**OUTLIERS**

What are the outliers?

An **outlier** is an observation that lies an abnormal distance from other values in a random sample from a population.

Outliers are not always dangerous they are useful sometimes let say in anomaly detection problems. Outliers are the anomaly if you remove the outliers than you can detect the anomaly.

Effects of Outliers on Algorithms:

1. Linear regression
2. Logistic regression
3. Ada boost
4. Deep learning

Immune to Outliers Algorithm

1. Tree algorithm
2. Random forest

How to treat outliers?

1. Trimming
2. Capping
3. Treat like missing value
4. Discretization

How to detect Outliers?

1. Normal distribution [ if your point is > µ + 3σ or < µ - 3σ will be treated as outliers]
2. Skewed distribution [ detected via Box plot]
3. Other distribution [ if point is < 5 percentile and > 95 percentile will be treated as outliers]

Techniques for Outlier Detection and Removal

1. Z-score treatment [ formula for calculating z-score = Xi = Xi-mean/SD]
2. IQR Based filtering
3. Percentile
4. Winsorization