

Pandas

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

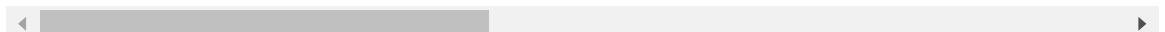
```
In [3]: store = pd.read_csv(r"D:\Sample - Superstore_Orders.csv")
```

```
In [4]: store
```

Out[4]:

| | 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 [5]: id(store)

```
Out[5]: 2234544007184
```

```
In [6]: len(store)
```

```
Out[6]: 10194
```

```
In [7]: store.shape
```

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

```
In [8]: store.columns
```

```
Out[8]: 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 [9]: len(store.columns)
```

```
Out[9]: 19
```

```
In [10]: store.dtypes
```

```
Out[10]: 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 [11]: store.info()
```

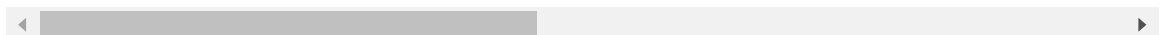
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10194 entries, 0 to 10193
Data columns (total 19 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Category              10194 non-null  object
1   City                  10194 non-null  object
2   Country/Region        10194 non-null  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 [12]: `store.isnull()`

Out[12]:

| | 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 × 19 columns



In [13]: `store.isnull().sum()`

```
Out[13]: 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 [14]: store.head()
```

```
Out[14]:
```

| | 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 [15]: store.tail()
```

Out[15]:

| | Category | City | Country/Region | Customer Name | Manufacturer | Order Date | Order |
|-------|-----------------|---------------|----------------|-------------------|--------------|------------|-------|
| 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 |

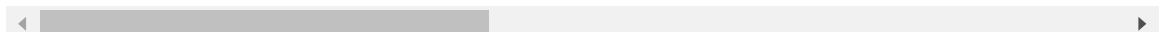
In [16]:

```
store[:]
```

Out[16]:

| | 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

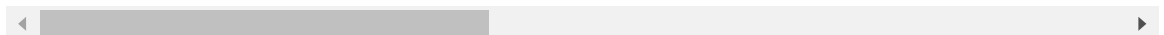


In [18]: store[:: -1]

Out[18]:

| | 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



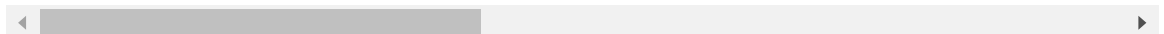
In [19]:

store[:,10]

Out[19]:

| | 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 | 164 |
| 10170 | Technology | New York City | United States | Jennifer Ferguson | Other | 28-12-2023 | 164 |
| 10180 | Office Supplies | New York City | United States | Michael Chen | Other | 29-12-2023 | 102 |
| 10190 | Office Supplies | Fairfield | United States | Erica Bern | GBC | 30-12-2023 | 115 |

1020 rows × 19 columns



In [20]: store[:10]

Out[20]:

| | Category | City | Country/Region | Customer Name | Manufacturer | Order Date | Order ID | Pos |
|---|-----------------|--------------|----------------|---------------|--------------|------------|----------------|-----|
| 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 |
| 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 |

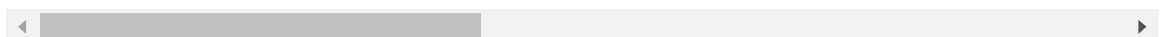
| | Category | City | Country/Region | Customer Name | Manufacturer | Order Date | Order ID | Po C |
|---|-----------------|-----------|----------------|---------------|--------------|------------|----------------|------|
| 9 | Office Supplies | Henderson | United States | Maria Etezadi | Alliance | 06-01-2020 | US-2020-167199 | 42 |

```
In [21]: store[10:]
```

Out[21]:

| | Category | City | Country/Region | Customer Name | Manufacturer | Order Date | |
|-------|-----------------|---------------|----------------|-------------------|---------------|------------|-----|
| 10 | Office Supplies | Henderson | United States | Maria Etezadi | Southworth | 06-01-2020 | 1 |
| 11 | Office Supplies | Los Angeles | United States | Lycoris Saunders | Xerox | 06-01-2020 | 1 |
| 12 | Technology | Henderson | United States | Maria Etezadi | Other | 06-01-2020 | 1 |
| 13 | Technology | Henderson | United States | Maria Etezadi | GE | 06-01-2020 | 1 |
| 14 | Furniture | Huntsville | United States | Vivek Sundaresam | Howard Miller | 07-01-2020 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 10189 | Office Supplies | New York City | United States | Patrick O'Donnell | Wilson Jones | 30-12-2023 | 1 |
| 10190 | Office Supplies | Fairfield | United States | Erica Bern | GBC | 30-12-2023 | 1 |
| 10191 | Office Supplies | Loveland | United States | Jill Matthias | Other | 30-12-2023 | 1 |
| 10192 | Technology | New York City | United States | Patrick O'Donnell | Other | 30-12-2023 | 1 |
| 10193 | Office Supplies | Charlottetown | Canada | Harry Olson | Wilson Jones | 30-12-2023 | 1 |

10184 rows × 19 columns



```
In [25]: store.dtypes
```

```
Out[25]: 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 [26]: store.columns
```

```
Out[26]: 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 [28]: store['Category']
```

```
Out[28]: 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 [30]: store[['Category', 'City']]
```

Out[30]:

| | 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 [33]: store_categorical_data = store [['Category', 'City', 'Country/Region', 'Customer
      'Order Date', 'Order ID', 'Postal Code', 'Product Name', 'Region',
      'Segment', 'Ship Date', 'Ship Mode', 'State/Province', 'Sub-Category',]]
```

```
In [34]: store_categorical_data.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
dtype: object
```

```
In [37]: store_numerical_data = store[['Discount', 'Profit', 'Quantity', 'Sales']]
```

```
In [39]: store_numerical_data.dtypes
```

```
Out[39]: Discount      float64
Profit                float64
Quantity              int64
Sales                 float64
dtype: object
```

In []:

