

Kaggie.com
UCI machine learning

Data Analysis Process

→ ASKING Question

Experience

Subject matter expertise

→ Data Wrangling / Munging

Available

① Gathering Data

Others - Web scraping

② Assessing Data

NOT available - Networking

③ Cleaning Data

CSV, excel, API - Rapid API,

*** Feature engineering

Web scraping - Hugging face

→ Exploratory Data Analysis

Semi-structured

Explore

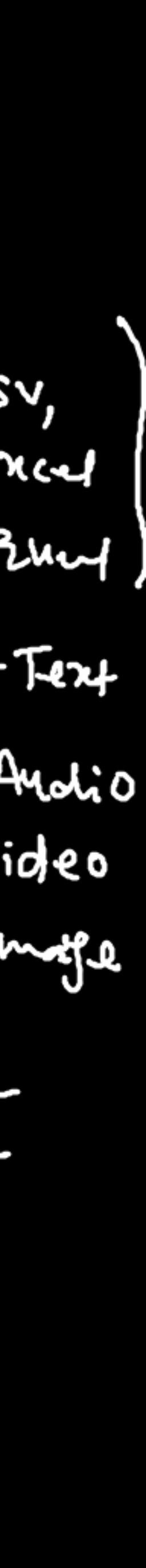
Augment

→ univar, BS-, m. on 11 -

Date		Year
Day	Month	Year
11	Dec	2023
8.45.23	11	12
	H	M
	M	S

MLops

CNN



→ Drawing conclusions

→ Descriptive Statistics

→ Inferential Statistics - Random

→ Machine Learning

↔ Communicating Results / Data Storytelling.

Date → Augmentation

11 Dec 2023

The diagram shows a horizontal line representing the date '11 Dec 2023'. Four arrows point downwards from this line to the words 'Day', 'Month', 'Year', and 'day of the week'. A fifth arrow points downwards from the end of the date line to the word 'Quarters'.

Day Month Year day of the week Quarters.

You are screen sharing

EN

Stop Share

milind pawar +2 others raised hands View X

Data Preprocessing Method - everywhere

→ ① Handling Missing value

Rule 1 :- Check threshold value (25%) missing data or more - Remove that

{ Please note, threshold may change participant variable basis process to process or basis your stakeholder requirement

You are screen sharing

Stop Share

Pritam raised hand View X

Rule 2 :- if data is missing - less than 25%. (Impute Method)

Approach :- Check variable/feature/column

uni-variate approach

Character

Number

Check - outlier

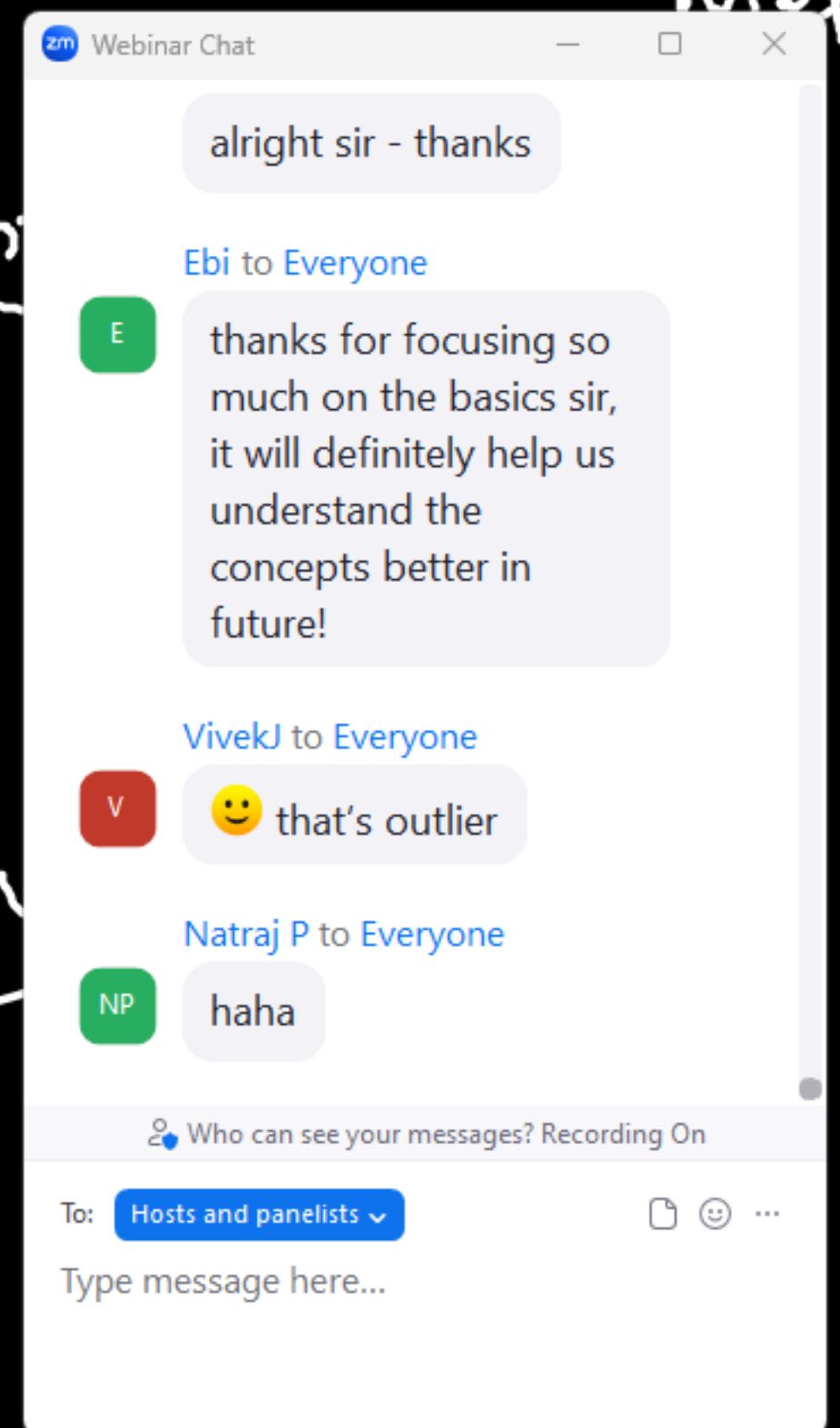
Yes

'Median'

KNN Imputation
MICE
iterative imputation

Impute

"Mode"



You are screen sharing

EN

Stop Share

	A	B	C	D
1	a	<u>Nan</u>	100	
2	b	<u>Nan</u>	200	
3	<u>Nan</u>	<u>Nan</u>	<u>Nan</u>	300
4				
5		<u>Nan</u>	10	<u>Nan</u>
6	c		20	
7	d		<u>Nan</u>	500
8	e		30	600
9	f		40	700
10	g		50	800
			<u>Nan</u>	900

X

$$\frac{2}{10}$$

How many missing

0%

A

20%

B

50%

C

10%

D

ASE

10

20

30

25

1000

Nan + 265

$$\frac{10+20+30+1000}{4}$$

$$\text{mean} = \frac{1060}{4} = \underline{\underline{265}}$$