

Intermedio - Big Data

PROGRAMA DE ESTUDIOS:

PYTHON APLICADO A LA MANIPULACIÓN Y VISUALIZACIÓN DE DATOS



Para empezar el curso:

1

Es recomendable contar con una base lógica de programación y familiaridad con el manejo de datos en hojas de cálculo.

2

El estudiante debe disponer de una computadora con capacidad para instalar la suite Anaconda o utilizar entornos en la nube.

3

Manejo básico de PC para la instalación de librerías y uso de las libretas Jupyter (Jupyter Notebooks).

Información Esencial:

Inversión Pago único de 450.000 Gs.

Inicio 23 de junio del 2026.

Duración 8 semanas (40 horas totales).

Carga Horaria 5 horas semanales: Clases virtuales e interactivas.

Horario Martes y Jueves de 19:00 a 21:30 h.

Modalidad 100% Online / Sincrónico a través de la plataforma EDUCA.

Este curso brinda una formación integral en tecnologías de procesamiento de información, combinando teoría sólida con un componente práctico intensivo.

La metodología se basa en el enfoque "Learning by Doing" y Aprendizaje Basado en Proyectos (PBL). Los estudiantes utilizarán el lenguaje Python y la poderosa librería Pandas para transformar datos crudos en información valiosa. Se explorarán técnicas de limpieza, fusión de datasets, análisis estadístico y la creación de visualizaciones impactantes para la toma de decisiones.

Además, el programa pone especial énfasis en la aplicabilidad real, permitiendo a los estudiantes enfrentarse a desafíos de curación de datos del mundo real. A través de sesiones prácticas, se potencia la pericia técnica y la capacidad analítica para validar hipótesis mediante evidencia estadística,



Objetivos del curso

- Comprender los fundamentos de la programación en Python aplicada específicamente al ecosistema de la ciencia de datos.
- Desarrollar habilidades avanzadas en la limpieza, manipulación y reestructuración de grandes volúmenes de datos con Pandas.
- Aplicar técnicas estadísticas y herramientas de visualización para comunicar hallazgos de manera efectiva y profesional.
- Integrar múltiples fuentes de información mediante técnicas de fusión y combinación de datasets.
- Validar hipótesis de negocio a través de análisis exploratorio y pruebas de significancia estadística.

Objetivos específicos



SINTAXIS

Utilizar las funciones y características de Python orientadas al análisis científico de datos.



ESTRUCTURA

Dominar el uso de Series y DataFrames para la indexación y consulta eficiente de información.



PROCESAMIENTO

Implementar técnicas de agrupación, tablas de resumen y manipulación de fechas en datasets.



VISUALIZACIÓN

Diseñar representaciones gráficas que faciliten la interpretación de métricas y pruebas estadísticas.



Perfil del egresado

El egresado será capaz de gestionar el ciclo completo de preparación de datos, desde la carga inicial hasta el análisis de importancia. Tendrá la competencia para limpiar datos ruidosos, combinar múltiples fuentes de información y generar informes visuales fundamentados, construyendo un portafolio sólido para su carrera profesional en tecnología.

Plantel Docente

El Prof. Alfredo Daniel Martínez Lara es Especialista en MLOps y Ciencia de Datos con más de 15 años de trayectoria en el sector tecnológico, liderando iniciativas críticas de automatización y despliegue de modelos de inteligencia artificial en entornos de producción.

Ha encabezado áreas de estrategia de datos y transformación digital, con un fuerte enfoque en la escalabilidad de soluciones, el diseño de arquitecturas en la nube (AWS) y la optimización de flujos de trabajo de CI/CD para Machine Learning.



Prof. Alfredo Daniel Martínez Lara

Cronograma

Semana	Módulo	Enfoque Principal
Semana 1	Fundamentos con Python	Introducción a Anaconda y Jupyter Notebooks.
Semana 2	Introducción a Pandas	Carga de datos, DataFrames y consultas básicas.
Semana 3	Limpieza de Datos	Manejo de valores nulos y formatos de datos.
Semana 4	Manipulación Avanzada	Indexación detallada y Examen Parcial.
Semana 5	Transformación de Datos	Fusión, combinación y agrupamiento de datasets.
Semana 6	Análisis Estadístico	Distribuciones, muestreo y validación de hipótesis.
Semana 7	Visualización de Datos	Creación de gráficos y dashboards informativos.
Semana 8	Proyecto Final	Proyecto integrador de limpieza y análisis real.

Contenido del curso

Exploración

Uso de entornos interactivos para el análisis de datos y configuración de librerías científicas.

Manipulación

Operaciones de limpieza, transformación de escalas y generación de métricas personalizadas.

Estadística

Aplicación de pruebas de importancia, análisis de distribuciones y muestreo de datos.

Gráficos

Diseño de visualizaciones claras para la presentación de resultados y soporte de hipótesis.

Fundamentos: Principios del análisis de datos, flujo de trabajo en Ciencia de Datos y mejores prácticas de código.

Colecciones: Documentación de Pandas, catálogos de visualización, datasets del mundo real y guías de estadística.

Distribución de la Calificación y Condiciones para Aprobar

La evaluación del curso se basa en un enfoque práctico:

Participación activa (10%):

Intervención y compromiso dentro de las clases.

Tareas prácticas (30%):

Desarrollo de laboratorios y ejercicios aplicados.

Examen parcial (30%):

Evaluación teórico-práctica sobre fundamentos y técnicas iniciales.

Examen final (30%):

Evaluación integral de todos los contenidos desarrollados durante el curso.

Requerimientos mínimos en cada módulo:

Asistencia: Las sesiones están diseñadas para enriquecer el aprendizaje, aunque la participación no es un requisito condicionante para la aprobación del curso.

Calificación Mínima: Obtener un promedio final superior a 70/100 puntos en el total de las evaluaciones.





cit.pol.una.py

