

```
#import pandas for Data Analysis
```

```
import pandas as pd
```

```
#read excel dataset into python
```

```
df = pd.read_excel(r"C:\Users\Syed Abdul Sami\Downloads\Ecommerce  
Sales Analysis.xlsx")
```

```
#view the top 5 rows of the data set to see the dataframe
```

```
df.head(5)
```

	Row ID	Order ID	Year	Order Date	Ship Date	Ship Mode
0	1	CA-2013-152156	2013	2013-11-09	2013-11-12	Second Class
1	2	CA-2013-152156	2013	2013-11-09	2013-11-12	Second Class
2	3	CA-2013-138688	2013	2013-06-13	2013-06-17	Second Class
3	4	US-2012-108966	2012	2012-10-11	2012-10-18	Standard Class
4	5	US-2012-108966	2012	2012-10-11	2012-10-18	Standard Class

	Customer ID	Customer Name	Segment	Country	Postal Code
0	CG-12520	Claire Gute	Consumer	United States	42420
1	CG-12520	Claire Gute	Consumer	United States	42420
2	DV-13045	Darrin Van Huff	Corporate	United States	90036
3	SO-20335	Sean O'Donnell	Consumer	United States	33311
4	SO-20335	Sean O'Donnell	Consumer	United States	33311

	Region	Product ID	Category	Sub-Category
0	South	FUR-BO-10001798	Furniture	Bookcases
1	South	FUR-CH-10000454	Furniture	Chairs
2	West	OFF-LA-10000240	Office Supplies	Labels
3	South	FUR-TA-10000577	Furniture	Tables
4	South	OFF-ST-10000760	Office Supplies	Storage

	Product Name	Sales
0	Bush Somerset Collection Bookcase	261.9600
1	Hon Deluxe Fabric Upholstered Stacking Chairs,...	731.9400
2	Self-Adhesive Address Labels for Typewriters b...	14.6200

```

3      Bretford CR4500 Series Slim Rectangular Table  957.5775
5
4      Eldon Fold 'N Roll Cart System  22.3680
2

```

```

      Discount  Profit
0      0.00  41.9136
1      0.00  219.5820
2      0.00   6.8714
3      0.45 -383.0310
4      0.20   2.5164

```

```
[5 rows x 22 columns]
```

*#information of the dataframe to evaluate the nature of the data*  
df.info()

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 22 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Row ID                9994 non-null  int64
1   Order ID              9994 non-null  object
2   Year                  9994 non-null  int64
3   Order Date            9994 non-null  datetime64[ns]
4   Ship Date             9994 non-null  datetime64[ns]
5   Ship Mode              9994 non-null  object
6   Customer ID           9994 non-null  object
7   Customer Name         9994 non-null  object
8   Segment               9994 non-null  object
9   Country               9994 non-null  object
10  City                  9994 non-null  object
11  State                 9994 non-null  object
12  Postal Code           9994 non-null  int64
13  Region                9994 non-null  object
14  Product ID            9994 non-null  object
15  Category              9994 non-null  object
16  Sub-Category          9994 non-null  object
17  Product Name          9994 non-null  object
18  Sales                 9994 non-null  float64
19  Quantity              9994 non-null  int64
20  Discount              9994 non-null  float64
21  Profit                9994 non-null  float64
dtypes: datetime64[ns](2), float64(3), int64(4), object(13)
memory usage: 1.7+ MB

```

*#to see the vector shape of the data; no. of rows by no. columns*  
df.shape

(9994, 22)

*#check for null values*

```
df.isnull().sum()
```

Row ID	0
Order ID	0
Year	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

dtype: int64

*#check for duplicates*

```
df.duplicated().sum()
```

0

*#check for unique values in each column*

```
df.nunique()
```

Row ID	9994
Order ID	5009
Year	4
Order Date	1238
Ship Date	1334
Ship Mode	4
Customer ID	793
Customer Name	793
Segment	3
Country	1
City	531
State	49
Postal Code	631
Region	4

```
Product ID      1862
Category         3
Sub-Category    17
Product Name    1841
Sales           6144
Quantity        14
Discount        12
Profit          7545
dtype: int64
```

```
#check for unique values in each column
df.nunique()
```

```
Row ID          9994
Order ID        5009
Year            4
Order Date      1238
Ship Date       1334
Ship Mode       4
Customer ID     793
Customer Name   793
Segment         3
Country         1
City            531
State           49
Postal Code     631
Region          4
Product ID      1862
Category         3
Sub-Category    17
Product Name    1841
Sales           6144
Quantity        14
Discount        12
Profit          7545
dtype: int64
```