

Big data

CONCESIÓN DE CRÉDITOS

10 DICIEMBRE 2021

AUTORES

Iñigo de Dios Gómez & Oier Mentxaka Gomez

Limpieza de datos

1. Borrado de ID y de Flags innecesarios excepto de los id de solicitud para saber cual de los créditos es concedido o no.
2. Borrado de columnas innecesarias o con información poco relevante.
3. Conversión de datos numéricos que estaban declarados como *factor* a *enteros*.
4. Suma de columnas para simplificar los datos.
5. En caso de haber columnas con *NA* y poder, sustituirlas con 0.
6. Exportación del dataset resultante para la próxima iteración del proceso.

Toma de decisiones

1. Gráfico de correlación para ver la relación entre las distintas variables y considerar o no su estudio.

Conclusión:

Hubo correlación directa entre ciertos parámetros.

Cantidad miembros familia	87%	Cantidad de hijos
Gastos fijos anuales	78%	Credito solicitado
Teléfono de empresa	62%	Edad

Estos parámetros dan conclusiones bastante obvias por lo que decidimos

ponernos a filtrar las peticiones denegándolas o aceptándolas.

2. Iteraciones

1. El caso en el que los gastos del usuario no superen el 30% de los ingresos anuales.
 - a. Si se superan los gastos con el crédito superan el 30% de los ingresos anuales -> se estudia.
 - b. Si se superan los gastos con el crédito NO superan el 30% de los ingresos anuales -> se acepta.
2. El caso en el que la cantidad solicitada sea inferior al 80% del valor de tasación del inmueble a hipotecar.
 - a. Si el valor de tasación es inferior al 80% del valor de tasación -> se acepta.
 - b. Si el valor de tasación NO es inferior al 80% del valor de tasación -> se estudia.
3. El caso en el que se tengan propiedades para poder avalar el crédito en caso de impago.
 - a. Si NO se tiene un inmueble con el que poder avalar -> se estudia.
 - b. Si se tiene un inmueble con el que poder avalar -> se acepta.
4. El caso en el que se tengan hijos o no a su cargo.
 - a. Si se tienen hijos -> se estudia.
 - b. Si NO se tienen hijos -> se acepta.
5. Se estudia la cantidad de trabajadores en el domicilio.
 - a. Si la cantidad de trabajadores en la familia es 4 o más -> se concede.
 - b. Si la cantidad de trabajadores es igual a 2 o a 3 -> se estudia.
 - c. Si la cantidad de trabajadores es uno -> se deniega.

6. Si se dispone de un coche para poder avalar el crédito en caso de impago.
 - a. Si no se tiene un coche -> se deniega.
 - b. Si se tiene un coche con más de 10 años -> se deniega.
 - c. Si se tiene un coche de menos de 5 años -> se acepta.
 - d. Si se tiene un coche entre 5 y 10 años -> se estudia.
7. El caso en el que el solicitante sea funcionario
 - a. Si se es funcionario -> se acepta.
 - b. Si no se es funcionario -> se estudia.
8. El caso en el que haya habido algún tipo de morosidad anteriormente.
 - a. Si se ha sido moroso -> se deniega.
 - b. Si NO se ha sido moroso -> se acepta.
3. Se hacen gráficas en Shiny para ver las distintas iteraciones y sacar conclusiones de los clientes del banco.

Visualizaciones en Shiny

1. Tipos de trabajo con cantidad solicitada.

En este gráfico se ven distintos boxplots donde se clasifican a partir de los tipos de trabajo la cantidad solicitada en cada crédito.

	Min	1st Qu	Median	Mean	3rd Qu	Max
Commercial associate	45000	315000	592560	680152	912240	2695500
Pensioner	47970	247500	450000	540399	755190	2250000
State servant	45000	315000	593327	706478	966600	2517300
Unemployed	270000	270000	270000	270000	270000	270000

Working	45000	270000	499736	592307	808650	2517300
---------	-------	--------	--------	--------	--------	---------

- a. Lo curioso después de ver esta tabla es que un desempleado ha pedido un crédito de 270000 euros.
- b. Todos menos los pensionistas han pedido como mínimo 45000 euros, en cambio los pensionistas han pedido 3000 euros más.
- c. En cambio, las medias rondan la general de 52000 euros, al igual que las máximas de 2.5M de euros.

2. Tipos cantidad de créditos aceptados

En estos diagramas podemos ver como la cantidad solicitada de los créditos es bastante inferior a la media calculada en el gráfico de datos. Vemos como aproximadamente la cantidad solicitada es un 20% inferior a lo que anteriormente se muestra.

3. Tipos cantidad de créditos denegados

En estos diagramas podemos ver como la cantidad solicitada de créditos es de un 30% más que en el gráfico general por lo que podemos intuir que los gastos de los solicitantes son superiores o que el valor solicitado es bastante superior a la tasación de la vivienda. Decidimos hacer un gráfico de barras con los aceptados y denegados por iteración para encontrar justificación a esta variación.

4. Aceptados y denegados por iteración

Podemos apreciar que la mayoría de los créditos se aceptan en la primera iteración, es decir 3600 solicitudes aproximadamente son aceptadas por ser la cantidad solicitada dividida en las cotas y los gastos fijos anuales no superan el 40% de los ingresos. Podemos ver como realmente nuestro cliente podría encajar en el perfil de un cliente “No soñador”, es decir son bastante realistas y saben que es realmente lo que pueden permitirse o no.

Además podemos apreciar que en la segunda iteración casi mil de las solicitudes ha sido denegada, es decir, estos clientes si han sido “soñadores”, ya que sin poder hacerle frente a los gastos han solicitado un crédito bastante superior a los que realmente vale la vivienda (en su 80%).

Por si fuese poco, también es relevante el hecho de tener o no una vivienda con la que poder avalar el crédito hipotecario en caso de impago, y se ve cómo 1000 de las solicitudes son aceptadas y 50 directamente rechazadas.

- a. Iteración 1: 3600 aceptados
- b. Iteración 2: 0 aceptados
- c. Iteración 3: 1050 aceptados
- d. Iteración 4: 0 aceptados
- e. Iteración 5: 28 aceptados
- f. Iteración 6: 5 aceptados
- g. Iteración 7: 6 aceptados
- h. Iteración 1: 550 aceptados
- i. Iteración 2: 1100 aceptados
- j. Iteración 3: 50 aceptados
- k. Iteración 4: 200 aceptados
- l. Iteración 5: 60 aceptados
- m. Iteración 6: 30 aceptados

5. Aceptados y denegados por iteración

El 10% de los usuarios que tenemos en nuestro dataset tienen propiedades y coche con las que avalar el crédito. Un 15% no tienen ni coches ni propiedades con los que avalar el crédito. El 20% tiene coche y el 65% tiene propiedades, por lo que en vista de estos datos, el criterio con el que avalar con este tipo de bienes es correcto.

6. Edad cantidad, estatus crédito y días trabajados

Respecto a la edad y cantidad solicitada, el estado civil de la persona y la cantidad solicitada, los días trabajados y la cantidad solicitada, y los días trabajados con los ingresos anuales, podemos ver como todos los clientes están en una media razonable.

7. Rendimiento

Respecto al rendimiento, podemos observar como el banco está realizando de forma correcta su trabajo no dejando de lado algún tipo de cliente o parecido. Respecto a los datos, este estudio se ha realizado a partir de los primeros 5000 solicitantes aproximadamente y el dataset contenía una cantidad de

usuarios de 50000, de manera que aunque estos gráficos den unos datos, siempre serán estimados.

La persona 104600, no ha tenido ningún tipo de deuda. Tiene unos ingresos anuales de 135000€ y solicita un crédito de 200000€ para una vivienda valorada en 270000€. Viendo que el cliente solicita una hipoteca que es el 100% del valor del inmueble, y viendo que los años a pagar son 25, quedarían como hipoteca 10800€ anuales. Teniendo en cuenta que está desempleada y que tiene un rendimiento anual de 135000€, podríamos considerarla una persona solvente. Sin embargo, debido a que es desempleado, vemos que a pesar de que cumple los requisitos, no concuerda y por lo tanto no se le otorga. Por lo tanto, es un cliente fraudulento.