

# Análisis de Datos

Clase 3 - Introducción y herramientas de SW

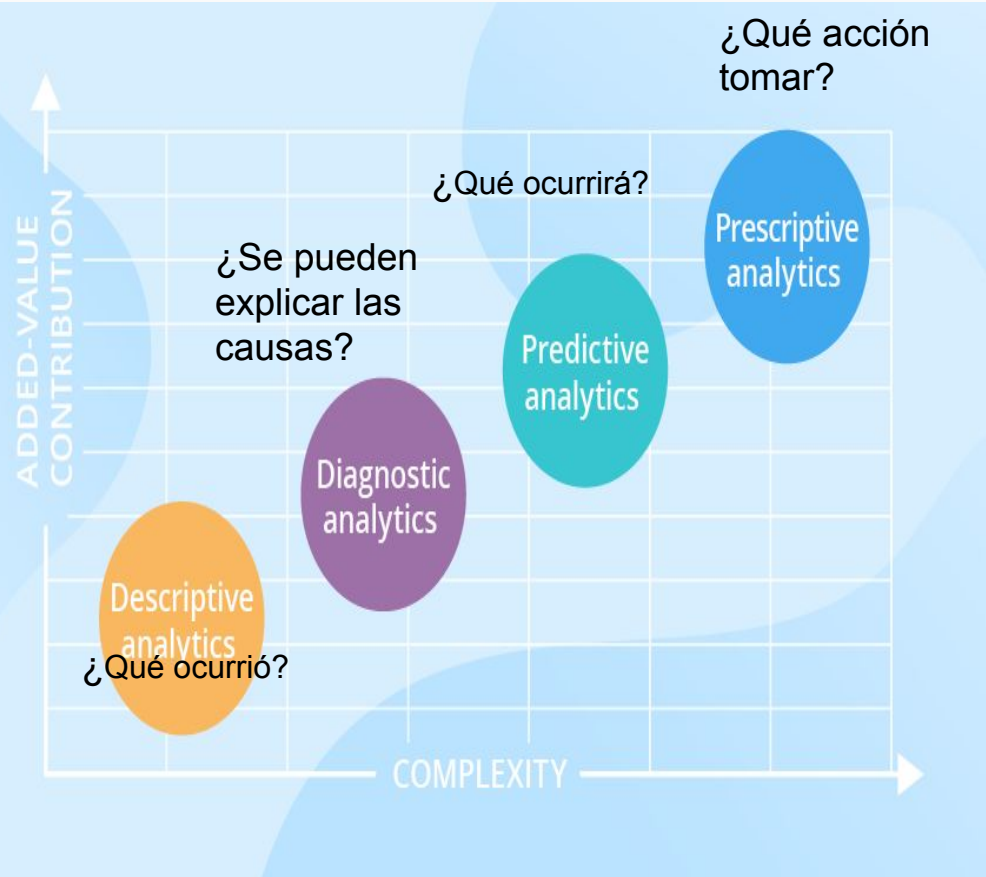
# ¿Qué es el análisis de datos?

- Es el proceso de explorar y analizar conjuntos de datos con el objetivo de hacer predicciones y contribuir a la toma de decisiones apoyada en datos.
  - Analizar datos
  - Tomar decisiones

# Aplicaciones

- Análisis de fraude
- Salud
- Administración de inventario
- Logística
- Marketing
- Planificación urbana

### 3. Tipos de análisis de datos



- Descriptivo
- Diagnóstico
- Predictivo
- Prescriptivo

### 3. Tipos de análisis de datos (ejemplos)

- **Análisis descriptivo:**
  - Estudiar la cantidad de unidades vendidas de un producto y el beneficio obtenido.
- **Análisis de diagnóstico:**
  - Hallar la correlación entre la contaminación del agua y una enfermedad.
- **Análisis predictivo:**
  - Estimar la demanda a futuro de un bien o servicio.
- **Análisis prescriptivo:**
  - Establecer los parámetros óptimos de una cadena de producción y distribución para suplir una demanda.

## 4. Pasos en el proceso de análisis de datos

1. Obtención de datos
2. Preparación de datos
3. Exploración de datos
4. Desarrollo de modelos
5. Interpretación de resultados

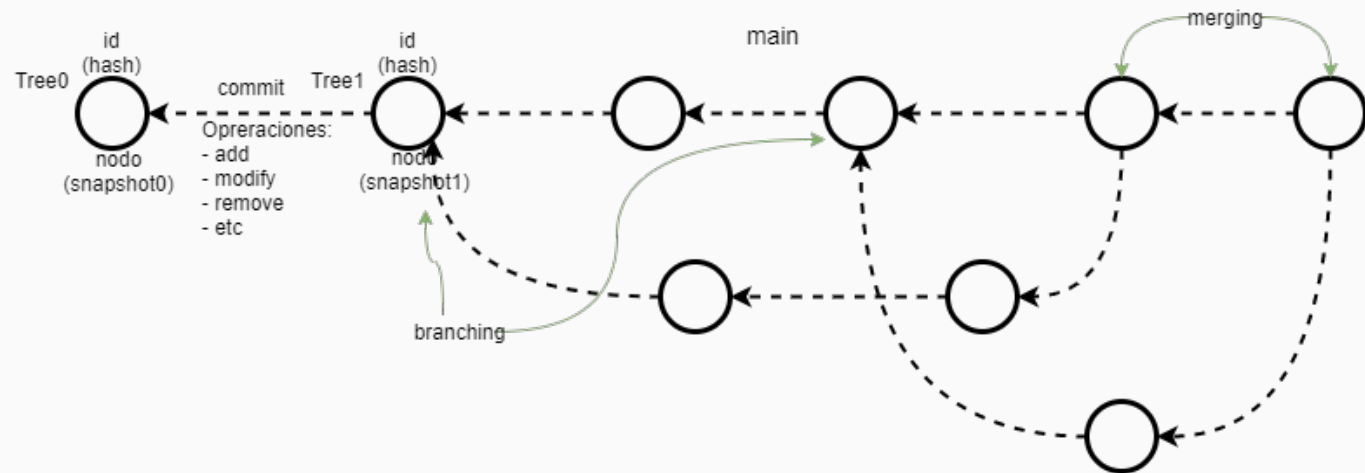
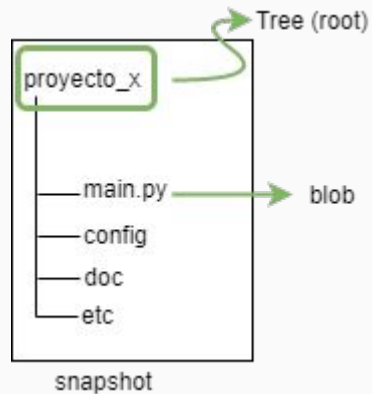


## 6. Python para análisis de datos



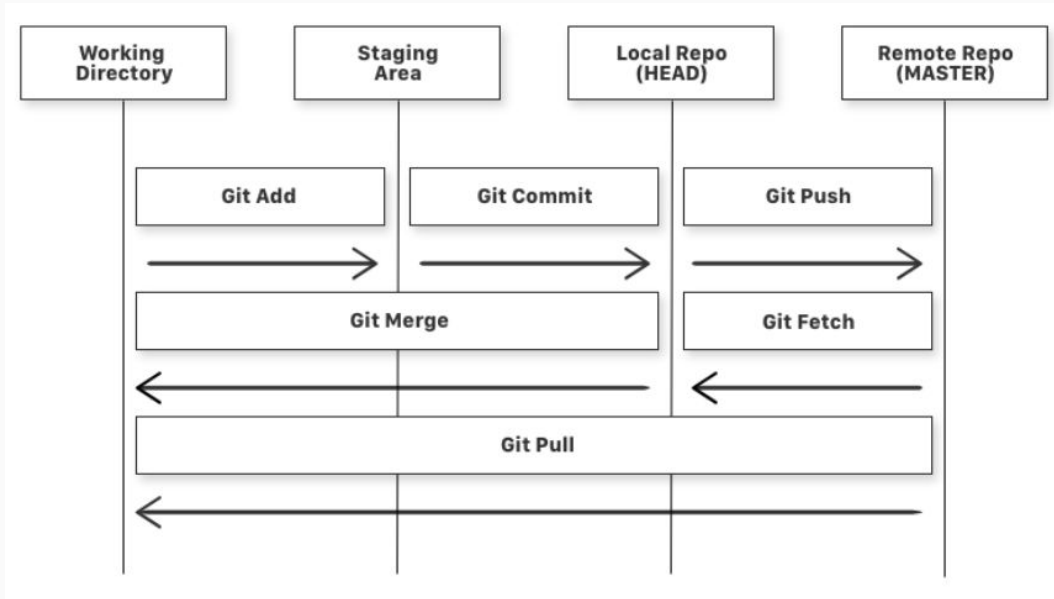
# 7. Git

- Git es un sistema de control de versiones distribuido.
- Modelo de datos: grafo dirigido acíclico (DAG)





## 8. Git



id_0	tree0
id_1	tree1
id_2	commit
id_3	tree2
...	...

id (hash)

TAGS

## 9. Git

Algunos comandos útiles

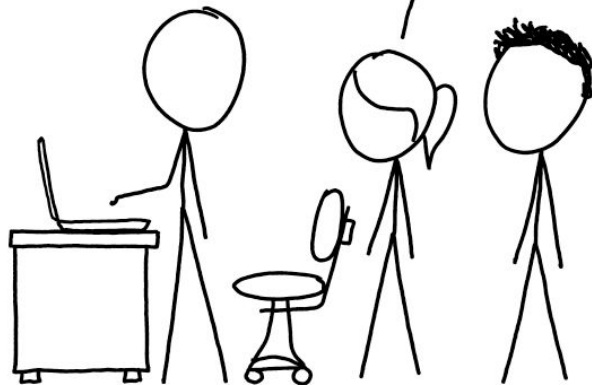
- `git clone <url>` #hace una copia local de un repo
- `git add <filename>` #agrega un archivo al *staging area*
- `git add .` # agrega todos los archivos al *staging area*
- `git commit -m "msj"` # hace el commit
- `git push` # envía los commit al repo remoto
- `git pull` # trae los cambios del repo remoto
- `git checkout -b <my branch name>` # crea una nueva rama y te posiciona allí

[Cheat Sheet + Instalación](#)

THIS IS GIT. IT TRACKS COLLABORATIVE WORK ON PROJECTS THROUGH A BEAUTIFUL DISTRIBUTED GRAPH THEORY TREE MODEL.

COOL. HOW DO WE USE IT?

NO IDEA. JUST MEMORIZE THESE SHELL COMMANDS AND TYPE THEM TO SYNC UP. IF YOU GET ERRORS, SAVE YOUR WORK ELSEWHERE, DELETE THE PROJECT, AND DOWNLOAD A FRESH COPY.



# 10. conda

- Ambientes virtuales (conda, virtualenv, etc.).
  - Posibilidad de tener múltiples configuraciones de paquetes.
- Sistema de paquetes conda.
  - Amplia comunidad de usuarios.
  - Librerías para Fortran, Python, C/C++, R, etc.

# 7. Hands-on

- Plan:
  - Repaso de GIT como sistema de control de versiones.
    - Modelo de datos de GIT.
  - Repaso de Python con algunos conceptos de Ingeniería de SW y algoritmos y estructuras de datos.
  - Introducción/repaso de Pandas.
  - Trabajo Práctico Nro. 1: Temperaturas en Europa en los últimos 500 años.