Comprehensive List of

 Pandas

Functions and methods

Rajendra Prasad

https://chlorinexxe.github.io/portfolio

1) DATA FRAME CREATION

- pd.DataFrame(data, index, columns) : Create a new DataFrame from various inputs.

2) VIEWING/INSPECTING DATA

- df.head(n), df.tail(n) : Returns first or last n rows.
- df.info() : Provides a concise summary of dataFrame.
- df.describe() : Generates descriptive statistics.
- df.shape : Returns dimensions (rows, columns).
- df.columns : Returns column labels.
- df.index : Returns row labels.

3)Selection / Filtering

- df[col], df[[col1, col2]] : Selects single or multiple columns.
- df.iloc[index], df.iloc[row, col] : Integer-location based indexing.
- df.loc[label], df.loc[row_label, col_label] : Label-based indexing.
- df.query('expression') : Query the DataFrame with a boolean expression.
- df.filter(items=['col1', 'col2']) : Filters by labels/columns.
- df.isin(values) : Checks whether elements are contained in DataFrame.
- df.where(cond, other) : Replace values where the condition is False.
- df.mask(cond, other) : Replace values where the condition is True.

4)Data Cleaning / manipulation

- df.drop(labels, axis) : Drops specified rows or columns.
- df.drop_duplicates() : Removes duplicate rows.
- df.fillna(value) : Fills NaN values with specified value.
- df.replace(to_replace, value) : Replaces values.
- df.rename(columns={'old_name': 'new_name'}) : Renames columns.
- df.sort_values(by='column_name') : Sorts by values.
- df.groupby('column_name') : Groups by column.
- df.pivot_table(values, index, columns, aggfunc) : Creates a pivot table.
- df.melt(id_vars, value_vars) : Unpivots a DataFrame from wide to long format.

5)Arithmetic and Statistical Operations

- df.mean(), df.median(), df.sum(), df.min(), df.max() : Aggregate functions.
- df.std(), df.var(), df.corr(), df.cov() : Statistical functions.
- df.diff(periods), df.pct_change(periods) : Difference and percent change

- *Std* - *Standard Deviation*
- *Var* - *Variance*
- *Corr* - *Corelation*
- *Cov* - *Covariance*
- *Diff* - *Returns dataframe with diff b/w current and specified element*
- *Pct_change* - *Returns the diff in percentage*

6) COMBINING DATAFRAME

- pd.concat(objs, axis) : Concatenates DataFrames.
- df.append(other) : Appends rows of other DataFrame.
- df.join(other, on='key') : SQLstyle join.
- Pd.merge([dataframes],on,how) : Merge on specific column

7) RESHAPING / MANIPULATION

- df.stack(), df.unstack() : Pivots a level of the (possibly hierarchical) index.
- df.transpose(), df.T : Transposes rows and columns.
- df.set_index(keys), df.reset_index() : Sets and resets index.

8) Time series

- pd.date_range(start, end, freq) : Generates date range.
- df.resample(rule) : Resamples time series data.

9) INPUT / OUTPUT

- pd.read_csv(), pd.read_excel(), pd.read_sql() : Reads data from various sources.
- df.to_csv(), df.to_excel(), df.to_sql() : Writes data to file or database.

10) Plotting

- df.plot(), df.hist(), df.boxplot() : Data visualization.

11) MISCELLANEOUS

- df.apply(func), df.applymap(func) : Applies a function to DataFrame.
- df.dtypes : Returns data types.
- df.memory_usage() : Returns memory usage.

“

I hope you found this information helpful!
Feel free to save this post for future reference.

Let's continue to learn and grow together! 🚀

”

Thank You for Your Support

Rajendra Prasad JM

https://chlorinexxe.github.io/portfolio