

CASO LENDING CLUB

Integrantes:

Frank Hernandez

Sara Peña

El propósito del siguiente informe es presentar el análisis realizado al caso de estudio de Lending Club.

1. Describa el modelo de negocio de las plataformas de créditos en línea.

Para una empresa como Lending Club, encontramos que está ubicada dentro del sector de las Fintech, dado que aprovecha la tecnología como mecanismo o herramienta de préstamo de sus servicios financieros. Dada su tecnología y sus aplicativos, es que elimina la intermediación de los bancos, los gastos que podrían generar, y logran conectar a prestamistas y prestatarios que requieren dinero con distintos propósitos.

a. ¿Cuáles son los interesados (stakeholders) y sus roles?

Su audiencia objetivo se divide en 2 partes. En primer lugar están los prestatarios, que son las personas que quieren pedir el préstamo de una cantidad de dinero por algún objetivo en específico. Por otro lado están los prestamistas, que son las personas que tienen capacidad económica de prestar parte de su dinero y a cambio obtener el retorno de su dinero con un rendimiento adicional dado ese préstamo, la temporalidad, entre otros factores. A la final, el problema que resuelven es que hay personas que no saben qué hacer con su dinero y por ende se les brinda la oportunidad de generar rendimientos a partir del préstamo del mismo, y por el otro lado ayudan al préstamo de dinero a personas que por alguna u otra circunstancia no pudieron tener un préstamo con condiciones que fueran favorables para ellos a través de un banco.

b. ¿Cuál es la ventaja de Lending Club sobre los demás competidores?

Su factor diferencial es la tecnología que usan para prestar el dinero de los prestamistas a prestatarios, de forma que el dinero prestado sea retornado según la persona asignada y la fecha de pago. Su modelo permite usar varias variables para detectar si una persona es apta o no para que se le preste el dinero, y así se aseguran de que el dinero prestado sea devuelto satisfactoriamente. No por nada tienen una tasa de rechazo del 90% de las solicitudes que se realizan en la plataforma para solicitar un préstamo.

c. ¿Cómo es la interacción de los inversores en la plataforma?

Los inversores deben indicar la cantidad que desean prestar, las características del solicitante, la tasa de interés esperado, el propósito para el que será usado ese dinero, entre otros factores. Adicionalmente, en dado caso que sientan que su dinero puede estar en una situación de riesgo, tienen la opción de poner sus inversiones en venta antes de que el préstamo sea totalmente financiado, y de esa forma reducir el riesgo de que pierdan su dinero.

d. ¿Cómo la plataforma crea rendimientos?

La plataforma genera rendimientos a partir de cobrarle a los solicitantes una cuota inicial para acceder al préstamo (depende del grado de crédito y abarca del 1,1 al 5 por ciento de la cantidad prestada) y a los inversores unos gastos de administración (1% de todas las cantidades pagadas por el solicitante). Adicionalmente, generar retornos entre el 6 y el 9 por ciento para los prestamistas, de manera que sí generan confianza e interés en la plataforma.

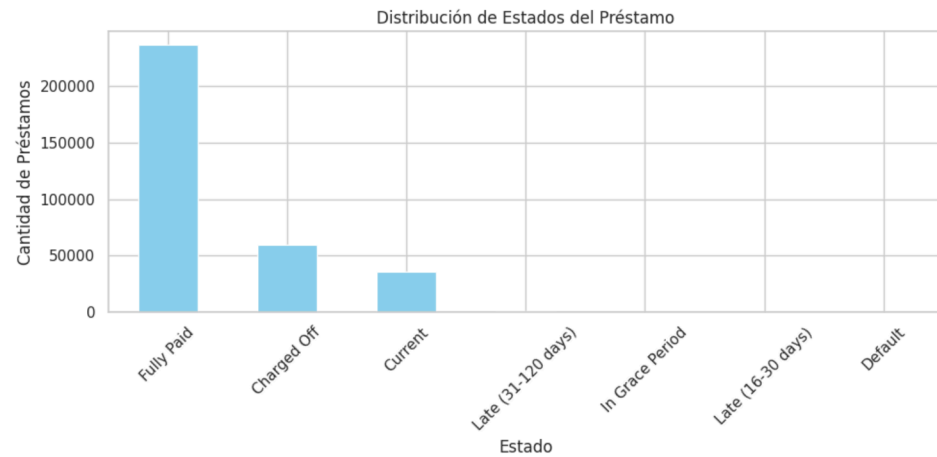
2. Explore los datos de los créditos.

- a. Cuál es la proporción de pagos completos vs impagos.

```

Distribución del estado del préstamo (loan_status):
loan_status
Fully Paid      70.805816
Charged Off     17.891731
Current         10.726529
Late (31-120 days)  0.349091
In Grace Period  0.155518
Late (16-30 days) 0.071017
Default         0.000300
Name: proportion, dtype: float64

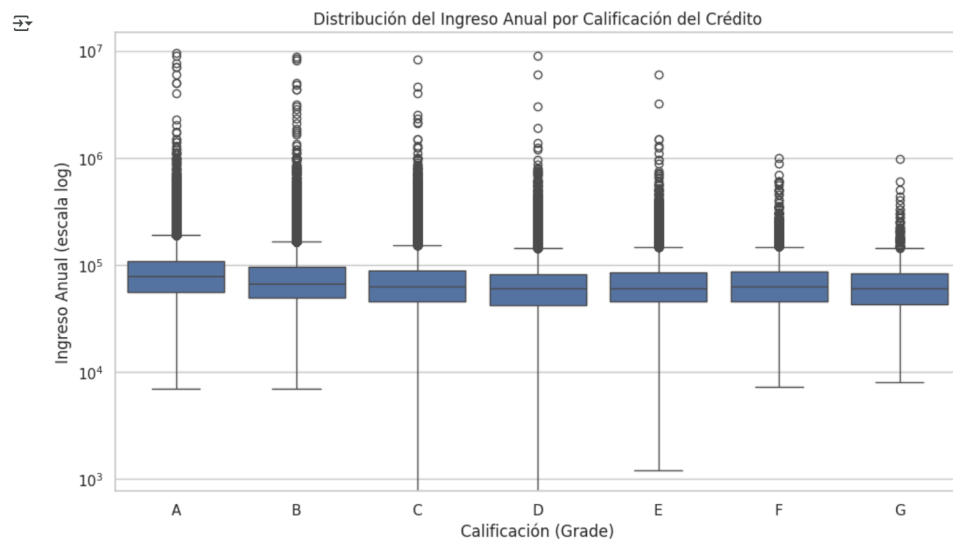
```



Los pagos completos corresponden a casi el 71% de los préstamos dados, mientras que los préstamos cancelados son cerca del 18% y los corrientes o actuales el 10.7%.

- b. Genere e implemente una hipótesis que explique la calificación.

Hipótesis: A mayor ingreso anual (annual_inc), mejor será la calificación del crédito (grade)



```

Ingreso promedio anual por calificación:
grade
A      91982.929933
B      79420.716855
C      73004.380654
D      68862.017774
E      71318.338438
F      72223.362823
G      70214.870315

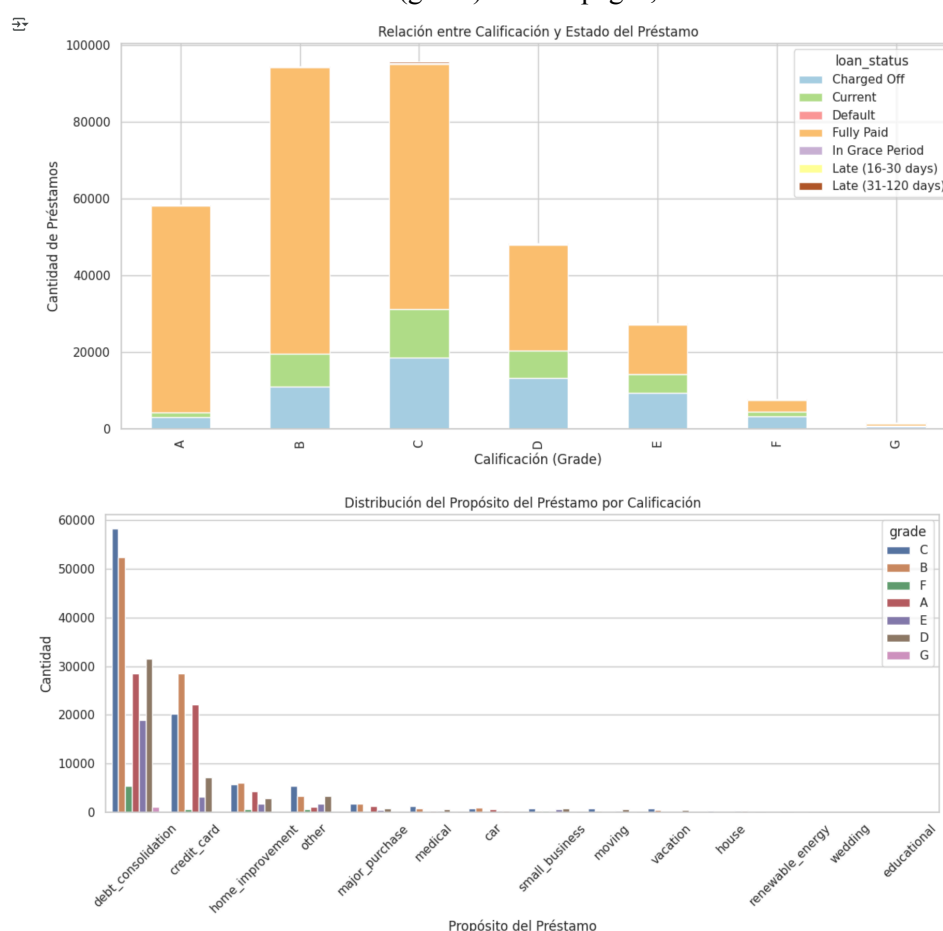
```

La calificación crediticia parece estar relacionada, en cierta medida, con el ingreso anual del solicitante. En el gráfico de cajas se observa que las calificaciones más altas (como A y B) tienden a concentrar solicitantes con ingresos ligeramente mayores en comparación con las calificaciones más bajas (como F y G). Aunque la mediana de

ingresos no varía drásticamente entre categorías, los ingresos promedio por calificación muestran una leve tendencia decreciente a medida que empeora la calificación (de A a G), lo que es consistente con la hipótesis inicial.

Esto sugiere que a mayor ingreso anual, existe una mayor probabilidad de recibir una mejor calificación crediticia, aunque el ingreso no es el único factor determinante. También es importante considerar otros atributos, como el historial de pagos, el nivel de endeudamiento y la estabilidad laboral. Por ello, la hipótesis es parcialmente respaldada por los datos, pero requiere análisis complementarios para confirmar su validez.

c. Cómo se relaciona la calificación (grade) con los pagos, el destino.

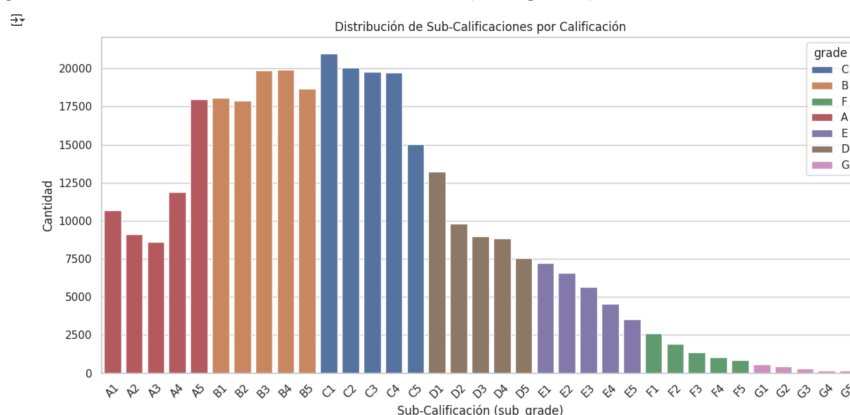


En cuanto al comportamiento de pago, se observa que los usuarios con mejor calificación crediticia tienden a recibir un mayor número de préstamos, y además son quienes más frecuentemente los pagan en su totalidad. Por el contrario, al analizar las calificaciones de D a G, se nota una disminución tanto en la cantidad de créditos otorgados como en el número de créditos pagados o cancelados, lo que indica un mayor riesgo de incumplimiento en estos grupos.

Respecto al propósito del crédito, los usuarios con mejores calificaciones suelen recibir montos más elevados, especialmente para fines como la consolidación de deudas o el pago de tarjetas de crédito. Sin embargo, al comparar con los niveles medios y bajos de calificación, aunque el monto del préstamo tiende a ser menor, los destinos del crédito se mantienen relativamente constantes. Esto sugiere que no existe

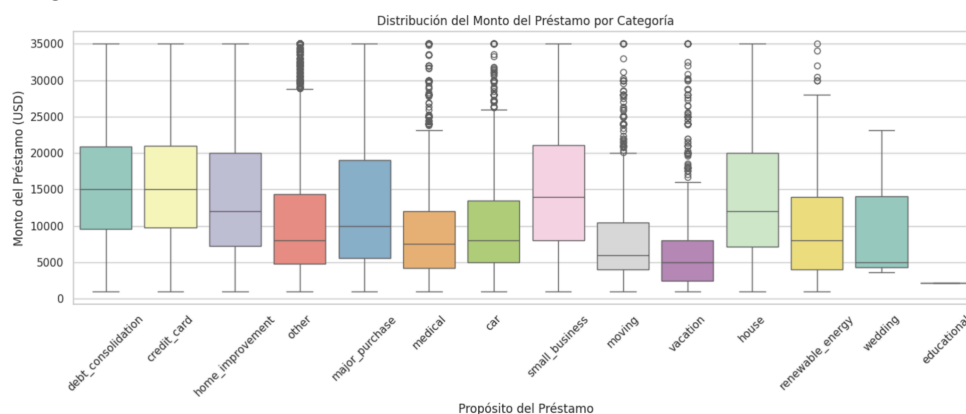
una fuerte correlación entre la calificación crediticia y el propósito declarado del préstamo, sino que la calificación parece estar más influenciada por el perfil de riesgo y comportamiento financiero del solicitante.

d. ¿Cómo se relaciona la sub-calificación (sub-grade)?



La sub-calificación (sub_grade) está directamente relacionada con la calificación principal (grade), ya que representa una subdivisión dentro de cada nivel de riesgo. En el gráfico se observa que cada grupo de calificación (A, B, C, etc.) contiene sus respectivas sub-calificaciones (A1–A5, B1–B5, etc.), mostrando una estructura jerárquica interna. Esta organización permite una evaluación más fina del riesgo crediticio, donde, por ejemplo, A1 indica menor riesgo que A5. Además, la frecuencia de ciertas sub-calificaciones dentro de cada grupo sugiere patrones en la asignación del riesgo que podrían estar influenciados por políticas de aprobación o comportamiento de los solicitantes.

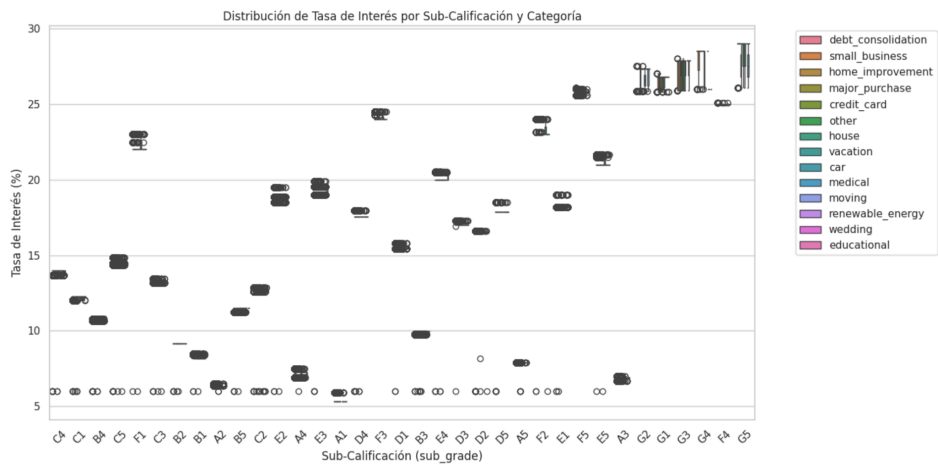
e. ¿Cuántos créditos hay por categoría? Y ¿cómo varía el monto del préstamo dentro de la categoría?



El número de créditos varía significativamente entre categorías, siendo más frecuentes los préstamos destinados a la consolidación de deudas y al pago de tarjetas de crédito. Estas dos categorías concentran la mayor parte del volumen total de solicitudes. En cuanto al monto del préstamo, se observa una amplia variación dentro de cada categoría. Algunas, como los préstamos educativos o para mejoras del hogar, tienden a tener montos mayores, mientras que otras como gastos médicos o bodas muestran montos más moderados. El análisis por la gráfica boxplot permite visualizar cómo la distribución del monto cambia por categoría, destacando diferencias

relevantes en las decisiones de financiación según el propósito declarado por los solicitantes.

- f. ¿La tasa de interés varía con la categoría y subcategoría? Analice el promedio, la desviación estándar, mínimos y máximos de la tasa de interés por categoría y subcategoría. ¿Este comportamiento es el esperado en su hipótesis, por qué?

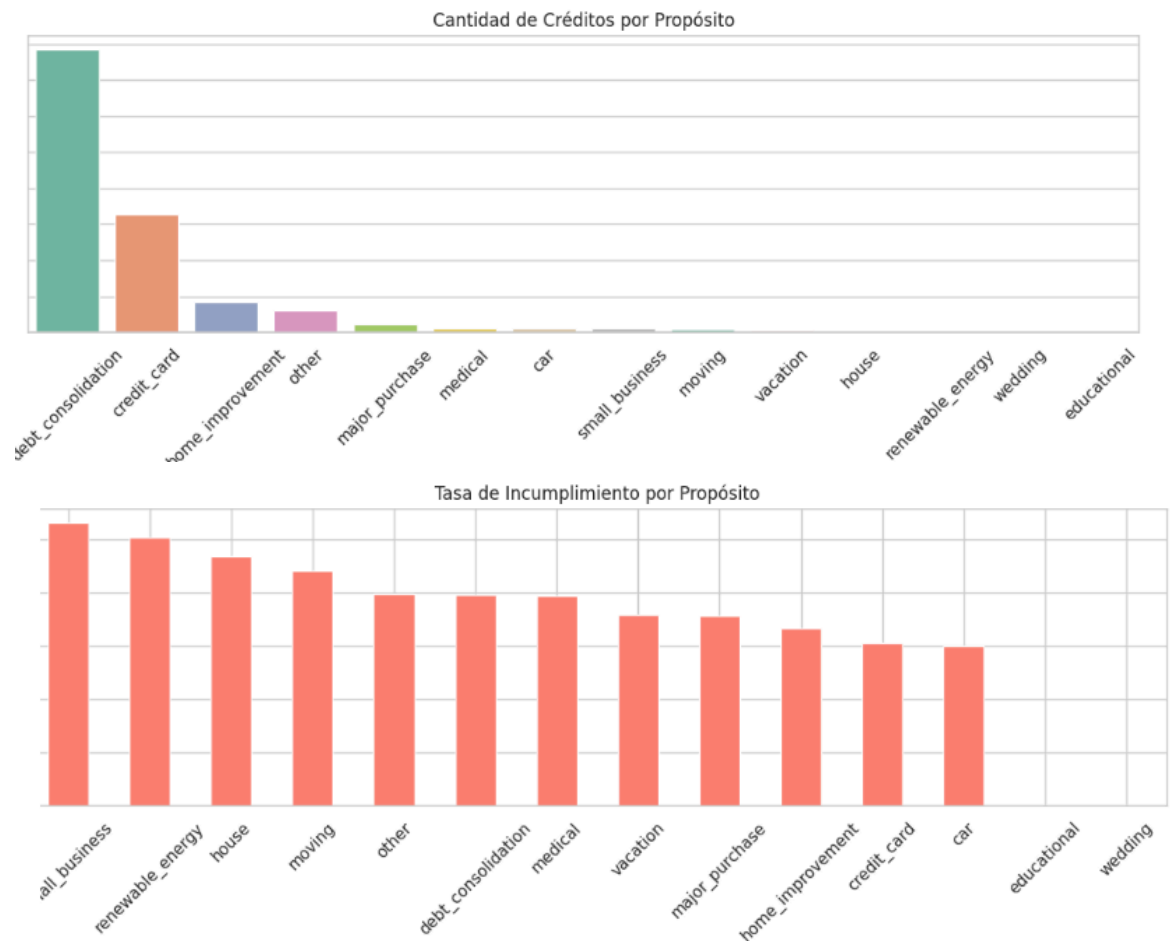


Estadísticas de tasa de interés por categoría y subcategoría:

	purpose	sub_grade	promedio	desviacion	minimo	maximo
0	car	A1	5.371915	0.171127	5.32	5.93
1	car	A2	6.262321	0.061454	6.24	6.49
2	car	A3	6.835045	0.095716	6.68	6.99
3	car	A4	7.214404	0.140062	6.92	7.49
4	car	A5	7.890296	0.002419	7.89	7.91
5	car	B1	8.217513	0.087705	8.18	8.49
6	car	B2	9.170000	0.000000	9.17	9.17
7	car	B3	9.946911	0.088409	9.76	9.99
8	car	B4	10.956707	0.097604	10.64	10.99
9	car	B5	11.486282	0.100561	11.22	11.53
10	car	C1	12.216349	0.122465	11.99	12.29
11	car	C2	12.600333	0.703827	6.00	12.88
12	car	C3	13.299324	0.062288	13.18	13.44
13	car	C4	13.933662	0.122309	13.67	13.99
14	car	C5	14.523333	0.813469	6.00	14.85
15	car	D1	15.557527	0.095276	15.41	15.77
16	car	D2	16.551311	0.007182	16.55	16.59
17	car	D3	16.995091	0.037755	16.99	17.27
18	car	D4	17.586667	0.080776	17.57	17.97
19	car	D5	17.872600	0.089095	17.86	18.49

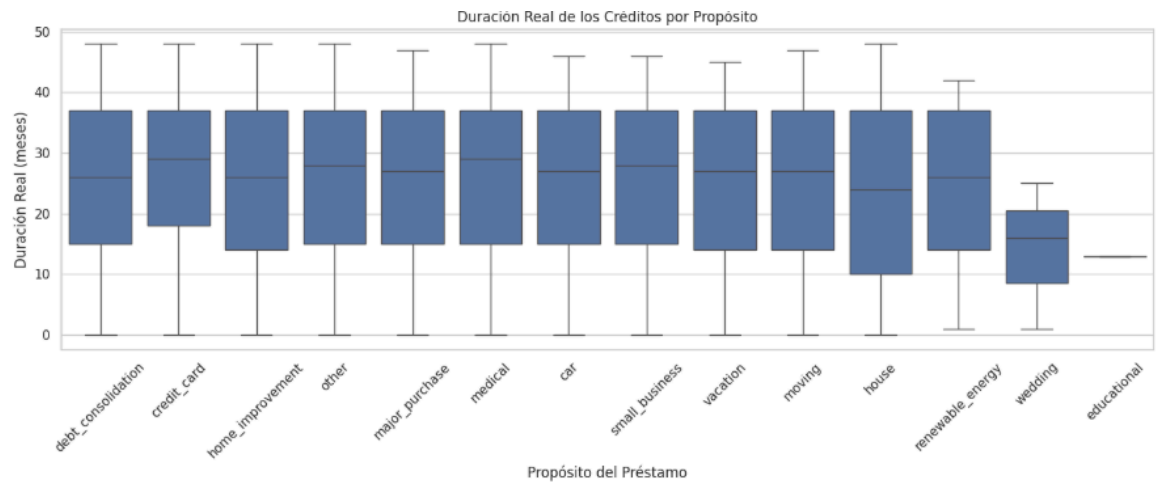
Podemos decir que las tasas de interés sí difieren según las categorías y las subcategorías, dado lo observado en los resultados. Esto va alineado con nuestra hipótesis, dado que, por ejemplo, no es lo mismo obtener un crédito para un carro que para consolidar una deuda, de forma que sí esperábamos que variaran las tasas de acuerdo a las categorías y a las subcategorías.

- g. ¿Cuál es el destino del crédito? Analice los montos, la cantidad de créditos asignados, los plazos según el propósito. ¿Las tasas de incumplimiento están relacionadas con el propósito?



El análisis por propósito del crédito revela que los préstamos destinados a la consolidación de deuda y al pago de tarjetas de crédito son los más solicitados, con montos promedio moderados y plazos estándar. Sin embargo, estas mismas categorías también presentan tasas de incumplimiento relativamente altas, lo que puede indicar un mayor riesgo asociado a quienes buscan refinanciar sus obligaciones previas. Por otro lado, propósitos como mejoras del hogar o adquisición de vehículos muestran una menor tasa de incumplimiento, lo cual podría reflejar un perfil de solicitante más solvente o con objetivos financieros más planificados. Estas diferencias son clave para evaluar el riesgo crediticio según el destino del préstamo.

- h. Para los créditos que están completamente pagados, ¿cómo se distribuye el tiempo y la cantidad de pagos? ¿Cuál es el plazo real?, fecha de asignación del crédito y fecha del último pago. ¿Cómo varían los plazos de acuerdo con la categoría del crédito? Explore el diagrama de box para analizar fácilmente los datos.



El análisis de la duración real de los créditos pagados completamente muestra que, si bien la mayoría de los préstamos se pagan dentro del plazo estipulado, existen diferencias según el propósito del crédito. Por ejemplo, los préstamos destinados a mejoras del hogar o a la compra de vehículos tienden a mantenerse durante todo el plazo oficial, mientras que otros, como los solicitados para consolidación de deuda o gastos médicos, a menudo se cancelan antes del tiempo estimado. Esto podría reflejar una mayor intención de cancelar rápidamente ciertas deudas o una mejor capacidad de pago en determinados perfiles. Comparar la duración real con la estipulada permite identificar comportamientos de pago por tipo de necesidad.

- i. ¿Qué sucede con los datos faltantes? ¿Cómo los maneja? ¿Cuál es el impacto de los datos faltantes en las conclusiones del análisis?

El manejo de los datos faltantes se realizó eliminando columnas con más del 80% de nulos y filas con pocos faltantes en variables clave, además de imputar con la mediana en casos específicos. Esto reduce el sesgo y mejora la calidad del análisis, aunque puede limitar la representatividad. Por tanto, los datos faltantes se manejaron según su impacto para no comprometer las conclusiones.

- j. ¿En cuáles casos los datos faltantes son normales en el negocio? Considere el caso de un atributo `monthsSinceLastDelinquency` tiene sentido en el negocio que tenga un valor nulo o vacío. Para cada caso, identifique cómo se debe reemplazar los datos faltantes y su significado.

En el contexto del negocio, el hecho de que haya valores nulos o vacíos en ese atributo importaría dado que indicaría que ese usuario nunca ha tenido valores de morosidad o de deuda, por lo que es una persona a la que se le puede seguir brindando créditos. En caso de que se evalúe con otras variables, como el tiempo que lleva empleado, se podría evaluar si sigue siendo meritorio para acceder a un crédito, o rechazarlo porque se desconoce si podría ser capaz financieramente de pagarlo.

- k. ¿Cuál variable debe descartarse del modelo porque contiene faltantes? ¿cómo afecta el comportamiento del modelo? ¿Cuáles son las suposiciones que se deben establecer con atención sobre las conclusiones del modelo?

El análisis identifica las variables con altos porcentajes de datos faltantes, destacando aquellas con más del 20% y señalando como candidatas para eliminación las que superan el 40%. Esto permite depurar el conjunto de datos eliminando columnas poco

informativas, lo cual mejora la calidad del modelo y evita conclusiones sesgadas por datos incompletos.

1. ¿Cuáles son los datos que deben estar disponibles cuando se desea invertir en un crédito?

Estructura de entrada para evaluación de crédito:						
	loan_amnt	term	int_rate	grade	sub_grade	purpose \
0	3600.0	36 months	13.99	C	C4	debt_consolidation
1	24700.0	36 months	11.99	C	C1	small_business
2	20000.0	60 months	10.78	B	B4	home_improvement
3	35000.0	60 months	14.85	C	C5	debt_consolidation
4	10400.0	60 months	22.45	F	F1	major_purchase

	emp_length	annual_inc	home_ownership	dti	fico_range_low \
0	10+ years	55000.0	MORTGAGE	5.91	675.0
1	10+ years	65000.0	MORTGAGE	16.06	715.0
2	10+ years	63000.0	MORTGAGE	10.78	695.0
3	10+ years	110000.0	MORTGAGE	17.06	785.0
4	3 years	104433.0	MORTGAGE	25.37	695.0

	fico_range_high	revol_util	mths_since_last_delinq	application_type \
0	679.0	29.7	30.0	Individual
1	719.0	19.2	6.0	Individual
2	699.0	56.2	NaN	Joint App
3	789.0	11.6	NaN	Individual
4	699.0	64.5	12.0	Individual

	annual_inc_joint	loan_status	issue_d	total_pymnt	last_pymnt_d
0	NaN	Fully Paid	Dec-2015	4421.723917	Jan-2019
1	NaN	Fully Paid	Dec-2015	25679.660000	Jun-2016
2	71000.0	Fully Paid	Dec-2015	22705.924294	Jun-2017
3	NaN	Current	Dec-2015	31464.010000	Feb-2019
4	NaN	Fully Paid	Dec-2015	11740.500000	Jul-2016

Este análisis identifica las variables esenciales que deben estar disponibles al evaluar una inversión en un crédito, como el monto, la tasa de interés, el ingreso del solicitante, su historial crediticio y el propósito del préstamo. Disponer de estos datos es clave para tomar decisiones informadas y mitigar riesgos, ya que permiten estimar la capacidad de pago y la probabilidad de incumplimiento del solicitante.

3. Modelo creado para identificar un crédito bueno y uno malo.

- Considere el rendimiento de una inversión correspondiente al retorno anual de la inversión.

Para evaluar el rendimiento de las inversiones, se reconstruyó la variable term a partir de la codificación binaria (term_ 60 months) para identificar si los préstamos tenían una duración de 36 o 60 meses. Posteriormente, se calculó el retorno bruto dividiendo el total pagado al inversionista (total_pymnt_inv) entre el monto prestado (loan_amnt) y restando 1. Este retorno se analizó considerando la duración real del préstamo expresada en años. El resultado muestra que el retorno anual promedio de los préstamos en el conjunto de datos fue de aproximadamente 0.0373%, lo que indica que, en promedio, las inversiones apenas superaron el punto de equilibrio, sugiriendo un bajo rendimiento anualizado para el inversionista en términos generales.

- ¿Los préstamos cancelados, generaron algún tipo de retorno?

Los préstamos cancelados ("Charged Off") no generaron retornos positivos para el inversionista, ya que la mayoría de ellos no fueron pagados o solo se recuperó una fracción mínima del monto prestado. Como consecuencia, no se calcula un retorno anual compuesto, ya que no se cumplió el ciclo completo de pago del préstamo.

Total Charged Off: 0

Con algún pago: 0

count	0.0
mean	NaN
std	NaN
min	NaN
25%	NaN
50%	NaN
75%	NaN
max	NaN

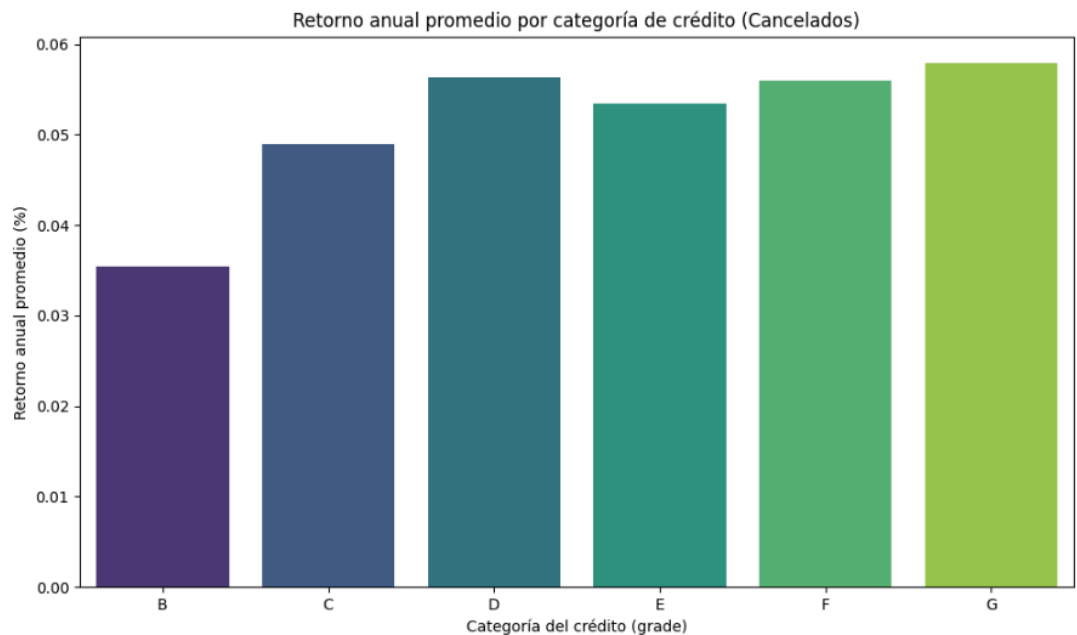
- ¿Cómo se explica la variación de los retornos de los créditos cancelados con respecto a la categoría del crédito?

La variación de los retornos de los créditos cancelados según la categoría del crédito (grade) se puede explicar a partir del nivel de riesgo asociado a cada categoría.

En el sistema de calificación utilizado (de A a G), los créditos de categoría A representan menor riesgo de incumplimiento, mientras que los de categoría G representan un riesgo significativamente mayor. Para compensar este mayor riesgo, los créditos con categorías más bajas (E, F, G) suelen ofrecer tasas de interés más altas, lo que se traduce en mayores retornos anuales para el inversionista, siempre y cuando el préstamo sea pagado completamente.

Según el análisis realizado:

- Los retornos anuales promedio aumentan conforme la categoría del crédito baja (de A hacia G).
- Esto refleja una relación directa entre riesgo y retorno esperado: los inversionistas que asumen un mayor riesgo potencial (categorías bajas) reciben una mejor recompensa, en promedio.
- Sin embargo, esta observación se basa solo en los préstamos cancelados exitosamente, por lo que no refleja el riesgo real de incumplimiento. Si se incluyeran los préstamos en mora o cargados como pérdida, los retornos de las categorías más bajas podrían verse negativamente afectados.



- ¿Existe diferencia entre el promedio de retorno de los créditos y la tasa de interés promedio? ¿Qué significa esta diferencia?

Tasa de interés promedio: 12.271014308840476
 Retorno anual promedio: 0.0003733111261247552
 Diferencia: 12.27064099771435

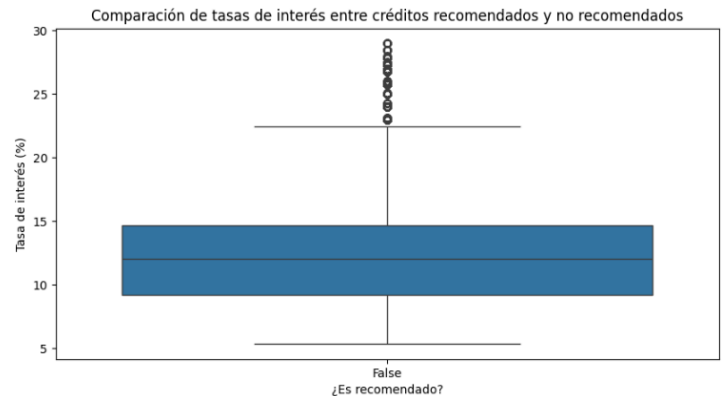
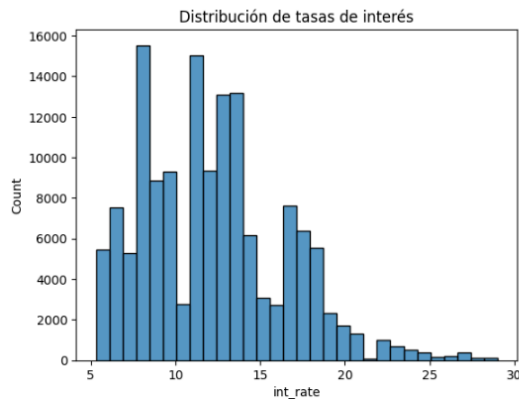
Con base en los resultados obtenidos, sí existe una diferencia significativa entre la tasa de interés promedio de los créditos y el retorno anual promedio para el inversionista. La tasa de interés promedio registrada fue de aproximadamente 12.27%, mientras que el retorno anual promedio real para el inversionista fue de apenas 0.037%. Esta brecha de más de 12 puntos porcentuales sugiere que, aunque los préstamos fueron emitidos con tasas de interés relativamente altas, los inversionistas no lograron capturar ese beneficio. Esta diferencia puede explicarse por diversos factores como pérdidas por incumplimientos (préstamos no pagados), comisiones, tiempos de inversión menores al plazo total, o reinversión no óptima del capital. Esto significa que la tasa de interés anunciada no refleja necesariamente el rendimiento real que obtiene el inversionista, lo cual es crucial al evaluar la rentabilidad de este tipo de instrumentos financieros.

- ¿Asumiendo el rol del inversionista, en cuáles créditos preferiría invertir y cuáles son las razones que motivan la decisión?

```

Total de créditos: 145784
Total recomendados: 0
count    145784.000000
mean      12.271814
std        4.202302
min        5.320000
25%        9.170000
50%       12.050000
75%       14.650000
max       28.990000
Name: int_rate, dtype: float64

```



Como inversionista preferiría invertir en créditos que cumplan con criterios estrictos de calidad y rendimiento. Específicamente, buscaría préstamos con tasas de interés de al menos 12% (el parámetro `tasa_min` predeterminado), un puntaje FICO promedio mínimo de 680, y una relación cuota/ingreso inferior al 30%. Además, priorizaría prestatarios con empleo estable, sin historial de morosidad (`delinq_yrs` = 0), sin quiebras públicas registradas, y cuyos préstamos estén al día (fully paid o current). Esta estrategia equilibra el rendimiento (altas tasas de interés) con la seguridad (buenos perfiles crediticios), lo que resulta crucial para minimizar el riesgo de incumplimiento mientras se maximiza el retorno de la inversión.

Analizando los resultados específicos mostrados en las imágenes, como inversionista tomaría decisiones basadas en los siguientes datos:

- El gráfico de distribución de tasas de interés muestra que la mayoría de los créditos se concentran entre 9% y 15%, con una tasa promedio de 12.27%. La comparación entre créditos recomendados y no recomendados (boxplot) indica que actualmente no hay créditos recomendados (el resultado muestra "Total recomendados: 0"), posiblemente porque los criterios de selección son demasiado estrictos.
- Ajustaría mi estrategia para priorizar créditos con tasas superiores al promedio ($>12.27\%$) pero no en el extremo superior ($>20\%$) donde el riesgo podría ser excesivo. Específicamente, buscaría préstamos en el tercer cuartil (12.85% a 14.65%) que mantengan puntajes FICO sobre 680, ya que estos ofrecen un equilibrio óptimo entre rendimiento y seguridad. El histograma muestra que hay suficientes créditos en este rango para crear una cartera diversificada.
- La ausencia de créditos recomendados sugiere que se deben recalibrar los parámetros en la función `seleccionar_creditos_preferidos`, posiblemente reduciendo ligeramente el puntaje FICO mínimo a 670 o ajustando la tasa mínima al 11%, para obtener un conjunto viable de oportunidades de inversión sin sacrificar significativamente la calidad.

4. Construcción de nuevos atributos.

- Construya al menos 3 atributos derivados que pueden utilizarse para predecir el impago del crédito. Justifique el atributo en términos del negocio. Justifique el nuevo atributo en términos del rendimiento del modelo.

Para mejorar la capacidad predictiva del modelo frente al **impago de créditos**, se construyeron tres atributos derivados que combinan lógica de negocio con fundamentos analíticos.

- El primer atributo es la relación cuota-ingreso (*cuota_ingreso*), calculada como la cuota mensual del préstamo dividida entre el ingreso mensual del solicitante. Este indicador es relevante desde la perspectiva del negocio porque mide la carga financiera del crédito sobre el ingreso del cliente, lo cual es un predictor natural del riesgo de impago: cuanto mayor sea esta proporción, más probable es que el cliente enfrente dificultades para cumplir con sus obligaciones. Desde el punto de vista del modelo, introduce una variable continua y directa sobre la capacidad de pago.
- El segundo atributo es empleo estable (*empleo_estable*), una variable binaria que toma el valor de 1 si el solicitante tiene al menos 5 años de experiencia laboral. En términos del negocio, la estabilidad laboral sugiere mayor confiabilidad y flujo constante de ingresos, lo que reduce la probabilidad de impago. Para el modelo, este atributo ayuda a capturar aspectos estructurales del historial laboral sin necesidad de múltiples variables dummy.
- El tercer atributo es el rango FICO promedio (*rango_fico_promedio*), una transformación que resume el intervalo de la puntuación crediticia en una sola métrica. En términos financieros, el puntaje FICO es un indicador consolidado del historial crediticio del solicitante, por lo que su valor promedio ofrece una forma más suave y generalizable de representar este factor en el modelo, mejorando la eficiencia del aprendizaje sin perder información relevante.

5. Variables con fuga de datos.

- ¿Cuáles variables debe excluir del análisis porque se consideran fuga de datos?
Con base en la detección de fugas realizada mediante la función *detectar_fugas_logicas*, se identificaron dos variables que deben ser excluidas del análisis porque representan fugas de datos:
 1. *loan_status* (correlación: -0.78): Esta variable indica el estado final del préstamo (por ejemplo, pagado, impago, cancelado), lo cual ocurre después del otorgamiento del crédito. Usarla en la predicción del rendimiento o impago sería una fuga grave, ya que da información que solo está disponible *una vez ha concluido el préstamo*, y no en el momento de tomar decisiones de inversión.
 2. *retorno* (correlación: 0.96): Este valor es calculado directamente a partir de datos posteriores al otorgamiento del crédito, como los pagos realizados (*total_pymnt_inv*), por lo tanto es redundante con la variable objetivo *retorno_anual*. Incluirlo en el modelo sería como anticipar la respuesta con información derivada de la misma.

Estas variables deben ser excluidas porque contienen información que no estaría disponible al momento de decidir si invertir o no, lo que llevaría a un modelo sobreajustado y poco realista en la práctica.

- ¿Cuáles variables individuales son particularmente útiles para predecir la variable dependiente del estado del crédito?

Los resultados muestran las variables numéricas que tienen mayor correlación individual con la variable objetivo `loan_status_original_Default`, es decir, aquellas que más se relacionan con el hecho de que un crédito termine o no en incumplimiento. Sin embargo, los valores de correlación son bastante bajos (todas por debajo de 0.01), lo que indica que **ninguna variable por sí sola tiene una relación fuerte con el estado del crédito**. Por ejemplo, `total_acc` y `mort_acc` tienen los valores más altos (0.0075 y 0.0071 respectivamente), lo que sugiere que el número total de cuentas o de hipotecas abiertas tiene cierta, aunque muy limitada, capacidad para distinguir entre prestatarios que incumplen y los que no.

<code>total_acc</code>	0.007497
<code>mort_acc</code>	0.007152
<code>loan_amnt</code>	0.006002
<code>loan_status</code>	0.005594
<code>int_rate</code>	0.005434
<code>installment</code>	0.005395
<code>revol_util</code>	0.005030
<code>total_pymnt_inv</code>	0.004079
<code>term_years</code>	0.003795
<code>open_acc</code>	0.003685

- Construya una medida que relacione cada una de las variables independientes con la variable predictora. Considere que la variable a predecir es binaria. Describa la interpretación de la medida.

Para responder a la pregunta, se utilizó el coeficiente de correlación biserial puntual (`pointbiserialr`) como medida estadística para relacionar cada variable independiente numérica con la variable binaria a predecir: `loan_status_original_Default`. Este coeficiente evalúa la fuerza y dirección de la relación lineal entre una variable numérica y una variable binaria. En este caso, los valores obtenidos indican qué tan útil es cada variable individualmente para diferenciar entre préstamos que cayeron en incumplimiento y los que no. Los resultados muestran que todas las correlaciones son bajas, con valores absolutos menores a 0.01, lo cual implica que ninguna variable numérica individual tiene una relación fuerte con el estado de incumplimiento del crédito. Esto sugiere que se requerirá un enfoque multivariado más complejo, como modelos de aprendizaje automático, para lograr una predicción efectiva. Además, también puede ser necesario incorporar y codificar adecuadamente variables categóricas para mejorar la capacidad predictiva del modelo.



0

total_acc	0.007497
mort_acc	0.007152
loan_amnt	0.006002
loan_status	0.005594
int_rate	0.005434
installment	0.005395
revol_util	0.005030
total_pymnt_inv	0.004079
term_years	0.003795
open_acc	0.003685