International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 4, pp: 516-522 www.ijaem.net ISSN: 2395-5252

A Qualitative Approach to Customer Segmentation and Customer Churn Application

Drashti Shrimal, Dr. Harshali Patil
Student - M.E CMPN Thakur College of Engineering & Technology Mumbai, Maharashtra
HOD - CMPN. Thakur College of Engineering & Technology Mumbai, Maharashtra
Corresponding Author: Drashti Shrimal

Date of Submission: 01-08-2020 Date of Acceptance: 15-08-2020

ABSTRACT: With the development of internet, various companies are coming up to the win the market. In such scenarios, it is very necessary to keep your company uptight. To aid this motive, we need to differentiate users in a meaningful way, and serve them accordingly - Customer Segmentation. Customer Segmentation is the practice of dividing a consumer base into businesses of individuals that are comparable in specific approaches applicable to marketing, which include age, gender, hobbies and spending behavior. Along with differentiating company needs to predict how many users are about to leave the company and join the competitor, possibly even know the reason - Customer Churn.
In this paper we are combining models like Cohort Analysis, RFM and K-Means Pro clustering techniques we are trying to build a better solution for the organization using it and hence producing better segments of customers. This paper thus provides a complete solution to Customer segmentation.
Keywords—Customer Segmentation, K means, Cohort analysis,RFM Model,Google Cloud, K Means Pro, Churn Rate.

I. INTRODUCTION

Ever visited a farmer's market? Usually, you'll walk long aisles of stall when stall of various farmers commercialism a set of manufacture on to shoppers. It is a great way to support your native economy, whereas attempting a number of the freshest manufacture around. A farmer's market is comparable in idea to a different style of marketplace, except this one lives in Internet.

A. What is an Online Market place?

A Web marketplace, typically referred to as an electronic marketplace, is an e-comm website wherever third-party firms will sell their merchandise or services to shoppers. All the transactions are processed through the web site owner. As an example, once you visit Amazon's on-line market, you'll purchase books and different merchandise directly from Amazon, however you'll conjointly notice - and obtain - merchandise from sellers everywhere the globe. Another example is sharing economy platforms, like eBay and Etsy, wherever sellers "share" within the area to form profits, assemble all sorts of freelance sellers into a one stop search that's convenient for shoppers to not solely check costs for the most effective deals, however do therefore all underneath one electronic roof. Currently that we've checked out samples of on-line marketplaces, let's verify an example of one thing that's not a web marketplace. Now that we've got checked out samples of on-line marketplaces, let's cross-check an example of one thing that's not an internet marketplace. When you purchase groceries from web your native market for home delivery, you're not searching in an on-line marketplace - even supposing you're searching online.

B. What are Markets?

Markets are primarily a bunch of potential patrons with wants and needs and also the getting power to satisfy them. Producers advertise merchandise and services to shoppers in an exceedingly market so as to come up with demand. Also, the term "market" is closely related to monetary assets and securities costs (for example, the stock exchange or the bond market).

C. What is Market/Customer Segmentation?

Also called as Client segmentation is that the follow of dividing a client base into teams of people that are similar in specific ways that relevant to promoting, like age, gender, interests and disbursement habits.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 4, pp: 516-522 www.ijaem.net ISSN: 2395-5252

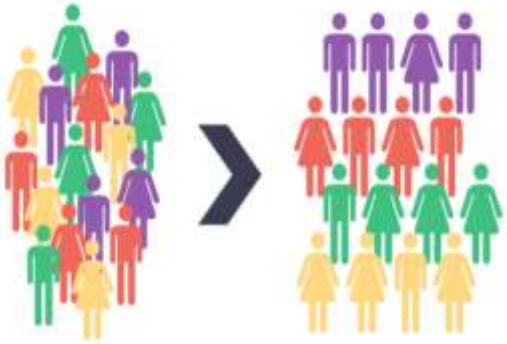


Figure 1: Customer Segmentation Logical Diagram

Figure 2: Types of Customer Segmentation

In business-to-business marketing, a company might segment customers according to a wide range of factors, as shown in the above diagram.

II. BACKGROUND STUDY - I

A. Need for Customer Segmentation:

i. Increase the focus of a firm

If you have better focus, obviously you will have better business. Many automobile firms have started concentrating on small car segments. This is their strategy of changing its focus for better returns. Thus, companies base their strategy completely on a new proof segment which increases its focus and gains. If a company knows what kind of customers its services, what are their characteristics, what they like the most, their spending patterns and so on; it can serve the customer much better. Basically, makes them more focused about their customers.

ii. Increase in competitiveness

If you are focusing on young people, your brand will be famous and equity with youngsters will be very high.. The brand loyalty will definitely increase. Thus market segmentation also increases competitiveness of a firm from a holistic view.

iii. Market expansion

Geographic segmentation is where expansion is any moment possible. If you have your market strategy on the basis of geography, then once you are serving a particular area, you can immediately expand to a nearby area. Similarly, if you are targeting customers based on their demography (Ex – Nike targets fitness enthusiasts) then you can expand in
similar products (Ex – Nike expanding with its fitness range of clothes and accessories).

iv. Customer retention

By performing segmentation, client retention is inspired through the life cycle of a client. The simplest example of this is often the auto and therefore the Airlines phase. You'll realize major example of client life cycle within the cordial reception segment whether or not they be hotels, airlines, or hospitals. In India, Titan is an example of merchandise that is planned through the life cycle of a client. From agency to serious music and therefore the high vary watches; Titan has them by worth phase likewise as life cycle phase. Thus, a watch is accessible for any client UN agency enters a Titan saleroom, no matter be his age.

v. Increase in brand loyalty

Customer loyalty and shared values; once you build a robust whole, you would like to convey these values to make an emotional reference to customers, whole loyalty usually lasts a time period and even transfers to future generations, increased quality and easy purchase. Higher sales volumes, premium valuation ability, and holding instead of seeking customers are the vital aspects of name loyalty. You'll be able to produce favorable whole attitudes by having a top-quality product that solves a drag a shopper is experiencing.

vi. Increase in Gains

Segmentation will increase fight, whole recall, whole equity, client retention, and communications. Therefore, if it's moving such a big amount of factors of your business, then positively it affects the profit of the firm. Does one ever see folks negotiating during a Greek deity, Gucci or BMW showrooms? You won't. One in every of the USP's of those wholes is their segmentation. They're really targeting segments that haven't any would like of talks or negotiation. Thus, their profit is high.

vii. Better Communication

One of the factors of selling combine that is totally obsessed with standard atmosphere is Promotions or communications. The communications

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 4, pp: 516-522 www.ijaem.net ISSN: 2395-5252

of an organization must be spot on for its TARGET market. Thus, if you would like a target market, you would like segmentation. Communication cannot be attainable while not knowing your target market.

Imagine that you just were ahead of a curtain. There was somebody on the opposite aspect. And you were asked to elucidate Politics to the person on the opposite aspect. You'd press on concerning ruling parties, states, countries and politicians. And once the curtain is taken aside, you discover that the person across the curtain may be a 5- year-old child.

Is there any use reprimand him concerning politics? This shows why communication wants segmentation. If you don't understand your market section, what's their sociology, what's their science, wherever they're from, what's their understanding level, and then however are you able to type a communication message?

III. BACKGROUND STUDY - II

A. What is Customer Churn/Loyalty?

Customer churn refers to once a client (player, subscriber, user, etc.) ceases his or her relationship with an organization. On-line businesses generally treat a client as churned once a selected quantity of your time has March on since the customer's last interaction with the positioning or service. The complete price of client churn includes each lost revenue and therefore the selling prices involved exchange those customers with new ones. Reducing client churn may be a key business goal of each on-line business.

B. What is Customer Retention?

Customer retention refers to the ability of a company or product to retain its customers over some specified period.

High customer retention means customers of the product or business tend to return to, continue to buy or in some other way not defect to another product or business, or to non-use entirely.

.

C. The Need of Predicting Customer Churn and Retention?

The ability to predict that a selected client is at a high risk of churning, whereas there's still time to try and do one thing concerning it, represents a large extra potential revenue supply for each on-line business. (In different words, feat that client might have really been a losing investment.)

What is more, it's perpetually tougher and big-ticket to accumulate a replacement client than it's to retain a current paying client. Also, what's the purpose of paying all of your resources, energy and time on the acquisition of latest customers whereas ignoring the prevailing ones?

IV. PROPOSED ARCHITECTURE

Figure 3: Architecture of Proposed solution

1. Google Cloud Platform:

Google Cloud Platform (GCP), offered by Google, is a suite of cloud computing services that runs on the same infrastructure that Google uses internally for its end-user products, such as Google Search, Gmail and YouTube. Alongside a set of management tools, it provides a series of modular cloud services including computing, data storage, data analytics and machine learning. Registration requires a credit card or bank account details.

Google Cloud Platform provides infrastructure as a service, platform as a service, and serverless computing environments.

2. Cloud Pub Sub:

It stands for Publisher Subscriber. It acts as buffer for the messages(json documents) created when the user interacts with website/app.

Google Cloud Pub/Sub is a fully-managed, scalable, global and secure messaging service that allows you to send and receive messages among applications and services. It is used to integrate decoupled systems and components hosted on Google Cloud Platform or elsewhere on the Internet.

3. Cloud Data Flow:

Google Cloud Dataflow is a cloud-based data processing service for both batch and real-time data streaming applications. It enables developers to set up processing pipelines for integrating, preparing and analyzing large data sets. Dataflow enables fast, simplified streaming data pipeline development with lower data latency.

4. Big Query:

Google Big Query is a cloud-based enterprise data warehouse that offers rapid SQL queries and interactive analysis of massive data sets. Big Query is a managed data warehouse, simply say it's a database. So your data will be stored in BigQuery, and you can access it by using SQL

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 4, pp: 516-522 www.ijaem.net ISSN: 2395-5252

queries. BigQuery manages the technical aspects of storing your structured data, including compression, encryption, replication, performance tuning, and scaling.

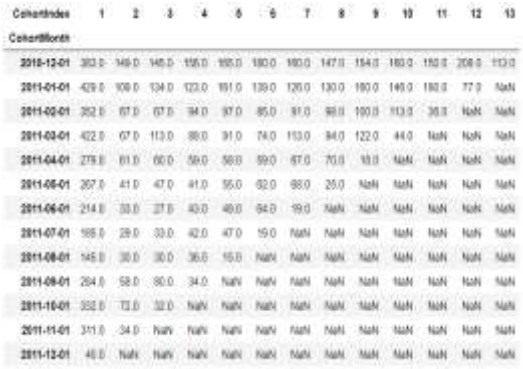
5. ML Models used:

5.1 Cohort Analysis-

Cohort analysis is a subset of behavioral analytics that takes the data from a given e-Commerce platform, web application, or online game and rather than looking at all users as one unit, it breaks them into related groups for analysis.

These related groups, or cohorts, usually share common characteristics or experiences within a defined time-span.

In our project we are using Cohort to check the retention rates of customers.



For Visualization of the above Customer Retention, Heat maps can be used which enables marketing analysts to view data in an informative way as follows:

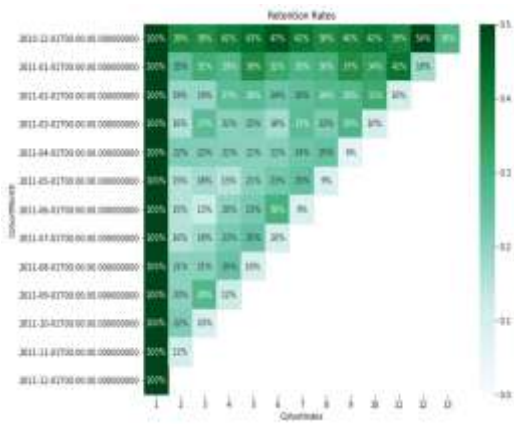


Figure 4: Heat Map for Cohort Analysis

5.2 RFM Model-

Recency, frequency, monetary value is a marketing analysis tool used to identify a company's or an organization's best customers by using certain measures.

The RFM model is based on three quantitative factors:

Recency: How recently a customer has made a purchase

Frequency: How often a customer makes a purchase

Monetary Value: How much money a customer spends on purchases

RFM analysis numerically ranks a customer in each of these three categories, generally on a scale of 1 to 5 (the higher the number, the better the result).

The "best" customer would receive a top score in every category.

CustomerID	Recency	Frequency	MonetaryValue	R	F	M
12747	3	25	948.70	4	4	4
12748	1	888	7046.16	4	4	4
12749	4	37	813.45	4	4	4
12820	4	17	268.02	4	3	3
12822	71	9	146.15	2	2	3

Figure 5: RFM example

CustomerID	Recency	Frequency	MonetaryValue	R	F	M	RFM_Score	RFM_Segment	General_Segment
12747	3	25	948.70	4	4	4	12.0	444	Gold
12748	1	888	7046.16	4	4	4	12.0	444	Gold
12749	4	37	813.45	4	4	4	12.0	444	Gold
12820	4	17	268.02	4	3	3	10.0	433	Gold
12822	71	9	146.15	2	2	3	7.0	223	Silver

Figure 6: RFM segment example

```
print("Best Customers: ",len(final_df[final_df['RFM_Score']==12]))
print('Loyal Customers: ',len(final_df[final_df['F']==4]))
print("Big Spenders: ",len(final_df[final_df['M']==4]))
print('Almost Lost: ', len(final_df[final_df['RFM_Segment']=='244']))
print('Lost Customers: ',len(final_df[final_df['RFM_Segment']=='144']))
print('Lost Cheap Customers: ',len(final_df[final_df['RFM_Segment']=='111']))
```

```
Best Customers:  28131
Loyal Customers:  47340
Big Spenders:  42925
Almost Lost:  2866
Lost Customers:  514
Lost Cheap Customers:  710
```

5.3 Proposed Algorithm: K-Means Pro

Let us first see the K-Means algorithm and its drawback-

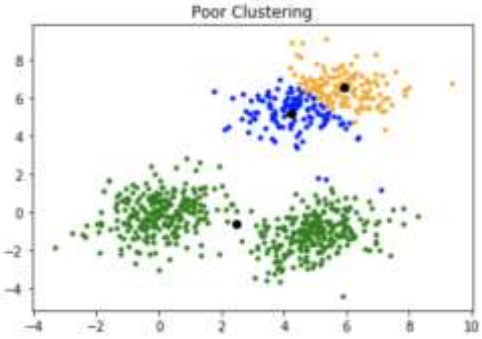
K-means is one of the most straightforward algorithms which is used to solve unsupervised clustering problems.

In these clustering problems we will use a DataSet of instances and the DataSet is defined with the help of some attributes. Each instance in the DataSet has some relevant values corresponding to those attributes. Goal is to categorize those instances into different clusters with the help of k-mean algorithm.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 4, pp: 516-522 www.ijaem.net ISSN: 2395-5252

Algorithm:
1.The first step involves the random initialization of k data points which are called means.
2.In this step we cluster each data point to its nearest mean and after that we update the mean of the current clusters. mean: is the average of a group of values.
3.This cycle continues for a given number of repetitions and after that we have our final clusters.

Drawback: Poor Clustering



The main drawback of k-means algorithm is that it is very much dependent on the initialization of the centroids or the mean points.

In this way, if a Centroid is introduced to be a "far away" point, it may very well wind up without any data point related with it and simultaneously more than one cluster may wind up connected with a solo Centroid. Likewise, more than one centroids may be introduced into a similar group bringing about poor clustering.

K-MEANS PRO-
The proposed algorithm can be used to overcome the drawback posed by the k-means algorithm. Lets name it 'K-Means Pro'

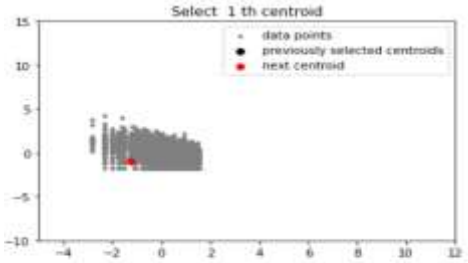
This algorithm guarantees a more intelligent introduction of the centroids and improves the nature of the clustering. Leaving the initialization of the mean points the K-means Pro algorithm is more or less the same as the conventional k-means algorithm.

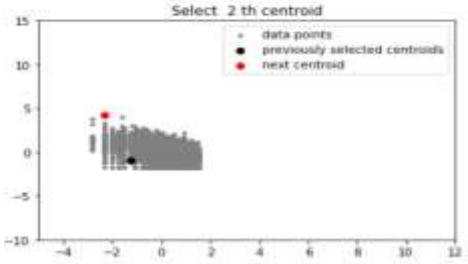
Algorithm:
1. In the starting we have to select a random first Centroid point from the given DataSet.
2. Now for every instance say 'i' in the DataSet calculate the distance say 'x' from 'i' to the closest, previously chosen Centroid.
3. Select the following Centroid from the DataSet with the end goal that the likelihood of picking a point as Centroid is corresponding to the distance from the closest, recently picked Centroid.
4. Last 2 steps are repeated until you get k mean points.

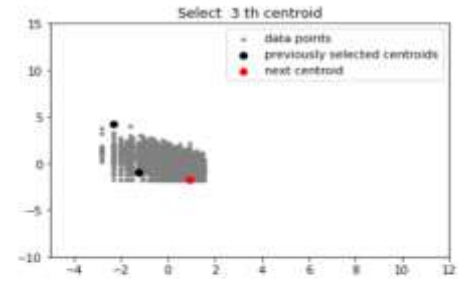
1. Data Studio:
Visualization of the segmented data is done using Data studio.

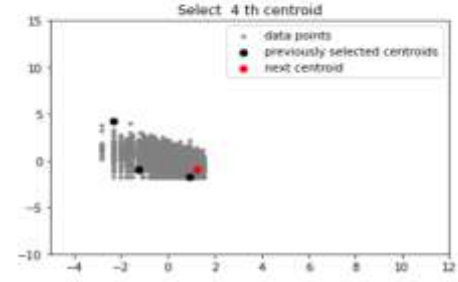
V. RESULTS

Following are outputs of K-Means Pro with k = 4:









International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 4, pp: 516-522 www.ijaem.net ISSN: 2395-5252

The comparisons of K-Means and K-Means Pro are given in the table below.

With synthetic examples K-Means does not perform well, because the random seeding will inevitably merge clusters together, and the algorithm will never be able to split them apart.

The careful seeding method of K-Means Pro avoids this problem altogether, and it almost always attains the optimal results on the synthetic datasets.

Following shows the results of a cloud data set from University of California - Irvene with 1024 data points and 10 attributes.

k	Average φ		Minimum φ	
	K-Means	K-Means Pro	K-Means	K-Means Pro
10	7533.5	6151.2	6139.45	5631.99
25	3626.1	2064.9	2568.2	1988.76
50	2004.2	1133.7	1344	1088

Table 1: Comparative Parameters of Result

VI. CONCLUSION

Our aim is to group the customers in such a way that it proves beneficial to customers and the organization by gaining maximum profits. Thus by combining models like Cohort Analysis, RFM and K-Means Pro clustering techniques we are trying to build a better solution for the organization using it and hence producing better segments of customers. The Cohort analysis will give us the Retention rates and Lost Customer segment gives us the Churned users. K - Means Pro terminates almost twice as fast while achieving potential function values about 20% better.

This application is a complete approach to Customer Segmentation starting from data collection to filtering and profiling data, to storing data in Big Query then using the ML Models to segment customers and finally visualizing data on Data Studio.

FUTURE SCOPE

In future Recommendation systems can be collaborated along with this application thus to produce a personalized effect to each and every customer. It will be extremely beneficial to the company for gaining max profits. For example: 35% of Amazon's profits is redeemed from Recommendation systems. On each customer segment recommendation systems can be applied which would thus produce products accordingly to specific users as well.

Based on the segment, loyal customers could be awarded with a gift and so on. Best customers could be given suitable discounts. Cheap customers could be given half rates so that they could become happy and come back to the shopping portal.

REFERENCES

[1]. SardjoeniMoedjiono, YosianusRobertusIsak, Aries Kusdaryono, "Customer Loyalty Prediction in Multimedia Service Provider Company With K-Means Segmentation and C4.5 Algorithm" Journal of marketing IEEE, Vol. 21, No. 1, July 2018.

[2]. A.S. M. Shahadat Hossain, "Customer Segmentation using Centroid Based and Density Based Clustering Algorithms,"2017 3rd International Conference on Electrical Information and Communication Technology (EICT), 7-9 December 2017, Khulna, Bangladesh.

[3]. Ms. E. Suganya, Dr. S. Vijayarani, "Customer Loyalty Prediction in Multimedia Service Provider Company With K-Means Segmentation and C4.5 Algorithm", IEEE, July 2017.

[4]. Dr. (Mrs) AnanthiSheshasaayee, L. Logeshwari," An Efficiency Analysis on the TPA Clustering Methods for Intelligent Customer Segmentation",ICIMIA 2017.

[5]. Data Mining", IEEE, 2016. [5]Surbhi Sharma,Dr. Arvind K Sharma, Dinesh Soni,", Enhancing DBSCAN Algorithm for Data Mining", ICECDS, 2017.

[6]. R. Revathy, R. Lawrance, "Comparative Analysis of C4.5 and C5.0 Algorithms on Crop Pest Data", IJIRCCE, 2017

[7]. Cheng-Fa Tsai, Yao Chiang,"Enhancement of Data Clustering using TSS DBSCAN approach for

[8]. Xixi He, Chen Li, "The Research and Application of Customer Segmentation on E-commerce Websites", IEEE, 2016.

[9]. Amandeep Kaur Mann, Navneet Kaur, "Grid Density Based Clustering Algorithm", IJARCET, 2013.

[10]. GuillemLefait, TaharKechadi, "Customer Segmentation Architecture Based on Clustering Techniques", IEEE, 2010.

[11]. HUANG Lan, ZHOU Chun-guang, ZHOU Yu-qin, WANG Zhe, "Research on Data Mining Algorithms for Automotive Customers' Behavior Prediction Problem", IEEE, 2008.

[12]. Andrew Kusiak, "Data Mining: Bagging and Boosting", Intelligent Systems Laboratory.
[13]. "Customer Segmentation." https:// pushalert.co/blog/audience-segmentation-ecommerce-conversions, 2018. https://bigdata-madesimple.com/top-5-best-data-visualization-techniques-for- 2018/, 2018.
[14]. "Types of Market Segmentation", https:// www.dreamstime.com/illustration/segmentation, 2017.
[15]. Andrew Kusiak, "Data Mining: Bagging and Boosting", Intelligent Systems Laboratory the University of Iowa Iowa City, 2016.
[16]. "Top 5 best data visualization techniques for 2018"

International Journal of Advances in Engineering and Management
ISSN: 2395-5252

IJAEM

Volume: 02 Issue: 01 DOI: 10.35629/5252

www.ijaem.net
Email id: ijaem.paper@gmail.com