



Customer Segmentation and Coupon generating system

Submitted by **Pranjal Ambulkar - 01 Naziya Inamdar - 24 Khushnuma Shaikh - 56**

Under the guidance of
ANKITA NAGMOTE

**USHA MITTAL INSTITUTE OF TECHNOLOGY S. N. D. T.WOMEN'S
UNIVERSITY**

Abstract

Customer segmentation is one of the key methods in marketing analytics and has been used to segment customers on various criteria and drive business results. One of the techniques used to segmenting the customers basis the behaviour they have exhibited in the past is RFM Analysis. RFM is an acronym for Recency, Frequency and Monetary value. With the increasing availability of past transactional data, RFM analysis can be effectively used to segment customers and drive subsequent business actions. This study performs customer segmentation on past transactional data using K-Means clustering algorithm in Python and basis the created segments, recommended course of actions is suggested.

Promotional codes are alphanumeric strings that online stores offer to encourage purchases on their website and are typically associated with an overarching promotional marketing strategy. Generating an automatic coupon generator will generate custom coupons for every group of customer according to the need. The parameters considered will help gaining more profit for the company.

Research Through Innovation

Chapter 1 Introduction

Understanding people's needs, behaviors, and expectations is key to communicating and connecting with them, promotional strategies. Rather than viewing potential and existing customers as a group, businesses should take the time to get to know them as individual groups with distinct perspectives

The segmentation process begins by grouping customers and potential customers into customer segments with similar characteristics, so that you can communicate to all of the individuals in that segment efficiently, effectively, and with a sense of personal attention, without contacting each individual individually. As a result, your marketing tactics will become more effective and efficient, saving you time and money while also increasing the benefits.

Segmentation helps businesses customize their marketing strategies to provide shoppers with the brand experience they're looking for. This can range from sending targeted email campaigns to designing in-store displays to serving personalized digital ads. As a result, customers feel more connected and engaged. Personalization can lead to customer loyalty and brand advocacy, especially when companies nurture the business-consumer relationship in the right ways.

When a store issues promotional codes, they're providing customers with an incentive to buy, which benefits both the customer and the business. These incentives even have a direct impact on the overall shopping experience, according to research conducted by the Center for Neuroeconomics Studies at Claremont Graduate University. The discount can either be a percentage or a specific dollar amount. Promo codes also provide customers with free shipping or gift-wrapping. This marketing strategy essentially gives customers yet another reason to buy your products.

Hence along with a successful marketing strategy of generating coupons for customer this system generates specialised custom coupons for customers. It analyses the database containing customer's purchase record and answers according to the database after performing RFM analyses special coupons are generated for specific group of customer.

1.1 Problem Statement

Any company in retail, no matter the industry, ends up collecting, creating, and manipulating data over the course of their lifespan. These data are produced and recorded in a variety of contexts, for example number of orders, amount of products purchased by an individual and so on. The wider one's access to data, the clearer the picture that the data can portray. With a clear picture created from data, previously unseen details begin to emerge, stimulating fresh perspectives and innovations.

It is crucial for them to have a sense of who their customers are. Not just the products they like to purchase, but when they like to purchase them, how often they want to purchase them, and what their lifetime value may be to the company. While some of these questions are more straight forward than others, it is clear that they all require data munging, analysis, and presentation that involve skills and techniques beyond what is required of a traditional analyst.

By integrating machine learning practices and conventional business understandings, the paths to answering these questions became more intertwined with that of a similar question: What segments or groups of customers do we have? After studying clustering and reading about it in numerous other contexts, it be-

came clear that segmenting retail customers became one way to investigate the purchasing patterns and behaviors of its customers.

Companies need to advance their marketing strategies with time. Selling products is not sufficient, customer must be attracted with various methods. Customers who are loyal to your brand should be rewarded for their good behaviour. Once the customer engagement is lost re-engagement must be initiated. Further more the current strategy of sending cold coupons, promo-code and offers to the customer is not strategic.

1.2 Existing System

The collection of three values for each customer is called an RFM cell. In a simple system, organizations average these values together, then sort customers from highest to lowest to find the most valuable customers. Some businesses, instead of simply averaging the three values, weigh the values differently. Further in this project shows how we can segment customers into different clusters using the K-means algorithm. But, we would be implementing RFM analysis to get the desired values and those features will be used as an input in K-means, to determine similarity.

1.3 Proposed System

In the Existing system, average of these values together are sorted into customers from highest to lowest to find the most valuable customers. This data is further utilized in generating an automatic coupon code generator which generates coupons on the basis of analysis extracted from RFM of the database.

When a store issues promotional codes, they're providing customers with an incentive to buy, which benefits both the customer and the business. These incentives even have a direct impact on the overall shopping experience, according to research conducted by the Center for Neuroeconomics Studies at Claremont Graduate University. The discount can either be a percentage or a specific dollar amount. Promo codes also provide customers with free shipping or gift-wrapping. This marketing strategy essentially gives customers yet another reason to buy your products.

Hence along with a successful marketing strategy of generating coupons for customer this system generates specialised custom coupons for customers. It analyses the database containing customer's purchase record and according to the database after performing RFM analyses special coupons are generated for specific group of customer.

Chapter 2

Review of Literature

According to Blattberg, R.C., Kim, B.D., Neslin, S.A. (2008). RFM Analysis. In: Database Marketing. International Series in Quantitative Marketing, vol 18. Springer, New York, NY. https://doi.org/10.1007/978-0-387-72579-6_12

Abstract Recency (R), Frequency (F), and Monetary Value (M) are the most popular database marketing metrics used to quantify customer transaction history. Recency is how recently the customer has purchases; frequency is how often the customer purchases, and monetary value is the dollar value of the purchases. RFM analysis classifies customers into groups according to their RFM measures, and relates these classifications to behaviors such as the likelihood of responding to a catalog or other offer. RFM analysis was probably the first “predictive model” used in database marketing.

How do you select customers for target mailing? Or whom should you send your catalogs or direct mail offers to? The need to mail smarter is always among the top concerns of direct marketers. The direct mail promotion that results in sales to 2 Percent of the mailed universe is considered a success. Identifying and targeting the customers who are most likely to respond are therefore of prime concern.

Because of the nature of their businesses, direct marketers including catalogers have been collecting customer data, analyzing them, and developing models for several decades to improve their business performance. One popular approach used to improve mailing efficiency is the RFM – Recency, Frequency, Monetary

amount – model. The primitive form of the RFM model was used about 50 years ago by catalogers of general merchandise. For example, as early as 1961, George Cullinan promoted the use and understanding of RFM customer data analysis. Recognizing his contribution in advancing the direct marketing industry, the DMA (Direct Marketing Association) inducted him into the DMA Hall of Fame in 1989. The core concept of the RFM model is based on the empirical evidence. Direct marketers have found that the response to a mailing offer is heterogeneous across customers. And they also found that customers who have responded well in the past are likely to respond in the future. More specifically, direct marketers have found that customers’ purchase response can be predicted using their previous purchase histories. The three most important variables to summarize customers’ purchase histories are recency (R), frequency (F), and monetary amount (M). That is, using RFM measures for each customer, one can predict his or her propensity to respond. Once identifying who is going to respond, the direct marketer sends catalogs to customers with high propensity.

It encounters several scoring challenges in the measure of frequency and is relatively sensitive, which leads to pulling apart customers who have identical behavior at the lower quintiles, but group customers together whose buying behaviors have significant differences (Alam and Khalifa, 2009).

There is another scoring method (behavior quintile scoring method) developed by John Wirth (The founder of Woodworker's Supply of New Mexico), which sorts customers based on their behavior and thus may have different number of customers in each quintile. The scoring scheme of frequency covers five intervals, including 0 to 3 months, 4 to 6 months, 7 to 12 months, 13 to 24 months and 25+ months, which are coded as 5, 4, 3, 2 and 1, respectively. This method is known as hard coding (McCarty and Hastak, 2007). For frequency, although this scoring method appears to solve the sensitivity problems, it still encounters similar problems as customer quintile method in the frequency measure.

Chapter 3 Methodology Planning

Methodology and planning is a systematic approach to designing, executing and delivering a project. It is a clearly defined combination of related practices, methods, and processes which determine how to best execute the project.

A summarized research and planning spread light upon the required tech stack and various algorithms, modules to be taken under consideration.

3.1 Hardware Software requirement

Hardware:

- Processor: Intel core i3 6th gen onwards
- RAM: 2 GB or higher
- Hard Disk Space: 20 GB or higher :Software
- Python

3.2 Algorithm

Step 1

We will start with a dataset containing information about individual orders performed on a platform. The data is represented in a transaction level i.e. each row of the dataset contains features associated with a single transaction, such as date, time, payment method, client id of the user that made the purchase, etc.

<i>order_date</i>	<i>order_id</i>	<i>customer</i>	<i>grand_total</i>
2022-01-01	IN-52653	Khushi	90
2022-01-02	IN-52654	Naziya	150
2022-01-03	IN-52655	Pranjal	70

Table 3.1: CSV file with all the orders

Step 2

The next step in building an RFM model is to assign Recency, Frequency and Monetary values to each customer

- Recency is simply the amount of time since the customer's most recent trans- action (most businesses use days, though for others it might make sense touse months, weeks or even hours instead).
- Frequency is the total number of transactions made by the customer (duringa defined period).
- Monetary is the total amount that the customer has spent across all trans- actions (during a defined period).

Step 3

The third step is to divide the customer list into tiered groups for each of thethree dimensions (R, F and M).

This results in 64 distinct customer a (4x4x4), into which customers will be segmented. Three tiers can also be used (resulting in 27 segments); using more

Recency	Frequency	Monetary
R-Tier-1 (most recent)	F-Tier-1 (most frequent)	M-Tier-1 (highest spend)
R-Tier-2	F-Tier-2	M-Tier-2
R-Tier-3	F-Tier-3	M-Tier-3
R-Tier-4 (least recent)	F-Tier-4 (only one transaction)	M-Tier-4 (lowest spend)

Table 3.2: RFM Segmentation

than four, however, is not recommended (because the difficulty in use outweighsthe small benefit gain from the extra granularity).

Step 4

The fourth step is to select groups of customers to whom specific types of communications will be sent, based on the RFM segments in which they appear.

It is helpful to assign names to segments of interest. Here are just a few examples to illustrate:

- **Best Customers** – This group consists of those customers who are found in R-Tier-1, F-Tier-1 and M-Tier-1, meaning that they transacted recently, do so often and spend more than other customers. A shortened notation for this segment is 1-1-1; we'll use this notation going forward.
- **High-spending New Customers** – This group consists of those customers in 1-4-1 and 1-4-2. These are customers who transacted only once, but very recently and they spent a lot.
- **Lowest-Spending Active Loyal Customers** – This group consists of those cus-tomers in segments 1-1-3 and 1-1-4 (they transacted recently and do so often, but spend the least).
- **Churned Best Customers** – This segment consists of those customers in groups 4-1-1, 4-1-2, 4-2-1 and 4-2-2 (they transacted frequently and spent alot, but it's been a long time since they've transacted).

Marketers should assemble groups of customers most relevant for their particular business objectives and retention goals.

Step 5

The fifth step actually goes beyond the RFM segmentation itself: crafting specific messaging that is tailored for each customer group. By focusing on the behavioral patterns of particular groups, RFM marketing allows marketers to communicate with customers in a much more effective manner.

Again, here are just some examples for illustration, using the groups we named above:

- 1. Core - Your Best Customers RFM Score: 111

Who They Are: Highly engaged customers who have bought the most recent, the most often, and generated the most revenue.

Marketing Strategies: Focus on loyalty programs and new product introductions. These customers have proven to have a higher willingness to pay, so don't use discount pricing to generate incremental sales. Instead, focus on value added offers through product recommendations based on previous purchases.

- 2. Loyal - Your Most Loyal Customers RFM Score: X1X

Who They Are: Customers who buy the most often from your store. Marketing Strategies: Loyalty programs are effective for these repeat visitors.

Advocacy programs and reviews are also common X1X strategies. Lastly, consider rewarding these customers with Free Shipping or other like benefits.

- 3. Whales - Your Highest Paying Customers RFM Score: XX1

Who They Are: Customers who have generated the most revenue for your store.

Marketing Strategies: These customers have demonstrated a high willingness to pay. Consider premium offers, subscription tiers, luxury products, or value add cross/up-sells to increase AOV. Don't waste margin on discounts.

- 4. Promising - Faithful customers RFM Score: X13, X14

Who They Are: Customers who return often, but do not spend a lot.

Marketing Strategies: You've already succeeded in creating loyalty. Focus on increasing monetization through product recommendations based on past purchases and incentives tied to spending thresholds (pegged to your store AOV).

3.3 Code Implementation

▼ RFM Analysis

```
[53] import numpy as np
import pandas as pd
```

```
[54] orders = pd.read_csv('sample-orders.csv',encoding = "ISO-8859-1")
```

```
[55] orders.head()
```

	order_date	order_id	customer	grand_total
0	9/7/11	CA-2011-100006	Dennis Kane	378
1	7/8/11	CA-2011-100090	Ed Braxton	699
2	3/14/11	CA-2011-100293	Neil Französisch	91
3	1/29/11	CA-2011-100328	Jasper Cacioppo	4
4	4/8/11	CA-2011-100363	Jim Mitchum	21

▼ the RFM Table

```
[56] import datetime as dt
NOW = dt.datetime(2014,12,31)
```

```
[57] # Make the date_placed column datetime
orders['order_date'] = pd.to_datetime(orders['order_date'])
```

Create the RFM Table

```
[58] rfmTable = orders.groupby('customer').agg({'order_date': lambda x: (NOW - x.max()).days, # Recency
'order_id': lambda x: len(x), # Frequency
'grand_total': lambda x: x.sum()}) # Monetary Value

rfmTable['order_date'] = rfmTable['order_date'].astype(int)
rfmTable.rename(columns={'order_date': 'recency',
'order_id': 'frequency',
'grand_total': 'monetary_value'}, inplace=True)
```



Validating the RFM Table

```
[59] rfmTable.head()
```

customer	recency	frequency	monetary_value
Aaron Bergman	415	3	887
Aaron Hawkins	12	7	1744
Aaron Smayling	88	7	3050
Adam Bellavance	54	8	7756
Adam Hart	34	10	3249

Customer **Aaron Bergman** has frequency:3, monetary value:\$887 and recency:415 days.

```
[60] aaron = orders[orders['customer']=='Aaron Bergman']
aaron
```

Creating the RFM segmentation table

```
[66] rfmSegmentation = rfmTable
```

We create two classes for the RFM segmentation since, being high recency is bad, while high frequency and monetary value is good.

```
[67] # Arguments (x = value, p = recency, monetary_value, frequency, k = quartiles dict)
def RClass(x,p,d):
    if x <= d[p][0.25]:
        return 1
    elif x <= d[p][0.50]:
        return 2
    elif x <= d[p][0.75]:
        return 3
    else:
        return 4

# Arguments (x = value, p = recency, monetary_value, frequency, k = quartiles dict)
def FMClass(x,p,d):
    if x <= d[p][0.25]:
        return 4
    elif x <= d[p][0.50]:
        return 3
    elif x <= d[p][0.75]:
        return 2
    else:
        return 1
```

International Research Journal
IJNRD
 Research Through Innovation

```
[68] rfmSegmentation['R_Quartile'] = rfmSegmentation['recency'].apply(RClass, args=('recency',quantiles,))
rfmSegmentation['F_Quartile'] = rfmSegmentation['frequency'].apply(FMClass, args=('frequency',quantiles,))
rfmSegmentation['M_Quartile'] = rfmSegmentation['monetary_value'].apply(MClass, args=('monetary_value',quantiles,))

[69] rfmSegmentation['RFMClass'] = rfmSegmentation.R_Quartile.map(str) \
+ rfmSegmentation.F_Quartile.map(str) \
+ rfmSegmentation.M_Quartile.map(str)

rfmSegmentation.head()
```

customer	recency	frequency	monetary_value	R_Quartile	F_Quartile	M_Quartile	RFMClass
Aaron Bergman	415	3	887	4	4	4	444
Aaron Hawkins	12	7	1744	1	2	3	123
Aaron Smayling	88	7	3050	3	2	2	322
Adam Bellavance	54	8	7756	2	2	1	221
Adam Hart	34	10	3249	2	1	2	212

```
# Uncomment any of the following lines to: copy data to clipboard or save it to a csv file.
# rfmSegmentation.to_clipboard()
rfmSegmentation.to_csv('rfm-table.csv',encoding = "ISO-8859-1")
```

DOWNLOAD SEGMENTED DATABASE

```
[ ] from google.colab import files
rfmSegmentation[rfmSegmentation['RFMClass']=='111'].to_csv('cluster_111.csv')
files.download('cluster_111.csv')
```

```
[ ] from google.colab import files
rfmSegmentation[rfmSegmentation['RFMClass']=='112'].to_csv('cluster_112.csv')
files.download('cluster_112.csv')
```

```
[ ] from google.colab import files
rfmSegmentation[rfmSegmentation['RFMClass']=='113'].to_csv('cluster_113.csv')
files.download('cluster_113.csv')
```

Mail System



```
[891] import smtplib, ssl
import csv

import string
import random

port = 587 # For starttls
smtp_server = "smtp.gmail.com"
login = "falangeregina01@gmail.com"
password = "ktqk dznc hepi ease"
code = ''.join(random.choices(string.digits, k=4))

message = """Subject: CONGRATULATIONS !! YOU JUST WON A COUPON
To: {recipient}
From: {sender}

Heyyy Luckyy you !! you just won a coupon for being out loyal customer !!

-----
YSx111
-----

Enter the coupon on checkout to avail your offer!!
"""
sender = "falangeregina01@gmail.com"

context = ssl.create_default_context()
with smtplib.SMTP(smtp_server, port) as server:
    server.connect('smtp.gmail.com', '587')
    server.ehlo() # Can be omitted
    server.starttls(context=context)
    server.ehlo() # Can be omitted
    server.login(login, password)

    with open("/content/cluster_111.csv") as file:
        reader = csv.reader(file)
        next(reader) # It skips the header row
        for email in reader:
            server.sendmail(
                sender,
                email,
                message.format(code = code, recipient=email, sender=sender)
            )
    print(f'Sent to')
```

Email Received

CONGRATULATIONS !! YOU JUST WON A COUPON Inbox x



falangeregina01@gmail.com

Sat, Feb 18, 11:02 PM (4 hours ag)

to ▾

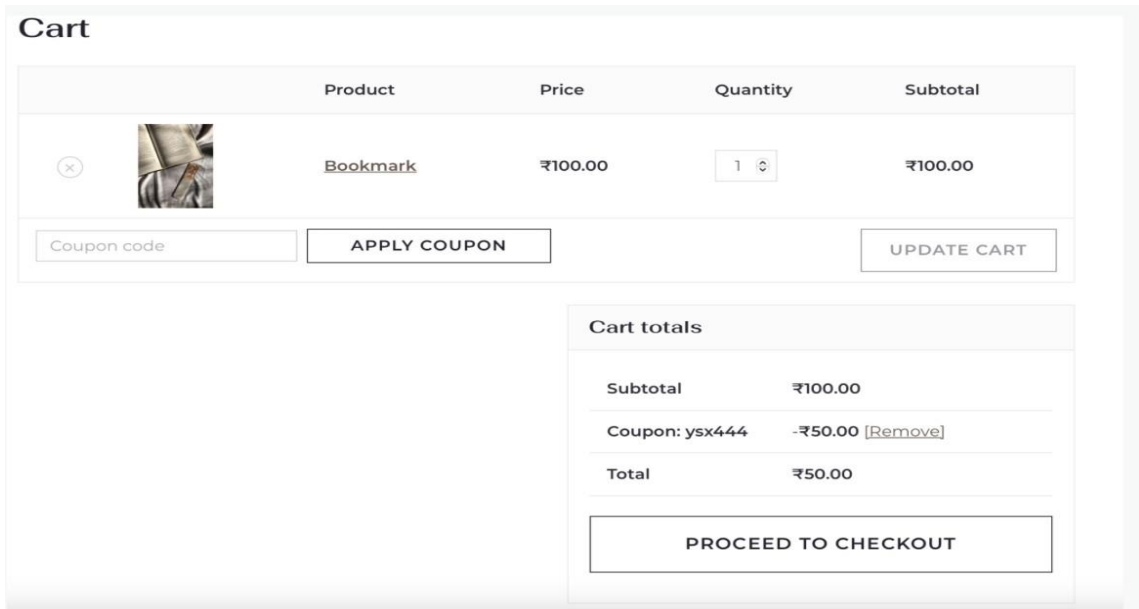
Heyyy Luckyy you !! you just won a coupon for being out loyal customer !!

YSx3576

Enter the coupon on checkout to avail your offer!!

Coupon on website





3.4 Work Flow

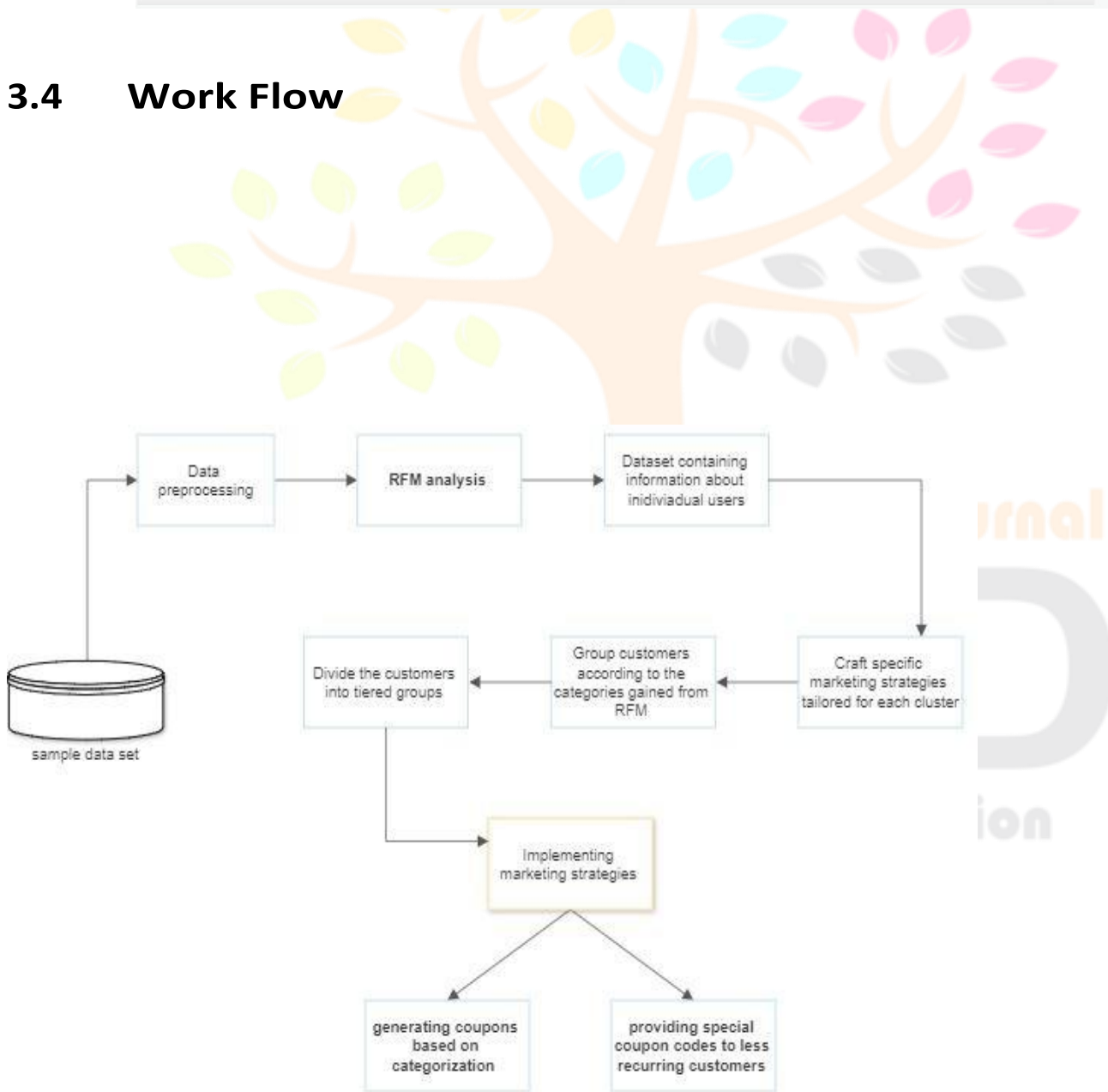


Figure 3.1: Project Roadmap

Chapter 4 Conclusion

The primary goal of this project is to develop a system that categorises the customers on the basis of the recency - that is how recent did they purchase an item, frequency - that is how frequent do they purchase and monetary - that is how much money are they willing to spend. This categorisation allows the business owners to develop unique and personalized marketing scheme for each of the cluster of customer segments. This project makes use of RFM analyses as well as k means clustering.

Furthermore this categorisation is used for generating coupons. These coupons will be generated specifically with respect to customers purchase history and RFM results leading to custom coupons for every group of customers.

Chapter 5 Future scope

A relevant implementation to the current project would be to add a marketing technique that profits the company's revenue. When a store issues promotional codes, they're providing customers with an incentive to buy, which benefits both the customer and the business. The coupons discount can either be a percentage or a specific dollar amount. Promo codes also provide customers with free shipping or gift-wrapping. This marketing strategy essentially gives customers yet another reason to buy your products. These incentives even have a direct impact on the overall shopping experience, according to research conducted by the Center for Neuroeconomics Studies at Claremont Graduate University.

Furthermore promotional codes are an effective way to track your marketing efforts to see which platforms are generating the most traffic or leading to conversions, according to Inc. magazine. Store owners can track their efforts by attaching a certain promo code to a specific marketing campaign or advertisement. Hence creating a database by analysing the user experience with the promotional codes can add more

References

Huang, Y., Zhang, M. and He, Y., 2020, June. Research on improved RFM customer segmentation model based on K-Means algorithm. In 2020 5th International Conference on Computational Intelligence and Applications (ICCIA) (pp. 24-27).

Marcus, C. (1998), 'A practical yet meaningful approach to customer segmentation approach to customer segmentation', *Journal of Consumer Marketing* 15, 494-504.

Sari, Juni Nurma; Nugroho, Lukito Edi; Ferdiana, Ridi; Santosa, P. Insap, 'Review on Customer Segmentation Technique on Ecommerce'

Jo-Ting Wei, Shih-Yen Lin, and Hsin-Hung Wu, "A Review of application of RFM model" (December 2010)

Abirami, M. Pattabiraman, V. (2016). Data mining approach for intelligent customer behavior analysis for a retail store. In: proceedings of the 3rd international symposium on big data and cloud computing challenges (ISBCC-16). 283-291.

Bowen, J. T. and Chen, S. (2001). *International Journal of Contemporary Hospitality Management*,

13(5), 213-217.

Bulut, Z. A. (2015). Determinants of repurchase intention in online shopping: a Turkish consumers' perspective. *International Journal of Business and Social Science*, 6(10), 55-63.

Chen, Y. L., Kuo, M. H., Wu, S. Y. Tang, K. (2009). *Electronic Commerce Research and Applications*, 8(5), 241-251.

Chen, Y. S., Cheng, C. H., Lai, C. J., Hsu, C. Y., Syu H. J. (2012). Identifying patients in target customer segments using a two-stage clustering-classification approach: A hospital-based assessment. *Computers in Biology and Medicine*, 42(2), 213-221.

Cheng, C. H. Chen, Y. S. (2009). *Expert Systems with Applications*, 36(3), 4176-4184.



Cho, Y. S., Moon, S. C., Jeong, S. P., Oh, I. B. Ryu, K. H. (2013). Clustering method using item preference based on RFM for recommendation system in u- commerce. In Ubiquitous Information Technologies and Applications, Springer Netherlands, pp. 353-362.

Coussement, K., Van den Bossche, F. A. M. De Bock, K. W. (2014). Data accuracy's impact on segmentation performance: benchmarking RFM analysis, logistic regression, and decision trees. Journal of Business Research, 67(1), 2751- 2758. Cui, G., Wong, M. L. Lui, H

Acknowledgemnt

I have a great pleasure to express my gratitude to all those who have I would like to express my deepest appreciation to all those who provided me the pos- sibility to complete this report. A special gratitude I give to our project guid, Ms Ankita Nagmote, whose contribution in stimulating suggestions and encour- agement, helped to coordinate project especially in writing this report and has invested her full effort in guiding the team in achieving the goal..

A special thanks goes to the team, Pranjali Ambulkar, Naziya Inamdar, Khush- numa Shaikh , who helped to assemble the project and gave suggestions to improve the project. Last but not least, I have to appreciate the guidance given by other supervisors as well as the panels especially in our project presentation that has improved our presentation skills thanks to their comment and advice.

Finally, we are very thankful to Usha Mittal institute of technology, for givingus a chance to complete our thesis project within the scheduled time.

Date:

Pranjali Ambulkar - 01Naziya Inamdar - 24 Khushnuma Shaikh - 56

