de los visitantes a los museos. La pestaña Column Selection le permite elegir los atributos de salida de la operación de Joiner, los atributos o columnas de la tabla final.

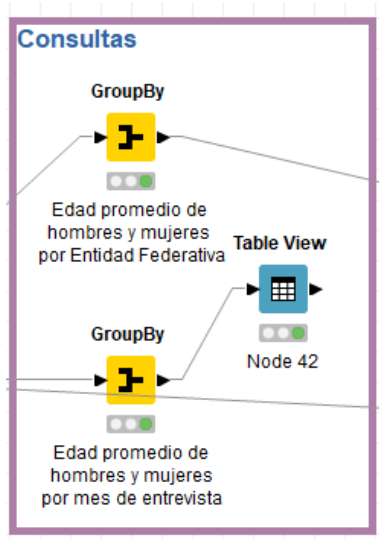


NODO COLUMN RENAME

Le permite cambiar el nombre y el tipo de los datos, selecciona el atributo y escribe los nombres (NOMBRE_ENT_RESID y NOMBRE_ENT_RESID) que se muestran en la imagen para distinguir las entidades y municipios de residencia de las entidades y municipios en donde se realizó la encuesta.



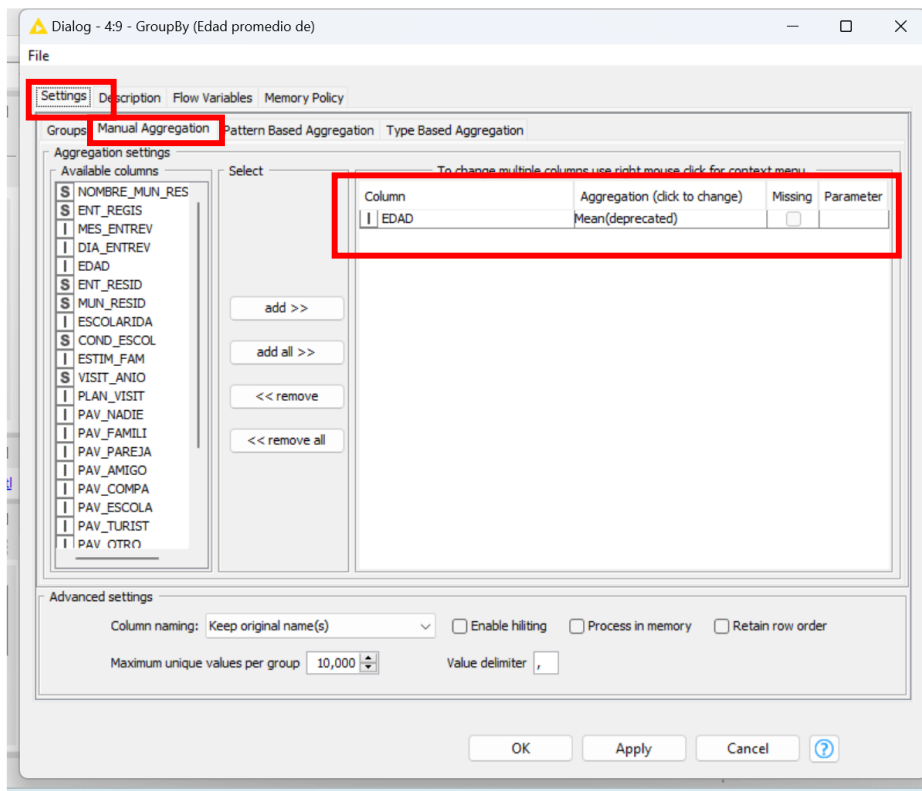
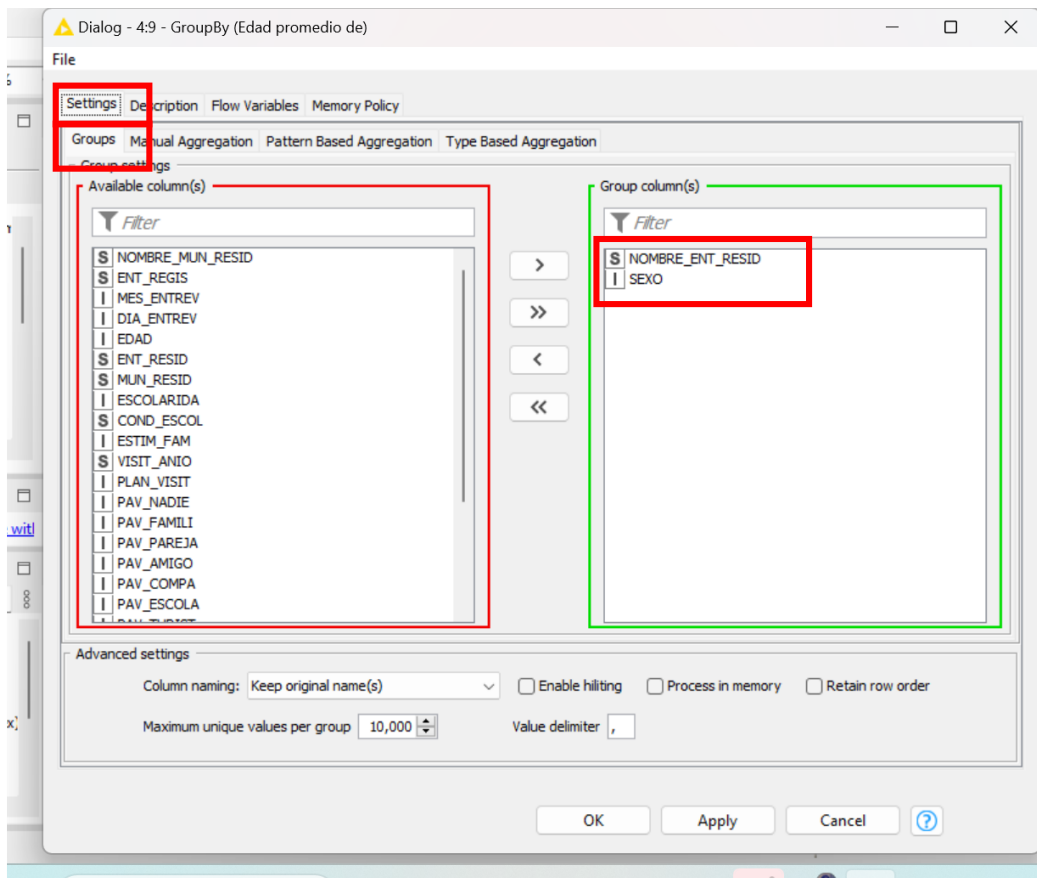
CONSULTAS

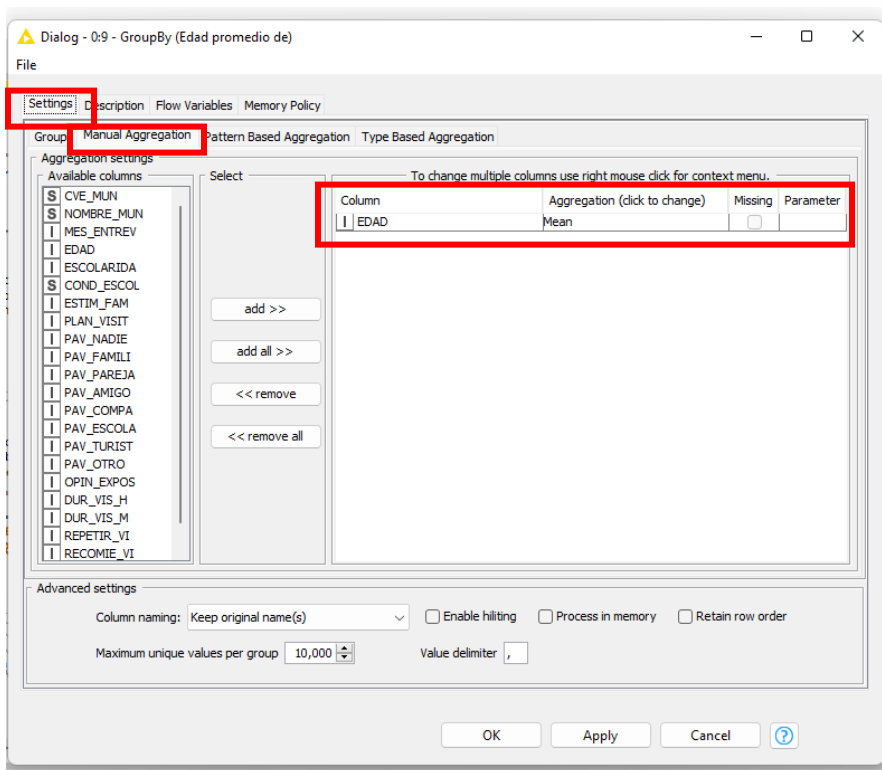


NODO GROUP BY. EDAD PROMEDIO DE HOMBRES Y MUJERES POR EF DE RESIDENCIA

Agrupar las filas de una tabla por los valores únicos en las columnas del grupo seleccionado (Groups). Se crea una fila para cada conjunto único de valores de la columna del grupo seleccionado.

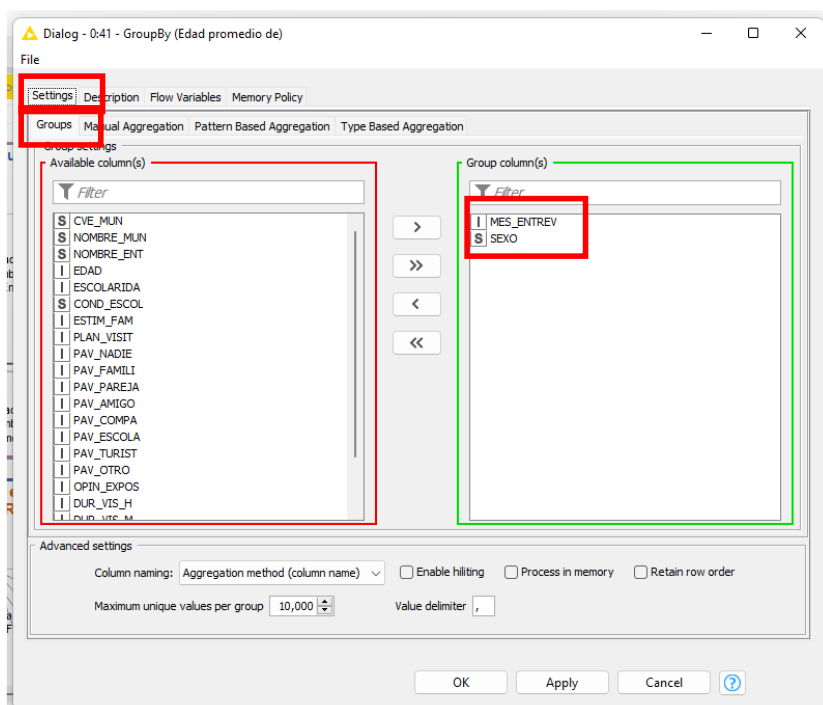
Esta consulta produce como resultado la edad promedio de hombres y mujeres de cada EF de residencia.

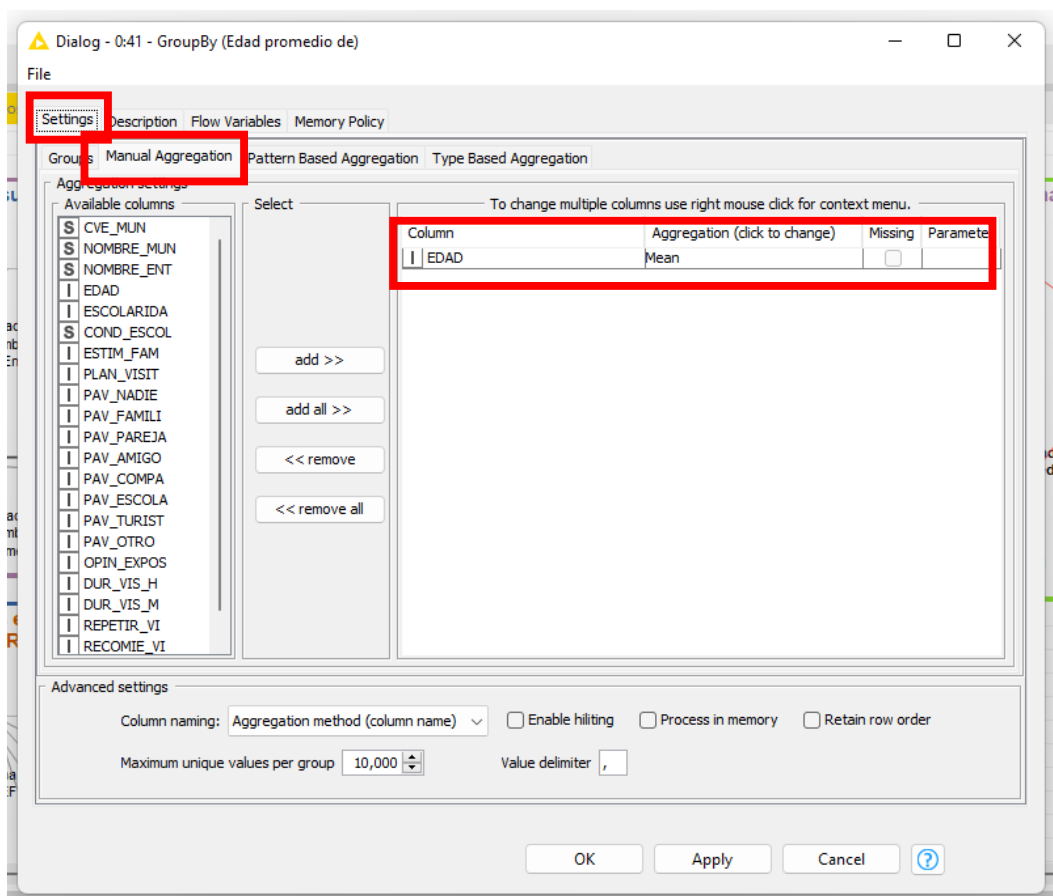




NODO GROUP BY. EDAD PROMEDIO DE HOMBRES Y MUJERES POR MES DE ENTREVISTA

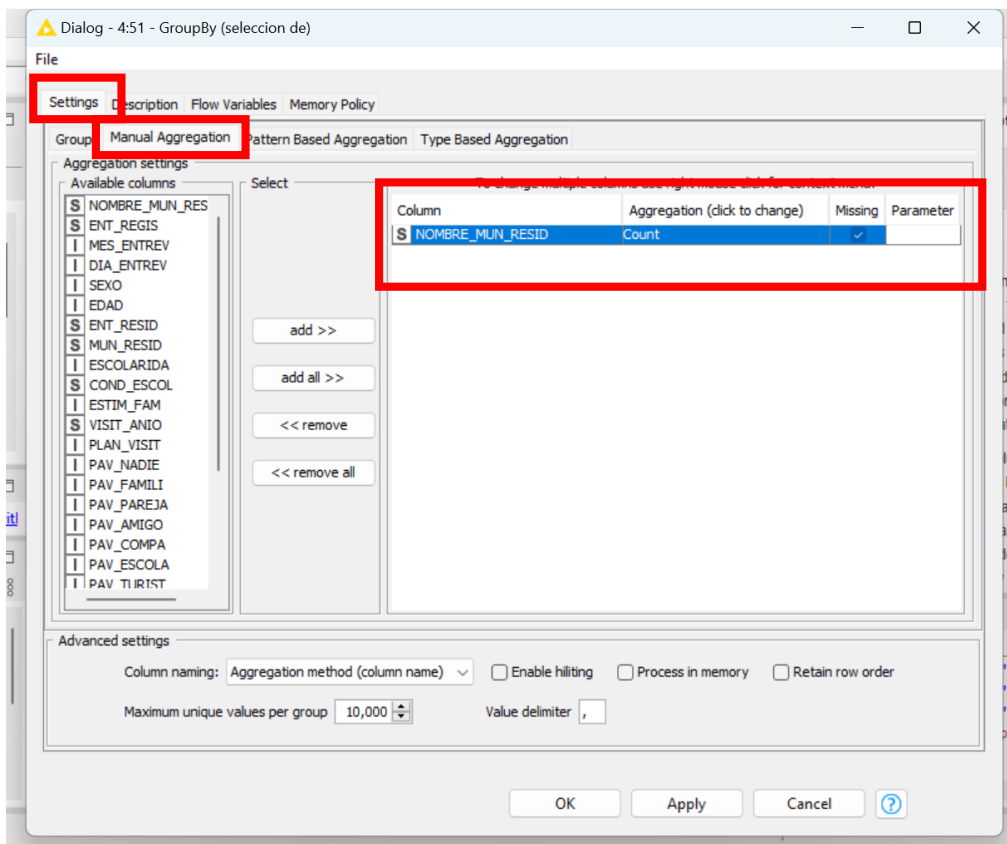
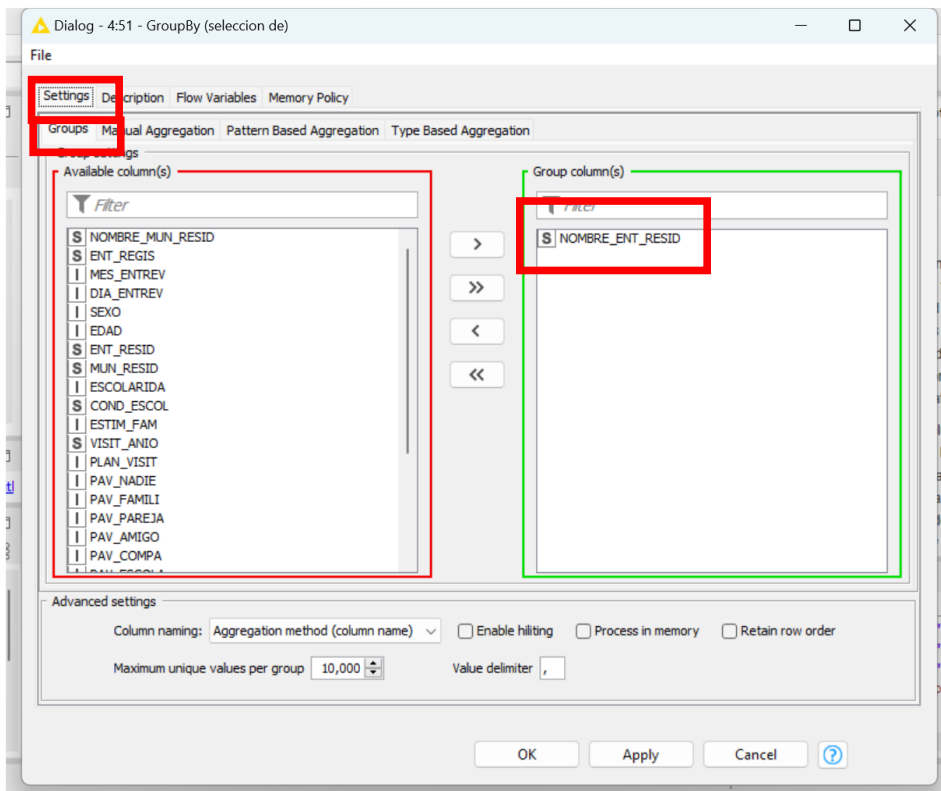
El resultado de esta consulta refleja la edad promedio de hombre y de mujeres en cada mes que se realizaron entrevistas.





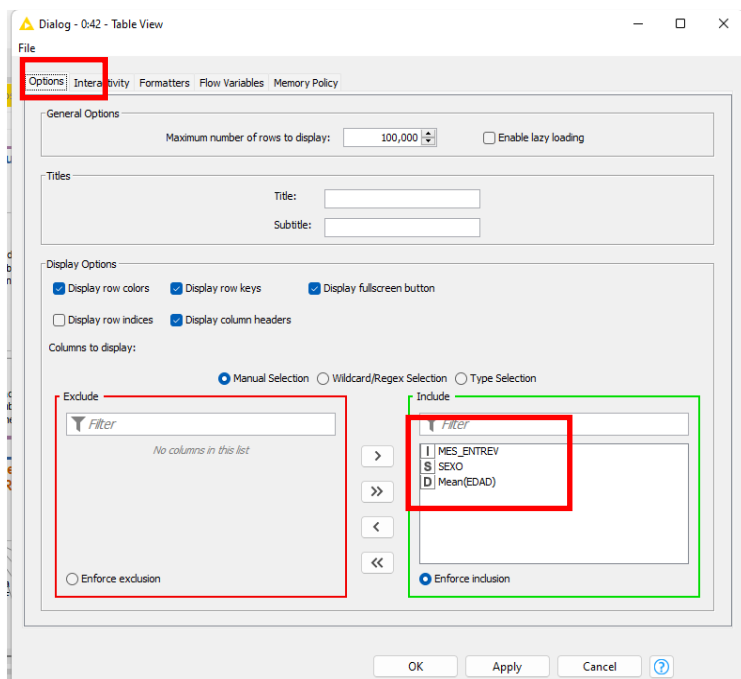
NODO GROUP BY. CUENTA EL NÚMERO DE ENTIDADES POR MUNICIPIO DE RESIDENCIA

Con esta consulta podrá conocer la cantidad de visitantes de cada municipio de residencia agrupadas por EF de residencia (NOMBRE_ENT_RESID).



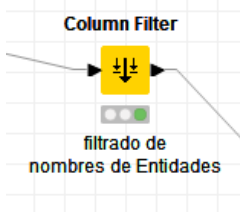
NODO TABLE VIEW

Muestra datos en una vista de tabla HTML. La vista ofrece varias funciones interactivas, así como la posibilidad de seleccionar filas.



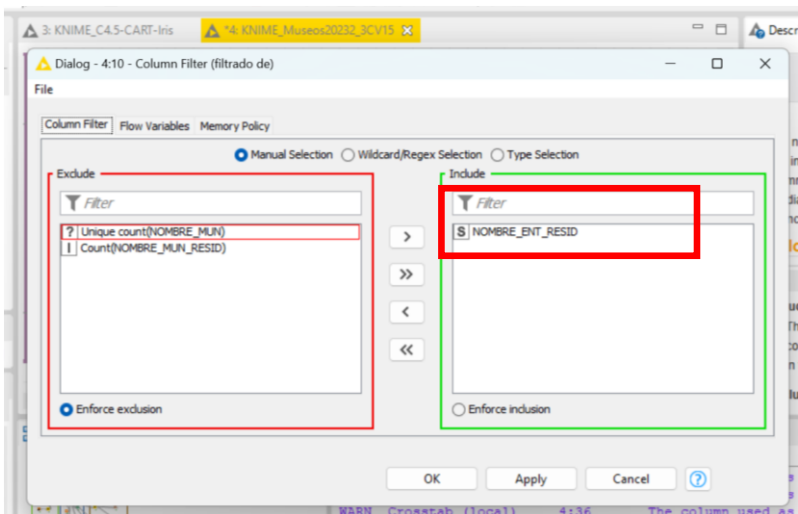
NODO COLUMN FILTER

Este nodo permite filtrar columnas de la tabla de entrada mientras que solo las columnas restantes se pasan a la tabla de salida. Dentro del cuadro de diálogo, las columnas se pueden mover entre la lista Incluir y Excluir.

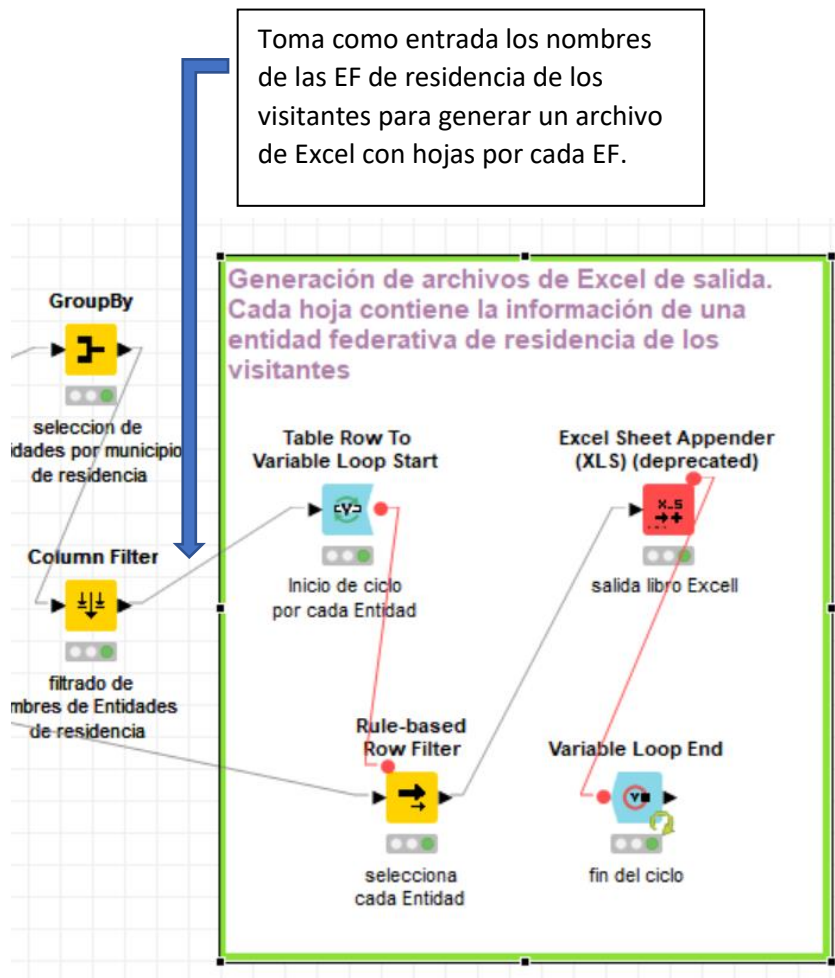


NODO COLUMN FILTER

Esta selección se realiza para generar posteriormente (en el ciclo) generar hojas de Excel que contengan los datos por EF de residencia de los visitantes al museo. Como verá es la entrada del inicio del ciclo.

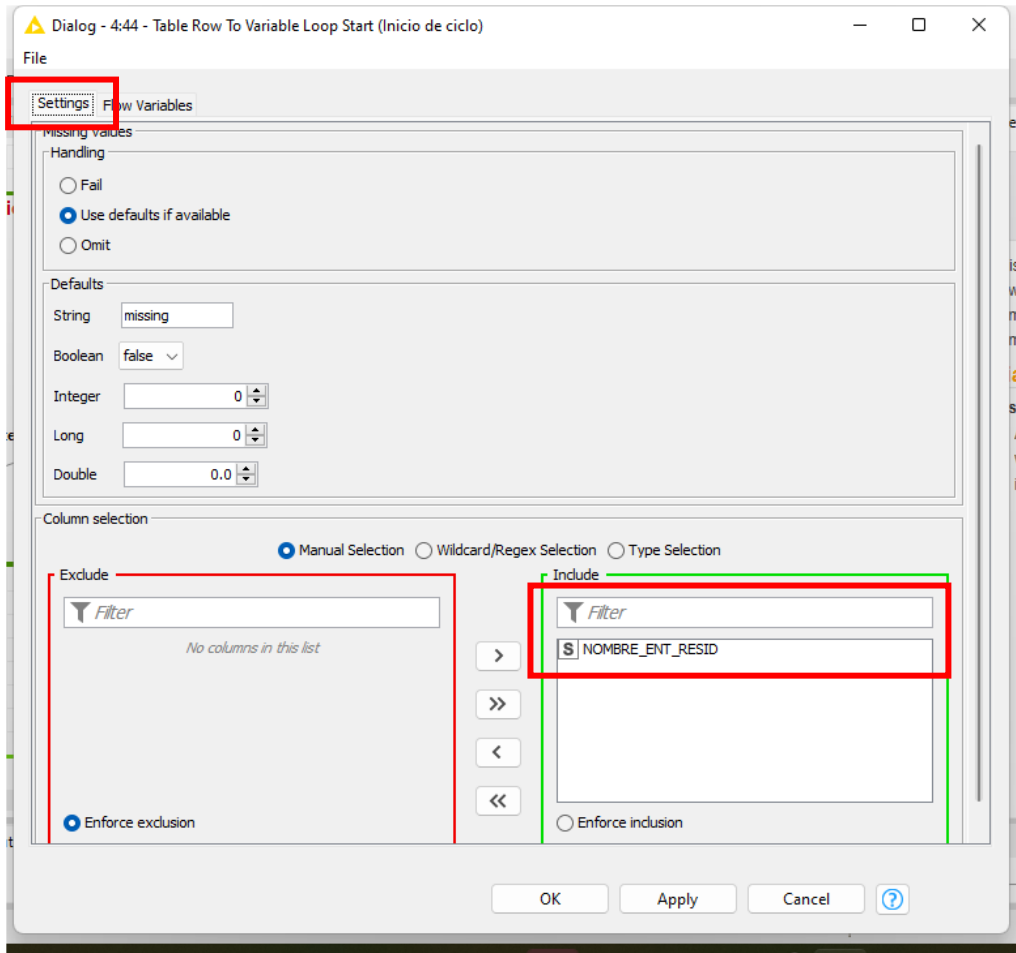


GENERACIÓN DE ARCHIVOS DE SALIDA EN EXCEL



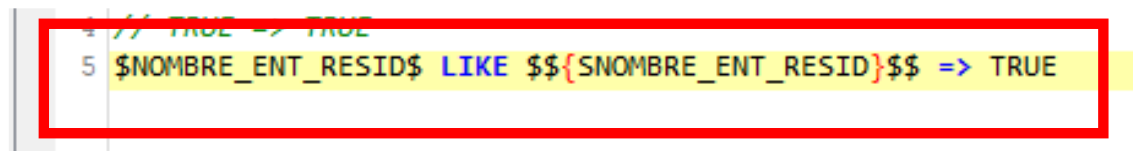
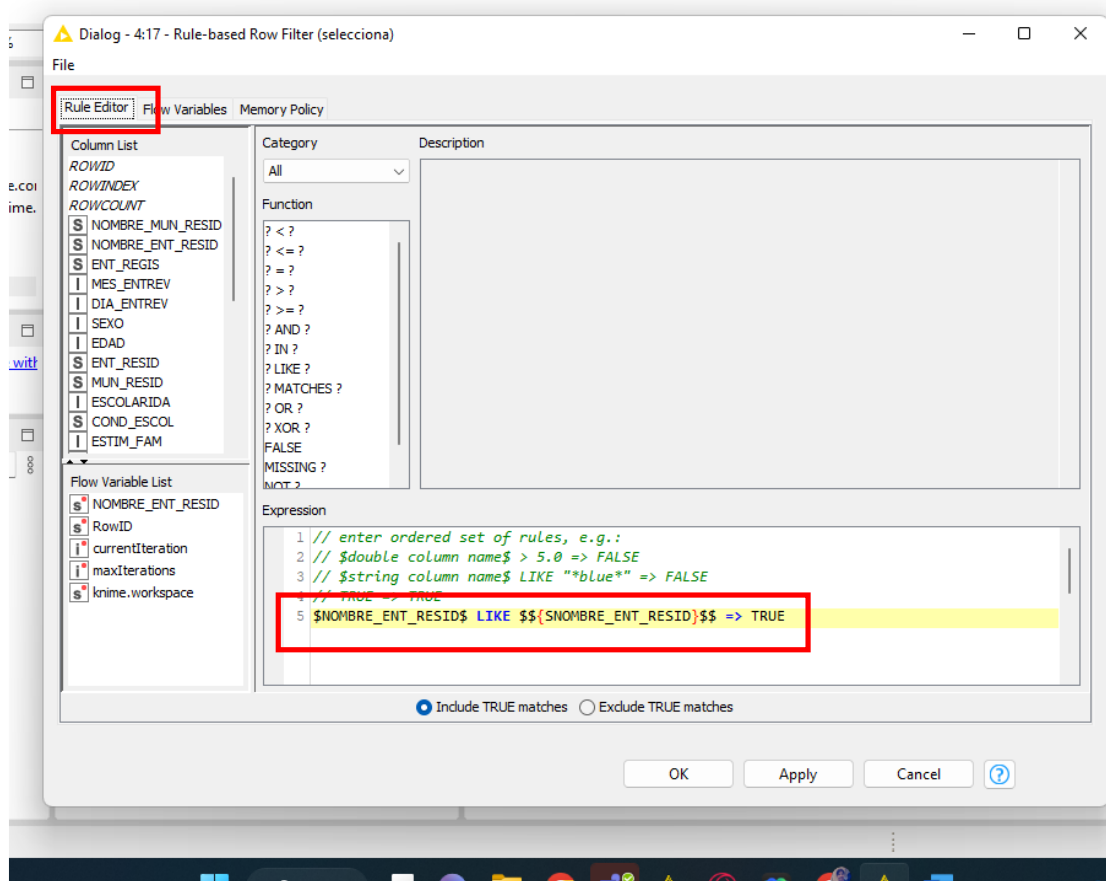
NODO TABLE ROW TO VARIABLE LOOP START

Este nodo usa cada fila de una tabla de datos para definir nuevos valores de variables para cada iteración de ciclo. Los nombres de las variables están definidos por los nombres de las columnas.



NODO RULE-BASED ROW FILTER

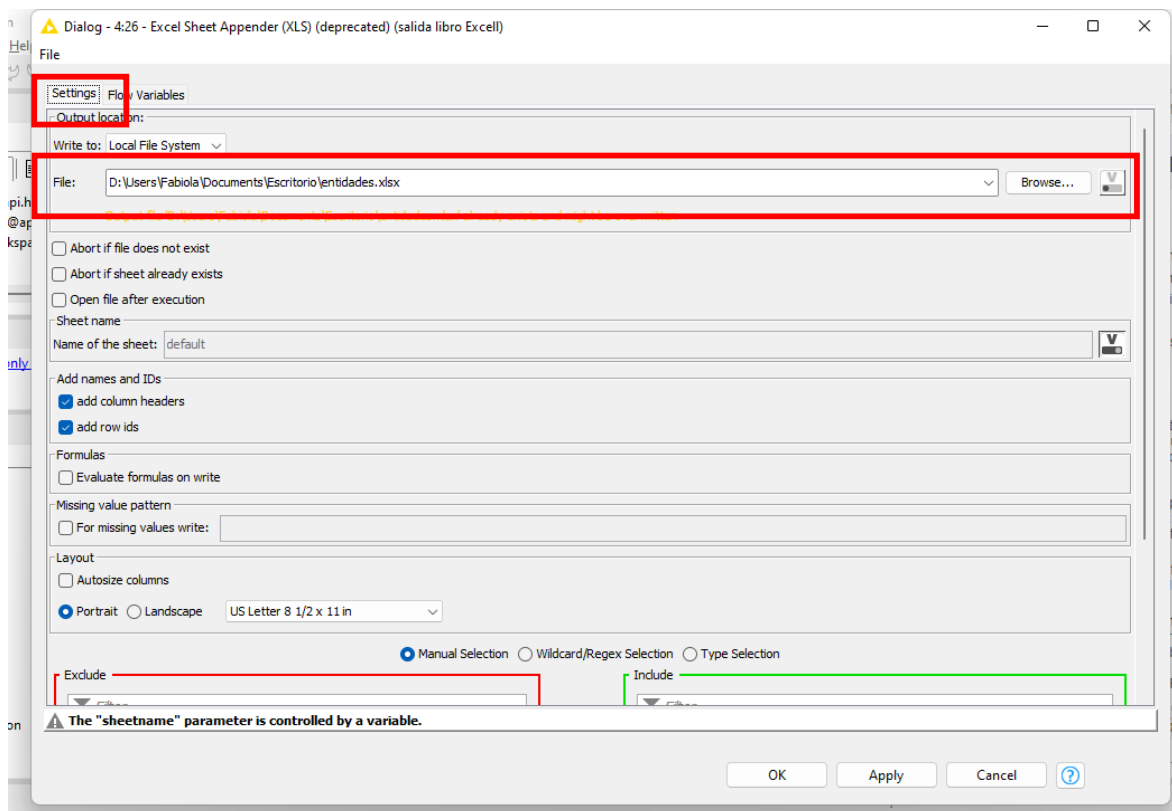
Este nodo toma una lista de reglas definidas por el usuario e intenta hacerlas coincidir con cada fila de la tabla de entrada. Si la primera regla coincidente tiene un resultado VERDADERO, la fila se seleccionará para su inclusión.



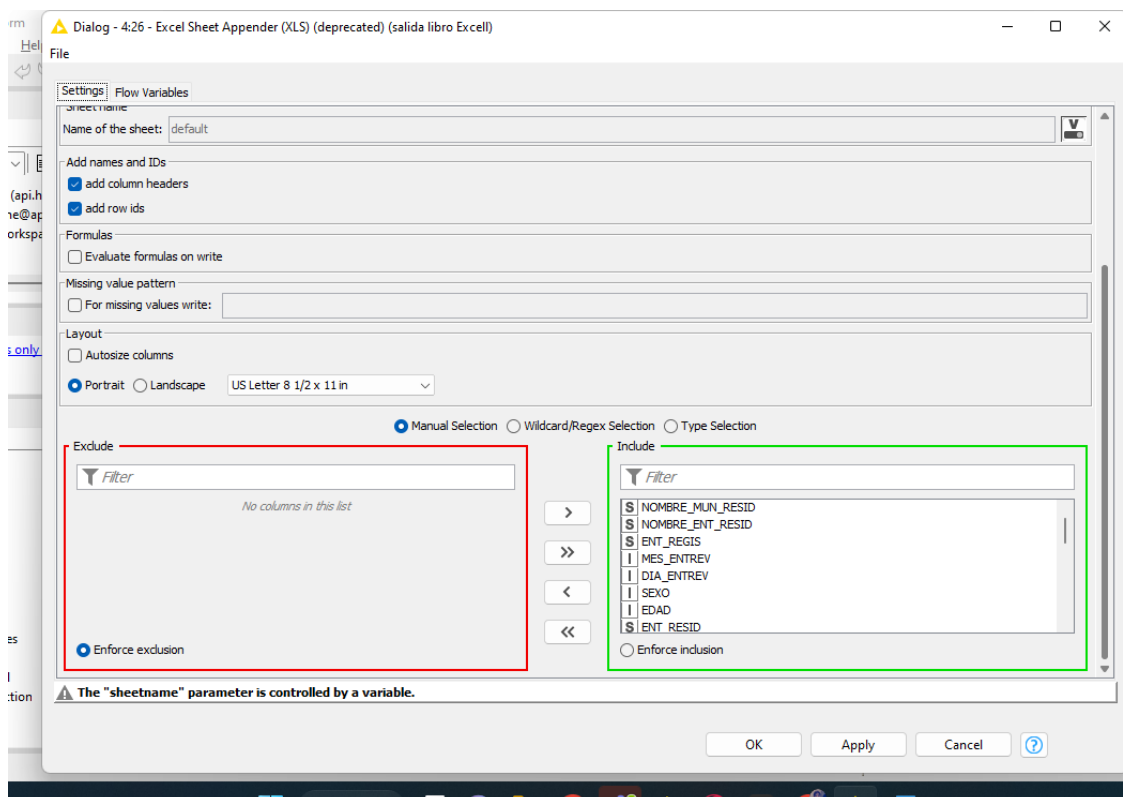
NODO EXCEL SHEET APPENDER (XLS) (DEPRECATED)

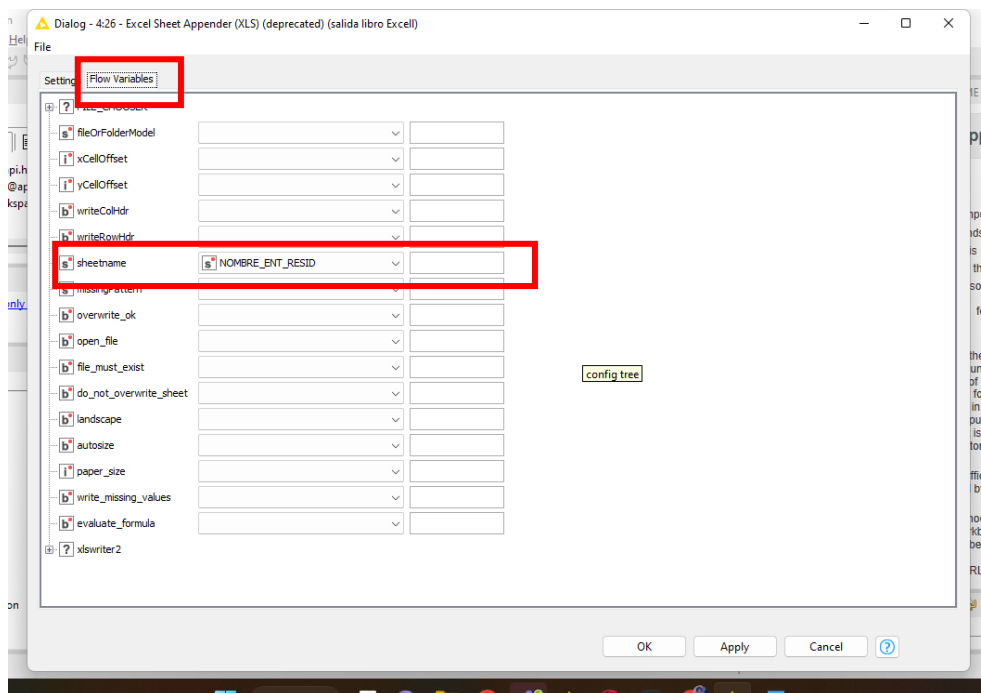
Este nodo escribe la tabla de datos de entrada en una hoja de cálculo y la agrega a un libro de trabajo existente. Si no hay ningún libro de trabajo presente, se creará uno nuevo. Luego, los archivos se pueden leer con otras aplicaciones, como Microsoft Excel.

Se establece la ruta y el nombre del archivo que contendrá los datos de salida.

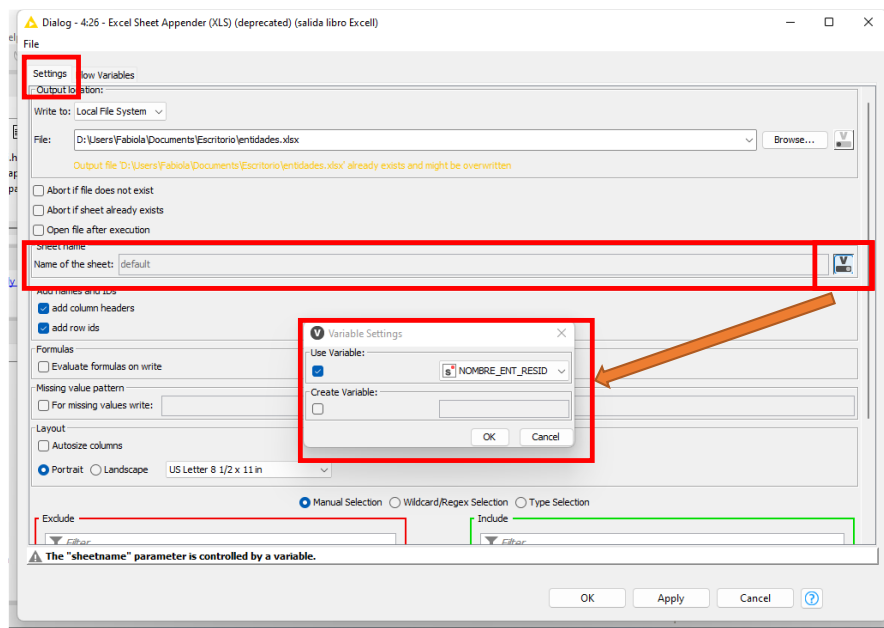


Continuación de la pantalla de configuración anterior.

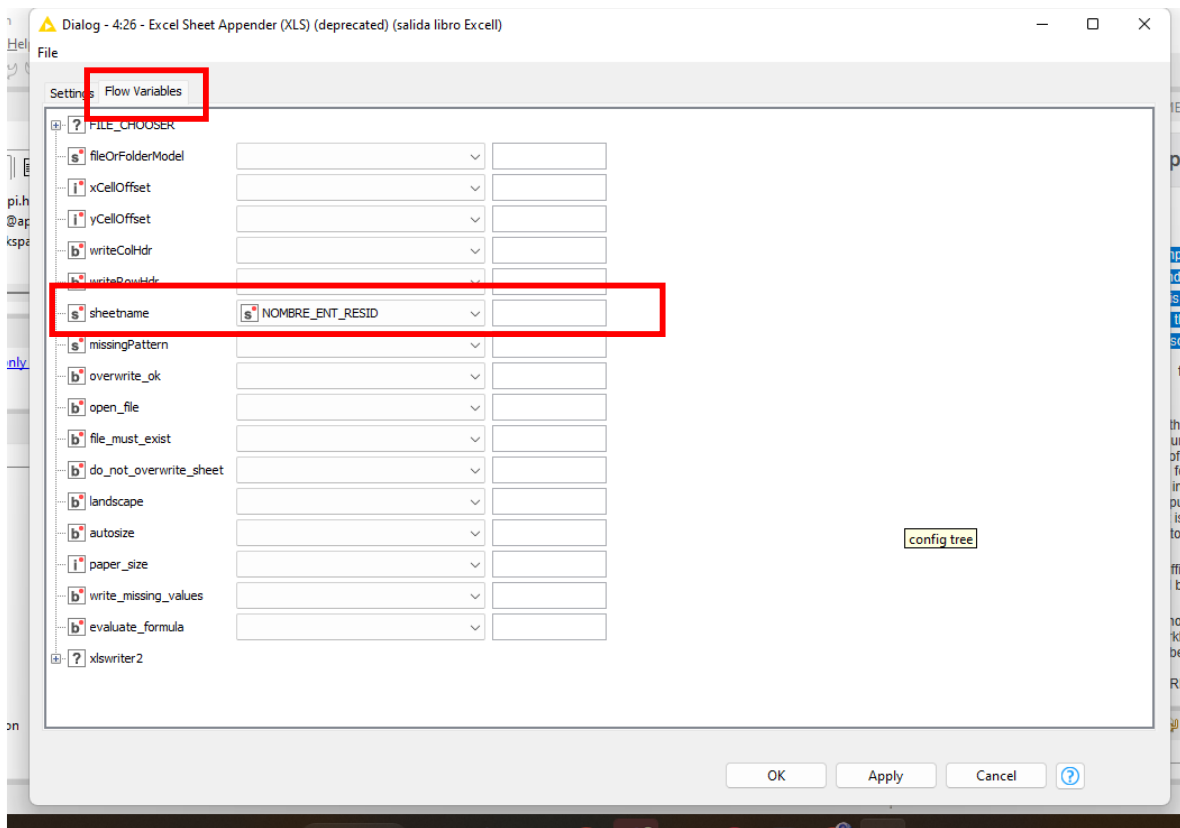




En la misma pestaña de Setting y en la opción Nombre de la hoja (Name of the sheet), presione el botón con la letra “V” para elegir el nombre de la cada hoja de Excel. Tal como se muestra en la figura.

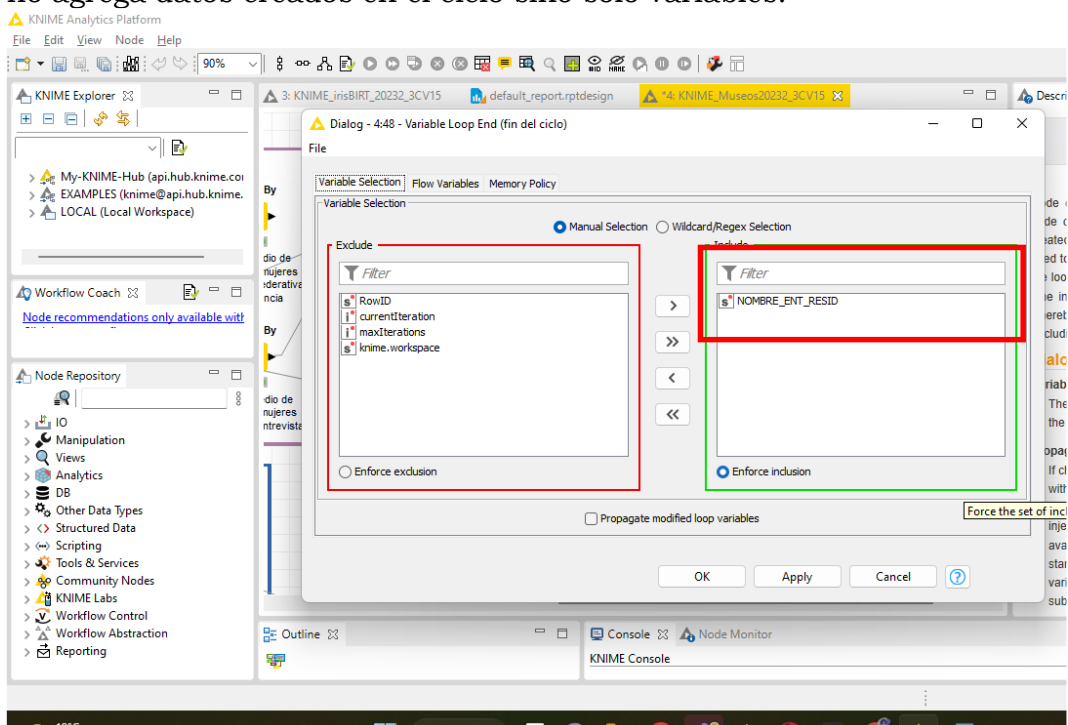


En la pestaña Flow Variable elija el nombre de la hoja de Excel como se muestra en la imagen.

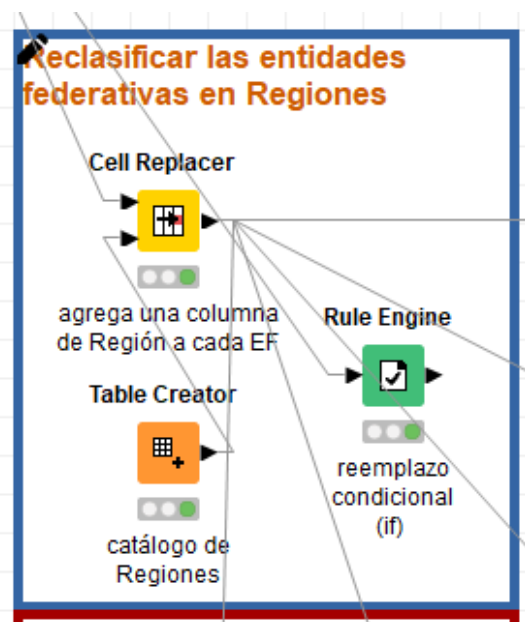


NODO VARIABLE LOOP END

Nodo que recopila variables creadas en un ciclo. Este nodo cierra un ciclo, pero no agrega datos creados en el ciclo sino solo variables.

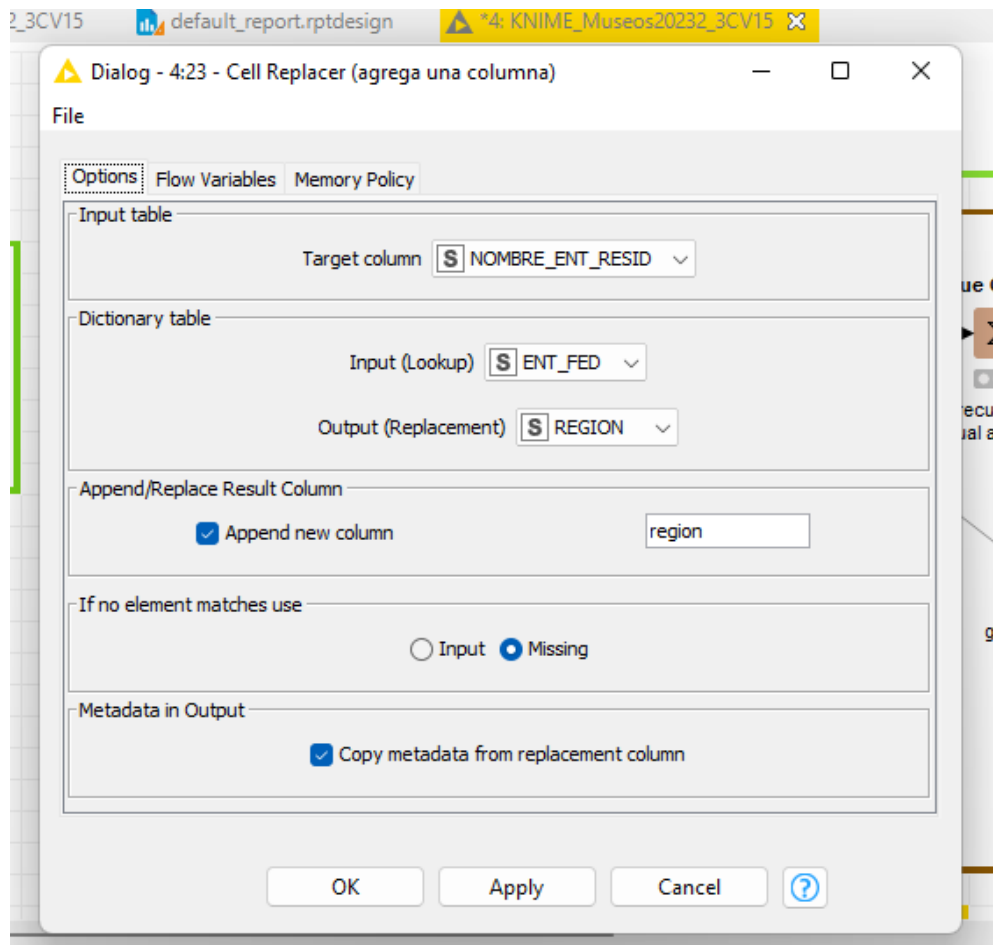


RECLASIFICACIÓN DE ENTIDADES FEDERATIVAS EN REGIONES



NODO CELL REPLACER

Reemplaza celdas en una columna de acuerdo con la tabla del diccionario (segunda entrada). El nodo tiene dos entradas: la primera entrada contiene una columna de destino cuyos valores se reemplazarán mediante la tabla de diccionario (segunda entrada). De la tabla del diccionario, elija una columna (o la columna ID de fila) que se usa como criterio de búsqueda. Cualquier aparición en la columna de destino (primera entrada) que coincida con el valor de búsqueda se reemplaza por el valor correspondiente de la columna de salida, que es otra columna en la tabla del diccionario.



NODO TABLE CREATOR

Permite la creación manual de una tabla de datos. Los datos se pueden ingresar en una hoja de cálculo como una tabla.

Dialog - 0:22 - Table Creator (catálogo de)

File

Table Creator Settings

Flow Variables

Memory Policy

Input line:

	S	ENT_FED	S	REGION				
Row0		Aguascalientes		Centronorte ...				
Row1		Baja California		Noroeste de ...				
Row2		Baja Californi...		Noroeste de ...				
Row3		Campeche		Sureste de M...				
Row4		Coahuila de Z...		Noreste de M...				
Row5		Colima		Oeste de Mé...				
Row6		Chiapas		Suroeste de ...				
Row7		Chihuahua		Noroeste de ...				
Row8		Ciudad de Mé...		Centrosur de...				
Row9		Durango		Noroeste de ...				
Row10		Guanajuato		Centronorte ...				
Row11		Guerrero		Suroeste de ...				
Row12		Hidalgo		Oriente de M...				
Row13		Jalisco		Oeste de Mé...				
Row14		México		Centrosur de...				
Row15		Michoacán de...		Oeste de Mé...				
Row16		Morelos		Centrosur de...				
Row17		Nayarit		Oeste de Mé...				
Row18		Nuevo León		Noreste de M...				
Row19		Oaxaca		Suroeste de ...				
Row20		Puebla		Oriente de M...				
Row21		Querétaro		Centronorte ...				
Row22		Quintana Roo		Sureste de M...				

The output table has 32 rows and 2 columns.

☐ Highlight output table

OK

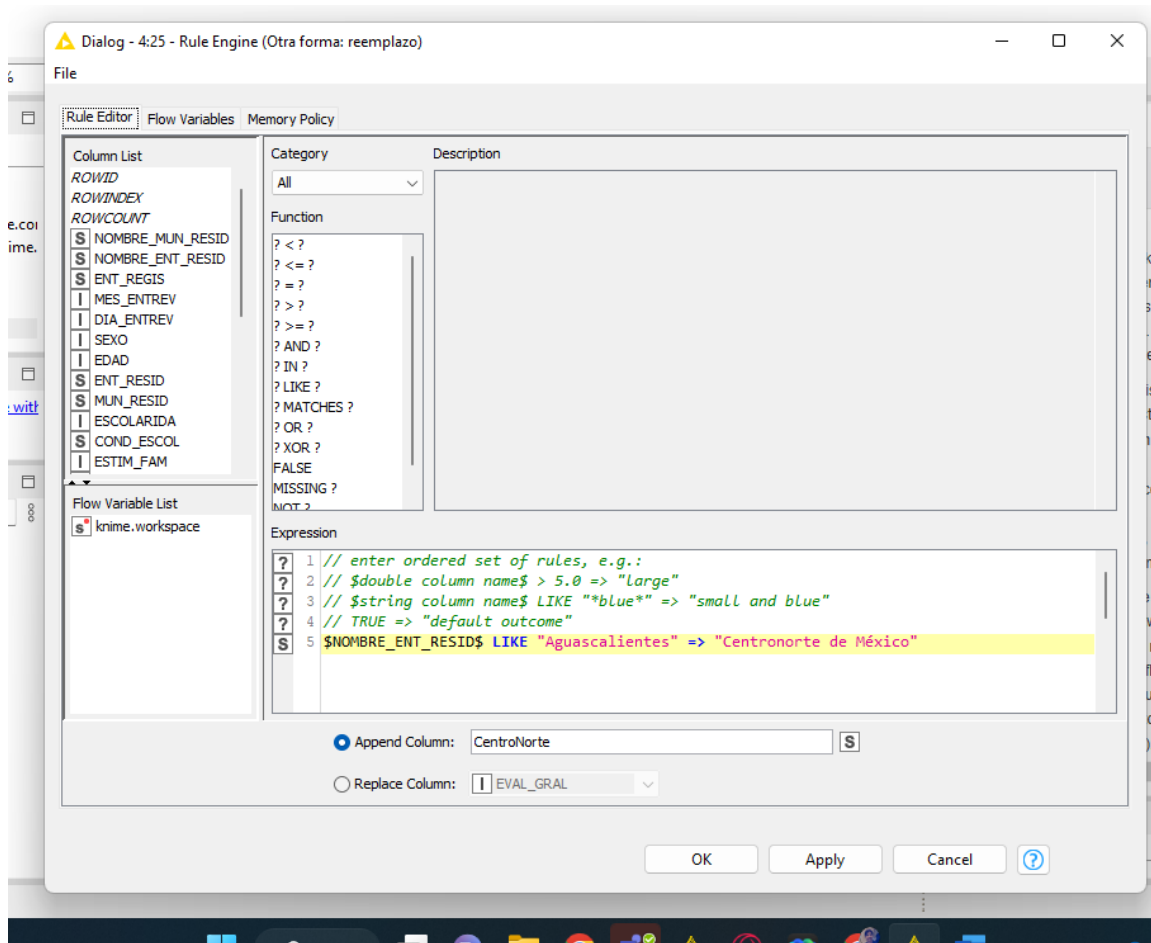
Apply

Cancel

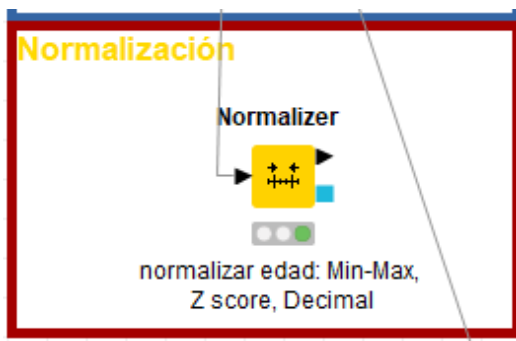
?

NODO RULE ENGINE

Este nodo toma una lista de reglas definidas por el usuario e intenta hacerlas coincidir con cada fila de la tabla de entrada. Si una regla coincide, su valor de resultado se agrega a una nueva columna. La primera regla coincidente en orden de definición determina el resultado.



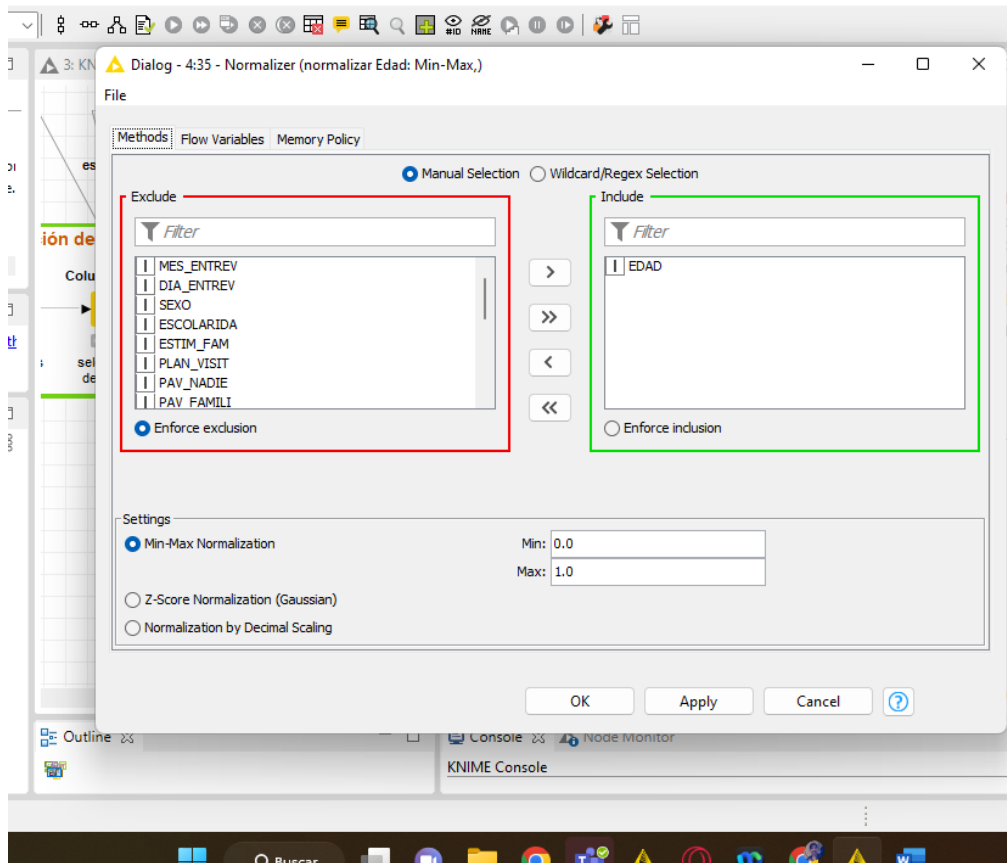
NORMALIZACIÓN



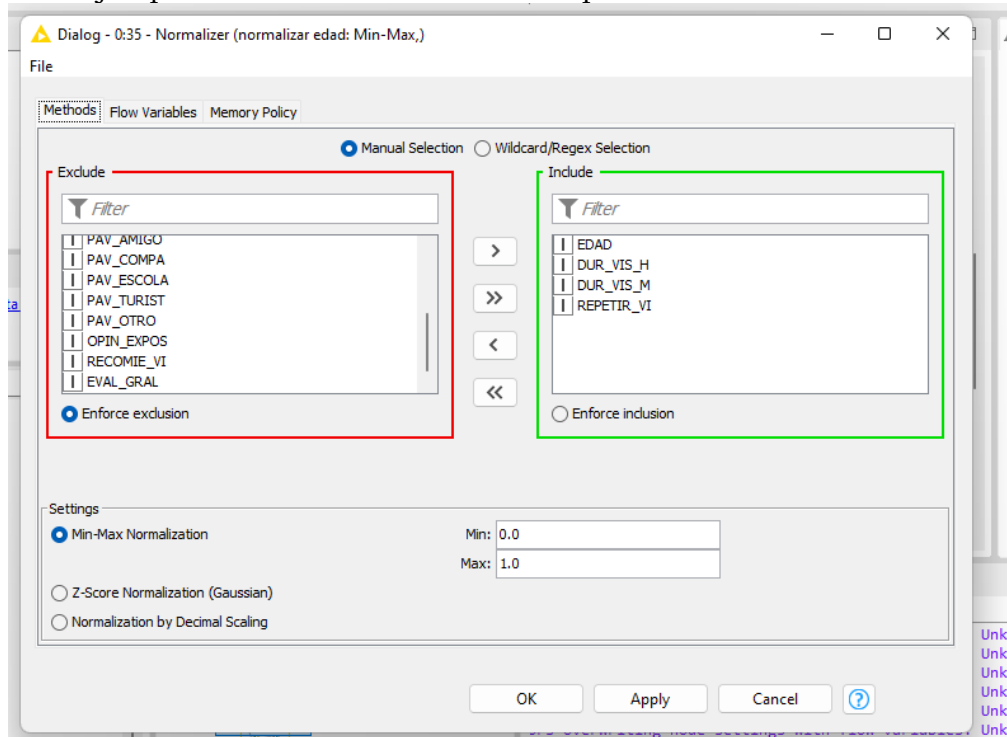
NODO NORMALIZER

Este nodo normaliza los valores de todas las columnas (numéricas). En el cuadro de diálogo, puede elegir las columnas en las que desea trabajar. Los siguientes métodos de normalización están disponibles en el cuadro de diálogo.

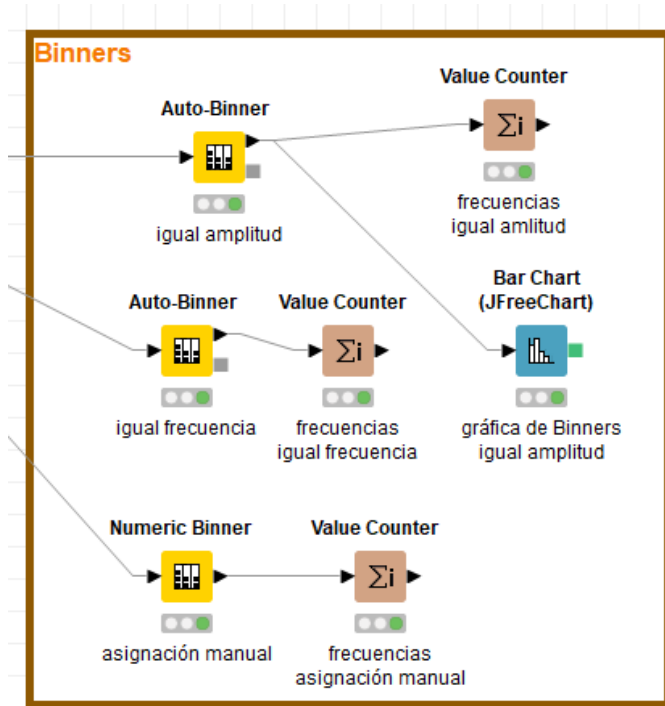
Un ejemplo de selección de un solo campo.



Otro ejemplo de selección de varios campos.

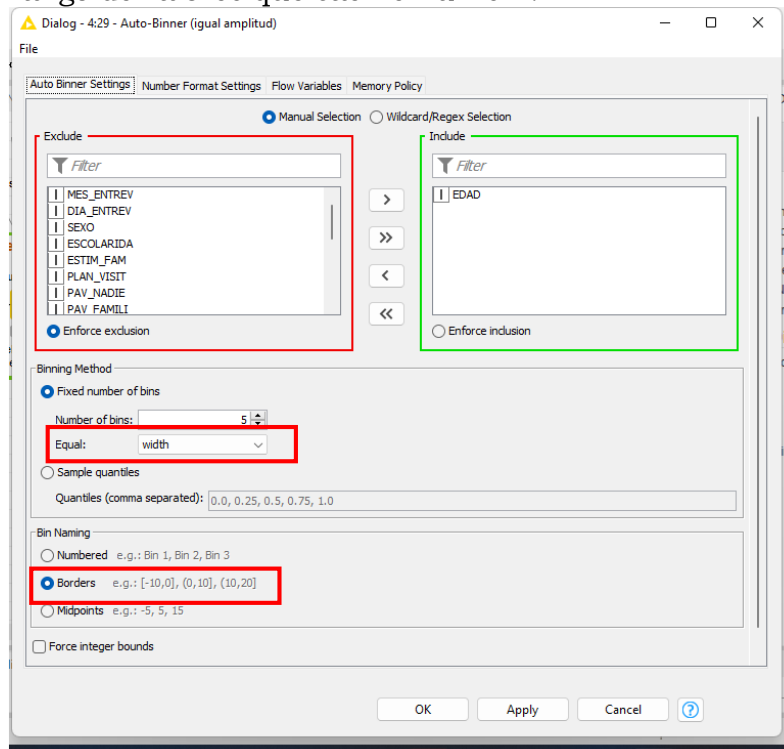


BINNERS



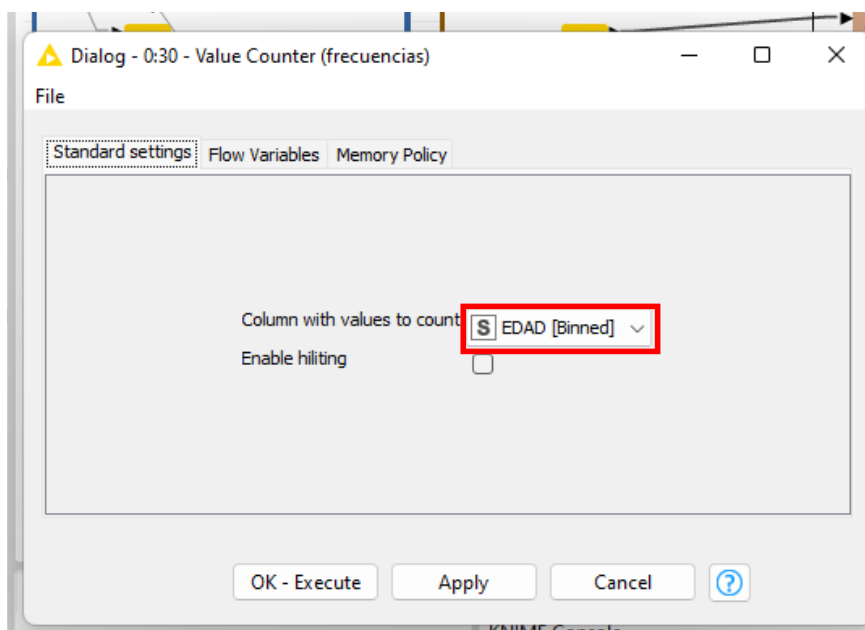
NODO AUTO-BINNER. IGUAL AMPLITUD

Este nodo permite agrupar datos numéricos en intervalos, llamados bins. Hay dos opciones de nombres para los bins y dos métodos que definen el número y el rango de valores que caen en un bin.



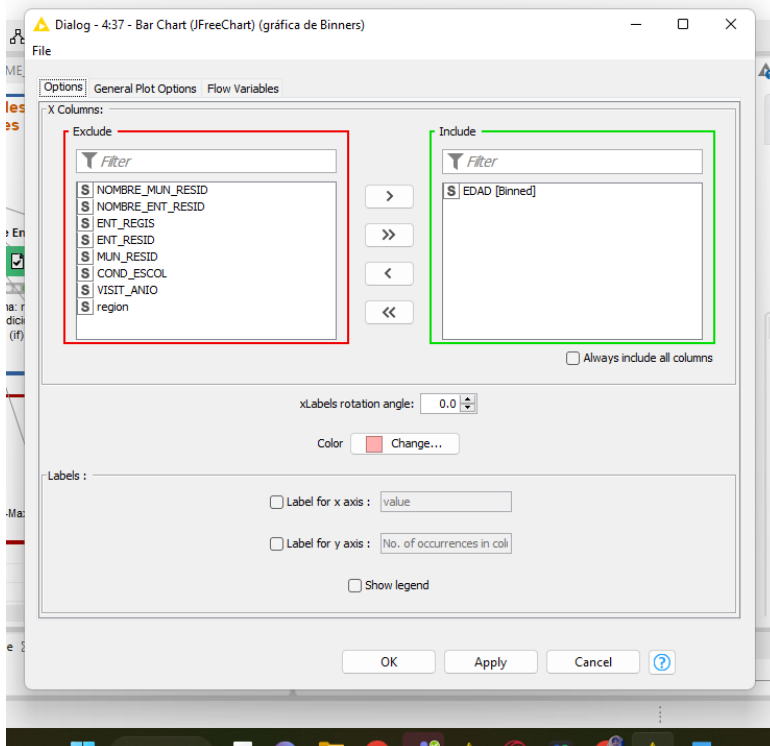
NODO VALUE COUNTER. FRECUENCIAS IGUAL AMPLITUD

Este nodo cuenta el número de ocurrencias de todos los valores en una columna seleccionada.

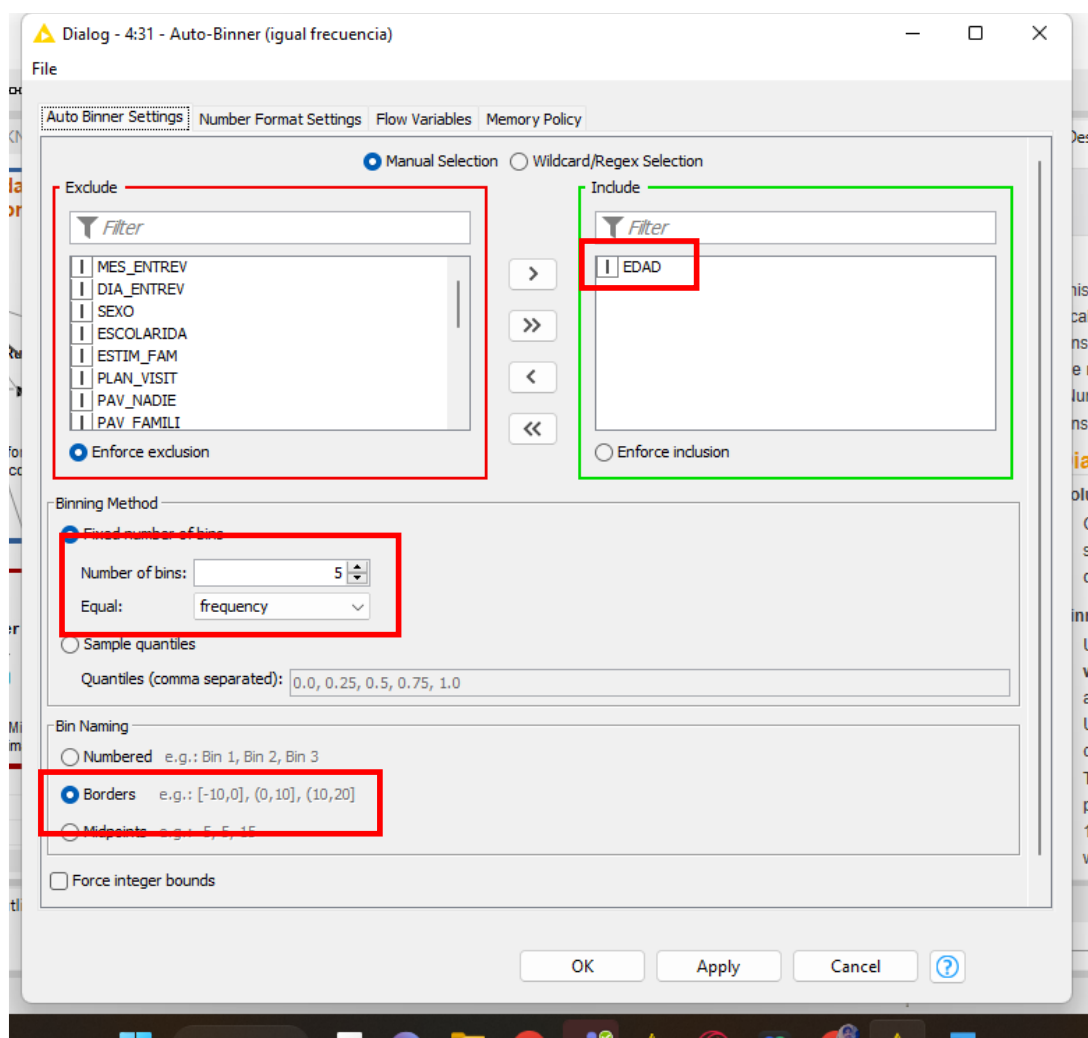


NODO BAR CHAR. GRÁFICA DE BINNERS IGUAL AMPLITUD

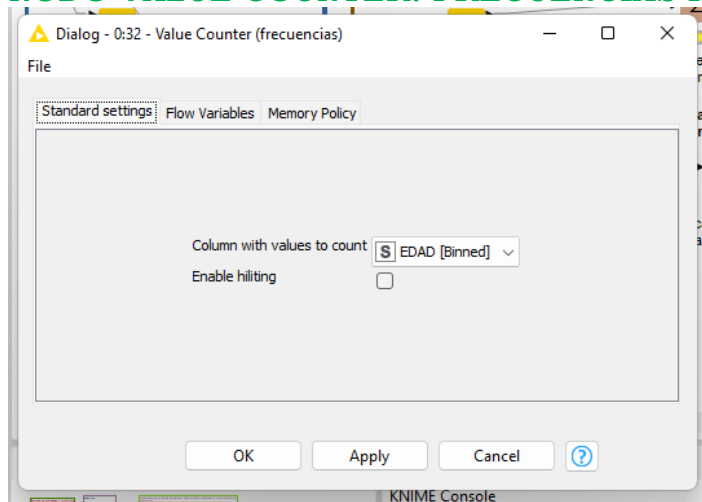
Este nodo muestra un gráfico de barras que visualiza el recuento de ocurrencias para cada valor y cada columna nominal seleccionada.



NODO AUTO-BINNER. IGUAL FRECUENCIA

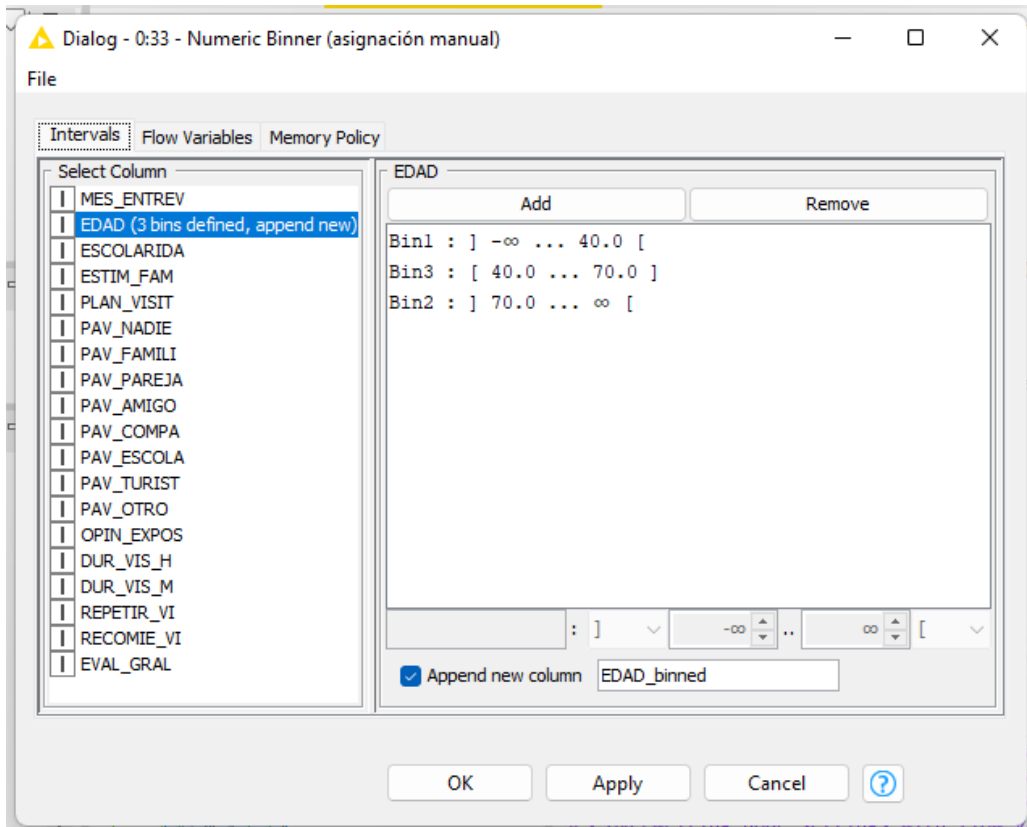


NODO VALUE COUNTER. FRECUENCIAS IGUAL FRECUENCIA

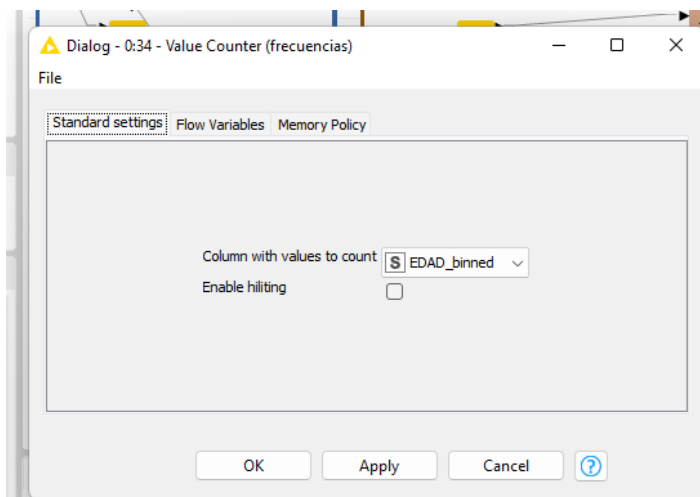


NODO NUMERIC BINNER. ASIGNACIÓN MANUAL

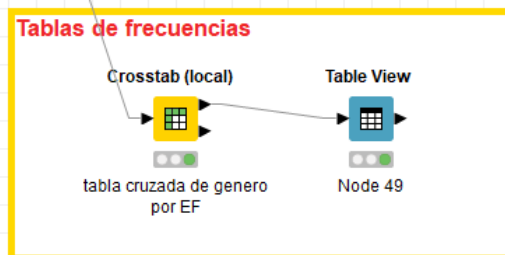
Para cada columna, se puede definir una cantidad de intervalos, conocidos como contenedores. A cada uno de estos contenedores se le asigna un nombre único (para esta columna), un rango definido y límites de intervalo abiertos o cerrados. Aseguran automáticamente que los rangos se definan en orden descendente y que los límites de los intervalos sean coherentes.



NODO VALUE COUNTER. FRECUENCIAS ASIGNACIÓN MANUAL

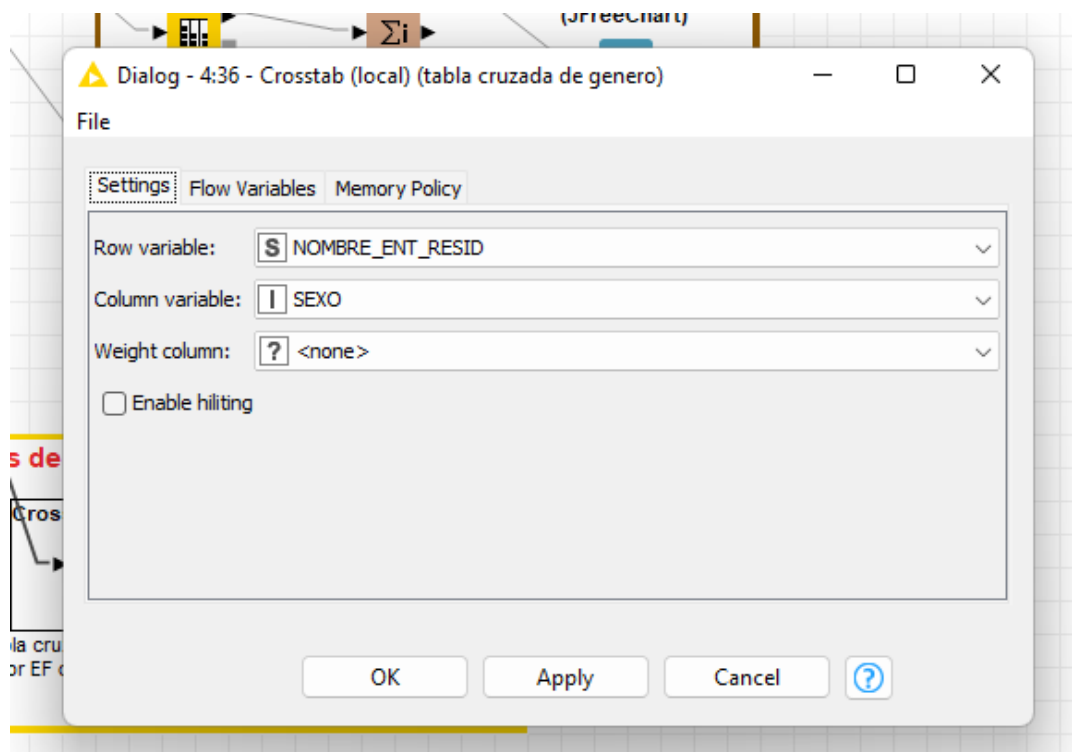


TABLAS DE FRECUENCIAS



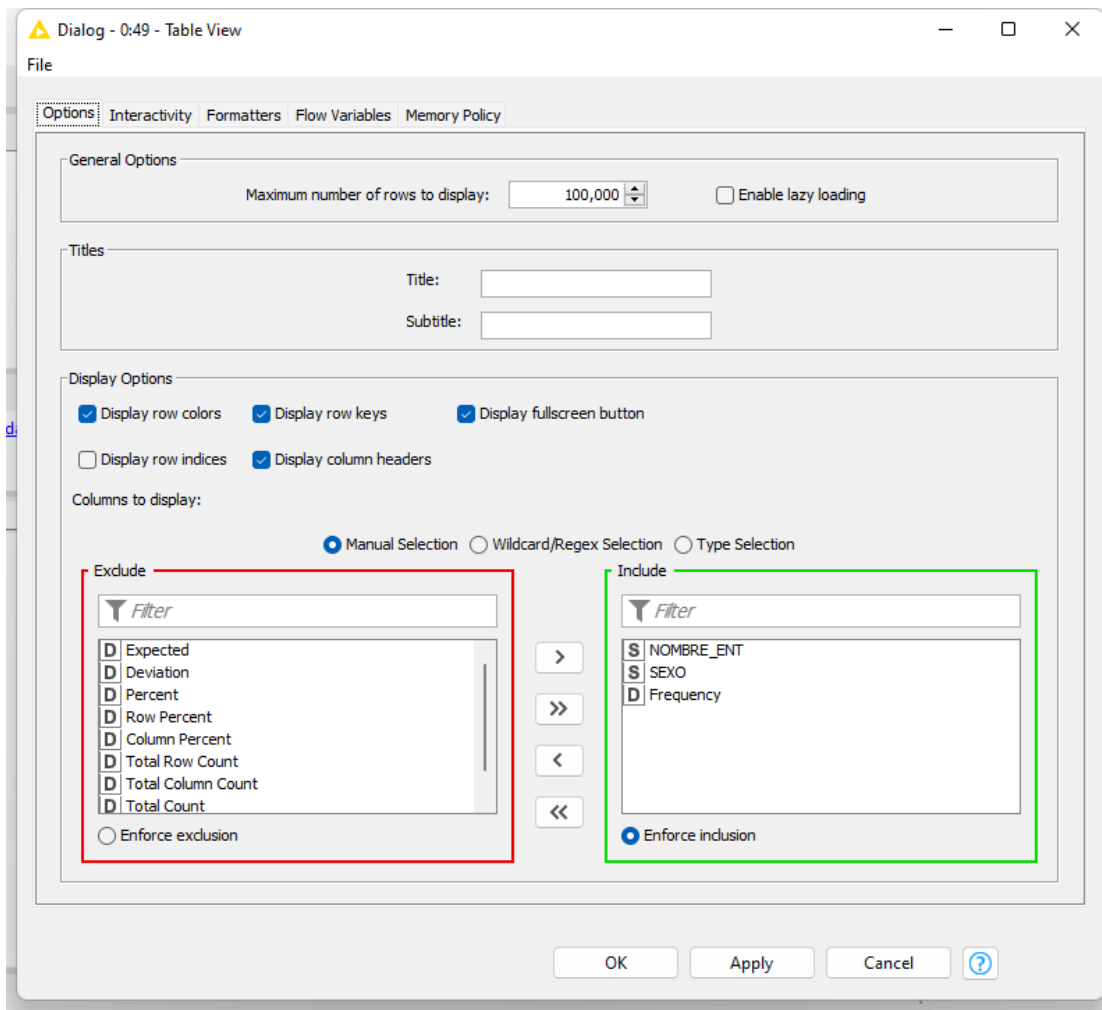
NODO CROSSTAB (LOCAL)

Crea una tabla cruzada (también conocida como tabla de contingencia o tabulación cruzada). Puede usarse para analizar la relación de dos columnas con datos categóricos y muestra la distribución de frecuencia de las variables categóricas en una tabla.



NODO TABLE VIEW

Muestra datos en una vista de tabla HTML. La vista ofrece varias funciones interactivas, así como la posibilidad de seleccionar filas.



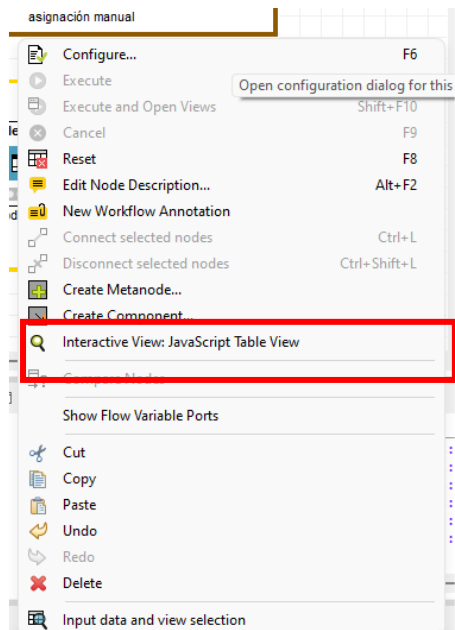


Table interface showing 10 entries. Search:

	RowID	NOMBRE_ENT	SEXO	Frequency
<input checked="" type="checkbox"/>	Row0	Durango	1	64
<input checked="" type="checkbox"/>	Row1	Durango	2	42
<input type="checkbox"/>	Row2	Guanajuato	1	11
<input type="checkbox"/>	Row3	Guanajuato	2	12
<input type="checkbox"/>	Row4	Guerrero	1	6
<input type="checkbox"/>	Row5	Guerrero	2	7
<input type="checkbox"/>	Row6	Hidalgo	1	15
<input type="checkbox"/>	Row7	Hidalgo	2	7
<input type="checkbox"/>	Row8	Jalisco	1	673
<input type="checkbox"/>	Row9	Jalisco	2	735

Showing 1 to 10 of 46 entries

Navigation: Previous 1 2 3 4 5 Next