



**TECNOLÓGICO  
NACIONAL DE MÉXICO**



# **TECNOLÓGICO NACIONAL DE MÉXICO**

## **INSTITUTO TECNOLÓGICO DE MORELIA**

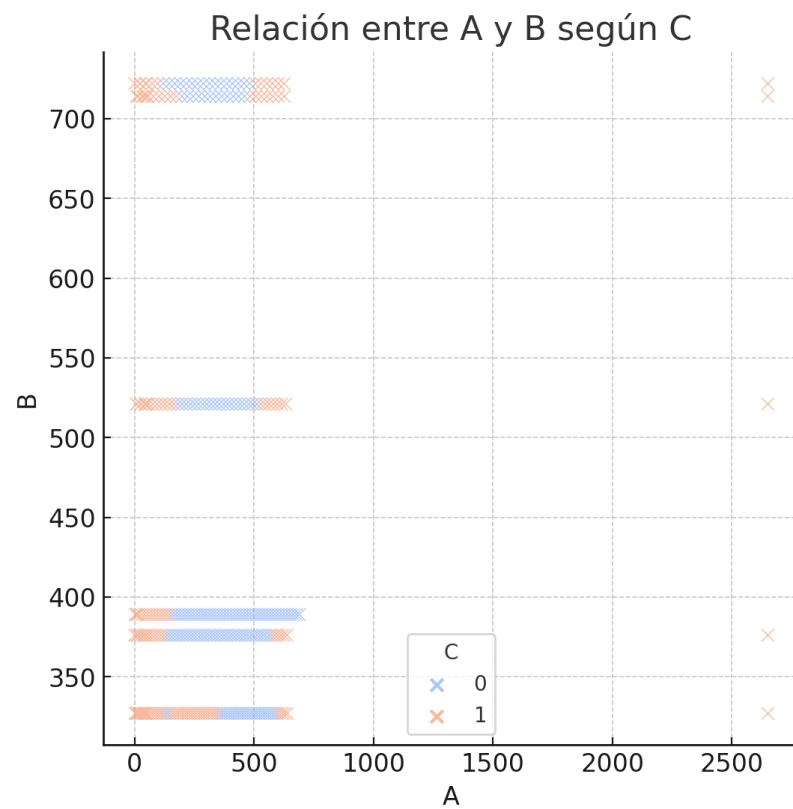
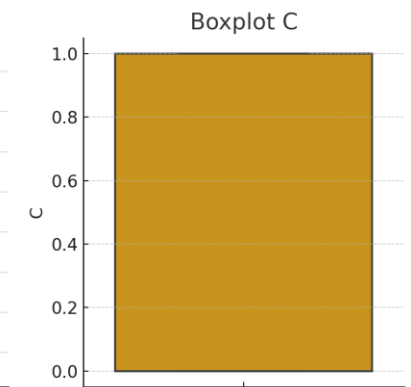
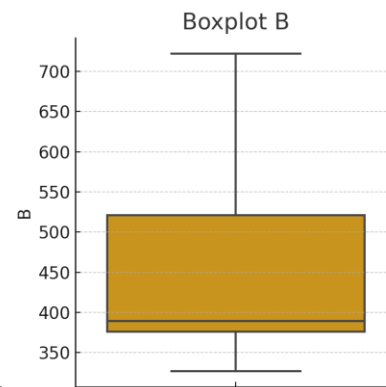
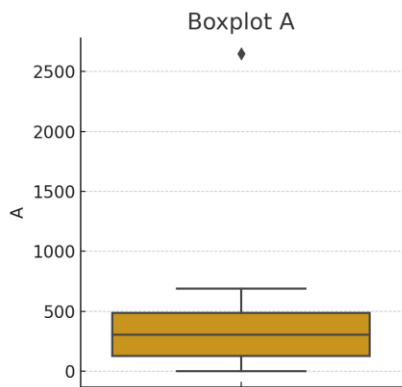
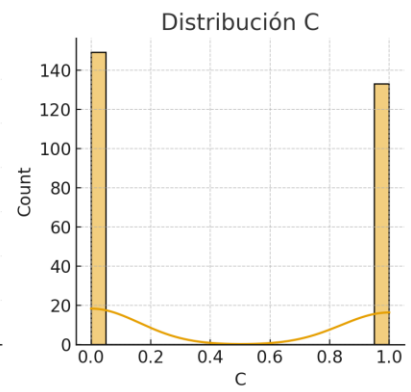
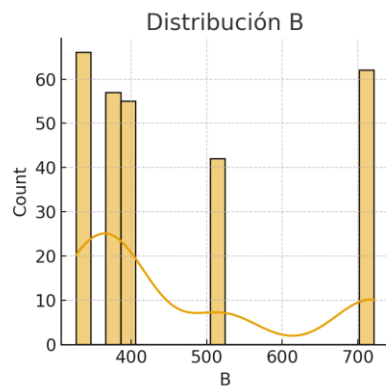
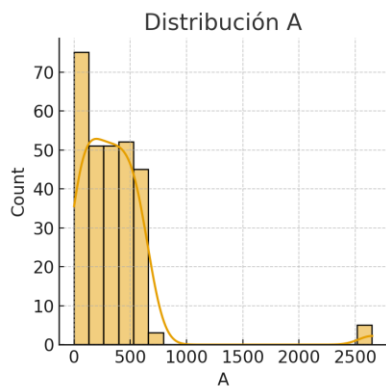
### **Inteligencia Artificial**

**Alumno:**

**Carlos Barajas Sánchez**

**Número de control: 20120083**

**Carrera: Ingeniería en Sistemas  
Computacionales**



## Estructura del dataset

- **Filas:** 282
- **Columnas:** 3 (las llamamos A, B, C porque no tenía encabezados).
- A: valores desde **1 hasta 2650**, con una gran dispersión.
- B: valores entre **327 y 722**, mucho más concentrados.
- C: solo **0 o 1**, parece una variable categórica binaria (clase o etiqueta).

## Gráficas

### 1. Histogramas

- a. A: distribución amplia con varios picos, hay valores grandes (outliers) comparados con la mayoría.
- b. B: se concentra mucho alrededor de 380–400.
- c. C: está más o menos equilibrada entre 0 y 1, aunque ligeramente más 0.

### 2. Boxplots

- a. A: tiene outliers muy altos (los que están cerca de 2650).
- b. B: más estable, sin valores extremos.
- c. C: solo 0 y 1, como esperábamos.

### 3. Diagrama de dispersión (A vs B, coloreado por C)

- a. Se ven dos grupos principales: los puntos con  $C=0$  y  $C=1$ .
- b. Parecen estar mezclados, pero con ciertas regiones donde domina un valor de C.

## Irregularidades

- **Outliers en columna A:** hay algunos valores desproporcionados (muy grandes) respecto al resto.
- **Columna B demasiado concentrada:** quizá venga de un sensor o medida fija, porque la mayoría de filas tienen valores casi iguales (389).
- **Columna C binaria:** no hay irregularidad, pero confirma que podría ser una **etiqueta de clasificación**.