# **TASCA 4: TRANSFORMACIÓ DE DADES**

### Després d’haver realitzat la **extracció de dades** de diverses fonts (bases de dades, APIs, fitxers), el següent pas clau en el procés ETL (Extracció, Transformació i Càrrega) és la **transformació**. Aquesta fase consisteix a convertir les dades en brut en un format net, coherent i útil per a les necessitats analítiques o operatives de l’organització.

### Durant la transformació, s’apliquen operacions com la neteja de valors nuls o incorrectes, la normalització de formats, la conversió de tipus de dades, l’agregació d’informació i l’enriquiment amb fonts externes.

### Objectiu

Aplicar tècniques de transformació de dades utilitzant **Pandas** per preparar dades extretes de diferents fonts (MySQL, API d’Odoo i arxiu CSV) amb l’objectiu de:

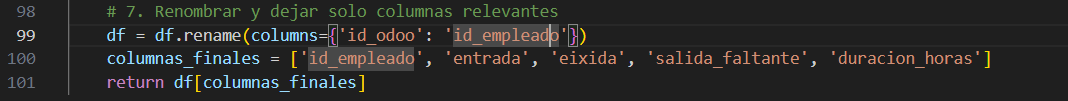
* Millorar la **qualitat de les dades**.
* Deixar les dades preparades per a la seva càrrega o anàlisi posterior.

### Requisits previs

* Entorn Python preparat (Docker ETL)
* Dispondre d’un conjunt de dades que prové de les diferents fonts d’informació ( Tasca3)
* Coneixements bàsics de Python y Pandas

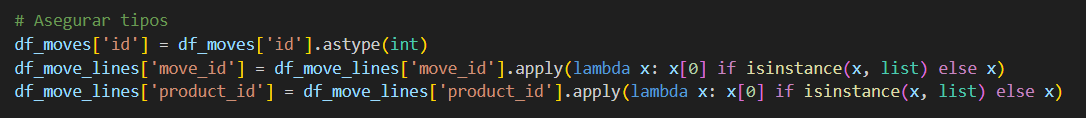
### Procediment pas a pas

1. Analitza i revisa les dades inicials obtingudes en la Tasca 3 (Extracció) i detecta anomalies, inconsistències, registres duplicats, etc. Defineix totes les operacions a realitzar en les dades
2. Fase de neteja de dades: Elimina i corregeix la informació incorrecta, incompleta o redundant



*Exemple: Eliminació de columnas no rellevants*

1. Aplica normalització en les dades: És necessari estandarditzar formats per tal de que les dades siguen consistents i comparables entre fonts.



*Exemple: Estandarització del tipus de data per las IDs*

1. Transformació de dades: Creació de camps derivats, agregacions d’informació, unió de taules que preparen les dades per al seu anàlisi o càrrega posterior en un Datawarehouse



*Exemple: Creació d’un camp derivat en un Dataframe*



*Exemple: Unio de dataframes mitjançant Pandas*

### Notes Importants

* **Planifica abans de transformar**: Abans d’aplicar cap canvi, és fonamental revisar a fons les dades extretes i decidir quines operacions cal aplicar per garantir la qualitat i coherència de la informació.
* **Comenta el codi**: El teu notebook o script ha d’incloure comentaris clars que expliquen què fa cada bloc de codi i amb quin propòsit.

# ENTREGA

La prova d’entrega de la Tasca 4 inclourà:

* Codi font .py, amb totes les operacions aplicades pas a pas.
* Informe amb el resum dels problemes detectats en les dades, la relació d’operacions realitzades en les 4 fases exposades en el Procediment pas a pas i la explicació de tècniques de programació utilitzades.