

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

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

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

```
In [32]: df = pd.read_csv(r'C:\Users\arsha_4tjdyqj\Downloads\Sample - Superstore_Orders.csv')
```

```
In [4]: df
```

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	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 [9]: `df.shape`

Out[9]: (10194, 19)

In [11]: `len(df)`

Out[11]: 10194

In [13]: `df.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]: `df.isnull()`

Out[15]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	F
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 [17]: `df.isnull().sum()`

```
Out[17]: 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 [19]: df.dtypes
```

```
Out[19]: 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 [21]: df['City']
```

```
Out[21]: 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 [23]: df[:]
```

Out[23]:

	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	Phyllina Ober	GBC	04-01-2020	US-2020-112326
2	Office Supplies	Naperville	United States	Phyllina Ober	Avery	04-01-2020	US-2020-112326
3	Office Supplies	Naperville	United States	Phyllina 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 [25]: df[1:10:2]
```

Out[25]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60540
5	Furniture	Henderson	United States	Maria Etezadi	Global	06-01-2020	US-2020-167199	42420
7	Office Supplies	Athens	United States	Jack O'Briant	Dixon	06-01-2020	US-2020-106054	30605
9	Office Supplies	Henderson	United States	Maria Etezadi	Alliance	06-01-2020	US-2020-167199	42420

```
In [27]: df[:, :-2]
```

Out[27]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	US-2023-143259
10187	Office Supplies	Columbus	United States	Chuck Clark	Eureka	30-12-2023	US-2023-126221
10185	Office Supplies	Quebec City	Canada	Bruce Galang	Acco	29-12-2023	CA-2023-141531
...
9	Office Supplies	Henderson	United States	Maria Etezadi	Alliance	06-01-2020	US-2020-167199
7	Office Supplies	Athens	United States	Jack O'Briant	Dixon	06-01-2020	US-2020-106054
5	Furniture	Henderson	United States	Maria Etezadi	Global	06-01-2020	US-2020-167199
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326

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

5097 rows × 19 columns

```
In [36]: df.columns
```

```
Out[36]: 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 [38]: df['Category']
```

```
Out[38]: 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 [42]: df[['Customer Name', 'City']]
```

Out[42]:

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

10194 rows × 2 columns

```
In [44]: df_catagorical=df[['Category', 'City', 'Country/Region', 'Customer Name', 'Manufacture',
                             'Order Date', 'Order ID', 'Postal Code', 'Product Name', 'Region',
                             'Segment', 'Ship Date', 'Ship Mode', 'State/Province', 'Sub-Category']]
df_catagorical
```

Out[44]:

	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	Phyllina Ober	GBC	04-01-2020	US-2020-112326
2	Office Supplies	Naperville	United States	Phyllina Ober	Avery	04-01-2020	US-2020-112326
3	Office Supplies	Naperville	United States	Phyllina 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 [46]: `df_catagorical.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 [50]: `df_numerical=df[['Discount', 'Profit', 'Quantity', 'Sales']]`
`df_numerical`

Out[50]:

	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 [52]: `df_numerical.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 []:

In []:

In []: