

DOCUMENTO GUÍA DE LA PRÁCTICA
DATA MINING (ISC) **GRUPO: 3CV15**

Equipo No. 4

Nombres:

Flores Ponce Alan Marcelo

García Cruz Octavio Arturo

Sampayo Hernández Mauro

Diccionario de Datos:

Nombre de variable	Tipo de dato	Dominio	Significado	Valores
ENT_REGIS	Nominal	01...31	Entidad federativa de registro. El territorio nacional se divide en 32 entidades. Hace referencia a la entidad federativa en donde se registra el dato.	01..32-Entidad Federativa
MES_ENTREV	Nominal	1..12	Mes de la entrevista. Aquel en el cual el visitante acudió al Establecimiento museográfico y fue entrevistado.	1-Enero 2-Febrero .. 12-Diciembre
SEXO	Nominal (String)	1, 2	Condición biológica que distingue a las mujeres y hombres	1-Hombre 2-Mujer
EDAD	Numérico	12..99	Años transcurridos entre la fecha de nacimiento de la persona y la fecha en la que se le encuesta.	12...97-Años 98 y más años no especifica
ESCOLARIDA	Nominal	1..10, 99	Último grado aprobado en el ciclo de instrucción que declare haber cursado la persona en el Sistema Educativo Nacional o su equivalente en el caso de	1-Ninguno 2-Preescolar 3-Primaria 4-Secundaria 5-Estudios técnicos con secundaria terminada 6-Normal básica

			estudios en el extranjero; al momento de levantar la información.	7-Preparatoria o bachillerato 8-Estudios técnicos con preparatoria terminada 9-Licenciatura 10-Maestría o Doctorado 99-No especificada
OCUPACIÓN	Nominal	1... 11, 98 y 99	Conjunto de trabajos cuyas principales tareas y cometidos se caracterizan por tener un alto grado de similitud, independientemente del lugar en el que se desarrollen y de las relaciones que establezca en el mercado laboral.	1-Funcionarios, directores y jefes 2-Profesionistas y técnicos 3-Trabajadores auxiliares en actividades administrativas 4-Comerciantes, empleados en ventas y agentes de ventas 5-Trabajadores en servicios personales y vigilancia 6-Trabajadores en actividades agrícolas, ganaderas, forestales, caza y pesca 7-Trabajadores artesanales 8-Operadores de maquinaria industrial, ensambladores, choferes y conductores de transporte 9-Trabajadores en actividades elementales y de apoyo 10-Busca trabajo 11-No trabaja 98-Insuficientemente especificada 99-No especificada
ESTIM_FAM	Nominal	1, 2, 9	Estímulo familiar en la infancia Impulso familiar que se obtuvo durante la infancia para lograr el gusto sobre el conocimiento cultural.	1-Si 2-No 9-No especificado
MEDIO_TRAN	Numérico	1..7,9	Medio de transporte utilizado para llegar a recinto. Es la forma o vehículo que usó el visitante para llegar al museo.	1-Vehículo particular 2-Transporte público 3-Transporte turístico 4-Taxi 5-Bicicleta 6-Caminando 7-Otro 9-No especificado

TIEMPO_TRA	Nominal	1.... 9	Planeación de la visita. Toma de decisiones para la visita al museo.	1-De 1 a 30 min 2-De 31 min a 1h 3-De 1.01 a 1.30 h 4-De 1.31 a 2 h 5-De 2.01 a 3 h 6-De 3.01 a 4 h 7-De 4.01 a 5 h 8-De 5.01 y más 9-No especificado
OPIN_EXPOS	Nominal	1..6, 9	Opinión sobre las exposiciones. Juicio o idea que se forma el visitante en relación a la exhibición vista.	1-Muy buenas 2-Buenas 3-Malas 4-Muy malas 5-Regulares 6-No vi la o las explicaciones 9-No especificado
DUR_VIS_H	Numérico	0..8, 98, 99	Duración de la visita en horas. Tiempo transcurrido en horas desde el momento que el visitante ingresa a la institución hasta el momento en que sale de ella.	0..8-De 0 a 8 horas 98-Más de 8 horas 99-No especificado
DUR_VIS_M	Numérico	0..59, 99	Duración de la visita en minutos. Tiempo transcurrido en minutos desde el momento que el visitante ingresa a la institución hasta el momento en que sale de ella.	0..59-Cantidad de minutos 99-No especificado
EVAL_GRAL	Numérico	0...10, 99	Evaluación general	0..10-Calificación del 0 al 10 99-No especificado

Proceso:

Carga y Selección de datos

En esta sección se cargará los datos del archivo de Excel (Visitas21) y a través del Nodo Column Filter seleccionamos las columnas de las tablas que deseamos que estén.

Row ID	ANDO...	ENT_R...	MES_E...	DIA_EN...	SEXO	EDAD	ENT_R...	MUN_R...	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

Figura 1 Carga de Archivo Vistas21

Después, mediante el Nodo de Python, haremos un tratamiento a nuestros datos para que sean más entendibles.

```
1 import knime.scripting.io as knio
2
3 # Convirtiendo la tabla de ENTRADA a dataframe
4 df = knio.input_tables[0].to_pandas()
5
6
7 # TRATAMIENTO DE DATOS
8
9 # Meses de entrevista
10 df['MES_ENTREV'] = df['MES_ENTREV'].replace(1, 'Enero')
11 df['MES_ENTREV'] = df['MES_ENTREV'].replace(2, 'Febrero')
12 df['MES_ENTREV'] = df['MES_ENTREV'].replace(3, 'Marzo')
13 df['MES_ENTREV'] = df['MES_ENTREV'].replace(4, 'Abril')
14 df['MES_ENTREV'] = df['MES_ENTREV'].replace(5, 'Mayo')
15 df['MES_ENTREV'] = df['MES_ENTREV'].replace(6, 'Junio')
16 df['MES_ENTREV'] = df['MES_ENTREV'].replace(7, 'Julio')
17 df['MES_ENTREV'] = df['MES_ENTREV'].replace(8, 'Agosto')
18 df['MES_ENTREV'] = df['MES_ENTREV'].replace(9, 'Septiembre')
19 df['MES_ENTREV'] = df['MES_ENTREV'].replace(10, 'Octubre')
20 df['MES_ENTREV'] = df['MES_ENTREV'].replace(11, 'Noviembre')
21 df['MES_ENTREV'] = df['MES_ENTREV'].replace(12, 'Diciembre')
22
23 # Dias de entrevista
24 df['DIA_ENTREV'] = df['DIA_ENTREV'].replace(1, 'Entre Semana')
25 df['DIA_ENTREV'] = df['DIA_ENTREV'].replace(2, 'Fin de Semana')
26 df['DIA_ENTREV'] = df['DIA_ENTREV'].replace(9, 'No Especificado')
27
28 # Sexo
```

Figura 1.1 Nodo Python



Figura 1.2 Carga y selección

Nombres de entidades y municipios de residencia de los visitantes a los museos

En esta sección se cargan los archivos Excel (CENTIDAD y MPIO2021), a través del nodo Excel Reader.

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

Preview with current settings

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

Row ID	S	CVE_ENT	S	CVE_MUN	S	NOMBR...
Row0	1	1		1		Aguascalientes
Row1	1	2		2		Asientos
Row2	1	3		3		Calvillo
Row3	1	4		4		Cosío
Row4	1	5		5		Jesús María
Row5	1	6		6		Pabellón de...
Row6	1	7		7		Rincón de R...
Row7	1	8		8		San José de...
Row8	1	9		9		Tepic
Row9	1	10		10		El Llano
Row10	1	11		11		San Francisco...
Row11	1	999				No especificado
Row12	2	1		1		Ensenada
Row13	2	2		2		Mexicali

OK Apply Cancel ?

Figura 2 Carga de archivos

Una vez cargado los archivos pasaran a nuestros nodos joiner los cuales se encargarán de juntar los datos de nuestras tablas, y por último pasara a nuestro Nodo Column Rename, el cual nos permitirá cambiar el calor de los datos.

Nombres de entidades y municipios de residencia de los visitantes a los museos

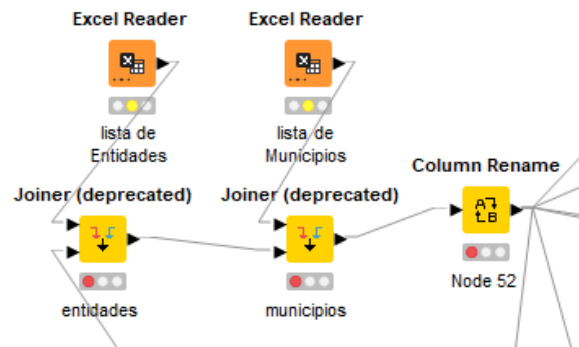


Figura 2.1 Nombre de entidades y Municipios

Consultas

Dentro del apartado consulta tendremos los nodos Group By los cuales se utilizarán para agrupar las filas de datos según un conjunto de criterios específicos y aplicar una función de agregación a cada grupo.

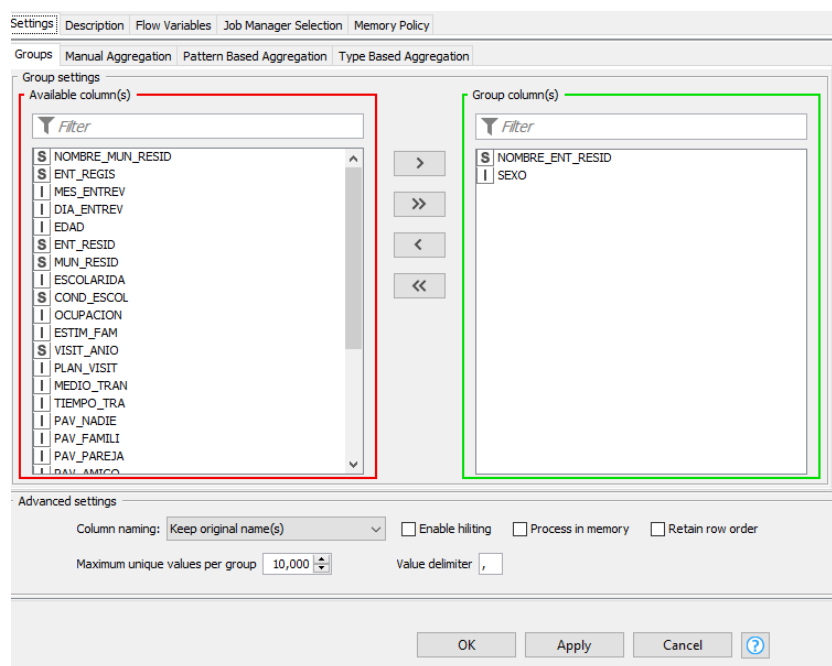


Figura 3 Group By Edad Promedio de Hombres y Mujeres

El nodo Table View nos permitirá tener una vista en HTML de la tabla.

File Edit Hilite Navigation View				
Table "default" - Rows: 46 Spec - Columns: 4 Properties Flow Variables				
Row ID	S NOMBRE...	I SEXO	D EDAD	B Selecte...
Row0	Durango	1	38.203	false
Row1	Durango	2	37.976	false
Row2	Guanajuato	1	34.727	false
Row3	Guanajuato	2	41.333	false
Row4	Guerrero	1	41	false
Row5	Guerrero	2	40.143	false
Row6	Hidalgo	1	33.867	false
Row7	Hidalgo	2	34.571	false
Row8	Jalisco	1	36.525	false
Row9	Jalisco	2	34.777	false
Row10	Michoacán d...	1	34.286	false
Row11	Michoacán d...	2	30.82	false
Row12	Morelos	1	24.75	false
Row13	Morelos	2	27.875	false
Row14	México	1	35.802	false
Row15	México	2	34.425	false
Row16	Nayarit	1	34.6	false
Row17	Nayarit	2	32.2	false
Row18	Nuevo León	1	32.167	false
Row19	Nuevo León	2	40.222	false
Row20	Oaxaca	1	35.447	false
Row21	Oaxaca	2	30.605	false
Row22	Puebla	1	36.059	false
Row23	Puebla	2	34.558	false
Row24	Querétaro	1	38.381	false

Figura 3.1 Nodo Table View

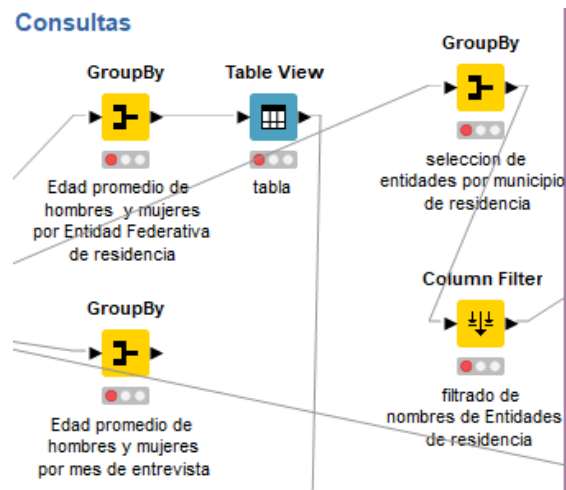


Figura 3.2 Consultas

Generacion de archivos de Excel de salida Cada hoja contiene la informacion de una entidad federativa de residencia de los visitantes

En el nodo Table Row se utiliza para iterar sobre las filas de una tabla y ejecutar un flujo de trabajo diferente para cada fila, el Nodo rule-based row filter SE ENCARGARA DE filtrar filas de una tabla según reglas predefinidas. El nodo permite la creacion de reglas personalizadas para filtrar datos extensos en condiciones especificas.

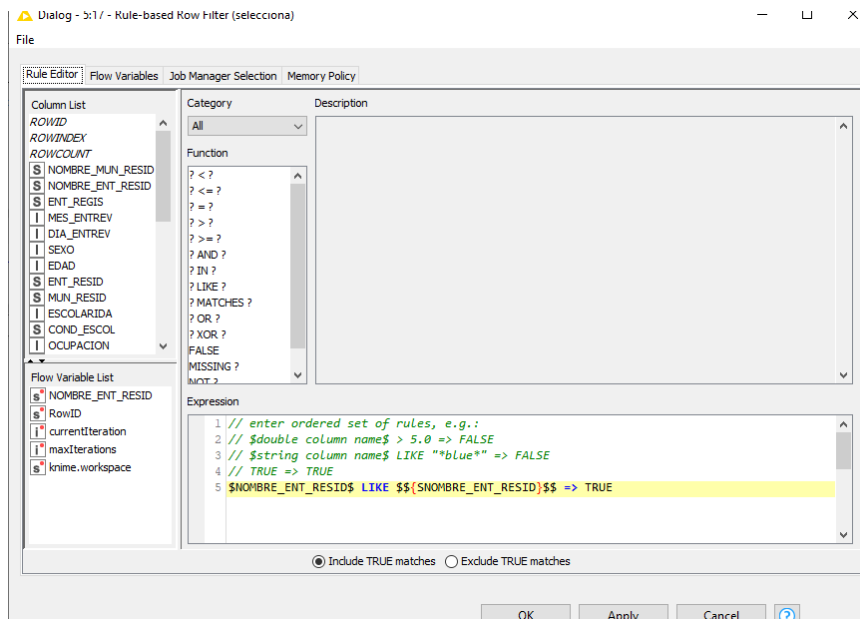


Figura 4 Nodo Row Filter

El nodo Excel Sheet Appender nos permitirá agregar nuevas filas a una hoja de cálculo de Excel existente, Variable Loop End nos permitirá marcar el final de un bucle que se ha creado utilizando el nodo "Table Row to Variable Loop Start".

Generación de archivos de Excel de salida.
Cada hoja contiene la información de una entidad federativa de residencia de los visitantes

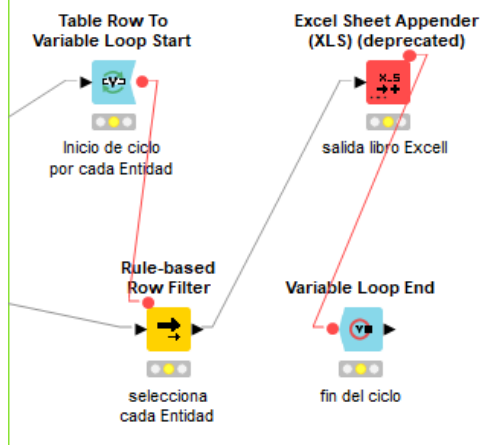


Figura 4.1 Generación de Archivos de Excel

Reclasificar las entidades federativas en Regiones

En esta sección se nos permitirá reclasificar las entidades federativas en Regiones esto a través del nodo Cell Replacer que se utilizará para reemplazar valores en una o varias celdas de una tabla, posteriormente tendremos el nodo Rule Engine el cual nos permitirá definir reglas Para la clasificación.

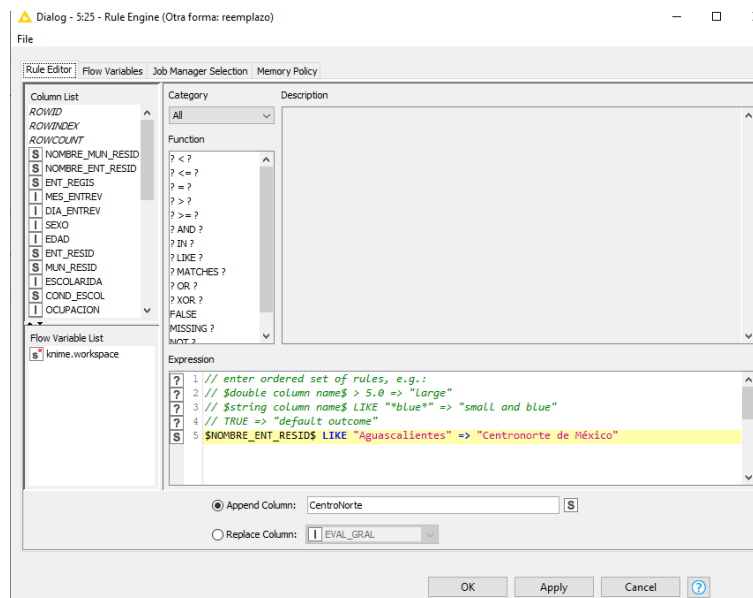


Figura 5 Nodo Rule Engine

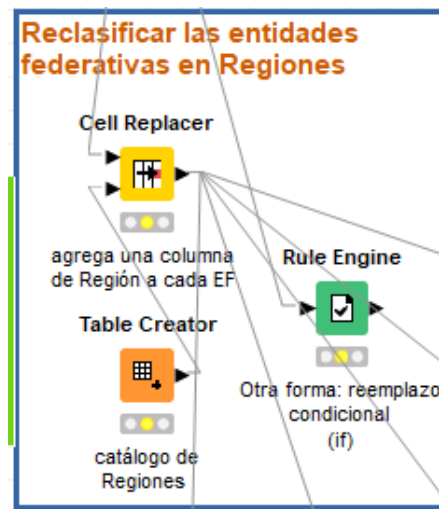


Figura 5.1 Reclasificar las entidades

Bidders del campo Edad

Se utiliza para crear una nueva variable categórica a partir de una variable numérica existente en una tabla, es decir nos permitirá agrupar en intervalos datos específicos que requiramos, Value counter se utilizará para contar el número de veces que aparece un valor específico en una columna de una tabla.

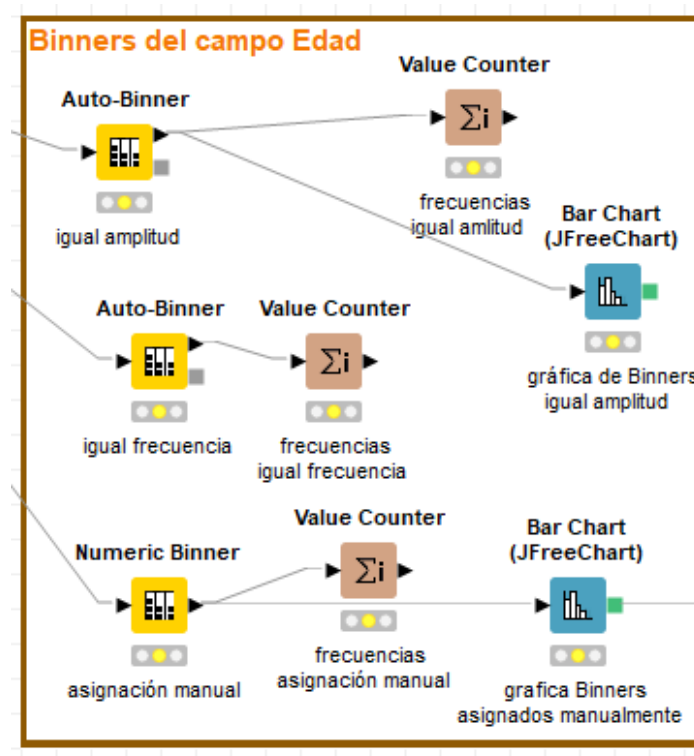


Figura 6

Código del Nodo “Python Script”:

Descripción del código:

El código implementado en este nodo reemplaza los valores numéricos presentes en algunas variables de tipo categórico (que hacen referencia a cada índice pertenecientes a un catálogo donde se enlistan los posibles valores que estas variables pueden tener), por los respectivos valores categóricos a los que se refieren.

Esto se realizó con el objetivo de hacer más clara y entendible la lectura de los datos de estas variables para su posterior análisis y presentación de resultados.

A continuación, se enlistan los catálogos pertenecientes a las variables a las que se les hicieron estos cambios:

Mes de Entrevista (MES_ENTREV): <table><tr><th>Valor</th><th>Etiqueta</th></tr><tr><td>1</td><td>Enero</td></tr><tr><td>2</td><td>Febrero</td></tr><tr><td>3</td><td>Marzo</td></tr><tr><td>4</td><td>Abril</td></tr><tr><td>5</td><td>Mayo</td></tr><tr><td>6</td><td>Junio</td></tr><tr><td>7</td><td>Julio</td></tr><tr><td>8</td><td>Agosto</td></tr><tr><td>9</td><td>Septiembre</td></tr><tr><td>10</td><td>Octubre</td></tr><tr><td>11</td><td>Noviembre</td></tr><tr><td>12</td><td>Diciembre</td></tr></table>	Valor	Etiqueta	1	Enero	2	Febrero	3	Marzo	4	Abril	5	Mayo	6	Junio	7	Julio	8	Agosto	9	Septiembre	10	Octubre	11	Noviembre	12	Diciembre	Día de Entrevista (DIA_ENTREV): <table><tr><th>Valor</th><th>Etiqueta</th></tr><tr><td>1</td><td>Entre semana</td></tr><tr><td>2</td><td>Fin de semana</td></tr><tr><td>9</td><td>No especificado</td></tr></table>	Valor	Etiqueta	1	Entre semana	2	Fin de semana	9	No especificado
Valor	Etiqueta																																		
1	Enero																																		
2	Febrero																																		
3	Marzo																																		
4	Abril																																		
5	Mayo																																		
6	Junio																																		
7	Julio																																		
8	Agosto																																		
9	Septiembre																																		
10	Octubre																																		
11	Noviembre																																		
12	Diciembre																																		
Valor	Etiqueta																																		
1	Entre semana																																		
2	Fin de semana																																		
9	No especificado																																		
Genero de la persona (SEXO): <table><tr><th>Valor</th><th>Etiqueta</th></tr><tr><td>1</td><td>Hombre</td></tr><tr><td>2</td><td>Mujer</td></tr></table>	Valor	Etiqueta	1	Hombre	2	Mujer	Estimulo Familiar (ESTIM_FAM): <table><tr><th>Valor</th><th>Etiqueta</th></tr><tr><td>1</td><td>SI</td></tr><tr><td>2</td><td>NO</td></tr><tr><td>9</td><td>No especificado</td></tr></table>	Valor	Etiqueta	1	SI	2	NO	9	No especificado																				
Valor	Etiqueta																																		
1	Hombre																																		
2	Mujer																																		
Valor	Etiqueta																																		
1	SI																																		
2	NO																																		
9	No especificado																																		
PERSONA QUE ACOMPAÑA AL VISITANTE																																			
Nadie (PAV_NADIE): <table><tr><th>Valor</th><th>Etiqueta</th></tr><tr><td>0</td><td>Con compañía</td></tr><tr><td>1</td><td>Sin compañía</td></tr></table>	Valor	Etiqueta	0	Con compañía	1	Sin compañía	Familia (PAV_FAMILIA): <table><tr><th>Valor</th><th>Etiqueta</th></tr><tr><td>0</td><td>Con compañía</td></tr><tr><td>1</td><td>Sin compañía</td></tr></table>	Valor	Etiqueta	0	Con compañía	1	Sin compañía																						
Valor	Etiqueta																																		
0	Con compañía																																		
1	Sin compañía																																		
Valor	Etiqueta																																		
0	Con compañía																																		
1	Sin compañía																																		

Pareja/Novio(a) (PAV_PAREJA):	Amigos/Conocidos (PAV_AMIGO):												
<table> <tr> <th>Valor</th><th>Etiqueta</th></tr> <tr> <td>0</td><td>Sin compañía</td></tr> <tr> <td>1</td><td>Con compañía</td></tr> </table>	Valor	Etiqueta	0	Sin compañía	1	Con compañía	<table> <tr> <th>Valor</th><th>Etiqueta</th></tr> <tr> <td>0</td><td>Sin compañía</td></tr> <tr> <td>1</td><td>Con compañía</td></tr> </table>	Valor	Etiqueta	0	Sin compañía	1	Con compañía
Valor	Etiqueta												
0	Sin compañía												
1	Con compañía												
Valor	Etiqueta												
0	Sin compañía												
1	Con compañía												
Compañeros del trabajo (PAV_COMPA):	Grupo Escolar (PAV_ESCOLA):												
<table> <tr> <th>Valor</th><th>Etiqueta</th></tr> <tr> <td>0</td><td>Sin compañía</td></tr> <tr> <td>1</td><td>Con compañía</td></tr> </table>	Valor	Etiqueta	0	Sin compañía	1	Con compañía	<table> <tr> <th>Valor</th><th>Etiqueta</th></tr> <tr> <td>0</td><td>Sin compañía</td></tr> <tr> <td>1</td><td>Con compañía</td></tr> </table>	Valor	Etiqueta	0	Sin compañía	1	Con compañía
Valor	Etiqueta												
0	Sin compañía												
1	Con compañía												
Valor	Etiqueta												
0	Sin compañía												
1	Con compañía												
Grupo de esparcimiento (turístico, recreativo, etc.) (PAV_TURIST):	Otro (PAV_OTRO):												
<table> <tr> <th>Valor</th><th>Etiqueta</th></tr> <tr> <td>0</td><td>Sin compañía</td></tr> <tr> <td>1</td><td>Con compañía</td></tr> </table>	Valor	Etiqueta	0	Sin compañía	1	Con compañía	<table> <tr> <th>Valor</th><th>Etiqueta</th></tr> <tr> <td>0</td><td>Sin compañía</td></tr> <tr> <td>1</td><td>Con compañía</td></tr> </table>	Valor	Etiqueta	0	Sin compañía	1	Con compañía
Valor	Etiqueta												
0	Sin compañía												
1	Con compañía												
Valor	Etiqueta												
0	Sin compañía												
1	Con compañía												

Código:

```
import knime.scripting.io as knio
Convirtiendo la tabla de ENTRADA a dataframe
df = knio.input_tables[0].to_pandas()

# TRATAMIENTO DE DATOS
# Meses de entrevista
df['MES_ENTREV']=df['MES_ENTREV'].replace(1, 'Enero')
df['MES_ENTREV']=df['MES_ENTREV'].replace(2, 'Febrero')
df['MES_ENTREV']=df['MES_ENTREV'].replace(3, 'Marzo')
df['MES_ENTREV']=df['MES_ENTREV'].replace(4, 'Abril')
df['MES_ENTREV']=df['MES_ENTREV'].replace(5, 'Mayo')
df['MES_ENTREV']=df['MES_ENTREV'].replace(6, 'Junio')
df['MES_ENTREV']=df['MES_ENTREV'].replace(7, 'Julio')
df['MES_ENTREV']=df['MES_ENTREV'].replace(8, 'Agosto')
df['MES_ENTREV']=df['MES_ENTREV'].replace(9, 'Septiembre')
df['MES_ENTREV']=df['MES_ENTREV'].replace(10, 'Octubre')
df['MES_ENTREV']=df['MES_ENTREV'].replace(11, 'Noviembre')
df['MES_ENTREV']=df['MES_ENTREV'].replace(12, 'Diciembre')
Dias de entrevista
df['DIA_ENTREV']=df['DIA_ENTREV'].replace(1, 'Entre Semana')
df['DIA_ENTREV']=df['DIA_ENTREV'].replace(2, 'Fin de Semana')
df['DIA_ENTREV']=df['DIA_ENTREV'].replace(9, 'No Especificado')
# Sexo
df['SEXO'] = df['SEXO'].replace(1, 'Hombre') df['SEXO'] =
df['SEXO'].replace(2, 'Mujer')
# Estimulo Familiar
df['ESTIM_FAM']=df['ESTIM_FAM'].replace(9, 'No Especificado')
df['ESTIM_FAM']=df['ESTIM_FAM'].replace(1, 'SI')
df['ESTIM_FAM']=df['ESTIM_FAM'].replace(2, 'NO')
# Persona que acompaña al visitante
df['PAV_NADIE']=df['PAV_NADIE'].replace(0, 'Con Compañía')
df['PAV_NADIE']=df['PAV_NADIE'].replace(1, 'Sin Compañía')
df['PAV_FAMILI']=df['PAV_FAMILI'].replace(0, 'Con Compañía')
df['PAV_FAMILI']=df['PAV_FAMILI'].replace(1, 'Sin Compañía')
```

```
df['PAV_PAREJA']=df['PAV_PAREJA'].replace(0, 'Sin Compañia')
df['PAV_PAREJA']=df['PAV_PAREJA'].replace(1, 'Con Compañia')
df['PAV_AMIGO']=df['PAV_AMIGO'].replace(0, 'Sin Compañia')
df['PAV_AMIGO']=df['PAV_AMIGO'].replace(1, 'Con Compañia')
df['PAV_COMPA']=df['PAV_COMPA'].replace(0, 'Sin Compañia')
df['PAV_COMPA']=df['PAV_COMPA'].replace(1, 'Con Compañia')
df['PAV_ESCOLA']=df['PAV_ESCOLA'].replace(0, 'Sin Compañia')
df['PAV_ESCOLA']=df['PAV_ESCOLA'].replace(1, 'Con Compañia')
df['PAV_TURIST']=df['PAV_TURIST'].replace(0, 'Sin Compañia')
df['PAV_TURIST']=df['PAV_TURIST'].replace(1, 'Con Compañia')
df['PAV_OTRO']=df['PAV_OTRO'].replace(0, 'Sin Compañia')
df['PAV_OTRO']=df['PAV_OTRO'].replace(1, 'Con Compañia')

# Convirtiendo el dataframe a un dataTable e igualandolo a la tabla de
SALIDA knio.output_tables[0] = knio.Table.from_pandas(df)
```