Usama Arif Rollno 14

Task 1: Read „Orders‟ Sheet from “Sample_Superstore.xls” and perform data cleaning. (Remove Duplicate Columns, Duplicate Rows, Blank Columns, and Blank Rows.)

In [41]:
```
import pandas as pd
df=pd.read_excel('Sample_Superstore.xls')
#removing duplicate coumns
df=df.loc[:,~df.columns.duplicated()]
df
```

Out[41]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	C
0	1.0	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henders
1	2.0	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henders
2	3.0	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Angel
3	4.0	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fl Lauderd
4	5.0	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fl Lauderd
...	...	...	...	...	...	...	...	...	...	
11415	9980.0	US-2016-103674	2016-12-06	2016-12-10	Standard Class	AP-10720	Anne Pryor	Home Office	United States	Angel
11416	9981.0	US-2015-151435	2015-09-06	2015-09-09	Second Class	SW-20455	Shaun Weien	Consumer	United States	Lafaye
11417	9982.0	CA-2017-163566	2017-08-03	2017-08-06	First Class	TB-21055	Ted Butterfield	Consumer	United States	Fairf
11418	9983.0	US-2016-157728	2016-09-22	2016-09-28	Standard Class	RC-19960	Ryan Crowe	Consumer	United States	Gra Rap
11419	9984.0	US-2016-157728	2016-09-22	2016-09-28	Standard Class	RC-19960	Ryan Crowe	Consumer	United States	Gra Rap

11420 rows × 41 columns

In [20]:
```
#removing rows
df=df.drop_duplicates()
```

In [21]:
```
#removing blank columns
df=df.dropna(axis=1,how='all')
#removing blank rows
df=df.dropna(axis=0,how='all')
```

Task 2: On which Order Date, the Super Store company generated highest profit.

In [39]:
```
1
2 df['Profit'] = pd.to_numeric(df['Profit'], errors='coerce').fillna(0)
3 profit=df.groupby('Order Date')['Profit'].sum()
4 max_profit=profit.max()
5 print('profit\n',max_profit)
6 date=profit.idxmax()
7 print('date\n',date)
8
```

```
profit
 8738.7971
date
 2016-10-02 00:00:00
```

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

In [51]:
```
1 avearge=df.groupby('State')['Sales'].mean()
2 avearge
```

Out[51]:
```
State
Alabama                    308.574844
Arizona                    151.143224
Arkansas                   221.123710
California                 231.220057
Colorado                   168.818168
Connecticut                160.225943
Delaware                   256.919418
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                   285.591868
North Carolina             227.001500
North Dakota               131.415714
Ohio                       162.341695
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
```

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

In [70]:
```
df['Profit'] = pd.to_numeric(df['Profit'], errors='coerce')

order_date=df.groupby('Order Date')['Profit'].sum()
min_profit_date=order_date.idxmin()
print(min_profit_date)
order=df[df['Order Date']==min_profit_date]
customer_name=order['Customer Name'].iloc[0]
id=order['Customer ID'].iloc[0]
print('cutomer name\n',customer_name,'id is \n',id)
```

```
2016-11-25 00:00:00
cutomer name
 Heather Jas id is
 HJ-14875
```

Task 5: Display the most used Ship Mode.

In [88]:
```
ship_mode=df.groupby('Ship Mode')['Ship Mode'].count()
most_used=ship_mode.idxmax()
print(f'{most_used} ship mode was used mostly that is ',ship_mode.max())
```

```
Standard Class ship mode was used mostly that is  6700
```

Task 6:Display name of the Arizona‟s customer who gave the highest profit to the company.

In [89]:
```
#Task 6: Display name of the Arizona‟s customer who gave the highest profi
#company.
arizona_customers = df[df['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_pr
```

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

Task 7 Display the name of Category that generated the most sales.

In [127]:
```
import pandas as pd
category_sales = df.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
```

Task 8: Display the name of the state that generated the most profits.

In [126]:
```
df=pd.read_excel('Sample_Superstore.xls')
profit = data.groupby("State")["Profit"].sum()
state_with_most_profit = profit.idxmax()
state_with_most_profit
```

Out[126]:
```
'New York'
```

In [ ]:
```

```