

Módulo: B1 Procesamiento de Datos con Python 2020

Objetivo general
Al finalizar el módulo serás capaz de utilizar Python para adquirir datos ya sea de archivos locales, APIs o Bases de Datos con el objetivo de limpiarlos, manejarlos, explorarlos para dejarlos de manera que queden en condiciones óptimas para su futuro análisis y visualización.
Temario
<ul style="list-style-type: none"> ● Sesión 1: Fundamentos de Python - Conoce las bases del lenguaje Python. ● Sesión 2: Estructuras de datos y funciones - Conoce las principales estructuras de datos y aprende a reutilizar código. ● Sesión 3: Programación funcional, operadores lógicos y funciones lambda - Identifica las estructuras de datos y las funciones necesarias para el procesamiento de datos. ● Sesión 4: Pandas y Análisis Exploratorio de Datos - Utiliza la librería de Data Analysis más importante para adquirir datos desde diferentes fuentes y homogeneizarlos y procesarlos. ● Sesión 5: Funciones vectorizadas y limpieza de datos - Utiliza diferentes fuentes para realizar proyectos de procesamiento de datos. ● Sesión 6: APIs, automatización y concatenación de DataFrames - Usa las herramientas básicas de exploración analiza a profundidad una base de datos. ● Sesión 7: Transformación, filtración y ordenamiento - Identifica los diferentes tipos de datos de un dataset y limpia o reordena estos para un mejor análisis y procesamiento. ● Sesión 8: Bases de datos, merge y agrupaciones - Utiliza métodos avanzados de reestructuración y división de datos para después seguir con pasos posteriores (análisis y visualización) del dataset.
Checkpoint
<p>Al finalizar este módulo serás capaz de aplicar los requerimientos a tu proyecto personal y cumplir con las actividades que se enlistan a continuación. Estos requerimientos y actividades se cumplirán dependiendo del alcance de tu proyecto; sin embargo, al final del módulo serás capaz de aplicar cualquiera de estos requerimientos y actividades para cualquier proyecto.</p> <p>Durante el Checkpoint, el experto te dará retroalimentación y orientación sobre cada entregables que hagas. Recuerda desarrollar y plantear de manera óptima tu proyecto para hacer que el proceso de retroalimentación y orientación sea más efectivo.</p> <p>Para esta sesión, mapea con la lista lo que estará incluido en tu proyecto y dáselo</p>

al experto para que te pueda guiar durante el desarrollo de este. Cualquier duda puedes consultarle al experto.

Sesión 1: Fundamentos de Python

- ☐ Identificar un problema (el primer paso en todo proyecto de ciencia de datos).

Sesión 2: Estructuras de datos y funciones

- ☐ Realizar planteamiento de preguntas (el paso #2 en todo proyecto de ciencia de datos).

Sesión 3: Programación funcional, operadores lógicos y funciones lambda

- ☐ Coleccionar datos.
- ☐ Realizar recolección de datos para responder a preguntas.

Sesión 4: Pandas y Análisis Exploratorio de Datos

- ☐ Leer un dataset en un DataFrame de pandas.
- ☐ Realizar Análisis Exploratorio de Datos básico de un conjunto de datos.

Sesión 5: Funciones vectorizadas y limpieza de datos

- ☐ Limpiar un dataset de NaNs.
- ☐ Reindexar, si es necesario.
- ☐ Renombrar columnas si es necesario.
- ☐ Experimentar la aplicación de agregaciones para explorar un dataset.

Sesión 6: APIs, automatización y concatenación de DataFrames

- ☐ Usar APIs.
- ☐ Realizar peticiones HTTP.
- ☐ Automatizar procesos de exploración y limpieza usando for loops.
- ☐ Proteger el código contra errores usando try except.

Sesión 7: Transformación, filtración y ordenamiento

- ☐ Utilizar transformaciones de datos para limpiar y preparar un dataset para análisis.
- ☐ Realizar filtraciones para obtener conjuntos de datos.
- ☐ Reordenar datos para ver el conjunto desde diferentes perspectivas.

Sesión 8: Bases de datos, merge y agrupaciones

- ☐ Crear un Jupyter Notebook para mostrar con claridad el proceso de procesamiento de datos.
- ☐ Identificar y analizar un dataset para crear visualizaciones o análisis futuros.