In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as snm
import numpy as np
```

In [3]:
```
sample_data=pd.read_csv(r"A:\LLM\Sample - Superstore.csv", encoding="ISO-8859-1")
sample_data.head()
```

Out[3]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Po C
0	1	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42
1	2	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42
2	3	CA-2016-138688	12-06-2016	16-06-2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	90
3	4	US-2015-108966	11-10-2015	18-10-2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33
4	5	US-2015-108966	11-10-2015	18-10-2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33

5 rows × 21 columns

In [4]:
```
sample_data[1:3]
```

Out[4]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Pos Co
1	2	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	424
2	3	CA-2016-138688	12-06-2016	16-06-2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	900

2 rows × 21 columns

In [5]:
```
#splitting order id based upon the -(hypen)

sample_data.columns
```

Out[5]:
```
Index(['Row ID', 'Order ID', 'Order Date', 'Ship Date', 'Ship Mode',
       'Customer ID', 'Customer Name', 'Segment', 'Country', 'City', 'State',
       'Postal Code', 'Region', 'Product ID', 'Category', 'Sub-Category',
       'Product Name', 'Sales', 'Quantity', 'Discount', 'Profit'],
      dtype='object')
```

In [9]:
```
sample_data["Order_ID_split"]=sample_data["Order ID"]
```

In [7]:
```
sample_data.columns
```

Out[7]:
```
Index(['Row ID', 'Order ID', 'Order Date', 'Ship Date', 'Ship Mode',
       'Customer ID', 'Customer Name', 'Segment', 'Country', 'City', 'State',
       'Postal Code', 'Region', 'Product ID', 'Category', 'Sub-Category',
       'Product Name', 'Sales', 'Quantity', 'Discount', 'Profit',
       'Order_ID_split'],
      dtype='object')
```

In [19]:
```
#splitting data based on str function and slicing
#manually hardcaode the splicing values

sample_data["Order_ID_region"]=sample_data["Order ID"].str[0:2]
sample_data["Order_ID_year"]=sample_data["Order ID"].str[3:7]
sample_data["Order_ID_id"]=sample_data["Order ID"].str[8:]
print(sample_data["Order_ID_region"].head(2))
print(sample_data["Order_ID_year"].head(2))
print(sample_data["Order_ID_id"].head(2))
```

```
0    CA
1    CA
Name: Order_ID_region, dtype: object
0    2016
1    2016
Name: Order_ID_year, dtype: object
0    152156
1    152156
Name: Order_ID_id, dtype: object
```

In [20]:
```
sample_data.head(2)
```

Out[20]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Ca
0	1	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	Boc
1	2	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	

2 rows × 25 columns

In [33]:
```
#Convert order date value in to datetime datatype
#here we have used the to_datetime

sample_data["Order_date_time_value"]=pd.to_datetime(sample_data["Order Date"])
sample_data["Order_date_time_value"]

sample_data.dtypes
```

Out[33]:
```
Row ID                           int64
Order ID                        object
Order Date                      object
Ship Date                       object
Ship Mode                       object
Customer ID                     object
Customer Name                   object
Segment                         object
Country                         object
City                            object
State                           object
Postal Code                      int64
Region                          object
Product ID                      object
Category                        object
Sub-Category                    object
Product Name                    object
Sales                          float64
Quantity                         int64
Discount                       float64
Profit                         float64
Order_ID_split                  object
Order_ID_region                 object
Order_ID_year                   object
Order_ID_id                     object
Order_date_time_value    datetime64[ns]
dtype: object
```

In [36]:
```
sample_data["order_year"]=sample_data["Order_date_time_value"].dt.year
sample_data["order_year"].head()
```

Out[36]:
```
0    2016
1    2016
2    2016
3    2015
4    2015
Name: order_year, dtype: int64
```

In [38]:
```
#before using the dt makesure you have converted that column in datatime

sample_data["Order_month"]=sample_data["Order_date_time_value"].dt.month
sample_data["Order_day"]=sample_data["Order_date_time_value"].dt.day
print(sample_data["Order_month"].head(2))
print(sample_data["Order_day"].head(2))
print(sample_data["Order_date_time_value"])
```

```
0    8
1    8
Name: Order_month, dtype: int64
```

```
0    11
1    11
Name: Order_day, dtype: int64
0      2016-08-11
1      2016-08-11
2      2016-12-06
3      2015-11-10
4      2015-11-10
          ...
9989   2014-01-21
9990   2017-02-26
9991   2017-02-26
9992   2017-02-26
9993   2017-04-05
Name: Order_date_time_value, Length: 9994, dtype: datetime64[ns]
```

In [39]:
```
sample_data.head(2)
```

Out[39]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Dis
0	1	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	
1	2	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	

2 rows × 29 columns

In [40]:
```
sample_data["Order ID_new"]=sample_data["Order ID"].str.split('-')
```

In [41]:
```
sample_data["Order ID_new"].head()
```

Out[41]:
```
0    [CA, 2016, 152156]
1    [CA, 2016, 152156]
2    [CA, 2016, 138688]
3    [US, 2015, 108966]
4    [US, 2015, 108966]
Name: Order ID_new, dtype: object
```

In [43]:
```
#Creating three columns according to split function
#make sure you are passing all three columns
#using str.split(-)
#please use expand=True , otherwise results are false

sample_data[["Order ID_part1","Order ID_part2","Order ID_part3"]]=sample_data["Order
```

In [44]:
```
sample_data["Order ID_part1"].head(2)
```

Out[44]:
```
0    CA
1    CA
Name: Order ID_part1, dtype: object
```

In [45]:
```
sample_data["Order ID_part2"].head(2)
```

Out[45]:
```
0    2016
1    2016
Name: Order ID_part2, dtype: object
```

In [46]:
```
sample_data["Order ID_part3"].head(2)
```

Out[46]:
```
0    152156
1    152156
Name: Order ID_part3, dtype: object
```

In [ ]:
```

```