



UNIVERSIDAD DE SONORA  
DEPARTAMENTO DE CIENCIAS EXACTAS Y NATURALES

## Iniciándose en Python

---

Por Álvarez Sánchez Francisco Eduardo  
Maestro: Dr. Carlos Lizárrga Celaya

Hermosillo, Sonora a 17 de febrero de 2017

# Índice

<b>1. Introducción</b>	<b>3</b>
<b>2. Desarrollo del estudio de datos</b>	<b>4</b>
2.1. Preparación de datos . . . . .	4
2.2. Jupyter Notebook . . . . .	4
2.3. CAPE . . . . .	5
2.4. Pricipitable Water (PW) . . . . .	9
<b>3. Conclusión</b>	<b>10</b>

## 1. Introducción

En el siguiente documento se presenta una interpretación de datos que hemos obtenido anteriormente en las actividades 1 y 2. Estos datos pertenecen a la Universidad de Wyoming, en donde se estudian diferentes localidades de Estados Unidos y otras repartidas por el mundo. En este caso en específico trabajaremos con los datos de la ciudad Del Rio en Texas. Utilizando como herramienta para manejar los datos utilizaremos la plataforma Jupyter Notebook con Python para realizar tablas y gráficas de diferentes tipos.

## 2. Desarrollo del estudio de datos

### 2.1. Preparación de datos

Utilizando los datos obtenidos con anterioridad acerca de las condiciones meteorológicas de la atmósfera en determinado sitio de observación, realizaremos una limpieza de estos mismos y obtendremos datos específicos de cada día del año.

Usaremos algunos comandos en bash, se unieron todos los archivos HTML en uno solo con formato .txt con el nombre de sondeos.txt. Con la terminal y utilizando los comandos para filtrar ciertas variables, separamos la información en 2 archivos; uno con la información de los sondeos realizados en el formato 12Z y otro con los sondeos en 00Z.

De los archivos queremos rescatar la información denominada como CAPE y otro llamado Precipitable water o PW. Para realizar esto, se utilizó el siguiente comando en la terminal:

```
cat sondeos.txt | egrep -i "Observations| CAPE | precipitable" |  
sed -e "/12Z/,+2d"> 00Zanual.txt
```

Para los días en los que se hicieron sondeos a las 00Z (GMT), y

```
cat sondeos.txt | egrep -i "Observations| CAPE | precipitable"  
| sed -e "/00Z/,+2d" > 12Zanual.txt
```

Para todos los días que se hicieron los sondeos a las 12Z (GMT). Con estos datos, podemos comenzar a utilizar Jupyter Notebook para realizar una interpretación de estos datos.

### 2.2. Jupyter Notebook

Ahora, iniciamos desde la terminal la plataforma Jupyter Notebook. Este abre una página en el navegador y damos inicio a un archivo nuevo. Para comenzar, necesitamos indicar las bibliotecas a utilizar. En este caso utilizaremos panda y para ello colocamos los siguientes comandos en distintas líneas cada uno para ejecutarlos presionando SHIFT+ENTER.

```
import pandas as pd  
import numpy as np  
import matplotlib as plt
```

Procederemos a leer el archivo desde Jupyter con el comando siguiente:

```
df= pd.read_csv("/home/franciscoeas/Actividad3/12ZAnual.csv"  
, names=['Fecha','CAPE', 'PW'])
```

Ya teniendo leídos los datos, pasaremos a realizar las tablas. Una contiene los primeros 10 datos del archivo con sus respectivas variables y en la segundo se observan datos estadísticos acerca de todos los datos. Los comandos utilizados fueron:

```
df.head(10)  
df.describe()  
df.apply(lambda x:sum(x.isnull()),axis=0)
```

	Fecha	CAPE	PW
0	01 Jan 2016	0.0	23.27
1	02 Jan 2016	0.0	25.33
2	03 Jan 2016	0.0	15.45
3	04 Jan 2016	0.0	10.93
4	05 Jan 2016	0.0	16.36
5	06 Jan 2016	0.0	20.46
6	07 Jan 2016	0.0	14.18
7	08 Jan 2016	0.0	9.02
8	09 Jan 2016	0.0	8.25
9	10 Jan 2016	0.0	7.87

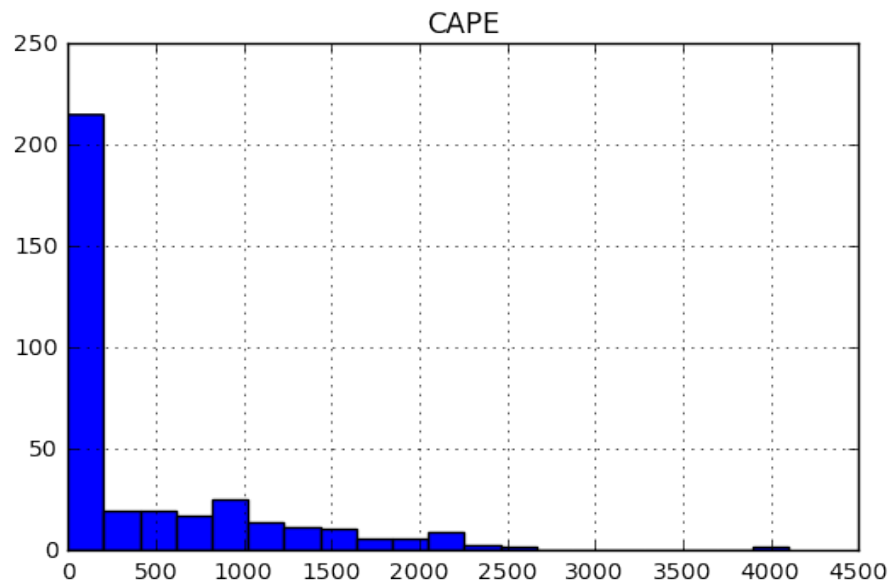
	CAPE	PW
count	362.000000	362.000000
mean	456.982956	28.669724
std	689.868515	13.559319
min	0.000000	0.000000
25%	0.000000	16.460000
50%	11.120000	29.255000
75%	810.912500	39.822500
max	4103.580000	63.750000

## 2.3. CAPE

Para iniciar los análisis estadísticos de la variable, utilizaremos Jupyter Notebook con Python para hacer un histograma, y dos diagramas de caja; uno para todo el conjunto de datos y el segundo será una comparativa de los datos organizados por un mes

```
df.columns
matplotlib inline
df.hist(u'CAPE',bins=20)
```

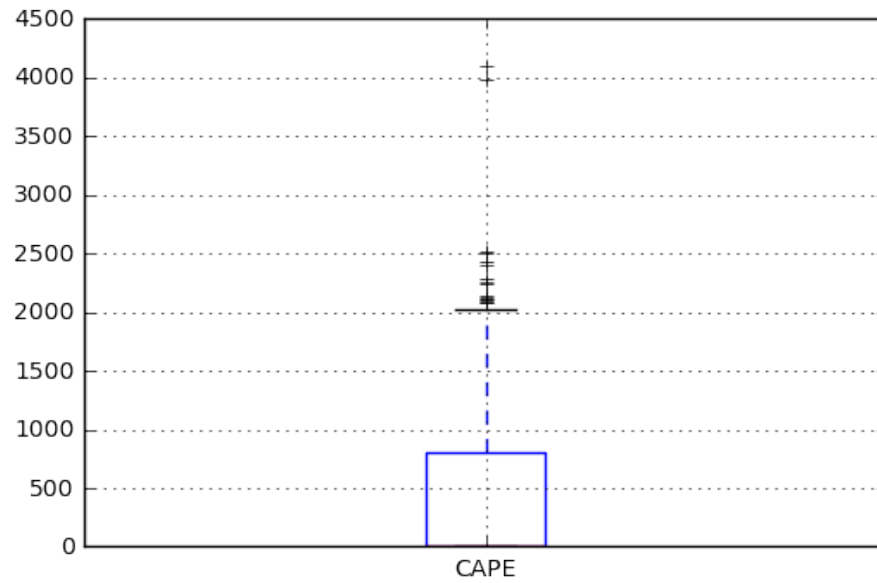
Son los comandos utilizados para generar el siguiente histograma de la variable CAPE



Para generar un diagrama de caja, utilizaremos el siguiente comando:

```
df.boxplot(column=u'CAPE')
```

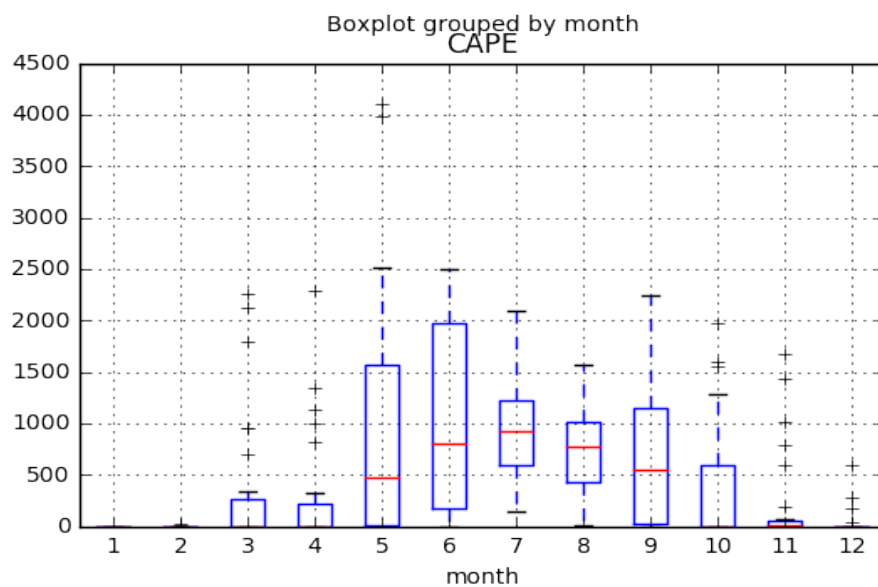
Y el resultado es el siguiente:



Ahora, para realizar una comparativa con mes, utilizaremos los siguientes comandos para realizar un diagrama:

```
df['month']=pd.DatetimeIndex(df[u'Fecha']).month
```

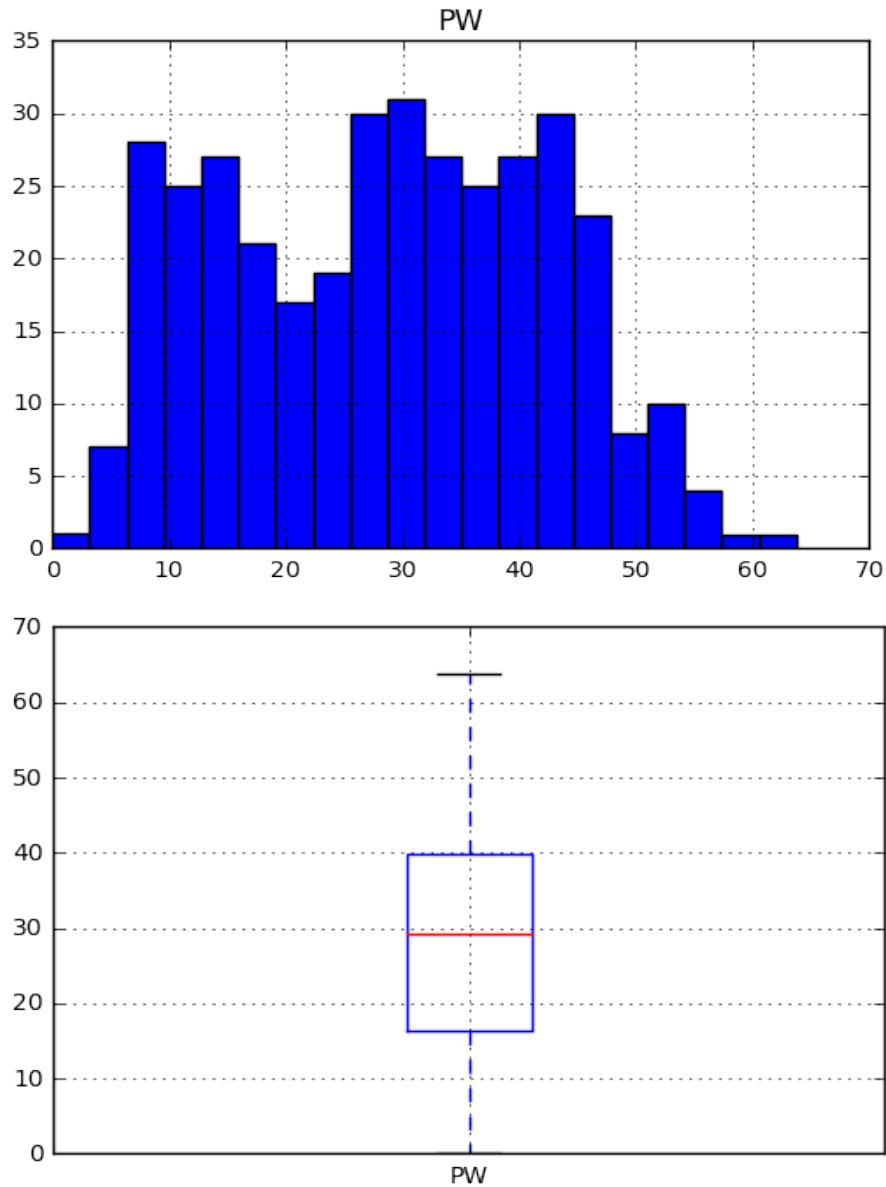
```
df.boxplot(column=u'CAPE',by=u'month')
```

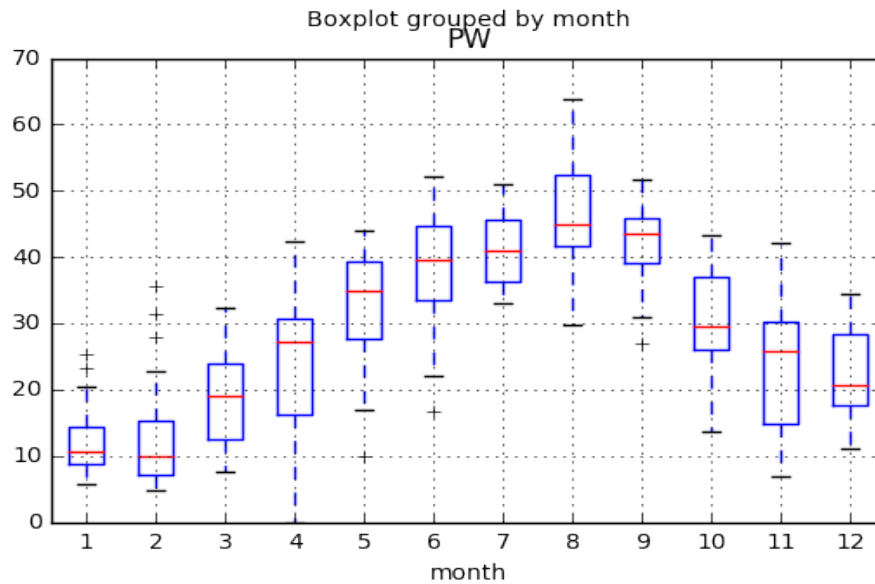




## 2.4. Pricipitable Water (PW)

Procederemos a realizar lo mismo pero ahora con los datos de Precipital Water. Repetiremos los pasos pero ahora con el conjunto de datos PW. Obtenemos como resultado:





### 3. Conclusión

Esta actividad resultó bastante útil para próximos trabajos. Fue investigar de los comandos necesarios de Python y ya con esto realizar la actividad. Resulto bastante entretenida y muy productiva, nos enseñó a que hacer una vez que tenemos los datos en nuestro poder. Python ofrece un sin fin de utilidades y aprenderemos a usarlo poco a poco. Esta fue una de las muchas capacidades de python.