

STORE DATA ANALYSIS USING PANDAS

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

```
In [5]: store = pd.read_csv(r'D:\DA ALL NOTES\DAY19\Sample - Superstore_Orders.csv')
```

```
In [7]: store
```

Out[7]:

	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 [9]: id(store)

Out[9]: 1804827076096

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

Out[11]: 10194

```
In [13]: store.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]: store.shape # 10194 rows & 13 columns
```

Out[15]: (10194, 19)

```
In [17]: len(store.columns)
```

Out[17]: 19

```
In [19]: store.isnull()
```

Out[19]:

	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 × 19 columns



```
In [21]: store.isna()
```

Out[21]:

	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 × 19 columns



In [23]: `store.isnull().sum()`

```
Out[23]: 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 [25]: `store.dtypes`

```
Out[25]: 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 [27]: 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 [29]: pd.__version__
```

```
Out[29]: '2.2.2'
```

```
In [31]: store.head()
```

Out[31]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Post Cost
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	7709
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	6054
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	6054
3	Office Supplies	Naperville	United States	Phillina Ober	SAFCO	04-01-2020	US-2020-112326	6054
4	Office Supplies	Philadelphia	United States	Mick Brown	Avery	05-01-2020	US-2020-141817	19

In [33]: store.head(3)

Out[33]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Post Cost
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	7709
1	Office Supplies	Naperville	United States	Phillina Ober	GBC	04-01-2020	US-2020-112326	6054
2	Office Supplies	Naperville	United States	Phillina Ober	Avery	04-01-2020	US-2020-112326	6054

In [35]: store.tail()

Out[35]:

	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 [37]: store.tail(4)

Out[37]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID
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 [39]: store[10:50:10]

Out[39]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
10	Office Supplies	Henderson	United States	Maria Etezadi	Southworth	06-01-2020	US-2020-167199	42
20	Furniture	Dover	United States	Seth Vernon	DAX	11-01-2020	US-2020-130092	19
30	Office Supplies	San Francisco	United States	Brian Dahlen	Tennsco	13-01-2020	US-2020-157147	94
40	Office Supplies	Scottsdale	United States	Toby Swindell	GBC	19-01-2020	US-2020-146591	85

In [41]: `store.head(1)`

Out[41]:

	Category	City	Country/Region	Customer Name	Manufacturer	Order Date	Order ID	Postal Code
0	Office Supplies	Houston	United States	Darren Powers	Message Book	03-01-2020	US-2020-103800	77095

In [43]: `store['Category']`

```
Out[43]: 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 [45]: `store[['Category', 'City']]`

Out[45]:

	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 × 2 columns

```
In [47]: store_num=store[['Discount','Profit','Quantity','Sales']]
store_num # numerical data only
```

Out[47]:

	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 [49]: store.columns
```

```
Out[49]: 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 [51]: store_cat=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',  
        'Discount', 'Profit', 'Quantity', 'Sales']]
```

```
In [53]: store_cat
```

Out[53]:

	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 [55]: store_cat.dtypes

```
Out[55]: 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 [57]: store['Profit'].mean()
```

```
Out[57]: 28.673417166960963
```

```
In [59]: store['Profit'].median()
```

```
Out[59]: 8.69
```

```
In [61]: store['Profit'].mode()
```

```
Out[61]: 0      0.0
Name: Profit, dtype: float64
```

```
In [63]: store['Profit'].std()
```

```
Out[63]: 232.46511505662147
```

```
In [65]: store['Profit'].var()
```

```
Out[65]: 54040.02971828826
```

STORE DATA ANALYSIS USING PANDAS COMPLETED