

Actividad 5

Meza Ruiz Humberto Eduardo
218208583

Febrero 2021

Introducción.

Durante esta actividad estaremos revisando las Series de tiempo utilizando la paquetería Pandas. Asimismo, vamos a agregar dos paqueterías extras a las que habíamos estado usando hasta la actividad pasada; *seaborn* y *statmodels.api*.

El objetivo de esta actividad será, utilizando los mismos datos climatológicos que hemos usado hasta ahora, explorar un análisis de datos en series de tiempo, esto para poder visualizar mejor la evolución de los datos a través del transcurso del tiempo en el cual se han obtenido.

Actividad.

En primera instancia nos dedicamos a corroborar que nuestros datos funcionan para realizar el análisis, esto se logra simplemente utilizando la función **df.plot()** la cual nos permitirá graficar los datos para visualizar si tenemos una continuidad de datos o no.

En caso de que nuestros datos sean continuos o al menos lo sea, en el periodo de los últimos 30 años analizaremos esos años y, para contrastar analizaremos un año en específico, para lo cual vamos a usar la función **df.loc()**

Haremos esto para la precipitación, para las temperaturas máximas y mínimas, y además, vamos a utilizar dos maneras de graficar la estacionalidad de los datos para la evaporación y temperaturas, utilizando las funciones **sns.barplot()** y **sns.boxplot** respectivamente para cada conjunto de datos.

Parte de la actividad pide hacer lo mismo para la evaporación pero, en nuestro caso, los datos de evaporación son demasiado pocos, y además, discontinuos, por lo que algunos comandos ni siquiera funcionan, aún así, lo logré graficar en algunas partes, quedando en evidencia la insuficiencia de datos de evaporación.

Ahora, una forma de analizar datos para comparar cambios en los datos es utilizar promedios móviles, para lo cual, usaremos gráficas que harán una comparación de la evolución de datos con sus promedios móviles de 7, 30 y de 365 días. Para ello, usaremos una función de Python, de la paquetería pandas llamada **df.rolling()**.

Y, como objetivo principal de esta actividad, realizaremos una serie de tiempo de los datos. Descompondremos los datos de tal forma que, cumpla que Serie de tiempo = Tendencia + Estacionalidad + Residuo. Para lograrlo usaremos la función **seasonal.descompose** de la paquetería *statsmodels.tsa*.

Por último, vamos a usar un rango de datos anterior a los últimos 30 años que hemos estado usando, para comparar la evolución en ambos casos. Como mis datos empiezan el 84, este rango será más corto, aún así, se logra ver una evolución bastante similar. Para ello, usaremos dos gráficas contiguas de histogramas y función de densidad usando la función **df.displot()**.

Conclusión.

Durante esta actividad se tomaron muchas ideas que están incluso fuera de dicha actividad las cuales dejaron en evidencia la capacidad de estas librerías para realizar análisis de datos. Estas herramientas son realmente útiles si se disponen de muchos datos y se quiere una eficiencia mayor en el análisis de estos, puede ser bastante complicado al principio pero la verdad es que están muy bien estructuradas y son, aunque no intuitivas, fáciles de dominar. Además, el desempeño del profesor al momento de exponer las librerías deja bastante en claro cómo usarlas para cada caso y cómo realizar caminos alternativos para visualizar lo mismo de distintas maneras.

En el siguiente enlace se encuentra el código del programa en Python resultante de esta actividad:
<https://github.com/HumbertoMezaRuiz/FisicaComputacional1/blob/main/Actividad5/Actividad5.ipynb>