Agenda

World bank CO2

easy - medium

1/2 → 1 hr

- Restructuring Data
- Dealing with Missing Values (NaN)
- String Method in Pandas
- Handling Datetime
- Writing to a File.

Assignments

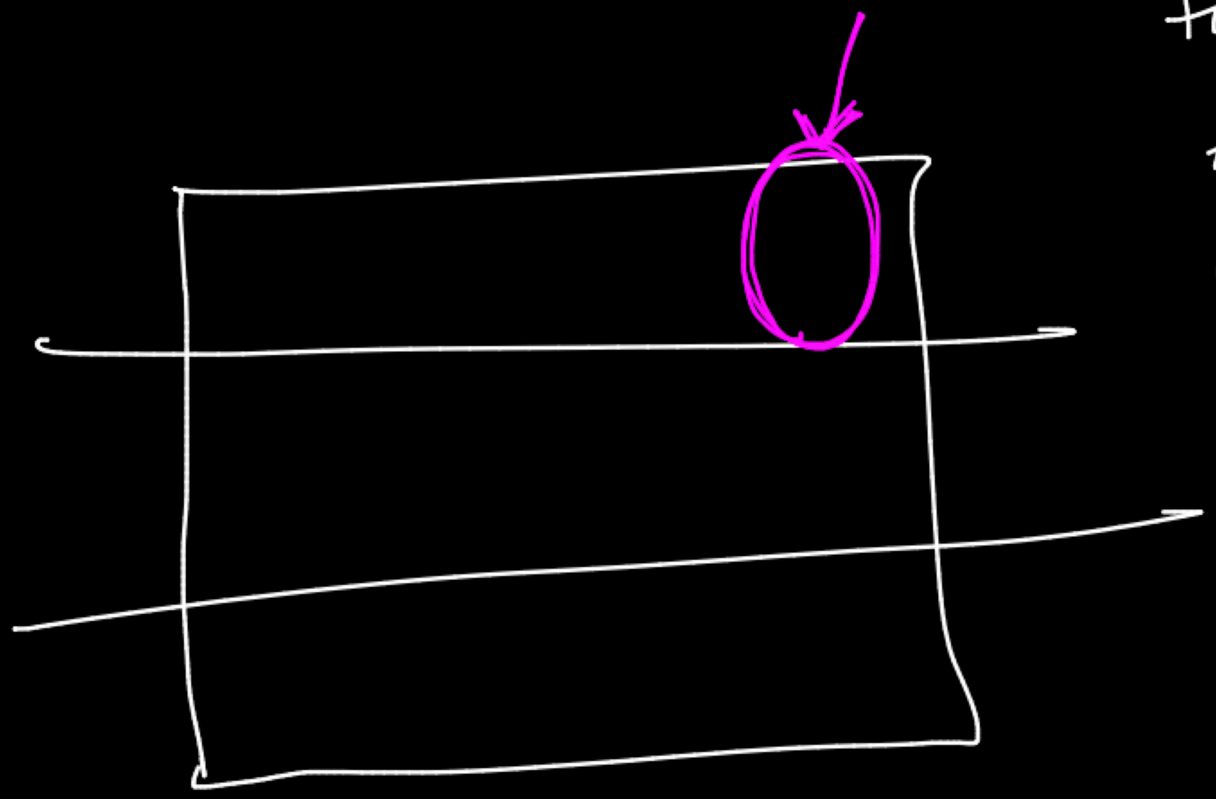
Goal → Data FA!

Grouping

Group based Aggr

filter

Apply



>=0

d1

60m

50m

budget - avg >=0

risky

60 - 50 >=0

True

False

d_1 60

m_1
m_2
m_3
50 m

Fall
not risky
40 — 50m
>=0

lst = [('title', 'count') , ('year', min) , ('year', ('max'))]

for i in lst:
print(i)

('title', 'count')
('year', min)

(_).join( )

risky

True

Restructering Data

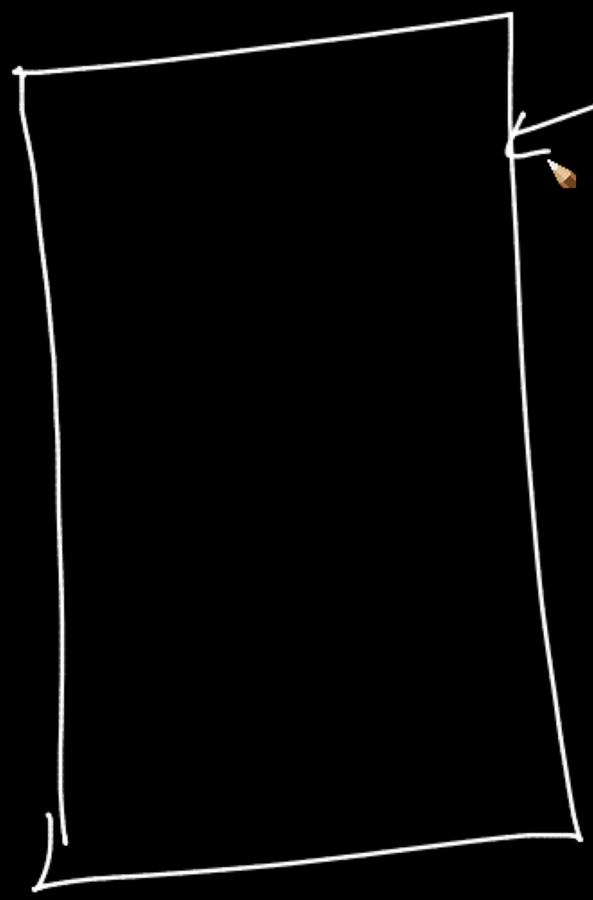
18 X 15

long format

wide format

1:30 1:45 2:0 - - - - - -

1:30 2:30 3:30



world bank co2

X

Name

1960 1961 - - - - - - 2003 2004 200

co2

Name

Year

CO2 Emissions

1960

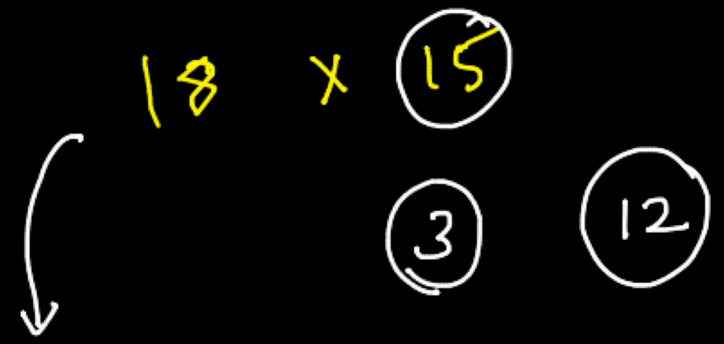
1961

Q 2.

Date Drug-Name Paramer

melt

1:30 2:30 3:30 - -- -



melt( )

2

18 x 12 => (216 , 5)

long → wide

data_melt

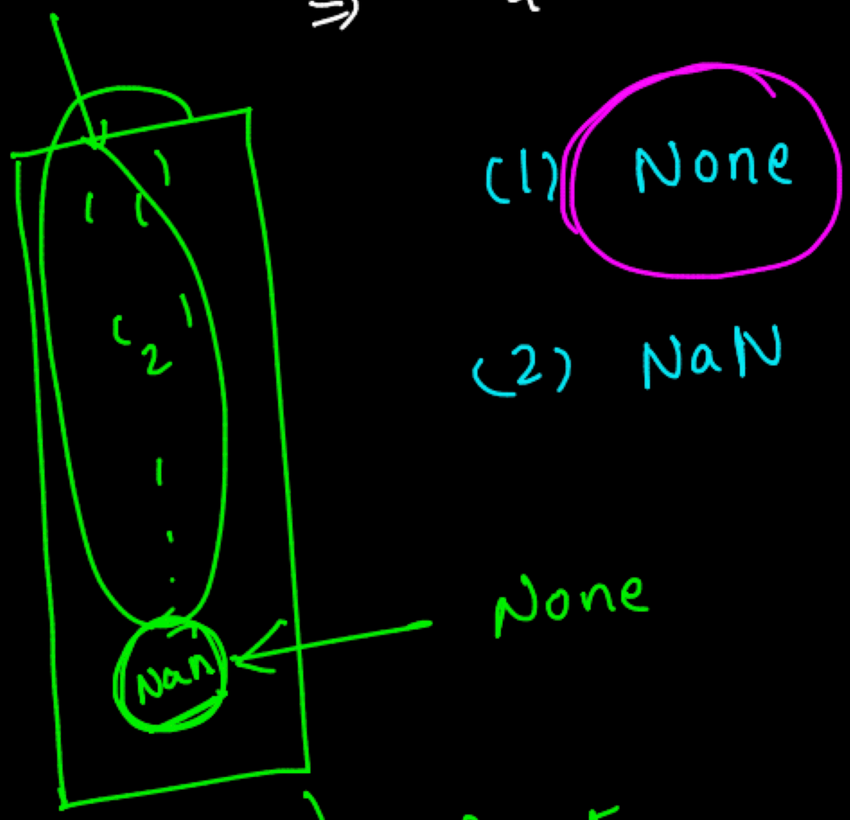
pivot( )

216 / 2 ⇒ 108

NaN → Missing values

empty cell / no data

⇒ 2 kinds missing values.



(1) None

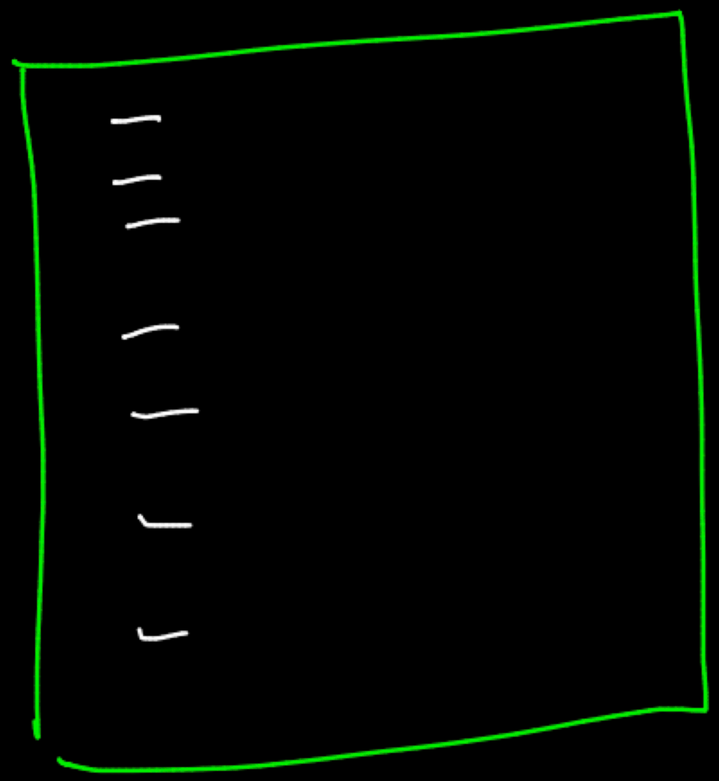
(2) NaN

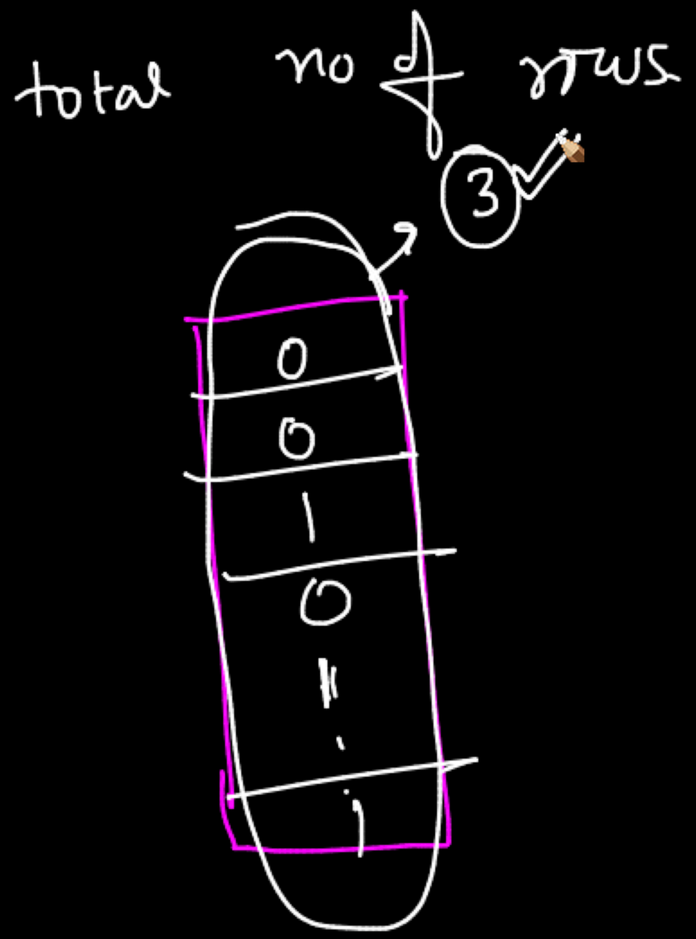
✓ None: int, float

✓ NaN: object

missing values in a column with non-number entries

Columns with number entries





(1) remove those rows / columns

(2) just fill missing values.

NaN

any()

all()

[F, F, T, F, F, F]