

# Customer Segment Prediction on Retail Transactional Data Using K-Means and Markov Model

A. S. Harish\* and C. Malathy

SRM Institute of Science and Technology, Kattankuluthur, Chennai, 603203, Tamil Nadu, India

\*Corresponding Author: A. S. Harish. Email: asharrish@gmail.com

Received: 04 May 2022; Accepted: 06 July 2022

**Abstract:** Retailing is a dynamic business domain where commodities and goods are sold in small quantities directly to the customers. It deals with the end user customers of a supply-chain network and therefore has to accommodate the needs and desires of a large group of customers over varied utilities. The volume and volatility of the business makes it one of the prospective fields for analytical study and data modeling. This is also why customer segmentation drives a key role in multiple retail business decisions such as marketing budgeting, customer targeting, customized offers, value proposition etc. The segmentation could be on various aspects such as demographics, historic behavior or preferences based on the use cases. In this paper, historic retail transactional data is used to segment the customers using K-Means clustering and the results are utilized to arrive at a transition matrix which is used to predict the cluster movements over the time period using Markov Model algorithm. This helps in calculating the futuristic value a segment or a customer brings to the business. Strategic marketing designs and budgeting can be implemented using these results. The study is specifically useful for large scale marketing in domains such as e-commerce, insurance or retailers to segment, profile and measure the customer lifecycle value over a short period of time.

**Keywords:** K-means; retail analytics; clustering; cluster prediction; Markov chain; transition matrix; RFM model; customer segmentation; segment prediction; Markov model; segment profiling

## 1 Introduction

Retail Business is one of the foremost domains in which marketing and the customized customer targeting is of critical value. It is an industry with the largest and most dynamic customer base, which also has to essentially provide finished goods and services that satisfy the needs of every single type of customer. This makes it a business with the most number of marketing challenges as well as an effective field for successful marketing explorations, unsurprisingly causing it to be one of the well sought out field fit for data analytics. Statistical and Predictive algorithms are being widely utilized in the retail industry for tracking the purchase patters, optimizing the inventories, budgeting for sectors, forecasting sales, targeted marketing etc. [1].



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Customer profiling or segmentation is one of the key models that play a critical role in any retail marketing decision. A retailer has to provide a *Unique Selling Proposition* (USP) to different kinds of customers which sets them apart from their competitors, this not only involves being ahead of the market trends or implementing new strategies but also understanding their customer base and ensuring the needs of their customer segments are satisfied. This is where customer profiling really helps, as the customer needs not only depend on their personal preferences but on numerous other financial, behavioral, historical, demographical metrics as well [2]. For example, a customer's need may change with time based on their life stages being a student, a parent etc. or their purchases may be impacted by the financial conditions. Understanding how each of these metrics impact a customer's retail purchasing interests and the purchase patterns will help in segmenting the larger base on these metrics provided we are able to acquire the data pointers based on their activity.

Transactional Data is one such mode of data acquisition in retail industry that provides steady inflow of metrics that could help us understand purchase patterns or preferences. It covers multiple aspects such as product preferences, association patterns, impact of sales offers, sales trends etc.

An effective methodology for customer segmentation would be clustering on transactional data [3], especially in instances where measuring the customer's financial worth and marketing budgeting comes into picture. It is a far simpler approach to establish how to balance the cost to revenue risk before taking strategic marketing decisions. This may or may not be largely influenced by the behavioral or demographical metrics and rather may be accurately reliant on historic data of purchase patterns. It is also notable that customers are likely to move across segments with time and therefore establishing an effective methodology to calculate their transitions and segment distributions will prove helpful in identifying valuable segments that has a higher net return rate with respect to the marketing cost. Businesses often have to invest more in dormant or moderate segments of customers as the scope of transition to better performing segments is higher for these clusters, this strategy is facilitation by segment prediction models.

Therefore, this study aims in clustering the sample dataset of retail transactional data by first deriving the RFM metrics (Recency, Frequency and Monetary) of the customers and repeating the same across  $t$  number of periods. Provided the cluster profiles are uniform and comparable, we are able to establish the Markov Transition Matrix and predict the movement of customers across clusters in  $t + 1$  time period.

Markov Matrix is a stochastic system that shows all possible states, transitions and the probabilities of each of them which can be leveraged to model the resultant segment distribution. This gives a measurable spread and therefore can be further used to make numeric decisions, specifically such as budgeting or value propositions. Alternative predictive strategies include regression models or decision trees which need extensive training and although may be useful for generic strategies, do not give a definitive mathematical result.

## 2 K-Means Clustering for Customer Segmentation

Unsupervised Clustering Algorithms are a major part of any large scale data analysis and therefore has an undeniable part in retail industry. It helps to segment customers without any necessary information on individual customers. Clustering serves as the first step for any analysis. There are multiple clustering algorithms widely in use with K-Means being the predominant type for larger numerical datasets. Although it has a weakness in accurately establishing the number of clusters, researches have established quite clearly that coupling it with elbow method to establish the cluster count is reliable enough as compared to other methods such as Hierarchical or Two-Step Clustering [4].

The steps in the K-Means method are as follows: [5]:

1. Determine the number of clusters  $n$ , using a separate methodology
2. Select  $n$  number of initial centroids randomly
3. Calculate the distance of a data point to the centroid using the Euclidean distance formula.

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \quad (1)$$

4. Update centroids by calculating the average value of metrics in each cluster
5. Return to stage 3 as long as there is a data point that moves clusters or if the value of the centroid changes

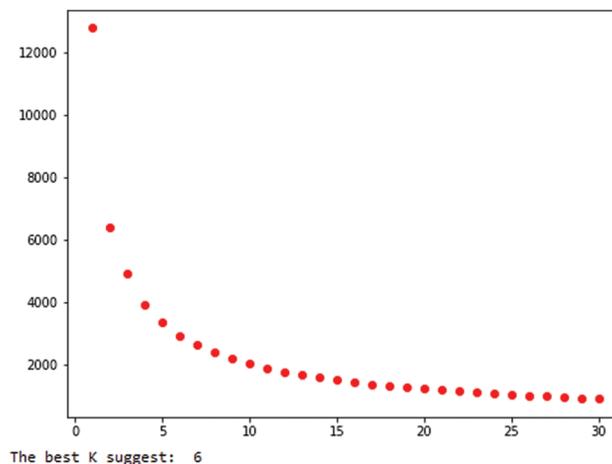
Although the initial cluster allocation plays a pivotal role in the performance of clustering algorithms such as K-Modes or K-Means [6], it does not impact the overall segmentation profiles when the datasets and metrics dealt with is larger and wider. In this case, retail transactional data would have fewer metrics with higher number of records making it trivial to worry over the minor accuracy factors of initial clusters as the volume would ensure the overall accuracy is intact.

### 2.1 Elbow Method for Number of Clusters

Elbow Method is used to determine the number of ideal clusters for a dataset by plotting the explained variation as a function of the number of clusters [7], the elbow or knee of the curve is then identified and picked as the ideal cluster count as it means adding another cluster doesn't necessary add to the modeling value [8].

It is based on the intuition that the fit increases with increased number of clusters but the variation flattens beyond a point when it's termed as over-fitting. This is graphically denoted by the elbow part of the curve.

The resultant elbow curve of the dataset taken for this study is seen as below Fig. 1.



**Figure 1:** Elbow method to determine cluster count

Realistically, it may be challenging to accurately establish the elbow of the curve like in this case where a prominent deviation in the curve may not be observed. However, the curve noticeably flattens beyond the cluster 6 which makes it the best  $k$ -value for our model

## ***2.2 Clustering on RFM Metrics***

The transactional data can be summarized simply in terms of three major metrics—Recency, Frequency and Monetary which in other words are referred to as RFM metrics. Recency is an indicator of the last date of transaction of a customer denoting how recently they were associated with the business. Frequency is a measure of the number of distinct purchases or visits done by a customer. Monetary is the amount of money that a customer has spent with the business during the analysis period. RFM Clustering has proved to be an effective methodology to segregate and focus on customer groups in terms of retention, incentives, targeted offers etc. It is based on a simple theory that customers who are regular or recent visitors and who spend more money on an average tend to bring more value to business or tend to revisit when compared to others.

Despite the simplicity of RFM models, they have some limitations: the customer scores can be predicted only for the given period, it does not predict the precise behavior of the customer or the ability of their needs to change with time, other customer variables are not taken into account and therefore may not be high on accuracy, the model does not output a measurable monetary value. To deal with the disadvantages, it is advised to mix this approach with stochastic ones (RFM with Markov Chain, for instance [9]). In the study, RFM variables are used as input for clustering and subsequently for Markov Modeling.

## ***2.3 Segment Profiling***

Cluster Profiling or Segment Profiling is the last and final stage of any clustering algorithm which involves utilizing the same variables taken as part of cluster analysis. This can be done by identifying the centroids of each cluster obtained as a result of the model implementation, observing the metrics of the centroid and trying to describe the same in business friendly terminology [10]. For example, when the customers are clustered using RFM metrics as in case of this study, we observe the recency, frequency and monetary metrics of the centroids and establish how much engagement or value the centroid provides to the business assuming the centroid value denotes an individual customer behavior. This process is called profiling and requires utmost care and business understanding. Assigning of profiles is how the model can be put to use, or rather be understood in business terms making it fit for real life implementation.

Profiling may sometimes be more complicated as explained above, where looking at the centroids may not prove to be as effective, especially in cases with multiple variables or with higher cluster counts. In such cases, visualization of the cluster distribution across the key variables may come in handy where the variables are identified based on the respective business scenarios.

## **3 Markov Model and Transition Matrix**

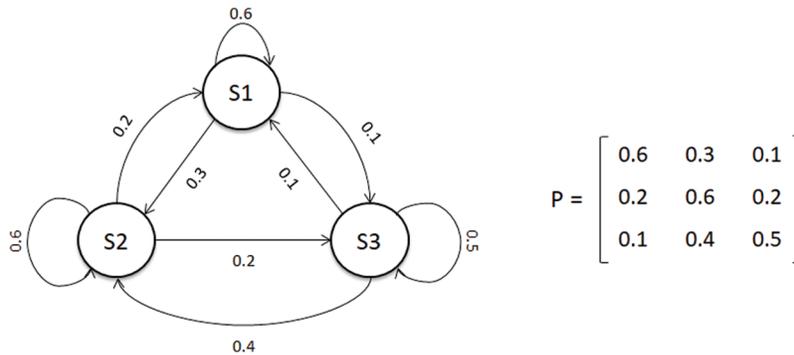
Markov Model is stochastic method for any system that randomly changes states in a stipulated time period, with an assumption that the future state does not depend on the past state but only on the present state [11]. The model shows the possible states along with the likelihood or probabilities of transition from one state to another in an incremental time period. If we are given some data describing the states of a customer and the inter-state relationships with time, then a Markov model can be made to describe the futuristic states of the customer.

### ***3.1 Transition Probabilities***

In a Markov Model, the probabilities of moving from one state to another in a single period are called transition probabilities and the matrix representing these as an  $S \times S$  matrix, where  $S$  denotes the number of states exhibited by the customer is called the Transition Matrix [11]. For finite state distributions, transition probability can be represented as matrix using the equation below-

$$p_{ij} = Pr(X_{n+1} = j | X_n = i) \tag{2}$$

In other words, this is the percentage of population from an initial state, being retained in same state or moving into other states. It measures the likelihood of transition of an entry into the finite set of states. Fig. 2 illustrates this with a sample scenario consisting of 3 states namely S1, S2 and S3. Over a given time period, the probability of an observation transitioning from S1 to S2 or S3 are 0.3 and 0.1 respectively and of remaining in S1 is 0.6. Similar values are measured for states 2 and 3 as well, which can then be denoted as a  $3 \times 3$  matrix, denoted by P in the Fig. 2, where the diagonal measures the likelihood of an entry remaining in the parent state.



**Figure 2:** Transition probabilities represented as transition matrix for 3 states

### 3.2 Transition Matrix and Markov Model

In Fig. 2, the matrix P is what can be called the transition matrix and is of high significance to analyze a Markov chain. The rows of a transition matrix represent the current state where the column represents the future state, where any entry in the matrix is a conditional probability of getting state j, provided the current state is I as shown in Eq. (2). The rows and columns must each sum up to 1, implying that the matrix covers all possible states of existence. This is the primary component of a Markov model which can then be used to recognize patterns, make predictions or to measure the future value of a sequential data. However, the model assumes a Markov property which indicates that the future state of an observation solely depends on the current state only and is not influenced by their past states or external factors. If this assumption does not hold true, then the resultant predictions and computations arrived using the model would be void of any value [12].

Markov Models have multiple use cases in real life such as weather forecasting, customer behavior, brand loyalty etc. and is an effective measure for larger datasets as it requires low or modest computational requirements and is easier to adapt to any series of data as long as it is sequential. The most common Markov Model used these days are text prediction algorithms which provide word suggestions based on most frequently associated words for every word entered.

Markov Chain is a specific scenario of Markov Models which deals with a discrete set of states over sequential periods of time according to the given transitional probability. The notable factor of Markov Chain algorithm is its convergence to equilibrium [13], which indicates that a Markov Chain, as time progresses, will attain a matrix state which can never relate to its original state and reaches a state of stability beyond which there are no transitions observed. This is observed irrespective of its initial state making it fit only for use cases with a stipulated period of time not longer than a few stages of transition.

## 4 Data Implementation of Markov Chain Cluster Prediction

For this study, we have obtained a transactional dataset from a retail departmental store that was collected over a span of 12 months and have passed it through the 4 phases of data preparation which includes data cleaning, feature generation, dimensionality reduction and pre-processing.

The resultant dataset devoid of extreme outliers has 375,617 records and 9 variables out of which Invoice No acts as the primary key. The dataset also holds a Customer ID which helps to identify distinct customers and therefore establish their RFM metrics by observing their purchases throughout the time period. The dataset consists of invoice entries generated over the 12 month period from Dec-2010 onwards until Nov-2011. The Invoice Date, Unit Price and Quantity fields are the key fields which are in turn used to arrive at the Recency, Frequency and Monetary calculated fields.

The dataset is then split into subsets of equal time intervals of 2 months each, resulting in 6 individual datasets. Each of these are then clustered using K-Means algorithm where the  $k$  was already established on the larger dataset using elbow curve. This results in 6 datasets with 6 clusters each, which are then profiled and compared to each other. It is necessary to ensure that the cluster profiles are comparable in nature across individual time periods which make it feasible to consider them as distinct stages for the Markov Model and to establish the future cluster states. This is explained in detail in the subsequent sections.

### 4.1 Data Manipulation and RFM Calculation

The first step of preparing any dataset is to explore the variables and ensure the nulls and outliers are handled appropriately. In this case, the invoices which have null Customer ID are removed as they cannot be associated with any individual customer's behavior. The Amount or Monetary metric at customer level is calculated by summing up individual purchase amounts, which is the product of Unit Price and Quantity. Therefore, any extreme outliers in quantity are also removed to acquire a uniform dataset. The Invoice Date, Invoice No and Amount fields are then used to arrive at the RFM metrics.

Recency is calculated as the number of days since last purchase of the customer, Frequency is the number of distinct purchases done over the given period and Monetary is the total amount spent over the given period. This is done by aggregating the invoices at individual Customer ID level. These measures are then grouped based on percentiles or quintiles, Pareto rule or business acumen based on the use case [14]. For the purpose of this study, we have retained the values as such but have normalized the metrics using log values to ensure they are given comparable weightage in terms of all 3 values.

Fig. 3 depicts the distribution of recency, frequency and amounts/monetary after log normalization and it can be observed how the fields do not skew the results individually and cause a bias.

An alternate method of normalization would be to assign comparable recency and frequency scores based on the percentiles or quartiles of the values calculated using invoice dates and counts. This would imply that a customer who has visited the store quite recently gets a higher recency score and therefore will require a descending quartile assignment during calculation. The customers who is a frequent purchaser gets a higher score for Frequency and this trend and distribution is observed in Fig. 4. However, it is notable to see that quartile distribution is considerably skewing the data for recency since most customers in our dataset have visited the store quite recently.

The conventional method goes on to utilize the RFM metrics to derive the RFM scores which is based on assigning customers into relative buckets of recency, frequency and monetary as shown in Fig. 5 and then defining a profile based on business use case such as frequent buyers, loyal customers, dormant base etc. [15]. This is however not as reliable as the log normalized clustering approach that we are going to implement in the next few steps for the reasons described above with respect to the recency skew observed where there are more number of customers with recency score 5 implying almost everyone visited recently.

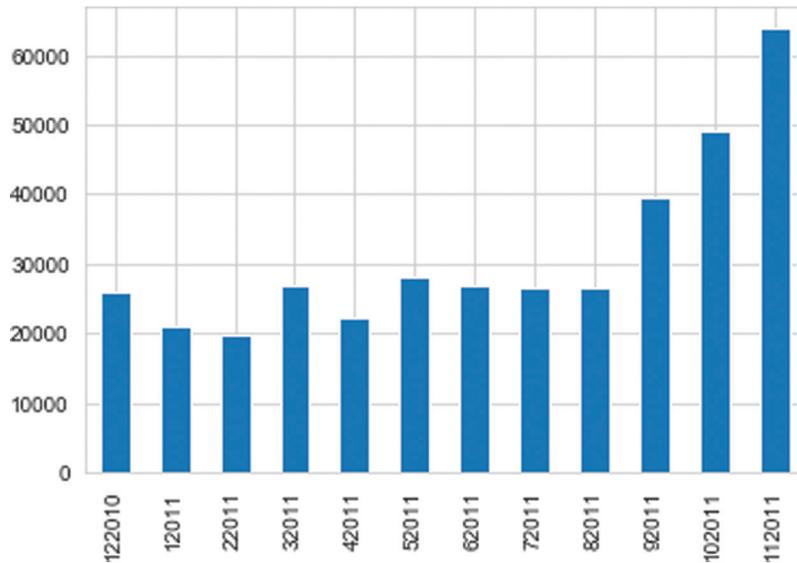


Figure 3: Invoice counts across 12 months

	count	mean	std	min	25%	50%	75%	max
recency_log	4259.0	-2.86e-17	1.0	-2.63	-0.69	0.11	0.86	1.50
frequency_log	4259.0	-1.72e-16	1.0	-1.02	-1.02	-0.24	0.54	4.93
amount_log	4259.0	6.36e-16	1.0	-4.39	-0.67	-0.04	0.67	4.53

Figure 4: Summarizing the log normalized RFM metrics

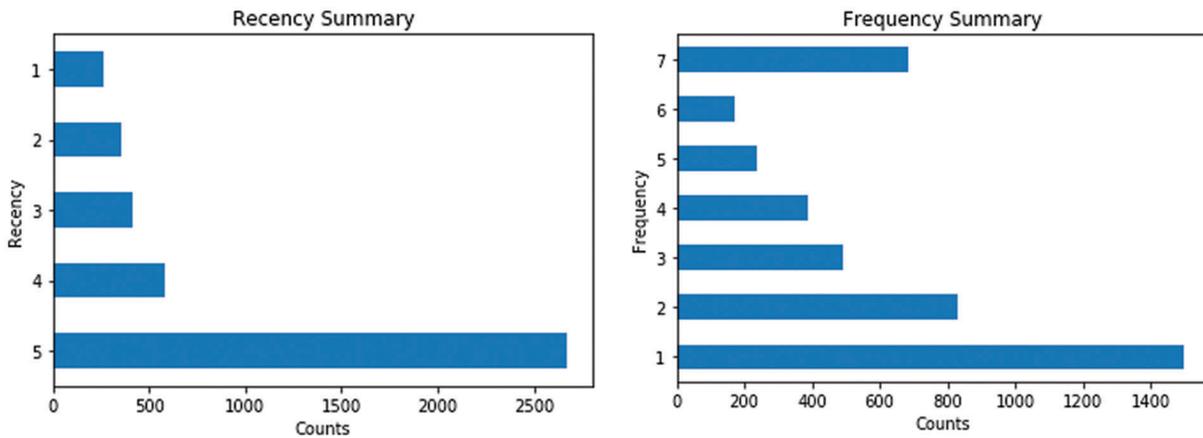


Figure 5: Recency and frequency distribution plots

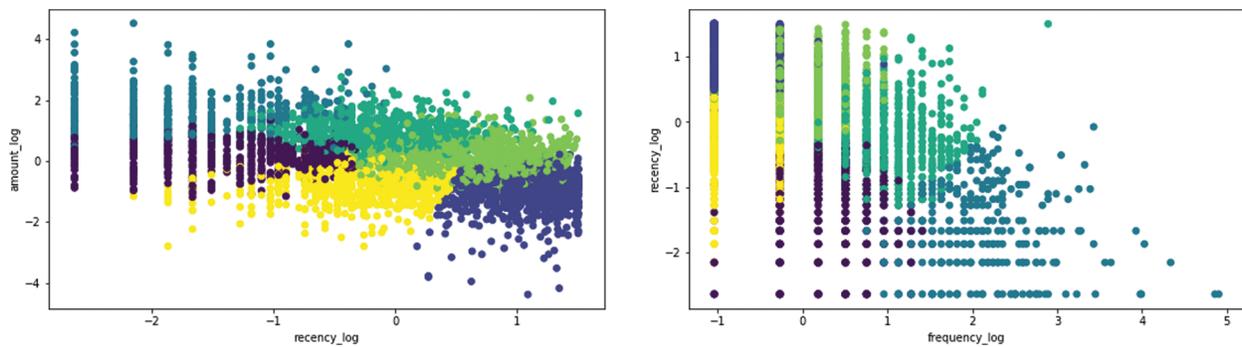
The RFM metrics are separately calculated for the six sets of datasets split for t1-t6 time periods of 2 months interval each. Calculating it separately is essential to capture comparable scales of recency and frequency for any given 2 month period which is unbiased by how latest the data is. These datasets are scalar transformed to pass through the K-Means algorithm which is detailed in the subsequent section.

#### 4.2 Cluster Implementation and Profiling

The RFM metrics for each datasets, referred as d1–d6 for the six sets of time periods, are passed through the k-means model to result in 6 clusters. The resultant centroids are observed and labeled individually. It is

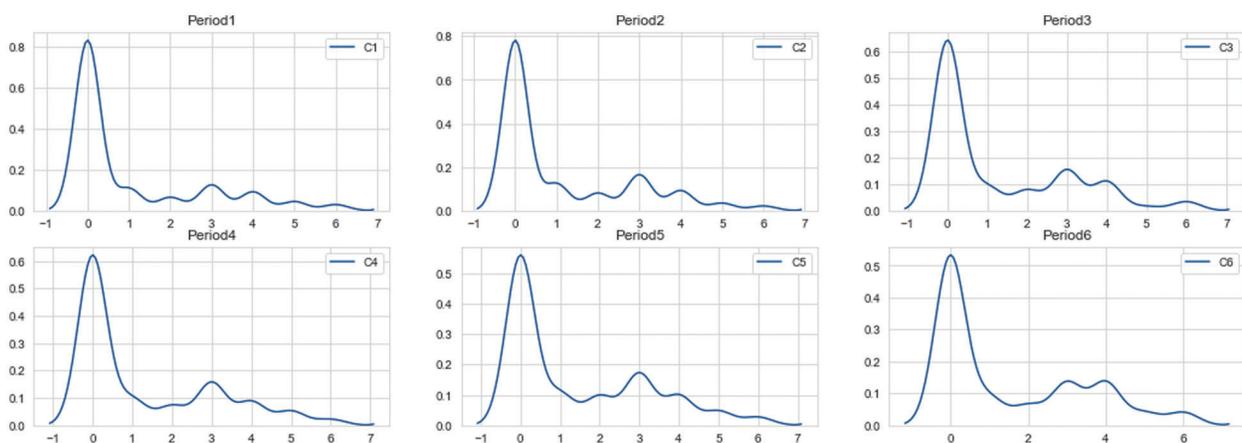
evident that the centroids are comparable and similar in profiles across all six periods since the dataset is obtained from the same uniform source. This is very important for the resultant clusters to be established as different states of a Markov Model which will be utilized in the next section. Also, an additional cluster 0 is added to all datasets for the missing customers who did not make a purchase in the given time period  $t$ , but has made a purchase at some point in the past or future. This is a state which signifies the absence of a customer in a given state, and a transition from cluster 0 to cluster 1 implies new arrival or a transition from cluster 1 to cluster 0 is considered to be an attrition of a customer from period  $t_1$  to  $t_2$ .

The scatter plots in Fig. 6 denote the spread of clusters across recency-amount and recency-frequency for the overall period of 12 months. It can be observed that the customers can be distinctly assigned to various sectors based on how recent, frequent or valuable their purchases are and the customers with lower recency value but with higher frequency or amount value, indicated by teal, are the ones with best business value.



**Figure 6:** Amount and frequency vs. recency spread across clusters

As important as it is that the profiles from each period are comparable to each other, it is also important to observe the density distribution across each cluster with time. Any extreme transition within clusters in a few periods would mean that there are external factors influencing the customer behavior with time and therefore may impact the transition probabilities and hence, we plot the cluster distribution for the six time periods as shown in Fig. 7. The curves show a similar trend but minor variations and transitions proving to be perfect for our further analysis.



**Figure 7:** Cluster distribution across 6 periods

Another notable observation is that the customers in cluster 0, which is the churned or yet to acquire base, are the highest in count. This means that the footfall of customers is a highly dispersed or a one-time occurrence and that the business is most often not seeing repeat purchases in any 2–4 months period. Cluster 3 and 4, where the customers are average or infrequent visitors is the segment with considerable density and the golden segment has the most minimal size of all. It is also the customers in 3 and 4 that have the most potential of transitioning to clusters 5 or 6, depending on the right marketing strategies of a business.

The profile tags and the descriptions are provided in details in [Tab. 1](#), where it is evident that cluster 3, 4 and 5 have a potential to become a better rewarding customer if nudged in the right direction with the right business strategies.

**Table 1:** Cluster profiling (Uniform over 6 periods)

Cluster	Customer profile	Profile description
0	Churned	Customers who have no spend in the period $t$ , but has a past or future activity
1	Rare visitors	Low engagement & low value customer base
2	Recently acquired	Lower recency, but engagement and value metrics are lower
3	Big spenders	Lower engagement, but higher monetary value
4	Moderate opportunity customers	Customer with average engagement and value
5	High opportunity customers	Moderate engagement with high value and a high potential for revenue
6	Golden customers	Best customers with highest engagement and value, most loyal

The study aims at measuring the likelihood of a customer to transition across the clusters across time period  $t$  to  $t + 1$  under natural circumstances since our datasets are derived from a continuous set of invoice data. In realistic studies, it would make more sense to study the transition states of a group of customers who were targeted with special offer or customized marketing strategies to then compare their matrices with the natural set who weren't influenced by the business. This would help us measure the chances of a customer transitioning to a higher desired segment if targeted by a marketing strategy.

### 4.3 Markov Modeling and Cluster Prediction

The Markov Model is a stochastic algorithm that results in higher accuracy with a larger dataset. This is a key driving factor in portioning the given dataset into 6 equal intervals, resulting in more number of transitions from one segment to another. The Markov Model is something that concerns the current state only and therefore, we do not require linking the series of transitions of a customer with time. The cluster transitions are then appended one below the other for the datasets d1 to d5. The final dataset d6 is then retained as a test base, which would help us validate how effective the model metrics are in establishing the cluster movements.

Transition Probabilities for each cluster are calculated by taking the count of transitions of each cluster to a given cluster state divided by the total number of occurrence of the original cluster. The transition probabilities are then arranged in a  $7 \times 7$  matrix, to signify the movement across Clusters 0 to 6 and this is called a Transition Matrix. The equation below explains how the probabilities are calculated for a sample scenario of cluster 0 to cluster 1 transition, where  $n$  is the number of observations–

$$P_{c0 \rightarrow c1} = n_{c0 \rightarrow c1} / n_{c0}$$

$$\begin{bmatrix} 0.6169 & 0.0946 & 0.0714 & 0.1464 & 0.0471 & 0.0199 & 0.0036 \\ 0.6678 & 0.1252 & 0.0638 & 0.0589 & 0.0614 & 0.0213 & 0.0016 \\ 0.6129 & 0.0945 & 0.0668 & 0.0979 & 0.0887 & 0.0288 & 0.0104 \\ 0.5347 & 0.0444 & 0.0716 & 0.1855 & 0.1183 & 0.035 & 0.0105 \\ 0.2846 & 0.0881 & 0.0767 & 0.1665 & 0.2194 & 0.0969 & 0.0678 \\ 0.3509 & 0.0688 & 0.0642 & 0.1353 & 0.2317 & 0.078 & 0.0711 \\ 0.0433 & 0.0155 & 0.0186 & 0.0619 & 0.2508 & 0.1146 & 0.4954 \end{bmatrix} \quad (3)$$

This implies that the probability of a customer under Cluster 1 churning out is more (67%) as compared to their chance of becoming a more valued customer even with all cluster probabilities put together. However, it is evident that the churn rate is reducing as we move up the segment ladder, with the chances of losing a customer being as low as 4% in the golden segment with high engaging, high value customers.

$$\begin{bmatrix} 1828 & 333 & 307 & 556 & 317 & 152 & 88 \end{bmatrix} \times \begin{bmatrix} 0.6169 & 0.0946 & 0.0714 & 0.1464 & 0.0471 & 0.0199 & 0.0036 \\ 0.6678 & 0.1252 & 0.0638 & 0.0589 & 0.0614 & 0.0213 & 0.0016 \\ 0.6129 & 0.0945 & 0.0668 & 0.0979 & 0.0887 & 0.0288 & 0.0104 \\ 0.5347 & 0.0444 & 0.0716 & 0.1855 & 0.1183 & 0.035 & 0.0105 \\ 0.2846 & 0.0881 & 0.0767 & 0.1665 & 0.2194 & 0.0969 & 0.0678 \\ 0.3509 & 0.0688 & 0.0642 & 0.1353 & 0.2317 & 0.078 & 0.0711 \\ 0.0433 & 0.0155 & 0.0186 & 0.0619 & 0.2508 & 0.1146 & 0.4954 \end{bmatrix} \quad (4)$$

Eq. (3) shows the matrix multiplication to acquire the predicted distribution of cluster in time period t6, where the left side of the equation has the density matrix depicting the counts of customers under each cluster in dataset d5 which is for the time period t5. Therefore, the resultant matrix will help us understand the distribution of customers across the 7 clusters in the t6 period as shown below.

$$D_{6_{pred}} = [1983 \quad 308 \quad 248 \quad 500 \quad 326 \quad 124 \quad 92] \quad (5)$$

This summarizes the cluster distribution of customers in t + 1 period, given the cluster distribution in period t. It is observed that there is an incremental churn observed, but also that cluster 4 and 6 are seeing a natural increment. In order to establish the accuracy of the above matrix, the actual cluster distribution in dataset d6 is depicted as below.

$$D_{6_{org}} = [1871 \quad 291 \quad 212 \quad 462 \quad 468 \quad 135 \quad 142] \quad (6)$$

Comparison of the D6 matrices from Markov prediction and actual test dataset proves to be comparable at first glance. It must be noted that the number of distinct customer data entries taken for the purpose of this study is a smaller sample and real use cases would be much larger and therefore, acquiring such comparable results is still appreciable for business decisions. Calculating numeric variance at each stage, it can be observed that some cluster transitions are not as accurately predicted such as the cluster 6 counts which have a variance of around 35% with respect to the test base. This is on one hand due to the fact a sample size of less than 100 records is inevitable to cause high variance, but on the other hand it could also be due to the natural influence in customer purchase behavior such as seasonality or financial influences in the market on the whole. However, the whole prediction on an average shows a 15% variance which is exceptionally helpful when it comes to future planning of business strategies.

## 5 Conclusion

The customers being clustered initially using RFM metrics across the set time intervals has proven effective to establish comparable customer profiles. This is a worthy tag which consists of a finite number of segments, 7 in this case, which helps with establishing the future state relationship using the Markov Transition Matrix. It is observed that the resultant predicted distribution across clusters is closer to the test dataset retained for period  $t6$  with an average variance of 15%. It is an acceptable accuracy for a sample size that is as small as the one taken for this study. In real life, we would be working with much larger datasets which would benefit from the prediction of cluster movements with time. The transition matrix can also be used for Markov Chain prediction which provides the cluster distribution for not only  $t + 1$ , but also for  $t + k$  periods by repeated utilization of the established matrix. This measure can be used to further predict the customer lifetime value of individual entries or a group of customers. It also helps the business to arrive at strategic decisions to pin point the valuable segments of customers who can be pushed across to beneficial segments by providing the right kind of offers.

Although predictive models such as regression or decision trees can be used for scoring or state prediction, it requires exhaustive training and a longer duration of study when compared to Markov Models. The Markov Model being stochastic in nature also gives a definitive probabilistic measure of future states thereby favoring applications that are broader business decisions such as budgeting or targeting. The transition matrix is also a valuable measure in cases where there are two independent sets of customer bases who are targeted with different offers or marketing methodologies. This would help the business establish the better strategy that proved to be more effective in transitioning the customer towards a segment of higher engagement and value. The Model can also be effectively utilized to predict purchase patterns, in recommendation models apart from segment prediction [16]. However, Markov models do not take the past states or historic trends into consideration and in cases such as this where the customers are concentrated predominantly in a single cluster, the results tend to move towards a single state showing no considerable movement with time. This is never observed in real life instances. Therefore, it is unadvisable to prefer Markov for use cases with repeated predictive cycle more than a few iterations.

**Funding Statement:** The authors received no specific funding for this study

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] D. Chen, S. Sain and K. Guo, "Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining," *Journal of Database Marketing & Customer Strategy Management*, vol. 19, no. 3, pp. 197–208, 2012.
- [2] T. Jiang and A. Tuzhilin, "Improving personalization solutions through optimal segmentation of customer bases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 3, pp. 305–320, 2008.
- [3] E. Wong and Y. Wei, "Customer online shopping experience data analytics: Integrated customer segmentation and customised services prediction model," *International Journal of Retail & Distribution Management*, vol. 46, no. 4, pp. 406–420, 2018.
- [4] Z. Huang, "Extensions to the K-Means algorithm for clustering," *Data Mining and Knowledge Discovery*, vol. 2, no. 3, pp. 283–304, 1998.
- [5] F. Cao, J. Liang and B. Liang, "A new initialization method for categorical data clustering," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10223–10228, 2009.
- [6] A. Ahmad and L. Dey, "A K-mean clustering algorithm for mixed numeric and categorical data," *Data and Knowledge Engineering*, vol. 63, no. 2, pp. 503–527, 2007. <http://dx.doi.org/10.1016/j.datak.2007.03.016>.

- [7] D. Marutho, S. H. Handaka, E. Wijaya and Muljono, "The determination of cluster number at K-mean using elbow method and purity evaluation on headline news," in *Int. Seminar on Application for Technology of Information and Communication*, Semarang, Indonesia, pp. 533–538, 2018.
- [8] F. Liu and Y. Deng, "Determine the number of unknown targets in open world based on elbow method," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 5, pp. 986–995, 2021.
- [9] P. Pfeifer and R. Carraway, "Modeling customer relationships as Markov chains," *Journal of Interactive Marketing*, vol. 14, no. 2, pp. 43–55, 2000.
- [10] M. R. Vela and E. M. García, "A segmentation analysis and segments profile of budget air travelers," *Cuadernos de Turismo*, vol. 26, pp. 235–253, 2010.
- [11] H. S. Hwang, "A stochastic approach for valuing customers: A case study," *International Journal of Software Engineering and Its Applications*, vol. 10, no. 3, pp. 67–82, 2016.
- [12] C. J. Cheng, S. W. Chiu, C. B. Cheng and J. Y. Wu, "Customer lifetime value prediction by a Markov chain based data mining model: Application to an auto repair and maintenance company in Taiwan," *Scientia Iranica*, vol. 19, no. 3, pp. 849–855, 2012.
- [13] Ari Freedman, "Convergence theorem for finite Markov chains," in *Proc. REU*, Chicago, IL, USA, 2017.
- [14] C. H. Cheng and Y. S. 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.
- [15] V. Aggelis and D. Christodoulakis, "Customer clustering using RFM analysis," in *Proc. 9th WSEAS Int. Conf. on Computers*, Wisconsin, WI, USA, pp. 1–5, 2005.
- [16] O. Netzer, J. M. Lattin and V. Srinivasan, "A hidden Markov model of customer relationship dynamics," *Marketing Science*, vol. 27, no. 2, pp. 185–204, 2008.