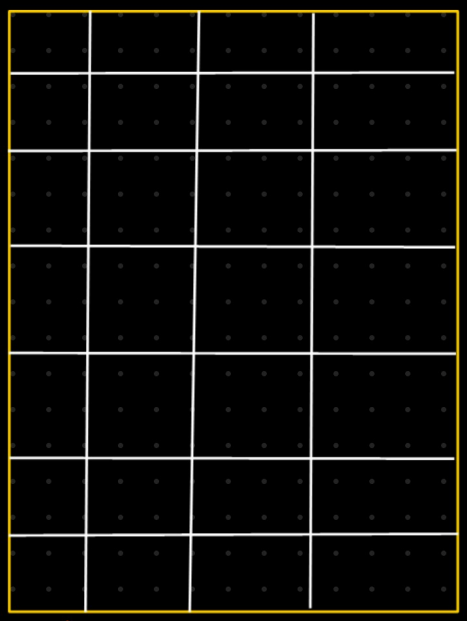
08-06-2025

Agenda - Pandas - II

subset = df.iloc[0:2]

↳ rows

we always get a copy (not a reference)



→ df

subset = df.loc[0:1]

subset → change → reflect

view

↓ reflect

another developer → swap

index - numeric

→ df → 150 columns → processed

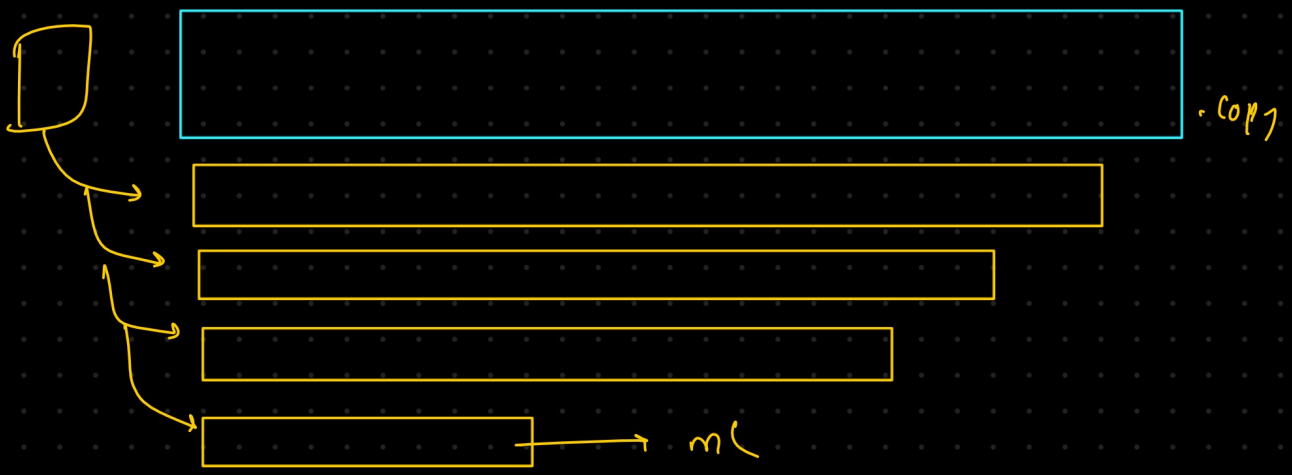
0, 1 → "name", "Age"

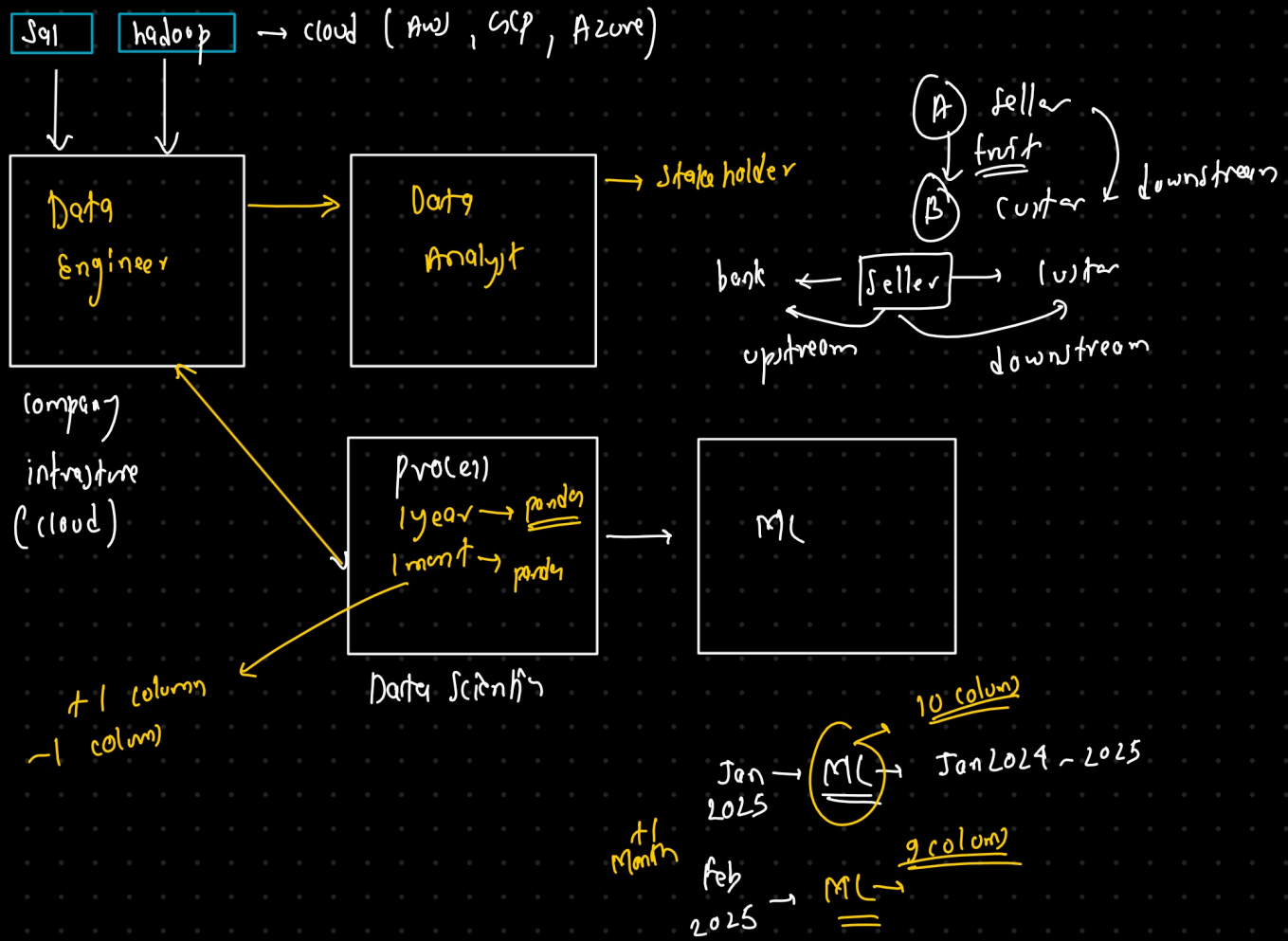
↓

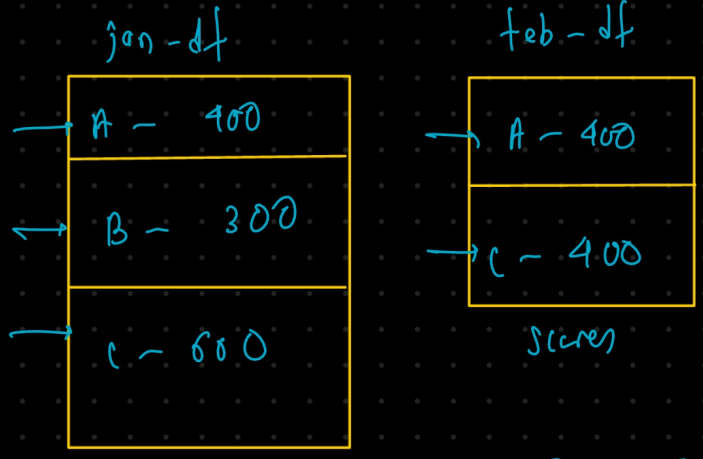
2 colomn → ~~iloc~~

loc → reference ↓ NO

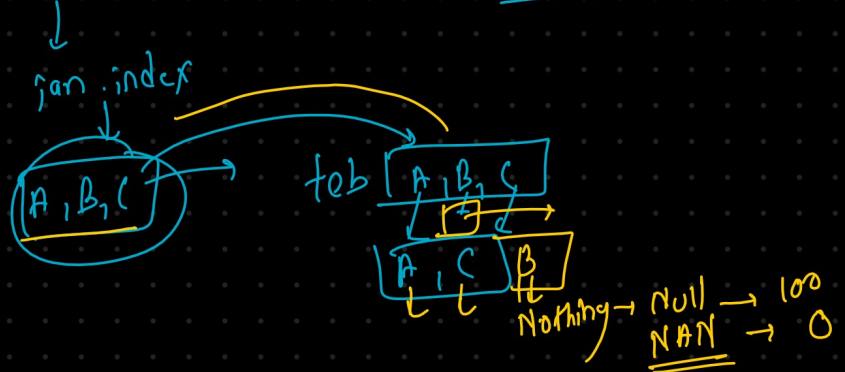
subset = df.loc[condition, column].copy()


· copy
ML


Sql
hadoop
→ cloud (AWS, GCP, Azure)
Data Engineer
Data Analyst
→ stake holder
company infrastructure (cloud)
process
1 year → pandas
1 month → pandas
ML
Data Scientist
+1 column
-1 column
A seller
first
B customer
downstream
bank ← Seller → customer
upstream
downstream
10 column
Jan 2025 → ML → Jan 2024 - 2025
+1 Month
Feb 2025 → ML
9 colomn



jan[scores] - feb[scores]

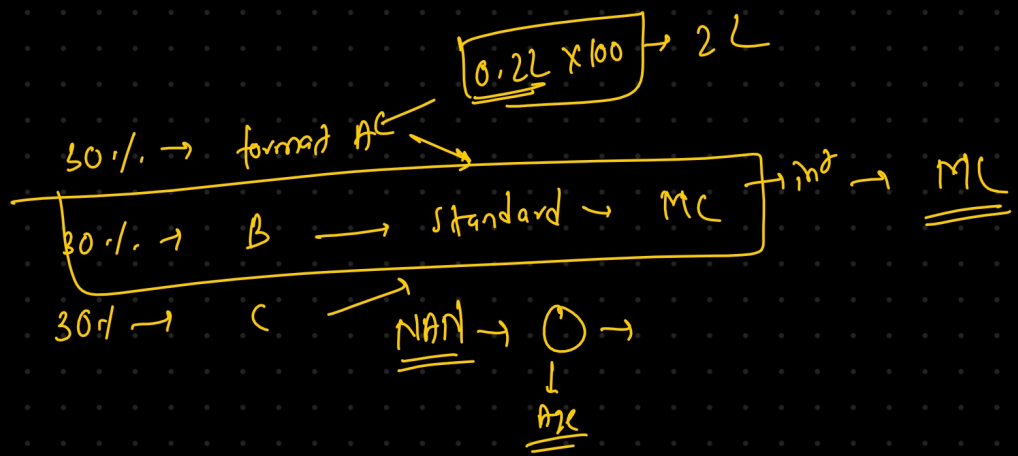


Jan	Feb
A	A
B	C
C	D

Common = index → A, B, C, D

→ Jan.reindex(common, 0)

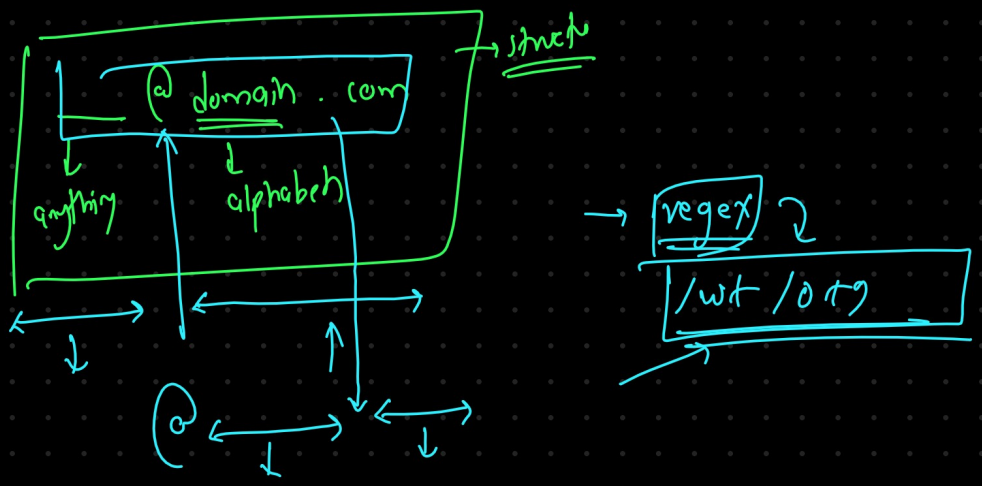
→ feb.reindex(common, 0)



Task → To check valid email

1) abcdefgh@gmail.com
2) defghijklmn@outlook.com
3) abc@yahoo.com
4) def@ya#.ai
5) abc@y.com.in.@gmail.coms

} suspicious



Assignment :—

https://github.com/Monalsingh/Basic---analysis

* Download as zip
* project-1 only
* Perform analysis on .csv file present
  * To under more about data go to url mentioned in notebook.
  * Analysis should be done on notebook.
    * Notebook should have image.
    * create a table & provide info about raw dataset.
    * Do some interesting analysis.