

1. Data Quality Check (Most Common Use)

Question to ask

“Is this data point even valid?”

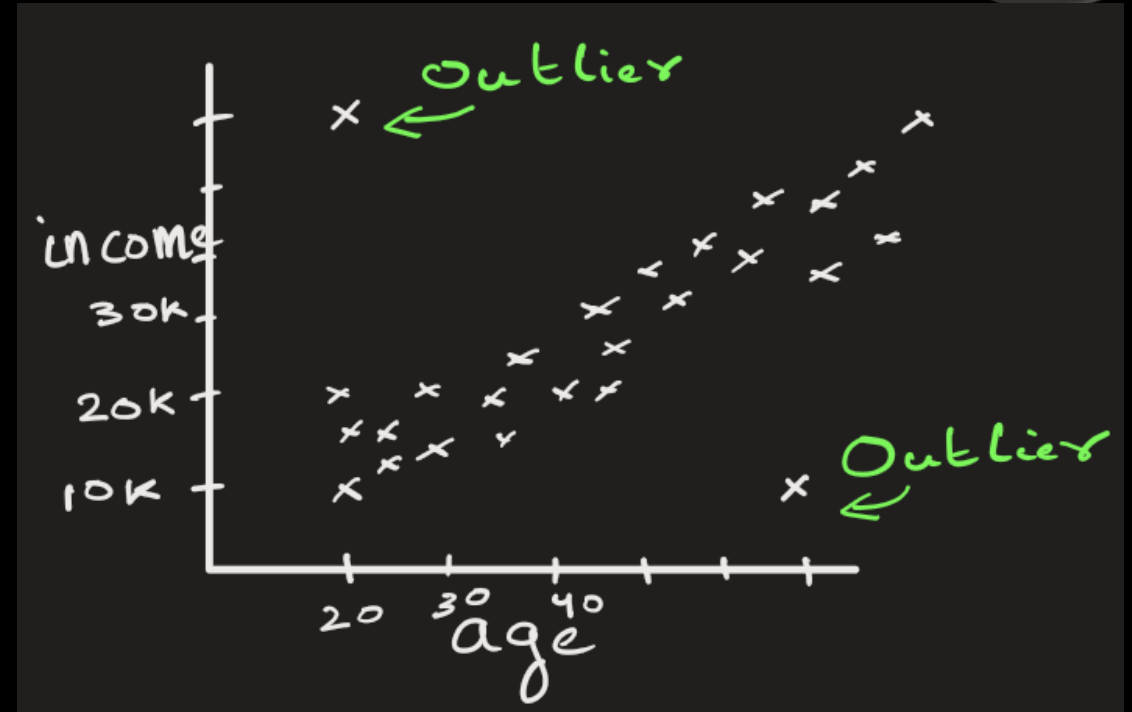
Actions

Check for:

- Measurement errors
- Sensor faults
- Data entry mistakes
- Unit mismatch

Example (Medical context):

- Tumor size = 500 mm (Not Possible)
- Missing decimals
- Wrong unit conversion

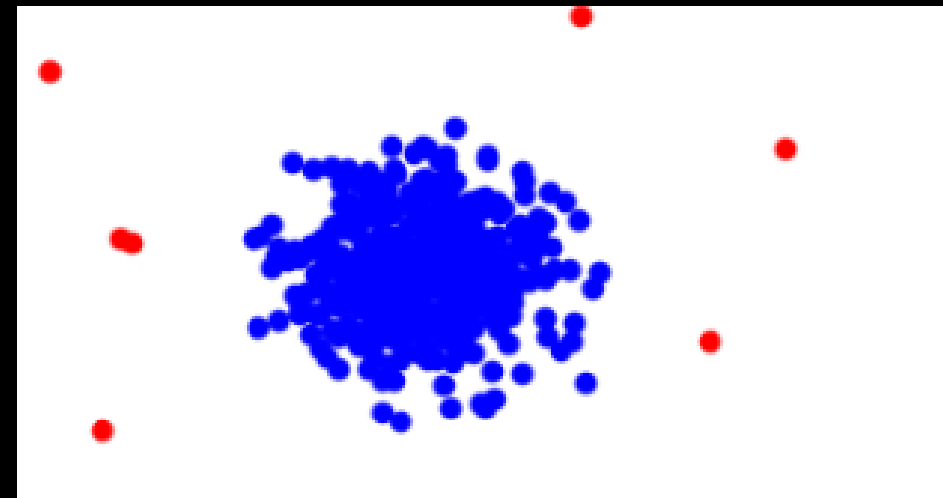




NOTE: “Outliers are the **first suspects** in data cleaning.”

Action in practice:

- Correct the value
- Or remove it **only if proven wrong**



2. Remove Outliers (Only in Specific Cases)

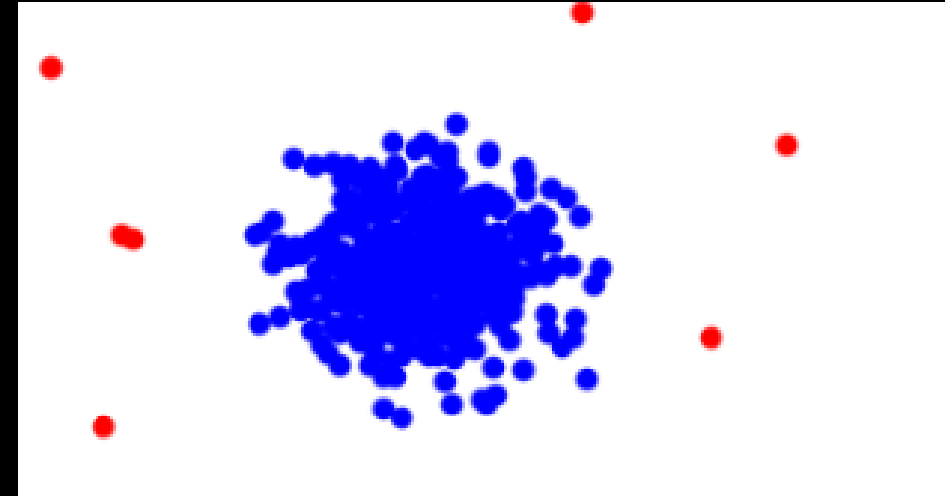
When removal is justified

- Training **distance-based models** (KNN, KMeans, Linear Regression)
- Outliers distort:
 - Mean
 - Variance
 - Distance calculations

Example

Predicting **average tumor size**

Fraud-free transaction modeling



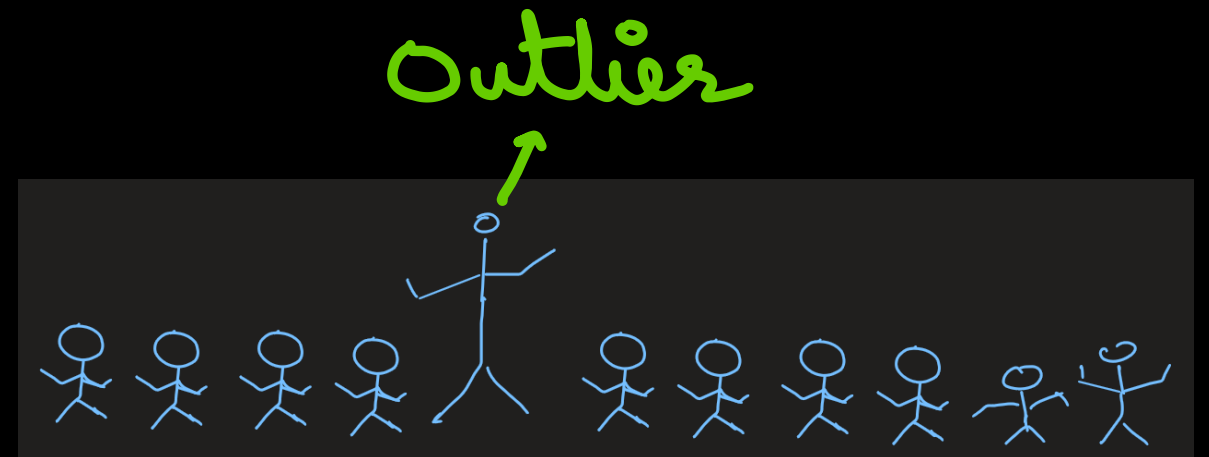
Warning: “Never delete outliers just because the model performs better.”

Check Model performance **with vs without outliers**

3. Keep Them – They Are the MOST Important Data

In real life:

- Fraud transactions
- Medical abnormalities
- Network intrusions
- Manufacturing defects
- In Breast cancer dataset, outliers could be:
 - Rare tumor patterns
 - Aggressive malignancies
 - Borderline cases

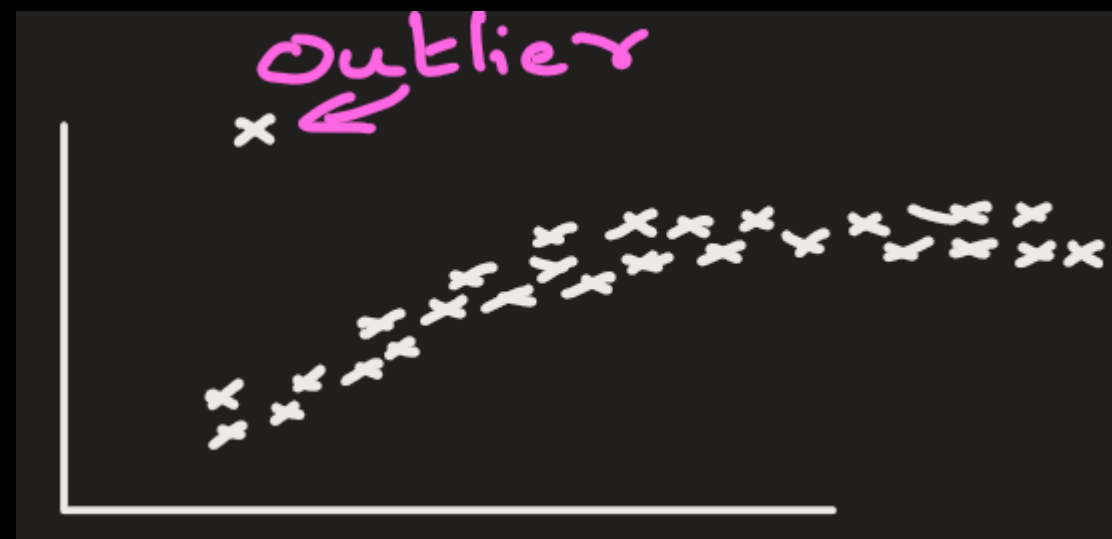




NOTE: “Outliers are often the cases doctors care about the most.”

How they're used

- Flag for **human review**
- Prioritize diagnosis
- Early warning systems



4. Convert Outliers into a Feature

Instead of deleting them:

Create a new feature

- `is_outlier_knn`
- `is_outlier_lof`
- `outlier_score`

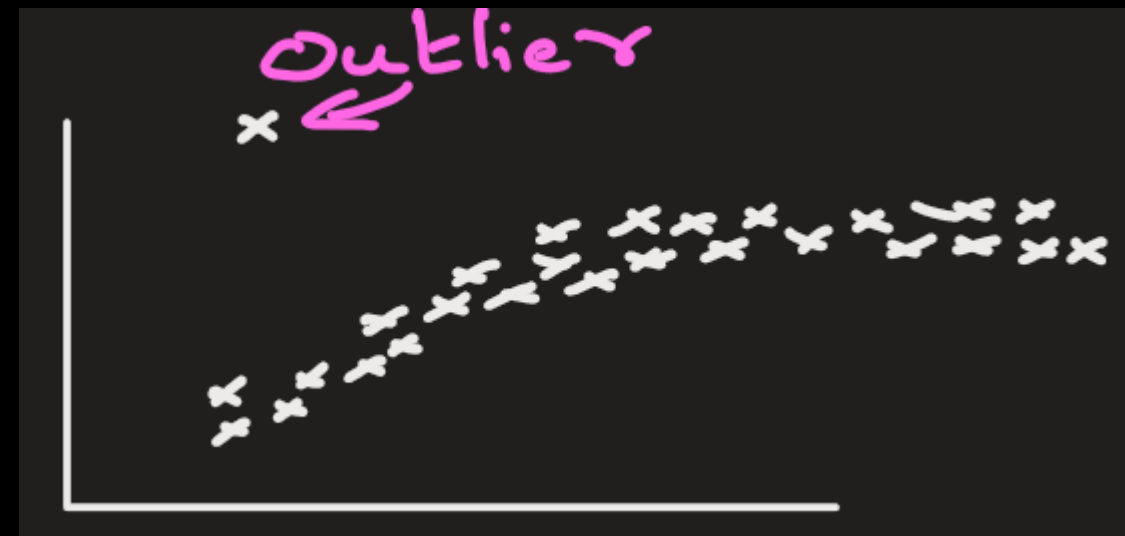
Why this is powerful

- Models learn *risk*.
- Create separate model with/without outliers.

Without outliers -> normal model

With outliers -> specialized models

- Doesn't lose information

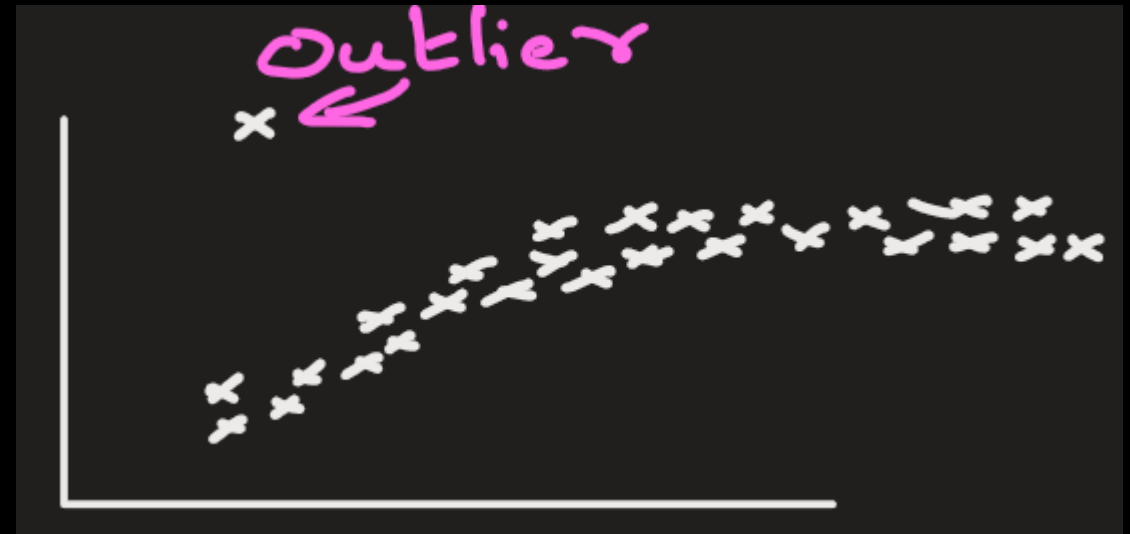




Example

- Credit risk modeling
- Medical risk prediction
- Fraud detection

Key point: “Outlier score becomes a risk score.”



5. Domain-Specific Examples (IMPORTANT)

Medical

- Rare cancer cases
- Unusual tumor growth patterns -> Data sent for senior doctor review

Finance

- Unusual transactions -> Investigate or Freeze account temporarily

Manufacturing

- Defective parts -> Machine needs recalibration
- Root cause analysis

