# Domain

Online Retail / Ecommerce

# About the business

An online retail company which sells unique gifting products primarily through online channel.

# Problem Statement

Segment the customer into different sections or segments based on various attributes like purchase value, kind of product purchased, geography etc.

These customer segmentation, rules and patterns can be used to make interesting and useful decisions as far as user interest is concerned.

# Primary Objective

* Higher Revenue
* Customer Understanding
* Target Marketing
* Optimal Product Placement

# Business Outcome

The deliverable of the project is to come up with a best model on “customer segmentation”. The right segmentation will be used for “personalised” communication instead of one for all which will result in more customer retention, better customer relationship & thus higher revenue/profit for the company.

# Database

The Online Retail a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. The company mainly sells unique all-occasion gifts.

**Dataset Kaggle:** [Link](https://www.kaggle.com/code/mgmarques/customer-segmentation-and-market-basket-analysis/data)

The Data set summary:-

First hand view

The Online Retail a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. The company mainly sells unique all-occasion gifts.

This data set is considerably large data with 541909 rows and 8 columns.

Column headers and the data type is as below-

1. InvoiceNo. – Both numeric and C series code (Possibly Credit Notes) – no missing value
2. Stockcode – Some generic numbers and some with special characters or terms- no missing
3. Description – Description of the stockcode – perhaps fed manually as well as multiple description of same stockcode available also some missing entries
4. Quantity – Both Positive and -ve entries visible (Credit note, sale reversal, discounts) impacting entry
5. InvoiceDate- transaction date for period 01-12-2010 to 09-12-2011
6. UnitPrice -generic term – different for some records / transactions
7. CustomerID – 135080 missing values identified (with +ve and -ve sold qty) and different Country
8. Country – 38 Country names of with no missing value are visible

We can build multiple stories with this data set.

|  |  |
| --- | --- |
| 1 | Product category |
| 2 | Geography |
| 3 | Low mid high |
| 4 | Loyal - Frequent |
| 5 | NLO- Next Logical product |
|  |  |