

Fecha	Ubicación	Valor
01/10/23	Girón	25.4
02/10/23	Bucaramanga	30.1
10/23	Floridablanca	28.5

Laboratorio de Ciencias de Datos - Pruebas Saber Girón

Guía 2: Los Datos del Mundo

Del análisis manual al poder de Python con Pandas.

```
import pandas as pd
data = pd.read_csv('datos_mundo.csv')
...
print(data.head())
...

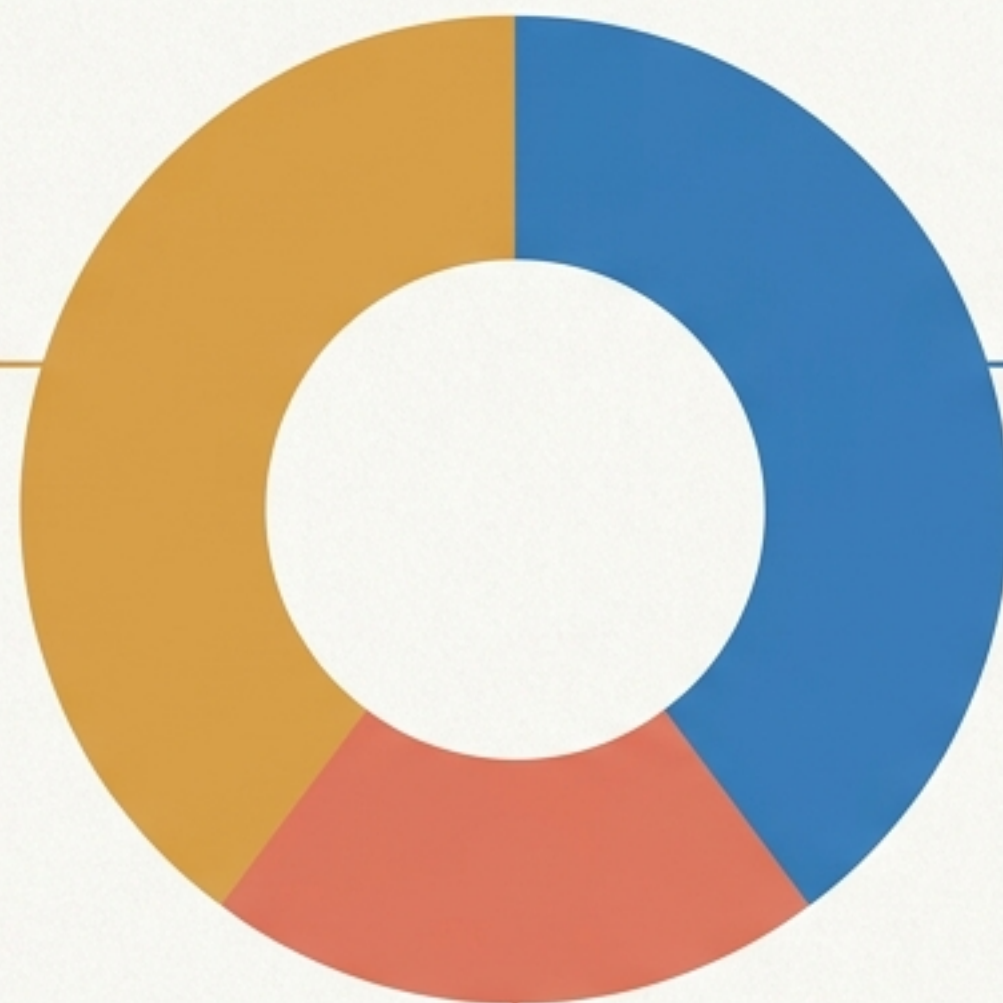
```

40%

Fase Desconectada



Analizar el problema a mano para entender el algoritmo.



40%

Fase Conectada



Traducir nuestro razonamiento a bucles y condicionales en Python.

20% Síntesis



Consolidar conceptos de Pandas e indexación.

Radiografía de nuestro Dataset

Índice (0)

Nuestra brújula numérica.
El identificador único de la fila.

Jornada

Bloque horario (Tarde,
Mañana, Única, Nocturna).

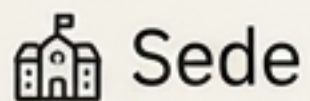
Promedio

El dato clave. El puntaje
obtenido a analizar.

Periodo

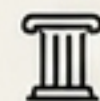
Año lectivo.

Índice:	Sede:	Jornada:	Naturaleza:	Promedio:	Periodo:
0	Principal	Tarde	Oficial	258	2020



Sede

Ubicación del colegio.



Naturaleza

Sector administrativo.

Las 5 Misiones Analíticas

1. **Mínimo Global:** Hallar el puntaje mínimo entre todos los colegios.

2. **Máximo 2022:** Hallar el puntaje máximo del 2022.

3. **Máximo Tarde:** Máximo puntaje en la jornada tarde.

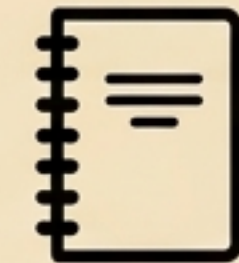
4. **Promedio No Oficiales:** Promedio de colegios No Oficiales.

5. **Promedio Nocturna:** Promedio de la jornada nocturna.

Protocolo de Equipo - Fase Desconectada



1 persona toma el tiempo de cada tarea.



2 personas toman nota de los pasos (El Algoritmo).

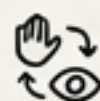


2 personas ejecutan la estrategia y calculan.

¿Por qué no hacer los cálculos a mano?

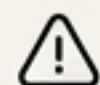
El Cerebro vs. La Máquina

Método Manual (Desconectado)



Acción:

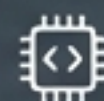
- Buscar con los ojos, comparar mentalmente, usar calculadora paso a paso, anotar en papel.



Riesgos:

- Lento (minutos por tarea), alta probabilidad de error humano, no es escalable si el anexo tuviera 10,000 filas.

Método Python (Conectado)



Acción:

- Lenguaje de programación leyendo la estructura del dataframe.

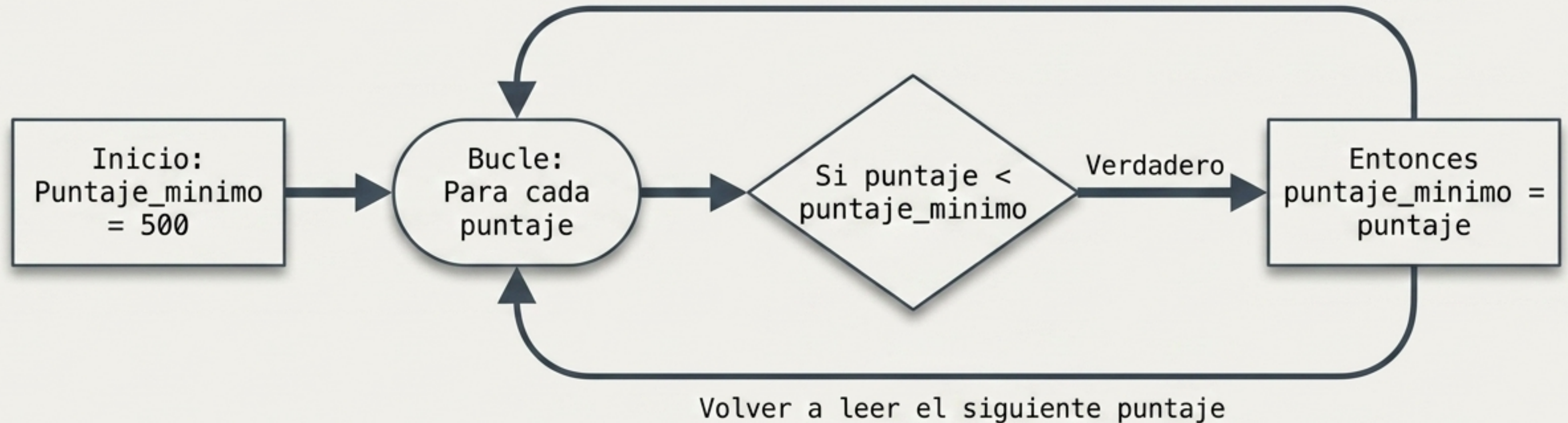


Ventajas:

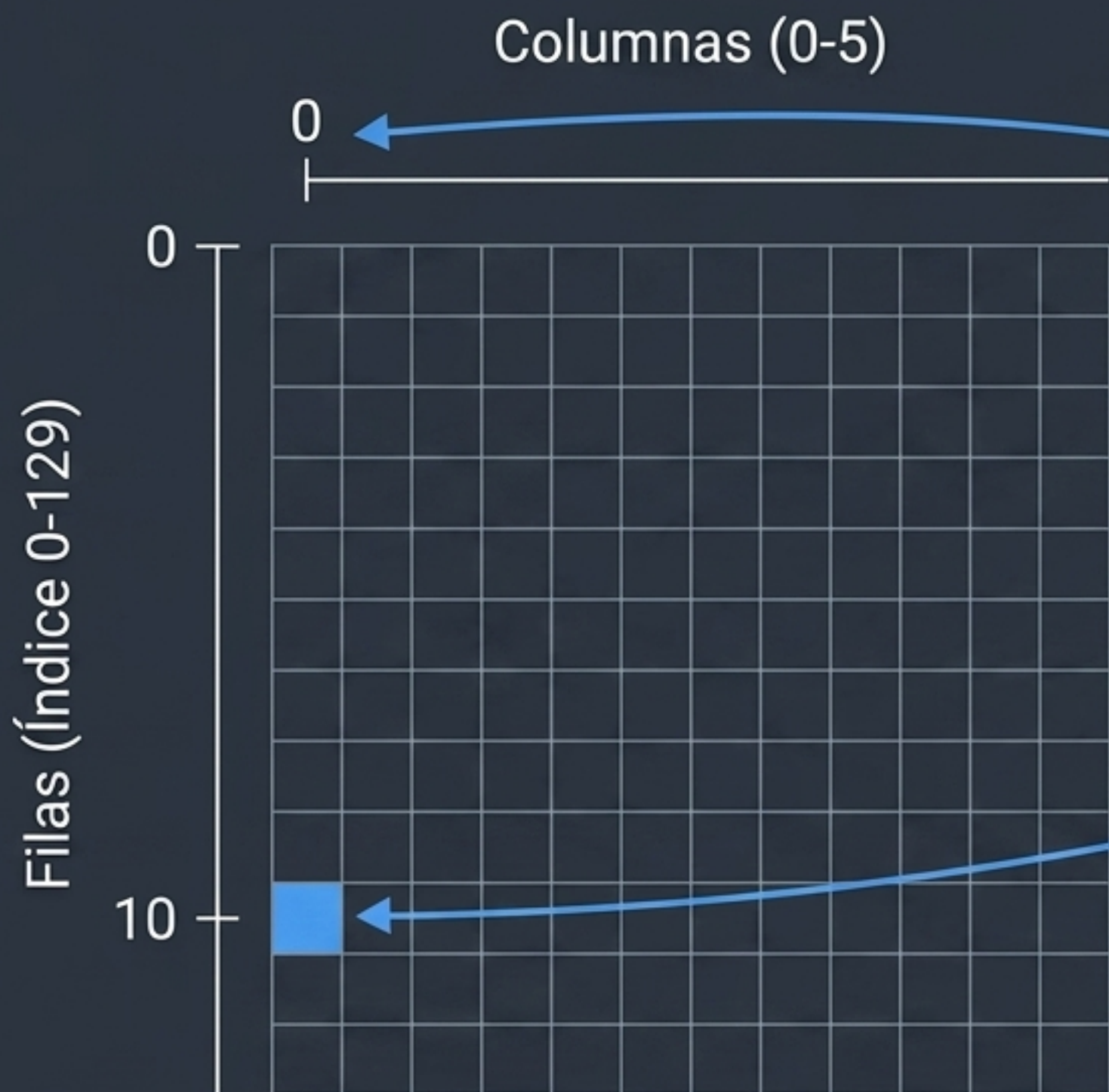
- Precisión absoluta, ejecución en milisegundos, escalabilidad ilimitada.

Entender cómo lo haces a mano es el primer paso para programarlo.

Diseñando el Algoritmo



Un algoritmo es simplemente la receta detallada de los pasos que tu cerebro ya hizo intuitivamente.



```
import pandas as pd
```

```
saber = pd.read_csv('pruebas_saber.csv')  
saber.iloc[10,0]
```

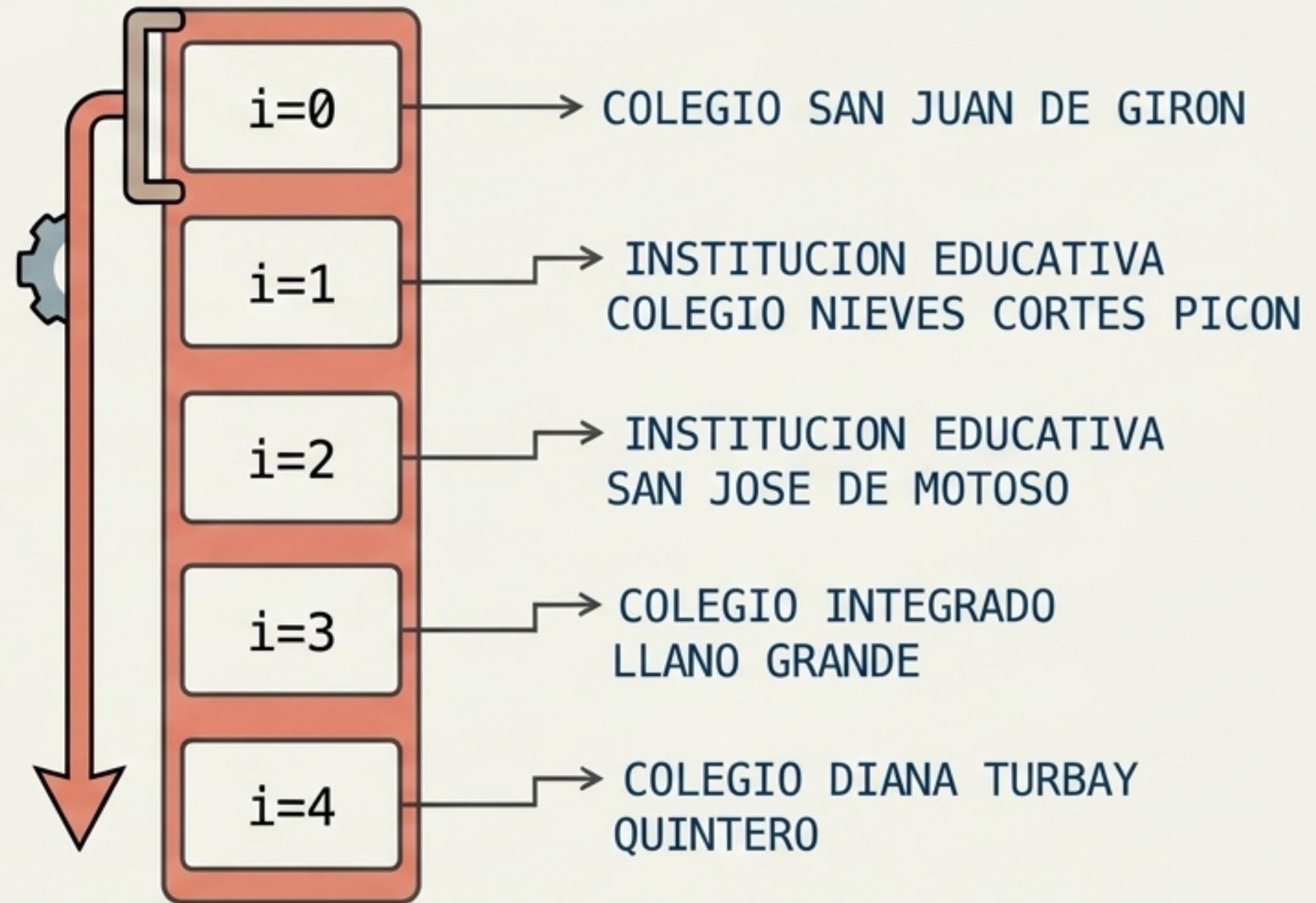
```
'INSTITUTO INTEGRADO FRANCISCO SERRANO MUÑOZ'
```

¿Qué función cumplen los números?
Son las coordenadas exactas del dato.

¿Y si usas paréntesis ()? ¡Error de sintaxis!
Pandas exige corchetes para esta indexación.

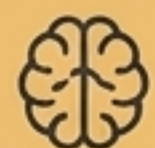
Ciclos Iterativos (El Motor)

```
for i in range(0,5):  
    print(saber.iloc[i,0])
```



El Motor Iterativo: El bucle automatiza el cambio de fila.
Reemplazamos el número fijo por la variable 'i'.

Traducción de Misiones (Condicionales)



La Lógica Humana

Filtrando por Año - Misión 2

Si el periodo de esta fila es igual a 2022...



El Código Python

```
if saber.iloc[i, 4] == 2022:
```

Filtrando por Jornada - Misión 3

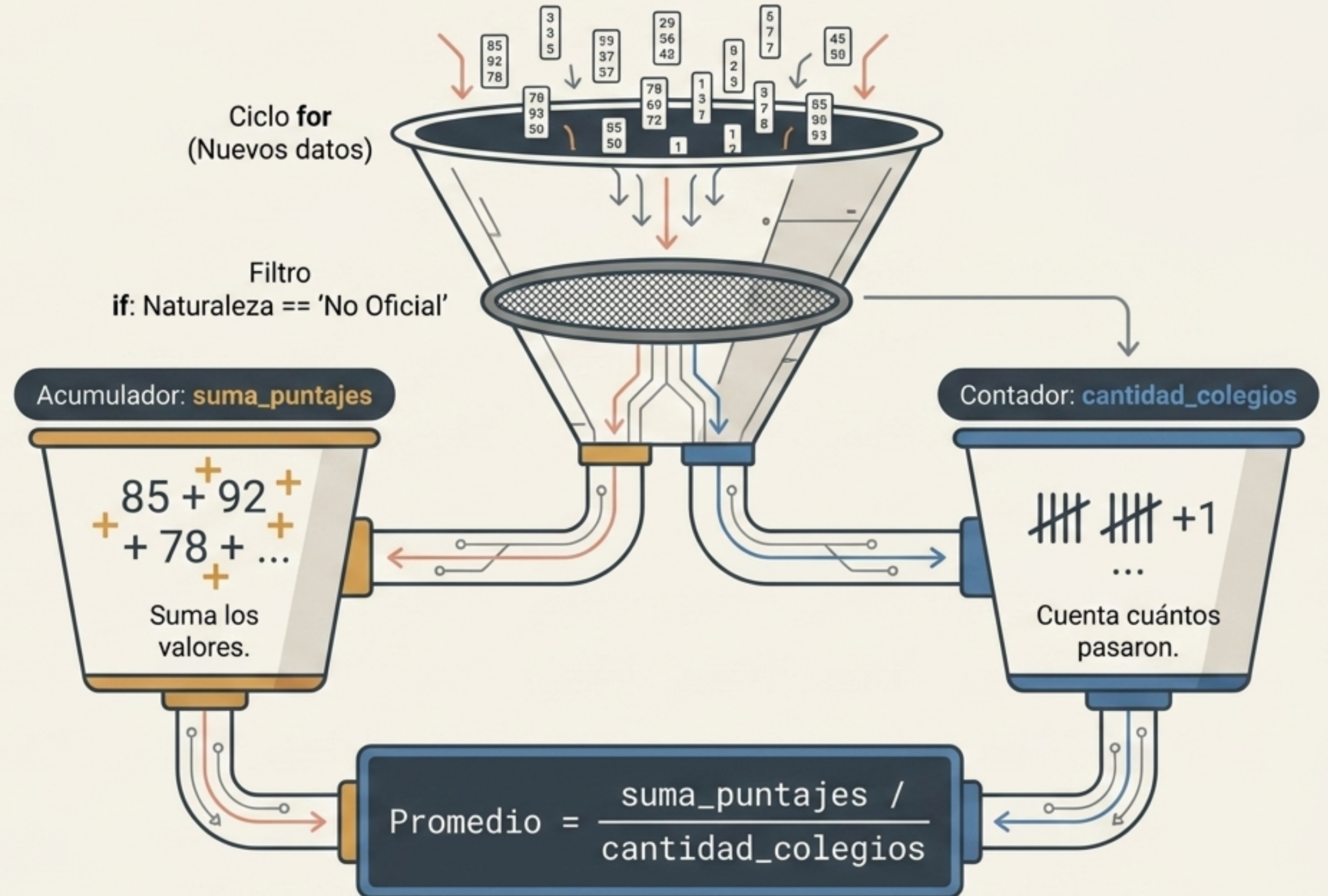
Si la jornada de esta fila dice 'Tarde'...

```
if saber.iloc[i, 2] == 'Tarde':
```

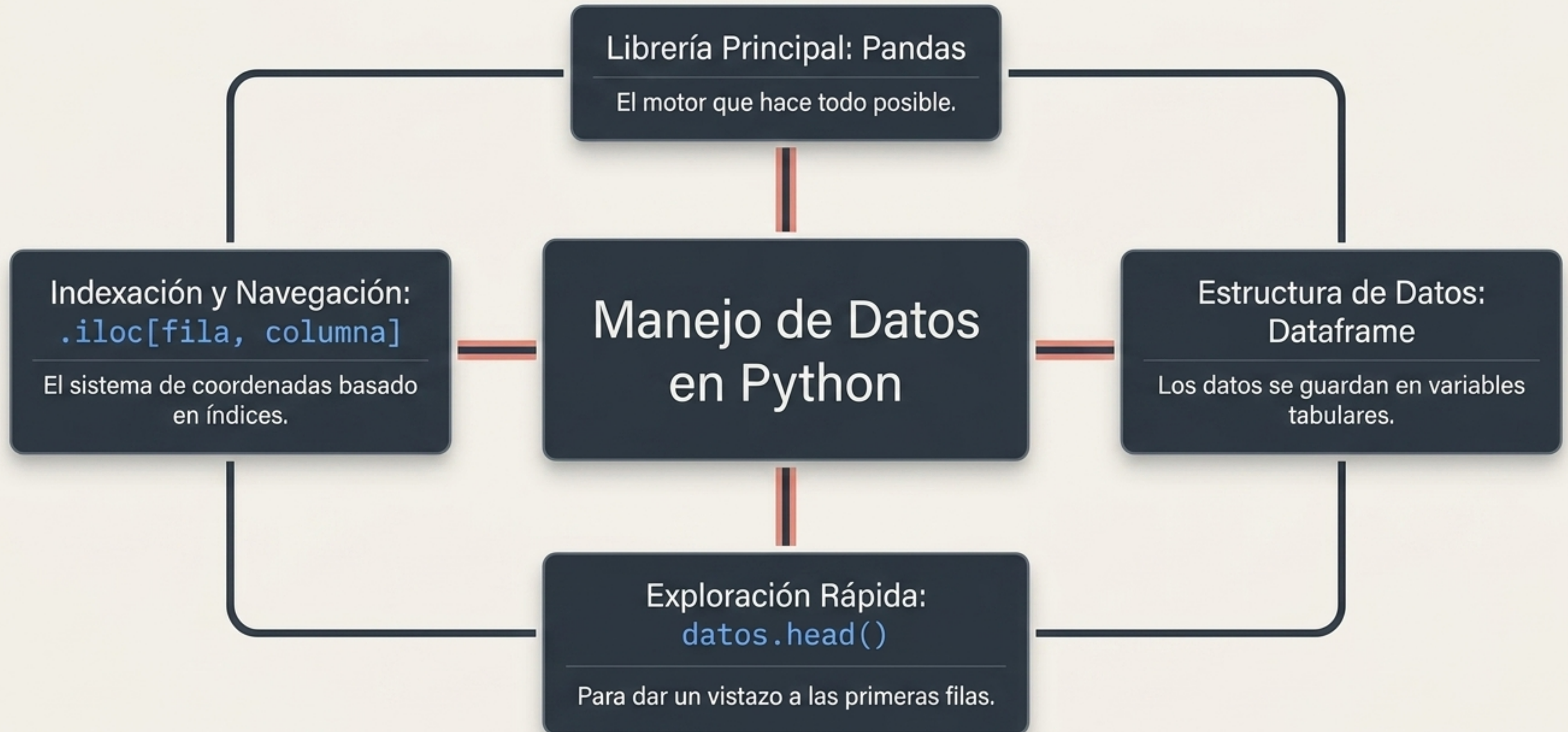
Combinando el ciclo 'for' con el condicional 'if', logramos que Python inspeccione las 130 filas y extraiga solo las que nos importan en milisegundos.

Traducción de Misiones (Promedios)

Para hallar promedios en programación, necesitamos dos variables que memoricen datos mientras el ciclo gira.



El Mapa Mental del Científico de Datos



Checklist de Aprendizaje: Antes de irnos.

¿Puedes filtrar un dataframe a partir de sus índices utilizando la función `.iloc` y condicionales?

 Sí Parcialmente Aún no

¿Puedes calcular el promedio de una serie de datos en Python utilizando ciclos, acumuladores y contadores?

 Sí Parcialmente Aún no

Has pasado de buscar datos con la vista a comandar a una máquina para que lo haga por ti.
El mundo de los datos está ahora a tu alcance.