



# ANÁLISIS PARA LA PREDICCIÓN DE CLIENTES QUE PODRÍAN CAER EN MORA DE LA AUTO FINANCIERA

Caso de negocio  
IMEC. Jimena G. Gordillo  
<sup>1</sup> Agosto 14, 2023

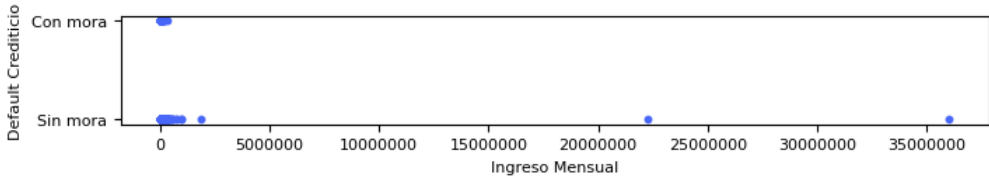
## Agenda

- › **ANÁLISIS EXPLORATORIO DE DATOS**
- › **PREPROCESAMIENTO DE DATOS**
- › **MODELADO**
- › **INTERPRETACIÓN**

## > Estructura del dataframe

Nom		Cont	Nom		Nom	Bin	Cont	Nom	Target Bin
ID_Credito	Modelo	Monto_prestamo	EntidadFederativa	ActividadEconomica	FLG_Empleo	IngresoMensual	Propiedad_Casa	Default_Crediticio	
1234	2326111	KICKS	418001.0	NAYARIT	NEGOCIO PROPIO	0	80000.0	FAMILIA O PADRES	0
2316	2410100	MARCH	189902.0	MORELOS	TAXISTA	0	220000.0	PROPIETARIO	0
11405	2352035	SENTRA	284801.0	COAHUILA	EMPLEADO PRIVADO	0	0.0	FAMILIA O PADRES	0
4052	2337108	SENTRA	289601.0	JALISCO	ECONOMIA INFORMAL	0	2500.0	PROPIETARIO	0
8451	2421003	VERSA	203900.0	NUEVO LEON	TAXISTA	0	21000.0	PROPIETARIO	0
13887	2358084	SENTRA	211774.0	CHIHUAHUA	EMPLEADO PRIVADO	0	18731.0	PROPIETARIO	0
7738	2412103	VERSA	196400.0	QUERETARO	TAXISTA	0	23500.0	RENTADA	0
6228	2302109	VERSA	205902.0	MICHOACAN	TAXISTA	0	25000.0	FAMILIA O PADRES	0
7025	2388063	VERSA	266800.0	SONORA	EMPLEADO PRIVADO	0	26000.0	PROPIETARIO	0
11258	2406027	VERSA	242100.0	JALISCO	EMPLEADO PRIVADO	0	23000.0	PROPIETARIO	0

De acuerdo al ingreso mensual, se puede observar que las personas que ganan aproximadamente menos de \$50,000 pesos mensuales, cayeron en mora.



Número de personas que cayeron en mora: 1984, 14.21% de la población total

### Variables cuantitativas

Cont: continua

Bin: Binaria

### Variables cuantitativas

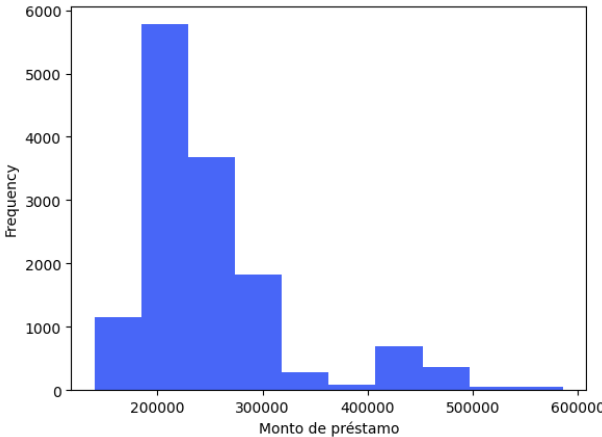
Nom: Nominal

## > Resumen numérico de variables cuantitativas

### Medidas de tendencia central para variables cuantitativas

	ID_Credito	Monto_prestamo	FLG_Empleo	IngresoMensual	Default_Crediticio
count	1.396000e+04	13960.000000	13960.000000	1.396000e+04	13960.000000
mean	2.753327e+06	253532.083295	0.000072	2.501755e+04	0.142120
std	1.468176e+06	72460.789827	0.008464	3.593882e+05	0.349186
min	2.137108e+06	140401.000000	0.000000	0.000000e+00	0.000000
25%	2.345036e+06	211400.000000	0.000000	5.000000e+03	0.000000
50%	2.378020e+06	229801.000000	0.000000	2.000000e+04	0.000000
75%	2.410112e+06	273200.000000	0.000000	2.500000e+04	0.000000
max	9.042001e+06	585576.000000	1.000000	3.602400e+07	1.000000

### Monto de préstamo



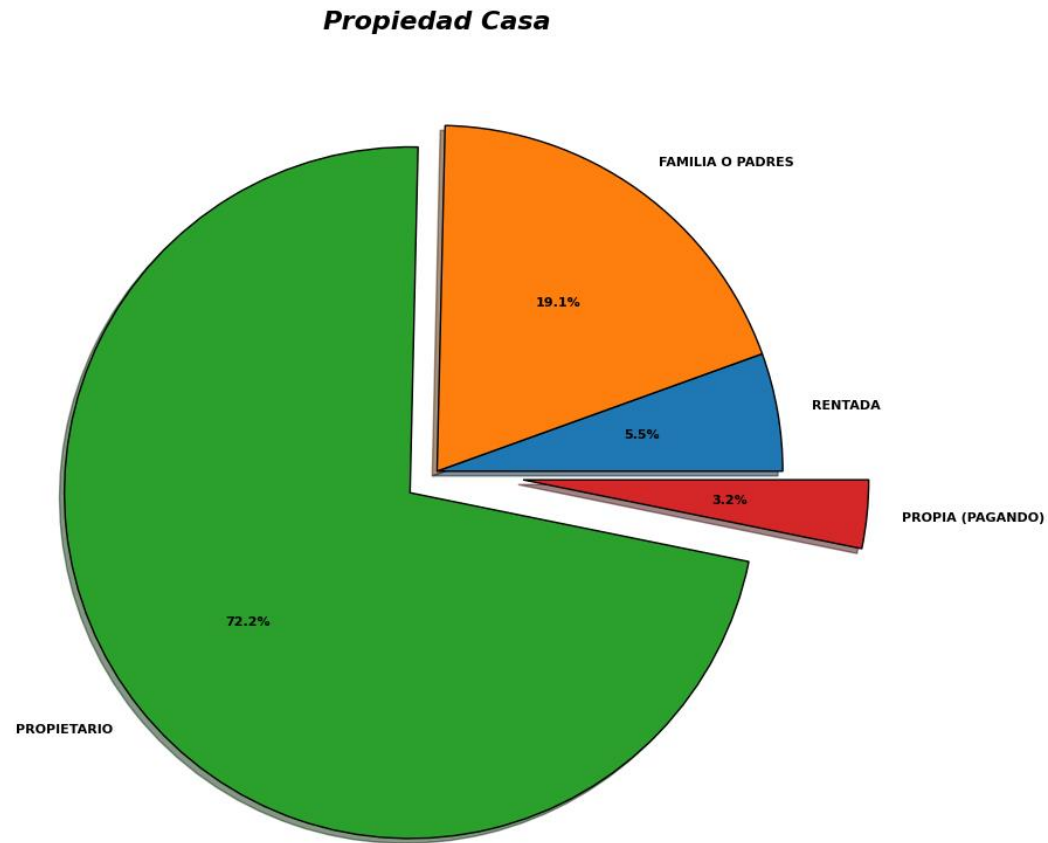
Cantidad en promedio de lo que se ha prestado

Mediana: 229801.0

Moda: 211902.0

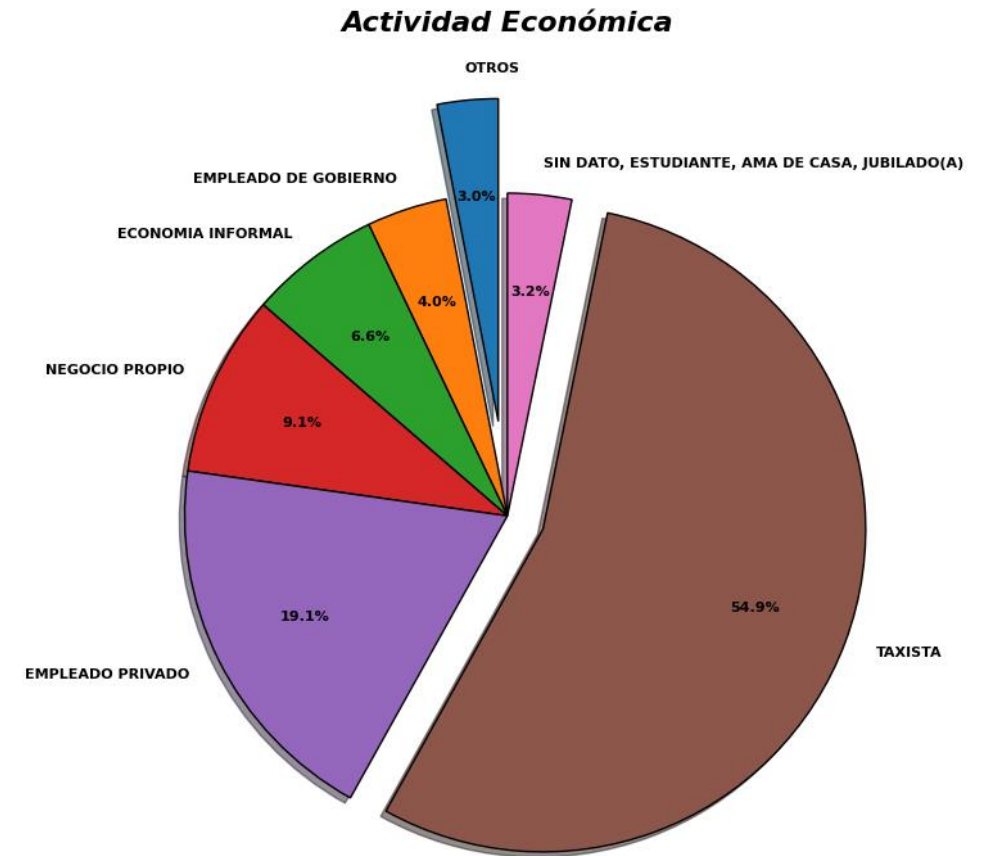
Rango de Dispersión: 445175.0

Porcentaje de los clientes respecto a la propiedad de la casa



- > El 72.2% de las personas son **propietarios de una casa**
- > Mientras que el 3.2% de personas han liquidado su propiedad

Porcentaje de los clientes respecto a la actividad económica que ejercen



- > El 54.9% de las personas su actividad económica registrada son **taxistas**
- > El 3.2% pertenece al conjunto de personas con el menor número de incidencias en las actividades económicas: estudiante, ama de casa, jubilado(a) o sin dato.

## > Tabla de contingencia con la prueba de chi cuadrado

Se generó una tabla *one-hot* a partir del dataframe original para pasar las variables categóricas a variables binarias. A continuación, se muestra una parte del dataframe final.

	Modelo_ALTIMA	Modelo_KICKS	Modelo_MARCH	Modelo_MAXIMA	Modelo_MURANO	Modelo_NOTE	Modelo_NP300	Modelo_OTROS
0	False	False	True	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	True	False	False	False	False	False
3	False	False	True	False	False	False	False	False
4	False	False	False	False	False	False	True	False
...	...	...	...	...	...	...	...	...
13955	False	False	False	False	False	False	False	False
13956	False	False	False	False	False	False	False	False
13957	False	False	True	False	False	False	False	False
13958	False	False	False	False	False	False	False	False
13959	False	False	True	False	False	False	False	False

Datos categóricos con mayor asociación con respecto al target

Cabe señalar que los propietarios y los empleados privados obtuvieron 0

Variable	P-valor
Propiedad_Casa_PROPIETARIO	0
Propiedad_Casa_FAMILIA O PADRES	0
ActividadEconomica_JUBILADO(A)	0
ActividadEconomica_EMPLEADO PRIVADO	0
EntidadFederativa_TAMAULIPAS	0
Propiedad_Casa_RENTADA	0
Modelo_VERSA	0
EntidadFederativa_VERACRUZ	0
ActividadEconomica_ESTUDIANTE	0
ActividadEconomica_NEGOCIO PROPIO	0
ActividadEconomica_EMPLEADO DE GOBIER	0.01
EntidadFederativa_JALISCO	0.01
EntidadFederativa_GUANAJUATO	0.02
EntidadFederativa_TABASCO	0.02
EntidadFederativa_DURANGO	0.03
EntidadFederativa_NAYARIT	0.03
Modelo_MARCH	0.04
ActividadEconomica_AMA DE CASA	0.04
EntidadFederativa_CHIHUAHUA	0.06
EntidadFederativa_CIUDAD DE MEXICO	0.06
ActividadEconomica_OTROS	0.06
EntidadFederativa_NUEVO LEON	0.07
Modelo_NP300	0.09
EntidadFederativa_HIDALGO	0.09
ActividadEconomica_ECONOMIA INFORMAL	0.09
Modelo_URVAN	0.1
EntidadFederativa_CAMPECHE	0.13
Modelo_KICKS	0.14
ActividadEconomica_SIN DATO	0.18
EntidadFederativa_OAXACA	0.2

Tabla 2. Los P-Valores ordenados de menor al mayor. Resultados en resultadosPorChiCuadrada.csv

### ¿Qué es la prueba de chi cuadrado?

La prueba de chi cuadrado se utiliza para **determinar si existe una asociación** estadísticamente significativa entre **dos conjuntos de datos categóricos**.

- > Si el valor p es **pequeño** (generalmente  $< 0.05$ ), se concluye que **hay una asociación significativa entre las variables**
- > Si el valor p es **grande**, **no se encuentra suficiente evidencia** para afirmar una asociación significativa.

El valor p no proporciona información sobre la fuerza de la asociación, solo indica si es estadísticamente significativa.

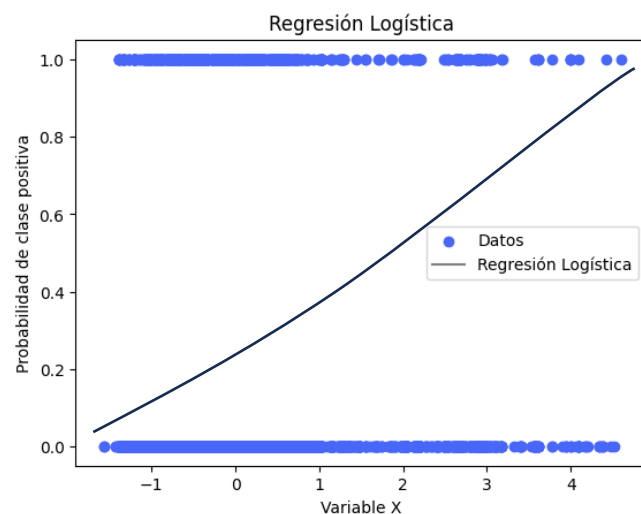
El valor p se generó con relación a cada atributo categórico con relación al target *Default\_Crediticio* para conocer la influencia que tiene cada atributo en dicha variable objetivo

## Informe de clasificación

	precision	recall	f1-score	support
0	0.87	1.00	0.93	3015
1	0.77	0.02	0.04	475
accuracy			0.87	3490
macro avg	0.82	0.51	0.48	3490
weighted avg	0.85	0.87	0.81	3490

- > Precisión: 87% clasificó correctamente las instancias con predicciones positivas.
- > Sensibilidad (recall): se clasificó correctamente el 100% de las instancias de la clase 0.
- > F1: 93%, lo que indica un buen equilibrio entre la precisión y el recall.
- > Support: 3015 instancias de la clase 0 y 475 instancias de la clase 1.

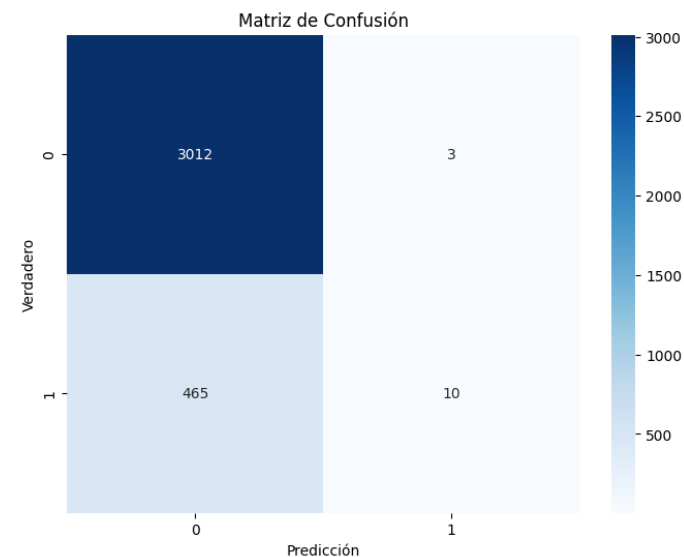
## Gráfica de regresión Logística



En la gráfica, el eje x representa el tiempo de estudio, mientras que el eje y representa la probabilidad de éxito.

La curva trazada en la gráfica muestra cómo cambia la probabilidad de éxito a medida que aumenta el tiempo de estudio.

## Matriz de confusión



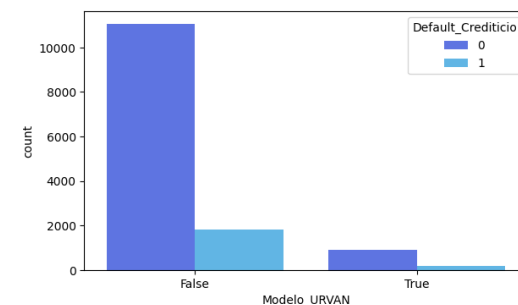
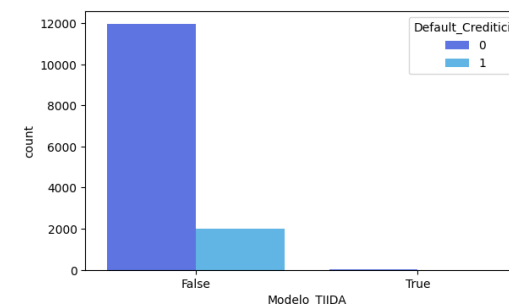
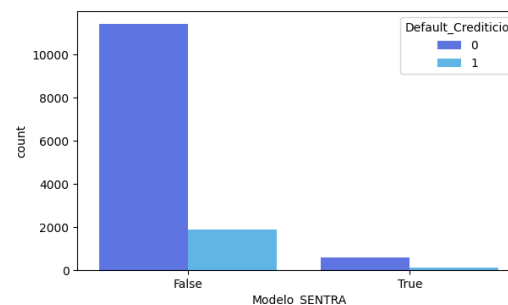
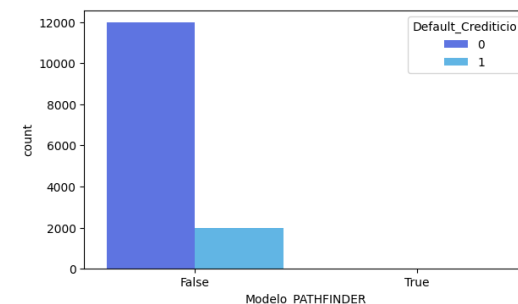
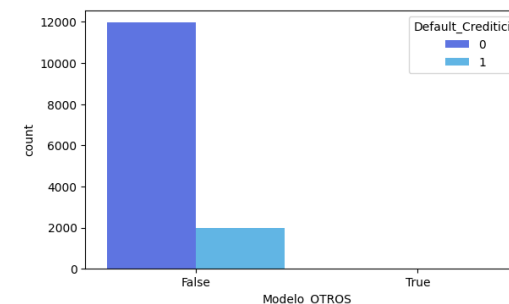
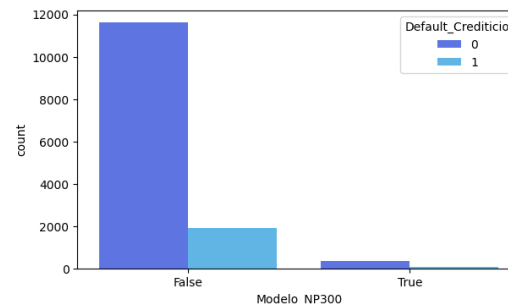
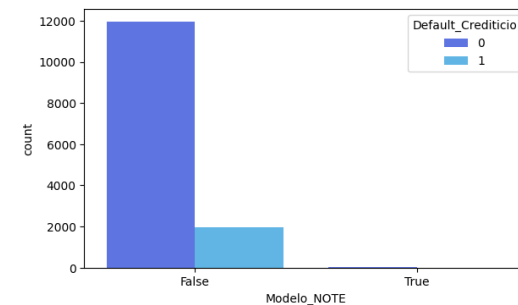
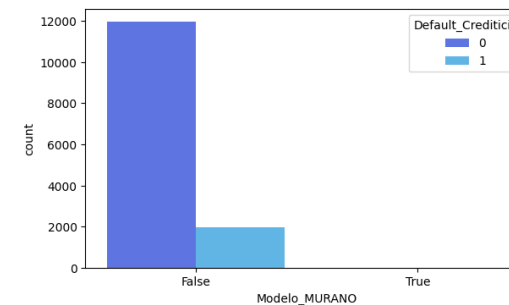
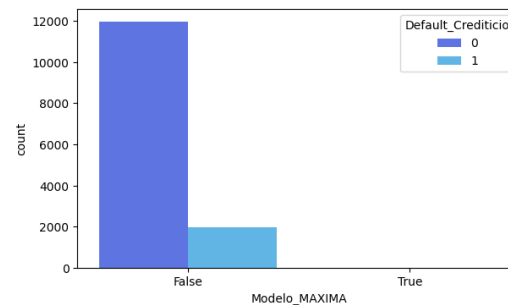
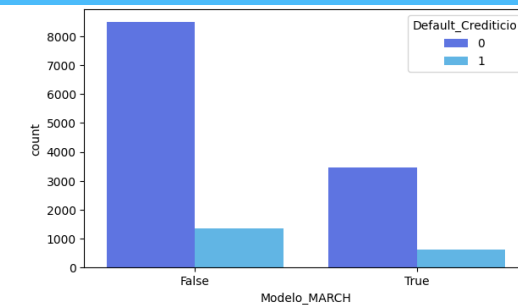
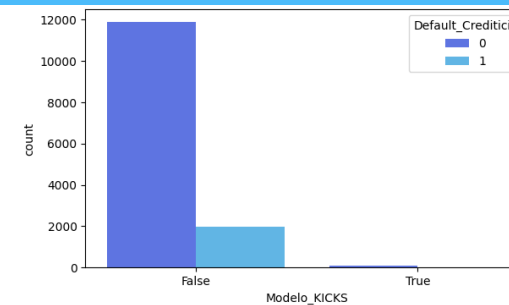
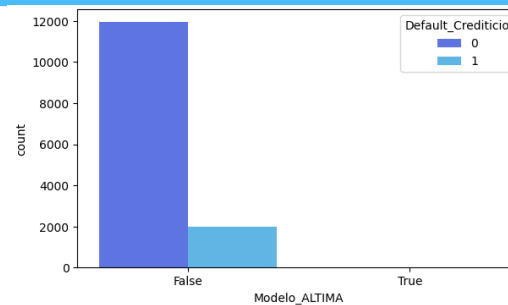
- > Verdaderos positivos (TP): Hay 3012 instancias que pertenecen a la clase positiva (1) y que el modelo ha clasificado correctamente como positivas.
- > Falsos negativos (FN): Hay 3 instancias que pertenecen a la clase positiva (1), pero que el modelo ha clasificado incorrectamente como negativas.
- > Falsos positivos (FP): Hay 465 instancias que pertenecen a la clase negativa (0), pero que el modelo ha clasificado incorrectamente como positivas.
- > Verdaderos negativos (TN): Hay 10 instancias que pertenecen a la clase negativa (0) y que el modelo ha clasificado correctamente como negativas.

## Insights por Modelo del auto

- Variable Default Crediticio indica si el cliente dejó de pagar el crédito, donde 1 es que lo dejó de pagar y 0 es que no tiene mora.
- Las barras que pertenecen a la sección False, indican el número de veces que no se compró el modelo y se eligió otro, caso contrario para las barras en True.

### Insights

- Se puede observar que, el modelo MARCH ha sido uno de los modelos más comprados y más de 3,000 personas no están en mora al obtener el modelo.
- El segundo lugar de los autos más comprados es el modelo URVAN.
- Los últimos lugares de compra son SENTRA y NP300.
- Los modelos que no fueron comprados fueron ALTIMA, KICKS, MAXIMA, MURANO, MODELO NORTE
- PATHFINDER, TIDA y OTROS. Aunque no han sido comprados estos modelos, las personas que obtuvieron el préstamo en su mayoría no cayeron en mora.





## > Coeficientes

### ¿Qué realiza el coeficiente en relación con la variable target?

- > Cuantificar la relación entre una variable dependientes (target )y una o más variables independientes.
- > Si el coeficiente es **positivo**, indica que a medida que la **variable independiente aumenta**, la **variable dependiente** tiende a **aumentar** también.
- > Si el coeficiente es **negativo**, indica que medida que la **variable independiente aumenta**, la **variable dependiente** tiende a **disminuir**

De la siguiente gráfica es apreciable ver que las de personas propietarias de una casa influye negativamente en un -2.03 (recordar que son el 72.2% de la población total). Esto significa que, las personas que son propietarias de una casa no tienden a caer en mora.

Mientras que el monto del préstamo indicó un coeficiente 0.37, lo que significa que, a medida que es más grande el monto del préstamo, aumenta las probabilidades de que la persona caiga en mora.

En el caso de la actividad económica más ejercida por el 54.9% de la población total, taxista, el coeficiente que obtuvo fue de 0.13, lo cual significa que las personas que son taxistas pueden caer en mora.

Los datos han sido guardados en el archivo *coefs.csv*

