RFM Analysis

# Technical document

Meghnad

RFM Analysis Document Version

|  |  |  |
| --- | --- | --- |
| ***Author*** | ***Version*** | ***Date*** |
| *Souvik* | *1* | *14-06-2022* |
| *Souvik* | *2* | *29-06-2022* |

Contents

[1. Introduction 3](#_Toc106620890)

[2. Prerequisites 4](#_Toc106620891)

[3. Config 4](#_Toc106620892)

*[RFM Definition](#_Toc106620893)* [4](#_Toc106620893)

*[Assumptions](#_Toc106620894)* [4](#_Toc106620894)

*[Customer groups & user input](#_Toc106620895)* [4](#_Toc106620895)

[4. Data connectors 5](#_Toc106620896)

[5. RFM Analyzer & Sample output 5](#_Toc106620897)

[6. RFM Code Module 7](#_Toc106620898)

*[Folder structure](#_Toc106620899)* [7](#_Toc106620899)

*[RFMAnalyzerConfig](#_Toc106620900)* [7](#_Toc106620900)

[](#_Toc106620901) *[get\_rfm\_configs(self)](#_Toc106620901)* [7](#_Toc106620901)

[](#_Toc106620902) *[get\_rfm\_group\_details(self)](#_Toc106620902)* [7](#_Toc106620902)

*[RFMDefinition](#_Toc106620903)* [7](#_Toc106620903)

[](#_Toc106620904) *[get\_rfm\_definition(self, rfm\_def\_path:str)](#_Toc106620904)* [7](#_Toc106620904)

[](#_Toc106620905) *[set\_rfm\_definition(self, rfm\_new:object, rfm\_original:object)](#_Toc106620905)* [7](#_Toc106620905)

*[RFMAnalyzer](#_Toc106620906)* [7](#_Toc106620906)

[](#_Toc106620907) *[config\_connectors(self, data\_path:str, data\_type:str, feature\_cols:str = [], target\_cols:str = [], rfm\_def\_path:str=None,dir\_to\_save\_result:str=None )](#_Toc106620907)* [7](#_Toc106620907)

[](#_Toc106620908) *[rfm\_analyzer(self)](#_Toc106620908)* [8](#_Toc106620908)

[7. Conclusion 8](#_Toc106620909)

# Introduction

RFM analysis allows eCommerce brands to segment and rank their customers by value over a time period, using three dimensions: Recency (how recently someone purchased), Frequency (how frequently someone purchased), and Monetary (how much they spent). In short, RFM analysis shows who are the best and worst customers are​​—and everyone in between.

# Prerequisites

The data which is to be trained and predicted in this module should follow some necessary formats. This module at present can only read csv files. The dataset should have the following features,

1. order date
2. revenue
3. order id
4. customer id

Sample input dataset is shown below,

|  |  |  |  |
| --- | --- | --- | --- |
| order\_id | grand\_total | cust\_uid | date\_formatted |
| JTI001022908 | 40 | 459cc24d24ceb3cd0f15fd2d17d90b54 | 10/27/2021 |
| JTI001015563 | 40 | e5f94d3506c6a6bb4819f7b9109a15c03 | 7/21/2021 |
| JTI001032425 | 40 | e638111eee63e8a2b45e44a1f7587aa67 | 1/25/2022 |
| JTI001012962 | 40 | fcc4821189a41aec45636941dc206f73c3 | 6/21/2021 |
| JTI001017771 | 40 | 4034e13da5e456fe2bb5474649a990b96 | 8/17/2021 |
| JTI001026458 | 40 | a0bd89d39536557f3a92d7150f7ed8b963 | 11/26/2021 |
| JTI001020721 | 40 | 938fa75e424e467a282c0c079de06f4183 | 10/1/2021 |
| JTI001007835 | 40 | bad2227be51957e8c1e88f946db3bcb38c | 4/14/2021 |
| JTI001019733 | 40 | 750bb86a57e00d060fcdfea197a0e2f317e | 9/22/2021 |
| JTI001032194 | 40 | 1be3833e4d50599285f978d5f03ee9137 | 1/22/2022 |
| JTI001019085 | 40 | 7e8f8cd140d0e02cf43703d61fe2816b04 | 9/9/2021 |
| JTI001009440 | 40 | 2893aa7898673cb7a8a5c25a9c7dea2b5f | 4/27/2021 |
| JTI001019220 | 40 | 5408751dba3514cacd63a856001df3b3c6 | 9/11/2021 |
| JTI001013043 | 40 | 5a09cc8b0a7cbc28fd20d96808013a3f77 | 6/23/2021 |

# Config

## *RFM Definition*

Recency, Frequency and Monetary – these three dimensions are key, as they correspond to core pillars of customer behaviour.

* **Recency:** Customers who have bought more recently are more likely to respond to marketing content, which means they may be more likely to take advantage of an offer or read more about your brand and its products.
* **Frequency:** Higher repurchase rate is a reliable indicator of customer enthusiasm and engagement.
* **Monetary:** Customer segments based on amount spent allows you to understand which customers are bigger spenders than others.

## *Assumptions*

* **No of days:** Time period in which the RFM analysis to be performed. In current module it is taken as 365 days.
* **Recency groups and labels:** Recency is divided into three groups as described below,
  + **R1:** Purchased between 0-60 days
  + **R2:** Purchased between 60-180 days
  + **R3:** Purchased between 180-365 days
* **Frequency groups and labels:** Frequency is divided into three groups as described below,
  + **F1:** Order count is greater than 4
  + **F2:** Order count is between 2 to 3
  + **F3:** Order count is 1
* **Quantile range & monetary labels:** Monetary is divided into three groups as described below,
  + **M1:** High, AOV is greater than quantile 0.8
  + **M2:** Medium, AOV is between 0.5 to 0.8 quantile
  + **M3:** Low, AOV is below 0.5 quantile.

## *Customer groups & user input*

Based on above assumptions customer groups are defined as follows,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **rfm\_group** | **customer\_group** | **recency** | **frequency** | **monetary** |
| R1-F3-M1 | Recent\_One\_timers\_High | Days\_lte\_60 | order\_count\_1 | High |
| R1-F3-M2 | Recent\_One\_timers\_Medium | Days\_lte\_60 | order\_count\_1 | Medium |
| R1-F3-M3 | Recent\_One\_timers\_Low | Days\_lte\_60 | order\_count\_1 | Low |
| R1-F2-M1 | Loyal\_regular\_consumers\_High | Days\_lte\_60 | order\_count\_2\_to\_3 | High |
| R1-F2-M2 | Loyal\_regular\_consumers\_Medium | Days\_lte\_60 | order\_count\_2\_to\_3 | Medium |
| R1-F2-M3 | Loyal\_regular\_consumers\_Low | Days\_lte\_60 | order\_count\_2\_to\_3 | Low |
| R1-F1-M1 | Loyal\_regular\_consumers\_High | Days\_lte\_60 | order\_count\_gte\_4 | High |
| R1-F1-M2 | Loyal\_regular\_consumers\_Medium | Days\_lte\_60 | order\_count\_gte\_4 | Medium |
| R1-F1-M3 | Loyal\_regular\_consumers\_Low | Days\_lte\_60 | order\_count\_gte\_4 | Low |
| R2-F3-M1 | One\_timers\_at\_risk\_High | Days\_61\_to\_lte\_180 | order\_count\_1 | High |
| R2-F3-M2 | One\_timers\_at\_risk\_Medium | Days\_61\_to\_lte\_180 | order\_count\_1 | Medium |
| R2-F3-M3 | One\_timers\_at\_risk\_Low | Days\_61\_to\_lte\_180 | order\_count\_1 | Low |
| R2-F2-M1 | Ex-Loyal\_regular\_consumers\_High | Days\_61\_to\_lte\_180 | order\_count\_2\_to\_3 | High |
| R2-F2-M2 | Ex-Loyal\_regular\_consumers\_Medium | Days\_61\_to\_lte\_180 | order\_count\_2\_to\_3 | Medium |
| R2-F2-M3 | Ex-Loyal\_regular\_consumers\_Low | Days\_61\_to\_lte\_180 | order\_count\_2\_to\_3 | Low |
| R2-F1-M1 | Ex-Loyal\_regular\_consumers\_High | Days\_61\_to\_lte\_180 | order\_count\_gte\_4 | High |
| R2-F1-M2 | Ex-Loyal\_regular\_consumers\_Medium | Days\_61\_to\_lte\_180 | order\_count\_gte\_4 | Medium |
| R2-F1-M3 | Ex-Loyal\_regular\_consumers\_Low | Days\_61\_to\_lte\_180 | order\_count\_gte\_4 | Low |
| R3-F3-M1 | One-timers\_Churners\_High | Days\_181\_to\_lte\_365 | order\_count\_1 | High |
| R3-F3-M2 | One-timers\_Churners\_Medium | Days\_181\_to\_lte\_365 | order\_count\_1 | Medium |
| R3-F3-M3 | One-timers\_Churners\_Low | Days\_181\_to\_lte\_365 | order\_count\_1 | Low |
| R3-F2-M1 | Ex-Loyal\_regular\_consumers\_High | Days\_181\_to\_lte\_365 | order\_count\_2\_to\_3 | High |
| R3-F2-M2 | Ex-Loyal\_regular\_consumers\_Medium | Days\_181\_to\_lte\_365 | order\_count\_2\_to\_3 | Medium |
| R3-F2-M3 | Ex-Loyal\_regular\_consumers\_Low | Days\_181\_to\_lte\_365 | order\_count\_2\_to\_3 | Low |
| R3-F1-M1 | Ex-Loyal\_regular\_consumers\_High | Days\_181\_to\_lte\_365 | order\_count\_gte\_4 | High |
| R3-F1-M2 | Ex-Loyal\_regular\_consumers\_Medium | Days\_181\_to\_lte\_365 | order\_count\_gte\_4 | Medium |
| R3-F1-M3 | Ex-Loyal\_regular\_consumers\_Low | Days\_181\_to\_lte\_365 | order\_count\_gte\_4 | Low |

User can change the ‘customer\_group’ in RFM definition as per input data set. The modified RFM definition table will be updated once user finalizes the changes.

# Data connectors

* **Data path:** Path of sales order dataset.
* **Data type:** Type of the data file. Currently, only csv is supported.
* **Feature columns:** List of columns which are required from the sales order dataset.
* **RFM definition path:** RFM definition table path.
* **Directory to save result:** Path where the result file is to be written.

# RFM Analyzer & sample output

RFM analyzer module is responsible for customer segmentations based on the config parameters. This module first puts each customer in respective recency, frequency and monetary buckets depending on their purchase pattern. Once it is done then with the help of RFM definition table each customer is assigned to a customer group.

The sample output is as follows,

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **cust\_uid** | **last\_order\_date** | **order\_cnt** | **revenue** | **aov** | **selected\_date** | **date\_diff** | **Customer\_group** |
| 001ce533cc4270df4478f05d50ca3a41 | 27-02-2022 | 4 | 193 | 48.2 | 15-03-2022 | 16 | Loyal\_regular\_consumers\_High |
| 0020c343272df025fb4ce2744ff49d4a | 15-04-2021 | 1 | 20 | 20 | 15-03-2022 | 334 | One-timers\_Churners\_Low |
| 00381ae80a7611c61a09711f46a643d | 26-04-2021 | 1 | 20 | 20 | 15-03-2022 | 323 | One-timers\_Churners\_Low |
| 0045da72e83284f5d55d1422322865d | 02-07-2021 | 1 | 40 | 40 | 15-03-2022 | 256 | One-timers\_Churners\_High |
| 005b0cc2999217eeb94ecf8c98d441eb | 30-06-2021 | 1 | 15 | 15 | 15-03-2022 | 258 | One-timers\_Churners\_Low |
| 005bd7c6b4eb934302c9d0cf23c99a4f | 14-11-2021 | 1 | 29 | 29 | 15-03-2022 | 121 | One\_timers\_at\_risk\_Medium |
| 005f39431a4a9e9706eb00a68570937 | 10-08-2021 | 1 | 15 | 15 | 15-03-2022 | 217 | One-timers\_Churners\_Low |
| 0078e480fe6c0e0231e2c2d340f6bf60 | 05-03-2022 | 1 | 20 | 20 | 15-03-2022 | 10 | Recent\_One\_timers\_Low |
| 0092fc780cce37ba002829175758fda1 | 01-02-2022 | 1 | 20 | 20 | 15-03-2022 | 42 | Recent\_One\_timers\_Low |
| 00a055e2ec70eb2545480986bf5e216 | 30-05-2021 | 1 | 20 | 20 | 15-03-2022 | 289 | One-timers\_Churners\_Low |
| 00a1643ddc8de9423fddb5df2a17a15 | 07-10-2021 | 1 | 15 | 15 | 15-03-2022 | 159 | One\_timers\_at\_risk\_Low |
| 00b5ef6a577ed831ecd26f81c1e44ecf | 28-10-2021 | 1 | 20 | 20 | 15-03-2022 | 138 | One\_timers\_at\_risk\_Low |
| 00bfad6ebf4759857a93b21bed27095 | 12-01-2022 | 2 | 69 | 34.5 | 15-03-2022 | 62 | Ex-Loyal\_regular\_consumers\_Medium |
| 00cf1c8cc86e8ae28cb9a528419c4a7c | 11-03-2022 | 2 | 57.4 | 28.7 | 15-03-2022 | 4 | Loyal\_regular\_consumers\_Medium |
| 00fe08ddebec85c84decd21eea8e1a6 | 05-11-2021 | 1 | 29 | 29 | 15-03-2022 | 130 | One\_timers\_at\_risk\_Medium |
| 010576eaa84edf7467607017a14c727 | 30-06-2021 | 1 | 20 | 20 | 15-03-2022 | 258 | One-timers\_Churners\_Low |
| 010e43da5929719065aa053a83c21b9 | 12-08-2021 | 2 | 47 | 23.5 | 15-03-2022 | 215 | Ex-Loyal\_regular\_consumers\_Low |

# Code Module

## *Folder structure*

## *RFMAnalyzerConfig*

This class belongs to ‘\rfm\cfg\config.py’. The details of the functions under this class are as follows,

### *get\_rfm\_configs(self)*

Returns a dictionary of RFM config parameters

### *get\_rfm\_group\_details(self)*

Returns a dictionary of RFM group parameters

## *RFMDefinition*

This class belongs to ‘\rfm\src\rfm\_definition.py’. The details of the functions under this class are as follows,

### *get\_rfm\_definition(self, rfm\_def\_path:str)*

Returns a dataframe consists of RFM definitions.

### *set\_rfm\_definition(self, rfm\_new:object, rfm\_original:object)*

This function updates the RFM definition based on user input. This update only happens if user changes customer groups column.

## *RFMAnalyzer*

This class belongs to ‘\rfm\src\rfm\_analyzer.py’. The details of the functions under this class are as follows,

### *config\_connectors(self, data\_path:str, data\_type:str, feature\_cols:str = [], target\_cols:str = [], rfm\_def\_path:str=None,dir\_to\_save\_result:str=None )*

This function sets all the data connector parameters passed from user.

### *rfm\_analyzer(self)*

This function first reads the sales order data and extracts the relevant columns. Then based on each customer’s purchase pattern it puts customers to different customer groups. Finally, it writes the result in the given folder path.

# Conclusion

This module can further be used to make strategic choices in the business. It engages marketers to rapidly distinguish and segment customers into similar clusters and target them with separated and personalized promoting methodologies