Análisis Exploratorio de Datos en Pitiquito, Sonora

Mario Antonio Martínez Cerda

5 de febrero de 2021

1 Introducción.

En estadística, el análisis exploratorio de datos es un enfoque para analizar conjuntos de datos para resumir sus características principales, a menudo con métodos visuales. Se puede usar o no un modelo estadístico, pero principalmente EDA es para ver qué nos pueden decir los datos más allá del modelado formal o la tarea de prueba de hipótesis.
John Tukey promovió el análisis de datos exploratorios para alentar a los estadísticos a explorar los datos y posiblemente formular hipótesis que podrían conducir a la recopilación de los mismos y experimentos nuevos.

Durante el siguiente documento, veremos el análisis de datos observados en el archivo de precipitaciones, temperaturas y evaporación en la ciudad de Pitiquito, Sonora. Proporcionados por la Conagua, con datos desde 1952. Realizando gráficas de cajas, de barra y muchos más estilos de visualización de datos, con ayuda de la biblioteca seaborn y matplotlib.
Redactaremos la actividad a realizar durante esta actividad en Física computacional y proporcionaremos las imágenes como evidencia de lo realizado.

2 Actividades y evidencias.

2.1 Actividad 1.

Crea un nuevo cuaderno de trabajo de Jupyter llamado Actividad4.ipynb en Google Colab.

Por favor, resume en una sola celda todas las funciones que aplicaste al DataFrame inicial y que concluya con la creación de un nuevo DataFrame para continuar con nuestro trabajo.

Se pide sintetizar las características principales del conjunto de datos que estas analizando, aplicando la siguiente secuencia de funciones de un proceso EDA

arriba mencionadas.

En los datos climatológicos con los que hemos trabajado anteriormente, podemos mencionar que hubo demasiados datos nulos en la columna de evaporación.

Tuvimos que realizar un dataframe con los datos, cambiando un poco los datos para que se ajusten a una lectura de datos fáciles de interpretar para Python. Entonces, realizamos varias impresiones de los datos para observar cuales cambios son los que realizaremos para el desarrollo de la actividad, entre ellos:

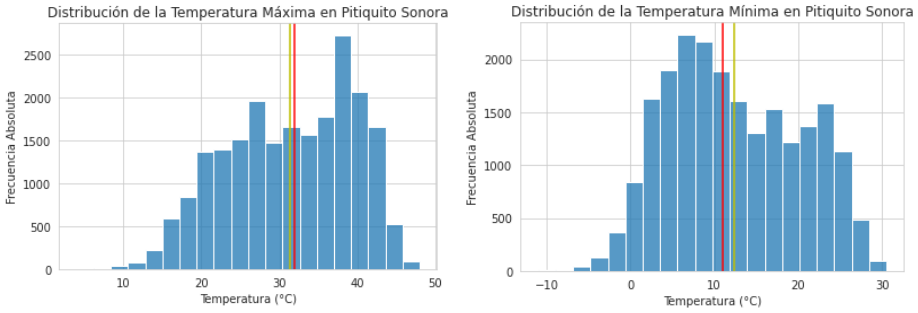
-Vimos las características.
-Vimos el contenido el dataframe.
-Creamos una copia de seguridad.
-Agregamos un Loop a las columnas que nos interesaban que fueran números flotantes.
-Vimos los datos nulos que habia en las columnas y donde se nos presentaorn más numeros nulos es en la evaporación.
-Pedimos una descripción de los datos con 4 cifras significativas.
-Los datos de las fechas las realizamos de tal forma para que se detectara en Python como un dato tipo fecha.
-Al final, pedimos la información para ver que los datos que tenemos están como pedimos.
-Con los datos, vemos que tienden a ser más presentes durante la segunda mitad del año. Ahí es donde realmente veremos algunos datos interesantes e interpretables.

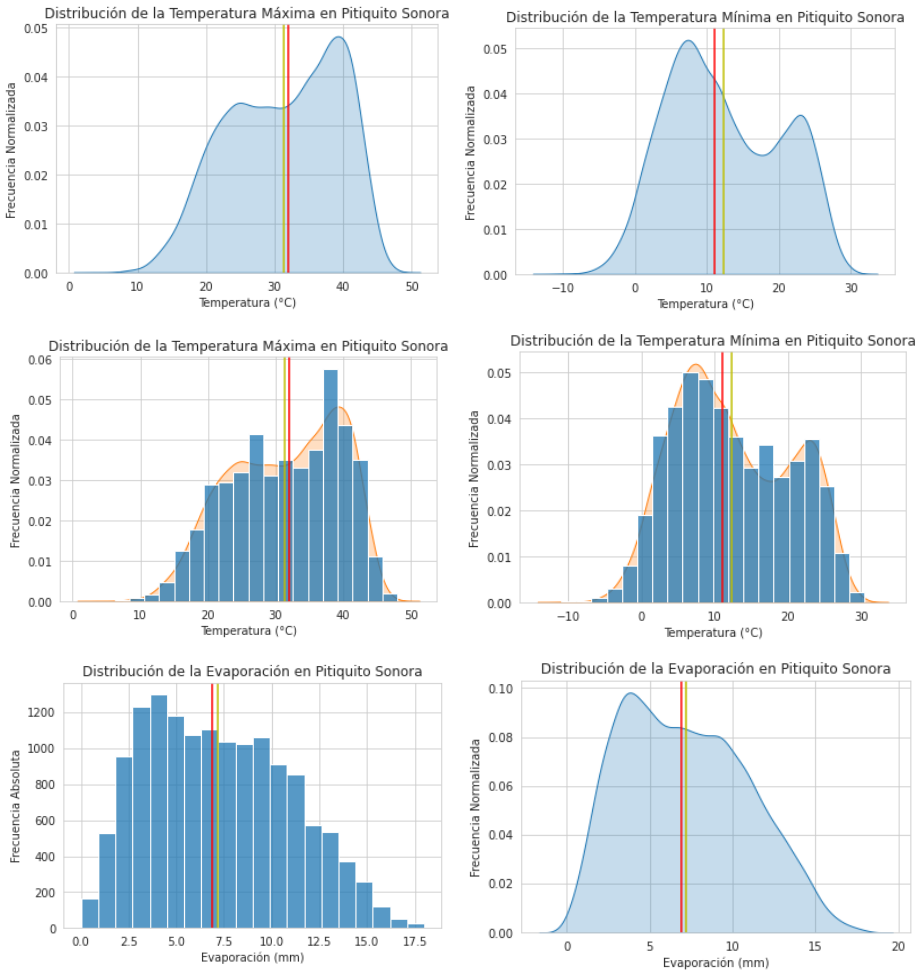
```
#Importamos las bibliotecas.
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
```

2.2 Actividad 2.

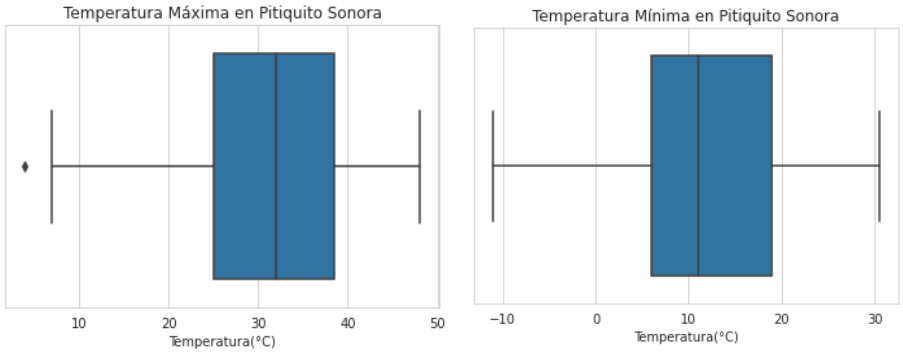
Crear Histogramas de las variables de Precipitación, Evaporación, Temperaturas Máxima y Mínima de el conjunto de datos que se están analizando (Función: sns.histplot()). Complementar en su caso con las gráficas de la función de densidad de probabilidad correspondiente (Función: sns.kdeplot())

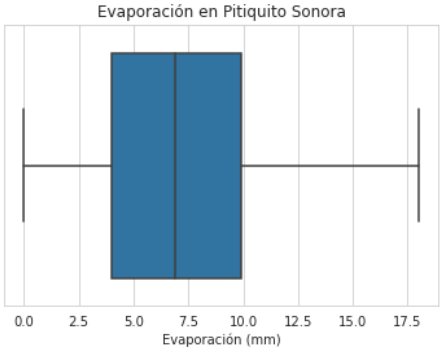




2.3 Actividad 3.

Crear las gráficas de cajas (Boxplot) para la Evaporación, Temperaturas Máxima y Mínima (Función: sns.boxplot())

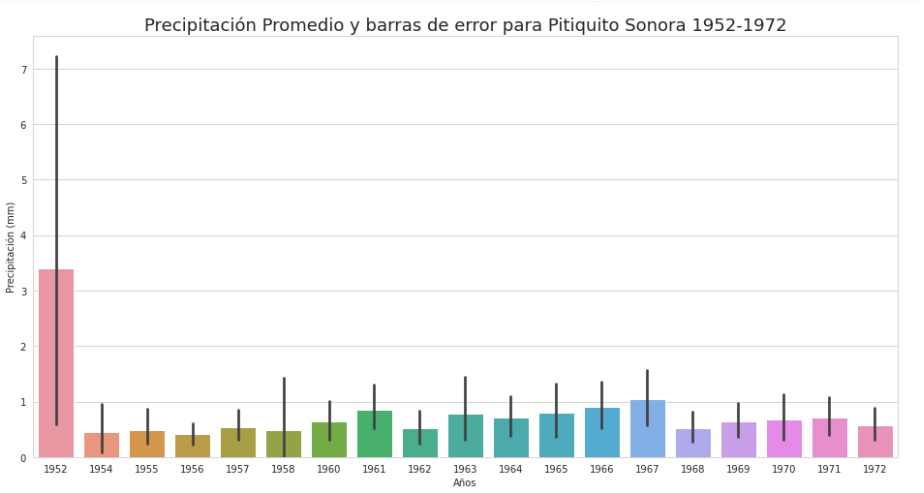


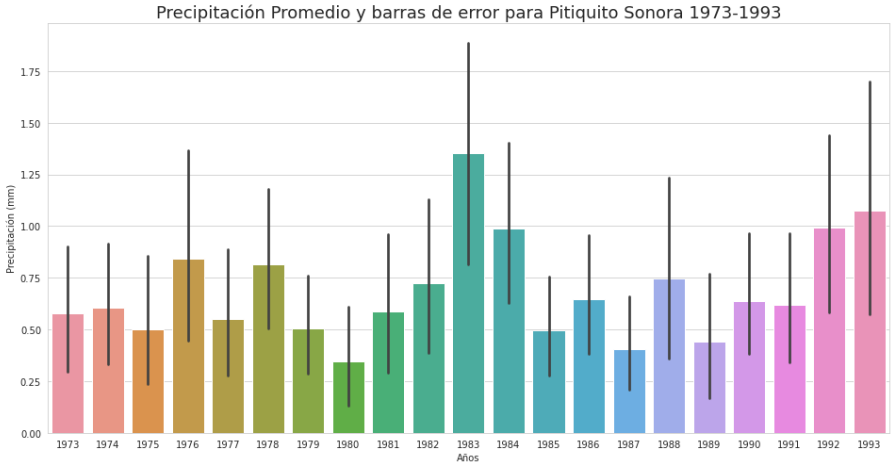


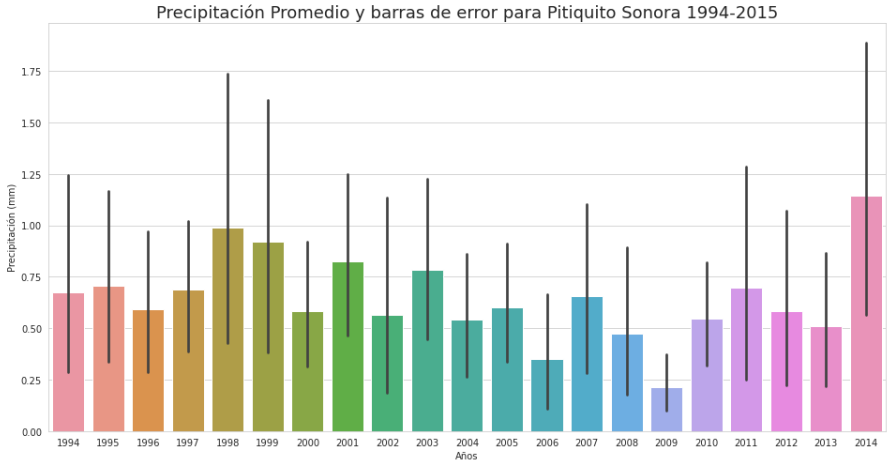
2.4 Actividad 4.

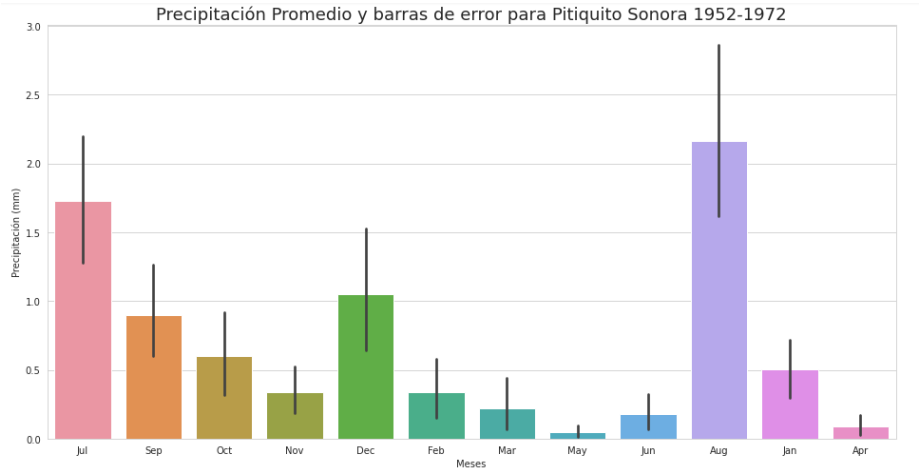
Produzca las gráficas de barras para la Precipitación agrupado por Años y después por meses (Función: sns.barplot())

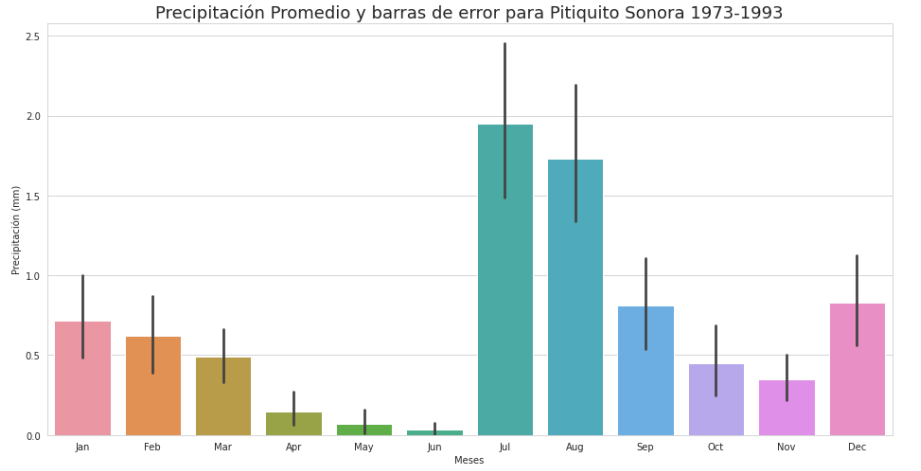
```
[25] #Como comenzamos en 1952, entonces seleccionamos datos a partir de ahi.
     #Lo más óptimo es de 21 datos a partir de 1952, para recolectar todos.
     df_52 = df_EDA[(df_EDA['Año'] >= 1952) & (df_EDA['Año'] < 1973)]
     df_73 = df_EDA[(df_EDA['Año'] >= 1973) & (df_EDA['Año'] < 1994)]
     df_94 = df_EDA[(df_EDA['Año'] >= 1994) & (df_EDA['Año'] < 2015)]
```

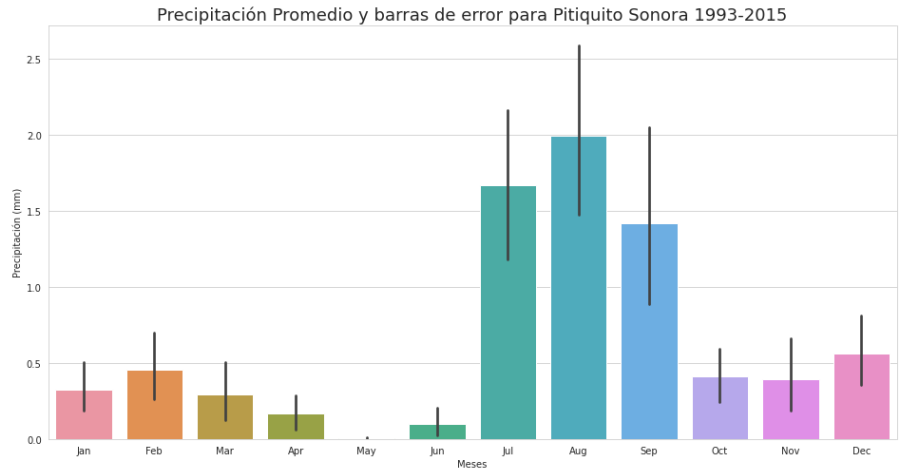

Precipitación Promedio y barras de error para Pitiquito Sonora 1952-1972
Precipitación (mm)
0
1
2
3
4
5
6
7
1952
1954
1955
1956
1957
1958
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
Años


Precipitación Promedio y barras de error para Pitiquito Sonora 1973-1993
Precipitación (mm)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
Años


Precipitación Promedio y barras de error para Pitiquito Sonora 1994-2015
Precipitación (mm)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
Años


Precipitación Promedio y barras de error para Pitiquito Sonora 1952-1972
Precipitación (mm)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Jul
Sep
Oct
Nov
Dec
Feb
Mar
May
Jun
Aug
Jan
Apr
Meses

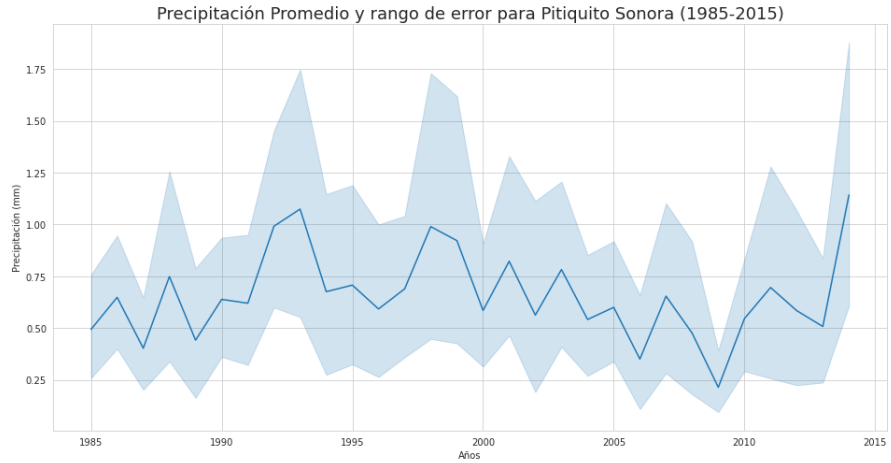

Precipitación Promedio y barras de error para Pitiquito Sonora 1973-1993
Precipitación (mm)
2.5
2.0
1.5
1.0
0.5
0.0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Meses

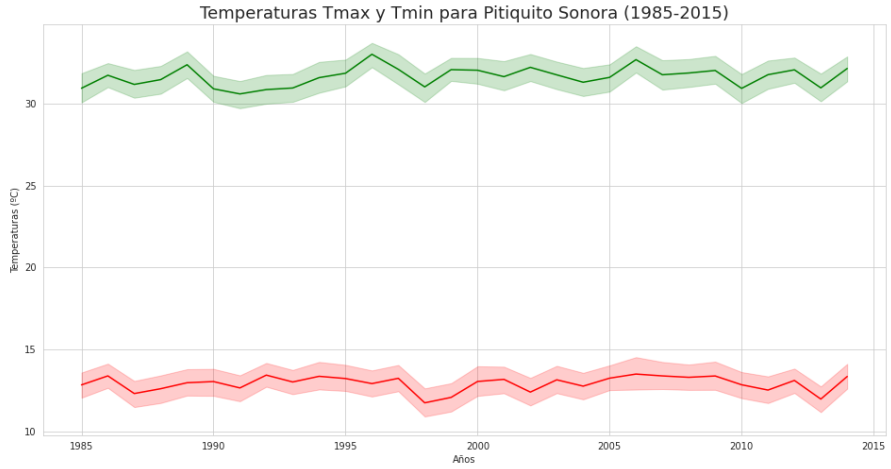

Precipitación Promedio y barras de error para Pitiquito Sonora 1993-2015
Precipitación (mm)
2.5
2.0
1.5
1.0
0.5
0.0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Meses

2.5 Actividad 5.

Por favor cree una colección de los últimos 30 años de datos, utilizando condiciones de filtrado por un rango de años. Crear las gráficas de línea de la Precipitación, Temperaturas Máxima y Mínima como funciones del tiempo (Últimos 30 Años). (Función: sns.lineplot()).

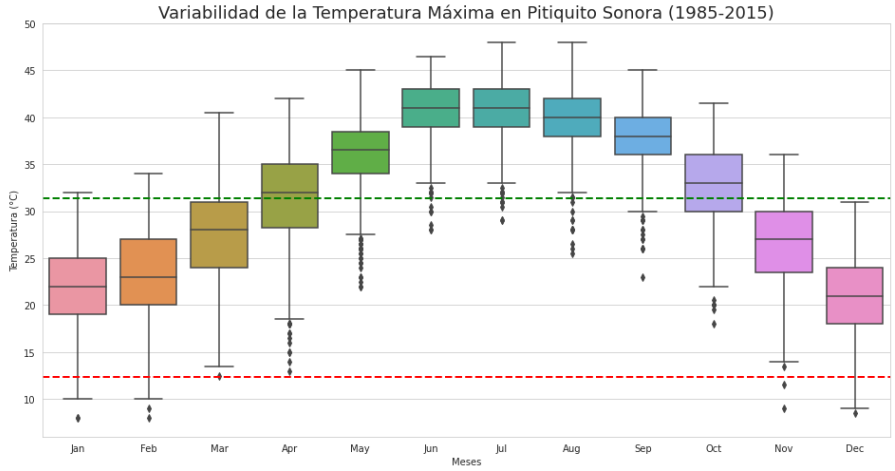
```
[36] #Definiré aqui, el periodo de años que me servirán para tener la gráfica de los ultimos
     #30 años.
     df_30 = df_EDA[(df_EDA['Año'] >= 1985) & (df_EDA['Año'] < 2015)]
```

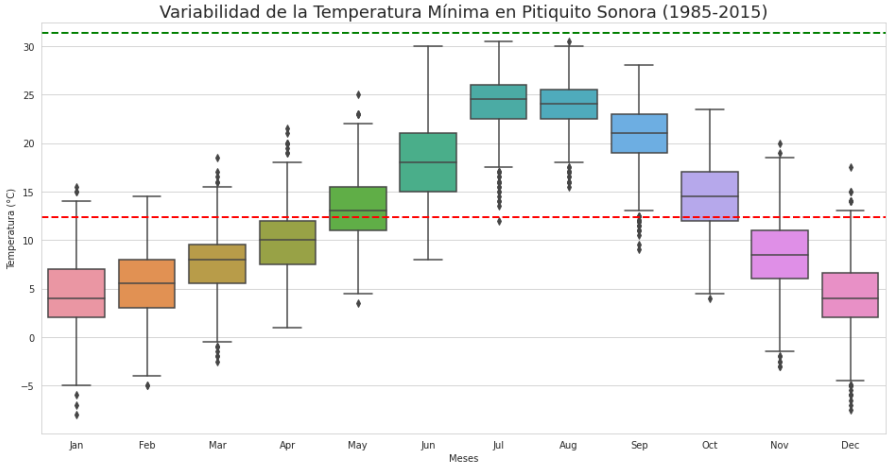


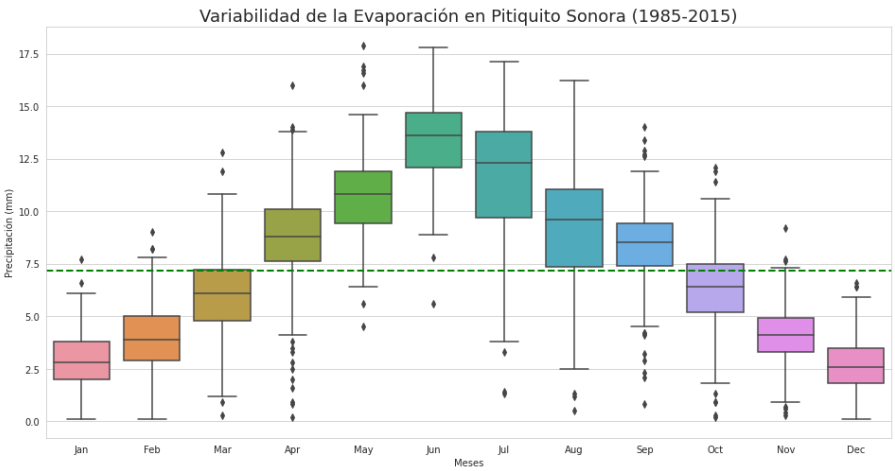


2.6 Actividad 6.

Con el conjunto de 30 años de datos, produzca diagramas de cajas (Función: sns.boxplot()) para observar la variabilidad de las Temperaturas (Max y Tmin) y la Evaporación agrupados por Mes.


Variabilidad de la Temperatura Máxima en Pitiquito Sonora (1985-2015)
Temperatura (°C)
Meses
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


Variabilidad de la Temperatura Mínima en Pitiquito Sonora (1985-2015)
Temperatura (°C)
Meses
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


Variabilidad de la Evaporación en Pitiquito Sonora (1985-2015)
Precipitación (mm)
Meses
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

3 Conclusión

Durante la realización del código en Python, se me dificultaron varias cosas, entre ellas es saber exactamente las funciones que tiene cada comando para las gráficas y la interpretación de las gráficas, debido a que no recordaba de buena manera como se comportaban las gráficas de cajas. Fue una actividad muy larga pero sencilla, donde el tiempo se me hizo un poco corto por cosas externas a la actividad. Lo que si no me quedo muy claro fue la Actividad 1, donde no supe si se tenía solamente que escribir o realizar un análisis más a detalle. El grado de complejidad de la actividad en cuestión lo colocaría como intermedio, sólo por que es larga.

4 Bibliografía.

1.-http://computacional1.pbworks.com/w/page/143144913/Actividad%204%20(2021-1)
2.-https://en.wikipedia.org/wiki/Exploratory_data_analysis