Case Study 2 (Data-driven Approach)

When it comes to the Retail Industry, meeting the customer needs is what plays a key role for the business to run successfully. So, in this case, I would be focusing on Customer Behavior to study and analyze what the customer needs. Customer Behavior mainly focuses on finding out what the customer purchases. These purchases depend on various factors such as Geographical Location, Community, Season/Weather and Preferences/Interests and coming up with offers/suggestions to meet customer satisfaction

The dataset which I would be using is Online Retail (OnlineRetail.csv) from Kaggle [URL: https://www.kaggle.com/datasets/tunguz/online-retail].

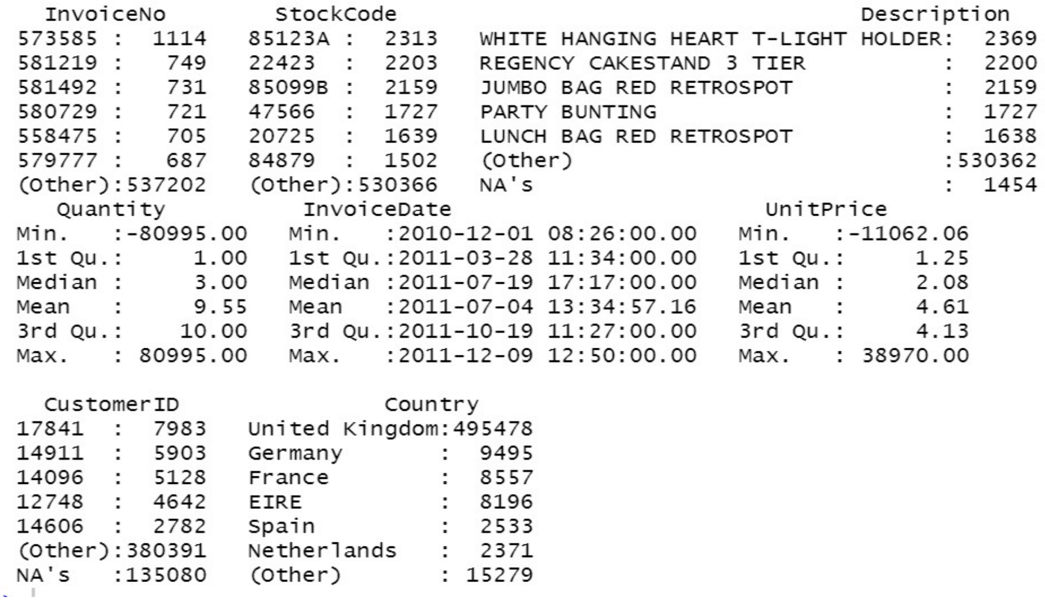
The data is in an Excel Workbook (XLSX) File with 8 Attributes and 541909 Observations.

The 8 attributes are as follows:

1. InvoiceNo - Invoice Number for the transaction/purchase - ***Numeric***
2. StockCode - Unique Code given to the Stock - ***Character / Alphanumeric***
3. Description - Description of the product - ***Character String***
4. Quantity - Quantity of the Products purchased - ***Numeric/Decimal***
5. InvoiceDate - Date on which the Invoice was generated (Purchase Date) - ***Date and Time - YYYY-MM-DD HH:MM: SS***
6. UnitPrice - Price per 1 unit of the Product - ***Numeric/Decimal***
7. CustomerID - ID of the Customer who did the transaction - ***Numeric (With NAs)***
8. Country - Country in which the transaction was made - ***Character String***

I'm planning to divide/filter the data based on the Country and then analyze the data. Instead of working on the whole data, I'll take the countries I'm interested in analyzing and look for insights. I'm planning to use R (4.2.1) and RStudio to work on the dataset. I'll change the data type (for convenience) and then plot graphs with various combinations of attributes.

Following is the Summary of the Dataset – OnlineRetail:



```
  InvoiceNo       StockCode                              Description    
 573585 :  1114   85123A :  2313   WHITE HANGING HEART T-LIGHT HOLDER:  2369  
 581219 :   749   22423  :  2203   REGENCY CAKESTAND 3 TIER          :  2200  
 581492 :   731   85099B :  2159   JUMBO BAG RED RETROSPOT           :  2159  
 580729 :   721   47566  :  1727   PARTY BUNTING                     :  1727  
 558475 :   705   20725  :  1639   LUNCH BAG RED RETROSPOT           :  1638  
 579777 :   687   84879  :  1502   (Other)                           :530362  
 (Other):537202   (Other):530366   NA's                              :  1454  
    Quantity           InvoiceDate                       UnitPrice        
 Min.   :-80995.00   Min.   :2010-12-01 08:26:00.00   Min.   :-11062.06  
 1st Qu.:     1.00   1st Qu.:2011-03-28 11:34:00.00   1st Qu.:     1.25  
 Median :     3.00   Median :2011-07-19 17:17:00.00   Median :     2.08  
 Mean   :     9.55   Mean   :2011-07-04 13:34:57.16   Mean   :     4.61  
 3rd Qu.:    10.00   3rd Qu.:2011-10-19 11:27:00.00   3rd Qu.:     4.13  
 Max.   : 80995.00   Max.   :2011-12-09 12:50:00.00   Max.   : 38970.00  

   CustomerID               Country      
 17841  :  7983   United Kingdom:495478  
 14911  :  5903   Germany       :  9495  
 14096  :  5128   France        :  8557  
 12748  :  4642   EIRE          :  8196  
 14606  :  2782   Spain         :  2533  
 (Other):380391   Netherlands   :  2371  
 NA's   :135080   (Other)       : 15279  
```

For cleaning and handling the data, *dplyr/tidyr* would be used, and for plotting, *ggplot2* would be helpful (*libraries and functions in R*).

We have NAs (Null Values) in the data for Customer ID. Customer ID is one of the important attributes that should not be empty. Since null values exist, working with them would be less efficient. We won't be a replace the data for the CustomerID since it is a unique attribute to identify who purchased the data and cannot be replaced with mean or random values.

After working on the data and plotting the graphs, we will be able to study the pattern and trends of the purchases made by the customer. With this, we would be able to analyze the customer behavior and come up with new business strategies to satisfy customer.