Mini Project Report on

MALL CUSTOMER SEGMENTATION

by

AVADHUT MANE(18CE8021)

AYUSH MISHRA(18CE8010)

ANKIT PAL(18CE7040)

Under the guidance of

MR. KAMLESH NENWANI

Department of Computer Engineering

Ramrao Adik Institute of Technology

Dr. D. Y. Patil Vidyanagar, Nerul, Navi Mumbai

University of Mumbai

May 2021

Ramrao Adik Institute of Technology

Dr. D. Y. Patil Vidyanagar, Nerul, Navi Mumbai

CERTIFICATE

This is to certify that Mini Project report entitled

Title of the Mini Project

By

AVADHUT MANE(18CE8021)

AYUSH MISHRA(18CE8010)

ANKIT PAL(18CE7040)

is successfully completed for Third Year Computer Engineering as prescribed by University of Mumbai.

Supervisor	Project Co-ordinator
(Mr. Kamlesh Nenwani)	**(Dr. Dakhane Dhananjay M.)**
Head of Department	Principal
(Dr. Leena Ragha)	**(Dr. Mukesh D. Patil)**

Mini Project Report Approval

This is to certify that the Mini Project entitled ***" MALL CUSTOMER SEGMENTATION"*** is a bonafide work done by ***AVADHUT MANE(18CE8021)***, ***ANKIT PAL(18CE7040)***, and ***AYUSH MISHRA(18CE8010)*** under the supervision of ***MR. KAMLESH NENWANI***. This Mini Project has been approved for Third Year Computer Engineering.

Internal Examiner :

1. ……

2. ……

External Examiners :

1. ……

2. ……

Date : …/…/……

Place : ……

DECLARATION

I declare that this written submission represents my ideas and does not invovle plagiarism. I have adequately cited and referenced the original sources wherever others' ideas or words have been included. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action against me by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: _____

AVADHUT MANE(18CE8021)

ANKIT PAL(18CE7040)

AYUSH MISHRA(18CE8010)

Abstract

The zeitgeist of the modern era is innovation, with everyone competing to be better than the others. Today's business is based on the ability of such innovation to enthral customers with products, but with such a diverse range of products, customers are unsure what to buy and what not to buy, and companies are unsure which segment of customers to target to sell their products. This is where machine learning comes in; various algorithms are used to uncover hidden patterns in data in order to make better decisions in the future. By using segmentation, the ambiguous concept of which segment to target becomes clear. The process of categorising customers who have similar behaviours into groups. This process is basically dividing the same segment into different other segments .

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Overview

Now by looking at the competitiveness and ever-growing needs of millennials these data mining has become has become an important part for solving this problem as it stores the pattern of behavior in which the customers shops for their requirements in the respective databases. The customer segmentation that we are seeing is one the applications of this data mining that helps the company to segment and cluster out the customers with similar kind of requirements and habits. What this does is that it makes the life of business industry a little easy to target their required audience which is present in huge numbers.

Objectives

The customer segmentation we do is trying to make and direct or say indirect impact on the already existing market strategies as this makes way for the newpaths that must be discovered like for which segmented clusters which products will be good , also they can customize their Marketing plan accordingly for these segmented clusters. It is basically an unsupervised learning that finds the clusters over unlabelled dataset.
There are number of clustering algorithms but in our project we havebasically used the k-means algorithm.

Motivation

we know that due to current increasing demands in the business industry day by day, we can say that businesses that exist from past are forcefully made to adapt to the new trending strategies going on in the current market by lookingat the competitiveness all these companies are facing. It is like do or die situation for all the firms inorder to survive these changes. By looking at the increasing population which results in increasing the base of customers for thecompanies it has become more and more challenging to fulfill the needs of each and everyone .

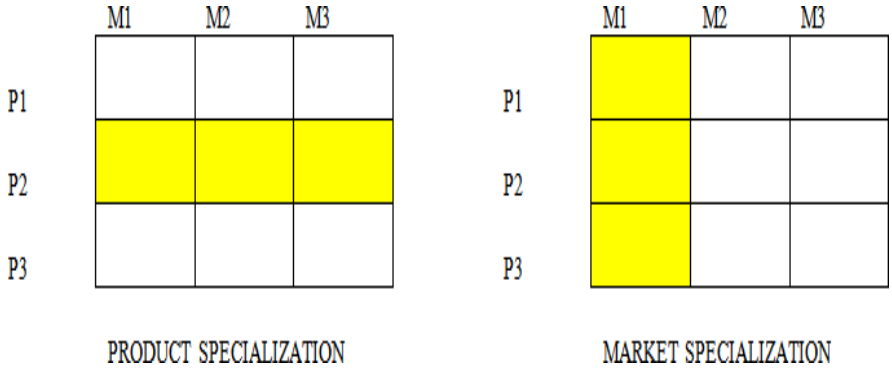
Organization of report

The introduction and the need of the project is discussed in Chapter 1. The survey of the existing system, and the research work carried out have been laid out in Chapter 2. The proposed model of the project, its design and implementation have been discussed in Chapter 3, which also encompasses a brief overview of our system. The results section in Chapter 4 contains screenshots and explains the functioning of the working computer model. Chapter 5 contains conclusion of the project and also, some scope of future works in the same space.

Chapter 2
Literature Survey

Existing Systems

In a traditional marketing strategy, a marketer creates a product and then seeks out customers who can purchase it. In the advanced stage of marketing, the marketer determines the demand for profitable products, after which he seeks out markets and a group of customers who intend to buy it and provide feedback to the marketers for improvisation of their needs, as well as willingness to pay a premium for meeting the extra needs. The marketers go through the process of identifying potential customers, profiling them, targeting them, and finally positioning them in their minds. Segmentation refers to how marketers classify or group heterogeneous markets into homogeneous markets, where a group of customers share a common need or demand pattern. Targeting explains how to target these markets, while segmentation explains who to target. The term "positioning" refers to a marketer's position.efforts in positioning the brand or offering while taking into account the customer’s thoughts.



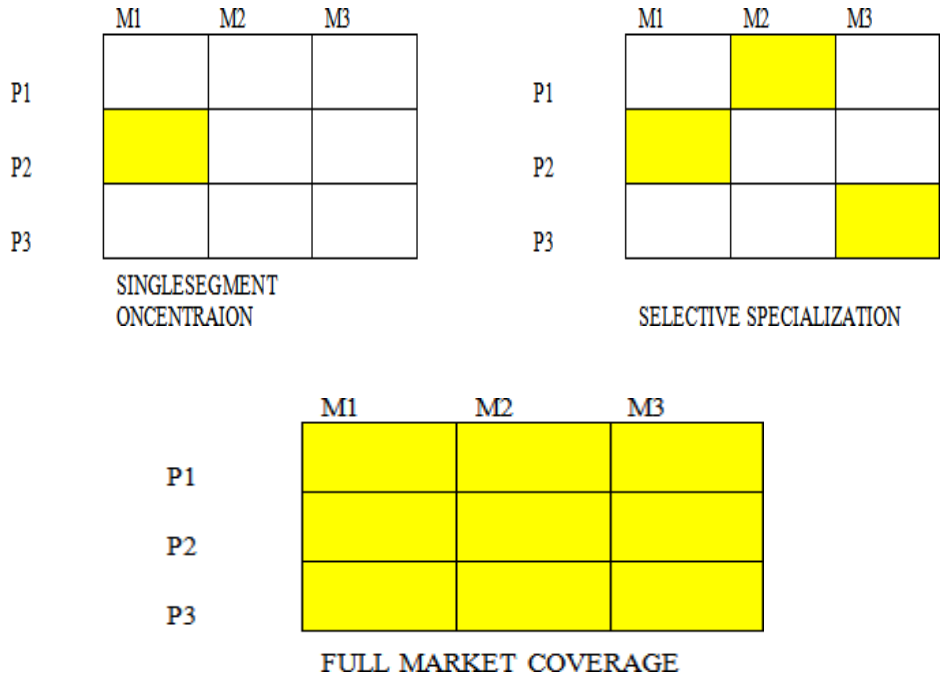


Figure 1: Positioning Substrates for Printed Paper Cups

Customers who prefer disposable cups for hot and cold beverages are represented by the red circle, while customers who prefer food-grade substrates are represented by the blue circle, and customers who value aesthetics prefer to have the brand logo printed on the cups are represented by the pink circle. The intersection of the three circles, where the customers have all of the desired needs for using the substrates being marketed, is targeted and positioned by a marketer who wants to sell paper cups with food-grade and high-quality printable substrates.

Limitaion of Existing System

Many businesses have failed to realise the full potential of their supply chains because they have not developed the performance measures and metrics required to completely integrate their supply chains for maximum effectiveness and efficiency. Gunasekaran et al. (2004) established a framework to help people comprehend the value of SCM performance metrics and measurement. (seven) Maltz et al. (2003) created the dynamic multi-dimensional performance (DMP) framework for evaluating performance .

Market segmentation and product differentiation are two independent approaches to product policy (Smith 1956). Product differentiation refers to product decisions made simply in relation to a firm's competitors, whereas market segmentation refers to product decisions made after analysing and describing the diversity of demands in a market place. This entails looking for groups of consumers in the general population, and occasionally even customers, who have similar within-group and across-group responses (see Frank, Massy, and Wind, 1972, Wedel and Kamakura, 2000).

For more than thirty years, comprehensive models of consumer behaviour have appeared in the marketing literature, describing a complicated, multi-event behavioural process (Nicosia 1966). Kamakura and Russell (1989) published a paper on statistical strategies for market segmentation that employ sales scanner data to estimate brand choice and attribute relevance.Segmenting" customers based on their buying habits or responses to marketing mix elements may appear out of touch with today's data mining capabilities.

Understanding where and how prospects shop is one facet of the task of market definition, but not market segmentation. Peacock (1998) identifies several potential uses that data mining has in the area of marketing including customer acquisition, customer retention and customer abandonment and market-basket analysis.

THE STUDY'S OBJECTIVES

The following are the study's goals:
To determine the value of Market Segmentation/Customer Segmentation, as well as its relevance and relationship to other marketing components.

Hypothesis

H(0): Segmentation has no bearing on other aspects of marketing; it has no influence on them.

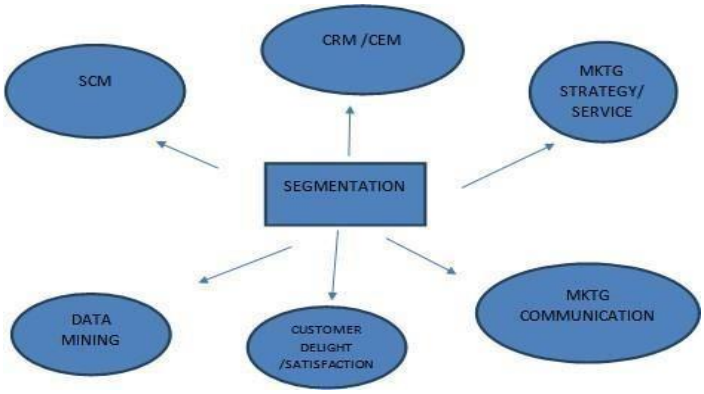
H(a): Segmentation is more relevant to other aspects of marketing and has a substantial impact on them.

Research Methodology

We conducted a thorough review of the literature on segmentation as well as the various aspects of marketing that are relevant to segmentation. We even gathered primary data to double-check the relevance of some of the components, such as customer satisfaction levels, and used statistical tools like Mintab to test certain hypotheses, but we couldn't share them due to confidentiality concerns and the need to limit the scope of the study to the stated goals.

Gathering of data

The Original Source With the help of structured questionnaires, data was collected directly from respondents given. Secondary data was gathered from the internet and other sources such as magazines and journals.



HYPOTHESIS EVALUATION

We evaluated the null hypothesis that segmentation has no relevance to other marketing components based on exploratory research on segmentation and its association and relevance to various other marketing components.
components of marketing and has no bearing on them, as well as the findings discussed in the following section of the study conclusions.

RESEARCH CONCLUSIONS AND FINDINGS:

According to our preliminary findings, the null hypothesis, which states that segmentation has no bearing on marketing components and has no impact on them, may be rejected. As a result, we can conclude that segmentation has a significant impact on various other components.

(As previously stated, statistical tools also supported our argument that customer satisfaction levels increased when the organisation followed the right segmentation and developed strategies for servicingthem specific to the segments - the results are not included in this paper but may be included in future research articles.)

Chapter 3

3. PROPOSED SYSTEM

PROBLEM STATEMENT

You own a mall and have some basic information on your customers on membership cards, such as Customer ID, age, gender, annual income, and so on. You want to understand your clients, such as who your targetclients are, so that you can inform your marketing staff and design yourapproach accordingly.

3.1 PROPOSED WORK:

The aim of our system is to generate a k number of clusters for a given dataset . We are going to implement this by K-means clustering . Datasetwill be uploaded and the we perform clustering using machine learning .Our further step will be to create a user interface for our project so the user can interact directly through the webpage without any complication.Customer segmentation is linked to a specific business goal.

The initial stage in segmentation is to decide on a business goal. The next stage is to gather the required information, such as data on demographics, transactions, and promotions then deciding on a consumer segmentation approach as well as measurement of standards.The last stage is to investigate the data by evaluating statistics and looking for correlations between variables.

The findings of the investigation can be used to determine how similar customers are measuring two points in a triangle with Euclidean distance
.The point is consumer data in a multidimensional space.

By computingthe ratio of the clusters, the cluster is certified.Variant that occurs between clusters. Finally we get our required number of clusters.
We create a webpage to display the graphs , tables to create a userfriendly environment.

The findings of the investigation can be used to determine how similar customers are measuring two points in a triangle with Euclidean distance
.The point is consumer data in a multidimensional space.By computingthe ratio of the clusters, the cluster is certified.Variant that occurs between clusters. Finally we get our required number of clusters.
We create a webpage to display the graphs , tables to create a userfriendly environment.

3.2 Proposed Methodology/Techniques

We are using k-means clustering for our project .
The K-Means clustering technique aims to divide a ‘n' number of observations into a specified number of ‘k' clusters (produces sphere- like clusters). The K-Means algorithm is an unsupervised learning technique and one of the most basic clustering algorithms. Without anycluster-internal structure, the K-Means separates the data into non- overlapping sections.

The numbers inside a cluster are quite similar to one another, whereasthe values between clusters vary dramatically. With medium and largedata sets, K-Means clustering performs extremely effectively.

Despite its simplicity, K-Means is nevertheless effective in data science for clustering instances.

The following steps have been Involved:

3.2.1 Obtaining the appropriate packages and importing them

3.2.2 Customer data is imported into the Python environment.

3.2.3 Analyze the data and come up with some relevant information

3.2.4 Adapting the data to our requirements

3.2.5 The K-Means algorithm was used to create the model.

3.2.6 Analyzing and visualising the K-Means model that has beendeveloped.

3.2.7 Now we implement our model on a webpage for bettervisualization and a user friendly interface.

3.2.8 We use HTML ,CSS, Java script , Django and Ajax for ourwebpage.

3.2.9 Finally our webpage is ready .

3.3 Design of the System :

Our project is a Webpage for implementation of customer segmentationusing k-means clustering.
Our user interface has different buttons like choose file, upload, generate , info ,csv and save as pdf.
The user interface asks the user to enter a csv file which has to be clustered. After we choose the file from our system we then press the upload button to upload the file. Now we can generate our cluster usingthe generate button. It runs our machine learning code and then we can see the graphs of the clusters formed.
The info button is used to visualize the tables that are formed.
The csv button is used to view the csv while which has been uploaded.The save as pdf button is used to save or print the present webpage.

Hardware/Software Requirement

3.3.1 python
The following packages will be required :

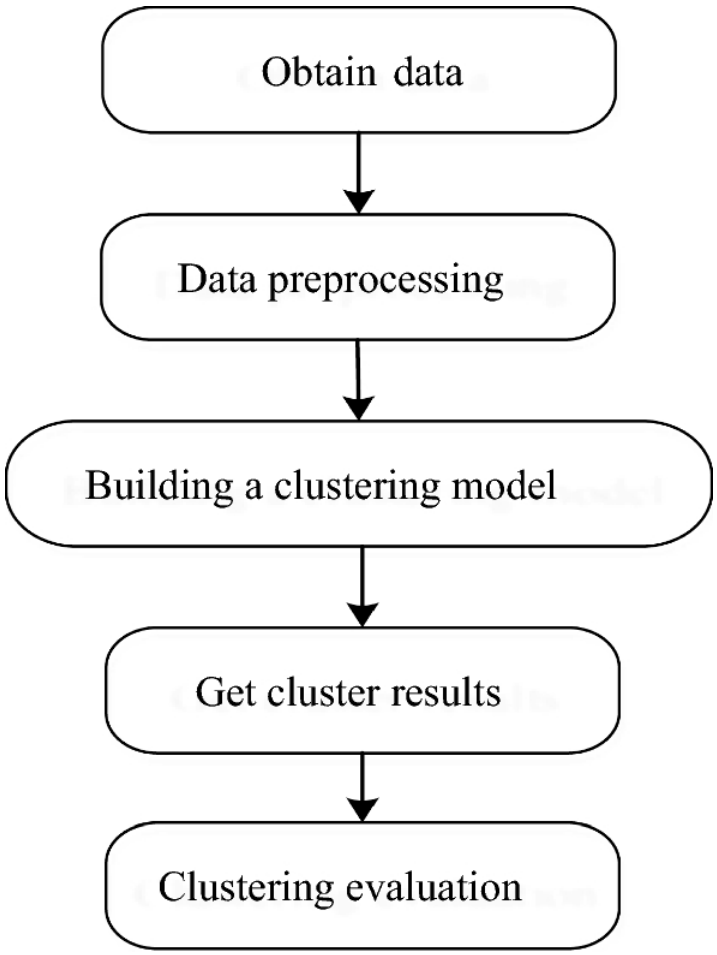
3.3.1.1 scikit-learn

3.3.1.2 seaborn

3.3.1.3 numpy

3.3.1.4 pandas

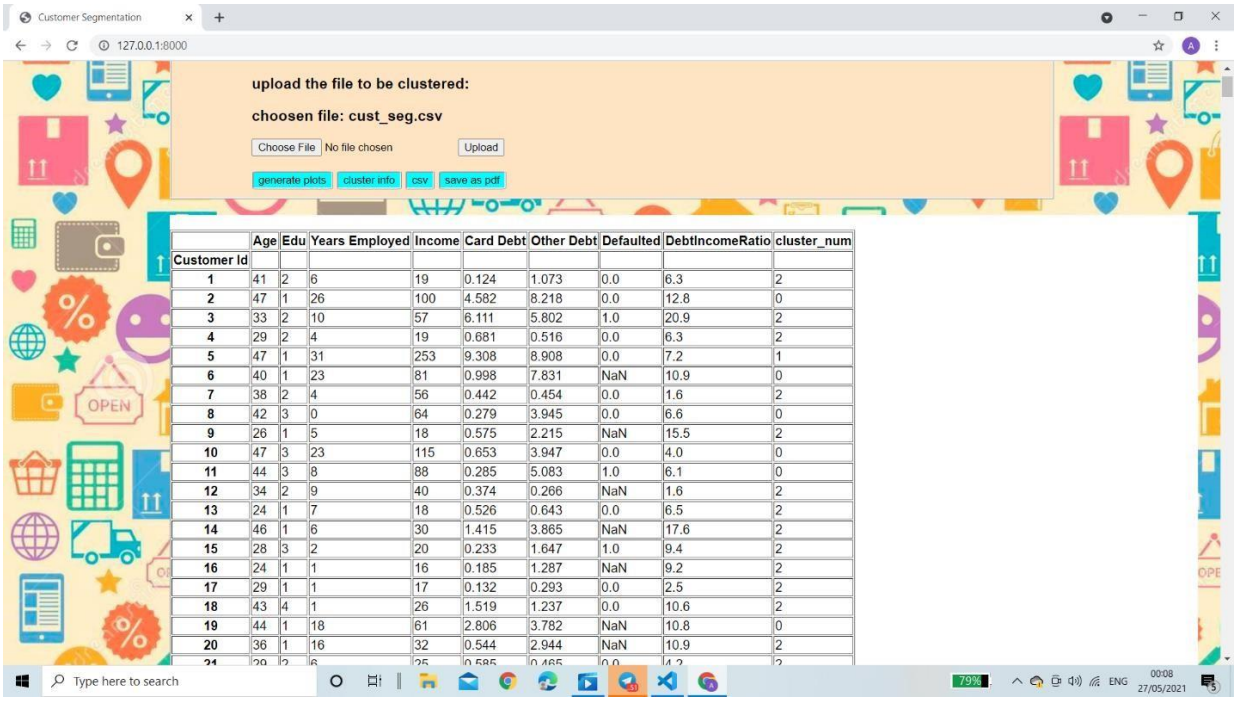
3.3.1.5 matplotlib


Obtain data
Data preprocessing
Building a clustering model
Get cluster results
Clustering evaluation

3.4 Implementation Details:

Every task must begin with the installation of the necessary packages in the appropriate environment (python in our case). Pandas is used to work with data, NumPy is used to work with arrays, matplotlib and seaborn are used for visualisation, mplot3d is used for three-dimensional visualisation, and scikit-learn is used to develop the K-Means model.

DATASET:

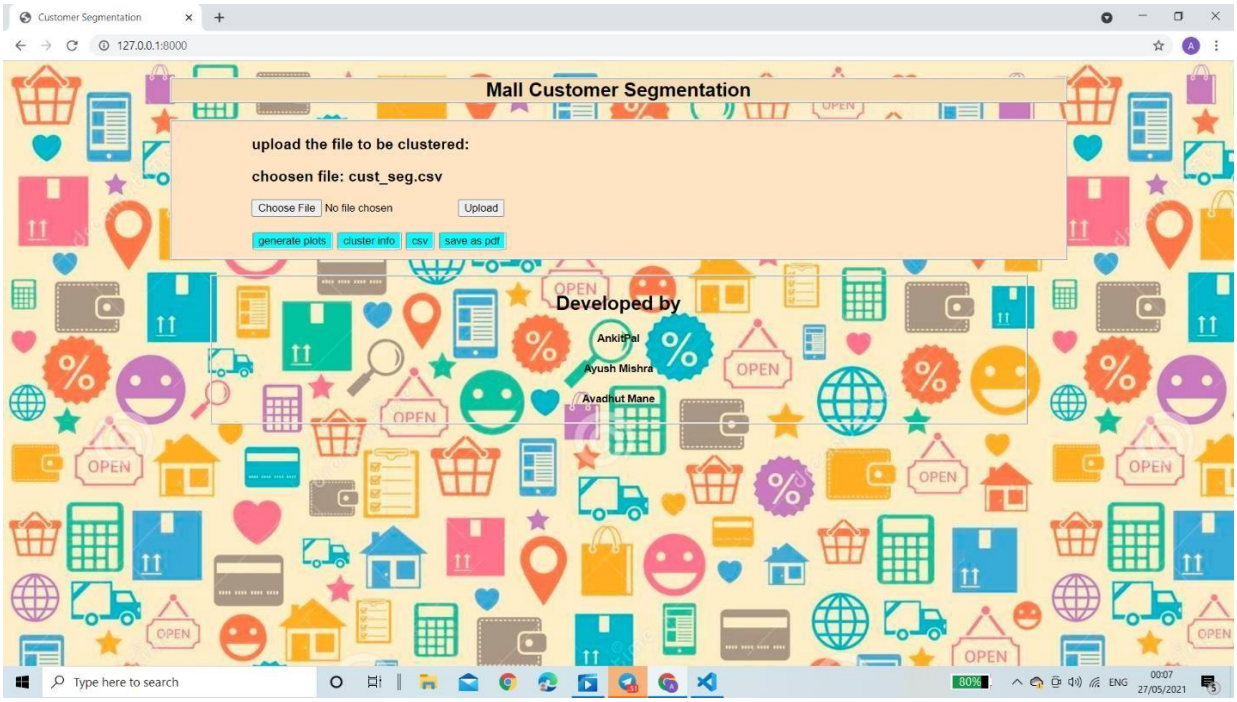


Customer Id	Age	Edu	Years Employed	Income	Card Debt	Other Debt	Defaulted	DebtIncomeRatio	cluster_num
1	41	2	6	19	0.124	1.073	0.0	6.3	2
2	47	1	26	100	4.582	8.218	0.0	12.8	0
3	33	2	10	57	6.111	5.802	1.0	20.9	2
4	29	2	4	19	0.681	0.516	0.0	6.3	2
5	47	1	31	253	9.308	8.908	0.0	7.2	1
6	40	1	23	81	0.998	7.831	NaN	10.9	0
7	38	2	4	56	0.442	0.454	0.0	1.6	2
8	42	3	0	64	0.279	3.945	0.0	6.6	0
9	26	1	5	18	0.575	2.215	NaN	15.5	2
10	47	3	23	115	0.653	3.947	0.0	4.0	0
11	44	3	8	88	0.285	5.083	1.0	6.1	0
12	34	2	9	40	0.374	0.266	NaN	1.6	2
13	24	1	7	18	0.526	0.643	0.0	6.5	2
14	46	1	6	30	1.415	3.865	NaN	17.6	2
15	28	3	2	20	0.233	1.647	1.0	9.4	2
16	24	1	1	16	0.185	1.287	NaN	9.2	2
17	29	1	1	17	0.132	0.293	0.0	2.5	2
18	43	4	1	26	1.519	1.237	0.0	10.6	2
19	44	1	18	61	2.806	3.782	NaN	10.8	0
20	36	1	16	32	0.544	2.944	NaN	10.9	2

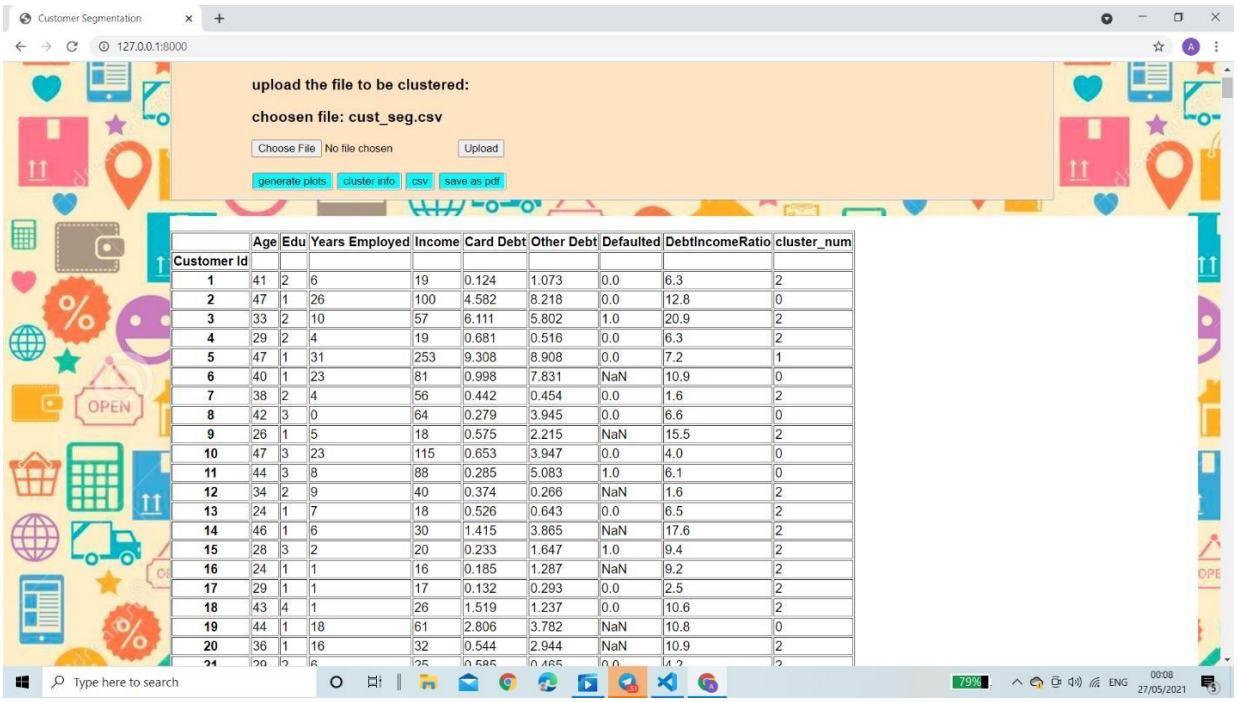
Chapter 4

Results and Discussion

Our system is webpage implemented in Django which segments the given dataset into k number of clusters .



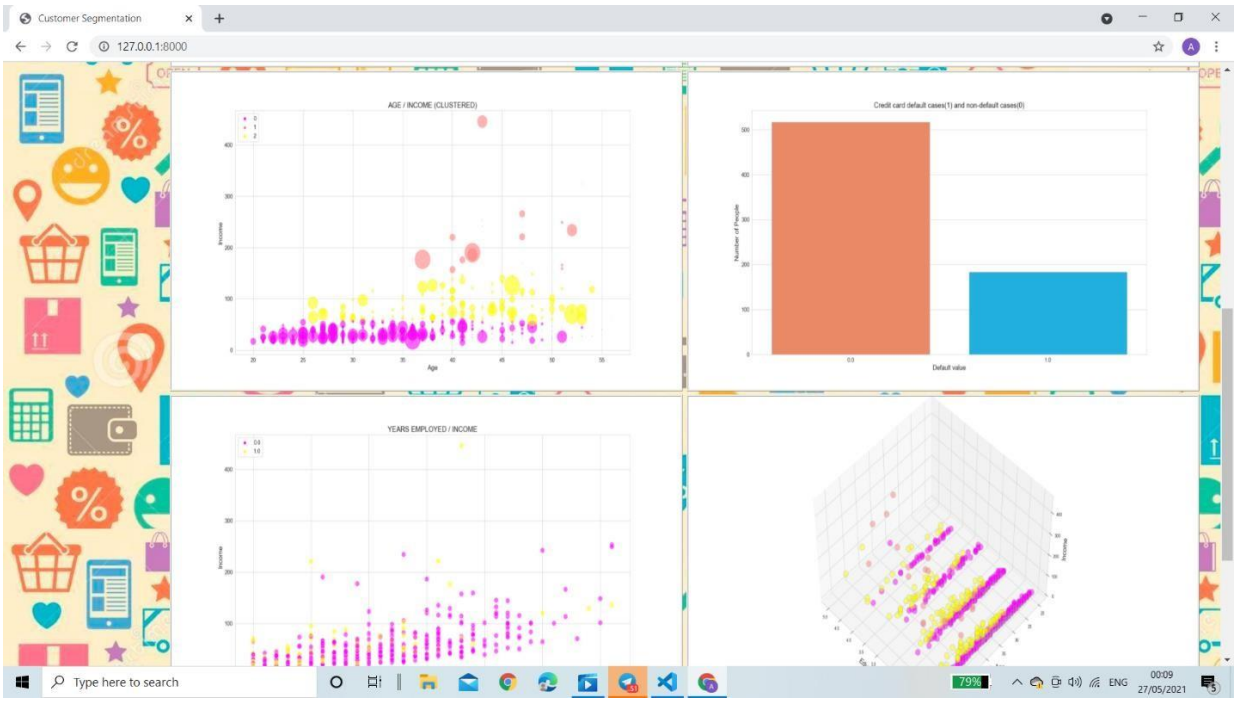
This is the frontend of our webpage. We can see the buttons like,upload,generate,csv etc.



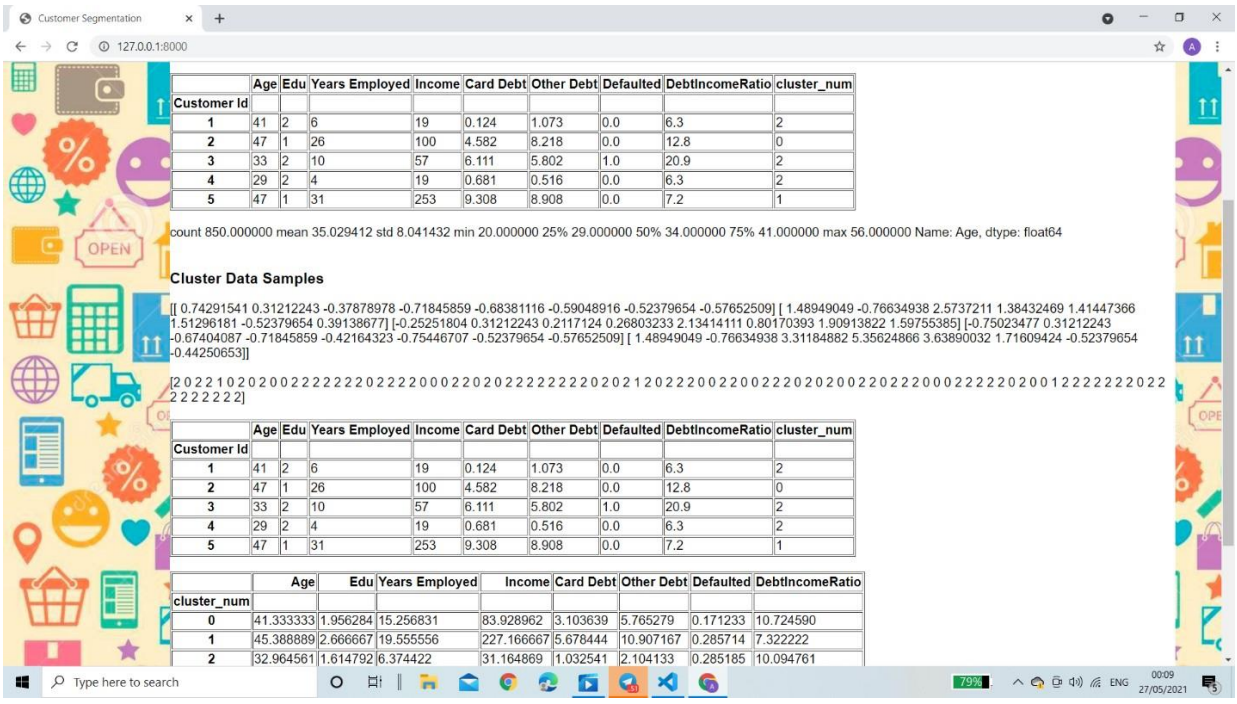
Customer Id	Age	Edu	Years Employed	Income	Card Debt	Other Debt	Defaulted	DebtIncomeRatio	cluster_num
1	41	2	6	19	0.124	1.073	0.0	6.3	2
2	47	1	26	100	4.582	8.218	0.0	12.8	0
3	33	2	10	57	6.111	5.802	1.0	20.9	2
4	29	2	4	19	0.681	0.516	0.0	6.3	2
5	47	1	31	253	9.308	8.908	0.0	7.2	1
6	40	1	23	81	0.998	7.831	NaN	10.9	0
7	38	2	4	56	0.442	0.454	0.0	1.6	2
8	42	3	0	64	0.279	3.945	0.0	6.6	0
9	26	1	5	18	0.575	2.215	NaN	15.5	2
10	47	3	23	115	0.653	3.947	0.0	4.0	0
11	44	3	8	88	0.285	5.083	1.0	6.1	0
12	34	2	9	40	0.374	0.266	NaN	1.6	2
13	24	1	7	18	0.526	0.643	0.0	6.5	2
14	46	1	6	30	1.415	3.865	NaN	17.6	2
15	28	3	2	20	0.233	1.647	1.0	9.4	2
16	24	1	1	16	0.185	1.287	NaN	9.2	2
17	29	1	1	17	0.132	0.293	0.0	2.5	2
18	43	4	1	26	1.519	1.237	0.0	10.6	2
19	44	1	18	61	2.806	3.782	NaN	10.8	0
20	36	1	16	32	0.544	2.944	NaN	10.9	2

First we upload the dataset .

On clicking the generate button the clusters are generated. Also we can visualize the graphs between different attributes like age vs income .



Further we get a 3D plot of our data and we can see that our clusters have been formed.

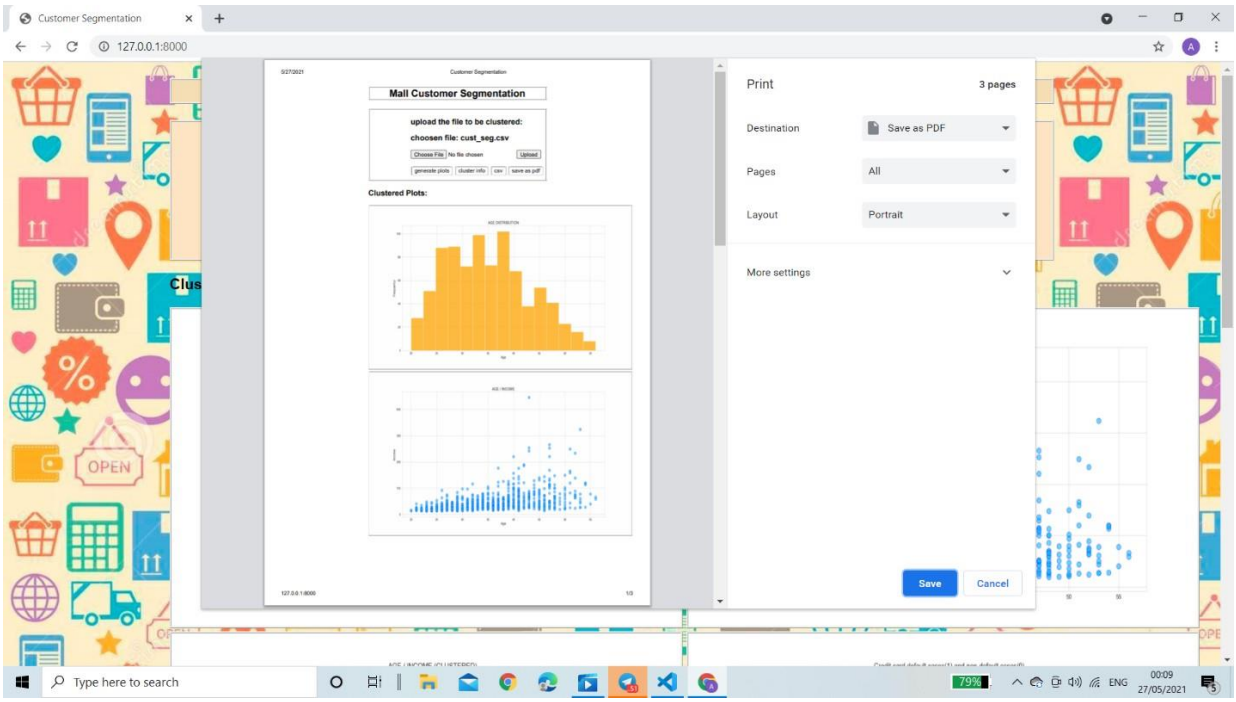


Customer Id	Age	Edu	Years Employed	Income	Card Debt	Other Debt	Defaulted	DebtIncomeRatio	cluster_num
1	41	2	6	19	0.124	1.073	0.0	6.3	2
2	47	1	26	100	4.582	8.218	0.0	12.8	0
3	33	2	10	57	6.111	5.802	1.0	20.9	2
4	29	2	4	19	0.681	0.516	0.0	6.3	2
5	47	1	31	253	9.308	8.908	0.0	7.2	1

Customer Id	Age	Edu	Years Employed	Income	Card Debt	Other Debt	Defaulted	DebtIncomeRatio	cluster_num
1	41	2	6	19	0.124	1.073	0.0	6.3	2
2	47	1	26	100	4.582	8.218	0.0	12.8	0
3	33	2	10	57	6.111	5.802	1.0	20.9	2
4	29	2	4	19	0.681	0.516	0.0	6.3	2
5	47	1	31	253	9.308	8.908	0.0	7.2	1

cluster_num	Age	Edu	Years Employed	Income	Card Debt	Other Debt	Defaulted	DebtIncomeRatio
0	41.333333	1.956284	15.256831	83.928962	3.103639	5.765279	0.171233	10.724590
1	45.388889	2.666667	19.555556	227.166667	5.678444	10.907167	0.285714	7.322222
2	32.964561	1.614792	6.374422	31.164869	1.032541	2.104133	0.285185	10.094761

On clicking the info button the tables are generated.



By clicking the save but we can download the given page or print it.

Our K-Means model divided the clients into three mutually exclusive groups, which we called clusters. The demographics of the clients in each cluster are comparable. We can now develop a profile for each group based on the cluster's shared traits. For instance, the three clusters could be:
Affluent, educated, and elderly
Middle-Aged and Middle-Income People
Low & Young

Chapter 5

Conclusion and Future Works

K means clustering is one of the most popular clustering methods, andit's frequently the first thing people do when they're working on a clustering problem to get a sense of the dataset's structure. K means
is used to divide data points into discrete, non-overlapping groupings**.**

Future scope will be to generalize the data so that it can be used on any dataset so that other businesses can implement our model too which willincrease the reach of our project.

We can further embark strategies on how to market the clustered segments. Segmentation of customers is one of the most demanded methods for companies to target advertisements . We can also design strategies on how to handle these clustered segments like what advertisements to show and which cluster is interested in what kind ofprojects.

Chapter 6

Bibliography

1. https://www.sciencedirect.com/science/article/pii/S0965856407000729
2. https://ijcrt.org/papers/IJCRT_196650.pdf
3. https://www.analyticsvidhya.com/blog/2021/01/profiling-market-segments-using-k-means-clustering/
4. https://www.kdnuggets.com/2019/11/customer-segmentation-using-k-means-clustering.html
5. https://www.sciencedirect.com/science/article/pii/S0957417403001337
6. https://dl.acm.org/doi/abs/10.1145/1089551.1089610
7. https://www.tandfonline.com/doi/abs/10.1300/j366v06n03_02
8. https://www.emerald.com/insight/content/doi/10.1108/07363769810235974/full/html
9. https://www.sciencedirect.com/science/article/pii/S0957417403001337

Acknowledgments:

We would like to thank our Principal Dr. MD Patil Sir and the Head of Computer Engineering Department, RAIT, Dr. Leena Ragha ma'am for their constant support to all students. We would also like to thank our project coordinator Dr. Dakhane Dhananjay M. and our project guide Mr. Kamlesh Nenwani for always guiding and motivating us throughout the course of this project. Last but not the least we would like to thank our friends and families who always stood by us, as an encouraging force to work on this project, even in the midst of a worldwide pandemic .

Date:_____