

FUNDAMENTOS DE PYTHON PARA LA CIENCIA DE DATOS

PRESENTACIÓN

El curso Fundamentos de Python para la Ciencia de Datos introduce el uso de Python como herramienta práctica para el análisis de datos. El participante aprenderá a preparar su entorno de trabajo, dominará la sintaxis esencial del lenguaje y aplicará librerías estándar del ecosistema de ciencia de datos (NumPy, Pandas, Matplotlib y Seaborn) para cargar, limpiar, transformar, explorar y visualizar datos. Al finalizar, el estudiante integrará lo aprendido en un proyecto básico con comunicación clara de resultados usando Markdown, priorizando un enfoque de “aprender haciendo”.

PÚBLICO OBJETIVO

- Estudiantes universitarios o técnicos que deseen iniciar una ruta en Ciencia de Datos con Python.
- Profesionales de áreas como ingeniería, administración, economía, estadística o afines que necesiten analizar datos en su trabajo.
- Personas con bases de programación que quieran migrar a analítica con Python, especialmente con Pandas y visualización.

CARACTERÍSTICAS

- HORAS TOTALES: 24
- SESIONES: De acuerdo al horario
- MATERIAL DEL CURSO: Formato digital
- CERTIFICADO DEL CURSO: Si cumple con los criterios de éxito

REQUISITOS

- Conocimientos básicos de programación: variables, condicionales, bucles y funciones.
- Nociones básicas de manejo de datos: tablas, conceptos generales de base de datos; no es indispensable SQL.

METODOLOGÍA

- Sesiones prácticas con ejercicios guiados y retos cortos.
- Trabajo colaborativo en actividades de análisis y discusión de resultados.
- Mini-proyecto integrador al final (estructura, limpieza, EDA y visualización).

MERCADO LABORAL

El dominio de Python aplicado al análisis de datos es una competencia transversal solicitada en múltiples roles. Este curso aporta bases para desempeñarse o evolucionar hacia perfiles como:

- Analista de Datos (Data Analyst): limpieza, transformación, análisis exploratorio y visualización.
- BI / Reporting Analyst: preparación de datasets y comunicación de hallazgos con gráficos.
- Data Scientist Junior (ruta inicial): fundamentos del stack (NumPy/Pandas/EDA) como base para modelado posterior.
- Analista de Riesgo / Operaciones / Marketing / Finanzas con enfoque en datos: automatización de análisis recurrentes y toma de decisiones.

Sectores típicos donde se demanda: banca/finanzas, retail, telecom, educación, logística, salud, sector público y startups orientadas a productos digitales, porque todas tienen datos y problemas.

EVALUACIÓN

El participante será evaluado de la siguiente manera:

1. Cuatro prácticas calificadas: PC1, PC2, PC3 y PC4
2. Un examen final
3. Promedio de prácticas (PP): Se promedian las 3 mejores notas de prácticas
4. Promedio Final (PF):

$$PF = \frac{PP + EF}{2}$$

5. Criterio de éxito: Asistencia mínima 80% y nota mínima aprobatoria 14

CONTENIDO

MOD	CONTENIDO	DESCRIPCION
1	FUNDAMENTOS DE CIENCIA DE DATOS: del concepto a la práctica con Python <ul style="list-style-type: none">▪ ¿Qué es Ciencia de Datos? Mercado laboral▪ Instalación y configuración del marco de trabajo▪ Primeros pasos en la Ciencia de Datos con Python	El participante comprenderá el alcance y aplicaciones de la Ciencia de Datos en el contexto profesional actual. Configuraré su entorno de desarrollo (Anaconda, Jupyter Notebook) y ejecutará sus primeros scripts en Python, familiarizándose con la sintaxis básica y el flujo de trabajo en notebooks.
2	ESTRUCTURAS FUNDAMENTALES DE PYTHON: sintaxis y control de flujo <ul style="list-style-type: none">▪ Operaciones y variables▪ Listas, tuplas y diccionarios▪ Estructuras selectivas de control▪ Funciones	El participante dominará los tipos de datos nativos de Python y las estructuras de control esenciales. Será capaz de manipular colecciones de datos (listas, tuplas, diccionarios), implementar lógica condicional e iterativa, y crear funciones reutilizables para organizar código de manera modular y eficiente.
3	COMPUTACIÓN NUMÉRICA CON NUMPY: arreglos multidimensionales <ul style="list-style-type: none">▪ Librerías: instalación e importación▪ Arreglos con Numpy (array & ndarray)▪ Atributos, métodos y clases	El participante aprenderá a trabajar con arreglos multidimensionales usando NumPy, la librería fundamental para cálculo numérico en Python. Comprenderá la diferencia entre listas y arrays, aplicará operaciones vectorizadas eficientes y explorará los atributos y métodos principales de ndarray para manipulación de datos numéricos.
4	ANÁLISIS DE DATOS ESTRUCTURADOS CON PANDAS: Series y DataFrames <ul style="list-style-type: none">▪ DataSeries▪ DataFrame▪ Lectura de diferentes fuentes de datos (offline, online)▪ Introducción al análisis de datos con Pandas	El participante dominará las estructuras fundamentales de Pandas (Series y DataFrame) para trabajar con datos tabulares. Cargará datos desde múltiples fuentes (CSV, Excel, APIs, bases de datos), realizará operaciones básicas de selección, filtrado e indexación, y ejecutará transformaciones iniciales para preparar datasets para análisis.

5	VISUALIZACIÓN DE DATOS: comunicando insights con Matplotlib y Seaborn <ul style="list-style-type: none"> ▪ Realizar gráficos con Matplotlib ▪ Personalización de gráficos ▪ Gráficos con Seaborn 	<p>El participante creará visualizaciones efectivas para comunicar patrones y hallazgos en los datos. Generará gráficos estadísticos con Matplotlib (líneas, barras, dispersión, histogramas), personalizará elementos visuales (colores, etiquetas, leyendas) y producirá visualizaciones avanzadas con Seaborn para análisis exploratorio y presentación profesional de resultados.</p>
6	PROYECTO INTEGRADOR: aplicando el ciclo completo de análisis de datos <ul style="list-style-type: none"> ▪ Presentación del proyecto y estructuración básica ▪ Tratamiento de datos (filtro, nulos, agrupamiento) ▪ Introducción al análisis exploratorio de datos ▪ Mostrando resultados con Markdown 	<p>El participante integrará todas las competencias desarrolladas en un proyecto real de análisis de datos. Aplicará técnicas de limpieza y tratamiento de datos (manejo de valores nulos, filtros, agrupaciones), realizará análisis exploratorio de datos (EDA) completo, generará visualizaciones relevantes y documentará todo el proceso y hallazgos usando Markdown en Jupyter Notebook, simulando un flujo de trabajo profesional.</p>