การจัดกลุ่มร้านค้าเบเกอรี่ กรณีศึกษาบริษัทแห่งหนึ่ง

CLSTERING OF BAKERY SHOPS :
A CASE STUDY OF A COMPANY

อารียา วังสิโก

สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาตรบัณฑิต (สถิติประยุกต์)
ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2566

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

CLUSTERING OF BAKERY SHOPS :
A CASE STUDY OF A COMPANY

AREEYA WANGSIGO

A COOPERATIVE EDUCATION SUBMITTED IN PARTIAL
FULFILLMENT OF THE REQUIREMENT FOR
THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS)
DEPARTMENT OF STATISTICS, FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
ACADEMIC YEAR 2023

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา	การจัดกลุ่มร้านค้าเบเกอรี่ กรณีศึกษาบริษัทแห่งหนึ่ง Clustering of Bakery Shops : A Case Study of a Company
ชื่อนักศึกษา	นางสาว อารียา วังสิโก รหัสนักศึกษา 63050694
ปริญญา	วิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชา	สถิติ
คณะ	วิทยาศาสตร์
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกลาเจ้าคุณทหารลาดกระบัง
ปีการศึกษา	2566
อาจารย์ที่ปรึกษา	ดร.สกุณา ศรีอโนมัย

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตร์บัณฑิต (สถิติประยุกต์) ประจำปีการศึกษา 2566

คณะกรรมการสอบ	ลายมือชื่อ
ผศ.ดร. สมศรี บัณฑิตวิไล ประธานกรรมการ	
นางสาวสมิตา ยงไพศาล กรรมการ	
ดร.สกุณา ศรีอโนมัย กรรมการและที่ปรึกษา	

ลิขสิทธิ์ของคณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา การจัดกลุ่มร้านค้าเบเกอรี่ กรณีศึกษาบริษัทแห่งหนึ่ง
Clustering of Bakery Shops : A Case Study of a Company
ชื่อนักศึกษา นางสาว อารียา วังสิโก รหัสนักศึกษา 63050694
ปริญญา วิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชา สถิติ
คณะ วิทยาศาสตร์
มหาวิทยาลัย สถาบันเทคโนโลยีพระจอมเกลาเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2566
อาจารย์ที่ปรึกษา ดร.สกุณา ศรีอโนมัย

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อที่จัดกลุ่มร้านเบเกอรี่ในเครือบริษัทแห่งหนึ่งและตรวจสอบว่ากลุ่มร้านค้าที่จำแนกมามีความเหมือนหรือแตกต่างกันซึ่งข้อมูลมีทั้งหมด 17 สดมภ์ 157 รายการ เนื่องจากผู้วิจัยต้องการจัดกลุ่มร้านค้าที่มีลักษณะคล้ายคลึงกันไว้ด้วยกันจึงเลือกใช้การจัดกลุ่มโดยเทคนิคเคมีน (K-Mean Clustering) และใช้วิธีจุดหักข้อศอก (Elbow Method) ในการหาจำนวนกลุ่มที่เหมาะสม ดำเนินการวิจัยด้วยภาษาโปรแกรมไพทอน (Python) และนำลักษณะของข้อมูลแต่ละกลุ่มมาเปรียบเทียบว่าเหมือนหรือแตกต่างกัน เพื่อให้งานวิจัยมีความน่าเชื่อถือจึงใช้ค่าสถิติในการวัดผลการจัดกลุ่มที่ได้ โดยการนำกลุ่มที่จำแนกไว้มาทดสอบการแจกแจงปรกติและนำมาทดสอบความแตกต่างระหว่างกลุ่มโดยสถิติของคลัสคาล-วอลลิส (Kruskal-Wallis Test) ด้วยโปรแกรม IBM SPSS Statistics 26 ผลการวิจัยพบว่าจำนวนกลุ่มที่เหมาะสมในการจัดครั้งนี้คือ 3 กลุ่ม แต่ละกลุ่มที่ได้มามีลักษณะที่แตกต่างกัน โดยสามารถระบุกลุ่มทั้ง 3 จากและกำหนดชื่อกลุ่มได้ดังนี้ คือกลุ่มร้านค้าที่มีคะแนนการประเมินต่ำ กลุ่มร้านค้าที่มีคะแนนการประเมินปานกลาง และกลุ่มร้านค้าที่มีคะแนนการประเมินสูงตามลำดับ ผลลัพธ์ที่ได้สามารถเป็นแนวทางไปใช้ในการพัฒนากลยุทธ์ธุรกิจของบริษัทกรณีศึกษาได้

คำสำคัญ : K-Means Clustering, จุดหักข้อศอก, กลยุทธ์ธุรกิจ, การจัดกลุ่มร้านเบเกอรี่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Clustering of Bakery Shops : A Case Study of a Company
Student	Miss Areeya Wangsigo Student ID 63050694
Degree	Bachelor of Science (Applied Statistics)
Department	Statistics
School	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Academic Year	2023
Advisor	Dr. Sakuna Sriarnomai

Abstract

This research aims to cluster bakery shops of a case study company and examine similarities and differences between the classified groups. The data consists of 17 columns and 157 entries. Since the researcher wants to group shops with similar characteristics together, K-Means Clustering is used for clustering, and the Elbow Method is used to determine the appropriate number of clusters. The research is conducted using the Python programming language, and the characteristics of each group are compared to determine similarities and differences. To ensure the reliability of the research, statistical measures are used to evaluate the clustering results. The classified groups are tested for normality and for differences between groups using the Kruskal-Wallis Test with IBM SPSS Statistics 26. The results show that the optimal number of clusters for this analysis is 3. Each of the resulting groups has different characteristics. The three groups can be identified and named as follows : Low evaluation score, Medium evaluation score, and High evaluation score. The results can be used to develop business strategies for the case study company.

Keywords: K-Means Clustering, Elbow Method, Business Strategy, Bakery Store Clustering

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

การทำสหกิจศึกษาฉบับนี้ สามารถดำเนินการสำเร็จลุล่วงไปได้ด้วยดี เนื่องจากได้รับความอนุเคราะห์จาก ดร.สกุณา ศรีอโนมัย อาจารย์ที่ปรึกษาสหกิจที่ให้ความรู้ คำปรึกษา คำแนะนำ เสนอแนวคิด ตลอดจนติดตามผลงานทุกขั้นตอน จนกระทั่งการดำเนินงานสหกิจศึกษานี้เสร็จสมบูรณ์จึงขอขอบพระคุณด้วยความเคารพเป็นอย่างสูงไว้ ณ ที่นี้

ขอขอบพระคุณ ผศ.ดร. สมศรี บัณฑิตวิไล ที่ให้เกียรติเป็นประธานกรรมการสอบสหกิจ ซึ่งได้ให้คำแนะนำแนวคิด วิธีการและข้อมูลเพิ่มเติมที่เป็นประโยชน์กับโครงงานสหกิจศึกษาฉบับนี้ทั้งหมด

ขอขอบพระคุณ สมิตา ยงไพศาล ที่ให้การสนับสนุนในด้านการศึกษา และการทำสหกิจศึกษาอีกทั้งสละเวลาในการตรวจสอบและให้คำปรึกษาตลอดการทำสหกิจศึกษา

ขอขอบพระคุณคณาจารย์ประจำภาควิชาสถิติ บุคลากรภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่ได้ประสิทธิ์ประสาทวิชาความรู้และได้ทำการเรียนการสอน ให้คำแนะนำรวมถึงการให้ความช่วยเหลือแก่ผู้วิจัยตลอดมา

ขอมอบความกตัญญูกตเวทิตาคุณ แด่บิดา มารดา ที่สนับสนุนและให้กำลังใจ ส่งเสริมและสนับสนุนในทุกๆด้าน รวมถึงเพื่อนๆและพี่ที่ร่วมงาน ที่ให้คำปรึกษา คำแนะนำ และคอยช่วยเหลือในการทำสหกิจศึกษาครั้งนี้และสุดท้ายนี้ผู้วิจัยขอขอบคุณบริษัทกรณีศึกษาที่อณุญาติให้นำข้อมูลมาใช้ในการทำสหกิจศึกษาฉบับนี้

อารียา วังสิโก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ฉ
สารบัญรูป	ช
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญ	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
1.5 นิยามศัพท์	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 ทฤษฎีเกี่ยวการวิเคราะห์การแบ่งกลุ่ม	4
2.1.1 ประเภทของการวิเคราะห์การแบ่งกลุ่ม	4
2.1.2 การคำนวณระยะห่างข้อมูล	6
2.1.3 การกำหนดจำนวนกลุ่มที่เหมาะสม	7
2.2 ทฤษฎีทางสถิติ	8
2.2.1 สถิติพรรณนา	8
2.2.2 สถิติอนุมาน	9
2.3 ทฤษฎีเกี่ยวกับการเรียนรู้ของเครื่อง	19
2.4 ภาษาโปรแกรมที่เกี่ยวข้อง	20
2.5 งานวิจัยที่เกี่ยวข้อง	21
บทที่ 3 วิธีการดำเนินงานวิจัย	23
3.1 ขั้นตอนการดำเนินงานวิจัย	23
3.2 การจัดเตรียมข้อมูล	24
3.2.1 การทำความสะอาดข้อมูล	25
3.2.2 การแปลงข้อมูล	26
3.2.3 การคัดเลือกข้อมูล	27
3.3 เครื่องมือที่ใช้ในการดำเนินงานวิจัย	27
3.3.1 ซอฟต์แวร์ที่ใช้ในงานวิจัย	27
3.3.2 ชุดคำสั่งไลบรารีที่ใช้ในงานวิจัย	27
3.4 การจัดกลุ่มด้วยเทคนิค K-Means	28
3.4.1 การวิเคราะห์จำนวนกลุ่มที่เหมาะสม	28
3.4.2 การจัดกลุ่มด้วยเทคนิค K-Means	29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
3.5 การวิเคราะห์ข้อมูล	30
3.5.1 สถิติเชิงพรรณนา	30
3.5.2 สถิติเชิงอนุมาน	30
บทที่ 4 ผลการวิจัย	33
4.1 ผลการวิเคราะห์การจัดกลุ่ม	33
4.2 ผลการวิเคราะห์ข้อมูลด้วยสถิติพรรณนา	34
4.2.1 ผลการวิเคราะห์ด้วยสถิติพรรณนาเปรียบเทียบทั้ง 3 กลุ่ม	34
4.2.2 ผลการวิเคราะห์ลักษณะร้านค้า	37
4.3 ผลการวิเคราะห์ข้อมูลด้วยสถิติอนุมาน	49
4.4 การอภิปรายผล	50
บทที่ 5 สรุปการวิจัยและข้อเสนอแนะ	51
5.1 สรุปผลงานวิจัย	51
5.2 ข้อจำกัด	52
5.3 ข้อเสนอแนะ	52
5.3.1 ข้อเสนอแนะในงานวิจัยครั้งถัดไป	52
5.3.2 ข้อเสนอแนะข้อเสนอแนะเพื่อนำไปใช้	53
เอกสารอ้างอิง	54
ภาคผนวก ก.	57
ภาคผนวก ข.	82

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่		หน้า
2.1	การทดสอบค่าเฉลี่ยของประชากร 2 ประชากรเมื่ออิสระต่อกัน	12
2.2	การทดสอบค่าเฉลี่ยของประชากร 2 ประชากรเมื่อไม่เป็นอิสระต่อกัน	13
2.3	ตัวอย่างตารางข้อมูลสำหรับการวิเคราะห์ความแปรปรวนทางเดียว	13
2.4	ตารางวิเคราะห์ความแปรปรวน	14
2.5	ตัวอย่างแสดงข้อมูลสำหรับการวิเคราะห์ความแปรปรวนสองทาง	15
2.6	ตารางวิเคราะห์ความแปรปรวนสองทาง	16
3.1	ตัวแปรและความหมายของข้อมูล	24
3.2	ตัวอย่างและรายละเอียดข้อมูล	25
3.3	ชุดคำสั่งสำหรับหาข้อมูลสูญหาย	25
3.4	ชุดคำสั่งสำหรับแปลงข้อมูลให้เป็นมาตรฐานเดียวกัน	26
3.5	การคัดเลือกข้อมูลในการจัดกลุ่ม	27
3.6	คำสั่งไลบรารี่ที่ใช้ในงานวิจัย	28
3.7	ชุดคำสั่งของ Elbow Method	28
3.8	ชุดคำสั่งการทำ K-Mean Clustering	29
4.1	ตารางแสดงลำดับค่าเฉลี่ยระหว่างกลุ่มร้านค้า	50
4.2	ตารางผลลัพธ์การทดสอบความแตกต่างระหว่างกลุ่ม	50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ

รูปที่		หน้า
2.1	รูปแสดงตัวอย่างการแบ่งกลุ่มโดยเทคนิค K-Means	5
2.2	รูปแสดงตัวอย่างการแบ่งกลุ่มโดยเทคนิค Hierarchical	6
2.3	รูปแสดงตัวอย่างจุดที่เหมาะสมของจำนวนกลุ่มโดยวิธี Elbow Method	7
3.1	ขั้นตอนการดำเนินงานวิจัย	23
3.2	รูปแสดงการค้นหาข้อมูลสูญหาย	26
3.3	รูปแสดง Elbow Method หา K ที่เหมาะสม	29
3.4	การวิเคราะห์การแจกแจงปรกติด้วยโปรแกรม SPSS	30
3.5	ผลการแจกแจงปรกติ	31
3.6	หน้าต่างการเลือกข้อมูลที่ใช้ทดสอบความแตกต่าง	31
4.1	แสดงผลลัพธ์การจัดกลุ่มร้านค้า	33
4.2	สัดส่วนของคะแนนการจัดร้านค้า	34
4.3	สัดส่วนของคะแนนการประเมินตนเอง	34
4.4	สัดส่วนของคะแนนการประเมินจากผู้จัดการร้านค้า	35
4.5	สัดส่วนของผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านค้า	35
4.6	สัดส่วนของคะแนนความน่าเข้ามาซื้อของ	36
4.7	สัดส่วนของตัวแปรคะแนนที่ตั้งร้านค้าทั้งหมด	36
4.8	ผลการวิเคราะห์ลักษณะที่ตั้งร้านค้าของร้านค้ากลุ่มที่ 1	37
4.9	ผลการวิเคราะห์พื้นที่ตั้งร้านค้าของร้านค้ากลุ่มที่ 1	38
4.10	ผลการวิเคราะห์ประเภทร้านค้าของร้านค้ากลุ่มที่ 1	38
4.11	ผลการวิเคราะห์เปอร์เซ็นต์ที่ขายได้ของเบเกอรี่แต่ละชนิด ร้านค้ากลุ่มที่ 1	39
4.12	ผลการวิเคราะห์ระดับร้านค้า ในร้านค้ากลุ่มที่ 1	39
4.13	ผลการวิเคราะห์การผลิตสินค้าเองของร้านค้ากลุ่มที่ 1	40
4.14	ผลการวิเคราะห์ลักษณะที่ตั้งร้านค้าของร้านค้ากลุ่มที่ 2	40
4.15	ผลการวิเคราะห์พื้นที่ตั้งร้านค้าของร้านค้ากลุ่มที่ 2	41
4.16	ผลการวิเคราะห์ประเภทร้านค้าของร้านค้ากลุ่มที่ 2	41
4.17	ผลการวิเคราะห์เปอร์เซ็นต์ที่ขายได้ของเบเกอรี่แต่ละชนิด ร้านค้ากลุ่มที่ 2	42
4.18	ผลการวิเคราะห์ระดับร้านค้า ในร้านค้ากลุ่มที่ 2	42
4.19	ผลการวิเคราะห์การผลิตสินค้าเองของร้านค้ากลุ่มที่ 2	43
4.20	ผลการวิเคราะห์ลักษณะที่ตั้งร้านค้าของร้านค้ากลุ่มที่ 3	43
4.21	ผลการวิเคราะห์พื้นที่ตั้งร้านค้าของร้านค้ากลุ่มที่ 3	44
4.22	ผลการวิเคราะห์ประเภทร้านค้าของร้านค้ากลุ่มที่ 3	44
4.23	ผลการวิเคราะห์เปอร์เซ็นต์ที่ขายได้ของเบเกอรี่แต่ละชนิด ร้านค้ากลุ่มที่ 3	45
4.24	ผลการวิเคราะห์ระดับร้านค้า ในร้านค้ากลุ่มที่ 3	45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ (ต่อ)

รูปที่		หน้า
4.25	ผลการวิเคราะห์การผลิตสินค้าเองของร้านค้ากลุ่มที่ 3.....	46
4.26	ผลการหาจำนวนกลุ่มที่เหมาะสม.....	46
4.27	ผลการจัดกลุ่มเพิ่มเติมจากร้านค้ากลุ่มที่ 1	47
4.28	ผลการวิเคราะห์ลักษณะร้านค้ากลุ่มที่ A.....	47
4.29	ผลการวิเคราะห์ลักษณะร้านค้ากลุ่มที่ B.....	48
4.30	ผลการวิเคราะห์ลักษณะร้านค้ากลุ่มที่ C.....	59

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

การสำรวจของหน่วยงานจัดเก็บและการวิเคราะห์ข้อมูลชั้นนำของโลก (Euromonitor International) เปิดเผยยอดขายเบเกอรี่ (Bakery) ของประเทศไทยว่ามีการขยายตัวขึ้นเป็น 38.8 พันล้านบาทในปี 2564 คิดเป็น 6% และคาดการณ์ว่ามีแนวโน้มที่จะเติบโตขึ้นอย่างต่อเนื่องเป็น 54.4 พันล้านบาทในปี 2567 และคาดว่าจะเติบโตมากขึ้นไปจนถึงปี 2569 นอกจากนี้ประเทศไทยยังถูกคาดการณ์ว่าจะเป็นที่ตั้งเชิงกลยุทธ์ในการขายสินค้าเบเกอรี่ไปทั่วตลาดเอเชีย (Euromonitor, 2021)

ผลิตภัณฑ์เบเกอรี่ เป็นผลิตภัณฑ์ที่สามารถบริโภคได้ทุกสถานที่และทุกโอกาสพฤติกรรมการบริโภคเบเกอรี่เปลี่ยนไปจากอดีตโดยเฉพาะในกลุ่มวัยทำงาน ผลิตภัณฑ์เบเกอรี่กลายเป็นปัจจัยหลักในการใช้ชีวิตของผู้คนที่อยู่ในเมืองที่ต้องใช้เวลาส่วนใหญ่กับการเดินทางไปทำงานในตอนเช้าที่เร่งรีบและผู้คนในเมืองต้องการความสะดวกสะดวกสบายจากอาหารพร้อมทาน ส่งผลให้ธุรกิจเบเกอรี่มีการเติบโตขึ้นอย่างต่อเนื่อง หากพิจารณาโอกาสการเติบโตของธุรกิจเบเกอรี่ในประเทศไทย โดยพิจารณาจากการบริโภคขนมปังของคนไทยยังมีเพียง 2-3 กิโลกรัมต่อคนต่อปี ซึ่งถือว่ายังน้อยเมื่อเทียบกับต่างประเทศจะเห็นได้ว่าธุรกิจเบเกอรี่ในประเทศไทยยังไม่ถึงจุดอิ่มตัว และยังมีโอกาสเติบโตได้มากในอนาคต (ศูนย์อัจฉริยะเพื่ออุตสาหกรรมอาหาร, 2559)

จากข้อมูลดังกล่าวถือเป็นการตอกย้ำว่าธุรกิจเบเกอรี่มีโอกาสเติบโตสูงในอนาคต อีกทั้งการแข่งขันของธุรกิจเบเกอรี่ในประเทศไทยมีการแข่งขันที่ค่อนข้างสูงถึงแม้จะมีลูกค้าจำนวนมากก็ตามแต่ก็ยังมีผู้ประกอบกิจการระดับสูงจำนวนมากเช่นกันซึ่งทำให้ลูกค้ามีทางเลือกมากขึ้น นอกจากนี้ยังมีผลิตภัณฑ์เบเกอรี่จากต่างประเทศที่เข้ามาตีตลาดในประเทศไทยอย่างต่อเนื่อง ในกลุ่มประเทศใกล้เคียงกันอย่างเช่นมาเลเซียเป็นประเทศที่มีศักยภาพในเชิงการแข่งขันมากที่สุดเพราะมีจุดแข็งในเรื่องต้นทุนที่ต่ำกว่าทำให้ขายได้ในราคาที่ไม่แพง และมีความสามารถในการยืดอายุของผลิตภัณฑ์ได้มากกว่าของไทยทำให้เกิดปัญหาสินค้าเหลือทิ้งน้อยกว่าของไทย ทั้งนี้การที่ธุรกิจเบเกอรี่จะอยู่รอดได้หรือไม่นั้น ผู้ประกอบการจะต้องเข้าใจความต้องการของลูกค้าในแต่ละกลุ่มเป้าหมายและสอดคล้องกับพฤติกรรมผู้บริโภคและลดจำนวนสินค้าเหลือทิ้งและปรับกลยุทธ์เพื่อเพิ่มกำไรตามสถานการณ์ที่เปลี่ยนแปลงไปเพื่อดึงดูดให้ลูกค้ากลับมาซื้อซ้ำเสมอ (กนิษฐ์กานต์, 2556)

ผู้วิจัยได้รับความอนุเคราะห์ในการศึกษาสหกิจกับร้านเบเกอรี่ในบริษัทแห่งหนึ่งของประเทศไทย ผู้วิจัยมองเห็นถึงปัญหาของธุรกิจเบเกอรี่ว่ามีการแข่งขันที่สูงมาก ดังนั้นจึงต้องมีปรับกลยุทธ์ที่ดีในการพัฒนาธุรกิจเบเกอรี่ โดยทางผู้วิจัยต้องการที่จะแก้ปัญหาผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านที่แตกต่างกันมากเกินไปให้เหลือผลต่างน้อยที่สุด เนื่องจากคะแนนการประเมินตนเองและคะแนนการประเมินจากผู้จัดการร้านของร้านที่มีประสิทธิภาพในการบริหารจัดการจะมีผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านต่ำ เพื่อให้ผลการประเมินมีประสิทธิภาพการประเมินตนเองจากร้านค้าและจากผู้จัดการร้านค้าจึงไม่ควรต่างกันมากเกินไปและต้องการเพิ่มกำไรให้กับทางบริษัทจึงมองเห็นถึงความสำคัญของการพัฒนากลยุทธ์ทางการตลาดโดยเฉพาะการจัดกลุ่มร้านค้า ผู้วิจัยสนใจที่จะจัด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลุ่มของร้านค้าที่คล้ายคลึงกันไว้ด้วยกันการจัดกลุ่มร้านค้าถือว่าเป็นเครื่องมือสำคัญในการจัดการธุรกิจ ช่วยให้ผู้บริหารเข้าใจความต้องการของลูกค้าแต่ละกลุ่มได้ดีขึ้น สามารถพัฒนากลยุทธ์ทางการตลาดได้ตรงจุดและบรรลุเป้าหมายทางธุรกิจได้ ซึ่งผู้วิจัยจะใช้เกณฑ์คะแนนการจัดร้านค้า คะแนนการประเมินตนเอง คะแนนการประเมินจากผู้จัดการร้าน ผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้าน คะแนนความน่าเข้ามาซื้อของ คะแนนที่ตั้งร้านค้า เปอร์เซ็นต์การขายของสินค้ากลุ่ม A ,B,C,D และ E เป็นตัวแปรหลักและเปรียบเทียบดูว่าแต่ละกลุ่มมีความเหมือนหรือแตกต่างกัน โดยใช้อาศัยวิธีการแบ่งกลุ่ม (Clustering) ในการจัดกลุ่ม เพื่อช่วยลดปัญหาสินค้าเหลือทิ้งมากเกินไปและเพิ่มกำไรให้มากขึ้นจากการวิเคราะห์กลุ่มร้านค้า ผู้วิจัยจึงสนใจที่จะจัดกลุ่มของร้านค้า เบเกอรี่ของบริษัทกรณีศึกษาแห่งนี้เพื่อให้บริษัทปรับกลยุทธ์ต่อไป

1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 เพื่อจัดกลุ่มของร้านค้าเบเกอรี่

1.2.2 เพื่อวิเคราะห์ความแตกต่างลักษณะของร้านค้าแต่ละกลุ่มที่จำแนกมา

1.3 ขอบเขตของงานวิจัย

1.3.1 ขอบเขตด้านข้อมูล ชุดข้อมูลที่ใช้สำหรับจัดกลุ่มร้านค้าเป็นข้อมูลจากทางบริษัทที่ผู้วิจัยทำสหกิจร่วมด้วยได้จัดการข้อมูลมาให้ซึ่งเป็นข้อมูลที่เก็บรวบรวมมา 1 ปี มีข้อมูลทั้งสิ้น 17 สดมภ์ (columns) 157 รายการ (rows)

1.3.2 ขอบเขตด้านระยะเวลา

ระยะเวลาในเวลาในการทำวิจัยตั้งแต่วันที่ 1 มกราคม 2567 ถึง 31 มีนาคม 2567

1.3.3 ขอบเขตด้านเครื่องมือ

1. ด้านซอฟแวร์ (Software)

1.1. ภาษาโปรแกรมไพทอน (Python)

1.1.1 Google Colab

1.2 ไลบรารี (Library) ที่ใช้

1.2.1 Pandas

1.2.2 Matplotlib

1.2.3 Scikit-learn

1.2.4 NumPy

1.3 IBM SPSS

2. เครื่องคอมพิวเตอร์มีคุณสมบัติดังนี้

CPU : Intel Core I3-8130U 2.2 Ghz With Turbo Boost Up To 3.4 Ghz

GPU : NVIDIA GeForce MX130 with 2 GB VRAM

Memory(Ram) : 4 GB

OS : Windows 10 Pro

System Type : 64-Bit Operating System, X64-Based Processor

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 จัดกลุ่มร้านค้าที่คล้ายคลึงกันไว้ด้วยกัน

1.4.2 ทราบกลุ่มของร้านค้าที่จำแนกมาว่ามีลักษณะเหมือนหรือแตกต่างกัน

1.5 นิยามศัพท์

1.5.1 ร้านค้า คือ สถานที่ขายของ (พจนานุกรมราชบัณฑิตสถาน, 2542)

1.5.2 เบเกอรี่ คือ ขนมปังและขนมอบต่างๆ เช่น คุ้กกี้ พาย (พจนานุกรมราชบัณฑิตสถาน, 2542)

1.5.3 กลยุทธ์ คือ การรบที่มีเล่ห์เหลี่ยมหรือวิธีการที่ต้องใช้กลอุบายต่าง ๆ (พจนานุกรมราชบัณฑิตสถาน, 2542)

1.5.4 การจัดกลุ่ม คือ การจัดข้อมูลที่มีลักษณะเหมือนกันไว้ในกลุ่มเดียวกันเป็นส่วนหลักของการทำเหมืองข้อมูล การเรียนรู้ของเครื่องแบบไม่มีผู้สอน (วิกิพีเดีย, 2565)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
งานวิจัยที่เกี่ยวข้อง

การแบ่งกลุ่มร้านค้าเบเกอรี่ กรณีศึกษาบริษัทแห่งหนึ่งผู้วิจัยได้ทำการศึกษาค้นคว้าแนวคิด ทฤษฎีและงานวิจัย ที่เกี่ยวข้องเพื่อเป็นพื้นฐานในการสนับสนุนงานวิจัยในครั้งนี้ โดยแบ่งประเด็นการทบทวนวรรณกรรมดังนี้

1. ทฤษฎีเกี่ยวกับการวิเคราะห์แบ่งกลุ่มข้อมูล (Clustering Analysis)
2. ทฤษฎีเกี่ยวกับการเรียนรู้ของเครื่อง (Machine learning)
3. ทฤษฎีทางสถิติ
4. ภาษาโปรแกรมที่เกี่ยวข้อง
5. งานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีเกี่ยวกับการวิเคราะห์การแบ่งกลุ่มข้อมูล (Clustering Analysis)

Clustering Analysis เป็นเทคนิคที่ใช้จัดกลุ่ม คน สัตว์ สิ่งของ ฯลฯ ที่เหมือนหรือคล้ายกันไว้ในกลุ่มเดียวกันเนื่องจากสามารถจำแนกกลุ่มย่อยได้ และรวบรวมระเบียนที่แตกต่างกันไว้ต่างกลุ่มกัน การจัดกลุ่มแตกต่างจากการจำแนกกลุ่ม คือ การจัดกลุ่มไม่มีตัวแปรเป้าหมายอัลกอริทึมของการจัดกลุ่มเป็นการค้นหาเพื่อจะแบ่งชุดข้อมูลทั้งหมดออกเป็นกลุ่ม ๆ โดยที่ระเบียนที่เหมือนหรือคล้ายคลึงกันมากที่สุดหรือมีความแตกต่างต่างกันน้อยที่สุดไว้ในกลุ่มเดียวกัน การจัดกลุ่มมีหลายวิธี แต่วิธีที่คนนิยมใช้กันนั้นมี 2 วิธี ได้แก่ การแบ่งกลุ่มแบบ K-Means และการจัดกลุ่มแบบลำดับขั้น (สายชล, 2558)

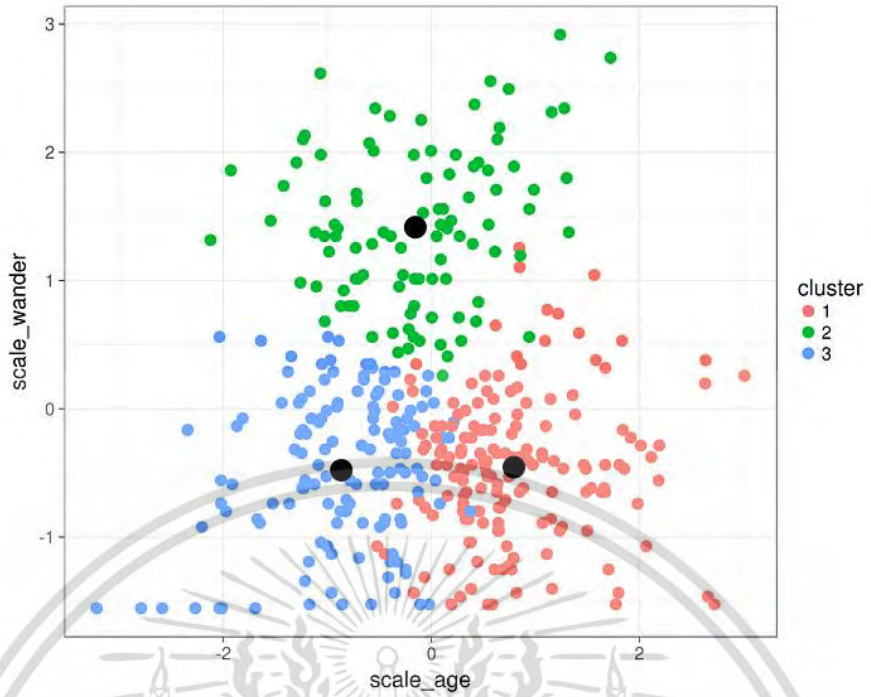
2.1.1 ประเภทของการวิเคราะห์การแบ่งกลุ่ม

1. K-mean Clustering โดยการแบ่งกลุ่มแบบ K-Meansเป็นวิธีที่นิยมใช้ในการแบ่งกลุ่มจากข้อมูล n จำนวน เป็น K กลุ่ม โดยเปรียบเทียบความคล้ายคลึงของข้อมูลกับจุดศูนย์กลางหรือค่าเฉลี่ยของแต่ละกลุ่ม เป็นการแบ่งกลุ่มย่อยตามจำนวนที่ต้องการ

ขั้นตอนการวิเคราะห์ของ K-Means มีดังนี้

1) แบ่งกลุ่มข้อมูลเป็น K กลุ่ม
2) ค้นหาจุดกึ่งกลางของกลุ่มแต่ละกลุ่ม
3) คำนวณระยะห่างกำลังสองของแต่ละหน่วยไปยังจุดกึ่งกลางกลุ่มที่หน่วยนั้นอยู่
4) การพิจารณาการย้ายกลุ่ม จะใช้เกณฑ์การย้ายตามค่าที่คำนวณได้ในขั้นตอนที่ 3 ถ้าขั้นที่ 4 ไม่มีการย้ายกลุ่มอีกแล้วแสดงว่ากลุ่มที่แบ่งได้นั้นเหมาะสมแล้วแต่ถ้าในขั้นที่ 4 มีการย้ายกลุ่มกลุ่มที่มีหน่วยย้ายเข้าหรือย้ายออกจะต้องทำการคำนวณหาจุดกลางกลุ่มใหม่ นั่นคือต้องกลับไปทำขั้นที่ 2 โดยตัวแปรที่ใช้ในเทคนิค K-Mean Clustering จะต้องเป็นตัวแปรเชิงปริมาณโดยไม่สามารถใช้กับข้อมูลที่อยู่ในรูปแบบความถี่เหมือนเทคนิค Hierarchical (เดชฤทธิ์, 2563) ซึ่งสามารถแสดงได้ดังรูป 2.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.1 รูปแสดงตัวอย่างการแบ่งกลุ่มโดยเทคนิคK-Means (GliScore6h, 2021)

ข้อดีของการทำ K-Means Clustering

1. เมื่อจำนวนข้อมูลมีจำนวนมากและมีจำนวนกลุ่มน้อยการหาค่าเฉลี่ยแบบ K-Means อาจจะคำนวณได้เร็วกว่าการจัดกลุ่มแบบอื่น ๆ (Hierarchical)
2. ขั้นตอนการหาค่าเฉลี่ยแบบ K-Means อาจจะได้สมาชิกภายในกลุ่มหนาแน่นกว่าการจัดกลุ่มแบบ Hierarchical โดยเฉพาะถ้ากลุ่มเป็นวงกลม

ข้อด้อยของการทำ K-Means Clustering

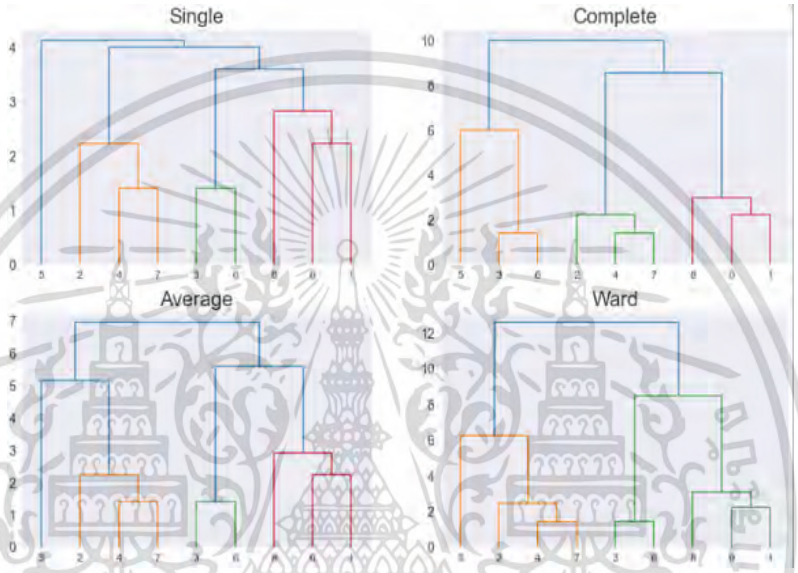
1. การหาค่า K ที่เหมาะสมคาดเดาได้ยาก
2. ทำงานได้ไม่ดีถ้ากลุ่มข้อมูลไม่เป็นรูปวงกลม
3. มีข้อจำกัดในเรื่องของขนาด ความหนาแน่น และรูปร่าง
4. ขั้นตอนวิธี K-Means จะใช้ได้ก็ต่อเมื่อข้อมูลในชุดข้อมูลนั้นสามารถนำมาคำนวณค่าเฉลี่ยได้อย่างมีความหมาย ในกรณีที่ข้อมูลแต่ละตัวประกอบด้วยฟีเจอร์ (Feature) ชนิดนามบัญญัติ เช่น ชุดข้อมูลลูกค้าบริษัทประกันภัยจะพบว่าระเบียนของลูกค้าประกอบด้วยเพศลูกค้า จังหวัดที่เป็นภูมิลำเนา ประเภทรถ ซึ่งเราจะเห็นได้ว่าการหาค่าศูนย์กลางของกลุ่มโดยใช้ฟีเจอร์เหล่านี้ทำให้ขาดความสมเหตุสมผล (นวลวรรณ, 2563)

2. Hierarchical Cluster Analysis เป็นวิธีการจัดกลุ่มแบบลำดับขั้น โดยจะพิจารณาการจัดกลุ่มจากระยะห่างของข้อมูลโดยข้อมูลที่มีความใกล้เคียงกันจะเชื่อมต่อกันเป็นกลุ่มก่อน จากนั้นค่อยเชื่อมข้อมูลที่มีระยะห่างอยู่ไกลออกไป ซึ่งระยะห่างที่แตกต่างกันออกไปสามารถอธิบายได้ด้วยแผนภาพเดนโดรแกรม (Dendrogram) วิธีจัดกลุ่มแบบลำดับขั้นสามารถทำได้ 2 วิธี โดยใช้วิธีจัดกลุ่มแบบรวมกัน (Agglomerative Clustering Method) หรือ วิธีการจัดกลุ่มแบบแยกกัน (Divisive Clustering Method) (นวลวรรณ, 2553)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1 วิธีจัดกลุ่มแบบรวมกัน (Agglomerative Clustering Method) จะค้นหาข้อมูลที่ใกล้เคียงกันมากที่สุดรวมไว้เข้ากันเป็นกลุ่มจากนั้นนำมารวมกันเป็นกลุ่มใหม่เป็นกลุ่มเดียว ทำให้จำนวนกลุ่มในชุดข้อมูลลดลงและท้ายที่สุดจะได้กลุ่มขนาดใหญ่เพียงกลุ่มเดียว ซึ่งวิธีการจัดกลุ่มแบบรวมกันเป็นวิธีที่นิยมมากที่สุด

2.2 วิธีการจัดกลุ่มแบบแยกกัน (Divisive Clustering Method) เริ่มต้นด้วยกลุ่มข้อมูลขนาดใหญ่และเริ่มแตกออกเป็นกลุ่มข้อมูลกลุ่มย่อย ๆ โดยจัดข้อมูลตามความไม่เหมือนกันของข้อมูล ซึ่งสามารถแสดงได้ดังรูปที่ 2.2



รูปที่ 2.2 รูปแสดงตัวอย่างการแบ่งกลุ่มโดยเทคนิค Hierarchical (Firdos, 2023)

2.1.2 การคำนวณระยะห่างข้อมูล

Euclidean Distance เป็นการวัดระยะห่างระหว่างจุดสองจุดเรียกว่าการคำนวณระยะห่างแบบยุคลิด คือความยาวของส่วนเส้นตรง pq ถ้า $p=(p_1,p_2,...,p_n)$ และ $q=(q_1,q_2,...,q_n)$ สามารถคำนวณได้จากสมการที่ 2.1 (นวลวรรณ, 2563)

$$d(p,q)=\sqrt{(p_1-q)^2+(p_2-q_2)^2+....+(p_n-q_n)^2}=\sqrt{\sum(p_i-q_i)^2} \qquad (2.1)$$

โดยที่ $d(p,d)$ คือ ระยะทางจาก p ไป q

q คือ จำนวนคุณลักษณะของข้อมูลข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.3 การกำหนดจำนวนกลุ่มที่เหมาะสม

การกำหนดจำนวนกลุ่มที่เหมาะสมมีทั้งหมด 2 วิธีที่เป็นที่นิยมในการกำหนดจำนวนกลุ่มก่อนที่จะทำการจัดกลุ่มได้แก่ วิธี Elbow Method และ Silhouette Coefficient

1. Elbow Method

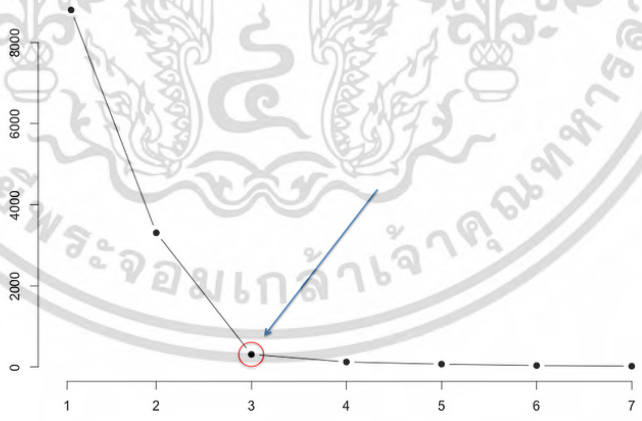
Elbow Method ซึ่งเป็นการพิจารณาข้อผิดพลาด ซึ่งถ้าจำนวนกลุ่มเพิ่มขึ้นก็จะทำให้ข้อผิดพลาดน้อยลงโดยเมื่อความผิดพลาดน้อยลงก็จะเริ่มอิ่มตัวและเกิดความชันของกราฟ ถ้าเกิดมุมที่มีลักษณะของจุดหักข้อศอก ๆ จะแสดงจำนวนกลุ่มที่เหมาะสมซึ่งอาจจะมีมากกว่า 1 จุดก็ได้ อธิบายได้ด้วยสมการที่ 2.2 (นวลวรรณ, 2563)

$$wcss = \sum_{c_k}^{c_n} \left(\sum_{d_i \, in \, c_i}^{d_m} distance\left(d_i, c_k\right) \right)^2 \tag{2.2}$$

โดยที่ c_k คือ จุดศูนย์กลางของกลุ่ม

d_i คือ จุดข้อมูลในแต่ละกลุ่ม

ในการจัดกลุ่มแต่ละรอบค่า WCSS จะมีค่าลดลงเรื่อย ๆ จากจำนวนกลุ่มที่เยอะขึ้น เพราะว่าสมาชิกในแต่ละกลุ่มจะลงลดไปเรื่อย ๆ ดังนั้นค่าความแปรปรวนภายในกลุ่มจะทำให้เกิดความโค้งที่เรียบขึ้นเรื่อย ๆ ดังภาพที่ 2.3



รูปที่ 2.3 รูปแสดงตัวอย่างจุดที่เหมาะสมของจำนวนกลุ่มโดยวิธี Elbow Method (ศศิวุฒิ, 2563)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จุดที่เหมาะสมของจำนวนกลุ่มคือจุดที่กราฟมีลักษณะ “หัก” ที่สุด ดังนั้นจากคำนวณหาจำนวนกลุ่มที่เหมาะสมที่สุดจึงสามารถหาได้จากระยะทางที่ไกลที่สุดนับจากเส้นตรงระหว่างสองจุดกับเส้นโค้ง

2. Silhouette Coefficient

Silhouette Coefficient เป็นเครื่องมือที่ใช้สำหรับการจัดกลุ่มโดยหาค่า Silhouette Coefficient โดยใช้ค่าเฉลี่ยกับระยะห่างระหว่างจุดกับกลุ่มที่อยู่ใกล้ชิดเทียบกับจุดที่อยู่ภายในกลุ่มเดียวกันซึ่งการคำนวณ Silhouette Coefficient นั้นเหมาะกับการจัดกลุ่มข้อมูลที่ข้อมูลแต่ละกลุ่มเกาะกลุ่มกัน ค่าที่ได้จากการจะอยู่ระหว่าง -1 ถึง 1 ค่าเข้าใกล้ 1 นั่นหมายความว่าเป็นการจัดที่เหมาะสมหากค่าที่ได้เข้าใกล้ -1 แสดงว่าข้อมูลไม่เหมาะสมกับกลุ่มที่จัดไว้ ควรจัดให้อยู่กลุ่มอื่นๆ แต่หากค่าที่เข้าใกล้ 0 นั่นหมายความว่าข้อมูลยังไม่เหมาะสมกับกลุ่มที่จัดไว้ โดยสามารถพิจารณาจากการคำนวณดังสมการที่ 2.3 (อุนนดา, 2564)

$$S_i = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{2.3}$$

โดยที่ a_i คือ ค่าเฉลี่ยระยะห่างระหว่างจุดที่อยู่ข้างในกลุ่มเดียวกัน

b_i คือ ค่าเฉลี่ยระยะห่างระหว่างจุดที่อยู่ข้างในกลุ่มกับข้อมูลทั้งหมด

2.2 ทฤษฎีทางสถิติ

2.2.1 สถิติพรรณนา (Descriptive Statistics) คือ สถิติที่ใช้วิเคราะห์ข้อมูลที่เก็บรวบรวมมาได้ไม่ว่าจะเป็นข้อมูลปฐมภูมิหรือทุติยภูมิเพื่อจำแนกและบรรยายข้อมูลให้เห็นถึงคุณลักษณะบางอย่างของกลุ่มใดกลุ่มหนึ่งไม่เกี่ยวข้องกับกลุ่มประชากรรวมทั้งวิเคราะห์ข้อมูลอย่างง่าย (สายชล, 2560) โดยสถิติพรรณนาประกอบด้วย

1. ค่าเฉลี่ย (Mean)

ค่าเฉลี่ย คือ เป็นการหาค่าเฉลี่ยของข้อมูลเชิงปริมาณที่อยู่ในระดับมาตราอัตราภาคชั้น (Interval Scale) หรืออัตราส่วน (Ratio Scale) การคำนวณหาค่าเฉลี่ยเลขคณิตสามารถคำนวณหาได้จากสมการที่ 2.4

$$\overline{X} = \frac{\sum x}{n} \tag{2.4}$$

โดยที่ $\sum x$ คือ ผลรวมของข้อมูลทั้งหมด

n คือ ขนาดตัวอย่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. พิสัย (Range)

พิสัย คือ ค่าที่ใช้วัดการกระจายของข้อมูล ที่ได้จากผลต่างระหว่างค่าจากการสังเกตที่มีค่าสูงสุดและต่ำสุด แต่พิสัยเป็นการวัดการกระจายที่ไม่ค่อยดีเมื่อเทียบกับตัวอื่นๆ เพราะเป็นค่าที่คำนวณจากค่าเพียงสองค่าเท่านั้น สามารถคำนวณหาได้จากสมการที่ 2.5

$$Range = Max - Min \tag{2.5}$$

โดยที่ Max คือ ค่าที่สูงสุดของชุดข้อมูล

Min คือ ค่าที่ต่ำสุดของชุดข้อมูล

3. ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation)

ค่าเบี่ยงเบนมาตรฐาน คือ ค่าที่วัดการกระจายของข้อมูลที่ได้จากรากที่สองที่ไม่เป็นจำนวนลบของค่าเฉลี่ยกำลังสองของผลต่างระหว่างค่าในข้อมูลกับค่าเฉลี่ยเลขคณิตของข้อมูลนั้น ค่าเฉลี่ยมาตรฐานเป็นวิธีการวัดการกระจายที่ดีที่สุด สามารถคำนวณหาได้จากสมการที่ 2.6

$$S.D. = \sqrt{\frac{\left(\sum_{i=1}^{n} x_i - \bar{x}\right)^2}{n-1}} \tag{2.6}$$

โดยที่ $S.D.$ คือ ค่าส่วนเบี่ยงเบนมาตรฐาน

x_i คือ คะแนนของตัวอย่างชุดข้อมูล

$\bar{x}$ คือ ค่าเฉลี่ยของชุดข้อมูล

n คือ ขนาดตัวอย่าง

4. ค่าร้อยละหรือเปอร์เซ็นต์ (Percentage)

ค่าร้อยละหรือเปอร์เซ็นต์ คือการคำนวณหาสัดส่วนข้อมูลแต่ละตัวเทียบกับข้อมูลทั้งหมดโดยให้ข้อมูลรวมทั้งหมดคือ 100 % สามารถคำนวณหาได้จากสมการที่ 2.7

$$\text{Percentage} = \frac{x}{n} \times 100 \tag{2.7}$$

โดยที่ x คือ จำนวนข้อมูลหรือความถี่

n คือ ขนาดตัวอย่าง

2.2.2 สถิติอนุมาน (Inference statistics) คือ สถิติที่กล่าวถึงเทคนิคต่างๆที่ใช้ในการวิเคราะห์ข้อมูลที่เก็บรวบรวมจากกลุ่มตัวอย่างและนำข้อมูลที่ได้จากตัวอย่างใช้เป็นตัวแทนประชากรและนำไปสู่ผลสรุปโดยอาศัยกระบวนการต่างๆ เช่น การประมาณค่า (สรชัย, 2559) ซึ่งในที่นี้ผู้วิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จะนำทฤษฎีที่เกี่ยวข้องสำหรับการทำวิจัยมาพิจารณาเท่านั้น ได้แก่ การทดสอบการแจกแจงปรกติ และการเปรียบเทียบค่าเฉลี่ยของประชากร

1. การทดสอบการแจกแจงปรกติ (Normality Test)

1) การทดสอบคอลโมโกรอฟ-สมีร์นอฟ (Kolmogorov-Smirnov Test)

การทดสอบคอลโมโกรอฟ-สมีร์นอฟ ใช้กับข้อมูลตั้งแต่มาตรวัดเชิงอันดับขึ้นไป และถือว่าประชากรเป็นแบบต่อเนื่อง โดยต้องกำหนดค่าเฉลี่ยของประชากร $\mathrm{u} = \mathrm{u}_0$ และความแปรปรวนของประชากร $\sigma^2 = {\sigma_0}^2$ ไว้ในสมมติฐานหลักเสมอ (สายชล, 2563)

ข้อจำกัดเบื้องต้น (Assumption)

ให้ $x_1,...,x_n$ เป็นตัวอย่างการสุ่มมาจากประชากรที่มีฟังก์ชันการแจกแจงสะสม $F(x)$ ที่ไม่ทราบค่า

สมมติฐาน

$$H_0 : F(x) = F_0(x)$$

$$H_0 : F(x) \neq F_0(x) \text{ หรือ}$$

$$H_1 : F(x) < F_0(x)$$

$$H_1 : F(x) > F_0(x)$$

ให้ $S(x) =$ ฟังก์ชันการแจกแจงสะสมที่ได้จากค่าสังเกต

$$= \frac{k}{n}$$

เมื่อ k คือจำนวน x_i ที่น้อยกว่าหรือเท่ากับ x

ตัวสถิติทดสอบ

$$D = Max\left|S(x) - F_0(x)\right| \tag{2.8}$$

2) การทดสอบลิลลีโฟรส์ (Lillefors Test)

การทดสอบลิลลีโฟรส์ (Lillefors Test) จะแตกต่างจากการทดสอบของการทดสอบคอลโมโกรอฟ-สมีร์นอฟ จะมีการกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของประชากร (สุจิตรา, 2565)

ข้อจำกัดเบื้องต้น

ข้อมูลประกอบด้วยค่าสังเกต $x_1,...,x_n$ ที่เป็นอิสระต่อกันจากฟังก์ชันการแจกแจงสะสม $F(x)$ ที่มีค่าเฉลี่ย μ ไม่ทราบค่าและค่าความแปรปรวน σ^2 ไม่ทราบค่า

สมมติฐาน

H_0 : ตัวอย่างมาจากประชากรที่มีการแจกแจงปรกติ

H_1 : ตัวอย่างมาจากประชากรที่ไม่ได้มีการแจกแจงปรกติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารทุกครั้งที่มีการนำไปใช้

ตัวสถิติทดสอบ

$$D = Max\left|S\left(x\right) - F_0\left(x\right)\right| \tag{2.9}$$

การสรุปผล

จะยอมรับ H_0 ถ้าค่า p-value มีค่ามากกว่าระดับนัยสำคัญที่กำหนด

3) การทดสอบชาพิโร-วิลก์ (Shapiro-Wilk Test)

การทดสอบชาพิโร-วิลก์ เป็นตัวสถิติที่ใช้ทดสอบการแจกแจงของตัวแปรเชิงปริมาณว่ามีการแจกแจงปรกติหรือไม่ และใช้ในกรณีที่ทราบหรือไม่ทราบค่าเฉลี่ยหรือความแปรปรวนของประชากรก็ได้ และตัวอย่างมีขนาดไม่เกิน 50 หน่วย ข้อมูลประกอบด้วยตัวอย่างสุ่ม $x_1,...,x_n$ ที่มาจากประชากรที่มีฟังก์ชันการแจกแจงสะสม $F\left(x\right)$ ไม่ทราบค่า (สายชล, 2563)

ข้อตกลงเบื้องต้น

ตัวอย่างเป็นตัวอย่างสุ่มมาจากประชากร

สมมติฐาน

H_0 : ตัวอย่างมาจากประชากรที่มีการแจกแจงปรกติ

H_1 : ตัวอย่างมาจากประชากรที่ไม่ได้มีการแจกแจงปรกติ

ตัวสถิติทดสอบ

$$T = \frac{1}{D}[\sum_{i=1}^{k} a_i(x^{(n-i+1)} - x^{(i)})]^2 \quad \text{โดยที่} \quad D = \sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{2.10}$$

เมื่อ $\bar{x}$ คือ ค่าเฉลี่ยของตัวอย่าง

a_i คือ ค่าสัมประสิทธิ์ที่ได้จากการเปิดตาราง

x^i คือ ค่าสังเกตตัวอย่างแบบเรียงลำดับ

การสรุปผล

ค่า T มีค่าเข้าใกล้ 1 นั่นหมายความว่าตัวอย่างมีการแจกแจงปรกติ แต่ถ้าค่า T มีค่าน้อยกว่า 1 มาก นั่นหมายความว่าตัวอย่างไม่มีการแจกแจงปรกติ

2. การเปรียบเทียบค่าเฉลี่ยของประชากร

การเปรียบเทียบค่าเฉลี่ยของประชากรต้องมีการทดสอบการแจกแจงปรกติของข้อมูลก่อน หากมีการแจกแจงปรกติให้เลือกใช้สถิติที่ใช้พารามิเตอร์ หากมีการแจกแจงไม่ปรกติให้เลือกใช้สถิติที่ไม่ใช้พารามิเตอร์ โดยเลือกตัวสถิติที่ใช้ทดสอบตามจำนวนกลุ่ม (สายชล, 2563)

กรณีข้อมูลมีการแจกแจงปรกติ

1) การเปรียบเทียบค่าเฉลี่ยประชากร 2 กลุ่ม กรณีข้อมูลมีการแจกแจงปรกติในการเปรียบเทียบค่าเฉลี่ยประชากร 2 กลุ่มต้องทดสอบด้วยว่าประชากรทั้งสองกลุ่มเป็นอิสระต่อกันหรือไม่เพื่อใช้สถิติทดสอบที่ถูกต้อง (สิทธิชัย, 2565)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สมมติฐาน

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

กรณีประชากรทั้งสองเป็นอิสระต่อกัน สามารถเลือกสถิติทดสอบตามข้อตกลงเบื้องต้นได้ ดังตาราง 2.1

ตารางที่ 2.1 การทดสอบค่าเฉลี่ยของประชากร 2 ประชากรเมื่อเป็นอิสระต่อกัน

ในกรณีที่	สถิติที่ใช้ทดสอบ
1. ทราบค่าความแปรปรวนของประชากรทั้งสองกลุ่ม	$z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
2. ไม่ทราบค่าความแปรปรวนของประชากรทั้งสองกลุ่ม แต่ขนาดตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 มากกว่า 30 ทั้งคู่	$z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
3. ไม่ทราบค่าความแปรปรวนของประชากรทั้งสองกลุ่ม และขนาดตัวอย่างของประชากรทั้งกลุ่มน้อยกว่า 30 แต่ทราบว่ามีความแปรปรวนเท่ากัน	$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \frac{1}{n_1} + \frac{1}{n_2}}}$ โดยที่ $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$
4. ไม่ทราบความแปรปรวนของประชากรทั้งสองกลุ่ม และขนาดตัวอย่างของประชากรทั้งสองกลุ่ม มีขนาดน้อยกว่า 30 ทั้งคู่ แต่ทราบความแปรปรวนของประชากรทั้งสองกลุ่มมีค่าไม่เท่ากัน	$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กรณีประชากรทั้งสองไม่เป็นอิสระต่อกัน สามารถเลือกสถิติทดสอบตามข้อตกลงเบื้องต้นได้ ดังตาราง 2.2

ตารางที่ 2.2 การทดสอบค่าเฉลี่ยของประชากร 2 ประชากรเมื่อไม่เป็นอิสระต่อกัน

กรณีที่	สถิติที่ใช้ในการทดสอบ
1. ค่าสังเกตุที่ได้มาจากวัตถุทดลองเดียวกัน หรือเป็นแบบจับคู่	$t = \frac{\bar{d} - \mu_D}{S_D / \sqrt{n}}$

2) การเปรียบเทียบค่าเฉลี่ยประชากรตั้งแต่ 3 กลุ่มขึ้นไป กรณีข้อมูลมีการแจกแจงปรกติ โดยจะต้องทำการทดสอบความเป็นอิสระต่อกันของประชากรแต่ละกลุ่มก่อน กรณีที่แต่ละประชากรเป็นอิสระต่อกัน จะใช้การทดสอบความแปรปรวนทางเดียว (One-way-ANOVA) ซึ่งมีลักษณะ ดังตารางที่ 2.3

ตารางที่ 2.3 ตัวอย่างตารางข้อมูลสำหรับการวิเคราะห์ความแปรปรวนทางเดียว

	คะแนนสอบนักศึกษาแต่ละชั้นปี		
1	2	3	4
4	3	2	4
10	15	6	9
3	9	6	5
3	4	5	6
ผลรวม 20	31	19	24

สมมติฐาน

$H_0: \mu_1 = \mu_2 = \mu_3 ... = \mu_k$

H_1: มีค่าเฉลี่ยประชากรอย่างน้อย 1 คู่ ที่แตกต่างกัน

การวิเคราะห์ค่าสถิติ

1. CT = $\frac{(\text{ข้อมูลทั้งหมด})^2}{\text{หน่วยทดลองทั้งหมด}}$
2. Sum of squares
   1) Total SS = ผลบวกของ $(\text{ข้อมูลจากแต่ละหน่วยทดลอง})^2$ - CT
   2) Treatments SS = $\frac{\text{ผลบวกของ } (\text{ผลรวมหน่วยทดลอง})^2}{\text{จำนวนซ้ำ}}$ − CT

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) $Error\ SS = Total\ SS - Treatments\ SS$

3. Mean squares

1) $Treatments\ MS = \frac{Treatments\ SS}{df\ Treatments}$

2) $Error\ MS = \frac{Error\ SS}{df\ Error}$

4. F-value

$F = \frac{Treatments\ MS}{Error\ MS}$; เป็นตัวสถิติที่ใช้ทดสอบ (2.11)

5. แสดงตารางวิเคราะห์ความแปรปรวน

ตารางการวิเคราะห์ความแปรปรวน แสดงดังตารางที่ 2.4

ตารางที่ 2.4 ตารางวิเคราะห์ความแปรปรวน

Source of Variation	df	SS	MS	F-ratio
ระหว่างกลุ่ม	t-1	Treatments SS	Treatments MS	F
ภายในกลุ่ม	t(r-1)	Error SS	Error MS	
Total	tr-1	Total SS		

โดยที่

Source of Variation	แสดงถึงแหล่งความแปรปรวนที่แยกออกมาวิเคราะห์
Df	ระดับหรือองศาของความเป็นอิสระ
SS	ผลรวมกำลังสองของค่าเบี่ยงเบนจากค่าเฉลี่ย
MS	ผลเฉลี่ยของผลรวมกำลังสองของค่าเบี่ยงเบนจากค่าเฉลี่ย
F	สถิติที่ใช้ในการทดสอบ
t	จำนวนสิ่งทดลอง
r	จำนวนซ้ำ

การสรุปผล

หากค่า F-value น้อยกว่าระดับนัยสำคัญที่กำหนด นั่นหมายความว่าปฏิเสธสมมติฐานหลัก ต้องมีการเปรียบเทียบเชิงซ้อนต่อไป เพื่อดูว่ากลุ่มใดที่แตกต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กรณีที่แต่ละประชากรไม่เป็นอิสระต่อกัน จะใช้การทดสอบความแปรปรวนสองทาง (Two-way-ANOVA) ซึ่งมีลักษณะ ดังตารางที่ 2.5

ตารางที่ 2.5 ตัวอย่างตารางข้อมูลสำหรับการวิเคราะห์ความแปรปรวนสองทาง

ตำแหน่งที่วางขนม	ร้านค้า					รวม
	A	B	C	D	E	
ชั้นบน	7	8	9	10	11	45
ชั้นกลาง	9	9	9	9	12	48
ชั้นล่าง	10	10	12	12	14	58
รวม	26	27	30	31	37	151

สมมติฐาน

$H_0: \mu_1 = \mu_2 = \mu_3 ... = \mu_k$

$H_1:$ มีค่าเฉลี่ยประชากรอย่างน้อย 1 คู่ ที่แตกต่างกัน

การวิเคราะห์ค่าสถิติ

1. CT = $\frac{(\text{ผลรวมทั้งหมดในการทดลอง})^2}{\text{จำนวนข้อมูลทั้งหมด}}$
2. Sum of squares
   1) Total SS = ผลบวกของ $(\text{ข้อมูลจากแต่ละหน่วยทดลอง})^2$ - CT
   2) Block SS = $\frac{\text{ผลบวกของ } (\text{ผลรวมของแต่ละบล็อค})^2}{\text{จำนวนสิ่งทดลอง}}$
   3) Treatments SS = $\frac{\text{ผลบวกของ } (\text{ผลรวมของแต่ละสิ่งทดลอง})^2}{\text{จำนวนบล็อค}} - \text{CT}$
   4) $Error\ SS = Total\ SS - Total\ Block - Treatments\ SS$
3. Mean squares
   1) $Block\ MS = \frac{Block\ SS}{df\ Block}$
   2) $Treatments\ MS = \frac{Treatments\ SS}{df\ Treatments}$
   3) $Error\ MS = \frac{Error\ SS}{df\ Error}$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. F-value

1) $F_{Block} = \frac{Block\ MS}{Error\ MS}$; เป็นตัวสถิติที่ใช้ทดสอบ (2.12)

2) $F_{treat} = \frac{Treatments\ MS}{Error\ MS}$ (2.13)

5. แสดงตารางวิเคราะห์ความแปรปรวน

ตารางการวิเคราะห์ความแปรปรวน แสดงดังตารางที่ 2.6

ตารางที่ 2.6 ตารางวิเคราะห์ความแปรปรวนสองทาง

Source of Variation	df	SS	MS	F-ratio
ระหว่างบล็อค	r-1	Block SS	Block MS	
ระหว่างกลุ่ม	t-1	Treatments SS	Treat MS	F ของ Treatments
ภายในกลุ่ม	(r-1) (t-1)	Error SS	Error MS	
Total	n-1	Total SS		

โดยที่ Source of Variation แสดงถึงแหล่งความแปรปรวนที่แยกออกมาวิเคราะห์

df ระดับหรือองศาของความเป็นอิสระ

SS ผลรวมกำลังสองของค่าเบี่ยงเบนจากค่าเฉลี่ย

MS ผลเฉลี่ยของผลรวมกำลังสองของค่าเบี่ยงเบนจากค่าเฉลี่ย

F สถิติที่ใช้ในการทดสอบ

t จำนวนสิ่งทดลอง

r จำนวนซ้ำ

n ขนาดตัวอย่าง

การสรุปผล

หากค่า F-value น้อยกว่าระดับนัยสำคัญที่กำหนด นั่นหมายความว่าปฏิเสธสมมติฐานหลัก ต้องมีการเปรียบเทียบเชิงซ้อนต่อไป เพื่อดูว่ากลุ่มใดบ้างที่แตกต่างกัน

3) การเปรียบเทียบเชิงซ้อน (Multiple Comparison)

การเปรียบเทียบเชิงซ้อน จะนำผลที่ได้จากการวิเคราะห์ความแปรปรวนที่ปฏิเสธสมมติฐานหลักมาดูว่าค่าเฉลี่ยตัวใดบ้างที่แตกต่างกัน จึงต้องมีการเลือกใช้ให้เหมาะสมกับข้อมูลซึ่งผู้วิจัยจะยกมาเพียงการทดสอบที่เกี่ยวข้องกับงานวิจัยเท่านั้น (ชูใจ, 2563)

1. การเปรียบเทียบเชิงซ้อน Dunnett's T3 Test

การเปรียบเทียบเชิงซ้อนของดันเนท จะใช้ในกรณีที่ความแปรปรวนไม่เท่ากัน ซึ่งอาจจะเกิดจากกลุ่มตัวอย่างที่เท่ากันหรือไม่เท่ากันก็ได้ (สุพัฒน์, 2555) สามารถคำนวณได้จากสมการที่ 2.14

$$\bar{d}_D = \frac{q_D\sqrt{2\left(MS_{\frac{S}{A}}\right)}}{\sqrt{S}} \quad (2.14)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่ $\bar{d}_D$ ค่าสถิติที่ใช้บรรยาย Dunnett's T3 Test

q_D ค่าจากตาราง Critical Values of the Dunnett Test

S ขนาดของกลุ่มตัวอย่าง

$MS_{\frac{S}{A}}$ ค่าความแปรปรวนภายในกลุ่ม

2. การเปรียบเทียบเชิงซ้อน Tukey

การเปรียบเทียบเชิงซ้อนของเทอกี้ ใช้เมื่อขนาดของกลุ่มตัวอย่างในแต่ละกลุ่มไม่เท่ากันและความแปรปรวนเท่ากัน สามารถคำนวณได้จากสมการที่ 2.15

$$W\left(HSD\right)=q_{\alpha}\left(t,v\right)\sqrt{\frac{MSE}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \qquad (2.15)$$

โดยที่ n_i และ n_j คือ จำนวนค่าสังเกตหรือขนาดตัวอย่าง

t คือ จำนวนสิ่งทดลองที่นำมาเปรียบเทียบค่าเฉลี่ย

MSE คือ ค่าความคลาดเคลื่อนจากตารางการวิเคราะห์ความแปรปรวนด้วยองศาความเป็นอิสระ v

$q_{\alpha}\left(t,v\right)$ คือ ค่าวิกฤตขอบเขตบนจากตารางค่าพิสัยของสติวเดนท์

กรณีข้อมูลไม่มีการแจกแจงปรกติ

หลังจากทดสอบการแจงแจงปรกติหากทดสอบแล้วพบว่าไม่มีการแจกแจงปรกติจะใช้สถิติไม่อิงพารามิเตอร์ในการทดสอบจะเลือกใช้การทดสอบตามจำนวนกลุ่มประชากร ดังนี้

1) การทดสอบของแมน-วิตนีย์ (The Mann-whitney Test)

การทดสอบของแมน-วิตนีย์ใช้เปรียบเทียบความแตกต่างประชากร 2 กลุ่มที่ไม่มีการแจกแจกปรกติและเป็นอิสระต่อกัน (สายชล, 2563)

ข้อกำหนดเบื้องต้น

1. ให้ค่าสังเกต X และ Y แต่ละตัวมีการแจกแจงเหมือนกันและเป็นอิสระกัน
2. ตัวอย่างทั้งสองกลุ่มเป็นอิสระกัน
3. ประชากรทั้งสองกลุ่มเป็นแบบต่อเนื่อง

สมมติฐาน

$H_0: M_X = M_Y$ $\quad H_0: M_X \geq M_Y$ $\quad H_0: M_X \leq M_Y$

$H_1: M_X = M_Y$ $\quad H_1: M_X < M_Y$ $\quad H_1: M_X > M_Y$

สถิติที่ใช้ทดสอบ

$$T = S - \frac{n_1\left(n_1+1\right)}{2} \qquad (2.16)$$

เมื่อ S คือ ผลรวมลำดับที่ของขนาดตัวอย่างในข้อมูลรวมทั้งหมดที่เรียบเรียงแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การสรุปผล

กรณีการทดสอบสองทาง จะปฏิเสธสมมติฐานหลักเมื่อค่า T มากหรือน้อยเกินไป

กรณีทดสอบสอบหางเดียวด้านซ้าย จะปฏิเสธสมมติฐานหลักเมื่อค่า T น้อยเกินไป

กรณีทดสอบสอบหางเดียวด้านขวา จะปฏิเสธสมมติฐานหลักเมื่อค่า T มากเกินไป

2) การทดสอบของวิลค็อกสัน (The Willcoxon Signed-Ranks Test)

การทดสอบของวิลค็อกสัน ใช้เปรียบเทียบความแตกต่างประชากร 2 กลุ่มที่ไม่มีการแจกแจงปรกติและไม่เป็นอิสระต่อกัน

ข้อกำหนดเบื้องต้น

1. ค่า X และ Y ของแต่ละคู่ต่าง ๆ ไม่ต้องทดสอบการคล้ายคลึงกันมากที่สุด
2. ผลต่างของข้อมูลแต่ละคู่เป็นตัวแปรสุ่มแบบต่อเนื่องและเป็นอิสระกัน
3. ผลต่างของข้อมูลมีการแจกแจงปรกติ
4. ผลต่างของข้อมูลมีมาตราวัดแบบช่วง

สมมติฐาน

$$H_0 : M_X = M_Y \qquad H_0 : M_X \geq M_Y \qquad H_0 : M_X \leq M_Y$$

$$H_1 : M_X = M_Y \qquad H_1 : M_X < M_Y \qquad H_1 : M_X > M_Y$$

สถิติที่ใช้ทดสอบ

กรณีตัวอย่างขนาดเล็ก

$$T = min(T_+, T_-) \tag{2.17}$$

กรณีตัวอย่างขนาดใหญ่

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \tag{2.18}$$

การสรุปผล

ตัวอย่างขนาดเล็ก

กรณีทดสอบสอบหางเดียวด้านซ้าย จะปฏิเสธสมมติฐานหลักเมื่อค่า $T_+ \leq T_{\alpha;n}$

กรณีทดสอบสอบหางเดียวด้านขวา จะปฏิเสธสมมติฐานหลักเมื่อค่า $T_- \leq T_{\alpha;n}$

กรณีการทดสอบสองทาง จะปฏิเสธสมมติฐานหลักเมื่อค่า $T \leq T_{\frac{\alpha}{2};n}$

ตัวอย่างขนาดใหญ่

กรณีทดสอบสอบหางเดียวด้านซ้าย จะปฏิเสธสมมติฐานหลักเมื่อค่า $Z > Z_\alpha$

กรณีทดสอบสอบหางเดียวด้านขวา จะปฏิเสธสมมติฐานหลักเมื่อค่า $Z < -Z_\alpha$

กรณีการทดสอบสองทาง จะปฏิเสธสมมติฐานหลักเมื่อค่า $Z > Z_{\frac{\alpha}{2}}$ หรือ $Z < -Z_{\frac{\alpha}{2}}$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) การทดสอบของคลัสคาล-วอลลิส (The Kruskal-Wallis Test)

การทดสอบของคลัสคาล-วอลลิส ใช้ทดสอบสมมติฐานว่างว่าไม่มีความแตกต่างของอิทธิพลในแต่ละกลุ่ม

สมมติฐาน

$H_0 : M_1 = M_2 = M_3 \ldots = M_k$

$H_1 : M_i \neq M_j$ แตกต่างกันอย่างน้อย 1 คู่

สถิติที่ใช้ทดสอบ

$$H = \frac{12}{n(n-1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} \qquad (2.19)$$

โดยที่ R_i ค่าลำดับที่ของค่าสังเกตในตัวอย่างที่ i

การสรุปผล

จะปฏิเสธสมมติฐานหลักเมื่อค่า $H \geq h_\alpha$

2.3 ทฤษฎีเกี่ยวกับการเรียนรู้ของเครื่อง (Machine learning)

ทอม มิทเชล (Tom Michell) ได้ให้นิยามการเรียนรู้ของเครื่อง คือโปรแกรมคอมพิวเตอร์ที่สามารถเรียนรู้ในงานจากชุดข้อมูลที่มีเป้าหมายเฉพาะ ที่ถูกกำหนดนรูปแบบของคลาสในชุดข้อมูลนั้น โดยโปรแกรมต้องสามารถแสดงให้เห็นถึงความสามารถในการเรียนรู้ได้ผ่านการวัดค่าประสิทธิภาพที่มีค่าสูงขึ้นเมื่อได้เรียนรู้จากข้อมูลที่มากขึ้น (MiScore6hell, 1997) ขั้นตอนการเรียนรู้ของเครื่องแบ่งได้ 4 ประเภท ดังนี้

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นขั้นตอนวิธีที่ต้องการชุดข้อมูลฝึกที่มีป้ายคลาส (Class Label) หรือฉลากซึ่งทำหน้าที่เป็นเฉลยของตัวอย่างแต่ละตัวเพื่อให้ขั้นตอนวิธีสามารถเรียนรู้จากชุดข้อมูลฝึกในแต่ละคลาสซึ่งนำไปสู่ความสามารถในการจำแนกคลาสหรือทำนายค่าผลลัพธ์ เช่น ถ้าเราต้องการสร้างแอปพลิเคชันที่สามารถจำแนกอีเมลออกเป็น 2 คลาส โดยต้องการใช้ขั้นตอนวิธีการเรียนรู้แบบมีผู้สอน ผู้วิจัยต้องเตรียมอีเมลจำนวนที่เหมาะสมเป็นชุดข้อมูลฝึกโดยอีมลแต่ละฉบับต้องมีป้ายคลาสที่เฉลยว่าเป็นอีเมลในคลาสใด เช่น อีเมลปกติหรือแสปมเมล ขั้นตอนการเรียนรู้แบบมีผู้สอนได้แก่ นาอีฟเบย์ ต้นไม้ตัดสินใจ เป็นต้น

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นขั้นตอนวิธีที่ไม่ต้องการป้ายคลาสในแต่ละตัวอย่าง การทำงานจะเน้นการจัดกลุ่มของตัวอย่างโดยใช้หลักความเหมือนของฟีเจอร์ในตัวอย่าง เช่น ขั้นตอนการจัดกลุ่มวิธี K-Means (Jones, 2008)

3. การเรียนรู้แบบกึ่งมีผู้สอน (Semi-Supervised Learning) เป็นวิธีที่เรียนรู้จากชุดข้อมูลที่บางตัวอย่างมีป้ายคลาสแต่ตัวอย่างส่วนใหญ่ไม่มีป้ายคลาส เช่น ปัญหาการจำแนกเนื้อหาของเว็บเพจ จะพบว่าการกำหนดป้ายคลาสเพื่อเฉลยประเภทของเนื้อหาโดยผู้เชี่ยวชาญต้องใช้เวลามากเมื่อเทียบกับจำนวนเว็บเพจที่มีจำนวนมาก ดังนั้นจึงต้องมีผู้วิจัยที่คิดค้นขั้นตอนวิธีที่สามารถเรียนรู้จากข้อมูลฝึกที่มีป้ายคลาสบางตัวอย่าง (Blum, 1998)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) เป็นขั้นตอนการเรียนรู้ที่ไม่ต้องมีป้ายคลาส แต่ขั้นตอนจะใช้กลไกการให้รางวัลและการลงโทษในระหว่างการเรียนรู้ (Silver, 2016)

2.4 ภาษาโปรแกรมที่เกี่ยวข้อง

2.3.1 ภาษาไพทอน (Python)

ภาษาไพทอน (Python) เป็นภาษาการเขียนโปรแกรมระดับสูงที่สร้างโดย Guido Van Rossum โดยเริ่มในปี พ.ศ. 2553 มีการพัฒนามาจาก Matlab จะประมวลผลคำสั่งทีละบรรทัดแต่ข้อแตกต่างระหว่างไพทอนและ Matlab คือ ภาษาไพทอนผู้สร้างไม่คิดมูลค่ากับสิ่งที่สร้างสรรค์ขึ้นมาและเน้นการเขียนที่ง่าย รวดเร็ว มีแนวคิดในการเขียนโปรแกรมเชิงวัตถุ ง่ายต่อการบำรุงและแก้ไขอีกทั้งยังเป็นภาษาที่มาพร้อมกับไลบรารีมาตรฐานจำนวนมากและยังเป็นที่นิยมในการทำงานด้านวิทยาการข้อมูล (จักรกฤษณ์, 2549)

ข้อดีของการใช้ภาษาไพทอนในการเขียนโปรแกรม

1) เป็นภาษาที่ง่ายต่อการเรียนรู้เมื่อเทียบกับภาษาอื่น
2) สามารถทำงานได้เร็วกว่าภาษาอื่น
3) หน่วยความจำน้อยในการประมวลผลโดยเฉพาะในการทำงานกับข้อมูลขนาดใหญ่
4) มีไลบรารีต่างๆมากมาย

โดยมีไลบรารีพื้นฐานดังนี้

1) Pandas เป็นไลบรารีการจัดการข้อมูล (Data Wrangling/ Data Cleaning) และการวิเคราะห์ข้อมูล (Data Analysis) ซึ่งถูกพัฒนามาเพื่อทำงานกับข้อมูลขนาดใหญ่

2) NumPy มีชื่อเต็มว่า “Numerical Python” โดดเด่นในด้านการคำนวณ และการทำงานกับตัวเลข นอกจากนี้ NumPy ยังมีความสามารถสำคัญในการสร้าง Array (โครงสร้างข้อมูล) อีกทั้ง NumPy ถูกนำไปใช้พัฒนา Library อื่นอีกด้วย เช่น Matplotlib และ pandas

3) Matplotlib เป็นไลบรารีอันดับหนึ่งในการสร้างกราฟ และทำ Data Visualization โดยที่ Matplotlib สามารถสร้างกราฟได้หลายประเภทเพื่อตอบโจทย์การทำงานของผู้ใช้ให้ได้หลากหลาย เช่น กราฟเส้น แผนภูมิจุดแบบกระจัดกระจาย (Scatter Plot) กราฟแท่งและอื่นๆ

4) Scikit Learn เป็นไลบรารีที่ถูกพัฒนาโดยใช้ไลบลารีสำคัญอื่นๆ ได้แก่ Numpy และ Matplotlib ซึ่งถูกพัฒนาขึ้นเพื่อทำ Machine Learning ทั้งในรูปแบบของ Unsupervised Learning และ Supervised Learning ซึ่ง Scikit Learn จะเน้นในส่วนของการสร้างโมเดลเพื่อทำนายและพยากรณ์ โดยสามารถทำ Spam Detection, Image Recognition, Clustering หรือ Regression ได้ จุดที่ต้องระวังของ Scikit Learn คือ หากต้องการผลที่ถูกต้องหรือมีประสิทธิภาพสูง น่าเชื่อถือ จำเป็นจะต้องมี Input Data ที่ดี

5) Seaborn เป็นไลบรารีสำหรับสร้างกราฟิกทางสถิติในภาษาไพทอน ถูกสร้างขึ้นบน Matplotlib และรวมเข้ากับโครงสร้างข้อมูลของ Pandas เป็นเครื่องมือที่แตกต่างจาก Matplotlib ตรงที่มีความเชี่ยวชาญในการแสดงภาพสถิติ (อรพิน, 2564)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ ดัดแปลง เผยแพร่ หรือนำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ลบข้อความนี้ออก และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5 งานวิจัยที่เกี่ยวข้อง

Agarwal et al. (2019) วิเคราะห์การเปรียบเทียบการจัดกลุ่มร้านค้าในอุตสาหกรรมค้าปลีก ในงานวิจัยนี้มีวัตถุประสงค์เพื่อจัดกลุ่มร้านค้าให้มีประสิทธิภาพมากยิ่งขึ้น ซึ่งจะช่วยให้เข้าใจธุรกิจในร้านค้าของตนเองได้ดีและกำหนดกรอบกลยุทธ์การตลาดที่เน้นร้านค้าเป็นเป้าหมายหลัก โดยมีเป้าหมายเพื่อเพิ่มรายได้สูงสุดมุ่งเน้นไปที่เทคนิคการจัดกลุ่มร้านค้า การเปรียบเทียบค่าเฉลี่ยและใช้เทคนิคการจัดกลุ่มต่างๆ ได้แก่ การทำคลัสเตอร์แบบลำดับชั้น SOM GMM FCM และการทำ K-Means Clustering วิจัยนี้ใช้เทคนิคการคลัสเตอร์ที่แตกต่างกันและแนะนำให้ควรเลือกเทคนิคที่มีผลลัพธ์ที่ดีที่สุด ผลงานวิจัยสรุปได้ว่าไม่มีเทคนิคที่ดีที่สุดสำหรับการจัดกลุ่มร้านค้าเนื่องจากเทคนิคที่แตกต่างกันทำงานได้ดีในข้อมูลที่แตกต่างกัน แต่การทำคลัสเตอร์แบบ K-Means เป็นเทคนิคที่ใช้กันทั่วไปสำหรับการจัดกลุ่มร้านค้าและพบว่าการทำคลัสเตอร์ด้วยเทคนิค K-Means ทำงานได้ดีเมื่อมีข้อมูลหลายมิติในชุดข้อมูลเดียวกัน งานวิจัยของ Agarwal et al. มีความเกี่ยวข้องกับงานวิจัยครั้งนี้ของผู้วิจัยเนื่องจากเป็นการจัดกลุ่มร้านค้าโดยใช้เทคนิค K-Means

Ahmed et al. (2020) การประเมินประสิทธิภาพของอัลกอริทึม K-Means งานวิจัยนี้นำเสนอภาพรวมข้อจำกัดของ K-mean Clustering มุ่งเน้นไปที่ปัญหาการจัดการประเภทข้อมูล พร้อมทั้งอภิปรายข้อจำกัดของการจัดกลุ่มด้วยวิธี K-Means ประเมินประสิทธิภาพของเทคนิค K-Means ในชุดข้อมูลต่าง ๆ วิจัยนี้ให้ภาพรวมที่ครอบคลุมของการทำคลัสเตอร์โดยวิธี K-Means โดยตั้งเป้าหมายที่จะแก้ไขปัญหาของการเริ่มต้นแบบสุ่มของเซ็นทรอยด์และการกำหนดจำนวนคลัสเตอร์ล่วงหน้าโดยมีการใช้เอกสารในการประเมินประสิทธิภาพ 6 ชุด และการวิเคราะห์เชิงทดลองในชุดข้อมูลที่หลากหลายเพื่อตรวจสอบประสิทธิภาพของอัลกอริทึมคลัสเตอร์แบบ K-Means งานวิจัยสรุปได้ว่ายังว่าไม่มีทางออกสำหรับปัญหาเกี่ยวกับการเริ่มต้นแบบสุ่มของเซ็นทรอยด์และการกำหนดจำนวนคลัสเตอร์ล่วงหน้า และการวิจัยในอนาคตควรมุ่งเน้นไปที่การพัฒนาต้นแบบสุ่มของเซ็นทรอยด์ และการกำหนดจำนวนคลัสเตอร์ ซึ่งงานวิจัยนี้ช่วยทำให้ผู้วิจัยทราบว่าปัญหาของการจัดกลุ่มด้วยวิธี K-Means และทราบข้อจำกัดของการจัดกลุ่มด้วยเทคนิค K-Means

Hemalatha et al. (2009) การแบ่งส่วนการตลาดของผู้ซื้อเสื้อผ้าอินเดียตามลักษณะของร้านค้า งานวิจัยนี้นำเสนอว่า ผู้ค้าปลีกแสวงหาความหลากหลายโดยใช้ประโยชน์จากส่วนผสมทางการตลาดค้าปลีกตามคุณลักษณะของร้านค้าโดยได้ทำการจัดกลุ่มโดยวิธี K-Means และระบุกลุ่มนักช็อปชาวอินเดีย หลังจากที่จัดกลุ่มได้ 3 กลุ่มแล้วได้มีการเปรียบเทียบความแตกต่างของแต่ละกลุ่มโดยใช้ ANOVA เพื่อดูความแปรปรวนผ่านโปรแกรม SPSS หลังจากการวิเคราะห์แสดงให้เห็นว่ามีความแตกต่างกันอย่างมีนัยสำคัญ และนำผลลัพธ์ที่ได้ไปปรับกลยุทธ์ยอดขาย งานวิจัยของ Hemalatha et al. เกี่ยวข้องกับงานวิจัยครั้งนี้ของผู้วิจัยเนื่องจากมีการแบ่งกลุ่มโดยวิธี K-Means และเปรียบเทียบความแตกต่างของแต่ละกลุ่มโดยใช้สถิติอนุมานผ่านโปรแกรม SPSS

Ogul et al. (2023) การแบ่งส่วนร้านค้าโดยใช้วิธีการเรียนรู้ของเครื่อง : กรณีศึกษาภายใต้บริษัทซูเปอร์มาร์เก็ต วิจัยนี้มุ่งเน้นไปที่แบ่งส่วนร้านค้าในอุตสาหกรรมค้าปลีกเพื่อรักษาความพึงพอใจของลูกค้าและการเติบโตตามเป้าหมาย การจัดกลุ่มร้านค้าโดย K–Mean Clustering ใช้เพื่อค้นหาการจัดกลุ่มที่ดีที่สุดของร้านค้าตามข้อมูลที่มีอยู่ งานวิจัยนี้ได้จัดระบบการจัดกลุ่มที่สมบูรณ์มีส่วนช่วยในการปรับปรุงประสิทธิภาพของร้านค้าการทำความเข้าใจพฤติกรรมของลูกค้าและการตัดสินใจเชิงกลยุทธ์ การใช้ K-Means เพื่อแบ่งกลุ่มร้านค้าส่งผลให้มี 12 กลุ่มร้านที่แตกต่างกันและ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารนี้ทุกครั้งที่มีการนำไปใช้

จัดกลุ่มร้านค้าตามความคล้ายคลึงกัน งานวิจัยของ Ogul et al. เกี่ยวข้องกับงานวิจัยครั้งนี้ของผู้วิจัยเนื่องจากเป็นการจัดกลุ่มร้านค้าโดยใช้เทคนิค K-Means

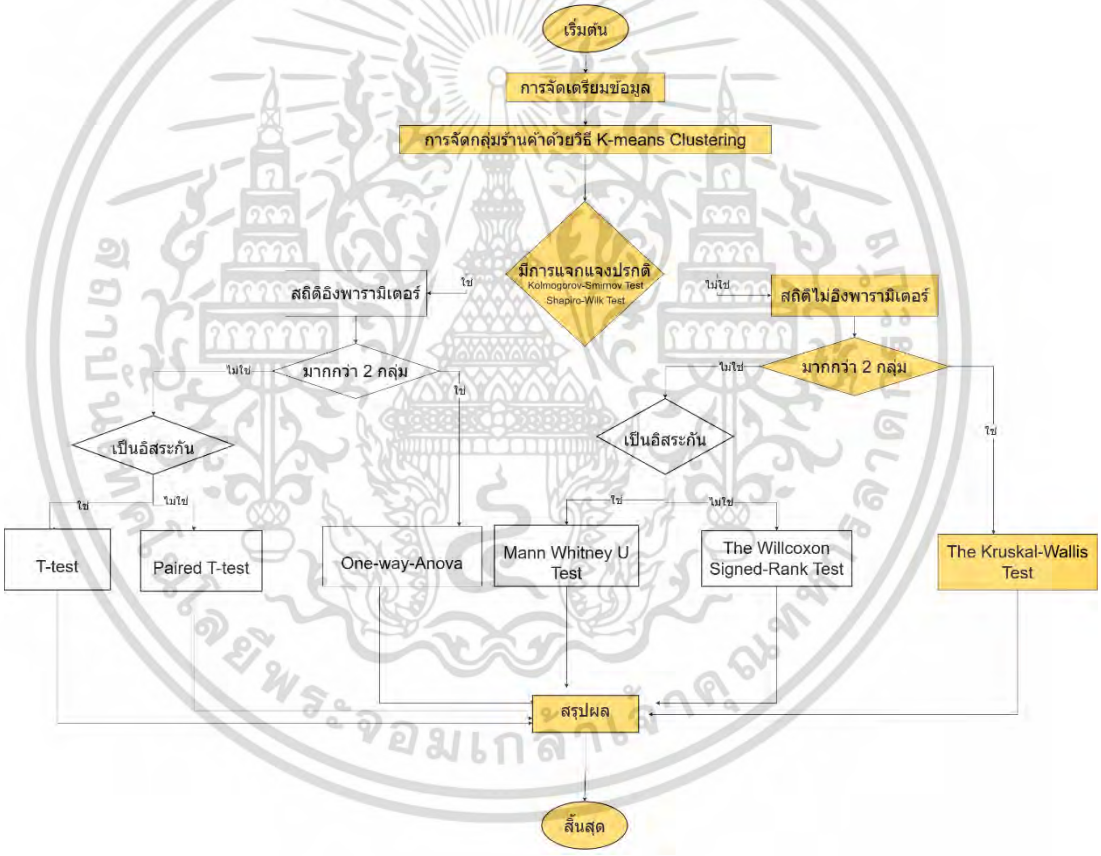
Pradana et al. (2021) การปรับปรุงกลยุทธ์สูงสุดในการแบ่งส่วนลูกค้าห้างสรรพสินค้าโดยใช้การจัดกลุ่ม K- mean วิจัยนี้นำเสนอการแบ่งกลุ่มลูกค้าโดยใช้ K-mean Clustering กลยุทธ์การตลาดห้างสรรพสินค้าเป็นตัวช่วยในการกำหนดเป้าหมายลูกค้าที่ทำกำไรในห้างสรรพสินค้า การทำความเข้าใจความต้องการของลูกค้าสำคัญสำหรับการจัดการห้างสรรพสินค้า โดยงานวิจัยนี้แบ่งลูกค้าออกเป็น 5 กลุ่ม ตามรายได้และการใช้จ่าย ระบุพฤติกรรมของลูกค้าตามคะแนนรายได้และการใช้จ่าย โดยระบุกลุ่มที่แตกต่างกันตามระดับรายได้และค่าใช้จ่าย เช่น กลุ่มที่ 1 คือกลุ่มลูกค้าที่มีรายได้สูงและคะแนนการใช้จ่ายสูง ซึ่งงานวิจัยนี้กำหนดจำนวนกลุ่มโดยวิธีจุดหักข้อศอก ซึ่งสอดคล้องกับงานวิจัยที่ผู้วิจัยได้ทำในครั้งนี้โดยใช้วิธีจุดหักข้อศอกในกำหนดจำนวนกลุ่มและจัดกลุ่มโดยวิธี K-Means เพื่อนำไปปรับกลยุทธ์ทางการตลาด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีดำเนินการวิจัย

ในงานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อจัดกลุ่มร้านค้าเบเกอรี่ จากการทบทวนวรรณกรรมในบทที่ 2 ผู้วิจัยนำมาสรุปได้ขั้นตอนดำเนินงานวิจัยดังภาพ 3.1 และผู้วิจัยได้เลือกการจัดกลุ่มร้านค้าโดยใช้เทคนิคการจัดกลุ่มแบบ K-means เนื่องจากต้องการแบ่งกลุ่มร้านค้าที่ลักษณะคล้ายคลึงกันไว้ด้วยกันจากนั้นวิเคราะห์ความแตกต่างโดยใช้สถิติพรรณนาและสถิติอนุมาน โดยมีรายละเอียดแยกเป็นหัวข้อย่อย ๆ ดังนี้

3.1 ขั้นตอนการดำเนินงานวิจัย



รูปที่ 3.1 ขั้นตอนการดำเนินงานวิจัย

ขั้นตอนการดำเนินงานวิจัยจะแสดงให้เห็นดังรูปที่ 3.1 โดยเริ่มจากการจัดเตรียมข้อมูล เมื่อจัดเตรียมข้อมูลแล้วจะนำข้อมูลที่ได้ไปทำการจัดกลุ่มโดยเทคนิควิธี K-means Clustering และทดสอบการแจกแจงของข้อมูลว่ามีการแจกแจงปรกติหรือไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยใช้การทดสอบคอลโมโกรอฟ-สมีร์นอฟ (Kolmogorov-Smirnov Test) สำหรับตัวอย่างขนาดใหญ่และการทดสอบของการทดสอบชาพิโร-วิลก์ (Shapiro-Wilk Test) สำหรับขนาดตัวอย่างที่น้อยกว่าหรือเท่ากับ 50 ถ้าหากมีการแจกแจงปรกติจะใช้สถิติอิงพารามิเตอร์ กรณีจัดกลุ่มได้ 2 กลุ่ม ที่ข้อมูลเป็นอิสระกันจะเลือกใช้ t-test แต่หากข้อมูลไม่เป็นอิสระกันจะเลือกใช้ paired t-test กรณีจัดกลุ่มได้มากกว่า 2 กลุ่มขึ้นไป จะเลือกใช้การวิเคราะห์ความแปรปรวนทางเดียว (One-way-ANOVA) แต่หากไม่เป็นไปตามข้อตกลงเบื้องต้น คือ ไม่มีการแจกแจงปรกติจะเลือกใช้สถิติไม่อิงพารามิเตอร์ กรณีจัดกลุ่มได้ 2 กลุ่มที่เป็นอิสระกัน จะเลือกใช้การทดสอบของแมนต์-วิตนีย์ (Mann Whitney U Test) แต่หากข้อมูลไม่เป็นอิสระกันจะเลือกใช้การทดสอบของวิลค็อกสัน (The Willcoxon Signed-Ranks Test) กรณีจัดกลุ่มได้มากกว่า 2 กลุ่มขึ้นไป จะเลือกใช้การทดสอบของคลัสคาล-วอลลิส (The Kruskal-Wallis Test) จากนั้นทำการสรุปผลการวิจัย เพื่ออธิบายคุณลักษณะของแต่ละกลุ่มร้านค้าว่าเหมือนหรือแตกต่างกัน

3.2 การจัดเตรียมข้อมูล

ในงานวิจัยครั้งนี้มีข้อมูลทั้งหมด 1 ปี ซึ่งมีตัวแปรทั้งหมด 17 ตัวแปร และร้านค้าจำนวนทั้งสิ้น 157 ร้านค้า ร้านเบเกอรี่กรณีศึกษามีเบเกอรี่ทั้งหมด 5 ชนิดประกอบด้วยชนิด A B C D E มีรายละเอียดข้อมูลดังตารางที่ 3.1

ตารางที่ 3.1 ตัวแปรและความหมายของข้อมูล

ตัวแปร	ความหมาย	ประเภท
Store	ร้านค้า	Num
Type	ประเภทร้านค้า	Category
Region	พื้นที่ตั้ง	Category
Rank	ระดับของร้านค้า	Category
Hall	ลักษณะที่ตั้งของร้านค้า	Category
Produce	ผลิตสินค้าได้เองไหม	Category
Score1	คะแนนการจัดร้านค้า	Num
Score2	คะแนนการประเมินตนเอง	Num
Score3	คะแนนการประเมินจากผู้จัดการร้านค้า	Num
Score4	ผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านค้า	Num
Score5	คะแนนความน่าเข้ามาซื้อของ	Num
Score6	คะแนนที่ตั้งร้านค้า	Num
A_Percentage	เปอร์เซ็นต์การขายของสินค้ากลุ่ม A	Num
B_Percentage	เปอร์เซ็นต์การขายของสินค้ากลุ่ม B	Num
C_Percentage	เปอร์เซ็นต์การขายของสินค้ากลุ่ม C	Num
D_Percentage	เปอร์เซ็นต์การขายของสินค้ากลุ่ม D	Num

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ตัวแปรและความหมายของข้อมูล (ต่อ)

ตัวแปร	ความหมาย	ประเภท
E_Percentage	เปอร์เซ็นต์การขายของสินค้ากลุ่ม E	Num

จากตารางที่ 3.1 ผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านเกิดจากการนำคะแนนการประเมินตนเองลบกับคะแนนการประเมินจากผู้จัดการร้าน เนื่องจากคะแนนการประเมินตนเองและคะแนนการประเมินจากผู้จัดการร้านของร้านที่มีประสิทธิภาพในการบริหารควรจะมีค่าใกล้เคียงกันและผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านค่ายิ่งน้อยยิ่งสรุปได้ได้ว่าเป็นร้านที่มีการบริหารดี

ตารางที่ 3.2 ตัวอย่างและรายละเอียดข้อมูล

ตัวแปร	ความหมาย	ตัวอย่างและรายละเอียดข้อมูล
Type	ประเภทร้านค้า	Tiny จำนวน 30 ร้านค้า Small จำนวน 65 ร้านค้า Medium จำนวน 35 ร้านค้า Large จำนวน 7 ร้านค้า Huge จำนวน 14 ร้านค้า Gigantic จำนวน 6 ร้านค้า
Region	พื้นที่ตั้ง	BKK จำนวน 95 ร้านค้า UPC จำนวน 62 ร้านค้า
Rank	ระดับของร้านค้า	Standard จำนวน 2 ร้านค้า Gold จำนวน 138 ร้านค้า Premium จำนวน 17 ร้านค้า
Hall	ลักษณะที่ตั้งของร้านค้า	Small จำนวน 7 ร้านค้า Medium จำนวน 35 ร้านค้า Huge จำนวน 65 ร้านค้า Gigantic จำนวน 50 ร้านค้า
Produce	ผลิตสินค้าได้เอง	Yes จำนวน 134 ร้านค้า No จำนวน 23 ร้านค้า

3.2.1 การทำความสะอาดข้อมูล (Data Cleaning)

เป็นกระบวนการตรวจสอบและแก้ไขข้อมูลที่ไม่ถูกต้องหรือหาข้อมูลสูญหาย ซึ่งเป็นหลักสำคัญของการวิเคราะห์ข้อมูลเมื่อพบข้อมูลที่ไม่ถูกต้องต้องมีการปรับปรุงหรือลบทิ้งไป

ตารางที่ 3.3 ชุดคำสั่งสำหรับหาข้อมูลสูญหาย

```
df.isnull().sum()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 3.3 เป็นการแสดงชุดคำสั่งสำหรับการหาข้อมูลสูญหาย เนื่องจากข้อมูลที่ได้รับมานั้นไม่มีข้อมูลสูญหายแสดงดังรูปที่ 3.2 ดังนั้นจึงนำข้อมูลไปใช้ในขั้นตอนถัดไปได้

```
df.isnull().sum()
```

```
Store Code        0
Type              0
Rank              0
Region            0
Hall              0
Produce           0
Score 1           0
Score 2           0
Score 3           0
Score 4           0
Score 5           0
A_Percentage      0
B_Percentage      0
C_Percentage      0
D_Percentage      0
E_Percentage      0
Score 6           0
```

รูปที่ 3.2 รูปแสดงการค้นหาข้อมูลสูญหาย

3.2.2 การแปลงข้อมูล (Data Transformation)

เป็นการเปลี่ยนข้อมูลให้เหมาะสมและถูกต้องในรูปแบบคะแนนมาตรฐาน ให้ข้อมูลอยู่ในมาตรฐานเดียวกัน ด้วยวิธี Standard Scaler โดยสามารถคำนวณได้ดังนี้

$$x_{stand} = \frac{x - \bar{x}}{SD(x)} \tag{3.1}$$

ตารางที่ 3.4 ชุดคำสั่งสำหรับการแปลงข้อมูลให้เป็นมาตรฐานเดียวกัน

from sklearn.preprocessing import StandardScaler
numeric_columns ['Score1', 'Score 2', 'Score3', 'Score4', 'Score5', 'SCORE6', 'A_Percentage', 'B_Percentage', 'C_Percentage', 'D_Percentage', 'E_Percentage']
scaler = StandardScaler()
X = scaler.fit_transform(df[numeric_columns])

จากตารางที่ 3.4 ได้ทำการแปลงข้อมูลให้อยู่ในมาตรฐานเดียวกันโดยเลือกเฉพาะตัวแปรที่ไม่ใช่ตัวแปรหุ่น โดยจะปรับจุดทศนิยมของแต่ละสดมภ์ให้มีมาตรฐานเดียวกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.3 การคัดเลือกข้อมูล (Data Management)

การคัดเลือกข้อมูล คือการคัดเลือกตัวแปรที่เหมาะสมสำหรับการจัดกลุ่มด้วยวิธี K-Means ซึ่งผู้วิจัยได้คัดเลือกตัวแปรในการจัดกลุ่มทั้งหมด 11 ตัว ซึ่งเป็นข้อมูลเชิงปริมาณทั้งหมด และเป็นข้อมูลในมาตรวัดอัตราส่วนและมาตรวัดอัตรภาคตามข้อกำหนดการจัดกลุ่มด้วยวิธี K-Means แสดงดังตารางที่ 3.5

ตารางที่ 3.5 การคัดเลือกข้อมูลในการจัดกลุ่ม

ชื่อตัวแปร	ความหมาย
Score1	คะแนนการจัดร้านค้า
Score2	คะแนนการประเมินตนเอง
Score3	คะแนนการประเมินจากผู้จัดการร้านค้า
Score4	ผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านค้า
Score5	คะแนนความน่าเข้ามาซื้อของ
Score6	คะแนนที่ตั้งร้านค้า
A_Percentage	เปอร์เซ็นต์การขายของสินค้ากลุ่ม A
B_Percentage	เปอร์เซ็นต์การขายของสินค้ากลุ่ม B
C_Percentage	เปอร์เซ็นต์การขายของสินค้ากลุ่ม C
D_Percentage	เปอร์เซ็นต์การขายของสินค้ากลุ่ม D
E_Percentage	เปอร์เซ็นต์การขายของสินค้ากลุ่ม E

3.3 เครื่องมือที่ใช้ในดำเนินงานวิจัย

3.3.1 ซอฟต์แวร์ที่ใช้ในงานวิจัย (Software)

1) Google Colab เป็นเครื่องมือสำหรับการจัดกลุ่มของร้านค้าด้วยเทคนิค K-Means

2) IBM SPSS 26 เป็นโปรแกรมสำเร็จรูปที่ใช้วิเคราะห์ข้อมูลทางสถิติ

3.3.2 คำสั่งไลบรารีที่ใช้ในงานวิจัย

การจัดกลุ่มร้านค้าในงานวิจัยครั้งนี้ได้ใช้ภาษาโปรแกรมไพทอนในการจัดกลุ่ม ซึ่งไลบรารีที่ใช้แสดงดังตารางที่ 3.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.6 คำสั่งไลบรารีที่ใช้ในงานวิจัย

Pandas	เป็นไลบรารีการจัดการข้อมูล (Data Wrangling/ Data Cleaning) และการวิเคราะห์ข้อมูล (Data Analysis) ซึ่งถูกพัฒนามาเพื่อทำงานกับข้อมูลขนาดใหญ่
Matplotlib	Matplotlib เป็นไลบรารีอันดับหนึ่งในการสร้างกราฟ และทำ Data Visualization โดยที่ Matplotlib สามารถสร้างกราฟได้หลายประเภทเพื่อตอบโจทย์การทำงานของผู้ใช้ให้ได้หลากหลาย เช่น กราฟเส้น แผนภูมิจุดแบบกระจัดกระจาย (Scatter Plot), กราฟแท่งและอื่นๆ
Scikit Learn	จะเน้นในส่วนของการสร้างโมเดลเพื่อทำนายและพยากรณ์ โดยสามารถทำ Spam Detection, Image Recognition, Clustering หรือ Regression ได้ จุดที่ต้องระวังของ Scikit Learn คือ หากต้องการผลที่ถูกต้องหรือมีประสิทธิภาพสูงน่าเชื่อถือ จำเป็นจะต้องมี Input Data ที่ดี
Numpy	โดดเด่นในด้านการคำนวณ และการทำงานกับตัวเลข นอกจากนี้ NumPy ยังมีความสามารถสำคัญในการสร้าง Array NumPy ถูกนำไปใช้พัฒนา Library อื่นอีกด้วย เช่น Matplotlib และ pandas

3.4 การจัดกลุ่มด้วยเทคนิค K-Means

3.4.1 การวิเคราะห์หาจำนวนกลุ่มที่เหมาะสม

หลังจากจัดการกับข้อมูลเรียบร้อยแล้ว จะวิเคราะห์จำนวนกลุ่มที่เหมาะสมโดยเลือกใช้วิธี Elbow ซึ่งเป็นการพิจารณาข้อผิดพลาด การคำนวณหาจำนวนกลุ่มที่เหมาะสมที่สุดสามารถหาได้จากระยะทางที่ไกลที่สุดนับจากเส้นตรงระหว่างสองจุดกับเส้นโค้งมุมที่มีลักษณะของจุดหักข้อศอก รายละเอียดดังตารางที่ 3.7

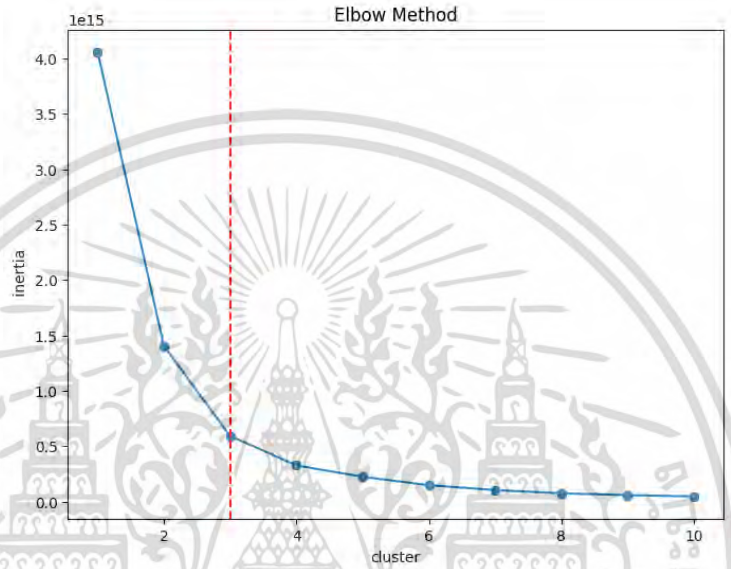
ตารางที่ 3.7 ชุดคำสั่งของ Elbow Method พร้อมสร้างกราฟ

```
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
X = df[['Score1', 'Score 2', 'Score3', 'Score4', 'Score5', 'SCORE6', 'A_Percentage',
'B_Percentage', 'C_Percentage', 'D_Percentage', 'E_Percentage',]].values
inertia = []
plt.figure(figsize=(8, 6))
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=0)
    kmeans.fit(X)
    inertia.append(kmeans.inertia_)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.7 ชุดคำสั่งของ Elbow Method พร้อมสร้างกราฟ (ต่อ)

plt.plot(range(1, 11), inertia, marker='o')
plt.xlabel('Number of clusters')
plt.ylabel('Inertia')
plt.title('Elbow Method')
plt.show()



รูปที่ 3.3 รูปแสดง Elblow หาค่า K ที่เหมาะสม

จากรูปที่ 3.3 จะพบว่าจำนวนที่เหมาะสมในการกลุ่มคือ 3 กลุ่ม (K = 3) เนื่องจากมองจากภาพจุดหักข้อศอกที่เหมาะสมที่สุดอยู่ที่ 3

3.4.2 การจัดกลุ่มด้วยวิธี K-Means

หลังจากจัดการกับข้อมูลเรียบร้อยแล้ว จะวิเคราะห์จำนวนกลุ่มที่เหมาะสมโดยเลือกใช้วิธี Elbow จำนวนกลุ่มที่เหมาะสม คือ 3 กลุ่ม จึงทำการจัดกลุ่มของร้านค้าเป็น 3 กลุ่ม โดยใส่ฟีเจอร์ด้วยตัวแปรที่คัดเลือกไว้ทั้ง 11 ตัวแปร ดังตารางที่ 3.8

ตารางที่ 3.8 ชุดคำสั่งการทำ K-means Clustering

from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
num_clusters = 3
kmeans = KMeans(n_clusters=num_clusters, random_state=42)
df['cluster'] = kmeans.fit_predict(X)
plt.figure(figsize=(16, 10))
for cluster_id in range(num_clusters):

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.8 ชุดคำสั่งการทำ K-means Clustering (ต่อ)

cluster_data = df[df['cluster'] == cluster_id]
plt.scatter(cluster_data['Score D'], cluster_data['Score E], label=df'Cluster {cluster_id}')
plt.xlabel('Score1')
plt.ylabel('Score 2')
plt.title('Scatter Plot with Store Code')
plt.legend()
plt.show()

3.5 การวิเคราะห์ข้อมูล

3.5.1 สถิติเชิงพรรณนา ใช้วิเคราะห์ข้อมูลแต่ละตัวแปรว่ามีค่าเฉลี่ย ร้อยละ ในแต่ละกลุ่มร้านค้าที่ได้จากการจัดกลุ่ม และดูว่าแต่ละกลุ่มเหมือนหรือแตกต่างกัน

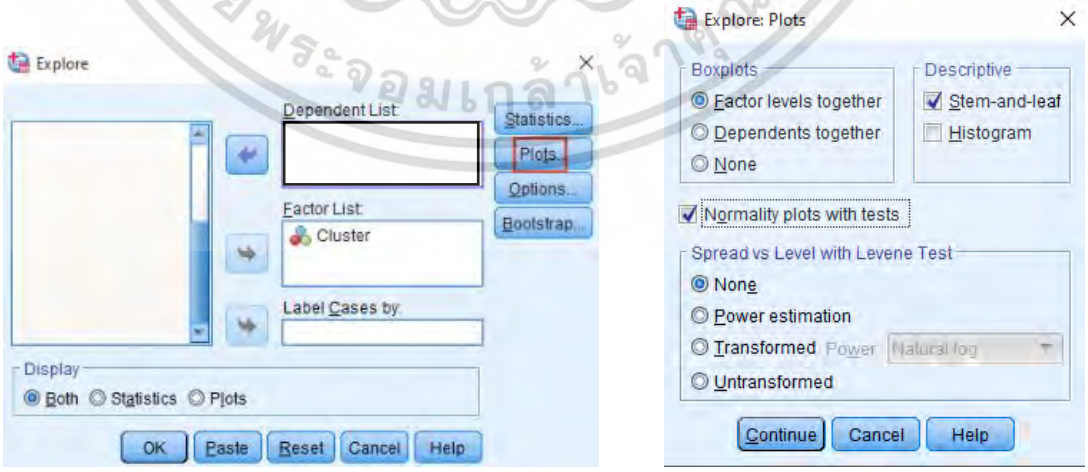
3.5.2 สถิติเชิงอนุมาน ใช้ในการวิเคราะห์ความแตกต่างระหว่างกลุ่มที่ได้จากการจัดกลุ่มด้วยโปรแกรมสำเร็จรูป IBM SPSS มีขั้นตอนดังต่อไปนี้

3.5.2.1 การทดสอบข้อจำกัดเบื้องต้น

เพื่อทดสอบว่าข้อมูลมีการแจกแจงปรกติหรือไม่ ถ้าหากข้อมูลมีการแจกแจงปรกติจะใช้สถิติอิงพารามิเตอร์ แต่ถ้าไม่มีการแจกแจงปรกติจะใช้สถิติไม่อิงพารามิเตอร์ มีขั้นตอนการวิเคราะห์ ดังนี้

1) Analyze
2) Descriptive
3) Explore

เลือกตัวแปรตามใส่ไว้ใน Dependent List และตัวแปรอิสระที่เป็นการแบ่งกลุ่มไว้ใน Factor List จากนั้นคลิกที่ Plots เลือก Normality Test ดังรูปที่ 3.4



รูปที่ 3.4 วิเคราะห์การแจกแจงปรกติด้วยโปรแกรม SPSS

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Tests of Normality

	Cluster	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
SumofNet_Total	0	.092	130	.009	.961	130	.001
	1	.159	23	.137	.882	23	.011
	2	.170	4	.	.988	4	.946

a. Lilliefors Significance Correction

รูปที่ 3.5 ผลการวิเคราะห์การแจกแจงปกรกติ

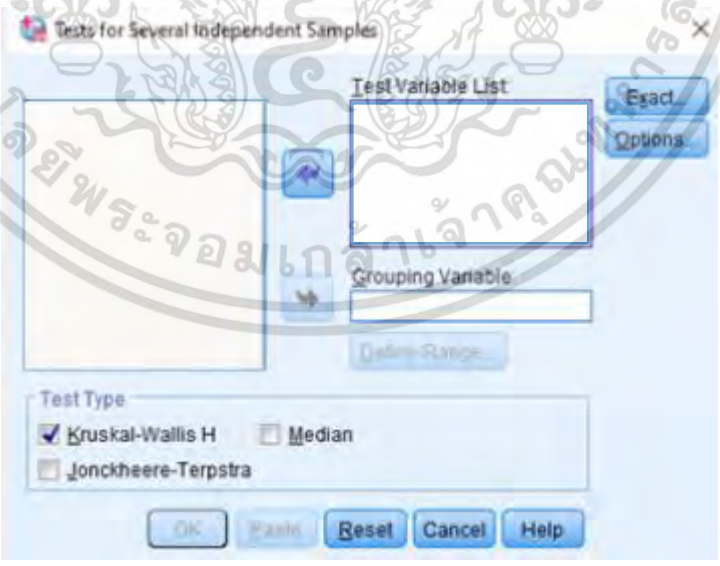
จากรูปที่ 3.5 จะพบว่าในแต่ละกลุ่มมีค่า Sig. น้อยกว่าระดับนัยสำคัญ 0.05 ดังนั้นจึงปฏิเสธสมมติฐานหลัก จึงสรุปได้ว่าแต่ละกลุ่มร้านค้ามีการแจกแจงไม่ปรกติ จึงเลือกใช้สถิติไม่อิงพารามิเตอร์

3.5.2.2 การทดสอบสมมติฐาน

เนื่องจากมีการจัดกลุ่มได้ 3 กลุ่ม และข้อมูลมีการแจกแจงที่ไม่ปรกติจึงต้องเลือกใช้การทดสอบของคลัสคาล-วอลลิส (The Kruskal-Wallis Test) เพื่อตรวจสอบว่าแต่ละกลุ่มมีความแตกต่างกันหรือไม่ มีขั้นตอนการทดสอบดังนี้

1) Analyze
2) Nonparametric Test
3) Legacy Dialogs
4) K Independent Sample...

จะได้ดังรูปที่ 3.6 จากนั้นกด OK



รูปที่ 3.6 หน้าต่างการเลือกข้อมูลที่ใช้ทดสอบความแตกต่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการดำงานวิจัยเริ่มจากการจัดเตรียมข้อมูล เมื่อจัดเตรียมข้อมูลแล้วจะนำข้อมูลที่ได้ไปทำการจัดกลุ่มโดยเทคนิควิธี K-means Clustering และทดสอบการแจกแจงปรกติของข้อมูล พบว่าไม่มีการแจกแจงปรกติ และจาก Elbow Method กำหนดจำนวนกลุ่มได้ 3 กลุ่ม จึงใช้การทดสอบของของคลัสคาล-วอลลิส ในการวิเคราะห์ความแตกต่างระหว่างกลุ่ม

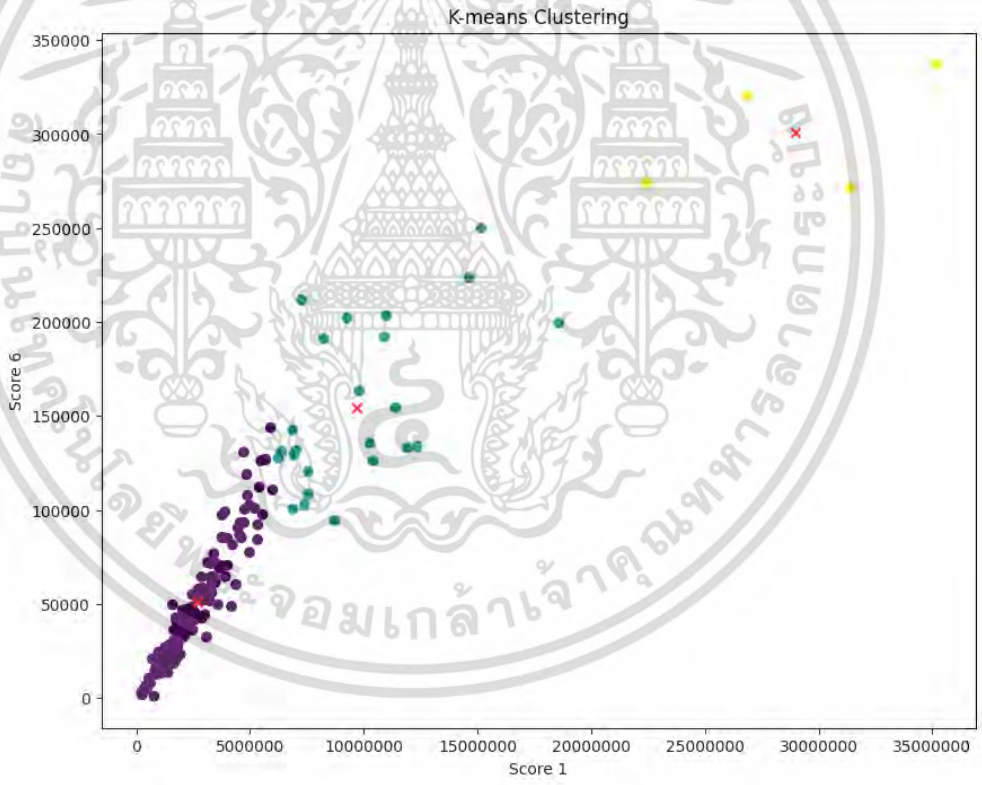
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการวิจัยและการอภิปราย

ในบทนี้จะกล่าวถึงผลของการวิเคราะห์จากการจัดกลุ่มร้านค้า โดยดูผลว่าเหมือนหรือต่างกันจากสถิติพรรณนาที่นำเสนอข้อมูลในรูปแบบกราฟ และผลการวิเคราะห์ความแตกต่างระหว่างกลุ่มโดยใช้สถิติอนุมาน เพื่อให้เกิดความเข้าใจถึงพฤติกรรมและลักษณะของแต่ละกลุ่มร้านค้าดังหัวข้อต่อไปนี้

4.1 ผลการวิเคราะห์การจัดกลุ่ม (Clustering Analysis)

โดยงานวิจัยครั้งนี้อาศัยเทคนิคการจัดกลุ่มข้อมูลด้วยวิธี K-Means โดยจะจัดกลุ่มร้านค้าตามที่มีลักษณะคล้ายคลึงกันไว้ด้วยกัน โดยผู้วิจัยมีการกำหนดกลุ่มด้วยวิธี Elbow เพื่อหาจำนวนกลุ่มที่เหมาะสมที่สุดและได้นำตัวแปรเข้าไปจัดกลุ่มทั้งหมด 11 ตัว ซึ่งผู้วิจัยนำเสนอผลลัพธ์ของการจัดกลุ่ม ดังรูปที่ 4.1



รูปที่ 4.1 แสดงผลลัพธ์จากการจัดกลุ่มร้านค้า

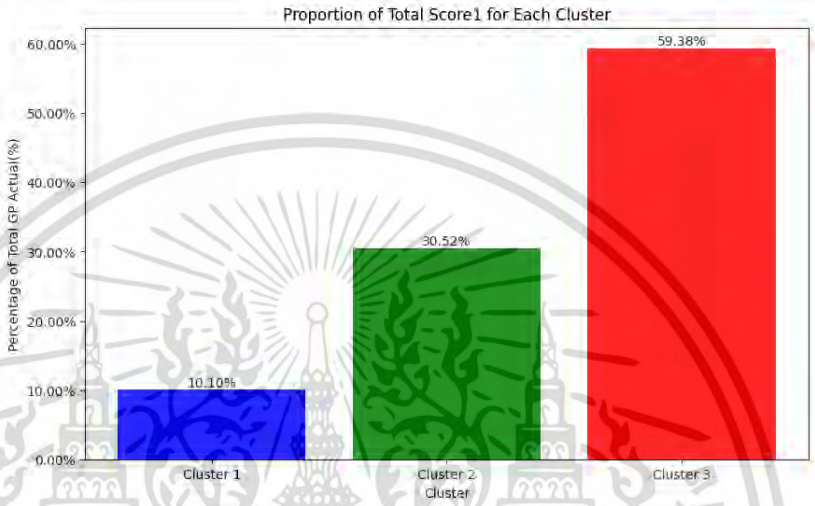
จากรูปที่ 4.1 จะเห็นได้ว่าหลังจากการจัดกลุ่มร้านค้าเป็น 3 กลุ่ม กำหนดให้สีม่วง ฟ้า เหลือง คือกลุ่มที่ 1,2,3 ตามลำดับ จากจำนวนร้านค้าทั้งหมด 157 ร้านค้า ในกลุ่มที่ 1 มีจำนวนทั้งหมด 129 ร้านค้า ในกลุ่มที่ 2 มีจำนวนทั้งหมด 24 ร้านค้า และกลุ่มที่ 3 จำนวนทั้งหมด 4 ร้านค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 ผลการวิเคราะห์ข้อมูลด้วยสถิติพรรณนา (Descriptive Statistics)

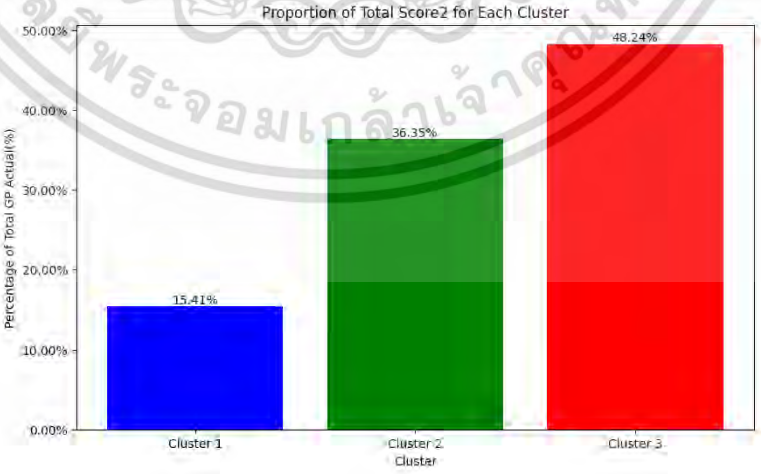
4.2.1 ผลการวิเคราะห์ข้อมูลด้วยสถิติพรรณนาเปรียบเทียบทั้ง 3 กลุ่ม

ผลการวิเคราะห์ข้อมูลด้วยสถิติพรรณนาเปรียบเทียบทั้ง 3 กลุ่ม จากตัวแปร 6 ตัว ได้แก่ คะแนนการประเมินตนเอง คะแนนความน่าเข้ามาซื้อของคะแนนการจัดร้านค้า คะแนนที่ตั้งร้านค้า คะแนนการประเมินจากผู้จัดการร้านค้าและผลต่างของการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านค้า ได้ผลลัพธ์ดังนี้



รูปที่ 4.2 สัดส่วนคะแนนการจัดร้านค้า

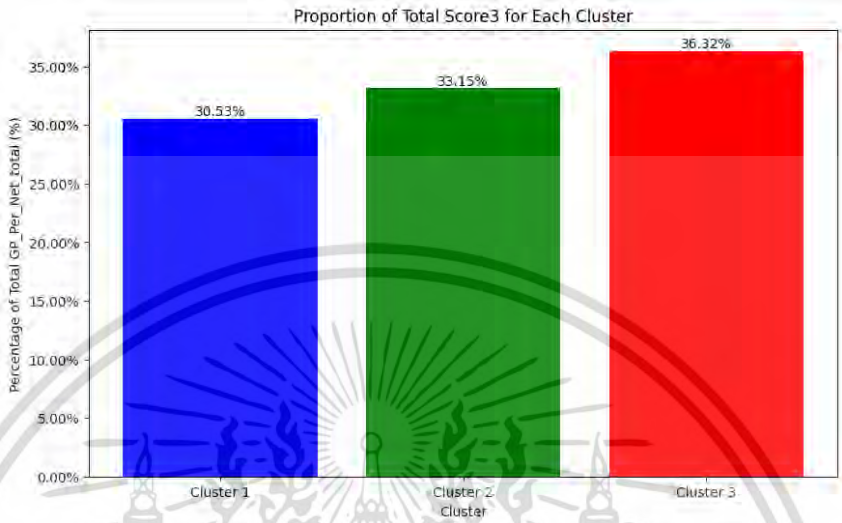
จากรูปที่ 4.2 สัดส่วนคะแนนการจัดร้านค้าทั้งหมดของทุกร้านค้า พบว่าร้านค้ากลุ่มที่ 1 มีคะแนนการจัดร้านค้าทั้งหมดคิดเป็น 10.10 เปอร์เซ็นต์ ร้านค้ากลุ่มที่ 2 คิดเป็น 30.52 เปอร์เซ็นต์ และร้านค้ากลุ่มที่ 3 คิดเป็น 59.38 เปอร์เซ็นต์



รูปที่ 4.3 สัดส่วนคะแนนการประเมินตนเอง

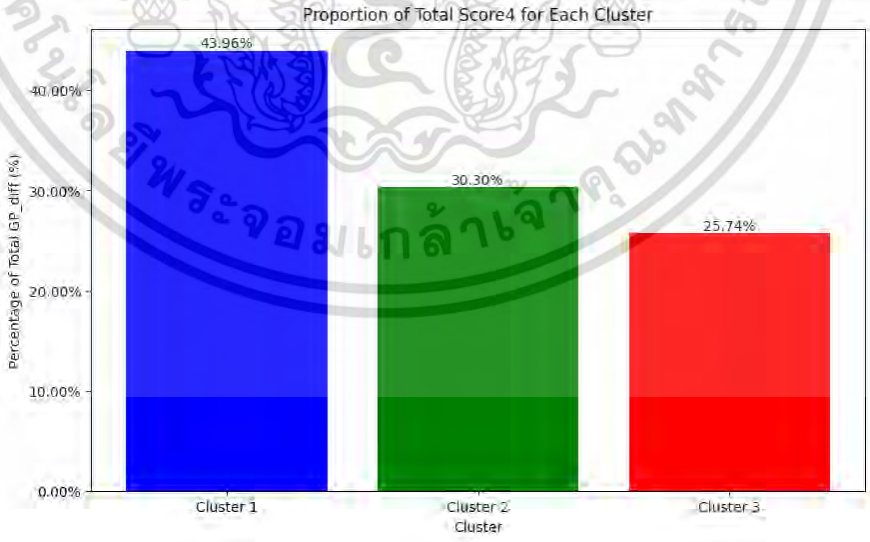
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.3 สัดส่วนคะแนนการประเมินตนเองของทุกร้านค้า พบว่าร้านค้ากลุ่มที่ 1 มีคะแนนการประเมินตนเองคิดเป็น 15.41 เปอร์เซ็นต์ ร้านค้ากลุ่มที่ 2 คิดเป็น 36.35 เปอร์เซ็นต์ และร้านค้ากลุ่มที่ 3 คิดเป็น 48.24 เปอร์เซ็นต์



รูปที่ 4.4 สัดส่วนคะแนนการประเมินจากผู้จัดการร้านค้า

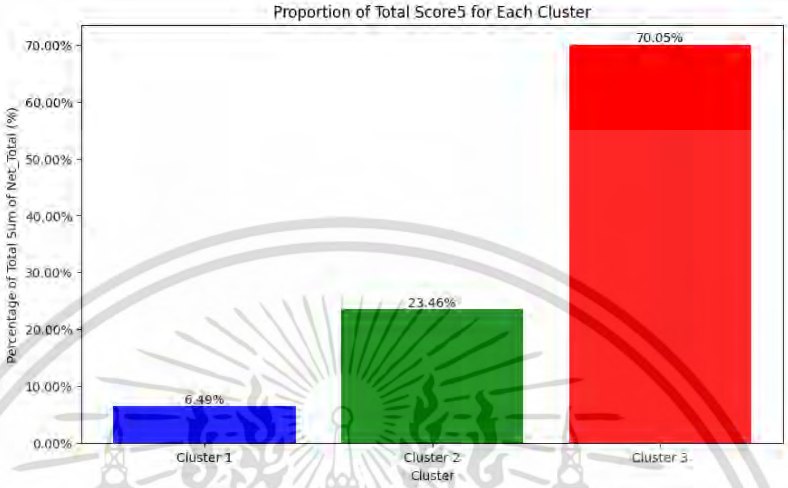
จากรูปที่ 4.4 สัดส่วนคะแนนการประเมินจากผู้จัดการร้านค้าของทุกร้านค้า พบว่าร้านค้ากลุ่มที่ 1 มีคะแนนการประเมินจากผู้จัดการร้านค้าคิดเป็น 30.53 เปอร์เซ็นต์ ร้านค้ากลุ่มที่ 2 คิดเป็น 33.15 เปอร์เซ็นต์ และร้านค้ากลุ่มที่ 3 คิดเป็น 36.32 เปอร์เซ็นต์



รูปที่ 4.5 สัดส่วนผลต่างของการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านค้า

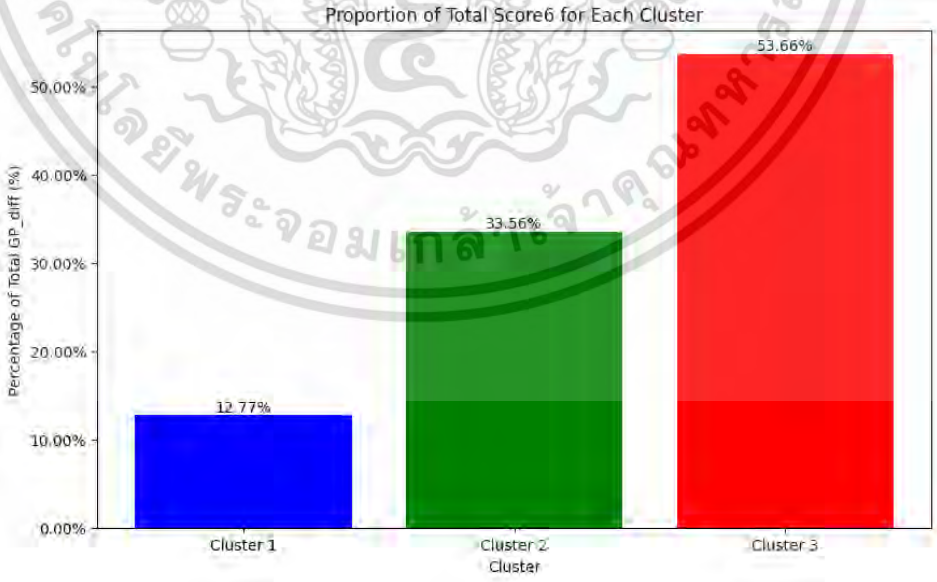
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.5 สัดส่วนผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านค้าของทุกร้านค้า พบว่าร้านค้ากลุ่มที่ 1 มีผลต่างของการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านค้า คิดเป็น 43.96 เปอร์เซ็นต์ ร้านค้ากลุ่มที่ 2 คิดเป็น 30.30 เปอร์เซ็นต์และร้านค้ากลุ่มที่ 3 คิดเป็น 59.38 เปอร์เซ็นต์



รูปที่ 4.6 สัดส่วนคะแนนความน่าเข้ามาซื้อของ

จากรูปที่ 4.6 สัดส่วนคะแนนความน่าเข้ามาซื้อของของทุกร้านค้า พบว่าร้านค้ากลุ่มที่ 1 มียอดขายรวมคิดเป็น 6.49 เปอร์เซ็นต์ ร้านค้ากลุ่มที่ 2 คิดเป็น 23.46 เปอร์เซ็นต์ และร้านค้ากลุ่มที่ 3 คิดเป็น 70.05 เปอร์เซ็นต์



รูปที่ 4.7 คะแนนที่ตั้งร้านค้า

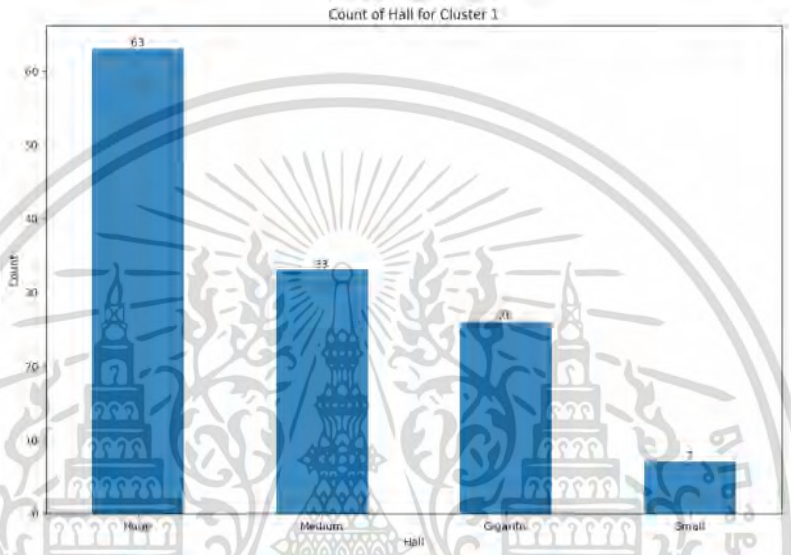
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.7 สัดส่วนคะแนนที่ตั้งร้านค้าของทุกร้านค้า พบว่าร้านค้ากลุ่มที่ 1 มีคะแนนที่ตั้งร้านค้าคิดเป็น 12.77 เปอร์เซ็นต์ ร้านค้ากลุ่มที่ 2 คิดเป็น 33.56 เปอร์เซ็นต์ และร้านค้ากลุ่มที่ 3 คิดเป็น 53.66 เปอร์เซ็นต์

4.2.2 ผลการวิเคราะห์ลักษณะร้านค้า

ผลการวิเคราะห์ลักษณะร้านค้าแต่ละกลุ่มว่าเหมือนหรือแตกต่างกันซึ่งสามารถแสดงผลได้ดังนี้

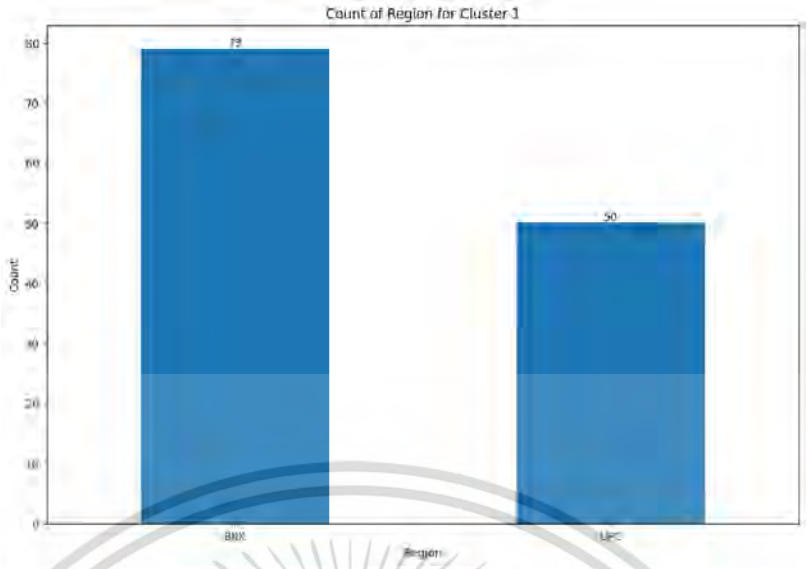
4.2.2.1 ผลลักษณะร้านค้ากลุ่มที่ 1



รูปที่ 4.8 ผลการวิเคราะห์ลักษณะที่ตั้งของร้านค้ากลุ่มที่ 1

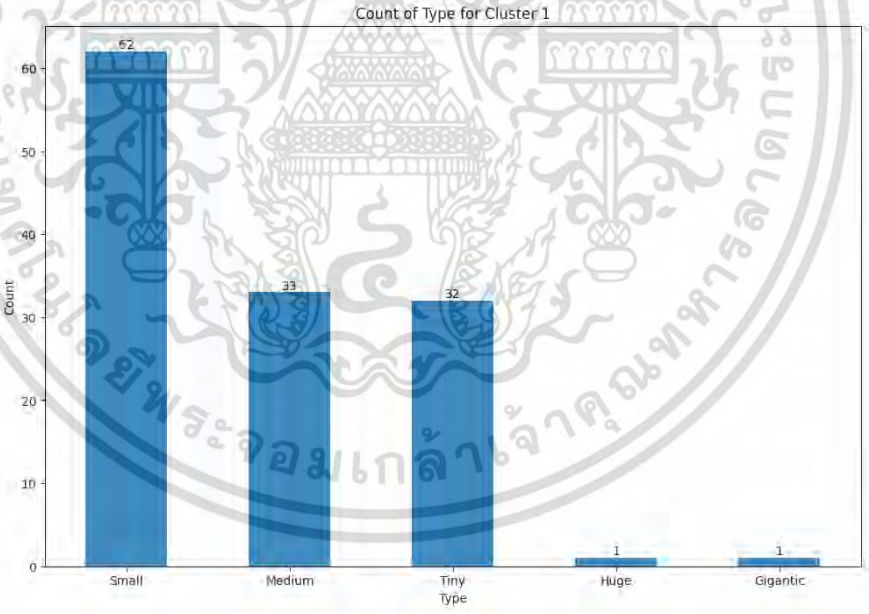
จากรูปที่ 4.8 ผลการวิเคราะห์ลักษณะที่ตั้งของร้านค้ากลุ่มที่ 1 พบว่าเป็นลักษณะที่ตั้งแบบ Huge จำนวน 63 ร้านค้า เป็นลักษณะที่ตั้งแบบ Medium จำนวน 33 ร้านค้า เป็นลักษณะที่ตั้งแบบ Gigantic จำนวน 26 ร้านค้า และเป็นลักษณะที่ตั้งแบบ Small จำนวน 7 ร้านค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.9 ผลการวิเคราะห์พื้นที่ตั้งของร้านค้ากลุ่มที่ 1

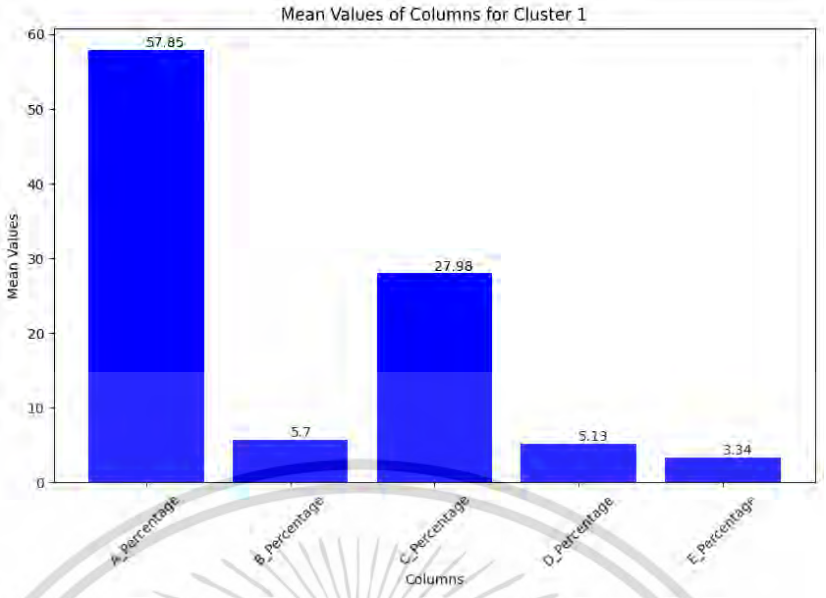
จากรูปที่ 4.9 ผลการวิเคราะห์พื้นที่ตั้งของร้านค้าของร้านค้ากลุ่มที่ 1 พบว่าเป็นพื้นที่ตั้งในกรุงเทพจำนวน 79 ร้านค้า และต่างจังหวัด จำนวน 50 ร้านค้า



รูปที่ 4.10 ผลการวิเคราะห์ประเภทร้านค้าของร้านค้ากลุ่มที่ 1

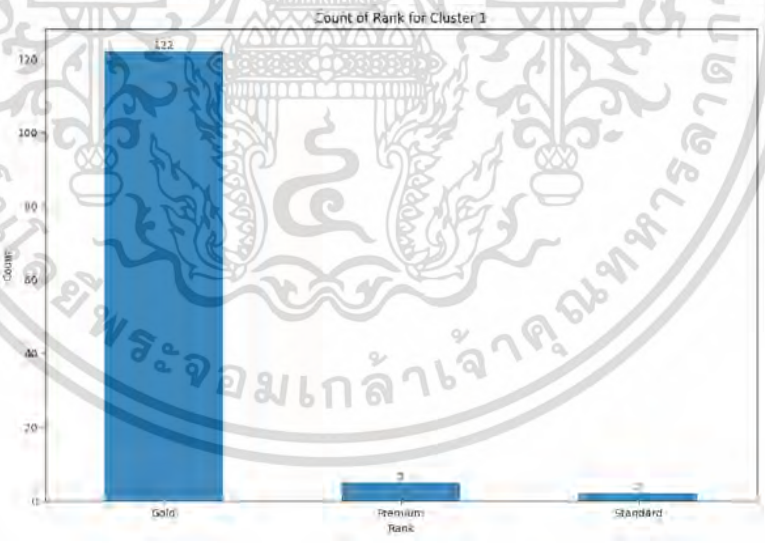
จากรูปที่ 4.10 ผลการวิเคราะห์ประเภทของร้านค้าในร้านค้ากลุ่มที่ 1 พบว่าเป็นร้านค้าประเภท Small จำนวน 62 ร้านค้า เป็นร้านค้าประเภท Medium จำนวน 33 ร้านค้า เป็นร้านค้าประเภท Tinyจำนวน 32 ร้านค้า เป็นร้านค้าประเภท Huge จำนวน 1 ร้านค้า และเป็นร้านค้าประเภท Gigantic จำนวน 1 ร้านค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.11 ผลการวิเคราะห์เปอร์เซ็นต์ที่ขายได้ของเบเกอรี่แต่ละชนิด ร้านค้ากลุ่มที่ 1

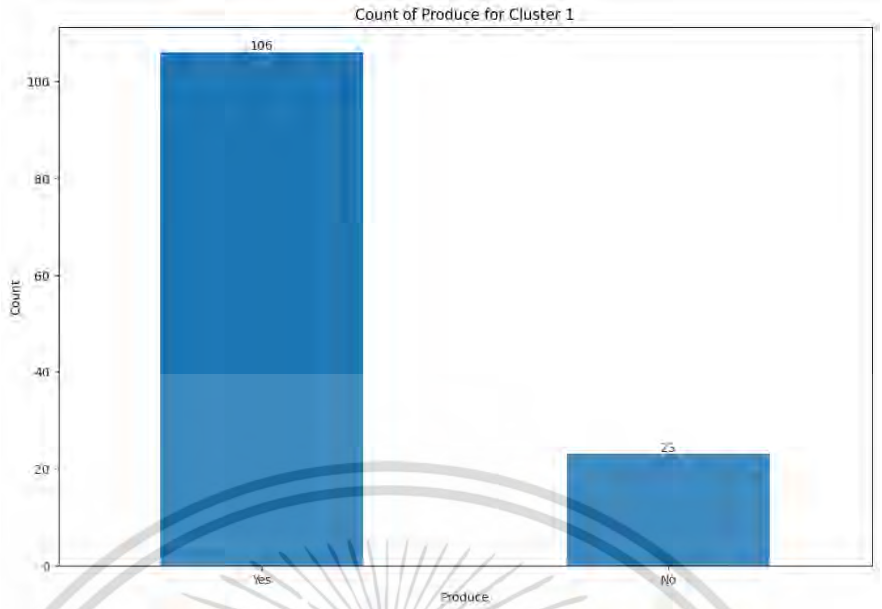
จากรูปที่ 4.11 ผลการวิเคราะห์เปอร์เซ็นต์ที่ขายได้ของเบเกอรี่แต่ละชนิดในร้านค้ากลุ่มที่ 1 พบว่าเป็นเบเกอรี่ชนิด A ร้อยละ 57.85 เป็นเบเกอรี่ชนิด B ร้อยละ 5.7 เป็นเบเกอรี่ชนิด C ร้อยละ 27.98 เป็นเบเกอรี่ชนิด D ร้อยละ 5.13 และเป็นเบเกอรี่ชนิด E ร้อยละ 3.34



รูปที่ 4.12 ผลการวิเคราะห์ระดับร้านค้าในร้านค้ากลุ่มที่ 1

จากรูปที่ 4.12 ผลการวิเคราะห์ระดับร้านค้าในร้านค้ากลุ่มที่ 1 พบว่าเป็นร้านค้าระดับ Gold จำนวน 122 ร้านค้า เป็นร้านค้าระดับ Premium จำนวน 5 ร้านค้า และเป็นร้านค้าระดับ Standard จำนวน 122 ร้านค้า

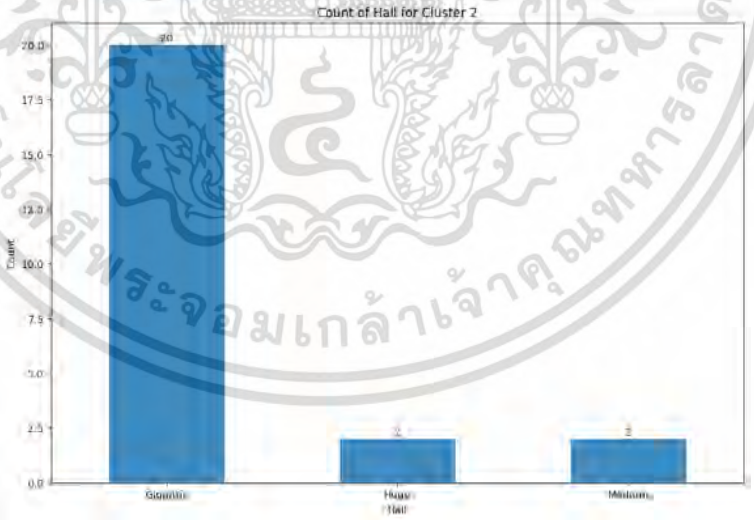
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.13 ผลการวิเคราะห์การผลิตสินค้าเองของร้านค้ากลุ่มที่ 1

จากรูปที่ 4.13 ผลการวิเคราะห์การผลิตสินค้าเองในร้านค้ากลุ่มที่ 1 พบว่าเป็นร้านค้าที่ผลิตสินค้าเองจำนวน 106 ร้านค้า และเป็นร้านค้าที่ไม่ได้ผลิตสินค้าเอง จำนวน 23 ร้านค้า

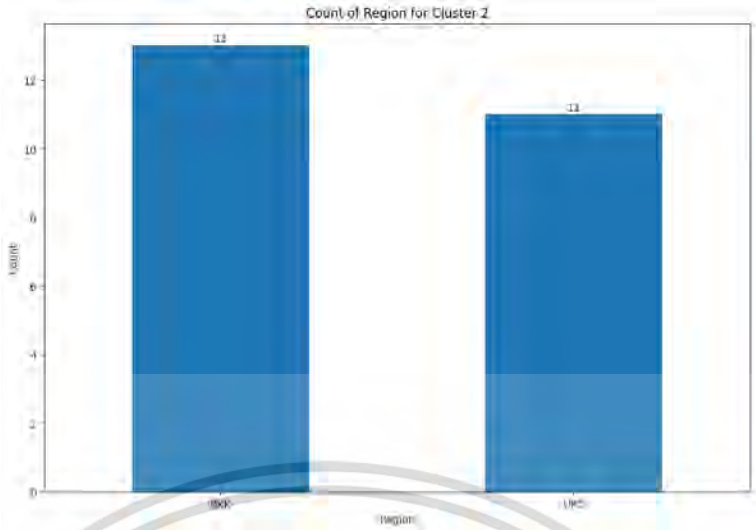
4.2.2.1 ผลลักษณะร้านค้ากลุ่มที่ 2



รูปที่ 4.14 ผลการวิเคราะห์ลักษณะที่ตั้งของร้านค้ากลุ่มที่ 2

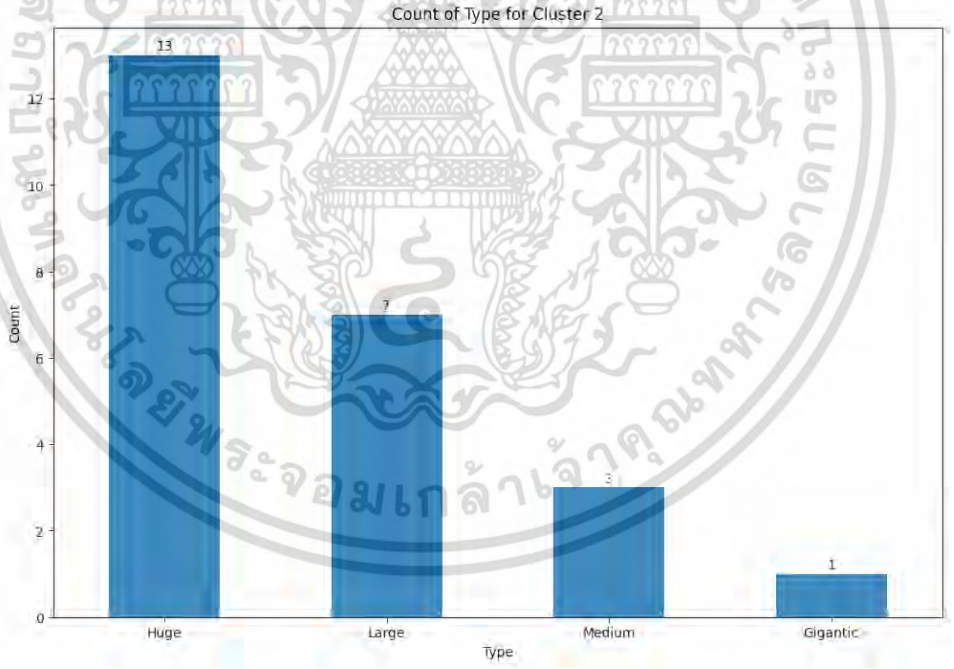
จากรูปที่ 4.14 ผลการวิเคราะห์ลักษณะที่ตั้งของร้านค้ากลุ่มที่ 2 พบว่าเป็นลักษณะที่ตั้งแบบ Gigantic จำนวน 20 ร้านค้า เป็นลักษณะที่ตั้งแบบ Huge จำนวน 2 ร้านค้า และเป็นลักษณะที่ตั้งแบบ Medium จำนวน 2 ร้านค้า

เอกสารนี้เป็นเอกสารสงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.15 ผลการวิเคราะห์พื้นที่ตั้งของร้านค้ากลุ่มที่ 2

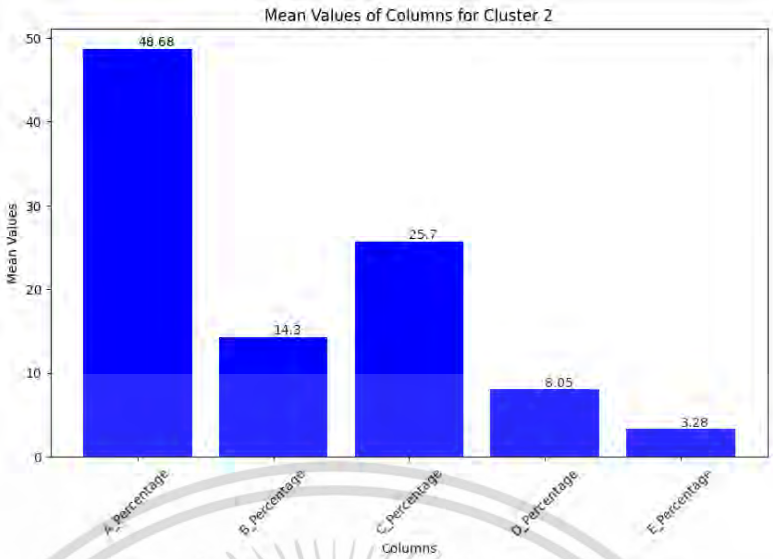
จากรูปที่ 4.15 ผลการวิเคราะห์พื้นที่ตั้งของร้านค้าในร้านค้ากลุ่มที่ 2 พบว่าเป็นพื้นที่ตั้งในกรุงเทพจำนวน 13 ร้านค้า และต่างจังหวัดจำนวน 11 ร้านค้า



รูปที่ 4.16 ผลการวิเคราะห์ประเภทร้านค้าของร้านค้ากลุ่มที่ 2

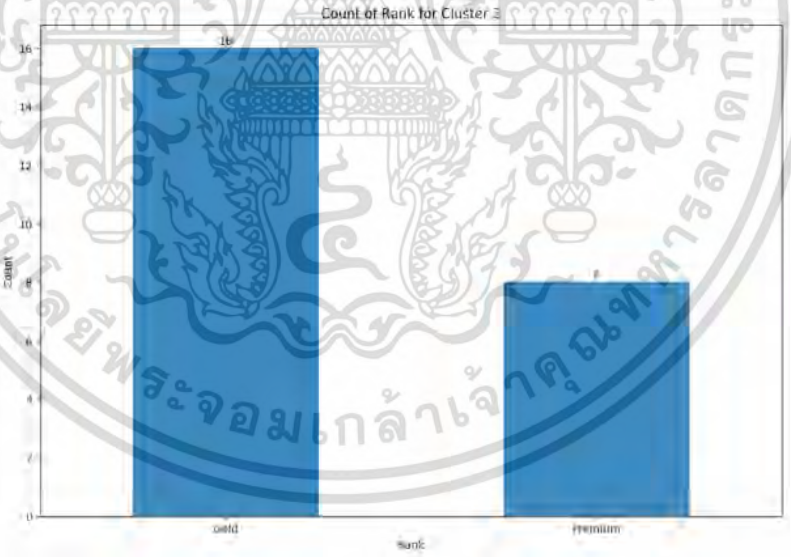
จากรูปที่ 4.16 ผลการวิเคราะห์ประเภทของร้านค้าในร้านค้ากลุ่มที่ 2 พบว่าเป็นร้านค้าประเภท Huge จำนวน 13 ร้านค้า เป็นร้านค้าประเภท Large จำนวน 7 ร้านค้า เป็นร้านค้าประเภท Medium จำนวน 3 ร้านค้า และเป็นร้านค้าประเภท Gigantic จำนวน 1 ร้านค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.17 ผลการวิเคราะห์เปอร์เซ็นต์ที่ขายได้ของเบเกอรี่แต่ละชนิด ร้านค้ากลุ่มที่ 2

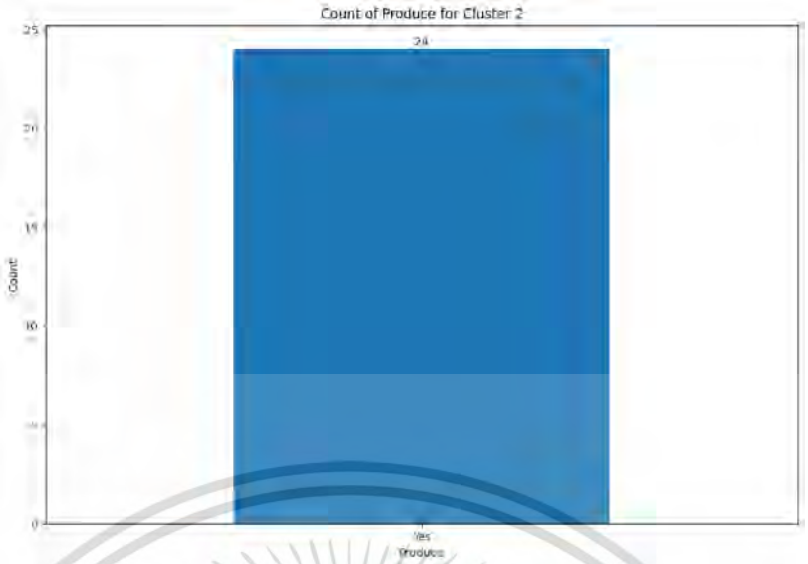
จากรูปที่ 4.17 ผลการวิเคราะห์เปอร์เซ็นต์ที่ขายได้ของเบเกอรี่แต่ละชนิดในร้านค้ากลุ่มที่ 2 พบว่าเป็นเบเกอรี่ชนิด A ร้อยละ 48.68 เป็นเบเกอรี่ชนิด B ร้อยละ 14.3 เป็นเบเกอรี่ชนิด C ร้อยละ 25.7 เป็นเบเกอรี่ชนิด D ร้อยละ 8.05 และเป็นเบเกอรี่ชนิด E ร้อยละ 3.28



รูปที่ 4.18 ผลการวิเคราะห์ระดับร้านค้าในร้านค้ากลุ่มที่ 2

จากรูปที่ 4.18 ผลการวิเคราะห์ระดับร้านค้าในร้านค้ากลุ่มที่ 2 พบว่าเป็นร้านค้าระดับ Gold จำนวน 16 ร้านค้า เป็นร้านค้าระดับ Premium จำนวน 8 ร้านค้า

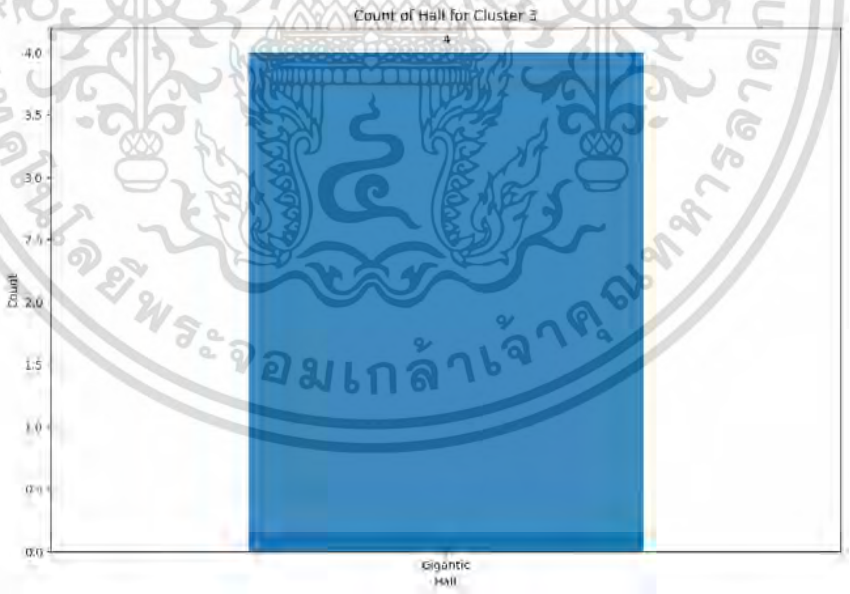
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.19 ผลการวิเคราะห์การผลิตสินค้าเองของร้านค้ากลุ่มที่ 2

จากรูปที่ 4.19 ผลการวิเคราะห์การผลิตสินค้าเองในร้านค้ากลุ่มที่ 2 พบว่าเป็นร้านค้าที่ผลิตสินค้าเองทั้งหมดจำนวน 24 ร้านค้า

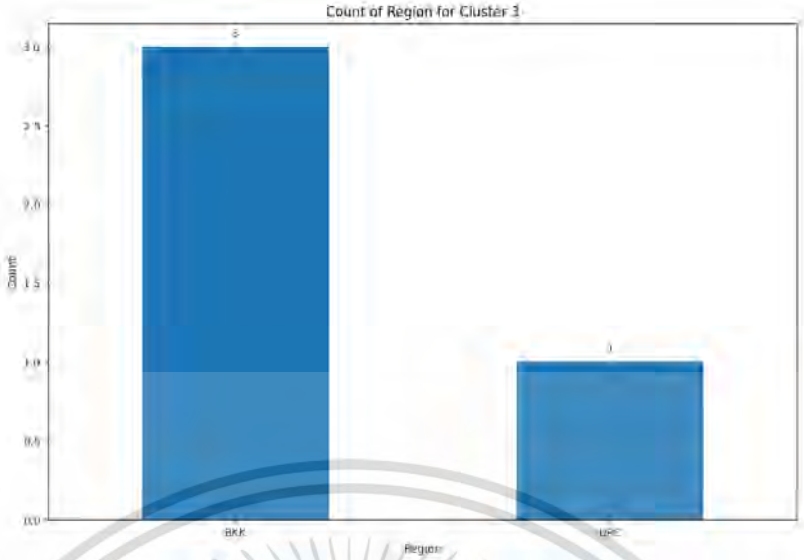
4.2.2.1 ผลลักษณะร้านค้ากลุ่มที่ 3



รูปที่ 4.20 ผลการวิเคราะห์ลักษณะที่ตั้งของร้านค้ากลุ่มที่ 3

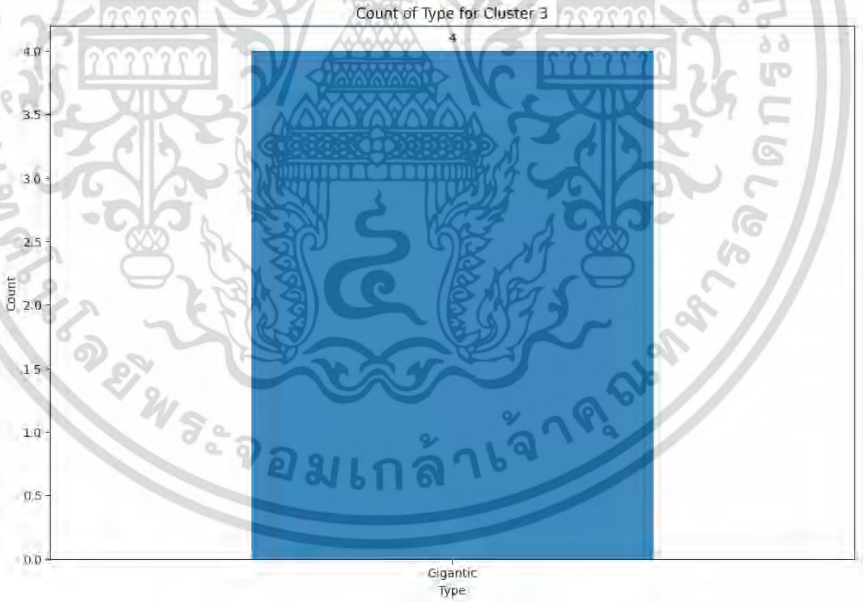
จากรูปที่ 4.20 ผลการวิเคราะห์ลักษณะที่ตั้งของร้านค้ากลุ่มที่ 3 พบว่าเป็นลักษณะที่ตั้ง Gigantic ทั้งหมดจำนวน 4 ร้านค้า

เอกสารนี้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.21 ผลการวิเคราะห์พื้นที่ตั้งของร้านค้ากลุ่มที่ 3

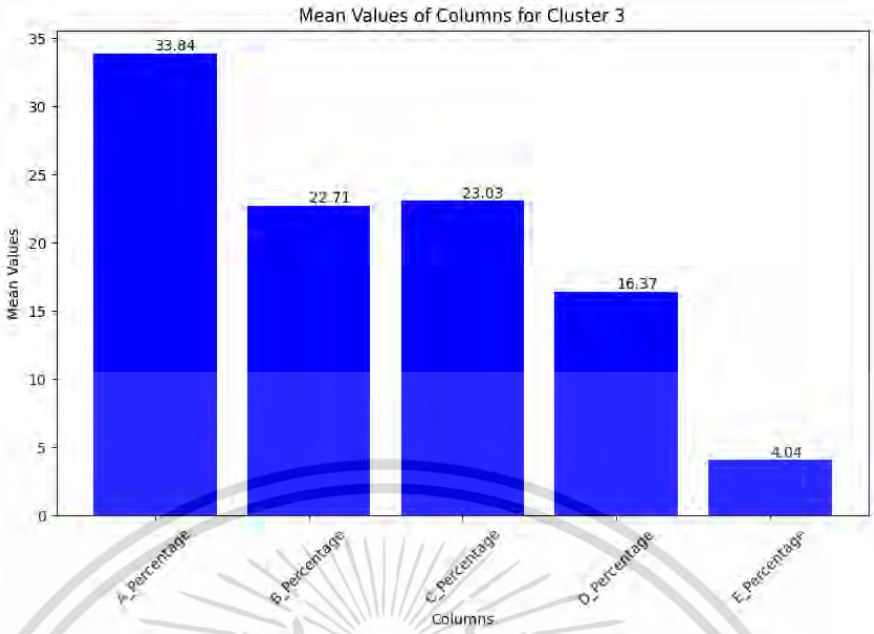
จากรูปที่ 4.21 ผลการวิเคราะห์พื้นที่ตั้งของร้านค้าในร้านค้ากลุ่มที่ 3 พบว่าเป็นพื้นที่ตั้งในกรุงเทพจำนวน 3 ร้านค้า และต่างจังหวัดจำนวน 1 ร้านค้า



รูปที่ 4.22 ผลการวิเคราะห์ประเภทร้านค้าของร้านค้ากลุ่มที่ 3

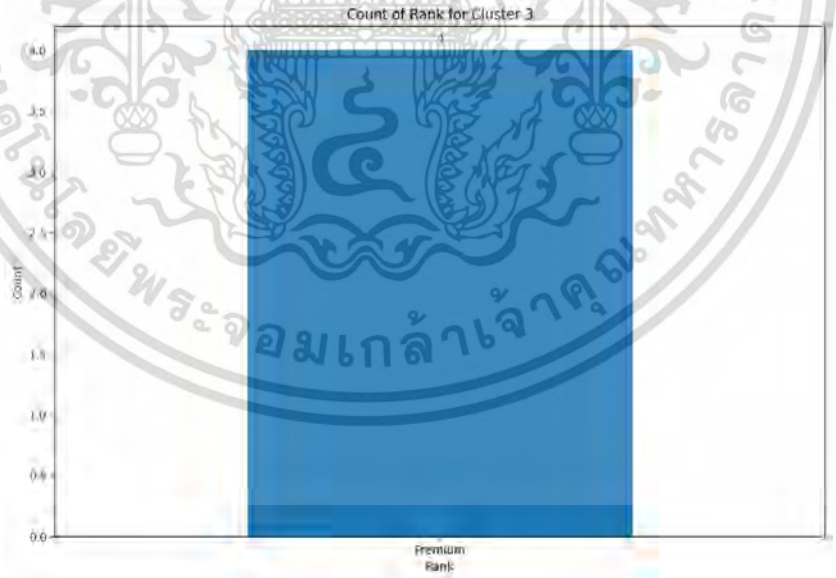
จากรูปที่ 4.22 ผลการวิเคราะห์ประเภทของร้านค้าในร้านค้ากลุ่มที่ 3 พบว่าเป็นร้านค้าประเภท Gigantic ทั้งหมดจำนวน 4 ร้านค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.23 ผลการวิเคราะห์เปอร์เซ็นต์ที่ขายได้ของเบเกอรี่แต่ละชนิด ร้านค้ากลุ่มที่ 3

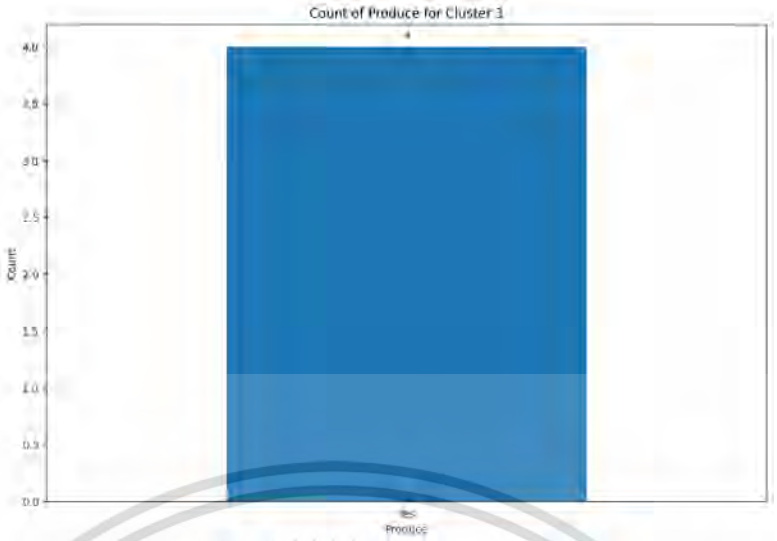
จากรูปที่ 4.23 ผลการวิเคราะห์เปอร์เซ็นต์ที่ขายได้ของเบเกอรี่แต่ละชนิดในร้านค้ากลุ่มที่ 3 พบว่าเป็นเบเกอรี่ชนิด A ร้อยละ 33.84 เป็นเบเกอรี่ชนิด B ร้อยละ 22.71 เป็นเบเกอรี่ชนิด C ร้อยละ 23.03 เป็นเบเกอรี่ชนิด D ร้อยละ 16.37 และเป็นเบเกอรี่ชนิด E ร้อยละ 4.04



รูปที่ 4.24 ผลการวิเคราะห์ระดับร้านค้าในร้านค้ากลุ่มที่ 3

จากรูปที่ 4.24 ผลการวิเคราะห์ระดับร้านค้าในร้านค้ากลุ่มที่ 3 พบว่าเป็นร้านค้า Premium ทั้งหมด จำนวน 4 ร้านค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

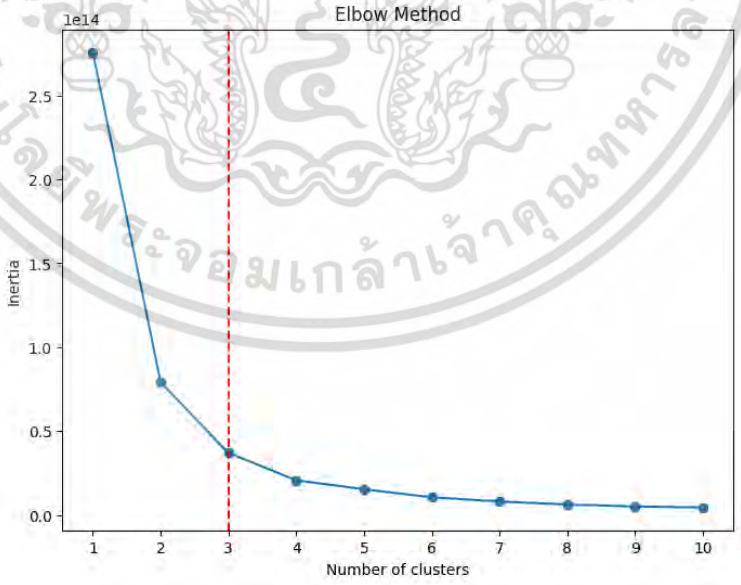


รูปที่ 4.25 ผลการวิเคราะห์การผลิตสินค้าเองของร้านค้ากลุ่มที่ 3

จากรูปที่ 4.25 ผลการวิเคราะห์การผลิตสินค้าเองในร้านค้ากลุ่มที่ 3 พบว่าเป็นร้านค้าที่ผลิตสินค้าเองทั้งหมดจำนวน 4 ร้านค้า

ผลการวิเคราะห์ลักษณะร้านค้ากลุ่มที่ 1 เพิ่มเติม

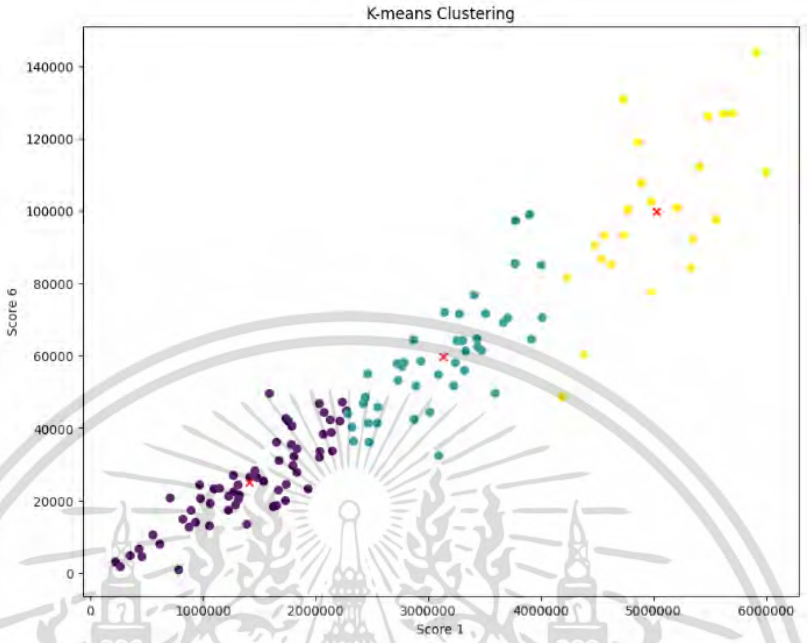
หลังจากทำการวิเคราะห์ลักษณะพบว่ากลุ่มที่ 1 มีจำนวนร้านค้ามากกว่ากลุ่มอื่นมากเกินไปจึงนำร้านค้ากลุ่มที่ 1 ไปทำการจัดกลุ่มเพื่อดูลักษณะเพิ่มเติมโดยใช้ Elbow Method กำหนดจำนวนกลุ่มได้ดังนี้



รูปที่ 4.26 ผลการหาจำนวนกลุ่มที่เหมาะสม

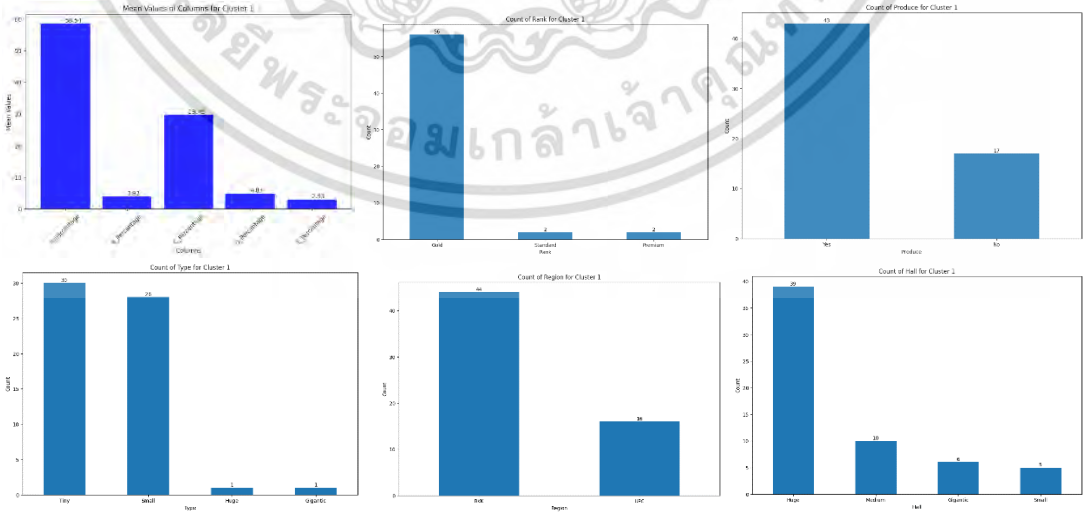
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.26 จะพบว่าจำนวนที่เหมาะสมในการกลุ่มคือ 3 กลุ่ม (K = 3) เนื่องจากมองจากภาพจุดหักข้อศอกที่เหมาะสมที่สุดอยู่ที่ 3



รูปที่ 4.27 ผลการจัดกลุ่มเพิ่มเติมจากร้านค้ากลุ่มที่ 1

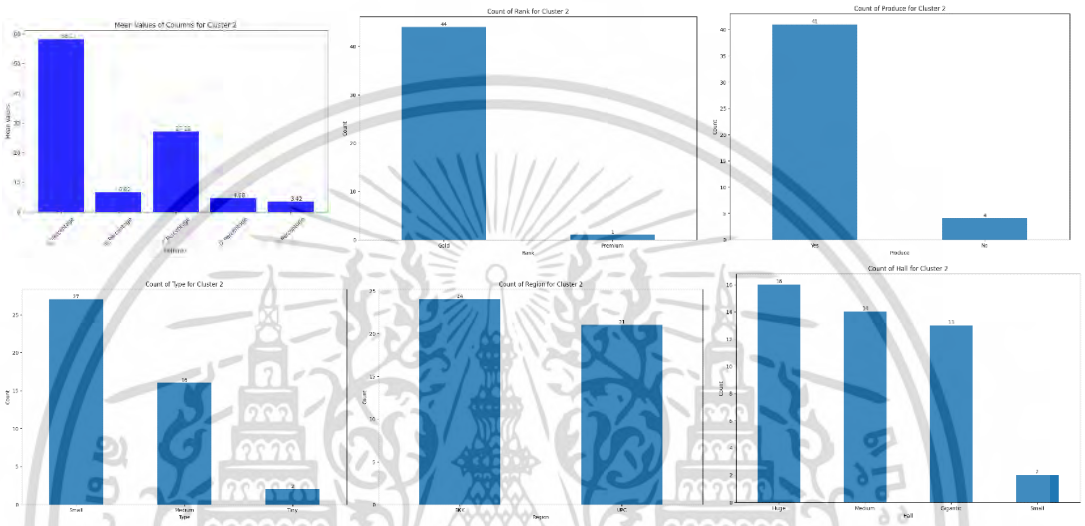
จากรูปที่ 4.27 จะเห็นได้ว่าหลังจากการจัดกลุ่มร้านค้าเป็น 3 กลุ่ม กำหนดให้สีม่วง ฟ้า เหลือง คือกลุ่มที่ A,B,C ตามลำดับ จากจำนวนร้านค้าในกลุ่มทั้งหมด 129 ร้านค้า ในกลุ่มที่ A มีจำนวนทั้งหมด 60 ร้านค้า ในกลุ่มที่ B มีจำนวนทั้งหมด 45 ร้านค้า และกลุ่มที่ C จำนวนทั้งหมด 24 ร้านค้า



รูปที่ 4.28 ผลการวิเคราะห์ลักษณะร้านค้ากลุ่มที่ A

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

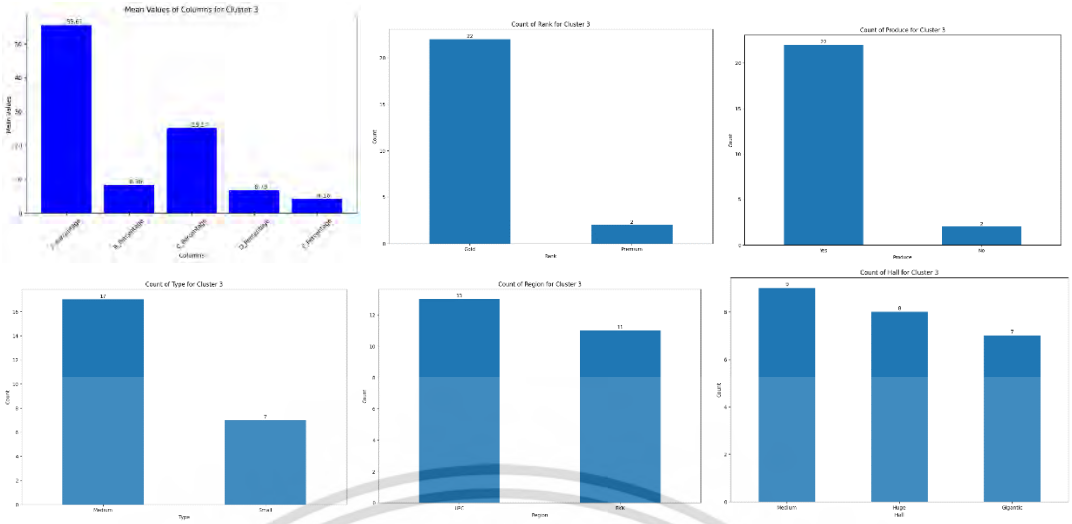
จากรูปที่ 4.28 ผลการวิเคราะห์ลักษณะร้านค้ากลุ่มที่ A จะพบว่าประเภทเบเกอรี่ที่ขายดียังคงเป็นประเภท A ระดับร้านค้าเป็นระดับ Gold จำนวน 56 ร้านค้า Standard จำนวน 2 ร้านค้า และ Premium จำนวน 2 ร้านค้า มีการผลิตสินค้าเองจำนวน 43 ร้านค้าและรับมาขายจำนวน 17 ร้านค้า ประเภทร้านค้าที่อยู่ในร้านค้ากลุ่ม A เป็นประเภท Tiny จำนวน 30 ร้านค้า Small 28 จำนวนร้านค้า Huge จำนวน 1 ร้านค้า Gigantic จำนวน 1 ร้านค้า พื้นที่ตั้งในกรุงเทพจำนวน 44 ร้านค้า ต่างจังหวัด 16 ร้านค้า และลักษณะที่ตั้งเป็น Hage จำนวน 39 ร้านค้า Medium จำนวน 10 ร้านค้า Gigantic จำนวน 6 ร้านค้า และ Small จำนวน 5 ร้านค้า



รูปที่ 4.29 ผลการวิเคราะห์ลักษณะร้านค้ากลุ่มที่ B

จากรูปที่ 4.29 ผลการวิเคราะห์ลักษณะร้านค้ากลุ่มที่ B จะพบว่าประเภทเบเกอรี่ที่ขายดียังคงเป็นประเภท A ระดับร้านค้าเป็นระดับ Gold จำนวน 44 ร้านค้า Premium จำนวน 1 ร้านค้าจะเห็นได้ว่าร้านค้ากลุ่ม B ไม่มีร้านค้าระดับ Standard มีการผลิตสินค้าเองจำนวน 41 ร้านค้าและรับมาขายจำนวน 4 ร้านค้า ประเภทร้านค้าที่อยู่ในร้านค้ากลุ่ม B เป็นประเภท Tiny จำนวน 2 ร้านค้า Small 27 และ Medium จำนวน 16 ร้านค้า พื้นที่ตั้งในกรุงเทพจำนวน 24 ร้านค้า ต่างจังหวัด 21 ร้านค้า และลักษณะที่ตั้งเป็น Hage จำนวน 16 ร้านค้า Medium จำนวน 14 ร้านค้า Gigantic จำนวน 13 ร้านค้า และ Small จำนวน 2 ร้านค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.30 ผลการวิเคราะห์ลักษณะร้านค้ากลุ่มที่ C

จากรูปที่ 4.30 ผลการวิเคราะห์ลักษณะร้านค้ากลุ่มที่ C จะพบว่าประเภทเบเกอรี่ที่ขายดียังคงเป็นประเภท A ระดับร้านค้าเป็นระดับ Gold จำนวน 22 ร้านค้า Premium จำนวน 2 ร้านค้าจะเห็นได้ว่าร้านค้ากลุ่ม C ไม่มีร้านค้าระดับ Standard มีการผลิตสินค้าเองจำนวน 22 ร้านค้าและรับมาขายจำนวน 2 ร้านค้า ประเภทร้านค้าที่อยู่ในร้านค้ากลุ่ม C คือประเภท Small 7 และ Medium จำนวน 17 ร้านค้า พื้นที่ตั้งในกรุงเทพจำนวน 11 ร้านค้า ต่างจังหวัด 13 ร้านค้า และลักษณะที่ตั้งเป็น Hage จำนวน 8 ร้านค้า Medium จำนวน 9 ร้านค้า Gigantic จำนวน 7 ร้านค้า

จากลักษณะร้านค้าทั้ง 3 กลุ่ม พบว่าในแต่ละกลุ่มร้านค้ามีลักษณะที่ใกล้เคียงกันมาก ผู้วิจัยจึงไม่ได้นำการวิเคราะห์ลักษณะร้านค้าเพิ่มเติมจากร้านค้ากลุ่มที่ 1 ไปใช้ในการอภิปรายและการสรุปผล

4.3 ผลการวิเคราะห์ข้อมูลด้วยสถิติอนุมาน (Inferential Statistics)

การวิเคราะห์ทางสถิติเพื่อตรวจสอบคุณลักษณะระหว่างกลุ่มว่ามีความเหมือนหรือแตกต่างกัน โดยใช้สถิติทดสอบ Kruskal-Wallis Test ด้วยโปรแกรมสำเร็จรูป IBM SPSS ได้ผลลัพธ์ดังต่อไปนี้

สมมติฐานการทดสอบ

$H_0 : M_1 = M_2 = M_3$

$H_1 : M_i \neq M_j$ แตกต่างกันอย่างน้อย 1 คู่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 ตารางแสดงลำดับค่าเฉลี่ยระหว่างกลุ่มร้านค้า

Rank			
Score 4	Cluster	N	Mean Rank
	1	78	98.59
	2	64	38.66
	3	15	149.27
Total	157		

ตารางที่ 4.2 ตารางผลลัพธ์การทดสอบความแตกต่างระหว่างกลุ่มร้านค้า

test statistics[a,b]	
	Score 4
Kruskal-Wallis H	100.699
df	2
Asymp. Sig.	0.000

เนื่องจากทำการทดสอบข้อกำหนดเบื้องต้นแล้ว พบว่าข้อมูลไม่มีการแจกแจงปรกติ สถิติที่ใช้ทดสอบ คือ Kruskal-Wallis Test

จากตารางที่ 4.2 พบว่าค่า Sig. คือ 0.000 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ 0.05 จึงปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.05 นั่นคือ ร้านค้าทั้ง 3 กลุ่มมีค่ามัธยฐานแตกต่างกันอย่างน้อย 1 คู่

4.4 การอภิปรายผล

จากการวิจัยโดยได้กลุ่มออกมาจำนวน 3 กลุ่ม โดยระบุชื่อกลุ่มที่ 1 คือกลุ่มร้านค้าเบเกอรี่ที่มีคะแนนการประเมินต่ำ กลุ่มที่ 2 คือกลุ่มร้านค้าเบเกอรี่ที่คะแนนการประเมินปานกลาง และกลุ่มที่ 3 คือกลุ่มร้านค้าเบเกอรี่ที่มีคะแนนการประเมินสูง หลังจากได้จำนวนกลุ่มและระบุจำนวนกลุ่มร้านค้าตามคะแนนการประเมินร้านค้าทั้ง 6 ตัว ทำให้ทราบว่าร้านค้ากลุ่มที่ 1 ส่วนใหญ่จะเป็นร้านค้าประเภท Small อีกทั้งตั้งอยู่ในลักษณะพื้นที่ Medium และมีร้านที่ไม่ได้ผลิตสินค้าเองถึง 23 ร้านค้า สำหรับร้านค้ากลุ่มที่ 2 พบว่าส่วนใหญ่เป็นร้านค้าประเภท Huge อีกทั้งตั้งอยู่ในลักษณะพื้นที่ Gigantic และกลุ่มที่ 3 พบว่าส่วนใหญ่เป็นร้านค้าประเภท Gigantic อีกทั้งตั้งอยู่ในลักษณะพื้นที่ Gigantic และส่วนใหญ่เป็นร้านค้าระดับ Premium การใช้ค่าสถิติในการวัดผลทำให้ทราบว่าข้อมูลทั้ง 3 กลุ่มที่จัดกลุ่มมาไม่มีการแจกแจงปรกติจึงได้สถิติไม่อิงพารามิเตอร์ในทดสอบ เนื่องจากมีจำนวนกลุ่ม 3 กลุ่มผู้วิจัยจึงใช้สถิติของ Kruskal-Wallis ในการวัดผลพบว่าค่า Sig. คือ 0.000 ซึ่งมีค่าน้อยกว่าระดับนัยสำคัญ 0.05 จึงปฏิเสธสมมติฐานหลักที่ระดับนัยสำคัญ 0.05 ทำให้ทราบว่ามัธยฐานของร้านค้าทั้ง 3 กลุ่มที่จัดกลุ่มมามีอย่างน้อย 1 คู่ที่แตกต่างกัน และจากการนำร้านค้ากลุ่มที่ 1 ไปจัดกลุ่มเพิ่มเติมจำนวน 3 กลุ่ม พบว่าในแต่ละกลุ่มร้านค้ามีลักษณะที่ใกล้เคียงกันมากหรือแทบไม่แตกต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

ในการศึกษาเรื่อง “การจัดกลุ่มร้านค้าเบเกอรี่” กรณีศึกษาบริษัทแห่งหนึ่งมีวัตถุประสงค์เพื่อจัดกลุ่มร้านค้าเบเกอรี่และวิเคราะห์ลักษณะของร้านค้าแต่ละกลุ่มที่จำแนกมาโดยสามารถสรุปผลได้ ดังนี้

5.1 สรุปผลการวิจัย

งานวิจัยครั้งนี้ทำการจัดกลุ่มร้านค้าเบเกอรี่สามารถจัดกลุ่มได้ 3 กลุ่มและพบว่าทั้ง 3 กลุ่มที่จำแนกมามีลักษณะที่แตกต่างกัน การวัดผลจากค่าสถิติของ Kruskal-Wallis ทำให้ทราบว่ามัธยฐานของร้านค้าทั้ง 3 กลุ่มที่จำแนกมามีอย่างน้อย 1 คู่ที่แตกต่างกัน การจัดกลุ่มร้านค้าในครั้งนี้ต้องการจัดกลุ่มร้านค้าที่มีลักษณะเหมือนหรือคล้ายคลึงกันมากที่สุดไว้ด้วยกัน จึงใช้เทคนิคการจัดกลุ่ม K-means โดยมีการใช้ Elbow Method ในการกำหนดจำนวนกลุ่มที่เหมาะสมและนำตัวแปรที่เลือกไว้ทั้งหมด 11 ตัว ซึ่งเป็นตัวแปรที่เป็นมาตรวัดอัตราส่วนและมาตรวัดอัตรภาคเข้าไปในโมเดลของ K-Means เพื่อจัดกลุ่ม และระบุชื่อกลุ่มจากผลการวิเคราะห์ข้อมูลด้วยสถิติพรรณนาเปรียบเทียบทั้ง 3 กลุ่ม จากตัวแปร 6 ตัว ได้แก่ คะแนนการประเมินตนเอง คะแนนความน่าเข้ามาซื้อของ คะแนนการจัดร้านค้า คะแนนที่ตั้งร้านค้า คะแนนการประเมินจากผู้จัดการร้านค้าและผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านค้า พบว่ากลุ่มที่ 1 มีสัดส่วนตัวแปรข้างต้นที่กล่าวมาต่ำที่สุดและมีผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านมากที่สุด จึงระบุชื่อกลุ่มที่ 1 เป็นกลุ่มร้านค้าเบเกอรี่ที่คะแนนการประเมินร้านค้าต่ำ พบว่ากลุ่มที่ 2 มีสัดส่วนตัวแปรข้างต้นที่กล่าวมาและมีผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านในระดับปานกลาง จึงระบุชื่อกลุ่มที่ 2 เป็นกลุ่มร้านค้าเบเกอรี่ที่มีคะแนนการประเมินร้านค้าในระดับปานกลาง และพบว่ากลุ่มที่ 3 มีสัดส่วนตัวแปรข้างต้นที่กล่าวมาสูงที่สุดและมีผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านน้อยที่สุด จึงระบุชื่อกลุ่มที่ 3 เป็นกลุ่มร้านค้าเบเกอรี่ที่มีคะแนนการประเมินร้านค้าสูง เนื่องจากคะแนนการประเมินตนเองและคะแนนการประเมินจากผู้จัดการร้านของร้านที่มีประสิทธิภาพในการบริหารควรจะมีค่าใกล้เคียงกันและผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านค่ายิ่งน้อยยิ่งสรุปได้ได้ว่าเป็นร้านที่มีการบริหารดี ดังนั้นร้านค้ากลุ่มที่มีผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านน้อยที่สุดยิ่งสรุปได้ว่ามีคะแนนการประเมินร้านค้าที่สูง หลังจากระบุกลุ่มแล้วทำการวิเคราะห์กลุ่มร้านค้าที่จำแนกมาว่ามีลักษณะเหมือนหรือแตกต่างกัน โดยการเปรียบเทียบลักษณะที่แตกต่างกันจากตัวแปรกลุ่มและเปอร์เซ็นต์ที่ขายได้ของเบเกอรี่แต่ละชนิด พบตัวแปรที่แตกต่างกันของแต่ละกลุ่มร้านค้าได้แก่ ประเภทร้านค้า ลักษณะพื้นที่ตั้ง การผลิตสินค้าเอง และระดับร้านค้า หลังจากนั้นใช้ค่าสถิติวัดผลการจัดกลุ่มจึงจำเป็นต้องทดสอบการแจกแจงปรกติก่อน

พบว่าส่วนใหญ่ไม่มีการแจกแจงปรกติจึงใช้การทดสอบของ Kruskal-Wallis ในการวัดผลและการทดสอบของ Kruskal-Wallis ปฏิเสธสมมติฐานหลัก ทำให้ทราบว่ามัธยฐานของร้านค้าทั้ง 3 กลุ่มที่จัดกลุ่มมามีอย่างน้อย 1 คู่ที่แตกต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

งานวิจัยครั้งนี้วัตถุประสงค์เพื่อจัดกลุ่มร้านเบเกอรี่ได้ผลลัพธ์เป็นจำนวน 3 กลุ่ม จากการใช้ Elbow Method ในการกำหนดจำนวนกลุ่มและนำตัวแปรที่เลือกไว้เข้าไปในโมเดลของ K-Means เพื่อจัดกลุ่มและระบุชื่อกลุ่มตามจุดเด่นของการทำ K-Means โดยดูจากสัดส่วนตัวแปรคะแนนการประเมินตนเอง คะแนนความน่าเข้ามาซื้อของ คะแนนการจัดร้านค้า คะแนนที่ตั้งร้านค้า คะแนนการประเมินจากผู้จัดการร้านค้าและผลต่างของคะแนนการประเมินตนเองเทียบกับคะแนนการประเมินจากผู้จัดการร้านได้ว่า กลุ่มที่ 1 เป็นกลุ่มร้านค้าเบเกอรี่ที่คะแนนการประเมินร้านค้าต่ำ จำนวน 129 ร้านค้า กลุ่มที่ 2 คือกลุ่มร้านค้าเบเกอรี่ที่คะแนนการประเมินร้านค้าในระดับปานกลางจำนวน 24 และกลุ่มที่ 3 เป็นกลุ่มร้านค้าเบเกอรี่ที่คะแนนการประเมินร้านค้าสูง จำนวน 4 ร้านค้า วัตถุประสงค์อีกข้อของงานวิจัยนี้คือเพื่อทราบลักษณะกลุ่มร้านค้าที่จำแนกมา หลังจากได้จำนวนกลุ่มและระบุกลุ่มร้านค้าตามวัตถุประสงค์ข้อที่ 1 แล้วผู้วิจัยได้นำตัวแปรเชิงคุณภาพและตัวแปรเปอร์เซ็นต์การขายได้ของสินค้าแต่ละชนิดมาทำการเปรียบเทียบลักษณะที่แตกต่างของแต่ละกลุ่มร้านค้าทำให้ทราบว่าร้านค้ากลุ่มที่มีคะแนนการประเมินร้านค้าต่ำหรือกลุ่มที่ 1 ร้านค้ากลุ่มที่มีคะแนนการประเมินร้านค้าปานกลางหรือกลุ่มที่ 2 ร้านค้ากลุ่มที่คะแนนการประเมินร้านค้าสูงหรือกลุ่มที่ 3 จากการวิเคราะห์ลักษณะของร้านค้าทั้ง 3 กลุ่มทำให้พบว่ามีตัวแปรที่มีลักษณะที่แตกต่างกัน คือประเภทร้านค้า ลักษณะพื้นที่ตั้ง การผลิตสินค้าเอง และระดับร้านค้า และได้มีการใช้ค่าสถิติในการวัดผลว่าข้อมูลมีความเหมือนหรือแตกต่างกัน ในการวิเคราะห์จากข้อกำหนดเบื้องต้นต้องทดสอบการแจกปรกติของข้อมูลทำให้ทราบว่าข้อมูลทั้ง 3 กลุ่มที่จัดกลุ่มมาไม่มีการแจกแจงปรกติจึงใช้สถิติไม่อิงพารามิเตอร์ในทดสอบเนื่องจากมีจำนวนกลุ่มถึง 3 กลุ่มผู้วิจัยจึงสถิติของ Kruskal-Wallis ในการวัดผลพบว่ามัธยฐานของร้านค้าทั้ง 3 กลุ่มที่จัดกลุ่มมามีอย่างน้อย 1 คู่ที่แตกต่างกันนั่นหมายความว่ากลุ่มที่จำแนกมามีลักษณะต่างกันอย่างน้อย 2 กลุ่ม

5.2 ข้อจำกัด

1) เนื่องจากข้อมูลที่ผู้วิจัยได้รับเป็นข้อมูลที่เป็นลิขสิทธิ์ของบริษัท จึงทำให้ไม่สามารถเปิดเผยชื่อตัวแปรที่แท้จริงได้ ในงานวิจัยนี้จึงมีการปรับเปลี่ยนชื่อตัวแปรให้มีความหมายที่สามารถสื่อสารให้เข้าใจได้

2) ข้อจำกัดในเรื่องของตัวแปรเนื่องจากมีตัวแปรไม่มากนัก ทำให้ไม่สามารถระบุความแตกต่างของกลุ่มร้านค้าได้ชัดเจน

5.3 ข้อเสนอแนะ

5.3.1 ข้อเสนอแนะในงานวิจัยครั้งถัดไป

1) ควรพิจารณาการจัดกลุ่มด้วยวิธีอื่นเพิ่มเติม เพื่อหาวิธีที่ให้ผลลัพธ์ดีที่สุดเนื่องจากการจัดกลุ่มนั้นมีหลายวิธี อาจจะพบวิธีที่ให้ประสิทธิภาพดีกว่าทำให้การจัดกลุ่มมีประสิทธิภาพและค้นพบความแตกต่างในแต่ละกลุ่มที่ชัดเจนมากขึ้น

2) ควรมีการพิจารณาการกำหนดจำนวนกลุ่มด้วยวิธีอื่นเพิ่มเติมสำหรับการจัดกลุ่มร้านค้าเบเกอรี่ เนื่องจากการจัดกลุ่มโดยวิธี K-Means มีข้อจำกัดเรื่องการกำหนดจำนวนกลุ่มที่เหมาะสม

3) ควรมีการเก็บรวบรวมข้อมูลตัวแปรที่เกี่ยวข้องเพิ่มเติมมาใช้ในงานวิจัยเพื่อวิเคราะห์ลักษณะของร้านค้าที่แตกต่างกันได้ดียิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ในเชิงพาณิชย์ ไม่ว่ากรณีใดๆ ทั้งสิ้น และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3.2 ข้อเสนอแนะเพื่อนำไปใช้

1) นำข้อมูลจากการจัดกลุ่มร้านค้าในกลุ่มที่มีการบริหารจัดการดีสามารถนำมาเป็นแนวทางในการพัฒนากลยุทธ์ทางธุรกิจ เพื่อเพิ่มกำไรของร้านค้าในกลุ่มอื่น ๆ ที่มีปัญหาในการบริหารจัดการ

2) ทำความเข้าใจเกี่ยวกับลักษณะและคุณสมบัติของกลุ่มร้านค้าที่แตกต่างกัน โดยการให้ความสำคัญกับจุดเด่นและปรับปรุงในส่วนที่มีข้อบกพร่อง เช่น การพัฒนาความสามารถในการบริหารจัดการ การสร้างยอดขาย หรือการลดผลต่างของการประเมินตนเองเทียบกับผู้จัดการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

กนิษฐ์กานต์ พรรหมราษฎร์. 2556. **กลยุทธ์การตลาดบริการสำหรับธุรกิจเบเกอรี่.** วารสารสุทธิปริทัศน์. มหาวิทยาลัยธุรกิจบัณฑิตย์.

จักรกฤษณ์ แสงแก้ว. 2549. **การเขียนโปรแกรมภาษาไพธอนด้วยตนเอง.** กรุงเทพมหานคร. สมาคมส่งเสริมเทคโนโลยี (ไทย-ญี่ปุ่น).

เดชฤทธิ์ มณีธรรม. 2563. **คัมภีร์การใช้งานระบบปัญญาประดิษฐ์และอินเตอร์เน็ตในทุกสรรพสิ่ง. กรุงเทพมหานคร.** ซีเอ็ดยูเคชั่น : 57.

ธนารักษ์ ธีระมั่นคง. 2563. **เครื่องมือพัฒนาโปรแกรมด้านปัญญาประดิษฐ์กับภาษาไพธอน.** สำนักวิทยาศาสตร์ราชบัณฑิตสภา. สถาบันเทคโนโลยีนานาชาติ มหาวิทยาลัยธรรมศาสตร์.

นวลวรรณ สุนทรภิษัช. 2563. **การเรียนรู้ของเครื่อง.** พิมพ์ครั้งที่ 1. กรุงเทพมหานคร. แดเน็กซ์อินเตอร์คอร์ปอเรชัน.

นวลวรรณ สุนทรภิษัช. 2553. **ปัญญาประดิษฐ์.** พิมพ์ครั้งที่ 2. กรุงเทพมหานคร. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

พจนานุกรมฉบับราชบัณฑิตยสถาน.**ราชบัณฑิตยสถาน.** 2542. นานมีบุ๊คส์พับลิเคชั่นส์

ชูใจ คูหารัตนไชย. 2563. **เอกสารประกอบการเรียนวิชาสถิติวิเคราะห์.** ภาควิชาสถิติ. คณะวิทยาศาสตร์. สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

วิกิพีเดีย. 2565. **การแบ่งกลุ่มข้อมูล.** [ออนไลน์]. เข้าถึงได้จาก https://th.m.wikipedia.org/wiki/การแบ่งกลุ่มข้อมูล.

ศศิวุฒิ ชัยเดชา . 2563. **หาจำนวน Clusters ที่เหมาะสมสำหรับ K-Means clustering ด้วย Elbow method.** [ออนไลน์]. เข้าถึงได้จาก https://lengyi.medium.com/หาจำนวน-clusters-ที่เหมาะสมสำหรับ-kmeans-clustering-ด้วย-elbow-method-85421efe9d.

ศูนย์อัจฉริยะเพื่ออุตสาหกรรมอาหาร. 2559. **อุตสาหกรรมเบเกอรี่ไทย** โดย คุณประเวศวุฒิ ไรวา ประธานเจ้าหน้าที่บริหาร บริษัท เอส แอนด์ พี ซินดิเคท จำกัด (มหาชน) . [ออนไลน์]. เข้าถึงได้จาก https://fic.nfi.or.th/area-based-Industry-detail.php?smid=1126.

สรชัย พิศาลบุตร. 2559. **หลักสถติ.** กรุงเทพมหานคร. วิทยพัฒน์.

สายชล สินสมบูรณ์ทอง. 2558. **การทำเหมืองข้อมูล.** กรุงเทพมหานคร. จามจุรีโปรดักส์.

สายชล สินสมบูรณ์ทอง. 2560. **สถิติเบื้องต้น.** พิมพ์ครั้งที่ 11. กรุงเทพมหานคร. จามจุรีโปรดักส์.

สายชล สินสมบูรณ์ทอง. 2563. **สถิติไม่อิงพารามิเตอร์.** พิมพ์ครั้งที่ 2. กรุงเทพมหานคร. จามจุรีโปรดักส์.

สิทธิชัย เจริญเศรษฐศิลป์. 2565. **เอกสารประกอบการเรียนวิชาการวางแผนการทดลอง.** ภาควิชาสถิติ. คณะวิทยาศาสตร์. สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

สุจิตรา สุคนธมัต. 2565. **เอกสารประกอบการเรียนวิชาโปรแกรมสำเร็จรูปทางสถิติ.** ภาควิชาสถิติ. คณะวิทยาศาสตร์. สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

สุพัฒน์ สุกมลสันต์. 2555. **การเปรียบเทียบก่อนและหลังการทดสอบรวมเพื่อการวิจัย.** วารสารภาษาปริทัศน์. สถาบันภาษา. จุฬาลงกรณ์มหาวิทยาลัย.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

อรพิน ประวัติบริสุทธิ์. 2564. **Python สำหรับ Data Science Data Visualization และ Machine Leaning.** พิมพ์ครั้งที่ 1. กรุงเทพ : โปรวิชั่น.

อุนนดา แพศรีวโรทัย. 2564**. การศึกษาพัฒนาการของจุดความร้อนจากไฟในภาคเหนือของประเทศไทยโดยใช้เทคนิคการจัดกลุ่มแบบดีบีสแกนและการประมาณความหนาแน่นแบบเคอร์เนล.** วิทยานิพนธ์คณะวิศวกรรมศาสตร์. จุฬาลงกรณ์มหาวิทยาลัย. 1 : 17

Agarwal, K., Jain, P., Rajnayak, M. 2019. **Comparative Analysis of Store Clustering Techniques in the Retail Industry.** SCITTERESS. Science and Technology Publications. Lda.

Ahmed, M., Seraj, R., Islam, S. 2020. **The k-means Algorithm: A Comprehensive Survey and Performance Evaluation.** Journals of Electronics.

Blum, A., & MiScore6hell, T. 1998. **Combining Labeled and Unlabeled Data with Co-Training.** InProc of 11th Annual Conf. on Computational Learning Theory, 92-100.

Brain GliScore6h. 2021. **K-mean clustering and its real use-case in the security domain.** [online]. เข้าถึงได้จาก https://medium.com/towardsdev/k-mean-clustering-and-its-real-use-case-in-the-security-domain-7009f490250a.

Euromonitor International. 2021. **Baked Goods in Thailand.** Industry Report.

Firdos, T. 2023. **Comparing Hierarchical, K-Means, and DBSCAN Clustering Algorithms in Python.** เข้าถึงได้จาก https://tahera-firdose.medium.com/understanding-clustering-algorithms-36f3652c8f33.

Jones, M., 2008. **Artificial Intelligence**. Infinity Science Press.

Hemalatha, M., Sivakumar, V. 2009. **Segmentation of Indian shoppers based on store attributes**. Management Studies National Institute of Technology Trichy. Tamilnadu India.

MiScore6hell, T.M. 1997. **Machine Learning, McGraw Hill**,1997.

Ogul, I., Zeybek, O. 2023**. Store Segmentation using Machine Leaning Methods : AnOrganized Supermarket Chain Case.** Intelligence and Fuzzy System. INFUS.

Padana, M., Ha, H. 2021. **Maximizing Strategy Improvement in Mall Customer Segmentation using K-means Clustering.** Journal of Applied Data Sciences. pp. 19-25.

Silver, D., Huang, A., Maddidson, C. et al. 2016. **Mastering the game of Go with deep neural networks and tree search.** Nature 529 : 484-489.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

ภาคผนวก ก ชุดคำสั่งไพทอนที่ใช้ในจัดเตรียมข้อมูล การกำหนดจำนวนกลุ่มที่เหมาะสม การจัดกลุ่มร้านค้าเบเกอรี่ และการเปรียบเทียบลักษณะของกลุ่มที่ได้จำแนกมาว่าเหมือนหรือแตกต่างกันของร้านค้าเบเกอรี่กรณีศึกษา เนื่องจากโปรแกรมภาษาไพทอนจะจัดกลุ่มได้กลุ่มที่ 0,1,2 แต่ความหมายคือกลุ่มที่ 1,2,3 ตามลำดับ ส่วนการวิเคราะห์ร้านค้ากลุ่มที่ 1 เพิ่มเติมใช้คำสั่งเดิมเพียงแต่เปลี่ยนข้อมูลกรองแค่ร้านค้ากลุ่มที่ 1 โดยการโหลดเป็นไฟล์ CSV.

ภาคนวก ก.1 ชุดคำสั่งที่ใช้ในการจัดเตรียมข้อมูล

```
# นำเข้าโมดูล pandas เพื่อใช้ในการจัดการข้อมูลแบบ DataFrame
import pandas as pd
import numpy as np
file_path = '/content/Data for clustering.xlsx'
# โหลดข้อมูลจากไฟล์ Excel เข้ามาในรูปแบบของ DataFrame
df = pd.read_excel('/content/Data for clustering.xlsx')
# การหาข้อมูลสูญหาย
df.isnull().sum()
# การแปลงข้อมูลให้อยู่ในมาตรฐานเดียวกัน
from sklearn.preprocessing import StandardScaler
numeric_columns = ['Score1', 'Score 2', 'Score3', 'Score4', 'Score5', 'SCORE6',
'A_Percentage', 'B_Percentage', 'C_Percentage', 'D_Percentage', 'E_Percentage',]
scaler = StandardScaler()
X = scaler.fit_transform(df[numeric_columns])
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคนวก ก.2 ชุดคำสั่งที่ใช้ในหาจำนวนกลุ่มที่เหมาะสม

```
# หาจำนวนกลุ่มที่เหมาะสมโดยวิธี Elbow Method
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
X = df[['Score1', 'Score 2', 'Score3', 'Score4', 'Score5', 'SCORE6', 'A_Percentage', 'B_Percentage', 'C_Percentage', 'D_Percentage', 'E_Percentage']].values
inertia = []
plt.figure(figsize=(8, 6))
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=0)
    kmeans.fit(X)
inertia.append(kmeans.inertia_)
# คำสั่งแสดงผลกราฟ Elbow Method
plt.plot(range(1, 11), inertia, marker='o')
plt.xlabel('Number of clusters')
plt.ylabel('Inertia')
plt.title('Elbow Method')
plt.show()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคนวก ก.3 ชุดคำสั่งที่ใช้ในการจัดกลุ่มด้วยวิธี K-Means

```
import pandas as pd
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
    numeric_columns = ['Score1', 'Score 2', 'Score3', 'Score4', 'Score5', 'SCORE6'
    'A_Percentage', 'B_Percentage', 'C_Percentage', 'D_Percentage', 'E_Percentage', ]
# Standardize the data
scaler = StandardScaler()
X = scaler.fit_transform(df[numeric_columns])
# ใช้เฉพาะฟีเจอร์ที่ต้องการ
    X = df[['Score1', 'Score 2', 'Score3', 'Score4', 'Score5', 'SCORE6', 'A_Percentage',
    'B_Percentage', 'C_Percentage', 'D_Percentage', 'E_Percentage']]
# ทำ K-means ด้วยจำนวน cluster = 3
kmeans = KMeans(n_clusters=3)
kmeans.fit(X)
# เพิ่ม label ของ cluster ลงใน DataFrame
df['Cluster'] = kmeans.labels_
# แสดงผลลัพธ์
print(df.head())
plt.figure(figsize=(10, 8))
# พล็อตข้อมูลและ centroid ของ cluster โดยใช้ฟีเจอร์ 'Score1' และ 'Score 2'
plt.scatter(X['Score1'], X['Score 2'], c=kmeans.labels_, cmap='viridis')
# พล็อต centroid ของ cluster แต่ละคลัสเตอร์
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], marker='x',
color='red')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
# ระบุสีและจำนวนของแต่ละคลัสเตอร์
for i, center in enumerate(kmeans.cluster_centers_):
    plt.text(center[0], center[1], f'Cluster {i}', fontsize=12, color='black', ha='center')
plt.title('K-means Clustering')
plt.xlabel('Score1')
plt.ylabel('Score 2')
plt.ticklabel_format(style='plain', axis='both')
plt.show()
#โหลดไฟล์สำหรับนำเข้าวิเคราะห์ต่อใน SPSS
df.to_csv('Clustering to spss1.csv', index=False)
```

ภาคนวก ก.4 ชุดคำสั่งที่ใช้ในการวิเคราะห์สัดส่วนของคะแนนการจัดร้านค้า

```
#หาผลรวมของ Score 2 ทั้งหมด
total_sum_of_quant = df['Score 2'].sum()
# หาค่าเฉลี่ยของ Score 2 ในแต่ละคลัสเตอร์
sum_of_quant_cluster_mean = df.groupby('Cluster')['Score 2'].mean()
# คำนวณสัดส่วนของ Score 2 ในแต่ละคลัสเตอร์
sum_of_quant_cluster_percentage = (sum_of_quant_cluster_mean /
total_sum_of_quant) * 100
# ค่าสัดส่วนของแต่ละคลัสเตอร์เท่ากับ 100 เปอร์เซ็นต์
# โดยหารค่าของแต่ละคลัสเตอร์ด้วยผลรวมของ Score 2 ทั้งหมดแล้วคูณด้วย 100
# ซึ่งจะทำให้ผลรวมของค่าสัดส่วนทั้งหมดเป็น 100%
sum_of_quant_cluster_percentage = (sum_of_quant_cluster_percentage /
sum_of_quant_cluster_percentage.sum()) * 100
# พล็อตกราฟแท่งแสดงสัดส่วนของ Score 2 ของแต่ละคลัสเตอร์โดยใส่สี
plt.figure(figsize=(10, 6))
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
bars = plt.bar(sum_of_quant_cluster_percentage.index,
sum_of_quant_cluster_percentage, color=colors)

plt.xlabel('Cluster')

plt.ylabel('Percentage of Total Score 2 (%)')

plt.title('Proportion of Total Score 2 for Each Cluster')

# เพิ่มค่าสัดส่วนในแต่ละแท่ง

for bar in bars:

    height = bar.get_height()

    plt.text(bar.get_x() + bar.get_width() / 2, height, f'{height:.2f}%', ha='center',
va='bottom')

# แสดงค่าเป็นทศนิยมสองตำแหน่งหลังจุดทศนิยม

plt.gca().yaxis.set_major_formatter(mtick.PercentFormatter(decimals=2))

# กำหนดชื่อของแกน x ใหม่

plt.xticks(sum_of_quant_cluster_percentage.index, ['Cluster 0', 'Cluster 1', 'Cluster 2'])

plt.show()
```

ภาคนวก ก.5 ชุดคำสั่งที่ใช้ในการวิเคราะห์สัดส่วนคะแนนการประเมินตนเอง

```
# หาผลรวมของ Score3ทั้งหมด

total_GP_actual = df['Score3'].sum()

# หาค่าเฉลี่ยของ Score3ในแต่ละคลัสเตอร์

GP_actual_cluster_mean = df.groupby('Cluster')['Score3'].mean()

# คำนวณสัดส่วนของ Score3ในแต่ละคลัสเตอร์

GP_actual_cluster_percentage = (GP_actual_cluster_mean / total_GP_actual) * 100

# ค่าสัดส่วนของแต่ละคลัสเตอร์เท่ากับ 100 เปอร์เซ็นต์

# โดยหารค่าของแต่ละคลัสเตอร์ด้วยผลรวมของ Score3ทั้งหมดแล้วคูณด้วย 100

# ซึ่งจะทำให้ผลรวมของค่าสัดส่วนทั้งหมดเป็น 100%
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
GP_actual_cluster_percentage = (GP_actual_cluster_percentage / GP_actual_cluster_percentage.sum()) * 100

# พล็อตกราฟแท่งแสดงสัดส่วนของ Score3ของแต่ละคลัสเตอร์โดยใส่สี

plt.figure(figsize=(10, 6))

bars = plt.bar(GP_actual_cluster_percentage.index, GP_actual_cluster_percentage, color=colors)

plt.xlabel('Cluster')

plt.ylabel('Percentage of Total GP Actual(%)')

plt.title('Proportion of Total Score3for Each Cluster')

# เพิ่มค่าสัดส่วนในแต่ละแท่ง

for bar in bars:

    height = bar.get_height()

    plt.text(bar.get_x() + bar.get_width() / 2, height, f'{height:.2f}%', ha='center', va='bottom')

# แสดงค่าเป็นทศนิยมสองตำแหน่งหลังจุดทศนิยม

plt.gca().yaxis.set_major_formatter(mtick.PercentFormatter(decimals=2))

# กำหนดชื่อของแกน x ใหม่

plt.xticks(GP_actual_cluster_percentage.index, ['Cluster 0', 'Cluster 1', 'Cluster 2'])

plt.show()
```

ภาคนวก ก.6 ชุดคำสั่งที่ใช้ในการวิเคราะห์สัดส่วนเปอร์เซ็นกำไรต่อยอดขายรวม

```
# หาผลรวมของ Score4 ทั้งหมด

total_Score4 = df['Score4'].sum()

# หาค่าเฉลี่ยของ Score4 ในแต่ละคลัสเตอร์

Score4_cluster_mean = df.groupby('Cluster')['Score4'].mean()

# คำนวณสัดส่วนของ Score4 ในแต่ละคลัสเตอร์
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
Score4_cluster_percentage = (Score4_cluster_mean / total_Score4) * 100
# ค่าสัดส่วนของแต่ละคลัสเตอร์เท่ากับ 100 เปอร์เซ็นต์
# โดยหารค่าของแต่ละคลัสเตอร์ด้วยผลรวมของ Score4 ทั้งหมดแล้วคูณด้วย 100
# ซึ่งจะทำให้ผลรวมของค่าสัดส่วนทั้งหมดเป็น 100%
Score4_cluster_percentage = (Score4_cluster_percentage /
Score4_cluster_percentage.sum()) * 100
# พล็อตกราฟแท่งแสดงสัดส่วนของ Score4 ของแต่ละคลัสเตอร์โดยใส่สี
plt.figure(figsize=(10, 6))
bars = plt.bar(Score4_cluster_percentage.index, Score4_cluster_percentage,
color=colors)
plt.xlabel('Cluster')
plt.ylabel('Percentage of Total Score4 ')
plt.title('Proportion of Total Score4 for Each Cluster')
# เพิ่มค่าสัดส่วนในแต่ละแท่ง
for bar in bars:
    height = bar.get_height()
    plt.text(bar.get_x() + bar.get_width() / 2, height, f'{height:.2f}%', ha='center',
va='bottom')
# แสดงค่าเป็นทศนิยมสองตำแหน่งหลังจุดทศนิยม
plt.gca().yaxis.set_major_formatter(mtick.PercentFormatter(decimals=2))
# กำหนดชื่อของแกน x ใหม่
plt.xticks(Score4_cluster_percentage.index, ['Cluster 0', 'Cluster 1', 'Cluster 2'])
plt.show()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคนวก ก.7 ชุดคำสั่งที่ใช้ในการวิเคราะห์สัดส่วนเปอร์เซ็นผลต่างของการประเมินตนเองเทียบกับผู้จัดการ

```
# หาผลรวมของ Score5 ทั้งหมด

total_Score5 = df['Score5'].sum()

# หาค่าเฉลี่ยของ Score5 ในแต่ละคลัสเตอร์

Score5_cluster_mean = df.groupby('Cluster')['Score5'].mean()

# คำนวณสัดส่วนของ Score5 ในแต่ละคลัสเตอร์

Score5_cluster_percentage = (Score5_cluster_mean / total_Score5) * 100

# ค่าสัดส่วนของแต่ละคลัสเตอร์เท่ากับ 100 เปอร์เซ็นต์

# โดยหารค่าของแต่ละคลัสเตอร์ด้วยผลรวมของ Score5 ทั้งหมดแล้วคูณด้วย 100

# ซึ่งจะทำให้ผลรวมของค่าสัดส่วนทั้งหมดเป็น 100%

Score5_cluster_percentage = (Score5_cluster_percentage /
Score5_cluster_percentage.sum()) * 100

# พล็อตกราฟแท่งแสดงสัดส่วนของ Score5 ของแต่ละคลัสเตอร์โดยใส่สี

plt.figure(figsize=(10, 6))

bars = plt.bar(Score5_cluster_percentage.index, Score5_cluster_percentage,
color=colors)

plt.xlabel('Cluster')

plt.ylabel('Percentage of Total Score5 (%)')

plt.title('Proportion of Total Score5 for Each Cluster')

# เพิ่มค่าสัดส่วนในแต่ละแท่ง

for bar in bars:

    height = bar.get_height()

    plt.text(bar.get_x() + bar.get_width() / 2, height, f'{height:.2f}%', ha='center',
va='bottom')

# แสดงค่าเป็นทศนิยมสองตำแหน่งหลังจุดทศนิยม
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plt.gca().yaxis.set_major_formatter(mtick.PercentFormatter(decimals=2))
# กำหนดชื่อของแกน x ใหม่
plt.xticks(Score5_cluster_percentage.index, ['Cluster 0', 'Cluster 1', 'Cluster 2'])
plt.show()
```

ภาคนวก ก.8 ชุดคำสั่งที่ใช้ในการวิเคราะห์สัดส่วนยอดขายรวม

```
# หาผลรวมของ Score1 ทั้งหมด
total_net_total = df['Score1'].sum()
# หาค่าเฉลี่ยของ Score1 ในแต่ละคลัสเตอร์
net_total_cluster_mean = df.groupby('Cluster')['Score1'].mean()
# คำนวณสัดส่วนของ Score1 ในแต่ละคลัสเตอร์
net_total_cluster_percentage = (net_total_cluster_mean / total_net_total) * 100
# ค่าสัดส่วนของแต่ละคลัสเตอร์เท่ากับ 100 เปอร์เซ็นต์
# โดยหารค่าของแต่ละคลัสเตอร์ด้วยผลรวมของ Score1 ทั้งหมดแล้วคูณด้วย 100
# ซึ่งจะทำให้ผลรวมของค่าสัดส่วนทั้งหมดเป็น 100%
net_total_cluster_percentage = (net_total_cluster_percentage /
net_total_cluster_percentage.sum()) * 100
# พล็อตกราฟแท่งแสดงสัดส่วนของ Score1 ของแต่ละคลัสเตอร์โดยใส่สี
plt.figure(figsize=(10, 6))
bars = plt.bar(net_total_cluster_percentage.index, net_total_cluster_percentage,
color=colors)
plt.xlabel('Cluster')
plt.ylabel('Percentage of Total Score1 ')
plt.title('Proportion of Total Score1 for Each Cluster')
# เพิ่มค่าสัดส่วนในแต่ละแท่ง
for bar in bars:
    height = bar.get_height()
    plt.text(bar.get_x() + bar.get_width() / 2, height, f'{height:.2f}%', ha='center',
va='bottom')
# แสดงค่าเป็นทศนิยมสองตำแหน่งหลังจุดทศนิยม
plt.gca().yaxis.set_major_formatter(mtick.PercentFormatter(decimals=2))
# กำหนดชื่อของแกน x ใหม่
plt.xticks(net_total_cluster_percentage.index, ['Cluster 0', 'Cluster 1', 'Cluster 2'])
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plt.show()
```

ภาคนวก ก.9 ชุดคำสั่งที่ใช้ในการวิเคราะห์สัดส่วนคะแนนที่ตั้งร้านค้า

```
# หาผลรวมของค่า SCORE6 ทั้งหมด
total_SCORE6 = df['SCORE6'].sum()
# หาค่าเฉลี่ยของ SCORE6 ในแต่ละคลัสเตอร์
SCORE6_cluster_mean = df.groupby('Cluster')['SCORE6'].mean()
# คำนวณสัดส่วนของ SCORE6 ในแต่ละคลัสเตอร์
SCORE6_cluster_percentage = (SCORE6_cluster_mean / total_SCORE6) * 100
# ค่าสัดส่วนของแต่ละคลัสเตอร์เท่ากับ 100 เปอร์เซ็นต์
# โดยหารค่าของแต่ละคลัสเตอร์ด้วยผลรวมของ SCORE6 ทั้งหมดแล้วคูณด้วย 100
# ซึ่งจะทำให้ผลรวมของค่าสัดส่วนทั้งหมดเป็น 100%
SCORE6_cluster_percentage = (SCORE6_cluster_percentage /
SCORE6_cluster_percentage.sum()) * 100
# พล็อตกราฟแท่งแสดงสัดส่วนของ SCORE6 ของแต่ละคลัสเตอร์โดยใส่สี
plt.figure(figsize=(10, 6))
bars = plt.bar(SCORE6_cluster_percentage.index, SCORE6_cluster_percentage,
color=colors)
plt.xlabel('Cluster')
plt.ylabel('Percentage of Total SCORE6 (%)')
plt.title('Proportion of Total SCORE6 for Each Cluster')
# เพิ่มค่าสัดส่วนในแต่ละแท่ง
for bar in bars:
    height = bar.get_height()
    plt.text(bar.get_x() + bar.get_width() / 2, height, f'{height:.2f}%', ha='center',
va='bottom')
# แสดงค่าเป็นทศนิยมสองตำแหน่งหลังจุดทศนิยม
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plt.gca().yaxis.set_major_formatter(mtick.PercentFormatter(decimals=2))
plt.show()
```

ภาคนวก ก.10 ชุดคำสั่งที่ใช้ในการวิเคราะห์ระดับร้านค้าของกลุ่มที่ 1

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 0
cluster_0_data = df[df['Cluster'] == 0]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Rank' สำหรับคลัสเตอร์ที่ 0
unique_format_counts_cluster_0 = cluster_0_data['Rank'].value_counts()
# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Format' สำหรับคลัสเตอร์ที่ 0
plt.figure(figsize=(12, 8))
bars = unique_format_counts_cluster_0.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('Rank')
plt.ylabel('Count')
plt.title('Count of Rank for Cluster 0')
plt.xticks(rotation=0)
plt.show()
```

ภาคนวก ก.11 ชุดคำสั่งที่ใช้ในการวิเคราะห์การผลิตสินค้าเองของกลุ่มที่ 1

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 0
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
cluster_0_data = df[df['Cluster'] == 0]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Produce' สำหรับคลัสเตอร์ที่ 0
unique_produce_counts_cluster_0 = cluster_0_data['Produce'].value_counts()

# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Produce' สำหรับคลัสเตอร์ที่ 0
plt.figure(figsize=(12, 8))
bars = unique_produce_counts_cluster_0.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('Produce')
plt.ylabel('Count')
plt.title('Count of Produce for Cluster 0')
plt.xticks(rotation=0)
plt.show()
```

ภาคนวก ก.12 ชุดคำสั่งที่ใช้ในการวิเคราะห์ประเภทร้านค้าของกลุ่มที่ 1

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 0
cluster_0_data = df[df['Cluster'] == 0]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Type' สำหรับคลัสเตอร์ที่ 0
unique_type_counts_cluster_0 = cluster_0_data['Type'].value_counts()
# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Type' สำหรับคลัสเตอร์ที่ 0
plt.figure(figsize=(12, 8))
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
bars = unique_type_counts_cluster_0.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('Type')
plt.ylabel('Count')
plt.title('Count of Type for Cluster 0')
plt.xticks(rotation=0)
plt.show()
```

ภาคนวก ก.13 ชุดคำสั่งที่ใช้ในการวิเคราะห์พื้นที่ตั้งร้านค้าของกลุ่มที่ 1

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 0
cluster_0_data = df[df['Cluster'] == 0]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Region' สำหรับคลัสเตอร์ที่ 0
unique_region_counts_cluster_0 = cluster_0_data['Region'].value_counts()
# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Region' สำหรับคลัสเตอร์ที่ 0
plt.figure(figsize=(12, 8))
bars = unique_region_counts_cluster_0.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
```

เอกสารนี้เป็นเอกสารที่สงวนสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plt.xlabel('Region')

plt.ylabel('Count')

plt.title('Count of Region for Cluster 0')

plt.xticks(rotation=0)

plt.show()
```

ภาคนวก ก.14 ชุดคำสั่งที่ใช้ในการวิเคราะห์ลักษณะที่ตั้งร้านค้าของกลุ่มที่ 1

```
import matplotlib.pyplot as plt

# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 0

cluster_0_data = df[df['Cluster'] == 0]

# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'hall' สำหรับคลัสเตอร์ที่ 0

unique_foodhall_counts_cluster_0 = cluster_0_data['hall'].value_counts()

# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'hall' สำหรับคลัสเตอร์ที่ 0

plt.figure(figsize=(12, 8))

bars = unique_foodhall_counts_cluster_0.plot(kind='bar', stacked=True)

# เพิ่มค่าจำนวนบนแต่ละแท่ง

for i, bar in enumerate(bars.paScore6hes):

    height = bar.get_height()

        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')

plt.xlabel('hall')

plt.ylabel('Count')

plt.title('Count of hall for Cluster 0')

plt.xticks(rotation=0)

plt.show()
```

ภาคนวก ก.15 ชุดคำสั่งที่ใช้ในการวิเคราะห์เปอร์เซ็นที่ขายได้ของเบเกอรี่แต่ละชนิดกลุ่มที่ 1

```
# สร้าง DataFrame ที่มีเฉพาะ cluster 0
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
cluster_0_data = df[df['Cluster'] == 0]

        columns_of_interest = ['A _Percentage', 'B_Percentage',      'C_Percentage',
    'D_Percentage', 'E_Percentage']

# หาค่าเฉลี่ยของคอลัมน์

mean_values0 = cluster_0_data[columns_of_interest].mean()

# พิมพ์ค่าเฉลี่ยของคอลัมน์ที่

print(mean_values0)

# พล็อตกราฟแท่งแสดงค่าเฉลี่ย

plt.figure(figsize=(10, 6))

bars = plt.bar(mean_values1.index, mean_values1, color='blue')

# เพิ่มค่าตัวเลขบนแต่ละแท่ง

for bar in bars:

    yval = bar.get_height()

    plt.text(bar.get_x() + bar.get_width()/2, yval, round(yval, 2), va='bottom')

plt.xlabel('Columns')

plt.ylabel('Mean Values')

plt.title('Mean Values of Columns for Cluster 1')

plt.xticks(rotation=45)

plt.show()
```

ภาคนวก ก.16 ชุดคำสั่งที่ใช้ในการวิเคราะห์ระดับร้านค้าของกลุ่มที่ 2

```
import matplotlib.pyplot as plt

# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 1

cluster_1_data = df[df['Cluster'] == 1]

# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Rank' สำหรับคลัสเตอร์ที่ 1

unique_format_counts_cluster_1 = cluster_1_data['Rank'].value_counts()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารฉบับนี้ทุกครั้งที่มีการนำไปใช้

```
# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Format' สำหรับคลัสเตอร์ที่ 1
plt.figure(figsize=(12, 8))
bars = unique_format_counts_cluster_1.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('Rank')
plt.ylabel('Count')
plt.title('Count of Rank for Cluster 1')
plt.xticks(rotation=0)
plt.show()
```

ภาคนวก ก.17 ชุดคำสั่งที่ใช้ในการวิเคราะห์การผลิตสินค้าเองของกลุ่มที่ 2

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 1
cluster_1_data = df[df['Cluster'] == 1]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Produce' สำหรับคลัสเตอร์ที่ 1
unique_produce_counts_cluster_1 = cluster_1_data['Produce'].value_counts()

# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Produce' สำหรับคลัสเตอร์ที่ 1
plt.figure(figsize=(12, 8))
bars = unique_produce_counts_cluster_1.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('Produce')
plt.ylabel('Count')
plt.title('Count of Produce for Cluster 1')
plt.xticks(rotation=0)
plt.show()
```

ภาคนวก ก.18 ชุดคำสั่งที่ใช้ในการวิเคราะห์ประเภทร้านค้าของกลุ่มที่ 2

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 1
cluster_1_data = df[df['Cluster'] == 1]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Type' สำหรับคลัสเตอร์ที่ 01
unique_type_counts_cluster_1 = cluster_1_data['Type'].value_counts()
# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Type' สำหรับคลัสเตอร์ที่ 1
plt.figure(figsize=(12, 8))
bars = unique_type_counts_cluster_1.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('Type')
plt.ylabel('Count')
plt.title('Count of Type for Cluster 1')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plt.xticks(rotation=0)
plt.show()
```

ภาคนวก ก.19 ชุดคำสั่งที่ใช้ในการวิเคราะห์พื้นที่ตั้งร้านค้าของกลุ่มที่ 2

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 1
cluster_1_data = df[df['Cluster'] == 1]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Region' สำหรับคลัสเตอร์ที่ 1
unique_region_counts_cluster_1 = cluster_1_data['Region'].value_counts()
# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Region' สำหรับคลัสเตอร์ที่ 1
plt.figure(figsize=(12, 8))
bars = unique_region_counts_cluster_1.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('Region')
plt.ylabel('Count')
plt.title('Count of Region for Cluster 1')
plt.xticks(rotation=0)
plt.show()
```

ภาคนวก ก.20 ชุดคำสั่งที่ใช้ในการวิเคราะห์ลักษณะที่ตั้งร้านค้าของกลุ่มที่ 2

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 1
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
cluster_0_data = df[df['Cluster'] == 1]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'hall' สำหรับคลัสเตอร์ที่ 1
unique_foodhall_counts_cluster_1 = cluster_1_data['hall'].value_counts()
# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'hall' สำหรับคลัสเตอร์ที่ 1
plt.figure(figsize=(12, 8))
bars = unique_foodhall_counts_cluster_1.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('hall')
plt.ylabel('Count')
plt.title('Count of hall for Cluster 1')
plt.xticks(rotation=0)
plt.show()
```

ภาคนวก ก.21 ชุดคำสั่งที่ใช้ในการวิเคราะห์เปอร์เซ็นที่ขายได้ของเบเกอรี่แต่ละชนิดกลุ่มที่ 2

```
# สร้าง DataFrame ที่มีเฉพาะ cluster 1
cluster_1_data = df[df['Cluster'] == 1]
        columns_of_interest = [A_Percentage', 'B_Percentage',      'C_Percentage',
    'D_Percentage', 'E_Percentage']
# หาค่าเฉลี่ยของคอลัมน์
mean_values1 = cluster_1_data[columns_of_interest].mean()
# พิมพ์ค่าเฉลี่ยของคอลัมน์ที่
print(mean_values1)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
# พล็อตกราฟแท่งแสดงค่าเฉลี่ย
plt.figure(figsize=(10, 6))
bars = plt.bar(mean_values1.index, mean_values1, color='blue')
# เพิ่มค่าตัวเลขบนแต่ละแท่ง
for bar in bars:
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2, yval, round(yval, 2), va='bottom')
plt.xlabel('Columns')
plt.ylabel('Mean Values')
plt.title('Mean Values of Columns for Cluster 1')
plt.xticks(rotation=45)
plt.show()
```

ภาคนวก ก.22 ชุดคำสั่งที่ใช้ในการวิเคราะห์ระดับร้านค้าของกลุ่มที่ 3

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 2
cluster_2_data = df[df['Cluster'] == 2]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Rank' สำหรับคลัสเตอร์ที่ 2
unique_format_counts_cluster_2 = cluster_2_data['Rank'].value_counts()
# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Format' สำหรับคลัสเตอร์ที่ 2
plt.figure(figsize=(12, 8))
bars = unique_format_counts_cluster_2.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('Rank')
plt.ylabel('Count')
plt.title('Count of Rank for Cluster 2')
plt.xticks(rotation=2)
plt.show()
```

ภาคนวก ก.23 ชุดคำสั่งที่ใช้ในการวิเคราะห์การผลิตสินค้าเองของกลุ่มที่ 3

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 2
cluster_2_data = df[df['Cluster'] == 2]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Produce' สำหรับคลัสเตอร์ที่ 2
unique_produce_counts_cluster_2 = cluster_2_data['Produce'].value_counts()

# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Produce' สำหรับคลัสเตอร์ที่ 3
plt.figure(figsize=(12, 8))
bars = unique_produce_counts_cluster_2.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('Produce')
plt.ylabel('Count')
plt.title('Count of Produce for Cluster 2')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plt.xticks(rotation=0)
plt.show()
```

ภาคนวก ก.24 ชุดคำสั่งที่ใช้ในการวิเคราะห์ประเภทร้านค้าของกลุ่มที่ 3

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 2
cluster_2_data = df[df['Cluster'] == 2]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Type' สำหรับคลัสเตอร์ที่ 2
unique_type_counts_cluster_2 = cluster_2_data['Type'].value_counts()
# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Type' สำหรับคลัสเตอร์ที่ 2
plt.figure(figsize=(12, 8))
bars = unique_type_counts_cluster_2.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('Type')
plt.ylabel('Count')
plt.title('Count of Type for Cluster 2')
plt.xticks(rotation=0)
plt.show()
```

ภาคนวก ก.25 ชุดคำสั่งที่ใช้ในการวิเคราะห์พื้นที่ตั้งร้านค้าของกลุ่มที่ 3

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 2
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
cluster_0_data = df[df['Cluster'] == 2]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Region' สำหรับคลัสเตอร์ที่ 2
unique_region_counts_cluster_2 = cluster_2_data['Region'].value_counts()
# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'Region' สำหรับคลัสเตอร์ที่ 2
plt.figure(figsize=(12, 8))
bars = unique_region_counts_cluster_2.plot(kind='bar', stacked=True)
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('Region')
plt.ylabel('Count')
plt.title('Count of Region for Cluster 2')
plt.xticks(rotation=0)
plt.show()
```

ภาคนวก ก.26 ชุดคำสั่งที่ใช้ในการวิเคราะห์ลักษณะที่ตั้งร้านค้าของกลุ่มที่ 3

```
import matplotlib.pyplot as plt
# กระจายข้อมูลเฉพาะในคลัสเตอร์ที่ 2
cluster_2_data = df[df['Cluster'] == 2]
# หาจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'hall' สำหรับคลัสเตอร์ที่ 2
unique_foodhall_counts_cluster_2 = cluster_2_data['hall'].value_counts()
# สร้างกราฟแท่งเพื่อแสดงจำนวนของค่าที่ไม่ซ้ำกันของคอลัมน์ 'hall' สำหรับคลัสเตอร์ที่ 2
plt.figure(figsize=(12, 8))
bars = unique_foodhall_counts_cluster_2.plot(kind='bar', stacked=True)
```

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
# เพิ่มค่าจำนวนบนแต่ละแท่ง
for i, bar in enumerate(bars.paScore6hes):
    height = bar.get_height()
        plt.text(bar.get_x() + bar.get_width() / 2, height + 0.05, f'{int(height)}',
    ha='center', va='bottom')
plt.xlabel('hall')
plt.ylabel('Count')
plt.title('Count of hall for Cluster 2')
plt.xticks(rotation=0)
plt.show()
```

ภาคนวก ก.27 ชุดคำสั่งที่ใช้ในการวิเคราะห์เปอร์เซ็นที่ขายได้ของเบเกอรี่แต่ละชนิดกลุ่มที่ 3

```
# สร้าง DataFrame ที่มีเฉพาะ cluster 2
cluster_2_data = df[df['Cluster'] == 2]
        columns_of_interest = ['A _Percentage', 'B_Percentage',     'C_Percentage',
    'D_Percentage', 'E_Percentage']
# หาค่าเฉลี่ยของคอลัมน์
mean_values0 = cluster_2_data[columns_of_interest].mean()
# พิมพ์ค่าเฉลี่ยของคอลัมน์ที่
print(mean_values0)
# พล็อตกราฟแท่งแสดงค่าเฉลี่ย
plt.figure(figsize=(10, 6))
bars = plt.bar(mean_values2.index, mean_values1, color='blue')
# เพิ่มค่าตัวเลขบนแต่ละแท่ง
for bar in bars:
    yval = bar.get_height()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    plt.text(bar.get_x() + bar.get_width()/2, yval, round(yval, 2), va='bottom')
plt.xlabel('Columns')
plt.ylabel('Mean Values')
plt.title('Mean Values of Columns for Cluster 2')
plt.xticks(rotation=45)
plt.show()
```


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

ภาคนวก ข.1 ผลการวิเคราะห์การแจกแจงปรกติ

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Cluster	Statistic	df	Sig.	Statistic	df	Sig.
SumofNet_Total	0	.092	130	.009	.961	130	.001
	1	.159	23	.137	.882	23	.011
	2	.170	4	.	.988	4	.946

a. Lilliefors Significance Correction

ภาคนวก ข.2 ผลการวิเคราะห์การเปรียบเทียบความแตกต่างระหว่างกลุ่ม

Ranks

	cluster	N	Mean Rank
Score D	0	78	98.59
	1	64	38.66
	2	15	149.27
	Total	157	

Test Statistics[a,b]

	Score D
Kruskal-Wallis H	100.699
df	2
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: cluster

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

งานทะเบียนคณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

คำรับรองเล่มสหกิจศึกษา

วันที่ 14 เดือน มิถุนายน พ.ศ. 2567

ข้าพเจ้า นางสาวอารียา วังสิโก รหัสนักศึกษา 63050694
นักศึกษาหลักสูตรวิทยาศาสตร์บัณฑิต สาขาวิชา สถิติประยุกต์ ภาควิชา สถิติขอรับรองว่าโครงงานสหกิจศึกษา เรื่อง

ชื่อภาษาไทย การแบ่งกลุ่มร้านค้าเบเกอรี่ กรณีศึกษาบริษัทแห่งหนึ่ง
ชื่อภาษาอังกฤษ Clustering of Bakery Shops : A Case Study of a Company
ปีการศึกษา 2566

เป็นผลงานวิจัยที่มิได้คัดลอกหรือละเมิดลิขสิทธิ์ของผู้อื่นและได้ผ่านการตรวจสอบความซ้ำซ้อนเรียบร้อยแล้วและได้แนบเอกสารการตรวจสอบการลอกเลียนงานวรรณกรรมที่ตรวจสอบจากเล่มโครงงานสหกิจศึกษาฉบับสมบูรณ์แล้ว
โปรแกรมอักขราวิสุทธิ์ 1.38 %

ลงชื่อ.....อารียา.....
(นางสาวอารียา วังสิโก)
นักศึกษา

ข้าพเจ้า ดร.สกุณา ศรีอโนมัย อาจารย์ที่ปรึกษาโครงงานสหกิจศึกษา ได้ตรวจสอบโครงงานสหกิจของนักศึกษาข้างต้นแล้ว ขอรับรองว่าเป็นผลงานวิจัยของนักศึกษาจริงและมีเนื้อหาสมบูรณ์ จึงลงชื่อไว้เป็นหลักฐาน

ลงชื่อ.....สกุณา.....
อาจารย์ที่ปรึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้