

# Data Shortcuts

Algunas ilustraciones clave que se tuvieron en cuenta  
para realizar la modelación en inteligencia artificial en  
Hevarán



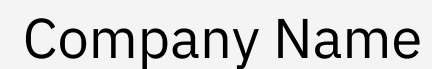
Juan José Pardo  
AV Vertical Data

Hevarán

# Análisis de Cohortes

El poder de segmentar los datos por usuarios

[BACK TO AGENDA PAGE](#)

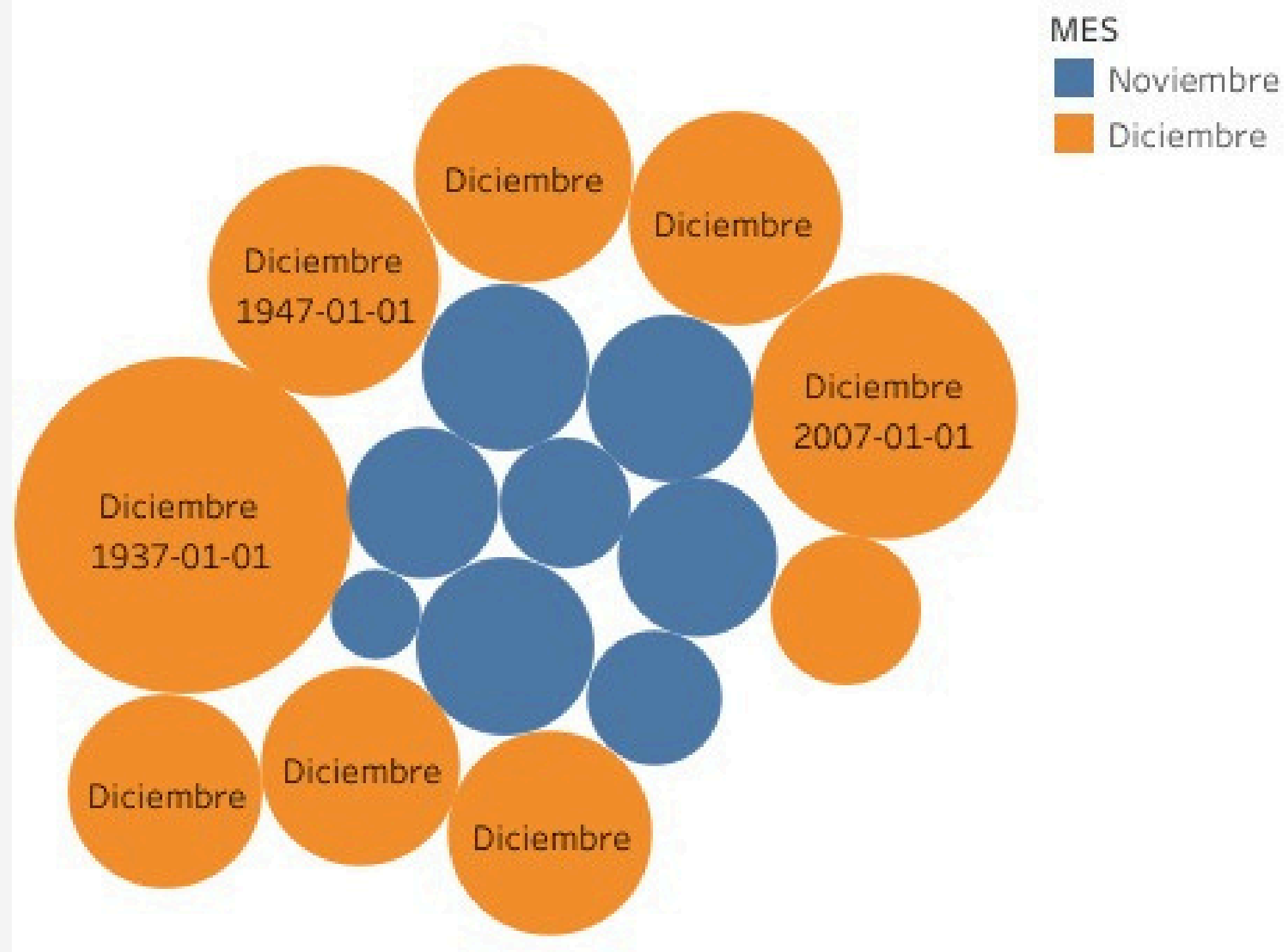


## 1977- el año de la cohorte mayor

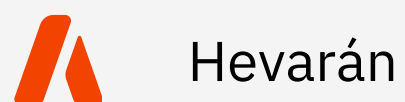
Presenta una representación del 31.3% de los datos en la categoría de saldo total general

**BACK TO AGENDA PAGE**

## Distribucion de saldos por cohorte de nacimiento y mes





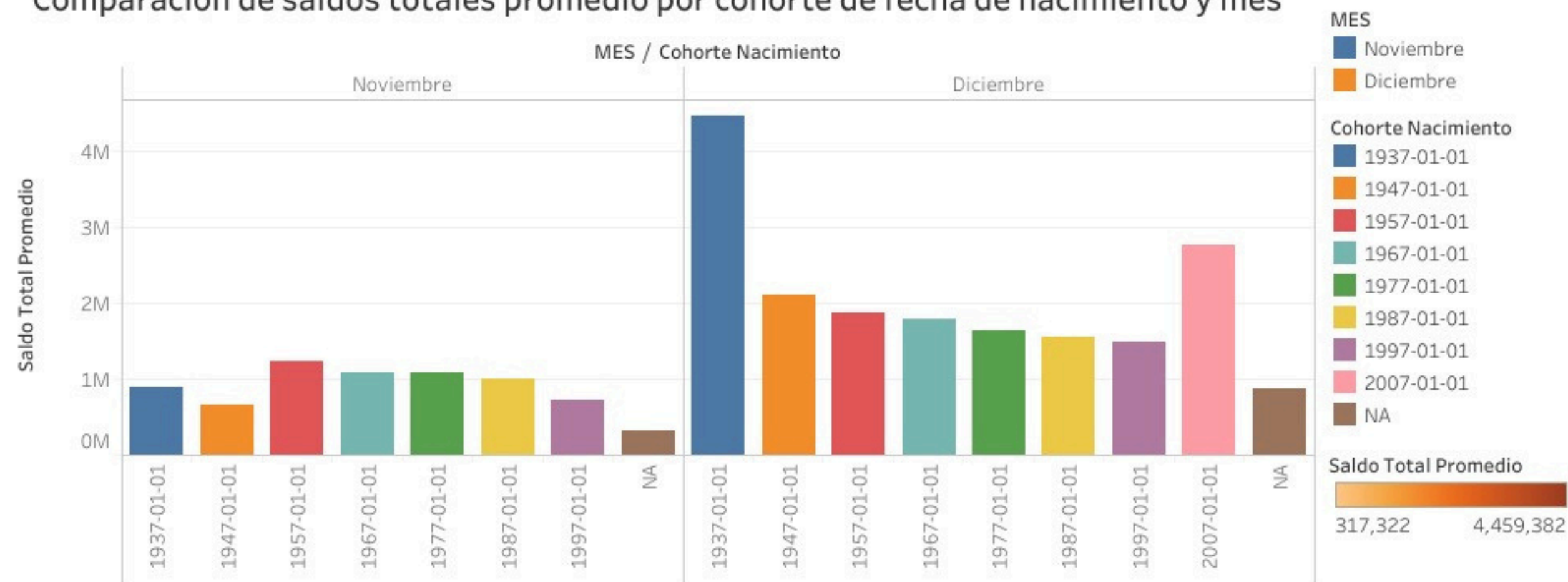


## Algunos valores adicionales?

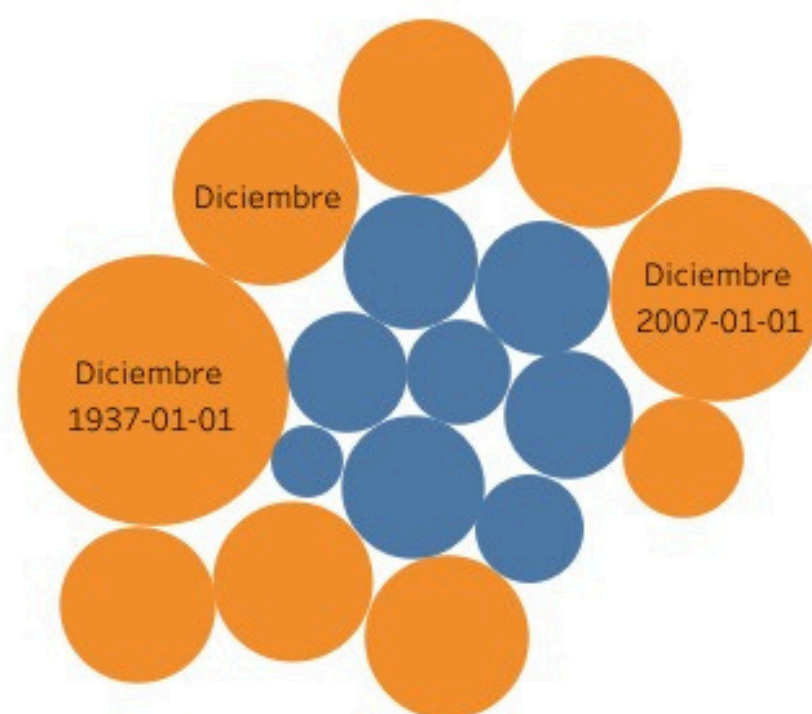
Encontramos que las edades que representan una proporción entre los 63 a los 87 años tienen un 23.57% mayor probabilidad de pago en contraste con medias de otras cohortes

**Insights Adicionales:** El modelo cuenta con una efectividad del 86.34% de los datos relacionados con su efectividad para medir la probabilidad de pago de un usuario respecto a la edad de sus clientes.

## Comparación de saldos totales promedio por cohorte de fecha de nacimiento y mes



## Distribucion de saldos por cohorte de nacimiento y mes



## Comparación de saldos promedio por cohorte y mes





Company Name

# Regresión Logística

Técnica de modelado de datos en inteligencia artificial con algoritmo del 82% de eficiencia a partir de la prueba de matriz de confusión

[BACK TO AGENDA PAGE](#)

Hevarán

```
rkspace  
s())
```

```
dataset in loan.csv
```

```
ran<-read.csv("/Users/mac/Desktop/ARCHIVOS HEVARAN/COLSUBSIDIO/DATOS HEVARAN/BasePagosCastigoCupo2.csv",header=TRUE,sep=";",as.is=TRUE)  
sHevaran)
```

```
r liquidación en base al sistema como factor
```

```
(DatosHevaran$liquida.sistema)
```

```
ran$Tipo.Cartera.Colsubsidio<-as.factor(DatosHevaran$liquida.sistema)
```

```
ción de modelo para predecir el efecto en la liquidación del sistema con factores como honorarios y comisión calculada
```

```
Cartera.Colsubsidio<-glm(liquida.sistema ~ Honorarios + Comisión.Calculada, data = DatosHevaran, family = "binomial")
```

```
ón de la categoría liquida.sistema usando el modelo
```

```
ct.liquida.sistema<- predict(glm.Tipo.Cartera.Colsubsidio,DatosHevaran, type = "response")
```

```
predict.liquida.sistema)
```

```
r los valores predictivos en Verdadero/Falso
```

```
ran$predict.liquida.sistema<-ifelse(glm.predict.liquida.sistema >=.5, "True", "False" )
```

```
sHevaran$predict.liquida.sistema,50)
```

```
ran$predict.liquida.sistema<-ifelse(glm.predict.liquida.sistema >=.5,"El valor tomado en honorarios superó los $76.000
```

```
sHevaran$predict.liquida.sistema,50)
```

```
ción del gráfico con los puntos originales en el dataset
```

```
sHevr)
```

# Fuente de Código

[BACK TO AGENDA PAGE](#)

La fuente de código utilizada par la generación del modelo de inteligencia artificial bajo la metodología de optimización fue presentado con el Software de Open Source de R generando para el cálculo de la regresión logística





Company Name

# Árboles de decisión

Técnica de modelado de datos en inteligencia artificial con algoritmo del 92.34% de eficiencia a partir de la prueba de matriz de confusión

[BACK TO AGENDA PAGE](#)

Hevarán

# Fuente de Código

[BACK TO AGENDA PAGE](#)

```
cantidad de valores faltantes en cada columna
issing)
la calidad de las variables y la presencia de outliers por medio de la función "summary()"
|
los datos faltantes en cada una de las variables aplicadas en el modelo (Porcentaje en cada columna)
<- colMeans(is.na(data)) * 100
missing_perc)
los nombres de las columnas que cuentan con mas del 25% de los valores faltantes
<- missing_perc[missing_perc > 25]
issing)
eliminación de outliers para eliminar el ruido de los valores extremos para Valor_Pago_Minimo: Identificación
le(data$Valor_Pago_Minimo, probs=c(.25, .75), na.rm = TRUE)
ile(data$Valor_Pago_Minimo, probs=c(.10, .90), na.rm = TRUE)
R(data$Valor_Pago_Minimo, na.rm = TRUE)
igo_Minimo[data$Valor_Pago_Minimo < (qnt[1] - H)] <- caps[1]
igo_Minimo[data$Valor_Pago_Minimo > (qnt[2] + H)] <- caps[2]
estadística relacionada con los datos generados: Presentación del output de los datos generados a p
Valor_Pago_Minimo)
multicolinealidad para identificar alta correlación entre las variables predictoras que afecten la pr
variables que son únicamente numéricas en la matriz de correlación
apply(data, is.numeric)
```

La fuente de código utilizada por la generación del modelo de inteligencia artificial bajo la metodología de optimización fue presentado con el Software de Open Source de R generando predictibilidad en los clusters

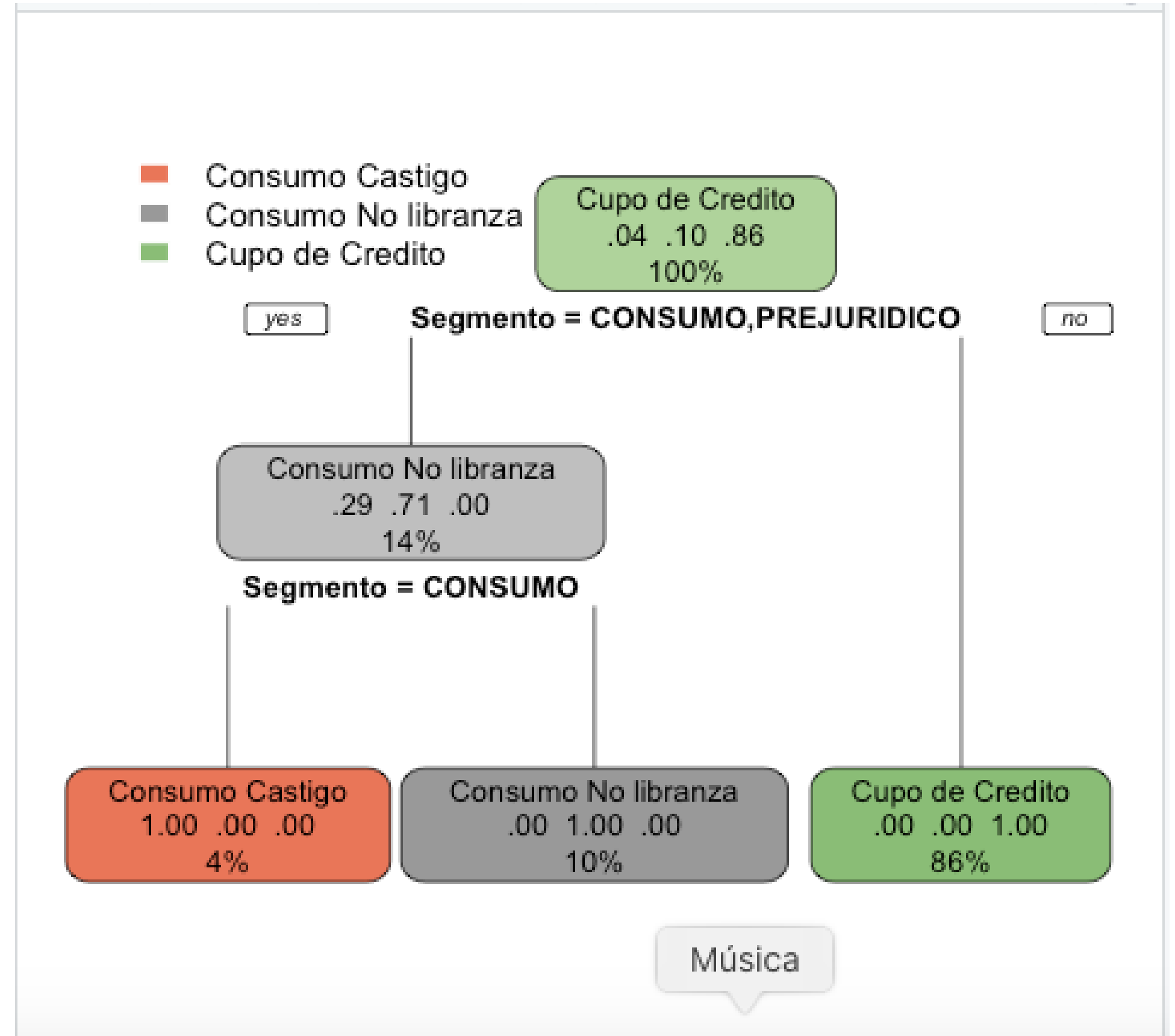


# Insights de Valor

**Foco en los clientes con saldos totales mayores a \$4.947.708**

Aquellos clientes que presentan especial atención con estado de mora superior a 59 días de mora o más

[BACK TO AGENDA PAGE](#)

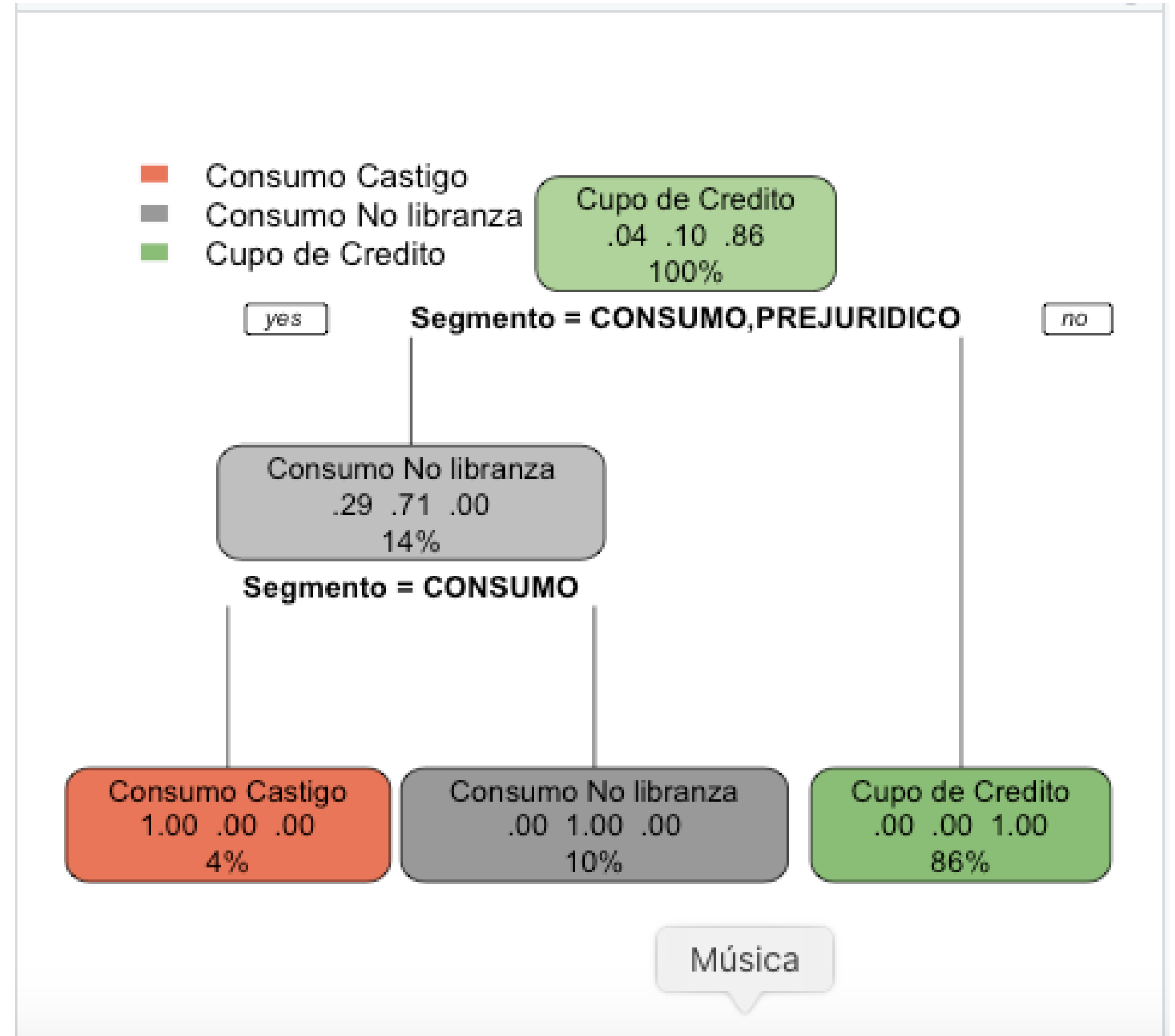


# Insights de Valor

Para el cluster de Consumo Castigo

Ofrecer planes de pago flexibles y presentar opciones de refinanciación para ayudarles a los mismos a ponerse al día con sus pagos y evitar que la cuenta pueda caer en estado de castigo

[BACK TO AGENDA PAGE](#)

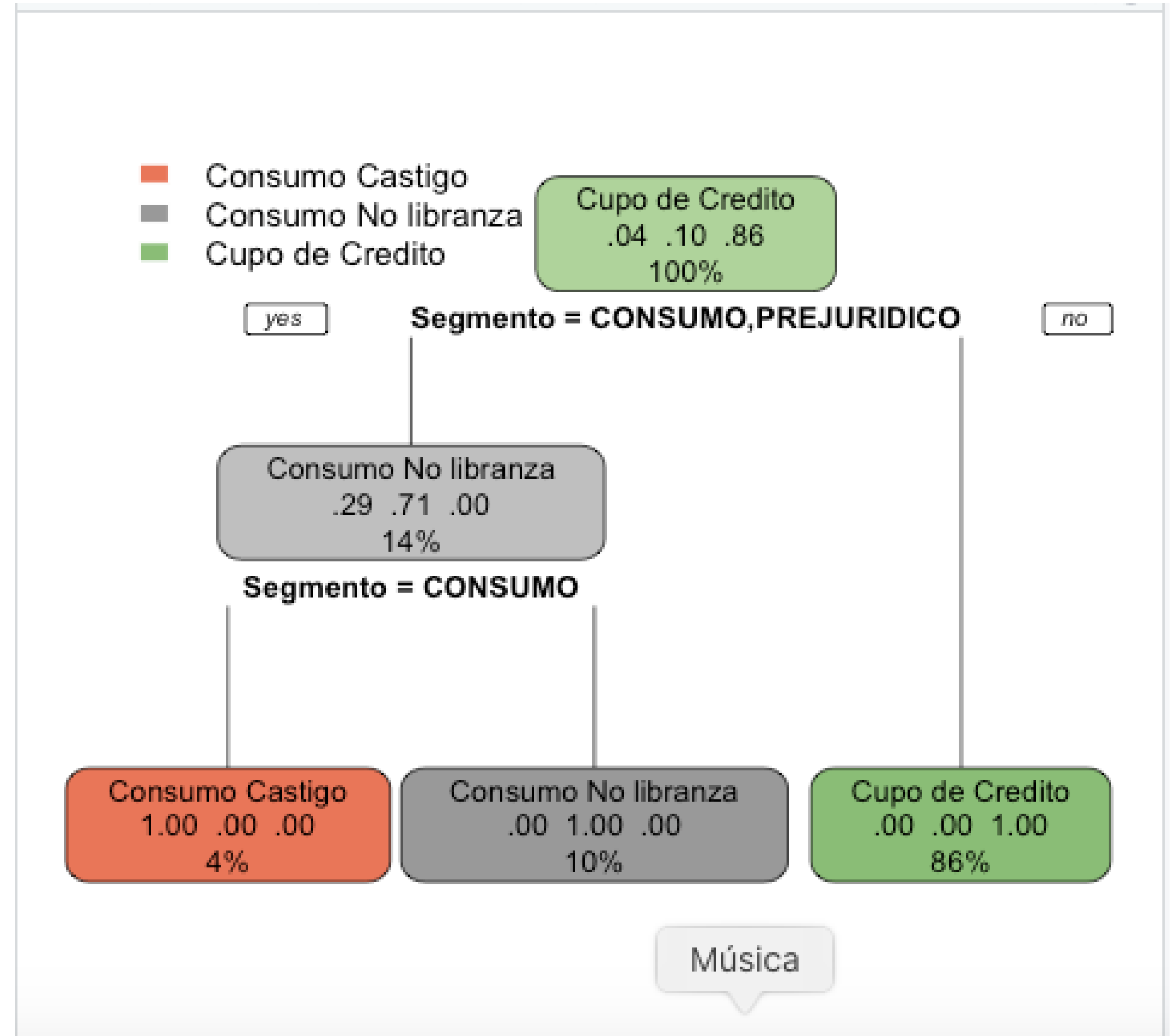


# Insights de Valor

Para el cluster de Cupo de Crédito

Bajo las hojas, los clientes que cuentan con un saldo total o menor que \$6,915,545 y que cuentan con un pago mínimo menor a \$281.021 cuentan con una alta probabilidad de estar ubicados en la categoría de castigo.

[BACK TO AGENDA PAGE](#)





# Decisiones clave a partir del árbol de decisión

[BACK TO AGENDA PAGE](#)



## **Priorizar la alta probabilidad de pago**

Presentar estrategias según el segmento generado en el código



## **Identificar las causas de impago**

Identificar las estrategias a partir de las bases operacionales del negocio



## **Asignación efectiva de clientes**

Establecer los agentes de mejor rendimiento en la prioridad de cuentas.