|  |  |  |
| --- | --- | --- |
| Technique | Description | When to Use |
| Handling Missing Values | - `isna()`, `isnull()`: Detect missing values.<br>- `fillna()`: Fill missing values with specified values.<br>- `dropna()`: Drop rows or columns with missing values.<br>- `interpolate()`: Interpolate missing values. | - When dealing with datasets containing missing values.<br>- When preparing data for analysis or modeling.<br>- When missing values can be handled through imputation, deletion, or interpolation without significantly affecting the analysis. |
| Handling Duplicates | - `duplicated()`: Identify duplicate rows.<br>- `drop\_duplicates()`: Remove duplicate rows. | - When dealing with datasets containing duplicate entries.<br>- When preparing data for analysis or modeling.<br>- When duplicate entries need to be removed to ensure data integrity. |
| Handling Outliers | - Statistical methods (e.g., Z-score, IQR) to detect and remove outliers.<br>- `clip()`: Clip outliers to a specified range.<br>- `winsorize()`: Winsorize outliers to a specified percentile range. | - When dealing with datasets containing outliers.<br>- When preparing data for analysis or modeling.<br>- When outliers need to be addressed to ensure accurate analysis or modeling results. |
| Data Transformation | - `map()`, `apply()`: Apply a function to each element.<br>- `cut()`, `qcut()`: Bin values into discrete intervals.<br>- `astype()`: Convert the data type of a column.<br>- `transform()`: Apply a function to each group separately. | - When transforming data to a different format or scale.<br>- When preparing data for analysis or modeling.<br>- When specific transformations are required for analysis or modeling algorithms.<br>- When applying group-wise transformations. |
| String Operations | - `str.strip()`, `str.lower()`, `str.upper()`: Modify string elements.<br>- `str.contains()`, `str.replace()`: Perform searches and replacements.<br>- `str.extract()`: Extract substrings using regular expressions.<br>- `str.split()`: Split strings into substrings. | - When dealing with string data that requires cleaning or manipulation.<br>- When preparing text data for analysis or modeling.<br>- When specific string operations are needed for data processing.<br>- When extracting information from text data. |
| Indexing and Selection | - `set\_index()`, `reset\_index()`: Set or reset the DataFrame index.<br>- `loc[]`, `iloc[]`: Label-based and integer-based indexing.<br>- `at[]`, `iat[]`: Fast scalar value access.<br>- `query()`: Filter rows using a boolean expression. | - When restructuring DataFrame indexes for analysis or visualization.<br>- When selecting specific rows or columns for analysis or visualization.<br>- When accessing individual elements quickly.<br>- When filtering rows based on a boolean condition. |
| Sorting Data | - `sort\_values()`: Sort rows by column values.<br>- `sort\_index()`: Sort by index labels. | - When arranging data in a specific order for analysis or visualization.<br>- When sorting data is necessary for further data processing. |
| Reshaping Data | - `melt()`: Unpivot DataFrame from wide to long format.<br>- `pivot()`, `pivot\_table()`: Reshape DataFrame from long to wide format.<br>- `stack()`, `unstack()`: Pivot index and columns.<br>- `merge()`, `join()`: Merge DataFrame objects by index or columns. | - When restructuring DataFrame layout for analysis or visualization.<br>- When converting data between wide and long formats.<br>- When dealing with hierarchical data structures.<br>- When combining multiple datasets based on common columns or indexes. |
| Combining DataFrames | - `concat()`: Concatenate DataFrames along rows or columns.<br>- `merge()`, `join()`: Merge DataFrame objects by index or columns.<br>- `append()`: Append rows of one DataFrame to another.<br>- `combine\_first()`: Combine DataFrame objects, filling NaN values with values from another DataFrame. | - When combining multiple datasets into one for analysis or modeling.<br>- When merging datasets based on common columns or indexes.<br>- When appending rows of one DataFrame to another.<br>- When filling missing values in one DataFrame with values from another DataFrame. |
| Handling Categorical Data | - `get\_dummies()`: Convert categorical variable into dummy/indicator variables.<br>- Label encoding and one-hot encoding techniques.<br>- `replace()`: Replace values in a DataFrame. | - When preparing categorical data for analysis or modeling.<br>- When converting categorical variables into a format suitable for machine learning algorithms.<br>- When replacing specific values in a DataFrame. |
| Sampling Data | - `sample()`: Randomly select rows or columns from a DataFrame. | - When creating a subset of data for analysis or modeling.<br>- When performing random sampling for statistical analysis. |
| Handling Time Series Data | - `pd.to\_datetime()`: Convert column to datetime type.<br>- Resampling methods (`resample()`) for changing frequency.<br>- `rolling()`: Calculate rolling statistics.<br>- `shift()`: Shift index by a specified number of periods. | - When working with time series data.<br>- When analyzing or modeling temporal patterns.<br>- When aggregating data over different time periods.<br>- When calculating rolling statistics for time series analysis.<br>- When creating lag features for time series modeling. |