


→ Problem Solving % → PSP

80 BS } sessions

→ 25 → Take home

↳ created by
Suraaj (POC)

70%

(30) → 70%
⇒ 21 BGC

First Point of Contact → TA → Suraaj

✓ Imp point → Scores do not matter

↳ Deadline (3.5 marks penalty
70% ↓ every day)

Basic → Netflix (7-8 deadline)

→ Only 1 submission

Netflix

Intro Call / Challenge ←

2 } work
Submit

Aerofit { Released }

Netflix Solⁿ, Aerofit intro | 1.5 hrs
1 hr 1/2 hour

Submit BC

- ① Peer Evaluation 9/10
 - ② TA ✓ 8/10
 - ③ BC session Business Case
-

Netflix Business Case

II Cost/Risk

- We won't have a Director
- NO Release / lock business
- They are overloading their system.

III benefit

- More profit
- More Bandwidth
- good content
- give more money to good producers

Problem St. I

get insights from data after processing

IV

Recommendations

Insights → Recommend

Insights

Recommend^u

① Karan Johar has the most movies in the Romantic Genre

① Netflix should create Collaborate with KT for Romantic Movies

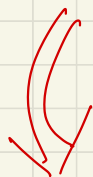
Challenges

① There is no Continuous Variable
(Views, Production expense, TRP)

(2.75, 300, 400,)

Limited Analysis

② Nested Data Cast (A, B, C, D)



Unnested

Cast
✓ SRK →
SRK, X, Y, Z

① Split → A, B, C, D

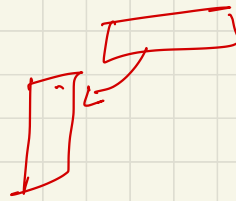
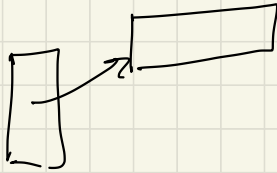
② Stack → Transpose

0	1	2	3
A	B	C	D

A
B
C
D

Stack \rightarrow Transposes

Unstack \rightarrow Pivot



Unnest

① split

② stack

(A, B, C, D) \rightarrow columns

cols \rightarrow rows

Title	✓	✓	✓	✓				
	A, B, C, D	X, Y	A, B, C, D	X, Y				

\rightarrow d.f. Cast. value - Counts

Split & Stack \rightarrow index \neq title

Country (2)

listed-in (2)

Title	Cast
DDLJ	A, B, C, D ✓

Split & Stack

Title	Cast
DDLJ	A
DDLJ	B
DDLJ	C
DDLJ	D

4 values

Title	Director
DDLJ	X, Y ✓

Split & Stack

Title	Director
DDLJ	X
DDLJ	Y

2 values

T	Cast
DDD	4

T	Director
	2

T	Country
	2

T	genre
	2

merge
 $8 \{4 \times 2\}$
 inner

C-D
 inner
 $8 \times 2 = 16$

inner
 16×2
 C-D-C-G

DDLJ	(A)
B	X
C	Y
D	

A	X	Ind	C
A	Y	Pak	R
B	X	Ind	C
B	X	Pak	R
C	Y		
C	Y		
D	X		
D	Y		

DDLJ

32 Rows

Title	Cast	Direct	Country	genre	Rating	Released	Date added

merge on Title

Time	Cast	Director	Duration	
	SR		90 Min	→ split Movie TV Shows
			90 Min	
			90 Min	

Assumptⁿ → whatever given is working for Netflix

Just Cast Insights

- Most popular actors
- Multi Star Count for every movie

df.groupby('Cast')[title].nunique()

SRX → title

→ cast

- Best cast in every genre
- famous actors of every country

→ Average sum time for every movie

original data from

df.groupby('title')[duration].mean()

original df → genre

↓

actor director

→ Average RT for every actor

" " genre →

→ for which genre we see max sum time

[describe()]

mean,
median,
count,
sum,

90, 90, 90, 90

90