Pandas 2

Monday, November 3, 2025 9:03 PM

Agenda

- Pandas → CSV
  - Series } index
  - DF }
  - head, tail, info, describe
  - col < [ ]
    - rename
    - add
    - drop
    } {inplace, axis}
  - unique
  - value_counts
  - Loc and iloc

- Missing value < filling, dropna
- Concat
- Merge
- Sort
- apply
- groupby < filter, apply
- Melt
- pivot_table
- Cut
- Strm
- DT
- .gg.

} 20 Min

Loc and Iloc

[7 , 4 , "Hello" , 3]

0 1 2 3

-4 -3 -2 -1

} iloc

df =

	C1	C2	C3	C4	C5
-5 / 0	R1				
-4 / 1	R2				
-3 / 2	R3				
-2 / 3	R4				
-1 / 4	R5				

implicit

implicit

0 1 2

Restaurant	Votes	Rating

implicit index

	Restaurant	Votes	Rating
0	R1	120	4.1
1	R2	240	4.5
2	R3	150	3.9
3	R2	240	4.5
4	R4	80	4.0
5	R1	120	4.1

From <https://ofe5cua09bb-496ff2e9c6d22116-0-colab.googleusercontent.com/outputframe.html?vrz=colab-external_20251030-060104_RC00_825959784>

R , C

2 : 4 , 0 : 2

10 T

Loc explict

iloc Hidde

	Restaurant	Votes	Rating
a	R1	120	4.1
b	R2	240	4.5
c	R3	150	3.9
d	R2	240	4.5
e	R4	80	4.0
f	R1	120	4.1

From <https://ofe5cua09bb-496ff2e9c6d22116-0-colab.googleusercontent.com/outputframe.html?vrz=colab-external_20251030-060104_RC00_825959784>

R , C

df.Loc[c : e , Rest : Rating]

x x

c : d , Rest : Votes

~~Loc~~

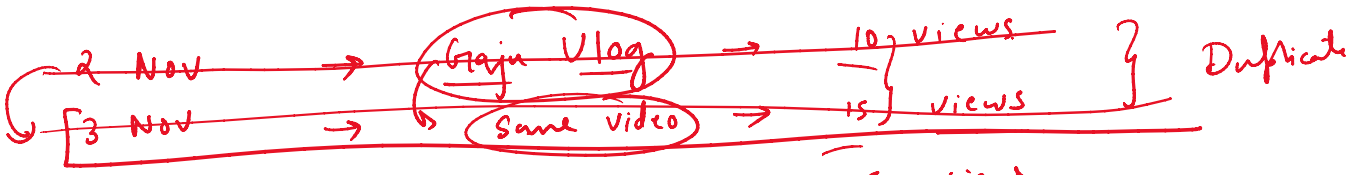
Loc and ILoc

index

Name

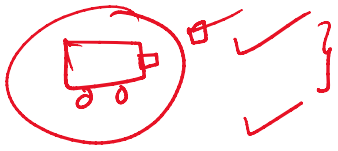
Hidden i

YOUTUBE



3 NOV same video → 15 views / 25 views ✗

Number ke →



duplicated

duplicated — drop duplicates

subset

inplace

keep — first → keep first

last →

False → remove all

agg

Sort_values

product_id

From <https://ofe5cua09bb-496ff2e9c6d22116-0-colab.googleusercontent.com/outputframe.html?vrz=colab-external_20251030-060104_RC00_825959784>

Ref
1 32 Sumit_cool 31 Super !!!!!!! Placed

product_id	user_id	user_name	review_id	review_title	review_content	order_timestamp

order.csv.

From <https://ofe5cua09bb-496ff2e9c6d22116-0-colab.googleusercontent.com/outputframe.html?vrz=colab-external_20251030-060104_RC00_825959784>

Ref
1 Sam 23 24 elec 36 2.5L 4.1 700 Des

product_id	product_name	category	discounted_price	actual_price	rating	rating_count	about_product	img_link	product_link

From <https://ofe5cua09bb-496ff2e9c6d22116-0-colab.googleusercontent.com/outputframe.html?vrz=colab-external_20251030-060104_RC00_825959784>

Product cs

	product_id	user_id	review_id

	product_id	product_name	category

	product_id	user_id	review_id
0	P001	U001	R001
1	P002	U002	R002
2	P003	U003	R003

From <https://ofe5cux09bb-496ff2e9c6d22116-0-colab.googleusercontent.com/outputframe.html?vrz=colab-external_20251030-060104_RC00_825959784>

pr

	product_id	product_name	category
0	P001	Product A	Category 1
1	P002	Product B	Category 2
2	P004	Product D	Category 3

From <https://ofe5cux09bb-496ff2e9c6d22116-0-colab.googleusercontent.com/outputframe.html?vrz=colab-external_20251030-060104_RC00_825959784>

(or, pr) on Productid

	product_id	user_id	review_id	product_name	category
0	P001	U001	R001	Product A	Category 1
1	P002	U002	R002	Product B	Category 2

From <https://ofe5cux09bb-496ff2e9c6d22116-0-colab.googleusercontent.com/outputframe.html?vrz=colab-external_20251030-060104_RC00_825959784>

	product_id	user_id	review_id
0	P001	U001	R001
1	P002	U002	R002
2	P003	U003	R003

From <https://ofe5cux09bb-496ff2e9c6d22116-0-colab.googleusercontent.com/outputframe.html?vrz=colab-external_20251030-060104_RC00_825959784>

	product_id	product_name	category
0	P001	Product A	Category 1
1	P002	Product B	Category 2
2	P004	Product D	Category 3

From <https://ofe5cux09bb-496ff2e9c6d22116-0-colab.googleusercontent.com/outputframe.html?vrz=colab-external_20251030-060104_RC00_825959784>

Left

Right

pd. merge(df1, df2, how="left")

	product_id	user_id	review_id	product_name	category
0	P001	U001	R001	Product A	Category 1
1	P002	U002	R002	Product B	Category 2
2	P003	U003	R003	NaN	NaN

From <https://ofe5cux09bb-496ff2e9c6d22116-0-colab.googleusercontent.com/outputframe.html?vrz=colab-external_20251030-060104_RC00_825959784>

empty

Vectorization

def f1

np. vec (f)

new

new – [array

apply

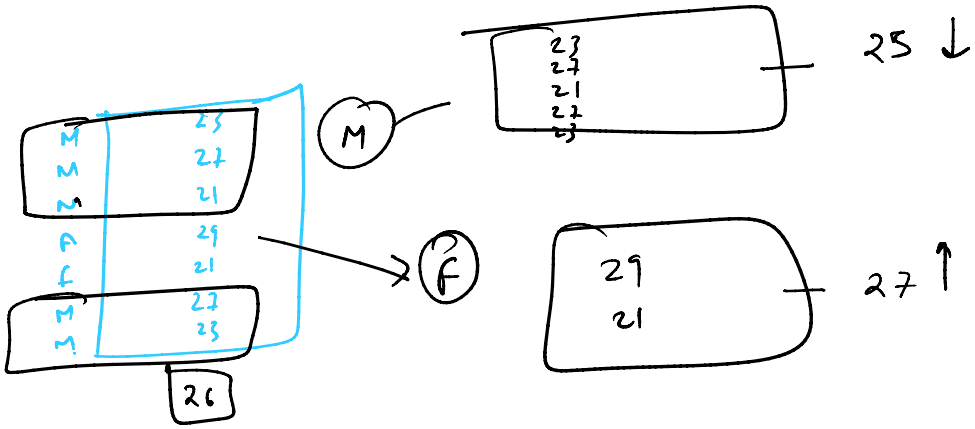
df['col'].apply(f1)

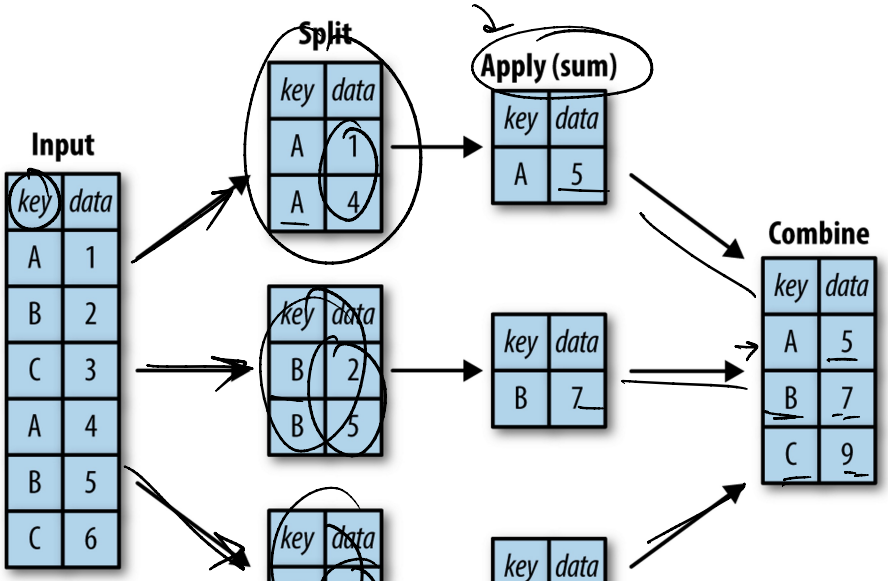
new -Larray J

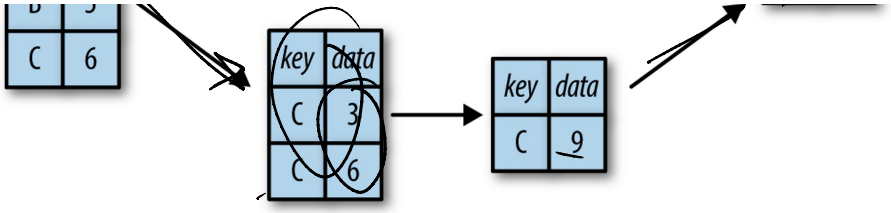
lambda

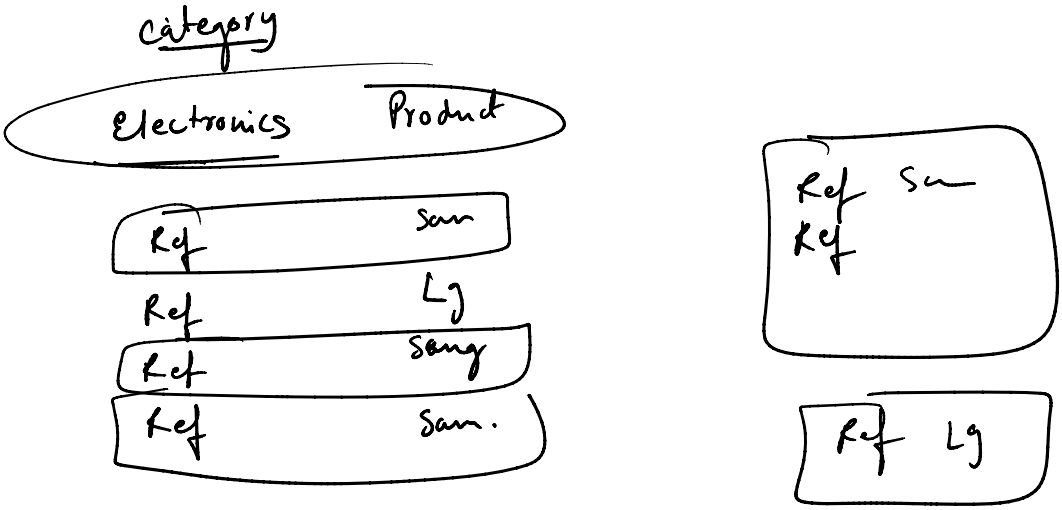
apply

Group by





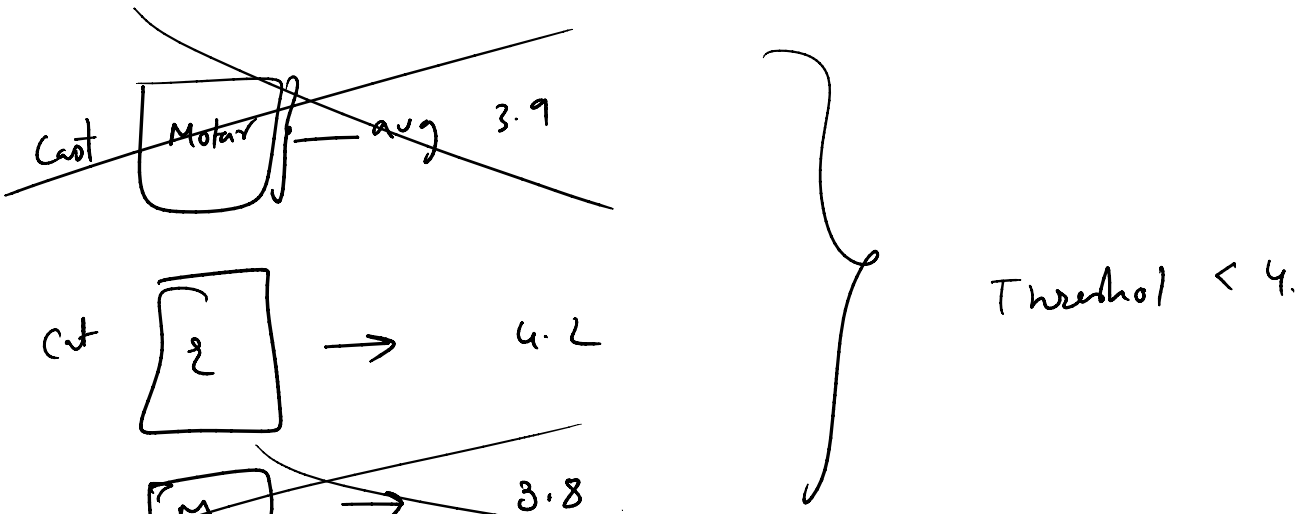


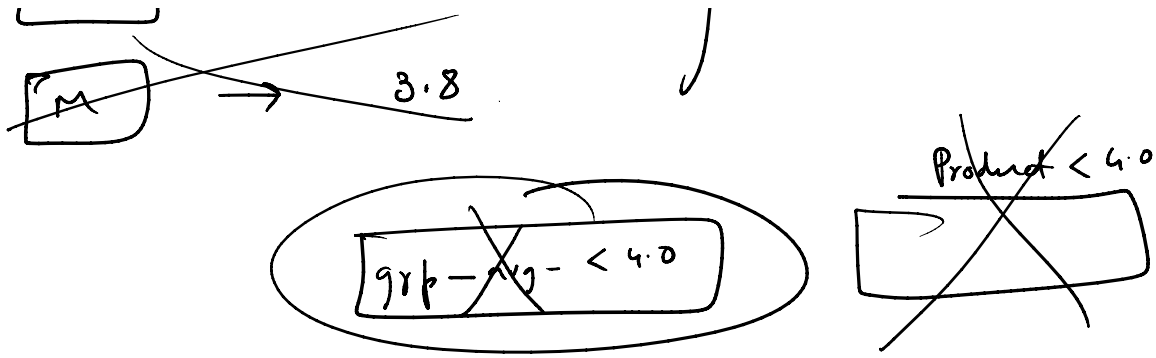


{Review analysis}

user should > 15 Review

1 Review





elec { Epb — 3.9 ; Lap — 4.1 ; M — 4.2 } 4.2 ✓

Hour { Mixer — 3.9 ; grider — 4.7 ; gas — 3.1 } 3.7 < 4.0

15 — 20 — channel

→ { Session — Pandas } optional
case study

M W F

+ Pandas Mrinal Ansul } Doubt