

```
In [1]: import pandas as pd
df=pd.read_csv('Sample - Superstore.csv',encoding='latin-1')
```

In [4]:

[4:1]	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City		Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit	
	0	1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gule	Consumer	United States	Henderson	--	42420	South	FUR-B0-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.9600	2	0.00	41.9136
	1	2	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gule	Consumer	United States	Henderson	--	42420	South	FUR-FU-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs	731.9400	3	0.00	219.5820
	2	3	CA-2016-138688	4/12/2016	6/16/2016	Second Class	DV-13045	Darin Van Huff	Corporate	United States	Los Angeles	--	90036	West	OFF-LA-10000340	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters	14.6200	2	0.00	6.8714
	3	4	US-2015-108866	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	--	33311	South	FUR-FL-10000577	Furniture	Tables	Bretford CH45SD Series Slim Rectangular Table	957.5775	5	0.45	-383.0310
	4	5	US-2015-108866	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	--	33311	South	OFF-ST-10000760	Office Supplies	Storage	Eldon Fold N Roll Cart System	22.3600	2	0.20	2.5164
	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	
	9989	9990	CA-2014-104042	1/21/2014	1/23/2014	Second Class	TB-21400	Tom Backenbauer	Consumer	United States	Miami	--	33180	South	FUR-FU-10001889	Furniture	Furnishings	Ultra Door Pull Handle	23.2480	3	0.20	4.1028
	9990	9991	CA-2017-121258	2/26/2017	3/3/2017	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	--	92627	West	FUR-FU-10000347	Furniture	Furnishings	Tenex BT-RE Series Chair Mats for Low Pile Carpet	91.9600	2	0.00	15.6332
	9991	9992	CA-2017-121258	2/26/2017	3/3/2017	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	--	92627	West	TEC-PA-10000845	Technology	Phones	Astra 571 VoIP phone	258.5760	2	0.20	19.3932
	9992	9993	CA-2017-121258	2/26/2017	3/3/2017	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	--	92627	West	OFF-PA-10000461	Office Supplies	Paper	It's Hot Message Books with Stickers, 2.34" x 5"	29.6000	4	0.00	13.3200
	9993	9994	CA-2015-119914	5/4/2017	5/6/2017	Second Class	CC-12220	Chris Cortes	Consumer	United States	Westminster	--	92683	West	OFF-AP-10000284	Office Supplies	Appliances	Acco 7-Outlet Masterpiece Power Center With USB	243.1600	2	0.00	72.9480

9994 rows x 21 columns

In [5]: df.isnull().sum()

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

In [6]: df.duplicated().sum()

Out[6]: 0

In [7]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 21 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Row ID      9994 non-null   int64
1   Order ID    9994 non-null   object
2   Order Date  9994 non-null   object
3   Ship Date   9994 non-null   object
4   Ship Mode   9994 non-null   object
5   Customer ID 9994 non-null   object
6   Customer Name 9994 non-null   object
7   Segment     9994 non-null   object
8   Country     9994 non-null   object
9   City        9994 non-null   object
10  State       9994 non-null   object
11  Postal Code  9994 non-null   int64
12  Region      9994 non-null   object
13  Product ID  9994 non-null   object
14  Category    9994 non-null   object
15  Sub-Category 9994 non-null   object
16  Product Name 9994 non-null   object
17  Sales       9994 non-null   float64
18  Quantity    9994 non-null   int64
19  Discount    9994 non-null   float64
20  Profit      9994 non-null   float64
dtypes: float64(3), int64(3), object(13)
memory usage: 1.6+ MB
```

In [9]: df['Order Date']=pd.to\_datetime(df['Order Date'])

df['Ship Date']=pd.to\_datetime(df['Ship Date'])

In [18]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 21 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Row ID      9994 non-null   int64
1   Order ID    9994 non-null   object
2   Order Date  9994 non-null   datetime64[ns]
3   Ship Date   9994 non-null   datetime64[ns]
4   Ship Mode   9994 non-null   object
5   Customer ID 9994 non-null   object
6   Customer Name 9994 non-null   object
7   Segment     9994 non-null   object
8   Country     9994 non-null   object
9   City        9994 non-null   object
10  State       9994 non-null   object
11  Postal Code  9994 non-null   int64
12  Region      9994 non-null   object
13  Product ID  9994 non-null   object
14  Category    9994 non-null   object
15  Sub-Category 9994 non-null   object
16  Product Name 9994 non-null   object
17  Sales       9994 non-null   float64
18  Quantity    9994 non-null   int64
19  Discount    9994 non-null   float64
20  Profit      9994 non-null   float64
dtypes: datetime64(2), float64(3), int64(3), object(13)
memory usage: 1.6+ MB
```

In [11]: df['Order Year']=df['Order Date'].dt.year

df['Order Month']=df['Order Date'].dt.month

df['Order Day of Week']=df['Order Date'].dt.dayofweek

In [12]: df.head()

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit	Order Year	Order Month	Order Day of Week
0	1	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gule	Consumer	United States	Henderson	42420	South	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.9600	2	0.00	41.9136	2016	11	1
1	2	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gule	Consumer	United States	Henderson	42420	South	Furniture	Chairs	Hon Delux Fabric Upholstered Seating Chair	731.9400	3	0.00	219.5820	2016	11	1
2	3	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DW-13045	Darin Van Huff	Corporate	United States	Los Angeles	90036	West	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters b...	14.6200	2	0.00	6.8714	2016	6	6
3	4	US-2015-108866	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	33311	South	Furniture	Tables	Bretford CH45SD Series Slim Rectangular Table	957.5775	5	0.45	-383.0310	2015	10	6
4	5	US-2015-108866	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	33311	South	Office Supplies	Storage	Eldon Fold N Roll Cart System	22.3600	2	0.20	2.5164	2015	10	6

5 rows x 24 columns

In [13]: df.describe()

	Row ID	Order Date	Ship Date	Postal Code	Sales	Quantity	Discount	Profit	Order Year	Order Month	Order Day of Week
count	9994	9994	9994	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000
mean	4997.500000	2016-04-30 00:07:122095568	2016-05-03 23:06:571142912	55190.379428	229.858001	3.789574	0.156263	-26.56896	2015.722213	7.809686	3.206124
min	1.000000	2014-01-01 00:00:000	2014-01-01 00:00:000	1040.000000	0.444400	1.000000	0.000000	-6599.978000	2014.000000	1.000000	0.000000
25%	2489.500000	2015-05-23 00:00:000	2015-05-27 00:00:000	23223.000000	17.280000	2.000000	0.000000	1.728750	2015.000000	5.000000	1.000000
50%	4997.500000	2016-06-26 00:00:000	2016-06-29 00:00:000	54490.500000	54.490000	3.000000	0.200000	8.666590	2016.000000	9.000000	4.000000
75%	7495.500000	2017-05-14 00:00:000	2017-05-18 00:00:000	90008.000000	209.940000	5.000000	0.200000	293460.00	2017.000000	11.000000	5.000000
max	9994.000000	2017-12-30 00:00:000	2018-01-05 00:00:000	99181.000000	22638.480000	14.000000	0.800000	8199.976000	2017.000000	12.000000	6.000000
std	1385.163629	NaN	NaN	32063.693350	623.245101	2.225110	0.206452	234.260108	1.323555	3.784654	2.122925

## Monthly sales analysis

```
In [28]: import plotly.express as px
import plotly.graph_objects as go
import plotly.io as pio
```

## Monthly sales analysis

In [20]: import matplotlib.pyplot as plt

import pandas as pd

import matplotlib.pyplot as plt

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd

import pandas as pd