LAB University of Applied Sciences

Data Analytics Applied in Customer Relationship Management in field of Segmentation, Retention in Start-ups

LAB University of Applied Sciences

Bachelor of Business Administration, International Business

Thesis 2021

Thoa Nguyen

Instructor: Senior Lecturer Jaani Väisänen, LAB university of Applied sciences.

Abstract

Author(s)	Publication type	Completion year
Thoa Nguyen	Thesis, UAS	2021
	Number of pages	
	40	

Data Analytics Applied in Customer Relationship Management in field of Segmentation, Retention in Start-ups

Degree Program in International Business (IB17)

Abstract

The research aims to figure out the ways data analytics applied in customer relationship management; especially in the field of segmentation and retention in start-ups. Furthermore, the research also focuses on the benefits and values that segmentation and retention offer. The research was granted permission by Eurosender and The Coffee House start-ups, located in Luxumbourg and Vietnam, respectively, to gather the information of the company in the process of implementation data analytics in segmentation and retention to improve the Customer Relationship Management.

The data was gathered from primary and secondary data. In terms of secondary data, the author extracted information from books, articles, and publications. Regarding to primary data, the author carried out the in – depth with two company representatives such as Eurosender and The Coffee House.

The outcome of the research could assist the start-ups in the beginning period of implementation of data analytics to have a clear picture of the ways data analytics application in segmentation and retention.

Keywords: data analytics, start-up, segmentation, retention, customer relationship management.

Table of Contents

1 Introduction

1.1 Theoretical Topics

According to 6-month experience working as a data analyst in 2 start-ups, the author comes up with ideas how data analytics would assist start-ups in marketing areas specialized in the areas of customer segmentation and retention. As start-ups in the first years of operation, start-ups mainly focus on attracting customers, increasing sales, etc. start-ups do not have resources, including human and financial resources to dig deeper into data analytics or start-ups do not know which solutions could assist them to analyse data in an efficient and affordable way. After few years, when data is bigger and business is on the rapid growth, the start-ups try to think about how to retain their current or loyal customers by understanding their behaviours and getting their insights, how customers would be divided into segmentation. The questions will be answered by applying data analytics. Regarding to recent research of a group of Norwegian students, data analytics would assist start-up companies in exploring their challenges and barriers (Berg et al. 2019). Data analytics is not valuable resource in big companies with an affluent budget; it could be applied in start-ups in an efficient and affordable way (Sherrie 2020). Moreover, not only large-sized companies demand in data analytics but also start-ups. Start-ups need data analytics to contribute to achieving (Kara 2020). Moreover, 67% of start-ups spend on data analytics with more than 10k in budget per year (Guta 2020).

To support the roles of data analytics in start-ups, Ladley (2020) highlighted the way how data would accelerate business by exploring the different perspectives of customer relationship management. There are many fields in customer relationship management, in the scope of this study, the author would focus on segmentation and retention. Thus, the author came up with writing thesis ideas in terms of “Data Analytics Applied in Customer Relationship Management in field of Segmentation, Retention in Start-ups”.

This study aims to assist start-ups in applying data analytics tools such as Tableau to divide customers into segments based on the company’s criteria and visualize customer retention. It would be understandable that start-ups at the first stage do not gain as much as big data compared to large companies; furthermore, start-ups mostly eager to get the simple data analysis dashboards, charts to know more about how

company's business is going – not require complex models compared to large companies. Based on the segments, start-ups would identify proper strategies to manage and satisfy customers. Regarding to retention, start-ups could gain general visualization how customers keep repeatedly purchasing or churning out. Data analytics application in start-ups in customer relationship would assist companies to have a transparent picture about relationship between the company and customers.

1.2 Objectives of the study

This research aims to help start-ups in applying data analytics to segment customers and retain customers in customer relationship management.

1.3 Delimitations of the study

When concentrating on data analytics, the author would focus mainly on data analytics, the study will not include data science in the field of machine learning or artificial intelligence.

The two main sections in customer relationship management the author would target:

- Segmentation
- Retention

There are various fields in customer relationship management. As the author was working in the marketing department, working closely in the field of customer relationship management, solved the case in segmentation and retention for start-ups few times, the author would choose segmentation and retention for delimitating topics.

The main research method of this thesis is the qualitative method. The author choose the qualitative method for the advantages. The author would dig deeper into how start-ups manage to implement segmentation and retention, what the values that segmentation and retention offer. Moreover, the number of samples are limited, leading to qualitative is the best option for the research.

1.4 Research Approach

The research approach would be divided into 2 main methods - deductive approach and inductive approach. In terms of deductive approach, deductive approach is the

proccess that the author would identify research issues; based on the issues, the author would form the hypothesis and examine with the data set and samples, after that, the author would identify and validate the hypothesis based on the result of the examination. In a similar assumption, deductive approach could be generated the concepts and models in theoretical framework from various sources and based on the theories to build the hypothesis and test the validation via empirical data (Saunders et al. 2007). Regarding inductive approach, inductive research refers to the theories or concepts that are formulated and assumed through observations (Saunders et al. 2007).

In this thesis, the author suggested applying inductive for the research approach. The information for this thesis would be collected from publications, books, articles and in–depth interview.

1.5 Research Design

The in–depth interview would be based on the exploratory research design. Exploratory research design is discovering updated information and the latest knowledge by communicating with experienced people in the major field. Exploratory would be flexible and feasible for the researcher to gain more new insights when interacting face to face with experts; furthermore, the research has the flexibility to guide the experts dig deeper into the problems or other direction when recognizing new insights during the interview.

1.6 Theoretical Structure

The thesis would be divided into 4 main parts:

Chapter 1: Introduction

Chapter 2: Theoretical Background

Chapter 3: Company's interview and findings analytics

Chapter 4: Recommendation and conclusion.

1.7 Research Methods

There are two main research methods, including quantitative research and qualitative research. Regarding the quantitative research method, the numerical or statistical data would be collected and demonstrated by the graphs, correlation, relationship among variables, trends of the data by the data analytics tools (Saunders et al. 2007). Quantitative would be effective and efficient when the researcher analyzes a large number of samples. The data is mainly collected from the survey contained various kinds of questionnaires filled by the group of samples. The research would utilize the non – numerical data or unstructured data to find out the insights or information to qualitative. To be specific, non – numerical or unstructured data could be defined as narrative expression, the data would be gathered during the in–depth interview face–to–face or an online list of opened questions (Saunders et al. 2007).

Depend on the different purposes of the study, the researcher would choose the suitable research method. In the thesis, the author would choose the qualitative research method. The number of samples is not significant; numerical data or statistical data in this research is not necessary, the concentration on the experience and the methods of implementation data analytics from experts. For all reason, qualitative is the most appropriate research method to collect data from in-depth interviews face–to–face with the experts.

In this research, the author would apply the qualitative methods. To be detailed, the secondary data for this study would be extracted from legal books and trustworthy articles and the primary data is from the in - depth face – to – face interviews with 3-4 start-ups. I would choose the exploratory research method for data collection mainly carried out via in-depth interviews.

1.8 Main research question

How data analytics would be applied in Customer Relationship Management in field of Segmentation, Retention in Start-ups?

Based on the main research questions, there are 3 sub-research questions. The research aims to assist start-ups to apply the segmentation and retetion in customer relationship management; so, the author would focus on the process of how to implement the segmentation and retention in data analytics, as well as the general

process. The author would like to guide the reader step by step from the base of data analytics process to the detail process such as segmentation and retention.

1.8.1 Sub-research questions

1. How is the general process of data analytics applied?
2. How is the process of segmentation implemented?
3. How is the process of retention implemented?

2 Theoretical Background

Data analytics has boomed in this recent decade. Moreover, data analytics plays an essential role in business, either large-sized or small-sized. Data analytics is a valuable resource in big companies with an affluent budget, but it is also applied in start-ups efficiently and affordably (Sherrie 2020). Nowadays, enterprises with affluent budget use data analytics and small-sized or medium-sized companies utilize the trend of data analytics in business. Start-ups gradually apply to data analytics in business with an affordable budget (Kara 2020). Guta reported 67% of start-ups to spend on data analytics with more than 10k in budget per year (Guta 2020).

2.1 Data analytics

Frankenfield (2020) defined data analytics as a process of transforming unprocessed data into meaningful and informative data for users or audience to get insights and make the conclusion. Regarding PwC (2020), business applying data analytics in operation achieves three times performance efficiency in decision–making capacity. Data analytics is a method of tracking the vital metrics of business (Croll & Yoskovitz 2012). The metrics could demonstrate the information and value to the company's business operation such as how the revenue is flowing, how many new customers, etc. According to Forbes (2020), data analytics offers the business a method to explore useful and valuable information to foster and map out the strategies.

To sum up, data analytics would be applied to assist business in making decisions or strategies in various fields such as finance, marketing, IT, etc. The company would utilize data analytics tools to analyze and process data from raw data into meaningful data. All of the information could be discussed in the next parts. In this study, the author would dig deeper into Customer Relationship Management - a part of marketing.

2.2 Customer Relationship Management

Customer relationship management includes variety of fields such as marketing, customer service activities. The main purposes of customer relationship management are to retain customers, attract new ones, nurture relationship and build customer loyalty; in an detail, business apply customer relationship manangement in order to

build effective customer relationship, meet the requirement and expectation of customers, deliver right value proposition for customers. Srivastava (2002) formulated the customer relationship management equaled to customer understanding and relationship management. He emphasized the importance of customer understanding and the application of data analytics in customer relationship management, leading to a better understanding of customers' needs and behavior. The combination of customer relationship management and data analytics could foster long-term customer relationships (Anshari et al. 2019).

The main purposes of data analytics of customer relationship management is to identify the right segmentation for products, services, customers for business to improve the performance and customer relationship (Anshari et al. 2019).

On the other hand, customer relationship management is an effective and proactive marketing strategy concentrating on managing customer purchase journey by understanding their needs and behavior (Baran et al. 2016).

Finally, to achieve successful and high performance in business operations, the company should have effective customer relationship management to have better customer understanding and behavior through the application of data analytics (Dolnicar et al. 2018). According to a PwC survey of 1,000 business representatives, companies that applied data analytics in operation outperform their opponents in productivity and profitability by 5% and 6%, respectively (PwC 2020).

2.3 Customer Segmentation

Customer segmentation is a separation of the whole group of customers into smaller ones. Each group or each segment must be homogeneous from the others. The segment would represent the distinct values and characteristics. The primary purpose of segmentation is to deliver the right values to the right target group of customers (Srivastava et al. 2002).

The benefits of customer segmentation are to assist the business in better understanding customers’ needs and preferences through the management of smaller groups. Those benefits would lead to appropriate and effective marketing strategies. It is remarkably efficient when concentrating on the subset of the segment rather than the whole group. This would reduce the cost of marketing and facilitate the marketing

resources more effectively (Dibb et al. 2008). It may be essential for small organizations to satisfy the very distinct needs of a small group of consumers because they lack the financial resources to serve a larger market or multiple market segments (Dolnicar et al. 2018). Market segmentation is effective in sales management because this leads the sales budget to direct target groups of consumers rather than each consumer individually (Dolnicar et al. 2018).

2.4 Segementation Criteria

It realizes the differences among the groups of customers leading to success. The more business meets customers' needs, the more business increases the customer's satisfaction. Along with the understanding of customers, company would offer the values exceeded those of competitors. Moreover, customer segmentation would assist business to focus on the resources allocation in the market and put effort into the potential market with high return on investment. It would be flexible and manageable for business to segment the customers' groups based on the criteria (Gavett 2014).

In the Market Segmentation Analysis book (Dolnicar et al. 2018), there are main criteria to divide the group into different segments. The criteria of segmentation would be discussed below.

2.4.1 Geographic segmentation

Geographic segmentation is the division of groups based on the distinct geographic locations, states, cities, regions, countries, etc. For instance, business need to identify the different demands among the groups of customers in terms of distinct geographic parts in order to offer the proper services or products to the target segment.

Regarding the small-sized or medium-sized business, the geographic segmentation would be ideal and feasible for the business to segment; this is because these kinds of business lack resources or belong to e-commerce platform; moreover, medium and small business tend to allocate the resources in local markets compared to international markets for the tight financial budget and profound understand of customers’.

Depend on the differences in geography, the business could provide the optimized distribution platform and marketing strategies.

2.4.2 Demographic segmentation

Demographic segmentation is the division of groups based on the variables of age, income, level of education, status, religion, etc. The application of demographic is considerably popular for flexible and easy identification. Significantly, the demographics units influence the customers demands and needs and the business had better to gain the knowledge of the characteristics of demographic segments to identify and measure the size of target segment (Armstrong 2013 & Kotler). Demographic data is easier to collect compared to others. Furthermore, applying demographic as a metric would be flexible and straightforward such as group of customers between ages of 18 and 25. On the other hand, the similiar demographic group does not infer that the customers in the group have the same needs, interest and hobby. The marketing message would be by a “one-size-fits-all” method to the group of customers by applying the demographic segmentation.

2.4.3 Psychographic Segmentation

Psychographic segmentation is the division of groups based on the units such as lifestyle, attitudes, characteristics, etc. The marketing messages would be more attractive and appropriate to target customers by the motives and needs identification; furthermore, psychographic segmentation would offer a overall understanding of the target customers; especially, customers are highly emotional and sensitive. On the other hand, psychographic data or information is quite difficult to collect compared to the others.

2.4.4 Behavioral Segementation

Behavioral Segmentation is the division of groups based on the number of usages, number of purchases and tendencies, overall knowledge of the company's brands, the attitude towards products or services. With the development of online tools, tracking and collecting the customers behaviors are more flexible and cost – effective for the real – time period; as the result, tracking and understanding the customers’ behaviors

would be more precise and power. Based on the customers' behavior proof, the marketing messages would be more precise and appropriate.

2.5 Effective Segmentation

To evaluate an effective segmentation, the segment should meet the following conditions (Armstrong & Kotler 2013):

First, each segmentation must be distinct from the other; however, the customers in each segment share similar values.

Second, the segment should be large enough; the business must measure the size of the segmentation, examine whether the segment is worth spending the resources.

Third, the segment must be suitable with the business's resources; the business needs to have enough capacity to meet the demands or requirements of the target segment.

Fourth, customers in the segment could be identifiable and reachable. These attributes of segments could assist the marketing mix regarding customer accessibility.

2.6 Customer Retention

Customer retention could be defined as a method of preventing customers from switching to the competitors or stopping purchase products or using the company's services. According to the research, it is up from 5 to 7 times more costly to attract new customers compared to retain the existing ones. A good group of loyal customers could be an influential asset of the company in terms of profit and high – qualified image (Wertz 2018). According to Bliemel, only 5 percent increase in retention could lead to 85 percent of the company's profitability. With appropriate retention strategies, business creates profitability and reduces the amount of nurturing – loyal customers (Bliemel 1998).

Healy (1999) assumed that a high retention rate could positively lead customers to spend more on purchases and offer word of mouth. Reichheld and Sasser (1990) supported that those kinds of customers would spread favorable recommendations to other potential customers, assist business in decreasing service cost, and increase spending budget and willingness to pay when prices are changing.

2.7 Data analytics tools

Data analytics tools allow the business to access, transform raw data, present data into vivid visualization such as dashboard, charts, graphs, maps to get actionable insights. Nowadays, data analytics tools are various, offer quick, easy–to–use, friendly interface, diverse in price range also (Fruhlinger et al. 2019); With the increase in data analytics, by 2025, business would get the most information from data-driven dashboard (Gartner 2020).

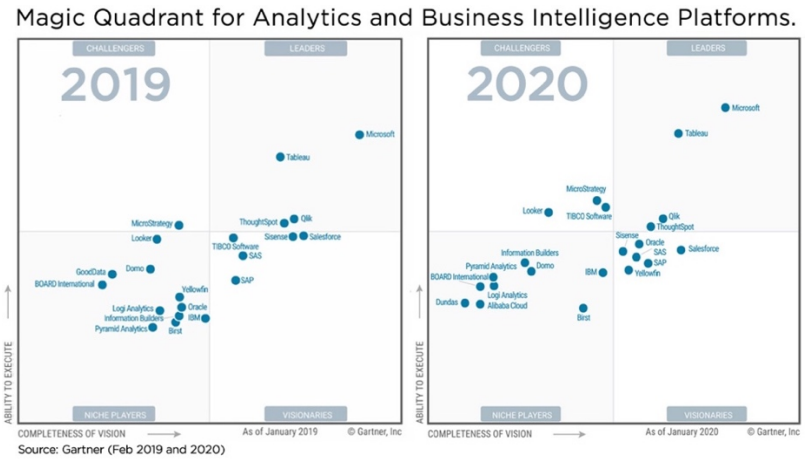

Figure 1. Magic Quadrant for Analaytics and Business Intelligence Platforms (Gartner 2020)

According to Microsoft, Tableau, Qlik are the leader in the data analytics tools market according to the magic quadrant chart above. There will be various prices and package for each data analytics tool for business to select. Depend on the financial budget and capacity, business would choose one tool to apply in data analytics (Widen 2018). Data analytics tools assist business utilizing the data exacted from different sources, and then data could be connected from the data warehouse.

2.8 Data analytics process model

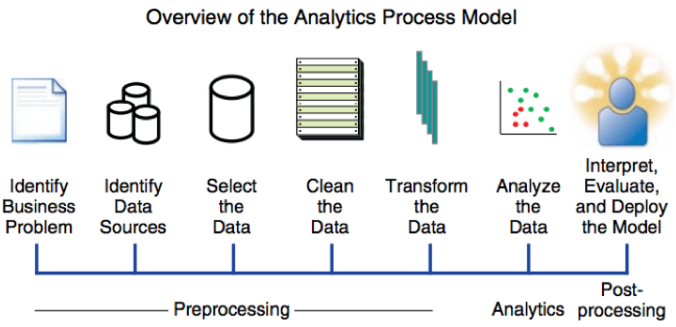

Figure 2. Overview of the Analytics Process Model (SaS 2019)

Hannah (2019) suggested the main five steps in the data analytics process model. The first step would business problem identification. With ineffectively – defined business issues, business could waste time, budget and failed to implement the campaign or plan (Spradlin 2012). Moreover, in 5 – Whys technique, business should keep asking opened and closed questions until discovering the cause of the issues. Among 5 phases of Lean Sigma, defining an issue plays an important role and is the first base to proceed to the next step.

The second step would be data sources identification and necessary data field selection. The company would identify where data sources come from. Data sources would be from the data warehouse, database, spreadsheet file, and type of csv or excel (SaS 2019). Data sources should be selected carefully to guarantee sufficient units and the availability of variables (Kornegay 2013). The primary key is that achieving the right and appropriate data that the company needs is to identify the well-defined issues and have a clear purpose understanding. The better the company defined the problems and the information, the easier and more flexible the company would locate the data resources and the necessary variables.

Brown (2016) pointed that the sources of a company would come from internal sources and new data collection. Regarding the new data collection or external data collection, when data does not exist in the company's internal database, the company would consider collecting data by carrying out the survey, installing the sensors, and outside observation. External data collection would be time–consuming and costly. However, external data would give the company more valuable and undiscovered insights.

The third step is cleaning and transforming data – called data preparation . In reality, data – preparation is accounted for nearly 80% of the data engineer's effort and time (Press 2016). Raw data in real life is usually complicated, incompleted and inconsistent. This is due to the fact that:

- Complicated data: aggregated data, inappropriate data types
- Incompleted data: missing values
- Inconsistent: duplicates, outliers

Data preparation would improve the quality of datasets. Tableau listed some benefits of cleaning data :

- Boost the efficiency and productivity of business daily activties in data process when there are no duplicates, errors or inappropriate data types.
- Achieve better decision – making with high - quality data.

After completion of cleaning data, company should check the criteria to ensure the quality of data:

- Validity : no empty columns, certain data type in a column, foreign key or primary key contraints, regular expression
- Accuracy: Ensure data close to true values.
- Completeness: the level of the known data
- Consistency: data consistent with the same dataset.
- Uniformity : level of which data in the same unit of measure

In data preparation process, a data analyst would process the data in the following patters, such as:

- Filling the missing values with techniques
- Handling wrong data types, deleting outliters and duplicate
- Joining other dataset into a completed dataset through primary key and foreign key.

The important thing to analyze data effectively and efficiently is to keep raising questions and asking the proper data analysis questions. Questions could be unlimited

to be raised; however, time and resources are restricted. Regarding data analytics, if data analysts have poor–qualified, ill-conceived or misleading questions, focus on creating charts or dashboards without asking or checking whether the data gives meaningful insights or not. It leads to the expensive and time–consuming data analytics process with any result. The more insightful and actionable data will be, the better quality of the question. The more specific question is, the more influential the data visualization process will be.

2.9 Clustering introduction

Clustering is one of the most appropriate models to apply in segmentation. (Kashwan 2013). Troy also suppose clustering is the standard model for segmentation flexibly (Troy 2019). Clustering is an analytical technique that gathers objects with similar attributes into a group. Furthermore, the clustering process is based on machine learning algorithms. The algorithms would utilize the distance technique to measure the similarity (Tableau 2021). There would be other segmentation methods such as Factor Analysis, Factor Segmentation, Discriminant Analysis, K-means or Hierarchical clustering, Latent Class Segmentation (Horn et al. 2021). In this research, the author would select K-means method.

2.10 K-means model

Clustering segmentation is applied k-means model and centroid approach. The k-means clustering is an unsupervised machine learning algorithm. In a specific way, the model would use the unlabeled data and group the data into clusters. In each cluster, there will be a centroid that is the center of each cluster (Tableau 2021). The number of clusters would depend on the Elbow method or business decision based on the statistics method (Troy 2019).

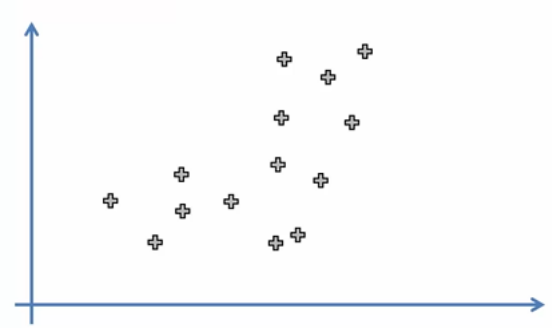

Figure 3. Data point represented in scatterplot (Medium 2021)

The k – means model process would be:

1. Number of clusters are chosen based on the Elbow method or other methods.

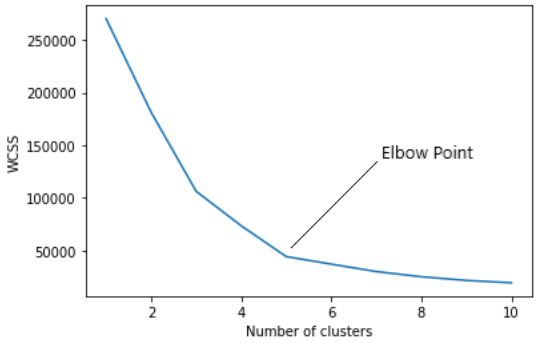

Figure 4. Elbow Method (Medium 2021)

2. The centroid is the mean value of all data in the segment. The model would calculate the means of the clusters and based on the means, the classification would be divided depending on the distances to the centers. In the figure, the number of the centroid is two, respectively, the blue centroid and the red centroid.

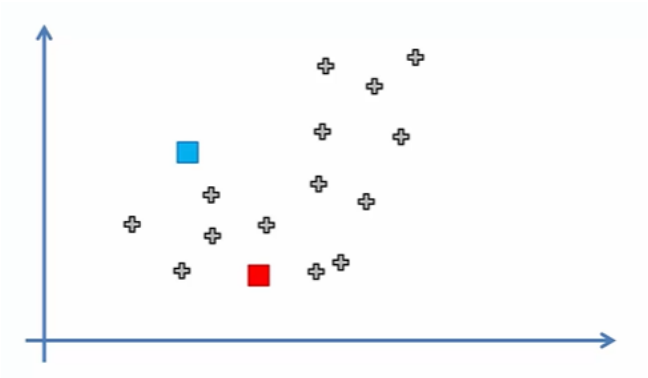

Figure 5. The two centroid – the red and blue one (Medium 2021)

3. By applying the Euclidean to calculate the distance of the points to the centroid, the data points with the closest distance to the centroid would be gathered into a cluster. On the other approach, the median line would be drawn by connecting two centroids with a line and find the center of the line. The data points would be divided.

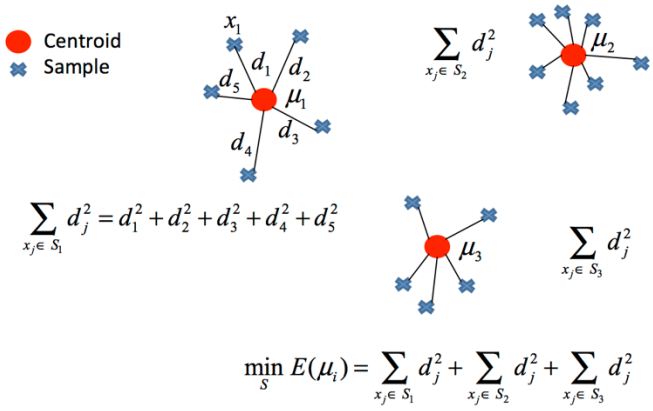

Figure 6. Euclidean distance measurement calculation (Medium 2021)

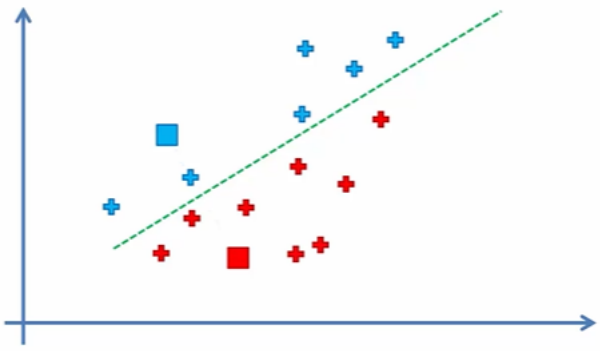

Figure 7. Median line drawed between 2 clusters (Medium 2021)

4. After the end of the process, the centroid of the clusters would be calculated and the classification is split again. The process could be replicated multiple times until the mean of the cluster will not be changed considerably after the iterations (Tableau 2021).

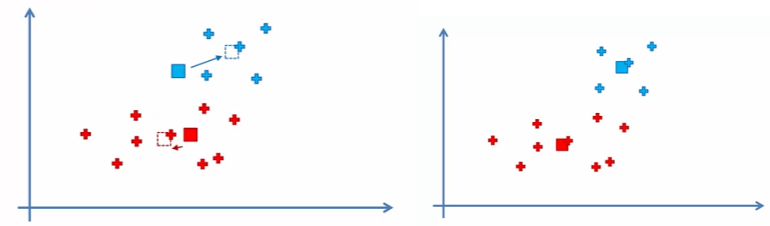

Figure 8. Centroids were calculated and classification is spli again (Medium 2021)

5. After the iterative process end, the model is ready. The data points are grouped into two clusters.

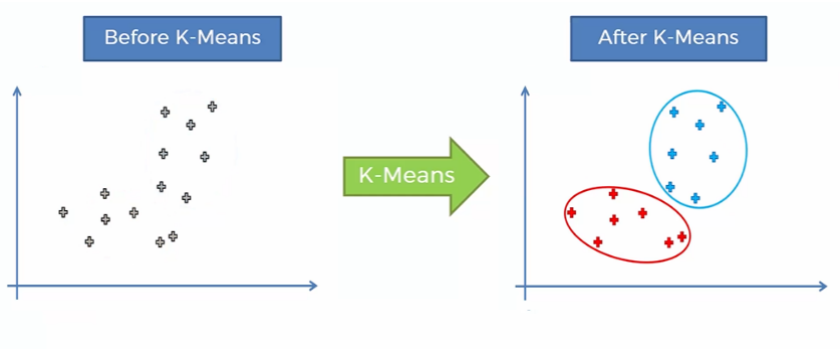

Figure 9. Before and after K - means application for segmentation (Medium 2021)

2.11 K-means cluster application case

There is a common case that applied K-means clustering. The case is to segment customers into groups based on the spending scores, annual income of a shopping mall dataset. Based on the cluster with the figures of spending score, annual income and age, the business would identify the customer segmentation and label the segmentation (Kaggle 2018).

First of all, the business had better determine the number of clusters. The number of cluster would be based on the Elbow method or other methods. After determining the number of clusters, the dataset would be processed by applying k – means clustering.

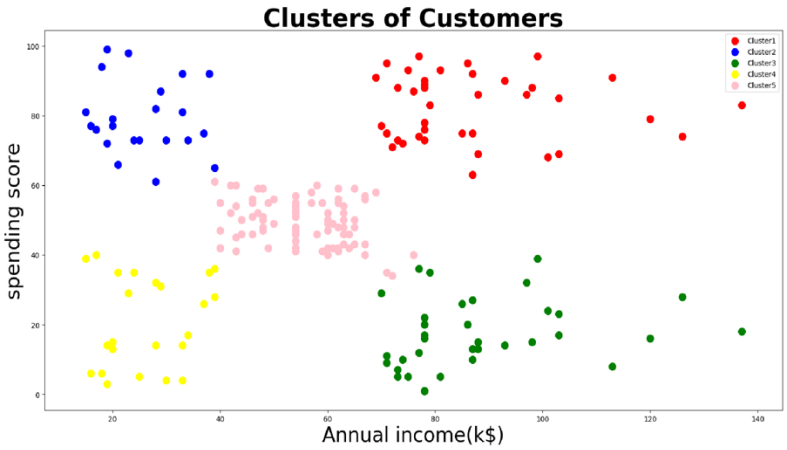

Figure 10. Clusters divided by the spending score and annual income measurement (Medium 2021)

From the figure 10, there are 5 segments with different colors in the model based on the measure of spending score and annual income.

Regarding cluster 4 – the yellow segment, the segment is categorized by the low annual income and low spending score. The business would infer that customers with a tight financial budget would spend less and save more.

Regarding cluster 2 – the blue segment, the segment is categorized by the low annual income and high spending score. The business would reason that these customers have a hobby to go shopping, they purchase products frequently even though the annual earning is low. The business may not target this group effectively since the group would not be stable in the financial budget.

Regarding cluster 5 – the pink segment, the segment is categorized by the average income and average spending score. This segment may not be the priority target of the business; however, the business would map out strategies to increase the spending score.

Regarding to cluster 1 – the red segment, the segment is categorized by the high income and high spending score. As a result, this segment would be an ideal and target group for the business. This segment would bring more values and profits. The business would try to enhance the budget to attract new customers in this segment and nurture long–term relationship with this segment.

Regarding to cluster 3 – the green segment, the segment is categorized by the high income but low spending score. The business would assume that this group may be unsatisfied with the business services or products. This group would be the potential if the business would figure out the reasons why the spending score is low; from the point, business would improve the business performance to gain the attraction and build relationship with the target group.

2.12 Cohort analysis introduction

Cohort Analysis is a model to demonstrate the purchase behavior of customers or a group of customers over time, illustrate the first time purchase of customers and their next following purchases. A cohort level of time could be based on year, quarter, month or week to observe the flow of number of customers back to business (Duggal 2021). Cohort analysis assists business in analyzing customers' behavior throughout the life cycle of each individual. Business utilizes the cohort analysis to get more insights and achieve the understanding of the trends of customers over the period of time (CFI 2021).

The date data field is vital and indispensable to apply the cohort analysis model. Date field would be order date, sign–in date, posting date, etc. The date field is necessary for this model to calculate the first date that the customer purchased, signed in, posted; based on the data, the next purchase calculation and the gap between the period of repeat purchase would be computed.

To track and supervise retention, a cohort retention chart is commonly used to illustrate how the retention rate is moving.

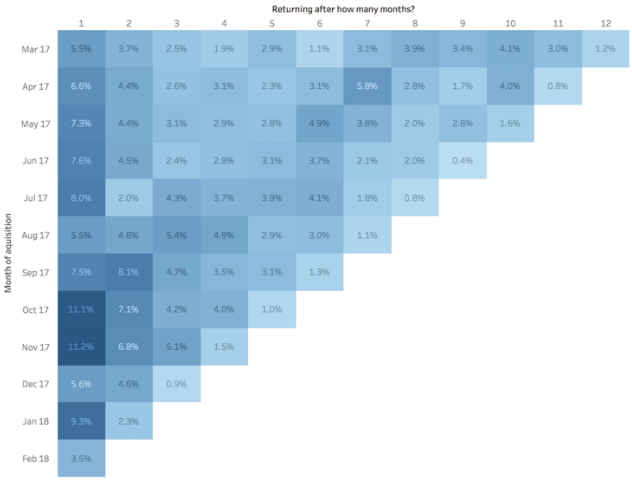

Figure 11. Cohort Retention chart visualized by Tableau (Banutoiu 2021)

The Y-axis illustrates the months that the customers carry out the first purchase.

The X-axis illustrates the number of months after the customers made their first purchase.

The figures in the square illustrates the rate of return customers in each month.

The insights could be read both based on the horizontal and vertical dimension - (Banutoniu 2019).

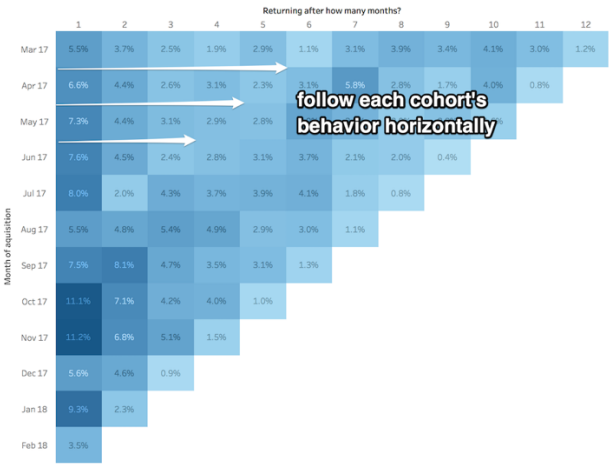

Figure 12. Cohort Retention Chart in horizontal axis (Banutoniu 2019)

Reading from the horizontal axis, the rate of repurchase would be distributed through months to track the rate of repurchase in specific month. From the rate, the business

would identify how many months after the first purchase, the customers would churn out (Banutoniu 2019).

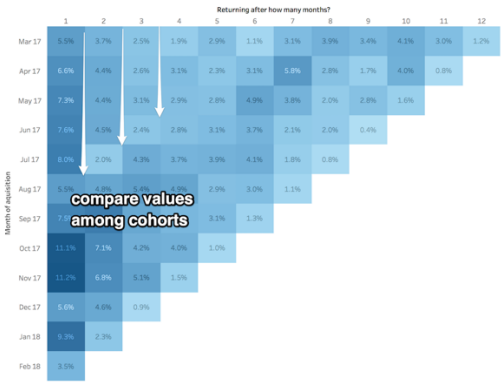

Figure 13. Cohort Retention Chart in vertical axis (Tableau 2019)

Regarding to vertical axis, business would compare retention rate among the cohorts in the month that customer repurchased (Banutoniu 2019). In general, cohort retention infers the flow of customers repeated purchase after the first purchase over the time period and compare the retention rate over each period of time and get the conclusion of how retention works.

2.13 Cohort analysis application in business

Su (2018) suggested a case that customer X was given a voucher of 50 percent discount entering to online store 6 months ago to purchase an item. Su (2021) supposed that a business owner would raised questions such as whether this kind of customers would come back to business after the purchase with a discount, how long it would take a customer to come back to business and the values of these customers bring to the company. To dig deeper and figure out the issues, the business implemented to track this kind of customers to examine if they came back to make a purchase and the frequency of their purchase.

First, the company would filter out the target customers visiting during the specific period of promotion time and made purchases with the discount program and the date timeframe of the promotion program. Consequently, after applying cohort analysis, the result is 70 percent of customers who made a purchase with voucher did not come back to business, while 20 percent came back to the online store at least a few times but did not make a purchase, and the rest purchased in these recent months.

The business would build assumptions such as:

- Customers may not be interested in the products.
- Customers may forget the company business because there are various businesses outside and a bunch of information.
- Customers may encounter the user experience in the website such as the shipping page.

After identifying the reasons that may lead to the low retention rate, the business would map out the strategies:

- Launching a campaign to retarget the churn out customers to remind customers to come back to business and make more purchases
- Investigating whether the shipping cost is the hinder for customers to make a purchase. If this assumption is correct, the business had better to launch a free-shipping campaign for those customers.

The case above is the application of cohort analysis in business, the cohort analysis assists business to track the life cycle of customers and carry out concrete actions to prevent the churn rate and improve the retention rate.

2.14 Types of Cohort Analysis

Acquisition Cohorts

The acquisition cohort classify the users or customers based on the acquisition date and time that the customers did sign up for the first time (Duggal 2021). The movement of the rate of retention customers in next following date could be tracked to gain the insights and measure the customer lifetime values, retention rate or churn rate (Qiao 2019).

Acquisition Date	Users	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Jan 25	1,098	100%	33.9%	23.5%	18.7%	15.9%	16.3%	14.2%	14.5%	13.3%	13.0%	12.1%
Jan 26	1,358	100%	31.1%	18.6%	14.3%	16.0%	14.9%	13.2%	12.9%	14.5%	11.3%	
Jan 27	1,257	100%	27.2%	19.6%	14.5%	12.9%	13.4%	13.0%	10.8%	11.4%		
Jan 28	1,587	100%	26.6%	17.9%	14.6%	14.8%	14.9%	13.7%	11.9%			
Jan 29	1,758	100%	26.2%	20.4%	16.9%	14.3%	12.7%	12.5%				
Jan 30	1,624	100%	26.4%	18.1%	13.7%	15.4%	11.8%					
Jan 31	1,541	100%	23.9%	19.6%	15.0%	14.8%						
Feb 01	868	100%	24.7%	16.9%	15.8%							
Feb 02	1,143	100%	25.8%	18.5%								
Feb 03	1,253	100%	24.1%									
All Users	13,487	100%	27.0%	19.2%	15.4%	14.9%	14.0%	13.3%	12.5%	13.1%	12.2%	12.1%

Figure 14. Acquisition cohort retention chart (Makhija 2020)

Behavioral Cohorts

The behavioral cohort segment the users or customers based on the behavior action such as the first date the customers carried out the transaction or made a purchase (Duggal 2021). According to Amplitude Analytics (2016) emphasized that knowing the date customers started is pivotal but the flow of customers actions and engagement could give the business more understanding of customers' behaviors and interactions with the company. For instance, the customers' activities could be sharing a photo, downloading songs, purchase coins, etc. within the period that the company could specify (Makhija 2020).

Cohort	Users	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Total Users who did a Transaction	1400	100%	43%	13%	20%	25%	0%	0%	0%
Jan 31	50	100%	0%	0%	0%	0%	0%	0%	0%
Feb 01	100	100%	100%	0%	100%	100%	0%	0%	
Feb 02	200	100%	50%	0%	0.00%	0%	0%		
Feb 03	400	100%	25%	25%	0.00%	0%			
Feb 04	200	100%	100%	50%	0%				
Feb 05	400	100%	25%	0%					
Feb 06	50	100%	0%						

Figure 15. Behavioral cohort retention chart (Makhija 2020)

In genreal, acquisition cohort illustrating when customers churned out after a period of time and behavioral cohort assist business to get insights what reasons customers churned out.

3 Company findings and analysis

3.1 Company Introduction

In this study, the author would invite two start-ups for an in-depth interview. The interviews were carried out face–to–face for almost 1 hour. All of the information during the interview was written down. These two start-ups have applied data analytics in business in recent years, especially in customer retention and segmentation to improve Customer Relationship Management performance.

The author interviewed with :

1. Mr. Nguyen: Head of Data Analytics in The Coffee House
2. Mr. Kavcic: Head of Finance in Eurosender

The Coffee House

The Coffee House is a coffee shop start-up in Vietnam. The Coffee House established in 2014, until now, there are around 150 The Coffee House branches throughout Vietnam. The Coffee House is one of the most popular coffee shops in Vietnam. The Coffee House offers a modern, convenient working atmosphere. The website of the company is https://www.thecoffeehouse.com.

The Coffee House applied the data analytics field into business five years ago. In the first period, the primary purpose of The Coffee House's data analytics application is similar to Eurosender's. Moreover, The Coffee House hired only one data engineer. After two years, with the achievement of data analytics application, The Coffee House spent a large amount of financial budget on data analytics. The data analytics team of The Coffee House included a data engineer, data scientist and data analyst. With the support of the high–skilled data analytics team, The Coffee House achieved successful results in data analytics in two recent years.

Eurosender

Eurosender is a start-up in the field of Logistics. The company was established in 2014. The headquarter is in Luxumbourg and the main offices are in Luxembourg, Ljubljana and Kiev. Eurosender offers a modern digital logistics platform. The website of the company is https://www.eurosender.com/en/.

Eurosender applied the data analytics field into the business three years ago. The primary purpose of Eurosender in the first period of data analytics is to create the daily dashboard to

track the vital KPI and analyze data to get insight based on the company's requirements. With a tight budget of finance and years of the business' beginning, Eurosender does not hire a data analytics team including a data engineer, data analyst or data scientist; instead, all of the data analytics tasks were carried out by the head of finance.

3.2 Findings and Analysis

In this section, the author carried out the in – depth interview with Head of Finance of Eurosender and Head of Data Analytics of The Coffee House.

3.2.1 The comparasion between the current and previous data analytics tools

Regarding to Eurosender, the current data analytics tools applied is Power BI. The reasons are:

- Power BI is one of the leading data analytics tools
- Power BI is one of the part of Microsoft suite. At this time, Eurosender is using the Microsoft ecosystem software, by applying the Power BI, the cost for data analytics tools would be an ideal for affordable budget.

Regarding to The Coffee House, the current data analytics tool is Tableau. The reasons why The Coffee House select this tool are:

- Tableau is one of the most popular data analytics tools
- Tableau offers machine learning library
- Tableau offers various connectivity data sources.
- Tableau offers library to builld the calculation easier

Eurosender and The Coffee House were using Excel as a data analytics tool in the past. By comparing the difference with PowerBI, Mr. Kavcic assumed that before Power BI, most of the analysis was completed in Excel, there were arising issues and difficulties. For example, it is quite hard to manage big data, connect with other datasets or data sources from the database. Moreover, compared to Excel, PowerBI offers function library such as DAX library to build formulas easily and flexibly. In Excel, to build a cohort retention model is quite complicated in building formulas. Regarding The Coffee House, Mr. Nguyen shared that Excel almost did data analytics. The

concept of data analytics is not common and quite new in Vietnam. The company carried out the data analytics tasks quite simply by using popular and basic charts along with the support of Pivot Table in Excel. Moreover, data analytics tasks would be completed more effectively with the support of the library function. Regarding the speed, Tableau solves complicated formulas and large datasets quicker than Excel. For example, the cohort retention would be done in Tableau easily and quickly with the support of Level of Detail library, compared to Excel; it would cost time and effort to write the complex formula. Moreover, if the dataset is large, the computer would be frozen when being implemented in Excel. On the other hand, segmentation is carried out in Excel is normal and not problematic.

3.2.2 The figures or information to identify the business issues

Based on which figures or information, the company would identify business problem. Mr. Kavcic denoted that by understanding the business operation and process and the business key drivers comprehensively, the company would diving into the key drivers individually to examine which insights or issues the business would explore. Additionally, Mr. Nguyen supported that a good understanding business operation is a vital base. Moreover, a data analytics team should work closely with other departments, generate information of issues and identify which issues are the target to dig deeper into. Depend on departments, there would be different KPIs and requirements. For example, the staff in different departments sometimes do not know what exactly the issues should be identified. When cooperating and discussing with the data analytics team, the data analytics would instruct and explore in various ways to have a better business issues identification.

3.2.3 Internal sources and external sources collection

In terms of internal data sources, Eurosender gathered data from order management software, multiple SaaS sources, google analytics, CRM tools, Pipedrive. For the external data sources, Eurosender would gather from the survey. In some cases, regarding customer service, the marketing department would cooperate with customer service to carry out the survey to gather to experience from customers. On the other hand, The Coffee House gathered internal data via the data warehouse by building a pipeline that gathers data from Point of Sales, Pipedrive, etc. For external data

collection, The Coffee House collected from surveys and feedback from customers. There was a case 1 year ago, The Coffee House installed the IoT such as Rasberry Pi to gather external data. By scanning the unique barcode of the cup of tea or cup of coffee to measure the time to complete and inform customers about the process. The main aim is to provide a better customer's experience with the shortest time of waiting.

Figure 16. Time Measurement of making beverage (Nguyen 2021)

3.2.4 Data preparation applied in company

Both Eurosender and The Coffee House using tools to process the raw data into clean and usable data. For Eurosender, the company uses PowerBI functions and internal self – written code in database to obtain the clean dataset, the process is almost done by the software developer. However, Eurosender does not own the complete the data warehouse to gather data sources from different sources from each department, most of the data files are extracted from the database and exported into csv or excel file. On the other hand, The Coffee house invested a complete data warehouse that pour data from different sources of each department. The Coffee House utilized data preparation library of Python, the dataset was processed and prepared when data was being poured into the pipeline of data warehouse.

For the criteria, Mr. Kavcic emphasized that the criteria to evaluate the clean and completed dataset are:

- No duplicates
- No outliers
- Data field figures in the correct format

On the other hand, Mr. Nguyen suggested that the size of the dataset should be large enough. Moreover, the dataset should cover the wide time range suited for the business questions and the data sources must be correct. Mr Nguyen explained that

there were many errors and bugs in system development that may lead to incorrect data records.

For Eurosender, the company does not have a complete the data warehouse or data engineer to process the complicated, raw data and combine datasets from different sources. It leads to the drawback in data analyzing. For example, the company want to analyze the data of calls with customers from the customer service department from the customer relationship management. However, the data sources and data field are quite complex to process to a clean and usable dataset. As a result, the project was delayed.

3.2.5 Dashboard Evaluation Criteria

According to Mr. Kavcic and Mr. Nguyen, the dashboard would clearly show the vital and necessary KPIs. The dashboard does not need to include a lot of charts or numbers. The dashboard should be clear and easy – to – use. Mr. Kavcic also emphasized that the dashboard could be shared to other staffs in different departments and some of staff are not professional at using the dashboard. Moreover, Mr. Nguyen constantly asked the business users to give feedbacks how they were using the dashboard and whether there were issues needed to improved the performance of the dashboard.

3.2.6 Skills required in data analytics

Mr. Kavcic assumed that besides being good at or fluent in technical skills, a data analyst should have a understanding of the business operation and process. From the base, the data analyst would identify the essential KPIs and information needed to analyze by keeping rasing questions. For Mr. Nguyen, first and foremost, the data analyst needs to understand the business domain he or she is in charge of. Only domain knowledge may lead to good data exploration and analysis. Mr. Nguyen also emphasized that the skills of asking questions or data exploration is considerably essential. A data analyst completes the requirement of data analysis from departments and offers more valuable uncovered insights. Other data skills in using software like Tableau, Power BI or Qlik are the tools to perform the tasks; data analysts would easily get used to the tools.

3.2.7 Segmentation application in company's case

Mr. Kavcic said that using a cluster chart is the best way to visualize the segmentation. Regarding figures or data that the company decides the number of segments, Mr. Kavcic said that by estimating the percentiles, the company would determine the number of segments. For example, the case of the company is dividing customers into segments based on the number of orders. Mr. Kavcic was created a percentile bar to examine how the number of orders is distributed. Based on the figures of percentile bar, Mr. Kavcic decided to divided customers into three segments, including groups of customers from 10 – 20 orders, group of customers from 21 – 40 orders, group of customers from 40 orders above.

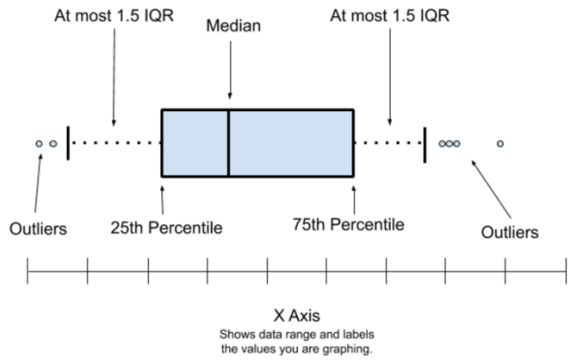

Figure 17. Box plot demontrates the percentile (medium 2021)

By dividing customers into segments, the company would easily manage the customers. Moreover, the company would identify groups with high lifetime value and strong potential; as a result, the company would spend more time and budget to nurture the relationship with those groups. For The Coffee House, Mr. Nguyen supposed that depends on the business situation analytics department will decide the number of K – means. Moreover, The Coffee House processed segmentation in Tableau, the team also carried out segmentation in Python with the K-mean cluster machine learning library. After the algorithm implementation, the elbow method gives the number of clusters. Most of the time, The Coffee House would be based on the elbow method number of clusters, but if the number is too high, the clusters would become too small and business would not be able to focus on the resources to each cluster.

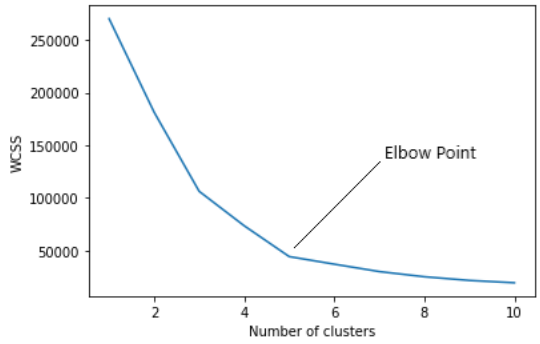

Figure 18. Number of Cluster in Elbow Method (Medium 2021)

There are two recent cases that The Coffee House apply to segment. The first case would be based on the delay time of beverage delivery to customers' location, the bonus of vouchers would be varied. The second case is segmentation based on the profit, there was a segment that turned out to be loyal but was not profitable since they constantly hunt for coupons. For this segment, the company reduced the number of coupons delivered to them to reduce our cost, while focused on a more profitable segment and spent more resources on valuable customers.

3.2.8 Customer's cohort retention application in company's case

Based on the cohort retention, Eurosender and The Coffee House calculated the average retention rate. Based on the retention rate, the Customer Lifetime Value could be calculated. There are many Customer Lifetime Value formulas, it depends on the company's decision, at Eurosender, the company chooses the Customer Lifetime Value formula the include the retention rate. Here is the formula :

$$CLV = \frac{1}{ChurnRate} \times ARPA$$

Figure 19. Customer Lifetime Value formula based on the churn rate (medium 2021)

As the result, Eurosender could figure out the reasons why customers churned. Before, Eurosender also listed the causes from logistics service third partners, customer services, or the experience of the orders from websites, price variety of couriers, etc. Now, based on the feedback of customers, Eurosender identified the reasons and mapped out the strategies to improve and reactivate churned customers.

Regarding to The Coffee House, the company also listed the reasons leading to churn rate such as customer service, time waiting for orders, the quality of beverage, the process of online orders. After collecting feedbacks from customers, the company launched strategies and offered the vouchers or bonus points if the customers encountered bad services as an appology. Consequently, both companies gain the effectiveness and better result when customers comeback to business, the companies also measure the rate of comback customer after the campaign, The Coffee House gained around from 15 % to 30% of. For Eurosender, the percentage was not revealed.

For The Coffee House, Mr. Nguyen did not reveal the formula. Both companies usually select the time period such as a week, month, a quarter or a year to track the retention. Moreover, the companies would identify the length of lifetime of customers. From the figures, the company would get insights and conclusions on how customers' repurchase behavior is, whether the company gains the ability to attract new customers, and pursuade old customers to carry out repurchase. The companies would map out the strategies to prevent the churn of a large number of customers, reactivate the churned customers. In case the retention is low, the marketing department cooperates customer service department to have the strategies to boost the retention. For example, Eurosender realized that there was a period that B2B customers churned out dramatically after 3 months of their first purchase, the customer service department got a list of those churned out B2B customers, the staff was sending email or give a call to customers with bonus and voucher to pursuade customers back to business. Similarly, Eurosender and The Coffee House also sent an email to ask for the reasons that customers churned. Eurosender and The Coffee House used acquisition and behavioral cohort. According to these companies, they want to know the sign–up and sign–in behavior and the transaction or purchase behavior. For Eurosender, based on the segmentation above, for each segmentation, Eurosender would create cohort retention and track the retention rate of each segmentation. Mr. Kavcic assumed that the cohort retention of each segmentation would give the company a detail picture of the customer's retention.

4 Conclusion and recommendation

4.1 Conclusion

Based on the in-depth interview above. It is quite obvious that the application of data analytics in segmentation and retention in start-ups brings values, a comprehensive understanding of customers' behaviors and fosters customer relationship management. Moreover, the interview and the sub-question answers provide the process of how to implement the data analytics process about business issues. The interview start–ups also provide real-life projects, the issues that start-ups encounter and the perspectives.

4.2 Answers to sub-research questions

4.2.1 How is the general process of data analytics applied?

Regarding sub-chapter 2.8, the data analytics process model was illustrated based on the SaS software document (2019). There were including primary 7 steps in the process. In the interview, both Eurosender and The Coffee House applied the model in the data analytics process, they assumed that the model was the common and the general one. The process followed step by step as a business in real-life was applied. The most important and first step is to identify the business problem. How to identify business problems effectively and correctly would stem from the deep and comprehensive understanding of business and the key drivers; based on the understanding, the company would map out raise questions. Regarding the data sources, it would depend on the development and the investment resources into the data warehouse. Start-ups with a tight budget mostly collected from the database; specifically, the data mostly was poured from api sources and operational system. Small start-ups lack of the generated and united data sources from different departments and customer relationship management software; consequently, small start-ups got difficulties in analyzing data and findings business insights when lacking of the united and complete dataset as Eurosender's case mentioned above. On the other hand, start-ups invested in data warehouse and data engineer human resources, the data would be generated from various sources such as websites, apps, sensor and Enterprise Resource Planning, Point of Sales system sensor and other operational system and processed into a complete dataset ready to extract. As a

result, the company would solve more cases by exploring and gathering external data via different tools.

Data preparation is essential in the data analytics process to analyze effectively. Depend on the company resources, the tool to process the data is different. Eurosender utilized the Power BI tool to process the raw data; however, Power BI has a limit to clean and transform data. The Coffee House applied Python library; moreover, the library in Python specialized in processing raw data into usable data. It helps The Coffee House analyze data more effectively and time–saving when the dataset is clean and usable. Consequently, Eurosender had a limit to solve complicated datasets to analyze data and some cases were pending. In the data analyzing stage, finding the answers to meet the business requirements and discovering uncovered data are important; moreover, understanding business and key drivers would be important in analyzing data; based on the understanding, the analyst would be able to customize the dashboard with main, concrete and vital information, leading to the insights would be received by the users more efficiently. After deploying the data analytics dashboard or model, the data analytics team evaluated the model or dashboard by asking for the feedback of the users whether the dashboard expressed vital key performance indicators or the dashboard was easy-to-use; moreover; the data analytics team would also examine the reliability, validity of the model or dashboard.

In conclusion to the first sub–question answer, to have an order and efficient data analytics process, it is necessary to follow step by step along with the knowledge of business operation and the investment in data analytics resources, the effectiveness in data analytics process would be optimized.

4.2.2 How is the process of segmentation implemented?

Based on the application case of segmentation above, the process of segmentation still follows the data analytics process. First, the business would identify the problem; in detail, the business requires to divide the customer segmentation and the number of segments. After identifying the issue, the business would collect the dataset from certain data sources and clean data to serve the analytics purpose. Depend on business requirements, business would decide to choose different models to segment. K – means model is one of the most common segmentation models and the mechanics

of K – means model is explained in the sub – chapter 2.10. To decide the number of segmentation, the number would be calculated and measured by the statistics method or Elbow method. Depend on business, the data analytics team would choose a different method to decide the number of segmentation. For instance, Eurosender selected percentile method to determine the number of segments; on the other hand, The Coffee House ran Elbow method to choose the number of segments; after identifying the appropriate number of segmentation, the K – means model is ready to run and give the result after the iteration completed.

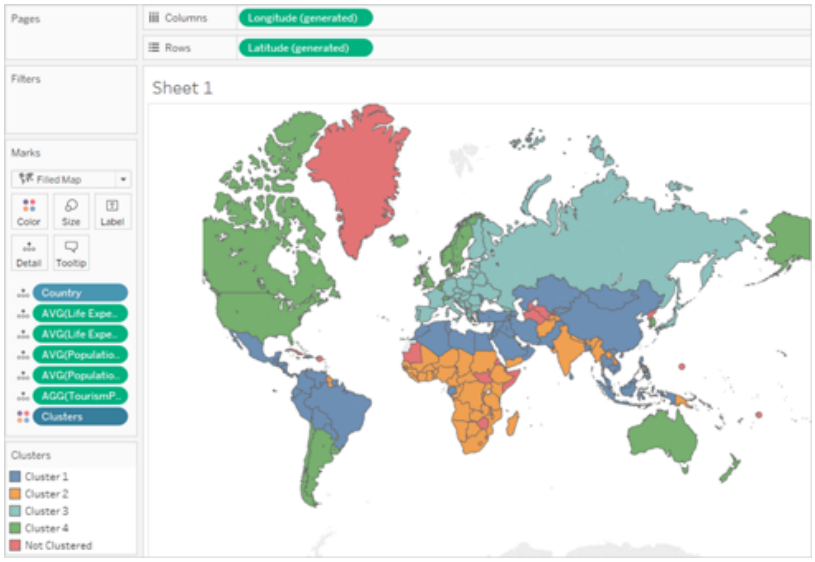

Figure 20. The segmentation illustrated in Tabelau (Tableau 2021)

In conclusion, the process of segmentation still follows the general data analytics process. The difference is the ways of analyzing the data, for different analytics purpose or result achievement, the model of segmentation would be varied based on the optimization and productivity. Business applied segmentation for the different cases. As the case of the two companies, segmentation assists companies to manage group of customers flexibly, launch appropriate program or campaign to groups and optimize the resources to certain groups.

4.2.3 How is the process of retention implemented?

The process of retention implementation follows the general data analytics process as well. The retention model would be adjusted based on the granularity of time period such as week, month, quarter and year. Both companies agreed that with the development of tools and models, creating a cohort retention model is not complicated

thanks to the library formula of the modern data analytics tools compared to Excel. For instance, in Tableau, the Level of Detail (LOD) library formula or the Data Analysis Expression (DAX) in Power BI are the powerful library to create cohort retention. Additionally, the main data field to create a cohort retention model are the number of customers and the order date; as a result, the data fields to create cohort analysis are standard and easy to extract from the database. Furthermore, nowadays, Google Analytics supported the cohort retention chart on the platform, the users could flexibly track the flow of customers over the time period. Based on the cohort retention chart, the business would able to calculate the retention rate, churn rate. From the calculation, both Eurosender and The Coffee House calculated the customer lifetime value to gain more insights about the customer groups. Based on the retention rate and customer tracking flow, the business would gain a comprehensive picture and insights to map out the optimized strategies.

4.3 Conclusions for the answers

Based on the answers to sub-questions above, it is obvious that data analytics process model is the standard process to be applied for any business data analytics cases. Moreover, with the development of data analytics software, the cases would be solved easiser and data would be processed faster, new models and charts applied to gain more insights. Additionally, the data analytics could be applied in start-ups easier and more flexibly. The start – ups do not have to invest a large amount of in data analytics at the first stage, they build dashboard and analyze data from the basics with the support of free or affordable data analytics softwares. However, with investment in data analytics resources as the case of The Coffee House, start-ups could gain better result after the beginning years. With all of the insights extracted from segmentation and retention, the business would increase the performance and productivity of customer relationship management. Furthermore, the cooperation among the departments is fostered. The data analytics team provided insights about the customers to marketing department, customer services department and even information technology department, those departments could get insights and understanding of customers. So, the departments could launch program or campaign to nurture and manage the relationship with customers. For example, sending the notification or email to churn out customers, vouchers given to customers based on

delay–time classification, resources and budget optimazation for the high-profit and valued customers, customer segmentation based on sales, cohort retention tracking per customer segmentation, etc. With the potential and benefits of data analytics offers values to the companies.

4.4 Reliability and validity

In the research, the author apply qualitative research method including the high-qualified data sources and up – to – date information to the time period of the research. Regarding the empirical section, the author apply qualitative research methods with the theoretical background gathered data from secondary sources and supported by the in-depth interviews for the primary data. For the secondary data, data was extracted from publication, articles, etc. The primary data collected from the in-depth interview via face-to-face and the information was written down. Thus, the reliability of the research is in acceptance level .

In terms of the short-comings, the information of the interview would change over the time, this is due to the fact that along with the development of data analytics application, the updated models or methods could be released. Moreover, the number of samples for the interview is small; including 2 people specialized in data analytics. It is not evident to reflect the comprehensive situation and represent for the whole population. The validity of the research is quite weak.

4.5 Recommendation

Generally, the research includes the information and steps that start-ups applied data analytics in segmentation and retention. Moreover, the ways that start-ups get the insights from the segmentation and retention to increase the performance of customer relationship management. The data analytics tools or the method of segmentation and retention implementation could be updated by the time. Moreover, the number of samples for the interview is limited, it would be better to have more samples for the interview. For further research, the research could focus on more the application of segmentation in machine learning and the ways to improve the retention rate.

List of figures

References

Amplitude Analytics. 2015. Why You Need Cohorts To Improve Your Retention. https://medium.com/. Accessed on 10 April 2021.

Anshari, M & Almumnawar N, M & Lim A, S & Al-Mudimigh, A. 2019. Customer Relationship Management and Big Data Enabled: Personalization & Customization of Services. Universiti Brunei Darussalam.

Armstrong, G. & Kotler, P. 2013. Marketing: An Introduction. The 11th revised edition. United States: Pearson Education Limited.

Banutoiu, D. 2019. Cohort Analysis: How To Study Cohorts For Actionable Insights. https://canonicalized.com/. Accessed on 10 April 2021.

Brill, B. 2000. The Close Connection Between Relationship Marketing and Marketing.

Brown, J. & Beltrani, F. 1989. Consumer Complaining and Word-of-Mouth Activities: Field Evidence.

Brown, M. 2016. How and Why To Get The Data That You Need. https://www.forbes.com/. Accessed on 5 April 2021.

Burns, S. 2020. 3 Ways Data Is Helping To Improve Business Efficiencies.

CFI. 2021. Cohort Analysis. https://corporatefinanceinstitute.com/. Accessed on 5 May 2021.

Clark, M. 1997. Modeling the Impact of Customer-Employee Relationships on Customer Retention Rates in Major UK Retail Bank, Management Decisions.

Croll, A. & Yoskovitz, B. 2013. Lean Analytics Use Data To Build A Better Startup Faster. O'Reilly Media.

Dar, P. 2020. What's the best Business Intelligence and Analytics Tool in The Market? https://www.analyticsvidhya.com/. Accessed on 25 Match 2021.

Data Cleaning: The Benefits And Steps To Creating and Using Clean Data. https://www.tableau.com/. Accessed on 4 April 2021.

Dolnicar, S & Grun, B & Leisch, F. 2018. Market Segementation Analysis Undertanding It, Doing It, Making It Useful. Springer nature Singapore.

Duggal, A. 2021. Understanding Cohort Analysis: A Comprehensive Guide. https://hevodata.com/. Accessed on 10 April 2021.

Dunkley, T. 2019. Customer Segmentation Using Clustering. https://medium.com/. Accessed on 3 April 2021.

Feihua, Q. 2011. Customer Retention in E-commerce Business. Haaga-Helia University of Applied Sciences.

Fieldler, L & GrobMab, T & Roth, M & Vetvik J, O. 2016. Why Customer Analytics Matter. https://www.mckinsey.com/. Accessed on 20 March 2021.

Find Cluster in Data. https://www.tableau.com/. Accessed on 4 April 2021.

Forbes Technology Council. 2020. 10 Data Analytics Myths That Can Hamper Your Business. Accessed on 30 March 2021.

Gavett, G. 2014. What You Need To Know About Segmentation.

Grissby, M 2016. Advanced Customer Analytics Targeting, Valuing, Segmenting and Loyalty Techniques. Marketing Science Series.

Haley, I. 1968. Benefit Segmentation: A Decision-Oriented Research Tool. Journal of Marketing.

Han, P. 2017. A Beginner's Guide to Cohort Analysis: the Most Actionable (and Underrated) Report on Google Analytics. https://medium.com/. Accessed on 5 May 2021.

Hannah, M. 2019. Big Data in Business Analytics: Talking About The Analaytics Process Model. https://blogs.sas.com/content/. Accessed on 30 Match 2021.

Hennig-Thurau, T & Hansen, U. 2000 . Relationship Marketing Gaining Competitive Advantage Through Customer Satisfaction and Customer Retention. Springer-Verlag Berlin Heidelberge.

Honebein, P. 1997. Strategies for Effective Customer Education.

https://www.forbes.com/. Accessed on 28 Match 2021.

Ladly, J. & Redman, T. 2020. Use Data to Accelerate Your Business Strategy. Harvard Business Review. 03 March 2020.

Lohrum, M. 2000. Relationship Marketing Success Through Investments in Products.

Lyons, B. 2015. Top 15 LOD Expressions. https://www.tableau.com/. Accessed on 08 April 2021.

Maheshwari, A. 2015. Data Analytics Made Accessible.

Maie, J & Saunders, J. 1990. The implementation process of segmentation in sales management. J Pers Sell Sales Manag.

Makhija, P. 2020. Cohort Analysis: Beginners Guide To Improving Retention. https://clevertap.com/blog/. Accessed on 8 April 2021.

Marr, B. 2020. The 9 Best Analytics Tools For Data Visualization Available Today. https://www.forbes.com/. Accessed on 28 Match 2021.

McDonald, S & Dunbar, I. 2012. Market Segmentation How To Do It and How To Profit From It. John Wily & Sons.

Meyer, A & Bliimelhuber, C. 2000). Relationship Marketing Success Through Investments in Services.

Normann, R. 1991. Service Management: Strategy and Leadership in Service Businesses.

Payne, A & Frow, P. 2000. Services Relationship Marketing: A Sector Case Study.

PwC. 2020. Analytics and AI Transformation. https://www.pwc.com/. Accessed on 25 Match 2021.

Reichheld, F & Kenny, D. 1990. The Hidden Advantages of Customer Retention. Journal of Retail Banking, p. 24.

Reichheld, R & Sasser, Jr. 1990. Zero Defections: Quality Comes to Services. Harvard Business Review.

Saunders, M., Lewis, P. & Thornhill, A. 2009. Research Methods for Business Students, 5th ed, pp. 124-155. London: the United Kingdom: Pearson.

Shackel, B. 1984. The Concept of Usability. Englewood Cliffs: Prentice Hall, p.45.

Singh, H. 2006. The Importance of Customer Satisfaction in Relation to Customer Loyalty and Retention. Asia Pacific University College of Technology & Innovation.

Singh, N. What is K-Means Clustering. https://ai.plainenglish.io/. Accessed on 5 May 2021.

Su. B. 2018. What is Cohort Analysis and How Should I Use it?. https://medium.com/. Accessed on 5 May 2021.

Zhang, S & Zhang, C & Yang, Q. 2003. Data Preparation For Data Mining. DOI: 10.1080/08839510390219264. Accessed on 5 April 2021.