

Panda Library Topics

1. Getting to know your data – In this, we shall understand the commands we can use to get the overview of the data inside the dataframe.

Commands –

- shape (Return the number of Rows & Columns)
- head, tail (Return the first 5 or last 5 records in the dataframe)
- info (information regarding each series in the dataframe)

2. Filtering and Sorting – In this, we shall understand the ways to filter and sort the data.

Commands –

- df['Series-Name'] (Return only single series at a time)
- df['Series-Name'][where] (Return only single series with specific condition)
- sort_values (sort the values in the dataset in ascending/descending order)
- sort_index (sort the index in the dataset in ascending/descending order)

3. Grouping – In this, we shall understand how to group our data in a specific format.

Commands –

- groupby('Series-Name') (Return the dataframe by grouping it based on specific series)
- agg(values) (Return the dataframe by computing special values like 'min', 'max' for each column)

4. Apply – In this, we shall understand how to apply a user defined function to a series to get specific results. E.g we can apply a function to convert the age of

Commands –

- apply(function-name) (Returns the series by applying user defined function)

5. Merge – Merge one of two series or dataframes.

Commands –

- df1.append(df2) (Returns the dataframe by appending second dataframe)
- pd.merge(left, right, how) (Returns the dataframe by merging two dataframes based on certain criteria)

6. Statistics – Used to get the statistics regarding the whole dataframe.

Commands –

- df['Series'].mean() (Returns the mean for the whole series. We can use functions like – min, max, mean, standard deviation etc.)

7. Visualization – Used to create visualization regarding the series or dataframe.

Commands –

- df['Series'].plot() (Returns the line plot for the series.)
- Some more types of visualizations - Scatter, Hist, Bar etc.

8. Time Series – Before working on time series, keep in mind that the format for the column should be datetime64.

Commands –

- to_datetime('Series') (Returns the series converted in DateTime format)
- resample('mode') (Resample the dataframe according to the mode specified. Make sure the index of the dataframe is datetime)