

Reporte sobre la Actividad 2

García Parra Pedro

Febrero 2019

En esta actividad aprendimos a utilizar la biblioteca pandas, la cual es una biblioteca para python que te permite realizar un analisis sobre una serie de datos. También se utilizó la biblioteca matplotlib, más específicamente pyplot para ayudarnos en la graficación de algunos datos.

Para la realización de esta actividad se respondieron las siguientes preguntas:

- ¿Cómo le podrás determinar cuáles son los meses más lluviosos?
- ¿Cuáles son los meses más fríos y cuáles son los más cálidos?
- ¿Cuáles han sido años muy húmedos?
- ¿Cuáles han sido años muy secos?
- ¿Cuáles años han tenido inviernos fríos?
- ¿Cuáles años han tenido veranos más cálidos?
- ¿Cómo ha venido siendo la temperatura mensual promedio en los últimos 20 años?
- ¿Qué ha pasado con la precipitación en los últimos 20 años de datos?

Para responder las primeras cuatro preguntas usé la siguiente función de pandas:

```
df.groupby(args)
```

FECHA	
June	0.393157
December	0.411810
January	0.457256
March	0.554882
November	0.562637
April	0.576316
May	0.670846
October	0.703374
February	0.738235
September	1.923795
August	2.229911
July	2.268429
Name: PRECIP, dtype: float64	

Figura 1: Output para la resolución de la primer pregunta

donde df es un objeto DataFrame y los argumentos que se le pasan es en base a qué se va a agrupar.

En particular para responder la primer pregunta se utilizó la siguiente única línea de código:

```
df.groupby(df['FECHA'].dt.strftime('%B'))['PRECIP'].mean().sort_values()
```

El resultado de ejecutar esa línea es la tabla que se muestra en la figura (1) y con ella podemos ver cual fue el mes en que hubo menos precipitación y el cual hubo más. Las tres preguntas siguientes son similares.

Para la pregunta cinco primero tuve que obtener la temperatura mínima solamente de los meses de invierno (Diciembre, Enero, Febrero y Marzo) para eso primero obtuve los índices en mi DataFrame en los cuales era invierno, luego cree un array donde guardé las temperaturas de estos días y por ultimo utilicé df.groupby() para ordenarlos de acuerdo al año; el código que utilicé fue el siguiente:

```
## Obtiene el índice de los datos que se encuentran en invierno
##(mes entre diciembre y marzo)
Invi = []
j = 0
```

```

for i in range(0,7790):
    if (df["MES"][i] == 12 or df["MES"][i] == 1 or df["MES"][i] == 2
        or df["MES"][i] == 3):
        Invi.append(i)

## Crea el array con los años y temperaturas
años = []
temp = []
inv = []
for i in Invi:
    años.append(df["AÑO"][i])
for i in Invi:
    temp.append(df["TMIN"][i])
inv = [años,temp]
inviernos = pd.DataFrame({
    'AÑOS': inv[0],
    'TMIN': inv[1]},
    index = range(0,len(años)))
## Muestra los valores de la media de la temperatura
inviernos.groupby("AÑOS")['TMIN'].mean().sort_values()

```

El resultado de ejecutar este código es similar a la figura(1) y se puede responder la pregunta de manera muy sencilla. La pregunta seis fue realizada de la misma manera.

Para realizar las preguntas 7 y 8 cree un dataframe nuevo donde se omitieran los datos de 1990 y 1991 para así tener solamente los últimos 20. Tras realizar eso solamente se volvió a utilizar la función `groupby()` y encontrar la respuesta de la pregunta fue muy sencillo.