

Reporte de la actividad 4

Daniela Olmos Velderrain

Grupo 3

17 de febrero de 2019

1. Introducción

En la actividad pasada se aplicó el uso de la biblioteca Pandas para análisis de datos sobre el registro meteorológico de Cajeme en las últimas décadas. En esta actividad se continuará trabajando con los mismos datos de precipitación, temperaturas máximas y mínimas, pero ahora utilizaremos una nueva herramienta para visualizar los datos: Matplotlib.

Mediante Matplotlib se elaborarán diversas gráficas para representar los datos de temperaturas y precipitación obtenidos mensual y anualmente.

2. Desarrollo

2.1. Metodología

Se importaron las siguientes librerías en el archivo de Jupyter Notebook para trabajar con los datos de Cajeme:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import calendar
import seaborn as sns
```

Estas nos ayudarán con los cálculos matemáticos, el análisis estadístico y la realización de gráficas.

Para evitar los valores nulos, se les asignó a estos una variable:

```
sentinels = {'PRECIP': ['Nulo'], 'EVAP': ['Nulo'], 'TMAX': ['Nulo'], 'TMIN': ['Nulo']}
```

Después se leyó el archivo mediante el comando "pd.read_csv" (indicando con na_values=sentinels que los valores en dicha variable son datos nulos), y posteriormente se creó un Data Frame mediante el comando "pd.DataFrame()".

Como el archivo cuenta con una columna de fechas, a esta se le dio formato de fecha mediante:

```
df['FECHAN'] = pd.to_datetime(df.apply(lambda x: x['FECHA'], 1), dayfirst=True)
```

```
df = df.drop(['FECHA'], 1)
```

Para poder definir columnas para meses y años, estos se extrajeron de la nueva variable "FECHAN".

```
df['MES'] = df['FECHAN'].dt.month
df['AÑO'] = df['FECHAN'].dt.year
```

Para averiguar el número de años que hay en el archivo, se contaron los datos diferentes de esta columna, obteniendo 32 años, los cuales se guardaron en la variable "NumA":

```
NumA = len(df['AÑO'].unique())
```

Para tener arreglos con años y meses se llenaron dos vectores mediante loops, el de los años de 0 a NumA, y el de los meses de 0 a 12. Para tener los nombres de los meses, también se creó un arreglo, utilizando la biblioteca calendar":

```
init = 1980
AÑOS = [init + i for i in range(0, NumA)]
init2 = 1
MESES = [init2 + i for i in range(0, 12)]
MESESlabel = calendar.month_name[1:13]
```

Estos arreglos servirán de etiquetas en el eje x más adelante.

Ahora, para calcular el promedio mensual de temperaturas y precipitaciones se crearon tres arreglos distintos, los cuales extraen información del DataFrame:

```
PRECIPPROPMES = [df[df.MES==(init2 + i)].PRECIP.sum()/NumA for i in range (0,12)]
TMAXPROMMES = [df[df.MES==(init2 + i)].TMAX.mean() for i in range (0,12)]
TMINPROMMES = [df[df.MES==(init2 + i)].TMIN.mean() for i in range (0,12)]
```

Este mismo procedimiento se siguió para los datos anuales, pero ahora con la columna 'AÑOS'.

Los arreglos realizados se utilizaron para hacer las gráficas de barras y de evolución.

Las gráficas de barras se realizaron de la siguiente manera:

```
X = MESESlabel
N = np.arange(len(X))
Y = PRECIPPROPMES

plt.bar(N, Y, width=0.5, align='center', alpha=0.7, color='b')
plt.xticks(N, X, size = 'small', color = 'k', rotation = 90)
```

Se llevó a cabo el mismo procedimiento para la gráfica de años.

Para la gráfica de evolución se realizaron dos gráficas en la misma figura, y así poder comparar la temperatura mínima y máxima a través del tiempo.

```
X = AÑOS
N = np.arange(len(X))
Y1 = TMINPROMAÑO
Y2 = TMAXPROMAÑO

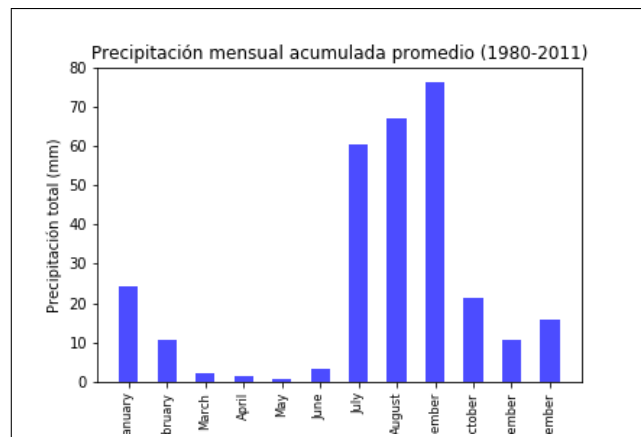
plt.plot(Y1, label = "Temperatura mínima", color = 'y')
plt.plot(Y2, label = "Temperatura máxima", color = 'b')
plt.xticks(N, X, size = 'small', color = 'k', rotation = 90)
```

Para hacer los gráficos de caja se recurrió a la biblioteca Seaborn, ya que no fue posible realizarla mediante Matplotlib. Se recurrió a los siguientes comandos para hacerlo a partir del DataFrame:

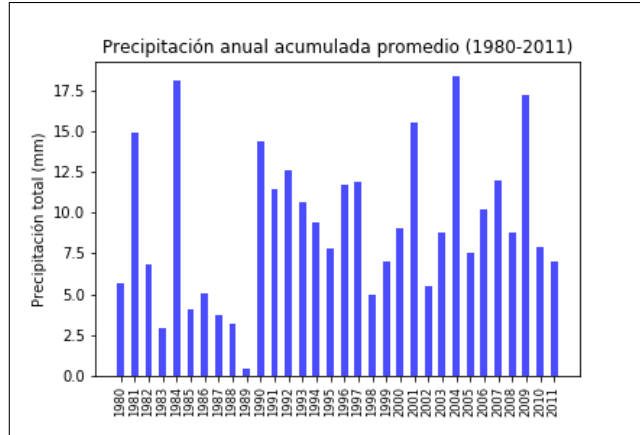
```
box = sns.boxplot(x="MES", y="TMAX", data=df)
box.set( xlabel= 'Meses', ylabel= 'Temperatura máxima (C)')
box.set_xticklabels(MESESlabel, rotation=90 )
```

2.2. Resultados

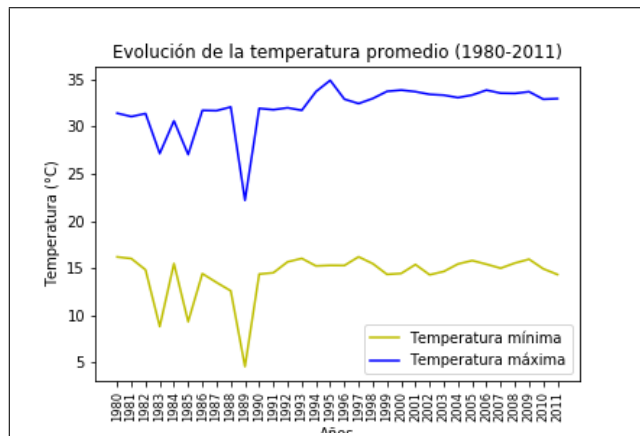
Las gráficas resultantes fueron las siguientes:



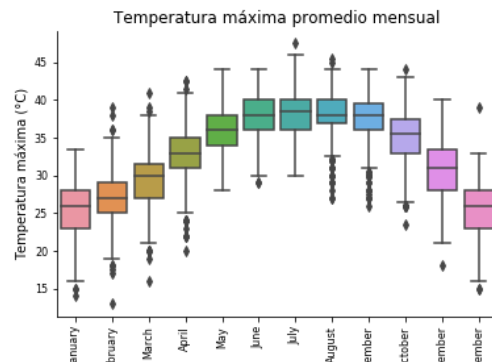
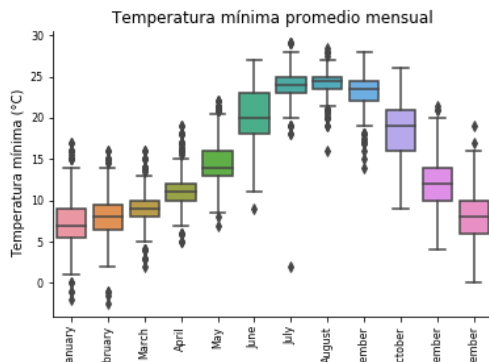
De acuerdo a la gráfica, la mayoría de las precipitaciones suelen encontrarse en los meses de julio, agosto y septiembre.



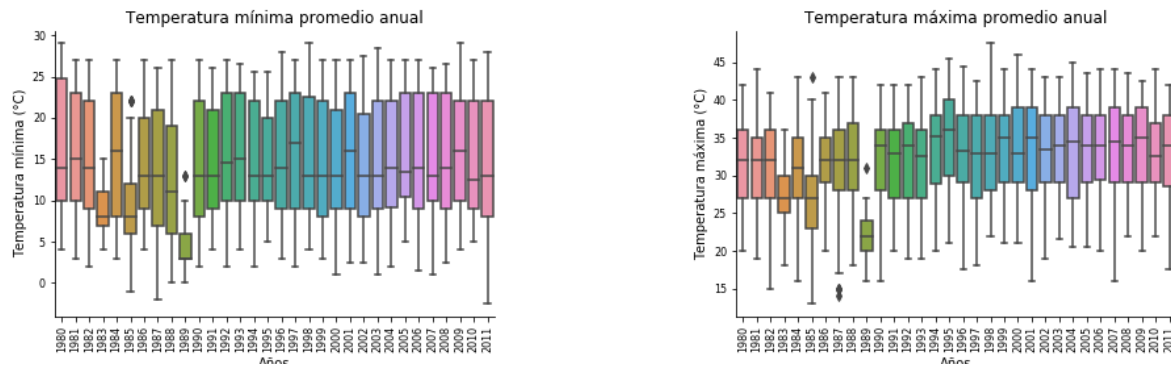
En esta gráfica se puede observar la precipitación promedio por año, destacando que las lluvias han aumentado en las últimas décadas.



Mediante esta gráfica podemos observar el progreso de las temperaturas a lo largo del año. Como se puede ver, estas avanzan de manera equidistantes entre ellas.



En estas gráficas se puede observar cómo las temperaturas mínimas y máximas parecen seguir el mismo patrón, siendo estas mayores en los meses de verano y menores en los meses invernales.



En estas gráficas se muestran los diagramas de caja de la temperatura a través de los años, tanto de temperatura mínima como máxima por separado.

3. Conclusiones

La biblioteca Matplotlib es una herramienta muy útil para visualizar datos y realizar un análisis de los mismos. Sin embargo, se presentaron dificultades al momento de elaborar gráficas de caja, ya que no se encontraron los comandos para extraer los arreglos necesarios del DataFrame.

Bibliografía

- Matplotlib. Recuperado el 17 de febrero de 2019 desde <https://matplotlib.org/>
- Boxplot. seaborn.pydata.org. Recuperado el 17 de febrero desde <https://seaborn.pydata.org/generated/seaborn.boxplot.html>