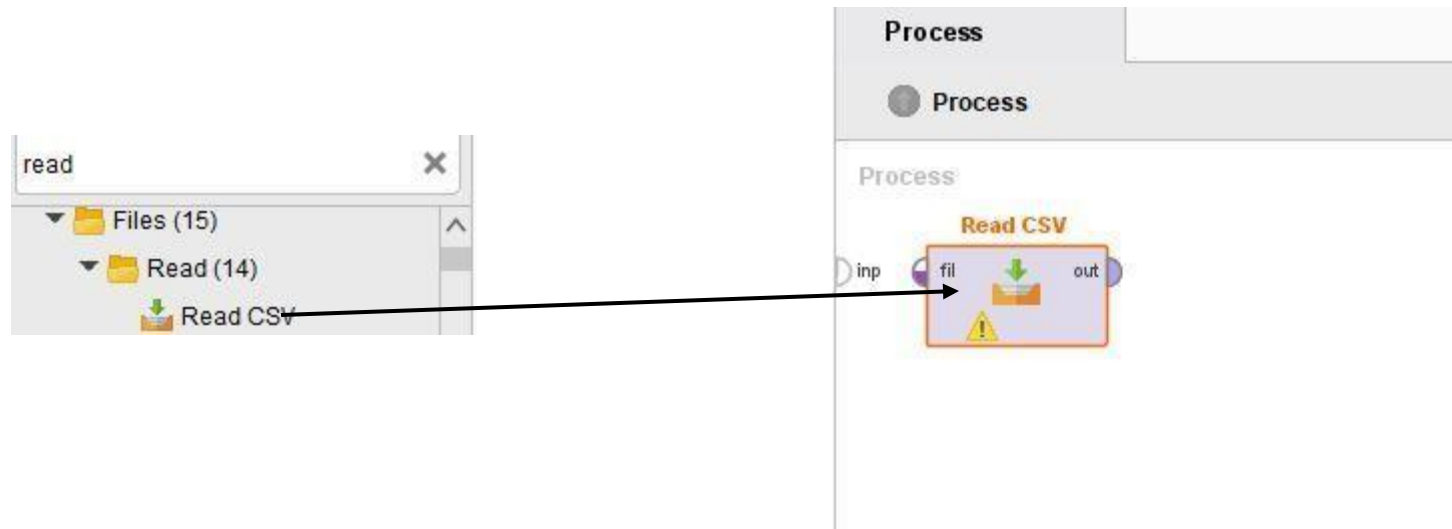


ANALISIS DE RESULTADOS-Guía 10


Para esta guía realizaremos el procedimiento de la guía anterior, a su vez utilizaremos el mismo archivo csv, para comparar las diferencias entre reglas de asociación y k-means, que es el procedimiento que se utilizará en esta guía.


Primero en el buscador de operadores escribimos read. Y seleccionamos el formato de archivo Excel .csv. Luego lo arrastramos a la plantilla y se vera de esta forma. Y nos pedirá que subamos un archivo del mismo al operador.




. Del lado derecho nos aparece la opción de importar la configuración del archivo .csv que vamos a utilizar. En este caso vamos a utilizar el de ferreteria.csv. Luego le damos Next.

Parameters ✕

 **Read CSV**

 **Import Configuration Wizard...** ⓘ

csv file  ⓘ

column separators ⓘ

☐ *trim lines* ⓘ

☒ *use quotes* ⓘ

quotes character ⓘ


escape character ⓘ


☐ *skip comments* ⓘ

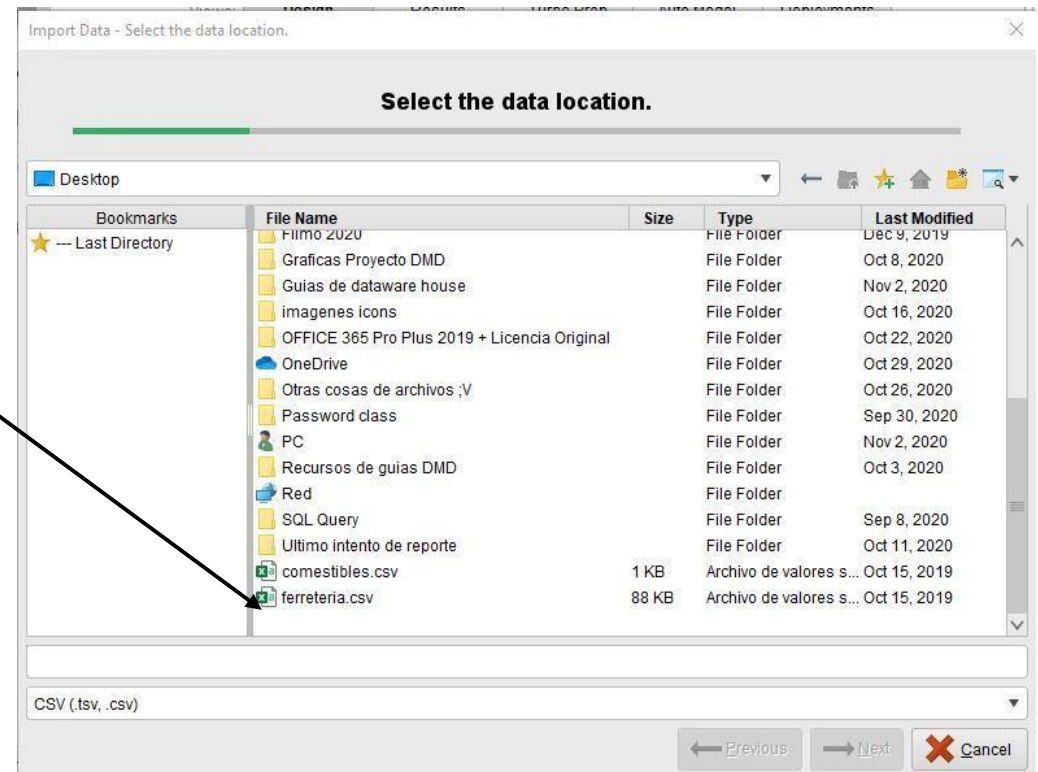
starting row ⓘ

☒ *parse numbers* ⓘ

decimal character ⓘ

 [Hide advanced parameters](#)

 [Change compatibility \(9.8.000\)](#)



. Luego tenemos que configurar algunos tipos de datos del formato, lo cual solo se enmarcara. Use quotes, trim lines, skip comments y por ultimo header row. En el caso aparece así porque se usó la versión 9.8 de rapidminer. Luego de eso nos aparecerá el formato de las columnas en este caso en esta versión ya esta predeterminado las opciones así que no hay problemas y solo le damos finish.

Import Data - Specify your data format

Specify your data format

☒ Header Row
Start Row
Column Separator
File Encoding
Escape Character
Decimal Character
☒ Use Quotes
☒ Trim Lines
☒ Skip Comments

	id	Pintura	Brocha	Rodillo	Bombilla	Destorni...	Madera	Abrazad...	Clavos	Ventila
1	id	Pintura	Brocha	Rodillo	Bombilla	Destorni...	Madera	Abrazad...	Clavos	Ventila
2	Charmia...	1	1	1	0	1	0	0	1	0
3	Deane V...	0	1	0	1	1	1	0	1	1
4	Willie Tong	0	1	1	0	1	1	0	0	1
5	Maire Br...	0	0	1	1	0	0	0	0	0
6	Augustin...	0	0	1	0	1	0	1	1	0
7	Frederik ...	0	0	0	1	1	1	0	0	0
8	Oralla B...	0	0	0	0	1	0	1	1	0
9	Chandle...	1	1	1	0	1	0	1	0	0
10	Ernesta ...	1	1	1	1	1	1	1	0	1

no problems.

Previous Next Cancel

Import Data - Format your columns:

Format your columns.

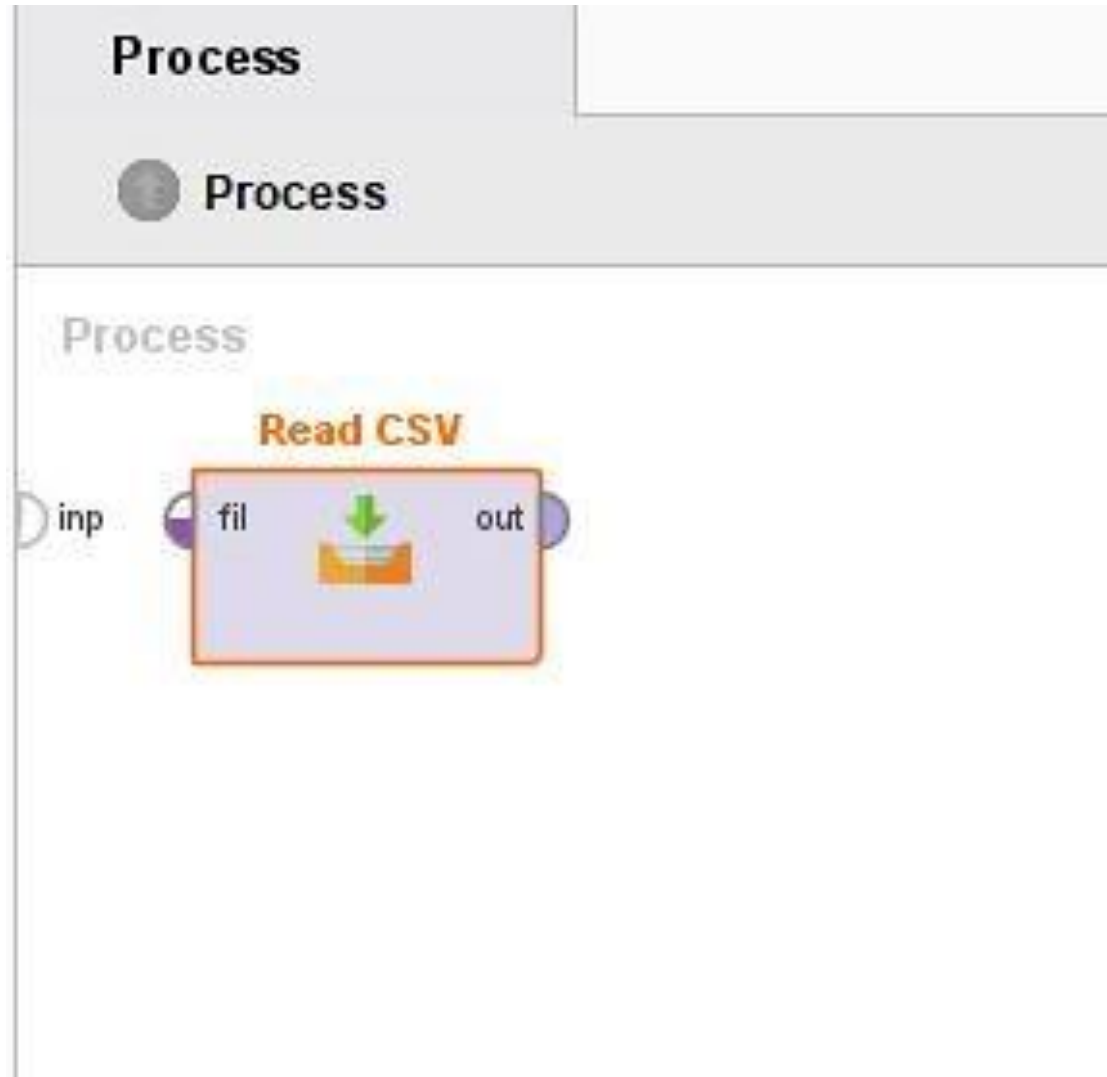
Date format ☐ Replace errors with missing values ⓘ

	id	Pintura	Brocha	Rodillo	Bombilla	Destornilla...
	polynomial	integer	integer	integer	integer	integer
1	Charmian Garritley	1	1	1	0	1
2	Deane Volk	0	1	0	1	1
3	Willie Tong	0	1	1	0	1
4	Maire Broomfield	0	0	1	1	0
5	Augustine Castel...	0	0	1	0	1
6	Frederik Janway	0	0	0	1	1
7	Oralla Bortoluzzi	0	0	0	0	1
8	Chandler Jemmi...	1	1	1	0	1
9	Ernesta Derricoat	1	1	1	1	1
10	Yvor Lavis	0	0	0	0	0
11	Ami Cantera	1	1	0	0	0

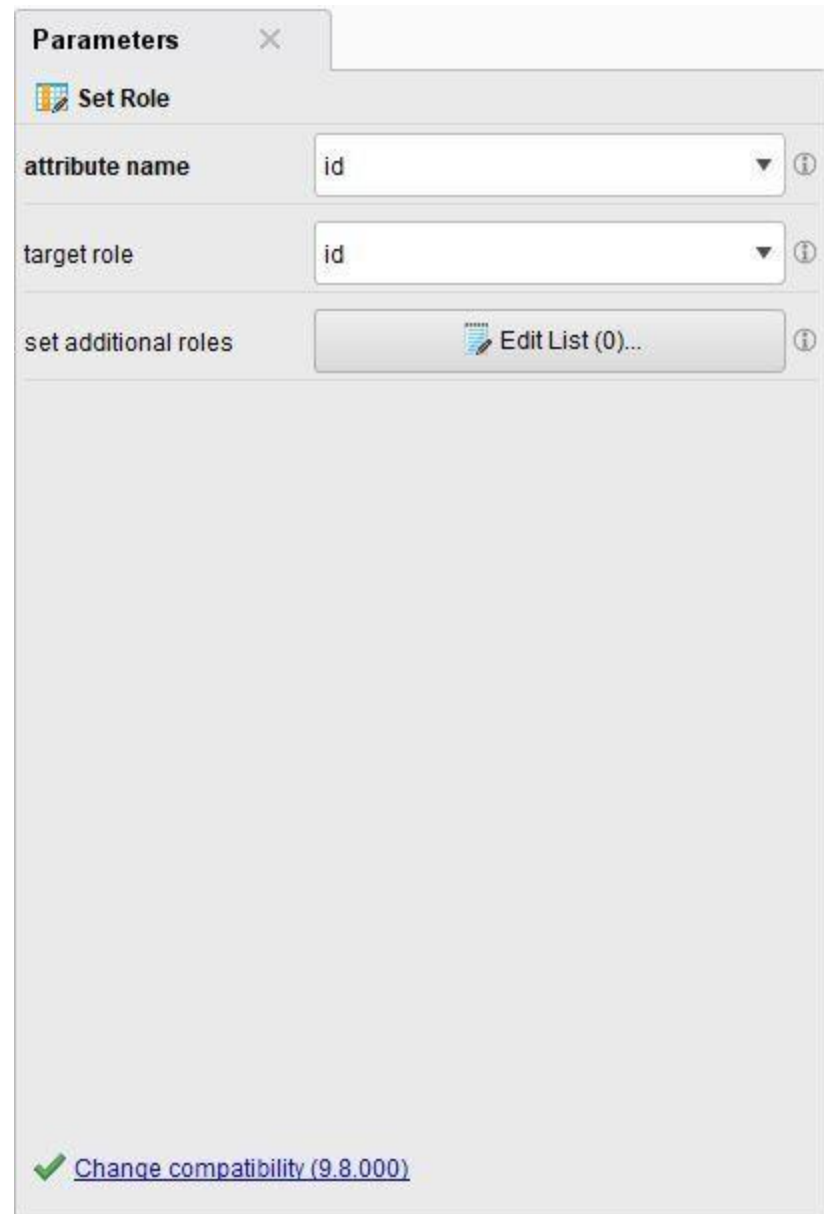
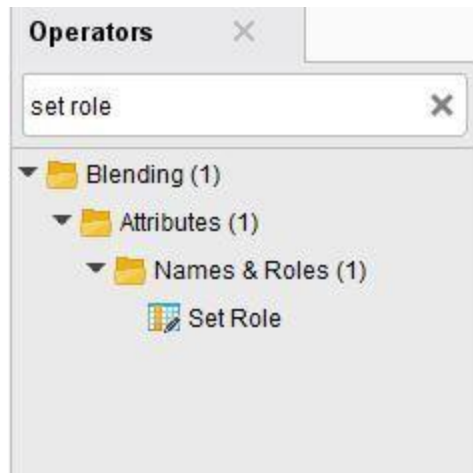
no problems.

Previous Finish Cancel

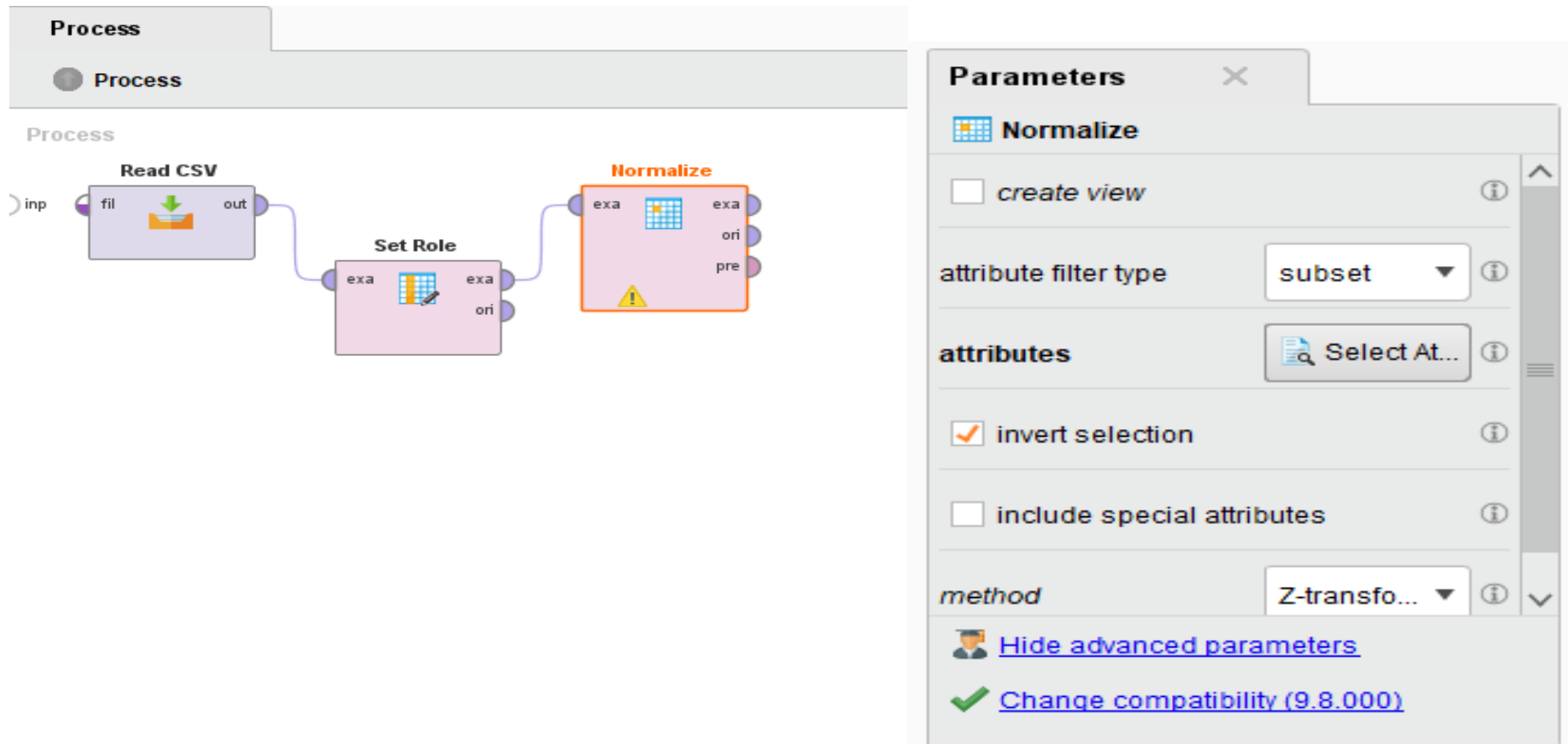
. Al final nos quedara el proceso del documento Excel ya añadido- Sin ningún problema.



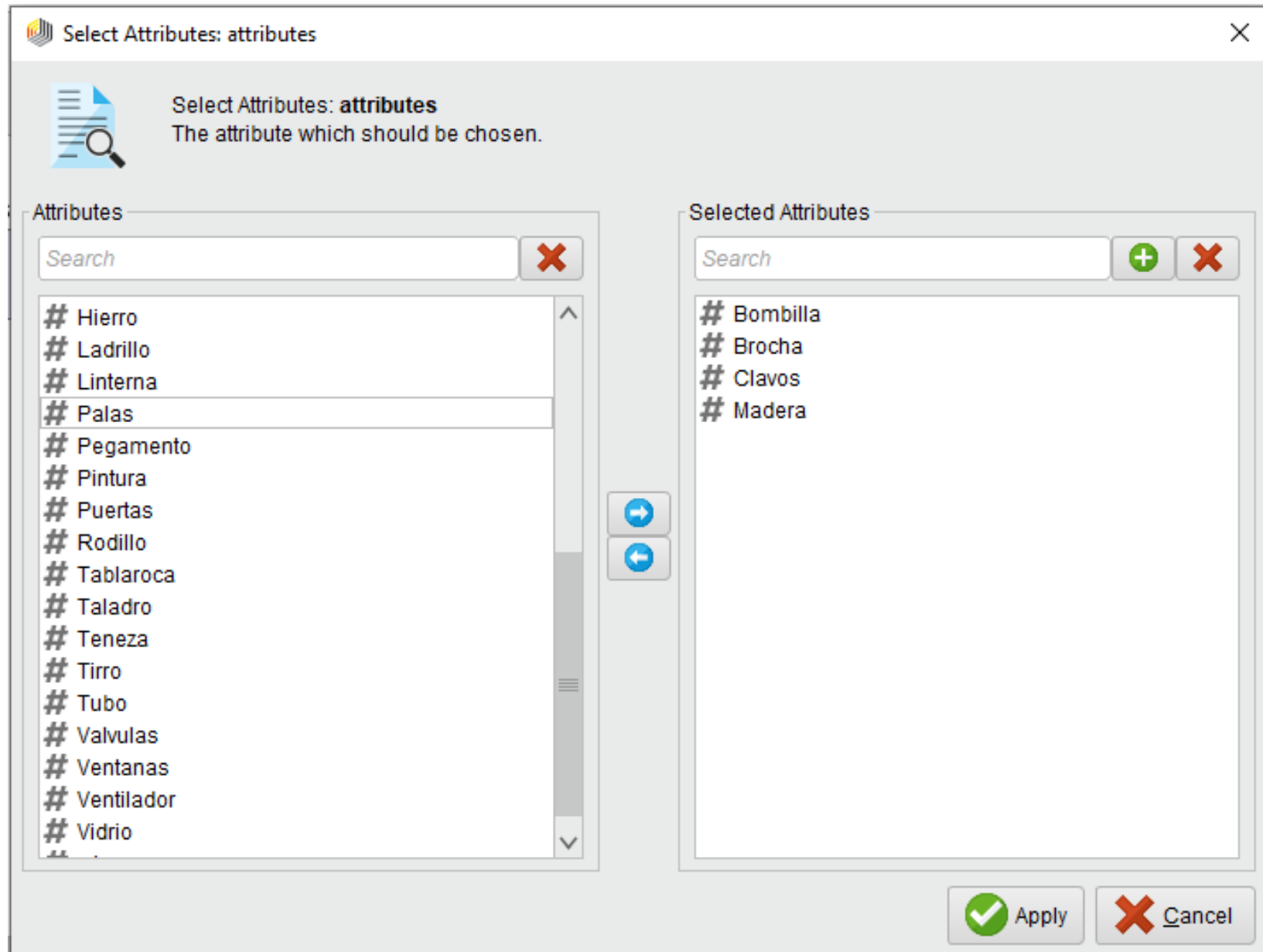
. Luego agregamos el segundo operador, en este caso el operador set role. Luego del lado derecho le configuramos algunos parámetros en este caso, sería que no tomara el id y en el apartado target role lo tome de esta manera.



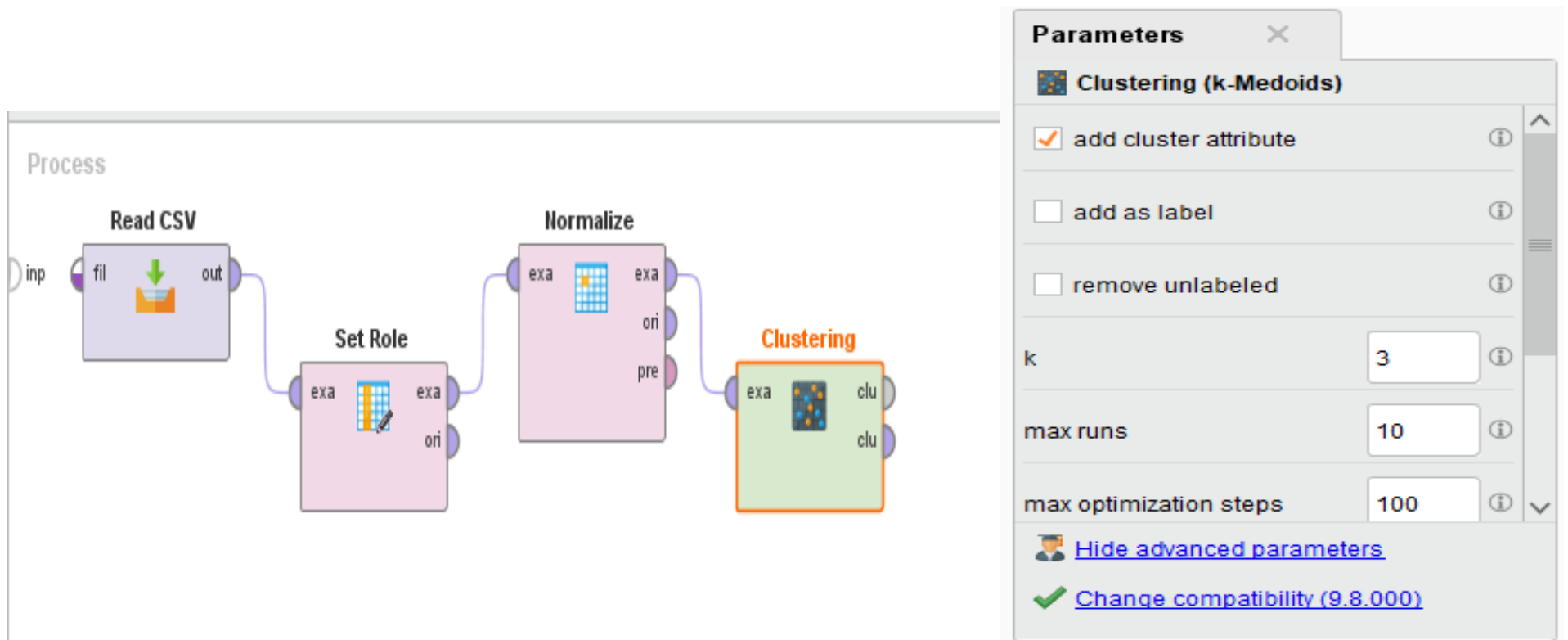
. Utilizamos el control Normalize para poder seleccionar los atributos necesarios para el análisis, para ello, primero en attribute filter type seleccionamos subset y marcamos invert selection



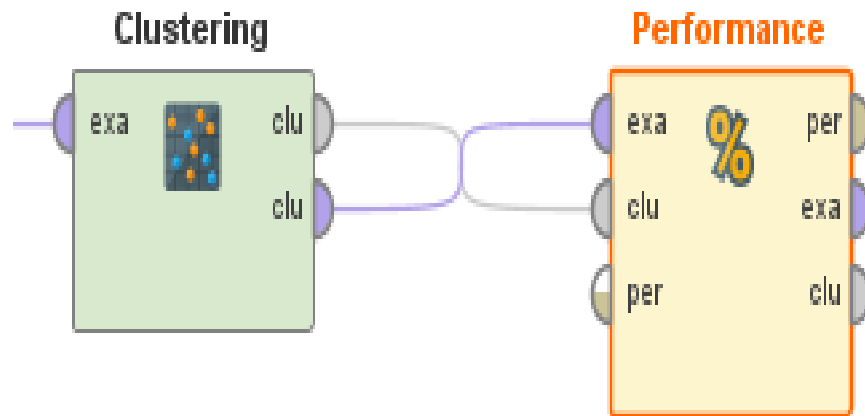
En la ventana Select Attributes: attributes seleccionamos los campos de preferencias para lograr el análisis objetivo, en este caso seleccionamos los datos de la columna bombilla, brocha, clavos y maderas.



Agregamos un nuevo control llamado clustering, que es básicamente la forma en la que nosotros agregamos los datos al modelos k-means.



Configuramos los criterios principales en Avg. Withing. Y por último proseguimos a ejecutar, el resultado se muestra en la siguiente página:



Parameters ✕

% Performance (Cluster Distance Performance)

main criterion Avg. within cent... ⓘ

☐ *main criterion only* ⓘ

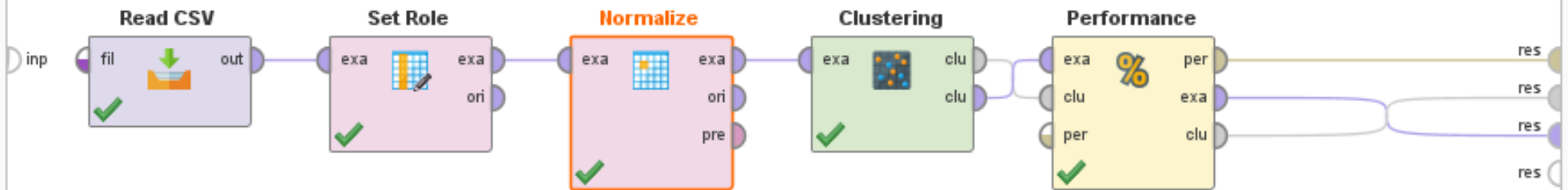
☐ *normalize* ⓘ

☐ *maximize* ⓘ

[Hide advanced parameters](#)

[Change compatibility \(9.8.000\)](#)

Process



Result History

ExampleSet (Clustering)

Cluster Model (Clustering)

PerformanceVector (Performance)

- Data
- Statistics
- Visualizations
- Annotations

Plot

Plot 1

Plot type

Step Area

X-Axis column

Pintura

Value columns

Brocha, Rodillo

Aggregate data

Color Group

-

Stacking

No stacking

Plot style

Add new plot

General

X-Axis

