



UNIVERSIDAD DE SONORA

FÍSICA COMPUTACIONAL 1

INICIÁNDOSE EN PYTHON

Campos Quiñonez Jorge Andres

29 de Enero de 2017

Índice

1. Introducción	4
2. Desarrollo	5
2.1. Preparación de los datos	5
2.2. Jupyter Notebook	6
2.3. Análisis de la variable CAPE	7
2.4. Análisis de datos de Precipitable Water	9
3. Conclusión	11

1. Introducción

En esta actividad, realizaremos una exploración e interpretación de los datos que hemos obtenido gracias a actividades anteriores de datos meteorológicos recopilados por la Universidad de Wyoming de distintos lugares de norte américa. En esta ocasión, al igual que en las anteriores, nos enfocaremos e interpretaremos los datos obtenidos de las observaciones realizadas en la Colonia JuanCarrasco. Utilizando una terminal y los archivos de datos anteriormente obtenidos, se limpiaran y se utilizarán en Jupyter Notebook con Python para realizar interpretaciones de los mismos.

2. Desarrollo

2.1. Preparación de los datos

Utilizando los datos obtenidos en la actividad anterior acerca de las condiciones meteorológicas de la atmósfera, en determinado sitio de observación, realizaremos una limpieza de estos mismos y buscar obtener datos específicos de cada día en todo el año.

Con determinados comandos en bash, se unieron todos los archivos .HTML descargados en un solo archivo .txt con el nombre de sondeos.txt. Con ayuda de la terminal y utilizando comandos para filtrar ciertas variables, separamos la información en 2 archivos; uno para todos los días que tuvieran sondeo a las 00Z horas y otro a las 12Z horas. Cada archivo debe poseer lo siguiente: Fecha, datos de CAPE y datos para Precipitable Water. 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 todos los días en 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 en que se hicieron 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, se inició con la terminal lo que es "Jupyter Notebook". Esta abre una página en un 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
```

Y realizamos una lectura del archivo creado con Emacs anteriormente al cual le hicimos limpieza con el siguiente comando:

```
df = pd.read_csv("/home/elmadafaka/Física_computacional/Actividad 3/12ZAnual.csv", names=['Fecha', 'CAPE', 'PW'])
```

Con los datos ya leídos, procedimos a realizar 2 tablas. Una contiene los 10 primeros datos del archivo con sus respectivas variables y en la segunda se observan datos estadísticos acerca de todos los datos. Los comandos utilizados fueron los siguientes:

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

	Fecha	CAPE	PW
0	01 Jan 2016	0.00	42.49
1	02 Jan 2016	13.71	38.46
2	03 Jan 2016	23.83	19.38
3	04 Jan 2016	84.00	19.69
4	05 Jan 2016	0.00	26.17
5	06 Jan 2016	137.93	25.85
6	07 Jan 2016	36.04	21.74
7	08 Jan 2016	0.00	16.92
8	09 Jan 2016	0.00	12.60
9	10 Jan 2016	0.00	13.10

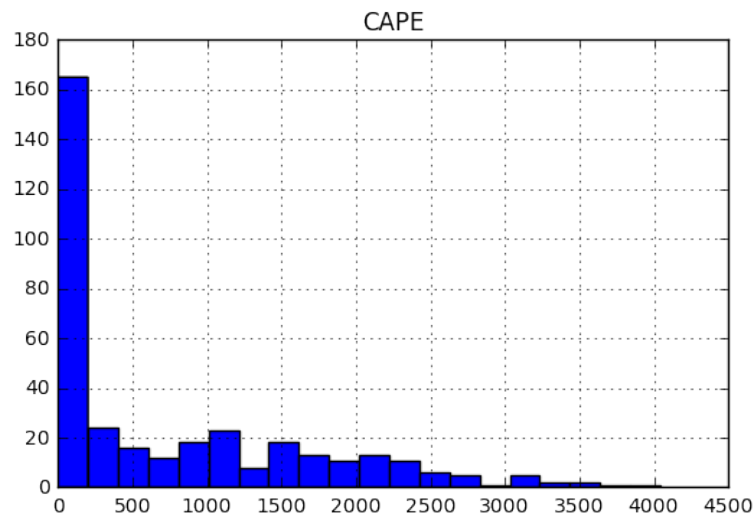
	CAPE	PW
count	355.000000	355.000000
mean	768.746197	38.657662
std	931.220413	17.065298
min	0.000000	8.970000
25%	0.000000	24.235000
50%	292.870000	36.570000
75%	1406.210000	55.115000
max	4043.510000	77.470000

2.3. Análisis de la variable 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 mes.

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

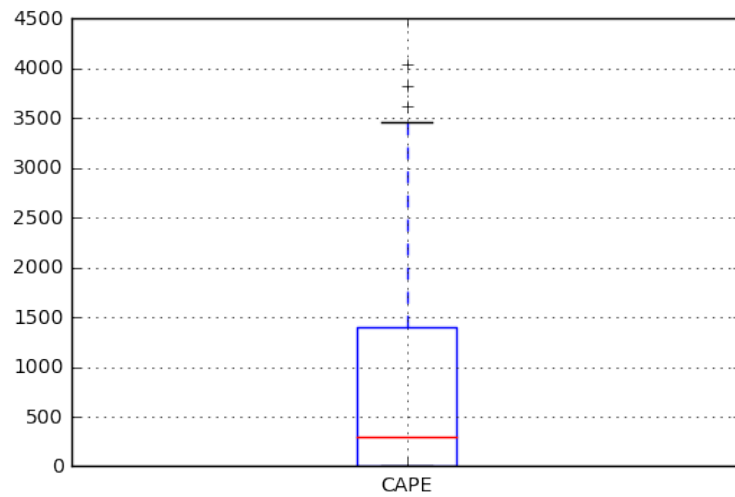
Son los comandos utilizados para generar el siguiente histograma de la variable CAPE.



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

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

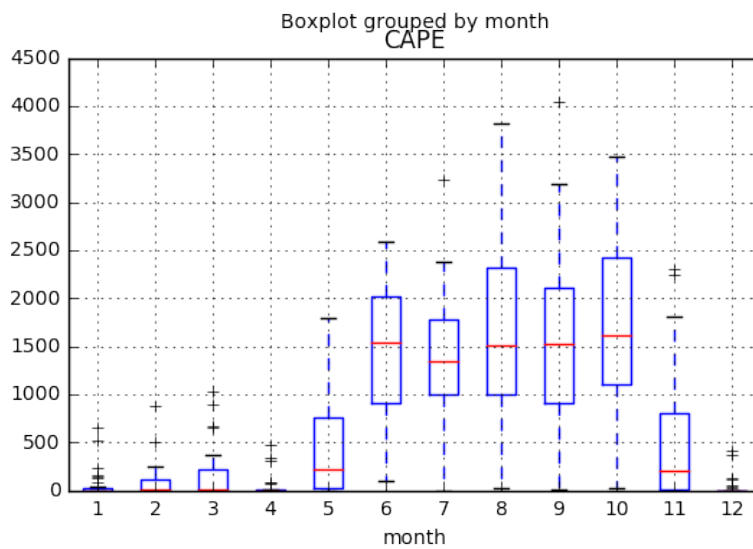
Y se forma el siguiente diagrama:



Ahora, para realizar una comparativa con mes, utilizaremos los siguientes comandos para realizar un diagrama de caja por mes de todos los datos.

```
df['month']=pd.DatetimeIndex(df[u'Fecha']).month
df.boxplot(column=u'CAPE',by=u'month')
```

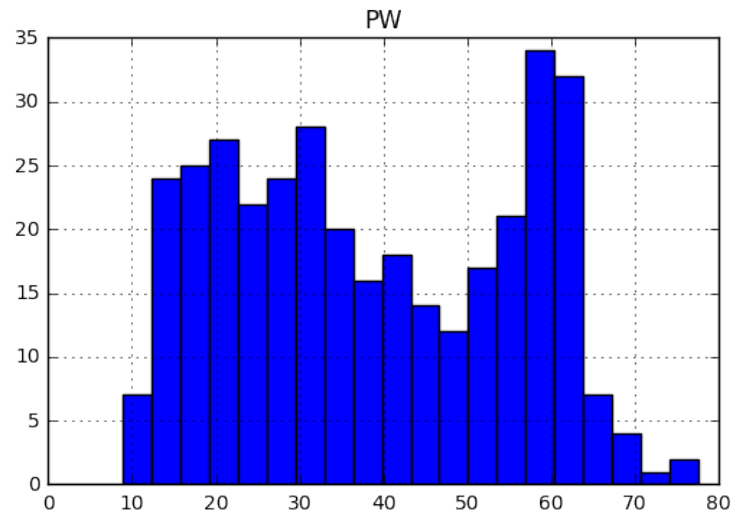
Y con ello, se formó el siguiente gráfico



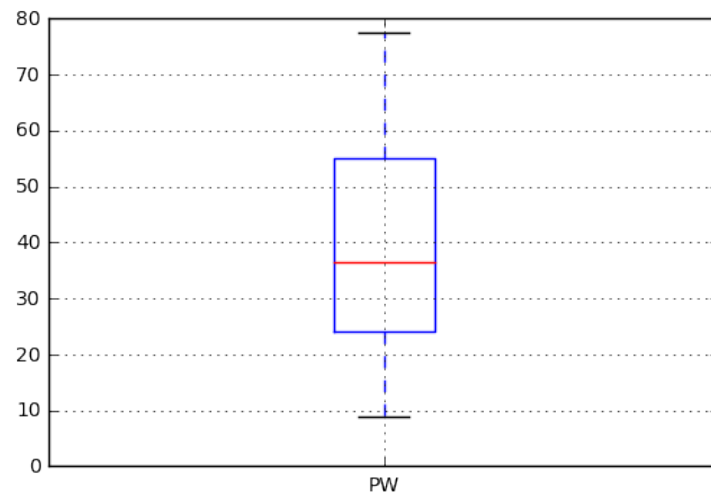
2.4. Análisis de datos de Precipitable Water

Ahora, realizaremos los mismos procedimientos para obtener gráficas para poder observar como se ha comportado durante todo el año en el sitio JuanCarrasco. Comenzando con el histograma obtenido con el siguiente comando

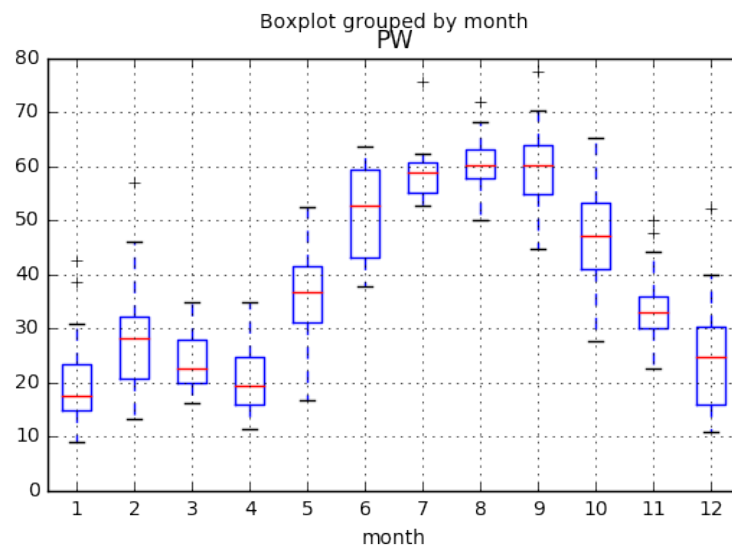
```
df.hist(u'PW',bins=20)
```



Procedemos a la observación de un diagrama de caja de todo el año



Y para finalizar, la comparativa de todos los meses utilizando diagramas de caja



3. Conclusión

Esta actividad resultó bastante entretenida, se que la limpieza de datos automatizada nos será de mucha utilidad debido a que constantemente trabajaremos con una innumerable cantidad de datos. La utilización de Python para realizar gráficos que nos permitan una interpretación de datos mucho más facilitada será de mucha utilidad también.