Most Used Pandas Functions for Data Analytics

Data Loading

pd.read_csv() - Load CSV files

pd.read_excel() - Load Excel files

pd.read_json() - Load JSON files

pd.read_sql() - Read from SQL database

df.to_csv() - Save dataframe to CSV

df.to_excel() - Save dataframe to Excel

Data Exploration & Summary

df.head(n) - View top n rows

df.tail(n) - View bottom n rows

df.info() - Summary of dataframe

df.describe() - Statistical summary

df.shape - Rows and columns

df.columns - Column names

df.dtypes - Data types

df.isnull().sum() - Count missing values

df.value_counts() - Count unique values in a Series

Data Selection & Filtering

df['column'] or df.column - Select a column

df[['col1', 'col2']] - Select multiple columns

df.loc[] - Label-based indexing

df.iloc[] - Position-based indexing

df[df['column'] > value] - Filter rows by condition

Data Cleaning

df.dropna() - Drop missing values

df.fillna(value) - Fill missing values

Most Used Pandas Functions for Data Analytics

df.drop(columns=['col']) - Drop column(s)

df.rename(columns={'old':'new'}) - Rename columns

df.duplicated() / df.drop_duplicates() - Handle duplicates

df.astype() - Convert data types

Data Transformation

df.apply() - Apply function to each row/column

df.map() - Map values in a Series

df.replace() - Replace values

df.sort_values(by='col') - Sort dataframe

df.reset_index() - Reset index

df.set_index('col') - Set column as index

Grouping and Aggregation

df.groupby('col') - Group by column

df.agg({'col': 'sum'}) - Aggregate

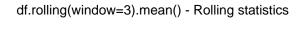df.pivot_table() - Create pivot table

df.crosstab() - Cross-tabulation

Merging and Joining

pd.concat([df1, df2]) - Concatenate dataframes

pd.merge(df1, df2, on='key') - Merge on key column

df.join(df2) - Join based on index

Time Series

pd.to_datetime(df['date']) - Convert to datetime

df.set_index('date') - Set datetime as index

df.resample('M').mean() - Resample by month

df.shift(1) - Shift data

Most Used Pandas Functions for Data Analytics

df.rolling(window=3).mean() - Rolling statistics

Visualization (with Pandas built-in)

df.plot()

df.hist()

df.boxplot()