**Realizado por:**

Juan Camilo Restrepo Velez  
William Leonardo Andrade Collazos  
Wilder Valencia Ocampo

**PRÁCTICA DE ANÁLISIS DESCRIPTIVO 10%**

Seleccionar una base de datos en <https://www.datos.gov.co/>y realizar un informe con todos los pasos de preparación de datos, utilizar pantallazos para documentar los resultados. Después de realizar todos los pasos, responder en el informe:

**Bank Marketing**

Los datos están relacionados con campañas de marketing directo de una institución bancaria portuguesa. Las campañas de marketing se basaron en llamadas telefónicas. A menudo, se requería más de un contacto con el mismo cliente, para poder acceder a si el producto (depósito bancario a plazo) sería ('sí') o no ('no') suscrito.

**Información de atributos**

Información Bancaria de los clientes

Age - Edad

Job - Trabajo: tipo de trabajo

Marital - Estado civil: estado civil

Education - Educación: Nivel educativo

Default - Incumplimiento: ¿tiene el crédito en mora?

Housing - Vivienda: ¿tiene un préstamo de vivienda?

Loan - Préstamo: ¿tiene préstamo personal?

Relacionado con la última llamada de la actual campaña

Contact - Contacto: tipo de comunicación

Month - Mes: último mes de contacto del año

DayofWeek - Día de la semana: último día de contacto de la semana

Duration - Duración: duración del último contacto, en segundos (numérico). Nota importante: este atributo afecta en gran medida al objetivo de salida (por ejemplo, si la duración = 0, entonces y = "no"). Sin embargo, no se conoce la duración antes de una llamada se realiza. Además, después del final de la llamada se conoce obviamente y. Por lo tanto, esta entrada sólo debe incluirse a efectos de referencia y debe descartarse si se pretende tener un modelo predictivo realista.

Otros

Campaign - Campaña: número de contactos realizados durante esta campaña y para este cliente

Pdays - pDías: número de días que pasaron después de que el cliente fue contactado por última vez en una campaña anterior. Nota, 999 significa que el cliente no fue contactado anteriormente

Previous - Anterior: número de contactos realizados antes de esta campaña y para este cliente

Poutcome: resultado de la anterior campaña de marketing

Atributos del contexto social y económico

Emp.var.rate - Tasa de variación del empleo - indicador trimestral

Cons.price.idx: Índice de Precios al Consumidor - Indicador mensual; el Índice de Precios al Consumidor o IPC mide los cambios en los precios pagados por los consumidores por una cesta de bienes y servicios cada mes.

Cons.conf.idx: Índice de confianza del consumidor - Indicador mensual; En Portugal, el índice de confianza del consumidor se basa en entrevistas con los consumidores sobre sus percepciones de la situación económica actual y futura del país y sus tendencias de compra. Se estima utilizando la diferencia entre la proporción de respuestas de evaluación positivas y las respuestas de evaluación negativas, pero no incluye la proporción de respuestas neutras

Euribor3m: euribor 3 meses - Euribor es la abreviatura de Euro Interbank Offered Rate. es un índice de referencia publicado diariamente que indica el tipo de interés promedio al que un gran número de bancos europeos dicen concederse préstamos a corto plazo entre ellos para prestárselo a terceros.

Nr.employed - Número de empleados: Número de empleados - Indicador trimestral; Número de personas empleadas para el trimestre.

y - ¿el cliente ha suscrito un depósito a plazo? (Variable objetivo)

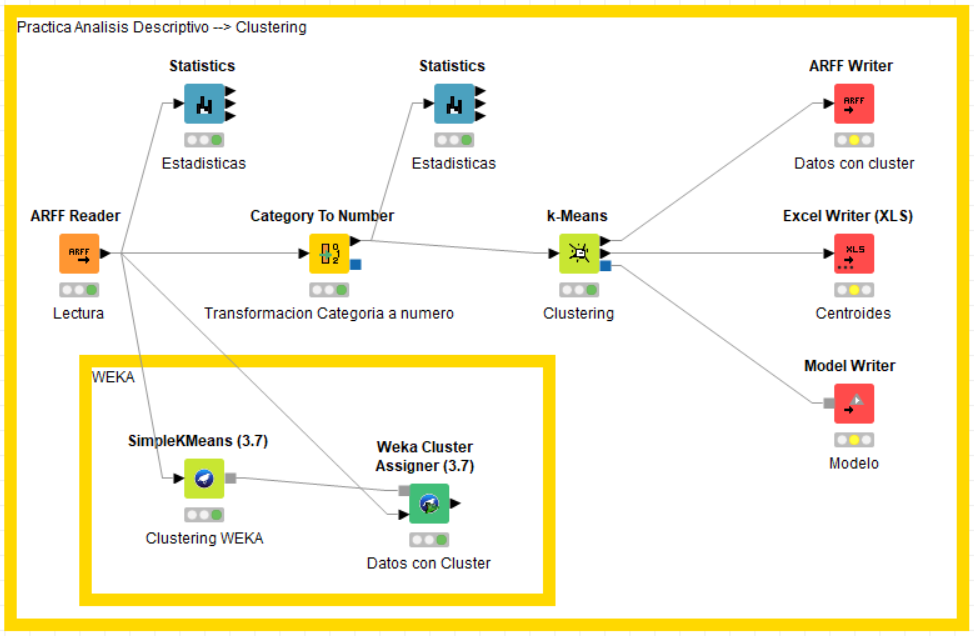
\*Tomado de <https://www.kaggle.com/henriqueyamahata/bank-marketing>

**PREPARACIÓN DE DATOS**

Ya se realizo debido a que se tomaron los datos utilizados en “Practica Análisis Predictivo”

**CLUSTERING**

**Aprendizaje- Modelo**



Nota: la conversión de categorías a número por parte de knime se hace a través de encoders (no de dummies que seria lo ideal para este caso) por lo cual el modelo arrojado por k-means de knime puede tener cierto grado de sesgo.

**Evaluacion**

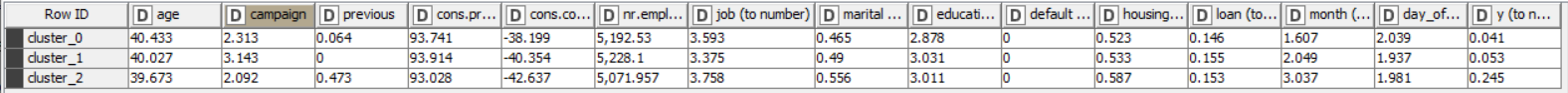
1. Cuánto es la cohesión?

Cohesión (weka): 143296.92463481252

**Perfilamiento**

2. Cuáles son los perfiles?

Knime



**Encoders**

Job: housemaid-0 services-1 admin.-2 Bluecollar-3 technician-4 retired- 5 management-6 unemployed-7 self-employed-8 entrepreneur-9 student-10

Marital: married-0 single-1 divorced-2

Educaction: basic.4y-0 high.school-1 basic.6y-2 basic.9y-3 professional.course-4 university.degree-5 illiterate-6

Default: no-0 si- 1

Housing: no-0 si-1

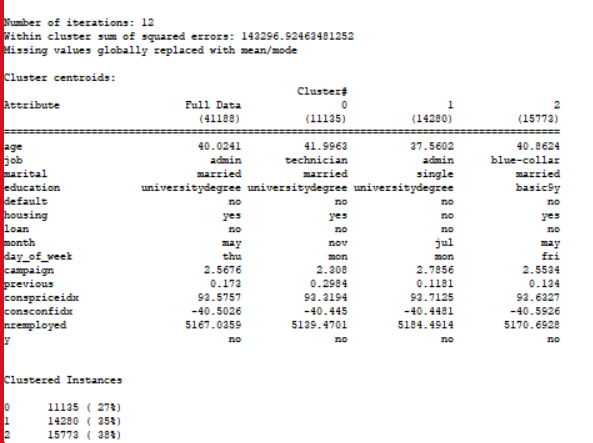
Loan: no-0 si-1

Month: may-0 jun-1 jul-2 aug-3 oct-4 nov-5 dec-6 mar-7 apr-8 sep-9

Day of week: mon-0 tue-1 wed-2 thu-3 fri-4

|  |  |  |
| --- | --- | --- |
| Clúster | Valores | Descripción |
| 0 |  | **Adultos casados sin vivienda propia:** Adultos graduados de bachillerato, que están casados, no tienen crédito de vivienda ni crédito personal, además no tienen créditos en mora. Fueron contactados por el banco en julio, un día martes, en la campaña pasada no fueron contactados y no se suscribió al depósito bancario a largo plazo. |
| 1 |  | **Adultos casados sin vivienda y bajo nivel escolar:** Adultos sin bachillerato, que están casados, no tienen crédito de vivienda ni crédito personal, además no tienen créditos en mora. Fueron contactados por el banco en julio, un día martes, en la campaña pasada no fueron contactados y no se suscribió al depósito bancario a largo plazo. |
| 2 |  | **Adultos Blue-Collar:**  Adultos trabajadores como Blue Collar y sin bachillerato, se encuentran casados, sin vivienda, no tienen crédito personal, además no tienen crédito en mora. Fueron contactados por el banco en agosto un día martes de la semana y no se suscribió a la campaña. |

Weka



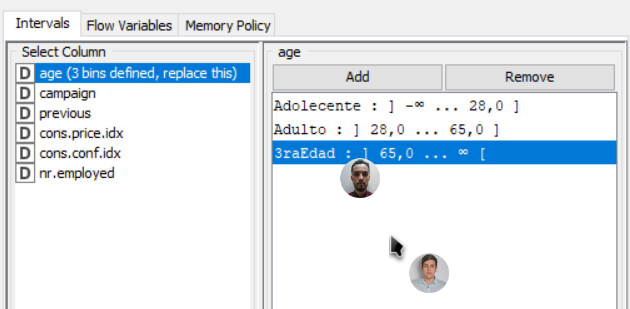
|  |  |
| --- | --- |
| Clúster | Descripción |
| 0 | **Adultos profesionales y casados:** Adultos técnicos graduados de la universidad que están casados, tienen crédito de vivienda, pero no crédito personal, además no tienen créditos en mora. Fueron contactados por el banco en noviembre, un día lunes, en la campaña pasada no fueron contactados y no se suscribió al depósito bancario a largo plazo. |
| 1 | **Adultos profesionales y solteros:** Adultos administrativos graduados de la universidad que están solteros, sin crédito de vivienda, no crédito personal y sin créditos en mora. Fueron contactados por el banco en el mes de julio un lunes, en la campaña pasada no fueron contactados y no se suscribieron al depósito bancario a largo plazo. |
| 2 | **Adultos bajo nivel de estudio y casados:** Adultos blue-collar sin graduarse del bachillerato, están casados, tienen crédito de vivienda, pero no crédito personal, además no tienen créditos en mora. Fueron contactados por el banco en el mes de mayo un día viernes, en la campaña pasada no fueron contactados y no se suscribieron al depósito bancario a largo plazo. |

**Nota**: Se analiza que para que los cluster no quedan tan diferenciados debido a la cantidad de iteraciones que se realizan, en este caso de 12, por lo que los centroides no se ajustan lo suficiente.

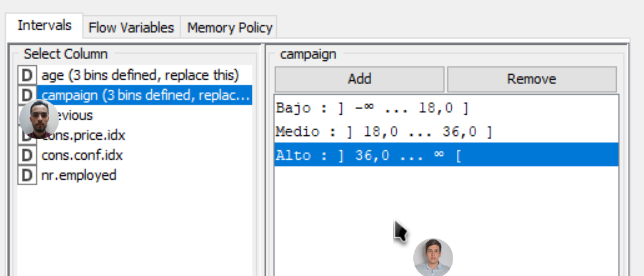
**REGLAS DE ASOCIACIÓN**

Conversión Números a categorías

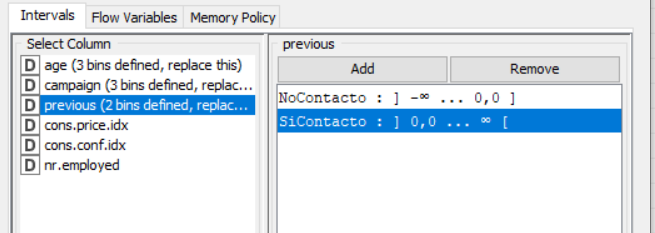
Age à 3 rangos que representan adolescencia(17-28], adultez(28,65] y tercera edad(65,98]



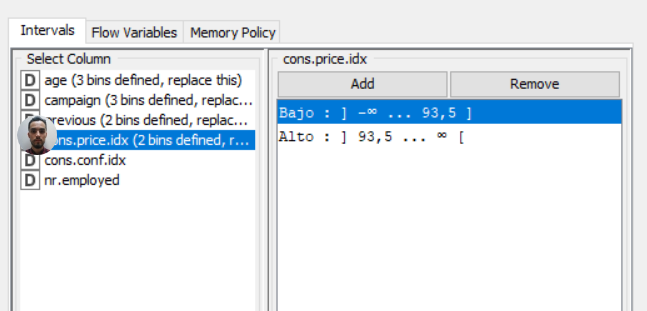
Campaign **à** 3 rangos que representan el grado de comunicación de la actual campaña



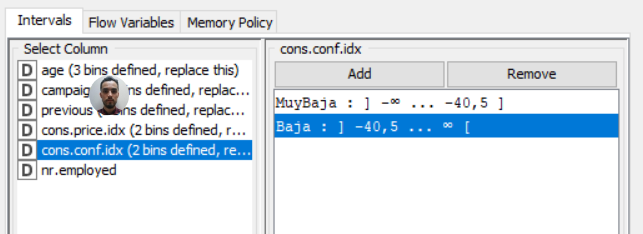
Previous à 2 rangos que representan la comunicación de la campaña anterior



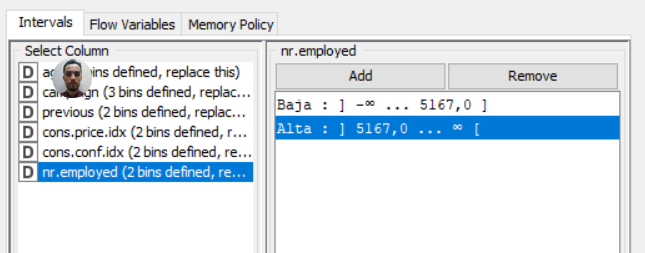
Cons.price.idx à 2 rangos que representan Alto IPC y bajo IPC



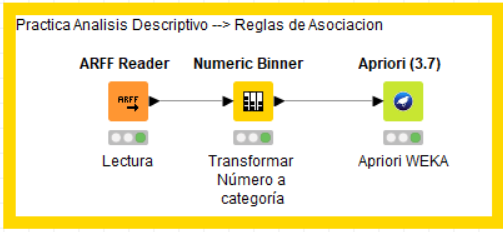
Cons.conf.idx à 2 rangos que representa la Baja confianza o Muy Baja confianza de los consumidores



Nr.employed à 2 rangos que representan Alta empleabilidad y baja Empleabilidad



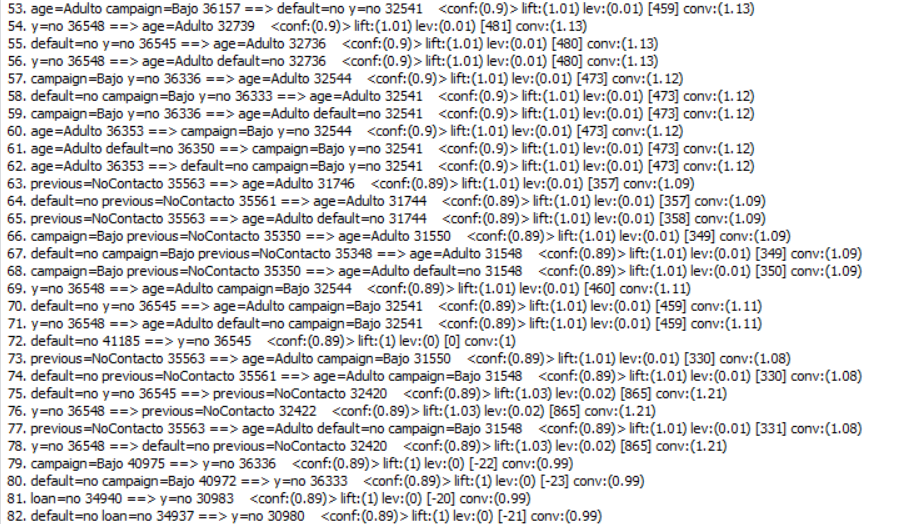
**Aprendizaje- Modelo**



**Evaluación & Interpretación**

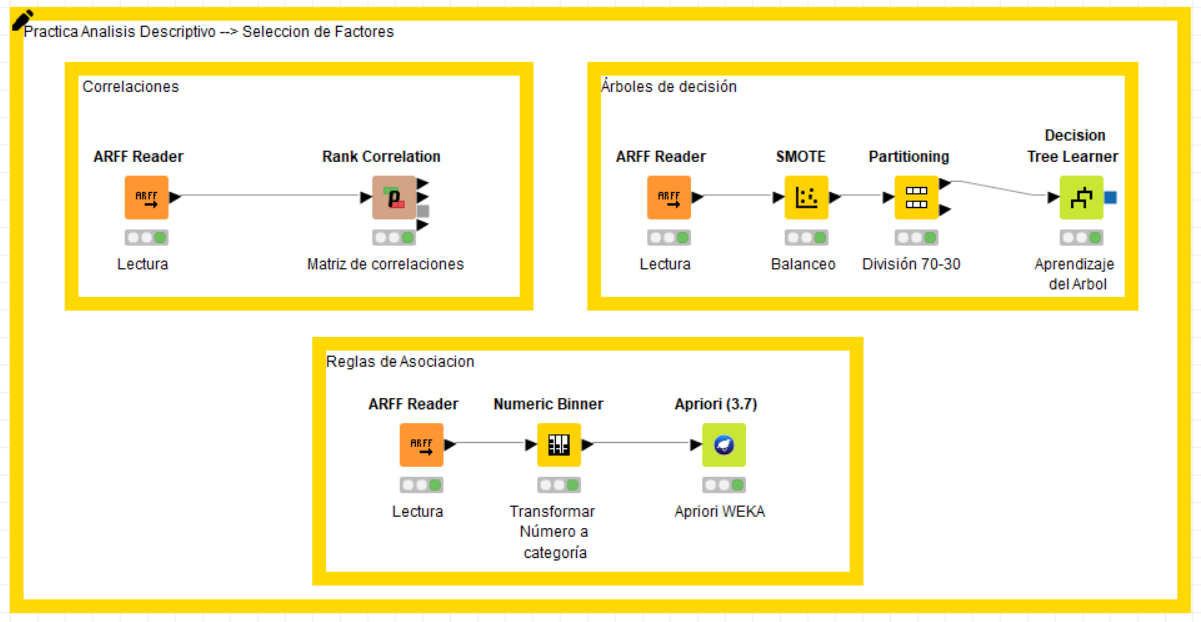
Se establece el soporte mínimo en: 0.3 y confianza mínima en: 0.5



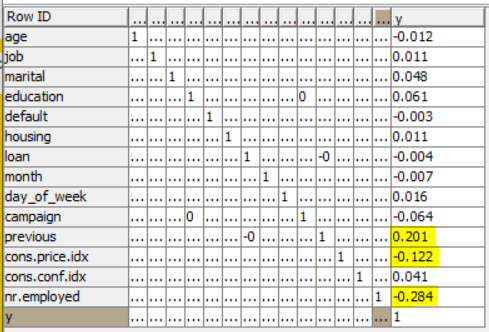


|  |  |  |
| --- | --- | --- |
| **Nro de regla** | **Descripción** | **Confianza** |
| 1 a la 16 | Los clientes adultos (age) que no tuvieron contacto en la campaña anterior (previous) y presentaron bajo contacto para la campaña presente (campaign), además de no tener préstamo personal (loan) y no han suscrito un depósito a plazo (y); son clientes que no tienen crédito en mora (default) | 1 |
| 39 a la 47 | Los clientes que no tuvieron contacto en la campaña anterior (previous), que no tienen crédito en mora (default) y presentaron bajo contacto para la campaña presente (campaign); son los clientes que no han suscrito un depósito a plazo (y) | 0.91 |
| 63 a la 71 | Los clientes que no tuvieron contacto en la campaña anterior (previous), que presentaron bajo contacto para la campaña presente (campaign), que no tienen crédito en mora (default) y no han suscrito un depósito a plazo (y); son clientes adultos (age) | 0.89 |

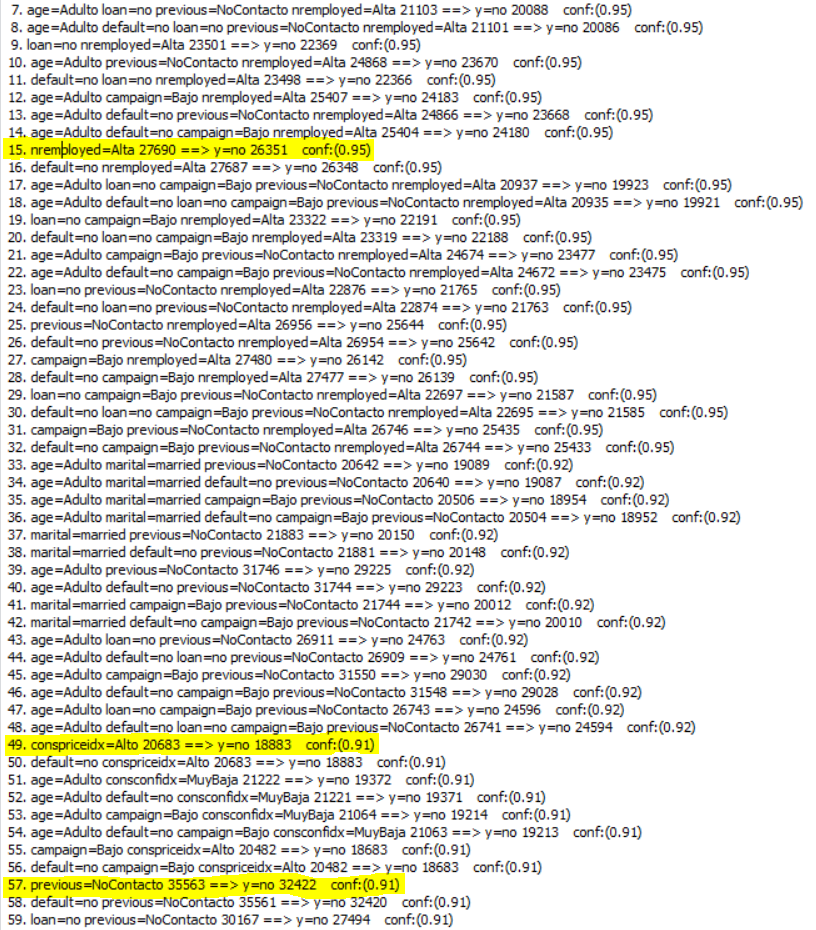
**SELECCIÓN DE FACTORES**



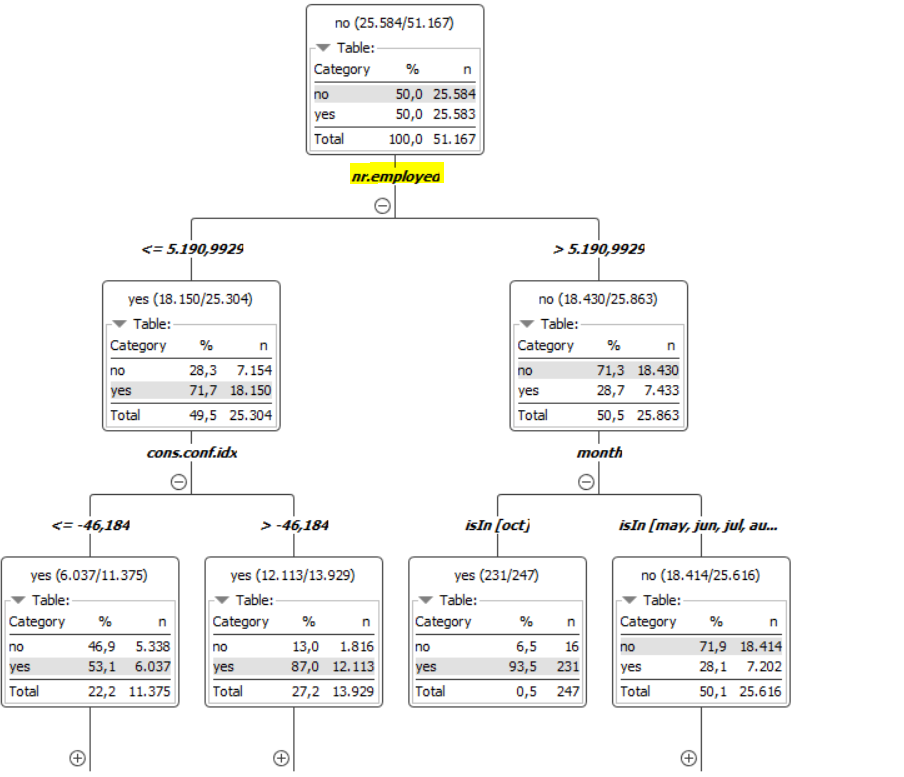
**Análisis de correlaciones**



**Reglas de asociación**



**Arboles de decisión**



4. Cuáles son las 3 variables seleccionadas?

1. Nr.employed
2. Previous
3. Cons.price.idx