



UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES

FÍSICA COMPUTACIONAL

---

# Iniciación en Python

---

*Autor:*

Carlos Alí Medina Leal

*Profesor:*

Carlos Lizárraga Celaya

30 de Agosto del 2016

## 1. Un breve resumen...

En esta práctica aprenderemos a usar Python, un lenguaje de programación que nos servirá para generar datos descriptivos, gráficas, etc. de cualquier archivo compatible con Python, como el tipo de archivo que usaremos en esta práctica, el *.csv*.

## 2. Introducción

Python es un lenguaje de programación de alto nivel, de propósito general e interpretado. Diseñado por el científico de la computación de origen Holandés, Guido van Rossum y administrado por la “Python Software Foundation”, que posee una licencia de código abierto, denominada “Python Software Foundation License”, que es compatible con la Licencia pública general de GNU a partir de la versión 2,1,1, e incompatible en ciertas versiones anteriores.

Existen 2 versiones de Python:

Python 2.7.x y Python 3.5.x.

Si se desea instalar Python, se hace a través de cualquiera de las 2 suites existentes:

“Canopy Python.” “Anaconda Python”

Ambas distribuciones contienen las bibliotecas básicas generales que nos interesan:

1. iPython - Interactive Python
2. SciPy - Scientific Python (Mathematics, Science & Engineering).
3. numPy - Numerical Python
4. Matplotlib - Biblioteca para gráficas
5. Pandas - Biblioteca para el análisis de datos.

En este caso, usaremos el Pandas, que nos sirve para analizar datos ágil y rápidamente.

### 3. Desarrollo

En esta práctica usaremos los archivos de la actividad anterior, y los usaremos de ejemplo para analizar los datos.

Comenzamos abriendo una sesión en la notebook, escribiendo “ipython notebook.”<sup>en</sup> la terminal. Se abrirá una ventana en el navegador predeterminado con la notebook de Python, a lo que creamos un archivo y empezaremos escribiendo:

```
mport pandas as pd
```

```
import numpy as np
```

```
import matplotlib as plt
```

```
df = pd.read_csv("/home/usuario/Carpeta/archivo.csv")
```

Que importará la biblioteca de *pandas* y la señala como *pd*, así como *numpy* a *np* y la librería de *Matplot* como *plt*.

Después, aplicaremos los siguientes comandos:

1. `df.head(20),`
2. `df.describe()`
3. `df.apply(lambda x: sum(x.isnull()),axis=0)`

El primer comando te muestra los primeros 20 renglones del archivo, dándote una idea de la estructura del archivo y las columnas de datos que vendrán.

El segundo comando te muestra una descripción resumida de los datos, mostrándote el número de datos que tiene, los cuartiles, los mínimos y los máximos.

El tercer comando te encuentra los valores faltantes en tu archivo contándolos por columna.

Puede que en el archivo hagan falta cientos o miles de datos, estos datos no se pueden reemplazar con ceros por ejemplo, porque modificaría los datos descriptivos de el archivo, así que se aplica el siguiente comando:

```
df_clean = df.dropna()
```

Este comando borra la fila en la que está el dato faltante, y así no se alteren las características descriptivas de los datos.

Ahora, se compara el nuevo resultado con:

```
df_clean.describe()
```

Y finalmente, usaremos los siguientes comandos:

1. `df.columns` (cuidando los nombres de las columnas exactamente)
2. `df_clean['CAPE'].hist(bins=100)`
3. `df_complete.boxplot(column='CAPE')` (anual y mensual)

El primer comando, te nombra exactamente el nombre de cada columna, el cual usaremos a continuación.

El segundo comando, te hace un histograma de los datos que tienes específicamente en la columna nombrada (Ojo: escribir exactamente el nombre de la columna como es mostrada en el primer comando).

El tercer comando sirve para crear un diagrama de caja de los datos de la columna especificada.

Y para concluir, todos estos comandos son sólo una rápida mirada a lo que podemos lograr con Python/Pandas, dándonos un manejo bastante fácil, fluído y rápido en el momento de analizar datos.

## 4. Bibliografía

University of Wyoming/ College of Engineering/ Department of Atmospheric Science : [http : //weather.uwyo.edu/upperair/sounding.html](http://weather.uwyo.edu/upperair/sounding.html), Recuperado el 14 de Septiembre del 2016.

Pydata/ Python Data Analysis Library : [http : //pandas.pydata.org/pandas-docs/stable/tutorials.html](http://pandas.pydata.org/pandas-docs/stable/tutorials.html), Recuperado el 14 de Septiembre del 2016.