Group By and Aggregate Functions in Pandas

Group by >> used when I want to filter based on a specific column or when I want to do some measurements on a specific column (used when you want to analyze some pandas series by some categories).

Aggregate functions in Pandas are used to perform operations that summarize data from a DataFrame or Series, typically reducing multiple values to a single number.

Common Aggregate Functions:

1: **sum()**: Returns the sum of the values.

2: **mean()**: Returns the mean (average) of the values.

3: **min()**: Returns the minimum value

4: **max()**: Returns the maximum value.

5: **count()**: Returns the number of non-observations.

6: **std()**: Returns the standard deviation of the values.

7: **median()**: Returns the median of the values.

 ----

describe(): Provides a summary of the most common aggregate functions.

Dates and Time Series Data

To convert a column to date time data we use: to_datetime()

to_datetime(the column you want to change, format=the way you want to read the date)

.dt.month >> to extract the month

.dt.year >> to extraxt the year