

Seminario: Analítica con Python

Explicación Casos de Estudio



Profesor: Juan Felipe Nájera

Fecha: Viernes, 17 de octubre de 2025

Agenda de Hoy

01

Visión del Caso de Estudio

Arquitectura y Entregables Finales.

03

Casos de Estudios

Reporte de Equipos y Próximos Pasos.

02

Asignación de Tareas

Presentación de Equipos y Casos de Estudio.

04

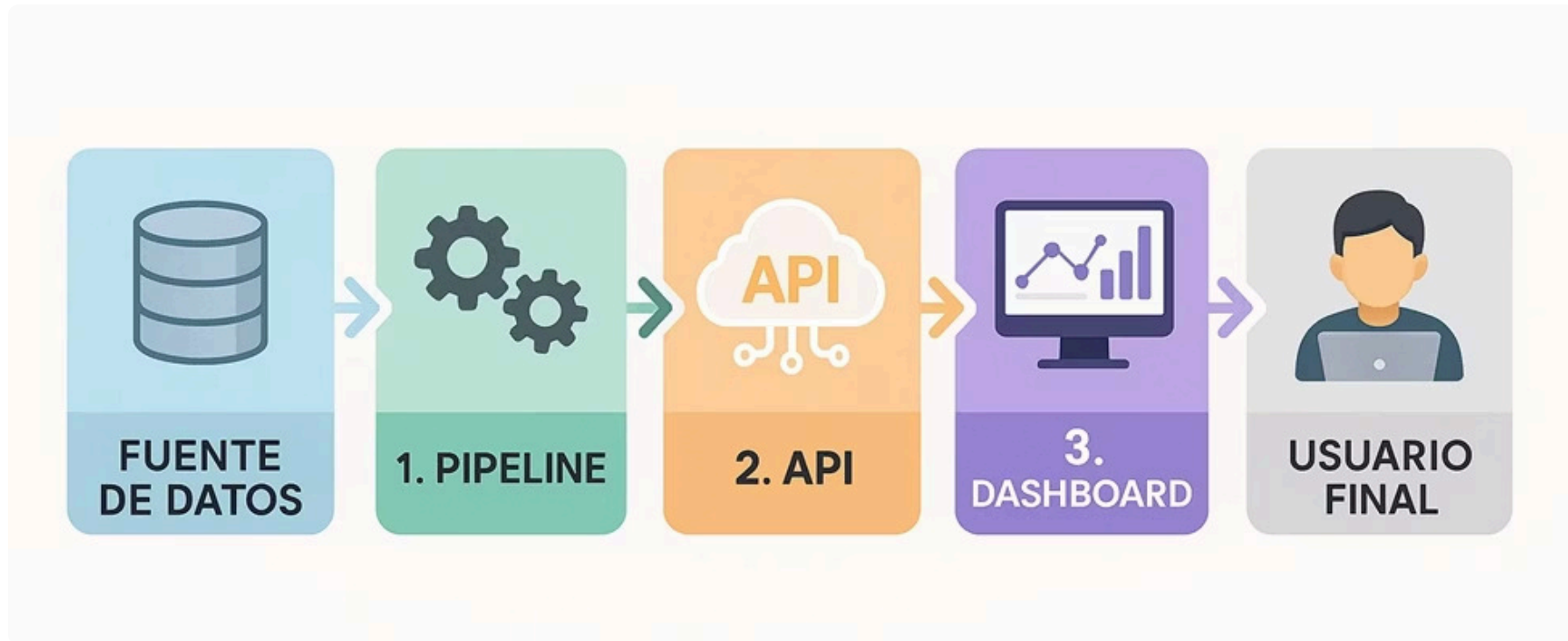
Taller Práctico

"Plan de Proyecto Inicial" en Equipos.

- ❏ **Objetivo Principal:** al finalizar esta sesión, cada equipo tendrá un plan de acción concreto, entendiendo qué van a construir y cómo van a empezar, todo documentado en su Plan de Proyecto Inicial.

La Arquitectura

Componentes clave de la arquitectura:



Pipeline de Datos

Responsable de la extracción, limpieza y transformación de los datos para prepararlos para el análisis.

API de Servicio

Construida con FastAPI, expone "endpoints" (URLs) que sirven los datos ya procesados de forma controlada y segura.

Dashboard Interactivo

Construido con Streamlit, consume datos exclusivamente desde la API para presentar visualizaciones interactivas al usuario.

Entregable final

El proyecto culmina con la entrega de un **único repositorio en GitHub**, diseñado para ser completo, portable y fácil de usar. Este repositorio contendrá todos los elementos esenciales para replicar y expandir su solución de analítica.



Código Fuente Completo

Organizado de forma modular, cada componente debe estar claramente definido y fácil de navegar. Se valorará la limpieza y los comentarios en el código. Aquí debe estar los scripts del Pipeline, el código API y el código del Dashboard Interactivo.



requirements.txt

Archivo fundamental que lista todas las dependencias del proyecto. Esto asegura que cualquier persona pueda recrear su entorno de desarrollo con un solo comando.



Dockerfile

Un Dockerfile bien estructurado para empaquetar la API. Esto demuestra la portabilidad de su solución y facilita su despliegue en diferentes entornos.



Documentación (README.md)

Un README.md profesional y claro. Debe incluir una descripción del proyecto, instrucciones detalladas para la instalación y ejecución, y ejemplos de uso.

Conformación de equipos y presentación de casos de estudio

Como saben, los equipos ya están formados. Tenemos un total de **15 grupos**, de los cuales 12 están compuestos por 3 personas y 3 por 2 personas. A continuación, se presenta la lista de los proyectos que cada equipo abordará.

Los casos son los siguientes:

Segmentación de Clientes para Retail	Construir un sistema que agrupe a clientes de un e-commerce en segmentos (VIP, Leales, En Riesgo) según su comportamiento de compra para optimizar el marketing.
Análisis Musical y Predicción de Popularidad	Desarrollar un modelo que analice las características sonoras de más de 170,000 canciones de Spotify para predecir si una nueva canción se convertirá en un "hit".
Análisis de Sentimiento en Reseñas de Hoteles	Crear un sistema de NLP para procesar 500,000 reseñas de hoteles, clasificar su sentimiento y extraer los temas clave (limpieza, servicio) que generan satisfacción o descontento.
Pronóstico de Ventas y Optimización de Inventario	Construir un sistema de series de tiempo que pronostique la demanda de productos de una gran tienda para ayudar a gestionar el inventario de forma proactiva.
Análisis Comparativo Bancario Ecuatoriano	Desarrollar una herramienta de inteligencia de negocios para consolidar y visualizar los boletines de la Superintendencia de Bancos, permitiendo comparar la salud financiera de los bancos del país.
Visualización de la Pandemia de COVID-19	Crear un dashboard de salud pública que procese el dataset global de "Our World in Data" para explorar y comparar la evolución de la pandemia a nivel mundial y local.
Scouting y Valoración de Jugadores de Fútbol	Construir un sistema que utilice datos de FIFA para analizar atributos de jugadores, predecir su valor de mercado e identificar talento potencialmente infravalorado.
Sistema de Recomendación Híbrido de Películas	Desarrollar un sistema que combine el análisis del contenido de las películas (géneros, tags) y el comportamiento de usuarios similares para generar recomendaciones personalizadas.
Análisis Geoespacial Movilidad Urbana NYC	Procesar millones de registros de viajes en taxi de NYC para crear un dashboard con mapas de calor y flujos que revelen los patrones de movilidad de la ciudad.
Análisis y Monitoreo Actividad Sísmica Ecuador	Transformar los catálogos sísmicos del Instituto Geofísico en un dashboard interactivo para explorar y visualizar la actividad sísmica histórica del país por fecha, magnitud y ubicación.
Predicción de Retrasos de Vuelos	Analizar más de 5 millones de registros de vuelos en EE.UU. para construir un modelo de machine learning que prediga la probabilidad de que un vuelo se retrase.
Análisis Ecosistema Stack Overflow	Construir un dashboard que analice cientos de miles de preguntas de Stack Overflow para visualizar las tendencias y la popularidad de diferentes tecnologías de programación a lo largo del tiempo.
Predicción Precios Alquileres Airbnb NYC	Desarrollar un modelo de regresión que prediga el precio por noche de un alquiler en Airbnb basándose en sus características y ubicación geográfica en Nueva York.
Análisis Reseñas E-Commerce y "Utilidad"	Utilizar más de 500,000 reseñas de Amazon para construir un modelo que prediga qué tan útil (helpful) será una nueva reseña para otros clientes basándose en su texto y calificación.
Predicción Precios Vehículos Segunda Mano	Crear un sistema de tasación automática que analice casi 500,000 anuncios de Craigslist para predecir el precio de mercado de un vehículo usado basándose en sus características.

Plan de Proyecto Inicial

Para asegurar un inicio exitoso y una dirección clara en sus proyectos, cada equipo deberá presentar un plan de proyecto inicial detallado, cubriendo los siguientes puntos clave:

Título del Proyecto y Miembros

Identificación del caso de estudio asignado y los nombres completos de todos los integrantes del equipo.

Definición del Problema

Una descripción concisa que articule el problema de negocio o la necesidad que el proyecto busca resolver.

Objetivo Principal

Una única frase, clara y medible, que establezca la función principal del producto de software a construir.

Columnas Clave del Dataset

Un listado de columnas del archivo de datos, seleccionadas tras una revisión visual preliminar, consideradas esenciales para el análisis.

Responsabilidades de los Componentes

Una breve descripción de las funciones específicas que desempeñarán el Pipeline, la API y el Dashboard dentro de la solución.

Primeras Tareas Técnicas (Sesión 5)

Un plan de acción concreto para la próxima sesión, incluyendo la creación del repositorio en GitHub, la redacción del README.md y el script inicial para la carga de datos.

Dudas y Riesgos Identificados

Una enumeración de cualquier pregunta, duda técnica o posible obstáculo anticipado durante el desarrollo del proyecto.

Investigación Adicional

Pueden comenzar a investigar también por su cuenta todo lo relacionado con limpieza, análisis y visualización de datos, como también temas de Machine Learning.

Presentación Ejecutiva de Planes

Al finalizar la sesión, cada equipo tendrá 90 segundos para una presentación ejecutiva de su plan inicial, enfocándose en dos aspectos clave:



Objetivo Principal

Indiquen la función principal del producto de software que construirán con una frase clara y medible.



Riesgo Más Importante

Mencionen el riesgo principal o el obstáculo más crítico que han identificado para el proyecto.

Próximos Pasos y Agenda

Aquí te presentamos las tareas inmediatas y un adelanto de lo que cubriremos en la próxima sesión:



Tarea Clave

Copia el contenido de tu Plan de Estudio al `README.md` de tu nuevo repositorio en GitHub.

Fecha Límite: Domingo.



Avance de la Sesión 5 (Mañana)

- Sesión 100% práctica.
- Configuración del Entorno Profesional (`venv`).
- Flujo de Colaboración con Git y GitHub (branch, commit, push, Pull Request).
- Primer Código Funcional: Introducción a Pandas y carga de tu dataset.

¿Preguntas?

The background of the slide is a light blue gradient. It is decorated with several faint, semi-transparent graphics. These include question marks of various sizes and colors (light blue and white) and speech bubbles of different shapes and sizes, some in white and some in light blue. The overall theme is one of inquiry and communication.