

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

```
In [3]: store = pd.read_csv(r"C:\Users\AKSHAY\OneDrive\Desktop\Code\Projects\Sample - Su
```

```
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 [11]: # HOW TO CHECK THE NUMBER OF COLUMNS AND ROWS IN THE DATASET

```
In [13]: store.shape # this gives number of rows and columns - dimesions of the dataset
```

```
Out[13]: (10194, 19)
```

```
In [15]: # CHECKING THE LENGTH OF THE DATASET
```

```
In [17]: len(store) # this returns the number of rows - Length of the rows
```

```
Out[17]: 10194
```

```
In [19]: # HOW TO CHECK THE VERSION OF PANDAS
```

```
In [21]: pd.__version__
```

```
Out[21]: '2.2.2'
```

```
In [23]: # HOW TO CHECK THE COLUMN NAMES IN THE DATASET
```

```
In [25]: store.columns
```

```
Out[25]: 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 [27]: # CHECKING THE NUMBER OF COLUMNS IN THE DATASET
```

```
In [29]: len(store.columns)
```

```
Out[29]: 19
```

```
In [31]: # HOW TO CHECK THE TYPE OF THE DATASET
```

```
In [33]: type(store)
```

```
Out[33]: pandas.core.frame.DataFrame
```

```
In [35]: # HOW TO CHECK THE ADDRESS OF THE DATASET
```

```
In [37]: id(store)
```

```
Out[37]: 2153386158624
```

```
In [39]: # HOW TO CHECK WHAT DATATYPE, MISSING VALUES etc. ARE THERE IN THE DATASET
```

```
In [43]: store.info() # this returns information about the dataframe
```

```

<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 [45]: # HOW TO RETURN THE VALUES THAT ARE IN THE FORM OF DATA TYPES
```

```
In [47]: store.dtypes
```

```

Out[47]: 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

```

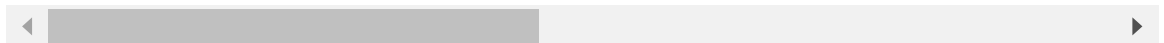
```
In [49]: # HOW TO CHECK IF THERE ARE ANY MISSING VALUES IN THE DATAFRAME
```

```
In [55]: store.isnull() # False means no missing values
```

Out[55]:

	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 × 9 columns

In [59]: `store.isnull().sum()` # this converts False into 0

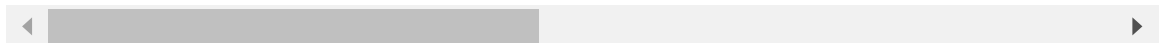
```
Out[59]: Category      0
City      0
Country/Region  0
Customer Name  0
Manufacturer  0
Order Date  0
Order ID  0
Postal Code  0
Product Name  0
Region  0
Segment  0
Ship Date  0
Ship Mode  0
State/Province  0
Sub-Category  0
Discount  0
Profit  0
Quantity  0
Sales  0
dtype: int64
```

In [61]: `# ANOTHER WAY TO CHECK IF THERE ARE ANY MISSING VALUES IN THE DATAFRAME`In [65]: `store.isna()` # False means no-null values - no missing values

Out[65]:

	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 × 9 columns

In [67]: `store.isna().sum()` *# this converts False into 0 or return the null values in int*

Out[67]:

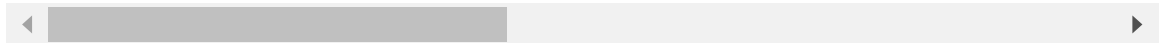
Category	0
City	0
Country/Region	0
Customer Name	0
Manufacturer	0
Order Date	0
Order ID	0
Postal Code	0
Product Name	0
Region	0
Segment	0
Ship Date	0
Ship Mode	0
State/Province	0
Sub-Category	0
Discount	0
Profit	0
Quantity	0
Sales	0

dtype: int64

In [69]: *# HOW TO GET FIRST 5 RECORDS IN THE DATASET*In [75]: `store.head()` *# as there is no parameter, by default it gives us 5 records*

Out[75]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Pos
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 [77]: `# HOW TO GET LAST 5 RECORDS IN THE DATASET`In [79]: `store.tail() # as there is no parameter, by default gives us bottom 5 records`

Out[79]:

	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	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 [81]: `store.tail(10) # returns bottom 10 records`

Out[81]:


	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
10184	Furniture	Quebec City	Canada	Bruce Galang	Other	29-12-2023	20141
10185	Office Supplies	Quebec City	Canada	Bruce Galang	Acco	29-12-2023	20141
10186	Furniture	New York City	United States	Patrick O'Donnell	Other	30-12-2023	20143
10187	Office Supplies	Columbus	United States	Chuck Clark	Eureka	30-12-2023	20126
10188	Office Supplies	Fairfield	United States	Erica Bern	Cardinal	30-12-2023	20115
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 [83]: # INDEXING AND SLICING THE DATASET
```

```
In [85]: store[5:11] # returns the values from 5 to 10 (n-1=11-1=10)
```

Out[85]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Po C
5	Furniture	Henderson	United States	Maria Etezadi	Global	06-01-2020	US-2020-167199	42
6	Office Supplies	Henderson	United States	Maria Etezadi	Rogers	06-01-2020	US-2020-167199	42
7	Office Supplies	Athens	United States	Jack O'Briant	Dixon	06-01-2020	US-2020-106054	30
8	Office Supplies	Henderson	United States	Maria Etezadi	Ibico	06-01-2020	US-2020-167199	42
9	Office Supplies	Henderson	United States	Maria Etezadi	Alliance	06-01-2020	US-2020-167199	42
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	US-2020-167199	42

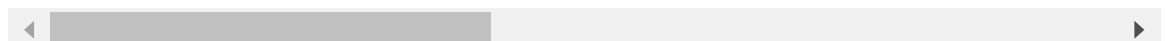


```
In [87]: store[:, -1]
```

Out[87]:

	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

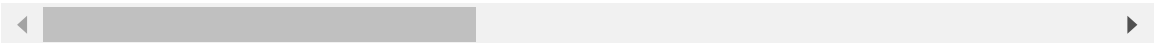
10194 rows × 19 columns

In [89]: `store[::-10] # reverses the dataset and gives values with every 10 step`

Out[89]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	1038
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	1677
20	Furniture	Dover	United States	Seth Vernon	DAX	11-01-2020	1306
30	Office Supplies	San Francisco	United States	Brian Dahlen	Tennsco	13-01-2020	1577
40	Office Supplies	Scottsdale	United States	Toby Swindell	GBC	19-01-2020	1465
...
10150	Office Supplies	Newark	United States	Dan Reichenbach	BIC	27-12-2023	1344
10160	Office Supplies	New York City	United States	Jennifer Ferguson	Storex	28-12-2023	1648
10170	Technology	New York City	United States	Jennifer Ferguson	Other	28-12-2023	1648
10180	Office Supplies	New York City	United States	Michael Chen	Other	29-12-2023	1026
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	1154

1020 rows × 19 columns



In [91]: store.columns

```
Out[91]: 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 [93]: # HOW TO PRINT THE COLUMN VALUES
```

```
In [97]: store['Category'] # returns the given parameter values
```

```
Out[97]: 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 [101... store[['Category', 'City']] # returns the given parameter values
```

```
Out[101...
      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 x 2 columns
```

```
In [103... store[['Category', 'Profit']] # returns the given parameter values
```

Out[103...

	Category	Profit
0	Office Supplies	5.5512
1	Office Supplies	-5.4870
2	Office Supplies	4.2717
3	Office Supplies	-64.7748
4	Office Supplies	4.8840
...
10189	Office Supplies	19.7910
10190	Office Supplies	6.4750
10191	Office Supplies	-0.6048
10192	Technology	2.7279
10193	Office Supplies	-0.6048

10194 rows × 2 columns

In [111...

HOW TO DIVIDE CATEGORICAL DATA AND NUMERICAL DATA

In [105...

```
store_cat = store[['Category', 'City', 'Country/Region', 'Customer Name', 'Manuf
    'Order Date', 'Order ID', 'Postal Code', 'Product Name', 'Region',
    'Segment', 'Ship Date', 'Ship Mode', 'State/Province', 'Sub-Category',
    ]]
store_num = store[['Discount', 'Profit', 'Quantity', 'Sales']]
```

In [109...

```
print(len(store.columns))
print(len(store_cat.columns))
print(len(store_num.columns))
```

19

15

4

In [113...

store_cat.head()

Out[113...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Pos
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 [115...

store_num.head()

Out[115...

	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

In [127...

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 [129... 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 [117... store_cat.describe()

Out[117...

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Po C
count	10194	10194	10194	10194	10194	10194	10194	10
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	10
freq	6128	915	9994	41	1940	38	14	

In [121... store_num.describe()

Out[121...

	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 [123...

```
store.describe()
```

Out[123...

	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 []: