|  |
| --- |
|  |
| Customer Segmentation  Online Retail Store |
| |  |  |  | | --- | --- | --- | | Kalpesh Bhurabhai Odedra | [Date] | Data Mining & Modelling | |

# Dataset Description

Name: Online Retail II Dataset

Dataset was retrieved from UCI ML Repository. Data contains transaction information of products of an online retail company from Dec 2009 to Dec 2010.

Source: Dr. Daqing Chen, Course Director: MSc Data Science. chend '@' lsbu.ac.uk, School of Engineering, London South Bank University, London SE1 0AA, UK.

|  |  |  |
| --- | --- | --- |
| Fields’ Description | | |
| No | Field Name | Description |
| 1 | Invoice | Invoice number for the transaction |
| 2 | StockCode | Code of the product purchased |
| 3 | Description | Text description of the product |
| 4 | Quantity | Number of units purchased |
| 5 | InvoiceDate | Date of transaction |
| 6 | Price | Price of product |
| 7 | Customer ID | Unique Id, given to each customer |
| 8 | Country | Country of Customer |

# Data cleaning and preparation

First I checked all the fields to analyze each field.

Table

Description automatically generated

Then I decided to check how many customers are in each country. Based on the result, I chose to target

The United Kingdom, since it has the most number of customers, therefore is a prime location for new products.

Table

Description automatically generated with low confidenceFiltering by country, I checked for missing values for customer ID and negative values in price and Quantity, since Price and Quantity would be the key fields for RFM Analysis based on Customer IDs.

Filtering out all negative and missing values, I calculated the Total Amount which is Price\*Quantity of each transaction. Graphical user interface, text, application

Description automatically generated

Now the file is ready for RFM Modelling. I chose the last invoice date as the recent date and the calculated Recency field as (recent date - last invoice date of a customer) in the number of days.

Frequency is calculated as the total number of transactions of a customer and Monetary as the total amount spent. This field gives us an idea of how much money a customer has spent in the business.

Table

Description automatically generated

Then I checked for descriptive stats of each of the fields of Recency, Frequency, and Monetary.

Text

Description automatically generatedText

Description automatically generatedText

Description automatically generated

I also created distribution plots for each of the R, F & M fields.

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Dividing each of the R, F&M in 4 quantiles and scoring them based on their values I created the RFM Group, and then adding the values I calculated the RFM Score. Labeling the scores as 'Platinum', 'Gold', 'Silver', 'Bronze' loyalty levels.

These segments give us an idea of how much value a customer provides.

Table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Total values in each segment combined with individual scores also give us an idea of which segment can be targeted more based on the total potential to move up a segment. Here, there is an almost similar number of Gold and Platinum members. The scatter plots further give an understanding of how the customer segments are related w.r.t each of the fields of Recency, Frequency, and Monetary.

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

# Recommendations

The analysis of quantity and price reveals that the products in question are not high-priced products, 75% of the products have a price lower than 3.75 pounds. Also, the segmentation analysis reveals the relationship that as frequency increases, a customer moves up the segmentation profile, and also the monetary value of the customer increase.

Therefore, in industries where products are low priced and transaction quantities are high, almost all the customer segments are important.

* The marketing towards platinum and Gold customers has to be designed to keep them buying. Here loyalty points, special privileges such as access to pre-sale items, referral bonuses may work very well. Most of the marketing should be targeted towards the exclusivity of products and special treatment.
* Since these products are low margin, it is important to analyze the cost of reaching a customer through marketing channels and converting them along with the lifetime value to make sure that it is worth targeting bronze customers. If it is, then they should be given and future purchase discounts. Packaged deals, free shipping, etc work well with bronze customers.
* For Silver customers, it is worth targeting them more than bronze customers since they are closer to move to gold customers. Therefore, the ROI of marketing is higher than that of bronze customers. Silver customers should also be provided with loyalty program deals, free shipping, and next purchase discounts, time-bound coupons, etc.