Python Fundamentals and Basic Statistics

Urban Data Analysis Course

Week- 2

- Python Basics and Statistics
- Content:
  - "This week, we'll dive into statistics and Python fundamentals."
  - Key focus:
    - Calculate basic statistics like mean and median.
    - Learn to work with CSV files using Pandas.
    - Understand data types and operations.

What Are Descriptive Statistics?

- Descriptive Statistics
- Content:
- **Mean:** The average of a dataset.
- **Median:** The middle value of a dataset when ordered.
- **Standard Deviation:** Measures how spread out the data is.
- **Example:**
  - Dataset: [10, 20, 30, 40, 50]
    - Mean: (10 + 20 + 30 + 40 + 50) / 5 = 30
    - Median: Middle Value = 30
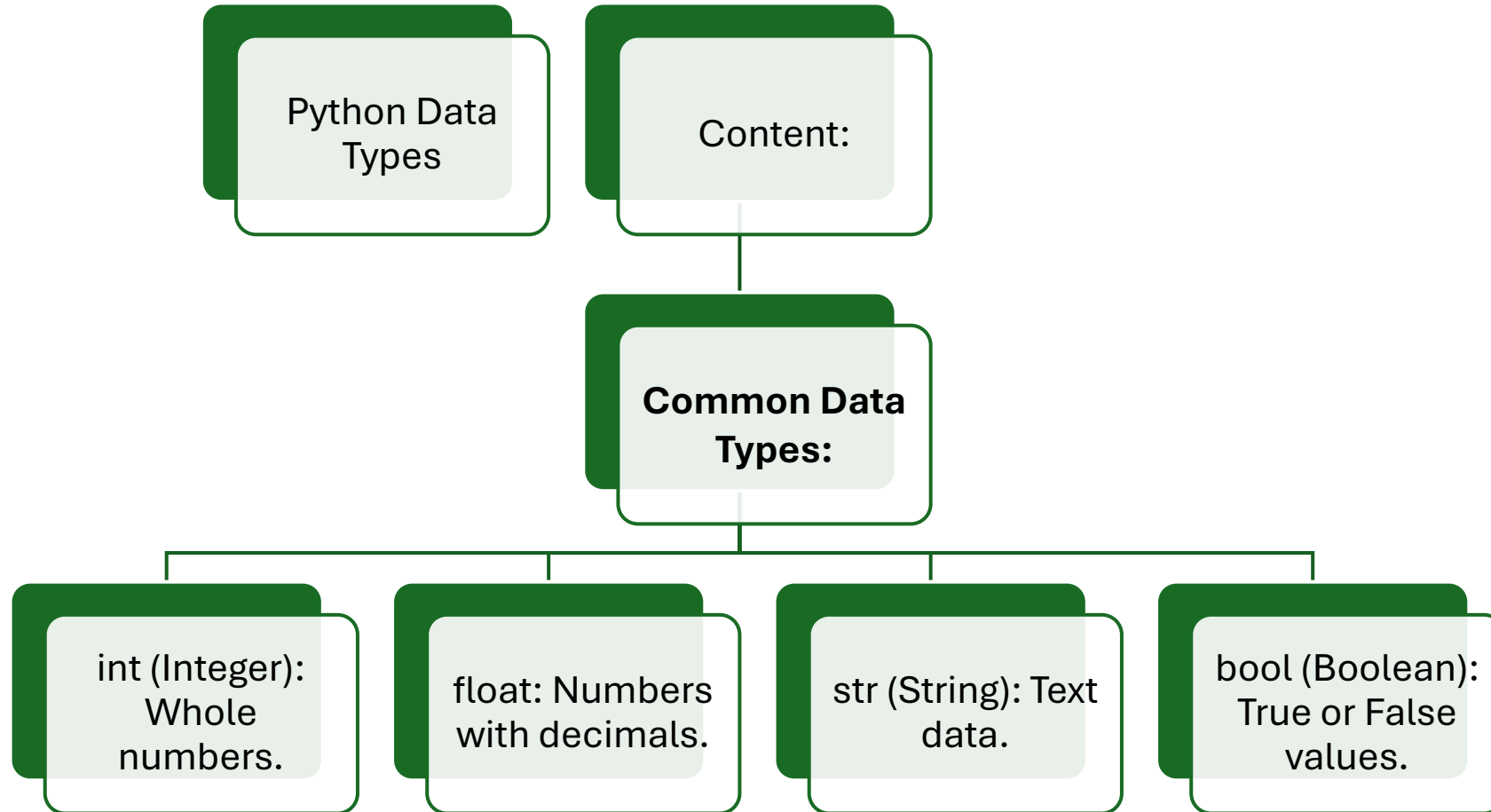   - Std. Dev: Measure of variability.

Data Types in Statistics

•Content:

•**Numerical Data:** Data expressed in numbers (e.g., age, income).

•**Categorical Data:** Data divided into groups or categories (e.g., gender, region).

•Practical Exercise:

•Identify examples of numerical and categorical data in your own datasets.

Python Data Types

Python Data Types

Content:

Common Data Types:

int (Integer): Whole numbers.

float: Numbers with decimals.

str (String): Text data.

bool (Boolean): True or False values.

Library

Library	Purpose
Pandas	Data manipulation and analysis (e.g., CSV processing, handling DataFrames).
Numpy	Numerical computations (e.g., arrays, statistics).
Seaborn	Data visualization (e.g., histograms, scatter plots, correlation matrices).
Matplotlib	Basic plotting and graphical representations.
Statsmodels	Statistical modeling and regression/ANOVA analysis.
Geopandas	Geospatial data analysis for location-specific datasets.

Introduction to Pandas

Getting Started with Pandas

•Content:

   •**What is Pandas?**

      •A Python library for data manipulation and analysis.

   •**Why Use Pandas?**

      •Simplifies working with structured data like tables.

   •**Core Features:**

      •Reading data from files (.csv, .xlsx).

      •Performing calculations on columns.

      •Manipulating and summarizing data.