

INTERNAL CREDIT SCORING

Project & Design

by David Vargas

Objetivos

Objetivo General

Diseño, modelización e implementación de herramienta analítica de riesgo, calificación crediticia y medición de default, como eje derivable para el desarrollo y mejoramiento de herramientas analíticas y de simulación, optimizando el otorgamiento de crédito, a partir de la explotación de Big Data y Data Mining.

Objetivos específicos:

- Dinamizar la analítica de datos aplicada al riesgo de default.
- Determinar los principales factores del riesgo crediticio y financiero.
- Optimización de oferta y demanda de crédito a través de la integración de analítica y modelización dinámica.
- Estandarización de Big Data y Data Mining intraempresarial de clientes y del mercado financiero.
- Perfeccionamiento en técnicas de visualización de datos automatizados a través de generación de Dashboards y proyecciones estadísticas.

Metodologia

Implementacion base de datos de corte transversal y series de tiempo de asociados deudores para el periodo enero - octubre del 2022. Siguiendo con el proceso ETL. Se ha realizado la limpieza y modificacion de la base de datos original para efectos del procesamiento en el software RStudio, teniendo en cuenta la relevancia de las variables para el presente analisis, por lo que la informacion a implementar, comparada con la base inicial, se ha estructurado en 36.192 datos/observaciones de periodicidad mensual acumulativa, constituida por 53 variables sociodemograficas y financieras.

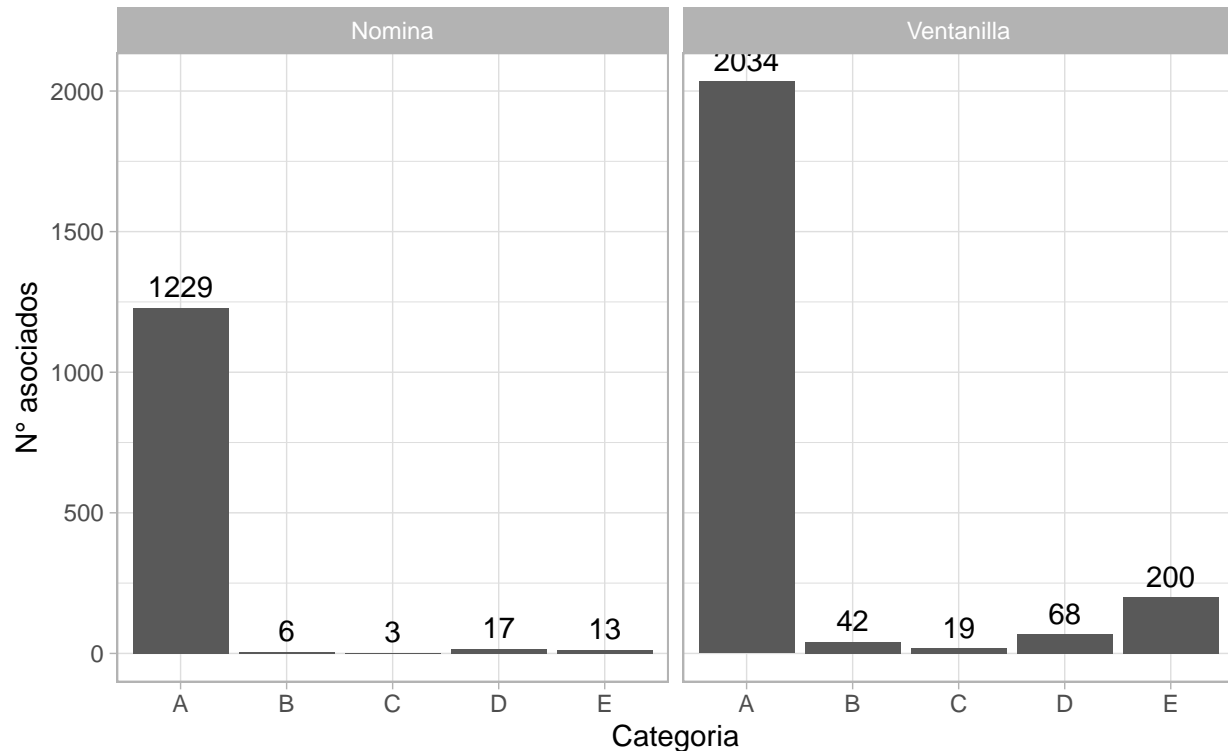
A partir del tipo de calificacion de riesgo para los tipos de obligaciones crediticias, en funcion de la altura demora (en dias), se presenta:

CALIFICACIONES

##	CATEGORIA	COMERCIAL	CONSUMO	VIVIENDA	MICROCREDITO
## 1	A	0 - 30	0 - 30	0 - 60	0 - 30
## 2	B	31 - 90	31 - 60	61 - 150	31 - 60
## 3	C	91 - 180	61 - 90	151 - 360	61 - 90
## 4	D	181 - 360	91 - 180	361 - 540	91 - 120
## 5	E	> 360	> 180	> 540	> 120

Categoría de riesgo

A corte de octubre 2022



Calificación antes de Arrastre (Cierre)

##	A	B	C	D	E
## 3303	38	22	72	196	

En cuanto a la metodología estadística, se opta por un modelo de regresión logística simple, donde se toma la calificación establecida de la línea de consumo como la variable dependiente. Esta variable, que inicialmente es de tipo cualitativa (A, B, C, D, E) y corresponderá a la explicativa del modelo, se transforma en variable dicotómica, tomando dos valores probabilísticos según el tipo de comportamiento de pago (0 y 1), como se especifica:

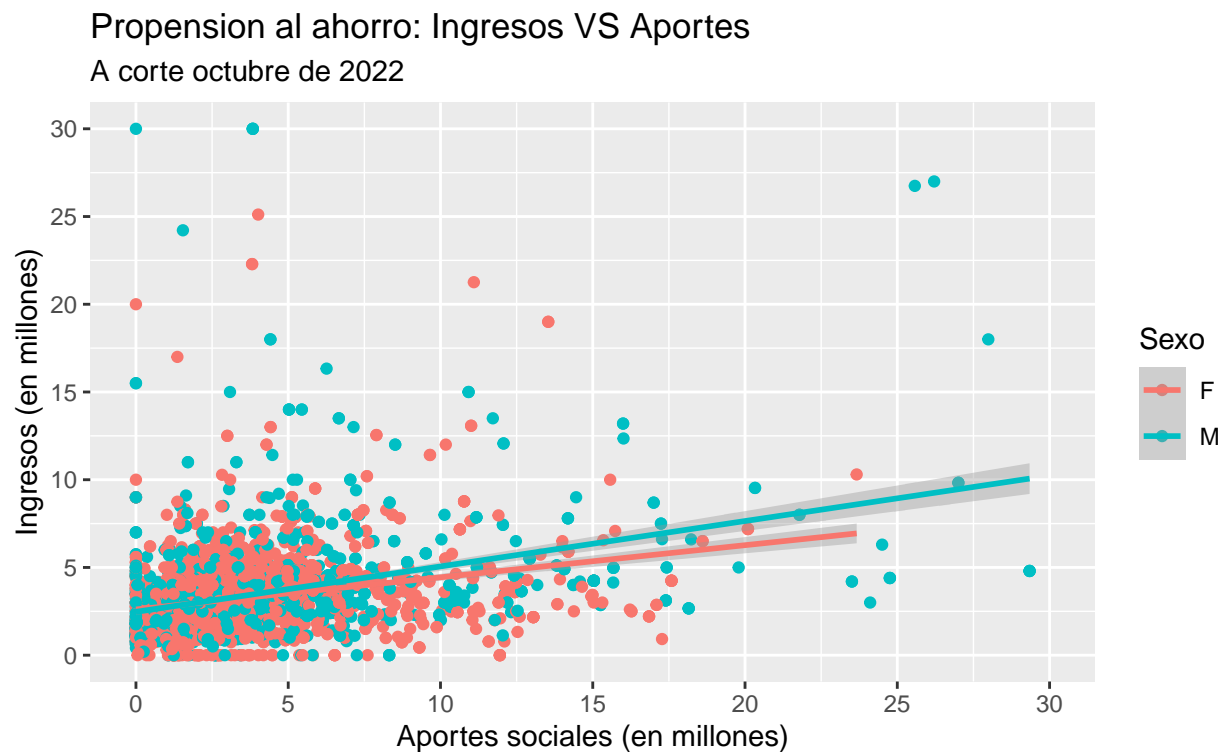
CRITERIOS DE TRANSFORMACION

##	COMPORTAMIENTO	CATEGORIA	VALOR
## 1	Bueno	A	1
## 2	Regular - Malo	B-C-D-E	0

En este caso, el valor “1” representa la probabilidad de buen comportamiento en el habito de pago, donde la obligacion no excede la altura de mora de 30 dias; y el valor “0” representa un regular o mal comportamiento en el pago de la obligacion, que es desde 31 dias en adelante.

Evidencia empirica

Algunos datos sociodemograficos



Distribucion del ingreso por edad y genero

A corte octubre de 2022

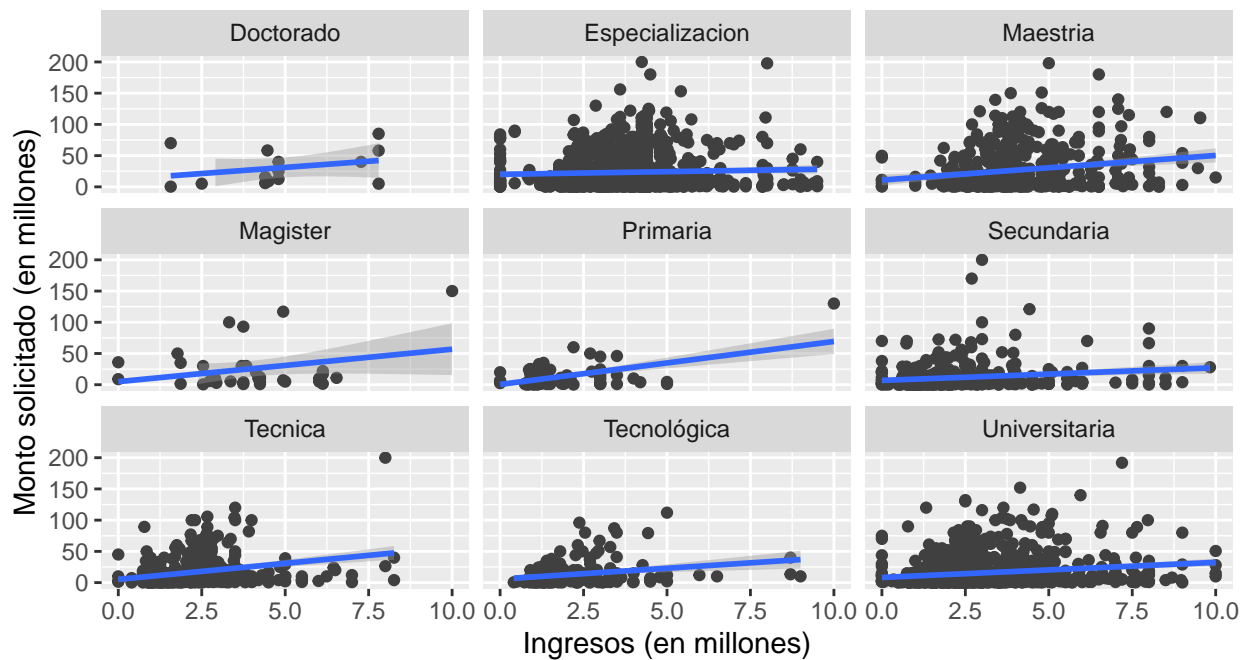


Informacion financiera Y crediticia

Propension al monto de solicitud de credito

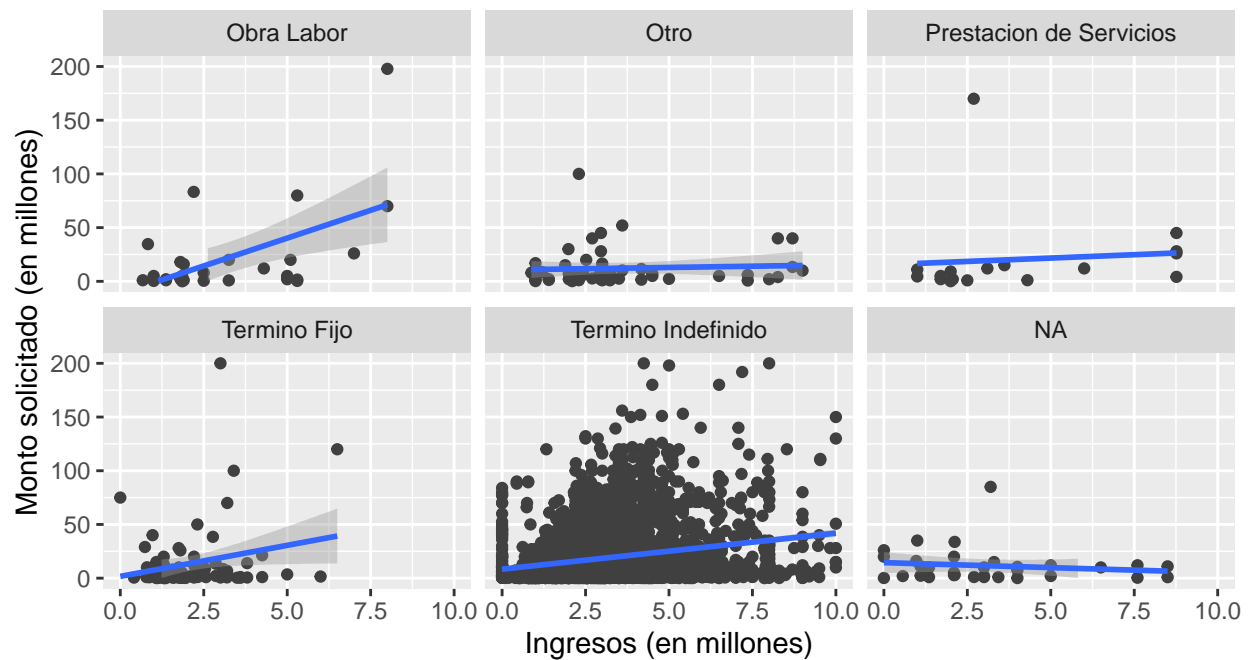
Propension del monto solicitado: por nivel de estudios

A corte octubre de 2022



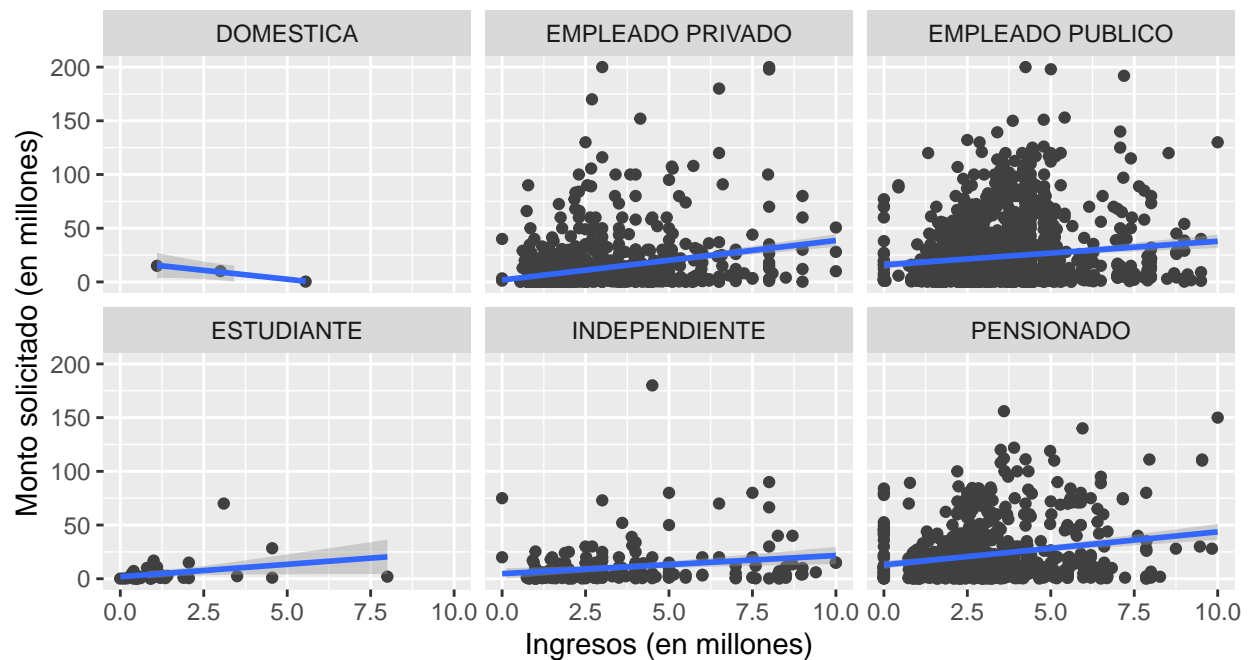
Propension del monto solicitado: por tipo de contrato laboral

A corte octubre de 2022



Propension del monto solicitado: por ocupacion

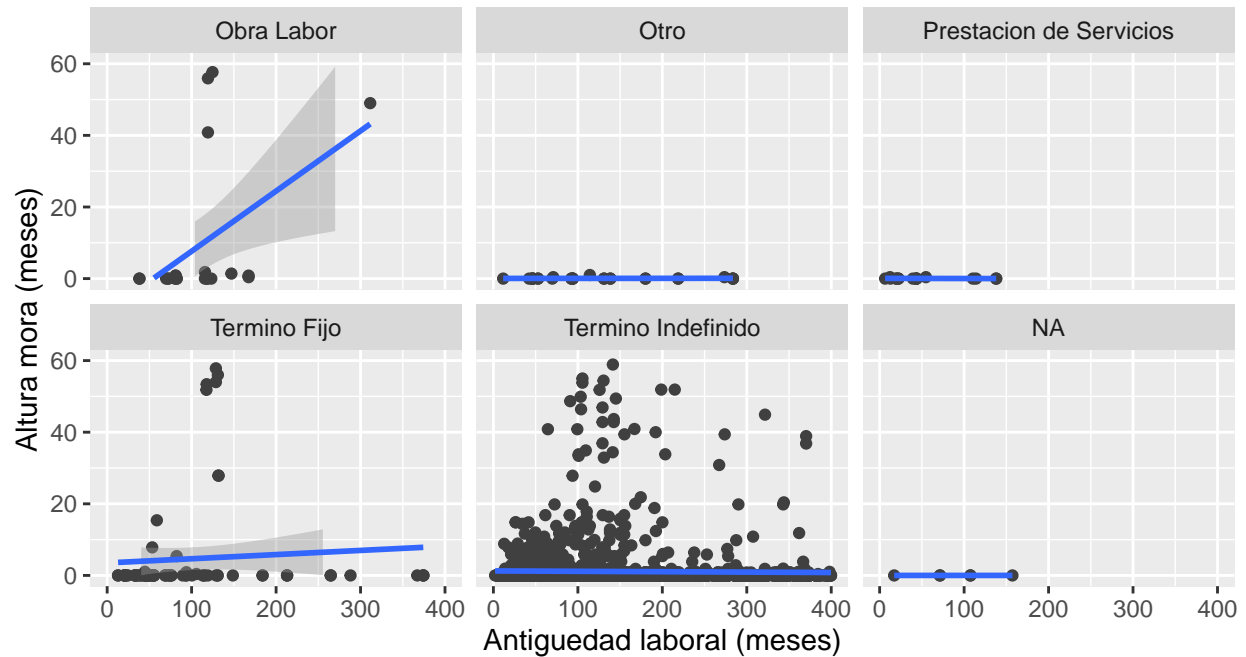
A corte octubre de 2022



Incidencia al impago

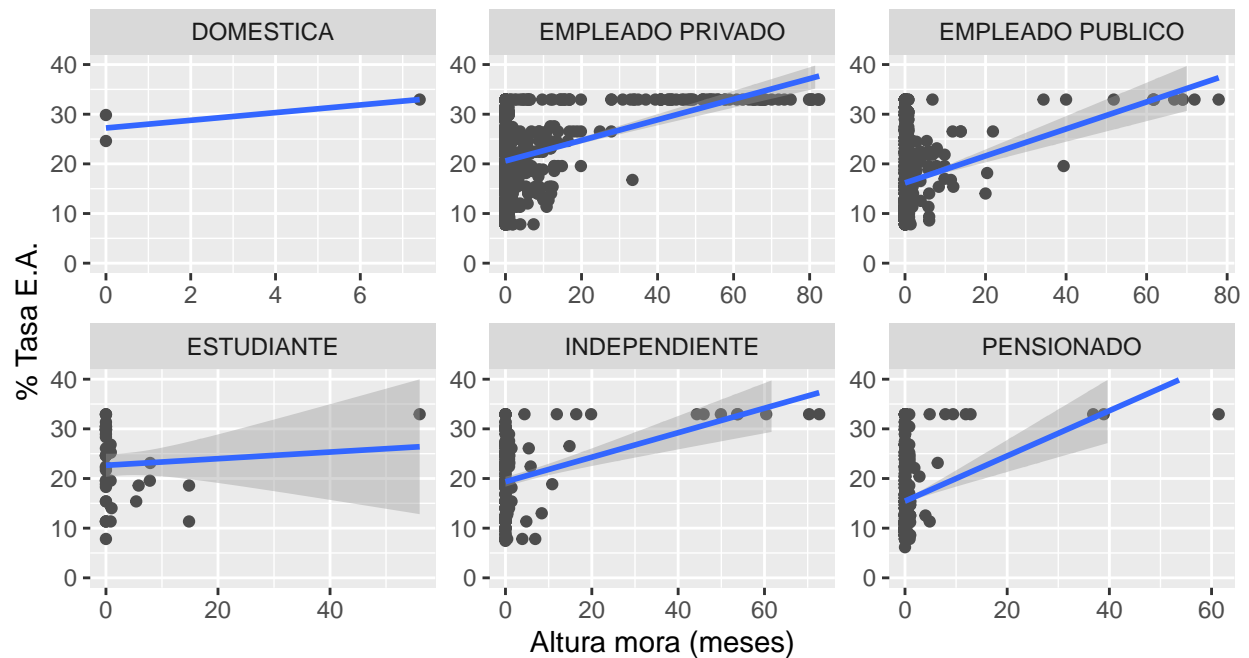
Incidencia de la antigüedad laboral sobre el impago : por tipo de contrato

A corte octubre de 2022



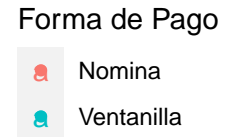
Sensibilidad a la tasa de interes: por ocupacion

A corte octubre de 2022

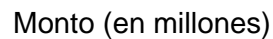


Elasticidad precio de la demanda del credito

Enero a octubre del 2022

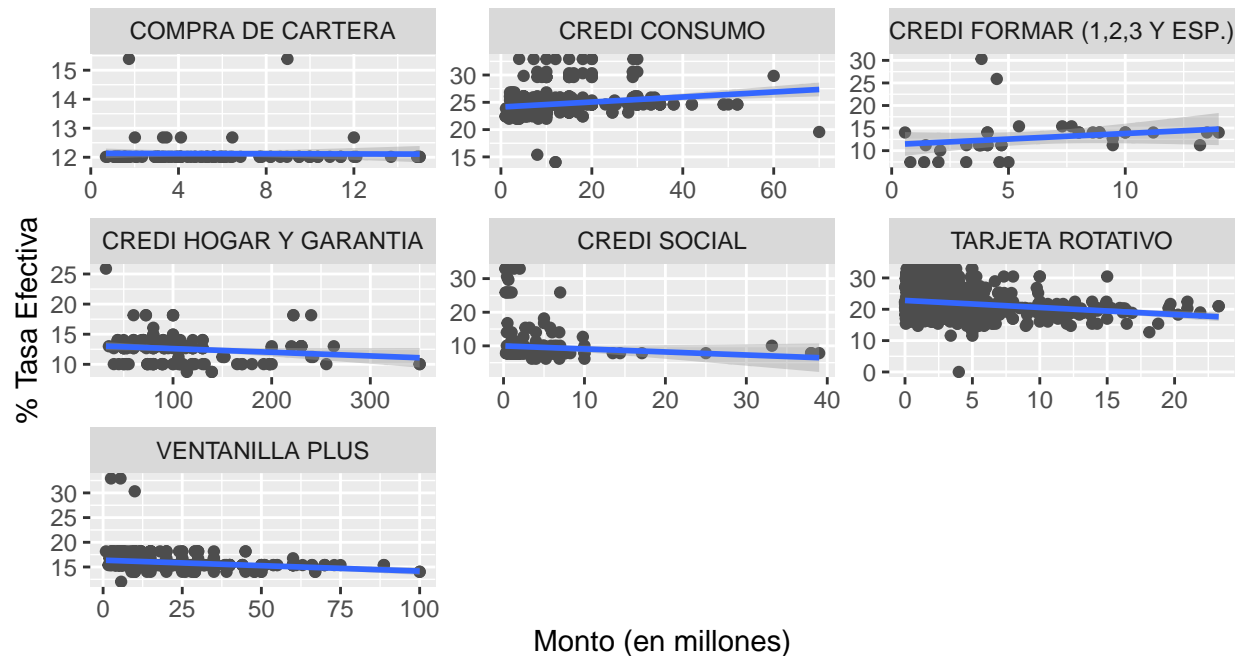


Enero a octubre del 2022



Elasticidad precio de la demanda: modalidades de ventanilla

Enero a octubre del 2022



Analitica de datos:

Transformacion inicial

Para este caso, se implementa un modelo Logit o de regresión logística, en el que la variable explicada corresponde al pago oportuno o inoportuno (1 o 0) que representa al Default, teniendo en cuenta la máxima altura de mora de la línea de consumo para una calificación de la categoría A.

Se realiza las transformaciones pertinentes para fines analíticos de los datos, convirtiendo en dicótomas a las variables explicativas originalmente categóricas: - Garantía: 1 = Con garantía real | 0 = Sin garantía real - Codeudor: 1 = Con codeudor | 0 = Sin codeudor - Sexo: 1 = Mujer | 0 = Hombre - Non Default 1 = Categoría A | 0 = demás categorías

Modelo estadístico

Estimacion del mejor modelo

El modelo estadístico inicial incluirá todas las variables que, según algunos supuestos financieros y teóricos iniciales, pueden explicar el Default de riesgo crediticio.

```
##
## Modelo 1: incluye todas las variables
## =====
##                               Dependent variable:
##                               -----
```


##	'Non Default'
## -----	
## Monto	-0.09**
##	(0.05)
##	
## Codeudor	-1.50
##	(1.07)
##	
## Garantia	10.91***
##	(3.74)
##	
## Edad	-0.01
##	(0.04)
##	
## Ingresos	0.43
##	(0.28)
##	
## Antigüedad laboral	-0.0001
##	(0.0001)
##	
## Antigüedad asociado	-0.001
##	(0.001)
##	
## Sexo	1.99*
##	(1.14)
##	
## Plazo	0.01
##	(0.03)
##	
## Altura mora	-0.36***
##	(0.08)
##	
## OcupacionEMPLEADO PRIVADO	-7.21
##	(29,232.44)
##	
## OcupacionEMPLEADO PUBLICO	-5.85
##	(29,232.44)
##	
## OcupacionESTUDIANTE	20.53
##	(32,384.30)
##	
## OcupacionINDEPENDIENTE	6.19
##	(29,242.67)
##	
## OcupacionPENSIONADO	-7.49
##	(29,232.44)
##	
## 'Saldo Capital'	0.11

##	(0.10)
##	
## 'Nivel Estudios'Especializacion	-14.38
##	(7,975.55)
##	
## 'Nivel Estudios'Maestria	-11.00
##	(7,975.55)
##	
## 'Nivel Estudios'Magister	-21.33
##	(7,975.55)
##	
## 'Nivel Estudios'Primaria	3.13
##	(10,529.68)
##	
## 'Nivel Estudios'Secundaria	-11.29
##	(7,975.55)
##	
## 'Nivel Estudios'Tecnica	-8.77
##	(7,975.55)
##	
## 'Nivel Estudios'Tecnológica	-12.11
##	(7,975.55)
##	
## 'Nivel Estudios'Universitaria	-12.35
##	(7,975.55)
##	
## 'Estado civil'Divorciado	-3.61*
##	(1.93)
##	
## 'Estado civil'Otro	5.19
##	(9,840.86)
##	
## 'Estado civil'Separado	10.34
##	(2,755.04)
##	
## 'Estado civil'Soltero	-0.91
##	(1.14)
##	
## 'Estado civil'Union libre	-2.08*
##	(1.25)
##	
## 'Estado civil'Viudo	-3.10
##	(4.78)
##	
## 'Tipo de contrato'Otro	19.08
##	(3,676.20)
##	
## 'Tipo de contrato'Prestacion de Servicios	9.66

```

## (5,581.06)
##
## 'Tipo de contrato'Termino Fijo -3.40
## (2.87)
##
## 'Tipo de contrato'Termino Indefinido 1.51
## (2.39)
##
## 'Tasa Efectiva Colocacion' -10.42
## (9.62)
##
## 'Total Aportes' 0.11
## (0.15)
##
## 'Forma de Pago'Ventanilla -4.86*
## (2.58)
##
## Constant 38.02
## (30,300.91)
##
## -----
## Observations 3,174
## Log Likelihood -26.89
## Akaike Inf. Crit. 129.77
## =====
## Note: *p<0.1; **p<0.05; ***p<0.01

```

Según estos resultados, se evidencia la irrelevancia de algunas variables dentro del modelo, como lo es el estado civil y la ocupación, por lo que es pertinente eliminarlas del análisis, con el fin de mejorar la validez estadística de las demás variables, así como del modelo en conjunto. Entonces se estima un segundo modelo:

```

##
## Modelo 2: sin nivel de estudio y ocupacion
## =====
## Dependent variable:
## -----
## 'Non Default'
## -----
## Monto -0.06*
## (0.04)
##
## Codeudor -1.69*
## (0.94)
##
## Garantia 7.40***
## (2.61)

```

##	
## Edad	-0.01
##	(0.03)
##	
## Ingresos	0.22
##	(0.21)
##	
## Antigüedad laboral	-0.0002*
##	(0.0001)
##	
## Antigüedad asociado	-0.001
##	(0.0004)
##	
## Sexo	0.77
##	(0.82)
##	
## Plazo	0.02
##	(0.02)
##	
## Altura mora	-0.26***
##	(0.04)
##	
## 'Estado civil'Divorciado	-1.90
##	(1.38)
##	
## 'Estado civil'Otro	8.32
##	(10,456.78)
##	
## 'Estado civil'Separado	12.59
##	(2,665.18)
##	
## 'Estado civil'Soltero	-1.27
##	(1.04)
##	
## 'Estado civil'Union libre	-1.95*
##	(1.04)
##	
## 'Estado civil'Viudo	-2.23
##	(3.73)
##	
## 'Tipo de contrato'Otro	17.68
##	(3,984.81)
##	
## 'Tipo de contrato'Prestacion de Servicios	11.29
##	(6,364.42)
##	
## 'Tipo de contrato'Termino Fijo	-1.97
##	(3.36)

```

##
## 'Tipo de contrato'Termino Indefinido          0.75
##                                           (3.08)
##
## 'Saldo Capital'                                0.05
##                                           (0.06)
##
## 'Tasa Efectiva Colocacion'                    -6.25
##                                           (7.95)
##
## 'Total Aportes'                                0.10
##                                           (0.13)
##
## 'Forma de Pago'Ventanilla                     -3.07*
##                                           (1.75)
##
## Constant                                     15.95***
##                                           (5.00)
##
## -----
## Observations                                3,174
## Log Likelihood                             -34.31
## Akaike Inf. Crit.                          118.63
## =====
## Note:                                     *p<0.1; **p<0.05; ***p<0.01

```

Hasta este punto, la significancia del Modelo 2 ha mejorado con respecto a la del Modelo 1; sin embargo aun se encuentran variables que no tienen un efecto importante, que para este caso son el estado civil y el tipo de contrato laboral. Teniendo en cuenta la evidencia empírica, además del efecto de cada una de estas variables de manera individual sobre el Default, se decide evaluar un modelo que incluya dos opciones de cada una de estas variables categóricas:

- ¿Contrato a termino indefinido? 1 = Si | 0 = No (demás casos)
- ¿Es casado/a? 1 = Si | 0 = No (demás casos)

```

##
## Modelo 3: incluye dicotomas Termino indefinido y Casado
## =====
##                               Dependent variable:
##                               -----
##                               'Non Default'
##                               -----
## Monto                        -0.07**
##                               (0.03)
##
## Codeudor                     -1.58*
##                               (0.87)

```

##	
##	Garantia
##	7.28***
##	(2.58)
##	
##	Edad
##	-0.01
##	(0.03)
##	
##	Ingresos
##	0.22
##	(0.20)
##	
##	Antigüedad laboral
##	-0.0002
##	(0.0001)
##	
##	Antigüedad asociado
##	-0.001
##	(0.0004)
##	
##	Sexo
##	0.74
##	(0.79)
##	
##	Plazo
##	0.02
##	(0.02)
##	
##	Altura mora
##	-0.25***
##	(0.04)
##	
##	'Termino Indefinido'
##	1.26
##	(1.50)
##	
##	'Saldo Capital'
##	0.06
##	(0.06)
##	
##	Casado
##	1.30
##	(0.84)
##	
##	'Tasa Efectiva Colocacion'
##	-6.99
##	(7.62)
##	
##	'Total Aportes'
##	0.05
##	(0.12)
##	
##	'Forma de Pago'Ventanilla
##	-2.52
##	(1.64)
##	
##	Constant
##	13.05***
##	(3.27)
##	
##	-----
##	Observations
	3,174

```
## Log Likelihood                -36.38
## Akaike Inf. Crit.            106.75
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

El Modelo 3, en comparación con el Modelo 1 y 2, presenta un mejor nivel significativo de las variables en conjunto, así mismo como del nivel explicativo del modelo en general.

Interpretacion de coeficientes

Los coeficientes de la regresión del modelo seleccionado no explican un efecto probabilístico sobre el Default, sin embargo el tipo de efecto si, sea positivo o negativo. Al tratarse de un modelo Logit, se debe realizar la transformación pertinente de los coeficientes de las variables, con el fin de determinar el efecto porcentual o el peso de cada variable. (El Exp(OR) mas alto, indica un mayor nivel explicativo de la variable dentro del modelo.

##	OR	X2.5..	X97.5..
## (Intercept)	463237.0201	1607.3784	721292338.9094
## Monto	0.9364	0.8790	1.0065
## Codeudor	0.2054	0.0358	1.1367
## '% Cobertura de la Garantia (Cierre)'	1448.2640	9.6844	328086.5173
## Edad	0.9882	0.9301	1.0553
## Sueldo	1.2446	0.8876	1.6941
## 'Antigüedad laboral dias'	0.9998	0.9996	1.0001
## 'Antigüedad asociado dias'	0.9994	0.9986	1.0002
## Sexo	2.0870	0.4153	10.0255
## Num.Cuotas	1.0247	0.9785	1.0769
## 'Dias vencidos'	0.7787	0.7108	0.8273
## 'Termino Indefinido'	3.5305	0.1315	51.5707
## 'Saldo Capital'	1.0574	0.9563	1.2082
## Casado	3.6792	0.7785	22.4495
## 'Tasa Efectiva Colocacion'	0.0009	0.0000	2157.0948
## 'Total Aportes'	1.0564	0.8480	1.3456
## 'Forma de Pago'Ventanilla	0.0803	0.0021	1.3313
##	(...)		
## (Intercept)	463237.0201000000		
## Monto	0.0679196924		
## Codeudor	3.8685491723		
## '% Cobertura de la Garantia (Cierre)'	1448.2640000000		
## Edad	0.0119409027		
## Sueldo	0.2446000000		
## 'Antigüedad laboral dias'	0.0002000400		
## 'Antigüedad asociado dias'	0.0006003602		
## Sexo	2.0870000000		
## Num.Cuotas	0.0247000000		
## 'Dias vencidos'	0.2841916014		
## 'Termino Indefinido'	3.5305000000		

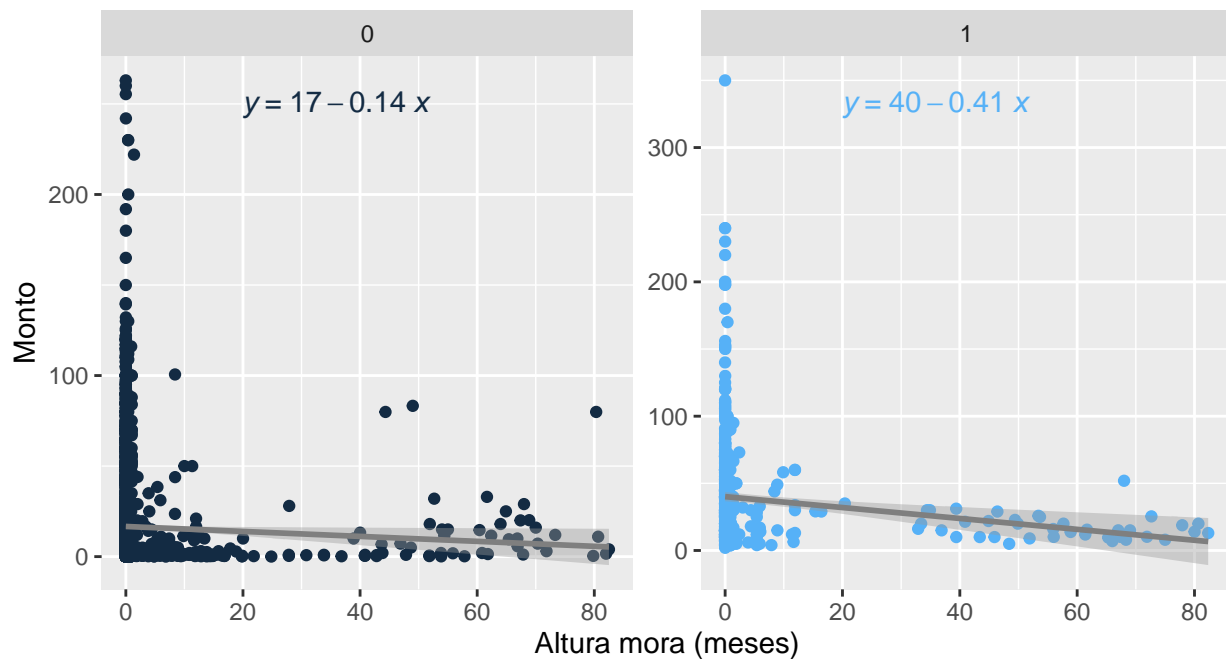
## 'Saldo Capital'	0.0574000000
## Casado	3.6792000000
## 'Tasa Efectiva Colocacion'	1110.1111111111
## 'Total Aportes'	0.0564000000
## 'Forma de Pago'Ventanilla	11.4533001245

Propuestas

1. Implementacion Internal Credit Scoring
2. Modificacion parametros estandar de validacion y analisis crediticio
3. Modificacion lineas y modalidades de credito
4. Optimizacion en la asignacion de productos crediticios potencialmente demandables (credit card)
5. Optimizacion en parametros de garantia (Requisito de codeudor):

Requisito de codeudor: Altura de mora VS Monto

A corte octubre de 2022



Teniendo en cuenta la regresión, la dispersión de los datos permite determinar la relación entre variables y por tanto la tendencia entre el monto de crédito y la probabilidad de altura de mora en meses, según se ha determinado en las siguientes funciones:

Sin codeudor:

$$y_0 = 17 - 0.14x \quad (1)$$

Con codeudor

$$y_1 = 40 - 0.41x \quad (2)$$

donde para ambos casos, “x” representa la variable altura de mora en meses. Siendo la altura de mora ideal máximo de 1 mes para obtener la calificación A en obligaciones de consumo, se sustituye $x = 1$ (un mes) en (1) y (2); y se determina:

Sin codeudor:

$$y_0 = 16.86 \quad (3)$$

Con codeudor

$$y_0 = 39.59 \quad (4)$$

Los resultados (3) y (4) se interpretan en millones de pesos, e indican que hasta estos montos existe un buen comportamiento de pago en una solicitud de crédito para los casos donde existe o no la figura de codeudor.

Conclusiones

- El nivel de estudios y el tipo de ocupación de los asociados no son relevantes en el modelo general, así como tampoco un modelo univariado, por lo que de manera concluyente, no logran explicar ningún efecto sobre el default.
- La variable de nivel de estudios y estado civil en su conjunto categórico, no tienen efecto sobre el default, sin embargo la categoría individual de si la relación laboral es con contrato a termino indefinido y el estado civil es casado, mejoran el nivel explicativo del modelo.
- Patrones de comportamiento y tendencia de consumo financiero obtenidos a través de Data Mining permite diseñar nuevos modelos financieros, con el fin de proponer proyectos encaminados a la optimización del portafolio de servicios financieros, así como mejorar y perfeccionar los procesos establecidos.
- Mayor conocimiento del mercado a partir de ETL y la explotación eficiente de datos, es posible la obtención de información y generación de conocimiento en el contexto de la inteligencia de negocios.