**CUSTOMER SEGMENTATION RECENCY FREQUENCY MONETARY ANALYSIS**

**CHIU LINH PHAT**

**OPEN UNIVERSITY MALAYSIA**

**2023**

**CUSTOMER SEGMEMENTATION RECENCY FREQUENCY MONETARY ANALYSIS**

**CHIU LINH PHAT**

A Final Year Project submitted in fulfilment of the requirements

for the degree of

Bachelor of Information Technology

OPEN UNIVERSITY MALAYSIA

2023

**DECLARATION**

Name:

Matric Number:

I hereby declare that this final year project is the result of my own work, except for quotations and summaries which have been duly acknowledged.

Signature: Date:

**CUSTOMER SEGMEMENTATION RECENCY FREQUENCY MONETARY ANALYSIS**

**ABSTRACT**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Keywords:

(Not more than 5 words/phrases)

**ACKNOWLEDGEMENT**

I would like to take this opportunity to express my gratitude and appreciation to my supervisor, Mrs. Le Thi Ngoc Tho guidance, patience and invaluable advice throughout this project. I also would like to express my appreciation to my family and friends for them endless support whenever I face problems. Without the mentioned parties, it is impossible for me to complete this project report successfully.

THANK YOU.

Chiu Linh Phat

26 December ,2023

TABLE OF CONTENTS (Times New Roman, size 14, Upper Case)

TITLE PAGE (Times New Roman, size 12, Lower Case)

DECLARATION ii

ABSTRACT iii

ABSTRAK iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF TABLES vii

LIST OF FIGURES viii

LIST OF ABBREVIATIONS ix

CHAPTER 1 INTRODUCTION

1.1 Background to the Study 1

1.2 Problem Statement 1

1.3 Objectives of the Study 2

1.4 Research Questions and Hypotheses 3

1.5 Significance of the Study 5

1.6 Scope of the Study 7

1.7 Definitions of Terms 9

CHAPTER 2 REVIEW OF LITERATURE

2.1 Previous Studies

2.1.1 Project success criteria

2.2 Methodological Issues

2.3 Conceptual Framework

CHAPTER 5 DISCUSSION AND CONCLUSION

5.1 Summary of Main Findings

5.2 Discussion

5.3 Implications

5.4 Limitations of the Study

5.5 Directions for Future Research

REFERENCES

APPENDICES

A Title 100

B Title 105

Outline

CHAPTER 1: INTRODUCTION

Background to the study:

* Nguyên nhân dẫn đến việc cần thiết customer segemation
* Nêu ra một số thông tin về tại sao lại cần customer segemation

Problem Statement:

* Nêu ra về vấn đề gặp phải trong customer segmentation

Objectives of the Study:

* Nêu ra về lợi ích sau khi thực hiện quá trình customer segmentation.

Research Questions / Hypotheses:

* Đưa ra một số câu hỏi xoáy vào nhất

Scope and Limitation:

* Đưa ra các giới hạn của data set

Implementation Plan:

* outlines the step-by-step plan for executing the study.

Definition of Terms:

* This section provides clear definitions for terms specific to the study on customer segmentation.

CHAPTER 2: LITERATURE REVIEW

**CHAPTER 1: INTRODUCTION**

* 1. **Background to the Study:**

Information technology and the internet have developed strongly in recent years, creating favorable conditions for e-commerce to develop. Consumers can easily access e-commerce websites to search and shop for products. Businesses can also easily build and operate e-commerce websites to reach customers. In that context, compared to traditional trading methods that are costly but less effective, Typical sales pages are growing rapidly and some stores are moving towards this market with significant changes.

Consumers increasingly tend to shop online. They can shop anytime, anywhere, as long as they have an internet connection. They can also easily co mpare prices, product quality and services of different stores.

Customer segmentation is the process of dividing customers into subgroups based on their common characteristics. This allows businesses to better understand their customers and provide them with more relevant products, services and experiences.

This section establishes the context for the study, emphasizing the importance of understanding and categorizing customers based on their behavior, purchasing patterns, and value to the business. It highlights the evolution from manual segmentation by marketing and sales teams to the contemporary use of data mining and machine learning for efficient and targeted customer segmentation.

**1.2 Problem Statement:**

Here, the problem statement identifies challenges inherent in traditional manual methods of customer segmentation. It suggests that these methods may struggle with large datasets and may not efficiently identify nuanced patterns among diverse customer groups. The study aims to address these limitations by proposing the use of unsupervised learning algorithms as a solution for automating and improving the efficiency of customer segmentation.

**1.3 Objectives of the Study:**

This section outlines the specific goals of the research related to customer segmentation. Objectives include exploring the effectiveness of unsupervised learning algorithms in automating segmentation, assessing their application in targeted marketing and sales strategies, and identifying common patterns and characteristics among different customer segments.

**1.4 Research Questions / Hypotheses:**

Here, the research questions or hypotheses are tailored to the context of customer segmentation. Questions may revolve around the effectiveness of unsupervised learning algorithms for segmentation, their impact on targeted marketing efforts, and the types of patterns and traits that can be identified across different customer segments.

* How effective are unsupervised learning algorithms in automating customer segmentation?
* Can automated customer segmentation contribute to more targeted and effective marketing and sales efforts?
* What common patterns and traits among customer segments can be identified through machine learning techniques?

**1.5 Scope and Limitation:**

Scope:

Focus Area: The study concentrates on the application of unsupervised learning algorithms for customer segmentation.

Dataset: The analysis primarily uses the Kaggle dataset provided by Dr. Omar Romero-Hernandez. This dataset has 2,240 rows of observations and 28 columns of variables. Among the variables, there are 5-character variables and 23 numerical variables.

Limitation:

Data Quality: Acknowledges that the findings may be influenced by the quality and representativeness of the Kaggle dataset.

External Factors: Recognizes that external factors, such as evolving market dynamics, may affect the generalizability of the study's results.

1.6 Implementation Plan:

This part outlines the step-by-step plan for executing the study on customer segmentation. It includes preprocessing the data, selecting suitable unsupervised learning algorithms, training the models on the dataset, evaluating their performance, and discussing the practical application of the findings in targeted marketing and sales efforts.

**1.7 Definition of Terms:**

This section provides clear definitions for terms specific to the study on customer segmentation. Key terms may include "Customer Segmentation," "Unsupervised Learning," and others relevant to the context of the research.

**CHAPTER 2: LITERATURE REVIEW**

….

**CHAPTER 3: RESEARCH METHODOLOGY**

**3.1 Research Design**

The research design for implementing RFM analysis involves outlining the overall plan and strategy for conducting the analysis. It encompasses decisions about the type of study, the scope of the research, and the methods employed. In the context of RFM analysis:

**3.1.1 Type of Study**

Determine whether the research is exploratory, descriptive, or explanatory in nature. Consider if the focus is on understanding customer behavior, describing patterns, or explaining the relationships between RFM metrics and business outcomes.

**3.1.2 Scope of Research**

Define the specific objectives of the RFM analysis. This could include understanding customer segments, identifying opportunities for targeted marketing, or improving overall customer engagement.

**3.1.3 Sampling Strategy**

Decide on the sampling method to select a representative subset of data for analysis. Consider factors such as data availability, time constraints, and the desired level of precision in the results.

**3.1.4 Time Horizon**

Specify the time period covered by the analysis. Determine whether the focus is on recent data, historical trends, or a combination of both. The time horizon can impact the recency aspect of RFM metrics.

**3.2 Data Collection Method**

The success of RFM analysis heavily relies on the quality and completeness of the data collected. The data collection method involves obtaining transactional data, customer information, and other relevant data points.

**3.2.1 Data Sources**

Identify the sources of data, which may include transaction records, customer databases, and any additional relevant datasets. Ensure that the data sources are reliable, up-to-date, and representative of the target customer population.

**3.2.2 Data Collection Techniques**

Specify the techniques used to collect the necessary data. This could involve data extraction from databases, API integration, or any other method suitable for gathering transactional and customer-related information.

**3.2.3 Ethical Considerations**

Address ethical considerations related to data collection, including customer privacy and consent. Ensure compliance with relevant data protection regulations and standards.

**3.3 Data Analysis Method**

The data analysis method involves the application of RFM scoring and subsequent segmentation. It includes techniques for interpreting results and deriving actionable insights.

**3.3.1 RFM Scoring Algorithm**

Define the algorithm used to calculate RFM scores based on recency, frequency, and monetary value. This may involve assigning numerical values, establishing score ranges, or applying weighted scores based on business priorities.

**3.3.2 Segmentation Criteria**

Specify the criteria used to segment customers based on their RFM scores. Determine the number and characteristics of segments (e.g., Champions, Loyal Customers, Loyal Customers, Recent Customers, Promising, Need Attention, About to Sleep, At Risk), and outline how these segments align with business objectives.

**3.3.3 Statistical Methods**

If applicable, describe any statistical methods employed for validating the significance of the identified segments or assessing the reliability of the RFM analysis.

Detail how the results of the RFM analysis will be interpreted and translated into actionable insights. This may involve recommendations for targeted marketing strategies, personalized customer experiences, or operational improvements.