

Actividad 4

Análisis Exploratorio de Datos en Python

Elizabeth Torres Torrecillas
Departamento de Física
Universidad de Sonora

February 5, 2021

1 Introducción

La actividad 4 tiene como fin el iniciar formalmente con los primeros pasos que se realizan en el Análisis Exploratorio de Datos. Esto a partir de datos climatológicos de la estación meteorológica de El Fresnal obtenidos por medio de la página web gubernamental de la CONAGUA

Una vez que culminamos los ejercicios propuestos en la actividad 3, en la cual tuvimos una primera interacción con la lectura de datos apropiadamente. Ahora, procedemos a continuar mediante un acercamiento más visual, siendo esto mediante representaciones gráficas para mejorar la percepción de los fenómenos o eventos que representan los datos.

2 Bibliotecas utilizadas

El análisis exploratorio de datos se realizó con ayuda de principalmente, dos bibliotecas que fueron base para lograr un acercamiento más visual con los fenómenos o eventos presentes en los mismos.

- Matplotlib
Biblioteca de Python desarrollada para la visualización de datos. Es muy potente y permite producir gráficas que deseemos.
Los tipos de gráficos más comunes con Matplotlib son:

1. Diagramas de barras
2. Histograma
3. Diagramas de sectores
4. Diagramas de caja y bigotes
5. Diagramas de violín

6. Diagramas de dispersión o puntos
 7. Diagramas de líneas
 8. Diagramas de áreas
 9. Diagramas de contorno
 10. Mapas de color
- Seaborn
Biblioteca de Python, permite generar fácilmente gráficos que son visualmente sencillos y elegantes. Seaborn se construyó sobre Matplotlib, pero con una estructura menos compleja, de mayor facilidad de manejo. Por lo que es capaz de proporcionar un interfaz de alto nivel que es realmente sencilla de aprender.

3 Actividades realizadas

Todo lo que se desarrolló en la presente actividad fue con base en lo que anteriormente se había realizado en la actividad 1 y 3, ya que en cada una de estas fuimos poco a poco interactuando con los datos meteorológicos de El Fresnal. Entonces, antes de continuar, sintetizamos las características principales del conjunto de datos que utilizamos en la actividad anterior. Así, resumiendo el DataFrame inicial y crear uno nuevo. Después con ello se crean histogramas para las variables de los datos, siendo "Precipitación, Evaporación, Temperaturas Máximas y Mínimas", como también las gráficas de la función de densidad de probabilidad

Continuamos creando las representaciones gráficas de cajas para las mismas variables mencionadas anteriormente. Para después realizar gráficas de barras para la variable "Precipitación", una que sea agrupada por años y otra por meses.

Se agrupa una colección de datos de los últimos 30 años, para con ello hacer gráficas de línea de las variables "Precipitación, Temperatura Máxima y Temperatura Mínima" en función del tiempo. Utilizamos la agrupación de los últimos 30 años pero agrupados por mes, realizada anteriormente para producir diagramas de cajas para observar variabilidad de las Temperaturas Máximas, Mínimas y Evaporación.

3.1 Observaciones

Podemos observar que la temperatura máxima comunmente ronda entre los 23°C y 33°C, como también que las temperatura mínima es común que se encuentre entre los 5°C y los 17°C.

Como también, por medio del análisis de datos podemos observar que aunque se espere que llueva durante todos los meses, especialmente julio, agosto y septiembre. Pero cuando esto ocurre es con poca cantidad, por lo que se espera como máximo una precipitación de alrededor de 2mm.

Además podemos observar una relación entre los meses donde se presenta mayormente la precipitación, es cuando se encuentran las temperaturas mayores, siendo esto con una máxima de 30°C a 40°C.

Algo que creo importante resaltar es que no se registraron datos sobre la evaporación de El Fresnal, por lo que busqué estaciones que fueran cercanas como Imuris, sin embargo esta no mantenía un registro continuo, los datos estaban tomados de 1 cada periodo significativo o si quiera registraban por un largo tiempo. Por lo que los datos tienen una gran incertidumbre. Es importante resaltar que no se presentan datos de la evaporación de El Fresnal porque los gráficos de este aparecen en blanco, debido al registro nulo.

4 Comentarios personales

La actividad realizada me pareció interesante, ya que exploramos con muchos y distintos tipos de gráficos que son de bastante utilidad en el día a día, más cuando se desea realizar un análisis estadístico. La manera de realizarlo e interaccionar con el mismo me pareció sencillo y el aspecto visualmente agradable. Creo que fue sencillo manipular los gráficos, junto con los elementos que los componen. Así, podemos tener un acercamiento hacia la exploración de los datos. En si, no me pareció del todo complicada, sin embargo me pareció bastante repetitiva y siento que me falta bastante aún como para que estos se vean más limpios y ordenados, supongo que con la práctica iré mejorando poco a poco. Me pareció bien, ya que conocí más bibliotecas con bastante utilidad y más porque sigo aprendiendo un poco más de Python, como utilizarlo para aprovechar al máximo sus herramientas. Sin embargo, no me fue tan divertido el realizarla, siendo que solo fue utilizar distintas funciones, aún así le asigno un nivel intermedio. Me hubiese gustado más ya tener algo aplicado, menor largo el trabajo o tal vez sentir que se realiza cierta gráfica para mostrar algo en específico. Aunque entiendo que la razón de ello es porque vamos empezando y tenemos que conocer las herramientas que tienen estos. La carga de trabajo esta semana siento que fue un poco más cargada al final, aunque avancé día con día, no fue distribuida la carga de trabajo de manera equitativa. Empero, es importante mencionar que el lunes fue asueto y también por eso se juntó la carga de trabajo.