# Read Data Functions

|  |  |  |
| --- | --- | --- |
| Function Name | Description | Link |
| read\_csv | Read delimited data from a file. Use Comma as default delimiter | https://pandas.pydata.org/docs/reference/api/pandas.read\_csv.html |
| read\_table | Read delimited data from a file. Use tab ('\') as default delimiter | https://pandas.pydata.org/docs/reference/api/pandas.read\_table.html |
| read\_excel | Read data from excel file | https://pandas.pydata.org/docs/reference/api/pandas.read\_excel.html |
| read\_fwf | Read data in fixed width column format | https://pandas.pydata.org/docs/reference/api/pandas.read\_fwf.html |
| read \_clipboard | Read data from clipboard. Useful for converting tables from web pages | https://pandas.pydata.org/docs/reference/api/pandas.read\_clipboard.html |
| Read\_json | Read data from JSON | https://pandas.pydata.org/docs/reference/api/pandas.read\_json.html |

# Data Frame Functions

| Function Name | Function Description | Parameters | Link |
| --- | --- | --- | --- |
| head() | Shows top 5 rows | * n: Number of rows to select | https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.head.html |
| Tail() | Show bot 5 rows | * n: Number of rows to select | https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.tail.html |
| Info() | Info about db |  |  |
| Describe() | Show data summary and statistical methods values | * Percentiles * Include * Exclude * Dateime\_is\_numeric : bool | https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.describe.html |
| Shape | No of rows and columns |  |  |
| Columns | Column names |  |  |
| Dtypes | Types of columns |  |  |
| isnull().sum() | Count sum values in each column | - | https://pandas.pydata.org/docs/reference/api/pandas.isnull.html |
| Value\_counts() | Count number of unique values in column | * Normalize : bool * Sort : bool * Ascending : bool * Bins : groups * Dropna : bool | https://pandas.pydata.org/docs/reference/api/pandas.Series.value\_counts.html |
| Unique() | Unique values of column | - | https://pandas.pydata.org/docs/reference/api/pandas.unique.html |
| Numique() | Count of unique values | * Axis : 0,1 * Dropna : bool | https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.nunique.html |
| Copy() | Copy dataframe | * Bool Deep : (a) True : changes will not be saved int the original dataframe  (B) False : changes will be saved to the original dataframe | <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.copy.html> |
| Rename() | Rename columns | * Mapper * Index * Columns : dictionary * Axis : {0 or ‘index’, 1 or ‘columns’}, * Copy : bool, default True * Inplace : bool, default False * Level : int or level name, default None * error : {‘ignore’, ‘raise’}, default ‘ignore’ | https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.rename.html |
| Drop() | Remove column/s | * labels : index to drop * axis : 0 rows 1 columns * index : row/s index to drop * columns : column/s name to drop * level : * inplace : true : current data frame , false returns a copy * errors : {‘ignore’, ‘raise’}, default ‘ignore’ | https://www.w3schools.com/python/pandas/ref\_df\_drop.asp |
| Sort\_values() | Sort data by column | * by : rows or columns to sort by * axis : 0 rows , 1 columns * ascending : true ,false * inplace : true , false * kind : {'quicksort' default , 'mergesort' ,  'heapsort'} * na\_position : {'first' ,  'last' default } where to put null values * ignore\_index * key : function to execute before sorting | https://www.w3schools.com/python/pandas/ref\_df\_sort\_values.asp |
| Dropna() | Remove null values from rows/col | * axis : 0 rows 1 columns * how : {‘all’ ,  ‘any’ default} all values are null or any value is null * tresh : number of nulls to remove * subset : where to look for null * inplace : true or false | https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.dropna.html |
| Groupby() | Group results by value | * by : rows or col to group * axis : 0 rows 1 columns * level : bool * as\_index : bool * sort : bool * group\_keys : bool * observed : if data is categorical * dropna : bool | https://www.geeksforgeeks.org/python-pandas-dataframe-groupby/ |
| Loc() | Access data by condition |  | <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.loc.html>  https://www.geeksforgeeks.org/python-pandas-dataframe-loc/ |
| Select() | Select data based on criteria | * criteria * assign values | NUMPY not Pandas |
| Pivot() | Create pivot table | * data * values * index * columns * aggfunc * fill\_value * margins * dropna * margins\_name * observed * sort | https://pandas.pydata.org/docs/reference/api/pandas.pivot\_table.html  https://www.geeksforgeeks.org/python-pandas-pivot\_table/ |
| Sort() | Depreciated |  |  |
| Drop\_duplicates() | Remove duplicates | * subset : column to check * keep : {'first' false 'last' False } which duplicates to keep * inplace * ignore\_index | <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.drop_duplicates.html>  https://www.w3schools.com/python/pandas/ref\_df\_drop\_duplicates.asp |
| Isnull() | Missing values from dataframe |  | https://www.w3schools.com/python/pandas/ref\_df\_isnull.asp  https://pandas.pydata.org/docs/reference/api/pandas.isnull.html |
| Fillna () | Fill missing value with a value | * value : value to replace * method : {'backfill' 'bfill' 'pad' 'ffill' None default} * axis : * inplace * limit : maximum number to fill * downcast : values to fill | <https://www.w3schools.com/python/pandas/ref_df_fillna.asp>  https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.fillna.html |
| Join() | Join two data frames | * other : dataframe * on : * how: {'left' default 'right' 'outer' 'inner' } * lsuffix * rsuffix * sort : bool | https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.join.html |
| Select\_dtypes() | Select only data types from data frame | * Include : data type to show * Exclude : data types to hide | https://www.geeksforgeeks.org/python-pandas-dataframe-select\_dtypes/ |
| Isna() | Check if it has value |  | https://pandas.pydata.org/docs/reference/api/pandas.isna.html |