

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).

Project Type, Geographical scope, Project manager, State, Project size, Project organization, BG, On-hold, Percent complete y Project Health.

VARIABLES ANALIZADAS

[TaxonName](#), [TaxonCode](#), [SamplingOperations_code](#), [CodeSite_SamplingOperations](#), [Date_SamplingOperation](#), [Abundance_nbcell](#), [TotalAbundance_SamplingOperation](#), [Abundance_pm](#) (combinaciones diversas).

Project Type, Geographical scope, Project manager, State, Project size, Project organization, BG, On-hold, Percent complete y Project Health.

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 |

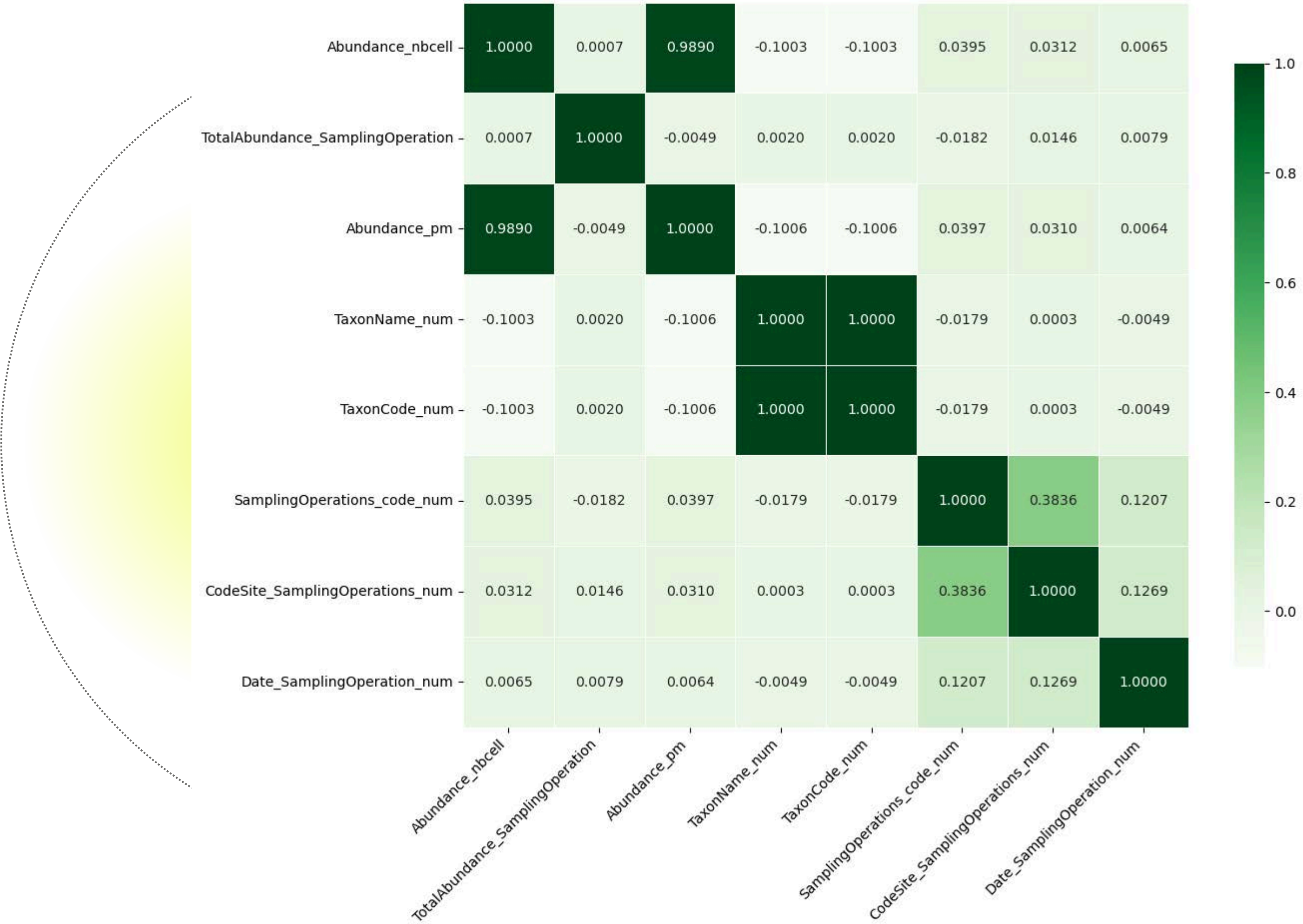
Fue necesario transformar las variables categóricas en variables numéricas. Para ello, se utilizó la jerarquía de frecuencias, asignando valores más bajos a las categorías con mayor frecuencia de aparición.

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 |

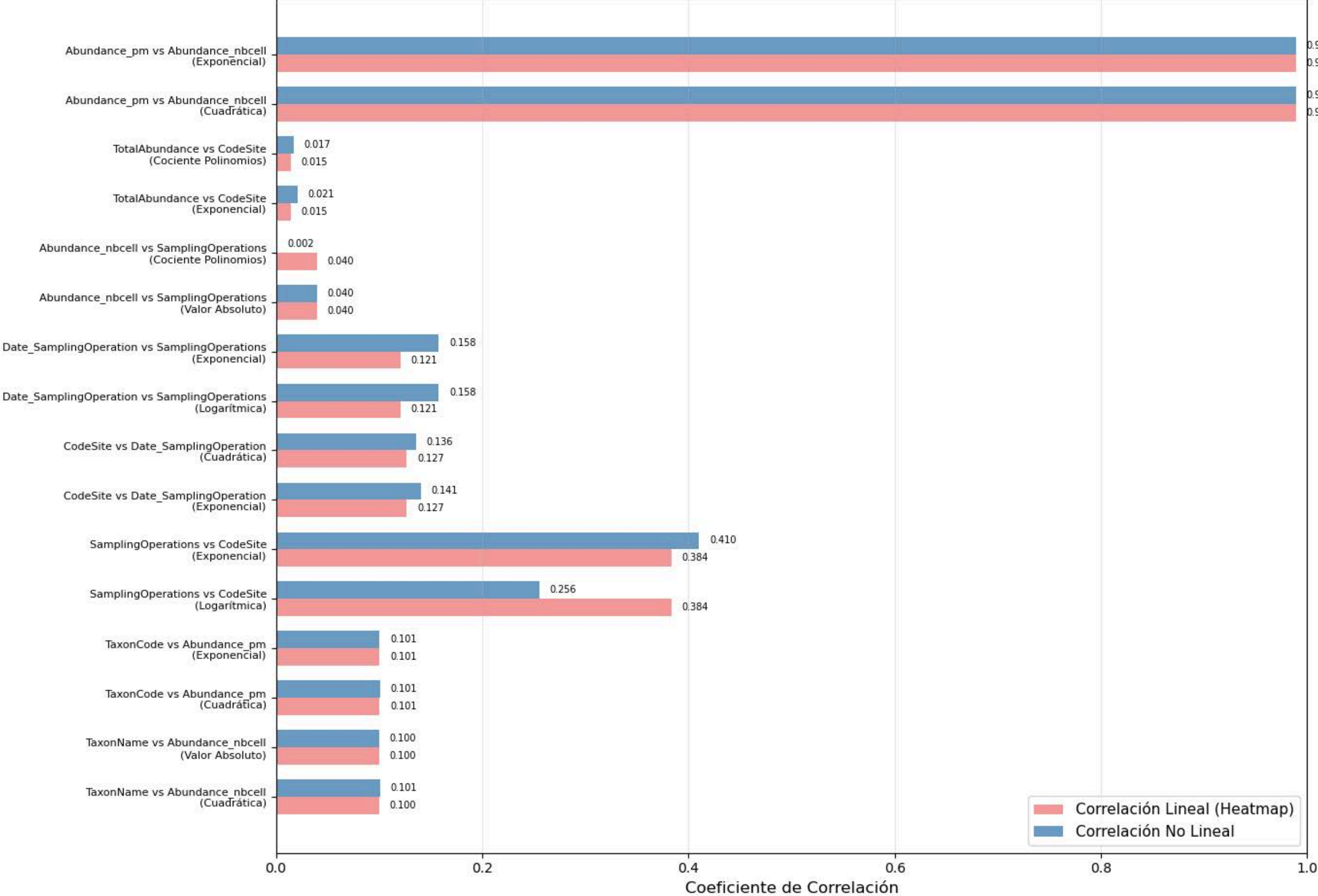
HEATMAP COMPLETO - Matriz de Correlación de Todas las Variables



ACTIVIDAD 3.1

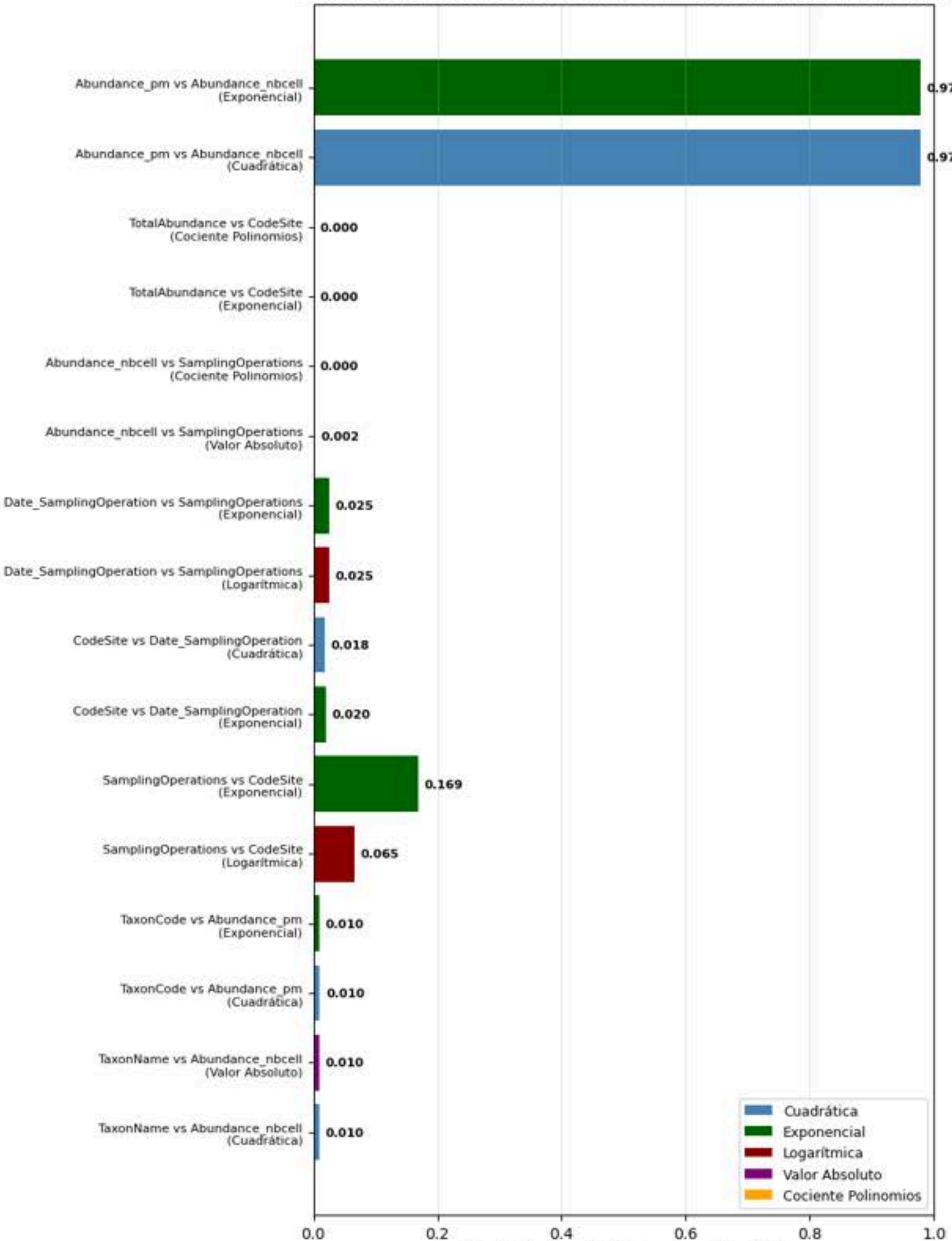
DATATHON

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> |

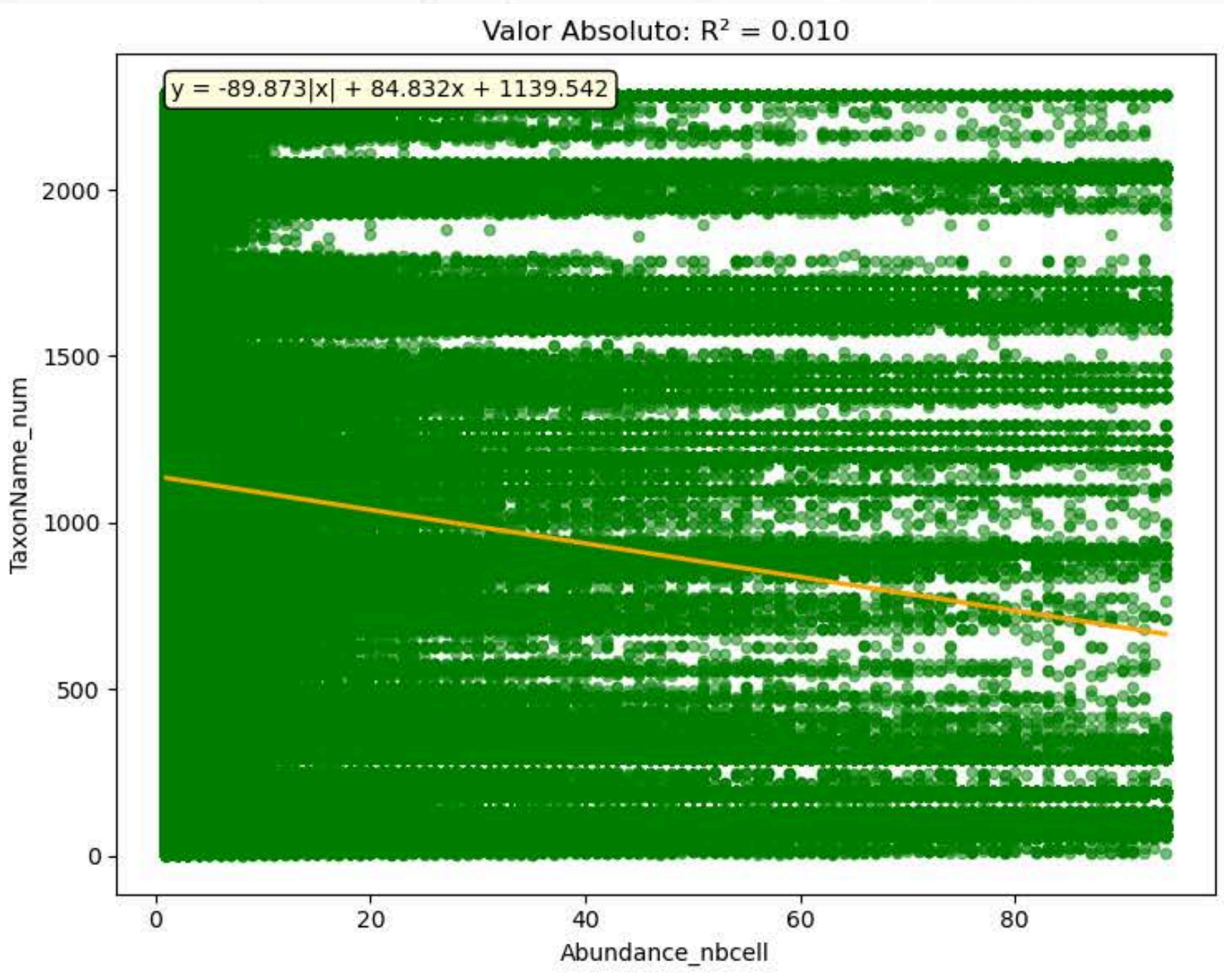
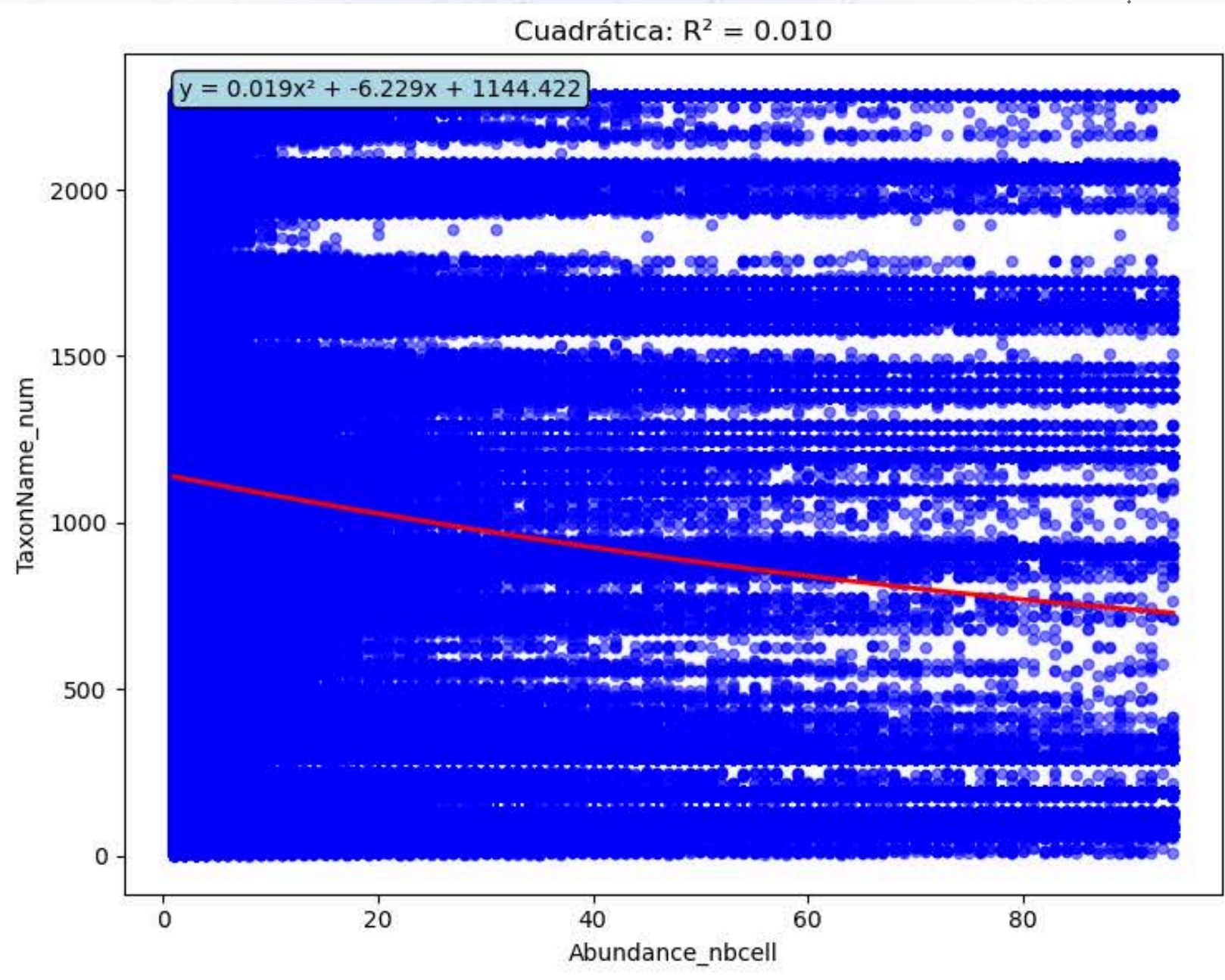
Coefficientes de Determinación - TODOS LOS MODELOS



| | |
|------------------------------|---|
| 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:** Bandas horizontales
- **Alta dispersión:** Los puntos están muy esparcidos
- **Concentración de datos:** Abundance_nbcell

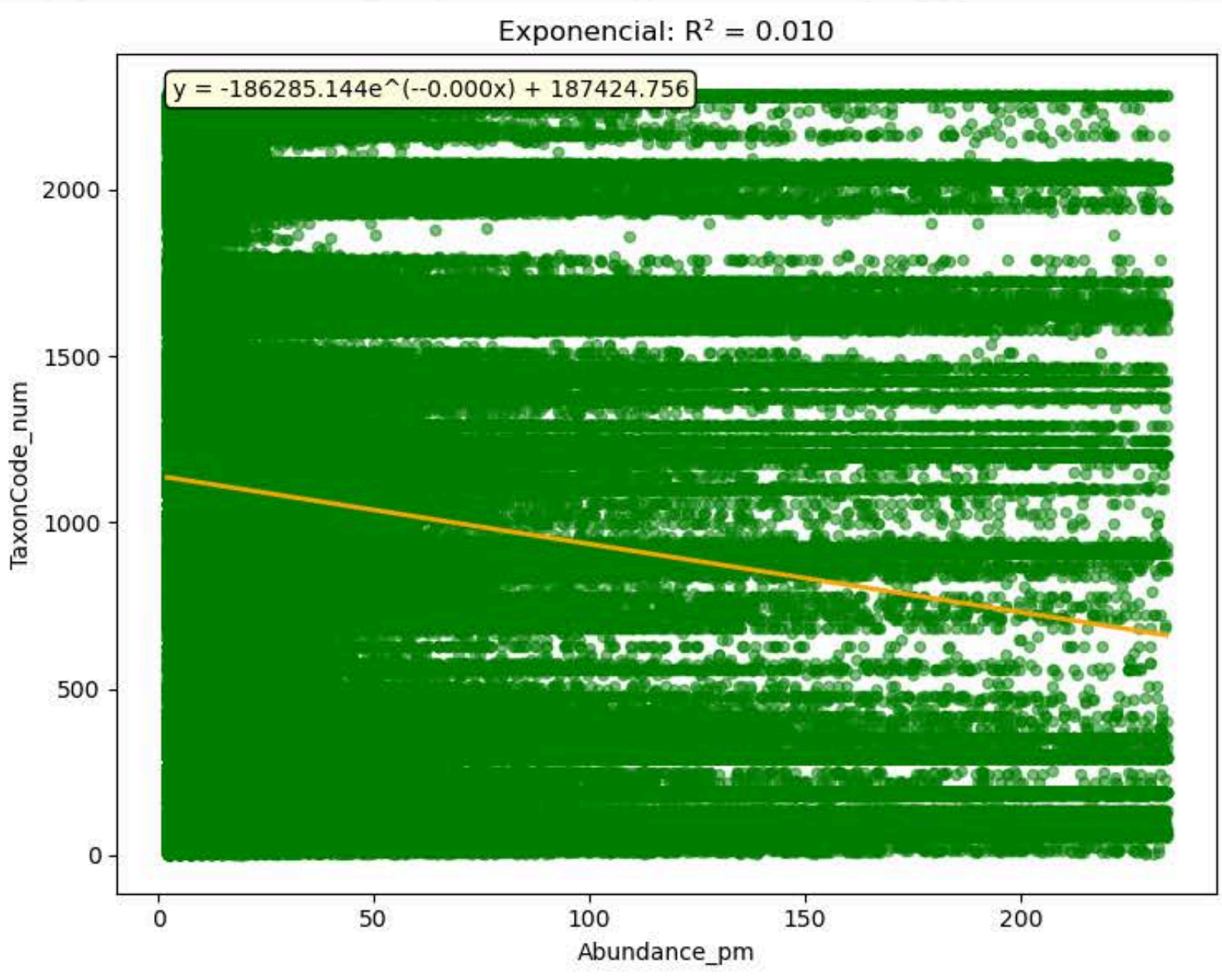
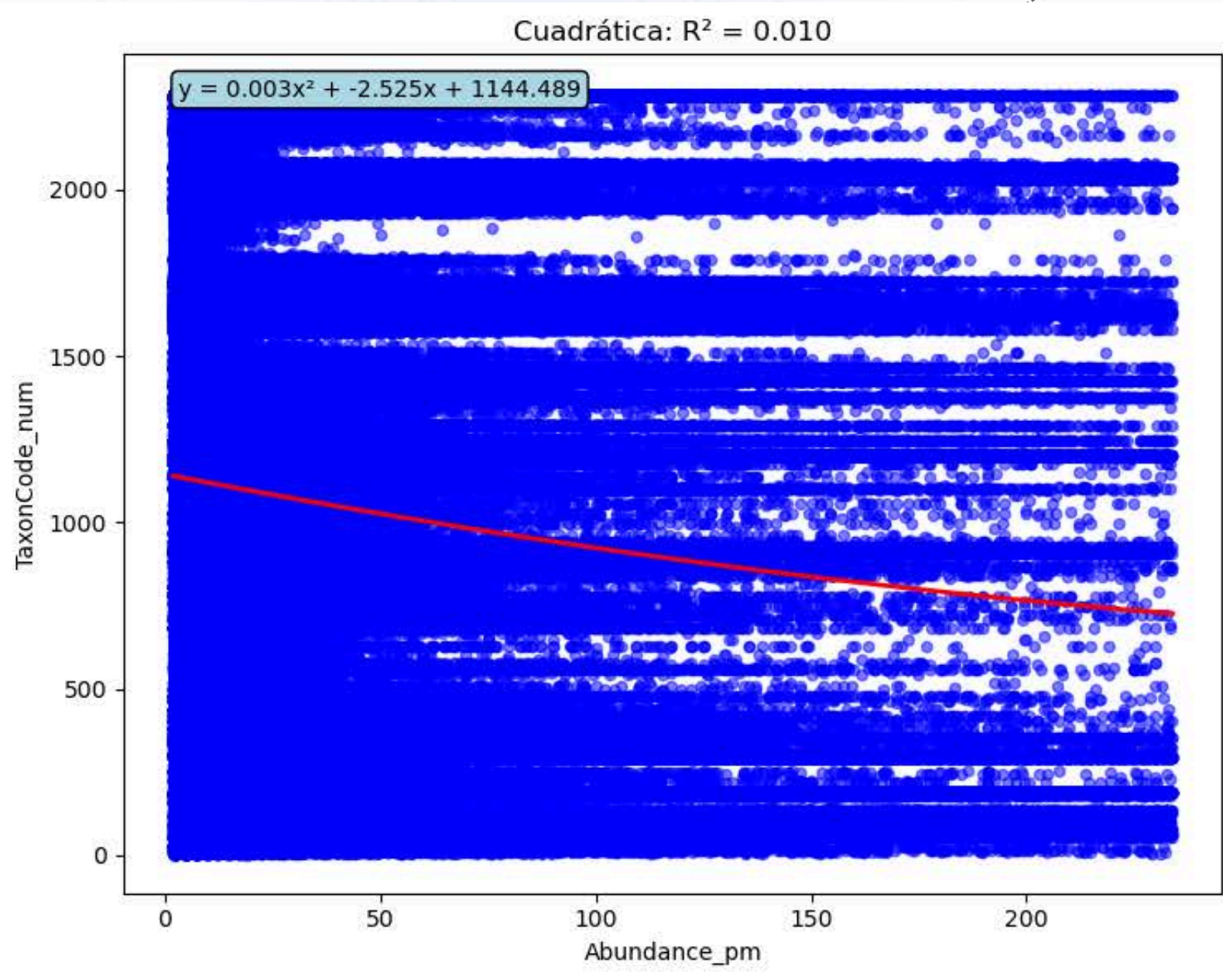


| | |
|--------------------------------------|------------------------|
| Correlación Lineal / Múltiple (r) | $r = -0.1003 / 0.1016$ |
| Correlación No Lineal (r) | $r = 0.1007$ |

| | |
|---------------------------|---------------|
| Correlación L (r) | $r = -0.1003$ |
| Correlación No Lineal (r) | $r = 0.1003$ |

TaxonCode vs Abundance_pm

- Bandas discretas más pronunciadas: TaxonCode_num
- Distribución uniforme: a lo largo del rango de Abundance_pm
- Mayor dispersión horizontal

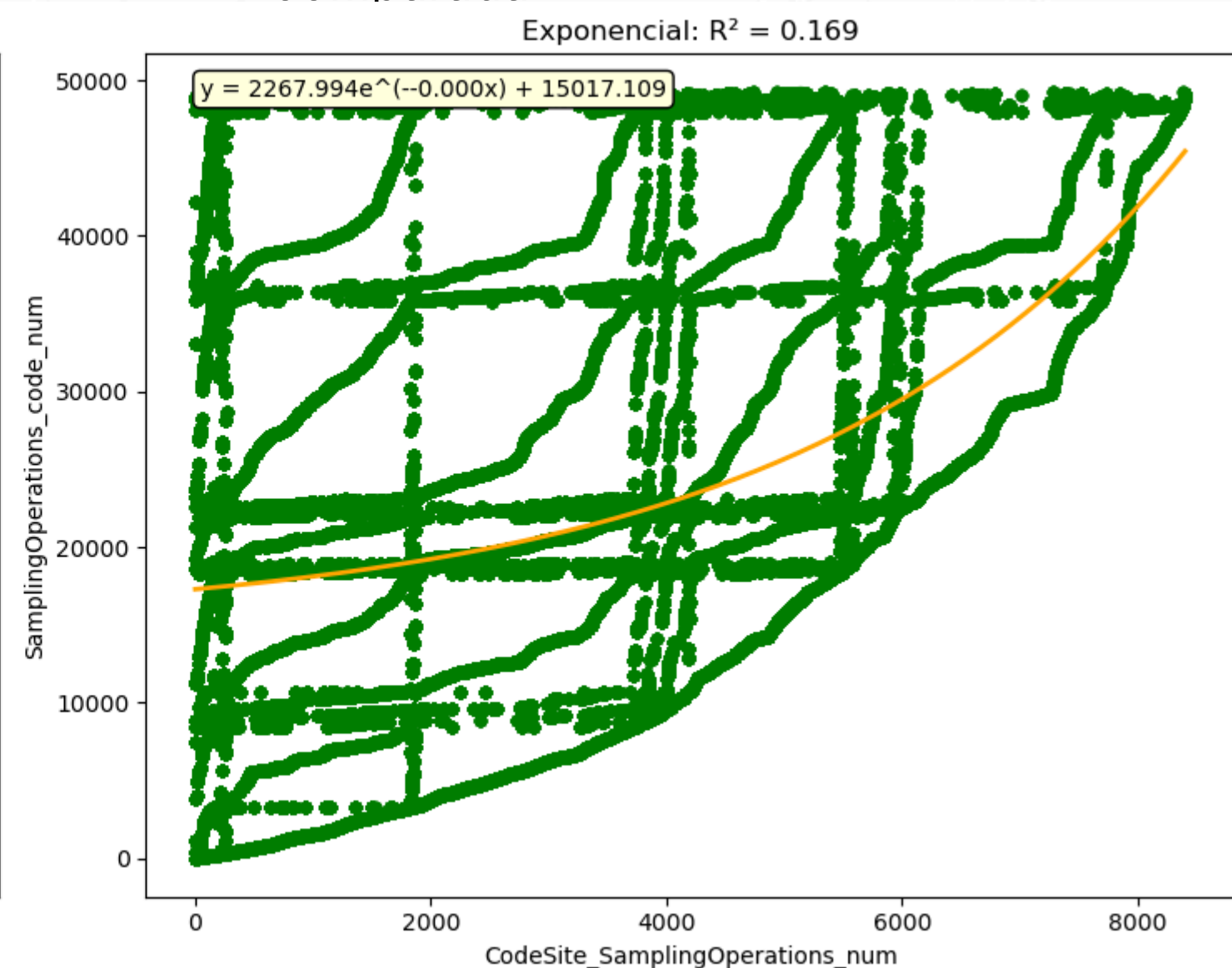
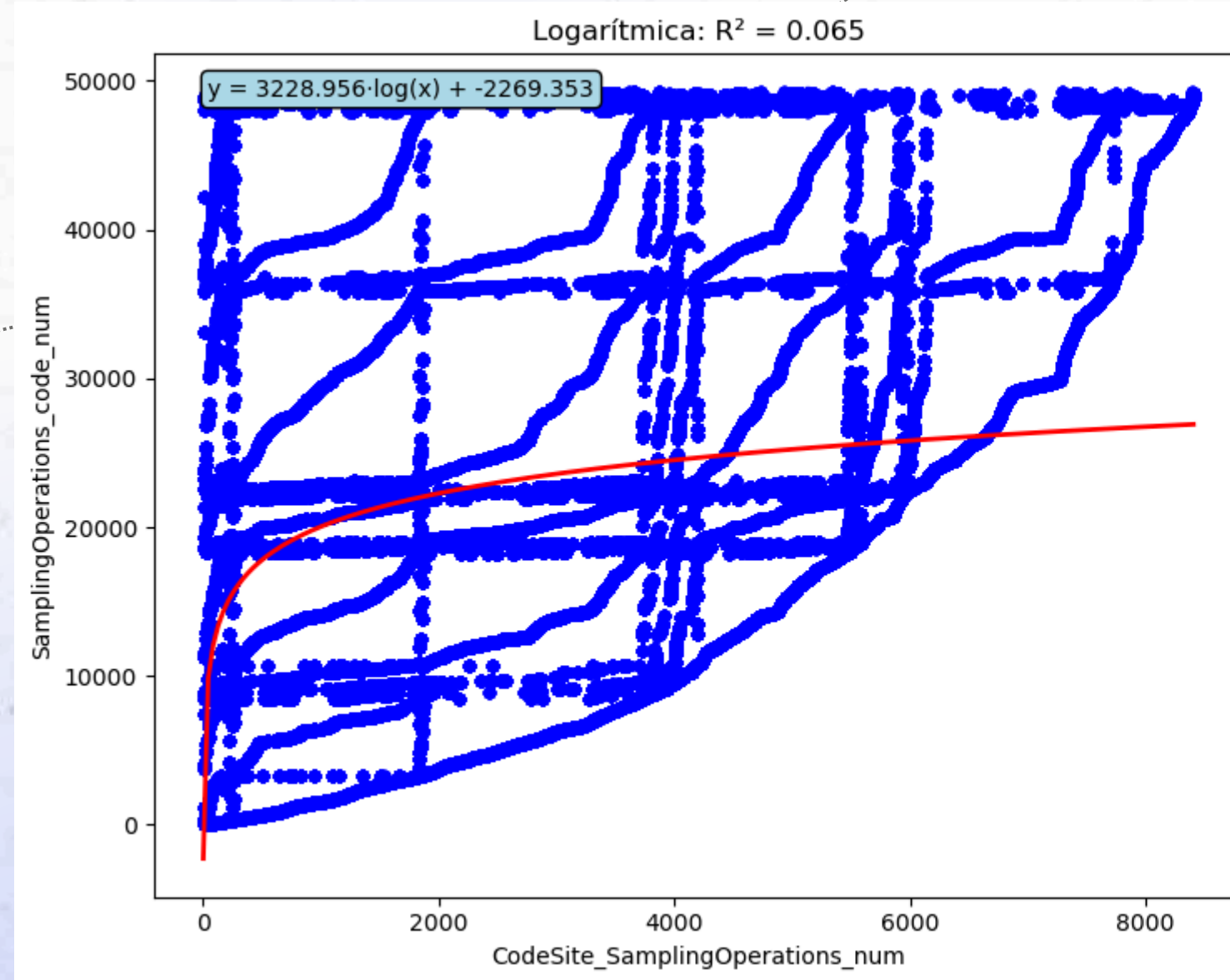


| | |
|--------------------------------------|------------------------|
| Correlación Lineal / Múltiple (r) | $r = -0.1006 / 0.1016$ |
| Correlación No Lineal (r) | $r = 0.1011$ |

| | |
|---------------------------|---------------|
| Correlación L (r) | $r = -0.1006$ |
| Correlación No Lineal (r) | $r = 0.1006$ |

SamplingOperations_code vs CodeSite_SamplingOperetions_num

- Líneas verticales marcadas: Los datos se agrupan en columnas verticales
- Distribución escalonada
- Patrón estructurado: Estructura más organizada comparada

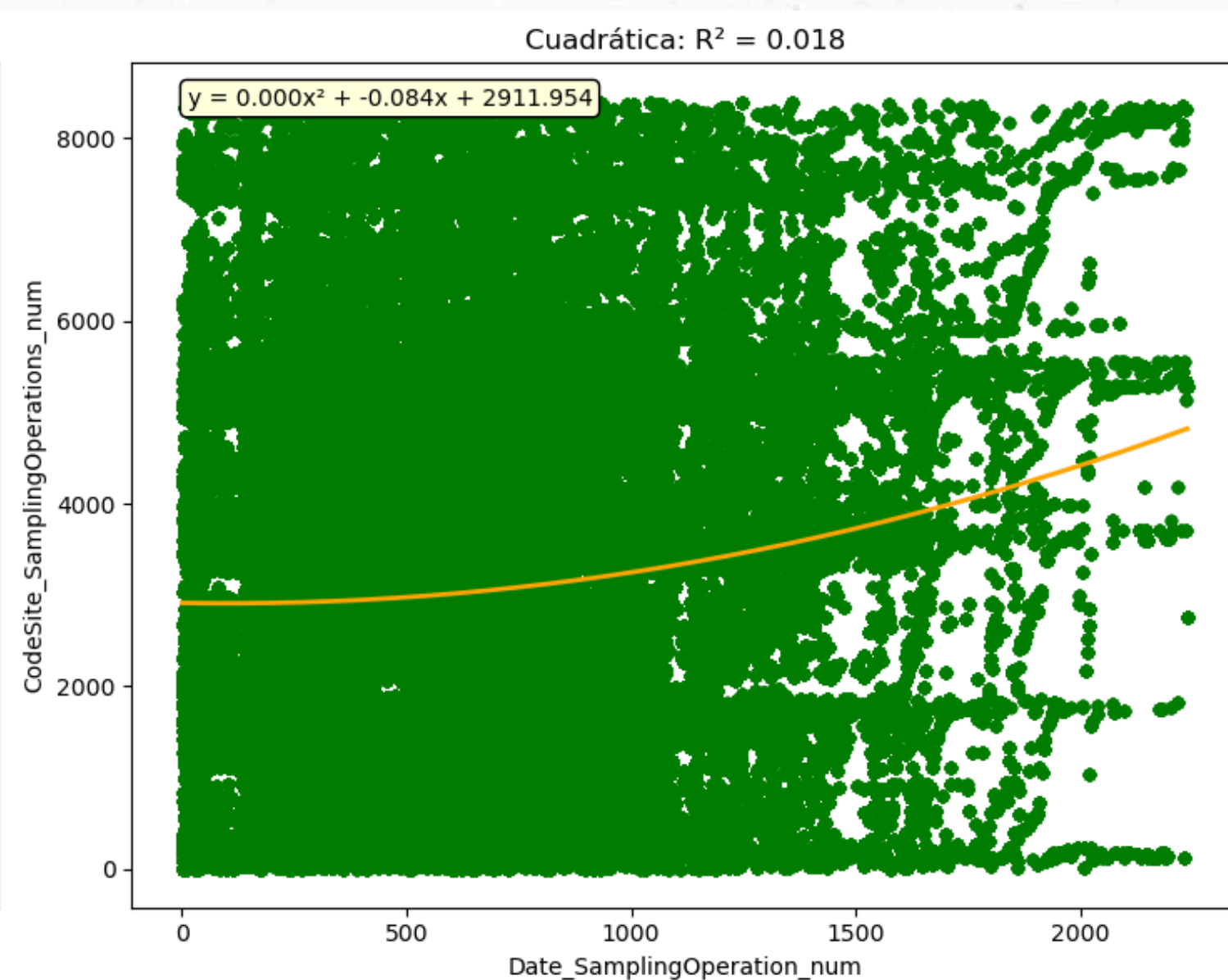
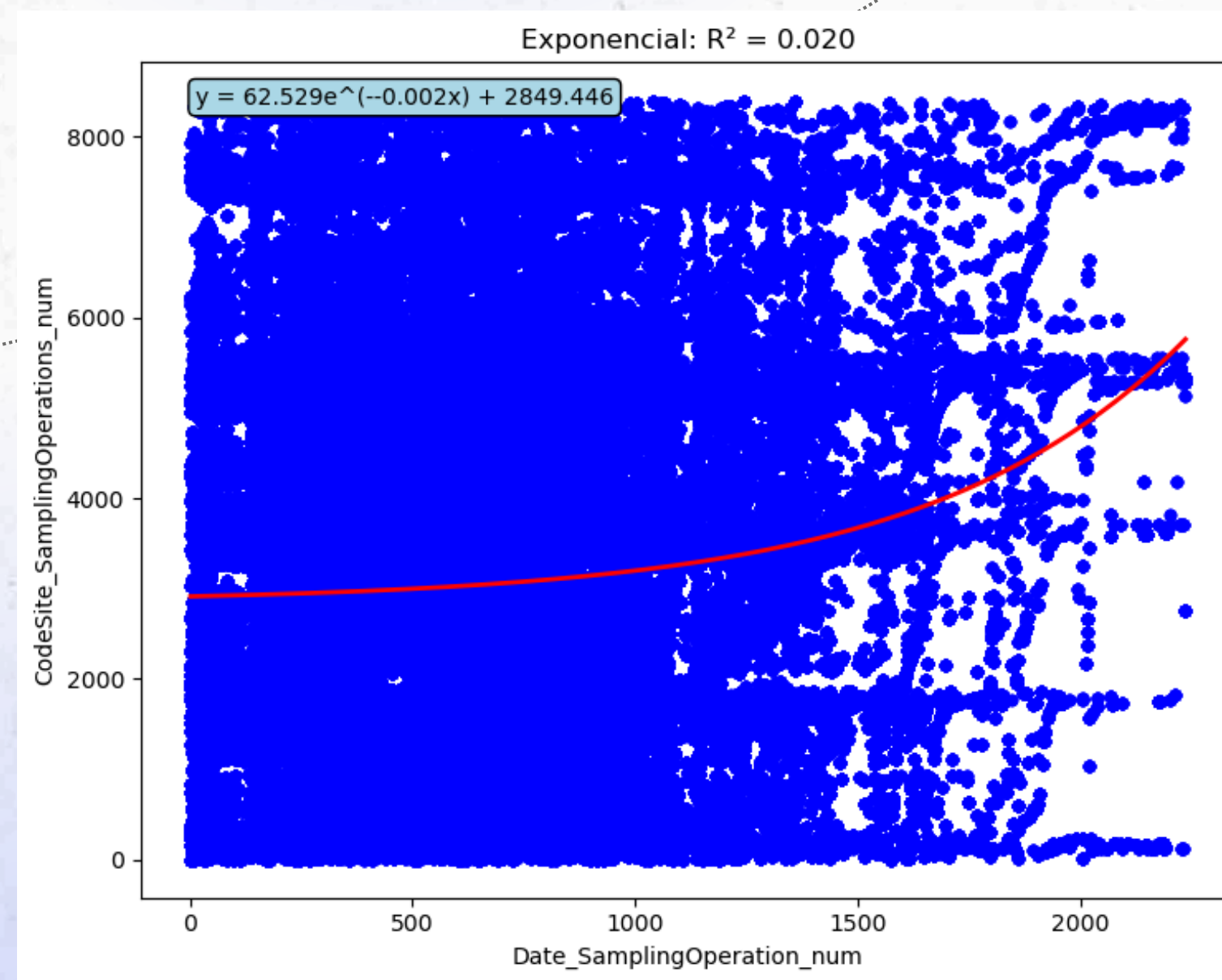


| | |
|--------------------------------------|-----------------------|
| Correlación Lineal / Múltiple (r) | $r = 0.3836 / 0.3903$ |
| Correlación No Lineal (r) | $r = 0.2558$ |

| | |
|---------------------------|--------------|
| Correlación L (r) | $r = 0.3836$ |
| Correlación No Lineal (r) | $r = 0.4105$ |

CodeSite_SamplingOperations vs Date_SamplingOperation_num

- Distribución triangular: Formación de puntos
- Mayor dispersión
- Concentración en valores bajos: Inferior izquierda

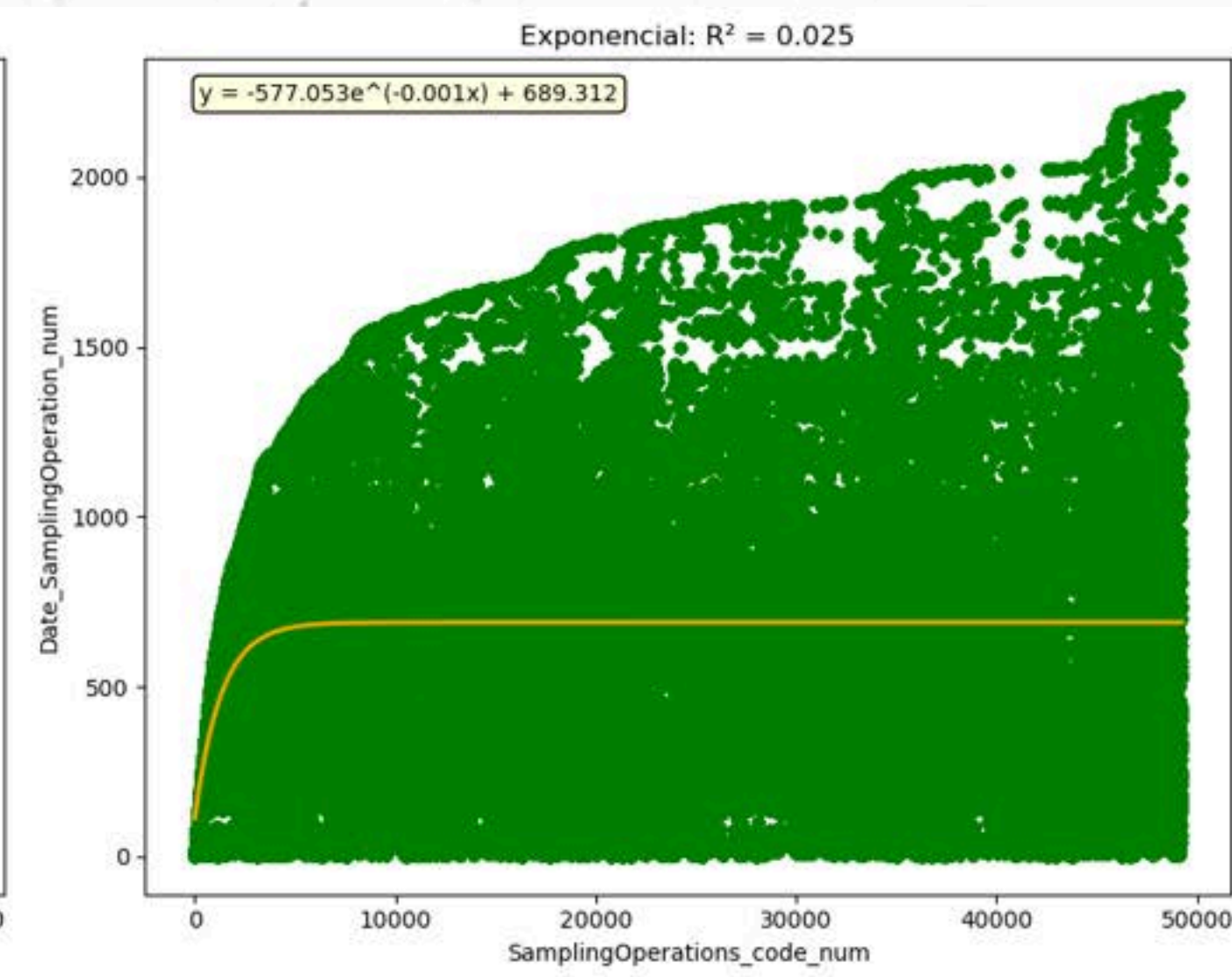
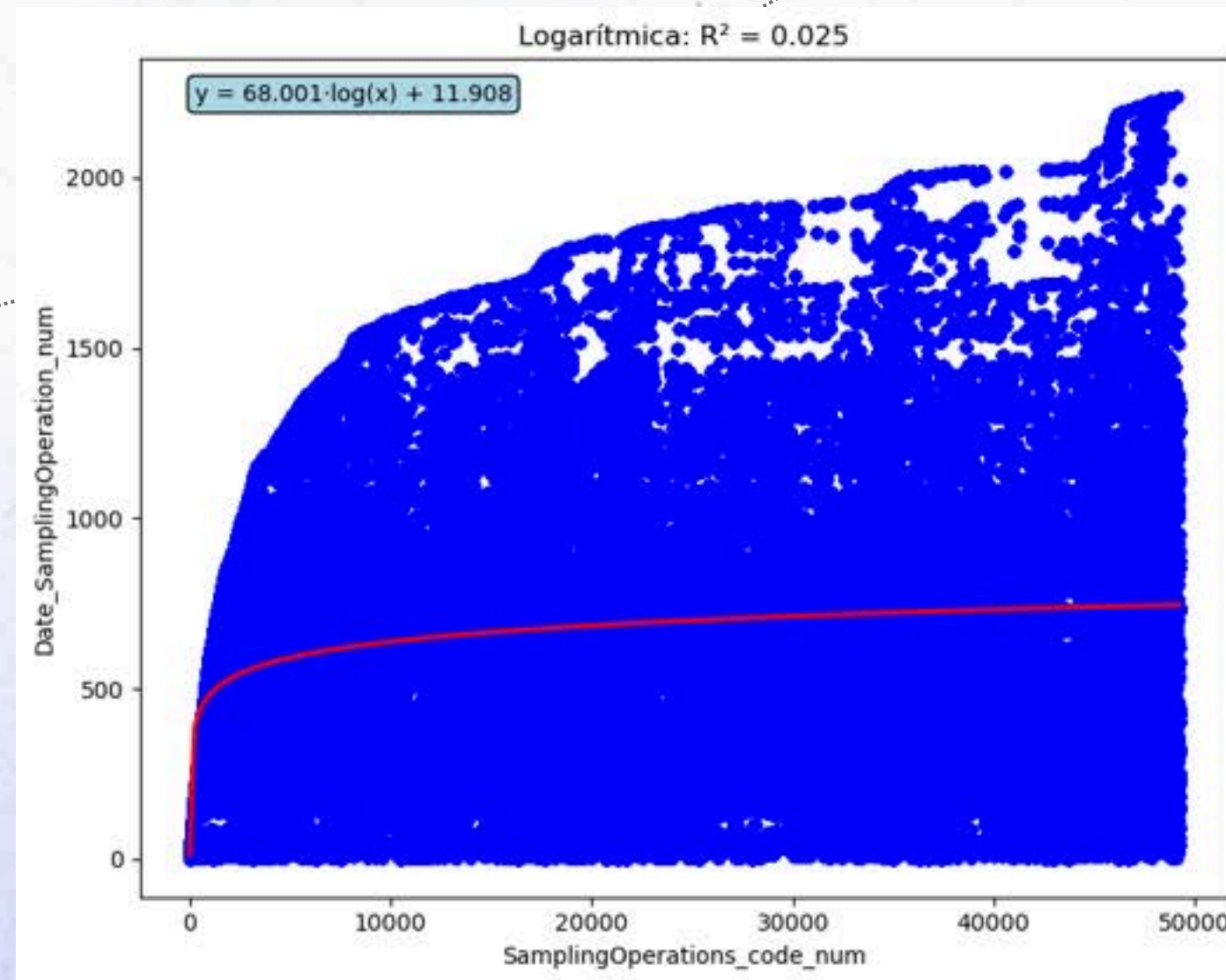


| | |
|--------------------------------------|-----------------------|
| Correlación Lineal / Múltiple (r) | $r = 0.1269 / 0.3920$ |
| Correlación No Lineal (r) | $r = 0.1409$ |

| | |
|---------------------------|--------------|
| Correlación L (r) | $r = 0.1269$ |
| Correlación No Lineal (r) | $r = 0.1360$ |

Date_SamplingOperation_num vs SamplingOperations_code_num

- **Bandas horizontales discretas:** Líneas horizontales muy marcadas
- **Concentración masiva**
- **Expansión gradual:** Conforme aumenta SamplingOperations_code, se van "abriendo" más fechas

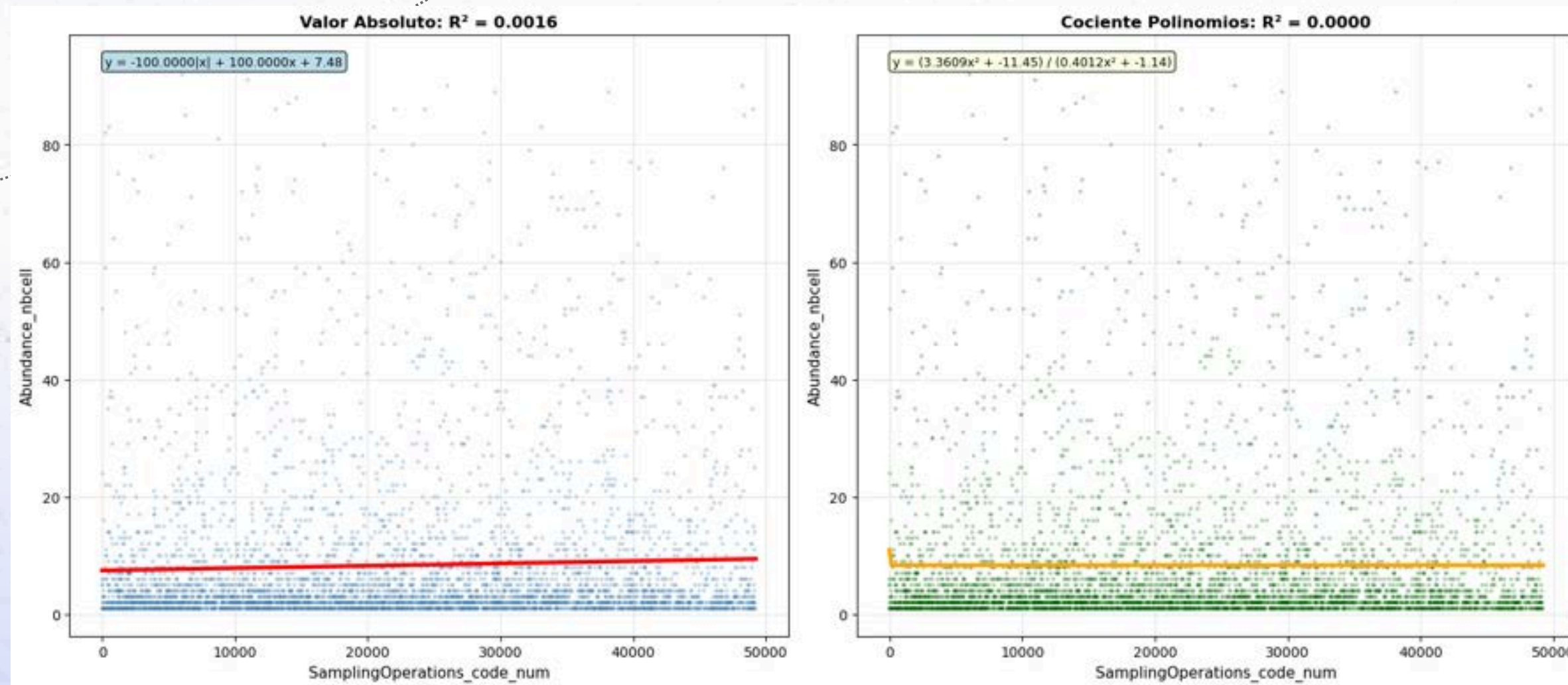


| | |
|-----------------------------------|-----------------------|
| Correlación Lineal / Múltiple (r) | $r = 0.1207 / 0.1490$ |
| Correlación No Lineal (r) | $r = 0.1580$ |

| | |
|---------------------------|--------------|
| Correlación L (r) | $r = 0.1207$ |
| Correlación No Lineal (r) | $r = 0.1580$ |

Abundance_nbcell vs SamplingOperations_code_num

- **Distribución horizontal dominante:** Abundance_nbcell
- **Patrón de bandas discretas:** Se agrupan en líneas horizontales separadas.
- **Concentración extrema en la base**

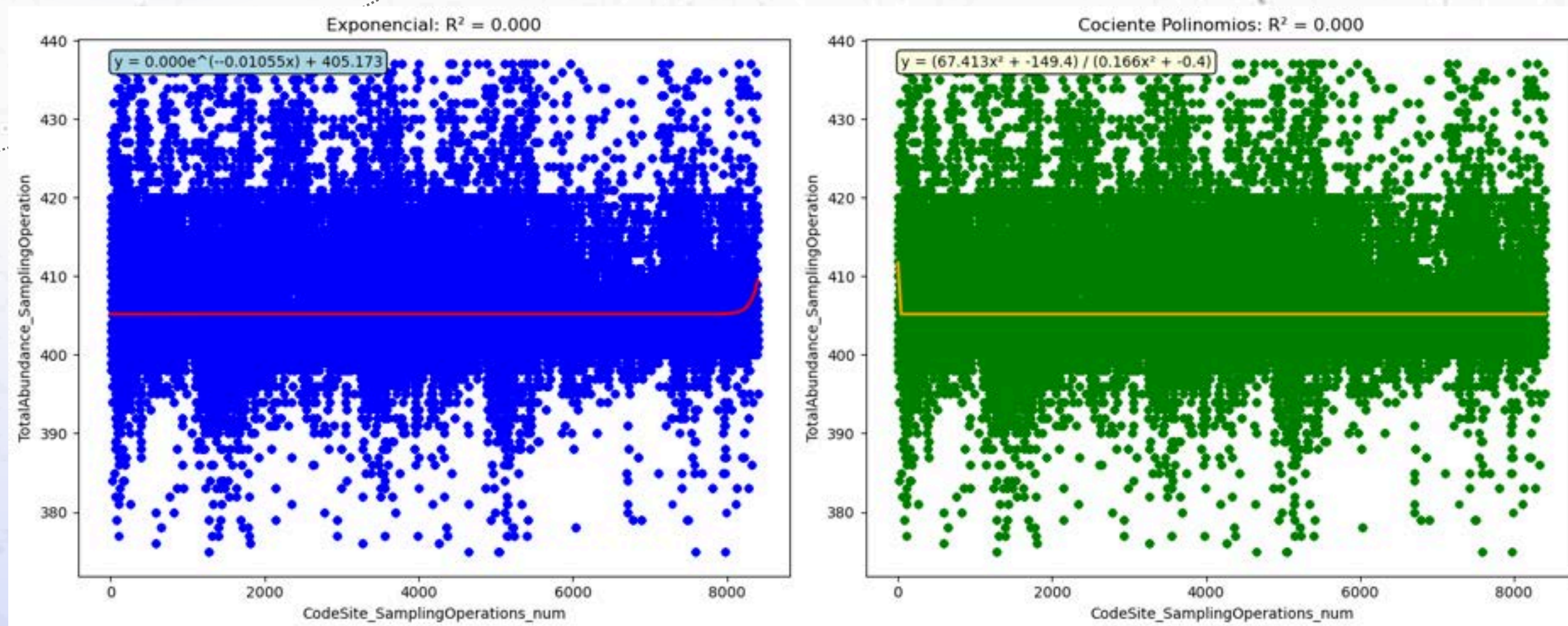


| | |
|--------------------------------------|-----------------------|
| Correlación Lineal / Múltiple (r) | $r = 0.0395 / 0.1040$ |
| Correlación No Lineal (r) | $r = 0.0395$ |

| | |
|---------------------------|--------------|
| Correlación L (r) | $r = 0.0395$ |
| Correlación No Lineal (r) | $r = 0.0015$ |

TotalAbundance_SamplingOperation vs CodeSite_SamplingOperations_num

- Distribución rectangular densa: Patrón rectangular muy denso con límites bien definidos.
- Concentración superior
- Banda principal dominante

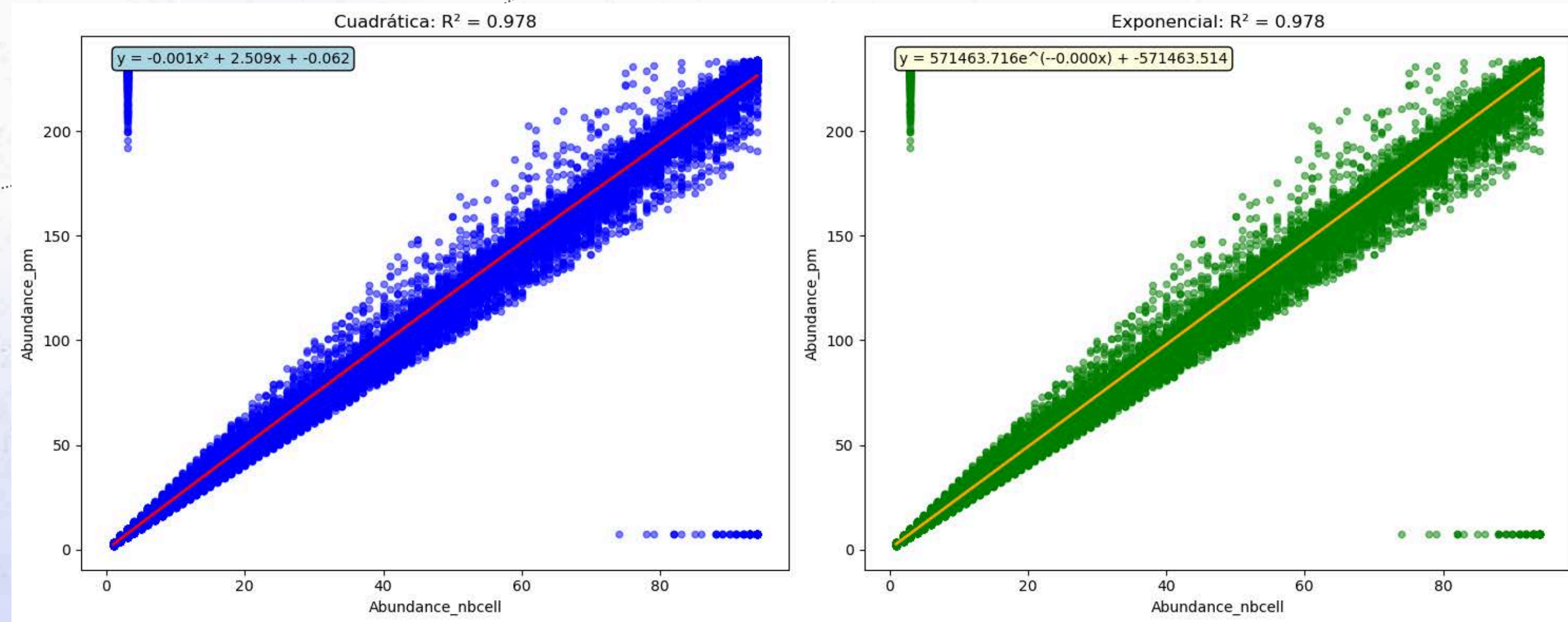


| | |
|--------------------------------------|-----------------------|
| Correlación Lineal / Múltiple (r) | $r = 0.0146 / 0.0290$ |
| Correlación No Lineal (r) | $r = 0.0211$ |

| | |
|---------------------------|--------------|
| Correlación L (r) | $r = 0.0146$ |
| Correlación No Lineal (r) | $r = 0.0174$ |

Abundance_pm vs Abundance_nbccl

- **Correlación lineal muy fuerte:** Nube alargada con una tendencia lineal positiva
- **Dispersión creciente:** Aumenta ligeramente con valores más altos
- **Concentración en origen:** Alta densidad de puntos



| | |
|--------------------------------------|-----------------------|
| Correlación Lineal / Múltiple (r) | $r = 0.9890 / 0.1080$ |
| Correlación No Lineal (r) | $r = 0.9891$ |

| | |
|---------------------------|--------------|
| Correlación L (r) | $r = 0.9890$ |
| Correlación No Lineal (r) | $r = 0.9890$ |

ANÁLISIS DE INSIGHTS

**Cada diatomea
está correctamente
identificada y
codificada**

TAXONNAME Y TAXONCODE:
CADA ORGANISMO TIENE IDENTIDAD BIOLÓGICA
PRECISA

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

CORRELACIÓN DIVERSIDAD-VARIABILIDAD: $R=-0.160$

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

COEFICIENTE DE VARIACIÓN TEMPORAL: 0.0469

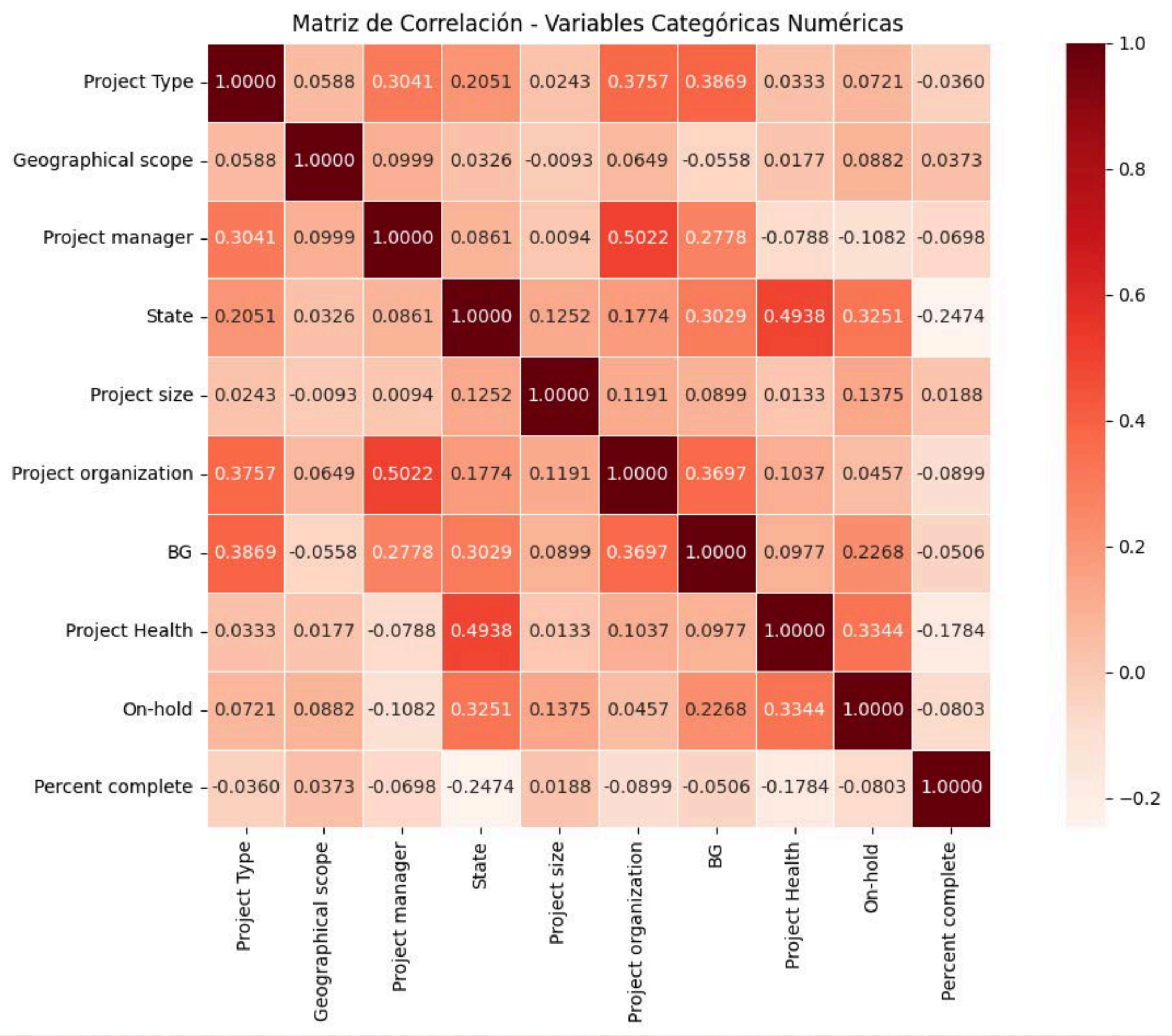
**La abundancia
total es
independiente de
qué especies hay**

CORRELACIÓN ABUNDANCIA TOTAL ↔
DIVERSIDAD: $R=-0.048$
SI UNA ESPECIE AUMENTA, OTRAS DEBEN
DISMINUIR PROPORCIONALMENTE

**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

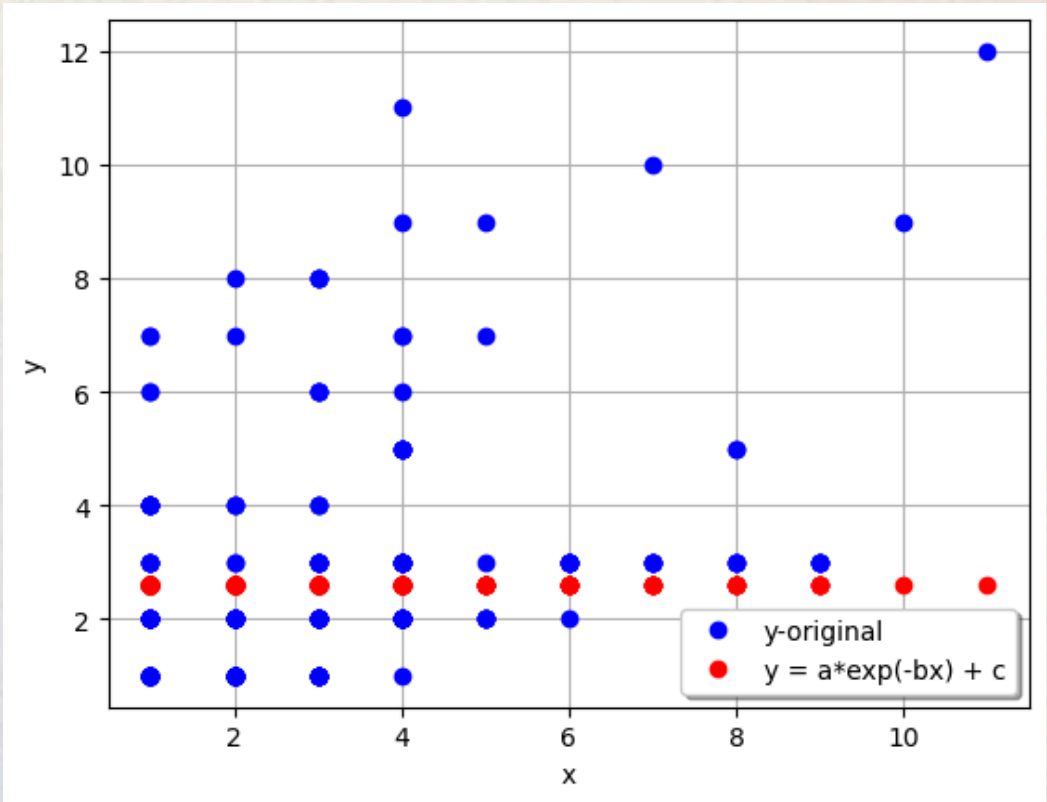
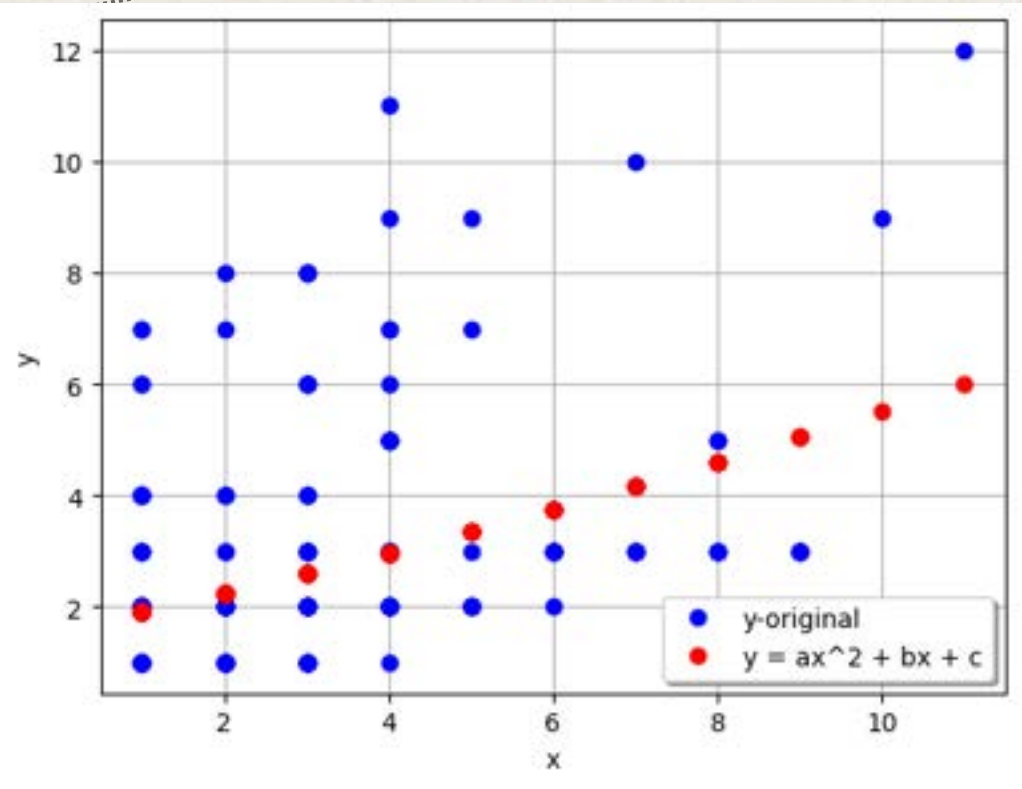


| | |
|------------------------------|--|
| Total de modelos analizados: | 20 |
| Mejor R²: | 0.4473 |
| Mejor correlación: | 0.6688 |
| R² promedio: | 0.1720 |
| Correlación promedio: | 0.3773 |
| MEJOR MODELO: | Variables: State vs Project Health Función: Senoidal R²: 0.4473 |
| TIPOS DE FUNCIÓN: | Cuadrática: 8 modelo(s) Valor Absoluto: 4 modelo(s) Cociente Polinomios: 3 modelo(s) Logarítmica: 2 modelo(s) Exponencial: 1 modelo(s) Senoidal: 1 modelo(s) Polinomial Inversa: 1 modelo(s) |

BG - Project Type

MODELOS NO LINEALES:

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



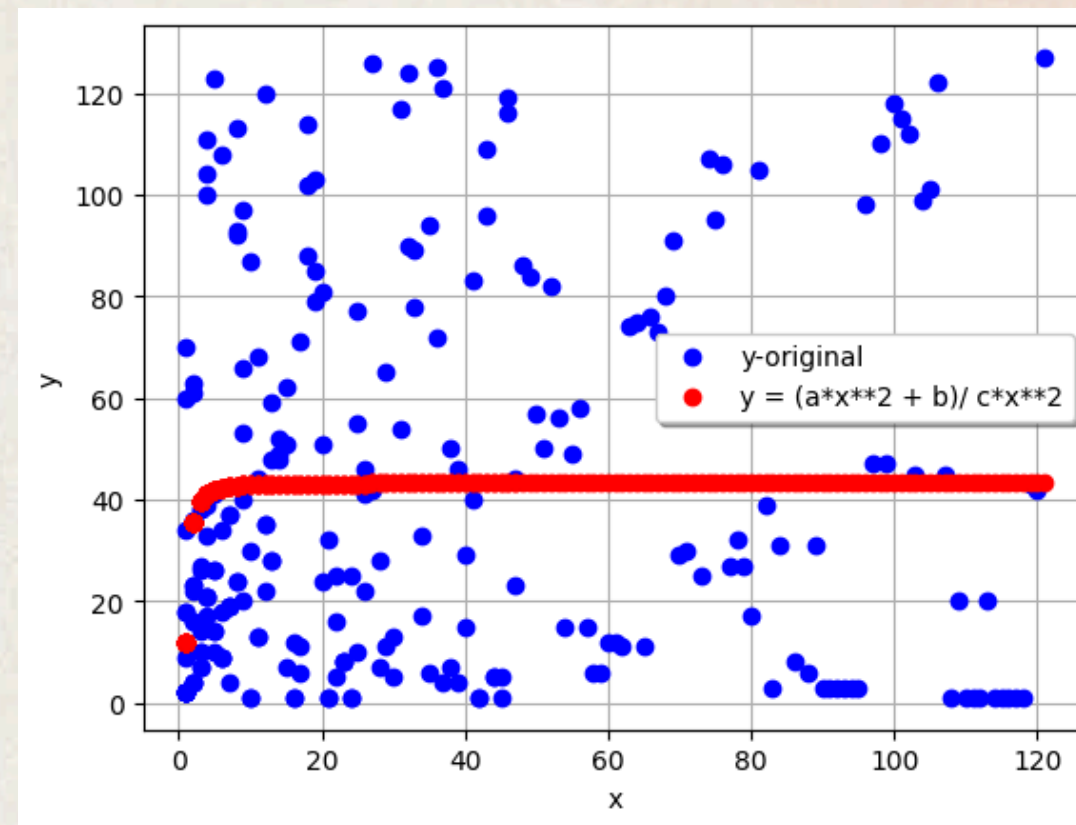
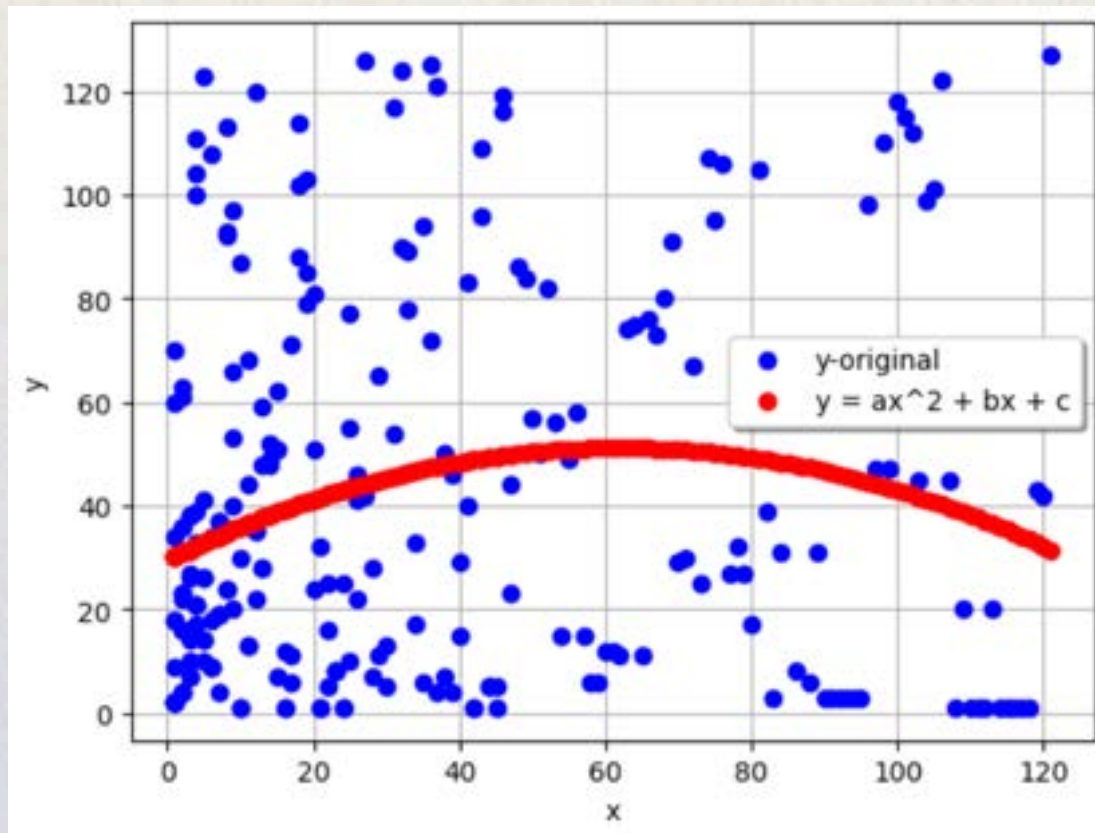
- **Resultados:**
- **Cuadrático:** $R^2 = 0.1502$, $r = 0.3876$
- **Exponencial:** $R^2 = 3.3306$, $r = 1.8250$
- La función cuadrática es la mejor ya que a pesar que esta no crece mucho los datos de la formula exponencial son exagerados para una correlación.
- **Visualmente:** alta dispersión sin tendencia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

| | |
|--------------------|--------|
| Funcion R2 | 0.1502 |
| Correlación NL (r) | 0.386 |
| Correlación L (r) | 0.3876 |

Project manager - Geographical scope

MODELOS NO LINEALES:

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



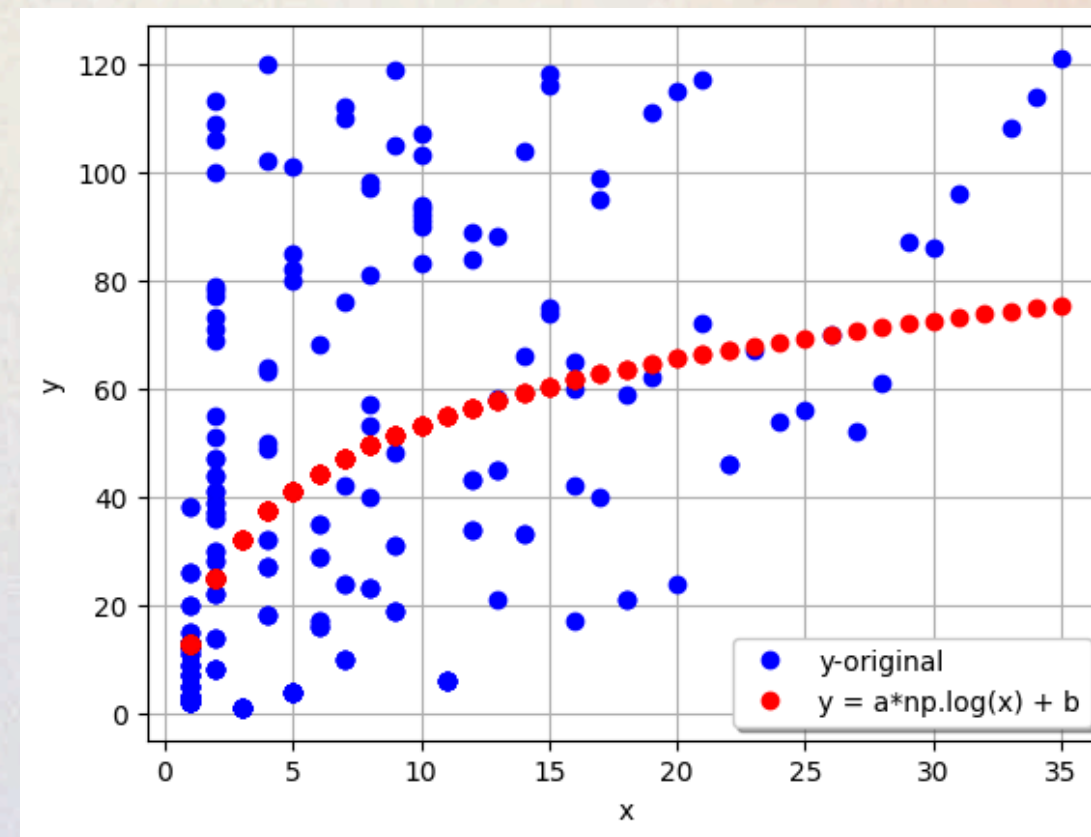
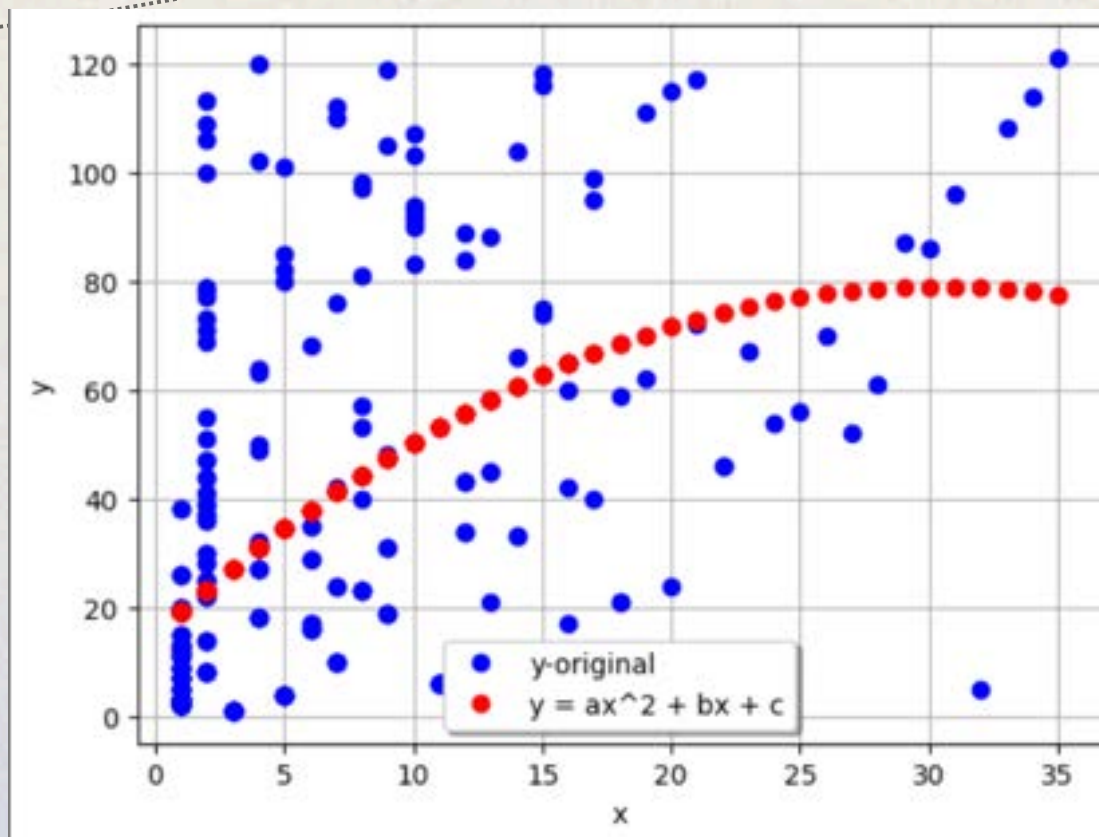
- **Resultados:**
- **Cuadrático:** $R^2 = 0.0365$, $r = 0.1912$
- **Co entre Polinomios:** $R^2 = 0.0407$, $r = 0.2018$
- **Interpretación:** La correlación en ambas es la mejor entre las demás ambas mejoran.
- **Visualmente:** alta dispersión con tendencia clara y determinada con agrupamientos definidos.

| | |
|--------------------|--------|
| Funcion R2 | 0.0407 |
| Correlación NL (r) | 0.0999 |
| Correlación L (r) | 0.2018 |

Project organization - Project manager

MODELOS NO LINEALES:

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



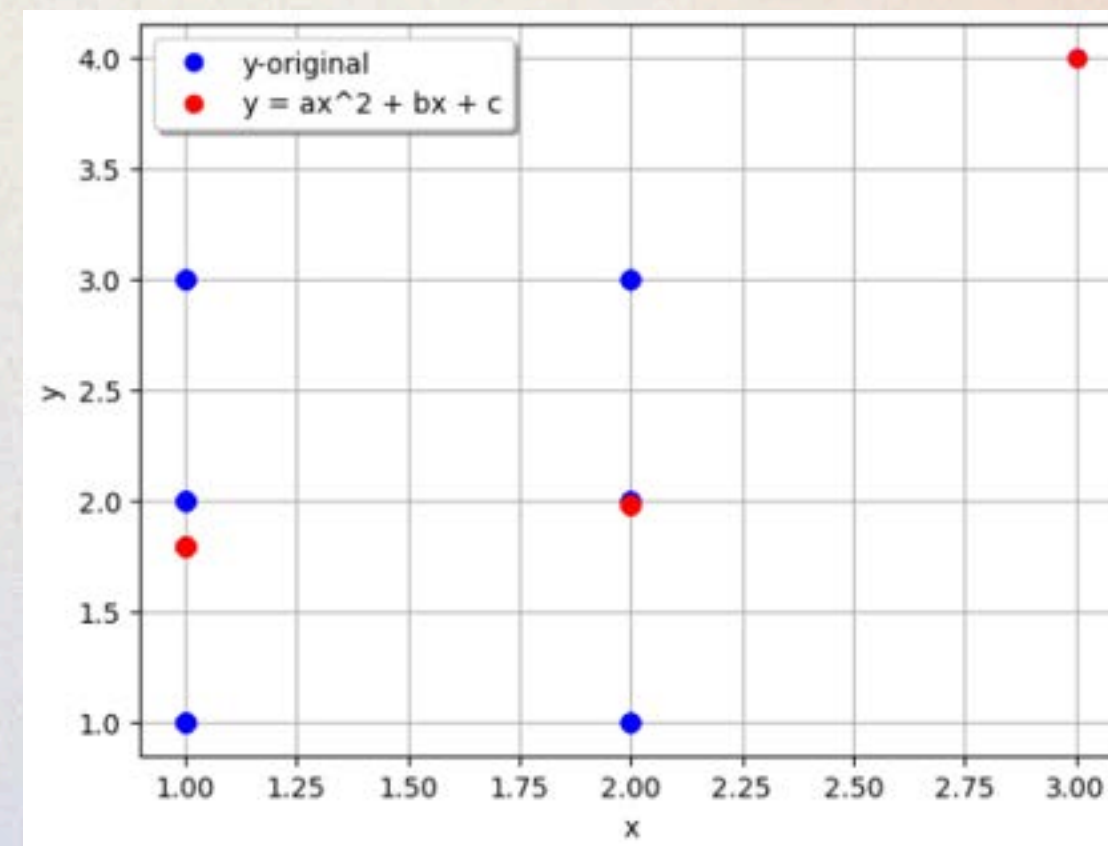
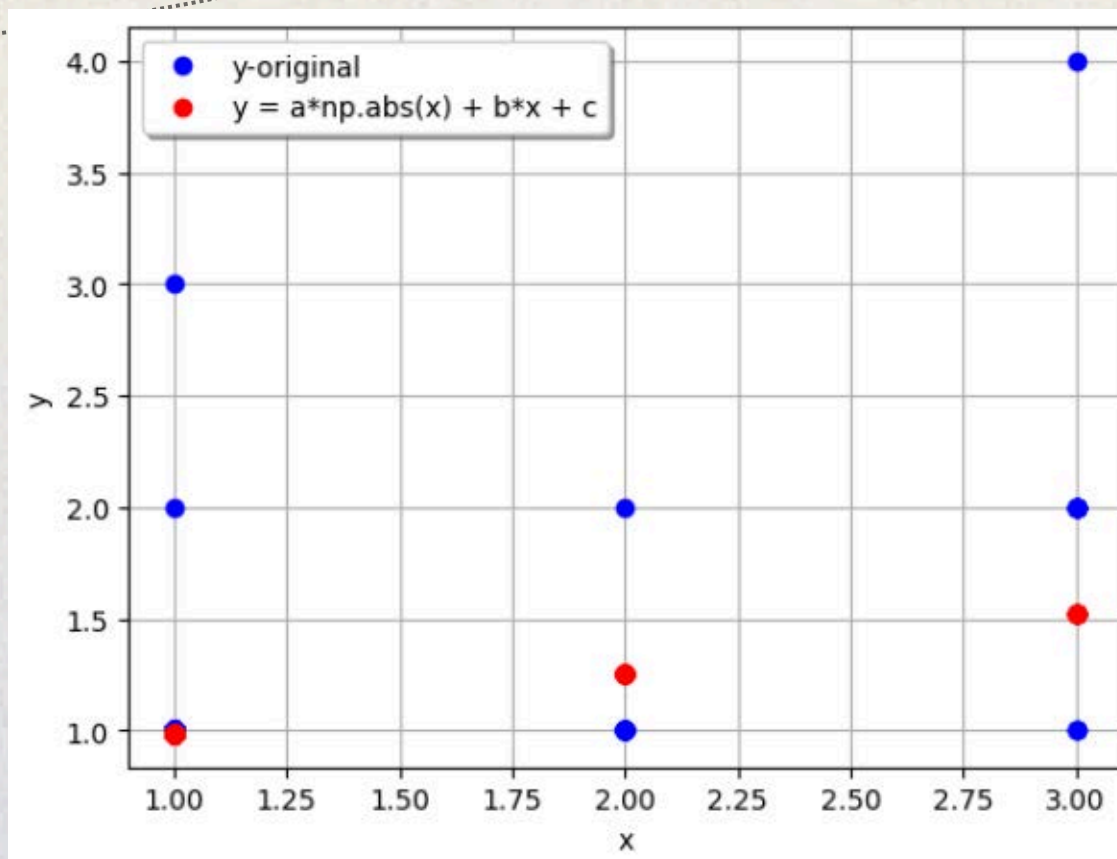
- **Resultados:**
- **Cuadrático:** $R^2 = 0.2723$, $r = 0.5219$
- **Logaritmica:** $R^2 = 0.2885$, $r = 0.5371$
- **Interpretación:** Las correlaciones mejoran en ambos casos.
- **Visualmente:** alta dispersión con tendencia clara y agrupamientos mejorados.

| | |
|--------------------|--------|
| Funcion R2 | 0.2885 |
| Correlación NL (r) | 0.5022 |
| Correlación L (r) | 0.5371 |

Project Health - State

MODELOS NO LINEALES:

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



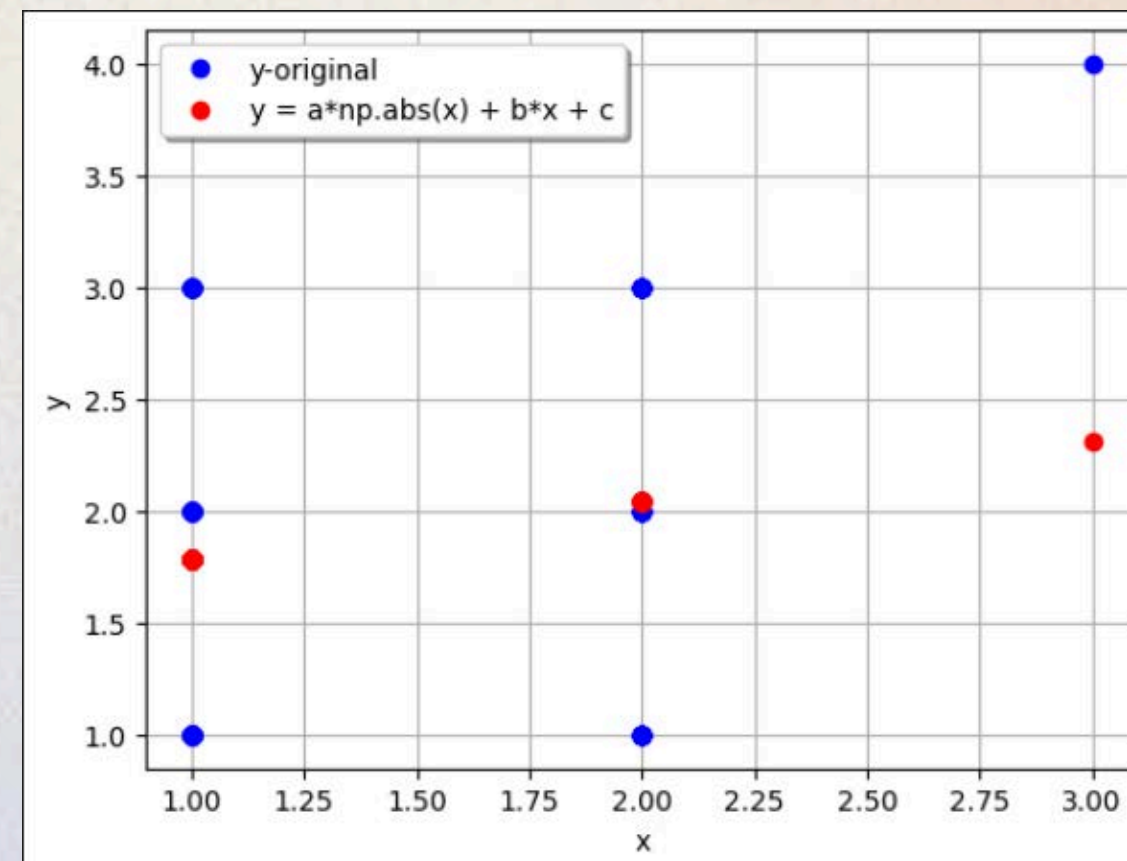
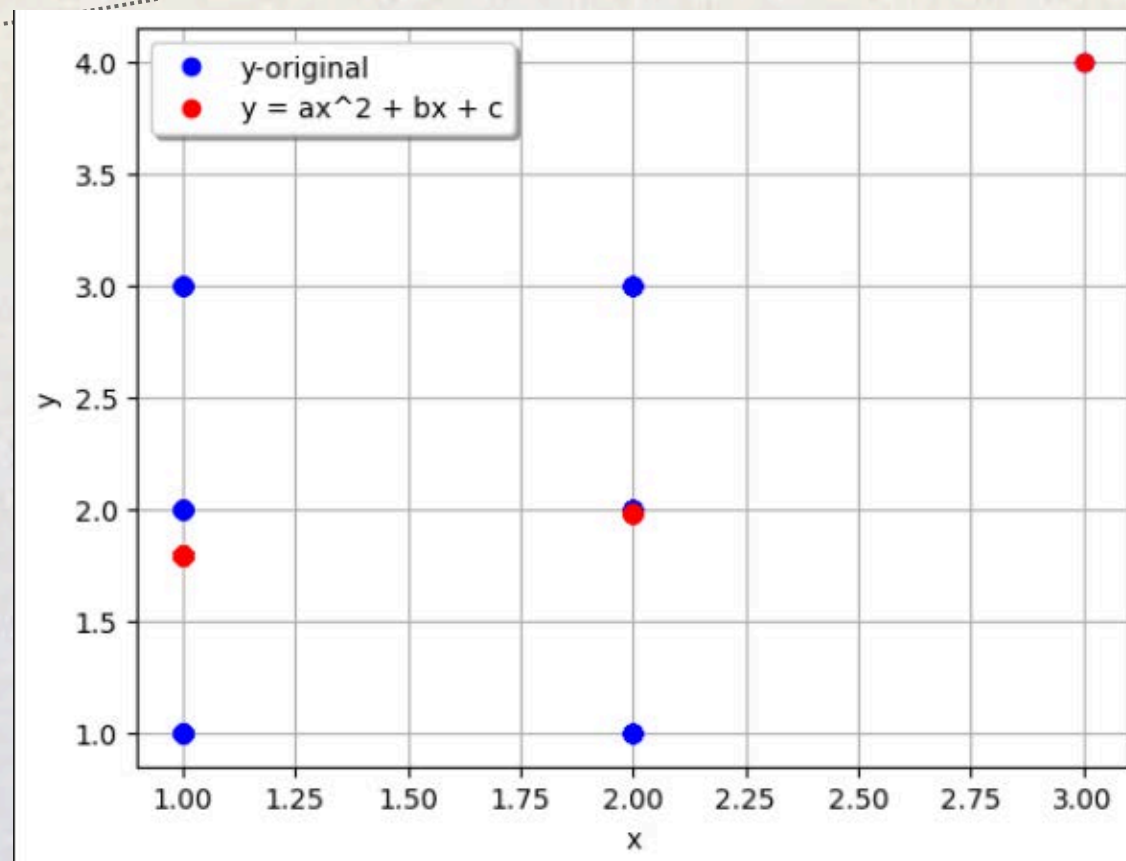
- **Resultados:**
- **Senoidal:** $R^2 = 0.4473$, $r = 0.6688$
- **Val. Absoluto:** $R^2 = 0.2438$ $r = 0.4937$
- **Interpretación:** De las dos mayores correlaciones la senodial es la unica coherente ya que es la unica que supera a la correlación inicial.
- **Visualmente:** alta dispersión sin tendencia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

| | |
|--------------------|--------|
| Funcion R2 | 0.4473 |
| Correlación NL (r) | 0.4937 |
| Correlación L (r) | 0.6688 |

On-hold - Project size

MODELOS NO LINEALES:

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



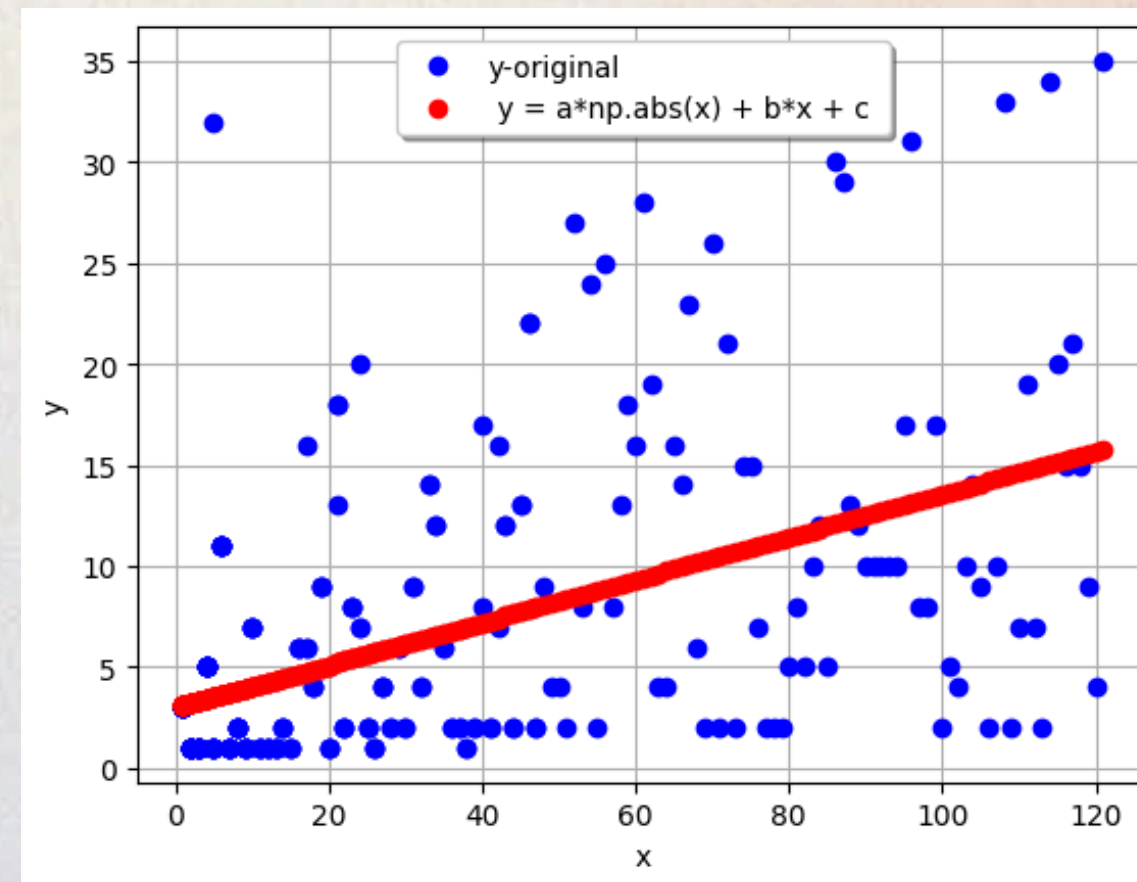
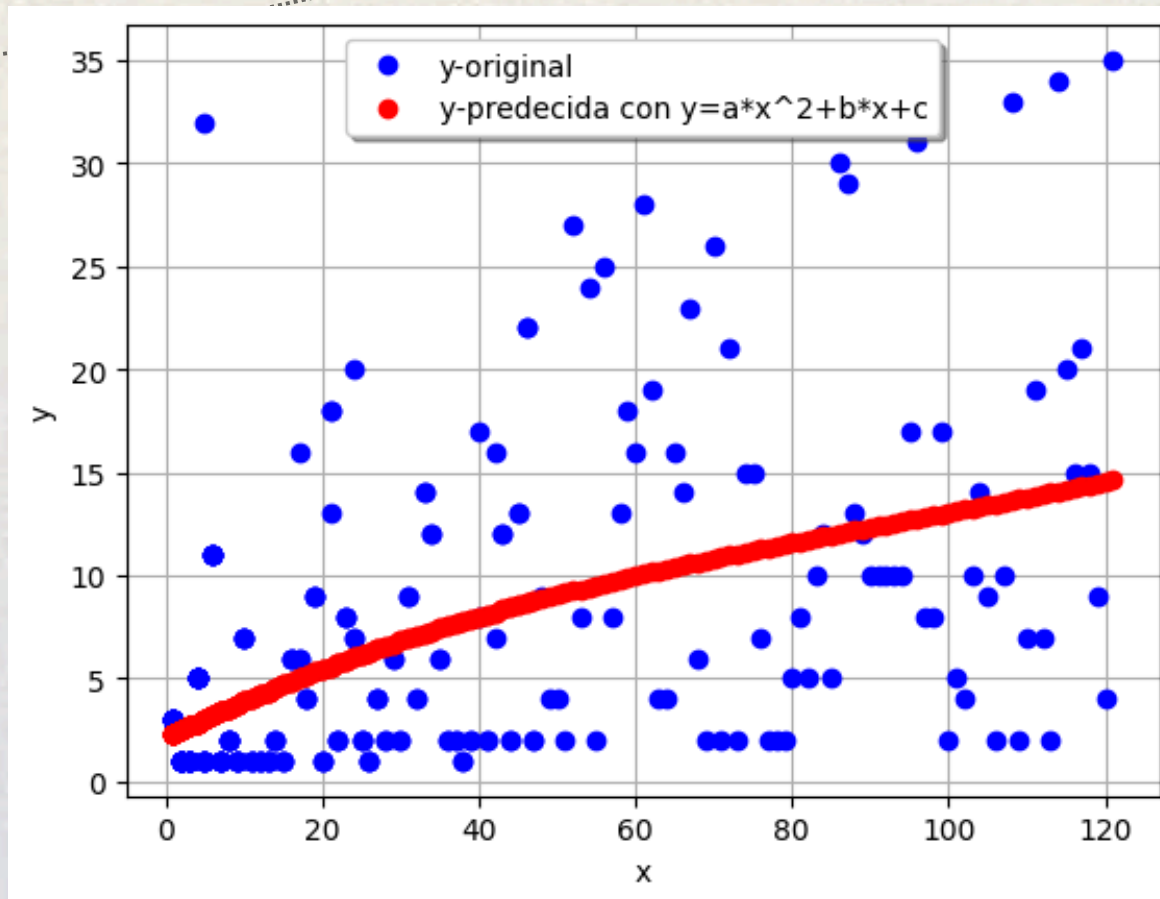
- **Resultados:**
- **Cuadrático:** $R^2 = 0.0379$, $r = 0.1949$
- **Valor A.:** $R^2 = 0.0188$, $r = 0.1374$
- **Interpretación:** La correlación de la regresión cuadrática es considerablemente mejor.
- **Visualmente:** alta dispersión sin tendencia clara ni agrupamientos definidos.

| | |
|--------------------|--------|
| Funcion R2 | 0.0379 |
| Correlación NL (r) | 0.1375 |
| Correlación L (r) | 0.1949 |

Project manager - Project organization

MODELOS NO LINEALES:

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



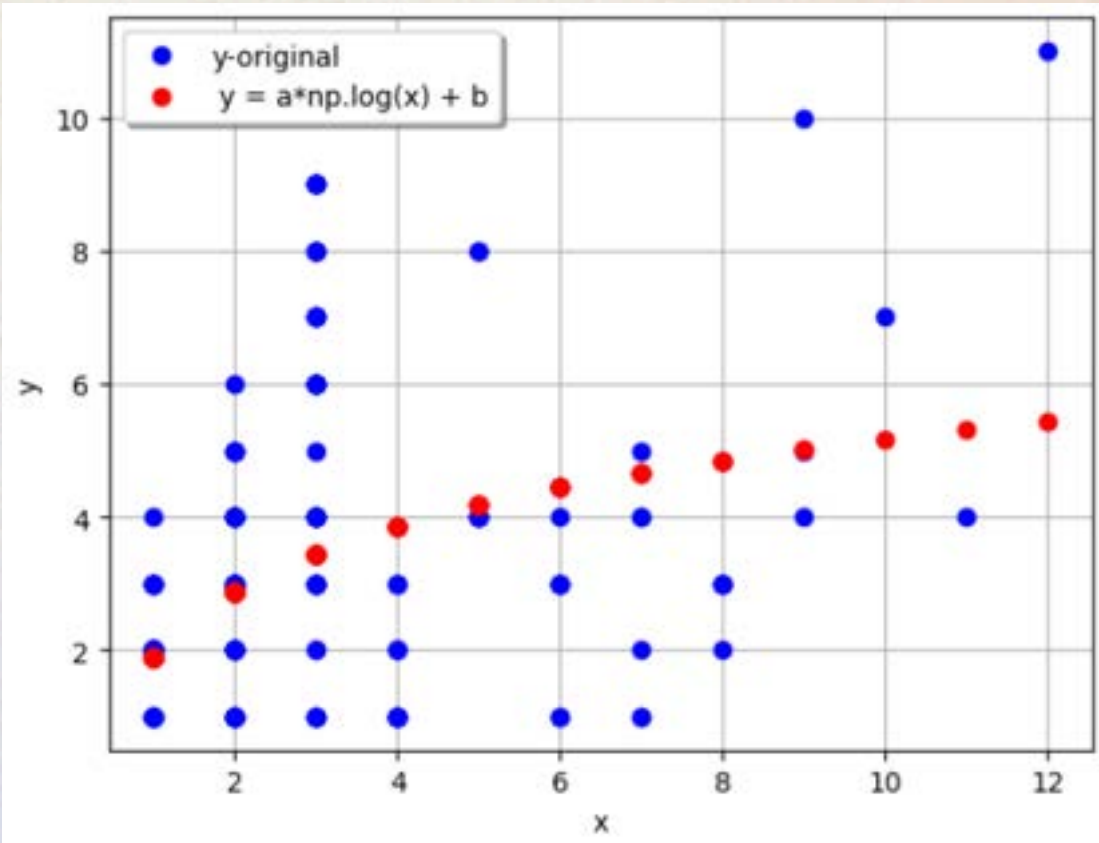
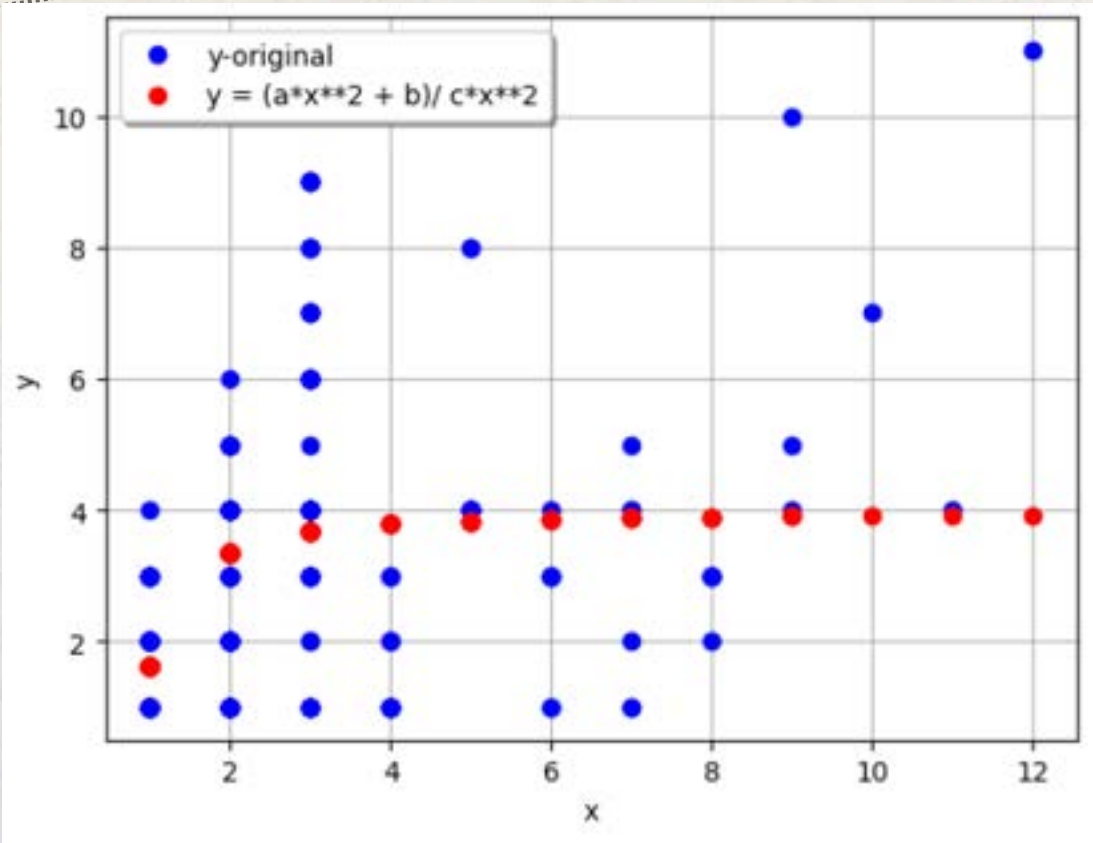
- Resultados:
- Cuadrático: $R^2 = 0.2579$, $r = 0.5078$
- Valor A.: $R^2 = 0.2522$, $r = 0.5022$
- Interpretación: Ambas regresiones ayudaron a crecer la correlación considerablemente siendo estas muy similares.
- Visualmente: alta dispersión con tendencia clara y agrupamientos definidos.

| | |
|--------------------|--------|
| Funcion R2 | 0.2579 |
| Correlación NL (r) | 0.5022 |
| Correlación L (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: $Y = A \cdot \text{NP.LOG}(X) + B$



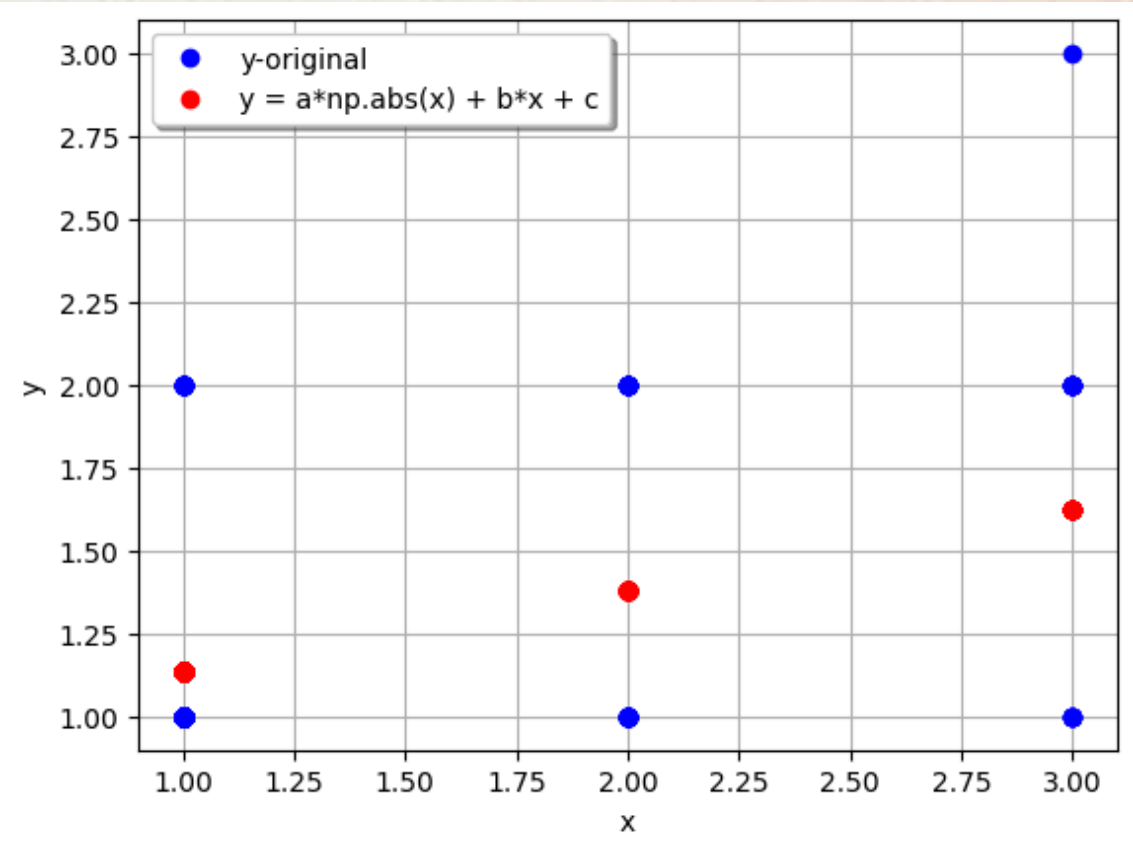
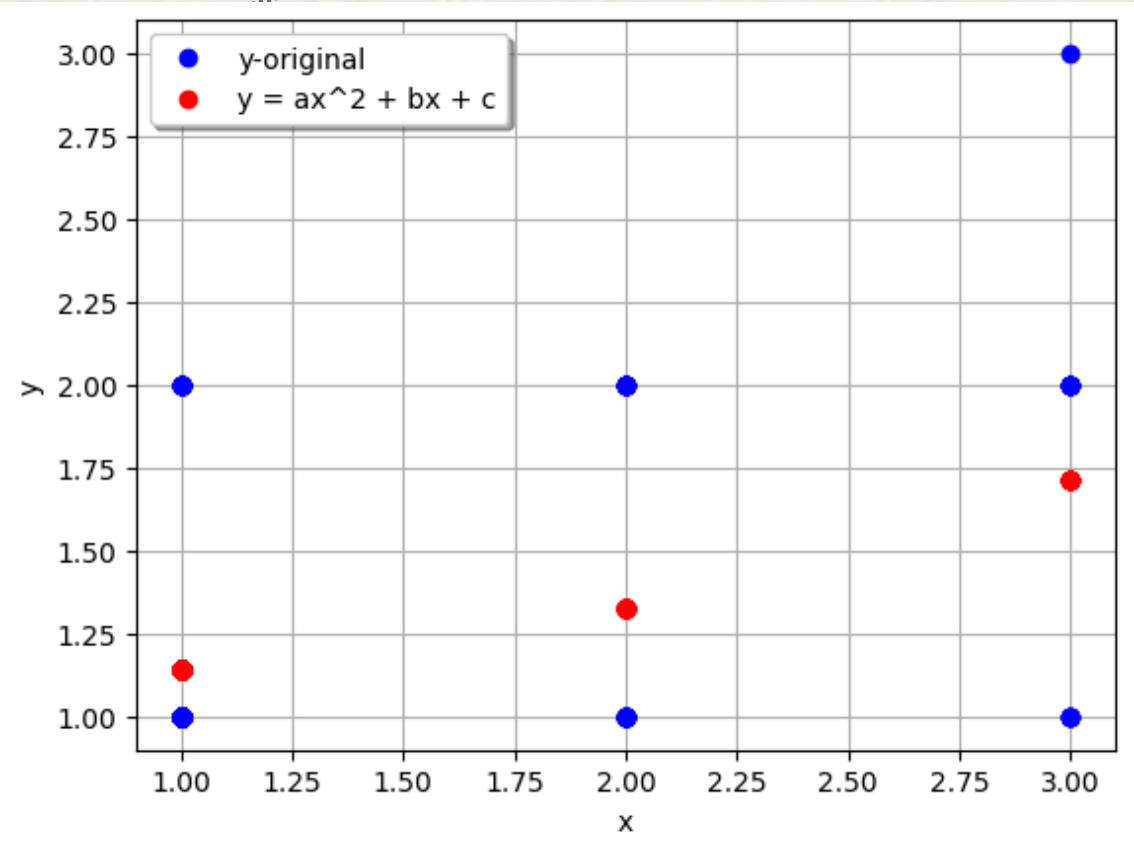
- **Resultados:**
- **Cociente:** $R^2 = 0.2128$, $r = 0.4613$
- **Logaritmica:** $R^2 = 0.2039$, $r = 0.4515$
- **Interpretación:** La correlacion mejora en ambos casos y tienen resultados similares respecto a la correlación
- **Visualmente:** alta dispersión con tendencia clara ni agrupamientos definidos.

| | |
|--------------------|--------|
| Funcion R2 | 0.2128 |
| Correlación NL (r) | 0.3869 |
| Correlación L (r) | 0.4613 |

Project Health - On-hold

MODELOS NO LINEALES:

- CUADRÁTICO: $Y=AX^2+BX+CY = AX^2 + BX + CY=AX^2+BX+C$
- FUNCION VALOR ABSOLUTO: $Y = A*NP.ABS(X) + B*X + C$



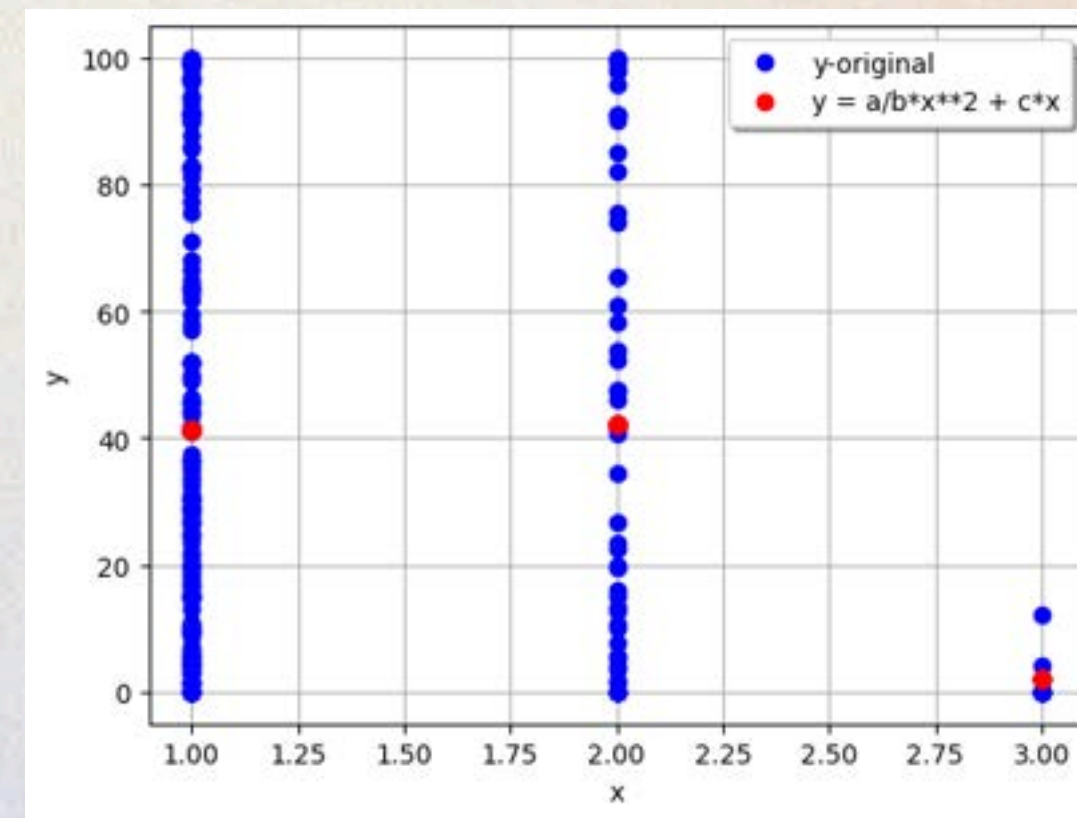
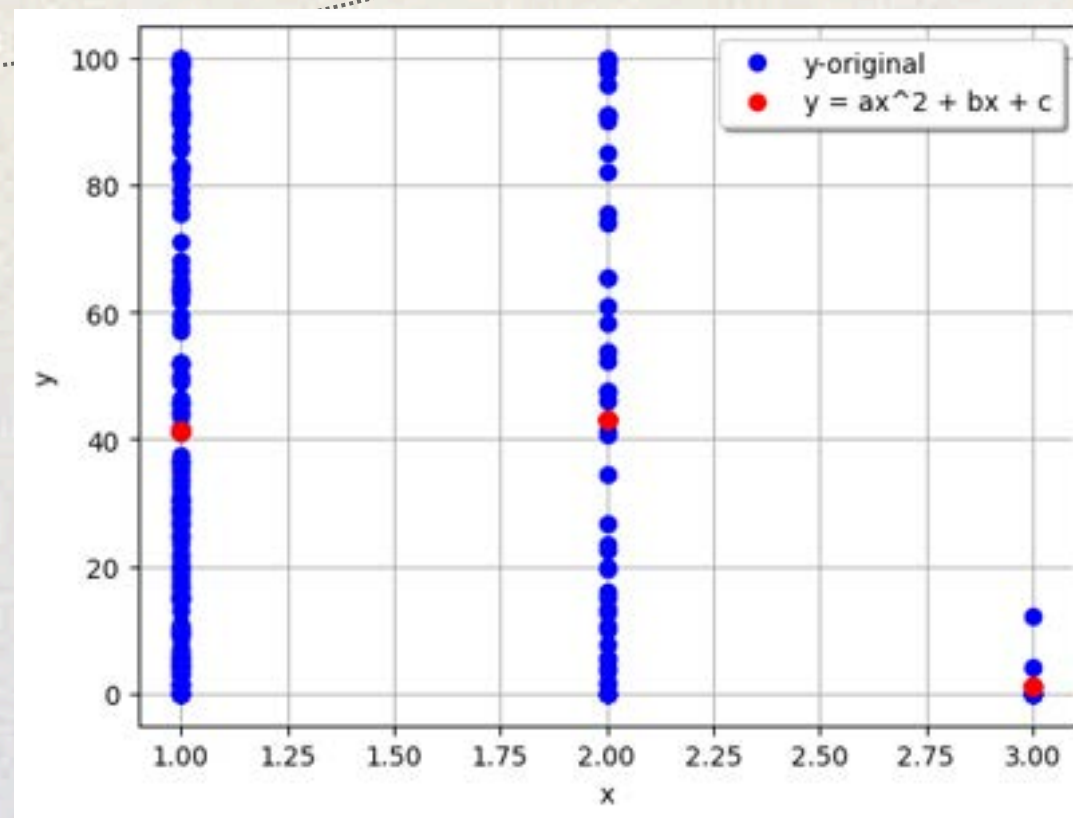
- Resultados:
- Cuadrático: $R^2 = 0.1178$, $r = 0.3433$
- Exponencial: $R^2 = 0.1118$, $r = 0.3343$
- Interpretación: Parecida R^2 y correlación mejoran por lo mínimo.
- Visualmente: alta dispersión con tendencias claras pero no hay agrupamientos definidos.

| | |
|--------------------|--------|
| Funcion R2 | 0.1178 |
| Correlación NL (r) | 0.3344 |
| Correlación L (r) | 0.3433 |

Project Health - Percent complete

MODELOS NO LINEALES:

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



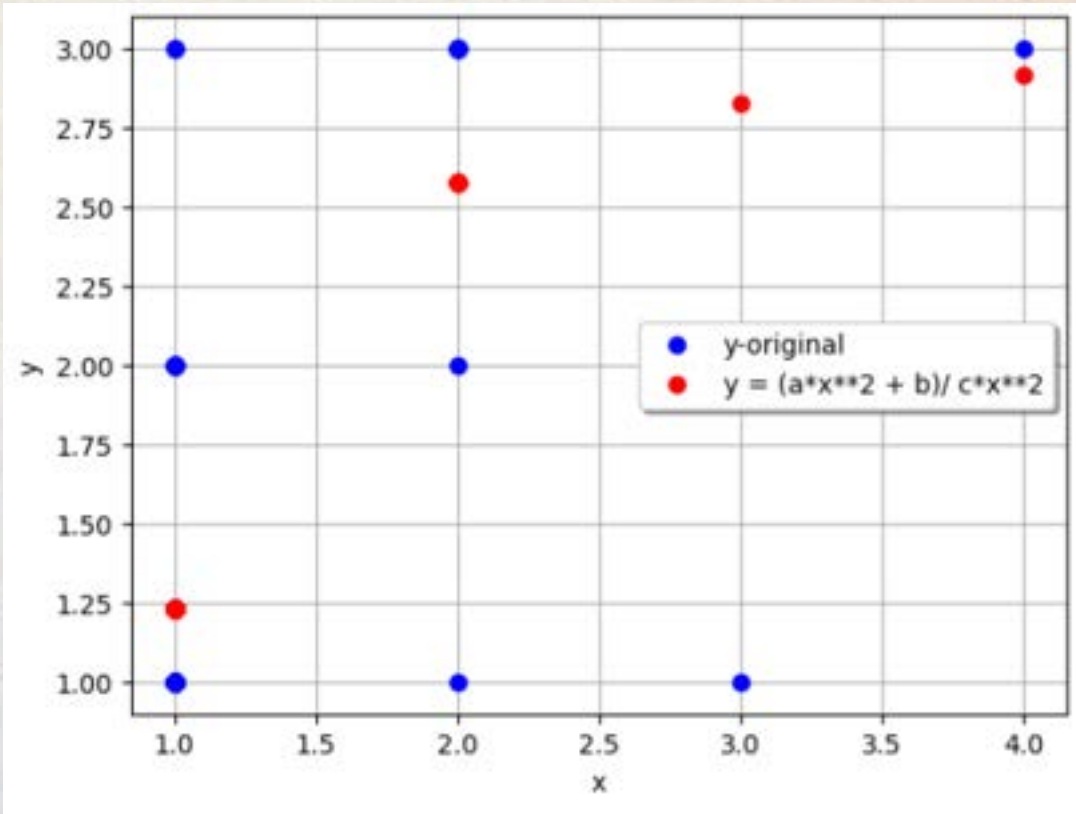
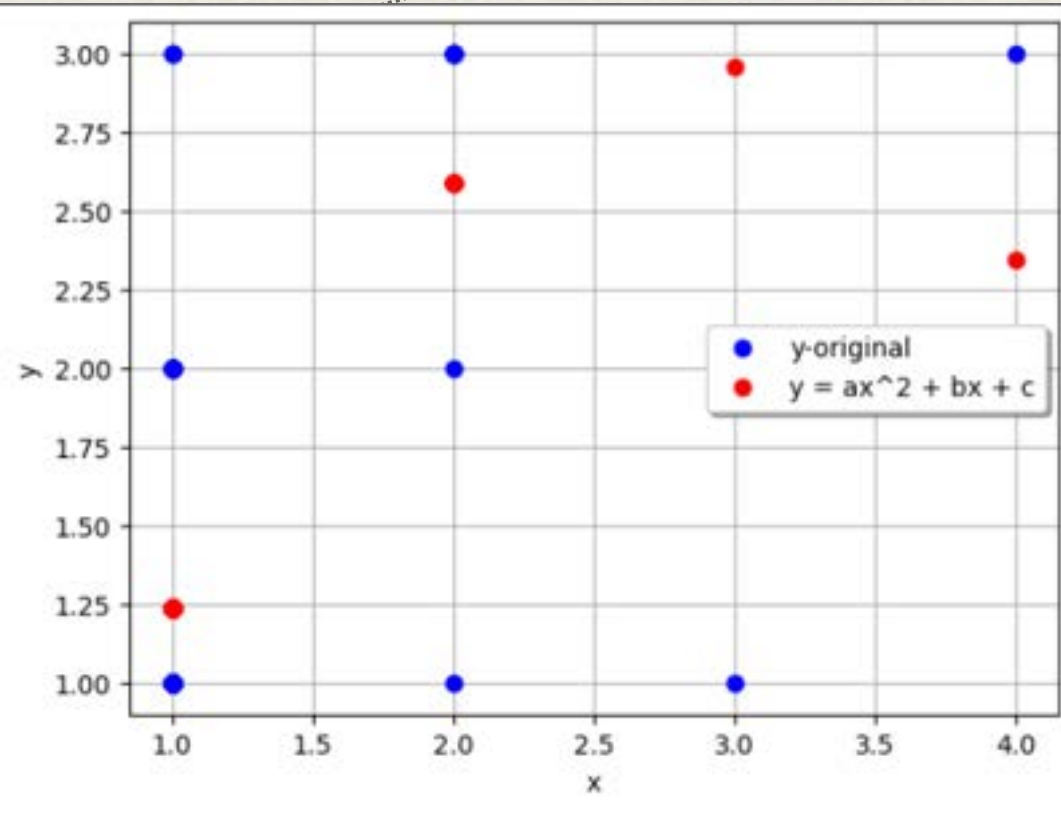
- **Resultados:**
- **Cuadrático:** $R^2 = 0.0709$, $r = 0.2663$
- **Exponencial:** $R^2 = 0.0707$, $r = 0.2660$
- **Interpretación:** Similares correlaciones y R^2 , ambas subieron considerablemente.
- **Visualmente:** alta dispersión con tendencias nada claras y no hay agrupamientos definidos.

| | |
|--------------------|---------|
| Funcion R2 | 0.0709 |
| Correlación NL (r) | -0.1784 |
| Correlación L (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 |
| Correlación NL (r) | 0.4938 |
| Correlación L (r) | 0.5616 |

ANÁLISIS DE INSIGHTS

Patrones de Registro

- DISTRIBUCIÓN CONTROLADA EN MÉTRICAS
- SISTEMAS DE REGISTRO MADUROS
- PROCESOS Estandarizados DE DOCUMENTACIÓN Y SEGUIMIENTO

Estandarización de Procesos

- PROCESOS ALTAMENTE ESTANDARIZADOS A NIVEL ORGANIZACIONAL
- MADUREZ OPERATIVA DEMOSTRADA POR DISTRIBUCIONES ESTRUCTURADAS
- LÍMITES OPERATIVOS BIEN DEFINIDOS Y RESPETADOS

Campos No Independientes

- COHERENCIA ESTRATÉGICA EN DECISIONES ORGANIZACIONALES
- PLANIFICACIÓN SOFISTICADA EVIDENCIADA POR PATRONES INTERCONECTADOS

Correlación: Project Manager ↔ Project Organization (0.505)

- ASIGNACIÓN ESTRATÉGICA DE GERENTES SEGÚN ESTRUCTURA ORGANIZACIONAL
- OPTIMIZACIÓN DE RECURSOS Y GESTIÓN DEL TALENTO SOFISTICADA

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