



Minería de Datos: Aprendizaje no Supervisado y Detección de Anomalías  
2024-2025

MASTER CIENCIA DE DATOS  
UNIVERSIDAD DE GRANADA

# Proyecto Final

MIGUEL GARCÍA LÓPEZ

# Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Sobre el conjunto de datos</b>	<b>2</b>
<b>3. Lectura de datos</b>	<b>3</b>
<b>4. Pre-procesamiento</b>	<b>4</b>
4.1. Eliminación de redundancia . . . . .	4
4.2. Nulos . . . . .	4
4.3. Discretización . . . . .	5
<b>5. Visualización</b>	<b>6</b>
<b>6. Extracción de reglas</b>	<b>7</b>
6.1. Agrupamientos interesantes . . . . .	10
6.2. Análisis de items negados . . . . .	23
<b>7. Conclusión</b>	<b>24</b>

## 1. Introducción

En el presente proyecto final de la asignatura de **Minería de Datos: Aprendizaje no Supervisado y Detección de Anomalías**, se llevará a cabo un análisis de un conjunto de datos por medio de la técnica de extracción de reglas de asociación.

El conjunto de datos elegido es *Market Segmentation in Insurance Unsupervised*. Este ha sido obtenido de la página de *Kaggle*:

<https://www.kaggle.com/datasets/jillanisofttech/market-segmentation-in-insurance-unsupervised>.

## 2. Sobre el conjunto de datos

El dataset contiene un resumen detallado del comportamiento de uso de tarjetas de crédito de aproximadamente 9000 titulares activos durante los últimos seis meses. Está estructurado a nivel de cliente e incluye 17 variables distintas relacionadas con el comportamiento. Las variables son las siguientes:

- **Balance:** El monto total pendiente en la tarjeta de crédito del cliente.
- **Balance Frequency:** La consistencia de las actualizaciones del saldo durante el período observado.
- **Purchases:** El valor monetario total de las compras realizadas con la tarjeta de crédito.
- **One-off Purchases:** El valor total de las compras individuales y no recurrentes.
- **Installment Purchases:** El valor total de las compras realizadas mediante planes de pago a plazos.
- **Cash Advance:** El monto total de efectivo retirado utilizando la tarjeta de crédito.
- **Purchases Frequency:** La frecuencia con la que se realizan compras con la tarjeta de crédito.
- **One-off Purchases Frequency:** La tasa a la que se realizan compras no recurrentes.
- **Purchases Installments Frequency:** La frecuencia de compras realizadas utilizando planes de pago a plazos.
- **Cash Advance Frequency:** La frecuencia de los retiros de efectivo utilizando la tarjeta de crédito.

- **Cash Advance TRX:** El número total de transacciones de adelantos en efectivo realizadas.
- **Purchases TRX:** El número total de transacciones de compra realizadas.
- **Credit Limit:** El crédito máximo disponible para el titular de la tarjeta.
- **Payments:** El monto total de los pagos realizados por el cliente hacia el saldo de su tarjeta de crédito.
- **Minimum Payments:** El monto mínimo de pago requerido por el emisor de la tarjeta de crédito.
- **PRC Full Payment:** La proporción de meses en los que el titular de la tarjeta pagó el saldo completo.
- **Tenure:** El número total de meses que el cliente ha tenido la cuenta de la tarjeta de crédito.

### 3. Lectura de datos

Se leen los datos del *csv* provisto por *Kaggle*. Acto seguido se obtiene un resumen de los datos para poder comprender qué tipo de procesamiento ha de aplicarse al *dataset*.

```
1 data <- read.csv("market_segmentation_insurance_unsupervised.csv")
2 head(data)
3 str(data)
4 dim(data)
5 summary(data)
```

Los datos obtenidos por el *summary* son los siguientes:

Cuadro 1: Key Financial Indicators

Metric	Minimum	Median	Mean	Maximum
Balance	0.0	873.4	1,564.5	19,043.1
Credit Limit	50	3,000	4,494	30,000
Payments	0.0	856.9	1,733.1	50,721.5

Cuadro 2: Purchase and Transaction Statistics

Category	Minimum	Median	Mean	Maximum
Total Purchases	0.00	361.28	1,003.20	49,039.57
One-off Purchases	0.0	38.0	592.4	40,761.2
Installment Purchases	0.0	89.0	411.1	22,500.0
Purchase Transactions	0.00	7.00	14.71	358.00

Cuadro 3: Transaction Frequency Patterns (0-1 scale)

Type	Minimum	Median	Mean	Maximum
Balance Frequency	0.0000	1.0000	0.8773	1.0000
Purchase Frequency	0.0000	0.5000	0.4904	1.0000
One-off Purchase Freq.	0.0000	0.0833	0.2025	1.0000
Installments Freq.	0.0000	0.1667	0.3644	1.0000

## 4. Pre-procesamiento

### 4.1. Eliminación de redundancia

Todas las variables descritas son continuas, por lo que será necesario discretizarlas. Además es necesario eliminar algunas variables que aportan información redundante, *Balance Frequency*, *Purchases Frequency*, *One-off Purchases Frequency*, *Purchases Installments Frequency* y *Cash Advance Frequency* ya que explican lo mismo que las originales, pero midiéndolo en frecuencia. También se debe eliminar el *ID*.

```
1 data$BALANCE_FREQUENCY <- NULL
2 data$PURCHASES_FREQUENCY <- NULL
3 data$ONEOFF_PURCHASES_FREQUENCY <- NULL
4 data$PURCHASES_INSTALLMENTS_FREQUENCY <- NULL
5 data$CASH_ADVANCE_FREQUENCY <- NULL
6 data$CUST_ID <- NULL
```

### 4.2. Nulos

Se tienen dos columnas con nulos, sobre todo en la columna de pagos mínimos. Se comprueba el porcentaje de nulos en esa columna correspondientemente. Se obtiene que solo un 3,5% de los datos totales de esa columna es nulo.

```
1 (sum(is.na(data$MINIMUM_PAYMENTS)) / nrow(data)) * 100
2
3 3.497207
```

No se eliminan los nulos de la columna de pagos mínimos pues es posible que puedan aportar información, ya que es un número elevado. Para ello se codifica como  $-1$ , ya que más tarde en la discretización se les dará el valor de *Unknown* a los valores negativos. El nulo en la columna de límite de crédito se elimina por ser solo uno.

```
1 data <- data %>% filter(!is.na(CREDIT_LIMIT))
2 data$MINIMUM_PAYMENTS[is.na(data$MINIMUM_PAYMENTS)] <- -1
```

### 4.3. Discretización

Se crean tres tipos de funciones de discretización de valores continuos. Aquellos que tienen valores y rangos muy amplios son *continuos high*, valores y rangos medios *continuos* y valores y rangos bajos *continuos low*. Para valores continuos y con valores negativos (−1) como en el caso de pago mínimo, se añade la categoría de desconocido, es decir, los antiguos nulos. El resto son por período (meses).

```

1 discretize_continuous <- function(variable) {
2   cut(
3     variable,
4     breaks = c(-Inf, 0, 500, 1000, 5000, Inf),
5     labels = c("Unknown", "Low", "Medium", "High", "Very High"),
6     right = FALSE
7   )
8 }
9
10 discretize_continuous_low <- function(variable) {
11   cut(
12     variable,
13     breaks = c(-Inf, 0, 5, 15, Inf),
14     labels = c("Unknown", "Low", "Medium", "High"),
15     right = FALSE
16   )
17 }
18
19 discretize_continuous_high <- function(variable) {
20   cut(
21     variable,
22     breaks = c(-Inf, 2000, 5000, 10000, Inf),
23     labels = c("Low", "Medium", "High", "Very High"),
24     right = FALSE
25   )
26 }
27
28 discretize_periods <- function(variable) {
29   cut(
30     variable,
31     breaks = c(-Inf, 6, 12, Inf),
32     labels = c("Short-term", "Medium-term", "Long-term"),
33     right = FALSE
34   )
35 }
36
37 discretize_frequency <- function(variable) {
38   cut(
39     variable,
40     breaks = c(-Inf, 0, 500, 1000, 5000, Inf),
41     labels = c("Unknown", "Low", "Medium", "High", "Very High"),
42     right = FALSE
43   )
44 }

```

Por último, se transforma el *dataset* en un conjunto de transacciones. De esta forma será posible extraer las reglas.

```
1 data_trans <- as(data, "transactions")
```

## 5. Visualización

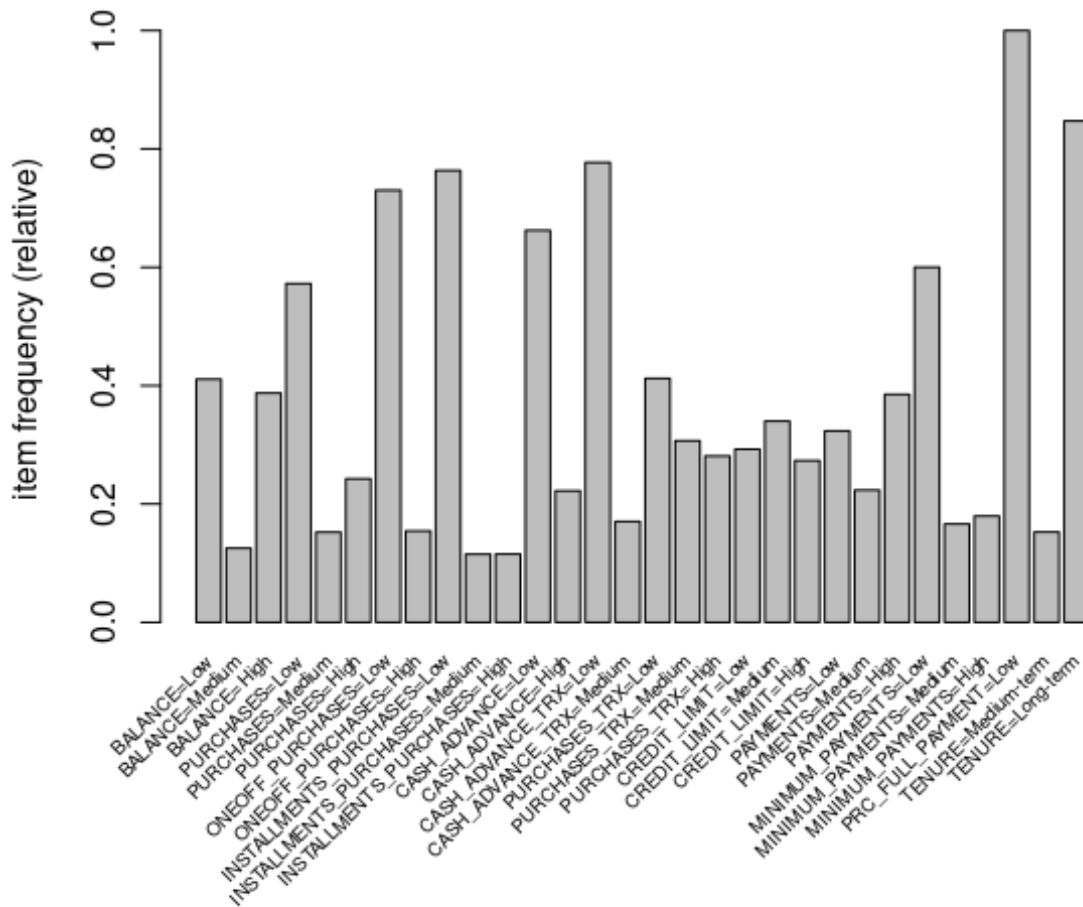


Figura 1: *Items* más frecuentes.

A continuación, se presenta un gráfico (fig 1) que muestra las frecuencias de los *items* en el *dataset*. Los *items* que destacan por su mayor frecuencia son:

- **tenure\_long\_term**: Indica que la mayoría de las observaciones en el *dataset* corresponden a titulares de tarjetas de crédito con un periodo de tenencia prolongado.

- **pcr\_full\_payment\_low**: Sugiere que los clientes tienden a no pagar el total del balance (es una proporción en meses), lo cual representa un mayor riesgo financiero.
- **cash\_advance\_frequency\_low**: Se refiere a una baja frecuencia de uso de la tarjeta de crédito para adelantos en efectivo, lo que implica un menor riesgo debido a las tasas de interés asociadas a estas operaciones.
- **balance\_frequency\_high**: Refleja que los clientes tienden a mantener un saldo en sus tarjetas de forma consistente.

Al extraer las primeras reglas, se obtiene el siguiente gráfico (fig 2) en el cual se observa que los *itemsets* más frecuentes son los de tamaño cuatro y cinco.

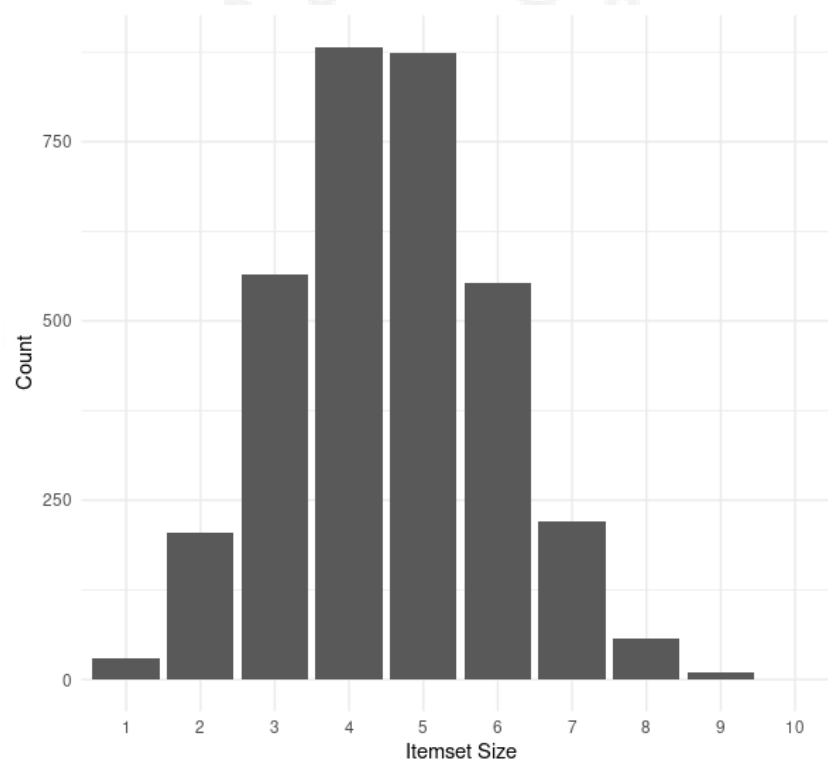


Figura 2: Tamaños de *itemsets* más frecuentes.

## 6. Extracción de reglas

Se procede a la extracción de reglas interesantes según ciertas métricas. Inicialmente se ha priorizado la extracción de reglas con un mínimo de soporte del 0,1 y una confianza de 0,8.

```
1 rules <- apriori(data_trans, parameter = list(
```



```

2  support = 0.1,
3  confidence = 0.8,
4  minlen = 2
5  ))
6
7  inspect(head(sort(rules, by = "lift"), 20))
8
9  lhs                                rhs                                support confidence
10 coverage lift count
11 [1] {ONEOFF_PURCHASES=High,
12     CASH_ADVANCE_TRX=Low} => {PURCHASES=High} 0.1177654 0.9016253
13     0.1306145 3.715261 1054
14 [2] {ONEOFF_PURCHASES=High,
15     CASH_ADVANCE_TRX=Low,
16     PRC_FULL_PAYMENT=Low} => {PURCHASES=High} 0.1177654 0.9016253
17     0.1306145 3.715261 1054
18 [3] {ONEOFF_PURCHASES=High} => {PURCHASES=High} 0.1392179 0.9015919
19     0.1544134 3.715123 1246
20 [4] {ONEOFF_PURCHASES=High,
21     PRC_FULL_PAYMENT=Low} => {PURCHASES=High} 0.1392179 0.9015919
22     0.1544134 3.715123 1246
23 [5] {ONEOFF_PURCHASES=High,
24     CASH_ADVANCE=Low} => {PURCHASES=High} 0.1062570 0.9005682
25     0.1179888 3.710905 951
26 [6] {ONEOFF_PURCHASES=High,
27     CASH_ADVANCE=Low,
28     PRC_FULL_PAYMENT=Low} => {PURCHASES=High} 0.1062570 0.9005682
29     0.1179888 3.710905 951
30 [7] {ONEOFF_PURCHASES=High,
31     CASH_ADVANCE=Low,
32     CASH_ADVANCE_TRX=Low} => {PURCHASES=High} 0.1049162 0.8994253
33     0.1166480 3.706195 939
34 [8] {ONEOFF_PURCHASES=High,
35     CASH_ADVANCE=Low,
36     CASH_ADVANCE_TRX=Low,
37     PRC_FULL_PAYMENT=Low} => {PURCHASES=High} 0.1049162 0.8994253
38     0.1166480 3.706195 939
39 [9] {ONEOFF_PURCHASES=High,
40     TENURE=Long-term} => {PURCHASES=High} 0.1246927 0.8928000
41     0.1396648 3.678895 1116
42 [10] {ONEOFF_PURCHASES=High,
43     PRC_FULL_PAYMENT=Low,
44     TENURE=Long-term} => {PURCHASES=High} 0.1246927 0.8928000
45     0.1396648 3.678895 1116
46 [11] {ONEOFF_PURCHASES=High,
47     CASH_ADVANCE_TRX=Low,
48     TENURE=Long-term} => {PURCHASES=High} 0.1062570 0.8921201
49     0.1191061 3.676093 951
50 [12] {ONEOFF_PURCHASES=High,
51     CASH_ADVANCE_TRX=Low,
52     PRC_FULL_PAYMENT=Low,
53     TENURE=Long-term} => {PURCHASES=High} 0.1062570 0.8921201
54     0.1191061 3.676093 951
55 [13] {CASH_ADVANCE=Low,

```

```

43     CASH_ADVANCE_TRX=Low,
44     PURCHASES_TRX=High,
45     PAYMENTS=High}      => {PURCHASES=High} 0.1017877 0.8251812
      0.1233520 3.400263 911
46 [14] {CASH_ADVANCE=Low,
47       CASH_ADVANCE_TRX=Low,
48       PURCHASES_TRX=High,
49       PAYMENTS=High,
50       PRC_FULL_PAYMENT=Low} => {PURCHASES=High} 0.1017877 0.8251812
      0.1233520 3.400263 911
51 [15] {CASH_ADVANCE=Low,
52       PURCHASES_TRX=High,
53       PAYMENTS=High}      => {PURCHASES=High} 0.1033520 0.8207631
      0.1259218 3.382058 925
54 [16] {CASH_ADVANCE=Low,
55       PURCHASES_TRX=High,
56       PAYMENTS=High,
57       PRC_FULL_PAYMENT=Low} => {PURCHASES=High} 0.1033520 0.8207631
      0.1259218 3.382058 925
58 [17] {CASH_ADVANCE_TRX=Low,
59       PURCHASES_TRX=High,
60       PAYMENTS=High}      => {PURCHASES=High} 0.1126257 0.8089888
      0.1392179 3.333540 1008
61 [18] {CASH_ADVANCE_TRX=Low,
62       PURCHASES_TRX=High,
63       PAYMENTS=High,
64       PRC_FULL_PAYMENT=Low} => {PURCHASES=High} 0.1126257 0.8089888
      0.1392179 3.333540 1008
65 [19] {CASH_ADVANCE_TRX=Low,
66       PURCHASES_TRX=High,
67       PAYMENTS=High,
68       TENURE=Long-term}    => {PURCHASES=High} 0.1081564 0.8086884
      0.1337430 3.332303 968
69 [20] {CASH_ADVANCE_TRX=Low,
70       PURCHASES_TRX=High,
71       PAYMENTS=High,
72       PRC_FULL_PAYMENT=Low,
73       TENURE=Long-term}    => {PURCHASES=High} 0.1081564 0.8086884
      0.1337430 3.332303 968

```

Se observa que todas las reglas tienen como consecuente `purchases_high`, es decir, un nivel de gasto con la tarjeta de crédito muy elevado.

La primera regla, que presenta el valor de *lift* más alto, es la siguiente:

$$\{\text{ONEOFF\_PURCHASES} = \text{High}, \text{PRC\_FULL\_PAYMENT} = \text{Low}\} \Rightarrow \{\text{PURCHASES} = \text{High}\}$$

Esta regla indica que, si un cliente realiza compras únicas (`ONEOFF_PURCHASES`) de alto valor y no suele efectuar pagos completos sobre el balance (`PRC_FULL_PAYMENT = Low`), es probable que también registre un monto elevado en compras generales (`PURCHASES = High`).

Sin embargo, esta regla no resulta particularmente interesante, ya que si el cliente tiene un número significativo de compras no recurrentes de alto valor, es lógico que el total

de compras también sea elevado. Por esta razón, se considera necesario explorar reglas adicionales que puedan ofrecer información más novedosa y menos evidente.

## 6.1. Agrupamientos interesantes

Se filtra para eliminar compras esporádicas altas y pagos altos, ya que suelen estar también muy relacionados, aunque no sean exactamente lo mismo.

```

1 rules_sorted <- head(sort(subset(
2   rules,
3   subset = !(lhs %in% "ONEOFF_PURCHASES=High") &
4     !(lhs %in% "PAYMENTS=High") &
5     lift > 1.5
6 ), by = "lift"), 20)
7
8 redundant <- is.redundant(x = rules_sorted, measure = "confidence")
9 rules_pruned <- rules_sorted[!redundant]
10 inspect(rules_pruned)
11
12 lhs                                rhs                                support
13 confidence coverage lift count
14 [1] {INSTALLMENTS_PURCHASES=Low,
15     MINIMUM_PAYMENTS=Medium,
16     TENURE=Long-term}              => {BALANCE=High}              0.1030168
17 0.9155909 0.1125140 2.362900      922
18 [2] {MINIMUM_PAYMENTS=Medium,
19     TENURE=Long-term}              => {BALANCE=High}              0.1326257
20 0.9151889 0.1449162 2.361863      1187
21 [3] {INSTALLMENTS_PURCHASES=Low,
22     MINIMUM_PAYMENTS=Medium}      => {BALANCE=High}              0.1188827
23 0.9032258 0.1316201 2.330989      1064
24 [4] {MINIMUM_PAYMENTS=Medium}      => {BALANCE=High}              0.1499441
25 0.9024882 0.1661453 2.329086      1342
26 [5] {PURCHASES=High,
27     CASH_ADVANCE_TRX=Low,
28     MINIMUM_PAYMENTS=Low,
29     TENURE=Long-term}              => {PAYMENTS=High}              0.1015642
30 0.8116071 0.1251397 2.106695      909
31 [6] {ONEOFF_PURCHASES=Low,
32     CASH_ADVANCE=Low,
33     CASH_ADVANCE_TRX=Low,
34     PURCHASES_TRX=Medium,
35     MINIMUM_PAYMENTS=Low}          => {BALANCE=Low}              0.1278212
36 0.8653555 0.1477095 2.105745      1144
37 [7] {PURCHASES=High,
38     MINIMUM_PAYMENTS=Low,
39     TENURE=Long-term}              => {PAYMENTS=High}              0.1063687
40 0.8095238 0.1313966 2.101287      952
41 [8] {PURCHASES=Low,
42     CASH_ADVANCE=High,
43     TENURE=Long-term}              => {PURCHASES_TRX=Low}          0.1056983
44 0.8655078 0.1221229 2.099836      946

```

La regla con mayor *lift* es:

$$\{\text{INSTALLMENTS\_PURCHASES} = \text{Low}, \text{MINIMUM\_PAYMENTS} = \text{Medium}, \\ \text{TENURE} = \text{Long-term}\} \Rightarrow \{\text{BALANCE} = \text{High}\}$$

Si un cliente tiene un monto total de pagos en plazos bajo, con un pago mínimo medio y una posesión de la tarjeta de crédito muy larga, es probable que el balance de su tarjeta sea alto. Un balance alto se refiere a una cantidad elevada de deuda o saldo pendiente. Pagos mínimos medianos sugieren que, aunque se están haciendo pagos, estos no son lo suficientemente grandes como para reducir el saldo rápidamente. Esto permite que los intereses se acumulen y el saldo permanezca alto. Un plazo largo significa que la persona está tomando tiempo para pagar la deuda, lo que contribuye a que el saldo permanezca alto durante más tiempo y siga acumulando intereses. Pese a que el valor total de compras financiadas a plazos sea bajo o nulo, es posible que este tipo de personas que prefieren realizar con poca frecuencia este tipo de compras, tiendan a acumular más deuda.

De hecho, la regla:

$$\{\text{INSTALLMENTS\_PURCHASES} = \text{Low}, \text{MINIMUM\_PAYMENTS} = \text{Medium}\} \Rightarrow \{\text{BALANCE} = \text{High}\}$$

Indica que solo con esas dos variables se suele tener un valor alto de balance. La regla:

$$\{\text{MINIMUM\_PAYMENTS} = \text{Medium}\} \Rightarrow \{\text{BALANCE} = \text{High}\}$$

Indica que solo con tener pagos mínimos de valor medio se tiene un balance alto. Esto quizá indica que esta variable tiene mucho peso sobre la variable de balance, por ello reglas con más *items* no empeoran el valor de *lift* y confianza.

Se analizan las reglas que solo tienen en la parte derecha un balance bajo.

```

1 rules_sorted <- head(sort(subset(
2   rules,
3   subset = !(lhs %in% "ONEOFF_PURCHASES=High") &
4     !(lhs %in% "PAYMENTS=High") &
5     (rhs %in% "BALANCE=Low") &
6     lift > 1.5
7 ), by = "lift"), 20)
8
9 redundant <- is.redundant(x = rules_sorted, measure = "confidence")
10 rules_pruned <- rules_sorted[!redundant]
11 inspect(rules_pruned)
12
13 lhs                                rhs                                support confidence coverage
14 lift count
15 [1] {ONEOFF_PURCHASES=Low,
16     CASH_ADVANCE=Low,
17     CASH_ADVANCE_TRX=Low,
18     PURCHASES_TRX=Medium,
19     MINIMUM_PAYMENTS=Low} => {BALANCE=Low} 0.1278212 0.8653555
20     0.1477095 2.105745 1144
21 [2] {ONEOFF_PURCHASES=Low,
22     CASH_ADVANCE=Low,

```

```

21 PURCHASES_TRX=Medium,
22 MINIMUM_PAYMENTS=Low} => {BALANCE=Low} 0.1286034 0.8621723
0.1491620 2.097999 1151
23 [3] {ONEOFF_PURCHASES=Low,
24 CASH_ADVANCE_TRX=Low,
25 PURCHASES_TRX=Medium,
26 MINIMUM_PAYMENTS=Low} => {BALANCE=Low} 0.1299441 0.8366906
0.1553073 2.035993 1163
27 [4] {CASH_ADVANCE=Low,
28 CASH_ADVANCE_TRX=Low,
29 PURCHASES_TRX=Medium,
30 MINIMUM_PAYMENTS=Low} => {BALANCE=Low} 0.1524022 0.8266667
0.1843575 2.011601 1364
31 [5] {CASH_ADVANCE=Low,
32 PURCHASES_TRX=Medium,
33 MINIMUM_PAYMENTS=Low} => {BALANCE=Low} 0.1531844 0.8219424
0.1863687 2.000105 1371

```

La regla con menos *items*, ya que el resto de reglas la contienen y tienen valores parecidos de confianza y *lift*, es:

$$\{\text{CASH\_ADVANCE} = \text{Low}, \text{PURCHASES\_TRX} = \text{Medium}, \text{MINIMUM\_PAYMENTS} = \text{Low}\} \Rightarrow \{\text{BALANCE} = \text{Low}\}$$

Esta regla relaciona una cantidad total de retirada de dinero físico de la tarjeta de crédito baja, una cantidad de compras por transacciones bajas y un pago mínimo bajo con un balance más sano. Esto hace ver que la variable de pagos mínimos no es tan importante como anteriormente se analizó, ya que en combinación con ciertos hábitos con la tarjeta de crédito, es posible disminuir esta deuda incluso pagando un mínimo bajo.

Se analizan los consecuentes con pagos mínimos bajos. Lo que se encuentra es que normalmente si se tiene un balance (deuda) bajo, los pagos mínimos son muy bajos.

```

1 rules_sorted <- head(sort(subset(
2   rules, subset = (rhs %in% "MINIMUM_PAYMENTS=Low") &
3     lift > 1.5
4 ), by = "lift"), 20)
5
6 redundant <- is.redundant(x = rules_sorted, measure = "confidence")
7 rules_pruned <- rules_sorted[!redundant]
8 inspect(rules_pruned)
9
10 lhs                                rhs                                support
11   confidence coverage lift count
12 [1] {BALANCE=Low,
13     CASH_ADVANCE_TRX=Low,
14     PURCHASES_TRX=High} => {MINIMUM_PAYMENTS=Low} 0.1004469
15   0.9771739 0.1027933 1.628017 899
16 [2] {BALANCE=Low,
17     PURCHASES_TRX=High} => {MINIMUM_PAYMENTS=Low} 0.1025698
18   0.9755579 0.1051397 1.625325 918
19 [3] {BALANCE=Low,
20     CASH_ADVANCE_TRX=Low,

```

```

18     PAYMENTS=High,
19     TENURE=Long-term}      => {MINIMUM_PAYMENTS=Low} 0.1071508
    0.9609218 0.1115084 1.600940 959
20 [4] {BALANCE=Low,
21     CASH_ADVANCE_TRX=Low,
22     PAYMENTS=High}      => {MINIMUM_PAYMENTS=Low} 0.1166480
    0.9604416 0.1214525 1.600140 1044
23 [5] {BALANCE=Low,
24     PAYMENTS=High}      => {MINIMUM_PAYMENTS=Low} 0.1234637
    0.9592014 0.1287151 1.598074 1105
25 [6] {BALANCE=Low,
26     INSTALLMENTS_PURCHASES=Low,
27     CASH_ADVANCE_TRX=Low,
28     PURCHASES_TRX=Medium} => {MINIMUM_PAYMENTS=Low} 0.1201117
    0.9331597 0.1287151 1.554687 1075
29 [7] {BALANCE=Low,
30     INSTALLMENTS_PURCHASES=Low,
31     PURCHASES_TRX=Medium} => {MINIMUM_PAYMENTS=Low} 0.1230168
    0.9322608 0.1319553 1.553190 1101

```

Se procede a analizar clientes que usen la tarjeta de crédito para retirar grandes cantidades de dinero en metálico.

```

1 rules_cash_advance <- apriori(
2   data_trans,
3   parameter = list(
4     support = 0.1,
5     confidence = 0.7,
6     minlen = 2
7   ),
8   appearance = list(lhs = "CASH_ADVANCE=High")
9 )
10
11 rules_cash_advance_sorted <- sort(rules_cash_advance, by = "lift")
12 inspect(head(rules_cash_advance_sorted))
13
14 lhs               rhs               support
15 confidence coverage lift      count
16 [1] {CASH_ADVANCE=High} => {PURCHASES=Low}      0.1597765
    0.7189542 0.2222346 1.2557846 1430
17 [2] {CASH_ADVANCE=High} => {INSTALLMENTS_PURCHASES=Low} 0.1912849
    0.8607340 0.2222346 1.1274066 1712
18 [3] {CASH_ADVANCE=High} => {ONEOFF_PURCHASES=Low}      0.1814525
    0.8164907 0.2222346 1.1180526 1624
19 [4] {CASH_ADVANCE=High} => {PRC_FULL_PAYMENT=Low}      0.2222346
    1.0000000 0.2222346 1.0000000 1989
20 [5] {CASH_ADVANCE=High} => {TENURE=Long-term}      0.1765363
    0.7943690 0.2222346 0.9374476 1580

```

Se obtienen reglas como:

$$\{\text{CASH\_ADVANCE} = \text{High}\} \Rightarrow \{\text{PURCHASES} = \text{Low}\}$$

Con una confianza del 70 % como mínimo y un lift de 1,25, que si los clientes que retiran

grandes cantidades de dinero suelen gastar poco usando la tarjeta de crédito, quizás por qué prefieren el metálico. Ocurre lo mismo con las compras a plazos usando la tarjeta.

Por último, se analizan las reglas que tengan pagos mínimos con valor desconocido. Inicialmente se codificaron los valores nulos de esa variable como valores *Unknown* para poder analizar si tienen un sentido o son valores faltantes aleatorios, que expresado de otra forma es, ver si los datos nulos son *missing at random* o *missing completely at random*.

```

1 rules_minimum_unknown <- apriori(
2   data_trans,
3   parameter = list(
4     support = 0.0001,
5     confidence = 0.7,
6     minlen = 2
7   ),
8   appearance = list(lhs = "MINIMUM_PAYMENTS=Unknown")
9 )
10
11 rules_rules_minimum_unknown_sorted <- sort(rules_minimum_unknown, by =
12   "lift")
13 inspect(head(rules_rules_minimum_unknown_sorted, 10))
14
15 lhs               rhs               support
16 confidence coverage lift count
17 [1] {MINIMUM_PAYMENTS=Unknown} => {PAYMENTS=Low}
18 0.03173539 0.9073482 0.03497597 2.805757 284
19 [2] {MINIMUM_PAYMENTS=Unknown} => {BALANCE=Low}
20 0.02771259 0.7923323 0.03497597 1.928360 248
21 [3] {MINIMUM_PAYMENTS=Unknown} => {PURCHASES_TRX=Low}
22 0.02223712 0.6357827 0.03497597 1.542739 199
23 [4] {MINIMUM_PAYMENTS=Unknown} => {PURCHASES=Low}
24 0.02748910 0.7859425 0.03497597 1.372906 246
25 [5] {MINIMUM_PAYMENTS=Unknown} => {CREDIT_LIMIT=Medium}
26 0.01564421 0.4472843 0.03497597 1.315395 140
27 [6] {MINIMUM_PAYMENTS=Unknown} => {INSTALLMENTS_PURCHASES=Low}
28 0.03207062 0.9169329 0.03497597 1.201059 287
29 [7] {MINIMUM_PAYMENTS=Unknown} => {ONEOFF_PURCHASES=Low}
30 0.02994748 0.8562300 0.03497597 1.172518 268
31 [8] {MINIMUM_PAYMENTS=Unknown} => {CASH_ADVANCE_TRX=Low}
32 0.03117667 0.8913738 0.03497597 1.147096 279
33 [9] {MINIMUM_PAYMENTS=Unknown} => {CASH_ADVANCE=Low}
34 0.02625992 0.7507987 0.03497597 1.133609 235
35 [10] {MINIMUM_PAYMENTS=Unknown} => {CREDIT_LIMIT=Low}
36 0.01050397 0.3003195 0.03497597 1.026178 94

```

Cuando los pagos mínimos son desconocidos, los pagos realizados con la tarjeta de crédito tienden a ser bajos. Aunque esta situación es poco común, cuando se presenta, la relación entre las premisas y la conclusión es sólida, como lo demuestra el elevado valor de *lift*, acompañado de una confianza del 90%.

Además, el conjunto restante de reglas obtenidas sugiere que, en todos los casos de valores faltantes, se asocian consistentemente con valores bajos en las variables analizadas. Esto podría indicar que los clientes en cuestión no utilizan su tarjeta de crédito, lo que



justificaría tanto los valores bajos como la presencia de datos faltantes, o posibles comportamientos anómalos.

Analizamos las reglas en las que los pagos mínimos faltantes aparezcan en la derecha como consecuente y filtramos por aquellas reglas que tengan más de 10 ocurrencias, pues la mayoría de reglas que aparecen solo tienen una ocurrencia, lo cual no desvela ningún patrón.

```

1 rules_minimum_unknown <- apriori(
2   data_trans,
3   parameter = list(
4     support = 0.0001,
5     confidence = 0.3,
6     minlen = 2
7   ),
8   appearance = list(rhs = "MINIMUM_PAYMENTS=Unknown")
9 )
10 total_transactions <- length(data_trans)
11
12 rules_minimum_unknown_filtered <- subset(
13   rules_minimum_unknown,
14   subset = (support * total_transactions) >= 10
15 )
16
17 rules_minimum_unknown_sorted <- sort(rules_minimum_unknown_filtered, by
18   = "lift")
19 inspect(head(rules_minimum_unknown_sorted, 50))
20 lhs                                rhs                                support
21 confidence    coverage    lift count
22 [1] {BALANCE=Low,
23     CASH_ADVANCE=High,
24     PURCHASES_TRX=Low,
25     CREDIT_LIMIT=Medium,
26     PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
27 0.001564421  0.6086957  0.002570120  17.40325    14
28 [2] {BALANCE=Low,
29     PURCHASES=Low,
30     CASH_ADVANCE=High,
31     PURCHASES_TRX=Low,
32     CREDIT_LIMIT=Medium,
33     PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
34 0.001564421  0.6086957  0.002570120  17.40325    14
35 [3] {BALANCE=Low,
36     ONEOFF_PURCHASES=Low,
37     CASH_ADVANCE=High,
38     PURCHASES_TRX=Low,
39     CREDIT_LIMIT=Medium,
40     PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
41 0.001564421  0.6086957  0.002570120  17.40325    14
42 [4] {BALANCE=Low,
43     INSTALLMENTS_PURCHASES=Low,
44     CASH_ADVANCE=High,
45     PURCHASES_TRX=Low,

```



```

42     CREDIT_LIMIT=Medium,
43     PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.6086957 0.002570120 17.40325 14
44 [5] {BALANCE=Low,
45     CASH_ADVANCE=High,
46     PURCHASES_TRX=Low,
47     CREDIT_LIMIT=Medium,
48     PAYMENTS=Low,
49     PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.6086957 0.002570120 17.40325 14
50 [6] {BALANCE=Low,
51     PURCHASES=Low,
52     ONEOFF_PURCHASES=Low,
53     CASH_ADVANCE=High,
54     PURCHASES_TRX=Low,
55     CREDIT_LIMIT=Medium,
56     PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.6086957 0.002570120 17.40325 14
57 [7] {BALANCE=Low,
58     PURCHASES=Low,
59     INSTALLMENTS_PURCHASES=Low,
60     CASH_ADVANCE=High,
61     PURCHASES_TRX=Low,
62     CREDIT_LIMIT=Medium,
63     PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.6086957 0.002570120 17.40325 14
64 [8] {BALANCE=Low,
65     PURCHASES=Low,
66     CASH_ADVANCE=High,
67     PURCHASES_TRX=Low,
68     CREDIT_LIMIT=Medium,
69     PAYMENTS=Low,
70     PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.6086957 0.002570120 17.40325 14
71 [9] {BALANCE=Low,
72     ONEOFF_PURCHASES=Low,
73     INSTALLMENTS_PURCHASES=Low,
74     CASH_ADVANCE=High,
75     PURCHASES_TRX=Low,
76     CREDIT_LIMIT=Medium,
77     PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.6086957 0.002570120 17.40325 14
78 [10] {BALANCE=Low,
79     ONEOFF_PURCHASES=Low,
80     CASH_ADVANCE=High,
81     PURCHASES_TRX=Low,
82     CREDIT_LIMIT=Medium,
83     PAYMENTS=Low,
84     PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.6086957 0.002570120 17.40325 14
85 [11] {BALANCE=Low,
86     INSTALLMENTS_PURCHASES=Low,
87     CASH_ADVANCE=High,
88     PURCHASES_TRX=Low,

```

```

89     CREDIT_LIMIT=Medium,
90     PAYMENTS=Low,
91     PRC_FULL_PAYMENT=Low}      => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.6086957  0.002570120  17.40325      14
92 [12] {BALANCE=Low,
93       PURCHASES=Low,
94       ONEOFF_PURCHASES=Low,
95       INSTALLMENTS_PURCHASES=Low,
96       CASH_ADVANCE=High,
97       PURCHASES_TRX=Low,
98       CREDIT_LIMIT=Medium,
99       PAYMENTS=Low}      => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.6086957  0.002570120  17.40325      14
100 [13] {BALANCE=Low,
101       PURCHASES=Low,
102       ONEOFF_PURCHASES=Low,
103       CASH_ADVANCE=High,
104       PURCHASES_TRX=Low,
105       CREDIT_LIMIT=Medium,
106       PAYMENTS=Low,
107       PRC_FULL_PAYMENT=Low}      => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.6086957  0.002570120  17.40325      14
108 [14] {BALANCE=Low,
109       PURCHASES=Low,
110       INSTALLMENTS_PURCHASES=Low,
111       CASH_ADVANCE=High,
112       PURCHASES_TRX=Low,
113       CREDIT_LIMIT=Medium,
114       PAYMENTS=Low,
115       PRC_FULL_PAYMENT=Low}      => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.6086957  0.002570120  17.40325      14
116 [15] {BALANCE=Low,
117       ONEOFF_PURCHASES=Low,
118       INSTALLMENTS_PURCHASES=Low,
119       CASH_ADVANCE=High,
120       PURCHASES_TRX=Low,
121       CREDIT_LIMIT=Medium,
122       PAYMENTS=Low,
123       PRC_FULL_PAYMENT=Low}      => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.6086957  0.002570120  17.40325      14
124 [16] {BALANCE=Low,
125       PURCHASES=Low,
126       ONEOFF_PURCHASES=Low,
127       INSTALLMENTS_PURCHASES=Low,
128       CASH_ADVANCE=High,
129       PURCHASES_TRX=Low,
130       CREDIT_LIMIT=Medium,
131       PAYMENTS=Low,
132       PRC_FULL_PAYMENT=Low}      => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.6086957  0.002570120  17.40325      14
133 [17] {PURCHASES=High,
134       ONEOFF_PURCHASES=High,
135       CREDIT_LIMIT=High,
136       PAYMENTS=Low,

```

```

137     TENURE=Long-term} => {MINIMUM_PAYMENTS=Unknown}
138     0.001340932 0.6000000 0.002234887 17.15463 12
139 [18] {PURCHASES=High,
140     ONEOFF_PURCHASES=High,
141     CREDIT_LIMIT=High,
142     PAYMENTS=Low,
143     PRC_FULL_PAYMENT=Low,
144     TENURE=Long-term} => {MINIMUM_PAYMENTS=Unknown}
145     0.001340932 0.6000000 0.002234887 17.15463 12
146 [19] {ONEOFF_PURCHASES=High,
147     CREDIT_LIMIT=High,
148     PAYMENTS=Low,
149     TENURE=Long-term} => {MINIMUM_PAYMENTS=Unknown}
150     0.001452676 0.5909091 0.002458375 16.89471 13
151 [20] {ONEOFF_PURCHASES=High,
152     CREDIT_LIMIT=High,
153     PAYMENTS=Low,
154     PRC_FULL_PAYMENT=Low,
155     TENURE=Long-term} => {MINIMUM_PAYMENTS=Unknown}
156     0.001452676 0.5909091 0.002458375 16.89471 13
157 [21] {PURCHASES=High,
158     ONEOFF_PURCHASES=High,
159     INSTALLMENTS_PURCHASES=Low,
160     CREDIT_LIMIT=High,
161     PAYMENTS=Low,
162     TENURE=Long-term} => {MINIMUM_PAYMENTS=Unknown}
163     0.001117443 0.5882353 0.001899654 16.81827 10
164 [22] {PURCHASES=High,
165     ONEOFF_PURCHASES=High,
166     INSTALLMENTS_PURCHASES=Low,
167     CREDIT_LIMIT=High,
168     PAYMENTS=Low,
169     PRC_FULL_PAYMENT=Low,
170     TENURE=Long-term} => {MINIMUM_PAYMENTS=Unknown}
171     0.001117443 0.5882353 0.001899654 16.81827 10
172 [23] {PURCHASES=High,
173     ONEOFF_PURCHASES=High,
174     CREDIT_LIMIT=High,
175     PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
176     0.001452676 0.5652174 0.002570120 16.16016 13
177 [24] {PURCHASES=High,
178     ONEOFF_PURCHASES=High,
179     CREDIT_LIMIT=High,
180     PAYMENTS=Low,
181     PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
182     0.001452676 0.5652174 0.002570120 16.16016 13
183 [25] {ONEOFF_PURCHASES=High,
184     CREDIT_LIMIT=High,
185     PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
186     0.001564421 0.5600000 0.002793608 16.01099 14
187 [26] {ONEOFF_PURCHASES=High,
188     CREDIT_LIMIT=High,
189     PAYMENTS=Low,
190     PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}

```

```

0.001564421 0.5600000 0.002793608 16.01099 14
182 [27] {ONEOFF_PURCHASES=High,
183       INSTALLMENTS_PURCHASES=Low,
184       CREDIT_LIMIT=High,
185       PAYMENTS=Low,
186       TENURE=Long-term} => {MINIMUM_PAYMENTS=Unknown}
0.001117443 0.5555556 0.002011398 15.88392 10
187 [28] {PURCHASES=High,
188       ONEOFF_PURCHASES=High,
189       CASH_ADVANCE_TRX=Low,
190       CREDIT_LIMIT=High,
191       PAYMENTS=Low,
192       TENURE=Long-term} => {MINIMUM_PAYMENTS=Unknown}
0.001117443 0.5555556 0.002011398 15.88392 10
193 [29] {ONEOFF_PURCHASES=High,
194       INSTALLMENTS_PURCHASES=Low,
195       CREDIT_LIMIT=High,
196       PAYMENTS=Low,
197       PRC_FULL_PAYMENT=Low,
198       TENURE=Long-term} => {MINIMUM_PAYMENTS=Unknown}
0.001117443 0.5555556 0.002011398 15.88392 10
199 [30] {PURCHASES=High,
200       ONEOFF_PURCHASES=High,
201       CASH_ADVANCE_TRX=Low,
202       CREDIT_LIMIT=High,
203       PAYMENTS=Low,
204       PRC_FULL_PAYMENT=Low,
205       TENURE=Long-term} => {MINIMUM_PAYMENTS=Unknown}
0.001117443 0.5555556 0.002011398 15.88392 10
206 [31] {PURCHASES=High,
207       ONEOFF_PURCHASES=High,
208       INSTALLMENTS_PURCHASES=Low,
209       CREDIT_LIMIT=High,
210       PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001229188 0.5500000 0.002234887 15.72508 11
211 [32] {ONEOFF_PURCHASES=High,
212       CASH_ADVANCE_TRX=Low,
213       CREDIT_LIMIT=High,
214       PAYMENTS=Low,
215       TENURE=Long-term} => {MINIMUM_PAYMENTS=Unknown}
0.001229188 0.5500000 0.002234887 15.72508 11
216 [33] {PURCHASES=High,
217       ONEOFF_PURCHASES=High,
218       INSTALLMENTS_PURCHASES=Low,
219       CREDIT_LIMIT=High,
220       PAYMENTS=Low,
221       PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001229188 0.5500000 0.002234887 15.72508 11
222 [34] {ONEOFF_PURCHASES=High,
223       CASH_ADVANCE_TRX=Low,
224       CREDIT_LIMIT=High,
225       PAYMENTS=Low,
226       PRC_FULL_PAYMENT=Low,
227       TENURE=Long-term} => {MINIMUM_PAYMENTS=Unknown}

```

```

0.001229188 0.5500000 0.002234887 15.72508 11
228 [35] {BALANCE=Low,
229       CASH_ADVANCE=High,
230       CREDIT_LIMIT=Medium,
231       PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.5384615 0.002905353 15.39518 14
232 [36] {BALANCE=Low,
233       PURCHASES=Low,
234       CASH_ADVANCE=High,
235       CREDIT_LIMIT=Medium,
236       PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.5384615 0.002905353 15.39518 14
237 [37] {BALANCE=Low,
238       ONEOFF_PURCHASES=Low,
239       CASH_ADVANCE=High,
240       CREDIT_LIMIT=Medium,
241       PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.5384615 0.002905353 15.39518 14
242 [38] {BALANCE=Low,
243       INSTALLMENTS_PURCHASES=Low,
244       CASH_ADVANCE=High,
245       CREDIT_LIMIT=Medium,
246       PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.5384615 0.002905353 15.39518 14
247 [39] {BALANCE=Low,
248       CASH_ADVANCE=High,
249       CREDIT_LIMIT=Medium,
250       PAYMENTS=Low,
251       PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.5384615 0.002905353 15.39518 14
252 [40] {BALANCE=Low,
253       PURCHASES=Low,
254       ONEOFF_PURCHASES=Low,
255       CASH_ADVANCE=High,
256       CREDIT_LIMIT=Medium,
257       PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.5384615 0.002905353 15.39518 14
258 [41] {BALANCE=Low,
259       PURCHASES=Low,
260       INSTALLMENTS_PURCHASES=Low,
261       CASH_ADVANCE=High,
262       CREDIT_LIMIT=Medium,
263       PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.5384615 0.002905353 15.39518 14
264 [42] {BALANCE=Low,
265       PURCHASES=Low,
266       CASH_ADVANCE=High,
267       CREDIT_LIMIT=Medium,
268       PAYMENTS=Low,
269       PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.5384615 0.002905353 15.39518 14
270 [43] {BALANCE=Low,
271       ONEOFF_PURCHASES=Low,
272       INSTALLMENTS_PURCHASES=Low,

```

```

273     CASH_ADVANCE=High,
274     CREDIT_LIMIT=Medium,
275     PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.5384615  0.002905353  15.39518      14
276 [44] {BALANCE=Low,
277       ONEOFF_PURCHASES=Low,
278       CASH_ADVANCE=High,
279       CREDIT_LIMIT=Medium,
280       PAYMENTS=Low,
281       PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.5384615  0.002905353  15.39518      14
282 [45] {BALANCE=Low,
283       INSTALLMENTS_PURCHASES=Low,
284       CASH_ADVANCE=High,
285       CREDIT_LIMIT=Medium,
286       PAYMENTS=Low,
287       PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.5384615  0.002905353  15.39518      14
288 [46] {BALANCE=Low,
289       PURCHASES=Low,
290       ONEOFF_PURCHASES=Low,
291       INSTALLMENTS_PURCHASES=Low,
292       CASH_ADVANCE=High,
293       CREDIT_LIMIT=Medium,
294       PAYMENTS=Low} => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.5384615  0.002905353  15.39518      14
295 [47] {BALANCE=Low,
296       PURCHASES=Low,
297       ONEOFF_PURCHASES=Low,
298       CASH_ADVANCE=High,
299       CREDIT_LIMIT=Medium,
300       PAYMENTS=Low,
301       PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.5384615  0.002905353  15.39518      14
302 [48] {BALANCE=Low,
303       PURCHASES=Low,
304       INSTALLMENTS_PURCHASES=Low,
305       CASH_ADVANCE=High,
306       CREDIT_LIMIT=Medium,
307       PAYMENTS=Low,
308       PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.5384615  0.002905353  15.39518      14
309 [49] {BALANCE=Low,
310       ONEOFF_PURCHASES=Low,
311       INSTALLMENTS_PURCHASES=Low,
312       CASH_ADVANCE=High,
313       CREDIT_LIMIT=Medium,
314       PAYMENTS=Low,
315       PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
      0.001564421  0.5384615  0.002905353  15.39518      14
316 [50] {BALANCE=Low,
317       PURCHASES=Low,
318       ONEOFF_PURCHASES=Low,
319       INSTALLMENTS_PURCHASES=Low,

```

```

320 CASH_ADVANCE=High ,
321 CREDIT_LIMIT=Medium ,
322 PAYMENTS=Low ,
323 PRC_FULL_PAYMENT=Low} => {MINIMUM_PAYMENTS=Unknown}
0.001564421 0.5384615 0.002905353 15.39518 14

```

Aparece la regla:

$$\{BALANCE = \text{Low}, CASH\_ADVANCE = \text{High}, PURCHASES\_TRX = \text{Low}, CREDIT\_LIMIT = \text{Medium}, PAYMENTS = \text{Low}\} \Rightarrow \{MINIMUM\_PAYMENTS = \text{Unknown}\}$$

Es escogida por ser la que mayor confianza tiene, además de un *lift* alto, y con menos variables. Si se analiza, indica que tiene un saldo bajo, compras con transacciones bajas, un límite de crédito medio y pagos bajos. Estas ocurrencias dan como consecuencia que los pagos mínimos tengan un valor faltante. Si se relaciona con el anterior análisis donde los pagos mínimos faltantes se situaban en el antecedente, es posible ver cierto patrón en los datos. Suelen ser tarjetas de crédito asociadas a valores muy bajos en las variables, aunque no todas las combinaciones de variables en su rango bajo a la vez suelen darse en las reglas, un ejemplo de ello es la regla encontrada donde **CASH\_ADVANCE** es alto. La hipótesis inicial de inactividad es pues menos favorable a ser cierta.



Figura 3: Grafo de relaciones con pagos mínimos nulos.

## 6.2. Análisis de items negados

Se añaden los negados o complementos de los dos *items* que más se repiten en las transacciones. La negación en este caso, al tratarse de una variable continua discretizada, agrupa los intervalos, haciendo que no se centre en intervalos como **HIGH** (por ejemplo) y lo haga en **LOW** y **MEDIUM** a la vez.

```

1 data_neg <- data %>%
2   mutate(
3     PRC_FULL_PAYMENT_NOT_HIGH = PRC_FULL_PAYMENT != "High",
4     PRC_FULL_PAYMENT_NOT_MEDIUM = PRC_FULL_PAYMENT != "Medium",
5     PRC_FULL_PAYMENT_NOT_LOW = PRC_FULL_PAYMENT != "Low",
6     TENURE_NOT_LONG_TERM = TENURE != "Long-term",
7     TENURE_NOT_MEDIUM_TERM = TENURE != "Medium-term",
8     TENURE_NOT_SHORT_TERM = TENURE != "Short-term",
9   )
10 data_trans_neg <- as(data_neg, "transactions")

```

Las reglas obtenidas son en gran parte bastante redundantes o que aportan poca información. Un ejemplo claro de esto es que si se no se es titular de una tarjeta de crédito en un término de tiempo largo ni tampoco en un tiempo medio, es de posesión corta. Pese a ello se pueden encontrar algunas reglas interesantes.

$$\{\text{BALANCE} = \text{High}, \text{CREDIT\_LIMIT} = \text{Medium}, \\ \text{TENURE\_NOT\_LONG\_TERM}\} \Rightarrow \{\text{CASH\_ADVANCE} = \text{High}\}$$

Los clientes con saldo alto, límite de crédito medio y tenencia corta o media tienden a utilizar adelantos en efectivo en gran medida.

$$\{\text{PURCHASES\_TRX} = \text{Low}, \text{MINIMUM\_PAYMENTS} = \text{Medium}, \\ \text{TENURE\_NOT\_LONG\_TERM}\} \Rightarrow \{\text{BALANCE} = \text{High}\}$$

Los clientes que realizan pocas compras con transacciones, realizan pagos mínimos de tamaño medio y tienen una tenencia corta con la tarjeta, suelen tener una deuda alta.

En cuanto a **PRC\_FULL\_PAYMENT\_NOT\_HIGH**, la mayoría de las reglas se asocian con saldos (deudas) muy altos, pagos y compras elevados, es decir, clientes que gastan y se endeudan considerablemente. Teniendo en cuenta que **PRC\_FULL\_PAYMENT\_NOT\_HIGH** indica que el cliente no tiene una proporción alta de pagos completos, esto sugiere que podrían ser clientes con un riesgo elevado de impago.

Una de las reglas más interesantes es la siguiente:

$$\{\text{BALANCE} = \text{Very High}, \text{PRC\_FULL\_PAYMENT\_NOT\_HIGH}\} \\ \Rightarrow \{\text{MINIMUM\_PAYMENTS} = \text{High}\}$$

Los clientes con saldo muy alto y un bajo porcentaje de pago completo tienden a realizar pagos mínimos elevados. Esto podría interpretarse como una medida compensatoria frente a incumplimientos.



## 7. Conclusión

Tras analizar todas las reglas obtenidas, se han recopilado las consideradas más interesantes en una lista:

1.  $\{\text{INSTALLMENTS\_PURCHASES} = \text{Low}, \text{MINIMUM\_PAYMENTS} = \text{Medium}, \text{TENURE} = \text{Long-term}\} \Rightarrow \{\text{BALANCE} = \text{High}\}$
2.  $\{\text{CASH\_ADVANCE} = \text{Low}, \text{PURCHASES\_TRX} = \text{Medium}, \text{MINIMUM\_PAYMENTS} = \text{Low}\} \Rightarrow \{\text{BALANCE} = \text{Low}\}$
3.  $\{\text{CASH\_ADVANCE} = \text{High}\} \Rightarrow \{\text{PURCHASES} = \text{Low}\}$
4.  $\{\text{BALANCE} = \text{High}, \text{CREDIT\_LIMIT} = \text{Medium}, \text{TENURE\_NOT\_LONG\_TERM}\} \Rightarrow \{\text{CASH\_ADVANCE} = \text{High}\}$
5.  $\{\text{PURCHASES\_TRX} = \text{Low}, \text{MINIMUM\_PAYMENTS} = \text{Medium}, \text{TENURE\_NOT\_LONG\_TERM}\} \Rightarrow \{\text{BALANCE} = \text{High}\}$
6.  $\{\text{CASH\_ADVANCE} = \text{Medium}, \text{PURCHASES\_TRX} = \text{High}, \text{MINIMUM\_PAYMENTS} = \text{High}\} \Rightarrow \{\text{BALANCE} = \text{High}\}$
7.  $\{\text{BALANCE} = \text{Low}, \text{CASH\_ADVANCE} = \text{High}, \text{PURCHASES\_TRX} = \text{Low}, \text{CREDIT\_LIMIT} = \text{Medium}, \text{PAYMENTS} = \text{Low}\} \Rightarrow \{\text{MINIMUM\_PAYMENTS} = \text{Unknown}\}$
8.  $\{\text{MINIMUM\_PAYMENTS} = \text{High}, \text{TENURE} = \text{Short-term}\} \Rightarrow \{\text{CASH\_ADVANCE} = \text{Low}\}$

Como posibles mejoras que se podrían haber realizado, me he dado cuenta que quizá al discretizar las variables habría sido más interesante generar rangos mucho más diversos. En vez de haber utilizado tres tipos de rangos, considero que habrían salido reglas más interesantes y específicas usando quizá hasta seis niveles por variable.