"Hello Everyone!"

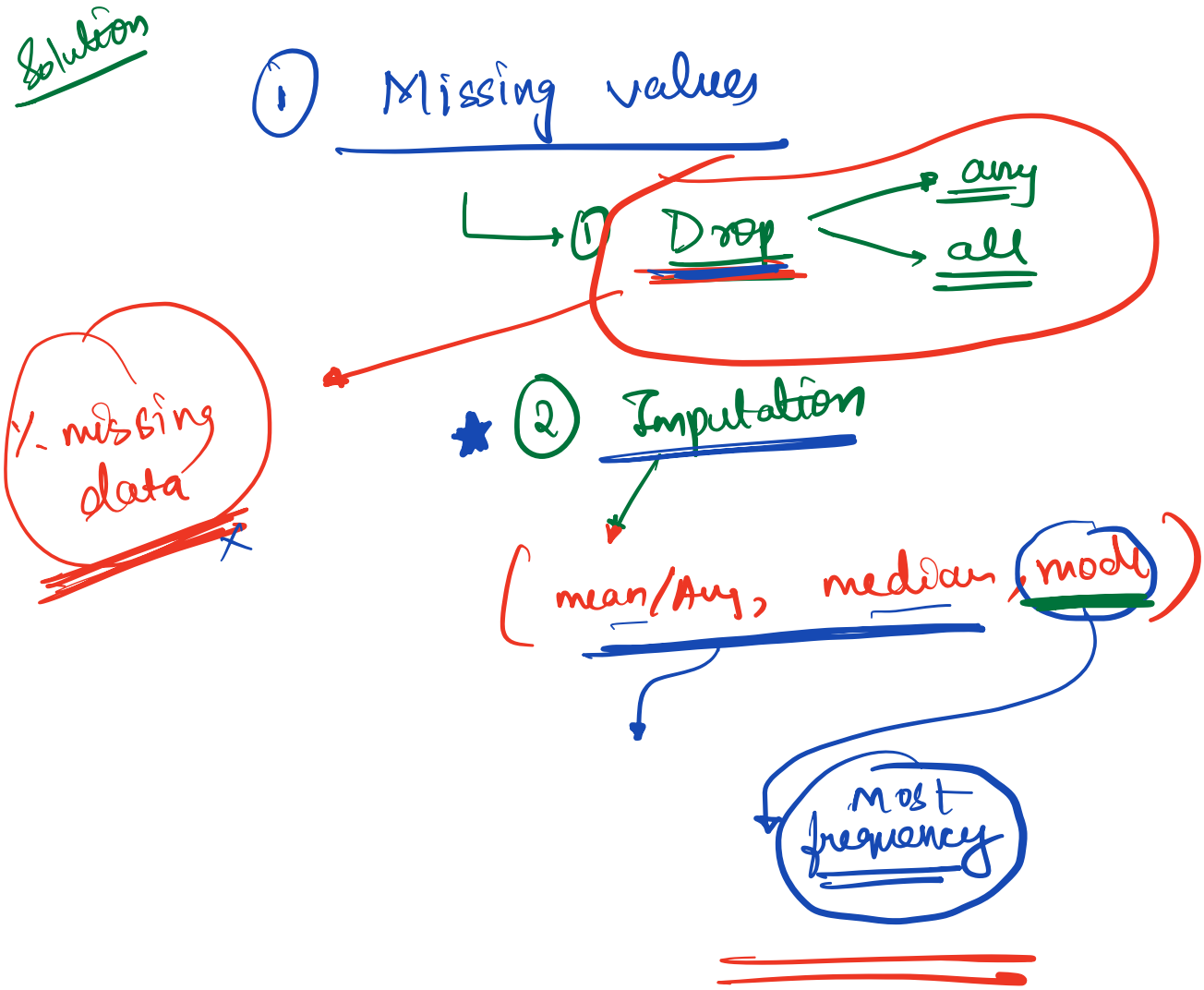
Netflix Review

+

Aerofit Intro

~ Aditya Jain
(AJ)

* Challenges
  1. Missing values
  2. Nested Problems
     - Cast, Director, Country, Listed-in.
  3. Type Conversion


Solution
1 Missing values
1 Drop
any
all
% missing data
2 Imputation
mean/Avg, median, mode
most frequency

* Dealing with Nested data :-

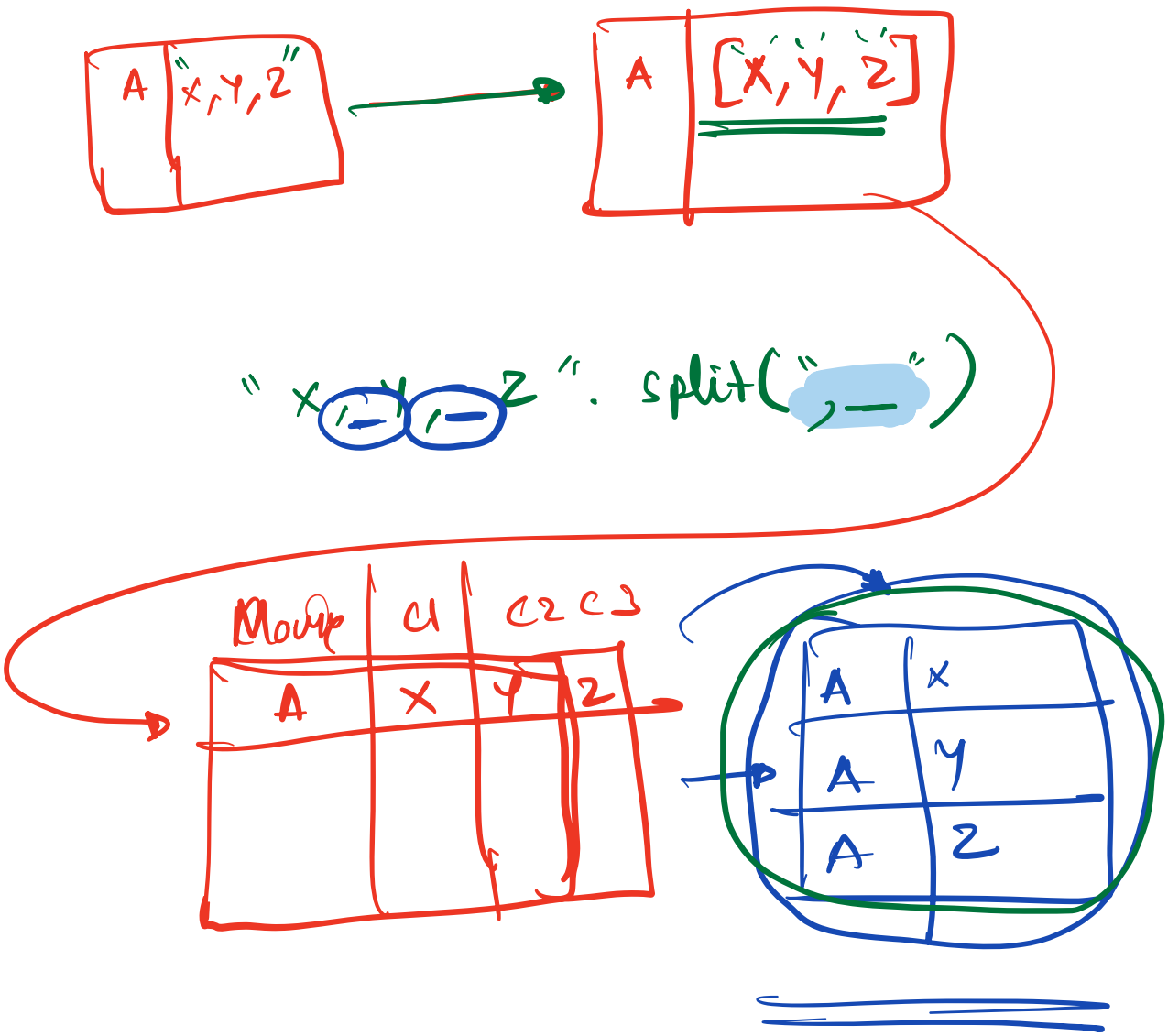
Given data :-

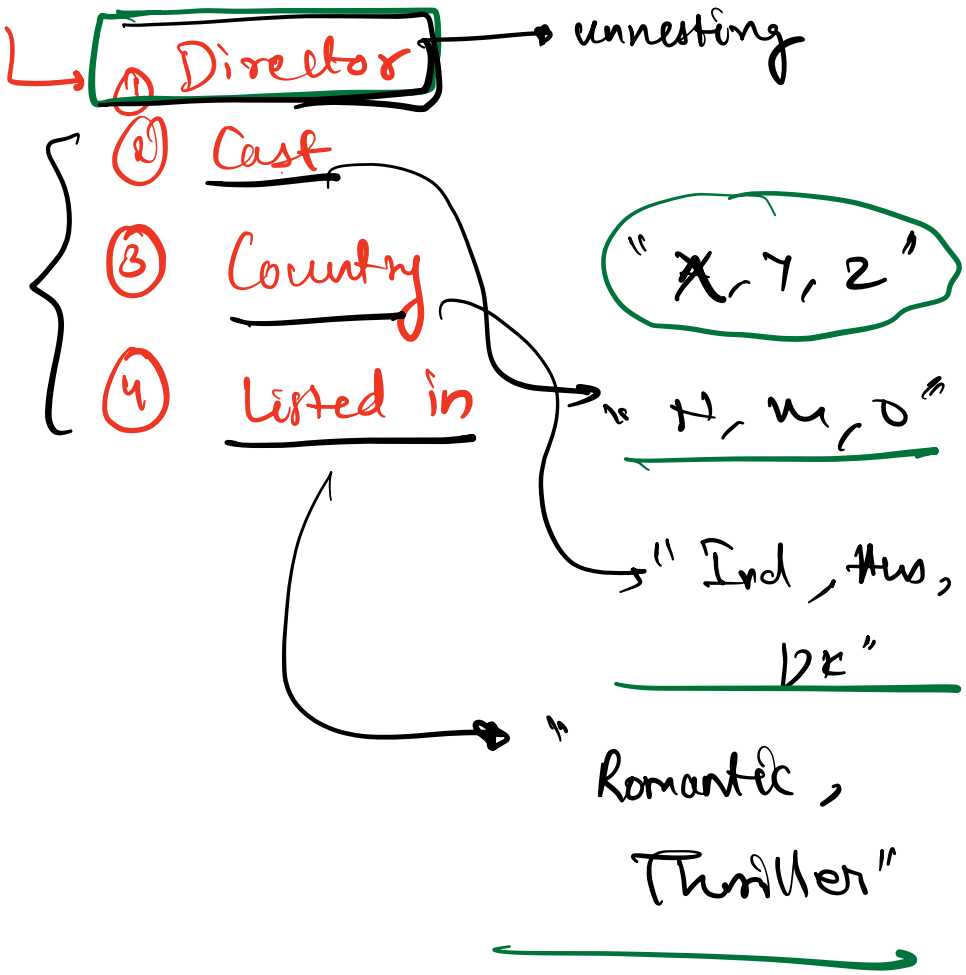
Movie	Cast
3 idiots	A, B, C

Nested ⟶ Unnested.

3i	A, B, C

Movie	Cast
3i	A
3i	B
3i	C


A
"x,y,z"
A
[x, y, z]
"x-y-z".split("-")
Молур
C1
C2 C3
A
x
y
z
A
x
A
y
A
z


1 Director
connecting
2 Cast
3 Country
4 Listed in
"X, Y, Z"
"M, N, O"
"Ind, Aus, UK"
"Romantic, Thriller"

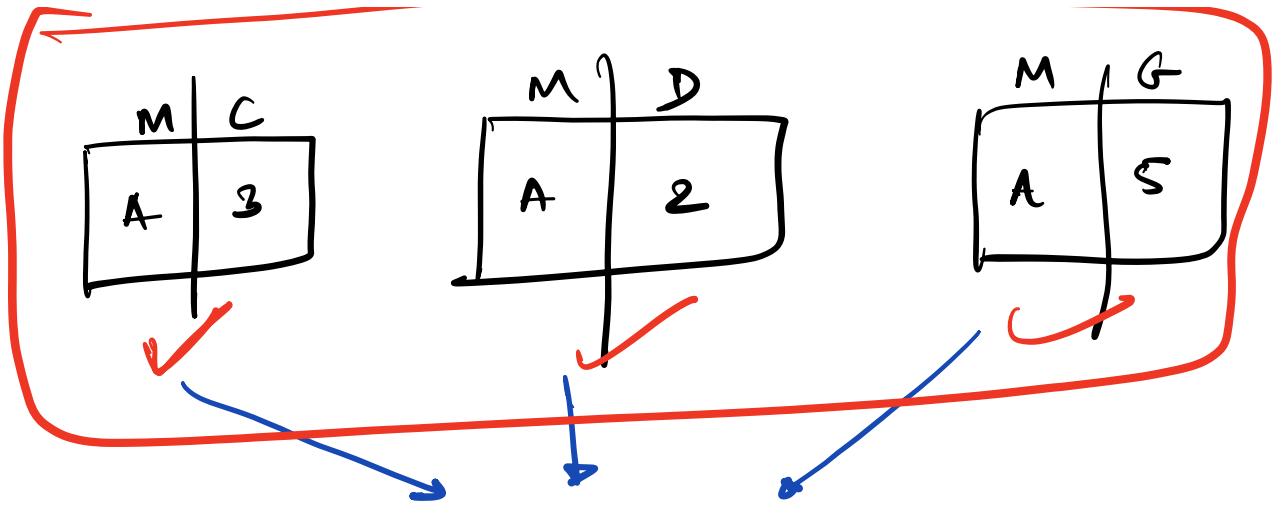
Title	Cast
A	3

*r 4r 2

Title	Director
A	2

Title	genre
A	5

Join

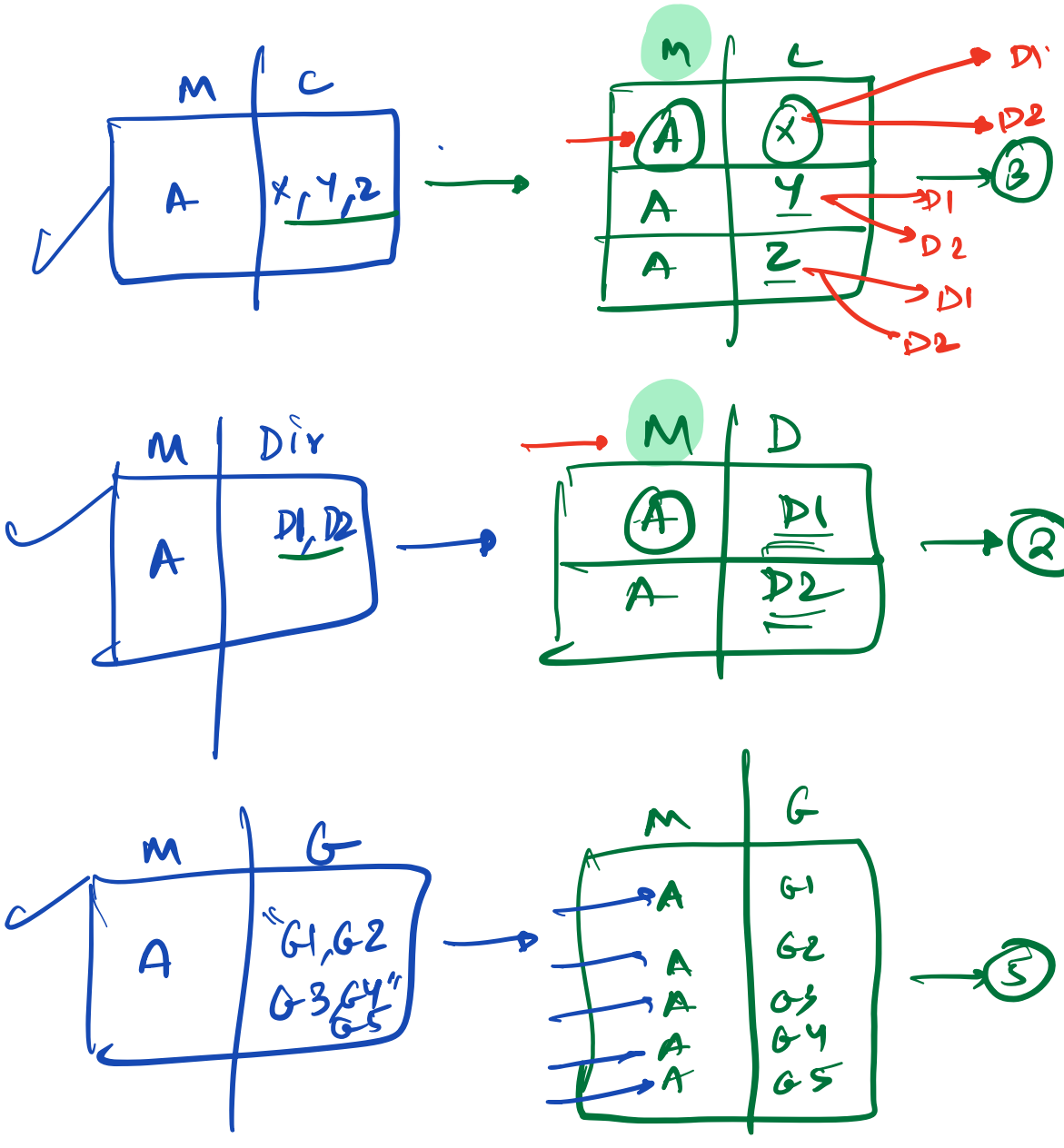


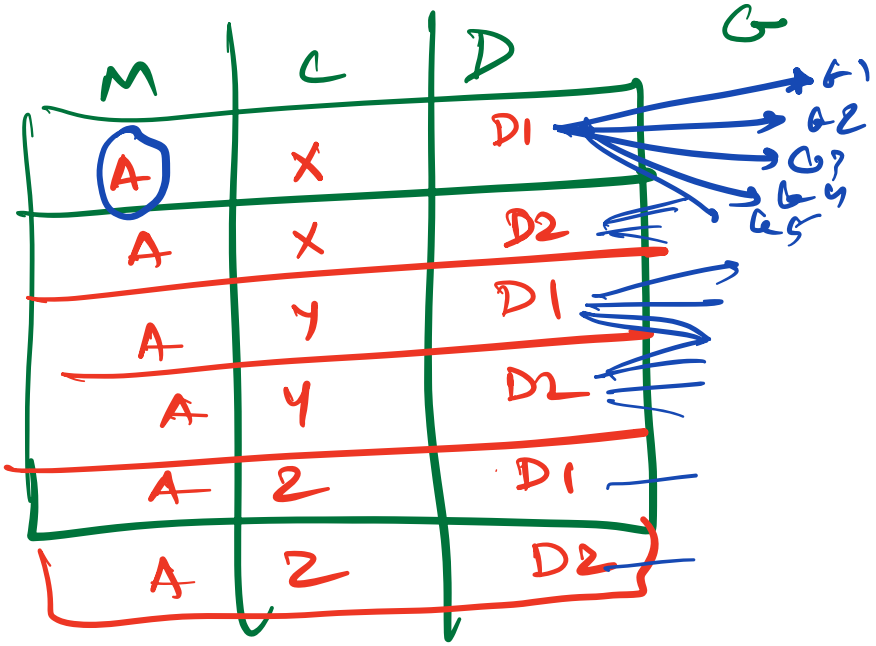
M	C	D	G

merge

$3 \times 2 \times 5 = 30$

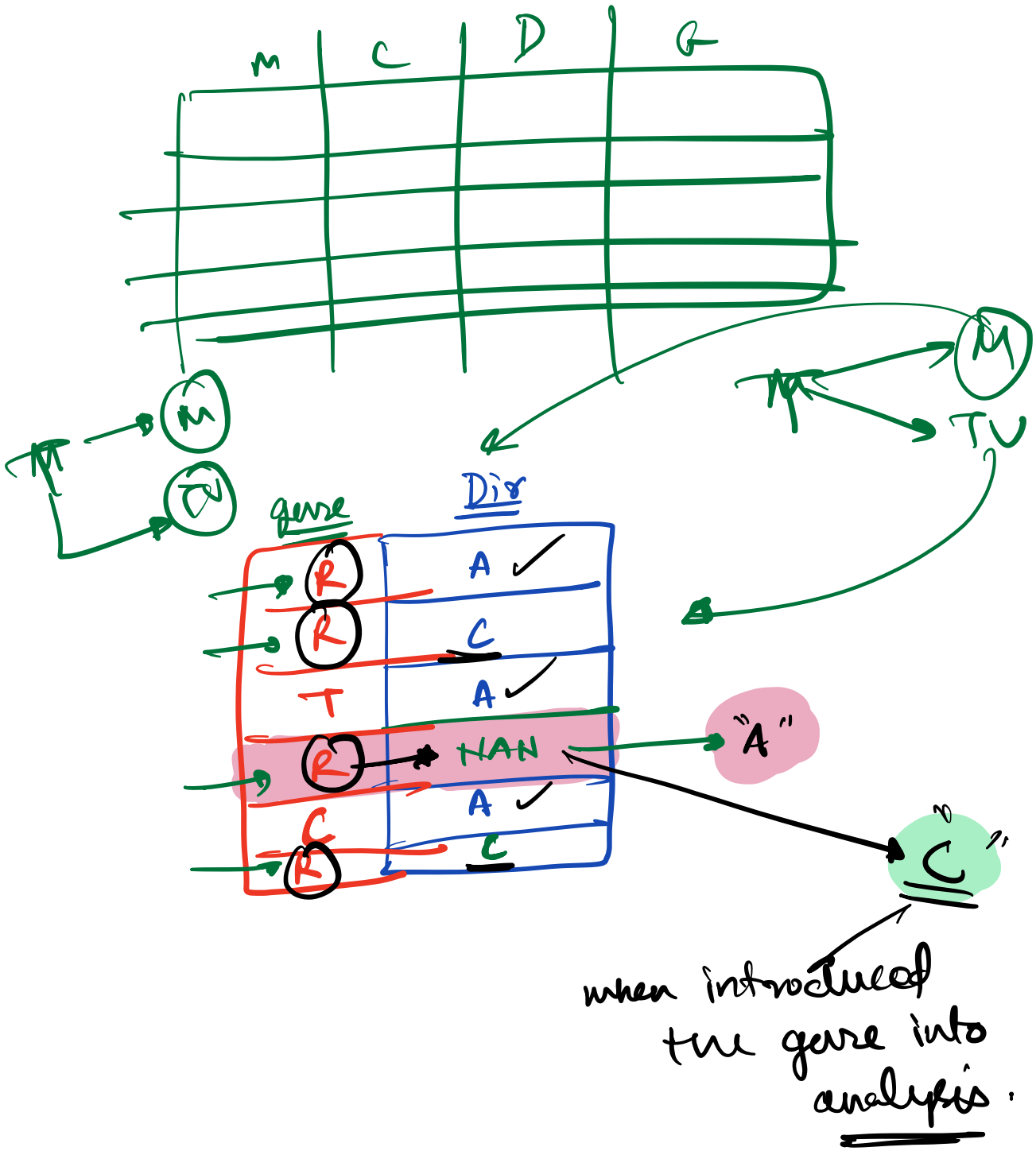
rows


M
C
A
X, Y, Z
M
L
A
X
A
Y
A
Z
D1
D2
3
D1
D2
D1
D2
M
Dir
A
D1, D2
M
D
A
D1
A
D2
2
M
G
A
"G1, G2
G3 G4"
G5
M
G
A
G1
A
G2
A
G3
A
G4
A
G5
5

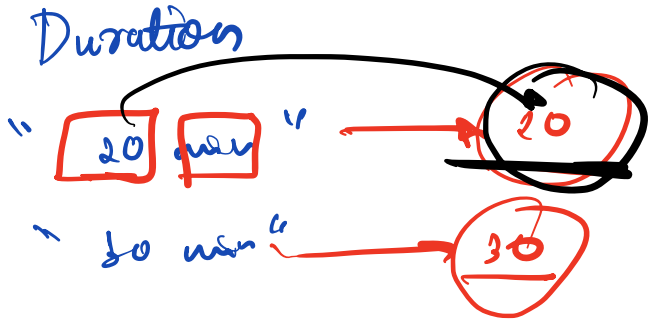


C, D, G

$3 \times 2 \times 5 = 30$ rows

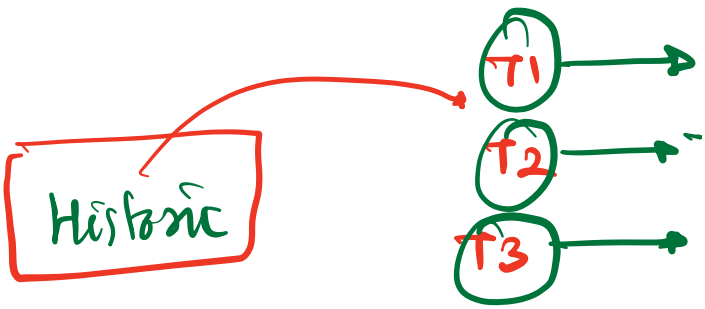

M
C
D
G
M
TV
M
TV
genre
Dir
R
A
R
C
T
A
R
NAN
"A"
A
C
R
C
when introduced
true genre into
analysis.

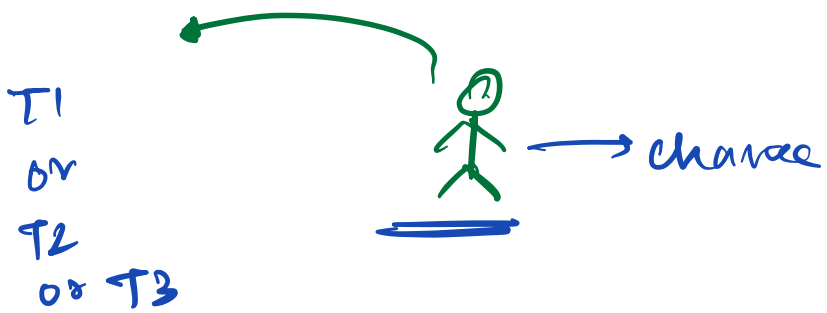
(8) what is the avg len per movie ?



For each type of product,
you have to come up with analysis
around which type of
customers are more likely
to purchase it.

* what kind of product is popular among what kind of Customers.



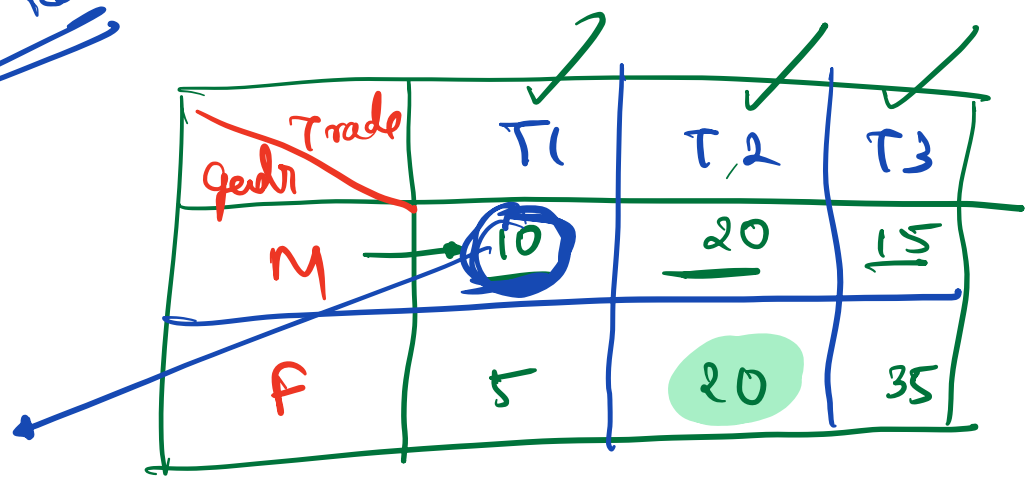


$1500 ← T1 → KP281 → Entry

$1750 ← T2 → KP481 → mid

$2500 ← T3 → KP781 → Advaced

Cross tab



M → 10 T1 → in part

20 T2 →

15 T3 →