→ NETFLIX INTRO

Streaming Platform

DataScientst

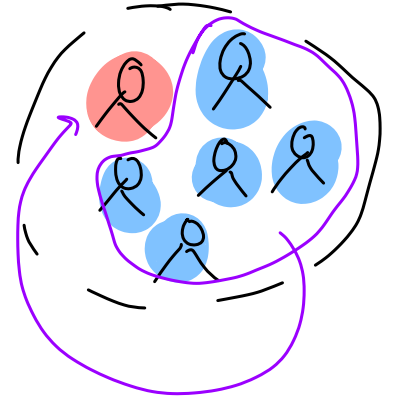
Recommend → persons
title
Genre

Recommendation Engines

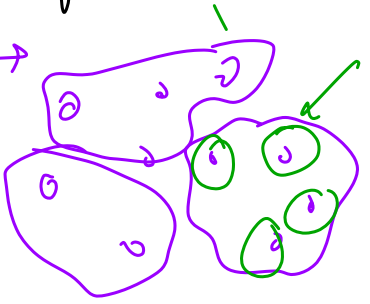
↳ Recommend

NETFLIX → Goal → {Increase Customer Subscribers.}

{* Collaborative filtering
* Content filtering} →

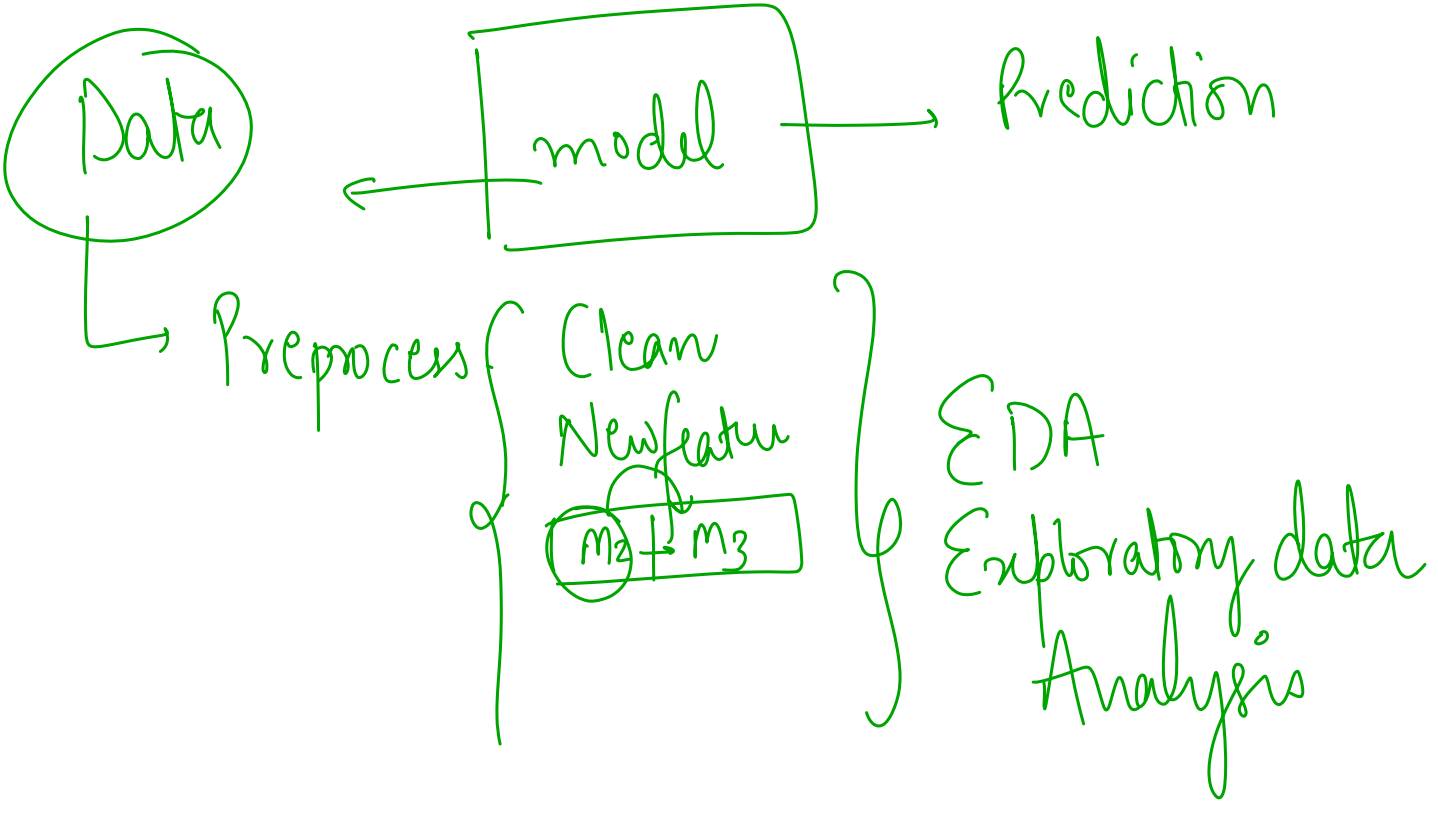


→ M1 →



Batman 1 → Batman 2

Dataset → Preprocess (Clean/make it usable)



Missing Values

1. Deletion / Remove — 5% 10%
2. Imputation ←

Nan → Not a Number → 5% ←

financial Trans

5% ← Imputation

Imputation → Numerical → mean/median

Categorical → Mode ←

SQL, Python, Numpy, Pandas, Matplotlib
Seaborn

Source of Truth: DOCS

datascientist

→ Preprocessing

Model

Deployment

Monitoring

Unnesting

title	Director	Cast	Genre
A	X, Y	SK, AK, DP	Drama, Movies
B	Z, E	Any VK, T	Document, TVshows
C	G, H	P Q, AK	Romantic

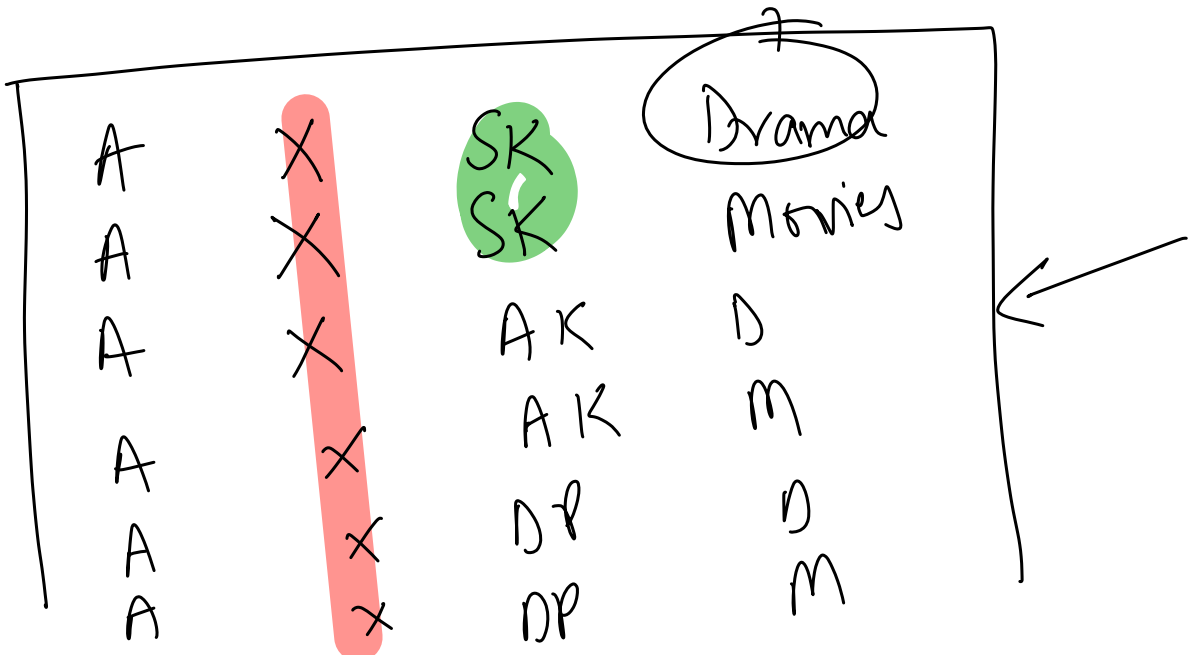
(1) Who is the most popular actor?:

["SK", "AK", DP"]

Unnesting

(2) Melting pivoting, Stack, unstack; explode

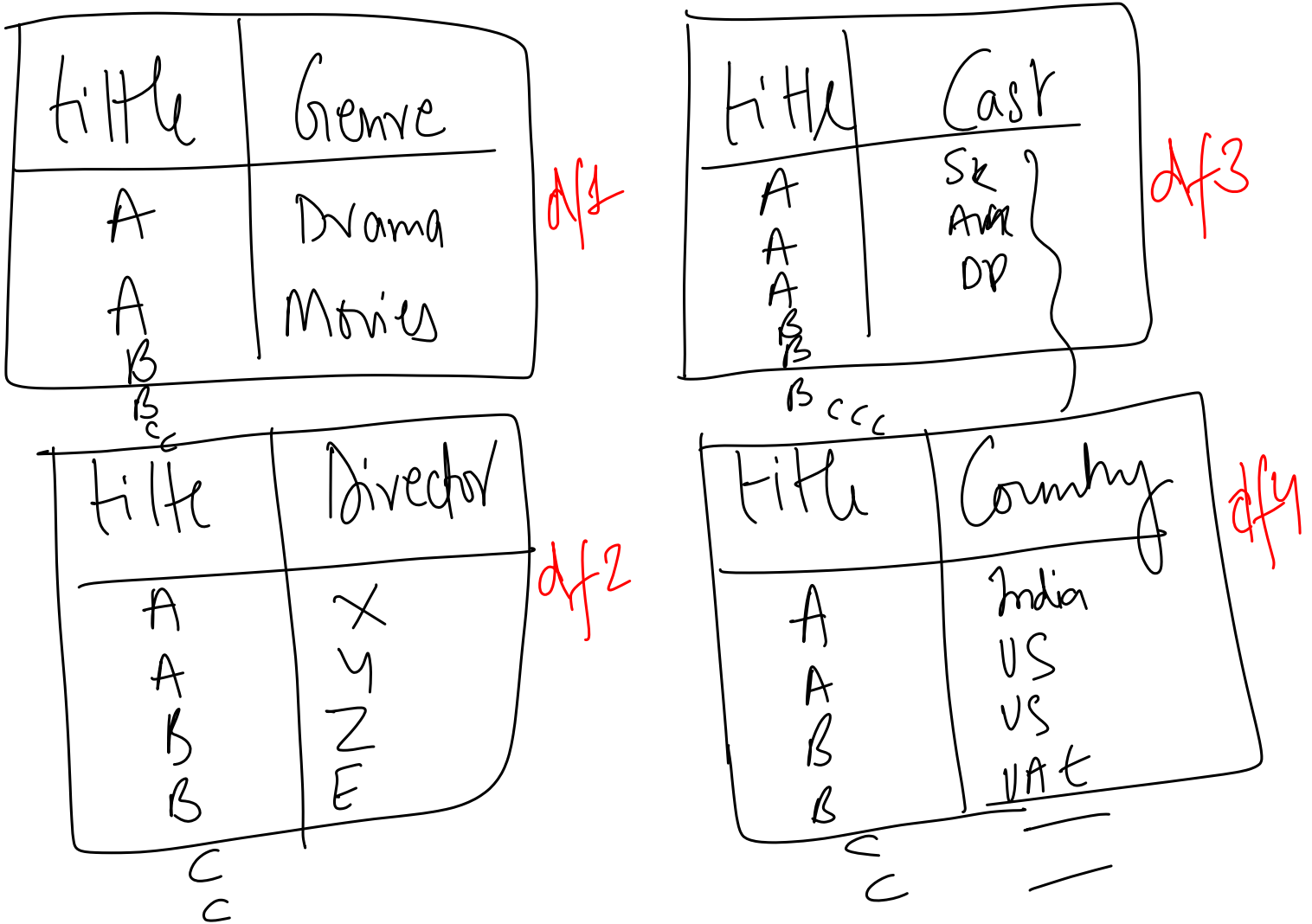
title	Director	Cast	Genre
A	X, Y	SK, AK, DP	Drama, Movies
B	Z, E	Any VK, T	Document, TV shows
C	G, H	P Q, AK	Romantic



A	Y	SK	Drama
A	Y	SK	Movies
A	Y	AK	D
A	Y	AK	M
A	Y	DP	D
A	Y	DP	M

"Actor Director" Combination

groupby (D, A) ~~.count()~~ .nunique()

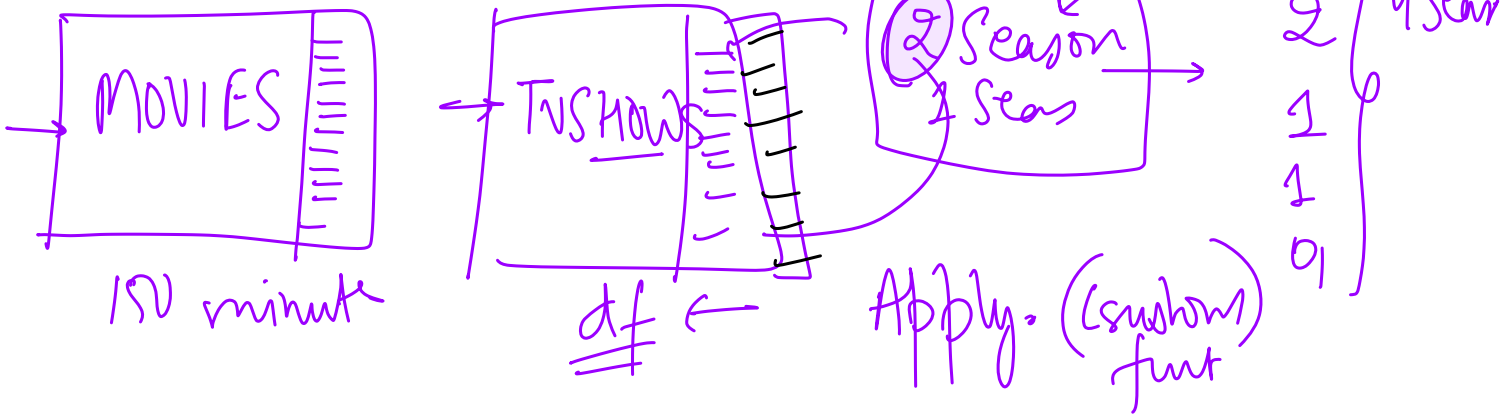

title
Genre
A
Drama
A
Movies
B
B
C
C
df1
title
Director
A
X
A
Y
B
Z
B
E
C
C
df2
title
Cast
A
SK
A
ANR
A
DP
B
B
B
C
C
C
df3
Title
Country
A
India
A
US
B
US
B
UAE
C
C
df4

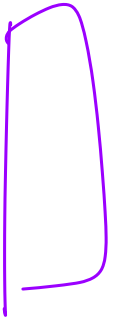
pd.merge(LD , RD , on → title

show

What is the avg Runtime of Show/Movies?

*





splitting " - "

2 season ✓

↳ ["2", 'season']

(1) Charts (Visualise)

(2) Be Creative in asking Question

(3) Is imputation making ~~no~~ sense here!
graphy

(4) Unnesting

1. Google / Docs
2. TA Help Req.
3. Reach out to me "nikhil.songhi@scaler.com".

→ Github

10 Mins

→ daily bash