

# Pandas

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

```
In [2]: store = pd.read_csv(r'E:\Sample - Superstore_Orders.csv')
```

```
In [3]: store
```

Out[3]:

	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 [4]: id(store)
```

```
Out[4]: 1940072788192
```

```
In [5]: len(store)
```

```
Out[5]: 10194
```

```
In [6]: store.shape
```

```
Out[6]: (10194, 19)
```

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

```
Out[7]: 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 [8]: len(store.columns)
```

```
Out[8]: 19
```

```
In [9]: store.dtypes
```

```
Out[9]: 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 [10]: 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 [11]: store.isnull()
```

Out[11]:

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

10194 rows × 19 columns



```
In [12]: store.isnull().sum()
```

```
Out[12]: 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 [13]: store.head()
```

Out[13]:

	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 [14]: store.head(2)

Out[14]:

	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

In [15]: store.tail()

Out[15]:

	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 [16]: store.tail(3)

Out[16]:

	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

# Slicing in Pandas

In [17]: store[:]

Out[17]:

	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 [18]: `store[:-1]`

Out[18]:

	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
...	...	...	...	...	...	...	...
10188	Office Supplies	Fairfield	United States	Erica Bern	Cardinal	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
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 rows × 19 columns

In [19]: `store[:,11]`

Out[19]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Pe (C
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	7
11	Office Supplies	Los Angeles	United States	Lycoris Saunders	Xerox	06-01-2020	US-2020-130813	9
22	Furniture	Mount Pleasant	United States	Natalie DeCherney	Global	13-01-2020	US-2020-109232	2
33	Technology	Roswell	United States	Erica Hackney	Logitech	15-01-2020	US-2020-103366	3
44	Furniture	Westland	United States	Xylona Preis	Ultra	20-01-2020	US-2020-167927	4
...	...	...	...	...	...	...	...	...
10142	Technology	Fairfield	United States	Susan Vittorini	Maxell	25-12-2023	US-2023-147669	4
10153	Furniture	Long Beach	United States	Jason Gross	Novimex	28-12-2023	US-2023-101322	9
10164	Office Supplies	New York City	United States	Jennifer Ferguson	Other	28-12-2023	US-2023-164826	1
10175	Furniture	Anaheim	United States	Ben Peterman	Nu-Dell	29-12-2023	US-2023-146626	9
10186	Furniture	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259	1

927 rows × 19 columns



In [20]: `store[:10]`

Out[20]:

	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	42420

In [21]: store[10:]

Out[21]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	U20216719
11	Office Supplies	Los Angeles	United States	Lycoris Saunders	Xerox	06-01-2020	U2021308
12	Technology	Henderson	United States	Maria Etezadi	Other	06-01-2020	U20216719
13	Technology	Henderson	United States	Maria Etezadi	GE	06-01-2020	U20216719
14	Furniture	Huntsville	United States	Vivek Sundaresam	Howard Miller	07-01-2020	U2021054
...	...	...	...	...	...	...	...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	U20214329
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	U2021154
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	U2021567
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	U20214329
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	C2021435



10184 rows × 19 columns

```
In [22]: store.dtypes
```

```
Out[22]: 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 [23]: store.columns
```

```
Out[23]: 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 [24]: store['Category']
```

```
Out[24]: 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 [25]: store[['Category', 'Customer Name']]
```

Out[25]:

	Category	Customer Name
0	Office Supplies	Darren Powers
1	Office Supplies	Phillina Ober
2	Office Supplies	Phillina Ober
3	Office Supplies	Phillina Ober
4	Office Supplies	Mick Brown
...	...	...
10189	Office Supplies	Patrick O'Donnell
10190	Office Supplies	Erica Bern
10191	Office Supplies	Jill Matthias
10192	Technology	Patrick O'Donnell
10193	Office Supplies	Harry Olson

10194 rows × 2 columns

```
In [26]: store[['Category', 'City']]
```

Out[26]:

	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 [27]: store_categorical_data = 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_data
```

Out[27]:

	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 [28]: store_categorical_data.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 [29]: store_numerical_data = store[['Discount', 'Profit', 'Quantity', 'Sales']]
store_numerical_data
```

```
Out[29]:
```

	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 [30]: store_numerical_data.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
```