**🧪 Taller de Análisis de Datos: Avance del Metro de Bogotá**

**Archivo base:** Taller\_MetroBogota.xlsx  
**Hoja:** Avances  
**Objetivo:** Aplicar habilidades de limpieza, transformación, análisis de cuartiles y detección de outliers en un conjunto de datos reales simulados.

**🔹 Parte 1: Limpieza de texto**

**Objetivo:** Estandarizar los nombres de la ciudad para facilitar el análisis.

**Instrucciones:**

1. Convierte todo el texto de la columna "Ciudad" a **mayúsculas** usando .str.upper().
2. Reemplaza todas las variantes para que se unifiquen como "BOGOTÁ".

**Pista en código:**

df["Ciudad"] = df["Ciudad"].str.upper()

df["Ciudad"] = df["Ciudad"].replace({"BOGOTA": "BOGOTÁ"})

**🔹 Parte 2: Análisis de cuartiles sobre AvanceObra (%)**

**Objetivo:** Calcular Q1, Q2 (mediana), Q3 e IQR para interpretar cómo avanzan las obras.

**Instrucciones:**

1. Calcula los **cuartiles** para la columna "AvanceObra (%)".
2. Interpreta en qué rango están el 25%, 50% y 75% de las estaciones.
3. Calcula el **IQR** y los límites inferior/superior.

**Pista en código:**

q1 = df["AvanceObra (%)"].quantile(0.25)

q2 = df["AvanceObra (%)"].quantile(0.5)

q3 = df["AvanceObra (%)"].quantile(0.75)

iqr = q3 - q1

**🔹 Parte 3: Detección de outliers en Presupuesto (millones)**

**Objetivo:** Identificar obras con presupuestos atípicos.

**Instrucciones:**

1. Usa el IQR de la columna "Presupuesto (millones)" para detectar outliers.
2. Filtra las estaciones que **superan o están por debajo** de los límites.

**Pista en código:**

q1\_p = df["Presupuesto (millones)"].quantile(0.25)

q3\_p = df["Presupuesto (millones)"].quantile(0.75)

iqr\_p = q3\_p - q1\_p

lim\_inf = q1\_p - 1.5 \* iqr\_p

lim\_sup = q3\_p + 1.5 \* iqr\_p

outliers = df[(df["Presupuesto (millones)"] < lim\_inf) | (df["Presupuesto (millones)"] > lim\_sup)]

**🔹 Parte 4: Agrupación**

**Objetivo:** Analizar el avance promedio y presupuesto total por ciudad.

**Instrucciones:**

1. Agrupa por "Ciudad" y calcula:
   * Promedio de "AvanceObra (%)"
   * Suma total de "Presupuesto (millones)"

**Pista en código:**

agrupado = df.groupby("Ciudad").agg({

"AvanceObra (%)": "mean",

"Presupuesto (millones)": "sum"

}).reset\_index()

**📊 Opcional (extra): Visualización**

* Haz un **boxplot** del avance de obras.
* Crea un gráfico de **barras por ciudad** del presupuesto total.