Summarization

Data Cleaning:

Handling Missing Values: Missing values can be dealt with by either removing the rows/columns with missing values or by filling them using statistical methods. Imputation techniques like mean, median, or more complex methods like KNN can be used to fill missing values.

Handling Outliers: Outliers can distort the analysis. Methods like IQR can be used to detect and remove outliers. Alternatively, one can use transformation methods to reduce their impact.

Data Type Conversion: Converting data types ensures that data is in the correct format for analysis. For example, dates should be in datetime format, and categorical data should be converted to category type. This helps in performing accurate computations and analyses.

Grouping Data:

Group By Single Column: Grouping by a single column allows summarizing the data by calculating statistics like mean, sum, count, etc., for each group. This helps in understanding the behavior of different categories within the dataset.

Group By Multiple Columns: Grouping by multiple columns provides a more detailed analysis by summarizing data across multiple dimensions. This is useful in identifying interactions between different variables.

Pivot Tables: Pivot tables offer a flexible way to summarize data. They can be used to display multiple statistics simultaneously and to see how data aggregates across different dimensions.

Time Datatype Summarization:

DateTime Conversion: Converting columns to datetime format is crucial for time-based analysis. This allows for operations like extracting components (year, month, day) and resampling.

Extracting Components: Extracting components from datetime columns enables detailed time-based analysis. For example, you can analyze sales by year, month, or day.

Resampling: Resampling is the process of converting time series data from one frequency to another, such as from daily to monthly. This is useful for identifying trends and seasonal patterns.

Time Series Analysis: Analyzing time series data involves looking for patterns like trends, seasonality, and cyclic behavior. This can include visualizing time series, decomposing it into trend and seasonal components, and applying forecasting models.