

Actividad 4: Análisis exploratorio de datos en Python

Antonio Reyes Montaña

03/03/2021

1. Introducción

En esta actividad se hizo un análisis de los datos recopilados por la estación meteorológica de La Paz y se hicieron gráficas de diferentes tipos para la evaporación, la precipitación, y las temperaturas máximas y mínimas utilizando las funciones de la librería Seaborn para Python.

Primero preparamos el archivo como lo hicimos en la actividad pasada en una sola celda para trabajar mejor con él. Al realizar las gráficas pudimos observar cuáles son los meses de mayor precipitación, los intervalos de temperaturas máximas y mínimas y también pudimos analizar la frecuencia con la cual ciertos valores aparecen en la base de datos para las propiedades antes mencionadas. También se revisó con los boxplots la distribución de datos y cómo podrían haber ciertos datos irregulares que es mejor no tomar en cuenta. Pudimos estudiar las tendencias presentadas para las diferentes propiedades a través de los años durante los cuáles la estación meteorológica ha estado activa, y también los valores promedio y barras de error por mes. Finalmente se estudiaron también las propiedades de los datos de los últimos 30 años.

2. Observaciones del análisis de datos

2.1. Histogramas

Del histograma realizado para la distribución de precipitación no pudimos extraer mucha información hasta que lo cambiamos a una escala logarítmica debido a que la mayoría de los datos estaban en el cero, y los demás datos en valores diferentes tenían una frecuencia muy baja. Esto indica claramente que La Paz es un lugar de muy poca precipitación y lo ha sido así por mucho tiempo. Por otro lado se registran algunas ocasiones en las cuales hubieron incluso hasta más de 250mm de precipitación.

Para el caso de la evaporación pudimos apreciar que varía bastante pero por lo general la evaporación tiende a ser en su mayoría de alrededor de 5mm a 6mm.

Para la temperatura máxima observamos que su promedio está un poco por encima de los 30 grados y la mediana un poco más arriba. Hay temperaturas máximas de hasta más de 40 grados aunque no con una frecuencia tan alta. Los valores más frecuentes están entre los 30 y los 40 grados (principalmente alrededor de los 35 grados), lo cual nos indica que La Paz tiende a ser un lugar caluroso.

Para las temperaturas mínimas observamos que el promedio está cerca de los 17 grados, y la mediana un poco por debajo del promedio. Las temperaturas mínimas más frecuentes están alrededor de los 15 grados, pero también hay otro pico cerca de los 23 grados. Las temperaturas mínimas van desde temperaturas menores a los 5 grados (las cuales no son muy frecuentes), hasta cerca de los 30 grados (que tampoco son muy frecuentes).

2.2. Diagramas de caja

Del diagrama de caja para la evaporación observamos que la mayoría de los datos están entre más o menos los 4mm y los 8mm de evaporación. Los demás datos no son tan frecuentes y por encima de alrededor de los 13mm los datos son despreciables para hacer un análisis estadístico.

Para la temperatura máxima observamos que la mayoría de los datos se encuentran entre alrededor de los 26 grados a los 35 grados, los demás datos son menos frecuentes pero existen datos que van desde los 15 grados hasta los casi 45 grados.

Para la temperatura mínima podemos ver que la mayoría de los datos están entre alrededor de los 13 grados hasta alrededor de los 22 grados. Los demás datos son menos frecuentes pero los datos van desde temperaturas cercanas a los 0 grados hasta temperaturas un poco mayores de los 30 grados.

2.3. Gráficas de barras por años y por meses

Revisando la precipitación promedio por año observamos algo muy peculiar, ya que podemos ver que en el año 2017 las precipitaciones son inusuales y mucho mayores que cualquier registro anterior.

Al observar la misma gráfica para los últimos 30 años, podemos ver que los promedios de precipitación anual son muy bajos, y como ya se mencionó antes, en el año 2017 el promedio aumentó drásticamente, ya que prácticamente se triplicó en comparación a los valores más altos registrados durante ese periodo.

Revisando la precipitación promedio podemos observar que los meses con más lluvia en el periodo de estudio de la estación, han sido agosto y septiembre, aunque parece que la época de lluvias empieza desde julio y cesa en octubre. Durante el invierno también se presentan algunas lluvias durante diciembre y enero pero con una frecuencia mucho menor, mientras que de marzo a junio es época sin lluvias. Si en cambio observamos el comportamiento de estos datos de precipitación mensual promedio de los últimos 30 años, podemos afirmar prácticamente lo mismo que se mencionó anteriormente

2.4. Gráficas de línea de los últimos 30 años

De la gráfica de línea de la precipitación promedio podemos concluir algo similar a lo que ya habíamos dicho de las gráficas de barras, que el promedio aumento considerablemente en el último año

De la gráfica de línea de temperaturas, podemos ver que ha habido una tendencia a través de los años al aumento de las temperaturas.

2.5. Diagramas de caja de los últimos 30 años

En la variabilidad de la temperatura máxima podemos ver que en los meses de primavera es donde existe una variabilidad mayor, por ejemplo, en marzo y abril, pero en los meses más calurosos las variabilidades son más pequeñas, aunque es donde existen más datos que se deshechan por estar muy alejados. También vemos que todos estos valores están alrededor de la temperatura promedio calculada.

Para la temperatura mínima vemos de nuevo que la menor variabilidad está en meses de verano como agosto, mientras que en meses más fríos como diciembre o enero, la variabilidad es más grande.

Para la variabilidad de la evaporación observamos que los meses de menor evaporación son los meses fríos como es de esperarse, y que la evaporación mayor es en los meses de junio y julio.

3. Comentarios generales de la información analizada

De todos los datos que se pudieron analizar, podemos fácilmente decir que La Paz es un lugar caluroso con épocas de lluvias en los veranos, y algunas lluvias en el invierno. También vimos que ha habido una tendencia a través de los años para el aumento de las temperaturas mínimas y máximas, lo cual podría ser fácilmente una repercusión del cambio climático.

4. Comentarios sobre la actividad realizada

La actividad fue un buen ejercicio para entender mejor el funcionamiento de la biblioteca de Seaborn que incluye muy buenas herramientas para graficar datos en diferentes formas. Además fue útil para desarrollar mejor la habilidad de interpretar la información que podemos extraer de las gráficas obtenidas. Por otro lado, el hacer gráficas de tantos tipos pudo ser un poco monótono.