**Context:**

This is 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. Many customers of the company are wholesalers.

**Data Description**:

**InvoiceNo**: Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation.

**StockCode**: Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.

**Description**: Product (item) name. Nominal.

**Quantity**: The quantities of each product (item) per transaction. Numeric.   
InvoiceDate: Invoice Date and time. Numeric, the day and time when each transaction was generated.

**UnitPrice**: Unit price. Numeric, Product price per unit in sterling.

**CustomerID**: Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.

**Country**: Country name. Nominal, the name of the country where each customer resides.

**Problem statement**

It is a business critical requirement to understand the value derived from a customer. RFM is a method used for analyzing customer value.

Perform customer segmentation using RFM analysis. The resulting segments can be ordered from most valuable (highest recency, frequency, and value) to least valuable (lowest recency, frequency, and value). Identifying the most valuable RFM segments can capitalize on chance relationships in the data used for this analysis.

**Approach:**

Following pointers will be helpful to structure your findings.

1. Perform a preliminary data inspection and Data cleaning
   1. Check for missing data and formulate apt strategy to treat them.
   2. Are there any duplicate data records? Remove them if present.
   3. Perform Descriptive analytics on the given data.
2. Cohort Analysis: A cohort is a group of subjects who share a defining characteristic. We can observe how a cohort behaves across time and compare it to other cohorts.
   1. Create month cohorts and analyse active customers for each cohort.
   2. Also Analyse the retention rate of customers. Comment.
3. Build a RFM model – Recency Frequency and Monetary based on their behaviour.

Recency is about when was the last order of a customer. It means the number of days since a customer made the last purchase. If it’s a case for a website or an app, this could be interpreted as the last visit day or the last login time.

Frequency is about the number of purchase in a given period. It could be 3 months, 6 months or 1 year. So we can understand this value as for how often or how many a customer used the product of a company. The bigger the value is, the more engaged the customers are. Could we say them as our VIP? Not necessary. Cause we also have to think about how much they actually paid for each purchase, which means monetary value.

Monetary is the total amount of money a customer spent in that given period. Therefore big spenders will be differentiated with other customers such as MVP or VIP.

* 1. Calculate RFM metrics.
     1. Recency as the time in no. of days since last transaction
     2. Frequency as count of purchases done
     3. Monetary value as total amount spend
  2. Build RFM Segments.
     1. Give Recency Frequency and Monetary scores individually by dividing them in to quartiles.

Note: Rate "Recency" for customer who have been active more recently better than the less recent customer, because each company wants its customers to be recent

Rate "Frequency" and "Monetary Value" higher label because we want Customer to spend more money and visit more often.

* + 1. Combine three ratings to get a RFM segment (as strings)
    2. Get the RFM score by adding up the three ratings.
  1. Analyse the RFM Segments by summarizing them and comment on the findings.

1. Create clusters using k means clustering algorithm.
   1. Prepare the data for the algorithm.
      1. If the data is Un Symmetrically distributed, manage the skewness with appropriate transformation.
      2. Standardize / scale the data.
   2. Decide the optimum number of clusters to be formed
   3. Analyse these clusters and comment on the results.
2. Create a dashboard in tableau by choosing appropriate chart types and metrics useful for the business. The dashboard must entail the following:
3. Country-wise analysis to demonstrate Average spend. Use a bar chart show monthly figures.
4. Bar graph of top 15 products which are mostly ordered by the users to show the number of products sold.
5. Bar graph to show the count of orders Vs. hours throughout the day. What are the peak hours per your chart?
6. Plot the distribution of RFM values using histogram and frequency-charts.
7. Plot error(cost) vs no of clusters selected
8. Visualize to compare the RFM values of the clusters using heatmap