Lab – 1

Definition: Introduction to Pandas Library Functions

The Pandas library is a powerful and popular Python library used for data manipulation and analysis. It provides data structures and functions designed to make working with structured data seamless and efficient. Here's a summarized overview of key functions in Pandas:

Data Structures:

- Series: A one-dimensional labeled array, similar to a NumPy array but with an associated index.
- DataFrame: A two-dimensional labeled data structure with columns of potentially different data types. It is like a spreadsheet or SQL table.

Reading and Writing Data:

- pd.read_csv(): Read data from a CSV file into a DataFrame.
- pd.read_excel(): Read data from an Excel file into a DataFrame.
- pd.read_sql(): Read data from a SQL database into a DataFrame.
- df.to_csv(): Write data from a DataFrame to a CSV file.
- df.to_excel(): Write data from a DataFrame to an Excel file.
- df.to_sql(): Write data from a DataFrame to a SQL database.

Data Inspection:

- df.head(): View the first few rows of the DataFrame.
- df.tail(): View the last few rows of the DataFrame.
- df.info(): Display summary information about the DataFrame.
- df.describe(): Generate descriptive statistics of the DataFrame.

Data Selection and Filtering:

- df[column]: Access a specific column in the DataFrame.
- df.iloc[]: Access rows by their integer location.
- df.loc[]: Access rows by labels or a boolean array.
- df.query(): Filter rows using a query expression.
- df.isnull(): Check for missing values.
- df.dropna(): Remove rows with missing values.
- df.fillna(): Replace missing values with specified values.

Data Manipulation:

- df.groupby(): Group data based on a specified column or columns.
- df.merge(): Merge two DataFrames based on common columns.
- df.sort_values(): Sort DataFrame by one or more columns.
- df.rename(): Rename columns or index labels.

- df.apply(): Apply a function along rows or columns.

Data Aggregation:

- df.sum(): Compute the sum of values in each column.
- df.mean(): Compute the mean of values in each column.
- df.count(): Count non-null values in each column.
- df.min()/df.max(): Find the minimum/maximum value in each column.
- df.groupby().agg(): Perform custom aggregation functions on grouped data.

Data Visualization:

- df.plot(): Create basic plots directly from the DataFrame.
- pd.plotting: Module for various plotting functions.

These are just a few highlights of the many functions and capabilities provided by the Pandas library. It is widely used in data science and analysis workflows due to its simplicity and versatility.