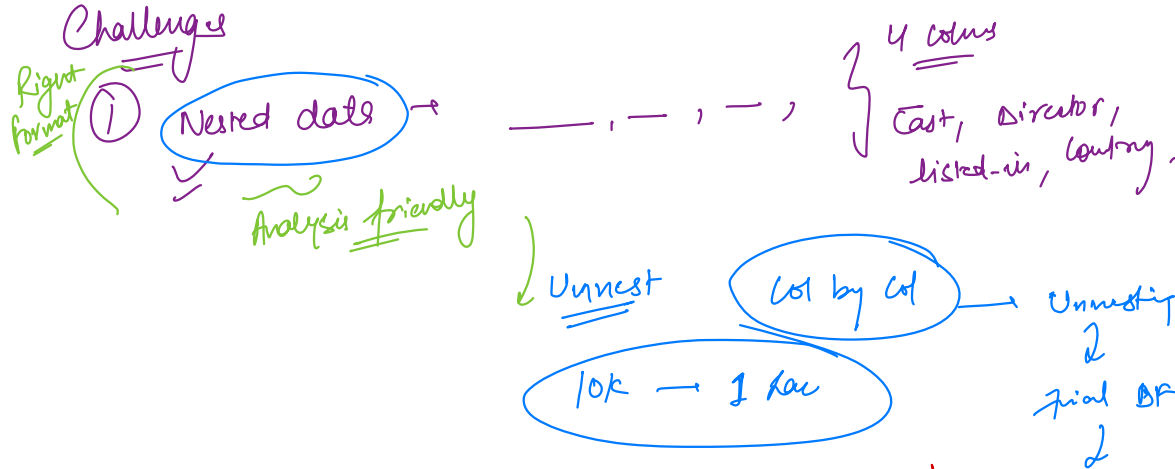


Quantitative → Sales
Profits | views
likes }

Qualitative measure



2) Missing values → replaced with mean, median, mode

Targeted imputing
using mode

df

Movie	Cast	Director	Mode
A	C ₁	D ₁	D ₁
B	C ₁	D ₂	D ₁
C	C ₁	D ₁	D ₁
D	C ₁	D ₁	D ₁
E	C ₂	D ₁	D ₁
F	C ₂	D ₂	D ₂
G	C ₂	D ₂	D ₂
H	C ₂	D ₂	D ₂
I	C ₂	D ₁	D ₂
J	C ₂	D ₂	D ₂
K	C ₂	D ₂	D ₂

Step 1

df.groupby('Cast')['Director'].mode

Cast	Mode
C ₁	D ₁
C ₂	D ₂
i	

ef

Step 2

join 'ef' & 'df' together on the basis of Cast.

3

Duration

→ 90 min → 2 seasons ✓

4

Date added

datetime

Step 2

Time	Cast
3 Idiots	C ₁ , C ₂

3 Idiots	C ₁
3 Idiots	C ₂

Split

Cast-0 Cast-1

3 Idiots	C ₁	C ₂

Stack

Type	Title	Duration	Duration
Movie	A	90 min	90
MOV	B	30 min	30
TV	C	1 sec	1
TV	D	2 sec	2
MOV	E	70 min	70

df.groupby('Type')['Duration'].mean()

Q Most popular actor-director pair.

final-df

final-df.groupby(['actor', 'director'])['Titles'].nunique().sort_values(ascending=False)

Title	Cast	Director	Country	genre
3 Idiots	C ₁ , C ₂ , C ₃	D ₁ , D ₂ , D ₃	Country 1, Country 2	G ₁ , G ₂ , G ₃ , G ₄
	3	3	2	4

90

90 rows ~~duplicate~~ at row level X
Unique

There was a question in the pdf given in hints -
Find After how many days the movie will be added to Netflix after the
release of
the movie (you can consider the recent past data)

