

Preparación de datos con Knime



**UNIVERSIDAD
DE GRANADA**

Pablo Morenilla Pinos
morenillapablo@correo.ugr.es
TID Prácticas Grupo 1

Índice

1. Factores de riesgo para el cancer de cuello uterino	3
1.1. Lectura e interpretación de los datos	5
1.2. Tratamiento de los datos/valores perdidos.	8
1.3. Visualización de los datos.	15
2. Descripción de la base de datos: gravedad en accidentes de tráfico	21
2.1. Gestión de Valores Desconocidos	21
2.2. Partición y entrenamiento de los datos	21
2.3. Construcción de la Variable Clase	21
2.4. Operaciones de Preprocesamiento	22
2.4.1. Discretización	22

Índice de figuras

1.	Lectura fichero csv en Knime.	5
2.	Lectura fichero CSV desde VS Code.	5
3.	Ejemplo datos DX.	7
4.	Double to Integer.	8
5.	Eliminar columnas con información perdida.	9
6.	Asignación valores perdidos parte 1.	10
7.	Asignación valores perdidos parte 2.	11
8.	Filtrado por correlación de 0.8.	12
9.	Filtrado por correlación de 0.8.	12
10.	Configuración de la PCA.	13
11.	Configuración feature selection Loop Start.	13
12.	Configuración del particionamiento de los datos.	14
13.	Configuración del Scorer.	14
14.	Filtrado de los datos con predicción del 0.6.	15
15.	Rangos para el campo edad.	16
16.	Rangos para el campo primer encuentro.	16
17.	Rango de número de embarazos.	17
18.	Colores por rango de edades.	17
19.	Parejas por embarazos.	18
20.	Parejas sexuales por edad.	18
21.	Frecuencia diagnóstico respecto parejas sexuales.	19
22.	Frecuencia de edades que usan DIU.	19
23.	Creación columna clase.	21
24.	Borrado columnas.	22
25.	Configuración decision tree learner.	23
26.	Scorer obtenido con la predicción de la columna clase.	23
27.	Nodo CAIM.	24
28.	Ha ingerido alcohol o no respecto al mes y en función de si es un accidente o no.	25
29.	Velocidades máximas respecto al mes y día de la semana.	25
30.	Tipo de accidentes respecto a las señales.	26

1. Factores de riesgo para el cancer de cuello uterino

Considera el fichero *Risk_factors_cervical_cancer.csv* conteniendo datos médicos de mujeres relativos a un estudio sobre los factores de riesgo para el cáncer de cuello uterino. En concreto, se investiga si los factores de riesgo sugieren una biopsia. Las características que se estudian son:

- Age (Edad)
- Number of sexual partners (Número de parejas sexuales)
- First sexual intercourse (Primera relación sexual)
- Num of pregnancies (Número de embarazos)
- Smokes (Fumadora)
- Smokes (years) (Años de fumadora)
- Smokes (packs/year) (Paquetes al año)
- Hormonal Contraceptives (Hormonas anticonceptivas)
- Hormonal Contraceptives (years) (Años tomando hormonas anticonceptivas)
- IUD (Usa DIU)
- IUD (years) (Número de años usando DIU)
- STDs (Enfermedades de Transmisión Sexual)
- STDs (number) (Número de ETS)
- STDs:condylomatosis (Condilomatosis)
- STDs:cervical condylomatosis (Condilomatosis cervical)
- STDs:vaginal condylomatosis (Condilomatosis vaginal)
- STDs:vulvo-perineal condylomatosis (Condilomatosis vulvo-perineal)
- STDs:syphilis (Sífilis)
- STDs:pelvic inflammatory disease (Enfermedad pélvica inflamatoria)
- STDs:genital herpes (Herpes genital)
- STDs:molluscum contagiosum (Molusco contagioso)
- STDs:AIDS (SIDA)
- STDs:HIV (Virus VIH)
- STDs:Hepatitis B (Hepatitis B)
- STDs:HPV (Papiloma humano)
- STDs: Number of diagnosis (Número de diagnósticos)

-
- STDs: Time since first diagnosis (Tiempo desde el primer diagnóstico)
 - STDs: Time since last diagnosis (Tiempo desde el último diagnóstico)
 - Dx:Cancer (Diagnóstico de cáncer)
 - Dx:CIN (Diagnóstico de neoplasia intraepitelial cervical)
 - Dx:HPV (Diagnóstico de virus del papiloma humano)
 - Dx (Algún diagnóstico de los anteriores)
 - Hinselmann (Se realizó colposcopia)
 - Schiller (Se realizó prueba de Schiller)
 - Cytology (Se realizó citología)
 - Biopsy (Se realizó biopsia)

Analiza qué acciones se pueden realizar y por qué, para una preparación de datos adecuada utilizando los nodos de KNIME existentes. Hay que tener en cuenta también, entre otros, los siguientes aspectos:

- Los tipos de datos. Se deben analizar qué tipos de datos son los que corresponden con cada característica, y si KNIME los ha identificado correctamente al leer el fichero. En caso contrario se deben cambiar.
- Tratamiento de valores perdidos.
- Visualizar los datos, para obtener información de las características.
- A partir de dicha visualización, posible discretización de algunas variables (hay varias técnicas), transformación de variables (también hay varias técnicas), estudiar correlaciones entre ciertas variables. Hay que tener en cuenta qué significan las variables.
- Posible eliminación de variables teniendo en cuenta lo anterior.

Comenzamos leyendo los datos del csv, para poder ver los tipos con los que trata Knime cada dato y tener una idea general de cuantos datos disponemos:



Si vemos los datos que hay en el CSV desde otro visor de ficheros como es por ejemplo Visual Studio Code se ven que hay valores vacíos como se ha mencionado anteriormente, estos valores vacíos tienen como valor un el símbolo `'?'`, y los datos que ha leído el nodo en Knime son valores numéricos correctamente:

Figura 2: Lectura fichero CSV desde VS Code.

Valores perdidos.

Hay que mencionar que desde Visual Studio Code, se puede ver en la imagen anterior, por ejemplo que los valores los muestra con su decimal, pero en Knime no, solo lo interpreta como tipo Number(Double), si tomamos como ejemplo la columna 'Num of pregnancies', desde VSCode lee los valores con decimales (1.0, 3.0, 4.0, etc.) pero desde Knime no, los interpreta como 1,3 o 4. En este caso sería apropiado transformar estos tipos de columnas a tipo entero, ya que no puede existir el caso de número de embarazos con decimales. Esto también se puede ver en la columna 'Number of sexual partners' que en Knime se interpretan como tipo Number(Double) pero no muestra el decimal y, de igual modo, debería de modificarse el tipo de dato para que fuera a tipo entero, ya que no puede darse el caso de tener media pareja sexual o una pareja y media, serían valores de tipo entero. Hay más ejemplos de esta modificación a realizar, más adelante se irán detallando cuáles son

Por otro lado, hay valores que si los visualizamos desde VSCode se pueden ver cómo tienen valores decimales de más de 4 cifras, pero en el caso de Knime los omite y muestra solo valores decimales de hasta 3 decimales. Ejemplo de ello es el caso de la columna '**smoked years**' donde la fila 9 tienen un valor de '1.266972909' pero desde Knime hace un redondeo y da el valor de '1.267', esto puede suponer algún problema, ya que se pierde información relevante por el camino y aquí si puede llegar a importar el tiempo que ha estado fumando. De igual modo, ocurre lo mismo para la columna '**smoked (packs/years)**' así que habría que tener en cuenta que se pierden decimales a la hora de leer los datos.

Valores booleanos.

Hay ciertas columnas que aunque Knime y Visual Studio Code los interpreta como números enteros y números decimales respectivamente, se podrían interpretar como valores de '**Sí/No**' o lo que es lo mismo, como booleanos, ya que son en las columnas que tienen como información si el paciente ha realizado, no realizado una prueba o si ha tenido diagnóstico de algún tipo, estos datos son: 'Smoked, IUD, STD's, Dx:Cancer, Dx:CIN, Dx:HPV, Dx, Hinselmann, Schiller, Citology, Biopsy'. De modo que, habría que tratar estos datos como si fueran tipos booleanos posteriormente o como en el caso de la práctica anterior con el caso de las familias, asignarles un valor de tipo cadena 'Si/No'.

Datos erróneos.

Tras lo mencionado anteriormente sobre los valores 0/1, se dan casos en los que el valor de la columna Dx, (si ha tenido al menos uno de los diagnósticos de cáncer, neoplasia intraepitelial cervical o virus de papiloma humano) no son adecuados, ya que por ejemplo si miramos los datos:

Dx:Cancer <i>Number (integer)</i>	Dx:CIN <i>Number (inte...</i>	Dx:HPV <i>Number (inte...</i>	Dx <i>Number (inte...</i>
0	0	0	0
0	0	0	0
0	0	0	0
1	0	1	0

Figura 3: Ejemplo datos DX.

Se puede ver que ese ejemplo, la 4 fila ha tenido diagnóstico de cáncer y de HPV, pero el valor de la columna Dx en este caso es 0 y no tendría que ser así, ya que la descripción de esta columna es si ha tenido alguno de los diagnósticos anteriores, de modo que, habría que modificar los valores de esta columna en los casos en los que estén mal y cambiar a 1 si la paciente ha tenido algún diagnóstico de los 3 anteriores mencionados anteriormente o 0 si no ha tenido ningún diagnóstico previo o de otra manera como se menciona en el apartado anterior 'Sí/No'.

Asociaciones entre columnas.

Tras lo mencionado con los valores de STD y diagnósticos se puede ver que hay relaciones entre los tipos de datos, ya que el campo STD será de tipo 1 si al menos ha tenido una de las STD mencionadas, de este modo, se puede rellenar los datos vacíos en caso de ser necesario, ya que, si entre todas las STD al menos hay una con valor igual a 1, la columna STD tendrá como valor también 1. Otro punto es que, si por ejemplo, todos los valores relacionados con STD estuvieran vacíos con el símbolo '?', se puede averiguar si es 0 o 1 con lo mencionado.

1.2. Tratamiento de los datos/valores perdidos.

Con todo lo analizado en el apartado anterior de lectura e interpretación de los datos, vamos a ver qué columnas son las que presentan mayor cantidad de información perdida, qué filas tienen valores perdidos, si son posibles obtener o asignarles alguno a partir de lo visto previamente y modificar el tipo de columna de los datos al tipo que mejor le conviene.

Double to Integer

Primero cambiamos el tipo de los valores descritos anteriormente de tipo Double a Integer con el nodo 'Double to Integer':

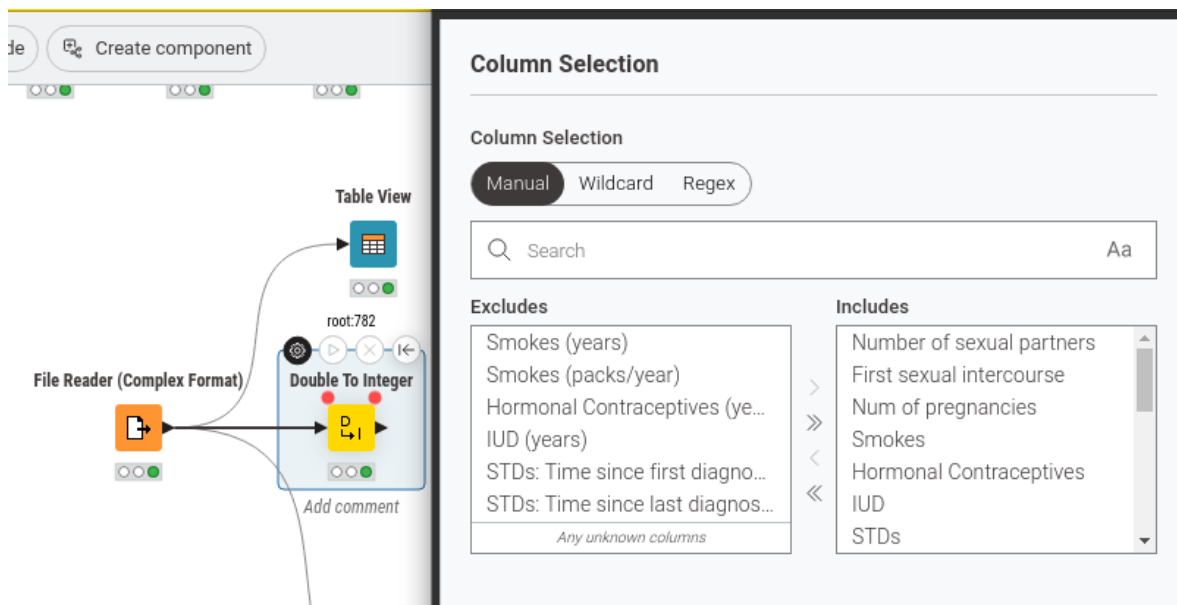


Figura 4: Double to Integer.

Los valores cambiados son: 'Number of sexual partners', 'First sexual intercourse', 'Num of pregnancies', 'Smokes', 'Hormonal Contraceptives', 'IUD', y todas las columnas 'STDs' a excepción de 'first and last diagnosis'.

Columnas con información perdida. Ahora vamos a eliminar aquellas columnas que tienen al menos más del 70 por ciento de información perdida, ya que no se puede tratar bien con tan poca información de las columnas, para ello usamos el nodo 'Missing Column Value Filter':

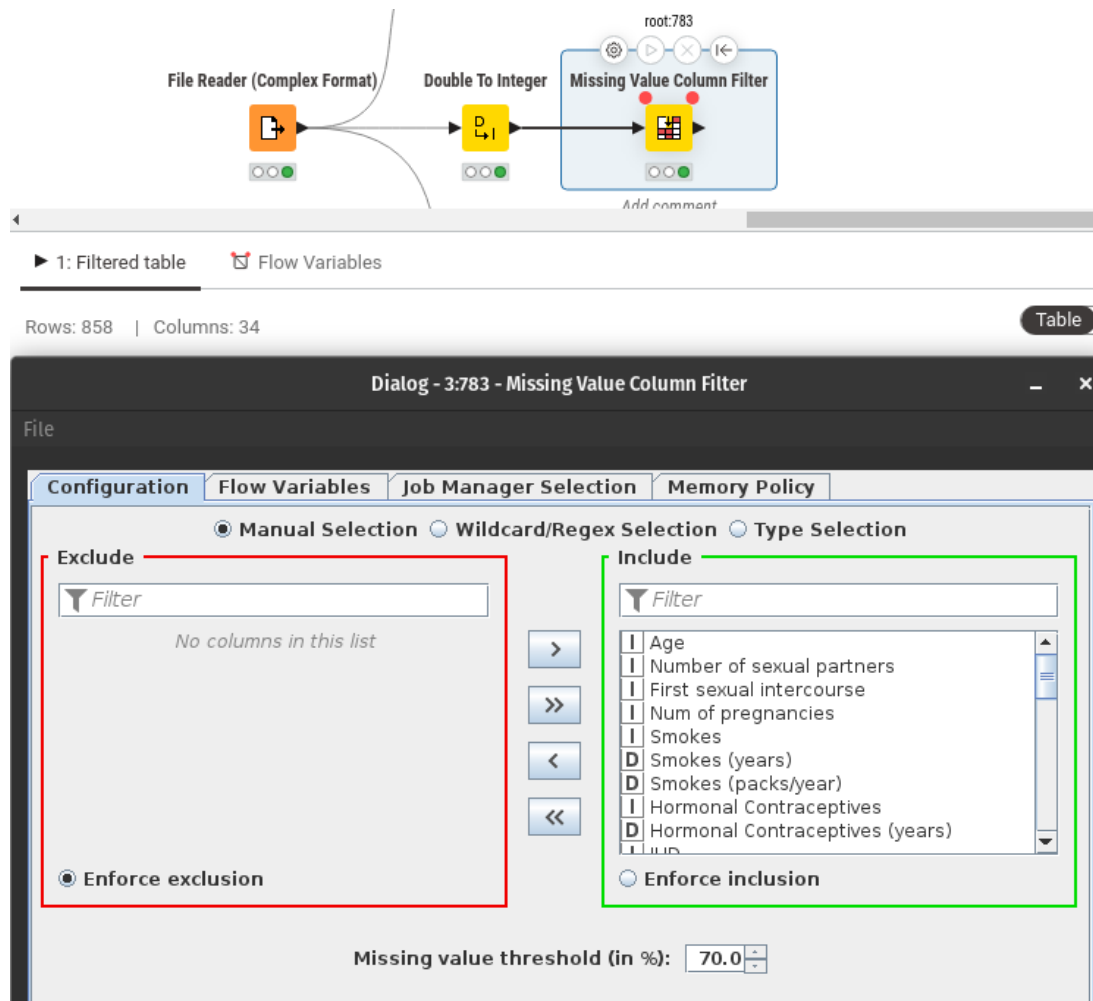


Figura 5: Eliminar columnas con información perdida.

En la imagen se puede ver que ha pasado de 36 columnas a tener 34 columnas, luego había dos columnas con un 70 % o más de información perdida, en este caso eran las tablas 'STDs: Time since first diagnosis' y 'STDs: Time since last diagnosis'.

Relleno valores perdidos.

Para rellenar los datos perdidos vamos a asumir una serie de criterios para los datos a rellenar:

- Todas las columnas STD tendrán el valor mínimo, es decir 0, ya que no se puede asumir que ha tenido ni una ni tiene todas ni ha tenido alguna de las enfermedades de transmisión sexual.
- IUD e IUD(years), no podemos asumir que tiene un DIU y del mismo modo, si no tiene, no puede tener durante los años, por lo que se le asigna el valor mínimo, en este caso 0 y 0.0 respectivamente (por ser de tipo entero y double).
- Smokes, smokes(years) y smokes (packs/year), de igual modo que en el caso anterior, no se puede asumir que haya fumado alguna vez, del mismo modo, si no ha fumado

nunca, no ha fumado en ningún año ni tiene una relación paquetes/año, por lo que se le asigna el valor mínimo, 0 en este caso para las tres columnas.

- Número de embarazos(num of pregnancies), se asume que no ha tenido ningún embarazo, puesto que si no tiene datos se puede asumir que tiene 0.
- Hormonal Contraceptives y Hormonal Contraceptives(years), del mismo modo que en el caso de los fumadores y del DIU, no se puede asumir que haya tenido, puesto que sería un dato muy importante a la hora de hacer las estimaciones posteriores, del mismo modo, si no ha usado hormonas anticonceptivas, no puede tener una relación de hormonas conceptivas por año.
- First sexual intercourse y Number of sexual partners, en este caso, no se puede dar un valor mínimo o 0 para estos casos, puesto que no se puede asumir que no haya tenido ni relaciones ni se puede asumir tampoco el número de parejas, por lo que se le asigna el valor con más frecuencia en cada caso.

Teniendo en cuenta estos criterios y, usando el nodo Missing Value establecemos los criterios de la siguiente manera:

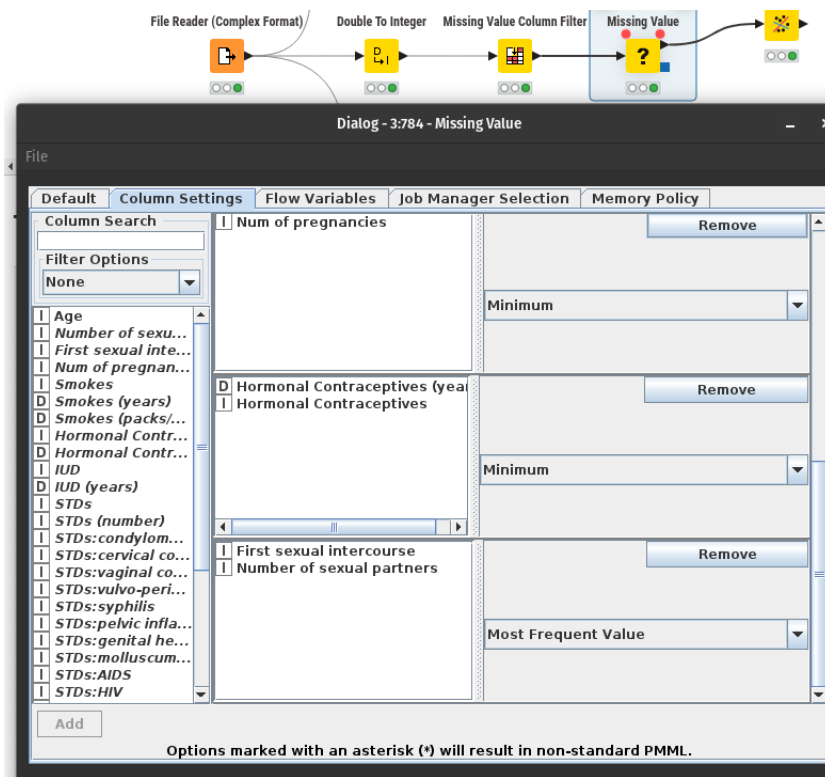


Figura 6: Asignación valores perdidos parte 1.

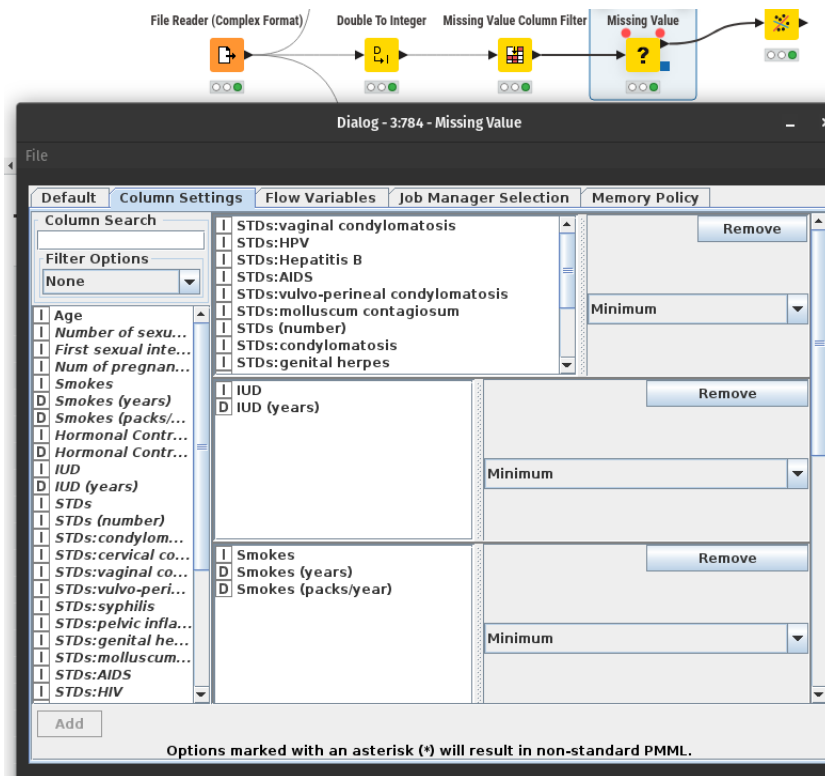


Figura 7: Asignación valores perdidos parte 2.

Una vez hecho esto, ya no hay valores perdidos o con el valor '??', y pueden hacerse más operaciones y visualización de los datos.

Dispersión de los datos.

Una vez tenemos todas las filas y columnas con datos, podemos hacer tratamiento de los datos y ver qué columnas podrían ser redundantes y eliminarlas de nuestro estudio, para ello usamos el nodo Linear Correlation donde establecemos todos los datos y luego con el nodo Correlation Filter establecemos un umbral por el cual eliminamos las columnas que tengan una correlación por encima del 80 por ciento.

Si vemos a través del nodo 'Linear Correlation' y ordenamos por los mayores resultados, vemos que hay 7 filas con relaciones entre ellas de más de 0.8:

#	Row...	First column name String	Second column name String	Correlation value ↓ Number (double)
354	Row...	STDs:condylomatosis	STDs:vulvo-perineal condylomatosis	0.988
309	Row...	STDs	STDs (number)	0.92
322	Row...	STDs	STDs: Number of diagnosis	0.908
331	Row...	STDs (number)	STDs:condylomatosis	0.9
343	Row...	STDs (number)	STDs: Number of diagnosis	0.898
334	Row...	STDs (number)	STDs:vulvo-perineal condylomatosis	0.891
535	Row...	Dx:Cancer	Dx:HPV	0.887
262	Row...	IUD	IUD (years)	0.749
560	Row...	Schiller	Biopsy	0.733
310	Row...	STDs	STDs:condylomatosis	0.73

Figura 8: Filtrado por correlación de 0.8.

Una vez obtenidas las correlaciones, se eliminan aquellas columnas que no sean necesarias para el estudio de los datos con el nodo 'Correlation Filter'.

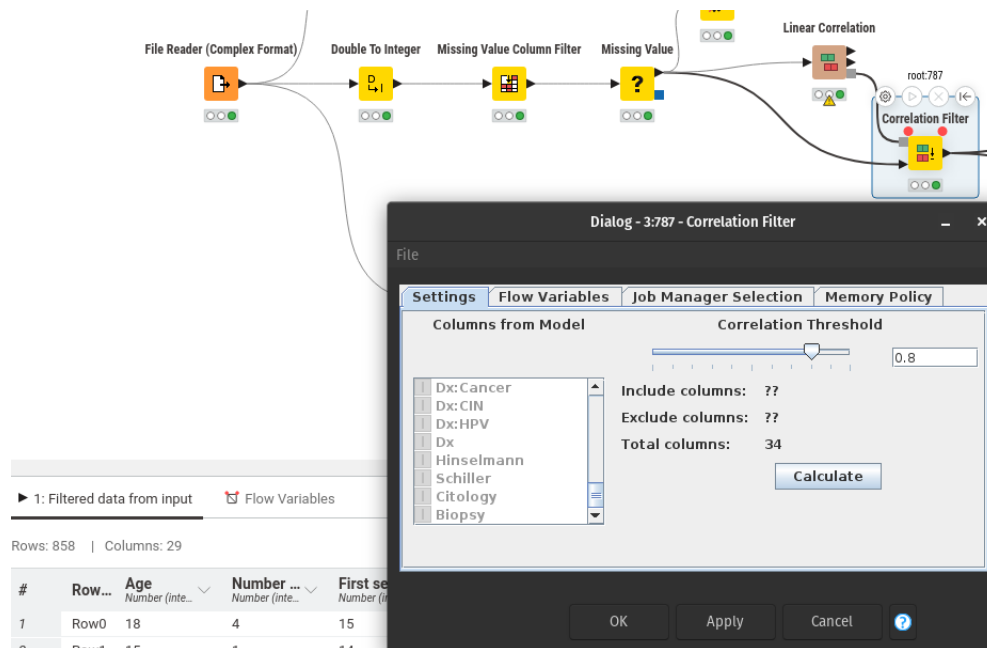


Figura 9: Filtrado por correlación de 0.8.

Como se puede ver en la imagen, se ha pasado de tener 34 columnas a tener 29, siendo 5 columnas las excluidas 'STD's condylomatis', 'STDs:vulvo-perineal condylomatosis', 'STDs: Number of diagnosis', 'STD' y 'Dx:HPV'.

Uso de PCA y nodo Backward Feature Elimination

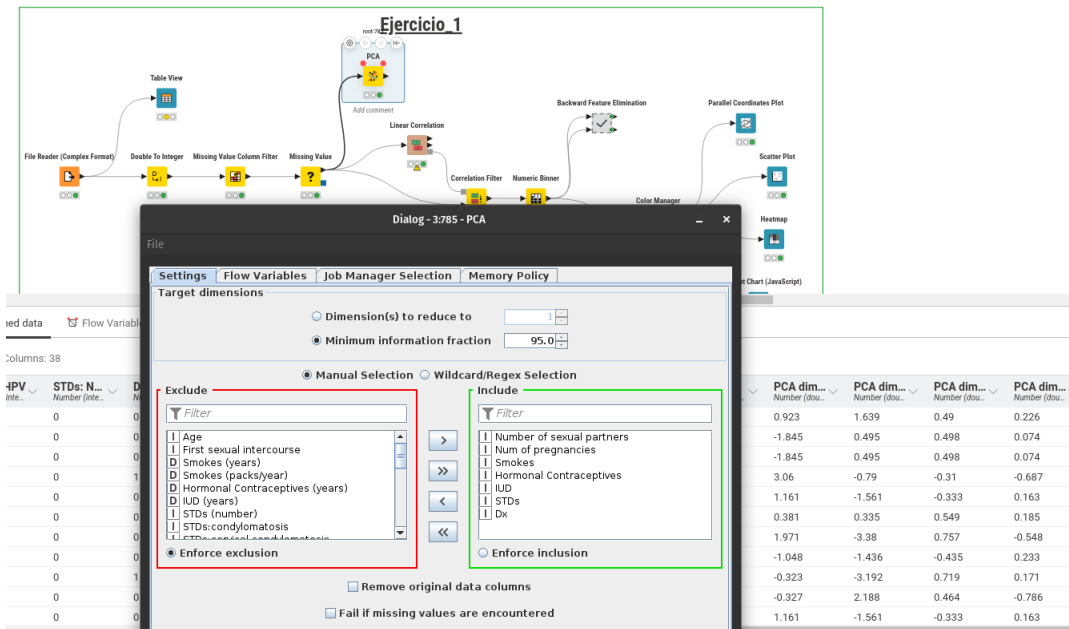


Figura 10: Configuración de la PCA.

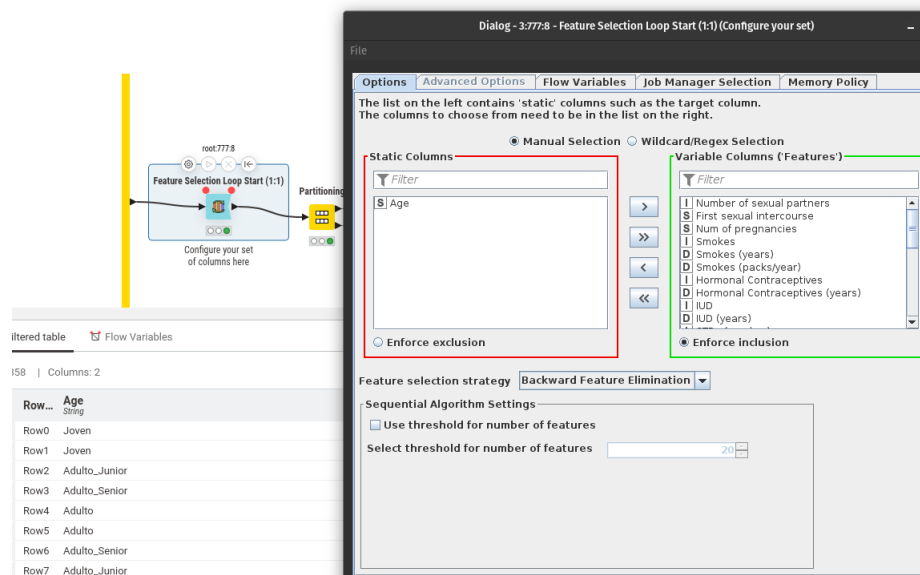


Figura 11: Configuración feature selection Loop Start.

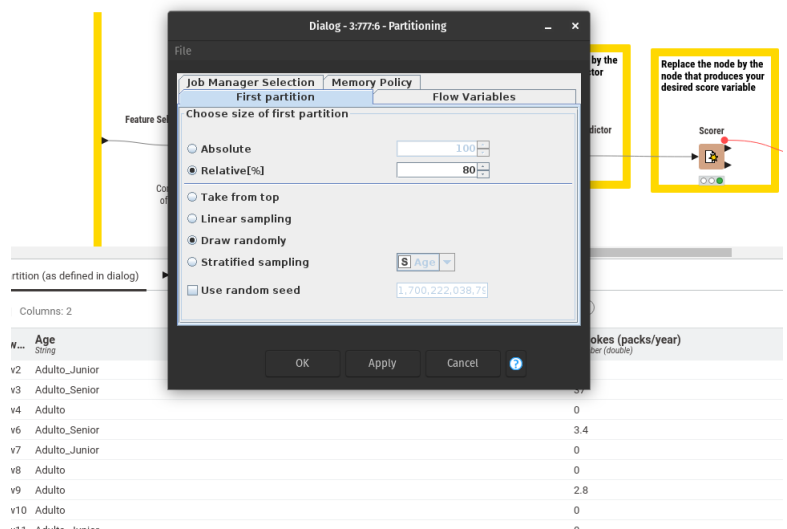


Figura 12: Configuración del particionamiento de los datos.

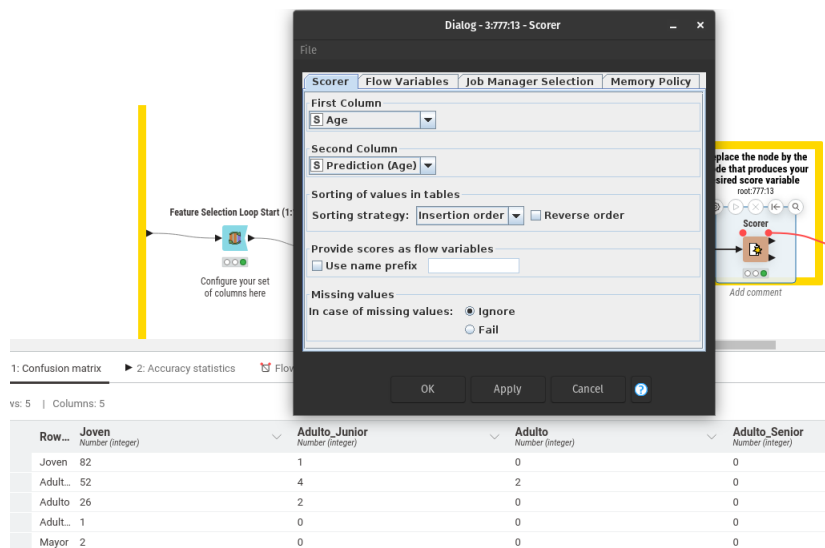


Figura 13: Configuración del Scorer.

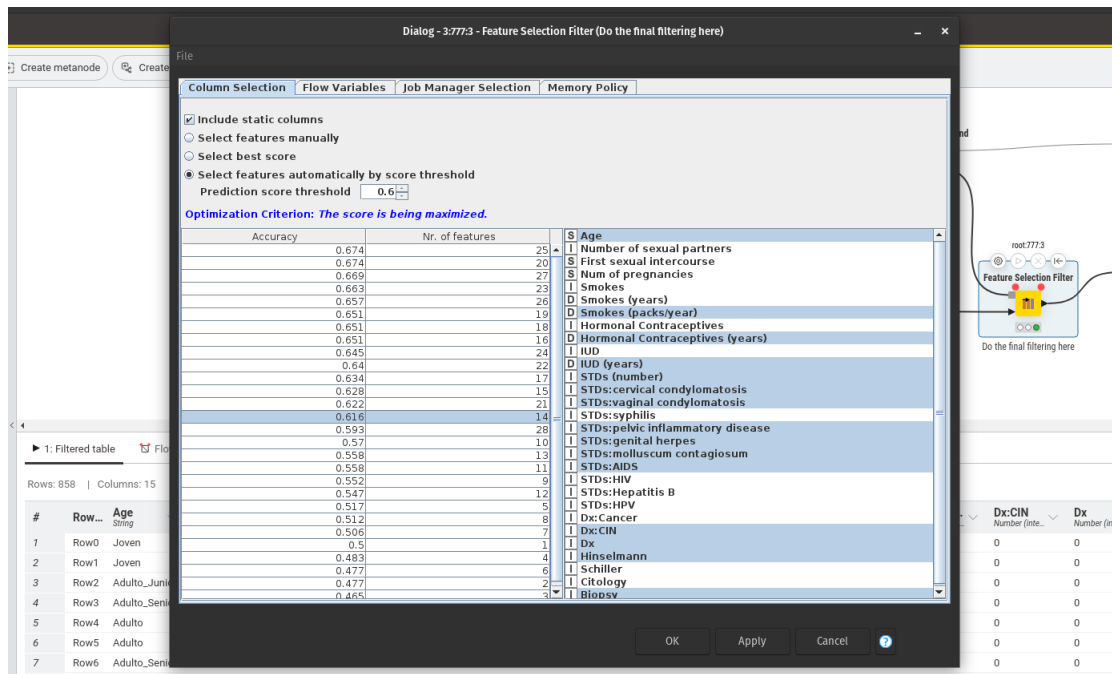


Figura 14: Filtrado de los datos con predicción del 0.6.

1.3. Visualización de los datos.

A continuación vamos a hacer una visualización de los datos con los nodos vistos en la práctica anterior, pero previamente a visualizarlos, vamos a hacer modificaciones sobre los datos para poder verlos de mejor manera, en este caso usando los nodos Numeric Binner y Auto-Binner. Con estos nodos establecemos rangos dentro los datos para poder tratarlos en menor cantidad, pero dividiendo estos en grupos, ejemplo de uso sería con los datos de la columna edad, que varía desde los 13 hasta los 84 años, por lo que se dividirá en 5 grupos, se podría usar el nodo Auto-Binner que automáticamente los divide en intervalos, pero el problema principal es que no le establece un nombre adecuado ni podemos establecer nosotros el rango, por lo que usamos Numeric Binner.

Como se puede ver en la siguiente imagen, se dividen los rangos de edad en cinco rangos:

- Joven: Desde el mínimo valor, 13 en este caso hasta los 25 años sin incluirlo.
- Adulto Junio: Desde los 25 hasta los 35 años sin incluirlo.
- Adulto: Desde los 35 hasta los 50 años sin incluirlo.
- Adulto Senior: Desde los 50 años hasta los 65 sin incluirlo.
- Mayor: De 65 años hasta la última edad, en este caso 84 años.

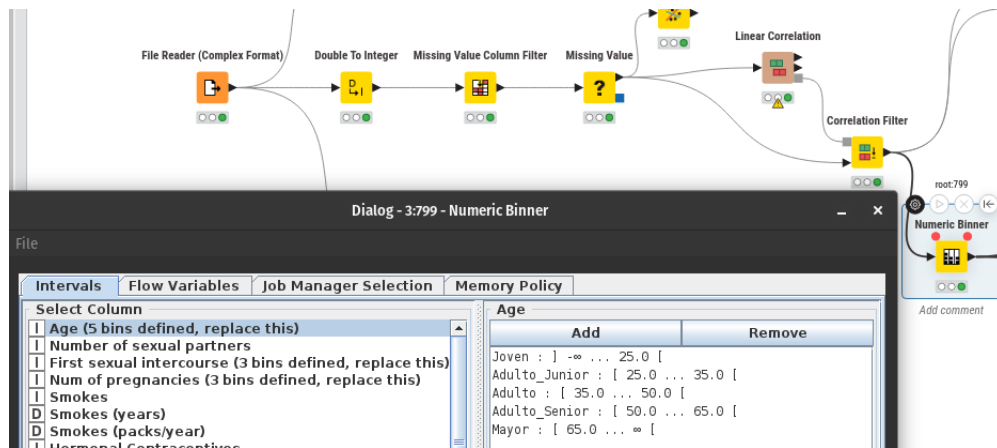


Figura 15: Rangos para el campo edad.

Del mismo modo, establecemos rangos para el primer encuentro sexual en 3 rangos, puesto que el menor valor es de 10 años y el mayor de 32:

- Menor de 18 años.
- Entre 18 y 25 años.
- Más de 25 años.

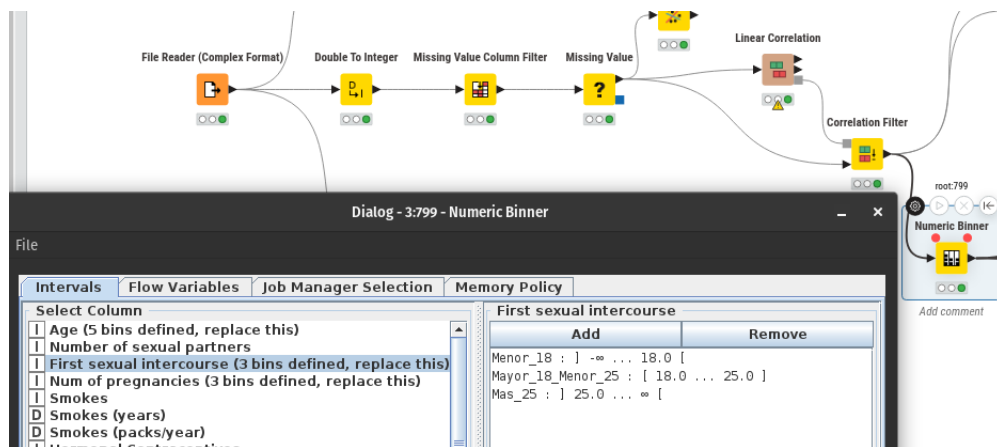


Figura 16: Rangos para el campo primer encuentro.

Por último, establecemos rangos del número de embarazos que ha tenido cada mujer, en este caso el mínimo es 0 y el máximo de 11, lo dividimos en 3 grupos, ya que se podría considerar menos de 2 poco, entre dos y 4 un estándar y más de 4 como por encima de la media.

- Menos de 2 embarazos.
- Entre dos y 4 embarazos.
- Más de 4 embarazos.

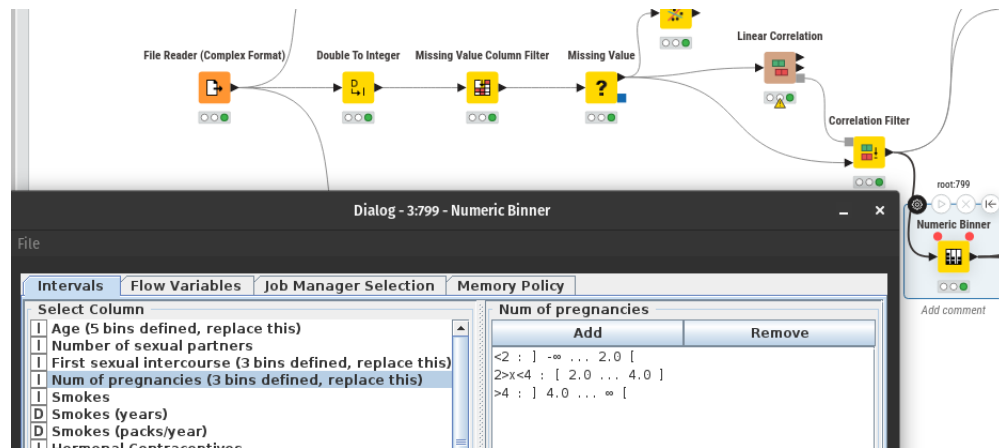


Figura 17: Rango de número de embarazos.

Una vez hemos establecido rangos adecuados para los tipos de datos mencionados anteriormente, usamos el nodo 'Color Manager' para establecer rangos de colores para cada valor de los que hemos subdividido y visualizar así los datos de mejor manera.

Por ejemplo, usando un mapa de calor podemos ver los rangos de edades que tienen más frecuencia con el uso de DIU, número de enfermedades, si han tenido o no alguna prueba de cáncer. Con los rangos de edades es más fácil interpretar los datos para posteriores operaciones que si, por el contrario, viéramos un mapa de calor de todos los datos que había.

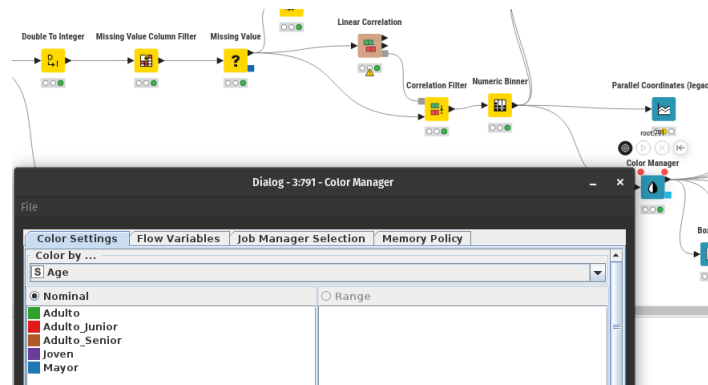


Figura 18: Colores por rango de edades.

Una vez hecho eso con las edades y con los otros dos parámetros que cambiamos a rangos con el nodo Numeric Binner, podemos visualizar por ejemplo el número de parejas sexuales que han tenido y la frecuencia de embarazo que han tenido en proporción.

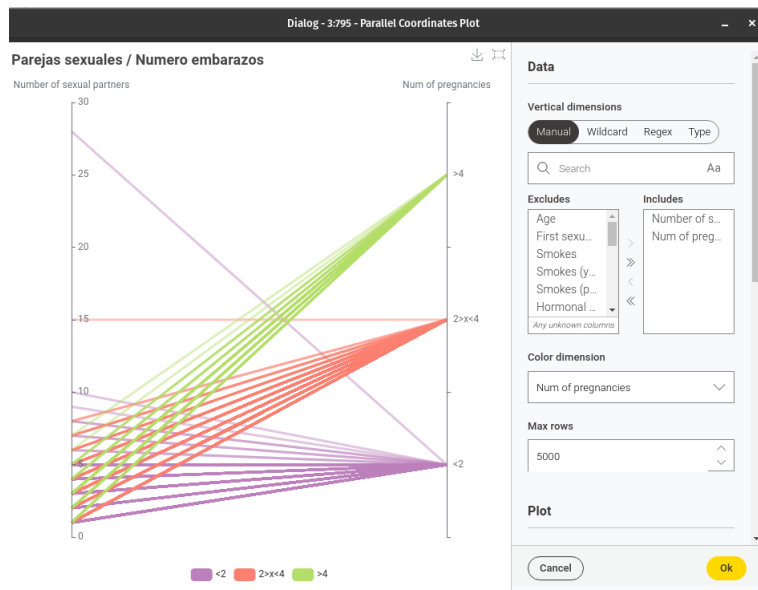


Figura 19: Parejas por embarazos.

Como se puede ver en la imagen, casi todos los casos que han tenido menos de 10 parejas sexuales ha tenido menos de dos embarazos y, a medida que se incrementa el número de parejas, la cantidad de embarazos aumenta, aunque suele estar bastante variado en este caso.

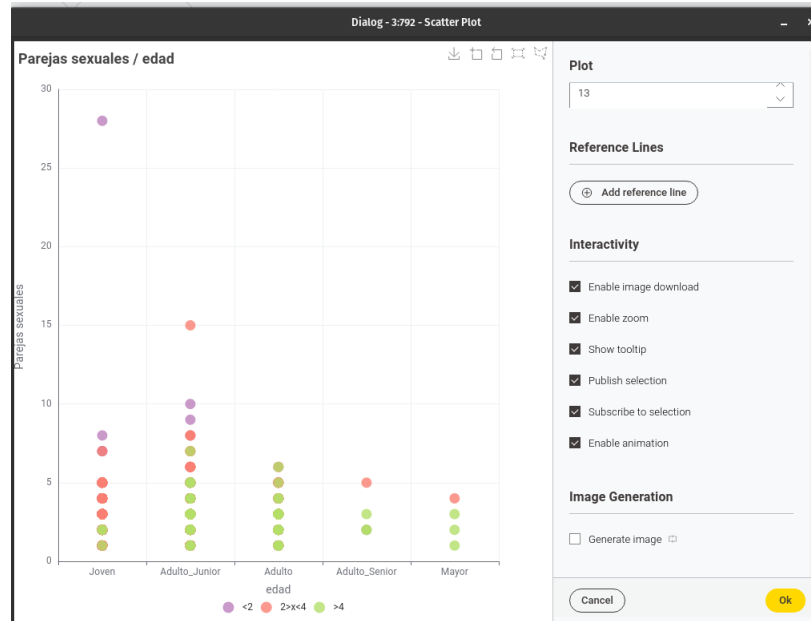


Figura 20: Parejas sexuales por edad.

En esta gráfica se ve que, a medida que la edad aumenta, el número de parejas sexuales es menor, mientras que, el máximo valor se tiene cuando la edad es de un Adulto Junior(entre 18 y 25 años)



Figura 21: Frecuencia diagnóstico respecto parejas sexuales.

En esta imagen se quiere hacer un mapa de calor con la frecuencia que han sido diagnosticadas o no, algún caso de cáncer, CIN o HPV. Como se puede ver, se concentra en casi toda la imagen cuando el número de parejas es entre 1 y 6 que no suelen haber sido diagnosticadas con ninguna de las 3 opciones, aunque también es verdad que, se concentran todos los casos de diagnósticos cuando se tienen ese rango de parejas sexuales.

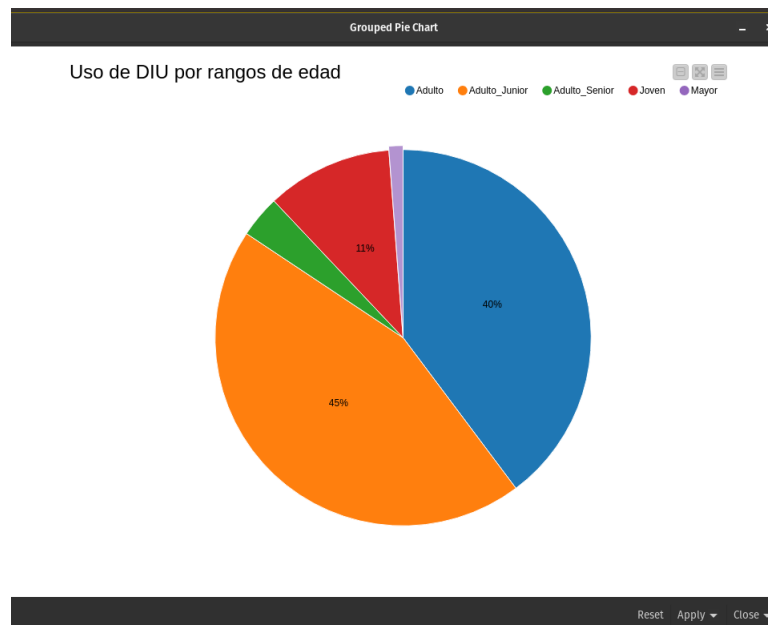


Figura 22: Frecuencia de edades que usan DIU.

En esta última gráfica se quiere mostrar la frecuencia en la que usan DIU las mujeres por edad, como se puede ver, casi todos los casos son en los rangos de Adulto Junior y Adulto, es decir, menos de 40 años son los casos que usan DIU y queda así un 15 por ciento que usa DIU siendo los otros 3 rangos de edades. Por lo que en casos de estudios posteriores, si se quiere usar la información del DIU, hay que tener en cuenta que el 85 por ciento de los datos de este, son en esos rangos de edades.

2. Descripción de la base de datos: gravedad en accidentes de tráfico

2.1. Gestión de Valores Desconocidos

Para el manejo de valores desconocidos en el archivo *.xls*, se deben dejar en blanco las celdas que contengan valores desconocidos, indicados generalmente con valores como "9." o "99". En las características que contienen valores perdidos, se incluyen dos variables: una que conserva los valores en bruto con los valores desconocidos y otra con valores imputados. El nombre de la segunda variable es igual al original (o su abreviatura) con la cadena ".I." al final. Nunca se deben utilizar ambas variables simultáneamente en ningún experimento.

Acción: Se ha utilizado Google Sheets para ir eliminando los valores de las columnas que tenían campos desconocidos como indica en la segunda página que tiene los valores de cada número, en este caso eran valores de 9, 99, 999, 9999 y 9998, por otro lado se han quitado también en aquellas columnas que precedían a las mismas con el nombre acabado en '.I.'.

2.2. Partición y entrenamiento de los datos

2.3. Construcción de la Variable Clase

La variable clase, que se utilizará para la clasificación, debe construirse como una combinación de las tres variables que describen la gravedad del accidente (*FATALITIES*, *INJURY CRASH* y *PRPTYDMG CRASH*). Cómo construir esta variable objetivo es a elección del alumno. Se pueden considerar diferentes enfoques, como considerar un accidente como grave solo si hay muertos, establecer niveles de gravedad, o definir condiciones específicas. Una vez construida la variable objetivo, se deben eliminar las columnas *FATALITIES*, *INJURY CRASH* y *PRPTYDMG CRASH*.

Acción: Para hacer esto se ha tenido en cuenta un criterio de los mencionados anteriormente, en este caso, si ha habido algún fallecido o no, es decir, con la columna 'Fatalities', desde google sheets establecemos la condición siguiente:

AK	AL	AM	AN
FATALITIES	INJURY_CRASH	PRPTYDMG_CRASH	
1	0	0	=SI(AK441;"Grave";"No grave")
0	1	0	No grave

Figura 23: Creación columna clase.

Una vez hecho esto, importamos el excel a Knime para eliminar las 3 columnas mencionadas anteriormente con el nodo 'Column Filter', en la siguiente imagen se puede ver en la parte inferior como ya está creada la columna clase que está establecida bajo el criterio anterior y cómo se han eliminado las tres columnas mencionadas con el nodo:

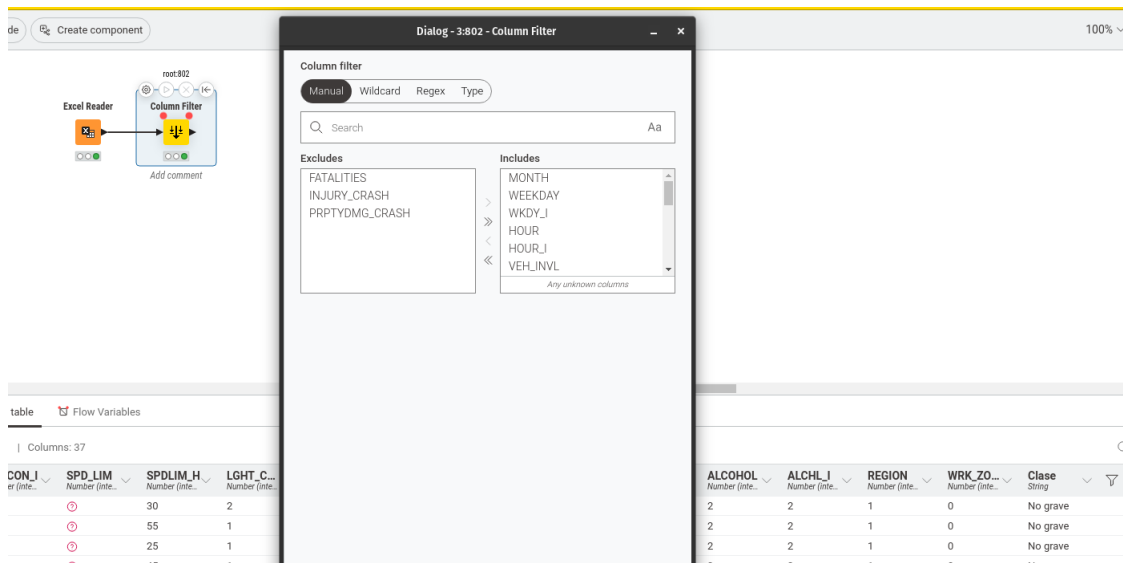


Figura 24: Borrado columnas.

2.4. Operaciones de Preprocesamiento

Para la preparación de datos en KNIME, se pueden realizar diversas operaciones por separado o en conjunto, con el objetivo de mejorar la clasificación. Es importante tomar decisiones basadas en resultados obtenidos y realizar un análisis exploratorio de datos. A continuación, se describen las operaciones de preprocesamiento:

2.4.1. Discretización

- a) Ejecutar el algoritmo de prueba para clasificación (C4.5) y estudiar cómo divide las características numéricas en los árboles de decisión aprendidos.

Usamos el nodo decision tree learner y usamos el algoritmo C4.5 con la columna creada previamente llamada clase, dividimos los datos en dos conjuntos, uno de 60 y otro de 40 para luego obtener un resultado o scorer:

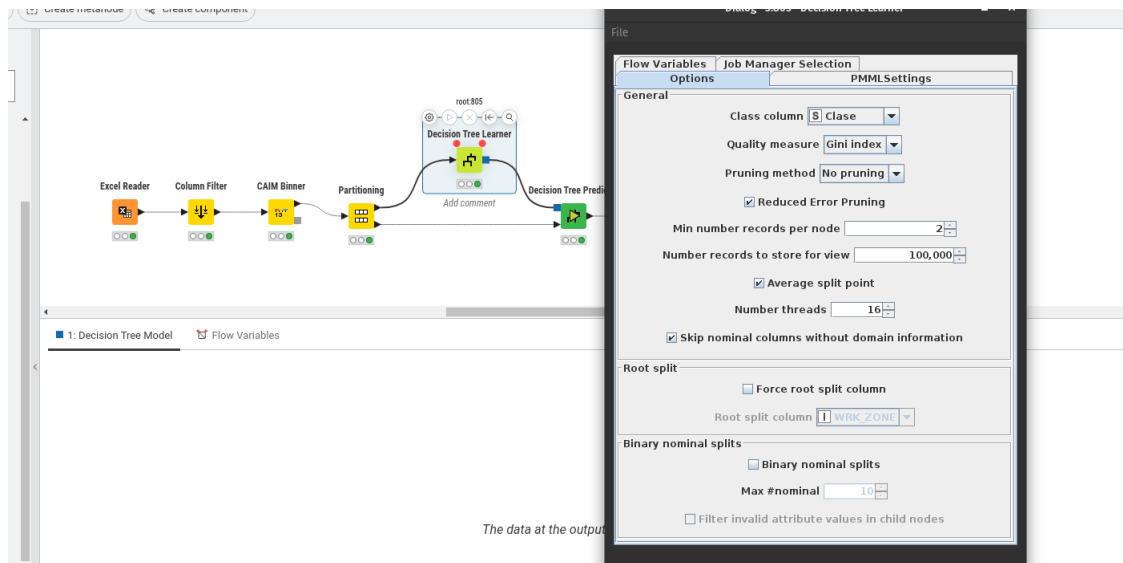


Figura 25: Configuración decision tree learner.

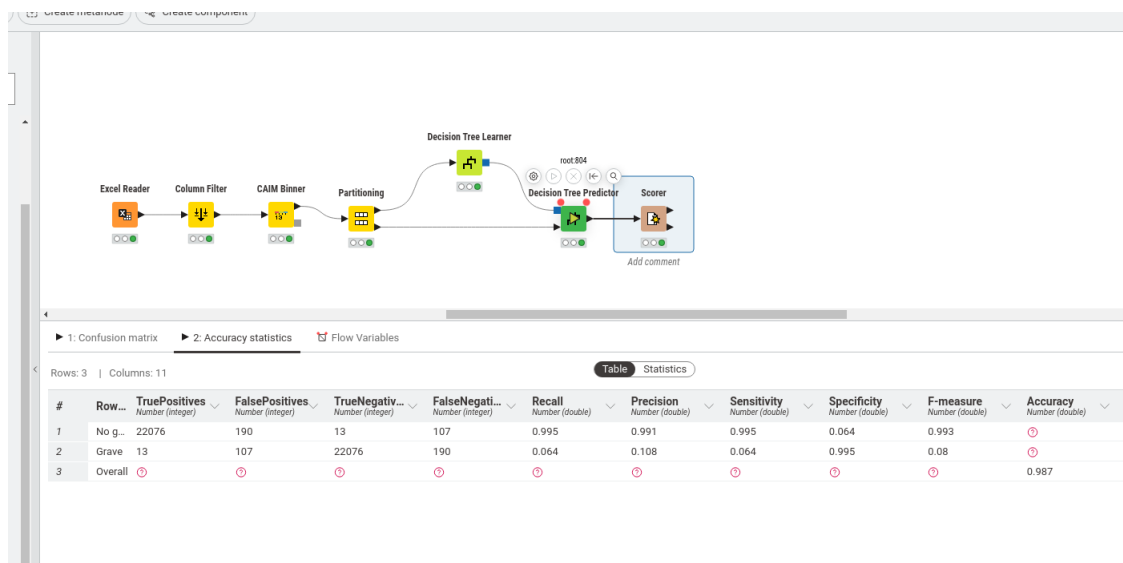


Figura 26: Scorer obtenido con la predicción de la columna clase.

- b) Aplicar el algoritmo de discretización top-down CAIM sobre las características numéricas y observar el comportamiento del algoritmo de prueba.

Se ha aplicado sobre las columnas que tienen establecidas al final el valor 'I'

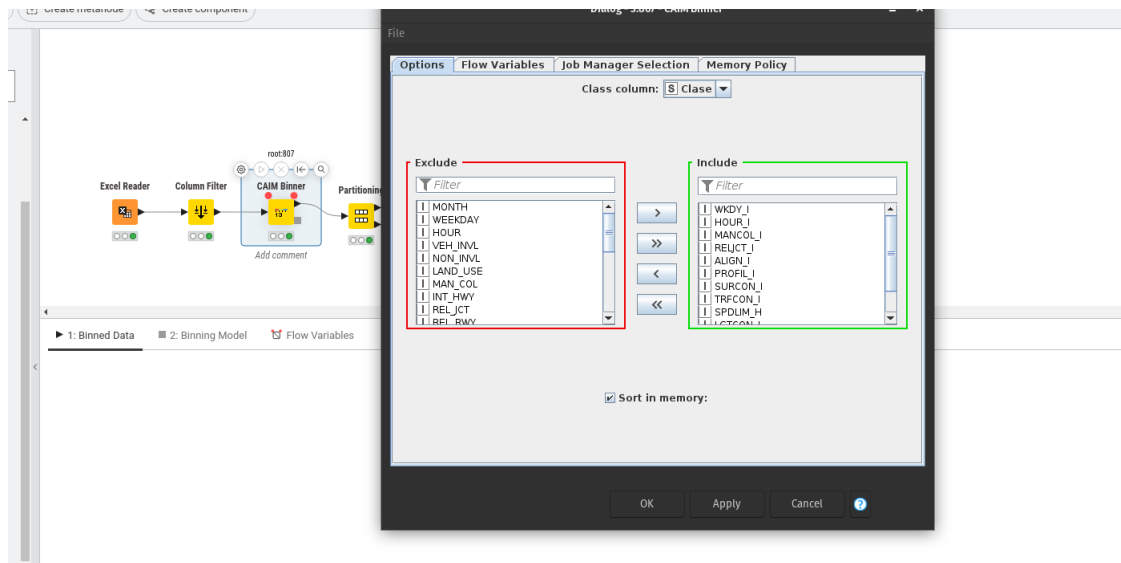


Figura 27: Nodo CAIM.

- c) Estudiar las características categóricas y proponer una discretización basada en el significado de la característica y la visualización de los datos. Realizar un estudio similar para las características numéricas.

A partir del uso del nodo CAIM, teniendo en cuenta que, la columna que hemos creado llamada clase está dividida en dos opciones, es un accidente grave o no, el nodo CAIM ha dividido todos los intervalos que le hemos pasado en dos opciones, intervalo 0 e intervalo 1, dejando los datos muy reducidos a esas dos opciones, si hacemos una visualización de los distintos datos podemos ver cómo los datos se concentran en ciertas horas de la noche, el número de accidentes se incrementa los fines de semana y lunes, y el tipo de accidente que ha ocurrido:



Figura 28: Ha ingerido alcohol o no respecto al mes y en función de si es un accidente o no.



Figura 29: Velocidades máximas respecto al mes y día de la semana.

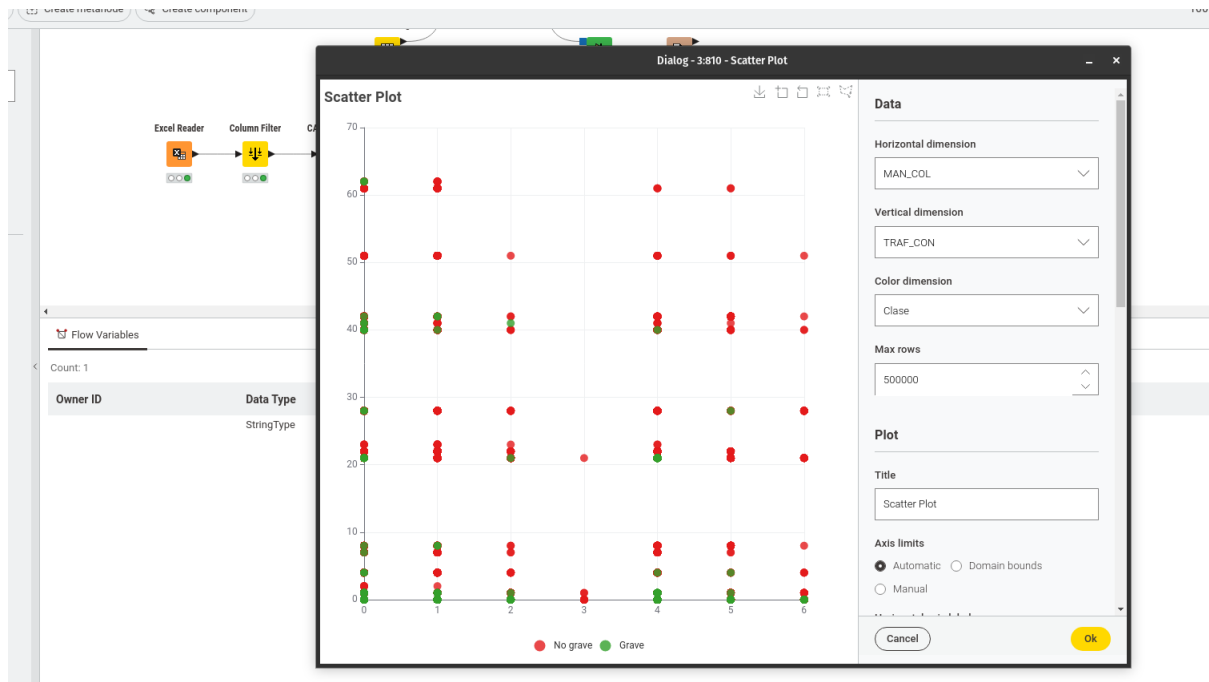


Figura 30: Tipo de accidentes respecto a las señales.