



Bloque 1.

Tema 3:
Manipulación y
análisis de datos
(NumPy y Pandas)

Para importar y analizar los datos de tu trabajo, negocio, empresa o industria como usuario o data scientist, es muy importante que cuentes con un paquete o herramienta que te facilite su manipulación (ya sea para convertirlos, modificarlos, expresarlos, agruparlos, graficarlos, etc.) para llegar al objetivo planteado.

Por lo tanto, en este tema te hablaremos sobre dos bibliotecas de la familia **Python** que te ayudarán a manipular y analizar los datos e información que desees de manera rápida y eficiente: NumPy y Pandas.

NumPy



Es una biblioteca de alto nivel de extensión con acceso libre y de licencia BSD, que fue desarrollada en el 2005 por Jim Hugunin. Asimismo, esta herramienta está focalizada en funciones matemáticas, la cual define matrices numéricas, tipos de matrices y las operaciones básicas realizadas en ellas.

Algunas de sus principales características son las siguientes (Numpy, s.f.):

- Biblioteca de matrices multidimensionales.
- Funciones sofisticadas.
- Escrito principalmente en C.
- Capaz de utilizar algebra lineal.
- Eficiente contenedor multidimensional de datos genéricos.
- Puede definir tipos de datos arbitrarios, lo que permite su rápida integración a una amplia variedad de bases de datos.

Para comenzar necesitas instalar el programa, conocer las bases y comprender por medio de ejemplos, por lo que te recomendamos leer el siguiente tutorial de instalación oficial:

Los siguientes enlaces son externos a la Universidad Tecmilenio, al acceder a estos considera que debes apegarte a sus términos y condiciones.



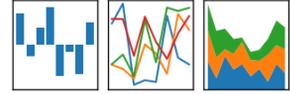
Numpy. (s.f.). *Quickstart tutorial*. Recuperado de <https://bit.ly/2Rrcjbb>

Pandas

Es un proyecto centrado en datos, es decir, en la importación y análisis de la información.

pandas

$$y_{it} = \beta^t x_{it} + \mu_i + \epsilon_{it}$$



Desarrollado en el 2008 por Wes McKinney, Pandas (que actualmente es patrocinado por NumFOCUS) cuenta con licencia BSD y es de código abierto. Asimismo, este se considera como una biblioteca de software de alto nivel con rendimiento de clase mundial, ya que proporciona estructuras y herramientas de análisis de datos que son fáciles de usar para el lenguaje de programación Python.

"Pandas es una biblioteca de software escrita como extensión de NumPy para manipulación y análisis de datos para Python. En particular, ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales" (Universidad de Alcalá, 2019).

Las multifacetas de Pandas la convierten en una herramienta muy popular de la familia Python, además de que su sencillez la hace muy útil para principiantes de programación. Algunas de sus características son las siguientes:

- Se basa en NumPy y Matplotlib, lo que facilita la manipulación y la visualización de la información.
- Complementa a Python en cuanto al análisis y a la modelación de datos, cubriendo así un área de oportunidad importante de dicha herramienta.
- Es compatible, por lo que se puede combinar con otras herramientas y bibliotecas de Python.
- Su estructura de datos clave es *DataFrame*, la cual permite almacenar y manipular datos tabulados en filas (observaciones) y columnas (variables).

Para instalarlo, te recomendamos leer la siguiente guía de instalación:



Pandas. (s.f.). *Installation*. Recuperado de <https://bit.ly/35itdMH>

Para saber más sobre **el manejo de la herramienta según su creador Wes McKinney**, revisa el siguiente video:



PyData. (2019, 17 de septiembre). *Wes McKinney: pandas in 10 minutes | Walkthrough* [Archivo de video]. Recuperado de <https://bit.ly/364lJgx>



Referencias

- NumPy. (s.f.). *NumPy*. Recuperado de <https://numpy.org/>
- Universidad de Alcalá. (s.f.). *Pandas: Herramienta básica para el Data Science en Python*. Recuperado de <https://www.master-data-scientist.com/pandas-herramienta-data-science/>

Para expandir tu conocimiento, te recomendamos los siguientes recursos:

Los siguientes enlaces son externos a la Universidad Tecmilenio, al acceder a estos considera que debes apegarte a sus términos y condiciones.

Podcasts

Para saber más sobre **NumPy**, te recomendamos escuchar los siguientes podcasts:

- Palen, B. (8 de febrero de 2011). *RCE 48: NumPy* [Audio podcast]. Recuperado de <https://bit.ly/2O5aPky>
- PodBean. (18 de enero de 2019). *Episode 8: Python NumPy – Edureka* [Audio podcast]. Recuperado de <https://bit.ly/2qym8c7>

Para saber más sobre **Pandas**, te recomendamos escuchar los siguientes podcasts:

- Podcast.__init__. (26 de febrero de 2017). *Pandas with Jeff Reback - Episode 98* [Audio podcast]. Recuperado de <https://bit.ly/339fGpl>
- Kennedy, M. (21 de febrero de 2019). *Episode #200: Escaping Excel Hell with Python and Pandas* [Audio podcast]. Recuperado de <https://bit.ly/2O5u2md>

MOOC

Para saber más sobre **NumPy y Pandas**, te invitamos a cursar los siguientes MOOC:

- Udemy. (s.f.). *Learn Core Python, Numpy and Pandas*. Recuperado de <https://bit.ly/2O6rKmY>
- Udemy. (s.f.). *Learn Data Analysis using Pandas and Python (Module 2/3)*. Recuperado de <https://www.udemy.com/course/learn-data-analysis-using-pandas-and-python/>

Para saber más sobre **NumPy**, te invitamos a cursar el siguiente MOOC:

- Udemy. (s.f.). *Learn NumPy Fundamentals (Python Library for Data Science)*. Recuperado de <https://www.udemy.com/course/python-numpy-fundamentals/>

La obra presentada es propiedad de ENSEÑANZA E INVESTIGACIÓN SUPERIOR A.C. (UNIVERSIDAD TECMILENIO), protegida por la Ley Federal de Derecho de Autor; la alteración o deformación de una obra, así como su reproducción, exhibición o ejecución pública sin el consentimiento de su autor y titular de los derechos correspondientes es constitutivo de un delito tipificado en la Ley Federal de Derechos de Autor, así como en las Leyes Internacionales de Derecho de Autor.

El uso de imágenes, fragmentos de videos, fragmentos de eventos culturales, programas y demás material que sea objeto de protección de los derechos de autor, es exclusivamente para fines educativos e informativos, y cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por UNIVERSIDAD TECMILENIO.

Queda prohibido copiar, reproducir, distribuir, publicar, transmitir, difundir, o en cualquier modo explotar cualquier parte de esta obra sin la autorización previa por escrito de UNIVERSIDAD TECMILENIO. Sin embargo, usted podrá bajar material a su computadora personal para uso exclusivamente personal o educacional y no comercial limitado a una copia por página. No se podrá remover o alterar de la copia ninguna leyenda de Derechos de Autor o la que manifieste la autoría del material.