

# Outliers

# Outliers

- Outliers are extreme values that fall a long way outside of the other observations.
- Outliers are those Values which are very very different from most of the Values.
- **Example:** There is a Column called “Age” for B.tech College.
  - Where students generally have age around 17 to 24.

# BAD for Building Predictive Models?

- Because these Outliers can Induce the state of Confusion in the Predictive Models, which in turn makes the Model very Weak for Making Decision Rules.
- Then there are High Chances that the Model will not perform Good.

# Reasons for having Outliers

- Data Entry Errors
- Measurement Errors
- Experimental Errors
- Sampling Errors
- Natural Occurrence.

# Outliers Sources

- In the process of producing, collecting, processing and analyzing data, outliers can come from many sources and hide in any dimension.
- **Conclusion:** Removal or Treatment of Outliers from the Data for Building a Predictive Model is very Important as many machine learning algorithms are sensitive to the range and distribution of attribute values in the input data.