**MarketBasedInsights\_phase 3**

import pandas as pd

**loading dataset**

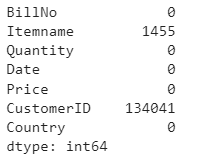
data=pd.read\_csv('Naan\_data.csv',encoding='iso-8859-1')



**counting number of values missed in each column**

missing\_values = data.isnull().sum()

missing\_values

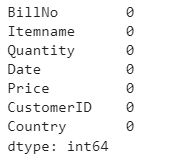


**As more data are missed in customerid, we use interpolate to fill data**

data\_cleaned = data.interpolate()

missing\_new\_values= data\_cleaned.isnull().sum()

missing\_new\_values



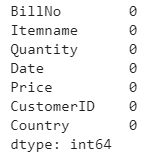
**Now we get missing values in itemname alone, hence we drop rows here**

df\_cleaned = data\_cleaned.dropna()

**final missing values**

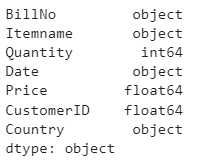
missing\_new\_values= df\_cleaned.isnull().sum()

missing\_new\_values



**printing data type of each column**

df\_cleaned.dtypes



**grouped by country**

grouped\_country = df\_cleaned.groupby('Country')

grouped\_country



Norway\_count = grouped\_country.size().get('Norway', 0)

Norway\_count



**grouped by items**

grouped\_items = df\_cleaned.groupby('Itemname')

metal\_latern\_count=grouped\_items.size().get('WHITE METAL LANTERN', 0)

metal\_latern\_count



**grouped by date**

grouped\_date = df\_cleaned.groupby('Date')

dec1 = grouped\_date.size().get('01-12-2010', 0)

dec1



**grouped by customerid**

grouped\_id = df\_cleaned.groupby('CustomerID')

id1 = grouped\_id.size().get(17850, 0)

id1



**rows, columns**

df\_cleaned.shape

