



UNIVERSITI M A L A Y A

*Faculty of Computer Science
and Information Technology*

WIA3002 ACADEMIC PROJECT I SEMESTER 2 2024/2025

Project Title:

**A WEB-BASED MODEL FOR MARKET
SEGMENTATION**

Prepared By:

CELINE ONG ZI YING

22004878

Supervised By:

ASSOCIATE PROFESSOR

DR. TUTUT HERAWAN

ABSTRACT

The purpose of developing a web-based market segmentation system is to help SMEs businesses better understand their customers. Discovering hidden patterns in the customer segments group allow the business to effectively personalize suitable marketing strategies to promote their products or services. Instead of just focusing on their own assumption to group customers, using the AI driven segmentation can help them better understand their customers. However, SMEs often struggle with complex, or overly technical market segmentation tools that require real-time data, technical expertise, or third-party integrations. There are a few limitations such as not supporting hybrid and a posteriori segmentation, and the report feature. Therefore, a new web-based market segmentation system has been proposed to address the issues they faced. The development process applied Agile methodology. Due to its flexibility and iterative approach, the system development can be improved, modified or enhanced based on the suggestion from the collaborator and stakeholder. The data collection was conducted via literature review, surveys and interviews with the SME stakeholder and a collaborator with relevant experience in business and economics from Indonesia. The new proposed system is defined based on a combination of research results to ensure it meets real business needs. The modules provided include a login module, registration account module, edit profile module, dataset management module, dataset selection module, segmentation module, report module, feedback module and admin module. The system aims to help SMEs businesses make informed marketing decisions through accessible and meaningful customer insights.

Keywords: segmentation, SMEs, customers, marketing, system

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all those who have supported me throughout my final year project.

Firstly, I would like to thank Associate Professor Dr. Tutut Herawan for giving me the opportunity to be one of his supervised students in the Academic Project. His suggestions and guidance along the way truly helped me to complete the project. His support was incredibly valuable to me.

Secondly, I would like to express my gratitude to both panelists, Dr. Hoo Wai Lam and Dr. Vimala Balakrishnan, for providing valuable feedback during the monitoring and viva sessions. Their evaluations and comments allowed me to further enhance my project.

Thirdly, I would like to extend my gratitude to the collaborator of this project, Dr. Lucia Ika Fitriastuti, who kindly collaborated and scheduled time to meet with me. During our discussions, she provided useful suggestions to improve my project and was always willing to help resolve my queries during system development.

Fourthly, I would like to thank the stakeholder of this project, Soapan Santun Enterprise, for agreeing to collaborate with me. Their support was incredibly helpful, especially in providing valuable information that contributed to the development of the system. Even though they were busy with their business operations, they still made time to meet and discuss the project with me.

Last but not least, to everyone who supported me along this journey, I am extremely grateful for your suggestions and guidance. I truly appreciate your time and unwavering support, which played a significant role in the successful completion of my project.

TABLE OF CONTENTS

ABSTRACT.....	2
ACKNOWLEDGEMENTS.....	3
CHAPTER 1: INTRODUCTION TO THE PROJECT.....	10
1.1 Background.....	10
1.2 Problem Statement.....	13
1.3 Project Objectives.....	16
1.4 Project Scope.....	17
1.5 Project Timelines.....	18
CHAPTER 2: LITERATURE REVIEW.....	19
2.1 Web-Based Model of Market Segmentation.....	19
2.1.1 A Priori Market Segmentation in Web Systems.....	19
2.1.2 A Posteriori Market Segmentation in Web Systems.....	20
2.2 Segmentation Criteria & Related Models.....	21
2.2.1 Demographic Segmentation.....	22
2.2.2 Behavioral Segmentation.....	23
2.2.3 Geographic Segmentation.....	25
2.2.4 Summary Table of Related Models and Their Evaluation.....	29
2.3 Comparative Analysis of Existing Systems.....	34
2.3.1 Amplitude.....	34
2.3.2 Hotjar.....	37
2.3.3 Google Analytics.....	41
2.3.4 Insightly.....	43
2.4 Proposed System.....	45
2.4.1 Improvements to the Recommended Framework with respect to Existing Systems	48
CHAPTER 3: RESEARCH METHODOLOGY.....	50
3.1 Agile Methodology.....	50
3.1.1 Development Phases Overview (Agile-Based).....	51
3.1.1.1 Planning Phase.....	51
3.1.1.2 Designing Phase.....	51
3.1.1.3 Developing Phase.....	52
3.1.1.4 Testing Phase.....	52
3.1.1.5 Deploying Phase.....	52
3.1.1.6 Review Phase.....	53
3.1.2 Sprint Breakdown Overview.....	53
3.2 Data Collection and Analysis.....	56
3.2.1 Survey.....	56
3.2.1.1 Demographic Information.....	56

3.2.1.2 Knowledge of Market Segmentation System.....	57
3.2.1.3 Ideas on Improved Version Web-based Market Segmentation Model.....	60
3.2.1.4 Additional Suggestion/Feedback.....	64
3.2.2 Interview.....	67
3.2.2.1 Interview With Collaborator.....	67
3.2.2.2 Interview with Stakeholders.....	70
3.2.3 Literatures Survey.....	72
CHAPTER 4: SYSTEM ANALYSIS AND DESIGN.....	73
4.1 System Requirements.....	73
4.1.1 Functional Requirement.....	73
4.1.2 Non-Functional Requirement.....	79
4.2 System Design.....	81
4.2.1 Functional Decomposition Diagram(FDD).....	81
4.2.2 Use-Case Diagram.....	82
4.2.3 Activity Diagram.....	83
4.2.3.1 Register Account.....	83
4.2.3.2 Login Account.....	84
4.2.3.3 User Profile Management.....	85
4.2.3.4 Dataset Management.....	86
4.2.3.5 Dataset Selection.....	87
4.2.3.6 Segmentation.....	88
4.2.3.7 Report Management.....	89
4.2.3.8 Submit Feedback.....	90
4.2.3.9 Admin Management.....	91
4.2.4 Entity Relationship Diagram (ERD).....	92
4.3 Interface Design.....	93
4.4 Hardware And Software Requirements.....	101
4.4.1 Hardware Requirements.....	101
4.4.2 Software Requirements.....	102
CONCLUSION.....	103
REFERENCES.....	104
APPENDIX.....	108
Appendix A: Logbook.....	108
Appendix B: Collaboration Letter.....	109
Appendix C: Meeting with Collaborator.....	111
Appendix D: Questionnaire validation form.....	112

List of Figures

Figure 2.1 Amplitude - Cohort Creation Page.....	34
Figure 2.2 Amplitude - Journal Analysis List Page.....	35
Figure 2.3 Amplitude - Dashboard Page.....	35
Figure 2.4 Amplitude - Upload Data Page.....	36
Figure 2.5 Hotjar - Session Replay Page.....	38
Figure 2.6 Hotjar - Survey Page.....	38
Figure 2.7 Hotjar - Dashboard Page.....	39
Figure 2.8 Hotjar - Filter Function Page.....	39
Figure 2.9 Hotjar - Integration Page.....	40
Figure 2.10 Google Analytics - Create Segment Page.....	42
Figure 2.11 Google Analytics - Dashboard Page (Google Analytics).....	42
Figure 2.12 Insightly - Lead Page.....	44
Figure 2.13 Insightly - Dashboard Page.....	44
Figure 3.1 Agile Methodology.....	50
Figure 3.2 Agile Planning.....	53
Figure 3.3 Survey - Industry Type.....	56
Figure 3.4 Survey - Company Size.....	57
Figure 3.5 Survey - Customer segmentation practices among respondents.....	57
Figure 3.6 Survey - Reasons Why Businesses Do Not Perform Customer Segmentation.....	58
Figure 3.7 Survey - Web-Based Segmentation Tools Used by Businesses.....	58
Figure 3.8 Survey - Challenges Faced When Using Current System.....	59
Figure 3.9 Survey - Preferred Method for Providing Customer Data.....	60
Figure 3.10 Survey - Dataset Deletion Preference.....	60
Figure 3.11 Survey - Relevant Segmentation Categories.....	61
Figure 3.12 Survey - User Preference for Visual and Interactive Dashboard.....	61
Figure 3.13 Survey - Perceived Benefits of Interactive Dashboard.....	62
Figure 3.14 Survey - Preference for Automatic Report Generation.....	62
Figure 3.15 Survey - Report Content Expectations in PDF.....	63
Figure 3.16 Survey - Importance of Having Access to Past PDF Reports.....	63
Figure 3.17 Survey - Standardized Dataset Template Suggestion.....	65
Figure 3.18 Survey - User Interface and Usability Suggestion.....	66
Figure 3.19 Survey - Feedback Feature Suggestion.....	66
Figure 4.1 Functional Decomposition Diagram.....	81
Figure 4.2 Use Case Diagram.....	82
Figure 4.3 Activity Diagram of User Register Account.....	83
Figure 4.4 Activity Diagram of User Login Account.....	84
Figure 4.5 Activity Diagram of Manage User Profile.....	85

Figure 4.6 Activity Diagram of Manage Dataset.....	86
Figure 4.7 Activity Diagram of Select Dataset.....	87
Figure 4.8 Activity Diagram of Perform Segmentation.....	88
Figure 4.9 Activity Diagram of Manage Report.....	89
Figure 4.10 Activity Diagram of Submit Feedback.....	90
Figure 4.11 Activity Diagram of Admin Handle Feedback.....	91
Figure 4.12 ERD.....	92
Figure 4.13 Mockup - Landing Page (Before Login).....	93
Figure 4.14 Mockup - User Dashboard.....	94
Figure 4.15 Mockup - Admin Dashboard.....	94
Figure 4.16 Mockup - Login.....	95
Figure 4.17 Mockup - Register Account.....	95
Figure 4.18 Mockup - Dataset Page.....	96
Figure 4.19 Mockup - Dataset Selection Page.....	96
Figure 4.20 Mockup - Mapping Result Page.....	97
Figure 4.21 Mockup - Select Criteria Page.....	97
Figure 4.22 Mockup - Segmentation Result Page (Table).....	98
Figure 4.23 Mockup - Segmentation Result Page (Dashboard).....	98
Figure 4.24 Mockup - Report Page.....	99
Figure 4.25 Mockup - My Profile Page.....	99
Figure 4.26 Mockup - Feedback Page.....	100
Figure 4.27 Mockup - Admin Page.....	100

List of Tables

Table 1.1 Project Schedule.....	18
Table 2.1 Past vs Current Market Segmentation.....	21
Table 2.2 Related Work.....	29
Table 2.3 Related Work Summary.....	32
Table 2.4 Proposed Web-Based Market Segmentation System.....	46
Table 2.5 Improvement of Proposed System Compared to Existing Systems.....	48
Table 2.6 Comparison Between Different Existing Market Segmentation Systems.....	49
Table 3.1 Feedback from Dr Lucia.....	67
Table 3.2 Interview with the Stakeholder.....	70
Table 4.1 Registration Module.....	73
Table 4.2 Login Module.....	74
Table 4.3 User Profile Module.....	74
Table 4.4 Dataset Management Module.....	75
Table 4.5 Dataset Selection Module.....	76
Table 4.6 Segmentation Module.....	76
Table 4.7 Reports Module.....	77
Table 4.8 Feedback Module.....	78
Table 4.9 Admin Management Module.....	78
Table 4.10 Non-Functional Requirements.....	79
Table 4.11 Device Details.....	101
Table 4.12 Software Tools.....	102

List of Symbols and Abbreviations

1. SME - Small and Medium Enterprise
2. RFM - Recency, Frequency, Monetary
3. CRM - Customer Relationship Management
4. STP - Segmentation, Targeting, and Positioning
5. AI - Artificial Intelligence
6. API - Application Programming Interface
7. LRFM - Length, Recency, Frequency, Monetary
8. RFMT - Recency, Frequency, Monetary, Tenure

CHAPTER 1: INTRODUCTION TO THE PROJECT

This chapter presents the introduction of this research. Section 1 presents a research background which discusses the importance of market segmentation in modern marketing strategies, the shift from traditional rule-based systems to more advanced data-driven web-based segmentation tools, and the limitations of current systems. Section 2 discusses the problem statement which highlights core issues that the current system and the problems faced by SME. Section 3 describes the motivation of this project and clearly defines the goals of this research. Under Section 4 will discuss the Project Scope which briefly explains the system features and capabilities included in the system development. In the last section, the Project Timeline presents a week-by-week schedule of key activities to be carried out during the course of the project.

1.1 Background

In the advance of development in the technology era, the use of information technology has become more sophisticated. Knowing customers' information regarding their background was essential to help companies make effective decisions in marketing strategies. Market segmentation, the process of dividing the target market into smaller, homogeneous groups based on the common characteristics (Luthfiandana et al., 2024). This is one of the crucial steps to help companies in making correct and effective decisions in marketing strategies. Customer information no matter from demographics, geographics or behavioral, they can help businesses to better understand customer preferences, allowing them to make informed marketing decisions by producing content that is relevant and appealing to their target market.

With the fast growth of data and rise of cloud computing, conventional rule-based segmentation methods have not been enough to optimize marketing strategies (Osakwe et al., 2023). They tend to be based on manual filtering or static conditions which provide little flexibility, understanding and outdated information. However, web-based segmentation systems have emerged as the modern alternative. Market segmentation has moved to data-driven, web-based applications that incorporate machine learning, a significant tool for building modelling methods for segmenting customers into segment groups. Additionally recent reviews revealed that predicting customer buying behaviors and patterns using traditional methods may not be comprehensive enough to extract correct information to guide decision making and provide a holistic support for accurate

marketing strategies (Osakwe et al., 2023). Most existing systems, however, fall short of integration of machine learning for effective decision making, mostly heavily rely on a priori segmentation logic, such as platforms like Google Analytics ([GA4] Segment Builder - Analytics Help, 2025), Hotjar (Hotjar, 2025), Insightly (Insightly, 2025) and Amplitude (Amplitude, 2025). They tend to be focusing on a priori segmentation logic by providing a certain condition or rules to create the segments and provide limited support on the automated clustering. In addition, they only provide basic report generation functions, typically allowing users to export segment results that are not comprehensive in PDF or CSV formats or some do not even have this function. These exports do not include full segmentation context such as the attributes selected, the segment group result, or insightful summaries of each segment. As a result, users are unable to trace back how the segments were created, making it difficult to compare the past segment groups with current segment groups which is essential for evaluating changes over time and making informed marketing decisions.

Several studies have emphasized the importance of using posteriori segmentation methods which are also called data-driven segmentation approaches. For example, many web-based decision systems lack dynamic personalization and adaptive segmentation logic (Chelioudakis et al., 2024). It shows the need for more advanced and intelligent tools to close the gap on this issue. Furthermore, unsupervised clustering techniques like RFM model in CRM was suggested to be applied on the system to improve the customer segmentation. (Ling et al., 2024) Their study found that traditional systems do not sufficiently use clustering to reveal deeper customer insights. This supports the argument where the existing platforms do not fully utilize the data-driven methods like unsupervised learning for segmentation process. In addition, the combining of RFM analysis with K-means clustering has been shown to help in derive more effective actionable marketing strategies (Akande et al., 2024). It was revealed that many current segmentation tools fail to provide integrated, data-backed outputs, making it difficult for businesses to confidently act on the insights. A similar conclusion was drawn in other CRM-based segmentation studies where it demonstrated that clustering-based tools outperformed the rule-based systems in the segmentation strategy (Al-Turjman & Karaderi, 2024). It is clear that current segmentation techniques are still rigid and have the potential to improve it to more advanced, automated ways to help businesses make right decisions.

Motivated from the above explanation, this project is designed to fill the huge gap existing between the potential of the available market segmentation platforms and the actual demands of SMEs. The suggested framework is specifically useful for SMEs, and academic researchers. The system will support hybrid segmentation, dynamic clustering, and automated report features, offering a smarter and more accessible solution for users in business and research contexts.

1.2 Problem Statement

Market segmentation systems are tools that help businesses split their large customer base into smaller, meaningful groups using categories like demographics, behavior, and geography. This helps businesses understand their customers' preferences and spot new market growth opportunities before tailoring effective marketing strategies. However, **many current segmentation platforms are mainly designed for large enterprises** such as Salesforce, HubSpot, and Tealium. These systems are often complex, feature-heavy, and more focusing on real-time data environments, requiring API integrations and technical infrastructure that most SMEs lack (Tawil et al., 2023). Even among more widely used tools for SMEs businesses like Amplitude, Google Analytics, Hotjar, and Insightly, the segmentation functions often require users to have a certain level of technical expertise. Without professionals who understand how to navigate and use these tools, SMEs may find it difficult to perform meaningful segmentation by using segmentation tools. As a result, SMEs tend to avoid adopting such segmentation tools which can limit their ability to obtain meaningful insights and target their customer groups effectively.

Furthermore, SMEs often rely on manually collected datasets from internal records or website-hosting platforms, rather than automated or real-time sources. They normally **lack expertise such as data scientists to manage, interpret, and analyze datasets effectively** (Omowole et al., 2024). In order to convert the manually collected dataset from internal records or website hosting platforms into a system that supports real-time data integration from multiple sources can be overwhelming for SMEs, as many small businesses do not have the capacity to hire specialized personnel or invest in the technical infrastructure that requires real time implementation. In addition to technological gaps, SMEs face internal challenges such as limited time, and financial resources for implementing effective segmentation strategies. Many SMEs may refuse to change their current organizational structure to adopt more advanced processes such as transitioning from manual workflows to advanced segmentation systems due to the time and effort required. Furthermore, the high costs associated with implementation no matter from hiring experts or investing in advanced segmentation tools can further hinder adoption. (Clarke & Freytag, 2023). These combined limitations highlight the need for a dedicated, web-based segmentation system designed specifically customized for SMEs, the one that supports manual

dataset uploads and provides simplified, user-friendly segmentation tools without requiring technical expertise.

Besides, existing segmentation systems that are most widely used such as Google Analytics, Hotjar, Insightly, and Amplitude **still heavily rely on rule-based or also called as priori segmentation**. This means that businesses have to manually define rules to create customer segments based on predefined attributes such as grouping customers by age ranges, incomes or specific behaviors. For instance, Google Analytics' Segment Builder allows users to create segments by specifying conditions like "users aged 25-34 who visited a particular page" ([GA4] Segment Builder - Analytics Help, 2025). Similarly, Hotjar allows segmentation by filtering visitors based on attributes like session duration or device type (Hotjar, 2025), while Insightly offers dynamic list segmentation through rule-based filters such as customer status or engagement levels (Insightly, 2025). Amplitude also supports user segmentation by using chart filters based on set criteria, like "users who completed a specific event within the last 30 days" (Amplitude, 2025b). While these rule-based methods are straightforward and easy to implement, they fail to uncover more complex or hidden patterns in customer data. However, a posteriori segmentation does the analysis of customer data before doing segmentation. It uses machine learning techniques like clustering algorithms to automatically group customers based on similarities found in their behaviour or characteristics without the need to specify the rules in advance. For example, using a clustering algorithm such as K-means clustering, a system might discover a segment consisting of customers aged 28–36 who live in urban areas, tend to shop on weekends without the need to define these rules in advance based on their own assumption and knowledge. These patterns would be hard to find with manual rule-based filters. They can reveal relationships in the data that were not noticed before. Relying on a priori methods definitely restricts the depth of insights companies can gain. In contrast, a posteriori segmentation offers more meaningful, data-driven groupings that can help shape better marketing strategies. Therefore, we need to improve the segmentation system that uses posteriori to identify valuable and previously unseen customer segments.

Lastly, existing systems **do not provide a clear and comprehensive reporting feature in PDF format that presents all the insights** generated by the system. While Amplitude, Hotjar,

Insightly, and Google Analytics offer basic export options like CSV or PDF, these exports typically lack segmentation context that include the specific attributes used for segmentation, the segment group result, and meaningful summaries of each segment. For instance, Amplitude allows users to export usage reports as PDF or PNG and event segmentation data as CSV files. However, these exports typically show high-level metrics like event counts, funnel charts, or user activity trends without including the segmentation logic and result. (Amplitude, 2025b). This limitation makes it hard for users to look back at how the segments were created and to compare past segment groups with the current ones. Those results are important for evaluating customer trends over time and for making informed marketing decisions. Existing research (Cacciolatti & Lee, 2015) shows that SMEs often rely on informal and unstructured methods for gathering marketing information, which can make systematic analysis and strategic planning difficult. Therefore, to help businesses, especially SMEs, make informed decisions in these situations, it is necessary to put formal processes in place for collecting, analyzing, and presenting marketing data. This is especially true for segmentation information that aids businesses in their marketing strategy choices. Even though the system is able to generate many different useful insights to companies, the user may struggle to always revisit and review back the recommended results easily. If the results are not saved, there is no way for them to retrieve back the results for reference. Without an efficient reporting function, companies need to spend more time manually recording down the segmentation results before they are able to communicate with their stakeholders. Therefore, developing a report generation function and report history viewing feature such as in PDF format into a web-based segmentation system need to be prioritized. This will allow companies to download the PDF report for their own reference and also review the past historical reports when they need to. This function is able to support more confident, efficient, and informed decision making in marketing strategy planning.

Overcoming such challenges is a need to develop a framework of segmentation that not only has a data-driven foundation but is also accessible and practical for SME. This framework should be flexible enough to work with manually uploaded datasets, have the ability in uncovering hidden customer insights through advanced clustering techniques, and equipped with clear, comprehensive reporting features that help users easily track and revisit past analyses.

1.3 Project Objectives

This project embarks the following objectives:

1. To identify the limitations and challenges faced by SME businesses in using current market segmentation systems.
2. To design a user-friendly web-based segmentation system specifically for SMEs businesses, focusing on core segmentation features without technical complexity.
3. To develop a system that supports hybrid and posteriori segmentation, and includes an automated reporting feature.

1.4 Project Scope

This study focuses solely on the development of a web-based system for market segmentation customized for SMEs businesses. The main objective is to assist businesses in identifying and grouping customers into distinct market segments using data-driven, *a posteriori* segmentation. The system also aims to provide meaningful insights into customer segments to help companies make their own marketing decisions.

It supports key functionalities such as uploading customer and order data in CSV format, automated data cleaning on the backend, and the selection of segmentation attributes. The system then performs clustering-based segmentation based on the selected attributes and presents the results in both table format and an interactive dashboard. Additionally, the system provides a reporting feature to generate downloadable PDF reports containing all segmentation details and results, along with a report history for easy access to past results.

In summary, the system is designed to provide simplified, insightful segmentation tools without requiring technical expertise from SMEs businesses. It helps them to gain clearer insights into their customer base. However, the final decision-making and strategy implementation based on the segmentation insights remain the responsibility of the business users.

1.5 Project Timelines

The following Table 1.1 describes the project schedule.

Table 1.1 Project Schedule

Activities	Week													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Title														
Problem Statement														
Project Objectives														
Literature Review														
Methodology														
Monitoring														
Functional Requirement														
Analysis And Design														
Technical Implementation														
Stakeholder Meeting														
Viva														
Report														

From Table 1.1 above, it shows the main phases of the research and system development planning over a 14-week period. It was designed to ensure smooth project execution by allocating appropriate time for each phase. This schedule also helps in monitoring progress, and ensuring the tasks are completed on time.

CHAPTER 2: LITERATURE REVIEW

This chapter presents an in-depth review of existing related market segmentation works. It explores the theoretical foundations of market segmentation, which distinguishes between a priori and a posteriori methods under Section 1, and three major segmentation categories which are demographic, geographic, and behavioral under Section 2. Each criterion was analyzed in terms of definitions, commonly used variables, methods, advantages, and limitations. Furthermore, it also discusses the comparative analysis of existing segmentation systems, including Amplitude, Hotjar, Insightly, Google Analytics under Section 3. To address the gaps identified in these systems, the proposed web-based segmentation system was discussed under Section 4.

2.1 Web-Based Model of Market Segmentation

Market segmentation has long been seen as an essential element of sound marketing strategy. Historically, it has been applied to segment customers based on similar characteristics, segmentation has allowed companies to personalize promotions for specific audience segments. Recently, the advent of web platforms and technologies has provided more dynamic, and data-driven segmentation.

The core theory of segmentation is derived from the STP model which is introduced early by Smith. In this model, the segmentation involves dividing the market based on shared characteristics, targeting refers to choosing high-potential segments, and positioning refers to the process of aligning products with the expectations of the chosen segments. Segmenting is about splitting an heterogeneous market into homogeneous segments that can be targeted more cost-effectively (Dolnicar, 2022). In the era of the internet, web-based models of segmentation extend these concepts by utilizing actual data, machine learning, and interactive dashboards to automate and optimize these tasks. Market segmentation methods can be broadly classified into a priori (pre-defined) and a posteriori (data-driven) approaches. These methods will be discussed in detail in Sections 2.1.1 and 2.1.2.

2.1.1 A Priori Market Segmentation in Web Systems

A priori segmentation involves grouping customers based on predefined characteristics like location or age before doing any real market research or data analysis. It's often based on

assumptions rather than actual customer behavior. (Matt, F., 2020) A priori segmentation is simple to implement but often lacks depth and may not reflect true diversity of customer preferences. It is typically based on past knowledge or assumptions regarding customer characteristics and preferences. For instance, customers may be segmented into separate age groups such as below 18 years old, 19-25 years old, 26-35 years old, and so on. This type of rule-based segmentation allows for faster decision making but may not identify new or hidden behavior patterns. In web-based systems, a priori segmentation is usually achieved through form-based filtering or list generation from static customer data. The segmentation criteria in this approach are largely determined by the marketer's or business owner's prior knowledge and assumptions about the target market, rather than insights derived from dynamic or behavioral data.

2.1.2 A Posteriori Market Segmentation in Web Systems

On the other hand, a posteriori segmentation uses machine learning techniques like clustering to discover natural groupings, offering deeper behavioral insight. It is derived after collecting and analyzing actual customer data. This approach uses statistical methods to uncover naturally occurring or hidden patterns in the data (Matt, F., 2020). There are few common clustering techniques including K-Means, Hierarchical Clustering, and DBSCAN, which group customers based on how closely they resemble one another across multiple variables. For example, rather than pre-defined customers by age group, a posteriori segmentation can be used to combine the age attributes with purchase frequency attributes to discover deeper segment insights. Web-based systems can combine these approaches by allowing user-defined filtering (a priori - rule-based classification) and algorithmic clustering (a posteriori - clustering) to co-exist, making segmentation both actionable and personalized. For example, a tourism segmentation model grouped users based on motivations and behavioral patterns like shopping or relaxing activities (Dolnicar, 2022). Such behavioral profiles, when automated in a web system, allow businesses to instantly adapt their campaigns based on clustering results. By applying these unsupervised learning techniques, web systems can bring more adaptive, intelligent segmentation by turning raw customer data into meaningful groupings that reflect actual behavior. This gives businesses the flexibility to adjust campaigns or offerings based on the recommended segmentation result as needed.

Table 2.1 below summarizes the a priori vs posteriori market segmentation in web-based systems.

Table 2.1 Past vs Current Market Segmentation

Aspect	Past Market Segmentation	Current Market Segmentation
Approach	A Priori (Pre-defined criteria)	A Posteriori (Data-driven, post-analysis)
Data Use	Limited manual data upload or no real-time data	Manually uploaded data or heavy use of real-time and big data
Techniques	Manual segmentation	Machine learning, clustering (e.g., K-means), AI
Customization	One-size-fits-all strategy	Hyper-personalized strategies
Tools	Surveys or basic analytics by using assumption or pre-existing knowledge and understanding of the market	Web analytics, CRM, APIs, behavioral tracking
Focus	Broad segments	Micro-segments, individual personalization

Table 2.1 above provides a comparative overview between the traditional (past) and modern (current) market segmentation approaches. It shows the key differences on different aspects. This comparison helps to illustrate the evolution from basic, rule-based segmentation to advanced, data-driven and personalized segmentation models.

2.2 Segmentation Criteria & Related Models

Segmentation allows companies to know more about their customers by grouping them into different segments based on their common characteristics. There are three main segmentation categories criteria which are demographic, behavioral, and geographic. They are commonly used in research and practice. Each of these types applies different variables and modeling techniques. In this section, models, methods, and algorithms related to each base are discussed.

2.2.1 Demographic Segmentation

Demographic segmentation involves categorizing consumers based on quantifiable attributes such as age, gender, income, education level, household size, and ethnicity (Olayinka, 2021) & (Rio et al., 2024). This is the most commonly used method due to its simplicity and the accessibility of relevant data. It is often considered a commonsense (priori) approach where segments are predefined based on easily perceptible differences such as nationality or gender.

For example, one past study used "country of origin" as the segmentation criterion (Dolnicar, 2022). This approach involves selecting the tourists' based on their country of residence based on practical knowledge that tourists from different countries often have distinct preferences, behaviors, and needs. The process begins by grouping all tourists from a specific country, such as Germany, into a single segment, Malaysia into a single segment, and so on.

Another example of priori segmentation can be found was segmentation by generation (Thomas & George, 2021). It focuses on differences between Gen Y, Z, and Alpha for customizing digital content. For instance, it does the segmentation based on age or birth year. It defines people born between 1981 and 1996 as Generation Y (Millennials), those born approximately between 1997 and 2012 as Generation Z, and individuals born from around 2013 onward as Generation Alpha.

A significant problem with demographic segmentation is that it has a fixed nature. Demographic variables may not totally represent the changes in consumer attitudes and preferences through births, while typical demographic segmentation could potentially result in marketing approaches that are obsolete (Thomas & George, 2021). Also, it tends to lose insight of the individual's differences within the group and also possible behavior variables that might more accurately predict consumer behavior. Consequently, it is important to use demographic segmentation with additional characteristics to help create richer and better understanding.

Models/Methods:

- Commonsense (a priori) segmentation - (Dolnicar, 2022) and (Thomas & George, 2021)

Advantages:

- Easy to implement.
- Widely understood by marketers.
- Useful for large-scale targeting.

Disadvantages:

- Static nature as it does not reflect evolving consumer interests.
- Risk of oversimplification (not all within the same demographic behave alike)

2.2.2 Behavioral Segmentation

Behavioral segmentation is mainly focus on a posteriori segmentation by using actual customer actions like buying frequency, loyalty, and promotion responsiveness to discover segments through data analysis (Yoseph et al., 2020). However, it can also be implemented a priori, using predefined rules such as categorizing customers as “frequent buyers,” “seasonal buyers,” or “one-time buyers” based on assumed behavior. These rules were created based on business assumptions rather than data-driven discovery. While this approach is straightforward and easy to use, it lacks flexibility and may not reveal hidden customer patterns, especially as consumer behavior changes over time.

Behavioral segmentation divides individuals based on customer behavior such as attitudes, frequency of buying, brand loyalty, and user status, decision-making process, and response to marketing efforts (Yoseph et al., 2020). It is especially useful in differentiating aspects such as loyalty to a customer, purchase intention, or channel preference because it reflects actual consumer behavior rather than assumptions. This makes it a lot more dynamic and predictive. For example, if one could identify those customers who always made purchases on discount periods or those who always selected premium-level options, stronger targeting of its marketing campaigns would be possible by the business. It can also help with forecasting by looking into possible trends, behavioral patterns, and so on in groups of customers, and helps a business make better data-knowledge decisions.

For instance, the LRFM model was applied to segment dental patients based on the frequency of visitation and spending so that the clinic could provide services appropriate to the customer needs (Torkzadeh et al., 2021). It analyzes based on their Length, Recency, Frequency, and Monetary behaviors. The model categorized different patient groups; for instance, patients who visited in the recent past and with higher spending were classified as high-value patients. Using these behavioral characteristics, the clinic could customize its services by rewarding

high-frequency patients with offers or sending personalized reminders to lower-frequency patients. This is an example of the true power of behavioral segmentation in providing targeted and effective services to the health tourism industry where patient actions and patterns are the basis of service delivery.

Furthermore, Clustering algorithms such as K-means and Ward's hierarchical clustering are also widely used in behavioral segmentation due to their ability to uncover hidden patterns in large datasets (Dolnicar, 2022; Torkzadeh et al., 2021). K-means is used because of its computational efficiency for large data, and Ward's algorithm is optimal for clustering based on pairwise similarity. Based on the example shown in this research, it shows the combination uses Ward's hierarchical clustering and K-means clustering to segment dental clinic patients in Taiwan. Initially, Ward's hierarchical method was employed to analyze the data and determine the optimal number of clusters by examining dendograms and similarity measures. Once established, K-means clustering was applied to efficiently arrange patients into the predefined segments, based on variables such as behaviors related to healthcare utilization and demographics. The two-step approach can lead to accurate and meaningful segmentation by using the strengths of the two methods toward enhancing validity of the resultant patient groups. Despite its power, behavioral segmentation does require accurate, well-organized, and updated data. If the data is filled with outliers or inconsistencies, the segmentation results may be adversely affected. Besides, K-means requires the user to pre-define the number of clusters to finalize. Sometimes this may cause the results to be oversimplified or misleading, if the matter is not handled with care.

Moreover, the application of AI-driven behavioral segmentation can also be observed in smart retail environments. For instance, digital assistants that utilize clustering algorithms have been employed to engage with customers in real time while dynamically personalizing product suggestions (Moore et al., 2022). By examining consumer interests and interactions, AI applications enable the personalization of suggestions real-time, thereby facilitating more sophisticated segmentation based on recent behaviors. All such techniques, although they have excellent results, can also present challenges with regard to the privacy of customer data, particularly when personal data are used to predict and influence consumer purchasing behavior.

Thus, while behavioral segmentation on AI can assist in the creation of more focused and powerful selling spaces, it has to be approached with caution, considered ethical, and fully detrimental consumers are made aware of how their data and behavior are being carried forth.

Models/Methods:

Posteriori Segmentation:

- LRFM model (Torkzadeh et al., 2021).
- Clustering Algorithms:
 - K-means (Dolnicar(2022) and Torkzadeh(2021))
 - Ward's Hierarchical Clustering (Dolnicar(2022) and Torkzadeh(2021))
 - AI-driven Behavioral Segmentation (Moore et al. (2022))

Advantages:

- Reflects actual customer behavior.
- Facilitates targeting high-value customers.
- More deep insights compared with demographic segmentation

Disadvantages:

- Requires regular data updates and high quality data.
- Algorithms sensitive to noise and outliers.
- Clustering outcomes may vary by initialization (especially in K-means)
- Ethical or privacy concerns

2.2.3 Geographic Segmentation

Geographic segmentation divides customers according to their location attributes, such as region, city, climate or population density (Osei et al., 2021). This is a common marketing practice used in addressing regional preferences, cultural dissimilarities, climate-based demands, and logistical issues. It normally takes an a priori approach in the sense that segmentation is done on pre-defined geographic borders such as based on national boundaries, provinces, urban or rural categories. Yet, geographic segmentation may also be enriched as well under hybrid segmentation by being combined with other variables, such as demographic, or behavioral. The

combination enables a more detailed understanding of the needs of consumers, even within the same geographic area.

One of the examples that use priori segmentation can be seen in tourism and advertising where location affects preferences, language, and product availability. This form of segmentation is chosen due to its easily observable characteristics. For example, targeting German tourists with specific promotional materials due to their known travel behaviors and accessible media channels is a common strategy (Dolnicar, 2022).

One past study also demonstrated the use of geographic segmentation through an a priori approach. This method assumes that consumer preferences, behaviors, and cultural norms vary by location. The segmentation was carried out by identifying different geographic units such as specific cities or neighborhoods within the Kumasi Metropolis where banking services were being delivered. These predefined regional divisions were then used to tailor marketing and service delivery strategies to the local context. (Osei et al., 2021). Besides, the study also shows more advanced segmentation to a new level by integrating geographical information with behavioral and demographic information to develop deeper insights. This allowed marketers to tailor campaigns on local customer needs, lifestyle behaviors, and geographic patterns of behavior which make the segmentation not only location-aware but also behaviorally and demographically relevant. However, there are few limitations of this geographic approach. It assumes homogeneity within geographic areas, which may overlook diverse customer needs. It also faces challenges in collecting accurate geographical and demographic data, and static segmentation may become outdated due to dynamic changes in regional characteristics.

Another study shows that geographic segmentation was conducted using a posteriori data-driven approach (McKercher et al., 2022). Tourists' home countries were collected through surveys, and then a two-step cluster analysis was applied to categorize respondents into different geographic groups. This process allowed the researchers to reveal eight geographic segments, including groups located in Oceania, North America, some segments from Asia, and several different regions in Europe. The segments were determined based on the patterns of the residence of where respondents lived, which helped determine how geographic origin affects both the

behavioral and demographic characteristics of tourists. The analysis involves two-step clustering, which uses rule-based pre-clustering and hierarchical clustering. Initially, the algorithm automatically scans the dataset to form pre-clusters based on similarity criteria, effectively reducing complexity and grouping similar cases. These pre-clusters are then hierarchically merged to generate the final, well-defined segments. This integrated, automated process enhances efficiency, particularly for large and mixed datasets, enabling the researchers to derive meaningful geographic segments without manual rule-setting, thereby shedding light on the influence of geographic origin on tourist behavior. However, there is a limitation where the segmentation results are highly dependent on the variables chosen and the context of the destination, and using too many variables can dilute the clarity and actionability of the resulting clusters.

In addition, other past research shows that it uses a combination of RFMT scoring model and geospatial analysis methods to do geographic segmentation (Ehsani & Hosseini, 2023). The RFMT model derives customer value from behavior variables like Recency, Frequency, Monetary value, and Tenure. These are then plotted on heatmaps according to geolocation information (latitude and longitude) so that researchers can visually determine areas of high concentrations of high-spending or loyal customers. For instance, the cities with the highest RFMT scores and consumer expenditures were identified as high-priority targets for special promotional campaigns. This approach offers a two-way view which incorporates both geographic and behavioral aspects of customer data to allow organizations to merge regional campaigns, gain insights into local tastes, and optimize resource allocation. However, this method also faces several limitations, such as dynamic market challenges due to its rapidly changing consumer locations or behaviors that require frequent data updates. Next, privacy concerns may arise because of using detailed location data, potentially restricted by regulations. Moreover, relying heavily on geographic indicators might oversimplify complex consumer behaviors and motivations.

Models / Methods Used:

- A Priori Geographic Segmentation (Dolnicar, 2022)
- Hybrid Geographic Segmentation (Osei et al., 2021)

- A Posteriori Geographic Segmentation (Data-Driven) (McKercher et al., 2022)
- RFMT + Geospatial Analysis (Ehsani & Hosseini, 2023)

Advantages:

- Easy to apply with known boundaries (a priori method)
- Hybrid models provide deeper, multi-dimensional insights
- A posteriori methods offer data-driven and context-sensitive segmentation
- Heatmaps in RFMT reveal high-value geographic zones visually

Disadvantages:

- May oversimplify diverse needs within the same area
- Geographic data can become outdated due to regional shifts
- Segmentation outcome depends heavily on selected variables and context
- Too many variables used can reduce clarity and interpretability of clusters
- Privacy concerns with collecting and using geolocation data

2.2.4 Summary Table of Related Models and Their Evaluation

Table 2.2 presents a summary related models and their evaluation on market segmentation models. It outlines key segmentation categories, methods, advantages, and disadvantages used in prior studies.

Table 2.2 Related Work

Author (Year)	Segmentation Category	Segmentation Type	Model / Method / Approach	Advantages	Disadvantages
Dolnicar (2022)	Demographic	Priori	Commonsense approach	Easy to apply	Inflexible, oversimplify,
	Behavioral	Posteriori	<ul style="list-style-type: none"> - K-means - Ward's hierarchical clustering 	Uncovers hidden patterns in behavior, supports data-driven targeting	Sensitive to noise
	Geographic	Priori	Pre-knowledge or assumption	Easy to apply	Assumes regional homogeneity, data may become outdated
Thomas & George (2021)	Demographic	Priori	Generational segmentation	Simple and intuitive	Static, may not reflect individual variation or evolving behaviors

Torkzadeh et al. (2021)	Behavioral	Posteriori	<ul style="list-style-type: none"> - LRFM model - K-means - Ward's clustering 	Reflects real behavior, good for high-value customer targeting, enables personalization	Requires high-quality and up-to-date data, clustering may be sensitive to outliers and initial values
Moore et al. (2022)	Behavioral	Posteriori	AI-driven segmentation using digital assistants and real-time clustering	Dynamic, enables real-time personalization, enhances user engagement	Raises ethical/privacy concerns, needs constant monitoring and consent management
Osei et al. (2021)	Geographic	Priori	Location-based segmentation using regional boundaries	Simple and easy to implement	Assumes uniformity within areas, geographic data may become outdated
	Geographic & Demographic & Behavioral (Hybrid)	Priori	Combined with demographic and behavioral data	Enables region-specific strategies, enriched by combining other variables	Needs more data, more complex implementation

McKercher et al. (2022)	Geographic (Hybrid)	Posteriori	Two-step cluster analysis	Data-driven and automated clustering	Outcome depends on chosen variables, too many variables can reduce clarity
Ehsani & Hosseini (2023)	Geographic & Behavioral (Hybrid)	Posteriori	<ul style="list-style-type: none"> - RFMT - Geospatial heatmap analysis 	Combines behavior and location, visually identifies high-value zones	Privacy risks, data may become outdated, overly reliant on location-based segmentation

The Table 2.2 above serves as a structured overview of previous research contributions related to market segmentation. It highlights not just the segmentation types and techniques used but also reveals how each approach performs in terms of practical benefits, and known shortcomings. This side-by-side comparison helps identify the foundational strengths and weaknesses of existing models which are later used to justify the features proposed in this project.

Table 2.3 below presents the summary of related works.

Table 2.3 Related Work Summary

Author (Year)	Segmentation Category	Segmentation Type	Model / Method	Hybrid / Data-Driven	Theoretically Supported	Practical	Deeper Insight
Dolnicar (2022)	Demographic	A Priori	Commonsense approach		✓	✓	
	Behavioral	Posteriori	K-means, Ward's hierarchical clustering	✓	✓	✓	✓
	Geographic	A Priori	Assumption-based regional grouping		✓	✓	
Thomas & George (2021)	Demographic	A Priori	Generational segmentation		✓	✓	
Torkzadeh et al. (2021)	Behavioral	Posteriori	LRFM model, K-means, Ward's clustering	✓	✓	✓	✓

Moore et al. (2022)	Behavioral	Posteriori	AI-driven real-time clustering	✓	✓		✓
Osei et al. (2021)	Geographic	A Priori	Location-based segmentation		✓	✓	
	Geographic + Demographic + Behavioral	A Priori (Hybrid)	Combined segmentation variables		✓	✓	✓
McKercher et al. (2022)	Geographic	Posteriori (Hybrid)	Two-step cluster analysis	✓	✓	✓	✓
Ehsani & Hosseini (2023)	Geographic + Behavioral	Posteriori (Hybrid)	RFMT model + Geospatial heatmap	✓	✓		✓

Table 2.3 above offers a high-level comparative matrix, focusing on the capabilities of each model across key evaluation dimensions. It examines whether the approaches support hybrid or data-driven methods, offer practical applicability, and follow an established theoretical foundation. This assessment helps identify which models align best with modern business needs. These insights directly inform the architecture and innovation direction of the proposed segmentation system.

2.3 Comparative Analysis of Existing Systems

2.3.1 Amplitude

Amplitude is a digital analytics platform that provides companies with insight into user behavior on websites and applications, allowing companies to analyze event data at scale. The primary intent of Amplitude is to analyze product analytics and user journeys so that companies can improve customer engagement and retention (Amplitude, 2025c). Amplitude is used primarily in product-led organizations. Amplitude gives teams the power to independently explore customer behavior trends, identify drop-off points, and gauge the adoption of features and effectiveness of campaigns. In terms of segmentation, Amplitude allows users to segment (cohort) users by applying a taxonomy of filters, event properties, and behavioral conditions. Amplitude allows organizations to capture predictive modeling and user journey data to augment their advanced behavioral analyses.

Segmentation Focus: It primarily centers on behavioral segmentation by grouping users according to their interactions, event triggers, and engagement patterns within digital products.

Below is the details explanation of segmentation function that provided by Amplitude:

Segmentation Features:

- Enables users to create segments (cohorts) using a combination of user properties, event triggers, session metrics, and time-based behavior.

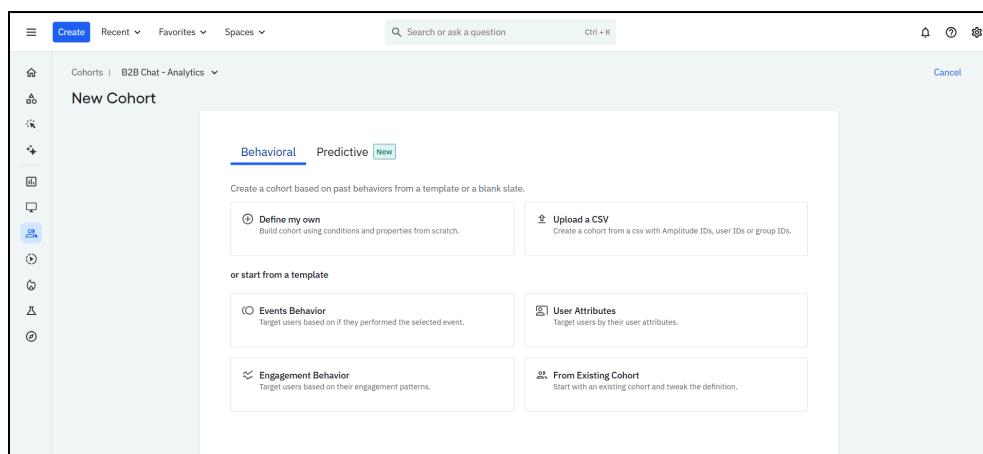


Figure 2.1 Amplitude - Cohort Creation Page

- Offers real-time segmentation and dynamic cohort updates as new user data is collected.

- Allows integration with machine learning-based predictive modeling to create data-driven segments.
- Provides a journey analysis tool to visualize user flows and behaviors before or after a specific event.

The screenshot shows the Amplitude platform's journal analysis list page. At the top, there is a navigation bar with 'Create', 'Recent', 'Favorites', and 'Spaces' buttons, and a search bar containing 'journey analysis'. On the left, a sidebar lists various product categories: Home, All Content, Live Events, Ask Amplitude, Product Analytics, Marketing Analytics, Users, Session Replay, Heatmaps, Experiment, and Guides and Surveys. The main area displays a table titled 'Search results for "journey analysis"' with 82 results. The columns in the table are Name, Last modified, Views, and Editors. The results include 'Campaign Conversion Analysis', 'Most Common Journey', 'Example Retention Analysis', 'Impact analysis', 'DAU analysis', 'Retention Analysis', 'Funnel Analysis Charts', 'My analysis', and 'YoY Analysis'. To the right of the table is a 'Filter' sidebar with sections for Editors (All editors), Space (In any space), Type (Chart, Cohort, Dashboard, Experiment, Flag, Notebook, Guide, Survey), and Content is... (Archived, Official content, Template).

Figure 2.2 Amplitude - Journal Analysis List Page

- Includes interactive dashboards for segment-based reporting, including retention, funnel, and path analyses.

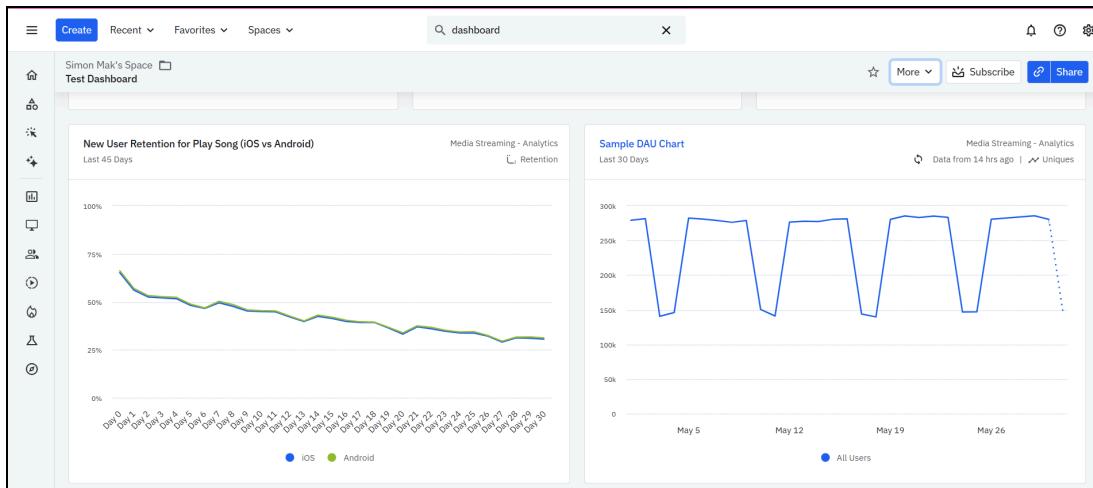


Figure 2.3 Amplitude - Dashboard Page

- Supports data export in CSV, JSON, or through API connections, and saves custom dashboard views within the platform.

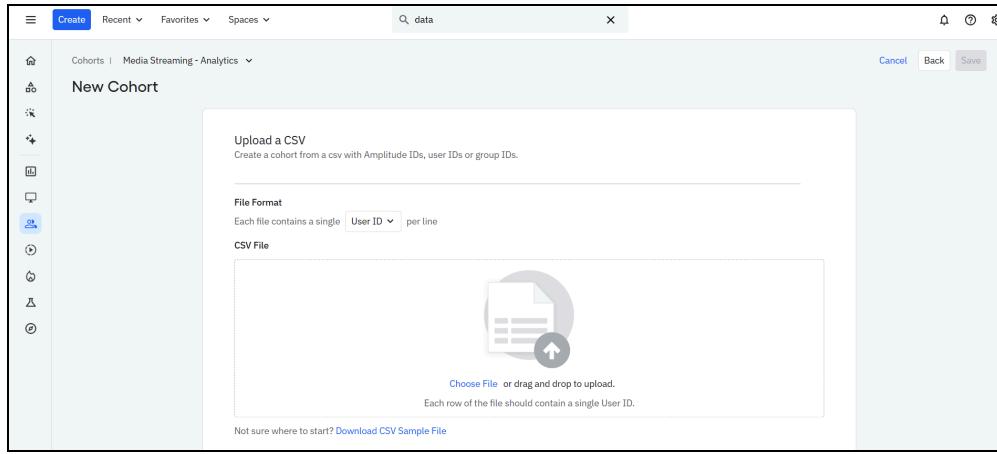


Figure 2.4 Amplitude - Upload Data Page

Constraints:

- Segmentation capabilities are highly advanced but require a steep learning curve for non-technical users unfamiliar with event-based data modeling
- Limited focus on demographic segmentation, mainly centers on behavioral and event-based metrics.
- While the dashboards are powerful, automated generation of comprehensive reports in a downloadable or printable format like PDF is limited unless integrated with external reporting tools such as Looker or Tableau.
- Requires manual effort to combine multiple segmentation criteria from different categories, for example combining geographic and behavioral dimensions
- No machine learning-based clustering algorithms for automatic segment clustering but it was done using rule-based if/then logic to segment the customer groups. Segments are predefined manually, which means it lacks a posteriori, data-driven clustering.
- It does not have history to save the PDF report. Users may not be able to view past segmentation reports for comparison with current results.

Main Gap:

Amplitude provides strong data-driven segmentation using behavioral cohorts and predictive analytics and therefore is in a good position for user engagement and product optimization.

However, Amplitude suffers from its strong complexity which makes it comparatively less feasible for SMEs or technical users who could struggle to implement an event tracking plan and best interpret the events inputted. Amplitude also does not provide the built-in feature for generating comprehensive, printable segmentation reports for easy reference or sharing. It also does not have clustering abilities and does not allow a posteriori segmentation formed using machine-learning algorithms. Instead, segmentation features rely on pre-defined behavioral filters and cohorts.

2.3.2 Hotjar

Hotjar is a user behavior analytics tool that provides qualitative insights into how users interact with websites through features such as heatmaps, session recordings, surveys, and feedback widgets (Hotjar, 2025). It is widely adopted by marketing and UX teams to complement quantitative analytics by uncovering the *why* behind user actions. Hotjar offers basic segmentation through behavior filters and user attributes, mainly for filtering recorded sessions and survey results.

Segmentation Focus: It mainly focuses on behavioral segmentation by analyzing user interactions to identify patterns and segment users based on their on-site behavior.

Below is the details explanation of segmentation function that provided by Hotjar:

Segmentation Features:

- Allows filtering of session recordings and heatmaps based on user attributes (e.g., country, device type, visited URL) and behavioral triggers (e.g., rage clicks, scroll depth, time on page).

The screenshot shows the 'Session Replay' section of the Hotjar interface. On the left, there's a sidebar with navigation links: Get started (0/4), Dashboards, Impact, Funnels, Session Replay (selected), Heatmaps, Surveys, Interviews, Tests, and Collapse. The main area is titled 'Session Replay' and shows a table of 6 results. The columns include Action (with a 'Play' button), Duration, Date, Score, Pageviews, Geolocation, Browser, OS, and Screen size. The data from the table is as follows:

Action	Duration	Date	Score	Pageviews	Geolocation	Browser	OS	Screen size
<button>Play</button>	0:00:24	May 31, 2025 at 7:18 AM	Ar 33	1	IRL	Chrome	Linux	2000 x 2000
<button>Play</button>	0:22:53	May 29, 2025 at 10:07 AM	Ar 41	9	ESP	Chrome	iOS	1512 x 982
<button>Play</button>	0:00:04	May 21, 2025 at 9:59 PM	Ar 33	1	COL	Chrome	Windows	1366 x 768
<button>Play</button>	0:00:06	May 17, 2025 at 7:07 AM	Ar 33	1	DEU	Chrome	Windows	5120 x 1440
<button>Play</button>	0:03:00	May 15, 2025 at 7:48 PM	Ar 20	7	USA	Chrome	Windows	1920 x 1080

Figure 2.5 Hotjar - Session Replay Page

- Supports real time data flow where it automatically collects user behavior data from your website such as mouse movements, clicks, scrolls, and responses to on-site surveys.
- Supports feedback segmentation by applying filters to user responses based on behavior and metadata such as referral source, browser, or operating system.

The screenshot shows the 'Surveys' section of the Hotjar interface. The sidebar includes links for Get started (0/4), Dashboards, Impact, Funnels, Session Replay, Heatmaps, Surveys (selected), Interviews, and Tests. The main area is titled 'Surveys' and features a 'GUIDES: NEW SURVEY OVERVIEW' section with a 'Take quick action with real-time survey insights' callout. Below this is a search bar labeled 'Filter by Survey name or creator'. A table lists surveys with columns for Status, Name, Responses, Created by, Date created, Type, and Actions. One survey is listed: 'Feedback.button' (Status: Active, Responses: 8, Created by: Anonymous, Date created: 30th October 2024, Type: Button). At the bottom, there's a 'Explore popular templates' section with the text 'Set up surveys in minutes and ask all the right questions'.

Figure 2.6 Hotjar - Survey Page

- Provides survey logic and targeting options to display questions to specific user segments, such as users from a specific location or those who visited a certain number of pages.
- Provides basic dashboards that display user behavior insights through tools like heatmaps, recordings, and surveys

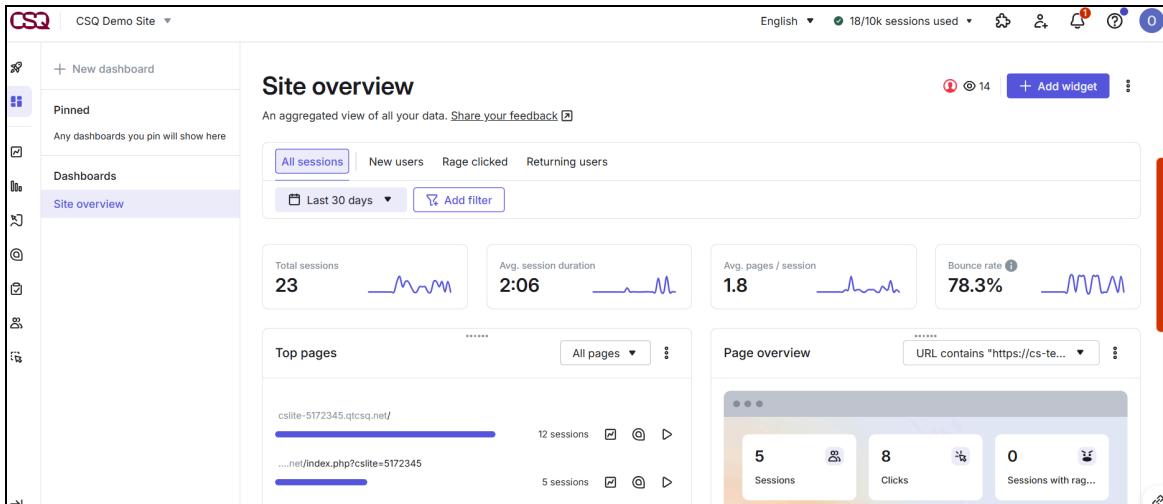


Figure 2.7 Hotjar - Dashboard Page

- Offers the ability to create “User Segments” to revisit specific visitor types based on predefined or custom behavioral conditions by using a filtering function.

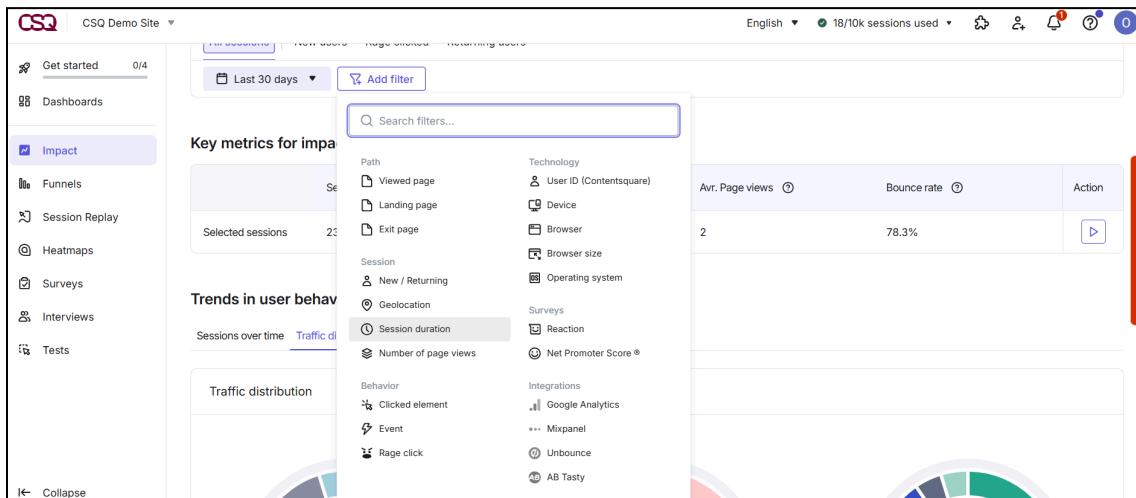


Figure 2.8 Hotjar - Filter Function Page

- Integration with external tools like HubSpot, this allows syncing user data for enhanced filtering.

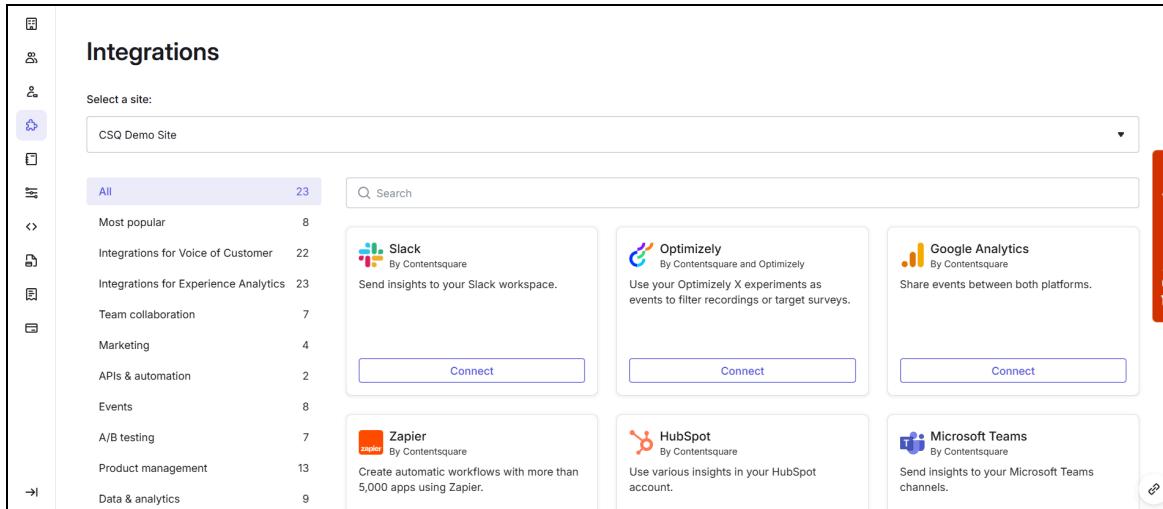


Figure 2.9 Hotjar - Integration Page

Constraints:

- Does not allow user to upload customer data
- Does not offer advanced dashboards for comparing customer segments or visualizing clustering results
- No machine learning or clustering-based segmentation, segment creation is rule-based and limited to predefined filters.
- Focuses mainly on behavioral data derived from web interactions, with little to no support for demographic, or demographic segmentation.
- Segments cannot be visualized in cluster-based dashboards or statistical summaries; insights are mostly qualitative and observational.
- Reporting is manual where users can export individual heatmaps or survey results, but Hotjar does not compile full segmentation reports in PDF or any unified format.
- Lacks cohort analysis and temporal comparison features that show how segments evolve over time.

Main Gap:

While Hotjar does excellently provide qualitative, user-focused insights from visual and behavioral data, it is limited in its automated segmentation. Users need to rely upon manual

filters. Additionally, Hotjar has fewer integrated user attributes to aggregate to find hidden patterns or conduct a posteriori segmentations. Although they have many great features, all so broad and powerful, they can create overwhelming complexities for non-technical users and SMEs that only need more simplistic segmentation tools. Furthermore, it may not be as useful for users engaging with larger, more complex market segmentation workflows given there is no integrated report generation, as structured reporting and multidimensional analyses can be quite beneficial.

2.3.3 Google Analytics

Google Analytics tracks user activity on web and mobile platforms (Google Analytics, 2025). Segmentation in these tools is mainly relying on predefined user attributes and behavioral data collected via user interactions. In Google Analytics, users create segments based on attributes such as geographic location, device type, source of user traffic, and user behavior metrics. Users can create a customized segment by including combinations of those attributes and logical conditions, for example users within a predetermined geographic area who completed a purchase during the logged session. These segments help companies understand user preference and identify patterns in site engagement and conversion. It updates segments dynamically whenever new user data flows in, supporting real-time or near-real-time analysis and ad delivery. However, segmentation is limited to the attributes Google tracks or that users upload in supported formats, and there are no automated clustering or advanced unsupervised segmentation methods available within these tools. Additionally, its interface and extensive features can be complex and challenging for SMEs or users without technical expertise.

Segmentation Focus: It mainly focuses on behavioral, by tracking how users interact with websites or apps such as pages visited, session duration, conversion actions. It also incorporates geographic segmentation by identifying users' locations based on IP addresses or device settings, and demographic segmentation through inferred or collected user data like age, gender, and interests.

Below are the details explanation of the segmentation features provided by Google Analytics:

Segmentation Features:

- Allows segmentation by predefined attributes such as device type, geographic location, traffic source, and user behavior metrics like session duration and conversion events.
- Supports custom segment creation within Google Analytics dashboards using available dimensions and metrics(Rule-based segmentation).

The screenshot shows the 'Segments' section in Google Analytics. On the left, there's a sidebar with navigation links like Admin, My preferences, Setup Assistant, Account settings, Property settings, Data display (Events, Key events, Audiences, Annotations, Comparisons), and a gear icon. The main area is titled 'Segments' and contains a table with the following data:

Segment name	Description	Created by	Last updated	Conditions
Untitled segment		ONG ZI YING CELINE	May 30, 2025	Age in "18-24" AND City = 1004133
All Users	Includes all your data.			
Direct traffic	Sessions acquired directly.			Session default channel group = Direct
Email, SMS & push notifications traffic	Sessions acquired via emails, SMS or push notifications.			Session default channel group in "Email, SMS, Mobile Push Notifications"
Mobile traffic	Traffic on mobile phones.			Device category = mobile
Organic traffic	Sessions acquired via organic channels.			Session default channel group in "Organic Search, Organic Video, Organic Social, Organic Shopping"
Paid traffic	Sessions acquired via paid channels.			Session default channel group in "Paid Shopping, Paid Search, Paid Social, Paid Other, Paid Video, Display, Cross-network..."
Referral & affiliates traffic	Sessions acquired via			Session default channel group in "Referral, Affiliate"

Figure 2.10 Google Analytics - Create Segment Page

- Provides interactive dashboards and reports for audience and campaign performance visualization.

The screenshot shows the 'Reports snapshot' dashboard in Google Analytics. The left sidebar includes links for Reports snapshot, Realtime overview, Realtime pages, Business objectives (Generate leads, Drive sales, Understand web and/or app t...), View user engagement & rete..., User (User attributes, Tech), and Library. The main area features a 'Reports snapshot' card with summary data:

Add to carts	Purchases	Purchase revenue	Active users
0	0	RMO.00	0

Below this is a 'Best sellers' table:

ITEM NAME	ITEMS ADDED TO CART	ITEMS PURCHASED	ITEM REVENUE
No data available			

Figure 2.11 Google Analytics - Dashboard Page (Google Analytics)

- Allows exporting of reports in PDF or CSV formats, though report exports are modular rather than comprehensive.

Constraints:

- Lacks clustering algorithms or machine learning models within the platforms to perform unsupervised segmentation or pattern discovery.
- Report management is basic, with no built-in history or centralized storage for past exported reports.
- Allows limited data import to augment existing tracked data, but it does not support easy CSV uploads of customer profiles for segmentation, making it complex and less suitable for SMEs.

Main Gap:

While Google Analytics excel at behavioral tracking and campaign-oriented segmentation, they lack advanced segmentation logic, such as algorithmic clustering to automatically discover hidden patterns in the data. It also does not provide the segmentation result review feature. These limitations restrict their use for comprehensive customer segmentation and strategic market analysis beyond campaign metrics. For SMEs, this creates a barrier to them since they often lack dedicated data teams or technical expertise, as they are difficult to uncover meaningful customer groups or derive deeper insights that could guide long-term marketing strategies and business growth.

2.3.4 Insightly

Insightly is a CRM and project management tool mainly designed for small to medium-sized enterprises (Insightly, 2025). It provides features such as managing contacts, keeping track and report on sales pipelines, automating workflow, and managing tasks. Insightly facilitates relationship management with leads, customers, and partners by bringing together customer information and touchpoints in one place. Insightly is mainly a CRM tool, but it does have some basic segmentation options that allow users to group and sort contact data based on designated criteria.

Segmentation Focus: Insightly mainly focuses on demographic segmentation, allowing users to group contacts based on attributes like age, gender, location, job title, and company-related demographics.

Below are the details explanation of the segmentation features provided by Insightly:

Segmentation Features:

- Allows users to create saved filters and custom contact lists based on standard or custom fields, such as lead source, location, or activity status.

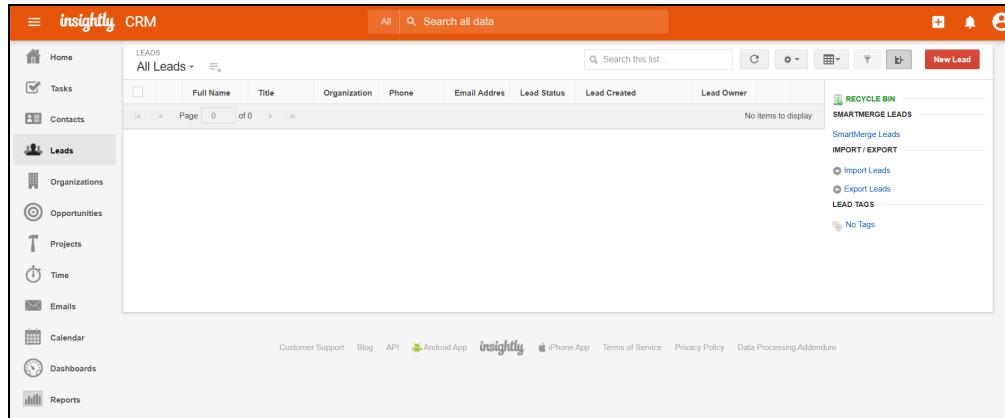


Figure 2.12 Insightly - Lead Page

- Offers tagging functionality to label contacts or organizations for basic group identification.
- Includes workflow automation based on segment-like conditions such as auto-assigning leads from a specific industry.
- Enables users to view filtered segments in dashboards with customizable charts and pipelines.

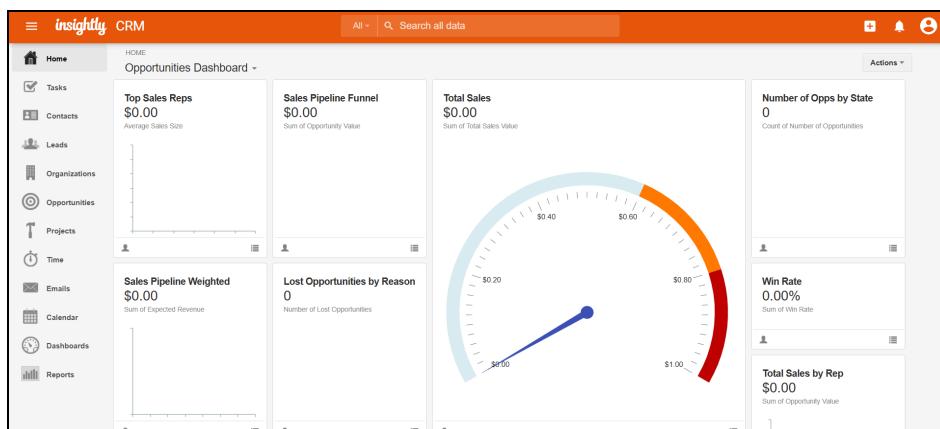


Figure 2.13 Insightly - Dashboard Page

Constraints:

- Segmentation is mostly manual and rule-based, using predefined filters and tags. There is no support for machine learning algorithms or automated clustering techniques.
- Reporting is functional but not tailored for segmentation analysis, and Insightly dashboards focus more on sales pipeline and CRM metrics than market segmentation views.
- No support for exporting segmented reports in formats like comprehensive PDFs that compile user-created segments over time.

Main Gap:

Insightly's segmentation features are limited to simple filtering, tagging, and automation of simple workflows. The platform does not have key advanced or automated segmentation features, such as clustering and multi-dimensional analysis. Though somewhat useful for CRM users who want to organize contacts or automate simple actions, Insightly does not provide options for a posteriori segmentation or deep customer insights using machine learning techniques. It is important to note that Insightly's reporting tools do not summarize all segmentation data into one report to make longitudinal decisions, or to support you in your activities. For SMEs this limitation means they cannot derive a more insightful analysis from their customer data, and it makes it more difficult to uncover meaningful patterns or trends that might be useful when constructing a targeted marketing and business strategy.

2.4 Proposed System

The proposed system aims to overcome the shortcomings of current market segmentation systems faced by SMEs by providing a simplified, automated and a user-friendly market segmentation system. It integrates both a posteriori segmentation through machine learning algorithm and flexible criteria selection to support a more deeper and accurate customer segmentation insights without relying solely on priori segmentation. Unlike current systems that mostly focus on single segmentation type. This solution introduces a hybrid segmentation function that allows users to combine demographic, behavioral and geographical in one workflow. SMEs users are allowed to choose relevant criteria from any of three segmentation

categories based on the attributes in their dataset. To accommodate the common SMEs manner, where data is often collected manually or through the export from the E-commerce website, the proposed system also includes a Data Importation module that supports .CSV file uploads. This ensures that SMEs are able to perform segmentation without needing real-time data and pipelines. SMEs users are also able to view the segmentation results in both table and dashboard format. The Report History Management feature also will be included to ensure SME users can view the past segmentation results and easily download for comparison or decision tracking over time. Table 2.4 below will explain more details on the key functionalities that will be included in the proposed system.

Table 2.4 Proposed Web-Based Market Segmentation System

Feature	Description
Data Import Function	Allows users to import CSV files of customers and orders data into the system.
Automated Data Cleaning	Uploaded data will be undergo a cleaning process to ensure it is suitable for accurate segmentation analysis which including process of handle missing values, remove duplicates, normalize or standardize values, and format columns correctly
Segmentation Criteria Selector	System will show the list of attributes and let users select which criteria to use based on their business objectives, which can promote personalized segmentation workflows.
Posteriori Segmentation	System will perform customer segmentation using machine learning algorithms such as K-means, hierarchical clustering or other suitable algorithms based on user selected segmentation criteria.
Interactive Visualization Dashboard	Provides an intuitive dashboard to compare cluster characteristics, visualize patterns, and allow users to interpret results quickly.

Automated Reports	All outputs, insights, and recommendations are compiled into a downloadable PDF report for business references and presentations.
Report History Management	This feature allows users to access and view previously generated PDF segmentation reports. Each time a segmentation is completed and a report is generated, it is saved in the system. Users can revisit these historical reports to compare current and past segmentation outcomes

System Purpose: Recommendation, Not Decision-Making

The system does not take on the role of final decision-making for the user but acts as a support tool in the decision-making process. It suggests the best customer segment and shows the evaluation result for each segment. This helps users to have a better understanding of their current customer base. However, the final decision making in personalizing marketing strategy for those segments will be left to the user to decide based on their business goals.

2.4.1 Improvements to the Recommended Framework with respect to Existing Systems

Table 2.5 below outlines the major limitations found in current market segmentation tools and explains how the proposed web-based system addresses each one.

Table 2.5 Improvement of Proposed System Compared to Existing Systems

Limitation in Current Systems	How Proposed System Solves It
Features too complex and mostly include additional advanced function beyond segmentation	Provides a more user-friendly and support for segmentation features only to avoid unnecessary complexity
Mostly are focusing on rule-based segmentation or the segments is created manually	Enables posteriori segmentation by applying machine learning for more flexible, data-driven segment discovery. Segments are auto-generated using machine learning.
No comprehensive report that stores all segmentation results	Automatically compiles segmentation results, including selected attributes, segment outputs, and dashboards, into a downloadable PDF report.
No report history or review function	Offers a report history feature where users can access and view previously generated reports through a dedicated interface.

Table 2.5 outlines the key limitations identified in existing market segmentation systems and summarizes how the proposed web-based system addresses each of these shortcomings. The proposed system enhances usability by simplifying the user interface, enhances intelligence by automating market segmentation using machine learning, and aims to remove technical expertise for SMEs. The suggested solution is built to address and fill practical gaps identified with current solutions.

The following Table 2.6 presents a comparative overview of existing market segmentation systems against the proposed system.

Table 2.6 Comparison Between Different Existing Market Segmentation Systems

Existing system/ Features	Amplitude	Hotjar	Google Analytics	Insightly	Proposed System
Data Import Function	/		/	/	/
Automated Data Cleaning					/
Segmentation Criteria Selector	/	/	/	/	/
Posteriori Segmentation					/
Interactive Visualization Dashboard	/	/	/	/	/
Exportable Segmentation Report with Insights		/ (Partially)	/ (Partially)	/ (Partially)	/
Report History Management					/

The Table 2.6 above provides a feature-by-feature comparison between the proposed web-based segmentation system and several existing systems such as Amplitude, Hotjar, Google Analytics, and Insightly. It illustrates which features are supported by current systems and identifies those introduced or enhanced by the proposed system. The table helps demonstrate how the proposed system fills functionality gaps and integrates advanced segmentation capabilities that are either missing or underdeveloped in existing tools.

CHAPTER 3: RESEARCH METHODOLOGY

This chapter presents the proposed research methodology used to guide the development of the system. It also discusses three key approaches used in this study for data collection and analysis which help in determining the design of the proposed system.

3.1 Agile Methodology

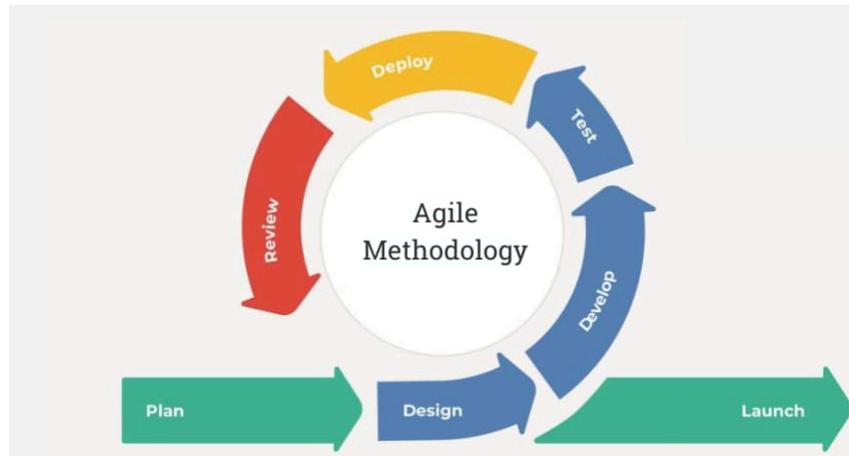


Figure 3.1 Agile Methodology

(Source:<https://interqualitybg.com/en/resources/scrum-and-agile-resources/agile-methodology>)

In this case, the Software Development Life Cycle (SDLC) methodology selected is the Agile methodology. The reason for selecting agile methodology is due to its flexibility, continuous delivery and improvement, and iterative progress. This methodology is best suited for developing the system as it is expected to evolve from feedback from the collaborator and stakeholder. Moreover, the system has multiple components including frontend interfaces, backend logic, machine learning models, and user inputs through CSV data. Agile methodology allows for each individual module or component to be developed, tested, and iterated in smaller sprints. This method allows to identify the problem in advance, regular validation along the way, and supports quick reactions based on new knowledge or insights as the development progresses.

The agile method is an iterative project management strategy that breaks a project into several open ended phases called sprints. Each sprint involves reviewing the current state of the work and identifying things to strengthen in order to continually improve in the next sprint. There are

6 different phases in a sprint: Plan, Design, Develop, Test, Deploy, Review, before launching. The following section describes each phase in detail.

3.1.1 Development Phases Overview (Agile-Based)

This section outlines the overall development workflow, on each six Agile phases, describing what activities occur at each stage of the system development life cycle.

3.1.1.1 Planning Phase

This phase will involve meeting with collaborators and stakeholders to get the specific needs from their perspectives and understanding the gaps of the current system.

1. Understanding the background, problem statement, doing the research regarding the current web-based model of market segmentation and literature review to identify the current gaps in this field.
2. Schedule meetings with collaborators and stakeholders to get the suggestions or ideas on what specific functions or needs are required for the improved version system.
3. Identify the project objectives, scope and timelines based on the research result/opinions from collaborators on the previous phase.
4. Create a questionnaire for data collection from stakeholders.
5. Define the functional requirements(FR) and non-functional requirements(Non-FR) of the system based on the data collection and feedback from both the collaborator and stakeholder.

3.1.1.2 Designing Phase

This phase will design the interface of the system.

1. Define the system architecture, how the flow of each component will interact between frontend, backend and machine learning components.
2. Design the use-case diagram, functional decomposition diagram(FDD), entity relationship diagram (ERD), and activity diagram based on the functional requirements.
3. Design the interface using Figma to ensure the interface is user-friendly.

3.1.1.3 Developing Phase

This phase will involve the coding parts which include the frontend and backend part.

1. Developing user-friendly frontend based on the designing interface on the previous phase to ensure the smooth interaction.
2. Developing the backend functionality that supports the database management and machine learning process for automated segmentation..

3.1.1.4 Testing Phase

This phase will test the developed system to ensure all the modules are worked. It is happening during every sprint, not just the end.

1. On each sprint, unit testing is conducted to test each module. If any bugs are found it can fix it to ensure the unit functions well.
2. Integration testing is conducted after the unit testing is performed where it groups all the units of a program into a group to test how each module's functions can work together smoothly. If bugs are found, fix it.
3. Lastly is user acceptance testing (UAT) where the program will let stakeholders test out the system to ensure it meets the requirements as determined by them.
4. Lastly will incorporate the feedback from supervisor, collaborator and stakeholders during sprint reviews to verify that the system remains aligned with expectations.

3.1.1.5 Deploying Phase

This phase will involve the end user to test the ready implemented system.

1. Basic training will be provided to guide them how to use the system.
 - a. Guide the user on how to upload a CSV file.
 - b. Demonstrate how to select segmentation criteria and interpret the dashboard results.
 - c. Explain how to look at the result shown and download the PDF report.
2. Monitor the deployed system for any performance issues or user difficulties.

3.1.1.6 Review Phase

This phase is called Retrospective or sprint reviews. The completed features will be present to the collaborator and stakeholder.

1. Show what was done, what was incomplete and what feedback they have, how the next sprint can be improved.
2. Collect feedback and suggestions from supervisor and collaborator.
3. Reflect on what went well and what can be improved in the next sprint.
4. Identify action items and improvements for the upcoming sprint based on feedback.
5. Ensure the development stays aligned with user expectations and project objectives.

3.1.2 Sprint Breakdown Overview

Sprint	1		2		3		4		5		6		7	
	Week													
Activities	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Title														
Problem Statement														
Project Objectives														
Literature Review														
Methodology														
Monitoring														
Functional Requirement														
Analysis And Design														
Technical Implementation														
Collaborator Meeting														
Viva														
Report														
Plan	Define scope, brainstorm problem and objectives	Re-defined objectives, doing research (LR)	Continue doing research and define methodology, Find collaborator	Design Questionnaire & Prepare monitoring slide	Define FR & non-FR. Survey collection	Re-defined system design	Wrap-up and fix bug							
Design	Draft problem statement and objective map	Draft objective and research result	Draft research result, and methodology, email collaborator	Draft Questionnaire & Monitoring Presentation Slide	Draft System design & collect survey respond	UI mockup	Improve frontend design							
Develop	N/A	N/A	N/A	Send Questionnaire & Complete Monitoring Presentation Slide	Design UI	Implement coding frontend and backend	Fix bug							
Test	N/A	N/A	N/A	N/A	N/A	Manual test prototype and system	Test the system							
Deploy	Supervisor check-in	Supervisor check-in	Supervisor check-in & collaborator and stakeholder meeting	Supervisor check-in & collaborator and stakeholder meeting	Supervisor check-in & collaborator and stakeholder meeting. Present to stakeholder and supervisor.	Supervisor check-in & collaborator and stakeholder meeting. Present to stakeholder and supervisor.	Supervisor check-in & collaborator and stakeholder meeting. Present to stakeholder and supervisor.							
Review	Need to refine the objective	Approved objective and need to improve read more research paper. Find collaborator	Collaborator approved collaboration. Research result approved.	Approved Monitoring Presentation Slide. Continue sending questionnaire	Continue improved system design	Approved	Approved							

Figure 3.2 Agile Planning

The figure above shows outline and documentation milestones. The details of the activities conducted on each sprint phase are described in the schedule. Each sprint includes the phases of Plan, Design, Develop, Test, Deploy, and Review, which provide for ongoing and iterative development and feedback from the collaborator and stakeholder. This sprint plan provides structure to incremental development and allows regular adjustments to accommodate changes in the way project needs evolve.

Sprint 1: Project Planning and Initial Problem Framing

- Defined the project scope and brainstormed the initial problem and objectives.
- Drafted the problem statement and created an objective mapping.
- Conducted supervisor check-in to confirm direction.
- Identified the need for objective refinement based on initial feedback.

Sprint 2: Objective Refinement and Early Research

- Re-defined project objectives and began literature review to explore existing systems and gaps.
- Drafted research aims and early findings.
- Continued supervisor check-ins to align progress.
- Recognized the need to expand research depth and find a suitable collaborator.

Sprint 3: Methodology Definition and Collaboration Setup

- Continued literature review and finalized the research methodology.
- Reached out to and confirmed collaboration with a collaborator.
- Drafted methodology section and emailed it for feedback.
- Supervisor and collaborator approved the revised objectives and research foundation.

Sprint 4: Questionnaire Design and Monitoring Preparation

- Designed a questionnaire and prepared monitoring presentation slides.
- Sent out the questionnaire and completed the monitoring presentation.
- Held meetings with both collaborator and stakeholder to validate direction.
- Monitoring slide was approved for presentation.

Sprint 5: Requirement Definition and System Design

- Defined functional and non-functional requirements based on survey responses.
- Drafted system design and collected additional survey responses.
- Began initial UI design planning.
- Stakeholder and collaborator meetings continued for feedback and validation.

Sprint 6: System Implementation and Testing

- Created UI mockups and implemented frontend and backend components.
- Conducted manual testing of prototype modules.
- Refined system design based on test results and feedback.
- System was presented to both supervisor and stakeholders.

Sprint 7: Final Refinement and Wrap-up

- Improved frontend design and fixed identified bugs.
- Conducted system testing to ensure functionality.
- Held final check-ins and sprint reviews with stakeholders and supervisor.
- All components and feedback were finalized and approved.

3.2 Data Collection and Analysis

To put in place a thorough theoretical and practical framework for the envisioned web-based market segmentation system, three main methodologies were used: survey, meeting with the collaborator and stakeholder, and a detailed review of past research papers. These methodologies were selected to obtain deep insights about the market segmentation systems existing in place today, along with the constraints faced and real expectations from the stakeholder.

3.2.1 Survey

A survey was designed and distributed to gather insights from SME business practitioners who operate in customer-facing sectors. The primary objective was to collect feedback regarding the challenges and expectations surrounding existing segmentation tools, and to validate the practicality of the proposed system's functionalities. The survey was created using Google Forms, and the link was distributed through email to relevant respondents. Participants were informed about the purpose of the research, the scope of the proposed system, and the expected completion time (approximately 5–8 minutes). Confidentiality and academic use of the data were clearly stated to ensure voluntary and informed participation. The data gathered from the survey was later analyzed and used to inform the definition of functional and non-functional requirements, as well as the refinement of the proposed system features. The Google Form was opened for 2 weeks and has collected a total of 29 responses.

3.2.1.1 Demographic Information

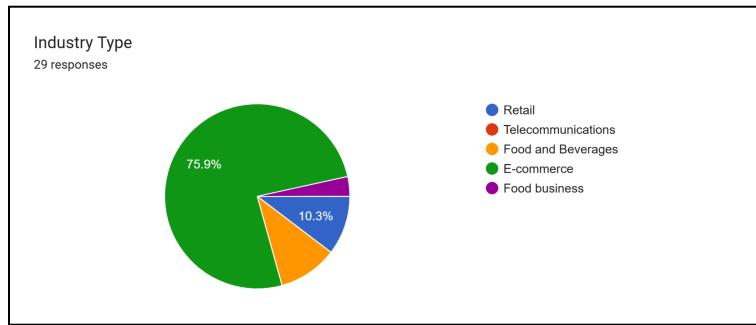


Figure 3.3 Survey - Industry Type

Figure 3.3 shows the industry type distribution among respondents. 75.9% of the respondents, majority are from the E-commerce sector. Other respondents include those from Retail(10.3%) and Food and Beverages (10.3%) and Food Business sectors.

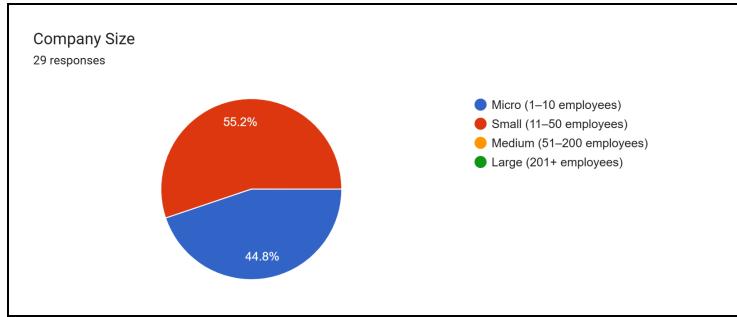


Figure 3.4 Survey - Company Size

Figure 3.4 shows the company size distribution among respondents. A majority of 55.2% are from small enterprises (11–50 employees), while the remaining 44.8% are from micro enterprises (1–10 employees).

3.2.1.2 Knowledge of Market Segmentation System

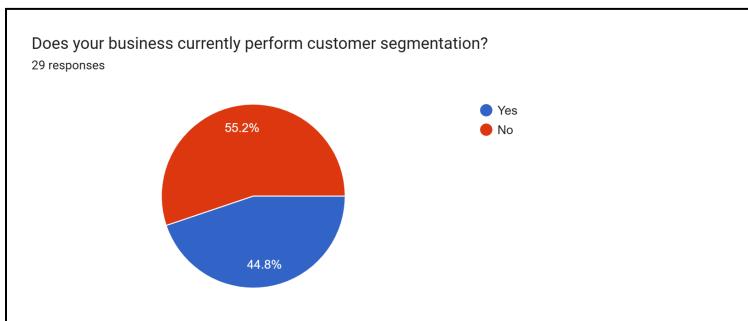


Figure 3.5 Survey - Customer segmentation practices among respondents

Figure 3.5 shows the responses to whether businesses currently perform customer segmentation. Out of 29 respondents, 55.2% indicated their business does not currently perform customer segmentation, while 44.8% indicated that they do.

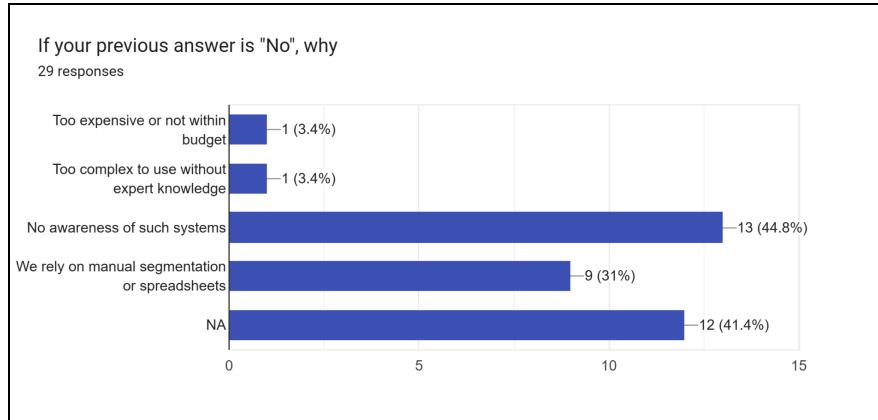


Figure 3.6 Survey - Reasons Why Businesses Do Not Perform Customer Segmentation

Figure 3.6 shows the reasons SME businesses gave for not conducting customer segmentation. The most common reason was lack of awareness of such systems (44.8%), followed by reliance on manual methods or spreadsheets (31%). A smaller number of respondents indicated issues such as high cost (3.4%) and complexity without expert knowledge (3.4%). Meanwhile, 41.4% selected NA, which may include those who answered "Yes" previously or did not specify a reason.

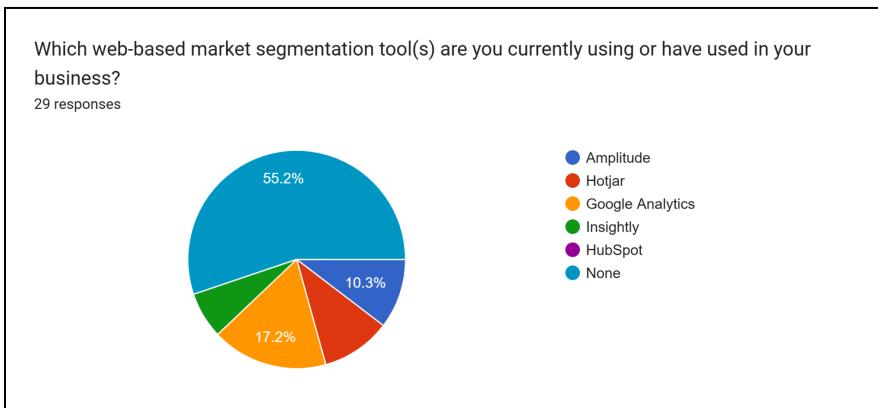


Figure 3.7 Survey - Web-Based Segmentation Tools Used by Businesses

Figure 3.7 above shows the pie chart that illustrates the responses of 29 participants regarding their use of web-based market segmentation tools. A majority of respondents (55.2%) not using any such tools in their business. Among the tools that were used, Google Analytics was the most common (17.2%), followed by Amplitude and Hotjar, each used by 10.3% of respondents. Insightly was used by 6.9%, while HubSpot was not selected by any respondent.

This suggests that while some businesses are adopting digital tools for market segmentation, a significant number still do not utilize any, indicating a potential gap in digital marketing practices among SMEs.

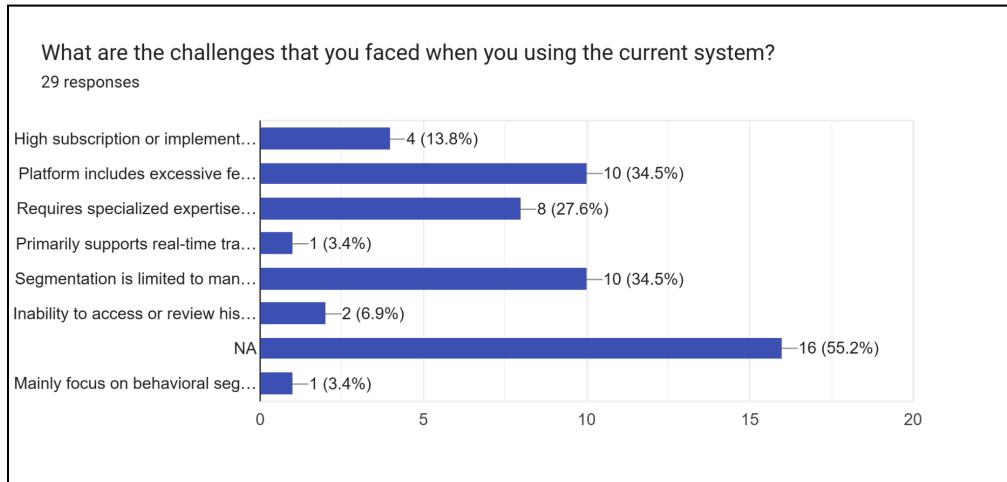


Figure 3.8 Survey - Challenges Faced When Using Current System

Figure 3.8 shows the challenges faced by respondents when using existing market segmentation systems. The most common issue, selected by 34.5% of respondents, is that the platform includes excessive features unrelated to segmentation, making it difficult to use for SMEs with limited technical resources. Similarly, 34.5% indicated that segmentation is limited to manual rules and filters, which focus more on a priori segmentation. Another 27.6% responded that the systems need specialized expertise to operate effectively, which may not be feasible for SMEs without data professionals. 13.8% cited high subscription or implementation costs as a concern. A few respondents (6.9%) mentioned the inability to access or review historical segmentation results, while others (3.4%) pointed out that current tools mainly focus on behavioral segmentation or mainly support real-time tracking.

3.2.1.3 Ideas on Improved Version Web-based Market Segmentation Model

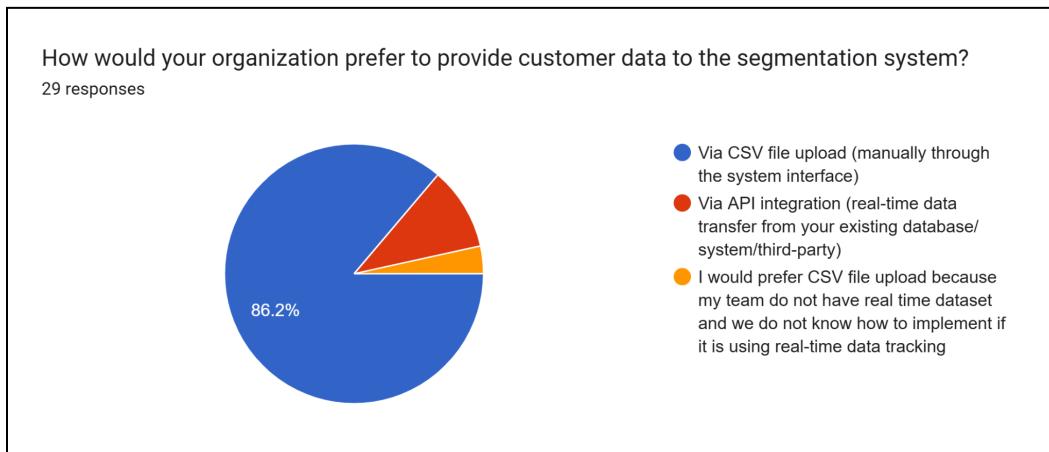


Figure 3.9 Survey - Preferred Method for Providing Customer Data

Figure 3.9 shows how SMEs prefer to supply customer data to the segmentation system. Majority of respondents ($86.2\% + 3.4\% = 89.6\%$) prefer uploading customer data manually via CSV files through the system interface. Only a few respondents (10.3%) expressed preference for API integration, probably because their organizations do have real-time datasets. This finding supports the decision to prioritize manual CSV upload functionality in the proposed system design to better accommodate SMEs' actual practices and limitations.

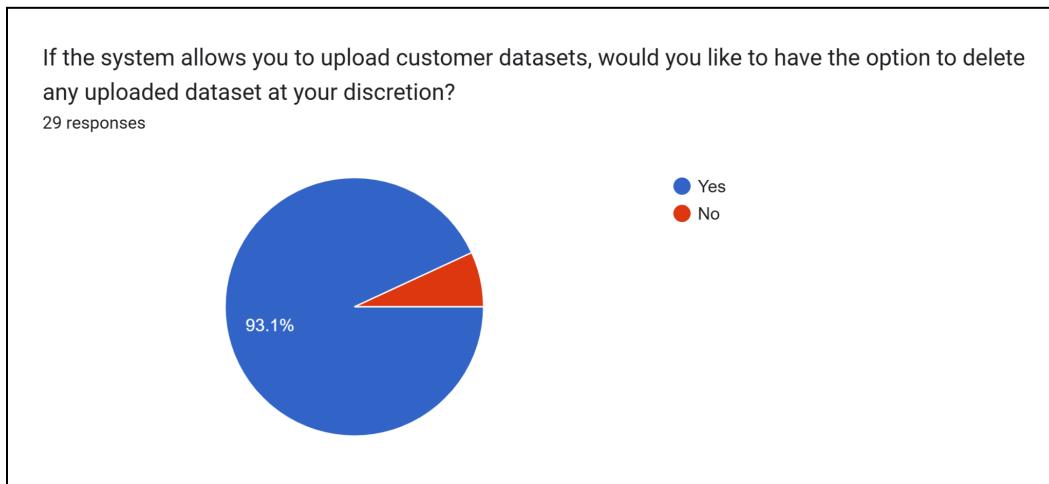


Figure 3.10 Survey - Dataset Deletion Preference

Figure 3.10 shows respondents' preferences regarding dataset management. Majority of 93.1% prefer to delete uploaded customer datasets at their discretion. This reflects a strong need for control over data management, particularly in cases where datasets may be outdated, incorrect, or

no longer relevant. Only 6.9% of respondents did not consider this function necessary. This feedback pointed out the importance of implementing a flexible data management module in the proposed system.

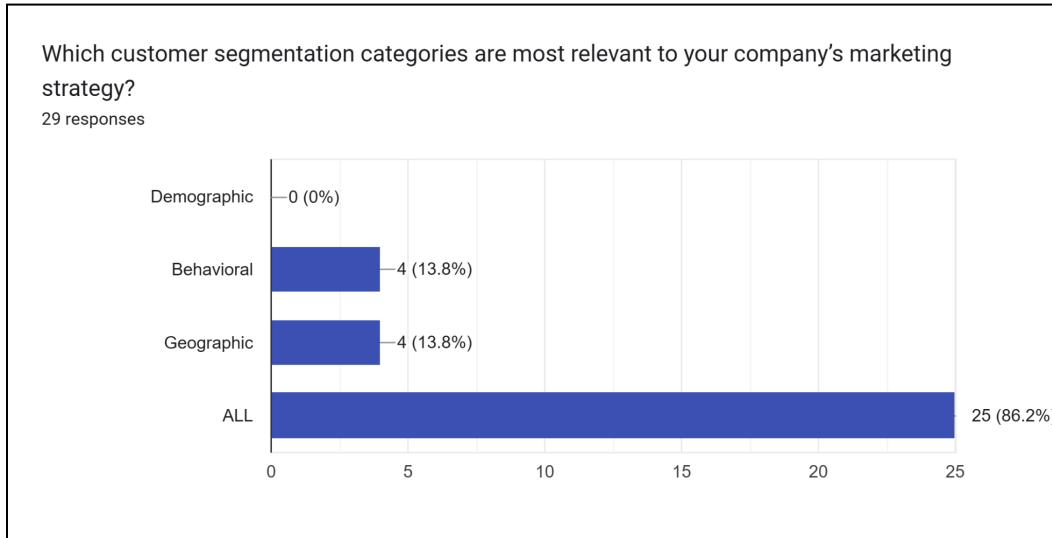


Figure 3.11 Survey - Relevant Segmentation Categories

Figure 3.11 shows the customer segmentation categories that the respondents most focus on in their marketing strategies. 86.2% of respondents indicated that all three segmentation types(demographic, behavioral, geographic) are important to their businesses marketing needs. This aligned with the past research paper result that mentioned hybrid segmentation was crucial to obtain a more deeper insight into their customer groups. While few respondents (13.8%) selected only behavioral or geographic segmentation.

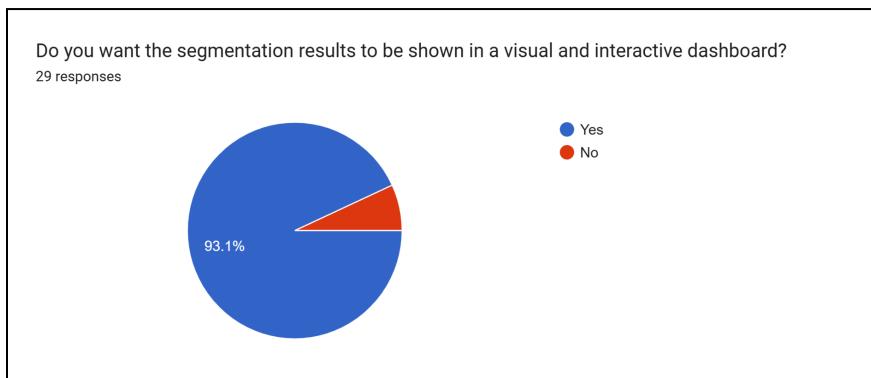


Figure 3.12 Survey - User Preference for Visual and Interactive Dashboard

Figure 3.12 presents respondents' preferences regarding the presentation format of segmentation results. 93.1% respondents prefer to display the segmentation results in a visual and interactive dashboard.

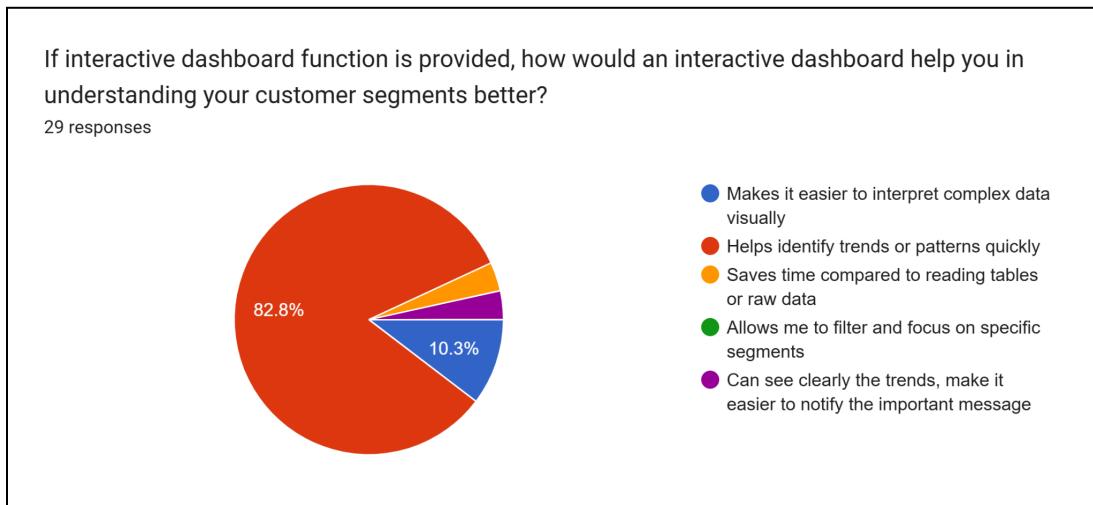


Figure 3.13 Survey - Perceived Benefits of Interactive Dashboard

Figure 3.13 further explores how an interactive dashboard would help users in understanding customer segments more effectively. The majority of respondents (82.8%) indicated that it would help identify trends or patterns quickly, suggesting that visual dashboards can improve analytical clarity. Other selected benefits included easier interpretation of complex data (10.3%), time-saving advantages(3.4%), filtering for specific segments, and clearer identification of key trends(3.4%).

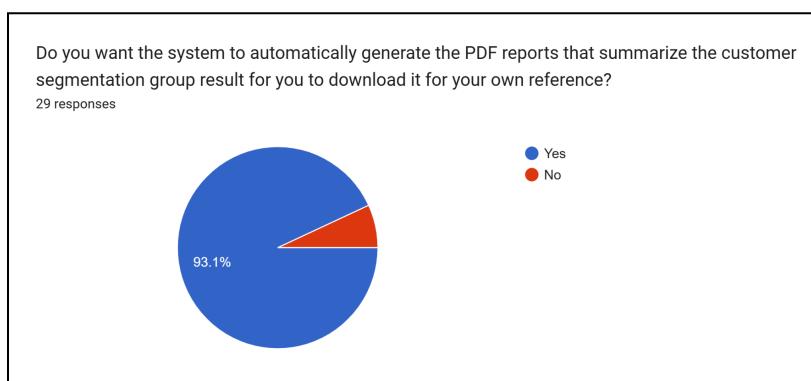


Figure 3.14 Survey - Preference for Automatic Report Generation

Figure 3.14 shows that 93.1% of respondents also preferred the system to automatically generate PDF reports summarizing the segmentation results. This feature is considered useful for record-keeping, internal reference, and decision-making, especially when comparing segmentation outcomes over time or sharing insights with team members.

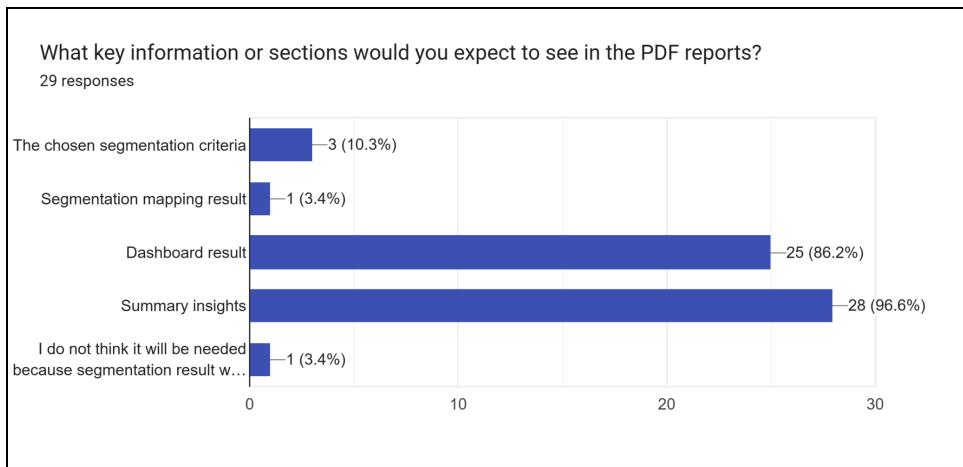


Figure 3.15 Survey - Report Content Expectations in PDF

Figure 3.15 shows the majority of respondents (96.6%) expect summary insights, while 86.2% value the dashboard results. A smaller portion considers chosen segmentation criteria (10.3%) and segmentation mapping results (3.4%) as necessary. One respondent indicated that a PDF report is not needed. However, the majority of these respondents highlight the importance of including high-level summaries and visualized results to ensure clarity and usefulness of the reports.

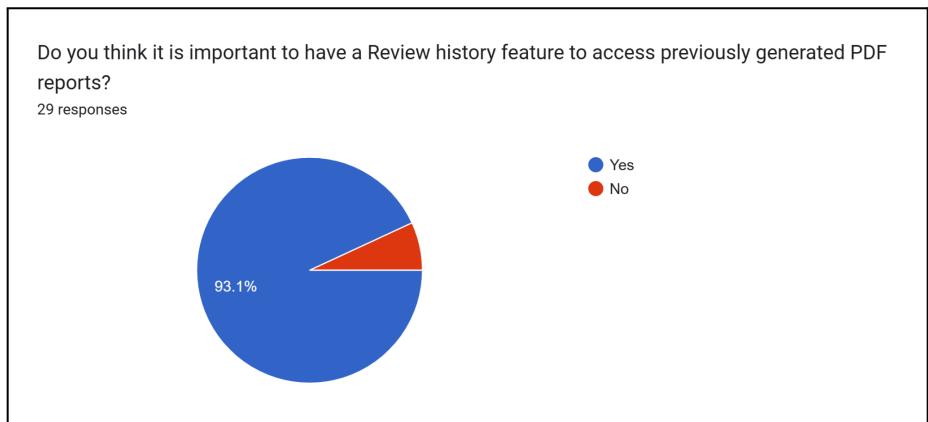


Figure 3.16 Survey - Importance of Having Access to Past PDF Reports

Figure 3.16 shows that the majority of respondents (93.1%) believe it is important to have a Review History feature that enables users to access previously generated PDF reports. This reflects the need for ongoing reference, historical comparison, and traceability of past segmentation results. Only a small portion of respondents (6.9%) did not consider this feature necessary. This feedback supports the inclusion of a report history module in the proposed system.

3.2.1.4 Additional Suggestion/Feedback

This section was an open-ended response which allowed respondents to provide additional feedback or their perspective on how the proposed system could be enhanced to better serve SME needs. Below are the details screenshot of the respondents feedback:

I hope you can guide us what dataset need to be used or what column or characteristic need to use because I do not know which data I need to provide and I also not sure which characteristics is belong to which categories.

Does uploaded dataset will have any format? Or is freely to upload any kind of customer dataset? In my point of view, I hope that it can have a fix template provided by your system because current system I use is they will give a fix templete, and we will need to follow their templete guideline. This will be better because if no have fix templete, we would not know what kind of dataset need to be uploaded. As my company also do not have real-time dataset, we do not know how to use real-time API like what you mention in your question. So, I will hope that the system will develop in a way that is using uploaded dataset manually.

For the dataset, if you were able to provide the sample templete will be better because we do not know which kind of dataset would you expected to use for doing segmentation, if have a templete, it can guide us to prepare our dataset. Most of the dataset we will manually recorded, and have different kind of the characteristic is recorded, so if you do not have dataset templete, I will upload all into your system, and not sure does all the column is related to your segmentation category

For the dataset, I would recommended that you can provide a example, what kind of dataset would you expect to use so that we can based on your template to upload the dataset

In my point of view, I wish to have a fix template to upload the dataset into your system because currently what I use is more on real-time data and is a bit difficult to use, I have hired a expert who know to use this google analytics tools. Next, I also have record my customer data manually, so I will upload into google analytics and create segment based on our knowledge. Comparing to real-time dataset that link to third party, I will more prefer on manually uploaded dataset into your system for further segmentation use. And I hope can have a templete for me to refer so that I can compile my recorded data to your templete or else I would not know which column or data can be use for segmentation process later, will struggle in determine which column need to include before upload the dataset.

The dataset have provide any example or have any template? We wish to have the guidance in dataset as we do not know what kind of dataset you want. Our company does no have real-time dataset and more rely on the manually recorded, so we more prefer to manually uploaded dataset first, we may follow your dataset template. If you no have template, we would difficult to know what dataset you expected to do segmentation.

In my POV, I would like to know how to use your system and easy to navaigate from one to another. For dataset, maybe you can provide in details what kind of information you want, because I would not know what data you will like to have. For your information, my company will more reply on manually recorded data so for real-time data may be a bit difficult because we do not know how to use real-time dataset since our company framework is very small, maybe if in future you may improve your system to include both manually upload dataset and link with real-time dataset but must support manually upload dataset first.

For dataset, if you were not define what kind of dataset would expect, it may be not consistent in term of the result given, so if best, it would be better if you can let user know what dataset you want/expected. What need to be include so that we can follow your template

Figure 3.17 Survey - Standardized Dataset Template Suggestion

Figure 3.17 shows that one of the common suggestions raised by respondents was the need for a **standardized dataset template**. Several SME users highlighted that having a fixed structure for uploading customer and order information would reduce the confusion and create ease of preparation. This is especially important for users without a technical background, as it ensures that they have consistency between the criteria that needed to be input to limit the risk of errors whilst uploading. Accordingly, the proposed system will take into account offering structured dataset templates to assist users in preparing and formatting the data correctly for import.

I hope that it can allow to look into details how your system work, such as how your system do the segmentation, and how to use your system.

If I am a first time user, I hope that I can know how to use your system and is easy to use such as easy to navigate from one page to another. And I also want to know how can I trust your system generated segmentation result can be trust

It would be better you guide us how to use your system, because we do not know how to use it. Our company do not have experience to use this segmentation tools before.

Figure 3.18 Survey - User Interface and Usability Suggestion

Figure 3.18 shows that respondents emphasized that they wanted the interface to be user friendly, with an intuitive design that makes navigation easy for new users who may not have prior experience with segmentation tools. Respondents emphasized a need for clear instructions on how to use the system, and wanted such information to include step by step directions on how to upload datasets, how to choose segmentation criteria, and how to interpret the results. Thus, this feedback will be taken into account when finalized the key function and design of the proposed system.

I hope that it can provide a feedback/complain part to let us deliver the problem we faced to you. Because I do not have experience in using this segmentation tools, so if have a feedback/complain part, I can reach out to your team when I faced any issues.

Figure 3.19 Survey - Feedback Feature Suggestion

Figure 3.19 shows one of the respondents suggesting including a feedback or complaint submission feature within the system. As many SME users may not have prior experience with segmentation tools, this feature would allow them to easily report issues, seek assistance, or share suggestions directly with the development team.

3.2.2 Interview

An interview was conducted with both a collaborator and a stakeholder to gain deeper insights into the current challenges they face and to gather feedback on how the proposed system could be enhanced to better meet real-world needs. This qualitative discussion helped identify practical requirements, system expectations, and key pain points from their perspectives. The detailed breakdown of the interview questions and responses is presented below.

3.2.2.1 Interview With Collaborator

An interview was conducted with Dr. Lucia, an academic collaborator with relevant experience in business and economics. The purpose of the interview was to evaluate the practicality and feasibility of the proposed functions of the system and to seek expert views about market segmentation strategies based on user perspectives.

Dr. Lucia pointed out the suitability and importance of the suggested system in relation to new-generation technological advancements and the increasing needs of the business environment for complex web-based platforms. She argued that it is practical to utilize a CSV-based market dataset in real-world applications, in particular for those SMEs with limited access to large automated systems.

Table 3.1 shows the feedback received from Dr. Lucia along with my responses to each point.

Table 3.1 Feedback from Dr Lucia

Feedback	Reply to the feedback
The topic is interesting and updated because it relates to the development of current technological advances and this is focused on the use of web/Web-Based Models that are connected to the business world, namely Market Segmentation. So the benefits of the	This feedback supports our intended contribution of the research, which seeks to help organizations (especially SMEs) improve their business processes with feasible and data-driven segmentation tools. The recognition further affirms the importance of the proposed system in meeting the practical challenges facing organizations in successfully adopting digital

<p>results of this research can help organizations in running their business.</p>	<p>solutions in a timely manner.</p>
<p>Let's have the same perception first. Is this research essentially making market segmentation using a CSV-based marketplace dataset? If so, the stages/steps/methodology seem to be okay. Maybe what needs to be clarified is the 4th step. How many clusters there are and what.</p>	<p>Yes, it is essential to use a CSV-based marketplace dataset.</p> <p>Regarding the question about clustering:</p> <ul style="list-style-type: none"> - In this example, we don't specify the number of clusters or the members of each cluster in advance. When a user uploads their CSV dataset and selects the attributes they want to segment by (i.e., Age, Spending Behavior, Country), the system would then run clustering algorithms, e.g. K-mean clustering, and evaluation tools (using something like the Elbow Method or Silhouette Analysis) to assess the optimal number of clusters based on the naturally occurring structure of the data provided by the user. - As an instance, if the user selected Age, Spending Behavior and Country, the clustering process would produce clusters based on commonality in these areas e.g. Younger high spenders in Malaysia, Older budget-conscious buyers in Singapore.

Regarding customer data in step 1 can be like that or maybe my input:

1. Demographic: Age, Gender, Income, Education, (it's okay but maybe it can be considered related to **marital status/marriage** because maybe the type of product needs can be different)
2. Geographic: Country, City (it's okay if customer data may be country **and city**, because it can later describe the customer's location, the characteristics of the rural or urban area that affect consumption patterns, the level of purchasing power in a certain area, weather conditions that require a certain group of products and others)
3. Behavioral: purchase history, product usage, website interaction (it's okay maybe you can consider **the time of purchase**)

To facilitate consistent data input and reduce the burden on SME users, a standardized dataset template for both customer and order data will be provided. The suggested additional attributes will be added into the template to ensure relevance and completeness.

3.2.2.2 Interview with Stakeholders

An interview was conducted with a stakeholder from Soapan Santun Enterprise, an SME specializing in handmade soap products. The company maintains customer and order datasets collected through its website and sales records. They have prior experience using Google Analytics, primarily for basic performance monitoring and simple segmentation tasks. However, their use of segmentation features has been limited due to several challenges. The following section presents the details of the discussion and findings. A series of questions were asked after the proposed system was introduced to the stakeholder. Based on their feedback, the functionalities of the proposed system were refined to better align with their needs and expectations.

Table 3.2 Interview with the Stakeholder

Section A: Understanding Current Challenges	
Question	Answer from Stakeholder
What difficulties have you encountered when trying to segment customers using your current tools or systems?	The segmentation capability of our tool has been difficult to use because of a lack of technical ability. The company has no data analytics expert, and no one on our team is fluent in Google Analytics. For this reason, we don't get much recourse to segmentation or meaningful segmentation when we do. The platform also has an unacceptable number of other features that have nothing to do with segmentation, such as the product analysis features and marketing tracking capabilities. Navigating around the full platform has made it unwieldy and hindered our scope of focus once we get into segmentation specific tasks. As of now, we only use the basic rule-based filters to create basic customer segments.
Have you found current tools to be overly complicated or difficult to use without	Yes, the current tools are overly complicated to use without technical support. The user interface is not beginner friendly and navigating through all the features to find the

technical support ? If yes, what is the most confusing or time consuming part?	segmentation functions is challenging. The time-consuming part is that I cannot understand how to use segmentation rules to apply and correctly configure segmentation rules, particularly when trying to combine all the conditions or when looking at behavioral insights.
Are you currently able to segment customers using multiple dimensions (e.g., demographic + behavioral)? Why or why not?	No, we mostly focus on demographic segmentation. We lack the knowledge and guidance to implement segmentation across multiple dimensions, such as behavioral or geographic data. Without clear tools or simplified guidance, it is difficult for us to explore beyond basic demographic filters.
Section B: Needs and Feature Suggestions	
What do you think I can improve more on my proposed system?	The overall objective for the proposed system appears to be strong, especially in terms of hybrid and a posteriori segmentation, which was noted to be a significant value over existing systems and actions towards obtaining better customer insights. I would like to suggest one more feature, it would be best if the feedback/complaint feature is included in your system. If users encounter any issues, they may reach the team easily.
Would it be useful if the system allowed you to upload your own dataset (e.g., CSV format)?	Yes, I believe most of the SME will use manually recorded data. While some may use real-time data, I would recommend that the system initially focus on supporting manual data upload in formats such as CSV. In the future, it could be enhanced to support both manually uploaded datasets and real-time data integration.
Do you require the ability to generate downloadable reports	Yes, that would be very helpful. Currently, no one system provides any reporting functionality, we are unable to revisit

<p>(e.g., in PDF) based on segmentation results for review or sharing with your team?</p>	<p>or review past segmentation results. If such a report feature were available, it would allow us to track and compare current segmentation outcomes with past results. This would help us better understand changing customer trends over time and support more informed business decisions.</p>
---	--

3.2.3 Literatures Survey

The literature survey was conducted to explore more about the conceptual framework and the real-world challenges with market segmentation systems. It helps in understanding what the current systems have, what the current segmentation used, what are the advantages and potential limitations. It also helped in providing a more clear and deeper understanding into the market segmentation. A comprehensive literature search of the articles was conducted to examine the market segmentation theory, the contrast between *a priori* and *a posteriori* market segmentation methods, and the use of segmentation criteria such as demographic, geographic, and behavioral factors.

Several research articles were thoroughly examined to gain better insights:

- The common methods and strategies used in segmentation are clustering methods, and traditional rule-based methods.
- The advantages and limitations behind each segment in various industrial uses.
- How segmentation can influence marketing strategies, customer personalization, and business decision-making.

This analysis helped define the system's architecture parameters and justified the need for a more flexible, interactive, and transparent segmentation system. It also provides insights to address the shortcomings of current web-based systems like Amplitude, Hotjar, Google Analytics, and Insightly. The insights obtained from academic literature ensures that the proposed system can meet the current market needs and offers meaningful improvements over existing systems.

CHAPTER 4: SYSTEM ANALYSIS AND DESIGN

This chapter discusses the analysis and design aspects of the proposed web-based market segmentation system. It defines the system requirements that include both functional and non-functional, which explain what the system should do and the performance levels it needs to meet. The next section presents several design models that explain the system's architecture and the interactions of its various components. They comprise the Functional Decomposition Diagram (FDD), the Use-Case Diagram, the Activity Diagram, and the Entity Relationship Diagram (ERD) that present greater details of the functionalities of the system, the interactions with the end-users, the processes involved, and the relationships among the various data components.

4.1 System Requirements

This section shows the main requirements of the system. It is divided into Functional Requirements (FR), which describe the features of the systems that will develop based on proposed ideas and comparisons with existing systems, and Non-Functional Requirements (Non-FR), which describe the quality attributes of the system such as usability and security.

4.1.1 Functional Requirement

The functional requirements describe what the system should do. These are derived from the identified needs, collaborator feedback, literature review, and the gaps found in existing systems (Amplitude, Hotjar, Google Analytics, and Insightly). The system is divided into several modules, each module consisting of features designed to support user workflows effectively and improve marketing segmentation processes. Below are the main functional requirements and description of the system:

Module 1: Registration Module

Table 4.1 Registration Module

FR	FR Description
Register Account	First time users can register and create their account by providing information such as username, email, and password.
Verification and Activate Account	Users will receive an email to verify and activate their account.

This module allows first-time users to create a new account by providing their personal information such as username, email, and password. For security purposes, the system requires new users to verify their account through an email verification process. Upon successful registration, a verification code will be sent to the user's registered email address. The account will not be activated unless the code is submitted correctly. This additional layer of security ensures that users are genuine and not spammers or fraudulent registrants.

Module 2: Login Module

Table 4.2 Login Module

FR	FR Description
Login Account	Users can log into their account by providing the correct email and password.
Forgot Password	Users can reset passwords through their email if they forgot.

This module handles user login, and password recovery functionalities. Users can log into their account using their registered email address and password. If users forgot their account password, a password reset feature is available. A reset password link that contains the verification code will be sent to the user email. Users need to provide the correct verification code to verify their authentication before they can reset the password.

Module 3: User Profile Module

Table 4.3 User Profile Module

FR	FR Description
Edit Profile Information	Users can edit their profile information such as username, email, and password to keep information current.
Delete Account	Users can delete their account if they no longer want to use our system

This module allows the users to edit their account if users need to keep their personal information up to date such as users are available to edit their profile username, email and password. In addition to editing their account, users have the option to permanently delete their

account. Once an account is deleted, all associated data including personal information, uploaded datasets and generated reports will be removed from the system and cannot be restored.

Module 4: Dataset Management Module

Table 4.4 Dataset Management Module

FR	FR Description
Upload Dataset	Users can upload customer and order data files in CSV format. The system validates file format and content structure for correctness and relevance.
Preview Dataset	Users can view the first 100 rows of the uploaded dataset to verify its content.
Delete Dataset	Users can delete their uploaded dataset from the system based on their need.
Search Dataset	Users can search for specific dataset by providing the dataset name in the provided search bar.
Download Dataset	Users can download the dataset template that provided under the dataset page

This module allows users to manage their customer and order datasets within the system. Users can upload customer and order dataset in CSV format and the system will automatically validate the uploaded files types and structure to ensure it aligns with the required format. The system verifies the file format type (CSV) as well as the structure of the uploaded dataset relative to the uploaded dataset template. If there are any failures of these conditions, an error message displays indicating to the users to upload a file that is valid. In addition, users can view the first 100 rows of uploaded dataset to review the dataset they have selected. Additionally, users are able to delete any of the uploaded dataset as needed. It provides a feature for data deletion that allows users to control the content of the uploaded datasets. Users can remove older datasets that are no longer needed for the purpose of efficient management of their storage space and sensitive data. This feature is especially useful for the users who are only interested in retaining the most recent or favorite datasets. Users can also search for specific dataset by entering the dataset name in the search bar, which improves the dataset search functionality. This is useful when the data collection has a lot of datasets. Eventually, users can download the standardized dataset template

from the dataset page, which will provide the users with a reference to ensuring that they format their data correctly before uploading.

Module 5: Dataset Selection Module

Table 4.5 Dataset Selection Module

FR	FR Description
Select Dataset	Users can select one dataset from both the customer and the order dataset list to use for segmentation.
View Dataset Attributes	Users can view into the detailed list of the dataset attributes that has been classified into correct segmentation categories.

This module allows users to select a dataset from both the customer and order dataset lists that they wish to use for segmentation. The system will perform the checking to ensure the dataset uploaded has undergone the data cleaning process. If the dataset has already gone through the data cleaning process, the system will skip that step and will simply present the categorized attributes to the user. If the dataset has not gone through the data cleaning process, the system will automatically complete the data cleaning tasks such as the handling of missing values, duplicates, normalizing or standardizing numeric data before storing the cleaned CSV file into the database for reuse. Once the dataset is selected, the system will merge the relevant customer and order datasets into one file, and will then categorize the attributes of the merged dataset for segmentation before displaying to the user for attribute selection.

Module 6: Segmentation Module

Table 4.6 Segmentation Module

FR	FR Description
Select Segmentation Criteria	Users can select two to three specific attributes they want to use for segmentation.
Start Segmentation Process	Users can click the “Start Segmentation” button to begin the segmentation process based on their selected attributes
View Segmentation Result	Users can view the segmentation results through a summary table and an interactive dashboard.
Refine Segmentation	Users can adjust their selected attributes and rerun the segmentation

Criteria	process to explore different segment groupings.
----------	---

This module helps users set up how they want to segment their customer. Users can choose between two or three attributes from any segmentation categories which reflect the hybrid segmentation can be worked in the system. According to these selections, the system will apply a machine learning clustering process such as K-means, hierarchical clustering or any other suitable algorithms and will group similar customers into meaningful segments. After the customers have been segmented, the system will provide the results in a summary table and in a graphical presentation including an interactive dashboard in which the users can visualize each segment and their characteristics. Finally, users will also be able to re-do the segmentation process by selecting different attributes to explore multiple segmentation strategies as needed.

Module 7: Reports Module

Table 4.7 Reports Module

FR	FR Description
Generate Report	Users can click the “Save” button to save the segmentation result and the system will generate a PDF report that summarizes customer segments results and save it in the report list.
Download Reports	Users can download reports in PDF format for their own use.
View Report History	Users can review past generated reports and can re-downloaded it when needed.
Delete Reports	Users can delete reports when needed.

This module focuses on making it easy for users to save and download the market segmentation results. The system automatically generates detailed PDF reports that include summaries on the results of the segment groups when the user clicks the "Save" button. Once generated, the reports can be downloaded to allow the user to keep a copy for use offline or to share with others. Furthermore, the previous reports are stored in a report history so the user is able to retrieve and re-download reports as they feel necessary in the future. Users have the opportunity to delete reports that they do not need to keep.

Module 8: Feedback Module

Table 4.8 Feedback Module

FR	FR Description
Submit Feedback	Users can provide any feedback and complain to our system.
View Status	Users can view whether the feedback they submitted is being solved or not.

Users can submit feedback or complaints about the system's functionality or usability in this module. Feedback submissions are tracked, so users can see the state of the feedback, whether it was viewed, not viewed, in processing, and handled. This feature ensures the system will continue to improve continuously based on real user feedback, and it is more transparent regarding how user feedback is treated.

Module 9: Admin Management Module

Table 4.9 Admin Management Module

FR	FR Description
Manage Feedback	Admin can view feedback submitted by users and update its status to "In Processing" or "Handled" based on its resolution progress.
Search Feedback	Admin can search for specific feedback by providing its specific feedback code to quickly search the feedback from the list.
View Feedback	Admin can view the detailed content of feedback submitted by the user.

This module is for Admin purposes only. We allow the admin to manage user feedback they submitted by reviewing the feedback information and updating the status. If it is a complaint, and it would take time to process, it will conveniently allow the admin to mark as "In Processing", which will inform the user and the admin of the status. Once the complaint is resolved, the feedback can be updated to "Handled". The admin can search for feedback by feedback entry's unique feedback code to help manage large volumes of feedback.

4.1.2 Non-Functional Requirement

The non-functional requirement defines the quality attributes of the system, focusing on aspects such as security, usability and performance. These requirements ensure that the system not only performs its intended functions but also operates securely and is accessible to non-technical users. Below table show the detailed explanation for each of the non-functional requirements for my system.

Table 4.10 Non-Functional Requirements

Non-FR	FR Description
Security	Security practices can provide end-users with protection against accessing user accounts, personal information, and uploaded information. The security of user's login information, as an example, will never be directly stored in plain text. Instead, a user's password will be converted into a secure string using hashing before being stored in the database. When users enter a password, the same kind of processes occur, and when the hashed passwords are matched, it authenticates the user. When a user logs in, the system assigns a secure token to establish that the user is logged in safely. Anytime a user submits a request to access, edit or update information, the system will first check the token to ensure that the request is legitimate. These methods can prevent unauthorized access of the user's personal information, but also prevent unauthorized access to a user's account. Additionally, users have the ability to manage their dataset and also delete their dataset whenever they see fit. Again, this is to give users improved privacy and allow for the individual the ability to manage their personal data. These efforts promote trust and provide a level of best practice in data security.
Usability	The system will be built with the primary focus on user accessibility and reducing the burden on those with limited technical experience to use it. Accordingly, the system will provide an easy-to-use interface that will allow users to interact with the application easily. Additionally, tooltips and contextual help will serve as quick instructions and assistance for users,

	<p>minimizing confusion. Lastly, the system will offer step-by-step instructions and prompts in every module, allowing users to perform any tasks seamlessly, and also fostering effective and confident use of the system.</p>
Performance	<p>The system architecture was also designed to enable it to stay responsive and run smoothly even with large customer and order datasets. The system applies some performance techniques to help achieve this outcome. For example, when users preview the uploaded datasets, it limits users' view to only the first 100 rows to minimize the load on their browsers. For more involved tasks like data cleaning and segmentation, we intend for the tasks to be as responsive as possible in the UI by executing them in the background. In addition to optimizing load and response time by limiting browser load, the system stores uploaded CSV files in MongoDB, as GridFS retrieves this type of data in smaller chunks and streams data as required from the Mongo storage layer, making it faster and more efficient while consuming less memory.</p>

4.2 System Design

Systems Design is the literal definition of architecture, components, or modules, interfaces, and overall structure of a system, to satisfy a set of requirements. It is the action of converting user needs or requirements to a specification of how the various components interact and work together as a system to provide the desired functionality, performance, and reliability. The main objective is to plan an organized and efficient structure to achieve the intended purpose. It will outline how the frontend and backend components work.

4.2.1 Functional Decomposition Diagram(FDD)

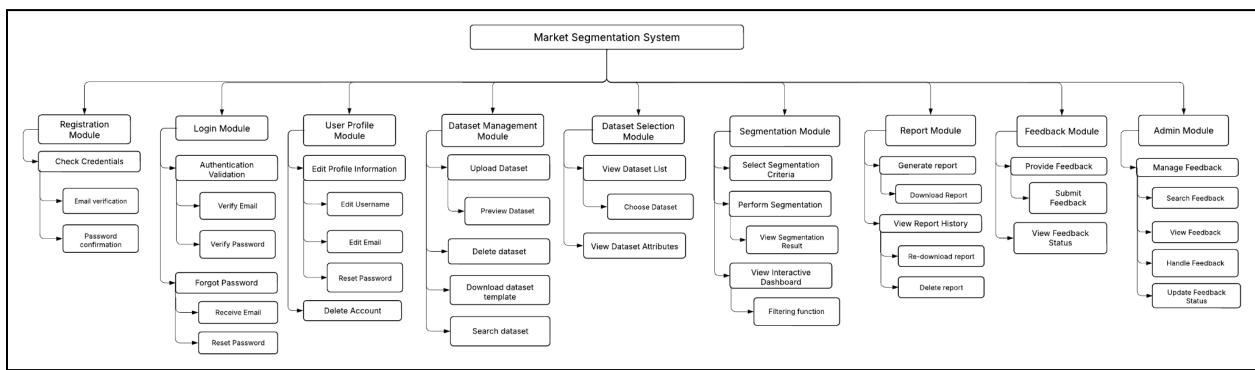


Figure 4.1 Functional Decomposition Diagram

Figure 4.1 shows the FDD of the proposed web-based market segmentation system. This diagram shows the breakdown of each module into smaller, sub-function to give a picture of how the system works. By showing this diagram, it allows the collaborator and stakeholder to quickly know what the system expected to do.

4.2.2 Use-Case Diagram

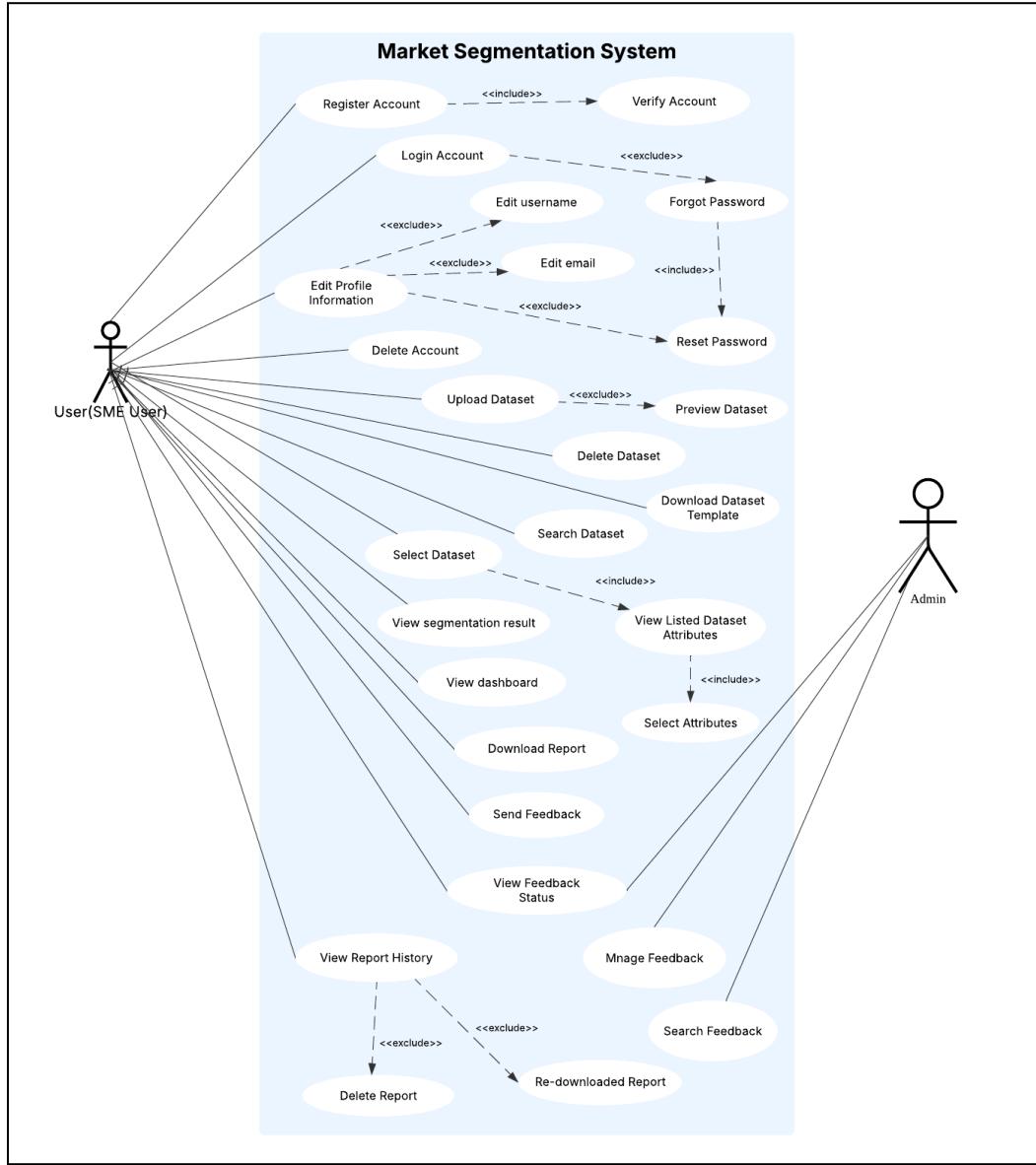


Figure 4.2 Use Case Diagram

Figure 4.2 shows the use case diagram of the system. It illustrates the interaction between two different actors which are SME users and Admin. It identifies and organizes the core functionalities of the system. It also includes the standard UML relationship such as <<include>> and <<exclude>> to represent the mandatory and optional flows within the system. This diagram allows the collaborator and stakeholder to understand the role of each user in the system and how they interact with various functions of the system.

4.2.3 Activity Diagram

This section outlines the activity diagrams for the main modules of the system. Each diagram explains the dynamic workflow related to a specific function, including register account, login account, user profile management, dataset management, segmentation processing, report generation, feedback management and admin management. The diagrams illustrate the steps from both the user's and system's perspective, emphasizing the starting points of tasks, points of decision, and achievement of results. By graphically illustrating each workflow, the activity diagrams allow validation of the fact that the system logic is correctly specified, consistent, and in conformance with user expectations in the development process.

4.2.3.1 Register Account

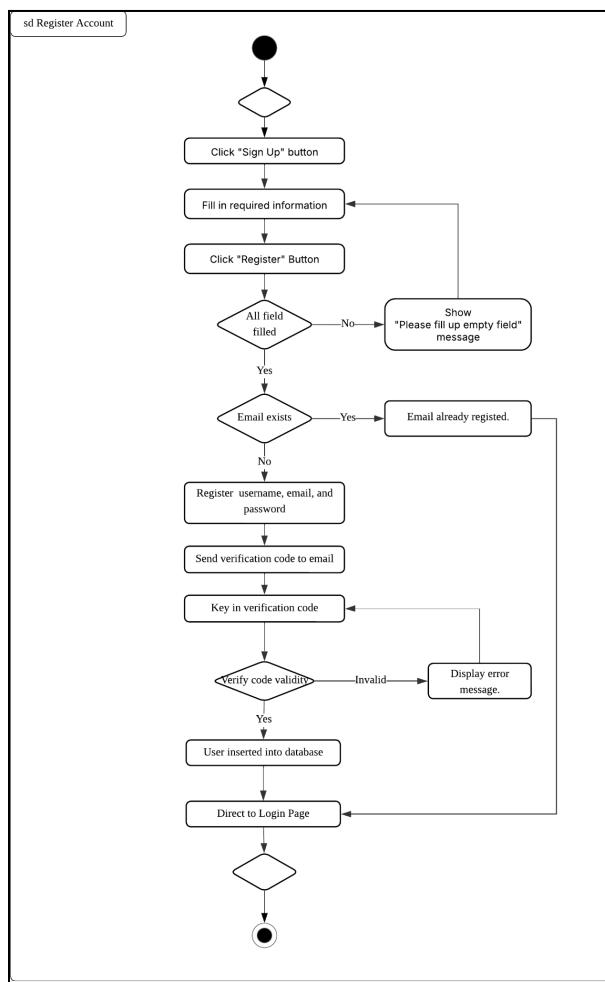


Figure 4.3 Activity Diagram of User Register Account

4.2.3.2 Login Account

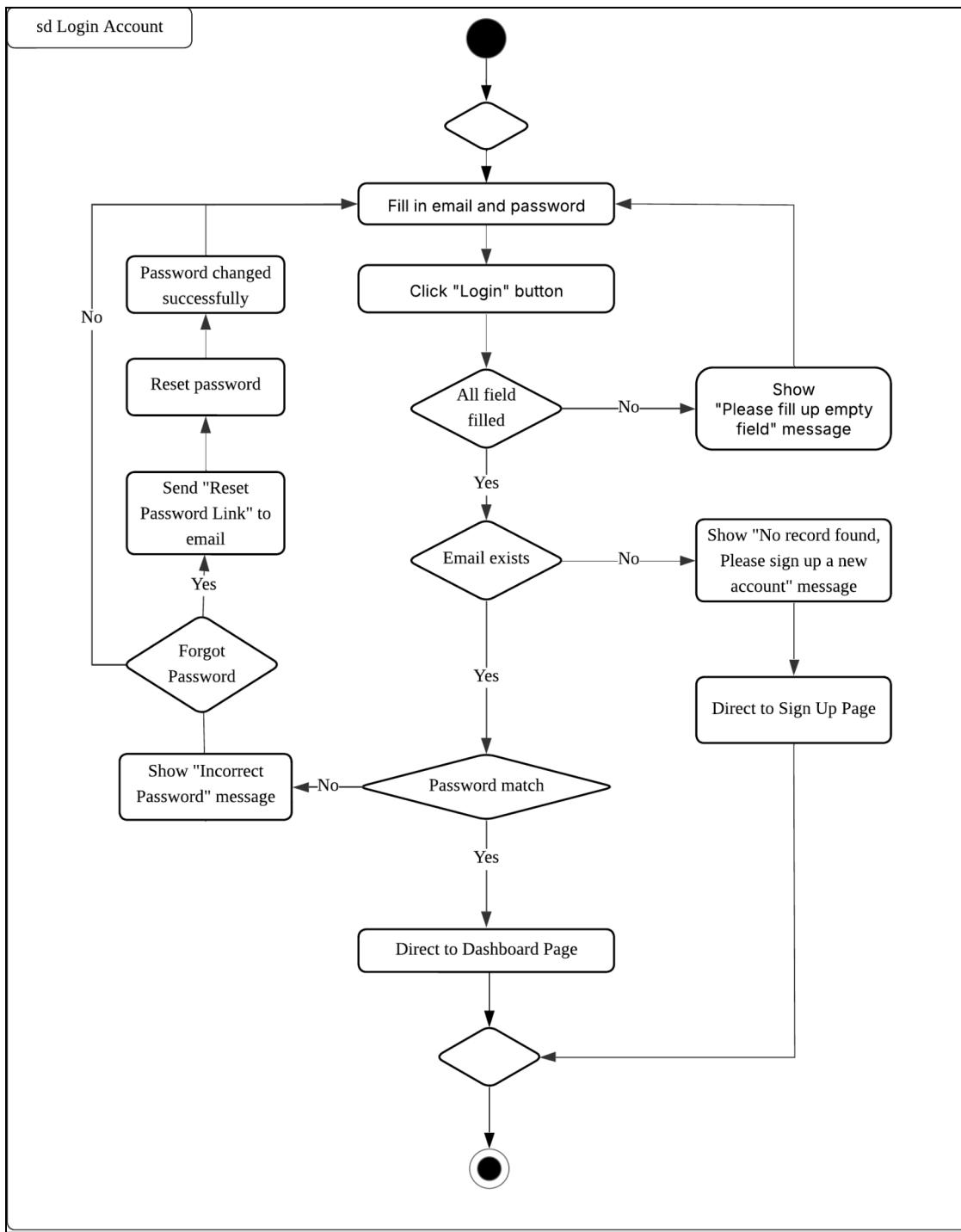


Figure 4.4 Activity Diagram of User Login Account

4.2.3.3 User Profile Management

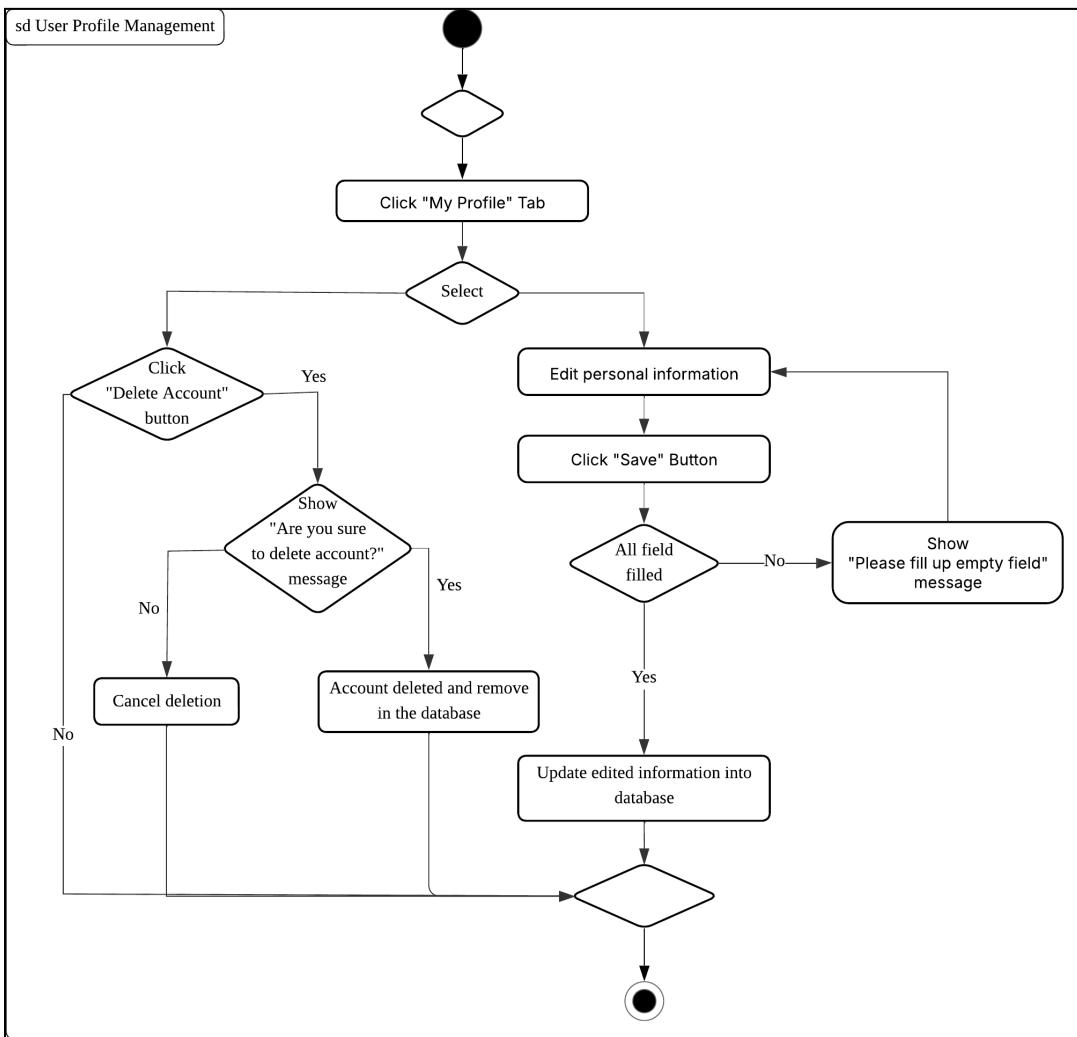


Figure 4.5 Activity Diagram of Manage User Profile

4.2.3.4 Dataset Management

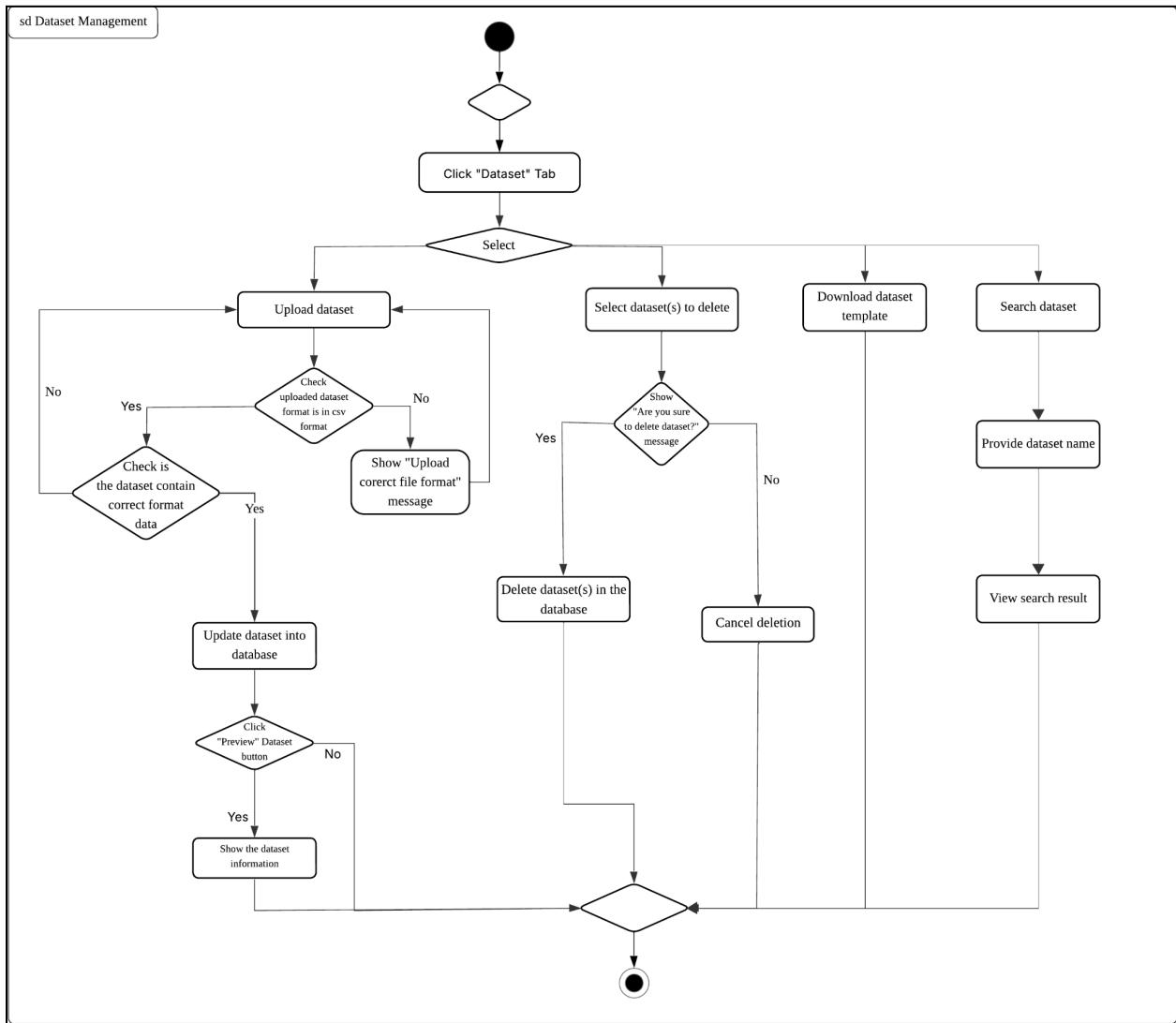


Figure 4.6 Activity Diagram of Manage Dataset

4.2.3.5 Dataset Selection

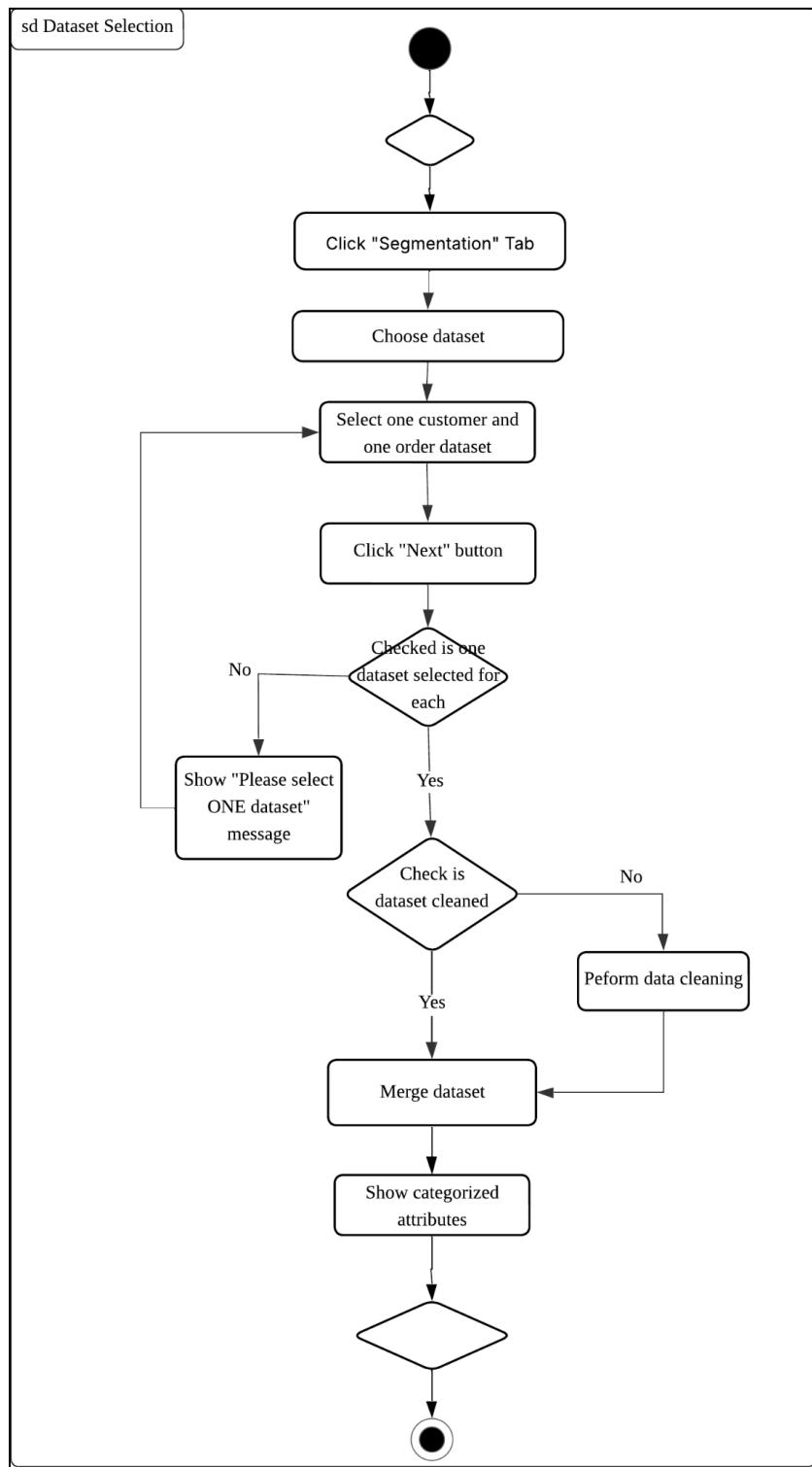


Figure 4.7 Activity Diagram of Select Dataset

4.2.3.6 Segmentation

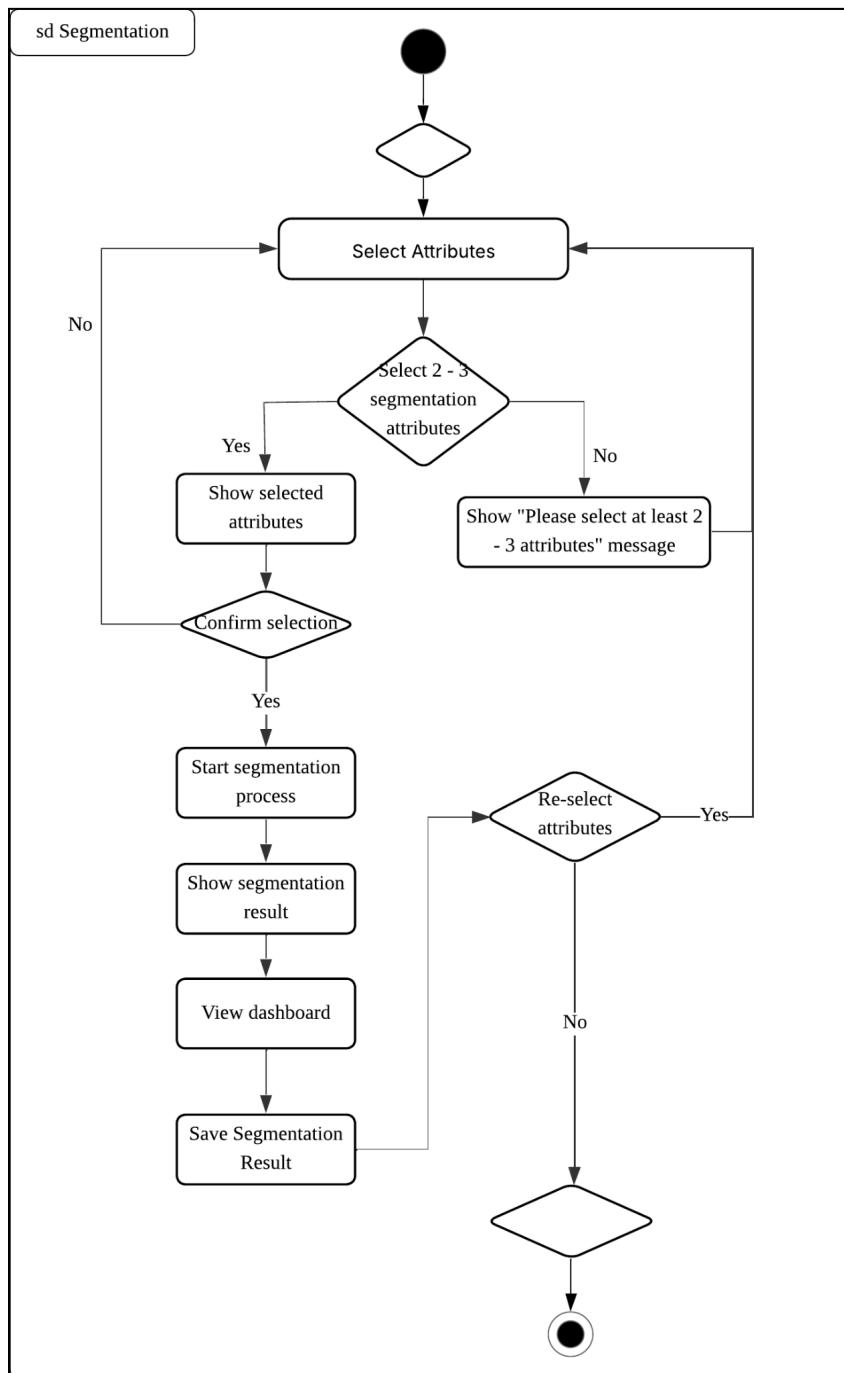


Figure 4.8 Activity Diagram of Perform Segmentation

4.2.3.7 Report Management

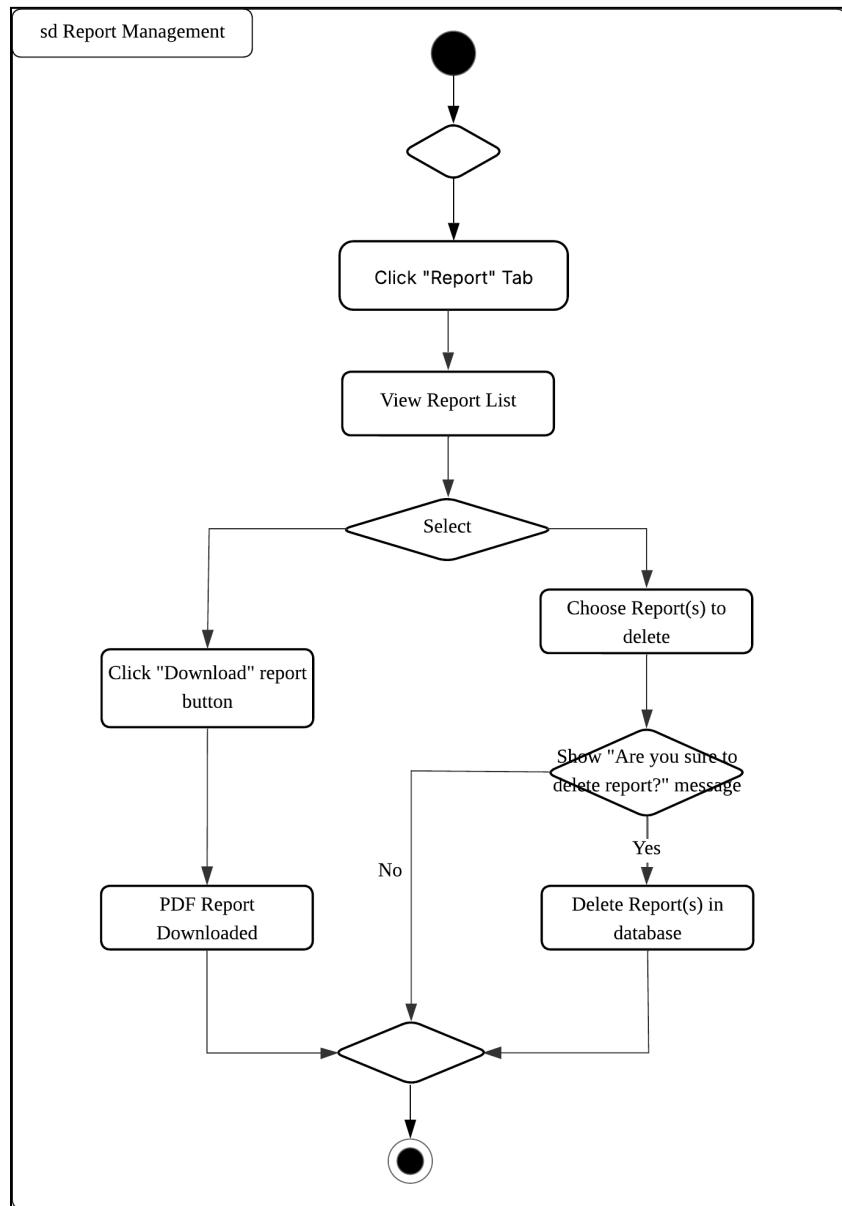


Figure 4.9 Activity Diagram of Manage Report

4.2.3.8 Submit Feedback

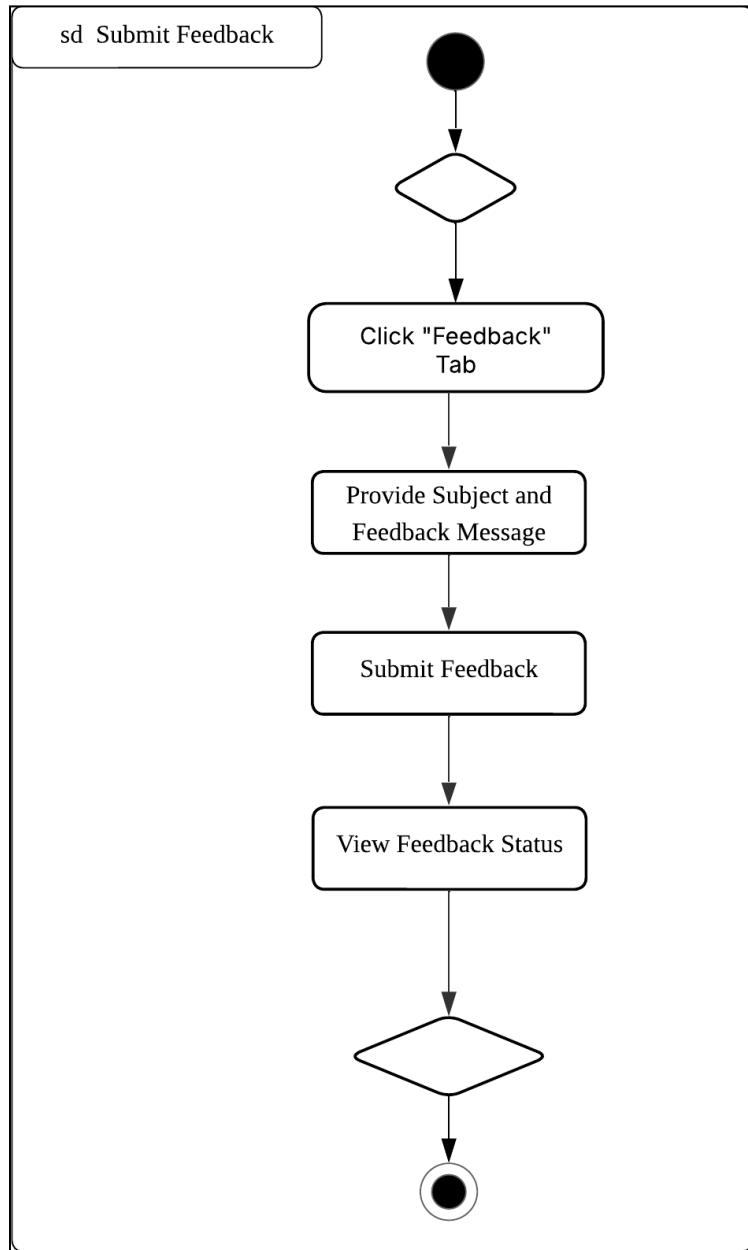


Figure 4.10 Activity Diagram of Submit Feedback

4.2.3.9 Admin Management

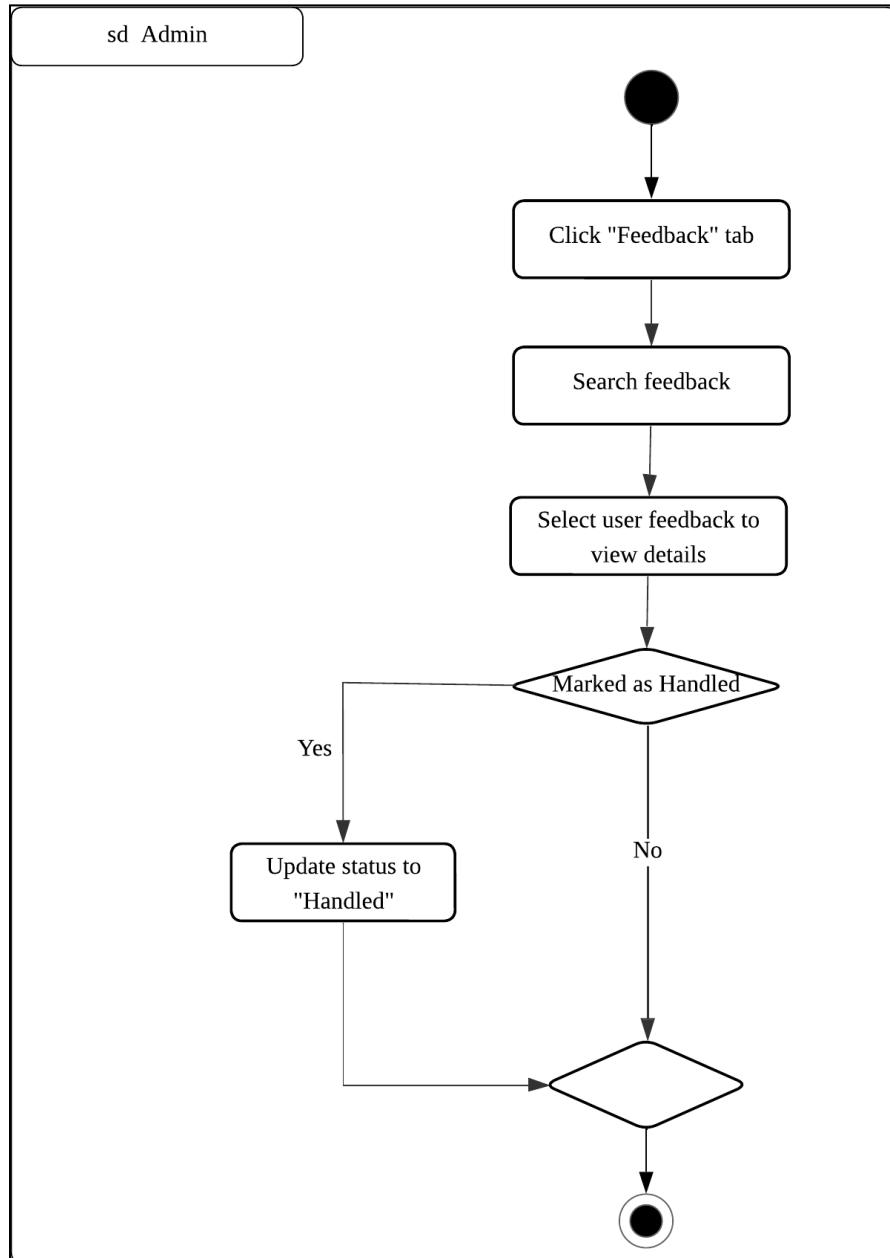


Figure 4.11 Activity Diagram of Admin Handle Feedback

4.2.4 Entity Relationship Diagram (ERD)

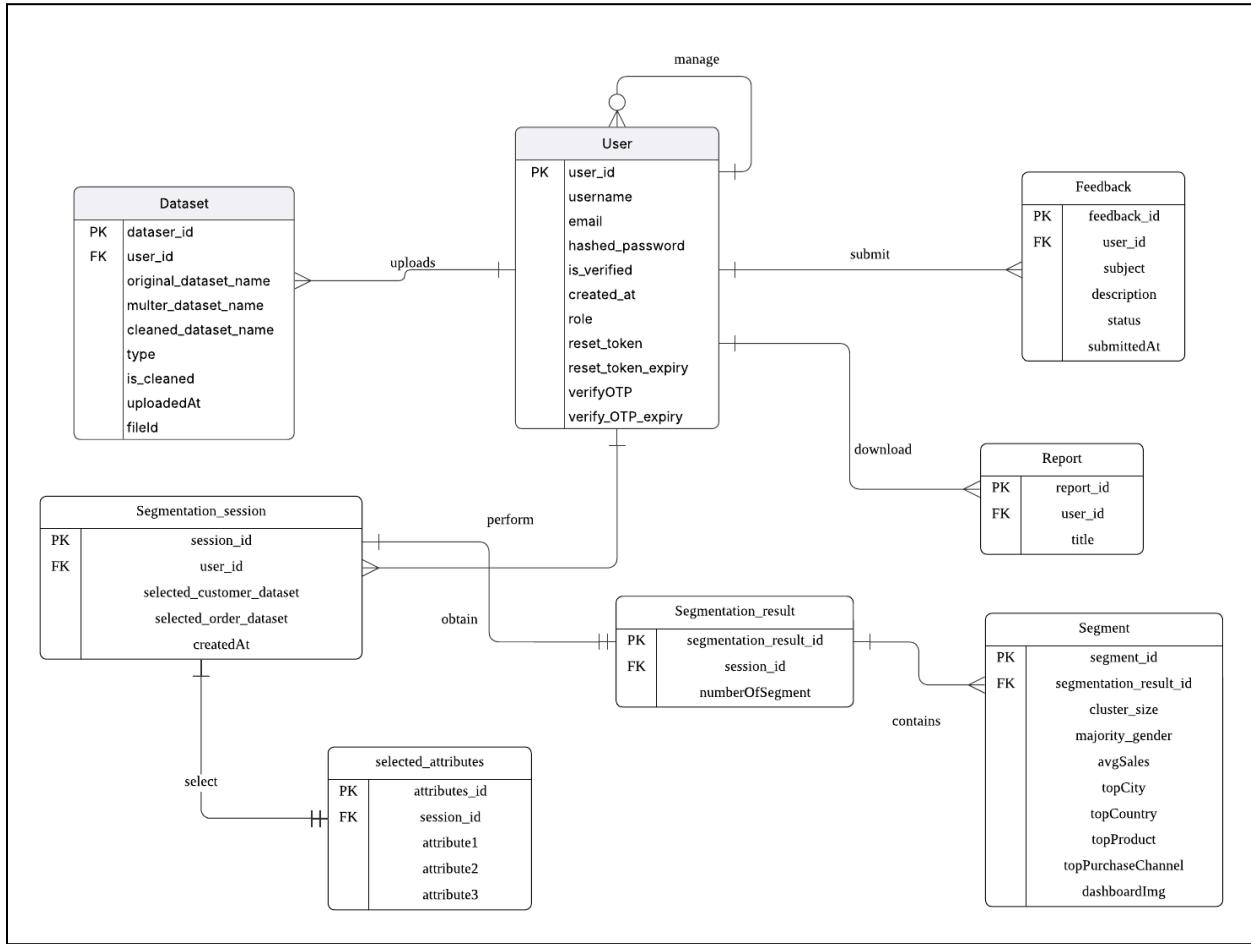


Figure 4.12 ERD

The Entity Relationship Diagram (ERD) illustrates the database structure of the proposed web-based market segmentation system. It defines the major entities in the system, including users, datasets, segmentation results, reports, and feedback, and shows the interrelationships among these elements. Each entity is described by relevant attributes, and the defined relationships maintain data integrity and logical consistency among the varied modules. This ERD serves as the foundation for the relational database design of the system and hence enables effective data structuring, retrieval, and scalability as the system grows. The ERD may be revised during the system development process, as the current version is a preliminary draft referenced from past research.

4.3 Interface Design

According to the system design, an initial mockup of the visual representation of the system was developed. This includes the visual layout, navigation structure, and the flow of the user interaction for all of the modules. The design is focused on keeping it as simple and as user friendly and usable, as possible so the user may use the system without overwhelming the user. This initial mockup allows the collaborator and stakeholder to see clearly how the system design and the flow look like. The final version of the design might be a bit different from the mock-up as it only serves as a prototype and the core functionalities will still remain the same. Below are the screenshot of the mockup UI:

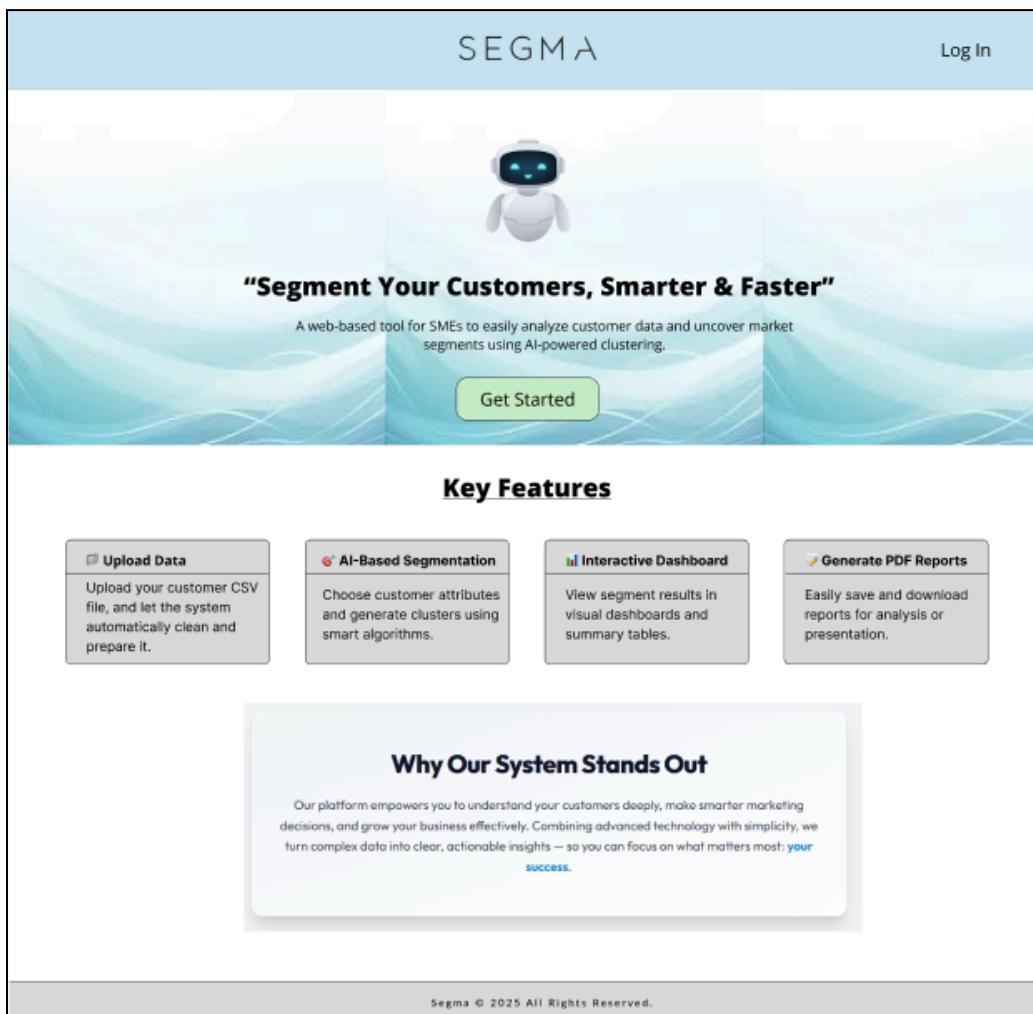


Figure 4.13 Mockup - Landing Page (Before Login)

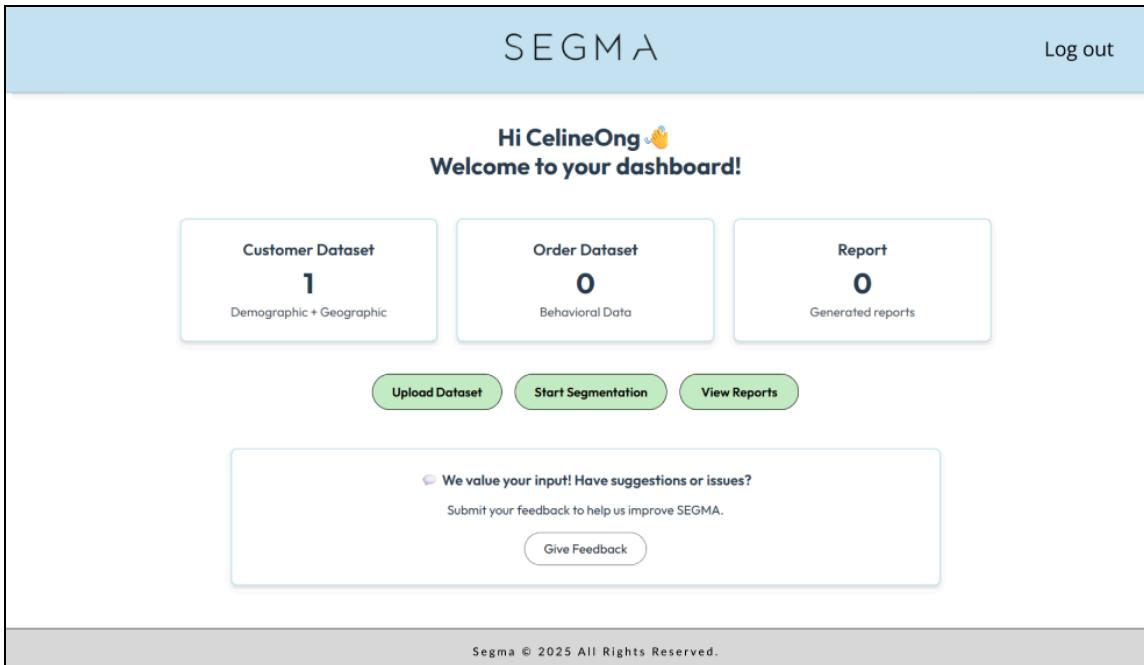


Figure 4.14 Mockup - User Dashboard

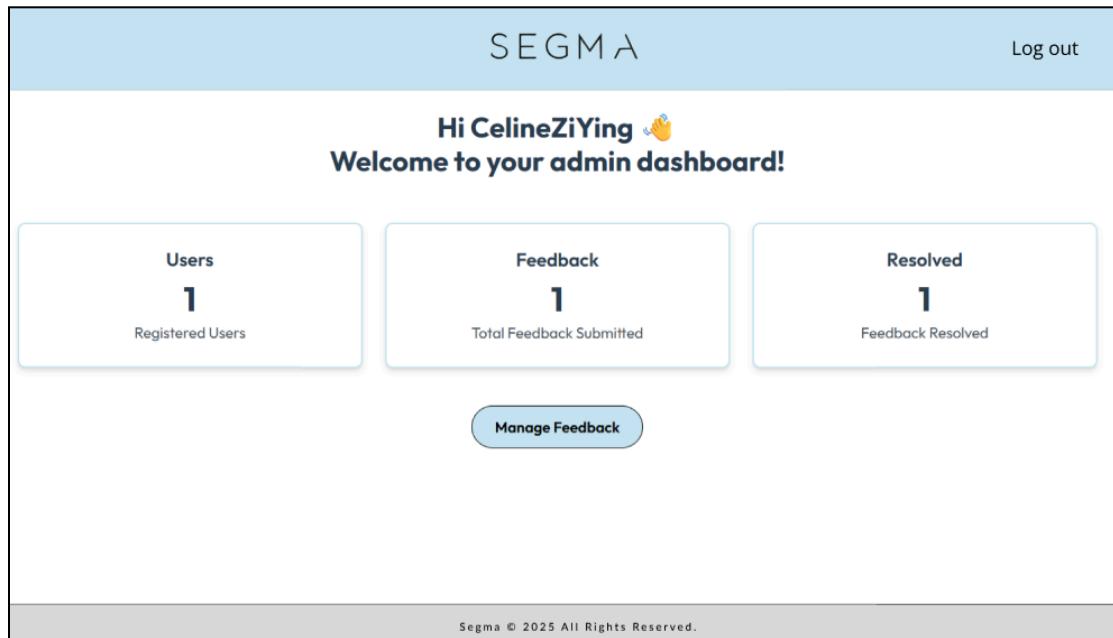


Figure 4.15 Mockup - Admin Dashboard

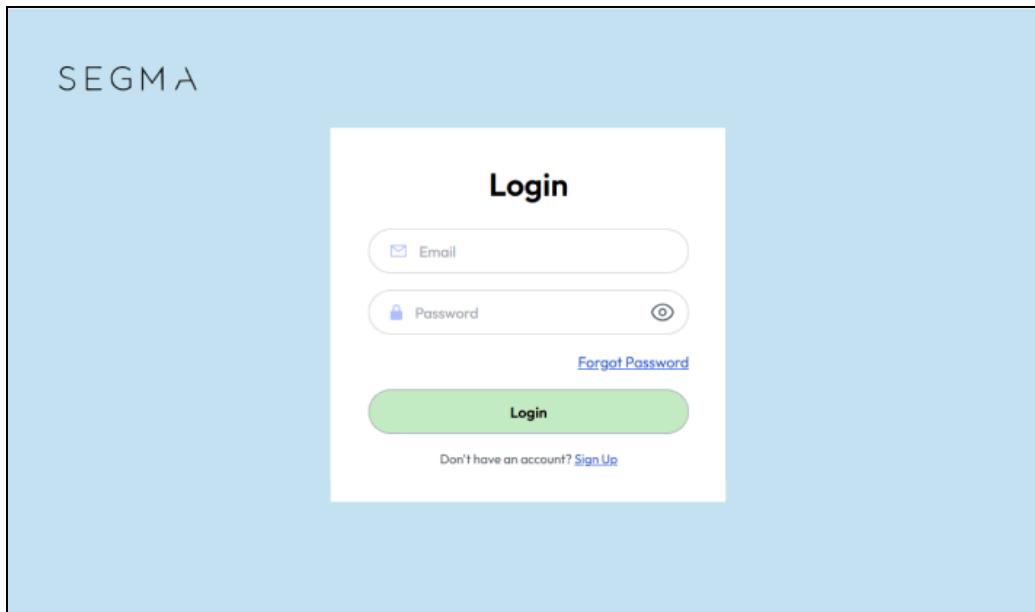


Figure 4.16 Mockup - Login

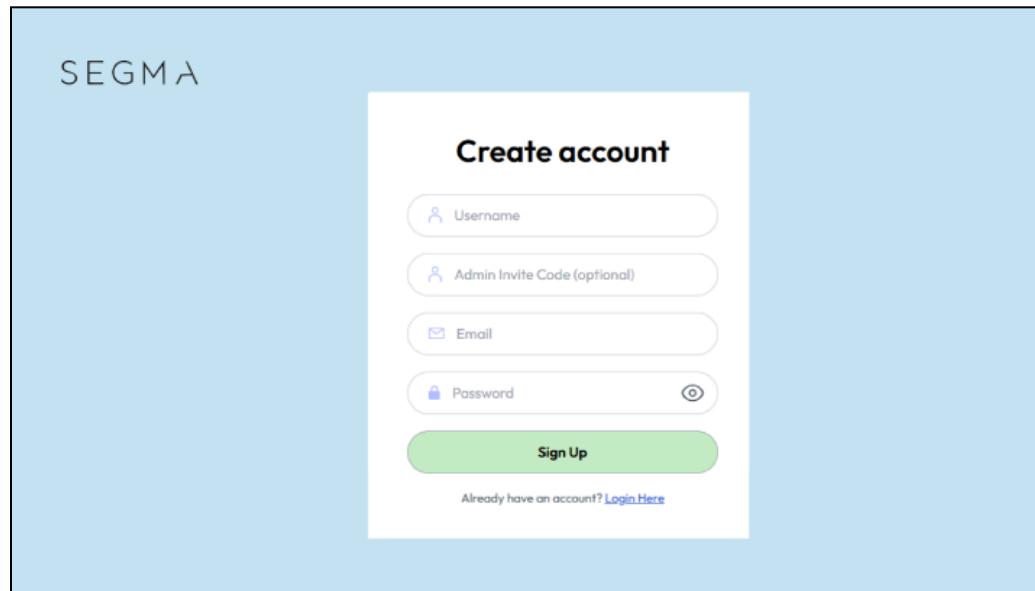


Figure 4.17 Mockup - Register Account

The screenshot shows the 'Dataset' page of the SEGMA application. The left sidebar contains navigation links: Home, Dataset (selected), Segmentation (with dropdown options: Selection, Result), Dashboard, Report, Feedback, and My Profile. The main content area has a title 'Dataset' and tabs for Customer, Orders, and a green 'Download Customer Template' button. A search bar says 'Search datasets...'. Below is a table with columns: NO., DATASET NAME, DATE uploaded, and ACTIONS (Preview, Delete). The table lists three datasets:

NO.	DATASET NAME	DATE uploaded	ACTIONS
1	customer_data_template (1).csv	6/17/2025, 12:03:23 PM	<button>Preview</button> <button>Delete</button>
2	customer_data_template (4).csv	6/17/2025, 12:03:10 PM	<button>Preview</button> <button>Delete</button>
3	customer_data_template (5).csv	6/17/2025, 11:33:27 AM	<button>Preview</button> <button>Delete</button>

A green button at the bottom right says '+ Upload Dataset'.

Figure 4.18 Mockup - Dataset Page

The screenshot shows the 'Dataset Selection' page of the SEGMA application. The left sidebar is identical to Figure 4.18. The main content area has a title 'Dataset Selection' and tabs for Customer, Orders, and a note '* Only one dataset can be selected'. A search bar says 'Search datasets...'. Below is a table with columns: SELECT, DATASET NAME, DATE uploaded, and ACTIONS (Preview). The table lists four datasets, each with a checkbox in the 'SELECT' column:

SELECT	DATASET NAME	DATE uploaded	ACTIONS
<input type="checkbox"/>	customer_data_template (5).csv	6/17/2025, 6:35:24 PM	<button>Preview</button>
<input type="checkbox"/>	customer_data_template (5).csv	6/17/2025, 12:03:58 PM	<button>Preview</button>
<input type="checkbox"/>	customer_data_template (4).csv	6/17/2025, 12:03:52 PM	<button>Preview</button>
<input type="checkbox"/>	customer_data_template (5).csv	6/17/2025, 12:03:46 PM	<button>Preview</button>

A green 'Next' button is at the bottom right.

Figure 4.19 Mockup - Dataset Selection Page

The screenshot shows the 'Customer Info' page of the SEGMA application. The left sidebar contains navigation links: Home, Dataset, Segmentation (with dropdown options: Dataset Selection, Mapping Result, Segmentation Result), Dashboard, Report, Feedback, and My Profile. The 'Mapping Result' link is highlighted. The main content area has a title 'Customer Info' and a subtitle 'Attributes assign according to segmentation categories'. It features three boxes: 'Demographic' (Age, Gender, Income, Marital Status, Occupation, Educational Level, Family Size, Nationality), 'Geographic' (Country, City, Region, Zip Code), and 'Behavioral' (Purchase History, Time of Purchase). A green 'Next' button is at the bottom right.

Figure 4.20 Mockup - Mapping Result Page

The screenshot shows the 'Select Customer Details' page of the SEGMA application. The left sidebar is identical to Figure 4.20. The main content area has a title 'Select Customer Details' and instructions: 'Choose how you want to group your customers.' and 'Pick 2 to 3 customer details below to help us create useful customer groups.' It features a box titled 'Customer Detail Categories:' with three columns: 'Demographic:' (Age, Gender, Income, Marital Status), 'Geographic:' (Country, City, Region, Zip Code), and 'Behavioral:' (Purchase History, Time of purchase). A green 'Next' button is at the bottom right.

Figure 4.21 Mockup - Select Criteria Page

SEGMA

Log out

- [Home](#)
- [Dataset](#)
- [Segmentation](#)
 - [Dataset Selection](#)
 - [Mapping Result](#)
 - [Segmentation Result](#)
- [Dashboard](#)
- [Report](#)
- [Feedback](#)

- [My Profile](#)

Customer Groups Result

← Back

Here is your customer grouping result:

Segment	No. of customer	Avg Orders	Top City
A	4233	3.3	KL, Malacca, Kelantan
B	1244	6.6	Johor

[Save Result](#)

[View Dashboard](#) ←

Figure 4.22 Mockup - Segmentation Result Page (Table)

SEGMA

Log out

- [Home](#)
- [Dataset](#)
- [Segmentation](#)
 - [Dataset Selection](#)
 - [Mapping Result](#)
 - [Segmentation Result](#)
- [Dashboard](#)
- [Report](#)
- [Feedback](#)

- [My Profile](#)

Dashboard

← Back

[Segment Again](#)
[View Report](#)

Number of segments
2

Total Customers
12031

Total Orders
2400

Total Sales
RM 50K

Sales vs Year

12,243
6,200
3,143
4,566
8,760

Segment A
Segment B

86%
14%

Figure 4.23 Mockup - Segmentation Result Page (Dashboard)

SEGMA

Log out

Home

Dataset

Segmentation ▾

Dataset Selection

Mapping Result

Segmentation Result

Dashboard

Report

Feedback

My Profile

Report History

Report	Date Generated	Customer Dataset
<input type="checkbox"/> Report_1	25/2/2024	Customer 2023.csv

Download

Figure 4.24 Mockup - Report Page

SEGMA

Log out

Home

Dataset

Segmentation ▾

Dataset Selection

Mapping Result

Segmentation Result

Dashboard

Report

Feedback

My Profile

Edit Profile Information



Username
CelineOng

Email
celineong1016@gmail.com

Reset Password

Delete Account &

Cancel

Save

Figure 4.25 Mockup - My Profile Page

SEGMA

Log out

[Home](#)
[Dataset](#)
[Segmentation ▾](#)
[Dataset Selection](#)
[Mapping Result](#)
[Segmentation Result](#)
[Dashboard](#)
[Report](#)
[Feedback](#)
[My Profile](#)

Feedback Form

Subject

Feedback

Send Feedback

Figure 4.26 Mockup - Feedback Page

SEGMA

Log out

[User](#)
[Feedback](#)
[My Profile](#)

Feedback

NO.	USERNAME	SUBJECT	DESCRIPTION	DATE SUBMITTED	STATUS	ACTION
1	CelineOng	Improve dataset page	Add on the search bar in dataset pages	6/16/2025, 1:03:42 PM	Solved	View Completed

Figure 4.27 Mockup - Admin Page

4.4 Hardware And Software Requirements

This section outlines the hardware and software requirements that will be used for developing and running the proposed web-based market segmentation system. The requirements are categorized into two parts which are hardware and software. Hardware requirements will cover the physical devices needed to operate and test the system while software requirements include the tools, programming languages, frameworks, libraries, and platforms used during system development.

4.4.1 Hardware Requirements

Hardware refers to the physical devices used to run and test the system. For this project, a standard laptop was used for the entire development process, including backend coding, frontend design, database handling, and model training.

Table 4.11 Device Details

Component	Specification Used
Device	Laptop
Processor (CPU)	Intel Core i7
RAM	8 GB
Storage	512 GB SSD
Internet Connection	Stable Wi-Fi connection
Display	15.6-inch screen with 1920×1080 resolution

This setup is sufficient for developing, testing, and running the system locally without the need for advanced computing hardware.

4.4.2 Software Requirements

Software includes all the programming tools, frameworks, and platforms used to build the system. The system was developed using both frontend and backend technologies, along with tools for data analysis and user interface design.

Table 4.12 Software Tools

Software/Tool	Purpose
Visual Studio Code	<ul style="list-style-type: none">Code editor used for writing and managing code for frontend and backend development.Supports extensions for Python, JavaScript, Tailwind, etc.
Hypertext Markup Language (HTML)	<ul style="list-style-type: none">Builds the structure of the web pages.
Cascading Style Sheets (CSS)	<ul style="list-style-type: none">Styles the layout and appearance of the web pages.
Tailwind CSS	<ul style="list-style-type: none">Utility-first CSS framework for fast and responsive UI styling.
JavaScript	<ul style="list-style-type: none">Frontend development to make the interface dynamic and responsive
React.js	<ul style="list-style-type: none">Build dynamic user interfaces
Node.js	<ul style="list-style-type: none">A JavaScript runtime for building scalable server-side applications.Handles backend logic like routing, data validation, and request handling.
Express.js	<ul style="list-style-type: none">Manages API endpoints, file uploads, and connects frontend to database
MongoDB	<ul style="list-style-type: none">NoSQL database to store user data, uploaded datasets, and segmentation results and so on.

Python	<ul style="list-style-type: none"> Used for segmentation logic and machine learning processing
Figma	<ul style="list-style-type: none"> To create visual mockups and interface design drafts
Power BI	<ul style="list-style-type: none"> For exploring and visualizing segmentation results

CONCLUSION

The project has successfully proposed a new, improved version of a market segmentation system for SME businesses. It addresses the limitations and challenges faced by SMEs when using existing segmentation systems, such as technical complexity, limited support for hybrid and a posteriori segmentation, and lack of report feature. With suggestions and feedback from the collaborator and SME stakeholder, the proposed system has been redefined into a version that fulfills real-world SME business needs and overcomes the identified limitations of existing systems. The new proposed system enables SME users to upload their customer and order datasets, choose from the attributes and perform automated segmentation using techniques from machine learning. The system is equipped with interactive dashboards, PDF reporting capabilities, and feedback submissions, all intended to increase the user experience and the user's decision-making abilities. It also closes the gaps found in the problem statement and problem faced by SMEs businesses when using the existing market segmentation tools. Overall, this project contributes a user-friendly, accessible, and insightful tool for SMEs to better understand their customers and make informed marketing decisions.

REFERENCES

- Akande, O., Asani, E. O., & Dautare, B. (2024). *Customer Segmentation Through RFM Analysis and K-Means Clustering: Leveraging Data-Driven Insights for Effective Marketing Strategy*. Ceddi Journal of Information System and Technology (JST), 3(1), 14–25.
<https://doi.org/10.56134/jst.v3i1.81>
- Al-Turjman, F., & Karaderi, A. (2024). *Customer Relationships Management (CRM) application for customer segmentation via RFM analysis and K-Means clustering*. In Advances in Science, Technology & Innovation (pp. 11–20).
https://doi.org/10.1007/978-3-031-63103-0_2
- Amplitude. (2025a). *Build charts in Amplitude: Add user segments*.
<https://amplitude.com/docs/analytics/charts/build-charts-add-user-segments>
- Amplitude. (2025b). *Usage reports: Understand how your organization uses Amplitude*.
<https://amplitude.com/docs/admin/billing-use/usage-reports>
- Amplitude. (2025c). *Product Analytics & event tracking platform*. <https://amplitude.com/>
- Cacciolatti, L., & Lee, S. H. (2015). *The role of Structured Marketing Information in SMEs' Decision-Making*. In Palgrave Macmillan UK eBooks (pp. 89–103).
https://doi.org/10.1057/9781137532589_6
- Cerquetti, M., & Romagnoli, A. (2023). Milieu and cultural heritage as a resource for digital marketing: Exploring web marketing strategies in the wine industry. *Corporate Governance and Organizational Behavior Review*, 7(3, special issue), 347–364.
<https://doi.org/10.22495/cgobrv7i3sip10>

Chelioudakis, D., Kalafati, F., Geramnis, E., & Matsatsinis, N. F. (2024). *A web-based multi-criteria decision support system for benchmarking marketing decisions alternatives*. *Operational Research*, 24(3). <https://doi.org/10.1007/s12351-024-00847-4>

Clarke, A. H., & Freytag, P. V. (2022). Implementation of new segments in small- and medium-sized enterprises (SMEs). *Journal of Business and Industrial Marketing*, 38(4), 930–942. <https://doi.org/10.1108/jbim-01-2021-0053>

Dolnicar, S. (2022). Market segmentation for e-Tourism. In *Springer eBooks* (pp. 849–863).
https://doi.org/10.1007/978-3-030-48652-5_53

Ehsani, F., & Hosseini, M. (2023). *Consumer Segmentation Based on Location and Timing Dimensions Using Big Data from Business-to-Customer Retailing Marketplaces*. *Big Data*. <https://doi.org/10.1089/big.2022.0307>

GA4. (2025). *Segment builder - Analytics Help*.
<https://support.google.com/analytics/answer/9304353>

Hotjar. (2025). *What Are Segments? Hotjar Documentation*.
<https://help.hotjar.com/hc/en-us/articles/360017762594-What-Are-Segments>

Insightly. (2025a). *Dynamic List Segmentation | Insightly CRM*.
<https://www.insightly.com/marketing/dynamic-list-segmentation/>

Insightly. (2025b). *Exporting data from Insightly to a CSV file - Insightly Help Center*.
https://support.insightly.com/en-us/Knowledge/article/1195/Exporting_data_from_Insightly_to_a_CSV_file

Ling, S. S., Too, C. W., Wong, W. Y., & Hoo, M. H. (2024). Customer Relationship Management System for Retail Stores Using Unsupervised Clustering Algorithms with RFM Modeling for Customer Segmentation. 2024 IEEE 14th Symposium on Computer Applications & Industrial Electronics (ISCAIE), 1–6. <https://doi.org/10.1109/iscaie61308.2024.10576353>

Matt, F. (2020). *Market segmentation analysis in the B2C health industry: A framework for determining the accurate target market of an international niche player based on a priori and a posteriori segmentation techniques* (Doctoral dissertation, FH Vorarlberg).

Moore, S.; Bulmer, S.; Elms, J. The Social Significance of AI in Retail on Customer Experience and Shopping Practices. *Journal of Retailing and Consumer Services* 2022, 64 (1), 102755. <https://doi.org/10.1016/j.jretconser.2021.102755>.

Olayinka, N. O. H. (2021). *Data driven customer segmentation and personalization strategies in modern business intelligence frameworks*. *World Journal of Advanced Research and Reviews*, 12(3), 711–726. <https://doi.org/10.30574/wjarr.2021.12.3.0658>

Omowole, N. B. M., Olufemi-Phillips, N. A. Q., Ofodile, N. O. C., Eyo-Udo, N. N. L., & Ewim, N. S. E. (2024). *Big data for SMEs: A review of utilization strategies for market analysis and customer insight*. *International Journal of Scholarly Research in Multidisciplinary Studies*, 5(2), 001–018. <https://doi.org/10.56781/ijsrms.2024.5.2.0044>

Osakwe, J., Shilongo, A., & Ziezo, M. (2023). *Optimising Customer segmentation in Digital Marketing Using Predictive Analytics: A Review of literature*. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4662191>

Osei, F., Ampomah, G., Kankam-Kwarteng, C., Bediako, D. O., & Mensah, R. (2021). *Customer Satisfaction Analysis of Banks: The role of Market Segmentation*. *Science Journal of Business and Management*, 9(2), 126. <https://doi.org/10.11648/j.sjbm.20210902.19>

Rio Luthfiandana, Gratia Atanka Barus, Nuraeni Nuraeni, Sri Yanthy, & Herni Pujiati. (2024). *Marketing Strategy: Segmenting, Targeting, and Positioning for Digital Business (Literature Review)*. Siber International Journal of Digital Business (SIJDB), 1(4), 11–22. <https://doi.org/10.38035/sijdb.v2i1.30>

Thomas, M. R., & George, G. (2021). Segmenting, targeting, and positioning (stp) of generational cohorts Y, Z and Alpha. IIMS Journal of Management Science, 12(2), 115–129. <https://doi.org/10.5958/0976-173x.2021.00008.7>

Torkzadeh, L., Jalilian, H., Zolfagharian, M., Torkzadeh, H., Bakhshi, M., & Khodayari-Zarnaq, R. (2021). *Market segmentation in the health tourism industry: a systematic review of approach and criteria*. Journal of Policy Research in Tourism Leisure and Events, 16(1), 69–88. <https://doi.org/10.1080/19407963.2021.1988622>

Tawil, A., Mohamed, M., Schmoor, X., Vlachos, K., & Haidar, D. (2023). *Trends and Challenges Towards an Effective Data-Driven Decision Making in UK SMEs: Case Studies and Lessons Learnt from the Analysis of 85 SMEs*. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2305.15454>

Yoseph, F., Malim, N. H. A. H., Heikkilä, M., Brezulianu, A., Geman, O., & Rostam, N. A. P. (2020). *The impact of big data market segmentation using data mining and clustering techniques*. Journal of Intelligent & Fuzzy Systems, 38(5), 6159–6173. <https://doi.org/10.3233/jifs-179698>

APPENDIX

Appendix A: Logbook

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY, UNIVERSITI MALAYA WIA3002/WIB3002: ACADEMIC PROJECT I PROJECT LOGBOOK				
No	Date&Time	Summary of Discussion	Platform F2F/ Online(e.g. gMeet, MsTeam, WhatsApp, Telegram, SocialMediaetc)	Supervisor Signature
1	Date: 26/3/2025 Wednesday Time: 9.00am - 9.30am	First meeting with my supervisor to inquire about suggestions for collaborators and stakeholders. Additionally, I sought ideas on how to start my project. Dr. Tutut also explained the steps I need to take, which include preparing letters for collaborator and stakeholders, proposing the report structure and logbook, and researching existing STP market segmentation systems.	Online Google Meet	 DR. TUTUT HERAWAN DEPT OF INFORMATION SYSTEMS FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA 50603 KUALA LUMPUR
2	Date: 9/4/2025 Thursday Time: 12.00pm - 1.00pm	Update the progress of collaborator and stakeholder issue, where I have found out two potential collaborators that were suggested by my supervisor. I also get feedback on literature reviews to improve it by adding on the comparison and summarization table.	Online Google Meet	 DR. TUTUT HERAWAN DEPT OF INFORMATION SYSTEMS FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA 50603 KUALA LUMPUR
3	Date: 18/4/2025 Friday Time: 9.00pm - 10.00pm	Brief the proposed system ideas to my supervisor and showing the comparison of existing systems with my proposed system. Recommendations were given by the supervisor to narrow the focus of the proposed system from segmentation, positioning and targeting(STP) to segmentation only.	Online Google Meet	 DR. TUTUT HERAWAN DEPT OF INFORMATION SYSTEMS FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA 50603 KUALA LUMPUR

4	Date: 1/5/2025 Thursday Time: 9.00am - 10.00am	Let my supervisor review my report on chapter 1 and chapter 2. A detailed discussion on the research methodology was also conducted. The supervisor provides feedback and recommendations to improve the structure of the report as well as to refine the selected methodological approach.	Online Google Meet	 DR. TUTUT HERAWAN DEPT OF INFORMATION SYSTEMS FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA 50603 KUALA LUMPUR
5	Date: 7/5/2025 Wednesday Time: 6.00pm - 7.00pm	We discussed Chapter 3 (Methodology) and addressed issues related to the meeting with the collaborator. The supervisor provided suggestions for improving my report and advised me to prepare questions for the collaborator to review and respond to.	Online Google Meet	 DR. TUTUT HERAWAN DEPT OF INFORMATION SYSTEMS FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA 50603 KUALA LUMPUR
6	Date: 15/5/2025 Wednesday Time: 12.30pm - 2.00pm	Discussion on methodology part.	Physical	 DR. TUTUT HERAWAN DEPT OF INFORMATION SYSTEMS FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA 50603 KUALA LUMPUR
7	Date: 16/5/2025 Thursday Time: 2.00pm - 3.00pm	Meeting with supervisor and collaborator to validate the content, structure, and relevance of the survey form.	Online Google Meet	 DR. TUTUT HERAWAN DEPT OF INFORMATION SYSTEMS FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA 50603 KUALA LUMPUR

Appendix B: Collaboration Letter

Collaborator: Dr Lucia Ika Fitriastuti

 UNIVERSITI
MALAYA

11/4/2025

Dr Lucia Ika Fitriastuti
Deputy Chairman
STIE Solusi Bisnis Indonesia
Jalan Ring Road Utara No. 17
Yogyakarta 55283, Indonesia

Sir/ Madam,

Collaborator for Academic Project Courses, Faculty of Computer Science and Information Technology, Universiti Malaya

With reference to the above matter,

2. The Faculty of Computer Science and Information Technology (FCSIT) University Malaya is pleased to invite you to be the collaborator for the project with title "Web-Based Model for Market Segmentation". This project is under the supervision of Associate Professor Dr. Tutut Herawan.

3. The Academic Project courses are mandatory for final year students of the Faculty of Computer Science and Information Technology as part of their graduation requirements. These courses are designed to nurture academically proficient and technically skilled graduates in the fields of Computer Science and Information Technology. During the course, students actively apply technical expertise, critical thinking, and problem-solving skills to develop innovative products. These projects are intended to meet collaborators' needs and enhance organizational efficiency.

4. This collaboration entails the involvement of the following student(s) from the Faculty of Computer Science and Information Technology working alongside you on the implementation of the aforementioned project. The faculty kindly requests your support in providing the student(s) with necessary information and/or other related resources to facilitate the successful completion of the project.

5. The following are the student(s) who will be involved in this collaboration

Name : Celine Ong Zi Ying
Matric Number : 22004878
Email : 22004878@siswa.um.edu.my

Faculty of Computer Science and Information Technology
Universiti Malaya, 50603 Kuala Lumpur, MALAYSIA
Tel (+603) 7967 6300/6301 • <http://fikrmum.um.edu.my>

6. We sincerely hope you will give thoughtful consideration to this invitation. We are confident that this collaboration will bring mutual benefits to both the faculty and your organization. Additionally, we kindly request your written consent to formalize your participation in this initiative. Should you require any further information, please do not hesitate to contact the project supervisor, **Associate Professor Dr. Tutut Herawan**, at tutut@um.edu.my.

Thank you.

Yours sincerely,



Professor Dr. Nor Liyana Mohd Shuib
Deputy Dean (Undergraduate)
Faculty of Computer Science and Information Technology
Universiti Malaya
Kuala Lumpur



Associate Professor Dr. Tutut Herawan
Project Supervisor
Faculty of Computer Science and Information Technology
Universiti Malaya
Kuala Lumpur

Faculty of Computer Science and Information Technology
Universiti Malaya, 50603 Kuala Lumpur, MALAYSIA
Tel (+603) 7967 6300/6301 • <http://fikrmum.um.edu.my>

Acceptance of Collaboration

I hereby acknowledge that I accept this proposal for collaboration with the Faculty of Computer Science & Information Technology, Universiti Malaya pertaining to the research project entitled **Web-Based Model for Market Segmentation**.

Collaborator signature



Dr Lucia Ika Fitriastuti
Deputy Chairman
STIE Solusi Bisnis Indonesia
Jalan Ring Road Utara No. 17
Yogyakarta 55283, Indonesia
Date:09/05/2025

Faculty of Computer Science and Information Technology
Universiti Malaya, 50603 Kuala Lumpur, MALAYSIA
Tel (+603) 7967 6300/6301 • <http://fikrmum.um.edu.my>

Stakeholder: Soapan Santun Enterprise

<p> UNIVERSITI MALAYA</p> <p>3/6/2025</p> <p>Soapan Santun Enterprise 28, Jalan Setia Permai U13/43D, Setia Alam, 40170 Shah Alam, Selangor</p> <p>Sir / Madam,</p> <p><u>Collaborator for Academic Project Courses, Faculty of Computer Science and Information Technology, Universiti Malaya</u></p> <p>With reference to the above matter,</p> <p>2. The Faculty of Computer Science and Information Technology (FCSIT) University Malaysia is pleased to invite you to be the collaborator for the project with title "Web-Based Model for Market Segmentation". This project is under the supervision of Associate Professor Dr. Tutut Herawan.</p> <p>3. The Academic Project courses are mandatory for final year students of the Faculty of Computer Science and Information Technology as part of their graduation requirements. These courses are designed to nurture academically proficient and technically skilled graduates in the fields of Computer Science and Information Technology. During the course, students actively apply technical expertise, critical thinking, and problem-solving skills to develop innovative products. These projects are intended to meet collaborators' needs and enhance organizational efficiency.</p> <p>4. This collaboration entails the involvement of the following student(s) from the Faculty of Computer Science and Information Technology working alongside you on the implementation of the aforementioned project. The faculty kindly requests your support in providing the student(s) with necessary information and/or other related resources to facilitate the successful completion of the project.</p> <p>5. The following are the student(s) who will be involved in this collaboration</p> <table><tr><td>Name : Celine Ong Zi Ying</td></tr><tr><td>Matric Number : 22004878</td></tr><tr><td>Email : 22004878@siswa.um.edu.my</td></tr></table> <p>6. We sincerely hope you will give thoughtful consideration to this invitation. We are confident that this collaboration will bring mutual benefits to both the faculty and your organization.</p> <p><small>Faculty of Computer Science and Information Technology Universiti Malaya, 50603 Kuala Lumpur, MALAYSIA Tel (+603) 7967 6300/6301 • http://faktm.um.edu.my</small></p>	Name : Celine Ong Zi Ying	Matric Number : 22004878	Email : 22004878@siswa.um.edu.my	<p>Additionally, we kindly request your written consent to formalize your participation in this initiative. Should you require any further information, please do not hesitate to contact the project supervisor, Associate Professor Dr. Tutut Herawan, at tutut@um.edu.my.</p> <p>Thank you.</p> <p>Yours sincerely,</p> <p></p> <p>Professor Dr. Nor Liyana Mohd Shuib Deputy Dean (Undergraduate) Faculty of Computer Science and Information Technology Universiti Malaya Kuala Lumpur</p> <p></p> <p>Associate Professor Dr. Tutut Herawan Project Supervisor Faculty of Computer Science and Information Technology Universiti Malaya Kuala Lumpur</p> <p><small>Faculty of Computer Science and Information Technology Universiti Malaya, 50603 Kuala Lumpur, MALAYSIA Tel (+603) 7967 6300/6301 • http://faktm.um.edu.my</small></p>
Name : Celine Ong Zi Ying				
Matric Number : 22004878				
Email : 22004878@siswa.um.edu.my				

<p>Acceptance of Collaboration</p> <p>I hereby acknowledge that I accept this proposal for collaboration with the Faculty of Computer Science & Information Technology, Universiti Malaya pertaining to the research project entitled Web-Based Model for Market Segmentation.</p> <p>Collaborator signature</p> <p></p> <p>Soapan Santun Enterprise 28, Jalan Setia Permai U13/43D, Setia Alam, 40170 Shah Alam, Selangor</p> <p>Date: 4/6/2025</p> <p><small>Faculty of Computer Science and Information Technology Universiti Malaya, 50603 Kuala Lumpur, MALAYSIA Tel (+603) 7967 6300/6301 • http://faktm.um.edu.my</small></p>

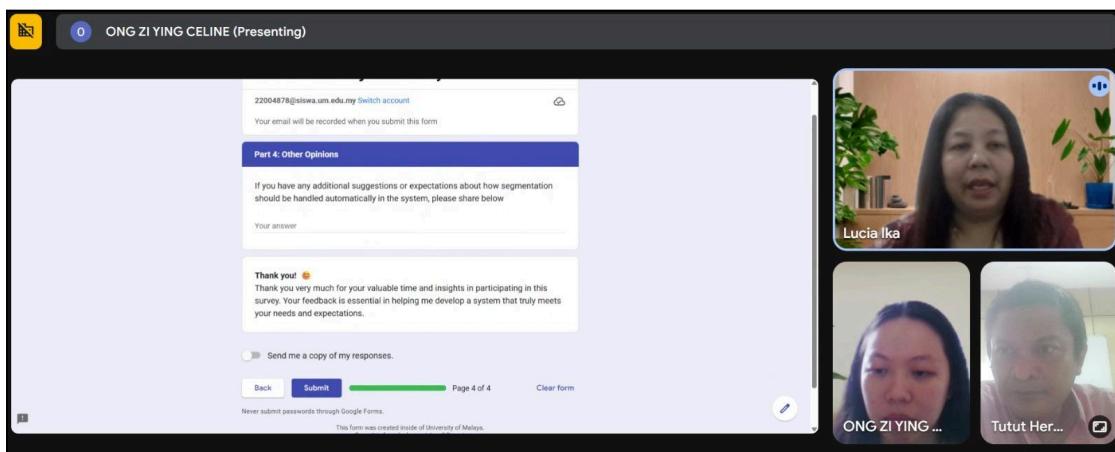
Appendix C: Meeting with Collaborator

Details of Meeting Arrangement:

Date: 16/5/2025 Thursday

Time: 2.00pm

People invited: Dr Lucia Ika Fitriastuti

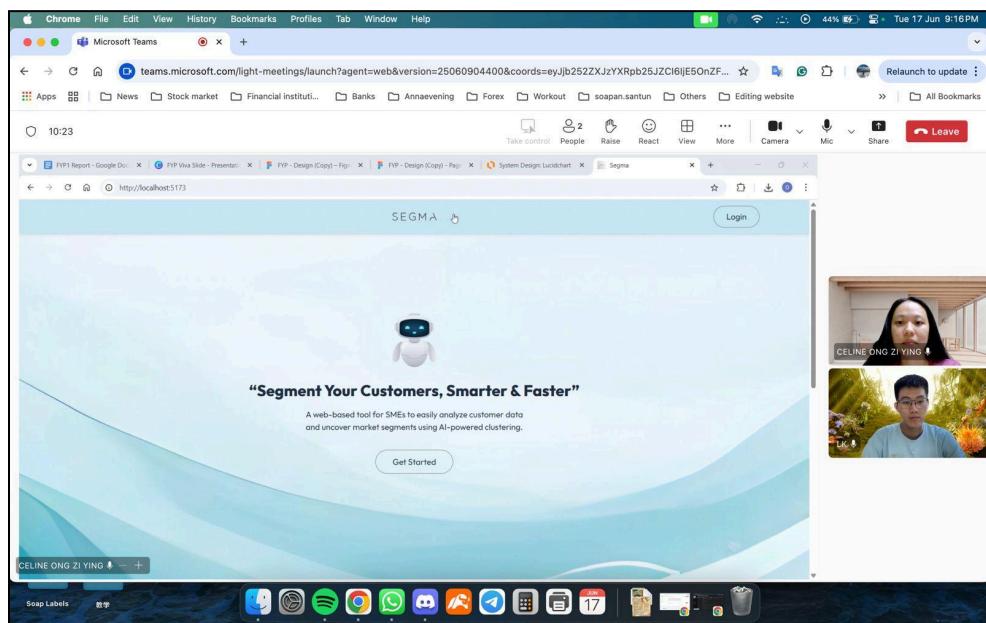


Details of Meeting Arrangement:

Date: 17/6/2025 Thursday

Time: 3.00pm

People invited: Soapan Santun Enterprise



Appendix D: Questionnaire validation form

Survey Link: <https://forms.gle/RCbHXnvjGzQKhMVA>

Part 1: Demographic

Part 1: Demographic	
Industry Type *	
<input type="radio"/> Retail	
<input type="radio"/> Telecommunications	
<input type="radio"/> Food and Beverages	
<input type="radio"/> E-commerce	
<input type="radio"/> Other:	_____
Company Size *	
<input type="radio"/> Micro (1–10 employees)	
<input type="radio"/> Small (11–50 employees)	
<input type="radio"/> Medium (51–200 employees)	
<input type="radio"/> Large (201+ employees)	

Part 2: Knowledge of Market Segmentation System

Does your business currently perform customer segmentation? *
<input type="radio"/> Yes
<input type="radio"/> No
If your previous answer is "No", why *
<input type="checkbox"/> Too expensive or not within budget
<input type="checkbox"/> Too complex to use without expert knowledge
<input type="checkbox"/> No awareness of such systems
<input type="checkbox"/> We rely on manual segmentation or spreadsheets
<input type="checkbox"/> NA
<input type="checkbox"/> Other: _____

Which web-based market segmentation tool(s) are you currently using or have used in your business? *

- Amplitude
- Hotjar
- Google Analytics
- Insightly
- HubSpot
- None
- Other: _____

What are the challenges that you faced when you using the current system? *

- High subscription or implementation costs
- Platform includes excessive features beyond segmentation needs
- Requires specialized expertise or external consultants
- Primarily supports real-time tracking only
- Segmentation is limited to manual, rule-based filtering
- Inability to access or review historical segmentation results
- NA
- Other: _____

Part 3: Ideas on Improved Version Web-based Market Segmentation Model

How would your organization prefer to provide customer data to the segmentation system? *

- Via CSV file upload (manually through the system interface)
- Via API integration (real-time data transfer from your existing database/system/third-party)
- Other: _____

If the system allows you to upload customer datasets, would you like to have the * option to delete any uploaded dataset at your discretion?

- Yes
- No

Which customer segmentation categories are most relevant to your company's * marketing strategy?

- Demographic
- Behavioral
- Geographic
- ALL
- Other: _____

Would you expect to select customer attributes (e.g., age, location, purchase history) before the system performs segmentation? *

- Yes
- No

Do you want the segmentation results to be shown in a visual and interactive dashboard? *

- Yes
- No

If interactive dashboard function is provided, how would an interactive dashboard * help you in understanding your customer segments better?

- Makes it easier to interpret complex data visually
- Helps identify trends or patterns quickly
- Saves time compared to reading tables or raw data
- Allows me to filter and focus on specific segments
- Other: _____

Do you want the system to automatically generate the PDF reports that * summarize the customer segmentation group result for you to download it for your own reference?

- Yes
- No

What key information or sections would you expect to see in the PDF reports? *

- The chosen segmentation criteria
- Segmentation mapping result
- Dashboard result
- Summary insights
- Other: _____

Do you think it is important to have a **Review history feature** to access previously generated PDF reports? *

- Yes
- No

Do you think **Review history feature** can help you with future marketing or business decisions? *

- Yes
- No

Part 4: Other Opinion

If you have any additional suggestions or expectations about how segmentation should be handled automatically in the system, please share below

Your answer
