

```
In [4]: # importing pandas  
# It is particularly well-suited for working with structured data, such as tables o  
import pandas as pd
```

```
In [8]: # Accessing a CSV file data from the local disk  
# read row string  
# (r'\filename.csv')  
store = pd.read_csv(r'E:\One_Drive(Microsoft)\OneDrive\Data_Science_cource\Module_1
```

```
In [10]: print(store)
```

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

	Manufacturer	Order Date	Order ID	Postal Code \
0	Message Book	03-01-2020	US-2020-103800	77015
1	GBC	04-01-2020	US-2020-112326	60540
2	Avery	04-01-2020	US-2020-112326	60540
3	SAFCO	04-01-2020	US-2020-112326	60540
4	Avery	05-01-2020	US-2020-141817	19143
...
10189	Wilson Jones	30-12-2023	US-2023-143259	10009
10190	GBC	30-12-2023	US-2023-115427	94533
10191	Other	30-12-2023	US-2023-156720	80538
10192	Other	30-12-2023	US-2023-143259	10009
10193	Wilson Jones	30-12-2023	CA-2023-143500	C0A

	Product Name	Region \
0	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central
1	GBC Standard Plastic Binding Systems Combs	Central
2	Avery 508	Central
3	SAFCO Boltless Steel Shelving	Central
4	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East
...
10189	Wilson Jones Legal Size Ring Binders	East
10190	GBC Binding covers	West
10191	Bagged Rubber Bands	West
10192	Gear Head AU3700S Headset	East
10193	Wilson Jones Impact Binders	East

	Segment	Ship Date	Ship Mode	State/Province \
0	Consumer	07-01-2020	Standard Class	Texas
1	Home Office	08-01-2020	Standard Class	Illinois
2	Home Office	08-01-2020	Standard Class	Illinois
3	Home Office	08-01-2020	Standard Class	Illinois
4	Consumer	12-01-2020	Standard Class	Pennsylvania
...
10189	Consumer	03-01-2024	Standard Class	New York
10190	Corporate	03-01-2024	Standard Class	California
10191	Consumer	03-01-2024	Standard Class	Colorado
10192	Consumer	03-01-2024	Standard Class	New York
10193	Consumer	03-01-2024	Standard Class	Prince Edward Island

	Sub-Category	Discount	Profit	Quantity	Sales
0	Paper	0.2	5.5512	2	16.448
1	Binders	0.8	-5.4870	2	3.540
2	Labels	0.2	4.2717	3	11.784

3	Storage	0.2	-64.7748	3	272.736
4	Art	0.2	4.8840	3	19.536
...
10189	Binders	0.2	19.7910	3	52.776
10190	Binders	0.2	6.4750	2	20.720
10191	Fasteners	0.2	-0.6048	3	3.024
10192	Phones	0.0	2.7279	7	90.930
10193	Binders	0.2	-0.6048	3	3.024

[10194 rows x 19 columns]

```
In [12]: store
```

Out[12]:

	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 [14]: # to know the shape of table use shape  
store.shape
```

```
Out[14]: (10194, 19)
```

```
In [16]: # columns function used to return the names of the fuctons  
store.columns
```

```
Out[16]: 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 [18]: # Lenght of dataset  
len(store)
```

```
Out[18]: 10194
```

```
In [22]: # total columns  
print("total columns in data set =",len(store.columns))
```

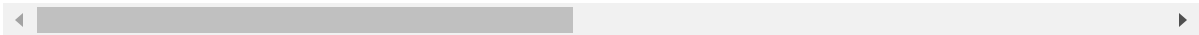
```
total columns in data set = 19
```

```
In [26]: # to know null values in the data set use isnull function  
store.isnull()  
# it return true when cell has no value => True else False
```

Out[26]:

	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 [34]: # We can count total missing values by using len() and sum() function
store.isnull().sum()
print(store.isnull().sum())
```

```
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 [48]: # dtypes
# this function return all columns data types
print("data types = \n",store.dtypes)
```

```
data types =  
  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 [52]: # info function  
         # info short info  
         # info() return details  
         print(store.info)
```

```
<bound method DataFrame.info of
Customer Name \
0      Office Supplies      Houston United States      Darren Powers
1      Office Supplies      Naperville United States      Phillina Ober
2      Office Supplies      Naperville United States      Phillina Ober
3      Office Supplies      Naperville United States      Phillina Ober
4      Office Supplies      Philadelphia United States      Mick Brown
...      ...      ...      ...      ...
10189 Office Supplies      New York City United States      Patrick O'Donnell
10190 Office Supplies      Fairfield United States      Erica Bern
10191 Office Supplies      Loveland United States      Jill Matthias
10192      Technology      New York City United States      Patrick O'Donnell
10193 Office Supplies      Charlottetown      Canada      Harry Olson
```

```
Manufacturer Order Date      Order ID Postal Code \
0      Message Book 03-01-2020 US-2020-103800      77015
1      GBC 04-01-2020 US-2020-112326      60540
2      Avery 04-01-2020 US-2020-112326      60540
3      SAFCO 04-01-2020 US-2020-112326      60540
4      Avery 05-01-2020 US-2020-141817      19143
...      ...      ...      ...      ...
10189 Wilson Jones 30-12-2023 US-2023-143259      10009
10190      GBC 30-12-2023 US-2023-115427      94533
10191      Other 30-12-2023 US-2023-156720      80538
10192      Other 30-12-2023 US-2023-143259      10009
10193 Wilson Jones 30-12-2023 CA-2023-143500      C0A
```

```
Product Name      Region \
0      Message Book, Wirebound, Four 5 1/2" X 4" Form... Central
1      GBC Standard Plastic Binding Systems Combs Central
2      Avery 508 Central
3      SAFCO Boltless Steel Shelving Central
4      Avery Hi-Liter EverBold Pen Style Fluorescent ... East
...      ...      ...
10189      Wilson Jones Legal Size Ring Binders East
10190      GBC Binding covers West
10191      Bagged Rubber Bands West
10192      Gear Head AU3700S Headset East
10193      Wilson Jones Impact Binders East
```

```
Segment Ship Date      Ship Mode      State/Province \
0      Consumer 07-01-2020 Standard Class      Texas
1      Home Office 08-01-2020 Standard Class      Illinois
2      Home Office 08-01-2020 Standard Class      Illinois
3      Home Office 08-01-2020 Standard Class      Illinois
4      Consumer 12-01-2020 Standard Class      Pennsylvania
...      ...      ...      ...
10189 Consumer 03-01-2024 Standard Class      New York
10190 Corporate 03-01-2024 Standard Class      California
10191 Consumer 03-01-2024 Standard Class      Colorado
10192 Consumer 03-01-2024 Standard Class      New York
10193 Consumer 03-01-2024 Standard Class      Prince Edward Island
```

```
Sub-Category Discount Profit Quantity Sales
0      Paper      0.2      5.5512      2      16.448
1      Binders      0.8      -5.4870      2      3.540
```



```

2      Labels      0.2   4.2717      3   11.784
3      Storage      0.2 -64.7748      3  272.736
4      Art          0.2   4.8840      3   19.536
...      ...      ...      ...      ...      ...
10189   Binders      0.2  19.7910      3   52.776
10190   Binders      0.2   6.4750      2   20.720
10191   Fasteners     0.2  -0.6048      3    3.024
10192   Phones        0.0   2.7279      7   90.930
10193   Binders      0.2  -0.6048      3    3.024

```

[10194 rows x 19 columns]>

In [54]: `print(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
None

```

In [58]: `# splitting`
`# text data == categorical data`
`# num data == numerical data`
`store.columns`

Out[58]: `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 [66]: `# id of the object can print by id() function`
`print("store id = ",id(store))`

store id = 2143214404528

```
In [70]: print(store[['City']])
print(store[['Customer Name', 'Order ID', 'Ship Date']])
```

```

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

```

```
[10194 rows x 1 columns]
```

	Customer Name	Order ID	Ship Date
0	Darren Powers	US-2020-103800	07-01-2020
1	Phillina Ober	US-2020-112326	08-01-2020
2	Phillina Ober	US-2020-112326	08-01-2020
3	Phillina Ober	US-2020-112326	08-01-2020
4	Mick Brown	US-2020-141817	12-01-2020
...
10189	Patrick O'Donnell	US-2023-143259	03-01-2024
10190	Erica Bern	US-2023-115427	03-01-2024
10191	Jill Matthias	US-2023-156720	03-01-2024
10192	Patrick O'Donnell	US-2023-143259	03-01-2024
10193	Harry Olson	CA-2023-143500	03-01-2024

```
[10194 rows x 3 columns]
```

```
In [60]: # text
store_text = store[['Category', 'City', 'Country/Region', 'Customer Name', 'Manufac
                'Order Date', 'Order ID', 'Postal Code', 'Product Name', 'Region',
                'Segment', 'Ship Date', 'Ship Mode', 'State/Province', 'Sub-Category']]
print(store_text)
```

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

	Manufacturer	Order Date	Order ID	Postal Code \
0	Message Book	03-01-2020	US-2020-103800	77015
1	GBC	04-01-2020	US-2020-112326	60540
2	Avery	04-01-2020	US-2020-112326	60540
3	SAFCO	04-01-2020	US-2020-112326	60540
4	Avery	05-01-2020	US-2020-141817	19143
...
10189	Wilson Jones	30-12-2023	US-2023-143259	10009
10190	GBC	30-12-2023	US-2023-115427	94533
10191	Other	30-12-2023	US-2023-156720	80538
10192	Other	30-12-2023	US-2023-143259	10009
10193	Wilson Jones	30-12-2023	CA-2023-143500	C0A

	Product Name	Region \
0	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central
1	GBC Standard Plastic Binding Systems Combs	Central
2	Avery 508	Central
3	SAFCO Boltless Steel Shelving	Central
4	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East
...
10189	Wilson Jones Legal Size Ring Binders	East
10190	GBC Binding covers	West
10191	Bagged Rubber Bands	West
10192	Gear Head AU3700S Headset	East
10193	Wilson Jones Impact Binders	East

	Segment	Ship Date	Ship Mode	State/Province \
0	Consumer	07-01-2020	Standard Class	Texas
1	Home Office	08-01-2020	Standard Class	Illinois
2	Home Office	08-01-2020	Standard Class	Illinois
3	Home Office	08-01-2020	Standard Class	Illinois
4	Consumer	12-01-2020	Standard Class	Pennsylvania
...
10189	Consumer	03-01-2024	Standard Class	New York
10190	Corporate	03-01-2024	Standard Class	California
10191	Consumer	03-01-2024	Standard Class	Colorado
10192	Consumer	03-01-2024	Standard Class	New York
10193	Consumer	03-01-2024	Standard Class	Prince Edward Island

	Sub-Category
0	Paper
1	Binders
2	Labels

```

3          Storage
4          Art
...
10189     Binders
10190     Binders
10191     Fasteners
10192     Phones
10193     Binders

```

[10194 rows x 15 columns]

In [74]: `store_text.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 [76]: `store_num = store[['Discount', 'Profit', 'Quantity', 'Sales']]`
`print(store_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 [78]: `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 [90]: # Lenghs of all store, store_text and store_num
print('length of store = ',len(store.columns))
print("length of store_text = ",len(store_text.columns))
print("length of store_num = ",len(store_num.columns))
len(store_num.columns)

```

```

length of store = 19
length of store_text = 15
length of store_num = 4

```

Out[90]: 4

```

In [92]: print(store.head())

```

	Category	City	Country/Region	Customer Name	Manufacturer	\
0	Office Supplies	Houston	United States	Darren Powers	Message Book	
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	

	Order Date	Order ID	Postal Code	\
0	03-01-2020	US-2020-103800	77015	
1	04-01-2020	US-2020-112326	60540	
2	04-01-2020	US-2020-112326	60540	
3	04-01-2020	US-2020-112326	60540	
4	05-01-2020	US-2020-141817	19143	

	Product Name	Region	Segment	\
0	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer	
1	GBC Standard Plastic Binding Systems Combs	Central	Home Office	
2	Avery 508	Central	Home Office	
3	SAFCO Boltless Steel Shelving	Central	Home Office	
4	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East	Consumer	

	Ship Date	Ship Mode	State/Province	Sub-Category	Discount	Profit	\
0	07-01-2020	Standard Class	Texas	Paper	0.2	5.5512	
1	08-01-2020	Standard Class	Illinois	Binders	0.8	-5.4870	
2	08-01-2020	Standard Class	Illinois	Labels	0.2	4.2717	
3	08-01-2020	Standard Class	Illinois	Storage	0.2	-64.7748	
4	12-01-2020	Standard Class	Pennsylvania	Art	0.2	4.8840	

	Quantity	Sales
0	2	16.448
1	2	3.540
2	3	11.784
3	3	272.736
4	3	19.536

In [96]: `print(store.tail())`

	Category	City	Country/Region	Customer Name \
10189	Office Supplies	New York City	United States	Patrick O'Donnell
10190	Office Supplies	Fairfield	United States	Erica Bern
10191	Office Supplies	Loveland	United States	Jill Matthias
10192	Technology	New York City	United States	Patrick O'Donnell
10193	Office Supplies	Charlottetown	Canada	Harry Olson

	Manufacturer	Order Date	Order ID	Postal Code \
10189	Wilson Jones	30-12-2023	US-2023-143259	10009
10190	GBC	30-12-2023	US-2023-115427	94533
10191	Other	30-12-2023	US-2023-156720	80538
10192	Other	30-12-2023	US-2023-143259	10009
10193	Wilson Jones	30-12-2023	CA-2023-143500	C0A

	Product Name	Region	Segment	Ship Date \
10189	Wilson Jones Legal Size Ring Binders	East	Consumer	03-01-2024
10190	GBC Binding covers	West	Corporate	03-01-2024
10191	Bagged Rubber Bands	West	Consumer	03-01-2024
10192	Gear Head AU3700S Headset	East	Consumer	03-01-2024
10193	Wilson Jones Impact Binders	East	Consumer	03-01-2024

	Ship Mode	State/Province	Sub-Category	Discount	Profit \
10189	Standard Class	New York	Binders	0.2	19.7910
10190	Standard Class	California	Binders	0.2	6.4750
10191	Standard Class	Colorado	Fasteners	0.2	-0.6048
10192	Standard Class	New York	Phones	0.0	2.7279
10193	Standard Class	Prince Edward Island	Binders	0.2	-0.6048

	Quantity	Sales
10189	3	52.776
10190	2	20.720
10191	3	3.024
10192	7	90.930
10193	3	3.024

In []: