

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

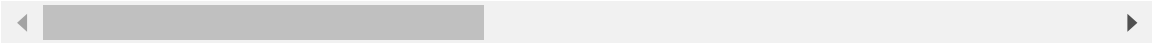
```
In [3]: store = pd.read_csv(r'C:\Users\rohit\OneDrive\Desktop\FSDS & GEN-AI 25TH-NOV-024
```

```
In [5]: store
```

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	20103
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
...	...	...	...	...	...	...	...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143

10194 rows × 19 columns



In [7]:

```
store.shape
```

```
Out[7]: (10194, 19)
```

```
In [9]: len(store)
```

```
Out[9]: 10194
```

```
In [11]: pd.__version__
```

```
Out[11]: '2.2.2'
```

```
In [13]: store.columns
```

```
Out[13]: 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 [17]: type(store)
```

```
Out[17]: pandas.core.frame.DataFrame
```

```
In [19]: id(store)
```

```
Out[19]: 2893920498320
```

```
In [21]: 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 [23]: store.dtypes
```

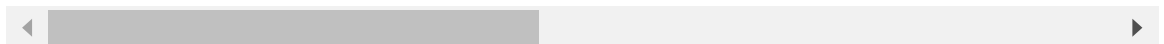
```
Out[23]: Category          object
City                    object
Country/Region          object
Customer Name           object
Manufacturer             object
Order Date              object
Order ID                object
Postal Code             object
Product Name            object
Region                  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 [25]: store.isnull()
```

```
Out[25]:
```

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal 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 × 9 columns



```
In [29]: store.isnull().sum()
```

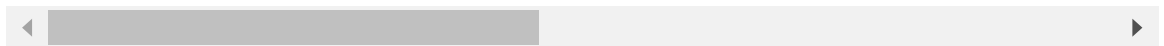
```
Out[29]: Category      0
City      0
Country/Region  0
Customer Name  0
Manufacturer  0
Order Date  0
Order ID  0
Postal Code  0
Product Name  0
Region  0
Segment  0
Ship Date  0
Ship Mode  0
State/Province  0
Sub-Category  0
Discount  0
Profit  0
Quantity  0
Sales  0
dtype: int64
```

```
In [31]: store.isna()
```

```
Out[31]:
```

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal 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 × 9 columns



```
In [33]: store.isna().sum()
```

```
Out[33]: 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 [35]: store.head()
```

```
Out[35]:
```

	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 [37]: store.tail()
```

Out[37]:

	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	20143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143

In [39]: store[5:11]

Out[39]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Po C
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
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	US-2020-167199	42

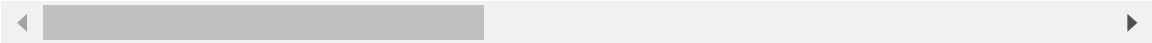
In [41]: store[:]



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	20103
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
...	...	...	...	...	...	...	...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143

10194 rows × 19 columns

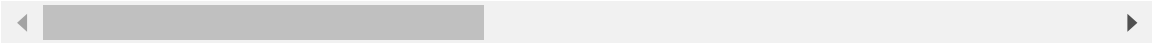


In [43]: store[:, -1]

Out[43]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
...	...	...	...	...	...	...	...
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	20103

10194 rows × 19 columns



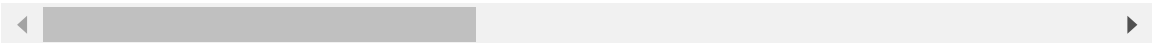
In [45]:

```
store[:, :10]
```

Out[45]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	1038
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	167
20	Furniture	Dover	United States	Seth Vernon	DAX	11-01-2020	1306
30	Office Supplies	San Francisco	United States	Brian Dahlen	Tennsco	13-01-2020	157
40	Office Supplies	Scottsdale	United States	Toby Swindell	GBC	19-01-2020	146
...	...	...	...	...	...	...	...
10150	Office Supplies	Newark	United States	Dan Reichenbach	BIC	27-12-2023	134
10160	Office Supplies	New York City	United States	Jennifer Ferguson	Storex	28-12-2023	1648
10170	Technology	New York City	United States	Jennifer Ferguson	Other	28-12-2023	1648
10180	Office Supplies	New York City	United States	Michael Chen	Other	29-12-2023	1026
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	115

1020 rows × 19 columns



```
In [47]: store.columns
```

```
Out[47]: 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 [51]: store['Category']
```

```
Out[51]: 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 [53]: store['Ship Mode']
```

```
Out[53]: 0      Standard Class  
         1      Standard Class  
         2      Standard Class  
         3      Standard Class  
         4      Standard Class  
         ...  
        10189 Standard Class  
        10190 Standard Class  
        10191 Standard Class  
        10192 Standard Class  
        10193 Standard Class  
        Name: Ship Mode, Length: 10194, dtype: object
```

```
In [55]: store[['Category', 'City']]
```

Out[55]:

	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 [57]: `store.columns`

```
Out[57]: 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 [61]: store_cat = store[['Category', 'City', 'Country/Region', 'Customer Name', 'Manuf  
              'Order Date', 'Order ID', 'Postal Code', 'Product Name', 'Region',  
              'Segment', 'Ship Date', 'Ship Mode', 'State/Province', 'Sub-Category']]  
  
store_num = store[['Discount', 'Profit', 'Quantity', 'Sales']]
```

In [63]: `store_cat`

Out[63]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	20103
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	20112
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	20112
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	20112
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	20141
...	...	...	...	...	...	...	...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	20143
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	20115
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	20156
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	20143

10194 rows × 15 columns



```
In [65]: print(len(store.columns))
print(len(store_cat.columns))
```

```
print(len(store_num.columns))
```

19

15

4

```
In [67]: store_cat.head(1)
```

```
Out[67]:
```

	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	77095



```
In [73]: store_cat.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10194 entries, 0 to 10193
Data columns (total 15 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
dtypes: object(15)
memory usage: 1.2+ MB
```

```
In [69]: store_num.head(1)
```

```
Out[69]:
```

	Discount	Profit	Quantity	Sales
0	0.2	5.5512	2	16.448

```
In [71]: store_num.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10194 entries, 0 to 10193
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Discount    10194 non-null  float64
1   Profit      10194 non-null  float64
2   Quantity    10194 non-null  int64
3   Sales       10194 non-null  float64
dtypes: float64(3), int64(1)
memory usage: 318.7 KB
```

In [75]: `store_cat.describe()`

Out[75]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
<b>count</b>	10194	10194	10194	10194	10194	10194	10194	10194
<b>unique</b>	3	542	2	800	183	1242	5111	
<b>top</b>	Office Supplies	New York City	United States	William Brown	Other	05-09-2022	US-2023-100111	10
<b>freq</b>	6128	915	9994	41	1940	38	14	

In [77]: `store_num.describe()`

Out[77]:

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

In [83]: `store.columns`

Out[83]: 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 [79]: `store.describe()`



Out[79]:

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

In [ ]: