

TALLER LENDING CLUB SELECCIÓN DE VARIABLES

Integrantes: Sara Peña y Frank Hernandez

1. Etapa del proceso para cada una de las variables del conjunto de datos:

- `acc_now_delinq`: Aprobación, es información que revisaría el prestamista o el banco.
- `acc_open_past_24mths`: Aprobación, es información que revisaría el prestamista o el banco.
- `addr_state`: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- `all_util`: Aprobación, es información que revisaría el prestamista o el banco.
- `annual_inc`: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- `annual_inc_joint`: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- `application_type`: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- `avg_cur_bal`: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- `bc_open_to_buy`: Aprobación, es información que revisaría el prestamista o el banco.
- `bc_util`: Aprobación, es información que revisaría el prestamista o el banco.
- `chargeoff_within_12_mths`: Aprobación, es información que revisaría el prestamista o el banco.
- `collection_recovery_fee`: Cierre del crédito, es un dato que podría darse al final del proceso.
- `collections_12_mths_ex_med`: No clasificada, no se encuentra una relación clara con alguna de las etapas.
- `delinq_2yrs`: Aprobación, es información que revisaría el prestamista o el banco.
- `delinq_amnt`: Aprobación, es información que revisaría el prestamista o el banco.
- `desc`: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- `dti`: Aprobación, es información que revisaría el prestamista o el banco.
- `dti_joint`: Aprobación, es información que revisaría el prestamista o el banco.
- `earliest_cr_line`: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- `emp_length`: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- `emp_title`: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- `fico_range_high`: Aprobación, es información que revisaría el prestamista o el banco.

- fico_range_low: Aprobación, es información que revisaría el prestamista o el banco.
- funded_amnt: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- funded_amnt_inv: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- grade: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- home_ownership: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- id: No clasificada
- il_util: Aprobación, es información que revisaría el prestamista o el banco.
- initial_list_status: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- inq_fi: Aprobación, es información que revisaría el prestamista o el banco.
- inq_last_12m: Aprobación, es información que revisaría el prestamista o el banco.
- inq_last_6mths: Aprobación, es información que revisaría el prestamista o el banco.
- installment: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- int_rate: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- issue_d: Cierre del crédito, es un dato que podría darse al final del proceso.
- last_credit_pull_d: Aprobación, es información que revisaría el prestamista o el banco.
- last_fico_range_high: Aprobación, es información que revisaría el prestamista o el banco.
- last_fico_range_low: Aprobación, es información que revisaría el prestamista o el banco.
- last_pymnt_amnt: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- last_pymnt_d: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- loan_amnt: Aprobación, es información que revisaría el prestamista o el banco.
- loan_status: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- max_bal_bc: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- member_id: Solicitud, es información que genera el prestamista al momento de recibir la solicitud del crédito.
- mo_sin_old_il_acct: Aprobación, es información que revisaría el prestamista o el banco.
- mo_sin_old_rev_tl_op: Aprobación, es información que revisaría el prestamista o el banco.
- mo_sin_rcnt_rev_tl_op: Aprobación, es información que revisaría el prestamista o el banco.

- mo_sin_rcnt_tl: Aprobación, es información que revisaría el prestamista o el banco.
- mort_acc: Aprobación, es información que revisaría el prestamista o el banco.
- mths_since_last_delinq: Aprobación, es información que revisaría el prestamista o el banco.
- mths_since_last_major_derog: Aprobación, es información que revisaría el prestamista o el banco.
- mths_since_last_record: Aprobación, es información que revisaría el prestamista o el banco.
- mths_since_rcnt_il: Aprobación, es información que revisaría el prestamista o el banco.
- mths_since_recent_bc: Aprobación, es información que revisaría el prestamista o el banco.
- mths_since_recent_bc_dlq: Aprobación, es información que revisaría el prestamista o el banco.
- mths_since_recent_inq: Aprobación, es información que revisaría el prestamista o el banco.
- mths_since_recent_revol_delinq: Aprobación, es información que revisaría el prestamista o el banco.
- next_pymnt_d: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- num_accts_ever_120_pd: Aprobación, es información que revisaría el prestamista o el banco.
- num_actv_bc_tl: Aprobación, es información que revisaría el prestamista o el banco.
- num_actv_rev_tl: Aprobación, es información que revisaría el prestamista o el banco.
- num_bc_sats: Aprobación, es información que revisaría el prestamista o el banco.
- num_bc_tl: Aprobación, es información que revisaría el prestamista o el banco.
- num_il_tl: Aprobación, es información que revisaría el prestamista o el banco.
- num_op_rev_tl: Aprobación, es información que revisaría el prestamista o el banco.
- num_rev_accts: Aprobación, es información que revisaría el prestamista o el banco.
- num_rev_tl_bal_gt_0: Aprobación, es información que revisaría el prestamista o el banco.
- num_sats: Aprobación, es información que revisaría el prestamista o el banco.
- num_tl_120dpd_2m: Aprobación, es información que revisaría el prestamista o el banco.
- num_tl_30dpd: Aprobación, es información que revisaría el prestamista o el banco.
- num_tl_90g_dpd_24m: Aprobación, es información que revisaría el prestamista o el banco.

- num_tl_op_past_12m: Aprobación, es información que revisaría el prestamista o el banco.
- open_acc: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- open_acc_6m: Aprobación, es información que revisaría el prestamista o el banco.
- open_il_12m: Aprobación, es información que revisaría el prestamista o el banco.
- open_il_24m: Aprobación, es información que revisaría el prestamista o el banco.
- open_act_il: Aprobación, es información que revisaría el prestamista o el banco.
- open_rv_12m: Aprobación, es información que revisaría el prestamista o el banco.
- open_rv_24m: Aprobación, es información que revisaría el prestamista o el banco.
- out_prncp: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- out_prncp_inv: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- pct_tl_nvr_dlq: Aprobación, es información que revisaría el prestamista o el banco.
- percent_bc_gt_75: Aprobación, es información que revisaría el prestamista o el banco.
- policy_code: No clasificada, no se encuentra una relación clara con alguna de las etapas.
- pub_rec: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- pub_rec_bankruptcies: Aprobación, es información que revisaría el prestamista o el banco.
- purpose: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- pymnt_plan: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- recoveries: Cierre del crédito, es un dato que podría darse al final del proceso.
- revol_bal: Aprobación, es información que revisaría el prestamista o el banco.
- revol_util: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- sub_grade: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- tax_liens: Aprobación, es información que revisaría el prestamista o el banco.
- term: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- title: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- tot_coll_amt: Aprobación, es información que revisaría el prestamista o el banco.

- tot_cur_bal: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- tot_hi_cred_lim: Aprobación, es información que revisaría el prestamista o el banco.
- total_acc: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- total_bal_ex_mort: Aprobación, es información que revisaría el prestamista o el banco.
- total_bal_il: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- total_bc_limit: Aprobación, es información que revisaría el prestamista o el banco.
- total_cu_tl: Aprobación, es información que revisaría el prestamista o el banco.
- total_il_high_credit_limit: Aprobación, es información que revisaría el prestamista o el banco.
- total_pymnt: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- total_pymnt_inv: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- total_rec_int: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- total_rec_late_fee: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- total_rec_prncp: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- total_rev_hi_lim: Aprobación, es información que revisaría el prestamista o el banco.
- url: No clasificada, no se encuentra una relación clara con alguna de las etapas.
- verification_status: Aprobación, es información que revisaría el prestamista o el banco.
- verified_status_joint: Aprobación, es información que revisaría el prestamista o el banco.
- zip_code: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- revol_bal_joint : Aprobación, es información que revisaría el prestamista o el banco.
- sec_app_fico_range_low: Aprobación, es información que revisaría el prestamista o el banco.
- sec_app_fico_range_high: Aprobación, es información que revisaría el prestamista o el banco.
- sec_app_earliest_cr_line: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- sec_app_inq_last_6mths: Aprobación, es información que revisaría el prestamista o el banco.
- sec_app_mort_acc : Aprobación, es información que revisaría el prestamista o el banco.

- sec_app_open_acc : Aprobación, es información que revisaría el prestamista o el banco.
- sec_app_revol_util : Aprobación, es información que revisaría el prestamista o el banco.
- sec_app_open_act_il: Aprobación, es información que revisaría el prestamista o el banco.
- sec_app_num_rev_accts : Aprobación, es información que revisaría el prestamista o el banco.
- sec_app_chargeoff_within_12_mths : Aprobación, es información que revisaría el prestamista o el banco.
- sec_app_collections_12_mths_ex_med : Aprobación, es información que revisaría el prestamista o el banco.
- sec_app_mths_since_last_major_derog : Aprobación, es información que revisaría el prestamista o el banco.
- hardship_flag: Aprobación, es información que revisaría el prestamista o el banco.
- hardship_type: Aprobación, es información que revisaría el prestamista o el banco.
- hardship_reason: Aprobación, es información que revisaría el prestamista o el banco.
- hardship_status: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- deferral_term: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- hardship_amount: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- hardship_start_date: Aprobación, es información que revisaría el prestamista o el banco.
- hardship_end_date: Amortización, porque es un dato que se define en la aprobación, pero puede variar mientras el cliente pague más que el valor de cuota definido..
- payment_plan_start_date: Aprobación, es información que revisaría el prestamista o el banco.
- hardship_length: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- hardship_dpd: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- hardship_loan_status: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- orig_projected_additional_accrued_interest: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- hardship_payoff_balance_amount: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- hardship_last_payment_amount: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- disbursement_method: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.

- `debt_settlement_flag`: Aprobación, es información que revisaría el prestamista o el banco.
- `debt_settlement_flag_date`: Aprobación, es información que revisaría el prestamista o el banco.
- `settlement_status`: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- `settlement_date`: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- `settlement_amount`: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.
- `settlement_percentage`: Amortización, porque es un dato que se da mientras el cliente haga los pagos.
- `settlement_term`: Solicitud, es información que brinda el usuario al momento de hacer la solicitud del crédito.

2. Listado de variables elegidas para cada etapa del proceso:

- Solicitud del crédito
 - `annual_inc`: ingreso anual
 - `avg_cur_bal`: promedio actual de todas las cuentas del usuario
 - `desc`: descripción del préstamo dado por el usuario
 - `emp_length`: Años de experiencia personal
 - `emp_title`: Título profesional del usuario que quiere solicitar un crédito
 - `home_ownership`: Propiedad de vivienda (propia, alquilada, hipotecada).
 - `purpose`: Categoría dada por el usuario para el crédito
 - `installment`: Pago mensual que podría acoger si el préstamo es aprobado.
 - `loan_amnt`: Monto solicitado en el préstamo
 - `title`: Título del crédito que el usuario creó
 - `zip_code`: Los primeros 3 dígitos del código postal del usuario que está creando la solicitud del crédito
 - `application_type`: Tipo de aplicación: individual o conjunta (por dos personas).
 - `verification_status`: Estado de verificación del ingreso.
 - `addr_state`: Estado o región de residencia.
 - `open_acc`: Número de cuentas de crédito abiertas.
 - `pub_rec`: Registros públicos adversos (ej: bancarrotas).
 - `total_acc`: Número total de líneas de crédito registradas.
 - `initial_list_status`: Estado inicial de listado (si fue directamente ofrecido al mercado).
 - `disbursement_method`: Método de desembolso solicitado (cash, directPay).
 - `earliest_cr_line`: Fecha de apertura de su primera línea de crédito.
 - `term`: Plazo solicitado (36 o 60 meses).
 - `int_rate`: Tasa de interés inicial ofrecida (relacionada al perfil de solicitud).
- Aprobación del crédito

- grade: Grado de riesgo asignado por Lending Club al préstamo (A, B, C, etc.).
 - sub-grade: Subgrado de riesgo más específico (A1, A2, ..., E5, etc.).
 - int_rate: Tasa de interés asignada al préstamo.
 - term: duración del prestamos (36 a 60 meses)
 - initial_list_status: Estado inicial de la publicación del préstamo (ej. aprobado o en revisión).
 - loan_amnt: Monto solicitado en el préstamo. (Aunque lo pide el usuario, LC puede ajustarlo)
 - funded_amnt: Monto finalmente aprobado (puede ser menor o igual al solicitado).
 - funded_amnt_inv: Monto financiado por los inversionistas.
 - application_type: Tipo de aplicación (individual o conjunta).
 - verification_status: Estado de verificación de ingresos del solicitante.
 - verification_status_joint: Estado de verificación conjunta si es un préstamo conjunto.
 - policy_code: Código de política de aprobación (interno de LC).
 - purpose: Propósito declarado del préstamo (influye en la evaluación).
- Otras Variables que se pueden tener en cuenta durante el análisis crediticio interno
 - dti: Relación deuda-ingreso (muy importante para decidir aprobación).
 - dti_joint: Relación deuda-ingreso conjunta.
 - fico_range_high: Rango alto del FICO Score inicial.
 - fico_range_low: Rango bajo del FICO Score inicial.
 - earliest_cr_line: Fecha de apertura de la primera línea de crédito.
 - home_ownership: Propiedad de vivienda (propia, alquilada, hipotecada).
 - emp_length: Años de experiencia laboral.
 - open_acc: Número de cuentas de crédito abiertas.
 - total_acc: Total de cuentas de crédito.
 - inq_last_6mths: Cantidad de consultas de crédito recientes (se usa para riesgo).
- Amortización del crédito
 - installment: Pago mensual que debe realizar el prestatario (valor fijo acordado).
 - last_pymnt_d: Fecha del último pago realizado.
 - last_pymnt_amnt: Monto del último pago realizado.
 - out_prncp: Capital pendiente (saldo de capital no pagado).
 - out_prncp_inv: Capital pendiente invertido (para los inversionistas).
 - total_pymnt: Total de pagos recibidos hasta el momento.
 - total_pymnt_inv: Total de pagos recibidos para los inversionistas.
 - total_rec_prncp: Total de capital recibido.
 - total_rec_int: Total de intereses recibidos.
 - total_rec_late_fee: Total de intereses por pagos tardíos recibidos.
 - recoveries: Monto recuperado después de default.

- `collection_recovery_fee`: Comisión cobrada por recuperación después de default.
 - `last_credit_pull_d`: Fecha en que Lending Club revisó el crédito por última vez (para actualizar riesgos).
 - `next_pymnt_d`: Fecha del próximo pago programado.
 - `loan_status`: Estado actual del préstamo (Fully Paid, Charged Off, etc.).
- Otras variables que pueden ayudar a mostrar el comportamiento del usuario después del desembolso:
 - `num_tl_90g_dpd_24m`: Número de líneas de crédito en mora 90 días en los últimos 24 meses.
 - `acc_now_delinq`: Número de cuentas actuales en mora.
 - `mths_since_last_delinq`: Meses desde la última morosidad.
 - `mths_since_recent_revol_delinq`: Meses desde la última mora en cuenta revolving (tarjeta de crédito)
- Cierre del crédito
 - `loan_status`: Estado final del préstamo: "Fully Paid", "Charged Off", etc.
 - `pymnt_plan`: Indica si el prestatario está en un plan de pagos especial (relacionado con cierre anticipado o modificaciones).
 - `last_pymnt_d`: Fecha del último pago realizado (muy importante para cierre).
 - `last_pymnt_amnt`: Monto del último pago realizado.
 - `total_pymnt`: Total de pagos hechos en la vida del préstamo.
 - `total_rec_prncp`: Capital total recuperado.
 - `total_rec_int`: Intereses totales recuperados.
 - `recoveries`: Recuperaciones obtenidas después de un default (si aplicó)
 - `collection_recovery_fee`: Comisiones cobradas por recuperación después de default.
 - `out_prncp`: Capital pendiente. Debe ser 0 si el crédito está completamente pagado.
 - `out_prncp_inv`: Capital pendiente para los inversionistas (igual, debe ser 0 en cierre exitoso).
 - `issue_d`: Fecha de desembolso del préstamo (sirve para calcular duración real)
 - `mths_since_last_major_derog`: Meses desde la última derogación importante (se usa para evaluar cierres complejos).
 - `next_pymnt_d`: Fecha del próximo pago programado — debería estar vacío (null) si el crédito está cerrado
- Otras variables que pueden ayudar a analizar cómo fue el cierre del crédito en los usuarios:
 - `mths_since_last_delinq`: Tiempo desde la última morosidad, es relevante si hubo problemas en el cierre.
 - `mths_since_recent_revol_delinq`: Tiempo desde última morosidad en cuentas revolving.

- `hardship_flag`: Indica si el préstamo tuvo un período de "hardship" (dificultad financiera) que podría haber afectado el cierre
- `disbursement_method`: Modo de desembolso del préstamo, que puede estar relacionado con el tipo de cierre (aunque es menos relevante).

3. Descripción a detalle de las variables relacionadas con la solicitud:

Para poder definir qué variables hacen parte de la etapa de solicitud, tuvimos en cuenta varios factores. En primer lugar, hicimos una revisión general de todas las variables, y nos fijamos en aquellos datos que debería brindar directamente el cliente, como sus ingresos, la profesión que desempeña y el código ZIP de donde vive. Posteriormente, empezamos a revisar las variables que estarían más relacionadas con el crédito que quisiera obtener el usuario, como son la cantidad de dinero que solicita, qué cuota podría pagar y el método en el que quisiera recibir el dinero, de forma que el prestamista pueda decidir bajo qué términos le presta el dinero si llegase a ocurrir. Y el tercer factor clave son variables que ayudan a entender qué tipo de cliente es el que solicita el dinero, y por ende establecer si se le debería prestar o no, como el tipo de aplicación, el número de cuentas de crédito abiertas y el plazo solicitado.

Posteriormente, pasamos a transformar las variables, y esto se hizo con el propósito de poder filtrar de mejor manera los datos en el modelo de predicción. Por ejemplo, consideramos la discretización por rangos de variables como el ingreso anual y la cantidad de dinero que solicita para el préstamo, puesto que son variables que podemos ajustar por rangos y así mismo interpretarlos como valores entre bajos y muy altos, de forma que decidamos si es viable o no brindarle un crédito a esa persona. Visto de otra manera, podríamos partir de variables como su ingreso anual y el balance financiero de la persona para saber si se le debería brindar un crédito bajo o alto, a partir de estas y de otras variables que se consideran en el modelo.

También se transformaron las variables categóricas, aquellas que contaban con información en texto y no numérica, de forma que a partir de valores enteros se pueda interpretar la información en el modelo. Y adicionalmente se cambiaron otros tipos de datos, como la fecha de apertura de la primera línea de crédito a un tipo de fecha, así como el título del crédito y el título profesional del cliente a valores binarios.

Ahora, todo esto se relaciona con el negocio en el sentido de que son datos clave para determinar si un crédito es bueno o malo, debido a que une el análisis financiero de la persona, con información como su profesión, si es dueño de su vivienda y su ingreso anual, con lo que respecta formalmente al crédito, como lo es el valor que solicita, la tasa de interés que se le brindaría, entre otras cosas. De esa manera, se predice de mejor manera el modelo, dándole las variables clave que ayuden a determinar si se debe o no ceder ese crédito.

4. Transformación de las variables de interés con las técnicas vistas en clase y con el apoyo de la lectura:

Recurrir al Colab: [P4 T Lending Club.ipynb](#)