



รายงานสหกิจศึกษาฉบับสมบูรณ์

การวิเคราะห์พฤติกรรมการซื้อของผู้บริโภคโดยการแบ่งกลุ่มลูกค้า

Analysis of shopping behavior by customer segmentation

นางสาวคัมภีร์ดา ภูทอง

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2565

ชื่อโครงการสหกิจศึกษา การวิเคราะห์พฤติกรรมการซื้อขายของโดยการแบ่งกลุ่มลูกค้า

ชื่อ-สกุล นักศึกษา นางสาวคัมภีรดา ภูทอง

คณะ วิศวกรรมศาสตร์ ภาควิชา วิศวกรรมคอมพิวเตอร์

ชื่อ-สกุล อาจารย์นิเทศ ผศ.ดร.ชมพูนุท เต็งเจริญ

ชื่อ-สกุล ผู้นิเทศงาน นายวุฒิพงศ์ มีจำรัส และนายพนพิชา การิกายจน์

สถานประกอบการ บริษัท เอ ไอ ซิสเต็ม จำกัด

บทคัดย่อ

ในปัจจุบัน มีแนวคิดในการประกอบธุรกิจต่างกันตามประเภทของธุรกิจที่จัดตั้งขึ้น การจำแนกประเภทของลูกค้า เป็นสิ่งที่จำเป็นต่อธุรกิจอย่างมาก โดยเฉพาะธุรกิจที่เน้นการผลิตสินค้าหรือผลิตภัณฑ์ เนื่องจากการทำธุรกิจทุกประเภทนั้น มีปัจจัยหลักคือ ค่าแรงและค่าต้นทุนในการผลิต ซึ่งหากสามารถวิเคราะห์ได้ว่า ธุรกิจนั้น มีกลุ่มลูกค้าใดบ้าง จะส่งผลให้เลือกลงทุนในการผลิตสินค้าหรือผลิตภัณฑ์ได้อย่างคุ้มค่า อีกทั้งยังสามารถวางแผนเพื่อหาวิธีการตลาดเพื่อความเหมาะสมกับกลุ่มลูกค้าในตลาดของธุรกิจนั้นๆ ซึ่งทางบริษัทที่ไปสหกิจ ได้เล็งเห็นปัญหาเหล่านี้ จึงได้จัดทำการวิเคราะห์และจำแนกลูกค้าโดยใช้ RFM model ซึ่งสามารถวิเคราะห์ความต้องการของลูกค้าได้สะดวกขึ้นตามข้อมูลที่ได้จากการรวบรวมในระหว่างการซื้อขายสินค้าหรือผลิตภัณฑ์

คำสำคัญ : ธุรกิจ, การตลาด, RFM model

Cooperative Title: Analysis of shopping behavior by customer segmentation

Student intern name: Ms.Khampeerada Phoothong

Faculty: School of Engineering **Department:** Computer Engineering

Advisor name: Asst.Prof.Dr.Chompoonuch Tengcharoen

Mentor name: Mr.Wuttipong Mejumras and Mr.Ponpicha Karikarn

Company: A I SYSTEM CO., LTD.

ABSTRACT

Business operations differ according to the type of business. However, customer segmentation is very important for all business types, especially for focusing on producing products. Refer to all business types pays attention to the wage and cost of production, the customer segmentation analysis will result in the direction of choosing the investment in which the products are worthwhile. Moreover, marketing methods that suit the customer group in that business market can plan. Therefore, this cooperative education work is analyzed and classifies customers using the RFM model, which can result in customer needs conveniently based on the information gathered during the trading.

Keywords : Business, Marketing, RFM model

กิตติกรรมประกาศ

ในรายงานสหกิจศึกษาฉบับนี้ สามารถสำเร็จลุล่วงได้อย่างสมบูรณ์ด้วยความกรุณาอย่างยิ่งจากบุคลากรในบริษัท เอ ไอ ซีเอสเต็ม จำกัด ซึ่งเป็นสถานที่ที่ข้าพเจ้าได้เข้ารับสหกิจและความรู้มากมายแก่ข้าพเจ้า และผศ. ดร. ชมพูนท เต็งเจริญ ผู้เป็นอาจารย์ที่ปรึกษาในช่วงระยะเวลาสหกิจนี้ ที่ได้สละเวลาอันมีค่าให้แก่ข้าพเจ้า เพื่อให้คำปรึกษาและแนะนำตลอดจนตรวจทานแก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่เป็นอย่างยิ่ง จนรายงานฉบับนี้สำเร็จสมบูรณ์ ข้าพเจ้าขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

สุดท้ายนี้ ข้าพเจ้าคาดหวังเป็นอย่างยิ่งว่ารายงานฉบับนี้ จะสามารถนำไปใช้งานและประยุกต์ได้ในอนาคตสำหรับผู้สนใจในเรื่องนี้

คัมภีรดา ภูทอง

สารบัญ

บทที่	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง	VI
สารบัญภาพ	VII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญ	1
1.2 วัตถุประสงค์ของการวิจัย	1
1.3 ขอบเขตของการวิจัย.....	1
1.4 วิธีการดำเนินการวิจัย	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	2
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	3
2.1 ทฤษฎีที่เกี่ยวข้อง.....	3
2.2 งานวิจัยที่เกี่ยวข้อง	9
2.3 เครื่องมือที่เกี่ยวข้อง.....	18
บทที่ 3 วิธีดำเนินการวิจัย	19
3.1 การออกแบบ Business Process Model and Notation	19
3.2 การออกแบบ User Experience และ User Interface	20
3.3 สร้างต้นแบบการจำลองการคำนวณข้อมูลเพื่อวิเคราะห์สูตรโดยอิงจาก RFM Model	21
บทที่ 4 ผลการวิจัย.....	27
4.1 ต้นแบบจำลอง RFM model.....	27
4.2 เว็บไซต์ RFM model.....	30
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	48
5.1 สรุปผลการวิจัย.....	48
5.2 ข้อเสนอแนะ	48

สารบัญ (ต่อ)

บทที่	หน้า
บรรณานุกรม.....	50
ภาคผนวก.....	52

สารบัญตาราง

ตารางที่	หน้า
1.1 การดำเนินการวิจัย	2
2.1 กลุ่มตัวอย่างในการทำ RFM model.....	5
2.2 ตัวอย่างชุดข้อมูลสังเคราะห์ของธุรกรรมลูกค้า.....	13
2.3 ค่า Silhouette สำหรับคลัสเตอร์ 3 จำนวน.....	14
2.4 ค่า Silhouette สำหรับคลัสเตอร์ 5 จำนวน.....	14
2.5 ผลวิเคราะห์ลูกค้ากับคลัสเตอร์ต่างๆ ที่นำมาเปรียบเทียบ	14
2.6 ค่า Silhouette สำหรับข้อมูลตามเวลาจริงสำหรับคลัสเตอร์ 3 จำนวน.....	17
2.7 ค่า Silhouette สำหรับข้อมูลตามเวลาจริงสำหรับคลัสเตอร์ 5 จำนวน.....	17
3.1 ตัวอย่างข้อมูลที่นำมาวิเคราะห์ RFM model.....	21
3.2 ตารางเปรียบเทียบระหว่าง PERCENTRANK.INC และ RANK.EQ/COUNT.....	23
3.3 ตัวอย่างข้อมูลหลังจากที่นำมาวิเคราะห์.....	24
3.4 กลุ่มลูกค้าเมื่อเทียบกับค่าระดับของ RFM	25
3.5 ลูกค้าที่ถูกจัดประเภทตามลำดับ.....	25
4.1 ตารางเปรียบเทียบวันเวลาจากโปรแกรม Excel และภาษา JavaScript.....	36
4.2 ตารางคำนวณความแตกต่างของวันเวลาเพื่อค้นหาสูตร	37
4.3 ตารางเปรียบเทียบก่อนและหลังตั้งค่าเวลา.....	38

สารบัญภาพ

ภาพที่	หน้า
2.1 การแบ่งกลุ่มระดับคะแนนของ RFM	3
2.2 ตัวอย่างการแบ่งกลุ่มลูกค้าตามระดับคะแนนของ RFM	4
2.3 ตารางเปรียบเทียบกลุ่มประเภทลูกค้า	6
2.4 กราฟเปรียบเทียบของหลักการพาเรโต	6
2.5 กราฟเปรียบเทียบระหว่าง PERCENTILE.INC และ PERCENTILE.EXC	8
2.6 รายละเอียดประกอบการอธิบายขั้นตอน	10
2.7 คำนวณระยะทางแบบยุคลิดสองมิติโดยใช้ทฤษฎีบทพีทาโกรัส	11
2.8 อัลกอริทึมของการคำนวณด้วยวิธี K-mean	12
2.9 แสดงการกระจายความใหม่ทางการขายที่สอดคล้องกับจำนวนธุรกรรมที่เกิดขึ้นบ่อย	13
2.10 การพล็อตจุดของค่า Silhouette สำหรับคลัสเตอร์ 3 จำนวน	15
2.11 การพล็อตจุดของค่า Silhouette สำหรับคลัสเตอร์ 5 จำนวน	15
2.12 การเปรียบเทียบความล่าช้ากับจำนวนเงินสำหรับคลัสเตอร์ 3 และ 5 จำนวน	15
2.13 แผนภาพแบบกล่องเปรียบเทียบความล่าช้าระหว่างคลัสเตอร์	16
2.14 แผนภาพแบบกล่องเปรียบเทียบความถี่ระหว่างคลัสเตอร์	16
2.15 แผนภาพแบบกล่องเปรียบเทียบจำนวนเงินที่ใช้ระหว่างคลัสเตอร์	16
3.1 BPMN ของระบบ RFM model	19
3.2 User Experience และ User Interface ของระบบ RFM model	20
4.1 ตัวอย่างข้อมูลจริงของลูกค้า	27
4.2 ตัวอย่างข้อมูลที่ผ่านการสังเคราะห์ข้อมูล	28
4.3 ข้อมูลตัวอย่างเมื่อใช้สูตร PERCENTRANK	28
4.4 ข้อมูลตัวอย่างเมื่อใช้สูตร RANK.EQ/COUNT	29
4.5 ข้อมูลตัวอย่างที่ผ่านการวิเคราะห์	29
4.6 แผนภาพส่วนที่นำเข้าไฟล์ข้อมูล	30
4.7 แผนภาพส่วนที่ส่งออกข้อมูล	32
4.8 ช่วงเวลาทั้งหมดที่สามารถเกิดขึ้นเมื่อเลือกข้อมูล	35
4.9 หน้าเว็บไซต์เมื่อเปิดครั้งแรก	42
4.10 หน้าเว็บไซต์ในส่วนตั้งค่า	42
4.11 กราฟหลังจากวิเคราะห์ RFM แสดงข้อมูลทั้งหมด	43
4.13 กราฟหลังจากวิเคราะห์ RFM แสดงข้อมูลเฉพาะช่วงเวลา	44
4.14 ตารางข้อมูลหลังจากวิเคราะห์ RFM แสดงข้อมูลเฉพาะช่วงเวลา	44
4.15 ลูกเล่นของกราฟแสดงผล	45

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
4.16 ตัวอย่างเมื่อกรอกข้อมูลในส่วนการแบ่งประเภทกลุ่มลูกค้า.....	45
4.17 หน้าต่างแจ้งเตือนเมื่อไม่ได้กรอกข้อมูล.....	45
4.18 หน้าต่างแจ้งเตือนเมื่อกรอกข้อมูลผิดพลาด	46
4.19 หน้าต่างแจ้งเตือนเมื่อกรอกข้อมูลสำเร็จ.....	46
4.20 ตารางข้อมูลก่อนการวิเคราะห์ RFM model	47
4.21 หน้าต่างที่แสดงผลเมื่อส่งออกข้อมูล.....	47
4.22 ข้อมูลภายในไฟล์ที่ส่งออกข้อมูลผลการวิเคราะห์.....	47

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญ

ในปัจจุบัน มีแนวคิดในการประกอบธุรกิจต่างกันตามประเภทของธุรกิจที่จัดตั้งขึ้น การจำแนกประเภทของลูกค้า เป็นสิ่งที่จำเป็นต่อธุรกิจอย่างมาก โดยเฉพาะธุรกิจที่เน้นการผลิตสินค้าหรือผลิตภัณฑ์ เนื่องจากการทำธุรกิจทุกประเภทนั้น มีปัจจัยหลักคือ ค่าแรงและค่าต้นทุนในการผลิต ซึ่งหากสามารถวิเคราะห์ได้ว่า ธุรกิจนั้น มีกลุ่มลูกค้าใดบ้าง จะส่งผลให้เลือกลงทุนในการผลิตสินค้าหรือผลิตภัณฑ์ได้อย่างคุ้มค่า อีกทั้งยังสามารถวางแผนเพื่อหาวิธีการตลาดเพื่อความเหมาะสมกับกลุ่มลูกค้าในตลาดของธุรกิจนั้นๆ ซึ่งทางบริษัทที่ไปสหกิจ ได้เล็งเห็นปัญหาเหล่านี้ จึงได้จัดทำการวิเคราะห์และจำแนกลูกค้าโดยใช้ RFM model ซึ่งสามารถวิเคราะห์ความต้องการของลูกค้าได้สะดวกขึ้นตามข้อมูลที่ได้จากการรวบรวมในระหว่างการซื้อขายสินค้าหรือผลิตภัณฑ์

1.2 วัตถุประสงค์ของการวิจัย

- 1.2.1 เพื่อวิเคราะห์และจำแนกประเภทของลูกค้าที่ส่งผลกระทบในการประกอบธุรกิจได้
- 1.2.2 เพื่อเป็นระบบที่ช่วยคาดการณ์แนวโน้มทางการตลาดแก่ธุรกิจได้

1.3 ขอบเขตของการวิจัย

1.3.1 Functional Requirement

- 1) ผู้ใช้สามารถนำเข้าข้อมูลที่เป็นนามสกุลไฟล์ .xlsx หรือ .xls และส่งออกข้อมูลที่เป็นนามสกุลไฟล์ .csv ได้
- 2) ผู้ใช้สามารถอ่านข้อมูลก่อนและหลังทำการวิเคราะห์ RFM model ได้
- 3) ระบบสามารถแสดงผลหลังจากวิเคราะห์ได้ว่า ลูกค้าคนใด อยู่ในประเภทไหน
- 4) ทุกครั้งที่เปลี่ยนข้อมูลที่จะอ่าน ระบบจะมีการลบข้อมูลเดิมทิ้งทุกครั้ง
- 5) ทุกครั้งที่มีความผิดพลาดของระบบ ต้องมีการแจ้งว่าไม่สำเร็จเพราะอะไร

1.3.2 Non - Functional Requirement

- 1) ผู้ใช้สามารถกำหนดช่วงระยะเวลาในการคำนวณได้
- 2) ผู้ใช้สามารถสร้างหรือแบ่งกลุ่มประเภทของลูกค้าตามช่วงระดับด้วยตัวเองได้
- 3) ระบบสามารถแสดงประเภทกลุ่มลูกค้าเป็นรูปแบบกราฟวงกลมได้

1.4 วิธีการดำเนินการวิจัย

ตารางที่ 1.1 การดำเนินการวิจัย

รายละเอียด	พ.ศ. 2565				
	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.
1. ศึกษา Business Model					
2. ศึกษา Data Analysis					
3. ศึกษา RFM Model					
4. สร้างต้นแบบการจำลองการคำนวณข้อมูลเพื่อวิเคราะห์สูตรโดยอิงจาก RFM Model					
5. ศึกษาและทดลองใช้งานเครื่องมือที่จำเป็น					
6. สร้างเว็บไซต์ RFM Model โดยใช้ข้อมูลที่กำหนดไว้					
7. สร้างเว็บไซต์ RFM Model โดยสามารถใช้ข้อมูลที่นำเข้าได้ตามที่ต้องการ					
8. จัดทำเอกสารโครงงานสหกิจ					

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 สามารถเกิดความสะดวกสบายแก่ผู้ใช้งานที่ทำธุรกิจได้

1.5.2 สามารถวิเคราะห์ และคาดการณ์แนวโน้มทางการตลาดได้ เพื่อให้หาวิธีรับมือกับปัญหาด้านความคุ้มค่าของธุรกิจได้

บทที่ 2

แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

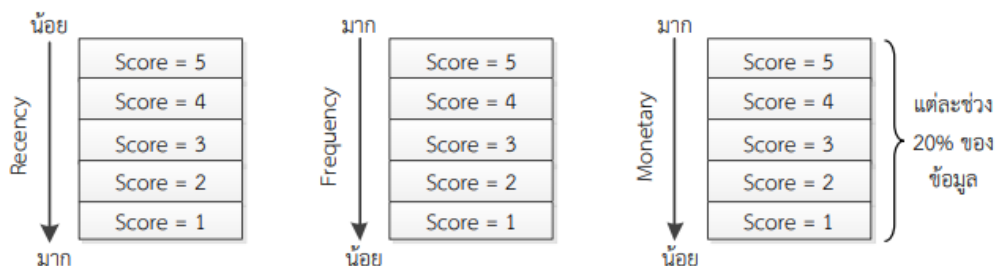
2.1.1 RFM model

RFM Model เป็นการนำประวัติการสั่งซื้อที่เกิดขึ้นจริงภายในธุรกิจมาใช้คำนวณเพื่อแบ่งลูกค้าออกเป็นกลุ่ม ๆ เพื่อให้สามารถจำแนกได้ว่า ลูกค้ากลุ่มไหน เป็นกลุ่มที่สร้างยอดขายให้ธุรกิจได้มากที่สุด, กลุ่มใดที่ธุรกิจต้องการดึงกลับมาให้ได้ในช่วงเวลาที่ห่างหาย, กลุ่มใดที่มีแนวโน้มจะกลายเป็นลูกค้าขาประจำ หรือกลุ่มใดที่ไม่จำเป็นต้องให้ความสนใจมากนัก และเมื่อเจ้าของธุรกิจสามารถระบุกลุ่มลูกค้าได้แล้ว ก็จะทำให้ความสำคัญกับตามกลุ่มลูกค้าที่สร้างยอดขาย 80% ในร้าน ตามหลักของพาเรโต ได้ไม่ยาก หรือสร้างกลยุทธ์การขายที่เหมาะสมแม่นยำและมีประสิทธิภาพมากขึ้น โดยมีกลุ่มเป้าหมายที่สนใจหรืออยู่ในช่วงกำลังตัดสินใจอยู่ ทยอยมาเป็นลูกค้าของธุรกิจที่เพิ่มขึ้น และมีค่าโฆษณาเพื่อดึงดูดลูกค้าที่ถูกละ

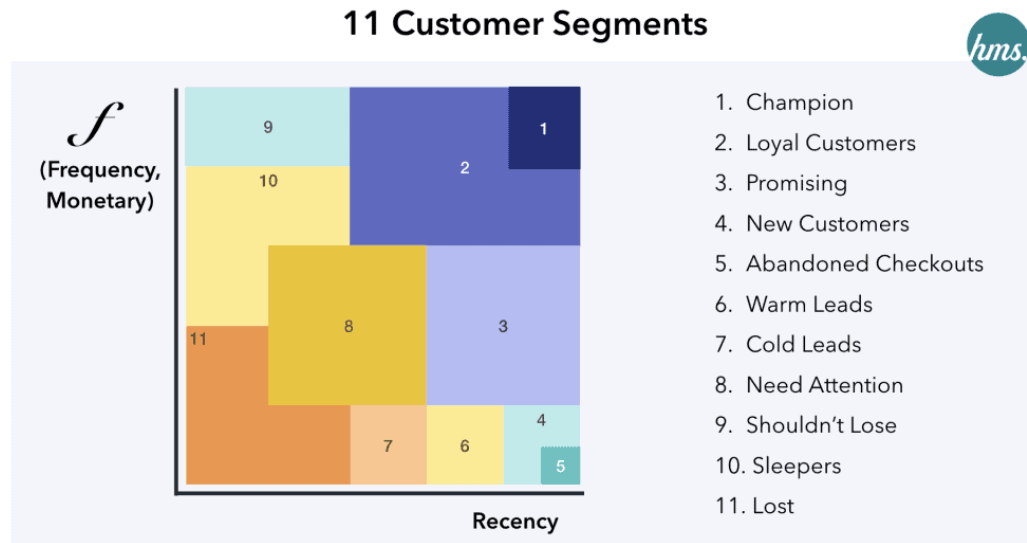
RFM Model เป็นการแบ่งกลุ่มลูกค้าโดยดูจาก 3 ตัวแปรหลัก ซึ่งประกอบด้วย

- 1) Recency คือ เวลาล่าสุดที่ใช้งานในการซื้อ – ขายสินค้า
- 2) Frequency คือ ความถี่ในการซื้อ – ขายสินค้ามีมากน้อยเพียงใด
- 3) Monetary คือ ปริมาณการใช้จ่ายสินค้าว่าใช้จำนวนเงินเท่าไร

โดยตัวแปรหลักทั้ง 3 นี้ จะถูกจัดระดับคะแนนเป็น 5 ระดับ โดยแบ่งจาก 100% เป็น 5 ส่วน สัดส่วนละ 20 % ซึ่งการแบ่งเป็นระดับคะแนนนั้น จะถูกนำไปใช้ในการแบ่งลูกค้าเป็น 11 ประเภทตามรูปแบบ RFM – model ดังภาพที่ 2.1 และภาพที่ 2.2



ภาพที่ 2.1 การแบ่งกลุ่มระดับคะแนนของ RFM



ภาพที่ 2.2 ตัวอย่างการแบ่งกลุ่มลูกค้าตามระดับคะแนนของ RFM

ในการแบ่งประเภทลูกค้านั้น สามารถปรับได้ตามความเหมาะสมของธุรกิจนั้นๆ เนื่องจากเจ้าของธุรกิจนั้น อาจต้องการทราบเพียงแค่กลุ่มลูกค้าที่ส่งผลต่อธุรกิจของเขาได้มากที่สุด อย่างไรก็ตามการกำหนดเป้าหมายเป้าหมายของธุรกิจว่าต้องการเพิ่มยอดขาย หรือการรักษาลูกค้าประจำ [1]

เมื่อการเตรียมข้อมูลได้อย่างครบถ้วนแล้ว ในขั้นตอนการวิเคราะห์ข้อมูลพฤติกรรมของลูกค้าจาก Transaction Data ที่ประกอบด้วยข้อมูลเกี่ยวกับการจำนวนเงินและเวลาที่ใช้การใช้จ่าย เพื่อแบ่งกลุ่มลูกค้า (Segmentation) ตามพฤติกรรมที่แตกต่างกัน

โดยการทำ RFM analysis สามารถวิเคราะห์ข้อมูลได้ 2 วิธีที่นิยมใช้กัน ได้แก่

1) Machine Learning เพื่อหา Macro segment

เป็นการใช้หลักสถิติในการจัดกลุ่มข้อมูล (clustering algorithms) โดยใช้วิธีคำนวณของ K-Means เพื่อแบ่งกลุ่มลูกค้าและดูสัดส่วนของลูกค้าแต่ละประเภท

โดยการคำนวณของ K-Means จะเริ่มจากการกำหนดจุดศูนย์กลางของกลุ่ม (centroid) มาจำนวน k จุด ซึ่ง k คือจำนวนกลุ่มที่คาดว่าจะได้จากการแบ่งกลุ่ม (ในกรณีที่นำมาประยุกต์กับธุรกิจ ให้กำหนดตามจำนวนของกลุ่มลูกค้าที่จะแบ่งประเภท) จากนั้นคำนวณระยะห่างระหว่างข้อมูลในแต่ละแถวกับ centroid ตามสูตรข้างต้น เพื่อใช้ในการจัดกลุ่มข้อมูล โดยระยะห่างนั้นสามารถคำนวณโดยใช้ Euclidean distance

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad \text{----- (2.1)}$$

เมื่อจัดกลุ่มครบทุกแถวข้อมูลแล้วจะมีการปรับตำแหน่งของ centroid ในแต่ละกลุ่มใหม่อีกครั้ง ซึ่งจะทวนซ้ำจนกว่าจะตรงตามเงื่อนไข

2) Traditional RFM scoring เพื่อหา Micro segment

เป็นการแบ่งกลุ่มลูกค้าด้วยวิธีการสร้าง Rule Based Segmentation หรือการกำหนดคุณสมบัติตาม RFM model ของลูกค้าแต่ละกลุ่ม ซึ่งตารางที่ 2 ในข้างต้นเป็นกลุ่มตัวอย่างที่นิยมพบเจอในการทำ RFM model [2]

ตารางที่ 2.1 กลุ่มตัวอย่างในการทำ RFM model

SEGMENT	คำอธิบาย
CHAMPIONS	เพิ่งซื้อไปไม่นานมานี้ ซื้อบ่อย และ ใช้จ่ายเยอะ
LOYAL CUSTOMERS	ใช้จ่ายเยอะ และซื้อบ่อย
NEED ATTENTION	กลุ่มที่อาจจะต้องยื่นข้อเสนอแบบจำกัดเวลาและให้ตรงกับความต้องการ หรือสิ่งที่กำลังจะต้องการ เพื่อกระตุ้นให้กลับมาซื้อซ้ำอีกครั้ง
SMALL BASKET SIZE	ใช้จ่ายไม่เยอะ แต่ซื้อบ่อย และ เพิ่งซื้อไปไม่นาน
POTENTIAL LOYALISTS	เพิ่งซื้อไปไม่นาน ใช้จ่ายปานกลางถึงสูง และซื้อมากกว่า 1 ครั้ง
NEW CUSTOMERS	เพิ่งซื้อไปไม่นาน และซื้อสินค้าไม่บ่อย
ABANDONED CHECKOUTS	กลุ่มที่อาจจะจะมีแนวโน้มที่จะอยู่ในช่วงกำลังตัดสินใจที่จะซื้อและจ่ายเงิน
PROMISING	เพิ่งซื้อไปไม่นาน และใช้จ่ายน้อย
LONG TIME BIG BUY	ใช้จ่ายเยอะ นานๆ ครั้ง
AT RISK	ใช้จ่ายเยอะ เมื่อนานมากๆ แล้ว
LOST	ใช้จ่ายน้อย ซื้อไม่บ่อย และ ซื้อครั้งล่าสุดเมื่อนานมาแล้ว

การสร้าง Rule Based Segmentation นิยมสร้างได้ 2 วิธี คือ การจัดระดับจากกลุ่มที่แบ่งเป็น 3 กลุ่มตามระดับค่าความล่าสุด ความถี่ และจำนวนเงินที่ใช้จ่าย (Recency (R) - Frequency (F) - Monetary (M)) และการจัดกลุ่มตามระดับที่อิงจากตาราง 5 x 5 จากระดับค่าความล่าสุด และความถี่รวมกับจำนวนเงินที่ใช้จ่าย (Recency (R) - Frequency and Monetary (FM))

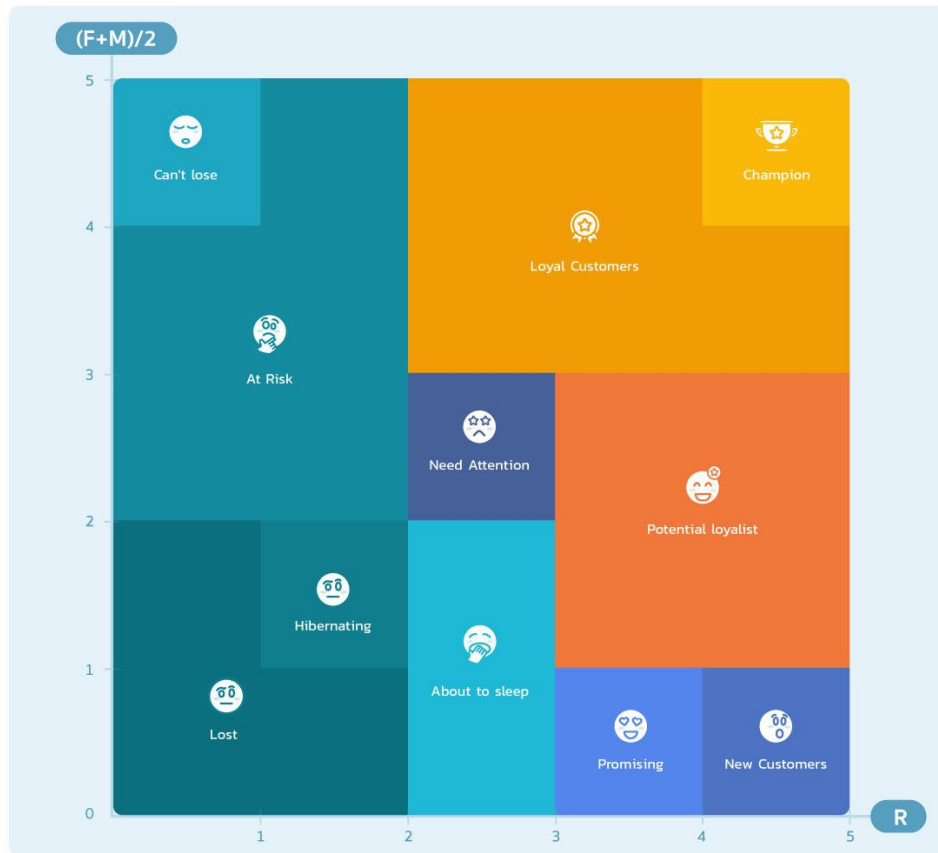
ในที่นี้ จะยกตัวอย่างหนึ่งในวิธีที่นำเสนอในข้างต้น คือ การจัดกลุ่มตามระดับที่อิงจากตาราง โดยเมื่อให้ระดับคะแนน R Score, F Score และ M Score กับลูกค้าแต่ละคน มาจัดเข้ากลุ่มด้วยตาราง 5 x 5 ช่อง โดยใช้ R แทนค่าในแกน X และใช้ (F+M)/2 แทนค่าในแกน Y ซึ่งพิกัดของแกน x และ y นั้นจะถูกจัดกลุ่มตามดังภาพที่ 2.3 [3]

ตัวอย่างการใช้คะแนน RFM ในการจัดกลุ่มลูกค้า

ลูกค้า A จัดอยู่ในกลุ่ม Potential Loyalist เมื่อมี R,F,M ที่ 4,2,1 โดยค่าแกน X คือ 4 และค่าแกน Y คือ 1.5 จาก $(2+1)/2$

ลูกค้า B จัดอยู่ในกลุ่ม At Risk เมื่อมี R,FM ที่ 2,3,5 โดยค่าแกน X คือ 2 และค่าแกน Y คือ 4 จาก $(3+5)/2$

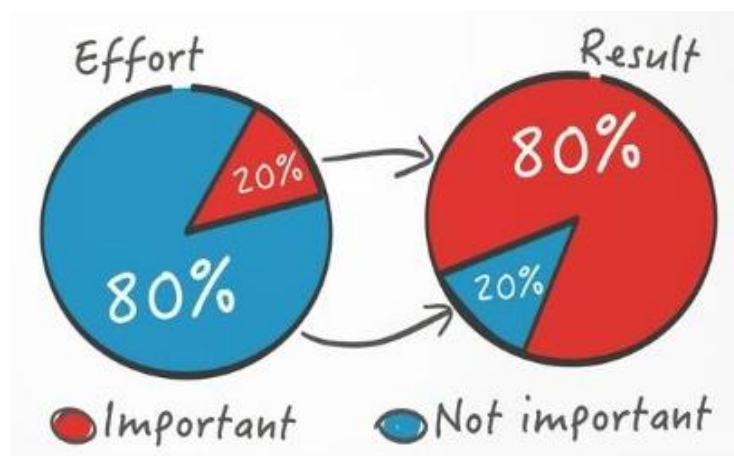
ลูกค้า C จัดอยู่ในกลุ่ม Champion เมื่อมี R,FM ที่ 5,5,5 โดยค่าแกน X คือ 5 และค่าแกน Y คือ 5 จาก $(5+5)/2$



ภาพที่ 2.3 ตารางเปรียบเทียบกลุ่มประเภทลูกค้า

2.1.2 กฎการพาเรโต (Pareto principle)

หลักการพาเรโต เป็นผลจากการสังเกตการณ์ของนักเศรษฐศาสตร์ชาวอิตาลี นามว่า Vilfredo Pareto ที่ได้เสนอแนวคิดที่ว่า การสนใจในการแก้ไขปัญหาที่มีเพียง 20% ที่สามารถส่งผลกระทบต่อผลลัพธ์ได้สูงสุดถึง 80% ตามภาพที่ 2.4



ภาพที่ 2.4 กราฟเปรียบเทียบของหลักการพาเรโต

ในการนำมาประยุกต์ใช้กับแนวคิดธุรกิจนั้น จะเป็นการให้ความสนใจกับสิ่งที่ส่งผลต่อผลลัพธ์ที่ต้องการมากที่สุด และทำสิ่งนั้นเป็นลำดับแรก เนื่องจากไม่สามารถให้ความสนใจในทุกด้านได้อย่างพร้อมกันและเป็นตัวเลือกในการตัดสินใจที่จะทำและแก้ไขปัญหาเหล่านั้น เพื่อให้ส่งผลต่อความก้าวหน้าของธุรกิจ [4]

2.1.3 เปอร์เซ็นไทล์ (Percentile)

ข้อมูลที่นำมาศึกษาในครั้งนี้ จะสามารถจัดแบ่งจำนวนข้อมูลที่มีกระจายเป็น 100 ส่วน โดยในครั้งนี้ จะใช้การคำนวณเปอร์เซ็นไทล์จากข้อมูลดิบโดยไม่สามารถทำการแจกแจงความถี่เป็นตารางได้ เนื่องจากเป็นข้อมูลที่ไม่มีความแน่นอน จึงไม่สามารถกำหนดอัตราภาคขั้นหรือช่วงห่างของข้อมูลได้อย่างชัดเจน [5]

1) วิธีคำนวณเปอร์เซ็นไทล์ที่ไม่ได้อยู่ในรูปแบบแจกแจงความถี่

- 1.1) เรียงลำดับข้อมูลที่มีค่าน้อยไปหาข้อมูลที่มีค่ามาก
- 1.2) แทนค่าตามสูตรเพื่อหาตำแหน่งของค่าเปอร์เซ็นไทล์
- 1.3) หาค่าเปอร์เซ็นไทล์จากค่าตำแหน่งของข้อมูลที่ได้ทำการเรียงลำดับ

$$P_r = r \frac{(N + 1)}{100} \text{----- (2.2)}$$

เมื่อ P_r คือ ค่าตำแหน่งของค่าเปอร์เซ็นไทล์ที่ต้องการ

r คือ ตำแหน่งที่ต้องการ โดยสามารถมีค่าได้ตั้งแต่ 1 ถึง 99

N คือ จำนวนของข้อมูลที่มีทั้งหมด

สมการ (2.2) เป็นวิธีคำนวณค่าเปอร์เซ็นไทล์ และเป็นต้นแบบของสูตรในโปรแกรม Excel คือ [6][7]

$$\text{PERCENTILE}(\text{array}, k) \text{----- (2.3)}$$

เมื่อ array คือ ช่วงของข้อมูลที่ต้องการ

k คือ ค่าเปอร์เซ็นไทล์ที่ต้องการ โดยอยู่ในช่วงเปิดที่ (0, 1)

2) วิธีคำนวณอันดับเปอร์เซ็นไทล์

ในบางกรณี ผู้ใช้งานอาจจะต้องการค้นหาค่าตำแหน่งของข้อมูลข้างต้น แต่ไม่สามารถใช้สูตรเปอร์เซ็นไทล์ (Percentile) ตามปกติได้ ซึ่งในโปรแกรม Excel ได้รับรองวิธีในการคำนวณนี้ และสามารถคำนวณได้หลายวิธี โดยการใช้แต่ละสูตรนั้น จะมีความแตกต่างกันเพียงไม่มาก

$$\text{PERCENTRANK.EXC}(\text{array}, x, [\text{significance}]) \text{----- (4)}$$

เมื่อ array คือ ช่วงของข้อมูลที่ต้องการ
 x คือ ค่าของข้อมูลที่ต้องการทราบตำแหน่ง
 significance คือ จำนวนเลขนัยสำคัญสำหรับค่าเปอร์เซ็นต์ที่ส่งกลับ

หรือ

$$\text{PERCENTRANK.INC}(\text{array}, x, [\text{significance}]) \quad \text{-----} \quad (2.5)$$

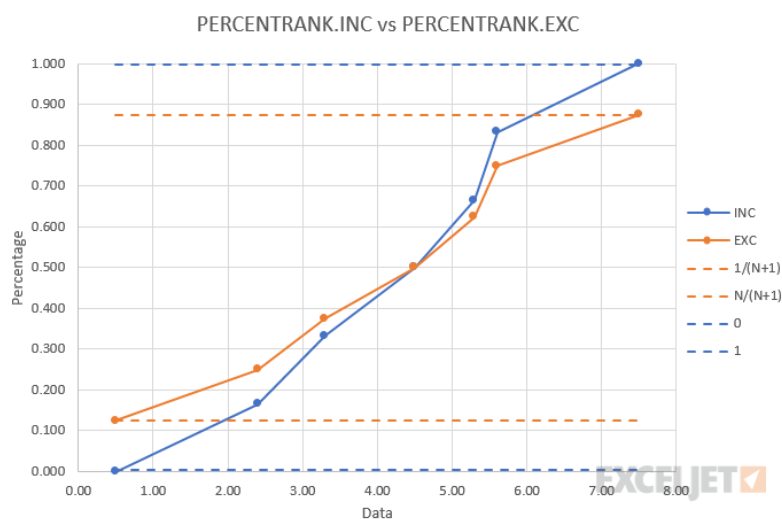
เมื่อ array คือ ช่วงของข้อมูลที่ต้องการ
 x คือ ค่าของข้อมูลที่ต้องการทราบตำแหน่ง
 significance คือ จำนวนเลขนัยสำคัญสำหรับค่าเปอร์เซ็นต์ที่ส่งกลับ

หรือ

$$\frac{\text{RANK.EQ}(\text{number}, \text{ref}, [\text{order}])}{\text{COUNT}(\text{ref})} \quad \text{-----} \quad (2.6)$$

เมื่อ number คือ ค่าตำแหน่งของข้อมูลที่ต้องการทราบ
 ref คือ ช่วงของข้อมูลที่ต้องการ
 order คือ ระบุวิธีจัดอันดับตัวเลข

RANK.EQ เป็นวิธีที่ประกอบด้วยการหาลำดับของข้อมูล (หากตัวเลขซ้ำ จะจัดอยู่ในลำดับเดียวกัน) และการนับจำนวนของกลุ่มข้อมูล ซึ่งการคำนวณด้วยวิธีนี้จะเป็นวิธีที่ค่าส่งกลับมีค่าใกล้เคียงกับสูตร PERCENTRANK.INC มากที่สุด และความแตกต่างระหว่าง PERCENTILE.INC (แบบ Inclusive) และ PERCENTILE.EXC (แบบ Exclusive) โดยการคำนวณแบบ Exclusive ทางนักสถิติมองว่ามีวิธีการคำนวณที่ตรงตามนิยามมากกว่า แต่มีปัญหาเกี่ยวกับ Percentile ที่ 0 กับ 100 เพราะไม่สามารถนำมาคำนวณได้ ซึ่งสามารถสังเกตได้จากกราฟเปรียบเทียบในภาพที่ 2.5 [8][9]



ภาพที่ 2.5 กราฟเปรียบเทียบระหว่าง PERCENTILE.INC และ PERCENTILE.EXC

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 แบบจำลอง RFM สำหรับพฤติกรรมการณ์ซื้อของลูกค้าโดยใช้อัลกอริทึม K-Means (RFM model for customer purchase behavior using K-Means algorithm)

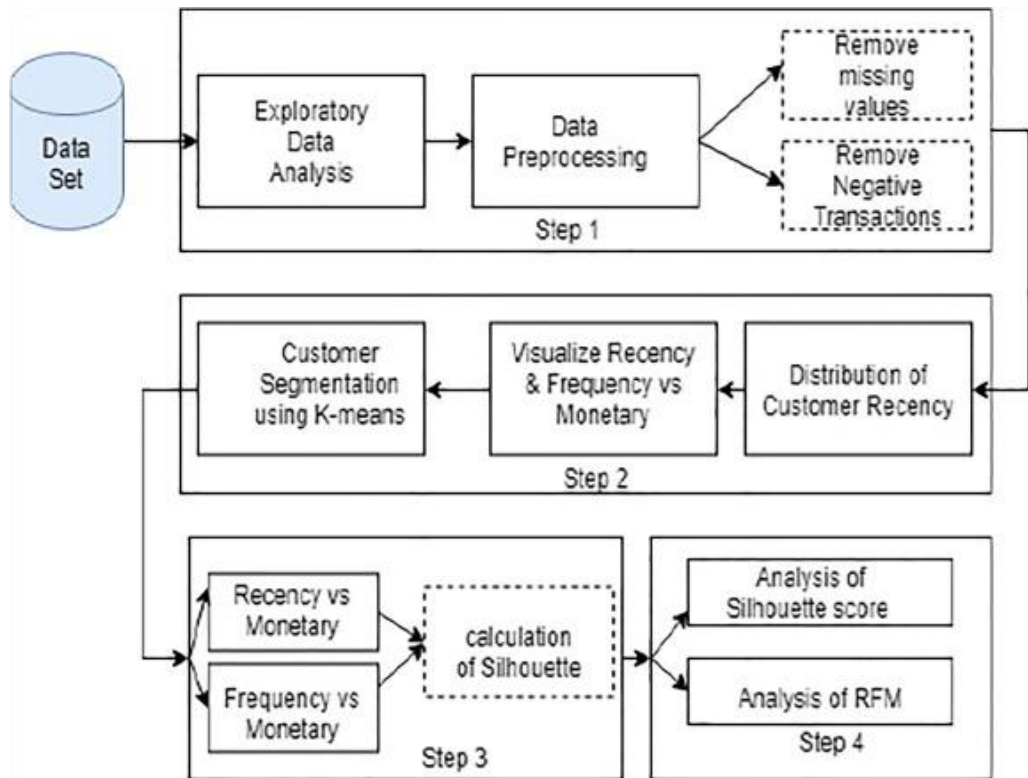
ในด้านการแบ่งส่วนข้อมูล ลูกค้าจะถูกแบ่งออกเป็นชุดของบุคคลที่มีความคล้ายคลึงกันอย่างชัดเจน คุณลักษณะบางอย่างที่เกี่ยวข้องกับการแบ่งกลุ่มลูกค้า ได้แก่ เพศ อายุ พฤติกรรมการณ์ซื้อ รายได้ และอื่นๆ ซึ่งข้อมูลดังกล่าวจัดประเภทตามพฤติกรรมการณ์ซื้อในอดีตที่สามารถนำไปสู่ผลลัพธ์เฉพาะ เช่น ยอดขายเพิ่มขึ้นและกำไรของบริษัท

ในการแข่งขันที่เพิ่มมากขึ้นและความซับซ้อนที่เพิ่มขึ้นของสภาพแวดล้อมทางธุรกิจ เป็นแนวทางที่ช่วยส่งเสริมในการดูแลลูกค้าและปรับปรุงธุรกิจสำหรับการสร้างความสัมพันธ์โดยการขยายฐานข้อมูลลูกค้าที่ทำการใดให้กว้างขึ้น โดยในการแบ่งส่วนที่โดดเด่นที่สุดใน 2 ประเภทที่ใช้ในอัลกอริทึม K-Means คือ ข้อมูลเชิงลึกเชิงคุณภาพและเชิงปริมาณ ซึ่งในขอบเขตของการศึกษาปัจจุบัน ข้อมูลเชิงลึกเชิงปริมาณถูกนำมาใช้เพื่อวัตถุประสงค์ในการจัดกลุ่มการแบ่งกลุ่ม การแบ่งกลุ่มลูกค้าที่ชัดเจนจะช่วยให้การจัดสรรทรัพยากรทางการตลาดอย่างมีประสิทธิภาพ ทำให้บริษัทสามารถกำหนดเป้าหมายกลุ่มลูกค้าเฉพาะและยังช่วยในการสร้างความสัมพันธ์กับลูกค้าในระยะยาว และอุตสาหกรรมหลักที่การแบ่งกลุ่มลูกค้าและการทำเหมืองข้อมูลสามารถนำไปใช้กับอุตสาหกรรมค้าปลีก เนื่องจากต้องใช้ข้อมูลจำนวนมากมาใช้ในการขาย การขนส่ง อัตราส่วนการบริโภค บริการจัดส่งซ้ำ และอื่นๆ อีกมากมาย นอกจากนี้ การทำเหมืองข้อมูลการค้าปลีกยังช่วยในการระบุและจำแนกพฤติกรรมของลูกค้าและรูปแบบที่เกี่ยวข้องอย่างมีประสิทธิภาพในระหว่างธุรกรรมทางธุรกิจทั้งหมด สิ่งนี้นำไปสู่การปรับปรุงการบริการลูกค้า กลยุทธ์การขายและการจัดจำหน่ายที่มีประสิทธิภาพ และอื่นๆ อีกมากมาย ในครั้งนี้มุ่งเน้นไปที่การติดตามพฤติกรรมการณ์ซื้อที่ผ่านมาของลูกค้าเป็นหลัก โดยมีจุดประสงค์เพื่อค้นหายอดขายสูงสุดที่เป็นไปได้ในพื้นที่เฉพาะ จากผลลัพธ์และตัวชี้วัดทางสถิติ บริษัทต่างๆ ในอุตสาหกรรมค้าปลีกสามารถออกแบบกลยุทธ์การขายและการตลาดต่างๆ เช่น แคมเปญส่งเสริมการขาย การขยายส่วนลดตามฤดูกาล หรือการขายแบบลอยตัว ซึ่งช่วยให้ผู้ประกอบสามารถเพิ่มยอดขายและปรับปรุงการรักษาลูกค้าได้

เพื่อให้บรรลุวัตถุประสงค์ข้างต้น การจัดกลุ่มลูกค้าและการแบ่งส่วนจะดำเนินการใช้อัลกอริทึม K-Means โดยอ้างอิงจากค่า RFM สามารถกำหนดให้เป็นการแบ่งส่วนการวิเคราะห์ลูกค้า ซึ่งไม่เพียงแต่ให้ข้อมูลเกี่ยวกับรูปแบบการซื้อบ่อยครั้งของลูกค้าเท่านั้น แต่ยังรวมถึงการซื้อล่าสุดและกำไรที่ได้รับด้วย ซึ่งในขั้นตอนแรก กลุ่มข้อมูลหรือคลัสเตอร์ จะได้รับการประเมินโดยวิธี Silhouette Analysis สำหรับเวลาความล่าช้าในการซื้อและวิธี K-Means สำหรับจำนวนเงินที่ใช้จ่าย ด้วยการวิเคราะห์การเปรียบเทียบเงินและคลัสเตอร์

การวิเคราะห์ Silhouette เป็นวิธีต้นแบบในการประเมินหรือตรวจสอบความถูกต้องของคลัสเตอร์ ซึ่งคลัสเตอร์สามารถเป็นในรูปแบบที่เป็นจัดกลุ่มรวมกันหรือกลุ่มที่แยกกระจายไป หรือทั้งสองอย่างรวมกัน ในที่นี้ ค่าสัมประสิทธิ์ของ Silhouette ผสมผสานทั้งการเกาะกลุ่มกันและการแยกจากกัน

ในส่วนนี้ อัลกอริทึมที่ใช้สำหรับผลลัพธ์ที่ต้องการของการศึกษานั้น จะสามารถนำเสนอออกเป็น 4 ขั้นตอนดังในภาพที่ 2.6



ภาพที่ 2.6 รายละเอียดประกอบการอธิบายขั้นตอน

ขั้นตอนที่ 1 การวิเคราะห์เชิงสำรวจและการประมวลผลข้อมูลล่วงหน้า

การวิเคราะห์ข้อมูลเชิงสำรวจ (EDA) หมายถึง การสำรวจข้อมูลเบื้องต้นเพื่อค้นหา รูปแบบเชิงสถิติ ในการสำรวจนี้ จะช่วยในการระบุลูกค้าที่ไม่ซ้ำกัน, เปอร์เซ็นต์ของค่าส่งซื้อตาม 10 อันดับ แรกหรือมากกว่า, ข้อมูลเกี่ยวกับข้อมูลหรือคำอธิบายที่ไม่ตรงกัน และตรวจสอบค่า Null นอกจากนี้ การประมวลผลข้อมูลล่วงหน้ายังใช้เพื่อระบุลูกค้าที่ขาดหายไปหรือข้อมูลที่ส่งผลต่อธุรกรรมเชิงลบ และอื่นๆ

ขั้นตอนที่ 2.1 การดำเนินการวิเคราะห์ RFM

หลังจากประมวลผลข้อมูลล่วงหน้าแล้ว ให้ตรวจสอบข้อมูลเกี่ยวกับความล่าสุด ความถี่ และจำนวนเงินที่ลูกค้าใช้เพื่อใช้ในการสร้างตัวแปร และกำหนดวันที่อ้างอิง ซึ่งอาจเป็น 1 วันก่อนการทำ ธุรกรรม (transaction) ครั้งล่าสุด ในการวิเคราะห์ RFM เป็นเทคนิคการแบ่งกลุ่มลูกค้าที่ได้รับความนิยม อย่างมากในตลาดฐานข้อมูล โดยลูกค้าที่อยู่ภายใต้ในการวิเคราะห์ RFM จะได้รับคะแนนตามปัจจัย 3 ประการ

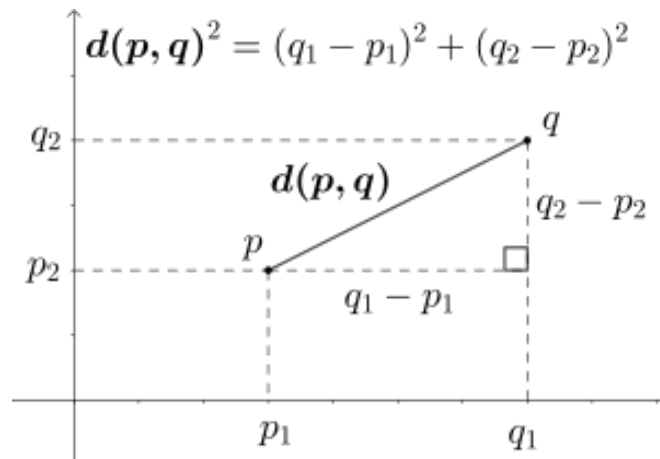
ความล่าสุด หมายถึง จำนวนวันก่อนวันที่อ้างอิงเมื่อลูกค้าทำการซื้อครั้งล่าสุด ซึ่งค่า ของความล่าสุดมีค่าน้อยลง จะทำให้การเข้าเยี่ยมชมร้านค้าก็จะยิ่งสูงขึ้น

- 1) ความถี่ หมายถึง จำนวนครั้งในช่วงเวลาระหว่างการซื้อของลูกค้าในครั้งต่อไป ซึ่ง ค่าความถี่ยิ่งสูง จะส่งผลให้ลูกค้าก็จะมาเยี่ยมชมร้านค้ามากขึ้น
- 2) การเงิน หมายถึง จำนวนเงินที่ลูกค้าใช้ไปในช่วงเวลาหนึ่ง ซึ่งมูลค่ายิ่งสูง จะ สามารถสร้างกำไรยิ่งมากขึ้นให้แก่บริษัท

ขั้นตอนที่ 2.2 ใช้อัลกอริทึม K-Means

โดยการใช้การคำนวณเมตริกระยะทางแบบยุคลิดเพื่อแบ่งลูกค้าสำหรับค่า RFM ซึ่ง K-Means จะใช้ 2 ครั้งเพื่อวิเคราะห์จำนวนเงินที่ได้รับสำหรับการทำธุรกรรมล่าสุดและบ่อยครั้งตามที่ระบุไว้ด้านล่าง

- 1) เพื่อแบ่งลูกค้าตามจำนวนที่สร้างขึ้นจากธุรกรรมล่าสุด
- 2) เพื่อจับกลุ่มลูกค้าตามยอดที่มีการทำธุรกรรมบ่อยๆ



ภาพที่ 2.7 คำนวณระยะทางแบบยุคลิดสองมิติโดยใช้ทฤษฎีบทพีทาโกรัส

ขั้นตอนที่ 3.1 การคำนวณคะแนน Silhouette คลัสเตอร์ที่ได้รับในขั้นตอนที่ 2

ขั้นตอนที่ 3.2 ได้รับการประเมินโดยใช้คะแนนจาก Silhouette

ซึ่งจะวิเคราะห์ว่าคลัสเตอร์ที่ถูกแยกออกจากกันได้ดีเพียงใด ซึ่งอยู่ในช่วงปิดที่ $[-1, +1]$ หากค่าอยู่ใกล้ $+1$ แสดงว่า กลุ่มคลัสเตอร์หรือในที่นี้คือ ลูกค้า จะถูกจัดกลุ่มให้ห่างจากคลัสเตอร์ข้างเคียง แต่ถ้ามีค่าเป็น -1 แสดงว่าอาจถูกกำหนดให้กับคลัสเตอร์ที่ไม่ถูกต้องหรือการประมวลผลข้อมูลล่วงหน้าไม่ถูกต้อง

ขั้นตอนที่ 4 การประเมินคลัสเตอร์

หากให้ K เป็นจำนวนคลัสเตอร์ จะทำให้ค่า Silhouette จะถูกเปรียบเทียบเพื่อระบุคลัสเตอร์ที่เหมาะสมที่สุดตามค่า หลังจากการวิเคราะห์ ให้เปรียบเทียบความล่าสุดกับยอดขายและความถี่ในการขายกับยอดขายจากคลัสเตอร์หนึ่งไปยังอีกคลัสเตอร์หนึ่งตามลำดับ ซึ่งจะช่วยในการระบุกลุ่มลูกค้าที่มียอดขายสูงสุดและความถี่ในการขาย

แบบจำลองทางคณิตศาสตร์โดยการจัดกลุ่มโดยใช้อัลกอริทึม K-mean ซึ่งอัลกอริทึมนี้จะระบุจุดศูนย์กลางของจำนวนของคลัสเตอร์ (K) จากชุดข้อมูล และกำหนดจุดข้อมูลที่ไม่ทับซ้อนกันให้กับแต่ละคลัสเตอร์ที่ใกล้ที่สุด ซึ่งระยะทางภายในคลัสเตอร์มีค่าสูงสุดเมื่อเทียบกับระยะทางระหว่างคลัสเตอร์ในอัลกอริทึม K-mean เนื่องจากเป็นวิธีการวนซ้ำ จะทำให้จุดข้อมูลจึงถูกย้ายไปยังคลัสเตอร์ต่างๆ ตามการคำนวณของจุดศูนย์กลาง

Input:	M: Dataset with 'n' instances K: clusters in number
Output:	Dataset partitioned into 'K' clusters
Algorithm:	<ol style="list-style-type: none"> 1. Choose arbitrarily 'k' random points from M as the cluster centers 2. repeat 3. Reassign each object to the clusters based on calculation of mean value. 4. Revise the cluster means that is recalculate the mean value of each cluster 5. Until 6. there is no change in the clusters obtained 7. Evaluation using silhouette coefficient: calculate average distance from objects in the same cluster and calculate average distance from objects to all other clusters 8. Calculate silhouette coefficient as below: $S_i = (b_i - a_i) / \max(a_i, b_i) \quad \text{for } a_i > b_i$ <p>Where, S_i represents silhouette coefficient a_i is average distance from i^{th} object to all other objects in a cluster. b_i is average distance from i^{th} object to any cluster not containing the object. Calculate the minimum such value with respect to all the clusters.</p>

ภาพที่ 2.8 อัลกอริทึมของการคำนวณด้วยวิธี K-mean

ตามแนวคิดของอัลกอริทึมที่แสดงในภาพที่ 2.8 แบบจำลองทางคณิตศาสตร์สำหรับการคำนวณ Silhouette มีดังต่อไปนี้ โดยพิจารณาจำนวนของคลัสเตอร์ (K) ซึ่งแต่ละกลุ่มมีตัวแปรเนื่องจากการใช้ค่าเฉลี่ย K สองครั้งในการทดลองปัจจุบัน ดังนั้นจึงถูกจัดกลุ่มตามข้อมูลธุรกรรมของลูกค้าสำหรับความใหม่เทียบกับจำนวนเงินและความถี่เทียบกับมูลค่าเงิน

$$K = \{ \{ (p_1, q_1), (p_2, q_2) \dots (p_x, q_x) \} \dots \{ (p_1, q_1), (p_2, q_2) \dots (p_y, q_y) \} \} \text{ ----- (2.7)}$$

เมื่อ K คือ จำนวนของคลัสเตอร์
(p, q) คือ กลุ่มของคลัสเตอร์

ในการระบุจุด สามารถระบุจุดใดก็ได้ เช่น $\{p_1, q_1\}$ ในคลัสเตอร์ที่ 1 โดยกลุ่มคลัสเตอร์จะแสดงถึงค่า RFM คำนวณระยะทางเฉลี่ยจาก $\{p_1, q_1\}$ ไปยังคลัสเตอร์ทั้งหมดในกลุ่มเดียวกัน (ค่าระยะทางภายใน ซึ่งกำหนดเป็น a_i) ซึ่งมีการคำนวณระยะทางเฉลี่ยตามที่กำหนดในสมการดังต่อไปนี้

$$\sqrt{\sum_{i=1}^n (p_1 - p_i) + (q_1 - q_i)^2} \text{ ----- (2.8)}$$

สำหรับคลัสเตอร์อื่น จะทำซ้ำในขั้นตอนเดียวกัน และค้นหาระยะทางเฉลี่ยต่ำสุดจาก $\{p_1, q_1\}$ ของคลัสเตอร์ที่ 1 ถึงคลัสเตอร์สุดท้ายของจำนวนทั้งหมด ค้นหาสัมประสิทธิ์ Silhouette กำหนดเป็น $s(i)$ สำหรับคลัสเตอร์ที่ 1 โดยใช้สมการต่อไปนี้

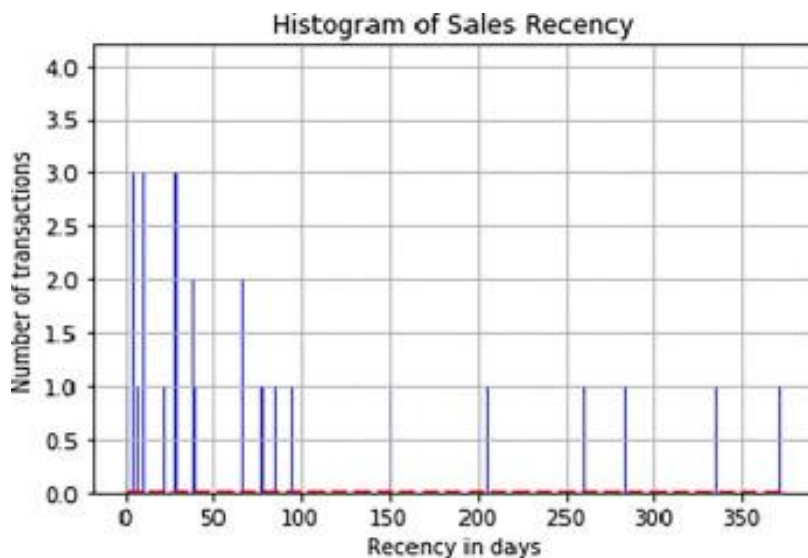
$$s(i) = \begin{cases} 1 - \frac{a_i}{b_i}, & a_i < b_i \\ 0, & a_i = b_i \\ \frac{a_i}{b_i} - 1, & a_i > b_i \end{cases} \text{ ----- (2.9)}$$

เมื่อ a_i คือระยะทางเฉลี่ยขั้นต่ำจาก $\{p_1, q_1\}$ ไปยังวัตถุอื่นๆ ในคลัสเตอร์เดียวกัน

b_i คือ ระยะทางเฉลี่ยขั้นต่ำจาก $\{p_1, q_1\}$ ไปยังคลัสเตอร์อื่นๆ ซึ่งไม่มี $\{p_1, q_1\}$

เมื่อเปรียบเทียบค่า Silhouette สำหรับคลัสเตอร์ที่ 2, 3 จนถึง n โดยทำซ้ำในขั้นตอนข้างต้น จะค้นพบว่าคลัสเตอร์ที่มีค่า Silhouette สูงสุดจะดีที่สุดตามวิธีการประเมิน ซึ่งคำนวณค่า Silhouette เฉลี่ยของคลัสเตอร์ทั้งหมดเพื่อประเมิน

วิธีการที่นำเสนอถูกนำไปใช้กับชุดข้อมูลสังเคราะห์ของธุรกรรมลูกค้าภายใน 1 ปีที่ได้รับที่เก็บข้อมูลที่ประกอบด้วยข้อมูลการซื้อของลูกค้าตั้งแต่วันที่ 1-12-2010 ถึง 09-12-2011 โดยมีค่าที่ขาดหายไป ธุรกรรมติดลบ รหัสสินค้า และคำอธิบายที่ไม่ตรงกัน จะได้รับการจัดการโดยใช้การประมวลผลข้อมูลล่วงหน้า สำหรับชุดข้อมูลที่แก้ไข ให้ใช้การวิเคราะห์ RFM และการจัดกลุ่ม K-Means วิธีการเดียวกันนี้ใช้กับชุดข้อมูลเรียลไทม์



ภาพที่ 2.9 แสดงการกระจายความใหม่ทางการขายที่สอดคล้องกับจำนวนธุรกรรมที่เกิดขึ้นบ่อย

ตารางที่ 2.2 ตัวอย่างชุดข้อมูลสังเคราะห์ของธุรกรรมลูกค้า

Index	Customer ID	Recency	Amount	Frequency
1	12413.0	67.0	758.101	38
2	12437	2.0	4951.411	200
3	12441.0	367.0	173.551	11
4	12488.0	10.0	1298.661	55
5	12489.0	336.0	334.931	15

ในที่นี้ จะยกตัวอย่างชุดข้อมูลสังเคราะห์ของธุรกรรมลูกค้าเพื่อเป็นชุดข้อมูลต้นแบบในการคำนวณได้ตามตารางที่ 2.2

ตารางที่ 2.3 ค่า Silhouette สำหรับคลัสเตอร์ 3 จำนวน

No. of Clusters	Recency_log	Frequency_log	Amount_log
0	161.191479	16.761959	291.852580
1	11.373230	209.371490	5316.800437
2	20.323096	48.877509	894.321423

ค่า Silhouette สำหรับคลัสเตอร์ 3 จำนวน คือ 0.362159752

ตารางที่ 2.4 ค่า Silhouette สำหรับคลัสเตอร์ 5 จำนวน

No. of Clusters	Recency_log	Frequency_log	Amount_log
0	210.331020	10.265222	170.590293
1	20.309899	95.243198	1913.827582
2	10.779640	280.086131	7456.390843
3	81.290783	32.515424	619.014428
4	3.847593	41.650355	708.639718

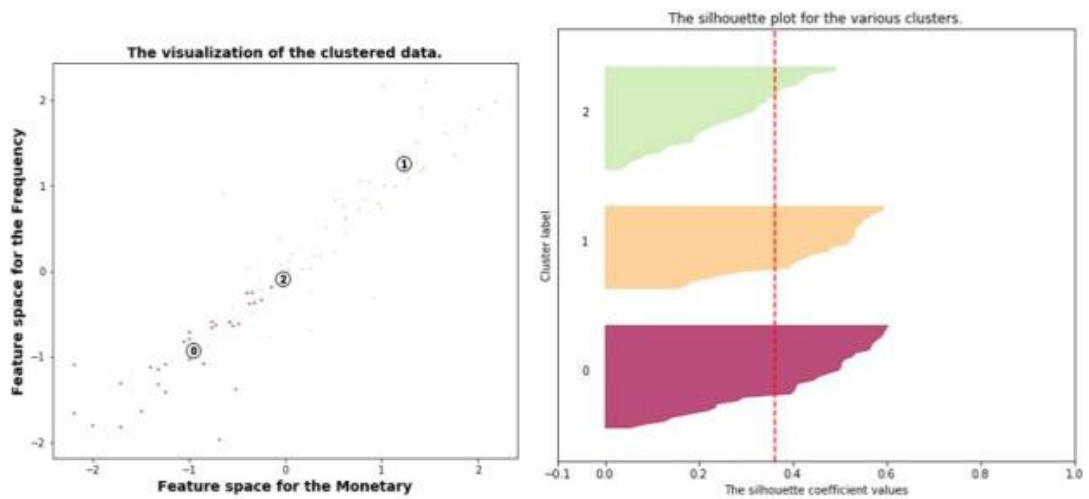
ค่า Silhouette สำหรับคลัสเตอร์ 5 จำนวน คือ 0.3490755342

จากตาราง จะบันทึกค่า RFM คำนวณสำหรับคลัสเตอร์ 3 และ 5 จำนวน โดยสังเกตได้ว่าค่า Silhouette สำหรับคลัสเตอร์ 5 จำนวน นั้นเหมาะสมน้อยกว่าเมื่อเปรียบเทียบกับคลัสเตอร์ 3 จำนวน ซึ่งค่าของ Silhouette ที่ใกล้ถึง +1 แสดงว่าเหมาะสมที่สุด เมื่อเทียบกับคลัสเตอร์อื่นๆ การกำหนดลูกค้าตามค่าในตารางที่ 2.2 ให้กับคลัสเตอร์ต่างๆ ดังในตารางที่ 2.5

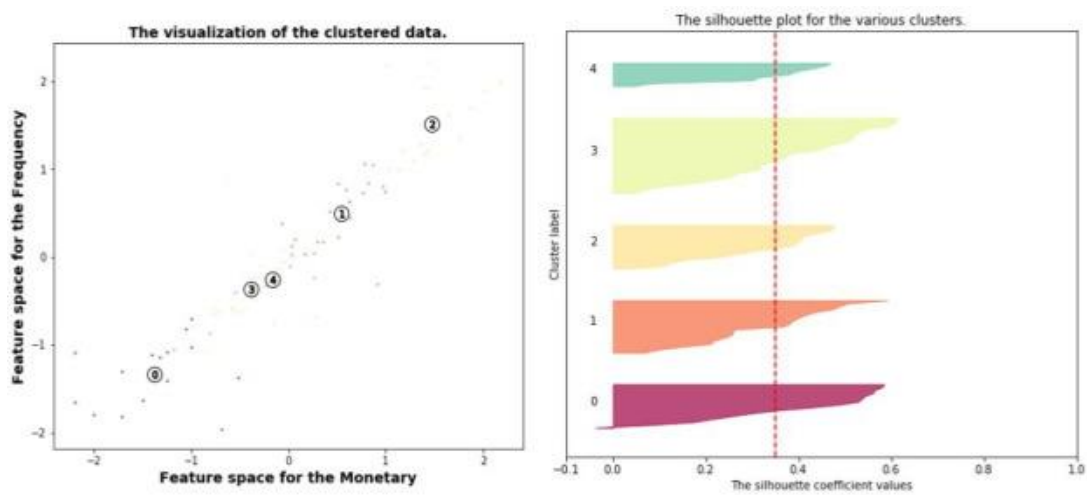
ตารางที่ 2.5 ผลวิเคราะห์ลูกค้ากับคลัสเตอร์ต่างๆ ที่นำมาเปรียบเทียบ

Customer ID	Recency_log	Frequency_log	Amount_log	Cluster5 labels	Cluster3 labels
12413.0	4.204693	3.637586	6.630817	3	2
12437.0	0.693147	5.298317	8.507428	2	1
12441.0	5.905362	2.397895	5.156472	0	0
12488.0	2.302585	4.007333	7.169089	1	2
12489.0	5.817111	2.708050	5.813925	0	0

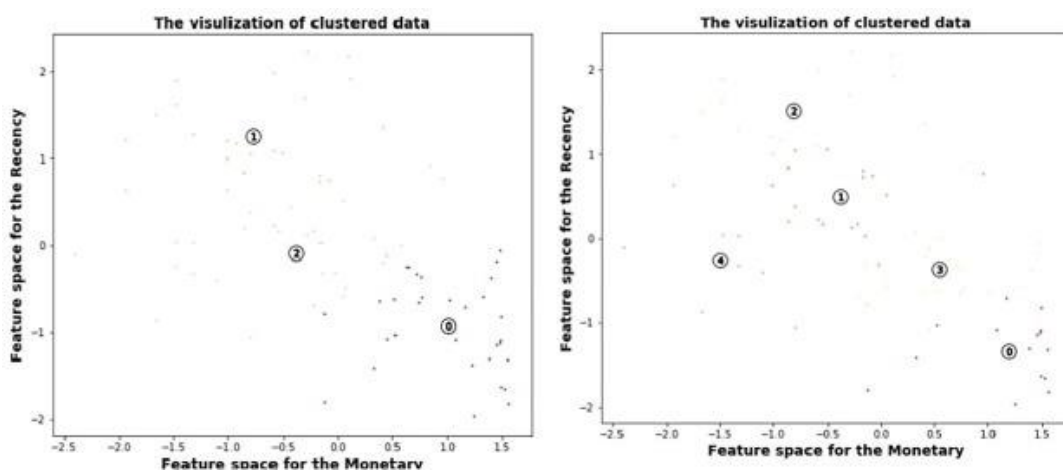
การพล็อตจุดของค่า Silhouette สำหรับคลัสเตอร์ 3 และ 5 จำนวน ที่ปรากฏในภาพที่ 2.9 และภาพที่ 2.10 ตามลำดับ การวิเคราะห์นี้ใช้กับค่า Recency และ Monetary ดังแสดงในภาพที่ 2.11 จากการวิเคราะห์ค่า Silhouette แสดงให้เห็นว่าการแบ่งกลุ่มลูกค้าสำหรับคลัสเตอร์ 3 จำนวนนั้นเหมาะสมกว่าคลัสเตอร์ที่มี 5 จำนวน ซึ่งจะนำเสนอในรูปแบบแผนภาพแบบกล่องสำหรับคลัสเตอร์ 3 และ 5 จำนวน ดังในภาพที่ 2.13, ภาพที่ 2.14 และภาพที่ 2.15



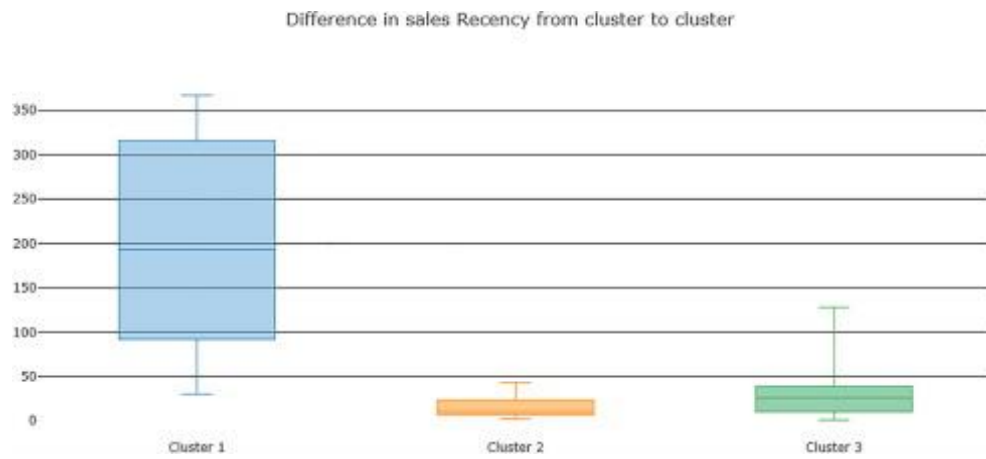
ภาพที่ 2.10 การพล็อตจุดของค่า Silhouette สำหรับคลัสเตอร์ 3 จำนวน



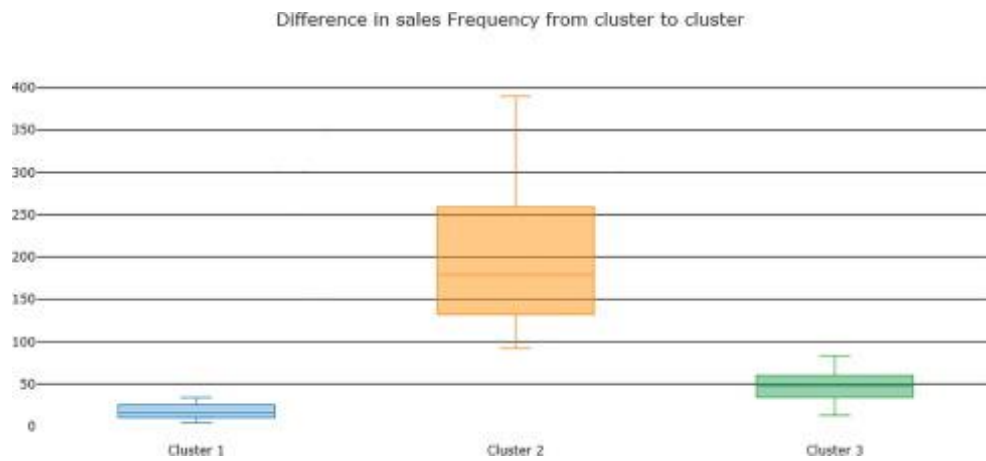
ภาพที่ 2.11 การพล็อตจุดของค่า Silhouette สำหรับคลัสเตอร์ 5 จำนวน



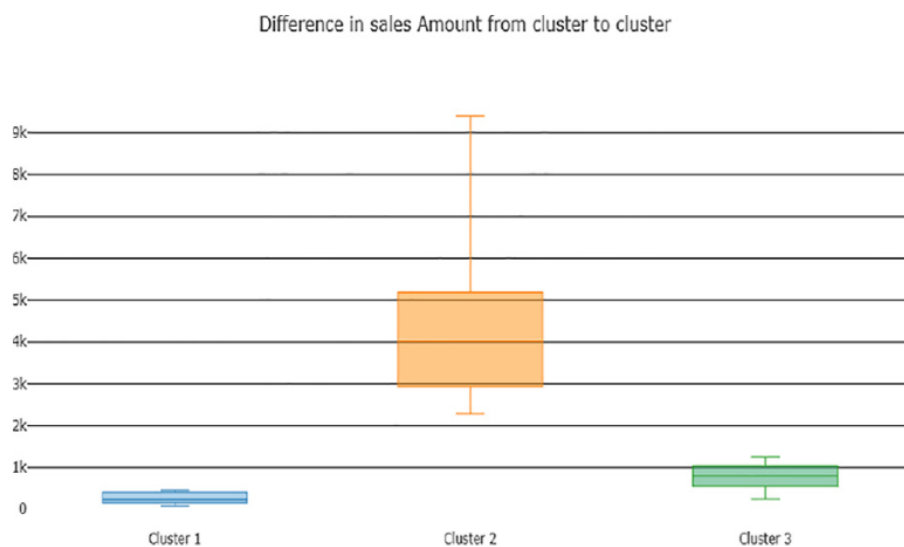
ภาพที่ 2.12 การเปรียบเทียบความล่าช้ากับจำนวนเงินสำหรับคลัสเตอร์ 3 และ 5 จำนวน



ภาพที่ 2.13 แผนภาพแบบกล่องเปรียบเทียบความล่าช้าระหว่างคลัสเตอร์



ภาพที่ 2.14 แผนภาพแบบกล่องเปรียบเทียบความถี่ระหว่างคลัสเตอร์



ภาพที่ 2.15 แผนภาพแบบกล่องเปรียบเทียบจำนวนเงินที่ใช้ระหว่างคลัสเตอร์

จากภาพที่ 2.12 คลัสเตอร์ที่ 1 มียอดขายสูงสุด และคลัสเตอร์ที่ 2 มีความถี่ในการขายสูงสุดดังในภาพที่ 2.13 และเป็นที่น่าสนใจคือ คลัสเตอร์ที่ 2 ก็มียอดขายสูงสุดเช่นกันเมื่อเทียบกับคลัสเตอร์อื่น ในการวิเคราะห์ที่คล้ายกันตาม RFM สำหรับข้อมูลตามเวลาจริง (Real Time) แสดงในตารางที่ 2.6 และตารางที่ 2.7

ตารางที่ 2.6 ค่า Silhouette สำหรับข้อมูลตามเวลาจริงสำหรับคลัสเตอร์ 3 จำนวน

No. of Clusters	Recency_log	Frequency_log	Amount_log
0	10.591432	46.571136	7395.997897
1	2.000000	115.430465	24945.010775
2	1.000000	1312.355135	612006.400895

ค่า Silhouette Score สำหรับคลัสเตอร์ 3 จำนวน คือ 0.362052482

ตารางที่ 2.7 ค่า Silhouette สำหรับข้อมูลตามเวลาจริงสำหรับคลัสเตอร์ 5 จำนวน

No. of Clusters	Recency_log	Frequency_log	Amount_log
0	6.633250	107.144762	4.006409e+04
1	1.414214	297.060600	1.413740e+05
2	10.458643	32.268573	4.026655e+03
3	1.414214	1.298395e+04	1.298395e+04
4	1.000000	5172.000000	1.228960e+06

ค่า Silhouette Score สำหรับคลัสเตอร์ 5 จำนวน คือ 0.381312431

ในการแบ่งกลุ่มลูกค้าตามรูปแบบการซื้อของลูกค้าแม้ว่าจะมีความสำคัญในเชิงกลยุทธ์ แต่ การรักษาลูกค้าเป็นสิ่งที่น่ากังวลสำหรับองค์กร ในปัจจุบัน มีการใช้โมเดล RFM สำหรับชุดข้อมูลสังเคราะห์และข้อมูลจริง เพื่อวิเคราะห์การแบ่งกลุ่มลูกค้า นอกจากนี้ คลัสเตอร์จะได้รับการประเมินโดยใช้ Silhouette Analysis สำหรับการจัดกลุ่ม K-Means ที่มีจำนวนคลัสเตอร์ต่างกัน จากค่า Silhouette สามารถวิเคราะห์ความต่ำสุด ความถี่ และจำนวนเงินในการขาย เพื่อพบแนวทางที่เหมาะสมที่สุด

ขอบเขตของงานด้านนี้ในอนาคตอยู่ที่การศึกษาและวิเคราะห์ผลิตภัณฑ์บางประเภท เช่น โทรศัพท์มือถือและอุปกรณ์เสริม สามารถศึกษาพารามิเตอร์ทางธุรกิจอื่น ๆ เช่น ผลิตภัณฑ์ที่ต้องการมากที่สุดหรือเทคนิคการขายที่มีประสิทธิภาพสูงสุดในระหว่างเหตุการณ์เฉพาะ หรือพารามิเตอร์เกณฑ์บางอย่างในภูมิภาคต่าง ๆ เพื่อการออกแบบการปรับปรุงธุรกิจที่มีประสิทธิภาพ ความก้าวหน้าและการพิจารณาอย่างรอบคอบในพื้นที่นี้จะช่วยให้องค์กรปรับปรุงธุรกิจโดยเสนอโปรโมชั่นและออกแบบกลยุทธ์ที่เป็นนวัตกรรมที่สามารถพิสูจน์ได้ว่าล้ำหน้าเหนือคู่แข่ง [10]

2.3 เครื่องมือที่เกี่ยวข้อง

2.3.1 Visual Studio Code

Visual Studio Code หรือ VS Code เป็นโปรแกรม Code Editor ที่ถูกพัฒนาขึ้นมาโดยบริษัท ไมโครซอฟต์ (Microsoft) ซึ่งสามารถแก้ไขและปรับแต่งได้หลายแพลตฟอร์ม (Platform) ทั้งระบบปฏิบัติการ Windows, macOS และ Linux และสามารถรองรับได้หลายภาษา รวมถึงสามารถติดตั้งเครื่องมือส่วนขยายต่างๆ (Extension) ได้อย่างอิสระ [11]

2.3.2 HyperText Markup Language

HyperText Markup Language หรือย่อเป็น HTML เป็นภาษาเขียนโปรแกรมที่ใช้สร้างหน้าเว็บในรูปแบบของไฟล์ที่มีนามสกุลเป็น .htm หรือ.html ซึ่งมีเว็บเบราว์เซอร์สามารถแปลงไฟล์ HTML เพื่อแสดงผลในรูปแบบของหน้าเว็บ

ไฟล์ HTML ประกอบด้วยแท็ก (Tag) ต่างๆ ที่เป็นคำสั่งของ HTML โดยแท็กจะอยู่ภายในเครื่องหมาย < และ > ซึ่งสามารถแบ่งเป็น 2 ประเภท คือ Container Tag และ Empty Tag โดยที่ Container Tag ประกอบไปด้วยแท็กเปิดและแท็กปิด โดยที่แท็กปิดจะมีเครื่องหมาย/ นำหน้าแท็ก เช่น <H1> . . . </H1> และส่วนแท็กเปล่าจะมีแท็กเปิดอย่างเดียว เช่น <hr> ซึ่งแท็กจะถูกเขียนด้วยตัวอักษรพิมพ์ใหญ่หรือพิมพ์เล็กก็ได้จะไม่มีผลต่อการแสดงผลของเว็บเบราว์เซอร์ เช่น <HR>, <hr>, <Hr> หรือ <hR> เว็บเบราว์เซอร์จะแสดงผลเหมือนกัน [12]

2.3.3 JavaScript

JavaScript หรือย่อเป็น JS เป็นภาษาเขียนโปรแกรมที่ถูกพัฒนา และปฏิบัติตามข้อกำหนดมาตรฐานของ ECMAScript ซึ่งภาษา JavaScript นั้นเป็นภาษาระดับสูง คอมไพล์ในขณะที่โปรแกรมรัน (Just In Time หรือ JIT) และเป็นภาษาเขียนโปรแกรมแบบหลาย Paradigm เช่น การเขียนโปรแกรมเชิงโครงสร้าง (Structured programming), การเขียนโปรแกรมเชิงวัตถุ (Object Oriented Programming หรือ OOP) หรือการเขียนโปรแกรมแบบฟังก์ชัน (Functional) นอกจากนี้ JavaScript ยังเป็นภาษาที่มีประเภทข้อมูลแบบไดนามิก (Dynamic) เป็นภาษาแบบ Prototype-based และ First-class function

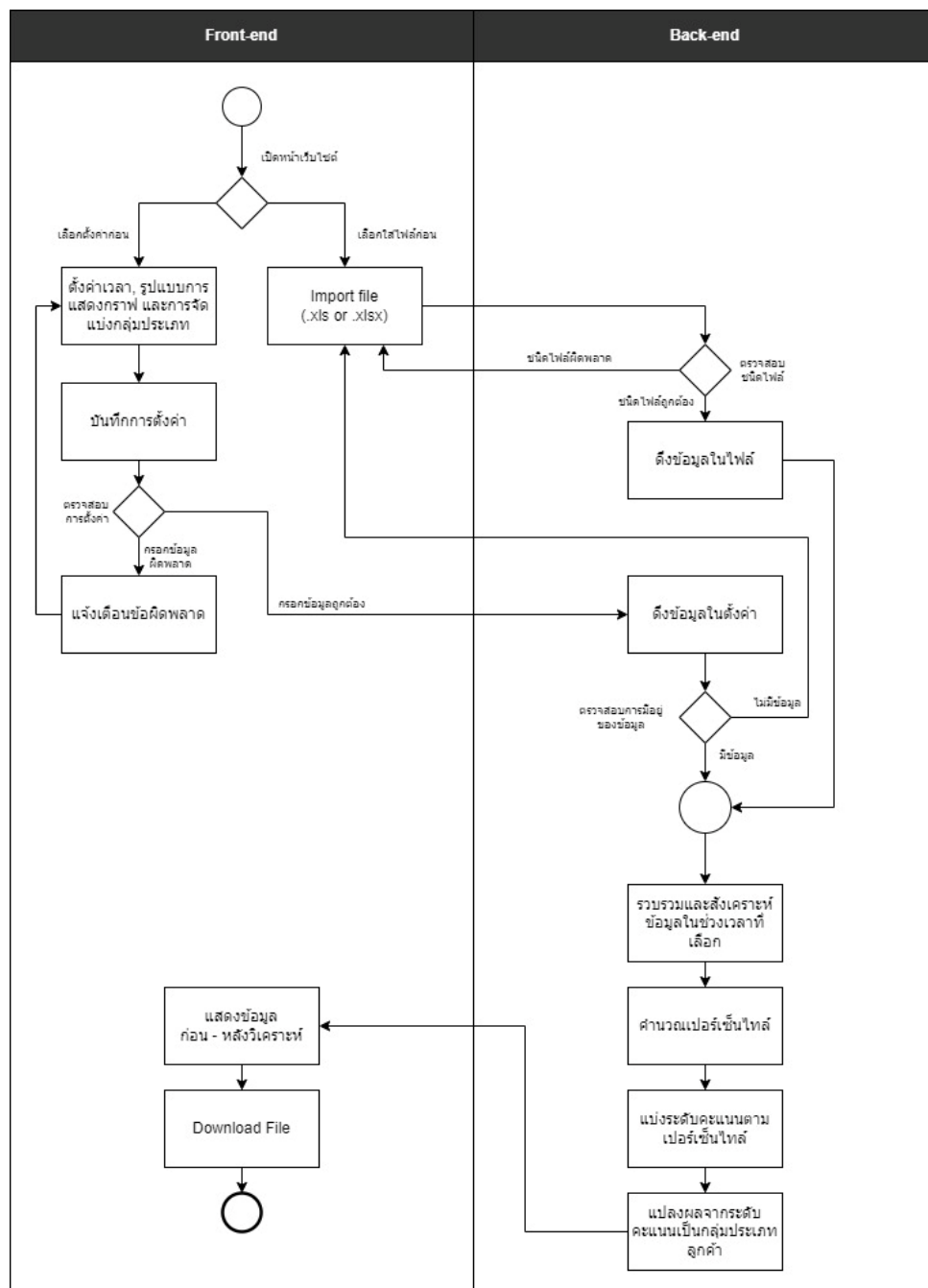
ภาษา JavaScript นั้นถือว่าเป็นมีประโยชน์ต่อพัฒนาเว็บไซต์ (World Wide Web) มันทำให้หน้าเว็บสามารถตอบโต้กับผู้ใช้ได้โดยไม่ต้องรีเฟรชหน้าใหม่ (Dynamic website) ซึ่งเว็บไซต์จำนวนมากใช้ภาษา JavaScript สำหรับควบคุมการทำงานที่ฝั่งผู้ใช้ (Client side) นั่นทำให้เว็บเบราว์เซอร์ต่างๆ มี JavaScript engine ที่ใช้สำหรับประมวลผลสคริปของภาษา JavaScript ที่รันบนเว็บเบราว์เซอร์ [13]

บทที่ 3

วิธีดำเนินการวิจัย

3.1 การออกแบบ Business Process Model and Notation

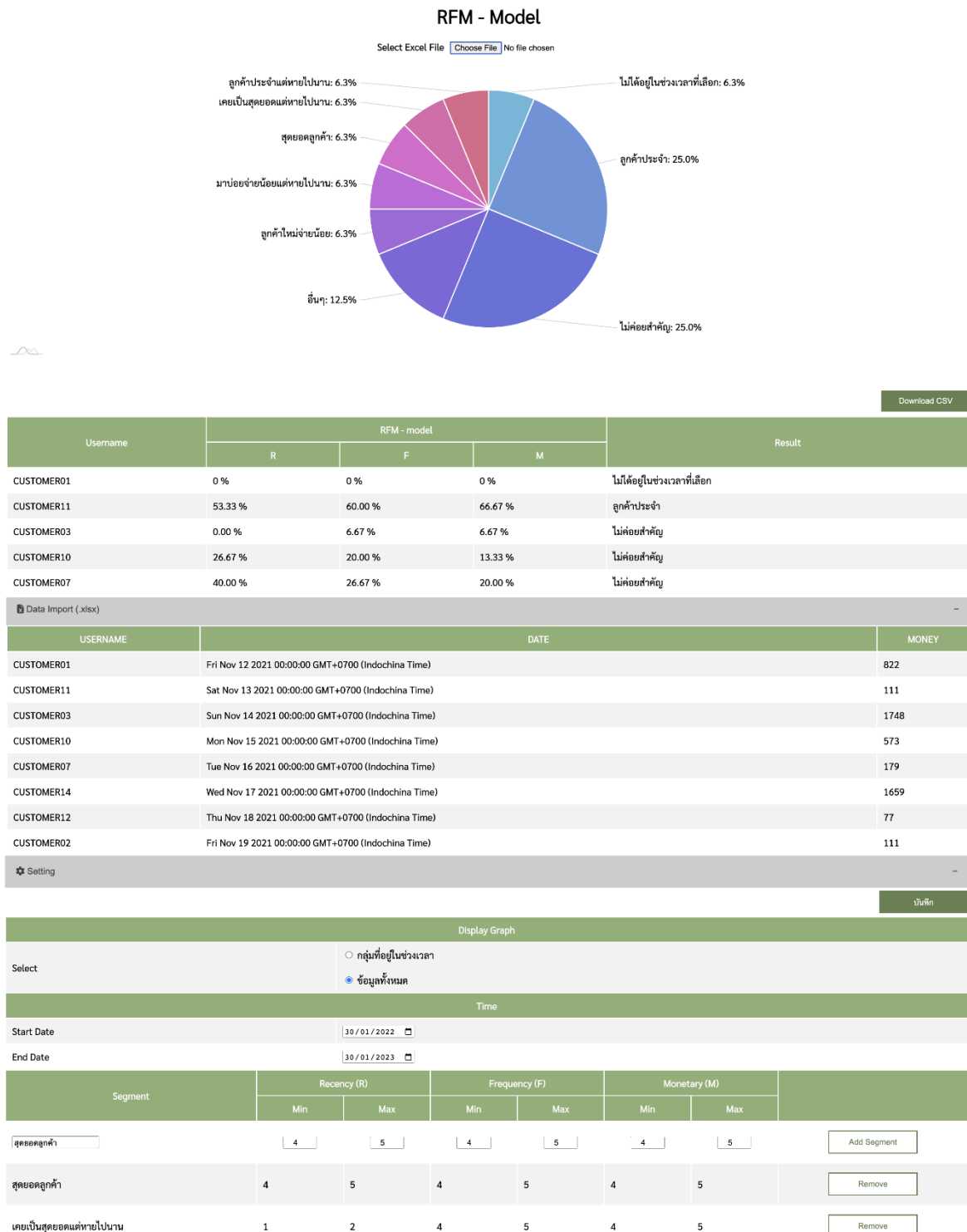
Business Process Model and Notation หรือย่อเป็น BPMN เป็นแผนผังที่ช่วยอธิบายกระบวนการทำงานของระบบต่างๆ ที่ทำขึ้น ซึ่งในที่นี้ จะเป็นการเขียนอธิบายระบบ RFM Model เพื่อให้ทราบว่าระบบที่ต้องการออกแบบนั้น ต้องมีขั้นตอนใดบ้างดังภาพที่ 3.1



ภาพที่ 3.1 BPMN ของระบบ RFM model

ในภาพที่ 3.1 อธิบายถึงระบบ RFM model เมื่อเปิดหน้าเว็บไซต์ จะพบเจอส่วนนำเข้าไฟล์ข้อมูลและส่วนที่ให้ผู้ใช้งานสามารถตั้งค่าได้ ก่อนที่จะทำการประมวลผลตามลำดับที่อ้างอิงจากข้อมูลที่ได้สืบค้นมา และหลังจากการวิเคราะห์จะมีการแสดงผลลัพธ์ดังกล่าว พร้อมทั้งสามารถส่งออกข้อมูลได้

3.2 การออกแบบ User Experience และ User Interface



ภาพที่ 3.2 User Experience และ User Interface ของระบบ RFM model

ในการออกแบบเว็บไซต์นั้น จะมีการออกแบบหน้าต่างแสดงผลดังภาพที่ 3.2 ซึ่งประกอบด้วย ข้อมูลก่อนการวิเคราะห์ผล, การแสดงผลหลังวิเคราะห์ด้วยกราฟและตาราง, ปุ่ม (Button) ที่สามารถนำเข้าและส่งออกข้อมูลได้ และการตั้งค่าการวิเคราะห์

3.3 สร้างต้นแบบการจำลองการคำนวณข้อมูลเพื่อวิเคราะห์สูตรโดยอิงจาก RFM Model

จากการศึกษาข้อมูลเกี่ยวกับ RFM Model แล้ว ในขั้นตอนวิเคราะห์ข้อมูลนั้น สามารถสร้างได้ด้วย 2 วิธี คือ การทำ Machine Learning โดยใช้วิธีคำนวณของ K-Means และการทำ Traditional RFM scoring ซึ่งในการสร้างต้นแบบการจำลองครั้งนี้ จะใช้วิธีการทำ Traditional RFM scoring เพื่อตอบสนองต่อขอบเขตที่ได้จัดตั้งไว้ อย่างเช่น การแบ่งกลุ่มประเภทลูกค้าตามระดับคะแนนได้อย่างอิสระ เนื่องจากวิธีคำนวณของ K-Means นั้น เมื่อผ่านการวิเคราะห์ออกมาแล้วจะแสดงออกมาเป็นกลุ่มคลัสเตอร์เพียงเท่านั้น ซึ่งไม่สามารถปรับเปลี่ยนกลุ่มประเภทลูกค้าได้ตามระดับคะแนน แต่ในการใช้วิธีการทำ Traditional RFM scoring นั้น จะค่อนข้างมีอิสระในการปรับเปลี่ยนได้มากกว่า

การทำต้นแบบจำลองเพื่อศึกษานั้น จะใช้โปรแกรม Excel ที่นิยมใช้ในการคำนวณ ซึ่งมีการจัดเตรียมข้อมูลตัวอย่างเพื่อสะดวกต่อการคำนวณ และสร้างเห็นความแตกต่างได้อย่างชัดเจน โดยข้อมูลที่นำมาใช้นั้นหรือเรียกว่า “transaction” จะประกอบด้วยชื่อของลูกค้า (username), เวลาที่ซื้อ (time), จำนวนเงินที่ใช้จ่าย (amount of money) [14]

ตารางที่ 3.1 ตัวอย่างข้อมูลที่นำมาวิเคราะห์ RFM model

username	time	amount of money
Customer01	12/11/2021	5980
Customer02	12/11/2021	706
Customer01	31/12/2021	1966
Customer01	31/12/2021	2039
Customer03	1/02/2022	3374
Customer04	6/02/2022	7920
Customer03	7/02/2022	992

จากข้างต้น ข้อมูลที่นำมาอ้างอิงสามารถเป็นข้อมูลที่มีรายละเอียดซ้ำกันได้ เพราะว่าในช่วงของการวิเคราะห์นำข้อมูลที่มีข้อมูลมารวบรวมและวิเคราะห์เพื่อให้มีข้อมูลที่จะนำไปวิเคราะห์เพียงข้อมูลสำหรับบุคคลคนนั้นๆ

ในการวิเคราะห์สามารถจัดทำได้หลายมุมมอง เช่น มุมมองในการวิเคราะห์เชิงบุคคล และมุมมองในการวิเคราะห์เชิงรวม โดยในแต่ละวิธีจะถูกนำมาวิเคราะห์ได้ขึ้นอยู่กับผู้วิเคราะห์ข้อมูลและเป้าหมายในการวิเคราะห์ ซึ่งในครั้งนี้ จะใช้วิธีการวิเคราะห์เชิงรวม เนื่องจากการวิเคราะห์เชิงบุคคลมีจุดด้อยอยู่ที่เมื่อข้อมูล transaction ของลูกค้านั้นๆ มีการเคลื่อนไหวเพียงครั้งเดียวเท่านั้นตามสูตรคำนวณที่ผู้วิเคราะห์ได้ค้นพบดังต่อไปนี้

$$\frac{(\text{ระยะเวลาทั้งหมด}) - \text{ระยะเวลาที่ห่าง}}{\text{ระยะเวลาทั้งหมด}} \times 100 \quad \text{-----} \quad (3.1)$$

เมื่อ ระยะเวลาที่ห่าง คือ วันสุดท้ายที่มี transaction เคลื่อนไหวจนถึงวันที่สนใจ
ระยะเวลาทั้งหมด คือ วันแรกที่มี transaction เคลื่อนไหวจนถึงวันที่สนใจ

ในเมื่อข้อมูล transaction ของลูกค้าคนๆ นั้น มีการเคลื่อนไหวเพียงครั้งเดียว จะส่งผลให้ระยะเวลาที่ห่างและระยะเวลาทั้งหมดมีค่าเท่ากัน ทำให้ค่าเปอร์เซ็นต์ที่ได้รับเท่ากับ 0% ซึ่งไม่ควรเกิดขึ้นอย่างมากในกรณีที่ลูกค้าทำการซื้อครั้งแรกในวันที่ผู้วิเคราะห์สนใจ

ส่วนในด้านการวิเคราะห์ของความถี่และปริมาณเงินที่ใช้จ่ายของลูกค้านั้น ในการวิเคราะห์เชิงบุคคล จะใช้รูปแบบเดียวกัน คือ การเปรียบเทียบอัตราส่วนโดยอิงจากค่าที่มากที่สุด โดยค่าต่างๆ นั้น จะต้องจัดกลุ่มเป็นแต่ละเดือนก่อนทำการเปรียบเทียบ จึงทำการจัดระดับคะแนนอีกครั้งจากการเฉลี่ยของค่าอัตราส่วนของบุคคลนั้นๆ

ส่วนการวิเคราะห์เชิงรวมจะทำการคำนวณโดยใช้วิธีเปอร์เซ็นต์ไทล์ (Percentile) โดยจัดเป็น 3 ประเภทตามกลุ่ม RFM (ความล่าสุด, ความถี่, ปริมาณของเงินที่ใช้จ่าย) โดยในการจัดต้นแบบนั้น จะคำนวณผ่านโปรแกรม Excel ก่อน ดังตัวอย่างตามตารางที่ 3.2

ในการพิจารณาข้อมูลที่ผ่านการจัดกลุ่มนั้น ข้อมูลดังตารางที่ 3.1 ที่เป็นข้อมูลที่ได้รับมา จะถูกใช้สูตรเพื่อจัดกลุ่มเพื่อต่อการวิเคราะห์และการแปลงค่าข้อมูลต่างๆ เป็นค่าเปอร์เซ็นต์ไทล์ โดยสูตรที่ใช้ในโปรแกรม Excel คือ

1) ความล่าสุด (Recency) ใช้เป็นค้นหาช่วงเวลาสุดท้ายของ transaction ของลูกค้านั้นๆ จึงเป็นสูตรหาค่าสุดท้ายตามเงื่อนไขที่กำหนดของข้อมูล คือ LOOKUP

ตัวอย่าง

$$\begin{aligned} &= \text{LOOKUP} (2, 1 / (\text{เงื่อนไข}), \text{ช่วงข้อมูลที่ต้องการแสดง}) [15] \\ &= \text{LOOKUP} (2, 1 / (A2 : A300 = E2), B2 : B300) \end{aligned}$$

2) ความถี่ (Frequency) ใช้สูตร COUNTIF ที่เป็นสูตรนับจำนวนตามเงื่อนไขที่กำหนด

ตัวอย่าง

$$\begin{aligned} &= \text{COUNTIF} (\text{ช่วงข้อมูลทั้งหมดในการนับจำนวน, ค่าที่เป็นเงื่อนไขในการนับจำนวน}) \\ &= \text{COUNTIF} (A2 : A300, E2) \end{aligned}$$

3) ปริมาณของเงินที่ใช้จ่าย (Monetary) ใช้สูตร SUMIF ที่เป็นการรวมค่าตามเงื่อนไข และสาเหตุที่เลือกใช้การรวมจำนวนเงิน เนื่องจากในมุมมองของบริษัทแล้ว จะให้ความสนใจแก่จำนวนเงินที่สะสมมากกว่าจำนวนเงินที่เฉลี่ย ซึ่งจำนวนเงินที่เฉลี่ยนั้น มักจะนำไปใช้งานในด้านการพยากรณ์มากกว่า

ตัวอย่าง

$$\begin{aligned} &= \text{SUMIF} (\text{ช่วงข้อมูลทั้งหมดที่นำมาเทียบกับเงื่อนไข, ค่าที่เป็นเงื่อนไขในการนับจำนวน, ช่วงข้อมูลทั้งหมดที่ต้องการรวมค่า}) \\ &= \text{SUMIF} (A2 : A300, E2, C2 : C300) \end{aligned}$$

4) เปอร์เซ็นไทล์ (Percentile) เป็นการแบ่งข้อมูลออกเป็นช่วงๆ ซึ่งสามารถใช้สูตร

ตัวอย่าง

= PERCENTRANK.INC (ช่วงข้อมูลทั้งหมด, ข้อมูลที่ต้องการรู้ตำแหน่งของ เปอร์เซ็นไทล์)

= PERCENTRANK.INC (G2 : G17, G2)

หรือ

= (RANK.EQ (ช่วงข้อมูลที่ต้องการทราบลำดับ, ในที่นี้ใช้ค่า 1 เนื่องจากต้องการ ลำดับที่เรียงจากน้อยไปมาก) - 1) / (COUNT (ช่วงข้อมูลทั้งหมดในการนับจำนวน) - 1)

= (RANK.EQ (G2, G2 : G\$17, 1) - 1)/(COUNT(G2 : G17) - 1)

จากแนวคิดของทั้งสองสูตรนี้ สามารถใช้แทนกันได้ แต่จะเกิดค่าข้อผิดพลาดหรือค่า Error เพียงเล็กน้อยที่อยู่ในระดับที่ยอมรับได้ตามตารางที่จะพิสูจน์ ณ ต่อไปนี้

ตารางที่ 3.2 ตารางเปรียบเทียบระหว่าง PERCENTRANK.INC และ RANK.EQ/COUNT

PERCENTRANK.INC	RANK.EQ/COUNT	Error
0.00%	0.00%	0.00%
86.60%	86.67%	0.07%
6.60%	6.67%	0.07%
40.00%	40.00%	0.00%
26.60%	26.67%	0.07%
93.30%	93.33%	0.03%
46.60%	46.67%	0.07%
66.60%	66.67%	0.07%
100.00%	100.00%	0.00%
33.30%	33.33%	0.03%
60.00%	60.00%	0.00%
73.30%	73.33%	0.03%
20.00%	20.00%	0.00%
13.30%	13.33%	0.03%
53.30%	53.33%	0.03%
80.00%	80.00%	0.00%

หากสังเกตดู จะพบว่า สูตร RANK.EQ/COUNT ที่นำมาแทนนั้น ถูกลบค่าด้วย 1 เนื่องจาก RANK.EQ นั้น จะเป็นระบุตำแหน่งของค่าตามการจัดเรียงลำดับไว้ หากเป็นค่าที่ซ้ำกันนั้น ตำแหน่งของค่าที่ซ้ำเป็นค่าแรกสุดจะถูกแสดงออกมา แต่ในสูตร PERCENTRANK.INC จะเป็นการกล่าวถึงเปอร์เซ็นไทล์ที่เริ่มตั้งแต่ 0 จนถึง 100 จึงจำเป็นต้องทำให้ค่าตำแหน่งแรกสุดเริ่มต้นด้วย 0 ไม่ใช่ 1 และส่วนกรณีที่ถูกนำมาเขียนโปรแกรมนี้ ค่าดัชนีหรือ index จะเริ่มต้นด้วย 0 เป็นส่วนใหญ่สำหรับภาษาโปรแกรม ซึ่งภาษา JavaScript ก็เป็นหนึ่งในภาษาเหล่านั้น และสูตร COUNT ที่ถูกลบด้วย 1 ก็เช่นกัน เป็นการป้องกัน

ไม่ให้เกิดจำนวนที่ผิดพลาด เช่น การกำหนดข้อมูลแรกด้วยตำแหน่งที่ 0 และมีจำนวนข้อมูลรวมเป็น 16 จำนวน แต่การคำนวณที่เป็นค่าตำแหน่งสุดท้ายในกรณีสูตร COUNT ไม่มีการลบ คือ 15/16 ซึ่งเป็นกรณีที่ไม่สมควรเกิดขึ้น เพราะว่า เปอร์เซ็นไทล์ที่ 100 นั้น ค่าของตำแหน่งและจำนวนของควรมีค่าเท่ากัน เว้นหากค่ามากที่สุดที่มีค่าเท่ากันหลายค่า

ตารางที่ 3.3 ตัวอย่างข้อมูลหลังจากที่นำมาวิเคราะห์

USERNAME	ข้อมูลที่ผ่านการจัดกลุ่ม			เปอร์เซ็นต์ไทล์		
	ความล่าช้า	ความถี่	ปริมาณเงิน	ความล่าช้า	ความถี่	ปริมาณเงิน
CUSTOMER01	12/11/2021	1	822	0.00%	0.00%	0.00%
CUSTOMER02	3/09/2022	14	16908	86.67%	20.00%	20.00%
CUSTOMER03	3/06/2022	13	23171	6.67%	13.33%	53.33%
CUSTOMER04	14/08/2022	18	33107	40.00%	46.67%	86.67%
CUSTOMER05	6/08/2022	20	14883	26.67%	60.00%	13.33%
CUSTOMER06	5/09/2022	23	23216	93.33%	73.33%	60.00%
CUSTOMER07	26/08/2022	17	17826	46.67%	33.33%	26.67%
CUSTOMER08	31/08/2022	24	27703	66.67%	80.00%	66.67%
CUSTOMER09	6/09/2022	17	38629	100.00%	33.33%	93.33%
CUSTOMER10	12/08/2022	16	22536	33.33%	26.67%	46.67%
CUSTOMER11	28/08/2022	24	29996	60.00%	80.00%	73.33%
CUSTOMER12	1/09/2022	21	22045	73.33%	66.67%	33.33%
CUSTOMER13	4/08/2022	27	32915	20.00%	93.33%	80.00%
CUSTOMER14	2/08/2022	12	12801	13.33%	6.67%	6.67%
CUSTOMER15	27/08/2022	18	22243	53.33%	46.67%	40.00%
CUSTOMER16	2/09/2022	34	60454	80.00%	100.00%	100.00%

ข้อมูลที่ผ่านการจัดกลุ่มมานั้น หากอยู่ในช่วงเวลาที่ถูกวิเคราะห์สนใจแล้ว จะสังเคราะห์ข้อมูลโดยใช้ชื่อลูกค้า (username) เป็นตัวหลักในการตัดสินใจ โดยความล่าช้าจะพิจารณาโดยเวลาซื้อสินค้าหรือบริการในครั้งสุดท้ายสำหรับลูกค้าคนๆ นั้น, ความถี่จะพิจารณาโดยการนับจำนวนของ transaction ของลูกค้าคนนั้น และจำนวนเงินจะพิจารณาจากเงินรวมของ transaction ของลูกค้าคนนั้น แล้วจึงนำมาวิเคราะห์ด้วยวิธีเปอร์เซ็นต์ไทล์ (Percentile)

โดยผลการวิเคราะห์ด้วยวิธีเปอร์เซ็นต์ไทล์ (Percentile) ที่ปรากฏข้างต้นนั้น ประมวลผลผ่านโปรแกรม Excel ด้วยสูตร PERCENTRANK.INC และเหตุผลที่เลือกสูตรนี้ เนื่องจากป้องกันข้อผิดพลาดจากสูตร PERCENTRANK.EXC ที่มีความใกล้เคียงกับนิยามมากที่สุด แต่มีปัญหาในการแทนค่าตำแหน่งที่ 0 และตำแหน่งที่ 100 (ค่าที่แทนในสูตรของโปรแกรมมีค่าเป็น 1)

เมื่อทราบเปอร์เซ็นต์ไทล์ของแต่ละค่า จะสามารถแบ่งเป็น 5 สัดส่วนได้ โดยกำหนดให้ 0% - 20% เป็นระดับ 1, 21% - 40% เป็นระดับ 2, 41% - 60% เป็นระดับ 3, 61% - 80% เป็นระดับ 4 และ 81% - 100% เป็นระดับ 5 จึงนำมาจัดเป็นกลุ่มลูกค้าตามระดับที่กำหนดดังตารางที่ 3.4

ตารางที่ 3.4 กลุ่มลูกค้าเมื่อเทียบกับค่าระดับของ RFM

SEGMENT	R	F	M
สุดยอดลูกค้า	4,5	4,5	4,5
เคยเป็นสุดยอดแต่หายไปนาน	1,2	4,5	4,5
ลูกค้าใหม่จ่ายเยอะ	4,5	1	4,5
ลูกค้าใหม่จ่ายน้อย	4,5	1	1,2,3
นานมาที่จ่ายเยอะ	3,4,5	1,2	4,5
นานมาที่จ่ายเยอะแต่หายไปนาน	1,2	1,2	4,5
มาบ่อยจ่ายน้อย	3,4,5	4,5	1,2
มาบ่อยจ่ายน้อยแต่หายไปนาน	1,2	4,5	1,2
ลูกค้าประจำ	3,4,5	3,4,5	1,2,3,4,5
ลูกค้าประจำแต่หายไปนาน	1,2	3,4,5	1,2,3,4,5
ไม่ค่อยสำคัญ	1,2	1,2	1,2
อื่นๆ	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5

ในการจัดระดับเป็น 5 ช่วงนั้น สามารถใช้สูตรในโปรแกรม Excel ในการคำนวณได้ คือ นำค่าเปอร์เซ็นต์มาคำนวณในสูตร $IF (CEILING.MATH (PERCENTRANK/0.2, 1) = 0, 1, CEILING.MATH (PERCENTRANK/0.2, 1))$ ซึ่งสูตร CEILING.MATH เป็นสูตรเกี่ยวกับการปัดเลข ในที่นี้เป็นการปัดเลขที่เข้าใกล้จำนวนเต็ม เช่น 3.12 จะกลายเป็น 4 และสาเหตุที่ใช้ IF นั้น ในบางครั้งจะมีค่าเปอร์เซ็นต์ที่เป็น 0 แม้ถูกแบ่งเป็น 5 ส่วนแล้ว ตามทฤษฎีบทคณิตศาสตร์ที่พิสูจน์มาแล้วนั้น เมื่อค่า 0 ถูกหารด้วยจำนวนใดๆ จะมีค่าเท่ากับ 0 ซึ่งตามสมมติฐานที่กำหนดไว้นั้น ค่า 0 จะถูกจัดเป็นระดับ 1 ซึ่งนั่นทำให้ต้องใช้ IF ในการตรวจสอบเงื่อนไข

สุดท้ายนี้ การแปลงข้อมูลจากระดับคะแนนเป็นกลุ่มประเภทลูกค้า ในโปรแกรม Excel สามารถสร้างเงื่อนไขเพื่อจัดแบ่งได้ โดยใช้สูตร IF (เงื่อนไข, แสดงผลเมื่อเป็นจริง, แสดงผลเมื่อเป็นเท็จ) และเงื่อนไขนั้น ต้องมีการเรียงลำดับของช่วงคะแนนที่ถูกต้อง

ตารางที่ 3.5 ลูกค้าที่ถูกจัดประเภทตามลำดับ

USERNAME	เปอร์เซ็นต์			ระดับคะแนน			แปลงข้อมูล
	ความล่าสุด	ความถี่	ปริมาณเงิน	R	F	M	
CUSTOMER01	0.00%	0.00%	0.00%	1	1	1	ไม่ค่อยสำคัญ
CUSTOMER02	86.67%	20.00%	20.00%	5	1	1	ลูกค้าใหม่จ่ายน้อย
CUSTOMER03	6.67%	13.33%	53.33%	1	1	3	อื่นๆ
CUSTOMER04	40.00%	46.67%	86.67%	2	3	5	ลูกค้าประจำแต่หายไปนาน
CUSTOMER05	26.67%	60.00%	13.33%	2	3	1	ลูกค้าประจำแต่หายไปนาน
CUSTOMER06	93.33%	73.33%	60.00%	5	4	3	ลูกค้าประจำ
CUSTOMER07	46.67%	33.33%	26.67%	3	2	2	อื่นๆ
CUSTOMER08	66.67%	80.00%	66.67%	4	4	4	สุดยอดลูกค้า

USERNAME	เปอร์เซ็นต์			ระดับคะแนน			แปลงข้อมูล
	ความล่าช้า	ความถี่	ปริมาณเงิน	R	F	M	
CUSTOMER09	100.00%	33.33%	93.33%	5	2	5	นานมาที่จ่ายเยอะ
CUSTOMER10	33.33%	26.67%	46.67%	2	2	3	อื่นๆ
CUSTOMER11	60.00%	80.00%	73.33%	3	4	4	ลูกค้าประจำ
CUSTOMER12	73.33%	66.67%	33.33%	4	4	2	มาบ่อยจ่ายน้อย
CUSTOMER13	20.00%	93.33%	80.00%	1	5	4	เคยเป็นสุดยอดแต่หายไปนาน
CUSTOMER14	13.33%	6.67%	6.67%	1	1	1	ไม่ค่อยสำคัญ
CUSTOMER15	53.33%	46.67%	40.00%	3	3	2	ลูกค้าประจำ
CUSTOMER16	80.00%	100.00%	100.00%	4	5	5	สุดยอดลูกค้า

บทที่ 4

ผลการวิจัย

4.1 ต้นแบบจำลอง RFM model

จากศึกษาเรื่อง RFM model และการใช้งานและสูตรของโปรแกรมคำนวณอย่างโปรแกรม Excel นั้น ทำให้สามารถสร้างสูตรการคำนวณเพื่อทำต้นแบบจำลองออกมา โดยขั้นตอนแรก คือ การจัดข้อมูลให้อยู่ในรูปแบบที่หัวข้อของคอลัมน์หรือแอตทริบิวต์ (Attribute) ให้เรียงตามลำดับ เช่น ชื่อลูกค้า วันที่ที่ใช้เงิน และปริมาณจำนวนเงิน

USERNAME	DATE	MONEY
CUSTOMER01	12/11/2021	822
CUSTOMER11	13/11/2021	111
CUSTOMER03	14/11/2021	1748
CUSTOMER10	15/11/2021	573
CUSTOMER07	16/11/2021	179
CUSTOMER14	17/11/2021	1659
CUSTOMER12	18/11/2021	77
CUSTOMER02	19/11/2021	111
CUSTOMER14	20/11/2021	1594
CUSTOMER06	21/11/2021	329
CUSTOMER12	22/11/2021	1594
CUSTOMER12	23/11/2021	180
CUSTOMER15	24/11/2021	111
CUSTOMER15	25/11/2021	1001
CUSTOMER06	26/11/2021	44
CUSTOMER14	27/11/2021	2646
CUSTOMER05	28/11/2021	825
CUSTOMER15	29/11/2021	2150
CUSTOMER12	30/11/2021	140
CUSTOMER10	1/12/2021	1865
CUSTOMER02	2/12/2021	1170
CUSTOMER16	3/12/2021	248
CUSTOMER12	4/12/2021	398

ภาพที่ 4.1 ตัวอย่างข้อมูลจริงของลูกค้า

หลังจากนั้น ทำการสังเคราะห์และจัดกลุ่มข้อมูลของลูกค้าทั้งหมด ซึ่งวันที่ที่ใช้เงินจะใช้เวลาล่าสุดของข้อมูล และปริมาณจำนวนเงินจะเป็นผลรวมของตัวเลขทั้งหมดของลูกค้าคนนั้นๆ

	R	F	M
CUSTOMER01	12 พฤศจิกายน 2021	1	822
CUSTOMER02	3 กันยายน 2022	14	16908
CUSTOMER03	3 มิถุนายน 2022	13	23171
CUSTOMER04	14 สิงหาคม 2022	18	33107
CUSTOMER05	6 สิงหาคม 2022	20	14883
CUSTOMER06	5 กันยายน 2022	23	23216
CUSTOMER07	26 สิงหาคม 2022	17	17826
CUSTOMER08	31 สิงหาคม 2022	24	27703
CUSTOMER09	6 กันยายน 2022	17	38629
CUSTOMER10	12 สิงหาคม 2022	16	22536
CUSTOMER11	28 สิงหาคม 2022	24	29996
CUSTOMER12	1 กันยายน 2022	21	22045
CUSTOMER13	5 สิงหาคม 2022	27	32915
CUSTOMER14	2 สิงหาคม 2022	12	12801
CUSTOMER15	27 สิงหาคม 2022	18	22243
CUSTOMER16	2 กันยายน 2022	34	60454

ภาพที่ 4.2 ตัวอย่างข้อมูลที่ผ่านการสังเคราะห์ข้อมูล

ในการค้นหาเปอร์เซ็นต์ไทล์นั้น ตามหลักการแล้ว จะใช้สูตร PERCENTRANK.INC ในการคำนวณ ซึ่งในการทดลองนี้จะนำสูตร PERCENTRANK.EXC มาเปรียบเทียบกับให้เห็นอย่างชัดเจนด้วย

	PERCENTRANK.EXC			PERCENTRANK.INC		
	R	F	M	R	F	M
CUSTOMER01	5.80%	5.80%	5.80%	0.00%	0.00%	0.00%
CUSTOMER02	82.30%	23.50%	23.50%	86.60%	20.00%	20.00%
CUSTOMER03	11.70%	17.60%	52.90%	6.60%	13.30%	53.30%
CUSTOMER04	41.10%	47.00%	82.30%	40.00%	46.60%	86.60%
CUSTOMER05	29.40%	58.80%	17.60%	26.60%	60.00%	13.30%
CUSTOMER06	88.20%	70.50%	58.80%	93.30%	73.30%	60.00%
CUSTOMER07	47.00%	35.20%	29.40%	46.60%	33.30%	26.60%
CUSTOMER08	64.70%	76.40%	64.70%	66.60%	80.00%	66.60%
CUSTOMER09	94.10%	35.20%	88.20%	100.00%	33.30%	93.30%
CUSTOMER10	35.20%	29.40%	47.00%	33.30%	26.60%	46.60%
CUSTOMER11	58.80%	76.40%	70.50%	60.00%	80.00%	73.30%
CUSTOMER12	70.50%	64.70%	35.20%	73.30%	66.60%	33.30%
CUSTOMER13	23.50%	88.20%	76.40%	20.00%	93.30%	80.00%
CUSTOMER14	17.60%	11.70%	11.70%	13.30%	6.60%	6.60%
CUSTOMER15	52.90%	47.00%	41.10%	53.30%	46.60%	40.00%
CUSTOMER16	76.40%	94.10%	94.10%	80.00%	100.00%	100.00%

ภาพที่ 4.3 ข้อมูลตัวอย่างเมื่อใช้สูตร PERCENTRANK

แต่ในการเขียนโปรแกรมนั้น ไม่อาจใช้ทุกสูตรคำนวณได้ เนื่องจากไม่มีโปรแกรมที่มารองรับสูตรที่จะนำมาอ้างอิงได้ ดังนั้น จึงได้มีการศึกษาเพิ่มเติมเพื่อให้ได้ผลลัพธ์ใกล้เคียงกับที่ต้องการ โดยในที่นี้จะนำสูตร RANK.EQ/COUNT มาใช้แทนสูตร PERCENTRANK.INC ซึ่งในภาพต่อไปนี้จะเป็นการแสดงผลลัพธ์จากสูตรดังกล่าว

	RANK.EQ			RANK.EQ/COUNT		
	R	F	M	R	F	M
CUSTOMER01	1	1	1	0.00%	0.00%	0.00%
CUSTOMER02	14	4	4	86.67%	20.00%	20.00%
CUSTOMER03	2	3	9	6.67%	13.33%	53.33%
CUSTOMER04	7	8	14	40.00%	46.67%	86.67%
CUSTOMER05	5	10	3	26.67%	60.00%	13.33%
CUSTOMER06	15	12	10	93.33%	73.33%	60.00%
CUSTOMER07	8	6	5	46.67%	33.33%	26.67%
CUSTOMER08	11	13	11	66.67%	80.00%	66.67%
CUSTOMER09	16	6	15	100.00%	33.33%	93.33%
CUSTOMER10	6	5	8	33.33%	26.67%	46.67%
CUSTOMER11	10	13	12	60.00%	80.00%	73.33%
CUSTOMER12	12	11	6	73.33%	66.67%	33.33%
CUSTOMER13	4	15	13	20.00%	93.33%	80.00%
CUSTOMER14	3	2	2	13.33%	6.67%	6.67%
CUSTOMER15	9	8	7	53.33%	46.67%	40.00%
CUSTOMER16	13	16	16	80.00%	100.00%	100.00%

ภาพที่ 4.4 ข้อมูลตัวอย่างเมื่อใช้สูตร RANK.EQ/COUNT

เมื่อสามารถได้ค่าจากตำแหน่งเปอร์เซ็นต์ไทล์แล้ว จะนำมาแบ่งเป็น 5 ส่วน เพื่อจัดมาเป็นระดับคะแนน โดยระดับคะแนนที่ได้จากทั้ง 3 ค่า นั้น จะมานำพิจารณาในการแบ่งกลุ่ม

	CEILING.MATH			CONVERT
	R	F	M	
CUSTOMER01	1	1	1	ไม่ค่อยสำคัญ
CUSTOMER02	5	1	1	ลูกค้าใหม่จ่ายน้อย
CUSTOMER03	1	1	3	อื่นๆ
CUSTOMER04	2	3	5	ลูกค้าประจำแต่หายไปนาน
CUSTOMER05	2	3	1	ลูกค้าประจำแต่หายไปนาน
CUSTOMER06	5	4	3	ลูกค้าประจำ
CUSTOMER07	3	2	2	อื่นๆ
CUSTOMER08	4	4	4	สุดยอดลูกค้า
CUSTOMER09	5	2	5	นานมาที่จ่ายเยอะ
CUSTOMER10	2	2	3	อื่นๆ
CUSTOMER11	3	4	4	ลูกค้าประจำ
CUSTOMER12	4	4	2	มาบ่อยจ่ายน้อย
CUSTOMER13	1	5	4	เคยเป็นสุดยอดแต่หายไปนาน
CUSTOMER14	1	1	1	ไม่ค่อยสำคัญ
CUSTOMER15	3	3	2	ลูกค้าประจำ
CUSTOMER16	4	5	5	สุดยอดลูกค้า

ภาพที่ 4.5 ข้อมูลตัวอย่างที่ผ่านการวิเคราะห์

4.2 เว็บไซต์ RFM model

จากการจำลอง RFM model ผ่านโปรแกรม Excel สำเร็จแล้ว จะนำมาเป็นต้นแบบในการสร้างเว็บไซต์ดังกล่าว ซึ่งเป็นเว็บไซต์ที่เน้นเขียนในภาษา JavaScript เพื่อให้สะดวกต่อการนำมาใช้งานต่อภายหลัง

โดยการสร้างเว็บไซต์นั้น จะอ้างอิงจากแบบจำลองการคำนวณ RFM model ซึ่งจะมีการปรับใช้สูตรให้เหมาะสมกับวิธีเขียนเว็บไซต์ตามภาษานั้นๆ ซึ่งในที่นี้ ภาษาที่สร้างจะเน้นไปที่ภาษา JavaScript เนื่องจากเป็นภาษาที่ถูกนำไปพัฒนาต่อร่วมกับภาษาอื่นได้อย่างหลากหลาย

ในขอบเขตการวิจัยที่กล่าวไว้ ณ บทที่ 1 ส่วนของ Functional Requirement ที่ผู้พัฒนาได้รับรายละเอียดมานั้น จะพัฒนาตามความต้องการของลูกค้าและบริษัท และในส่วน Non - Functional Requirement นั้น เป็นเพียงความคิดเห็นของผู้พัฒนาที่เห็นสมควรว่า ฟังก์ชันดังกล่าวที่ยกมานั้น เป็นส่วนเสริมที่พัฒนาเพื่อความสะดวกสบาย และตอบสนองต่อผู้ใช้ที่ต้องการระบบที่ปรับใช้ได้อย่างอิสระ

4.2.1 Functional Requirement

4.2.1.1 ผู้ใช้สามารถนำเข้าข้อมูลที่เป็นนามสกุลไฟล์ .xlsx หรือ .xls และส่งออกข้อมูลที่เป็นนามสกุลไฟล์ .csv ได้

ในส่วนนี้ จะแบ่งเป็น 2 ส่วน คือ ส่วนที่นำเข้าและส่วนที่ส่งออกของไฟล์ข้อมูล

1) ส่วนที่นำเข้าไฟล์ข้อมูล

เป็นการนำเข้าข้อมูล transaction ของลูกค้าเอาเข้ามาผ่านการอัปโหลดไฟล์ที่มีนามสกุลไฟล์ .xlsx หรือ .xls เมื่อผู้ใช้ทำการกดปุ่ม (Button) บนเว็บไซต์ดังภาพที่ 4.1 และหลังจากการที่นำเข้าข้อมูลสำเร็จแล้ว จะมีการแสดงผลออกมาในรูปแบบตาราง



ภาพที่ 4.6 แผนภาพส่วนที่นำเข้าไฟล์ข้อมูล

ตามภาพที่ 4.6 แสดงว่า หลังจากที่นำเข้าข้อมูลสู่เว็บไซต์แล้ว จะมีการดึงข้อมูลที่อยู่ในไฟล์เข้าสู่ตัวแปรโดยบันทึกชื่อลูกค้า และข้อมูล transaction ทั้งหมดของลูกค้าคนนั้นๆ โดยสามารถเขียนฟังก์ชันผ่านภาษา JavaScript ได้ดังนี้

โปรแกรมที่ 4.1 ฟังก์ชันการอ่านข้อมูลที่น่าเข้าและแสดงผลในรูปแบบตาราง

```

const excel_file = document.getElementById('excel_file');
excel_file.addEventListener('change', (event) => {
    if (!['application/vnd.openxmlformats-officedocument.spreadsheetml.sheet', 'application/vnd.ms-excel'].includes(event.target.files[0].type)) {
        document.getElementById('excel_data').innerHTML =
        '<div class="alert alert-danger">Only .xlsx or .xls file
        format are allowed</div>';
        excel_file.value = '';
        return false;
    }
    var reader = new FileReader();
    reader.readAsArrayBuffer(event.target.files[0]);
    reader.onload = function (event) {
        var data = new Uint8Array(reader.result);
        var work_book = XLSX.read(data, { type: 'array' });
        var sheet_name = work_book.SheetNames;
        var sheet_data =
        XLSX.utils.sheet_to_json(work_book.Sheets[sheet_name[0]], {
        header: 1 });
        if (sheet_data.length > 0) {
            var table_output = '<table class="table table-
            striped table-bordered">';
            for (var row = 0; row < sheet_data.length;
            row++) {
                table_output += '<tr>';
                for (var cell = 0; cell <
                sheet_data[row].length; cell++) {
                    if (row == 0) {
                        table_output += '<th>' +
                        sheet_data[row][cell] + '</th>';
                    }
                    else {
                        table_output += '<td>' +
                        sheet_data[row][cell] + '</td>';
                    }
                }
                table_output += '</tr>';
            }
            table_output += '</table>';
            document.getElementById('excel_data').innerHTML
            = table_output;
        }
        excel_file.value = '';
    }
});

```


โปรแกรมที่ 4.2 โปรแกรมบันทึกข้อมูลเข้าสู่ตัวแปร

```
if (row != 0 && typeof sheet_data[row][0] !== 'undefined') {
  if (user.indexOf(sheet_data[row][0]) == -1) {
    user.push(sheet_data[row][0]);
    dataofuser.push({ name: sheet_data[row][0],
transactions: [[date, sheet_data[row][2]] ]});
  }
  else {
    for (let i = 0; i < user.length; i++) {
      if (dataofuser[i].name == sheet_data[row][0]) {
        dataofuser[i].transactions.push([date,
sheet_data[row][2]]);
      }
    }
  }
}
}
```

จากข้างต้น โปรแกรมที่ 4.2 นั้น ในความจริงจะถูกแทรกอยู่ในระหว่างโปรแกรมที่ 4.1 เพื่อเก็บข้อมูลของตัวแปร เมื่อมีการอ่านค่าข้อมูลของไฟล์ ซึ่งในที่นี้ จะแยกแยะโปรแกรมเพื่อให้ทราบถึงจุดประสงค์ของโปรแกรมที่เขียนขึ้น โดยสามารถอ่านเพิ่มเติมได้ที่ภาคผนวก

2) ส่วนที่ส่งออกข้อมูล

เป็นส่วนที่จะปรากฏหลังจากนำข้อมูลเข้า ซึ่งเมื่อผู้ใช้ทำการกดปุ่ม (Button) บนเว็บไซต์ดังภาพที่ 4.2 จะส่งออกข้อมูลหลังผ่านการวิเคราะห์แล้ว ซึ่งประกอบด้วยข้อมูลชื่อลูกค้าและกลุ่มประเภทลูกค้าในรูปแบบนามสกุลไฟล์ .csv



ภาพที่ 4.7 แผนภาพส่วนที่ส่งออกข้อมูล

จากภาพที่ 4.7 นั้น จะเป็นการเก็บค่าการวิเคราะห์ผล RFM model และบันทึกลงในตัวแปรดังกล่าว หลังจากที่ทำกรกดปุ่มเรียกใช้ฟังก์ชันส่งออกข้อมูล โปรแกรมจะทำงานโดยการอ่านค่าภายในตัวแปรแล้วเขียนในรูปแบบที่สามารถส่งออกเป็นไฟล์ .csv ได้ดังโปรแกรมที่ 4.3

โปรแกรมที่ 4.3 ฟังก์ชันส่งออกข้อมูลในรูปแบบไฟล์ .csv

```
function download_csv_file() {
  //define the heading for each row of the data
  var csv = 'Username, Segmentresult\n';

  //merge the data with CSV
```

```

    if (csvFileData != null) {
        csvFileData.forEach(function (row) {
            csv += row.join(',');
            csv += "\n";
        });
    }

    var hiddenElement = document.createElement('a');
    hiddenElement.href = 'data:text/csv;charset=utf-8,%EF%BB%BF' + encodeURIComponent(csv);
    hiddenElement.click();
}

```

4.2.1.2 ผู้ใช้สามารถอ่านข้อมูลก่อนและหลังทำการวิเคราะห์ RFM model ได้

ส่วนนี้ จะเป็นเพียงแสดงค่าข้อมูลที่ได้จัดเก็บไว้ หากเป็นข้อมูลก่อนการวิเคราะห์นั้น หมายความว่าข้อมูลจริงที่ได้จากไฟล์ตั้งโปรแกรมที่ 4.1 ที่มีการแสดงในรูปแบบตาราง และข้อมูลหลังจากการวิเคราะห์นั้น จะเกี่ยวข้องกับ Functional Requirement ข้อที่ 4.2.1.3 ที่กล่าวถึงลูกค้าคนใด อยู่ในประเภทไหน แต่ในที่นี้จะเพิ่มข้อมูลระหว่างวิเคราะห์อย่างเปอร์เซ็นต์ไทม์ของแต่ละข้อมูล (ความล่าช้า ความถี่และปริมาณเงินที่ใช้ของลูกค้า)

4.2.1.3 ระบบสามารถแสดงผลหลังจากวิเคราะห์ได้ว่า ลูกค้าคนใด อยู่ในประเภทไหน

เป็นการวิเคราะห์จากค่าระดับคะแนนที่ได้ผ่านการวิเคราะห์เทียบกับช่วงระดับการจัดกลุ่มที่ผู้ใช้ได้กำหนดไว้ โดยค่าที่ส่งเข้าผ่านโปรแกรมที่ 4.3 นั้น จะเป็นค่าที่ผ่านการจัดช่วง 5 ช่วง เว้นค่าที่เป็น 0 ที่ต้องมีการเปลี่ยนเป็นค่าระดับ 1 ใหม่อีกครั้ง เนื่องจากการคำนวณก่อนส่งค่าเข้าฟังก์ชันเป็น $\text{Math.ceil}(\text{temp} / 20)$ ซึ่งเป็นคำสั่งที่คล้ายคลึงกับสูตรในโปรแกรม Excel คือ CEILING.MATH ดังนั้นจะมีโอกาสกรณีที่ผลลัพธ์เท่ากับ 0 เช่นเดียวกัน

โปรแกรมที่ 4.4 ฟังก์ชันแปลงค่าจาก RFM เป็นกลุ่มลูกค้า

```

function convertRFM(R, F, M) {
    R = R == 0 ? 1 : R;
    F = F == 0 ? 1 : F;
    M = M == 0 ? 1 : M;

    console.log("RFM:", R, F, M);
    for (let i = 3; i < table.rows.length; i++) {
        if ((R >= sendvalue(i, 1) && R <= sendvalue(i, 2))
            && (F >= sendvalue(i, 3) && F <= sendvalue(i, 4)) && (M >=
            sendvalue(i, 5) && M <= sendvalue(i, 6))) {
            return table.rows[i].cells[0].innerHTML;
        }
    }
    return 'ไม่จัดอยู่ในกลุ่มใด';
}

```

จากโปรแกรมที่ 4.4 นั้น จะอ่านจากตารางที่บันทึกไว้ที่ระดับคะแนนไว้ ซึ่งได้ทำฟังก์ชัน sendvalue ดังโปรแกรมที่ 4.5 เพื่อสะดวกต่อการเขียนโปรแกรมเท่านั้น ซึ่งเป็นโปรแกรมที่จะช่วยในการดึงค่าจากตารางตามแถวและคอลัมน์ที่ระบุไว้

โปรแกรมที่ 4.5 ฟังก์ชันอ่านค่าจากตาราง

```
function sendvalue(r, c) {
    return table.rows[r].cells[c].innerHTML;
}
```

4.2.1.4 ทุกครั้งที่เปลี่ยนข้อมูลที่จะอ่าน ระบบจะมีการลบข้อมูลเดิมทิ้งทุกครั้ง

ในการเขียนโปรแกรมนั้น จำเป็นที่จะต้องมีการประกาศชนิดตัวแปรทุกครั้ง トラบใดที่ไม่มีการโหลดหน้าเดิมซ้ำอีกครั้ง (refresh) ข้อมูลภายในตัวแปรจะยังคงอยู่ เมื่ออัปเดตไฟล์ใหม่เข้าไปอีกครั้ง จำเป็นที่จะต้องลบค่าจากไฟล์ก่อนหน้านั้นทิ้ง เหลือเพียงค่าใหม่เท่านั้น ซึ่งจะแก้ไขปัญหาเกี่ยวกับการประกาศค่าว่างหรือลบข้อมูลภายในตัวแปรก่อนรับข้อมูลใหม่อีกครั้ง

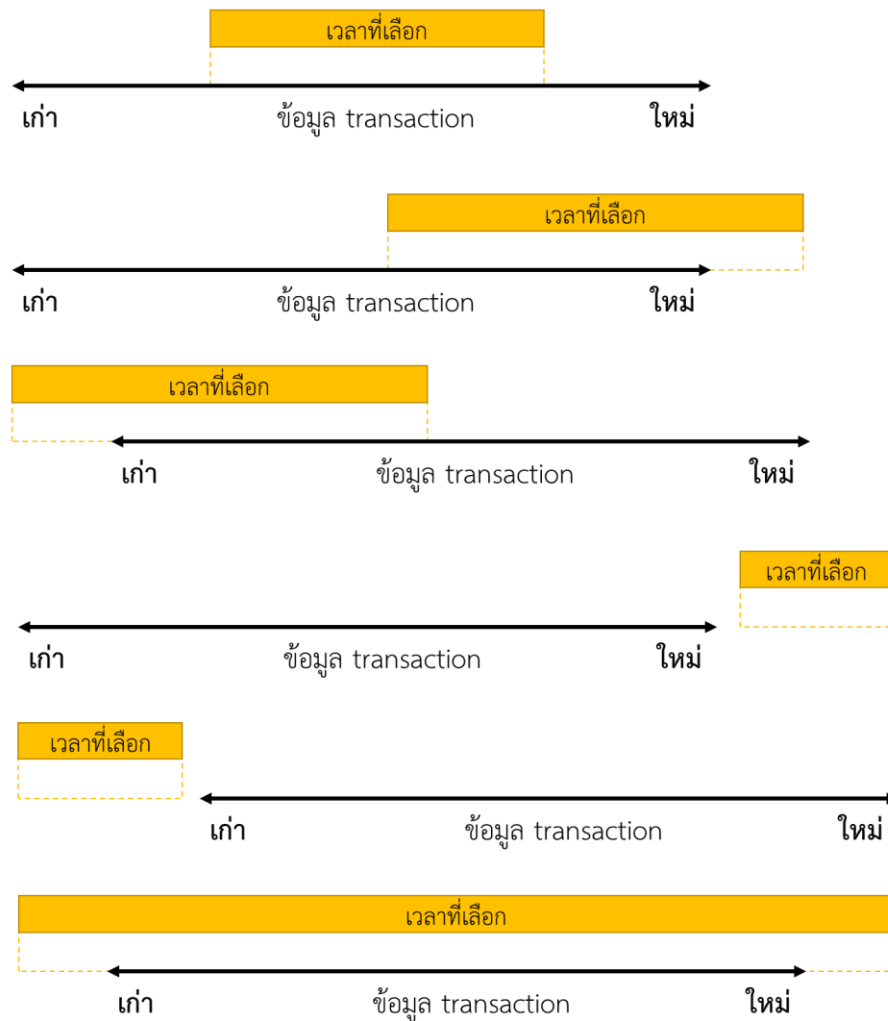
4.2.1.5 ทุกครั้งที่มีความผิดพลาดของระบบ ต้องมีการแจ้งว่าไม่สำเร็จเพราะอะไร

ในการเขียนโปรแกรมนั้น เมื่อผู้ใช้ออกค่าผิดพลาดจะต้องมีการแจ้งเตือนโดยใช้คำสั่ง alert ตามด้วยข้อความที่ผู้เขียนต้องการแจ้งเตือน ซึ่งจะไม่รวมการอ่านข้อมูลในไฟล์หากพบเจอกรณีที่หัวข้อของคอลัมน์หรือแอตทริบิวต์ (Attribute) ไม่ตรงตามที่กำหนดไว้

4.2.2 Non - Functional Requirement

4.2.2.1 ผู้ใช้สามารถกำหนดช่วงระยะเวลาในการคำนวณได้

ในเว็บไซต์นั้น สามารถตั้งค่าเวลาเพื่อดูข้อมูลในช่วงเวลานั้นได้ ซึ่งเวลาที่ไมเกี่ยวข้องหรืออยู่นอกกรอบเวลาที่ต้องการค้นหา จะไม่ถูกนำมาประมวลผลในการวิเคราะห์ แต่หากกรณีที่ไม่มีข้อมูลโดยอยู่ในช่วงเวลานั้นดังภาพที่ 4.3 จะมีการส่งค่า -1 กลับมาในทุกค่าอย่างความล่าช้า ความถี่ และปริมาณเงินที่ใช้ โดยข้อมูลทุกค่า นั้น ต้องมีการเรียงลำดับเวลาจากเก่าไปสู่อใหม่ (หากคิดในอีกมุมหนึ่งคือ เรียงจากน้อยไปหามาก)



ภาพที่ 4.8 ช่วงเวลาทั้งหมดที่สามารถเกิดขึ้นเมื่อเลือกข้อมูล

จากภาพที่ 4.8 นั้น แสดงถึงช่วงเวลาที่ผู้ใช้ต้องเลือกค้นหาเพื่อวิเคราะห์ข้อมูล ซึ่งหากไม่ได้กำหนดช่วงเวลาไว้ ระบบจะมีการกำหนดอัตโนมัติเป็นช่วงระยะเวลา 1 ปีก่อน ณ เวลาปัจจุบันถึงเวลาปัจจุบันตามที่โปรแกรมที่ 4.5 ได้ตั้งค่าไว้เป็นค่าเริ่มต้น

โปรแกรมที่ 4.6 โปรแกรมตั้งเวลาเริ่มต้น

```
let start = new Date();
start.setFullYear(start.getFullYear() - 1);
document.getElementById('start_select').value =
start.toISOString().substring(0, 10);
let end = new Date();
document.getElementById('end_select').value =
end.toISOString().substring(0, 10);
```

โปรแกรมที่ 4.7 โปรแกรมวิเคราะห์ในส่วนตรวจสอบเวลา

```
data.sort(function (a, b) { return new Date(a[0]) - new
Date(b[0]); });
    var start_select = new
Date(document.getElementById('start_select').value);
    start_select.setHours(0, 0, 0, 0);
    var end_select = new
Date(document.getElementById('end_select').value);
    end_select.setHours(0, 0, 0, 0);
    let start_datareal = data[0][0];
    let end_datareal = data[data.length - 1][0];

    if (((start_datareal < start_select) &&
(end_datareal < start_select)) || ((start_datareal >
end_select) && (end_datareal > end_select))) {
        return { Name: name, Lastdate: -1.0, Total: -
1.0, Freq: -1.0 };
    }
```

โปรแกรมที่ 4.8 โปรแกรมแปลงเวลาจากโปรแกรม Excel เป็นภาษา JavaScript

```
var date = new Date(Math.round((sheet_data[row][cell] -
(25568 + 1)) * 86400 * 1000));
date.setHours(0, 0, 0, 0);
```

ในข้างต้น เป็นคำสั่งเพิ่มเติมสำหรับกรณีที่มีการดึงค่าจากโปรแกรม Excel ซึ่งเป็นคำสั่งที่ควรเขียนก่อนที่จะนำมาคำนวณร่วมกับโปรแกรมในภาษา JavaScript เนื่องจากการนำค่ามาใช้โดยไม่ผ่านคำสั่งนี้จะมีส่งผลให้เกิดข้อผิดพลาดขึ้น ดังตารางที่ 3.6

ตารางที่ 4.1 ตารางเปรียบเทียบวันเวลาจากโปรแกรม Excel และภาษา JavaScript

วันและเวลาใน		วันและเวลาในภาษา JavaScript		
โปรแกรม Excel		เวลาโดยอิงตาม Excel		เวลาที่ควรจะเป็น
12 พฤศจิกายน 2021	44512	Thu Jan 01 1970 07:00:44 GMT+0700 (Indochina Time)	1636675200000	Fri Nov 12 2021 07:00:00 GMT+0700 (Indochina Time)
13 พฤศจิกายน 2021	44513	Thu Jan 01 1970 07:00:44 GMT+0700 (Indochina Time)	1636761600000	Sat Nov 13 2021 07:00:00 GMT+0700 (Indochina Time)
14 พฤศจิกายน 2021	44514	Thu Jan 01 1970 07:00:44 GMT+0700 (Indochina Time)	1636848000000	Sun Nov 14 2021 07:00:00 GMT+0700 (Indochina Time)

วันและเวลาในโปรแกรม Excel		วันและเวลาในภาษา JavaScript		
		เวลาโดยอิงตาม Excel	เวลาที่ควรจะเป็น	
15 พฤศจิกายน 2021	44515	Thu Jan 01 1970 07:00:44 GMT+0700 (Indochina Time)	1636934400000	Mon Nov 15 2021 07:00:00 GMT+0700 (Indochina Time)
16 พฤศจิกายน 2021	44516	Thu Jan 01 1970 07:00:44 GMT+0700 (Indochina Time)	1637020800000	Tue Nov 16 2021 07:00:00 GMT+0700 (Indochina Time)
17 พฤศจิกายน 2021	44517	Thu Jan 01 1970 07:00:44 GMT+0700 (Indochina Time)	1637107200000	Wed Nov 17 2021 07:00:00 GMT+0700 (Indochina Time)

จากตารางที่ 4.1 นั้น ในส่วนวันเวลาของภาษา JavaScript จะแบ่งเป็น 2 ส่วน คือ เวลาที่ถูกการแปลงเป็นตัวเลขด้วยคำสั่ง Date.parse() และเวลาที่แปลงกลับจากตัวเลขที่ผ่านคำสั่ง Date.parse() ซึ่งจากการสังเกตนั้น เมื่อเวลาที่มาจากการไฟล์ไม่ตรงกับสิ่งที่ควรจะเป็น ดังนั้นจะต้องหาจุดร่วมในการคำนวณแปลงค่าให้เหมาะสมกับภาษา JavaScript

ตัวเลขในข้างต้น จึงได้ข้อสรุปจากการหาส่วนต่าง คือ เวลาโดยอิงตาม Excel จะเพิ่มขึ้นทีละ 1 ทุกครั้งเมื่อเปลี่ยนวัน และเวลาที่ควรจะเป็นในภาษา JavaScript จะต่างกัน 86400000 เมื่อห่างกัน 1 วัน

ตารางที่ 4.2 ตารางคำนวณความแตกต่างของวันเวลาเพื่อค้นหาสูตร

วันที่	เวลาที่ควรเป็นหารด้วย 86400000 ที่เป็นส่วนต่างของวัน	ส่วนต่างของวันที่ผ่านการหารกับวัน อิงตาม Excel
12 พฤศจิกายน 2021	= 1636675200000/86400000 = 18943	= 44512 - 18943 = 25569
13 พฤศจิกายน 2021	= 1636761600000/86400000 = 18944	= 44513 - 18944 = 25569
14 พฤศจิกายน 2021	= 1636848000000/86400000 = 18945	= 44514 - 18945 = 25569
15 พฤศจิกายน 2021	= 1636934400000/86400000 = 18946	= 44515 - 18946 = 25569
16 พฤศจิกายน 2021	= 1637020800000/86400000 = 18947	= 44516 - 18947 = 25569
17 พฤศจิกายน 2021	= 1637107200000/86400000 = 18948	= 44517 - 18948 = 25569

ตารางที่ 4.3 ตารางเปรียบเทียบก่อนและหลังตั้งค่าเวลา

วันที่	เวลาก่อน setHours	เวลาหลัง setHours (0, 0, 0, 0)	ส่วนที่แตกต่าง
12 พฤศจิกายน 2021	1636675200000	1636650000000	25200000
13 พฤศจิกายน 2021	1636761600000	1636736400000	25200000
14 พฤศจิกายน 2021	1636848000000	1636822800000	25200000
15 พฤศจิกายน 2021	1636934400000	1636909200000	25200000
16 พฤศจิกายน 2021	1637020800000	1636995600000	25200000
17 พฤศจิกายน 2021	1637107200000	1637082000000	25200000

4.2.2.2 ผู้ใช้สามารถสร้างหรือแบ่งกลุ่มประเภทของลูกค้าตามช่วงระดับด้วยตัวเองได้

เป็นส่วนเสริมที่จะสามารถจัดกลุ่มได้อย่างอิสระ เนื่องจากในบางครั้ง ผู้ใช้อาจจะต้องการกลุ่มหรือประเภทของลูกค้าที่มากกว่าที่ส่งมอบทั่วไปได้ใช้งาน แต่ทั้งนี้ ได้มีการกำหนดรายละเอียดในการเพิ่มกลุ่มประเภทด้วยตนเอง คือ ชื่อกลุ่ม และระดับน้อยที่สุด มากที่สุดของมูลค่าสูงสุด ความถี่ และปริมาณการใช้เงิน ส่วนระดับนั้นสามารถมีค่าได้ตั้งแต่ 1 ถึง 5 เท่านั้น และสามารถลบระดับคะแนนที่สร้างได้ ซึ่งสามารถเขียนโปรแกรมได้ดังโปรแกรมที่ 4.8 และโปรแกรมที่ 4.9 ที่เป็นฟังก์ชันเพิ่มหรือลบข้อมูลที่สร้างไว้

โปรแกรมที่ 4.9 โปรแกรมต้นแบบในการแบ่งกลุ่มลูกค้า

```
<td><input type="text" id="segment" required></td>
<td class="center"><input type="number" min="1" max="5"
class="input_rfm" id="Rmin" required></td>
<td class="center"><input type="number" min="1" max="5"
class="input_rfm" id="Rmax" required></td>
<td class="center"><input type="number" min="1" max="5"
class="input_rfm" id="Fmin" required></td>
<td class="center"><input type="number" min="1" max="5"
class="input_rfm" id="Fmax" required></td>
<td class="center"><input type="number" min="1" max="5"
class="input_rfm" id="Mmin" required></td>
<td class="center"><input type="number" min="1" max="5"
class="input_rfm" id="Mmax" required></td>
<td class="center"><input class="btn_secondary"
type="submit" value="Add Segment"
class="record-submit"></td>
```

โปรแกรมที่ 4.10 ฟังก์ชันเพิ่มและลบกลุ่มลูกค้า

```

form.addEventListener("submit", addItem);
function addItem(e) {
    e.preventDefault();

    var segment =
document.getElementById("segment").value;
    var Rmin = document.getElementById("Rmin").value;
    var Rmax = document.getElementById("Rmax").value;
    var Fmin = document.getElementById("Fmin").value;
    var Fmax = document.getElementById("Fmax").value;
    var Mmin = document.getElementById("Mmin").value;
    var Mmax = document.getElementById("Mmax").value;

    if ((Rmin > Rmax) || (Fmin > Fmax) || (Mmin > Mmax))
{
        alert("อาจมีข้อผิดพลาด โปรดตรวจสอบการจัดระดับ RFM อีกครั้ง");
    }
    else if (model.indexOf(m) != -1) {
        alert("อาจมีข้อผิดพลาด โปรดตรวจสอบการจัดระดับ RFM อีกครั้ง");
    }
    else{
        //create rows
        var row = table.insertRow(3);
        //create row cells
        var segmentCell = row.insertCell(0);
        segmentCell.innerHTML = segment;

        var RminCell = row.insertCell(1);
        RminCell.innerHTML = Rmin;

        var RmaxCell = row.insertCell(2);
        RmaxCell.innerHTML = Rmax;

        var FminCell = row.insertCell(3);
        FminCell.innerHTML = Fmin;

        var FmaxCell = row.insertCell(4);
        FmaxCell.innerHTML = Fmax;

        var MminCell = row.insertCell(5);
        MminCell.innerHTML = Mmin;

        var MmaxCell = row.insertCell(6);
        MmaxCell.innerHTML = Mmax;

        model.push(m);

        var DeleteCell = row.insertCell(7);

```



```

        DeleteCell.innerHTML = "<button
class='btn_secondary'
onclick='removeRow(this) '>Remove</button>";
        DeleteCell.style.textAlign = "center";
    }
}
function removeRow(e) {
    table.deleteRow(e.parentNode.parentNode.rowIndex);
    model.splice((e.parentNode.parentNode.rowIndex - 3),
1);
    // button -> td -> tr.
}

```

4.2.2.3 ระบบสามารถแสดงประเภทกลุ่มลูกค้าเป็นรูปแบบกราฟวงกลมได้

หากข้อมูลลูกค้าที่ผ่านการวิเคราะห์มาแล้ว มีข้อมูลเยอะเกินไปนั้น การสร้างข้อมูลออกมาในรูปแบบอื่นๆ ที่ไม่ใช่ตาราง จะทำให้สามารถดูผลลัพธ์ได้ง่ายยิ่งขึ้นอย่างการสร้างกราฟวงกลม ซึ่งในระบบนั้น สามารถเลือกให้ข้อมูลเกี่ยวกับกลุ่มประเภทที่ไม่ได้อยู่ในช่วงเวลาออกได้

โปรแกรมที่ 4.11 ฟังก์ชันการสร้างกราฟวงกลม

```

<script
src="https://cdn.amcharts.com/lib/4/core.js"></script>
<script
src="https://cdn.amcharts.com/lib/4/charts.js"></script>
<script
src="https://cdn.amcharts.com/lib/4/themes/animated.js"></sc
ript>
function chart_show() {
    am4core.ready(function () {

        // Themes begin
        am4core.useTheme(am4themes_animated);
        // Themes end

        // Create chart instance
        var chart = am4core.create("chartdiv",
am4charts.PieChart);

        // Add data
        chart.data = segment_findnumof();
        console.log("CHART : ", chart.data);

        // Add and configure Series
        var pieSeries = chart.series.push(new
am4charts.PieSeries());
        pieSeries.dataFields.value = "numberof";
        pieSeries.dataFields.category = "segment";

```

```

        pieSeries.slices.template.stroke =
am4core.color("#fff");
        pieSeries.slices.template.strokeWidth = 2;
        pieSeries.slices.template.strokeOpacity = 1;

        // This creates initial animation
        pieSeries.hiddenState.properties.opacity = 1;
        pieSeries.hiddenState.properties.endAngle = -90;
        pieSeries.hiddenState.properties.startAngle = -
90;

    }); // end am4core.ready()
}

```

นอกจากนี้ ส่วนที่สำคัญในการพัฒนาอย่างการวิเคราะห์เพื่อตั้งทำ RFM model นั้น ในการเขียนด้วยภาษา JavaScript นั้น ไม่สามารถสร้าง PERCENTRANK.INC ได้ ดังนั้น จึงเปลี่ยนไปใช้วิธีทางอ้อมอย่างสูตร RANK.EQ/COUNT ที่ได้ผลลัพธ์ใกล้เคียงกันมากกว่า ซึ่งสามารถเขียนโปรแกรมได้ดังนี้ (ในที่นี้ จะยกตัวอย่างเพียงบางส่วน โดยสามารถอ่านเพิ่มเติมได้ที่ภาคผนวก)

โปรแกรมที่ 4.12 ฟังก์ชันที่อิงตามสูตร RANK.EQ

```

function PERCENTRANK(data) {
    var Setrank = {};
    data.sort(function (a, b) { return a - b });
    for (let i = 0; i < data.length; i++) {
        const item = data[i];
        if (!Setrank[item]) {
            Setrank[item] = i;
        }
    }
    return Setrank;
}

```

โปรแกรมที่ 4.13 ฟังก์ชันที่อิงตามสูตร PERCENTRANK.INC

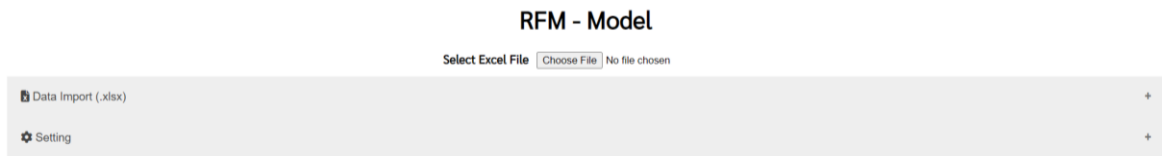
```

function PERCENTRANK_INC(val, Setrank, len) {
    return (Setrank[val] / (len - 1) * 100).toFixed(2);
}

```

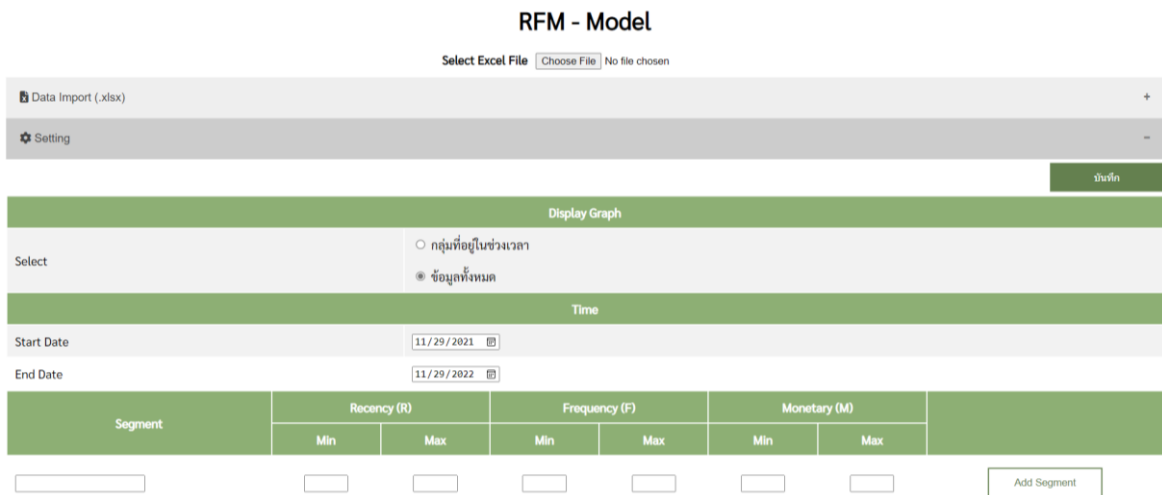
ในโปรแกรมที่ 4.11 เป็นการเก็บค่าตำแหน่งหรือ index ของตัวแปรซึ่ง คล้ายคลึงกับ สูตร RANK.EQ และในโปรแกรมที่ 4.12 เป็นการจำลองสูตร PERCENTRANK.INC โดยอิงจาก RANK.EQ/COUNT ในโปรแกรม Excel ทั้งนี้ ตัวแปรที่จะใช้ในการวิเคราะห์จะถูกรวบรวมและจัดเรียงให้ถูกต้องก่อน

จากที่กล่าวมาทั้งหมดนั้น เมื่อทราบ Requirement และแนวทางการเขียนโปรแกรม จึงสามารถจัดทำเว็บไซต์ได้ดังต่อไปนี้



ภาพที่ 4.9 หน้าเว็บไซต์เมื่อเปิดครั้งแรก

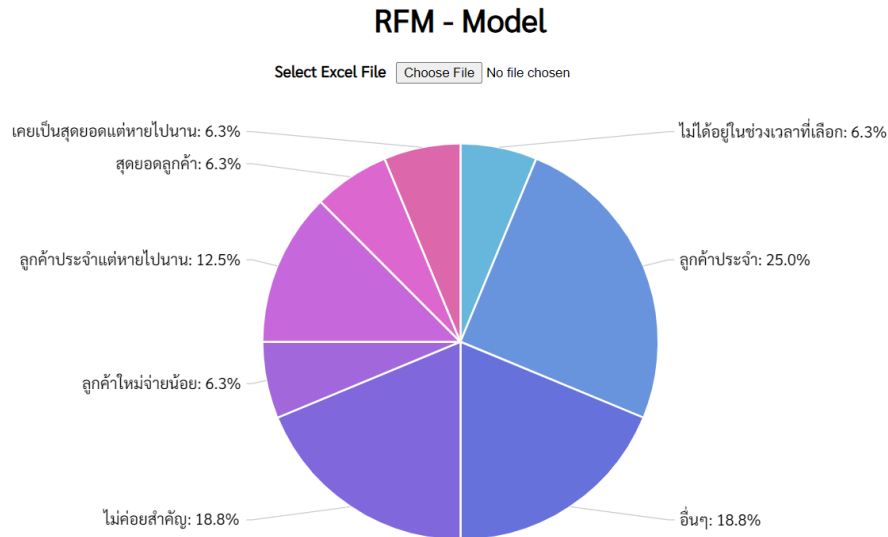
ในขั้นต้น เป็นหน้าแรกเมื่อเปิดเว็บไซต์ขึ้นมา ซึ่งจะพบปุ่ม (Button) สำหรับการนำเข้าข้อมูลมาวิเคราะห์ และแท็บย่อหน้าสำหรับข้อมูลที่นำเข้าไฟล์ (Data Import) และตั้งค่า (Setting)



ภาพที่ 4.10 หน้าเว็บไซต์ในส่วนตั้งค่า

ในขั้นต้น เป็นส่วนตั้งค่าระบบ RFM model ซึ่งประกอบด้วยประเภทข้อมูลที่นำมาแสดงในรูปแบบกราฟ ซึ่งมี 2 รูปแบบให้เลือก คือ เฉพาะข้อมูลที่อยู่ในช่วงเวลาที่กำหนด และข้อมูลทั้งหมดซึ่งอยู่ไม่จำกัดเพียงในช่วงเวลาที่เลือกไว้เท่านั้น โดยเวลานั้น สามารถตั้งค่าจุดเริ่มต้นและจุดสุดท้ายได้ และในส่วนของการแบ่งกลุ่มประเภทลูกค้าก็สามารถทำได้เช่นกัน

โดยกราฟวงกลมนั้น จะปรากฏก็ต่อเมื่อข้อมูลผ่านการวิเคราะห์เสร็จสิ้นแล้ว ซึ่งตัวอย่างจะเป็นดังภาพที่ 4.11 และต้องสอดคล้องกับข้อมูลบนตารางตามภาพที่ 4.12



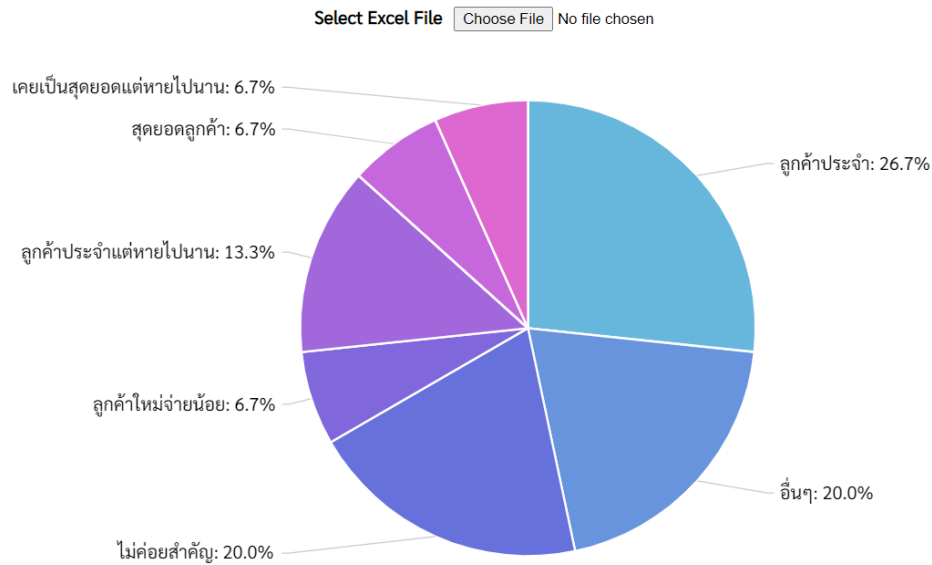
ภาพที่ 4.11 กราฟหลังจากวิเคราะห์ RFM แสดงข้อมูลทั้งหมด

Username	RFM - model			Result
	R	F	M	
CUSTOMER01	0 %	0 %	0 %	ไม่ได้อยู่ในช่วงเวลาที่เลือก
CUSTOMER11	53.33 %	66.67 %	66.67 %	ลูกค้าประจำ
CUSTOMER03	0.00 %	13.33 %	46.67 %	อื่นๆ
CUSTOMER10	26.67 %	20.00 %	26.67 %	ไม่ค่อยสำคัญ
CUSTOMER07	40.00 %	20.00 %	13.33 %	ไม่ค่อยสำคัญ
CUSTOMER14	6.67 %	0.00 %	0.00 %	ไม่ค่อยสำคัญ
CUSTOMER12	66.67 %	40.00 %	40.00 %	อื่นๆ
CUSTOMER02	80.00 %	6.67 %	6.67 %	ลูกค้าใหม่จ่าน้อย
CUSTOMER06	86.67 %	66.67 %	53.33 %	ลูกค้าประจำ
CUSTOMER15	46.67 %	33.33 %	33.33 %	อื่นๆ
CUSTOMER05	20.00 %	60.00 %	20.00 %	ลูกค้าประจำแต่หายไปนาน
CUSTOMER16	73.33 %	93.33 %	93.33 %	สุดท้ายลูกค้า
CUSTOMER13	13.33 %	86.67 %	73.33 %	เคยเป็นสุดท้ายแต่หายไปนาน
CUSTOMER09	93.33 %	46.67 %	86.67 %	ลูกค้าประจำ
CUSTOMER08	60.00 %	80.00 %	60.00 %	ลูกค้าประจำ
CUSTOMER04	33.33 %	53.33 %	80.00 %	ลูกค้าประจำแต่หายไปนาน

ภาพที่ 4.12 ตารางข้อมูลหลังจากวิเคราะห์ RFM แสดงข้อมูลทั้งหมด

และเมื่อตั้งค่าให้แสดงข้อมูลเฉพาะส่วนที่อยู่ในช่วงเวลาที่เลือกไว้ นั่น ข้อมูลส่วนที่อยู่นอกเหนือจะถูกตัดออก ซึ่งหากเทียบกับภาพที่ 4.11 และภาพที่ 4.13 แล้วนั้น อาจจะมีแตกต่างเล็กน้อยเมื่อตัดข้อมูลส่วนที่ไม่ต้องการออกไป

RFM - Model

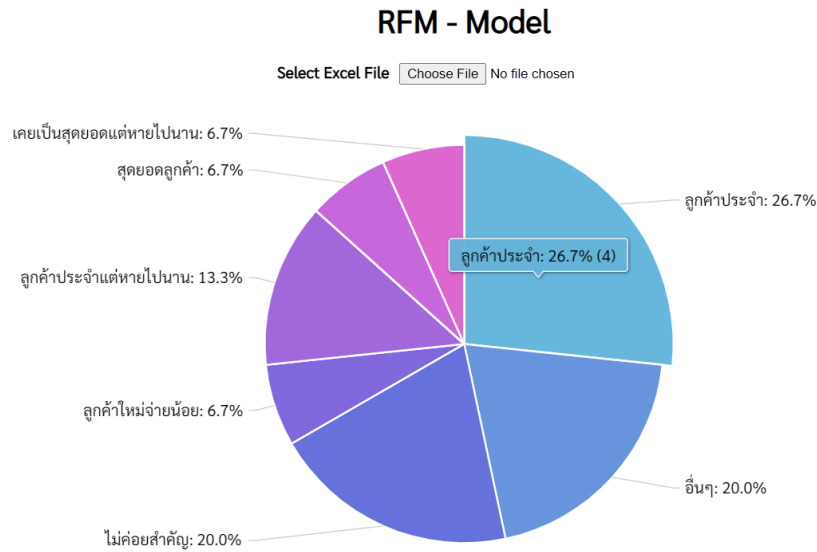


ภาพที่ 4.13 กราฟหลังจากวิเคราะห์ RFM แสดงข้อมูลเฉพาะช่วงเวลา

Username	RFM - model			Result
	R	F	M	
CUSTOMER11	53.33 %	66.67 %	66.67 %	ลูกค้าประจำ
CUSTOMER03	0.00 %	13.33 %	46.67 %	อื่นๆ
CUSTOMER10	26.67 %	20.00 %	26.67 %	ไม่ค่อยสำคัญ
CUSTOMER07	40.00 %	20.00 %	13.33 %	ไม่ค่อยสำคัญ
CUSTOMER14	6.67 %	0.00 %	0.00 %	ไม่ค่อยสำคัญ
CUSTOMER12	66.67 %	40.00 %	40.00 %	อื่นๆ
CUSTOMER02	80.00 %	6.67 %	6.67 %	ลูกค้าใหม่จ่ายน้อย
CUSTOMER06	86.67 %	66.67 %	53.33 %	ลูกค้าประจำ
CUSTOMER15	46.67 %	33.33 %	33.33 %	อื่นๆ
CUSTOMER05	20.00 %	60.00 %	20.00 %	ลูกค้าประจำแต่หายไปนาน
CUSTOMER16	73.33 %	93.33 %	93.33 %	สุดท้ายลูกค้า
CUSTOMER13	13.33 %	86.67 %	73.33 %	เคยเป็นสุดท้ายแต่หายไปนาน
CUSTOMER09	93.33 %	46.67 %	86.67 %	ลูกค้าประจำ
CUSTOMER08	60.00 %	80.00 %	60.00 %	ลูกค้าประจำ
CUSTOMER04	33.33 %	53.33 %	80.00 %	ลูกค้าประจำแต่หายไปนาน

ภาพที่ 4.14 ตารางข้อมูลหลังจากวิเคราะห์ RFM แสดงข้อมูลเฉพาะช่วงเวลา

และจุดเด่นอีกอย่าง คือ กราฟที่นำมาแสดงนั้น เมื่อนำ cursor มาวางบนส่วนของกราฟ จะมีข้อความเกี่ยวข้องกลุ่มประเภท เพอร์เซ็นต์เมื่อเทียบกับข้อมูลทั้งหมด และจำนวนของข้อมูลนั้นๆ ดังภาพที่ 4.15



ภาพที่ 4.15 ลูกเล่นของกราฟแสดงผล

ในลำดับต่อมา จะเป็นส่วนในการเพิ่มหรือลบกลุ่มประเภทของลูกค้า ซึ่งสามารถรอกชื่อกลุ่มประเภท และระดับของความล่าสุด ความถี่ และปริมาณเงินที่ใช้ได้อย่างตามใจชอบดังภาพที่ 4.16 แต่มีข้อกำหนดในด้านการเพิ่มกลุ่มที่ต้องกรอกข้อมูลในทุกช่องว่าง มิเช่นนั้นจะมีการแจ้งเตือนไปตามภาพที่ 4.17 และเมื่อมีการกรอกข้อมูลที่ไม่ถูกต้อง ก็จะมีการแจ้งเตือนเช่นกันตามภาพที่ 4.18

RFM - Model

Select Excel File No file chosen

Data Import (.xlsx)

Setting

Display Graph

Select

☐ กลุ่มที่อยู่ในช่วงเวลา
☒ ข้อมูลทั้งหมด

Time

Start Date

12/14/2021

End Date

12/14/2022

Segment	Recency (R)		Frequency (F)		Monetary (M)		
	Min	Max	Min	Max	Min	Max	
สุดยอดลูกค้า	4	5	4	5	4	5	Add Segment
สุดยอดลูกค้า	4	5	4	5	4	5	Remove

ภาพที่ 4.16 ตัวอย่างเมื่อกรอกข้อมูลในส่วนการแบ่งประเภทกลุ่มลูกค้า

Segment	Recency (R)		Frequency (F)		Monetary (M)		
	Min	Max	Min	Max	Min	Max	
	1	5	1	5	1	5	Add Segment

Please fill out this field.

ภาพที่ 4.17 หน้าต่างแจ้งเตือนเมื่อไม่ได้กรอกข้อมูล

This page says
อาจมีข้อผิดพลาด โปรดตรวจสอบการเชื่อมต่อ RFM อีกครั้ง

OK

Data Import (.xlsx)

Setting

บันทึก

Display Graph

Select

☐ กลุ่มที่อยู่ในช่วงเวลา

☒ ข้อมูลทั้งหมด

Time

Start Date

12/14/2021

End Date

12/14/2022

Segment	Recency (R)		Frequency (F)		Monetary (M)	
	Min	Max	Min	Max	Min	Max
แสดงลูกค้า	5	4	4	5	4	5

Add Segment

ภาพที่ 4.18 หน้าต่างแจ้งเตือนเมื่อกรอกข้อมูลผิดพลาด

เมื่อผู้ใช้ตั้งค่าข้อมูลทุกอย่างเสร็จสิ้นแล้ว จำเป็นที่ต้องกดปุ่มบันทึกก่อนทุกครั้ง ถึงจะส่งผลให้ระบบได้ทำการวิเคราะห์ใหม่อีกครั้งตามสิ่งที่ได้กำหนดไว้ และทุกครั้งที่กดบันทึกนั้น จะมีระบบแจ้งเตือนเสมอ ดังภาพที่ 4.19

CUSTOMER13 13.33 %

CUSTOMER09 93.33 %

CUSTOMER08 60.00 %

CUSTOMER04 33.33 % 53.33 % 80.00 % ลูกค้าประจำแต่หายไปนาน

This page says
Setting Change!

OK

บันทึก

Data Import (.xlsx)

Setting

บันทึก

Display Graph

Select

☒ กลุ่มที่อยู่ในช่วงเวลา

☐ ข้อมูลทั้งหมด

Time

Start Date

12/14/2021

End Date

12/14/2022

Segment	Recency (R)		Frequency (F)		Monetary (M)	
	Min	Max	Min	Max	Min	Max
แสดงลูกค้า	4	5	4	5	4	5

Add Segment

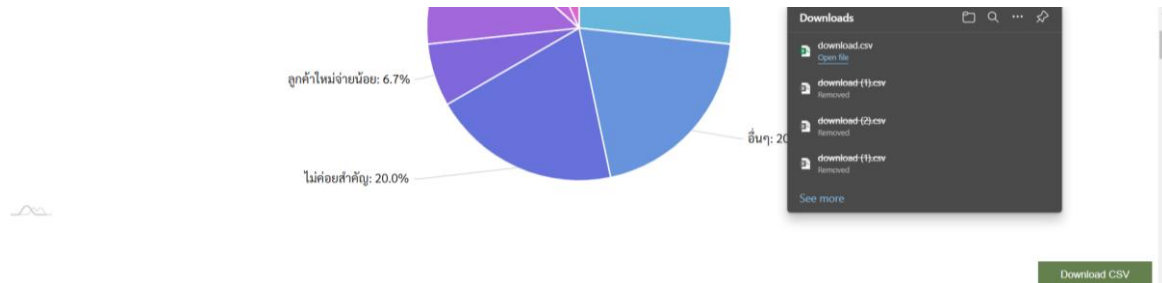
ภาพที่ 4.19 หน้าต่างแจ้งเตือนเมื่อกรอกข้อมูลสำเร็จ

ในบางครั้งนั้น ผู้ใช้อาจจะมีความต้องการในการทราบข้อมูลก่อนวิเคราะห์เกิดขึ้น ดังนั้น ทางระบบจึงได้รองรับการแสดงผลทุกครั้งเมื่อมีการนำเข้าไฟล์ เพื่อให้สามารถอ่านข้อมูลผ่านหน้าเว็บไซต์โดยไม่ต้องเปิดไฟล์ และยังเป็น การตรวจสอบด้วยอีกว่า ไฟล์ข้อมูลได้นำเข้าอย่างถูกต้องหรือไม่

Data Import (.xlsx)		
USERNAME	DATE	MONEY
CUSTOMER01	Fri Nov 12 2021 00:00:00 GMT+0700 (Indochina Time)	822
CUSTOMER11	Sat Nov 13 2021 00:00:00 GMT+0700 (Indochina Time)	111
CUSTOMER03	Sun Nov 14 2021 00:00:00 GMT+0700 (Indochina Time)	1748
CUSTOMER10	Mon Nov 15 2021 00:00:00 GMT+0700 (Indochina Time)	573
CUSTOMER07	Tue Nov 16 2021 00:00:00 GMT+0700 (Indochina Time)	179
CUSTOMER14	Wed Nov 17 2021 00:00:00 GMT+0700 (Indochina Time)	1659
CUSTOMER12	Thu Nov 18 2021 00:00:00 GMT+0700 (Indochina Time)	77

ภาพที่ 4.20 ตารางข้อมูลก่อนการวิเคราะห์ RFM model

หลังจากการวิเคราะห์ทุกอย่างและได้ผลลัพธ์ตามที่ตั้งค่าตั้งที่ผู้ใช้งานต้องการแล้วนั้น สามารถส่งออกหรือดาวน์โหลดข้อมูลออกมาได้เพียงกดที่ปุ่ม Download CSV ซึ่งไฟล์จะอยู่ในรูปแบบไฟล์ที่มีนามสกุล .csv และส่วนไฟล์ที่ปรากฏหลังจากกดปุ่มนั้น จะมีชื่อไฟล์ว่า download.csv ตามภาพที่ 4.21 และเมื่อเปิดไฟล์นั้น จะได้ผลลัพธ์ข้อมูลเพียงแค่ชื่อลูกค้า และกลุ่มประเภทตามภาพที่ 4.22 เท่านั้น



ภาพที่ 4.21 หน้าต่างที่แสดงผลเมื่อส่งออกข้อมูล

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Username	Segment	result																		
2	CUSTOMER11	ลูกค้าประจำ																			
3	CUSTOMER03	อื่นๆ																			
4	CUSTOMER10	ไม่ค่อยสำคัญ																			
5	CUSTOMER07	ไม่ค่อยสำคัญ																			
6	CUSTOMER14	ไม่ค่อยสำคัญ																			
7	CUSTOMER12	อื่นๆ																			
8	CUSTOMER02	ลูกค้าใหม่จ่ายน้อย																			
9	CUSTOMER06	ลูกค้าประจำ																			
10	CUSTOMER15	อื่นๆ																			
11	CUSTOMER05	ลูกค้าประจำแต่หายไปนาน																			
12	CUSTOMER16	ไม่ค่อยสำคัญ																			
13	CUSTOMER13	เคยเป็นลูกค้าแต่หายไปนาน																			
14	CUSTOMER09	ลูกค้าประจำ																			
15	CUSTOMER08	ลูกค้าประจำ																			
16	CUSTOMER04	ลูกค้าประจำแต่หายไปนาน																			
17																					
18																					
19																					
20																					
21																					
22																					
23																					
24																					
25																					
26																					
27																					
28																					
29																					
30																					

ภาพที่ 4.22 ข้อมูลภายในไฟล์ที่ส่งออกข้อมูลผลการวิเคราะห์

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

จากผลการใช้งานเว็บไซต์วิเคราะห์ประเภทลูกค้าด้วย RFM model แล้วนั้น พบว่าสามารถนำไปประยุกต์ใช้ เพื่อแบ่งประเภทของลูกค้าตามระดับที่ผู้ใช้ได้กำหนดไว้ และยังเป็นแนวทางในการวางแผนในเชิงกลยุทธ์ได้ ซึ่งในปัจจุบัน เว็บไซต์ที่สร้างขึ้นนี้เป็นเพียงส่วนหนึ่งในการวิเคราะห์ข้อมูลของลูกค้าที่ต้องนำเข้าสู่ข้อมูลจากข้อมูลจริงที่ผ่านระบบการซื้อ – ขายของลูกค้ามาก่อน จึงจะสามารถใช้ระบบนี้ได้ ทั้งนี้สามารถใช้การวิเคราะห์ด้วยวิธีอื่นเช่นการประเมินโดยใช้ Silhouette Analysis ดังงานวิจัยที่เคยมีมาก่อนหน้า ทั้งนี้ขึ้นกับความต้องการและความเหมาะสมกับผู้ใช้งาน

ในที่นี้ ผู้พัฒนาจะกล่าวถึงถึงข้อดี ข้อเสีย รวมถึงข้อจำกัดของระบบที่ได้จัดทำขึ้นด้วยวิธี Traditional RFM scoring ที่ได้ศึกษาและนำมาประยุกต์ใช้กับเว็บไซต์ มีดังต่อไปนี้

5.1.1 ข้อดีของระบบ RFM model

- 1) สามารถเลือกช่วงเวลาที่ต้องการทราบได้อย่างอย่างอิสระ
- 2) สามารถแบ่งกลุ่มประเภทของลูกค้าที่ต้องการได้อย่างอย่างอิสระ
- 3) สามารถส่งออกข้อมูลในรูปแบบไฟล์ได้
- 4) เว็บไซต์มีความสวยงามและง่ายต่อการใช้งาน

5.1.2 ข้อเสียและข้อจำกัดของระบบ RFM model

- 1) มีนามสกุลไฟล์ในส่วนการนำเข้าและส่งออกไฟล์ไม่หลากหลาย
- 2) ไม่สามารถเก็บผลลัพธ์จากไฟล์ข้อมูลก่อนหน้านี้ได้ หากไม่ได้ทำการส่งออก
- 3) ไม่มีการแสดงระดับคะแนนของความล่าช้า ความถี่ และปริมาณที่ใช้เงิน
- 4) ไม่สามารถตรวจสอบการแบ่งระดับคะแนนได้ในกรณีที่เกิดความซ้อนทับกันของช่วงระดับคะแนน
- 5) หากข้อมูลมาอยู่ในรูปแบบไม่ถูกต้อง จะไม่สามารถวิเคราะห์ได้ ซึ่งจำเป็นต้องมีการตรวจสอบไฟล์ข้อมูลก่อน

5.2 ข้อเสนอแนะ

ในที่นี้ จะกล่าวถึงข้อเสนอแนะที่มีต่อเว็บไซต์ RFM model ที่ได้จัดทำขึ้น

- 1) ควรสร้างเว็บไซต์ที่สามารถมีการประมวลผลที่รวดเร็วเพื่อรับรองข้อมูลที่มีปริมาณมหาศาล
- 2) ควรสร้างเว็บไซต์ที่ไม่จำกัดรูปแบบในการนำเข้าข้อมูล หรือสามารถระบุหัวข้อของคอลัมน์หรือแอตทริบิวต์ (Attribute)

- 3) ควรมีระบบตรวจสอบเมื่อมีการแบ่งกลุ่มประเภทลูกค้าตามระดับด้วยตนเอง เนื่องจากการทำงานนั้น จะทำงานเรียงจากบนลงล่าง และหากผู้ใช้งานจัดกลุ่มด้านบนที่มีช่วงระดับกว้างกว่าด้านล่างนั้น จะเกิดข้อผิดพลาดได้

บรรณานุกรม

- [1] Nattapon Muangtum. 2564. “Pareto Marketing จาก 80/20 สู่ RFM Model – Smart Customer Segmentation” [Online]. Available : <https://www.everydaymarketing.co/business-and-marketing-case-study/data/rfm-model-smart-customer-segmentation-from-data-pareto-marketing/>.
- [2] Chanokchon, Nalyn. 2565. “เทคนิคการใช้ RFM Analysis แบ่งกลุ่มลูกค้า เพื่อวางกลยุทธ์สื่อสารได้ตรงจุด” [Online]. Available : <https://predictive.co.th/blog/rfm-analysis/>.
- [3] BentoWeb. 2564. “แบ่งกลุ่มลูกค้าด้วย RFM คืออะไร? แล้วทำไมถึงมีความสำคัญ?” [Online]. Available : <https://help.bentoweb.com/th/article/rfm-75paf0/>.
- [4] HR NOTE.asia. 2565. “กฎ 80/20 ทำน้อยให้ได้มาก : 6 แนวทางที่ใช้ได้จริงในองค์กรเพื่อผลลัพธ์ที่ดีขึ้น” [Online]. Available : <https://th.hrnote.asia/orgdevelopment/8020rule-paretoprinciple-01182021/>.
- [5] ระบบจัดการเรียนการสอน มหาวิทยาลัยมหาสารคาม. 2565. “การวิเคราะห์ค่าสถิติเบื้องต้น” [Online]. Available : http://www.wbi.msu.ac.th/file/1349/doc_41218.658969907456.pdf.
- [6] Microsoft. 2022. “Excel functions” [Online]. Available : <https://support.microsoft.com/en-us/office/excel-functions-alphabetical-b3944572-255d-4efb-bb96-c6d90033e188>.
- [7] นรินาม. 2565. “วิธีคำนวณอันดับเปอร์เซ็นต์ของรายการใน Excel” [Online]. Available : <https://th.extendoffice.com/documents/excel/4794-excel-rank-percentile.html>.
- [8] เทพเอ็กเซล. 2563. “ค่าสถิติที่สำคัญ” [Online]. Available : <https://www.thepexcel.com/stats-01-descriptive-statistics/>.
- [9] Dave Bruns. 2022. “PERCENTRANK.INC Function” [Online]. Available : <https://exceljet.net/functions/percentrank.inc-function>.
- [10] P. Anitha, Malini M. Patil. 2022. “RFM model for customer purchase behavior using K-Means algorithm” [Online]. Available : <https://www.sciencedirect.com/science/article/pii/S1319157819309802>.

- [11] Microsoft. 2565. “ใช้ส่วนขยาย Visual Studio Code” [Online]. Available :
<https://learn.microsoft.com/th-th/power-apps/maker/portals/vs-code-extension>.
- [12] พิชิต วิจิตรบุญรักษ์. 2565. “HTML: ภาษาเขียนเว็บ” [Online]. Available :
https://www.bu.ac.th/knowledgecenter/executive_journal/july_sep_11/pdf/aw32.pdf.
- [13] MarcusCode. 2565. “แนะนำภาษา JavaScript” [Online]. Available :
<http://marcuscode.com/lang/javascript/introducing-to-javascript>.
- [14] เทพเอ็กเซล. 2565. “แบ่ง Segment ลูกค้าด้วย RFM Analysis” [Online]. Available :
<https://www.thepexcel.com/segment-rfm-analysis-excel/>.
- [15] เทพเอ็กเซล. 2565. “สารพัดวิธี Lookup หาค่าตัวสุดท้าย” [Online]. Available :
<https://www.thepexcel.com/lookup-last-value/>.

ภาคผนวก

โปรแกรมเว็บไซต์ RFM model

```
<!DOCTYPE html>
<html>

<head>
  <meta name="viewport" content="width=device-width,
initial-scale=1">
  <script type="text/javascript"
src="https://unpkg.com/xlsx@0.15.1/dist/xlsx.full.min.js"></
script>
  <script src="http://code.jquery.com/jquery-
1.10.2.js"></script>
  <script src="http://code.jquery.com/ui/1.11.2/jquery-
ui.js"></script>
  <link rel="stylesheet"
href="https://cdnjs.cloudflare.com/ajax/libs/font-
awesome/6.2.0/css/all.min.css">
  <link rel="stylesheet" href="styles.css">
</head>

<body>
  <h1 class="center">RFM - Model</h1>
  <div style="text-align: center; margin: 0 0 10px
0;"><span style="font-weight: 600; margin-right:
10px;">Select
      Excel File</span><input type="file"
id="excel_file" />
    </div>
    <div id="result_rfm" style="display: none; height:
auto;">
      <div id="chartdiv"></div>
      <button class="btn_primary"
onclick="download_csv_file()"> Download CSV </button>
      <div id="rfm_data"></div>
    </div>
    <button class="accordion"><i class="fa-solid fa-file-
excel"></i> Data Import (.xlsx)</button>
    <div class="panel">
      <div id="excel_data"></div>
    </div>
    <button class="accordion"><i class="fa-solid fa-
gear"></i> Setting</button>
    <div class="panel">
      <form action="test02.html" method="post"
id="form_setting">
```

```

        <button class="btn_primary" type="submit"
id="form_save">บันทึก</button>
        <table class="table table-striped table-
bordered" id="table_time">
            <tr>
                <th colspan="4">Display Graph</th>
            </tr>
            <tr>
                <td>
                    <p>Select</p>
                </td>
                <td colspan="3" style="padding: 10px;">
                    <input type="radio" id="type1"
name="typeshow" value="only">
                    <label for="data1">กลุ่มที่อยู่ในช่วงเวลา
</label><br><br>
                    <input type="radio" id="type2"
name="typeshow" value="all" checked>
                    <label for="data2">ข้อมูลทั้งหมด
</label><br>
                </td>
            </tr>
            <tr>
                <th colspan="4">Time</th>
            </tr>
            <tr>
                <td>
                    <p>Start Date</p>
                </td>
                <td colspan="3" style="padding: 0
10px;"><input type="date" id="start_select"
name="start_select">
                </td>
            </tr>
            <tr>
                <td>
                    <p>End Date</p>
                </td>
                <td colspan="3" style="padding: 0
10px;"><input type="date" id="end_select"
name="end_select"></td>
            </tr>
        </table>
    </form>
    <form action="test02.html" method="post" id="form">
        <table class="table table-striped table-
bordered" id="table_setrmf">
            <tr>
                <th rowspan="2">Segment</th>

```

```

        <th colspan="2">Recency (R)</th>
        <th colspan="2">Frequency (F)</th>
        <th colspan="2">Monetary (M)</th>
        <th rowspan="2"></th>
    </tr>
    <tr>
        <th>Min</th>
        <th>Max</th>
        <th>Min</th>
        <th>Max</th>
        <th>Min</th>
        <th>Max</th>
    </tr>
    <tr>
        <td><input type="text" id="segment"
required></td>
        <td class="center"><input type="number"
min="1" max="5" class="input_rfm" id="Rmin" required></td>
        <td class="center"><input type="number"
min="1" max="5" class="input_rfm" id="Rmax" required></td>
        <td class="center"><input type="number"
min="1" max="5" class="input_rfm" id="Fmin" required></td>
        <td class="center"><input type="number"
min="1" max="5" class="input_rfm" id="Fmax" required></td>
        <td class="center"><input type="number"
min="1" max="5" class="input_rfm" id="Mmin" required></td>
        <td class="center"><input type="number"
min="1" max="5" class="input_rfm" id="Mmax" required></td>
        <td class="center"><input
class="btn_secondary" type="submit" value="Add Segment"
class="record-submit"></td>
    </tr>
</table>
</form>
</div>
</body>

<!-- Resources -->
<script
src="https://cdn.amcharts.com/lib/4/core.js"></script>
<script
src="https://cdn.amcharts.com/lib/4/charts.js"></script>
<script
src="https://cdn.amcharts.com/lib/4/themes/animated.js"></sc
ript>

<script>
    var csvFileData = new Array();
    var model = new Array();
    var form = document.getElementById("form");

```

```

var table = document.getElementById("table_setrmf");
var acc = document.getElementsByClassName("accordion");
for (let i = 0; i < acc.length; i++) {
    acc[i].addEventListener("click", function () {
        this.classList.toggle("active");
        var panel = this.nextElementSibling;
        if (panel.style.maxHeight) {
            panel.style.maxHeight = null;
        } else {
            panel.style.maxHeight = panel.scrollHeight +
"px";
        }
    });
}

const one_day = 1000 * 60 * 60 * 24;
let start = new Date();
start.setFullYear(start.getFullYear() - 1);
document.getElementById('start_select').value =
start.toISOString().substring(0, 10);
let end = new Date();
document.getElementById('end_select').value =
end.toISOString().substring(0, 10);
//console.log("INPUT: ", start, " - ", end);
var dataofuser = new Array();

function RFM_show() {
    var dataFields = new Array();
    csvFileData = new Array();
    let rfm_output = '';
    rfm_output += '<table id="rfm_output" class="table
table-striped table-bordered">';
    rfm_output += '<tr><th rowspan="2">Username</th><th
colspan="3">RFM - model</th><th
rowspan="2">Result</th></tr>';
    rfm_output +=
'<tr><th>R</th><th>F</th><th>M</th></tr>';

    //summary
    let allData = { allRecency: [], allFrequency: [],
allMonetary: [] }
    for (let i = 0; i < dataofuser.length; i++) {
        var a = PRIMARYanalysis(dataofuser[i].name,
dataofuser[i].transactions);
        if (a.Lastdate != -1) {
            allData.allRecency.push(a.Lastdate);
            allData.allFrequency.push(a.Freq);
            allData.allMonetary.push(a.Total);
        }
        dataFields.push(a);
    }
}

```



```

    }
    let rank = { R_rank:
PERCENTRANK(allData.allRecency), F_rank:
PERCENTRANK(allData.allFrequency), M_rank:
PERCENTRANK(allData.allMonetary) };
    //select type show
    let typeshow =
document.getElementsByName('typeshow');
    let selectshow = '';
    for (i = 0; i < typeshow.length; i++) {
        if (typeshow[i].checked)
            selectshow = typeshow[i].value;
    }

    //display
    //return { Name: name, Lastdate: -1.0, Total: -1.0,
Freq: -1.0 };
    for (let i = 0; i < dataFields.length; i++) {
        if (dataFields[i].Lastdate == -1 && selectshow
== "all") {
            rfm_output += '<tr>';
            rfm_output += "<td>" + dataFields[i].Name +
"</td>";

            rfm_output += "<td>0 %</td>";
            rfm_output += "<td>0 %</td>";
            rfm_output += "<td>0 %</td>";
            rfm_output += "<td>ไม่ได้อยู่ในช่วงเวลาที่เลือก</td>";
            rfm_output += '</tr>';

            csvFileData.push([dataFields[i].Name, "ไม่ได้อยู่ใน
ช่วงเวลาที่เลือก"]);
        }
        else if (dataFields[i].Lastdate == -1 &&
selectshow == "only") {
            continue;
        }
        else {
            var tempUser = rankRFM(dataFields[i], rank,
dataFields.length);
            rfm_output += '<tr>';
            rfm_output += "<td>" + dataFields[i].Name +
"</td>";

            rfm_output += "<td>" + tempUser.Recency + "
%</td>";

            rfm_output += "<td>" + tempUser.Frequency + "
" %</td>";

            rfm_output += "<td>" + tempUser.Monetary + "
%</td>";

            var tempRFM =

```

```

convertRFM(Math.ceil(tempUser.Recency / 20),
Math.ceil(tempUser.Frequency / 20),
Math.ceil(tempUser.Monetary / 20));
    rfm_output += "<td>" + tempRFM + "</td>";
    rfm_output += '</tr>';
    console.log(dataofuser[i].name, " : ",
tempUser, " >> ", tempRFM);
    csvFileData.push([dataFields[i].Name,
tempRFM]);
    }

    }
    //console.log("csvFile",csvFileData);
    rfm_output += '</table>';
    document.getElementById("rfm_data").innerHTML =
rfm_output;
    chart_show();
}

function PRIMARYanalysis(name, data) {
    var Result = { Name: name, Lastdate: 0.0, Total:
0.0, Freq: 0.0 };

    //Time
    data.sort(function (a, b) { return new Date(a[0]) -
new Date(b[0]); });
    var start_select = new
Date(document.getElementById('start_select').value);
    start_select.setHours(0, 0, 0, 0);
    var end_select = new
Date(document.getElementById('end_select').value);
    end_select.setHours(0, 0, 0, 0);
    let start_datareal = data[0][0];
    let end_datareal = data[data.length - 1][0];

    if (((start_datareal < start_select) &&
(end_datareal < start_select)) || ((start_datareal >
end_select) && (end_datareal > end_select))) {
        return { Name: name, Lastdate: -1.0, Total: -
1.0, Freq: -1.0 };
    }

    //Recency_percent + Frequency_percent +
Monetary_percent
    for (let i = 0; i < data.length; i++) {
        if (start_select <= data[i][0] && data[i][0] <=
end_select) {
            Result.Lastdate = data[i][0];
            Result.Total += data[i][1];
            Result.Freq++;
        }
    }
}

```

```

    }
    }
    Result.Total = Result.Total.toFixed(2);
    return Result;
}

function rankRFM(datafields, rank, len) {
    console.log("datafields", datafields, "rank", rank,
"len", len)
    return {
        Recency: PERCENTRANK_INC(datafields.Lastdate,
rank.R_rank, len),
        Frequency: PERCENTRANK_INC(datafields.Freq,
rank.F_rank, len),
        Monetary: PERCENTRANK_INC(datafields.Total,
rank.M_rank, len)
    };
}

function sendvalue(r, c) {
    return table.rows[r].cells[c].innerHTML;
    //return document.getElementById(input).value;
}

function segment_findnumof() {
    var result = document.getElementById("rfm_output");
    var chart_data = new Array();
    var segment_name = new Array();

    for (let i = 2; i < result.rows.length; i++) {
        if
(segment_name.indexOf(result.rows[i].cells[4].innerHTML) ==
-1) {
            //chart_data.findIndex(x => x.segment ===
tempRFM) == -1

            segment_name.push(result.rows[i].cells[4].innerHTML);
            chart_data.push({ segment:
result.rows[i].cells[4].innerHTML, numberof: 1 });
        }
        else {
            for (let j = 0; j < segment_name.length;
j++) {
                if (chart_data[j].segment ==
result.rows[i].cells[4].innerHTML) {
                    chart_data[j].numberof++;
                }
            }
        }
    }
    return chart_data;
}

```

```

}
function convertRFM(R, F, M) {
    R = R == 0 ? 1 : R;
    F = F == 0 ? 1 : F;
    M = M == 0 ? 1 : M;

    console.log("RFM:", R, F, M);
    for (let i = 3; i < table.rows.length; i++) {
        //console.log(table.rows[i].cells[1].innerHTML);
        if ((R >= sendvalue(i, 1) && R <= sendvalue(i,
2)) && (F >= sendvalue(i, 3) && F <= sendvalue(i, 4)) && (M
>= sendvalue(i, 5) && M <= sendvalue(i, 6))) {
            return table.rows[i].cells[0].innerHTML;
        }
    }
    return 'ไม่จัดอยู่ในกลุ่มใด';
}

//Statistics
function PERCENTRANK(data) {
    var Setrank = {};
    data.sort(function (a, b) { return a - b });

    for (let i = 0; i < data.length; i++) {
        const item = data[i];

        if (!Setrank[item]) {
            Setrank[item] = i;
        }
    }
    return Setrank;
}

function PERCENTRANK_INC(val, Setrank, len) {
    return (Setrank[val] / (len - 1) * 100).toFixed(2);
}

function chart_show() {
    am4core.ready(function () {

        // Themes begin
        am4core.useTheme(am4themes_animated);
        // Themes end

        // Create chart instance
        var chart = am4core.create("chartdiv",
am4charts.PieChart);

        // Add data
        chart.data = segment_findnumof();
        console.log("CHART : ", chart.data);
    });
}

```

```

        // Add and configure Series
        var pieSeries = chart.series.push(new
am4charts.PieSeries());
        pieSeries.dataFields.value = "numberof";
        pieSeries.dataFields.category = "segment";
        pieSeries.slices.template.stroke =
am4core.color("#fff");
        pieSeries.slices.template.strokeWidth = 2;
        pieSeries.slices.template.strokeOpacity = 1;

        // This creates initial animation
        pieSeries.hiddenState.properties.opacity = 1;
        pieSeries.hiddenState.properties.endAngle = -90;
        pieSeries.hiddenState.properties.startAngle = -
90;

    }); // end am4core.ready()
}

//add event listener to form
form.addEventListener("submit", addItem);
function addItem(e) {
    e.preventDefault();

    var segment =
document.getElementById("segment").value;
    var Rmin = document.getElementById("Rmin").value;
    var Rmax = document.getElementById("Rmax").value;
    var Fmin = document.getElementById("Fmin").value;
    var Fmax = document.getElementById("Fmax").value;
    var Mmin = document.getElementById("Mmin").value;
    var Mmax = document.getElementById("Mmax").value;

    let m = Rmin + "," + Rmax + "," + Fmin + "," + Fmax
+ "," + Mmin + "," + Mmax;

    if ((Rmin > Rmax) || (Fmin > Fmax) || (Mmin > Mmax))
{
        alert("อาจมีข้อผิดพลาด โปรดตรวจสอบการจัดระดับ RFM อีกครั้ง");
    }
    else if (model.indexOf(m) != -1) {
        alert("อาจมีข้อผิดพลาด โปรดตรวจสอบการจัดระดับ RFM อีกครั้ง");
    }
    else {
        //create rows
        var row = table.insertRow(3);
        //create row cells
        var segmentCell = row.insertCell(0);
        segmentCell.innerHTML = segment;
    }
}

```

```

        var RminCell = row.insertCell(1);
        RminCell.innerHTML = Rmin;

        var RmaxCell = row.insertCell(2);
        RmaxCell.innerHTML = Rmax;

        var FminCell = row.insertCell(3);
        FminCell.innerHTML = Fmin;

        var FmaxCell = row.insertCell(4);
        FmaxCell.innerHTML = Fmax;

        var MminCell = row.insertCell(5);
        MminCell.innerHTML = Mmin;

        var MmaxCell = row.insertCell(6);
        MmaxCell.innerHTML = Mmax;

        model.push(m);
        //console.log("ADD", model);
        var DeleteCell = row.insertCell(7);
        DeleteCell.innerHTML = "<button
class='btn_secondary'
onclick='removeRow(this)'>Remove</button>";
        DeleteCell.style.textAlign = "center";
    }
}

function removeRow(e) {
    table.deleteRow(e.parentNode.parentNode.rowIndex);
    model.splice((e.parentNode.parentNode.rowIndex - 3),
1);

    //console.log("REMOVE", model);
    // button -> td -> tr.
}

const excel_file =
document.getElementById('excel_file');
excel_file.addEventListener('change', (event) => {
    if (!['application/vnd.openxmlformats-
officedocument.spreadsheetml.sheet', 'application/vnd.ms-
excel'].includes(event.target.files[0].type)) {
        document.getElementById('excel_data').innerHTML
= '<div class="alert alert-danger">Only .xlsx or .xls file
format are allowed</div>';
        excel_file.value = '';
        return false;
    }
}

```

```

var reader = new FileReader();
reader.readAsArrayBuffer(event.target.files[0]);
reader.onload = function (event) {
    var data = new Uint8Array(reader.result);
    var work_book = XLSX.read(data, { type: 'array'
});

    var sheet_name = work_book.SheetNames;
    var sheet_data =
XLSX.utils.sheet_to_json(work_book.Sheets[sheet_name[0]], {
header: 1 });
    if (sheet_data.length > 0) {
        var table_output = '<table class="table
table-striped table-bordered">';
        let user = new Array();
        dataofuser = new Array();
        for (var row = 0; row < sheet_data.length;
row++) {
            table_output += '<tr>';
            for (var cell = 0; cell <
sheet_data[row].length; cell++) {
                if (row == 0) {
                    table_output += '<th>' +
sheet_data[row][cell] + '</th>';
                }
                else {
                    if (cell == 1) {
                        var date = new
Date(Math.round((sheet_data[row][cell] - (25568 + 1)) *
86400 * 1000));
                        date.setHours(0, 0, 0, 0);
                        //ไม่ใส่จะต่างกัน 25,200,000
                        table_output += '<td>' +
date + '</td>';
                    }
                    else {
                        table_output += '<td>' +
sheet_data[row][cell] + '</td>';
                    }
                }
            }

            if (row != 0 && typeof
sheet_data[row][0] !== 'undefined') {
                if (user.indexOf(sheet_data[row][0])
== -1) {
                    user.push(sheet_data[row][0]);
                    dataofuser.push({ name:
sheet_data[row][0], transactions: [[date,
sheet_data[row][2]]] });
                }
            }
        }
    }
}

```

```

        else {
            for (let i = 0; i < user.length;
i++) {
                if (dataofuser[i].name ==
sheet_data[row][0]) {
                    dataofuser[i].transactions.push([date, sheet_data[row][2]]);
                }
            }
            table_output += '</tr>';
        }
        table_output += '</table>';

document.getElementById('excel_data').innerHTML =
table_output;
        excel_file.value = '';

document.getElementById('result_rfm').style.display =
"initial";

        //dataofuser.sort(function (a, b) { return new
Date(a.transactions) - new Date(b.transactions); });
        console.log(dataofuser);

        RFM_show();
    }
});

//https://www.encodedna.com/javascript/dynamically-add-
remove-rows-to-html-table-using-javascript-and-save-data.htm

const formset = document.querySelector("#form_setting");
formset.addEventListener("submit", function (event) {
    alert("Setting Change!");
    // stop form submission
    event.preventDefault();

    RFM_show();
});

function download_csv_file() {
    //REF.: https://www.javatpoint.com/javascript-
create-and-download-csv-file

    //define the heading for each row of the data

```



```
var csv = 'Uername, Segmentresult\n';
//merge the data with CSV
if (csvFileData != null) {
    csvFileData.forEach(function (row) {
        csv += row.join(',');
        csv += "\n";
    });
}

var hiddenElement = document.createElement('a');
hiddenElement.href = 'data:text/csv;charset=utf-8,%EF%BB%BF' + encodeURIComponent(csv);
hiddenElement.click();
}

</script>

</html>
```