



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

```
In [2]: pd.__version__
```

```
Out[2]: '2.2.2'
```

```
In [3]: store = pd.read_csv(r"C:\Users\Admin\Downloads\Data Science\Daily work\19th, 2
```

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

Out[4]:

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

10194 rows × 19 columns

In [5]: `id(store)`

Out[5]: 1950809253040

```
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 [11]: store.isnull()
```

Out[11]:

	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 [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[:,]
```

Out[13]:

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

10194 rows × 19 columns

In [14]: store[0:10]

Out[14]:

	Category	City	Country/ Region	Customer Name	Manufacturer	Order Date	C
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020
5	Furniture	Henderson	United States	Maria Etezadi	Global	06-01-2020	US-2020
6	Office Supplies	Henderson	United States	Maria Etezadi	Rogers	06-01-2020	US-2020
7	Office Supplies	Athens	United States	Jack O'Briant	Dixon	06-01-2020	US-2020
8	Office Supplies	Henderson	United States	Maria Etezadi	Ibico	06-01-2020	US-2020
9	Office Supplies	Henderson	United States	Maria Etezadi	Alliance	06-01-2020	US-2020

Category	City	Country/ Region	Customer Name	Manufacturer	Order Date	C

In [15]: store[0:20:5]

Out[15]:

	Category	City	Country/ Region	Customer Name	Manufacturer	Order Date	C
<b>0</b>	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-202
<b>5</b>	Furniture	Henderson	United States	Maria Etezadi	Global	06-01-2020	US-202
<b>10</b>	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	US-202
<b>15</b>	Office Supplies	Huntsville	United States	Vivek Sundaresam	Acco	07-01-2020	US-202

In [16]: store.head()

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	US-2020-1
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-1
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-1
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-1
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-1

In [17]: `store.head(3)`

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-2020-1
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-1
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-1

In [18]: `store.tail()`



Out[18]:

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

In [19]: store.isna()

Out[19]:

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

10194 rows × 19 columns

# Introduce statistical concept in pandas

## ==>

```
In [20]: store.describe()
```

	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

```
In [21]: store.head()
```

	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-2020-0000000001
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-0000000002
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-0000000003
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-0000000004
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-0000000005

```
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 [26]: store[['Category', 'City', 'Customer Name']]
```

```
Out[26]:
```

	Category	City	Customer Name
0	Office Supplies	Houston	Darren Powers
1	Office Supplies	Naperville	Phillina Ober
2	Office Supplies	Naperville	Phillina Ober
3	Office Supplies	Naperville	Phillina Ober
4	Office Supplies	Philadelphia	Mick Brown
...	...	...	...
10189	Office Supplies	New York City	Patrick O'Donnell
10190	Office Supplies	Fairfield	Erica Bern
10191	Office Supplies	Loveland	Jill Matthias
10192	Technology	New York City	Patrick O'Donnell
10193	Office Supplies	Charlottetown	Harry Olson

10194 rows x 3 columns

# Can you divide the dataset into categorical & numerical

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

```
Out[27]: 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_cat = store[['Category', 'City', 'Country/Region', 'Customer Name', 'Mar  
              'Order Date', 'Order ID', 'Postal Code', 'Product Name', 'Region',  
              'Segment', 'Ship Date', 'Ship Mode', 'State/Province', 'Sub-Category']]
```

```
In [32]: store_cat
```

Out[32]:

	Category	City	Country/ Region	Customer Name	Manufacturer	Order Date
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020
...	...	...	...	...	...	...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023
10194 rows × 15 columns						

In [33]: `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 [34]: 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 [38]: len(store.columns)  #1.5 MB
```

Out[38]: 19

```
In [39]: len(store_cat.columns) # 1.2 MB
```

Out[39]: 15

```
In [40]: store_num = store[['Discount', 'Profit', 'Quantity', 'Sales']]
```

```
In [42]: 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 [46]: print(len(store.columns))
print(len(store_cat.columns))
print(len(store_num.columns))
```

19  
15  
4

```
In [48]: store_cat.describe()
```

Out[48]:

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

```
In [49]: store_num.describe()
```

Out[49]:

	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 [54]: rev = store[::-1] # Slicing
rev
```



Out[54]:

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

10194 rows × 19 columns

In [55]: `rev.head()`

Out[55]:

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

In [57]: `rev.tail()`

Out[57]:

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

In [ ]: `# Pandas Introduction Completed`

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: