SNS COLLEGE OF TECHNOLOGY

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade

Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

Department of Computer Science and Engineering

Course Code & Title : 23ADO201 - Data Science Fundamentals

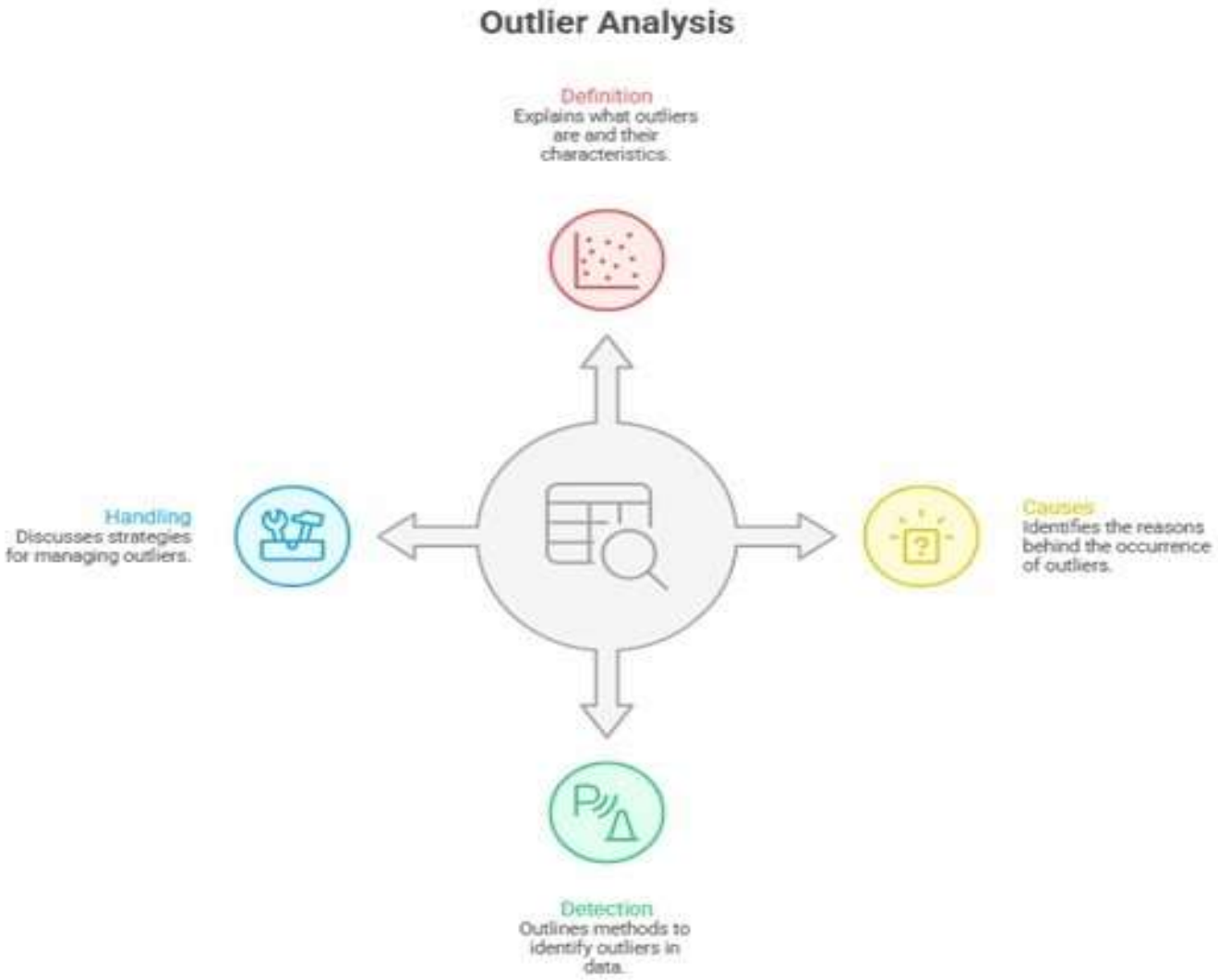
III YEAR / VI SEMESTER - EEE

Unit 2 - DESCRIPTIVE ANALYTICS

Topic : OUTLIERS

Mr.K.Karthikeyan AP /CSE

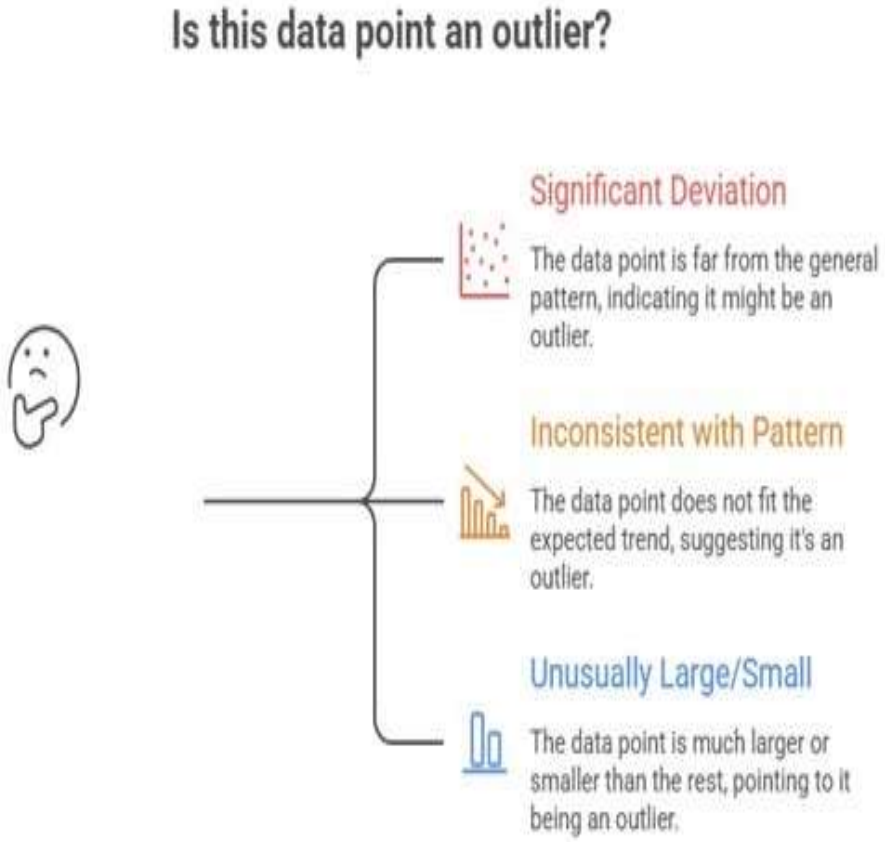
A Comprehensive Overview Outliers



www.snsgroups.com

What are Outliers?

Outliers are data points that significantly deviate from the other values in a dataset. They are observations that appear to be inconsistent with the general pattern of the data. Outliers can be unusually large or small compared to the rest of the dataset.



What is the cause of the outlier?

Measurement Errors

Faulty equipment or incorrect data entry can lead to inaccurate measurements.

Data Entry Errors

Human errors during data collection or input can introduce incorrect values.

Sampling Errors

If the sample is not representative, it may contain extreme values.

Natural Variation

Outliers may represent genuine extreme values within the population.

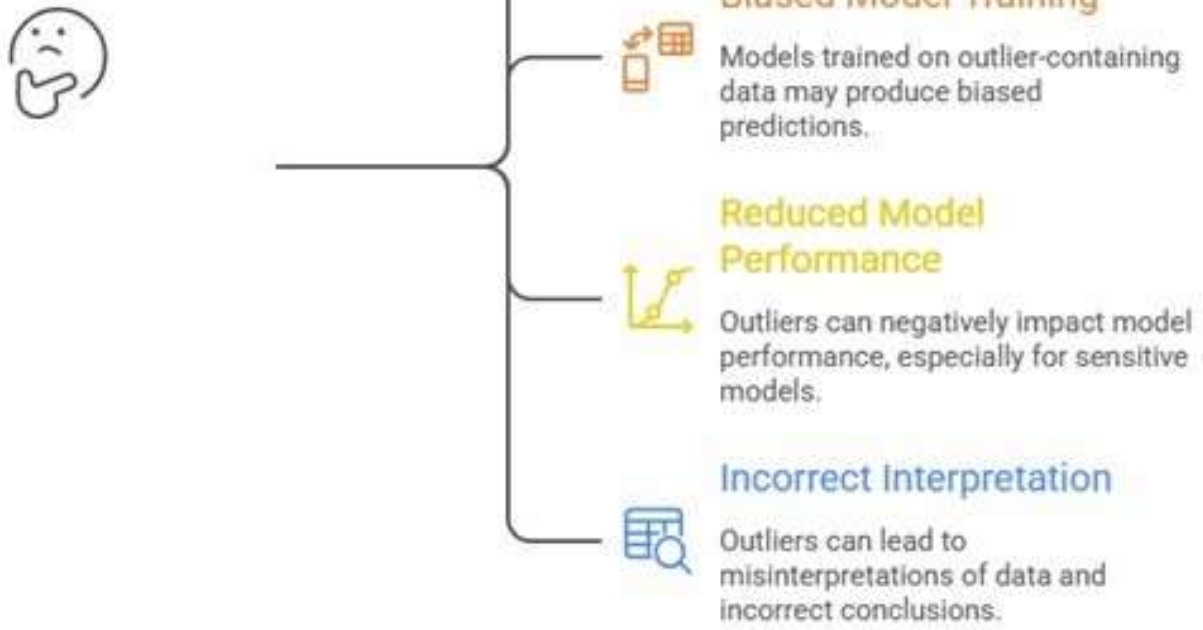
Novelty Events

Outliers can signal new or unusual events.

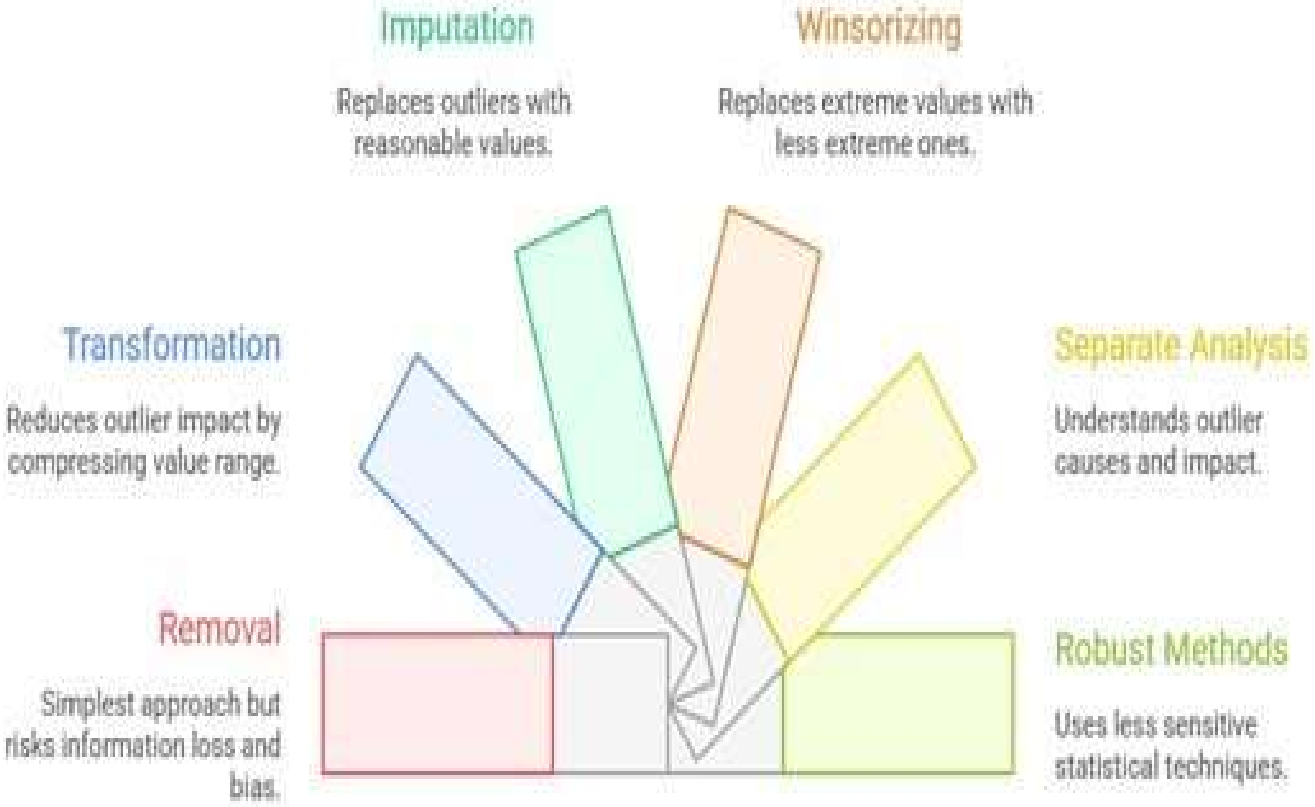
Fraudulent Outliers

Outliers can be intentionally created to manipulate data.

How do outliers impact data analysis and model performance?



How to handle outliers in a dataset?



www.snsgroups.com

🎯 1. “Spot the Odd One Out”

Objective: Identify outliers visually

Activity:

Write this dataset on the board:

10, 12, 11, 13, 12, 14, 95

Ask students:

Which value looks unusual?

Why is it different?

The Decision Matrix: Drop or Keep?

DROP IF:

- Impossible values (Age=200)
- Entry Errors.

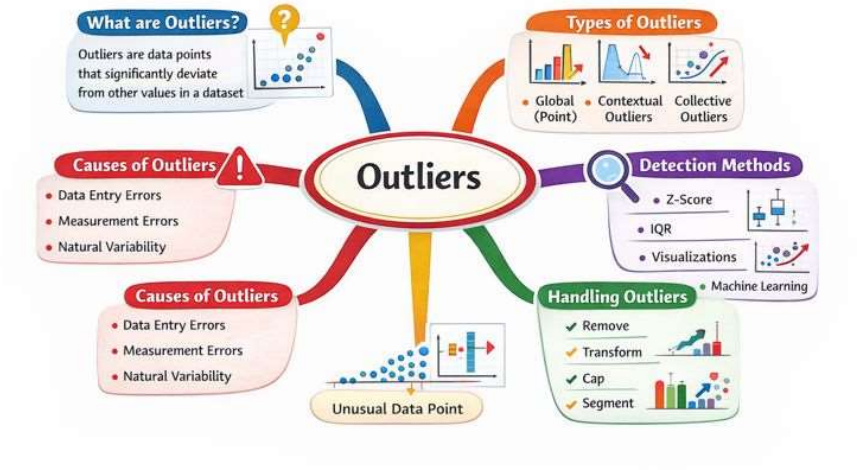
KEEP IF:

- Fraud Detection
- High-risk applications
- Large % of data.

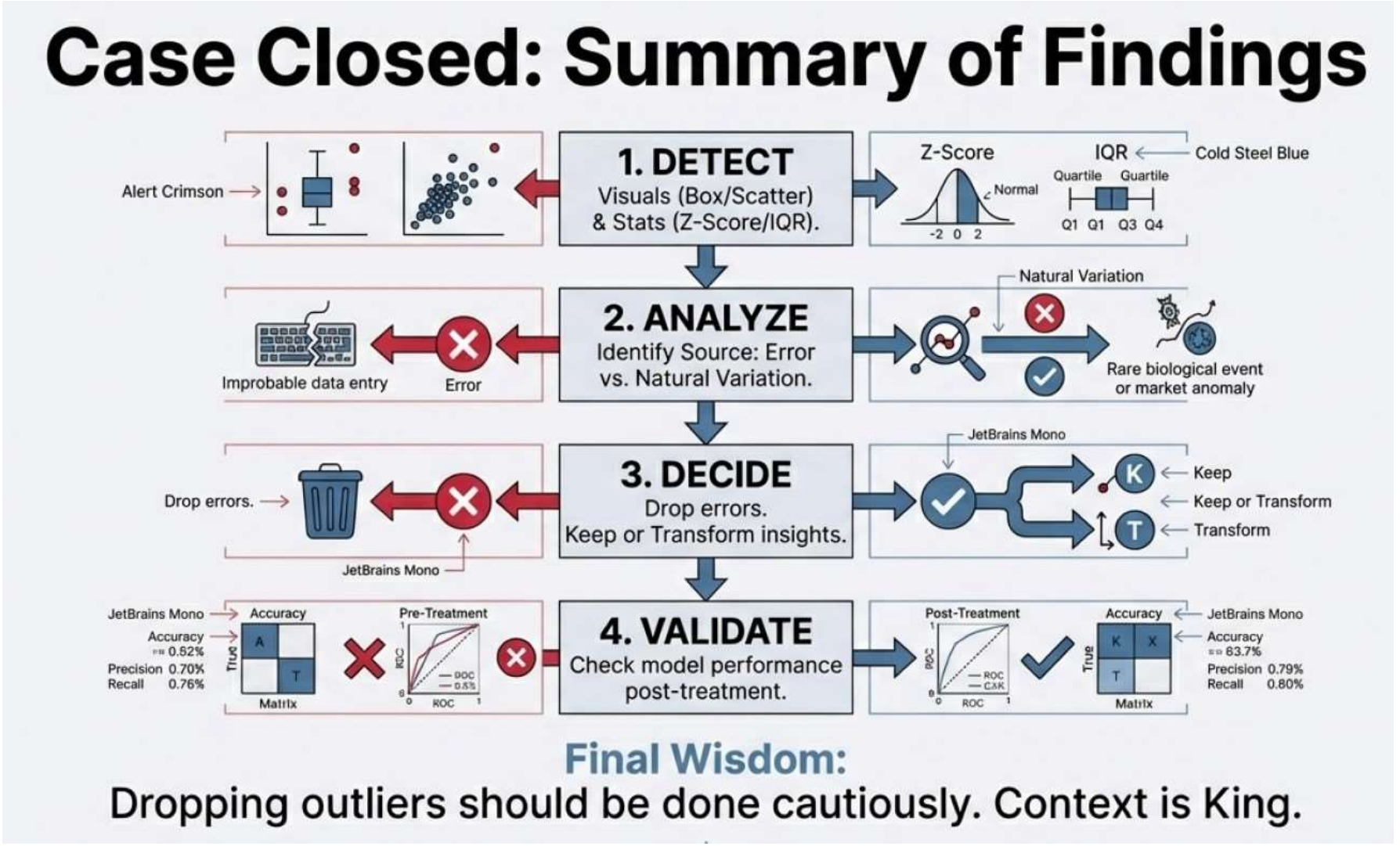
KEEP IF:

- Fraud Detection
- High-risk applications
- Large % of data.

MIND MAP –(Ex.Project Management with Outliers)



Case Closed: Summary of Findings



Final Wisdom:

Dropping outliers should be done cautiously. Context is King.

ASSESSMENT

www.snsgroups.com

MCQs: Outliers

1. An outlier is a data point that:

A. Occurs frequently in the dataset
B. Is equal to the mean
C. Lies far away from other observations
D. Always improves model accuracy

✅ **Answer:** C

2. Which of the following is a common reason for outliers?

A. Data entry error
B. Measurement error
C. Rare but valid event
D. All of the above

✅ **Answer:** D

REFERENCE BOOKS

1.Allen B. Downey, “Think Stats: Exploratory Data Analysis in Python”, Green Tea Press, 2014.
2.Sanjeev J. Wagh, Manisha S. Bhende, Anuradha D. Thakare, “Fundamentals of Data Science”, CRC Press, 2022.
3.Chirag Shah, “A Hands-On Introduction to Data Science”, Cambridge University Press, 2020.
4.Vineet Raina, Srinath Krishnamurthy, “Building an Effective Data Science Practice: A Framework to Bootstrap and Manage a Successful Data Science Practice”, A press, 2021.

www.snsgroups.com