

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

METODOLOGÍA PARA LA OBTENCIÓN DE UN MODELO DE GESTIÓN DE COBRANZAS DE CRÉDITOS MASIVOS. DESARROLLO E IMPLEMENTACIÓN DE UN MODELO SCORE

TESIS PREVIA A LA OBTENCIÓN DEL TÍTULO DE MAGISTER EN ESTADÍSTICA
APLICADA

DIEGO ORIOL VARGAS LARA

lorio71@hotmail.com

DIRECTOR: Msc. Ruth Cecilia Vallejo Idrovo

cvallejo@solidario.fin.ec

CODIRECTOR: Dr. Walter Polo Vaca Arellano

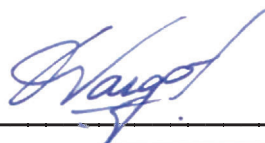
polo.vaca@epn.edu.ec

Quito, septiembre 2015

DECLARACIÓN

Yo Diego Oriol Vargas Lara, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

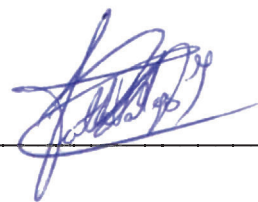
A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.



Diego Oriol Vargas Lara

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Diego Oriol Vargas Lara, bajo mi supervisión.

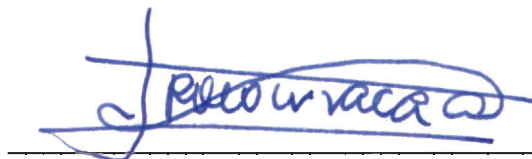
A handwritten signature in blue ink, consisting of several loops and flourishes, positioned above a horizontal line.

Msc. Ruth Cecilia Vallejo Idrovo

DIRECTORA DE LA TESIS DE GRADO

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Diego Oriol Vargas Lara, bajo mi supervisión.



Dr. Walter Polo Vaca Arellano

CODIRECTOR DE LA TESIS DE GRADO

RESUMEN

La evolución del consumo en el Ecuador, ha permitido que la oferta de crédito en diferentes líneas de negocio se masifique, generando en algunos sectores el sobre endeudamiento de los consumidores, lo que ha dificultado la recuperación del capital por parte de las originadoras de crédito, que han delegado la responsabilidad a empresas de Cobranzas para que gestionen y recuperen su cartera. Este incremento de volúmenes obliga a las empresas de cobro a optimizar sus recursos para poder recuperar la mayor cantidad del capital en el menor tiempo posible, por lo que se hace necesario la implementación de un modelo conceptual de gestión, así como también el desarrollo de modelos estadísticos que permitan determinar la probabilidad de recuperar la obligación crediticia. Este trabajo describe las características del modelo de gestión y el desarrollo de un modelo de scoring de comportamiento, aplicado con el fin optimizar los recursos de la empresa y mejorar los resultados de la recuperación.

ABSTRACT

The permanent growth of consumption in Ecuador has increased the supply of credit in different business lines generating the increase of consumer indebtedness in some sectors, which has hampered the recovery of capital by credit originators, which have delegated the responsibility of collection to specialized companies to manage and recover their portfolio. This increase in volume requires collection companies to optimize their resources to recover as much capital as quickly as possible, so that the implementation of a conceptual model of management is necessary, as well as the development of statistical models for determining the likelihood of recovering the credit obligation. This paper describes the characteristics of the management model and development of a behavioral scoring model, applied to optimize business resources and improve the results of recovery.

PRESENTACIÓN

Con la dolarización en el país se ha generado una gran estabilidad en los ingresos de la población, así como una inflación controlada, lo que ha desembocado en que la capacidad adquisitiva de las personas sea practicante constate y con un mercado cada vez más orientado al consumo, debido a la penetración de los medios de comunicación, los avances tecnológicos y las estrategias de marketing alineados a promocionar crédito en pequeñas cuotas, ha provocado un enorme crecimiento del volumen de operaciones, así como la inclusión de sujetos de crédito que anteriormente no estaban atendidos. Estos portafolios masivos por su gran cantidad de operaciones y heterogeneidad de perfiles de deudores, deben ser segmentados de tal forma que la gestión realizada sea diferenciada por perfiles para mejorar la colocación y disminuir la pérdida por riesgo crediticio.

Para poder manejar un negocio de consumo masivo, es necesario disponer de metodologías de administración, que permita conocer el rendimiento esperado de la colocación de crédito, con información que ayude a tomar decisiones oportunas a corto, mediano y largo plazo, para así lograr el balance riesgo retorno propuesto, ejecutando los correctivos necesarios a los procesos y políticas de originación y la mejora continua en la recuperación.

Debido a la gran cantidad de estudios realizados para la concesión de crédito, se planteó la necesidad de analizar lo que sucede una vez concedido el mismo, desde la óptica de la recuperación, como una oportunidad de mejora y dado que el tema está muy poco difundido y estudiado. El desarrollar una metodología para la Administración de los procesos de Cobranza masivos, apalancados en los análisis estadísticos que conduzcan optimizar los recursos y mejorar las recuperaciones, permitirá a las empresas que lo ejecuten, entre otros objetivos, disminuir las pérdidas por riesgo crediticio, incrementar la rentabilidad fruto de la optimización de los recursos y el

mejoramiento en la recuperación, controlar el resultado de la gestión del portafolio, monitorear el proceso integral de la gestión y estandarizar los procesos en la gestión de cobro.

La implementación de un modelo de estadístico en la gestión de recuperación provee al administrador información de estadísticas tanto descriptivas como inferenciales que le conducen a optimizar los procesos de análisis de datos y toma de decisiones en las estrategias de gestión de portafolios masivos, así como en la administración de los recursos, al remplazar criterios subjetivos de los gestores al momento del contacto con el cliente.

Se genera eficiencias en la evaluación de grandes lotes de información en el menor tiempo posible, manteniendo el control de los procesos de cobranza a través de parámetros y políticas según el perfil de riesgo del cliente, lo que permite tomar decisiones más asertivas en las estrategias de recuperación.

Esta visión en la administración del crédito cada vez más está siendo estudiada a nivel regional, prácticamente en todos los seminarios de Crédito y Cobranza, se enfocan en las características de la administración de la cobranza, enfatizando en la optimización de los recursos mediante la utilización de técnicas estadísticas que permitan segmentar, ordenar y rentabilizar los portafolios en mora, como son los modelos de cluster, modelos logísticos, redes neuronales, así como también precautelar los intereses de los originadores para el cuidado de sus clientes.

En el presente trabajo se abordarán conceptos del proceso de cobranza de créditos masivos, como son: la asignación de Carga, distribución de recursos, utilización de canales de contacto, tecnologías de operación, indicadores de gestión, indicadores de recuperación, enfocados a la aplicación de las estrategias que incluyan criterios de ordenamiento y segmentación apalancados en datos estadísticos. Adicionalmente el modelo de scoring desarrollado permite clasificar y ordenar individuos que cumplen

determinadas características de un fenómeno estudiado, obteniendo una fórmula matemática que otorga un puntaje a la probabilidad de recuperación de la operación concedida.

El estudio contempla una serie de fases metodológicas que se deben alcanzar con el fin de garantizar la consistencia del modelo a aplicarse. Inicia con las definiciones relacionadas con el negocio, el análisis descriptivo de los datos de un portafolio de créditos masivos, la generación de reportes de control que contienen indicadores estadísticos para la toma de decisiones; a partir de esta información se realiza el método estadístico de discriminación, evaluando sus beneficios desde un punto de vista estadístico y de gestión, y finalmente se realiza su implementación, considerando las estrategias que puede emplear la institución con el fin de maximizar la recuperación de la cartera.

ÍNDICE

1	Introducción: Características de los portafolios masivos, principales productos en el mercado, relación riesgo rentabilidad, el ciclo de crédito.	1
1.1	Crecimiento de los créditos de consumo en Ecuador.	1
1.1.1	Principales productos en el Mercado	2
1.2	Relación Riesgo Retorno	3
2	Aspectos generales en la gestión de cobranzas, tecnología utilizada, tipos de gestión, estrategias usuales.....	9
2.1	Aspectos generales.	9
2.2	Tecnología utilizada.....	11
2.3	Tipos de Gestión.....	13
2.4	Estrategias usuales.....	16
3	Análisis descriptivo de datos de un portafolio de créditos masivos, generación de reportes de control, definición de indicadores de gestión y de recuperación de cartera.....	18
3.1	Análisis de datos.....	18
3.2	Generación de reportes de control	23
3.3	Definición de indicadores de gestión y de recuperación	25
4	Desarrollo de un modelo de score de comportamiento.....	31
4.1	Definición del objetivo del modelo	34

4.2	Construcción del Modelo	39
4.2.1	Resultados	41
4.2.2	Validación.....	42
5	Implementación en la gestión de recuperación.....	45
6	Conclusiones y Recomendaciones.	48
7	Anexos	51
7.1	Anexo 1: Análisis descriptivo de datos.....	51
7.2	Anexo 2: Modelo de Regresión Logística	73
7.3	Anexo 3: Código de programación SAS	88

1 Introducción: Características de los portafolios masivos, principales productos en el mercado, relación riesgo rentabilidad, el ciclo de crédito.

1.1 Crecimiento de los créditos de consumo en Ecuador.

La dolarización en Ecuador en el año 1999, fue un punto de partida crucial en el desarrollo del Sistema Financiero y de Retail en el país, y al tiempo en que se estabilizó la economía, empezó a tomar mucha fuerza la colocación de créditos masivos. Anteriormente era muy complicado el poder generar créditos en montos pequeños, por los costos operativos y las altas tasas de interés que se manejaban, por ende el rendimiento no era el esperado.

La dolarización abrió el camino para que muchas empresas vean un nicho de negocio interesante, al generar altos volúmenes de colocación con una tasa de interés fija y un mercado ávido de consumir. Los primeros en detectar esta oportunidad fueron las casas comerciales, que vendían sus productos a crédito con un alto margen financiero, así como también las tiendas por departamentos (ropa, artículos personales, menaje) que comenzaron a generar tarjetas de crédito cerradas para compras a crédito en sus locales.

La reacción de la Banca y las instituciones financieras no se hizo esperar, con lo que se agregó nuevos competidores al mercado, básicamente en créditos directos y tarjetas de crédito a todos los sectores de la población, pero en particular a los segmentos que tradicionalmente no estaban siendo atendidos.

Así lo podemos evidenciar en los gráficos a continuación, de la evolución del crédito de consumo partir del año 2002.

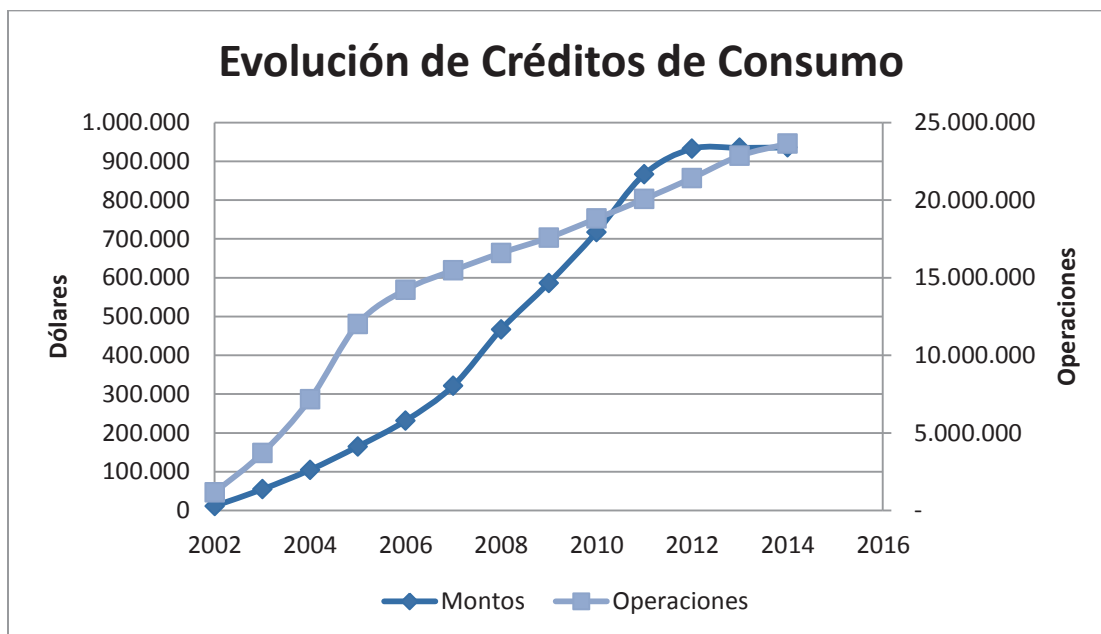


GRÁFICO 1.1 : EVOLUCIÓN DEL CRÉDITO DE CONSUMO EN ECUADOR

Fuente: Superintendencia de Bancos y Seguros

1.1.1 Principales productos en el Mercado

Con la estabilización de precios y la múltiple gama de productos en el mercado nacional e internacional, comenzaron a generarse ofertas de crédito especializadas, las mismas que permitían libertad al consumidor de comprar elementos que nunca podrían pagarlos sin meses o años de ahorro y que cada vez estaban más a su alcance mediante el uso de tarjetas de crédito, que permitieron adquirir desde zapatos deportivos de marca, muebles, electrodomésticos y hasta comida, tanto en los locales comerciales como por internet, también se han desarrollado productos de crédito más especializados para adquirir autos, remodelaciones de vivienda, estudios, electrodomésticos, tecnología,

etc., acordes al nivel socioeconómico, plazos, condiciones y de colaterales, como se muestra en la siguiente tabla.

	Condiciones			
Producto	Inicial	Plazo	Monto Máximo USD	Colateral
Tarjetas	N/A	Rotativo 24 meses	25.000,00	Quirografaria
Autos	25%	48 meses	40.000,00	Prendaria
Consumo	N/A	36 meses	20.000,00	Quirografaria
Hipotecario	30%	240 meses	150.000,00	Hipotecaria

Tabla 1.1 Productos tradicionales de la Banca

Fuente: Políticas Generales de Crédito Banco Pichincha.

1.2 Relación Riesgo Retorno

Normalmente se piensa que en la administración de Riesgos, el objetivo es minimizar la pérdida por Riesgo Crediticio, lo cual puede ser un error desde el punto de vista que el mejor riesgo es aquel que no es tomado, es decir, se puede cumplir el objetivo dejando de generar crédito (en el extremo de los casos) o peor aun perdiendo oportunidad en la colocación y en la participación de mercado, por lo que de acuerdo con Lawrence¹, en el que indica que “la existencia del equilibrio adecuado entre el riesgo y el retorno, y la optimización de los beneficios es mejor que simplemente minimizar las pérdidas”

El objetivo general de cualquier negocio es obtener ganancias que cumplan los objetivos de gestión y genere beneficios a sus accionistas. En el caso del crédito de consumo no es minimizar los préstamos incobrables, sino hacer todo lo posible para generar un equilibrio entre el riesgo (pérdida) con el retorno (rentabilidad). Un buen indicador que permite comprender este concepto, se basa en analizar el rendimiento del producto de forma integral, es decir, tomando los beneficios totales relacionados con la

¹ Lawrence David, Managing a Consumer Lending Business

colocación del crédito y comparándolas con las pérdidas que se han generado, fruto de los incobrables. Recordando que la relación riesgo/retorno es individual de cada producto, algunos productos son más riesgosos que otros por varios motivos, entre los cuales podemos destacar, el mercado objetivo al que fueron orientados, el plazo, los colaterales, entre otros. Por ejemplo la tasa de pérdida en tarjetas de crédito oscila entre de 2 y 5%, mientras que en los créditos hipotecarios varia de 0.2 a 1% o en autos entre un 1 y 3%, si solo analizáramos esta variable, probablemente nos veríamos tentados a direccionar el crédito al de menor pérdida, sin embargo, la tarjeta de crédito tiene otros atributos (cobros por mantenimiento, comisiones por transacciones, comisiones a establecimientos) que supera ampliamente el rendimiento del producto hipotecario.

Pérdida Promedio	
Producto	90+
Tarjetas	2,89%
Autos	2,46%
Consumo	2,46%
Hipotecario	1,57%

Tabla 1.3 Pérdida por producto en mora 90+, no se considera ejecución de garantías

Fuente: Unidad de Riesgos Personas Banco Pichincha, análisis mensual de portafolio dic 2014.

En general para cada producto se puede calcular una tasa mínima de rentabilidad basada en el precio que cobra por cada producto, en Ecuador al ser fijada por norma la tasa máxima de crédito por destino de la operación es mucho más complejo estructurar un producto, pero es mucho más fácil el definir el costo relacionado con los incobrables.

En resumen se puede agrupar esta relación bajo el siguiente análisis de la combinatoria de los posibles cruces entre riesgo retorno.

- Alto Riesgo - Bajo Retorno: Créditos a largo plazo sin colaterales.
- Bajo Riesgo - Bajo Retorno: Créditos para remodelaciones de viviendas con colateral de la hipoteca, créditos hipotecarios para vivienda.
- Alto Riesgo - Alto Retornos: Créditos de Consumo, Tarjeta de Crédito, principalmente segmentadas a clientes de estratos socioeconómicos altos (Platinum, Black, Zafiro), tarjetas de crédito privada, emitidas por tiendas por departamentos, farmacias, ropa, etc.
- Bajo Riesgo-Alto Retorno: Créditos para autos, Tarjetas de Débito.

En los últimos 15 años, las instituciones que colocan crédito han incorporado dentro de sus estrategias y procesos estadísticos para crecer su exposición y participación de mercado, con las herramientas matemáticas y la capacidad de las computadoras de procesar datos en la actualidad, se puede predecir el comportamiento con un alto grado de confianza. El término general utilizado para este efecto es el scoring. El reto se ha concentrado en obtener la suficiente información de sus clientes o sus potenciales clientes para que puedan ser calificados y obtener la relación de la cantidad de buenos versus la cantidad de malos, conocida como ODDS, según su comportamiento.

En la actualidad y fruto del desarrollo de modelos para la originación de créditos, recibimos diariamente ofertas de las instituciones relacionada con productos crediticios, tarjetas de crédito, préstamos de consumo o la utilización de los productos que ya tenemos con el banco, un re tanqueo de un crédito antiguo, incrementar el consumo en la tarjeta mediante promociones o avances en efectivo, promociones de incrementos de cupo por temporada, etc., y que cada vez son más agresivas en relación a los montos y plazos propuestos. Sin embargo, muy poco se ha desarrollado el concepto de la administración integral de los créditos en su ciclo de vida, en muchas instituciones todavía se considera únicamente el tema de la originación como el único parámetro a evaluar en los resultados de comportamiento de la cartera.

El ciclo de crédito detallado a continuación², provee las etapas que se deberían implementar en la administración para poder controlar de forma integral el riesgo generado en la colocación de un crédito, si bien es cierto está basado en el negocio de consumo, con ciertos ajustes se puede extrapolar a todo tipo de crédito destinado a las personas.



GRAFICO 1.2: CICLO DE CRÉDITO

Fuente: Managing a Consumer Lending Business

Debido a que el enfoque de este trabajo está relacionado con la cobranza, rápidamente revisaremos cada etapa del ciclo y en los capítulos posteriores profundizaremos en la etapa de recuperación:

² Lawrence David, Managing a Consumer Lending Business

Planificación:

- Definición de las características del Producto en relación a los términos y condiciones tasa de interés, plazo, tipo de amortización, periodos de gracia, colaterales, pagos mínimos, todo esto orientado al mercado objetivo.
- Definición de la Estructura del Negocio.
- Análisis de Rentabilidad (Riesgo - Retorno).
- Definición de la Estructura organizacional que soporta la operación.
- Elaboración de los Manuales de Políticas y procedimientos para todas las etapas del ciclo de crédito.

Adquisición:

- Definición del proceso de evaluación y aprobación de crédito, determinando la documentación necesaria para la evaluación de los perfiles, construcción e implementación de los modelos de Score de Iniciación, construcción de los procesos de que permitan la validación de Políticas y verificaciones de datos de los clientes solicitantes, algoritmos para el cálculo de la capacidad de endeudamiento, análisis del carácter de pago y evaluación de los colaterales o garantías.
- SLA's (Tiempos de Servicio, Plan de Capacidad).
- Reportes y controles del proceso Integral.
- Dotación del Área (Infraestructura de operación).

Mantenimiento

- Renovaciones (re tanqueos), Upgrades, incrementos de cupos por temporada, bloqueos de tarjetas de crédito.

- Profundización de clientes (Venta Cruzada).
- Soporte a actividades de Mercadeo y Venta.
- Generación de Campañas de Retención y Fidelización.
- Generación de Campañas de Colocación.
- Generación de Campañas de Actualización de Datos.

Recaudación

- Habilitación de puntos de pago.
- Procesos de débitos automáticos y órdenes de prelación.
- Cobranzas
 - Definición del modelo, si se lo realiza internamente o se contrata una casa de cobro.
 - Definición de las etapas de cobranza, asignación de cartera y estrategias.
 - Metas y objetivos de la gestión de cobranza, definición de indicadores de recuperación del portafolio.
 - Definición de reportes para el control del proceso, relacionados con la eficiencia, eficacia y calidad.
 - Dotación del Área (Infraestructura de operación).

Perdida

- Análisis del costo de crédito, en valores de provisiones, pérdidas por castigos, por días mora, por definiciones de cartera irrecuperable, por pérdida de recurso y/o ejecución de colaterales.

Para poder administrar este modelo, es necesario contar con un sistema de información gerencial robusto, que permita ir supervisando cada paso del proceso. El resultado de un paso en el proceso conduce al siguiente paso en el ciclo de crédito. La

administración de un negocio de créditos masivos resume el conocimiento del negocio que comienza con el nuevo ciclo. Esto cubre algunos temas que el administrador debe conocer y la importancia de incorporar suficiente cantidad de buenas operaciones como para compensar las que produzcan pérdidas, controlar los tamaños de las líneas de crédito incentivando el uso de las mismas a los buenos clientes y controlando o cerrando las líneas a los malos clientes.

2 Aspectos generales en la gestión de cobranzas, tecnología utilizada, tipos de gestión, estrategias usuales.

2.1 Aspectos generales.

La gestión de cobranza es una etapa fundamental en la administración de portafolios de crédito masivos, no solo desde el punto de vista del proceso de contacto con el cliente, sino desde la perspectiva operativa de la recaudación, dado que si no se cuenta con las herramientas que permitan un proceso ágil y efectivo, probablemente se genere un desincentivo del cliente para el pago de sus obligaciones. En la actualidad y gracias a la bancarización de personas en Ecuador, los prestamistas intentan concatenar la obligación (crédito) con un proceso automatizado de recaudación para reducir su riesgo, mediante débitos de cuentas en los bancos o a su vez integrando puntos de pago cada vez más asequibles al cliente, sea o no de su red de oficinas o almacenes.

Sin embargo y a pesar de que estos servicios se han incrementado considerablemente, la cobranza por la gran cantidad de endeudamiento en el mercado, exige una premisa: “cobrar más rápido y mejor”; es decir, existe una necesidad imperiosa de generar estrategias que apalanquen un mejor resultado en la cobranza y que permitan adelantarse al resto de acreedores para poder tener la prioridad.

Una operación cae en mora, cuando no llega el pago de la obligación hasta la fecha máxima de vencimiento de la cuota o del pago del estado de cuenta. Es conocido en el mundo crediticio que a mayor edad de mora, la probabilidad de recuperación se reduce, es por esto que se hace necesaria una planificación de las estrategias para evitar que esto suceda.

Las estrategias usuales en la gestión de cobranza se basan en la segmentación de portafolios y la utilización de los canales de contacto con los clientes, generalmente, la base está en los días mora y en el tipo de producto, lo que ha generado criterios para las acciones a ejecutarse en estas etapas, nominándolas generalmente como Administrativas, Prejurídicas (Prejudiciales) o Jurídicas (Judiciales), relacionadas con el tipo de gestión y/o de producto, gestionándolos mediante llamadas telefónicas, visitas de campo, envío de mensajes de texto vía celular, faxes, mails o cartas que permitan obtener una respuesta positiva del cliente para la regularización de sus obligaciones. Es claro comprender que a medida que la edad de mora de un cliente avanza, la probabilidad de recuperación disminuye, por una simple lógica de la forma como se conceden los créditos; esto tiene que ver con la capacidad de pago del cliente ya que la gran mayoría de empresas que otorgan crédito, utilizan el indicador del valor de la cuota sobre el ingreso neto como una variable del nivel de endeudamiento, por ende, con el incremento de los días mora este indicador se multiplica por cada vencimiento de cuota, incluso normativamente está restringido a un valor máximo del 50%³.

Finalmente y como el otorgamiento de crédito es una actividad de Riesgo, se deben tener planes de “normalización” de operaciones, para aquellos casos en los que las condiciones de la originación se han modificado, ha existido un cambio en la economía o ha sucedido un evento fortuito con el prestatario (enfermedad, accidente, pérdida de su trabajo), pero que su carácter de pago le permita un cambio de condiciones en las variables de crédito (plazo, tasa) para su regularización.

³Superintendencia de Bancos y Seguros del Ecuador

2.2 Tecnología utilizada.

El desarrollo de la industria de cobranzas y de los sistemas de comunicación en los últimos años, han permitido que cada vez más la tecnología sea implementada en los procesos de segmentación y contacto con el cliente, basados siempre en la optimización de los recursos.

Existen muchas alternativas en el mercado de soluciones completas para la gestión de Cobranzas, estas plataformas permiten la optimización de tiempos y capacidad instalada, aplicando conceptos utilizados por la cobranza tradicional, pero automatizándolos y generando soluciones robustas para utilizarlos con grandes volúmenes de datos. Los avances más importantes se dan en relación al manejo de la información mediante algoritmos en bases de datos que disminuyen considerablemente el tiempo de procesamiento y la carga de diversas fuentes, tipos y formatos, llegando en algunos casos a disponer de la información en línea de los sistemas transaccionales, así como también las mejoras sustanciales en los procesos de marcación telefónica en el Call Center o la optimización de rutas mediante dispositivos geo referenciados.

En general, la mayoría de los sistemas permite automatizar procesos como:

- Carga de Datos: Las soluciones pueden ser parametrizadas para leer diferentes tipos de archivos y formatos, proveen validadores de carga y procesos automáticos tanto de extracción de la/las fuentes de datos, como de transformación de los mismos para la utilización de la data en los siguientes procesos.
- Segmentación de Cartera: Permiten segmentar la cartera por una o más variables definidas por el usuario, generalmente se segmenta en función del

producto, saldo, días mora o región geográfica. La idea es generar especialización en la gestión de los grupos definidos.

- **Asignación y Estrategias:** Tienen funcionalidades para asignar o distribuir la carga en función estrategias puntuales, definidas ya sea por variables de segmentación o de tipo de contacto con el cliente, además permite generar estrategias relacionadas con los recursos asignados como son porcentajes de saldos, equitativamente por número de operaciones, ordenamiento por variables, etc.
- **Módulos Particulares:** Debido al nivel de especialización, algunas soluciones incorporan módulos adicionales, relacionados con la segmentación o con el canal de contacto, por ejemplo, en el tema de la segmentación existen módulos particulares que manejan la cobranza Judicial o la administración de bienes, de igual manera, se puede observar aplicaciones con módulos puntuales para telefonía o visitas de campo.
- **Reportería:** Generan reportes automáticos relacionados con la administración de la cobranza, basados en las variables de segmentación o de nivel de contacto.

Estas soluciones se las puede adquirir en diferentes modalidades:

- **Compra de Licencia:** Involucra la instalación del software y la implementación del modelo de gestión de cobranza propio, generalmente es la solución más costosa, pero a su vez más controlada para el tipo de gestión a implementar.
- **Alquiler de Licencia:** Puede generarse con la instalación del software o a su vez con el envío de la data al proveedor. Sea físicamente o en la nube, pero con la particularidad de que es parametrizable únicamente a la funcionalidad propia del sistema, sin ningún tipo de personalización.

Modelo conceptual:

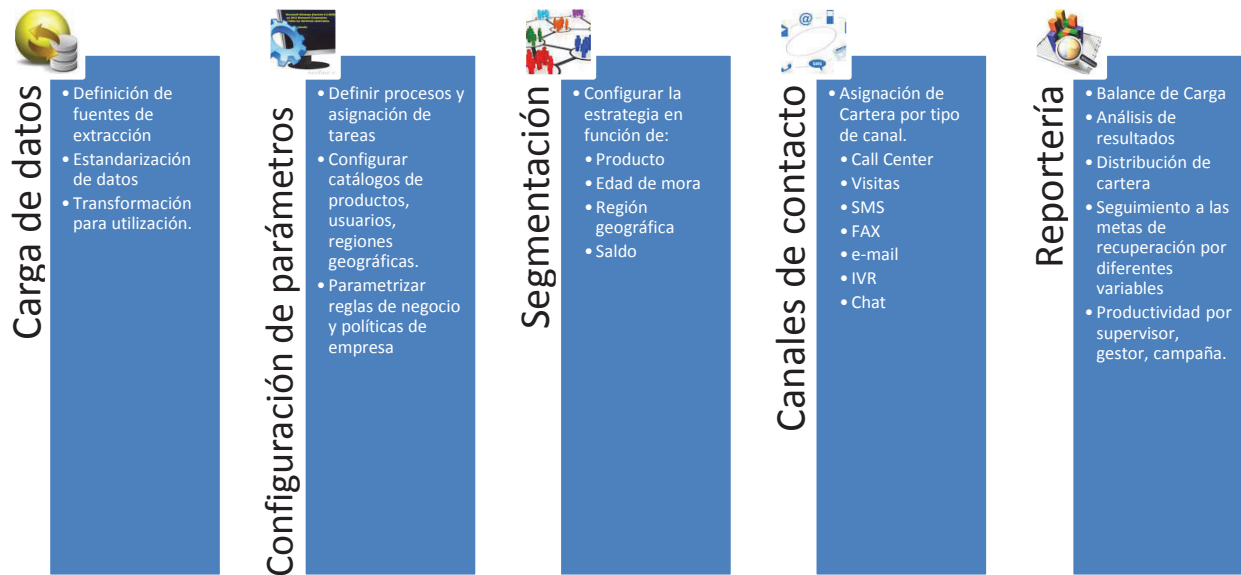


GRÁFICO 2.1: Modelo Conceptual de Sistemas de Cobranza

Fuente: Autor.

El disponer de una herramienta tecnológica sofisticada, no garantiza el éxito de la operación, depende mucho del modelo conceptual definido para la segmentación y los canales de contacto, así como también de la estructura funcional que apalanque las estrategias definidas, en resumen los resultados exitosos se obtienen cuando se logra plasmar la estrategia con la ejecución.

2.3 Tipos de Gestión

La gestión de cobranza está relacionada con la estrategia, los canales de contacto y la segmentación, por ello es importante definir claramente las acciones sobre estas tres variables para obtener los resultados.

Si analizamos los canales de contacto, debemos considerar el nivel de impacto sobre el cliente, tomando en cuenta la forma como le llega el mensaje, ya que existen múltiples alternativas, llamadas telefónicas, visitas, mensajes de texto al celular, e-mail o chat en línea. Cada una de ellas tiene un impacto diferente y un costo diferente para la empresa, es por ello que se planificar la estrategia de las acciones a ejecutarse y las resumimos en el siguiente cuadro.

Canal de Contacto	Costo	Interacción	Impacto
Telefonía Gestor	Medio	Alta	Alto
Telefonía IVR	Bajo	Baja	Bajo
Visita	Alto	Alta	Alto
e-mail	Bajo	Media	Bajo
SMS	Medio	Baja	Medio
Chat	Bajo	Media	Alto

Tabla 2.1: Matriz Canal de contacto por Costo interacción e impacto

Fuente: Autor

Adicionalmente, las respuestas a la gestión deben ser consideradas para el siguiente paso, ya que es información valiosa para asegurar que la misma sea efectiva. Para este efecto, se definen los denominados “árboles de gestión” que vienen incorporados en las herramientas de cobranzas y que son parametrizables por tipo de empresa. Un árbol de gestión permite la parametrización de los diálogos, acciones y reportería de la herramienta, por ejemplo en el caso de la gestión telefónica con un operador, el árbol a parametrizar incorpora acciones desde el proceso de marcación hasta el manejo de las objeciones por parte del gestor. Con esto se logra una estandarización tanto en los

procesos de llamadas como en los resultados de la gestión y se obtienen indicadores para cada uno de los canales de contacto. A continuación un ejemplo de un árbol de gestión para el canal de telefonía.

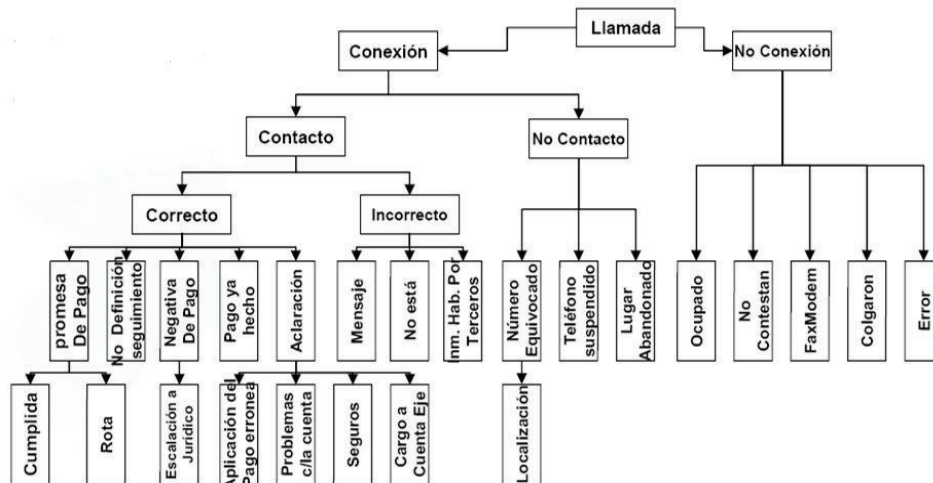


GRAFICO 2.2: EJEMPLO DE ÁRBOL DE GESTIÓN

Al igual que en los otros tipos canal de contacto para la gestión, la información de cada ramal provee indicadores que permiten validar los procesos involucrados en la gestión y administrar la estrategia tanto en cantidad de recursos como en la especialización de la gestión por las variables de segmentación.

2.4 Estrategias usuales

Como se mencionó en los párrafos anteriores las estrategias están enfocadas en la segmentación y el canal de contacto, estas se realizan en intervalos de tiempo o como campañas, definiendo la frecuencia (intensidad) y los objetivos a alcanzar. Parte del modelo conceptual de la gestión integral de la empresa, donde claramente debe estar definida la segmentación y los canales de contacto a utilizar y decanta en la operativa durante el intervalo de tiempo a utilizar.

Lo mayor parte de empresas de cobranzas utilizan segmentación por estas tres variables, edad de mora, producto y región geográfica y destina los canales de recuperación en función del costo y el nivel de impacto. Adicionalmente incorpora una variable de priorización en la gestión que puede ser relacionada con el monto del cliente. Para clarificar los conceptos ejemplificamos un modelo de gestión, segmentado por edad de mora y producto, entonces, la macro estrategia sería:

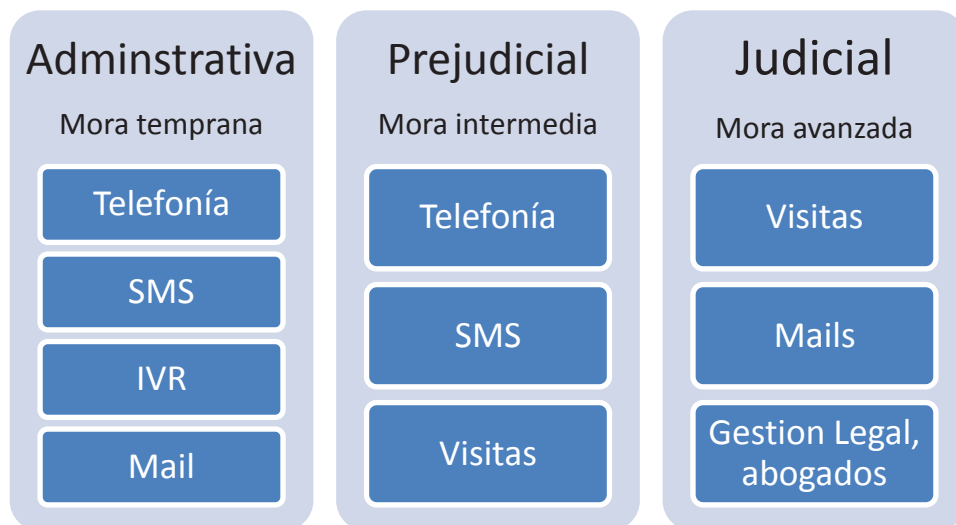


GRAFICO 2.3: ESQUEMA DE UN MODELO GENERAL DE GESTIÓN DE COBRANZAS

Fuente: Autor

Una vez definida la estrategia global, se define la estrategia operativa, para el segmento de cartera administrativa en mora temprana, está definido el canal de contacto, si incorporamos la variable producto, por ejemplo crédito a sola firma y tarjeta de crédito, se podría definir una estrategia para una semana de la siguiente manera:

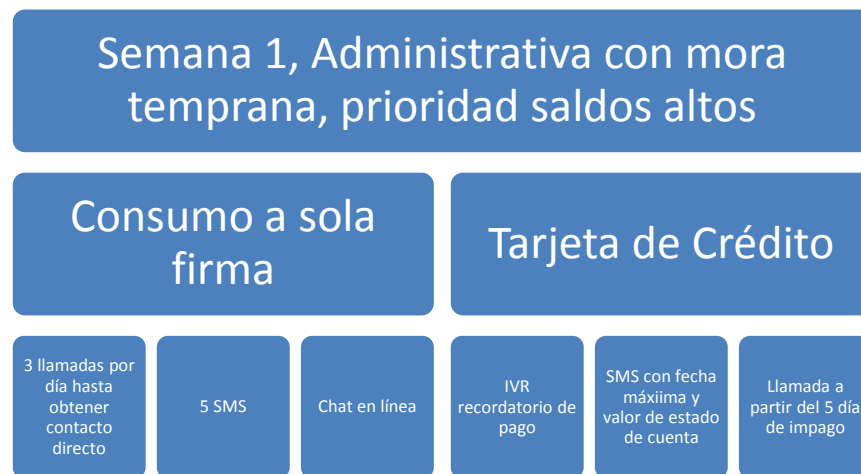


GRAFICO 2.4: EJEMPLO DE ESTRATEGIA DE COBRO

Fuente: Autor

Es importante analizar la variable por la cual se prioriza la gestión, dado que se pueden cometer errores en esta planificación, la gran mayoría de créditos son generados en “cuotitas”, algunas de las empresas comerciales que no manejan el concepto de la amortización, generan estrategias relacionadas con los montos en cuotas y no se analiza el Riesgo Total, esto produce que el portafolio en general deteriore y no exista una administración sobre la obligación completa.

3 Análisis descriptivo de datos de un portafolio de créditos masivos, generación de reportes de control, definición de indicadores de gestión y de recuperación de cartera.

3.1 Análisis de datos

La característica esencial en los portafolio de créditos masivos el volumen de operaciones que se generan, por lo tanto el análisis de datos es fundamental para tomar decisiones en tendencias y generar pronósticos, que por la homogeneidad de las operaciones son altamente confiables.

La información que genera la cobranza es muy rica en diferentes aspectos, pero lo más importante es que en general, las variables que son transmitidas por quien origina la cartera (cedente) son validadas previamente, ya que obedecen normalmente a estructuras estandarizadas por los organismos de control o a su vez pasan por un proceso de revisión de su consistencia para la publicación de los balances contables. Es por ello que la calidad de los datos relacionada con variables del producto, son el insumo primario para generar la segmentación y las estrategias, sin embargo existe otro conjunto de variables que en oposición a lo anterior son de muy mala calidad, como las direcciones y teléfonos de contacto, así como las variables socio demográficas y las que resultan de la captura de los sistemas de gestión cuando no están definidos claramente los árboles y las acciones para cada ramal, ya que en la gran mayoría de negocios los procesos crediticios y de cobranza, adolecen de controles para la captura, validación e ingreso de datos en los sistemas transacciones e incluso son actualizados frecuentemente, sin guardar el histórico o la consistencia en la validación.

Bajo esta premisa, las variables con las que se cuenta netamente son de comportamiento y de producto, por lo que es importante definir una metodología que permita generar una segmentación que utilice únicamente este tipo de variables y a futuro combine la estrategia con las variables de menor calidad.

El proceso de cobranza al primer día de asignación de carga debe segmentarse para la asignación de la estrategia, este proceso se basa en distribución de las cuentas a los diferentes gestores, en función fundamentalmente de la carga operativa que puede administrar. Por ejemplo, para un gestor telefónico que se estima que en un día puede hacer máximo 120 llamadas manuales a un promedio de 3 minutos por llamada ($120 \times 3 / 60 = 6$ horas laborales) se le podría asignar como máximo 1200 operaciones para el mes, si se define una intensidad de 2 llamadas en promedio para cada cliente en un mes de 20 días de gestión y que llame al 100% de los clientes.

Sin embargo esto que parece tan simple, se torna complejo al momento de definir la estrategia, ya que no todos los clientes reaccionan con 2 llamadas o a su vez el nivel de contacto no es efectivo en diferentes portafolios y edades de mora, así como también existen pagos espontáneos que podrían reducir la carga.

Para esto, las mejores prácticas recomiendan generar un modelo de segmentación, que permita el enfoque en los diferentes productos, regiones y edades de mora, basados en el comportamiento de estas variables, generalmente se utiliza un árbol de decisión en función del promedio de la cartera como se muestra en el ejemplo:

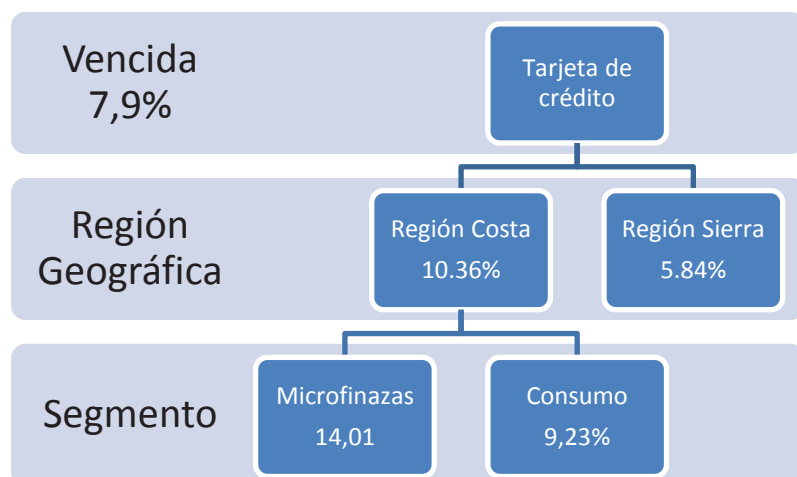


GRAFICO 3.1 : EJEMPLO DE MODELO DE SEGMENTACIÓN

Fuente: Autor

La correcta segmentación es el mejor apalancamiento para la definición de la estrategia, como lo mencionamos en el párrafo anterior, las variables de comportamiento son las más confiables y por ende las que deberían servir en primera instancia para la segmentación. En general si realizamos un análisis descriptivo de datos nos enfocamos en el proceso de recuperación, vemos que la relación de pagos con la edad de mora es inversamente proporcional, es decir, a mayor edad de mora, menor la recuperación.

Edad de mora	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC	Promedio
01. 1 A 30	87,75%	85,56%	90,14%	88,67%	87,40%	89,66%	87,44%	87,38%	89,05%	88,92%	89,78%	86,09%	88,15%
02. 31 A 60	77,05%	74,28%	78,80%	79,25%	80,23%	78,09%	73,26%	78,52%	79,40%	78,07%	81,68%	72,50%	77,59%
03. 61 A 90	50,78%	52,46%	49,50%	52,98%	52,08%	58,43%	52,83%	53,92%	52,37%	47,01%	52,38%	38,93%	51,14%
04. 91 A 120	27,48%	29,55%	32,28%	26,95%	30,37%	31,42%	33,23%	33,15%	31,01%	31,88%	34,53%	26,46%	30,69%
05. 121 A 150	21,27%	16,18%	18,09%	20,10%	18,42%	18,79%	20,37%	16,86%	17,32%	19,09%	19,22%	16,31%	18,50%
06. 151 A 180	13,00%	11,14%	14,34%	11,04%	14,96%	15,01%	17,73%	12,35%	16,14%	21,28%	11,53%	11,75%	14,19%

Tabla 3.1: Porcentajes de recuperación por banda de mora

Fuente: Unidad de Riesgos Banco Pichincha, 2014

Este resultado invita a plantear la segmentación primaria en función de la edad de mora. Para ello y utilizando teoría de procesos Estocásticos y el concepto de las matrices de transición podemos adaptar esta metodología, tanto para la segmentación como para el tema de seguimiento, dado que nos permite controlar el deterioro día a día, la recuperación y controlar el proceso de gestión con variables adicionales.

La utilización de este concepto nos permite representar las probabilidades de transición en n pasos:

$$P^{(n)} = \begin{array}{c|cccc} \text{Estado} & \mathbf{0} & \mathbf{1} & \mathbf{2} & \dots & \mathbf{M} \\ \hline 0 & p_{00}^{(n)} & p_{01}^{(n)} & p_{02}^{(n)} & \dots & p_{0M}^{(n)} \\ 1 & p_{10}^{(n)} & p_{11}^{(n)} & p_{12}^{(n)} & \dots & p_{1M}^{(n)} \\ 2 & p_{20}^{(n)} & p_{21}^{(n)} & p_{22}^{(n)} & \dots & p_{2M}^{(n)} \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ M & p_{M0}^{(n)} & p_{M1}^{(n)} & p_{M2}^{(n)} & \dots & p_{MM}^{(n)} \\ \hline \end{array}$$

Tabla 3.2: Matriz de transición en n pasos

Fuente: Autor

Bajo esta idea general y como la gran mayoría de los créditos de consumo son generados en cuotas mensuales, se construye la matriz de transición de las operaciones con sus días de mora mes a mes, teniendo en cuenta que existe operaciones que van a cambiar en el transcurso del mes y que aparecen sin morosidad (al día) al momento de la asignación.

		Mes actual								
Mes anterior	Estado	Al día	1-30	31-60	61-90	91-120	121-150	151-180	...	TOTAL
		Al día	$\text{Al día}_{t-1}/\text{Al día}_t$	$\text{Al día}_{t-1}/1-30_t$
	1-30	$1-30_{t-1}/\text{Al día}_t$	$1-30_{t-1}/1-30_t$	$1-30_{t-1}/61-60_t$	100%
	31-60	$31-60_{t-1}/\text{Al día}_t$	$31-60_{t-1}/1-30_t$	$31-60_{t-1}/31-60_t$	$31-60_{t-1}/61-90_t$	100%
	61-90
	91-120
	121-150
	151-180

Tabla 3.3: Representación de la transición por edad de mora

Fuente: Autor

Con esta forma matricial de representar los cambios mensuales (se puede revisar a diario) del estado de la cartera, claramente se observa los estados de mejora del portafolio o de deterioro, analizando simplemente a partir de la diagonal de la matriz, todo lo que está a la izquierda de la diagonal (matriz triangular inferior) sería un concepto de recupero y el complemento, un tema de deterioro. Para ser más precisos, la diagonal indica un proceso de recuperación relacionado con la contención hacia un rango de mora superior, en cambio, un salto de dos rangos hacia abajo se considera una recuperación de más de una cuota, que generalmente se la conoce como liberación, por el concepto asociado a las provisiones de cartera por edad de mora.

		Mes actual								
Mes anterior	Estado	Al día	1-30	31-60	61-90	91-120	121-150	151-180	...	TOTAL
		Al día	$\text{Al día}_{t-1}/\text{Al día}_t$	$\text{Al día}_{t-1}/1-30_t$
	1-30	$1-30_{t-1}/\text{Al día}_t$	$1-30_{t-1}/1-30_t$	$1-30_{t-1}/31-60_t$	100%
	31-60	$31-60_{t-1}/\text{Al día}_t$	$31-60_{t-1}/1-30_t$	$31-60_{t-1}/31-60_t$	$31-60_{t-1}/61-90_t$	100%
	61-90
	91-120
	121-150
	151-180

$$\begin{array}{rcl}
 \text{Contención} & + & \text{Liberación} = \text{Recuperación} \\
 \text{Recuperación} & + & \text{Deterioro} = 100\%
 \end{array}$$

Tabla 3.4: Definición del indicador de Recuperación en función de la Matriz de transición

Fuente: Autor

Entonces, una vez que se estudian las matrices de transición de los portafolios, se puede definir las segmentaciones por banda de mora, dado que cuando existen grandes volúmenes de datos podemos asumir que las proporciones generadas se asemejan a las probabilidades condicionales y por lo cual se podría aplicar la propiedad markoviana⁴, es decir, la probabilidad condicional de cualquier evento futuro, depende única y exclusivamente de cualquier evento pasado y su estado actual es independiente del anterior, depende únicamente de su estado actual. Con esta aplicación podemos definir las estrategias de recuperación, los procesos de control y la segmentación de la cartera.

3.2 Generación de reportes de control

La gestión de cobranza en los productos masivos, involucra un seguimiento diario de las tendencias y pronósticos del portafolio, tanto por los sistemas de información como por las personas que controlan los procesos.

Como definimos anteriormente en la sección 3.1, la utilización del concepto de la matriz de transición, permite establecer las estrategias diferenciadas por banda de mora y a su vez construir reportes diarios de recuperación que ayudan a tomar decisiones oportunas en relación a la estrategia en función de los resultados y las estimaciones de los pronósticos.

⁴ Luis Rincón, Introducción a los procesos Estocásticos.

Los reportes contienen los días hábiles de gestión y la recuperación acumulada por banda de mora, así como un pronóstico del resultado sobre los reales ejecutados. En el ejemplo a continuación vemos un gráfico para el producto tarjeta de crédito en la edad de mora de 31-60 días, con 22 días hábiles para la recuperación, transcurridos 20 días de gestión, así como también una línea de tendencia, como parte del ejemplo se generó una línea de tendencia polinómica de orden 3.

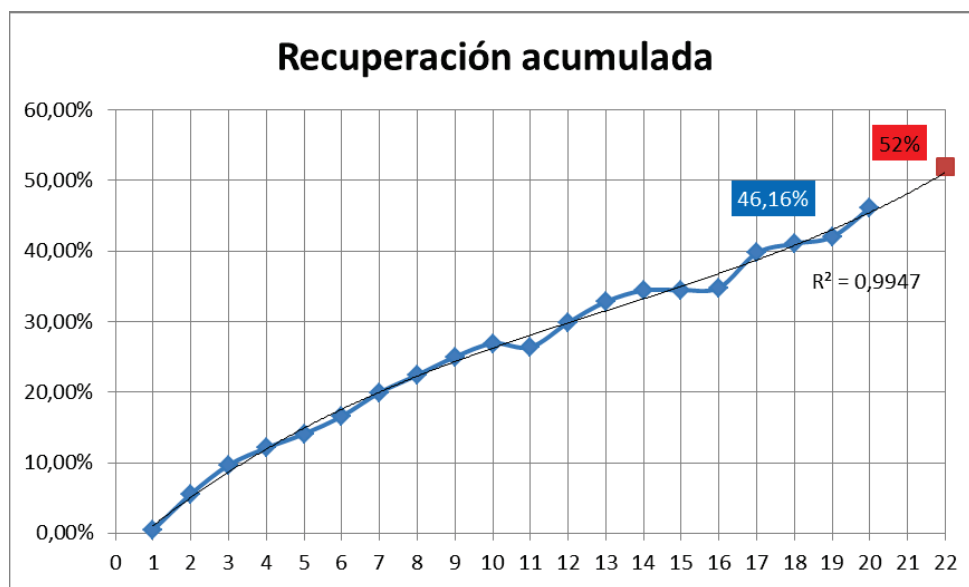


GRAFICO 3.2: EVOLUCIÓN DE LA RECUPERACIÓN

La reconstrucción de los reportes históricos, contribuye en la detección de estacionalidades o comportamientos inusuales al comparar fechas similares, meses de diferentes años o los resultados del mes actual con lo sucedido en el mes anterior, situación que aporta en la generación de alertas tempranas y en la toma de decisiones sobre la estrategia de gestión intra mes.

Finalmente si se desea completar el monitoreo del portafolio, se debe incorporar el concepto de las variables de segmentación para alinearlos a la recuperación, es decir,

incorporar variables de producto, región geográfica o alguna dimensión de la segmentación, bajo el mismo concepto de la recuperación acumulada por bandas de mora que permitan determinar diferencias en los procesos de recuperación y enfocar la estrategia.

3.3 Definición de indicadores de gestión y de recuperación

Los indicadores de gestión están relacionados con el proceso realizado en el canal asignado, por lo que la definición de los mismos se basa en las respuestas al árbol construido. Como en el ejemplo mostrado en 2.3, para la gestión telefónica, los indicadores nacen a raíz del resultado de la llamada, sin embargo, es importante que se disponga estadísticas desde lo macro a lo micro. Intuitivamente, lo primero que se debe garantizar en la operación, es que toda la cartera sea gestionada, por lo tanto, el primer indicador obedece a la cantidad de clientes gestionados, que se lo conoce como “barrido de cartera” y se lo calcula como el número de casos gestionados sobre el número de casos asignados.

$$\%barrido = \frac{\#casos\ gestionados}{\#casos\ asignados} * 100\%$$

Basados en el ejemplo del árbol de gestión, es importante revisar la calidad del nivel de conexión, que permita analizar la estrategia de mejor hora mejor día, manejo de agenda, actualización de información por número de contacto o reporte de avería en la línea, se lo mide como la proporción entre el número de llamadas conectadas sobre el número de llamadas totales.

$$\%tasa\ de\ conexión = \frac{\#conexiones\ efectivas}{\#llamadas\ totales} * 100\%$$

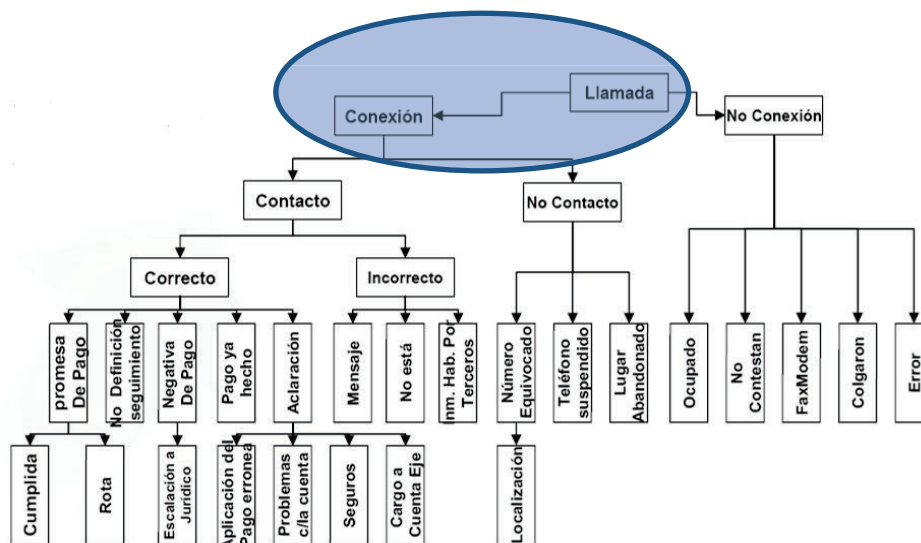


GRAFICO 3.3: CÁLCULO DE LA TASA DE CONEXIÓN

Un nivel más abajo, se puede determinar el nivel de contacto con el titular de la deuda, esta tasa de contacto es una de las variables más importantes en la estrategia de gestión, ya que a mayor nivel de contacto directo, la recuperación y/o la negociación se incrementa. Este indicador se lo obtiene como el número de contactos directos obtenidos sobre el total de contactos realizados.

$$\% \text{tasa de contactos directos} = \frac{\# \text{contactos directos}}{\# \text{contactos}} * 100\%$$

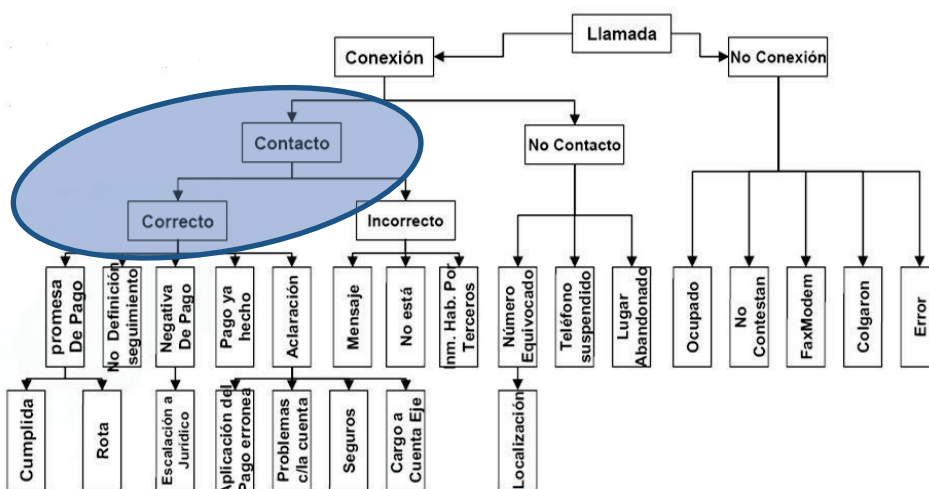


GRAFICO 3.4: CÁLCULO DE LA TASA DE CONTACTO DIRECTO

Al continuar con el proceso y como se puede observar en este punto del ramal, se evalúa las competencias del gestor para la negociación, aquí es importante medir la capacidad del gestor para llegar a un acuerdo de pago, a través de una promesa, el complemento del árbol permitirá tomar otro tipo de decisiones que optimicen el flujo. El indicador clave a medir es el porcentaje de promesas de pago realizadas (eliminado los otros ramales de errores operativos), medido como la cantidad de promesas de pago sobre el número de contactos directos realizados.

$$\% \text{ tasa de promesas} = \frac{\# \text{ promesas conseguidas}}{\# \text{ contactos directos}}$$

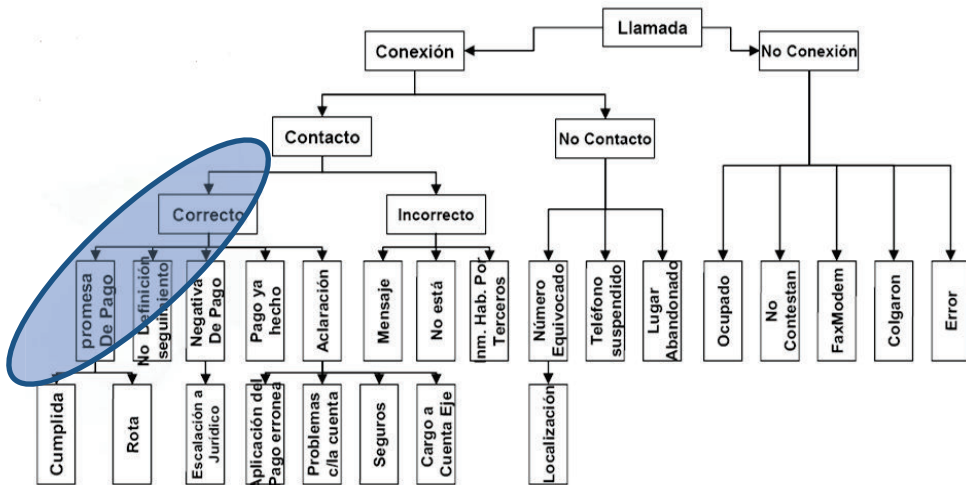


GRAFICO 3.5: CÁLCULO DE LA TASA PROMESAS DE PAGO

Finalmente el resultado del árbol de gestión se cierra con la recuperación efectiva, mediante el cumplimiento de las promesas de pago y el análisis del causal de las promesas rotas. El indicador de cumplimiento de promesas permite mejorar las competencias del gestor, así como también planificar la estrategia en función de los

históricos, este indicador se lo obtiene de la proporción entre el total de los pagos realizados en las fechas acordadas versus las promesas registradas.

$$\% \text{ efectividad} = \frac{\# \text{ promesas cumplidas}}{\# \text{ promesas totales}} * 100\%$$

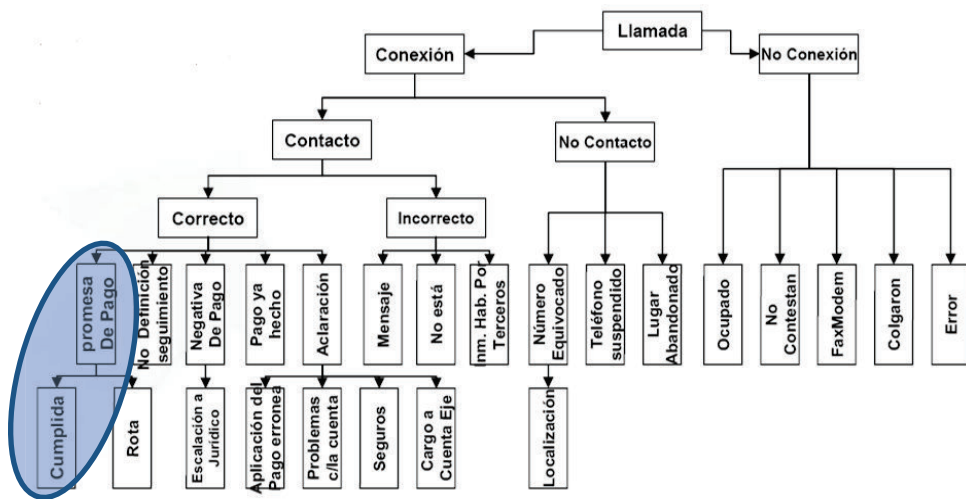


GRAFICO 3.6: CÁLCULO DE LA TASA DE EFECTIVIDAD

Como se puede deducir, la definición del árbol de gestión es crucial para la generación de las estrategias de recuperación en los diferentes canales, sin embargo, la ejecución de la estrategia es en orden inverso, arranca midiéndose desde abajo hacia arriba y determina los valores que deben alcanzar los indicadores para obtener los resultados de la recuperación.

Esta tarea no es fácil de implementar, ya que se necesita un sistema robusto que guarde la información diaria de todos los ramales del árbol. Este tipo de estadísticas se

las puede encontrar a nivel mundial y permiten estandarizar la operación, generalmente son conocidas como EWM (Early Warning Monitor) por sus siglas en inglés.

Los indicadores de recuperación, están más asociados a definir las estrategias que permitan generar el resultado esperado por banda de mora. La industria de cobranzas mantiene algunos estándares de medición dependiendo el producto y el canal de gestión, que se resumen a continuación:

- **Contención:** En valores de saldo recuperado, clientes, operaciones, calculado como el valor que permanece en la diagonal de la matriz de transición, esto indica básicamente que un cliente pagó al menos un dividendo y no deterioró su estado de morosidad o su calificación cuando la normativa coincide con los días de mora de la transición.

		Mes actual							TOTAL	
Estado		Al día	1-30	31-60	61-90	91-120	121-150	151-180		...
Mes anterior	Al día	Al día _{t-1} /Al día _t	Al día _{t-1} /1-30 _t	100%
	1-30	1-30 _{t-1} /Al día _t	1-30 _{t-1} /1-30 _t	1-30 _{t-1} /61-60 _t	100%
	31-60	31-60 _{t-1} /Al día _t	31-60 _{t-1} /1-30 _t	31-60 _{t-1} /31-60 _t	31-60 _{t-1} /61-90 _t	100%
	61-90
	91-120
	121-150
	151-180

Tabla 3.5: Definición del indicador de Contención en función de la Matriz de transición

Fuente: Autor

- **Liberación:** En valores de saldo recuperado o clientes u operaciones, calculado como el valor por debajo de la diagonal de la matriz de transición, permite analizar la recuperación de más de un dividendo, el nombre se adquiere ya que generalmente al recuperar más de una cuota y el cliente cambia de rango en la

matriz, por lo que se liberan provisiones, sin embargo, no en todos los lugares la norma del organismo de control obedece a estos rangos, pero como se mencionó anteriormente, la mayoría de las operaciones de consumo son en cuotas mensuales, en caso particular se podría adaptar la temporalidad de las cuotas a los rangos de morosidad.

		Mes actual							
Estado	Al día	1-30	31-60	61-90	91-120	121-150	151-180	...	TOTAL
Mes anterior	Al día	$\text{Al día}_{t-1}/\text{Al día}_t$	$\text{Al día}_{t-1}/1-30_t$	100%
	1-30	$1-30_{t-1}/\text{Al día}_t$	$1-30_{t-1}/1-30_t$	$1-30_{t-1}/61-60_t$	100%
	31-60	$31-60_{t-1}/\text{Al día}_t$	$31-60_{t-1}/1-30_t$	$31-60_{t-1}/31-60_t$	$31-60_{t-1}/61-90_t$	100%
	61-90
	91-120
	121-150
	151-180

Tabla 3.6: Definición del indicador de Liberación en función de la Matriz de transición

Fuente: Autor

- Efectivo Recuperado: Los dos indicadores anteriores se complementan con este concepto, también relacionado con la matriz de transición, sin embargo, el indicador se lo utiliza más para un concepto de pago de incentivos variables, del punto de partida en el que se asigna la cartera a recuperar, es mucho más fácil para los gestores comprender en un tablero de control, el efectivo que ha sido recuperado de su cartera que los porcentajes de contención y liberación que para llegar a las metas deben ser ponderados.

4 Desarrollo de un modelo de score de comportamiento.

En la administración de créditos masivos, la herramienta más importante que se ha utilizado en los últimos 20 años han sido los modelos de scoring, este método permite mediante la información histórica, realizar pronósticos de comportamientos futuros y provee al administrador una enorme cantidad de información estadística que le permite controlar los portafolios con un alto grado de oportunidad.

Un modelo de scoring se desarrolla con técnicas estadísticas para identificar una clasificación en función de un objetivo de los clientes o la actividad de los clientes. El objetivo se lo define de antemano por el analista y puede ser por ejemplo, el cálculo de la probabilidad de que un cliente pague su obligación, o que tan riesgoso puede ser si se le otorga un crédito, o a su vez que tan rentable resulta un mix de productos. De igual forma se pueden definir objetivos de respuesta de aceptación de producto en una campaña, voluntad de pago a sus obligaciones, voluntad para renegociar si está con problemas de pago, etc. Algunos modelos se han construido para identificar clientes con potencial sobreendeudamiento, o identificar cuentas en deserción, ya sea por no tener movimientos o por patrones transaccionales en detrimento de la relación. Los más utilizados en el mercado identifican a los clientes para aumentos o reducciones de cupos en tarjetas de crédito, de cancelación anticipada o no renovación de plásticos y prospección de nuevos clientes.

Los modelos de scoring en otros países también son utilizados en otros negocios diferentes a los crediticios, se los encuentra en compañías de seguros para establecer un orden de rango para los candidatos por longevidad, buena o mala salud, o para el riesgo de accidentes automovilísticos. Aunque no es publicitado, los sistemas de scoring también son usados en servicios como el pago de impuestos, telefonía, servicios de televisión por cable y hasta en la búsqueda de parejas.

Un modelo de scoring considera muchas variables simultáneamente de tal forma que se pueda definir un patrón donde todos los miembros dentro de un grupo tienen las mismas probabilidades de que el evento se produzca.

Los modelos de scoring tanto de originación como de comportamiento funcionan cuando existen altos volúmenes de datos con valores relativamente homogéneos, por lo que las grandes líneas de crédito o los créditos con hipotecas y otros préstamos de alto valor requieren una revisión más exhaustiva e individual, por cantidad de datos y lo que involucra el otorgar créditos de alta cuantía y con garantías particulares.

Existen muchos beneficios en la utilización de modelos de scoring, como la evaluación objetiva del riesgo a asumir, ya que se eliminan los prejuicios personales en el proceso de decisiones y permite ejecutarlas consistentemente y de forma estandarizada. Así mismo, se genera eficiencia en el costo de procesamiento, dado que se puede detectar perfiles no deseados en las aplicaciones y ser eliminados rápidamente de la gestión. Adicionalmente permite disponer de un control estadístico de los portafolios, dado que el riesgo de cada operación puede ser descrita a un nivel alto de detalle y con alta confiabilidad, con la particularidad de que todos los resultados obtenidos en el tiempo pueden ser contrastados con los resultados de pronóstico de los modelos para asegurar que el scoring siga funcionando como se esperaba.

Normalmente los modelos de scoring permiten generar experimentación controlada, los experimentos se pueden ejecutar para realizar pruebas de Champion-Challenger (Campeón-Retador), en diferentes actividades como ofrecer cupos de más grandes en tarjetas de crédito, para colocar menos requisitos a nuevas operaciones y generar eficiencias en varios procesos por lo que una empresa puede ser más rentable en el largo plazo.

De igual forma, el modelo de scoring presenta algunas limitaciones, como el tiempo que toma el proceso del desarrollo e implementación, por la recolección de datos, el procesamiento de los mismos y las herramientas que permitan las automatizaciones de la implementación. La predicción es limitada, ya que el modelo no puede identificar una buena o una mala operación o cliente, lo que identifica es la probabilidad de que pueda ser bueno o mala, según la definición en la construcción. Además, todos los modelos pierden predicción en el tiempo, debido a varios factores, como los cambios en la economía, la población o mercado objetivo que se evalúa, por lo que los modelos deben ser constantemente monitoreados y evaluados para determinar si sus niveles de discriminación y ordenamiento siguen siendo óptimos.

Se ha comprobado a lo largo del tiempo que los beneficios superan ampliamente las limitaciones, y que la implementación de los mismos genera eficiencias en los procesos, ya que al definir correctamente la característica del evento bueno y malo, la técnica maximiza la distancia entre las distribuciones, permitiendo al administrador tomar mejores decisiones.

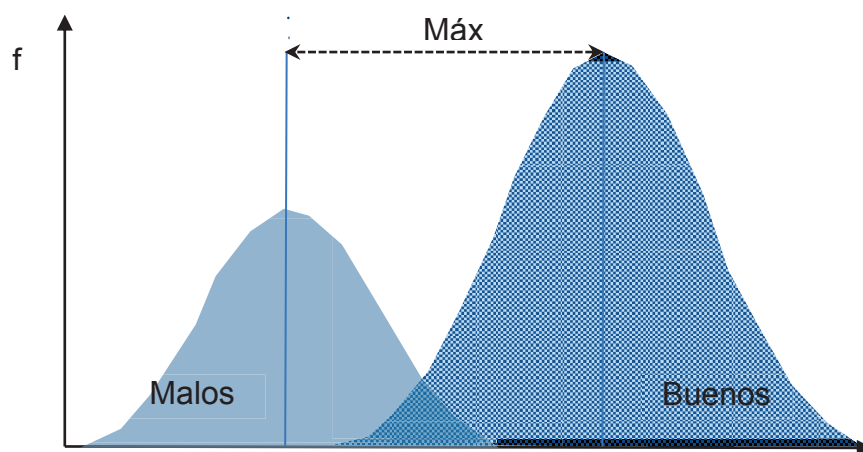


GRAFICO 4.1: MODELO CONCEPTUAL EN LA CONSTRUCCIÓN DEL SCORING

Fuente: Autor

4.1 Definición del objetivo del modelo

Debido al volumen de operaciones concedidas, la idea principal es enfocarse en aquellos clientes que presenten problemas para cumplir sus obligaciones, en general y por el crecimiento de la colocación de crédito en todas las entidades financieras y comerciales, la gestión de cobranza debe estar enfocada en llegar más rápido al cliente y con la mejor estrategia, por lo que se hace necesario la construcción de un modelo de comportamiento que calcule la probabilidad de que un cliente sea rentable con un comportamiento de pago, que no involucre un costo excesivo de recursos por parte de la cobranza, es decir, un buen pagador.

Se analizaron tres alternativas para definir el umbral de bueno o malo (default), en función de la información del portafolio del año 2012, que permita pronosticar un año hacia adelante su comportamiento:

4.1.1 Por Amortización:

El análisis radica en encontrar el punto de mora en el cual la recuperación de Capital + Interés proyectada a un año es menor al Capital Absoluto.

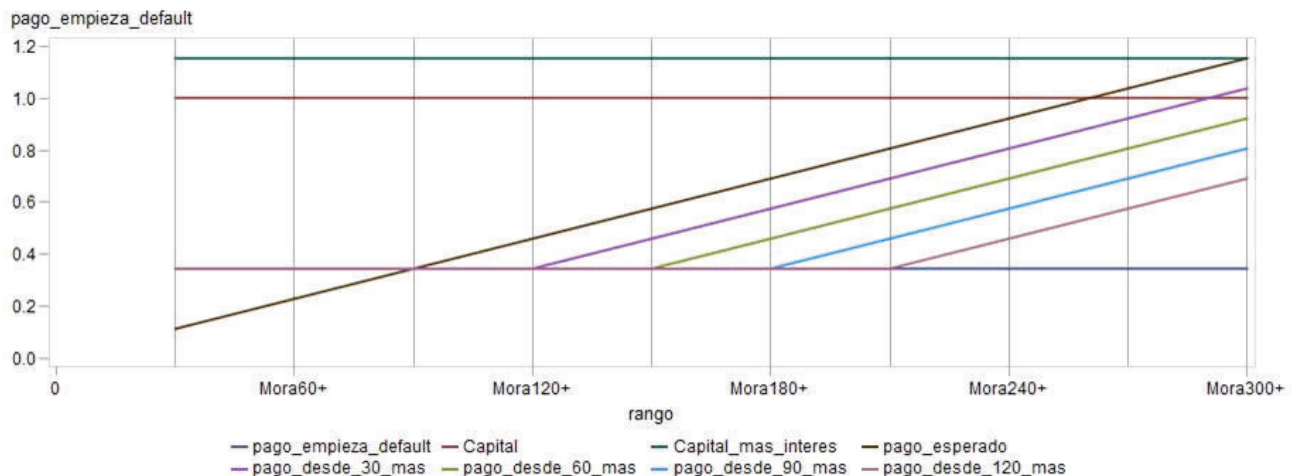


GRAFICO 4.2: ANÁLISIS DE LA MOROSIDAD EN FUNCIÓN CAPITAL MÁS INTERÉS

El resultado de este análisis nos sugiere tomar como default 60+, que significa que todas las obligaciones que llegaron a tener al menos una mora de 60 días o más en la ventana de observación se le asigna una marca fija, independientemente que su morosidad disminuya, estos indicadores se los conocen como “mora ever”.

4.1.2 Por recuperación:

El análisis por banda de mora en la matriz de transición que indica cuando la recuperación es mayor que el deterioro.

Rango	% de Recuperación
02. 31 A 60	80.4%
03. 61 A 90	63.6%
04. 91 A 120	43.3%
05. 121 A 150	41.2%
06. 151 A 180	28.3%
07. 181 A 210	27.6%
08. 211 A 240	23.0%
09. 241 A 270	16.8%
10. 271 A 300	19.2%
11. 301 A 330	16.6%
12. 331 A 360	93.1%
13. MAS DE 360	17.6%
Total general	47.8%

Tabla 4.1: Porcentajes de recuperación por banda de mora

Nota: Se descarta el rango 331 A 360, ya que por una definición tecnológica para Tarjeta de Crédito, los días mora se truncan a partir del día 360.

Bajo este análisis, se determina que un punto de corte para el default debería ser mora 90+.

4.1.3 Por recuperación Acumulada:

El siguiente análisis permite verificar por banda de mora, si la recuperación de la cartera es superior o igual a la de esa banda. El promedio se lo determina como la recuperación acumulada desde la banda de mora 30+. El punto en el cual la recuperación acumulada es mayor que la recuperación promedio se lo toma para determinar el punto de default.

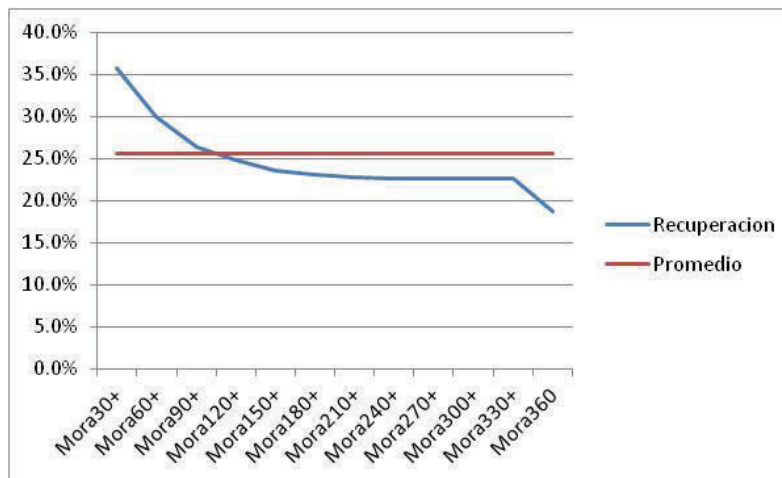


GRAFICO 4.3: ANÁLISIS DE LA RECUPERACIÓN ACUMULADA POR BANDA DE MORA

En este método la recuperación acumulada de mora 120+ es la primera en estar por debajo de la mora promedio, por lo que se define que mora 90+ es el punto adecuado para el default.

En función de estos análisis se toma como definición de default como mora 90+, es decir, todas las operaciones que tengan una marca de mora 90+ se catalogan como “malos” y su complemento al utilizar un modelo binario como “buenos”.

4.2 Definición y extracción de datos para el desarrollo

El score es una herramienta de decisión que permite ordenar jerárquicamente la población, en función de que ocurra el default, en este caso que las operaciones hayan alcanzado en su ciclo, una mora superior a 90 días. Para la construcción de este modelo se define una ventana de observación de 24 meses y una ventana de desempeño de 12 meses.

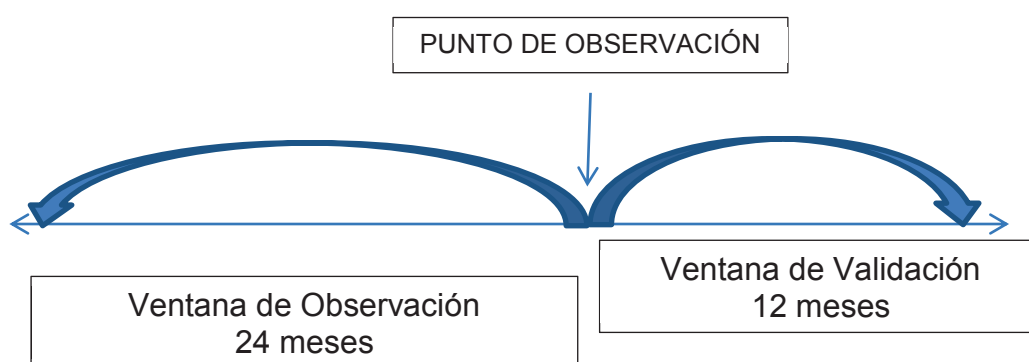


GRAFICO 4.4: DIAGRAMA CONCEPTUAL DE CONSTRUCCIÓN DEL MODELO

Se recogen variables explicativas del cliente que se basan en su comportamiento financiero principalmente con la institución que originó la operación.

Con el fin de evitar estacionalidades, se toma como variables de construcción los cortes a cierre de cada mes en los siguientes períodos: Marzo 2012, Abril 2012, Mayo 2012, Junio 2012, que corresponde a una base de 846.942 clientes.

Y como referente comparativo el real del máximo estado de mora del cliente durante el siguiente año, como lo indicamos en mora 90+, realizando un análisis de correlación

para determinar las variables nominales y categóricas significativas que nos permitan explicar la transición al estado de default.

Las variables que ingresan al modelo son las detalladas a continuación, la validación estadística se las encuentra en el Anexo 1:

Atraso promedio Últimos 3 meses

Atraso máximo últimos 6 meses

Atraso promedio últimos 6 meses

Atraso máximo últimos 9 meses

Atraso promedio últimos 9 meses

Atraso máximo últimos 12 meses

Atraso promedio últimos 12 meses

Número de atrasos mayores 30 días último año

Número de atrasos mayores a 60 días último año

Días mora máximos

Saldo insoluto vencido final

Número de cuota en mora

Numero de cuotas vencidas

Numero pagos en los últimos 3 meses

Número de pagos en los últimos 6 meses

Género

Edad

Estado civil

Provincia de domicilio

Antigüedad laboral dependiente

Nivel de educación

Ingresos

Indicador si actividad económica es independiente

Número de créditos en el Sistema Financiero

Se utilizaron métodos de conteo, evoluciones y trasposiciones para generar las matrices de transición. El modelo se enfoca en determinar los rangos de mora para generar las calificaciones de los clientes que permitirán en la implementación, generar una estrategia ya que la probabilidad de entrar en default es independiente y parametrizable, basado en las calificaciones. Los clientes que alguna vez fueron castigados contablemente se los considera directamente como 1 su probabilidad de default, esto se lo conoce como los malos evidentes.

4.2 Construcción del Modelo

El modelo fue realizado en SAS® E.G. 9.2. y se resume en el siguiente flujo:

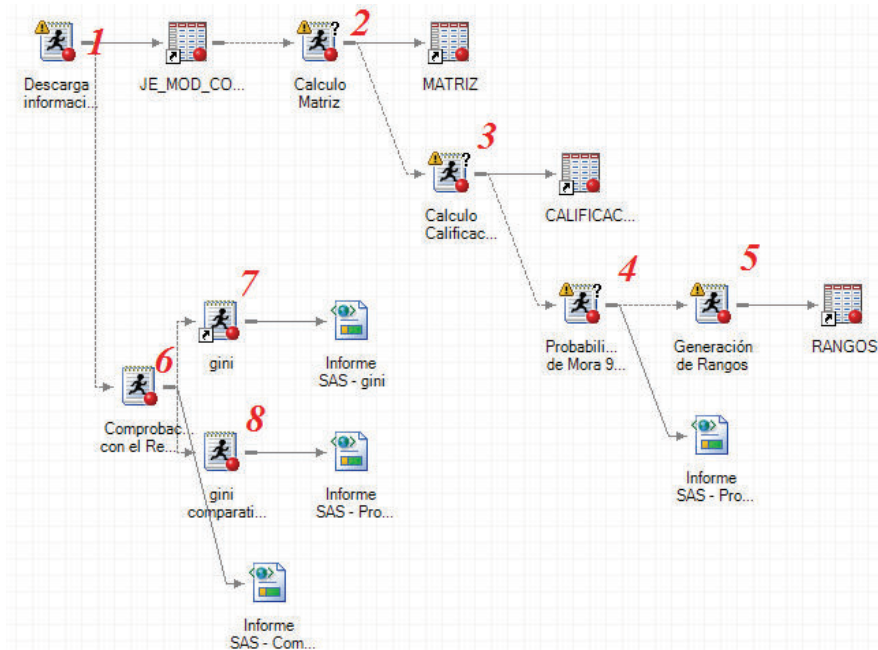


GRAFICO 4.5: DIAGRAMA MODELO ENTREPRICE GUIDE SAS

En el cual:

1. Proceso ETL (extracción, transformación y carga de datos) que trae la información original de los clientes en los cortes señalados.
2. Calcula la matriz de transición (valores entre los cuales se determina el default).
3. Ejecuta la regresión logística para las variables seleccionadas.
4. Otorga un Score a cada Cliente
5. Determina los puntos de corte del Score
6. Validación del modelo con datos reales
7. Grafico del coeficiente Gini para el modelo realizado
8. Grafico del coeficiente Gini para el modelo sin mejoras

Los métodos de regresión logística para una variable cualitativa (dependiente), engloban diferentes tipos de modelos que tratan de explicar y predecir una característica cualitativa a partir de los datos de otras variables, sean estas cuantitativas o cualitativas, que actúan como variables explicativas (independientes). Debido al tipo de variables explicativas que se dispone para el análisis y dado el interés de conocer la probabilidad de que un cliente pague o no pague su obligación en una ventana de tiempo (variable binaria), nos sugiere que el método más aconsejable a utilizar es Regresión Logística, adicionalmente existen múltiples factores que se han considerado y que los destacamos a continuación:

- Se utiliza en el estudio de variables que no tienen una distribución conjunta normal multivariante.
- Permite el uso de variables explicativas tanto cualitativas como cuantitativas. Adicionalmente las variables cuantitativas podrían no ser continuas.
- Ejecuta un doble rol, explicar y predecir los resultados.
- No asume una relación lineal entre la variable dependiente e independientes.
- La variable dependiente no necesita ser normalmente distribuida.

- Se elimina el criterio de homocedasticidad de la variable dependiente para cada nivel las variables independientes (homogeneidad de la varianza).
- No se estudia la normalidad de los términos del error.

Los detalles de la ejecución de la regresión logística⁵, se lo puede revisar en el Anexo 2.

4.2.1 Resultados

Cómo resultados de la ejecución del modelo se obtienen:

- La Calificación determinada por la probabilidad de default para cada cliente, que simplemente etiqueta de datos para comprender mejor el concepto del puntaje, relacionado con el valor promedio de los deciles.

IdentificacionCliente	rangomoraini	p
00001711178275	A-2	5.20%
00001711178283	A-1	4.67%
00001711178374	A-1	4.86%
00001711178390	A-1	6.21%
00001711178465	A-1	5.45%
00001711178705	A-1	5.34%
00001711178754	A-1	7.47%
00001711178812	A-2	6.58%
00001711179125	A-3	18.2%

Tabla 4.2: Asignación de Calificaciones por scoring promedio

- Los puntos de corte de la tabla de validación en deciles, que permitirán tomar las decisiones en función de la tabla de validación, analizando la cantidad de Buenos y Malos (ODDS):

⁵ Iñiguez C, Morales G, Selección de perfiles de clientes mediante regresión logística para muestras desproporcionadas, validación, monitoreo y aplicación en la proyección de provisiones 2009, E.P.N

Decil	De	Hasta
1	-	91
2	92	97
3	98	108
4	109	128
5	129	150
6	151	169
7	170	185
8	186	241
9	242	399
10	400	1.000

Tabla 4.3: Porcentajes de recuperación por boda de mora

4.2.2 Validación

Para la validación del modelo se consideró el comportamiento previsto de los clientes con datos a julio de 2012 y se los comparó con los días de mora reales al cierre de julio 2013.

Con lo cual se realizó la prueba de Gini obteniendo los siguientes resultados para la regresión logística tradicional:

Gini coefficient	
Obs	gini
972	0.35743

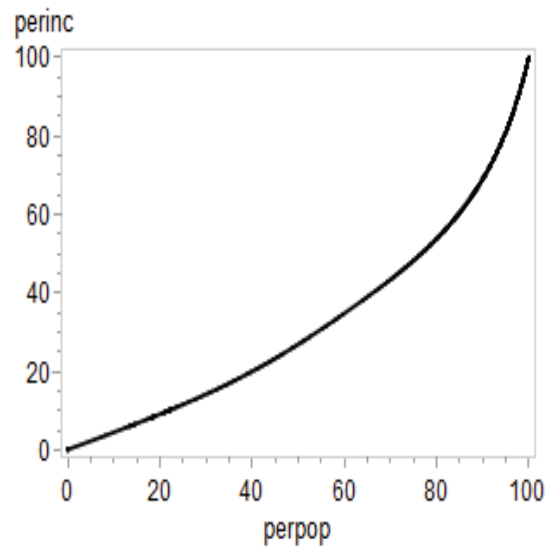


GRAFICO 4.6: COEFICIENTE DE GINI MUESTRA DE CONSTRUCCIÓN

Se ejecutó la misma prueba para el modelo depurado, obteniendo los siguientes resultados:

Gini coefficient

Obs	gini
1636	0.44260

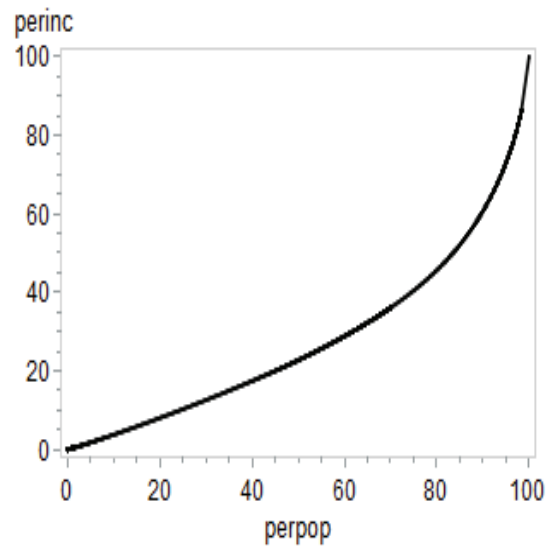


GRAFICO 4.7: COEFICIENTE DE GINI MODELO DEPURADO

Con lo cual se concluye que se gana 8.5 puntos en el coeficiente dando una mejor aproximación con el modelo propuesto.

Adicionalmente para el mismo período se realizó la comparación con la misma definición de default en saldo la siguiente aproximación:

Saldo Estimado Cae A Default	Saldo Real Cae a Default
\$371,242,423.84	\$339,443,332.50

Como se puede notar el valor de la estimación es bastante cercano al ejecutado realmente, lo que confirma que el modelo aporta en la predicción del cálculo del default y por ende la identificación de los clientes malos, que permitirá generar estrategias diferenciadas para su recuperación.

5 Implementación en la gestión de recuperación.

La implementación en el modelo de recuperación se basa en utilizar la variable de scoring como parte fundamental de la estrategia, cruzando los conceptos antes definidos de segmentación y canales de contacto, teniendo en cuenta que el objetivo es recuperar más rápido y con menos costo.

Basado en la tabla de validación del modelo, podemos observar que hasta el tercer decil tenemos una relación de ODDS de 7:1, y que ordena perfectamente a medida que el scoring tiene mayor puntuación, por lo que la estrategia debe estar relacionada con el puntaje por cliente y el canal de contacto para obtener eficiencia tanto en el resultado de la gestión como en costo.

Percentil	De	Hasta	No. Clientes	% Clientes	No. Clientes Buenos	% Buenos (Columna)	% Acumulado Buenos	% Buenos (Fila)	No. Clientes Malos	% Malos (Columna)	% Acumulado Malos	% Malos (Fila)	Odds
1	-	91	106036	12,5%	100.225	17,4%	17,4%	95%	5.811	2,1%	2,1%	5%	17:1
2	92	97	67991	8,0%	61.745	10,7%	28,2%	91%	6.246	2,3%	4,4%	9%	10:1
3	98	108	80504	9,5%	61.789	10,8%	38,9%	77%	18.715	6,9%	11,3%	23%	3:1
4	109	128	86817	10,3%	61.477	10,7%	49,6%	71%	25.340	9,3%	20,6%	29%	2:1
5	129	150	84201	9,9%	58.517	10,2%	59,8%	69%	25.684	9,4%	30,0%	31%	2:1
6	151	169	82640	9,8%	56.164	9,8%	69,6%	68%	26.476	9,7%	39,7%	32%	2:1
7	170	185	85786	10,1%	55.272	9,6%	79,2%	64%	30.514	11,2%	51,0%	36%	2:1
8	186	241	83611	9,9%	50.070	8,7%	87,9%	60%	33.541	12,3%	63,3%	40%	1:1
9	242	399	84697	10,0%	42.024	7,3%	95,2%	50%	42.673	15,7%	78,9%	50%	1:1
10	400	1.000	84699	10,0%	27.311	4,8%	100,0%	32%	57.388	21,1%	100,0%	68%	:1
			846.982	100%	574.594	100%			272.388	100,0%			2:1

Tabla 5.1: Tabla de validación del modelo de scoring

Adicionalmente es necesario adelantarse a la cobranza de otros acreedores, por lo que en los peores puntajes incluso es conveniente intervenir con una gestión “preventiva”.

Como se menciona en la sección 2.4, la definición de la estrategia cruza los canales con la segmentación, es así que en función de la variable de scoring, se pueden definir diferentes estrategias para asignar los recursos y optimizarlos.

La segmentación de la cartera en primera instancia genera un concepto de grado de dificultad de cobro, apalancado en la matriz de transición y en la definición de bueno y malo. Para esta aplicación, se definió el umbral de mora a los 90 días considerando el análisis de la proporción de cartera que se deteriora versus la que mejora, entonces la estrategia de segmentación por edad de mora se define de la siguiente forma:



GRAFICO 5.1: DEFINICIÓN DE LA SEGMENTACIÓN POR EDAD DE MORA

Fuente: Autor

Se han colocado rótulos a los diferentes estados de segmentación, acorde con estándares internacionales de denominación de gestión, “Preventiva” cuando el cliente está al día en sus obligaciones pero se realiza una gestión de cobro de la obligación próxima a vencer, “Administrativa” cuando la probabilidad de recuperación es mayor que la de deterioro y “Judicial” dado que la probabilidad de deterioro es más alta, las acciones de recuperación tienen otra connotación y otro costo asociado.

Una vez definida la estrategia mediante la segmentación debemos integrarla a los canales de gestión y la variable de scoring que permita una optimización tanto en recursos como en costo, retomando el concepto descrito en la sección 2.4, planteamos como ejemplo la siguiente estrategia.

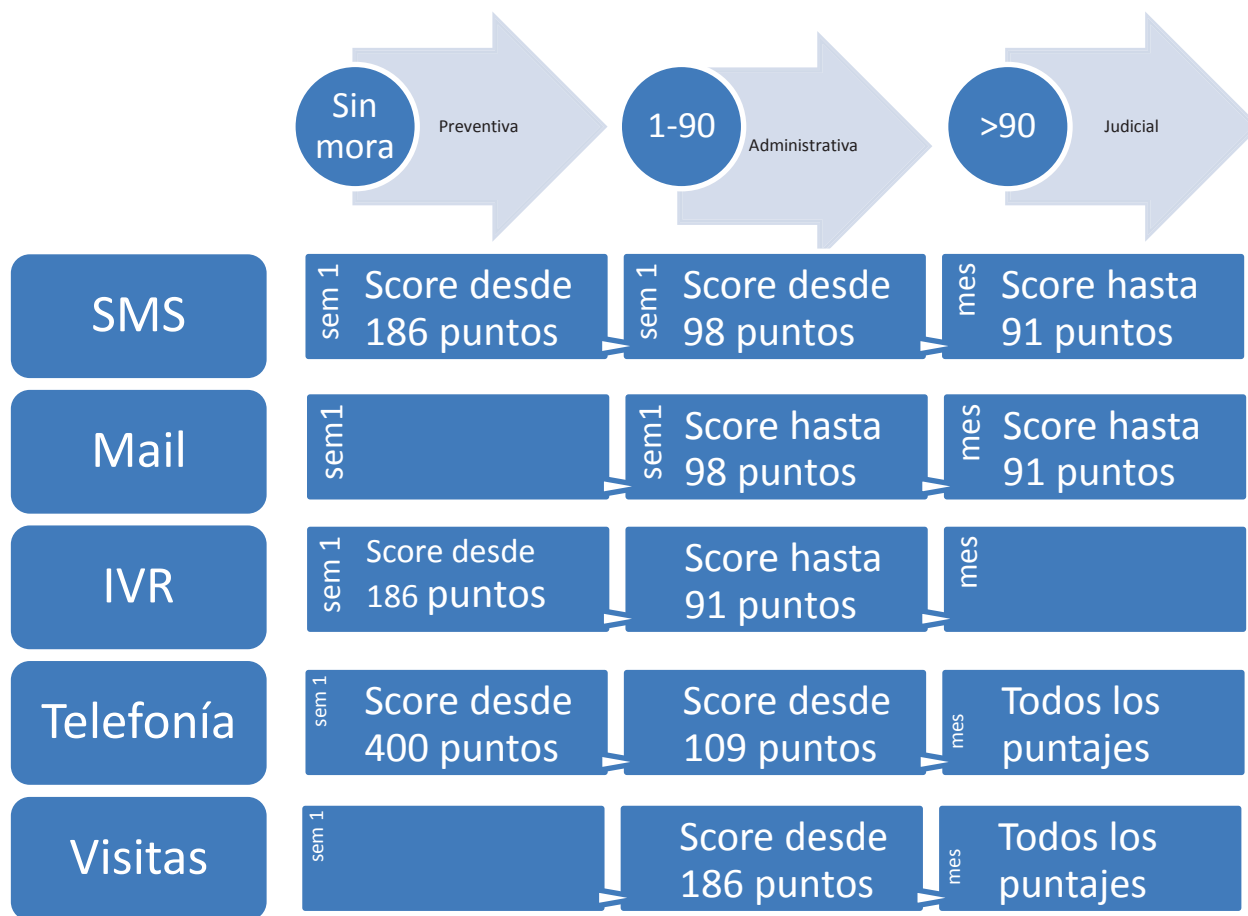


GRAFICO 5.2: ESTRATEGIA DE GESTIÓN POR SCORING

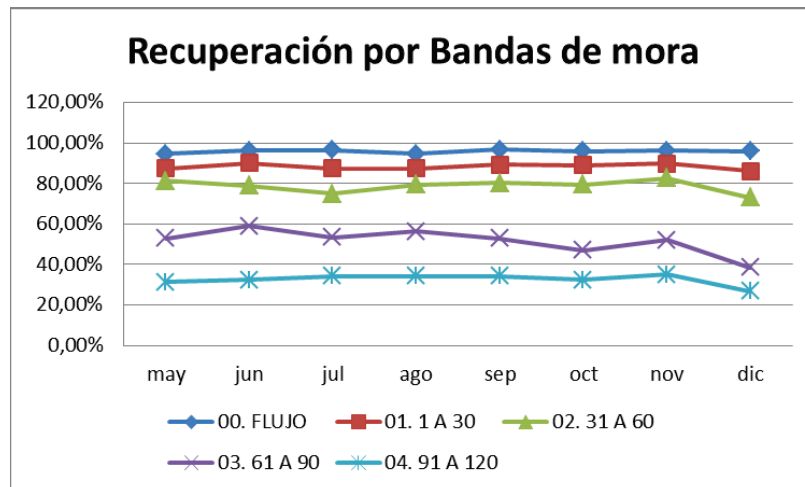
Fuente: Autor

Una vez definida la estrategia, se define los recursos a asignarse, así como la intensidad de gestión.

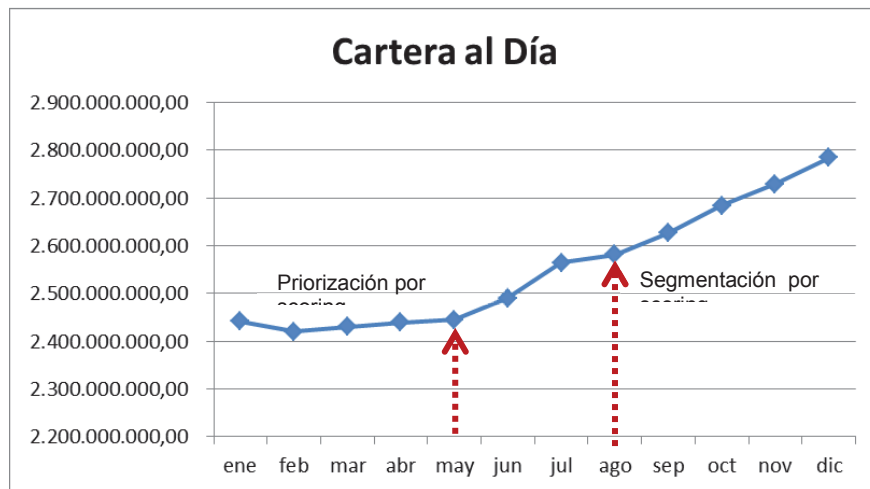
Como se muestra en el esquema anterior, esta sería la estrategia para la semana uno y en función de los resultados obtenidos, se genera la estrategia para la siguientes semanas, incluyendo o excluyendo rangos de score, teniendo en cuenta que la estrategia principal es cobrar más rápido y con mayor intensidad a los clientes determinados como malos, es decir, cuya probabilidad de deterioro sea mayor.

6 Conclusiones y Recomendaciones.

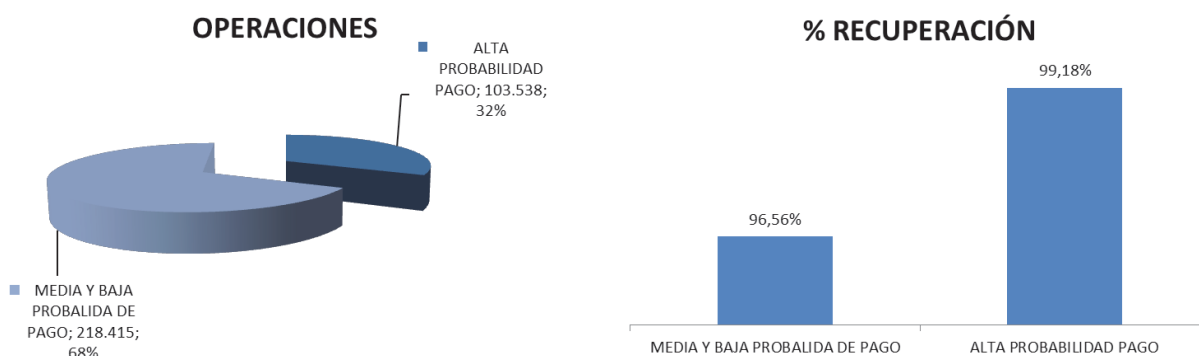
- La segmentación de la cartera para recuperación en función de los rangos de mora, generó estabilidad en la recuperación, así como estandarización de los procesos de gestión.
- Debido a que la recuperación por rangos de mora, evidencia una disminución en sus porcentajes a medida que la edad crece, se pudo determinar estrategias y mensajes diferenciados para cada uno de ellos.



- La incorporación de la etapa Preventiva, permitió adelantarse al cobro de obligaciones de clientes compartidos, cuya preferencia de pago no era la obligación con el banco, llegando a mejorar en 14% la recuperación en dólares, en 6 meses de aplicación del modelo.
- Debido a que el siguiente rango de morosidad es de 1 a 30 días, la intensidad de las llamadas mejoró al contener el número de casos a gestionar con la preventiva y enfocarse en aquellos que presentan un deterioro.



- La incorporación del modelo de scoring dentro de las estrategias de gestión, mejoraron un 7% la recuperación en todas las bandas de mora en una ventana de 4 meses de implementación.
- En edades tempranas de mora, la estrategia permitió una eficiencia en los gestores de cobranza de un 32%, al enfocarse desde el día -7 a gestionar a los clientes malos, dejando por fuera los clientes con alta probabilidad de pago (buenos), lo que generó una mejora tanto en la recuperación como en la relación con el cliente final al no gestionarlo por tener unos pocos días de mora y que pagan espontáneamente.



- Se debe realizar una gestión diferenciada a los clientes con alta probabilidad de recuperación, con el fin de que no afecte los indicadores financieros de la empresa, en cuanto a la generación de ingresos por honorarios de cobranza.
- Se recomienda que para la actualización del modelo, se incorporen variables de gestión, resultado de la implementación de los árboles en los diferentes canales, que permitan la optimización de los recursos, al disparar la mejor gestión por el mejor canal de contacto en función de la probabilidad de recuperación.
- Se recomienda un análisis de costos sobre la cartera catalogada por default como malo, con el fin de calcular la rentabilidad de la recuperación versus el costo asociado a la misma, con la finalidad de sacarla de la gestión a un tercero o a su vez una venta de cartera por parte del cedente.
- Al igual que en el área de Cobranza, existe un enorme potencial de lograr eficiencias en el Ciclo de Crédito, mediante la implementación de metodologías y modelos estadísticos que optimicen el rendimiento de la concesión de operaciones y que no han sido explotados.
- Finalmente, se optimizan los recursos con herramientas que pueda automatizar la segmentación por diferentes variables, la incorporación de modelos de score y cuya reportería permita medir constantemente el resultado de la estrategia, con múltiples integradores que apalanquen la incorporación de funcionalidades adicionales a los descritos como básicos en el software base, software o hardware de marcación, módulos de administración de bienes o control de calidad de datos, etc.

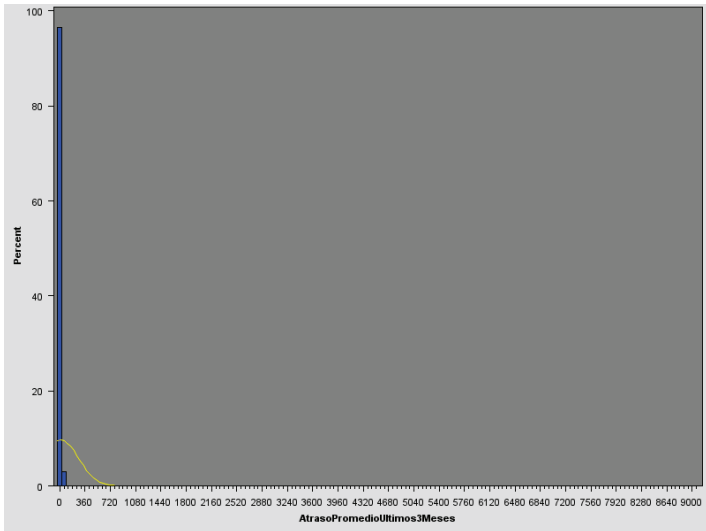
7 Anexos

7.1 Anexo 1: Análisis descriptivo de datos

Atraso promedio Últimos 3 meses

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	19.11542	STD DEVIATION	248.31832
MEDIAN	1.00000	VARIANCE	61662
MODE	0.00000	RANGE	9086
		INTERQUARTILE RANGE	4.00000

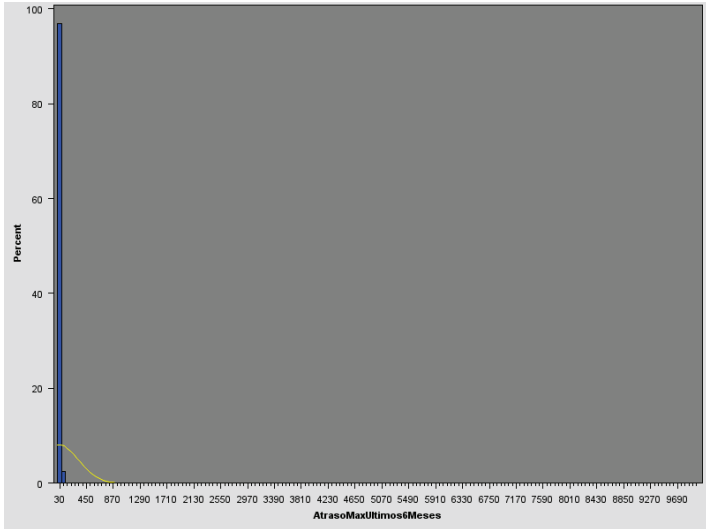
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	142038	11.44	100.00



Atraso máximo últimos 6 meses

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	29.20870	STD DEVIATION	299.79583
MEDIAN	4.00000	VARIANCE	89878
MODE	0.00000	RANGE	9999
		INTERQUARTILE RANGE	12.00000

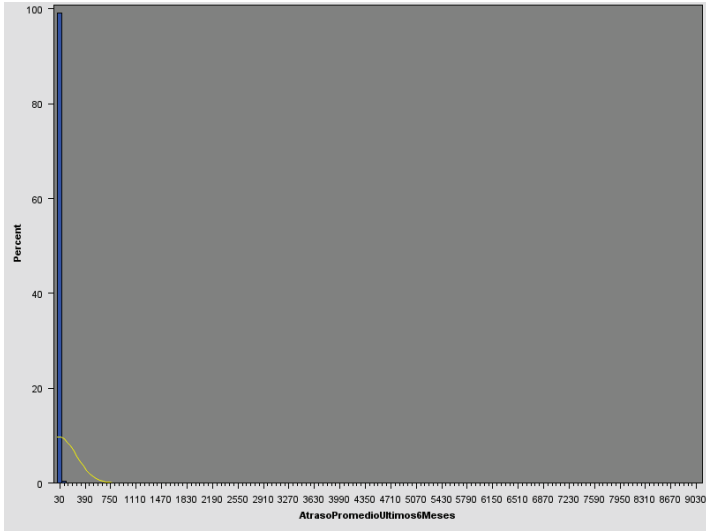
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	141230	11.38	100.00



Atraso promedio últimos 6 meses

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	18.29334	STD DEVIATION	245.75042
MEDIAN	1.00000	VARIANCE	60393
MODE	0.00000	RANGE	9040
		INTERQUARTILE RANGE	4.00000

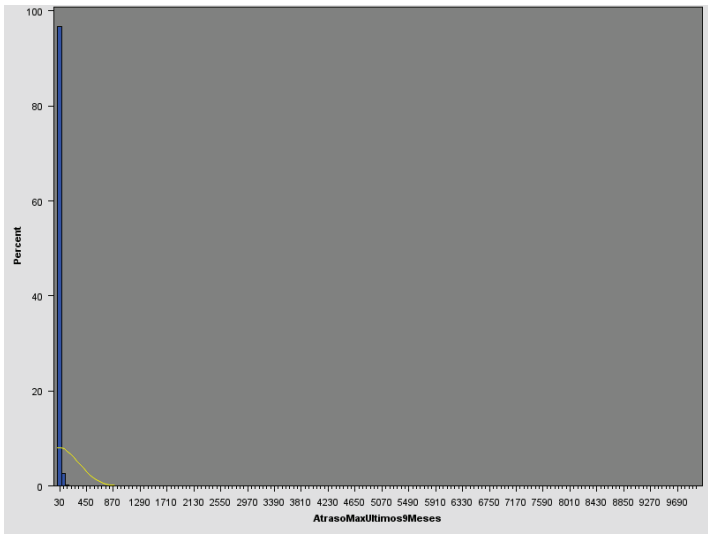
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	142082	11.45	100.00



Atraso máximo últimos 9 meses

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	30.14727	STD DEVIATION	299.75114
MEDIAN	5.00000	VARIANCE	89851
MODE	0.00000	RANGE	9999
		INTERQUARTILE RANGE	13.00000

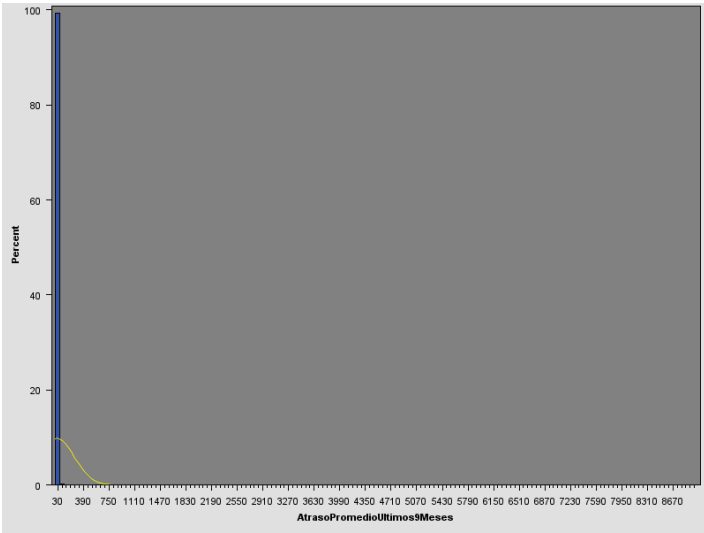
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	141249	11.38	100.00



Atraso promedio últimos 9 meses

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	17.87911	STD DEVIATION	244.28450
MEDIAN	1.00000	VARIANCE	59675
MODE	0.00000	RANGE	8995
		INTERQUARTILE RANGE	4.00000

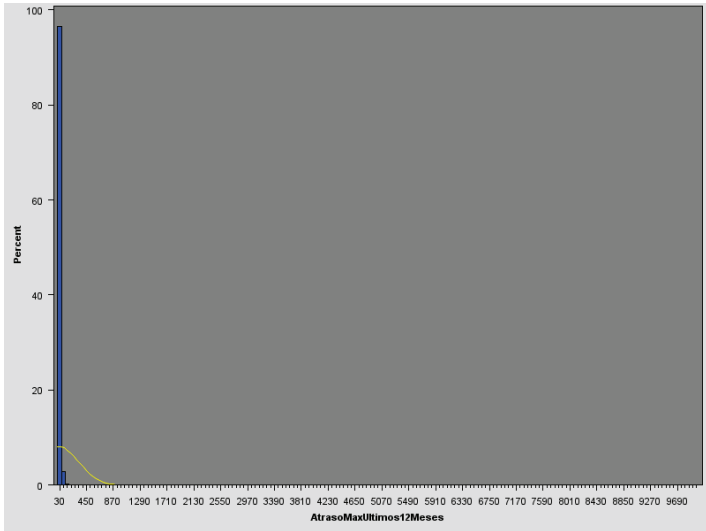
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	142097	11.45	100.00



Atraso máximo últimos 12 meses

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	30.78684	STD DEVIATION	299.66993
MEDIAN	5.00000	VARIANCE	89802
MODE	0.00000	RANGE	9999
		INTERQUARTILE RANGE	14.00000

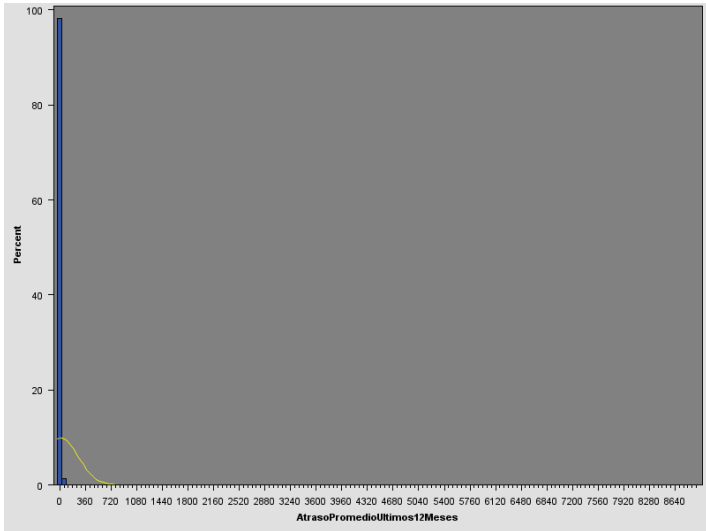
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	141261	11.38	100.00



Atraso promedio últimos 12 meses

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	17.49988	STD DEVIATION	241.86008
MEDIAN	1.00000	VARIANCE	58496
MODE	0.00000	RANGE	8949
		INTERQUARTILE RANGE	3.00000

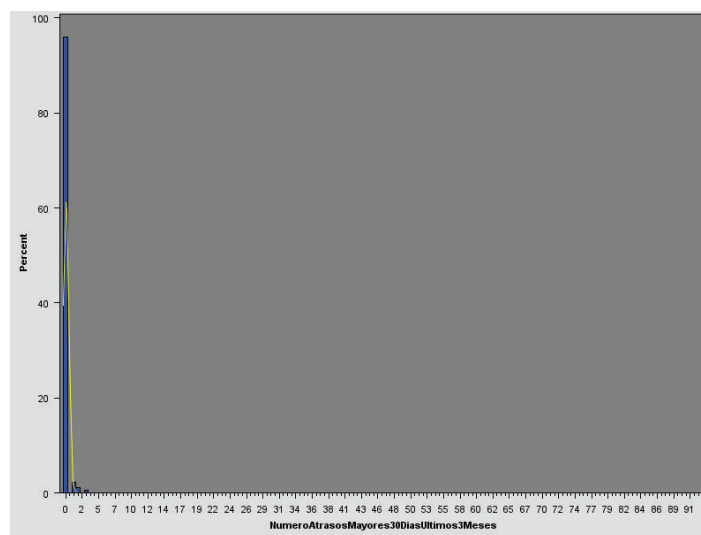
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	142132	11.45	100.00



Número de atrasos mayores 30 días último año

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	0.066718	STD DEVIATION	0.39200
MEDIAN	0.000000	VARIANCE	0.15366
MODE	0.000000	RANGE	93.00000
		INTERQUARTILE RANGE	0

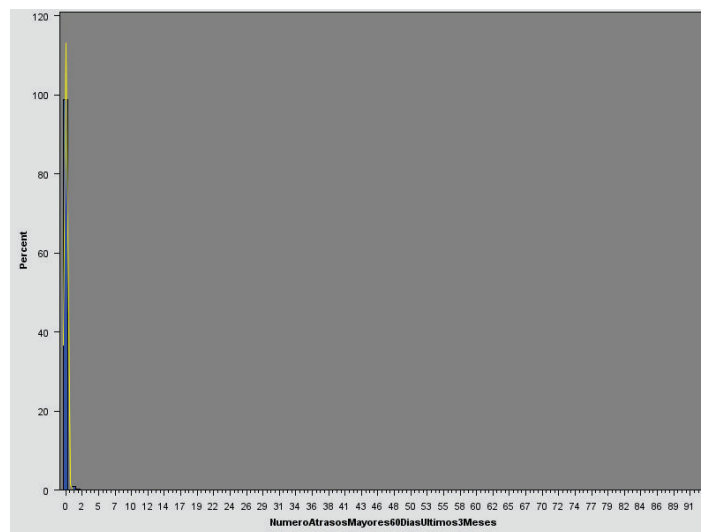
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	147318	11.87	100.00



Número de atrasos mayores a 60 días último año

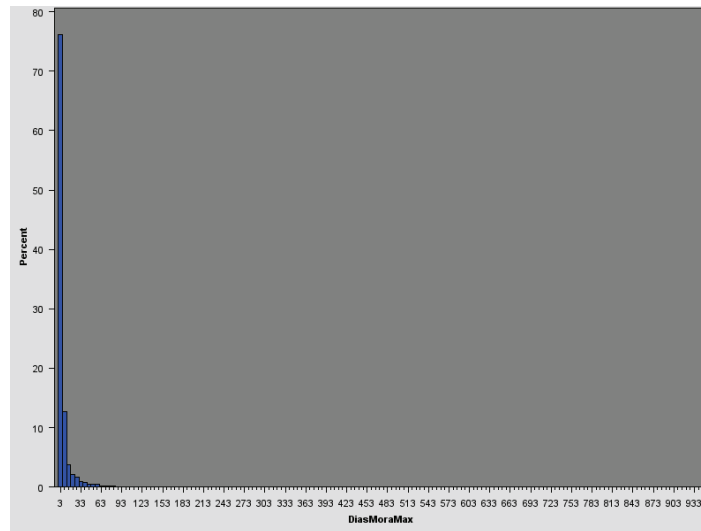
BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	0.017200	STD DEVIATION	0.21162
MEDIAN	0.000000	VARIANCE	0.04478
MODE	0.000000	RANGE	93.00000
		INTERQUARTILE RANGE	0

MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	147316	11.87	100.00



Días mora máximos

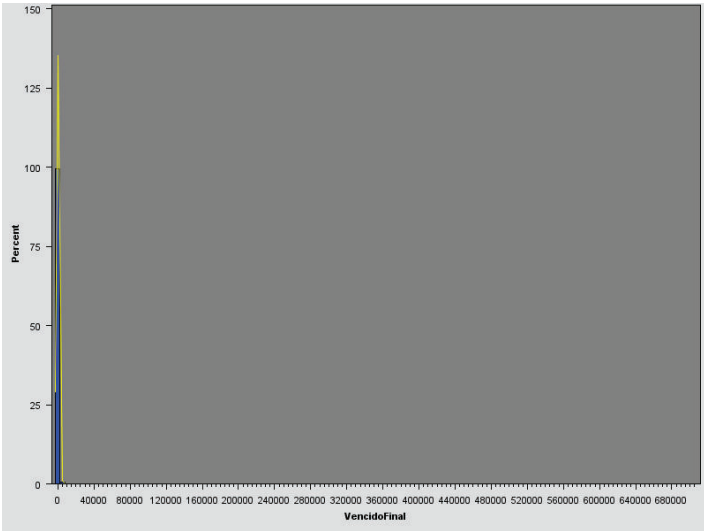
BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	4.879895	STD DEVIATION	11.47998
MEDIAN	0.000000	VARIANCE	131.78984
MODE	0.000000	RANGE	945.00000
		INTERQUARTILE RANGE	5.00000



Vencido final

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	85.95370	STD DEVIATION	1476
MEDIAN	0.00000	VARIANCE	2179064
MODE	0.00000	RANGE	706000
		INTERQUARTILE RANGE	0

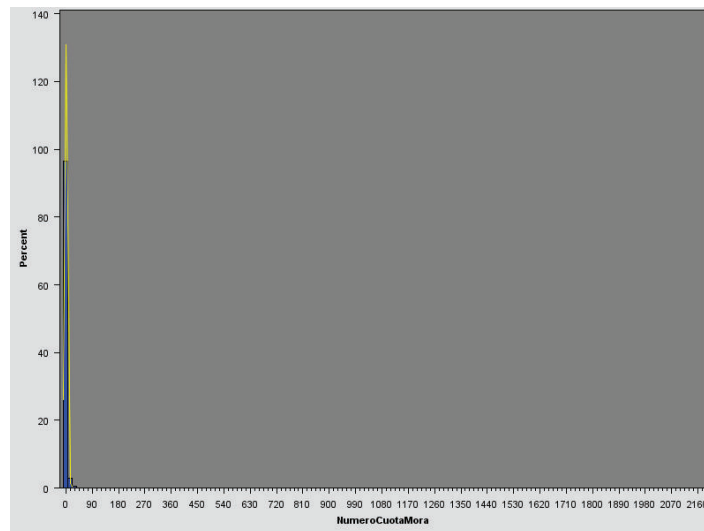
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	94813	7.64	100.00



Número de cuota en mora

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	0.715485	STD DEVIATION	4.57331
MEDIAN	0.000000	VARIANCE	20.91516
MODE	0.000000	RANGE	2180
		INTERQUARTILE RANGE	0

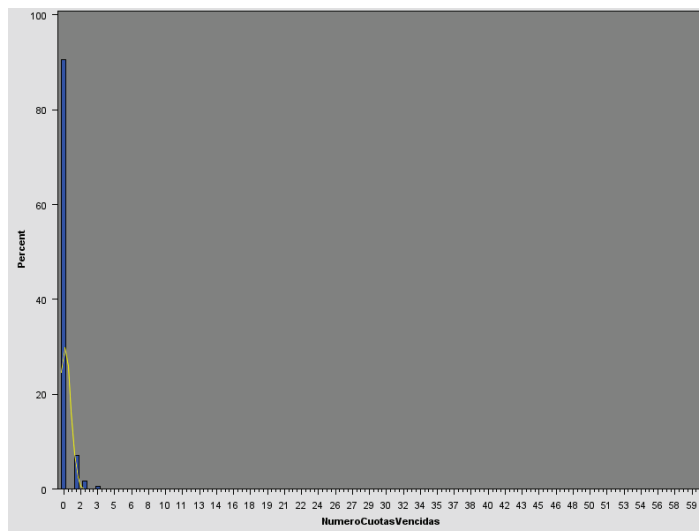
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	147323	11.87	100.00



Numero de cuotas vencidas

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	0.130790	STD DEVIATION	0.53682
MEDIAN	0.000000	VARIANCE	0.28818
MODE	0.000000	RANGE	60.00000
		INTERQUARTILE RANGE	0

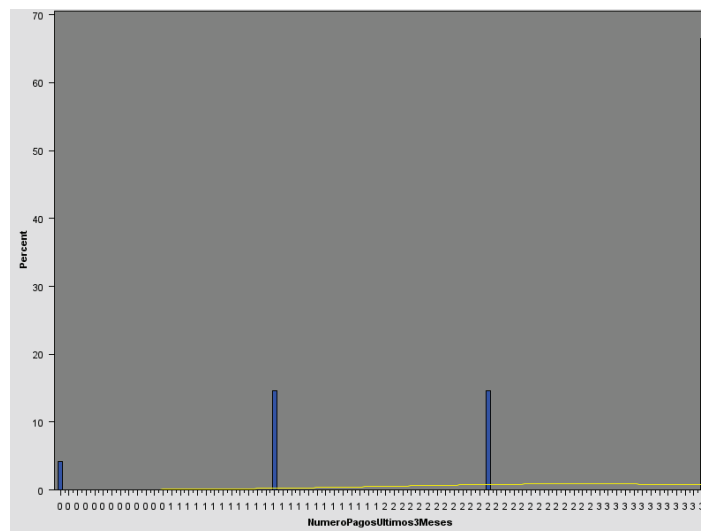
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	147306	11.87	100.00



Numero pagos en los últimos 3 meses

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	2.435946	STD DEVIATION	0.88790
MEDIAN	3.000000	VARIANCE	0.78837
MODE	3.000000	RANGE	3.00000
		INTERQUARTILE RANGE	1.00000

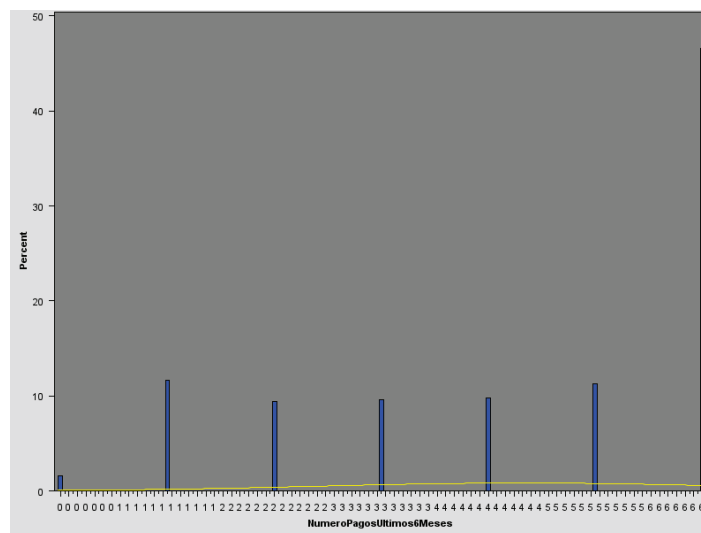
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	147331	11.87	100.00



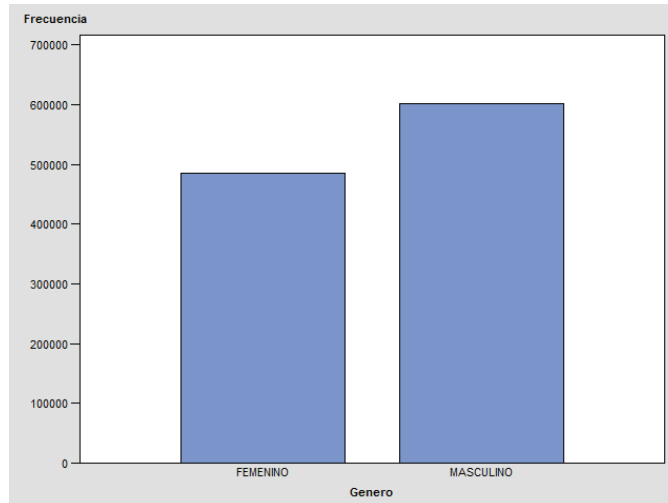
Número de pagos en los últimos 6 meses

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	4.347108	STD DEVIATION	1.90723
MEDIAN	5.000000	VARIANCE	3.63752
MODE	6.000000	RANGE	6.00000
		INTERQUARTILE RANGE	3.00000

MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	147331	11.87	100.00



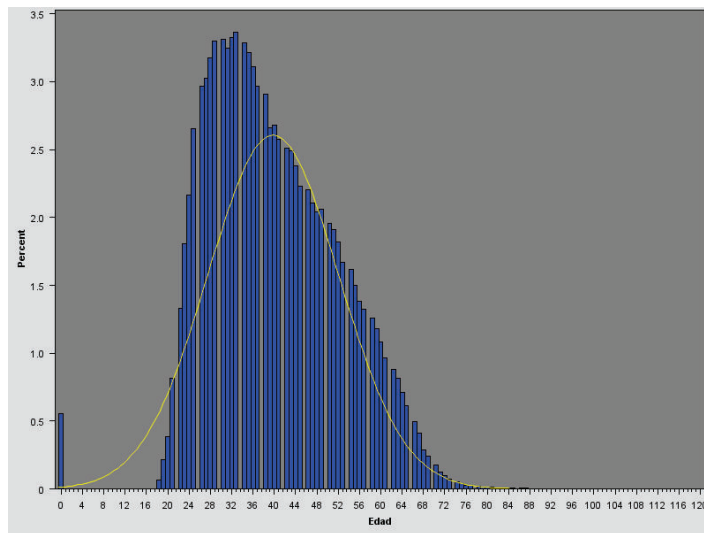
Género



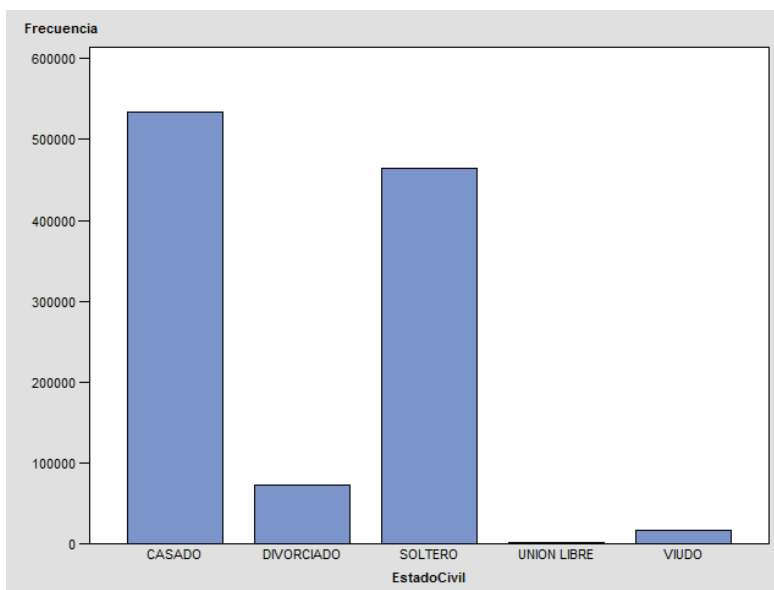
Edad

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	39.80010	STD DEVIATION	12.24922
MEDIAN	38.00000	VARIANCE	150.04331
MODE	33.00000	RANGE	120.00000
		INTERQUARTILE RANGE	18.00000

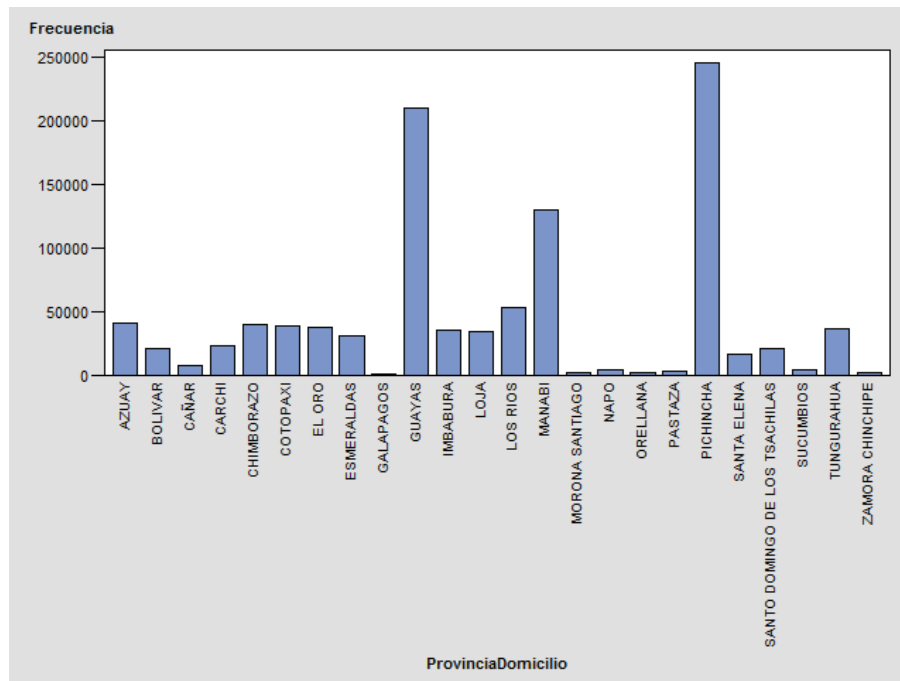
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	147323	11.87	100.00



Estado civil



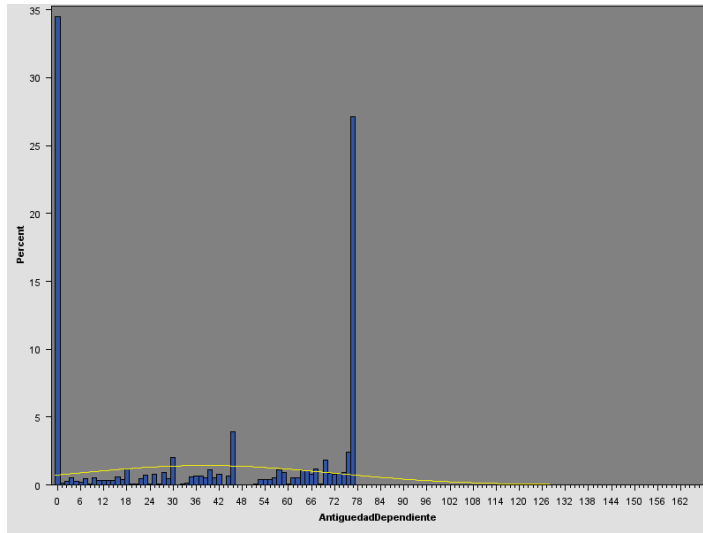
Provincia de domicilio



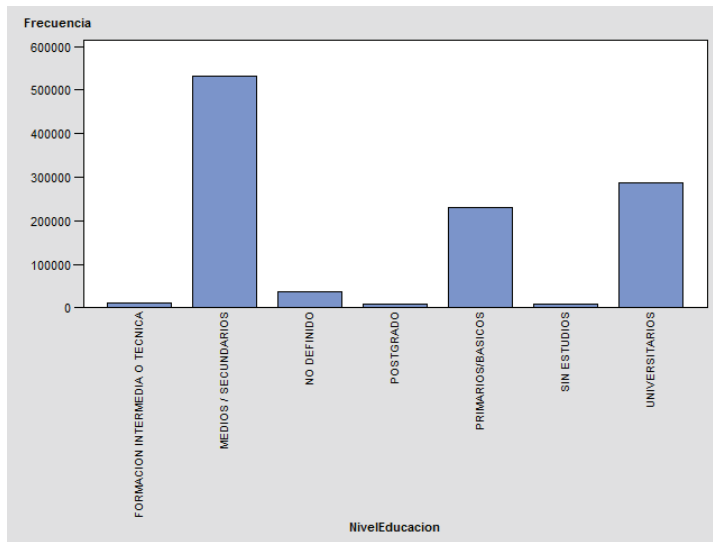
Antigüedad laboral dependiente

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	38.37283	STD DEVIATION	33.32992
MEDIAN	41.00000	VARIANCE	1111
MODE	0.00000	RANGE	167.00000
		INTERQUARTILE RANGE	77.00000

MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	147241	11.86	100.00



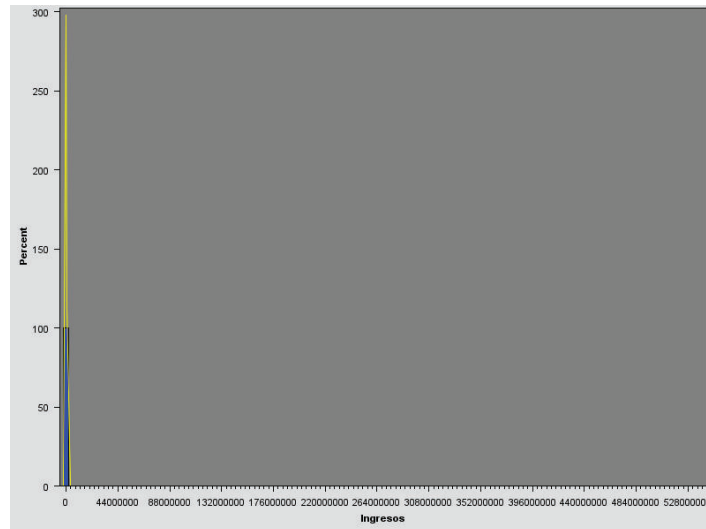
Nivel de educación



Ingresos

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	2728.183	STD DEVIATION	536024
MEDIAN	0.000	VARIANCE	2.87322E11
MODE	0.000	RANGE	540000000
		INTERQUARTILE RANGE	0

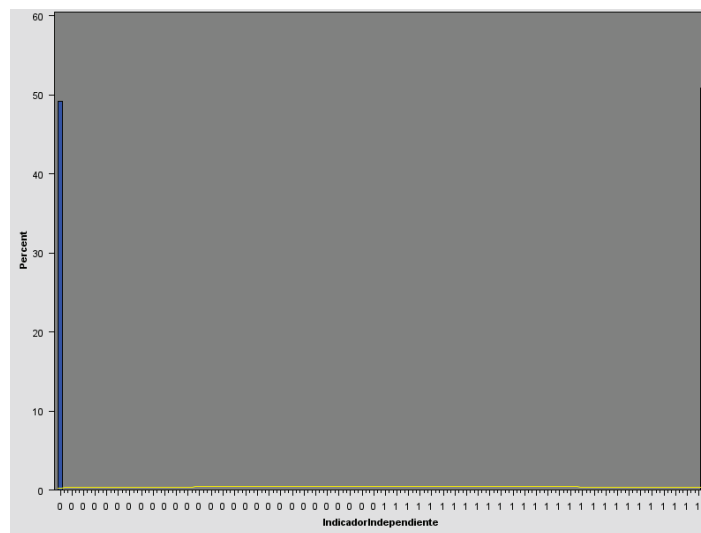
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	98096	7.90	100.00



Indicador si actividad económica es independiente

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	0.508321	STD DEVIATION	0.49993
MEDIAN	1.000000	VARIANCE	0.24993
MODE	1.000000	RANGE	1.00000
		INTERQUARTILE RANGE	1.00000

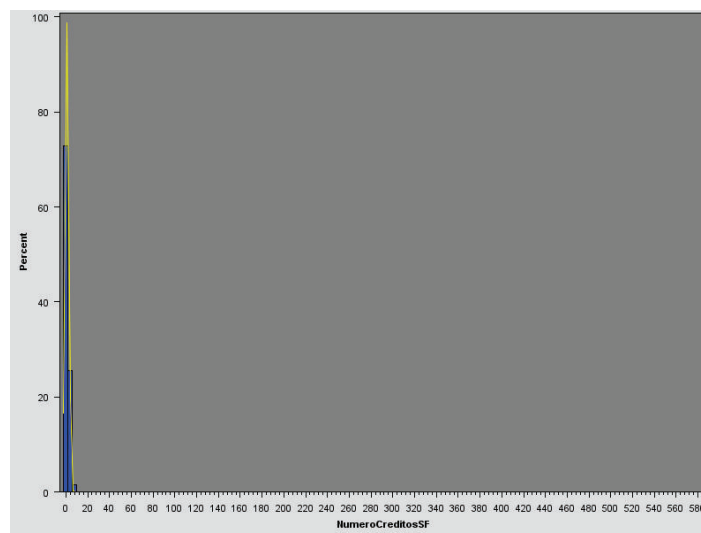
MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	147331	11.87	100.00



Número de créditos en el Sistema Financiero

BASIC STATISTICAL MEASURES			
LOCATION		VARIABILITY	
MEAN	1.052049	STD DEVIATION	1.61743
MEDIAN	1.000000	VARIANCE	2.61607
MODE	0.000000	RANGE	584.00000
		INTERQUARTILE RANGE	2.00000

MISSING VALUES			
MISSING VALUE	COUNT	PERCENT OF	
		ALL OBS	MISSING OBS
.	147310	11.87	100.00



7.2 Anexo 2: Modelo de Regresión Logística

The LOGISTIC Procedure

MODEL INFORMATION	
DATA SET	WORK.TMPMOD
RESPONSE VARIABLE	malo
NUMBER OF RESPONSE LEVELS	2
MODEL	binary logit
OPTIMIZATION TECHNIQUE	Fisher's scoring

NUMBER OF OBSERVATIONS READ	882396
NUMBER OF OBSERVATIONS USED	683076

RESPONSE PROFILE		
ORDERED VALUE	MALO	TOTAL FREQUENCY
1	0	548610
2	1	134466

Probability modeled is malo=1.

Note: 199320 observations were deleted due to missing values for the response or explanatory variables.

Forward Selection Procedure

Step 0. Intercept entered:

MODEL CONVERGENCE STATUS
Convergence criterion (GCONV=1E-8) satisfied.

-2 LOG L	=	677624.49
----------	---	-----------

RESIDUAL CHI-SQUARE TEST		
CHI-SQUARE	DF	PR > CHISQ
108904.550	12	<.0001

Step 1. Effect AtrasoMaxUltimos3Mes entered:

MODEL CONVERGENCE STATUS
Convergence criterion (GCONV=1E-8) satisfied.

MODEL FIT STATISTICS		
CRITERION	INTERCEPT ONLY	INTERCEPT AND COVARIATES
AIC	677626.49	598377.17
SC	677637.93	598400.04
-2 LOG L	677624.49	598373.17

TESTING GLOBAL NULL HYPOTHESIS: BETA=0			
TEST	CHI-SQUARE	DF	PR > CHISQ
LIKELIHOOD RATIO	79251.3190	1	<.0001
SCORE	92210.4722	1	<.0001
WALD	54454.0346	1	<.0001

RESIDUAL CHI-SQUARE TEST		
CHI-SQUARE	DF	PR > CHISQ
16323.8419	11	<.0001

Step 2. Effect NumeroPagosUltimos6M entered:

MODEL CONVERGENCE STATUS
Convergence criterion (GCONV=1E-8) satisfied.

MODEL FIT STATISTICS		
CRITERION	INTERCEPT ONLY	INTERCEPT AND COVARIATES
AIC	677626.49	590609.84
SC	677637.93	590644.15
-2 LOG L	677624.49	590603.84

TESTING GLOBAL NULL HYPOTHESIS: BETA=0			
TEST	CHI-SQUARE	DF	PR > CHISQ
LIKELIHOOD RATIO	87020.6508	2	<.0001
SCORE	100059.524	2	<.0001
WALD	60070.6620	2	<.0001

RESIDUAL CHI-SQUARE TEST		
CHI-SQUARE	DF	PR > CHISQ
8519.4721	10	<.0001

Step 3. Effect AtrasoMaxUltimos6Mes entered:

MODEL CONVERGENCE STATUS
Convergence criterion (GCONV=1E-8) satisfied.

MODEL FIT STATISTICS		
CRITERION	INTERCEPT ONLY	INTERCEPT AND COVARIATES
AIC	677626.49	588162.72
SC	677637.93	588208.45
-2 LOG L	677624.49	588154.72

TESTING GLOBAL NULL HYPOTHESIS: BETA=0			
TEST	CHI-SQUARE	DF	PR > CHISQ
LIKELIHOOD RATIO	89469.7777	3	<.0001
SCORE	102607.144	3	<.0001
WALD	61961.2843	3	<.0001

RESIDUAL CHI-SQUARE TEST		
CHI-SQUARE	DF	PR > CHISQ
5893.2221	9	<.0001

Step 4. Effect NumeroAtrasosMayores entered:

MODEL CONVERGENCE STATUS
Convergence criterion (GCONV=1E-8) satisfied.

MODEL FIT STATISTICS		
CRITERION	INTERCEPT ONLY	INTERCEPT AND COVARIATES
AIC	677626.49	586897.23
SC	677637.93	586954.40
-2 LOG L	677624.49	586887.23

TESTING GLOBAL NULL HYPOTHESIS: BETA=0			
TEST	CHI-SQUARE	DF	PR > CHISQ
LIKELIHOOD RATIO	90737.2660	4	<.0001
SCORE	104944.593	4	<.0001
WALD	67967.9222	4	<.0001

RESIDUAL CHI-SQUARE TEST		
CHI-SQUARE	DF	PR > CHISQ
3432.0528	8	<.0001

Step 5. Effect NumeroCuotasVencidas entered:

MODEL CONVERGENCE STATUS
Convergence criterion (GCONV=1E-8) satisfied.

MODEL FIT STATISTICS		
CRITERION	INTERCEPT ONLY	INTERCEPT AND COVARIATES
AIC	677626.49	585458.30
SC	677637.93	585526.91
-2 LOG L	677624.49	585446.30

TESTING GLOBAL NULL HYPOTHESIS: BETA=0			
TEST	CHI-SQUARE	DF	PR > CHISQ
LIKELIHOOD RATIO	92178.1892	5	<.0001
SCORE	107373.364	5	<.0001
WALD	67362.7709	5	<.0001

RESIDUAL CHI-SQUARE TEST		
CHI-SQUARE	DF	PR > CHISQ
2425.5443	7	<.0001

Step 6. Effect NumeroAtrasosMayores entered:

MODEL CONVERGENCE STATUS
Convergence criterion (GCONV=1E-8) satisfied.

MODEL FIT STATISTICS		
CRITERION	INTERCEPT ONLY	INTERCEPT AND COVARIATES
AIC	677626.49	584490.94
SC	677637.93	584570.98
-2 LOG L	677624.49	584476.94

TESTING GLOBAL NULL HYPOTHESIS: BETA=0			
TEST	CHI-SQUARE	DF	PR > CHISQ
LIKELIHOOD RATIO	93147.5544	6	<.0001
SCORE	107446.042	6	<.0001
WALD	69410.9303	6	<.0001

RESIDUAL CHI-SQUARE TEST		
CHI-SQUARE	DF	PR > CHISQ
1387.1174	6	<.0001

Step 7. Effect AtrasoPromedioUltimo entered:

MODEL CONVERGENCE STATUS
Convergence criterion (GCONV=1E-8) satisfied.

MODEL FIT STATISTICS		
CRITERION	INTERCEPT ONLY	INTERCEPT AND COVARIATES
AIC	677626.49	583590.54
SC	677637.93	583682.02
-2 LOG L	677624.49	583574.54

TESTING GLOBAL NULL HYPOTHESIS: BETA=0			
TEST	CHI-SQUARE	DF	PR > CHISQ
LIKELIHOOD RATIO	94049.9480	7	<.0001
SCORE	108165.132	7	<.0001
WALD	68597.6752	7	<.0001

RESIDUAL CHI-SQUARE TEST		
CHI-SQUARE	DF	PR > CHISQ
498.0761	5	<.0001

Step 8. Effect NumeroCreditosSF entered:

MODEL CONVERGENCE STATUS
Convergence criterion (GCONV=1E-8) satisfied.

MODEL FIT STATISTICS		
CRITERION	INTERCEPT ONLY	INTERCEPT AND COVARIATES
AIC	677626.49	583323.70
SC	677637.93	583426.61
-2 LOG L	677624.49	583305.70

TESTING GLOBAL NULL HYPOTHESIS: BETA=0			
TEST	CHI-SQUARE	DF	PR > CHISQ
LIKELIHOOD RATIO	94318.7971	8	<.0001
SCORE	108409.068	8	<.0001
WALD	68873.5264	8	<.0001

RESIDUAL CHI-SQUARE TEST		
CHI-SQUARE	DF	PR > CHISQ
224.6407	4	<.0001

Step 9. Effect NumeroPagosUltimos3M entered:

MODEL CONVERGENCE STATUS
Convergence criterion (GCONV=1E-8) satisfied.

MODEL FIT STATISTICS		
CRITERION	INTERCEPT ONLY	INTERCEPT AND COVARIATES
AIC	677626.49	583185.62
SC	677637.93	583299.96
-2 LOG L	677624.49	583165.62

TESTING GLOBAL NULL HYPOTHESIS: BETA=0			
TEST	CHI-SQUARE	DF	PR > CHISQ
LIKELIHOOD RATIO	94458.8769	9	<.0001
SCORE	108441.996	9	<.0001
WALD	68321.3497	9	<.0001

RESIDUAL CHI-SQUARE TEST		
CHI-SQUARE	DF	PR > CHISQ
84.4704	3	<.0001

Step 10. Effect AtrasoPromedioUltimo entered:

MODEL CONVERGENCE STATUS
Convergence criterion (GCONV=1E-8) satisfied.

MODEL FIT STATISTICS		
CRITERION	INTERCEPT ONLY	INTERCEPT AND COVARIATES
AIC	677626.49	583106.82
SC	677637.93	583232.60
-2 LOG L	677624.49	583084.82

TESTING GLOBAL NULL HYPOTHESIS: BETA=0			
TEST	CHI-SQUARE	DF	PR > CHISQ
LIKELIHOOD RATIO	94539.6726	10	<.0001
SCORE	108833.378	10	<.0001
WALD	68443.9213	10	<.0001

RESIDUAL CHI-SQUARE TEST		
CHI-SQUARE	DF	PR > CHISQ
3.1696	2	0.2050

Note: No (additional) effects met the 0.05 significance level for entry into the model.

SUMMARY OF FORWARD SELECTION						
STEP	EFFECT ENTERED	DF	NUMBER IN	SCORE CHI-SQUARE	PR > CHISQ	VARIABLE LABEL
1	ATRASOMAXULTIMOS3MES	1	1	92210.47 22	<.000 1	AtrasoMaxUltimos3Meses
2	NUMEROPAGOSULTIMOS6M	1	2	8115.555 8	<.000 1	NumeroPagosUltimos6Meses
3	ATRASOMAXULTIMOS6MES	1	3	2595.943 1	<.000 1	AtrasoMaxUltimos6Meses
4	NUMEROATRASOSMAYORES	1	4	2020.654 3	<.000 1	NumeroAtrasosMayores60DiasUltimos3Meses
5	NUMEROCUOTASVENCIDAS	1	5	1239.614 2	<.000 1	NumeroCuotasVencidas

SUMMARY OF FORWARD SELECTION						
STEP	EFFECT ENTERED	DF	NUMBER IN	SCORE CHI-SQUARE	PR > CHISQ	VARIABLE LABEL
6	NUMEROATRASMAYORES	1	6	1049.0746	<.0001	NumeroAtrasosMayores30DiasUltimos3Meses
7	ATRASOPROMEDIOULTIMO	1	7	895.3438	<.0001	AtrasoPromedioUltimos3Meses
8	NUMEROCREDITOSSF	1	8	273.1414	<.0001	NumeroCreditosSF
9	NUMEROPAGOSULTIMOS3M	1	9	140.1960	<.0001	NumeroPagosUltimos3Meses
10	ATRASOPROMEDIOULTIMO	1	10	81.3789	<.0001	AtrasoPromedioUltimos12Meses

ANALYSIS OF MAXIMUM LIKELIHOOD ESTIMATES					
PARAMETER	DF	ESTIMATE	STANDARD ERROR	WALD CHI-SQUARE	PR > CHISQ
INTERCEPT	1	-1.5008	0.0120	15645.5961	<.0001
ATRASOMAXULTIMOS3MES	1	0.00808	0.000930	75.4686	<.0001
ATRASOPROMEDIOULTIMO	1	0.0466	0.00167	772.9327	<.0001
ATRASOMAXULTIMOS6MES	1	0.0281	0.000746	1415.5590	<.0001
ATRASOPROMEDIOULTIMO	1	-0.0156	0.00173	81.3690	<.0001
NUMEROPAGOSULTIMOS3M	1	0.1029	0.00760	183.3015	<.0001
NUMEROCREDITOSSF	1	0.0352	0.00200	309.4107	<.0001
NUMEROPAGOSULTIMOS6M	1	-0.1913	0.00331	3335.8923	<.0001

ANALYSIS OF MAXIMUM LIKELIHOOD ESTIMATES					
PARAMETER	DF	ESTIMATE	STANDARD ERROR	WALD CHI-SQUARE	PR > CHISQ
NUMEROCUOTASVENCIDAS	1	0.4612	0.0107	1855.9830	<.0001
NUMEROATRASOSMAYORES	1	-0.5770	0.0157	1342.3820	<.0001
NUMEROATRASOSMAYORES	1	-0.5519	0.0279	392.0799	<.0001

ODDS RATIO ESTIMATES			
EFFECT	POINT ESTIMATE	95% WALD CONFIDENCE LIMITS	
ATRASOMAXULTIMOS3MES	1.008	1.006	1.010
ATRASOPROMEDIOULTIMO	1.048	1.044	1.051
ATRASOMAXULTIMOS6MES	1.028	1.027	1.030
ATRASOPROMEDIOULTIMO	0.985	0.981	0.988
NUMEROPAGOSULTIMOS3M	1.108	1.092	1.125
NUMEROCREDITOSSF	1.036	1.032	1.040
NUMEROPAGOSULTIMOS6M	0.826	0.821	0.831
NUMEROCUOTASVENCIDAS	1.586	1.553	1.620
NUMEROATRASOSMAYORES	0.562	0.545	0.579
NUMEROATRASOSMAYORES	0.576	0.545	0.608

ASSOCIATION OF PREDICTED PROBABILITIES AND OBSERVED RESPONSES			
PERCENT CONCORDANT	75.2	SOMERS' D	0.515
PERCENT DISCORDANT	23.7	GAMMA	0.520
PERCENT TIED	1.0	TAU-A	0.163
PAIRS	73769392260	C	0.757

Anexo 3: Código de programación SAS

A. Cálculo Matrices de Transición

```
data a;
input mes $ 1-5 mes_name $ 7-9 periodocorteactual $ 11-16
periodocortepasado $18-23;
datalines;
marzo mar 201303 201203
abril apr 201304 201204
mayo may 201305 201205
junio jun 201306 201206
;run;

%macro meses;
proc sort data= trab_rie.cas_castigos_view out= castigados (keep=
IdentificacionBanco
rename= (IdentificacionBanco = Identificacioncliente));
by IdentificacionBanco;
run;

proc sql noprint;
select count(*) into: N_mes from a
;quit;

%do i = 1 % to &N_mes;
options firstobs=&i obs=&i;
data b;
set a ;
options firstobs=1 obs=max;
proc sql noprint;
```

```

select mes into : mes from b;
select mes_name into : mes_name from b;
select periodocorteactual into : periodocorteactual from b;
select periodocortepasado into : periodocortepasado from b;
;quit;

data &mes_name.12 (keep=identificacioncliente atrasomaxultimos3meses
AtrasopromedioUltimos3meses atrasomaxultimos6meses
atrasopromedioultimos6meses atrasomaxultimos9meses
atrasopromedioultimos9meses atrasomaxultimos12meses
atrasopromedioultimos12meses numeroatrasosmayores30diasultimo
numeroatrasosmayores60diasultimo diasmoramax vencidofinal
numerocuotamora numerocuotasvencidas numeropagosultimos3meses
numeropagosultimos6meses genero edad estadocivil provinciadicilio
antiguedaddependiente
niveleduccion ingresos indicadorindependiente numerocreditosSF
VencidoTotalFinal)
&mes_name.13 (keep=identificacioncliente diasmoramax);
set trab_rie.JE_MOD_COB_CLT_2013_&mes_name ;
if periodocorte = "&periodocortepasado" then output &mes_name.12;
if periodocorte = "&periodocorteactual" then output &mes_name.13;
run;

proc sql;
create table res_&mes_name.12 as
select identificacioncliente, atrasomaxultimos3meses,
AtrasopromedioUltimos3meses, atrasomaxultimos6meses,
atrasopromedioultimos6meses, atrasomaxultimos9meses,
atrasopromedioultimos9meses, atrasomaxultimos12meses,
atrasopromedioultimos12meses, numeroatrasosmayores30diasultimo,
numeroatrasosmayores60diasultimo, max(diasmoramax) as moramax,
vencidofinal,
numerocuotamora, numerocuotasvencidas, numeropagosultimos3meses,

```

```

numeropagosultimos6meses, genero, edad, estadocivil,
provinciadomicilio,
antiguedaddependiente, niveleduccion, ingresos,
indicadorindependiente, numerocreditosSF, sum(VencidoTotalFinal) as
RiesgoTotal
from &mes_name.12
group by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
order by 1
;quit;

```

```

data res_&mes_name.12 (drop=moramax);
set res_&mes_name.12;
if moramax = 1*&A1 then rangomoraini = "A-1";
if 1*&A2_min le moramax le 1*&A2_max then rangomoraini = "A-2";
if 1*&A3_min le moramax le 1*&A3_min then rangomoraini = "A-3";
if 1*&B1_min le moramax le 1*&B1_min then rangomoraini = "B-1";
if 1*&B2_min le moramax le 1*&B2_min then rangomoraini = "B-2";
if 1*&C1_min le moramax le 1*&C1_min then rangomoraini = "C-1";
if 1*&C2_min le moramax le 1*&C2_min then rangomoraini = "C-2";
if 1*&D_min le moramax le 1*&D_min then rangomoraini = "D";
if moramax gt 1*&E then rangomoraini = "E";
run;

```

```

proc sql;
create table res_&mes_name.13 as
select identificacioncliente, (diasmoramax) as moramax
from &mes_name.13
group by 1
order by 1
;quit;

```

```

data res_&mes_name.13 (drop=moramax);
set res_&mes_name.13;

```

```

if moramax = 1*&A1 then rangomorafin = "A-1";
if 1*&A2_min le moramax le 1*&A2_max then rangomorafin = "A-2";
if 1*&A3_min le moramax le 1*&A3_min then rangomorafin = "A-3";
if 1*&B1_min le moramax le 1*&B1_min then rangomorafin = "B-1";
if 1*&B2_min le moramax le 1*&B2_min then rangomorafin = "B-2";
if 1*&C1_min le moramax le 1*&C1_min then rangomorafin = "C-1";
if 1*&C2_min le moramax le 1*&C2_min then rangomorafin = "C-2";
if 1*&D_min le moramax le 1*&D_min then rangomorafin = "D";
if moramax gt 1*&E then rangomorafin = "E";
run;

data union_&mes_name;
merge res_&mes_name.12 (in=a)
      res_&mes_name.13(in=b );
by identificacioncliente;
if a;
run;

data union_&mes_name;
merge union_&mes_name (in=a)
      castigados (in=b );
by identificacioncliente;
if b then rangomorafin = "E";
if a;
run;

proc sql;
create table para_matriz_&mes_name as
select rangomoraini, rangomorafin, sum(RiesgoTotal) /*count(*)*/ as
cuanto
from union_&mes_name
group by 1,2
order by 1,2

```

```

;quit;

proc sql;
create table pre_matriz_&mes_name as
select rangomoraini, rangomorafin, cuanto / sum(cuanto) format
percent7.2 as probabilidad
from para_matriz_&mes_name
group by 1
order by 1,2
;quit;

%end;

data promedios;
merge pre_matriz_mar (rename= (probabilidad = p1))
pre_matriz_apr (rename= (probabilidad = p2))
pre_matriz_may (rename= (probabilidad = p3))
pre_matriz_jun(rename= (probabilidad = p4));
by rangomoraini rangomorafin;
p = max(p1,p2,p3,p4);
format p percent7.2;
run;

proc transpose data= promedios out= matriz (drop=_name_);
var p;
by rangomoraini;
id rangomorafin;
run;

%do i = 1 % to &N_mes;
options firstobs=&i obs=&i;
data b;
set a ;

```

```

options firstobs=1 obs=max;
proc sql noprint;
select mes into : mes&i from b;
select mes_name into : mes_name&i from b;
;quit;
%end;

proc sql;
%do i = 1 % to &N_mes;
drop table &&mes_name&i..12;
drop table &&mes_name&i..13;
drop table union_&&mes_name&i;
drop table para_matriz_&&mes_name&i;
drop table pre_matriz_&&mes_name&i;
drop table res_&&mes_name&i..13;
drop table res_&&mes_name&i..12;
%end;

drop table a;
drop table b;
drop table promedios;
quit;

%mend;

%meses

data matriz;
set matriz;
suma = sum("A-1"n,"A-2"n, "A-3"n, "B-1"n, "B-2"n, "C-1"n, "C-2"n, D,
E);
run;

data matriz (Drop=suma);

```

```

set matriz;
"A-1"n = "A-1"n /suma;
"A-2"n = "A-2"n /suma;
"A-3"n = "A-3"n/suma;
"B-1"n = "B-1"n/suma;
"B-2"n = "B-2"n/suma;
"C-1"n = "C-1"n/suma;
"C-2"n = "C-2"n/suma;
D= D/suma;
E = E/suma;
run;

```

B. CALIFICACION POR CLIENTE

```

data a;
input mes $ 1-5 mes_name $ 7-9 periodocorteactual $ 11-16
periodocortepasado $18-23;
datalines;
marzo mar 201303 201203
abril apr 201304 201204
mayo may 201305 201205
junio jun 201306 201206
;run;

```

```

%macro meses;

```

```

data para_calificar (keep=identificacioncliente atrasomaxultimos3meses
AtrasopromedioUltimos3meses atrasomaxultimos6meses
atrasopromedioultimos6meses atrasomaxultimos9meses
atrasopromedioultimos9meses atrasomaxultimos12meses
atrasopromedioultimos12meses numeroatrasosmayores30diasultimo
numeroatrasosmayores60diasultimo diasmoramax vencidofinal

```



```

numerocuotamora numerocuotasvencidas numeropagosultimos3meses
numeropagosultimos6meses genero edad estadocivil provinciadicilio
antiguedaddependiente
niveleducacion ingresos indicadorindependiente numerocreditosSF);
set trab_rie.JE_MOD_COB_CLT_2014_OCT ;
identificacioncliente = identificacion;
run;

```

```

data para_calificar (drop=diasmoramax);
set para_calificar;
if diasmoramax = 1*&A1 then rangomoraini = "A-1";
if 1*&A2_min le diasmoramax le 1*&A2_max then rangomoraini = "A-2";
if 1*&A3_min le diasmoramax le 1*&A3_max then rangomoraini = "A-3";
if 1*&B1_min le diasmoramax le 1*&B1_max then rangomoraini = "B-1";
if 1*&B2_min le diasmoramax le 1*&B2_max then rangomoraini = "B-2";
if 1*&C1_min le diasmoramax le 1*&C1_max then rangomoraini = "C-1";
if 1*&C2_min le diasmoramax le 1*&C2_max then rangomoraini = "C-2";
if 1*&D_min le diasmoramax le 1*&D_max then rangomoraini = "D";
if moramax gt 1*&E then rangomoraini = "E";
run;

```

```

%do i = 1 % to 4;
options firstobs=&i obs=&i;
data b;
set a ;
options firstobs=1 obs=max;
proc sql noprint;
select mes into : mes from b;
select mes_name into : mes_name from b;
select periodocorteactual into : periodocorteactual from b;
select periodocortepasado into : periodocortepasado from b;
;quit;

```

```

data &mes_name.12 (keep=identificacioncliente atrasomaxultimos3meses
AtrasopromedioUltimos3meses atrasomaxultimos6meses
atrasopromedioultimos6meses atrasomaxultimos9meses
atrasopromedioultimos9meses atrasomaxultimos12meses
atrasopromedioultimos12meses numeroatrasosmayores30diasultimo
numeroatrasosmayores60diasultimo diasmoramax vencidofinal
numerocuotamora numerocuotasvencidas numeropagosultimos3meses
numeropagosultimos6meses genero edad estadocivil provinciadicilio
antiguedaddependiente
niveleduccion ingresos indicadorindependiente numerocreditosSF)
&mes_name.13 (keep=identificacioncliente diasmoramax);
set trab_rie.JE_MOD_COB_CLT_2013_&mes_name ;
if periodocorte = "&periodocortepasado" then output &mes_name.12;
if periodocorte = "&periodocorteactual" then output &mes_name.13;
run;

```

```

proc sql;
create table res_&mes_name.12 as
select identificacioncliente, atrasomaxultimos3meses,
AtrasopromedioUltimos3meses, atrasomaxultimos6meses,
atrasopromedioultimos6meses, atrasomaxultimos9meses,
atrasopromedioultimos9meses, atrasomaxultimos12meses,
atrasopromedioultimos12meses, numeroatrasosmayores30diasultimo,
numeroatrasosmayores60diasultimo, max(diasmoramax) as moramax,
vencidofinal,
numerocuotamora, numerocuotasvencidas, numeropagosultimos3meses,
numeropagosultimos6meses, genero, edad, estadocivil,
provinciadomicilio,
antiguedaddependiente, niveleduccion, ingresos,
indicadorindependiente, numerocreditosSF
from &mes_name.12
group by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
order by 1

```

```

;quit;

data res_&mes_name.12 (drop=moramax);
set res_&mes_name.12;
if moramax = 1*&A1 then rangomoraini = "A-1";
if 1*&A2_min le moramax le 1*&A2_max then rangomoraini = "A-2";
if 1*&A3_min le moramax le 1*&A3_min then rangomoraini = "A-3";
if 1*&B1_min le moramax le 1*&B1_min then rangomoraini = "B-1";
if 1*&B2_min le moramax le 1*&B2_min then rangomoraini = "B-2";
if 1*&C1_min le moramax le 1*&C1_min then rangomoraini = "C-1";
if 1*&C2_min le moramax le 1*&C2_min then rangomoraini = "C-2";
if 1*&D_min le moramax le 1*&D_min then rangomoraini = "D";
if moramax gt 1*&E then rangomoraini = "E";
run;

proc sql;
create table res_&mes_name.13 as
select identificacioncliente, (diasmoramax) as moramax
from &mes_name.13
group by 1
order by 1
;quit;

data res_&mes_name.13 (drop=moramax);
set res_&mes_name.13;
if moramax = 1*&A1 then rangomorafin = "A-1";
if 1*&A2_min le moramax le 1*&A2_max then rangomorafin = "A-2";
if 1*&A3_min le moramax le 1*&A3_min then rangomorafin = "A-3";
if 1*&B1_min le moramax le 1*&B1_min then rangomorafin = "B-1";
if 1*&B2_min le moramax le 1*&B2_min then rangomorafin = "B-2";
if 1*&C1_min le moramax le 1*&C1_min then rangomorafin = "C-1";
if 1*&C2_min le moramax le 1*&C2_min then rangomorafin = "C-2";
if 1*&D_min le moramax le 1*&D_min then rangomorafin = "D";

```

```

if moramax gt 1*&E then rangomorafin = "E";
run;

data union_&mes_name;
merge res_&mes_name.12 (in=a)
      res_&mes_name.13(in=b );
by identificacioncliente;
if a;
if rangomorafin = "" then rangomorafin = "E";
run;

proc sql noprint;
select name into : var1 - : var24 from
dictionary.columns
where libname = "WORK" and memname = "UNION_MAR"
and name not in ("IdentificacionCliente" "rangomoraini" "rangomorafin")
;quit;

%do j = 1 %to 24

proc sql;
create table para_prob_v&j._&mes_name as
select &&var&j, count(*) as cuantos
from union_&mes_name
group by 1
;Quit;

proc sql;
create table malos_prob_v&j._&mes_name as
select &&var&j, count(*) as malos
from union_&mes_name
where rangomorafin = "E"

```

```

group by 1
;Quit;

data prob_v&j.&mes_name;
merge para_prob_v&j._&mes_name malos_prob_v&j._&mes_name;
by &&var&j;
p_v&j = malos / cuantos;
format p_v&j percent7.2;
run;

%end;

%end;

%do k = 1 %to 24

data promedios_v&k (keep= &&var&k p_v&k);
merge prob_v&k.mar (rename= (p_v&k = p1))
prob_v&k.apr (rename= (p_v&k = p2))
prob_v&k.may (rename= (p_v&k = p3))
prob_v&k.jun (rename= (p_v&k = p4));
by &&var&k;
p_v&k = mean(p1,p2,p3,p4);
format p_v&k percent7.2;
run;

proc sort data=para_calificar;
by &&var&k;
run;

data para_calificar;
merge para_calificar promedios_v&k;
by &&var&k;

```

```

run;

%end;

%mend;

%meses

data calificacion (keep = identificacioncliente rangomoraini p pmin
pmax);
set para_calificar (keep= IdentificacionCliente rangomoraini p_v1 -
p_v24) ;
if identificacioncliente ne "";
p =
(sum(p_v1,p_v2,p_v3,p_v4,p_v5,p_v6,p_v7,p_v8,p_v9,p_v10,p_v11,p_v12,
p_v13,p_v14,p_v15,p_v16,p_v17,p_v18,p_v19,p_v20,p_v21,p_v22,p_v23,p_v24
))/23;
pmin =
min(p_v1,p_v2,p_v3,p_v4,p_v5,p_v6,p_v7,p_v8,p_v9,p_v10,p_v11,p_v12,
p_v13,p_v14,p_v15,p_v16,p_v17,p_v18,p_v19,p_v20,p_v21,p_v22,p_v23,p_v24
);
pmax =
max(p_v1,p_v2,p_v3,p_v4,p_v5,p_v6,p_v7,p_v8,p_v9,p_v10,p_v11,p_v12,
p_v13,p_v14,p_v15,p_v16,p_v17,p_v18,p_v19,p_v20,p_v21,p_v22,p_v23,p_v24
);
pmean =
mean(p_v1,p_v2,p_v3,p_v4,p_v5,p_v6,p_v7,p_v8,p_v9,p_v10,p_v11,p_v12,
p_v13,p_v14,p_v15,p_v16,p_v17,p_v18,p_v19,p_v20,p_v21,p_v22,p_v23,p_v24
);
p = p * pmin * pmax * pmean;
format p percent7.2;

```

```
run;
```

```
proc sort data=calificacion;
```

```
by identificacioncliente;
```

```
run;
```

```
DATA WORK.TMPMod;
```

```
    SET UNION_JUN;
```

```
malo = 0;
```

```
if rangomorafin = "E" then malo = 1;
```

```
RUN;
```

```
PROC LOGISTIC DATA=WORK.TMPMod PLOTS (ONLY)=ALL;
```

```
    MODEL malo (Event = '1')=AtrasoMaxUltimos3Meses  
AtrasoPromedioUltimos3Meses AtrasoMaxUltimos6Meses  
AtrasoMaxUltimos9Meses AtrasoMaxUltimos12Meses  
AtrasoPromedioUltimos12Meses NumeroPagosUltimos3Meses  
IndicadorIndependiente NumeroCreditosSF Ingresos  
NumeroPagosUltimos6Meses NumeroCuotaMora NumeroCuotasVencidas  
NumeroAtrasosMayores30DiasUltimo NumeroAtrasosMayores60DiasUltimo  
AtrasoPromedioUltimos9Meses AtrasoPromedioUltimos6Meses/
```

```
    SELECTION=FORWARD SLE=0.05 INCLUDE=0 LINK=LOGIT;
```

```
OUTPUT OUT=WORK.PRED PREDPROBS=INDIVIDUAL;
```

```
RUN; QUIT;
```

```
proc sql;
```

```
select max(ip_1) into : max from PRED;
```

```
select min(ip_1) into : min from PRED
```

```
;quit;
```

```
proc sql;
```

```
select max(p) into : max_mio from calificacion;
```

```
select min(p) into : min_mio from calificacion
```

```

;quit;

data calificacion;
set calificacion;
p = ((p + &min - &min_mio)/((1*&max_mio-1*&min_mio)/(1*&max-1*&min)));
run;

data calificacion (drop=pmin pmax);
merge calificacion (in=a)
       castigados (in=b );
by identificacioncliente;
if b then p = 1;
if a;
if p gt 1 then p = 1;
run;

%macro limpieza;
%do i = 1 % to 4;
options firstobs=&i obs=&i;
data b;
set a ;
options firstobs=1 obs=max;
proc sql noprint;
select mes into : mes&i from b;
select mes_name into : mes_name&i from b;
;quit;
%end;

proc sql;
%do i = 1 % to 4;
drop table &&mes_name&i..12;
drop table &&mes_name&i..13;
drop table union_&&mes_name&i;

```



```

drop table res_&&mes_name&i..13;
drop table res_&&mes_name&i..12;
%do j = 1 %to 24;
drop table malos_prob_v&j._&&mes_name&i;
drop table prob_v&j.&&mes_name&i;
drop table para_prob_v&j._&&mes_name&i;

%end;

%end;

proc sql;
%do k = 1 % to 24;
drop table promedios_v&k;
%end;
drop table a;
drop table b;
drop table tmpmod;
quit;
%mend;
%limpieza

```

C. GENERACION DE RANGOS

```

proc sql;
create table q1 as
select rangomoraini, mean(p) as minimo
from calificacion
group by 1
;quit;

proc sql;
select rangomoraini, mean(p) as minimo

```

```
from calificacion
group by 1
;quit;
```

```
proc univariate data=calificacion;
output out= rangos PCTLPTS= 11 22 33 44 55 66 77 88 99 PCTLPRE= P;
var p;
run;
```

```
proc sql;
select p11 into : p11 from rangos;
select p22 into : p22 from rangos;
select p33 into : p33 from rangos;
select p44 into : p44 from rangos;
select p55 into : p55 from rangos;
select p66 into : p66 from rangos;
select p77 into : p77 from rangos;
select p88 into : p88 from rangos;
quit;
```

```
proc sql;
select minimo into : p11 from q1 where rangomoraini = "A-1" ;
select minimo into : p22 from q1 where rangomoraini = "A-2";
select minimo into : p33 from q1 where rangomoraini = "A-3";
select minimo into : p44 from q1 where rangomoraini = "B-1";
select minimo into : p55 from q1 where rangomoraini = "B-2";
select minimo into : p66 from q1 where rangomoraini = "C-1";
select minimo into : p77 from q1 where rangomoraini = "C-2";
select minimo into : p88 from q1 where rangomoraini = "D";
quit;
```

```
data rangos_calificacion_jul;
```

```

set calificacion;
if p le &p11 then rango= "A-1";
if &p11 lt p le &p22 then rango= "A-2";
if &p22 lt p le &p33 then rango= "A-3";
if &p33 lt p le &p44 then rango= "B-1";
if &p44 lt p le &p55 then rango= "B-2";
if &p55 lt p le &p66 then rango= "C-1";
if &p66 lt p le &p77 then rango= "C-2";
if &p77 lt p le &p88 then rango= "D";
if p gt &p88 then rango= "E";
run;

proc sql;
create table q2 as
select rango, count(Identificacioncliente) as cuantos
from rangos_calificacion_jul
group by 1
;quit;

data trab_Rie.JE_Modelo_PD_Personas_201410; /**/
set rangos_calificacion_jul;
run;

```

REFERENCIAS BIBLIOGRÁFICAS

- PEREZ, C. Estadística Aplicada a través de Excel. Editorial Prentice Hall, 2002.
- LIND, Marshall,. Estadística Aplicada a los Negocios y la Economía. Editorial Mc. Graw Hill, 2008.
- MONTGOMERY, D, ; RUNGER, G. Probabilidades y Estadística aplicadas a la Ingeniería Editorial Mc. Graw Hill, 1998.
- SOTO, Federico. Hacia una efectiva gestión de Cobranzas. Negril Ediciones 2006.
- LAWRENCE, Davis; SOLOMON Arlene. Manging a Consumer Lending. Solomon Lawrence Partners 2002.
- SCHEAFFER, R; MENDENHALL, W; LYMAN, O. Elementos de Muestreo. Grupo Editorial Iberoamérica 1986.
- RINCÓN, Luis; Introducción a los Procesos Estocásticos. Departamento de Matemáticas UNAM 2011.
- II Congreso Andino de Call Center, Contact Centers y CRM, Conferencia e Lilian Simbaqueba, Abril 2005.
- I Congreso de crédito y Cobranza. Quito 2009