Customer Analytics in Fashion Retail Industry

Chandadevi Giri, Sebastien Thomassey and Xianyi Zeng

Abstract This paper aims to give an overview of customer analytics in fashion retail industry in the era of big data. Fashion retail industry has been facing significant challenges since last few years due to rapidly varying customer demands. Nowadays, customers are much more informed and connected because of social media and other channels on the Internet. They demand more personalized services, and perception is not sufficient to understand our customers. Therefore, we need data to understand our customers and meet their expectation. We will discuss how customer analytics can create value in fashion retail industry, strategies and methodology to examine the consumer data. Employing and investing in these methods and technologies, industry will benefit from improved revenues, improve in sales, higher customer retention rates and thereby it will sustain in the uncertain markets. Segments are created using recency value of the customers, and their future behavior is predicted using transition matrix.

Keywords Customer analytics · Big data · Segmentation · Consumer behavior · Fashion retail industry

1 Introduction

The present study is part of a Ph.D. project in sustainable design and management of textiles, focusing on e-commerce and consumer analytics to understand consumer behavior, and it is necessary to understand data apart from the perception of the consumer. The contemporary fashion industry is challenging, and integrating customer analytics with its business models will enable it to achieve business goals and high

C. Giri (✉) · S. Thomassey · X. Zeng
Gemtex, Ensait, 2 Allée Louise et Victor Champier, 59056 Roubaix, France
e-mail: chandadevi.giri@ensait.fr

C. Giri
University of Boras, 501 90 Boras, Sweden

Soochow, College of Textile and Clothing Engineering, Suzhou 21506, China

A. Majumdar et al. (eds.), *Functional Textiles and Clothing*,
https://doi.org/10.1007/978-981-13-7721-1_27

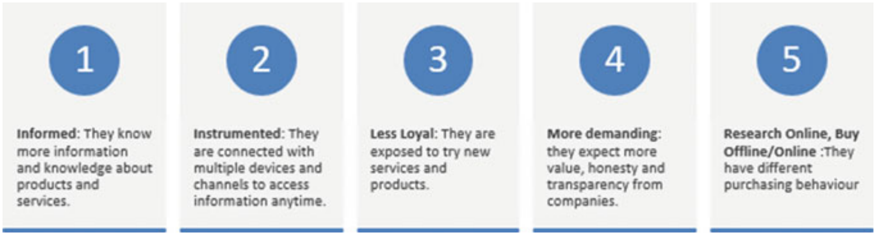

Fig. 1 Features of new customers

revenue growth. To illustrate how customer analytics can be effective, we have created customer segments on fashion apparel retail data to study customer behavior and the revenue generated by them. This will be discussed in detail in Sect. 5, customer data analysis.

Understanding consumer preferences has been a major challenge for fashion retailers. As per the market research of Statista [1], the revenue generated in the year 2017 from fashion globally is accounted for US$406,476 million globally and it is estimated to increase by 11.6% by 2022. In the fashion industry, the largest market segment is "Clothing" that contributes a market volume of US$272,599 million in 2017. Revenue generated in China itself contributed around US$164,219 million in 2017. Also, it is estimated that large no. of users will be buying fashion products online by 2022, and the clothing industry will have the maximum market share. Currently, China, the USA, and UK are the major players in the fashion industry, and they are expected to grow in the future. It is important for the retailers to invest in big data analytics to understand customer preferences in real time. Big data provides an opportunity to understand customers in more precise way, and this leads to the emergence of new analytics area which is termed as "customer analytics." "Customer analytics refers to the collection, management, analysis, and strategic leverage of a firm's granular data about the behavior(s) of its customers" [2]. A new type of consumers has evolved. The evolution of customer behavior can be seen in Fig. 1. Therefore, it is essential to study the customer insights from the generated data to understand the fashion market trends.

2 Research Problem and Goal

The premise of the research problem emerges from the fact that fashion retailers are continuously facing many challenges in terms of predicting customer behavior in real time and adopting new strategies to fulfill customer demands. In line with this premise, the goal of this research was to study the following:

1. How to classify customers based on their purchasing behavior evolving over time?
2. How to predict future customer movement and revenue generated by them?

The goal of the customer analytics is to address the above problems. It is important to understand whether consumer is creating value to the business, or whether they are satisfied with the services provided by the retail company, and their preferences in order to take appropriate actions to improve the services and products. Customer analytics can help retail companies to retain their customer base, increase revenue growth, and to predict consumer behavior, and eventually, the objective of value creation by each consumer could be achieved. Therefore, it is imperative to segment different group of the consumers as per their preferences and the value they create.

3 Customer Analytics

This section presents an overview of customer analytics strategy, scope, and methodology as shown in Fig. 2.

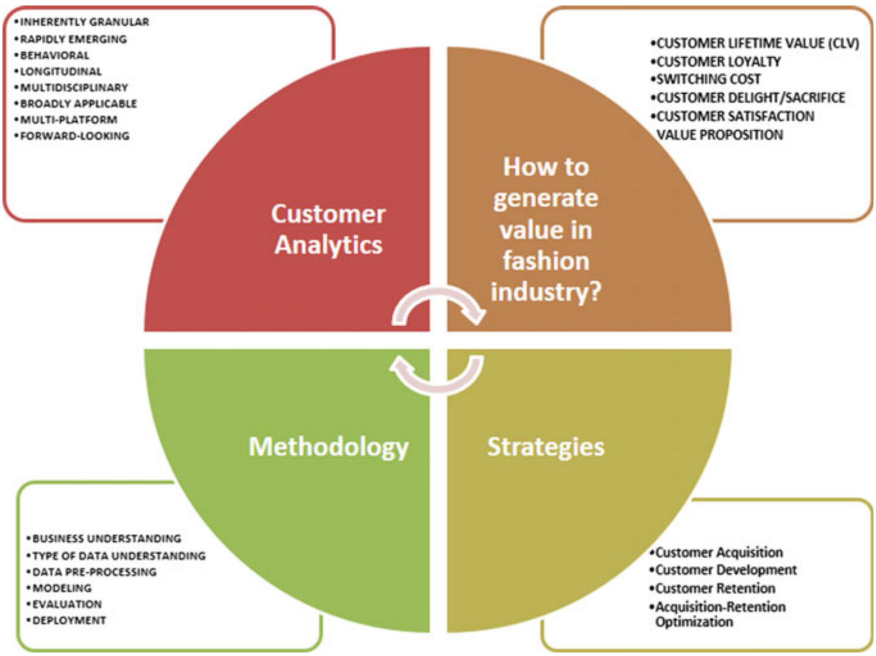

Fig. 2 Customer analytics, strategy, scope, and methodology

3.1 Feature of Customer Analytics

Company's business goals can define how customer analytics can be incorporated into their business intelligence. It could be focused on the prediction of customer behavior at an individual level without considering any other information, or it could be used by merging information from different systems to analyze, or it could be based on behavioral or longitudinal social network analysis. In other words, it depends on the business goals of the companies and which problem they are trying to address.

3.2 Value Generation in Fashion Retail Industry

Fashion retail industry is complicated, and it is quite difficult to understand the consumer choices toward the product. We can generate a high business value by identifying consumer lifetime value which can in turn help the fashion industry to achieve their profit goals. Customer Life Time Value (CLV) is the predicted value that businesses will derive from their entire relationship with their customers [3]. Well-known machine learning (ML) and probabilistic models such as Bayesian Inferences, Moving Averages, Regressions, and Pareto/NBD (Negative Binomial Distribution) can be used to predict CLV [4]. Customer segmentation can be done considering the demographic, geographic, behavioral factors using K-means clustering [5]. Association analysis can also be used for building recommendation system [6].

3.3 Strategy

All industries are consumer oriented, and consumers are crucial for their business success. The main strategy for fashion industry is to expand their customer base. As mentioned in the introduction, the dynamic nature of customers' buying pattern drives retailers to improve their strategy for customer acquisition and retention.

3.4 Customer Analytics Methodologies

Methodology for customer analytics often depends on the business problems that we are trying to address. Fashion social network data can be collected from tweets, boards, blogs to understand the hot topics, current trends, brands, events, criticism, etc. in fashion industry [7]. Internal sources are ERP, CRM, fashion e-commerce, etc., and external sources are cookies, plug-ins Adobe flash, etc. [8]. Data can be pre-processed to get into structured form; then, analytical method such as descriptive and predictive can be applied, and models can be evaluated. According to PwC and

Table 1 Customer analytics with ML methods and statistics

Analysis	ML methods and statistics
Future profitability	Neural networks
CLV	Statistics
Potential CLV	Multi-regression
CLV profiling	Supervised clustering
Churn	Decision trees/Classification
RFM profiling	Decision trees
Acquisition modeling	Neural networks
Response analysis and modeling	Neural networks
Response index	Statistics
ROI	Structured procedures
Campaigns	Regression and structured procedures

SAP retailer survey [9], 39% of retailers ranked "Ability to turn customer data into intelligent and actionable insight" one of their greatest challenges. There is a huge gap between the big data and fashion retail industry. Retailers are more concerned with the data collection. After collecting data, they do not know what to do with such a huge and highly complex customer data. There is a lack of systems for tracking their minute-wise inventory. Therefore, understanding business, data, and customers is very important but retailers have to invest on analytics for creating valuable insights from the chunks of data that could benefit the industry to improve their products and services. Fashion industry is in a greater need to use advanced business intelligence tools, data analytics platforms, big data tools for capturing and processing data. Table 1 lists different machine learning method which can be used for customer analytics to predict profitability, life cycle of the consumers, loyalty, and campaigns.

4 Customer Data Analysis

For this study, we used dataset from the apparel industry, which spans from 2015-10-01 to 2016-12-01. To create segments, we used "Recency" value of each customer as the main indicator. Segments are created based on the recency value for six months, and each segment is further classified as "Inactive," "Less Active," "Active," "Highly Active," and "New Customers." Given the 14-month time span of our dataset, we created two segments: one is 01-06-2016 to 01-12-2016, and the other is from 01-12-2015 to 01-06-2016. Using these two segments, we computed the transition matrix which is used for predicting the behavior and movement of customers within the five categories of each segment in the next one year. Based on the no. of the customers in each segment, revenue and cumulative revenue are calculated and predicted for next one year. No discount rate is considered for this study, but full price only.

4.1 Methodology

Transition matrix is used to compute the likelihood of the future occurrence depending on the current state. Let us assume that fashion retailers have X_n customers where X represents the total number of customers in a given segment at current state n, $X \in S$, where $S = \{$Inactive, Less Active, Highly Active, Active, New Customers], then the probability P_{nm} of the customer in next state m will be given as

$$P_{nm} = P(X_{t+1} = m | X_t = n) \tag{1}$$

Thus, the transition matrix is conditioned on the present state, and the previous and the upcoming conditions are independent. In our study, we predicted the number of customers in each segment in an organization and their transition to the next state. Suppose the states are 1, 2, 3, … r, where r represents the row, then the transition matrix for the different segments can be represented in matrix form as shown in Eq. 2. Thus, the probability of the customers in a segment in state m conditioned on state n can be represented by Eq. 3, and sum of all probabilities in a row will be equal to 1.

$$P(S) = \begin{bmatrix} S_{11} & S_{12} & S_{13} & \dots & S_{1r} \\ S_{21} & S_{22} & S_{23} & \dots & S_{2r} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ S_{r1} & S_{r2} & S_{r3} & \dots & S_{rr} \end{bmatrix} \tag{2}$$

$$\sum_{m=1}^{r} P(S_{nm}) = \sum_{m=1}^{r} P(X_{t+1} = m | X_t = n) = 1 \tag{3}$$

4.2 Descriptive Analysis of the Data

Dataset is comprised of 5,770,844 customer transaction data with 1,020,923 distinct customers. For the customer analysis, we considered three variables: "Customer ID," "Purchase Amount," and "Date of Purchase", see Table 2. The statistics show that average spending by each customer is 38.60 units, and maximum spending for each transaction is 199.90 units. The maximum time lapse with last transactions is 426.77 days (approx. 14 months).

We have calculated three variables for customer data, i.e. recency, frequency, and monetary Value (RFM). Recency is the no. of the days lapsed between customers' recent transaction date and the last transaction date. The bigger the recency value, the less active the customers are. Frequency is defined as the no. of the purchases made by the customers in a given period of time, and monetary value is the amount of money customers spent in each transaction. RFM is calculated for each customer. We can see from Table 3 that the average recency of a customer is 180 days, and

Table 2 Summary statistics of customer data

Purchase_amount	Date_of_purchase	Days_since
Min.: −4.08	Min: 2015-10-01	Min.: −0.2292
1st Qu.: 19.95	1st Qu.: 2016-01-10	1st Qu.: 120.7708
Med Median: 34.95	Median: 2016-04-22	Median: 222.7708
Mean: 38.60	Mean: 2016-04-25	Mean: 219.4961
3rd Qu.: 53.15	3rd Qu.: 2016-08-02	3rd Qu.: 325.7708
Max.: 199.90	Max.: 2016-12-01	Max.: 426.7708

Table 3 Summary statistics of recency, frequency, and monetary (amount) for customer

Recency	Frequency	Amount
Min.: −0.2292	Min.: 1.00	Min.: 0.00
1st Qu.: 80.7708	1st Qu.: 2.00	1st Qu.: 25.82
Median: 155.7708	Median: 3.00	Median: 37.84
Mean: 180.4872	Mean: 5.65	Mean: 40.93
3rd Qu.: 288.7708	3rd Qu.: 6.00	3rd Qu.: 51.35
Max.: 426.7708	Max.: 1555.00	Max.: 199.90

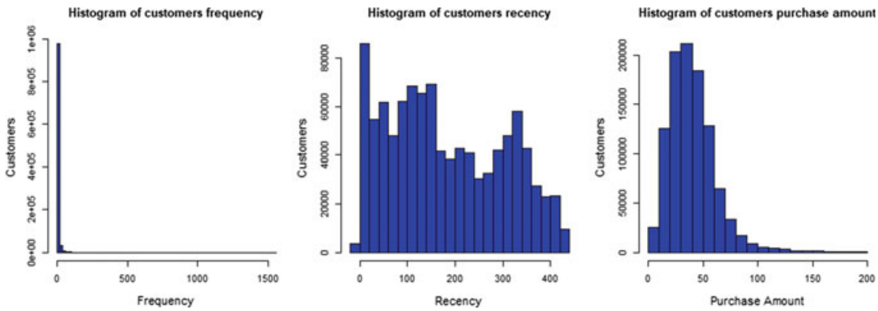

Fig. 3 Histogram of recency, frequency, and purchase amount

maximum recency is 426 days. The average spent amount per transaction is 40.93 units while the maximum spent amount for each transaction is 199.90 units. According to the frequency, customers had made at least one purchase and on an average five purchases. Histogram of RFM values is depicted in Fig. 3.

4.3 Segmentation Based on Recency

Two segments were created for the six-month interval based on recency values. In each segment, classes were assigned as Inactive, Less Active, Active, and Highly

Table 4 Segmentation class based on recency value

Recency (days)	Customer class
>180	Inactive
≤180 and >90	Less active
≤90 and >50	Active
≤50	Highly active

Table 5 Average recency, frequency, and purchase amount for each class within Segment 1

	Segment 1	Recency	First_purchase	Frequency	Amount
1	Inactive	296.47006	317.42518	3.286801	40.69390
2	Less active	131.86588	217.59175	5.252555	34.16796
3	Active	69.98107	195.52441	6.867103	44.75028
4	Highly active	21.90170	301.30038	15.947734	43.73544
5	New customers	23.62203	25.01742	3.271787	62.07015

Table 6 Average recency, frequency, and purchase amount for each class within Segment 2

	Segment 2	Recency	First_purchase	Frequency	Amount
1	Inactive	212.44278	214.87399	2.388128	52.89656
2	Less active	135.97081	150.84615	3.346653	33.82017
3	Active	68.74415	117.36808	4.974818	43.37889
4	Highly active	26.68699	175.74538	12.261362	39.70714
5	New customers	24.02350	25.32776	2.695572	43.46406

Active customers based on the recency values as shown in Table 4. The class "New Customer" is calculated as Customer in segment = "Highly Active" and first purchase $\leq$ 50. This will help us to identify the customers who were not present in the Segment 2 but are newly added in the Segment 1. Highly Active customers are those whose recency is less than or equal to 50 days while inactive customers are those whose recency is more than 180 days. Recency, frequency, and average purchase amount for Segments 1 and 2 can be seen in Tables 5 and 6.

In Fig. 4, we have depicted how new customers who were absent in Segment 2 are now evolved in the Segment 1. As the logic behind our segmentation is to identify the similarity between the customers in different segments and understanding their behaviors after 6 months, we grouped them together according to their recency criteria for each segment. From the Segment 1, it is evident that new customers have been acquired that were not present in the Segment 2. Those who were new in Segment 2 transferred into another category "Inactive" and "Less Active" in Segment 1. Same can be observed for the highly active consumer, no. of highly active customers reduced in the segmentation 1, which are current customers.

The concept of transition matrix is employed to identify the probabilities of customers changing their segments. In other words, it is important to measure the likeli-

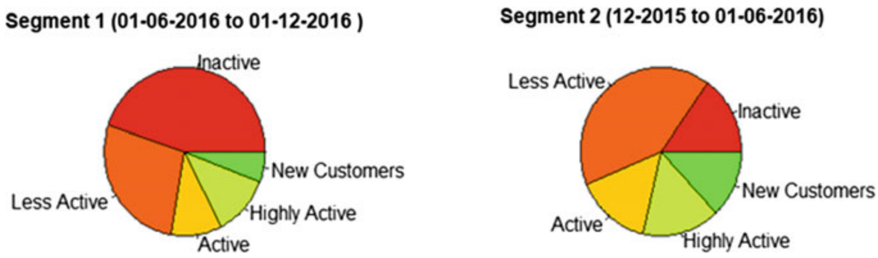

Fig. 4 Number of customers of each class in Segment 1 and Segment 2

Table 7 Transition matrix as per class of customers in Segment 1 and Segment 2

	Inactive less	Active	Active	Highly active	New customers
Inactive	83,293	12,661	4869	8986	0
Less active	208,564	48,854	15,118	24,620	0
Active	64,694	19,288	8322	16,226	0
Highly active	37,687	25,518	13,197	36,721	0
New customers	64,861	14,182	5430	9628	0

hood of inactive customers from the Segment 2 to moving to the Less Active, Active, and Highly Active classes of the Segment 1. Therefore, transition matrix enables us to find the number of customers in each class of the Segment 1. The probabilities from transition matrix are crucial in directing the future marketing campaigns and in targeting the potential customers. Transition matrix showing the probabilities of customers evolving in the next segment (Segment 1) is depicted in Fig. 4.

We have created transition matrix for the two segments to understand how the customers have changed their status from Segment 2 to Segment 1 in Table 7. The classes shown horizontally are Segment 1, and the one vertically is Segment 2. Based on vertical classes, we will see the movement of the customer in the horizontal classes. For example, in the first row, the customers who were inactive in segment 2, 83,293 customers remain inactive, 12,661 becomes less active, 8986 becomes active, and 4869 becomes highly active in Segment 1. So, we can say that very less inactive customers joined the groups: "Active" and "Highly Active." Now, if we see what happened to the new customers in Segment 2, the fifth row of the transition table, 64,861 customers became inactive and 14,182 customers became less active, while only few remained in the groups: "Active" and "Highly Active."

After we divided the rows in Table 7 by the sum of the customers for a given class, we get the transition probabilities as shown in Table 8. From Table 8, we can interpret that 75% customers who were inactive in Segment 2 will remain inactive in Segment 1, while about 8% inactive customers in Segment 2 will become highly active in Segment 1. For new customers in Segment 2, about 10% of them will remain

Table 8 Transition matrix by probability

	Inactive	Less active	Active	Highly active	New customers
Inactive	0.75852617	0.11530020	0.04434063	0.08183300	0.00000000
Less active	0.70186703	0.16440523	0.05087563	0.08285210	0.00000000
Active	0.59609325	0.17772045	0.07667926	0.14950705	0.00000000
Highly active	0.33315064	0.22557747	0.11666063	0.32461126	0.00000000
New customers	0.68927004	0.15071041	0.05770396	0.10231560	0.00000000

highly active, 5% will remain active, and about 69% will remain inactive in Segment 1. Likewise, we can interpret the results for other classes too.

4.4 Prediction Based on Transition Matrix for the Next 6 Months and 12 Months

It is often necessary to predict the no. of new customers in the next time period as they can potentially contribute to the revenue growth. Based on the prediction, managers design their marketing campaigns targeting new customers and the customers that are less likely to evolve, or in other words who are not exhibiting any movement to the active segments can be removed from the target group. Based on the results of transition matrix, we predicted the no. of the customers for the next 6 and 12 months and the revenue generated by them as shown in Figs. 5 and 6. Table 9 shows the predicted value of total number of customers in each class of the segments after 6 and 12 months. We can perceive that the number of the inactive users will be increasing in next one year, which means that new customers, active customers, and highly active customers from Segment 1 (current customers) will move to either Inactive or Less Active group of the same segment. As we have seen that the probability of "Inactive" customers shifting to "Active" group is quite less. New customers will be evolving in the future for 6 and 12 months of prediction. However, Fig. 5 shows the probability of the customer's transition to other groups after 6 and 12 months. We can see from Table 9 that the no. of the customers in "Inactive" and "Less Active" class is increasing, while in "Active" group is decreasing. No. of "Highly Active" users remain approximately the same.

Intuitively, if the number of customers becoming active in the Segment 1 is higher, it means that the revenue generated by them will be higher. The future revenue or the revenue that will be generated in the next segment or in future segments can be predicted by looking at the classes of segments to which they belong. Table 10 shows revenue generated by each class in a segment currently and after 6 and 12 months. From Table 10, it is evident that the inactive customers cannot generate revenue as indicated by values "0," whereas new customers are significant to high revenue generation. However, we could not predict the number of new customers and revenue

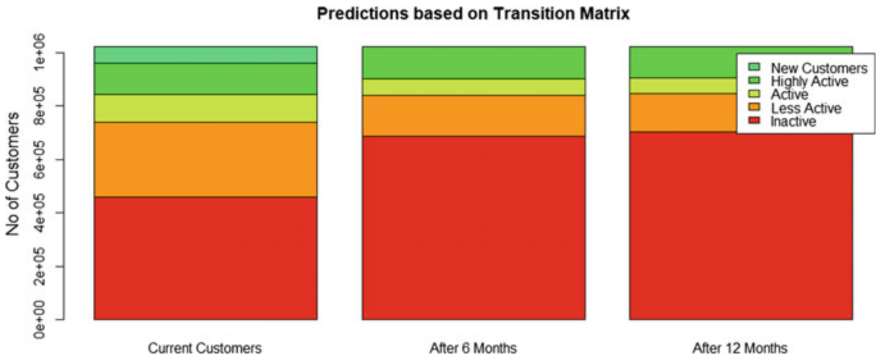

Fig. 5 Prediction using transition matrix for no. of customers for each class for 6 and 12 months

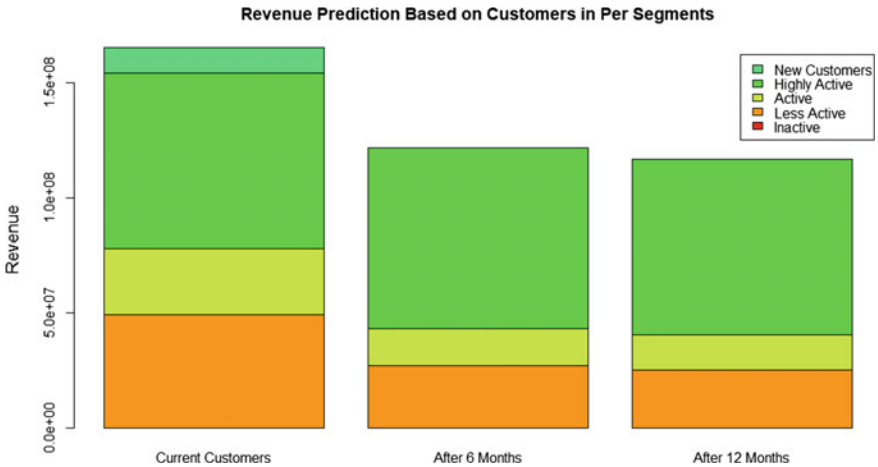

Fig. 6 Prediction of revenue generated by each class after 6 and 12 months

Table 9 Number of customers in each class after 6 and 12 based on Segment 1

	Current customers	After 6 months	After 12 months
Inactive	459,099	687,299.59	704,646.68
Less active	279,777	153,078.56	142,277.96
Active	103,584	59,794.53	56,935.07
Highly active	118,107	120,750.32	117,063.29
New customers	60,356	0.00	0.00

Table 10 Revenue generated by each class after 6 and 12 months based on Segment 1

	Current customers	After 6 months	After 12 months
Inactive	0	0	0
Less active	49,552,927	27,112,632	25,199,675
Active	26,369,141	16,376,269	15,593,133
Highly active	76,394,171	78,103,929	75,719,075
New customers	10,850,470	0	0

generated by them in the future time periods (6 and 12 months) because the Customer IDs of the new customers will be unique and different from the Customer IDs in the dataset we used for the study.

5 Results and Conclusion

Segment is created on 14-month customer data, and five different classes were defined based on the calculated recency value. As a result, we could identify five different customer classes purchase behavior over time, including revenues, accordingly. Transition matrix is used to predict the number of the customers in each segment class and revenue generated by them for the next 6 and 12 months. This kind of segmentation in a fashion apparel industry would help us to identify which segment of customers generates high value to the organization and how they can be retained for a long period. Besides, we can also analyze consumer behavior in detail by studying their purchasing behavior. As this segmentation is created by considering only one parameter "recency" value of the customer, it could be further improved by including other parameters: "Frequency" and "Monetary Value." We will consider machine learning methods for predicting the customer's behavior in our future work.

The fashion industry is dynamic and sensitive to quick changes in the customer behavior with the seasons and trends. Customer analytics will help the fashion industry to quickly respond to changing customer preferences. By applying customer analytics, we can analyze the buying behavior of the consumers and their preferences. Furthermore, it will also benefit supply chain and inventory management and it will be easy for the retailers to make decisions based on real-time tracking systems, reducing losses and helping the company to operate more environmentally friendly. Customer Analytics in fashion retail industry will help to customize the profiles of their consumer, enhance the personalized recommendation services, more loyalty programs, will give the opportunity to know their customers better than before and will help the business to create value from it. Thus, focusing on customer analytics in the era of big data, industries could be benefitted more than ever in the history.

Acknowledgements We are grateful to SMDTex—Sustainable Design and Management of Textiles, Erasmus Mundus Joint Doctoral commission for creating good research environment to take out this research and their consistent support. We would like to show our deep gratitude to Mr. Giuseppe Craparotta, Senior Data Scientist, Evo Pricing for helping us with the data. Without him, it would have been difficult for us to execute this research work. We would also like to thank you Evo Pricing and their innovative team for their contribution to the research.

References

1. Statista—The Statistics Portal for Market Data, Market Research and Market Studies. (n.d.). Retrieved from https://www.statista.com 15 Oct 2017
2. Home—Wharton CAI. (n.d.). Retrieved from http://wcai.wharton.upenn.edu/ 16 Oct 2017
3. Strickland, P.J.: Data Science and Analytics for Ordinary People. S.l.: Lulu com (2015)
4. Wilkinson, J.W., Trinh, G., Lee, R., Brown, N.: Can the negative binomial distribution predict industrial purchases? J. Bus. Ind. Market. **31**(4), 543–552 (2016). https://doi.org/10.1108/jbim-05-2014-0105
5. Li, Z.: Research on customer segmentation in retailing based on clustering model. In: 2011 International Conference on Computer Science and Service System (CSSS) (2011). https://doi.org/10.1109/csss.2011.5974496
6. Tan, P., Steinbach, M., Kumar, V.: Introduction to Data Mining. Pearson, Nueva Delhi (India) (2016)
7. Data is Fashionable [Web log post]. (n.d.) (2016)
8. Curto, J., Braulio, N.: Customer Analytics. Editorial UOC (2015)
9. Verhoef, P., Kooge, E., Walk, N.: Creating Value with Big Data Analytics. Routledge (2016)