



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 3, March 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.512

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

A Survey on Customer Segmentation System

Girish Kumar Singh¹, Sushil Krishna Patil², S.P.Godse³

U.G. Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, Maharashtra, India¹

U.G. Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, Maharashtra, India²

Assistant Professor, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune,

Maharashtra, India³

ABSTRACT: Customer segmentation is the process of dividing customers into discrete groups based on certain traits such as personality, interests, habits and factors such as demographics, industry, and income. Through customer segmentation analysis, marketers can come up with the right messages, using the right words, to promote their products. Here Recency, Frequency and Monetary variables of customers are used for performing the Customer Segmentation. Customer Segmentation is done in roughly two stages. First is to calculate the RFM score for each customer and second is to create clusters using K means.

KEYWORDS: Customer segmentation, RFM, K-Means, Market, Machine learning, Clustering.

I. INTRODUCTION

Customer Segmentation is a methodology using which we can divide our customer base into group of individuals who are similar in terms of gender, spending behaviour, and frequency to visit, age or other demographics.

Customer Segmentation allows companies to precisely target the customers who have specific needs and desires. This way companies can design the targeted campaigns for the right group or audience, it may also be the case that during the creation of these individual groups or clusters company might identify new market segment on which company can focus more. Machine learning techniques are used to identify the most loyal customers and the customers who are on the verge of churning out.

Clustering is the classification of objects into different groups, or more precisely, the partitioning of a dataset into subset (clusters), so that the data in each subset share some common trait- often according to some defined distance measure.

There are two types of clustering Hierarchical Clustering and Partitional Clustering.

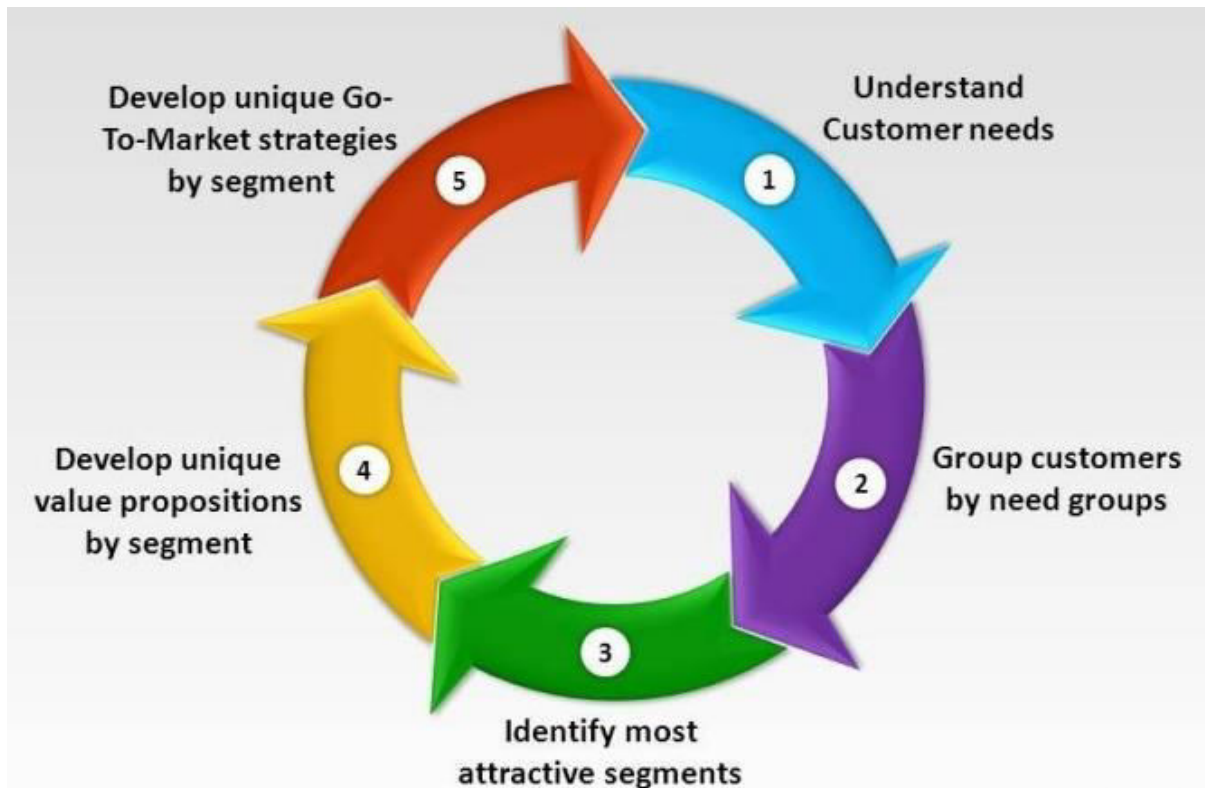
We are going to use K- means clustering algorithm which comes under Partitional Clustering. K- Means clustering is a classical clustering algorithm that uses expectation maximization like technique to partition number of data points into K clusters.

RFM analysis is applied here with Python, exhibiting its simplicity and use of most basic set of information available with purchasing records.

Finally a data score will be generated which will be displayed according to the output the algorithm on the console.

II. RELATED WORK

Smith is often known as the first to introduce the concept of market segregation in marketing books in 1956 with the publication of his article, "*Product Differentiation and Market Segmentation as Alternative Marketing Strategies.*"



1) Steps to Customer Segmentation

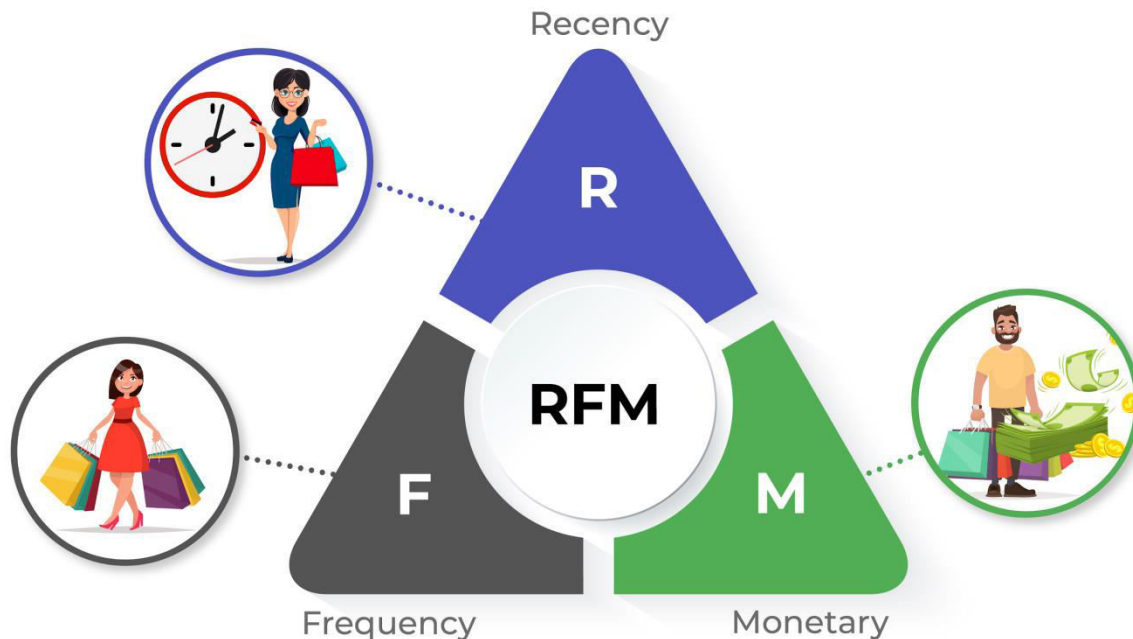
The experimental outputs of many studies show that the agglomerative clustering needs a long processing time for large dataset in comparison with k-means and advanced version of k-means clustering. However one crucial advantage of agglomerative clustering is, it does not require appointing the total number of clusters initially.

The segmentation outcomes of various studies also exhibit that the advanced version of k-means can effectively reduce the total running time by 27.8% and 97.8% compared to standard k-means and agglomerative clustering respectively and so effectively increase the speed of clustering. It can also produce better clustering results than standard k-means in respect of intra cluster distance and inter-cluster distance.

III. METHODOLOGY

When we are provided with raw data from the database, it can be confusing and instructive to look at individual records. RFM analysis is used to present data at an integrated level and is used to classify customers into related groups. It has been accepted in business for a long time, especially as part of a marketing effort.

RFM Segmentation:



2) RFM Model

Step 1: Assign Recency, Frequency and Monetary values to each customer.

- Recency is simply the amount of time since the customer's most recent transaction took place.
- Frequency is the total number of transactions made by a customer during a specific period.
- Monetary is nothing but the total amount that the customer has spent across all the transactions.

Step 2: Here customer list is divided into tiered groups for each of the three dimensions (R, F, and M).

Step 3: Select groups of customers to whom specific types of communications will be sent, based on the RFM segments in which they appear.

Step 4: Creating special messaging that is customised for each customer group. By focusing on behavioural patterns of particular groups RFM analysis allows marketers to communicate with customers in a much more effective manner.

The three main differences as suggested, **R**-recency, **F**-frequency, and **M**-monetary, are defined and calculated. These three values are important as F and M indicate customer value, and R reflects customer engagement and satisfaction.

RFM analysis is applied here with Python, exhibiting its simplicity and use of most basic set of information available with purchasing records. RFM analysis can segment customers into homogenous group quickly with set of minimum variables.

K-means Algorithm:

Step 1: To decide the numbers of clusters select the number K.

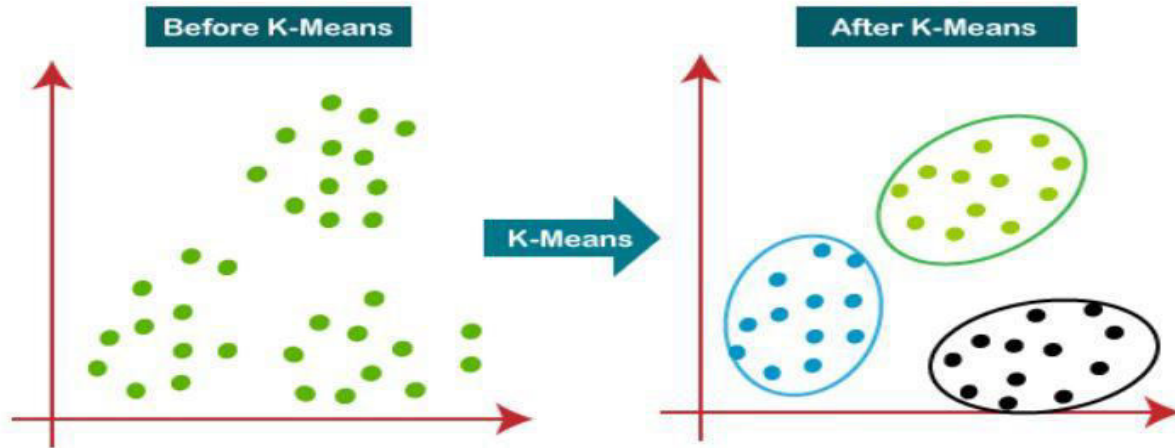
Step 2: Select random K points or centroids.

Step 3: Assign each data point to their closest centroid, which will form the predefined clusters.

Step 4: Calculate the variance and place a new centroid of each cluster.

Step 5: Repeat the third step.

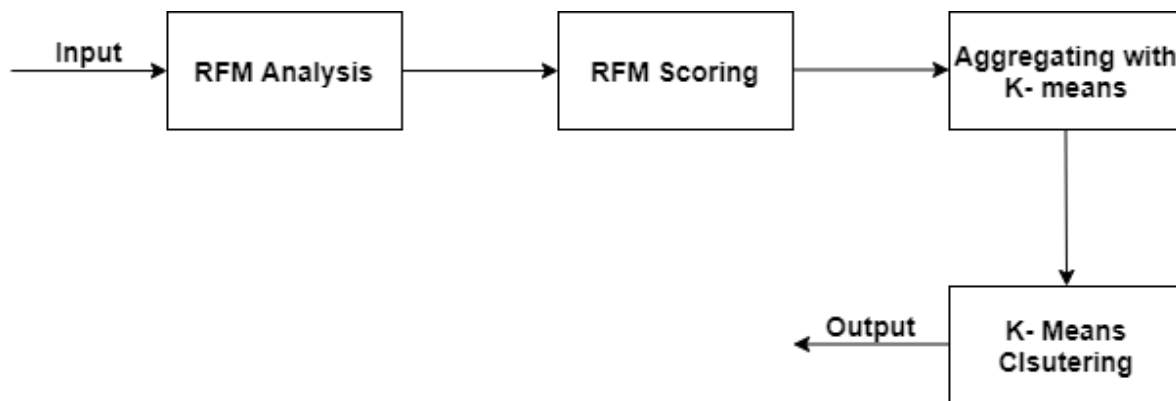
Step 6: If any reassignment occurs then go to Step 4 else got to FINISH.



3) K-means Algorithm.

We are going to use K- means clustering algorithm which comes under Partitional Clustering. K- Means clustering is a classical clustering algorithm that uses expectation maximization like technique to partition number of data points into K clusters.

Finally a data score will be generated which will be displayed according to the output of the algorithm on the console.



4) Flow Diagram

REFERENCES

[1] RFM Based Market Segmentation Approach Using Advanced K-means and Agglomerative Clustering: A Comparative Study by Sabbir Hossain, Shihab Shyla Afroge, Sadia Zaman Mishu 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019

[2] W. Pride, O. Ferrell, B. Lukas, S. Schembri, O. Niininen and R. Casidy, Marketing principles, 3rd ed. South Melbourne, Victoria : Cengage,2018, p. 200.

[3] P. Spring, P. Leeflang and T. Wansbeek, Journal of Market-Focused Management: The Combination Strategy to Optimal Target Selection and Offer Segmentation in Direct Mail. Kluwer Academic Publishers, 1999, pp. 187-203.

[4] O. Etzion, A. Fisher, and S. Wasserkrug, “e-CLV: a modelling approach for customer lifetime evaluation in e-commerce domains, with an application and case study for online auctions,” IEEE International Conference on e-Technology, e-Commerce and e-Service, 2004. IEEE 04. 2004, 2004.



- [5] C. Cheng and Y. Chen, "Classifying the segmentation of customer value via RFM model and RS theory", Expert Systems with Applications, vol. 36, no. 3, pp. 4176-4184, 2009.
- [6] Y. Chen, C. Cheng, C. Lai, C. Hsu and H. Syu, "Identifying patients intarget customer segments using a two-stage clustering-classification approach: A hospital-based assessment", Computers in Biology and Medicine, vol. 42, no. 2, pp. 213-221, 2012



INNO  **SPACE**
SJIF Scientific Journal Impact Factor

**Impact Factor:
7.512**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details