






## **Clase 1: Introducción**

# **LABORATORIO DE PROGRAMACIÓN CIENTÍFICA PARA CIENCIA DE DATOS**

MDS7202-1 – Otoño 2024

# Objetivo de la clase

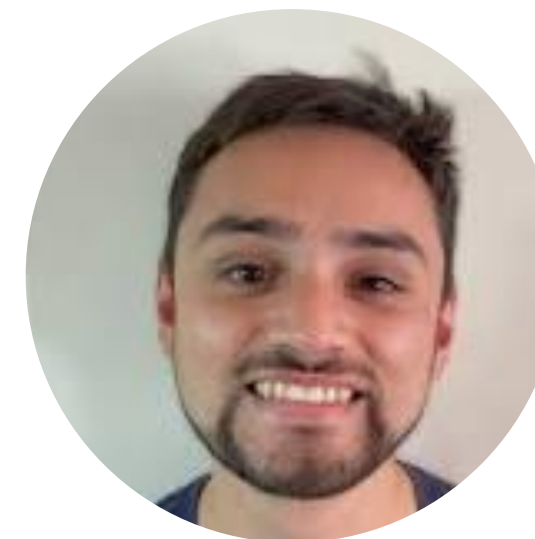
- Conocer al equipo docente 
- ¿Para que es este curso? 
- Metodología 
- Información Administrativa 
- Calendario 

# Equipo Docente



**Ignacio Meza**  
**Profesor**

- Ingeniero Civil Eléctrico
- Candidato en Msc. Ciencias de la Computación
- Data Scientist @BCI-MACH
- 6 iteraciones del curso



**Sebastian Tinoco**  
**Profesor**

- Ingeniero Comercial
- Candidato en Msc Ciencia de Datos
- Data Scientist @AB-INBEV
- 3 iteraciones del curso

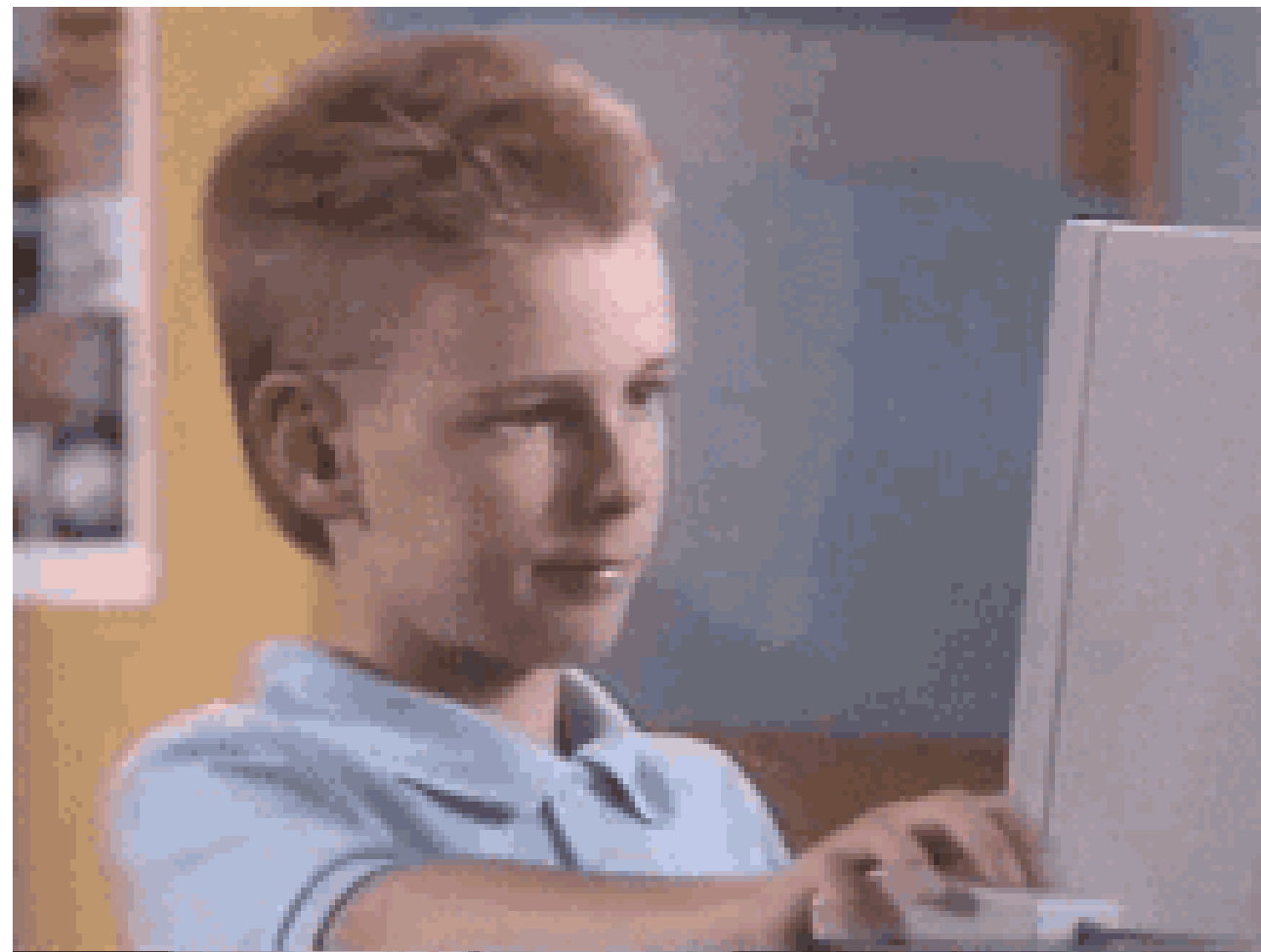
# ¿Para que es este curso?



**Respondamos en Menti!**

# ¿Qué **si** es este curso?

El curso estará enfocado en entregar las **herramientas necesarias**, tanto teóricas como prácticas para el: análisis, modelamiento, resolución y **puesta en marcha de proyectos en ciencia de datos.**



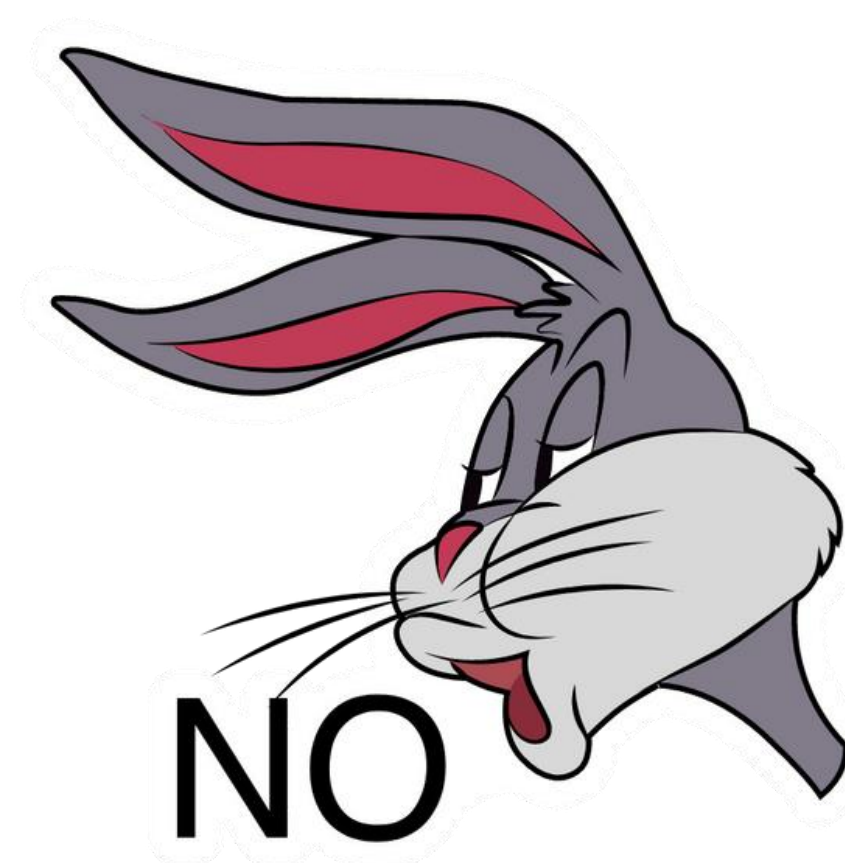
# ¿Qué **no** es este curso?

## NO es un curso teórico:

- Veremos **aspectos teóricos sólo en cuanto se necesite**. Pero se pondrá énfasis en la práctica.
- Múltiples opciones para profundizar (<http://mds.uchile.cl/programa/>):

### **Electivos (catálogo 2020)**

- EL7006 Redes Neuronales y Teoría de Información para el Aprendizaje
- EL7007 Introducción al Procesamiento Digital de Imágenes
- EL7024 Teoría de Información: Fundamentos y Aplicaciones
- EL7037 Computación Evolutiva
- EL7014 Diagnóstico y Pronóstico de Fallas
- EL7031 Robotics, Sensing and Autonomous Systems
- EL7021 Seminario de Robótica y Sistemas Autónomos
- CC6204 Deep Learning
- CC5212 Procesamiento Masivo de Datos
- CC5213 Recuperación de Información Multimedia
- CC7220 La Web de Datos
- CC5208 Visualización de Información
- CC5113 Aprendizaje Automático Bayesiano
- CC5615 Business Analytics
- CC5509 Reconocimiento de Patrones





# Metodología del Curso



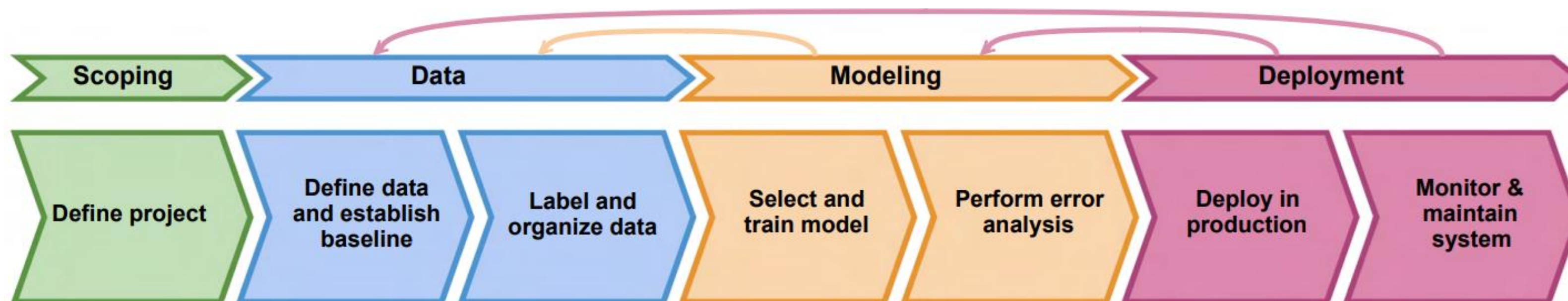


**Respondamos en Menti!**

# Metodología del Curso 1.0

- **Unidad 1:** Introducción a las Herramientas para Ciencia de los Datos (3 sem).
- **Unidad 2:** Manejo de Arreglos Multidimensionales y computación Científica (1 sem).

- **Unidad 4:** Modelos de Aprendizaje Automático (5 sem).
- **Unidad 5:** Optimización de Código y Despliegue (3 sem).



- **Unidad 3:** Manejo de Datos Tabulares y con Pandas (2 sem).
- **Unidad 4:** Visualizaciones y Análisis Exploratorio de Datos (1 sem).

# Metodología del Curso 2.0

## Fundamentos de Programación

- Python
- OOP
- Numpy
- Pandas
- Pyspark
- Plotly

## Machine Learning

- Aprendizaje Supervisado
- Aprendizaje No Supervisado

## Tópicos de Machine Learning

- Optimización de Hiperparámetros
- Forecasting
- Sistemas de Recomendación

## Interpretabilidad

- Métodos locales
- Métodos globales

## Agentes Autónomos

- Reinforcement Learning
- LLM

## MLOps

- Despliegue
- Pipelines productivos
- Monitoreo

# Cronograma del Curso



## Fundamentos de Programación (11 clases)



Semana 1

# Cronograma del Curso



## Fundamentos de Programación (11 clases)



Semana 6

Semana 1

## Machine Learning (5 clases)





# Cronograma del Curso



**Fundamentos de Programación (11 clases)**

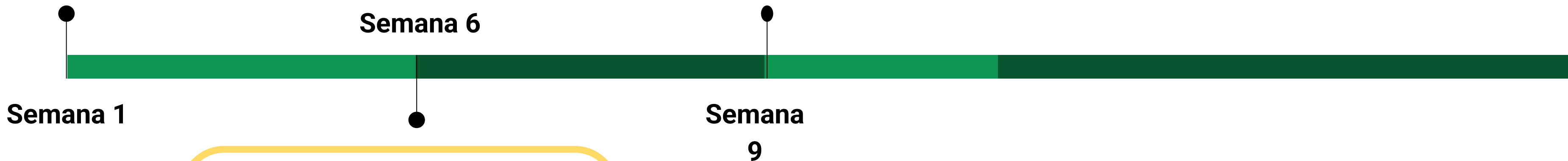
  





**Tópicos de Machine Learning (3 clases)**



 

 **O P T U N A**



**Machine Learning (5 clases)**

# Cronograma del Curso



## Fundamentos de Programación (11 clases)



## Tópicos de Machine Learning (3 clases)



Semana 1

Semana 6

Semana 9

Semana 10

## Machine Learning (5 clases)



## Interpretabilidad (3 clases)



# Cronograma del Curso



## Fundamentos de Programación (11 clases)



## Tópicos de Machine Learning (3 clases)



## Agentes Autónomos (2 clases)



Semana 1

Semana 6

Semana 9

Semana 10

Semana 12

## Machine Learning (5 clases)



## Interpretabilidad (3 clases)



# Cronograma del Curso



**Fundamentos de Programación (11 clases)**

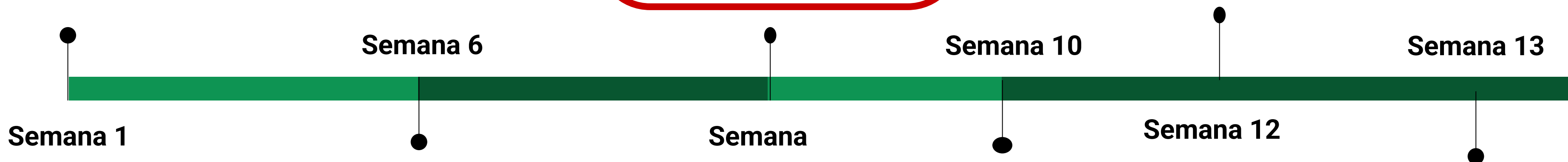
     
   
 

**Tópicos de Machine Learning (3 clases)**





   


**Agentes Autónomos (2 clases)**

 **Hugging Face**  
 **LlamaIndex**  

**Machine Learning (5 clases)**

**Interpretabilidad (3 clases)**

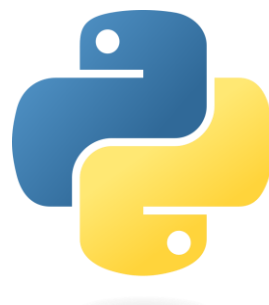
  
SHAP

**MLOps (4 clases)**

 **Apache Airflow**  **docker**  
 **FastAPI**  **gradio**

# Stack Tecnológico





**Hugging Face**

PROPHET

# Stack Tecnológico



Visual Studio Code



Apache  
Airflow



LlamaIndex



surprise



SHAP



PYCARRET



XGBoost



gradio



OPTUNA



# Información Administrativa

# Reglas del curso



# Reglas del curso

## Curso Presencial

Todas las cátedras serán tutoriales en donde resolveremos problemas desafiantes usando distintas tecnologías aplicadas a la ciencia de los datos.

### **¡La idea es que participen!**

- Cátedras de 1.30 hora.
- Pausas de 5 minutos a los 40 minutos de clases.
- No hay asistencia obligatoria a las cátedras.
- La colaboración es fundamental para resolver problemas complejos. Este será un concepto fundamental en el curso. La mayor parte de las evaluaciones serán grupales.

# Evaluaciones

Las evaluaciones este semestre serán:

 **12 Laboratorios**



 **1 Proyecto**

 **Repositorio**



# Labs

Evaluación que evalúa las herramientas vistas en la semana.

- Serán por ZOOM  en horario de auxiliar.
- La Auxiliar  estará presente para explicar enunciado y contestar dudas.
- La asistencia a esta sesión **no es obligatoria.**
- Estarán diseñados para que se puedan contestar en el horario del auxiliar más un pequeño tiempo de trabajo en casa.

# Labs

Reglas:

- **Grupos de 2** (puede ser individual, pero no se recomienda).
- **6 días de plazo** con descuento de **1 punto por día**. Entregas el Jueves a las 23.59.

# Labs

## Reglas:

- Se pide **respeto por los horarios hábiles**. No se responderá en horarios de noche.
- **Consultas por correo y/o foro U-cursos**.
- Consultas del lab anterior permitidas en el lab de la semana actual, pero evitarlas.
- **Errores graves** de contenidos vistos en Minería de Datos o Aprendizaje de Máquinas serán **descontados**.

Incentivamos conversar con sus compañeros acerca de la soluciones de los labs, pero:

- **Está prohibida la copia** (obviamente).
- También que solo uno del equipo trabaje. La idea es que ambos resuelvan en conjunto el lab!
- **Si existen problemas con sus compañeros, comentarlo al equipo docente**.

# Proyecto

**Evalúa el contenido general del curso.**

- Entregable: Notebook con informe + código

## **Reglas:**

- Grupos de 2.
- Plazo y problema por determinar. Sin atrasos.

# Repositorio

**Contiene todas las evaluaciones del curso.**

El objetivo es que incorporen esta herramienta en sus metodologías de trabajo.

**Se evaluará:**

- **Rama master o main** en donde estarán todas las evaluaciones ya terminadas. Las entregas serán revisadas por este medio.
- **Ramas lab{n}** en donde deberán tener el desarrollo del lab y proyecto actuales. La idea es que a medida que trabajen vayan guardando su lab/proyecto no finalizado en esta rama.
- **Orden y prolijidad:** Cada laboratorio debe estar guardado en una carpeta diferente. Además, se debe incluir un readme explicando la estructura del repositorio.

**Se irá registrando en cada evaluación el correcto uso de git.**



# Evaluaciones

La nota final se calcula como:

$$\text{NF} = 70\% \text{ labs} + 25\% \text{ proyecto} + 5\% \text{ repositorio}$$

Observación: Se debe aprobar labs y proyecto por separado para pasar de curso.



¡Recuerden que el objetivo es aprender!

Por ende, está **totalmente prohibida la copia**

# Canales de Comunicación



## **Foro de U-cursos:**

- Cualquier duda de los contenidos del curso y administrativas.
- Todos los correos con consultas al equipo docente serán redirigidos al foro. (Nuevamente, colaboración...)



**GitHub:** Repositorio Oficial: Todo el material del curso se encontrará ahí, incluido el calendario.

# Calendario



🐱 Podrá estar **sujeto a cambios** frente a cualquier eventualidad. La versión actual la podrán encontrar en el repositorio del curso.

Semana	Módulo	Fecha	Temas	Detalle	Evaluación
1	Introducción a las Herramientas para Ciencia de los Datos	3/13/2023	Introducción al curso. IDEs y Jupyter. Markdown. Ambientes de Ejecución	Primera parte: Explicación del objetivo del curso, los contenidos, los métodos de evaluación y las reglas. Segunda parte: Trabajo con Jupyter Notebook/Lab y Markdown como entorno de trabajo para Data Science. Tercera Parte: Creación y administración de ambientes de ejecución de Python.	
		3/15/2023	Control de versiones con Git	Fundamentos de Git, repositorios, ciclos de vida de los archivos en un repositorio, Commits, Branches y colaboración y repositorios remotos.	
		3/17/2023			
2		3/20/2023	Python 1: Introducción a Python	Introducción a Python: Sintaxis, variables, tipos de datos básicos, expresiones y operaciones, control de flujo, colecciones (listas, tuplas, conjuntos, diccionarios), iteraciones.	
2		3/22/2023	Python 2: Programación Funcional y Modular.	Funciones, scope, unit testing. Decoradores. librerías built-in y programación modular.	
2		3/24/2023			Lab 1: Git
3	Manejo de Arreglos Multidimensionales y computación Científica	3/27/2023	Python 3: Programación Orientada a Objetos y Excepciones	Clases, objetos, constructores, abstracción y encapsulación, herencia, polimorfismo. Excepciones y manejo de estas. Quizas generadores.	
3		3/29/2023	Arreglos Multidimensionales con Numpy	Creación de arreglos multidimensionales, vectorización de operaciones, atributos de los arreglos, indexado, operaciones básicas, documentación, funciones universales,	
3		3/31/2023			Lab 2: Python

## Clase 1: Introducción

# LABORATORIO DE PROGRAMACIÓN CIENTÍFICA PARA CIENCIA DE DATOS

MDS7202-1 – Otoño 2024