การแบ่งกลุ่มลูกค้าด้วยการวิเคราะห์อาร์เอฟเอ็ม กรณีศึกษาร้านขายสินค้าออนไลน์

CUSTOMER SEGMENTATION BY USING RFM ANALYSIS: CASE STUDY OF ONLINE SHOP

กิตติคุณ มิ่งมงคล

KITTIKUN MINGMONGKHOL

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์

ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2566

KMITL-2023-SC-M-017-023

CUSTOMER SEGMENTATION BY USING RFM ANALYSIS: CASE STUDY OF ONLINE SHOP

KITTIKUN MINGMONGKHOL

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE
IN DATA SCIENCE AND ANALYTICS
KMITL DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2023
KMITL-2023-SC-M-017-023

COPYRIGHT 2023
SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

หัวข้อการค้นคว้าอิสระ การแบ่งกลุ่มลูกค้าด้วยการวิเคราะห์อาร์เอฟเอ็ม กรณีศึกษาร้านขายสินค้าออนไลน์
ชื่อนักศึกษา กิตติคุณ มิ่งมงคล
รหัสประจำตัว 64605013
ปริญญา วิทยาศาสตรมหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์) ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
พ.ศ. 2566
อาจารที่ปรึกษาการค้นคว้าอิสระ ผู้ช่วยศาสตราจารย์ ดร.บุษยมาส พิมพ์พรรณชาติ

บทคัดย่อ

พฤติกรรมการซื้อสินค้าผ่านทางช่องทางออนไลน์มีการเพิ่มขึ้นอย่างต่อเนื่องในปัจจุบัน ส่งผลให้ผู้ขายต้องปรับตัวเข้ากับแนวโน้มที่เกิดขึ้นเพื่อเพิ่มโอกาสการขาย จนตลาดออนไลน์เกิดการแข่งขันที่สูง ดังนั้นผู้ขายจึงต้องปรับกลยุทธ์การตลาดเพื่อให้สามารถแข่งขันได้ งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์พฤติกรรมของลูกค้าจากข้อมูลการสั่งซื้อสินค้าบนแพลตฟอร์มออนไลน์ ด้วยวิธีวิเคราะห์ RFM ซึ่งเป็นเครื่องมือทางธุรกิจที่ใช้ในการวิเคราะห์ลูกค้าโดยใช้ข้อมูลสามปัจจัยหลักคือ ระยะเวลาที่ลูกค้าสั่งซื้อครั้งล่าสุด, จำนวนธุรกรรมในการสั่งซื้อ, และ ผลรวมของยอดใช้จ่ายจนถึงวันปัจจุบัน จากนั้นจะทำการแบ่งกลุ่มลูกค้าออกเป็นกลุ่มๆ ตามลักษณะพฤติกรรมที่แตกต่างกัน โดยในการแบ่งกลุ่มลูกค้าจะใช้วิธีการแบ่งกลุ่มด้วยการเรียนรู้ของเครื่องด้วยวิธีค่าเฉลี่ย k กลุ่ม วิเคราะห์ร่วมกับการแบ่งกลุ่มลูกค้าด้วยค่าคะแนนแบบดั้งเดิม ซึ่งจะช่วยให้วิเคราะห์กลุ่มลูกค้าได้ง่ายขึ้น เข้าใจถึงพฤติกรรมของลูกค้ากลุ่มย่อยในลูกค้ากลุ่มใหญ่ที่ได้จากการแบ่งกลุ่มด้วยวิธีค่าเฉลี่ย k กลุ่ม ผลการวิจัยสามารถแบ่งกลุ่มลูกค้าออกมาได้ทั้งหมด 5กลุ่ม ที่มีพฤติกรรมแตกต่างกัน ได้ดังนี้ กลุ่ม 1.Best Customer, 2.Loyal Big Spenders Customers, 3.At Risk, 4.Short-Term Inactive Customer, 5.Long-Term Inactive Customer การแบ่งกลุ่มลูกค้าเหล่านี้จะช่วยให้ธุรกิจเข้าใจพฤติกรรมลูกค้าของตัวเอง และสามารถปรับกลยุทธ์การตลาดให้เหมาะสมกับความต้องการของกลุ่มลูกค้าแต่ละกลุ่มที่แตกต่างกัน ช่วยรักษาความสัมพันธ์และสร้างความพึงพอใจของลูกค้าให้เพิ่มขึ้น นอกจากนี้ยังช่วยเพิ่มโอกาสการขายกระตุ้นการกลับมาซื้อซ้ำ เพิ่มยอดขายให้กับร้านค้าได้

คำสำคัญ : การเรียนรู้ของเครื่อง การแบ่งกลุ่มด้วยวิธีค่าเฉลี่ย k กลุ่ม การแบ่งกลุ่มลูกค้า การวิเคราะห์RFM

Independent Study Title	Customer Segmentation By Using RFM Analysis: Case Study Of Online Shop
Student Name	Kittikun Mingmongkhol
Student ID	64605013
Degree	Master of Science (Data Science and Analytics) Kmitl Digital Analytics and Intelligence Center
Year	2023
Independent Study Advisor	Asst.Prof.Dr.Busayamas Pimpunchat

Abstract

The trend of purchasing products through online channels has been steadily rising, forcing sellers to adapt and capitalize on these emerging trends in order to increase their sales opportunities. Consequently, the online market has become extremely competitive, compelling sellers to adjust their marketing strategies to stay competitive. This research aims to examine customer behavior by utilizing data from online purchases and employing the RFM analysis method. RFM is a business tool that assesses customers based on three key factors: recency (the time since the customer's last purchase), frequency (the number of purchases within a specified period), and monetary value (total spending). Subsequently, customers are classified into different segments based on their distinct behavioral characteristics. The K-means clustering technique is employed in conjunction with traditional RFM scoring to facilitate the segmentation process and gain a better understanding of subgroup behavior within larger customer groups derived from the k-means clustering method. The results enable the categorization of customers into five groups, each displaying different behavioral patterns: 1. Best customers, 2. Loyal big spenders’ customers, 3. At risk, 4. Short-Term Inactive Customers, and 5. Long-Term Inactive Customers. The segmentation of these customer groups enables businesses to comprehend their customers' behavior and tailor their marketing strategies according to the specific needs of each group. This approach helps maintain customer relationships,

enhance customer satisfaction, and ultimately increase sales opportunities by encouraging repeat purchases and driving sales for businesses.

Keywords : Customer Segmentation, K-Means Clustering, Machine learning, RFM analysis, RFM Scoring

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงไปด้วยดีด้วยความอนุเคราะห์ของบุคคลหลายท่าน ซึ่งไม่อาจจะนำมากล่าวได้ทั้งหมด ท่านแรกที่ผู้ทำการศึกษาใคร่ขอกราบขอบพระคุณคือ ผศ.ดร.บุษยมาส พิมพ์พรรณชาติ อาจารย์ที่ปรึกษาการค้นคว้าอิสระ ซึ่งท่านได้ให้ความรู้ คำแนะนำตรวจทาน ข้อคิดเห็นและแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเอาใจใส่ทุกขั้นตอนอันเป็นประโยชน์อย่างยิ่งในการทำวิจัย อีกทั้งยังช่วยแก้ปัญหาต่าง ๆ ที่เกิดขึ้นระหว่างการดำเนินงานอีกด้วย

ขอขอบคุณกรรมการในการสอบการค้นคว้าอิสระ ที่ชี้แนะและให้คำปรึกษาในจุดบกพร่องต่าง ๆ ที่เกิดขึ้น ทำให้โครงงานสำเร็จลุล่วงไปได้ด้วยดี

สุดท้ายนี้ขอขอบคุณเพื่อนๆ ในภาควิทยาการข้อมูลและการวิเคราะห์ สจล. ที่คอยให้ความช่วยเหลือและสนับสนุนในงานค้นคว้าอิสระฉบับนี้ และยังเป็นกำลังใจในการทำการค้นคว้าอิสระฉบับนี้จนสำเร็จลุล่วงไปได้ด้วยดี

นายกิตติคุณ มิ่งมงคล

สารบัญ

หน้า

บทคัดย่อภาษาไทย ข

บทคัดย่อภาษาอังกฤษ ค

กิตติกรรมประกาศ จ

สารบัญ ฉ

สารบัญตาราง ช

สารบัญรูป ซ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของงานวิจัย/ปัญหา 1

1.2 วัตถุประสงค์ของงานวิจัย 1

1.3 ขอบเขตของงานวิจัย 2

1.4 ประโยชน์ที่คาดว่าจะได้รับ 2

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 3

2.1 แนวคิดและทฤษฎีเกี่ยวข้องกับการวิเคราะห์ลูกค้า 3

2.1.1 การบริหารลูกค้าสัมพันธ์ 3

2.1.2 การแบ่งกลุ่มลูกค้า 3

2.1.3 มูลค่าตลอดช่วงชีวิตของลูกค้า 4

2.1.4 การวิเคราะห์ RFM 4

2.1.5 การจัดกลุ่มลูกค้าจากค่าคะแนน RFM แบบดั้งเดิม 4

2.2 แนวคิดและทฤษฎีเกี่ยวข้องกับการทำเหมืองข้อมูล 6

2.2.1 การจัดกลุ่ม 6

2.2.2 การจัดกลุ่มข้อมูลวิธีเฉลี่ย k กลุ่ม 6

2.2.3 ฟังก์ชันวัดระยะห่างของข้อมูล 7

2.2.4 การหาจำนวนกลุ่ม (k) ที่เหมาะสม 8

2.3 ข้อมูลทั่วไปเกี่ยวกับ Shopee 9

2.4 งานวิจัยที่เกี่ยวข้อง 9

สารบัญ (ต่อ)

บทที่ 3 วิธีการดำเนินงานวิจัย 11
3.1 ข้อมูลการซื้อขายสินค้า 11
3.2 เครื่องมือที่ใช้ในงานวิจัย 11
3.3 การตรวจสอบข้อมูล และ พิจารณาข้อมูลมาใช้ในการวิเคราะห์ 12
3.4 การแปลงข้อมูลเพื่อเตรียมทำการคำนวณค่า RFM 13
3.5 การจัดกลุ่มลูกค้าจากค่า RFM โดยใช้วิธี ค่าเฉลี่ย K กลุ่ม 16
3.6 การจัดกลุ่มลูกค้าจากค่าคะแนน RFM 17
3.7 สรุปขั้นตอนการดำเนินงาน 20
บทที่ 4 ผลการวิจัยและการอภิปรายผล 21
4.1 ผลการจัดกลุ่มลูกค้า 21
4.2 กำหนดกลยุทธ์ทางการตลาด 26
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 29
เอกสารอ้างอิง 31
ภาคผนวก 33
ภาคผนวก ก 34
ภาคผนวก ข 46
ประวัติผู้เขียน 50

สารบัญตาราง

ตารางที่

	หน้า
2.1 เกณฑ์คะแนน RFM ในการแบ่งกลุ่มลูกค้า	5
4.1 ค่าสถิติเบื้องต้นกลุ่มลูกค้าทั้ง 5 กลุ่ม	22
4.2 การรวมผลการจัดกลุ่มวิธีค่าเฉลี่ย k กลุ่มและวิธีค่าคะแนนแบบดั้งเดิมของกลุ่มที่ 0	27
4.3 การรวมผลการจัดกลุ่มวิธีค่าเฉลี่ย k กลุ่ม และวิธีค่าคะแนนแบบดั้งเดิมของกลุ่มที่ 1	28

สารบัญรูป

รูปที่	หน้า
2.1 การเลือกจำนวนกลุ่มด้วยวิธี Elbow Method	8
3.1 ตัวอย่างโค้ดนำเข้าไลบรารี่ที่จำเป็นและไฟล์ข้อมูล	12
3.2 ข้อมูลรายการคำสั่งซื้อ	12
3.3 ประเภทของข้อมูลของตัวแปร	12
3.4 จำนวนค่าว่างแต่ละคอลัมน์	13
3.5 การตัดแถวของข้อมูลที่มีค่าซ้ำกัน	13
3.6 เลือกข้อมูลเฉพาะสดมภ์ CustomerName, DateTime, OrderNo, Sales	14
3.7 กำหนดวันที่ล่าสุดในข้อมูลเป็น today_date	14
3.8 แสดงการคำนวณค่า RFM	15
3.9 เปลี่ยนชื่อสดมภ์ Recency, Frequency, Monetary	15
3.10 แผนภาพกล่องของค่า Recency, Frequency, Monetary	16
3.11 การทำข้อมูลเป็นมาตรฐานของค่า Recency, Frequency, Monetary	16
3.12 ค่า K ที่เหมาะสม โดยวิธี Elbow แสดงการนำจัดกลุ่มและกำหนดค่า k = 5	17
3.13 การแบ่งกลุ่มและกำหนดค่า k = 5	17
3.14 การแบ่งช่วงข้อมูลของ Recency	17
3.15 ค่า Frequency แต่ละค่าเปอร์เซนไทล์	18
3.16 การแบ่งค่า Frequency ออกเป็น 5 ช่วง	18
3.17 การแบ่งช่วงของ Monetary ออกเป็น 5 ช่วง	19
3.18 การแบ่งกลุ่มลูกค้าตาม Traditional RFM scoring	19
3.19 ขั้นตอนการดำเนินการ	20
4.1 กราฟสามมิติแสดงการกระจายตัวของแต่ละกลุ่มลูกค้าจากวิธีค่าเฉลี่ย k กลุ่ม	21
4.2 แผนภาพกล่องค่า R, F และ M ของกลุ่มที่ 0	22
4.3 แผนภาพกล่องค่า R, F และ M ของกลุ่มที่ 1	22
4.4 แผนภาพกล่องค่า R, F และ M ของกลุ่มที่ 2	23
4.5 แผนภาพกล่องค่า R, F และ M ของกลุ่มที่ 3	23
4.6 แผนภาพกล่องค่า R, F และ M ของกลุ่มที่ 4	24
4.7 แผนภูมิต้นไม้แสดงสัดส่วนของกลุ่มลูกค้าทั้ง 5กลุ่ม	25

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันที่พฤติกรรมการซื้อสินค้าของผู้บริโภคมีการเปลี่ยนแปลงเพิ่มขึ้นในการซื้อสินค้าผ่านทางช่องทางออนไลน์นั้น สร้างความสำคัญในการปรับตัวของผู้ขายเพื่อเข้ากับแนวโน้มใหม่นี้ เนื่องจากการซื้อสินค้าออนไลน์เป็นที่นิยมและสะดวกสบายต่อผู้บริโภค ซึ่งส่งผลให้ผู้ขายจำนวนมากปรับตัวโดยการเปิดขยายช่องทางการขายตามแพลตฟอร์มออนไลน์เพิ่มขึ้น เป็นการตอบสนองต่อความต้องการของลูกค้าที่มีการสั่งซื้อผ่านช่องทางนี้มากขึ้น แต่เนื่องด้วยสินค้าส่วนใหญ่มักไม่ได้ถูกผูกขาดทางการค้า ส่งผลให้มีหลายร้านค้าที่จำหน่ายสินค้าแบบเดียวกันหลายร้าน ทำให้การแข่งขันในตลาดมีระดับสูงขึ้นเรื่อย ๆ ดังนั้นร้านค้าจึงต้องหาวิธีเพื่อให้รอดต่อไปในตลาดและมีโอกาสเพิ่มการขาย การมีกลยุทธ์ที่ดีกลายเป็นสิ่งสำคัญที่จะช่วยพาไปสู่ความสำเร็จตามเป้าหมายที่ตั้งไว้ได้ เพื่อทำให้สามารถเพิ่มโอกาสการขายให้กับร้านค้า จึงควรพัฒนาความสัมพันธ์กับลูกค้าและเข้าใจพฤติกรรมของพวกเขา เพื่อเสนอสินค้าหรือบริการที่ตอบโจทย์ต่อพฤติกรรมดังกล่าว ซึ่งเป็นสิ่งสำคัญที่ช่วยในการเพิ่มโอกาสการขายให้กับร้านค้า และเพื่อทำให้รายได้ของร้านค้าเพิ่มขึ้นในตลาดที่มีการแข่งขันสูง (Meta for Business, 2023 ; Mohammadreza. et al, 2019)

นอกจากนี้ในปัจจุบันหลายบริษัทได้เริ่มนำเอากระบวนการทางด้านวิทยาศาสตร์ข้อมูลเข้ามาช่วยในการพัฒนาธุรกิจตัวเอง ดังนั้นในงานวิจัยนี้จะนำมาช่วยให้เข้าใจถึงพฤติกรรมของลูกค้า โดยการนำเอาข้อมูลการสั่งซื้อสินค้า มาผ่านกระบวนการวิเคราะห์ RFM แล้วแบ่งลูกค้าออกมาเป็นกลุ่มๆ โดยจะทำการศึกษาการใช้การเรียนรู้ของเครื่องด้วยวิธีเฉลี่ย K กลุ่ม ในการจัดกลุ่มลูกค้า วิเคราะห์ร่วมกับวิธีการแบ่งกลุ่มจากคะแนน RFM แบบดั้งเดิม (Traditional RFM Scoring) ลูกค้าที่มีพฤติกรรมใกล้เคียงกันหรือเหมือนกันจะถูกจัดไว้ในกลุ่มเดียวกัน ส่วนลูกค้าที่มีพฤติกรรมแตกต่างกันจะถูกจัดไว้คนละกลุ่ม ผลที่ได้จากการวิเคราะห์นี้จะช่วยให้เข้าใจลูกค้าและทำการตลาดได้ตรงกลุ่มมากขึ้น และเป็นการกระตุ้นให้ลูกค้ากลับมาซื้อซ้ำ โดยไม่ต้องเสียงบการตลาดที่ไม่จำเป็น ดังนั้นการพัฒนาความสัมพันธ์กับลูกค้า การเข้าใจพฤติกรรมและเสนอสินค้าหรือบริการที่ตรงต่อพฤติกรรมเป็นสิ่งสำคัญในการเพิ่มโอกาสการขายและความสำเร็จในธุรกิจที่ตั้งไว้ (Think with Google, 2023)

1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 นำเสนอวิธีการแบ่งกลุ่มลูกค้าโดยใช้การเรียนรู้เรื่องเครื่องด้วยวิธีค่าเฉลี่ย k กลุ่ม วิเคราะห์กลุ่มลูกค้าร่วมกับการแบ่งกลุ่มด้วยค่าคะแนนอาร์เอฟเอ็มแบบดั้งเดิม

1.2.2 เป็นแนวทางการวิเคราะห์พฤติกรรมการซื้อสินค้าและแบ่งกลุ่มลูกค้า โดยใช้เทคนิคอาร์เอฟเอ็ม

1.2.3 เป็นแนวทางในการนำผลลัพธ์จากการแบ่งกลุ่มลูกค้า มากำหนดกลยุทธ์ทางการตลาดให้เหมาะสมกับลูกค้าแต่ละกลุ่มให้กับร้านค้า

1.3 ขอบเขตของงานวิจัย

1.3.1 ข้อมูลการซื้อขายสินค้าของร้านขายสินค้าออนไลน์จากแพลตฟอร์มช้อปปี้ ที่เกิดขึ้นในช่วงวันที่ 10 กุมภาพันธ์ 2020 จนถึง 31 ธันวาคม 2022 เป็นรายวัน ประกอบด้วย 7 ตัวแปร มีจำนวนข้อมูลทั้งหมด 88,238 แถว

1.3.2 ตัวแปรที่ใช้ในงานวิจัย หมายเลขรหัสคำสั่งซื้อของลูกค้า, วันและเวลาที่ลูกค้าทำการสั่งซื้อ, ชื่อลูกค้า, สินค้า, ตัวเลือกสินค้า, จำนวนสินค้าที่ลูกค้าสั่งซื้อ และ ราคาสินค้า

1.3.3 ในการแบ่งกลุ่มลูกค้าจะใช้การเรียนรู้ของเครื่องวิธีค่าเฉลี่ย k กลุ่ม และ การแบ่งกลุ่มด้วยค่าคะแนนอาร์เอฟเอ็มแบบดั้งเดิม มาวิเคราะห์ร่วมกัน

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 เพื่อเป็นแนวทางในการศึกษาพฤติกรรมการสั่งซื้อสินค้าจากข้อมูลการขายโดยใช้เทคนิคอาร์เอฟเอ็ม

1.4.2 สามารถแบ่งกลุ่มลูกค้าและปรับกลยุทธ์ที่เหมาะสมกับลูกค้าแต่ละกลุ่มที่มีพฤติกรรมการสั่งซื้อที่แตกต่างกันได้

1.4.3 สามารถรู้ถึงกลุ่มลูกค้าที่สำคัญและคุ้มค่าต่อการทำกลยุทธ์การตลาด ตลอดจนช่วยลดค่าใช้จ่ายทางการตลาดที่ไม่จำเป็น

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาวิจัยนี้ ผู้วิจัยได้ทำการทบทวน ศึกษา ค้นคว้าแนวคิดทฤษฎี และงานวิจัยที่มีความสัมพันธ์ด้วยเนื้อหา หรือบทความทางวิชาการที่เผยแพร่ต่อสาธารณะ โดยมีทฤษฎีที่มีความสัมพันธ์กับงานวิจัย เพื่อนำมาเป็นแนวทางการทำวิจัยนี้ เอกสารและงานวิจัยที่เกี่ยวข้อง โดยกำหนดประเด็นศึกษา แบ่งเป็นหัวข้อดังนี้

2.1 แนวคิดและทฤษฎีเกี่ยวข้องกับการวิเคราะห์ลูกค้า

2.1.1 การบริหารลูกค้าสัมพันธ์ (Customer Relationship Management : CRM)

คือการบริหารความสัมพันธ์กับลูกค้า เป็นกลยุทธ์ที่ใช้ในการสร้างความสัมพันธ์กับลูกค้าให้ยั่งยืนยาวนานทั้งกับกลุ่มลูกค้าปัจจุบัน และผู้ที่มีโอกาสเป็นลูกค้า โดยมีเป้าหมายเพื่อสร้างความพึงพอใจของลูกค้าโดยการนำเสนอสินค้าหรือบริการ ไปจนถึงการทำให้ลูกค้ากลับมาซื้อซ้ำ โดยเน้นที่ความต้องการของลูกค้ามาเป็นอันดับหนึ่ง ทำความเข้าใจกลุ่มลูกค้าเป้าหมายหรือลูกค้าว่าต้องการอะไร เพื่อเพิ่มประสิทธิภาพในการขายให้มากกว่าเดิม และเน้นให้ลูกค้าเก่าเกิดความสัมพันธ์ที่ดีต่อธุรกิจ ผ่านการเก็บข้อมูลและพฤติกรรมของลูกค้าแล้วนำมาปรับใช้วางแผนการตลาด จนทำให้เกิดความภักดีต่อแบรนด์ (K. Tsiptsis, 2009)

2.1.2 การแบ่งกลุ่มลูกค้า (Customer Segmentation)

เมื่อบริษัทมีข้อมูลลูกค้าเป็นปริมาณมาก ดังนั้นการพิจารณาลูกค้าทีละรายคงเป็นเรื่องที่ก็เป็นเรื่องที่ยากและใช้เวลานาน ดังนั้นเพื่อให้บริษัทเข้าใจลูกค้าแต่ละรายและเข้าใจภาพรวมได้ง่ายมากขึ้น จึงมีความจำเป็นต้องแบ่งกลุ่มลูกค้า การที่จัดกลุ่มลูกค้าออกเป็นกลุ่มๆ โดยแต่ละกลุ่มจะมีความต้องการแตกต่างกัน และจะมีพฤติกรรมเฉพาะตัว โดยการจัดกลุ่มเป็นเทคนิคที่นักการตลาดนิยมใช้จัดการลูกค้าตามลักษณะของพฤติกรรมที่คล้ายๆกันให้อยู่กลุ่มเดียวกัน เช่น เพศ อายุ ที่อยู่ พฤติกรรมการซื้อ เป็นต้น เพื่อที่จะวางแผนหรือออกแบบการมีปฏิสัมพันธ์กับลูกค้าให้ง่ายขึ้น เพราะมีความชอบและความต้องการที่คล้ายคลึงกัน ซึ่งวิธีที่นิยมในการนำมาแบ่งกลุ่มลูกค้าคือ การทำ Clustering เทคนิคที่นิยมใช้ก็คือ K-Mean Clustering, DBSCAN เป็นต้น (Thanachart, 2020)

2.1.3 มูลค่าตลอดช่วงชีวิตของลูกค้า (Customer Lifetime Value)

เป็นการประเมินมูลค่าของลูกค้าออกมาเป็นตัวเลข คือมูลค่าที่คาดว่าจะได้รับจากลูกค้าแต่ละรายตั้งแต่ลูกค้าเริ่มเป็นลูกค้าจนกระทั่งเลิกเป็นลูกค้า ช่วยให้นักการตลาดสามารถประเมินความคุ้มค่าในการใช้จ่ายสำหรับการบริหารจัดการลูกค้า เช่นการใช้งบในการรักษาลูกค้าไว้ ในบางรายอาจจะไม่คุ้มค่าที่จะลงทุน ซึ่งวิธีที่นิยมใช้ในการคำนวณ ได้แก่ RFM Analysis , Regression Model , BSJ 2008 Model เป็นต้น (Thanachart, 2019)

2.1.4 การวิเคราะห์ RFM (RFM Analysis)

เป็นการแบ่งกลุ่มลูกค้าโดยการนำพฤติกรรมการซื้อมาวิเคราะห์ เพื่อแบ่งกลุ่มของลูกค้าออกเป็นกลุ่มๆ โดยขึ้นกับปัจจัย 3 อย่าง คือ 1. ระยะเวลาที่ลูกค้ามาใช้บริการครั้งสุดท้าย (Recency) หมายถึง จำนวนวันนับตั้งแต่ลูกค้าหรือกลุ่มเป้าหมายเกิดการกระทำครั้งล่าสุดขึ้น เช่น ซื้อสินค้าครั้งล่าสุด 2. ความถี่ในการซื้อ (Frequency) หมายถึง ความถี่หรือจำนวนครั้งที่เกิดการกระทำ ในช่วงระยะเวลาหนึ่งๆ เช่น ความถี่ในการซื้อสินค้า เป็นต้น 3.จำนวนเงินทั้งหมดที่ลูกค้าใช้จ่าย (Monetary) หมายถึง จำนวนเงินที่ลูกค้าจ่ายตลอดอายุการเป็นลูกค้า

วิธีการคำนวณค่า RFM ของลูกค้าแต่ละคน สามารถคำนวณได้ดังนี้

Recency = ระยะเวลาที่ลูกค้าสั่งซื้อครั้งสุดท้ายคิดจาก วันที่ล่าสุดในข้อมูล - วันที่ซื้อครั้งล่าสุด

Frequency = จำนวนธุรกรรมที่ลูกค้ามีการสั่งซื้อ

Monetary = ผมรวมจำนวนเงินที่ลูกค้าจ่ายให้ร้านค้าตั้งแต่รายการซื้อแรกจนถึงวันปัจจุบัน

ประโยชน์จากการวิเคราะห์ RFM จะช่วยให้เข้าใจลูกค้าว่าคนใดเป็นลูกค้าที่มีคุณค่า สร้างรายได้ให้กับร้านค้า หรือใครเป็นลูกค้าประจำที่ซื้อสินค้าบ่อย ลูกค้าคนใดมีโอกาสที่จะเลิกใช้บริการ ลูกค้าคนใดมีโอกาสเป็นลูกค้าที่สนใจสินค้าตัวใหม่ เป็นต้น (Ananthi, 2018)

2.1.5 การจัดกลุ่มลูกค้าจากค่าคะแนน RFM (Traditional RFM Scoring)

การแบ่งกลุ่มจากคะแนนจากการให้คะแนน RFM นั้น มีขั้นตอนดังนี้

1. คำนวณค่า Recency, Frequency และ Monetary ของลูกค้าแต่ละราย

2. ทำการแบ่งข้อมูล Recency ของลูกค้าทั้งหมดเป็น 5 ช่วง โดยที่กลุ่มลูกค้าที่มีการซื้อสินค้าไปไม่นานจะเป็นกลุ่มที่ได้คะแนนสูงที่สุดคือ 5 คะแนน และลดลงตามลำดับ กลุ่มที่ระยะเวลาสั่งซื้อนานที่สุดจะได้คะแนน Recency เท่ากับ 1 คะแนน

3. ทำการแบ่งข้อมูล Frequency ของลูกค้าทั้งหมดเป็น 5 ช่วง โดยที่กลุ่มลูกค้าที่มีการซื้อสินค้าบ่อยมากที่สุดจะเป็นกลุ่มที่ได้คะแนนสูงที่สุดคือ 5 คะแนน และลดลงตามลำดับ กลุ่มที่ความถี่ในการซื้อสินค้าต่ำที่สุดจะได้คะแนน Frequency เท่ากับ 1 คะแนน

4. ทำการแบ่งข้อมูล Monetary ของลูกค้าทั้งหมดเป็น 5 ช่วง โดยที่กลุ่มลูกค้าที่มีการซื้อสินค้าบ่อยมากที่สุดจะเป็นกลุ่มที่ได้คะแนนสูงที่สุดคือ 5 คะแนน และลดลงตามลำดับ กลุ่มที่ความถี่ในการซื้อสินค้าต่ำที่สุดจะได้คะแนน Frequency เท่ากับ 1 คะแนน

5. ทำการรวมคะแนนโดยการเรียงคะแนนของ Recency , Frequency , Monetary ตามลำดับ เช่น คะแนน R=5, F=5, M=5 ลูกค้าคนนั้นจะมีคะแนนรวมเท่ากับ 555 คะแนน โดยเกณฑ์ในการแบ่งกลุ่มคะแนนที่ได้สามารถแบ่งได้ 11 กลุ่ม มีดังนี้ (Connectif, 2020)

ตารางที่ 2.1 เกณฑ์คะแนน RFM ในการแบ่งกลุ่มลูกค้า

ชื่อกลุ่มลูกค้า	**RFM SCORE**
Champions	555, 554, 544, 545, 454, 455, 445
Loyal Customers	543, 444, 435, 355, 354, 345, 344, 335
Potential Loyalist	553, 551, 552, 541, 542, 533, 532, 531, 452, 451, 442, 441, 431, 453, 433, 432, 423, 353, 352, 351, 342, 341, 333, 323
New Customers	512, 511, 422, 421 412, 411, 311.
Promising	525, 524, 523, 522, 521, 515, 514, 513, 425,424, 413,414,415, 315, 314, 313
Need Attention	535, 534, 443, 434, 343, 334, 325, 324
About To Sleep	331, 321, 312, 221, 213
At Risk	255, 254, 245, 244, 253, 252, 243, 242, 235, 234, 225, 224, 153, 152, 145, 143, 142, 135, 134, 133, 125, 124
Can’t Lose Them	155, 154, 144, 214,215,115, 114, 113
Hibernating	332, 322, 231, 241, 251, 233, 232, 223, 222, 132, 123, 122, 212, 211
Lost	111, 112, 121, 131, 141, 151

6. ทำการตั้งชื่อกลุ่มลูกค้าให้ลูกค้าแต่ละรายตามเกณฑ์คะแนน จะได้กลุ่มลูกค้าที่มีลักษณะพฤติกรรมคล้ายๆ กันอยู่ในกลุ่มเดียวกัน และสามารถที่จะกำหนดนโยบายการตลาดให้แต่ละกลุ่มต่อไป

2.2 แนวคิดและทฤษฎีเกี่ยวข้องกับการทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining) คือ กระบวนการค้นหาความสัมพันธ์ รูปแบบและแนวโน้มใหม่ๆ โดยใช้ข้อมูลจำนวนมากที่เก็บไว้ในคลังข้อมูล แล้วใช้วิธีการทางคณิตศาสตร์และสถิติในการวิเคราะห์ข้อมูล (สายชล, 2560)

2.2.1 การจัดกลุ่ม (Clustering)

คือการจัดกลุ่มข้อมูลที่ไม่เคยมีการจัดกลุ่มมาก่อนหน้า แต่จะแบ่งกลุ่ม (Cluster) โดยการรวบรวมข้อมูลที่คล้ายคลึงกันไว้ในกลุ่มเดียวกันมากที่สุดหรือมีความแตกต่างน้อยที่สุด และข้อมูลที่ต่างกันไว้ต่างกลุ่มกัน ซึ่งข้อมูลที่อยู่ต่างกลุ่มกันจะมีความคล้ายคลึงกันน้อยที่สุดหรือมีความแตกต่างกันมากที่สุด การจัดกลุ่มแตกต่างจากการจำแนกกลุ่ม (Classification) คือ การจัดกลุ่มไม่มีตัวแปรเป้าหมาย งานในการจัดกลุ่มไม่ได้ทำการจำแนกกลุ่ม ประมาณค่า หรือทำนายค่าของตัวแปรเป้าหมาย ขั้นตอนวิธีของการจัดกลุ่มเป็นการค้นหาเพื่อที่จะแบ่งชุดข้อมูลทั้งหมดออกเป็นกลุ่มๆ การจัดกลุ่มเป็นเทคนิคของการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ขั้นตอนวิธีของการจัดกลุ่มเป็นการค้นหาเพื่อที่จะแบ่งชุดข้อมูลออกเป็นกลุ่มๆ (สายชล, 2560)

ตัวอย่างของงานในการจัดกลุ่มในการวิจัยและธุรกิจ เช่น

1) การจัดกลุ่มลูกค้าจากพฤติกรรมการซื้อสินค้า

2) การตรวจสอบบัญชี แบ่งพฤติกรรมทางการเงินออกเป็นลูกค้าชั้นดีและไม่ดี

3) การจัดกลุ่มคนไข้ตามอาการหรือความรุนแรงของโรค เป็นต้น

การจัดกลุ่มแบ่งออกเป็น 2 ประเภท คือ

1) การจัดกลุ่มแบบเป็นขั้นตอน (Hierarchical Clustering Methods) โครงสร้างกลุ่มเหมือนต้นไม้ เรียกว่า เดนโดแกรม(Dendrogram) สร้างโดยใช้วิธีการจัดกลุ่มแบบรวมกัน (Agglomerative clustering method) หรือ วิธีการจัดกลุ่มแบบแยกกัน (Divisive clustering medthod) เช่น การจัดกลุ่มเชื่อมโยงแบบเดี่ยว (Single linkage Clustering) การจัดกลุ่มเชื่อมโยงแบบสมบูรณ์ (Complete linkage Clustering) การเชื่อมโยงแบบเฉลี่ย (Average linkage Clustering) วิธีการจัดกลุ่มของวอร์ด (Ward’s Method)

2) การจัดกลุ่มแบบไม่เป็นขั้นตอน (Non-Hierarchical Clustering Methods) จะต้องกำหนดเองว่าต้องการแบ่งเป็นกี่กลุ่ม เช่น k กลุ่ม เช่น วิธีเฉลี่ย k กลุ่ม (K-Means Clustering)

2.2.2 การจัดกลุ่มข้อมูลวิธีเฉลี่ย k กลุ่ม (K-Mean Clustering)

เป็นวิธีการหนึ่งในการทำเหมืองข้อมูล นับเป็นวิธีการที่นิยมใช้ เนื่องจากมีขั้นตอนการทำงานที่ไม่ซับซ้อน และเข้าใจได้ง่าย โดยมีหลักการคือ จำแนกข้อมูลออกเป็นกลุ่มย่อย จะใช้เมื่อมีจำนวนข้อมูลมาก ต้องกำหนดจำนวนกลุ่มหรือจำนวน Cluster ตามที่ต้องการ เช่น กำหนดจำนวนมี k กลุ่ม โดย K-Means จะทำงานหลายรอบ (Iteration) ซึ่งการทำงานแต่ละครั้งจะมีการรวมข้อมูลออกไปอยู่

ในกลุ่มใดกลุ่มหนึ่ง การเลือก Case ใช้คำนวณระยะห่างจากค่ากลางที่น้อยที่สุดของกลุ่ม โดยจะคำนวณหาค่ากลางซ้ำๆจนค่ากลางของกลุ่มคงที่หรือครบตามจำนวนที่กำหนด (สายชล, 2560)

ชนิดของตัวแปรที่ใช้ในเทคนิค (K-Mean Clustering) จำเป็นตัวแปรเชิงปริมาณ เพราะเป็นรูปแบบสเกลอันตรภาค (Interval Scale) หรือรูปแบบสเกลอัตราส่วน (Ration Scale) โดยจะไม่สามารถใช้ข้อมูลความถี่หรือแบบ Binary โดยมีขั้นตอนการทำงานดังนี้

ขั้นตอนที่ 1 เริ่มต้นจากการกำหนดค่า K หรือจำนวนกลุ่มข้อมูลที่ต้องการ
ขั้นตอนที่ 2 สุ่มวางตำแหน่งจุดศูนย์กลางของข้อมูลแต่ละกลุ่ม หรือเซ็นทรอยด์ (Centroid)
ขั้นตอนที่ 3 จากจุดศูนย์กลางแต่ละจุดจะมีการคำนวณระยะทางกับทุกข้อมูลในชุดข้อมูล ซึ่งคำนวณโดยวิธียูคลิเดียน จากนั้นแต่ละข้อมูลจะถูกจัดอยู่ในกลุ่มของจุดศูนย์กลางที่มี ระยะทางใกล้ที่สุด
ขั้นตอนที่ 4 หลังจากทำการจัดกลุ่มข้อมูลใหม่แล้วทำการคำนวณค่าเฉลี่ยของสมาชิกในกลุ่ม เพื่อกำหนดเป็นจุดศูนย์กลางของกลุ่มข้อมูลใหม่
ขั้นตอนที่ 5 ทำซ้ำข้อ 3 ถึง 4 จนกระทั่งค่าจุดศูนย์กลางของกลุ่มข้อมูลใหม่ได้ค่าไม่ต่างหรือต่างเพียงเล็กน้อยจากค่าจุดศูนย์กลางรอบก่อนหน้า

2.2.3 ฟังก์ชันวัดระยะห่างของข้อมูล (Dissimilarity Function)

ใช้ในการพิจารณาความเหมือน หรือความคล้ายกันของข้อมูลโดยฟังก์ชันที่นิยมใช้ในเทคนิคการจัดกลุ่มสามารถแบ่งออกได้เป็น 3 วิธี ดังนี้ (สายชล, 2560)

1.ระยะห่างยูคลิเดียน (Euclidean distance) เป็นฟังก์ชันที่นิยมใช้ในวิธีเฉลี่ย k กลุ่ม โดยมีสูตรการคำนวณดังนี้

$$d_{Euclidean}(x, y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2}$$

X คือข้อมูลใดๆที่ต้องการหาระยะห่างระหว่างข้อมูลกับจุดศูนย์กลางของกลุ่ม K

Y คือข้อมูลที่เป็นจุดศูนย์กลางของกลุ่ม K

m คือ จำนวนตัวแปร

2.ระยะห่างแมนฮัทตันหรือระยะห่างระหว่างเมืองในรูปบล็อก (Manhattan or City-block distance) เป็นฟังก์ชันวัดระยะห่างโดยพิจารณาความแตกต่างของมิติค่าเฉลี่ยโดยมีสูตรการคำนวณดังนี้

$$d_{Manhattan}(x,y) = \sum_i |(x_i - y_i)|$$

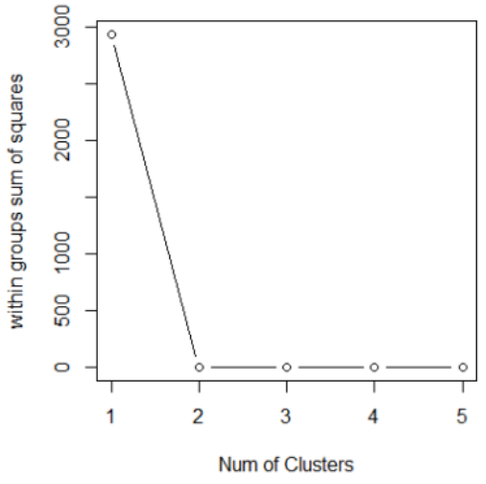
3.ระยะห่างมินโคสกี้ (Minkowski distance) เป็นกรณีทั่วไปของระยะห่างยูคลิเดียนและระยะห่างแมนฮัทตัน

$$d_{Minkowski}(x,y) = (\sum_i |(x_i - y_i)|^p)^{\frac{1}{p}}$$

โดยที่ p เป็นจำนวนเต็มบวก ถ้า p=1 จะได้ระยะห่างแมนฮัทตัน และถ้า p=2 จะได้ระยะห่างยูคลิเดียน

2.2.4 การหาจำนวนกลุ่ม (k) ที่เหมาะสม Elbow Method

เป็นการวัดค่าความคลาดเคลื่อนของผลรวมระยะห่าง ระหว่างข้อมูล กับจุดศูนย์กลางของกลุ่ม หรือเรียกว่า ผลรวมกำลังสองภายในกลุ่ม (Within Groups Sum of Squares: WGSS) ซึ่งการวนซ้ำแต่ละรอบจะทำให้ค่าของ WGSS ลดลงจากจำนวนกลุ่มที่เพิ่มขึ้นสมาชิกในแต่ละกลุ่มก็จะลดลง ค่าผลบวกของค่าคลาดเคลื่อนกำลังสอง (ESS) เป็นค่าที่ทำให้เส้นกราฟเกิดความโค้ง เมื่อวิเคราะห์เส้นกราฟแล้วจะทำให้ได้จำนวนกลุ่ม (k) ที่เหมาะสม คือจุดที่กราฟมีลักษณะหักโค้งมากที่สุด (Mulika H, 2020)



รูปที่ 2.1 การเลือกจำนวนกลุ่มด้วยวิธี Elbow Method

2.3 ข้อมูลทั่วไปเกี่ยวกับ Shopee

Shopee คือ แพลตฟอร์มอีมาร์เก็ตเพลส (E-Marketplace) ชนิดหนึ่งที่เป็นตลาดกลางในการซื้อขายสินค้าผ่านช่องทางอินเตอร์เน็ต ปัจจุบันเป็นแพลตฟอร์มอีคอมเมิร์ซชั้นนำในเอเชียตะวันออกเฉียงใต้และไต้หวัน เปิดตัวในปี 2015 เป็นแพลตฟอร์มที่ได้รับการออกแบบให้เหมาะกับภูมิภาค บริการของ Shopee สินค้าที่นำเสนอขายบนแพลตฟอร์มมีความหลากหลาย ตอบรับความต้องการผู้บริโภคในทุกๆไลฟ์สไตล์ ด้วยการมอบประสบการณ์การช้อปปิ้งออนไลน์ที่ปลอดภัยและรวดเร็ว ผ่านการชำระเงินและการขนส่งที่น่าเชื่อถือ (Shopee Thailand, 2022)

2.4 งานวิจัยที่เกี่ยวข้อง

Yash Parikh and Eman Abdelfattah (2020) ศึกษาอัลกอริทึมการจัดกลุ่มลูกค้า จากค่า Recency , Frequency และ Monetary (RFM Analysis) ที่อยู่ในรายการคำสั่งซื้อ เพื่อที่นำเสนอกลยุทธให้เหมาะสมกับพฤติกรรมการซื้อของลูกค้าแต่ละราย โดยใช้ทั้งหมด 4 เทคนิค ได้แก่ 1. Mean-shift, 2. DBSCAN, 3. Agglomerative Clustering, 4. K-Means โดยทำการเปรียบเทียบทั้ง 4 เทคนิค ผลที่ได้ทั้ง 4 เทคนิคสามารถจัดกลุ่มได้แตกต่างกันตั้งแต่ 2 – 6 กลุ่ม ซึ่งเทคนิค K-Means ใช้เวลาในการประมวลผลเร็วที่สุด และ ทั้ง4เทคนิคมีความสามารถในการระบุกลุ่มลูกค้าที่แตกต่างกัน ซึ่งจะช่วยให้การทำการตลาดได้ตรงกลุ่มกับลูกค้ามากขึ้น เช่น การรักษาให้ลูกค้ายังคงอยู่ หรือรักษาลูกค้าที่จงรักภักดี (Royal Customer) ให้มีความสุข

Ananthi Sheshasaayee and L. Logeshwari (2018) ศึกษาการจัดกลุ่มลูกค้าโดยใช้เทคนิคการทำเหมืองข้อมูลเนื่องจาก กระบวนการของการบริหารจัดการความสัมพันธ์ของลูกค้า ข้อมูลนี้จะช่วยสร้างความพึงพอใจและการรักษาลูกค้าให้ได้มากขึ้น แต่ทุกวันนี้อัลกอริทึมในการทำ CRM นั้นล้าสมัยและไม่เพียงพอเนื่องจากข้อมูลมีความผิดพลาดและปริมาณข้อมูลที่มีมากขึ้น ในงานวิจัยนี้จะทำการใช้เทคนิคการทำเหมืองข้อมูลในการแก้ปัญหาข้อมูลลูกค้า โดยใช้วิธีค่าเฉลี่ย k กลุ่ม ร่วมกับการวิเคราะห์พฤติกรรมของลูกค้าผ่านวิธี RFM ผลที่ได้สามารถแบ่งกลุ่มลูกค้าได้ทั้งหมด 2 กลุ่ม กลุ่มแรกจะเป็นกลุ่มลูกค้าที่มีการสั่งซื้อสินค้าบ่อย เป็นกลุ่มลูกค้าที่จงรักภักดี ส่วนกลุ่มที่ 2 มีการสั่งซื้อน้อยครั้ง ดังนั้นกลุ่มลูกค้าเป้าหมายคือลูกค้ากลุ่มที่ 2 เพราะต้องมีการโฆษณาเพื่อให้ลูกค้าจดจำได้ รวมไปถึงเสนอส่วนลดเพื่อให้ลูกค้ากลุ่มนี้กลับมาซื้อสินค้าและบรรลุเป้าหมายทางการตลาด

Sabbir Hossain Shihab, Shyla Afroge and Sadia Zaman Mishu (2019) ศึกษาการแบ่งกลุ่มลูกค้าด้วยวิธี RFM เป็นวิธีที่นิยมอย่างมากในปัจจุบัน เพื่อที่จะแบ่งลูกค้าจากพฤติกรรมการซื้อ นอกจากนี้มีการนำเทคนิคการจัดกลุ่มหลายอย่างมาใช้เพื่อแบ่งกลุ่มลูกค้าออกเป็นกลุ่มเพื่อให้ได้ผลลัพธ์ที่ดีขึ้น ในการวิจัยนี้จะวิเคราะห์เปรียบระหว่างวิธี Agglomerative, K-means และ Advance

K-means ตามแนวทางการวิเคราะห์ RFM ผลที่ได้แสดงให้เห็นถึงวิธี Agglomerative ต้องใช้เวลาในการประมวลผลที่นานสำหรับข้อมูลที่มีขนาดใหญ่ เมื่อเทียบกับวิธี K-means และ Advance K-Means อย่างไรก็ตามข้อดีของวิธี Agglomerative คือไม่ต้องมีการกำหนดกลุ่ม (ค่าK) ในตอนแรก ผลลัพธ์จากการจัดกลุ่มแสดงให้เห็นว่าวิธี Advance K-means สามารถลดเวลาการทำงานได้อย่างมีประสิทธิภาพได้ถึง 27.8% และ 97.8% เมื่อเทียบกับวิธี K-means และ Agglomerative ตามลำดับ และยังให้ผลลัพธ์การจัดกลุ่มที่ดีกว่า K-means ในแง่ของระยะห่างภายในกลุ่ม และระยะห่างระหว่างกลุ่ม ดังนั้นวิธี Advance K-means ที่ทางผู้วิจัยนำเสนอนั้นจะช่วยลดเวลาและสามารถทำงานได้ดีสำหรับข้อมูลขนาดใหญ่

Alexandrina-Mirela Pater, Stefan Vári-Kakas, Otto Poszet and Ionel Gabriel Pintea (2019) ได้กล่าวถึงกลยุทธ์ทางการตลาดที่มีประสิทธิภาพขึ้นอยู่กับการวิเคราะห์ข้อมูลจากประวัติการทำธุรกรรมอย่างเหมาะสม มีการวิเคราะห์ข้อมูลจำนวนมากด้วยวิธีการทำเหมืองข้อมูล ซึ่งช่วยสามารถแบ่งกลุ่มลูกค้าได้ตามเกณฑ์บางอย่าง ในงานวิจัยนี้ได้นำกระบวนการวิเคราะห์ที่นิยมมาใช้เป็นกรณีศึกษากับธุรกิจขนาดเล็ก เช่นร้านค้าออนไลน์ โดยนำวิธีค่า เฉลี่ย k กลุ่ม มาช่วยจัดกลุ่มลูกค้า ผ่านการวิเคราะห์พฤติกรรมด้วยเทคนิค RFM ผลที่ได้สามารถวิเคราะห์รูปแบบการซื้อของลูกค้าได้ แม้จะเป็นธุรกิจขนาดเล็กก็ตาม และยังสามารถทำการตลาดเช่นการโฆษณาหรือโปรโมชั่นตามพฤติกรรมการซื้อ เพื่อเพิ่มยอดขายและประหยัดงบของบริษัท

Mohammadreza Tavakoli, Mohammadreza Molavi, Vahid Masoumi, Majid Mobini, Sadegh Etemad and Rouhollah Rahmani (2018) ได้ศึกษาการทำ RFM เพื่อการแบ่งกลุ่มลูกค้าโดยคำนึงถึงการเปลี่ยนแปลงทางธุรกิจและใช้วิธีการจัดกลุ่มด้วยวิธีค่า เฉลี่ย k กลุ่ม ได้ทดลองใช้โมเดลที่พัฒนากับบริษัท Digikala ซึ่งเป็นร้านค้าออนไลน์ที่ใหญ่ที่สุดในภูมิภาคเอเชียตะวันออกกลาง และเปรียบเทียบผลลัพธ์ของโมเดลก่อนนำไปใช้และหลังนำไปใช้ นอกจากนี้ได้พัฒนากลยุทธ์สำหรับแต่ละกลุ่มลูกค้าและทำการแคมเปญส่งข้อความตามกลยุทธ์ดังกล่าว ผลลัพธ์ของแคมเปญแสดงให้เห็นว่าโมเดลการแบ่งกลุ่มได้ช่วยเพิ่มจำนวนการซื้อสินค้า และยอดเงินเฉลี่ยของตะกร้าสินค้าได้อย่างมีประสิทธิภาพมากขึ้น

บทที่ 3

วิธีการดำเนินงานวิจัย

งานวิจัยการศึกษาพฤติกรรมของลูกค้าในการซื้อสินค้าบนแพลตฟอร์มออนไลน์ โดยใช้วิธีวิเคราะห์ RFM ซึ่งจะทำให้ทราบถึงพฤติกรรมของลูกค้าและสามารถจัดกลุ่มลูกค้าออกเป็นกลุ่มๆ ตามแต่พฤติกรรมการซื้อสินค้าของลูกค้าแต่ละคน ในงานวิจัยนี้ผู้วิจัยได้ดำเนินงานวิจัยตามขั้นตอนดังนี้

3.1 ข้อมูลการซื้อขายสินค้า

งานวิจัยนี้ได้ใช้ข้อมูลการซื้อขายของร้านค้าออนไลน์ที่เกิดขึ้นในช่วงวันที่ 10 กุมภาพันธ์ 2020 จนถึง 31 ธันวาคม 2022 ประกอบด้วย 7 สดมภ์ และมีจำนวนข้อมูลทั้งหมด 88,238 แถว โดยแต่ละสดมภ์มีความหมายดังนี้

1. Order no คือ หมายเลขรหัสคำสั่งซื้อของลูกค้า
2. Date time คือ วันและเวลาที่ลูกค้าทำการสั่งซื้อ
3. Customer name คือ ชื่อลูกค้า
4. Product Name คือ ชื่อสินค้า
5. Product Variant คือ ตัวเลือกสินค้า เช่น ไซส์ สี ขนาด เป็นต้น
6. Qty คือ จำนวนสินค้าที่ลูกค้าสั่งซื้อ
7. Sales คือ ราคาสินค้า

3.2 เครื่องมือที่ใช้ในงานวิจัย

3.2.1 โปรแกรมสำหรับการประมวลผลภาษาไพทอน

3.2.2 ไลบรารี่ในการเตรียมข้อมูล เช่น Pandas , Numpy

3.2.3 ไลบรารี่ในการทำแบบจำลอง เช่น statmodel , sklearn

3.2.4 ไลบรารี่ในการแสดงผลข้อมูล matplotlib , seaborn

3.2.5 โปรแกรมสำหรับแสดงผลข้อมูล Tableau

3.3 การตรวจสอบข้อมูลและพิจารณาข้อมูลมาใช้ในการวิเคราะห์

ในงานวิจัยนี้ใช้ภาษาไพทอน (Python) ในการดำเนินการเริ่มต้นจากการนำเข้าที่เก็บรวบรวมชุดคำสั่งที่จำเป็นในการวิจัย และนำเข้าข้อมูลคำสั่งซื้อ ดังรูปที่ 3.1

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 from datetime import datetime, timedelta,date
6 from sklearn.preprocessing import StandardScaler
7
8 df=pd.read_csv('/content/Transaction data for RFM.csv')
```

รูปที่ 3.1 ตัวอย่างโค้ดนำเข้าไลบรารี่ที่จำเป็นและไฟล์ข้อมูล

	OrderNo	DateTime	Qty	Sales	CustomerName	ProductName	ProductVariants
0	2311115XKEY57V	31-12-22 23:46	1	344	Cust03967	PRD10071	PRV10194
1	2311115V4JMUTY	31-12-22 23:06	1	498	Cust43385	PRD10171	PRV10342
2	2212315NKFN1EG	31-12-22 21:05	1	205	Cust07387	PRD10030	PRV10027
3	2212315MX6MEPB	31-12-22 20:58	1	1094	Cust23450	PRD10003	PRV10029
4	2212315MXYH8RJ	31-12-22 20:58	1	452	Cust28334	PRD10100	PRV10615

รูปที่ 3.2 ข้อมูลรายการคำสั่งซื้อ

ทำการตรวจสอบประเภทของข้อมูลแต่ละสดมภ์ พบว่าสดมภ์ DateTime เป็นประเภทตัวเลข ดังนั้นจึงต้องทำการแก้ไขให้เป็นประเภทวันที่ให้ถูกต้อง

```
1 #ตรวจสอบ Type ของข้อมูล
2 df.dtypes
```

```
OrderNo           object
DateTime          object
Qty                int64
Sales              int64
CustomerName      object
ProductName       object
ProductVariants   object
dtype: object
```

รูปที่ 3.3 ประเภทของข้อมูลของตัวแปร

ทำการตรวจสอบค่าว่างพบว่าค่าในสดมภ์ CustomerName มีค่าว่างทั้งหมด 319 แถว ดังนั้นจึงทำการตัดแถวที่มีค่าว่างทิ้ง เพราะไม่สามารถระบุตัวตนของลูกค้าได้ ในส่วนสดมภ์ ProductVariant มีค่าว่าง 14449 แถว ดังรูปที่ 3.4 จะไม่ทำการตัดทิ้งเพราะในขั้นตอนการทำ RFM นั้นจะใช้ข้อมูลเพียงสดมภ์ DateTime, Qty และ Sales เท่านั้น ในส่วนสดมภ์ ProductVariant จะถูกนำมาใช้ประกอบในการตัดแถวของข้อมูลที่มีค่าซ้ำกัน ดังรูปที่ 3.5

```
#ตรวจสอบค่า Null
df.isnull().sum()
```

```
OrderNo              0
DateTime             0
Qty                  0
Sales                0
CustomerName       319
ProductName          0
ProductVariants    14449
dtype: int64
```

รูปที่ 3.4 จำนวนค่าว่างแต่ละสดมภ์

```
# Remove duplicate
print("Before remove :", df.duplicated().sum())
df.drop_duplicates(inplace=True)
print("After remove :", df.duplicated().sum())
```

```
Before remove : 247
After remove : 0
```

รูปที่ 3.5 การตัดแถวของข้อมูลที่มีค่าซ้ำกัน

3.4 การแปลงข้อมูลเพื่อเตรียมทำการคำนวณค่า RFM

ทำการเลือกข้อมูลเฉพาะสดมภ์ CustomerName, DateTime, OrderNo, Sales เก็บค่าข้อมูลในตัวแปร df2 เพื่อนำไปใช้ในการคำนวณค่า RFM ดังรูปที่ 3.6

```
#เลือกเฉพาะคอลัมน์ 'CustomerName','DateTime','OrderNo','Sales'
df2 = df[['CustomerName','DateTime','OrderNo','Sales']]
df2
```

	CustomerName	DateTime	OrderNo	Sales
0	Cust03967	2022-12-31 23:46:00	2311115XKEY57V	344
1	Cust43385	2022-12-31 23:06:00	2311115V4JMUTY	498
2	Cust07387	2022-12-31 21:05:00	2212315NKFN1EG	205
3	Cust23450	2022-12-31 20:58:00	2212315MX6MEPB	1094
4	Cust28334	2022-12-31 20:58:00	2212315MXYH8RJ	452

รูปที่ 3.6 เลือกข้อมูลเฉพาะสดมภ์ CustomerName, DateTime, OrderNo และ Sales

ในการเตรียมข้อมูลเพื่อทำการคำนวณค่า Recency สามารถคำนวณได้จากวันที่ล่าสุดในข้อมูลการขาย (รูปที่ 3.7) - วันที่ล่าสุดที่ลูกค้ารายนั้นมีการสั่งซื้อสินค้า จะได้ช่วงระยะเวลาที่ลูกค้าสั่งซื้อครั้งล่าสุด ในส่วนการคำนวณค่า Frequency จะทำการคำนวณโดยการนับจำนวนของรหัสคำสั่งซื้อ (OrderNo) ที่ไม่ซ้ำกันของลูกค้ารายนั้นๆ เพื่อหาจำนวนครั้งที่มีการสั่งซื้อทั้งหมด และในการคำนวณค่า Monetary จะคิดจากผลรวมของยอดสั่งซื้อ (Sales) ทั้งหมดของลูกค้ารายนั้นๆ ดังรูปที่ 3.8 และทำการเปลี่ยนชื่อสดมภ์เป็นชื่อ Recency, Frequency และ Monetary ตามรูปที่ 3.9

```
#ให้วันที่ล่าสุดในข้อมูลการขายเป็น Today date
today_date = df2['DateTime'].max()
today_date
```

```
Timestamp('2022-12-31 23:46:00')
```

รูปที่ 3.7 กำหนดวันที่ล่าสุดในข้อมูลเป็น today_date

```
#คำนวณค่า RFM

df_cust = df2.groupby("CustomerName").agg({"DateTime": lambda x: (today_date - x.max()).days,
                                           "OrderNo": lambda x: x.nunique(),
                                           "Sales": lambda x: x.sum()}).reset_index()

df_cust
```

	CustomerName	DateTime	OrderNo	Sales
0	Cust00001	698	1	1132
1	Cust00002	95	1	617
2	Cust00003	144	1	286
3	Cust00004	780	1	1235
4	Cust00005	510	1	182

รูปที่ 3.8 การคำนวณค่า RFM

```
df_cust.columns = ["CustomerName","Recency", "Frequency", "Monetary"]
df_cust
```

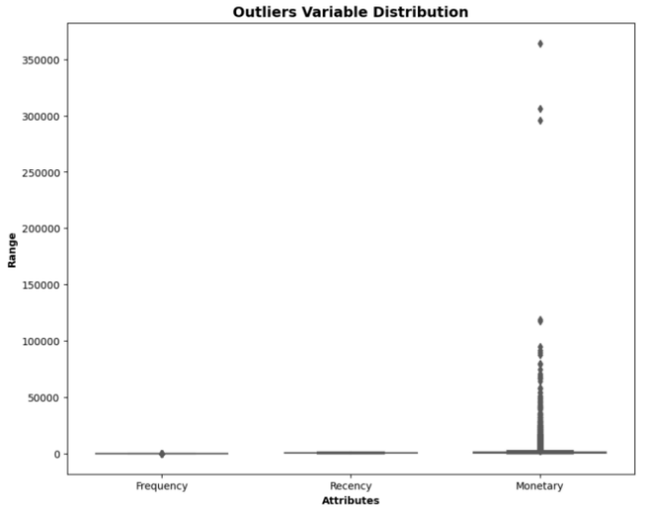
	CustomerName	Recency	Frequency	Monetary
0	Cust00001	698	1	1132
1	Cust00002	95	1	617
2	Cust00003	144	1	286
3	Cust00004	780	1	1235
4	Cust00005	510	1	182

รูปที่ 3.9 เปลี่ยนชื่อสดมภ์ Recency, Frequency และ Monetary

นำค่า Recency, Frequency และ Monetary สร้างแผนภาพกล่องเพื่อดูการกระจายตัวและค่าผิดปกติ (ดังรูป 3.10) แต่จะไม่ทำการตัดค่าผิดปกติออกเนื่องจากจะทำให้ผลการวิเคราะห์นั้นขาดกลุ่มลูกค้าที่มีค่าสูงของ Recency, Frequency และ Monetary หายไปจากการวิเคราะห์

```
# Outlier Analysis of Frequency Recency and Monetary

attributes = ['Frequency','Recency','Monetary']
plt.rcParams['figure.figsize'] = [10,8]
sns.boxplot(data = df_cust[attributes], orient="v", palette="Set2" ,whis=1.5,saturation=1, width=0.7)
plt.title("Outliers Variable Distribution", fontsize = 14, fontweight = 'bold')
plt.ylabel("Range", fontweight = 'bold')
plt.xlabel("Attributes", fontweight = 'bold')
```

```
Text(0.5, 0, 'Attributes')
```



รูปที่ 3.10 แผนภาพกล่องของค่า Recency, Frequency, Monetary

3.5 การจัดกลุ่มลูกค้าจากค่า RFM โดยใช้วิธีค่าเฉลี่ย k กลุ่ม

ก่อนนำข้อมูลทำการจัดกลุ่มโดยใช้วิธีค่าเฉลี่ย k กลุ่ม สังเกตการกระจายตัวของข้อมูลทั้ง 3 ค่า Recency, Frequency และ Monetary จากรูปที่ 3.10 จะเห็นได้ว่าข้อมูลมีค่าที่แตกต่างกันมาก ดังนั้นจะต้องทำให้ข้อมูลเป็นมาตรฐาน (Standard scaler) ก่อนนำข้อมูลจัดกลุ่มโดยวิธีค่าเฉลี่ย k กลุ่ม เพื่อลดความเอนเอียงของผลลัพธ์ ดังรูป 3.11

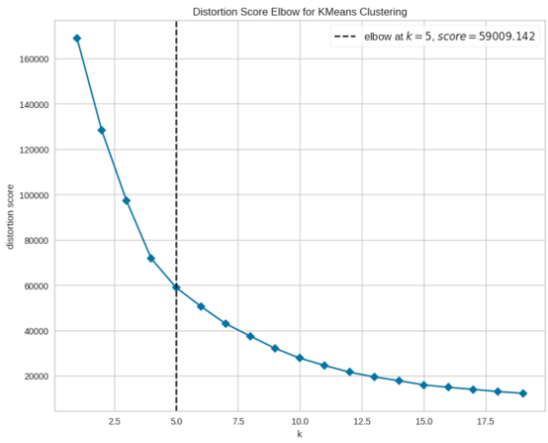
```
# Standard scaler

df_scaled = df_cust[['Recency', 'Frequency', 'Monetary']]

df_cust_scaled = StandardScaler().fit_transform(df_scaled)

```

```
df_cust_scaled = pd.DataFrame(df_cust_scaled)
df_cust_scaled.columns = ['Recency', 'Frequency', 'Monetary']
df_cust_scaled.head()
```

	Recency	Frequency	Monetary
0	1.192838	-0.201759	0.033856
1	-1.197326	-0.201759	-0.112740
2	-1.003100	-0.201759	-0.206959
3	1.517868	-0.201759	0.063175
4	0.447646	-0.201759	-0.236563

รูปที่ 3.11 การทำข้อมูลเป็นมาตรฐานของค่า Recency, Frequency และ Monetary

ทำการหาจำนวนกลุ่มที่เหมาะสมในการจัดกลุ่มลูกค้าโดยใช้ค่า Elbow method ซึ่งจากผลพบว่าจำนวนกลุ่มที่ 5 กลุ่มมีความเหมาะสมที่สุด (ดังรูป 3.12) ดังนั้นจะกำหนดค่า K=5 โดยกำหนดให้ทำการแบ่งกลุ่มลูกค้าออกเป็น 5 กลุ่ม ดังรูป 3.13



รูปที่ 3.12 ค่า K ที่เหมาะสม โดยวิธี Elbow method

```
1 from sklearn.cluster import KMeans
2 model = KMeans(n_clusters=5)
3 model.fit(X)
```

รูปที่ 3.13 การแบ่งกลุ่มและกำหนดค่า K=5

3.6 การจัดกลุ่มลูกค้าจากค่าคะแนน RFM (Traditional RFM Scoring)

นำข้อมูลที่ผ่านการคำนวณค่า Recency, Frequency และ Monetary มาทำการแบ่งช่วงโดยเริ่มจากการนำค่า Recency มาแบ่งช่วงควอไทล์เป็น 5 ช่วง ดังรูป 3.14 โดยช่วงของข้อมูลที่มีค่ามากจะได้คะแนนเท่ากับ 5 คะแนน และลดลงตามลำดับช่วงของข้อมูลที่มีค่าน้อยสุดจะมีคะแนนเท่ากับ 1 คะแนน

```
1 #แบ่งช่วงของค่าของข้อมูลเป็น 5 กลุ่ม
2
3 rfm["RecencyScore"] = pd.qcut(rfm['Recency'], 5, labels = [5, 4, 3, 2, 1])
```

รูปที่ 3.14 การแบ่งช่วงข้อมูลของ Recency

หาค่า Frequency ที่ค่าเปอร์เซ็นไทล์ต่างๆ พบว่าความถี่ในการซื้อสินค้าของร้านค้าที่เปอร์เซ็นไทล์ที่ 90 มีค่าเท่ากับ 1 ครั้ง เปอร์เซ็นไทล์ที่ 95 มีค่าเท่ากับ 2 ครั้ง เปอร์เซ็นไทล์ที่ 98 มีค่าเท่ากับ 3 ครั้ง และ เปอร์เซ็นไทล์ที่ 99 มีค่าเท่ากับ 4 ครั้ง และลูกค้ามีการสั่งซื้อซ้ำมากที่สุดที่ 79 ครั้ง (ดังรูป 3.15) ดังนั้นผู้วิจัยจะทำการแบ่งข้อมูลออกเป็น 5 ช่วง ตามค่าความถี่คือ 1,2,3,4,79 ครั้ง โดยกลุ่มลูกค้าที่มีช่วงค่าความถี่ตั้งแต่ 5-79 ครั้ง จะได้คะแนนเท่ากับ 5 คะแนน, ความถี่เท่ากับ 4 ครั้ง มีคะแนนเท่ากับ 4 คะแนน และลดลงตามลำดับ โดยที่คนที่มีความถี่ 1 ครั้ง จะได้คะแนนเท่ากับ 1 คะแนน (ดังรูป 3.16) ซึ่งจำนวนลูกค้าที่มีการสั่งซื้อ 1 ครั้ง มีจำนวนเท่ากับ 50740 คน สั่งซื้อ 2 ครั้ง มีจำนวน 4177 คน สั่งซื้อ 3 ครั้ง มีจำนวน 856 คน สั่งซื้อ 4 ครั้ง มีจำนวน 261 คน และสั่งซื้อมากกว่า 4 ครั้งมีจำนวน 310 คน

```
1 rfm["Frequency"].describe([0.01,0.05,0.10,0.25,0.50,0.90,0.95,0.98,0.99]).T
```

```
count    56344.000000
mean         1.155030
std          0.768397
min          1.000000
1%           1.000000
5%           1.000000
10%          1.000000
25%          1.000000
50%          1.000000
90%          1.000000
95%          2.000000
98%          3.000000
99%          4.000000
max         79.000000
Name: Frequency, dtype: float64
```

รูปที่ 3.15 ค่า Frequency แต่ละค่าเปอร์เซนไทล์

```
1 #แบ่งช่วง Frequency ออกเป็น 5ช่วง
2
3 cut_bins = [0,1,2,3,4,79]
4
5 rfm["FrequencyScore"] = pd.cut(rfm["Frequency"], bins = cut_bins, labels = [1, 2, 3, 4, 5])
6 rfm["FrequencyScore"].value_counts()
```

```
1    50740
2     4177
3      856
5      310
4      261
```

รูปที่ 3.16 การแบ่งค่า Frequency ออกเป็น 5 ช่วง

ทำการแบ่งช่วงของข้อมูลในสดมภ์ Monetary ออกเป็น 5 ช่วง โดยลูกค้าที่มีช่วงของค่า Monetary มากที่สุดจะได้คะแนน 5 คะแนน และลดลงตามลำดับ ดังแสดงในรูป 3.17 และทำการรวมคะแนนทั้ง 3 ไว้ด้วยกันตั้งชื่อว่า RFM_score สุดท้ายนำคะแนน RFM score ไปจัดกลุ่มตาม

เกณฑ์คะแนน RFM ในการแบ่งกลุ่มลูกค้าตามตารางที่ 2.1 และจะได้กลุ่มลูกค้าทั้งหมด 11กลุ่ม ตามรูปที่ 3.18

```
#แบ่งช่วง Monetaryออกเป็น 5ช่วง

rfm["MonetaryScore"] = pd.qcut(rfm['Monetary'], 5, labels = [1, 2, 3, 4, 5])
rfm.head()
```

	CustomerName	Recency	Frequency	Monetary	RecencyScore	FrequencyScore	MonetaryScore
0	Cust00001	698	1	1132	1	1	4
1	Cust00002	95	1	617	5	1	4
2	Cust00003	144	1	286	5	1	2
3	Cust00004	780	1	1235	1	1	4
4	Cust00005	510	1	182	2	1	2

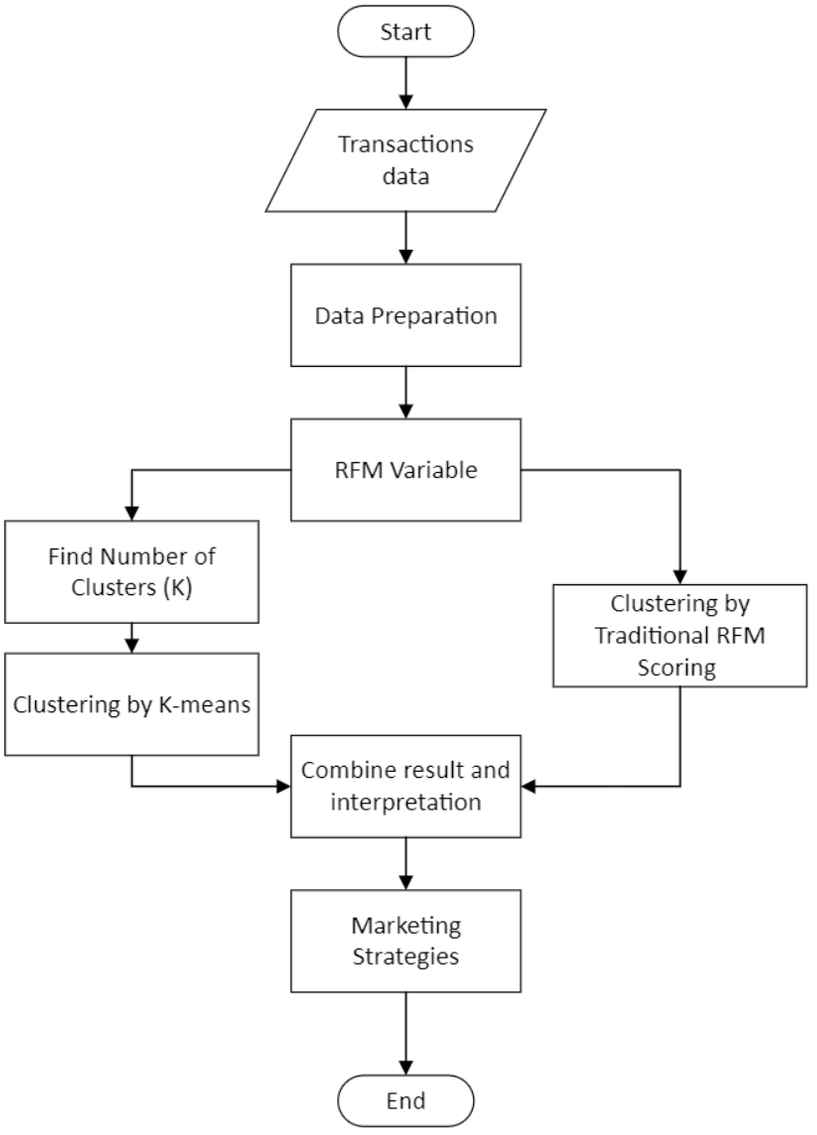
รูปที่ 3.17 แสดงการแบ่งช่วงของ Monetary ออกเป็น 5 ช่วง

```
rfm['Segment'] = rfm['Segment'].replace(seg_map, regex=True)
rfm.head()
```

	CustomerName	Recency	Frequency	Monetary	RecencyScore	FrequencyScore	MonetaryScore	RFM_SCORE	Segment
0	Cust00001	698	1	1132	1	1	4	114	Hibernating
1	Cust00002	95	1	617	5	1	4	514	New Customers
2	Cust00003	144	1	286	5	1	2	512	New Customers
3	Cust00004	780	1	1235	1	1	4	114	Hibernating
4	Cust00005	510	1	182	2	1	2	212	Hibernating

รูปที่ 3.18 แสดงการแบ่งกลุ่มลูกค้าตาม Traditional RFM scoring

3.7 สรุปขั้นตอนการดำเนินงาน



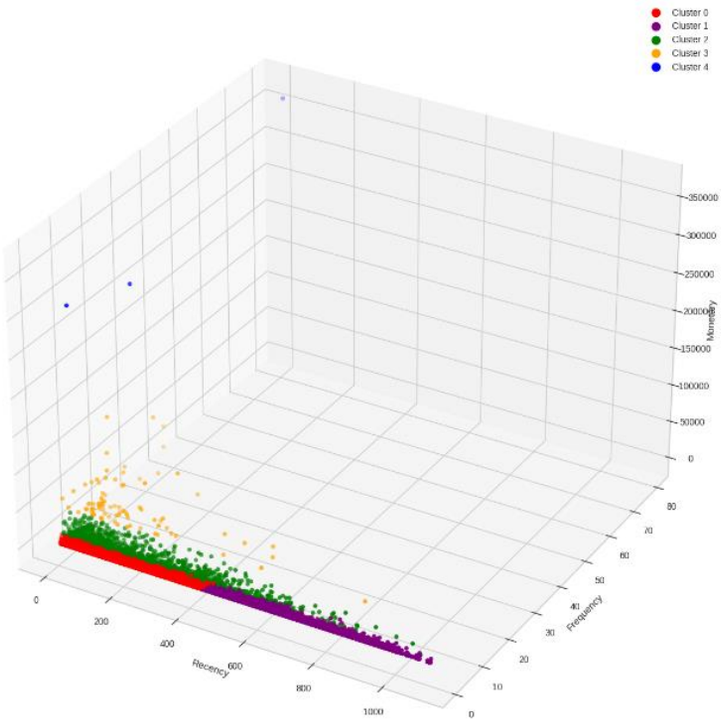
รูปที่ 3.19 แสดงขั้นตอนการดำเนินการ

แผนผังแสดงขั้นตอนการดำเนินการ (รูปที่ 3.19) เริ่มจากการนำข้อเข้ามูลการซื้อขายสินค้าออนไลน์ ผ่านกระบวนการทำความสะอาดข้อมูล และเตรียมให้พร้อมสำหรับขั้นตอนต่อไปคือ การคำนวณตัวแปร R, F, และ M หลังจากนั้นเมื่อได้ตัวแปรทั้ง 3 ตัว จะแบ่งออกเป็น 2 ส่วนแรกคือ นำข้อมูลไปหาจำนวนค่า k ที่เหมาะสม และ ทำการจัดกลุ่มด้วยวิธีค่าเฉลี่ย k กลุ่ม ส่วนที่สองคือทำการคำนวณคะแนนและจัดกลุ่มตามขั้นตอนวิธีการจัดกลุ่มด้วยค่าคะแนนแบบดั้งเดิม หลังจากเสร็จทั้งสองส่วนแล้วจึงนำผลที่ได้มารวมกันเพื่อตีความและวิเคราะห์ผล และกำหนดกลยุทธ์ทางการตลาดที่เหมาะสมกับลูกค้าแต่ละกลุ่ม

บทที่ 4
ผลการวิจัยและการอภิปรายผล

ในงานวิจัยนี้ได้ศึกษาการวิเคราะห์ลูกค้าด้วยวิธี RFM และได้ทำการแบ่งกลุ่มโดยใช้การเรียนรู้ของเครื่องโดยวิธีค่าเฉลี่ย k กลุ่ม วิเคราะห์ประกอบร่วมกับการใช้วิธีการแบ่งกลุ่มแบบดั้งเดิมคือการแบ่งกลุ่มลูกค้าตามเกณฑ์คะแนน (Traditional RFM scoring) ซึ่งได้ผลจากการแบ่งกลุ่มดังนี้

4.1 ผลการแบ่งกลุ่มลูกค้า

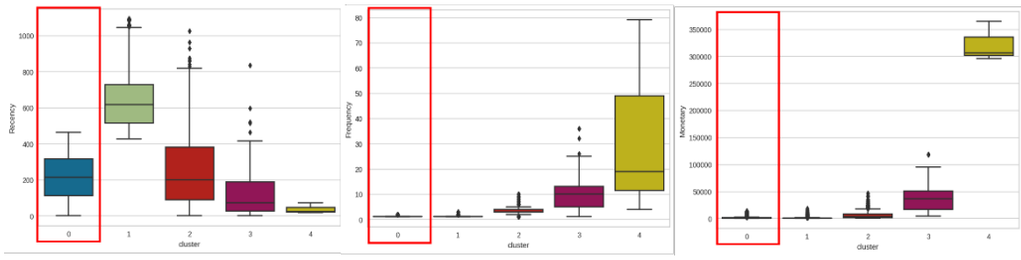


รูปที่ 4.1 กราฟสามมิติแสดงการกระจายตัวของแต่ละกลุ่มลูกค้าจากวิธีค่าเฉลี่ย k กลุ่ม

ผลจากการแบ่งกลุ่มจากวิธีค่าเฉลี่ย k กลุ่ม นั้นจำนวนกลุ่มที่เหมาะสมจากวิธี Elbow จะได้จำนวนกลุ่มที่เหมาะสมคือจำนวน 5 กลุ่ม ดังรูปที่ 4.1 ผลจากการแบ่งกลุ่มด้วยวิธีค่าเฉลี่ย k กลุ่ม สามารถแบ่งกลุ่มลูกค้าที่มีความคล้ายคลึงกันออกมาได้ผลลัพธ์ดังนี้ กลุ่มที่ 0 30,785 คน คิดเป็น 54.65% กลุ่มที่ 1 23,773 คน คิดเป็น 42.18% กลุ่มที่ 2 1,695 คน คิดเป็น 0.16% กลุ่ม 3 88คน คิดเป็น 3.01% และกลุ่มที่ 4 3คน คิดเป็น 0.01% เมื่อนำค่า Recency, Frequency และ Monetary ของแต่ละกลุ่มนำมาสร้างแผนภาพกล่องและค่าสถิติเบื้องต้น (รูป 4.2-4.6) เพื่อทำการวิเคราะห์ลักษณะกลุ่มลูกค้าทั้ง 5 กลุ่ม ได้ดังนี้

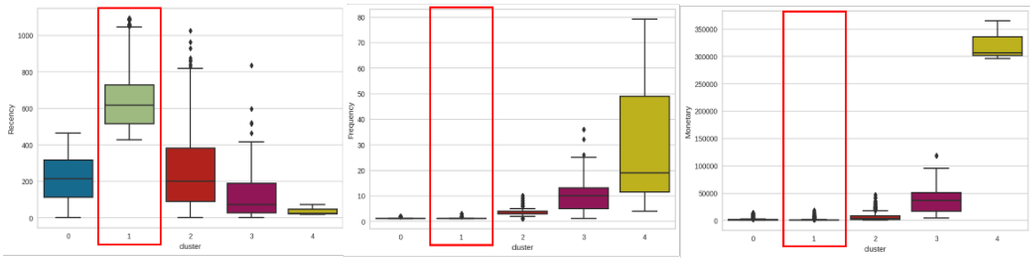
ตารางที่ 4.1 ค่าสถิติเบื้องต้นกลุ่มลูกค้าทั้ง 5 กลุ่ม

Cluster	count	Recency			Frequency			Monetary		
		mean	max	min	mean	max	min	mean	max	min
0	30788	216.07	464	0	1.09	2	1	902.63	13856	0
1	23770	643.50	1092	427	1.05	3	1	611.76	18672	0
2	1695	243.51	1026	0	3.30	10	1	6139.72	46572	75
3	88	128.33	834	1	10.28	36	1	38351.02	118913	3874
4	3	37.67	71	18	34.00	79	4	322138.33	364336	295754



รูปที่ 4.2 แสดงแผนภาพกล่องค่า R, F และ M ของกลุ่มที่ 0

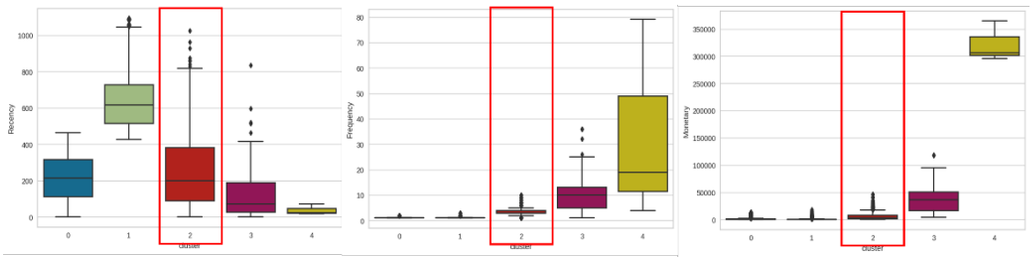
กลุ่มที่ 0 : มี Recency เฉลี่ยที่ 216 วัน หมายถึงลูกค้าในกลุ่มนี้มีการทำธุรกรรมครั้งล่าสุดค่อนข้างนาน จำนวนที่นับได้ในกลุ่มนี้มี 30,788 คน ความถี่เฉลี่ยของการทำธุรกรรมในกลุ่มนี้เป็น 1.09 ครั้ง และมูลค่าเฉลี่ยของการทำธุรกรรมในกลุ่มนี้เท่ากับ 902 บาท จากพฤติกรรมจะตั้งชื่อลูกค้ากลุ่มนี้ว่า Short-Term Inactive Customer คือกลุ่มลูกค้าที่ไม่มีการซื้อสินค้ามาเป็นระยะเวลานานและมีความถี่ในการซื้อสินค้ารวมถึงยอดซื้อค่อนข้างน้อย



รูปที่ 4.3 แสดงแผนภาพกล่องค่า R, F และ M ของกลุ่มที่ 1

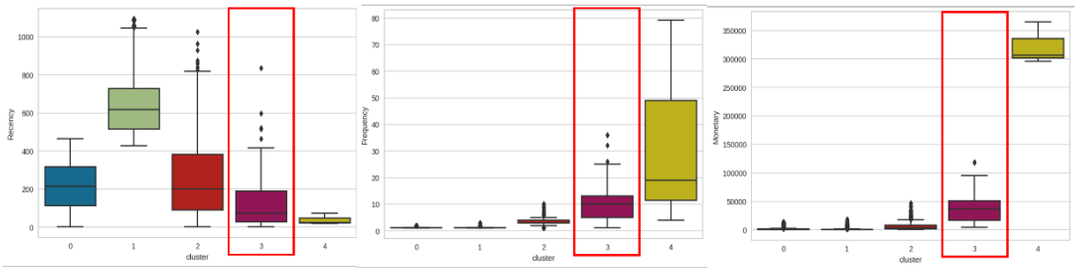
กลุ่มที่ 1 : มี Recency เฉลี่ยที่ 643.5 วัน หมายถึงลูกค้าในกลุ่มนี้มีการทำธุรกรรมครั้งล่าสุดไปเมื่อเวลาผ่านไปมากกว่ากลุ่มที่ 1 จำนวนที่นับได้ในกลุ่มนี้มี 23,770 คน ความถี่เฉลี่ยของการทำธุรกรรมในกลุ่มนี้เป็น 1.05 ครั้ง และมูลค่าเฉลี่ยของการทำธุรกรรมในกลุ่มนี้เท่ากับ 611 บาท จาก

พฤติกรรมจะตั้งชื่อลูกค้ากลุ่มนี้ว่า Long-Term Inactive Customer คือกลุ่มลูกค้าที่ไม่มีการซื้อสินค้ามาเป็นระยะเวลานานมาก และมีความถี่ในการซื้อสินค้ารวมถึงยอดซื้อน้อยที่สุดในทุกกลุ่ม



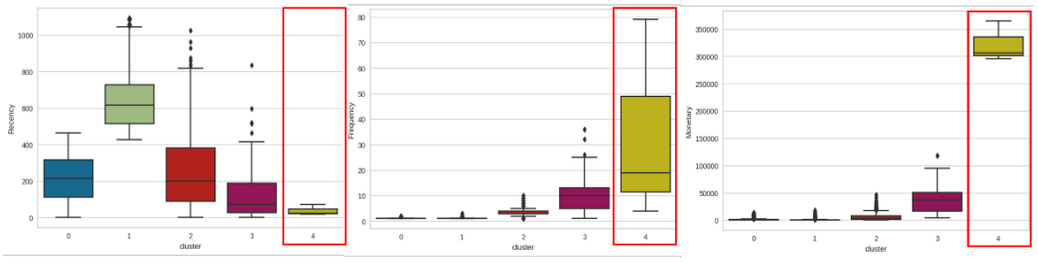
รูปที่ 4.4 แสดงแผนภาพกล่องค่า R, F และ M ของกลุ่มที่ 2

กลุ่มที่ 2 : มี Recency เฉลี่ยที่ 243.51 วัน หมายถึงลูกค้าในกลุ่มนี้มีการทำธุรกรรมครั้งล่าสุดค่อนข้างนาน จำนวนที่นับได้ในกลุ่มนี้มี 1,695 คน ความถี่เฉลี่ยของการทำธุรกรรมในกลุ่มนี้เป็น 3.3 ครั้งและมูลค่าเฉลี่ยของการทำธุรกรรมในกลุ่มนี้เท่ากับ 6,139.72 บาท จากพฤติกรรมจะตั้งชื่อลูกค้ากลุ่มนี้ว่า At Risk คือเป็นกลุ่มลูกค้าที่เคยซื้อสินค้าบ่อยและใช้จ่ายสินค้าสูงกว่าลูกค้ากลุ่มที่ 1 และ 2 แต่ไม่ได้มีการกลับมาซื้อสินค้าเป็นระยะเวลานานแล้ว ดังนั้นจึงมีแนวโน้มที่จะเลิกใช้บริการหรือหยุดการซื้อสินค้ากับร้านค้า



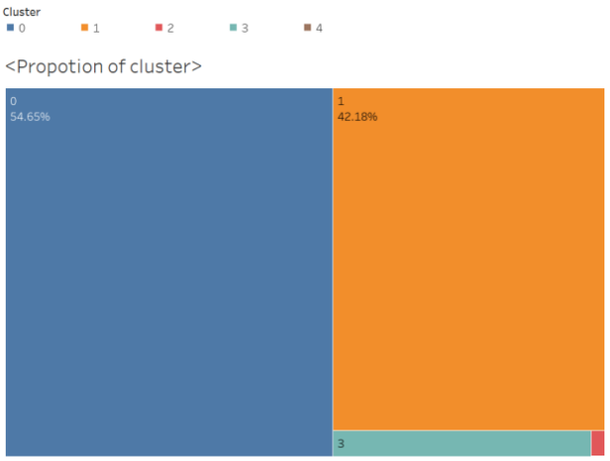
รูปที่ 4.5 แสดงแผนภาพกล่องค่า R, F และ M ของกลุ่มที่ 3

กลุ่มที่ 3 : มี Recency เฉลี่ยที่ 128.33 วัน หมายถึงลูกค้าในกลุ่มนี้มีการทำธุรกรรมครั้งล่าสุดไปไม่นานมานัก จำนวนที่นับได้ในกลุ่มนี้มี 88 คน ความถี่เฉลี่ยของการทำธุรกรรมในกลุ่มนี้เป็น 10.28 ครั้ง และมูลค่าเฉลี่ยของการทำธุรกรรมในกลุ่มนี้เท่ากับ 38,351 บาท จากพฤติกรรมจะตั้งชื่อลูกค้ากลุ่มนี้ว่า Loyal Big Spenders Customers คือกลุ่มลูกค้าที่มีการทำธุรกรรมไปไม่นาน และเป็นกลุ่มลูกค้าประจำที่มีการซื้อสินค้าบ่อยครั้ง และมีกำลังซื้อสูงเป็นอันดับที่สองรองจากกลุ่มที่ 4



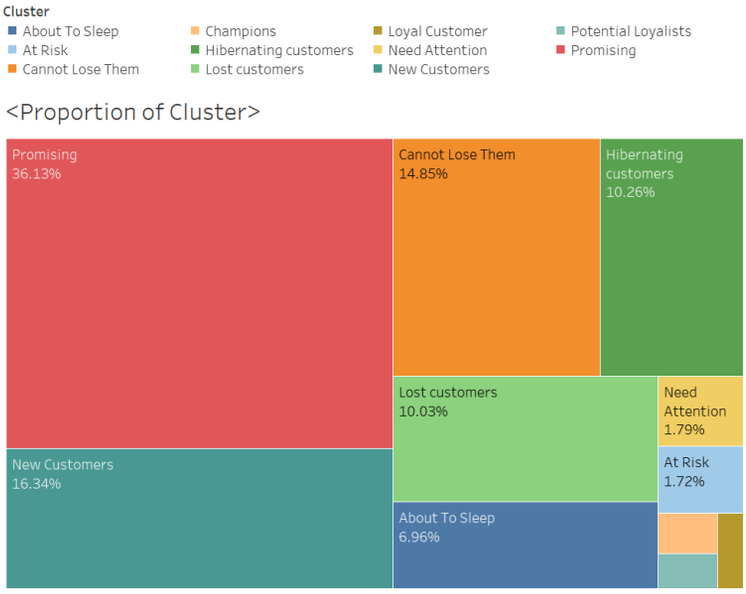
รูปที่ 4.6 แสดงแผนภาพกล่องค่า R, F และ M ของกลุ่มที่ 4

กลุ่มที่ 4 : มี Recency เฉลี่ยที่ 37.66 วัน หมายถึงลูกค้าในกลุ่มนี้มีการทำธุรกรรมครั้งล่าสุดไปเมื่อเวลาผ่านไปไม่นานมานัก จำนวนที่นับได้ในกลุ่มนี้มี 3 คน ความถี่เฉลี่ยของการทำธุรกรรมในกลุ่มนี้เป็น 34 ครั้ง และมูลค่าเฉลี่ยของการทำธุรกรรมในกลุ่มนี้เท่ากับ 322,138 บาท จากพฤติกรรมจะตั้งชื่อลูกค้ากลุ่มนี้ว่า Best Customer คือกลุ่มลูกค้าที่ทำธุรกรรมไปไม่นาน และ มีความถี่ในการซื้อสินค้าบ่อยมาก และมียอดซื้อสูงที่สุดในทุกกลุ่ม



รูปที่ 4.8 แผนภูมิต้นไม้แสดงสัดส่วนของกลุ่มลูกค้าทั้ง 5กลุ่ม

ผลจากการแบ่งกลุ่มลูกค้าจากค่าคะแนนแบบดั้งเดิม นั้นสามารถที่จะแบ่งกลุ่มลูกค้าได้ทั้งหมด 11 กลุ่มและสามารถตั้งชื่อกลุ่มได้ตามตารางที่ 2.1 ซึ่งได้ผลที่ได้พบว่ากลุ่มลูกค้าส่วนใหญ่ในร้านค้า 3 อันดับแรกจะเป็นกลุ่มลูกค้าประเภท Promising คิดเป็น 20355 คน คิดเป็น 36.13% กลุ่ม New Customers 9206 คน คิดเป็น 16.34% และกลุ่ม Cannot lose them 8365 คน คิดเป็น 14.85% ดังรูปที่ 4.9



รูปที่ 4.9 แผนภูมิต้นไม้แสดงสัดส่วนของกลุ่มลูกค้าทั้ง11กลุ่ม

จากวิธีการแบ่งกลุ่มแบบดั้งเดิมนั้นจะเริ่มจากการให้คะแนนลูกค้าโดยดูจากค่าของ Recency, Frequency และ Monetary และแบ่งลูกค้าออกด้วยจำนวนที่เท่ากัน 5 กลุ่ม และให้คะแนน 1-5 คะแนน จากนั้นรวมคะแนนและเทียบคะแนนกับเกณฑ์ จะแบ่งลูกค้าได้ทั้งหมด 11กลุ่ม วิธีนี้จะไม่ได้มีการหาความสัมพันธ์ระหว่างลูกค้าด้วยกันจากค่าของ Recency, Frequency และ Monetary สืบเนื่องจากการแบ่งกลุ่มลูกค้าโดยไม่ได้สนใจความสัมพันธ์ของข้อมูล ทำให้มีโอกาสที่กลุ่มลูกค้าอาจจะมีได้น้อยกว่าหรือมากกว่า 11 กลุ่ม ตามความสัมพันธ์ที่แท้จริง เมื่อนำกลุ่มที่แบ่งได้ไปกำหนดกลยุทธ์ทางการตลาดต่อ ผู้วิเคราะห์อาจจะต้องรวมกลุ่มลูกค้าใหม่ด้วยตนเอง เพราะแต่ละกลุ่มไม่มีความหมายแตกต่างกันมากนัก แต่ในส่วนการใช้วิธีการเรียนรู้ของเครื่องโดยวิธีค่าเฉลี่ย k กลุ่ม สามารถที่จะแบ่งกลุ่มลูกค้าออกมาโดยพิจารณาจากความสัมพันธ์ของข้อมูลที่แท้จริง จากนั้นต้องอาศัยการตีความของผู้วิเคราะห์ว่าแต่ละกลุ่มมีพฤติกรรมอย่างไร ดังนั้นในการแบ่งกลุ่มลูกค้าในงานวิจัยนี้จะเลือกกลุ่มลูกค้าที่แบ่งได้จากวิธีค่าเฉลี่ย k กลุ่ม เป็นหลัก และจะนำกลุ่มที่แบ่งได้จากวิธีแบบดั้งเดิมมารวมและช่วยวิเคราะห์ร่วมกันในการทำกลยุทธ์ทางการตลาด

4.2 กำหนดกลยุทธ์ทางการตลาด

เมื่อทำการแบ่งกลุ่มลูกค้าออกเป็นกลุ่มสำเร็จแล้ว หลังจากนั้นเราควรกำหนดกลยุทธ์ให้เหมาะสมกับลูกค้าแต่ละกลุ่ม ซึ่งสามารถทำการกำหนดแนวทางได้ดังนี้

กลุ่มที่ 4 เป็นกลุ่มลูกค้า Best Customer มีจำนวน 3 คน จัดเป็นกลุ่มลูกค้าที่มีความสำคัญมากที่สุด ดังนั้นควรให้ความสำคัญในการรักษาลูกค้ากลุ่มนี้ไว้ทำให้ลูกค้าในกลุ่มนี้รู้สึกพิเศษที่สุด

ได้รับการบริการที่พิเศษด้วยคุณภาพสูงสุด รวมถึงการให้สิทธิพิเศษในการแจ้งสินค้าใหม่ๆ ของร้านค้า หรือ สินค้ารุ่นพิเศษที่มีจำนวนจำกัด และเชิญชวนลูกค้าแสดงความคิดเห็นเกี่ยวกับสินค้าให้กับร้านค้า เพื่อดึงดูดให้ลูกค้ารายอื่นมีความมั่นใจในตัวสินค้าเพิ่มมากขึ้น

กลุ่มที่ 3 เป็นกลุ่มลูกค้า Loyal Big Spenders Customers มีจำนวน 88 คน จัดเป็นกลุ่มลูกค้าที่มีความสำคัญควรสร้างความสัมพันธ์รวมถึงบริการที่ดีกับลูกค้า เข้าใจความต้องการและแนะนำสินค้าที่ตรงกับความต้องการโดยดูจากประวัติคำสั่งซื้อของลูกค้าแต่ละราย จัดโปรโมชั่นเฉพาะกลุ่มให้โค้ดส่วนลดโดยกำหนดราคาขั้นต่ำในการสั่งซื้อให้มากกว่าค่าเฉลี่ยการใช้จ่ายต่อครั้ง หากไม่มีการแบ่งกลุ่มลูกค้าทางผู้ขายจะเห็นแค่ข้อมูลภาพรวมว่ายอดใช้จ่ายต่อครั้งเฉลี่ยของลูกค้าทั้งร้านนั้นมีค่าเท่าใด ดังนั้นการแบ่งกลุ่มก็จะช่วยให้ร้านค้าทำโปรโมชั่นได้ตรงกลุ่มลูกค้ามากขึ้น เพื่อสร้างความพึงพอใจของลูกค้า และดึงดูดให้ลูกค้ากลับมาซื้อซ้ำและใช้จ่ายเพิ่มมากขึ้น ต้องพยายามให้ลูกค้ากลุ่มนี้เพิ่มจำนวนสินค้าในตระกร้าสินค้าให้เพิ่มขึ้น เช่นการทำ Add-on Deal คือโปรโมชั่นที่ช่วยให้ผู้ขายสามารถมอบส่วนลดสำหรับการซื้อสินค้ารองเมื่อผู้ซื้อสั่งซื้อสินค้าหลักจากร้าน โดยโปรโมชั่นนี้จะช่วยกระตุ้นให้ผู้ซื้อสั่งซื้อสินค้าจำนวนมากขึ้นต่อการซื้อ 1 ครั้ง เนื่องจากสามารถซื้อสินค้ารองเพิ่มเติมได้ในราคาพิเศษ หรือ การทำ Bundle deal คือโปรโมชั่นที่ช่วยให้ร้านค้าสามารถกำหนดส่วนลดสำหรับสินค้าที่มีการสั่งซื้อเป็นชุด ซึ่งจะสามารถดึงดูดผู้ซื้อที่ต้องการซื้อสินค้าจำนวนมาก หรือสินค้าที่สามารถใช้ร่วมกับสินค้าอื่น ๆ ได้ในราคาพิเศษ ซึ่งจะนำไปสู่ยอดขายที่เพิ่มขึ้นจากการเพิ่มจำนวนการซื้อสินค้าต่อ 1 คำสั่งซื้อ

กลุ่มที่ 2 เป็นกลุ่มลูกค้า At Risk มีจำนวน 1695คน จัดเป็นกลุ่มที่เสี่ยงที่จะเลิกใช้บริการหรือซื้อสินค้าของร้าน แต่เคยเป็นลูกค้ากลุ่มที่มีศักยภาพในการซื้อเพราะดูจากความถี่และยอดใช้จ่าย ดังนั้นลูกค้ากลุ่มนี้ควรทำการดึงกลับมา สาเหตุอาจเกิดจากการที่ลูกค้ากลุ่มนี้มีการซื้อสินค้าในลักษณะเดียวกันกับที่ซื้อจากร้านแต่ไปซื้อที่ร้านค้าคู่แข่งแทน ควรวิเคราะห์ดูถึงสินค้าที่ลูกค้ากลุ่มนี้ซื้อและเปรียบเทียบกับราคาในตลาด ควรที่จะนำเสนอราคาสินค้าที่แข่งขันได้ในตลาด หรือนำเสนอโค้ดส่วนลดพิเศษสำหรับลูกค้ากลุ่มนี้ โดยอ้างอิงจากพฤติกรรมการสั่งซื้อสินค้าที่ผ่านมา เพราะเป็นกลุ่มที่สร้างรายได้ให้กับร้านค้าเยอะดังนั้นจึงเป็นกลุ่มที่ควรลงทุนเพื่อให้กลับมาเป็นลูกค้าของเรา

จากการกำหนดกลยุทธ์ทางการตลาดของกลุ่มที่ 4, กลุ่มที่ 3 และ กลุ่มที่ 2 นั้น จะมีจำนวนลูกค้าทั้งหมดรวมกัน 1786 คน จากทั้งหมด 56,344 คน คิดเป็น 3.17% จากลูกค้าทั้งหมด แต่ในกลุ่มลูกค้าที่ 0 และ กลุ่มลูกค้าที่ 1 นั้น มีลูกค้าอยู่จำนวน 30,785 คน และ 23,770 คน ซึ่งเป็นจำนวนที่ค่อนข้างสูง ดังนั้นทางผู้วิจัยจะวิเคราะห์และเลือกลูกค้าบางส่วนจากทั้ง 2 กลุ่ม เพื่อมากำหนดกลยุทธ์ทางการตลาดสำหรับลูกค้าที่มีโอกาสตอบรับกลับมาซื้อสินค้า

กลุ่มที่ 0 เป็นกลุ่มลูกค้า Short-Term Inactive Customer มีจำนวน 30,785 คน เป็นกลุ่มลูกค้าที่เริ่มมีการเลิกใช้บริการไปแล้วเป็นระยะเวลาหนึ่ง ประกอบกับมีจำนวนลูกค้าในกลุ่มนี้ที่ค่อนข้างมาก ดังนั้นเราจะทำการเลือกลูกค้าแค่บางส่วนที่คาดว่ามีโอกาสและคุ้มค่าในการดึงกลับมาเป็นลูกค้าโดยวิเคราะห์ร่วมกับการจัดกลุ่มแบบดั้งเดิม คือการจัดกลุ่มลูกค้าตามเกณฑ์คะแนน (Traditional RFM scoring) จากตารางที่ 4.1

ตารางที่ 4.2 การรวมผลการจัดกลุ่มวิธีค่าเฉลี่ย k กลุ่มและวิธีค่าคะแนนแบบดั้งเดิมของกลุ่มที่ 0

K-MEANS_Segment	Traditional_RFM_Segment	Q'ty	% No. of Total
0	Promising	19410	63.04%
	New Customers	8779	28.51%
	About To Sleep	1724	5.60%
	Need Attention	512	1.66%
	Potential Loyalists	294	0.95%
	Hibernating customers	67	0.22%
	Cannot Lose Them	2	0.01%

จะทำการตลาดดึงลูกค้าในกลุ่มของ Potential Loyalists, Need Attention, Cannot Lose Them กลับมาเป็นลูกค้า เพราะเป็นกลุ่มที่มีศักยภาพโดยพิจารณา จากค่าคะแนน RFM ที่มีค่าสูงจากขั้นตอนการแบ่งกลุ่ม โดยจะทำการเสนอข้อเสนอพิเศษในการซื้อสินค้า การแจ้งการจัดโปรโมชั่นแบบจำกัดเวลา การให้โค้ดส่วนลด เพื่อดึงดูดให้ให้เกิดการกลับมาซื้อซ้ำ

กลุ่มที่ 1 เป็นกลุ่มลูกค้า Long-Term Inactive Customer จำนวน 23770 คน เป็นกลุ่มลูกค้าที่เลิกใช้บริการไปนานแล้วดูได้จากค่า Recency มีค่าน้อยที่สุดที่ 427 วัน ดังรูปที่ 4.4 ประกอบกับมีจำนวนลูกค้าในกลุ่มนี้ที่ค่อนข้างมาก ดังนั้นเราจะทำการเลือกลูกค้าแค่บางส่วนที่คาดว่ามีโอกาสและคุ้มค่าในการดึงกลับมาเป็นลูกค้าโดยวิเคราะห์ร่วมกับการแบ่งกลุ่มแบบดั้งเดิมคือการจัดกลุ่มลูกค้าตามเกณฑ์คะแนน (Traditional RFM scoring) จากตารางที่ 4.2 โดยจะทำการตลาดดึงลูกค้าในกลุ่มของ At Risk, Cannot Lose Them, Potential Loyalists และ Need Attention เพราะเป็นกลุ่มที่มีศักยภาพโดยพิจารณา จากค่าคะแนน RFM ที่มีค่าสูงจากพิจารณาจากขั้นตอนการแบ่งกลุ่ม ในกลุ่มที่1 นี้จะไม่ได้เลือกลูกค้าที่อยู่ในกลุ่ม New customers เพราะลูกค้าในกลุ่มที่ 1 ระยะเวลาที่สั่งซื้อน้อยที่สุดคือ 427 วัน ดังนั้นในกลุ่ม New customers ลูกค้าจะมีพฤติกรรมที่ซื้อไม่บ่อย และยอดใช้จ่ายต่ำ ดังนั้นจึงไม่ได้ถูกเลือกทำการตลาด โดยลูกค้าส่วนที่ถูกเลือกจะทำการเสนอส่วนลด เพื่อดึงให้เกิดการกลับมาซื้อซ้ำแต่ควรประเมินถึงผลลัพธ์ที่ได้กลับมา เพราะมีโอกาสที่จะไม่คุ้มค่าและไม่ได้รับการตอบรับกลับมา

ตารางที่ 4.3 การรวมผลการจัดกลุ่มวิธีค่าเฉลี่ย k กลุ่ม และวิธีค่าคะแนนแบบดั้งเดิมของกลุ่มที่ 1

K-MEANS_Segment	Traditional_RFM_Segment	Q'ty	% No. of Total
1	Cannot Lose Them	8341	35.09%
	Hibernating customers	5696	23.96%
	Lost customers	5652	23.78%
	About To Sleep	2193	9.23%
	At Risk	734	3.09%
	Promising	686	2.89%
	New Customers	427	1.80%
	Need Attention	28	0.12%
	Potential Loyalists	13	0.05%

การนำวิธีการแบ่งกลุ่มด้วยค่าคะแนนแบบดั้งเดิมมาร่วมในการวิเคราะห์เลือกลูกค้าที่จะทำกลยุทธ์การตลาดด้วยนั้น จะช่วยให้เข้าใจถึงพฤติกรรมของลูกค้า และเข้าใจลำดับความสำคัญของลูกค้าจากกลุ่มที่แบ่งได้จากวิธีค่าเฉลี่ย k ตัว ซึ่งเป็นวิธีที่ง่าย เพราะในกลุ่มที่ 0 และกลุ่มที่ 1 มีจำนวนลูกค้าที่เยอะ ไม่ต้องทำการวิเคราะห์ลูกค้าในกลุ่มเพิ่มเติม

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

การวิเคราะห์อาร์เอฟเอ็ม เป็นเทคนิคที่มีประโยชน์และสามารถใช้งานได้จริงในการแบ่งกลุ่มลูกค้าตามพฤติกรรมจากข้อมูลการซื้อขายสินค้าผ่าน 3 ตัวแปร คือ ระยะเวลาที่ลูกค้าสั่งซื้อครั้งล่าสุด จำนวนธุรกรรมในการสั่งซื้อ และ ผลรวมของยอดใช้จ่ายจนถึงวันปัจจุบัน ยิ่งเราทำการแบ่งกลุ่มลูกค้าออกมาได้แม่นยำ เราก็จะสามารถทำการตลาดที่ตรงใจได้ถูกคน ถูกเวลา และในปัจจุบันที่ข้อมูลการซื้อขายได้ถูกจัดเก็บอย่างเป็นระบบ การได้นำแนวคิดนี้ไปใช้จะทำให้การขายของออนไลน์นั้นยิ่งมีประสิทธิภาพที่ดีขึ้นและคุ้มค่ามากขึ้น ในงานวิจัยนี้ทำการศึกษาพฤติกรรมของลูกค้าจากข้อมูลการสั่งซื้อสินค้าและทำการวิเคราะห์พฤติกรรมโดยใช้เทคนิควิเคราะห์อาร์เอฟเอ็ม และทำการแบ่งกลุ่มลูกค้าโดยใช้การเรียนรู้ของเครื่องด้วยวิธีค่าเฉลี่ย k กลุ่ม วิเคราะห์ร่วมกับวิธีการจัดกลุ่มแบบดั้งเดิมคือการจัดกลุ่มลูกค้าตามเกณฑ์คะแนน (Traditional RFM scoring) ซึ่งจะช่วยให้วิเคราะห์กลุ่มลูกค้าได้ง่ายขึ้น เข้าใจถึงพฤติกรรมของลูกค้ากลุ่มย่อยของลูกค้าในกลุ่มใหญ่ที่แบ่งออกมา จากวิธีค่าเฉลี่ย k กลุ่ม สามารถจัดกลุ่มลูกค้าออกมาได้ทั้งหมด 5 กลุ่มที่แตกต่างต่างกัน คือ กลุ่ม Best Customer, Loyal Big Spenders Customers, At Risk, Short-Term Inactive Customer และ Long-Term Inactive Customer

ผลจากงานวิจัยช่วยให้เข้าใจว่าลูกค้าแต่ละคนมีพฤติกรรมอย่างไร และภาพรวมพฤติกรรมลูกค้าของร้านค้าว่ามีลักษณะเป็นอย่างไร ทราบถึงกลุ่มลูกค้าที่มีศักยภาพที่ควรให้ความสำคัญ เพื่อพัฒนากลยุทธ์การตลาดที่เหมาะสม สร้างความสัมพันธ์ที่ดีและเพิ่มโอกาสการขาย โดยการให้รางวัลทางการตลาดที่เหมาะสมตามความสนใจของกลุ่มลูกค้านั้นๆ เช่น การให้ส่วนลดพิเศษที่อาจจะแตกต่างกันตามความสำคัญของลูกค้า หรือแนะนำโปรโมชั่นพิเศษ เป็นต้น รวมถึงสามารถจัดเรียงลำดับความสำคัญของลูกค้าแต่ละกลุ่มได้ เพื่อสร้างความพึงพอใจของลูกค้า เพื่อให้ลูกค้ากลับมาสั่งซื้อสินค้าซ้ำและเพิ่มยอดใช้จ่ายและสามารถเพิ่มสัดส่วนลูกค้าที่มีศักยภาพในร้านค้าให้มากขึ้น

5.2 ข้อเสนอแนะ

5.2.1 ในงานวิจัยนี้จัดกลุ่มลูกค้าโดยใช้เทคนิค K-Means ซึ่งในปัจจุบันมีหลายเทคนิคที่ช่วยในการจัดกลุ่มข้อมูล ดังนั้นแนะนำให้ทดลองใช้เทคนิคอื่นๆในการจัดกลุ่มซึ่งอาจจะมีประสิทธิภาพที่ดีกว่า

5.2.2 ในงานวิจัยนี้เป็นการจัดกลุ่มลูกค้าโดยการใช้ค่า Recency, Frequency, Monetary เพียง 3 ตัวแปร ซึ่งในการจัดกลุ่มลูกค้านั้นอาจจะเพิ่มตัวแปรอื่นๆเข้าไปในการจัดกลุ่มเพื่อให้ได้ประสิทธิภาพที่มากขึ้น หรือมุมมองของลูกค้าที่มากขึ้น เช่น การเพิ่มระยะเวลาการเป็นลูกค้าเข้ามา

(Tenure) เพราะลูกค้าแต่ละคนนั้นมีระยะเวลาในการเริ่มเป็นลูกค้าที่แตกต่างกัน หรือที่เรียกว่า RFMT

เอกสารอ้างอิง

1. Meta for Business. 2023. **Facebook Digital Consumer Insight**
   สืบค้นเมื่อ 20 มิถุนายน 2565, [online]. Available From:
   https://www.facebook.com/business/m/sync-southeast-asia
2. Think with Google. 2023. **Retail guide**
   สืบค้นเมื่อ 20 มิถุนายน 2565, [online]. Available From:
   https://www.thinkwithgoogle.com/intl/th-th/consumer-insights/consumer-journey/retail-guide/
3. Thanachart Ritbumroong. 2020 **Customer Segmentation with Clustering**.
   สืบค้นเมื่อ 15 พฤศจิกายน 2564, [online]. Available From:
   https://medium.com/madt-and-bads-nida/customer-analytics
4. Thanachart Ritbumroong. 2019 **Customer Lifetime Value**.
   สืบค้นเมื่อ 15 พฤศจิกายน 2564 [online]. Available From:
   https://medium.com/madt-and-bads-nida/customer-lifetime-value
5. Thanachart Ritbumroong. 2017 **Building Customer Single View**.
   สืบค้นเมื่อ 15 พฤศจิกายน 2564 [online]. Available From:
   https://thanachart-rit.medium.com/building-customer-single-view
6. Shopee Thailand. 2022. **About Shopee**
   สืบค้นเมื่อ 15 พฤศจิกายน 2564 [online]. Available From: https://careers.shopee.co.th/about
7. S Siddharth S. Singh, , Dipak C. Jain. 2017. **“Measuring Customer Lifetime Value” In Review of Marketing Research.**
8. Mulika H. 2020. **การทำ Machine Learning ด้วย Clustering Model**. สืบค้นเมื่อ 15 พฤศจิกายน 2564 [online]. Available From:
   https://medium.com/tni-university/machine-learning--clustering-model
9. สายชล สินสมบูรณ์ทอง. 2560. **การทำเหมืองข้อมูลเล่ม 1** พิมพ์ครั้งที่ 2. กรุงเทพฯ จามจุรีโปรดักส์
10. Everitt B, Landau S, Leese Morven M and Stahl D. 2011. **Cluster Analysis 5 th Edition**. UK: John Wiley& Sons.

เอกสารอ้างอิง (ต่อ)

11. Yash Parikh and Eman Abdelfattah. 2020. **Clustering Algorithms and RFM Analysis Performed on Retail Transactions.** Published in: 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON).
12. Ananthi Sheshasaayeel and L. Logeshwari. 2018. **Implementation of Clustering Technique Based RFM Analysis for Customer Behavior** in Online Transactions. Published in:**2018** 2nd International Conference on Trends in Electronics and Informatics (ICOEI)
13. Sabbir Hossain Shihab, Shyla Afroge and Sadia Zaman Mishu. 2019. **RFM Based Market Segmentation Approach Using Advanced K-means and Agglomerative Clustering: A Comparative Study** Published in: **2019** International Conference on Electrical, Computer and Communication Engineering (ECCE)
14. Mohammadreza Tavakoli, Mohammadreza Molavi, Vahid Masoumi, Majid Mobini, Sadegh Etemad and Rouhollah Rahmani. 2018. **Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study** Published in: IEEE 15th International Conference on e-Business Engineering (ICEBE)
15. **Segmenting Users of an Online Store Using Data Mining Techniques**

    Alexandrina-Mirela Pater; Stefan Vári-Kakas; Otto Poszet; Ionel Gabriel Pintea**. Published in:** 2019 15th International Conference on Engineering of Modern Electric Systems (EMES)
16. Connectif. 2020. **What Are RFM Scores and How To Calculate Them.** สืบค้นเมื่อ 15 พฤศจิกายน 2564 [online]. Available From: https://connectif.ai/en/what-are-rfm-scores-and-how-to-calculate-them/
17. K. Tsiptsis and A. Chorianopoulos. 2009. **Data Mining Techniques in CRM: Inside Customer Segmentation.**

ภาคผนวก

ภาคผนวก ก

โปรแกรมภาษาไพทอนแสดงขั้นตอนการนำเข้าข้อมูล

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 from datetime import datetime, timedelta,date
6 from sklearn.preprocessing import StandardScaler
7
8 df=pd.read_csv('/content/Transaction data for RFM analysis test311222.csv')
9 df.head()
```

	OrderNo	DateTime	Qty	Sales	CustomerName	ProductName	ProductVariants
0	2301015XAEY57V	31-12-22 23:46	1	344	Cust03967	PRD10071	PRV10194
1	2301015V4JKUTY	31-12-22 23:06	1	498	Cust43385	PRD10171	PRV10342
2	2212315NAFN0EG	31-12-22 21:05	1	205	Cust07387	PRD10030	PRV10027
3	2212315MX6MEPB	31-12-22 20:58	1	1094	Cust23450	PRD10003	PRV10029
4	2212315MXYH8RJ	31-12-22 20:58	1	452	Cust28334	PRD10100	PRV10615

โปรแกรมภาษาไพทอนแสดงขั้นตอนการเตรียมข้อมูล

```
[3]  1 #ตรวจสอบ Type ของข้อมูล
     2 df.dtypes
```

```
OrderNo           object
DateTime          object
Qty                int64
Sales              int64
CustomerName      object
ProductName       object
ProductVariants   object
dtype: object
```

```
[4]  1 #เปลี่ยน Type ของคอลัมน์ Datetime
     2 df['DateTime']=pd.to_datetime(df['DateTime'])
```

```
[5]  1 #ตรวจสอบค่า Null
     2 df.isnull().sum()
```

```
OrderNo               0
DateTime              0
Qty                   0
Sales                 0
CustomerName        319
ProductName           0
ProductVariants   14449
dtype: int64
```

โปรแกรมภาษาไพทอนแสดงขั้นตอนการเตรียมข้อมูล (ต่อ)

```
[6]  1 #ลบค่า Null ออกจาก CustomerName
     2 df = df.dropna(subset=['CustomerName'])
```

```
     1 #ตรวจสอบค่า Null หลังจากทำการลบออก
     2 df.isnull().sum()
```

```
OrderNo              0
DateTime             0
Qty                  0
Sales                0
CustomerName         0
ProductName          0
ProductVariants    14384
dtype: int64
```

```
[8]  1 #แสดงค่า Duplicate
     2 df[df.duplicated(keep=False)]
```

	OrderNo	DateTime	Qty	Sales	CustomerName	ProductName	ProductVariants
1634	221213HTHXE623	2022-12-13 13:37:00	2	2495	Cust22761	PRD10358	NaN
1635	221213HTHXE623	2022-12-13 13:37:00	2	2495	Cust22761	PRD10358	NaN
1636	221213HTHXE623	2022-12-13 13:37:00	2	2495	Cust22761	PRD10358	NaN
3947	221122QAX3XMWY	2022-11-22 14:40:00	1	622	Cust22871	PRD10004	PRV10004
3950	221122QAX3XMWY	2022-11-22 14:40:00	1	622	Cust22871	PRD10004	PRV10004
...	...	...	...	...	...	...	...
87462	200522RWNC00JW	2020-05-22 21:21:00	1	972	Cust55024	PRD10895	NaN
87463	200522RWNC00JW	2020-05-22 21:21:00	1	972	Cust55024	PRD10895	NaN
87464	200522RWNC00JW	2020-05-22 21:21:00	1	972	Cust55024	PRD10895	NaN
88182	200401C3VB24GJ	2020-01-04 12:01:00	1	140	Cust45882	PRD10378	NaN
88183	200401C3VB24GJ	2020-01-04 12:01:00	1	140	Cust45882	PRD10378	NaN

394 rows × 7 columns

```
[9]  1 # Remove duplicate
     2 print("Before remove :", df.duplicated().sum())
     3 df.drop_duplicates(inplace=True)
     4 print("After remove :", df.duplicated().sum())
```

```
Before remove : 247
After remove : 0
```

โปรแกรมภาษาไพทอนแสดงขั้นตอนการเตรียมข้อมูล (ต่อ)

```
[20]  1 #เลือกเฉพาะคอลัมน์ 'CustomerName','DateTime','OrderNo','Sales'
      2 df2 = df[['CustomerName','DateTime','OrderNo','Sales']]
      3 df2
```

	CustomerName	DateTime	OrderNo	Sales
0	Cust03967	2022-12-31 23:46:00	2301015XAEY57V	344
1	Cust43385	2022-12-31 23:06:00	2301015V4JKUTY	498
2	Cust07387	2022-12-31 21:05:00	2212315NAFN0EG	205
3	Cust23450	2022-12-31 20:58:00	2212315MX6MEPB	1094
4	Cust28334	2022-12-31 20:58:00	2212315MXYH8RJ	452
...	...	...	...	...
88233	Cust52940	2020-02-18 21:55:00	200218NE9KVA39	170
88234	Cust07556	2020-02-17 12:26:00	200217HX2PBCDC	200
88235	Cust47604	2020-10-02 11:43:00	20021008CBW2BB	240
88236	Cust38566	2020-01-30 17:33:00	2001302A5C0KW1	2600
88237	Cust38566	2020-01-30 17:33:00	2001302A5C0KW1	2600

87672 rows × 4 columns

```
[21]  1 df2.dtypes
```

```
CustomerName            object
DateTime        datetime64[ns]
OrderNo                 object
Sales                    int64
dtype: object
```

โปรแกรมภาษาไพทอนแสดงขั้นตอนการเตรียมข้อมูล (ต่อ)

```
[22]  1 #ให้วันที่ล่าสุดในข้อมูลการขายเป็น Today date
      2 today_date = df2['DateTime'].max()
      3 today_date
```

```
Timestamp('2022-12-31 23:46:00')
```

```
[23]  1 #จำนวนลูกค้าทั้งหมดที่มีการสั่งซื้อ
      2 df2.CustomerName.nunique()
```

```
56344
```

```
[24]  1 df2.isnull().sum()
```

```
CustomerName    0
DateTime        0
OrderNo         0
Sales           0
dtype: int64
```

```
1 #คำนวณค่า RFM
2
3 df_cust = df2.groupby("CustomerName").agg({"DateTime": lambda x: (today_date - x.max()).days,
4                                            "OrderNo": lambda x: x.nunique(),
5                                            "Sales": lambda x: x.sum()}).reset_index()
6
7 df_cust
```

	CustomerName	DateTime	OrderNo	Sales
0	Cust00001	698	1	1132
1	Cust00002	95	1	617
2	Cust00003	144	1	286
3	Cust00004	780	1	1235
4	Cust00005	510	1	182
...	...	...	...	...
56339	Cust58202	219	1	366
56340	Cust58203	319	1	99
56341	Cust58204	870	1	1370
56342	Cust58205	663	1	82
56343	Cust58206	746	1	226

56344 rows × 4 columns

โปรแกรมภาษาไพทอนแสดงขั้นตอนการเตรียมข้อมูล (ต่อ)

```
1 df_cust.columns = ["CustomerName","Recency", "Frequency", "Monetary"]
2 df_cust
```

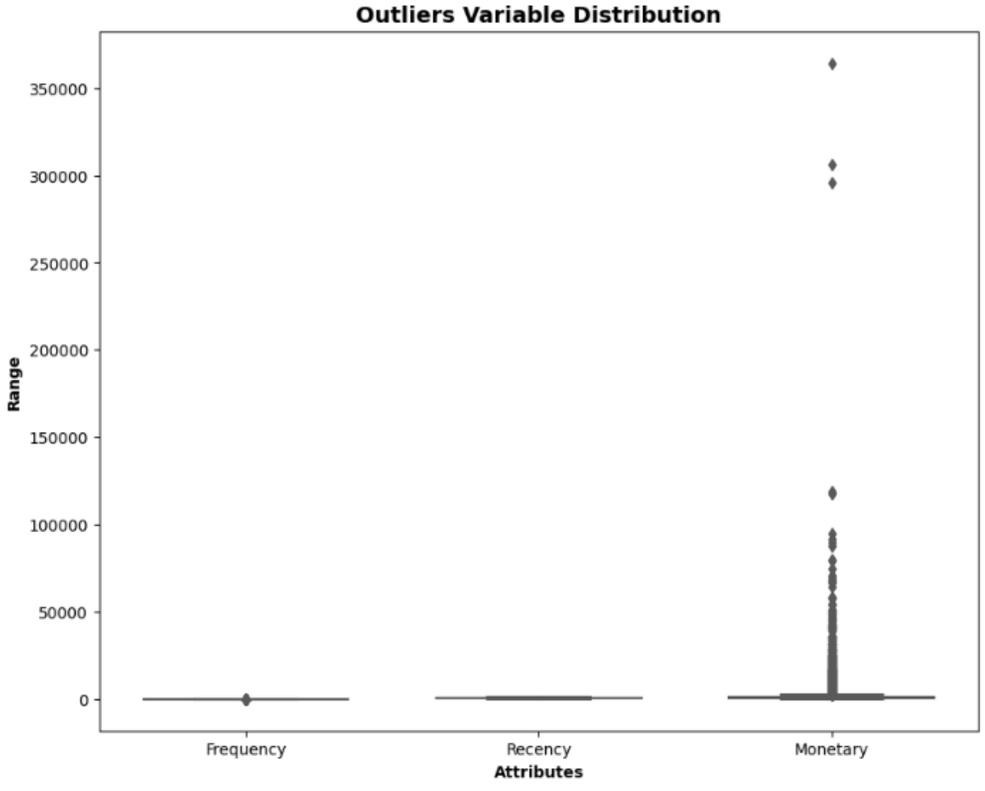
	CustomerName	Recency	Frequency	Monetary
0	Cust00001	698	1	1132
1	Cust00002	95	1	617
2	Cust00003	144	1	286
3	Cust00004	780	1	1235
4	Cust00005	510	1	182
...	...	...	...	...
56339	Cust58202	219	1	366
56340	Cust58203	319	1	99
56341	Cust58204	870	1	1370
56342	Cust58205	663	1	82
56343	Cust58206	746	1	226

56344 rows × 4 columns

โปรแกรมภาษาไพทอนแสดงขั้นตอนการเตรียมข้อมูล (ต่อ)

```python
# Outlier Analysis of Frequency Recency and Monetary

attributes = ['Frequency','Recency','Monetary']
plt.rcParams['figure.figsize'] = [10,8]
sns.boxplot(data = df_cust[attributes], orient="v", palette="Set2" ,whis=1.5,saturation=1, width=0.7)
plt.title("Outliers Variable Distribution", fontsize = 14, fontweight = 'bold')
plt.ylabel("Range", fontweight = 'bold')
plt.xlabel("Attributes", fontweight = 'bold')
```

```
Text(0.5, 0, 'Attributes')
```



โปรแกรมภาษาไพทอนแสดงขั้นตอนการเตรียมข้อมูล (ต่อ)

```
1 df_cust.shape
```

```
(56344, 4)
```

```
[35] 1 # Standard scaler
     2
     3 df_scaled = df_cust[['Recency', 'Frequency', 'Monetary']]
     4
     5 df_cust_scaled = StandardScaler().fit_transform(df_scaled)
     6
```

```
[36] 1 df_cust_scaled = pd.DataFrame(df_cust_scaled)
     2 df_cust_scaled.columns = ['Recency', 'Frequency', 'Monetary']
     3 df_cust_scaled.head()
```

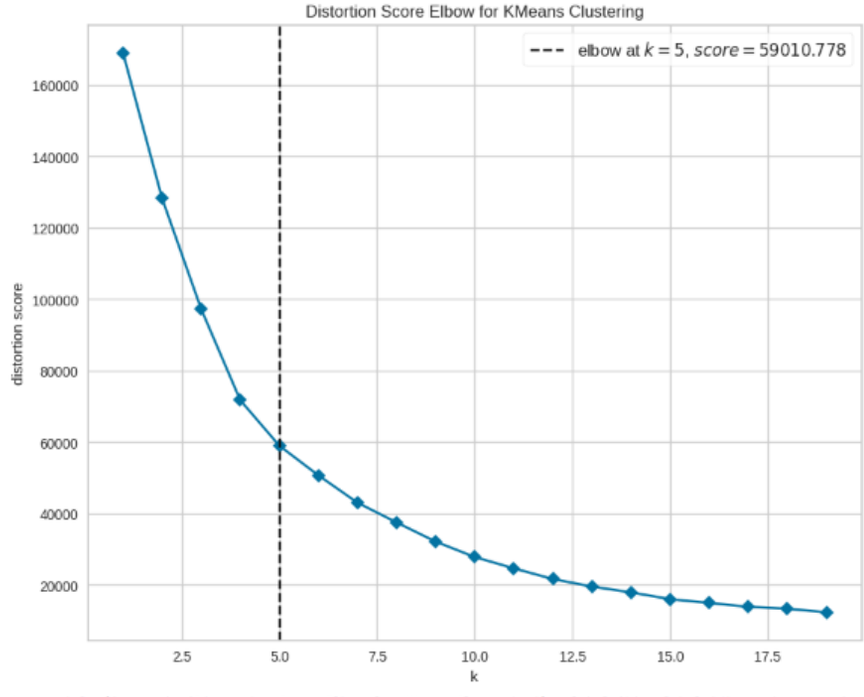
	Recency	Frequency	Monetary
0	1.192838	-0.201759	0.033856
1	-1.197326	-0.201759	-0.112740
2	-1.003100	-0.201759	-0.206959
3	1.517868	-0.201759	0.063175
4	0.447646	-0.201759	-0.236563

โปรแกรมภาษาไพทอนแสดงขั้นตอนการแบ่งกลุ่มด้วยวิธีค่าเฉลี่ย k กลุ่ม

```
[37]  1 X=df_cust_scaled[['Recency', 'Frequency', 'Monetary']]
```

```
1 from yellowbrick.cluster import KElbowVisualizer
2 from sklearn.cluster import KMeans
3 plt.figure(figsize= (10,8))
4 # Instantiate the clustering model and visualizer
5 model = KMeans()
6 visualizer = KElbowVisualizer(model, k=(1,20), timings=False)
7
8 visualizer.fit(X)        # Fit the data to the visualizer
9 visualizer.show()
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to '
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to '
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to '
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to '
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to '
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to '
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to '
  warnings.warn(
```




```
<Axes: title={'center': 'Distortion Score Elbow for KMeans Clustering'}, xlabel='k', ylabel='distortion score'>
```

โปรแกรมภาษาไพทอนแสดงขั้นตอนการแบ่งกลุ่มด้วยวิธีค่าเฉลี่ย k กลุ่ม (ต่อ)

```
1 from sklearn.cluster import KMeans
2 model = KMeans(n_clusters=5,random_state=18)
3 model.fit(X)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning:
  warnings.warn(
```

```
KMeans(n_clusters=5, random_state=18)
```

```
[40]  1 model.cluster_centers_
```

```
array([[-7.17539338e-01, -8.63459626e-02, -3.14585078e-02],
       [ 9.76690978e-01, -1.36778468e-01, -1.14239589e-01],
       [-6.08567814e-01,  2.79243275e+00,  1.45915024e+00],
       [-1.06521447e+00,  1.18807633e+01,  1.06283092e+01],
       [-1.42458278e+00,  4.27451752e+01,  9.14088125e+01]])
```

```
[41]  1 model.labels_
```

```
array([1, 0, 0, ..., 1, 1, 1], dtype=int32)
```

```
[42]  1 #เพิ่มคอลัมน์ Cluster เข้าใน df_cust
      2
      3 df_cust['cluster']=model.labels_
      4
      5 df_cust.head(10)
```

	CustomerName	Recency	Frequency	Monetary	cluster
0	Cust00001	698	1	1132	1
1	Cust00002	95	1	617	0
2	Cust00003	144	1	286	0
3	Cust00004	780	1	1235	1
4	Cust00005	510	1	182	1
5	Cust00006	654	1	285	1
6	Cust00007	200	1	932	0
7	Cust00008	191	1	79	0
8	Cust00009	373	1	69	0
9	Cust00010	340	1	265	0

โปรแกรมภาษาไพทอนแสดงขั้นตอนการแบ่งกลุ่มด้วยวิธีค่าเฉลี่ย k กลุ่ม (ต่อ)

```
1 df_cust.describe()
```

	Recency	Frequency	Monetary	cluster
count	56344.000000	56344.000000	56344.000000	56344.000000
mean	397.066129	1.155030	1013.061515	0.486937
std	252.286254	0.768397	3513.098058	0.565695
min	0.000000	1.000000	0.000000	0.000000
25%	189.000000	1.000000	197.000000	0.000000
50%	371.000000	1.000000	390.000000	0.000000
75%	580.000000	1.000000	994.000000	1.000000
max	1092.000000	79.000000	364336.000000	4.000000

[45]
```
1 df_cust.groupby("cluster").agg(["mean","median","count","max","min"])
```

```
<ipython-input-45-9e6208e5b8db>:1: FutureWarning: ['CustomerName'] did not aggregate successfully. If any error is raised this will ra
  df_cust.groupby("cluster").agg(["mean","median","count","max","min"])
```

	Recency					Frequency					Monetary				
	mean	median	count	max	min	mean	median	count	max	min	mean	median	count	max	min
cluster															
0	216.065220	215.0	30788	464	0	1.088703	1.0	30788	2	1	902.631610	419.0	30788	13856	0
1	643.496887	616.0	23770	1092	427	1.049937	1.0	23770	3	1	611.762137	304.0	23770	18672	0
2	243.510914	200.0	1695	1026	0	3.301475	3.0	1695	10	1	6139.719764	3468.0	1695	46572	75
3	128.329545	72.5	88	834	1	10.284091	10.0	88	36	1	38351.022727	35969.0	88	118913	3874
4	37.666667	24.0	3	71	18	34.000000	19.0	3	79	4	322138.333333	306320.0	3	364336	295759

[46]
```
1 df_cust['cluster'].count()
```

```
56344
```

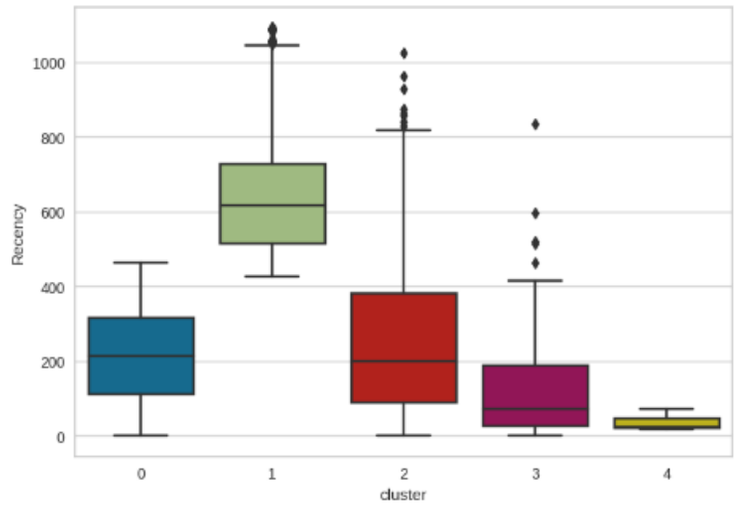
[47]
```
1 df_cust['cluster'].value_counts()
```

```
0    30788
1    23770
2     1695
3       88
4        3
Name: cluster, dtype: int64
```

โปรแกรมภาษาไพทอนแสดงขั้นตอนการแบ่งกลุ่มด้วยวิธีค่าเฉลี่ย k กลุ่ม (ต่อ)

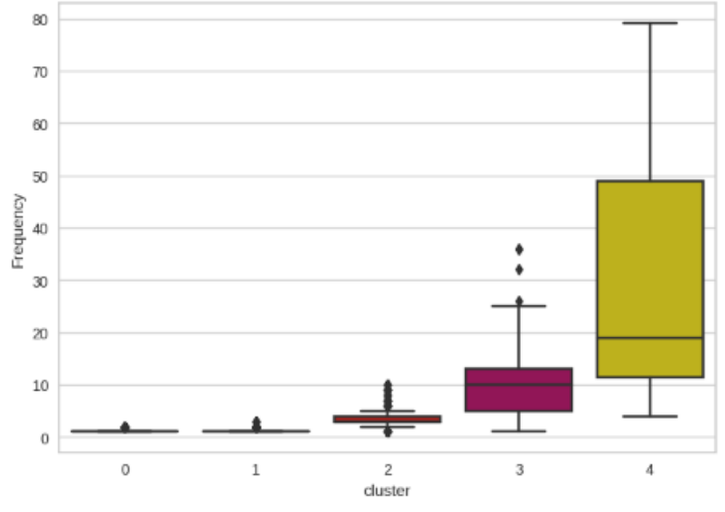
```
1 sns.boxplot(x='cluster',y='Recency', data=df_cust)
```

```
<Axes: xlabel='cluster', ylabel='Recency'>
```



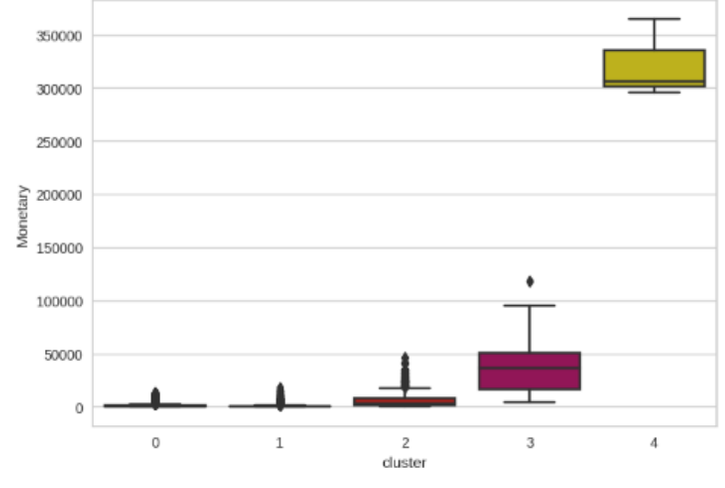

```
[49] 1 sns.boxplot(x='cluster',y='Frequency', data=df_cust)
```

```
<Axes: xlabel='cluster', ylabel='Frequency'>
```



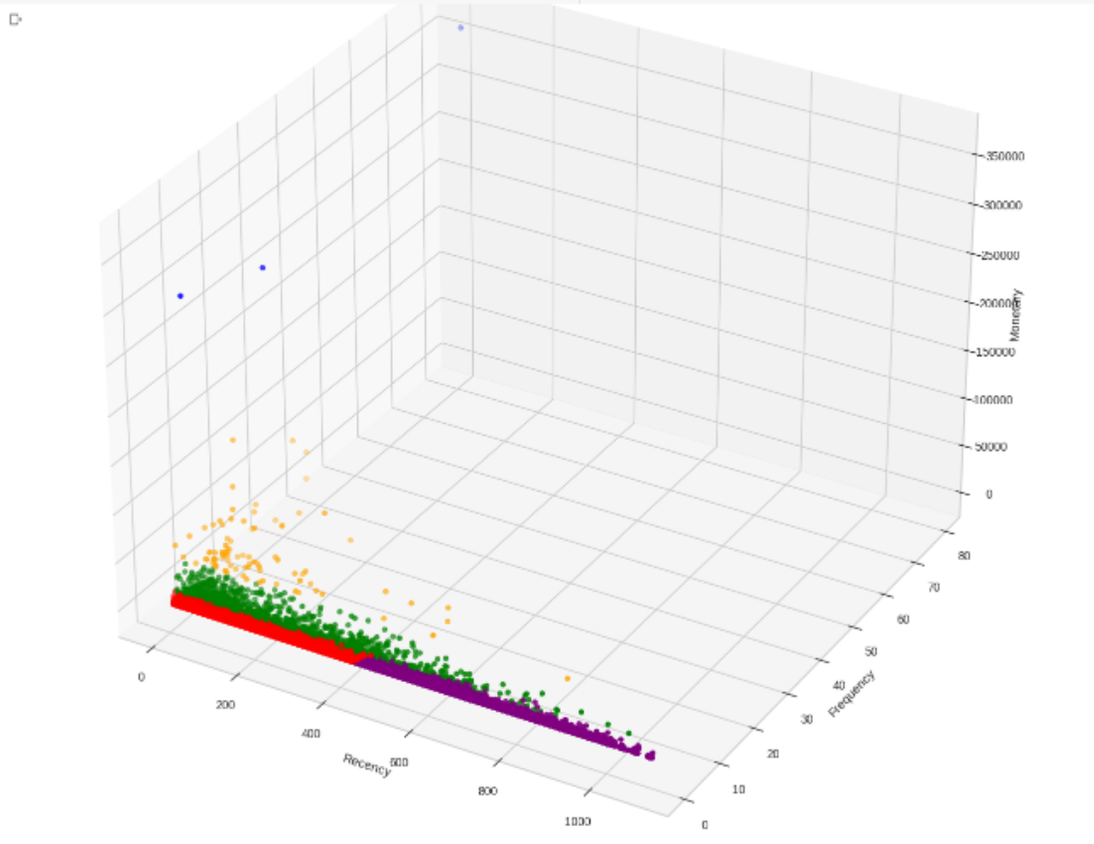

```
1 sns.boxplot(x='cluster',y='Monetary', data=df_cust)
```

```
<Axes: xlabel='cluster', ylabel='Monetary'>
```



โปรแกรมภาษาไพทอนแสดงขั้นตอนการแบ่งกลุ่มด้วยวิธีค่าเฉลี่ย k กลุ่ม (ต่อ)

```python
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import numpy as np

colors = np.array(["red", "purple", "green", "orange","blue"])
fig = plt.figure(figsize=(15, 15))
fig.suptitle('Plot of Customer\'s Distribution', fontsize=25)
ax = plt.axes(projection='3d')
scatter = ax.scatter3D(df_cust["Recency"], df_cust["Frequency"], df_cust["Monetary"], marker='o', c=colors[df_cust["cluster"].tolist()])

ax.set_xlabel('Recency')
ax.set_ylabel('Frequency')
ax.set_zlabel('Monetary')

# Create custom legend with cluster names
legend_labels = ['Cluster 0', 'Cluster 1', 'Cluster 2', 'Cluster 3', 'Cluster 4']
custom_legend = [plt.Line2D([0], [0], marker='o', color='w', markerfacecolor=color, markersize=10) for color in colors]
ax.legend(custom_legend, legend_labels)

plt.show()

```



ภาคผนวก ข

โปรแกรมภาษาไพทอนแสดงขั้นตอนการแบ่งกลุ่มด้วยค่าคะแนนแบบดั้งเดิม

```
[ ]  1 rfm = df_cust[['CustomerName','Recency', 'Frequency', 'Monetary']]
```

```
1 rfm
```

	CustomerName	Recency	Frequency	Monetary
0	Cust00001	698	1	1132
1	Cust00002	95	1	617
2	Cust00003	144	1	286
3	Cust00004	780	1	1235
4	Cust00005	510	1	182
...	...	...	...	...
56339	Cust58202	219	1	366
56340	Cust58203	319	1	99
56341	Cust58204	870	1	1370
56342	Cust58205	663	1	82
56343	Cust58206	746	1	226

56344 rows × 4 columns

```
[58]  1 rfm["Frequency"].describe([0.01,0.05,0.10,0.25,0.50,0.90,0.95,0.98,0.99]).T
```

```
count    56344.000000
mean         1.155030
std          0.768397
min          1.000000
1%           1.000000
5%           1.000000
10%          1.000000
25%          1.000000
50%          1.000000
90%          1.000000
95%          2.000000
98%          3.000000
99%          4.000000
max         79.000000
Name: Frequency, dtype: float64
```

```
[57]  1 #แบ่งช่วง Frequency ออกเป็น 5ช่วง
      2
      3 cut_bins = [0,1,2,3,4,79]
      4
      5 rfm["FrequencyScore"] = pd.cut(rfm["Frequency"], bins = cut_bins, labels = [1, 2, 3, 4, 5])
      6 rfm["FrequencyScore"].value_counts()
```

```
1    50740
2     4177
3      856
5      310
4      261
Name: FrequencyScore, dtype: int64
```

โปรแกรมภาษาไพทอนแสดงขั้นตอนการแบ่งกลุ่มด้วยค่าคะแนนแบบดั้งเดิม(ต่อ)

```
#แบ่งช่วง Monetaryออกเป็น 5ช่วง

rfm["MonetaryScore"] = pd.qcut(rfm['Monetary'], 5, labels = [1, 2, 3, 4, 5])
rfm.head()
```

	CustomerName	Recency	Frequency	Monetary	RecencyScore	FrequencyScore	MonetaryScore
0	Cust00001	698	1	1132	1	1	4
1	Cust00002	95	1	617	5	1	4
2	Cust00003	144	1	286	5	1	2
3	Cust00004	780	1	1235	1	1	4
4	Cust00005	510	1	182	2	1	2

```
[59] (rfm['RecencyScore'].astype(str) +
      rfm['FrequencyScore'].astype(str) +
      rfm['MonetaryScore'].astype(str)).head()
```

```
0    114
1    514
2    512
3    114
4    212
dtype: object
```

```
[60] rfm["RFM_SCORE"] = rfm['RecencyScore'].astype(str) + rfm['FrequencyScore'].astype(str) + rfm['MonetaryScore'].astype(str)
     rfm.head()
```

	CustomerName	Recency	Frequency	Monetary	RecencyScore	FrequencyScore	MonetaryScore	RFM_SCORE
0	Cust00001	698	1	1132	1	1	4	114
1	Cust00002	95	1	617	5	1	4	514
2	Cust00003	144	1	286	5	1	2	512
3	Cust00004	780	1	1235	1	1	4	114
4	Cust00005	510	1	182	2	1	2	212

```
rfm.dtypes
```

```
CustomerName        object
Recency              int64
Frequency            int64
Monetary             int64
RecencyScore      category
FrequencyScore    category
MonetaryScore     category
RFM_SCORE           object
dtype: object
```

```
[62] rfm["RFM_SCORE"] = rfm["RFM_SCORE"].astype(int)
```

```
[63] rfm.dtypes
```

```
CustomerName        object
Recency              int64
Frequency            int64
Monetary             int64
RecencyScore      category
FrequencyScore    category
MonetaryScore     category
RFM_SCORE            int64
dtype: object
```

```
# Define the mapping of scores to segments
segment_mapping = {
    "Champions": [555, 554, 544, 545, 454, 455, 445],
    "Loyal Customer": [543, 444, 435, 355, 354, 345, 344, 335],
    "Potential Loyalists": [553, 551, 552, 541, 542, 533, 532, 531, 452, 451, 442, 441, 431, 453, 433, 432, 423, 353, 352, 351, 342, 341, 333, 323],
    "New Customers": [512, 511, 422, 421, 412, 411, 311],
    "Promising": [525, 524, 523, 522, 521, 515, 514, 513, 425, 424, 413, 414, 415, 315, 314, 313],
    "Need Attention": [535, 534, 443, 434, 343, 334, 325, 324],
    "About To Sleep": [331, 321, 312, 221, 213, 231, 241, 251],
    "At Risk": [255, 254, 245, 244, 253, 252, 243, 242, 235, 234, 225, 224, 153, 152, 145, 143, 142, 135, 134, 133, 125, 124],
    "Cannot Lose Them": [155, 154, 144, 214, 215, 115, 114, 113],
    "Hibernating customers": [332, 322, 233, 232, 223, 222, 132, 123, 122, 212, 211],
    "Lost customers": [111, 112, 121, 131, 141, 151]
}

# Function to map score to segment name
def map_segment(score):
    for segment, scores in segment_mapping.items():
        if score in scores:
            return segment
    return "Unknown"  # In case a score doesn't match any segment

# Apply the mapping function to create the "segment" column
rfm["RFM_CLUSTER"] = rfm["RFM_SCORE"].apply(map_segment)

# Print the updated DataFrame with the new "segment" column
print(rfm)
```

```
      CustomerName  Recency  Frequency  Monetary RecencyScore FrequencyScore  \
0        Cust00001      698          1      1132            1              1
1        Cust00002       95          1       617            5              1
2        Cust00003      144          1       286            5              1
3        Cust00004      780          1      1235            1              1
4        Cust00005      510          1       182            2              1
...            ...      ...        ...       ...          ...            ...
56339    Cust58202      219          1       366            4              1
56340    Cust58203      319          1        99            3              1
56341    Cust58204      870          1      1370            1              1
56342    Cust58205      663          1        82            1              1
56343    Cust58206      746          1       226            1              1

      MonetaryScore  RFM_SCORE            RFM_CLUSTER
0                 4        114       Cannot Lose Them
1                 4        514             Promising
2                 2        512          New Customers
3                 4        114       Cannot Lose Them
4                 2        212  Hibernating customers
...             ...        ...                   ...
56339             3        413             Promising
56340             1        311          New Customers
56341             5        115       Cannot Lose Them
56342             1        111         Lost customers
56343             2        112         Lost customers
```

```
rfm
```

	CustomerName	Recency	Frequency	Monetary	RecencyScore	FrequencyScore	MonetaryScore	RFM_SCORE	RFM_CLUSTER
0	Cust00001	698	1	1132	1	1	4	114	Cannot Lose Them
1	Cust00002	95	1	617	5	1	4	514	Promising
2	Cust00003	144	1	286	5	1	2	512	New Customers
3	Cust00004	780	1	1235	1	1	4	114	Cannot Lose Them
4	Cust00005	510	1	182	2	1	2	212	Hibernating customers
...	...	...	...	...	...	...	...	...	...
56339	Cust58202	219	1	366	4	1	3	413	Promising
56340	Cust58203	319	1	99	3	1	1	311	New Customers
56341	Cust58204	870	1	1370	1	1	5	115	Cannot Lose Them
56342	Cust58205	663	1	82	1	1	1	111	Lost customers
56343	Cust58206	746	1	226	1	1	2	112	Lost customers

56344 rows × 9 columns

```
rfm["K-MEANS_CLUSTER"]=df_cust['cluster']
rfm.head(10)
```

	CustomerName	Recency	Frequency	Monetary	RecencyScore	FrequencyScore	MonetaryScore	RFM_SCORE	RFM_CLUSTER	K-MEANS_CLUSTER
0	Cust00001	698	1	1132	1	1	4	114	Cannot Lose Them	1
1	Cust00002	95	1	617	5	1	4	514	Promising	0
2	Cust00003	144	1	286	5	1	2	512	New Customers	0
3	Cust00004	780	1	1235	1	1	4	114	Cannot Lose Them	1
4	Cust00005	510	1	182	2	1	2	212	Hibernating customers	1
5	Cust00006	654	1	285	1	1	2	112	Lost customers	1
6	Cust00007	200	1	932	4	1	4	414	Promising	0
7	Cust00008	191	1	79	4	1	1	411	New Customers	0
8	Cust00009	373	1	69	3	1	1	311	New Customers	0
9	Cust00010	340	1	265	3	1	2	312	About To Sleep	0

```
[80] pd.DataFrame(rfm[["RFM_CLUSTER", "K-MEANS_CLUSTER"]].groupby(["K-MEANS_CLUSTER", "RFM_CLUSTER"])["RFM_CLUSTER"].agg("count"))
```

		RFM_CLUSTER
K-MEANS_CLUSTER	RFM_CLUSTER	
0	About To Sleep	1724
	Cannot Lose Them	2
	Hibernating customers	67
	Need Attention	512
	New Customers	8779
	Potential Loyalists	294
	Promising	19410
1	About To Sleep	2193
	At Risk	734
	Cannot Lose Them	8341
	Hibernating customers	5696
	Lost customers	5652
	Need Attention	28
	New Customers	427
	Potential Loyalists	13
	Promising	686

ประวัติผู้เขียน

ชื่อ	นายกิตติคุณ มิ่งมงคล
วัน เดือน ปีเกิด	27 สิงหาคม 2536
ที่อยู่ปัจจุบัน	3 เจริญนคร14 แขวงคลองต้นไทร เขตคลองสาน กทม 10600
ประวัติการศึกษา	(2558) วิศวกรรมศาสตร์บัณฑิต สาขาวิศวกรรมอุตสาหการ เกรดเฉลี่ย 3.20 (มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี)
ทุนการศึกษาที่ได้รับ	ไม่มี
ผลงานทางวิชาการ	ไม่มี