

# REGRESIÓN NO LINEAL

DATAFORGE

10 OCT, 2025



# TEAM MEMBERS (DATAFORGE)



**JESÚS EDUARDO VALLE  
VILLEGAS**

FINANZAS  
A01770616



**DIEGO ANTONIO OROPEZA  
LINARTE**

BGB  
A01733018



**MANUEL EDUARDO  
COVARRUBIAS RODRÍGUEZ**

ITC  
A01737781



**ITHANDEHUI JOSELYN  
ESPINOZA**

ITC  
A01734547



**MAURICIO GRAU GUTIERREZ  
RUBIO**

LEM  
A01734914



# OBJETIVO

Analizar la relación entre las variables **TaxonName**, **TaxonCode**, **SamplingOperations\_code**, **CodeSite\_SamplingOperations**, **Date\_SamplingOperation**, **Abundance\_nbcell**, **TotalAbundance\_SamplingOperation** y **Abundance\_pm** del conjunto de datos **01\_DiatomInventories\_GTstudentproject\_B.csv**, aplicando y comparando dos modelos de regresión no lineal para determinar el grado de correlación y la capacidad explicativa de cada modelo mediante los coeficientes de determinación ( $R^2$ ) y correlación.



# METODOLOGÍA

Dataset1: [01\\_DiatomInventories\\_GTstudentproject\\_B.csv](#)  
Dataset2: [proyectos\\_forvia.csv](#)

## LIMPIEZA Y PREPARACIÓN

Revisión de estructura y nulos; codificación numérica de [TaxonName](#), [TaxonCode](#), [SamplingOperations\\_code](#), [CodeSite\\_SamplingOperations](#); [Date\\_SamplingOperation](#) → numérico (ordinal/continuo).

## VARIABLES ANALIZADAS

[TaxonName](#), [TaxonCode](#), [SamplingOperations\\_code](#), [CodeSite\\_SamplingOperations](#), [Date\\_SamplingOperation](#), [Abundance\\_nbcell](#), [TotalAbundance\\_SamplingOperation](#), [Abundance\\_pm](#) (combinaciones diversas).

## MODELOS NO LINEALES:

- Polinómico (grados 2–3) para capturar curvatura.
- Exponencial/Potencial para explorar patrones de crecimiento/decrecimiento.

## ANÁLISIS COMPARATIVO DE RESULTADOS

- Evaluación del ajuste: se calcularon  $R^2$  y correlaciones (Pearson/Spearman según el tipo de variable) para medir fuerza y dirección de la relación; además, se revisaron residuales para verificar la adecuación del modelo y detectar desviaciones.
- Comparativo de resultados: se elaboró una tabla resumen con  $R^2$  y correlaciones por cada relación evaluada, identificando el mejor desempeño.



# Transformacion de variables

Mapeo con un ciclo for

Index	TaxonName_num	TaxonCode_num	SamplingOperations_code_num	CodeSite_SamplingOperations_num	Date_SamplingOperation_num
0	1	1	1	1	1
1	1	1	2	2	2
2	2	2	2	3	3
3	2	2	3	4	4
4	2	2	4	5	5
5	2	2	5	6	6
6	2	2	6	7	7
7	2	2	7	8	8
8	2	2	8	9	9
9	2	2	9	10	10

Fue necesario transformar las variables categóricas en variables numéricas. Para ello, se utilizó la jerarquía de frecuencias, asignando valores más bajos a las categorías con mayor frecuencia de aparición.



# Transformacion de variables

## Mapeo con un ciclo for

Index	Project Type	Geographical scope	Project manager	State	Project size	Project Org	BG	Project Health	On-Hold
0	1	63	2	1	3	1	1	1	1
1	1	62	15	1	2	1	2	2	2
2	1	51	20	1	1	1	2	2	1
3	1	51	15	1	3	1	2	1	2
4	1	61	2	1	1	1	2	1	1
...	...	...	...	...	...	...	...	...	...
241	6	43	119	1	2	9	3	1	1
242	8	126	27	1	1	4	3	1	1
243	8	42	27	1	1	4	3	1	1
244	1	42	120	1	3	4	3	1	1
245	12	127	121	4	4	35	11	3	3



HEATMAP COMPLETO - Matriz de Correlación de Todas las Variables



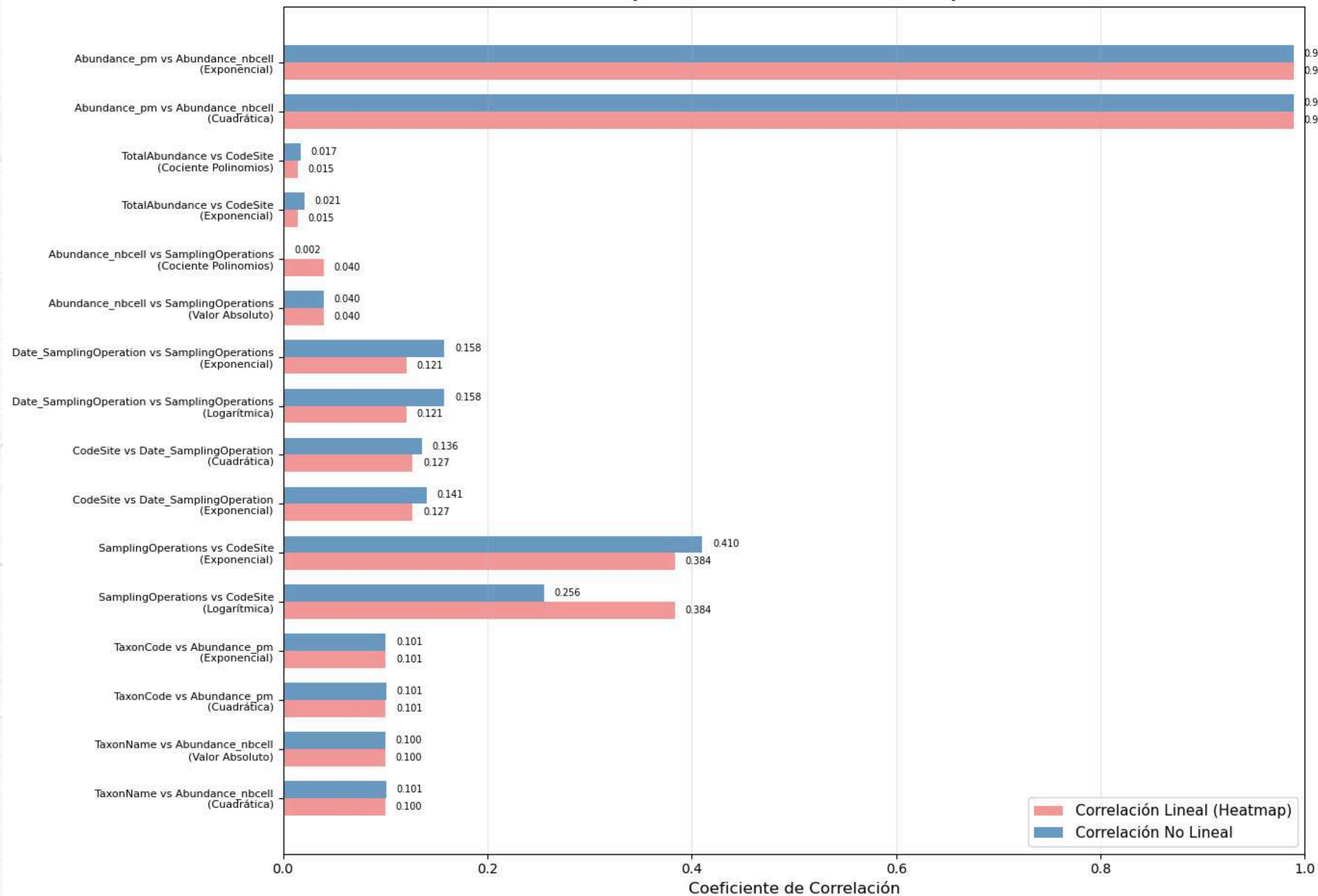


# ACTIVIDAD 3.1

DATATHON



## Comparación Completa: Correlación Lineal vs No Lineal (Todos los Modelos Analizados)

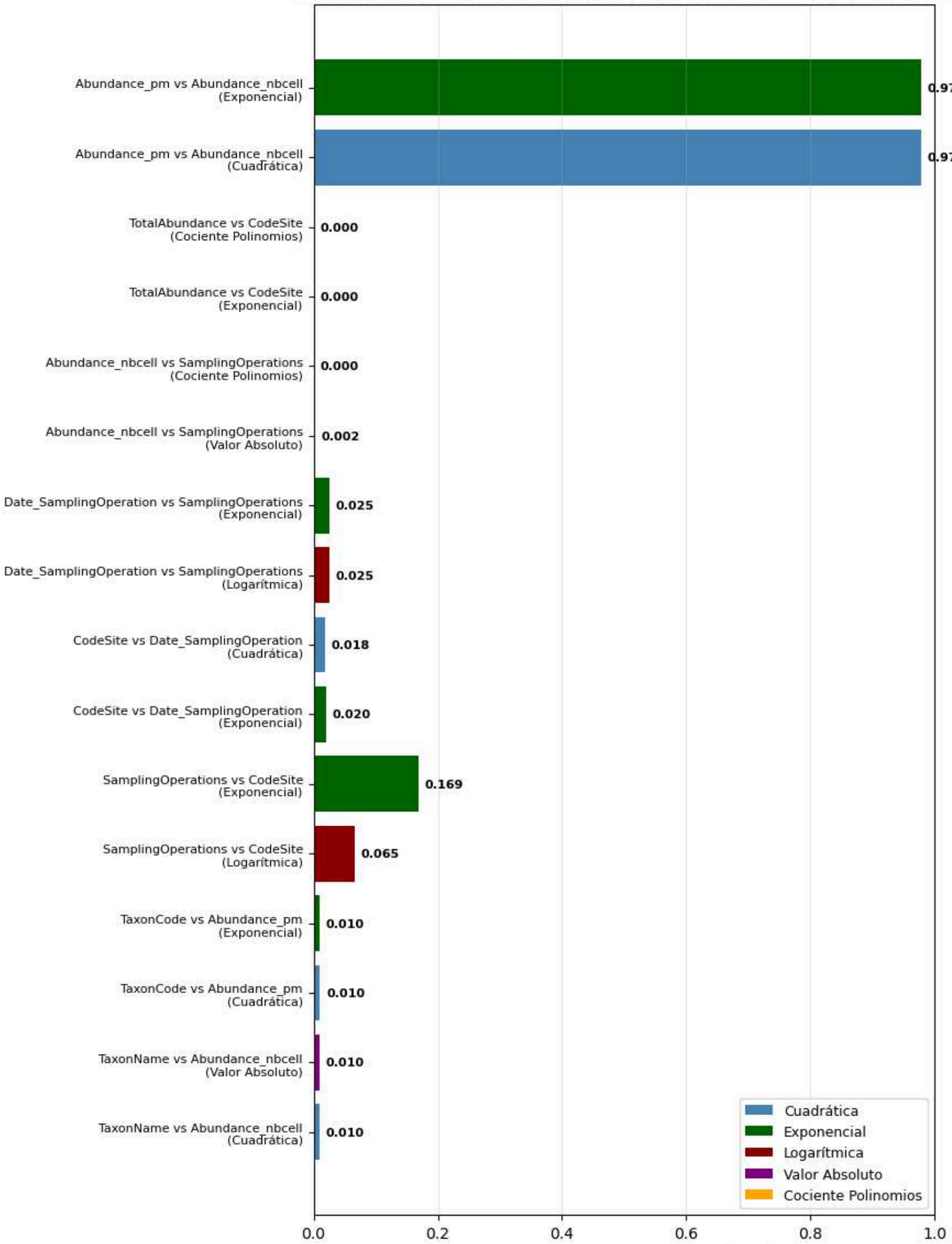




Total de modelos analizados:	16
Correlación lineal promedio:	0.2344
Correlación no lineal promedio:	0.2325
Modelos que mejoraron:	11/16 (68.8%)
TOP 3 MODELOS CON MAYOR MEJORA:	<div>1. Date_SamplingOperation vs SamplingOperations (Logarítmica) Lineal: 0.1207 → No Lineal: 0.1580 (Mejora: +0.0373)</div> <div>2. Date_SamplingOperation vs SamplingOperations (Exponencial) Lineal: 0.1207 → No Lineal: 0.1580 (Mejora: +0.0373)</div> <div>3. SamplingOperations vs CodeSite (Exponencial) Lineal: 0.3836 → No Lineal: 0.4105 (Mejora: +0.0269)</div>



Coefficientes de Determinación - TODOS LOS MODELOS

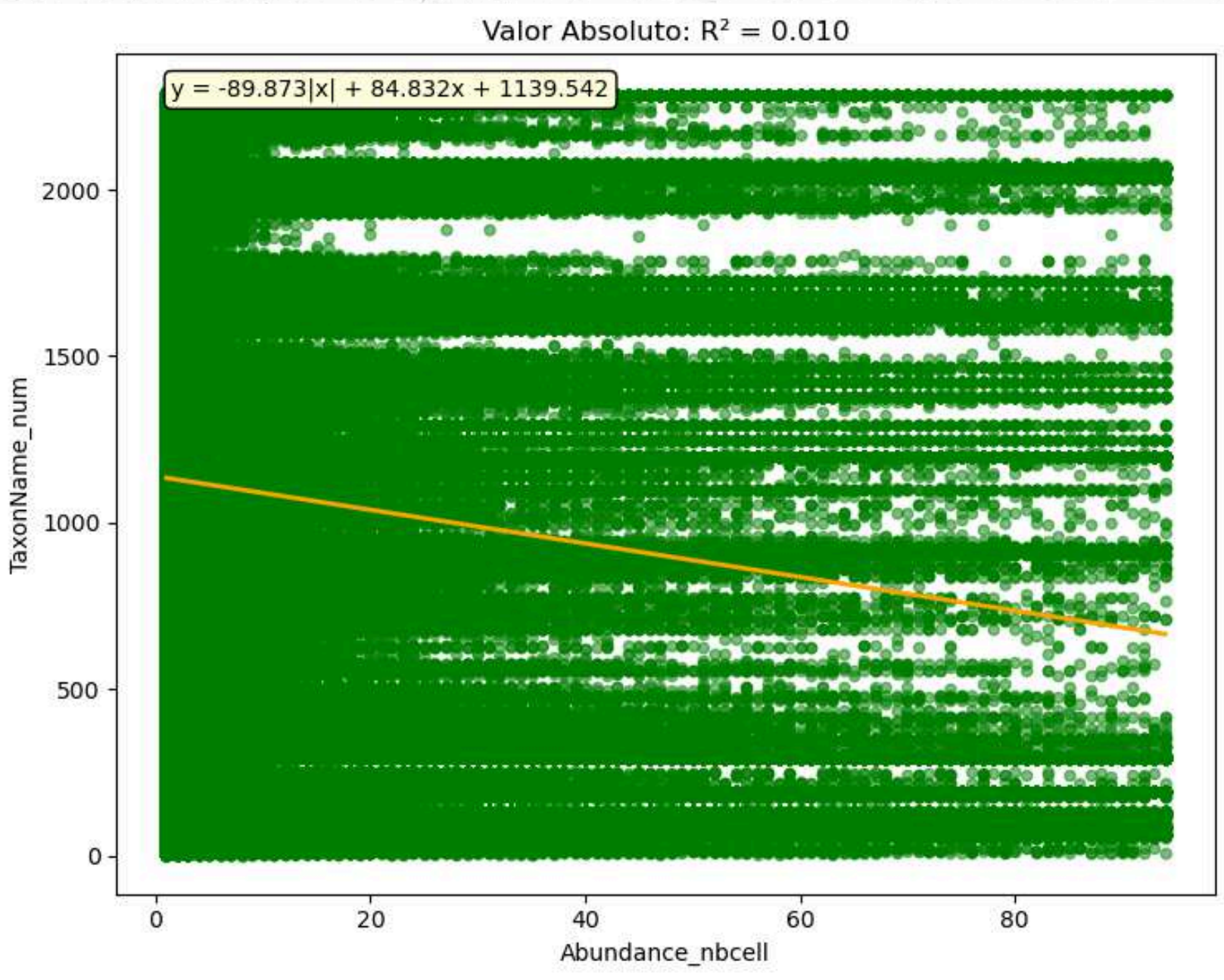
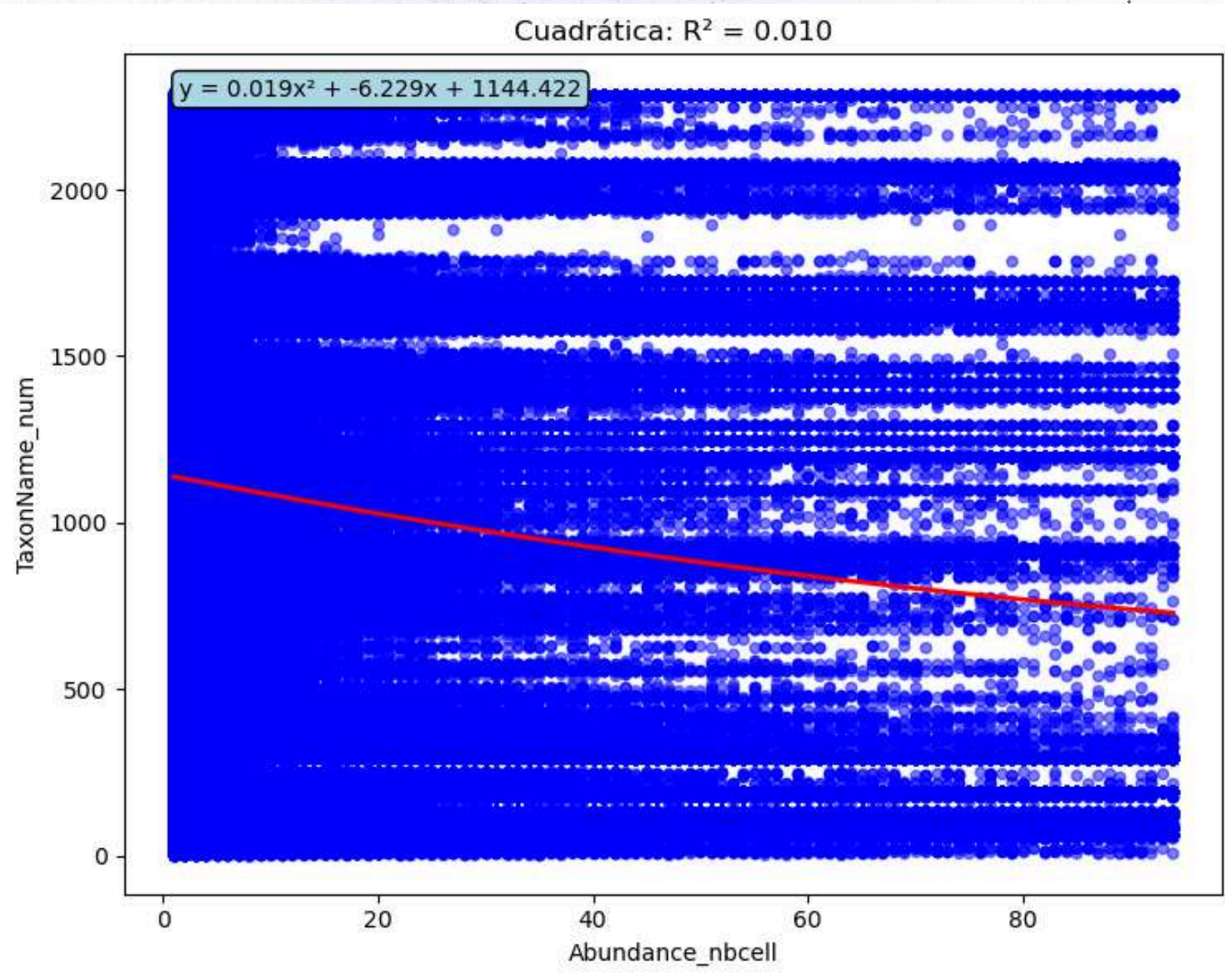


Total de modelos analizados:	16
Mejor R²:	0.9783
Mejor Correlación:	0.9891
R² Promedio:	0.1451
Correlación Promedio:	0.2325
TOP 5 MEJORES MODELOS:	Abundance_pm vs Abundance_nbcell (Cuadrática) Abundance_pm vs Abundance_nbcell (Exponencial) SamplingOperations vs CodeSite (Exponencial) SamplingOperations vs CodeSite (Logarítmica) Date_SamplingOperation vs SamplingOperations (Logarítmica)
ANÁLISIS POR TIPO DE FUNCIÓN	Exponencial: R² promedio = 0.2004 (6 modelos)



# TaxonName vs Abundance\_nbcell

- Distribución discreta
- Alta dispersión
- Concentración de datos



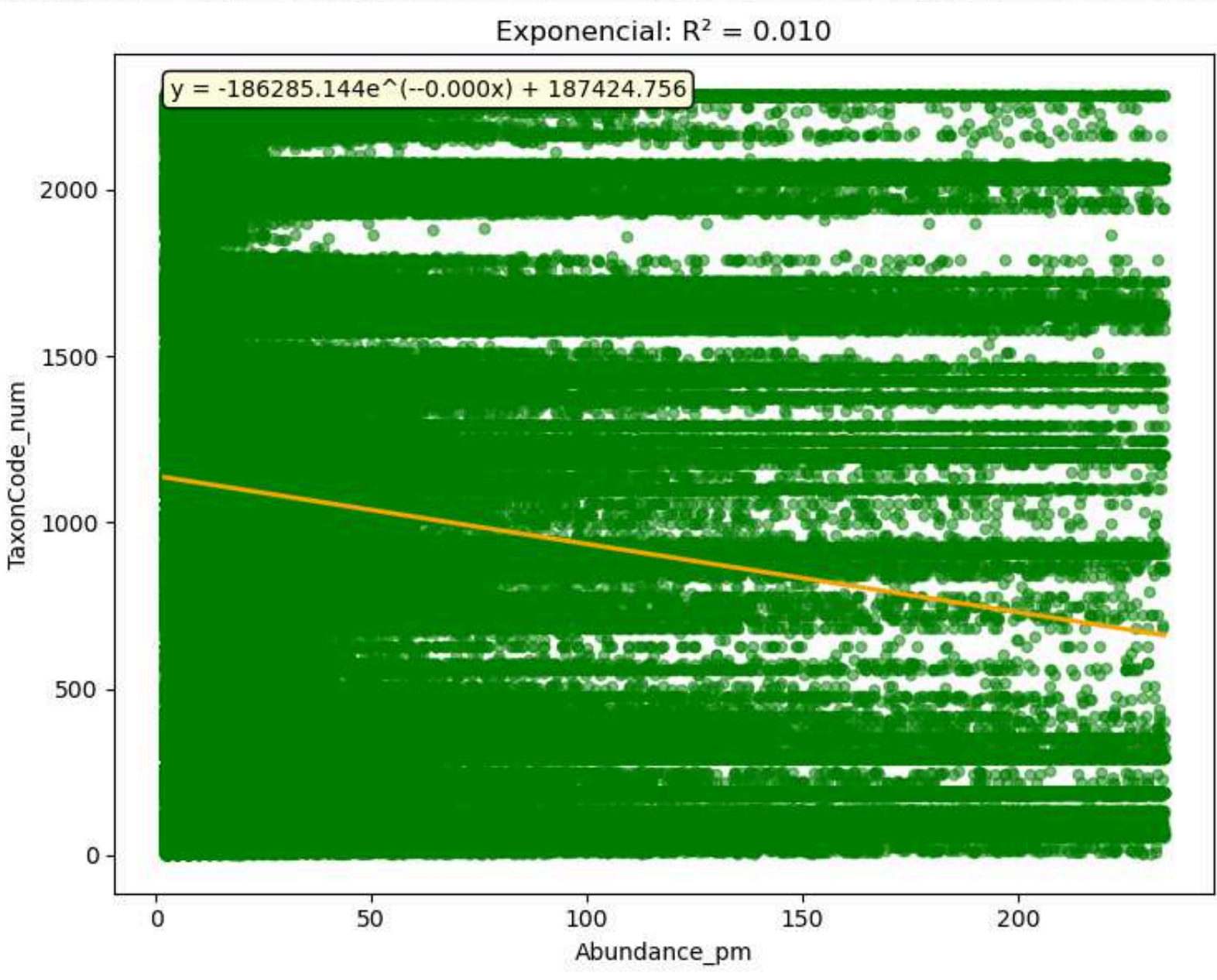
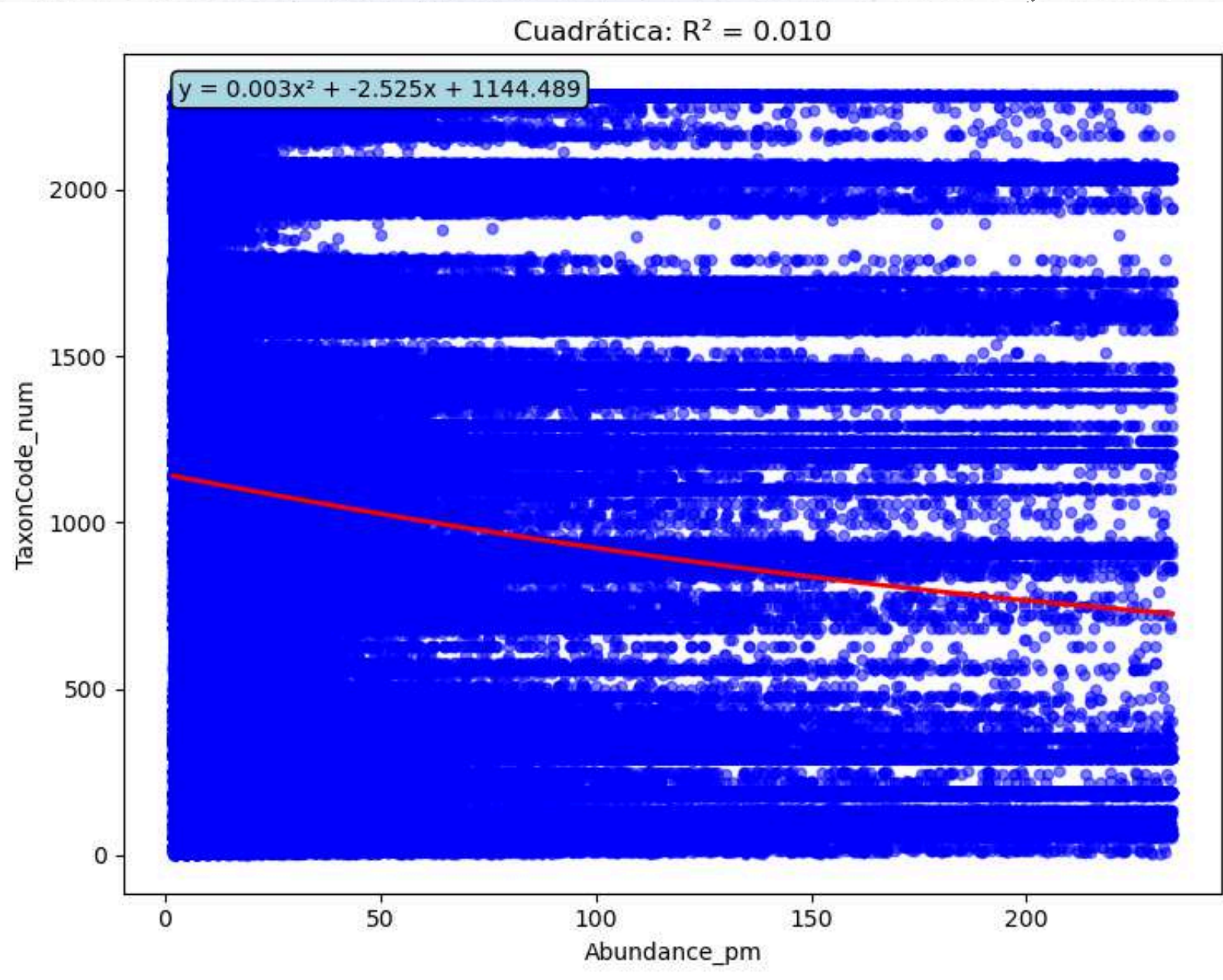
Correlacion L (r)	$r = -0.1003$
Correlacion No Lineal (r)	$r = 0.1007$

Correlacion L (r)	$r = -0.1003$
Correlacion No Lineal (r)	$r = 0.1003$



# TaxonCode vs Abundance\_pm

- Bandas discretas más pronunciadas
- Distribución uniforme
- Mayor dispersión horizontal



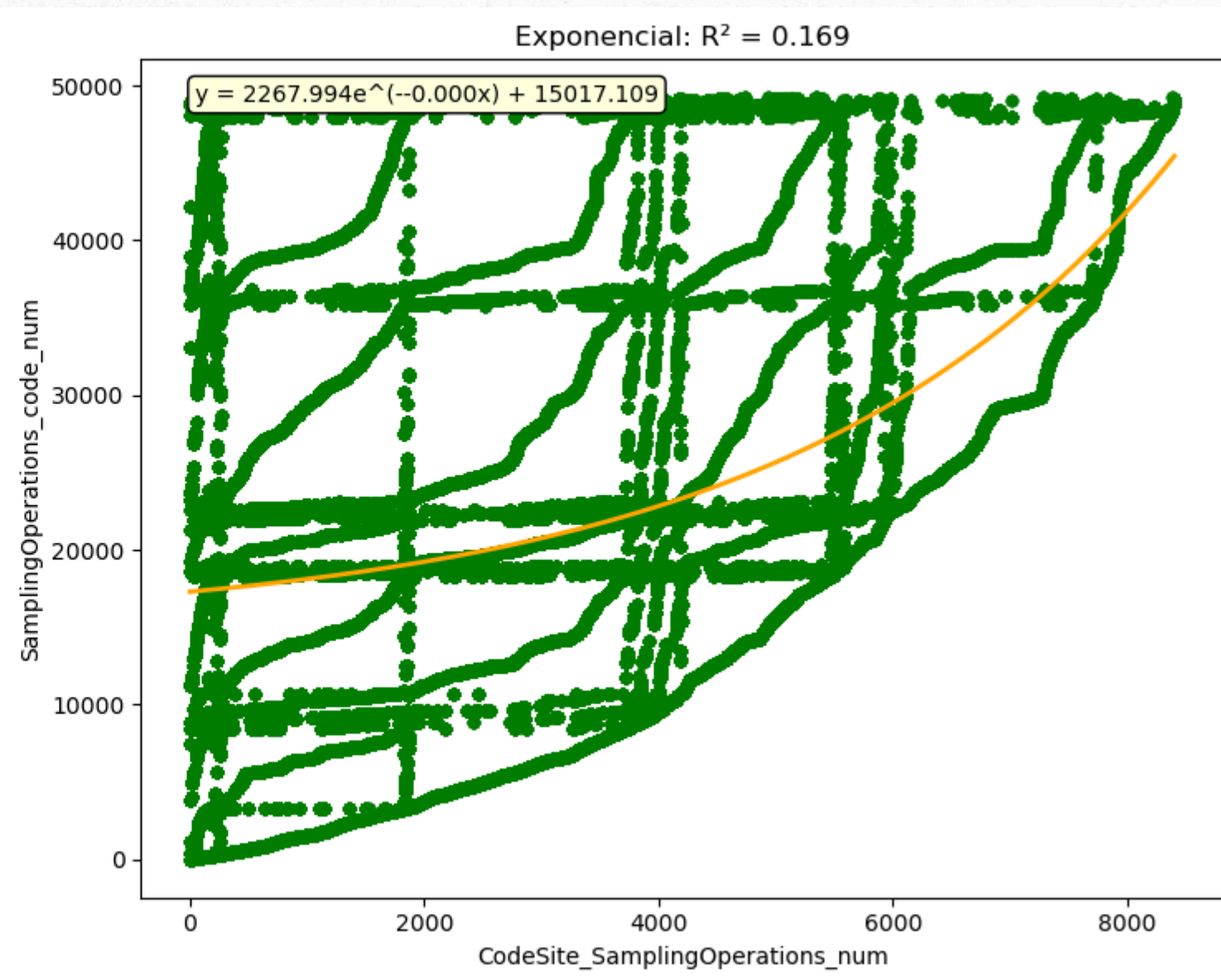
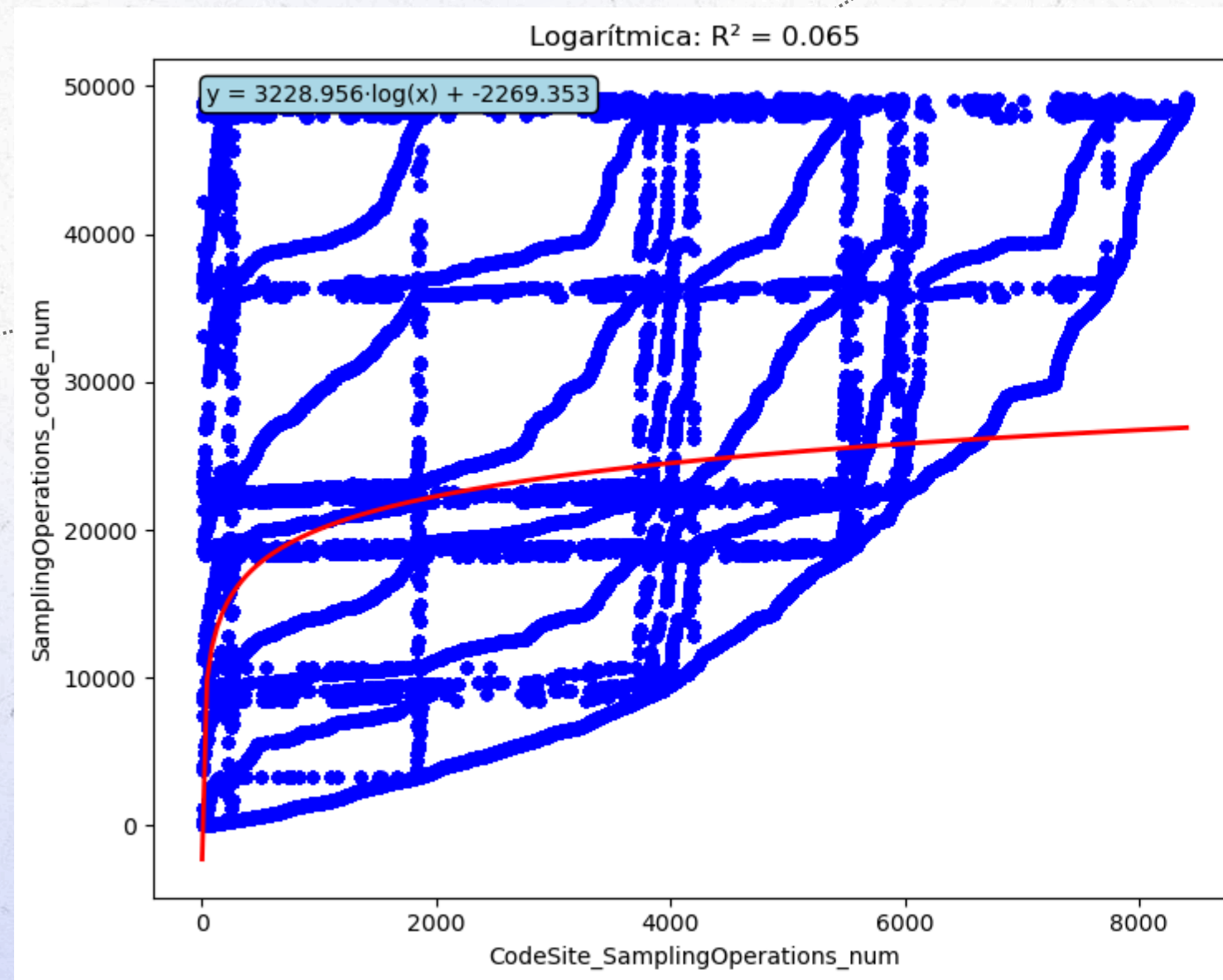
Correlacion L (r)	$r = -0.1006$
Correlacion No Lineal (r)	$r = 0.1011$

Correlacion L (r)	$r = -0.1006$
Correlacion No Lineal (r)	$r = 0.1006$



# SamplingOperations\_code vs CodeSite\_SamplingOperetions\_num

- Líneas verticales marcadas
- Distribución escalonada
- Patrón estructurado



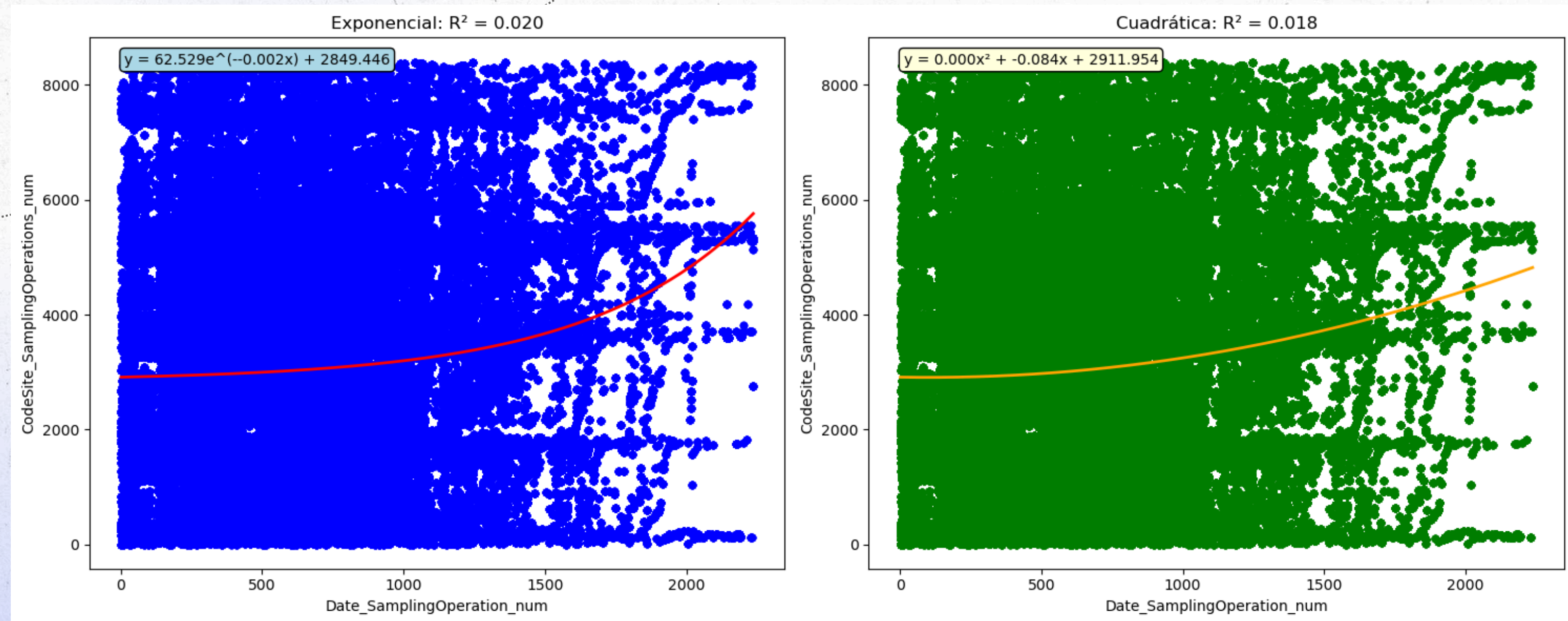
Correlacion L (r)	$r = 0.3836$
Correlacion No Lineal (r)	$r = 0.2558$

Correlacion L (r)	$r = 0.3836$
Correlacion No Lineal (r)	$r = 0.4105$



# CodeSite\_SamplingOperations vs Date\_SamplingOperation\_num

- Distribución triangular
- Mayor dispersión
- Concentración en valores bajos



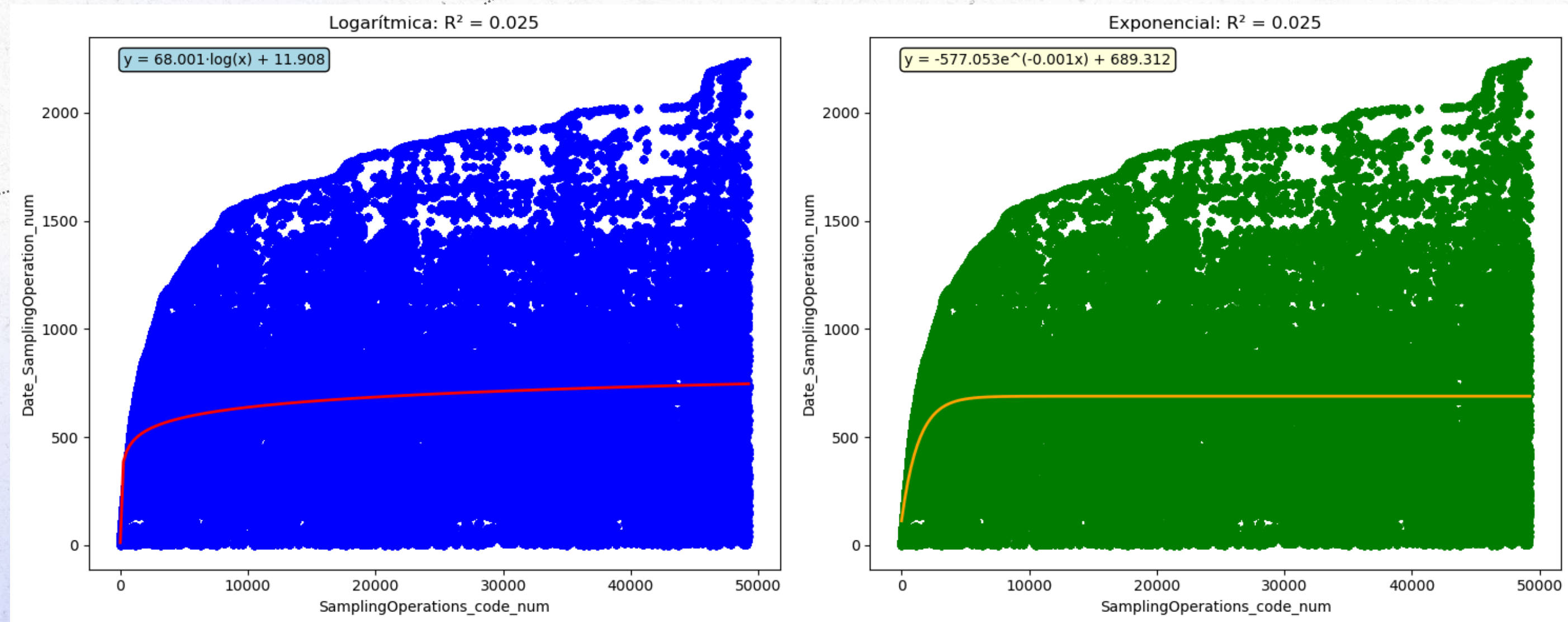
Correlacion L (r)	$r = 0.1269$
Correlacion No Lineal (r)	$r = 0.1409$

Correlacion L (r)	$r = 0.1269$
Correlacion No Lineal (r)	$r = 0.1360$



# Date\_SamplingOperation\_num vs SamplingOperations\_code\_num

- Bandas horizontales discretas
- Concentración masiva
- Expansión gradual



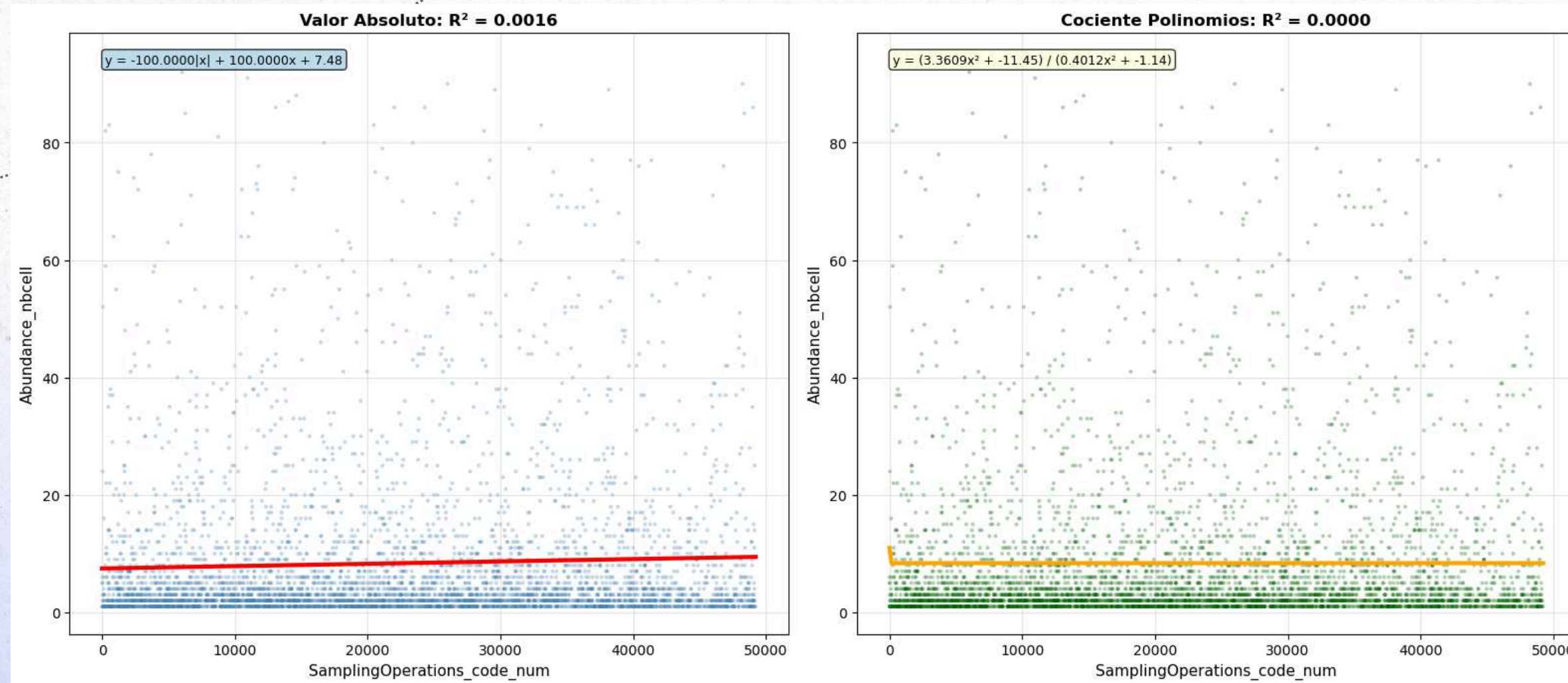
Correlacion L (r)	r = 0.1207
Correlacion No Lineal (r)	r = 0.1580

Correlacion L (r)	r = 0.1207
Correlacion No Lineal (r)	r = 0.1580



# Abundance\_nbcell vs SamplingOperations\_code\_num

- Distribución horizontal dominante
- Patrón de bandas discretas
- Concentración extrema en la base



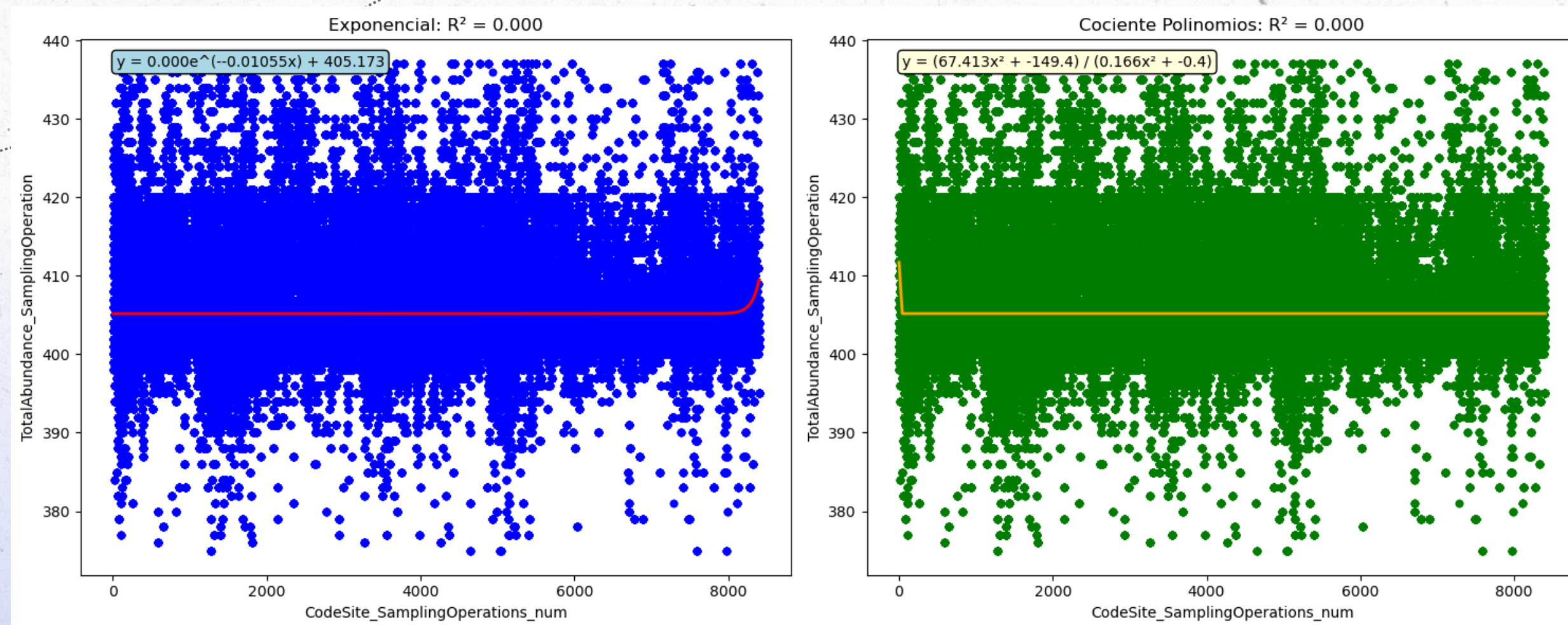
Correlacion L (r)	r = 0.0395
Correlacion No Lineal (r)	r = 0.0395

Correlacion L (r)	r = 0.0395
Correlacion No Lineal (r)	r = 0.0015



# TotalAbundance\_SamplingOperation vs CodeSite\_SamplingOperations\_num

- Distribución rectangular densa
- Concentración superior
- Banda principal dominante



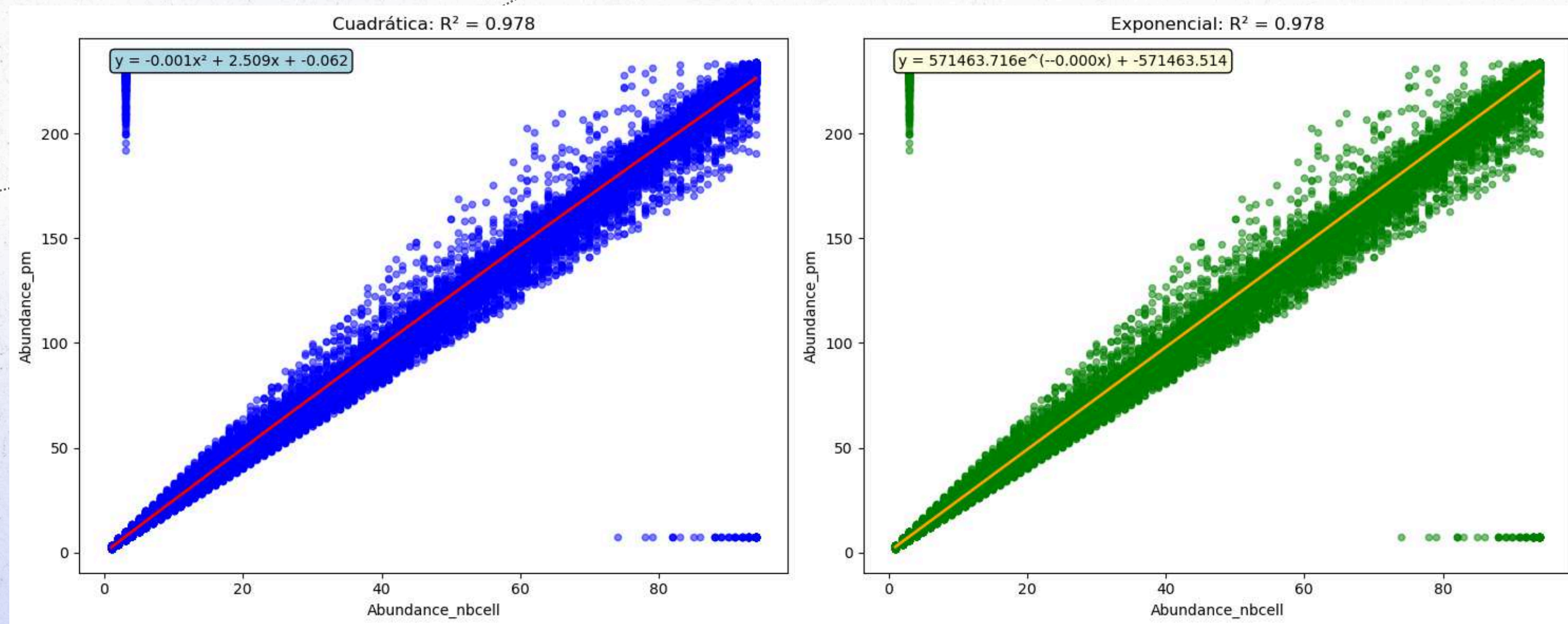
Correlacion L (r)	$r = 0.0146$
Correlacion No Lineal (r)	$r = 0.0211$

Correlacion L (r)	$r = 0.0146$
Correlacion No Lineal (r)	$r = 0.0174$



# Abundance\_pm vs Abundance\_nbccl

- Correlación lineal muy fuerte:
- Dispersión creciente
- Concentración en origen



Correlacion L (r)	r = 0.9890
Correlacion No Lineal (r)	r = 0.9891

Correlacion L (r)	r = 0.9890
Correlacion No Lineal (r)	r = 0.9890



# ANÁLISIS DE INSIGHTS

**Cada diatomea  
está correctamente  
identificada y  
codificada**

**Hay guerra  
ecológica entre  
especies por los  
recursos**

**Las diatomeas viven en  
ambientes muy  
predecibles, estables y  
consistente a lo largo  
del tiempo**

**La abundancia  
total es  
independiente de  
qué especies hay**

**Tienes datos de alta calidad  
de un ecosistema con  
patrones ecológicos claros  
que revelan competencia,  
estabilidad temporal, y  
gradientes espaciales  
interesantes para investigar.**

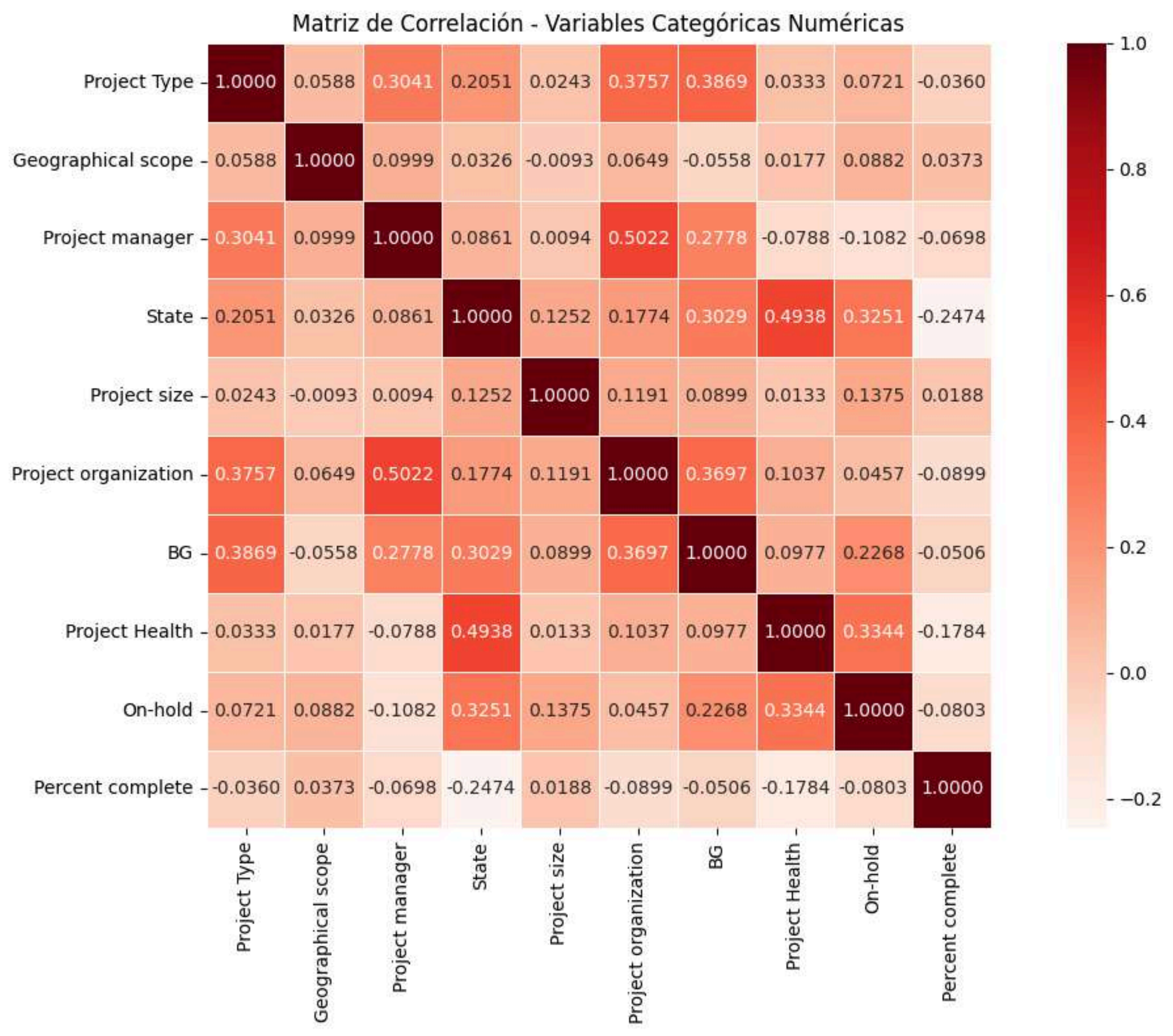




# ACTIVIDAD 3.2

**FORVIA**  
faurecia



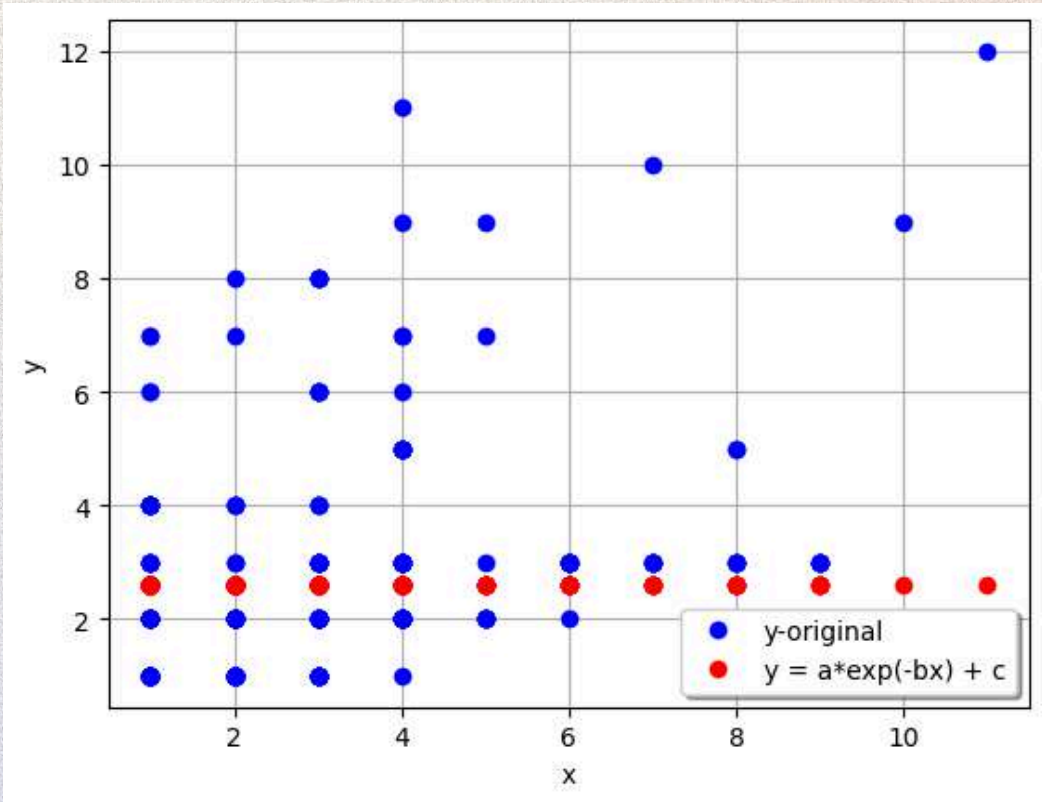
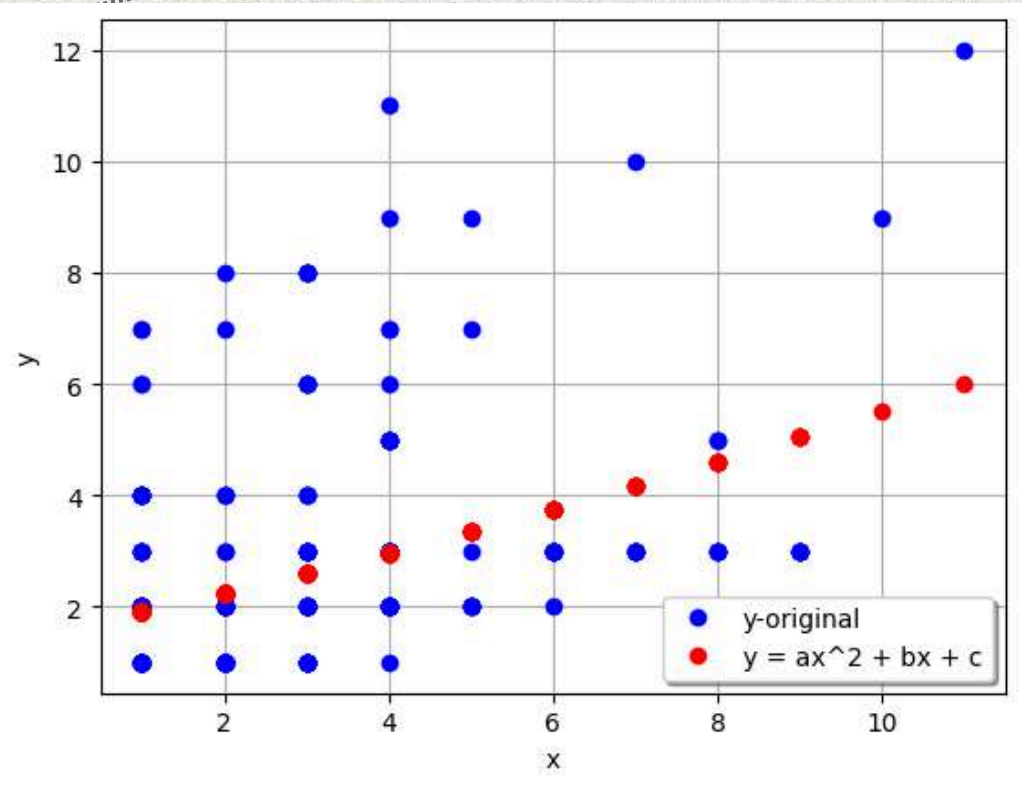




# BG - Project Type

## MODELOS NO LINEALES:

- CUADRÁTICO:  $Y=AX^2+BX+CY = AX^2 + BX + C$
- FUNCIÓN EXPONENCIAL:  $Y = A*EXP(-BX) + C$



- **Resultados:**
- **Cuadrático:**  $R^2 = 0.1502$ ,  $r = 0.3876$
- **Exponencial:**  $R^2 = 3.3306$ ,  $r = 1.8250$
- La función cuadrática es la mejor ya que a pesar que esta no crece mucho los datos de la formula exponencial son exagerados para una correlación.
- **Visualmente:** alta dispersión sin tendencia clara ni agrupamientos definidos.

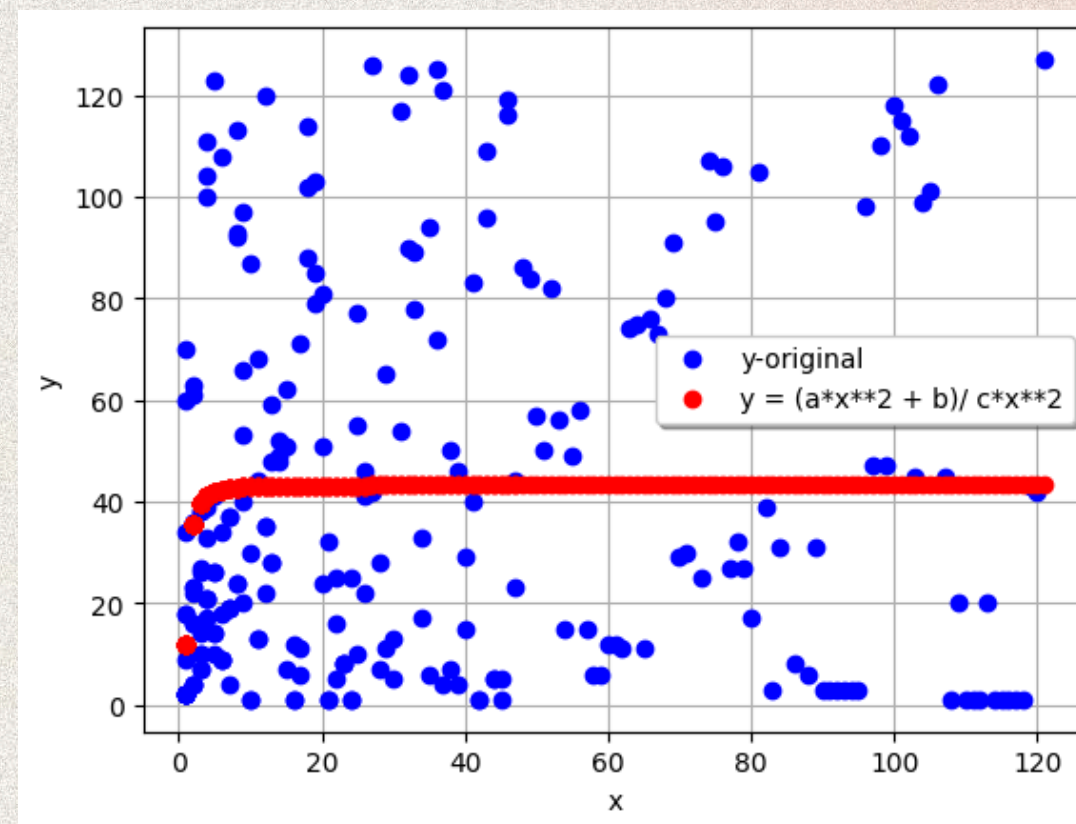
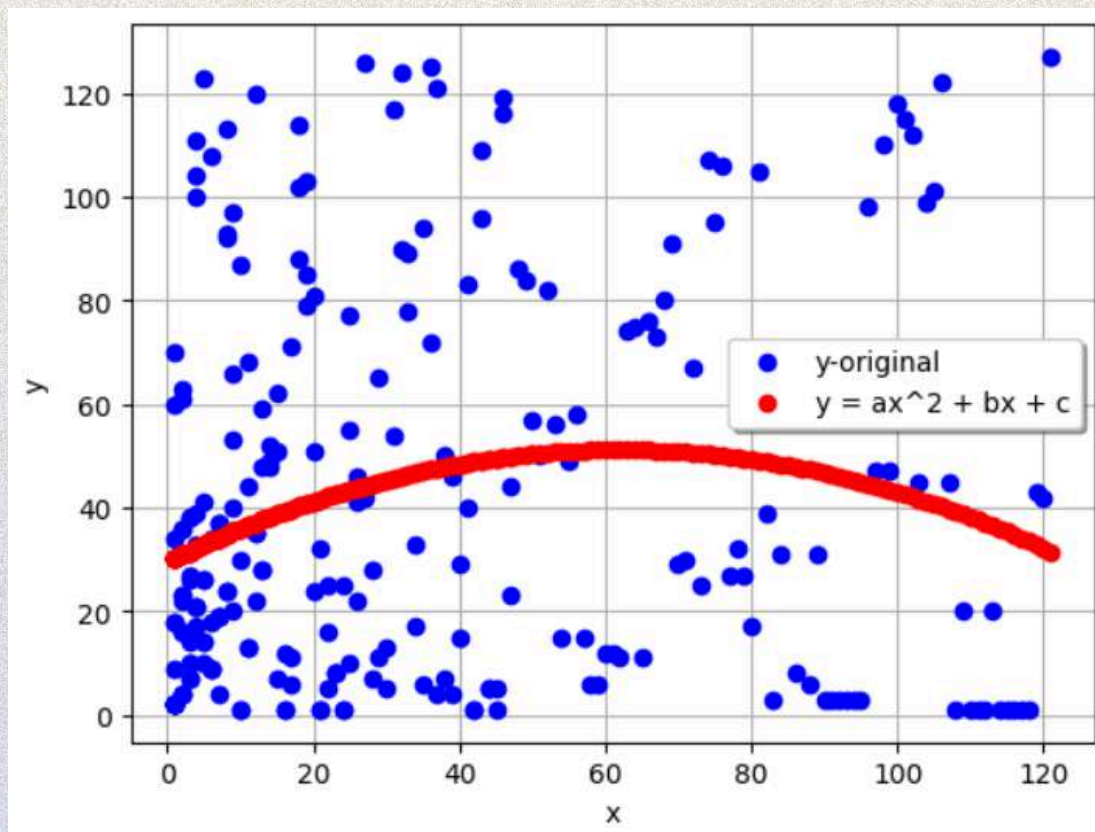
Funcion R2	0.1502
Correlacion (r)	0.386
Correlacion Final (r)	0.3876



# Project manager - Geographical scope

## MODELOS NO LINEALES:

- CUADRÁTICO:  $Y = AX^2 + BX + C$
- FUNCIÓN COCIENTE ENTRE POLINOMIOS:  $Y = (A \cdot X^{**2} + B) / C \cdot X^{**2}$



- **Resultados:**
- **Cuadrático:**  $R^2 = 0.0365$ ,  $r = 0.1912$
- **Co entre Polinomios:**  $R^2 = 0.0407$ ,  $r = 0.2018$
- **Interpretación:** La correlación en ambas es la mejor entre las demás ambas mejoran.
- **Visualmente:** alta dispersión con tendencia clara y determinada con agrupamientos definidos.

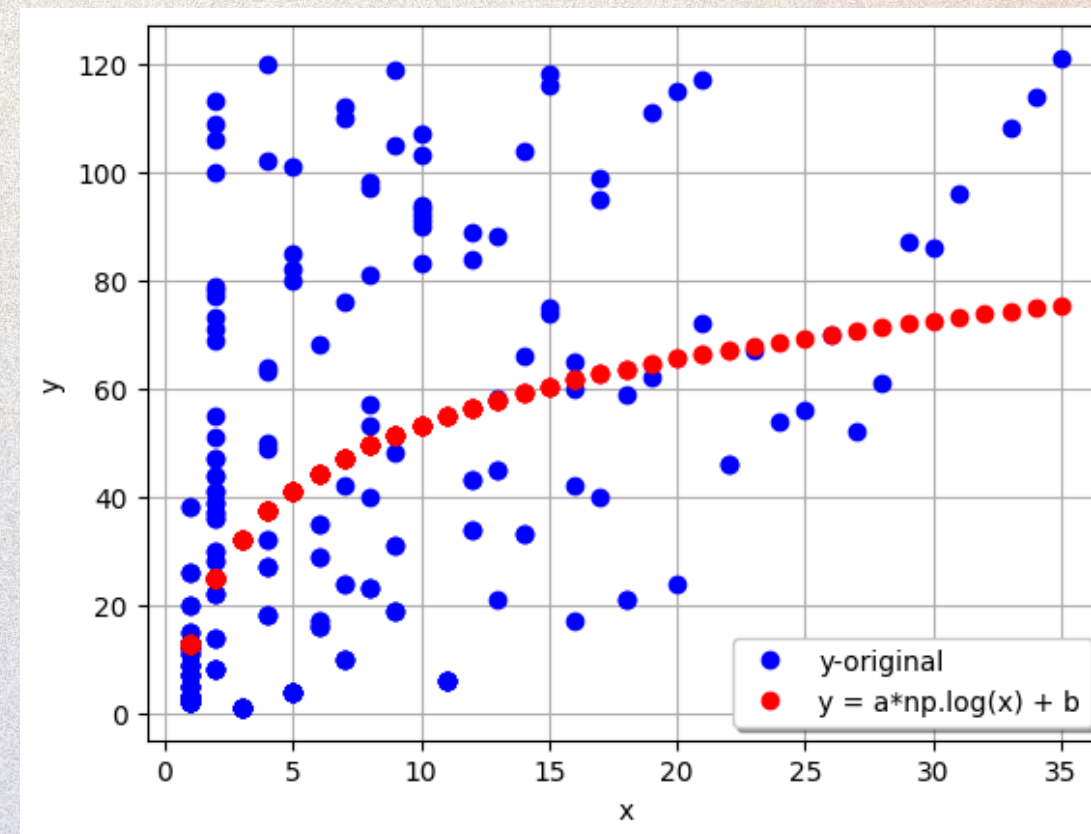
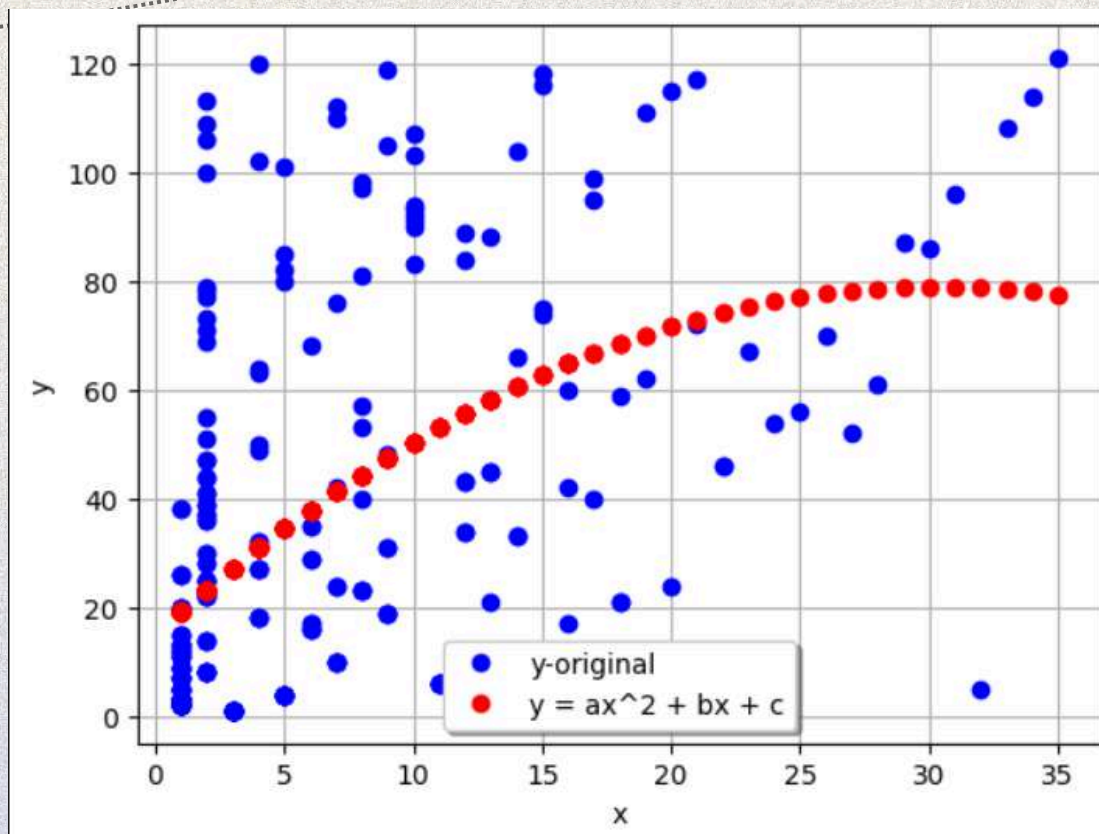
Funcion R2	0.0407
Correlacion (r)	0.0999
Correlacion Final (r)	0.2018



# Project organization - Project manager

## MODELOS NO LINEALES:

- CUADRÁTICO:  $Y = AX^2 + BX + C$
- LOGARITMICA:  $Y = A \cdot E^{-BX} + C$



- Resultados:
- Cuadrático:  $R^2 = 0.2723$ ,  $r = 0.5219$
- Logaritmica:  $R^2 = 0.2885$ ,  $r = 0.5371$
- Interpretación: Las correlaciones mejoran en ambos casos.
- Visualmente: alta dispersión con tendencia clara y agrupamientos mejorados.

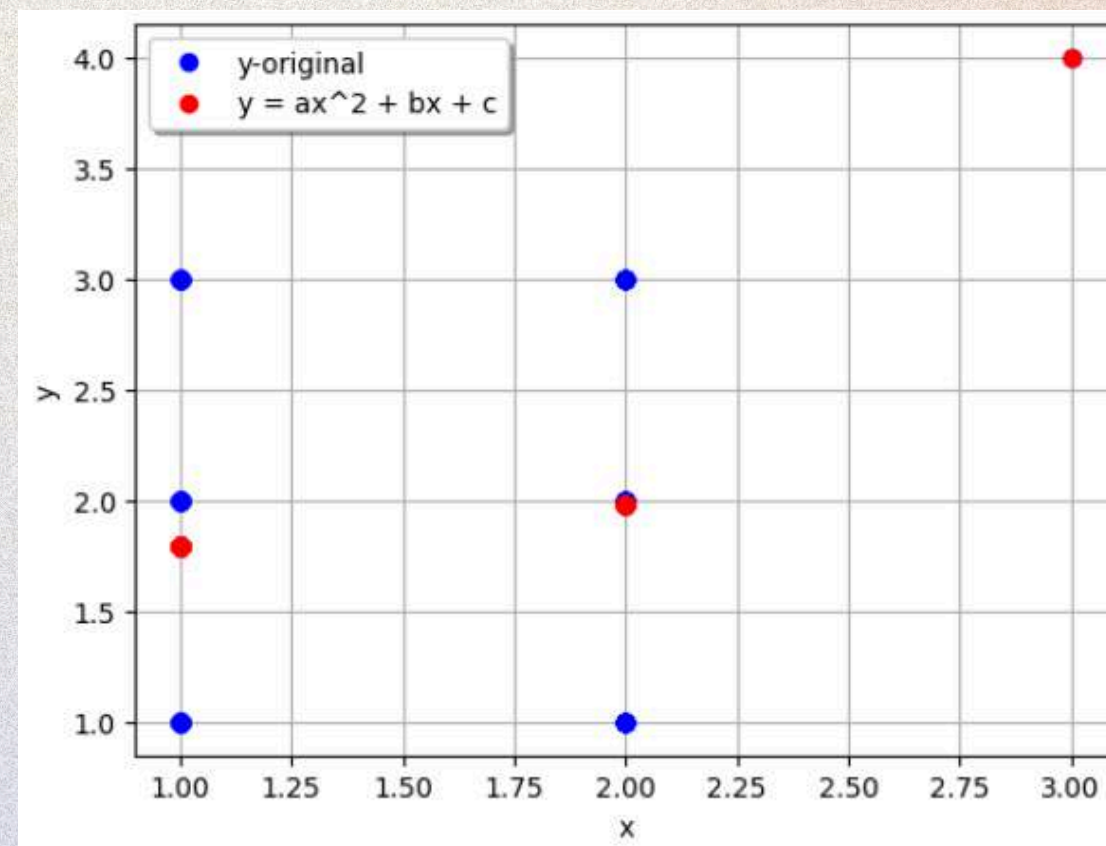
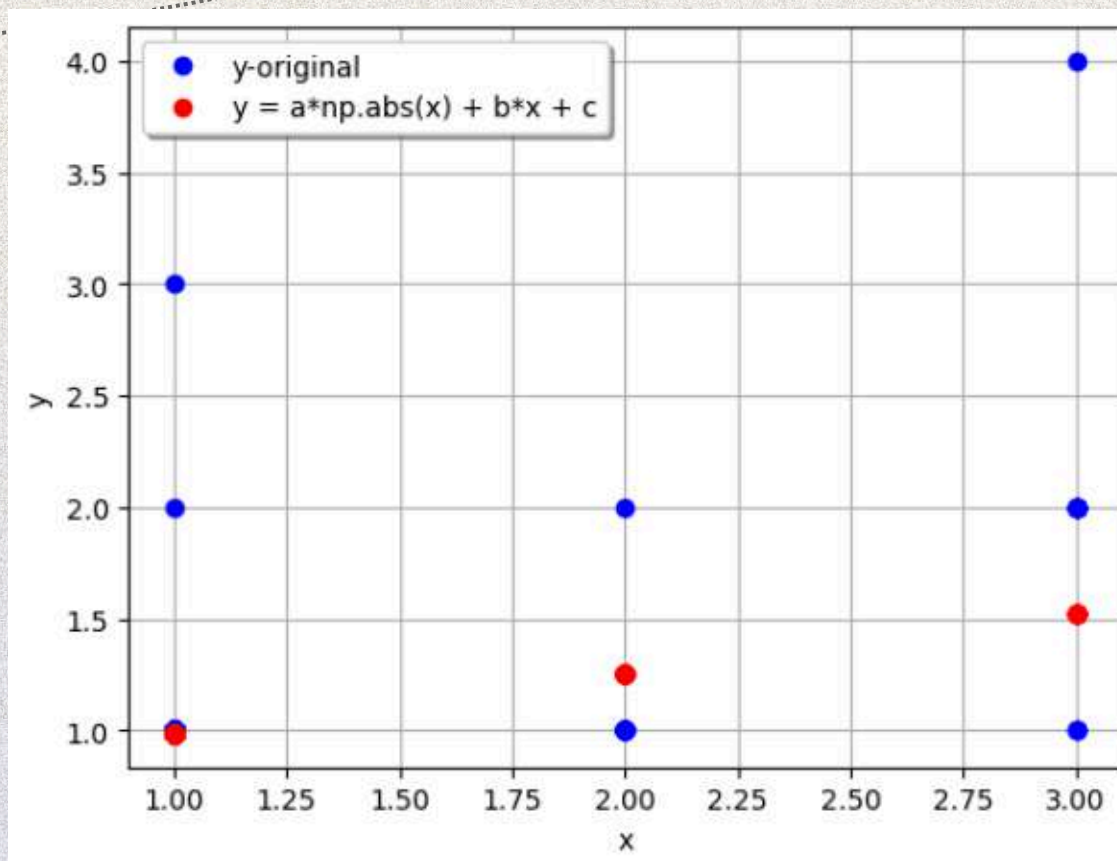
Funcion R2	0.2885
Correlacion (r)	0.5022
Correlacion Final (r)	0.5371



# Project Health - State

## MODELOS NO LINEALES:

- **FUNCIÓN SENOIDAL:**  $Y = A * \text{NP.SIN}(X) + B$
- **FUNCIÓN VALOR ABSOLUTO:**  $Y = A * \text{NP.ABS}(X) + B * X + C$



- **Resultados:**
- **Senoidal:**  $R^2 = 0.4473$ ,  $r = 0.6688$
- **Val. Absoluto:**  $R^2 = 0.2438$   $r = 0.4937$
- **Interpretación:** De las dos mayores correlaciones la senodial es la unica coherente ya que es la unica que supera a la correlación inicial.
- **Visualmente:** alta dispersión sin tendencia clara ni agrupamientos definidos.

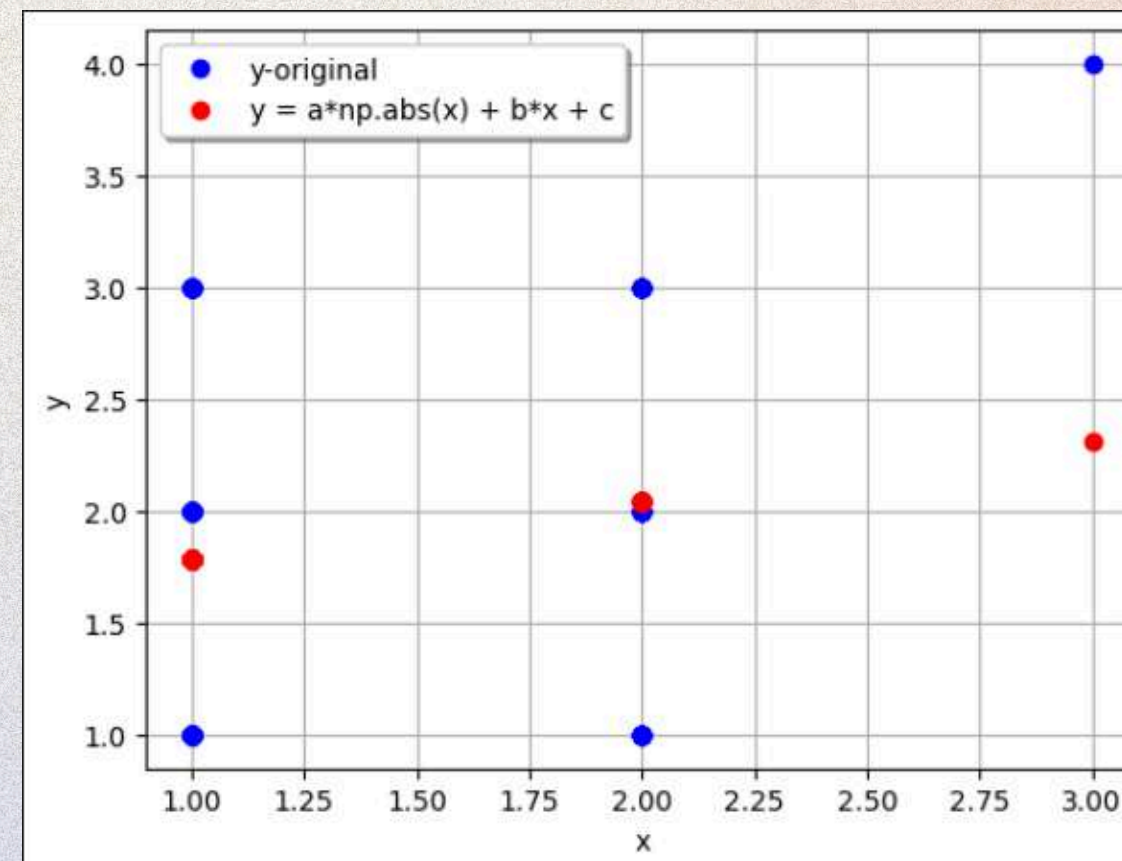
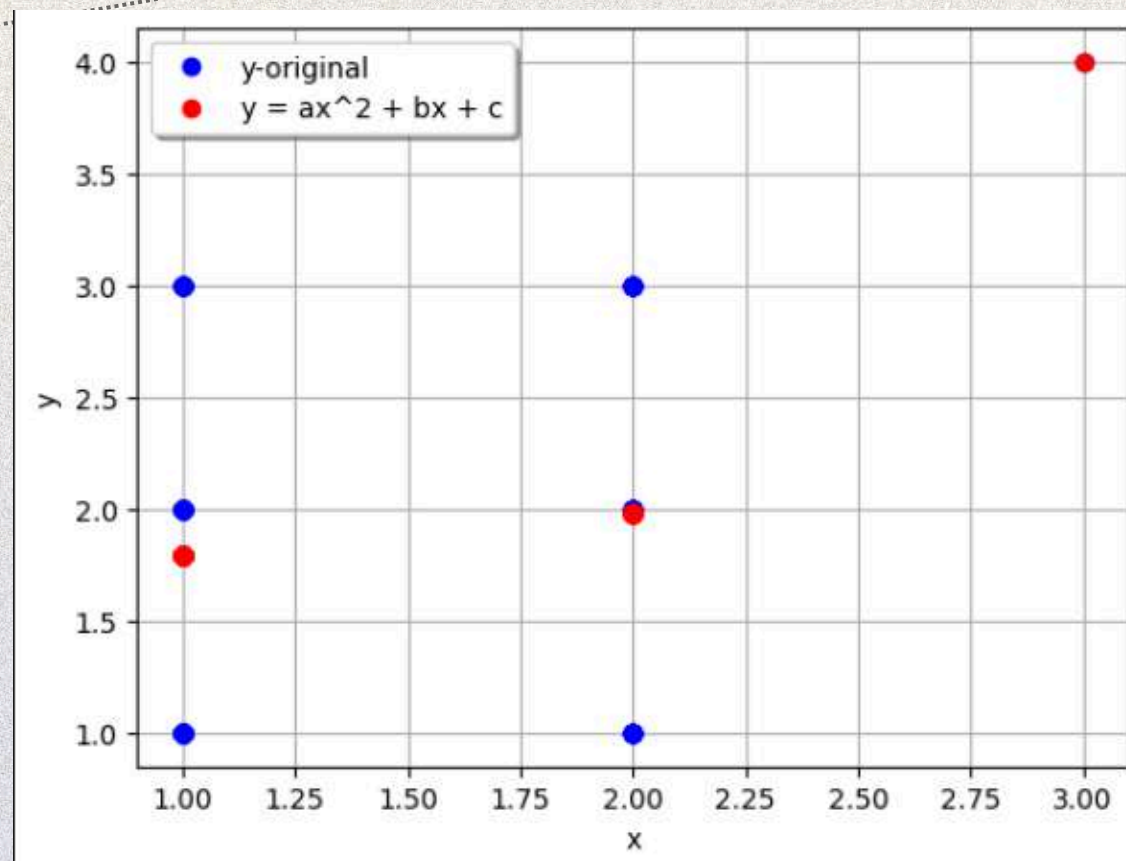
Funcion R2	0.4473
Correlacion (r)	0.4937
Correlacion Final (r)	0.6688



# On-hold - Project size

## MODELOS NO LINEALES:

- CUADRÁTICO:  $Y = AX^2 + BX + C$
- FUNCION VALOR ABSOLUTO:  $Y = A \cdot \text{NP.ABS}(X) + B \cdot X + C$



- **Resultados:**
- **Cuadrático:**  $R^2 = 0.0379$ ,  $r = 0.1949$
- **Valor A.:**  $R^2 = 0.0188$ ,  $r = 0.1374$
- **Interpretación:** La correlación de la regresión cuadrática es considerablemente mejor.
- **Visualmente:** alta dispersión sin tendencia clara ni agrupamientos definidos.

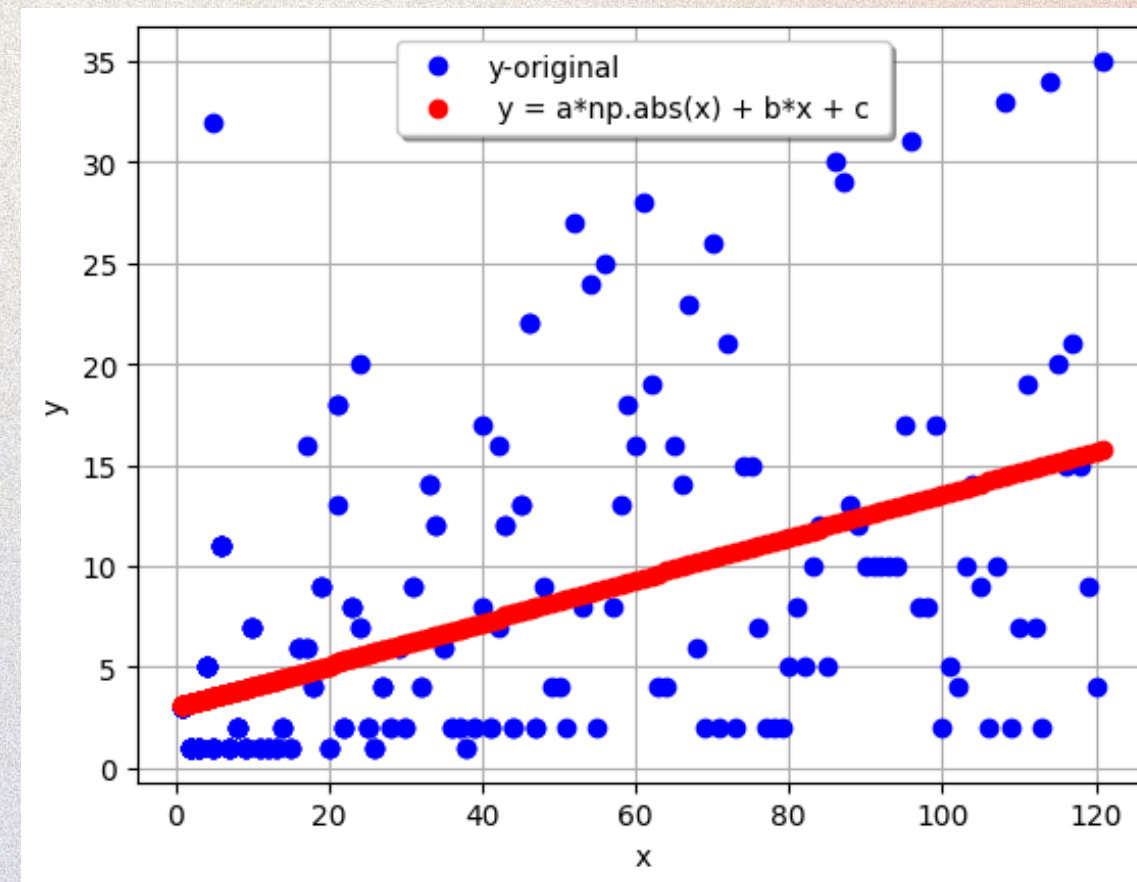
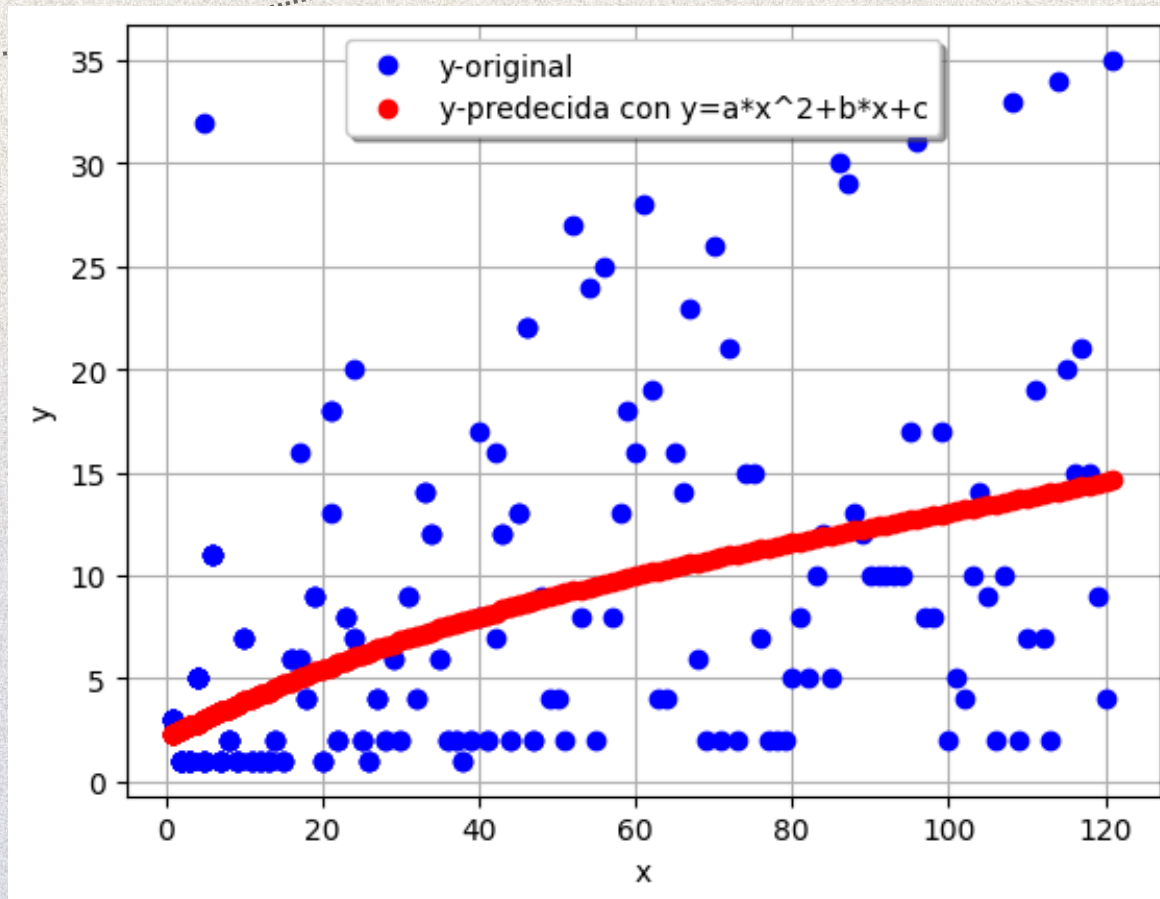
Funcion R2	0.0379
Correlacion (r)	0.1375
Correlacion Final (r)	0.1949



# Project manager - Project organization

## MODELOS NO LINEALES:

- CUADRÁTICO:  $Y = AX^2 + BX + C$
- FUNCION VALOR ABSOLUTO:  $Y = A \cdot \text{NP.ABS}(X) + B \cdot X + C$



- Resultados:
- Cuadrático:  $R^2 = 0.2579$ ,  $r = 0.5078$
- Valor A.:  $R^2 = 0.2522$ ,  $r = 0.5022$
- Interpretación: Ambas regresiones ayudaron a crecer la correlación considerablemente siendo estas muy similares.
- Visualmente: alta dispersión con tendencia clara y agrupamientos definidos.

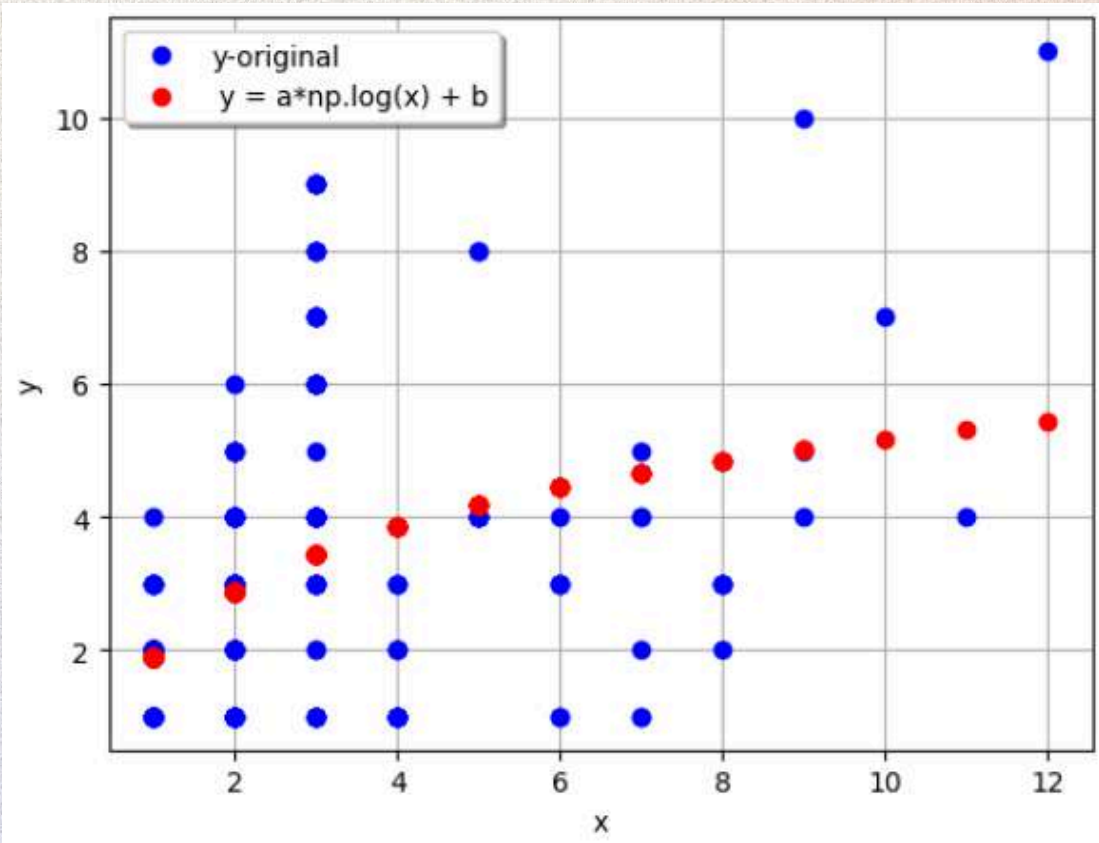
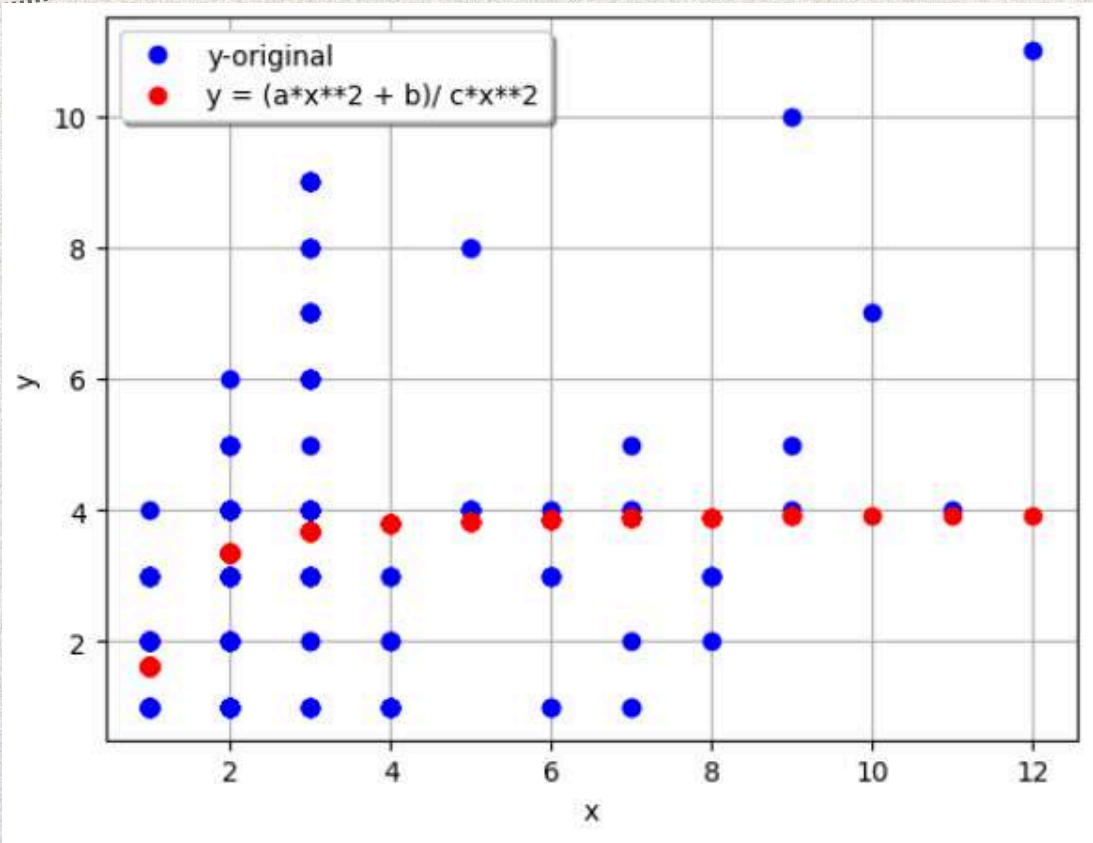
Funcion R2	0.2579
Correlacion (r)	0.5022
Correlacion Final (r)	0.5078



# Project Type - BG

## MODELOS NO LINEALES:

- FUNCIÓN COCIENTE ENTRE POLINOMIOS:  $Y = (A \cdot X^{**2} + B) / C \cdot X^{**2}$
- FUNCIÓN LOGARITMICA:  $Y = A \cdot \text{NP.LOG}(X) + B$



- **Resultados:**
- **Cociente:**  $R^2 = 0.2128$ ,  $r = 0.4613$
- **Logaritmica:**  $R^2 = 0.2039$ ,  $r = 0.4515$
- **Interpretación:** La correlacion mejora en ambos casos y tienen resultados similares respecto a la correlación
- **Visualmente:** alta dispersión con tendencia clara ni agrupamientos definidos.

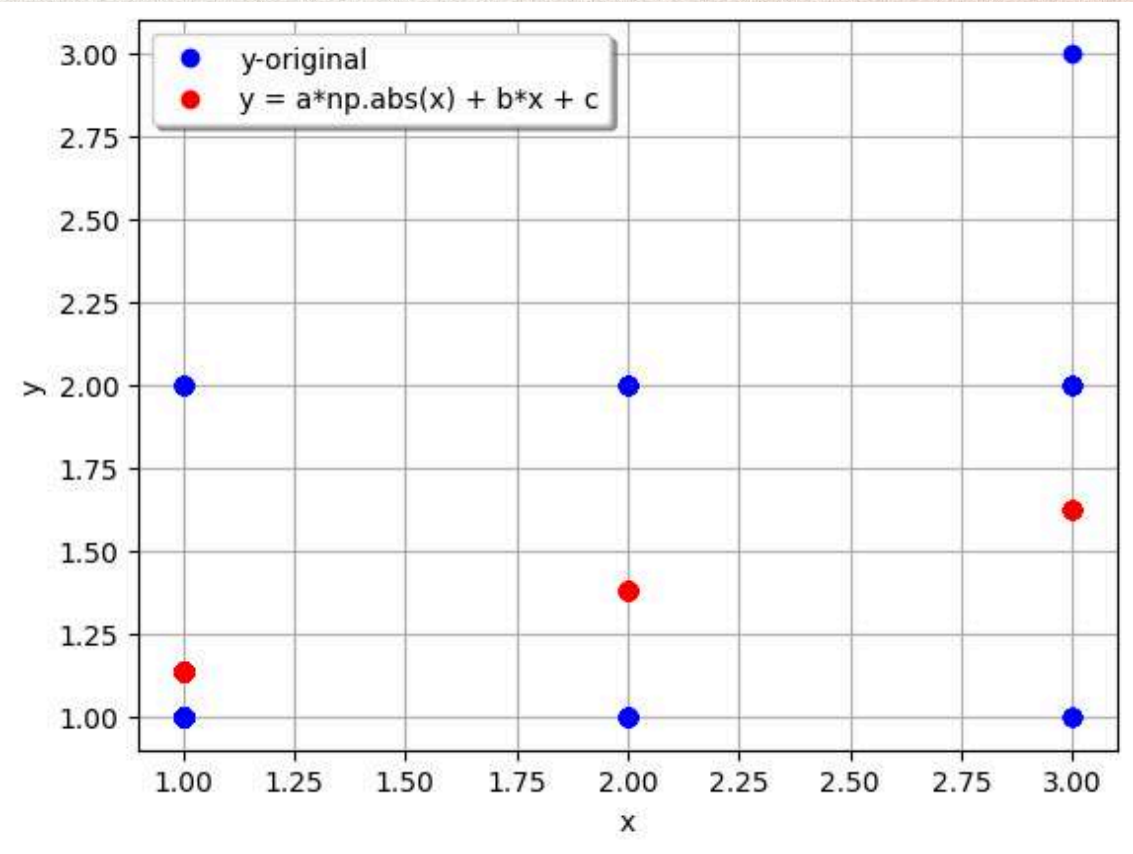
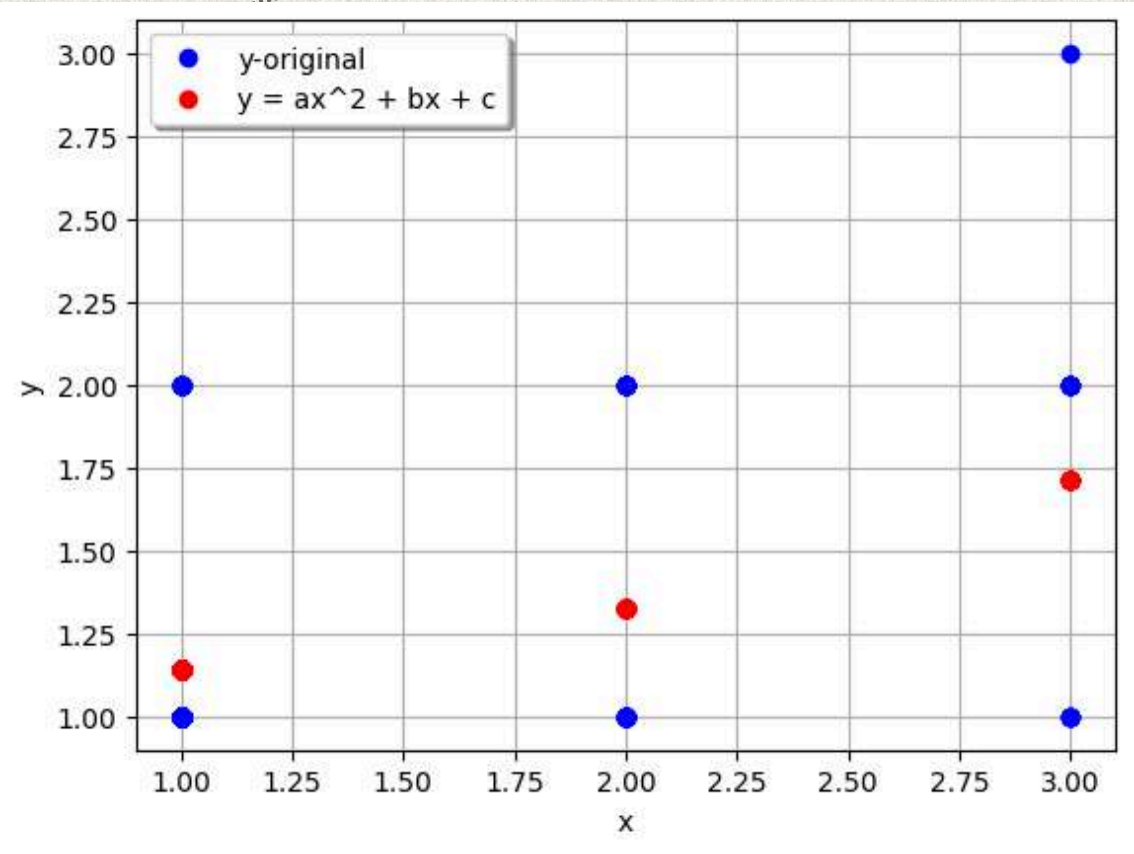
Funcion R2	0.2128
Correlacion (r)	0.3869
Correlacion Final (r)	0.4613



# Project Health - On-hold

## MODELOS NO LINEALES:

- CUADRÁTICO:  $Y = AX^2 + BX + C$
- FUNCION VALOR ABSOLUTO:  $Y = A \cdot \text{NP.ABS}(X) + B \cdot X + C$



- Resultados:
- Cuadrático:  $R^2 = 0.1178$ ,  $r = 0.3433$
- Exponencial:  $R^2 = 0.1118$ ,  $r = 0.3343$
- Interpretación: Parecida  $R^2$  y correlación mejoran por lo mínimo.
- Visualmente: alta dispersión con tendencias claras pero no hay agrupamientos definidos.

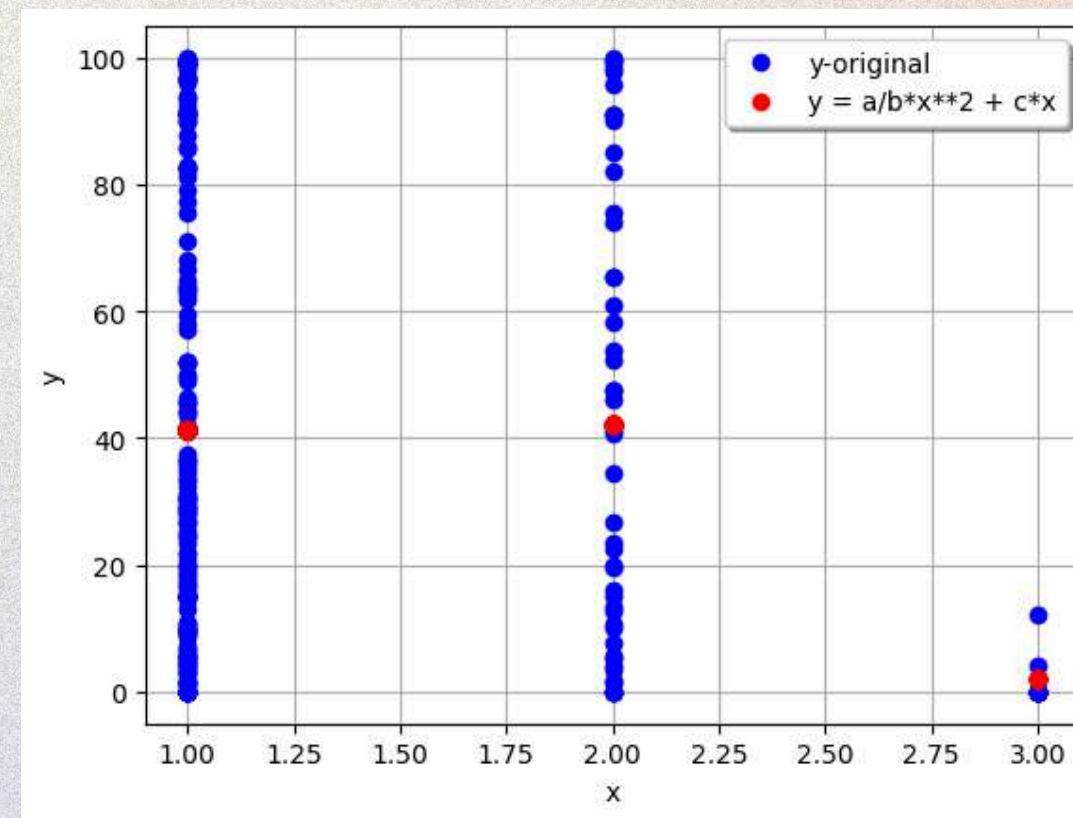
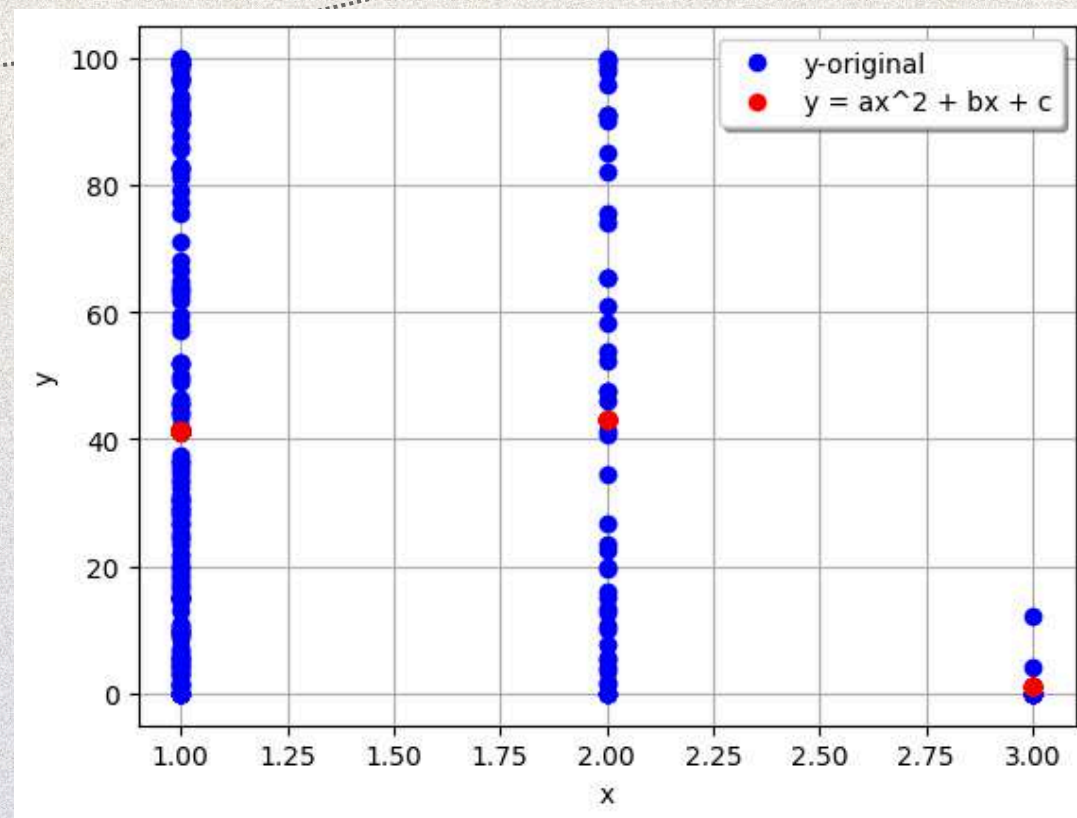
Funcion R2	0.1178
Correlacion (r)	0.3344
Correlacion Final (r)	0.3433



# Project Health - Percent complete

## MODELOS NO LINEALES:

- CUADRÁTICO:  $Y = AX^2 + BX + C$
- FUNCIÓN POLINOMIAL INVERSA:  $Y = A/B * X^{**2} + C * X$



- **Resultados:**
- **Cuadrático:**  $R^2 = 0.0709$ ,  $r = 0.2663$
- **Exponencial:**  $R^2 = 0.0707$ ,  $r = 0.2660$
- **Interpretación:** Similares correlaciones y  $R^2$ , ambas subieron considerablemente.
- **Visualmente:** alta dispersión con tendencias nada claras y no hay agrupamientos definidos.

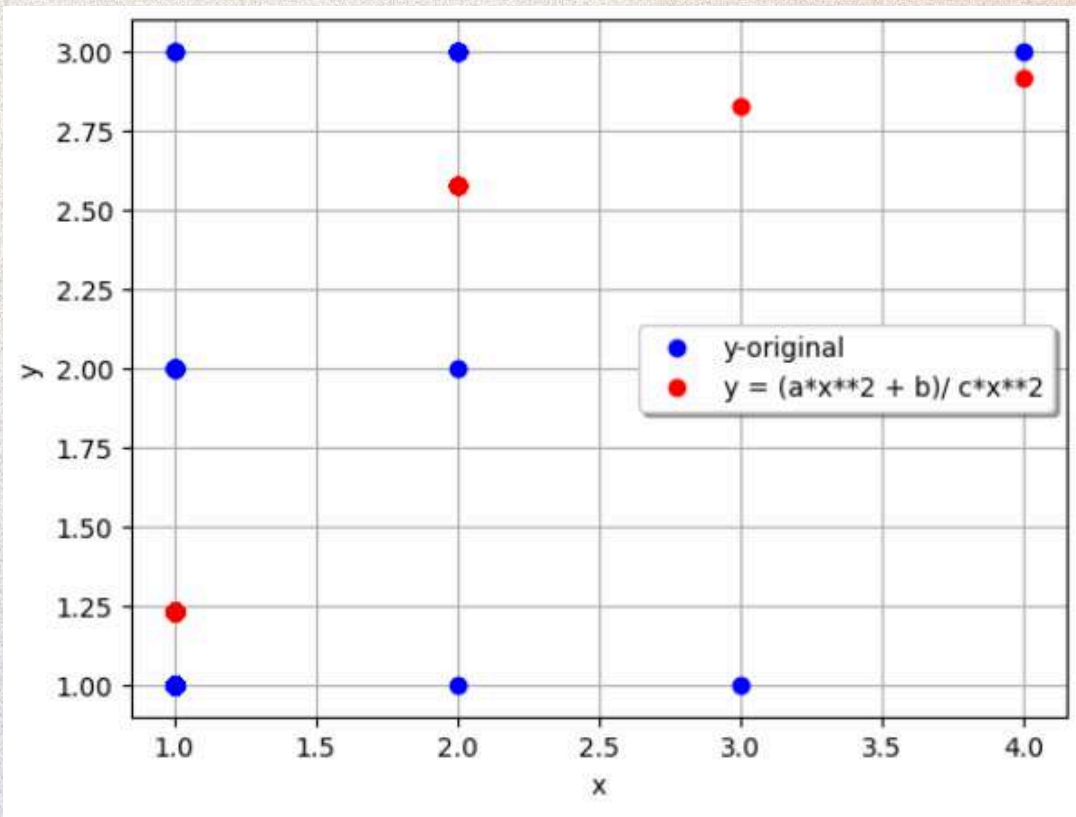
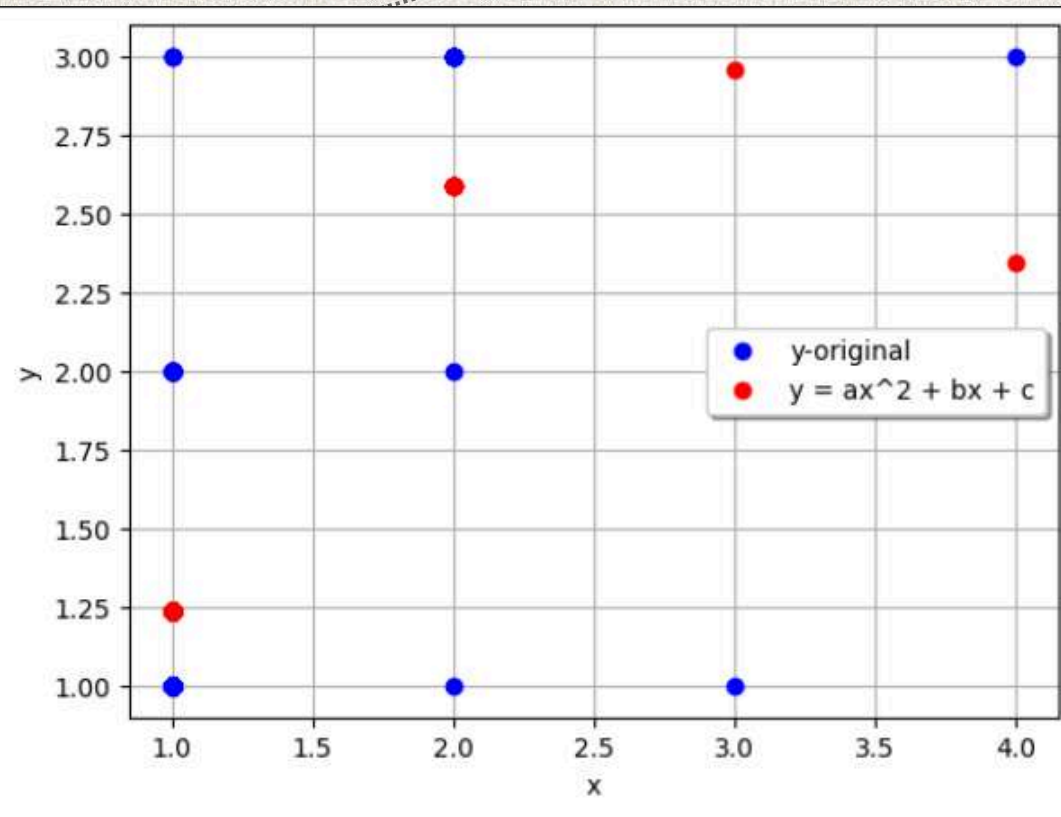
Funcion R2	0.0709
Correlacion (r)	-0.1784
Correlacion Final (r)	0.2663



# State - Project Health

## MODELOS NO LINEALES:

- CUADRÁTICO:  $Y=AX^2+BX+CY = AX^2 + BX + CY=AX^2+BX+C$
- FUNCIÓN COCIENTE ENTRE POLINOMIOS:  $Y = (A*X^{**2} + B)/ C*X^{**2}$



- **Resultados:**
- **Cuadrático:**  $R^2 = 0.3045$ ,  $r = 0.5518$
- **Polinomios:**  $R^2 = 0.3154$ ,  $r = 0.5616$
- **Interpretación:** relación extremadamente débil (~1% de variabilidad explicada) con ligera pendiente negativa casi nula.
- **Visualmente:** alta dispersión sin tendencia clara ni agrupamientos definidos.

Funcion R2	0.3154
Correlacion (r)	0.4938
Correlacion Final (r)	0.5616



# ANÁLISIS DE INSIGHTS

**Patrones de  
Registro**

**Estandarización de  
Procesos**

**Campos No  
Independientes**

**Correlación #1:  
Project Manager ↔  
Project  
Organization  
(0.505)**

Forvia es una organización madura con sistemas sofisticados de gestión de proyectos, operaciones globales diversificadas, y una estructura organizacional adaptable que les permite manejar eficientemente un portfolio complejo y variado de proyectos.





GRACIAS POR  
SU ATENCION