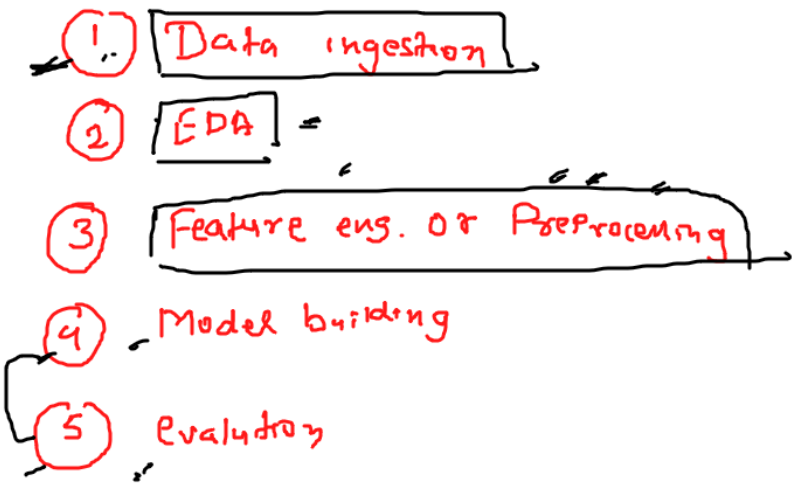
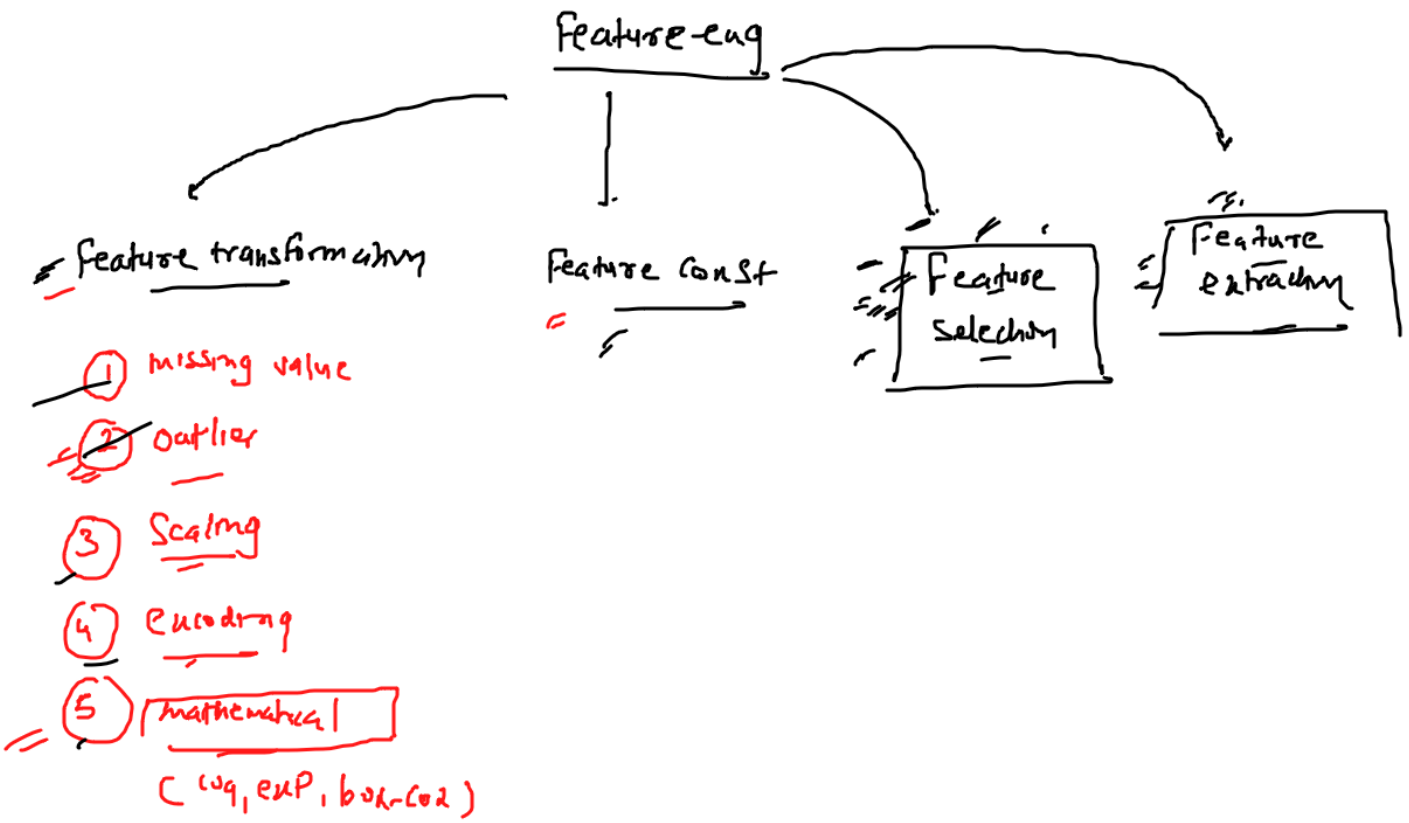
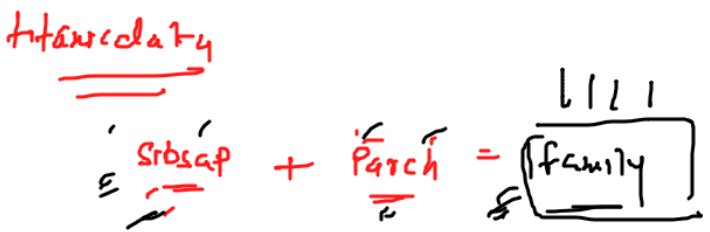
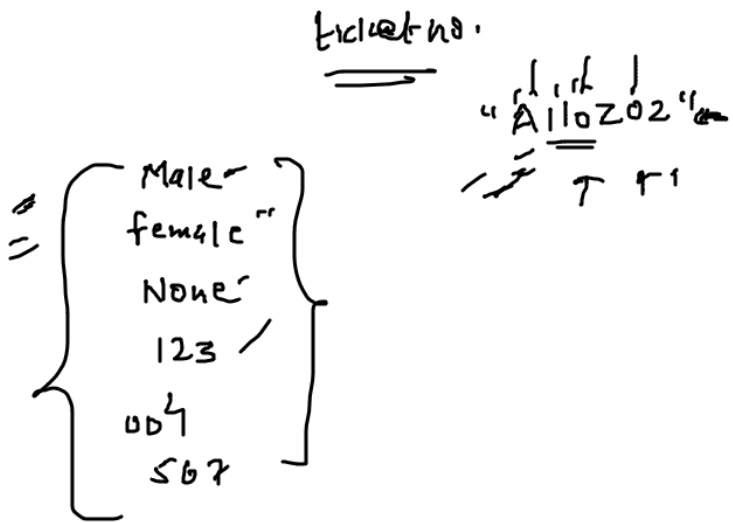
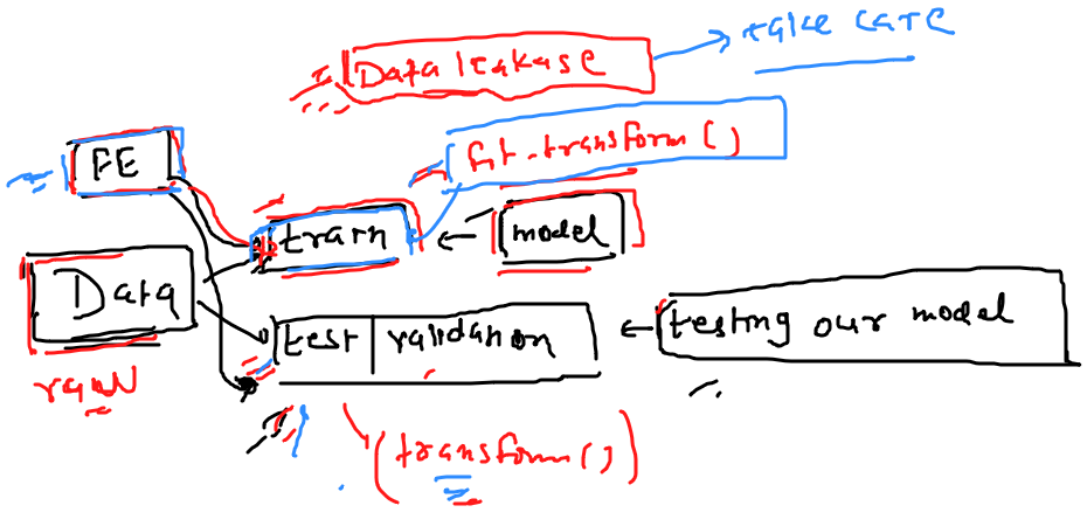
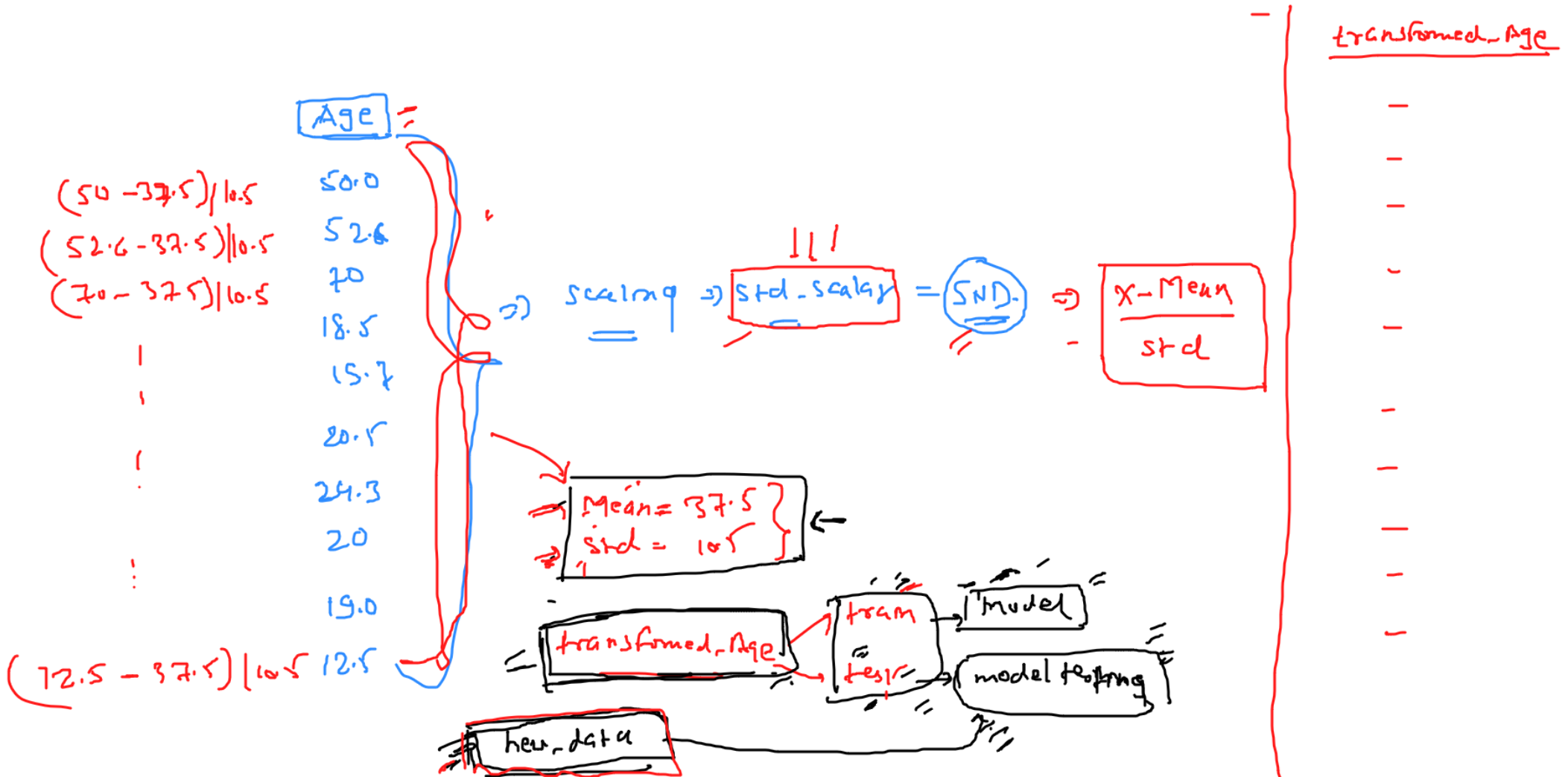

1
Data ingestion
2
EDA
3
Feature eng. or Preprocessing
4
Model building
5
Evaluation

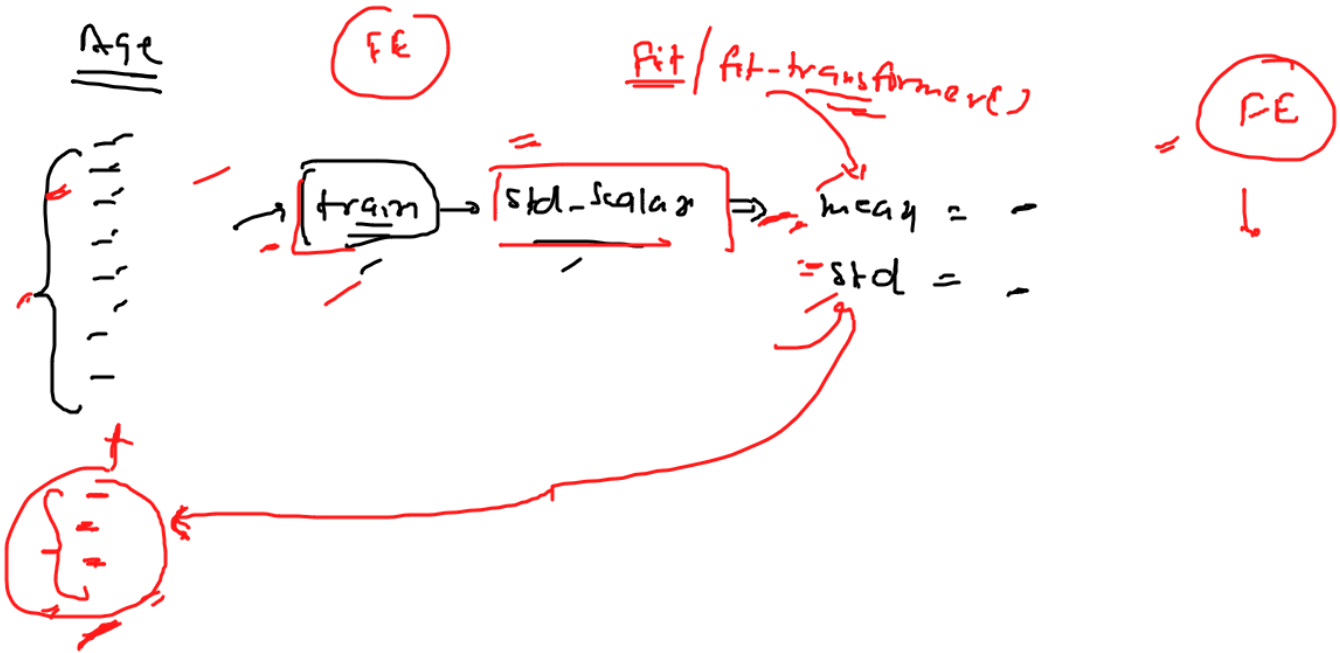

Feature-eng
Feature transformation
1 missing value
2 outlier
3 Scaling
4 encoding
5 mathematical
(log, exp, box-cox)
Feature const
Feature Selection
Feature extraction

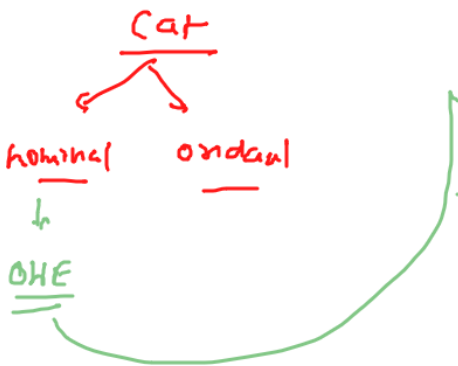

Srbsap
+
Parch
=
family


Male
female
None
123
567


Data leakage
FE
fit_transform()
train
model
Data
test
validation
testing our model
raw
transform()


Age
50.0
52.6
70
18.5
15.7
20.5
24.3
20
19.0
12.5
(50 - 37.5)/10.5
(52.6 - 37.5)/10.5
(70 - 37.5)/10.5
(72.5 - 37.5)/10.5
scaling
std_scalar
SND.
X - Mean / std
Mean = 37.5
std = 10.5
transformed_Age
train
test
model
model testing
new_data
transformed_Age


Age
FE
fit / fit-transformer()
FE
train
std_scalar
mean = -
std = -



col	Male	female	None
Male	1	0	0
female	0	1	0
None	0	0	1
Male →	1	0	0
None →	0	0	1
female →	0	1	0