

# Customer Segmentation on Online Retail using RFM Analysis: Big Data Case of Bukku.id

Mohamad Abdul Kadir, Adrian Achyar, Abhishek Puppala

October 23, 2022

## 1 The main idea

The objective of this research is to analyze consumers' buying habits, build consumer groups, and identify customer addresses for Bukku.id. This research provides utilisation of Bukku.co.id's consumer's buying information from the days of September 1, 2017, to september 17, 2018. To identify consumer segmentation, the RFM methodology and clustering are applied. It would also be advantageous for the business to prioritize any treatments relating to authors and publishers in its customer profiles. Understanding location analysis will help you create better offline marketing strategies.

## 2 The methodology

There are two pre-processing algorithms used in this study. They are data-preprocessing which contains rebuilding necessary tables using data error elimination and dates, months information were extracted from the dataset. The second pre-processing method is done with regarding to RFM technique. The transaction table consists of 19,813 rows, 344 book titles, 173 authors, and 37 publishers. RFM Analysis use recency, frequency, and monetary attributes to characterize customers.

The clustering was done by the K- means clustering analysis to obtain customer segmentation. The R, F, or M value is analyzed to see if it is above or below the average value by designating high or low in order to identify the RFM profile from one another. Recency is below the mean value when the sign is high, and above the mean value when the sign is low. Due to the customer's recent purchase, the smaller number would be given a higher rating for recentness. However, the high sign indicates that the frequency or monetary component is above the mean value, while the low sign indicates that it is below the mean value. The ideal client has a low recency—they recently purchased the product, a high frequency of shopping, and a high monetary value, they spend a lot of money. The Pareto Analysis can be employed to pinpoint any authors or distributors who sell more volumes than some others. A corporation may use pareto analysis to advertise and sell rapidly moving commodities in order to maximize profits. Consumer profiles are used for customer location analysis after which it can be examined. This will facilitate cooperation between the business, the publisher, the author, and other important stakeholders as well as the development of any online-offline marketing initiatives. Create marketing strategies based on customer profiles so that you may develop marketing plans with more targeted marketing actions based on the geography and customer profiles.

## 3 The results

All datasets are subjected to RFM analysis. Recency, frequency, and monetary calculations come first. Recency, Frequency, and Monetary values are used to group the RFM data. The numerical output of the RFM method using R studio. Each consumer has a unique RFM value. Through the use of clustering analysis based on RFM numeric value, this new dataset will generate clusters. After the dendrogram cluster, the number of clusters is established. To create a dendrogram cluster, K Means analysis is used. Each customer ID is assigned to a cluster group by understanding the K Means-based cluster of customers. In order for each customer to share a given trait with other customers in the same group. Any consumer in a distinct category will have different traits, and they may also be treated differently. g. Having knowledge of the set cluster, The average RFM value for each cluster is compared to the average RFM value overall.

Each cluster group has a unique RFM value feature. Group 1 can be referred to as platinum because it has a higher mean value for recentness, frequency, and money. Group 3 can be referred to as gold because of its moderate mean values for currency, frequency, and recentness. Group 2 can be referred to as iron because it has a low mean value for recentness, frequency, and money. Prioritizing marketing efforts according to publisher and author can be done by using Pareto analysis.

## 4 Recommendation

Based on client purchasing information, this study looks into customer segmentation. Sadly, it omits demographic and loyalty information. Three clusters make up the customer segmentation that has been established. The three clusters

show how frequently they made purchases, how recently they did so, and how much money they spent with the business during that time. Demographic and loyalty information can be combined with the results of future study.