## **Summary:**

**Data Bias:**

A type of error that systematically skews results in a certain direction.

**According to Google's definition of data bias, there are four types of data bias:**

1. **Sampling bias**: This occurs when the sample of data that is collected is not representative of the population as a whole.

Avoid by using Random sapmpling!

1. **Observer bias**: The tendency for different people to observe things differently. It’s also called as Experimenter/Research Bias

Ex. Scientist see different things under the microscope

1. **Interpretation bias:** The tendency to always interpret ambiguous situations in a positive or negative way.

Ex. Two people seeing and hearing the exact same thing and interpreting it in a variety of ways.

1. **Confirmation bias:** This occurs when data is selected or interpreted in a way that confirms the preconceptions or expectations of the person collecting or analyzing the data, leading to a biased interpretation of the results.

“People see what they want to see”

It is important to be aware of these types of data bias and to try to minimize their impact when collecting and analyzing data. This can be done through techniques such as random sampling, using reliable and accurate measurement methods, and being aware of any preconceptions or expectations that may influence the interpretation of the data.

Before jumping into the Data Analytics process, first we should check the credibility of datasets: Using the ROCCC formula

A dataset should be:

Reliable

Original

Comprehensive

Current

Cited

### 