

UNIVERSIDAD DE SONORA



"El saber de mis hijos
hará mi grandeza"

FÍSICA COMPUTACIONAL I

Actividad 2

Iniciando el análisis de datos con la biblioteca Pandas de Python

1. Introducción

En la presente práctica de Python, hicimos uso de un cuadernillo de trabajo generado en Jupyter notebook y trabajamos con las bibliotecas de Panda y Matplotlib para analizar los datos recabados por las Estaciones Meteorológicas Automáticas (EMAS)¹ del Servicio Meteorológico Nacional de CONAGUA.

Hay que tener en cuenta que el objetivo principal de esta actividad consiste en aprender a manejar series de datos haciendo uso de Pandas para finalizar creando dos gráficos distintos y poder analizar los datos tras esto.

2. Desarrollo del código

A continuación veremos paso a paso cómo fue que se manejaron los datos y posteriormente, la creación de las gráficas con los mismos.

2.1. Importación de bibliotecas y datos

En el bloque de código [1] podemos ver que importamos las bibliotecas necesarias y además, hacemos uso del comando mágico `%matplotlib inline` para que los gráficos que creemos se muestren dentro del cuadernillo; el bloque [2] muestra la configuración utilizada para leer el archivo .CSV que contiene las lecturas de la estación.

```
In [1]: 1 import pandas as pd
        2 import matplotlib.pyplot as plt
        3 %matplotlib inline

In [2]: 1 #Lectura del archivo (dirección, separador, salto de líneas, encabezado)
        2 Estacion = pd.read_csv('UNISON_90.csv', skiprows=9, header=None)
```

Figura 1: Importando bibliotecas y leyendo el archivo .CSV.

¹<http://smn.cna.gob.mx/es/estaciones-meteorologicas-automaticas-2>

2.2. Reordenando datos y formatos

Lo siguiente a tratar son los encabezados de cada columna pues no cuentan con un nombre, así que habremos de asignárselos nosotros mismos.

```
In [4]: 1 #Asignaremos los nombres de cada columna
        2 Estacion.columns = ['Fecha', 'DirRaf', 'DirVien', 'HumRel', 'Precip',
        3                       'RadSolar', 'Tempe', 'PresAtm']
```

Figura 2: Cada columna es nombrada apropiadamente.

Al verificar los tipos de cada elemento en el marco de datos podremos ver que la columna **Fecha** es simplemente marcada como un objeto cuando debería ser reconocido como **datetime**, así que modificaremos eso en el bloque [7].

```
In [6]: 1 Estacion.dtypes

Out[6]: Fecha          object
        DirRaf         int64
        DirVien        int64
        HumRel         float64
        Precip         float64
        RadSolar       float64
        Tempe          float64
        PresAtm        float64
        dtype: object

In [7]: 1 #La columna FECHA se marca como OBJECT, ahora lo convertiremos a DATETIME
        2 Estacion['Fecha'] = pd.to_datetime(Estacion['Fecha'])
```

Figura 3: Modificamos el tipo de la columna **Fecha**.

2.3. Nuevo marco de datos

Nuestro próximo paso consiste en obtener un nuevo marco de datos que contenga la información de 10 días, los cuales nosotros podremos elegir al momento de introducir la fecha en la condición lógica para delimitar la selección del bloque [10] (Véase la figura 4).

```
In [10]: 1 #Solamente necesitamos los 'ultimos 10 dias, debemos seleccionarlos.
2 DiezDias = Esta2[(Esta2['Fecha'] > '2018-08-24 23:59:59') &
3             (Esta2['Fecha'] < '2018-09-04 00:00:00')]
4
5 print(DiezDias)
```

Figura 4: Delimitamos la selección con una condición lógica.

2.4. Gráficos

En este punto tenemos bien ordenados los datos y simplemente resta configurar cada una de las gráficas a crear.

La primer gráfica es generada con el código del bloque [11].

```
In [11]: 1 #Gr'afica de Temperatura durante los 'ultimos 10 d'ias
2 plt.title('Temperatura de 10 días', fontweight='bold')
3
4 plt.xlabel('Días', fontweight='bold')
5 plt.xticks(rotation=45)
6 plt.ylabel('Temperatura (°C)', fontweight='bold')
7
8 plt.grid(True)
9
10 x = DiezDias['Fecha']
11 y = DiezDias['Tempe']
12
13 plt.plot(x,y)
14
15 plt.savefig('Tempe10', bbox_inches='tight')
```

Figura 5: Configuración de la gráfica de temperatura.

Con los comandos mostrados podemos genera una gráfica igual a la que veremos en la figura 6.

Los comandos para generar la segunda gráfica se muestran en la figura 7, la cual corresponde al bloque de código [12], el mismo que permite generar una gráfica que podemos apreciar en la figura 8.

3. Conclusión

Tras llevar a cabo este código podemos ver lo útil que resulta la biblioteca de Pandas a la hora de manejar marcos de datos pues cuentan con una amplia

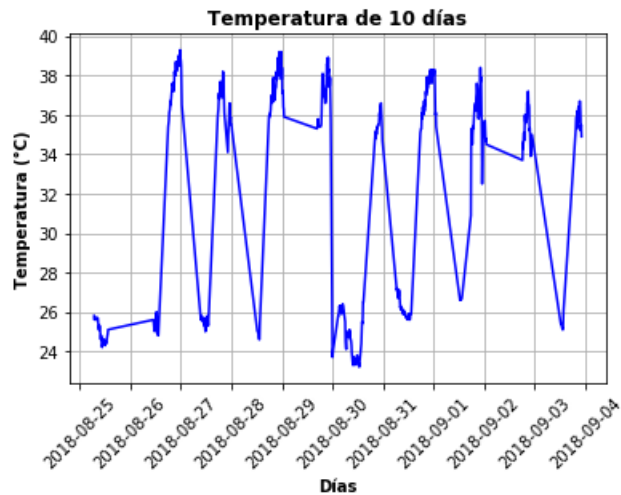


Figura 6: Temperatura contra tiempo durante 10 días.

```
In [16]: 1 #Gr'afica de Presi'on durante los 'ultimos 10 d'ias
2 plt.title('Presión de 10 días', fontweight='bold')
3
4 plt.xlabel('Días', fontweight='bold')
5 plt.xticks(rotation=45)
6 plt.ylabel('Presión (hPa)', fontweight='bold')
7
8 plt.grid(True)
9
10 x = DiezDias['Fecha']
11 y = DiezDias['PresAtm']
12
13 plt.plot(x,y, 'g')
14
15 plt.savefig('Pres10', bbox_inches='tight')
```

Figura 7: Configuración de la gráfica de presión.

lista de comandos con los cuales podemos resolver y aplicar un tratamiento de "limpieza." a los conjuntos de información que obtengamos.

Si bien queremos dar una opinión acerca de lo visto en las gráficas, podemos mencionar que parecen estar inversamente relacionados los datos analizados pues cuando el valor de la presión disminuye la temperatura sufre un incremento.

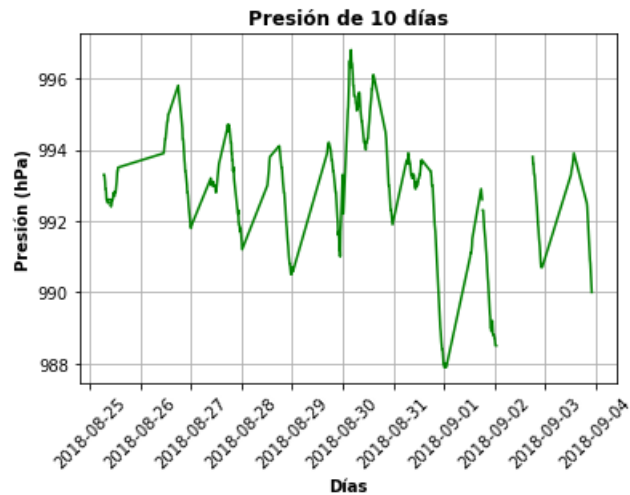


Figura 8: Presión contra tiempo durante 10 días.

4. Apéndice

1. ¿Cuál es tu impresión de trabajar con Python en Jupyter?

Parece hacer todo mucho más sencillo por la gran cantidad de ventajas que ofrece, además de hacer un poco más visual todo, es bastante amigable con el usuario.

2. ¿Cuál es tu primera impresión de trabajar con Pandas?

La biblioteca de Pandas cuenta con una gran cantidad de funciones muy útiles cuando se trata de manejar los marcos de datos, parece ser más que suficiente por el momento.

3. ¿Qué partes se te dificultaron más?

Lo más complicado siento que fue aplicarle un "tratamiento" a los datos para que fuera más sencillo trabajarlos, en la forma y formato adecuado.

4. ¿Qué te pareció esta Actividad? ¿Alcanzó el tiempo?

Creo que la actividad es bastante apropiada para dar una introducción al análisis de datos pues la biblioteca de Pandas veo que cuenta con un gran respaldo por parte de muchos usuarios, haciéndola un elemento indispensable para este tema.

5. ¿Qué le agregarías o quitarías a esta actividad?

Me gustaría que la actividad tuviera otra cosa más por hacer, así como una gráfica doble para apreciar mejor el comportamiento de los datos.

6. ¿Qué cosas te parecieron interesantes y qué se te hizo aburrido?

Lo interesante sería el comenzar a manejar un nuevo lenguaje y utilizar Jupyter Notebook; lo aburrido, podría decirse que es al momento de buscar información para realizar lo que se desea pues en ocasiones se visita una página que no ofrece buena información, ejemplos o aquello que buscamos.

Referencias

- [1] Mapa de Estaciones Meteorológicas Automáticas (EMAS)(versión 2014-2017). Recuperado de <http://smn.cna.gob.mx/es/estaciones-meteorologicas-automaticas-2>