Pandas Library

- Pandas is a Python library.Pandas is used to analyze data.
- Pandas is a powerful Python library used for data manipulation and analysis. It provides data structures like DataFrames and Series, which are essential for handling and analyzing structured data. Here are some common data manipulation tasks you can perform using Pandas
- Methods/pands used here -
- (**version**) ==> To check the version of pandas that we imported
- read_csv("file_path") ==> to read the csv file into dataframes
- len => len of ddf
- id => id of the df in location
- dataframe.shape => it will give no.of rows,columns
- df.columns => it print all the coulmn names
- len(df.columns) => give the lenght of columns in df
- df[:] => Slicing similar to list and arrays
- df.head() => print 1st 5 rows of df
- df.tail() => print last 5rows in df
- df.isna() => check whethr any value is having NA in df
- df.describe() => It will give statastical data like mean,median,mode,cound,max and min on the numerical data in dataframe

In [1]:
```
import pandas as pd
```

In [2]:
```
pd.__version__ # Checking pandas version
```

Out[2]:
```
'2.2.2'
```

In [5]:
```
store = pd.read_csv(r"C:\Users\KonaSowjanya\Desktop\Python\data_sets\Sample Supe
```

In [21]:
```
store.head() # it will give top 5rows
```

Out[21]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Po C
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19

In [7]:
```
id(store)
```

Out[7]:
```
2178646840848
```

In [8]:
```
len(store)
```

Out[8]:
```
10194
```

In [9]:
```
store.shape # Gives rows and columsn count
```

Out[9]:
```
(10194, 19)
```

In [11]:
```
store.columns # Gives all the column names in data frame
```

Out[11]:
```
Index(['Category', 'City', 'Country/Region', 'Customer Name', 'Manufacturer',
       'Order Date', 'Order ID', 'Postal Code', 'Product Name', 'Region',
       'Segment', 'Ship Date', 'Ship Mode', 'State/Province', 'Sub-Category',
       'Discount', 'Profit', 'Quantity', 'Sales'],
      dtype='object')
```

In [12]:
```
len(store.columns)
```

Out[12]:
```
19
```

In [15]:
```
store.isnull() # To check if any null values are there. If yes it will give True
```

Out[15]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Posta Code
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	..
10189	False	False	False	False	False	False	False	False
10190	False	False	False	False	False	False	False	False
10191	False	False	False	False	False	False	False	False
10192	False	False	False	False	False	False	False	False
10193	False	False	False	False	False	False	False	False

10194 rows × 19 columns

In [16]:
```
store[:] # Slicing
```

Out[16]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Or
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	20 103
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20 112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20 112
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20 112
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20 141
...	...	...	...	...	...	...	
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20 143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20 115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20 156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20 143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20 143

10194 rows × 19 columns

In [19]:
```
store[1:10] # Row 1 to 9th row
```

Out[19]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Po C
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19
5	Furniture	Henderson	United States	Maria Etezadi	Global	06-01-2020	US-2020-167199	42
6	Office Supplies	Henderson	United States	Maria Etezadi	Rogers	06-01-2020	US-2020-167199	42
7	Office Supplies	Athens	United States	Jack O'Briant	Dixon	06-01-2020	US-2020-106054	30
8	Office Supplies	Henderson	United States	Maria Etezadi	Ibico	06-01-2020	US-2020-167199	42
9	Office Supplies	Henderson	United States	Maria Etezadi	Alliance	06-01-2020	US-2020-167199	42

In [20]:
```
store[0:10:5] # row 0 to 9 with step 5
```

Out[20]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Pos Co
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	770
5	Furniture	Henderson	United States	Maria Etezadi	Global	06-01-2020	US-2020-167199	424

In [23]:
```
store.tail() # last 5 rows
```

Out[23]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Or
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20 143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20 115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20 156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20 143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20 143

In [24]:
```
store.isna()
```

Out[24]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Posta Code
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	..
10189	False	False	False	False	False	False	False	False
10190	False	False	False	False	False	False	False	False
10191	False	False	False	False	False	False	False	False
10192	False	False	False	False	False	False	False	False
10193	False	False	False	False	False	False	False	False

10194 rows × 19 columns

Statastical Concepts

In [27]:
```
store.describe() # descriptive statistics like numerical data
```

Out[27]:

	Discount	Profit	Quantity	Sales
count	10194.000000	10194.000000	10194.000000	10194.000000
mean	0.155385	28.673417	3.791838	228.225854
std	0.206249	232.465115	2.228317	619.906839
min	0.000000	-6599.978000	1.000000	0.444000
25%	0.000000	1.760800	2.000000	17.220000
50%	0.200000	8.690000	3.000000	53.910000
75%	0.200000	29.297925	5.000000	209.500000
max	0.800000	8399.976000	14.000000	22638.480000

In [ ]:
```
```