



ENHANCING DIGITAL GOVERNMENT AND ECONOMY

DATA ANALYSIS WITH PYTHON

SESSION-02: INTRODUCTION TO STATISTICS

Md. Maruf Al Hossain Prince

Lecturer

Dept. of Computer Science and Engineering

German University Bangladesh

Mobile: 01577-185858, 01752-276886

E-mail: mah-prince@ieee.org





INTRODUCTION TO STATISTICS & TERMINOLOGIES RELATED TO STATISTICS



LEARNING OBJECTIVES

- 01 UNDERSTAND THE ROLE OF STATISTICS IN DATA ANALYSIS
- 02 LEARN KEY STATISTICAL TERMS AND CONCEPTS
- 03 DIFFERENTIATE BETWEEN DESCRIPTIVE AND INFERRENTIAL STATISTICS.

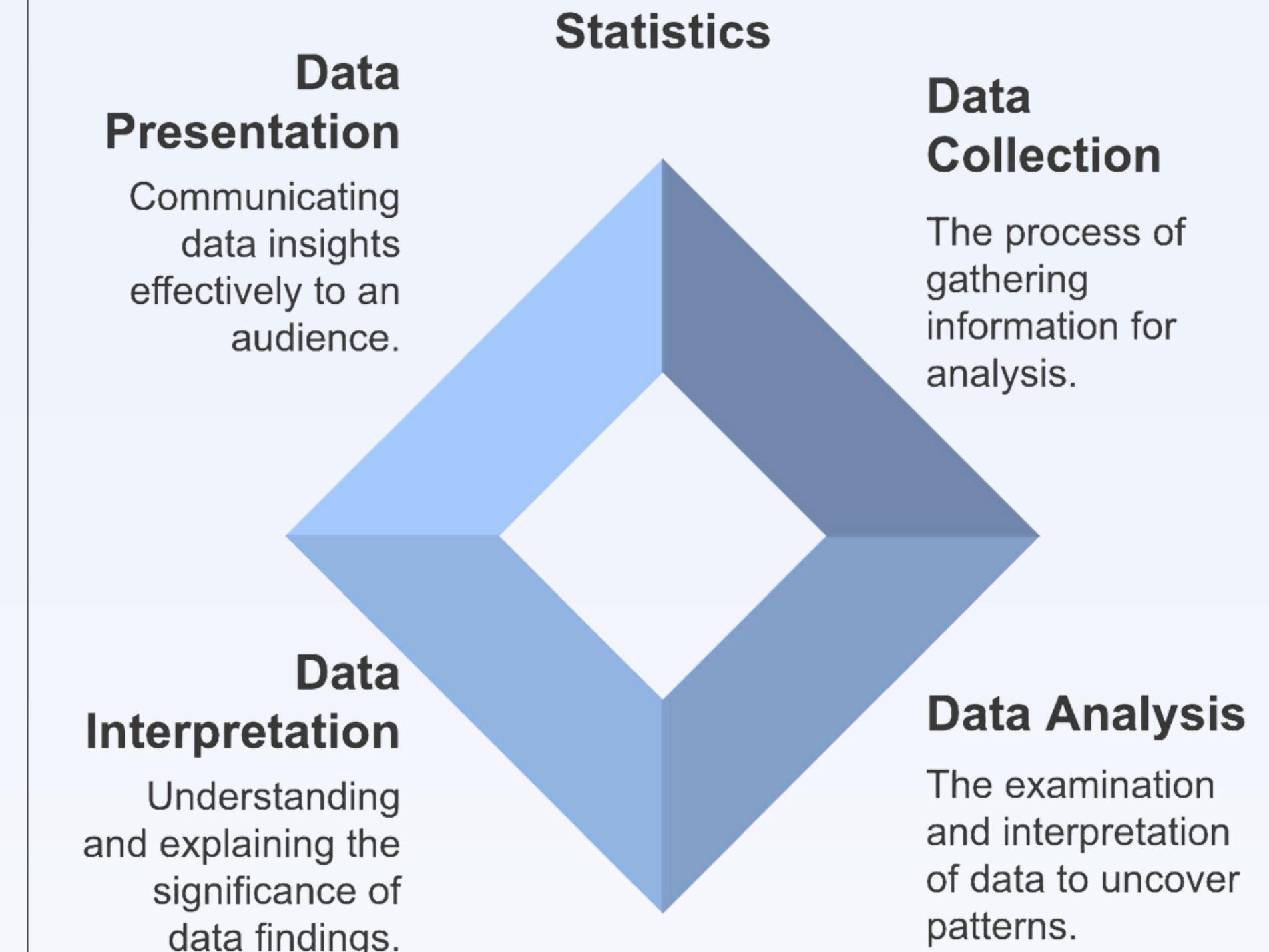
STATISTICS

WHAT IS STATISTICS

Definition: Statistics is the science of collecting, analyzing, interpreting, and presenting data.

Importance:

- Identifying trends.
- Making informed decisions.
- Solving real-world problems.

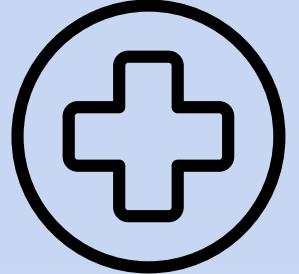


APPLICATION OF STATISTICS



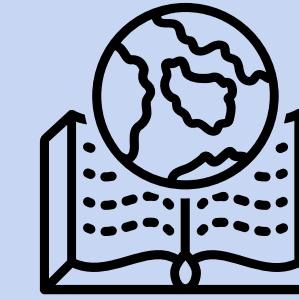
BUSINESS

Market trend analysis



HEALTHCARE

Predicting patient
outcome



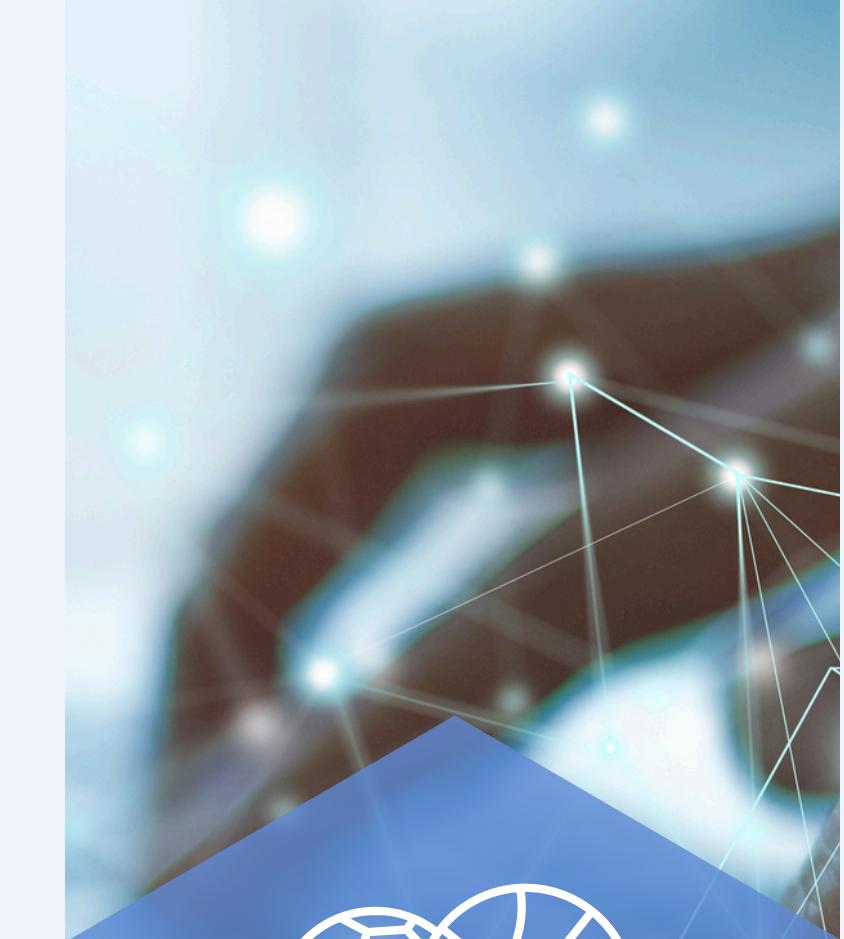
SOCIAL SCIENCES

Understanding
population behaviour



SPORTS

Performance
analytics



KEY STATISTICAL TERMINOLOGY

POPULATION

Definition

The complete set of items or individuals under study.



Examples

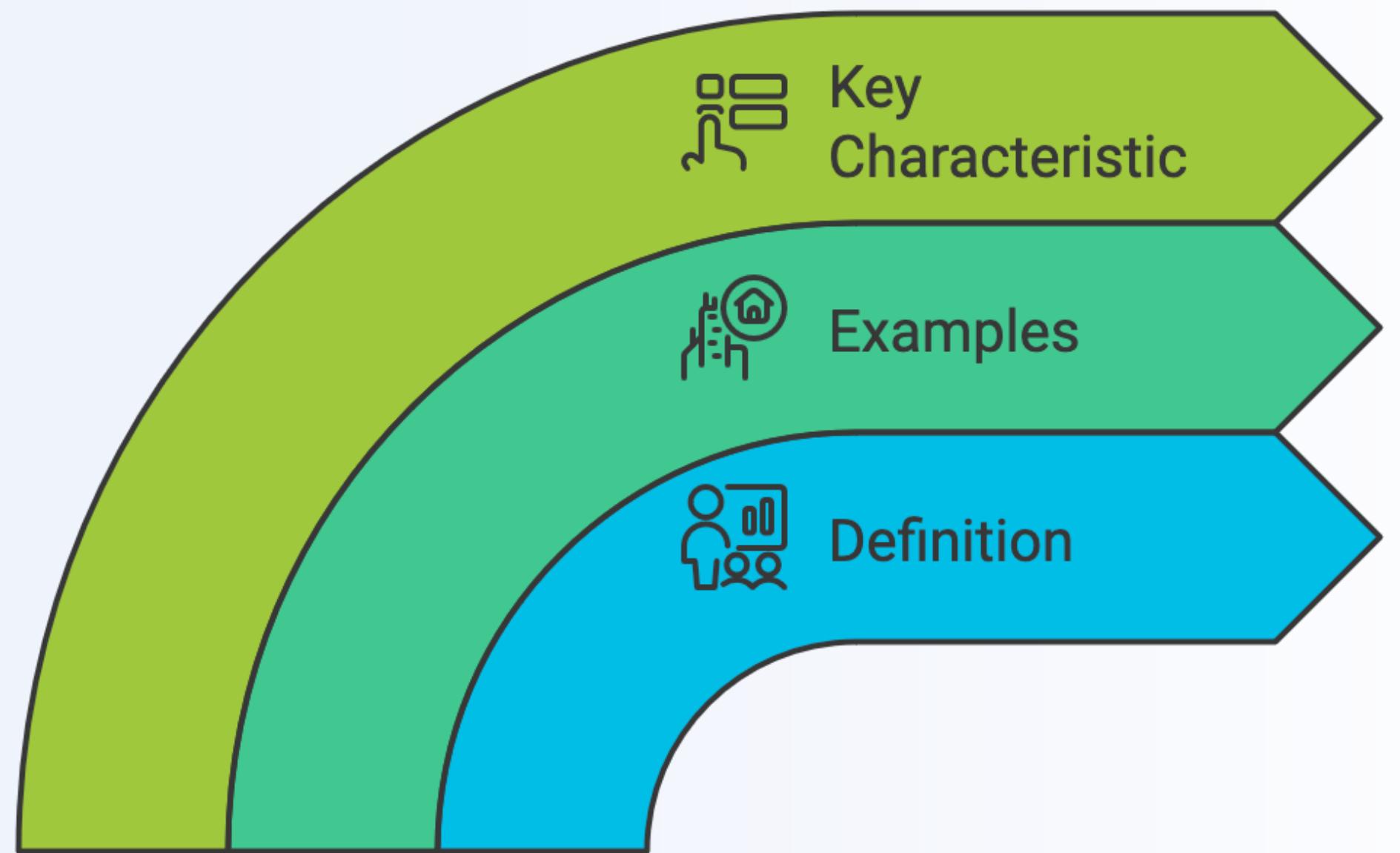
Illustrative cases like citizens or factory products.

Key Characteristic

Represents the whole group, typically large.

KEY STATISTICAL TERMINOLOGY

SAMPLE



Should be representative of the population.

A survey of 1,000 citizens from a country or inspection of 100 products from a batch of 10,000.

A smaller subset of the population, used for analysis.

KEY STATISTICAL TERMINOLOGY

SAMPLE



Population

Represents the entire group



Sample

Represents a smaller,
manageable subset

IMPORTANCE OF SAMPLING

Enables Estimation

Allows for estimation of population characteristics.



Saves Time and Resources

Sampling is more efficient than studying the whole.

Feasibility

Analyzing the entire population is often impractical.

FACTORS INFLUENCING SAMPLING

Research Goals

Aligns sampling strategy with the objectives of the study



Available Resources

Balances time and budget constraints in the sampling process

Sample Size

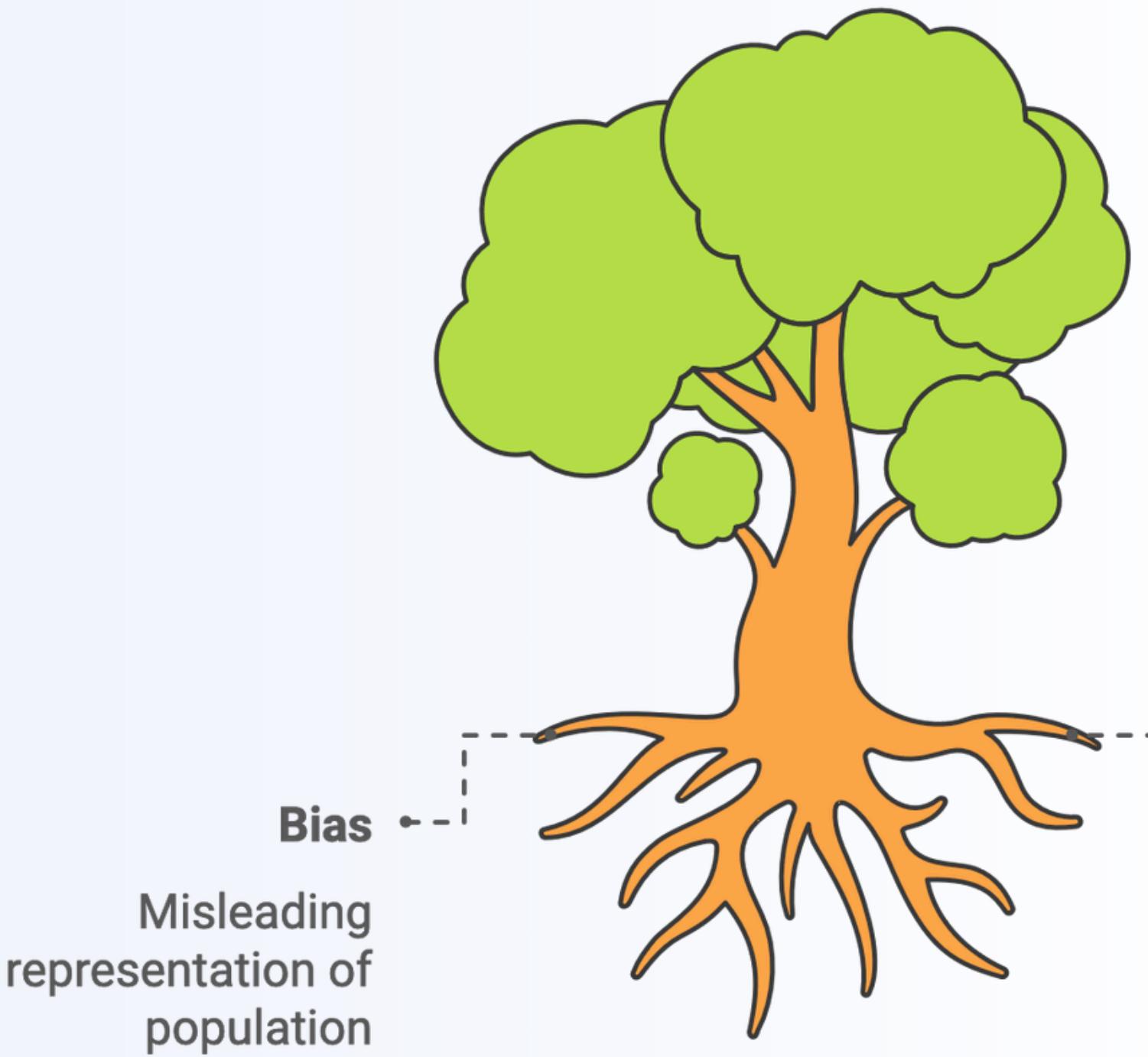
Determines the number of participants needed for reliable results

Population Diversity

Ensures representation of different subgroups in the sample

PITFALSS IN SAMPLING

Inaccurate Sampling Results



Sampling Error

Distortion of true population metrics

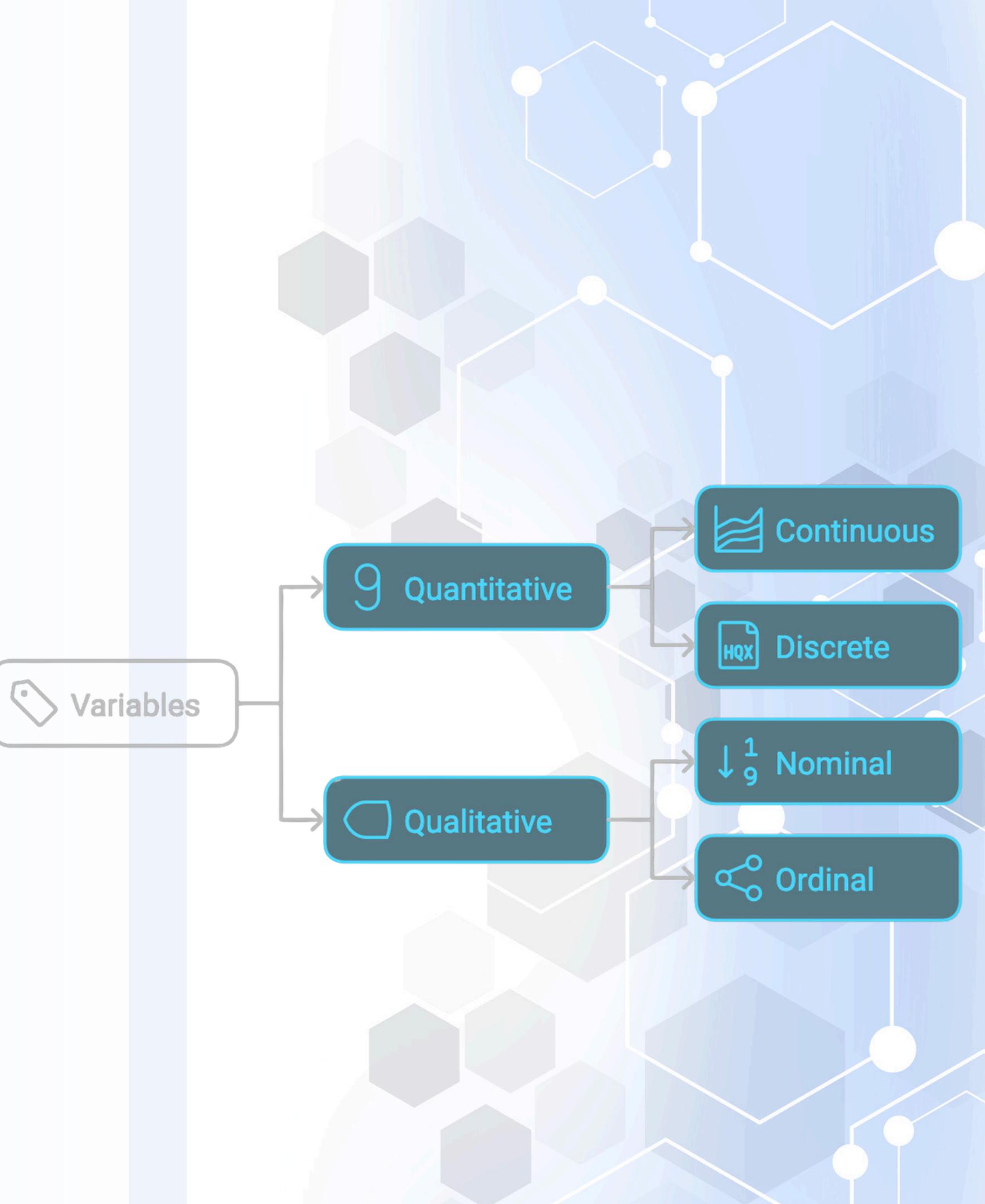


VARIABLE

Definition: Attributes or characteristics that can take on different values.

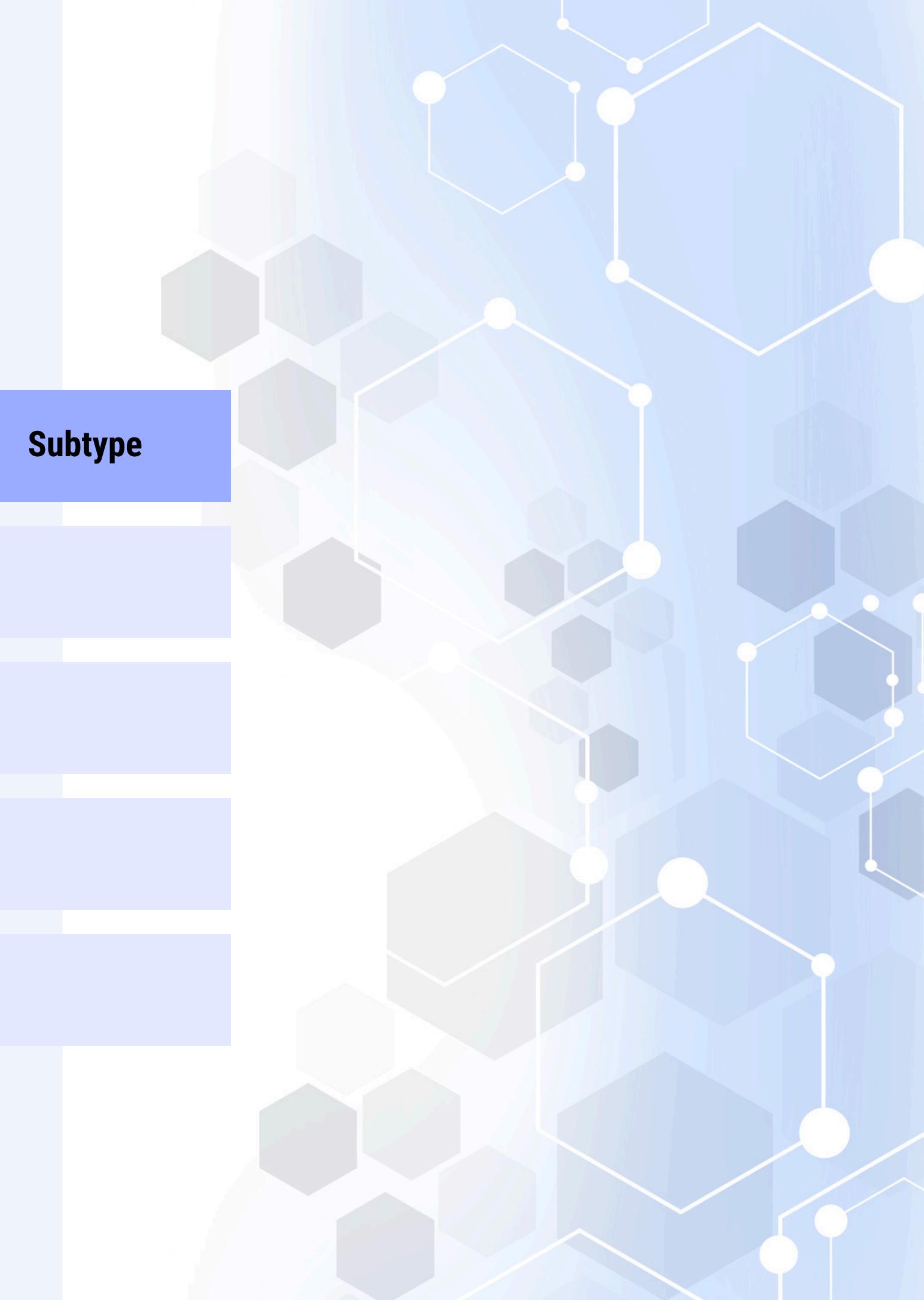
Types of Variables:

1. **Quantitative:** Represents numeric values
 - a. **Continuous:** Can take any value within a range (e.g., weight, height).
 - b. **Discrete:** Takes specific values (e.g., number of children in a family).
2. **Qualitative:** Represents categories or labels.
 - a. **Nominal:** Categories without order (e.g., gender, colors).
 - b. **Ordinal:** Categories with a meaningful order (e.g., education level).



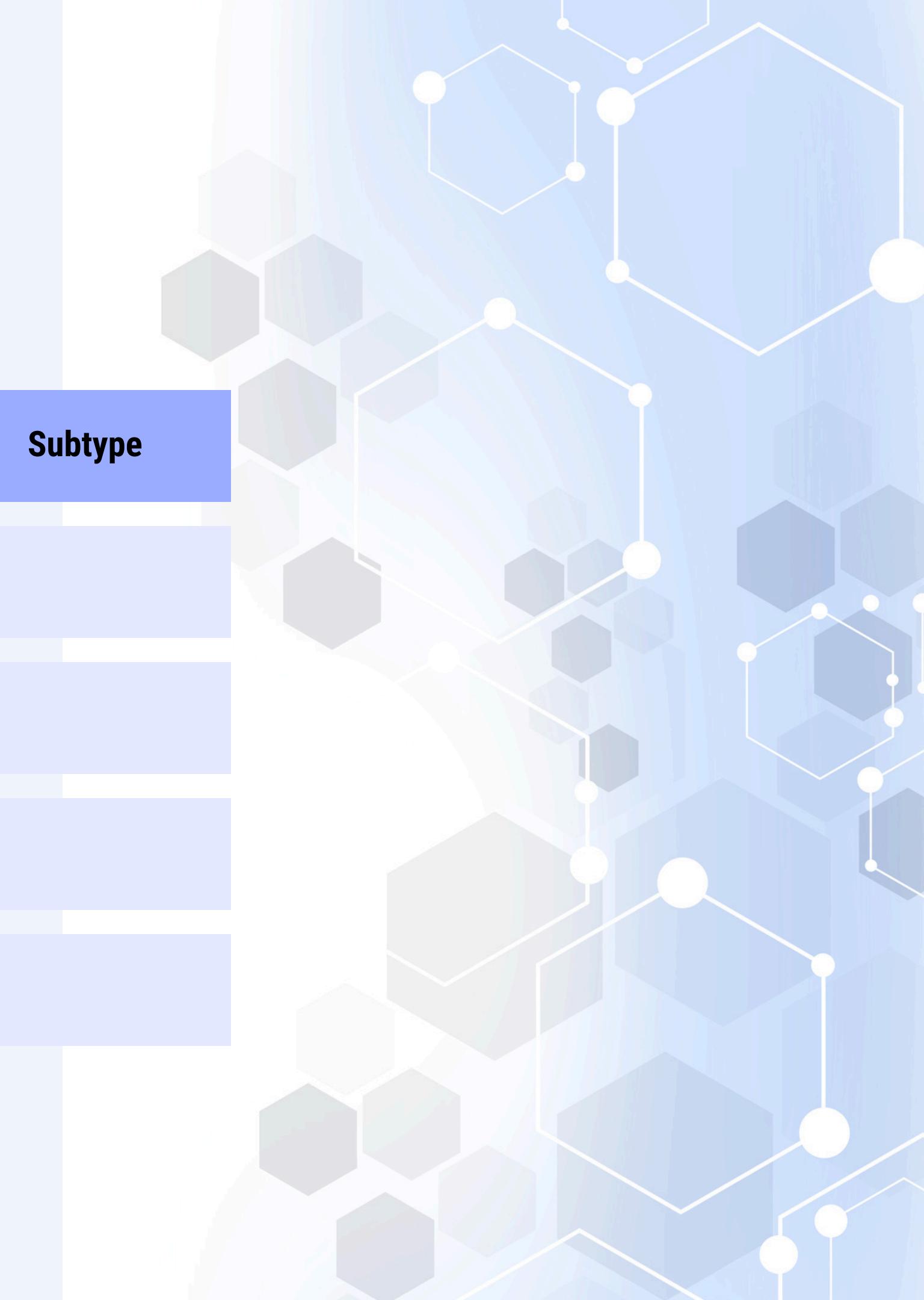
VARIABLE

Variable	Type	Subtype
Income		
Number of cars		
Marital status		
Satisfaction level		



VARIABLE

Variable	Type	Subtype
Income	Quantitative	
Number of cars		
Marital status		
Satisfaction level		



VARIABLE

Variable	Type	Subtype
Income	Quantitative	Continuous
Number of cars		
Marital status		
Satisfaction level		

VARIABLE

Variable	Type	Subtype
Income	Quantitative	Continuous
Number of cars	Quantitative	
Marital status		
Satisfaction level		

VARIABLE

Variable	Type	Subtype
Income	Quantitative	Continuous
Number of cars	Quantitative	Discrete
Marital status		
Satisfaction level		

VARIABLE

Variable	Type	Subtype
Income	Quantitative	Continuous
Number of cars	Quantitative	Discrete
Marital status	Qualitative	
Satisfaction level		

VARIABLE

Variable	Type	Subtype
Income	Quantitative	Continuous
Number of cars	Quantitative	Discrete
Marital status	Qualitative	Nominal
Satisfaction level		

VARIABLE

Variable	Type	Subtype
Income	Quantitative	Continuous
Number of cars	Quantitative	Discrete
Marital status	Qualitative	Nominal
Satisfaction level	Qualitative	

VARIABLE

Variable	Type	Subtype
Income	Quantitative	Continuous
Number of cars	Quantitative	Discrete
Marital status	Qualitative	Nominal
Satisfaction level	Qualitative	Ordinal



REAL WORLD CHALLENGES IN DATA

01

NOISY DATA

Irregularities and errors.

02

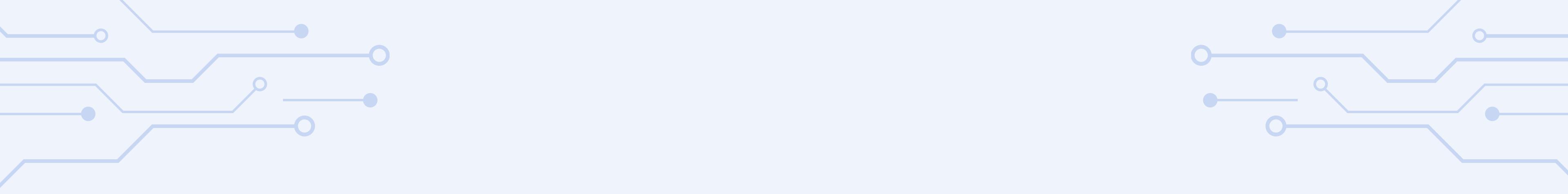
MISSING VALUES

Incomplete datasets.

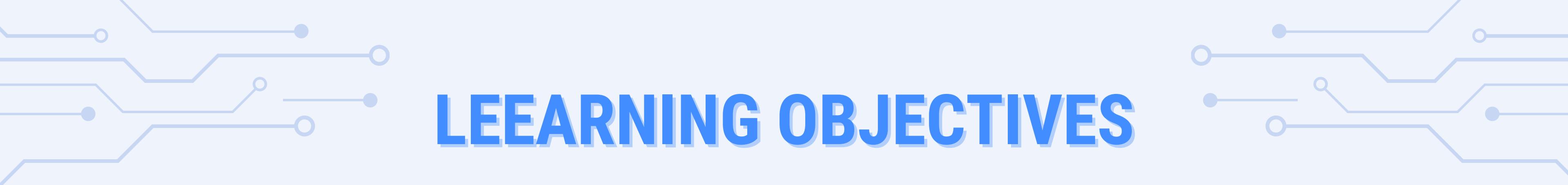
03

BIAS

Misleading trends.



DIFFERENT TYPES OF DATA, DATA COLLECTION METHODS & SAMPLING METHODS



LEARNING OBJECTIVES

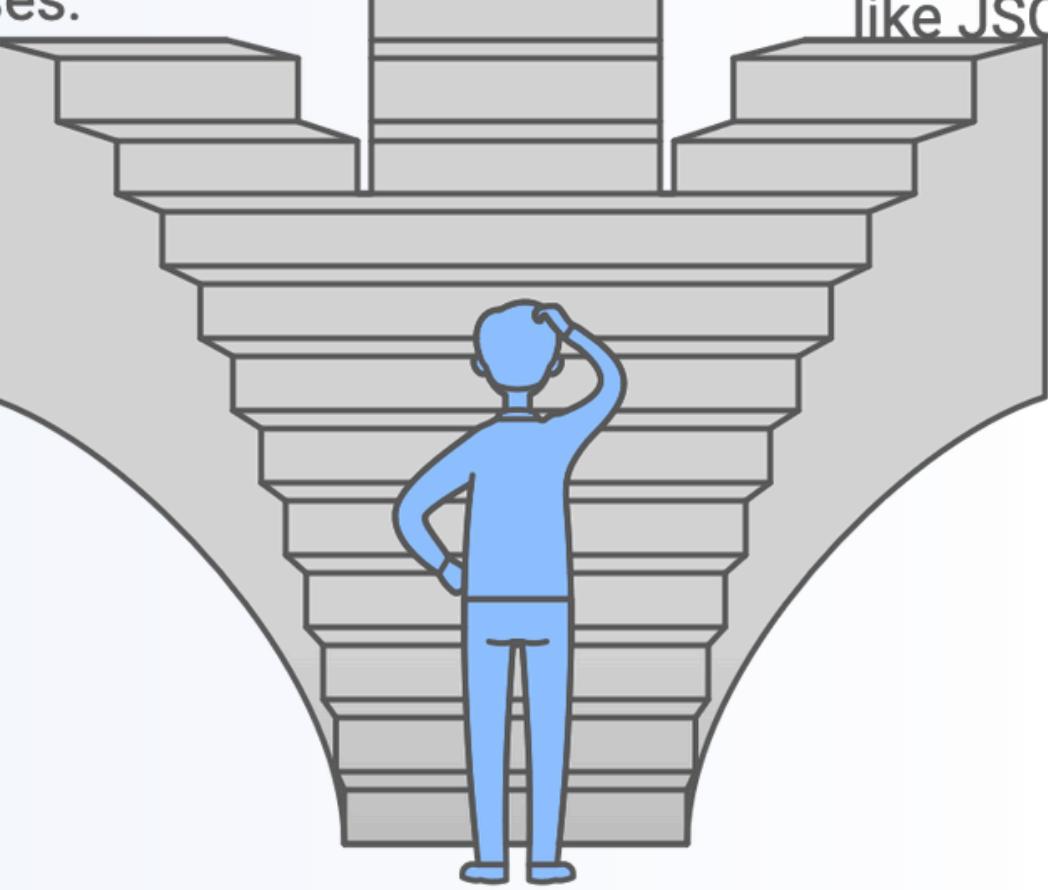
- 01 UNDERSTAND THE DIFFERENT TYPES OF DATA
- 02 EXPLORE VARIOUS DATA COLLECTION METHODS
- 03 LEARN COMMON SAMPLING TECHNIQUES AND THEIR IMPORTANCE

TYPES OF DATA

What type of data is being dealt with?

Structured Data

Organized in rows and columns, suitable for relational databases.



Unstructured Data

Lacks a predefined format, includes diverse formats like images and videos.

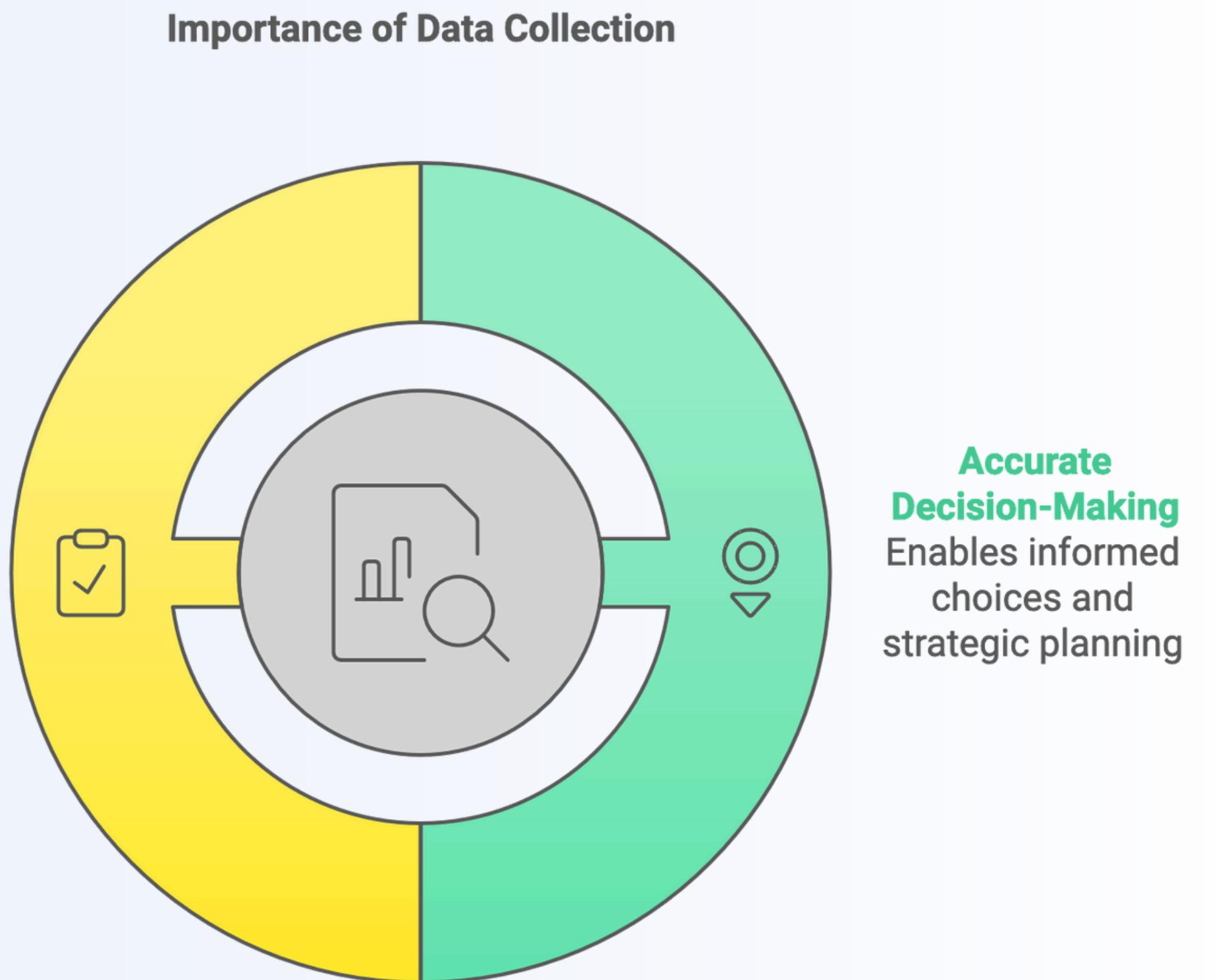


Semi-Structured Data

Contains elements of both structured and unstructured data, like JSON or XML.

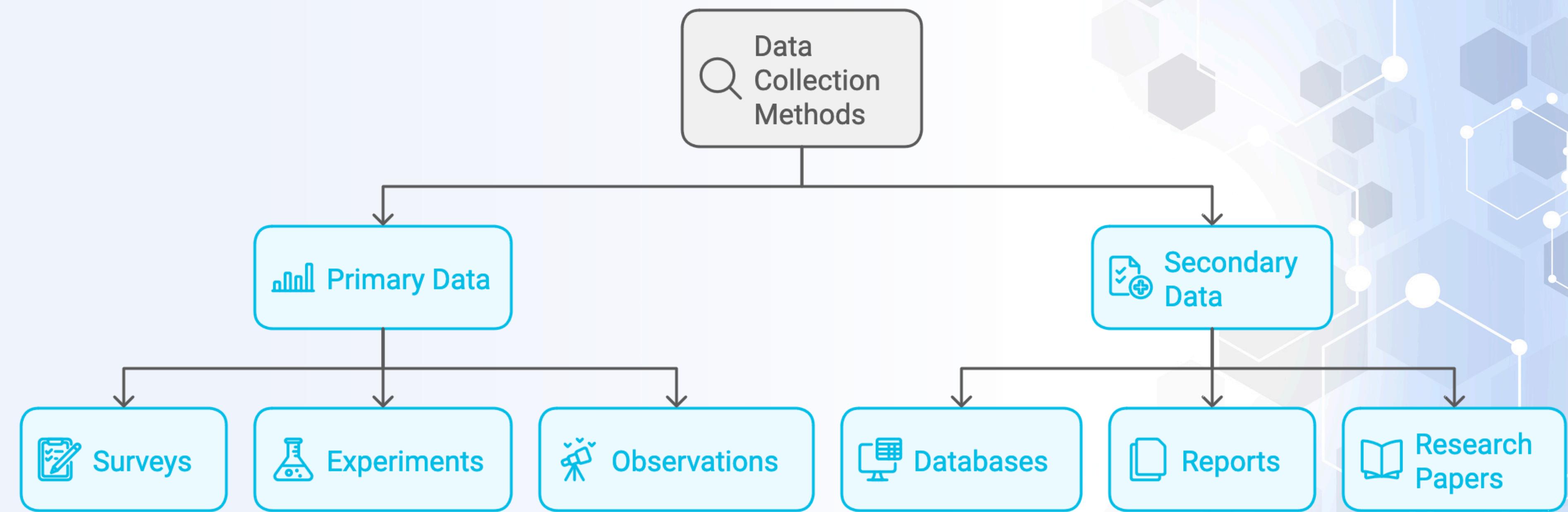


OVERVIEW OF DATA COLLECTION

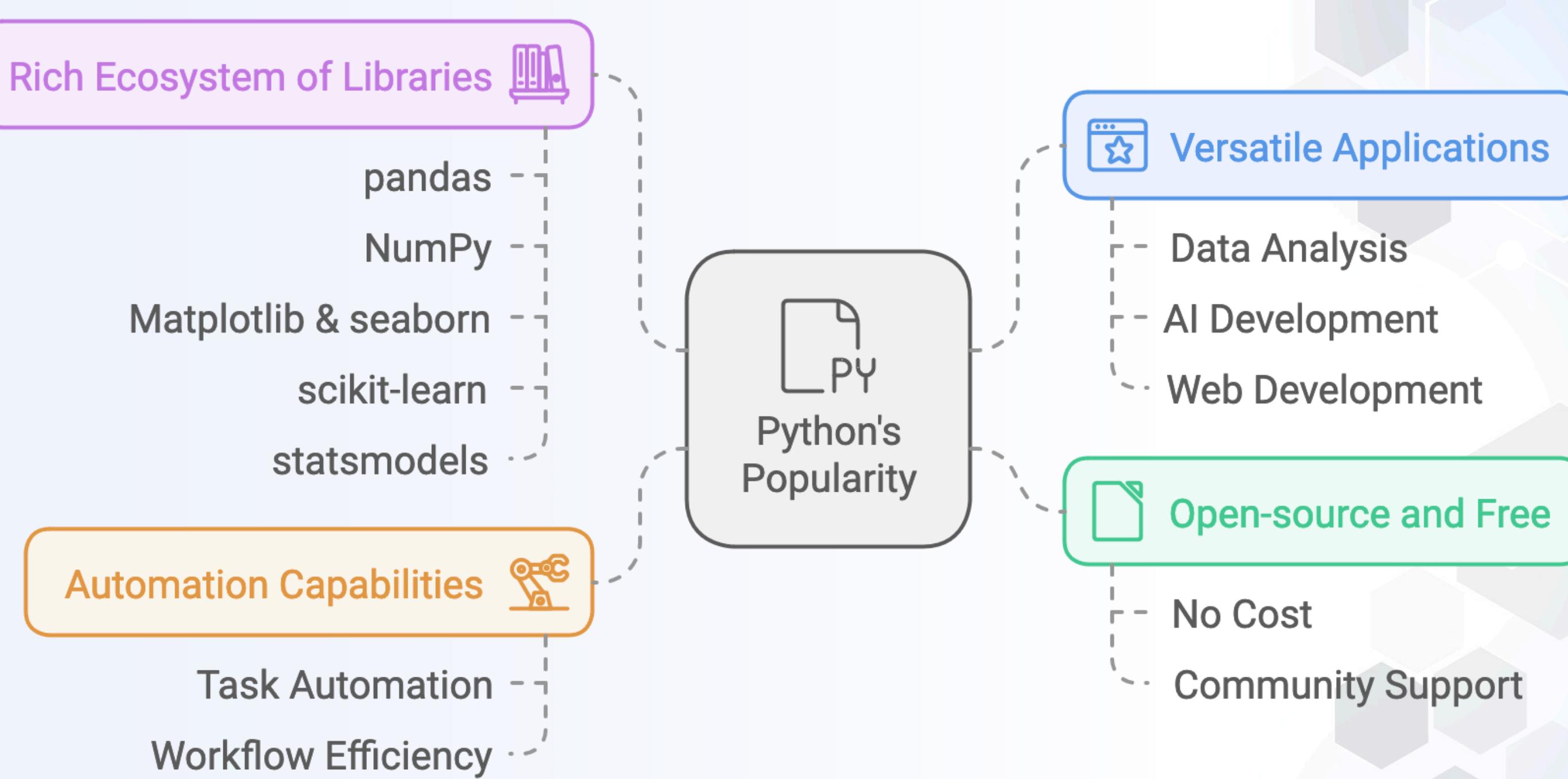


Definition: The process of gathering and measuring information.

DATA COLLECTION METHOD

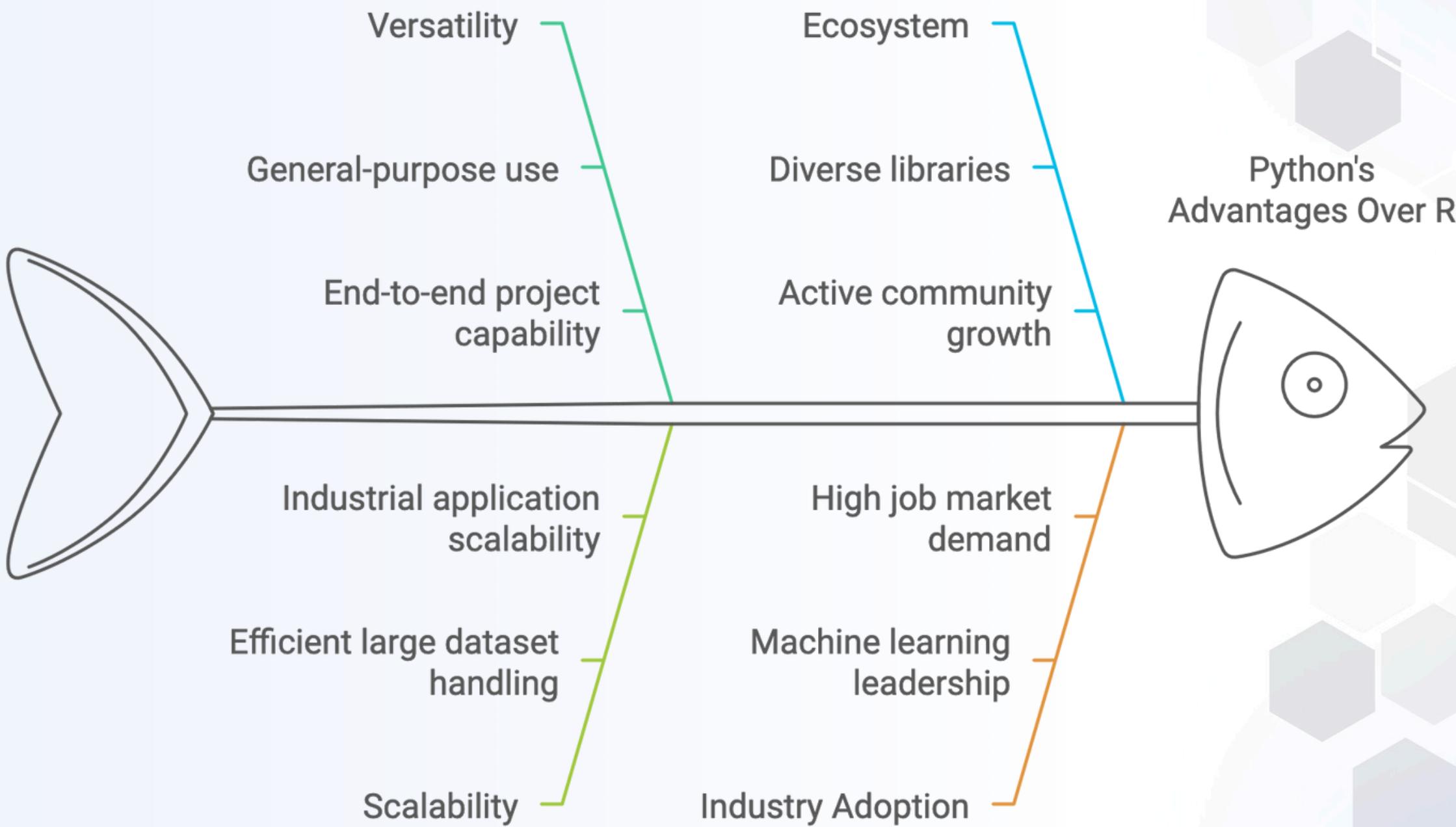


PYTHON: THE FUTURE OF DATA ANALYSIS



PYTHON'S SUPERIORITY OVER R

Python's Superiority Over R in Data Analysis



THANK YOU

GET IN
TOUCH

📞 01577-185858, 01752-276886

✉️ mah-prince@ieee.org

▶ www.youtube.com/@gyandotbd

>f www.facebook.com/gyandotbd