

Octubre 2025

Análisis exploratorio de
proyectos Datathon y FORVIA

REGRESIÓN LOGÍSTICA

FORVIANOS.py

FORVIANOS.PY



Maria Matanzo

A01737554



Jorge Cortes

A01736236



Marco Cornejo

A01276411



Eduardo Torres

A01734935



Laisha Puan

A01736397

OBJETIVOS

- **Analizar 5** casos de correlación logística que existe entre diferentes variables de nuestra base de datos
- **Aplicar** el ajuste de correlación logística que existe entre las variables dicotómicas de nuestra base de datos
- **Convertir** las variables **que sean necesarias** en variables de tipo dicotómica con las categorías que se consideren pertinentes, aplicando la herramienta de “**Regresión Logística**”

METODOLOGÍA

Preparación de los datos



01

Primero seleccionamos 5 variables de nuestro Data

02

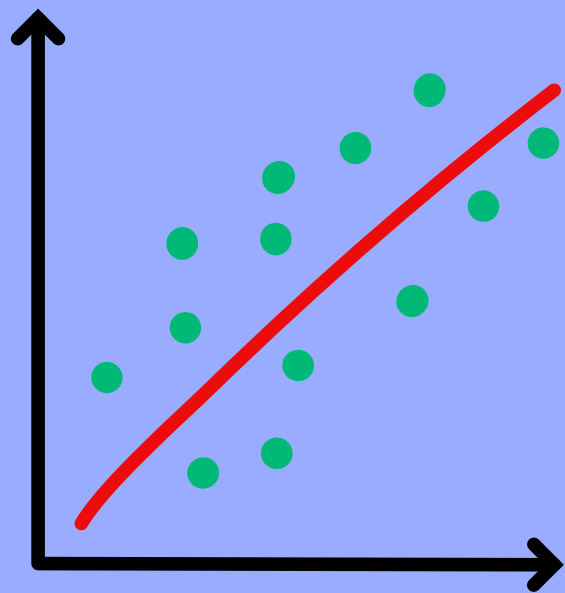
Posteriormente las transformamos a dicotómicas, obteniendo dos clases diferentes, definiendo rangos según su frecuencia.

03

???

METODOLOGÍA

Regresión Lineal



01

Cálculo de correlaciones.

02

Visualización mediante heatmap.

METODOLOGÍA

Regresión No Lineal



01

Construcción de modelos de correlación No lineal de las variables.

02

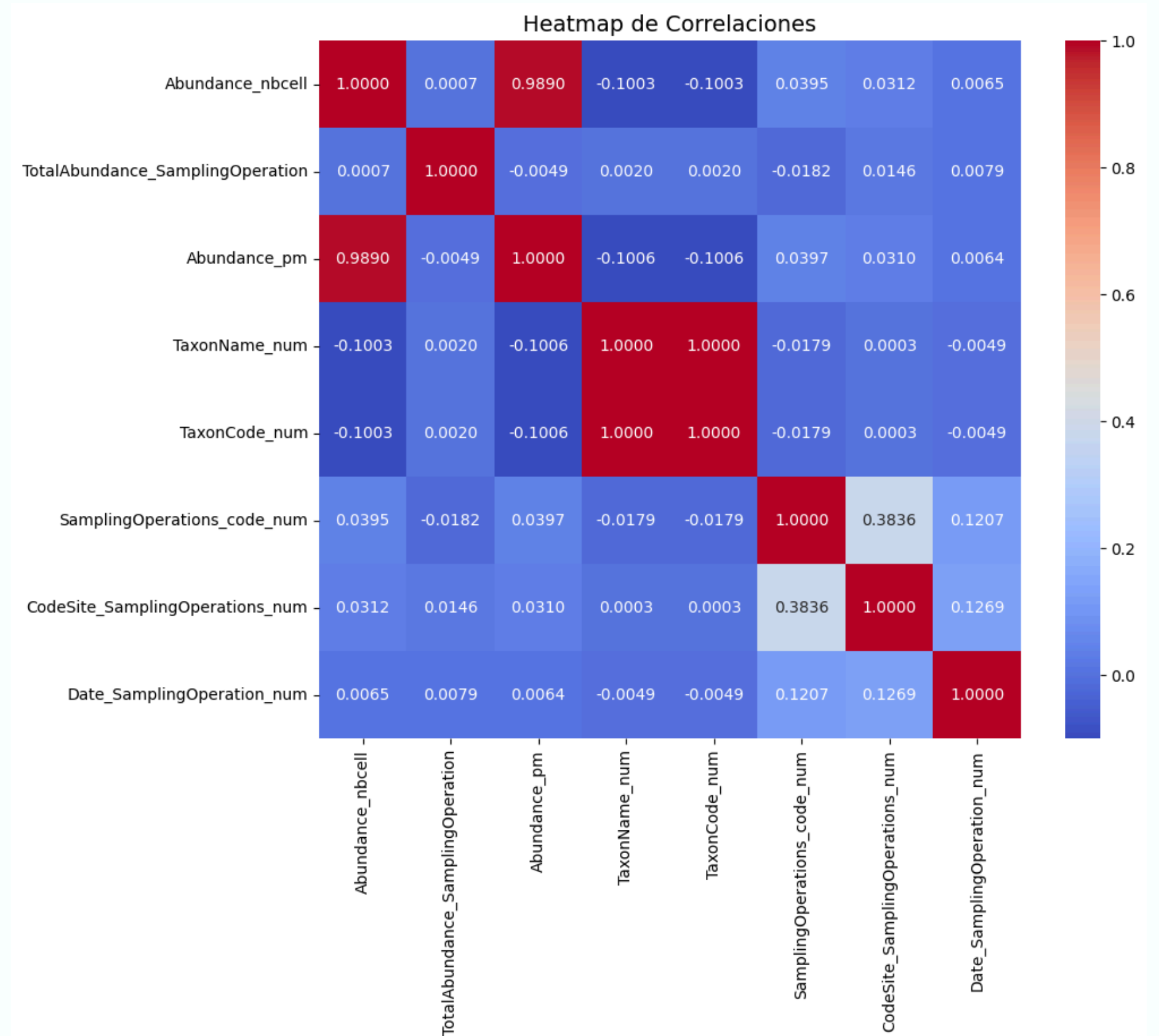
Analizar la correlación y comparar.

ACTIVIDAD 4.1

forvianos.py

MODELO PREDICTIVO

- **Gráfica:** Heatmap de correlaciones
- **Hallazgos:**
 - Solo existen dos pares de variables altamente correlacionados (abundancias y taxones).
 - La mayoría de las variables son independientes.



MODELO TAXON NAME

Variable dependiente (Y):

- TaxonName

Variables independientes (X):

- Abundance_nbcell

correlación heatmap

- -0.1003

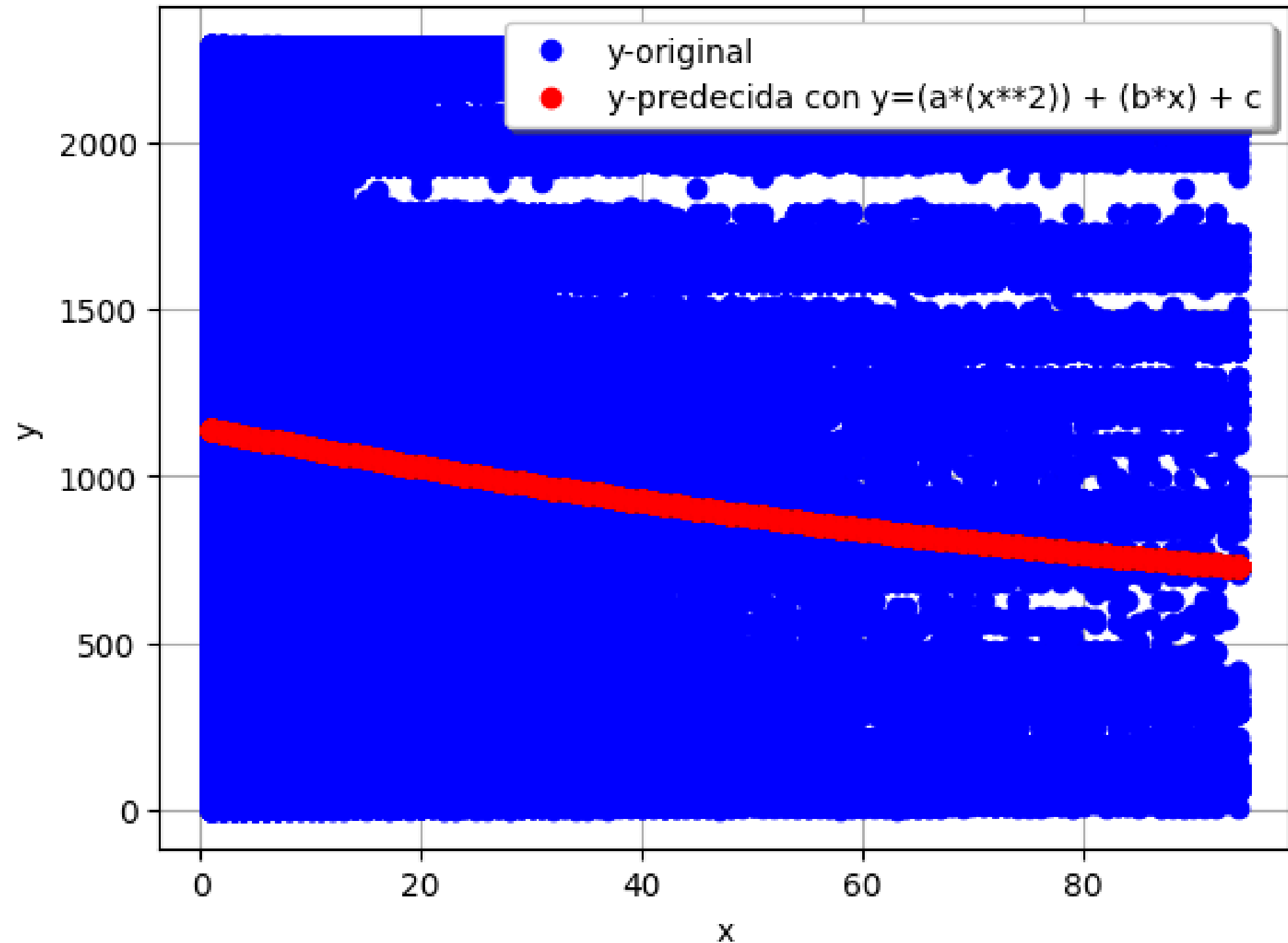
correlación modelo no lineal

- 0.1007

correlación modelo lineal

múltiple:

- 0.1016



MODELO SAMPLING OPERATIONS CODE

Variable dependiente (Y):

- SamplingOperations_code

Variables independientes (X):

- CodeSite_SamplingOperations

correlacion heatmap:

- 0.3836

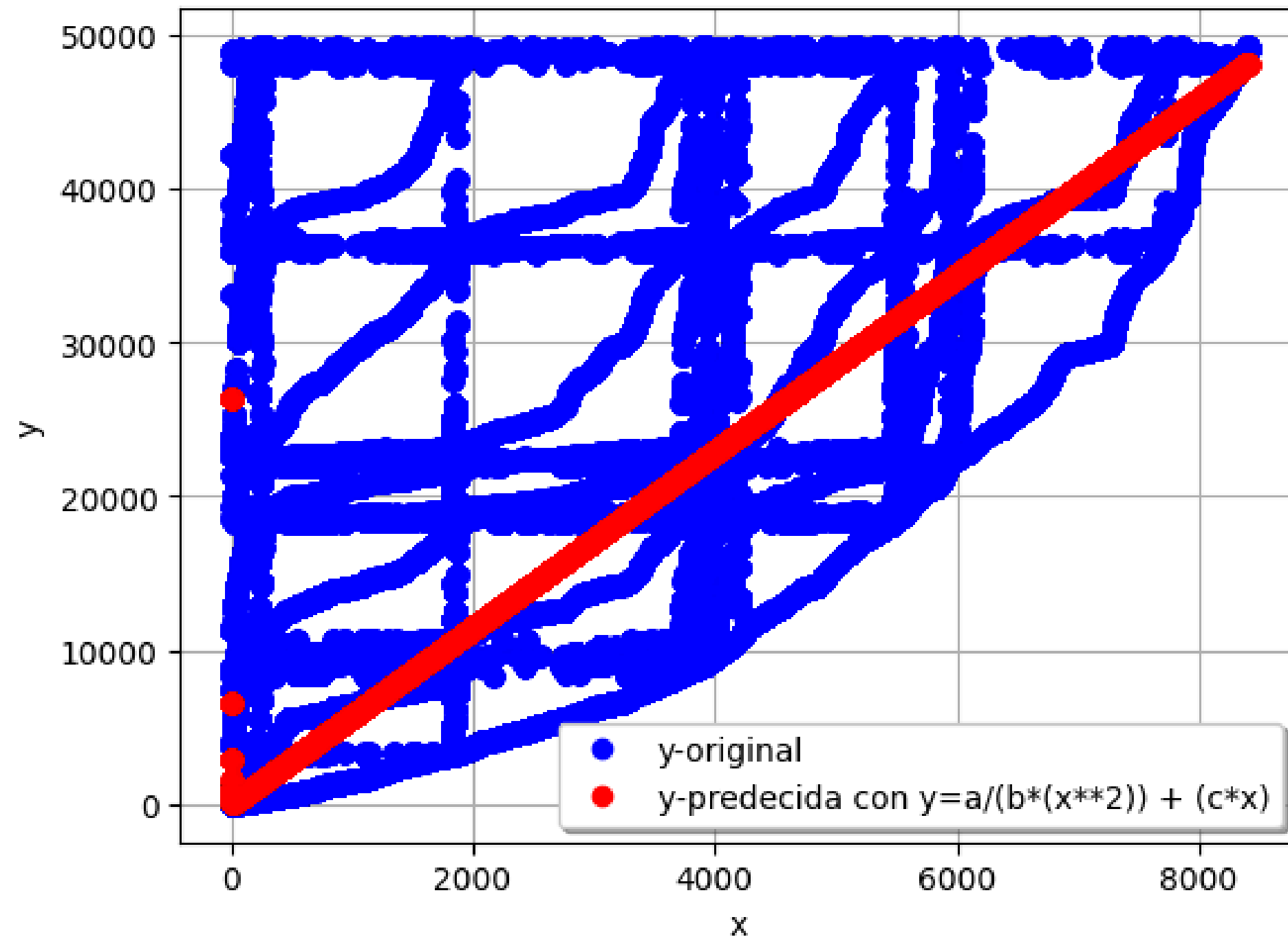
correlacion modelo no lineal:

- 0.4595

correlación modelo lineal

múltiple:

- 0.3913



MODELO CODESITE SAMPLING OPERATIONS

Variable dependiente (Y):

- CodeSite_SamplingOperations

Variables independientes (X):

- SamplingOperations_code

correlación heatmap:

- 0.3836

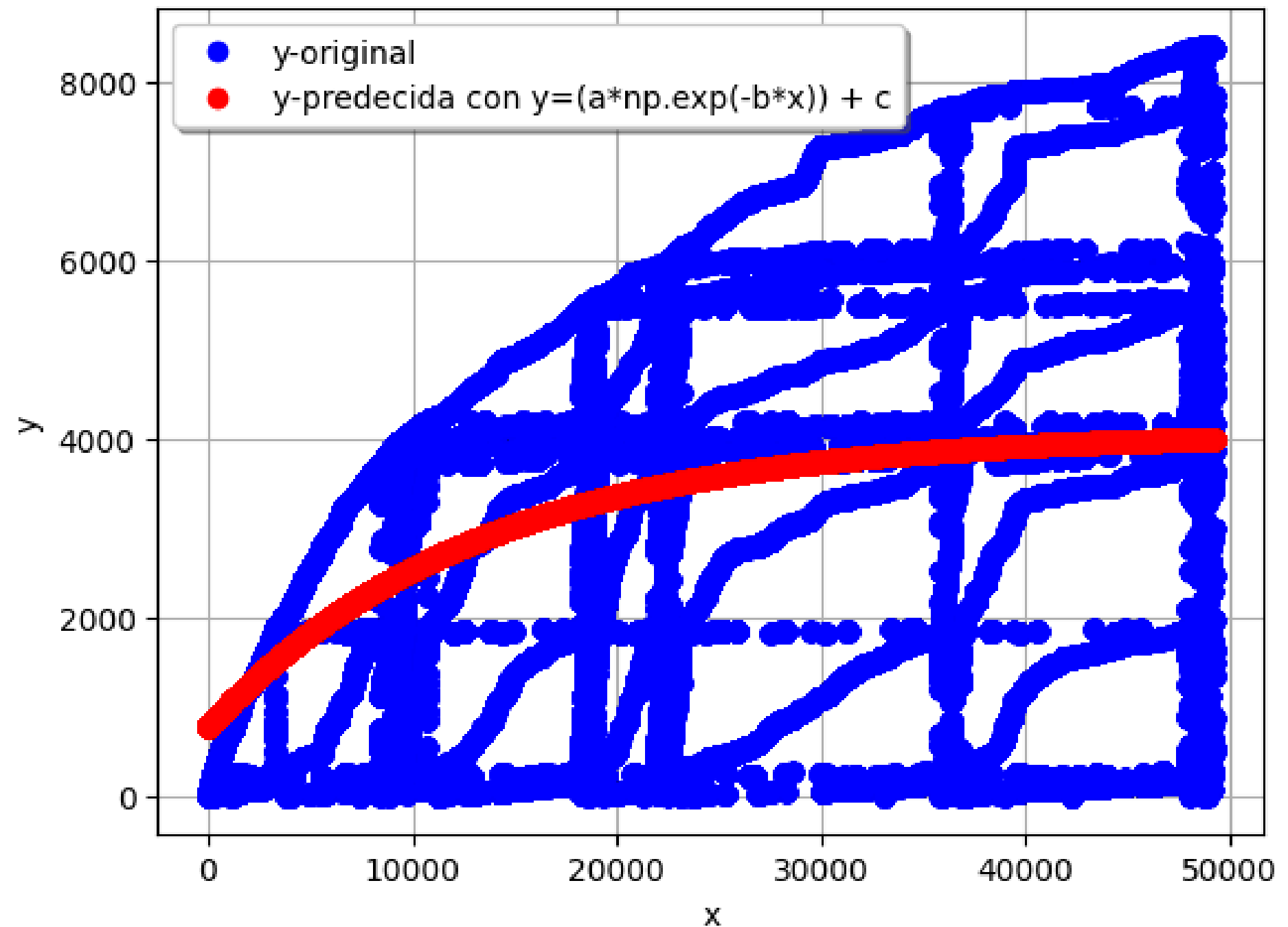
correlación modelo no lineal:

- 0.4113

correlación modelo lineal

múltiple:

- 0.3924



MODELO DATE SAMPLING OPERATION

Variable dependiente (Y):

- Date_SamplingOperation

Variables independientes (X):

- SamplingOperations_code

correlación heatmap:

- 0.1269

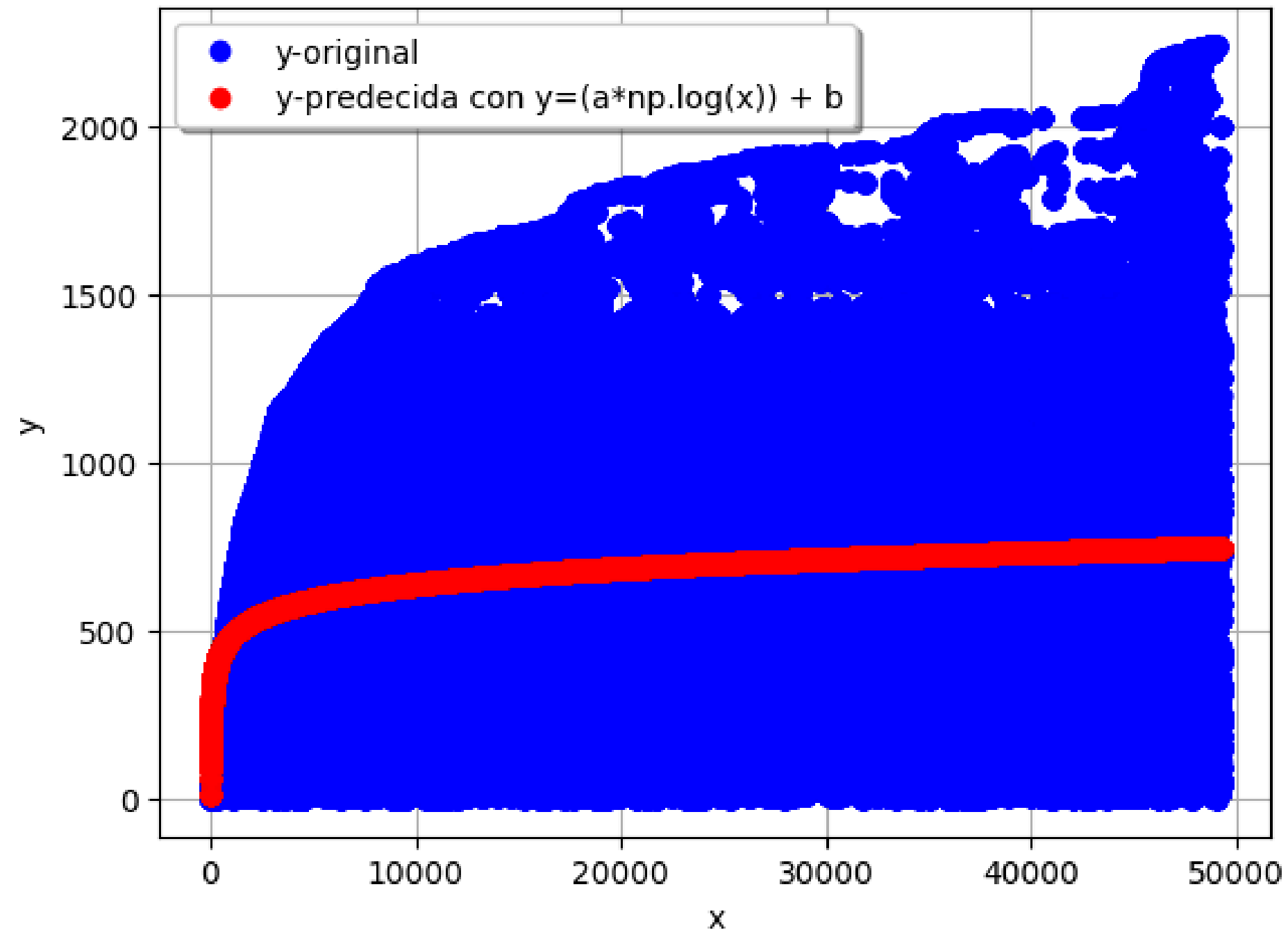
correlación modelo no lineal:

- 0.1580

correlación modelo lineal

múltiple:

- 0.1491



MODELO ABUNDANCE NBCELL

Variable dependiente (Y):

- Abundance_nbcell

Variables independientes (X):

- TaxonName

correlación heatmap:

- -0.1003

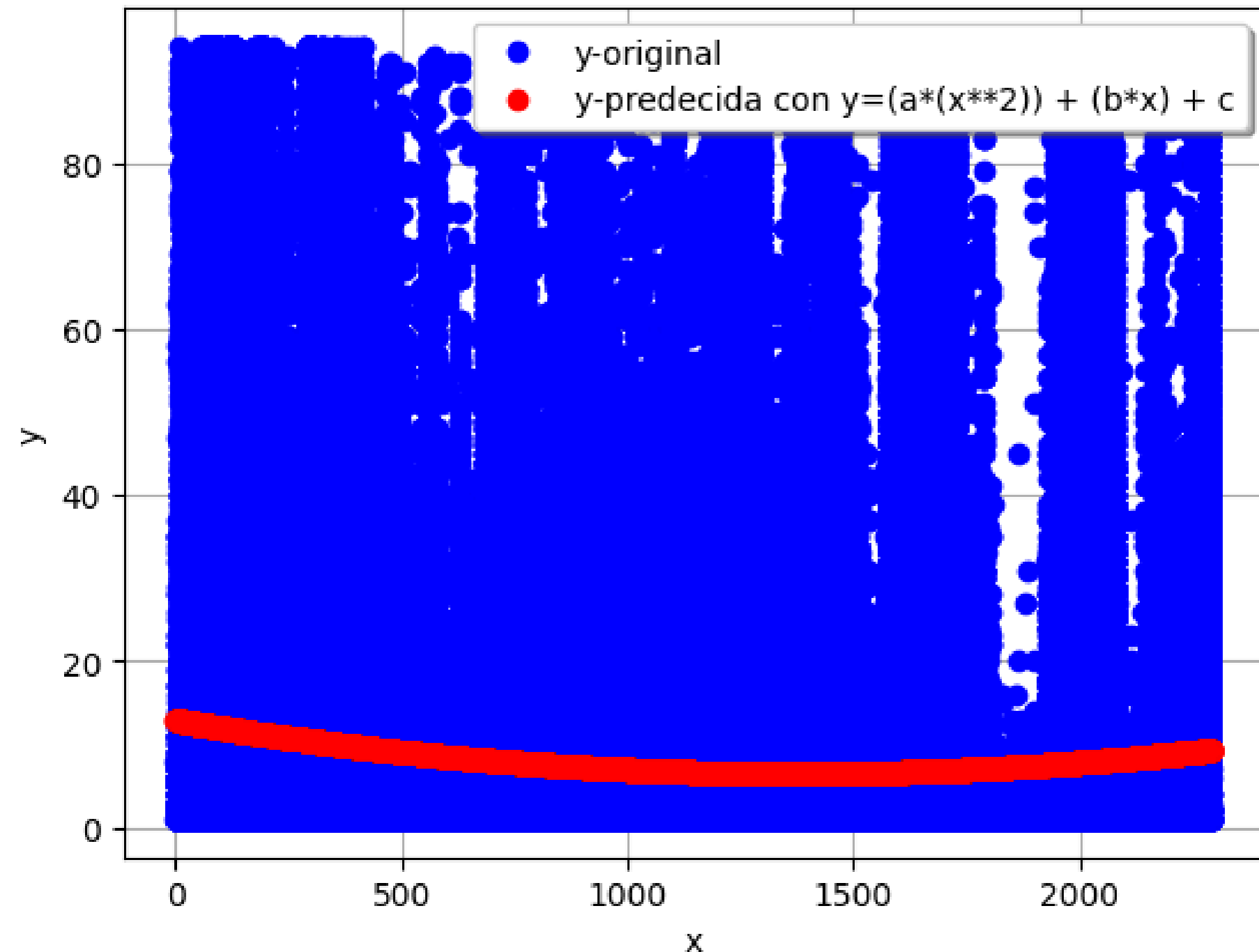
correlación modelo no lineal:

- 0.1431

correlación modelo lineal

múltiple:

- 0.1071



MODELO TOTAL ABUNDANCE SAMPLING OPERATION

Variable dependiente (Y):

- TotalAbundance_SamplingOperation

Variables independientes (X):

- SamplingOperations_code

correlación heatmap:

- -0.0182

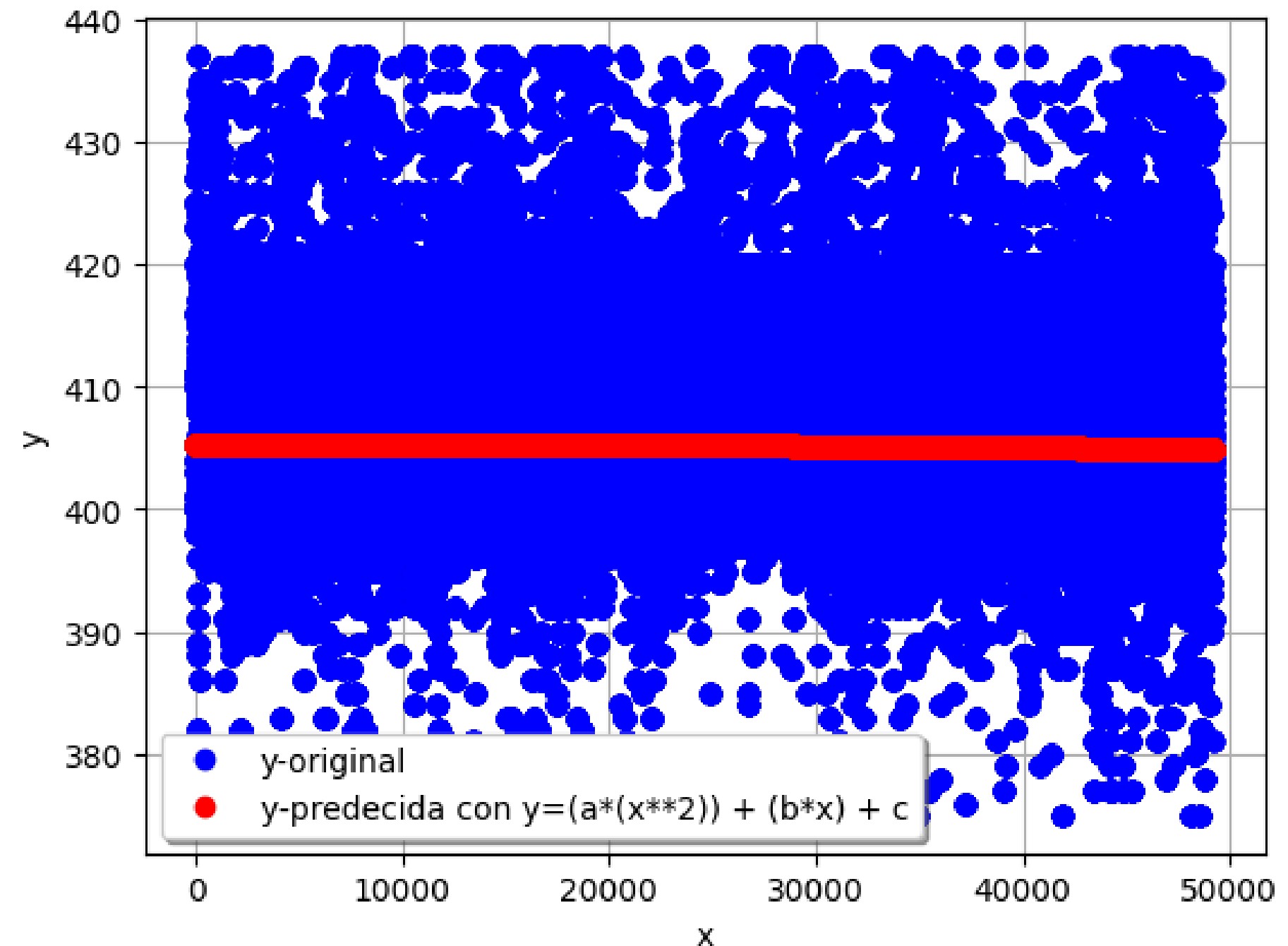
correlación modelo no lineal:

- 0.0187

correlación modelo lineal

múltiple:

- 0.0306



MODELO ABUNDANCE P M

Variable dependiente (Y):

- Abundance_pm

Variables independientes (X):

- TaxonName

correlación heatmap

- -0.1006

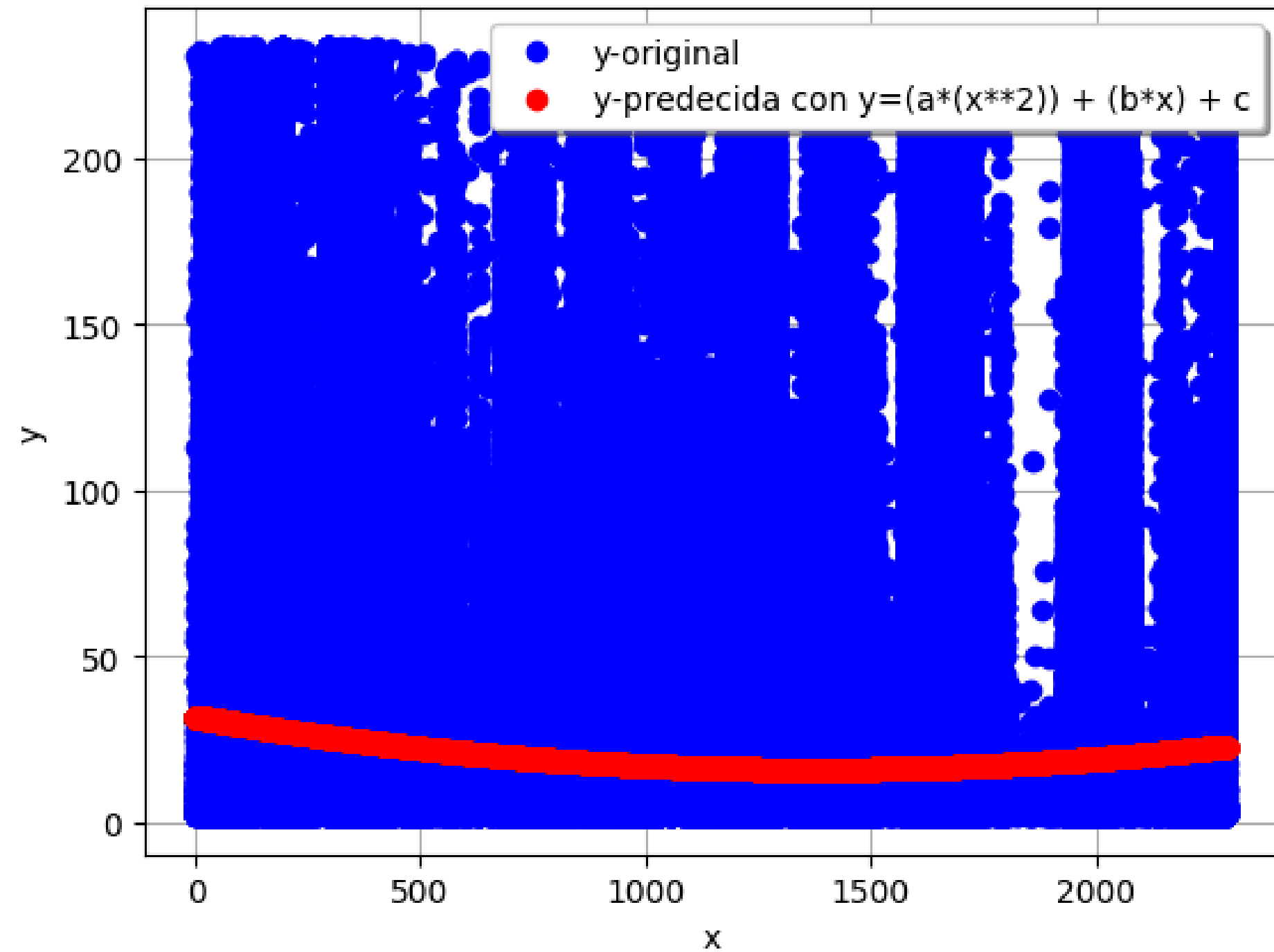
correlación modelo no lineal:

- 0.1434

correlación modelo lineal

múltiple:

- 0.1075



COEFICIENTES DE DETERMINACIÓN Y CORRELACIÓN

	Variable_dependiente	Variable_independiente	r	r	R2
18	TaxonName_num	TaxonCode_num	1.000000	1.000000	1.000000
1	Abundance_nbcell	Abundance_pm	0.989024	0.989024	0.978168
25	SamplingOperations_code_num	CodeSite_SamplingOperations_num	0.383587	0.383587	0.147139
27	CodeSite_SamplingOperations_num	Date_SamplingOperation_num	0.126912	0.126912	0.016107
26	SamplingOperations_code_num	Date_SamplingOperation_num	0.120650	0.120650	0.014557
14	Abundance_pm	TaxonCode_num	-0.100609	0.100609	0.010122
13	Abundance_pm	TaxonName_num	-0.100609	0.100609	0.010122
2	Abundance_nbcell	TaxonName_num	-0.100268	0.100268	0.010054
3	Abundance_nbcell	TaxonCode_num	-0.100268	0.100268	0.010054
15	Abundance_pm	SamplingOperations_code_num	0.039678	0.039678	0.001574

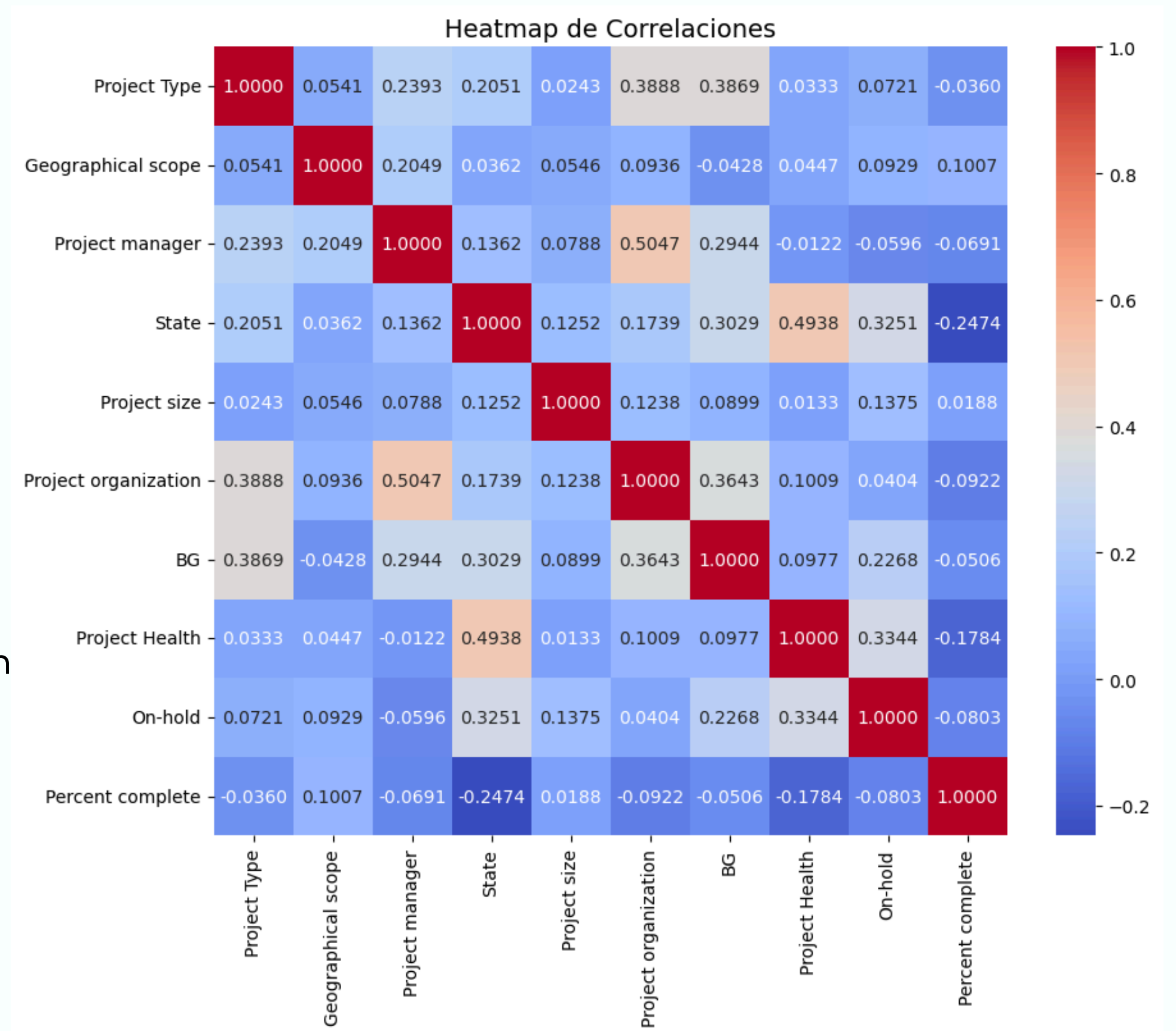
- Solo dos pares de variables muestran correlaciones fuertes o perfectas, lo que indica posibles redundancias.
- El resto de las variables son independientes, lo cual es positivo para análisis multivariado.

ACTIVIDAD 3.2

forvianos.py

MODELO PREDICTIVO

- **Gráfica:** Heatmap de correlaciones
- **Hallazgos:**
 - El heatmap revela que las variables más interconectadas son las administrativas (manager, organización, tipo), mientras que las métricas de desempeño son más independientes.
 - La baja correlación de Percent complete con otras métricas indica que el avance de un proyecto no siempre refleja su salud o eficiencia, por lo que podrían considerarse indicadores adicionales para evaluar desempeño real.



MODELO PROJECT MANAGER

Variable dependiente (Y):

- Project manager

Variables independientes (X):

- Project organization

correlación heatmap:

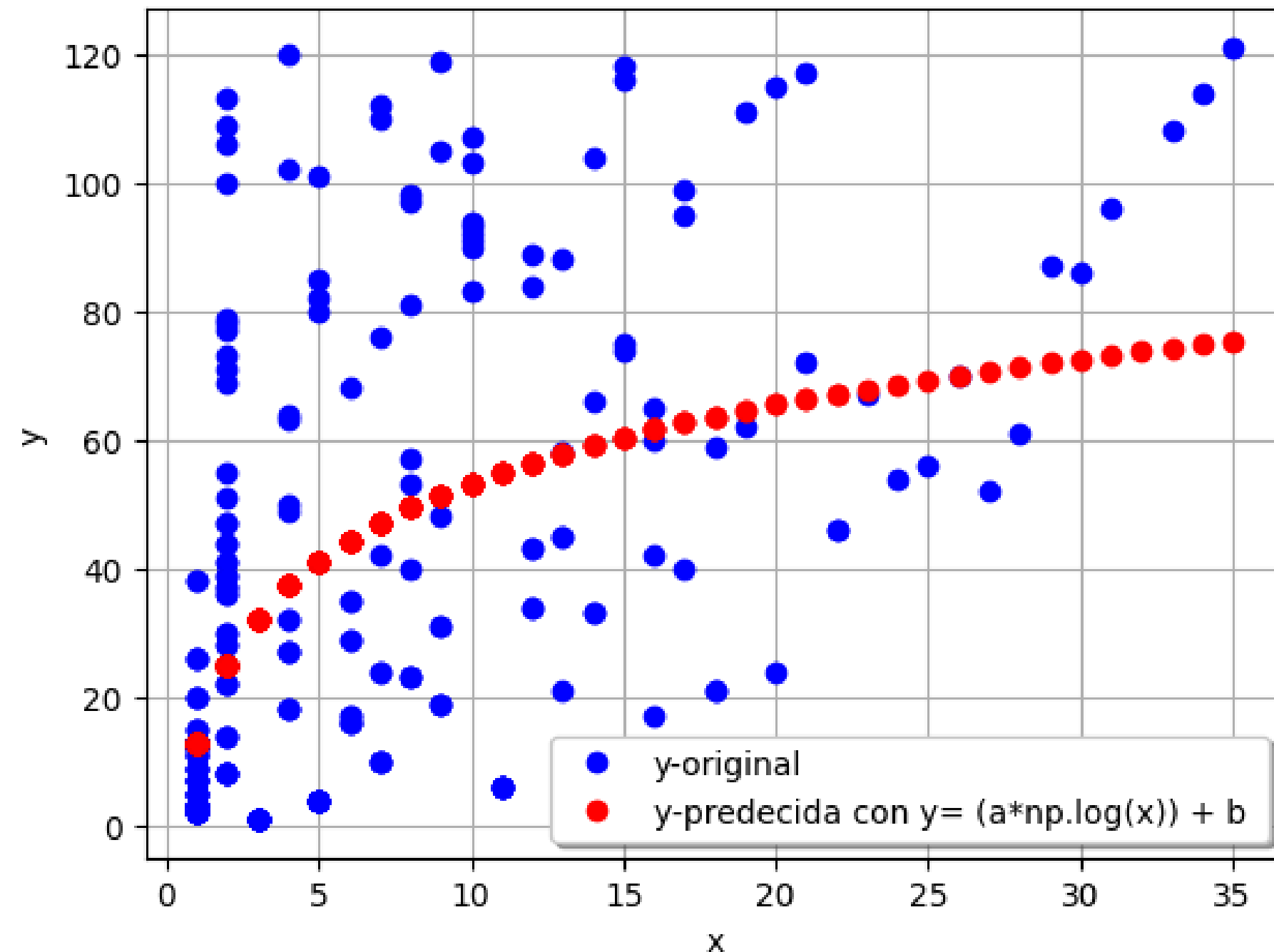
- 0.5022

correlación modelo no lineal:

- 0.5371

correlación lineal múltiple:

- 0.5216



MODELO PROJECT HEALTH

Variable dependiente (Y):

- Project Health

Variables independientes (X):

- State

correlación heatmap:

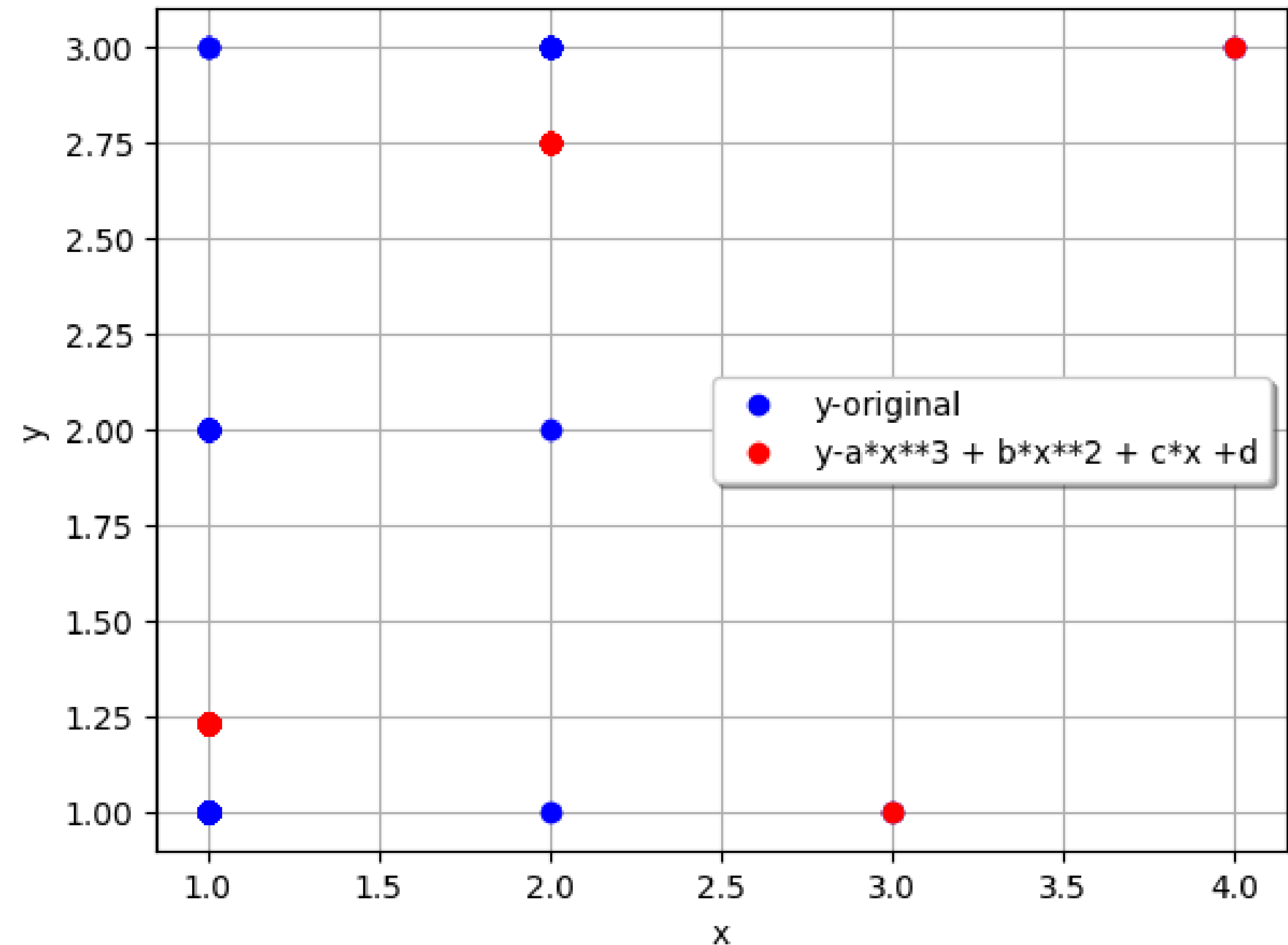
- 0.4938

correlación modelo no lineal:

- 0.6009

correlación lineal múltiple:

- 0.5455



MODELO BG

Variable dependiente (Y):

- BG

Variables independientes (X):

- Project Type

correlación heatmap:

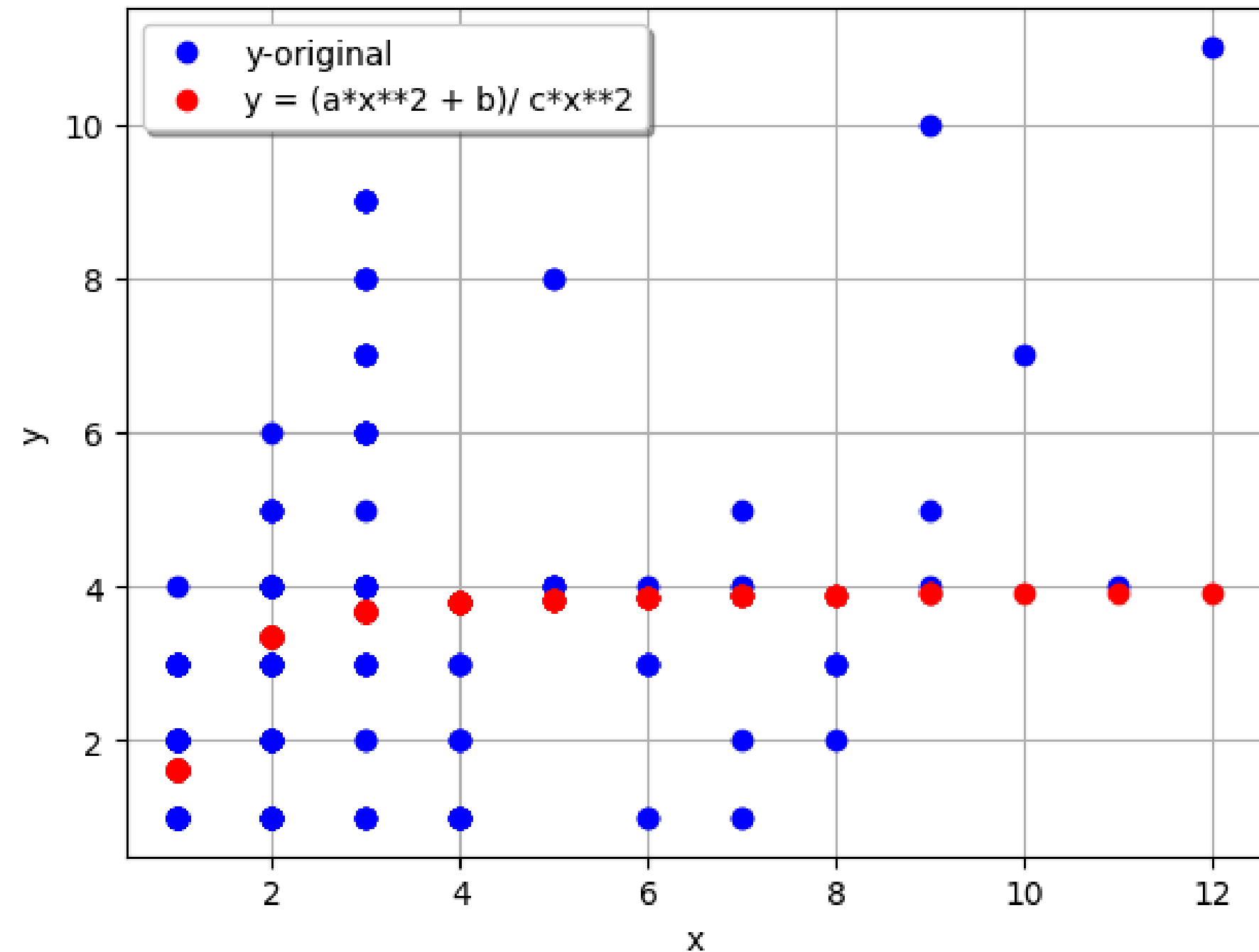
- 0.3869

correlación modelo no lineal:

- 0.4613

correlación lineal múltiple:

- 0.4994



MODELO ON HOLD

Variable dependiente (Y):

- On_hold

Variables independientes (X):

- Project Health

correlación heatmap:

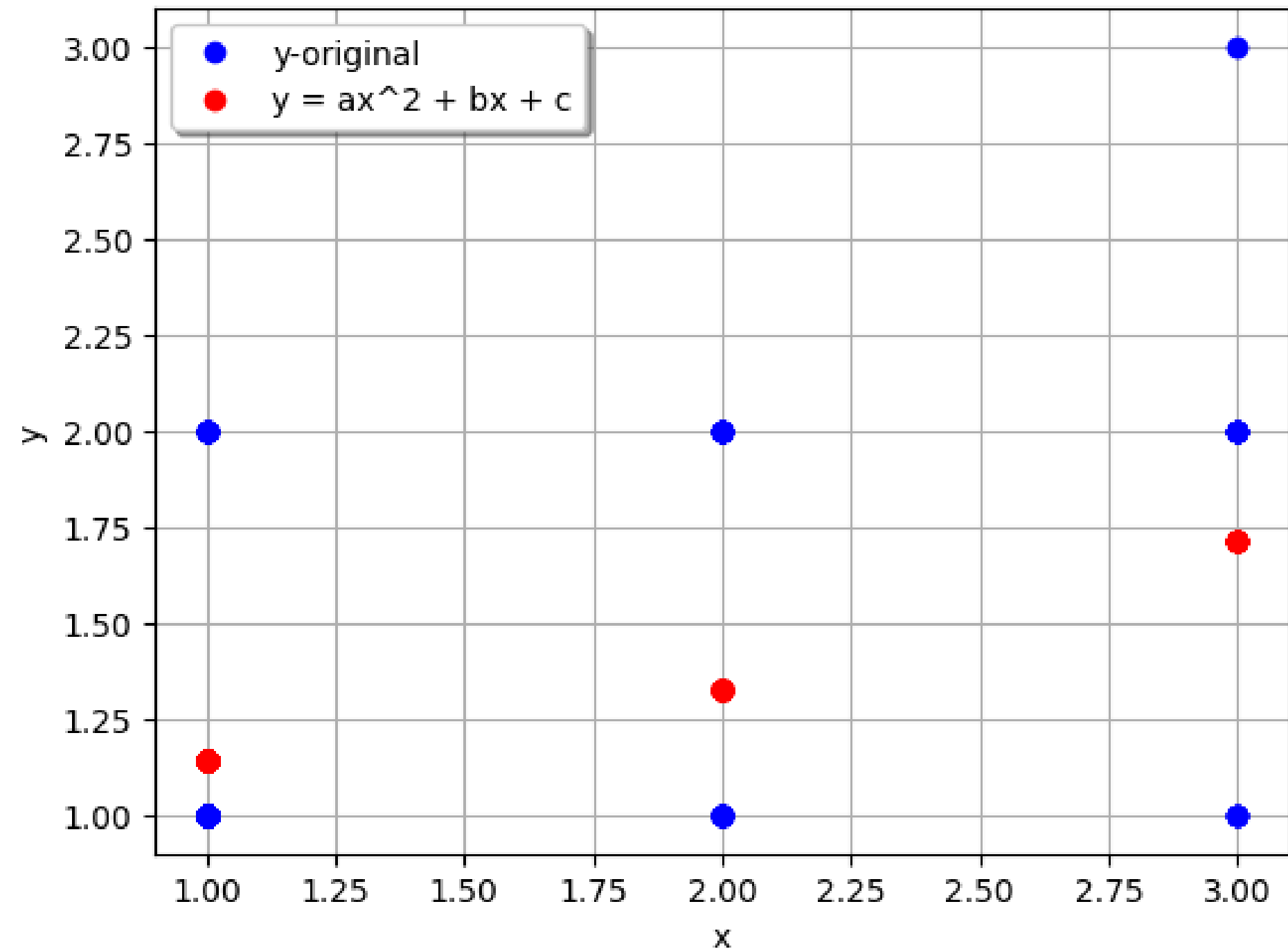
- 0.3343

correlación modelo no lineal:

- 0.3433

correlación lineal múltiple:

- 0.4091



MODELO PERCENT COMPLETE

Variable dependiente (Y):

- Percent complete

Variables independientes (X):

- State

correlación heatmap:

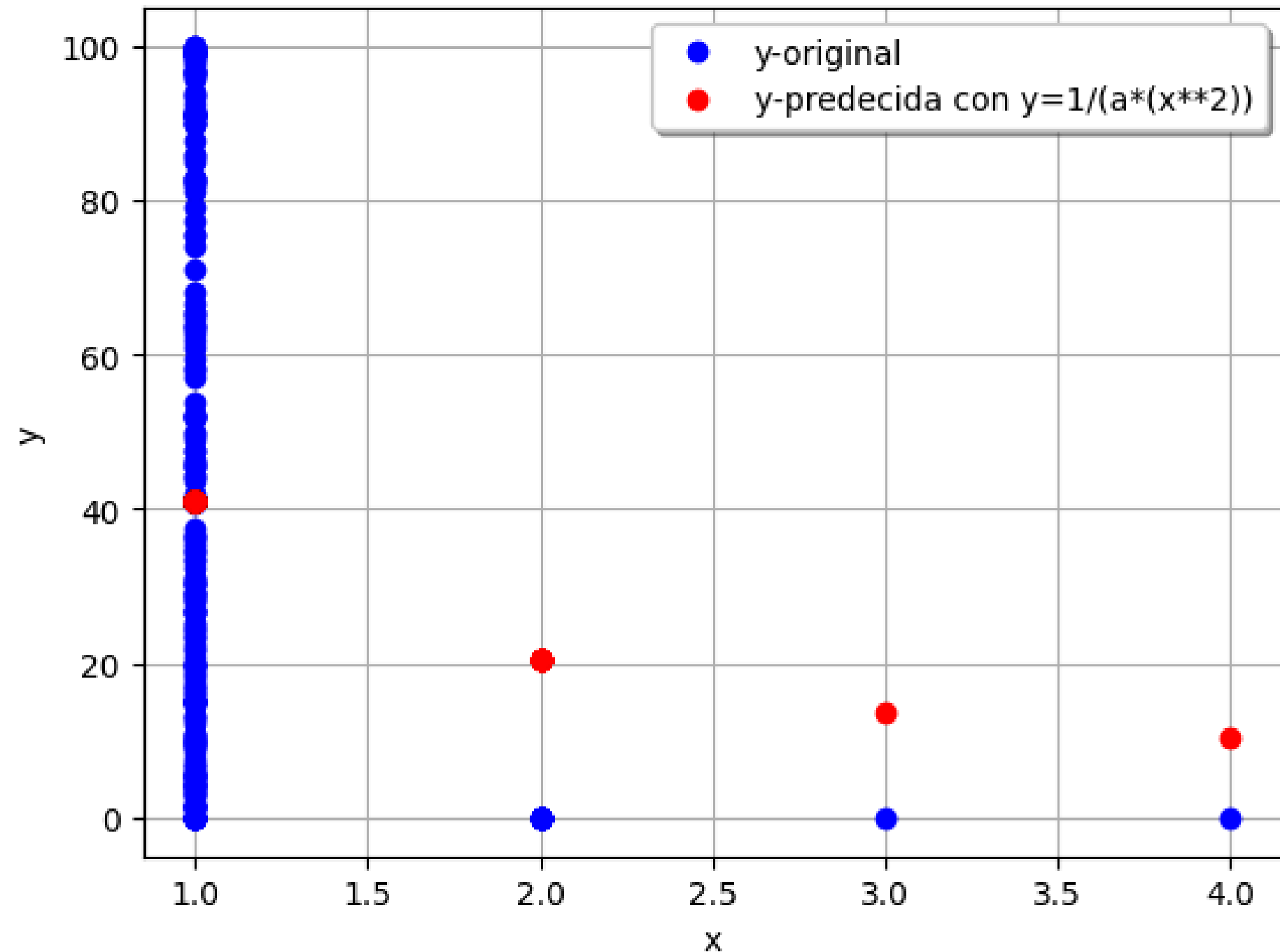
- -0.2474

correlación modelo no lineal:

- 0.2390

correlación lineal múltiple:

- 0.2597



COEFICIENTES DE DETERMINACIÓN Y CORRELACIÓN

	Variable_dependiente	Variable_independiente	r	r	R2
17	Project manager	Project organization	0.502209	0.502209	0.252214
24	State	Project Health	0.493797	0.493797	0.243835
5	Project Type	BG	0.386941	0.386941	0.149723
4	Project Type	Project organization	0.375739	0.375739	0.141180
30	Project organization	BG	0.369726	0.369726	0.136697
35	Project Health	On-hold	0.334380	0.334380	0.111810
25	State	On-hold	0.325096	0.325096	0.105687
1	Project Type	Project manager	0.304068	0.304068	0.092457
23	State	BG	0.302922	0.302922	0.091762
18	Project manager	BG	0.277785	0.277785	0.077164

- No hay una variable dominante: Ninguna combinación muestra fuerza suficiente para construir un modelo confiable con solo estas variables.
- Las correlaciones más altas involucran gestión, estado y salud de proyectos, lo que puede reflejar prácticas internas o políticas de seguimiento.

CONCLUSIONES

Los datos no siguen una sola regla ni dependen de una sola variable. Aunque probamos modelos no lineales que mejoraron un poco las correlaciones, no fue suficiente para hacer predicciones fuertes. Esto nos demuestra que los proyectos y los datos reales son más complejos y dependen de varios factores al mismo tiempo. Por eso, para entenderlos bien, necesitamos usar análisis más completos y no solo fijarnos en una relación.

Octubre 2030

MUCHAS
GRACIAS

Bruno Lago