



# Introducción al Manejo de Datos con Python

Oscar A. Esquivel Flores

Escuela de Gobierno y Transformación Pública

## PRESENTACIÓN

El lenguaje de programación Python no es nuevo, tiene un largo y exitoso camino recorrido. Quienes han adoptado y aplicado este lenguaje coinciden en varias de sus ventajas: es una alternativa libre, clara, amigable y divertida a lenguajes altamente técnicos y/o con enfoque al análisis estadístico. Python y su entorno de paquetes ofrecen una potente flexibilidad para desarrollar cualquier tipo de aplicaciones, además de presentar una curva de aprendizaje sutil y facilita la productividad inmediata. Desde aplicaciones científicas hasta pequeños scripts, Python puede considerarse como una navaja suiza para automatizar tareas, manejar, visualizar y analizar datos, además de enfrentar los retos de la ciencia pura y dura.

## Índice

<b>1. Objetivos y propósitos</b>	<b>1</b>
1.1. Objetivo . . . . .	1
1.2. Requisitos . . . . .	1
1.3. Enfoques . . . . .	1
<b>2. Contenidos</b>	<b>2</b>
2.1. Python . . . . .	3
2.2. Scipy . . . . .	4
2.3. Pandas . . . . .	5
2.4. Caso de estudio . . . . .	6

# 1. Objetivos y propósitos

## 1.1. Objetivo

El objetivo del curso consiste en que el participante conozca los principios del lenguaje y utilice los paquetes básicos para el cálculo numérico y de manejo de datos, de forma tal que aumenten sus habilidades computacionales y elabore estrategias de solución a problemas comunes que se presentan en su área de experiencia.

El curso abarca los siguientes temas:

- Elementos básicos del lenguaje de programación Python
- Librerías y paquetes del entorno científico de Python.
- El paquete para el manejo y análisis de datos.
- Caso de estudio y/o aplicación.

Estos temas se agruparán en 4 sesiones cada una. Al final del curso Se ofrecerá al participante el código desarrollado en cada una de las sesiones.

## 1.2. Requisitos

- Se asume que el participante tiene conocimientos básicos de algún lenguaje de programación.
- Se solicita que los participantes cuenten con computadora personal en la cual tendrán instalada la versión 5.3 de la distribución Anaconda (<https://www.anaconda.com/download/>) o superior.
- Diseño de un caso de estudio en su área de experiencia en el cual aplicar los conocimientos adquiridos.

La distribución Anaconda 5.3 instalará la versión 3.7 de Python y funcionará como un entorno de desarrollo e instalación de paquetes del lenguaje. Se sugiere el uso de un sistema operativo tipo unix (linux, mac) aunque no es indispensable.

## 1.3. Enfoques

El **enfoque didáctico** del diplomado está orientado al logro de los siguientes aspectos:

- *Teórico*: Conocer los conceptos básicos relativos al lenguaje de programación Python v.3.5+.
- *Práctico*: Adquirir habilidades computacionales por medio del lenguaje de programación Python.

- *Analítico*: Aplicar algoritmos básicos desarrollados en Python para analizar información.

El uso de estas herramientas computacionales coadyuvará al participante a:

- Identificar en el software libre herramientas para el desarrollo tecnológico-computacional contemporáneo.
- Resolver problemas por medio del lenguaje de programación Python y su entorno científico.
- Plantear y desarrollar soluciones a problemas en su área de experiencia de manera rápida y clara.
- Adquirir nuevas habilidades mediante el uso de herramientas computacionales actuales.
- Desarrollar un pensamiento lógico, reflexivo, crítico auxiliado del uso de nuevas tecnologías.
- Ampliar sus habilidades computacionales.

## 2. Contenidos

El curso está organizado de la siguiente manera:

- La **Sesión I** proporciona una visión general del lenguaje de programación Python.
- La **Sesión II** profundiza en el entorno científico de Python. Se revisan las herramientas mas representativas de este ecosistema.
- La **Sección III** introduce al análisis de datos por medio del uso del paquete *Pandas*.
- En la **Sesión IV** se aborda un caso de estudio y se aplican los conocimientos adquiridos.

La propuesta de contenidos organizados por temática, aprendizajes y estrategias didácticas es la siguiente:

## 2.1. Python

Sesión I		
<b>Propósito:</b> al finalizar la sesión el participante tendrá una visión general del lenguaje de programación Python y las ventajas de su uso.		
Temática	Aprendizajes	Estrategias
Conceptos básicos de la programación en Python: <ul style="list-style-type: none"> <li>■ Instalación de Python</li> <li>■ Variables y operaciones</li> <li>■ Operadores</li> <li>■ Funciones</li> </ul> “Codeando” con Python: <ul style="list-style-type: none"> <li>■ Tipos de datos</li> <li>■ Listas, cadenas</li> <li>■ Control de flujo</li> <li>■ Módulos, atributos y métodos</li> <li>■ Archivos</li> </ul>	El participante: <ul style="list-style-type: none"> <li>■ Conoce las distribuciones e instalación del ecosistema Python</li> <li>■ Compara los tipos de ambientes de trabajo de Python</li> <li>■ Comprende la sintaxis básica de Python</li> <li>■ Usa, define e implementa funciones <i>Pythonicas</i></li> <li>■ Diseña e implementa algoritmos utilizando los tipos de datos estudiados</li> <li>■ Conoce los módulos, propiedades y métodos de objetos</li> <li>■ Conoce la manera de abrir, leer y escribir en archivos</li> </ul>	<ul style="list-style-type: none"> <li>■ Exposición de las distribuciones de Python y la correspondiente instalación.</li> <li>■ Propiciar una discusión colectiva sobre las ventajas del lenguaje después de explorar su sintaxis básica.</li> <li>■ Elaborar distintos ejercicios de programación en los que se utilicen los elementos básicos del lenguaje y ejercitar el diseño de funciones para resolver y automatizar acciones variadas.</li> </ul>

La bibliografía y referencias sugeridas para esta sección son:

- *Python*, recuperado el 7 de enero de 2018 de <https://www.python.org/>
- *Python Docs*, recuperado el 10 de enero de 2018 de <https://docs.python.org/3/>
- *The Python Wiki*, recuperado el 7 de enero de 2018 de <https://wiki.python.org/moin/BeginnersGuide/Overview>
- BEAZLEY, D. y Jones, B., *Python Cookbook*. 3a ed. New York, O’Reilly, 2013.
- BAHIT, Eugenia, *Curso: Python para Principiantes*, Buenos Aires, Safe Creative, 2012.
- BARRY, Paul, *Head First Python*. Boston, O’Reilly, 2011.
- DOWNEY, Allen, *Think Python. How to Think Like a Computer Scientist*, Green Tea Press, 2013.
- GONZALEZ, Raúl, *Python para todos*, Barcelona, Creative Commons.
- GRIES, P., et. al., *Practical Programming. An Introduction to Computer Science Using Python 3*. 2a ed. The Pragmatic Bookshelf, 2013.
- LUTZ, Mark, *Python Pocket Reference. Python in your Pocket*, O’Reilly, 2014.

## 2.2. Scipy

Sesión II		
<b>Propósito:</b> al finalizar la sesión el participante conocerá y utilizará las librerías científicas del ecosistema <i>SciPy</i> mediante la codificación de segmentos de código con Ipython notebooks.		
Temática	Aprendizajes	Estrategias
Python en la ciencia: <ul style="list-style-type: none"> <li>■ Jupyter-Notebooks</li> <li>■ Python Científico</li> <li>■ NumPy</li> <li>■ Matplotlib</li> <li>■ Seaborn</li> </ul>	El participante: <ul style="list-style-type: none"> <li>■ Conoce Python interactivo con Jupyter-Notebooks <i>notebooks</i></li> <li>■ Conoce los paquetes y librerías diseñados para el cómputo científico en Python</li> <li>■ Utiliza los elementos más importantes de la librería <i>NumPy</i></li> <li>■ Estudia los elementos básicos de <i>Matplotlib</i></li> <li>■ Diseña e implementa código para obtener gráficas en 2D y 3D</li> <li>■ Conoce los elementos sustanciales de la librería <i>Seaborn</i></li> <li>■ Elabora gráficos con <i>Seaborn</i></li> </ul>	<ul style="list-style-type: none"> <li>■ Realizar ejercicios utilizando el entorno Jupyter-Notebooks</li> <li>■ Elaborar un mapa de librerías dentro del entorno científico de Python (<i>SciPy</i>)</li> <li>■ Describir las características principales de diversas librerías para el cómputo científico</li> <li>■ Practicar el cálculo de operaciones matriciales utilizando <i>Numpy</i></li> <li>■ Ejemplificar con segmentos de código la elaboración de gráficos con <i>Matplotlib</i></li> <li>■ Exposición de las funcionalidades de librería <i>Seaborn</i></li> </ul>

La bibliografía y referencias sugeridas para esta unidad son:

- *IPython. Interactive Computing*, consultado el 7 de enero de 2018 de <https://www.python.org/>
- *Jupyter. The Jupyter Notebook*, consultado el 10 de enero de 2018 de <http://jupyter.org/>
- *Matplotlib*, consultado el 10 de enero de 2017 de <https://matplotlib.org>
- *Scipy*, consultado el 10 de enero de 2018 de <http://www.scipy.org/>
- *Numpy*, consultado el 10 de enero de 2018 de <http://www.numpy.org/>
- *Python Scientific Lecture Notes (Spanish translation)*, consultado el 7 de septiembre de 2018 de <https://claudiovz.github.io/scipy-lecture-notes-ES/index.html>
- FANGOHR, H., *Introduction to Python for Computational Science and Engineering*, University of Southampton, 2015.
- LANGTANGEN, H., *Python Scripting for Computational Science*, 3rd edition, Springer, 2007.

- LANGTANGEN, H., *A primer on Scientific Programming with Python*, 2nd edition, Springer, 2011.
- TOSI, Sandro, *Matplotlib for Python Developers*, Packt, U.K., 2009.

## 2.3. Pandas

Sesión III		
<b>Propósito:</b> al finalizar la sesión el participante comprenderá y aplicará las funcionalidades del paquete <i>Pandas</i> para el manejo de datos		
Temática	Aprendizajes	Estrategias
Python en la ciencia: <ul style="list-style-type: none"> <li>■ Pandas</li> <li>■ Lectura de archivos</li> <li>■ Creación de DataFrames</li> <li>■ Indexado</li> <li>■ Análisis estadístico básico</li> <li>■ Graficar datos</li> </ul>	El participante: <ul style="list-style-type: none"> <li>■ Conoce el paquete <i>Pandas</i> y sus funcionalidades para el manejo de datos</li> <li>■ Adquiere habilidades para extraer y organizar datos con <i>Pandas</i></li> <li>■ Crea estructuras de datos DataFrame</li> <li>■ Selecciona y extrae información de DataFrames</li> <li>■ Desarrolla operaciones estadísticas con los datos</li> <li>■ Grafica datos contenidos en DataFrames</li> </ul>	<ul style="list-style-type: none"> <li>■ Exposición de los fundamentos teóricos del paquete.</li> <li>■ Extraer información relevante de distintos tipos de archivos de datos</li> <li>■ Elegir información relevante para hacer operaciones estadísticas básicas sobre los datos.</li> <li>■ Propiciar una discusión colectiva sobre la utilización de esta librería en casos de estudio particulares.</li> </ul>

La bibliografía sugerida para esta unidad es:

- *Pandas. Python Data Analysis Library*, consultado el 10 de enero de 2018 de <http://pandas.pydata.org>
- McKINNEY, W., *Python for Data Analysis*, O'Reilly, 2013.
- NELLI, F., *Python Data Analytics*, Apress, 2015.
- MÜLLER, A., Guido, S., *Introduction to Machine Learning with Python*, O'Reilly, 2016.

## 2.4. Caso de estudio

Sesión IV		
<b>Propósito:</b> al finalizar la sesión el participante diseñó una estrategia de solución a un caso de estudio particular en las que utiliza las herramientas computacionales del entorno científico de Python.		
Temática	Aprendizajes	Estrategias
<ul style="list-style-type: none"><li>■ Definición de caso de estudio</li><li>■ Diseño de estrategia de solución</li><li>■ Selección de herramientas del lenguaje a utilizar</li><li>■ Codificación</li><li>■ Conclusiones cualitativas y cuantitativas</li></ul>	El participante: <ul style="list-style-type: none"><li>■ Define un caso de aplicación</li><li>■ Elabora alguna estrategias de solución con base en los conocimientos adquiridos</li><li>■ Desarrolla el código de la solución planteada</li><li>■ Valora las habilidades adquiridas</li><li>■ Evalúa la eficacia de python y su ecosistema científico como herramienta para analizar datos</li><li>■ Elabora conclusiones cualitativas del análisis cuantitativo de datos</li></ul>	<ul style="list-style-type: none"><li>■ Se crean grupos de trabajo para elegir un problema particular y diseñar distintas soluciones computacionales.</li><li>■ Trabajo en conjunto para llevar acabo el flujo de tareas.</li><li>■ Evaluar de manera grupal los resultados obtenidos.</li><li>■ Fomentar la discusión colectiva para llegar a conclusiones.</li></ul>

La bibliografía sugerida para esta unidad es:

- *Kaggle: Your Home for Data Science*, consultado el 10 de enero de 2018 de <https://www.kaggle.com/>
- OJEDA, T., *et al.*, *Practical Data Science Cookbook*, Packt Publishing, 2014.
- BOWLES, M., *Machine Learning in Python. Essential Techniques for Predictive Analysis*, Wiley, 2015.
- AVILA, J., Hauck, T., *Scikit-learn Cookbook*, 2nd. editon, Packt, 2017.