① Number ⟶ Calc : sum(), max() _ _ _ , Filter

② String / Object ⟶ Filter, count, groupby()

↳ string function can be applied

df['Col'].str.func()

lower()

split()

⋮

③ Date

↳ df['Date'].dt.func()

year

month

Current date

import datetime

Lifecycle

70%

1. Problem understanding
2. Data Collection
3. Data Preprocessing
   - a. Fixing Rows and Columns
   - b. Handling Duplicates
   - c. —— Missing values
   - d. —— Outliers

a. Rows/Columns

1. Delete → df.drop(columns=['Col1','Col2']) axis = 0
2. Add → df['New Col'] = Calc on existing col
3. Rename → df.rename(columns={'old':'New'})
4. Datatype
   - astype(int/float)
   - to_datetime( )

Store	Profit
INDIA	1000 INR
USA	10 USD
CANADA	50 CAD
~	100 USD
=	10 USD
=	1000 INR

→ complex datatype

↳ Rule to make it atomic (number or text)

Step 1 : Split

Step 2 : Datatype Conversion

Step 3 : Currency conversion

- replace()