

Python: Análisis de Datos y Ecosistemas Modernos de Analítica

Tiempo estimado: 40 horas (20 horas teoría / 20 horas proyecto)

Curso enfocado en el aprovechamiento de Python para el análisis de datos, con énfasis en bibliotecas como *pandas* para la manipulación, limpieza, análisis exploratorio y visualización de datos. Incluye una exploración avanzada del manejo eficiente de grandes volúmenes de información mediante DuckDB, así como la orquestación y gestión de pipelines de datos utilizando herramientas como Dagster. También se abordará la automatización de reportes y el manejo de proyectos de datos en contextos de producción. Como parte integral del contenido, se incluyen principios fundamentales de gobernanza de datos, trazabilidad y linaje de datos, aseguramiento de la calidad, cumplimiento normativo y aspectos éticos en la gestión de información. Como proyecto final el estudiante deberá aplicar integralmente los conocimientos en un problema de análisis de datos, considerando gobernanza y calidad.

Módulo 1: Fundamentos de Python Moderno

- Python 3.12, novedades y operaciones
- Entornos de desarrollo aislados y gestión de dependencias (uv)
- Empaquetado y gestión de proyectos (librerías, paquetes y módulos)
- Tipado estático y testeo
- Github code spaces

Módulo 2: Análisis de Datos con Python

- Pandas (limpieza, transformación, manipulación de datos)
- Testeo con datos y su importancia
- Análisis Exploratorio de Datos (estadísticas)

Módulo 3: Gestión de Grandes Conjuntos de Datos y Orquestación

- Tipos de archivos y datos para distribución
- Trabajando con CSV, parquet y bases de datos
- DuckDB: Base de datos para grandes datos
- Dagster: Orquestación de Pipelines y sistemas de datos complejos

Módulo 4: Despliegue, Reportes y Gobernanza de Datos

- Calidad de datos automatizada
- CI/CD Pipelines y su importancia
- Conceptos de Gobernanza de Datos (Linaje, Calidad, Catálogos)

Evaluación del Curso

Componente	Total	Requisitos Adicionales
Proyecto Final	100%	Para aprobar el curso, se requiere una asistencia mínima del 70% a las sesiones. Las inasistencias superiores al 30% deben ser justificadas adecuadamente.

