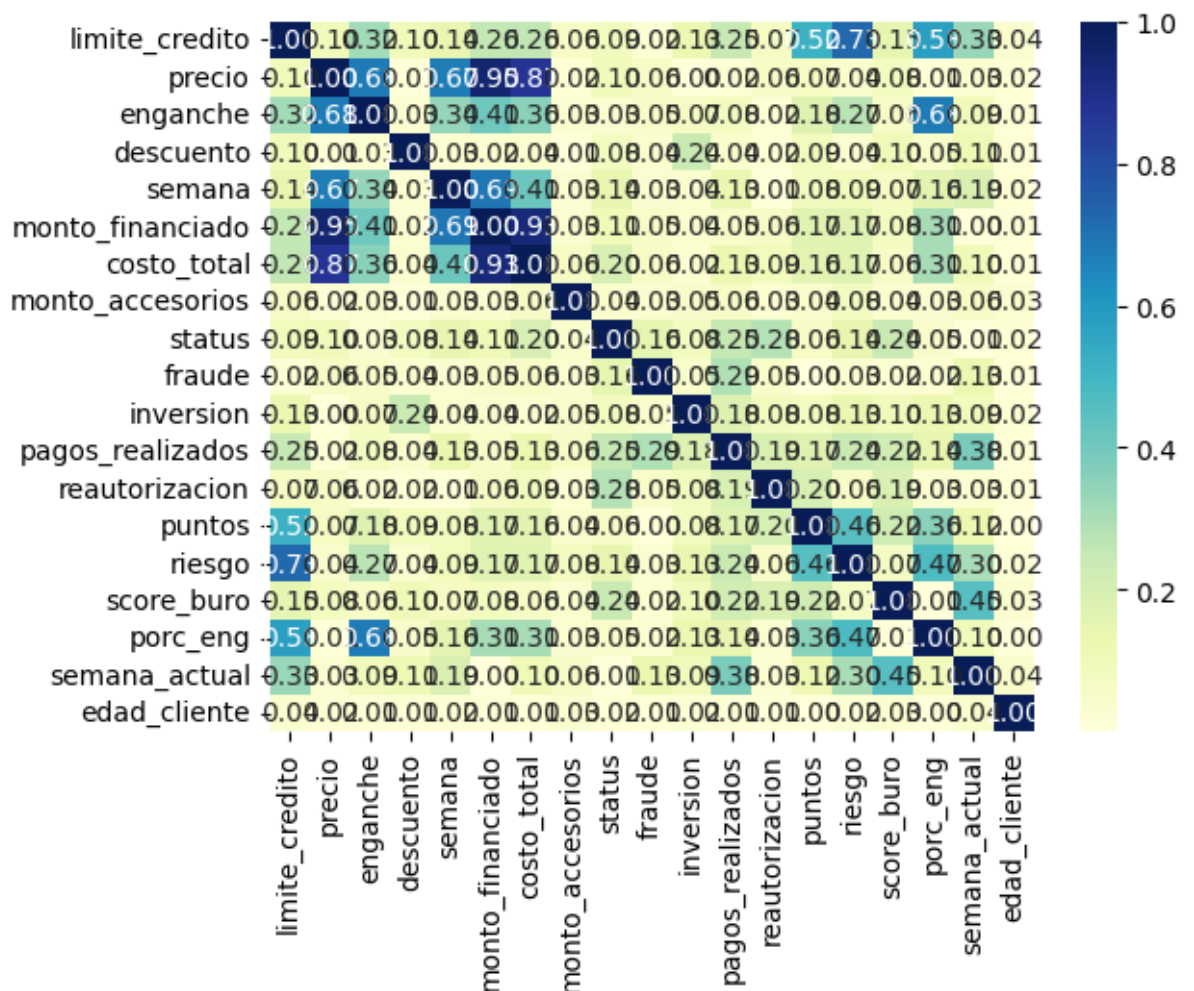


## Introducción

Durante la realización de la actividad de modelos de regresión no lineales se usaron distintas variables: riesgo, por\_eng, score\_buro, limite\_credito. Se intentó encontrar alguna relación no lineal, sin embargo esta fue una labor bastante difícil, pues al realizar un 'pairplot' se notaba una dispersión dentro de los datos ya sea o lineales o sin relación alguna, sin embargo aún así se decidió intentar y experimentar con distintas funciones pretendiendo encontrar algún comportamiento predecible.

Aunado a lo anterior, con apoyo de un mapa de calor, detectamos las correlaciones más fuertes entre variables.



- riesgo

Dentro del análisis de esta variable se utilizaron regresiones de tipo no lineal, cuadráticas y exponenciales en su mayoría, dichas regresiones se aplicaron a las variables riesgo, límite de crédito, porcentaje de enganche y calificación en el buró.

Como hallazgos tenemos que la relación más fuerte es entre la variable riesgo y la variable límite de crédito, representando una línea de predicción del .53 siendo un score débil, es decir, las predicciones que genere este modelo serán imprecisas en su mayoría.

El resto de modelos presentan variaciones en el score que se muestran anormales, como valores negativos y valores extremadamente grandes que salen del rango.

Como conclusión agregamos que los datos tienen un comportamiento lineal, por lo tanto no pueden ser analizados mediante regresiones no lineales.

- **por\_eng**

Para esta variable, se realizaron 7 modelos, entre ellos cuadráticos, cúbicos y con función de seno, la mayoría de ellos con la variable “riesgo” (ya que en el mapa muestra la segunda mayor significancia de 0.47), y en segundo lugar “costo total” (con una significancia de 0.31 en el mapa de calor). Sin embargo, ningún modelo fue lo suficientemente significativo (el “mejor” obtuvo un modelo de 0.223), esto, por la poca correlación que presentaban las variables desde un inicio.

- **score\_buro**

Se intentaron varios modelos, sin embargo ninguno fue realmente exitoso, la variable de score\_buro se usó como variable dependiente y se comparó con otras variables, exponiendo la función usada y  $r^2$  obtenida en la siguiente tabla:

Función	Variable Independiente	$R^2$
Tangencial	riesgo	0.000273007426971
Cúbica	porc_tasa	0.307943174356557
Cociente entre polinomios	porc_enganche	0.004616039155868
Inversa	monto_financiado	-0.00292118588380
Senoidal	monto_accesorios	-2.5263617418058715e-06

Se hizo un mapa de calor al finalizar y ver si realmente existía alguna relación, y como se vió en los resultados la relación era muy baja, coincidió que el mejor modelo que salió con una  $r^2$  más alta es el mismo en el que se indica mayor relación.

En lo que respecta a esta variable se puede concluir que dado que había muy poca relación entre las variables era difícil construir un buen modelo, además de que la distribución de los datos en esta variable era sin relación.

- **limite\_credito**

Directamente en la parte de limite\_credito, la variable que se seleccionó para el análisis fue la de Riesgo, esto ya que en el mapa de calor se encuentra explicada en un 70% nuestra variable por el riesgo (Existe correlación).

Como tal hubo 3 que considera para el análisis, esto tomando en cuenta que ninguno salió tan alto como tal, solo salieron 3 que se diferencian del resto por lo cual se puede definir que esos 3 son los mejores modelos (1 sobre los 3 sobresale también) pero para definirlo, antes se muestra la tabla de resultados

Función	Variable Independiente	$R^2$
Modelo Cuadrático	Riesgo	0.5722140006829264
Senoidal	Riesgo	0.0067525436487431145
Tangencial	Riesgo	0.0816378209083981
Valor Absoluto	Riesgo	0.5411064439198444
Cociente Entre Polinomios	Riesgo	0.6460066838492988

Como se puede observar, el valor absoluto, el modelo cuadrático y el cociente entre polinomios se diferencian exponencialmente del resto pero el que se puede determinar como el mejor modelo debido a su coeficiente de determinación es el cociente entre polinomios ya que su  $r^2$  es de 0.64 por lo que significa que el 64% de la variable está explicada por el modelo, como tal no es tan fuerte la correlación pero con la información brindada creo que es de lo mejor que se pudo sacar por lo tanto la fórmula sería:

00000173404516 x1, -0306135625 x2, 117.962451 x3, 17229.8278 x4,  
1472.13406 int