RFM Analysis

# Assignment

## What is RFM Analysis?

RFM (recency, frequency, monetary) analysis is a marketing technique used to determine quantitatively which customers are the best ones by examining how recently a customer has purchased (recency), how often they purchase (frequency), and how much the customer spends (monetary).

RFM can also be understood as a method used for analyzing customer value. It is commonly used in database marketing and direct marketing and has received particular attention in retail and professional services industries.

RFM stands for the three dimensions:

Recency – How recently did the customer purchase?

Frequency – How often do they purchase?

Monetary Value – How much do they spend?

## Problem Statement

Using the Available retail dataset, perform RFM Analysis to answer following customers –

* On How many customers the marketing budget for retaining customers should be spent and what are the Customer Id of those customers?
* Who are the best customers and how many such best customers we have?
* Who are the almost lost customers and for how many such customers the marketing budget for reacquiring the customers must be spent?

|  |  |
| --- | --- |
| Skills to be tested: | Python Basics, Data aggregation, Pandas – groupby operation, lambda function |
| Dataset: | Online Retail Dataset |
| Expected Challenges: | Missing Values, Garbage Values, Date Handling, Feature Handling |
| Dataset Format: | Xlsx (Excel) |
| Dataset download Link | <https://archive.ics.uci.edu/ml/machine-learning-databases/00352/Online%20Retail.xlsx> |

## Dataset Information

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.

### Attribute Information:

* 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: Invice 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.

## Submissions

* Python Code File
* CSV with Customer ID of the Customers for 3 questions in the Problem Statement.