

Instituto Tecnológico y de Estudios Superiores de Monterrey



Analítica de Datos y
Herramientas de Inteligencia Artificial

Reporte de actividad 5.1

Profesor: Alfredo García Suárez

Camila Trujillo Beristain | A01737170
Bernardo Quintana López | A01658064
Fernando Guadarrama González | A01379340
Mauricio Goris García | A01736428

Campus Puebla

2 de mayo de 2025

Comenzamos cargando las librerías necesarias

```
#Cargamos librerías
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
from statsmodels.formula.api import ols

✓ 1.4s
```

Python

Después cargamos los datos y limpiamos los datos nulos

```
#Cargar archivo csv desde seaborn
df = pd.read_csv("DataAnalytics.csv")
df.head()
```

✓ 0.0s

Python

```
#Rellenamos nulos
df = df.fillna(method="bfill")
df = df.fillna(method="ffill")
df = df.fillna(method="bfill")
```

✓ 0.0s

Python

Posteriormente identificamos las variables numéricas y categóricas para su comparación ANOVA y comenzamos con el primero, teniendo tiempo de interacción como variable y Juego y Dificultad como factores de comparación.

```
# Variables numéricas y factores categóricos
numeric_vars = df.select_dtypes(include=[np.number]).columns.tolist()
factores = ['Juego', 'dificultad', 'mini juego']
```

✓ 0.0s

Python

```
# Seleccionar variable y factores
variable = 'tiempo de interacción'
factor1 = 'Juego'
factor2 = 'dificultad'

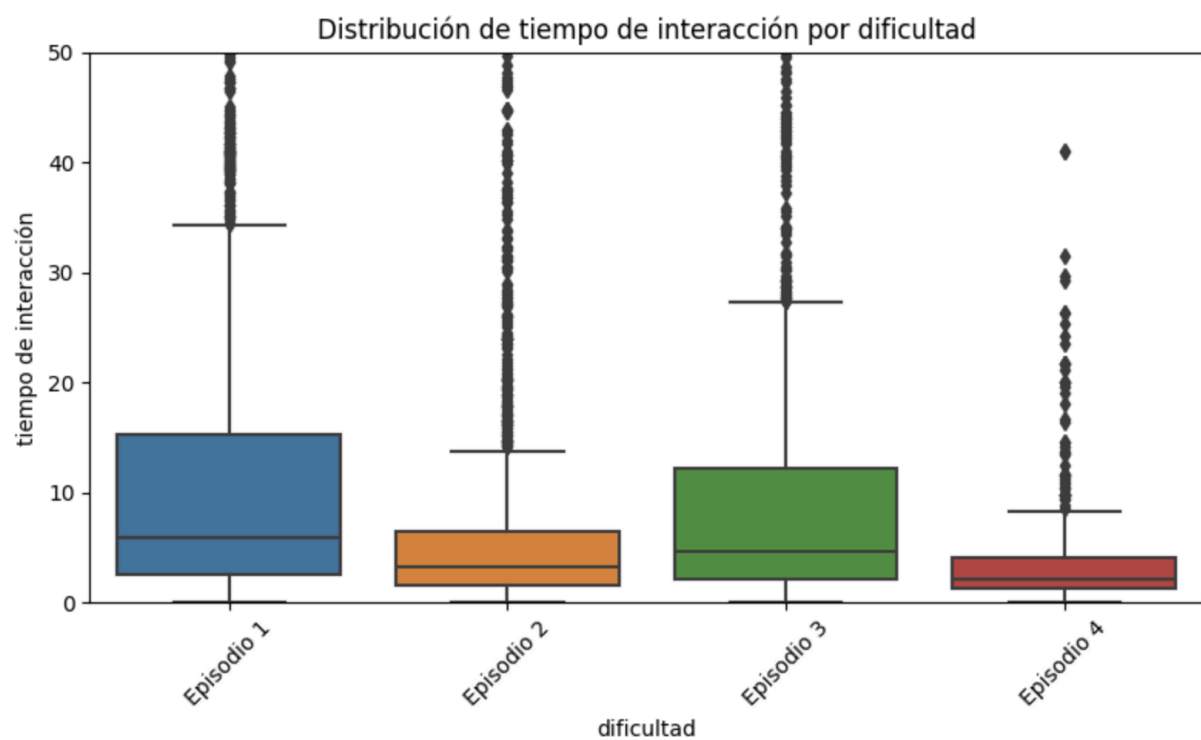
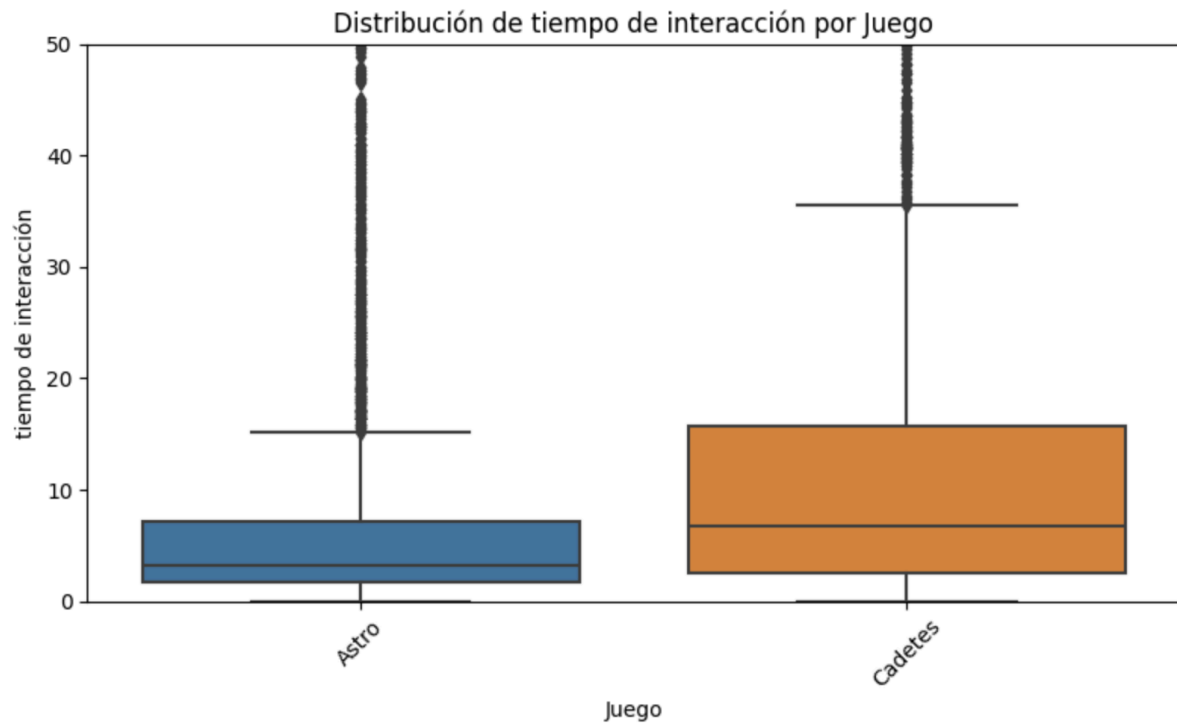
for factor in [factor1, factor2]:
    print(f"\nANÁLISIS DE VARIANZA PARA: {variable} ~ {factor}")
    try:
        formula = f'Q("{variable}") ~ C({factor})'
        model = ols(formula, data=df).fit()
        anova_table = sm.stats.anova_lm(model, typ=2)
        print(anova_table)

        p_val = anova_table["PR(>F)"][0]
        if p_val < 0.05:
            print(f"→ Efecto significativo de {factor} (p = {p_val:.4f})")
        else:
            print(f"→ No hay diferencias significativas para {factor} (p = {p_val:.4f})")

    except Exception as e:
        print(f"Error al analizar {factor}: {e}")

# Boxplot por Juego
plt.figure(figsize=(8, 5))
sns.boxplot(x=factor1, y=variable, data=df)
plt.title(f'Distribución de {variable} por {factor1}')
plt.xticks(rotation=45)
plt.ylim(0, 50)
plt.tight_layout()
plt.show()

# Boxplot por Dificultad
plt.figure(figsize=(8, 5))
sns.boxplot(x=factor2, y=variable, data=df)
plt.title(f'Distribución de {variable} por {factor2}')
plt.xticks(rotation=45)
plt.ylim(0, 50)
```



Después número de interacción como variable y Juego y Dificultad como factores de comparación.

```

# Seleccionar variable y factores
variable = 'número de interacción'
factor1 = 'Juego'
factor2 = 'dificultad'

for factor in [factor1, factor2]:
    print(f"\nANÁLISIS DE VARIANZA PARA: {variable} ~ {factor}")
    try:
        formula = f'Q("{variable}") ~ C({factor})'
        model = ols(formula, data=df).fit()
        anova_table = sm.stats.anova_lm(model, typ=2)
        print(anova_table)

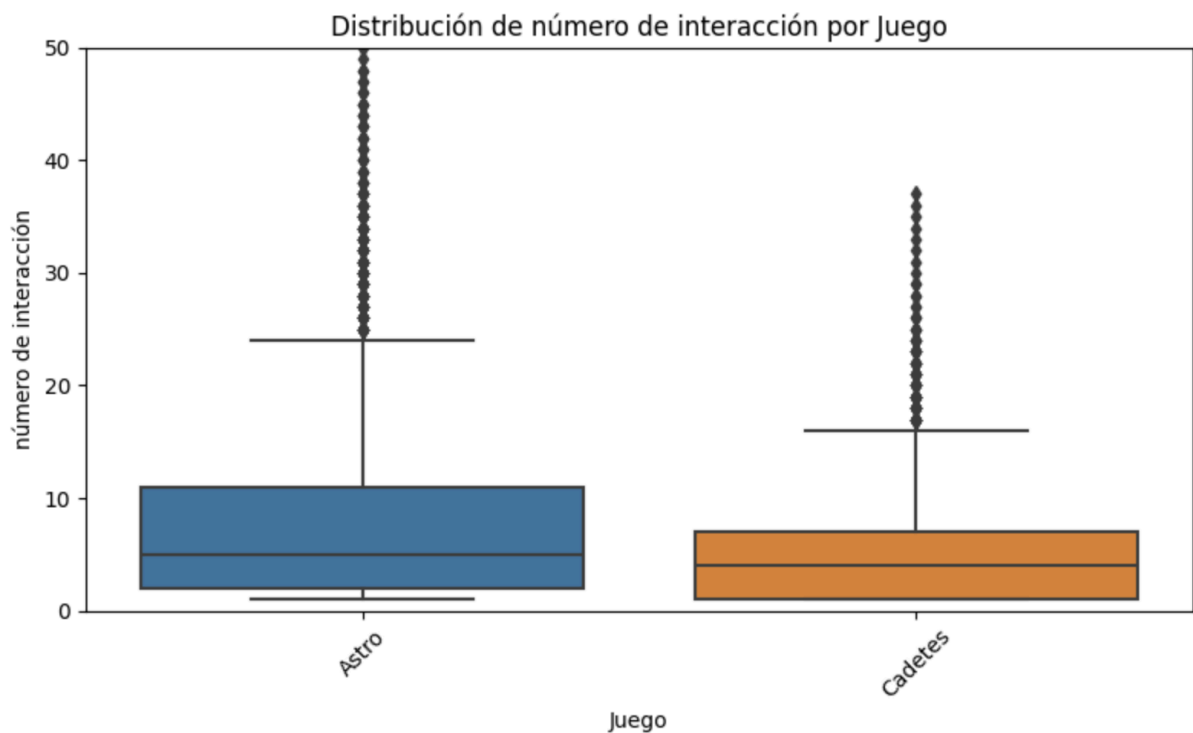
        p_val = anova_table["PR(>F)"][0]
        if p_val < 0.05:
            print(f"~ Efecto significativo de {factor} (p = {p_val:.4f})")
        else:
            print(f"~ No hay diferencias significativas para {factor} (p = {p_val:.4f})")

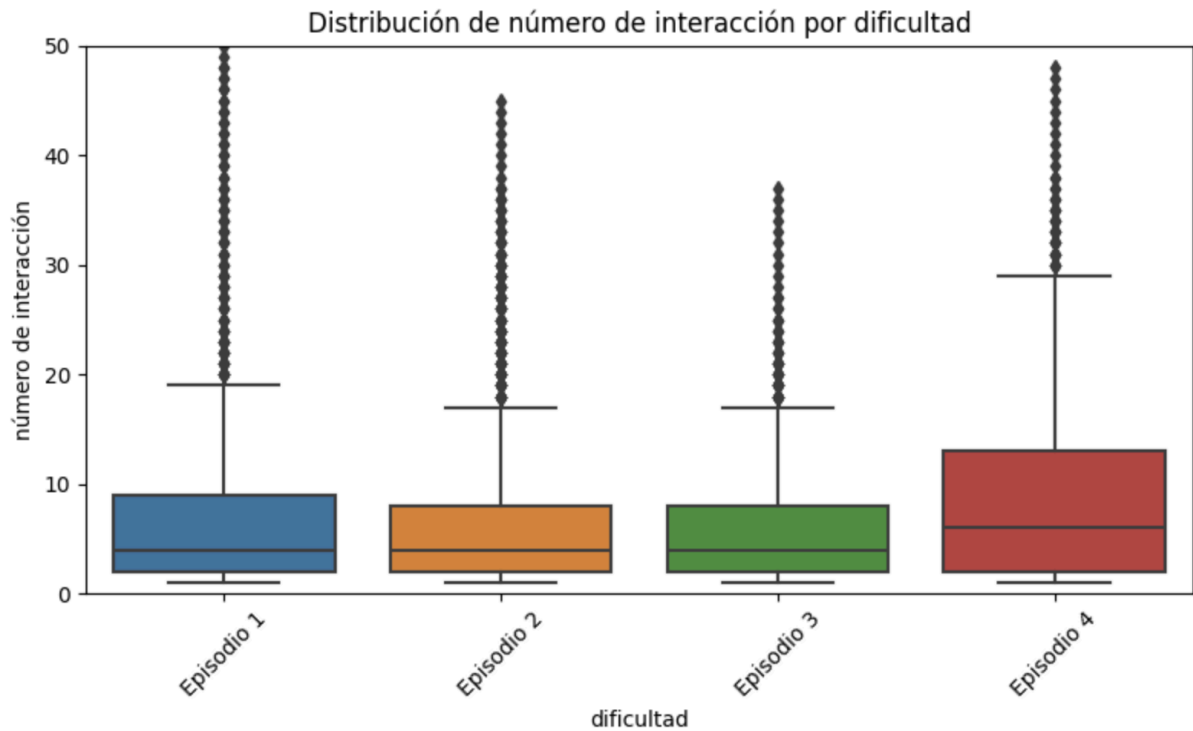
    except Exception as e:
        print(f"Error al analizar {factor}: {e}")

# Boxplot por Juego
plt.figure(figsize=(8, 5))
sns.boxplot(x=factor1, y=variable, data=df)
plt.title(f'Distribución de {variable} por {factor1}')
plt.xticks(rotation=45)
plt.ylim(0, 50)
plt.tight_layout()
plt.show()

# Boxplot por Dificultad
plt.figure(figsize=(8, 5))
sns.boxplot(x=factor2, y=variable, data=df)
plt.title(f'Distribución de {variable} por {factor2}')
plt.xticks(rotation=45)
plt.ylim(0, 50)
plt.show()

```





Después auto push como variable y Juego y Dificultad como factores de comparación.

```
# Seleccionar variable y factores
variable = 'auto push'
factor1 = 'Juego'
factor2 = 'dificultad'

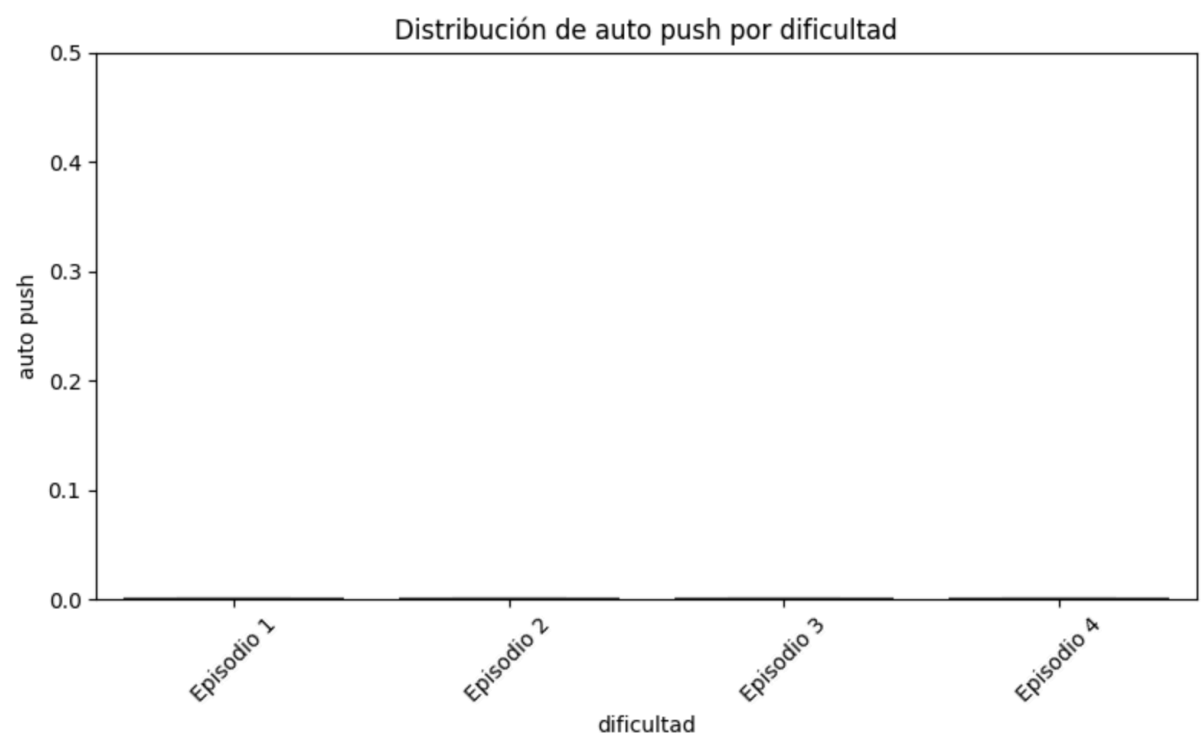
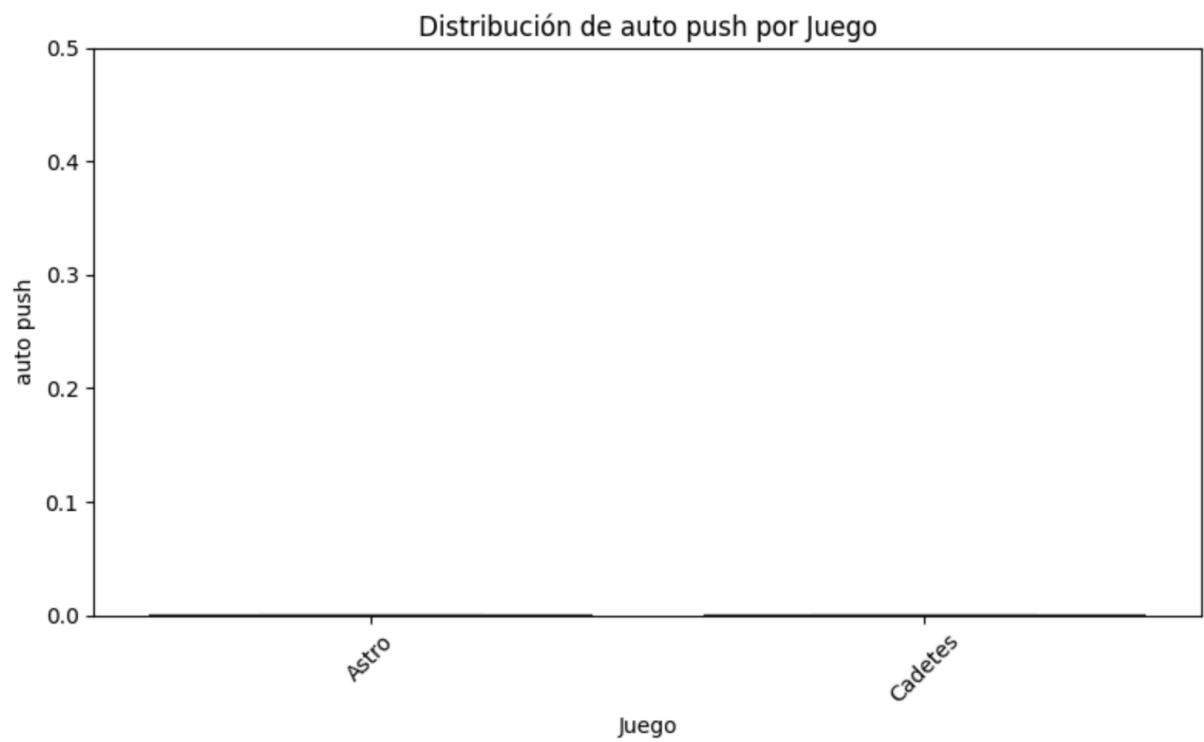
for factor in [factor1, factor2]:
    print(f"\nANÁLISIS DE VARIANZA PARA: (variable) ~ {factor}")
    try:
        formula = f'Q("{variable}") ~ C({factor})'
        model = ols(formula, data=df).fit()
        anova_table = sm.stats.anova_lm(model, typ=2)
        print(anova_table)

        p_val = anova_table["PR(>F)"][0]
        if p_val < 0.05:
            print(f"↪ Efecto significativo de {factor} (p = {p_val:.4f})")
        else:
            print(f"↪ No hay diferencias significativas para {factor} (p = {p_val:.4f})")

    except Exception as e:
        print(f"Error al analizar {factor}: {e}")

# Boxplot por Juego
plt.figure(figsize=(8, 5))
sns.boxplot(x=factor1, y=variable, data=df)
plt.title(f'Distribución de {variable} por {factor1}')
plt.xticks(rotation=45)
plt.ylim(0, 0.5)
plt.tight_layout()
plt.show()

# Boxplot por Dificultad
plt.figure(figsize=(8, 5))
sns.boxplot(x=factor2, y=variable, data=df)
plt.title(f'Distribución de {variable} por {factor2}')
plt.xticks(rotation=45)
plt.ylim(0, 0.5)
plt.show()
```



Después tiempo de lección como variable y Juego y Dificultad como factores de comparación.

```
# Seleccionar variable y factores
variable = 'tiempo de lección'
factor1 = 'Juego'
factor2 = 'dificultad'

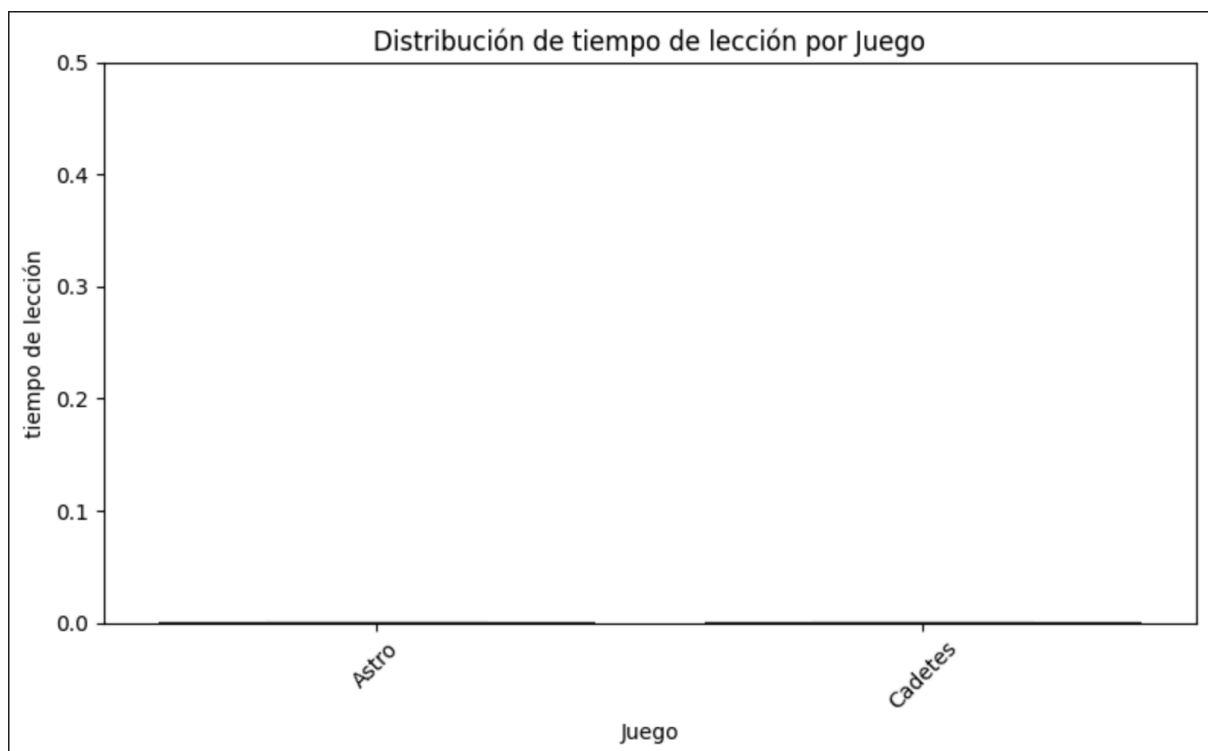
for factor in [factor1, factor2]:
    print(f"\nANÁLISIS DE VARIANZA PARA: {variable} ~ {factor}")
    try:
        formula = f'Q("{variable}") ~ C({factor})'
        model = ols(formula, data=df).fit()
        anova_table = sm.stats.anova_lm(model, typ=2)
        print(anova_table)

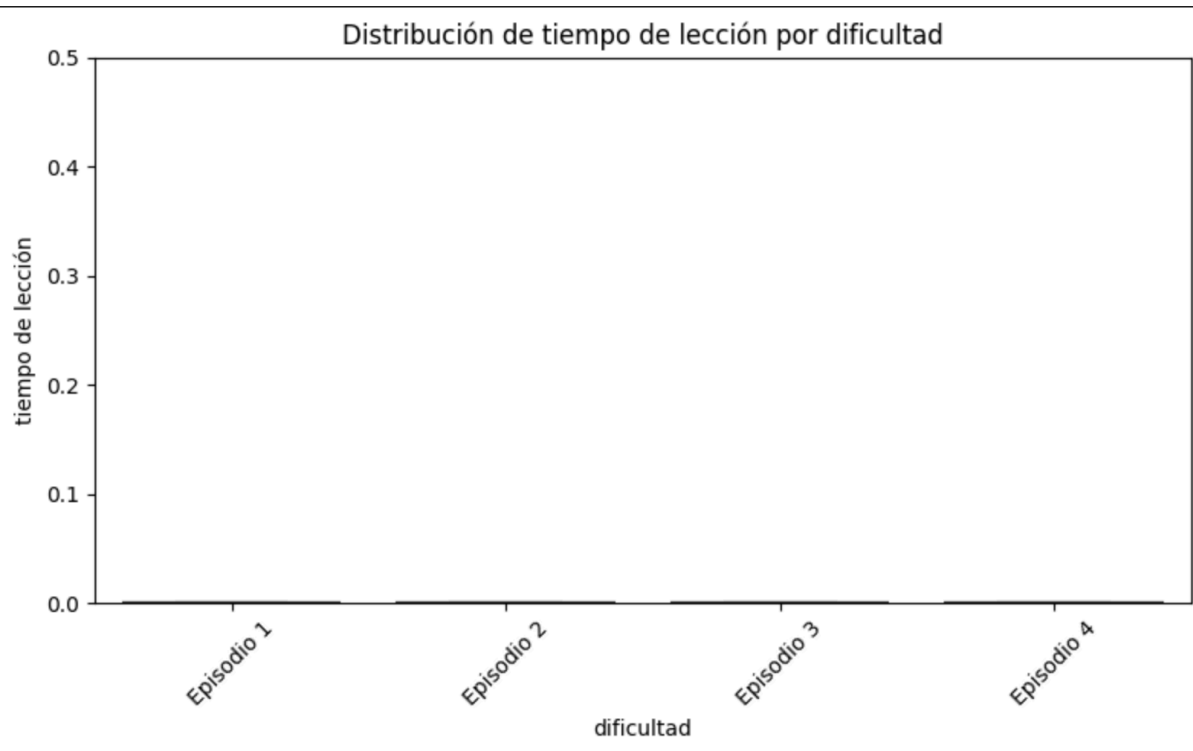
        p_val = anova_table["PR(>F)"][0]
        if p_val < 0.05:
            print(f"→ Efecto significativo de {factor} (p = {p_val:.4f})")
        else:
            print(f"→ No hay diferencias significativas para {factor} (p = {p_val:.4f})")

    except Exception as e:
        print(f"Error al analizar {factor}: {e}")

# Boxplot por Juego
plt.figure(figsize=(8, 5))
sns.boxplot(x=factor1, y=variable, data=df)
plt.title(f'Distribución de {variable} por {factor1}')
plt.xticks(rotation=45)
plt.ylim(0, 0.5)
plt.tight_layout()
plt.show()

# Boxplot por Dificultad
plt.figure(figsize=(8, 5))
sns.boxplot(x=factor2, y=variable, data=df)
plt.title(f'Distribución de {variable} por {factor2}')
plt.xticks(rotation=45)
plt.ylim(0, 0.5)
```





Después tiempo de sesión como variable y Juego y Dificultad como factores de comparación.

```
# Seleccionar variable y factores
variable = 'tiempo de sesión'
factor1 = 'Juego'
factor2 = 'dificultad'

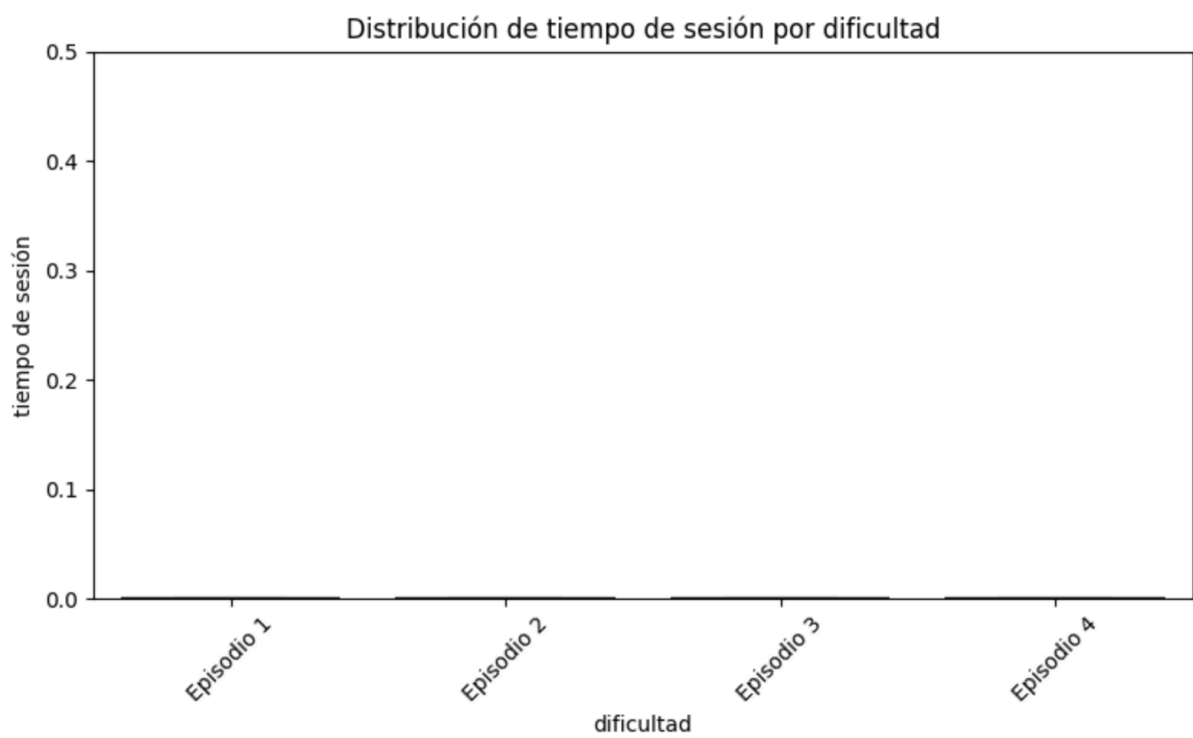
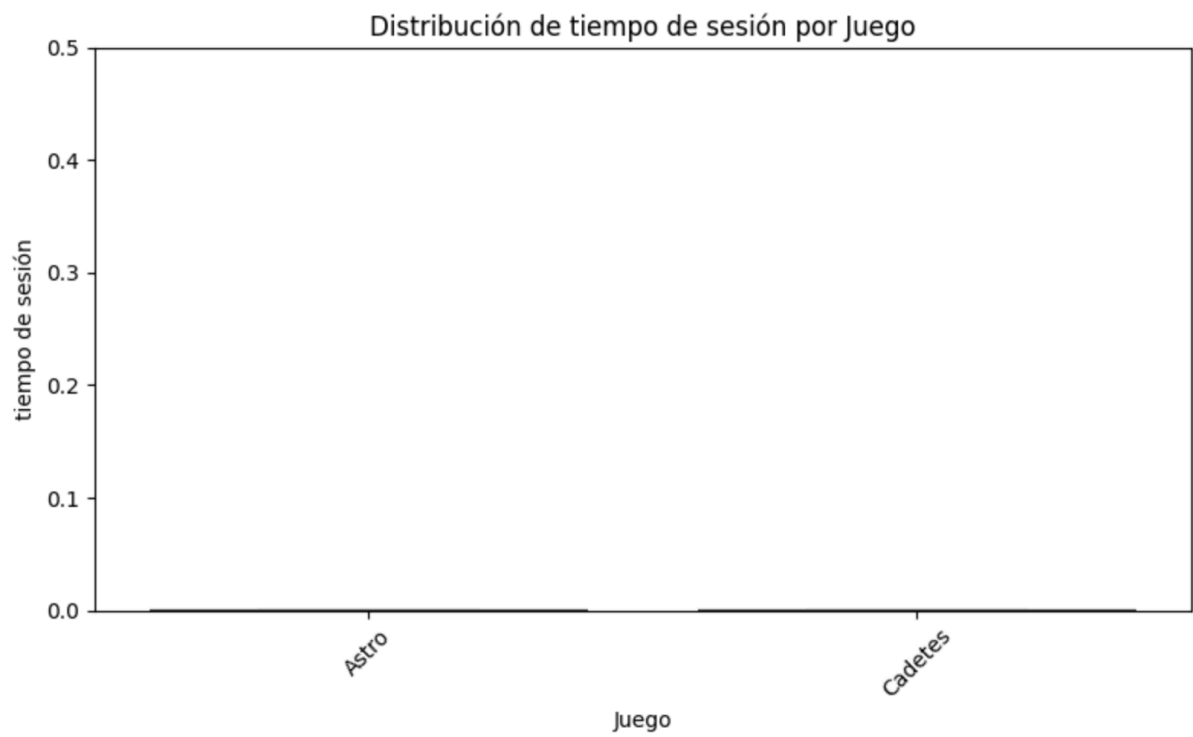
for factor in [factor1, factor2]:
    print(f"\nANÁLISIS DE VARIANZA PARA: {variable} ~ {factor}")
    try:
        formula = f'Q("{variable}") ~ C({factor})'
        model = ols(formula, data=df).fit()
        anova_table = sm.stats.anova_lm(model, typ=2)
        print(anova_table)

        p_val = anova_table["PR(>F)"][0]
        if p_val < 0.05:
            print(f"↪ Efecto significativo de {factor} (p = {p_val:.4f})")
        else:
            print(f"↪ No hay diferencias significativas para {factor} (p = {p_val:.4f})")

    except Exception as e:
        print(f"Error al analizar {factor}: {e}")

# Boxplot por Juego
plt.figure(figsize=(8, 5))
sns.boxplot(x=factor1, y=variable, data=df)
plt.title(f'Distribución de {variable} por {factor1}')
plt.xticks(rotation=45)
plt.ylim(0, 0.5)
plt.tight_layout()
plt.show()

# Boxplot por Dificultad
plt.figure(figsize=(8, 5))
sns.boxplot(x=factor2, y=variable, data=df)
plt.title(f'Distribución de {variable} por {factor2}')
plt.xticks(rotation=45)
plt.ylim(0, 0.5)
```

En cuanto a los resultados y valores de p, fueron los siguientes:

Variable	Factor	Valor p	Resultado
tiempo de interacción	Juego	3.286152e-08	Muy alejado, diferencia significativa
tiempo de interacción	Dificultad	6.377241e-61	Muy alejado, diferencia significativa
número de interacción	Juego	9.235165e-34	Muy alejado, diferencia significativa
número de interacción	Dificultad	2.006597e-48	Muy alejado, diferencia significativa
auto push	Juego	1.795800e-22	Muy alejado, diferencia significativa
auto push	Dificultad	0.032854	No tan alejado, pero menor a 0.05. Poca diferencia
tiempo de lección	Juego	4.409445e-08	Muy alejado, diferencia significativa
tiempo de lección	Dificultad	0.000095	No tan alejado, pero menor a 0.05. Poca diferencia
tiempo de sesión	Juego	0.376944	Mayor a 0.05, no hay mucha diferencia
tiempo de sesión	Dificultad	5.029391e-22	Muy alejado, diferencia significativa

Recomendaciones

Hay algunas variables como autopush, tiempo de interacción y tiempo de sesión, que tienen una gran cantidad de valores 0, lo cual dificulta su visualización en gráficas de caja; sin embargo, es posible analizar por medio del método ANOVA si son variables significativas para las conclusiones y posteriores análisis. Se podrían convertir en otras variables sintéticas o cambiar sus valores para poder manejarlos mejor con las bibliotecas de visualización de datos.

Conclusiones

El juego influye en la interacción:

Variables como "tiempo de interacción" y "auto push" mostraron diferencias significativas ($p = 0.0000$) entre juegos, indicando que el diseño del juego afecta el comportamiento del usuario.

La dificultad impacta el rendimiento:

El "número de interacción" y "tiempo de lección" variaron significativamente ($p = 0.0000$) según la dificultad, sugiriendo que retos más complejos requieren más tiempo y esfuerzo.

Tiempo de sesión depende de la dificultad, no del juego:

No hubo diferencias en "tiempo de sesión" entre juegos ($p = 0.3769$), pero sí entre niveles de dificultad ($p = 0.0000$), destacando que la complejidad es clave en la duración.