



UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES

FÍSICA COMPUTACIONAL

# Actividad 4: Análisis Exploratorio de Datos en Python.

Grupo 2.

María José Ramos Murillo

---

## Introducción

En esta actividad hemos retomado el archivo de datos meteorológicos que leímos en la Actividad 3, usando biblioteca Pandas, y realizamos un Análisis Exploratorio de Datos (EDA, Exploratory Data Analysis).

Algunas de las funciones que utilizamos anteriormente nos dieron información sobre la estructura y patrones de los datos, el número de datos faltantes, datos anómalos y valores extremos, extracción de variables importantes y nos permitieron probar alguna teoría de comportamiento de datos.

Encontramos que la descripción estadística es la siguiente:

<b>count</b>	19037.00	16923.00	19030.00	19004.00
<b>mean</b>	0.90	6.00	32.72	18.35
<b>std</b>	6.15	2.53	5.14	5.68
<b>min</b>	0.00	0.00	13.50	-1.50
<b>25%</b>	0.00	4.00	29.00	13.50
<b>50%</b>	0.00	6.00	33.50	18.00
<b>75%</b>	0.00	7.90	37.00	24.00
<b>max</b>	211.00	18.00	48.00	31.00

Figura 1: Descripción Estadística de Los Mochis, Sinaloa.

Así, lo que exploramos visualmente fueron estos resultados y obtuvimos más información.

La biblioteca que usamos principalmente fue el matplotlib y seaborn, pues desplegamos las gráficas que muestran los datos de temperaturas, evaporación y precipitación, según las indicaciones, ya se por meses o años.

Realizamos histogramas, gráficas de caja (BoxPlots), y distribuciones de probabilidad.

Analizamos el número de días en que no llovió (precipitación 0) mediante histogramas, y como el número de días no permite ver el comportamiento de los días, lo que hubo que hacer fue cambiar la escala de la gráfica mediante la escala logarítmica, de esta manera pudimos hacer un análisis más detallado de los días. Después obtuvimos la gráfica de la distribución y las comparamos.

Repetimos lo mismo para las Temperaturas Máxima y Mínima, y la evaporación, y

---

probamos el resto del tipo de gráficas.

## **Primeras Impresiones**

Esta actividad me pareció muy útil y entretenida, pues pudimos visualizar las gráficas y pudimos entender mejor lo que sucede en el dataframe. El nivel me pareció intermedio, pues no fue muy complicado comprender el código, pero el contenido que estamos realizando sí me parece que es avanzado y que es muy útil tanto para otras materias como en el área laboral. Me pareció divertido aprender esta información, y me sirvió mucho para repasar los conocimientos de Estadística