

Mastering Pandas Library and EDA (Part-3)

Handling Missing Values:

- If a column has a large number of missing values, we should remove that column from the dataset.
- If only a few values are missing in a column, we can either:
 - > Remove the rows with missing values
 - > Fill the missing values using methods like mean, median, or mode

Key Questions to Ask in EDA:

- Who gathered the data?
- What is this data about?
- What is the metadata of the data? (column descriptions, data types, units, etc.)
- What are the data dimensions? (number of rows and columns)

Now look for the four things:

- Composition of data -> Understand types, nulls, unique values
- Correlation of data -> Identify relationships between variables
- Comparison -> Compare different categories or groups
- Distribution -> Analyze how values are spread

Types of Analysis in EDA:

- Descriptive analysis: Summarizing the dataset
- Diagnostic analysis: Understanding reasons behind trends or patterns

Common EDA Tasks:

- Data Exploration
- Data Cleaning
- Data Wrangling
- Data Munging
- Data Pre-processing