|  |  |
| --- | --- |
| **Utility** | **Functions** |
| Extract Column Names | df.columns |
| Select first 2 rows | df.iloc[:2] |
| Select first 2 columns | df.iloc[:,:2] |
| Select columns by name | df.loc[:,["col1","col2"]] |
| Select random no. of rows | df.sample(n = 10) |
| Select fraction of random rows | df.sample(frac = 0.2) |
| Rename the variables | df.rename( ) |
| Selecting a column as index | df.set\_index( ) |
| Removing rows or columns | df.drop( ) |
| Sorting values | df.sort\_values( ) |
| Grouping variables | df.groupby( ) |
| Filtering | df.query( ) |
| Finding the missing values | df.isnull( ) |
| Dropping the missing values | df.dropna( ) |
| Removing the duplicates | df.drop\_duplicates( ) |
| Creating dummies | pd.get\_dummies( ) |
| Ranking | df.rank( ) |
| Cumulative sum | df.cumsum( ) |
| Quantiles | df.quantile( ) |
| Selecting numeric variables | df.select\_dtypes( ) |
| Concatenating two dataframes | pd.concat() |
| Merging on basis of common variable | pd.merge( ) |