

# Credit Risk Scoring

Como estrategia en la banca masiva

# Agenda



- 1 Banca Persona
- 2 Modelos de Credit Scoring
- 3 Aplicación del Modelo
- 4 ¿Cómo genera valor?

# Banca Persona



**Tradicional**



**Crédito vehicular**



**Crédito Hipotecario**



**Masivo**



**Tarjeta de Crédito**



**Préstamo personal**



**Compras**



**Disposición de efectivo**



**Línea paralela**

# ¿Cuál es la problemática?

## GESTIÓN

Suscríbete

### En el 2018 Aumentó la Morosidad Crediticia en el Perú

Autora: Elena Bobadilla

in Compartir

f Compartir

G+ Compartir

+



Wichaya Prasongsin / 123RF

25.08.2019 - 02:52 PM

Según el informe de la Asociación de Bancos del Perú (Asbanc) hasta el mes de octubre se concluyó que el índice de morosidad promedio en el 2018 alcanzó el 3.14 %. Esta cifra supera el ratio de morosidad bancaria que se cerró en el 2017 con 3.04 %, la cual ya

## GESTIÓN

### Jóvenes deben en promedio S/ 1,815 en tarjetas y su deuda vencida es la más alta

Por estilos de vida, las personas conservadoras tienen la mayor tasa de deuda vencida, según información a noviembre del 2018.

in

f

WhatsApp

+



La tasa de deudas vencidas en jóvenes es más alta en el uso de tarjeta de crédito que en el de préstamos personales. (Foto: Manuel Melgar)



### 523,000 peruanos tiene deudas mayores a seis veces su sueldo en tarjetas de crédito

Un reciente estudio de la consultora Experian detalla que de este grupo 130 mil peruanos adeuda montos **50 veces mayores a sus ingresos**.

24 de julio del 2018 - 1:00 PM Redacción

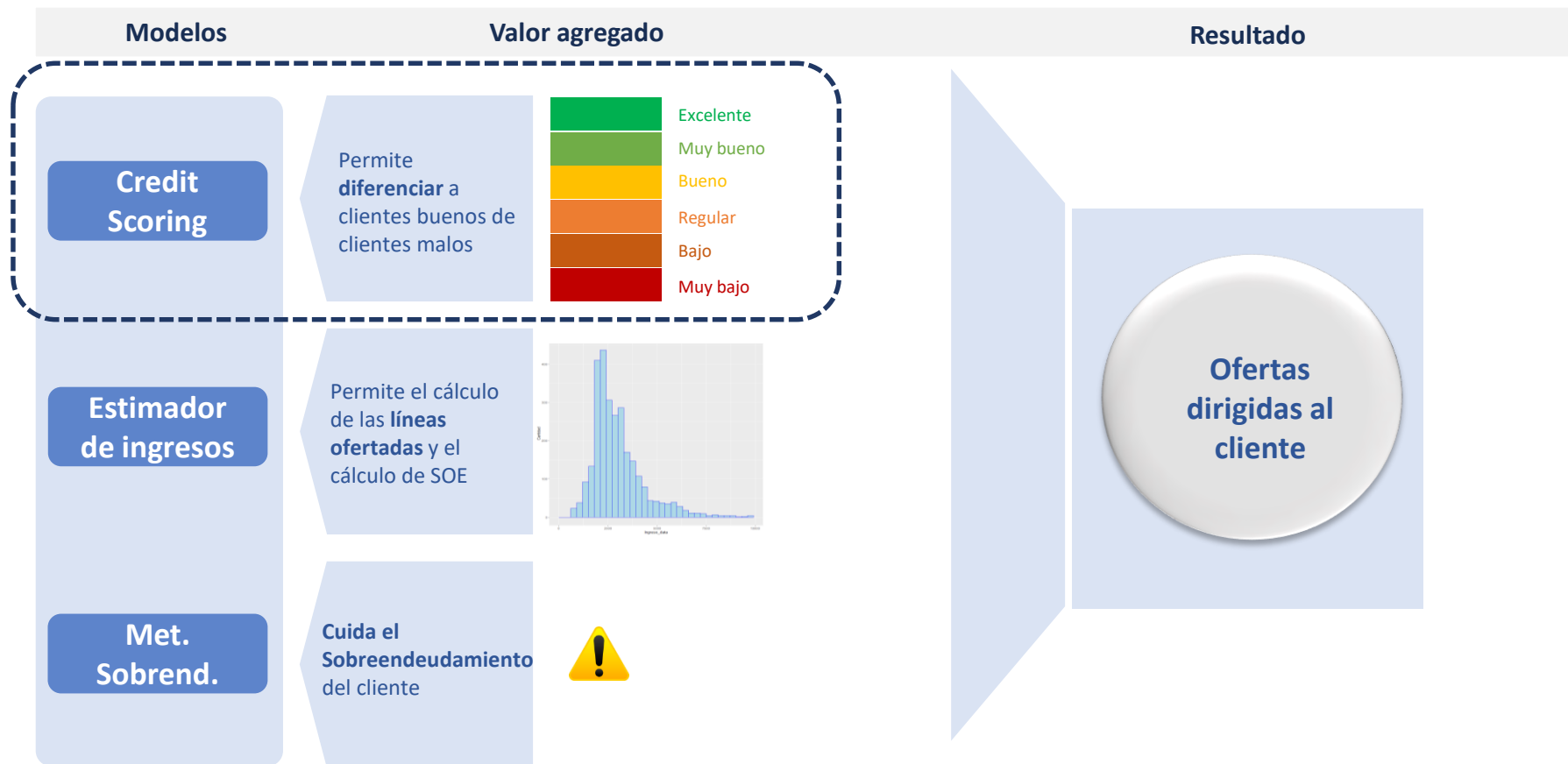


Más de medio millón de peruanos debe más de seis veces su sueldo, de acuerdo con la consultora. | Fuente: Andina

- No hay diferencias de condiciones de riesgos
- Decisiones subjetivas, mala gestión de riesgos

Identificar aquellos clientes con buen comportamiento crediticio.

# Pilares de las estrategias de Riesgo de Crédito



# ¿Qué es Credit Scoring?



**Credit  
Scoring**

Son todas las técnicas y modelos estadísticos que ayudan a las instituciones financieras para el otorgamiento de un préstamo o una tarjeta de crédito.

## VENTAJAS



Facilita las decisiones y estrategias al identificar al cliente por su grado de riesgo de impago.

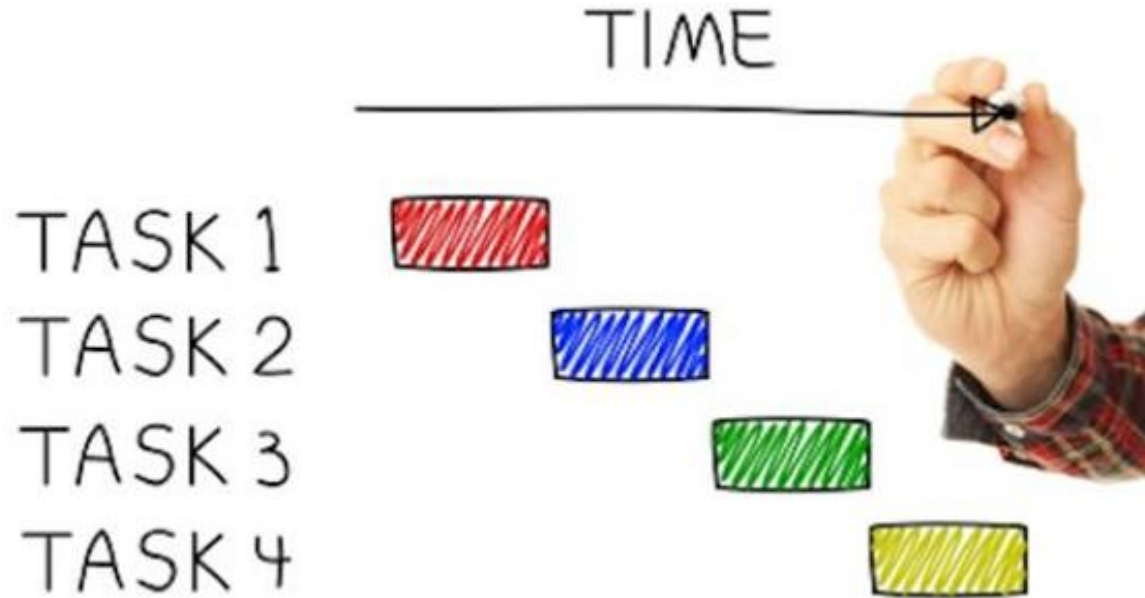
Permite conocer a la persona y su grado de exposición a créditos previo a la originación.



Disminuye el riesgo de las operaciones al dar crédito.

## ¿Cómo calculamos el puntaje?

# Fase 1: Comprensión del Negocio



# ¿Cuál es la finalidad del estudio?



Bancarizado	
	Nº cliente: <b>10 MM</b>
	Saldo S/ <b>127,000 MM</b>

## Objetivos

Discriminar a los clientes con alto y bajo riesgo de incumplimiento en el Sistema Financiero para la generación de nuevos prospectos.

## Población

Persona Natural del RCC ( 4 meses del año 2017).



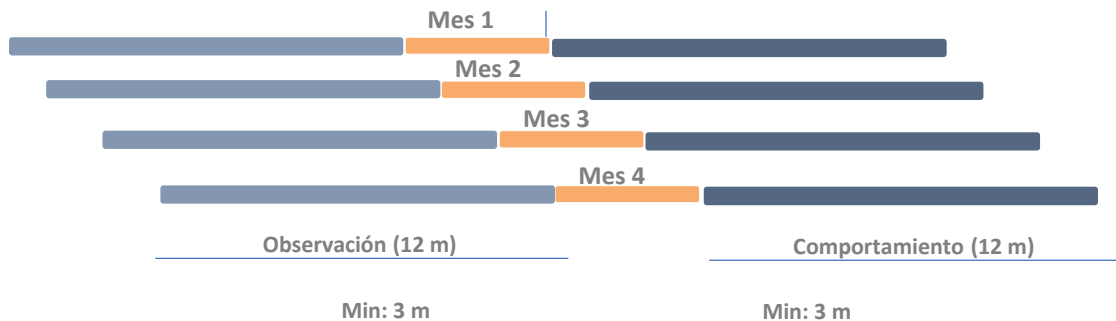
Fase 2:  
Comprensión  
de los datos



## Filtramos clientes



## Definición del Target (PD):



## Roll Rates Analysis

Morosidad T0	Morosidad A 12 meses					
	0.[0]	0.[1-8]	1.[9-30]	2.[31-60]	3.[61-120]	4.>120
0.[0]	67%	7%	13%	4%	4%	6%
0.[1-8]	12%	27%	19%	12%	10%	20%
1.[9-30]	19%	2%	27%	18%	15%	18%
2.[31-60]	13%	4%	7%	11%	27%	39%
3.[61-120]	7%	1%	8%	2%	12%	71%
4.>120	4%	0%	0%	0%	0%	95%
<b>Total</b>	<b>65%</b>	<b>7%</b>	<b>13%</b>	<b>4%</b>	<b>4%</b>	<b>7%</b>

- Bueno: durante los 12 meses tuvo 0 días de atraso.
- Indeterminado: clientes en la frontera de buenos y malos.
- Malo : en alguno de los próximos 12 meses tuvo más de 60 días de atraso.

Con ello hemos definido nuestro grupo objetivo ...

# Tenemos diversos grupos de variables...

## 30% Historial de Pagos

¿Con qué frecuencia paga sus cuentas a tiempo?

- Información de pagos de las deudas.

## 10% Nuevos créditos

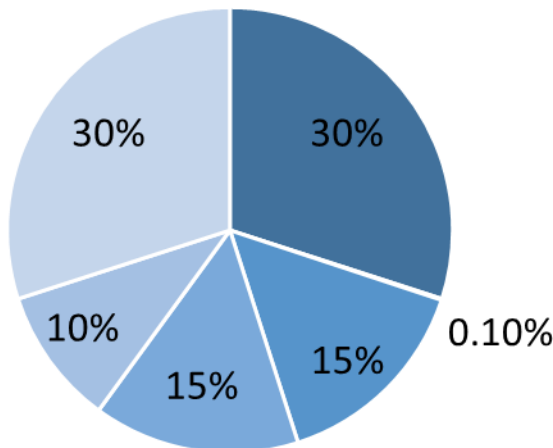
¿Has abierto una nueva línea de crédito?

- Número de entidades.
- Línea de TC.

## 15% Créditos utilizados

¿Tienes más de un crédito?

- Número de entidades con saldo.
- Número de entidades en estado castigado, vencido o refinanciado.
- Tipo de crédito.
- Líneas no utilizadas.



## 30% Montos adeudados

¿Cuánto de tu crédito total has sado?

Deudas por tipo de crédito.  
Montos castigados, vencidos, refinanciados.

## 0.1% Demográficos

¿Cuál es tu edad?

- Edad.
- Estado Civil.
- Ubigeo.

## 15% Historial crediticio

¿Hace cuanto tiempo tienes un crédito bueno o malo?

- Antigüedad en el SS.FF.
- Días de mora.
- Clasificación crediticia (Normal, CPP, deficiente, dudoso, pérdida)

## Partición de la muestra



**Entrenamiento 75%**  
**PD: 13.14%**



**Testeo 25%**  
**PD: 13.14%**

### Código en R

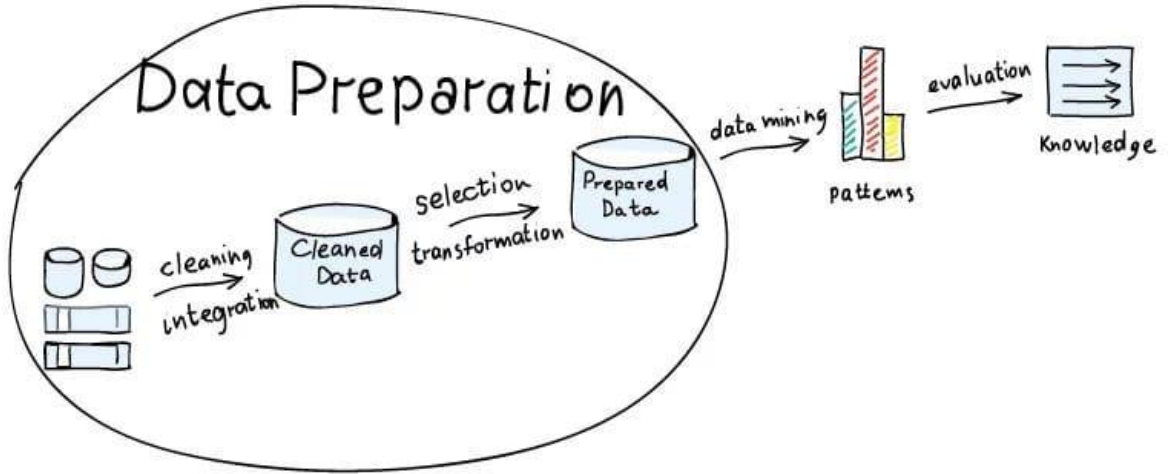
```
##Base_Modelo: Base con los registros
##Base_Modelo$FLG_MALO: Columna del Target

library(caTools)
set.seed(100)
muestra=sample.split(Base_Modelo$FLG_MALO,SplitRatio=0.75)

data.train=subset(Base_Modelo,muestra==TRUE)
data.test=subset(Base_Modelo,muestra==FALSE)
```

¿Por qué particionamos la base de datos?

## Fase 3: Preparación de los datos



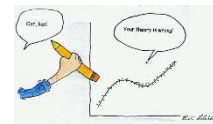
# Conociendo las Variables...

## Análisis Univariado de Variables Continuas

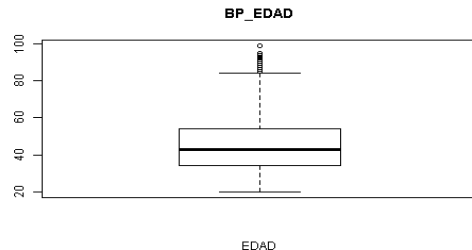
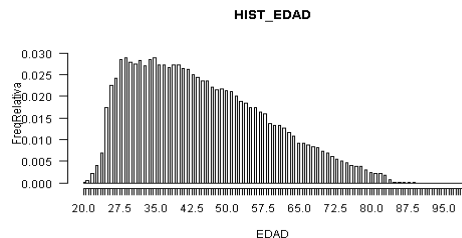
Valores Perdidos



Valores Extremos



Codigo	Variable	Obs	MValues	Min	Max	Media	P1	P2	P3	P4	P5	P95	P96	P97	P98	P99	P100
1	EDAD	364636	23941	20	99	44.94	24	25	25	26	26	71	73	74	77	80	99



### Código en R

```
library(XLConnect)
library(lattice)
library(StatMeasures)
library(agricolae)
##Creación de un libro de Excel
LIBRO_DES <- loadWorkbook(file="D:/UNIVARIADO/Var_Continuo1.xlsx", create =
TRUE)
##Creación de las hojas en el libro creado
createSheet(LIBRO_DES, name = "Estadisticas")
createSheet(LIBRO_DES, name = "Graficos")
##Creación de las estadísticas descriptivas de las variables en la Hoja
"Estadisticas"
Cuanti_Gen2_1(Base_Modelo, Variables, LIBRO_DES, "Estadisticas", " Graficos")
##Creación de gráficos descriptivos de las variables en la Hoja "Graficos"
Cuanti_Gen2_2(Base_Modelo, Variables, LIBRO_DES, "Estadisticas", " Graficos")
```

# ¿Cómo se relacionan las variables con el target?

## Poder Predictivo de la Categoría (WOE)

$$WOE = \ln\left(\frac{Prop\ Bueno}{Prop\ Malo}\right)\%$$

Categorías	PD		WOE
1	21%	↓	-0.571551898
2	16%	↓	-0.216400846
3	13%	→	-0.028657016
4	10%	→	0.292313947
5	6%	↑	0.861482203

## Código en R

```
library(XLConnect)
library(lattice)
##Creación de un libro de Excel
LIBRO_DES <-
loadWorkbook(file="D:/BIVARIADO/
3.Variables_Bivariadas.xlsx", create =
TRUE)

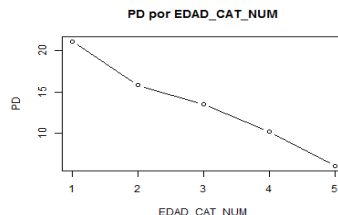
##Crear la hoja "Variables" en el libro
creado
createSheet(LIBRO_DES, name =
"Variables")
saveWorkbook(LIBRO_DES)
```

```
###Funciones
Calculo_Stat_Gen_Cuali
(Base_Modelo,FLG_MALO,variables,
LIBRO_DES,Variables,tipo)
```

## Poder Predictivo de la Variable (IV)

$$\sum_{i=1}^n (Prop\ Bueno_i - Prop\ Malo_i) * \ln\left(\frac{Prop\ Bueno_i}{Prop\ Malo_i}\right)$$

Intervalo de IV	Nivel de Predicción
< 2%	No Predictora
[2%-10%>	Débil
[10%-30%>	Medio
[30%-50%>	Fuerte
>= 50%	Posible Sobre Predictora

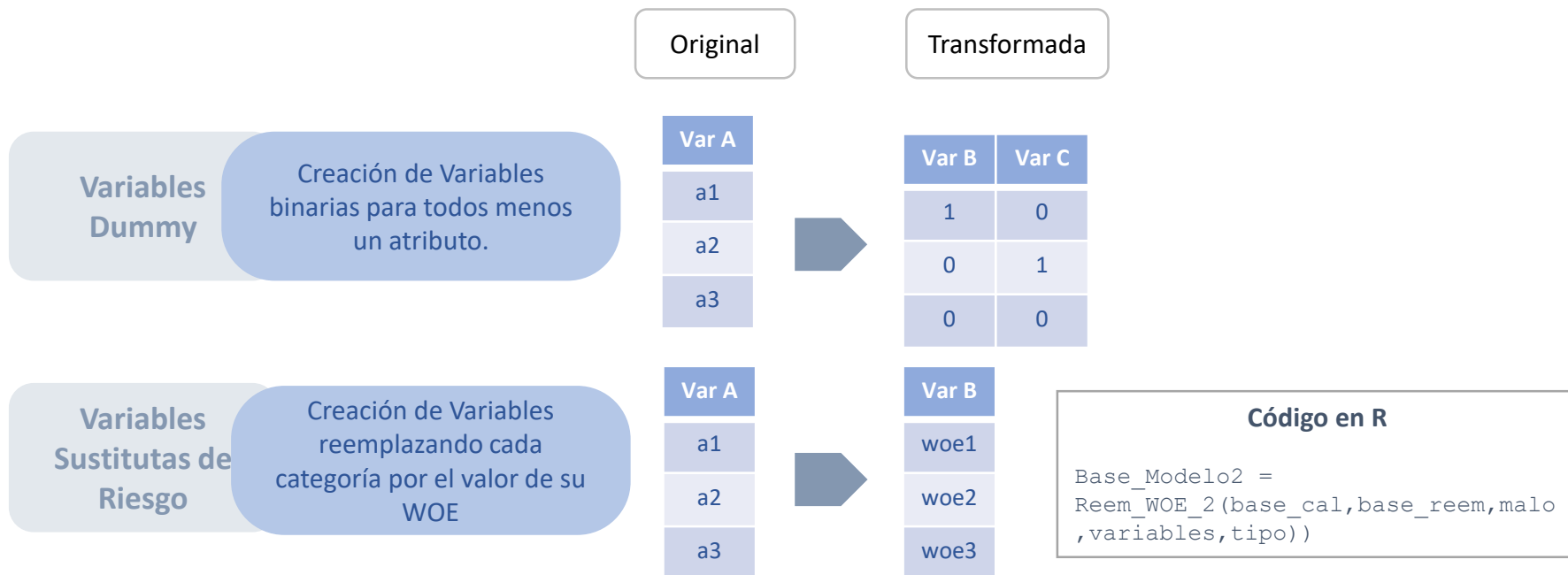


## Tendencia

*Tendencia Monotónica Creciente*  
*Tendencia Monotónica Decreciente*



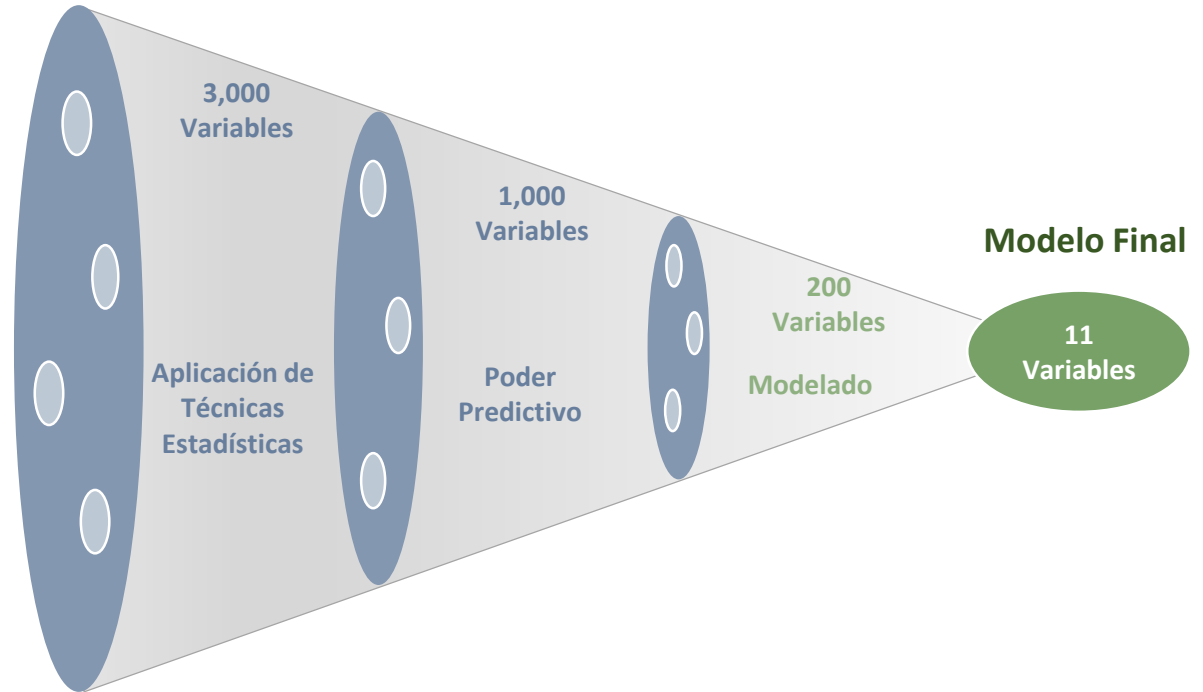
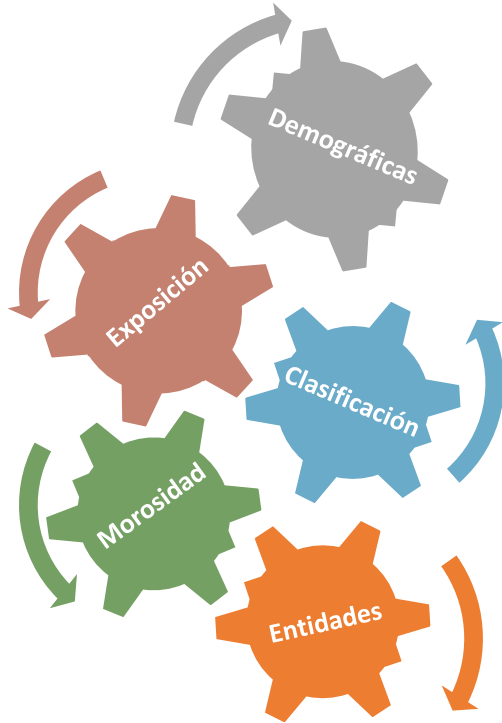
## ...algunas opciones de transformación de variables...



## Fase 4: Modelamiento



# Selección de Variables



# Regresión Logística

Fórmula

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

Donde

$p$  : Probabilidad de que un evento suceda.  
 $x$  : Variables predictoras.  
 $\beta_0$  : Intercepto.  
 $\beta_i$  : Parámetros.

¿Qué es?

Técnica supervisada de clasificación que en base a un conjunto de variables predice la probabilidad de ocurrencia de un evento, siendo en este caso el default de un cliente.

¿Por qué?

- Técnica más utilizada en Credit Scoring.

- Mejor Interpretación de los resultados.

- De fácil implementación

## Código en R

```
Modelo1=glm(FLG_MALO~Var1_WOE + Var2_WOE +  
Var3_WOE + Var4_WOE + Var5_WOE + Var6_WOE +  
Var7_WOE + Var8_WOE + Var9_WOE + Var10_WOE +  
Var11_WOE, data =  
data.train,family=binomial(link=logit))  
  
summary(Modelo1)
```

# ¿Cómo calcular el puntaje?

## Tabla de Puntuación - Scorecard



Convertir la probabilidad estimada del Modelo, a una escala de valores enteros (puntaje)

ODDS: Ventaja de N° Buenos respecto a N° Malos:  
**32 : 1**

$$Score = Offset + Factor * \ln(odds)$$

Factor: pdo/ $\ln(m)$   
**pdo = 80 m = 2**

$$Score = - \left( \sum_{j,i}^{k,n} WOE_j * \beta_i + \frac{\beta_0}{n} \right) * factor + offset$$

$$Score = - \left( \sum_{j,i}^{k,n} WOE_j * \beta_i + \frac{\beta_0}{11} \right) * \frac{80}{\ln(2)} + 600$$

EDAD	WOE	SCORE	PD
1	-0.5715519	-9	21%
2	-0.21640085	9	16%
3	-0.02865702	18	13%
4	0.29231395	35	10%
5	0.8614822	63	6%

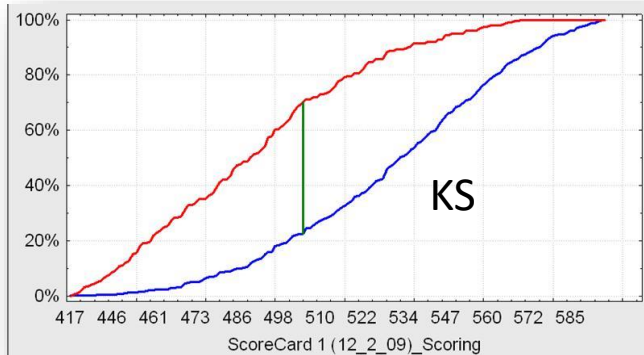
## Fase 5: Evaluación



# Indicador de Discriminación

**Kolgomorov - Smirnov  
(KS)**

Diferencia Máxima  
entre la distribución  
de clientes buenos y malos



## Código en R

```
Validacion= function(modelo,data,malo)
{
  MI_score <- predict(modelo,type='response',data)
  MI_pred <- prediction(MI_score, data[,malo])
  MI_perf <- performance(MI_pred,"tpr","fpr")
  MI_KS <- round(max(attr(MI_perf,'y.values')[[1]]-
                    attr(MI_perf,'x.values')[[1]])*100, 0)
  indicadores=cbind(MI_KS)
  return(indicadores)
}
```

```
Val.train=Validacion(Modelo1,data.train,"FLG_MALO")
```

```
Val.test=Validacion(Modelo1,data.test,"FLG_MALO")
```

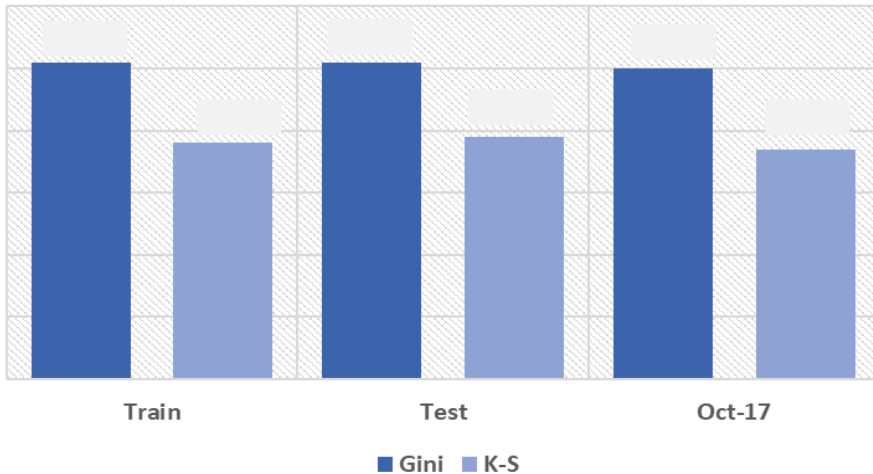
K-S Bueno

Poder de  
Discriminación

Semáforo	K - S
● Débil	< 30%
● Aceptable	[30% - 40%]
● Bueno	> 40%

### Backtesting:

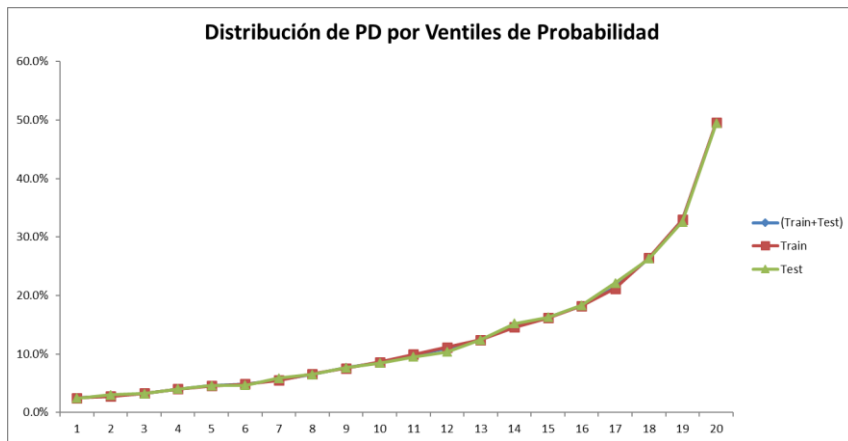
Se realizó la validación del modelo en una muestra “Out of time”. Los resultados indican un adecuado nivel de predictibilidad del riesgo de incumplimiento.





...así se refleja la distribución por segmentos...

## Distribución de PD por Ventiles

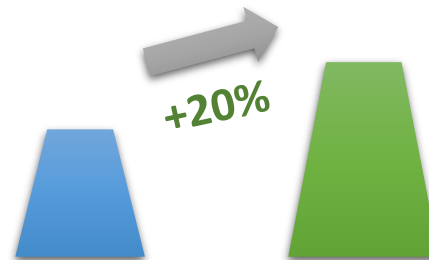


## Segmentos de Riesgos

Segmento	% Total	PD
Excepcional	22%	3%
Muy bueno	27%	7%
Bueno	25%	12%
Regular	10%	19%
Bajo	4%	25%
Muy bajo	5%	29%
Crítico	7%	45%

## ¿Qué se logró?

- Incremento en la Base de Prospección de Clientes



Segmento	Antes	Nuevo
Excepcional	3.6%	3.2%
Muy Bueno	6.8%	6.5%

- Disminución de Riesgo en los Mejores segmentos

- Incremento en los niveles de ROE y RORAC, en 22% y 14% respectivamente.

