

Octubre 2025

Análisis exploratorio de proyectos Datathon y FORVIA

ANOVA Y MANOVA

FORVIANOS.py

FORVIANOS . PY



Maria Matanzo

A 01737554



Jorge Cortes

A 01736236



Marco Cornejo

A 01276411



Eduardo Torres

A 01734935



Laisha Puan

A 01736397

OBJETIVOS

- Aplicación del ANOVA (Análisis de Varianza univariado) y del ANOVA FACTORIAL y se analizó si se rechaza o no la hipótesis nula
- Comparar los resultados del ANOVA y ANOVA FACTORIAL.
- Visualizar los patrones de dispersión de cada variable objetivo mediante gráficos (boxplot, violín y líneas)

METODOLOGÍA DE LAS ACTIVIDADES

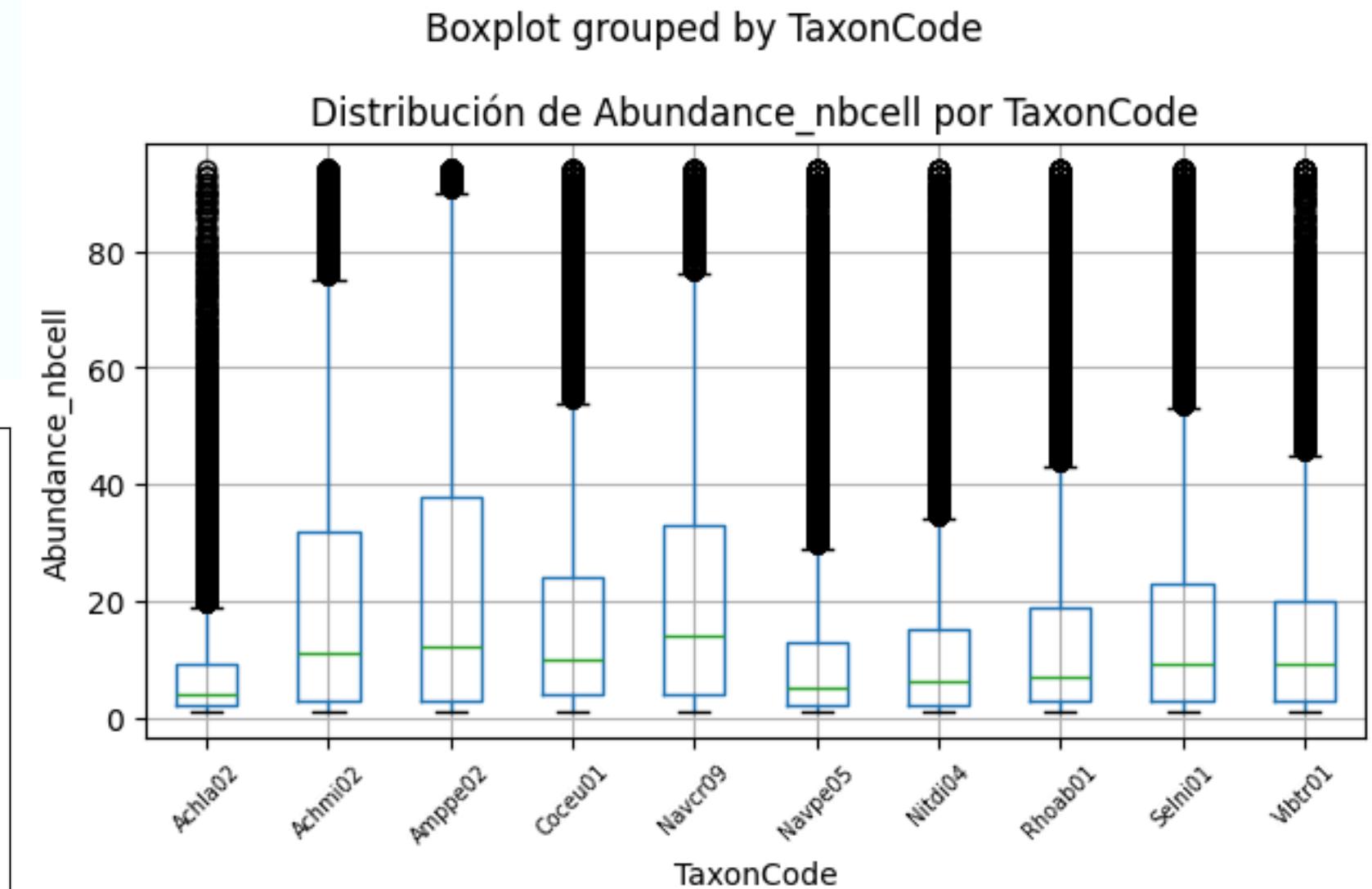
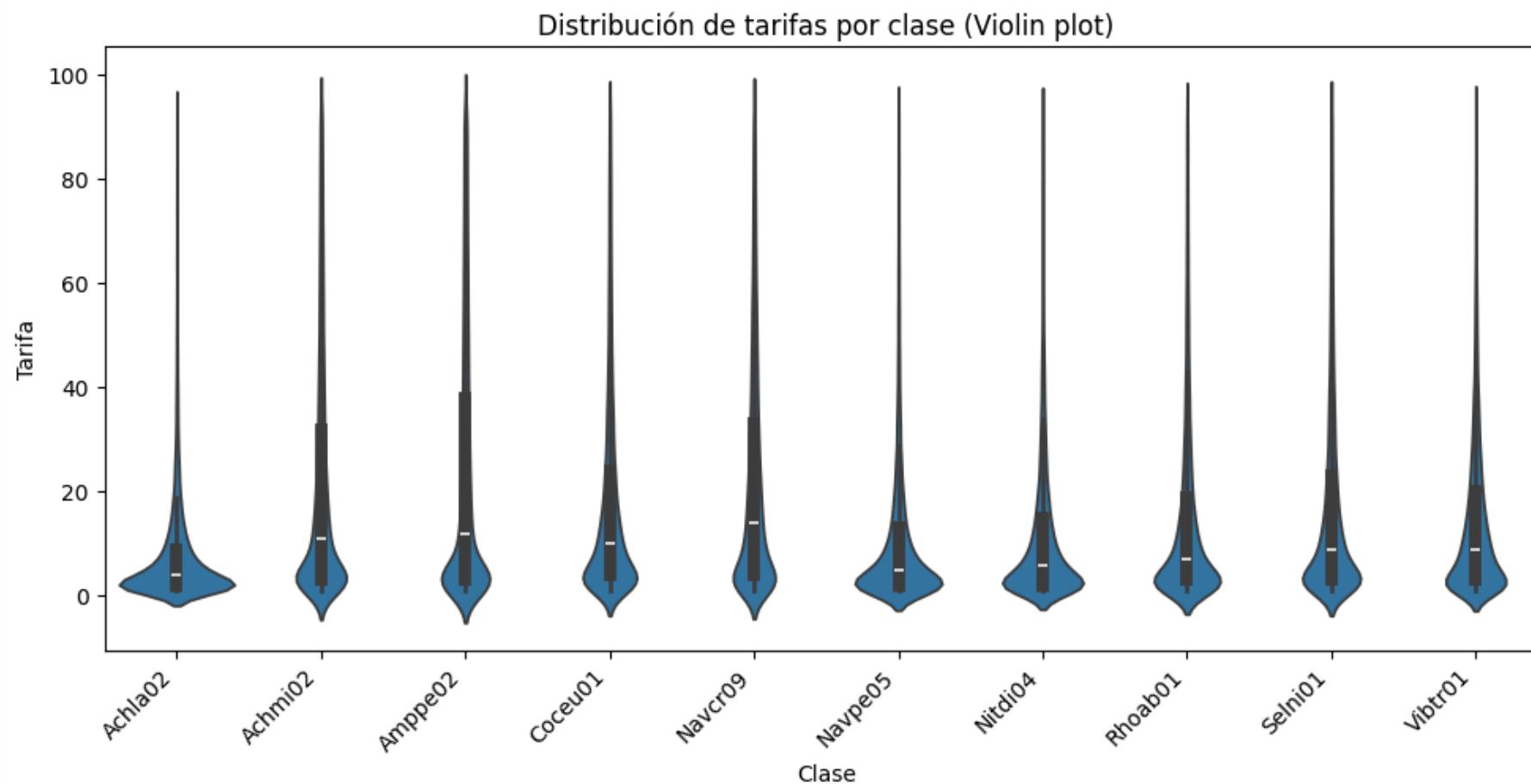
1. Se identificaron las variables objetivo y las categoricas.
2. Se elaboraron gráficos para observar la dispersión de las variables categoricas con la variable objeto.
3. Aplicación del ANOVA (Análisis de Varianza univariado) y del ANOVA FACTORIAL y se analizó si se rechaza o no la hipótesis nula
4. Se elaboro una grafica Lineplots para observar tendencias o variaciones entre las variables.

ACTIVIDAD 5.1

fORVIANOS.py

ANOVA ABUNDANCE_NBCELL

- **Variable objeto:** Abundance_nbcell
- **Variable Categorica:** TaxonCode



ANOVA ABUNDANCE_NBCELL

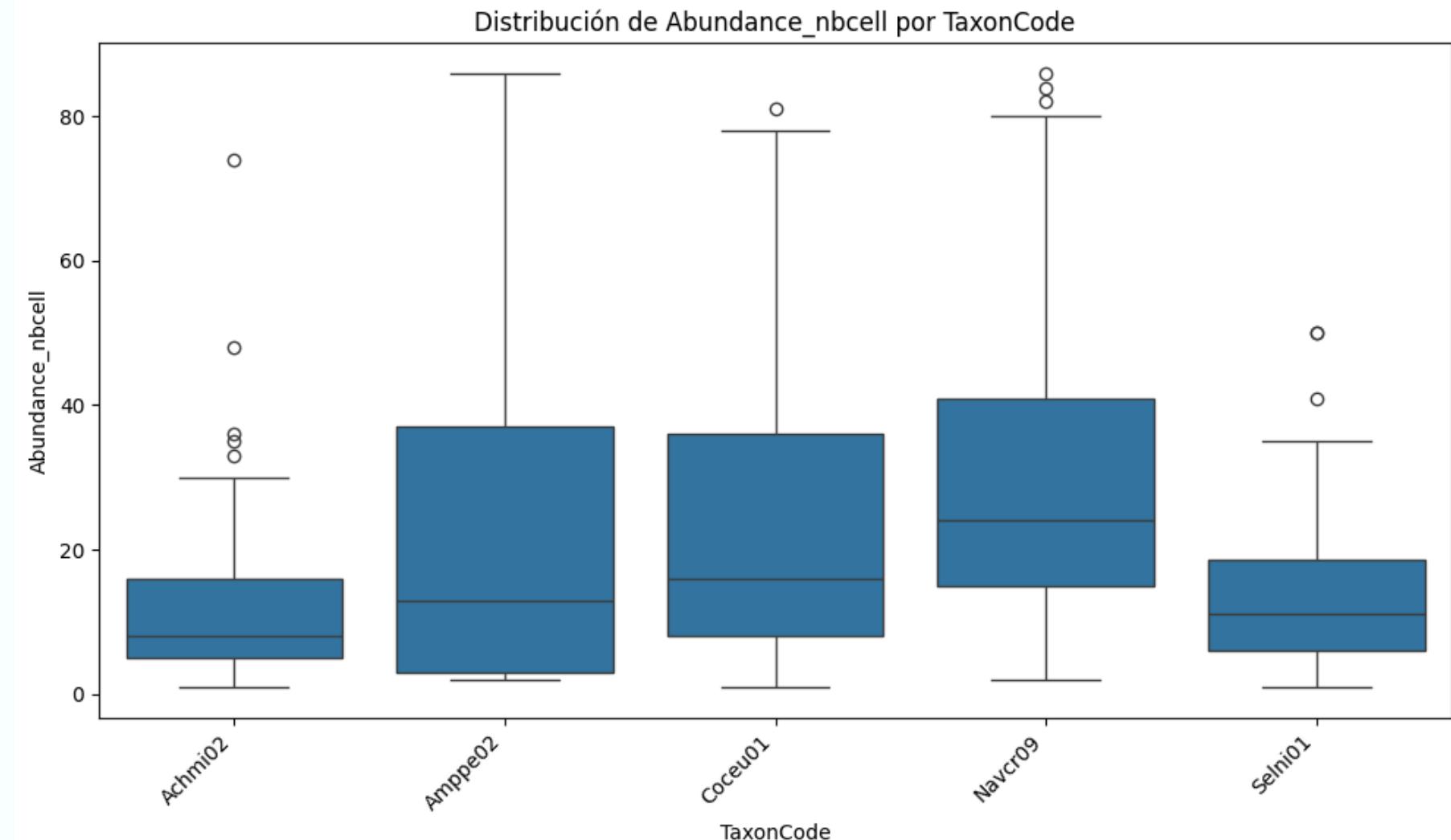
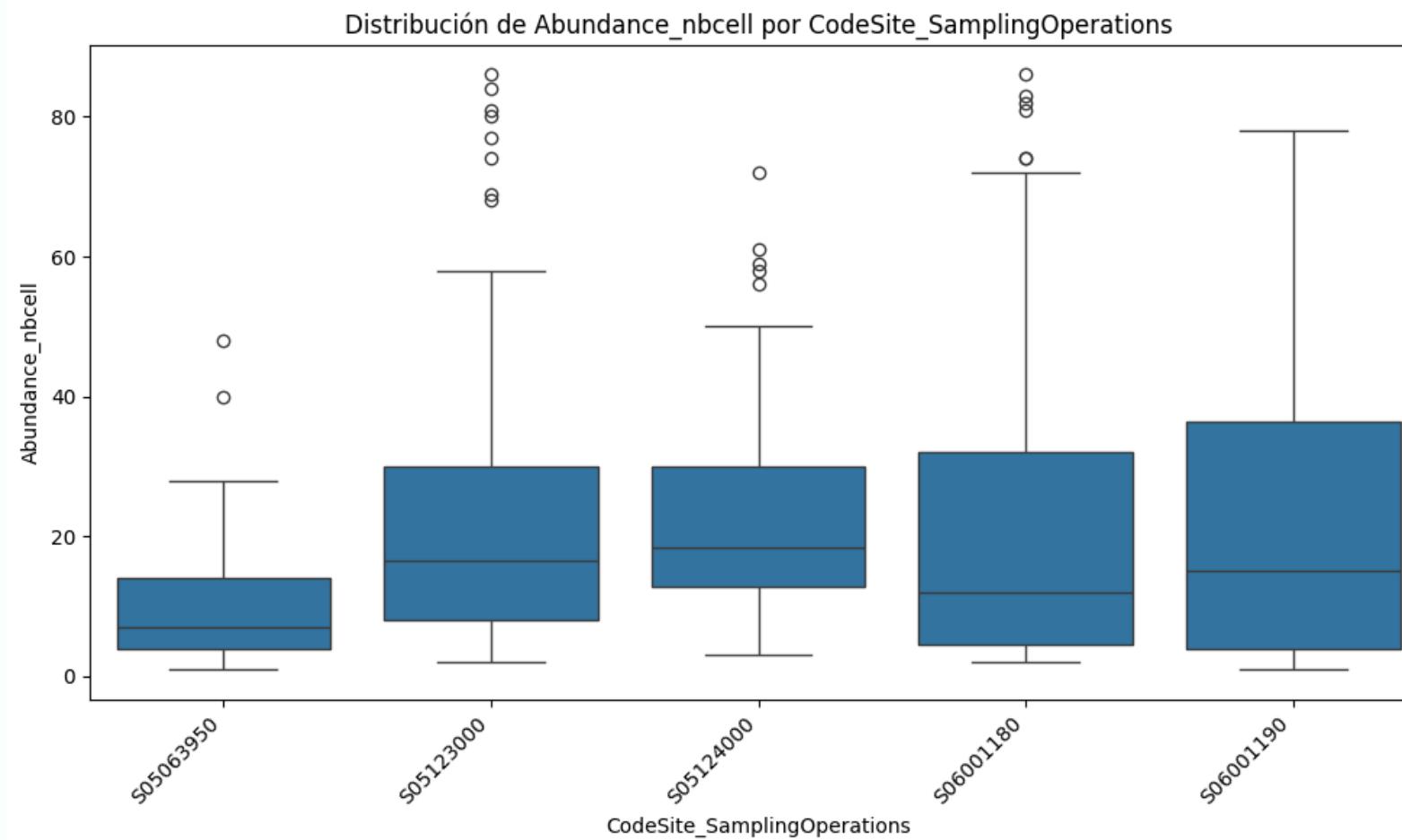
Variable objeto: Abundance_nbcell
Variables Categorica: TaxonCode

	estadístico de prueba F	p-value
TaxonCode	2470.22	0.0

Se rechaza la hipótesis nula.

ANOVA FACTORIAL ABUNDANCE_NBCELL

- **Variable objeto:** Abundance_nbcell
- **Variable Categorica:** TaxonCode,
CodeSite_SamplingOperations



ANOVA FACTORIAL ABUNDANCE_NBCELL

Variable objeto: Abundance_nbcell

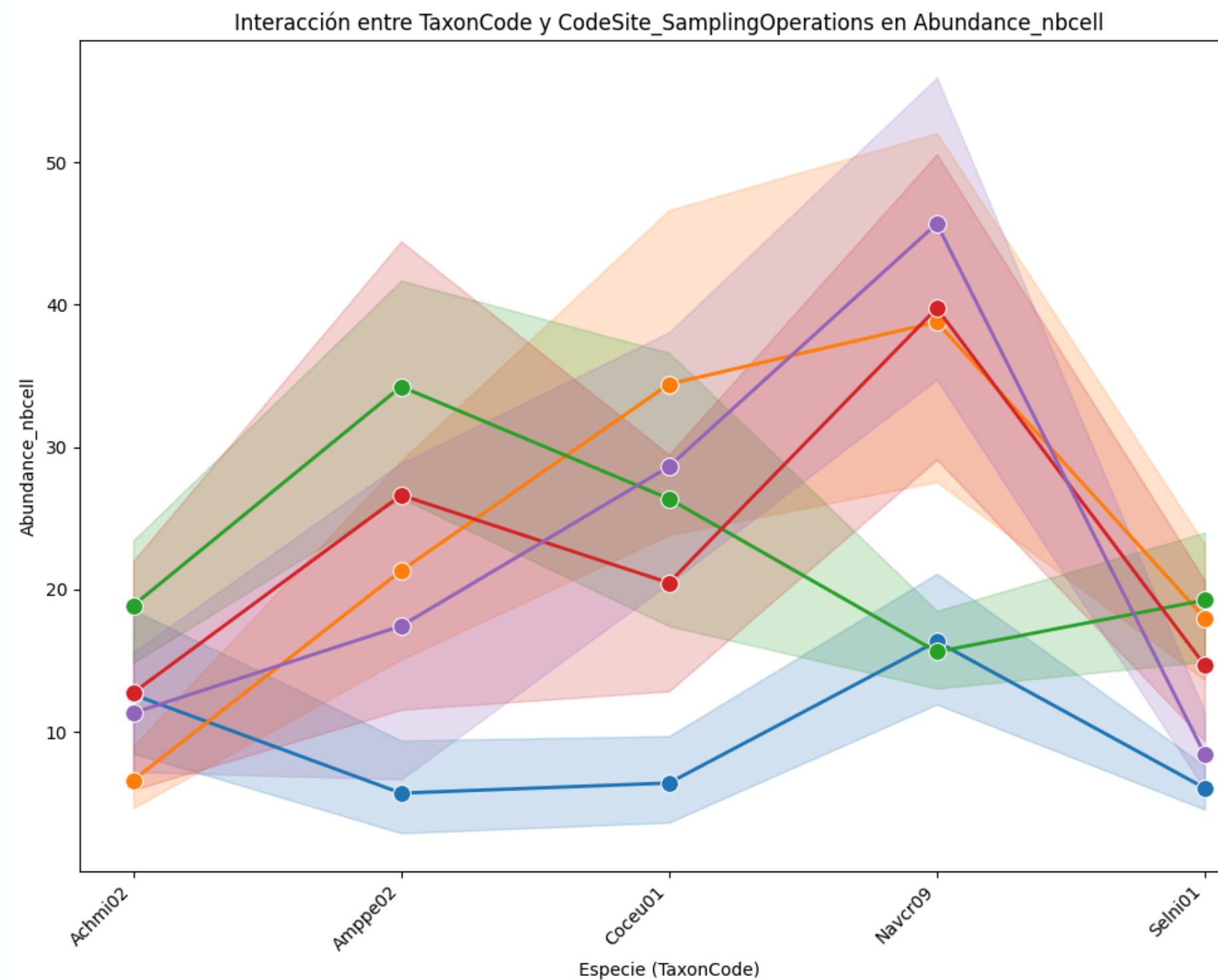
Variables Categorica: TaxonCode,
CodeSite_SamplingOperations

	estadístico de prueba F	p-value
TaxonCode	17.46	3.62e-13
CodeSite_SamplingOperations	9.62	2.021e-07
TaxonCode: CodeSite_SamplingOperations	4.17	1.82e-07

Se rechaza la hipótesis nula

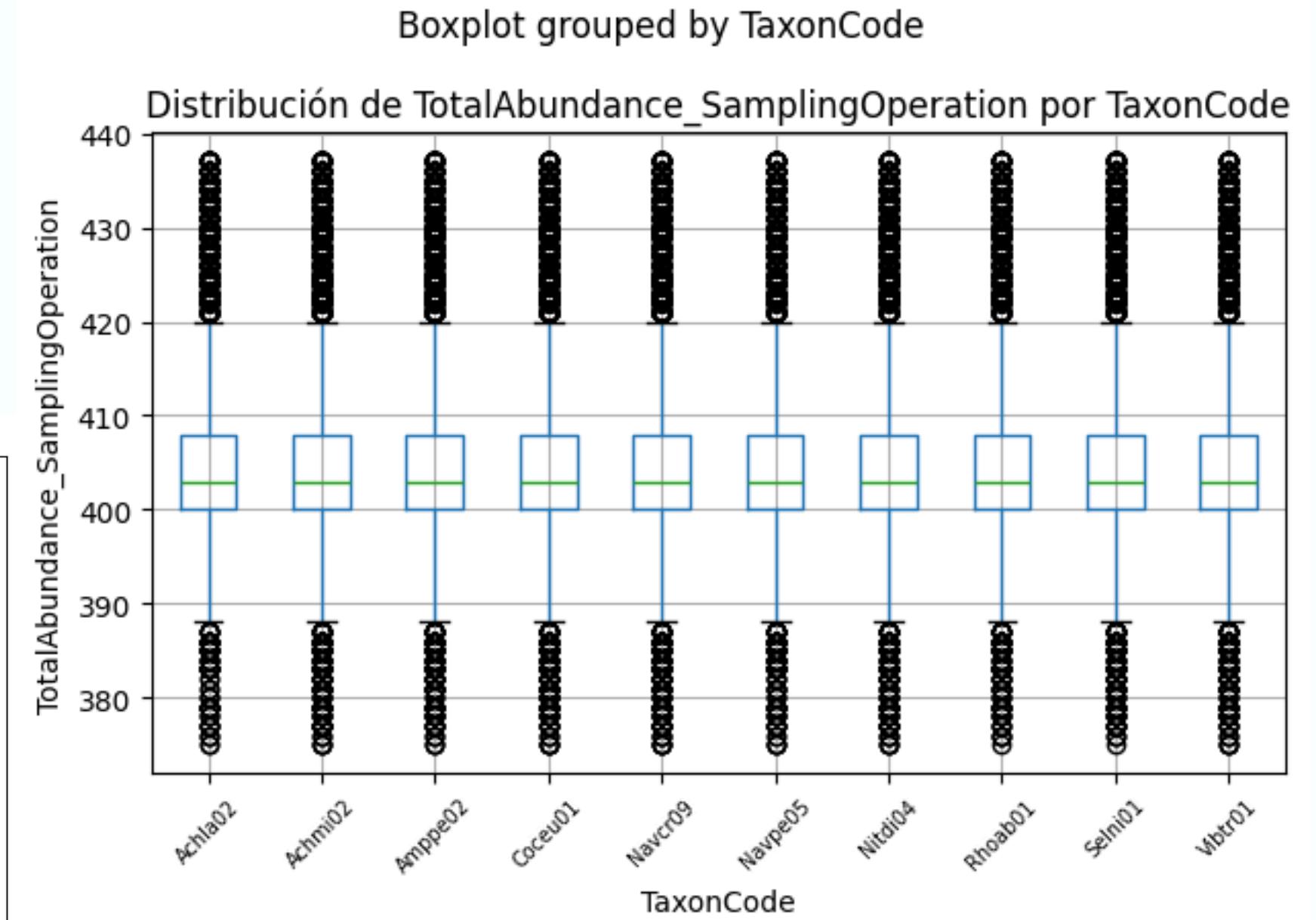
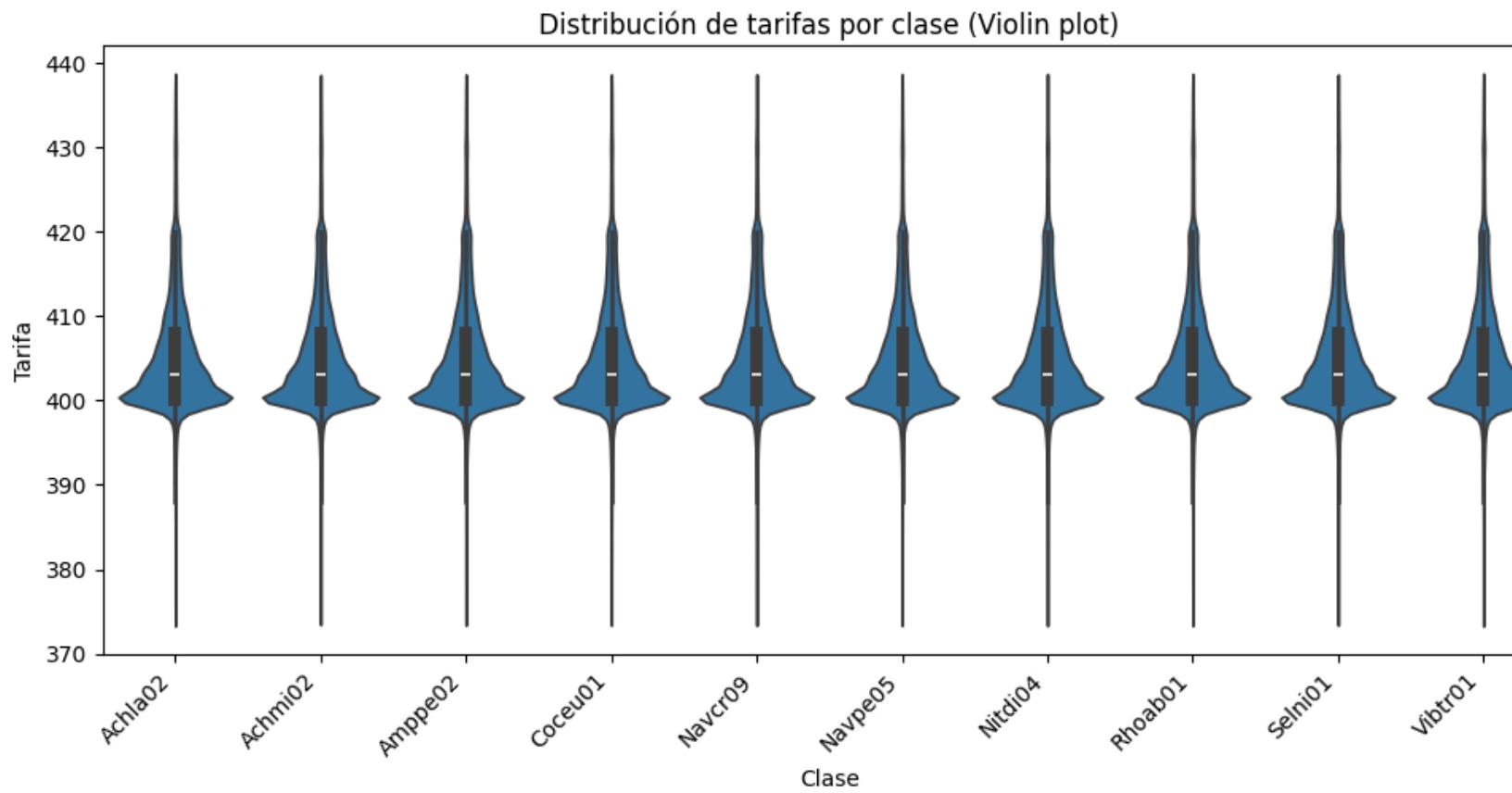
ANOVA FACTORIAL ABUNDANCE_NBCELL

- **Variable objeto:** Abundance_nbcell
- **Variable Categorica:** TaxonCode, CodeSite_SamplingOperations



ANOVA TOTAL ABUNDANCE_SAMPLING OPERATION

- **Variable objeto:**
TotalAbundance_SamplingOperation
- **Variable Categorica:** TaxonCode



ANOVA TOTAL ABUNDANCE_SAMPLING OPERATION

Variable objeto:

TotalAbundance_SamplingOperation

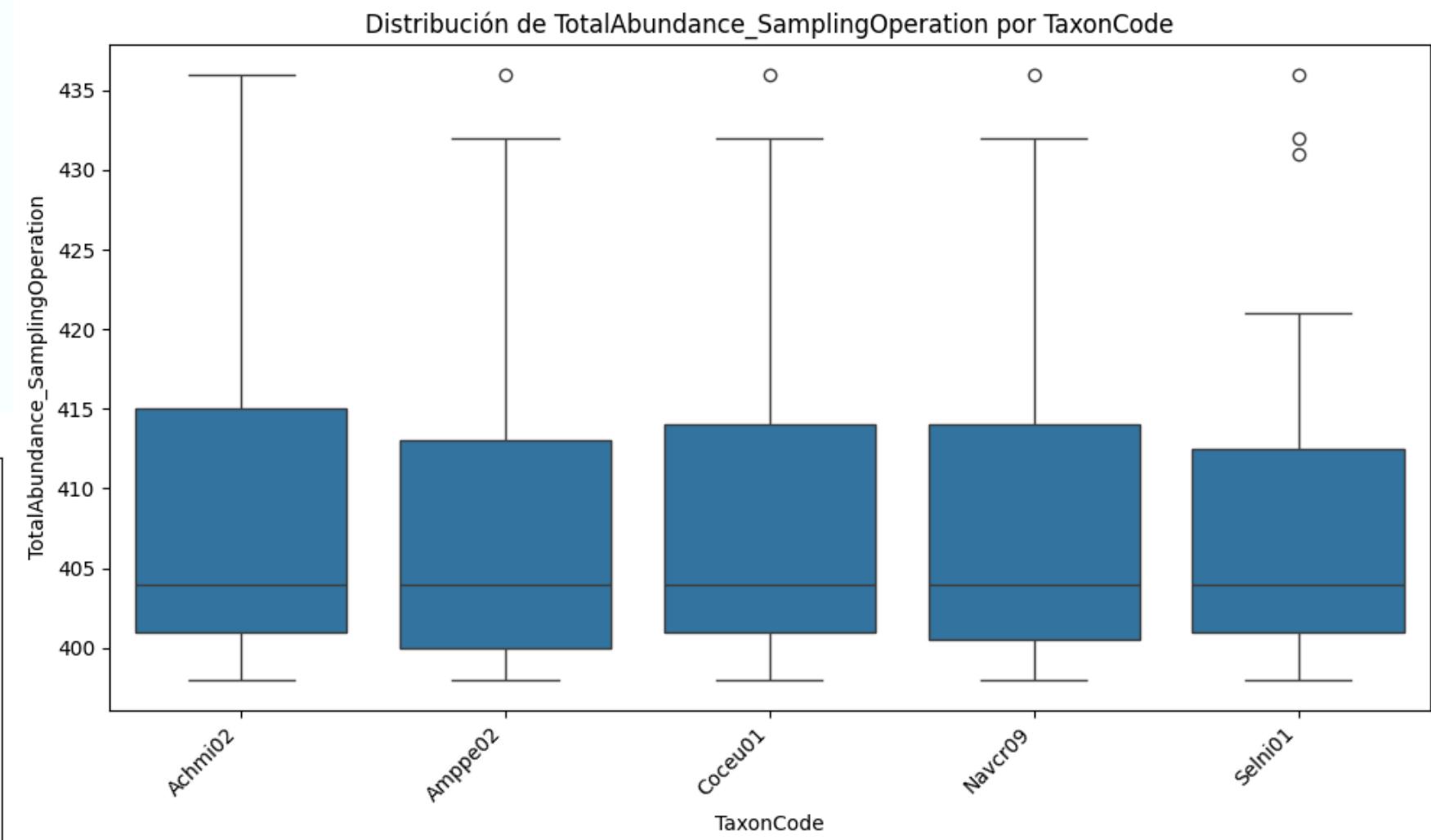
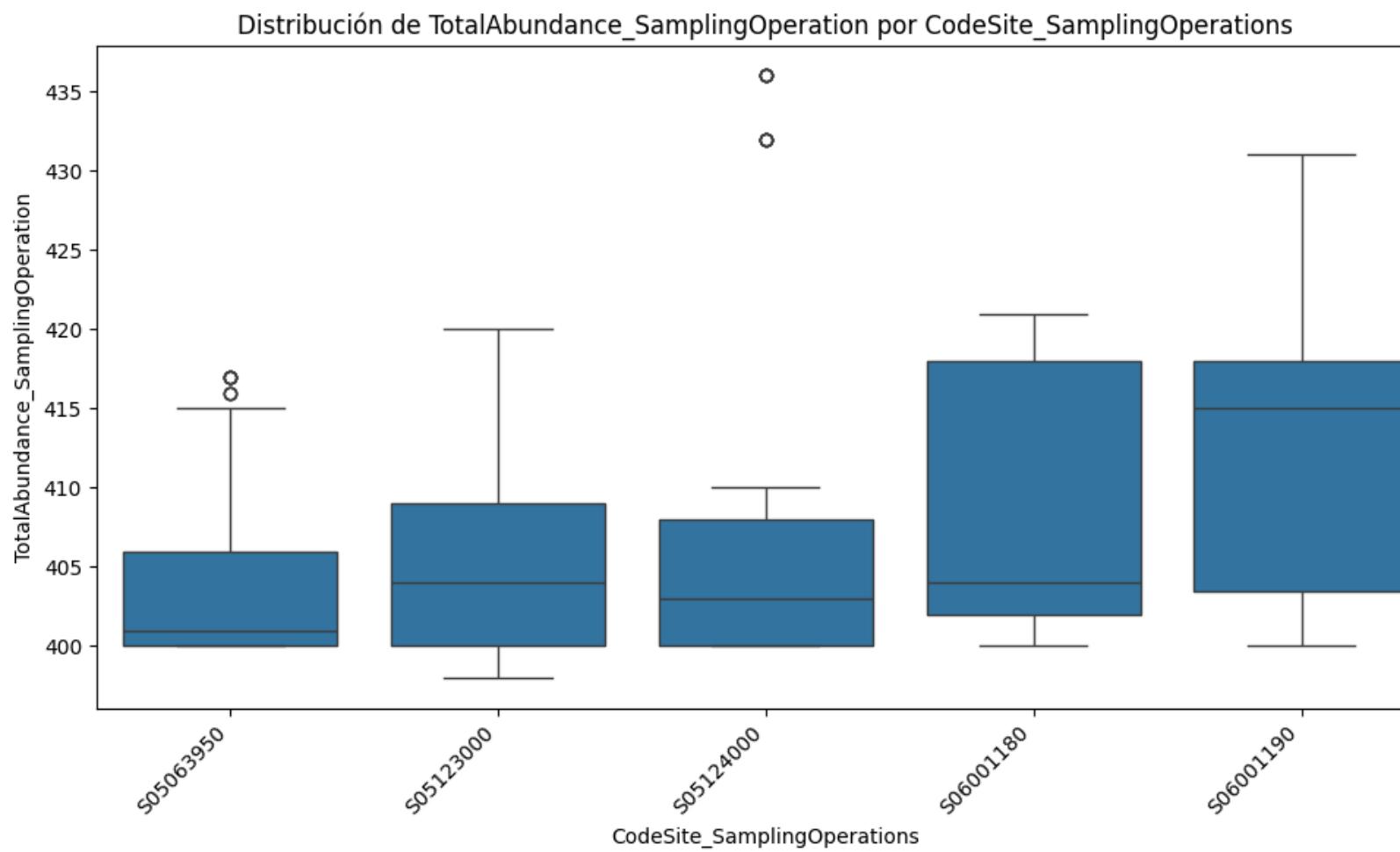
Variables Categorica: TaxonCode

	estadístico de prueba F	p-value
TaxonCode	2.29	0.01

Se rechaza la hipótesis nula

ANOVA FACTORIAL TOTAL ABUNDANCE_SAMPLING OPERATION

- **Variable objeto:**
CodeSite_SamplingOperations
- **Variable Categorica:** TaxonCode,
CodeSite_SamplingOperations



ANOVA FACTORIAL TOTAL ABUNDANCE_SAMPLING OPERATION

Variable objeto: Abundance_nbcell

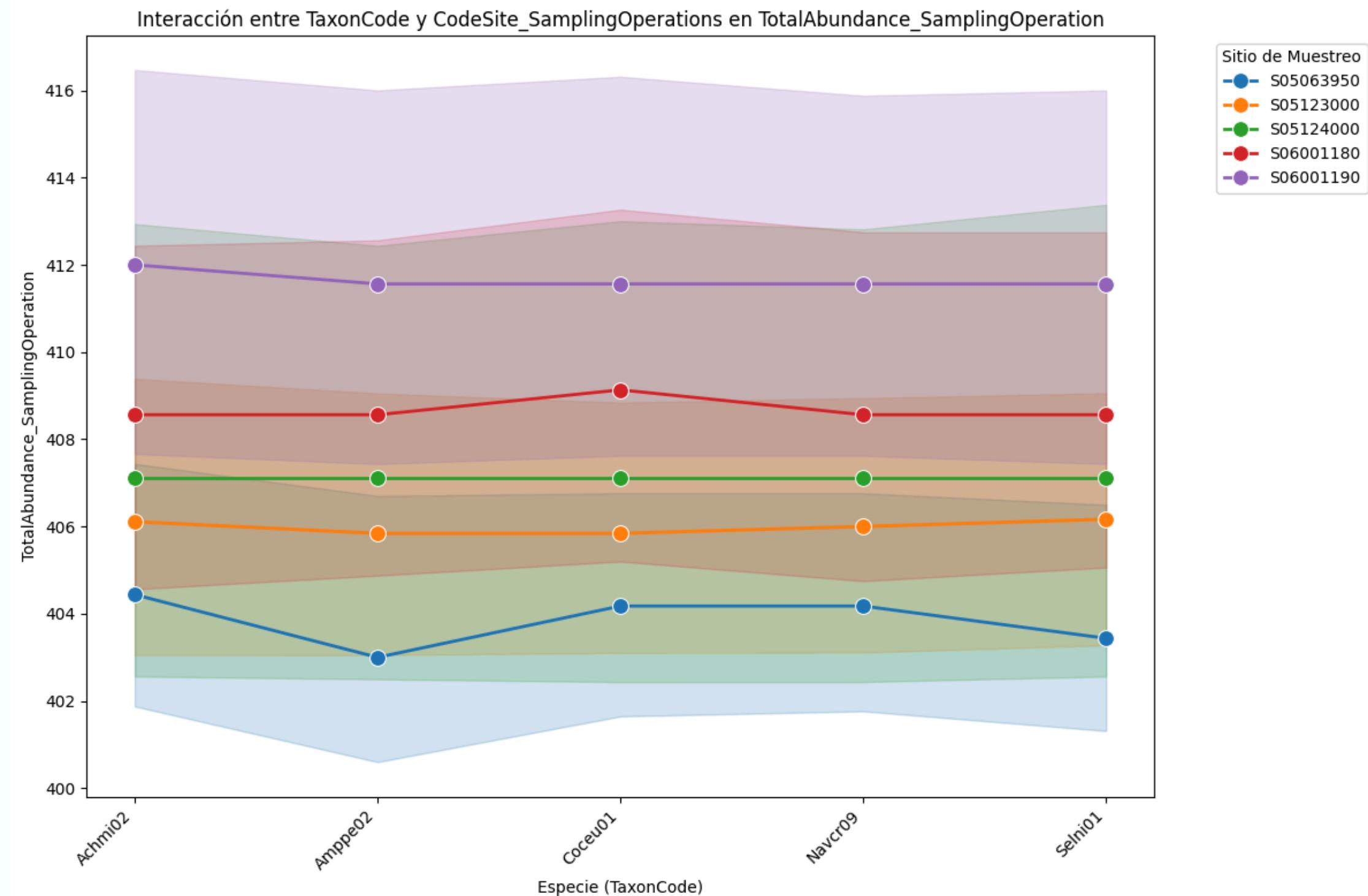
Variables Categorica: TaxonCode,
CodeSite_SamplingOperations

	estadístico de prueba F	p-value
TaxonCode	0.00	9.99e-01
CodeSite_SamplingOperations	9.58	2.15e-07
TaxonCode: CodeSite_SamplingOperations	0.01	1

Solo se rechaza la hipótesis con la variable
codeSite_SamplingOperations

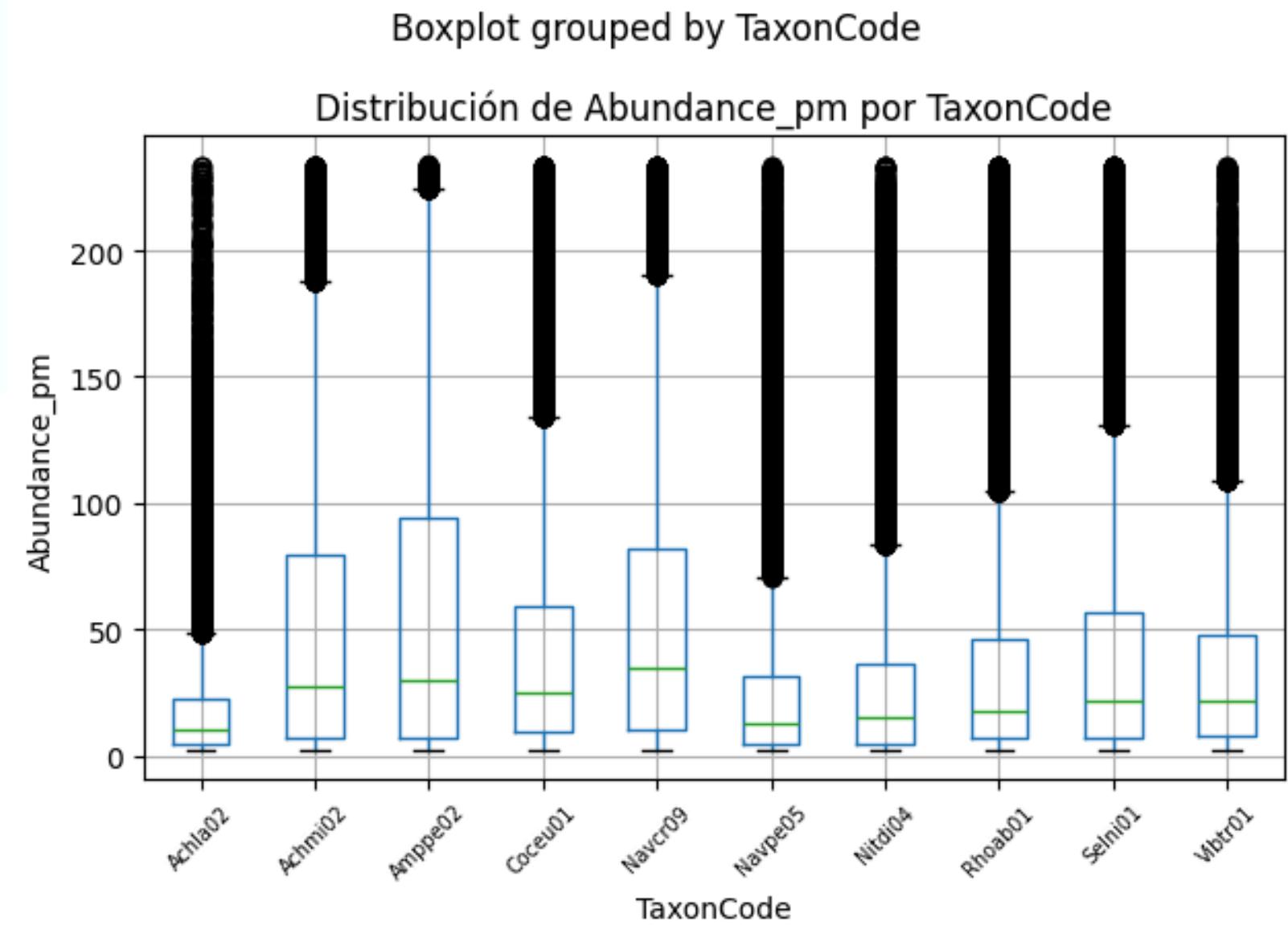
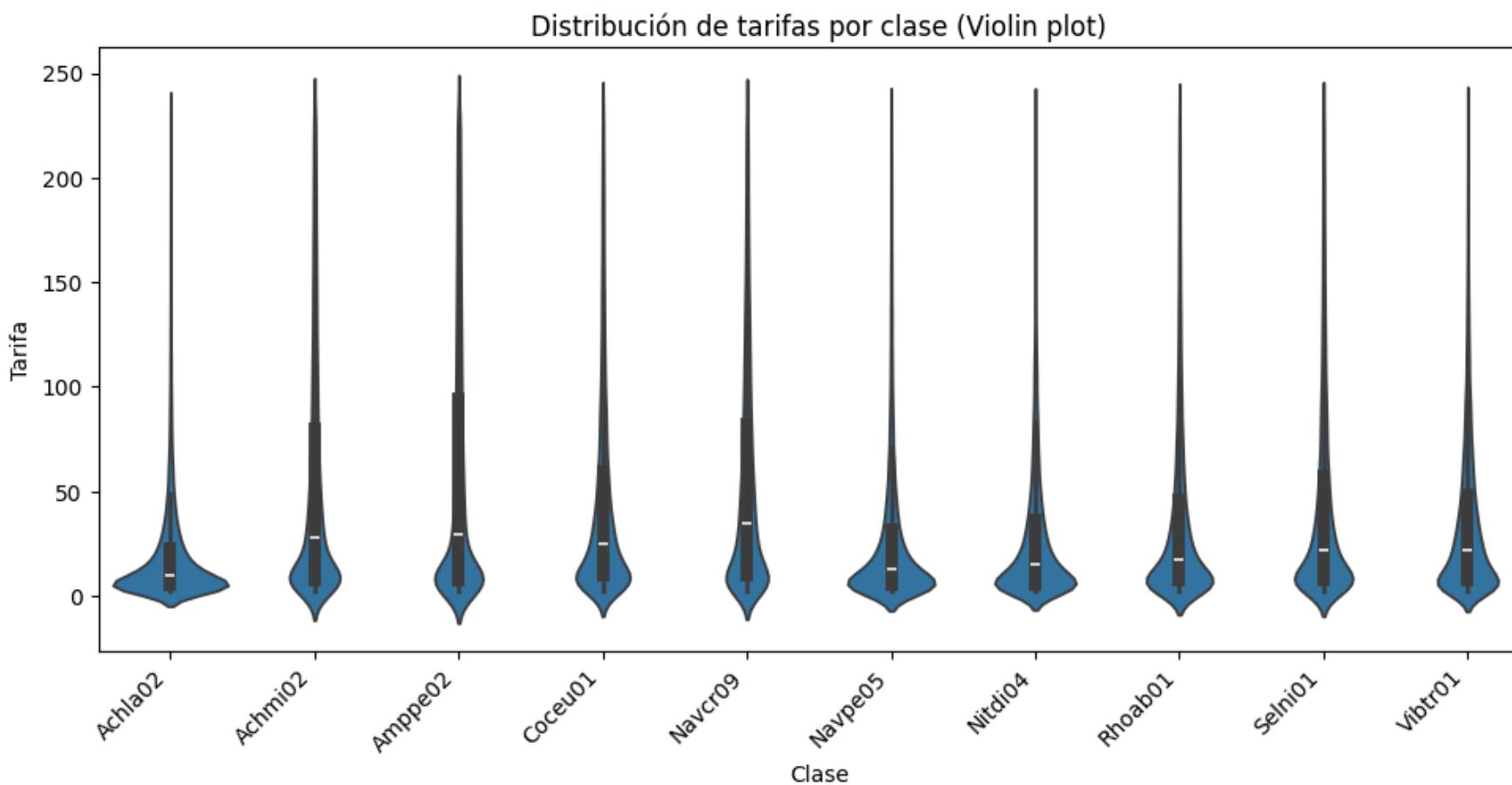
ANOVA FACTORIAL TOTAL ABUNDANCE_SAMPLING OPERATION

- **Variable objeto:**
TotalAbundance_SamplingOperation
- **Variable Categorica:** TaxonCode,
CodeSite_SamplingOperations



ANOVA ABUNDANCE_PM

- **Variable objeto:** Abundance_PM
- **Variable Categorica:** TaxonCode



ANOVA ABUNDANCE_PM

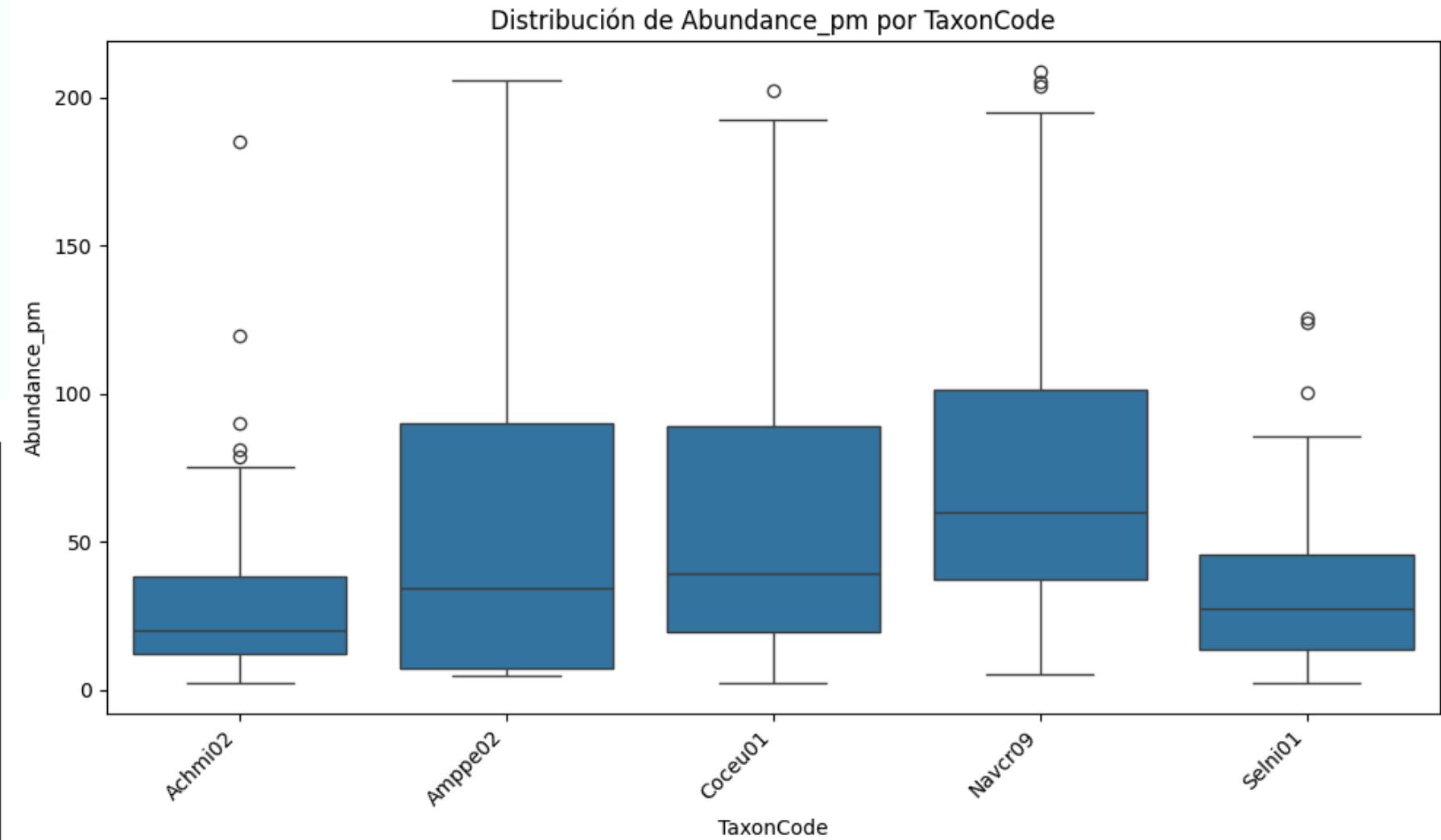
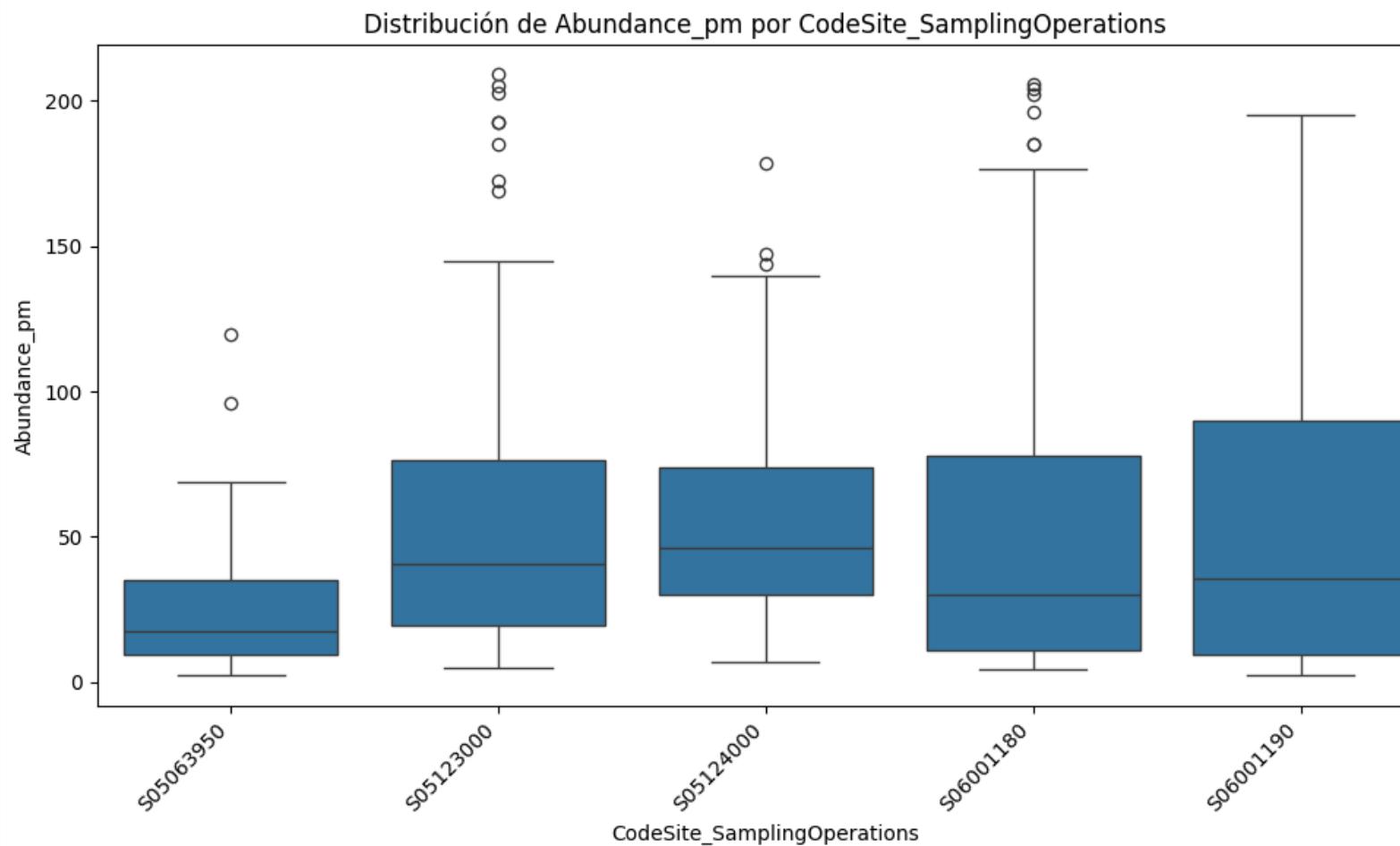
Variable objeto: Abundance_pm
Variables Categorica: TaxonCode

	estadístico de prueba F	p-value
TaxonCode	2484.64	0.0

Se rechaza la hipótesis nula

ANOVA FACTORIAL ABUNDANCE_PM

- **Variable objeto:** Abundance_pm
- **Variable Categorica:** TaxonCode,
CodeSite_SamplingOperations



ANOVA FACTORIAL ABUNDANCE_PM

Variable objeto: Abundance_pm

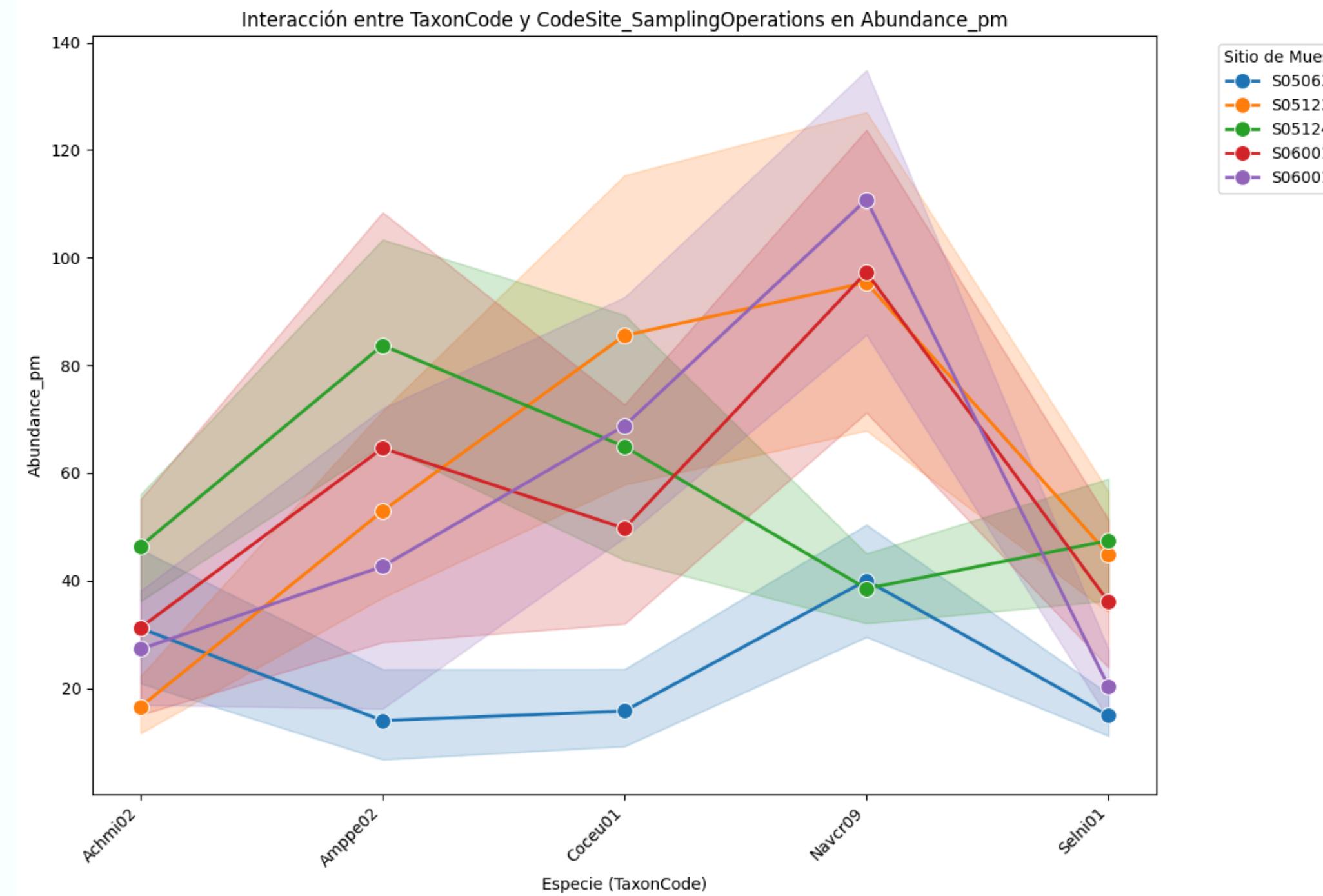
Variables Categorica: TaxonCode,
CodeSite_SamplingOperations

	estadístico de prueba F	p-value
TaxonCode	17.32	4.54e-13
CodeSite_SamplingOperations	9.70	1.75e-07
TaxonCode: CodeSite_SamplingOperations	4.16	1.97e-07

Se rechaza la hipótesis nula

ANOVA FACTORIAL ABUNDANCE_NBCELL

- **Variable objeto:** Abundance_nbcell
- **Variable Categorica:** TaxonCode, CodeSite_SamplingOperations



CONCLUSIONES DATASET DATATHON

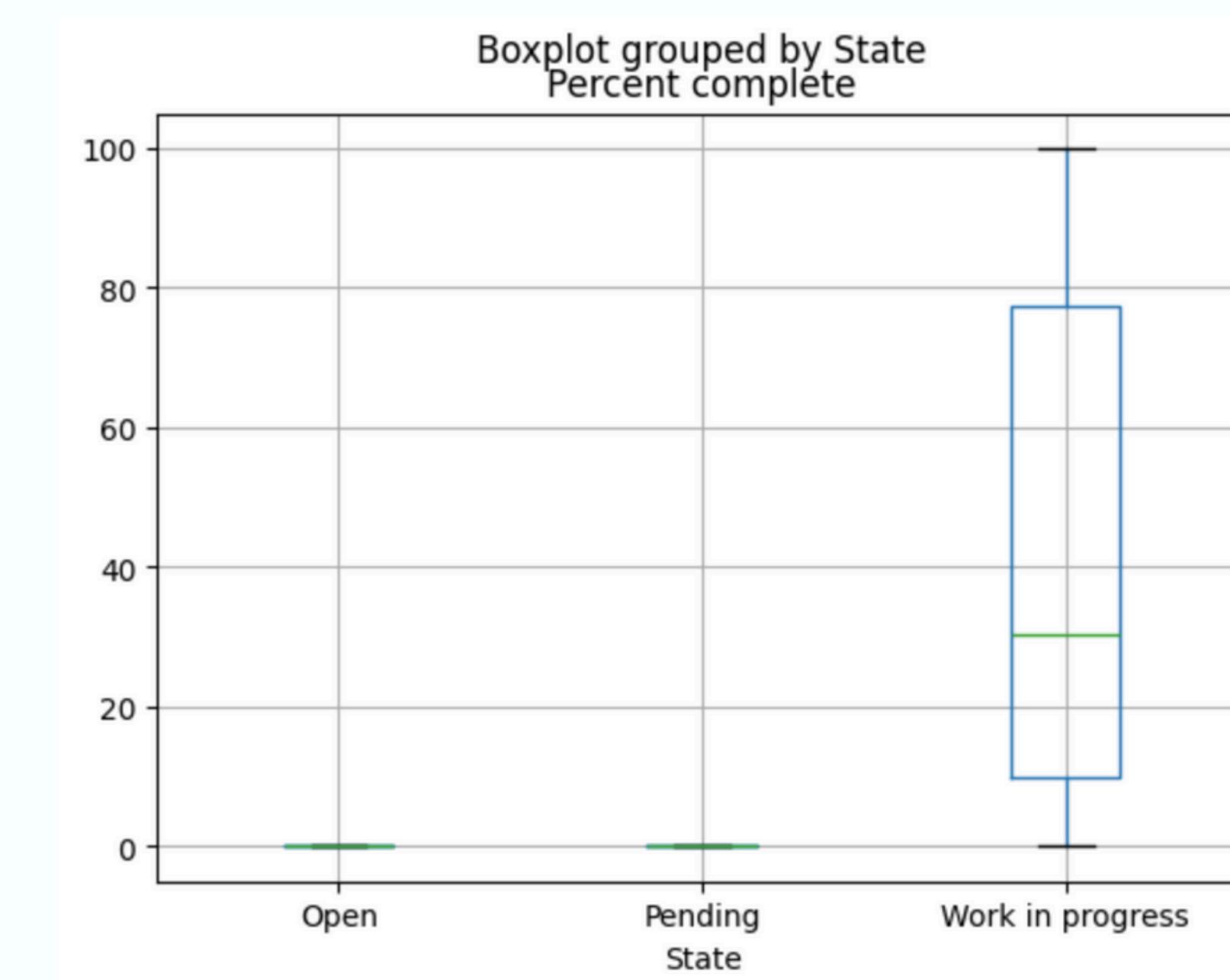
- Existe una interacción significativa entre el código taxonómico (TaxonCode) y el sitio de muestreo (CodeSite_SamplingOperations) para Abundance_nbcell y Abundance_pm.
- El ANOVA confirma diferencias importantes en la media de abundancia entre las distintas especies (TaxonCode) para Abundance_nbcell y Abundance_pm.
- Para la Abundancia Total por Operación (TotalAbundance_SamplingOperation), solo el Sitio de Muestreo resulta ser un factor significativo.
- Las pruebas estadísticas (ANOVA Factorial) rechazan la hipótesis nula, confirmando que las medias de abundancia son diferentes a través de los grupos analizados.

ACTIVIDAD 5.2

fORVIANOS.py

ANOVA PERCENT COMPLETE

- **Variable objeto:** Percent Complete
- **Variable Categorica:** State



ANOVA PERCENT COMPLETE

Variable objeto: Percent complete

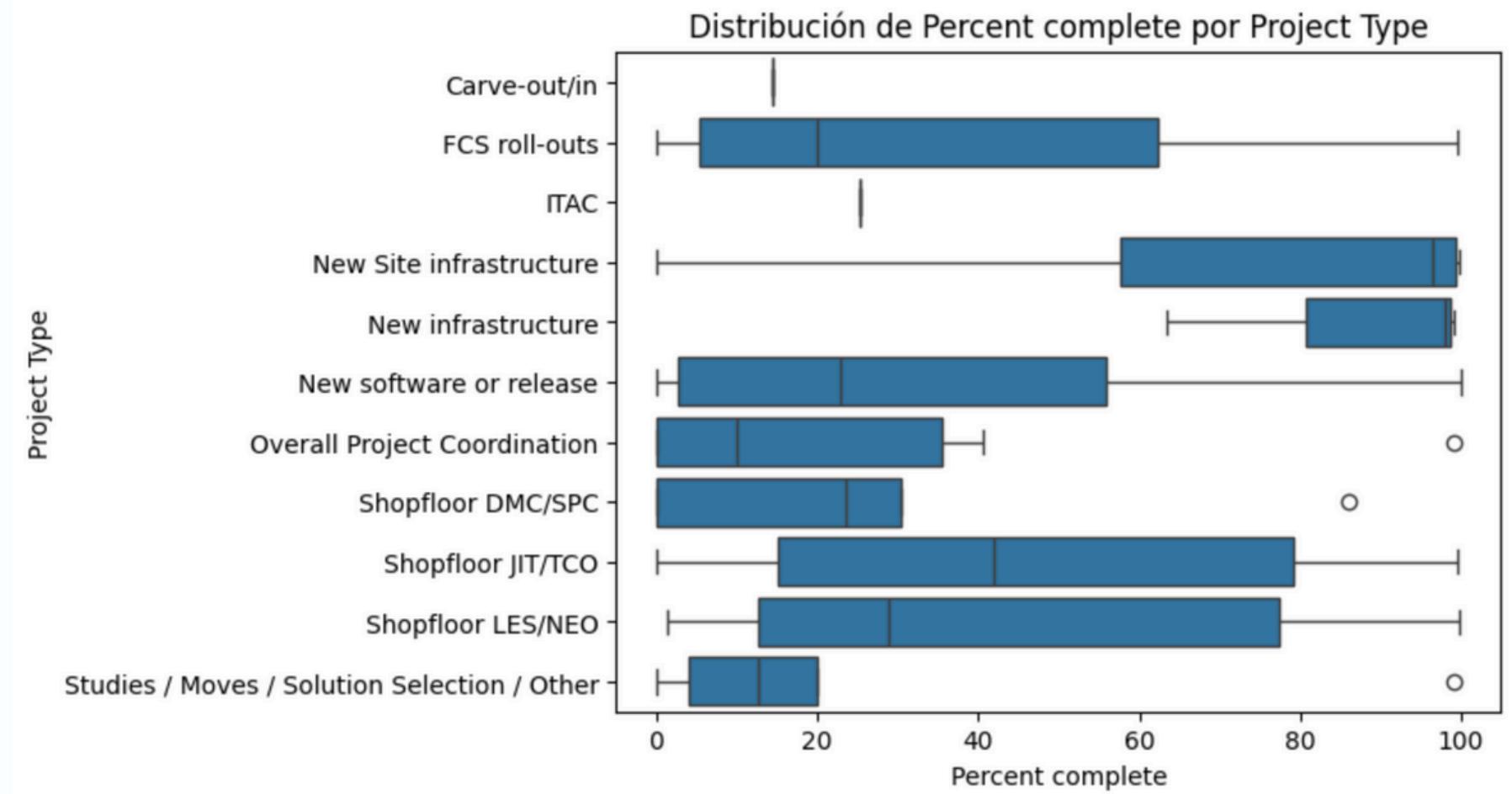
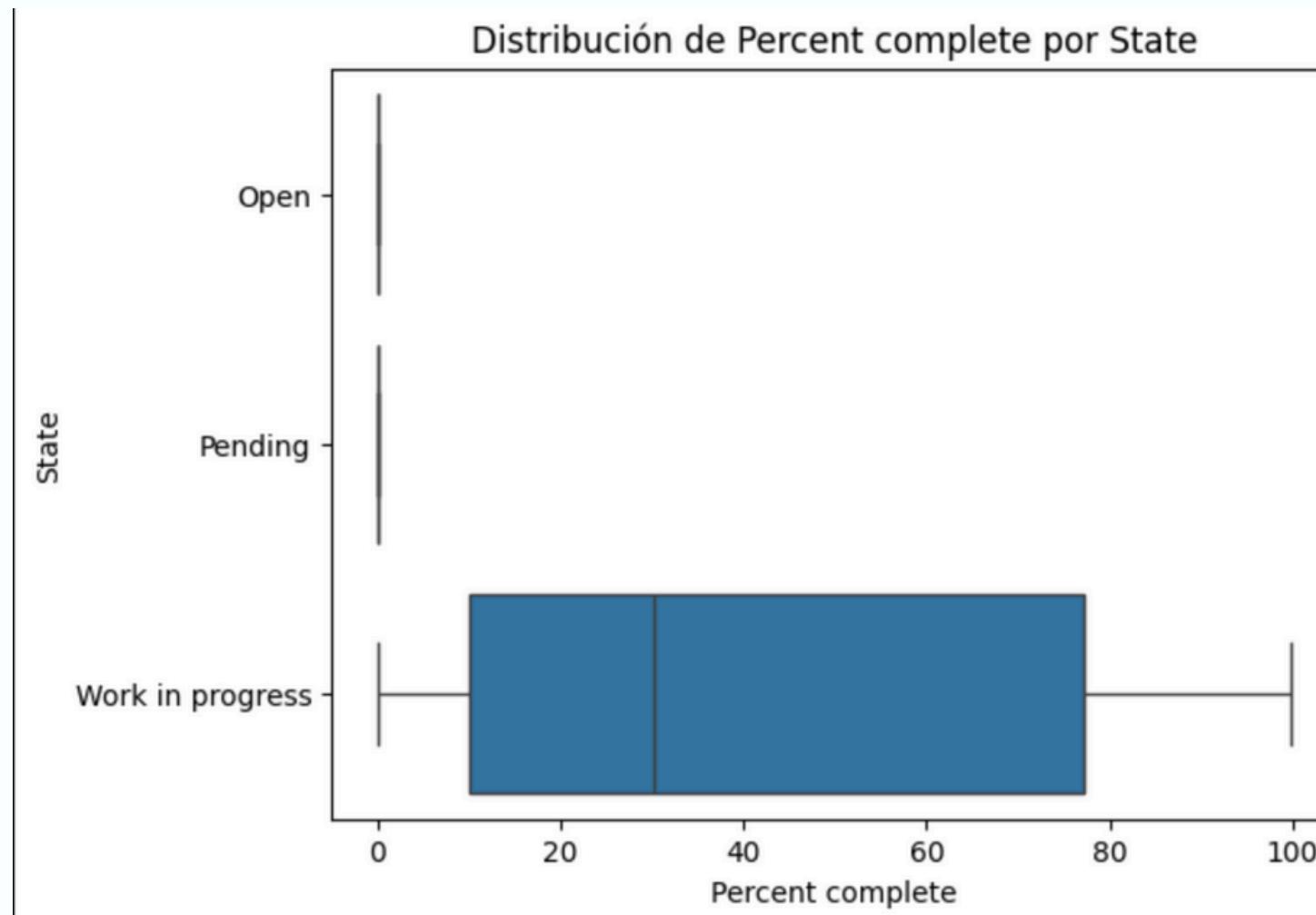
Variables Categorica: State

	estadístico de prueba F	p-value
State	9.043084	0.000163

Se rechaza la hipótesis nula

ANOVA FACTORIAL PERCENT COMPLETE

- **Variable objeto:** Percent complete
- **Variable Categorica:** State, Project Type



ANOVA FACTORIAL PERCENT COMPLETE

Variable objeto: Percent complete

Variables Categorica: State, Project Type

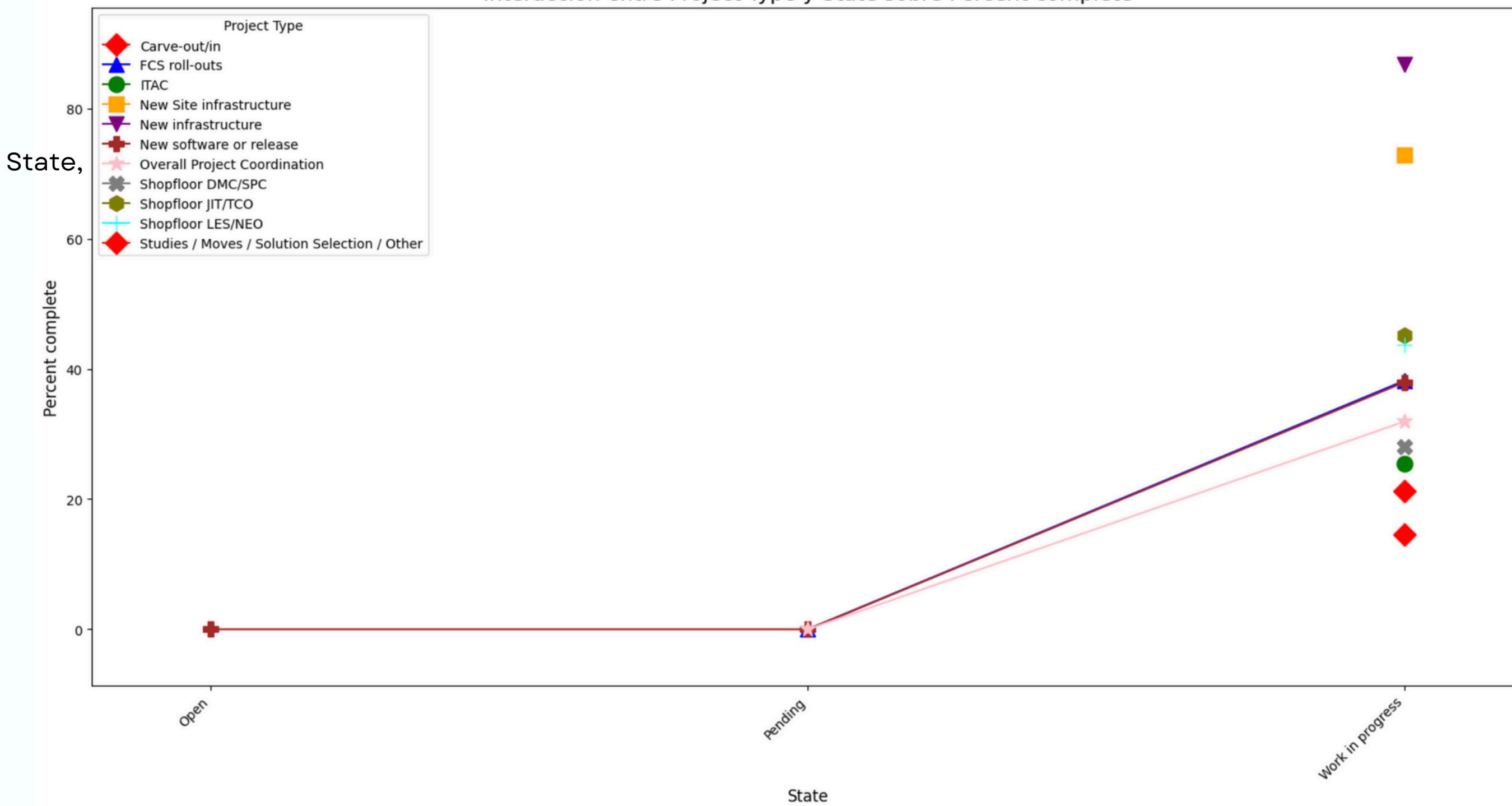
	estadístico de prueba F	p-value
State	9.285982	0.000132
Project Type	1.850123	0.053258
State: Project Type	1.016204	0.443813

Solo se rechaza la hipótesis con la variable
State

ANOVA FACTORIAL TOTAL PERCENT COMPLETE

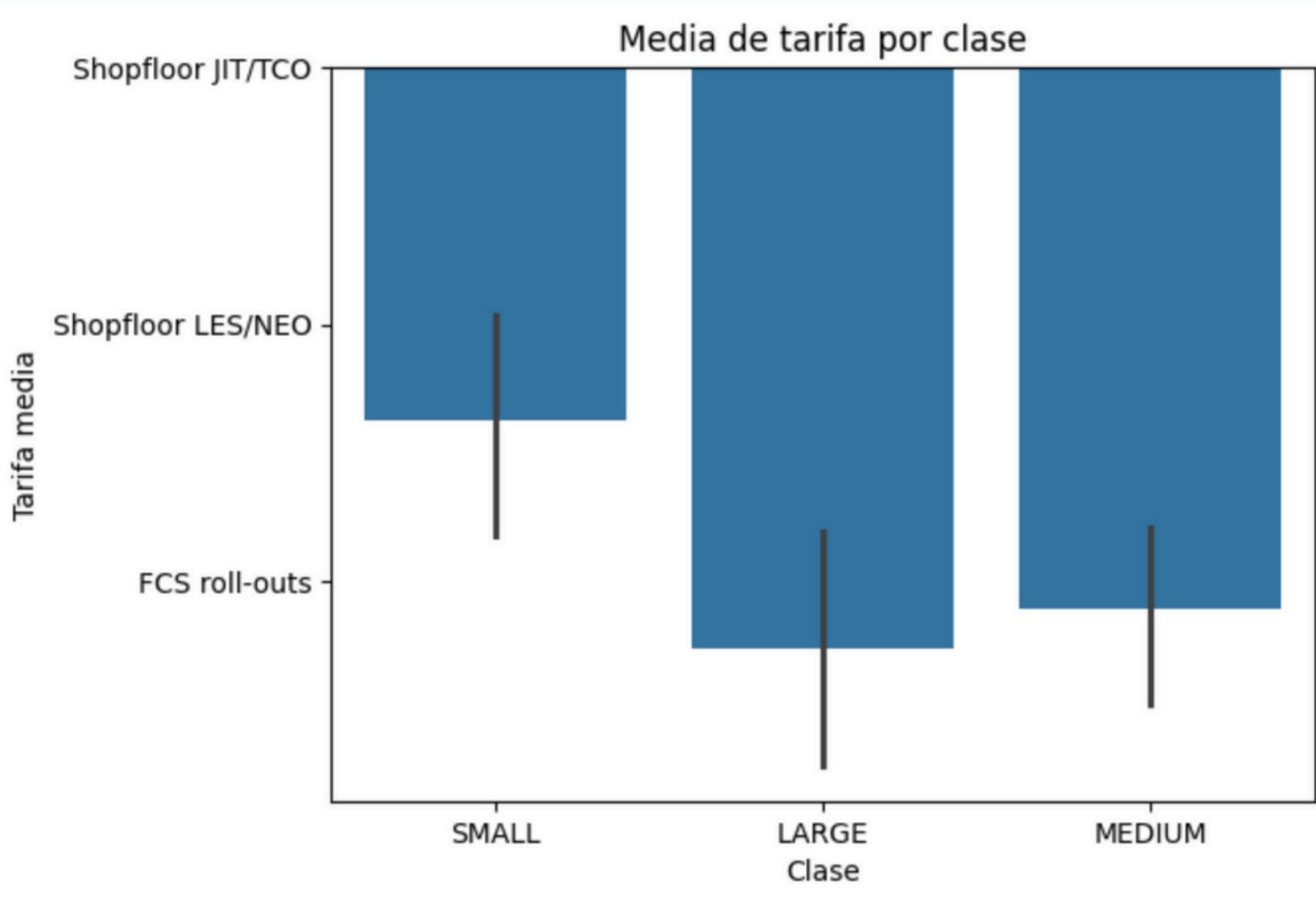
Interacción entre Project Type y State sobre Percent complete

- **Variable objeto:** Percent complete
- **Variable Categorica:** State, Project Type



ANOVA PROJECT SIZE

- **Variable objeto:** Project Size
- **Variable Categorica:** Project Type



ANOVA PROJECT SIZE

Variable objeto: Project size

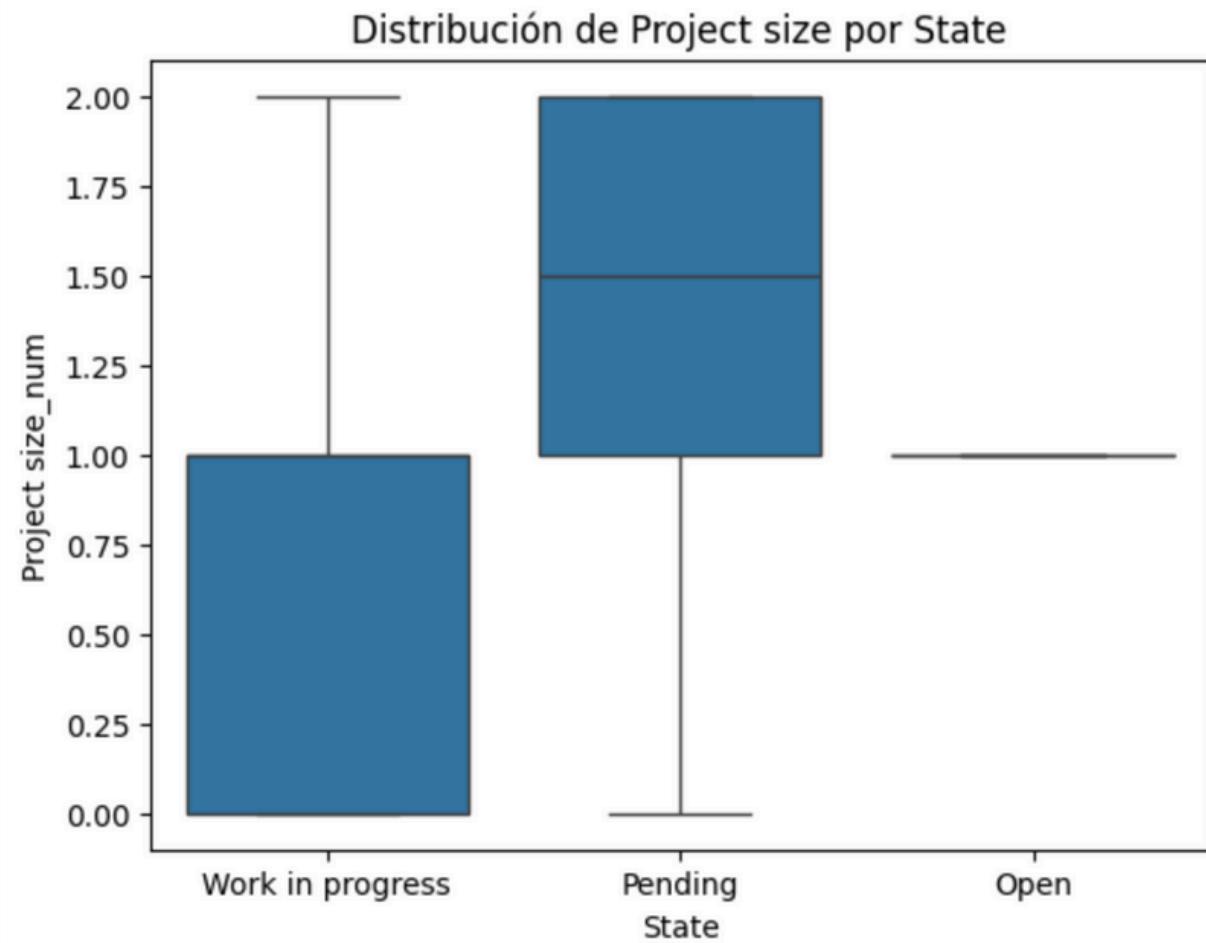
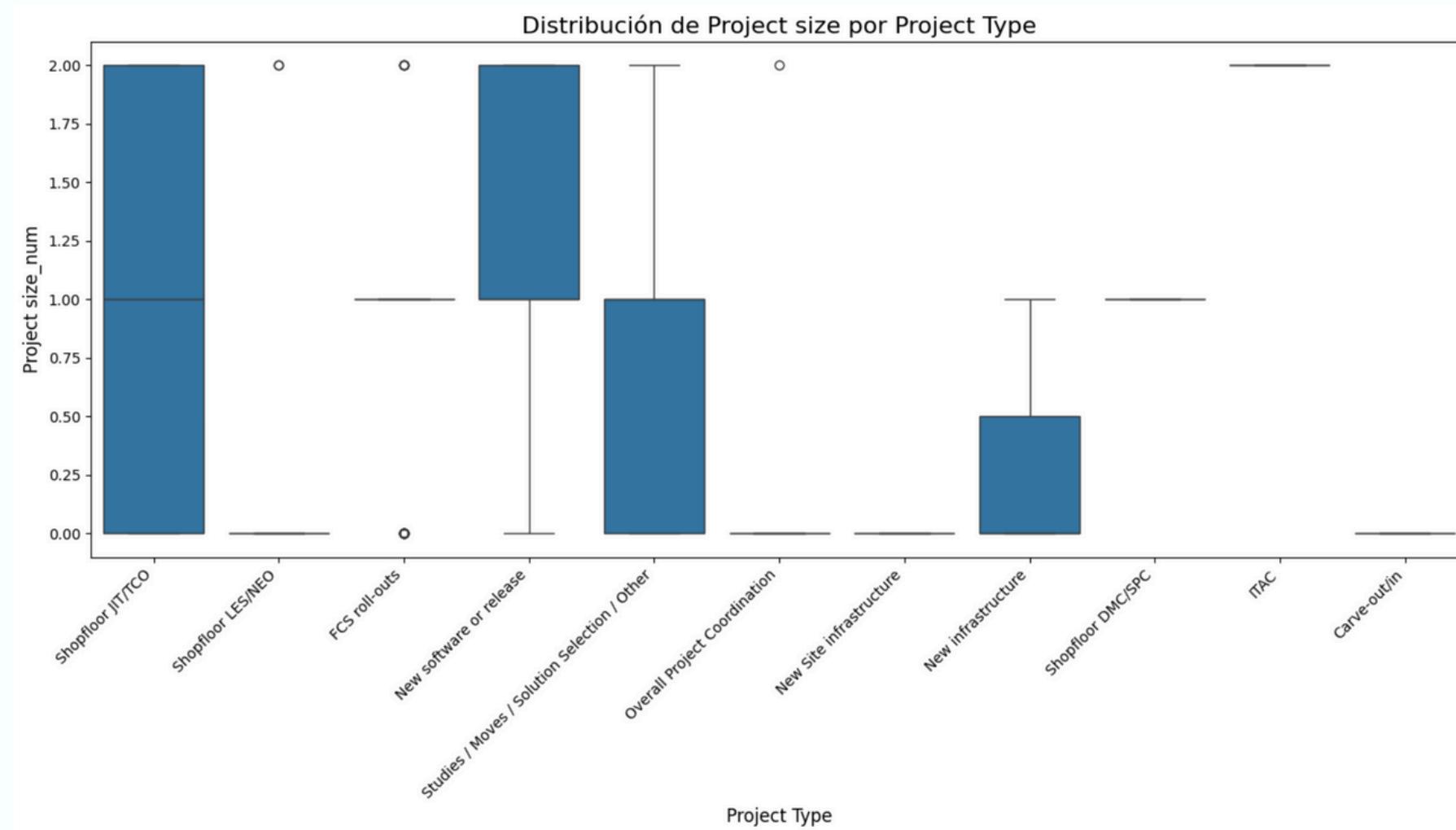
Variables Categorica: Project Type

	estadístico de prueba F	p-value
Project Type	4.741425	0.000003

Se rechaza la hipótesis nula

ANOVA FACTORIAL PROJECT SIZE

- **Variable objeto:** Project size
- **Variable Categorica:** State, Project Type



ANOVA FACTORIAL PROJECT SIZE

Variable objeto: Project size

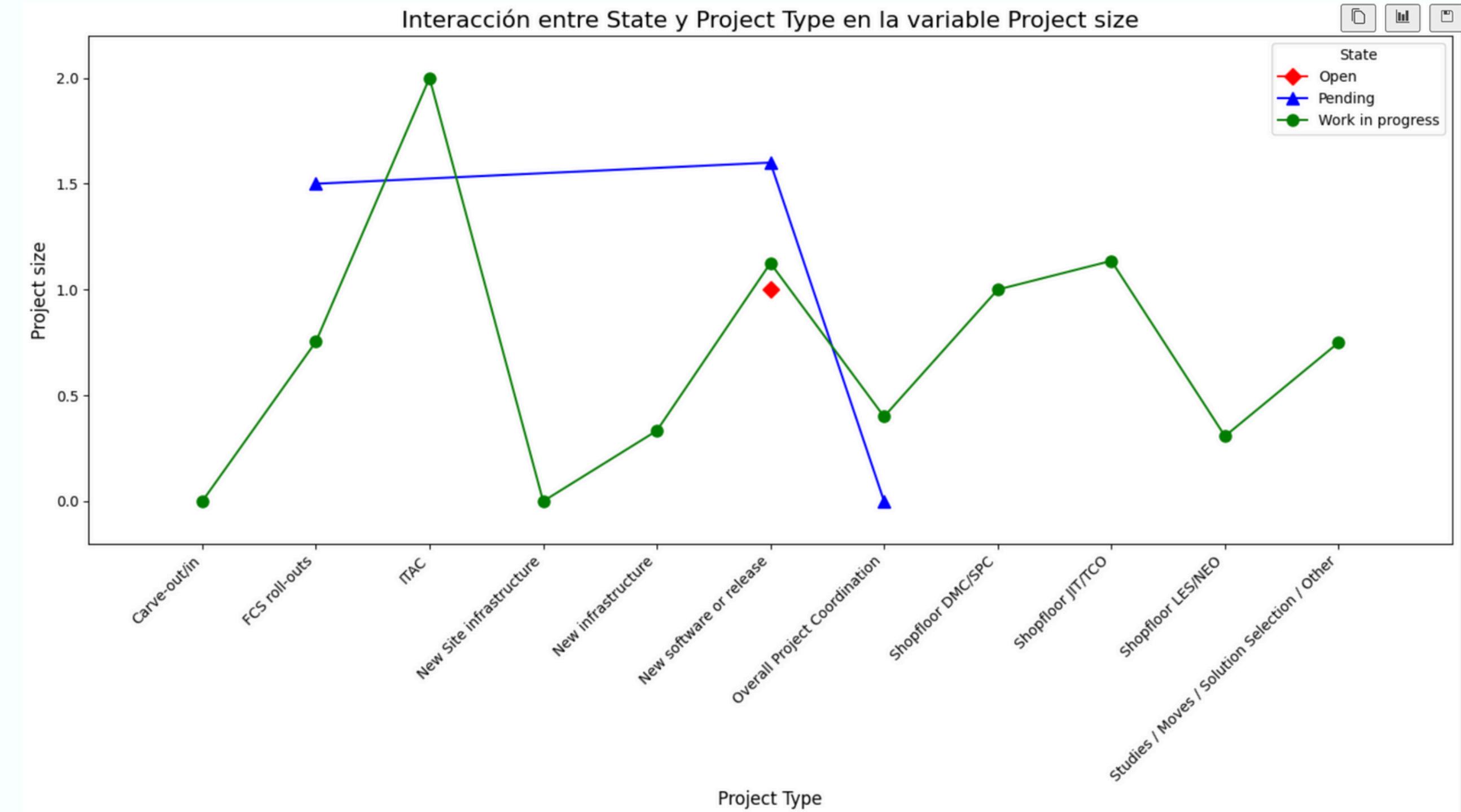
Variables Categorica: State, Project Type

	estadístico de prueba F	p-value
State	3.141577	0.045067
Project Type	4.868311	0.000002
State: Project Type	1.391721	0.177566

Solo se rechaza la hipótesis con las variables en conjunto

ANOVA FACTORIAL TOTAL PROJECT SIZE

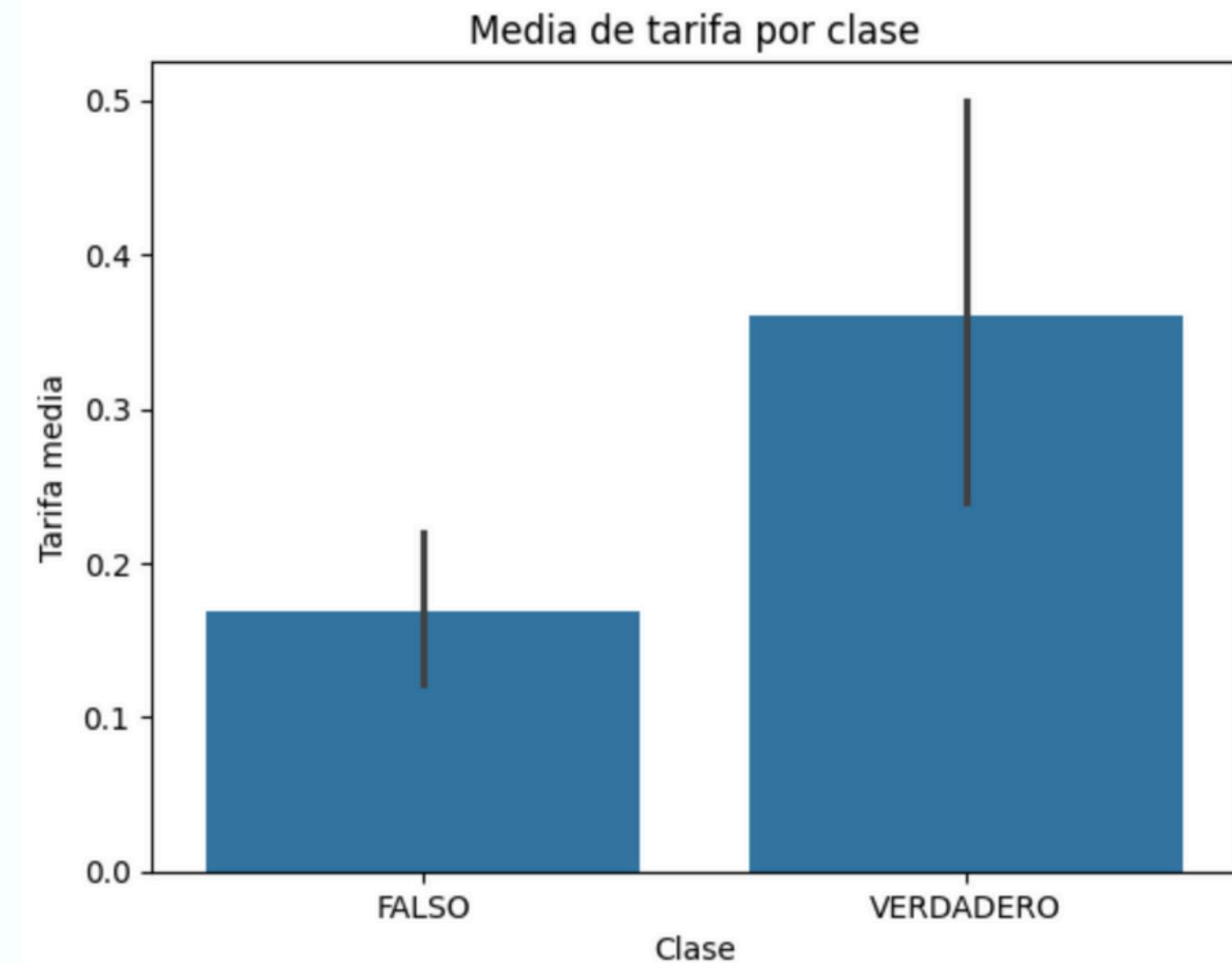
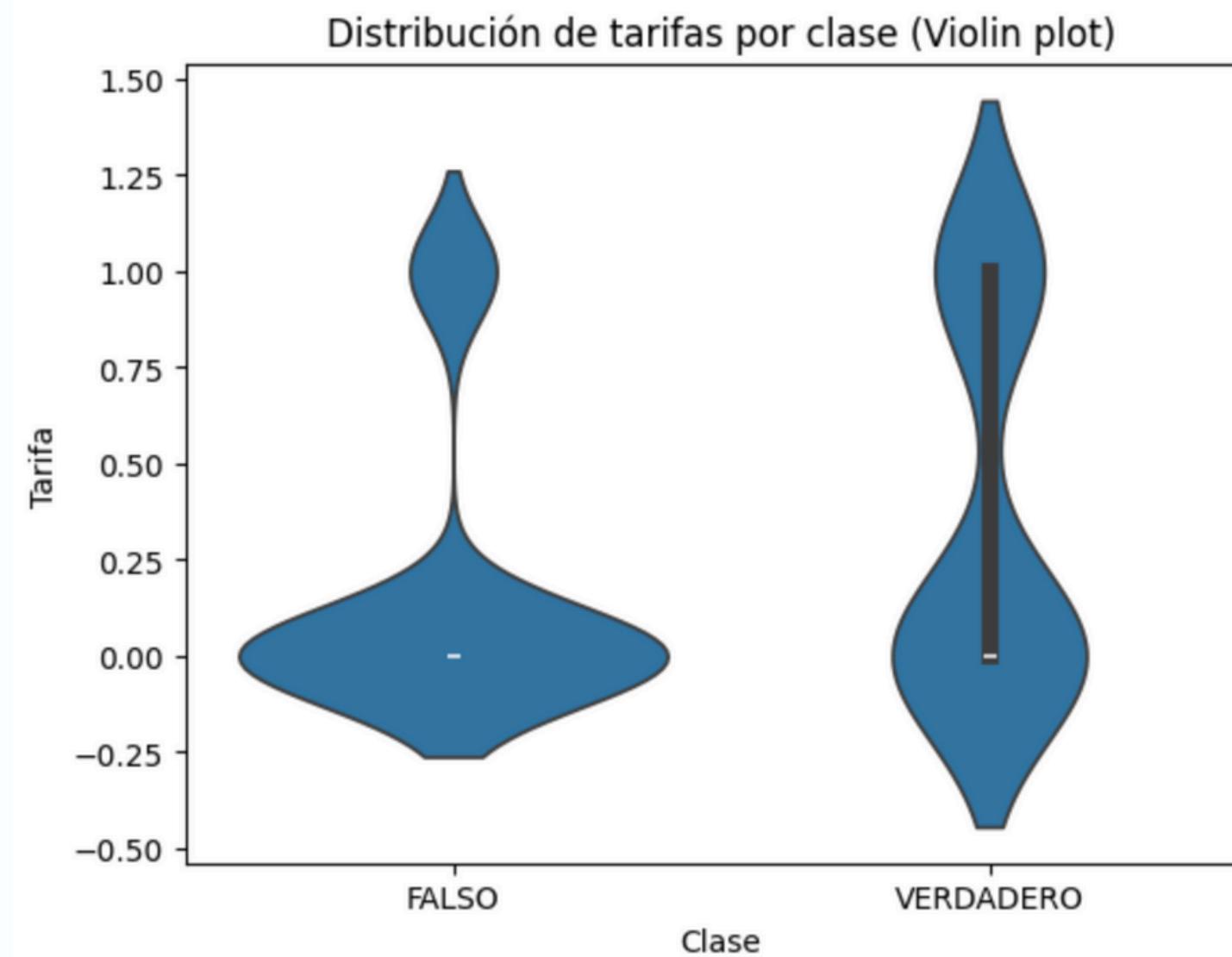
- **Variable objeto:** Percent complete
- **Variable Categorica:** State, Project Type



ANOVA PROJECT

HEALTH

- **Variable objeto:** Project Health
- **Variable Categorica:** On-hold



ANOVA PROJECT HEALTH

Variable objeto: Project Health

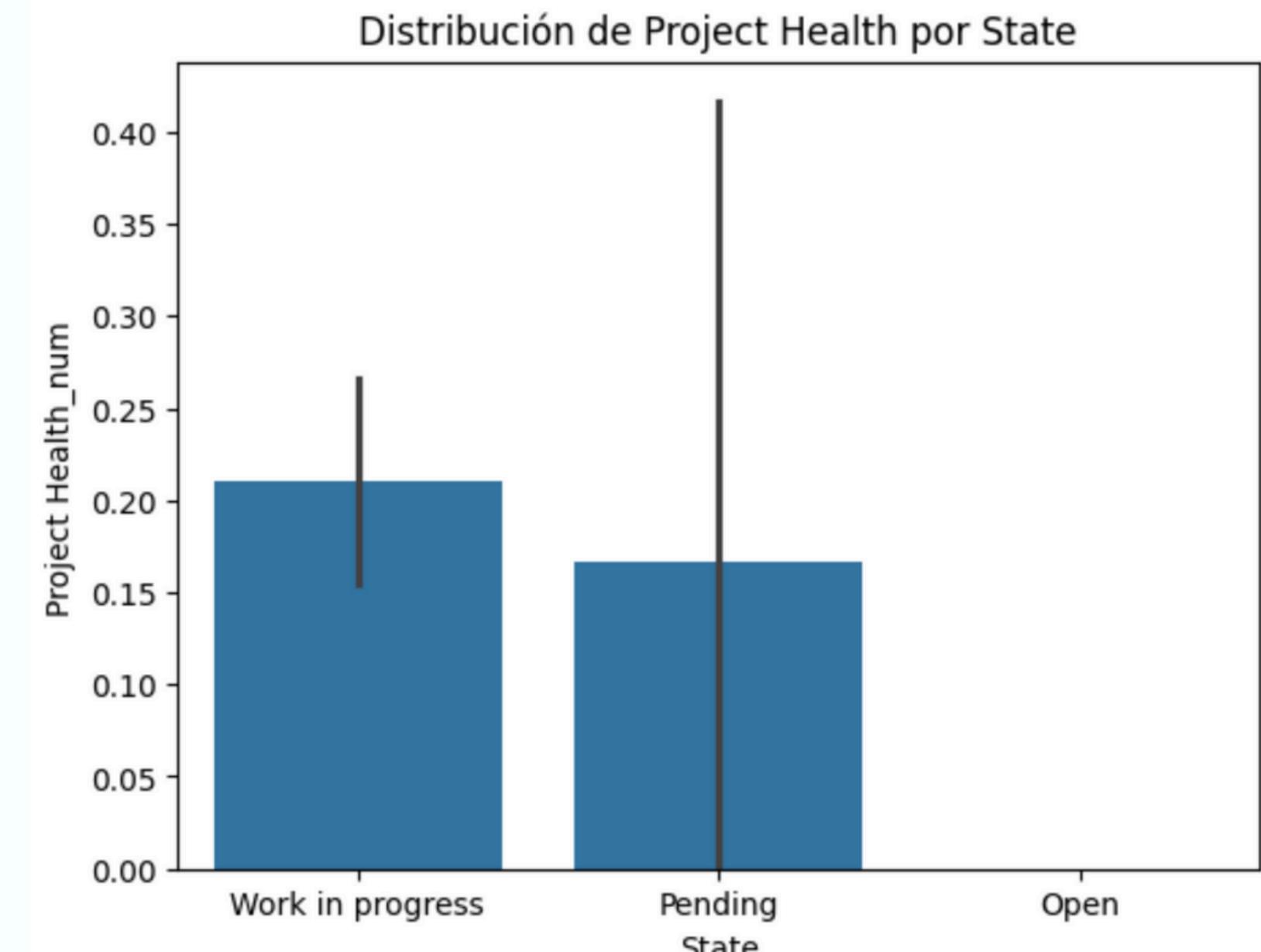
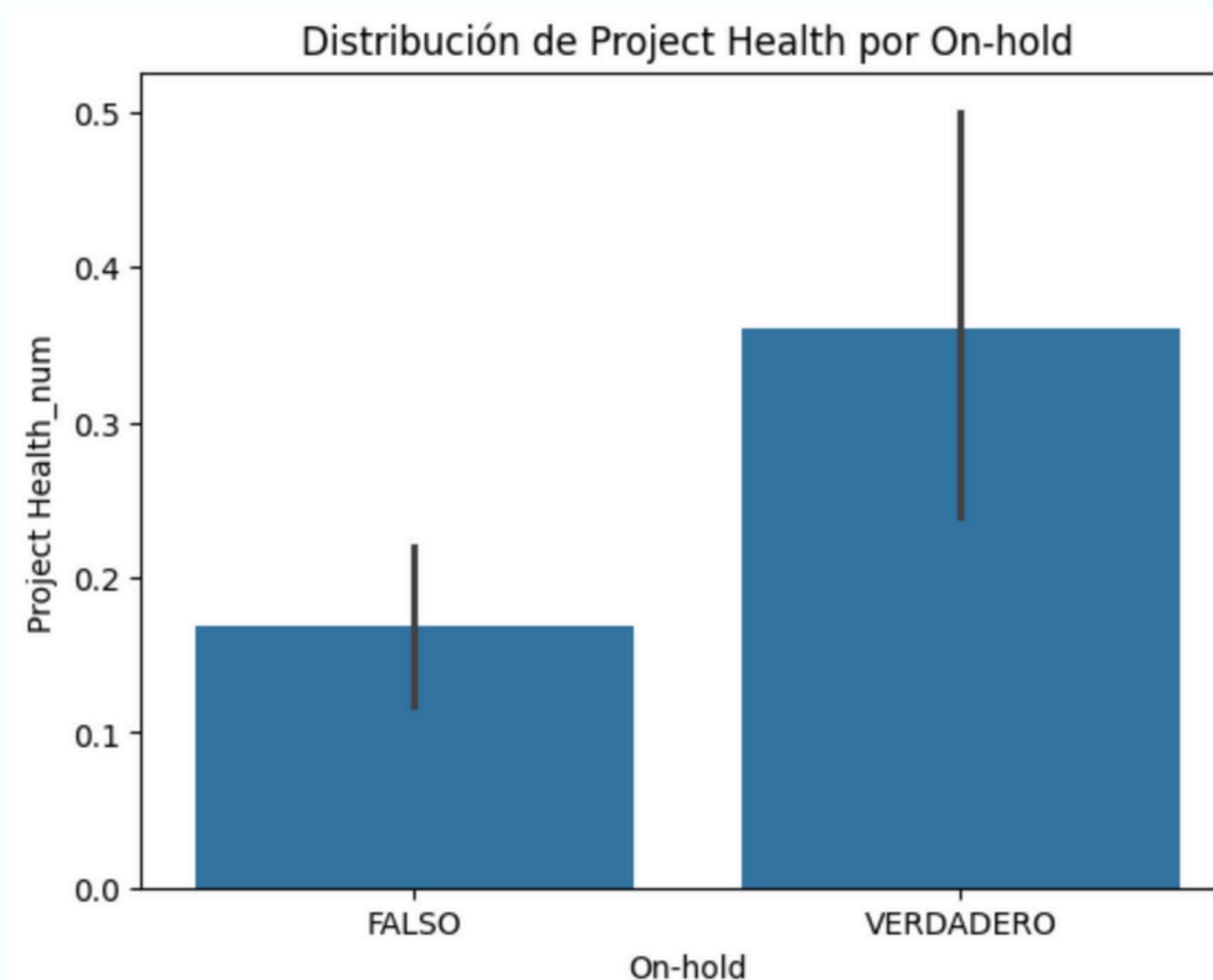
Variables Categorica: On-hold

	estadístico de prueba F	p-value
On-hold	9.161311	0.002737

Se rechaza la hipótesis nula

ANOVA FACTORIAL PROJECT HEALTH

- **Variable objeto:** Project Health
- **Variable Categorica:** On-hold, State



ANOVA FACTORIAL PROJECT HEALTH

Variable objeto: Project Health

Variables Categorica: State, On-hold

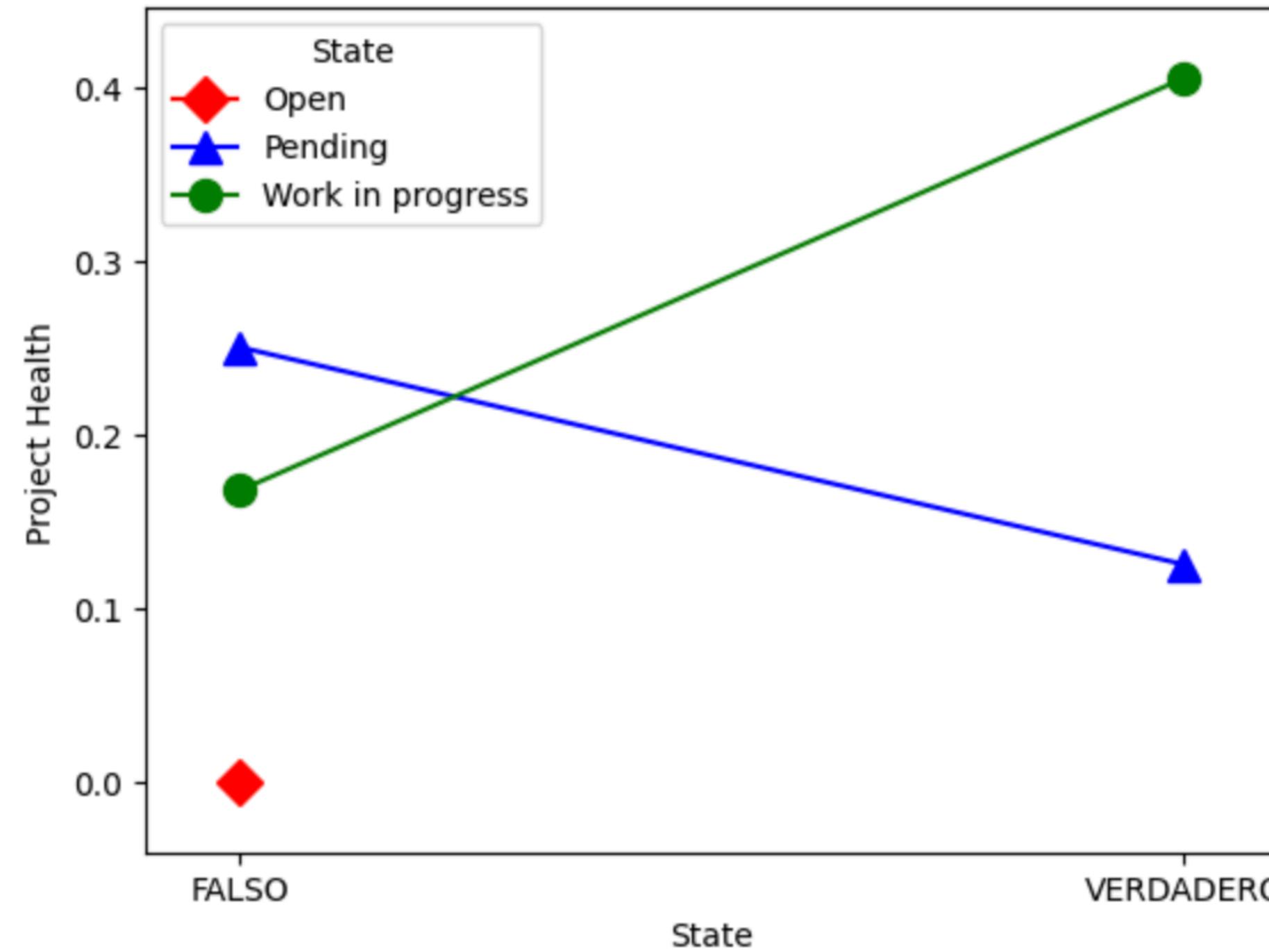
	estadístico de prueba F	p-value
State	0.203697	0.815849
On-hold	10.387898	0.001444
State: On-hold	1.258750	0.285869

Solo se rechaza la hipótesis con la variable
State

ANOVA FACTORIAL TOTAL PROJECT HEALTH

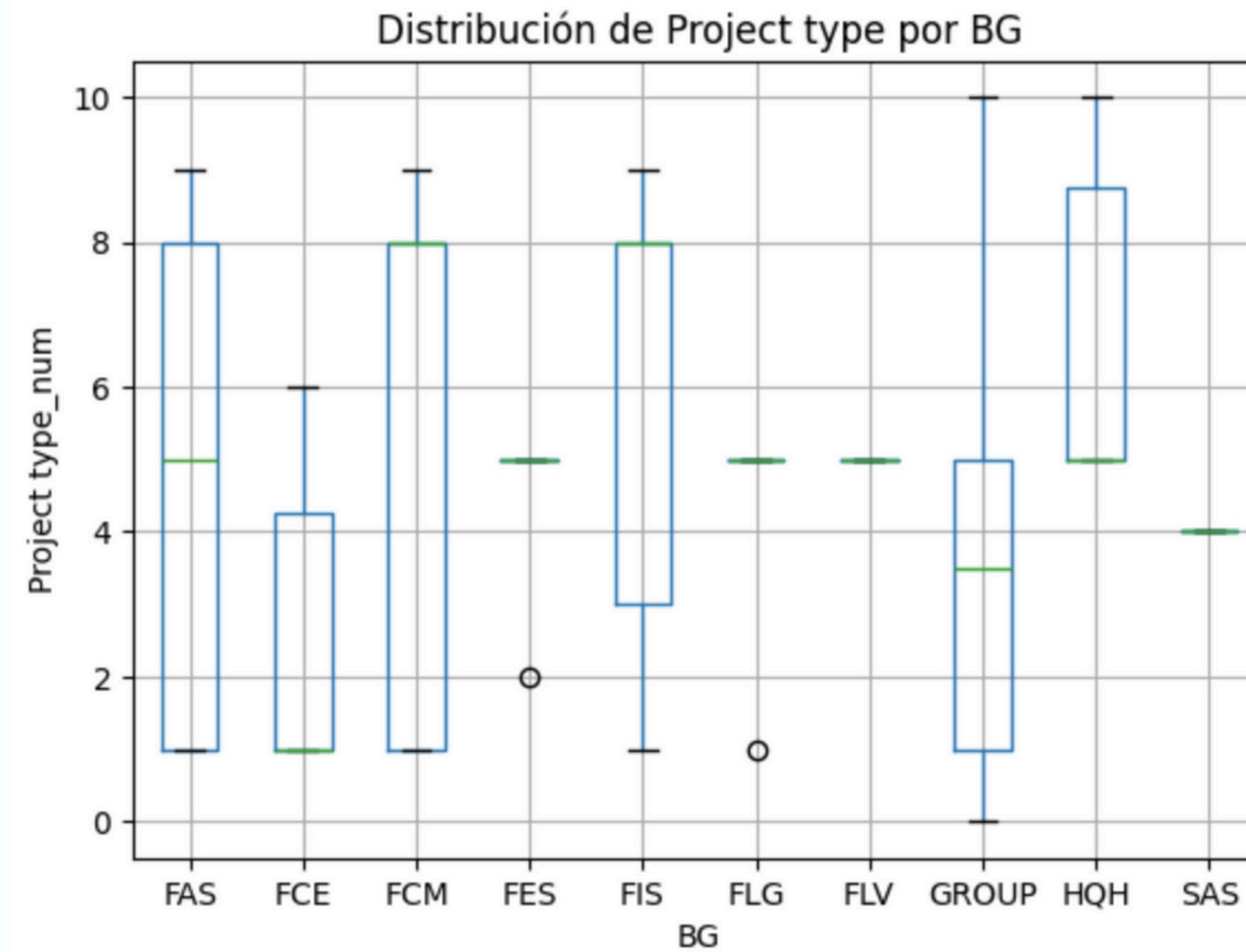
- **Variable objeto:** Project Health
- **Variable Categorica:** State, On-hold

Interacción entre la columna On-hold y State en la columna Project Health



ANOVA PROJECT TYPE

- **Variable objeto:** Project type
- **Variable Categorica:** BG



ANOVA PROJECT TYPE

Variable objeto: Project type

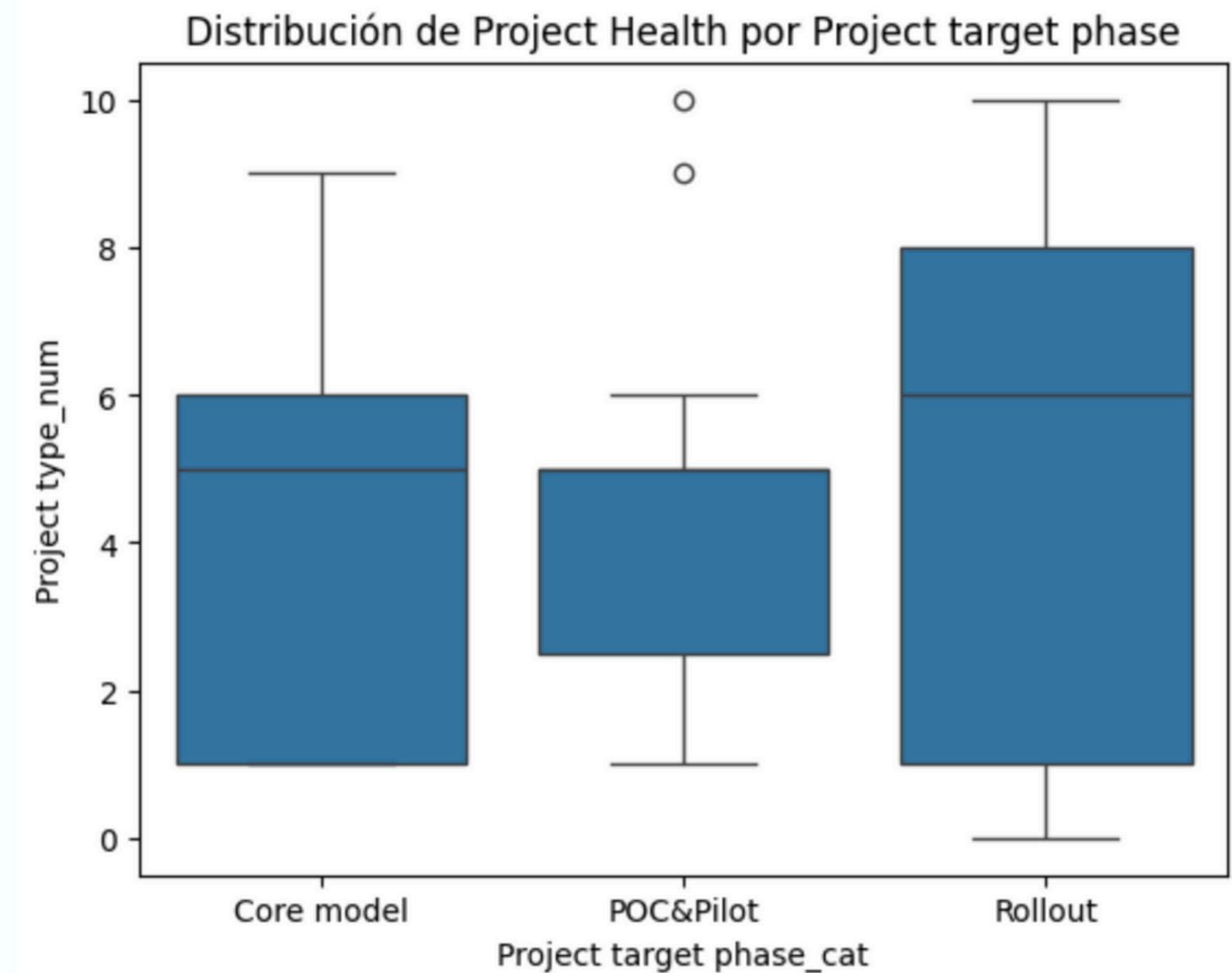
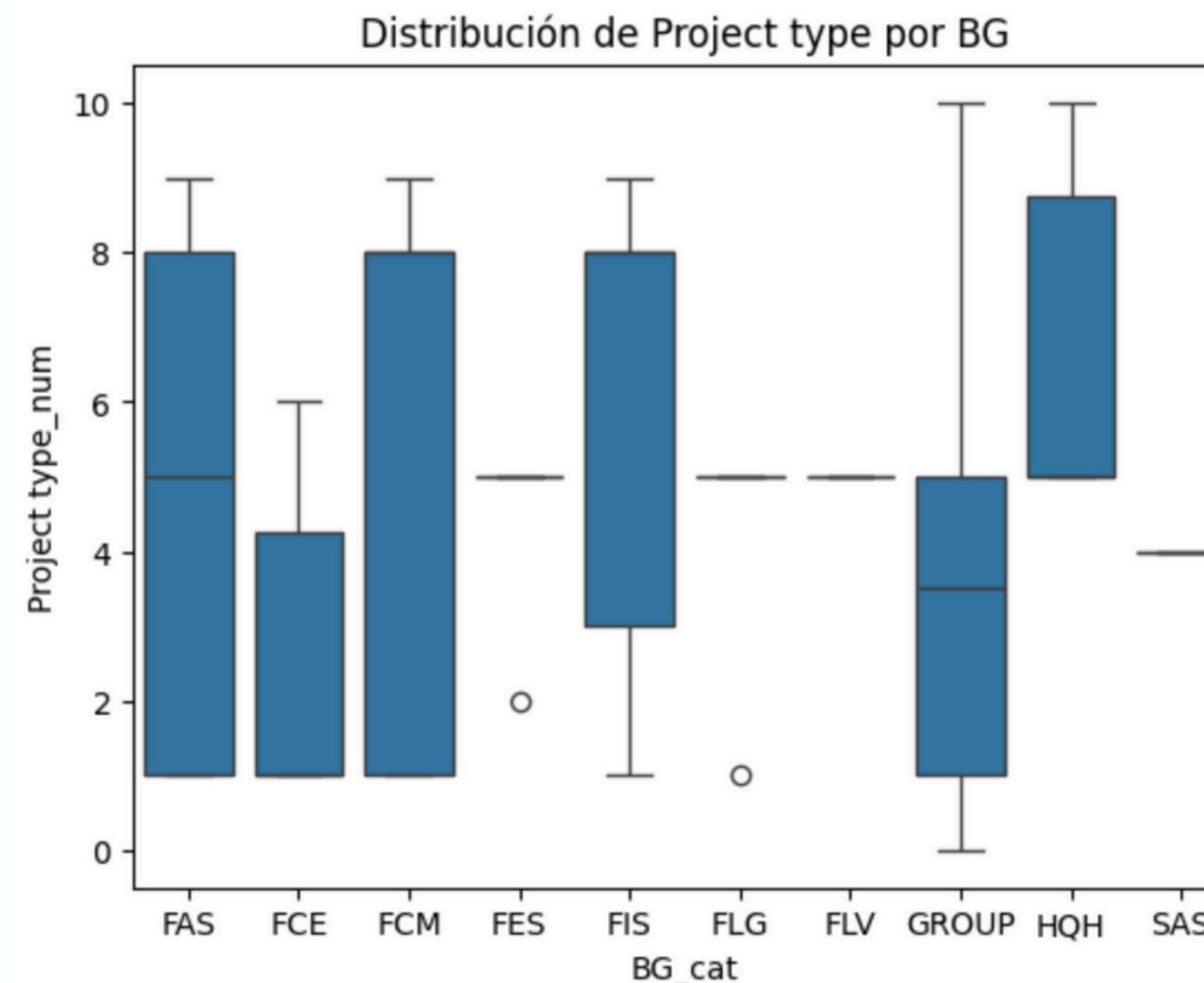
Variables Categorica: BG

	estadístico de prueba F	p-value
BG	2.806317	0.003786

Se rechaza la hipótesis nula

ANOVA FACTORIAL PROJECT TYPE

- **Variable objeto:** Project type
- **Variable Categorica:** BG, Project target phase



ANOVA FACTORIAL PROJECT TYPE

Variable objeto: Project type

Variables Categorica: BG, Project target phase

	estadístico de prueba F	p-value
BG	1.359896	0.258780
Project target phase	2.583652	0.007499
Project target phase: BG	1.412848	0.126610

Solo se rechaza la hipótesis con la variable
Project target phase

ANOVA FACTORIAL TOTAL PROJECT TYPE

- **Variable objeto:** Project type
- **Variable Categorica:** BG, Project target phase



CONCLUSIONES DATASET FORVIA

- El Estado del Proyecto (State) es el factor más significativo para determinar el Porcentaje de Completado.
- El Tipo de Proyecto (Project Type) es el principal factor que afecta el Tamaño del Proyecto seguido por el Estado.
- La condición de "En espera" (On-hold) tiene el impacto más significativo en la Salud del Proyecto (Project Health).
- La Fase Objetivo del Proyecto (Project target phase) influye significativamente en el Tipo de Proyecto.
- Los análisis de varianza (ANOVA/MANOVA) resultaron en el rechazo de la hipótesis nula para los factores clave en cada variable objetivo, indicando que hay diferencias estadísticamente significativas en las medias.

CONCLUSIONES

- Los análisis ANOVA y MANOVA fueron efectivos para identificar los impulsores estadísticamente significativos en cada conjunto de datos.
- En el Datathon, se identificó una interacción compleja en la abundancia biológica, mostrando que los diferentes factores varían por la ubicación de muestreo.
- Para FORVIA, se aisló el impacto de variables específicas, como: el Estado para el porcentaje de completado, el Tipo para el tamaño, y la condición "On-hold" para la salud del proyecto.
- Los resultados demuestran que las variables categóricas sí tienen una influencia diferenciada en los resultados numéricos de interés para ambos estudios.
- La visualización de datos (Boxplots, Violin Plots y Lineplots de Interacción) fue clave para confirmar la dispersión y las tendencias observadas en los resultados estadísticos.

Octubre 2030

MUCHAS
GRACIAS

Bruno Lago