UNIVERSIDAD DE CORDOBA

FACULTAD DE INGENIERIA

INGENIERIA DE SISTEMAS

TALLER DE CLUSTERING CON KNIME

ELECTIVA APRENDIZAJE AUTOMATICO

INTEGRANTES:

DEIMER YAMIT HERNANDEZ DE LA ROSA

DEYSER OROZCO YEPES

PROFESOR:

OSWALDO VELEZ LANGS

2020

**Desarrollo**

**1.Comprensión de los Datos**.

Identificar la estructura de los datos y su composición.

|  |  |  |
| --- | --- | --- |
| variable | descripción | tipo de dato |
| curso | curso | String |
| dia\_semana | dia\_semana | Number |
| solo\_hora | solo\_hora | Number |
| programa | programa | String |
| genero | genero | String |

**3. Compresión Datos:** Implementar un nodo “Statistics” y realizar una descripción Estadística de los datos.

**Atributos numéricos**

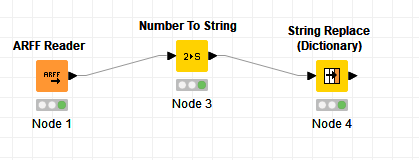
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| atributo | min | Mean | Max | Std. Dev. | varianza | histograma |
| dia\_semana | 1 | 3,5159 | 7 | 1,9282 | 3.71 |  |
| solo\_hora | 0 | 14,8368 | 23 | 5,2159 | 27.2 |  |

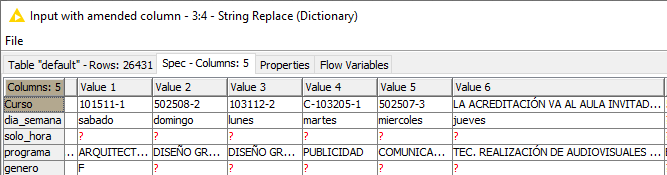
**Atributos nominales**

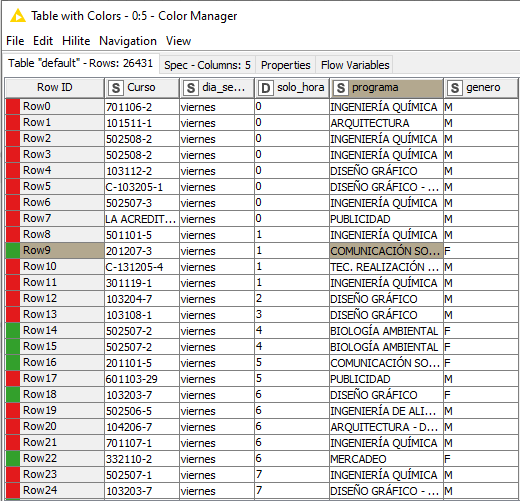
|  |  |
| --- | --- |
| Atributo | Histograma |
| curso |  |
| programa |  |
| genero |  |

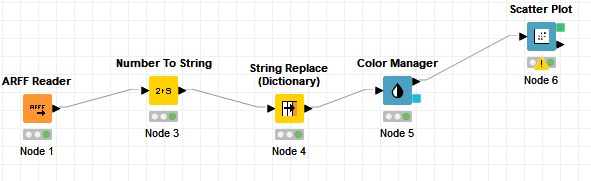
**4. Preparación de los datos:** Reemplazar los valores de la variable *Dia\_Seman,* por sus correspondientes nombres (etiquetas) de acuerdo al día de la semana. Implementar nodo “String replace (Dictionary)”. En la configuración del nodo para el reemplazo de palabras, utilizar el archivo “DictionaryDays.txt”

**R/** Para la preparación de los datos, se utilizó el nodo “Number To String”, para poder convertir el atributo de tipo Numérico *Dia\_semana* A tipo Nominal y de esta manera poder usar el Nodo “String Replace (Dictionary) ”



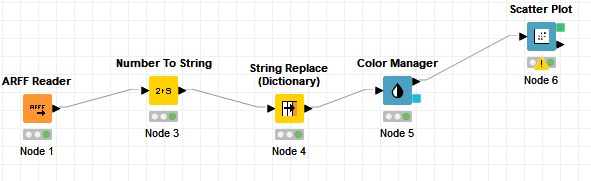


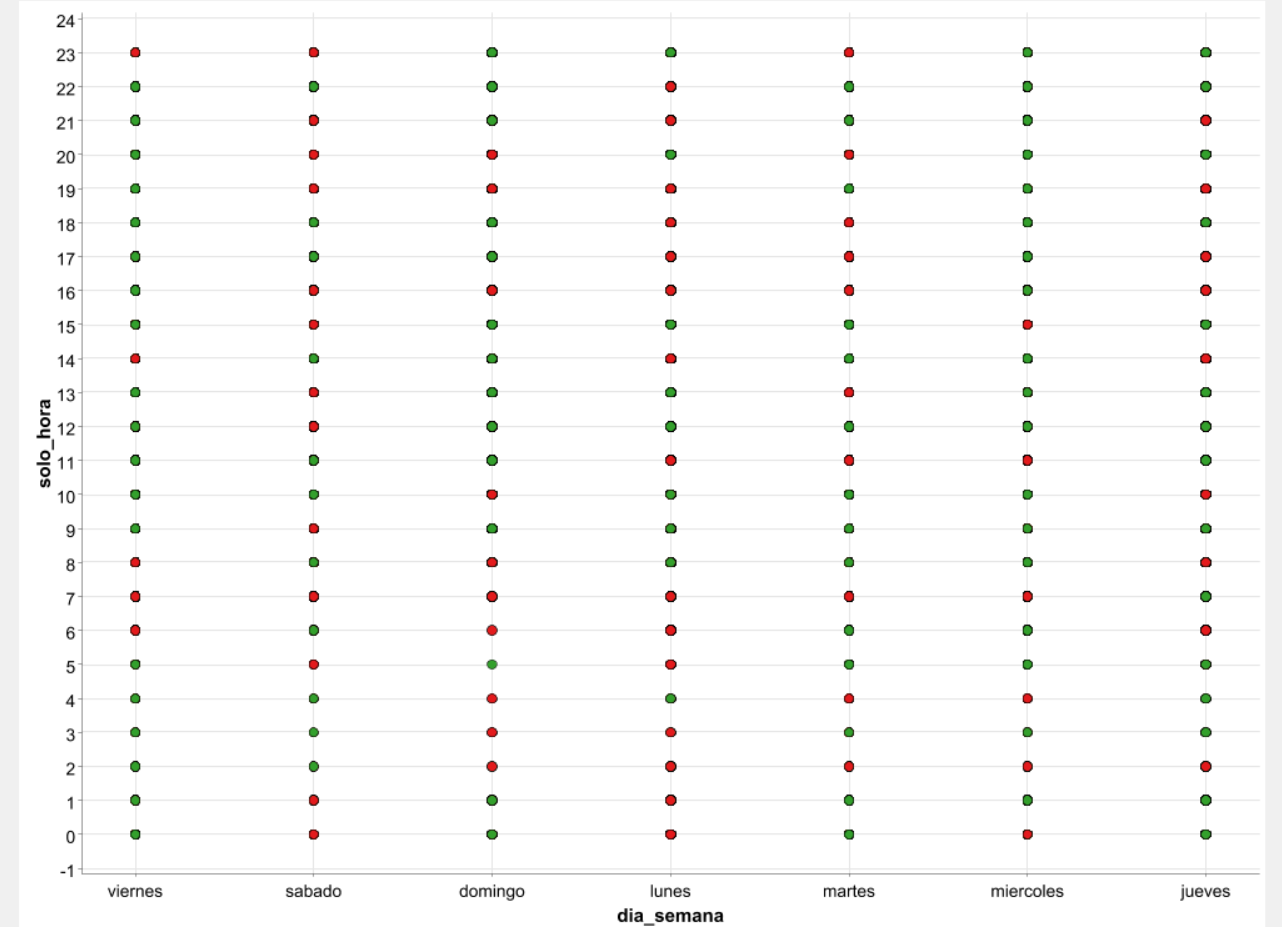
**5. Preparación de los datos.** Implementar nodo “Color Manager” para asignar “Etiquetas de color” a los registros de la base. 



**6. Análisis.** Implementar un nodo “Color manager” y conectarlo a un nodo “ScatterPlot” y reportar hallazgos relevantes en cuanto al comportamiento de los datos en su estado natural. Los datos salida del nodo “color manager” serán la entrada del nodo “Normalizer”.

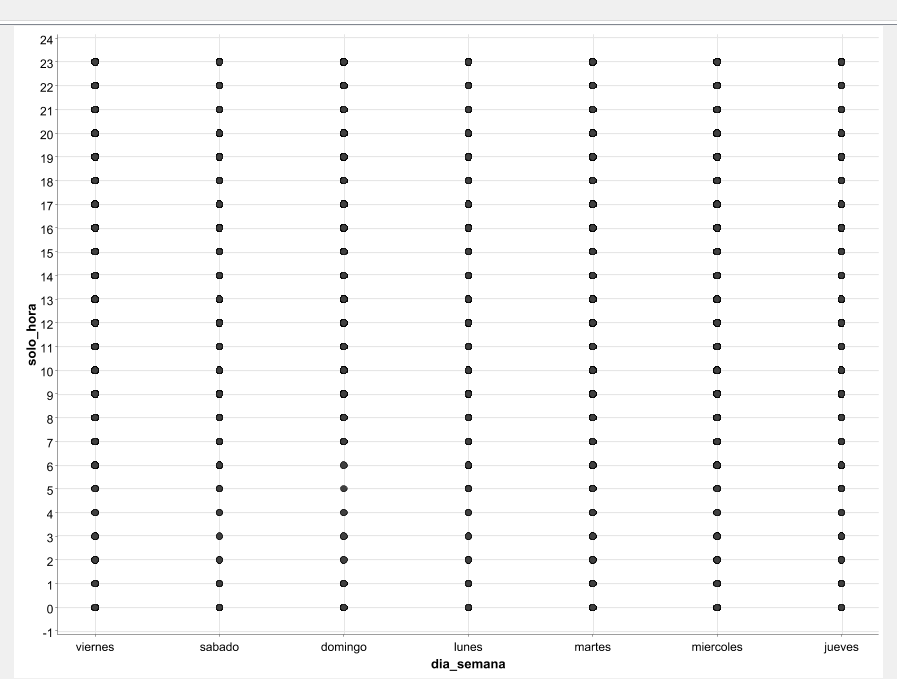
***Color manager aplicado a la columna “genero”***





Del análisis podemos concluir a simple vista que este nodo nos permite ver de manera clara datos relevantes en los datos de entrada, como puede ser la jerarquía de géneros que hay en un día en especifico a una hora en especifica. Como un ejemplo podemos ver que los sábados entre las 15hrs y las 22hrs la jerarquía de género es de Mujeres

***Metodo Scatter Plot desconectado del nodo Color Manager***



**7. Modelado.**

**7.1:** Los algoritmos *“EM y K-means”* No reciben datos nominales, por esta razón trabajamos solo con los datos de la columna “solo\_horas” que es la única después de normalizar y aplicar el nodo *“String Replace ( Dictionary )”*que sigue siendo numerica