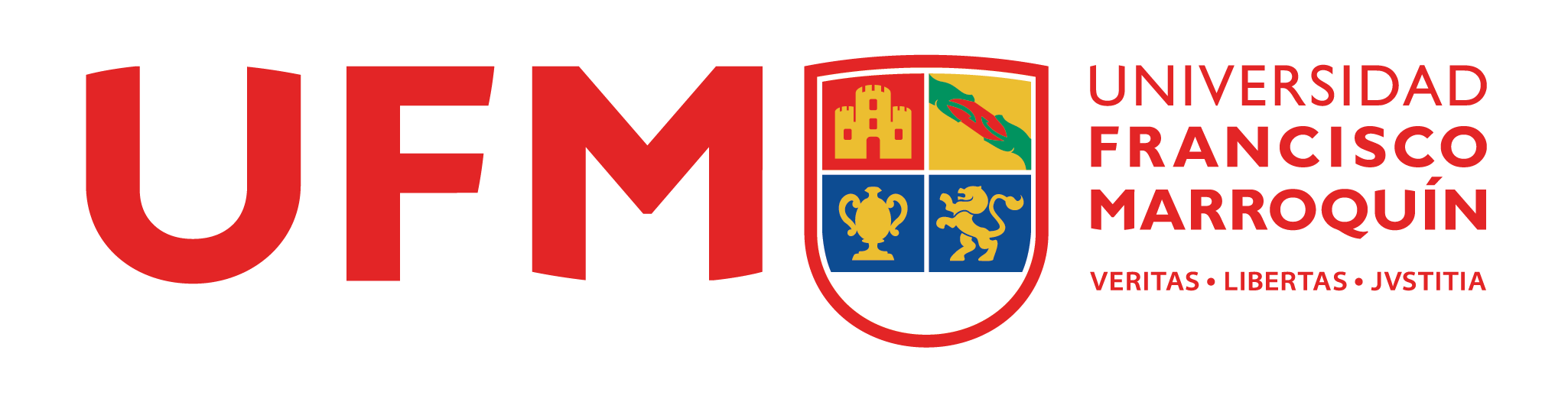
Universidad Francisco Marroquín

María Isabel Avila Rigalt

Data Science for Finance

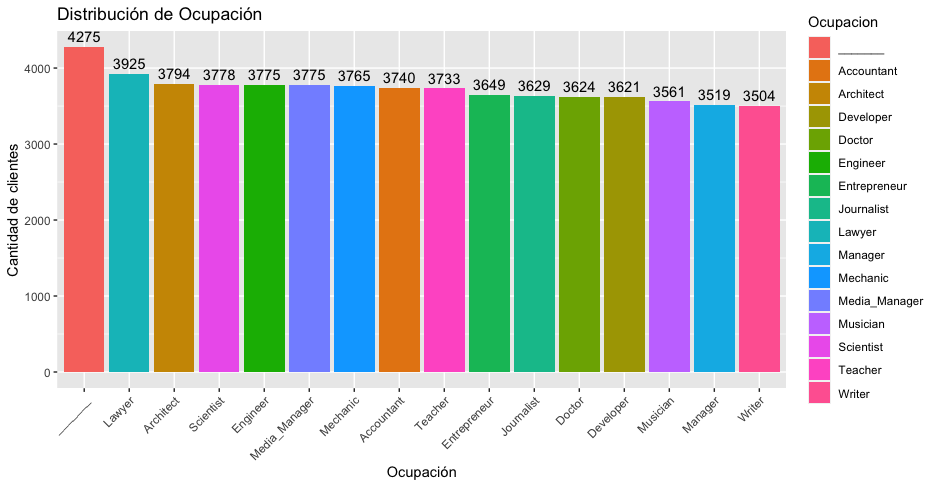
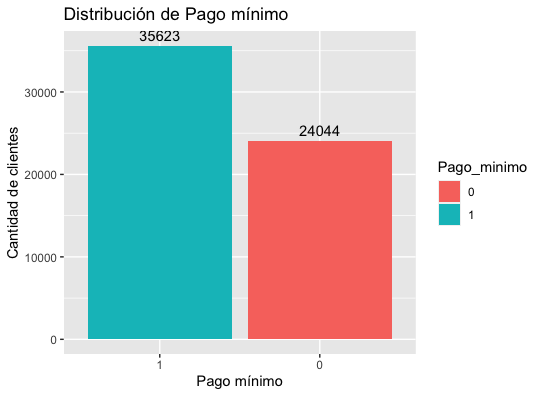
**CASO PRÁCTICO NO. 3**

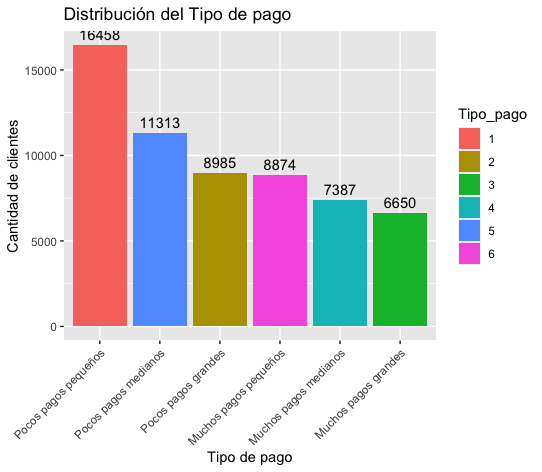
**PREDICCIÓN DE CALIFICACIÓN CREDITICIA CON ALGORITMOS DE CLASIFICACIÓN**

Steph Grotewold - 20210567

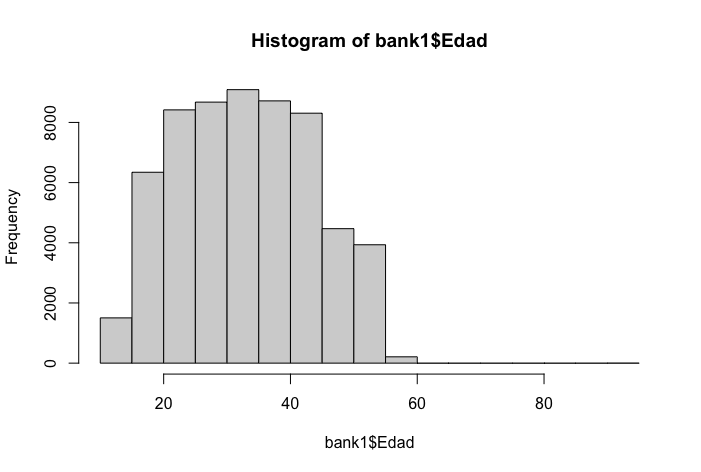
03/07/2023

# Exploración de Variables:

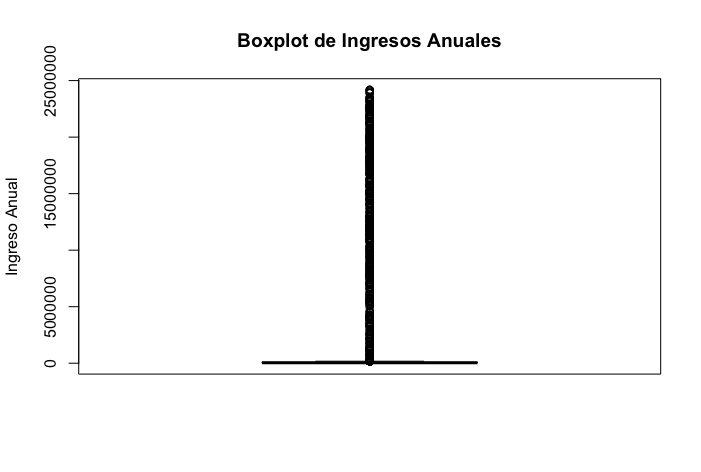
* **Variables Categóricas:**
  + Ocupación: tiene 16 niveles en el cual el más frecuente es \_\_\_, luego sigue con Lawyer 3,925 y el menos frecuente siendo Writer con 3,504, por lo que podemos notar que está bastante uniforme la distribución. Por otra parte, el área bajo la curva es de 50.23% que no es una buena variable predictora porque está muy cerca a que sea aleatorio.
  + Pago Mínimo: Tiene más instancias que su último pago fue el mínimo requerido por el banco, siendo un 59.7% aproximadamente. Sin embargo, el area bajo la curva para esta variable es de 72.75% lo cual nos dice que es una buena predictora.
  + Tipos de pagos: El tipo de pago más frecuente es de pocos pagos pequeños con un 16,458. Ahora bien, el área bajo la curva es de 46.11% por lo que se asume que su capacidad de predecir es moderada.



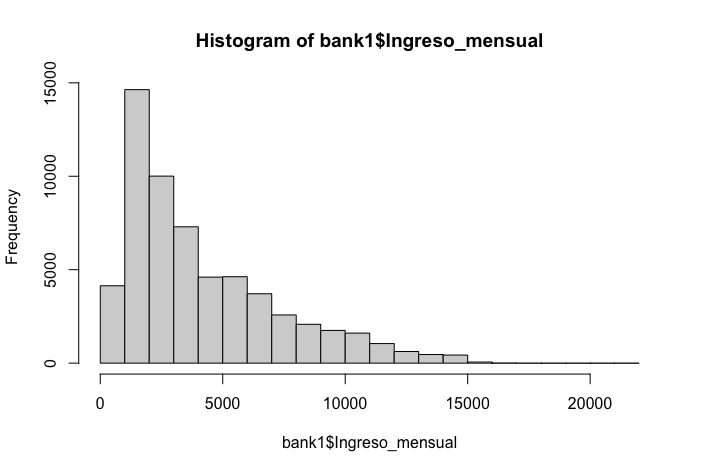
* **Variables Numéricas:**
  + Edad: Tiene edades desde 14 hasta 95 años, con una media de 33 años.



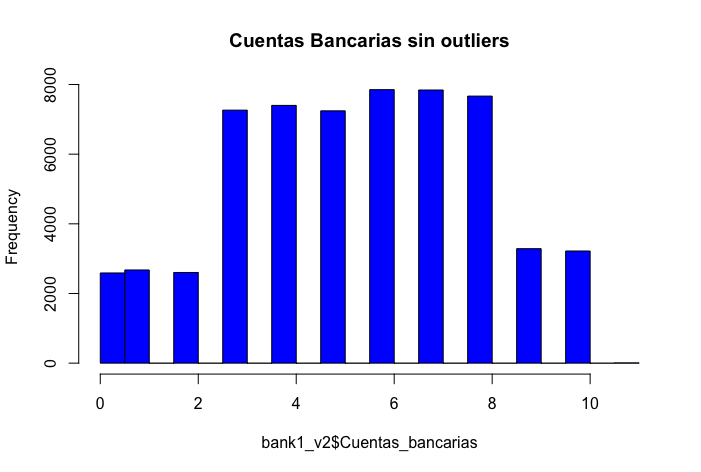
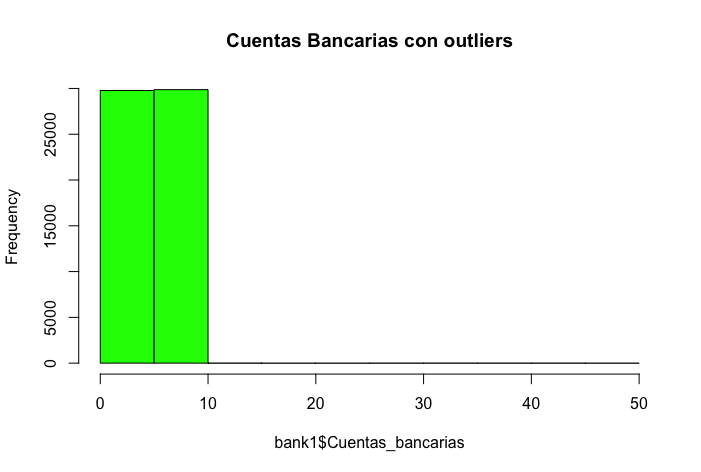
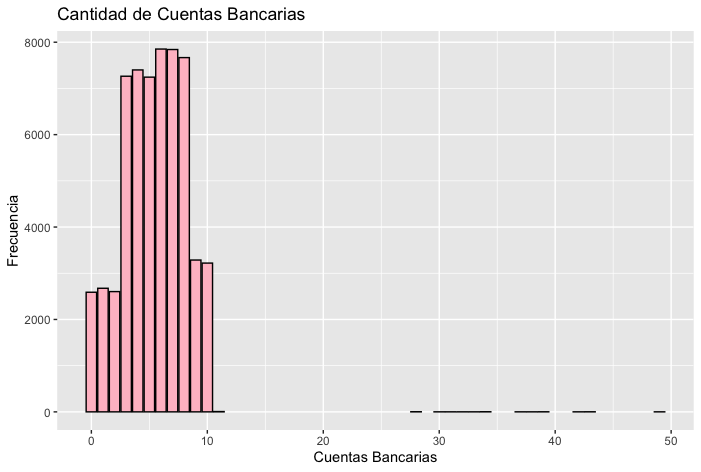
* + Ingreso Anual: Este varía mucho, el mínimo siendo de 7,006 y llegar a un máximo de 24,198,062. La media es de 183,663.



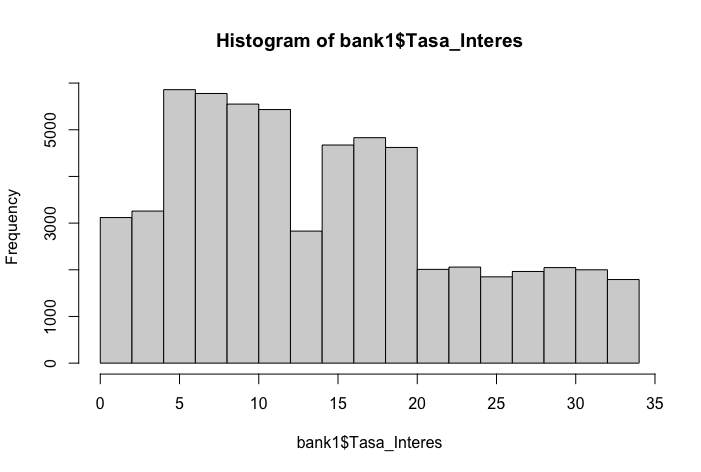
* + Ingresos Mensuales: Este tiene un rango de 0 a 21,167, con una media de 4,179. Este tiene un sesgo positivo como se puede ver en la siguiente grafica.

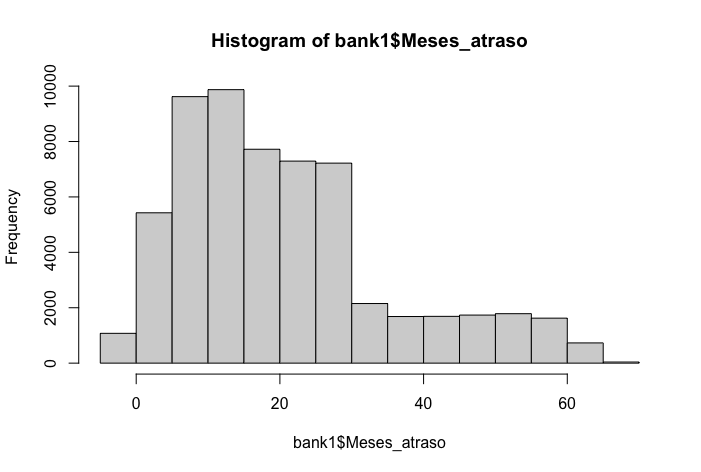
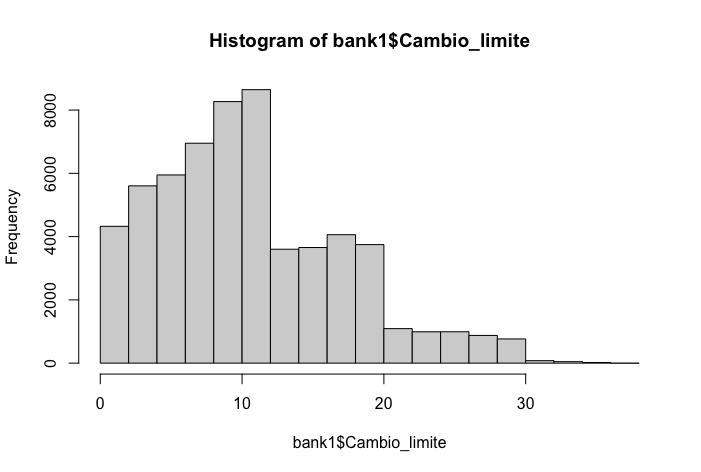
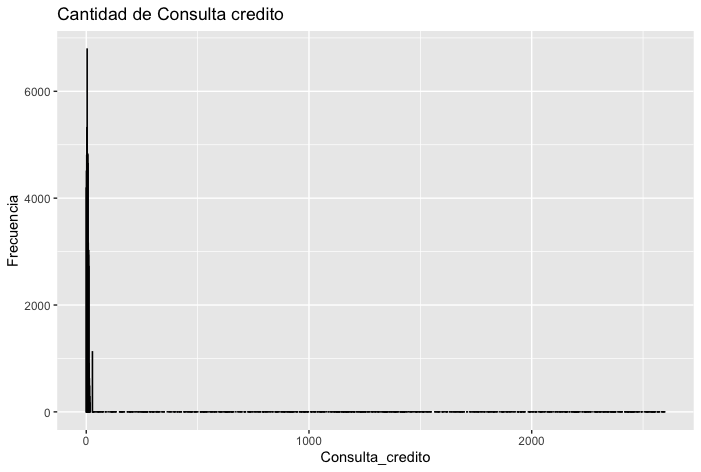


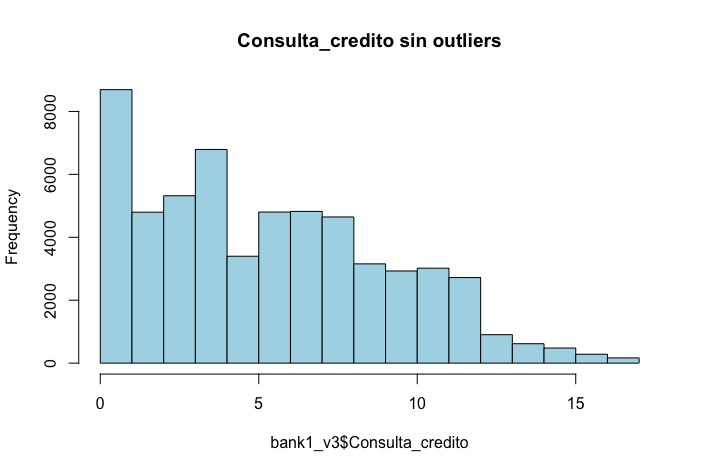
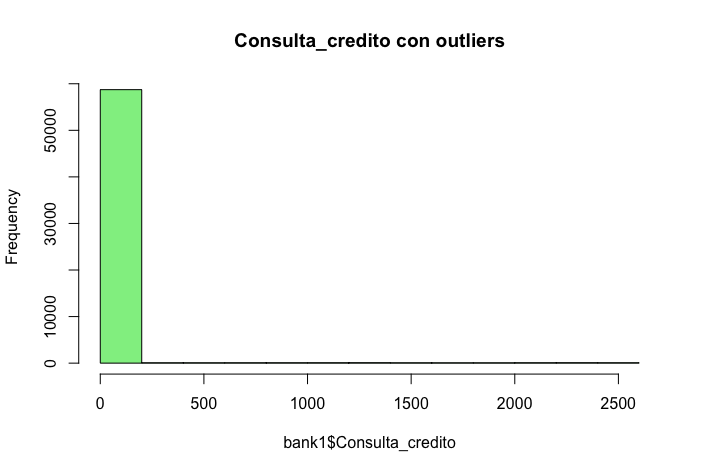
* + Cuentas Bancarias: tiene una persona que tiene 0 cuentas hasta 49 con una media de 5 cuentas por persona. Ahora bien, a esta variable se le quitaron los outliers lo cual ayudo a mejorar la predicción.



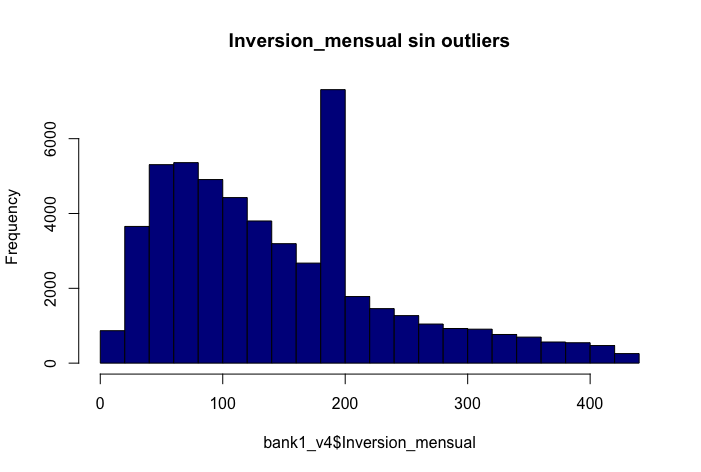
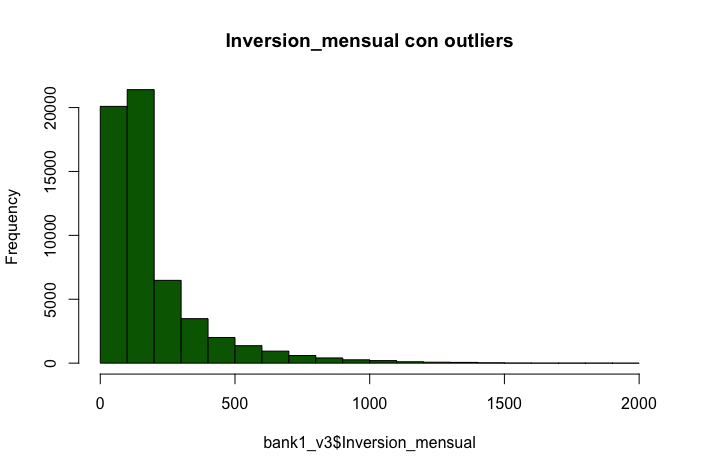
* + Tasa de interés: está varía de un 1% a un 34%, el promedio siendo de 14.55%



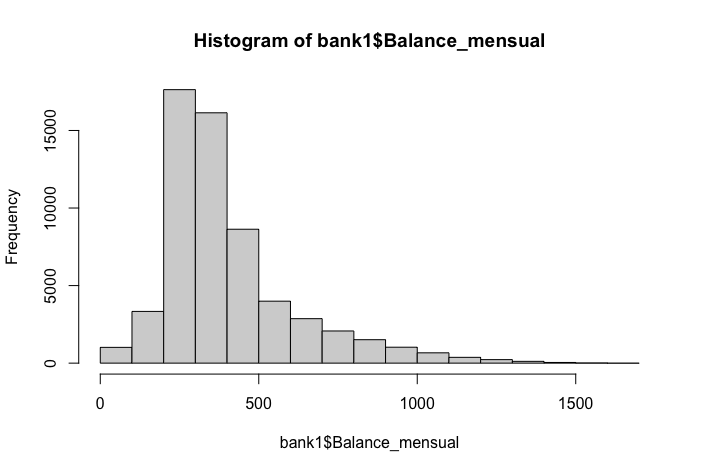
* + Meses Atraso: Esta variable tiene números negativos, entonces tiene un rango de -5 a 67 meses y su media es de 21. 
  + Cambio límite: Este varía de 0.01 a 36.09 con una media de 10.61. 
  + Consulta crédito: la cantidad de veces que ha consultado sobre su crédito varían de 0 a 2,597 con un promedio de 28 veces. A este también se le quitaron los outliers para mejorar la predicción. 



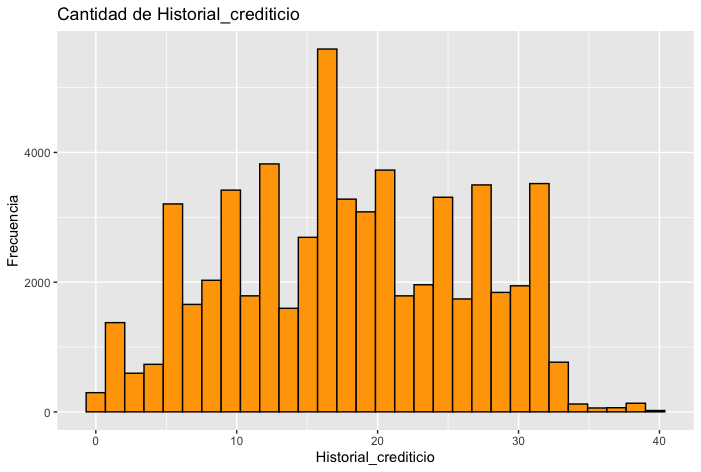
* + Inversión Mensual: La inversión mensual varía de 0 a 1,977.33, con una media de 194.37. Se puede notar un sesgo positivo previo a que se le quitaran los outliers para mejorar la predicción.



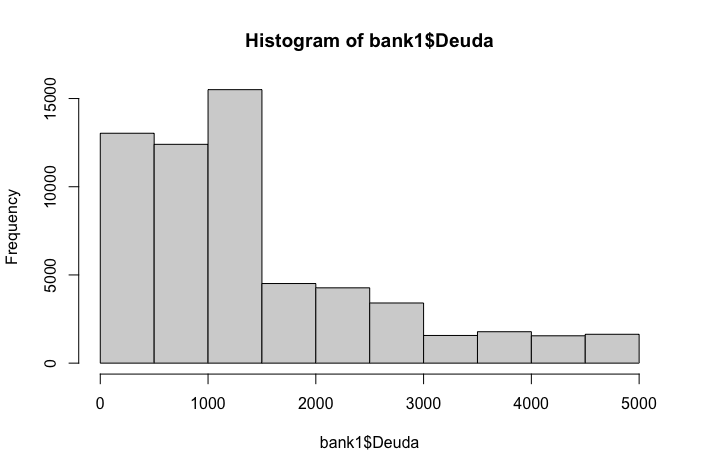
* + Balance Mensual: Este varía de 0.089 a 1,602.04 con un promedio de 402.51.



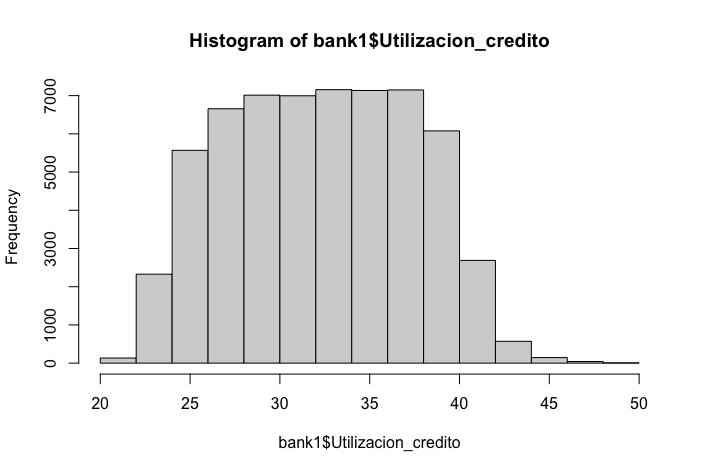
* + Historial Crediticio: Tiene un rango de 0 a 39.7 años, con un promedio de 17 años por cliente.



* + Deuda: el monto de la deuda varía de 0.23 a 4,998.07 con un promedio de 1,425.41.



* + Utilización de crédito: El porcentaje va de 20% a un 50% con un promedio de 32.28%



# Resultados de Modelos:

Se probaron 9 modelos, sin embargo, se eligió el siguiente modelo porque tenía el akaike más bajo, dependiendo de la segmentación que realice, el train está muy cerca del modelo 3, en donde el model3 daba un AIC de 57179.69 y luego el modelo final daba un AIC de 57180.16. Ahora bien, este modelo final tiene un accuracy mayor de 63.14 a un 63.096 en la parte de test, lo cual es mínimo, pero se prefirió este modelo por esa pequeña diferencia. Ahora bien, se descartaron las variables dummy de ocupación porque no tenían un poder predictivo muy alto y hacían que el akaike aumentara.

Modelo final:

Interpretación:

* Hay dos interceptos, porque el tercero se toma como referencia implícita. El primer nivel es de -1.10 y el segundo nivel es de -3.085, que es la probabilidad de que sea cada nivel de riesgo en caso el resto de las variables sean cero. El implicito es 0, el cual significa bueno, nivel 1 es regular y el nivel de riesgo 2 es malo.
* Edad: Por cada año extra, la probabilidad de calificación de riesgo 1 disminuye en 0.00333, mientras que para el nivel 2 disminuye en 0.003415
* Ingreso anual: por cada unidad que aumente el ingreso anual, la probabilidad de que el nivel de riesgo sea 1 aumenta en 0.00000001160936, por el otro lado para el nivel 2 disminuye en 0.00000001154682
* Cuentas Bancarias: Por cada cuenta extra que tenga el individuo, la probabilidad que sea nivel 1 aumenta en 0.091, mientras que para el nivel 2 aumenta en 0.052.
* Tasa interés: por cada unidad que aumente la tasa de interés la probabilidad de ser nivel 1 aumenta en 0.06386559, por el otro lado para el nivel 2 aumenta en 0.116.
* Meses Atraso: Por cada mes de atraso, la probabilidad de ser nivel 1 aumenta en 0.04, mientras que para ser nivel 2 aumenta en 0.055.
* Tipo Pago 1: tener este tipo de pago (pocos pagos de valor pequeño) hace que la probabilidad de ser nivel 1 aumente en de 0.48 y nivel 2 sea de 0.7978.
* Tipo Pago 2: tener este tipo de pago (pocos pagos de valor mediano) hace que la probabilidad de ser nivel 1 aumente en de 0.2775 y nivel 2 sea de 0.495.
* Tipo Pago 4: tener este tipo de pago (muchos pagos de valor pequeño) hace que la probabilidad de ser nivel 1 aumente en de 0.01706 y nivel 2 disminuye en -0.042.
* Tipo Pago 5: tener este tipo de pago (muchos pagos de valor mediano) hace que la probabilidad de ser nivel 1 disminuya en de -0.086 y nivel 2 disminuye en -0.19.
* Tipo Pago 6: tener este tipo de pago (muchos pagos de valor grande) hace que la probabilidad de ser nivel 1 disminuya en de -0.262 y el nivel 2 disminuye en -0.54.