

CIENCIA DE DATOS EN PYTHON PARA O&G

Carga Horaria

28 horas

Introducción

El lenguaje de programación Python se desarrolló para proporcionar una forma de desarrollar código que sea fácil de programar y entender. Si bien Python contiene las mismas estructuras básicas que otros lenguajes, también ofrece funcionalidades, facilita la programación de la que especialistas de diferentes disciplinas pueden beneficiarse. Este curso guiará a los estudiantes para crear estructuras de programación básicas como reglas de decisión y bucles. Se focaliza en ejemplos y en aplicaciones útiles. Luego pasará a temas más avanzados como estructuras de datos únicas de Python, como tuplas y diccionarios. Crear funciones específicas y utilizar librerías existentes para cálculos específicos y visualización de datos.

Objetivo

Aprender a programar en lenguaje Python con foco aplicado en las necesidades más habituales para los profesionales de la industria. Los participantes podrán aplicar Python para resolver problemas específicos de su disciplina que requieran del análisis de datos. Quedarán además capacitados para realizar el curso de Ciencia de Datos Avanzado para O&G donde se resolverán problemas complejos de industria utilizando modelos de aprendizaje automático

A quién esta dirigido

Aplica a ingenieros, geocientistas y otros profesionales de la industria (de áreas como Finanzas, Negocios y RRHH) experimentados en resolución de problemas por medio de técnicas matemáticas, entendiendo que el análisis avanzado de datos cruza muchas de las disciplinas profesionales en las que puedan desarrollarse.

Programa**Módulo 1**

- Por qué programar
- Instalar y usar Python
- Variables y Expresiones.
- Sintaxis. Strings, manipulación de strings. Ejercicios.
- Archivos: Lectura y escritura de archivos. Ejercicios escribir programas que lean y procesan datos reales.
- Listas: Se introducen variables más complejas comenzando con las listas donde se pueden guardar más de un valor en una sola variable utilizando un índice.
- Diccionarios: Introducción a diccionarios en Python usando key y valor en pares.
- Tuplas: introducción a las tuplas como una versión simplificada de las listas. Uso de tuplas en conjunción con diccionarios.
- Ejemplos con set de datos reales

Módulo 2

- Reglas de decision If-else-elif
- Python while loops
- Python for loops. Ejemplos.
- Funciones. Escribir funciones definidas por el usuario. Incluyendo: Como definir una función, incorporar parámetros, como devolver datos de una función, variables locales y globales.

Módulo 3

- Librerías.
- Introducción a Pandas. Por qué usar pandas para trabajar con grandes volúmenes de datos.
- Escribir y leer archivos (Excel, csv, etc)
- Funcionalidades de Python Pandas: filtrado de datos, agrupado, operación entre columnas (sort, filter, groupby). Ejemplo con set de datos real.

Módulo 4

- Conceptos avanzados de funciones: parámetros opcionales, y funciones lambda.
- Visualización de datos: Conceptos básicos de visualización de datos, mejores prácticas.
- Librería Matplotlib: Historia y Arquitectura de matplotlib. Gráficos básicos de esta librería
- Tipos de gráficos. Gráficos areales, histogramas, gráficos de barra, de torta, boxplots, scatter plots y gráficos de burbujas y como crearlos usando matplotlib. Todos los conceptos se ilustrarán de forma teórica y práctica utilizando un set de datos real.

Importante: Los participantes deben llevar sus laptops al curso. Es requisito técnico para la participación que la computadora de cada uno tenga instalados Python y Jupyter Notebook. Los instructores podrán guiarlos vía mail previo al inicio del curso.

Instructores**Julio Rodríguez Martino**

Julio cuenta con más de 25 años de experiencia en análisis de datos. Comenzó su carrera como físico experimental, trabajando en 2 grandes colaboraciones internacionales enfocadas en la construcción de experimentos de astrofísica de partículas. En la industria, desarrolló análisis y visualizaciones de datos para Inteligencia de Negocios. Su trabajo fue evolucionando hacia la Ciencia de Datos y el desarrollo de modelos predictivos de Machine Learning. Fue responsable de construir y liderar equipos de Ciencia de Datos y Machine Learning para resolver problemas de negocio en las industrias de Gas y Petróleo, Banca, Finanzas, Auditoría, Logística y Energías Renovables. Actualmente se desempeña como COO de Infolytics.

Germán Malfatti

Doctor en Física por la UNLP, Magíster en Física por el Instituto Balseiro, Licenciado en Física por el Instituto Balseiro. Actualmente es Científico de Datos Infolytics, donde se especializa en modelado con algoritmos de machine learning e ingeniería de datos. Previamente trabajó en IT en los sectores de administración de redes y desarrollo. Durante la carrera de Física se dedicó mayormente a modelos matemáticos de interiores de estrellas de neutrone.

Alfonso Barrios

Alfonso Barrios es Ingeniero Geofísico, MS en Computación y Diplomado en Ciencia de Datos con más de 15 años de experiencia en software y data management aplicados a la industria petrolera. Hoy es Científico de Datos en Infolytics donde aplica sus conocimientos en proyectos de variadas industrias.

para más Información solicitarla a cursos@iapg.org.ar