

Actividad 4

Moreno Murrieta Yazmin Yamileth
Departamento de Física
Universidad de Sonora

05 Febrero 2021

1 Introducción

En la actividad 3 estuvimos explorando inicialmente algunas funciones de Pandas para guiarnos en el proceso de limpieza y preparación de los datos con DataFrame.

Muchas de las funciones que hemos utilizado son parte del proceso de Análisis Exploratorio de Datos, donde buscamos conocer la estructura, contenidos y características del conjunto de datos que se desean analizar.

Inicialmente se realizó la exploración general de DataFrame, ya que se necesitaba tener exportado el archivo a utilizar.

Una vez empezada la etapa de exploración, continuamos con un proceso de visualización de datos para mejorar la percepción de los fenómenos o eventos que representan los datos.

En esta actividad se utilizaron dos bibliotecas: Matplotlib y Seaborn.

1. Matplotlib: es la biblioteca de Python desarrollada para la visualización de datos. Es muy potente y permite producir gráficas.
2. Seaborn: se construyó sobre Matplotlib, pero con una estructura menos compleja, de mayor facilidad de manejo.

Así mismo, esta actividad consistió en 6 pasos.

1. En el primer paso se pidió sintetizar las características principales del conjunto de datos analizados.
2. En el paso dos, se crearon histogramas sobre la precipitación, evaporación y de las temperaturas (máxima y mínima).
3. En el paso tres, se pide crear gráficas de cajas para la evaporación y temperaturas.
4. Para el paso cuatro, se debía producir gráficas de barras para la precipitación agrupada por años y meses.

5. En el paso cinco, se solicitó crear una colección de los últimos 30 años de datos para la precipitación y temperaturas.
6. Y por último, se produjeron diagramas de cajas para observar la variabilidad de las temperaturas y la evaporación, agrupados por meses.

2 Comentarios

Me puse a investigar sobre que es el EDA, ya que nunca antes lo había escuchado y menos utilizado, y me pareció una herramienta muy útil, ya que nos ayuda a vislumbrar posibles patrones o a reconocer distribuciones estadísticas que puedan ser útiles.

3 Primeras Impresiones

Me pareció un tema muy interesante, ya que nunca antes lo había escuchado, es una herramienta demasiado útil pues nos permite analizar datos de manera mas detallada y a nuestra conveniencia.

En lo personal, se me hizo una actividad bastante larga y un poco enredosa, pero, una vez teniendo en claro las ideas resulta más sencillo de lo que parece. Por lo tanto, le asigno un grado de complejidad Avanzado, pues es una actividad que requiere de tiempo y bastante entendimiento.