

María Ana Ortiz
Alvaro Pabón
Jorge López

CLASIFICACIÓN DE CLIENTES: CASO DE UNA FINANCIERA DE CRÉDITOS DE LIBRANZAS

Resumen.

El objetivo de este caso de estudio es generar una segmentación de los clientes actuales de una compañía financiera que ofrece créditos de libranzas.

A pesar de que las libranzas son un instrumento de crédito relativamente estandarizado y bien reconocido en la industria financiera, el perfil de las personas que los solicitan es, por decirlo menos, poco homogéneo. Derivado de la necesidad de poder ofrecer productos de crédito que se adapten cada vez más a las características y necesidades de los clientes, resulta indispensable segmentar eficientemente a clientes potenciales.

Introducción.

¿Cómo puede una compañía de financiamiento tener una mayor penetración de mercado haciendo uso de la analítica de datos?

Con la crisis económica producto de la pandemia por Covid-19, la entidad de financiamiento “Tu Libranza” ha tenido complicaciones para regresar al volumen previo de colocación de créditos. Visto que el sector financiero ha tenido una importante reactivación económica, la Junta Directiva de la entidad se pregunta si sus productos estandarizados se ajustan realmente a las características de su población objetivo.

“Tu Libranza” ha determinado que a través de una segmentación clara de sus clientes potenciales podría generar una oferta de productos más atractiva que conduciría, finalmente, a lograr una mayor penetración de mercado.

El problema que se pretende resolver corresponde a una clasificación de clientes que utilizará herramientas y algoritmos de aprendizaje no supervisado.

Revisión preliminar de la literatura.

No se encontró alguna referencia pública que incluyera variables crediticias.

Descripción de los datos.

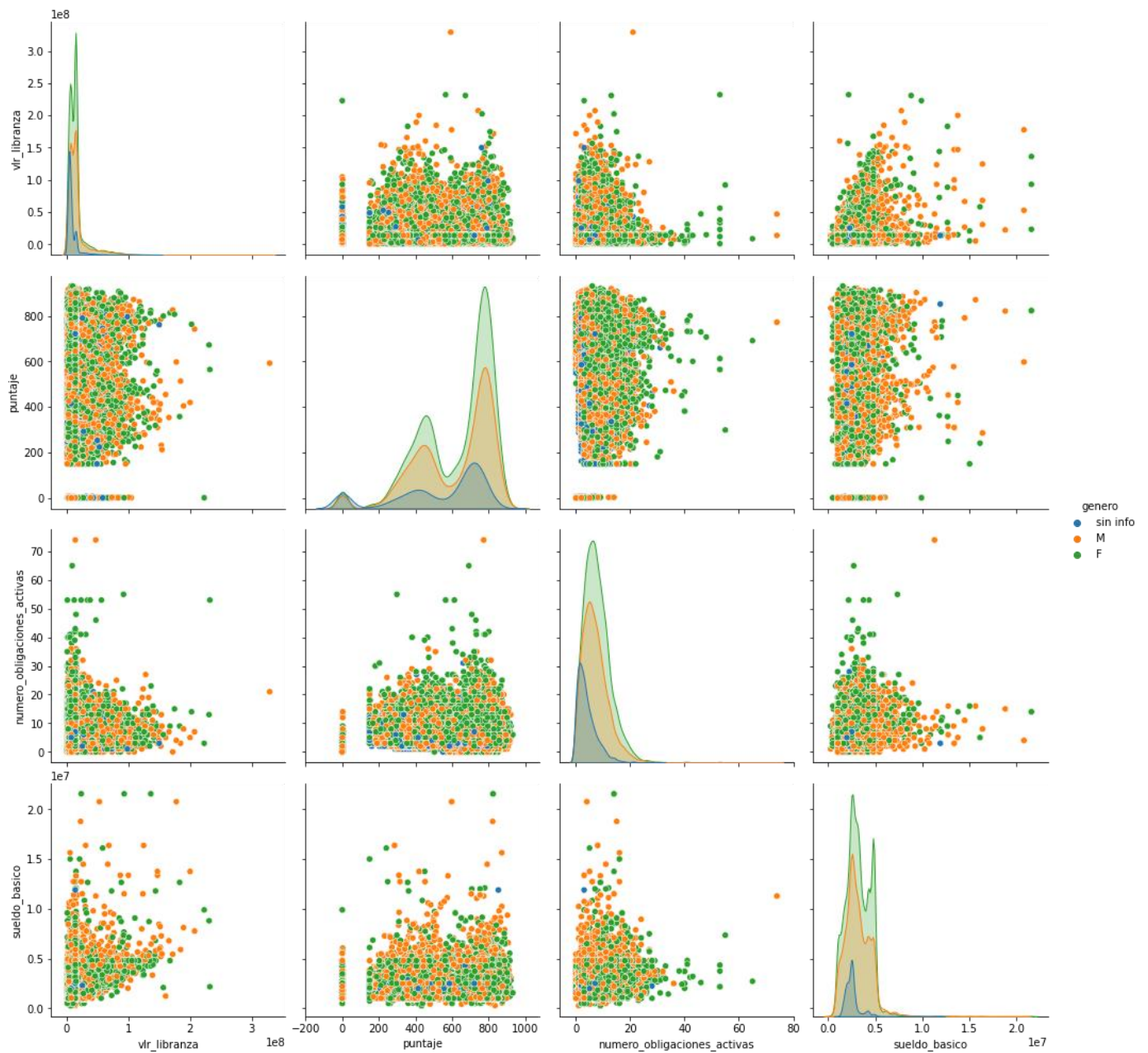
La información con la cual se dispone para el estudio corresponde a la base de clientes actuales de la financiera. En total son 30.776 registros de créditos de libranza con 37 variables cada uno, de cuales, en su mayoría corresponden a variables de información de crédito.

Las variables, al ser la mayoría tan específica, se vuelven complejas de describir una a una. Por esta razón se decidió describir únicamente las variables cuya interpretación es más sencilla de entender:

- Valor de la libranza
- Puntaje de crédito
- Número de obligaciones
- Género
- Rango de edad
- Sueldo básico

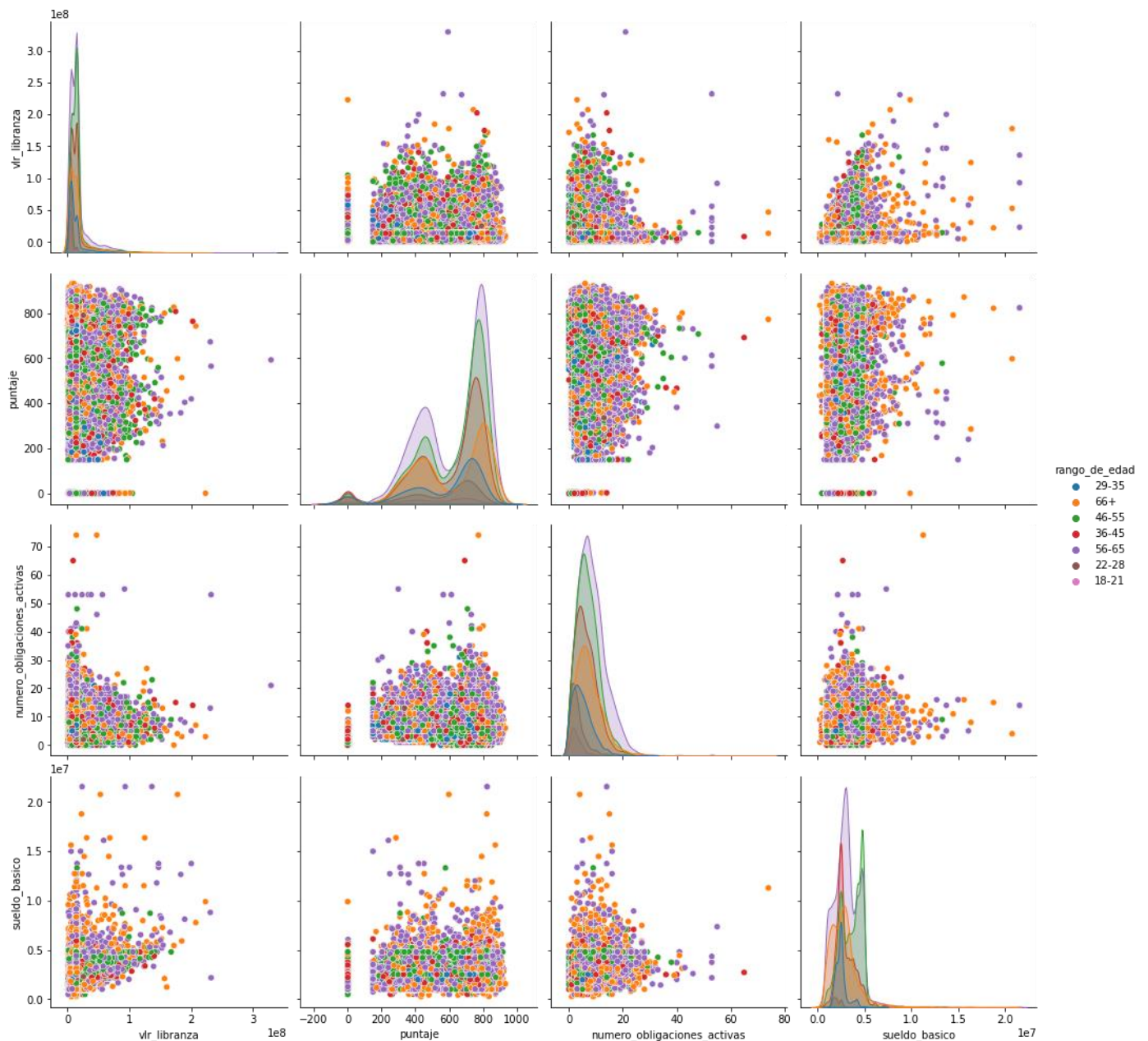
Del total de 30,776 registros, el 51% son mujeres, el 35% son hombres y el 14% carece de información de género.

A continuación, se presentan gráficos de dispersión sobre las variables mencionadas, de las cuales se puede observar que, independientemente del género, las variables se distribuyen de una forma muy similar.



Adicionalmente, se realizó un análisis por 7 grupos etarios, para los cuales la proporción de registros es la siguiente:

18-21	0.015353
22-28	0.044223
29-35	0.073617
36-45	0.173611
46-55	0.239289
56-65	0.313792
66+	0.140115



Más allá de la diferencia por el número de registros en cada grupo etario, se puede observar que el grupo con un mayor sueldo básico es aquel de personas entre 46 y 55 años.

37 Variables, a continuación mostramos la descripción de cada uno:

ID: Cédula anonimizada

vlr_cuota: Valor cuota del crédito

vlr_libranza: Valor presente del crédito

Puntaje: Score de bureau de crédito

numero_obligaciones_activas: Cantidad de obligaciones financieras son saldo mayor a cero

genero: Femenino o Masculino

rango_de_edad: Edad de la persona en rangos

cartera_bancaria: pendiente por identificar

cartera_bancaria_numero_creditos: Cantidad total de créditos en el sector financiero.

cartera_bancaria_valor_inicial: Valor total inicial de créditos en el sector financiero.

cartera_bancaria_valor_saldo: Valor saldo total de créditos en el sector financiero.

cartera_bancaria_valor_cuotas: Valor total cuotas de créditos activos en el sector financiero.

cartera_bancaria_valor_mora: Valor mora total de créditos activos en el sector financiero.

informacion_tdc_numero_tdc: Cantidad total de tarjeta de créditos activas.

informacion_tdc_valor__cupos: Cupo total de tarjeta de créditos activas.

informacion_tdc_valor__utilizado: Saldo utilizado en la tarjeta de crédito.

informacion_tdc_valor__cuotas: Cuota total a pagar de las tarjetas de crédito activas.

informacion_tdc_valor_mora: Valor total en mora de las tarjetas de crédito.

sector__real_numero_creditos: Cantidad de créditos activos en el sector real.

sector__real_valor_inicial: Valor total inicial de créditos activos en el sector cooperativo.

sector__real_valor_saldo: Valor saldo total de créditos activos en el sector real.

sector__real_valor_cuotas: Valor total cuotas de créditos activos en el sector real.

sector__real_valor_mora: Valor mora total de créditos activos en el sector real.

sector__cooperativo_numero_creditos: Cantidad de créditos activos en el sector cooperativo.

sector__cooperativo_valor_inicial: Valor total inicial de créditos activos en el sector cooperativo.

sector__cooperativo_valor_saldo: Valor saldo total de créditos activos en el sector cooperativo.

sector__cooperativo_valor_cuotas: Valor total cuotas de créditos activos en el sector cooperativo.

sector__cooperativo_valor_mora: Valor mora total de créditos activos en el sector cooperativo.

obligaciones_mora_30: Número de obligaciones en mora de 30

obligaciones_mora_60 Número de obligaciones en mora de 60

obligaciones_mora_90 Número de obligaciones en mora de 90

obligaciones_mora_120 Número de obligaciones en mora de 120

ultimo_ano_moras_30: Cantidad de productos en mora 30 en el último año.

ultimo_ano_moras_60 Cantidad de productos en mora 60 en el último año.

ultimo_ano_moras_90 Cantidad de productos en mora 90 en el último año.

ultimo_ano_moras_120 Cantidad de productos en mora 120 en el último año.

sueldo_basico: Salario de la persona

- **Propuesta metodológica.** Se define un algoritmo/técnica que se planea utilizar y se plantea qué otros algoritmos/técnicas son candidatos a utilizarse para solucionar la pregunta de interés. Se argumenta por qué el elegido es el adecuado dado el contexto del problema y de la organización. [15 puntos]

1. Revisión viabilidad recodificación de variables categóricas.
2. Estandarización de variables numéricas.
3. Reducción de variables a través de un análisis por componentes principales.
4. Se espera compararán los resultados implementado 2 métodos de clasificación:
 - a. K medias / k medoides
 - b. Clasificación jerárquica

- **Bibliografía:** Se citan los artículos mencionados en el texto, usando de forma correcta y consistente el estilo de referencia que se hay escogido usar (Chicago, APA, MLA, etc.). [5 puntos]

Depende de la revisión preliminar de literatura.