

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

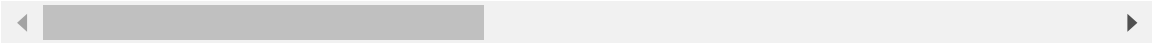
```
In [3]: store = pd.read_csv(r'C:\Users\Sample - Superstore_Orders.csv')
```

```
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]: `store.columns`

Out[9]: 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 [11]: `len(store)`

Out[11]: 10194

In [15]: `len(store.columns)`

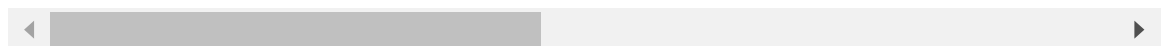
Out[15]: 19

In [17]: `store.isnull()` *## detect missing value*

Out[17]:

	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 [19]: `store[5:10]`

Out[19]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Pos Co
5	Furniture	Henderson	United States	Maria Etezadi	Global	06-01-2020	US-2020-167199	424
6	Office Supplies	Henderson	United States	Maria Etezadi	Rogers	06-01-2020	US-2020-167199	424
7	Office Supplies	Athens	United States	Jack O'Briant	Dixon	06-01-2020	US-2020-106054	306
8	Office Supplies	Henderson	United States	Maria Etezadi	Ibico	06-01-2020	US-2020-167199	424
9	Office Supplies	Henderson	United States	Maria Etezadi	Alliance	06-01-2020	US-2020-167199	424

In [21]: `store.isnull().sum()`

Out[21]:

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 [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]: id(store)
```

```
Out[25]: 2793518934672
```

```
In [27]: 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 [29]: store.columns
```

```
Out[29]: 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 [31]: store['City']
```

```
Out[31]: 0          Houston
         1          Naperville
         2          Naperville
         3          Naperville
         4          Philadelphia
         ...
        10189    New York City
        10190          Fairfield
        10191          Loveland
        10192    New York City
        10193    Charlottetown
        Name: City, Length: 10194, dtype: object
```

```
In [33]: store[['City', 'Category']]
```

```
Out[33]:
```

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

10194 rows × 2 columns

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

Out[37]:

	City	Category	Order ID
0	Houston	Office Supplies	US-2020-103800
1	Naperville	Office Supplies	US-2020-112326
2	Naperville	Office Supplies	US-2020-112326
3	Naperville	Office Supplies	US-2020-112326
4	Philadelphia	Office Supplies	US-2020-141817
...	...	...	...
10189	New York City	Office Supplies	US-2023-143259
10190	Fairfield	Office Supplies	US-2023-115427
10191	Loveland	Office Supplies	US-2023-156720
10192	New York City	Technology	US-2023-143259
10193	Charlottetown	Office Supplies	CA-2023-143500

10194 rows × 3 columns

```
In [39]: store_text = [['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']]
```

```
In [41]: store_text
```

```
Out[41]: [['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']]
```

```
In [43]: store_text = store[['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']]
```

```
In [45]: store_text
```

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	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 [47]: store\_text.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 [51]: store_number = store[['Discount', 'Profit', 'Quantity', 'Sales']]
store_number
```

```
Out[51]:
```

	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 [53]: store.columns
```

```
Out[53]: 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 [57]: len(store.columns)
```

Out[57]: 19

```
In [59]: store_text.columns
```

```
Out[59]: 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'],  
              dtype='object')
```

```
In [61]: len(store_text.columns)
```

Out[61]: 15

```
In [63]: store_number.columns
```

```
Out[63]: Index(['Discount', 'Profit', 'Quantity', 'Sales'], dtype='object')
```

```
In [65]: len(store_number.columns)
```

Out[65]: 4

```
In [67]: print(len(store.columns))  
         print(len(store_text.columns))  
         print(len(store_number.columns))
```

19

15

4

```
In [69]: store.head()
```

Out[69]:

	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 [71]:

```
store.tail()
```

Out[71]:

	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

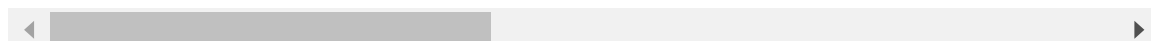
In [73]:

```
store[:]
```

Out[73]:

	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

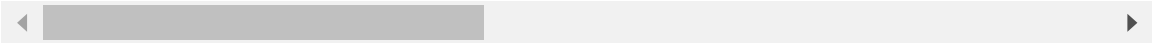


In [75]: store[:, :-1]

Out[75]:

	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 [77]: store[0:100:10] ## 10 step count

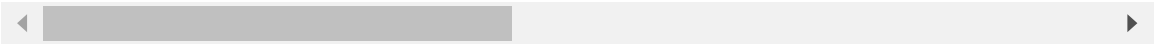
Out[77]:

	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
50	Office Supplies	Jonesboro	United States	Hunter Lopez	Newell	20-01-2020	US-2020-147627
60	Furniture	Miami	United States	Tom Boeckenhauer	Ultra	21-01-2020	US-2020-110422
70	Furniture	Alexandria	United States	Shirley Daniels	GE	26-01-2020	US-2020-155502
80	Office Supplies	New York City	United States	Jasper Cacioppo	Pressboard	28-01-2020	US-2020-100328
90	Office Supplies	Escondido	United States	Mark Haberlin	GBC	04-02-2020	US-2020-104808

In [79]: store[10:11]

Out[79]:

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



In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: