|  |
| --- |
| import pandas as pd |
| nums=pd.Series([1,2,3,4]) |
| nums.index=[‘a’,’b’,’c’,’d’] |
| df=pd.DataFrame(lst) # lst is a numpy list |
| df[2] # Get the Column with Column Name 2 |
| df.columns=[‘A’,’B’] # Set the column Names |
| df.shape |
| df.head(n=3) #Get the top 3 Rows |
| df.tail(n=3) # Last 3 Rows |
| df[[‘A’,’B’]] #Extracting multiple Columns |
| df[‘A+B’]=df[‘A’]+df[‘B’] #Add a new Column with is a result of operations on other columns |
| df.drop(columns=[‘A+B’]) #Drop a column |
| df.index=[‘P’,’Q’,’R’,’S’] #Changing the Row Name |
| df.loc[‘P’] #Extract Row ‘P’ |
| df.iloc[-2:] #Extract the Row by index not by RowName |
| df.iloc[[0,1,4]] #Extract Multiple Rows |
| df.iloc[0][[‘A’,’B’]] #Extract A and B Column of First Row. |
| mask=df[‘B’]>30 #Mask Creation |
| df[mask] #Extract Rows with Column B value >30 |
| dfArray=df.value #Convert DataFrame into numpy matrix |
| df.dropna() #Droping all the Rows with Nan values |
| df[‘B’]=df[‘B’].fillna(value=10) #Fill the Nan value in B column with Number 10. |
| pd.concat([df1,df2]) #Concat df1 and df2 DataFrames |
| df1.merge(df2,how=’inner’) #Inner Join df1 & df2 |
| df1.merge(df2,how=’outer’) #OuterJoin df1 & df2 |
| df1.merge(df2,how=’left’) #LeftOuterJoin df1 & df2 |
| df1.merge(df2,how=’right’) #RightOuterJoin df1 & df2 |
| df=pd.read\_csv(“./xyz.csv”) #Read a csv file |
| df.desribe() #Get all the aggregate values for each column |
| df[‘species’].nunique() #Count of Unique names in Species Column |
| df[‘species’].unique() #Name of Unique names in Species Column |
| df.sort\_values(by=[‘Sepal’,’Petal’]) #Sort the Data by Sepal colmn and if 2 rows have the same sepal value then sort on the basis of Petal value |
| df[‘species’].apply(func) #Apply a function to the species Column of the DataSet |
| df.groupby(‘species’) #Group the data according to the species column |
|  |
|  |