```
# Task1
!pip install xlrd
```

In [52]:

```
#Task 1: Read „Orders“ Sheet from “Sample_Superstore.xls” and perform data
#cleaning.
#(Remove Duplicate Columns, Duplicate Rows, Blank Columns, and Blank Rows.)
import pandas as pd
data = pd.read_excel('Sample_Superstore.xls', sheet_name='Orders')

# Remove duplicate columns
data = data.loc[:, ~data.columns.duplicated()]

# Remove duplicate rows
data = data.drop_duplicates()

# Remove blank columns
data = data.dropna(axis=1, how='all')

# Remove blank rows
data = data.dropna(axis=0, how='all')

# Display the cleaned data
data
```

Out[52]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Sub-Category.2	Product Name.1	Sale
0	1.0	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	Bookcases	Bush Somerset Collection Bookcase	261.96
1	2.0	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs,...	731.94
2	3.0	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	Labels	Self-Adhesive Address Labels for Typewriters b...	14.62
3	4.0	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	Tables	Bretford CR4500 Series Slim Rectangular Table	957.57
4	5.0	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	Storage	Eldon Fold 'N Roll Cart System	22.36
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
11415	9980.0	US-2016-103674	2016-12-06	2016-12-10	Standard Class	AP-10720	Anne Pryor	Home Office	United States	Los Angeles	...	Binders	NaN	N
11416	9981.0	US-2015-151435	2015-09-06	2015-09-09	Second Class	SW-20455	Shaun Weien	Consumer	United States	Lafayette	...	Tables	NaN	N
11417	9982.0	CA-2017-163566	2017-08-03	2017-08-06	First Class	TB-21055	Ted Butterfield	Consumer	United States	Fairfield	...	Labels	NaN	N
11418	9983.0	US-2016-157728	2016-09-22	2016-09-28	Standard Class	RC-19960	Ryan Crowe	Consumer	United States	Grand Rapids	...	Paper	NaN	N
11419	9984.0	US-2016-157728	2016-09-22	2016-09-28	Standard Class	RC-19960	Ryan Crowe	Consumer	United States	Grand Rapids	...	Phones	NaN	N

11249 rows × 38 columns

# Task2

In [50]:

```
#Task 2: On which Order Date, the Super Store company generated highest profit.
import pandas as pd
data['Profit'] = pd.to_numeric(data['Profit'], errors='coerce').fillna(0)
profit = data.groupby("Order Date")["Profit"].sum()

p=profit.max()
date = profit.idxmax()

print('Date :',date)
print('Max Profits' ,p)
```

```
Date : 2016-10-02 00:00:00
Max Profits 8738.7971
```

# Task3

In [55]:

```
#Task 3: Calculate average sales w.r.t. States.

average = data.groupby("State")["Sales"].mean()
average
```

```
Out[55]: State
         Alabama                 308.574844
         Arizona                 151.143224
         Arkansas                221.123710
         California              231.220057
         Colorado                168.818168
         Connecticut             160.225943
         Delaware                262.923173
         District of Columbia     27.832857
         Florida                 222.605841
         Georgia                 257.549550
         Idaho                   210.268667
         Illinois                159.018820
         Indiana                 343.774651
         Iowa                    133.825429
         Kansas                  121.429583
         Kentucky                279.690000
         Louisiana               232.291915
         Maine                   171.604545
         Maryland                213.927219
         Massachusetts           215.781369
         Michigan                310.400550
         Minnesota               327.102737
         Mississippi             199.539273
         Missouri                398.180533
         Montana                 372.623467
         Nebraska                192.800976
         Nevada                  401.319826
         New Hampshire           263.350483
         New Jersey              251.269169
         New Mexico              120.023561
         New York                274.621869
         North Carolina          227.001500
         North Dakota            131.415714
         Ohio                    162.373207
         Oklahoma                310.660000
         Oregon                  144.847500
         Pennsylvania            200.413271
         Rhode Island            386.232983
         South Carolina          201.945476
         South Dakota            109.630000
         Tennessee               174.095042
         Texas                   174.357128
```

```
Utah                        203.244929
Vermont                     846.925500
Virginia                    289.636743
Washington                  266.461544
West Virginia               376.633600
Wisconsin                   284.248231
Wyoming                    1603.136000
Name: Sales, dtype: float64
```

```
# Task4
```

In [104]:

```
#Task 4: On which Order Date, the Super Store Company generated lowest profit
#and display the
#customer name and ID.

import pandas as pd

data['Profit'] = pd.to_numeric(data['Profit'], errors='coerce')

# Group the data by 'Order Date' and calculate the total profit
profit_by_date = data.groupby('Order Date')['Profit'].sum()

# Find the order date with the lowest profit
min_profit_date = profit_by_date.idxmin()
min_profit = profit_by_date.min()

# Find the customer name and ID for the order with the lowest profit
order_profit = data[data['Order Date'] == min_profit_date]
customer_name = order_profit['Customer Name'].iloc[0]
customer_id = order_profit['Customer ID'].iloc[0]

print("Order Date:", min_profit_date)
print("Lowest Profit:", min_profit)
print("Customer Name:", customer_name)
print("Customer ID:", customer_id)
```

```
Order Date: 2016-11-25 00:00:00
Lowest Profit: -6247.396900000001
Customer Name: Heather Jas
Customer ID: HJ-14875
```

In [ ]:

# Task5

In [90]:

```
#Task 5: Display the most used Ship Mode.

max_ships_used = data.groupby('Ship Mode')['Ship Mode'].count()
max_ships_used = max_ships_used.sort_values(ascending=False)
max_ships_used
```

Out[90]:

```
Ship Mode
Standard Class    6694
Second Class      2190
First Class       1745
Same Day           620
Name: Ship Mode, dtype: int64
```

# Task6

In [80]:

```
#Task 6: Display name of the Arizona“s customer who gave the highest profit to the
#company.
arizona_customers = data[data['State'] == 'Arizona']
arizona_profit = arizona_customers.groupby('Customer Name')['Profit'].sum()
customer_with_highest_profit = arizona_profit.idxmax()

print('Customer with highest profit in Arizona:', customer_with_highest_profit)
```

```
Customer with highest profit in Arizona: John Murray
```

# Task7

In [76]:

```
#Task 7: Display the name of the state that generated the most profits.

profit = data.groupby("State")["Profit"].sum()
state_with_most_profit = profit.idxmax()

sorted_data = profit.sort_values(ascending=False)
print(sorted_data)
```

```
State
California              75778.7378
New York                72375.6788
Washington              32731.3589
Michigan                23694.1123
Indiana                 18312.9685
Virginia                17125.9785
Georgia                 16250.0433
Kentucky                11162.8274
Minnesota               10823.1874
Delaware                 9971.2708
New Jersey               9772.9138
Wisconsin                8401.8004
Rhode Island             7285.6293
Maryland                 7031.1788
Massachusetts            6785.5016
Missouri                 6329.4194
Alabama                  5786.8253
Oklahoma                 4853.9560
Arkansas                 3935.4247
Connecticut              3511.4918
Nevada                   3316.7659
Mississippi              3129.5262
Utah                     2546.5335
Vermont                  2244.9783
Louisiana                2196.1023
Nebraska                 2037.0942
Montana                  1833.3285
South Carolina           1769.0566
New Hampshire            1706.5028
Iowa                     1108.7821
New Mexico               1102.7195
Kansas                    836.4435
Idaho                     826.7231
Maine                     454.4862
South Dakota              388.1983
North Dakota              230.1497
West Virginia             185.9216
Wyoming                   100.1960
District of Columbia       85.2758
Oregon                  -1189.6785
Arizona                 -2533.2013
Florida                 -3549.4853
```

```
Tennessee                -5358.2296
Colorado                 -6660.1519
North Carolina           -7528.6402
Illinois                -12613.8860
Pennsylvania            -15115.7802
Ohio                    -16996.6104
Texas                   -25775.4835
Name: Profit, dtype: float64
```

# Task8

In [77]:

```
#Task 8: Display the name of Category that generated the most sales.

import pandas as pd

category_sales = data.groupby('Category')['Sales'].sum()
category_with_most_sales = category_sales.idxmax()
print("Category with the Most Sales:", category_with_most_sales)
```

```
Category with the Most Sales: Technology
```

In [ ]: