

# CP004 - Préstamos y GLM

*Mayra Goicochea Neyra*

*4/11/2019*

## 1. Introducción

Se tiene un conjunto de datos de todos los préstamos emitidos en el periodo 2008-2015, que incluye el estado actual del préstamo (Current, Late, Fully Paid, etc.), los datos del prestatario, y la información más reciente de pago. La toma de decisiones con respecto a los riesgos crediticios es una de las operaciones cruciales para los negocios de banca y finanzas. En este análisis, se realiza un análisis exploratorio de datos, visualizaciones, predicción de resultados de préstamos y la representación de la matriz de confusión.

### 1.1 Dataset Overview

El dataset se encuentra en un archivo database.sqlite. Incluye 887383 casos con 75 variables, entre las que se consideran relevantes para el modelo predictivo:

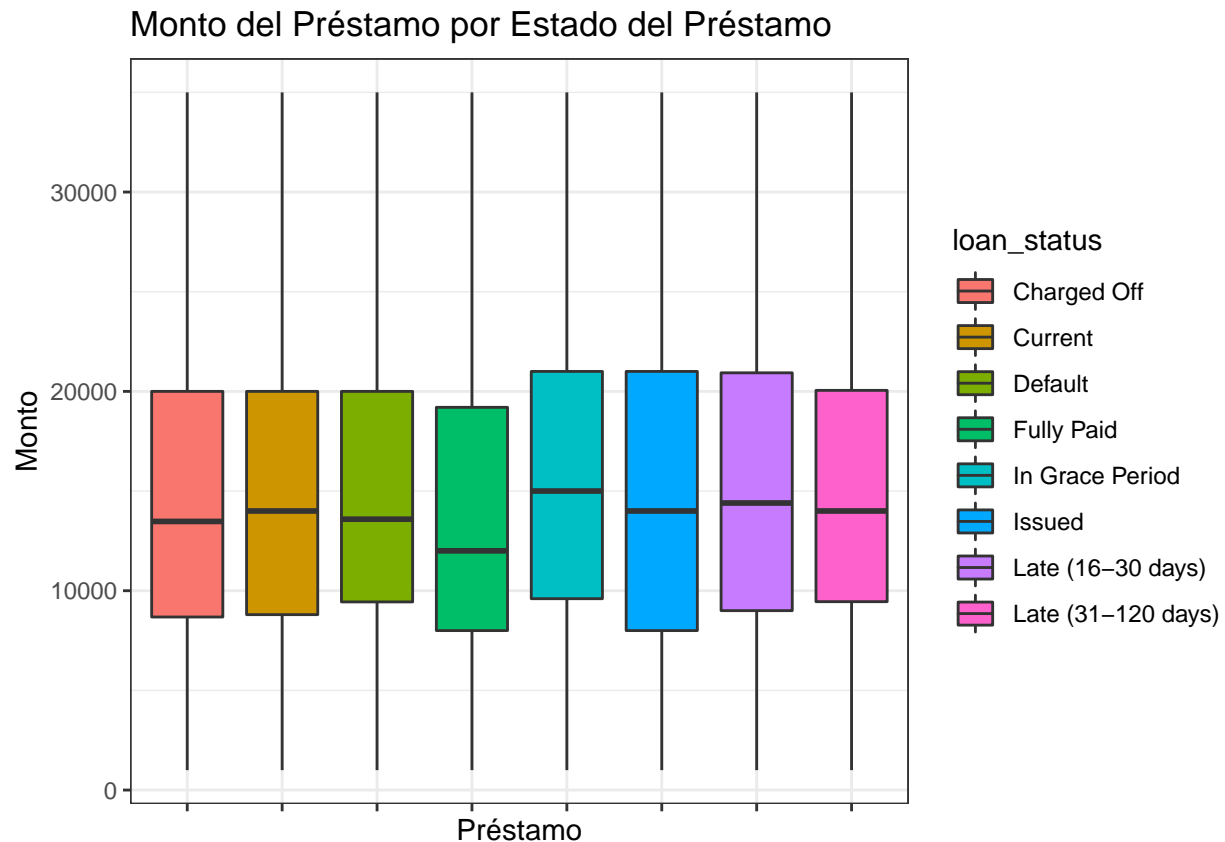
- `loan_amnt` : El monto del préstamo solicitado por el prestatario.
- `term` : El número de pagos del préstamo.
- `int_rate` : Tasa de interés del préstamo.
- `grade` : Grado de préstamo.
- `sub_grade` : Préstamo asignado.
- `emp_length` : Duración del empleo en años.
- `home_ownership` : El estado de propiedad de la vivienda proporcionado por el prestatario durante el registro.
- `annual_inc` : El ingreso anual autoinformado proporcionado por el prestatario durante el registro.
- `verification_status` : Si el registro ha pasado por el proceso de verificación.
- `issue_d` : Mes en que se financió el préstamo.
- `loan_status` : Estado actual del préstamo.
- `purpose` : Propósito del préstamo proporcionado por el prestatario para la solicitud de préstamo.
- `addr_state` : Código del estado de residencia proporcionado por el prestatario.
- `dti` : Ratio calculado en base a los pagos mensuales totales de la deuda sobre las obligaciones de la deuda total, excluyendo la hipoteca y el préstamo LC solicitado, dividido por el ingreso mensual autoinformado del prestatario.
- `delinq_2yrs` : Número de incidencias de morosidad (mayor a 30 días) durante los 2 últimos años.
- `inq_last_6mths` : Número de consultas en los últimos 6 meses (excluyendo consultas de automóviles e hipotecas).
- `mths_since_last_delinq` : Número de meses desde la última morosidad del prestatario.
- `mths_since_last_record` : Número de meses desde el último registro público.
- `open_acc` : Número de líneas de crédito abiertas en el archivo de crédito del prestatario.
- `pub_rec` : Número de registros públicos derogados.

- `revol_bal` : Saldo rotativo total de crédito.
- `revol_util` : Tasa de utilización de línea rotativa, o la cantidad de crédito que el prestatario está usando en relación con todo el crédito rotativo disponible.
- `total_acc` : Número total de líneas de crédito actualmente en el archivo de crédito del prestatario.
- `out_prncp` : Capital pendiente restante por el monto total financiado.
- `application_type` : Indica si el préstamo es una solicitud individual o una solicitud conjunta con dos coprestatarios.
- `tot_cur_bal` : Saldo total actual de todas las cuentas.

### 1.1.1. Loan Status

Existen 8 estados de préstamo. Para realizar el modelo de riesgo crediticio se considera sólo la información histórica (estados diferente a Current). La variable Objetivo será creada en base a ésta información. Los casos default serán aquellos con etiqueta Default o Charged Off (que se considera créditos impagos).

- 1 (Yes) : Default o Charged Off.
- 0 (No) : Fully Paid, In Grace Period, Issued, Late.

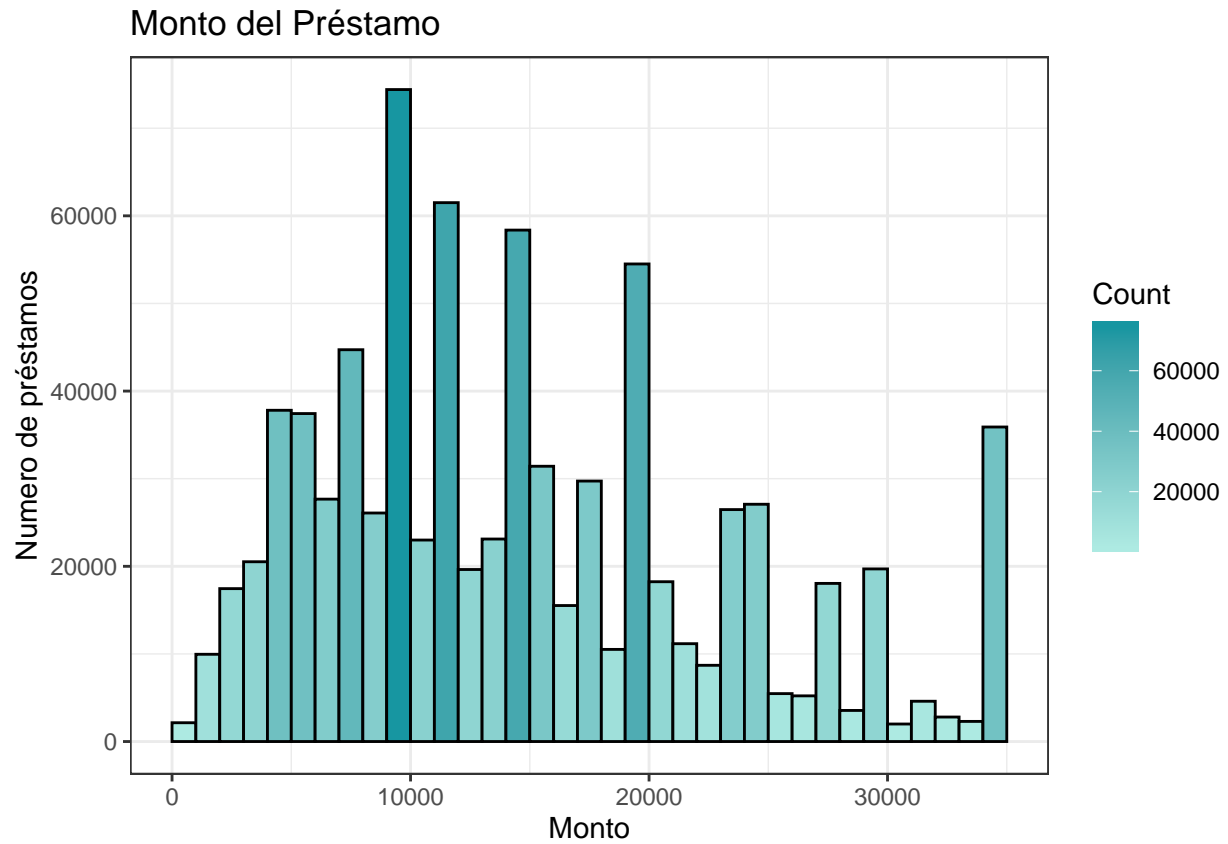


Los casos que se tienen en el dataset son los siguientes:

```
## loan_status
##      No    Yes
## 182370 36406
```

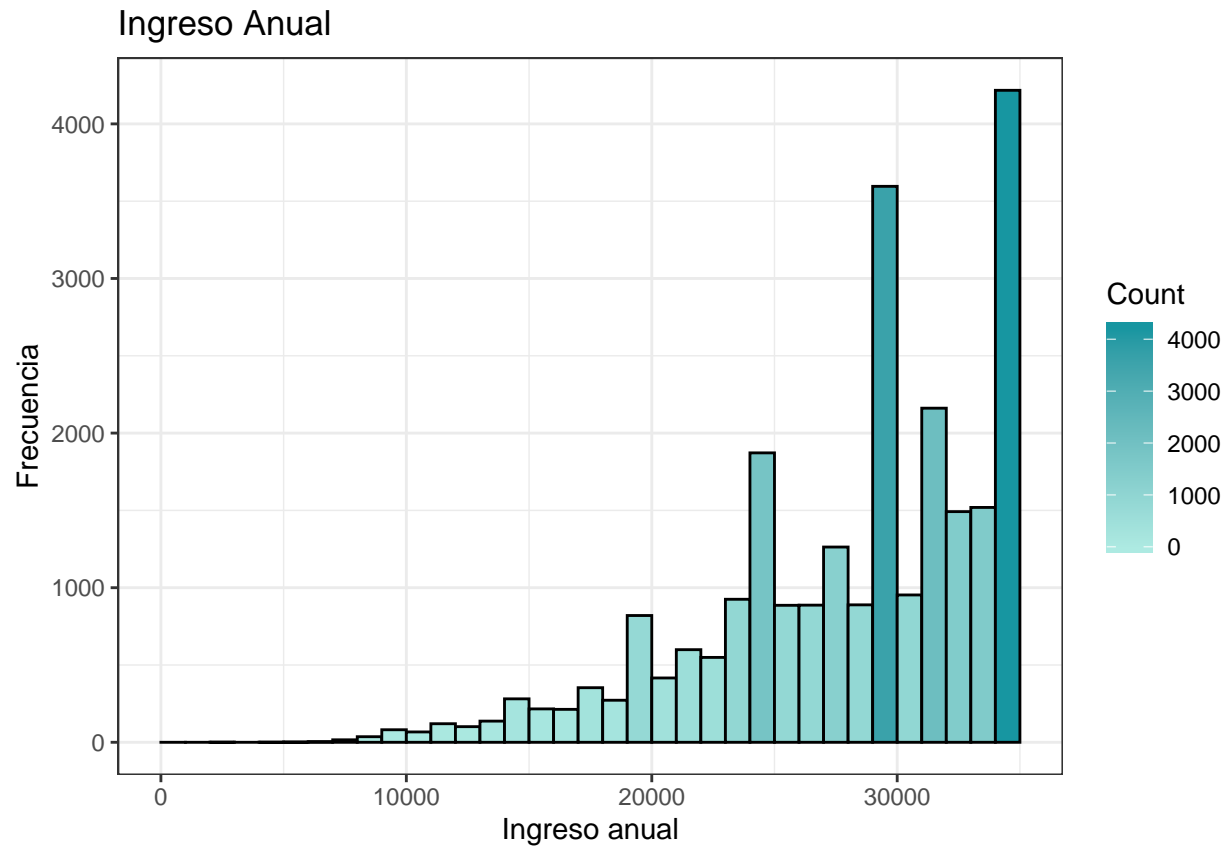
### 1.1.2. Loan Amount

Se tiene mucha variabilidad en los montos otorgados de préstamos con una desviación típica de 8446.32. Siendo el monto promedio de 14998.54.



### 1.1.3. Annual Income

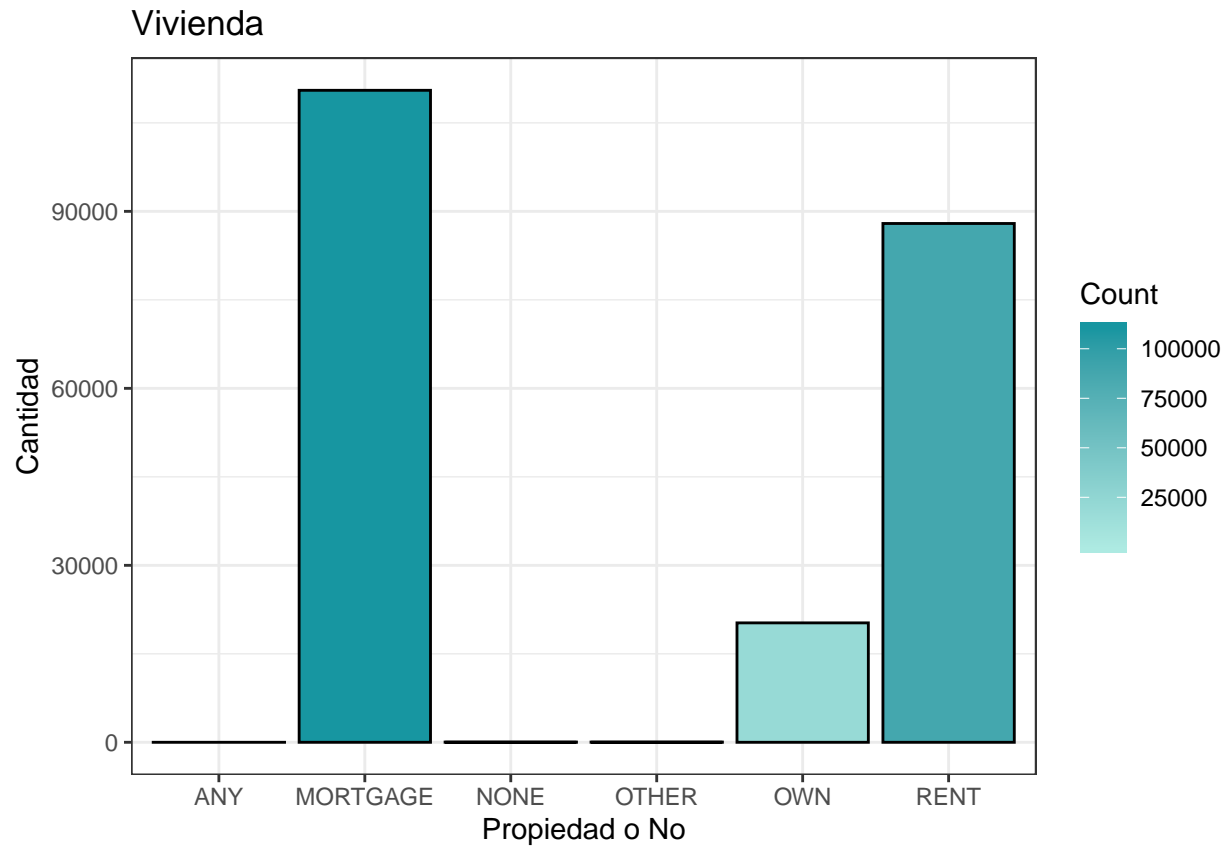
También tiene alta variabilidad (desviación típica de 65307.91) e ingreso anual promedio de 75544.93.



#### 1.1.4. Home Ownership

Hay 6 tipos de propiedad: Vivienda propia, renta, vivienda hipotecada, ninguna vivienda, otros. Se clasificarán en dos grupos:

- 1 (Yes) : OWN o Mortgage.
- 0 (No) : Rent, Any, Other, None.



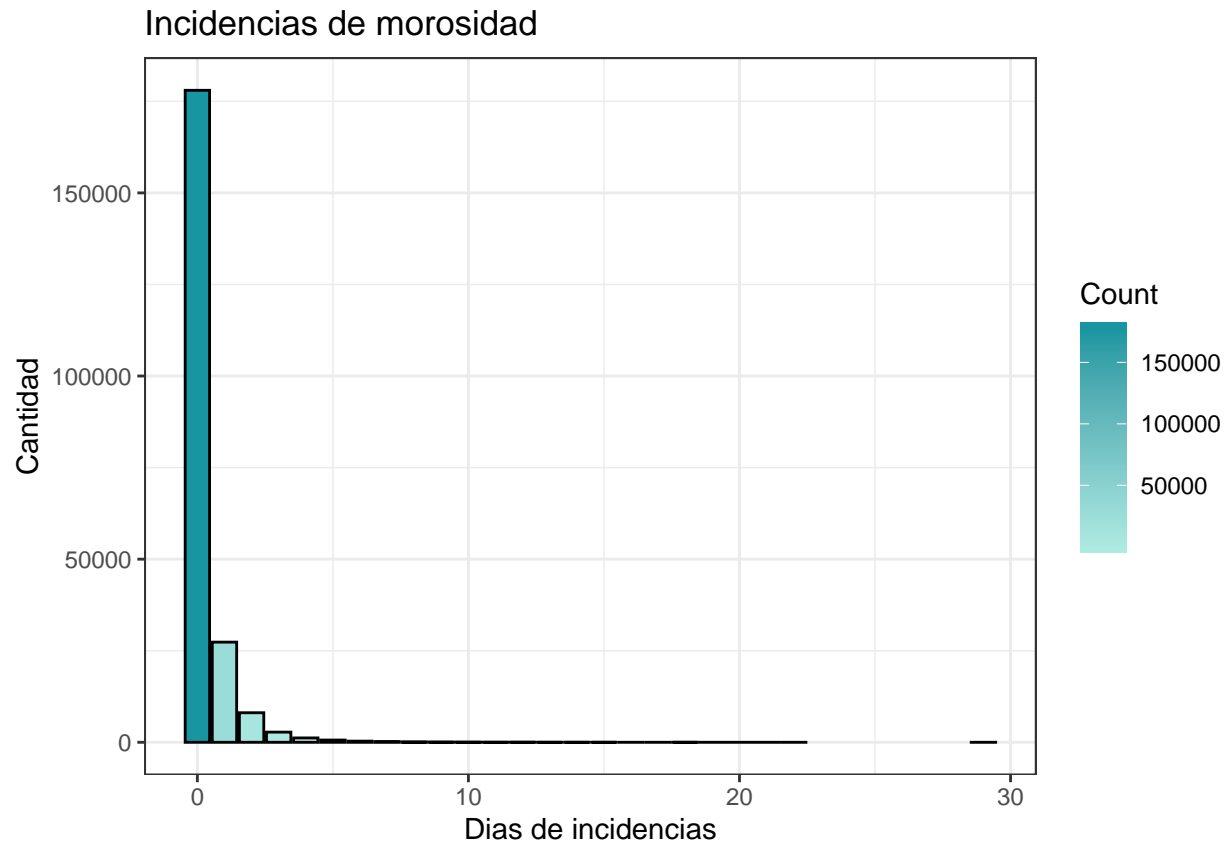
Se tienen los siguientes casos:

```
## home_ownership
##      No      Yes
## 88024 130752
```

#### 1.1.5. Delinq 2yrs

Se clasificara los casos en dos grupos:

- 1 (Yes) : Hay o hubo incidencias de morosidad.
- 0 (No) : Ninguna incidencia.



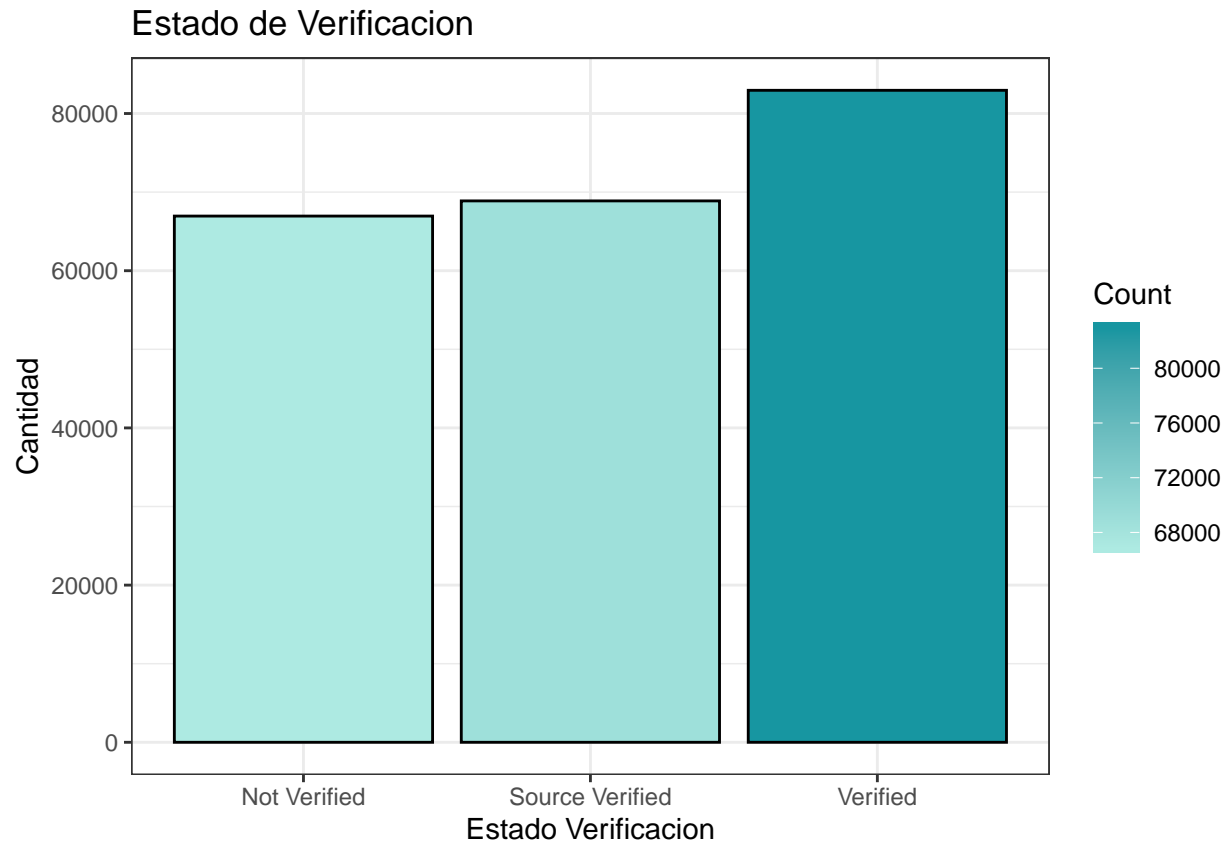
Los casos se distribuyen de la siguiente manera:

```
## delinq_2yrs
##      No      Yes
## 40776 178000
```

#### 1.1.6. Verification status

Se clasificara los casos en dos grupos:

- 1 (Yes) : Verified, Source Verified.
- 0 (No) : Not Verified.



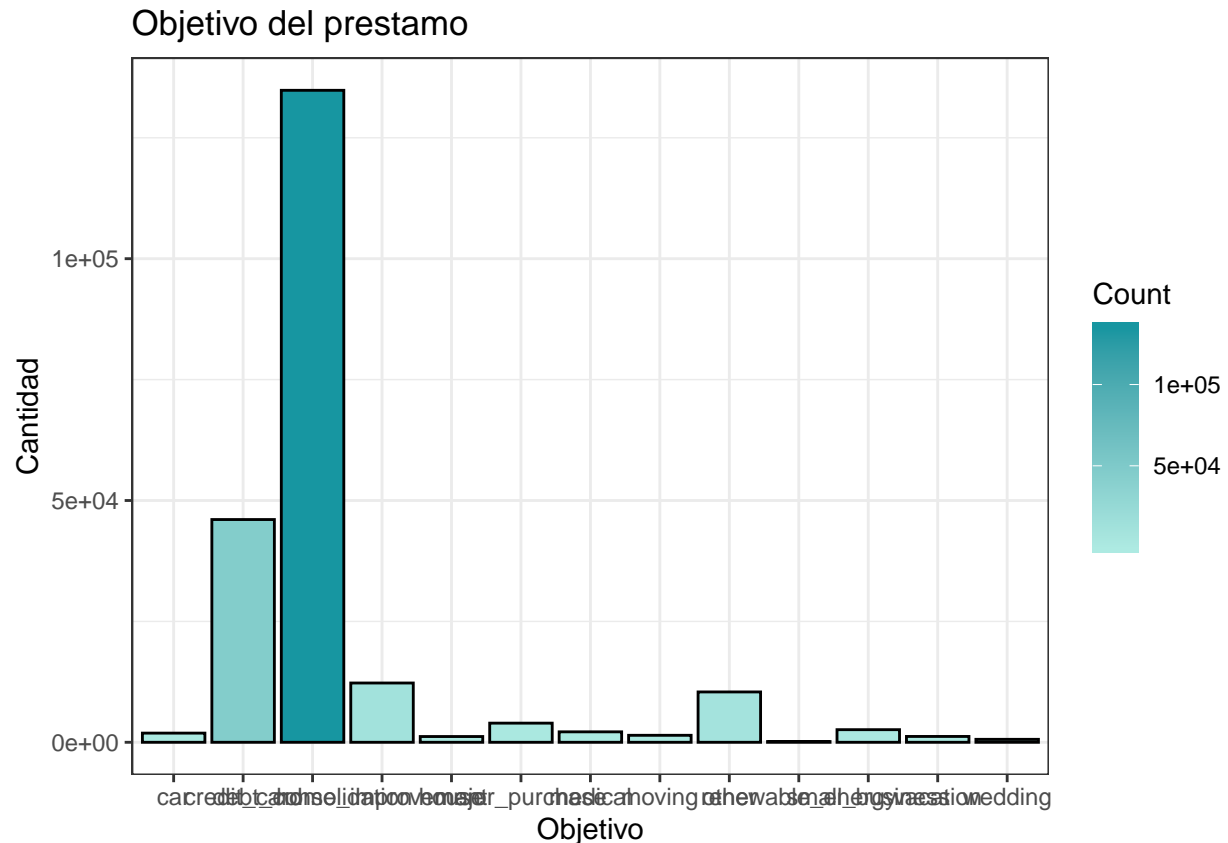
Los casos se distribuyen de la siguiente manera:

```
## verification_status
##      No      Yes
## 66948 151828
```

#### 1.1.7. Purpose

Se clasificara los casos en tres grupos:

- 0 : Moving, small business, renewable energy.
- 1 : Car, major purpose, home improvement, credit card.
- 2 : Educational, debt consolidation, medical, house, vacation, wedding, other.



Los casos se distribuyen de la siguiente manera:

```
## purpose
##      0      1      2
##  4210  64175 150391
```

## 2. Modelos GLM

Se elaboran tres modelos logísticos:

- **Modelo 0** : Todas las variables explicativas
- **Modelo 1** : Se consideran que las variables loan\_amnt, term, grade, home\_ownership, annual\_inc, purpose, dti, delinq\_2yrs, inq\_last\_6mths, open\_acc, revol\_util, total\_acc, out\_prncp y tot\_cur\_bal son suficientes para explicar variable default.
- **Modelo 1** : Se consideran que las variables loan\_amnt, term, grade, home\_ownership, annual\_inc, purpose, dti, delinq\_2yrs, inq\_last\_6mths, open\_acc, revol\_util, total\_acc, out\_prncp y tot\_cur\_bal son suficientes para explicar la variable default.

Se tienen los subsets \* **Train**:

```
## [1] 175020      26
```

- **Test**:

```
## [1] 43756      26
```



## 2.1. Estimando modelos logísticos

### 2.1.1. Modelo 0

##	Estimate	Std. Error	z value	Pr(> z )
## (Intercept)	-3.2368e+00	2.1784e-01	-14.8581	< 2.2e-16
## loan_amnt	1.6767e-05	1.1215e-06	14.9508	< 2.2e-16
## term 60 months	2.9609e-01	1.9211e-02	15.4123	< 2.2e-16
## int_rate	1.4308e+00	1.6148e+00	0.8861	0.375588
## gradeB	1.2654e+00	1.6298e-01	7.7644	8.204e-15
## gradeC	1.6150e+00	2.0025e-01	8.0649	7.327e-16
## gradeD	1.7669e+00	2.4453e-01	7.2257	4.986e-13
## gradeE	1.9993e+00	2.9546e-01	6.7668	1.316e-11
## gradeF	2.0641e+00	3.4079e-01	6.0569	1.388e-09
## gradeG	2.3863e+00	3.9333e-01	6.0671	1.303e-09
## sub_gradeA2	3.0685e-01	1.4323e-01	2.1423	0.032168
## sub_gradeA3	3.9458e-01	1.3600e-01	2.9014	0.003715
## sub_gradeA4	6.9981e-01	1.2571e-01	5.5669	2.593e-08
## sub_gradeA5	8.0727e-01	1.2626e-01	6.3937	1.620e-10
## sub_gradeB1	-3.0951e-01	7.7792e-02	-3.9787	6.929e-05
## sub_gradeB2	-2.9315e-01	6.2811e-02	-4.6672	3.053e-06
## sub_gradeB3	-1.2946e-01	5.0865e-02	-2.5452	0.010921
## sub_gradeB4	-7.3881e-02	4.4319e-02	-1.6670	0.095512
## sub_gradeC1	-2.7588e-01	5.7302e-02	-4.8145	1.476e-06
## sub_gradeC2	-1.9962e-01	4.9675e-02	-4.0185	5.858e-05
## sub_gradeC3	-1.0848e-01	4.3940e-02	-2.4689	0.013553
## sub_gradeC4	-4.9313e-02	4.0061e-02	-1.2310	0.218337
## sub_gradeD1	-1.1616e-01	6.0304e-02	-1.9263	0.054069
## sub_gradeD2	-6.4115e-02	5.4702e-02	-1.1721	0.241171
## sub_gradeD3	-8.4731e-02	5.2053e-02	-1.6278	0.103566
## sub_gradeD4	4.1082e-02	4.9093e-02	0.8368	0.402690
## sub_gradeE1	-1.9905e-01	7.7345e-02	-2.5735	0.010068
## sub_gradeE2	-4.4907e-02	7.1460e-02	-0.6284	0.529729
## sub_gradeE3	-1.0920e-01	6.9292e-02	-1.5759	0.115042
## sub_gradeE4	-7.5932e-02	6.7824e-02	-1.1195	0.262907
## sub_gradeF1	-1.6936e-01	1.0254e-01	-1.6517	0.098592
## sub_gradeF2	-1.7222e-01	1.0430e-01	-1.6513	0.098681
## sub_gradeF3	9.2505e-02	1.0354e-01	0.8935	0.371612
## sub_gradeF4	-2.1926e-02	1.0912e-01	-0.2009	0.840744
## sub_gradeG1	-3.5270e-01	2.0256e-01	-1.7412	0.081641
## sub_gradeG2	-4.6148e-01	2.1252e-01	-2.1715	0.029894
## sub_gradeG3	-2.5054e-01	2.2231e-01	-1.1270	0.259746
## sub_gradeG4	-5.3641e-01	2.5339e-01	-2.1169	0.034265
## emp_length1 year	-5.0017e-02	3.5872e-02	-1.3943	0.163222
## emp_length10+ years	-1.2240e-01	2.7403e-02	-4.4667	7.943e-06
## emp_length2 years	-1.0906e-01	3.3168e-02	-3.2881	0.001009
## emp_length3 years	-8.2818e-02	3.4261e-02	-2.4172	0.015639
## emp_length4 years	-1.0296e-01	3.7426e-02	-2.7510	0.005940
## emp_length5 years	-8.8244e-02	3.5733e-02	-2.4695	0.013530
## emp_length6 years	-6.6150e-02	3.7105e-02	-1.7828	0.074624
## emp_length7 years	-8.6093e-02	3.7083e-02	-2.3217	0.020252
## emp_length8 years	-6.1776e-02	3.8917e-02	-1.5874	0.112427
## emp_length9 years	-1.1751e-02	4.1346e-02	-0.2842	0.776241
## emp_lengthn/a	3.3410e-01	3.8757e-02	8.6203	< 2.2e-16
## home_ownershipYes	-1.5898e-01	1.6998e-02	-9.3525	< 2.2e-16

## annual_inc	-3.7361e-06	2.7012e-07	-13.8313	< 2.2e-16
## verification_statusYes	3.2287e-02	1.7328e-02	1.8633	0.062425
## issue_dApr-2014	-6.6991e-02	5.1677e-02	-1.2964	0.194853
## issue_dApr-2015	-8.0297e-01	7.3713e-02	-10.8932	< 2.2e-16
## issue_dAug-2012	-2.1776e-01	1.1033e-01	-1.9738	0.048409
## issue_dAug-2013	-1.0957e-01	5.2147e-02	-2.1011	0.035634
## issue_dAug-2014	-1.9275e-01	5.8772e-02	-3.2796	0.001040
## issue_dAug-2015	-2.7214e+00	2.0203e-01	-13.4701	< 2.2e-16
## issue_dDec-2012	-3.6811e-01	5.8280e-02	-6.3162	2.680e-10
## issue_dDec-2013	-1.6670e-01	5.1973e-02	-3.2074	0.001339
## issue_dDec-2014	-4.5705e-01	8.5860e-02	-5.3232	1.019e-07
## issue_dDec-2015	-1.4979e+01	4.2720e+01	-0.3506	0.725859
## issue_dFeb-2013	-2.9942e-02	5.6229e-02	-0.5325	0.594376
## issue_dFeb-2014	-1.6708e-02	5.2805e-02	-0.3164	0.751686
## issue_dFeb-2015	-4.4801e-01	7.1493e-02	-6.2665	3.693e-10
## issue_dJan-2013	-2.6994e-01	5.6467e-02	-4.7804	1.749e-06
## issue_dJan-2014	-1.1541e-01	5.2233e-02	-2.2095	0.027141
## issue_dJan-2015	-4.9212e-01	6.1404e-02	-8.0144	1.107e-15
## issue_dJul-2013	-2.1809e-02	5.2024e-02	-0.4192	0.675064
## issue_dJul-2014	-1.6796e-01	5.3646e-02	-3.1309	0.001743
## issue_dJul-2015	-1.9701e+00	1.1729e-01	-16.7967	< 2.2e-16
## issue_dJun-2013	-1.1194e-02	5.1794e-02	-0.2161	0.828897
## issue_dJun-2014	-1.0569e-01	5.8400e-02	-1.8097	0.070336
## issue_dJun-2015	-1.4316e+00	1.0693e-01	-13.3888	< 2.2e-16
## issue_dMar-2013	-9.4756e-02	5.5427e-02	-1.7096	0.087345
## issue_dMar-2014	-1.4315e-01	5.3109e-02	-2.6953	0.007032
## issue_dMar-2015	-5.8205e-01	7.4011e-02	-7.8644	3.710e-15
## issue_dMay-2013	-3.0680e-02	5.2129e-02	-0.5885	0.556170
## issue_dMay-2014	-1.0097e-01	5.4612e-02	-1.8488	0.064487
## issue_dMay-2015	-1.0157e+00	8.3002e-02	-12.2374	< 2.2e-16
## issue_dNov-2012	-3.5524e-01	5.6099e-02	-6.3324	2.414e-10
## issue_dNov-2013	-5.7449e-02	5.1387e-02	-1.1180	0.263581
## issue_dNov-2014	-3.5857e-01	6.2608e-02	-5.7272	1.021e-08
## issue_dNov-2015	-1.5983e+01	1.4257e+02	-0.1121	0.910739
## issue_dOct-2012	-3.6751e-01	5.6309e-02	-6.5266	6.727e-11
## issue_dOct-2013	-1.5766e-01	5.1531e-02	-3.0595	0.002217
## issue_dOct-2014	-2.4787e-01	5.3726e-02	-4.6137	3.956e-06
## issue_dOct-2015	-1.6134e+01	1.1012e+02	-0.1465	0.883516
## issue_dSep-2012	-2.9483e-01	5.6815e-02	-5.1893	2.111e-07
## issue_dSep-2013	-1.3191e-01	5.1989e-02	-2.5373	0.011170
## issue_dSep-2014	-1.3050e-01	6.9453e-02	-1.8789	0.060253
## issue_dSep-2015	-4.2815e+00	5.0419e-01	-8.4918	< 2.2e-16
## purpose1	-3.0690e-01	4.6560e-02	-6.5914	4.356e-11
## purpose2	-3.0669e-01	4.4990e-02	-6.8169	9.305e-12
## addr_stateAL	4.8315e-01	1.4992e-01	3.2228	0.001270
## addr_stateAR	3.6552e-01	1.5994e-01	2.2853	0.022294
## addr_stateAZ	2.9319e-01	1.4571e-01	2.0122	0.044197
## addr_stateCA	2.4481e-01	1.4008e-01	1.7477	0.080514
## addr_stateCO	-5.5343e-03	1.4779e-01	-0.0374	0.970129
## addr_stateCT	2.9202e-01	1.5121e-01	1.9313	0.053445
## addr_stateDC	-4.2096e-02	2.0557e-01	-0.2048	0.837746
## addr_stateDE	3.2388e-01	1.9242e-01	1.6832	0.092333
## addr_stateFL	4.2463e-01	1.4128e-01	3.0057	0.002650
## addr_stateGA	2.2357e-01	1.4458e-01	1.5463	0.122034

## addr_stateHI	2.2615e-01	1.6251e-01	1.3916	0.164048
## addr_stateIA	1.9183e+01	3.9562e+03	0.0048	0.996131
## addr_stateIL	2.9471e-01	1.4368e-01	2.0512	0.040250
## addr_stateIN	4.8501e-01	1.4854e-01	3.2652	0.001094
## addr_stateKS	2.4587e-01	1.5811e-01	1.5551	0.119931
## addr_stateKY	4.1819e-01	1.5488e-01	2.7000	0.006933
## addr_stateLA	4.6962e-01	1.5175e-01	3.0946	0.001971
## addr_stateMA	3.5503e-01	1.4655e-01	2.4226	0.015411
## addr_stateMD	3.6437e-01	1.4589e-01	2.4976	0.012504
## addr_stateME	-1.2045e+01	5.4760e+02	-0.0220	0.982451
## addr_stateMI	4.2683e-01	1.4516e-01	2.9404	0.003278
## addr_stateMN	3.6007e-01	1.4763e-01	2.4390	0.014727
## addr_stateMO	4.5356e-01	1.4856e-01	3.0529	0.002266
## addr_stateMS	4.8991e-01	1.9735e-01	2.4825	0.013048
## addr_stateMT	-5.9828e-02	1.9763e-01	-0.3027	0.762101
## addr_stateNC	2.9897e-01	1.4454e-01	2.0683	0.038609
## addr_stateND	-9.9662e+00	5.4602e+02	-0.0183	0.985438
## addr_stateNE	2.9037e-01	1.0723e+00	0.2708	0.786556
## addr_stateNH	-8.3350e-02	1.8054e-01	-0.4617	0.644325
## addr_stateNJ	4.4023e-01	1.4334e-01	3.0711	0.002132
## addr_stateNM	4.1836e-01	1.6442e-01	2.5445	0.010945
## addr_stateNV	4.2432e-01	1.4829e-01	2.8614	0.004218
## addr_stateNY	4.1691e-01	1.4101e-01	2.9567	0.003109
## addr_stateOH	3.8353e-01	1.4399e-01	2.6635	0.007732
## addr_stateOK	4.6847e-01	1.5546e-01	3.0134	0.002583
## addr_stateOR	1.7973e-01	1.5145e-01	1.1867	0.235340
## addr_statePA	3.9831e-01	1.4373e-01	2.7713	0.005583
## addr_stateRI	2.0185e-01	1.7736e-01	1.1381	0.255092
## addr_stateSC	1.3964e-01	1.5439e-01	0.9045	0.365749
## addr_stateSD	2.0464e-01	2.0253e-01	1.0104	0.312303
## addr_stateTN	5.8694e-01	1.4940e-01	3.9288	8.538e-05
## addr_stateTX	2.2229e-01	1.4132e-01	1.5729	0.115731
## addr_stateUT	2.6870e-01	1.5735e-01	1.7076	0.087707
## addr_stateVA	3.8930e-01	1.4405e-01	2.7025	0.006882
## addr_stateVT	2.6049e-01	2.2289e-01	1.1687	0.242531
## addr_stateWA	2.0140e-01	1.4635e-01	1.3762	0.168767
## addr_stateWI	2.7778e-01	1.5198e-01	1.8278	0.067587
## addr_stateWV	-8.8971e-03	1.7708e-01	-0.0502	0.959929
## addr_stateWY	6.9063e-02	2.0479e-01	0.3372	0.735937
## dti	2.6889e-02	1.0105e-03	26.6101	< 2.2e-16
## delinq_2yrsYes	-8.0443e-02	1.8024e-02	-4.4630	8.081e-06
## inq_last_6mths	4.7217e-02	6.5931e-03	7.1615	7.978e-13
## mths_since_last_delinq	-9.2563e-04	3.0905e-04	-2.9951	0.002744
## mths_since_last_record	-8.5255e-04	3.6119e-04	-2.3604	0.018258
## open_acc	1.9958e-02	1.9557e-03	10.2046	< 2.2e-16
## pub_rec	-9.0876e-03	2.2050e-02	-0.4121	0.680242
## revol_bal	-5.2268e-07	5.1610e-07	-1.0128	0.311173
## revol_util	2.6244e-01	3.4472e-02	7.6130	2.678e-14
## total_acc	-1.3475e-02	8.6821e-04	-15.5198	< 2.2e-16
## out_prncp	-1.0952e-04	3.0016e-06	-36.4864	< 2.2e-16
## application_typeJOINT	-2.8789e-01	4.8480e+02	-0.0006	0.999526
## tot_cur_bal	-5.7685e-07	7.5388e-08	-7.6517	1.983e-14
##				
## n = 175020 p = 156				

```
## Deviance = 138598.68996 Null Deviance = 157405.05401 (Difference = 18806.36405)
```

AIC

```
## [1] 138910.7
```

BIC

```
## [1] 140482
```

### 2.1.2. Modelo 1

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -3.2196e+00  5.8912e-02 -54.6507 < 2.2e-16
## loan_amnt      1.3340e-05  1.0104e-06  13.2022 < 2.2e-16
## home_ownershipYes -1.7182e-01  1.3927e-02 -12.3375 < 2.2e-16
## annual_inc    -5.5038e-06  2.5043e-07 -21.9772 < 2.2e-16
## verification_statusYes 6.8713e-02  1.6264e-02  4.2247 2.392e-05
## purpose1      -1.6690e-01  4.4808e-02 -3.7247 0.0001955
## purpose2      -1.8514e-01  4.3320e-02 -4.2737 1.922e-05
## dti           1.8221e-02  9.2907e-04  19.6116 < 2.2e-16
## delinq_2yrsYes -1.3449e-02  1.7345e-02 -0.7754 0.4381161
## int_rate       1.0694e+01  1.7213e-01  62.1241 < 2.2e-16
## inq_last_6mths  6.2650e-02  6.3329e-03  9.8929 < 2.2e-16
## mths_since_last_delinq -1.8607e-03  2.9932e-04 -6.2165 5.083e-10
## revol_bal      1.5591e-07  4.4808e-07  0.3480 0.7278670
## revol_util      3.4422e-01  3.1423e-02  10.9545 < 2.2e-16
## total_acc      -5.8826e-03  6.5956e-04 -8.9189 < 2.2e-16
##
## n = 175020 p = 15
## Deviance = 147454.03875 Null Deviance = 157405.05401 (Difference = 9951.01526)
```

AIC

```
## [1] 147484
```

BIC

```
## [1] 147635.1
```

### 2.1.3. Modelo 2

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.4982e+00  6.0946e-02 -40.9908 < 2.2e-16
## loan_amnt      1.9588e-05  1.0572e-06  18.5282 < 2.2e-16
## term 60 months  3.3061e-01  1.8442e-02  17.9271 < 2.2e-16
## gradeB         5.6633e-01  3.3249e-02  17.0328 < 2.2e-16
## gradeC         9.3380e-01  3.3393e-02  27.9638 < 2.2e-16
## gradeD         1.1900e+00  3.4995e-02  34.0040 < 2.2e-16
## gradeE         1.3705e+00  3.8707e-02  35.4080 < 2.2e-16
## gradeF         1.4999e+00  4.5672e-02  32.8399 < 2.2e-16
## gradeG         1.5554e+00  6.7977e-02  22.8808 < 2.2e-16
## home_ownershipYes -1.4809e-01  1.5895e-02 -9.3165 < 2.2e-16
## annual_inc    -4.7973e-06  2.6733e-07 -17.9457 < 2.2e-16
## purpose1      -3.3990e-01  4.5462e-02 -7.4766 7.625e-14
## purpose2      -3.4088e-01  4.3925e-02 -7.7606 8.451e-15
```

```
## dti                2.3033e-02  9.6138e-04  23.9585 < 2.2e-16
## delinq_2yrsYes     -9.8004e-02  1.7531e-02  -5.5902 2.268e-08
## inq_last_6mths     8.3436e-02  6.3998e-03  13.0373 < 2.2e-16
## open_acc           2.0488e-02  1.8825e-03  10.8831 < 2.2e-16
## revol_util         5.0347e-01  3.2274e-02  15.5998 < 2.2e-16
## total_acc          -1.3593e-02  8.3425e-04 -16.2935 < 2.2e-16
## out_prncp          -1.4874e-04  2.9890e-06 -49.7634 < 2.2e-16
## tot_cur_bal        -7.1269e-07  7.2664e-08  -9.8080 < 2.2e-16
##
## n = 175020 p = 21
## Deviance = 142899.18953 Null Deviance = 157405.05401 (Difference = 14505.86448)
```

## AIC

```
## [1] 142941.2
```

## BIC

```
## [1] 143152.7
```

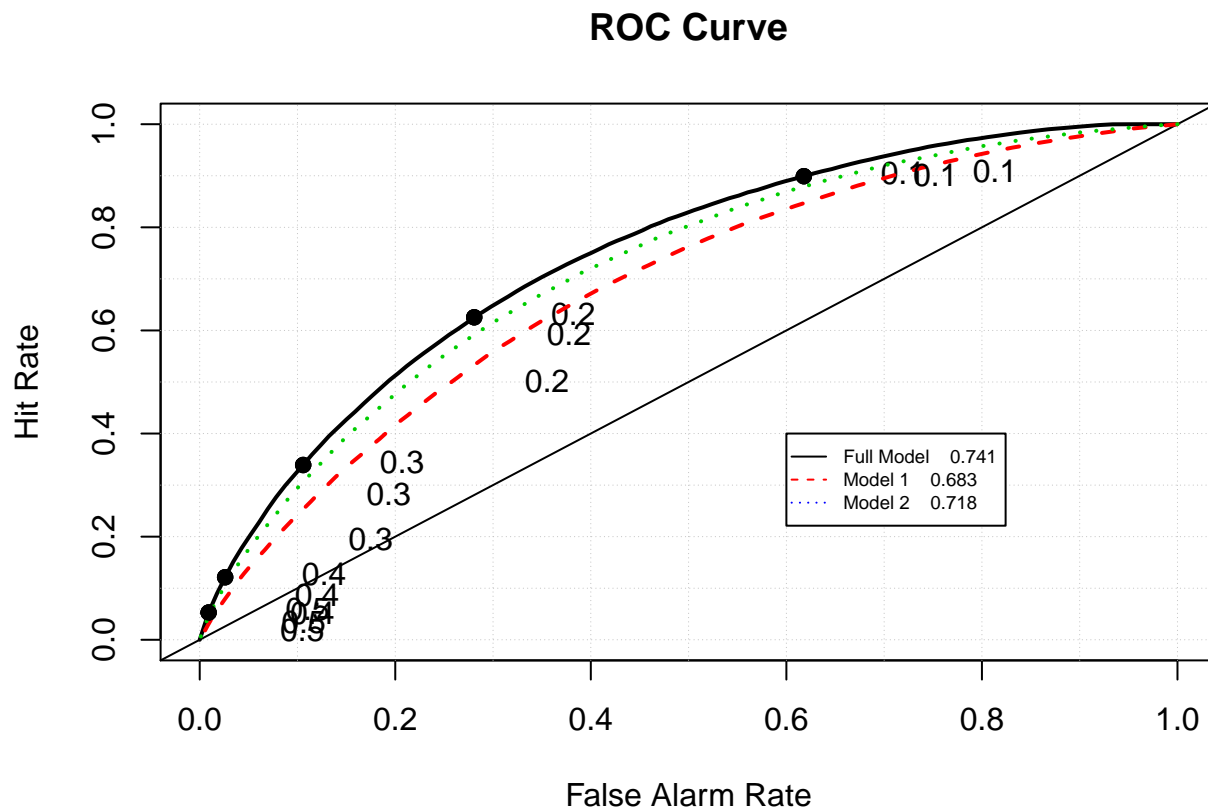
## Conclusiones

El modelo con todas las variables (modelo 0) parece ser el mejor según los ratios BIC y AIC.

## 2.2. Pruebas Insample y Curva ROC

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in roc.plot.default(x = loan.train$loan_status == 1, pred =
## cbind(prob.glm0.insample, : More than 10,000 unique predictions are used as
## thresholds. 100 equally spaced thresholds are used
```



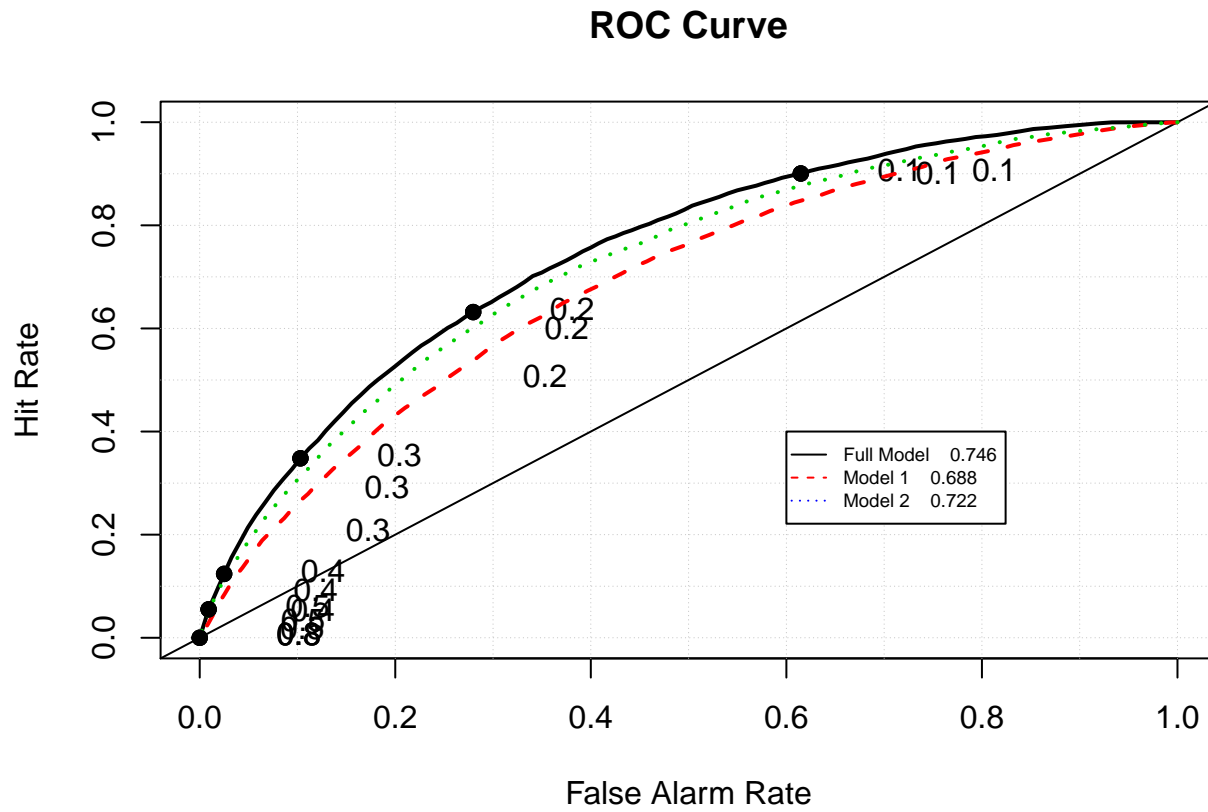
```
##      Model      Area p.value binorm.area
## 1 Model 1 0.7408225      0      NA
## 2 Model 2 0.6827418      0      NA
## 3 Model 3 0.7175603      0      NA
```

Según la curva ROC, el modelo con todas las variables es el que mejor explica la información.

### 2.3. Pruebas Outsample y Curva ROC

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in roc.plot.default(x = loan.test$loan_status == 1, pred =
## cbind(prob.glm0.outsample, : More than 10,000 unique predictions are used
## as thresholds. 100 equally spaced thresholds are used
```



```
##      Model      Area p.value binorm.area
## 1 Model 1 0.7462003      0      NA
## 2 Model 2 0.6879516      0      NA
## 3 Model 3 0.7222007      0      NA
```

El modelo con todas las variables continua siendo el mejor en la explicación de la variable default en las pruebas del dataset test.

## 2.4. Cross Validation

La tecnica de cross validation entrega más detalle sobre el error medio de los modelos logísticos:

**Modelo 0:**

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading
```

```

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

```



```

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

```

```

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## [1] 0.1544534

```

[illegible]

```
## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## [1] 0.1473703
```

#### Modelo 1:

```
## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor
```

```
## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in y - yhat: longitud de objeto mayor no es múltiplo de la longitud
## de uno menor

## [1] 0.1508977
```

El modelo 0 es el que tiene menor error medio.

### 3. Estimación del Cut Off y Matriz de Confusión

Con un cutoff de 0.4663192, se tiene una menor cantidad de casos.

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading

## [1] 0.4663192
```

### Bibliografía

- Machine learning logistic regression for credit modelling in R, Datascience+, en sitio web: <https://datascienceplus.com/machine-learning-logistic-regression-for-credit-modelling-in-r/>
- Logistic Regression, r-statistics.co, en sitio web: <http://r-statistics.co/Logistic-Regression-With-R.html>
- Devolución de préstamos, RPUBS, en sitio web: <https://rpubs.com/perevales/83897>