



การศึกษาการแบ่งกลุ่มผู้โดยสารสายการบินตามแบบประเมินความพึงพอใจด้วยเทคนิคทาง
วิทยาการข้อมูล

A STUDY OF AIRLINE PASSENGER SEGMENTATION BASED ON SATISFACTION
EVALUATION USING DATA SCIENCE TECHNIQUES

ปวิภาณ เกริกชัยวัน

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2567

การศึกษาการแบ่งกลุ่มผู้โดยสารสายการบินตามแบบประเมินความพึงพอใจด้วยเทคนิคทาง
วิทยาการข้อมูล



สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2567
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

A STUDY OF AIRLINE PASSENGER SEGMENTATION BASED ON SATISFACTION
EVALUATION USING DATA SCIENCE TECHNIQUES



A Master's Project Submitted in Partial Fulfillment of the Requirements
for the Degree of MASTER OF SCIENCE
(Data Science)

Faculty of Science, Srinakharinwirot University

2024

Copyright of Srinakharinwirot University

สารนิพนธ์

เรื่อง

การศึกษาการแบ่งกลุ่มผู้โดยสารสายการบินตามแบบประเมินความพึงพอใจด้วยเทคนิคทาง

วิทยาการข้อมูล

ของ

ปฏิภาณ เกริกชัยวัน

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าสารนิพนธ์

ที่ปรึกษาหลัก

(อาจารย์ ดร.โสภณ มงคลลักษณ์)

ประธาน

(รองศาสตราจารย์ ดร.อรรณพ หมั่นสกุล)

กรรมการ

(อาจารย์ ดร.ศุภร คนธภักดิ์)

ชื่อเรื่อง	การศึกษาการแบ่งกลุ่มผู้โดยสารสายการบินตามแบบประเมินความพึงพอใจด้วยเทคนิคทางวิทยาการข้อมูล
ผู้วิจัย	ปริญญ์ เกริกชัยวัน
ปริญญา	วิทยาศาสตร์มหาบัณฑิต
ปีการศึกษา	2567
อาจารย์ที่ปรึกษา	อาจารย์ ดร. โสภณ มงคลลักษณ์

ในอุตสาหกรรมสายการบินที่มีการแข่งขันสูง การทำความเข้าใจความพึงพอใจของผู้โดยสารเป็นปัจจัยสำคัญต่อความสำเร็จทางธุรกิจ งานวิจัยนี้ใช้เทคนิคการเรียนรู้ของเครื่องเพื่อระบุปัจจัยหลักที่ส่งผลต่อความพึงพอใจของผู้โดยสารและแบ่งกลุ่มลูกค้าสายการบินเพื่อปรับปรุงบริการให้เหมาะสมยิ่งขึ้น โดยใช้ชุดข้อมูลความพึงพอใจของผู้โดยสารสายการบินในสหรัฐอเมริกาที่เปิดเผยต่อสาธารณะ ผู้วิจัยใช้ Logistic Regression เพื่อวิเคราะห์ปัจจัยสำคัญที่ส่งผลต่อความพึงพอใจ และใช้ K-means clustering ในการแบ่งกลุ่มลูกค้า ผลการวิเคราะห์ชี้ให้เห็นว่า การเชื่อมต่ออินเทอร์เน็ตและบริการ Wi-Fi บนเที่ยวบินเป็นปัจจัยสำคัญที่สุดที่มีผลต่อความพึงพอใจ Logistic Regression แสดงประสิทธิภาพสูงในหลายตัวชี้วัด (Accuracy: 88%, Precision: 87.2%, Recall: 84.6%, F1-score: 85.9%) นอกจากนี้ การวิเคราะห์ยังระบุถึงกลุ่มผู้โดยสารที่แตกต่างกัน 2 กลุ่มซึ่งมีความต้องการและความคาดหวังที่แตกต่างกัน โดยทั้งสองกลุ่มแสดงความไม่พึงพอใจในบริการ Wi-Fi บนเที่ยวบิน ข้อมูลเชิงลึกเหล่านี้ช่วยให้สายการบินสามารถวางแผนปรับปรุงบริการ จัดสรรทรัพยากรอย่างมีประสิทธิภาพ และสร้างประสบการณ์การเดินทางที่เหมาะสมยิ่งขึ้นสำหรับผู้โดยสาร งานวิจัยในอนาคตควรพิจารณาข้อมูลพฤติกรรม ข้อมูลเฉพาะช่วงเวลา และโมเดลการวิเคราะห์อื่น ๆ เพื่อเพิ่มความเข้าใจเกี่ยวกับพฤติกรรมและความพึงพอใจของผู้โดยสาร

คำสำคัญ : ความพึงพอใจของผู้โดยสาร, การเรียนรู้ของเครื่อง, ความสำคัญของปัจจัย, การแบ่งกลุ่ม

Title	A STUDY OF AIRLINE PASSENGER SEGMENTATION BASED ON SATISFACTION EVALUATION USING DATA SCIENCE TECHNIQUES
Author	PATIPAN KRIRKCHAIWAN
Degree	MASTER OF SCIENCE
Academic Year	2024
Thesis Advisor	Dr. Sophon Mongkolluksamee

In the highly competitive airline industry, understanding passenger satisfaction is crucial for business success. This study leverages machine learning techniques to identify key drivers of passenger satisfaction and segment airline customers for targeted service improvements. Using a publicly available U.S. airline passenger satisfaction dataset, Logistic Regression was employed to identify significant factors affecting passenger satisfaction while K-means clustering was used to segment customers. The analysis revealed that online boarding and in-flight Wi-Fi service are the most significant factors influencing satisfaction. Logistic Regression demonstrated high performance across multiple metrics (accuracy: 88%, precision: 87.2%, recall: 84.6%, F1-score: 85.9%). Furthermore, two distinct passenger clusters were identified each with differing needs and expectations, though both expressed dissatisfaction with in-flight Wi-Fi. These insights enable airlines to prioritize service enhancements, optimize resource allocation, and personalize the customer experience. Future research should explore the integration of behavioral variables, time-specific data, and alternative models to enhance the understanding of passenger behavior and satisfaction.

Keyword : Passenger Satisfaction, Machine Learning, Feature Importance, Clustering

กิตติกรรมประกาศ

การจัดทำวิจัยฉบับนี้สำเร็จลุล่วงไปได้ด้วยดีจากการสนับสนุน ความรู้ความช่วยเหลือ คำแนะนำ ตลอดจนแนวทางในการทำวิจัยและจัดทำสารนิพนธ์ของ อ.ดร.โสภณ มงคลลักษณ์ (อาจารย์ที่ปรึกษา) รศ.ดร.อรรณพ หมั่นสกุล (ผู้ทรงคุณวุฒิภายนอกมหาวิทยาลัย) และคณาจารย์ทุกท่านในหลักสูตรวิทยาการข้อมูล ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ รวมทั้งการสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการนำเสนอผลงานวิจัย ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้



ปฎิภาณ เกริกชัยวัน

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ	ช
สารบัญตาราง.....	ญ
สารบัญรูปภาพ	ฎ
บทที่ 1 บทนำ.....	13
1.1 ที่มาและความสำคัญ	13
1.2 วัตถุประสงค์	13
1.3 ความสำคัญของงานวิจัย	13
1.4 ขอบเขตและข้อจำกัดงานวิจัย.....	14
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	14
บทที่ 2 ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง	16
2.1 การวิเคราะห์การถดถอยแบบโลจิสติก (Logistic Regression)	16
2.2 ทฤษฎีการวัดผลประสิทธิภาพของแบบจำลองทำนาย.....	18
2.3 การแบ่งกลุ่มลูกค้า (Customer Segmentation)	20
2.3.1 แบ่งตามประชากรศาสตร์ (Demographic).....	20
2.3.2 แบ่งตามพื้นที่ (Geographic).....	20
2.3.3 แบ่งตามหลักจิตวิทยา (Psychographic)	20
2.3.4 แบ่งตามพฤติกรรม (Behavioral)	20
2.4 ทฤษฎีการแบ่งกลุ่ม (Clustering Method).....	20

2.5 ทฤษฎีการวัดประสิทธิภาพและความเหมาะสมของการแบ่งกลุ่มของข้อมูล	22
2.5.1 Elbow Method.....	22
2.5.2 Silhouette Score	23
2.5.3 Davies Bouldin	23
2.6 งานวิจัยที่เกี่ยวข้อง	24
2.7 เครื่องมือและอัลกอริทึมที่ใช้ในงานวิจัย	31
2.7.1 เครื่องมือที่ใช้ในงานวิจัย	31
2.7.2 อัลกอริทึมที่ใช้ในงานวิจัย	31
บทที่ 3 วิธีการดำเนินงานวิจัย.....	33
3.1 กระบวนการทำงานของแบบจำลอง	33
3.2 การสืบค้นกลุ่มตัวอย่าง (กลุ่มผู้โดยสารที่มาใช้บริการสายการบิน)	35
3.3 การสำรวจข้อมูล (Exploratory Data Analysis: EDA).....	37
3.4 การเตรียมข้อมูล (Data Preprocessing)	45
3.4.1 การแปลงค่าประเภทของข้อมูล (Encoding)	45
3.4.2 การปรับช่วงค่าของพีเจอร์ (Normalization)	46
3.5 อัลกอริทึมแบบจำลองในการหาปัจจัยที่สำคัญ	46
3.6 อัลกอริทึมแบบจำลองในการแบ่งกลุ่มข้อมูล	47
3.7 อัลกอริทึมของแบบจำลองการทำนาย	48
บทที่ 4 ผลการศึกษา	49
4.1 ผลจากการหาปัจจัยที่สำคัญจาก Logistic Regression	49
4.2 ผลจากการแบ่งกลุ่มทำ K-Means Clustering	52
4.2.1 ผลลัพธ์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Elbow	52
4.2.2 ผลลัพธ์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Silhouette Score	52

4.2.3 ผลลัพธ์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Davies-Bouldin Index.....	53
4.3 ผลลัพธ์ค่าที่เหมาะสมต่อการแบ่งกลุ่ม.....	54
บทที่ 5 สรุปผลการวิจัย อภิปรายผลการวิจัย และข้อเสนอแนะ	55
5.1 สรุปผลการวิจัย.....	55
5.1.1 ผลการวิจัยจากการหาปัจจัยสำคัญจาก Logistic Regression.....	55
5.1.2 ผลจากการแบ่งกลุ่มด้วยวิธี K-Means	55
5.1.3 ผลลัพธ์ข้อมูลส่วนตัวและปัจจัยที่ควบคุมไม่ได้ของแต่ละกลุ่ม.....	55
5.1.4 ผลลัพธ์การให้บริการของแต่ละกลุ่ม	59
5.2 อภิปรายผลการวิจัย	61
5.2.1 อภิปรายผลลัพธ์การหาปัจจัยที่สำคัญต่อความพึงพอใจ.....	61
5.2.2 อภิปรายผลลัพธ์การแบ่งกลุ่มผู้โดยสารที่ไม่พึงพอใจตามพฤติกรรมการให้คะแนน	61
5.2.3 อภิปรายผลลัพธ์ที่สอดคล้องกัน	62
5.3 ข้อเสนอแนะ	63
ภาคผนวก.....	64
บรรณานุกรม	68
ประวัติผู้เขียน.....	71

สารบัญตาราง

	หน้า
ตาราง 1 แสดง Confusion Matrix.....	18
ตาราง 2 แสดงตัวแปรของข้อมูลส่วนตัวของผู้โดยสาร.....	36
ตาราง 3 แสดงตัวแปรของไฟล์ท.....	36
ตาราง 4 แสดงตัวแปรของการบริการของสายการบิน.....	37
ตาราง 5 แสดงผลลัพธ์การทำนายของ Logistic Regression.....	49
ตาราง 6 แสดงผลลัพธ์การทำนายของ Logistic Regression.....	50
ตาราง 7 แสดงค่าสัมประสิทธิ์ของแต่ละปัจจัย.....	51
ตาราง 8 แสดงประสิทธิภาพการแบ่งกลุ่มของจำนวน K กลุ่ม.....	54

สารบัญรูปภาพ

หน้า

ภาพประกอบ 1 หลักการตัดสินใจของ Logistic Regression	17
ภาพประกอบ 2 เทคนิค Elbow Method สำหรับแบบจำลอง K-Means.....	23
ภาพประกอบ 3 แสดงขั้นตอนการสร้างแบบจำลอง เพื่อหาปัจจัยที่สำคัญต่อความพึงพอใจของ ผู้โดยสาร	34
ภาพประกอบ 4 แสดงขั้นตอนการสร้างแบบจำลองการแบ่งกลุ่ม.....	35
ภาพประกอบ 5 แสดงจำนวน Missing Value ในแต่ละตัวแปร	38
ภาพประกอบ 6 แสดงความหนาแน่นของข้อมูล (Destiny) ของ Arrival Delay in Minutes.....	38
ภาพประกอบ 7 แสดงกราฟ Box Plot ของระยะทางของเที่ยวบิน	39
ภาพประกอบ 8 แสดงกราฟ Box Plot ของความล่าช้าในการออกเดินทาง.....	39
ภาพประกอบ 9 แสดงจำนวนของผู้โดยสารที่ไม่พึงพอใจตามระดับชั้นที่นั่ง (Class)	40
ภาพประกอบ 10 แสดงจำนวนของผู้โดยสารที่ไม่พึงพอใจตามเพศ (Gender)	40
ภาพประกอบ 11 แสดงจำนวนของผู้โดยสารที่ไม่พึงพอใจตามประเภทการท่องเที่ยว	41
ภาพประกอบ 12 แสดงจำนวนของผู้โดยสารที่ไม่พึงพอใจตามประเภทของผู้โดยสาร	41
ภาพประกอบ 13 แสดงความหนาแน่นของจำนวนของผู้โดยสารที่ไม่พึงพอใจตามอายุ.....	42
ภาพประกอบ 14 แสดงความหนาแน่นของจำนวนของผู้โดยสารที่ไม่พึงพอใจตามความล่าช้าใน การออกเดินทาง.....	42
ภาพประกอบ 15 แสดงจำนวนคะแนนของบริการภาคพื้นที่ได้จากการประเมินของผู้โดยสาร	43
ภาพประกอบ 16 แสดงจำนวนคะแนนของบริการบนไฟล์ที่ได้จากการประเมินของผู้โดยสาร ...	44
ภาพประกอบ 17 แสดงความสัมพันธ์กันของแต่ละตัวแปร.....	44
ภาพประกอบ 18 แสดงตัวอย่างข้อมูลหลังการทำ Encoding	45
ภาพประกอบ 19 แสดงข้อมูลหลังจากผ่าน StandardScaler	46

ภาพประกอบ 20	ปัจจัยที่ส่งผลกระทบต่อความพึงพอใจของผู้โดยสาร.....	46
ภาพประกอบ 21	แสดงการหาจำนวนกลุ่มที่ดีที่สุดด้วยวิธี Elbow Method.....	47
ภาพประกอบ 22	แสดงตัวอย่างข้อมูลหลังจากการทำ Normalization	49
ภาพประกอบ 23	แสดงปัจจัยที่สำคัญที่ส่งผลกระทบต่อความพึงพอใจ	50
ภาพประกอบ 24	แสดงผลลัพธ์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Elbow Method	52
ภาพประกอบ 25	แสดงผลลัพธ์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Silhouette Score	53
ภาพประกอบ 26	แสดงผลลัพธ์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Davies-Bouldin Index.....	53
ภาพประกอบ 27	จำนวนผู้โดยสารของทั้ง 2 กลุ่ม	56
ภาพประกอบ 28	จำนวนผู้โดยสารของทั้ง 2 กลุ่ม ตามประเภทของชั้นโดยสาร	56
ภาพประกอบ 29	จำนวนผู้โดยสารของทั้ง 2 กลุ่ม ตามประเภทการเดินทาง.....	57
ภาพประกอบ 30	จำนวนผู้โดยสารของทั้ง 2 กลุ่ม ตามเพศของผู้โดยสาร	58
ภาพประกอบ 31	ค่าเฉลี่ยอายุและระยะทางของเที่ยวบินของทั้ง 2 กลุ่ม.....	58
ภาพประกอบ 32	ค่าเฉลี่ยคะแนนการให้บริการภาคพื้นต่อผู้โดยสารคลัสเตอร์ 0.....	59
ภาพประกอบ 33	ค่าเฉลี่ยคะแนนการให้บริการภาคพื้นต่อผู้โดยสารคลัสเตอร์ 1.....	59
ภาพประกอบ 34	ค่าเฉลี่ยคะแนนการให้บริการบนเครื่องต่อผู้โดยสารคลัสเตอร์ 0	60
ภาพประกอบ 35	ค่าเฉลี่ยคะแนนการให้บริการบนเครื่องต่อผู้โดยสารคลัสเตอร์ 1	60

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญ

ในปัจจุบันการเดินทางด้วยเครื่องบินมีความสำคัญในชีวิตประจำวันอย่างมาก เนื่องจากความสะดวก รวดเร็ว และปลอดภัย อีกทั้งการแข่งขันในด้านอุตสาหกรรมการบินที่เพิ่มสูงขึ้นส่งผลให้ราคาค่าโดยสารลดลงซึ่งเปิดโอกาสให้ผู้โดยสารมีตัวเลือกมากขึ้นและกระตุ้นให้ธุรกิจสายการบินต้องปรับกลยุทธ์เพื่อรักษาความได้เปรียบในตลาด การพัฒนาคุณภาพการบริการจึงกลายเป็นกลยุทธ์ที่สำคัญในการตอบสนองต่อความคาดหวังของผู้โดยสารและเพื่อให้สายการบินสามารถสร้างความโดดเด่นจากคู่แข่งได้โดยเฉพาะในด้านการบริการที่ดีเยี่ยมซึ่งรวมถึงความสะดวกสบาย การตรงต่อเวลา คุณภาพของบริการทั้งภาคพื้นและบนเครื่องบิน รวมถึงประสบการณ์ที่ผู้โดยสารได้รับโดยการให้บริการที่ดีมากกว่าคู่แข่งสามารถเป็นปัจจัยสำคัญในการดึงดูดผู้โดยสารและสร้างความภักดีแก่ลูกค้าได้

1.2 วัตถุประสงค์

1.2.1 เพื่อจัดกลุ่มผู้โดยสารตามความคล้ายคลึงกันในการประเมินการให้บริการของสายการบิน

1.2.2 วิเคราะห์ปัจจัยหลักที่ส่งผลต่อความพึงพอใจของผู้โดยสาร ด้วยเทคนิคการเรียนรู้ของเครื่อง (Supervised Machine Learning)

1.2.3 เพื่อสามารถนำผลจากการจัดกลุ่มไปให้สายการบินใช้ในการวางแผนงานด้านการพัฒนาการบริการ และแผนงานการดูแลลูกค้าหลังการเดินทางได้

1.3 ความสำคัญของงานวิจัย

งานวิจัยนี้มีความสำคัญในการทำความเข้าใจปัจจัยที่มีผลต่อความพึงพอใจของผู้โดยสารสายการบินรวมถึงการแบ่งกลุ่มผู้โดยสารตามการประเมินบริการซึ่งช่วยให้เราเข้าใจถึงความต้องการของผู้โดยสารในแง่ของบริการได้อย่างชัดเจนการแบ่งกลุ่มผู้โดยสารที่ไม่พึงพอใจตามลักษณะการประเมินช่วยให้สามารถระบุเจาะจงถึงปัญหาและข้อจำกัดต่างๆที่เกิดขึ้นในแต่ละกลุ่ม เช่น ปัญหาเกี่ยวกับคุณภาพการบริการ ซึ่งเป็นข้อมูลที่สำคัญสำหรับการปรับปรุงและพัฒนาคุณภาพบริการของสายการบินการใช้ข้อมูลเหล่านี้สำหรับการวางแผนพัฒนาและปรับปรุงการบริการจึงช่วยเพิ่มประสิทธิภาพและการตอบสนองต่อความต้องการของผู้โดยสารได้ดียิ่งขึ้น

ในงานวิจัยนี้ผู้วิจัยใช้เทคนิคการเรียนรู้ในรูปแบบที่มีผู้สอนเพื่อระบุปัจจัยสำคัญที่มีผลต่อความพึงพอใจของผู้โดยสารที่มาใช้บริการสายการบินและใช้การเรียนรู้ในรูปแบบที่ไม่มีผู้สอนในการจัดกลุ่มผู้โดยสารซึ่งทำให้สามารถตีความและเข้าใจลักษณะของปัจจัยที่สำคัญในแต่ละกลุ่มได้อย่างชัดเจนผลลัพธ์จากการวิจัยนี้จะช่วยให้สายการบินเข้าใจถึงความต้องการของผู้โดยสารที่มาใช้บริการและสามารถนำข้อมูลเหล่านี้ไปใช้ในการพัฒนาการบริการและรักษาผู้โดยสารให้พึงพอใจและทำให้ผู้โดยสารกลับมาใช้บริการอีกครั้งทั้งยังช่วยเพิ่มประสิทธิภาพในการแข่งขันในตลาดการบินที่มีความท้าทายสูง โดยการนำเสนอข้อมูลที่ชัดเจนและน่าเชื่อถือจะช่วยสร้างความมั่นใจในผลลัพธ์และการตัดสินใจสำหรับการบริการสายการบินได้ดียิ่งขึ้น

1.4 ขอบเขตและข้อจำกัดงานวิจัย

งานวิจัยนี้มุ่งเน้นการศึกษาปัจจัยสำคัญที่ส่งผลต่อความพึงพอใจของผู้โดยสารในสายการบินที่ให้บริการเต็มรูปแบบ (Full Service) โดยการวิเคราะห์ข้อมูลจากกลุ่มผู้โดยสารที่มีความคล้ายคลึงกันในการประเมินบริการเพื่อให้ได้ข้อมูลที่สามารถนำไปใช้ในการปรับปรุงและวางแผนพัฒนาการบริการของสายการบิน ผู้วิจัยใช้ฐานข้อมูลของผู้โดยสารสายการบินในสหรัฐฯ ที่มีข้อมูลการให้คะแนนความพึงพอใจซึ่งข้อมูลนี้ได้รับการยอมรับในวงกว้างและช่วยให้สามารถเปรียบเทียบผลการศึกษากับงานวิจัยอื่นๆ ได้อย่างมีประสิทธิภาพนอกเหนือจากนี้การใช้ข้อมูลที่มีอยู่แล้วช่วยลดเวลาและทรัพยากรในการเก็บข้อมูลใหม่

ในการศึกษาผู้วิจัยใช้เทคนิคต่างๆ เช่น การใช้แบบจำลอง Logistic Regression เพื่อหาปัจจัยต่างๆ ที่มีผลสำคัญความพึงพอใจของผู้โดยสารและการใช้เทคนิคการแบ่งกลุ่ม (Clustering) เพื่อจำแนกกลุ่มผู้โดยสารตามคุณลักษณะการประเมินบริการซึ่งจะช่วยให้สามารถวิเคราะห์และเข้าใจความพึงพอใจของผู้โดยสารในแต่ละกลุ่มได้ดีขึ้นข้อมูลที่ได้จากการวิเคราะห์นี้จะเป็นประโยชน์สำหรับการพัฒนาสายการบินในการบริการและวางแผนพัฒนาการบริการให้ดียิ่งขึ้น

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 การแบ่งกลุ่มผู้โดยสารทำให้สายการบินสามารถกำหนดกลยุทธ์และแผนงานในการดูแลและบริการลูกค้าในแต่ละกลุ่มได้อย่างเหมาะสมซึ่งช่วยเพิ่มประสิทธิภาพการเดินทางที่ดีและตอบโต้ความต้องการของแต่ละกลุ่มได้อย่างตรงจุด

1.5.2 การแบ่งกลุ่มผู้โดยสารช่วยให้เข้าใจปัจจัยต่าง ๆ ที่มีผลต่อความพึงพอใจของผู้โดยสาร เช่น การบริการเช็คอิน การจัดการสัมภาระ คุณภาพอาหารและเครื่องดื่ม ความ

สะดวกสบายของที่นั่งโดยสาร รวมถึงการบริการจากพนักงานต้อนรับบนเครื่องบินข้อมูลเหล่านี้จะเป็นข้อมูลที่มีคุณค่าสำหรับการปรับปรุงและเพิ่มประสิทธิภาพการบริการของสายการบิน

1.5.3 การแบ่งกลุ่มช่วยให้สายการบินสามารถนำข้อมูลเชิงลึกมาใช้ในการวางแผนการตลาดและการดูแลลูกค้าหลังการเดินทาง โดยสามารถสร้างความสัมพันธ์อันดีกับผู้โดยสารและเสนอสิ่งจูงใจต่าง ๆ เพื่อปรับปรุงประสบการณ์การเดินทางของผู้โดยสารและเพิ่มโอกาสในการกลับมาใช้บริการในอนาคต



บทที่ 2

ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

งานวิจัยในบทนี้จะกล่าวถึงทฤษฎีวรรณกรรมและงานวิจัยที่เกี่ยวข้อง เพื่อเป็นแนวทางในการศึกษาเรียนรู้และวิเคราะห์ข้อมูลของผู้โดยสารสายการบินในธุรกิจการบิน เริ่มต้นจากการศึกษาแนวทางการหาปัจจัยที่มีความสำคัญด้วยวิธีการของ Supervised Machine Learning ซึ่งจะช่วยในการระบุฟีเจอร์ที่ส่งผลกระทบต่อความพึงพอใจของผู้โดยสารต่อการแบ่งกลุ่มลูกค้าผ่านการวิเคราะห์ Clustering เพื่อหากลุ่มเป้าหมายที่มีลักษณะและพฤติกรรมคล้ายคลึงกัน

งานวิจัยนี้ยังให้ความสำคัญกับทฤษฎีการวัดผลประสิทธิภาพของการทำนายและความเหมาะสมของการแบ่งกลุ่ม สุดท้ายกล่าวถึงงานวิจัยที่เกี่ยวข้องในด้านนี้และเครื่องมือรวมถึงอัลกอริทึมที่ใช้ในการวิเคราะห์และการวิจัยเพื่อการศึกษาเป็นไปโดยมีความน่าเชื่อถือและมีประสิทธิภาพ

2.1 การวิเคราะห์การถดถอยแบบโลจิสติก (Logistic Regression)

แบบจำลอง Logistic Regression เป็นเทคนิคการจำแนกหมวดหมู่ (Classification) ที่ใช้ในอัลกอริทึมการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยอาศัยข้อมูลที่มีเป้าหมายหรือผลลัพธ์ที่กำหนดไว้ล่วงหน้า Logistic Regression เป็นวิธีการที่มาจากทฤษฎีความน่าจะเป็น ซึ่งใช้อัลกอริทึมที่เหมาะสมสำหรับการทำนายผลลัพธ์ที่เป็นประเภท (Class) เช่น การพยากรณ์ว่าลูกค้าจะพึงพอใจหรือไม่พึงพอใจ โดยแบบจำลองนี้จะใช้ Sigmoid Function หรือ Logistic Function เพื่อแปลงค่าผลลัพธ์ให้อยู่ในช่วง 0 ถึง 1 ซึ่งสะท้อนถึงความน่าจะเป็นแต่ละประเภทในการจำแนกข้อมูล และ Logistic Regression สามารถแบ่งออกเป็นสองประเภทหลักตามชนิดของตัวแปรตาม ดังนี้

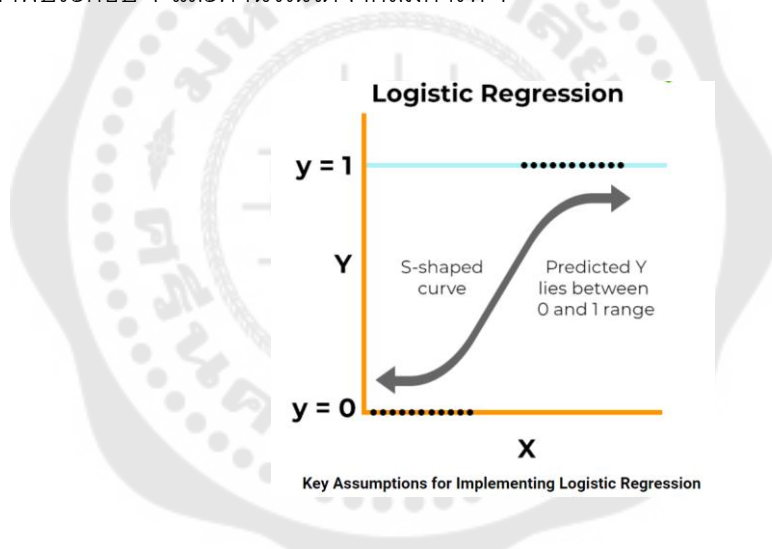
2.1.1 Binary Logistic Regression

ใช้ในกรณีที่ตัวแปรตามมีสองกลุ่มย่อย (Dichotomous Variable) เช่น ตัวแปรที่มีค่าเป็น 0 หรือ 1 เช่น การประเมินความพึงพอใจของผู้โดยสาร เป็นรูปแบบการวิเคราะห์พื้นฐานที่เหมาะสมสำหรับข้อมูลที่ต้องการแบ่งออกเป็นสองประเภทชัดเจน

2.1.2 Multinomial Logistic Regression

ใช้เมื่อข้อมูลมีตัวแปรตามมากกว่าสองกลุ่ม (Polytomous Variable) เช่น การประเมินคุณภาพบริการของธุรกิจการบริการที่แบ่งเป็นระดับพอใจ ปานกลาง และไม่พอใจ เหมาะกับการจัดการกับตัวแปรที่มีหลายคลาส

วิธีการทำงานของ Logistic Regression คือการสร้างสมการเส้นตรงเพื่อทำการประเมินความสัมพันธ์ระหว่างตัวแปรต้น (Features) และผลลัพธ์ (Target) โดยเริ่มต้นที่การกำหนดค่าสัมประสิทธิ์ (Coefficients) สำหรับแต่ละตัวแปรต้นในสมการ จากนั้นแปลงสมการให้เป็นความน่าจะเป็นผ่านฟังก์ชัน Sigmoid ซึ่งเมื่อค่าความน่าจะเป็นสูงกว่า 0.5 จะถูกกำหนดให้เป็นประเภทหนึ่ง และหากต่ำกว่าหรือเท่ากับ 0.5 จะถูกกำหนดให้เป็นอีกประเภทหนึ่ง (Wang et al., 2023) ดังตัวอย่างภาพประกอบ 1 และคำนวณได้จากสมการที่ 1



ภาพประกอบ 1 หลักการตัดสินใจของ Logistic Regression

ที่มา <https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/>

$$P(y = 1|x_i) = \frac{e^{-(w^T x_i + b)}}{1 + e^{-(w^T x_i + b)}} \quad (1)$$

สมการที่ 1 แสดงสมการฟังก์ชันของ Logistic Regression

โดยที่ $P(Y = 1|X)$ คือ ความน่าจะเป็นที่ X อยู่ในคลาส 1

X คือ ข้อมูลที่ต้องการทำนาย

ตัวอย่างเช่น ในการแบ่งกลุ่มลูกค้าเพื่อการให้บริการสายการบิน เราสามารถสร้างแบบจำลองเพื่อพิจารณาว่าลูกค้าจะพึงพอใจหรือไม่ โดยอาจใช้ข้อมูลเช่น อายุ ชั้นโดยสาร และค่าโดยสารเป็นตัวแปรต้น และผลลัพธ์เป็น "พึงพอใจ" หรือ "ไม่พึงพอใจ" แบบจำลองจะช่วยให้สายการบินสามารถวิเคราะห์และเข้าใจได้ว่าปัจจัยใดที่ส่งผลต่อความพึงพอใจของลูกค้ามากที่สุด เช่น หากลูกค้าอายุน้อยกว่า 30 ปี และนั่งชั้น Business Class ก็มีแนวโน้มที่จะพึงพอใจ ซึ่งจะช่วยให้สายการบินปรับปรุงเกี่ยวกับการบริการให้ตอบโต้ตามความต้องการของลูกค้าแต่ละกลุ่มได้อย่างมีประสิทธิภาพ

2.2 ทฤษฎีการวัดผลประสิทธิภาพของแบบจำลองทำนาย

การวัดผลการทำงานของโมเดล Logistic Regression สำหรับจำแนกประเภท งานวิจัยนี้ นำเสนอการสร้างโมเดล Logistic Regression เพื่อแก้ปัญหการจำแนกประเภท ในกรณีที่ตัวแปรที่ต้องการทำนายอยู่ในรูปแบบของหมวดหมู่หรือคลาส การประเมินประสิทธิภาพของโมเดลดังกล่าวทำได้โดยการพิจารณาค่าต่าง ๆ เช่น Accuracy, Precision, Recall และ F1-Score ซึ่งทั้งหมดนี้สามารถคำนวณได้จาก Confusion Matrix ดังตาราง 1 แสดงให้เห็นถึงการใช้ Confusion Matrix

ตาราง 1 แสดง Confusion Matrix

	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

TP (True Positive)	คือ ทำนายว่าเป็นความจริงตรงกับสิ่งที่เกิดขึ้นว่าจริง
TN (True Negative)	คือ ทำนายว่าไม่เป็นความจริงตรงกับสิ่งที่เกิดขึ้นว่าไม่จริง
FP (False Positive)	คือ ทำนายว่าเป็นความจริงแต่ผลที่เกิดขึ้นคือไม่จริง
FN (False Negative)	คือ ทำนายว่าไม่เป็นความจริงแต่ผลที่เกิดขึ้นคือจริง

การวัดความแม่นยำและการวัดประสิทธิภาพโดยรวมของโมเดล ในการประเมินแบบนี้ค่า Accuracy ช่วยให้เห็นภาพรวมว่าความแม่นยำของโมเดลเป็นอย่างไรซึ่งคำนวณได้จากสมการที่ 2

$$Accuracy = \frac{TP + TN}{N} \quad (2)$$

สมการที่ 2 แสดงสมการการคำนวณ Accuracy

ในขณะที่ Precision ชี้ให้เห็นถึงความแม่นยำของการทำนายคลาสที่กำหนดดังสมการที่ 3

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

สมการที่ 3 แสดงสมการการคำนวณ Precision

Recall อัตราส่วนที่แบบจำลองที่ทำนายว่าจริงกับจำนวนข้อมูลที่เป็นจริงทั้งหมด ดังสมการที่ 4

$$Recall = \frac{TP}{(TP + FN)} \quad (4)$$

สมการที่ 4 แสดงสมการการคำนวณ Recall

ส่วน F1-Score เป็นการรวมกันของ Precision และ Recall เพื่อแสดงให้เห็นความสมดุลระหว่างทั้งสองอย่างคือเป็นค่าเฉลี่ยแบบฮาร์โมนิกระหว่าง Precision และ Recall ดังสมการที่ 5

$$F1\ Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (5)$$

สมการที่ 5 แสดงสมการการคำนวณ F1-Score

การวิเคราะห์ผลลัพธ์จากโมเดล Logistic Regression ด้วยวิธีนี้ทำให้สามารถวัดประสิทธิภาพของการทำนายได้อย่างละเอียดและช่วยให้เข้าใจความถูกต้องและความแม่นยำของการจำแนกประเภทในบริบทของงานวิจัยนี้

2.3 การแบ่งกลุ่มลูกค้า (Customer Segmentation)

การแบ่งกลุ่มลูกค้าเป็นขั้นตอนสำคัญในการวางแผนการตลาด ตามทฤษฎี STP (Segmentation, Targeting, Positioning) โดยเราเสนอการแบ่งกลุ่มลูกค้าหรือ Segmentation ที่เป็นกระบวนการที่ทำให้ธุรกิจสามารถแบ่งลูกค้าออกเป็นกลุ่มย่อย ๆ โดยมีลักษณะที่คล้ายคลึงกัน หรือเหมือนกัน (Khandelwal et al., 2020)

มีประเภทการแบ่งออกเป็น 4 ประเภทดังนี้

2.3.1 แบ่งตามประชากรศาสตร์ (Demographic) การแบ่งกลุ่มลูกค้าตามลักษณะทางสถิติเช่น เพศ, อายุ, ระดับการศึกษา, สถานะ และอาชีพ เพื่อให้เข้าใจลูกค้าในแง่มุมทางประชากร

2.3.2 แบ่งตามพื้นที่ (Geographic) การแบ่งกลุ่มลูกค้าตามพื้นที่ที่อยู่อาศัย เช่น ตำบล, อำเภอ, ประเทศ และความหนาแน่นของประชากรในแต่ละพื้นที่ เพื่อเข้าใจบริบททางภูมิศาสตร์ที่สำคัญต่อธุรกิจ

2.3.3 แบ่งตามหลักจิตวิทยา (Psychographic) การแบ่งกลุ่มลูกค้าตามลักษณะพฤติกรรม, ความเชื่อ, นิสัย, และความชอบส่วนตัว เพื่อเข้าใจประสบการณ์และความต้องการที่ลึกซึ้งของลูกค้า

2.3.4 แบ่งตามพฤติกรรม (Behavioral) การแบ่งกลุ่มลูกค้าตามพฤติกรรมการซื้อสินค้าของลูกค้า เช่น แรนด์สินค้าที่เลือกซื้อ, ความถี่ในการซื้อสินค้า, เวลาในการซื้อสินค้า, และประเภทของสินค้าที่ซื้อ เพื่อเข้าใจพฤติกรรมการเลือกซื้อสินค้าของลูกค้า

การแบ่งกลุ่มลูกค้าช่วยให้ธุรกิจมีข้อมูลที่ลึกซึ้งเกี่ยวกับลูกค้าแต่ละกลุ่มทำให้สามารถวางแผนจัดการการตลาดได้อย่างมีประสิทธิภาพ เช่น การเสนอแบรนด์ผลิตภัณฑ์หรือการบริการที่เข้ากับความต้องการของลูกค้าในแต่ละกลุ่ม, การสร้างความสัมพันธ์ที่ดี, การมองเห็นโอกาสในการขยายธุรกิจด้านลูกค้าและการสร้างรายได้เพิ่มให้กับธุรกิจ

2.4 ทฤษฎีการแบ่งกลุ่ม (Clustering Method)

Clustering เป็น Machine Learning Model ซึ่งอยู่ในประเภท Unsupervised ที่ไม่มี Target ซึ่งเป็นโมเดลที่สามารถนำไปใช้ในการจัดกลุ่มสำหรับข้อมูลที่ไม่เคยมีการจัดกลุ่มมาก่อน โดยแบ่งกลุ่มจากชุดข้อมูลตามความคล้ายกัน เช่น จัดกลุ่มลูกค้าจากพฤติกรรมหรือความต้องการที่คล้ายคลึงกันของลูกค้าตามลักษณะที่คล้ายคลึงกันจะเป็นกลุ่มลูกค้าประเภทเดียวกัน ซึ่งประเภทของ Clustering ที่ใช้ในการแบ่งกลุ่มมี 6 ประเภท ดังนี้

1. การแบ่งกลุ่มแบบมีลำดับชั้น (Hierarchical Clustering)
2. การแบ่งกลุ่มจากจุดข้อมูล (Centroid-based Clustering)
3. การแบ่งกลุ่มในรูปแบบการกระจาย (Distribution-based Clustering)
4. การแบ่งกลุ่มตามการกระจุกตัวของข้อมูล (Density-based Clustering)
5. การแบ่งกลุ่มแบบคลุมเครือ (Fuzzy Clustering)
6. การแบ่งกลุ่มของข้อมูลจากข้อจำกัด (Constraint-based Clustering)

ในงานวิจัยนี้ใช้เทคนิคการจัดกลุ่มข้อมูลโดยอิงตามจุดศูนย์กลาง (Centroid-based Clustering) ซึ่งวิธีนี้จะพิจารณาระยะห่างระหว่างแต่ละจุดข้อมูลกับจุดศูนย์กลางที่เป็นตัวแทนของแต่ละกลุ่ม (Centroid) ในการศึกษาี้ ใช้แบบจำลอง K-Means ซึ่งเป็นวิธีที่ได้รับความนิยมในการจัดกลุ่มข้อมูล เนื่องจากความเรียบง่ายของการแบ่งกลุ่มและประสิทธิภาพในการทำงาน จึงขออธิบายเฉพาะแบบจำลอง K-means Clustering ที่ใช้ในงานนี้

K-means Clustering เป็นเทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Machine Learning) โดยจะมีวิธีดำเนินการดังนี้ โดยเริ่มจากการกำหนดค่าของจำนวนกลุ่มก่อนด้วยแทนค่าจำนวนกลุ่มแทนค่า K หลังจากนั้นและในขั้นตอนถัดมาคือการคำนวณหาจุดข้อมูลแต่ละ Centroid ด้วยการกำหนดจุดขึ้นมาตามจำนวนกลุ่มก่อนหน้าที่เรากำหนดแทนค่า K กำหนดหลังจากนั้นทำการคำนวณระยะห่างตามตำแหน่งของข้อมูลและจุดศูนย์กลางซึ่งสามารถคำนวณได้จากสมการ Euclidean distance โดยแสดงจากสมการที่ 1 หลังจากนั้นจุด Centroid จะเคลื่อนตามค่าเฉลี่ยที่คำนวณ แล้วทำการหาค่าเฉลี่ยและปรับตำแหน่ง Centroid ใหม่ไปเรื่อยๆ จนกว่าค่าเฉลี่ยไม่มีการเปลี่ยนแปลงแล้วก็จะได้กลุ่มและข้อมูลในกลุ่มตามจำนวนที่กำหนด

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (6)$$

สมการที่ 6 แสดงสมการการคำนวณ Euclidean Distance

โดยที่ $d(\mathbf{x}, \mathbf{y})$ คือ ระยะห่างแบบ Euclidean ระหว่างจุด $\mathbf{x}$ และ $\mathbf{y}$

x_i และ y_i คือ คือพิกัดของจุด $\mathbf{x}$ และ $\mathbf{y}$ ในมิติที่ i

2.5 ทฤษฎีการวัดประสิทธิภาพและความเหมาะสมของการแบ่งกลุ่มของข้อมูล

ตัววัดประสิทธิภาพโมเดลเป็นวิธีการที่ใช้สำหรับการประเมินคุณภาพในการจัดกลุ่มข้อมูลรวมไปถึงประเมินและวัดประสิทธิภาพของโมเดลภายใน (Internal Metric) และภายนอก (External Metric) ในงานวิจัยนี้การวัดผลประสิทธิภาพของแบบจำลองแบบจัดกลุ่ม ด้วยตัวชี้วัดภายใน (Internal Metric) และเทคนิคการเลือกจำนวนคลัสเตอร์ที่เหมาะสม (Validation Method) จึงขออธิบายเฉพาะการวัดผลที่ใช้ในงานนี้

2.5.1 Elbow Method

ในโมเดล K-Means จะต้องระบุจำนวนคลัสเตอร์ที่จะใช้ และ Elbow Method เป็นวิธีที่นิยมในการหาจำนวนคลัสเตอร์ที่เหมาะสมกับชุดข้อมูล เทคนิคนี้ทำงานโดยการวิเคราะห์การลดค่าความคลาดเคลื่อน โดยวัดค่าความคลาดเคลื่อนรวม (Sum of Square Error) ซึ่งเกิดจากระยะห่างระหว่างจุดศูนย์กลางของกลุ่ม (Centroid) และข้อมูลนั้น จุดที่ความชันของกราฟเริ่มเปลี่ยนเป็นแนวราบแสดงถึงตำแหน่งที่เหมาะสมสำหรับการเลือกจำนวนคลัสเตอร์ (k) ที่เหมาะสมที่สุด โดยในจุดนี้ ค่า SSE (Sum of Squared Errors) จะมีค่าต่ำสุดสำหรับชุดข้อมูลที่กำลังวิเคราะห์ การใช้ Elbow Method เพื่อหาจำนวนกลุ่มที่เหมาะสมช่วยให้สามารถแบ่งกลุ่มลูกค้าได้อย่างมีประสิทธิภาพ โดยลดความเบี่ยงเบนในกลุ่มให้น้อยที่สุด ซึ่งส่งผลให้การวิเคราะห์ข้อมูลด้วย K-Means Clustering มีความเหมาะสมและแม่นยำสูงขึ้น (Nandapala & Jayasena, 2020)

$$SSE = \sum_{i=1}^n (y_i - y')^2$$

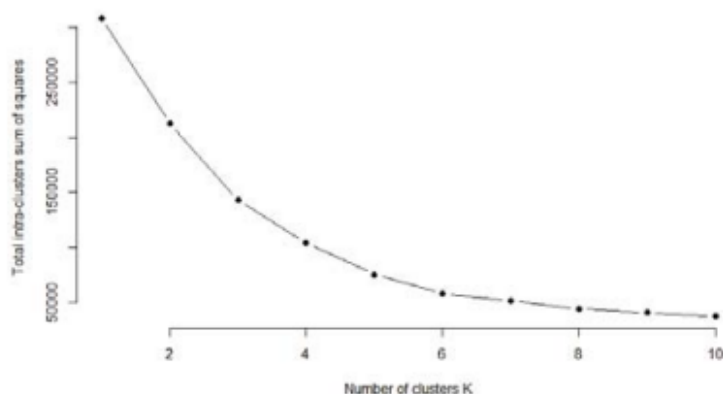
สมการที่ 7 แสดงสมการการคำนวณ Sum of Squared Errors (SSE)

โดยที่ n คือ จำนวนอินสแตนซ์

y_i คือ ค่าของอินสแตนซ์ที่ i

y' คือ ค่าเฉลี่ยของอินสแตนซ์

เทคนิค Elbow Method มักใช้ภาพเพื่อช่วยสำหรับการตัดสินใจเกี่ยวกับจำนวนคลัสเตอร์ที่เหมาะสม โดยจุดที่เหมาะสมกับการเลือกจำนวนคลัสเตอร์ มักจะอยู่ตรงจุดที่กราฟแสดงการ "หักศอก" นั่นคือ จุดที่ค่า SSE (Sum of Squared Errors) เริ่มลดลงน้อยลงเมื่อเพิ่มจำนวนคลัสเตอร์ ดังที่แสดงในภาพประกอบ 2



ภาพประกอบ 2 เทคนิค Elbow Method สำหรับแบบจำลอง K-Means

ที่มา <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9342639>

2.5.2 Silhouette Score

เทคนิคการหาคำตอบว่าควรแบ่งข้อมูลเป็นกี่กลุ่ม (ค่า k) ซึ่งเป็นองค์ประกอบพื้นฐานหลักของวิธีการจัดกลุ่ม ถึงแม้ว่าในบางเทคนิคไม่จำเป็น และมีการพัฒนาอัลกอริทึมที่ไม่ต้องกำหนดค่า k ก่อนแล้วก็ตาม แต่ในที่สุดผู้ใช้ออมต้องการระบุว่าจะแบ่งเป็นกี่กลุ่มจึงจะเหมาะสมสำหรับบางข้อมูล ซึ่ง Silhouette Score เป็นมาตรวัดที่อาศัยทั้งการยึดเหนี่ยวภายในและการแยกกันระหว่างกลุ่ม โดยใช้การหาค่าเฉลี่ยของระยะห่างระหว่างจุดกับกลุ่มที่อยู่ใกล้เคียงเทียบกับจุดที่อยู่ภายในกลุ่มเดียวกัน (Wang et al., 2023) ซึ่งคำนวณได้จากสมการที่ 8

$$S(i) = \begin{cases} 1 - \frac{a(i)}{b(i)}, & a(i) < b(i) \\ 0, & a(i) = b(i) \\ \frac{a(i)}{b(i)} - 1, & a(i) > b(i) \end{cases}$$

สมการที่ 8 แสดงสมการการคำนวณของ Silhouette Score

โดยที่ a คือ ค่าเฉลี่ยระยะห่างระหว่างตัวอย่างกับจุดข้อมูลอื่น ๆ ทั้งหมดในกลุ่มเดียวกัน

b คือ ค่าเฉลี่ยระยะห่างระหว่างตัวอย่างกับจุดข้อมูลทั้งหมดในกลุ่มถัดไปที่อยู่ใกล้ที่สุด

2.5.3 Davies Bouldin

เกณฑ์ที่ใช้ในการวัดคุณภาพในการจัดกลุ่มที่ใช้ในการวิเคราะห์เพื่อการแบ่งกลุ่มข้อมูล โดย Davies-Bouldin index คือ เป็นอัตราส่วนของข้อมูลระหว่างผลรวมของของการ

กระจายตัวของจุดข้อมูลในกลุ่มและระยะห่างระหว่างกลุ่ม จะเห็นว่าในการแบ่งกลุ่มที่ดีนั้น การกระจายตัวของข้อมูลภายในกลุ่มจะต้องน้อยและระยะห่างระหว่างแต่ละกลุ่มจะต้องมาก (Wang et al., 2023)

โดยสมการด้านล่างนี้วัดความถูกต้องของการจัดกลุ่ม โดยเป้าหมายคือการทำให้ระยะห่างระหว่างจุดข้อมูลในกลุ่มเดียวกันมีค่าน้อยที่สุด ในขณะที่ระยะห่างระหว่างกลุ่มต่างๆ มีค่ามากที่สุด ดังสมการที่ 9

$$db = \frac{1}{k} \sum_{i=1}^k R_i \quad (9)$$

สมการที่ 9 แสดงสมการการคำนวณประสิทธิภาพการจัดกลุ่มด้วยวิธี Davies Bouldin

โดยที่ k คือ กลุ่มข้อมูล

R_{ij} คือ อัตราส่วนระหว่างกลุ่ม i และกลุ่ม j

2.6 งานวิจัยที่เกี่ยวข้อง

การทบทวนวรรณกรรมของงานวิจัยนี้ได้ทำการศึกษางานวิจัยที่เกี่ยวข้องกับผลกระทบ คุณภาพการบริการภาคพื้นและบนเครื่อง ความคาดหวังของผู้โดยสาร และการแบ่งกลุ่มผู้โดยสาร มีรายละเอียดดังต่อไปนี้

บทความวิจัยเรื่อง Airline Service Quality Improvement Approach for Aviation Business (ชลิต เญียบพิมาย & วสันต์ นิลมัย, 2562)

บทความนี้เป็นการวิจัยเชิงคุณภาพโดยใช้วิธีการศึกษาจากเอกสาร หนังสือ ตำรา บทความ และงานวิจัยที่เกี่ยวข้องทำการวิเคราะห์ข้อมูลด้วยการวิเคราะห์เชิงเนื้อหาตามประเด็นที่ศึกษา คือ AIRQUAL Concept ประกอบด้วย การบริการบนเครื่องบินที่จับต้องได้ (Airline Tangibles: ATANG), การบริการภาคพื้นในสนามบินที่จับต้องได้ (Terminal tangibles: TTANG), บุคลากรผู้ให้บริการ (Personnel: PER), การเอาใจใส่ (Empathy: EMP), ภาพลักษณ์ (Image: IMG), การรับรู้คุณภาพการบริการ (Perceived Service Quality: PSQ) และความพึงพอใจของผู้มาใช้บริการ (Customer Satisfaction: CSAT) เพื่อเสนอแนวทางการพัฒนาคุณภาพการบริการในเชิงธุรกิจ

การศึกษาคุณภาพการบริการที่จับต้องได้ (Airline Tangibles: ATANG) ของสายการบินไทยพบว่าปัจจัยต่างๆ ที่เกี่ยวข้องกับสภาพแวดล้อมบนเครื่องบินส่งผลต่อพฤติกรรมการใช้บริการ

ของผู้โดยสาร ในประเด็นด้านคุณภาพบริการ ผู้โดยสารมีความคิดเห็นเชิงบวกเกี่ยวกับทุกหัวข้อ เช่น ความสะอาดและสุขอนามัยของห้องโดยสารและห้องนํ้าบนเครื่องบิน แรงดันอากาศและอุณหภูมิที่เหมาะสม ขนาดและความกว้างของที่นั่งโดยสารที่สบาย อุปกรณ์อำนวยความสะดวกที่อยู่ในสภาพดี อาหารและเครื่องดื่มบนเครื่องบินที่มีรสชาติดี มีมาตรฐาน และหลากหลาย รวมถึงปริมาณที่เหมาะสม นอกจากนี้ พื้นที่เก็บสัมภาระยังมีเพียงพอสำหรับการใช้งานในทุกเที่ยวบิน

ส่วนการศึกษาในการบริการภาคพื้นในสนามบินที่จับต้องได้ (Terminal Tangibles: TTANG) โดยการบริการประเภทนี้ต้องมีคุณลักษณะ 4 ประการดังนี้คือ ความรวดเร็ว ความสะดวก ความสบาย และต้องสร้างความมั่นใจให้แก่ผู้ใช้บริการ และการศึกษาเกี่ยวกับความพึงพอใจของผู้มาใช้บริการ (Customer Satisfaction: CSAT) ได้แบ่งระดับของทัศนคติของผู้ใช้บริการหรือผู้ใช้บริการเป็น 4 ระดับ ดังนี้

ระดับ 1 ความไม่พึงพอใจ คือ การที่ผู้ใช้บริการไม่ได้รับการบริการตามความคาดหวัง

ระดับ 2 ความพึงพอใจ คือ การที่ผู้ใช้บริการได้รับการที่เป็นไปตามความคาดหวัง

ระดับ 3 ความประทับใจ คือ การที่ผู้ใช้บริการได้รับการที่เกินกว่าความคาดหวัง

ระดับ 4 ความเชื่อมั่น คือ การที่ผู้ใช้บริการได้รับการ ที่สร้างความพึงพอใจและความประทับใจอย่างต่อเนื่อง จนเกิดความเชื่อมั่นในบริการที่สะท้อนถึงความตั้งใจที่จะกลับมาใช้บริการซ้ำ และมีแนวโน้มที่จะแนะนำบริการให้แก่ผู้อื่น

จากงานวิจัยนี้สามารถใช้ประกอบเป็นข้อมูลเกี่ยวกับการพัฒนาเชิงคุณภาพของการบริการในธุรกิจการบินโดยองค์กรหรือหน่วยงานทั้งภาครัฐและภาคเอกชนที่ดูแลรับผิดชอบในอุตสาหกรรมการบิน

บทความวิจัยเรื่อง Investigating Airline Passenger Satisfaction: Data Mining Method (Noviantoro & Huang, 2022)

งานวิจัยนี้ศึกษาแนวทางการทำ Data miningโดยใช้การเลือกฟีเจอร์ (Feature Selection) เพื่อกำหนดความพึงพอใจของผู้โดยสารสายการบิน โดยศึกษาจากข้อมูลความพึงพอใจของผู้โดยสารที่มาใช้บริการสายการบินในประเทศอเมริกา งานวิจัยนี้จะใช้เทคนิคการเลือกฟีเจอร์แบบเพิ่มฟีเจอร์ทีละตัว (Forward Selection) เพื่อคัดกรองฟีเจอร์ที่ไม่จำเป็นออกจากชุดข้อมูล

งานวิจัยนี้ได้ทดลองใช้อัลกอริทึมการจำแนกที่หลากหลาย ได้แก่ Decision Tree, Random Forest, Gradient Boosted Tree, K-NN, Naïve Bayes, Rule Induction, Logistic Regression, Neural Network, Deep Learning และ Support Vector Machine (SVM) และเทียบประสิทธิภาพด้วย Accuracy, Precision, Recall, F-Score และ ROC

ผลการศึกษาพบว่า พีเจอาร์การเช็คอินออนไลน์เป็นบริการที่สำคัญที่สุดสำหรับผู้โดยสารสายการบิน โดยได้จากการทำ Feature Importance โดยศึกษาจากค่า coefficient ที่ส่งผลในแบบจำลอง Logistic regression และ Deep learning เป็นอัลกอริทึมการเรียนรู้ของเครื่องจักรที่เหมาะสมที่สุดในแง่ของ Accuracy, F-score และกราฟ Receiver Operating Characteristic (ROC)

บทความวิจัยเรื่อง Understanding the Dynamics of the Quality of Airline Service Attributes: Satisfiers and Dissatisfiers (Park et al., 2020)

งานวิจัยนี้ศึกษาคุณลักษณะของบริการสายการบินที่ทำหน้าที่เป็นตัวสร้างความพึงพอใจหรือผู้ไม่พอใจสำหรับผู้โดยสาร เพื่อให้ข้อมูลเชิงลึกสำหรับผู้จัดการสายการบินเพื่อพัฒนากลยุทธ์การตลาดที่มุ่งเน้นลูกค้าและจัดลำดับความสำคัญของคุณลักษณะที่ก่อให้เกิดความพึงพอใจของลูกค้า โดยงานวิจัยได้รวบรวมและวิเคราะห์บทวิจารณ์ออนไลน์และการให้คะแนนของสายการบินในสหรัฐอเมริกา 20 สายการบินจาก TripAdvisor โดยใช้โปรแกรมรวบรวมข้อมูลอัตโนมัติที่พัฒนาใน Python

การศึกษานี้ใช้ตัวแปรขึ้นอยู่กับสองตัวแปรคือ การเบี่ยงเบนเชิงลบ (Negative Deviations) และการเบี่ยงเบนเชิงบวก (Positive Deviations) ซึ่งกำหนดโดยความแตกต่างระหว่างการจัดอันดับของแต่ละบุคคลสำหรับการเดินทางที่เฉพาะเจาะจงและโหมดการจัดอันดับโดยรวมสำหรับสายการบินและเส้นทางเดียวกัน และใช้แบบจำลอง Tobit เพื่อวิเคราะห์ปัจจัยกำหนดของการเบี่ยงเบนการจัดอันดับ และใช้ค่า p-value และ Wald Test บ่งบอกความสำคัญของการบริการได้

โดยผลจากการวิจัยพบว่า คุณลักษณะบางอย่างทำหน้าที่เป็นตัวสร้างความพึงพอใจ (เช่น ความสะอาด อาหารและเครื่องดื่ม ความบันเทิงระหว่างเที่ยวบิน) ในขณะที่บางอย่างยังเป็นบริการที่ไม่พึงพอใจพอ (เช่น การบริการลูกค้า การเช็คอินและการขึ้นเครื่อง) และพบว่า การโดยสารเที่ยวบินภายในประเทศเทียบกับระหว่างประเทศและชั้นการบริการ (Economy Class , Business Class) เป็นปัจจัยที่มีผลต่อการเบี่ยงเบนการจัดอันดับ

บทความวิจัย Airline Customer Satisfaction and Loyalty: Impact of In-Flight Service Quality Model (An & Noh, 2009)

การศึกษามีวัตถุประสงค์เพื่อระบุปัจจัยต่าง ๆ ของคุณภาพการบริการในเที่ยวบินที่มีความสำคัญตามระดับที่นั่งผู้โดยสารได้แก่ชั้นธุรกิจ (Business Class) และชั้นประหยัด

(Economy Class) เพื่อตรวจสอบผลกระทบของคุณภาพการบริการในเที่ยวบินต่อความพึงพอใจและความภักดีของลูกค้า (Loyalty) ซึ่งได้รวบรวมข้อมูลจากผู้โดยสารของบริษัทสายการบินชั้นนำระดับโลกซึ่งมีสำนักงานใหญ่ในเกาหลีใต้ โดยผู้โดยสารทั้งหมดเป็นลูกค้าที่ใช้เส้นทางอเมริกาเหนือ-เกาหลี หรือเส้นทางยุโรป-เกาหลี ซึ่งต้องใช้เวลาบินมากกว่า 10 ชั่วโมง แบบสอบถามทั้งหมด 150 รายการถูกรวบรวมจากชั้นธุรกิจ (Business Class) และแบบสอบถาม 370 รายการจากชั้นประหยัด (Economy Class) ผู้วิจัยแยกแบบสอบถามที่ไม่สมบูรณ์ออก 25 แบบสอบถาม และใช้แบบสอบถามที่ใช้งานได้ทั้งหมด 494 แบบสอบถาม โดยแบ่งเป็น 139 แบบสอบถามสำหรับชั้นธุรกิจ (Business Class) และ 355 แบบสอบถามสำหรับชั้นประหยัด (Economy Class)

ผู้วิจัยได้เลือกศึกษาความคาดหวังของผู้โดยสารที่มีต่อสายการบิน โดยเน้นในด้านความน่าเชื่อถือ (Reliability) การตอบสนอง การสร้างความมั่นใจ และความเข้าอกเข้าใจ (Responsiveness, Assurance, and Empathy) นอกจากนี้ ยังได้พัฒนาชุดตัวชี้วัดใหม่ที่ออกแบบมาเฉพาะเพื่อประเมินบริการด้านอาหารและเครื่องดื่มในเที่ยวบิน เช่น คุณภาพของอาหาร คุณภาพของเครื่องดื่มแอลกอฮอล์ และเครื่องดื่มไม่มีแอลกอฮอล์ เพื่อเพิ่มความครอบคลุมและความเหมาะสมในการประเมิน โดยแยกการวิเคราะห์ข้อมูลระหว่างกลุ่มผู้โดยสารชั้นธุรกิจ (Business Class) และชั้นประหยัด (Economy Class) เพื่อสร้างความแม่นยำในการวิเคราะห์คุณภาพการบริการในเที่ยวบิน และศึกษาผลกระทบที่เกิดขึ้นต่อความพึงพอใจและความภักดีของลูกค้า สำหรับกระบวนการวิเคราะห์ ผู้วิจัยได้นำการวิเคราะห์ถดถอย (Regression Analysis) มาใช้ โดยอ้างอิงทฤษฎีของ Parasuraman, Zeithaml และ Berry ซึ่งเป็นผู้พัฒนารูปแบบการประเมินคุณภาพการบริการในปี 1988 ซึ่งแบบจำลองดังกล่าวประกอบด้วยองค์ประกอบหลัก 4 ด้าน ได้แก่ ความน่าเชื่อถือ การตอบสนอง การสร้างความมั่นใจ และความเข้าอกเข้าใจ ซึ่งยังคงถูกนำมาใช้อย่างแพร่หลายในปัจจุบันโดยสายการบินต่างๆ สำหรับการประเมินคุณภาพการบริการของพวกเขา (Parasuraman et al., 1988)

ผลการทดลองพบว่าในชั้นธุรกิจ (Business Class) คุณภาพของเครื่องดื่มแอลกอฮอล์และไม่มีแอลกอฮอล์มีผลกระทบต่อความพึงพอใจของลูกค้าอย่างมาก คุณภาพอาหารมีอิทธิพลต่อความพึงพอใจของลูกค้ามากกว่ารูปแบบการนำเสนอ ในส่วนของความภักดี (Loyalty) ของผู้โดยสารชั้นธุรกิจ (Business Class) ส่วนสำคัญคือการตอบสนองและการเอาใจใส่ของพนักงาน ซึ่งส่งผลให้การบริการที่เอาใจใส่จากพนักงานต้อนรับบนเครื่องบินมีบทบาทสำคัญในการเลือกบริษัทสายการบินของผู้โดยสาร รวมถึงคุณภาพของเครื่องดื่มแอลกอฮอล์และไม่มี

แอลกอฮอล์ในส่วนของผู้โดยสารชั้นประหยัด (Economy Class) คุณภาพอาหารมีผลกระทบต่อความพึงพอใจของลูกค้ามากที่สุด ถัดจากคุณภาพอาหาร และเครื่องดื่มแอลกอฮอล์และไม่มีแอลกอฮอล์มีอิทธิพลต่อความพึงพอใจของลูกค้ารองลงมา ในส่วนของความภักดี (Loyalty) ผู้โดยสารชั้นประหยัด (Economy) ดูเหมือนจะรู้สึกภักดีต่อสายการบิน โดยปัจจัยสำคัญคือการบริการที่เอาใจใส่ของพนักงานต้อนรับบนเครื่องบินมากกว่าบริการในเที่ยวบิน (อาหารและเครื่องดื่ม) เนื่องจากใช้เวลาเดินทางนานหลายชั่วโมงในพื้นที่ที่แคบกว่าชั้นธุรกิจ (Business Class)

ผลลัพธ์เหล่านี้บ่งชี้ว่าการรับรู้คุณภาพการบริการค่อนข้างแตกต่างกันตามชั้นที่นั่ง (Class) โดยส่วนใหญ่เกิดจากการที่ผู้โดยสารชั้นธุรกิจ (Business Class) มีประสบการณ์เดินทาง และโปรไฟล์ (Profile) ที่แตกต่างจากผู้โดยสารชั้นประหยัด ปัจจัยด้านคุณภาพการบริการหลักที่ส่งผลต่อความพึงพอใจ (Satisfaction) และความภักดี (Loyalty) ของผู้โดยสารก็แตกต่างกันระหว่างชั้นธุรกิจ (Business Class) และชั้นประหยัด (Economy Class) ดังนั้นบริษัทสายการบินจึงต้องสร้างความแตกต่างให้กับกลยุทธ์ประเภทต่าง ๆ

บทความวิจัย Performance Comparison of K-Means and DBSCAN Methods for Airline Customer Segmentation (Sahinbas, 2022)

งานวิจัยนี้เล็งเห็นว่าการได้รับบริการที่ดีเยี่ยมในอุตสาหกรรมการบินสามารถสร้างความได้เปรียบทางการแข่งขันได้ การศึกษานี้มีวัตถุประสงค์เพื่อช่วยด้านการบินบริษัทต่าง ๆ เพื่อทราบว่าบริการของตนตอบสนองความต้องการของผู้โดยสารหรือไม่และเพื่อให้ผู้โดยสารได้รับความพึงพอใจ โดยการทดลองนี้ใช้ข้อมูลที่เป็นสาธารณะที่ได้จาก Kaggle ประกอบด้วยชุดข้อมูลจากตัวอย่างผู้โดยสารทั้งหมด 25,976 ราย โดยเป็นผู้โดยสารที่พอใจ 11,403 คน และผู้โดยสารที่ไม่พอใจ 14,573 คน เพื่อใช้ประเมินระดับความพึงพอใจของผู้โดยสาร

การศึกษานี้มีเป้าหมายเพื่อวิเคราะห์ความพึงพอใจของผู้โดยสารสายการบิน โดยนำวิธีการวิเคราะห์ข้อมูล (Data Mining) มาใช้ พร้อมทั้งใช้อัลกอริทึม K-Means และ DBSCAN เพื่อระบุปัจจัยสำคัญที่ผู้โดยสารให้ความสำคัญในด้านคุณภาพการบริการ การประเมินประสิทธิภาพของการจัดกลุ่มจะดำเนินการผ่านเมตริกต่างๆ เช่น Silhouette Coefficient, Calinski-Harabasz Index, Davies-Bouldin Index และ Elbow Method ซึ่งช่วยให้สามารถเลือกอัลกอริทึมที่เหมาะสมที่สุดได้อย่างแม่นยำ

ผลการทดสอบเปรียบเทียบประสิทธิภาพของอัลกอริทึมคลัสเตอร์ของ K-Means และ DBSCAN สำหรับการแบ่งกลุ่มลูกค้าของสายการบินอัลกอริทึม K-Means ได้ผลลัพธ์ที่ดีกว่าอัลกอริทึม DBSCAN เล็กน้อยที่มีค่า Silhouette 0.1450671 และ K-Means ยังได้ผลลัพธ์ที่ดีขึ้น

เล็กน้อยในค่าสัมประสิทธิ์ Silhouette, Calinski-Harabasz และ Davies Bouldin และได้ค่าอัลกอริทึมจำนวนกลุ่ม(K)ที่ดีที่สุดเท่ากับ 8 โดยแต่ละกลุ่มสามารถตรวจสอบจำนวนผู้โดยสารที่พึงพอใจ (Satisfier) และไม่พึงพอใจ (Unsatisfied) รวมถึงการบริการที่ได้คะแนนรวมมากที่สุดและน้อยที่สุด

บทความ A Study on Service Quality and Passenger Satisfaction on Indian Airlines (Archana & Subha, 2012)

งานวิจัยนี้มุ่งเน้นศึกษาปัจจัยด้านคุณภาพการบริการที่ส่งผลต่อระดับความพึงพอใจของผู้โดยสารในเที่ยวบิน โดยวิเคราะห์ว่าการบริการในแต่ละมิติ (Dimensions) มีผลกระทบเชิงบวกต่อคุณภาพการบริการอย่างไร และระบุว่ามิติใดมีอิทธิพลมากที่สุดและน้อยที่สุดต่อคุณภาพการบริการสำหรับการเดินทางทางอากาศระหว่างประเทศ โดยการศึกษาที่ใช้ข้อมูลจากกลุ่มตัวอย่างผู้โดยสารจำนวน 270 คน ซึ่งตอบแบบสอบถามเกี่ยวกับประสบการณ์การใช้บริการสายการบิน Indian Airlines โดยครอบคลุมระดับชั้นบริการ 3 ประเภท ได้แก่ ชั้นประหยัด (Economy Class), ชั้นธุรกิจ (Business Class) และชั้นพรีเมียม (First Class)

ผู้วิจัยได้ใช้หลักการ Factor Analysis (FA) ซึ่งสามารถระบุความสำคัญและอธิบายความแปรปรวนของข้อมูลได้ โดยทดสอบกับการบริการด้านต่าง ๆ (Dimensions) ได้แก่ การบริการบนเครื่อง (In-flight Services) , บริการเครื่องเล่นบนเครื่อง (In-flight Digital Services) และการบริการหลังบ้านของสายการบิน (Airline Back office Operations)

ผลการวิจัยระบุว่าคุณภาพการบริการบนเที่ยวบินมีความสำคัญแตกต่างกันตามระดับชั้นที่นั่ง โดยการบริการที่ผู้โดยสารให้ความสำคัญมากที่สุดคือการบริการบนเครื่อง (In-flight Services) ซึ่งได้ค่า Eigen Value เท่ากับ 11.262 โดยปัจจัยที่มีผลต่อความพึงพอใจสูงสุดคือความสบายของที่นั่ง (Seat Comfort) ตามด้วยบริการด้านดิจิทัลบนเครื่อง (In-flight Digital Services) และการดำเนินงานหลังบ้าน (Airline Back Office Operations) นอกจากนี้ การสำรวจยังชี้ว่าผู้โดยสารส่วนใหญ่เลือกใช้สายการบิน Jet Airways (9W) โดยเฉพาะในกลุ่มผู้โดยสารชั้นประหยัด ชั้นธุรกิจ และชั้นพรีเมียม ข้อมูลส่วนบุคคลแสดงให้เห็นว่าผู้โดยสารที่มีการศึกษาและรายได้สูงมักเดินทางบ่อยขึ้น ซึ่งชี้ให้เห็นว่าคุณภาพการบริการที่เหมาะสมสามารถสนับสนุนการวางกลยุทธ์ของสายการบินได้อย่างมีประสิทธิภาพ

บทความ The Practical Approach in Customer Segmentation by Using the K-Means Algorithm (Nandapala & Jayasena, 2020)

งานวิจัยนี้ผู้วิจัยเล็งเห็นความสำคัญระหว่างธุรกิจและผู้บริโภคมีความสำคัญมากขึ้นตามการเติบโตทางเทคโนโลยี จำเป็นต้องบริหารจัดการความสัมพันธ์นี้เพื่อการเติบโตของบริษัทในอนาคต กลไกการสื่อสารระหว่างบริษัทและลูกค้าเรียกว่า Customer Relations Management (CRM) ซึ่ง CRM มีบทบาทสำคัญในภาคธุรกิจ นอกจากนี้ ธุรกิจอาจจัดประเภททัศนคติคุณลักษณะ ฯลฯ ของลูกค้าตามวิธี CRM ที่เหมาะสม ด้วยการใช้ความรู้นี้ บริษัทต่าง ๆ สามารถระบุได้ว่าผู้บริโภครายใดมีผลต่อกำไรของบริษัทมากที่สุด ซึ่งสามารถทำได้ด้วยวิธีที่เรียกว่าการแบ่งส่วนลูกค้า (Clustering) โดยผู้วิจัยได้เก็บข้อมูลลูกค้าจากห้างสรรพสินค้าแห่งหนึ่งซึ่งผู้วิจัยไม่ได้ระบุไว้ ในงานวิจัยนี้ K-Means ถูกใช้เป็นอัลกอริทึมสำหรับกระบวนการแบ่งส่วนเพื่อแบ่งกลุ่มลูกค้า ในการศึกษาี้ให้กำหนดกลุ่มทั้ง 6 กลุ่มตามคะแนนรายได้และการใช้จ่ายต่อปี

โดยได้จากการสรุปผลของ 3 เทคนิคได้แก่ Elbow Method, Silhouette Method และ Gap Statistic Method หลังจากนั้นหาปัจจัยต่าง ๆ ของแต่ละกลุ่ม โดยใช้วิธี K-means ภายใน Principal Component Analysis (PCA) และหลังจากการจัดกลุ่มด้วย K-means ในพื้นที่ Principal Component แล้ว สามารถทำการวิเคราะห์ผลลัพธ์เพื่อดูความหมายของกลุ่มที่ได้และความสัมพันธ์ระหว่าง Principal Components กับกลุ่มที่เกิดขึ้น โดยเป็นประโยชน์ต่อบริษัทในการจัดการกับกลุ่มลูกค้าที่มีความแตกต่างกัน เพื่อประโยชน์สูงสุดของบริษัท

บทความวิจัย Customer Segmentation and Profiling for Life Insurance using K-Modes Clustering and Decision Tree Classifier (Abdul-Rahman et al., 2021)

บทความนี้นำเสนอวิธีการที่ฝ่ายการตลาดในธุรกิจประกันสามารถวางกลยุทธ์ที่เหมาะสมกับลูกค้าแต่ละกลุ่ม เพื่อสร้างประโยชน์สูงสุดให้แก่ลูกค้า โดยให้ลูกค้าได้รับแผนประกันที่เหมาะสมกับความต้องการและสร้างความพึงพอใจให้กับลูกค้า โดยใช้เทคนิค K-Modes ในการจัดกลุ่มและ Decision Tree Classifier ในการทำนายกลุ่มลูกค้า

จากการทดลองในงานวิจัยนี้ใช้ชุดข้อมูลจากบริษัทประกันชีวิตหนึ่งในประเทศมาเลเซีย ที่รวบรวมข้อมูลตั้งแต่เดือนมกราคม 2561 ถึงธันวาคม 2562 ประกอบด้วยข้อมูลทั้งหมด 37,181 รายการ และเลือกใช้แอตทริบิวต์ที่เกี่ยวข้องทั้งหมด 12 แอตทริบิวต์

ผลการวิจัยพบว่าการจัดกลุ่มลูกค้าด้วย K-Modes สามารถแบ่งกลุ่มลูกค้าออกเป็น 3 กลุ่ม คือ กลุ่มลูกค้าที่มีศักยภาพสูง (Potential High-Value Customer) กลุ่มลูกค้าที่มีค่าต่ำ (Low Value Customer) และกลุ่มลูกค้าที่ไม่น่าสนใจ (Disinterested Customer) และการทำนายกลุ่มลูกค้าด้วย Decision Tree Classifier ได้ค่า Accuracy สูงสุดที่ 81.30% เมื่อปรับค่าพารามิเตอร์

ด้วย Gini โดยตั้งค่า Leaf node สูงสุดไม่เกิน 50 และใช้เทคนิคการแบ่งข้อมูลแบบ 10-Fold Cross Validation

2.7 เครื่องมือและอัลกอริทึมที่ใช้ในงานวิจัย

2.7.1 เครื่องมือที่ใช้ในงานวิจัย

งานวิจัยนี้ใช้เครื่องมือและซอฟต์แวร์ที่หลากหลายเพื่อสนับสนุนการวิเคราะห์ข้อมูล และการพัฒนาโมเดลให้มีความถูกต้องและประสิทธิภาพสูง โดยแพลตฟอร์มหลักที่ใช้คือ Google Colab ซึ่งรองรับการตั้งค่าฮาร์ดแวร์ขั้นสูง เช่น GPU (T4 GPU) และบริการ Colab Pro ที่ช่วยเพิ่มความเร็วในการประมวลผล อีกทั้ง Microsoft Excel ยังถูกใช้เพื่อการวิเคราะห์เชิงเปรียบเทียบ การจัดทำตาราง และการสรุปผลในรูปแบบที่เข้าใจง่าย ใช้ Google Scholar ในการศึกษางานวิจัยที่เกี่ยวข้อง และใช้ Endnote ในการทำการอ้างอิงงานวิจัยในบทความเล่มนี้

สำหรับการพัฒนาแบบจำลองและการวิเคราะห์ข้อมูล งานวิจัยนี้ได้ใช้ไลบรารี Python ที่มีความหลากหลายและเป็นประโยชน์ โดย Numpy ถูกใช้สำหรับการจัดการข้อมูลเชิงตัวเลขและการคำนวณทางคณิตศาสตร์ Pandas ช่วยในการจัดระเบียบและวิเคราะห์ข้อมูลในรูปแบบ DataFrame ที่มีความยืดหยุ่น ส่วน Matplotlib และ Seaborn ถูกนำมาใช้ในการสร้างกราฟและแสดงผลข้อมูลในลักษณะที่เข้าใจง่าย สำหรับการพัฒนาแบบจำลอง Logistic Regression ใช้ Sklearn.linear_model เพื่อวิเคราะห์หาปัจจัยสำคัญ และ Sklearn.metrics ใช้ในการวัดประสิทธิภาพของแบบจำลอง นอกจากนี้ Sklearn.preprocessing ถูกนำมาใช้สำหรับการปรับค่าข้อมูลให้เหมาะสม และ Sklearn.cluster ถูกใช้ในกระบวนการจัดกลุ่มข้อมูลด้วย K-Means Clustering โดย Matplotlib.pyplot ช่วยสร้างกราฟเพื่อสรุปผลการวิเคราะห์อย่างชัดเจน

การใช้เครื่องมือเหล่านี้ช่วยให้งานวิจัยสามารถดำเนินการวิเคราะห์ข้อมูลได้อย่างมีประสิทธิภาพและแม่นยำ อีกทั้งยังช่วยให้การนำเสนอผลลัพธ์มีความชัดเจนและน่าเชื่อถือ ซึ่งเป็นปัจจัยสำคัญในการอภิปรายผลและการนำเสนองานวิจัย

2.7.2 อัลกอริทึมที่ใช้ในงานวิจัย

ในงานวิจัยนี้ผู้วิจัยได้นำอัลกอริทึม Logistic Regression และ K-Means Clustering มาใช้เป็นเครื่องมือหลัก โดย Logistic Regression ถูกเลือกเนื่องจากมีประสิทธิภาพในการจำแนกข้อมูลออกเป็นสองกลุ่ม เช่น กลุ่มที่พึงพอใจและไม่พึงพอใจ ซึ่งสอดคล้องกับแนวทางที่ได้รับการสนับสนุนจากงานวิจัยก่อนหน้านี้เกี่ยวกับการวิเคราะห์ข้อมูลในบริบทเชิงพาณิชย์

ในขณะที่ K-Means Clustering ถูกนำมาใช้ในการแบ่งกลุ่มผู้โดยสารตามพฤติกรรม และลักษณะการประเมินบริการ ซึ่งงานวิจัยก่อนหน้านี้ได้แสดงให้เห็นถึงประสิทธิภาพของ K-Means ในการจัดกลุ่มข้อมูลที่มีความหลากหลายการใช้ K-Means ช่วยให้เข้าใจลักษณะเฉพาะของแต่ละกลุ่มผู้โดยสาร และสามารถพัฒนากลยุทธ์การบริการที่เหมาะสมได้ โดยการเลือกใช้ อัลกอริทึมเหล่านี้ได้รับการสนับสนุนจากผลการวิจัยที่เกี่ยวข้อง ซึ่งช่วยเสริมความน่าเชื่อถือในการวิเคราะห์และแบ่งกลุ่มผู้โดยสารอย่างมีประสิทธิภาพ



บทที่ 3

วิธีการดำเนินงานวิจัย

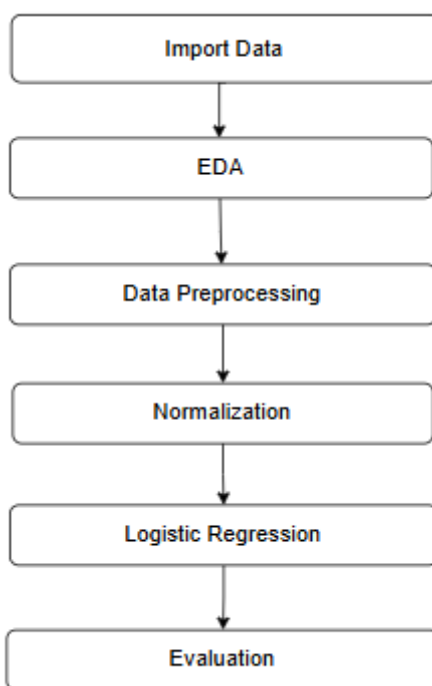
งานวิจัยนี้อธิบายขั้นตอนการวิเคราะห์ข้อมูลและการพัฒนาแบบจำลองเพื่อระบุปัจจัยสำคัญและจัดกลุ่มลูกค้า กระบวนการเริ่มจากการเก็บรวบรวมข้อมูล (Data Collection) ซึ่งเป็นขั้นตอนสำคัญในการจัดหาข้อมูลที่เหมาะสมและเพียงพอ จากนั้นจึงทำการสำรวจข้อมูลเบื้องต้น (Exploratory Data Analysis: EDA) เพื่อทำความเข้าใจลักษณะของข้อมูล ตามด้วยการเตรียมข้อมูล (Data Preprocessing) เพื่อจัดการข้อมูลที่อาจขาดหายหรือไม่สมบูรณ์ สำหรับการวิเคราะห์ข้อมูลในเชิงลึก ได้ใช้โมเดล Supervised Learning เพื่อค้นหาปัจจัยที่มีผลต่อความพึงพอใจของลูกค้า และนำโมเดล Clustering มาใช้ในการจัดกลุ่มลูกค้าตามพฤติกรรมและข้อมูลประชากรศาสตร์ สุดท้ายมีการประเมินผลโมเดลการทำนายเพื่อวัดความแม่นยำและประสิทธิภาพของการวิเคราะห์ข้อมูล

3.1 กระบวนการทำงานของแบบจำลอง

การศึกษานี้มุ่งเน้นการวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อระดับความพึงพอใจของผู้โดยสารสายการบิน โดยใช้ข้อมูลการให้คะแนนของผู้โดยสารเพื่อตรวจสอบปัจจัยที่เกี่ยวข้องกับความพึงพอใจหรือไม่พึงพอใจ กระบวนการเริ่มต้นจากการเตรียมข้อมูลและการวิเคราะห์ด้วย Logistic Regression เพื่อระบุปัจจัยสำคัญที่ส่งผลกระทบต่อความพึงพอใจ ขั้นตอนนี้ช่วยคัดเลือกตัวแปรที่มีผลต่อการจำแนกกลุ่มผู้โดยสารที่พึงพอใจและไม่พึงพอใจ หลังจากนั้นผลลัพธ์จาก Feature Importance ของ Logistic Regression ถูกนำไปใช้ร่วมกับ K-Means Clustering เพื่อแบ่งกลุ่มผู้โดยสารตามการประเมินบริการ ซึ่งการใช้เทคนิคทั้งสองร่วมกันช่วยระบุปัจจัยที่สำคัญในแต่ละกลุ่มได้อย่างชัดเจน

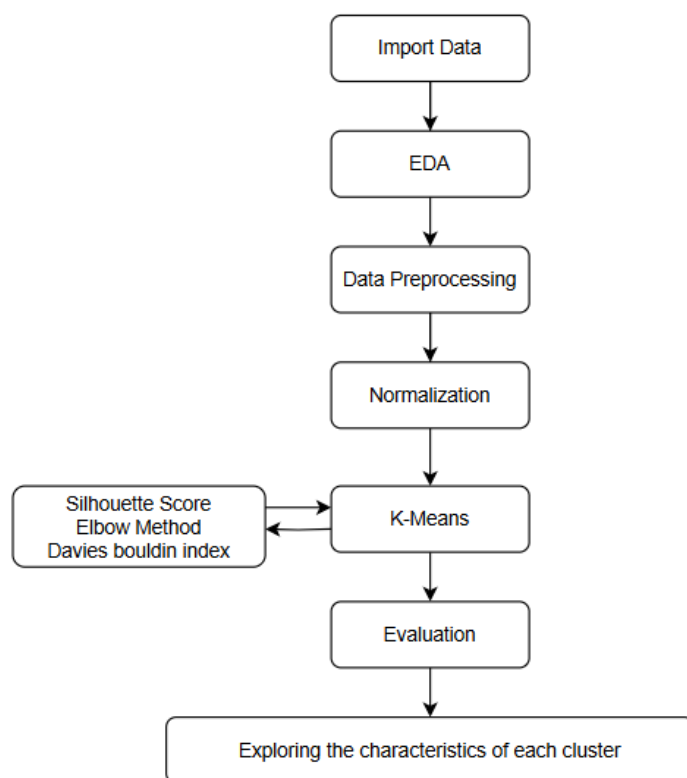
ในกระบวนการเบื้องต้น ข้อมูลจากการสำรวจความคิดเห็นผู้โดยสารถูกตรวจสอบเพื่อความครบถ้วน เช่น การจัดการข้อมูลที่ขาดหาย (Missing Data) และข้อมูลที่ผิดปกติ (Outliers) เพื่อป้องกันผลกระทบเชิงลบต่อการวิเคราะห์ จากนั้นทำการ Normalize ข้อมูลเพื่อลดความแตกต่างในหน่วยและขนาดของฟีเจอร์ และเตรียมข้อมูลสำหรับ Logistic Regression ที่ใช้วิเคราะห์ความสัมพันธ์ระหว่างคุณลักษณะและความพึงพอใจหรือไม่พึงพอใจ Logistic Regression ถูกเลือกเพื่อช่วยระบุฟีเจอร์ที่มีอิทธิพลต่อการตัดสินใจของผู้โดยสารในด้านความพึงพอใจ โดยคะแนนการให้บริการในแต่ละด้านเป็นตัวแปรหลัก เมื่อพัฒนาแบบจำลองแล้ว มีการประเมินประสิทธิภาพด้วยเกณฑ์ เช่น Accuracy เพื่อวัดความแม่นยำโดยรวม, Precision

เพื่อวัดความถูกต้องในการจำแนกผู้โดยสารที่พึงพอใจ, Recall ที่บ่งบอกถึงความสามารถในการตรวจพบผู้โดยสารที่พึงพอใจ และ F1-score ซึ่งเป็นตัววัดสมดุระหว่าง Precision และ Recall ซึ่งขั้นตอนการสร้างแบบจำลองแสดง ดังภาพประกอบ 3



ภาพประกอบ 3 แสดงขั้นตอนการสร้างแบบจำลอง เพื่อหาปัจจัยที่สำคัญต่อความพึงพอใจของผู้โดยสาร

นอกจากการวิเคราะห์ปัจจัยสำคัญที่ส่งผลต่อความพึงพอใจแล้ว ผู้วิจัยยังได้แบ่งกลุ่มผู้โดยสารที่ไม่พึงพอใจออกเป็นกลุ่มย่อยเพิ่มเติมด้วยวิธี K-Means เพื่อเข้าใจลักษณะเฉพาะของแต่ละกลุ่ม โดยใช้ Elbow Method ,Silhouette Score และ Davies-Bouldin Score ในการกำหนดจำนวนกลุ่มที่เหมาะสม จากนั้นประเมินประสิทธิภาพการแบ่งกลุ่มด้วย Silhouette Score และ Davies-Bouldin Score ผลที่ได้ช่วยให้สามารถระบุความต้องการเฉพาะของผู้โดยสารในแต่ละกลุ่ม



ภาพประกอบ 4 แสดงขั้นตอนการสร้างแบบจำลองการแบ่งกลุ่ม

3.2 การสืบค้นกลุ่มตัวอย่าง (กลุ่มผู้โดยสารที่มาใช้บริการสายการบิน)

งานวิจัยนี้ใช้ชุดข้อมูลชื่อ The U.S. Airline Passenger Satisfaction Dataset จากข้อมูลสาธารณะของเว็บไซต์ Kaggle (<https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>) โดยชุดข้อมูลนี้อธิบายเกี่ยวกับความพึงพอใจของผู้โดยสารในสายการบินในสหรัฐฯ โดยการดำเนินการสำรวจในปี 2015 โดยมีข้อมูลที่รวบรวมจากตัวอย่างผู้โดยสารจำนวน 129,880 รายคนที่ใช้บริการสายการบินแบบเต็มรูปแบบ (Full Service) แบบซึ่งมีผู้โดยสารที่พึงพอใจต่อการใช้บริการ 56,428 ราย และผู้โดยสารที่ไม่พึงพอใจต่อการใช้บริการ 73,452 ราย ข้อมูลชุดนี้ประกอบไปด้วยข้อมูลส่วนตัวของผู้โดยสาร ดังตาราง 2, ข้อมูลของไฟล์ทโดยสาร ดังตาราง 3 และข้อมูลการบริการต่าง ๆ ของสายการบิน ดังตาราง 4

ตาราง 2 แสดงตัวแปรของข้อมูลส่วนตัวของผู้โดยสาร

Variable	Description
Satisfaction	ความพึงพอใจของผู้โดยสาร
Id	รหัสผู้โดยสาร
Age	อายุ
Gender	เพศ
Customer Type	ประเภทของลูกค้า

ตาราง 3 แสดงตัวแปรของไฟล์ท์

Variable	Description
Type of Travel	ประเภทของการท่องเที่ยว
Class	ระดับที่นั่ง
Flight distance	ระยะทางของเที่ยวบิน (Mile)
Departure Delay in Minutes	ความล่าช้าในการออกเดินทาง (นาที)
Arrival Delay in Minutes	ความล่าช้าในการไปถึงปลายทาง (นาที)

ตาราง 4 แสดงตัวแปรของการบริการของสายการบิน

Variable	Description
Inflight WiFi Service	การบริการอินเทอร์เน็ตบนเครื่อง
Ease of Online booking	การจองแบบออนไลน์
Inflight service	การให้บริการบนเครื่องของพนักงาน
Online boarding	การเช็คอินออนไลน์
Inflight entertainment	ความบันเทิงบนเครื่อง
Food and drink	อาหารและเครื่องดื่ม
Seat comfort	ความสบายของที่นั่ง
On-board service	บริการที่มีบนเครื่อง
Leg room service	พื้นที่วางขา
Departure/Arrival time convenient	ความเหมาะสมของเที่ยวบิน
Baggage handling	การจัดการสัมภาระกระเป๋า
Gate location	ตำแหน่งทางขึ้นเครื่องบินโดยสาร
Cleanliness	ความสะอาด
Check-in service	การบริการเช็คอิน

3.3 การสำรวจข้อมูล (Exploratory Data Analysis: EDA)

การวิเคราะห์ข้อมูลในเชิงการตรวจสอบเป็นการทำความเข้าใจชุดข้อมูลที่มีการใช้เทคนิคต่าง ๆ เพื่อเป็นการทำความเข้าใจกับข้อมูลที่ใช้งานได้ดีขึ้น โดยงานวิจัยนี้ใช้ภาษา Python ในการวิเคราะห์ข้อมูลและทำ Machine Learning จากการสำรวจข้อมูล ข้อมูลชุดนี้มีค่า Missing Value โดยจากการตรวจสอบพบว่า มีจำนวน 227 ตัวที่ Feature Arrival Delay in Minutes ดังภาพประกอบ 5

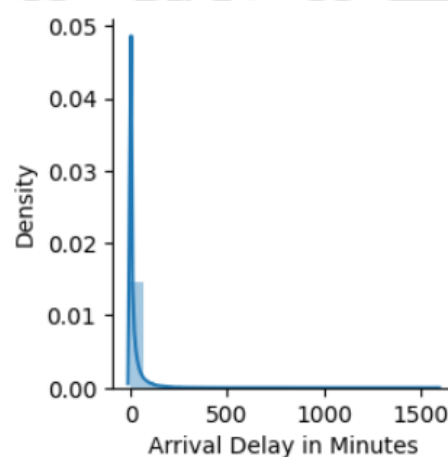
```

Gender                                0
Customer Type                         0
Age                                   0
Type of Travel                        0
Class                                 0
Flight Distance                       0
Inflight wifi service                 0
Departure/Arrival time convenient    0
Ease of Online booking                0
Gate location                         0
Food and drink                        0
Online boarding                       0
Seat comfort                          0
Inflight entertainment                0
On-board service                     0
Leg room service                     0
Baggage handling                     0
Checkin service                       0
Inflight service                      0
Cleanliness                           0
Departure Delay in Minutes            0
Arrival Delay in Minutes              227
satisfaction                          0
dtype: int64

```

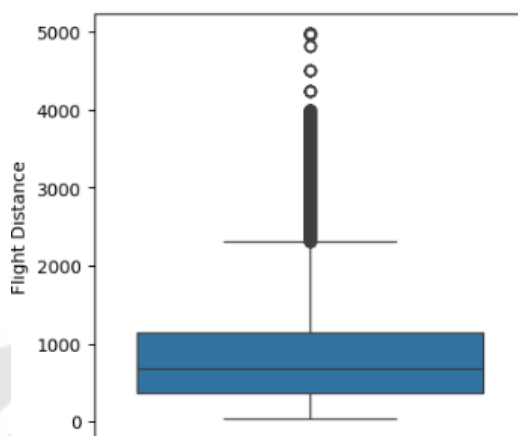
ภาพประกอบ 5 แสดงจำนวน Missing Value ในแต่ละตัวแปร

ผู้วิจัยได้จัดการกับค่า Missing Value โดยพิจารณาจากการวิเคราะห์ความหนาแน่นของข้อมูล (Density Analysis) ดังภาพประกอบ 6 จากการวิเคราะห์พบว่าค่าข้อมูลส่วนใหญ่กระจุกตัวอยู่ที่ศูนย์ ดังนั้นในการเติมข้อมูลที่ขาดหาย (Missing Values) ผู้วิจัยจึงเลือกเติมด้วยค่าศูนย์เพื่อรักษาลักษณะการกระจายตัวของชุดข้อมูล

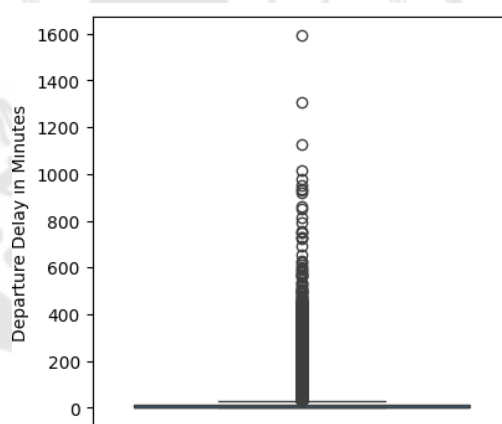


ภาพประกอบ 6 แสดงความหนาแน่นของข้อมูล (Destiny) ของ Arrival Delay in Minutes

หากใช้กราฟ Box Plot พิจารณากระจายตัวของตัวแปร พบว่า ระยะทางของเที่ยวบิน และความล่าช้าในการออกเดินทาง มี Outlier ดังภาพประกอบ 7 และ 8 ตามลำดับ ซึ่งผู้วิจัยเลือกที่กำจัดค่านั้นทิ้งเนื่องจากอาจเป็นตัวแปรนอกเหนืองานวิจัยที่ส่งผลต่อผลการทดลอง



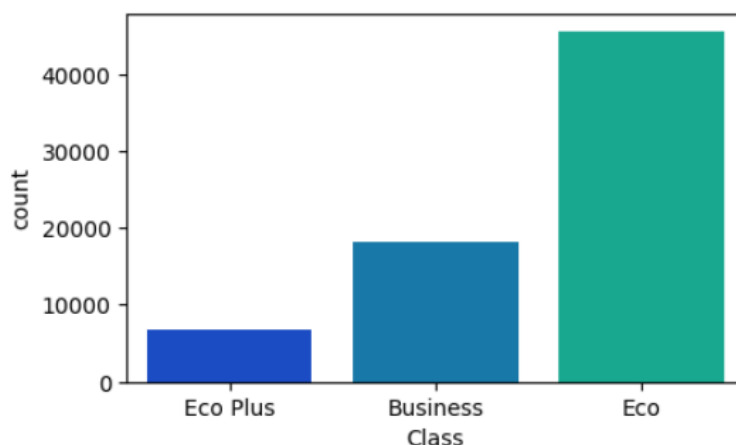
ภาพประกอบ 7 แสดงกราฟ Box Plot ของระยะทางของเที่ยวบิน



ภาพประกอบ 8 แสดงกราฟ Box Plot ของความล่าช้าในการออกเดินทาง

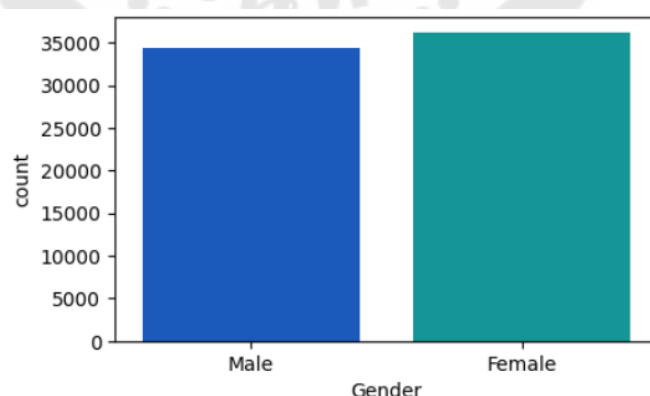
ทำการดูภาพรวมของข้อมูลว่าผู้โดยสารที่มาใช้บริการที่ไม่พอใจกับการบริการของสายการบิน และเนื่องจากสายการบินต้องการคุณภาพที่สูง จึงมีมุมมองให้ผู้โดยสารที่เฉย ๆ ต่อการให้บริการ รวมกับผู้โดยสารที่ไม่พอใจกับการให้บริการเป็นมุมมองที่สายการบินให้ความสนใจเพื่อการพัฒนาความพึงพอใจของผู้โดยสารในกลุ่มนั้น ๆ

โดยเริ่มต้นจากการวิเคราะห์ระดับชั้นที่นั่ง (Class) ที่มีผู้โดยสารแสดงความไม่พึงพอใจสูงสุด ผลลัพธ์ชี้ให้เห็นว่าชั้นที่นั่งแบบประหยัด (Eco Class) มีจำนวนผู้โดยสารที่ไม่พึงพอใจมากกว่าชั้นที่นั่งแบบธุรกิจ (Business Class) และชั้นที่นั่งประหยัดแบบพรีเมียม (Eco Plus Class) อย่างชัดเจน ดังที่แสดงในภาพประกอบ 9



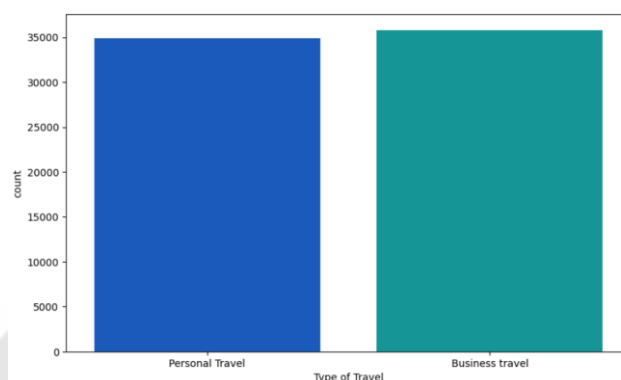
ภาพประกอบ 9 แสดงจำนวนของผู้โดยสารที่ไม่พึงพอใจตามระดับชั้นที่นั่ง (Class)

ในส่วนเพศ (Gender) ของผู้โดยสารที่ไม่พึงพอใจต่อการบริการสายการบินนั้นมีจำนวนเพศชายและหญิงที่ใกล้เคียงกันดังภาพประกอบ 10



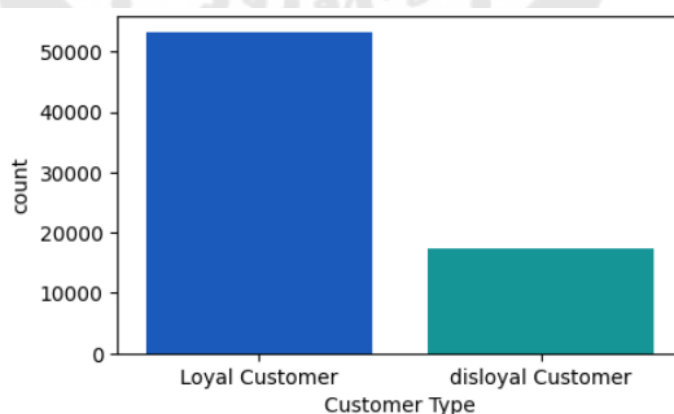
ภาพประกอบ 10 แสดงจำนวนของผู้โดยสารที่ไม่พึงพอใจตามเพศ (Gender)

ในส่วนประเภทการท่องเที่ยวของผู้โดยสาร (Type of Travel) ของผู้โดยสารที่ไม่พึงพอใจต่อบริการสายการบินนั้นมีจำนวนผู้โดยสารเดินทางด้วยเหตุผลส่วนตัวและเดินทางด้วยเหตุผลทางธุรกิจที่ใกล้เคียงกันดังภาพประกอบ 11



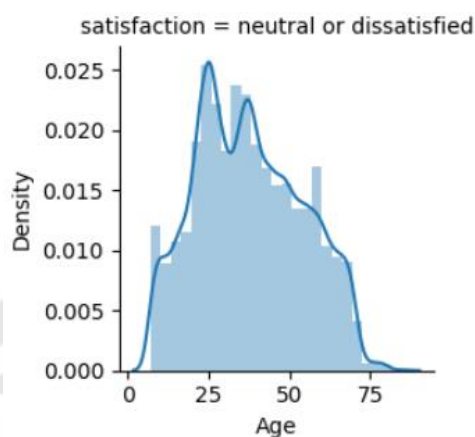
ภาพประกอบ 11 แสดงจำนวนของผู้โดยสารที่ไม่พึงพอใจตามประเภทการท่องเที่ยว

สำหรับประเภทของผู้โดยสาร (Customer Type) ที่แสดงความไม่พึงพอใจต่อบริการสายการบิน พบว่ากลุ่มผู้โดยสารที่เป็นลูกค้าประจำ (Loyal Customer) มีจำนวนผู้โดยสารที่ไม่พึงพอใจต่อบริการมากกว่ากลุ่มผู้โดยสารที่ไม่ได้เดินทางกับสายการบินเป็นประจำ (Disloyal Customer) อย่างชัดเจน ดังภาพประกอบ 12



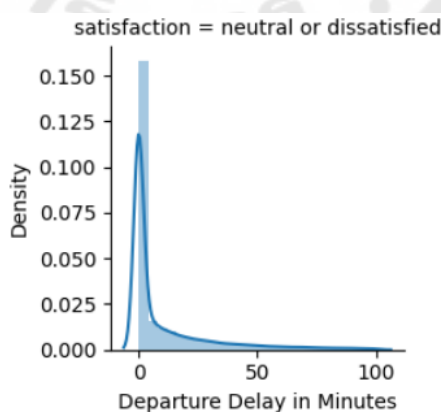
ภาพประกอบ 12 แสดงจำนวนของผู้โดยสารที่ไม่พึงพอใจตามประเภทของผู้โดยสาร

ในส่วนอายุ (Age) ของผู้โดยสารที่ไม่พึงพอใจต่อบริการสายการบินนั้นมีจำนวนมากในช่วง 25-30 ปี รองลงมาคือช่วงอายุ 35-40 ปี และลดลงเรื่อย ๆ หลังจากช่วงอายุ 40 ปีขึ้นไป ดังภาพประกอบ 13



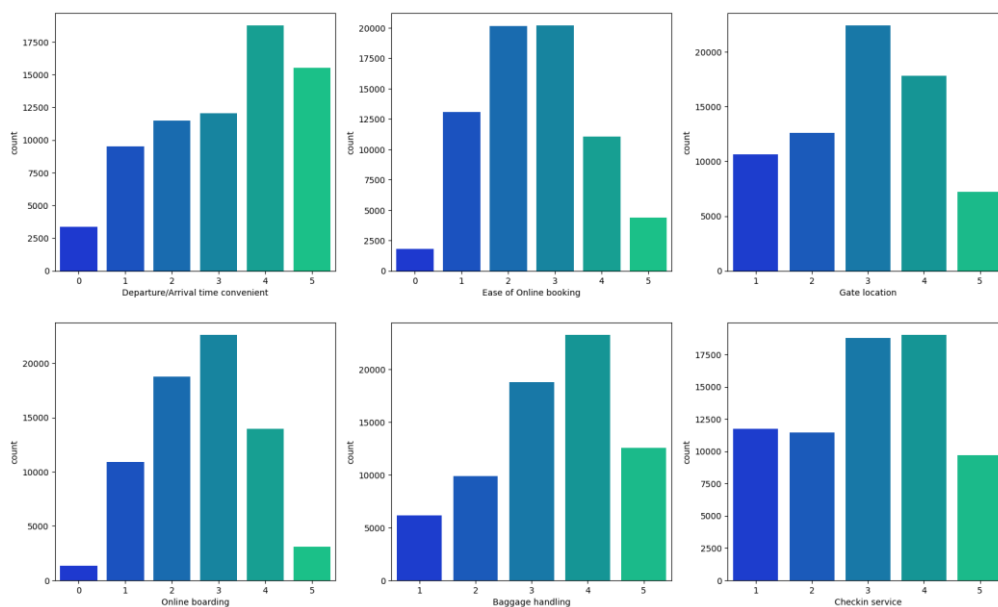
ภาพประกอบ 13 แสดงความหนาแน่นของจำนวนของผู้โดยสารที่ไม่พึงพอใจตามอายุ

ในส่วนความล่าช้าในการออกเดินทาง (Departure Delay in Minutes) ของผู้โดยสารที่ไม่พึงพอใจต่อบริการสายการบินนั้นส่วนมากไม่มีความล่าช้าในเที่ยวบินหรือมีความล่าช้าที่น้อยมากไม่เกิน 10 นาที ดังภาพประกอบ 14



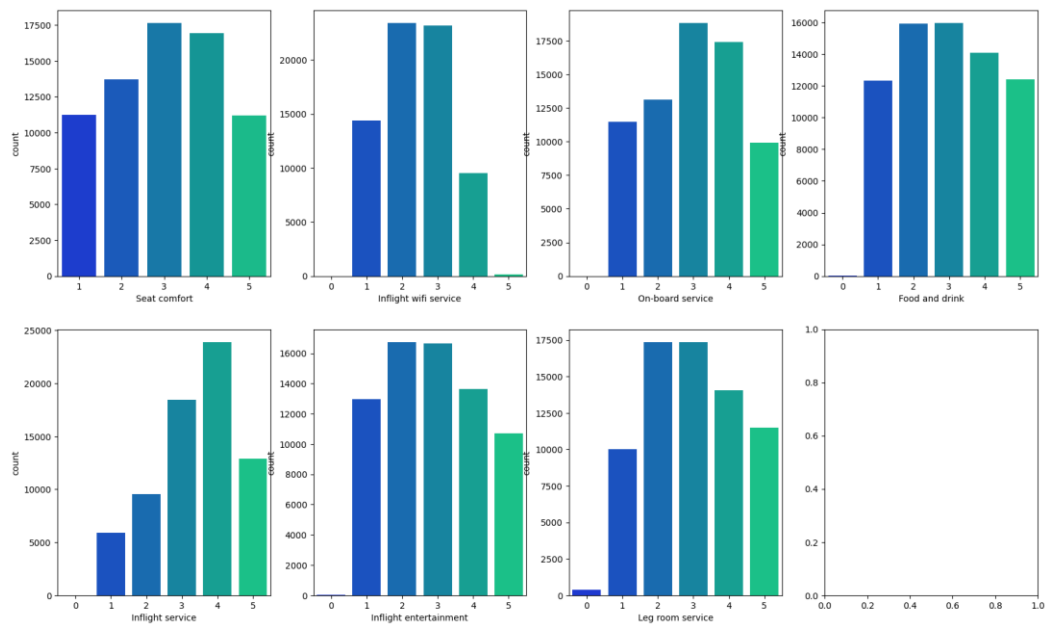
ภาพประกอบ 14 แสดงความหนาแน่นของจำนวนของผู้โดยสารที่ไม่พึงพอใจตามความล่าช้าในการออกเดินทาง

เมื่อดูภาพรวมจากการให้คะแนนการบริการภาคพื้นโดยประกอบไปกับการเช็คอินออนไลน์, การเช็คอิน, ตำแหน่ง Gate โดยสาร, การจัดการเรื่องกระเป๋า, ความเหมาะสมของเวลาเที่ยวบิน และการจองแบบออนไลน์ ด้วยพบว่าคะแนนส่วนมากที่ผู้โดยสารประเมินคือ 3 และ 4 ดังภาพประกอบ 15



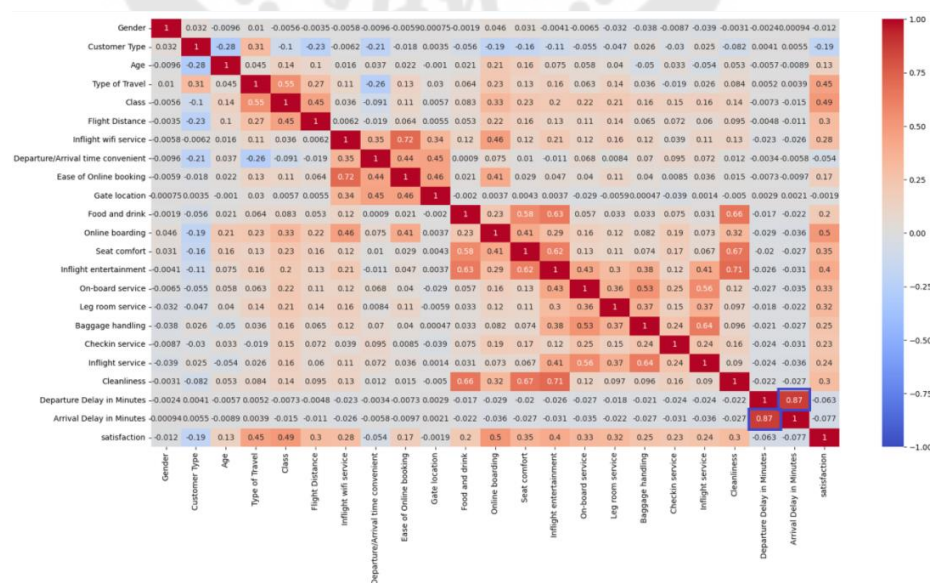
ภาพประกอบ 15 แสดงจำนวนคะแนนของบริการภาคพื้นที่ได้จากการประเมินของผู้โดยสาร

ภาพรวมจากการให้คะแนนการบริการบนไฟล์ทโดยประกอบไปความสบายของที่นั่ง, ความบันเทิงบนเครื่อง, พื้นที่วางขา, การบริการบนเที่ยวบิน, ความสะอาด, บริการที่มีบนเครื่อง, การให้บริการอินเทอร์เน็ตบนเครื่อง, การให้บริการบนเครื่องของพนักงานและอาหารและเครื่องดื่ม ด้วยพบว่าคะแนนส่วนมากที่ผู้โดยสารประเมิน คือ 2 - 4 ดังภาพประกอบ 16



ภาพประกอบ 16 แสดงจำนวนคะแนนของบริการบนไฟล์ท์ที่ได้จากการประเมินของผู้โดยสาร

จากการทดสอบค่าความสัมพันธ์กันของข้อมูลพบว่าความล่าช้าในการออกเดินทางและความล่าช้าที่ไปถึงปลายทางมีค่าความสัมพันธ์กันที่สูงมากดังภาพประกอบ 17 จึงเลือกใช้เพียงความล่าช้าในการออกเดินทางเนื่องจากครอบคลุมต่อผลกระทบการบริการภาคพื้นและบนไฟล์ท์ของสถานีต้นทางซึ่งสายการบินสามารถดูแลผู้โดยสารเพิ่มเติมเพื่อรักษาความพึงพอใจไว้ได้



ภาพประกอบ 17 แสดงความสัมพันธ์กันของแต่ละตัวแปร

สรุปผลลัพธ์ที่ได้จากการทำ EDA

1. จากข้อมูลเมื่อพิจารณาจากระดับชั้นที่นั่งพบว่าผู้โดยสารที่ไม่พึงพอใจส่วนมากโดยสารในระดับชั้นที่นั่งแบบประหยัด (Economy Class)
2. ส่วนของปัจจัยเรื่องเพศและประเภทการท่องเที่ยวมีจำนวนผู้โดยสารที่ไม่พึงพอใจในแต่ประเภทใกล้เคียงกัน
3. ส่วนของปัจจัยเรื่องอายุ (Age) ผู้โดยสารที่ไม่พึงพอใจส่วนมากอยู่ในช่วงวัยรุ่น (25 - 30 ปี)
4. ส่วนของปัจจัยเรื่องความล่าช้าในการออกเดินทางของผู้โดยสารที่ไม่พึงพอใจส่วนมากมีความล่าช้าเพียงเล็กน้อยไม่เกิน 10 นาที
5. การประเมินการให้บริการส่วนมากของผู้โดยสารที่ไม่พึงพอใจต่อการบริการสายการบินอยู่ที่ 3 และ 4 คะแนน

3.4 การเตรียมข้อมูล (Data Preprocessing)

3.4.1 การแปลงค่าประเภทของข้อมูล (Encoding)

การเตรียมข้อมูลสำหรับการฝึกฝนของแบบจำลองเริ่มจากการแปลงค่าประเภทของข้อมูล (Encoding) เนื่องจากชุดข้อมูลมีฟีเจอร์ประเภทหมวดหมู่ ได้แก่ Gender, Customer Type, Type of Travel และ Class จึงใช้วิธีแปลงค่าฟีเจอร์ประเภทหมวดหมู่เป็นค่าตัวเลขด้วยวิธีการ One-Hot Encoding ซึ่งจะได้ข้อมูลดังภาพประกอบ 18

	Gender	Customer Type	Age	Type of Travel	Class
0	0	0	13	0	1
1	0	1	25	1	2
3	1	0	25	1	2
5	1	0	26	0	0
6	0	0	47	0	0

ภาพประกอบ 18 แสดงตัวอย่างข้อมูลหลังการทำ Encoding

3.4.2 การปรับช่วงค่าของพีเจอร์ (Normalization)

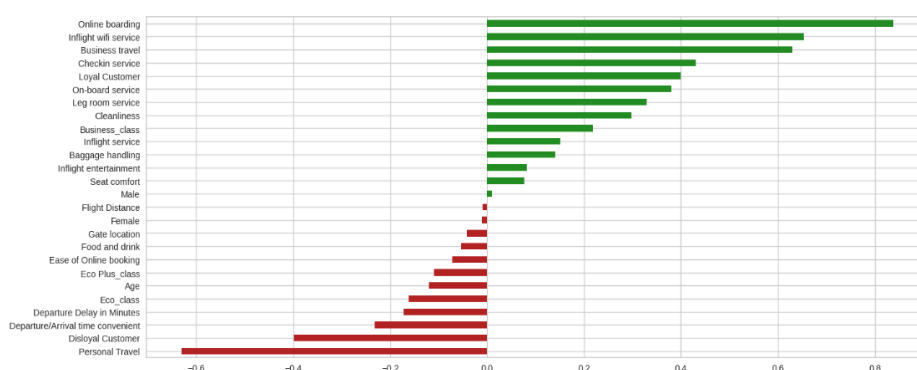
หลังจากเสร็จสิ้นการทำ Encoding ผู้วิจัยใช้ StandardScaler เพื่อช่วยปรับค่าของคุณลักษณะให้อยู่ในช่วงค่าที่มีขนาดเท่ากันหรือใกล้เคียงกัน ซึ่งจะช่วยลดผลกระทบจากความแตกต่างในขนาดของคุณลักษณะ ดังตัวอย่างภาพประกอบ 19

	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	...
0	-1.015116	-0.472890	-1.748931	-1.492699	-0.032583	-0.731595	0.202627	0.617189	0.173019	-1.544273	...
1	-1.015116	2.114655	-0.955265	0.669927	1.006254	-0.957040	0.202627	-0.691840	0.173019	0.019405	...
2	0.985109	-0.472890	-0.889126	0.669927	1.006254	-0.048246	-0.548437	-0.691840	-0.539902	-0.762434	...
3	0.985109	-0.472890	-0.955265	0.669927	1.006254	-0.629393	-0.548437	1.271703	1.598860	1.583082	...
4	-1.015116	-0.472890	1.425733	0.669927	1.006254	-0.978082	0.202627	-0.037325	0.173019	0.019405	...

ภาพประกอบ 19 แสดงข้อมูลหลังจากผ่าน StandardScaler

3.5 อัลกอริทึมแบบจำลองในการหาปัจจัยที่สำคัญ

งานวิจัยนี้ใช้เทคนิค Logistic Regression เพื่อจำแนกกลุ่มผู้โดยสารตามปัจจัยต่าง ๆ ที่ส่งผลกระทบต่อความพึงพอใจ โดย Logistic Regression เป็นโมเดลการถดถอยที่นิยมใช้ในการทำนายผลลัพธ์ที่เป็นประเภทต่าง ๆ โดยเฉพาะเมื่อมีความสัมพันธ์เชิงเส้นระหว่างคุณลักษณะ (Features) และผลลัพธ์ที่เป็นไปได้ (Binary Outcome) สำหรับการหาปัจจัยสำคัญ (Feature Importance) ในโมเดลนี้ ผู้วิจัยใช้การวิเคราะห์ค่าสัมประสิทธิ์ (Coefficient) ของพีเจอร์แต่ละตัว ค่าสัมประสิทธิ์จะบ่งชี้ถึงความสำคัญและทิศทางของผลกระทบต่อผลลัพธ์ โดยค่าบวกแสดงถึงปัจจัยที่เพิ่มโอกาสให้เกิดผลลัพธ์บวก ขณะที่ค่าลบหมายถึงปัจจัยที่ลดโอกาสของผลลัพธ์นั้น ดังตัวอย่างภาพประกอบ 20

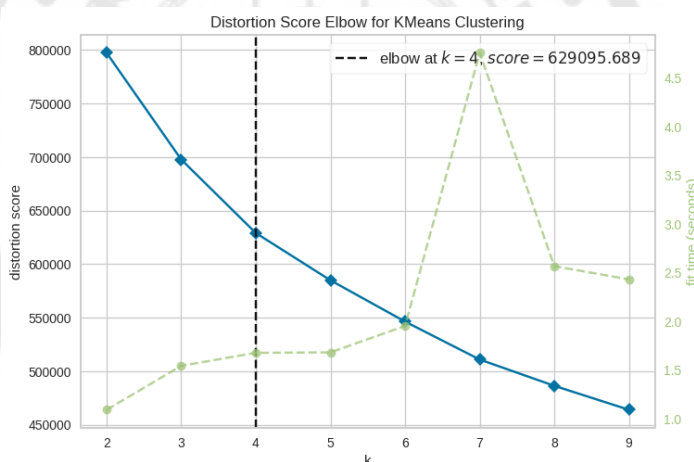


ภาพประกอบ 20 ปัจจัยที่ส่งผลกระทบต่อความพึงพอใจของผู้โดยสาร

3.6 อัลกอริทึมแบบจำลองในการแบ่งกลุ่มข้อมูล

ในการวิจัยนี้ ผู้วิจัยใช้เทคนิค K-Means Clustering เพื่อจัดกลุ่มผู้โดยสาร โดย K-Means เป็นวิธีการจัดกลุ่มข้อมูลที่นิยม ซึ่งในขั้นตอนแรกจะต้องกำหนดจำนวนคลัสเตอร์ (K) ที่ต้องการล่วงหน้า เพื่อหาค่า K ที่เหมาะสมที่สุด ผู้วิจัยได้ใช้ Elbow Method , Silhouette Score และ Davies-Bouldin Index ซึ่งช่วยในการเลือกจำนวนคลัสเตอร์ที่เหมาะสม โดยการคำนวณค่า Cost Function ซึ่งในกรณีของ K-Means คือการหาผลรวมระยะห่างระหว่างแต่ละจุดข้อมูลกับจุดศูนย์กลางของคลัสเตอร์ วิธีนี้ช่วยหาจุดที่เหมาะสมที่สุดในการเลือกจำนวนคลัสเตอร์

ในวิธี K-Means แบบดั้งเดิม จุดศูนย์กลางของคลัสเตอร์ (Centroid) มักถูกเลือกแบบสุ่ม ซึ่งอาจทำให้เกิดความไม่เสถียรในการจัดกลุ่ม แต่ในงานวิจัยนี้ ผู้วิจัยได้นำเทคนิค K-Means++ มาใช้ ซึ่งเป็นการปรับปรุงวิธีการสุ่มในการเลือกจุดศูนย์กลางแรก เพื่อให้การจัดกลุ่มมีความเสถียรมากขึ้นและลดโอกาสเกิดความไม่เหมาะสมในการจัดกลุ่มดังภาพประกอบ 21



ภาพประกอบ 21 แสดงการหาจำนวนกลุ่มที่ดีที่สุดด้วยวิธี Elbow Method

K-Means++ มีคุณสมบัติในการเลือกจุดเริ่มต้นที่ดีขึ้น ซึ่งช่วยเพิ่มความแม่นยำในการจัดกลุ่มและลดปัญหาจากการเริ่มต้นที่ไม่เหมาะสม นอกจากนี้ การใช้ Elbow Method ร่วมกับ K-Means++ ช่วยในการกำหนดจำนวนคลัสเตอร์ที่เหมาะสม ซึ่งทำให้การจัดกลุ่มลูกค้ามีประสิทธิภาพมากขึ้นและผลลัพธ์จากการวิจัยมีความน่าเชื่อถือสูงขึ้น

3.7 อัลกอริทึมของแบบจำลองการทำนาย

งานวิจัยนี้มุ่งเน้นในการเพิ่มประสิทธิภาพและความน่าเชื่อถือของการจำแนกกลุ่มลูกค้าด้วย Logistic Regression ซึ่งเป็นโมเดลการถดถอยที่ใช้สำหรับการทำนายผลลัพธ์ในรูปแบบของการจำแนกประเภท โดยคำนึงถึงความน่าจะเป็นของข้อมูลในแต่ละคลาส ข้อมูลถูกแบ่งอย่างเหมาะสมระหว่างการฝึกสอนและการทดสอบเพื่อเพิ่มประสิทธิภาพความแม่นยำในการสร้างโมเดลนี้ ต่อมาผู้วิจัยได้ปรับพารามิเตอร์ต่าง ๆ เพื่อให้ได้ค่าที่เหมาะสมที่สุดสำหรับ Logistic Regression ผ่านการทดสอบหลาย ๆ ชุดพารามิเตอร์โดยแต่ละพารามิเตอร์มีความหมายดังนี้

1. Penalty คือ กำหนดค่าโทษในการปรับลดขนาดน้ำหนักของพีเจอร์เพื่อป้องกัน Overfitting โดยเลือกค่าจาก l1, l2, Elasticnet, หรือ none ตามลักษณะของข้อมูล
2. C คือ ค่าคงที่ที่ควบคุมความเข้มข้นในการลงโทษ ค่าที่สูงกว่าจะให้ความสำคัญกับการทำนายให้แม่นยำมากขึ้น แต่เสี่ยงต่อการ Overfitting
3. Solver คือ เลือกวิธีในการหาค่าเหมาะสมของพารามิเตอร์ โดยมีตัวเลือกเช่น liblinear, saga, newton-cg, และ lbfgs ซึ่งแต่ละวิธีมีข้อดีในการจัดการกับข้อมูลและการเลือกพีเจอร์ที่แตกต่างกัน
4. Max_iter คือ จำนวนรอบสูงสุดในการเรียนรู้เพื่อให้แน่ใจว่าโมเดลจะเรียนรู้ข้อมูลครบถ้วนและไม่หยุดก่อน
5. Class_weight คือ ปรับน้ำหนักของแต่ละคลาสเพื่อจัดสมดุลในกรณีที่ข้อมูลมีความไม่สมดุลระหว่างกลุ่มมากเกินไป
6. Multi_class คือ เลือกรูปแบบของการจำแนกประเภทในกรณีที่ข้อมูลมีหลายกลุ่ม เช่น ovr (One-vs-Rest) หรือ multinomial สำหรับข้อมูลหลายประเภท

หลังจากที่ปรับพารามิเตอร์จนได้ค่าที่เหมาะสมและให้โมเดล Logistic Regression ได้เรียนรู้ข้อมูลอย่างครบถ้วน ผู้วิจัยได้ทำการประเมินประสิทธิภาพของแบบจำลองโดยใช้ตัวชี้วัดต่าง ๆ เช่น Accuracy, Precision, Recall, F1-Score รวมถึงการแสดงผลของ Confusion Matrix เพื่อประเมินความถูกต้องและความน่าเชื่อถือของผลลัพธ์

บทที่ 4

ผลการศึกษา

การศึกษาวិธีการหาปัจจัยสำคัญโดยใช้โลจิสติกส์รีเกรสชัน และการแบ่งกลุ่มลูกค้าโดยใช้เทคนิคเคมีน ดำเนินการตามขั้นตอนต่างๆ เพื่อให้เป็นไปตามวัตถุประสงค์ที่กำหนดไว้ ดังนี้

1. ผลจากการหาปัจจัยที่สำคัญจาก Logistic Regression
2. ผลจากการแบ่งกลุ่มทำ K-Means Clustering

4.1 ผลจากการหาปัจจัยที่สำคัญจาก Logistic Regression

แบบจำลองการจำแนกประเภทด้วย Logistic Regression ใช้ข้อมูลผู้โดยสารดำเนินการ Normalization ดังภาพที่ 22 ซึ่งได้ผลการทำนายดังตาราง 5 และประสิทธิภาพการจำแนกประเภท ได้แก่ Accuracy Precision Recall และ F1 score ซึ่งวัดประสิทธิภาพทั้งในส่วนของ Positive class และ Negative class แสดงดังตาราง 6 และผลจากการหาปัจจัยที่สำคัญ แสดงดังภาพที่ 23 โดยค่าสัมประสิทธิ์ของแต่ละปัจจัยแสดงดังตาราง 7

	Age	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	Online boarding	Seat comfort	Inflight entertainment	...
0	-1.287045	0.197078	0.971450	-1.465208	0.900545	0.798764	0.601334	0.557519	0.425875	0.484785	...
1	-0.228415	-0.167199	-0.594541	-0.769528	-0.603869	0.015854	1.354346	-0.978440	0.425875	1.234177	...
2	0.036242	-0.964807	-1.377537	-1.465208	1.652753	-1.549967	1.354346	-1.746419	-0.332396	0.484785	...
3	0.697886	2.585899	-0.594541	-0.073848	0.148338	0.015854	-1.657700	-0.210460	-0.332396	-1.013998	...
4	1.690351	0.083992	-0.594541	0.621831	-0.603869	1.581674	-1.657700	-0.978440	-1.848939	-1.763390	...

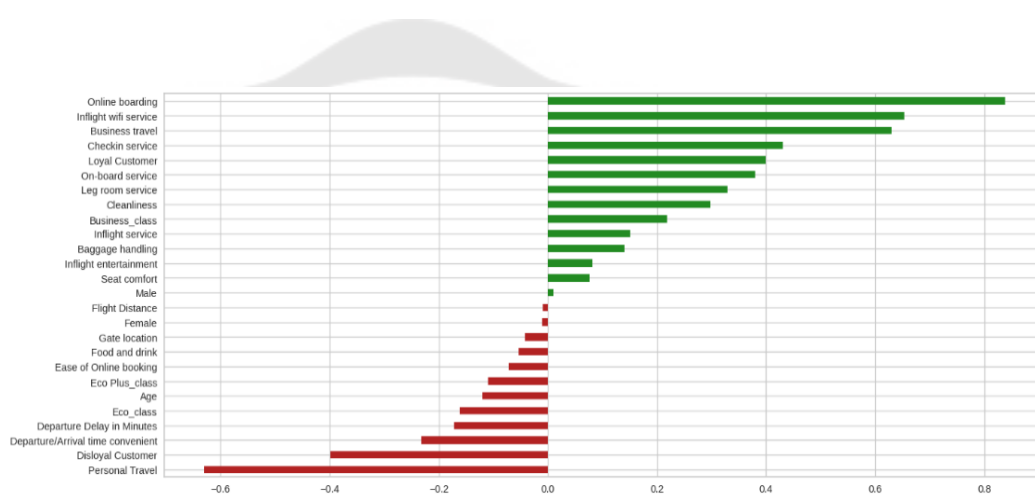
ภาพประกอบ 22 แสดงตัวอย่างข้อมูลหลังจากการทำ Normalization

ตาราง 5 แสดงผลลัพธ์การทำนายของ Logistic Regression

	Y_pred_satisfied	Y_pred_dissatisfied
Y_satisfied	19941	2095
Y_dissatisfied	2599	14329

ตาราง 6 แสดงผลลัพธ์การทำนายของ Logistic Regression

Accuracy score	88.0%	Accuracy score	88.0%
Precision (Positive)	87.2%	Precision (Negative)	88.4%
Recall (Positive)	84.6%	Recall (Negative)	90.4%
F1 (Positive)	85.9%	F1 (Negative)	89.4%



ภาพประกอบ 23 แสดงปัจจัยที่สำคัญที่ส่งผลต่อความพึงพอใจ

ตาราง 7 แสดงค่าสัมประสิทธิ์ของแต่ละปัจจัย

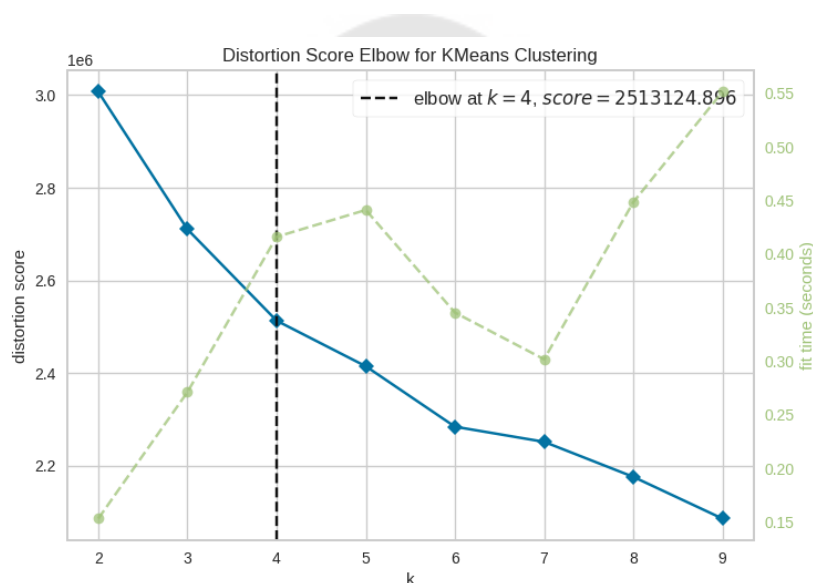
Feature	Feature Coefficient
Online boarding	0.837754
Inflight WiFi service	0.653675
Business travel	0.629881
Check in service	0.431069
Loyal Customer	0.399155
On-board service	0.379828
Leg room service	0.329539
Cleanliness	0.297204
Business class	0.217912
Inflight service	0.150124
Baggage handling	0.139762
Inflight entertainment	0.081703
Male	0.076466
Flight Distance	-0.009479
Female	-0.010478
Gate location	-0.041931
Food and drink	-0.054277
Ease of Online booking	-0.072125
Eco Plus_class	-0.109469
Age	-0.120046
Eco_class	-0.161675
Departure Delay in Minutes	-0.171585
Departure/Arrival time convenient	-0.231970
Disloyal Customer	-0.399155
Personal Travel	-0.629881

4.2 ผลจากการแบ่งกลุ่มทำ K-Means Clustering

แบบจำลองการแบ่งกลุ่มด้วย K-Means โดยใช้ข้อมูลผู้โดยสารที่ไม่พึงพอใจต่อการให้บริการสายการบิน โดยจะเป็นการแบ่งกลุ่มข้อมูล que เลือกค่า K ที่ดีที่สุด โดยใช้ Elbow Method, Silhouette Score และ Davies-Bouldin Index ดังภาพประกอบ 24 - 26 ตามลำดับ

4.2.1 ผลลัพธ์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Elbow

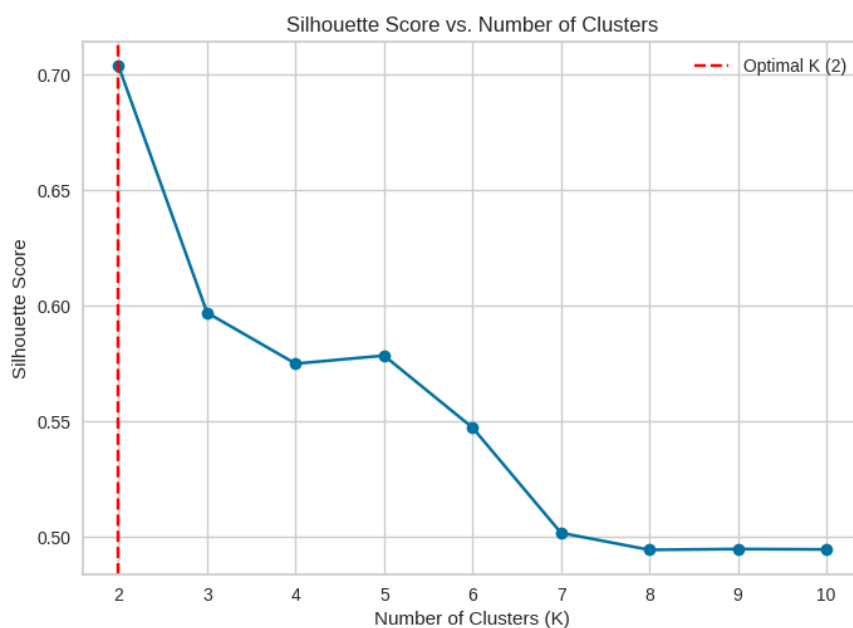
จำนวนกลุ่มที่เหมาะสมของลูกค้า เมื่อพิจารณาจากการหักศอก ได้จำนวนเท่ากับ 4 กลุ่ม ดังภาพประกอบ 24



ภาพประกอบ 24 แสดงผลลัพธ์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Elbow Method

4.2.2 ผลลัพธ์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Silhouette Score

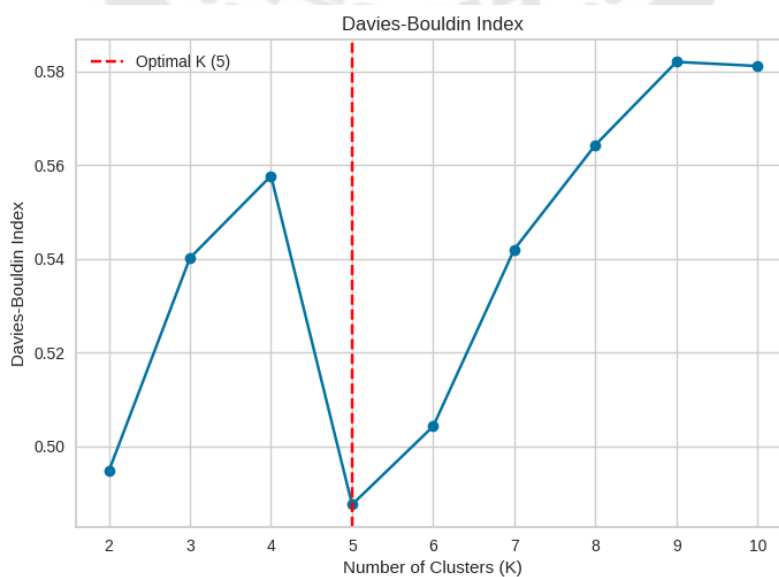
จำนวนกลุ่มที่เหมาะสมของลูกค้า เมื่อพิจารณาจากค่า Silhouette Score ที่ สูงและใกล้เคียง 1 ที่สุด ได้จำนวนเท่ากับ 2 กลุ่ม ดังภาพประกอบ 25



ภาพประกอบ 25 แสดงผลลัพธ์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Silhouette Score

4.2.3 ผลลัพธ์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Davies-Bouldin Index

จำนวนกลุ่มที่เหมาะสมของลูกค้า เมื่อพิจารณาจากค่า Davies-Bouldin index ใกล้เคียง 0 ที่สุด ได้จำนวนเท่ากับ 2,5 กลุ่ม ดังภาพประกอบ 26



ภาพประกอบ 26 แสดงผลลัพธ์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Davies-Bouldin Index

4.3 ผลลัพธ์ค่าที่เหมาะสมต่อการแบ่งกลุ่ม

ในส่วนของการประเมินผลลัพธ์การแบ่งกลุ่มข้อมูล ผู้วิจัยได้ใช้ตัวชี้วัดที่สำคัญ ได้แก่ ค่า SSE (Sum of Squared Errors), Silhouette Score และ Davies-Bouldin Index เพื่อวัดประสิทธิภาพ และความเหมาะสมของการแบ่งกลุ่ม โดยผลลัพธ์จากตัวชี้วัด แสดงดังตาราง 8

ตาราง 8 แสดงประสิทธิภาพการแบ่งกลุ่มของจำนวน K กลุ่ม

K	SSE (Sum of Squared Errors)	Silhouette Score	Davies-Bouldin Index
2	13378094223.94	0.70	0.49
3	6629755030.01	0.60	0.54
4	4036020393.57	0.57	0.56
5	2462338947.35	0.58	0.49
6	1648895402.45	0.55	0.50
7	1391287762.86	0.50	0.54
8	1108486408.03	0.49	0.56
9	922749496.43	0.49	0.58
10	743037488.22	0.49	0.58

บทที่ 5

สรุปผลการวิจัย อภิปรายผลการวิจัย และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

5.1.1 ผลการวิจัยจากการหาปัจจัยสำคัญจาก Logistic Regression

การวิเคราะห์การทำนายความพึงพอใจของผู้โดยสารสายการบินในสหรัฐฯ ที่ได้ข้อมูลจากการให้คะแนนจำนวน 129,880 คนในปี 2015 โดยใช้เทคนิค Logistic Regression เพื่อทำนายระดับความพึงพอใจและระบุปัจจัยที่มีผลต่อความพึงพอใจของผู้โดยสาร

ผลการวิจัยพบว่าแบบจำลองจาก Logistic Regression ที่สร้างขึ้นมีผลลัพธ์ความแม่นยำในการทำนายความพึงพอใจของผู้โดยสารอยู่ในระดับที่น่าพอใจ โดยมีค่าประสิทธิภาพ Accuracy: 88%, Precision: 87.2%, Recall: 84.6% และ F1-score: 85.6%

5.1.2 ผลจากการแบ่งกลุ่มด้วยวิธี K-Means

การวิเคราะห์แบ่งกลุ่มผู้โดยสารที่ไม่พึงพอใจของสายการบินในสหรัฐฯ ตามการให้คะแนนจำนวน 73,452 คน จัดเก็บข้อมูลในปี ค.ศ. 2015 โดยใช้เทคนิค K-Means ในการทำ Clustering เพื่อแบ่งกลุ่มลูกค้า กำหนดจำนวนกลุ่มที่ต้องการโดยพิจารณาจากตาราง 8

จำนวนกลุ่มที่เหมาะสมด้วยวิธี Elbow ค่าเฉลี่ยของ Silhouette Score และค่าเฉลี่ยของ Davies Bouldin โดยจากการวิเคราะห์ผลลัพธ์ของ Elbow Method, Silhouette Method และ Davies-Bouldin Index พบว่า $K = 2$ เป็นค่าที่เหมาะสมที่สุดสำหรับการทำ K-Means Clustering ด้วยเหตุผลดังต่อไปนี้ Silhouette Method และ Davies-Bouldin Index สอดคล้องกัน โดยทั้ง 2 ค่าเน้นการประเมินข้อมูลภายในกลุ่มและเปรียบเทียบกับกลุ่มอื่นจึงให้ผลไปในทิศทางเดียวกัน

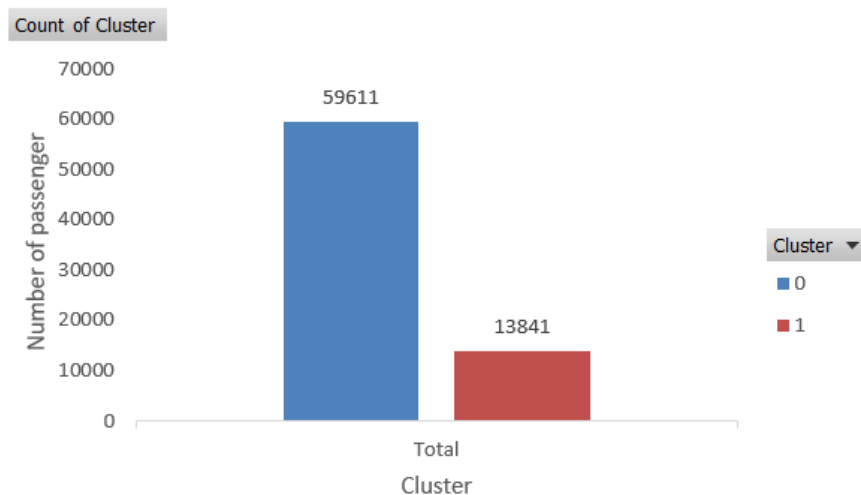
Silhouette Score สูงสุดอยู่ที่ $K = 2$ ซึ่งแสดงถึงการแบ่งกลุ่มที่ดี โดยข้อมูลในแต่ละคลัสเตอร์มีความคล้ายคลึงกันสูง และมีความแตกต่างจากคลัสเตอร์อื่น ๆ ชัดเจน Davies-Bouldin Index ต่ำสุดอยู่ที่ $K = 2$ ซึ่งแสดงถึงการแบ่งกลุ่มที่ดีเช่นกัน โดยคลัสเตอร์มีความกระจายตัวน้อย และมีความห่างจากคลัสเตอร์อื่น ๆ มาก การที่ทั้งสอง Metrics ให้ผลลัพธ์สอดคล้องกัน ช่วยเพิ่มความมั่นใจในการเลือก $K = 2$

5.1.3 ผลลัพธ์ข้อมูลส่วนตัวและปัจจัยที่ควบคุมไม่ได้ของแต่ละกลุ่ม

ผลลัพธ์ที่ได้จากการแบ่งกลุ่ม ผู้โดยสารโดยใช้วิธี K-Means Clustering ดังภาพประกอบ 27 แบ่งได้เป็น 2 กลุ่ม โดยมีลักษณะ ดังนี้

คลัสเตอร์ที่ 0 มีสมาชิก 59,611 คน

คลัสเตอร์ที่ 1 มีสมาชิก 13,841 คน



ภาพประกอบ 27 จำนวนผู้โดยสารของทั้ง 2 กลุ่ม

ผลลัพธ์ที่ได้จากการแบ่งกลุ่มผู้โดยสารตามประเภทชั้นที่นั่ง ดังภาพประกอบ 28 และมีรายละเอียด ดังนี้

คลัสเตอร์ที่ 0 ผู้โดยสารในกลุ่มนี้ส่วนมากเป็นผู้โดยสารที่นั่งชั้น Economy Class ซึ่งมีจำนวน 48,650 คน และ Business Class จำนวน 10,960 คน

คลัสเตอร์ที่ 1 ผู้โดยสารในกลุ่มนี้ส่วนมากเป็นผู้โดยสารที่นั่งชั้น Business Class ซึ่งมีจำนวน 8,033 คน และ Eco Class จำนวน 5,808 คน



ภาพประกอบ 28 จำนวนผู้โดยสารของทั้ง 2 กลุ่ม ตามประเภทของชั้นโดยสาร

ผลลัพธ์ที่ได้จากการแบ่งกลุ่มผู้โดยสารตามประเภทการเดินทาง ดังภาพประกอบ 29 และมีรายละเอียด ดังนี้

คลัสเตอร์ที่ 0 ผู้โดยสารในกลุ่มนี้ส่วนมากเดินทางด้วยเหตุผลส่วนตัวซึ่งมีจำนวน 31,470 คน และเดินทางด้วยเหตุผลทางธุรกิจจำนวน 28,141 คน

คลัสเตอร์ที่ 1 ผู้โดยสารในกลุ่มนี้ส่วนมากเดินทางด้วยเหตุผลทางธุรกิจจำนวน 9,196 คนและ เดินทางด้วยเหตุผลส่วนตัวซึ่งมีจำนวน 4,645 คน

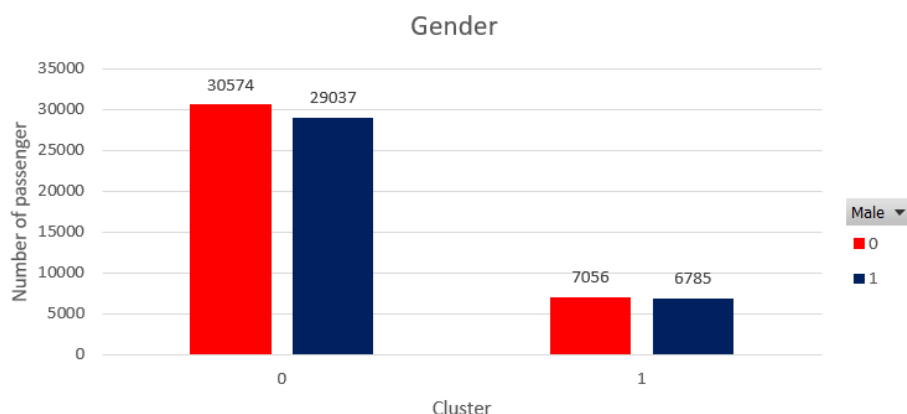


ภาพประกอบ 29 จำนวนผู้โดยสารของทั้ง 2 กลุ่ม ตามประเภทการเดินทาง

ผลลัพธ์ที่ได้จากการแบ่งกลุ่มผู้โดยสารตามเพศ ดังภาพประกอบ 30 และมีรายละเอียด ดังนี้

คลัสเตอร์ที่ 0 ผู้โดยสารในกลุ่มนี้เป็นเพศหญิงซึ่งมีจำนวน 30,574 คน และเพศชายจำนวน 29,037 คน

คลัสเตอร์ที่ 1 ผู้โดยสารในกลุ่มนี้เป็นเพศหญิงซึ่งมีจำนวน 7,056 คน และเพศชายจำนวน 6,785 คน



ภาพประกอบ 30 จำนวนผู้โดยสารของทั้ง 2 กลุ่ม ตามเพศของผู้โดยสาร

ผลลัพธ์ที่ได้จากการหาค่าเฉลี่ยของอายุและระยะทางของเที่ยวบิน ดังภาพประกอบ 31 และมีรายละเอียด ดังนี้

คลัสเตอร์ที่ 0 ผู้โดยสารในกลุ่มนี้มีค่าเฉลี่ยอายุอยู่ที่ 37 ปี เดินทางในระยะสั้น โดยค่าเฉลี่ยอยู่ที่ 607.945 ไมล์

คลัสเตอร์ที่ 1 ผู้โดยสารในกลุ่มนี้มีค่าเฉลี่ยอายุอยู่ที่ 39 ปี เดินทางในระยะไกล โดยค่าเฉลี่ยอยู่ที่ 2,315 ไมล์

Row Labels	Average of Age	Average of Flight Distance
0	37.29400614	607.9456308
1	39.18849794	2315.527057

ภาพประกอบ 31 ค่าเฉลี่ยอายุและระยะทางของเที่ยวบินของทั้ง 2 กลุ่ม

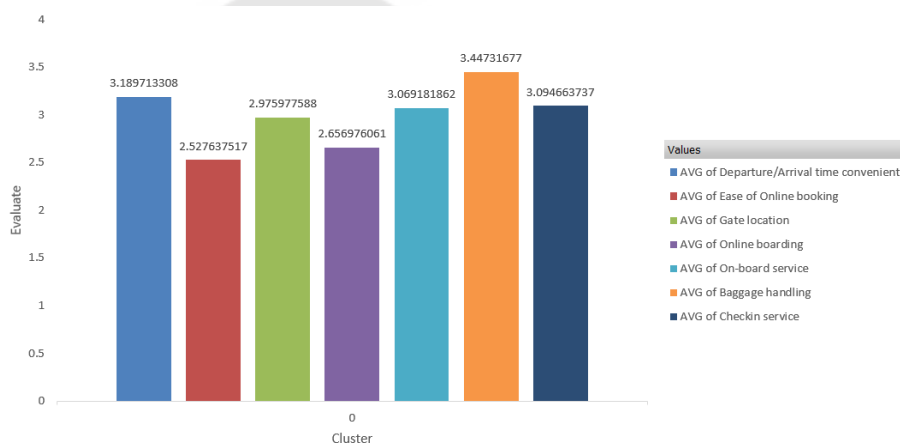
ค่าเฉลี่ยของ Departure Delay in Minutes ของคลัสเตอร์ที่ 0 และคลัสเตอร์ที่ 1 อยู่ที่ 81.8 นาทีและ 57.4 นาที ตามลำดับ

5.1.4 ผลลัพธ์การให้บริการของแต่ละกลุ่ม

ผลลัพธ์ที่ได้จากการประเมินความพึงพอใจของการบริการภาคพื้น โดยรายละเอียดคลัสเตอร์ที่ 0 ดังภาพประกอบ 32 และคลัสเตอร์ที่ 1 ดังภาพประกอบ 33 และมีรายละเอียด ดังนี้

คลัสเตอร์ที่ 0 ผู้โดยสารในกลุ่มไม่พึงพอใจต่อการจองแบบออนไลน์ มากที่สุด โดยการประเมินเฉลี่ยอยู่เพียง 2.52 คะแนน

คลัสเตอร์ที่ 1 ผู้โดยสารในกลุ่มไม่พึงพอใจต่อการบอร์ดผู้โดยสารขึ้นเครื่อง มากที่สุด โดยการประเมินเฉลี่ยอยู่เพียง 2.76 คะแนน



ภาพประกอบ 32 ค่าเฉลี่ยคะแนนการให้บริการภาคพื้นต่อผู้โดยสารคลัสเตอร์ 0

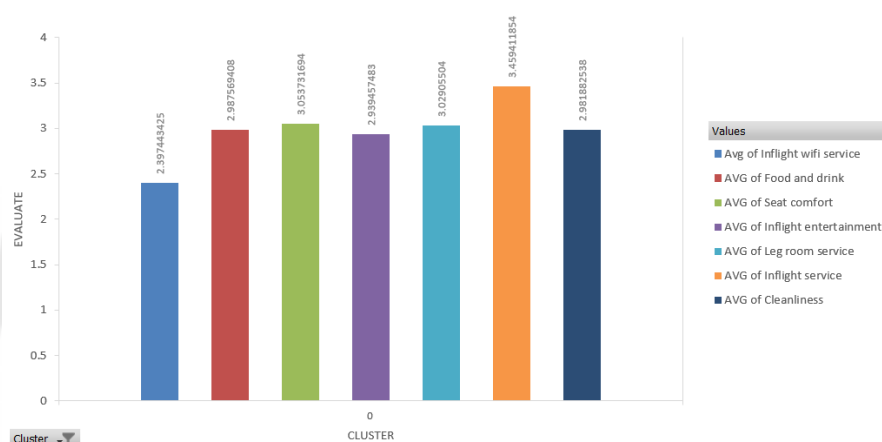


ภาพประกอบ 33 ค่าเฉลี่ยคะแนนการให้บริการภาคพื้นต่อผู้โดยสารคลัสเตอร์ 1

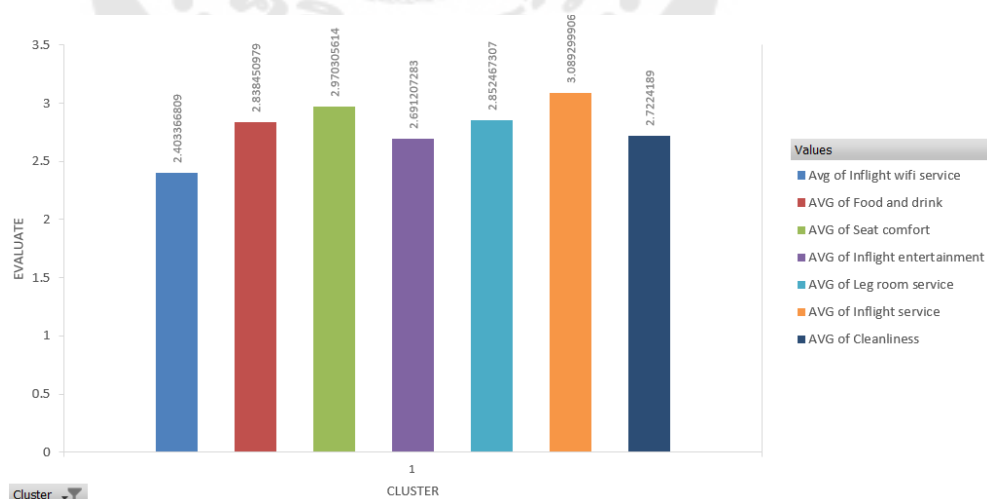
ผลลัพธ์ที่ได้จากการประเมินความพึงพอใจของการบริการบนเครื่องบิน โดยรายละเอียดคลัสเตอร์ที่ 0 ดังภาพประกอบ 34 และคลัสเตอร์ที่ 1 ดังภาพประกอบ 35 และมีรายละเอียด ดังนี้

คลัสเตอร์ที่ 0 ผู้โดยสารในกลุ่มไม่พึงพอใจต่อการให้บริการอินเทอร์เน็ตบนเครื่องบินมากที่สุดโดยการประเมินเฉลี่ยอยู่เพียง 2.39 คะแนน

คลัสเตอร์ที่ 1 ผู้โดยสารในกลุ่มไม่พึงพอใจต่อการให้บริการอินเทอร์เน็ตบนเครื่องบินมากที่สุดโดยการประเมินเฉลี่ยอยู่เพียง 2.40 คะแนน



ภาพประกอบ 34 ค่าเฉลี่ยคะแนนการให้บริการบนเครื่องบินต่อผู้โดยสารคลัสเตอร์ 0



ภาพประกอบ 35 ค่าเฉลี่ยคะแนนการให้บริการบนเครื่องบินต่อผู้โดยสารคลัสเตอร์ 1

5.2 อภิปรายผลการวิจัย

งานวิจัยนี้ผู้วิจัยได้มุ่งเน้นที่การศึกษาวิธีการหาปัจจัยที่สำคัญต่อความพึงพอใจของผู้โดยสารด้วยแบบจำลอง Logistic Regression และการศึกษาวิธีการแบ่งกลุ่มผู้โดยสารที่ไม่พึงพอใจตามพฤติกรรมการให้คะแนน โดยใช้เทคนิค K-Means ในการทำ Clustering เพื่อแบ่งกลุ่มผู้โดยสาร พิจารณาจำนวนกลุ่มที่เหมาะสมด้วยวิธี Elbow ค่าของ Silhouette Score และค่าของ Davies-Bouldin Index และนำผลมาวิเคราะห์เปรียบเทียบระหว่างผลลัพธ์ โดยมีผลลัพธ์ ดังนี้

5.2.1 อภิปรายผลลัพธ์การหาปัจจัยที่สำคัญต่อความพึงพอใจ

ผลลัพธ์ของการหา Feature Importance จากแบบจำลอง Logistic Regression ซึ่งพบว่าปัจจัยการบริการที่มีอิทธิพลที่ส่งผลให้ผู้โดยสารพึงพอใจมากที่สุดได้แก่ การเช็คอินออนไลน์ (Online Boarding), การให้บริการอินเทอร์เน็ตบนเครื่อง (Inflight Wi-Fi Service) และการให้บริการเช็คอิน (Check in Service) ตามลำดับ โดยปัจจัยการบริการที่มีอิทธิพลที่ส่งผลให้ผู้โดยสารไม่พึงพอใจได้มากที่สุดได้แก่ ความเหมาะสมของเวลาเที่ยวบิน (Departure/Arrival time convenient), การจองแบบออนไลน์ (Ease of Online Booking) และอาหารและเครื่องดื่ม (Food and Drink) ตามลำดับ

ปัจจัยที่มีผลกระทบต่อความพึงพอใจของผู้โดยสารนอกเหนือจากการบริการของสายการบิน พบว่า ผู้โดยสารที่เดินทางด้วยเหตุผลทางธุรกิจ (Business Travel) และผู้โดยสารที่เดินทางประจำ (Loyal Customer) มีแนวโน้มที่จะพึงพอใจมากที่สุด ส่วนปัจจัยที่ส่งผลให้ผู้โดยสารไม่พึงพอใจมากที่สุด ได้แก่ การเดินทางด้วยเหตุผลส่วนตัว (Personal Travel), ผู้โดยสารที่เดินทางไม่ประจำ (Disloyal Customer), และความล่าช้าของการออกเดินทาง (Departure Delay in Minutes) ตามลำดับ

ผลการวิเคราะห์เหล่านี้สามารถใช้เป็นข้อมูลในการพัฒนากลยุทธ์เพื่อปรับปรุงการบริการและประสบการณ์การเดินทางของผู้โดยสาร โดยการมุ่งเน้นไปที่ปัจจัยที่มีอิทธิพลสำคัญต่อความพึงพอใจของผู้โดยสารเพื่อให้การบริการดีขึ้นและตอบสนองความต้องการของลูกค้าได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

5.2.2 อภิปรายผลลัพธ์การแบ่งกลุ่มผู้โดยสารที่ไม่พึงพอใจตามพฤติกรรมการให้คะแนน

ผลลัพธ์ของ Elbow Method ได้จำนวนเท่ากับ 4 กลุ่ม ค่าของ Silhouette Score อยู่ที่ 0.704 ได้จำนวนเท่ากับ 2 กลุ่ม และค่าของ Davies-Bouldin Index อยู่ที่ 0.49 ได้จำนวนเท่ากับ 2 และ 5 กลุ่ม

ผู้วิจัยเลือกแบ่งกลุ่มลูกค้าออกเป็น 2 กลุ่มโดยให้เหตุผลว่าค่าที่เหมาะสมที่สุดสำหรับการทำ K-Means Clustering โดย Silhouette Method และ Davies-Bouldin Index สอดคล้องกัน Silhouette Score สูงสุดอยู่ที่ $K = 2$ ซึ่งแสดงถึงการแบ่งกลุ่มที่ดี โดยข้อมูลในแต่ละคลาสเตอร์มีความคล้ายคลึงกันสูง และมีความแตกต่างจากคลาสเตอร์อื่น ๆ ชัดเจน โดยแต่ละกลุ่มมีลักษณะดังนี้

คลาสเตอร์ที่ 0 มีสมาชิก 59,611 คน ผู้โดยสารในกลุ่มนี้ส่วนมากเป็นผู้โดยสารที่นั่งชั้น Economy Class โดยมีจำนวน 48,650 คน โดยมีชั้น Business Class จำนวน 10,961 คน เดินทางด้วยเหตุผลส่วนตัว (Personal Travel) จำนวน 31,470 คน, เดินทางด้วยเหตุผลทางธุรกิจ จำนวน 28,141 คน เป็น Loyal Customer จำนวน 42,847 คน และเป็น Disloyal Customer จำนวน 16,764 คน, เป็นเพศหญิงซึ่งมีจำนวน 30,574 คน และเพศชายจำนวน 29,037 คน, ค่าเฉลี่ยของการออกเดินทางล่าช้าอยู่ที่ 81.8 นาที, ค่าเฉลี่ยอายุอยู่ที่ 37 ปี และการบริการภาคพื้นที่ไม่พึงพอใจมากที่สุดคือการจองแบบออนไลน์ (Ease of Online booking) โดยการประเมินเฉลี่ยอยู่เพียง 2.52 คะแนน และการบริการบนเครื่องบินที่ได้คะแนนน้อยที่สุดคือการบริการอินเทอร์เน็ตบนเครื่องบิน (Inflight Wi-Fi Service) ซึ่งได้คะแนนเฉลี่ย 2.39 คะแนน

คลาสเตอร์ที่ 1 มีสมาชิก 13,841 คน ผู้โดยสารในกลุ่มนี้ส่วนมากเป็นผู้โดยสารที่นั่งชั้น Business class โดยมีจำนวน 8,033 คน โดยมีชั้น Economy Class จำนวน 5,808 คน เดินทางด้วยเหตุผลทางธุรกิจ (Business Travel) จำนวน 9,196 คน เดินทางด้วยเหตุผลส่วนตัว (Personal Travel) จำนวน 4,645 คน เป็น Loyal Customer จำนวน 12,525 คน และเป็น Disloyal Customer จำนวน 1,316 คน ค่าเฉลี่ยของการออกเดินทางล่าช้าอยู่ที่ 57.4 นาที ค่าเฉลี่ยอายุอยู่ที่ 39 ปี ไม่พึงพอใจต่อการบอร์ดผู้โดยสารขึ้นเครื่อง (Online boarding) มากที่สุดโดยการประเมินเฉลี่ยอยู่เพียง 2.76 คะแนน และการบริการบนเครื่องบินที่ได้คะแนนน้อยที่สุดคือการบริการอินเทอร์เน็ตบนเครื่องบิน (Inflight Wi-Fi service) ซึ่งได้คะแนนเฉลี่ย 2.40 คะแนน

5.2.3 อภิปรายผลลัพธ์ที่สอดคล้องกัน

ผลการวิจัยแสดงให้เห็นถึงปัจจัยที่มีผลกระทบต่อความพึงพอใจและไม่พึงพอใจของผู้โดยสารสายการบิน ซึ่งเชื่อมโยงกับผลลัพธ์จากการแบ่งกลุ่ม (Clustering) ด้วยเทคนิค K-Means พบว่าปัจจัยที่มีผลต่อความพึงพอใจมากที่สุด ได้แก่ การเช็คอินออนไลน์ (Online Boarding), การให้บริการอินเทอร์เน็ตบนเครื่องบิน (Inflight Wi-Fi Service), และการให้บริการเช็คอิน (Check-in Service) ซึ่งมีผลบวกต่อประสบการณ์ของผู้โดยสารในกลุ่มที่มีความพึงพอใจสูง ขณะที่ใน

คลัสเตอร์ 0 และ 1 ผู้โดยสารมีความพึงพอใจต่ำสุดในเรื่องการบริการบนเครื่อง โดยเฉพาะการให้บริการอินเทอร์เน็ตบนเครื่อง (Inflight Wi-Fi Service)

ในขณะเดียวกัน ปัจจัยที่ส่งผลต่อความไม่พึงพอใจ เช่น ความเหมาะสมของเวลาเที่ยวบิน (Departure/Arrival Time Convenient) การจองแบบออนไลน์ (Ease of Online Booking) และอาหารและเครื่องดื่ม (Food and Drink) เชื่อมโยงกับกลุ่มผู้โดยสารที่มีความไม่พึงพอใจ โดยเฉพาะคลัสเตอร์ที่ 0 ซึ่งส่วนใหญ่เป็นผู้โดยสารชั้นประหยัด (Eco Class) และเดินทางด้วยเหตุผลส่วนตัว(Personal Travel) ผู้โดยสารกลุ่มนี้ให้คะแนนต่ำในเรื่องการจองแบบออนไลน์ (Ease of Online booking) สะท้อนถึงความไม่สอดคล้องระหว่างการให้บริการกับความคาดหวัง

ผลการแบ่งกลุ่มด้วย K-Means ยังแสดงให้เห็นว่า คลัสเตอร์ที่ 1 ซึ่งประกอบด้วยผู้โดยสารชั้นธุรกิจที่เดินทางเพื่อธุรกิจ มีความคาดหวังในคุณภาพการบริการที่สูงกว่า แม้ว่าจะให้คะแนนสูงกว่าคลัสเตอร์ที่ 0 ในบางด้าน แต่ยังพบว่าการใช้คอนออนไลน์ (Online Boarding) และการให้บริการอินเทอร์เน็ตบนเครื่อง (Inflight Wi-Fi Service) ได้รับการประเมินค่อนข้างต่ำ (2.76 และ 2.40 คะแนนตามลำดับ) ซึ่งผลลัพธ์นี้สอดคล้องกับการวิเคราะห์จาก Logistic Regression ที่ชี้ให้เห็นว่าปัจจัยเหล่านี้มีความสำคัญอย่างยิ่งต่อความพึงพอใจของผู้โดยสาร

จากผลลัพธ์ดังกล่าว การเชื่อมโยงข้อมูลปัจจัยสำคัญกับการแบ่งกลุ่มช่วยให้เข้าใจถึงความแตกต่างในพฤติกรรมและความคาดหวังของผู้โดยสารแต่ละกลุ่มได้ชัดเจนยิ่งขึ้น สายการบินสามารถใช้ข้อมูลนี้เป็นแนวทางในการปรับปรุงบริการเฉพาะกลุ่ม เช่น การพัฒนาคุณภาพการจองออนไลน์สำหรับผู้โดยสารชั้นประหยัด และการเสริมประสิทธิภาพการใช้งานอินเทอร์เน็ตบนเครื่องสำหรับผู้โดยสาร เพื่อเพิ่มความพึงพอใจในทุกกลุ่มอย่างครอบคลุม

5.3 ข้อเสนอแนะ

เพื่อการพัฒนางานวิจัยในอนาคต อาจประยุกต์ใช้วิธีการนี้ และสามารถปรับปรุงแบบจำลองในการสำรวจเพิ่มเติมได้โดยเพิ่มตัวแปรที่มีผลต่อพฤติกรรมของผู้โดยสาร รวมทั้งเวลาที่ชัดเจนของข้อมูล หรือใช้แบบจำลองประเภทอื่นในการเปรียบเทียบเพื่อประสิทธิภาพที่ดีกว่า



ภาคผนวก

```

from sklearn.preprocessing import StandardScaler

# สร้าง StandardScaler
s_scaler = StandardScaler()

# ใช้ StandardScaler เพื่อ Standardization
df_scaled = pd.DataFrame(s_scaler.fit_transform(df_clean), columns=df_clean.columns)
df_scaled.head()

```

	Age	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	Online boarding	Seat comfort	Inflight entertainment	...	satisfaction	Male	Female	Loyal Customer	Disloyal Customer	Personal Travel	Business travel	Eco_class
0	-1.747961	-0.732184	0.188780	0.619987	0.150052	-1.546291	1.351205	-0.212367	1.181437	1.230957	...	-0.876487	0.104878	-1.014878	0.473422	-0.473422	1.493951	-1.493951	-0.902608
1	-0.954274	-0.957780	0.188780	-0.771740	0.150052	0.018043	-1.661163	-0.212367	-1.850550	-1.768113	...	-0.876487	0.104878	-1.014878	-2.112281	2.112281	-0.669366	0.669366	-0.902608
2	-0.888133	-0.048440	-0.594042	-0.771740	-0.602556	-0.764124	1.351205	1.324589	1.181437	1.230957	...	1.140918	-0.985340	0.985340	0.473422	-0.473422	-0.669366	0.669366	-0.902608
3	-0.954274	-0.029624	-0.594042	1.315851	1.655269	1.582376	-0.908071	-0.980845	-1.029553	-1.018346	...	-0.876487	-0.985340	0.985340	0.473422	-0.473422	-0.669366	0.669366	-0.902608
4	1.426788	-0.978814	0.188780	-0.075876	0.150052	0.018043	0.598113	1.324589	1.181437	-0.268578	...	1.140918	0.104878	-1.014878	0.473422	-0.473422	-0.669366	0.669366	-0.902608

5 rows x 27 columns

3. รายละเอียดการหาปัจจัยที่สำคัญ (Feature Importance)

```
[ ] import pandas as pd

# สร้าง DataFrame จากค่าสัมประสิทธิ์ของ feature
feature_importance = pd.DataFrame({'Feature': X_test.columns, 'Coefficient': log_model.coef_[0]})

# เรียงลำดับ DataFrame ตามค่าสัมประสิทธิ์ (absolute value)
feature_importance = feature_importance.sort_values(by='Coefficient', ascending=False)

# แสดงผลลัพธ์
print(feature_importance)
```

	Feature	Coefficient
7	Online boarding	0.837754
2	Inflight wifi service	0.653675
22	Business travel	0.629881
13	Checkin service	0.431069
19	Loyal Customer	0.399155
10	On-board service	0.379828
11	Leg room service	0.329539
15	Cleanliness	0.297204
25	Business class	0.217912
14	Inflight service	0.158124
12	Baggage handling	0.139762
9	Inflight entertainment	0.081703
8	Seat comfort	0.076466
17	Male	0.019478
1	Flight Distance	-0.009479
18	Female	-0.018478
5	Gate location	-0.041931
6	Food and drink	-0.054277
4	Ease of Online booking	-0.072125
24	Eco Plus class	-0.109469
0	Age	-0.120046
23	Eco class	-0.161675

4. รายละเอียดการหาจำนวนกลุ่มที่เหมาะสมกับข้อมูล

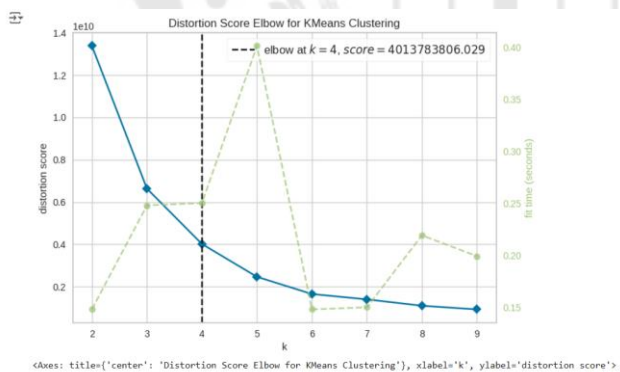
```
[ ] from sklearn.cluster import KMeans
from yellowbrick.cluster import KElbowVisualizer

# สร้าง KMeans model
model = KMeans()

# ใช้ Elbow method เพื่อหาจำนวน clusters ที่เหมาะสม
visualizer = KElbowVisualizer(model, k=(2,10))

# Fit ข้อมูลเข้ากับ visualizer
visualizer.fit(df_cluster)

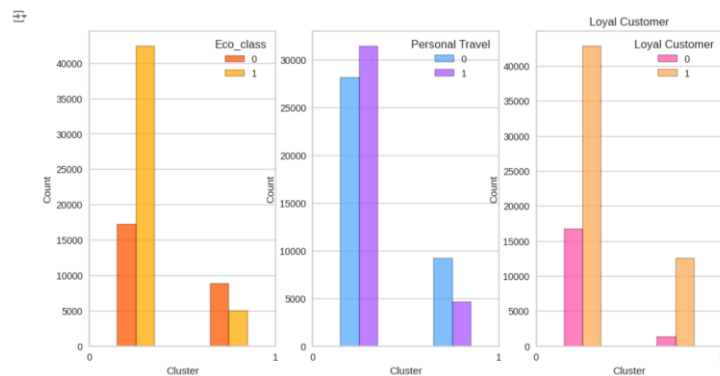
# แสดงกราฟ
visualizer.show()
```



5. รายละเอียดการแบ่งกลุ่มด้วยวิธีการ K-means

```
[ ] pal=['autumn','cool','spring']
fig,axs=plt.subplots(ncols=3,figsize =(12, 6))

for idx, i in enumerate(Info.columns):
    plt.title(i)
    ax = sns.histplot(data=df_cluster, x='Cluster', hue=i, palette=pal[idx], binwidth=.6, ax=axs[idx], multiple="dodge", shrink=.4)
    ax.set_xticks(range(df_cluster['Cluster'].nunique())) # กำหนด scale ให้เป็นจำนวนเต็ม
plt.show()
```



บรรณานุกรม

- Abdul-Rahman, S., Arifin, N. F. K., Hanafiah, M., & Mutalib, S. (2021). Customer segmentation and profiling for life insurance using k-modes clustering and decision tree classifier. *International Journal of Advanced Computer Science and Applications*, 12(9), 434-444. <https://doi.org/10.14569/IJACSA.2021.0120950>
- An, M., & Noh, Y. (2009). Airline customer satisfaction and loyalty: impact of in-flight service quality. *Service Business*, 3, 293-307.
- Archana, R., & Subha, M. (2012). A study on service quality and passenger satisfaction on Indian airlines. *International Journal of Multidisciplinary Research*, 2(2), 50-63.
- Khandelwal, K., Jakhar, T., & Khandelwal, T. (2020). Segmentation, Targeting and Positioning. *International Research Journal of Engineering and Technology (IRJET)*, 3909-3914.
- Nandapala, E., & Jayasena, K. (2020). The practical approach in Customers segmentation by using the K-Means Algorithm. 2020 IEEE 15th international conference on industrial and information systems (ICIIS),
- Noviantoro, T., & Huang, J.-P. (2022). Investigating airline passenger satisfaction: Data mining method. *Research in Transportation Business & Management*, 43, 100726. <https://doi.org/https://doi.org/10.1016/j.rtbm.2021.100726>
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12-40.
- Park, S., Lee, J.-S., & Nicolau, J. L. (2020). Understanding the dynamics of the quality of airline service attributes: Satisfiers and dissatisfiers. *Tourism Management*, 81, 104163. <https://doi.org/https://doi.org/10.1016/j.tourman.2020.104163>
- Sahinbas, K. (2022). Performance Comparison of K-Means and DBSCAN Methods for Airline Customer Segmentation. *Black Sea Journal of Engineering and Science*, 5(4), 158-165. <https://doi.org/10.34248/bsengineering.1170943>
- Wang, J., Wang, H., Nie, F., & Li, X. (2023). Feature selection with multi-class logistic

regression. *Neurocomputing*, 543, 126268.

<https://doi.org/https://doi.org/10.1016/j.neucom.2023.126268>

ชลิต เจียบพิมาย, & วสันต์ นิลมัย. (2562). แนวทางการพัฒนาคุณภาพการบริการในธุรกิจการบิน. (Airline Service Quality Improvement Approach for Aviation Business). การประชุม วิชาการระดับชาติ การท่องเที่ยวและการโรงแรมร่วมสมัย ครั้งที่ 6,



ประวัติผู้เขียน

ชื่อ-สกุล

ปฏิภาณ เกริกชัยวัน

