

What is Pandas?

Pandas is a powerful open-source library in Python used for data analysis and manipulation. It provides flexible and efficient data structures, primarily Series and DataFrame, to handle structured data effectively.

- ◆ Key Features of Pandas

1. DataFrame & Series: Tabular (2D) and single-column (1D) data structures
2. Data Cleaning: Handle missing values, duplicate data, and transformations
3. Data Aggregation: Grouping, sorting, and summarization
4. Data Visualization: Works with Matplotlib & Seaborn for plotting
5. Integration: Works well with NumPy, SQL, and Excel

```
In [2]: import numpy as np  
import pandas as pd
```

```
In [5]: store = pd.read_csv(r'E:\FSDS With GEN AI_NIT\1st Feb - 28th Feb 2025\12th - Pandas
```

Out[5]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817
...							
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	US-2023-143259
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	US-2023-115427
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500

10194 rows × 19 columns

```
In [7]: id(store) # address of the memory location
```

```
Out[7]: 2035676436304
```

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

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

```
In [11]: store.shape #returns the no. of rows and columns (diemsionality)
```

```
Out[11]: (10194, 19)
```

```
In [14]: store.columns #returns features or columns
```

```
Out[14]: 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 [15]: len(store.columns)
```

```
Out[15]: 19
```

```
In [16]: store.dtypes #if data type is object it means it is categorical data
```

```
Out[16]: Category          object
          City              object
          Country/Region    object
          Customer Name    object
          Manufacturer     object
          Order Date        object
          Order ID          object
          Postal Code       object
          Product Name      object
          Region            object
          Segment           object
          Ship Date         object
          Ship Mode         object
          State/Province    object
          Sub-Category      object
          Discount          float64
          Profit            float64
          Quantity          int64
          Sales             float64
          dtype: object
```

```
In [18]: store.info() #Schema
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10194 entries, 0 to 10193
Data columns (total 19 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Category        10194 non-null   object  
 1   City             10194 non-null   object  
 2   Country/Region  10194 non-null   object  
 3   Customer Name   10194 non-null   object  
 4   Manufacturer    10194 non-null   object  
 5   Order Date      10194 non-null   object  
 6   Order ID         10194 non-null   object  
 7   Postal Code     10194 non-null   object  
 8   Product Name    10194 non-null   object  
 9   Region           10194 non-null   object  
 10  Segment          10194 non-null   object  
 11  Ship Date       10194 non-null   object  
 12  Ship Mode        10194 non-null   object  
 13  State/Province  10194 non-null   object  
 14  Sub-Category    10194 non-null   object  
 15  Discount         10194 non-null   float64 
 16  Profit            10194 non-null   float64 
 17  Quantity          10194 non-null   int64   
 18  Sales             10194 non-null   float64 
dtypes: float64(3), int64(1), object(15)
memory usage: 1.5+ MB
```

In [19]: `store.isnull()`

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

10194 rows × 19 columns



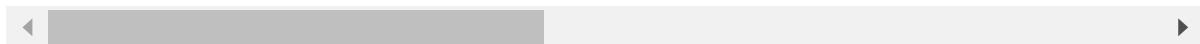
```
In [20]: store.isnull().sum()
```

```
Out[20]: Category      0  
City          0  
Country/Region 0  
Customer Name 0  
Manufacturer   0  
Order Date    0  
Order ID      0  
Postal Code   0  
Product Name  0  
Region        0  
Segment       0  
Ship Date    0  
Ship Mode    0  
State/Province 0  
Sub-Category  0  
Discount      0  
Profit        0  
Quantity      0  
Sales         0  
dtype: int64
```

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

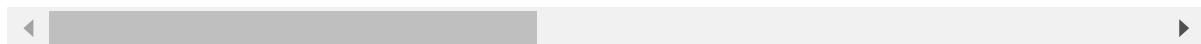
Out[21]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77015
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60540
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60540
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19143

In [22]: `store.tail() #it returns bottom 5 rows`

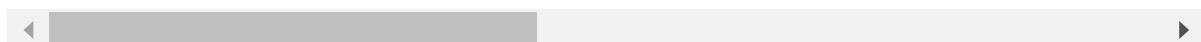
Out[22]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	US-2023-143259
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	US-2023-115427
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500

In [23]: `store.tail(3) #it returns bottom 5 rows`

Out[23]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500

In [24]: `store[:]`

Out[24]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817
...							
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	US-2023-143259
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	US-2023-115427
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500

10194 rows × 19 columns

```
In [25]: store[::-1]
```

Out[25]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	US-2023-115427
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	US-2023-143259
...
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800

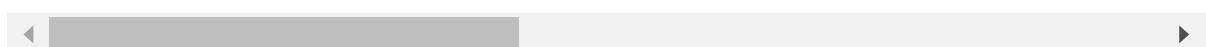
10194 rows × 19 columns

```
In [26]: store[::10]
```

Out[26]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	US-2020-167199
20	Furniture	Dover	United States	Seth Vernon	DAX	11-01-2020	US-2020-130092
30	Office Supplies	San Francisco	United States	Brian Dahlen	Tennsco	13-01-2020	US-2020-157147
40	Office Supplies	Scottsdale	United States	Toby Swindell	GBC	19-01-2020	US-2020-146591
...
10150	Office Supplies	Newark	United States	Dan Reichenbach	BIC	27-12-2023	US-2023-134404
10160	Office Supplies	New York City	United States	Jennifer Ferguson	Storex	28-12-2023	US-2023-164826
10170	Technology	New York City	United States	Jennifer Ferguson	Other	28-12-2023	US-2023-164826
10180	Office Supplies	New York City	United States	Michael Chen	Other	29-12-2023	US-2023-102638
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	US-2023-115427

1020 rows × 19 columns



```
In [27]: store[:10]
```

Out[27]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77015
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60540
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60540
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19143
5	Furniture	Henderson	United States	Maria Etezadi	Global	06-01-2020	US-2020-167199	42420
6	Office Supplies	Henderson	United States	Maria Etezadi	Rogers	06-01-2020	US-2020-167199	42420
7	Office Supplies	Athens	United States	Jack O'Briant	Dixon	06-01-2020	US-2020-106054	30605
8	Office Supplies	Henderson	United States	Maria Etezadi	Ibico	06-01-2020	US-2020-167199	42420

Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
9	Office Supplies	Henderson	United States	Maria Etezadi	Alliance	06-01-2020	US-2020-167199

In [28]: `store[10:]`

Out[28]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Ord
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	U 202 1671
11	Office Supplies	Los Angeles	United States	Lycoris Saunders	Xerox	06-01-2020	U 202 1308
12	Technology	Henderson	United States	Maria Etezadi	Other	06-01-2020	U 202 1671
13	Technology	Henderson	United States	Maria Etezadi	GE	06-01-2020	U 202 1671
14	Furniture	Huntsville	United States	Vivek Sundaresam	Howard Miller	07-01-2020	U 202 1054
...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	U 202 1432
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	U 202 1154
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	U 202 1567
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	U 202 1432
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	C 202 1435

10184 rows × 19 columns

```
In [30]: store.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10194 entries, 0 to 10193
Data columns (total 19 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Category        10194 non-null    object  
 1   City             10194 non-null    object  
 2   Country/Region  10194 non-null    object  
 3   Customer Name   10194 non-null    object  
 4   Manufacturer    10194 non-null    object  
 5   Order Date      10194 non-null    object  
 6   Order ID         10194 non-null    object  
 7   Postal Code     10194 non-null    object  
 8   Product Name    10194 non-null    object  
 9   Region           10194 non-null    object  
 10  Segment          10194 non-null    object  
 11  Ship Date       10194 non-null    object  
 12  Ship Mode       10194 non-null    object  
 13  State/Province  10194 non-null    object  
 14  Sub-Category    10194 non-null    object  
 15  Discount         10194 non-null    float64 
 16  Profit           10194 non-null    float64 
 17  Quantity         10194 non-null    int64   
 18  Sales            10194 non-null    float64 
dtypes: float64(3), int64(1), object(15)
memory usage: 1.5+ MB
```

```
In [34]: store.dtypes
```

```
Out[34]: Category        object  
          City           object  
          Country/Region  object  
          Customer Name   object  
          Manufacturer    object  
          Order Date      object  
          Order ID         object  
          Postal Code     object  
          Product Name    object  
          Region           object  
          Segment          object  
          Ship Date       object  
          Ship Mode       object  
          State/Province  object  
          Sub-Category    object  
          Discount         float64 
          Profit          float64 
          Quantity         int64  
          Sales            float64 
          dtype: object
```

```
In [35]: store.columns
```

```
Out[35]: 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 [36]: store['Category']
```

```
Out[36]: 0      Office Supplies
1      Office Supplies
2      Office Supplies
3      Office Supplies
4      Office Supplies
...
10189    Office Supplies
10190    Office Supplies
10191    Office Supplies
10192        Technology
10193    Office Supplies
Name: Category, Length: 10194, dtype: object
```

```
In [37]: store[['Category', 'City']]
```

	Category	City
0	Office Supplies	Houston
1	Office Supplies	Naperville
2	Office Supplies	Naperville
3	Office Supplies	Naperville
4	Office Supplies	Philadelphia
...
10189	Office Supplies	New York City
10190	Office Supplies	Fairfield
10191	Office Supplies	Loveland
10192	Technology	New York City
10193	Office Supplies	Charlottetown

10194 rows × 2 columns

```
In [41]: store_categorical_list = store[['Category', 'City', 'Country/Region', 'Customer Name',
   'Order Date', 'Order ID', 'Postal Code', 'Product Name', 'Region',
   'Segment', 'Ship Date', 'Ship Mode', 'State/Province', 'Sub-Category']]
store_categorical_list
```

Out[41]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817
...							
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	US-2023-143259
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	US-2023-115427
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500

10194 rows × 15 columns

In [43]: `store_categorical_list.dtypes`

```
Out[43]: Category      object
          City         object
          Country/Region  object
          Customer Name   object
          Manufacturer    object
          Order Date     object
          Order ID       object
          Postal Code    object
          Product Name   object
          Region         object
          Segment        object
          Ship Date      object
          Ship Mode      object
          State/Province  object
          Sub-Category   object
          dtype: object
```

In [42]: `store_numeric_list = store[['Discount', 'Profit', 'Quantity', 'Sales']]
store_numeric_list`

	Discount	Profit	Quantity	Sales
0	0.2	5.5512	2	16.448
1	0.8	-5.4870	2	3.540
2	0.2	4.2717	3	11.784
3	0.2	-64.7748	3	272.736
4	0.2	4.8840	3	19.536
...
10189	0.2	19.7910	3	52.776
10190	0.2	6.4750	2	20.720
10191	0.2	-0.6048	3	3.024
10192	0.0	2.7279	7	90.930
10193	0.2	-0.6048	3	3.024

10194 rows × 4 columns

In [44]: `store_numeric_list.dtypes`

```
Out[44]: Discount      float64
          Profit       float64
          Quantity     int64
          Sales        float64
          dtype: object
```

In [45]: `store.describe()`

Out[45]:

	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