

UNIVERSIDAD REGIONAL AUTÓNOMA DE LOS ANDES

“UNIANDES”



Facultad de Sistemas Mercantiles

Carrera de Software

Ambato

**MÓDULO PARA EL PROCESAMIENTO DE TITULACIÓN COMPLEXIVO -
SEMINARIO**

Profesor: Ing. Juan Felipe Nájera Puente, Msc.

Modalidad: En línea

Módulo: Analítica con Python

Octubre 2025 – Noviembre 2025

1. OBJETIVO

Proporcionar a los estudiantes los conocimientos y habilidades necesarios para diseñar, construir y desplegar sistemas de analítica de datos con Python. Los estudiantes aprenderán a transformar datos crudos en insights accionables a través de pipelines de procesamiento, APIs robustas y dashboards interactivos, aplicando técnicas de visualización y Machine Learning que aporten a la práctica profesional del Ingeniero en Software.

2. ESQUEMA DE CONTENIDO

- **Semana 1: Introducción a Python y Kick-Off del Proyecto**
 - **Temas:** Fundamentos de Python (variables, estructuras de datos, control de flujo), funciones, y herramientas de desarrollo profesional (Git, GitHub, entornos virtuales).
 - **Entregable:**
 - "Plan de Ataque" de los Casos de Estudio con el análisis estratégico inicial del caso asignado.
 - Repositorio de GitHub configurado, incluyendo el plan y la colaboración inicial del equipo.
- **Semana 2: Manipulación y Análisis de Datos**
 - **Temas:** Análisis exploratorio, limpieza (nulos, duplicados), transformación de datos (feature engineering, manejo de fechas) y agregaciones (groupby) utilizando la librería Pandas.
 - **Entregable:** Pipeline de datos V1.0: Un script modular y ejecutable que procesa los datos crudos del proyecto y genera un archivo de salida limpio.
- **Semana 3: Visualización y Análisis Exploratorio de Datos**
 - **Temas:** Principios de visualización (Seaborn, Plotly), desarrollo de dashboards web con Streamlit, creación de interactividad con widgets y optimización de rendimiento.
 - **Entregable:** Prototipo de Dashboard Interactivo V1.0: Una aplicación web funcional construida con Streamlit que visualiza los datos del proyecto de forma interactiva.
- **Semana 4: Ciencia de Datos Aplicada y Escalabilidad**
 - **Temas:** Desarrollo de APIs con FastAPI, integración cliente-servidor, implementación de modelos de Machine Learning, empaquetado de aplicaciones con Docker y despliegue en la nube.
 - **Entregable:**
 - Sistema integrado (Dashboard consumiendo la API).
 - Despliegue de la aplicación en una plataforma en la nube con URL pública.

3. BIBLIORAFÍA RECOMENDADA

- McKinney, W. (2022). Python for Data Analysis, 3rd Edition. O'Reilly Media.
- VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.
- Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow, 3rd Edition. O'Reilly Media.

4. RESULTADOS ESPERADOS

- Aplicar los fundamentos de Python para manipular, limpiar y transformar datasets complejos de manera eficiente y reproducible.
- Concebir, desarrollar y mantener sistemas de analítica de datos compuestos por un pipeline de procesamiento, una API y un dashboard interactivo.
- Comunicar hallazgos y insights a través de visualizaciones de datos efectivas, utilizando librerías como Plotly y plataformas como Streamlit.
- Implementar soluciones de Machine Learning (clasificación, regresión, clustering) y exponer como servicios a través de una API, demostrando una comprensión de la arquitectura de software moderna.

5. EVALUACIÓN

- Evaluación – Examen de Módulos
- Defensa del Caso de Estudio

6. CASO DE ESTUDIO

El Caso de Estudio consistirá en el desarrollo y despliegue de un sistema de analítica de datos completo basado en uno de los 15 casos de estudio asignados. El objetivo es transformar un problema de negocio en un producto de datos funcional, demostrando una comprensión práctica de todo el ciclo de vida de un proyecto de analítica.

El sistema deberá seguir la arquitectura enseñada en el curso, consistiendo en tres componentes principales:

1. Un pipeline de datos para el procesamiento
2. Una API con FastAPI para servir los datos y/o modelos
3. Un dashboard interactivo con Streamlit como interfaz de usuario.

A continuación, se detallan las características que deberá cumplir el proyecto:

- **Arquitectura de 3 Componentes:** El sistema debe estar claramente desacoplado. El dashboard de Streamlit debe consumir todos sus datos a través de llamadas a la API de FastAPI, no leyendo archivos locales.
- **Pipeline de Datos Robusto:** El proyecto debe incluir un pipeline de datos modular y ejecutable que transforme los datos crudos en un formato limpio y analizable, listo para ser consumido por la API.

- **Interactividad y Storytelling:** El dashboard debe permitir al usuario final explorar los datos a través de al menos dos widgets interactivos (filtros, sliders, etc.) que actualicen las visualizaciones en tiempo real, contando una historia clara que responda a las preguntas clave del caso de estudio.
- **Código de Calidad:** El código debe estar organizado en módulos, seguir buenas prácticas (funciones, comentarios) y estar alojado en un repositorio de GitHub con contribuciones regulares de todos los miembros del equipo y un README.md actualizado.
- **Despliegue Público:** El proyecto final debe estar completamente implementado y accesible públicamente a través de una URL funcional (utilizando plataformas como Streamlit Community Cloud), demostrando habilidades en el despliegue de aplicaciones de datos.
- **Presentación Final (Pitch):** El proyecto se presentará en un formato de pitch, donde se expondrá el problema de negocio, la solución técnica implementada y los principales insights descubiertos.