

COURSE GLOSSARY

Data Manipulation with pandas

- **DataFrame**: A two-dimensional labeled data structure in pandas, similar to a spreadsheet or SQL table, that holds rows and columns of data.
- **Series**: A one-dimensional array-like pandas object that contains data for a single variable, with an index to label each value.
- **Index**: The labels or row identifiers of a pandas Series or DataFrame that allow for easy data alignment and selection.
- **Mask**: A pandas Series of Boolean values (**True/False**) used to filter rows of a DataFrame.
- **Boolean Indexing**: Selecting rows or columns in a DataFrame using logical conditions that return **True** or **False**.
- **loc[]**: A pandas method used to select rows and columns by label or conditional statements.
- **iloc[]**: A pandas method used to select rows and columns by integer position.
- **Slicing**: Extracting a subset of rows or columns from a DataFrame or Series using index ranges.
- **Aggregation**: The process of summarizing data values, such as calculating the sum, mean, or count.
- **GroupBy**: A pandas method to split data into groups based on one or more variables, and then compute summary statistics.
- **Pivot Table**: A DataFrame summarization tool that allows data to be reorganized and aggregated across multiple dimensions.
- **apply()**: A pandas function that applies a given function along an axis of a DataFrame, often slower than built-in methods.
- **Imputation**: The process of handling missing values by replacing them with substituted values such as a mean, median, or placeholder.
- **fillna()**: A pandas method used to replace missing values with specified values.
- **dropna()**: A pandas method used to remove rows or columns with missing values.
- **set_index()**: A method to assign one or more columns as the index of a DataFrame.
- **reset_index()**: A method to restore the default integer index of a DataFrame, moving any existing index back into columns.
- **sort_values()**: A method to sort a DataFrame by one or more column values.
- **sort_index()**: A method to sort a DataFrame by its index values.
- **describe()**: A method to generate summary statistics (mean, median, count, etc.) for numerical columns in a DataFrame.
- **head()**: A method that displays the first five rows of a DataFrame by default.
- **info()**: A method to print concise details about a DataFrame, including column data types and non-null values.
- **read_csv()**: A pandas function to import data from a CSV file into a DataFrame.
- **to_csv()**: A pandas function to export a DataFrame to a CSV file.
- **Matplotlib**: A Python plotting library often used with pandas to create visualizations such as line charts, bar plots, and scatter plots.