

working on Pandas

```
In [3]: import pandas as pd
```

```
In [5]: Orders = pd.read_csv(r"/Users/babarhussain/Documents/Store_Orders.csv")
```

```
In [7]: Orders
```

Out[7]:	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 [9]: Orders.shape
```

```
Out[9]: (10194, 19)
```

```
In [11]: len(Orders)
```

```
Out[11]: 10194
```

```
In [15]: pd.__version__
```

```
Out[15]: '2.2.2'
```

```
In [17]: Orders.columns
```

```
Out[17]: 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 [21]: len(Orders.columns)
```

```
Out[21]: 19
```

```
In [23]: type(Orders)
```

```
Out[23]: pandas.core.frame.DataFrame
```

```
In [25]: id(Orders)
```

```
Out[25]: 5307218352
```

```
In [31]: Orders.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 [33]: Orders.dtypes
```

```

Out[33]: 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 [37]: Orders.isnull()
```

Out[37]:

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

10194 rows × 19 columns

In [41]: `Orders.isnull().sum()`

```
Out[41]: 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 [43]: `Orders.isna()`

Out[43]:

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

10194 rows × 19 columns

In [45]: `Orders.isna().sum()`

```
Out[45]: 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 [47]: `Orders.head()`

Out[47]:

	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

In [49]: `Orders.head(7)`

Out[49]:

	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
5	Furniture	Henderson	United States	Maria Etezadi	Global	06-01-2020	US-2020-167199
6	Office Supplies	Henderson	United States	Maria Etezadi	Rogers	06-01-2020	US-2020-167199

In [51]: `Orders.tail()`

Out [51]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date
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

In [57]: `Orders.tail(7)`

Out[57]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date
10187	Office Supplies	Columbus	United States	Chuck Clark	Eureka	30-12-2023
10188	Office Supplies	Fairfield	United States	Erica Bern	Cardinal	30-12-2023
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

In [59]: Orders[10:21]

Out[59]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Or
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	167
11	Office Supplies	Los Angeles	United States	Lycoris Saunders	Xerox	06-01-20	

							2020	130	
12	Technology	Henderson	United States	Maria Etezadi	Other		06-01-2020	20167	
13	Technology	Henderson	United States	Maria Etezadi	GE		06-01-2020	20167	
14	Furniture	Huntsville	United States	Vivek Sundaresam	Howard Miller		07-01-2020	20105	
15	Office Supplies	Huntsville	United States	Vivek Sundaresam	Acco		07-01-2020	20105	
16	Office Supplies	Laredo	United States	Melanie Seite	Newell		09-01-2020	20135	
17	Technology	Laredo	United States	Melanie Seite	Memorex		09-01-2020	20135	
18	Furniture	Springfield	United States	Anthony Jacobs	Howard Miller		10-01-2020	20149	
19	Office Supplies	Springfield	United States	Anthony Jacobs	Avery		10-01-2020	20149	
20	Furniture	Dover	United States	Seth Vernon	DAX		11-01-2020	20130	

In [61]: `Orders[:::-1]`

Out[61]:

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

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

10194 rows × 19 columns

In [65]: Orders[::10]

Out[65]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020
20	Furniture	Dover	United States	Seth Vernon	DAX	11-01-2020
30	Office Supplies	San Francisco	United States	Brian Dahlen	Tennsco	13-01-2020
40	Office Supplies	Scottsdale	United States	Toby Swindell	GBC	19-01-2020
...
10150	Office Supplies	Newark	United States	Dan Reichenbach	BIC	27-12-2023
10160	Office Supplies	New York City	United States	Jennifer Ferguson	Storex	28-12-2023
10170	Technology	New York City	United States	Jennifer Ferguson	Other	28-12-2023
10180	Office Supplies	New York City	United States	Michael Chen	Other	29-12-2023
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023

1020 rows × 19 columns

```
In [67]: Orders.columns
```

```
Out[67]: 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 [75]: Orders[['City', 'Country/Region']]
```

```
Out[75]:
```

	City	Country/Region
--	------	----------------

0	Houston	United States
1	Naperville	United States
2	Naperville	United States
3	Naperville	United States
4	Philadelphia	United States
...
10189	New York City	United States
10190	Fairfield	United States
10191	Loveland	United States
10192	New York City	United States
10193	Charlottetown	Canada

10194 rows × 2 columns

```
In [91]: Orders_Cate = Orders[['Category', 'City', 'Country/Region', 'Customer Name',  
                             'Order Date', 'Order ID', 'Postal Code', 'Product Name', 'Region',  
                             'Segment', 'Ship Date', 'Ship Mode', 'State/Province', 'Sub-Category',  
                             'Discount', 'Profit', 'Quantity', 'Sales']]  
Orders_num = Orders[['Discount', 'Profit', 'Quantity', 'Sales']]  
Orders_Cate
```

```
Out[91]:
```

	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 x 15 columns

```
In [93]: print(Orders_num)
```

	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 x 4 columns]

In []:

```
In [89]: print(len(Orders_Cate.columns))
print(len(Orders_num.columns))
```

15
4

```
In [97]: Orders_Cate.head(1)
```

```
Out[97]:
```

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	P
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	7

```
In [99]: Orders_num.head(1)
```

```
Out[99]:
```

	Discount	Profit	Quantity	Sales
0	0.2	5.5512	2	16.448

```
In [101... Orders_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 [103... Orders_Cate.describe()

Out[103...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
count	10194	10194	10194	10194	10194	10194	10194
unique	3	542	2	800	183	1242	5111
top	Office Supplies	New York City	United States	William Brown	Other	05-09-2022	US-2023-100111
freq	6128	915	9994	41	1940	38	14

In [105... Orders.describe()

Out[105...

	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 []: