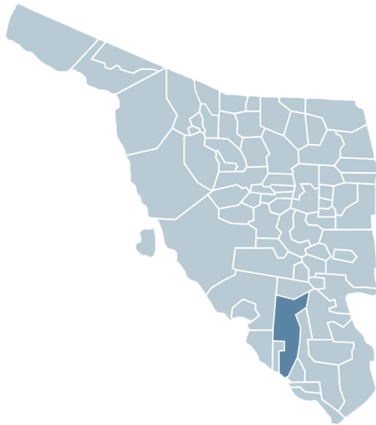


Reporte de la actividad 3

Daniela Olmos Velderrain
Grupo 3

10 de febrero de 2019

1. Introducción



En esta actividad se buscó aplicar el uso de la biblioteca Pandas en análisis de datos. El análisis se realizó sobre los datos meteorológicos de Cajeme, un municipio del estado de Sonora, los cuales fueron proporcionados por el Servicio Meteorológico Nacional. Se eligió este municipio en particular por contar con un amplio registro de datos, lo cual arrojará mejores resultados al hacer un análisis estadístico.

En la actividad se aplicaron distintas funciones de la biblioteca de Pandas, así como otras bibliotecas auxiliares para la realización de gráficas.

En base al análisis, se buscó responder distintas preguntas, sobre los niveles de precipitación y temperaturas registradas en las últimas décadas:

1. ¿Cómo podrás determinar cuáles son los meses más lluviosos?
2. ¿Cuáles son los meses más fríos y cuáles son los más cálidos?
3. ¿Cuáles han sido años muy húmedos?
4. ¿Cuáles han sido años muy secos?
5. ¿Cuáles años han tenido inviernos fríos?
6. ¿Cuáles años han tenido veranos más cálidos?
7. ¿Cómo ha venido siendo la temperatura mensual promedio en los últimos 20 años?
8. ¿Qué ha pasado con la precipitación en los últimos 20 años de datos?

2. Desarrollo

2.1. Metodología

Una vez obtenido el archivo de texto con los datos de cajeme, se creó un nuevo archivo en Jupyter Notebook, al cual se importaron las siguientes librerías:

```
import plotly.plotly as py
import plotly.tools as tls
import matplotlib.pyplot as plt; plt.rcParamsDefaults()
import pandas as pd
import numpy as np
```

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

Para poder leer los datos, evitando los valores nulos, a estos se les asignó una variable.

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

Después se leyó el archivo mediante el comando "pd.read_csv", 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 dar solución a las preguntas planteadas se emplearon diversas funciones sobre el Data Frame.

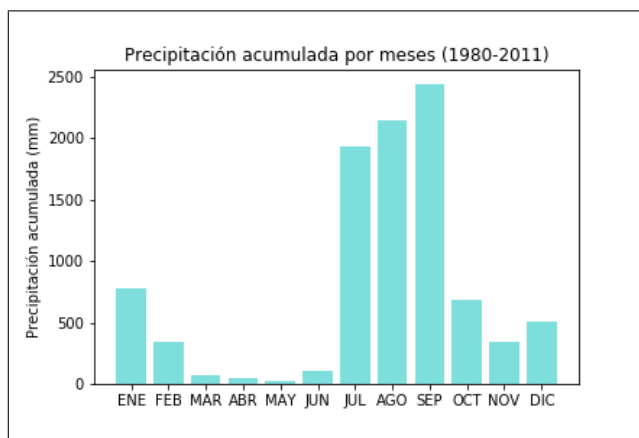
Cuadro 1: Funciones de Data Frame empleadas.

Función	Acción
df.head()	Encabezado
df.tail()	Final
df.dtypes()	Tipos de variables
df.mean()	Promedio
df.std()	Desviación Estándar
df.median()	Mediana
df.max()	Máximo
df.min()	Mínimo
df.describe()	Resumen estadístico
df.drop()	Eliminar
df.apply()	Aplicar función
df.idxmax()	Máximo (índice)
df.idxmin()	Mínimo (índice)
df.iloc()	Mostrar índice
df.loc()	Accesar
df.isin()	Selección
df.unique()	Valores diferentes
df.nsmallest()	N menores
df.nlargest()	N mayores

También se utilizó polyplot para realiza gráficas, y así lograr una mejor interpretación de los datos.

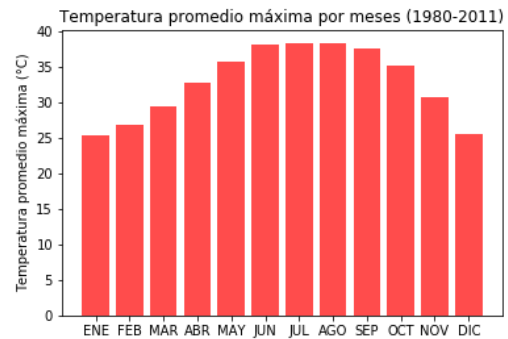
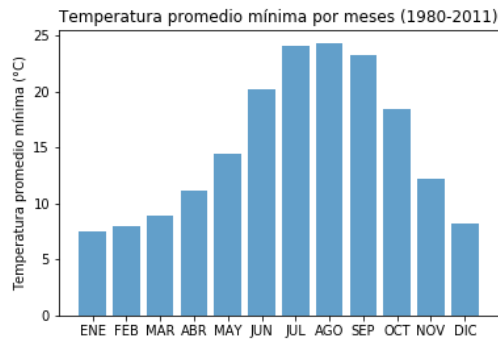
2.2. Resultados

2.2.1. ¿Cómo podrás determinar cuáles son los meses más lluviosos?



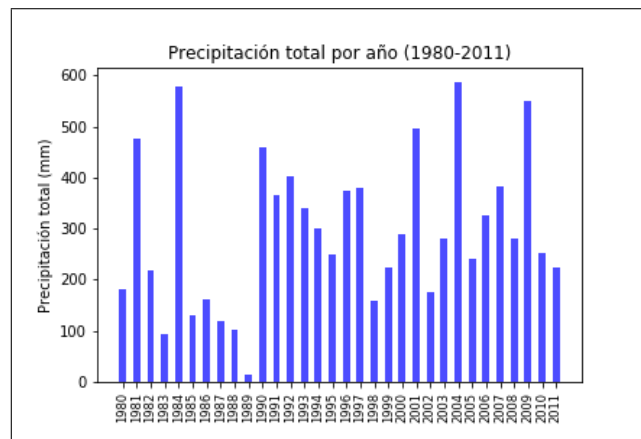
Para determinar esto se calculó la suma de precipitaciones totales para cada mes de todos los años. Se realizó una gráfica de barras con los resultados. Aquí se puede apreciar que los meses más lluviosos han sido agosto, septiembre y octubre.

2.2.2. ¿Cuáles son los meses más fríos y cuáles son los más cálidos?



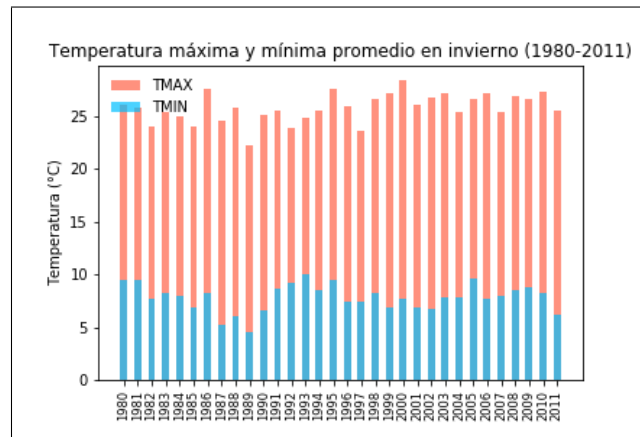
Para saber cuáles fueron los meses más fríos y cálidos se calculó el promedio de las temperaturas máximas y mínimas de manera mensual. En base a ellas podemos afirmar que los meses con temperaturas más bajas son diciembre, enero febrero y marzo. En cambio, los meses con temperaturas más altas son junio, julio, agosto y septiembre.

2.2.3. ¿Cuáles años han sido años muy húmedos?, ¿cuáles muy secos?, ¿qué ha pasado con la precipitación en los últimos 20 años?



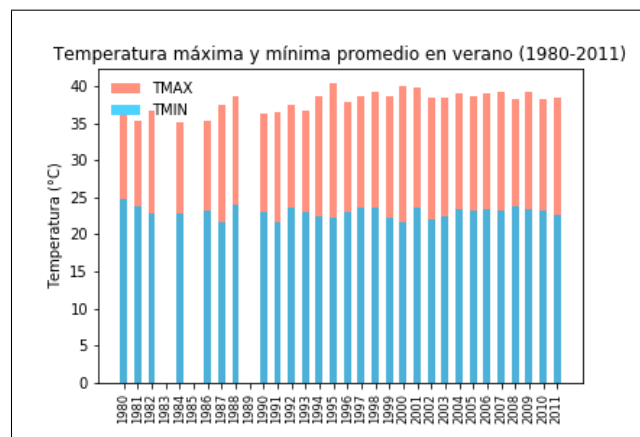
Para responder a esto se realizó un nuevo Data Frame con dos columnas. La primera columna contiene un arreglo de los años, y la segunda contiene la suma de precipitaciones para cada año. Al observar la gráfica correspondiente notamos que 1989, 1983 y 1988 fueron años muy secos, mientras que 2004, 1984 y 2009 han sido los años más lluviosos. Observando la información, vemos que las precipitaciones han ido aumentando en los últimos años.

2.2.4. ¿Cuáles años han tenido inviernos fríos?



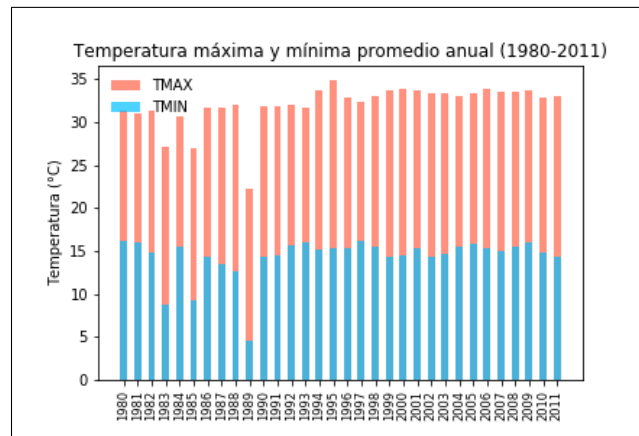
Para responder a la pregunta, se seleccionaron los datos de los meses de invierno de cada año, y se calculó el promedio de temperaturas máximas y mínimas. En la gráfica se observa que 1989 fue el año con el invierno más frío. Otros años con temperaturas mínimas muy bajas fueron 1987 y 1989.

2.2.5. ¿Cuáles años han tenido veranos cálidos?



En este caso, se seleccionaron los datos de los meses de verano de cada año, y se calculó el promedio de temperaturas máximas y mínimas. De aquí podemos observar que los años con temperaturas más altas han sido 1995, 2000 y 2001.

2.2.6. ¿Cómo ha venido siendo la temperatura mensual promedio en los últimos 20 años?



Debido a la extensión de los datos, se realizó un promedio de las temperaturas máximas y mínimas de manera anual. Aquí se observa que las temperaturas mínimas han sido cada vez mayores, al igual que las máximas.

3. Conclusiones

A partir de las gráficas observamos que desde 1980 hasta 2011 han aumentado la temperatura y las precipitaciones en el municipio de cajeme. También vemos que para este municipio, la época de lluvias se da en los meses de julio, agosto y septiembre. Estos resultados son congruentes con lo esperado, ya los cambios de temperatura y precipitación observados pueden atribuirse al calentamiento del planeta debido al cambio climático.