

**MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY**

**SCHOOL OF COMPUTING AND INFORMATICS**

**DEPARTMENT OF COMPUTER SCIENCE.**

**MARKET GROUPING SYSTEM**

By

Aduke James

A Project Proposal Submitted to the School of Computing and Informatics in

Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in

Computer Science of Meru University of Science and Technology

- **January 2020 -**

**DECLARATION**

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at Meru University of Science and Technology.

Aduke, James CT201/0086/2016

Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Certification**

The undersigned certify that he has read and hereby recommend for acceptance of Meru University of Science and Technology a Project Proposal entitled Bursary Allocation System.

Dr. Makau Mutua, Ph.D.

**Signature:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Date: \_\_\_\_\_**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Computer Science Department

Meru University of Science and Technology

P.O Box 972-60200, Meru, Kenya**Abstract**

Market segmentation has been one of the best innovations in marketing since its inception. It has had vast benefits for both the businesses and the consumers. Segmentation can be seen in most aspect of various types of businesses today, especially those that exist in the internet.

This research study seeks to find out how you can integrate market segmentation to local retail store such as supermarkets. Local retails stores are at the heart of identifying human and social needs of a particular locale. Segmentation in these stores would mean that the stores would get maximum return on their investments and their customers would also get the products they need instead of going to drastic measure such as online shopping which is time consuming before the product gets to the customer.

I seek to perform this segmentation through machine learning and in particular data science. I would simulate this relationship through a dataset of some London retails stores sourced from Kaggle. Through analysis of these datasets using data science modules, we should be able to get descriptive analysis of how to create segments that are a fit for the local retail store.

Table of Contents

[CHAPTER 1: INTRODUCTION 1](#_Toc30064535)

[1.0 Overview 1](#_Toc30064536)

[1.1 Background Study 1](#_Toc30064537)

[1.2 Problem statement 2](#_Toc30064538)

[1.3 Objectives 2](#_Toc30064539)

[1.3.1 Main Objective 2](#_Toc30064540)

[1.3.2 Specific objectives 3](#_Toc30064541)

[Justification 3](#_Toc30064542)

[1.4 Significance 4](#_Toc30064543)

[1.5 Scope 4](#_Toc30064544)

[1.6 Limitations 5](#_Toc30064545)

[1.7 Expected Outcomes 5](#_Toc30064546)

[CHAPTER 2: LITERATURE REVIEW 6](#_Toc30064547)

[2.1 Types of market segmentation 7](#_Toc30064548)

[2.2 Steps in Market Segmentation 8](#_Toc30064549)

[2.3 Tools and techniques in market segmentation 10](#_Toc30064550)

[2.3.1Cluster Analysis 10](#_Toc30064551)

[2.3.2 Segmentation Trees 11](#_Toc30064552)

[2.4 Application of market segmentation 11](#_Toc30064553)

[2.4 Market segmentation in local retail stores 12](#_Toc30064554)

[2.4.1 Significance of market segmentation to retailers 13](#_Toc30064555)

[2.5 Related Work 13](#_Toc30064556)

[2.5.1Dispalyr (formerly Data Cracker) 13](#_Toc30064557)

[2.5.2IBM SPSS 14](#_Toc30064558)

[2.6 Proposed Project Framework 14](#_Toc30064559)

[CHAPTER THREE: RESEARCH METHODOLOGY 15](#_Toc30064560)

[3.1 Introduction 15](#_Toc30064561)

[3.1.1Requirements analysis 15](#_Toc30064562)

[3.1.2 Design 16](#_Toc30064563)

[3.1.3 Implementation 16](#_Toc30064564)

[3.1.4 Testing 16](#_Toc30064565)

[3.1.5 Delivery 16](#_Toc30064566)

[References 17](#_Toc30064567)

# CHAPTER 1: INTRODUCTION

## 1.0 Overview

In this chapter we will be discussing the current state of market segmentation and define the exact problem with market segmentation in the local business community. We will then define the solution proposed to solve this problem.

## 1.1 Background Study

Segmentation refers to breaking the market into smaller groups so that like customers may be served with products that meet their needs. Kenya’s economy is relatively strong however it has been plagued by corruption and reliance upon primary goods whose prices have remained re­latively low. Factors have led to a disparity in the cost of living for many citizens as it has kept increasing causing some of the population to have to change their way of lives entirely. These changes have not been reflected by many businesses in the country as a lot of businesses are unable to meet their customer’s needs. A large number have not embarked on the practice of customer segmentation.

Some of the businesses have been able to do so are mostly online e-commerce stores. This owes to the large of user data they possess and also the ease of accessibility of this data as most of this transaction and dealings of e-commerce store are done on one platform so everything is readily available for manipulation.

Most businesses like brick and mortar stores are not able to do this due to the lack of this kind of customer data. Organizations are environment dependent. They receive inputs from the environment and they sell or distribute their products to the environment. Companies are created to produce either goods or services that meet the demands of consumers. Companies perform different functions in the process of producing goods or services. They include production, marketing, financial activities, and the management of human resources. Marketing is a social and managerial process by which individuals and organizations obtain what they need and want through creating and exchanging value with others. In a narrower business context, marketing involves building profitable, value-laden exchange relationships with customers. Hence, we define marketing as the process by which companies create value for customers and build strong customer relationships in order to capture value from customers in return.

Companies now recognize the need to appeal to all buyers in different ways. Buyers, due to the increasing s divisions caused by the state of the economy, are too numerous, widely scattered and varied in their needs and buying practices. Moreover, companies themselves vary widely to service different markets. This lack of ability to be able to properly appeal to local buyers has led to buyers turning to foreign businesses who better suit the customer’s needs and are at times cheaper. This thus shows there is an urgent need to for many companies and businesses locally to move away from mass marketing to identifying market segments, selecting some segments and developing products and marketing programs tailored for each of those segments.

## 1.2 Problem statement

The inability of companies to be able to appeal to all buyers in the same market due to different factors such as mass marketing strategies, differences of income in the market, government regulation. These factors have led for the need for them to be a more tailored way to be able to meet the customer’s needs and also making the business focus on market that will offer the most returns to the business.

The approach I intend to use to solve this problem are using data science modules to analyze the data that the business is meant to collect from their customers and thus deriving different segments for the business’ market.

The solution would be dynamically change with different changing attribute values of the market this is because it will be able to learn through the machine learning technologies that will be applied to the data.

## 1.3 Objectives

### 1.3.1 Main Objective

The main objective of this project will be enabling the ability to segment the market for local businesses. This will help the business have clear decision-making processes in regards to what products the customer needs. This will also help business make better returns on their investment due to clear business goals the results that the solution will provide.

This will help the business have proper customer categories such that the business will be able to better marketing decisions

### 1.3.2 Specific objectives

Created market groups based on dataset given

Provide visualization for customer data

Provide insight from the segmentation results

## Justification

The purpose of this project was to examine market segmentation in local businesses in particular retail stores. While some e-commerce store already performs some type of market segmentation a large number of businesses do not have any kind of segmentation approach. This has shown as a lot of the market is being lost to foreign businesses which in turn leads to a lack of promotion of the local economy. These foreign businesses use segmentation and other marketing strategies such as online marketing to capture different customers.

Market segmentation as a solution for this problem would prove very useful for businesses. There will be a better understanding of customer’s needs since they tend to differ depending on a number of factors such as income, cultural beliefs, government regulation and environmental factors

It would also prove very useful for businesses as they will be able to achieve one if its core functions more effectively. That is making returns on investments by stakeholders in the business. Customers have different disposable income. They are therefore different on how sensitive they are to price. By segmenting markets businesses can alter average prices and subsequently enhance profits

## 1.4 Significance

The marketing landscape all over the world has changed significantly. The rise of e-commerce stores means that marketing is no longer targeted to just local customers but international customers too. Amazon, as an example, have been very successful in reaching and gaining market all over the world. This means that small local businesses and industries have to put in extra effort to be able to retain the local market that they need for their businesses to flourish. Because of the internet, the world has become more connected and the needs and wants that people had ten years ago are not the same as the needs and wants of today.

The availability of data will be a key factor for market segmentation locally. The process of using market segmentation with the data given starts before a business’ products even enter the market as the business will develop qualities and features that meet the customer’s needs. For example, a clothing business may develop clothes for plus size women specifically thus getting a whole new market through their segmentation analysis.

As the business develops clear understanding of who your customers are and what they specifically want, it will be able to specifically target marketing and advertising towards them. For example, if a business sells nutritional supplements, you make products targeted towards men, children, expectant women and older women. It will take a different strategy to reach each of these diverse groups. To advertise children’s vitamins, you may run commercial on children’s television shows or in parenting magazines. Understanding each of the market segments is an important tool for using advertising funds effectively and casting a smaller successful net.

## 1.5 Scope

The project would cover local retailers in the Central and Eastern Region. It seeks to take data such as the store inventory stock identification of different products and user data if available to use. It will also use data science modules Pandas, NumPy and TensorFlow. The major issue would be to collect authentic data from the retailers so there will be use of realistic similar datasets from other organizations.

## 1.6 Limitations

1. If implemented it might increase operation costs. This is because the firm is trying to serve different segments.
2. If the characteristics of a certain segment change, investment made initially in that segment might become void

## 1.7 Expected Outcomes

* Segmentation of the market based on the datasets given
* Clear descriptive analysis of the output
* Clear predictive analysis of the output

# CHAPTER 2: LITERATURE REVIEW

Throughout history customer needs have kept growing and have become increasingly diverse and organizations needed to figure out a way to meet these complexities. Segmentation provides organizations with a way to handle this compleity by identifying homogenous groups of customers or target markets (McDonald & Dunbar , 2010). Segmentation was initially designed to assist with consumer goods, but it has since become available for any products or services in all markets including business to business (Clarke & Freytag, 2008)

Market segmentation is the dividing of a customer base into groups that have similarities in different ways that are relative to marketing such as spending habits, gender, age and interests. The premise of segmentation is that all of an organization’s customers are not the same in their goals, needs, expectations and behaviors. It is only occasionally that there are such similarities. Organizations are also not able to cater to a customer individually since their resources rarely permit such tailored servicing. This is the same in digital enable one to one marketing as decisions on where to focus resources and priority target segments need to be identified.

Segmentation considers how to best structure a market, chooses the most attractive customers groups on which to focus its resources. This enables the organization to develop a persuasive case for themselves to the targeted customers and ensures that the organization goes to market in a far more prepared and engaging manner with clear positioning. Segmentation seeks to understand what the needs of customers and how they will buy it, grouping like-minded ones together and then tailoring a proposition together for the customers they choose to go after.

However, Segmentation goes further than the division of customers in a market base. It is at the core of creating a target market strategy, deciding which customers to serve and which not to focus on, understanding how to attract and retain the most suitable customers and knowing through what means to compete successfully. Not only is market segmentation at the core of marketing strategy, it is part of an organization overall strategy. Organizations use segmentation for many reasons. Smart target marketing and efficient utilization of resources is the core benefit (Dibb & Simkin, 1996).

## 2.1 Types of market segmentation

Four segmentation types have emerged as the most popular in segmentation studies (Kotler, Armstrong, Saunders, & Wong, 2002):

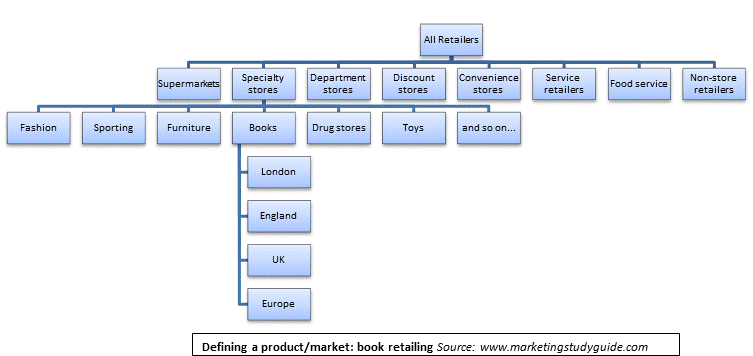
1. **Demographic segmentation**, such as age, gender, income, has been widely used. That works well, when demographics are highly associated with needs and wants. However, such an association may often not be the case, as two people with the exact same demographic characteristics may have very different needs and therefore exhibit different buying behaviors.
2. **Psychographic segmentation** has become more popular as it reflects people’s lifestyles, attitudes and aspirations. Psychographic segmentation can be very useful in strengthening brand identity and creating an 11 emotional connection with the brand, but may not necessarily result in sales.
3. **Behavioral segmentation** is based on product consumption-related behaviors and can include frequency, volume and type of product usage. This type of segmentation can be very powerful for firms that have a membership-type relationship with customers, for example, via a contract such as banks and telecommunications providers, or via loyalty programs. Here, firms can exactly observe consumption behavior. A drawback is that firms typically can only observe the behavior with regard to their own products, but not those of their competitors.
4. **Needs-based segmentation** groups customers based on similar needs and wants, or benefits sought, with regards to a particular product or consumption context. Needs-based segmentation is perhaps the segmentation truest to the marketing concept, that is, satisfying customers’ needs and wants. For companies to increase their sales, segmentation requires understanding customer needs, including those that are underserved or even unmet.

The segmentation base chosen to subdivide a market will depend on many factors such as “the type of product, the nature of demand, the method of distribution, the media available for market communication, and the motivation of the buyers”.

## 2.2 Steps in Market Segmentation

**Step1- Define the market**

The first step is to clearly define the market of interest. For instance, if we want to segment the market for the market for a firm that operates a chain of book stores. If we were to define the market as all retailing consumers, it would be unlikely that we would make any meaningful segmentation. We need to split the broad market into sub markets as shown in the figure below .

We can also sub divide further if the markets are too broad

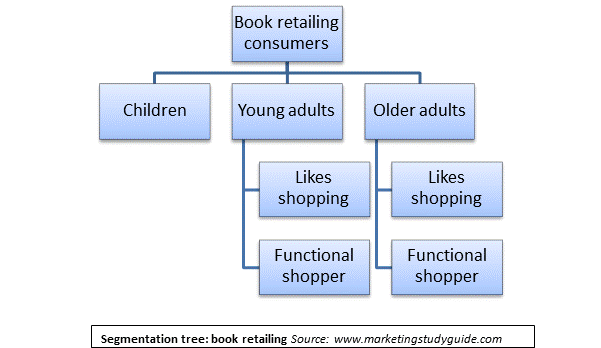
*Figure 2.1: Define a book retailing retail market*

**Step 2 - Create market segments**

Now that we have defined the market, we need to deteremine what type of different customers that form that overall market. To do this we need to review the types of market segmentations and choose two or three of those types

|  |  |  |
| --- | --- | --- |
| Main Category | Segmentation Base | Example/s |
| Demographic | Age group | Pre-teens, teens, young adults, older adults |
| Behavioral | Shopping style | Enjoys shopping, functional, avoids |

Now that we have chosen the segmentation types we can map out segmentations as shown below.



*Figure 2.2: Define a book retailing market segments*

**Step 3 – Evaluate the proposed market segments for viability.**

Now that we have created some market segments, we may be required to evaluate them to check their viability to ensure they are usable and logical. To do this you would need to evaluate them against a check list of factors.

If the segments do not meet the evaluation criteria then you would have to revisit the second step.

**Evaluation Criteria Our Assessment**

Homogeneous Segments should be similar in needs

Homogeneous Assumes that age groups vary in needs, which is likely in this market

Measurable Market research data can be utilized

Substantial Given the segments are relatively broad, they should be individually substantial

Accessible various merchandising techniques can be used to promote the reach of these segments

Practical the organization is able to market to each segment, if required

**Step Four – Evaluate the attractiveness of each segment**

If you were to select one target market from the list of market segments, then some form of objective assessment would be needed. Some of the factors to be considered would be:

**Financial Issues:** Segment size, Segment growth rate, Profit margins

**Structural Attractiveness:** Competitors, Distribution channels

**Strategic Direction:** Fit with firm’s strategy, Fit with firm’s goals

**Marketing Expertise:** Resources, Capability, Branding

**Step 5 – select target market**

Using the assessment information, you have just created, you can select the most appropriate target market for the organization. While there are many factors to consider, you should take into account the organizational strategy, competition rivalry for the segment and the firm’s ability to successfully compete.

## 2.3 Tools and techniques in market segmentation

### 2.3.1Cluster Analysis

This is a statistical technique that can be defined as an analysis designed to categorize objects to a pre-defined number of different groups, with each group being relatively similar on a range of selected attributes. The resultant groups are referred to as clusters. Therefore, it allows marketers to look at the information they have gathered and use it to identify segments they can market to. However, in order for one to be able to use cluster analysis, they first have to know the customers’ demographics and characteristics. Once one has this information, you are then able to enter various variables such as customer service as well as a demographic such as age into a customer analysis program.

The results show segments that will act as potential customers for your product or service. One of the most commonly used cluster analysis software programs is SPSS by IBM.

### 2.3.2 Segmentation Trees

The second technique involves a marketer creating a database from scratch. A segmentation tree can, therefore, be defined as a variation of a decision tree, which visually shows the division of a market into smaller possible market segments. This second technique is very helpful especially when there is lack of valid statistical or research data. Similar to a family tree, a segmentation tree starts with the name of the product or service at the top. It (the tree) is then broken down into branches made up of potential markets. Below is an example of a segmentation tree with the product being cereal.

Once a segmentation tree has been created, the end result is a list of segments one can evaluate for potential marketing opportunities. However, it is advisable for one to create multiple drafts of  
segmentation trees before settling on the one best fit for your task.

## 2.4 Application of market segmentation

A good example of market segmentation and how an organization markets to those customer groups is in the banking industry. Most commercial banks service a wide range of customers, many of whom have relatable monetary goals and life situations. If for example a bank wants to market to individual who have just recently become parents, it would condict research and find out that retirement planning is the most important aspect of their financial needs. The bank would therefore market tax-deferred accounts to this particular segment.

Furthermore, if the same bank wants to effectively market to millennials ( individuals who were born mid-1980s to early 1990s ), it would conduct research and find that most millennials want to have a family. The bank uses this data to make college friendly savings and investments accounts to this customer segment.

In the sports and entertainment industry, market segmentation strategies are often used to fill seats. Instead of a uniform price for all tickets, many theatres and stadiums segment their offerings by price point. For example, a fan who attends a basketball game may find standard seats at a low price. He/ she may also find court-side seats which give them better and closer access to the actions and players. For a more luxurious experience, there may be private suites for the game that feature a catered meal, air conditioning and flat screen TVs to watch the action. By offering different levels of service and amenity to each segment of the market, the venue ensures that all consumers will be satisfied with the experience.

## 2.4 Market segmentation in local retail stores

Retail marketing deals with identification of human and social needs and meeting those needs. It is typically seen creating promotion and delivering goods and services to retail consumers. The marketer has two options to be able to satisfy these needs – first they could approach the customers with an identical marketing approach which is known as mass marketing. Secondly, they could approach different set of customers which is market segmentation. Like many other sorts of businesses retail markets may enjoy the benefits of segmenting their market. Increased competition makes mass marketing not feasible all the time.

Customers have diverse range of retails formats to shop and factors such as distance are not an obstacle these days. A customer could, for example buy an electronic item from a nearby shop, he may also visit an electronic shop to buy the same. In order to attract and retain customers into smaller groups and approaches them with different set of promotional programs.

In evaluating different market segments, a retailer considers two different market segments:

1. The market segment’s attractiveness
2. The organizations objectives and overall resources

### 2.4.1 Significance of market segmentation to retailers

1. **Deciding store location**

Segmentation would assist the retailer in deciding on where to place the organization’s new outlets in the case of expansion. The retail store may be set up as per the concentration of the target segments.

1. **Understanding consumer behavior**

It will enable a retailer to understand why consumers behave differently in a set of marketing and promotional offers. When a heterogenous market is divided into homogenous groups, the retailer can easily develop an effective marketing and promotional strategy.

1. **Deciding merchandise assortment**

One of the decisions that always bothers a retailer is which inventory should be displayed and showed on shelves. Once a market is segmented, this decision becomes easier to make. If a merchandise decision is to be successful good understanding of the particular market is essential.

1. **Deciding promotional campaigns**

Retailers will be able to decide and develop accurate promotional campaigns that hit the target at the right and the right place.

## 2.5 Related Work

### 2.5.1Dispalyr (formerly Data Cracker)

This is a tool used for online retailers to get feedback from customers through surveys. It then weighs the survey data, clean the data and give valuable results base on the survey data

This system is very insightful when it comes to capturing behavioral data from the customers as you gain valuable insights spend their time and money in an online setting. You are able to uncover more specific insight as opposed to data on what the user buys

This system is very time consuming as you have to make surveys that are very objective and also to get the users to participate online might not be easy. It would also require a lot of resources on the production environment of the system

### 2.5.2IBM SPSS

This is a software platform offers advanced statistical analysis, a vast library of machine-learning algorithms, text analysis, open-source extensibility, integration with big data and seamless deployment into applications. Its ease of use, flexibility and scalability make SPSS accessible to users with all skill levels and outfits projects of all sizes and complexity to help you and your organization find new opportunities, improve efficiency and minimize risk.

This tool is very user friendly by providing its end results in a compact form. If you are to analyze your data with this tool it results are easily understandable and it generates very comprehensive reports. A user can easily make decisions from such data.

The major downside with software is it is expensive. Documentation on the software is sometimes very difficult to find.

## Proposed Project Framework

My proposed solution is collecting data from the customers activities in the business, for example, in a retail store the products the users by their stock identification, the time that the product is bought, pricing of the product. Using this data, a dataset is derived. Using data science modules and machine learning we would train a model to read the dataset and derive different customer segments for a give business. But in my research project I intend to use am dataset from Kaggle which has information from local retails stores in London.

# CHAPTER THREE: RESEARCH METHODOLOGY

## 3.1 Introduction

In this project I used agile methodology. Agile Methodology combines iterative and incremental process models with attention to process adaptability and customer satisfaction by rapid delivery of working software product. Agile methodology divides the product into small incremental builds. The builds are later provided in iterations.

I intend to use agile in this project due to the refining nature of machine learning.



*Figure 3.1: Agile methodology steps*

### 3.1.1Requirements analysis

In this stage the requirements of the iteration based on the product backlog, sprint backlog, customer and stakeholder feedback. The functional and non-functional requirements are determined in this phase

### 3.1.2 Design

After the requirements are clearly stated, design of the software begins. U Unified Modeling Language (UML) diagrams and system architectural designs were used in this phase. The tools used will include:

* **Data flow Diagrams –** this will be used to model the flow the system the process aspects graphically.
* **Activity Diagram –** this diagram is used to show how operations in the system are connected and will relate with each other.

### 3.1.3 Implementation

The system will be implemented based on the design came up with at each iteration of the process. I intend to use different machine learning tools to implement this system. The system will be a web-based system. The tools would include:

* **Python –** due to its data cleaning and data processing capabilities. I will use it in the backend to connect it with database if need be.
* **MySQL –** to store my data and data sets
* **TensorFlow –** this will be used for simulating my neural networks

### 3.1.4 Testing

The system will be tested with regards to the functional requirements. Each module will be tested separately and the system will then be tested as a whole. The following will be done:

* **Unit testing –** the smallest unit in each module will be tested from bottom up
* **Functional testing –** the system will be tested against its functional requirements. The input and the output will all be examined
* **Acceptance** **testing** – the adaption of the system will be tested also. The ease of use of the system will be tested and if the output is satisfactory to the users

### 3.1.5 Delivery

Deliver the working iteration into production. The system will be deployed for use by various stakeholders.

## 3.2 Research Design

The research was based on local stores and supermarkets in Meru town and how they try to form relationships with their customers. The main reason I chose this area because I observed there is a disparity between different stores and supermarkets. Some stores were performing visibly and immensely better than others. While others struggled to retain a regular customer base. This was also a very familiar area for me and I would be easier to understand since I frequent some of this stores and supermarkets.

### 3.2.1 Sample of the stores

For some studies, the store may be small enough to warrant the inclusion of all of them in the study. But this study entailed a large population which could not all be studied. That portion of the population that was studied is called a sample of the stores. For some of the supermarkets the sample number of customers monthly is not enough to warrant any kind of segmentation. Some of the supermarkets the customer base is generally the same so customer segmentation does not make sense to perform any kind of segmentation to them

|  |  |  |
| --- | --- | --- |
| **Supermarkets** | **Number of customers(monthly)** | **Location** |
| Mathais | 3000 | Meru |
| Magunas | 6000 | Makutano |
| Budget | 2300 | Meru and Makutano |

### 3.2.1 Sampling Technique

The study used both purposive sampling. Purposive sampling was used to select the head of section (Supermarket manager). I would then issue the supermarket managers with questionnaires for them to answer relevant questions to my study and information needed

A sample size of 30 percent is a good representation of the population (Alreck & Settle, 2014). The study had a sample size of 32 percent of the target population.

### 3.2.2 Data collection methods

Prior to data collection I decided to carry out a survey on a sample of 20 members of staff from different stores and supermarkets in Meru who were not in any way involved with the main study. This survey made sure that the questionnaires were very clear to respondents and helped to estimate the response rate and completion time. This survey also helped to estimate the validity of some of the questions and the reliability of data to be collected.

Questionnaires were issued and picked from respondents taking approximately 2 days to achieve a minimum sample of 20 respondents.

The study used both primary and secondary data. Primary data was collected by use of Questionnaires. The four groups of respondents were issued with questionnaires each particular to their role. Questionnaires had both closed and open-ended questions.

### 3.2.3 System configurations

#### 3.2.3.1 Software configurations

Operating system        : Windows 10

Front End : JavaScript

Back End :Python,

Text Editors : Visual Studio Code, Sublime Text.

Software used : Jupyter Notebook, Streamlit

#### 3.2.3.2 Hardware system requirements.

**Processor** - Intel Atom® Processor Z2520 1.2 GHz, or faster processor

**Storage** - Between 850 MB and 1.2 GB, depending on the language version

**RAM**  - Minimum of 512 MB, 2 GB is recommended

**Video**  - 1280 x 800 pixels or higher on a 10-inch device

## 3.3 System Design

System design is process of defining elements of system like modules, architecture, component and their interfaces and data for a system based on the specified requirements. For this research it was implemented through a flow chart which is a graphical modeling language that involves a schematic or stepwise representation of an algorithm.

I used the flowchart is because flowcharting involves a sequence of steps that require decisions, furthermore it’s easier to learn how the overall system operate since distinction can be made easily on different categories of users in the system.

Analyst

Model Evaluation

Segment

Upload file

Segmented data and charts

## 3.4. Data processing and analysis

Considering the quantitative and qualitative nature of the study, data was analyzed both qualitatively and quantitatively. Quantitative data upon collection was cleaned, thermalized and then analyzed. Analyzed data was presented in form of tables and graphical presentations. Content analysis was used to analyze narrative data obtained from the open-ended questions.

### 3.4.1 Ethical considerations

This related to moral standards that the researcher should consider in all research methods in all stages of research design after approval from Meru University of Science and Technology was obtained to conduct the study. The research was based on the principles of Belmont report namely beneficence, respect for human dignity as well as justice.

### 3.4.2 Principle of beneficence

The principle means “above all does no harm”. The principle contains broad dimensions such as freedom from harm and exploitation as well as the researcher’s duty to evaluate the risk or benefit ratio.

### 3.4.3 Principle of human dignity

This principle includes the right to self-determination and full disclosure. Right to self-determination means that prospective participants should not be coerced into taking part in the study that is participants have the right to decide whether to participate without incurring any penalty. Right to full disclosure means the researcher has fully explained the nature of the study and the person’s right to refuse participation.

### 3.4.5 Principle of justice

This principle includes participant right to fair treatment and privacy. Fair treatment includes that the selected participant inclusion was based on requirements of research. The right to privacy means that the information provided by participants will not be shared without their will. An injustice occurs when some benefit to which a person is entitled is denied without good reason or when some burden is imposed unduly.

# CHAPTER 4: SYSTEM ANALYSIS

## 4.1 Overview

System analysis is the process of identifying the problems of the system, collecting and interpreting facts and decomposition of system into components in order to achieve a specific goal.

In this chapter we studied the current customer segmentation systems that are applied in the marketing industry and study them in depth. We then describe the proposed system in terms of functionality. We will then find out the user requirements both functional and non-functional requirements.

## 4.2 Description of the current systems

One of the current systems of segmenting customers for a business is segmenting them on the fly. For example, if an e-commerce store would like to segment its users on the fly it would do it for example how the customers go to the website in the first place. One segment would consist of customers who came to the website using affiliate links, advertisements and website notifications. The other segment would consist of customers who came to the website through influencers, for example, a linked in advertisement, google advertisements, online coupon or word of mouth

One of the software that is used is IBM SPSS statistical software program. SPSS would use the decision trees algorithms to segment market. In this algorithm they need a target variable that was added to the data after a classification from the organization itself. They then feed the spreadsheet of customer information into their Data Editor. They define the variable properties and some more variables are added. Variable which are selected are mostly variable which represent some type of ordering. Variable which will be used to make the tree model are then moved to the independent variable section. The target variable is classified as a dependent variable. The decision tree output format is chosen and the rules around it being specified. The program then segments with the decision tree and returns the segmented data results.

## 4.3 Feasibility study

This is the preliminary study that I conducted to determine whether the project was financially, technically and operationally practical enough to be pursued. It would also determine whether the project might be profitable to a business or makes sense in the long run.

The other objective of the feasibility study is to determine the specific reasons for developing the software. Can the software that is to be made acceptable to the users, would be adaptable to change in technology would it conform to company standards. Some of the information considered during the feasibility study include cost of developing and maintaining the software, benefits of the software to the organization.

### 4.3.1 Technical feasibility

In this part of the study we assessed whether the resources available such as software and hardware and the technology required to accomplish the requirements for the software within the allocated time and within the budget for the project. The software developer ascertained whether the current resources and technology available would be enough or some resources needed to be added to complete the work in time.

It also involved performing some of the following tasks:

* Analyze the technical skills of the personnel meant to work on the project
* Ascertain whether the technology chosen is stable enough and constantly maintained by their creators

### 4.3.2 Schedule feasibility

We analyzed the projects constraints and whether they could be reasonably navigated and met. Given the technical expertise that the software developer has were some of the project deadlines reasonable. Which deadlines would be mandatory and which would be desirable but not as important to the project as a whole? Depending on the answer provided on the previous question the analyst would propose alternative schedules. It is preferable and always recommended to deliver a properly functioning information system a couple of months later than to deliver and error prone and useless system at the desired and agreed upon time

### 4.3.3 Economic feasibility

In this part of the feasibility study we determine whether the system to be developed is capable of producing some financial gains for the targeted organization. Cost incurred for the software development team, estimated cost of software and hardware, cost of performing all necessary preliminary studies, so on and so forth are considered in economic feasibility. Expenses made on purchased and services offered such as hiring of software developers. Then we also considered the benefits that could be drawn from the developing of the proposed software to the organization and parties involved.

It focuses on the issues below:

* Total cost on software development with a view to produce long terms gains
* Total cost required to perform a full preliminary study such as requirement analysis and requirement elicitation
* Cost of hardware, software development team and training

### 4.3.4 Operational feasibility

We determined the extent to which the proposed system carries out steps to help solve real business problems and the user requirements. This feasibility study is mostly dependent on human resource part of the project and it involves whether the software will operate after it is developed and is operative once it is installed.

The following tasks are also performed in operational feasibility:

* Determines whether the problems anticipated in user requirements are of high priority.
* Determines whether the solution suggested by the software development team is acceptable.
* Analyzes whether users will adapt to a new software (drivers and the mechanics).
* Determines whether the organization is satisfied by the alternative solutions proposed by the software development team.

# CHAPTER 5: SYSTEM ANALYSIS

## 5.0 Overview

In this chapter we give a detailed description of the proposed system. Some of the components used are diagrams such as Data Flow Diagrams. We also discuss the functional and non-functional requirements of the proposed system as well as the complete logic and physical design of the proposed system

## 5.1 Description of the proposed system

The proposed system seeks to perform market segmentation with the use of machine learning algorithms to segment the significant market data. Development of the application is currently on the web. Other platforms such as Windows and MacOS would be considered for the future depending on feedback from the users.

The proposed system would take customer data in whatever form but preferably csv format. The machine learning model then takes over at this point and check for any null and NA values to make sure the data provided is well formatted. The model then starts forming features to be used. For this specific case three features from the dataset provided would be used as per the identification of the customer. These features would also help us with visualization and algorithm explanations in later steps of the program. From my sample dataset from Kaggle which is and e-commerce dataset with sales data from November 2018 to April 2019, the following features were chosen Total spending, Average return rate of the product and number of products ordered. We will prepare this data with pandas and numpy libraries from python.

After some calculations we merge these three features in one data frame which we name the customers data frame. We perform some visualizations to see the distribution of the features. We check for the skewness of the distributions and then begin scaling the features best on the skewness results from some of these distributions.

### 5.1.1 Scaling

K-means algorithm interprets each row in the customers data frame as a point in a 3-dimensional space. When grouping them, it uses the [euclidian distance](https://en.wikipedia.org/wiki/Euclidean_distance) between the data points and the center of the group. With highly varying ranges, algorithm may perform poorly and not be able to form the groups as expected. For K-means to perform effectively, we are going to scale the data using logarithmic transformation which is a suitable transformation for skewed data. This will scale down proportionally the 3D space which our data is spread, yet preserving the proximity between the points. After applying logarithmic transformation, the customers data frame is ready to be fed into the k-means clustering algorithm.

### 5.1.2 Segmentation with k-means clustering

We are going to use [K-means algorithm from scikit-learn](https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html). The k-means clustering algorithm works in this way:

1. Initialize k=n centroids=number-of-clusters randomly or smartly
2. Assign each data point to the closest centroid based on euclidian distance, thus forming the groups
3. Move centers to the average of all points in the cluster
4. Repeat steps 2 and 3 until [convergence](https://www.mathsisfun.com/definitions/converge.html).

While running the steps through, the algorithm checks the sum of squared distances between clustered-point and center for each cluster. Mathematically speaking, it tries to minimize — optimize the within-cluster sum-of-squared-distances *or* inertiaof each cluster**.** When inertiavalue does not minimize further, algorithm converges. Thus, iteration stops. One important consideration is the selection of k. In other words, how many groups should be formed? For example, K-means applied above uses k=8 as a default value.

### 5.1.3 Hyperparameter tuning

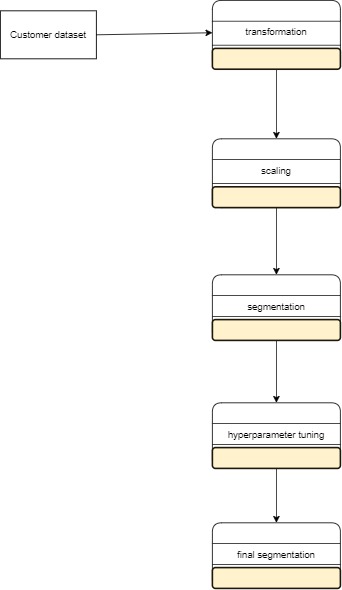
While selecting k, we are going to decide against the optimization criteria of the K-means, inertia, using [elbow method](https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/). We then build a different K-means models with k values 1 to 15, and save the corresponding inertia values. With the elbow method, we are going to select the k value where the decrease in the inertia stabilizes. When k=1 inertia is at the highest, meaning data is not grouped yet. Inertia decreases steeply until k=2. Between k=2 and 4, the curve continues to decrease fast. At k=4, the descent stabilizes and continues linearly afterwards, forming an elbow at k=4. This points out the optimal number of customer group is 4.

### Data points are shown in spheres and centroids of each group are shown with cubes. 4 customer groups are as follows:

* **Blue:** Customers who ordered at least one product, with maximum total spending of 100 and having the highest average return rate. They might be the newcomers of the e-commerce website.
* **Red:** Customers who ordered 1 to 4 products, with average total spending of 150 and a maximum return rate of 0.5.
* **Purple:** Customers who ordered 1 to 4 products, with average total spending of 300 and a maximum return rate of 0.5.
* **Green:** Customers who ordered 1 to 13 products, with average total spending of 600 and average return rate as 0. It makes the most favourable customer group for the company.

The overall strategy would be preserving the most favorable customer

### 5.1.4 Data flow diagram



# CHAPTER 6 : IMPLEMENTATION

# References

(2020). Retrieved from https://www.displayr.com.

Baker, M., & Saren, M. (2000). *Marketing theory: A student text.* SAGE Publications Ltd.

Clarke, A. H., & Freytag, P. (2008). An intra- and inter-organisational perspective on industrial segmentation: A segmentation classification framework. *European Journal of Marketing*.

Dibb, S., & Simkin, L. (1996). *The Market Segmentation Workbook: Target Marketing for Marketing Managers.* Cengage Learning EMEA.

Geoff , F. (2017). *A Step-by-step Guide to Segmenting a Market*. Retrieved from Segmentation Study Guide: https://www.segmentationstudyguide.com/understanding-market-segmentation/a-step-by-step-guide-to-segmenting-a-market/

Kotler, P., & Wong, V. (2008). *Principles of Marketing.* Financial Times/ Prentice Hall.

McDonald, M., & Dunbar , I. (2010). *Market Segmentation: How to Do It, How to Profit from It.* Goodfellow Publishers, 2010.

*SDLC - Quick Guide*. (n.d.). Retrieved from Tutorials Point: https://www.tutorialspoint.com/sdlc/sdlc\_quick\_guide.htm

# Appendices

## Appendix A: Budget of resources

|  |  |
| --- | --- |
| **ITEM** | **ESTIMATED** **COST** |
| Printing paper | 400 |
| Server cloud storage | 2000 |
| Ubuntu server hosting | 4000 |
| Laptop computer 4gb ram 500gb rom 2.4ghz | 35000 |
| Research cost | 2000 |
|  | 43400 |

## Appendix B: time SCHEDULE

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SEP 2019 | OCT 2019 | NOV  2019 | DEC  2019 | JAN  2019 | FEB  2019 | MAR  2019 | APRIL  2019 |
| Problem definition |  |  |  |  |  |  |  |  |
| Requirement  Analysis |  |  |  |  |  |  |  |  |
| System  design |  |  |  |  |  |  |  |  |
| System implementation |  |  |  |  |  |  |  |  |
| Testing |  |  |  |  |  |  |  |  |
| Documentation |  |  |  |  |  |  |  |  |