

Super store data analysis using pandas

```
In [2]: # Importing required package "pandas" to Load data
import pandas as pd
```

```
In [3]: # importing data using pandas
store = pd.read_csv("Superstore_Orders.csv")
store
```

Out[3]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77095	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540	GBC Standard Plastic Binding Systems Combs	Central	Home Office
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60540	Avery 508	Central	Home Office
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60540	SAFCO Boltless Steel Shelving	Central	Home Office
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19143	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East	Consumer
...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	US-2023-143259	10009	Wilson Jones Legal Size Ring Binders	East	Consumer
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	US-2023-115427	94533	GBC Binding covers	West	Corporate
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720	80538	Bagged Rubber Bands	West	Consumer
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259	10009	Gear Head AU3700S Headset	East	Consumer
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500	C0A	Wilson Jones Impact Binders	East	Consumer

10194 rows × 19 columns



In [4]: `# For more clarification view print first 5 rows
store.head(5)`

Out[4]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment	Ship Date
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77095	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer	07-01-2020
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540	GBC Standard Plastic Binding Systems Combs	Central	Home Office	08-01-2020
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60540	Avery 508	Central	Home Office	08-01-2020
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60540	SAFCO Boltless Steel Shelving	Central	Home Office	08-01-2020
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19143	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East	Consumer	12-01-2020

In [5]: `# For identify how many rows of the data print last 5 rows
store.tail(5)`

Out[5]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	US-2023-143259	10009	Wilson Jones Legal Size Ring Binders	East	Consumer
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	US-2023-115427	94533	GBC Binding covers	West	Corporate
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720	80538	Bagged Rubber Bands	West	Consumer
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259	10009	Gear Head AU3700S Headset	East	Consumer
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500	C0A	Wilson Jones Impact Binders	East	Consumer

Form the beginning row of the dataset identify as row number 0, but row number 0 contain values, So python identify as row. That means dataset contains 10194 rows

In [7]: `## Using shape function to find total number of rows and columns
store.shape`

Out[7]: (10194, 19)

```
In [8]: # Display columns
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 [10]: # checking is there any null values present in our dataset
store.isnull()
```

Out[10]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment	Ship Date	Ship Mode
0	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False
...
10189	False	False	False	False	False	False	False	False	False	False	False	False	False
10190	False	False	False	False	False	False	False	False	False	False	False	False	False
10191	False	False	False	False	False	False	False	False	False	False	False	False	False
10192	False	False	False	False	False	False	False	False	False	False	False	False	False
10193	False	False	False	False	False	False	False	False	False	False	False	False	False

10194 rows × 14 columns



Seems like there is no null values present in our data it better cross the null by different function.

```
In [11]: # sum() function help to identify number null values present in the dataset of each individual columns
store.isnull().sum()
```

```
Out[11]: 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
```

Now we can say there is no null value in dataset. Then we can see what type data present in each column in the dataset

```
In [12]: # Using dtypes to find the present datatypes of each columns
store.dtypes
```

```
Out[12]: 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 [11]: 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 [12]: store.describe()
```

```
Out[12]:
```

	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 [13]: id(store)
```

```
Out[13]: 2762700850960
```

```
In [14]: type(store)
```

Out[14]: pandas.core.frame.DataFrame

```
In [16]: store['City']
```

Out[16]: 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 [17]: store[['City', 'Category']]
```

Out[17]:

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

10194 rows × 2 columns

```
In [18]: store[['City', 'Category', 'Order ID']]
```

Out[18]:

	City	Category	Order ID
0	Houston	Office Supplies	US-2020-103800
1	Naperville	Office Supplies	US-2020-112326
2	Naperville	Office Supplies	US-2020-112326
3	Naperville	Office Supplies	US-2020-112326
4	Philadelphia	Office Supplies	US-2020-141817
...
10189	New York City	Office Supplies	US-2023-143259
10190	Fairfield	Office Supplies	US-2023-115427
10191	Loveland	Office Supplies	US-2023-156720
10192	New York City	Technology	US-2023-143259
10193	Charlottetown	Office Supplies	CA-2023-143500

10194 rows × 3 columns

```
In [20]: store_text = store[['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' ]]
store_text
```

Out[20]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77095	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540	GBC Standard Plastic Binding Systems Combs	Central	Home Office
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60540	Avery 508	Central	Home Office
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60540	SAFCO Boltless Steel Shelving	Central	Home Office
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19143	Avery Hi-Liter EverBold Pen Style Fluorescent	East	Consumer
...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	US-2023-143259	10009	Wilson Jones Legal Size Ring Binders	East	Consumer
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	US-2023-115427	94533	GBC Binding covers	West	Corporate
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720	80538	Bagged Rubber Bands	West	Consumer
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259	10009	Gear Head AU3700S Headset	East	Consumer
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500	C0A	Wilson Jones Impact Binders	East	Consumer

10194 rows × 15 columns



```
In [21]: 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 [22]: store_number = store[['Discount', 'Profit', 'Quantity', 'Sales']]
store_number
```

```
Out[22]:
```

	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 [23]: store_text.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'],
              dtype='object')
```

```
In [24]: len(store_text.columns)
```

```
Out[24]: 15
```

In [25]: store_number.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 [26]: store_number.columns

Out[26]: Index(['Discount', 'Profit', 'Quantity', 'Sales'], dtype='object')

In [27]: len(store_number.columns)

Out[27]: 4

In [30]: store[:]

Out[30]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77095	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540	GBC Standard Plastic Binding Systems Combs	Central	Home Office
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60540	Avery 508	Central	Home Office
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60540	SAFCO Boltless Steel Shelving	Central	Home Office
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19143	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East	Consumer
...
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	US-2023-143259	10009	Wilson Jones Legal Size Ring Binders	East	Consumer
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	US-2023-115427	94533	GBC Binding covers	West	Corporate
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720	80538	Bagged Rubber Bands	West	Consumer
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259	10009	Gear Head AU3700S Headset	East	Consumer
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500	C0A	Wilson Jones Impact Binders	East	Consumer

10194 rows × 19 columns

In [31]:

store[::-1]

Out[31]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment
10193	Office Supplies	Charlottetown	Canada	Harry Olson	Wilson Jones	30-12-2023	CA-2023-143500	C0A	Wilson Jones Impact Binders	East	Consumer
10192	Technology	New York City	United States	Patrick O'Donnell	Other	30-12-2023	US-2023-143259	10009	Gear Head AU3700S Headset	East	Consumer
10191	Office Supplies	Loveland	United States	Jill Matthias	Other	30-12-2023	US-2023-156720	80538	Bagged Rubber Bands	West	Consumer
10190	Office Supplies	Fairfield	United States	Erica Bern	GBC	30-12-2023	US-2023-115427	94533	GBC Binding covers	West	Corporate
10189	Office Supplies	New York City	United States	Patrick O'Donnell	Wilson Jones	30-12-2023	US-2023-143259	10009	Wilson Jones Legal Size Ring Binders	East	Consumer
...
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19143	Avery Hi-Liter EverBold Pen Style Fluorescent ...	East	Consumer
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60540	SAFCO Boltless Steel Shelving	Central	Home Office
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60540	Avery 508	Central	Home Office
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540	GBC Standard Plastic Binding Systems Combs	Central	Home Office
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77095	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer

10194 rows × 19 columns

◀

▶

In [32]: store[0:100:10]

Out[32]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77095	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	US-2020-167199	42420	Southworth 25% Cotton Granite Paper & Envelopes	South	Home Office
20	Furniture	Dover	United States	Seth Vernon	DAX	11-01-2020	US-2020-130092	19901	DAX Value U-Channel Document Frames, Easel Back	East	Consumer
30	Office Supplies	San Francisco	United States	Brian Dahlen	Tennsco	13-01-2020	US-2020-157147	94109	Tennsco 6- and 18-Compartment Lockers	West	Consumer
40	Office Supplies	Scottsdale	United States	Toby Swindell	GBC	19-01-2020	US-2020-146591	85254	GBC Standard Recycled Report Covers, Clear Pla...	West	Consumer
50	Office Supplies	Jonesboro	United States	Hunter Lopez	Newell	20-01-2020	US-2020-147627	72401	Newell 327	South	Consumer
60	Furniture	Miami	United States	Tom Boeckenhauer	Ultra	21-01-2020	US-2020-110422	33180	Ultra Door Pull Handle	South	Consumer
70	Furniture	Alexandria	United States	Shirley Daniels	GE	26-01-2020	US-2020-155502	22304	GE General Use Halogen Bulbs, 100 Watts, 1 Bul...	South	Home Office
80	Office Supplies	New York City	United States	Jasper Cacioppo	Pressboard	28-01-2020	US-2020-100328	10024	Pressboard Covers with Storage Hooks, 9 1/2" X...	East	Consumer
90	Office Supplies	Escondido	United States	Mark Haberlin	GBC	04-02-2020	US-2020-104808	92025	GBC Standard Recycled Report Covers, Clear Pla...	West	Corporate

In [33]: store[10:11]

Out[33]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment	Ship Date
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	US-2020-167199	42420	Southworth 25% Cotton Granite Paper & Envelopes	South	Home Office	10-01-2020

```
In [34]: store_text[:4]
```

Out[34]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code	Product Name	Region	Segment	Ship Date
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77095	Message Book, Wirebound, Four 5 1/2" X 4" Form...	Central	Consumer	07-01-2020
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	60540	GBC Standard Plastic Binding Systems Combs	Central	Home Office	08-01-2020
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	60540	Avery 508	Central	Home Office	08-01-2020
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	60540	SAFCO Boltless Steel Shelving	Central	Home Office	08-01-2020

```
In [ ]:
```