

**GESTIÓN GLOBAL DEL RIESGO - SCORING 2023/24**

**\*\*\***

# **PRÁCTICA DE CREDIT SCORING UN CASO DE ESTUDIO DE TELCO**

**MAI ANH VÕ**

## Contenidos

|   |    |
|---|----|
| 1. INTRODUCCIÓN.....  | 2  |
| 1.1 Contexto.....   | 2  |
| 1.2. Objetivos .....  | 2  |
| 1.3. Recursos .....   | 2  |
| 1.4. Fases principales del trabajo .....                        | 2  |
| 2. EDA.....   | 4  |
| 2.1. TrainCS.....   | 4  |
| 2.2. CS_Rejected .....  | 9  |
| 2.3. Scoring_CS.....  | 10 |
| 3. TRANSFORMACIÓN DE DATOS .....                                | 10 |
| 4. SELECCIÓN DE DATOS.....                                      | 11 |
| 5. ESTIMACIÓN DEL PRIMER MODELO Y EL SCORECARD PRELIMINAR ..... | 12 |
| 6. INFERENCIA DE DENEGADOS .....                                | 16 |
| 7. ESTIMACIÓN DEL MODELO Y EL SCORECARD DEFINITIVO .....        | 17 |
| 8. VALIDACIÓN DEL SCORECARD DEFINITIVO Y CONCLUSIÓN. ....       | 20 |

## 1. INTRODUCCIÓN

### 1.1 Contexto

Tras 5 años de irrupción en el mercado español la telco “MovilTV” dispone de una importante cartera conformada por casi un millón de clientes. El plan es seguir creciendo la base de clientes en España pero incrementando la rentabilidad de las nuevas altas, Para mantener el ritmo de crecimiento es necesario no perjudicar las condiciones contractuales que se ofrecen a los nuevas altas, por tanto sólo queda optimizar los procesos de negocio.

Entre los procesos de negocio destaca principalmente el que se ocupa de prever el comportamiento de pago de los clientes, ya que un cliente que te impaga exige a la compañía el afrontar un importante monto económico, MovilTV calcula que cada impago equivale al valor que generan 20 contrataciones nuevas.

Durante estos cinco años la política de movilTV de riesgos se ha fundamentado en la aplicación de reglas duras de parada en base a la experiencia del departamento de riesgo.

### 1.2. Objetivos

En esta tarea, desarrollamos una estrategia predictiva que permita realizar ofertas únicamente a los clientes que tengan una alta propensión a pagar, por lo cual, el documento se trata de cumplir 3 objetivos principales:

- Traducir el modelo a una tarjeta de puntuación para poder integrarlo en la fuerza de ventas.
- Presentar la estrategia seguida y los resultados esperados, junto con los programas utilizados.
- Realizar una prueba ciego, puntuando un set de datos para decidir a quién otorgaríamos crédito antes de integrar la tarjeta de puntuación en producción la empresa.

Para el desarrollo de esta tarea vamos a utilizar el software SAS Enterprise Miner, utilizando la metodología de análisis SEMMA.

### 1.3. Recursos

El archivo de datos proporcionado por MovilTV se compone de 3 conjuntos de datos:

- **TrainCS:** Datos de líneas que se conoce su comportamiento de pago.
- **CS\_Rejected:** Base de datos que cuenta con los usuarios a los que el departamento de riesgos no se les aprobó la entrada a la compañía.
- **Scoring\_CS:** Datos ciegos que tienes que puntuar y seleccionar que observaciones son las seleccionadas.

Aparte de los datos, disponemos de un “Diccionario de variables” que explica el significado de las variables contenidas en el conjunto de datos.

### 1.4. Fases principales del trabajo

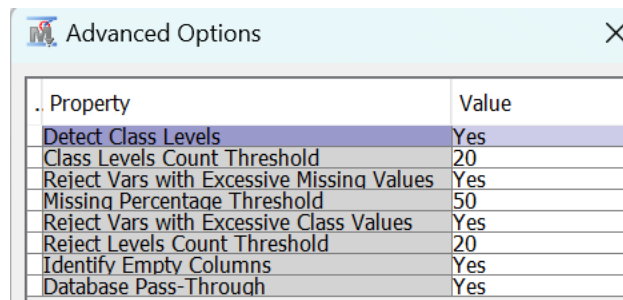
La finalidad de este trabajo es crear una tarjeta de puntuación para seleccionar los clientes que tengan una alta propensión a pagar, por lo cual, el trabajo se compone de los siguientes apartados principales:

- **Análisis descriptivo exploratorio de los datos:** Realizar una exploración formal y gráfica para los 3 conjuntos de datos.
- **Transformación de datos:** Imputar los ausentes, tratar los atípicos.

- Selección de datos: Hacer el muestreo, tramificar los datos y seleccionar las variables importantes para construir el scorecard.
- Estimación del primer modelo y obtención del scorecard preliminar.
- Inferencia de denegados.
- Estimación del modelo definitivo y obtención del scorecard definitivo.
- Validación del scorecard definitivo y conclusión.

## 1. EDA

En el primer lugar, importamos los datos al SAS Miner y establecer correctamente los roles utilizando el asesor avanzado para facilitarnos el trabajo.



| Property                                  | Value |
|---|-------|
| Detect Class Levels                       | Yes   |
| Class Levels Count Threshold              | 20    |
| Reject Vars with Excessive Missing Values | Yes   |
| Missing Percentage Threshold              | 50    |
| Reject Vars with Excessive Class Values   | Yes   |
| Reject Levels Count Threshold             | 20    |
| Identify Empty Columns                    | Yes   |
| Database Pass-Through                     | Yes   |

Vamos a realizar un análisis exploratorio de los tres ficheros de datos con el objetivo de entender mejor las variables de las que disponemos, y realizar una depuración de los mismos, si fuera necesario (detección de atípicos, ausentes, selección de variables, transformación, etc.) en el punto que se desarrolla a continuación.

### 2.1. TrainCS

El fichero TRAINS contiene los datos de líneas que se conoce su comportamiento de pago con 100,061 observaciones, 73 variables de tipo categórica, fecha y numérica. De las cuales, la variable objetivo es Target\_i\_ECInvol\_12m que muestra la decisión de early churn involuntario en 12 meses.

Al importar el fichero “TrainCS” eliminamos 2 variables unarias: “**Doc\_i\_PersonaJuridica**”, “**CB\_i\_EntTelefonialIntegrada**”, y la variable categórica “**PS\_Cod\_CP**” que tiene el más que 20 niveles (125 niveles en total). Al final quedamos con 70 variables con los roles de datos asignados automáticamente como se muestra en la pantalla de resumen.

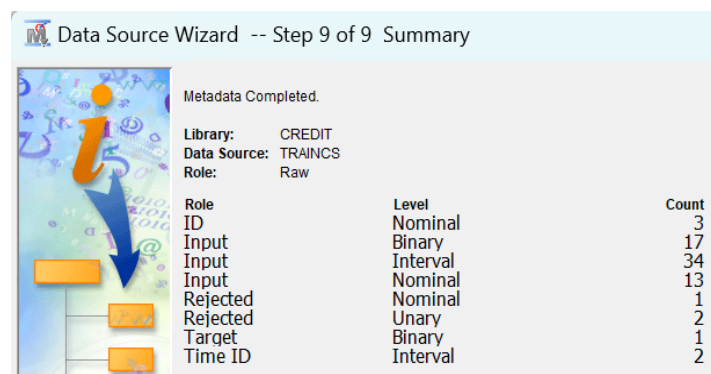
| Obs # | Variable Name         | Label                         | Type | Perc... | Minim... | Maxi... | Mean |
|-------|-----------------------|-------------------------------|------|---------|----------|---------|------|
| 1     | Doc i PersonaJuridica | Indicador de Persona Jurídica | VAR  | 0       | 0        | 0       | 0    |

| Obs # | Variable Name ▲             | Label                                   | Type | Perc... | Minim... | Maxi... | Mean |
|-------|-----------------------------|---|------|---------|----------|---------|------|
| 1     | CB i EntTelefonialIntegrada | Indicador Cedente Ent. Telef. Integrada | VAR  | 2.59    | 0        | 0       | 0    |

| Obs # | Variable Name | Label ▲              | Type  | Percent Missing | Number of Levels | Mode Percentage | Mode      |
|-------|---------------|----------------------|-------|-----------------|------------------|-----------------|-----------|
| 1     | PS Cod CP     | CP Punto de Servicio | CLASS |                 | 0125             |                 | 4.0346022 |



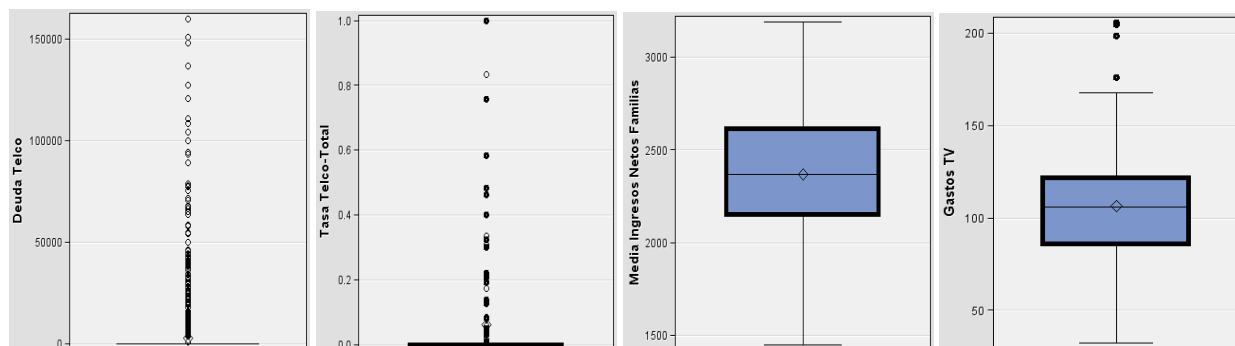
| Role     | Level    | Count |
|----------|----------|-------|
| ID       | Nominal  | 3     |
| Input    | Binary   | 17    |
| Input    | Interval | 34    |
| Input    | Nominal  | 13    |
| Rejected | Nominal  | 1     |
| Rejected | Unary    | 2     |
| Target   | Binary   | 1     |
| Time ID  | Interval | 2     |

Utilizamos el nodo StatExplore in SAS Miner para sacar la tabla los estadísticos descriptivos para las variables numéricas y categóricas y obtenemos los siguientes resultados:

- Exploración de variables numéricas:

| Interval Variable Summary Statistics<br>(maximum 500 observations printed) |       |          |                       |                |         |          |          |          |          |          |
|--|-------|----------|-----------------------|----------------|---------|----------|----------|----------|----------|----------|
| Data Role=TRAIN  |       |          |                       |                |         |          |          |          |          |          |
| Variable   | Role  | Mean     | Standard<br>Deviation | Non<br>Missing | Missing | Minimum  | Median   | Maximum  | Skewness | Kurtosis |
| CB_Int_DeudaDescubiertoCC  | INPUT | 1033.59  | 38668.8               | 97174          | 2826    | 0        | 0        | 9277854  | 165.0824 | 35854.41 |
| CB_Int_DeudaMedia  | INPUT | 35407.46 | 709549.3              | 97174          | 2826    | 0        | 0        | 1.5566E8 | 131.6433 | 25589.98 |
| CB_Int_DeudaNoTelco  | INPUT | 89303.56 | 1653681               | 97174          | 2826    | 0        | 0        | 2.5362E8 | 83.86757 | 10242.87 |
| CB_Int_DeudaPrestHipo  | INPUT | 39700.11 | 1388792               | 97174          | 2826    | 0        | 0        | 2.0544E8 | 96.70746 | 12039.7  |
| CB_Int_DeudaPrestPerso   | INPUT | 17057.69 | 467013.5              | 97174          | 2826    | 0        | 0        | 70630097 | 81.06222 | 9068.916 |
| CB_Int_DeudaPrestamos  | INPUT | 56757.8  | 1570940               | 97174          | 2826    | 0        | 0        | 2.5291E8 | 95.54696 | 12446.08 |
| CB_Int_DeudaTarjetas   | INPUT | 7946.629 | 96383.45              | 97174          | 2826    | 0        | 0        | 11197778 | 37.49435 | 2730.559 |
| CB_Int_DeudaTelco  | INPUT | 3802.755 | 20145.45              | 97174          | 2826    | 0        | 0        | 1277960  | 14.5108  | 448.5088 |
| CB_Int_DeudaTotal  | INPUT | 93106.32 | 1654938               | 97174          | 2826    | 0        | 0        | 2.5363E8 | 83.71385 | 10214.73 |
| CB_Int_MaxSaldoImpagado  | INPUT | 68671.63 | 1181067               | 97174          | 2826    | 0        | 0        | 1.9921E8 | 86.33556 | 11933.62 |
| CB_Num_Cedentes  | INPUT | 0.245786 | 0.918547              | 97174          | 2826    | 0        | 0        | 42       | 7.136097 | 104.2619 |
| CB_Num_DiasCB  | INPUT | 97.62521 | 319.0854              | 97174          | 2826    | 0        | 0        | 2040     | 3.504049 | 11.59335 |
| CB_Num_DiasDsdUltimaDeuda  | INPUT | 71.05232 | 254.3766              | 97174          | 2826    | 0        | 0        | 1963     | 4.137982 | 17.42658 |
| CB_Num_DiasEntreDeudas   | INPUT | 26.57289 | 149.162               | 97174          | 2826    | 0        | 0        | 1885     | 6.932336 | 52.59544 |
| CB_Tasa_TelcoTotal   | INPUT | 0.0668   | 0.243872              | 97174          | 2826    | 0        | 0        | 1        | 3.494447 | 10.34967 |
| Dda_Int_DeudaCli   | INPUT | 0.402447 | 91.35628              | 100000         | 0       | -243.32  | 0        | 28743.52 | 311.5608 | 97995.5  |
| Dda_Int_DeudaPS  | INPUT | 18.3028  | 129.5308              | 100000         | 0       | -1531.72 | 0        | 14042.82 | 32.78849 | 2492.714 |
| PS_Cod_SecCensal   | INPUT | 2.5034E9 | 1.5771E9              | 99060          | 940     | 0        | 2.8079E9 | 5.0297E9 | -0.14952 | -1.21765 |
| SD_Int_GtoAlqTV  | INPUT | 70.14272 | 36.34809              | 73888          | 26112   | 1.034182 | 64.17216 | 531.2285 | 1.794179 | 9.16496  |
| SD_Int_GtoAzar   | INPUT | 2.219844 | 5.077483              | 73888          | 26112   | 8.869E-7 | 1.022521 | 133.4034 | 9.963254 | 168.5078 |
| SD_Int_GtoCineTeatro   | INPUT | 8.34919  | 10.74391              | 73888          | 26112   | 0.000308 | 5.323318 | 192.2175 | 4.938361 | 45.59346 |
| SD_Int_GtoEqTelefno  | INPUT | 12.49741 | 6.522398              | 73888          | 26112   | 1.163517 | 11.68273 | 189.2487 | 4.367368 | 76.68185 |
| SD_Int_GtoFiestas  | INPUT | 99.56489 | 46.2113               | 73888          | 26112   | 11.3576  | 92.55899 | 666.5952 | 1.462785 | 4.189563 |
| SD_Int_GtoLibrosText   | INPUT | 415.3077 | 203.3129              | 73888          | 26112   | 41.98216 | 373.6051 | 2783.062 | 1.132936 | 2.628188 |
| SD_Int_GtoMaterialInf  | INPUT | 91.36146 | 34.74441              | 73888          | 26112   | 5.218622 | 86.18784 | 501.4241 | 1.432793 | 5.353812 |
| SD_Int_GtoRestoLibro   | INPUT | 38.79742 | 22.87558              | 73888          | 26112   | 1.945219 | 34.31812 | 377.1638 | 2.432997 | 12.91184 |
| SD_Int_GtoTV   | INPUT | 106.6581 | 32.56456              | 73888          | 26112   | 16.73346 | 102.6264 | 414.3101 | 1.176886 | 4.112285 |
| SD_Int_GtoTelefMovil   | INPUT | 486.6218 | 93.33984              | 73888          | 26112   | 199.1691 | 474.6068 | 1055.017 | 0.928533 | 1.841184 |
| SD_Int_GtoVacacionT  | INPUT | 79.22354 | 48.81303              | 73888          | 26112   | 1.620914 | 69.61651 | 538.2929 | 1.860646 | 6.380221 |
| SD_Int_IngrActivos   | INPUT | 2469.457 | 411.2973              | 73888          | 26112   | 1288.808 | 2416.884 | 4833.343 | 0.797691 | 0.911244 |
| SD_Int_IngrNetos   | INPUT | 2214.139 | 391.5988              | 73888          | 26112   | 1039.75  | 2172.044 | 4061.926 | 0.681862 | 0.63214  |
| SD_Int_IngrPensionistas  | INPUT | 1967.368 | 434.5524              | 73888          | 26112   | 690.7104 | 1912.426 | 5289.701 | 0.884485 | 1.815276 |
| SD_Int_IngrXPers   | INPUT | 1378.913 | 211.7965              | 73888          | 26112   | 582.2794 | 1369.664 | 2528.815 | 0.346907 | 0.412779 |
| SD_Tasa_HabExtr  | INPUT | 0.116499 | 0.086206              | 73888          | 26112   | 0        | 0.094054 | 0.717263 | 1.22753  | 1.697584 |

En la tabla de resumen, podemos ver que hay ausentes en casi todas las variables numéricas, por lo cual, es necesario imputar los datos antes de construir los modelos. Los valores grandes de kurtosis nos indican que hay más datos en las colas de la distribución, combinado con la gran diferencia entre la media y la mediana sugiere la presencia de valores extremos o atípicos. Por lo tanto, vamos a sacar gráficos de caja y bigote para comprobar si hay valores atípicos en los datos de “TrainCS”.

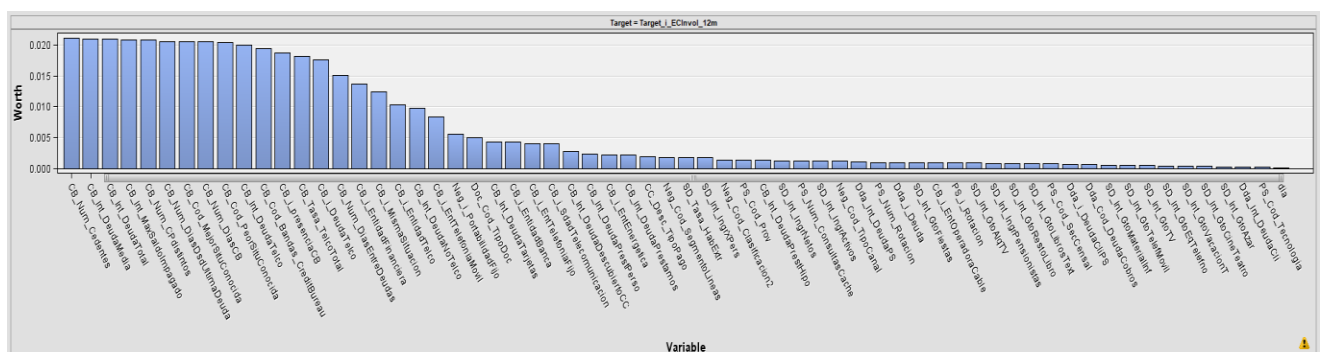


Comprobamos que existe valores muy extremos en algunas de las variables y necesitamos tratar los valores atípicos a continuación.

- Exploración de variables categóricas:

| Data  |                            |        | Number |         |                 |            |               |            |
|-------|----------------------------|--------|--------|---------|-----------------|------------|---------------|------------|
| Role  | Variable Name              | Role   | Levels | Missing | Mode            | Percentage | Mode2         | Mode2      |
|       |                            |        |        |         |                 |            |               | Percentage |
| TRAIN | CB_Cod_Bandas_CreditBureau | INPUT  | 6      | 0       | 00_No_en_ASNEF  | 86.45      | 04_Mas_de_300 | 8.44       |
| TRAIN | CB_Cod_MejorSituConocida   | INPUT  | 8      | 2826    | BL              | 86.00      | 01            | 8.08       |
| TRAIN | CB_Cod_PeorSituConocida    | INPUT  | 9      | 2826    | BL              | 86.00      | 01            | 5.36       |
| TRAIN | CB_Num_CPdistintos         | INPUT  | 8      | 2826    | 0               | 86.00      | 1             | 8.89       |
| TRAIN | CB_i_DeudaTelco            | INPUT  | 3      | 2826    | 0               | 88.76      | 1             | 8.42       |
| TRAIN | CB_i_EntEnergetica         | INPUT  | 3      | 2826    | 0               | 96.74      | .             | 2.83       |
| TRAIN | CB_i_EntOperadoraCable     | INPUT  | 3      | 2826    | 0               | 97.09      | .             | 2.83       |
| TRAIN | CB_i_EntTelefoniaFijo      | INPUT  | 3      | 2826    | 0               | 96.06      | .             | 2.83       |
| TRAIN | CB_i_EntTelefoniaMovil     | INPUT  | 3      | 2826    | 0               | 93.14      | 1             | 4.04       |
| TRAIN | CB_i_EntidadBanca          | INPUT  | 3      | 2826    | 0               | 94.34      | 1             | 2.84       |
| TRAIN | CB_i_EntidadFinanciera     | INPUT  | 3      | 2826    | 0               | 91.24      | 1             | 5.94       |
| TRAIN | CB_i_EntidadTelco          | INPUT  | 3      | 2826    | 0               | 92.09      | 1             | 5.08       |
| TRAIN | CB_i_MismaSituacion        | INPUT  | 3      | 2826    | 0               | 89.13      | 1             | 8.04       |
| TRAIN | CB_i_SdadTelecomunicacion  | INPUT  | 3      | 2826    | 0               | 95.84      | .             | 2.83       |
| TRAIN | CB_i_presenciaCB           | INPUT  | 2      | 0       | 0               | 86.00      | 1             | 14.00      |
| TRAIN | CC_Desc_TipoPago           | INPUT  | 5      | 24981   | DOMICILIADO     | 73.65      |               | 24.98      |
| TRAIN | Dda_Cod_DeudaCobros        | INPUT  | 3      | 0       | No_Deuda_Cobros | 95.27      | Deuda_PS      | 4.63       |
| TRAIN | Dda_i_Deuda                | INPUT  | 2      | 0       | 0               | 88.21      | 1             | 11.79      |
| TRAIN | Dda_i_DeudaClipS           | INPUT  | 2      | 0       | 0               | 91.43      | 1             | 8.57       |
| TRAIN | Doc_Cod_TipoDoc            | INPUT  | 3      | 0       | NIF             | 83.78      | NIE           | 15.25      |
| TRAIN | Neg_Cod_Clasificacion2     | INPUT  | 3      | 0       | RESIDENCIAL     | 91.17      | PYME          | 8.65       |
| TRAIN | Neg_Cod_SegmentoLineas     | INPUT  | 18     | 0       | RES_SOHO_0      | 89.10      | PYME_0        | 7.40       |
| TRAIN | Neg_Cod_TipoCanal          | INPUT  | 2      | 0       | CFY             | 75.02      | MOB           | 24.98      |
| TRAIN | Neg_i_PortabilidadFijo     | INPUT  | 2      | 0       | 0               | 59.79      | 1             | 40.21      |
| TRAIN | PS_Cod_Prov                | INPUT  | 45     | 0       | 28              | 21.03      | 08            | 12.48      |
| TRAIN | PS_Cod_Tecnologia          | INPUT  | 2      | 0       | CABLE           | 91.56      | ULL           | 8.44       |
| TRAIN | PS_Num_ConsultasCache      | INPUT  | 9      | 0       | 0               | 98.24      | 1             | 1.54       |
| TRAIN | PS_Num_Rotacion            | INPUT  | 10     | 0       | 0               | 55.85      | 1             | 29.96      |
| TRAIN | PS_i_Rotacion              | INPUT  | 2      | 0       | 0               | 55.85      | 1             | 44.15      |
| TRAIN | dia                        | INPUT  | 7      | 0       | 4               | 20.26      | 5             | 19.27      |
| TRAIN | Target_i_ECIInvol_12m      | TARGET | 2      | 0       | 0               | 89.70      | 1             | 10.30      |

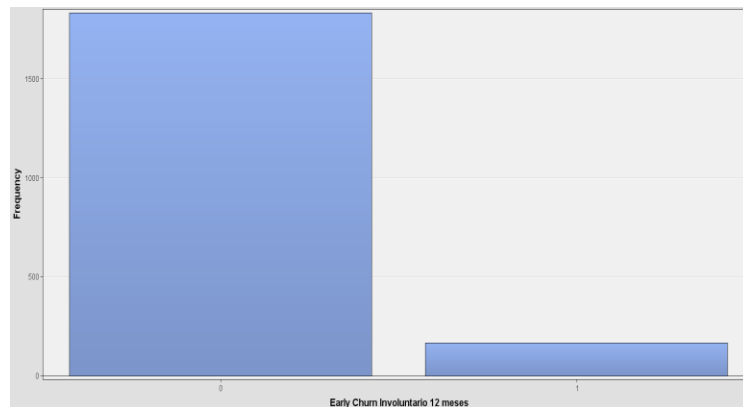
Mirando a los estadísticos básicos de las variables categóricas, vemos que también hay datos perdidos en estas variables. Con el mismo nodo de StatExplore, obtenemos gráficos en los que podemos observar cómo se comportan las variables en función de la variable objetivo. Observamos que las variables que tienen gran influencia en la variable objetivo son las variables del grupo bureau crédito, por ejemplo, *“CB\_Num\_Cedentes”*, *“CB\_Int\_DeudaMedia”*, *“CB\_Deuda\_Total”* y *“CB\_Int\_MaxSaldolmpagado”*.



Si profundizamos en el análisis de la variable objetivo, vemos que en toda la muestra con 100,061 observaciones, la decisión anticipada de churn ocupa un 10.304% de los casos, según la tabla a continuación:

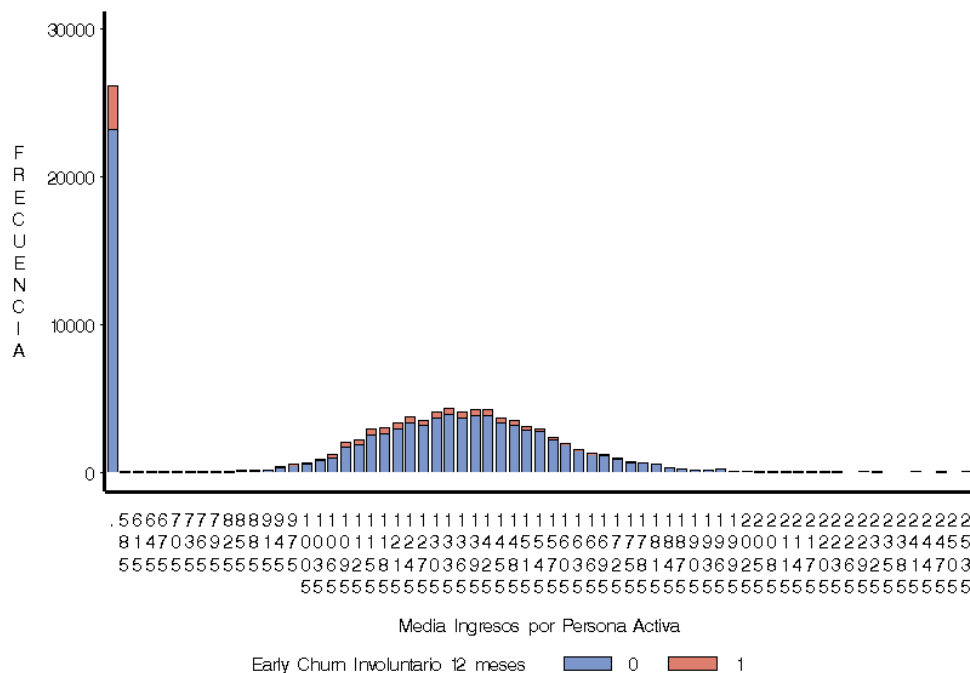
| Data  |                      |        |       | Frequency |         |
|-------|----------------------|--------|-------|-----------|---------|
| Role  | Variable Name        | Role   | Level | Count     | Percent |
| TRAIN | Target_i_ECInvol_12m | TARGET | 0     | 89696     | 89.696  |
| TRAIN | Target_i_ECInvol_12m | TARGET | 1     | 10304     | 10.304  |

Por otro lado, podemos hacer histogramas para ver la distribución de las variables categóricas usando el nodo GraphExplore. Por ejemplo, obtenemos un histograma que nos represente como se distribuye la variable objetivo “**Early Churn Involuntario 12 meses**”.



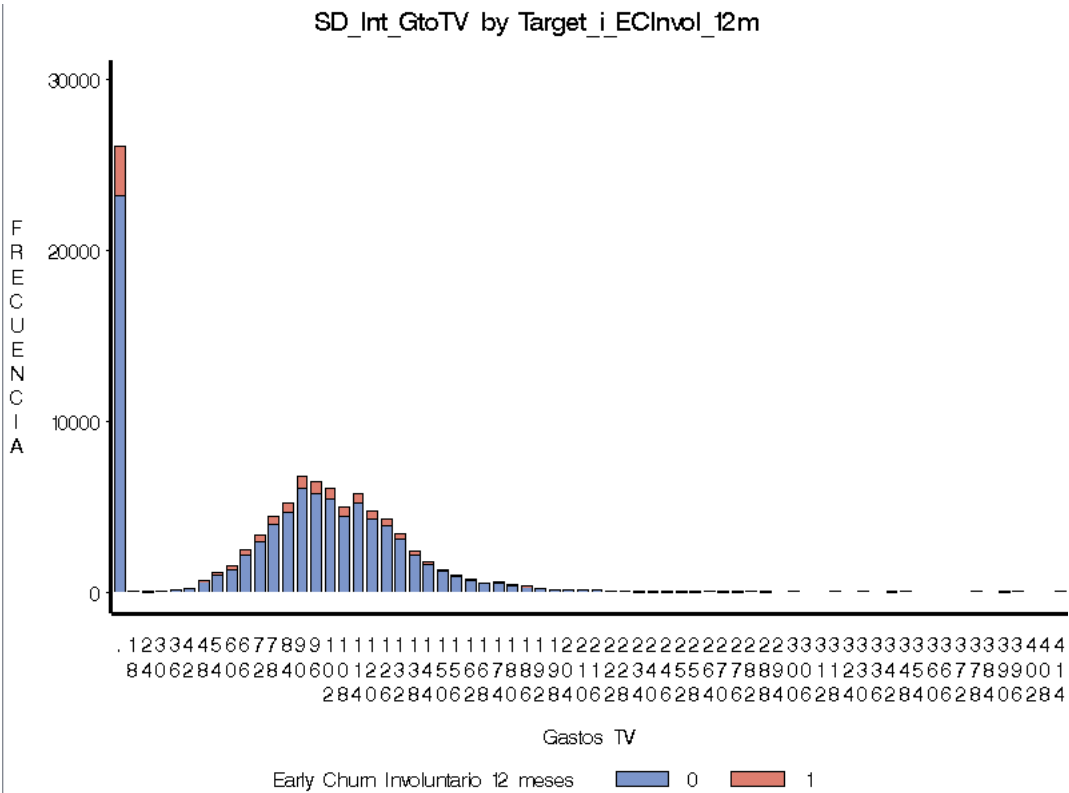
Con el nodo multigráfico, podemos analizar qué variables a priori influirán más en la predicción de la variable objetivo. Para ello, vamos a construir varios gráficos de barras para observar si existen diferencias significativas entre los niveles de la variable objetivo.

SD\_Int\_IngrXPers by Target\_i\_ECInvol\_12m

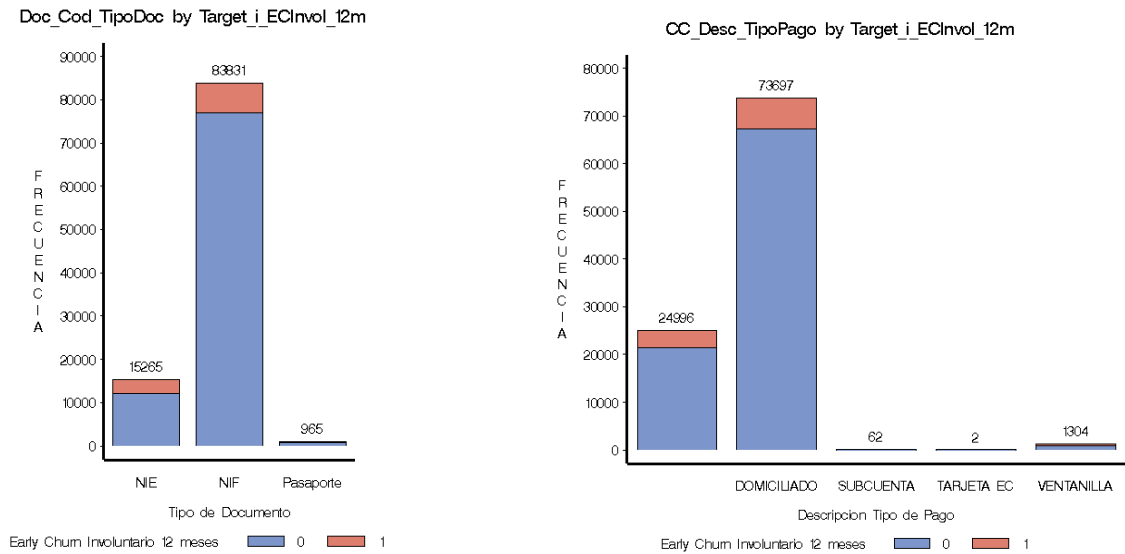


A primera vista, se observa que la “**Media Ingresos por Persona Activa**” influye a la decisión de cancelación anticipada, ya que el porcentaje de cancelación aumenta según va bajando el rango de ingreso medio.





La variable de “**Gastos TV**” tiene un comportamiento similar a la variable “**Media Ingresos por Persona Activa**”, lo que nos sugiere una influencia inversa a la decisión de cancelación anticipada.



En cuanto al tipo de documento y la descripción tipo de pago de los clientes, encontramos diferencias entre los grupos distintos como se muestran en los gráficos.

## 2.2. CS\_Rejected

Este conjunto de datos contiene información de los usuarios a los que el departamento de riesgos no se les aprobó la entrada a la compañía. Es importante añadir esta información al modelo para evitar el sesgo de la muestra. Como característica principal de este conjunto de datos, la variable objetivo no existe, puesto que les hemos rechazado desde principio. Por lo tanto, lo hemos de importar con el rol de puntuación y quitamos la variable.

Al importar el fichero de datos, también eliminamos una variable unaria “*Doc\_i\_PersonaJuridica*” y 2 variables categóricas con más de 20 niveles: “*PS\_Cod\_CP*” y “*PS\_Cod\_Prov*”

Para este fichero, realizamos el análisis exploratorio con el nodo StatExplore en Sas EM y comprobamos que también tiene ausentes y valores atípicos, ya que las variables tienen el valor de curtosis grande y hay mucha diferencia entre la media y la mediana.

| Variable                  | Role  | Mean     | Standard<br>Deviation | Non<br>Missing | Missing | Minimum  | Median   | Maximum  | Skewness | Kurtosis |
|---------------------------|-------|----------|-----------------------|----------------|---------|----------|----------|----------|----------|----------|
| CB_Int_DeudaDescubiertoCC | INPUT | 1808.722 | 47157.07              | 27959          | 1150    | 0        | 0        | 4926337  | 88.14072 | 8829.03  |
| CB_Int_DeudaMedia         | INPUT | 50206.45 | 523567.6              | 27959          | 1150    | 0        | 0        | 32622829 | 31.37087 | 1323.08  |
| CB_Int_DeudaNoTelco       | INPUT | 125877   | 1365363               | 27959          | 1150    | 0        | 0        | 1.0265E8 | 32.75063 | 1678.927 |
| CB_Int_DeudaPrestHipo     | INPUT | 51826.66 | 1058067               | 27959          | 1150    | 0        | 0        | 66791165 | 33.8751  | 1436.7   |
| CB_Int_DeudaPrestPerso    | INPUT | 22042.98 | 409128.4              | 27959          | 1150    | 0        | 0        | 32884511 | 42.03906 | 2396.101 |
| CB_Int_DeudaPrestamos     | INPUT | 73869.65 | 1157245               | 27959          | 1150    | 0        | 0        | 66791165 | 29.10992 | 1091.431 |
| CB_Int_DeudaTarjetas      | INPUT | 11723.56 | 98557.33              | 27959          | 1150    | 0        | 0        | 6042476  | 20.655   | 754.3701 |
| CB_Int_DeudaTelco         | INPUT | 7746.54  | 33060.73              | 27959          | 1150    | 0        | 0        | 2429983  | 22.07969 | 1203.851 |
| CB_Int_DeudaTotal         | INPUT | 133623.5 | 1367613               | 27959          | 1150    | 0        | 0        | 1.0265E8 | 32.63273 | 1669.235 |
| CB_Int_MaxSaldoImpagado   | INPUT | 101627.8 | 1045907               | 27959          | 1150    | 0        | 0        | 53267226 | 26.09382 | 892.6366 |
| CB_Num_Cedentes           | INPUT | 0.446332 | 1.264408              | 27959          | 1150    | 0        | 0        | 48       | 7.147745 | 137.5036 |
| CB_Num_DiasCB             | INPUT | 171.3357 | 412.9785              | 27959          | 1150    | 0        | 0        | 2045     | 2.427133 | 4.687171 |
| CB_Num_DiasDsdUltimaDeuda | INPUT | 121.5355 | 325.8096              | 27959          | 1150    | 0        | 0        | 1943     | 3.002189 | 8.468419 |
| CB_Num_DiasEntreDeudas    | INPUT | 49.80017 | 201.8852              | 27959          | 1150    | 0        | 0        | 1855     | 4.945821 | 25.94975 |
| CB_Tasa_TelcoTotal        | INPUT | 0.114091 | 0.308561              | 27959          | 1150    | 0        | 0        | 1        | 2.448127 | 4.103899 |
| Dda_Int_DeudaCli          | INPUT | 0.118559 | 7.351173              | 29109          | 0       | -420.67  | 0        | 515.94   | 29.9477  | 2892.694 |
| Dda_Int_DeudaPS           | INPUT | 24.53801 | 147.4736              | 29109          | 0       | -417.6   | 0        | 10241.35 | 21.5222  | 1009.098 |
| NumAleat                  | INPUT | 0.720554 | 0.095426              | 29109          | 0       | 0.630002 | 0.684987 | 0.999997 | 1.480963 | 1.027444 |
| PS_Cod_SecCensal          | INPUT | 2.4945E9 | 1.5883E9              | 28853          | 256     | 0        | 2.8079E9 | 5.0297E9 | -0.13997 | -1.24293 |
| SD_Int_GtoAlgTV           | INPUT | 68.63054 | 35.6798               | 21379          | 7730    | 1.034182 | 62.94084 | 514.386  | 1.621252 | 6.816122 |
| SD_Int_GtoAzar            | INPUT | 2.28756  | 5.654173              | 21379          | 7730    | 8.869E-7 | 1.003254 | 133.4034 | 10.60227 | 176.8289 |
| SD_Int_GtoCineTeatro      | INPUT | 8.104636 | 10.2723               | 21379          | 7730    | 0.000421 | 5.235306 | 192.2175 | 4.936845 | 46.92916 |
| SD_Int_GtoEqTelefno       | INPUT | 12.38377 | 6.090992              | 21379          | 7730    | 1.163517 | 11.54605 | 104.9603 | 1.916603 | 11.36491 |
| SD_Int_GtoFiestas         | INPUT | 97.87378 | 45.7853               | 21379          | 7730    | 14.13053 | 90.5444  | 666.5952 | 1.448443 | 4.279795 |
| SD_Int_GtoLibrosText      | INPUT | 407.7138 | 202.0603              | 21379          | 7730    | 41.98216 | 365.4352 | 2392.748 | 1.193916 | 2.750437 |
| SD_Int_GtoMaterialInf     | INPUT | 90.22103 | 33.83846              | 21379          | 7730    | 6.132817 | 85.28632 | 501.4241 | 1.33799  | 4.55045  |
| SD_Int_GtoRestoLibro      | INPUT | 37.99693 | 22.33673              | 21379          | 7730    | 2.854787 | 33.48124 | 299.6282 | 2.394796 | 12.36581 |
| SD_Int_GtoTV              | INPUT | 105.7196 | 32.22877              | 21379          | 7730    | 16.73346 | 101.7844 | 414.3101 | 1.215022 | 4.449948 |
| SD_Int_GtoTelefMovil      | INPUT | 485.1578 | 92.62423              | 21379          | 7730    | 234.6248 | 473.5163 | 1055.017 | 0.886704 | 1.594909 |
| SD_Int_GtoVacacionT       | INPUT | 79.01601 | 48.43275              | 21379          | 7730    | 1.620914 | 69.61651 | 538.2929 | 1.970037 | 7.614344 |
| SD_Int_IngrActivos        | INPUT | 2449.636 | 409.0351              | 21379          | 7730    | 1288.808 | 2397.94  | 4915.61  | 0.795355 | 0.922145 |
| SD_Int_IngrNetos          | INPUT | 2195.694 | 389.458               | 21379          | 7730    | 1107.258 | 2156.33  | 4061.926 | 0.667979 | 0.616536 |
| SD_Int_IngrPensionistas   | INPUT | 1953.622 | 437.1663              | 21379          | 7730    | 690.7104 | 1893.809 | 5289.701 | 0.964294 | 2.325333 |
| SD_Int_IngrXPers          | INPUT | 1364.876 | 211.5252              | 21379          | 7730    | 582.2794 | 1352.738 | 2422.049 | 0.346146 | 0.342689 |
| SD_Tasa_HabExtr           | INPUT | 0.122284 | 0.089828              | 21379          | 7730    | 0        | 0.099046 | 0.717263 | 1.169763 | 1.459708 |

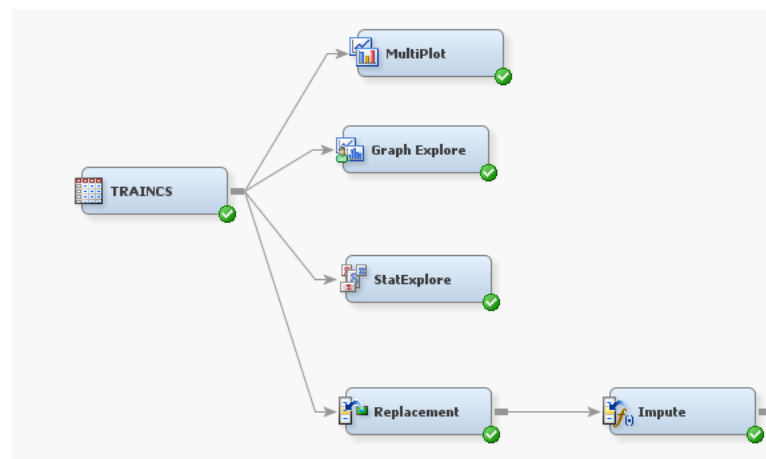
| Data Role | Variable Name              | Role  | Number of Levels | Missing | Mode            | Mode Percentage | Mode2         | Mode2 Percentage |
|-----------|----------------------------|-------|------------------|---------|-----------------|-----------------|---------------|------------------|
| TRAIN     | CB_Cod_Bandas_CreditBureau | INPUT | 6                | 0       | 00_No_en_ASNEF  | 78.10           | 04_Mas_de_300 | 14.26            |
| TRAIN     | CB_Cod_MejorSituConocida   | INPUT | 8                | 1150    | BL              | 77.22           | 01            | 13.99            |
| TRAIN     | CB_Cod_PeorSituConocida    | INPUT | 9                | 1150    | BL              | 77.22           | 01            | 8.90             |
| TRAIN     | CB_Num_CPdistintos         | INPUT | 7                | 1150    | 0               | 77.22           | 1             | 14.44            |
| TRAIN     | CB_i_DeudaTelco            | INPUT | 3                | 1150    | 0               | 81.52           | 1             | 14.52            |
| TRAIN     | CB_i_EntEnergetica         | INPUT | 3                | 1150    | 0               | 95.32           | .             | 3.95             |
| TRAIN     | CB_i_EntOperadoraCable     | INPUT | 3                | 1150    | 0               | 95.88           | .             | 3.95             |
| TRAIN     | CB_i_EntTelefoniaFijo      | INPUT | 3                | 1150    | 0               | 93.84           | .             | 3.95             |
| TRAIN     | CB_i_EntTelefoniaIntegrada | INPUT | 3                | 1150    | 0               | 96.05           | .             | 3.95             |
| TRAIN     | CB_i_EntTelefoniaMovil     | INPUT | 3                | 1150    | 0               | 89.07           | 1             | 6.98             |
| TRAIN     | CB_i_EntidadBanca          | INPUT | 3                | 1150    | 0               | 91.42           | 1             | 4.63             |
| TRAIN     | CB_i_EntidadFinanciera     | INPUT | 3                | 1150    | 0               | 85.52           | 1             | 10.53            |
| TRAIN     | CB_i_EntidadTelco          | INPUT | 3                | 1150    | 0               | 87.28           | 1             | 8.77             |
| TRAIN     | CB_i_MismaSituacion        | INPUT | 3                | 1150    | 0               | 83.03           | 1             | 13.02            |
| TRAIN     | CB_i_SdadTelecomunicacion  | INPUT | 3                | 1150    | 0               | 93.51           | .             | 3.95             |
| TRAIN     | CB_i_presenciaCB           | INPUT | 2                | 0       | 0               | 77.22           | 1             | 22.78            |
| TRAIN     | CC_Desc_TipoPago           | INPUT | 4                | 7984    | DOMICILIADO     | 70.70           |               | 27.43            |
| TRAIN     | Dda_Cod_DeudaCobros        | INPUT | 3                | 0       | No_Deuda_Cobros | 94.04           | Deuda_PS      | 5.84             |
| TRAIN     | Dda_i_Deuda                | INPUT | 2                | 0       | 0               | 85.89           | 1             | 14.11            |
| TRAIN     | Dda_i_DeudaCliPS           | INPUT | 2                | 0       | 0               | 89.54           | 1             | 10.46            |
| TRAIN     | Doc_Cod_TipoDoc            | INPUT | 3                | 0       | NIF             | 78.67           | NIE           | 19.64            |
| TRAIN     | Neg_Cod_Clasificacion2     | INPUT | 3                | 0       | RESIDENCIAL     | 89.26           | PYME          | 10.50            |
| TRAIN     | Neg_Cod_SegmentoLineas     | INPUT | 12               | 0       | RES_SOHO_0      | 87.40           | PYME_0        | 9.32             |
| TRAIN     | Neg_Cod_TipoCanal          | INPUT | 2                | 0       | CFY             | 72.57           | MOB           | 27.43            |
| TRAIN     | Neg_i_PortabilidadFijo     | INPUT | 2                | 0       | 0               | 67.08           | 1             | 32.92            |
| TRAIN     | PS_Cod_Tecnologia          | INPUT | 2                | 0       | CABLE           | 90.83           | ULL           | 9.17             |
| TRAIN     | PS_Num_ConsultasCache      | INPUT | 9                | 0       | 0               | 97.51           | 1             | 2.10             |
| TRAIN     | PS_Num_Rotacion            | INPUT | 9                | 0       | 0               | 52.51           | 1             | 31.71            |
| TRAIN     | PS_i_Rotacion              | INPUT | 2                | 0       | 0               | 52.51           | 1             | 47.49            |
| TRAIN     | dia                        | INPUT | 7                | 0       | 4               | 20.03           | 5             | 19.22            |

### 2.3. Scoring\_CS

El último conjunto de datos contiene 66647 observaciones, con las que vamos a hacer una prueba aplicando el modelo scorecard definitivo, y en función de los resultados decidiremos si conceder o no el crédito. Por eso, lo importamos con el rol de puntuación para sacar la predicción si ocurre el evento de “Early Churn Involuntario 12 meses” o no.

## 2. TRANSFORMACIÓN DE DATOS

Hemos comprobado en la fase exploratorio de los datos que existe ausentes y valores atípicos en el fichero de TrainCS. Por lo cual, usamos el nodo Reemplazo para recortar la distribución de las variables y imputamos los datos perdidos con el nodo Imputar utilizando el método de árbol.



### 3. SELECCIÓN DE DATOS

Antes de seleccionar las variables importantes para el scorecard, es necesario hacer la partición de los datos en 70% - 30%, entrenamiento y validación sucesivamente con el nodo Data Partition para evitar el sobreajuste. El siguiente paso será utilizar el nodo de agrupación interactiva (Interactive Bining) con el objetivo de eliminar características débiles o aquellas que no se ajustan a una buena lógica empresarial y agrupar las características más fuertes (es decir, crear niveles de atributos) para producir un modelo en formato de cuadro de mando.

Para la construcción de una tarjeta de puntuación es necesario la tramificación de las variables de intervalo, puesto que tendríamos sino tantos niveles como valores distintos existieran en las observaciones de la variable. Con el nodo agrupación interactiva podemos tramificar las variables, configurándolo para que el número máximo de tramos o niveles sea 5. Como resultados, obtenemos cada variable de intervalo y de clase agrupada en 5 grupos como hemos especificado.

Siguiente se calcula y examina las métricas de evaluación importantes, WOE, Gini e IV. El valor WOE (Weights of evidence) se utiliza para evaluar el grado de discriminación de los atributos para cada característica y el valor IV (Information value) se utiliza para evaluar el poder predictivo global de una característica, es decir, su capacidad para separar los casos buenos de los malos. El estadístico de Gini es otro estadístico utilizado para evaluar el poder predictivo global de una variable. poder predictivo global de una variable. Puede utilizarse como alternativa a IV (la opción por defecto) para seleccionar variables variables para introducirlas en el cuadro de mando.

Obtenemos los siguientes resultados al ejecutar el nodo agrupación interactiva:

| Variable               | Gini Statistic | Information Value | Level for Interactive | Calculated Role | New Role | Pre-Defined Grouping | Level    | Label                    | Information Value Ordering |
|------------------------|----------------|-------------------|-----------------------|-----------------|----------|----------------------|----------|--------------------------|----------------------------|
| CB i presenciaCB       | 36.916         | 0.76              | BINARY                | Input           | Default  |                      | BINARY   | Indicador presencia...   | 1                          |
| IMP CB Cod Mejor...    | 36.916         | 0.76              | NOMINAL               | Input           | Default  |                      | NOMINAL  | Imputed: Mejor Situa...  | 2                          |
| IMP CB Cod PeorS...    | 36.916         | 0.76              | NOMINAL               | Input           | Default  |                      | NOMINAL  | Imputed: Peor Situa...   | 3                          |
| IMP REP CB Num...      | 36.916         | 0.76              | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 4                          |
| IMP CB Num CPdi...     | 36.889         | 0.759             | NOMINAL               | Input           | Default  |                      | NOMINAL  | Imputed: CP distintos... | 5                          |
| CB Cod Bandas Cr...    | 35.505         | 0.718             | NOMINAL               | Input           | Default  |                      | NOMINAL  | Bandas Bureau Cred...    | 6                          |
| IMP REP CB Int D...    | 30.925         | 0.654             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 7                          |
| IMP REP CB Int M...    | 30.847         | 0.651             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 8                          |
| IMP REP CB Int D...    | 30.724         | 0.646             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 9                          |
| IMP REP CB Num...      | 29.193         | 0.59              | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 10                         |
| IMP REP CB Num...      | 28.74          | 0.573             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 11                         |
| Ned i PortabilidadFilo | 28.353         | 0.417             | BINARY                | Input           | Default  |                      | BINARY   | Indicador Portabilidad   | 12                         |
| Doc Cod TipoDoc        | 19.552         | 0.221             | NOMINAL               | Input           | Default  |                      | NOMINAL  | Tipo de Documento        | 13                         |
| IMP REP SD Int In...   | 12.732         | 0.062             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 14                         |
| PS Num Rotacion        | 12.318         | 0.061             | NOMINAL               | Input           | Default  |                      | NOMINAL  | Rotacion punto de s...   | 15                         |
| PS i Rotacion          | 12.316         | 0.061             | BINARY                | Input           | Default  |                      | BINARY   | Indicador Rotacion       | 16                         |
| Ned Cod TipoCanal      | 10.105         | 0.05              | BINARY                | Input           | Default  |                      | BINARY   | Tipo de canal de con...  | 17                         |
| PS Cod Prov            | 11.289         | 0.049             | NOMINAL               | Input           | Default  |                      | NOMINAL  | Código Provincia Pu...   | 18                         |
| IMP REP SD Int In...   | 10.623         | 0.043             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 19                         |
| IMP REP SD Int In...   | 10.437         | 0.04              | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 20                         |
| IMP REP SD Tasa...     | 10.611         | 0.039             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 21                         |
| IMP REP SD Int G...    | 10.158         | 0.034             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 22                         |
| IMP REP SD Int G...    | 9.803          | 0.029             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 23                         |
| IMP REP SD Int G...    | 7.551          | 0.026             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 24                         |
| IMP REP SD Int In...   | 8.532          | 0.024             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 25                         |
| IMP REP SD Int G...    | 8.186          | 0.023             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...    | 26                         |

Seleccionamos 26 variables con el poder predictivo que tengan el criterio IV mayor que 0.02, puesto que un valor menor de 0.02 implica que la variable no tiene influencia en la variable objetivo.

En este paso, probamos la opción de retocar manualmente los tramos obtenidos en las variables con el nodo de agrupación interactiva basando en el valor de WOE para conseguir el mejor poder predictivo medido por el criterio Gini y IV.

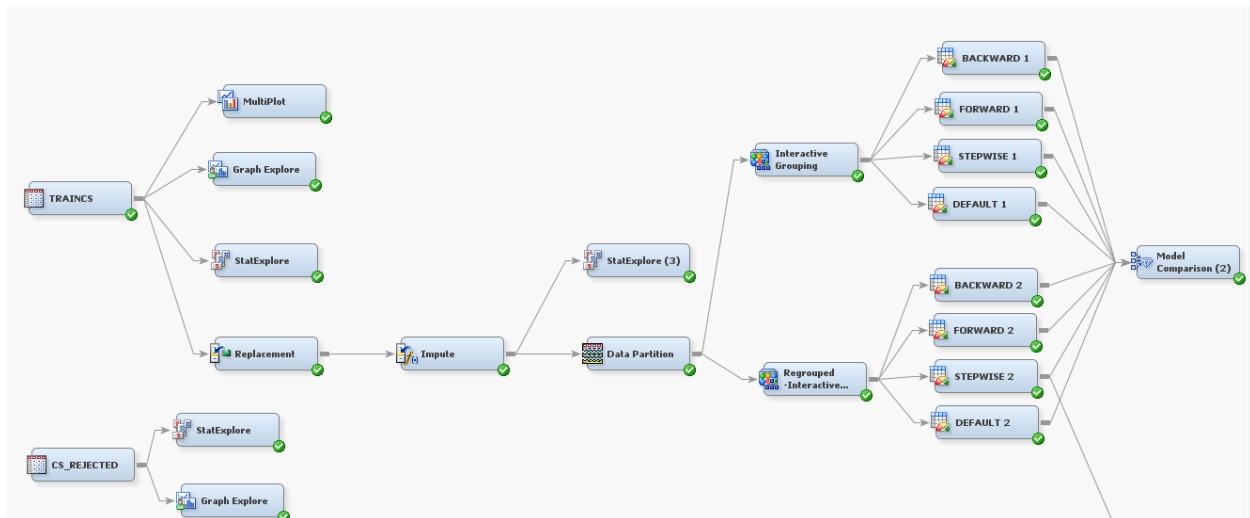
|                      |     |
|----------------------|-----|
| Train                |     |
| Variables            | ... |
| Interactive Grouping | ... |

| Variable               | Gini Statistic | Information Value | Level for Interactive | New Role | Calculated Role | Pre-Defined Grouping | Measurement Level | Label                    | Information Value Ordering |
|------------------------|----------------|-------------------|-----------------------|----------|-----------------|----------------------|-------------------|--------------------------|----------------------------|
| IMP CB Num CPdi...     | 37.393         | 0.781             | NOMINAL               | Default  | Input           |                      | NOMINAL           | Imputed: CP distintos... | 1                          |
| IMP CB Cod PeorS...    | 37.31          | 0.77              | NOMINAL               | Default  | Input           |                      | NOMINAL           | Imputed: Peor Situa...   | 2                          |
| IMP CB Cod Mejor...    | 37.024         | 0.762             | NOMINAL               | Default  | Input           |                      | NOMINAL           | Imputed: Mejor Situa...  | 3                          |
| CB i presenciaCB       | 36.916         | 0.76              | BINARY                | Default  | Input           |                      | BINARY            | Indicador presencia ...  | 4                          |
| IMP REP CB Num...      | 37.004         | 0.76              | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 5                          |
| CB Cod Bandas Cr...    | 35.865         | 0.731             | NOMINAL               | Default  | Input           |                      | NOMINAL           | Bandas Bureau Cred...    | 6                          |
| IMP REP CB Int D...    | 31.02          | 0.655             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 7                          |
| IMP REP CB Int M...    | 30.847         | 0.651             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 8                          |
| IMP REP CB Int D...    | 30.724         | 0.646             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 9                          |
| IMP REP CB Num...      | 29.328         | 0.592             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 10                         |
| IMP REP CB Num...      | 28.832         | 0.575             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 11                         |
| Ned i PortabilidadFilo | 28.353         | 0.417             | BINARY                | Default  | Input           |                      | BINARY            | Indicador Portabilidad   | 12                         |
| Doc Cod TipoDoc        | 19.685         | 0.224             | NOMINAL               | Default  | Input           |                      | NOMINAL           | Tipo de Documento        | 13                         |
| IMP REP SD Int In...   | 13.749         | 0.074             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 14                         |
| PS Num Rotacion        | 13.617         | 0.068             | NOMINAL               | Default  | Input           |                      | NOMINAL           | Rotación punto de s...   | 15                         |
| PS i Rotacion          | 12.316         | 0.061             | BINARY                | Default  | Input           |                      | BINARY            | Indicador Rotacion       | 16                         |
| Ned Cod TipoCanal      | 10.105         | 0.05              | BINARY                | Default  | Input           |                      | BINARY            | Tipo de canal de con...  | 17                         |
| PS Cod Prov            | 11.364         | 0.05              | NOMINAL               | Default  | Input           |                      | NOMINAL           | Código Provincia Pu...   | 18                         |
| IMP REP SD Tasa...     | 11.072         | 0.048             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 19                         |
| IMP REP SD Int In...   | 11.132         | 0.047             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 20                         |
| IMP REP SD Int In...   | 11.301         | 0.046             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 21                         |
| IMP REP SD Int G...    | 10.625         | 0.039             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 22                         |
| IMP REP SD Int G...    | 9.703          | 0.038             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 23                         |
| IMP REP SD Int G...    | 9.283          | 0.037             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 24                         |
| IMP REP SD Int G...    | 8.051          | 0.03              | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 25                         |
| IMP REP SD Int In...   | 8.728          | 0.026             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...    | 26                         |

Después de retocar la agrupación, vemos que obtenemos los valores Gini y IV mayor que antes.

#### 4. ESTIMACIÓN DEL PRIMER MODELO Y EL SCORECARD PRELIMINAR

Para obtener el scorecard preliminar, probamos diferentes modelos como “default”, “stepwise”, “forward”, “backward” con los datos originales y retocados obtenidos del nodo agrupación interactiva y los comparamos entre ellos utilizando los criterios de AIC, SBIC, RMSE, el índice ROC y Gini.



Con el nodo comparación de modelos, obtenemos la siguiente tabla de resultados:

| Modelo     | Entrenar: AIC | Entrenar: SBIC | Entrenar: RMSE | Validar: RMSE | Train: Roc Index | Valid: Roc Index | Train: Gini Coefficient | Valid: Gini Coefficient |
|------------|---------------|----------------|----------------|---------------|------------------|------------------|-------------------------|-------------------------|
| STEPWISE 2 | 37852.16331   | 38016.98661    | 0.279461151    | 0.28115419    | 0.805            | 0.798            | 0.611                   | 0.597                   |
| BACKWARD 2 | 37852.16331   | 38016.98661    | 0.279461151    | 0.28115419    | 0.805            | 0.798            | 0.611                   | 0.597                   |
| FORWARD 2  | 37854.04689   | 38028.02705    | 0.279462408    | 0.281140641   | 0.805            | 0.798            | 0.611                   | 0.597                   |
| DEFAULT 2  | 37863.52983   | 38110.76478    | 0.279442349    | 0.281151583   | 0.805            | 0.798            | 0.611                   | 0.596                   |
| BACKWARD 1 | 38088.228     | 38225.58076    | 0.28051683     | 0.281430102   | 0.803            | 0.798            | 0.605                   | 0.595                   |
| FORWARD 1  | 38088.89573   | 38217.09164    | 0.280525266    | 0.281442086   | 0.803            | 0.798            | 0.605                   | 0.595                   |
| STEPWISE 1 | 38088.89573   | 38217.09164    | 0.280525266    | 0.281442086   | 0.803            | 0.798            | 0.605                   | 0.595                   |
| DEFAULT 1  | 38098.57317   | 38318.33758    | 0.280503918    | 0.281620426   | 0.803            | 0.797            | 0.605                   | 0.594                   |

⇒ Según la tabla, el mejor scorecard en esta fase se trata de un modelo “Stepwise” o “Backward” utilizando las variables retocadas manualmente en el paso anterior, puesto que estos dos modelos tienen la mayor área bajo la curva ROC, mayor Índice de Gini y menor AIC, menor SBIC

y menor RMSE. Por lo cual, elegimos el modelo STEPWISE 2 para construir el scorecard preliminar.

- El scorecard preliminar sacado con el modelo STEPWISE 2

El scorecard obtenido se compone de 18 variables significativas como se puede observar en la tabla de resumen:

| Step | Entered                          | Effect<br>Removed                | DF | Number<br>In | Score<br>Chi-Square | Wald<br>Chi-Square | Pr > ChiSq | Criterio de<br>información<br>de Akaike |
|------|----------------------------------|----------------------------------|----|--------------|---------------------|--------------------|------------|---|
| 1    | WOE_IMP_CB_Num_CPdistintos       |                                  | 1  | 1            | 7648.1665           |                    | <.0001     | 40959.6                                 |
| 2    | WOE_Neg_i_PortabilidadFijo       |                                  | 1  | 2            | 1436.2000           |                    | <.0001     | 39385.9                                 |
| 3    | WOE_Doc_Cod_TipoDoc              |                                  | 1  | 3            | 814.8141            |                    | <.0001     | 38646.6                                 |
| 4    | WOE_Neg_Cod_TipoCanal            |                                  | 1  | 4            | 231.2559            |                    | <.0001     | 38426.4                                 |
| 5    | WOE_IMP_REP_SD_Int_IngrXPers     |                                  | 1  | 5            | 183.8435            |                    | <.0001     | 38239.2                                 |
| 6    | WOE_IMP_REP_CB_Int_DeudaTotal    |                                  | 1  | 6            | 89.2194             |                    | <.0001     | 38149.8                                 |
| 7    | WOE_PS_Cod_Prov                  |                                  | 1  | 7            | 86.2008             |                    | <.0001     | 38067.2                                 |
| 8    | WOE_IMP_REP_SD_Int_GtoAlqTV      |                                  | 1  | 8            | 48.9455             |                    | <.0001     | 38020.0                                 |
| 9    | WOE_PS_Num_Rotacion              |                                  | 1  | 9            | 43.0302             |                    | <.0001     | 37979.2                                 |
| 10   | WOE_IMP_REP_SD_Tasa_HabExtr      |                                  | 1  | 10           | 26.0634             |                    | <.0001     | 37955.4                                 |
| 11   | WOE_IMP_REP_SD_Int_IngrPensionis |                                  | 1  | 11           | 26.9256             |                    | <.0001     | 37930.2                                 |
| 12   | WOE_IMP_REP_SD_Int_GtoLibrosText |                                  | 1  | 12           | 16.8198             |                    | <.0001     | 37915.2                                 |
| 13   | WOE_IMP_REP_SD_Int_GtoRestoLibro |                                  | 1  | 13           | 7.7456              |                    | 0.0054     | 37909.4                                 |
| 14   | WOE_IMP_REP_CB_Num_DiasDsdUltima |                                  | 1  | 14           | 4.9673              |                    | 0.0258     | 37906.4                                 |
| 15   | WOE_IMP_CB_Cod_PeorSituConocida  |                                  | 1  | 15           | 5.1862              |                    | 0.0228     | 37903.2                                 |
| 16   | WOE_CB_i_presenciaCB             |                                  | 1  | 16           | 6.8764              |                    | 0.0087     | 37898.4                                 |
| 17   | WOE_IMP_CB_Cod_MejorSituConocida |                                  | 1  | 17           | 23.2375             |                    | <.0001     | 37876.6                                 |
| 18   | WOE_IMP_REP_CB_Num_Cedentes      |                                  | 1  | 18           | 24.6209             |                    | <.0001     | 37854.0                                 |
| 19   |                                  | WOE_IMP_REP_CB_Num_DiasDsdUltima | 1  | 17           |                     | 0.1165             | 0.7329     | 37852.2                                 |
| 20   |                                  | WOE_IMP_CB_Cod_PeorSituConocida  | 1  | 16           |                     | 3.4424             | 0.0635     | 37853.6                                 |

Los parámetros incluidos se detallan a continuación:

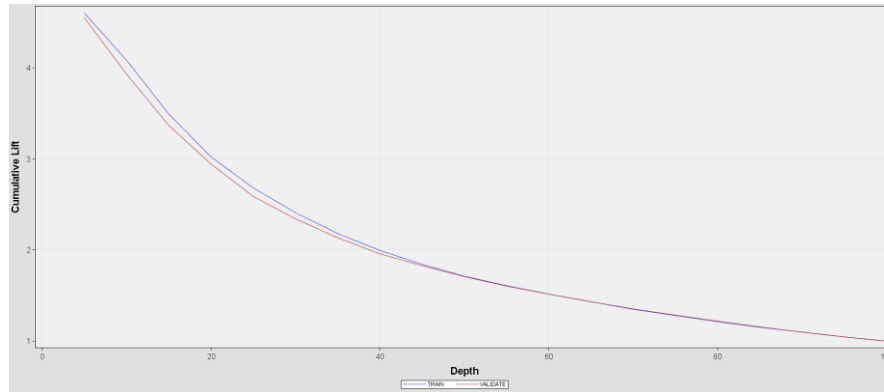
| Parameter                        | DF | Estimate | Standard<br>Error | Wald<br>Chi-Square | Pr > ChiSq | Estimador<br>estandarizado | Exp(Est) |
|----------------------------------|----|----------|-------------------|--------------------|------------|----------------------------|----------|
| Intercept                        | 1  | -2.1668  | 0.0145            | 22276.37           | <.0001     |                            | 0.115    |
| WOE_CB_i_presenciaCB             | 1  | 5.3981   | 0.7450            | 52.50              | <.0001     | 2.1272                     | 220.987  |
| WOE_Doc_Cod_TipoDoc              | 1  | -0.7064  | 0.0273            | 670.48             | <.0001     | -0.1620                    | 0.493    |
| WOE_IMP_CB_Cod_MejorSituConocida | 1  | -1.3648  | 0.2681            | 25.92              | <.0001     | -0.5381                    | 0.255    |
| WOE_IMP_CB_Cod_PeorSituConocida  | 1  | -0.2453  | 0.1322            | 3.44               | 0.0635     | -0.0969                    | 0.782    |
| WOE_IMP_CB_Num_CPdistintos       | 1  | -0.5028  | 0.0940            | 28.62              | <.0001     | -0.1991                    | 0.605    |
| WOE_IMP_REP_SD_Int_IngrXPers     | 1  | -0.2145  | 0.0648            | 10.97              | 0.0009     | -0.0334                    | 0.807    |
| WOE_Neg_i_PortabilidadFijo       | 1  | -0.7486  | 0.0247            | 921.98             | <.0001     | -0.2979                    | 0.473    |
| WOE_PS_Num_Rotacion              | 1  | -0.3574  | 0.0526            | 46.16              | <.0001     | -0.0504                    | 0.700    |
| WOE_IMP_REP_CB_Int_DeudaTotal    | 1  | -0.3187  | 0.0302            | 111.58             | <.0001     | -0.1115                    | 0.727    |
| WOE_IMP_REP_CB_Num_Cedentes      | 1  | -3.9261  | 0.6993            | 31.52              | <.0001     | -1.5473                    | 0.020    |
| WOE_IMP_REP_SD_Int_GtoAlqTV      | 1  | -0.4712  | 0.0787            | 35.81              | <.0001     | -0.0507                    | 0.624    |
| WOE_IMP_REP_SD_Int_GtoLibrosText | 1  | -0.2635  | 0.0896            | 8.64               | 0.0033     | -0.0294                    | 0.768    |
| WOE_IMP_REP_SD_Int_GtoRestoLibro | 1  | -0.2621  | 0.0942            | 7.74               | 0.0054     | -0.0264                    | 0.769    |
| WOE_IMP_REP_SD_Int_IngrPensionis | 1  | -0.3813  | 0.0935            | 16.62              | <.0001     | -0.0350                    | 0.683    |
| WOE_IMP_REP_SD_Tasa_HabExtr      | 1  | -0.3688  | 0.0672            | 30.17              | <.0001     | -0.0426                    | 0.692    |
| WOE_Neg_Cod_TipoCanal            | 1  | -0.7940  | 0.0599            | 175.53             | <.0001     | -0.0937                    | 0.452    |
| WOE_PS_Cod_Prov                  | 1  | -0.5050  | 0.0637            | 62.74              | <.0001     | -0.0596                    | 0.604    |



La tarjeta de puntuación preliminar con un número total de 200 puntos se muestra en la tabla siguiente

| Scorecard   |  |                  |
|---|--|------------------|
|   |  | Scorecard Points |
| Indicador presencia Bureau Credito                          | 1  | 247              |
|   | 0, _MISSING_, _UNKNOWN_  | -74              |
| Tipo de Documento   | NIF, _MISSING_, _UNKNOWN_  | 14               |
|   | NIE  | -8               |
|   | PASAPORTE  | -18              |
| Imputed: Mejor Situación (Bureau Credito)                   | BL, _MISSING_, _UNKNOWN_   | 30               |
|   | 03, 08   | -37              |
|   | 01, 02, 04   | -52              |
|   | 99   | -62              |
| Imputed: Peor Situación (Bureau Credito)                    | BL, _MISSING_, _UNKNOWN_   | 13               |
|   | 01, 02, 03, 08   | -1               |
|   | 04, 99   | -4               |
| Imputed: CP distintos en Bureau Credito                     | 1  | -12              |
|   | 0, _MISSING_, _UNKNOWN_  | 16               |
|   | 2, 3, 4  | -22              |
|   | 5  | -46              |
| Imputed: Replacement: Deuda Total                           | IMP_REP_CB_Int_DeudaTotal< 10543, _MISSING_  | 13               |
|   | 10543<= IMP_REP_CB_Int_DeudaTotal< 74891   | -6               |
|   | 74891<= IMP_REP_CB_Int_DeudaTotal  | -8               |
| Imputed: Replacement: Número Cedentes                       | IMP_REP_CB_Num_Cedentes< 1, _MISSING_  | 69               |
|   | 1<= IMP_REP_CB_Num_Cedentes< 2.4   | -161             |
|   | 2.4<= IMP_REP_CB_Num_Cedentes  | -169             |
| Imputed: Replacement: Gastos Alquiler TV                    | IMP_REP_SD_Int_GtoAlqTV< 35.84   | 4                |
|   | 35.84<= IMP_REP_SD_Int_GtoAlqTV< 47.64   | 8                |
|   | 47.64<= IMP_REP_SD_Int_GtoAlqTV< 67.58, _MISSING_  | 8                |
|   | 67.58<= IMP_REP_SD_Int_GtoAlqTV< 103.85  | 10               |
|   | 103.85<= IMP_REP_SD_Int_GtoAlqTV   | 15               |
| Imputed: Replacement: Gastos Libros Texto                   | IMP_REP_SD_Int_GtoLibrosText< 178.48, _MISSING_  | 6                |
|   | 178.48<= IMP_REP_SD_Int_GtoLibrosText< 212.3   | 7                |
|   | 212.3<= IMP_REP_SD_Int_GtoLibrosText< 295.58   | 8                |
|   | 295.58<= IMP_REP_SD_Int_GtoLibrosText< 521.87  | 9                |
|   | 521.87<= IMP_REP_SD_Int_GtoLibrosText< 745.57  | 11               |
|   | 745.57<= IMP_REP_SD_Int_GtoLibrosText  | 14               |
| Imputed: Replacement: Gastos Resto Libros                   | IMP_REP_SD_Int_GtoRestoLibro< 21.95  | 7                |
|   | 21.95<= IMP_REP_SD_Int_GtoRestoLibro< 42.11  | 9                |
|   | 42.11<= IMP_REP_SD_Int_GtoRestoLibro< 45.93, _MISSING_                                       | 8                |
|   | 45.93<= IMP_REP_SD_Int_GtoRestoLibro< 58.81  | 10               |
|   | 58.81<= IMP_REP_SD_Int_GtoRestoLibro   | 12               |
| Imputed: Replacement: Media Ingresos Netos Fam.Pensionistas | IMP_REP_SD_Int_IngrPensionistas< 1697.92   | 6                |
|   | 1697.92<= IMP_REP_SD_Int_IngrPensionistas< 1941.68   | 9                |
|   | 1941.68<= IMP_REP_SD_Int_IngrPensionistas< 2082.86, _MISSING_                                | 8                |
|   | 2082.86<= IMP_REP_SD_Int_IngrPensionistas< 2674.66   | 11               |
|   | 2674.66<= IMP_REP_SD_Int_IngrPensionistas  | 14               |
| Imputed: Replacement: Media Ingresos por Persona Activa     | IMP_REP_SD_Int_IngrXPers< 1150.75  | 6                |
|   | 1150.75<= IMP_REP_SD_Int_IngrXPers< 1231.54  | 8                |
|   | 1231.54<= IMP_REP_SD_Int_IngrXPers< 1417.99  | 9                |
|   | 1417.99<= IMP_REP_SD_Int_IngrXPers< 1455.73, _MISSING_                                       | 10               |
|   | 1455.73<= IMP_REP_SD_Int_IngrXPers< 1604.53  | 11               |
|   | 1604.53<= IMP_REP_SD_Int_IngrXPers   | 13               |
| Imputed: Replacement: Tasa Habitantes Extranjeros           | IMP_REP_SD_Tasa_HabExtr< 0.05  | 11               |
|   | 0.05<= IMP_REP_SD_Tasa_HabExtr< 0.09   | 11               |
|   | 0.09<= IMP_REP_SD_Tasa_HabExtr< 0.16, _MISSING_  | 9                |
|   | 0.16<= IMP_REP_SD_Tasa_HabExtr< 0.22   | 8                |
|   | 0.22<= IMP_REP_SD_Tasa_HabExtr   | 3                |
| Tipo de canal de contratación                               | MOB  | 1                |
|   | CFY, _MISSING_, _UNKNOWN_  | 12               |
| Indicador Portabilidad                                      | 1  | 32               |
|   | 0, _MISSING_, _UNKNOWN_  | 0                |
| Código Provincia Punto Servicio                             | 02, 03, 12, 15, 17, 25, 27, 30, 32, 33, 35, 36, 38, 49                                       | 3                |
|   | 00, 07, 08, 11, 19, 23, 24, 26, 37, 42, 43, 45, 46   |                  |
|   | 04, 05, 09, 13, 14, 16, 18, 21, 22, 28, 29, 31, 34, 39, 40, 41, 47, 60, _MISSING_, _UNKNOWN_ | 12               |
| Rotación punto de servicio                                  | 0, 6, _MISSING_, _UNKNOWN_   | 11               |
|   | 1, 4, 5  | 7                |
|   | 2, 3, 7, 8   | 4                |

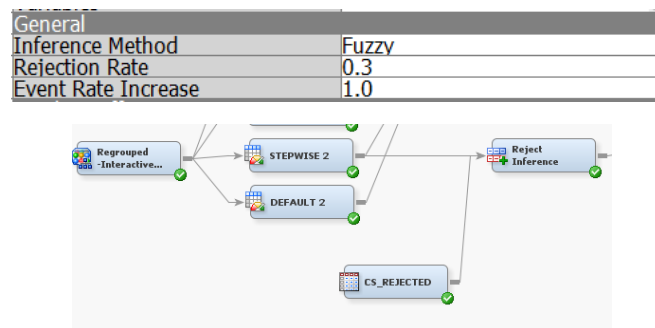




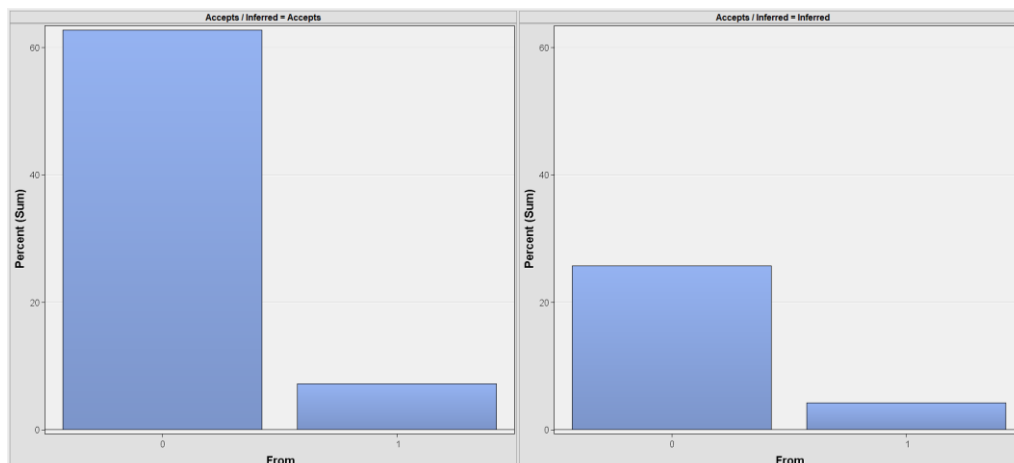
Observamos que las líneas lift acumulada de los datos del entrenamiento y validación son muy parecidas, lo que nos implica un buen modelo.

## 5. INFERENCIA DE DENEGADOS

En este apartado, utilizamos el nodo Inferencia de rechazados con el método Fuzzy, que nos ayuda incluir en el modelo la información sobre los clientes rechazados en el fichero “CS\_Rejected”.



Una vez ejecutamos este nodo, sacamos el siguiente gráfico de los resultados

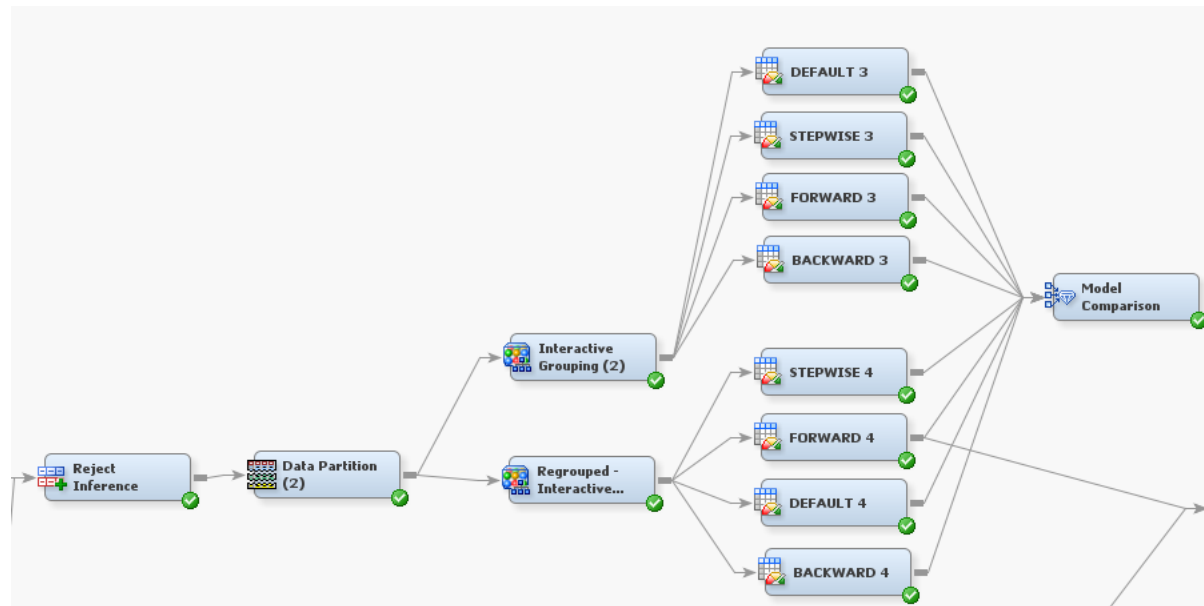


En el gráfico podemos observar el total de la muestra dividida en los dos conjuntos de datos: a la izquierda, los clientes que se le concedió el crédito de la empresa y, a la derecha, observamos como el nodo inferencia de rechazados ha asignado a los clientes que no se le concedió el crédito la probabilidad de

Early Churn Involuntario 12 meses o no. Se puede ver que hay un comportamiento similar en los dos gráficos.

## 6. ESTIMACIÓN DEL MODELO Y EL SCORECARD DEFINITIVO

Para construir el scorecard definitivo repetimos los pasos anteriores de partición de datos, agrupación interactiva y comparación de distintos modelos.



Con el nodo agrupación interactivo conseguimos 12 variables con el poder predictivo según el criterio IV.

| Variable               | Gini Statistic | Information Value | Level for Interactive | Calculated Role | New Role | Pre-Defined Grouping | Level    | Label                  | Information Value Ordering |
|------------------------|----------------|-------------------|-----------------------|-----------------|----------|----------------------|----------|------------------------|----------------------------|
| Neg I PortabilidadPilo | 27.883         | 0.425             | BINARY                | Input           | Default  |                      | BINARY   | Indicador Portabilidad | 1                          |
| IMP REP SD Int In...   | 12.012         | 0.055             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...  | 2                          |
| PS I Rotacion          | 11.649         | 0.055             | BINARY                | Input           | Default  |                      | BINARY   | Indicador Rotacion     | 3                          |
| PS Num Rotacion        | 11.626         | 0.054             | NOMINAL               | Input           | Default  |                      | NOMINAL  | Rotacion punto de s... | 4                          |
| IMP REP SD Int In...   | 9.473          | 0.045             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...  | 5                          |
| IMP REP SD Int In...   | 10.347         | 0.042             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...  | 6                          |
| PS Cod Prov            | 10.115         | 0.036             | NOMINAL               | Input           | Default  |                      | NOMINAL  | Código Provincia Pu... | 7                          |
| IMP REP SD Tasa...     | 8.766          | 0.034             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...  | 8                          |
| IMP REP SD Int G...    | 8.228          | 0.031             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...  | 9                          |
| IMP REP SD Int In...   | 7.677          | 0.025             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...  | 10                         |
| IMP REP SD Int G...    | 7.919          | 0.024             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...  | 11                         |
| IMP REP SD Int G...    | 7.713          | 0.024             | INTERVAL              | Input           | Default  |                      | INTERVAL | Imputed: Replaceme...  | 12                         |

Volvemos a retocar el nodo para mejorar el poder predictivo de las variables cogidas:

| Variable               | Gini Statistic | Information Value | Level for Interactive | New Role | Calculated Role | Pre-Defined Grouping | Measurement Level | Label                  | Information Value Ordering |
|------------------------|----------------|-------------------|-----------------------|----------|-----------------|----------------------|-------------------|------------------------|----------------------------|
| Neg I PortabilidadPilo | 27.883         | 0.425             | BINARY                | Default  | Input           |                      | BINARY            | Indicador Portabilidad | 1                          |
| IMP REP SD Int In...   | 13.461         | 0.071             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...  | 2                          |
| IMP REP SD Int In...   | 12.03          | 0.055             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...  | 4                          |
| PS Num Rotacion        | 11.661         | 0.055             | NOMINAL               | Default  | Input           |                      | NOMINAL           | Rotacion punto de s... | 3                          |
| PS I Rotacion          | 11.649         | 0.055             | BINARY                | Default  | Input           |                      | BINARY            | Indicador Rotacion     | 5                          |
| IMP REP SD Int In...   | 11.268         | 0.05              | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...  | 6                          |
| PS Cod Prov            | 11.399         | 0.05              | NOMINAL               | Default  | Input           |                      | NOMINAL           | Código Provincia Pu... | 7                          |
| IMP REP SD Tasa...     | 12.084         | 0.048             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...  | 8                          |
| IMP REP SD Int G...    | 10.926         | 0.043             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...  | 9                          |
| IMP REP SD Int G...    | 11.164         | 0.043             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...  | 10                         |
| IMP REP SD Int G...    | 9.705          | 0.034             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...  | 11                         |
| IMP REP SD Int G...    | 9.733          | 0.033             | INTERVAL              | Default  | Input           |                      | INTERVAL          | Imputed: Replaceme...  | 12                         |

Con el nodo comparación de modelos, sacamos esta tabla de resultados:

| Modelo     | Entrenar: AIC | Entrenar: SBIC | Entrenar: RMSE | Validar: RMSE | Train: Roc Index | Valid: Roc Index | Train: Gini Coefficient | Valid: Gini Coefficient |
|------------|---------------|----------------|----------------|---------------|------------------|------------------|-------------------------|-------------------------|
| FORWARD 4  | 66121.23388   | 66197.3381     | 0.310703514    | 0.312687295   | 0.692            | 0.692            | 0.385                   | 0.384                   |
| BACKWARD 4 | 66121.23388   | 66197.3381     | 0.310703514    | 0.312687295   | 0.692            | 0.692            | 0.385                   | 0.384                   |
| STEPWISE 4 | 66121.23388   | 66197.3381     | 0.310703514    | 0.312687295   | 0.692            | 0.692            | 0.385                   | 0.384                   |
| DEFAULT 4  | 66127.93872   | 66251.60808    | 0.310704173    | 0.312701335   | 0.692            | 0.692            | 0.385                   | 0.384                   |
| BACKWARD 3 | 66292.38786   | 66378.0051     | 0.31104597     | 0.313035544   | 0.687            | 0.688            | 0.375                   | 0.375                   |
| FORWARD 3  | 66294.11598   | 66370.2202     | 0.311049046    | 0.313034038   | 0.687            | 0.688            | 0.374                   | 0.375                   |
| STEPWISE 3 | 66294.11598   | 66370.2202     | 0.311049046    | 0.313034038   | 0.687            | 0.688            | 0.374                   | 0.375                   |
| DEFAULT 3  | 66298.43783   | 66422.10718    | 0.311047229    | 0.313037856   | 0.687            | 0.688            | 0.375                   | 0.375                   |

⇒ Al comparar los modelos, vemos que el mejor scorecard en esta fase se trata de un modelo “Forward”, “Stepwise” o “Backward” utilizando las variables retocadas manualmente en el paso anterior, puesto que estos 3 modelos tienen la mayor área bajo la curva ROC, mayor Índice de Gini y menor AIC, menor SBIC y menor RMSE. Por lo cual, elegimos el modelo FORWARD 4 para construir el scorecard definitivo.

El scorecard obtenido con el modelo FORWARD 4 se compone de 7 variables significativas como se puede observar en la tabla de resumen:

| Summary of Forward Selection |                                  |    |           |                  |            |                                   |
|------------------------------|----------------------------------|----|-----------|------------------|------------|-----------------------------------|
| Step                         | Effect Entered                   | DF | Number In | Score Chi-Square | Pr > ChiSq | Criterio de información de Akaike |
| 1                            | WOE_Neg_i_PortabilidadFijo       | 1  | 1         | 3343.2393        | <.0001     | 67213.1                           |
| 2                            | WOE_IMP_REP_SD_Int_IngrXPers     | 1  | 2         | 532.6774         | <.0001     | 66656.1                           |
| 3                            | WOE_IMP_REP_SD_Tasa_HabExtr      | 1  | 3         | 189.4770         | <.0001     | 66469.3                           |
| 4                            | WOE_IMP_REP_SD_Int_GtoAlqTV      | 1  | 4         | 159.0073         | <.0001     | 66309.8                           |
| 5                            | WOE_PS_Num_Rotacion              | 1  | 5         | 88.0295          | <.0001     | 66223.6                           |
| 6                            | WOE_PS_Cod_Prov                  | 1  | 6         | 75.7276          | <.0001     | 66150.3                           |
| 7                            | WOE_IMP_REP_SD_Int_IngrPensionis | 1  | 7         | 30.6471          | <.0001     | 66121.2                           |

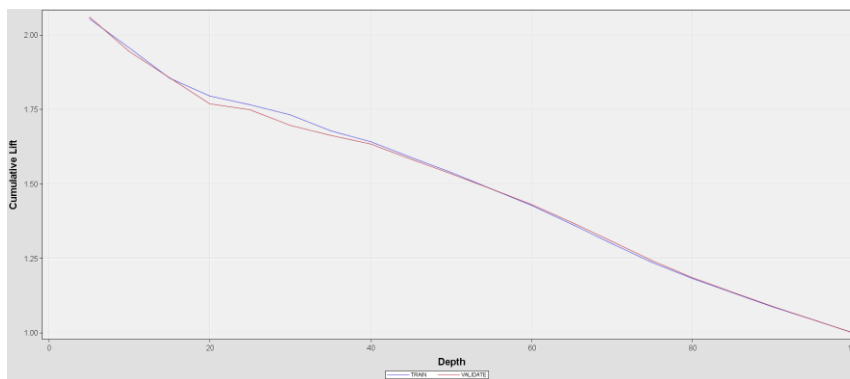
Los parámetros incluidos se detallan a continuación:

| Analysis of Maximum Likelihood Estimates |    |          |                |                 |            |                         |
|--|----|----------|----------------|-----------------|------------|-------------------------|
| Parameter                                | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq | Estimador estandarizado |
| Intercept                                | 1  | -2.0516  | 0.0107         | 36691.07        | <.0001     |                         |
| WOE_Neg_i_PortabilidadFijo               | 1  | -0.9424  | 0.0191         | 2442.60         | <.0001     | -0.3842                 |
| WOE_PS_Num_Rotacion                      | 1  | -0.4023  | 0.0444         | 81.93           | <.0001     | -0.0517                 |
| WOE_IMP_REP_SD_Int_GtoAlqTV              | 1  | -0.4910  | 0.0553         | 78.83           | <.0001     | -0.0574                 |
| WOE_IMP_REP_SD_Int_IngrPensionis         | 1  | -0.3470  | 0.0627         | 30.63           | <.0001     | -0.0361                 |
| WOE_IMP_REP_SD_Int_IngrXPers             | 1  | -0.4967  | 0.0456         | 118.66          | <.0001     | -0.0759                 |
| WOE_IMP_REP_SD_Tasa_HabExtr              | 1  | -0.6846  | 0.0490         | 195.33          | <.0001     | -0.0817                 |
| WOE_PS_Cod_Prov                          | 1  | -0.3852  | 0.0477         | 65.27           | <.0001     | -0.0466                 |

La tarjeta de puntuación definitivo con un número total de 200 puntos se muestra en la tabla siguiente:

## Scorecard

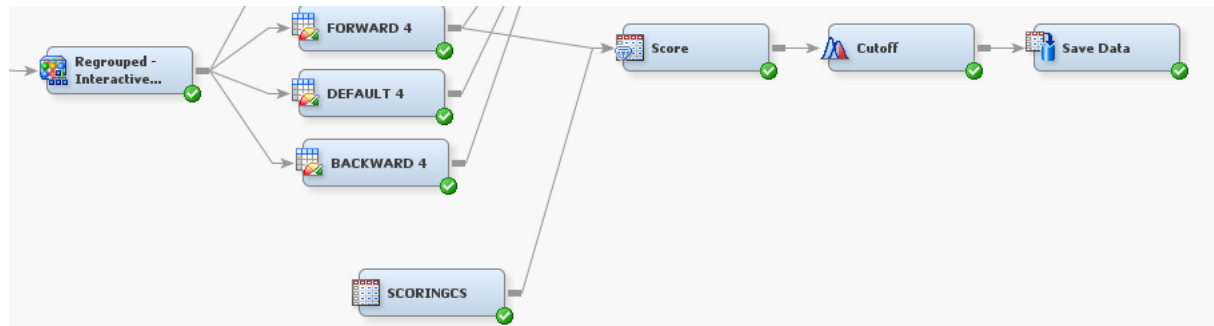
|   |   | Scorecard Points |
|---|---|------------------|
| Imputed: Replacement: Gastos Alquiler TV                    | IMP_REP_SD_Int_GtoAlqTV< 35.45                                | 16               |
|   | 35.45<= IMP_REP_SD_Int_GtoAlqTV< 55.02                        | 21               |
|   | 55.02<= IMP_REP_SD_Int_GtoAlqTV< 56.41                        | 19               |
|   | 56.41<= IMP_REP_SD_Int_GtoAlqTV< 67.23, _MISSING_             | 22               |
|   | 67.23<= IMP_REP_SD_Int_GtoAlqTV< 103.43                       | 23               |
|   | 103.43<= IMP_REP_SD_Int_GtoAlqTV                              | 28               |
| Imputed: Replacement: Media Ingresos Netos Fam.Pensionistas | IMP_REP_SD_Int_IngrPensionistas< 1692.98                      | 19               |
|   | 1692.98<= IMP_REP_SD_Int_IngrPensionistas< 2008.12, _MISSING_ | 21               |
|   | 2008.12<= IMP_REP_SD_Int_IngrPensionistas< 2012.28            | 20               |
|   | 2012.28<= IMP_REP_SD_Int_IngrPensionistas< 2667.98            | 23               |
|   | 2667.98<= IMP_REP_SD_Int_IngrPensionistas                     | 26               |
| Imputed: Replacement: Media Ingresos por Persona Activa     | IMP_REP_SD_Int_IngrXPers< 1147.74                             | 15               |
|   | 1147.74<= IMP_REP_SD_Int_IngrXPers< 1267.06                   | 19               |
|   | 1267.06<= IMP_REP_SD_Int_IngrXPers< 1413.86, _MISSING_        | 20               |
|   | 1413.86<= IMP_REP_SD_Int_IngrXPers< 1601.52                   | 25               |
|   | 1601.52<= IMP_REP_SD_Int_IngrXPers                            | 30               |
| Imputed: Replacement: Tasa Habitantes Extranjeros           | IMP_REP_SD_Tasa_HabExtr< 0.03                                 | 24               |
|   | 0.03<= IMP_REP_SD_Tasa_HabExtr< 0.09                          | 27               |
|   | 0.09<= IMP_REP_SD_Tasa_HabExtr< 0.1, _MISSING_                | 19               |
|   | 0.1<= IMP_REP_SD_Tasa_HabExtr< 0.18                           | 22               |
|   | 0.18<= IMP_REP_SD_Tasa_HabExtr                                | 14               |
| Indicador Portabilidad                                      | 0, _MISSING_, _UNKNOWN_                                       | 10               |
|   | 1   | 52               |
| Código Provincia Punto Servicio                             | 15, 27, 32, 36  | 8                |
|   | 00, 02, 03, 07, 12, 17, 25, 30, 33, 35, 38, 42, 49            | 17               |
|   | 05, 08, 11, 19, 23, 24, 37, 45, 46, _MISSING_, _UNKNOWN_      | 21               |
|   | 04, 09, 14, 18, 21, 26, 28, 29, 39, 41, 43, 47, 50            | 23               |
|   | 13, 16, 22, 31, 34, 40  | 27               |
| Rotación punto de servicio                                  | 0, 11, 8, _MISSING_, _UNKNOWN_                                | 24               |
|   | 1, 2, 3, 4, 5, 6  | 18               |
|   | 14, 7   | 5                |



Observamos que las líneas lift acumulada de los datos del entrenamiento y validación son muy parecidas, lo que nos implica un buen modelo.

## 7. VALIDACIÓN DEL SCORECARD DEFINITIVO Y CONCLUSIÓN.

El último paso antes de integrar la tarjeta de puntuación en producción la empresa sería realizar una prueba ciego para puntuar 66647 clientes en el fichero Scoring\_CS usando el nodo puntuación para sacar la puntuación total de cada cliente y el nodo corte para sacar la predicción del evento “Early Churn Involuntario 12 meses”.



Para el nodo corte, utilizamos el método de Máximo KS Estadístico para elegir el punto de corte automáticamente. Guardamos la tabla de resultados en el formato Excel, que contiene las siguientes columnas creadas:

| Variables        | Descripción  |
|------------------|--|
| SCORECARD_POINTS | La puntuación total del cliente  |
| SCORECARD_BIN    | Es el bin de la scorecard en el que se encuentra cada cliente basado en sus puntos totales.  |
| EM_PROBABILITY   | Probabilidad de impago del cliente   |
| EM_CUTOFF        | La Decisión final que establecerá el criterio: <ul style="list-style-type: none"> <li>▪ '1' - impago, por tanto, no se admite</li> <li>▪ '0' – Buen pagador, solicitud aprobada</li> </ul> |

La tabla de resultados obtenida al final sería como muestra en continuación:

| id_cli      | SCORECARD_POINTS | SCORECARD_BIN | EM_EVENTPROBABILITY | EM_CUTOFF_DECISIÓN |
|-------------|------------------|---------------|---------------------|--------------------|
| 00000280305 | 188              | 21            | 0.0304002           | 0                  |
| 00004810104 | 148              | 11            | 0.11129363          | 1                  |
| 00008190204 | 148              | 11            | 0.11129363          | 1                  |
| 00008280202 | 190              | 22            | 0.0289092           | 0                  |
| 00008960201 | 190              | 22            | 0.0289092           | 0                  |
| 00009290102 | 190              | 22            | 0.0289092           | 0                  |
| 00011050205 | 164              | 15            | 0.066674312         | 0                  |
| 00012030103 | 194              | 23            | 0.0253176           | 0                  |
| 00014430104 | 185              | 21            | 0.0340246           | 0                  |
| 00014650601 | 137              | 9             | 0.151715157         | 1                  |
| 00014930201 | 179              | 19            | 0.040782            | 0                  |

En la tabla, podemos observar en cada fila los clientes a evaluar, con su puntos obtenidos en total según el scorecard, la probabilidad de impago y la decisión final. Los clientes que tienen asignado un uno en la columna EM\_CUTOFF, corresponden a los que tiene alta probabilidad de impago, por lo cual, no se los

admite. Al contrario, aprobamos la solicitud de los clientes marcados con 0 en la columna EM\_CUTOFF\_DECISIÓN, puesto que un valor de 0 nos implica que el cliente es un buen pagador.

Por ejemplo, aprobamos el primer cliente en la tabla tiene el valor de EM\_Cutoff igual que 0 y rechazamos el segundo cliente con el valor de EM\_Cutoff igual que 1.