



Think Ahead



Dataset overview: My Grosto

Machine learning with Python
for finance professionals

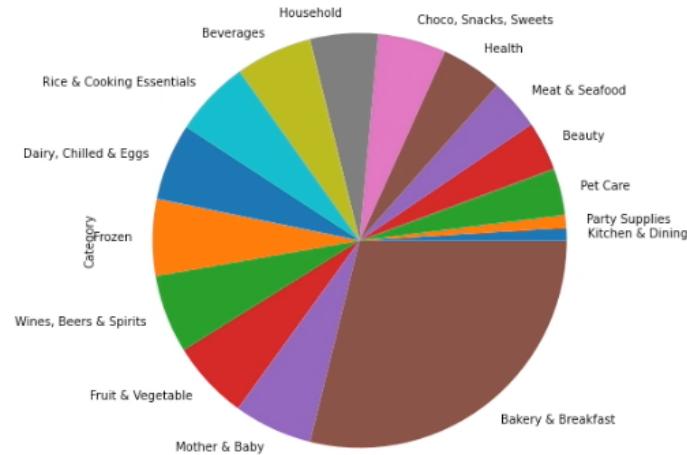
Case Study – My Grosto – Factsheet

❑ **Name / location / industry:** *My Grosto, Singapore, grocery retailer*

❑ **Market:** 5 cities across Singapore (Bedok & Woodlands), Indonesia (Jakarta), Malaysia (Kuala Lumpur) and Philippines (Manila)

❑ **Stats:**

- 20K+ products across 16 different product categories
- In 2019, gross sales of SGD 200K with 22% YoY growth
- In last four years, sold 78K+ products to 2.6K+ customers, net sales of SGD 600K, 50K+ transactions
- The company has 45 full-time employees
- Bedok store averages 40 customers/day; average age is 36 years; 80% are female.
- Average customer visits 4 times/month and purchases 6 products with average billing of 41 SGD



Case Study – My Grosto – Data

The database provided is a consolidation of My Grosto's POS and Finance systems.

- Attributes:** Receipt details, location, customer demography, date/time, product information, price, payment details, and more.
- Size:** 50K records from 5 stores across 4 countries. Each record maps to one of 13K receipts.

Column Name
Membership ID
Country
City
Age
Gender
QTY
Frequency
Monetary
Recency (Days)
GST
Sum of Gross Bill Amount

We will initially be working with this aggregated dataset containing the following customer sales metrics for each customer:

- Membership ID** to identify each customer
- Recency** is the number of days since the customer was last seen
- Frequency** is the time interval between visits
- Monetary** is the total net sales per customer

Column Name	Description of the Column names
Receipt Number	Transaction Number
Date	Date of Transaction
Year	Year of Transaction
Month	Month of Transaction
Time	Time of the transaction
Mobile Number	Mobile number of the customer
Membership ID	Unique Membership ID of the Customer
Loyalty Card Points	Loyalty Points earned by customer (10 Points/100 SGD Spent)
Age	Age of the Customer in the Company Database
Gender	Gender of the customer
City	City Name of store
Country	Country Name of store
Category	Category of item bought by customer
Sub_Category	Sub-Category of item bought by customer
Items	Name if the item bought by customer
Brand	Brand of the item bought
Description	Description of the item (weight/volume/pack size/unit)
Price	Price per unit of the item
QTY	Quantity bought by the customer
DISC	Discount offered by store
Amount	Discounted Price
Net Bill Amount	Amount exclusive of tax payable to customer
GST	Tax Payable
Gross Bill Amount	Gross Bill to be paid by customer inclusive of tax
Payment Mode	Mode of payment opted by customer
Bank Name	Name of the Bank
% Profit Margin	Percentage profit margin of store
% Operating Cost	Percentage operating cost of store
% Product Cost	Percentage product cost to store
Profit Margin	Profit Margin earned by store
Operating Cost	Operating cost borne by the store
Product Cost	Product Cost for the store

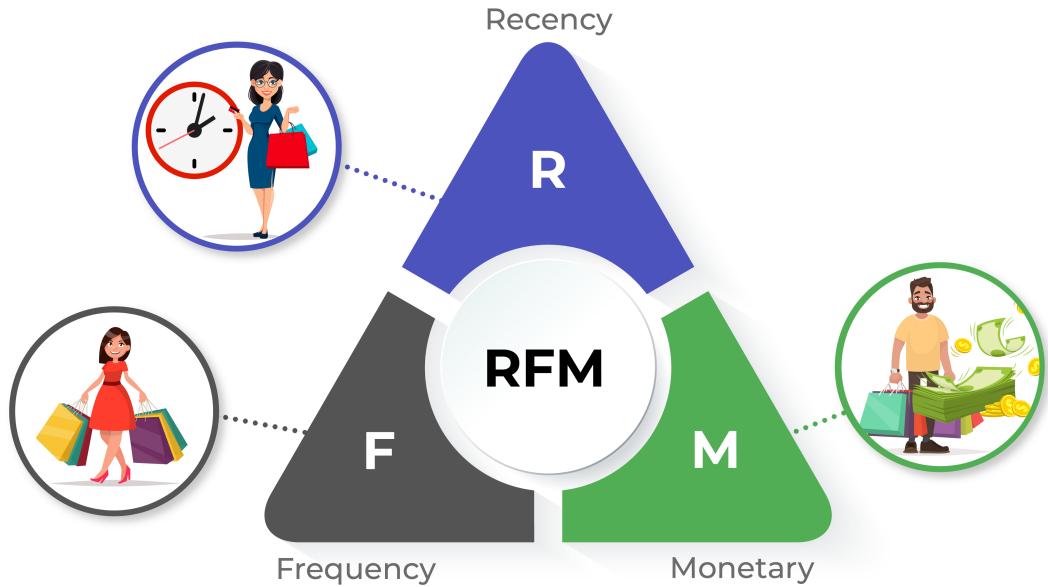
Case Study – My Grosto – Problem Statement

- *My Grosto* is a small Singapore based retailer looking to expand their operations by opening 4 new stores in the existing countries, expand into 5 other South East Asian countries, and IPO on the Singapore Exchange to raise capital.
- The ability to understand and predict customer behaviour is key to commercial success. One way that data analysis and machine learning can help is by creating effective marketing and promotional campaigns.
- In this context, a Recency-Frequency-Monetary (RFM) model can help to create a marketing strategy by suggesting optimized target groups for different promotional campaigns.



What is RFM Analysis?

- Recency, Frequency & Monetary value (RFM) analysis allows you to see how recently, frequently and how much a consumer has spent within a business.
- Although it has been used since the 1970's, it's one of the best techniques for performing market segmentation.
- For example, ideal customers have **low recency, high frequency** and **high monetary value**.



Why use RFM analysis?



RFM analysis provides information to marketers for:

- Creating **personalized marketing** campaigns
- Increasing **customer engagement**
- **Cohort analysis** for increasing customer retention



Specifically for company growth, RFM analysis allows you to **cluster consumers into groups based on their behavior**, improving strategic campaigns for business expansion



Other techniques are available, such as **clustering algorithms** that can group customers based on **demographics** such as age, as well as spending patterns. Python's **scikit-learn** library comes with a variety of clustering algorithms.

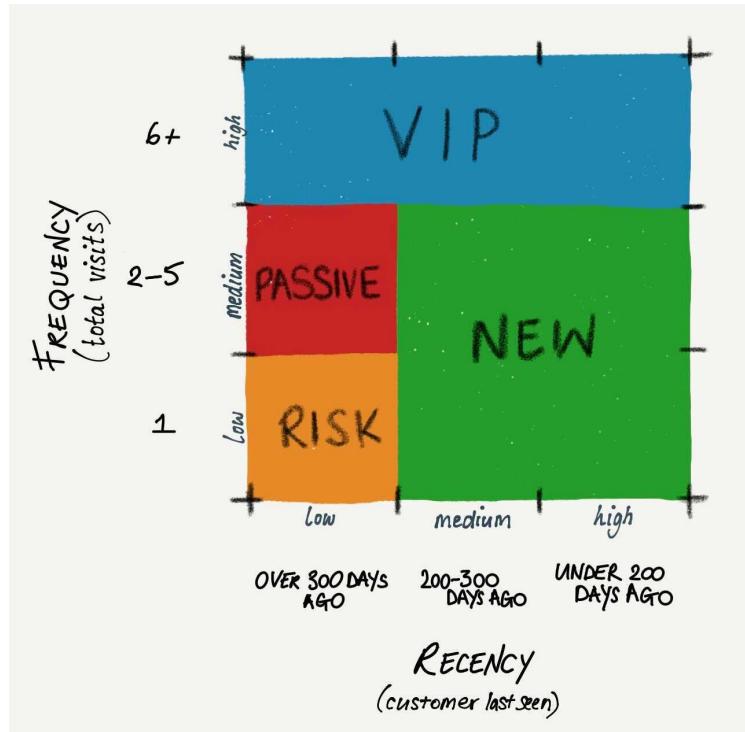
Breakdown of RFM

In our dataset we have calculated the following 3 fields:

- Recency**: when was this customer last seen?
- Frequency**: number of visits?
- Monetary**: total sales generated from each customer.

The transactions which observe **low recency**, **high frequency** & **high monetary value** suggest most fruitful customers.

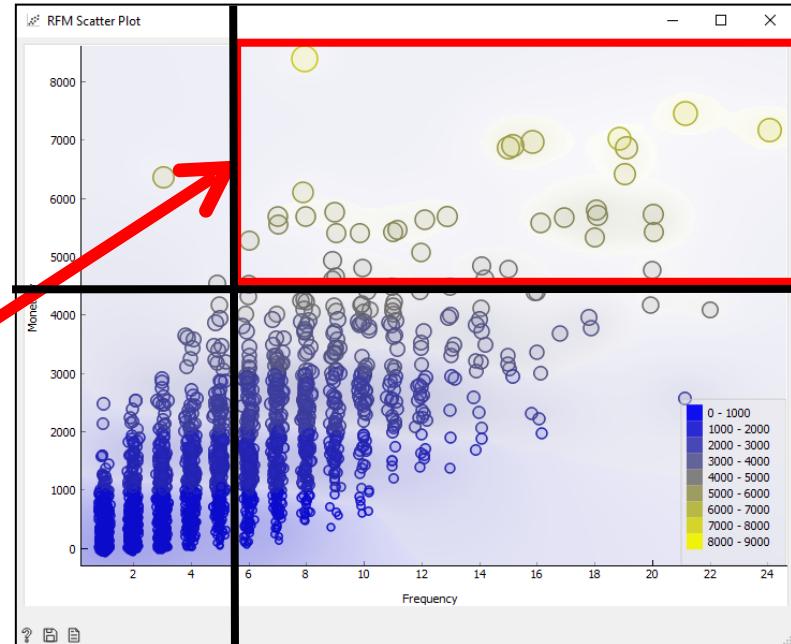
- VIP** customers are recent customers with the strongest rate of high value engagement.
- New** customers are recent but too new to develop a high frequency – the goal is to shift these towards VIP.
- Passive** customers are potentially at risk by showing inactivity.
- Risk** customers are inactive and possibly already lost.



Monetary vs Frequency

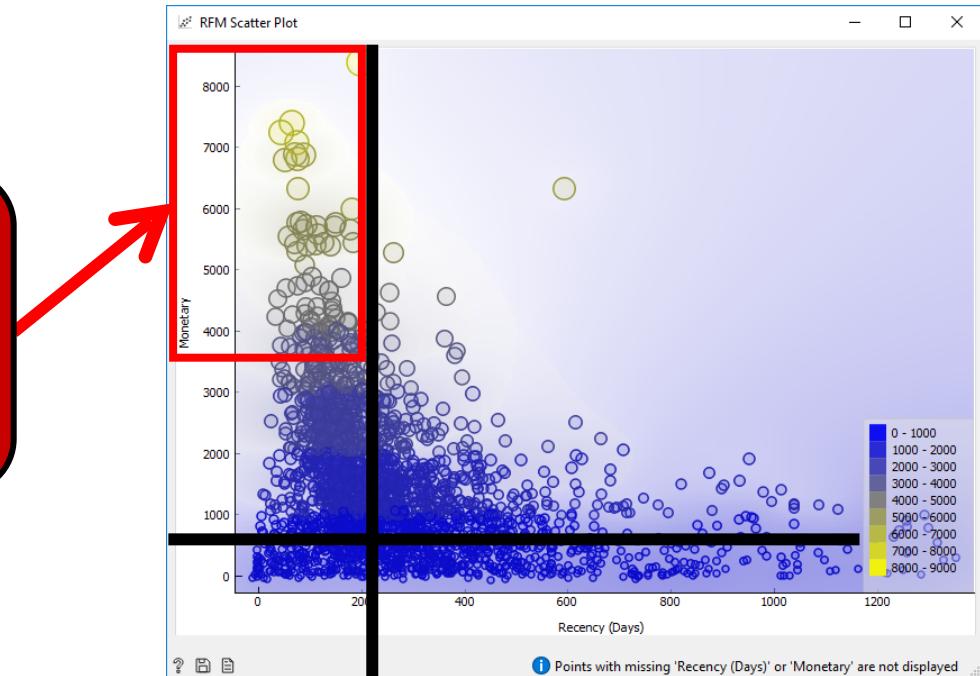
Consumers with a high frequency & high monetary value are the most profitable and should be targeted for the greater benefits.

We can cross-reference their Membership ID in order to communicate offers & promotions directly through company media communications.



Monetary vs Recency

The consumers with low recency & high monetary value are the most profitable. These can also be targeted for the greater benefits.



Recency vs Frequency

Consumers with high monetary value need to be evaluated for each specific store to target them with great offers.

