

1. PROBLEMA:

Las empresas que se dedican a otorgar préstamos siempre tienen presente el riesgo de crédito, el cual se define como el potencial incumplimiento generado por la imposibilidad de un cliente para cumplir sus compromisos.

La principal pérdida que tienen estas empresas prestamistas es la pérdida crediticia, la cual es la cantidad de dinero que pierde el prestamista cuando el prestatario incumple con el pago. Este tipo de prestatarios son un riesgo para estas empresas, ya que causan la mayor cantidad de pérdidas para las empresas prestamistas, por lo que para evitar este riesgo, es indispensable identificar y predecir estos posibles prestamistas que pueden caer en mora.

2. OBJETIVO:

El objetivo de esa práctica será generar una tabla estructurada y limpia para que en un futuro se pueda predecir cuando un cliente caerá en mora a partir de sus características y comportamiento pasado.

Como estamos interesados en predecir si el cliente pagará el préstamo en el futuro (fully paid) o se le cancelaría la cuenta por no pagar (charged off), se debe restringir la información solo a esos estatus en la variable loan status (variable objetivo)

3. INFORMACIÓN:

Contamos con información de la empresa Lending Club, la cual es una empresa estadounidense de préstamos, con sede en San Francisco, California. En esta empresa los prestatarios pueden acceder fácilmente a préstamos con tasas de interés bajas.

El dataset contiene datos de todos los préstamos emitidos durante el período 2007-2015, donde contiene el estado actual del préstamo (actual, cancelado, completamente pagado) y la información de pago más reciente. Las características adicionales incluyen puntajes de crédito, número de consultas financieras y cobros, entre otros.

****El diccionario de datos se anexa al final de este archivo****

4. CAMPOS A CUBRIR

- Calidad de los datos - Revisión a detalle de la información para descartar casos particulares o erróneos, la tabla después de calidad de datos debe ser precisa.
- Normalización de variables, en caso de ser necesario
- Análisis exploratorio - Dentro del PDF agregar aquellas visualizaciones o análisis que muestren un comportamiento particular o de importancia para el objetivo, las visualizaciones deben estar acompañadas de una breve descripción.
- Outliers - Agregar visualizaciones dentro del PDF pre eliminación de valores atípicos y post eliminación de valores atípicos, además de agregar que porcentaje se ha eliminado y que métodos se utilizaron.
- Missings - Describir los métodos utilizados y el valor a imputar de cada variable.
- Ingeniería de variables - Las columnas de la tabla final deben ser de tipo numérico, donde las variables categóricas, texto, entre otras se han transformado para utilizarlas dentro de un modelo. La sofisticación de la ingeniería será lo que se premiará,

es decir, la creación de nuevas variables deben reflejar aquellos datos que pueden influenciar al solicitar un préstamo.

5. ENTREGABLES

- PDF con resultados, se premiará las descripciones claras y las visualizaciones así como las secciones trabajadas. (puede ser el mismo colab en formato de pdf con su explicación en celdas de texto)
- Código en python (Notebook) , limpio, ordenado , comentado y bien estructurado, sin errores en el código. En el notebook deben mostrarse todas las gráficas que contiene el PDF.
- Recursos adicionales en caso de existir.

Diccionario

1. **acc now delinq**: El número de cuentas en las que el prestatario está ahora en mora.
2. **application_type**: El tipo de aplicación que se realizó, si fue individual o en conjunto
3. **addr state**: El estado proporcionado por el prestatario en la solicitud de préstamo.
4. **all util**: Límite de saldo a crédito en todas las operaciones.
5. **annual inc**: Los ingresos anuales autoinformados proporcionados por el prestatario durante el registro.
6. **annual inc joint**: Los ingresos anuales autoinformados combinados proporcionados por los co-prestatarios durante el registro.
7. **collection recovery fee**: cargo posterior a la tarifa de cobro.
8. **collections 12 mths ex med**: Número de colecciones en 12 meses, excluidas las colecciones médicas.
9. **delinq 2yrs**: El número de más de 30 días de incidencias vencidas de morosidad en el archivo de crédito del prestatario durante los últimos 2 años.
10. **desc**: Descripción del préstamo proporcionada por el prestatario.
11. **dti**: Una relación calculada utilizando los pagos mensuales totales de la deuda del prestatario sobre las obligaciones totales de la deuda, excluyendo la hipoteca y el préstamo LC solicitado, dividido por los ingresos mensuales autoinformados del prestatario
12. **dti joint**: Una relación calculada utilizando los pagos mensuales totales de los co-prestatarios sobre las obligaciones totales de deuda, excluyendo las hipotecas y el préstamo LC solicitado, dividido por los ingresos mensuales autoinformados
13. **earliest cr line**: El mes en que se abrió la línea de crédito reportada más temprana del prestatario.
14. **emp length**: Duración del empleo en años. Los valores posibles están entre 0 y 10, donde 0 significa menos de un año y 10 significa diez o más años.
15. **emp title**: El título del trabajo proporcionado por el Prestatario al solicitar el préstamo.*
16. **funded amnt**: El importe total comprometido para ese préstamo en ese momento.
17. **funded amnt inv**: La cantidad total comprometida por los inversores para ese préstamo en ese momento.
18. **grade**: LC asignó grado de préstamo.
19. **home ownership**: El estado de propiedad de la vivienda proporcionado por el prestatario durante el registro. Nuestros valores son: ALQUILER, PROPIO, HIPOTECA, OTRO.
20. **id**: Una LC única asignada ID para el listado de préstamos.
21. **il util**: Relación entre el saldo corriente total y el límite de crédito/crédito alto en todos los valores de la instalación.
22. **initial list status**: El estado de listado inicial del préstamo. Los valores posibles son – W, F.
23. **inq fi**: Número de consultas de finanzas personales.
24. **inq last 12m**: Número de consultas de crédito en los últimos 12 meses.
25. **inq last 6mths**: El número de consultas en los últimos 6 meses (excluyendo las consultas de automóviles e hipotecas).
26. **installment**: El pago mensual adeudado por el prestatario si el préstamo se origina.
27. **int rate**: Tasa de interés del préstamo.

28. **issue d**: El mes en que se financió el préstamo
29. **last credit pull d**: El mes más reciente LC sacó crédito para este préstamo.
30. **last pymnt amnt**: Último importe total de pago recibido.
31. **last pymnt d**: El mes pasado se recibió el pago.
32. **loan amnt**: El monto indicado del préstamo solicitado por el prestatario. Si en algún momento, el departamento de crédito reduce el monto del préstamo, entonces se reflejará en este valor.
33. **loan status**: Estado actual del préstamo.
34. **max bal bc**: Saldo corriente máximo adeudado en todas las cuentas rotativas.
35. **member id**: Una LC única asignada id para el miembro prestatario.
36. **mths since last delinq**: El número de meses desde la última morosidad del prestatario.
37. **mths since last major derog**: Meses desde la calificación más reciente de 90 días o peor.
38. **mths since last record**: El número de meses transcurridos desde el último registro público.
39. **mths since rcnt il**: Meses desde que se abrieron las cuentas a plazos más recientes.
40. **next pymnt d**: Próxima fecha de pago programada.
41. **open acc**: El número de líneas de crédito abiertas en el archivo de crédito del prestatario.
42. **open acc 6m**: Número de operaciones abiertas en los últimos 6 meses
43. **open il 12m**: Número de cuentas a plazos abiertas en los últimos 12 meses.
44. **open il 24m**: Número de cuentas a plazos abiertas en los últimos 24 meses.
45. **open il 6m**: Número de operaciones a plazos actualmente activas.
46. **open rv 12m**: Número de operaciones rotativas abiertas en los últimos 12 meses.
47. **open rv 24m**: Número de operaciones rotativas abiertas en los últimos 24 meses.
48. **out prncp**: Capital pendiente restante para el monto total financiado.
49. **out prncp inv**: Capital pendiente restante por parte del importe total financiado por los inversores.
50. **policy code**: disponible públicamente policy code=1 nuevos productos no disponibles públicamente policy code=2.
51. **pub rec**: Número de registros públicos despectivos.
52. **purpose**: Una categoría proporcionada por el prestatario para la solicitud de préstamo.
53. **pymnt plan**: Indica si se ha establecido un plan de pago para el préstamo.
54. **recoveries**: cargo posterior a la recuperación bruta.
55. **revol bal**: Saldo rotatorio de crédito total.
56. **revol util**: Tasa de utilización de la línea rotativa, o la cantidad de crédito que el prestatario está utilizando en relación con todo el crédito rotatorio disponible.
57. **sub grade**: Subsuelo de préstamo asignado LC.
58. **term**: El número de pagos del préstamo. Los valores están en meses y pueden ser 36 o 60.
59. **title**: El título del préstamo proporcionado por el prestatario.
60. **total acc** : Número de cuentas totales
61. **tot coll amt**: Montos totales de cobro adeudados.
62. **tot cur bal**: Saldo corriente total de todas las cuentas.
63. **total bal il**: Saldo actual total de todas las cuentas a plazos.
64. **total cu tl**: Número de operaciones financieras.
65. **total pymnt**: Pagos recibidos hasta la fecha por el importe total financiado.

- 66. **total pymnt inv:** Pagos recibidos hasta la fecha por una parte del importe total financiado por los inversores.
- 67. **total rec int:** Intereses recibidos hasta la fecha.
- 68. **total_rec_late_fee:** Monto total recibido hasta la fecha por cargos por mora
- 69. **total rec prncp:** Principal recibido hasta la fecha.
- 70. **total rev hi lim:** Límite de crédito/crédito revolving total alto.
- 71. **url:** Dirección URL de la página lc con datos de listado.
- 72. **verification status:** Indica si lc verificó los ingresos, no los verificó o si se verificó la fuente de ingresos.
- 73. **verified status joint:** Indica si los ingresos conjuntos de los co-prestatariosMfueron verificados por LC, no verificados, o si la fuente de ingresos fue verificada.
- 74. **zip code:** Los primeros 3 números del código postal proporcionado por el prestatario en la solicitud de préstamo.