

# 7. Pandas

Introducción a Python

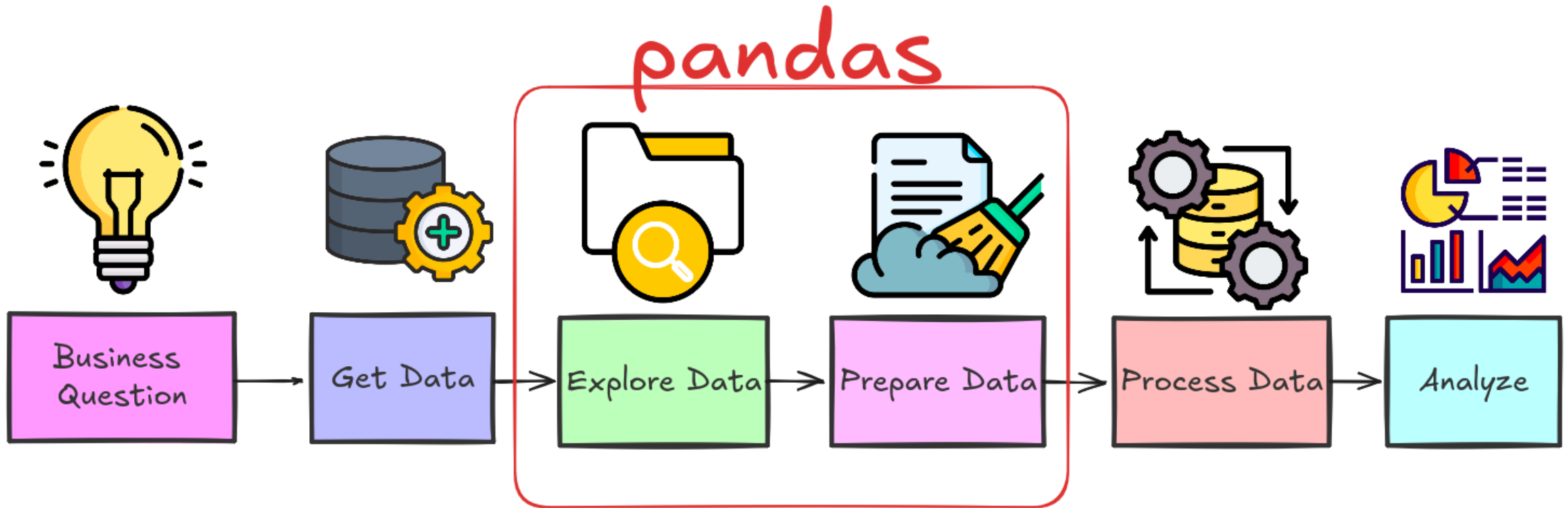
# Contenidos

1. Introducción
2. Entrada, salida y exploración
3. Transformación de datos
4. Referencias

# Contenidos

1. Introducción
2. Entrada, salida y exploración
3. Transformación de datos
4. Referencias

# 1. Introducción: Transformaciones de datos



# 1. Introducción: ¿Qué es?

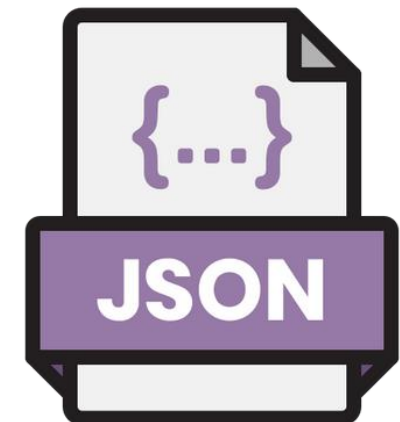
## Librería de python

- Manipulación de datos.
- Análisis de datos.
- Estructuras de datos eficientes.



## Propósito

- Análisis exploratorio rápido
- Grandes volúmenes de datos estructurados



# 1. Introducción: Elementos, datos tabulares

**Jupyter format**

**Standard Python format**

**Column labels**

**Index**

**Data**

**Pandas Series**

**Pandas DataFrame**

`pandas.core.frame.DataFrame`

	YEARMODA	TEMP	MAX	MIN
0	20160601	65.5	73.6	54.7
1	20160602	65.8	80.8	55.0
2	20160603	68.4	77.9	55.6
3	20160604	57.5	70.9	47.3
4	20160605	51.4	58.3	43.2
5	20160606	52.2	59.7	42.8
6	20160607	56.9	65.1	45.9
7	20160608	54.2	60.4	47.5
8	20160609	49.4	54.1	45.7
9	20160610	49.5	55.9	43.0

	YEARMODA	TEMP	MAX	MIN
0	20160601	65.5	73.6	54.7
1	20160602	65.8	80.8	55.0
2	20160603	68.4	77.9	55.6
3	20160604	57.5	70.9	47.3
4	20160605	51.4	58.3	43.2
5	20160606	52.2	59.7	42.8
6	20160607	56.9	65.1	45.9
7	20160608	54.2	60.4	47.5
8	20160609	49.4	54.1	45.7
9	20160610	49.5	55.9	43.0

**Standard Python format**

**Index**

**Data**

**Column label**

**Data type**

**Pandas Series**

`pandas.core.series.Series`

0	65.5
1	65.8
2	68.4
3	57.5
4	51.4
5	52.2
6	56.9
7	54.2
8	49.4
9	49.5

Name: TEMP, dtype: float64

# 1. Introducción: Instalación de la librería

(En nuestro entorno virtual de python)

```
pip install pandas
```

```
import pandas as pd
```

**¡AHORA HACEDLO VOSOTROS!**

# Contenidos

1. Introducción
2. Entrada, salida y exploración
3. Transformación de datos
4. Referencias



# Contenidos

1. Introducción
2. Entrada, salida y exploración
3. Transformación de datos
4. Referencias

## 2. Importar y exportar datos

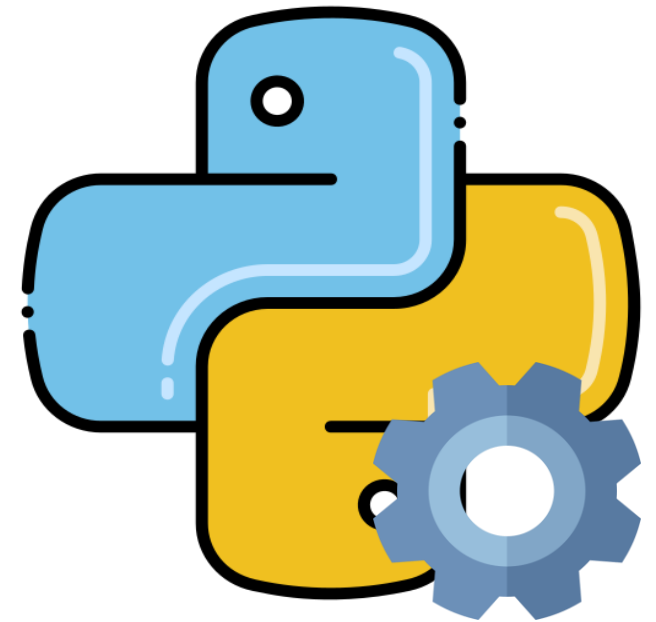
Archivo



`pd.read_ + format (csv, excel)`



Transformaciones



`pd.to_ + format (csv, excel)`



## 2. Importar y exportar datos: Consejos

### Lectura y salida

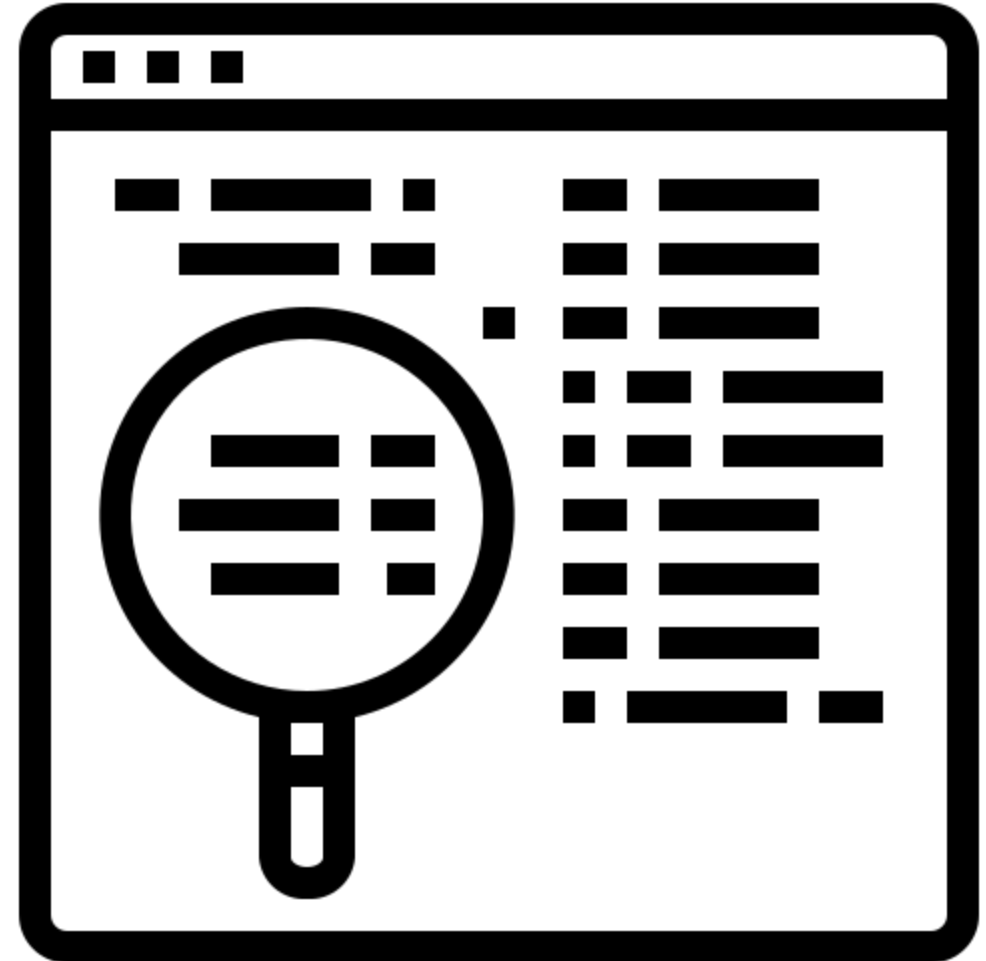
- Especificar correctamente la ruta al archivo
- Tener en cuenta el "encoding" (ver [metadatos](#))
- Tener en cuenta el delimitador (Ver CSV)
- Especificar índice
- En caso de usar datos muy grandes usar chunksize.

# Contenidos

1. Introducción
2. Entrada, salida
3. Exploración
4. Transformación de datos
5. Referencias

## 2. Exploración de datos: ¿Qué es?

- Es el análisis de un dataset.
- Implica comprender:
  - Estructura
  - Contenido
  - Características
  - Posibles errores
  - Valores atípicos



## 2. Exploración de datos: ¿Por qué hacerlo?

La exploración de datos es crucial porque:

- Ayuda a familiarizarse con el dataset.
- Permite identificar problemas
- Garantiza la calidad del análisis

De esta exploración surgirán luego las distintas transformaciones sobre los datos...

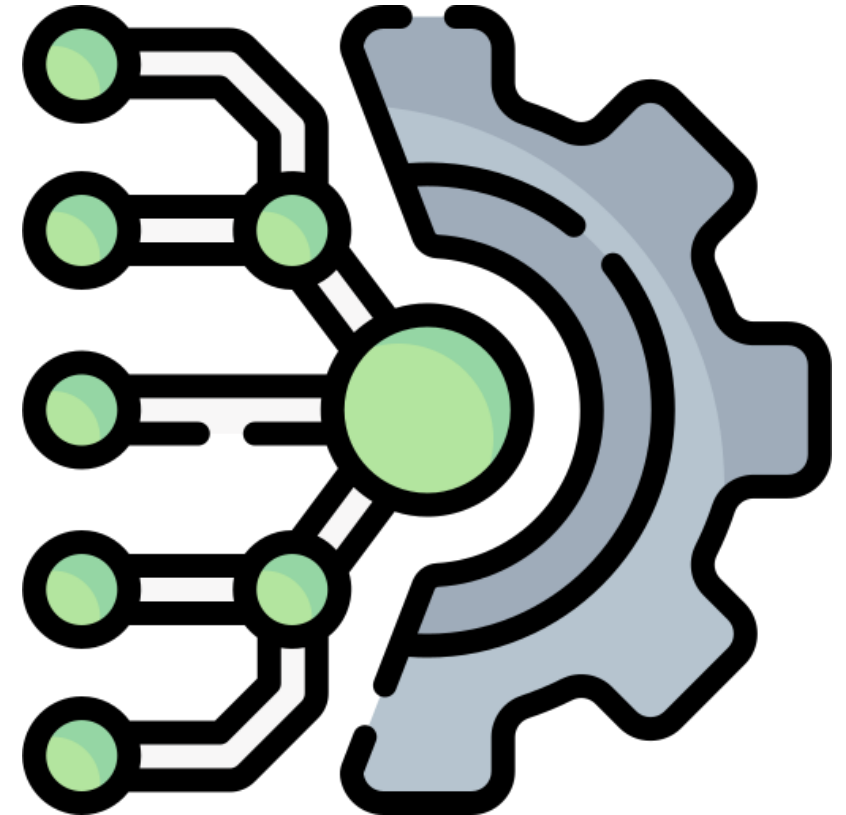


### 3. Transformaciones: ¿Qué cambiar?

Algunas cosas que pueden suponer un cambio son:

- Duplicados
- Sucios
- Errores
- Campos vacíos/nulos
- Información redundante
- Formato

¿Podría haber más?



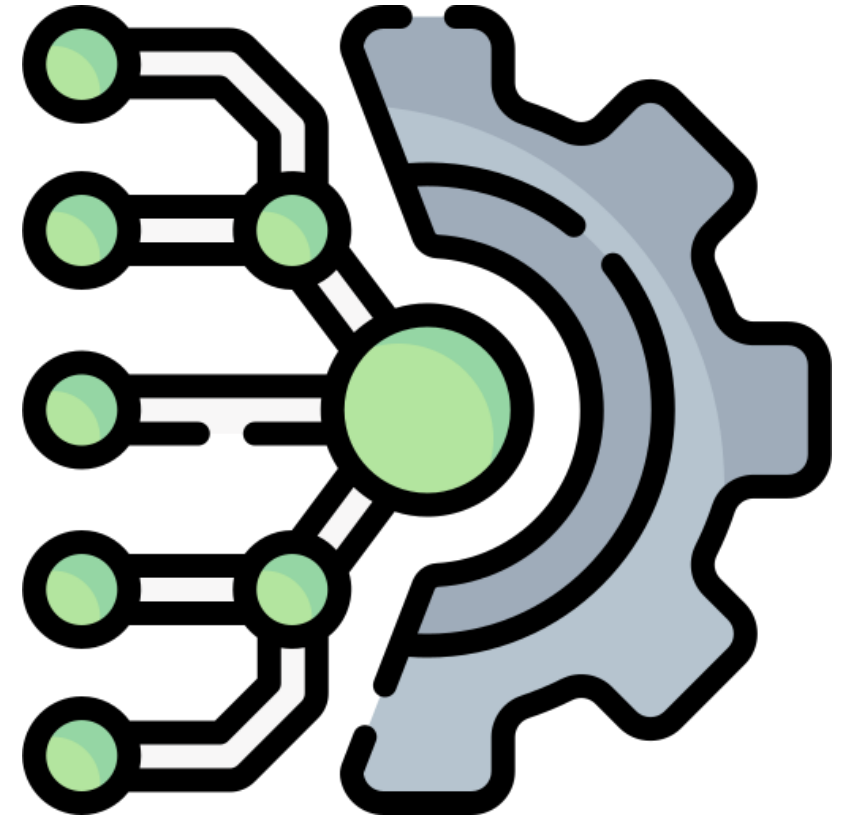
### 3. Transformaciones: ¿Qué cambiar?

Algunas cosas que pueden suponer un cambio son:

- Duplicados
- Sucios
- Errores
- Campos vacíos/nulos
- Información redundante
- Formato

Por supuesto:

One-hot encoding, normalización,...





### 3. Transformaciones: Duplicados

	Nombre	Edad	Ciudad	Salario Mensu	Salario Anuāl	Fecha
	Ana	23	Madrid	2500	30000	2022-10-12
	Luis	34	Barcelon	3000	36000	10/12/2022
	Pedro	-5	Madird	1500	18000	12/10/2022
	Luis	34	Barcelona	3000	36000	
	Marta	1000	Sevilla			2022/12/10
	Ana	45	Madrid		24000	2022-01-05
	Clara	29	Valencia	2700	32400	12/10/2020
	Pedro	29	Sevilla	1500	18000	2021-08-09
	Marta	50	Madrid	4000	48000	2022-02-12
	Carlos	60	Vigo	3500	42000	2022-03-15
	Lucía	27	Sevilla	2200	26400	2022-04-10
	Juan	32	Barcelona	2800	33600	2022-05-20
	Elena	-8	Valencia	3200	38400	2021-06-18
	Lucas	45	Madrid	4100	49200	2022-07-22
	Raúl	37	Barcelona	2800	33600	2022-08-30
	Clara	29	Madrid	2700	32400	2022-09-12
	Sofía	31	Sevilla	3200	38400	2022-10-11
	Jorge	28	Vigo	3400	40800	2023-01-15
	María	33	Madrid	3000	36000	2022-12-31
	Jorge	28	Vigo	3400	40800	2023-01-15

### 3. Transformaciones: Duplicados

	Nombre ▾	Edad ▾	Ciudad ▾	Salario Mensual	Salario Anual	Fecha ▾
	Ana	23	Madrid	2500	30000	2022-10-12
	Luis	34	Barcelon	3000	36000	10/12/2022
	Pedro	-5	Madird	1500	18000	12/10/2022
	Luis	34	Barcelona	3000	36000	
	Marta	1000	Sevilla			2022/12/10
	Ana	45	Madrid		24000	2022-01-05
	Clara	29	Valencia	2700	32400	12/10/2020
	Pedro	29	Sevilla	1500	18000	2021-08-09
	Marta	50	Madrid	4000	48000	2022-02-12
	Carlos	60	Vigo	3500	42000	2022-03-15
	Lucía	27	Sevilla	2200	26400	2022-04-10
	Juan	32	Barcelona	2800	33600	2022-05-20
	Elena	-8	Valencia	3200	38400	2021-06-18
	Lucas	45	Madrid	4100	49200	2022-07-22
	Raúl	37	Barcelona	2800	33600	2022-08-30
	Clara	29	Madrid	2700	32400	2022-09-12
	Sofía	31	Sevilla	3200	38400	2022-10-11
	Jorge	28	Vigo	3400	40800	2023-01-15
	María	33	Madrid	3000	36000	2022-12-31
	Jorge	28	Vigo	3400	40800	2023-01-15

# 3. Transformaciones: Sucios

	Nombre ▾	Edad ▾	Ciudad ▾	Salario Mensual	Salario Anual	Fecha ▾
	Ana	23	Madrid	2500	30000	2022-10-12
	Luis	34	Barcelon	3000	36000	10/12/2022
	Pedro	-5	Madird	1500	18000	12/10/2022
	Luis	34	Barcelona	3000	36000	
	Marta	1000	Sevilla			2022/12/10
	Ana	45	Madrid		24000	2022-01-05
	Clara	29	Valencia	2700	32400	12/10/2020
	Pedro	29	Sevilla	1500	18000	2021-08-09
	Marta	50	Madrid	4000	48000	2022-02-12
	Carlos	60	Vigo	3500	42000	2022-03-15
	Lucía	27	Sevilla	2200	26400	2022-04-10
	Juan	32	Barcelona	2800	33600	2022-05-20
	Elena	-8	Valencia	3200	38400	2021-06-18
	Lucas	45	Madrid	4100	49200	2022-07-22
	Raúl	37	Barcelona	2800	33600	2022-08-30
	Clara	29	Madrid	2700	32400	2022-09-12
	Sofía	31	Sevilla	3200	38400	2022-10-11
	Jorge	28	Vigo	3400	40800	2023-01-15
	María	33	Madrid	3000	36000	2022-12-31
	Jorge	28	Vigo	3400	40800	2023-01-15

### 3. Transformaciones: Sucios

	Nombre ▾	Edad ▾	Ciudad ▾	Salario Mensu	Salario Anual	Fecha ▾
	Ana	23	Madrid	2500	30000	2022-10-12
	Luis	34	Barcelon	3000	36000	10/12/2022
	Pedro	-5	Madird	1500	18000	12/10/2022
	Luis	34	Barcelona	3000	36000	
	Marta	1000	Sevilla			2022/12/10
	Ana	45	Madrid		24000	2022-01-05
	Clara	29	Valencia	2700	32400	12/10/2020
	Pedro	29	Sevilla	1500	18000	2021-08-09
	Marta	50	Madrid	4000	48000	2022-02-12
	Carlos	60	Vigo	3500	42000	2022-03-15
	Lucía	27	Sevilla	2200	26400	2022-04-10
	Juan	32	Barcelona	2800	33600	2022-05-20
	Elena	-8	Valencia	3200	38400	2021-06-18
	Lucas	45	Madrid	4100	49200	2022-07-22
	Raúl	37	Barcelona	2800	33600	2022-08-30
	Clara	29	Madrid	2700	32400	2022-09-12
	Sofía	31	Sevilla	3200	38400	2022-10-11
	Jorge	28	Vigo	3400	40800	2023-01-15
	María	33	Madrid	3000	36000	2022-12-31
	Jorge	28	Vigo	3400	40800	2023-01-15

# 3. Transformaciones: Errores

	Nombre ▾	Edad ▾	Ciudad ▾	Salario Mensual	Salario Anual	Fecha ▾
	Ana	23	Madrid	2500	30000	2022-10-12
	Luis	34	Barcelon	3000	36000	10/12/2022
	Pedro	-5	Madird	1500	18000	12/10/2022
	Luis	34	Barcelona	3000	36000	
	Marta	1000	Sevilla			2022/12/10
	Ana	45	Madrid		24000	2022-01-05
	Clara	29	Valencia	2700	32400	12/10/2020
	Pedro	29	Sevilla	1500	18000	2021-08-09
	Marta	50	Madrid	4000	48000	2022-02-12
	Carlos	60	Vigo	3500	42000	2022-03-15
	Lucía	27	Sevilla	2200	26400	2022-04-10
	Juan	32	Barcelona	2800	33600	2022-05-20
	Elena	-8	Valencia	3200	38400	2021-06-18
	Lucas	45	Madrid	4100	49200	2022-07-22
	Raúl	37	Barcelona	2800	33600	2022-08-30
	Clara	29	Madrid	2700	32400	2022-09-12
	Sofía	31	Sevilla	3200	38400	2022-10-11
	Jorge	28	Vigo	3400	40800	2023-01-15
	María	33	Madrid	3000	36000	2022-12-31
	Jorge	28	Vigo	3400	40800	2023-01-15

### 3. Transformaciones: Errores

	Nombre ▾	Edad ▾	Ciudad ▾	Salario Mensual	Salario Anual	Fecha ▾
	Ana	23	Madrid	2500	30000	2022-10-12
	Luis	34	Barcelon	3000	36000	10/12/2022
	Pedro	-5	Madird	1500	18000	12/10/2022
	Luis	34	Barcelona	3000	36000	
	Marta	1000	Sevilla			2022/12/10
	Ana	45	Madrid		24000	2022-01-05
	Clara	29	Valencia	2700	32400	12/10/2020
	Pedro	29	Sevilla	1500	18000	2021-08-09
	Marta	50	Madrid	4000	48000	2022-02-12
	Carlos	60	Vigo	3500	42000	2022-03-15
	Lucía	27	Sevilla	2200	26400	2022-04-10
	Juan	32	Barcelona	2800	33600	2022-05-20
	Elena	-8	Valencia	3200	38400	2021-06-18
	Lucas	45	Madrid	4100	49200	2022-07-22
	Raúl	37	Barcelona	2800	33600	2022-08-30
	Clara	29	Madrid	2700	32400	2022-09-12
	Sofía	31	Sevilla	3200	38400	2022-10-11
	Jorge	28	Vigo	3400	40800	2023-01-15
	María	33	Madrid	3000	36000	2022-12-31
	Jorge	28	Vigo	3400	40800	2023-01-15

### 3. Transformaciones: Campos nulos

	Nombre ▾	Edad ▾	Ciudad ▾	Salario Mensual	Salario Anual	Fecha ▾
	Ana	23	Madrid	2500	30000	2022-10-12
	Luis	34	Barcelon	3000	36000	10/12/2022
	Pedro	-5	Madird	1500	18000	12/10/2022
	Luis	34	Barcelona	3000	36000	
	Marta	1000	Sevilla			2022/12/10
	Ana	45	Madrid		24000	2022-01-05
	Clara	29	Valencia	2700	32400	12/10/2020
	Pedro	29	Sevilla	1500	18000	2021-08-09
	Marta	50	Madrid	4000	48000	2022-02-12
	Carlos	60	Vigo	3500	42000	2022-03-15
	Lucía	27	Sevilla	2200	26400	2022-04-10
	Juan	32	Barcelona	2800	33600	2022-05-20
	Elena	-8	Valencia	3200	38400	2021-06-18
	Lucas	45	Madrid	4100	49200	2022-07-22
	Raúl	37	Barcelona	2800	33600	2022-08-30
	Clara	29	Madrid	2700	32400	2022-09-12
	Sofía	31	Sevilla	3200	38400	2022-10-11
	Jorge	28	Vigo	3400	40800	2023-01-15
	María	33	Madrid	3000	36000	2022-12-31
	Jorge	28	Vigo	3400	40800	2023-01-15

### 3. Transformaciones: Campos nulos

	Nombre ▾	Edad ▾	Ciudad ▾	Salario Mensual	Salario Anual	Fecha ▾
	Ana	23	Madrid	2500	30000	2022-10-12
	Luis	34	Barcelon	3000	36000	10/12/2022
	Pedro	-5	Madird	1500	18000	12/10/2022
	Luis	34	Barcelona	3000	36000	
	Marta	1000	Sevilla			2022/12/10
	Ana	45	Madrid		24000	2022-01-05
	Clara	29	Valencia	2700	32400	12/10/2020
	Pedro	29	Sevilla	1500	18000	2021-08-09
	Marta	50	Madrid	4000	48000	2022-02-12
	Carlos	60	Vigo	3500	42000	2022-03-15
	Lucía	27	Sevilla	2200	26400	2022-04-10
	Juan	32	Barcelona	2800	33600	2022-05-20
	Elena	-8	Valencia	3200	38400	2021-06-18
	Lucas	45	Madrid	4100	49200	2022-07-22
	Raúl	37	Barcelona	2800	33600	2022-08-30
	Clara	29	Madrid	2700	32400	2022-09-12
	Sofía	31	Sevilla	3200	38400	2022-10-11
	Jorge	28	Vigo	3400	40800	2023-01-15
	María	33	Madrid	3000	36000	2022-12-31
	Jorge	28	Vigo	3400	40800	2023-01-15



### 3. Transformaciones: Campos redundantes

	Nombre ▾	Edad ▾	Ciudad ▾	Salario Mensu.	Salario Anual	Fecha ▾
	Ana	23	Madrid	2500	30000	2022-10-12
	Luis	34	Barcelon	3000	36000	10/12/2022
	Pedro	-5	Madird	1500	18000	12/10/2022
	Luis	34	Barcelona	3000	36000	
	Marta	1000	Sevilla			2022/12/10
	Ana	45	Madrid		24000	2022-01-05
	Clara	29	Valencia	2700	32400	12/10/2020
	Pedro	29	Sevilla	1500	18000	2021-08-09
	Marta	50	Madrid	4000	48000	2022-02-12
	Carlos	60	Vigo	3500	42000	2022-03-15
	Lucía	27	Sevilla	2200	26400	2022-04-10
	Juan	32	Barcelona	2800	33600	2022-05-20
	Elena	-8	Valencia	3200	38400	2021-06-18
	Lucas	45	Madrid	4100	49200	2022-07-22
	Raúl	37	Barcelona	2800	33600	2022-08-30
	Clara	29	Madrid	2700	32400	2022-09-12
	Sofía	31	Sevilla	3200	38400	2022-10-11
	Jorge	28	Vigo	3400	40800	2023-01-15
	María	33	Madrid	3000	36000	2022-12-31
	Jorge	28	Vigo	3400	40800	2023-01-15

### 3. Transformaciones: Campos redundantes

	Nombre ▾	Edad ▾	Ciudad ▾	Salario Mensu	Salario Anual	Fecha ▾
	Ana	23	Madrid	2500	30000	2022-10-12
	Luis	34	Barcelon	3000	36000	10/12/2022
	Pedro	-5	Madird	1500	18000	12/10/2022
	Luis	34	Barcelona	3000	36000	
	Marta	1000	Sevilla			2022/12/10
	Ana	45	Madrid		24000	2022-01-05
	Clara	29	Valencia	2700	32400	12/10/2020
	Pedro	29	Sevilla	1500	18000	2021-08-09
	Marta	50	Madrid	4000	48000	2022-02-12
	Carlos	60	Vigo	3500	42000	2022-03-15
	Lucía	27	Sevilla	2200	26400	2022-04-10
	Juan	32	Barcelona	2800	33600	2022-05-20
	Elena	-8	Valencia	3200	38400	2021-06-18
	Lucas	45	Madrid	4100	49200	2022-07-22
	Raúl	37	Barcelona	2800	33600	2022-08-30
	Clara	29	Madrid	2700	32400	2022-09-12
	Sofía	31	Sevilla	3200	38400	2022-10-11
	Jorge	28	Vigo	3400	40800	2023-01-15
	María	33	Madrid	3000	36000	2022-12-31
	Jorge	28	Vigo	3400	40800	2023-01-15

### 3. Transformaciones: Formato

	Nombre ▾	Edad ▾	Ciudad ▾	Salario Mensu	Salario Anual	Fecha ▾
	Ana	23	Madrid	2500	30000	2022-10-12
	Luis	34	Barcelon	3000	36000	10/12/2022
	Pedro	-5	Madird	1500	18000	12/10/2022
	Luis	34	Barcelona	3000	36000	
	Marta	1000	Sevilla			2022/12/10
	Ana	45	Madrid		24000	2022-01-05
	Clara	29	Valencia	2700	32400	12/10/2020
	Pedro	29	Sevilla	1500	18000	2021-08-09
	Marta	50	Madrid	4000	48000	2022-02-12
	Carlos	60	Vigo	3500	42000	2022-03-15
	Lucía	27	Sevilla	2200	26400	2022-04-10
	Juan	32	Barcelona	2800	33600	2022-05-20
	Elena	-8	Valencia	3200	38400	2021-06-18
	Lucas	45	Madrid	4100	49200	2022-07-22
	Raúl	37	Barcelona	2800	33600	2022-08-30
	Clara	29	Madrid	2700	32400	2022-09-12
	Sofía	31	Sevilla	3200	38400	2022-10-11
	Jorge	28	Vigo	3400	40800	2023-01-15
	María	33	Madrid	3000	36000	2022-12-31
	Jorge	28	Vigo	3400	40800	2023-01-15

### 3. Transformaciones: Formato

	Nombre ▾	Edad ▾	Ciudad ▾	Salario Mensual	Salario Anual	Fecha ▾
	Ana	23	Madrid	2500	30000	2022-10-12
	Luis	34	Barcelon	3000	36000	10/12/2022
	Pedro	-5	Madird	1500	18000	12/10/2022
	Luis	34	Barcelona	3000	36000	
	Marta	1000	Sevilla			2022/12/10
	Ana	45	Madrid		24000	2022-01-05
	Clara	29	Valencia	2700	32400	12/10/2020
	Pedro	29	Sevilla	1500	18000	2021-08-09
	Marta	50	Madrid	4000	48000	2022-02-12
	Carlos	60	Vigo	3500	42000	2022-03-15
	Lucía	27	Sevilla	2200	26400	2022-04-10
	Juan	32	Barcelona	2800	33600	2022-05-20
	Elena	-8	Valencia	3200	38400	2021-06-18
	Lucas	45	Madrid	4100	49200	2022-07-22
	Raúl	37	Barcelona	2800	33600	2022-08-30
	Clara	29	Madrid	2700	32400	2022-09-12
	Sofía	31	Sevilla	3200	38400	2022-10-11
	Jorge	28	Vigo	3400	40800	2023-01-15
	María	33	Madrid	3000	36000	2022-12-31
	Jorge	28	Vigo	3400	40800	2023-01-15

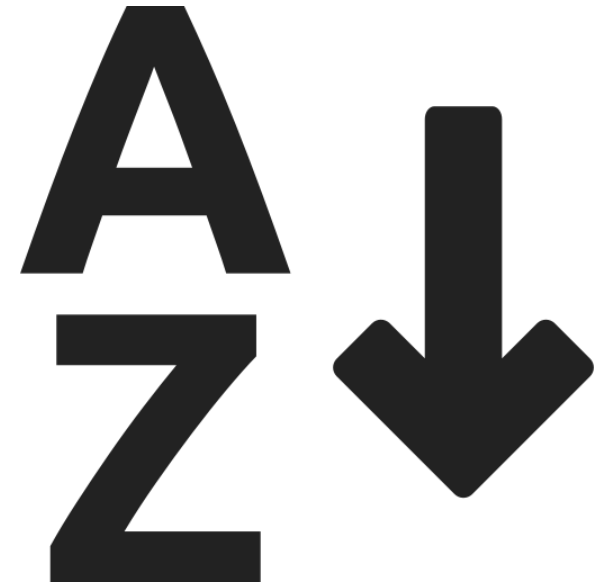
# 3. Transformaciones: Importancia del orden

Es importante el orden de las transformaciones.

Ejemplo de nulos:

- Rellenar vacíos de sueldos con la media
- Eliminar filas con vacíos

¿Qué pasa si cambiamos el orden?



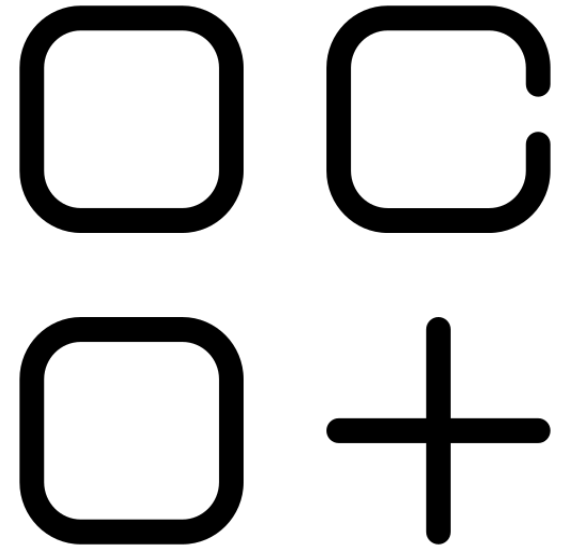
### 3. Transformaciones: Hay más

Las transformaciones que hemos visto no son las únicas. ¡¡Hay muchas más!!

La idea es saber cuáles hacer y por qué.

Dos últimas interesantes usadas en IA:

**Normalización y One-hot encoding**



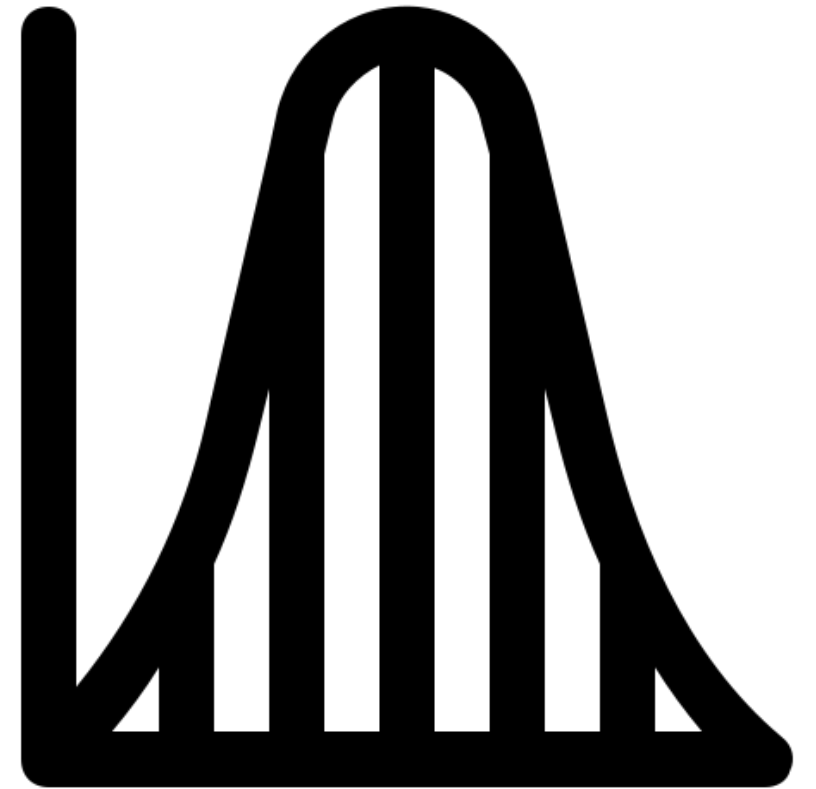
# 3. Transformaciones: Normalización

Ajustar rango de valores numéricos a una escala común.

Útil para ML. Modelos sensibles a magnitud de datos

Hay muchas formas, pero la que vamos a usar es la fórmula min-máx

$$norm(x) = \frac{x - min(x)}{max(x) - min(x)}$$



# 3. Transformaciones: One-hot encoding

- Transformar variables categóricas a numéricas.
- Una columna por categoría y usamos binario (0 ó 1)
- Adecuado para ML.
- Tener en cuenta problema de **ordinalidad falsa**.

Color
Verde
Azul
Rojo

Verde	Azul	Rojo
1	0	0
0	1	0
0	0	1



# X. Referencias finales

- [Documentación oficial de Pandas](#)
- [Kaggle Learn Pandas](#)
- [W3schools Pandas tutorial](#)
- [FreeCodeCamp Youtube Tutorial](#)

# 7. Pandas

Introducción a Python