

การวิเคราะห์ข้อมูลจากการเดินทางของรถแท็กซี่

## DATA ANALYTICS ON TAXI TRAJECTORIES

ศาสตราจารย์ ดร. ชัยวงศ์ ศรีอรุณ

ชนพล คำเพราะ

ปริญญาานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2565

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญาานิพนธ์ปีการศึกษา 2565

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การวิเคราะห์ข้อมูลจากการเดินทางของรถแท็กซี่

DATA ANALYTICS ON TAXI TRAJECTORIES

ผู้จัดทำ

1. นายตาน ชัยวงศ์ศรีอรุณ รหัสนักศึกษา 62010309

2. นายชนพล คำเพราะ รหัสนักศึกษา 62010355



  
(อาจารย์รัฐชัย ชาวอุทัย)

อาจารย์ที่ปรึกษา

## การวิเคราะห์ข้อมูลจากการเดินทางของรถแท็กซี่

นายตาด ชัยวงศ์ศรีอรุณ 62010309  
 นายชนพล คำเพราะ 62010355  
 ผศ.ดร.รัฐชัย ชาวอุทัย อาจารย์ที่ปรึกษา  
 ปีการศึกษา 2565

### บทคัดย่อ

โครงการนี้มีวัตถุประสงค์เพื่อให้ข้อมูลเชิงลึกเกี่ยวกับความหนาแน่นของผู้โดยสารในพื้นที่ต่างๆ ของกรุงเทพฯ โดยเน้นการระบุจุดรับ-ส่งที่ได้รับความนิยม โดยใช้ข้อมูลจีพีเอส (GPS) จากการเดินทางของรถแท็กซี่ในปี ค.ศ. 2020 โครงการนี้จะใช้การวิเคราะห์ข้อมูลเพื่อวิเคราะห์จุดรับ-ส่งที่ได้รับความนิยมและแนะนำจุดรับที่เหมาะสมในแต่ละช่วงเวลา นอกจากนี้โครงการยังมีวัตถุประสงค์เปรียบเทียบจุดรับ-ส่งที่ได้รับความนิยมโดยดูข้อมูลระยะทาง เวลาเดินทาง และรายได้ ข้อมูลเหล่านี้สามารถช่วยในการวางแผนและจัดการการขนส่งได้ ช่วยให้ผู้ให้บริการแท็กซี่สามารถวางแผนการวิ่งเพื่อเพิ่มรายได้ได้

## Data Analytics on Taxi Trajectories

Mr. Tan Chaiwongsriaroon 62010309

Mr. Tanapol Kampoh 62010355

Asst.Prof. Rathachai Chawuthai Advisor

Academic Year 2022

### ABSTRACT

This project aims to provide insights into the passenger density in different areas of Bangkok, specifically focusing on identifying popular pickup/dropoff clusters. Using GPS data from taxi trajectories in 2020, the study utilizes data analytics to find clusters and identify time-based pickup points recommendations. Additionally, the project aims to compare popular pickup and dropoff clusters based on insights on distance, travel time, and fare. These insights can be useful for transportation planning and management, enabling taxi operators to optimize their routes and increase their profitability.

## กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้สามารถสำเร็จได้ด้วยการช่วยเหลือจากบุคคลที่มีพระคุณหลาย ๆ ท่านที่  
ให้ทรัพยากร คำปรึกษาด้านวิชาการที่เป็นประโยชน์ จึงขอขอบพระคุณในการช่วยเหลือต่าง ๆ ณ ที่นี้

ผู้จัดทำขอขอบพระคุณ ผศ.ดร.รัฐชัย ชาวอุทัย อาจารย์ที่ปรึกษาของโครงการผู้ให้แนวทางการ  
การดำเนินโครงการความรู้ที่จำเป็นต่อโครงการ และผู้จัดทำขอขอบพระคุณครอบครัว คณาจารย์ภาค  
วิศวกรรมคอมพิวเตอร์สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุกท่าน และข้อมูลเปิด  
ของมูลนิธิศูนย์ข้อมูลจราจรอัจฉริยะไทย (iTIC Open Data Archives) ที่ให้การสนับสนุนในด้านต่างๆ  
ทำให้ปริญญานิพนธ์ฉบับนี้สำเร็จได้

ผู้จัดทำหวังว่าปริญญานิพนธ์ฉบับนี้จะมีประโยชน์ต่อผู้ที่สนใจไม่มากก็น้อยหากมีข้อผิดพลาด  
ไม่ว่าด้านใด ผู้จัดทำขออภัยในข้อผิดพลาดนั้นมา ณ ที่นี้ด้วย

ตาม ชัยวงศ์ศรีอรุณ  
ธนพล คำเพราะ

# สารบัญ

	หน้า
บทคัดย่อ.....	I
ABSTRACT.....	II
สารบัญ.....	IV
สารบัญรูปภาพ.....	VI
สารบัญตาราง.....	VIII
บทที่ 1 บทนำ.....	1
1.1    ความเป็นมาของปัญหา.....	1
1.2    วัตถุประสงค์ของโครงการ.....	2
1.3    ขอบเขตของโครงการ.....	2
1.4    ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.5    แผนการดำเนินงาน.....	3
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	4
2.1    งานวิจัยที่เกี่ยวข้อง.....	4
2.2    เทคนิคที่นำมาใช้ในงานวิจัย.....	8
บทที่ 3 การออกแบบและการพัฒนา.....	15
3.1    รายละเอียดของข้อมูลที่ใช้.....	15
3.2    ขั้นตอนการพัฒนา.....	18
3.3    เป้าหมายของการพัฒนา.....	20
3.4    การเตรียมข้อมูล.....	22

3.5	การวิเคราะห์ข้อมูล .....	28
บทที่ 4 ผลการทดลอง.....		33
4.1	คลัสเตอร์จัดรับส่งผู้โดยสารขออนियม .....	33
4.2	การหาความสัมพันธ์ของข้อมูล .....	34
4.3	การแนะนำจุดรับผู้โดยสารตามช่วงเวลา .....	39
4.4	ผลลัพธ์จากการศึกษาอื่นๆ.....	44
บทที่ 5 สรุปผลและปัญหาของโครงการ .....		48
5.1	สรุปผลการดำเนินงาน.....	48
5.2	ปัญหา อุปสรรค และแนวทางการแก้ปัญหา.....	49
5.3	แนวทางการพัฒนาต่อ .....	50
บรรณานุกรม.....		51
ภาคผนวก.....		53

## สารบัญรูปภาพ

รูปภาพ	หน้า
รูปที่ 1.1 การกระจายตัวของพื้นที่ที่สนใจ (ที่มา : Zheng, et al., 2017) .....	7
รูปที่ 2.1 ดีบีเอสแกน (DBSCAN) (ที่มา : geeksforgeeks, 2019) .....	9
รูปที่ 2.2 กฎของฮาเวอร์ไซน์ (Haversine Formula) (ที่มา : Wikipedia, 2022) .....	10
รูปที่ 2.3 รูปข้อมูลพิกัดพื้นที่ (Polygon) ของกรุงเทพมหานคร (ที่มา : openstreetmap.org, 2023).....	14
รูปที่ 3.1 ตัวอย่างข้อมูลคู่พิกัดพื้นที่ (polygon) (ที่มา : polygons.openstreetmap.fr, 2022) .....	17
รูปที่ 3.2 แผนที่กรุงเทพมหานครจากข้อมูลพิกัดพื้นที่ (polygon) (ที่มา : polygons.openstreetmap.fr, 2022) .....	17
รูปที่ 3.3 ภาพรวมการพัฒนางานวิจัย .....	18
รูปที่ 3.4 สูตรของฮาเวอร์ไซน์ (Haversine Law) .....	25
รูปที่ 3.5 สูตรการคำนวณเงิน (price) จากระยะทาง (distance) .....	27
รูปที่ 4.1 รูปแสดงจุด cluster ของจุดรับผู้โดยสาร (pickup points).....	33
รูปที่ 4.2 รูปแสดงจุด cluster ของจุดส่งผู้โดยสาร (dropoff points).....	34
รูปที่ 4.3 รูปแสดงกราฟรูปแผนภูมิ (Box Plot) แสดงความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอดนิยมกับระยะทาง (Distance).....	35
รูปที่ 4.4 รูปแสดงกราฟรูปแผนภูมิ (Box Plot) แสดงความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอดนิยมกับค่าโดยสาร (Price).....	36
รูปที่ 4.5 รูปแสดงกราฟรูปแผนภูมิ (Box Plot) แสดงความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอดนิยมกับเวลาเดินทาง (travel time).....	38
รูปที่ 4.6 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 06:00-12:00 ในวันธรรมดา (weekday) .....	40
รูปที่ 4.7 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 12:00-16:00 ในวันธรรมดา (weekday) .....	40
รูปที่ 4.8 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 16:00-20:00 ในวันธรรมดา (weekday) .....	41
รูปที่ 4.9 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 20:00-06:00 ในวันธรรมดา (weekday) .....	41
รูปที่ 4.10 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 06:00-12:00 ในวันหยุดเสาร์ (weekend) .....	42
รูปที่ 4.11 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 12:00-16:00 ในวันหยุดเสาร์ (weekend) .....	42
รูปที่ 4.12 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 16:00-20:00 ในวันหยุดเสาร์ (weekend) .....	43
รูปที่ 4.13 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 20:00-06:00 ในวันหยุดเสาร์ (weekend) .....	43
รูปที่ 4.14 รูปแสดงการเดินทางระหว่างจุดคลัสเตอร์ (clusterID) ของรถแท็กซี่ .....	44
รูปที่ 4.15 รูปแสดงระยะทาง (distance) รวม ในแต่ละจุดคลัสเตอร์ (clusterID) (หน่วย : เมตร).....	45

รูปที่ 4.16 รูปแสดงเวลาเดินทาง (traveltime) รวม ในแต่ละจุดคลัสเตอร์ (clusterID) (หน่วย : นาที) ..... 46

รูปที่ 4.17 การแสดงผลการศึกษางานวิจัยผ่านไมโครซอฟท์เพาเวอร์บีไอ (Microsoft Power BI) ..... 47



## สารบัญตาราง

ตาราง	หน้า
ตารางที่ 1.1 แผนการดำเนินงานโครงการ .....	3
ตารางที่ 2.1 ตารางแสดงผลลัพธ์ของโมเดล Time-Location-Relationship (TLR) (ที่มา : Kong, et al., 2017).....	6
ตารางที่ 3.1 ตัวอย่างข้อมูลที่ใช้ในงานวิจัย.....	15
ตารางที่ 3.2 ตารางตัวอย่างข้อมูลก่อนการคัดแยก (6 คอลัมน์).....	23
ตารางที่ 3.3 ตารางตัวอย่างข้อมูลหลังคัดแยกให้มีแต่ในจังหวัดกรุงเทพมหานคร (7 คอลัมน์).....	23
ตารางที่ 3.4 ตารางตัวอย่างข้อมูลหลังคัดแยกจุดรับผู้โดยสาร (pick up points) (8 คอลัมน์) .....	24
ตารางที่ 3.5 ตารางตัวอย่างข้อมูลหลังคัดแยกจุดส่งผู้โดยสาร (drop off points) (9 คอลัมน์) .....	24
ตารางที่ 3.6 ตารางตัวอย่างข้อมูลหลังคำนวณระยะทาง (distance) (10 คอลัมน์).....	26
ตารางที่ 3.7 ตารางตัวอย่างข้อมูลหลังคำนวณเวลาเดินทาง (travel time) (11 คอลัมน์).....	26
ตารางที่ 3.8 ตารางตัวอย่างข้อมูลหลังคำนวณเงิน (price) (12 คอลัมน์) .....	28
ตารางที่ 3.9 ตารางตัวอย่างข้อมูลหลังเพิ่มคอลัมน์วัน (day) และชั่วโมง (hour) (14 คอลัมน์) .....	32

# บทที่ 1

## บทนำ

ปฏิญานี้จัดทำขึ้นเพื่อสร้างโมเดลการวิเคราะห์หาพื้นที่ที่แท็กซี่ควรจะไปรับผู้โดยสารในแต่ละช่วงเวลา เพื่ออำนวยความสะดวกแก่คนขับรถแท็กซี่ในกรุงเทพมหานครโดยมีรายละเอียดของระบบ ดังต่อไปนี้

### 1.1 ความเป็นมาของปัญหา

รูปแบบการเดินทางของประชาชนในเมืองนั้นๆจะให้ข้อมูลเชิงลึกที่เป็นประโยชน์เกี่ยวกับกลไกทางเศรษฐกิจและสังคมมากมาย และด้วยการพัฒนาทางด้านเทคโนโลยีนั้นทำให้นักวิจัยนั้นสามารถเข้าถึงข้อมูลการเดินทางจำนวนมากได้โดยง่าย แต่เนื่องจากข้อจำกัดความเป็นส่วนตัวนั้นทำให้การศึกษาส่วนใหญ่เกี่ยวกับการค้นพบรูปแบบการเดินทางนั้นจึงอิงตามข้อมูลขนส่งสาธารณะ เช่น ข้อมูลเส้นทางรถแท็กซี่, ข้อมูลธุรกรรมสมาร์ตการ์ดสำหรับรถไฟฟ้าใต้ดิน และรถประจำทาง ซึ่งในโครงการนี้จะใช้ข้อมูลการเดินทางของรถแท็กซี่

แท็กซี่เป็นหนึ่งในวิธีการเดินทางที่สำคัญที่สุดในการเดินทางของคนกรุงเทพฯ ซึ่งจะแตกต่างจากรถประจำทาง รถไฟ และวิธีการขนส่งอื่นๆที่มีเส้นทางและสถานีที่แน่นอน ตัวแท็กซี่นั้นสามารถให้บริการการเดินทางส่วนบุคคลที่สะดวกและกำหนดสถานที่ที่จะไปเองได้ เส้นทางของรถแท็กซี่นั้นเป็นลำดับของพิกัดอ้างอิงทางภูมิศาสตร์ที่ต่อเนื่องกันของรถแท็กซี่และมีการประทับเวลาที่เกี่ยวข้อง เส้นทางเดินรถของแท็กซี่นั้นจะแสดงให้เห็นถึงประสบการณ์ของคนขับรถแท็กซี่และความต้องการในการเดินทางของผู้โดยสารคนขับรถแท็กซี่นั้นจะมีความเข้าใจเชิงลึกเกี่ยวกับการจราจรภายในเมืองมักจะพิจารณาปัจจัยหลายประการก่อนทำการเลือกเส้นทาง ดังนั้นข้อมูลที่ได้จึงได้ความรู้เกี่ยวกับสถานะการจราจรและในเวลาเดียวกันข้อมูลต้นทาง-ปลายทางที่ดึงมาจากการเดินทางโดยรถแท็กซี่จำนวนมากสามารถสะท้อนถึงกฎการเดินทางของผู้อยู่อาศัยและการกระจายความต้องการในการเดินทางได้อย่างเต็มที่

โดยจุดมุ่งหมายของโครงการวิเคราะห์ข้อมูลจากการเดินทางของรถแท็กซี่ (Data Analytics on Taxi Trajectories) คือการช่วยเหลือบรรดารถแท็กซี่ที่มีการรับ-ส่งผู้โดยสารในสถานที่ๆตนเองนั้นไม่คุ้นชินทำให้การที่จะรับ-ส่งผู้โดยสารแต่ละครั้งนั้นเป็นเรื่องที่ยาก ทำให้เกิดแนวคิดของโครงการนี้ที่จะช่วยเหลือเหล่าบรรดารถแท็กซี่เหล่านั้น โดยการนำข้อมูล จีพีเอส (GPS) การรับ-ส่งผู้โดยสารของ

แท็กซีในปี ค.ศ. 2020 มาทำการคาดการณ์ว่าหากแท็กซีไปสถานที่ไหน ณ ช่วงเวลาไหนถึงจะได้รับผู้โดยสาร เพื่อให้รถแท็กซีนั้นสามารถรับส่งคนได้ต่อเนื่องตลอดทั้งวัน

## 1.2 วัตถุประสงค์ของโครงการ

- 1) แนะนำจุดที่แท็กซีควรไปรอรับผู้โดยสารตามช่วงเวลาต่างๆ
- 2) พิสูจน์สมมติฐานที่ว่าแท็กซีมีการรับผู้โดยสาร ณ จุดที่มีความหนาแน่นของผู้โดยสารมาก จะมีประสิทธิภาพในการรับผู้โดยสารต่อวันมากกว่า

## 1.3 ขอบเขตของโครงการ

- 1) นำข้อมูลทั้งหมดที่ใช้มาจากข้อมูลเปิดของมูลนิธิศูนย์ข้อมูลจราจรอัจฉริยะไทย (ITIC Open Data Archives)
- 2) ข้อมูลที่นำมาวิเคราะห์คือข้อมูลในปี 2020
- 3) ข้อมูลที่นำมาวิเคราะห์จะจำกัดขอบเขตอยู่ในกรุงเทพมหานครเท่านั้น

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ได้จุดที่เป็นสถานที่ได้รับความนิยมซึ่งมีความหนาแน่นของผู้โดยสาร
- 2) สามารถแนะนำจุดที่แท็กซีควรไปรอรับผู้โดยสารตามช่วงเวลาต่างๆได้

## 1.5 แผนการดำเนินงาน

แผนการดำเนินงาน		2565					2566			
		ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.
1	คิดหัวข้อโครงการและปรึกษาอาจารย์									
2	ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง									
3	นำเข้าข้อมูลที่ใช้ในการศึกษา									
4	เตรียมข้อมูลการเดินทางของรถแท็กซี่									
5	เตรียมข้อมูลพิกัดพื้นที่ (polygon)									
6	คัดแยกข้อมูลให้อยู่ในกรุงเทพมหานคร									
7	คัดแยกข้อมูลรับ-ส่งผู้โดยสาร									
8	คำนวณระยะทาง ค่าโดยสารและเวลาเดินทาง									
9	คัดแยกข้อมูลด้วยดีบีสแกน (DBSCAN)									
10	สรุปข้อมูลหลังการคัดแยกข้อมูล									
11	หาความสัมพันธ์ระหว่างความถี่ในการรับ-ส่งผู้โดยสารกับเวลาเดินทาง									
12	แนะนำจุดรับผู้โดยสารตามช่วงเวลา									
13	แสดงผลข้อมูลด้วย Power BI									
14	จัดทำโครงร่างงานวิจัย									

ตารางที่ 1.1 แผนการดำเนินงานโครงการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 2

### ทฤษฎีที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงรายละเอียดของข้อมูลที่ใช้ในโครงการนี้ เช่น รายละเอียดหรือความหมายของข้อมูล รายละเอียดของงานวิจัยที่นำมาใช้อ้างอิงและเปรียบเทียบผลการทดลอง รวมถึงทฤษฎีและเครื่องมือที่คณะผู้จัดทำได้ศึกษาค้นคว้าเพื่อนำมาประกอบ โครงการนี้ ซึ่งสามารถดูรายละเอียดได้ดังนี้

#### 2.1 งานวิจัยที่เกี่ยวข้อง

หัวข้อนี้ จะกล่าวถึงงานวิจัยที่ได้ทำการศึกษาและนำมาอ้างอิงพัฒนาต่อ ดังนี้

##### 2.1.1 Time-Location-Relationship Combined Service Recommendation Based on Taxi Trajectory Data

ในงานวิจัยนี้ได้มีการเสนอคำแนะนำบริการแท็กซี่โดยใช้แบบจำลองความสัมพันธ์เชิงเวลากับตำแหน่ง (Time-Location-Relationship: TLR) เพื่อช่วยจัดการกับปัญหาการที่แท็กซี่นั้น ไม่มีผู้โดยสารและความยากลำบากในการหารถแท็กซี่สำหรับผู้โดยสาร ด้วยการบูรณาการ ประมวลผล และวิเคราะห์ข้อมูลเส้นทางเดินรถของแท็กซี่ จะสามารถระบุความสัมพันธ์เชิงปริมาณระหว่างการขึ้นและลงของผู้โดยสารในเขตพื้นที่เมืองได้และสามารถรับรูปแบบการเคลื่อนย้ายของผู้คนได้อย่างมีประสิทธิภาพ เช่น สามารถอนุมานได้ว่าคนส่วนใหญ่มักใช้เวลาประมาณ 3 ชั่วโมงในการเยี่ยมชมสถานที่ทางประวัติศาสตร์ ในช่วงเวลา 11.00 - 14.00 นาฬิกา

ในส่วนของการเริ่มต้นนั้น จะมีการใช้รูปแบบการเคลื่อนที่ในแบบจำลองความสัมพันธ์เชิงเวลากับตำแหน่ง (Time-Location-Relationship: TLR) ซึ่งพิจารณาคุณสมบัติสามมิติ (เวลา ตำแหน่ง และความสัมพันธ์) ของการเปลี่ยนแปลงของเมือง จากนั้นจะสามารถคาดการณ์จำนวนผู้โดยสารที่จะขึ้นแท็กซี่ด้วยความช่วยเหลือจากจำนวนผู้โดยสารที่ขึ้นรถแท็กซี่ในพื้นที่ใช้งาน โดยจะใช้แบบจำลองความสัมพันธ์เชิงเวลากับตำแหน่ง (Time-Location-Relationship: TLR) ที่แนะนำไปต่อจากนั้น จะคำนวณระยะทางการเดินทางเฉลี่ยและเวลาเดินทางเฉลี่ยระหว่างพื้นที่ทำงานสองแห่ง นอกจากนี้ยังมีการแนะนำพื้นที่ยอดนิยม N ตำแหน่ง (Top-N) สำหรับผู้ขับขี่ตามรูปแบบที่เสนอ ซึ่งช่วย

ให้ผู้ขับขี่สามารถรับรู้โดยสารได้มากขึ้นเพื่อประหยัดพลังงาน และลดแรงกดดันในการขนส่งในเมือง โดยแบบจำลองความสัมพันธ์เชิงเวลากับตำแหน่ง (Time-Location-Relationship: TLR) นั้นจะใช้ความสัมพันธ์เชิงปริมาณ (Quantitative Relationship) ระหว่างการขึ้นและลงของแท็กซี่ของผู้โดยสารเป็นหลัก ในขณะที่รุ่นดั้งเดิมใช้เฉพาะจำนวนผู้โดยสารที่เพิ่มขึ้นในภูมิภาคการทำงานต่างๆ ซึ่งมีการขยายการทดลองต้นแบบและมีการใช้เกณฑ์การประเมินเพิ่มเติมเพื่อตรวจสอบประสิทธิภาพของแบบจำลองความสัมพันธ์เชิงเวลากับตำแหน่ง (Time-Location-Relationship: TLR) เมื่อเทียบกับแบบจำลองการวิเคราะห์และจัดการกับข้อมูลชุดชั้นเวลา (Autoregressive Integrated Moving Average : ARIMA), แบบจำลองของโครงข่ายประสาทเทียม (Backpropagation Neural Network :BPNN), ซัพพอร์ตเวกเตอร์แมทชีน (Support Vector Machine : SVM) และแบบจำลองทางสถิติที่ใช้ในการทำนายหรือจัดกลุ่มข้อมูล (Gradient Boosting Decision Tree Model : GBDT) ซึ่งนำไปสู่การปรับปรุงความแม่นยำในการทำนาย ซึ่งจะได้ผล ดังนี้

- 1) นำเสนอแบบจำลองความสัมพันธ์เชิงเวลากับตำแหน่ง (Time-Location-Relationship: TLR) ซึ่งสามารถระบุคุณสมบัติสามมิติของเมืองแบบไดนามิกเพื่อคาดการณ์การกระจายตัวของผู้โดยสารสำหรับภูมิภาคและหน้าที่ทางสังคมต่างๆ
- 2) ประเมินและเปรียบเทียบประสิทธิภาพของแบบจำลองที่นำเสนอ แบบจำลองความสัมพันธ์เชิงเวลากับตำแหน่ง (Time-Location-Relationship: TLR), แบบจำลองการวิเคราะห์และจัดการกับข้อมูลชุดชั้นเวลา (Autoregressive Integrated Moving Average : ARIMA), แบบจำลองของโครงข่ายประสาทเทียม (Backpropagation Neural Network :BPNN), ซัพพอร์ตเวกเตอร์แมทชีน (Support Vector Machine : SVM) และแบบจำลองทางสถิติที่ใช้ในการทำนายหรือจัดกลุ่มข้อมูล (Gradient Boosting Decision Tree Model : GBDT) โดยใช้เมตริกต่อไปนี้อย่างน้อยถึงค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient : CC), ค่ารากที่สองของผลเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMSE), ข้อผิดพลาดสัมบูรณ์เฉลี่ย (Mean Absolute Error : MAE) และค่าเชิงปริมาณที่ใช้วัดประสิทธิภาพของโมเดลทางสถิติ (Normalized Mean Absolute Error : NMAE)
- 3) ซึ่งจะสามารถกำหนดประสิทธิภาพและความเสถียรของแบบจำลองที่นำเสนอได้
- 4) มีการแนะนำพื้นที่ยอดนิยม N ตำแหน่ง (Top-N) ให้กับผู้ใช้โดยพิจารณาจากผลการคาดคะเน ระยะทางเฉลี่ยในการเดินทาง และเวลาเดินทางโดยเฉลี่ย จากนั้นพวกเขาสามารถ

ตัดสินใจได้ว่าจะรับผู้โดยสารที่ไหนเพื่อเพิ่มผลกำไรสูงสุด โดยแบบจำลองความสัมพันธ์เชิงเวลากับตำแหน่ง (Time-Location-Relationship: TLR) ที่นำเสนอ นั้นมีผลลัพธ์ความแม่นยำในการทำนาย 90.9% ในวันธรรมดาและ 80.4% ในวันหยุดสุดสัปดาห์ตามลำดับตามตารางที่ 2.1 ซึ่งหมายความว่าแบบจำลองที่นำเสนอมีความสัมพันธ์ที่มีประสิทธิภาพดี

Dataset#	Time	ARIMA	BPNN	SVM	GBDT	Proposed
Dataset1&2	weekdays	0.7680	0.8332	0.8910	0.9065	0.9087
Dataset3&4	weekends	0.5635	0.6958	0.7498	0.7575	0.8041

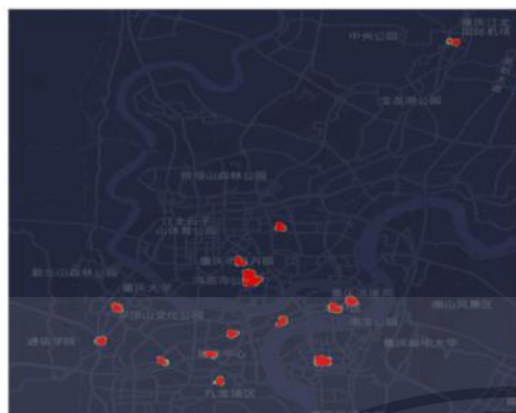
ตารางที่ 2.1 ตารางแสดงผลลัพธ์ของโมเดล Time-Location-Relationship (TLR)

(ที่มา : Kong, et al., 2017)

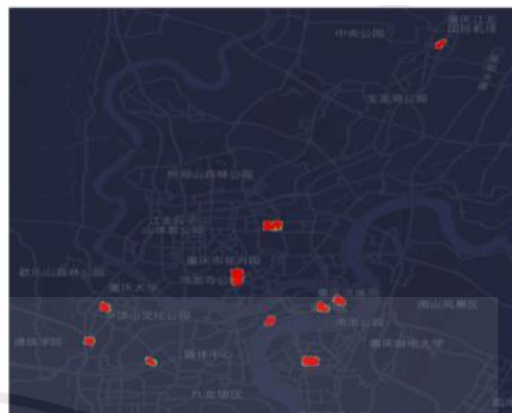
### 2.1.2 Mining trip attractive areas using large-scale taxi trajectory data

เนื้อหาการวิจัยหลักของบทความนี้ประกอบด้วย 3 ด้าน ได้แก่ การประมวลผลข้อมูลสำหรับข้อมูลเส้นทางรถแท็กซี่ อัลกอริทึมการจัดกลุ่มสำหรับพื้นที่ที่น่าสนใจในการเดินทาง และการทดลองที่เกี่ยวข้อง ซึ่งส่วนหลักของงานวิจัยนี้คือการเสนออัลกอริทึม (Algorithm) กริดบีสแกน (GRIDBSCAN) ที่เหมาะสำหรับข้อมูลดิบขนาดใหญ่ ซึ่งจะให้ผลออกมาคล้ายกับตัว ดีบีสแกน (DBSCAN) แต่จะมีระยะเวลาในการคำนวณนั้นน้อยกว่ามาก

ซึ่งการค้นพบพื้นที่ที่น่าสนใจสำหรับการเดินทางภายในเมืองนั้นมีความสำคัญอย่างยิ่งต่อการสร้างระบบการจราจร ซึ่งพื้นที่ที่น่าสนใจนั้นมักจะเป็นพื้นที่ที่มีการจราจรหนาแน่น จากการศึกษาจะได้ว่าพื้นที่ที่น่าสนใจในตอนเช้าและตอนบ่ายนั้นมีน้อยและค่อนข้างรวมกันดังรูป 1.1(a) – 1.1(d) ในขณะที่ตอนกลางคืนนั้นมีพื้นที่ที่น่าสนใจมากและมีความกระจัดกระจาย ดังรูป 1.1(e) – 1.1(f) ซึ่งสอดคล้องกับการใช้ชีวิตประจำวันของคนในช่วงกลางวันนั้นส่วนมากจะเป็นช่วงเวลาของการทำงาน ทำให้พื้นที่ส่วนใหญ่จะอยู่บริเวณย่านธุรกิจ และในส่วนของตอนกลางคืนนั้นจะมีกิจกรรมต่างๆเพิ่มมากขึ้นและเป็นเวลาที่ผู้คนเลิกงานและหากิจกรรมอะไรทำเพื่อผ่อนคลายจึงทำให้มีการกระจายตัวสูง



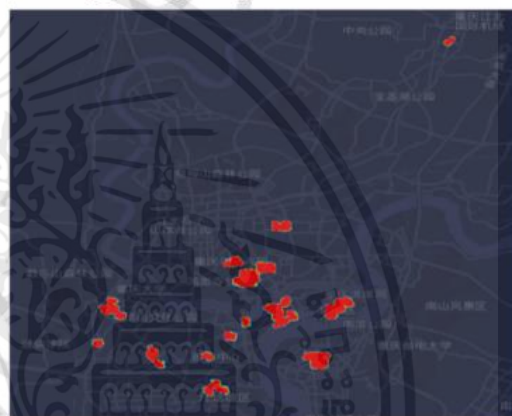
(a) Morning passenger pick-up attractive area distribution



(b) Morning passenger drop-off attractive area distribution



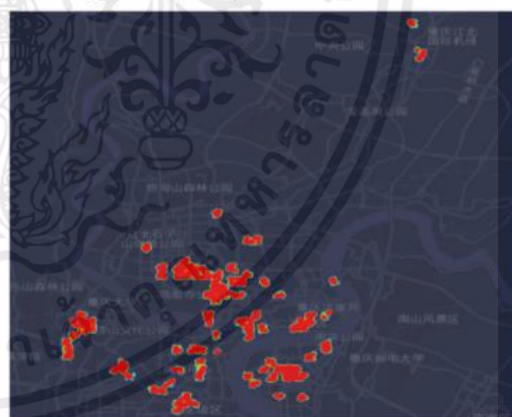
(c) Midday passenger pick-up attractive area distribution



(d) Midday passenger drop-off attractive area distribution



(e) Night passenger pick-up attractive area distribution.



(f) Night passenger drop-off attractive area distribution

### รูปที่ 1.1 การกระจายตัวของพื้นที่ที่สนใจ

(ที่มา : Zheng, et al., 2017)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 2.1.3 A Big Data Analysis on Efficiency of Bangkok Taxi System

งานวิจัยนี้วิเคราะห์ประสิทธิภาพของระบบแท็กซี่กรุงเทพโดยใช้ข้อมูลการติดตามแท็กซี่จริง ข้อมูลที่ใช้ในงานวิจัยนี้เป็นชุดข้อมูลที่จัดทำโดยศูนย์ข้อมูลจราจรอัจฉริยะไทย (ITIC) ประกอบด้วยข้อมูล GPS ของแท็กซี่ในกรุงเทพมหานคร มากกว่า 114 ล้านรายการ ตลอดเดือนธันวาคม 2562 รวมถึงตำแหน่งของรถแท็กซี่แต่ละคันที่รวบรวมโดยประมาณทุกๆ 1-3 นาที ผลการศึกษาพบว่าระบบทำงานในระดับประสิทธิภาพต่ำ โดยมีแท็กซี่ว่างจำนวนมากบนท้องถนนและผู้โดยสารรอเป็นเวลานาน

กระบวนการการสกัดคุณลักษณะ (Feature Extraction) ของงานวิจัยนี้ ได้ออกมาเป็น 5 คุณลักษณะ ประกอบด้วยเส้นทางการเดินทางของรถแท็กซี่ (Trip) ระยะทางการเดินทาง (Trip distance) ระยะเวลาในการเดินทาง (Trip duration) จุดรับและส่งผู้โดยสาร (Pickup and Drop-off Locations) และราคาค่าโดยสาร (Trip Fare)

สิ่งที่ได้จากงานวิจัยนี้ คือ

- การวัดประสิทธิภาพโดยรวมของระบบรถแท็กซี่ในกรุงเทพฯ โดยพิจารณาจากข้อมูลตำแหน่งรถแท็กซี่ในเดือนธันวาคม 2562
- แสดงให้เห็นว่าประสิทธิภาพโดยรวมของระบบแท็กซี่กรุงเทพฯ อยู่ที่ประมาณ 47.25% เท่านั้น
- การตรวจสอบประสิทธิภาพระดับตัวแทนของระบบแท็กซี่กรุงเทพฯ.
- ให้ข้อมูลเชิงลึกเกี่ยวกับการกระจายอัตราเพิ่มรายได้ในระบบแท็กซี่กรุงเทพฯ

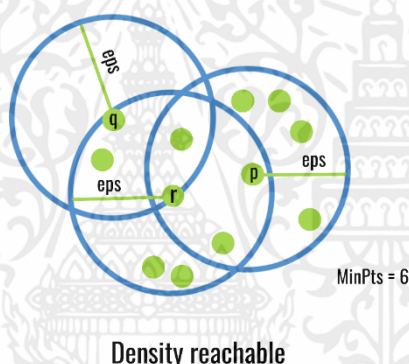
## 2.2 เทคนิคที่นำมาใช้ในงานวิจัย

ในหัวข้อนี้จะเป็นรายละเอียดของทฤษฎีที่เกี่ยวข้องและสูตรทางคณิตศาสตร์ที่มีการนำมาใช้ โดยมีรายละเอียดตามหัวข้อย่อต่อไปนี้ ดังนี้

### 2.2.1 ดีบีสแกน (DBSCAN)

การจัดกลุ่มเชิงพื้นที่ที่ตามความหนาแน่นของข้อมูลพร้อมด้วยข้อมูลรบกวน (Density-Based Spatial Clustering of Applications with Noise : DBSCAN) เป็นอัลกอริทึม (Algorithm) ในการจัดกลุ่มข้อมูล (Clustering) โดยใช้การจัดกลุ่มจุดของข้อมูลที่มีความหนาแน่นและอยู่ใกล้กันอยู่ด้วยกัน พร้อมด้วยการแยกจุดที่มีความหนาแน่นน้อยออกไป

ดีบีสแกน (DBSCAN) ใช้พารามิเตอร์หลักสองตัว  $\epsilon$  และ MinPts



รูปที่ 2.1 ดีบีสแกน (DBSCAN)

(ที่มา : geeksforgeeks, 2019)

- 1)  $\epsilon$  คือ ระยะทางที่กำหนดรัศมีของพื้นที่ใกล้เคียงรอบๆ จุดข้อมูลแต่ละจุด
  - 2) MinPts คือ จำนวนจุดข้อมูล (Data Points) ขั้นต่ำสำหรับการสร้างคลัสเตอร์
- ดีบีสแกน (DBSCAN) จะระบุจุดสามประเภทในชุดข้อมูล ดังนี้
- 1) จุดหลัก (core points) คือ จุดที่มีจำนวนจุดข้อมูล (Data Points) อย่างน้อยเท่ากับ MinPts และอยู่ภายในรัศมี  $\epsilon$  ซึ่งหมายความว่า เป็นพื้นที่ที่ผู้คนอยู่เรื่อย ๆ และเป็นพื้นที่ที่คนมักจะไปด้วยกัน
  - 2) จุดขอบ (border points) คือ จุดที่อยู่ภายในรัศมี  $\epsilon$  ของจุดหลัก (core points) แต่มีจำนวนข้อมูลน้อยกว่า MinPts ภายในรัศมีเดียวกัน

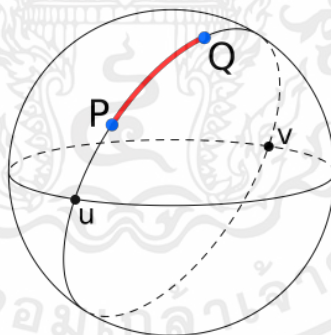
3) จุดรบกวน (noise points) คือ จุดที่ไม่ใช่จุดหลัก (core points) หรือจุดขอบ (border points)

กระบวนการจัดกลุ่มใน DBSCAN เริ่มต้นด้วยการเลือกจุดข้อมูลและระบุพื้นที่ใกล้เคียงตามค่า  $\epsilon$  หากจำนวนจุดข้อมูลในละแวกนี้มากกว่าหรือเท่ากับค่า  $\text{MinPts}$  ก็จะถือว่าเป็นจุดหลัก (core points) และจะมีการสร้างคลัสเตอร์ใหม่ จากนั้นอัลกอริทึมจะขยายคลัสเตอร์โดยเพิ่มจุดหลัก (core points) ที่อยู่ใกล้เคียงทั้งหมดและจุดข้อมูลที่เกี่ยวข้องภายในระยะ  $\epsilon$  และจะทำกระบวนการนี้ต่อไปจนกว่าจะไม่มีจุดหลัก (core points) ในพื้นที่ใกล้เคียงของคลัสเตอร์

หากจุดข้อมูลอยู่ภายในระยะ  $\epsilon$  ของจุดหลัก (core points) แต่ไม่มีจุดข้อมูลเพียงพอภายในระยะ  $\epsilon$  ของตัวเองที่จะเป็นจุดหลัก (core points) จุดนั้นจะถูกจัดประเภทเป็นจุดขอบ (border points) และเพิ่มไปยังคลัสเตอร์ มิฉะนั้นจะจัดเป็นจุดรบกวน (noise points)

### 2.2.2 สูตรของฮาเวอร์ไซน์ (Haversine Formula)

สูตรของฮาเวอร์ไซน์ (Haversine Formula) เป็นสูตรการคำนวณเพื่อใช้หาระยะทางที่สั้นที่สุดระหว่างจุดสองจุดบนพื้นผิวทรงกลมหรือพื้นผิวของโลก ซึ่งคำนวณผ่านค่าละติจูด (latitude) และลองจิจูด (longitude) ของจุดสองจุดที่แตกต่างกัน



รูปที่ 2.2 กฎของฮาเวอร์ไซน์ (Haversine Formula)

(ที่มา : Wikipedia, 2022)

### 2.2.3 แพนดาส (Pandas)

แพนดาส (Pandas) เป็นไลบรารี (Library) ในไพทอน (Python) ที่มีความนิยมในการนำมาใช้เกี่ยวกับ วิทยาศาสตร์ข้อมูล (Data Science) การวิเคราะห์ข้อมูล (Data Analysis) และงานเกี่ยวกับการเรียนรู้ของเครื่อง (Machine Learning) เป็นเครื่องมือที่เหมาะสมในการจัดการกับข้อมูล เช่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทำความสะอาดข้อมูล (Data Cleaning) และใช้งานง่าย แสดงผลข้อมูลออกมาในรูปแบบตาราง (Data Frame) ซึ่งสามารถนำไปใช้ประโยชน์กับแพ็คเกจ (Package) อื่นๆ ได้มากมาย

#### 2.2.4 ไลบรารี (Scikit-learn)

ไลบรารี (Scikit-learn) เป็นไลบรารี (Library) ในไพธอน (Python) ที่มีประโยชน์และโดดเด่นในด้านการเรียนรู้ของเครื่อง (Machine Learning) ประกอบไปด้วยเครื่องมือมากมายในการสร้างแบบจำลองทางสถิติ (Statistical Modeling) ทั้ง การจำแนกข้อมูล (Classification) การวิเคราะห์การถดถอย (Regression analysis) และการจัดกลุ่มข้อมูล (Clustering)

#### 2.2.5 umpy (NumPy)

คือ โมดูลของภาษาไพธอน (Python) ที่เกี่ยวกับการคำนวณทางคณิตศาสตร์และวิทยาศาสตร์ข้อมูล (scientific computing) โดยเฉพาะการคำนวณด้านอาร์เรย์หรือเมทริกซ์ (array/matrix) ซึ่งเป็นโครงสร้างข้อมูลที่สำคัญในการแก้ปัญหาในงานวิทยาศาสตร์ข้อมูลและเทคโนโลยีต่างๆ umpy (NumPy) มีความสามารถในการจัดการข้อมูลอาร์เรย์หรือเมทริกซ์อย่างมีประสิทธิภาพ และมีฟังก์ชันทางคณิตศาสตร์และวิทยาศาสตร์ข้อมูลที่เขียนเป็นฟังก์ชันคอมพิวเตอร์ (computer functions) ที่ช่วยให้ผู้ใช้สามารถทำงานร่วมกับอาร์เรย์หรือเมทริกซ์ได้โดยเร็วและง่ายต่อการใช้งาน นอกจากนี้ umpy (NumPy) ยังสามารถใช้งานร่วมกับโมดูลอื่นๆ เช่น แพนดาส (Pandas), แมทพล็อตลิบ (Matplotlib), ไลบรารี (Scikit-learn) เพื่อให้การคำนวณและการแสดงผลข้อมูลทางคณิตศาสตร์และวิทยาศาสตร์ข้อมูลเป็นไปได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

#### 2.2.6 แมทพล็อตลิบ (Matplotlib)

เป็นไลบรารี (Library) ในไพธอน (Python) ที่ใช้สำหรับการสร้างกราฟและแผนภูมิต่างๆ โดย Matplotlib เป็น ไลบรารี (Library) ที่นิยมใช้กันอย่างแพร่หลายในการทำงานวิเคราะห์ข้อมูล และวิจัยต่างๆ เช่น การทำนายผลของการวิเคราะห์ข้อมูล การเปรียบเทียบข้อมูลระหว่างกลุ่มต่างๆ หรือการสร้างกราฟเพื่อแสดงผลข้อมูลที่สื่อความหมายได้อย่างชัดเจน โดยแมทพล็อตลิบ (Matplotlib) สามารถสร้างกราฟและแผนภูมิที่หลากหลายได้ เช่น กราฟเส้น (Line plot) กราฟแท่ง (Bar plot) กราฟเส้นแท่ง

รวม (Stacked bar plot) กราฟเส้นแท่งแบ่ง (Grouped bar plot) ฮิสโตแกรม (Histogram) และอื่น ๆ อีกมากมาย

### 2.2.7 โมดูลโอเอส (Operating System)

คือ โมดูลที่ใช้ในภาษาไพธอน (Python) เพื่อจัดการระบบปฏิบัติการ (Operating System) โดยใช้งานได้หลายฟังก์ชัน เช่น สร้าง/ลบไดเรกทอรี อ่าน/เขียนไฟล์ จัดการกับโปรเซส (process) และทำงานร่วมกับโมดูลอื่นๆ เช่น ชูติล(shutil) เพื่อคัดลอกและย้ายไฟล์ หรือ กลอบ (glob) เพื่อค้นหาไฟล์ในโฟลเดอร์ การใช้งาน โอเอส (os) ช่วยให้โปรแกรมสามารถทำงานได้สอดคล้องกับระบบปฏิบัติการที่ใช้งานอยู่บนเครื่องของผู้ใช้ และทำให้ง่ายต่อการพัฒนาโปรแกรมที่เป็นพลังงาน และมีประสิทธิภาพมากยิ่งขึ้นด้วย

### 2.2.8 ไลบรารีโกลบ (Glob)

เป็นฟังก์ชันในไพธอน (Python) ที่ใช้สำหรับค้นหาพาธเนมส์ (pathnames) ในรูปแบบของสตริง (string) ที่ตรงกับแพทเทิร์น (pattern) ที่กำหนด โดยแพทเทิร์น (pattern) จะใช้ไวลด์การ์ด (wildcard) เช่น \* หรือ ? เพื่อแทนที่ตัวอักษรหรือตัวเลขในชื่อไฟล์หรือพาธ (path) ที่ต้องการค้นหา ซึ่งช่วยให้สามารถค้นหาไฟล์หลายไฟล์ที่มีลักษณะคล้ายกันได้ง่ายขึ้น และช่วยในการทำงานกับไฟล์หลายไฟล์ได้อย่างมีประสิทธิภาพ

### 2.2.9 เชปลี่ (Shapely)

เป็นไลบรารี (Library) ในไพธอน (Python) ที่ใช้สำหรับจัดการและวิเคราะห์วัตถุทางเรขาคณิตในระบบพิกัดคาร์ทีเซียน (Cartesian coordinate system) โดยมีชุดเครื่องมือที่ทรงพลังและยืดหยุ่นสำหรับการทำงาน

คุณสมบัติหลักของเชปลี่ (Shapely) ที่นำมาใช้ คือ ความสามารถในการทำงานกับวัตถุทางเรขาคณิตประเภทต่างๆ รวมถึง จุด (Point) ข้อมูลพิกัดพื้นที่ (Polygon) มัลติพอยท์ (MultiPoint) และ มัลติโพลิโกน (MultiPolygon)

### 2.2.10 ไมโครซอฟต์เพาเวอร์บีไอ (Microsoft Power BI)

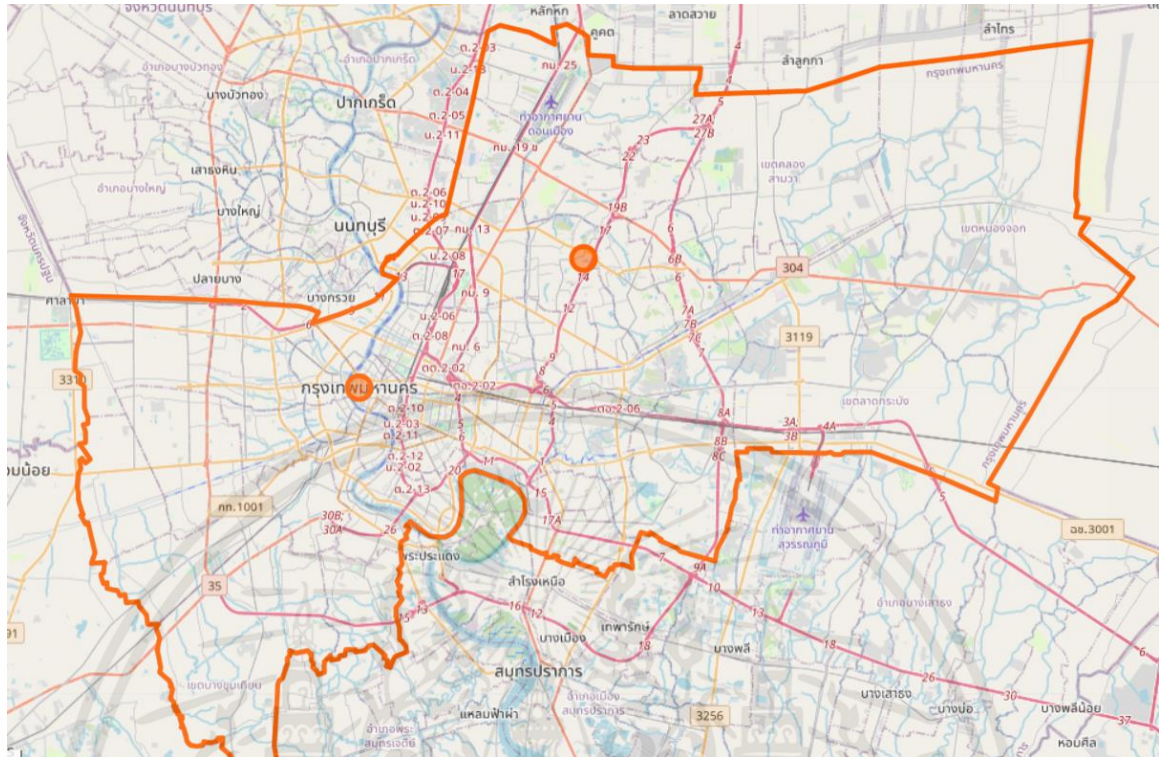
ไมโครซอฟต์เพาเวอร์บีไอ (Microsoft Power BI) เป็นบริการวิเคราะห์ธุรกิจที่ใช้งานบนคลาวด์ ช่วยให้ผู้ใช้วิเคราะห์และแสดงผลข้อมูลจากแหล่งต่าง ๆ ได้ โดยเพาเวอร์บีไอ (Power BI) จะมีการนำเสนอแดชบอร์ด (dashboard) และรายงานที่สามารถใช้งานได้อย่างสื่อสารและมีประสิทธิภาพสูง สามารถแชร์ได้ทั้งภายในองค์กร

ผู้ใช้สามารถเชื่อมต่อกับแหล่งข้อมูลต่าง ๆ เช่น สเปรดชีต Excel, ฐานข้อมูลหรือเว็บเซอร์วิส และใช้เครื่องมือของเพาเวอร์บีไอ (Power BI) เพื่อแปลงและทำความสะอาดข้อมูลก่อนสร้างการแสดงผล รวมถึงรองรับการแสดงผลแบบต่าง ๆ เช่น แผนภูมิ ตาราง แผนที่ และการแสดงผลที่กำหนดเองโดยนักพัฒนา

นอกจากนี้เพาเวอร์บีไอ (Power BI) ยังสามารถทำงานร่วมกัน เช่น การแชร์รายงานและแดชบอร์ด (dashboard) กับเพื่อนร่วมงาน และการทำงานร่วมกันในโปรเจกต์วิเคราะห์ข้อมูลแบบเรียลไทม์ (real-time) สามารถเข้าถึงเพาเวอร์บีไอ (Power BI) ผ่านเว็บเบราว์เซอร์ หรือผ่านแอปพลิเคชันบนอุปกรณ์ iOS และ Android ได้

### 2.2.11 ข้อมูลพิกัดพื้นที่ (Polygon)

เป็นรูปหลายเหลี่ยมที่ใช้ในการแสดงพื้นที่หรือพิกัดต่าง ๆ บนแผนที่ โดยที่แต่ละด้านของข้อมูลพิกัดพื้นที่ (Polygon) จะมีค่าพิกัด (latitude, longitude) ที่สอดคล้องกับเส้นของรูปหลายเหลี่ยมนั้น ๆ โดยรูปข้อมูลพิกัดพื้นที่ (Polygon) นี้จะใช้ในการบอกขอบเขตของพื้นที่หรือขอบเขตที่สนใจ เช่น แสดงพื้นที่ที่มีอาณาเขตในการจัดส่งสินค้า หรือแสดงพื้นที่ที่เกิดเหตุการณ์น้ำท่วม และอื่น ๆ ซึ่งในการศึกษานี้ จะนำมาใช้ในการแบ่งข้อมูลให้อยู่ในอาณาเขตของกรุงเทพมหานคร ตัวอย่างรูปข้อมูลพิกัดพื้นที่ (Polygon) ดังรูปที่ 2.3



รูปที่ 2.3 รูปข้อมูลพิกัดพื้นที่ (Polygon) ของกรุงเทพมหานคร  
(ที่มา : [openstreetmap.org](https://openstreetmap.org), 2023)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 3

### การออกแบบและการพัฒนา

ในบทนี้จะกล่าวถึงการออกแบบและการพัฒนาโครงการ ซึ่งจะมีการอธิบายและมีภาพประกอบเพื่อให้ผู้อ่านเข้าใจได้มากขึ้น

#### 3.1 รายละเอียดของข้อมูลที่ใช้

ข้อมูลที่นำมาใช้ในโครงการมี ดังนี้

##### 3.1.1 ข้อมูลการเดินทางของรถแท็กซี่ในประเทศไทย

VehicleID	gpsvalid	lat	lon	timestamp	speed	heading	for_hire	light	engine_acc
RdZ+aLuit7HgakstnAm8wlCD3ll	1	13.82811	100.63003	1/1/2020 0:00	30	188	1	1	1
Z/EHO1TlpBQhRPkBWuppPrKXXwg	1	13.74802	100.54831	1/1/2020 0:00	0	12	1	1	1
C5PcZWjRll+RjvHicnZUYVWFaQw	1	13.7924	100.52455	1/1/2020 0:00	19	291	0	1	1
rwunHqTWzTc+VnvKNwA/NFXZDql	1	13.74041	100.61557	1/1/2020 0:00	0	315	0	1	1
N8Ugs7SfidT3Ufa/fQUYEV3HtPA	1	13.97559	100.62237	1/1/2020 0:00	0	251	0	0	0
D/LVUIrv6DH83WfnutAVohCy0c	1	13.79365	100.56549	1/1/2020 0:00	0	354	0	1	1
Aapcp+loB/8AEFeCBxwwIVSyMZl	1	13.74886	100.56496	1/1/2020 0:00	47	97	0	1	1
eMZ9uY1qhh7+Qf335E9viqzeCHw	1	13.56443	100.61017	1/1/2020 0:00	0	63	1	0	0
ytp6fM8b17wVCFjgbdCSn7zGH2l	1	13.82148	100.72432	1/1/2020 0:00	0	267	0	0	0
WsQDFMebgg4ZcbRi8Uc4NAUwvOk	1	13.71986	100.33091	1/1/2020 0:00	0	350	0	0	0

ตารางที่ 3.1 ตัวอย่างข้อมูลที่ใช้ในงานวิจัย

ข้อมูลจากตารางที่ 3.1 มีการจัดเก็บอยู่ในรูปแบบของ TGZ (.tar.gz) โดยจะมีรายละเอียดของข้อมูล ดังนี้

- 1) คอลัมน์ VehicleID หมายถึง ข้อมูลรหัสประจำตัวของรถแท็กซี่แต่ละคัน
- 2) คอลัมน์ gpsvalid หมายถึง ข้อมูลแสดงสถานะของสัญญาณดาวเทียม
  - 1 = มีสัญญาณดาวเทียมเพียงพอ
  - 0 = มีสัญญาณดาวเทียมไม่เพียงพอ
- 3) คอลัมน์ lat หมายถึง ค่าพิกัดละติจูด (latitude) ในรูปแบบทศนิยม 5 ตำแหน่ง
- 4) คอลัมน์ lon หมายถึง ค่าพิกัดลองจิจูด (longitude) ในรูปแบบทศนิยม 5 ตำแหน่ง

- 5) คอลัมน์ timestamp หมายถึง วันที่และเวลาที่เก็บข้อมูลอยู่ในรูปแบบ ปี-เดือน-วันที่ ชั่วโมง:นาที:วินาที
- 6) คอลัมน์ speed หมายถึง ความเร็วของรถแท็กซี่ในหน่วย กิโลเมตร/ชั่วโมง
- 7) คอลัมน์ heading หมายถึง ทิศการเดินทางของรถยนต์ (0-360 องศา) โดยมีทิศเหนืออยู่ที่ 0 องศา
- 8) คอลัมน์ for\_hire\_light หมายถึง ค่าของไฟสัญญาณแสดงสถานะผู้โดยสาร  
1 = ไฟติด หมายถึง คาดว่าไม่มีผู้โดยสาร  
0 = ไฟดับ หมายถึง คาดว่ามีผู้โดยสาร
- 9) คอลัมน์ engine\_acc หมายถึง สถานะแสดงการทำงานของเครื่องยนต์  
1 = เครื่องยนต์มีการทำงาน จะมีการเก็บข้อมูลทุกๆ 1 นาที  
0 = เครื่องยนต์ไม่มีการทำงาน จะมีการเก็บข้อมูลทุกๆ 3 นาที

### 3.1.2 ข้อมูลพิกัดพื้นที่ (polygon) ของกรุงเทพมหานคร

ลักษณะทางภูมิศาสตร์ที่มีพื้นที่เดียวกันจะถูกล้อมรอบด้วยเส้นเพื่อแสดงขอบเขต ตัวอย่างข้อมูลที่เป็นพื้นที่ โดยในการศึกษาครั้งนี้ จะมีการกำหนดขอบเขตของการศึกษาเป็นข้อมูลที่อยู่ในเขตกรุงเทพมหานคร ซึ่งสามารถกำหนดขอบเขตได้จากการนำข้อมูลพิกัดพื้นที่ (polygon) มากำหนดได้

#### 3.1.2.1 ข้อมูลค่าพิกัดพื้นที่ (polygon)

จะเป็นข้อมูลค่าพิกัดละติจูด (latitude) และ ลองจิจูด (longitude) จำนวน 189 คู่พิกัด ซึ่งสามารถนำมาเขียนแผนที่รวมกันเป็นรูปพื้นที่ของกรุงเทพมหานครได้ ตัวอย่างดังรูปที่ 3.1

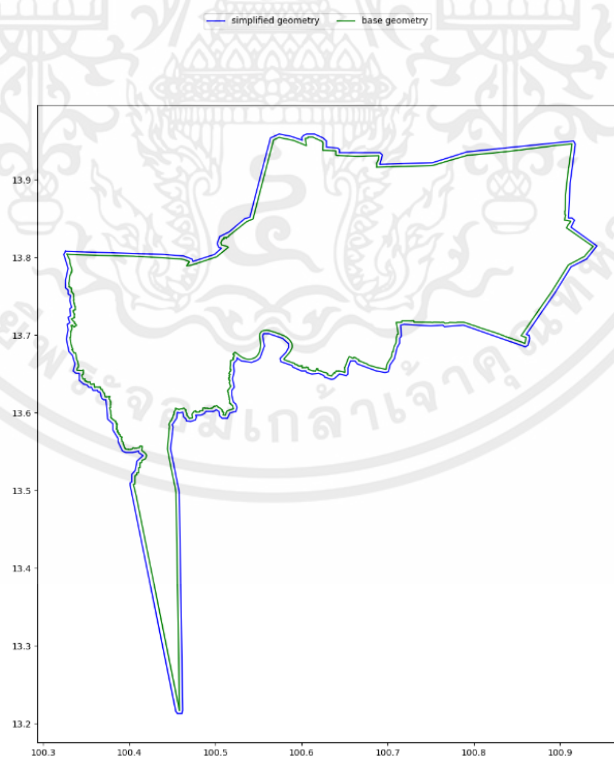
100.326	13.808
100.439	13.804
100.463	13.802
100.473	13.798
100.474	13.795
100.499	13.804
100.507	13.812
100.504	13.813
100.502	13.818
100.505	13.827
100.516	13.833
100.534	13.85

รูปที่ 3.1 ตัวอย่างข้อมูลพิกัดพื้นที่ (polygon)

(ที่มา : [polygons.openstreetmap.fr](https://polygons.openstreetmap.fr), 2022)

### 3.1.2.2 รูปแผนที่ของกรุงเทพมหานครที่ได้จากค่าพิกัดพื้นที่ (polygon)

สามารถนำค่าพิกัดพื้นที่ (polygon) มาวาดเป็นแผนที่ได้ ดังรูปที่ 3.2



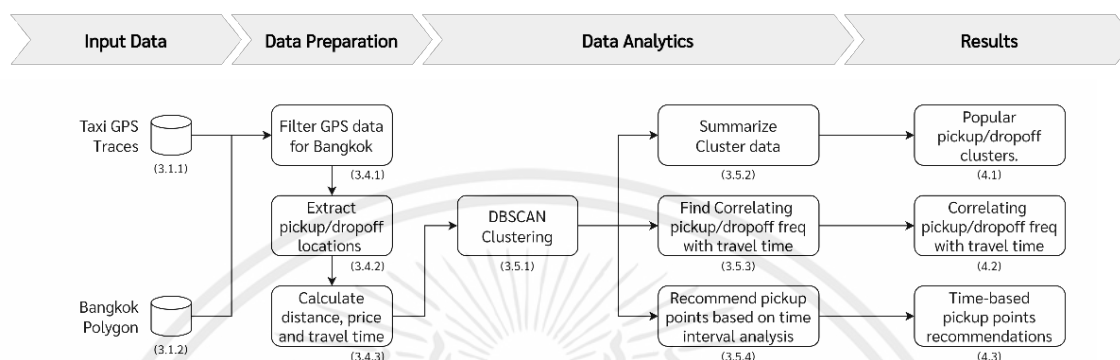
รูปที่ 3.2 แผนที่กรุงเทพมหานครจากข้อมูลพิกัดพื้นที่ (polygon)

(ที่มา : [polygons.openstreetmap.fr](https://polygons.openstreetmap.fr), 2022)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 3.2 ขั้นตอนการพัฒนา

ส่วนของการอธิบายถึงขั้นตอนการพัฒนา ตั้งแต่ข้อมูล การเตรียมข้อมูล การวิเคราะห์ข้อมูล และผลลัพธ์ที่คาดว่าจะได้จากการวิเคราะห์ข้อมูล ดังรูปที่ 3.3



รูปที่ 3.3 ภาพรวมการพัฒนางานวิจัย

### 3.2.1 ข้อมูลขาเข้า (Input Data)

#### 3.2.1.1 ข้อมูลการเดินทางรถแท็กซี่ (Taxi GPS Traces)

ข้อมูลการเดินทางของรถแท็กซี่ในประเทศไทย ปี 2020

#### 3.2.1.2 ข้อมูลพิกัดพื้นที่กรุงเทพมหานคร (Bangkok Polygon)

ข้อมูลพิกัดพื้นที่ (polygon) ของกรุงเทพมหานคร

### 3.2.2 การเตรียมข้อมูล (Data Preparation)

#### 3.2.2.1 คัดแยกข้อมูลให้อยู่ในกรุงเทพมหานคร (Filter GPS data for Bangkok)

นำข้อมูลจีพีเอส (GPS) ของแท็กซี่และข้อมูลพื้นที่ (Polygon) ของกรุงเทพมหานคร มากรองข้อมูลโดยใช้พิกัดพื้นที่ (polygon) ของกรุงเทพมหานคร มาใช้คัดกรองข้อมูลจีพีเอส (GPS) ของแท็กซี่ที่อยู่ในประเทศไทย ให้เหลือแค่ข้อมูลที่อยู่ในกรุงเทพมหานครเท่านั้น

#### 3.2.2.2 คัดแยกข้อมูลรับ-ส่งผู้โดยสาร (Extract pickup/dropoff locations)

นำข้อมูลจีพีเอส (GPS) ของแท็กซี่ที่อยู่ในกรุงเทพมหานครแล้ว มาทำการวิเคราะห์ข้อมูล โดยจะใช้ข้อมูลสัญญาณไฟของรถแท็กซี่มาทำการวิเคราะห์ว่าแท็กซี่นั้นมีการรับผู้โดยสารที่จุดใด (pickup points) ละทำการส่งผู้โดยสารที่จุดใด (dropoff points)

### 3.2.2.3 คำนวณหาระยะทาง ค่าโดยสารและเวลาเดินทาง

#### (Calculate distance, price and travel time)

คำนวณหาระยะทาง (distance) จากจุดรับผู้โดยสาร (pickup points) ไปยังจุดส่งผู้โดยสาร (dropoff points) จำนวนเงิน (price) ที่แท็กซี่จะได้รับในแต่ละครั้งที่ทำการส่งผู้โดยสารในแต่ละรอบ โดยจะคิดจากระยะทาง (distance) ที่ใช้ในการเดินทาง และ จำนวนเวลาเดินทาง (travel time) ที่แท็กซี่ใช้ในการเดินทางจากจุดรับผู้โดยสาร (pickup points) ไปยังจุดส่งผู้โดยสาร (dropoff points)

### 3.2.3 การวิเคราะห์ข้อมูล (Data Analytics)

#### 3.2.3.1 การคัดแยกข้อมูลด้วยดีบีสแกน (DBSCAN clustering)

ใช้ดีบีสแกน (DBSCAN) เป็นเครื่องมือสำหรับการจัดกลุ่มจุดรับผู้โดยสาร (pickup points) และจุดส่งผู้โดยสาร (dropoff points) เพื่อช่วยระบุพื้นที่ที่มีความต้องการในการใช้บริการรถแท็กซี่สูง

#### 3.2.3.2 สรุปข้อมูลหลังการคัดแยก (Summarize Cluster data)

เพื่อสรุปการจัดกลุ่มจุดรับผู้โดยสาร (pickup points) และจุดส่งผู้โดยสาร (dropoff points) ทั้งจำนวนของกลุ่มและเนื้อหาของแต่ละกลุ่ม นอกจากนี้ยังควรเน้นไปที่ความสำคัญของข้อมูลที่ได้รับจากกระบวนการแบ่งกลุ่มเช่นแนวโน้มหรือลักษณะเฉพาะ เพื่อใช้ในการจัดการข้อมูลในอนาคต สามารถนำเสนอผลลัพธ์ของการแบ่งกลุ่มด้วยภาพแสดงผลเพื่อช่วยในการแสดงความสัมพันธ์ระหว่างข้อมูลที่ได้แบ่งกลุ่มกันได้ดียิ่งขึ้น

#### 3.2.3.3 หาความสัมพันธ์ระหว่างความถี่ในการรับ-ส่งผู้โดยสารกับเวลาเดินทาง

#### (Find Correlating pickup/dropoff freq with travel time)

หาความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup points) ขอดนิยมกับเวลาเดินทาง (travel time)

#### 3.2.3.4 วิเคราะห์แนะนำจุดรับผู้โดยสารตามช่วงเวลา

#### (Recommend pickup points based on time interval analysis)

วิเคราะห์ตามช่วงเวลาของข้อมูลเพื่อแนะนำจุดรับผู้โดยสาร (pickup points) ที่เหมาะสมสำหรับรถแท็กซี่ในช่วงเวลานั้นๆ

### 3.2.4 ผลลัพธ์ (Results)

#### 3.2.4.1 จุดรับ-ส่งผู้โดยสารยอดนิยม (Popular pickup/dropoff clusters)

ผลลัพธ์จากการใช้ดีบีสแกน (DBSCAN) เป็นเครื่องมือสำหรับการจัดกลุ่มจุดรับผู้โดยสาร (pickup points) และจุดส่งผู้โดยสาร (dropoff points) เพื่อช่วยระบุพื้นที่ที่มีความต้องการในการใช้บริการรถแท็กซี่สูง

#### 3.2.4.2 ความสัมพันธ์ระหว่างความถี่ในการรับ-ส่งผู้โดยสารกับเวลาเดินทาง (Correlating pickup/dropoff freq with travel time)

ผลลัพธ์จากการหาความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup points) และจุดส่งผู้โดยสาร (dropoff points) ยอดนิยมกับเวลาเดินทาง (travel time)

#### 3.2.4.3 แนะนำจุดรับผู้โดยสารตามช่วงเวลา (Time-based pickup points recommendations)

ผลลัพธ์จากการวิเคราะห์ตามช่วงเวลาของข้อมูลเพื่อแนะนำจุดรับผู้โดยสาร (pickup points) ที่เหมาะสมสำหรับรถแท็กซี่ในช่วงเวลานั้นๆ

### 3.3 เป้าหมายของการพัฒนา

ส่วนของการอธิบายถึงเป้าหมายของการพัฒนา ทั้งเป้าหมายหลัก และเป้าหมายย่อยระหว่างทาง ตามที่ได้วางแผนไว้ในขั้นตอนการพัฒนา

#### 3.3.1 การเตรียมข้อมูล

3.3.1.1 ใช้พิกัดพื้นที่ (polygon) ของกรุงเทพมหานครมาใช้คัดกรองข้อมูลจีพีเอส (GPS) ของแท็กซี่ที่อยู่ในประเทศไทย ให้เหลือแค่ข้อมูลที่อยู่ในกรุงเทพมหานครเท่านั้น

3.3.1.2 ใช้ข้อมูลสัญญาณไฟของรถแท็กซี่มาทำการหาว่าแท็กซี่นั้นมีการรับผู้โดยสารที่จุดใด (pickup points) ละทำการส่งผู้โดยสารที่จุดใด (dropoff points) เพื่อให้สามารถแยกว่าข้อมูลนั้นๆเป็นการรับผู้โดยสาร (pickup) หรือส่งผู้โดยสาร (dropoff)

### 3.3.2 การหาความสัมพันธ์ของข้อมูล

3.3.2.1 การหาจุดรับผู้โดยสาร (pickup points) จุดส่งผู้โดยสาร (dropoff points) จากข้อมูลละติจูด (latitude) และ ลองจิจูด (longitude)

3.3.2.2 การคำนวณหาระยะทาง (distance) จากจุดรับผู้โดยสาร (pickup points) ไปยังจุดส่งผู้โดยสาร (dropoff points)

3.3.2.3 การคำนวณเงิน (price) ที่แท็กซี่จะได้รับในแต่ละครั้งที่ทำการส่งผู้โดยสารในแต่ละรอบ โดยจะคิดจากระยะทาง (distance) ที่ใช้ในการเดินทาง

3.3.2.4 การคำนวณเวลาเดินทาง (travel time) ที่แท็กซี่ใช้ในการเดินทางจากจุดรับผู้โดยสาร (pickup points) ไปยังจุดส่งผู้โดยสาร (dropoff points)

### 3.3.3 การจัดกลุ่มข้อมูล

3.3.3.1 การจัดกลุ่มของข้อมูลจุดรับผู้โดยสาร (pickup points) และจุดส่งผู้โดยสาร (dropoff points) ด้วยดีบีเอสแกน (DBSCAN) เป็นเครื่องมือในการจัดกลุ่มข้อมูล เพื่อช่วยระบุพื้นที่ที่มีความต้องการใช้บริการรถแท็กซี่สูง 300 จุดในกรุงเทพมหานคร

3.3.3.2 ได้พื้นที่ที่มีความต้องการใช้บริการรถแท็กซี่สูง 300 จุดในกรุงเทพมหานคร จาก จุดที่เป็นจุดรับผู้โดยสาร (pickup points) และจุดส่งผู้โดยสาร (dropoff points)

3.3.3.3 สามารถนำเสนอผลลัพธ์ของการแบ่งกลุ่มด้วยภาพแสดงผลเพื่อช่วยในการแสดงความสัมพันธ์ระหว่างข้อมูลที่ได้แบ่งกลุ่มกันได้ง่ายขึ้น

### 3.3.4 การวิเคราะห์ข้อมูล

3.3.4.1 การหาความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup points) และจุดส่งผู้โดยสาร (dropoff points) ขอดนิยมนกับเวลาเดินทาง (travel time)

3.3.4.2 วิเคราะห์ข้อมูลเพื่อแนะนำจุดรับผู้โดยสาร (pickup points) ที่เหมาะสมสำหรับรถแท็กซี่ในช่วงเวลานั้นๆ

### 3.3.5 การแสดงผลพีชการวิเคราะห์ด้วย Power BI

3.3.5.1 นำข้อมูลที่ได้หลังจากการวิเคราะห์ข้อมูล มาแสดงผลเป็น แผนที่ที่สามารถแสดงผลความหนาแน่นของผู้โดยสารในแต่ละพื้นที่ พร้อมด้วยข้อมูลที่เกี่ยวข้องได้ ผ่าน Power BI

ตามเป้าหมายของการพัฒนาที่ได้กำหนดขึ้นนั้น ผู้จัดทำได้ทำการเตรียมข้อมูล การหาความสัมพันธ์ของข้อมูล การจัดกลุ่มข้อมูล การวิเคราะห์ข้อมูล และจัดทำการแสดงผลพีชการวิเคราะห์ด้วย Power BI เป็นที่เรียบร้อยแล้ว

## 3.4 การเตรียมข้อมูล

ส่วนของการอธิบายถึงการเตรียมข้อมูลในการพัฒนา ประกอบไปด้วยการเลือกข้อมูลพื้นที่เฉพาะกรุงเทพมหานคร การเตรียมข้อมูลในรูปการรับผู้โดยสารและการส่งผู้โดยสาร และการคำนวณระยะทาง เวลาเดินทางและค่าโดยสาร ในแต่ละการรับส่งผู้โดยสาร

### 3.4.1 การเลือกข้อมูลพื้นที่เฉพาะกรุงเทพมหานคร

ข้อมูลตั้งต้น (ตารางที่ 3.2) เป็นข้อมูลจากทั่วประเทศ ข้อมูลที่นำมาวิเคราะห์จะจำกัดขอบเขตอยู่ในกรุงเทพมหานครเท่านั้น ในการจำกัดข้อมูลให้อยู่เฉพาะในกรุงเทพมหานครนั้นจะช่วยให้การวิเคราะห์ง่ายและมีความแม่นยำขึ้น โดยอิงข้อมูลพิกัดพื้นที่ (polygon) ของกรุงเทพมหานคร ที่เป็นรูปร่างที่สร้างจากคู่อันดับละติจูด (latitude) และลองจิจูด (longitude) จำนวน 189 คู่พิกัด หากข้อมูลไหนพบว่าอยู่ในพิกัดพื้นที่ (polygon) นี้ก็จะถือว่าข้อมูลนั้นอยู่ในกรุงเทพมหานคร ผลลัพธ์ของขั้นตอนนี้จะได้เป็นคอลัมน์ BKK ที่แสดงว่าอยู่ในกรุงเทพมหานคร (True) และไม่อยู่ในกรุงเทพมหานคร (False) ดังตารางที่ 3.3

	VehicleID	lat	lon	timestamp	for_hire_light	engine_acc
0	++iY830QzVC	13.770	100.473	2020-01-01 00:01:28	0	0
1	++iY830QzVC	13.7708	100.473	2020-01-01 00:04:28	0	0
2	++iY830QzVC	13.7708	100.473	2020-01-01 00:07:28	0	0
...	...	...	...	...	...	...
147437	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:54:35	1	0
147438	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:57:35	1	0

ตารางที่ 3.2 ตารางตัวอย่างข้อมูลก่อนการคัดแยก (6 คอลัมน์)

	VehicleID	lat	lon	timestamp	for_hire_light	engine_acc	BKK
0	++iY830QzVC	13.770	100.473	2020-01-01 00:01:28	0	0	True
1	++iY830QzVC	13.7708	100.473	2020-01-01 00:04:28	0	0	True
2	++iY830QzVC	13.7708	100.473	2020-01-01 00:07:28	0	0	True
...	...	...	...	...	...	...	...
147437	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:54:35	1	0	False
147438	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:57:35	1	0	False

ตารางที่ 3.3 ตารางตัวอย่างข้อมูลหลังคัดแยกให้มีแต่ในจังหวัดกรุงเทพมหานคร (7 คอลัมน์)

### 3.4.2 การเตรียมข้อมูลในรูปการรับผู้โดยสารและการส่งผู้โดยสาร

กระบวนการในการหาจุดรับผู้โดยสาร (pickup points) และจุดส่งผู้โดยสาร (dropoff points) จากข้อมูลไฟแสดงสัญญาณรับผู้โดยสาร (for\_hire\_light) ของรถแท็กซี่ที่จะแตกต่างกันไปตามการเปลี่ยนแปลงจาก 0 เป็น 1 หรือ จาก 1 เป็น 0 โดยทั่วไปแล้ว ไฟแสดงสัญญาณรับผู้โดยสารในรถแท็กซี่จะถูกเปิดเมื่อรถพร้อมให้บริการ (หมายถึงไม่มีผู้โดยสารอยู่ภายใน) หรือเมื่อมีผู้โดยสารอยู่ในรถแล้ว ซึ่งสามารถนำข้อมูลนี้มาคัดแยกข้อมูลการเดินทางได้ ดังนี้

#### 3.4.2.1 จุดรับผู้โดยสาร (pick up points)

กระบวนการในการหาจุดรับผู้โดยสาร (pickup points) จากข้อมูลไฟแสดงสัญญาณรับผู้โดยสาร (for\_hire\_light) ทำโดยการตรวจสอบว่าจุดการเดินทางไหนที่มีการเปลี่ยนแปลงของสัญญาณรับผู้โดยสาร (for\_hire\_light) จากค่า 1 เป็น 0 จะนับว่าจุดนั้นเป็นจุดรับผู้โดยสาร (pickup points) ผลลัพธ์ของขั้นตอนนี้จะได้เป็นคอลัมน์ pickup ที่แสดงว่าเป็นจุดรับผู้โดยสาร (pickup points) (True) และไม่เป็นจุดรับผู้โดยสาร (pickup points) (False) ดังตารางที่ 3.4

	VehicleID	lat	lon	timestamp	for_hire_light	engine_acc	BKK	pickup
0	++iY830QzVC	13.770	100.473	2020-01-01 00:01:28	0	0	True	False
1	++iY830QzVC	13.7708	100.473	2020-01-01 00:04:28	0	0	True	False
2	++iY830QzVC	13.7708	100.473	2020-01-01 00:07:28	0	0	True	False
...	...	...	...	...	...	...	...	...
147437	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:54:35	1	0	False	False
147438	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:57:35	1	0	False	False

ตารางที่ 3.4 ตารางตัวอย่างข้อมูลหลังคัดแยกจุดรับผู้โดยสาร (pick up points) (8 คอลัมน์)

### 3.4.2.2 จุดส่งผู้โดยสาร (drop off points)

กระบวนการในการหาจุดรับผู้โดยสาร (pickup points) และจุดส่งผู้โดยสาร (dropoff points) จากข้อมูลไฟแสดงสัญญาณรับผู้โดยสาร (for\_hire\_light) ทำโดยการตรวจสอบว่าการเดินทางไหนที่มีการเปลี่ยนแปลงของสัญญาณรับผู้โดยสาร (for\_hire\_light) จากค่า 0 เป็น 1 จะนับว่าจุดนั้นเป็นจุดส่งผู้โดยสาร (dropoff points) ผลลัพธ์ของขั้นตอนนี้จะได้เป็นคอลัมน์ dropoff ที่แสดงว่าเป็นจุดส่งผู้โดยสาร (dropoff points) (True) และไม่เป็นส่งผู้โดยสาร (dropoff points) (False) ดังตารางที่ 3.5

	VehicleID	lat	lon	timestamp	for_hire_light	engine_acc	BKK	pickup	dropoff
0	++iY830QzVC	13.770	100.473	2020-01-01 00:01:28	0	0	True	False	False
1	++iY830QzVC	13.7708	100.473	2020-01-01 00:04:28	0	0	True	False	False
2	++iY830QzVC	13.7708	100.473	2020-01-01 00:07:28	0	0	True	False	False
...	...	...	...	...	...	...	...	...	...
147437	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:54:35	1	0	False	False	True
147438	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:57:35	1	0	False	False	False

ตารางที่ 3.5 ตารางตัวอย่างข้อมูลหลังคัดแยกจุดส่งผู้โดยสาร (drop off points) (9 คอลัมน์)

### 3.4.3 การคำนวณระยะทาง เวลาเดินทางและค่าโดยสาร ในแต่ละการรับส่งผู้โดยสาร

#### 3.4.3.1 การคำนวณระยะทาง (distance)

ในขั้นตอนนี้จะเป็นวิธีการคำนวณระยะทาง (distance) ระหว่างสองจุดบนพื้นผิวของโลกโดยใช้สูตรของฮาเวอร์ไซน์ (Haversine Formula) ซึ่งเป็นวิธีการคำนวณระยะทาง (distance) รอบวงกลมระหว่างสองจุดบนพื้นผิวของทรงกลม สูตรนี้นิยมใช้ในการนำทางและมีประโยชน์มากในการคำนวณระยะทาง (distance) ระหว่างสถานที่บนพื้นผิวของโลก สูตรนี้คำนึงถึงรัศมีของโลกและพิกัดละติจูดและลองจิจูดของสองจุด โดยใช้ฟังก์ชันตรีโกณมิติเพื่อคำนวณระยะทาง (distance) ระหว่างสองจุด ซึ่งมีสูตรดังรูปที่ 3.4

```

1 def haversine(lon1, lat1, lon2, lat2):
2     lon1, lat1, lon2, lat2 = map(radians, [lon1, lat1, lon2, lat2])
3     dlon = lon2 - lon1
4     dlat = lat2 - lat1
5     a = sin(dlat/2)**2 + cos(lat1) * cos(lat2) * sin(dlon/2)**2
6     c = 2 * asin(sqrt(a))
7     km = 6371 * c
8     m = km * 1000
9     return m

```

รูปที่ 3.4 สูตรของฮาเวอร์ไซน์ (Haversine Law)

นำสูตรของฮาเวอร์ไซน์ (Haversine Formula) ดังรูปที่ 3.4 มาคำนวณระยะ (distance) แต่ละครั้งเริ่มจาก จุดรับผู้โดยสาร (pickup points) จนถึง จุดส่งผู้โดยสาร (dropoff points) ซึ่งการคิดระยะทางจะคิดที่ละแถวของข้อมูลจนถึงจุดส่งผู้โดยสาร (dropoff points) ผลลัพธ์ของขั้นตอนนี้จะได้เป็นคอลัมน์ distance ที่แสดงระยะทางระหว่างแต่ละคู่ของจุดในแต่ละแถวของข้อมูล ตั้งแต่จุดรับผู้โดยสาร (pickup points) ไปจนถึงจุดส่งผู้โดยสาร (dropoff points) ดังตารางที่ 3.6

	VehicleID	lat	lon	timestamp	for_hire_light	engine_acc	BKK	pickup	dropoff	distance
0	++iY830QzVC	13.770	100.473	2020-01-01 00:01:28	0	0	True	False	False	0
1	++iY830QzVC	13.7708	100.473	2020-01-01 00:04:28	0	0	True	False	False	0
2	++iY830QzVC	13.7708	100.473	2020-01-01 00:07:28	0	0	True	False	False	0
...	...	...	...	...	...	...	...	...	...	...
147437	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:54:35	1	0	False	False	True	0
147438	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:57:35	1	0	False	False	False	0

ตารางที่ 3.6 ตารางตัวอย่างข้อมูลหลังคำนวณระยะทาง (distance) (10 คอลัมน์)

### 3.4.3.2 การคำนวณเวลาเดินทาง (travel time)

ในขั้นตอนนี้จะเป็นวิธีการคำนวณเวลาเดินทาง (travel time) จากจุดรับผู้โดยสาร (pickup points) ไปยังจุดส่งผู้โดยสาร (dropoff points) โดยวิธีการจะคิดจากส่วนต่างระหว่างการประทับเวลา (timestamp) ของทั้งสองจุด ผลลัพธ์ของขั้นตอนนี้จะได้เป็นคอลัมน์ travel time ที่แสดงเวลาเดินทาง (travel time) จากจุดรับผู้โดยสาร (pickup points) ไปยังจุดส่งผู้โดยสาร (dropoff points) ดังตารางที่ 3.7

	VehicleID	lat	lon	timestamp	for_hire_light	engine_acc	BKK	pickup	dropoff	distance	travel time
0	++iY830QzVC	13.770	100.473	2020-01-01 00:01:28	0	0	True	False	False	0	0
1	++iY830QzVC	13.7708	100.473	2020-01-01 00:04:28	0	0	True	False	False	0	0
2	++iY830QzVC	13.7708	100.473	2020-01-01 00:07:28	0	0	True	False	False	0	0
...	...	...	...	...	...	...	...	...	...	...	...
147437	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:54:35	1	0	False	False	True	0	0
147438	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:57:35	1	0	False	False	False	0	0

ตารางที่ 3.7 ตารางตัวอย่างข้อมูลหลังคำนวณเวลาเดินทาง (travel time) (11 คอลัมน์)

### 3.4.3.3 การคำนวณค่าโดยสาร (price)

ในขั้นตอนนี้จะเป็นวิธีการคำนวณค่าโดยสาร (price) โดยวิธีการจะคำนวณด้วยฟังก์ชันในรูปที่ 3.5 ซึ่งเป็นการคำนวณค่าโดยสารค่าแท็กซี่ตามระยะทาง (distance) ที่เดินทางเป็นกิโลเมตร ค่าโดยสาร (price) จะคำนวณตามชุดของอัตราที่กำหนดไว้ล่วงหน้าสำหรับระยะทาง (distance) ต่างๆ

```

1  def taxi_price(kilo):
2
3      price = 0
4
5      if kilo > 80:
6          x_kilo = (kilo - 80)
7          price += x_kilo * 10.50
8          kilo -= x_kilo
9
10     if kilo > 60:
11         x_kilo = (kilo - 60)
12         price += x_kilo * 9
13         kilo -= x_kilo
14
15     if kilo > 40:
16         x_kilo = (kilo - 40)
17         price += x_kilo * 8
18         kilo -= x_kilo
19
20     if kilo > 20:
21         x_kilo = (kilo - 20)
22         price += x_kilo * 7.50
23         kilo -= x_kilo
24
25         price += x_kilo * 7.50
26         kilo -= x_kilo
27
28     if kilo > 10:
29         x_kilo = (kilo - 10)
30         price += x_kilo * 6.50
31         kilo -= x_kilo
32
33     if kilo > 1:
34         x_kilo = (kilo - 1)
35         price += x_kilo * 5.50
36         kilo -= x_kilo
37
38     if kilo == 1:
39         price += 35
40
41     if round(price) % 2 == 0:
42         price += 1
43
44     return price

```

รูปที่ 3.5 สูตรการคำนวณเงิน (price) จากระยะทาง (distance)

จากฟังก์ชันในรูปที่ 3.5 คำสั่ง if แต่ละคำสั่งจะตรวจสอบว่าระยะทาง (distance) ที่เดินทางนั้นมากกว่าเกณฑ์ที่กำหนดหรือไม่ และถ้าเป็นเช่นนั้น จะคำนวณค่าโดยสารตามอัตราสำหรับช่วงระยะทางนั้น ค่าเกณฑ์และอัตราที่ใช้มีดังนี้

- ระยะทาง > 80 กิโลเมตร: อัตราค่าบริการ = 10.50 บาทต่อกิโลเมตร
- ระยะทาง > 60 กิโลเมตร: อัตราค่าบริการ = 9.00 บาทต่อกิโลเมตร
- ระยะทาง > 40 กิโลเมตร: อัตราค่าบริการ = 8.00 บาทต่อกิโลเมตร
- ระยะทาง > 20 กิโลเมตร: อัตราค่าบริการ = 7.50 บาทต่อกิโลเมตร
- ระยะทาง > 10 กิโลเมตร: อัตราค่าบริการ = 6.50 บาทต่อกิโลเมตร

- ระยะทาง > 1 กิโลเมตร: อัตราค่าบริการ = 5.50 บาทต่อกิโลเมตร
- ระยะทาง = 1 กิโลเมตร: อัตราค่าบริการคงที่ 35 บาท

ผลลัพธ์ของขั้นตอนนี้จะได้เป็นคอลัมน์ price ที่แสดงค่าโดยสาร (price) ค่าแท็กซี่ตามระยะทาง (distance) ดังตารางที่ 3.8

	VehicleID	lat	lon	timestamp	for_hire_light	engine_acc	BKK	pickup	dropoff	distance	travel time	price
0	++iY830QzVC	13.770	100.473	2020-01-01 00:01:28	0	0	True	False	False	0	0	0
1	++iY830QzVC	13.7708	100.473	2020-01-01 00:04:28	0	0	True	False	False	0	0	0
2	++iY830QzVC	13.7708	100.473	2020-01-01 00:07:28	0	0	True	False	False	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
147437	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:54:35	1	0	False	False	True	0	0	0
147438	zzLYPcDONa2	13.5958	100.313	2020-01-01 00:57:35	1	0	False	False	False	0	0	0

ตารางที่ 3.8 ตารางตัวอย่างข้อมูลหลังคำนวณเงิน (price) (12 คอลัมน์)

### 3.5 การวิเคราะห์ข้อมูล

ส่วนของการอธิบายถึงการวิเคราะห์ข้อมูลหลังจากที่ได้ทำการเตรียมข้อมูลเสร็จสิ้นแล้ว ประกอบไปด้วย การจัดกลุ่มข้อมูล การสรุปการจัดกลุ่มข้อมูล การหาความสัมพันธ์ของข้อมูลและการวิเคราะห์การแนะนำจุดรอรับผู้โดยสารตามช่วงเวลา

#### 3.5.1 การจัดกลุ่มข้อมูลจุดรับส่งผู้โดยสารยอดนิยม

การจัดกลุ่มข้อมูลจุดรับส่งผู้โดยสารยอดนิยมด้วย ดีบีเอสแกน (DBSCAN) จะช่วยให้สามารถแบ่งข้อมูลออกเป็นกลุ่มตามลักษณะการเดินทางของผู้โดยสารและสามารถหาพื้นที่หรือเส้นทางที่มีการใช้บริการขนส่งสูงสุดได้ โดย DBSCAN (Density-Based Spatial Clustering of Applications with Noise) คือ เทคนิคการจัดกลุ่มข้อมูลที่ใช้ระบบจุดและความหนาแน่นของข้อมูลในการจัดกลุ่ม

### 3.5.1.1 การกำหนดค่าพารามิเตอร์ (Parameter) ให้กับการจัดกลุ่มข้อมูล

ดีบีสแกน (DBSCAN) จะจัดกลุ่มข้อมูลด้วยการกำหนดพารามิเตอร์ที่สำคัญโดยใช้การจัดกลุ่มจุดของข้อมูลที่มีความหนาแน่นและอยู่ใกล้กันอยู่ด้วยกัน พร้อมด้วยการแยกจุดที่มีความหนาแน่นน้อยออกไป

โดยจะมีพารามิเตอร์ (Parameter) ที่สำคัญอยู่ด้วยกัน 2 ตัว คือ

1) ค่า eps คือ รัศมีของวงกลมรอบจุดรับส่งที่จะนับเป็นจุดใกล้เคียง (neighbor) ซึ่งจะตัดข้อมูลออกไปจนกว่าจะเหลือเฉพาะจุดรับส่งที่มีจำนวนจุดใกล้เคียง (neighbor) ตามเกณฑ์ที่กำหนด ค่าที่กำหนดในการศึกษา คือ  $eps = 0.002$

2) ค่า MinPts คือ จำนวนจุดข้อมูล (Data Points) ขั้นต่ำสำหรับการสร้างคลัสเตอร์ ค่าที่กำหนดในโครงการนี้ คือ  $MinPts = 16$

การจัดกลุ่มข้อมูลในการศึกษานี้จะใช้ข้อมูลตั้งต้นที่นำมาจัดกลุ่มทั้งหมด 84 วัน ซึ่งประกอบด้วย 12 เดือน นำข้อมูลมาจัดกลุ่มเดือนละ 7 วัน ดังนี้

- มกราคม : วันที่ 5-11
- กุมภาพันธ์ : วันที่ 16-22
- มีนาคม : วันที่ 15-21
- เมษายน : วันที่ 19-25
- พฤษภาคม : วันที่ 17-23
- มิถุนายน : วันที่ 7-13
- กรกฎาคม : วันที่ 12-18
- สิงหาคม : วันที่ 16-22
- กันยายน : วันที่ 6-12
- ตุลาคม : วันที่ 4-10
- พฤศจิกายน : วันที่ 8-14
- ธันวาคม : วันที่ 13-19

โดยหลักการเลือกวัน 7 วันในแต่ละเดือน คือ

- 1) จะต้องเป็นช่วงเวลา 7 วันติดกันที่เป็นวันอาทิตย์-วันเสาร์

2) จะต้องเป็นวันที่ไม่ใช่วันหยุดนักขัตฤกษ์

### 3.5.1.2 การจัดกลุ่มข้อมูลจุดรับผู้โดยสาร (pickup points)

จัดกลุ่มข้อมูลจุดรับผู้โดยสาร (pickup points) ตามความหนาแน่น จะได้ผลลัพธ์เป็นจุดคลัสเตอร์ (cluster) รับผู้โดยสารยอดนิยม

### 3.5.1.3 การจัดกลุ่มข้อมูลจุดส่งผู้โดยสาร (dropoff points)

จัดกลุ่มข้อมูลจุดส่งผู้โดยสาร (dropoff points) ตามความหนาแน่น จะได้ผลลัพธ์เป็นจุดคลัสเตอร์ (cluster) ส่งผู้โดยสารยอดนิยม

## 3.5.2 การสรุปการจัดกลุ่มข้อมูล

จากจุดทั้งหมดที่วิเคราะห์ได้ พบว่า สถานที่ส่วนใหญ่ที่มีการกระจุกตัวของรถแท็กซี่อย่างหนาแน่น เป็นสถานที่ที่เกี่ยวข้องกับการขนส่งเป็นจำนวนมาก เช่น สถานีรถไฟฟ้า BTS สถานีรถไฟฟ้าใต้ดิน MRT ขนส่งเอกมัย สนามบินดอนเมือง เป็นต้น

ข้อมูลของกลุ่มของข้อมูล (cluster) ที่ได้ จะประกอบไปด้วย

1. เลขประจำกลุ่มข้อมูล (clusterID)
2. ละติจูด (latitude) และ ลองจิจูด (longitude) ที่เป็นจุดกึ่งกลาง (centroid)
3. จำนวนจุดทั้งหมด (counts) ที่อยู่ในกลุ่มของข้อมูล (cluster) นั้นๆ
4. วันในสัปดาห์ของกลุ่มคลัสเตอร์ (day)
5. ชั่วโมงของกลุ่มคลัสเตอร์ (hour)

## 3.5.3 การหาความสัมพันธ์ของข้อมูล

หาความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอดนิยมกับเวลาเดินทาง (travel time)

### 3.5.3.1 จำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency)

คือ ความถี่ว่ารถ 1 คันมีการเดินทางผ่าน จุดรับผู้โดยสาร (pickup points) ยอดนิยม เป็นจำนวนกี่ครั้งใน 1 วัน

### 3.5.3.2 เวลาเดินทาง (travel time)

คือ เวลาจากจุดรับผู้โดยสาร (pickup points) ไปยังจุดส่งผู้โดยสาร (dropoff points) โดยวิธีการจะคิดจากส่วนต่างระหว่าง การประทับเวลา (timestamp) ของทั้งสองจุด

จากข้อมูลทั้งสองนี้ จะทำการหาความสัมพันธ์ของข้อมูล วิเคราะห์ความสัมพันธ์ของ จำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) และเวลาเดินทาง (travel time) โดยดูทิศทางของ กราฟว่าเคลื่อนที่ไปทางไหน ถ้ากราฟเคลื่อนที่ขึ้น แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับเวลาเดินทาง (travel time) มีความสัมพันธ์กันเชื่อมโยงกันในทิศทางเดียวกัน แต่ถ้ากราฟ เคลื่อนที่ลง แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับเวลาเดินทาง (travel time) ไม่มีความสัมพันธ์กันหรือมีความสัมพันธ์ที่แตกต่างกันในทิศทางตรงกันข้าม

### 3.5.4 การแนะนำจุดรับผู้โดยสาร (pickup points) ตามแต่ละช่วงเวลา

#### 3.5.4.1 เพิ่มคอลัมน์วัน (day) และชั่วโมง (hour)

การแปลงข้อมูลวัน (day) และเวลา (hour) โดยใช้ฟังก์ชันหรือ โมดูลเพื่อแปลง ข้อมูลวันและเวลาในข้อมูลแท็กชี้ให้อยู่ในรูปแบบที่เหมาะสมสำหรับการวิเคราะห์ ในกรณีนี้ จะใช้ โมดูล datetime ของภาษาไพธอน (Python) เพื่อแปลงข้อมูลวันและเวลาเป็นวันในอาทิตย์และชั่วโมงได้

การแปลงข้อมูลวัน (day) โดยนำข้อมูลจาก timestamp มาแยกเป็นวันใน อาทิตย์ ซึ่งกำหนดค่าของแต่ละวันเป็น ดังนี้

- 1) วันอาทิตย์ : 1-Sun
- 2) วันจันทร์ : 2-Mon
- 3) วันอังคาร : 3-Tue
- 4) วันพุธ : 4-Wed
- 5) วันพฤหัสบดี : 5-Thu
- 6) วันศุกร์ : 6-Fri
- 7) วันเสาร์ : 7-Sat

การแปลงข้อมูลชั่วโมง (hour) โดยนำข้อมูลจาก timestamp มาแยกเป็นชั่วโมง ในแต่ละวัน ซึ่งกำหนดค่าของแต่ละชั่วโมงเป็นตัวเลข 0-23 จากเวลาที่เที่ยงคืน-ห้าทุ่ม หลังการคัดแยกจะ ได้คอลัมน์ใหม่เป็นข้อมูลวัน (day) และเวลา (hour) ดังตารางที่ 3.9

4	VehicleID	lat	lon	timestamp	for_hire_light	engine_acc	BKK	pickup	dropoff	distance	travel time	price	day	hour
0	++iY830QzV C	13.770	100.47 3	2020-01- 01 00:01:28	0	0	True	False	False	0	0	0	4- Wed	0
1	++iY830QzV C	13.770 8	100.47 3	2020-01- 01 00:04:28	0	0	True	False	False	0	0	0	4- Wed	0
2	++iY830QzV C	13.770 8	100.47 3	2020-01- 01 00:07:28	0	0	True	False	False	0	0	0	4- Wed	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14743 7	zzLYPcDON a2	13.595 8	100.31 3	2020-01- 01 00:54:35	1	0	Fals e	False	True	0	0	0	4- Wed	0
14743 8	zzLYPcDON a2	13.595 8	100.31 3	2020-01- 01 00:57:35	1	0	Fals e	False	False	0	0	0	4- Wed	0

ตารางที่ 3.9 ตารางตัวอย่างข้อมูลหลังเพิ่มคอลัมน์วัน (day) และชั่วโมง (hour) (14 คอลัมน์)

หลังจากทำการเพิ่มคอลัมน์วัน (day) และชั่วโมง (hour) แล้ว จะทำการวิเคราะห์ตามช่วงเวลาของข้อมูลเพื่อแนะนำจุดรับผู้โดยสาร (pickup points) ที่เหมาะสมสำหรับรถแท็กซี่ในช่วงเวลานั้นๆ ผ่านตัวแปรสภาพแวดล้อมอย่าง ระยะทาง (distance) ค่าโดยสาร (price) และความหนาแน่นของการรับผู้โดยสาร (pickup)

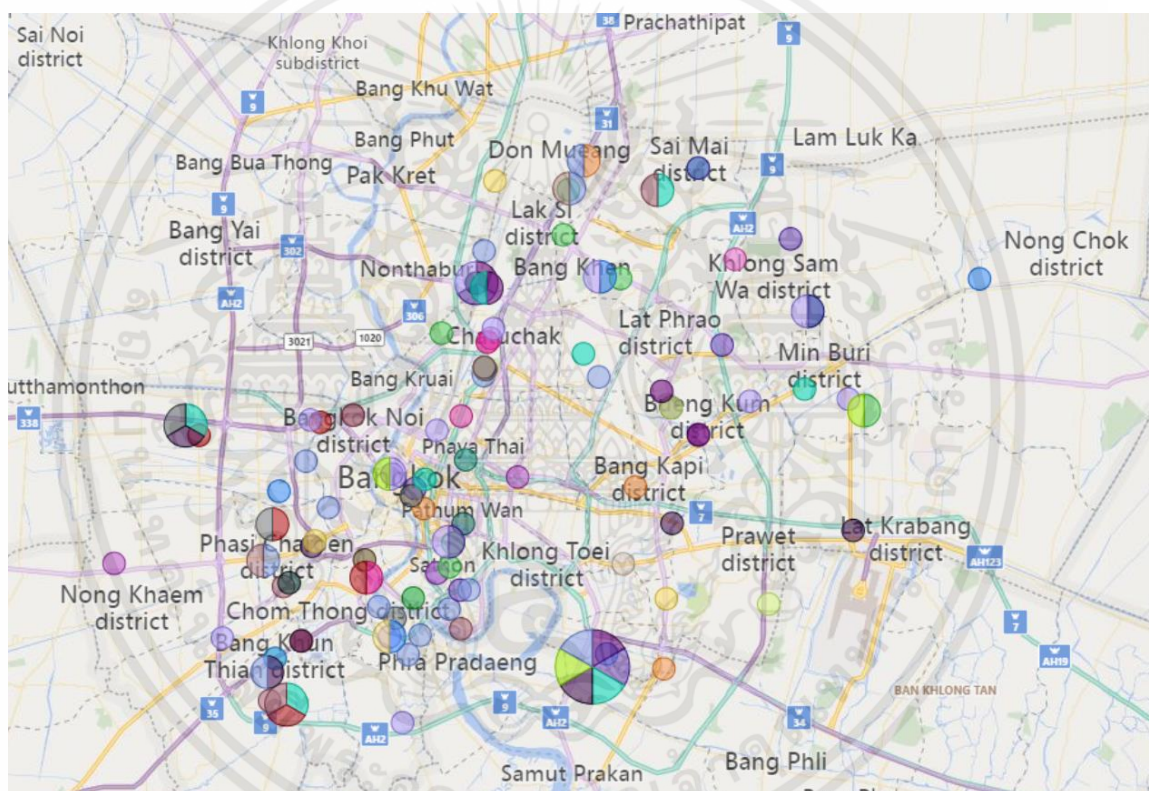
## บทที่ 4

### ผลการทดลอง

#### 4.1 คลัสเตอร์จุดรับส่งผู้โดยสารโดยสายรถโดยสาร

##### 4.1.1 การจัดกลุ่มข้อมูลจุดรับส่งผู้โดยสาร (pickup points)

จัดกลุ่มข้อมูลจุดรับส่งผู้โดยสาร (pickup points) ตามความหนาแน่น ได้ผลลัพธ์เป็นจุดคลัสเตอร์ (cluster) รับส่งผู้โดยสารโดยสายรถโดยสาร ดังรูปที่ 4.1

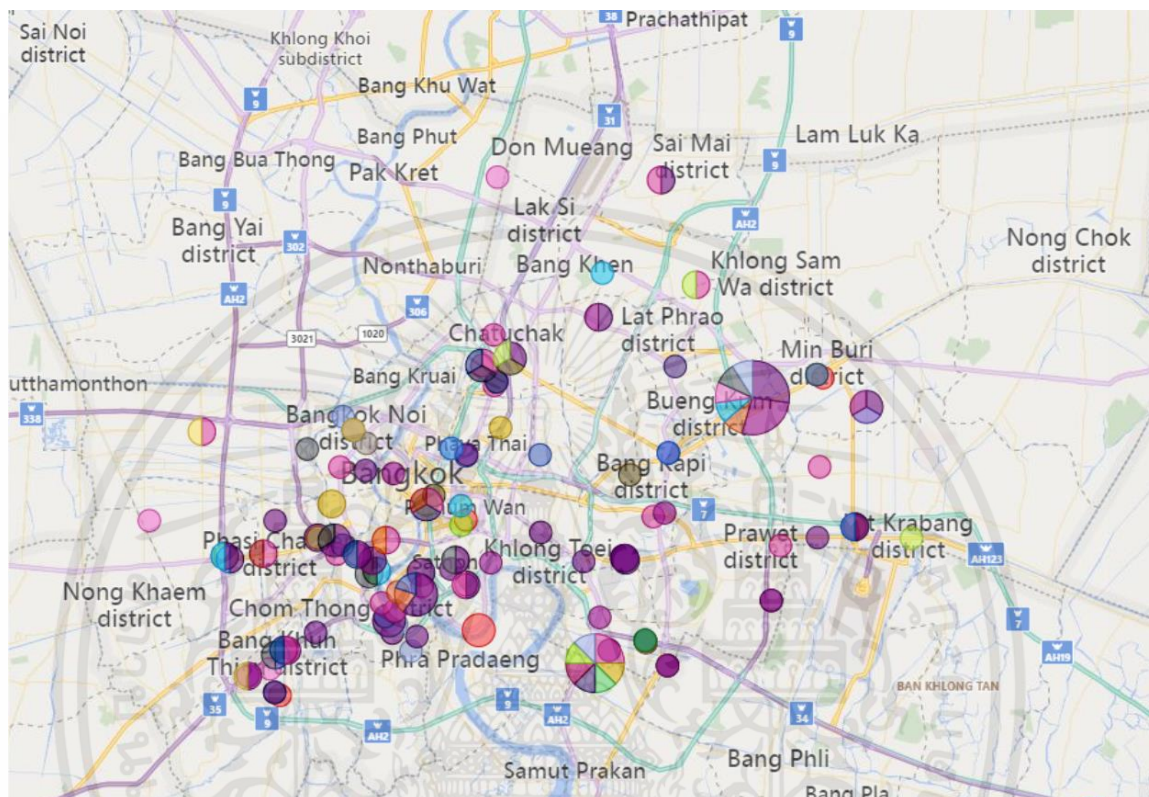


รูปที่ 4.1 รูปแสดงจุด cluster ของจุดรับส่งผู้โดยสาร (pickup points)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 4.1.2 การจัดกลุ่มข้อมูลจุดส่งผู้โดยสาร (dropoff points)

จัดกลุ่มข้อมูลจุดส่งผู้โดยสาร (dropoff points) ตามความหนาแน่น ได้ผลลัพธ์เป็น จุดคลัสเตอร์ (cluster) ส่งผู้โดยสารยอดนิยม ดังรูปที่ 4.2



รูปที่ 4.2 รูปแสดงจุด cluster ของจุดส่งผู้โดยสาร (dropoff points)

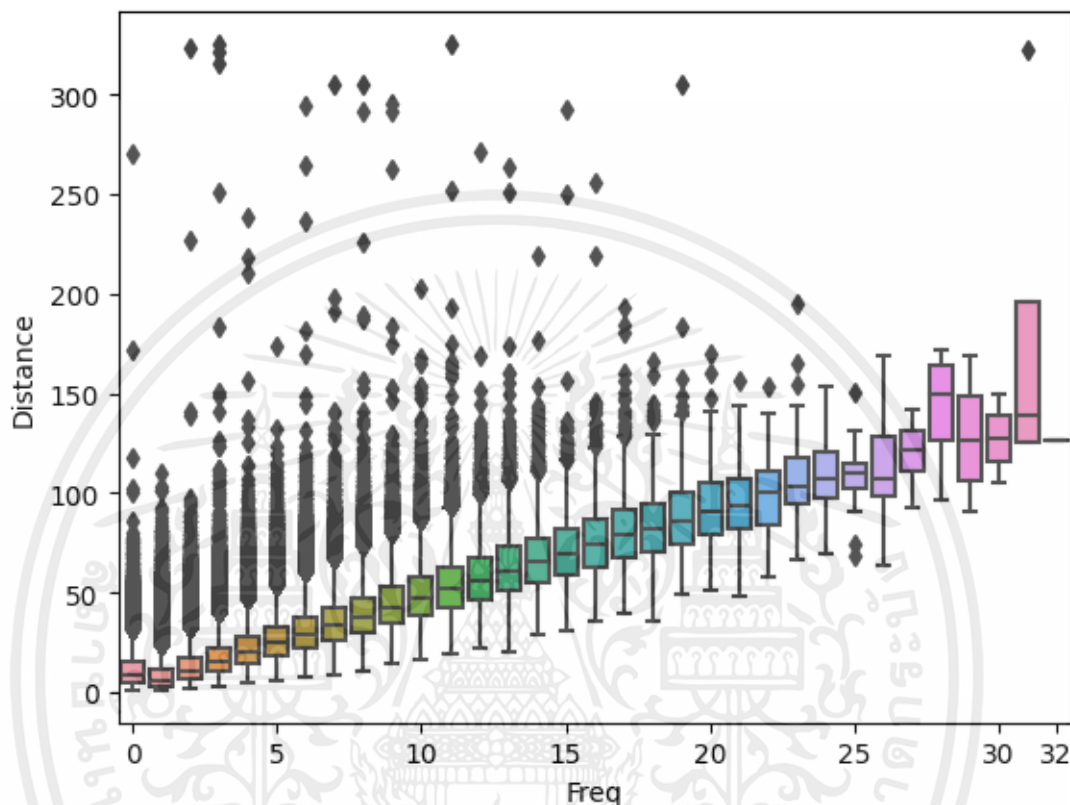
## 4.2 การหาความสัมพันธ์ของข้อมูล

### 4.2.1 หาความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอดนิยมกับระยะทาง (Distance)

จากข้อมูลทั้งสองนี้ จะทำการหาความสัมพันธ์ของข้อมูล วิเคราะห์ความสัมพันธ์ของจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) และระยะทาง (Distance) โดยดูทิศทางของกราฟว่าเคลื่อนที่ไปทางไหน ถ้ากราฟเคลื่อนที่ขึ้น แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับระยะทาง (Distance) มีความสัมพันธ์กันเชื่อมโยงกันในทิศทางเดียวกัน แต่ถ้ากราฟเคลื่อนที่ลง

แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับระยะทาง (Distance) ไม่มีความสัมพันธ์กันหรือมีความสัมพันธ์ที่แตกต่างกันในทิศทางตรงกันข้าม

ได้ผลลัพธ์เป็นแผนภูมิรูปกล่อง (Box Plot) ดังรูปที่ 4.3



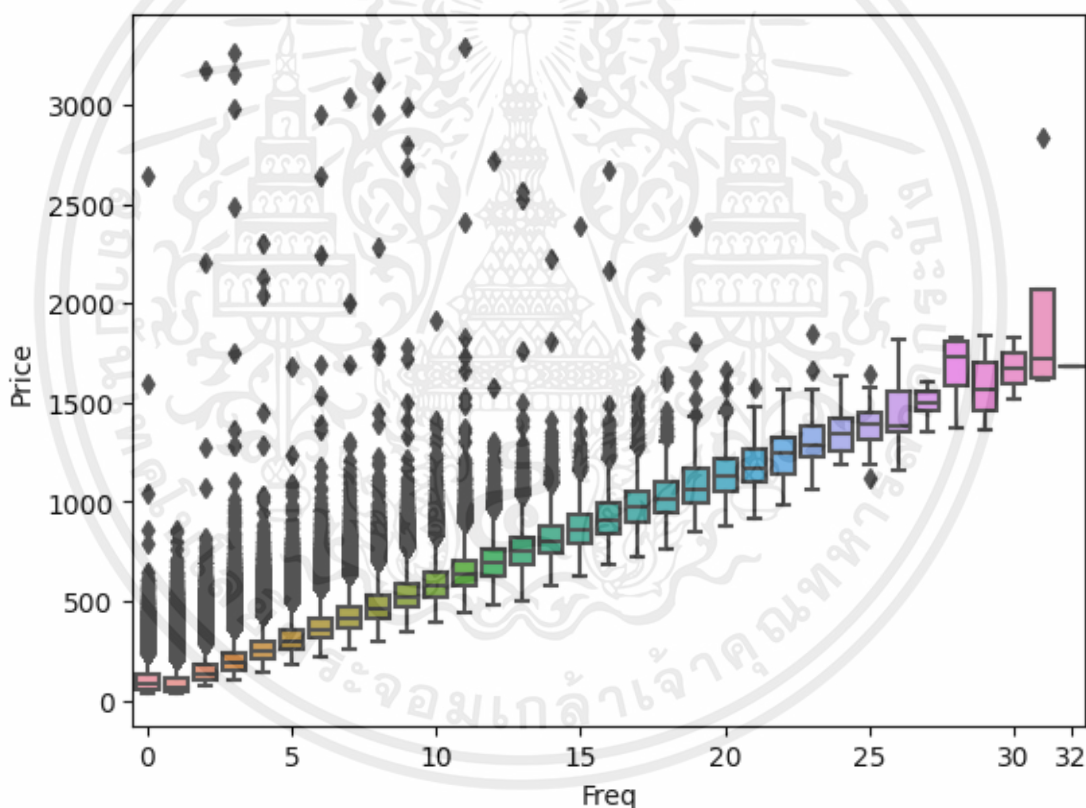
รูปที่ 4.3 รูปแสดงกราฟรูปแผนภูมิ (Box Plot) แสดงความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอดนิยมกับระยะทาง (Distance)

จากกราฟรูปกล่อง (Box Plot) ในรูปที่ 4.3 สามารถแสดงให้เห็นกราฟเคลื่อนที่ขึ้น แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับระยะทาง (Distance) มีความสัมพันธ์กัน เชื่อมโยงกันในทิศทางเดียวกัน นั่นหมายความว่า สมมติฐานที่ว่าแท็กซี่มีการรับผู้โดยสาร ณ จุดที่มีจุดรับผู้โดยสารยอดนิยม จะมีประสิทธิภาพในการรับผู้โดยสารต่อวันมากกว่า เป็นความจริง

#### 4.2.2 หาคความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอดนิยกับค่าโดยสาร (Price)

จากข้อมูลทั้งสองนี้ จะทำการหาคะสัมพันธ์ของข้อมูล วิเคราะห์ความสัมพันธ์ของจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) และค่าโดยสาร (Price) โดยดูทิศทางของกราฟว่าเคลื่อนที่ไปทางไหน ถ้ากราฟเคลื่อนที่ขึ้น แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับค่าโดยสาร (Price) มีความสัมพันธ์กันเชื่อมโยงกันในทิศทางเดียวกัน แต่ถ้ากราฟเคลื่อนที่ลง แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับค่าโดยสาร (Price) ไม่มีความสัมพันธ์กันหรือมีความสัมพันธ์ที่แตกต่างกันในทิศทางตรงกันข้าม

ได้ผลลัพธ์เป็นแผนภูมิรูปกล่อง (Box Plot) ดังรูปที่ 4.4



รูปที่ 4.4 รูปแสดงกราฟรูปแผนภูมิ (Box Plot) แสดงความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอดนิยกับค่าโดยสาร (Price)

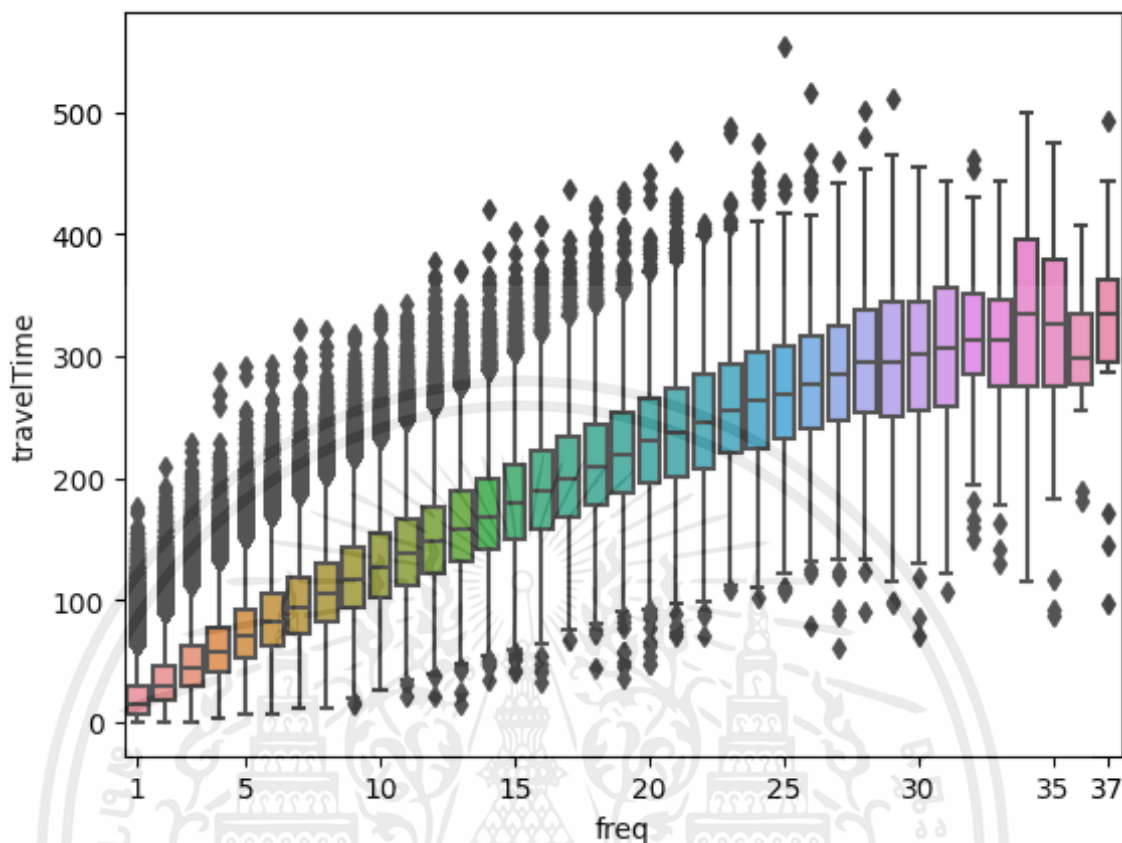
จากกราฟรูปกล่อง (Box Plot) ในรูปที่ 4.4 สามารถแสดงให้เห็นกราฟเคลื่อนที่ขึ้น แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับค่าโดยสาร (Price) มีความสัมพันธ์กัน

เชื่อมโยงกันในทิศทางเดียวกัน นั่นหมายถึงว่า สมมติฐานที่ว่าแท็กซี่มีการรับผู้โดยสาร ณ จุดที่มีจุดรับผู้โดยสารยอดนิยม จะมีประสิทธิภาพในการรับผู้โดยสารต่อวันมากกว่า เป็นความจริง

#### 4.2.3 หาความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอ นียมกับเวลาเดินทาง (travel time)

จากข้อมูลทั้งสองนี้ จะทำการหาความสัมพันธ์ของข้อมูล วิเคราะห์ความสัมพันธ์ของจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) และเวลาเดินทาง (travel time) โดยดูทิศทางของกราฟว่าเคลื่อนที่ไปทางไหน ถ้ากราฟเคลื่อนที่ขึ้น แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับเวลาเดินทาง (travel time) มีความสัมพันธ์กันเชื่อมโยงกันในทิศทางเดียวกัน แต่ถ้ากราฟเคลื่อนที่ลง แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับเวลาเดินทาง (travel time) ไม่มีความสัมพันธ์กันหรือมีความสัมพันธ์ที่แตกต่างกันในทิศทางตรงกันข้าม

ได้ผลลัพธ์เป็นแผนภูมิรูปกล่อง (Box Plot) ดังรูปที่ 4.5



รูปที่ 4.5 รูปแสดงกราฟรูปแผนภูมิ (Box Plot) แสดงความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอดนิยมกับเวลาเดินทาง (travel time)

จากกราฟรูปกล่อง (Box Plot) ในรูปที่ 4.5 สามารถแสดงให้เห็นกราฟเคลื่อนที่ขึ้น แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับเวลาเดินทาง (travel time) มีความสัมพันธ์กันเชื่อมโยงกันในทิศทางเดียวกัน นั่นหมายความว่า สมมติฐานที่ว่าแท็กซี่มีการรับผู้โดยสาร ณ จุดที่มีจุดรับผู้โดยสารยอดนิยม จะมีประสิทธิภาพในการรับผู้โดยสารต่อวันมากกว่า เป็นความจริง

### 4.3 การแนะนำจุดรับผู้โดยสารตามช่วงเวลา

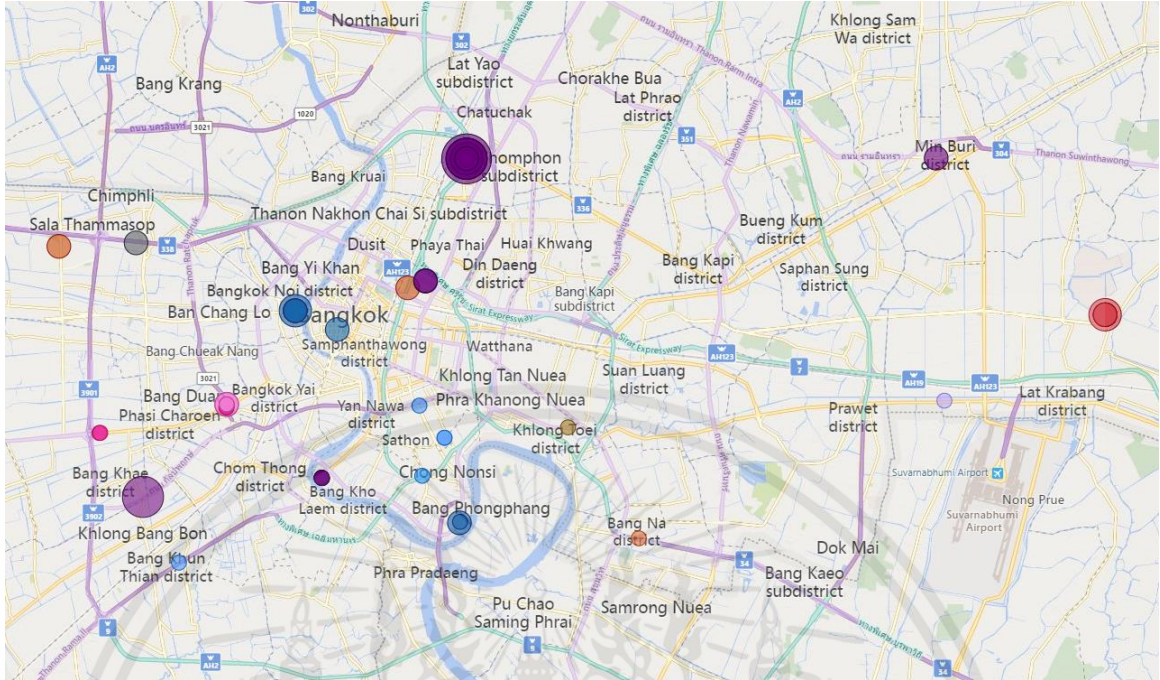
แสดงการแนะนำจุดรับผู้โดยสารตามช่วงเวลา (Time-based pickup points recommendations) ผลลัพธ์จากการวิเคราะห์ตามช่วงเวลาของข้อมูลเพื่อแนะนำจุดรับผู้โดยสาร (pickup points) ที่เหมาะสมสำหรับรถแท็กซี่ในช่วงเวลานั้นๆ โดยจากการศึกษาได้แบ่งช่วงเวลาที่น่าสนใจออกเป็น 4 ช่วง คือ

- 1) ช่วงเช้า-เที่ยง : 06:00-12:00 น.
- 2) ช่วงเที่ยง-เย็น : 12:00-16:00 น.
- 3) ช่วงเย็น-ค่ำ : 16:00-20:00 น.
- 4) ช่วงค่ำ-เช้า : 20:00-06:00 น.

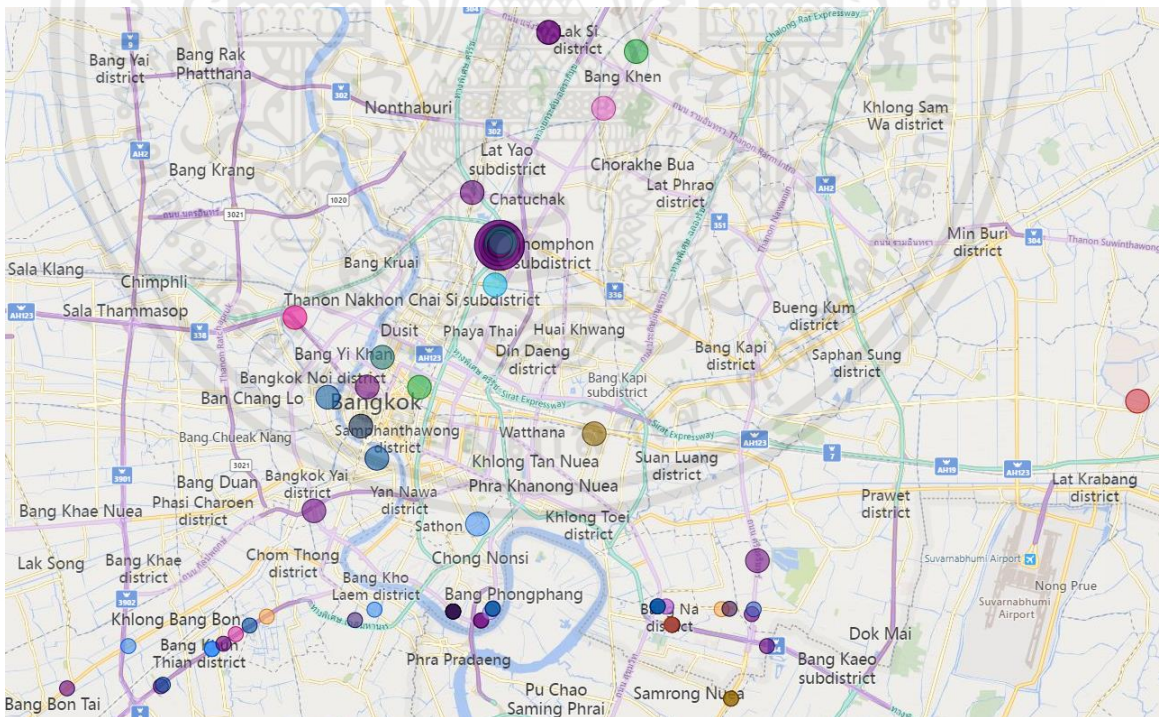
โดยทำการศึกษาตาม 4 ช่วงเวลานี้ และแบ่งวันออกเป็น 2 กลุ่ม

- 1) วันธรรมดา (weekday) : วันจันทร์-วันศุกร์
- 2) วันสุดสัปดาห์ (weekend) : วันเสาร์และวันอาทิตย์

จากการศึกษาได้ผลลัพธ์ของจุดรับผู้โดยสาร (pickup points) ในวันธรรมดา (weekday) เป็นตามรูปที่ 4.6-4.9 และได้ผลลัพธ์ของจุดรับผู้โดยสาร (pickup points) ในวันสุดสัปดาห์ (weekend) เป็นตามรูปที่ 4.10-4.13 ซึ่งแต่ละสีแสดงถึงจุดคลัสเตอร์ (clusterID) ที่ต่างกัน และขนาดของรัศมีจะผันตรงกับจำนวนของรถแท็กซี่ในจุดคลัสเตอร์ (clusterID) นั้นๆ

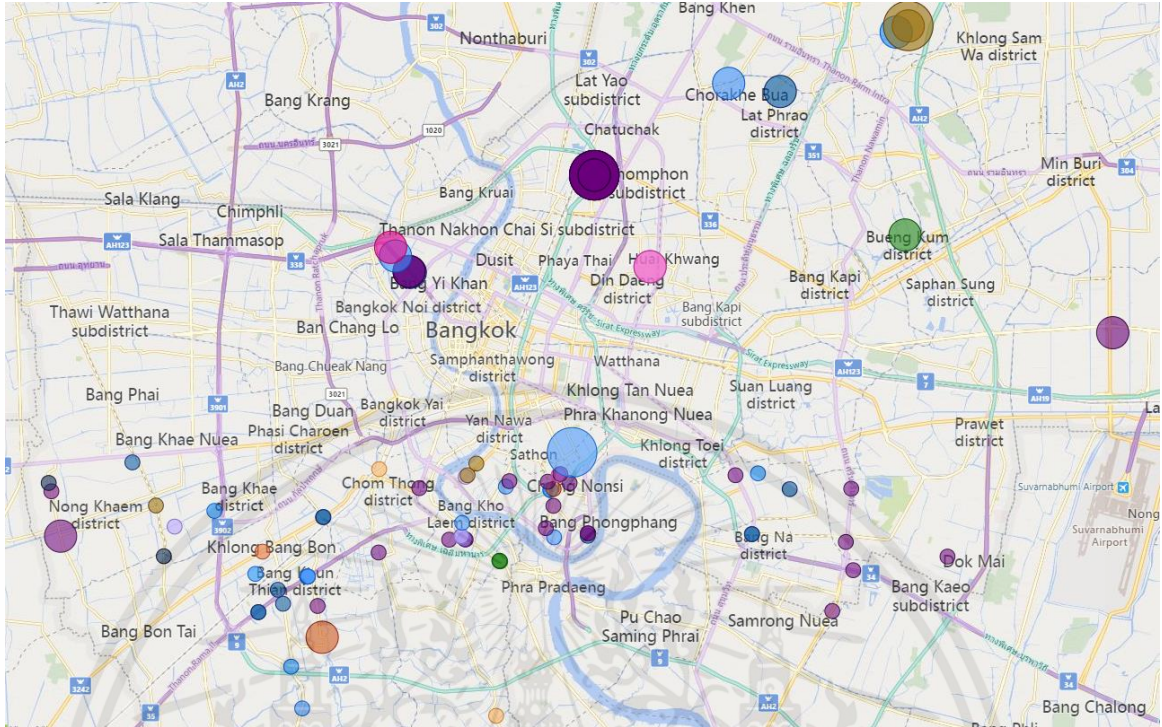


รูปที่ 4.6 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 06:00-12:00 ในวันธรรมดา (weekday)

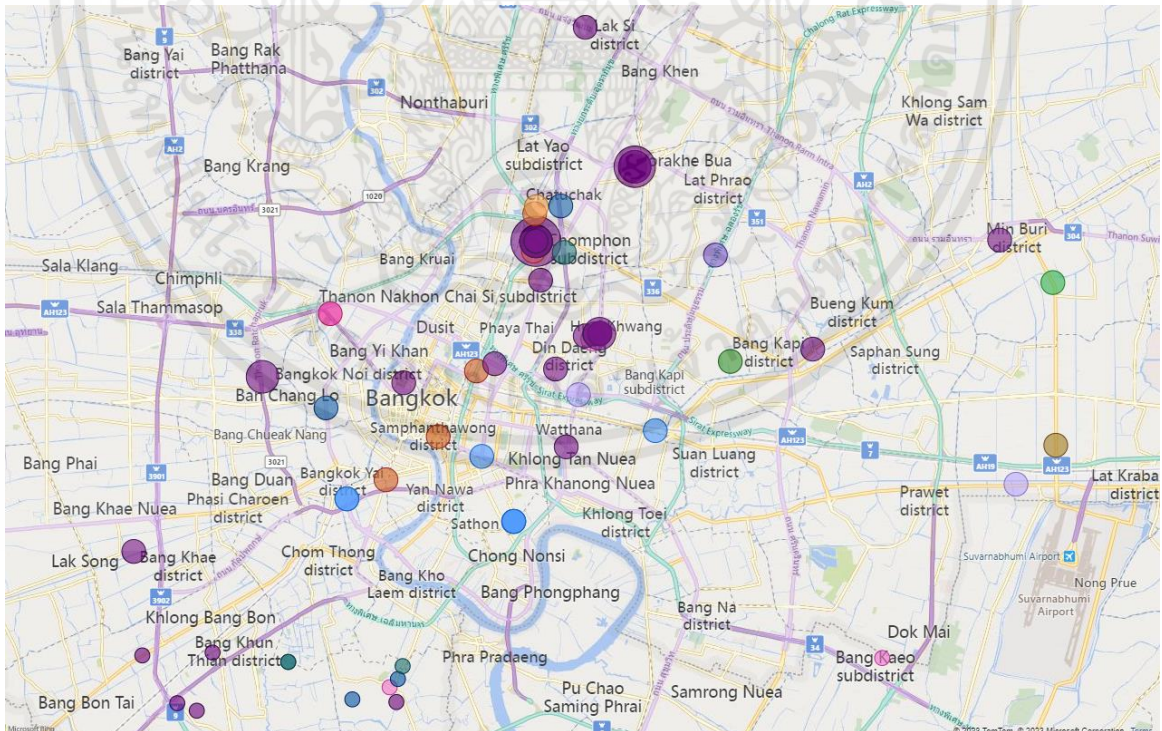


รูปที่ 4.7 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 12:00-16:00 ในวันธรรมดา (weekday)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

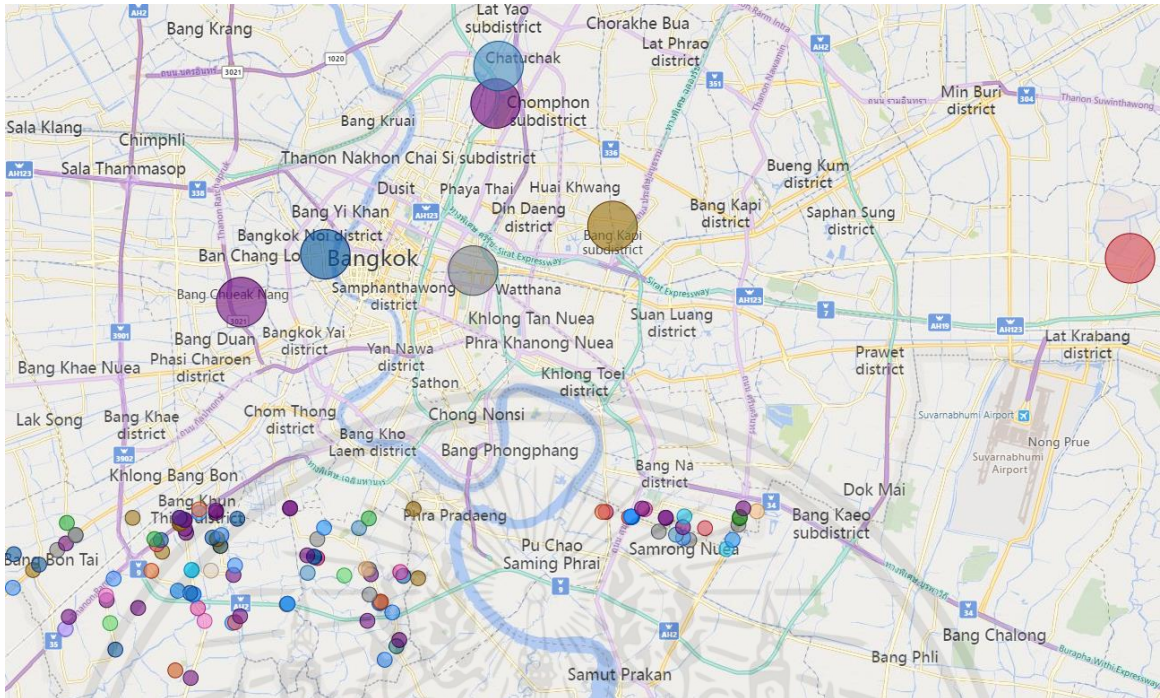


รูปที่ 4.8 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 16:00-20:00 ในวันธรรมดา (weekday)

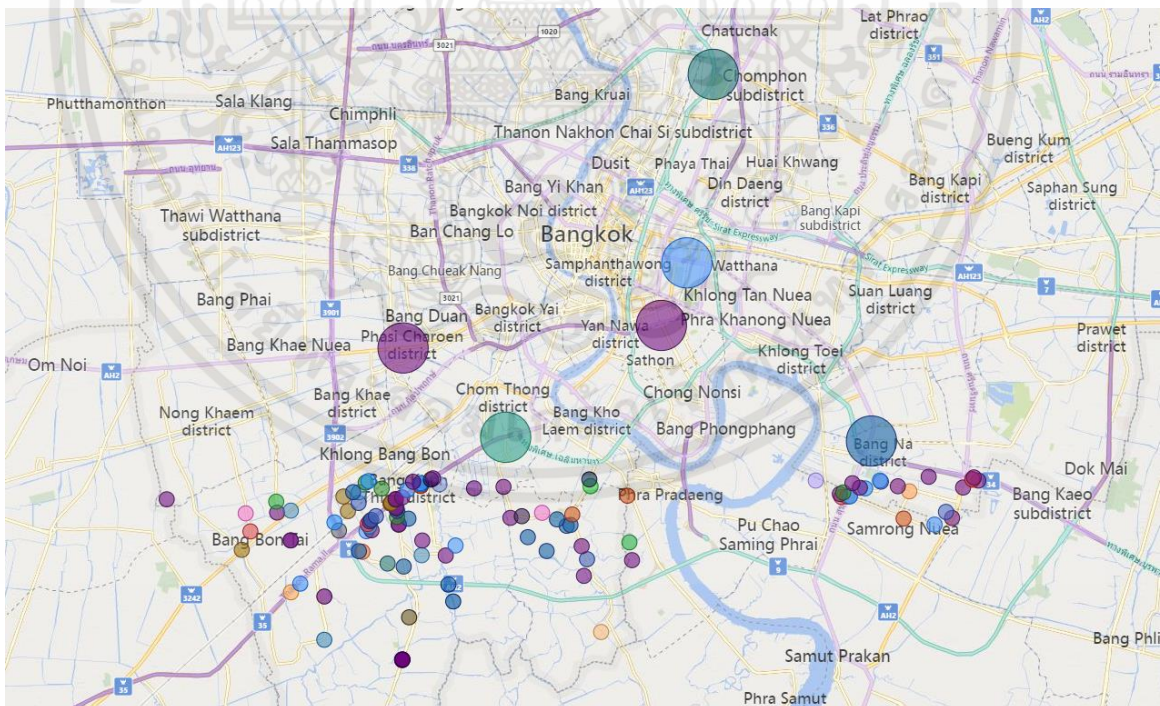


รูปที่ 4.9 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 20:00-06:00 ในวันธรรมดา (weekday)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

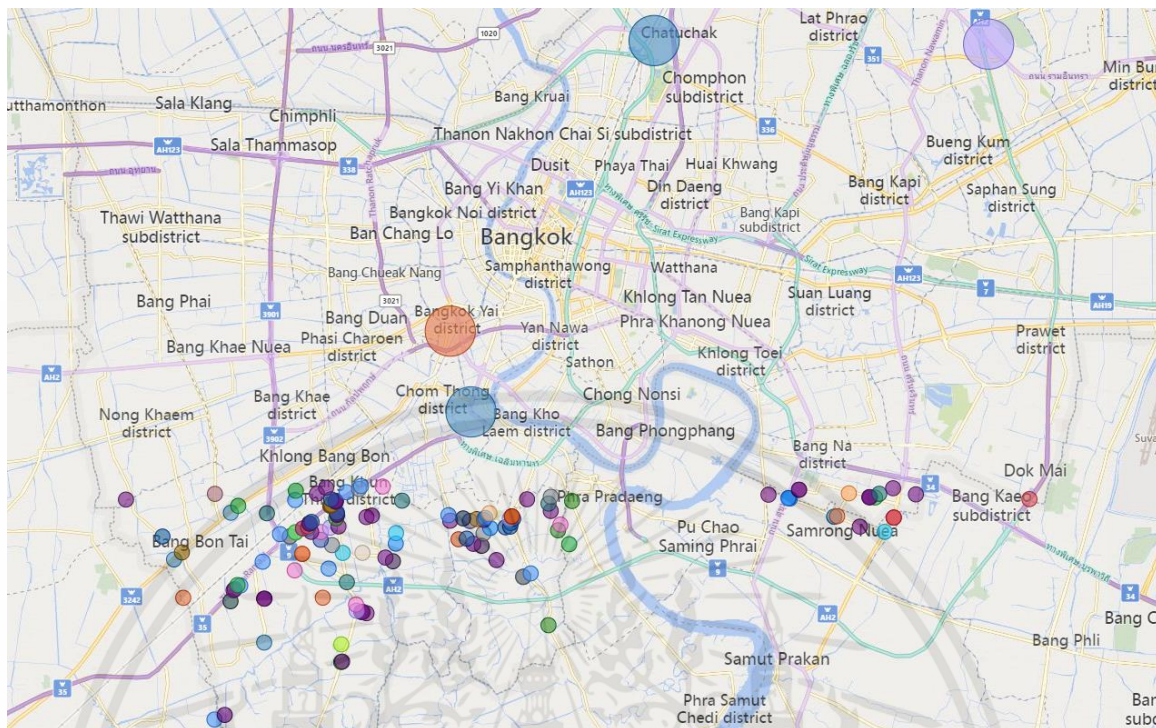


รูปที่ 4.10 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 06:00-12:00 ในวันสุดสัปดาห์ (weekend)

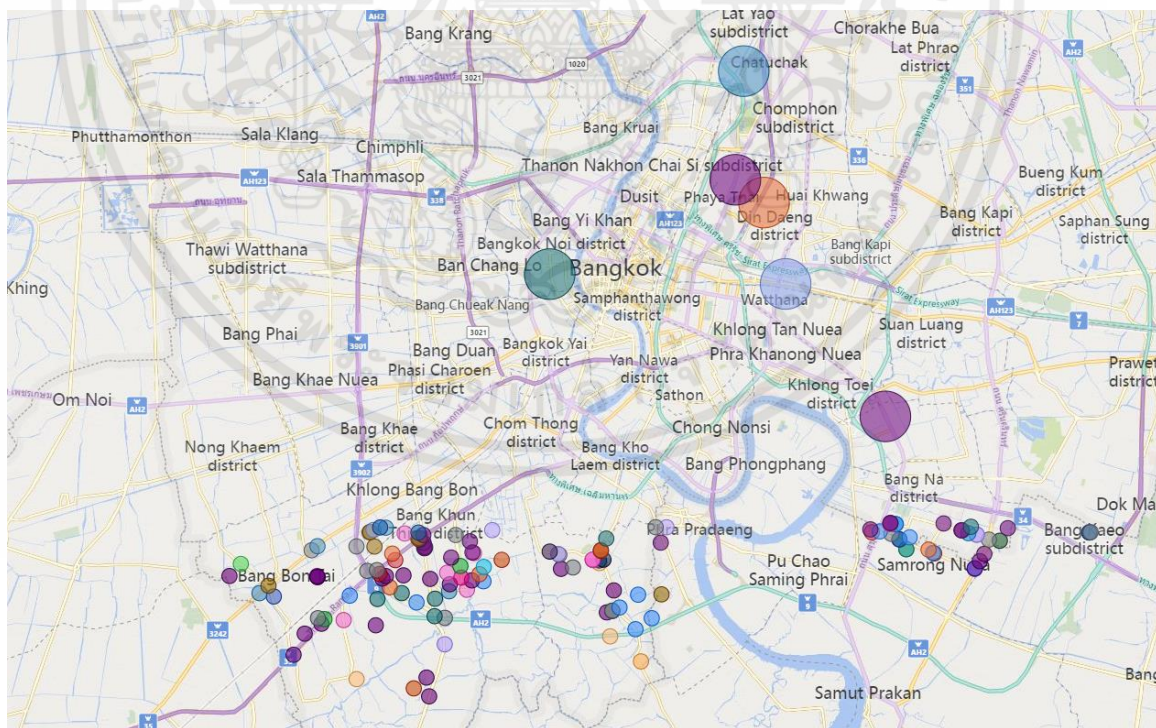


รูปที่ 4.11 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 12:00-16:00 ในวันสุดสัปดาห์ (weekend)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



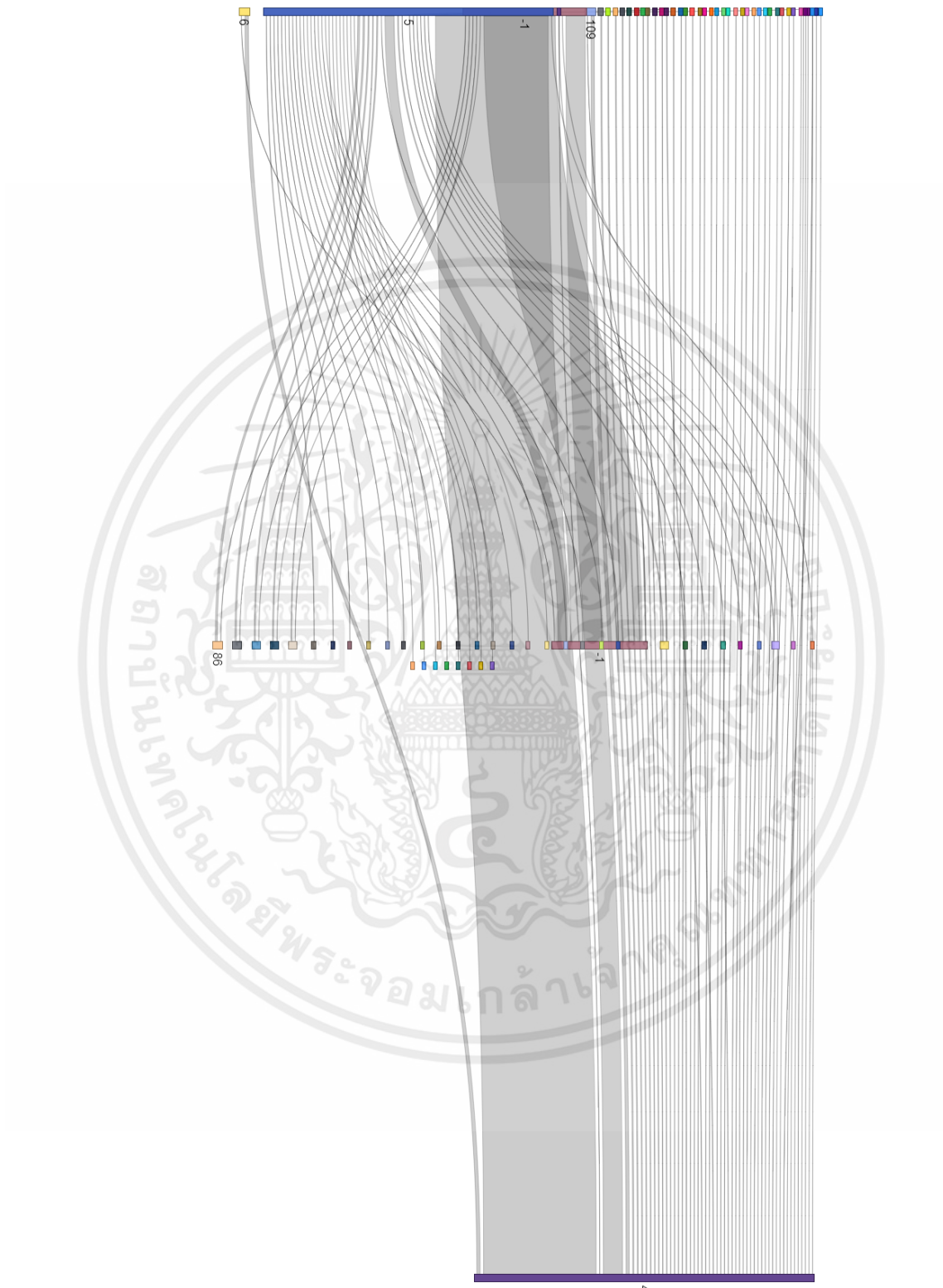
รูปที่ 4.12 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 16:00-20:00 ในวันสุดสัปดาห์ (weekend)



รูปที่ 4.13 รูปแสดงจำนวนการรับผู้โดยสาร (pickup) เวลา 20:00-06:00 ในวันสุดสัปดาห์ (weekend)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.4 ผลลัพธ์จากการศึกษาอื่นๆ

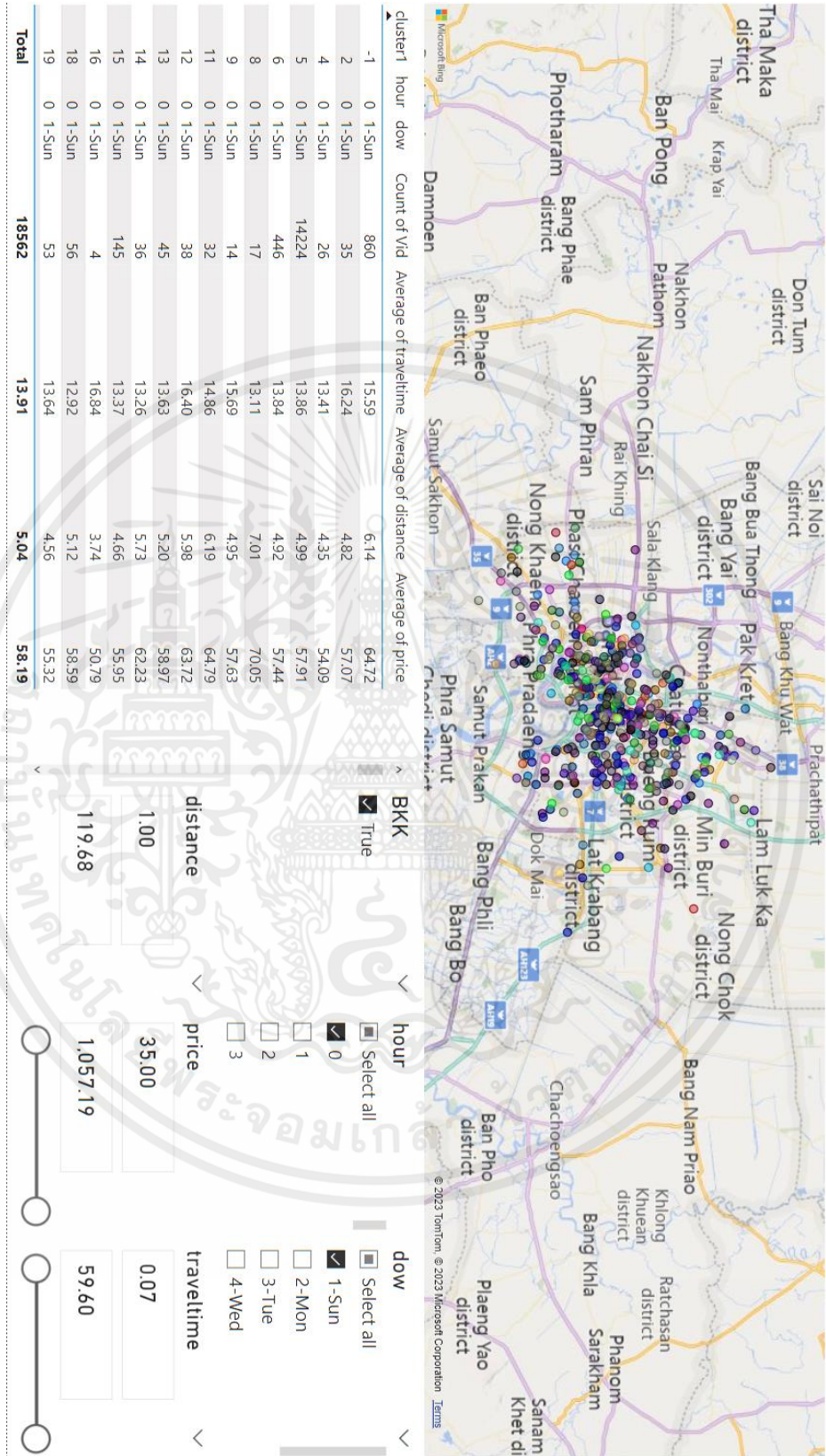


รูปที่ 4.14 รูปแสดงการเดินทางระหว่างจุดคลัสเตอร์ (clusterID) ของรถแท็กซี่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้







รูปที่ 4.1 การแสดงผลการศึกษางานวิจัยผ่านไมโครซอฟท์เพาเวอร์บีไอ (Microsoft Power BI)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 5

### สรุปผลและปัญหาของโครงการ

ในบทนี้จะกล่าวถึงผลสรุปจากการพัฒนาโครงการ เป็นส่วนสรุปผลของการพัฒนาทั้งหมดที่ทางผู้จัดทำได้ทำการทดลอง ปัญหาที่พบระหว่างการทำงาน แนวทางการแก้ปัญหา และแนวทางพัฒนาต่อยอดโครงการนี้ต่อไป

#### 5.1 สรุปผลการดำเนินงาน

สิ่งที่โครงการนี้ได้ทำการพัฒนาขึ้น แบ่งออกเป็น 3 ส่วน ดังนี้

##### 5.1.1 การเตรียมข้อมูล

- ข้อมูลตั้งต้นที่ใช้ในการศึกษาเป็นข้อมูลเปิดของมูลนิธิศูนย์ข้อมูลจราจรอัจฉริยะไทย (iTIC Open Data Archives) ซึ่งเป็นข้อมูลการเดินทางของรถแท็กซี่ในประเทศไทยปี 2020
- จากการที่ผู้จัดทำได้ทำการคัดแยกข้อมูลให้อยู่ในกรุงเทพมหานครนั้นพบว่า จากข้อมูลจำนวน 1,222,450,816 แถว สามารถคัดแยกลงมาเหลือ 718,658,390 แถว ในขณะเดียวกันจำนวนรถแท็กซี่ทั้งประเทศจาก 28,231 คัน คัดแยกได้เหลือ 27,510 คัน
- จากวัตถุประสงค์ที่ต้องการแนะนำจุดที่แท็กซี่ควรไปรอรับผู้โดยสารตามช่วงเวลาต่างๆ และพิสูจน์สมมติฐานที่ว่าแท็กซี่มีการรับผู้โดยสาร ณ จุดที่มีความหนาแน่นของผู้โดยสารมาก จะมีประสิทธิภาพในการรับผู้โดยสารต่อวันมากกว่า
- ผู้จัดทำจึงทำการคัดแยกข้อมูลรับ-ส่งผู้โดยสาร ด้วยข้อมูลสัญญาณไฟผู้โดยสารของรถแท็กซี่ที่มีการเปลี่ยนแปลง และคำนวณหาระยะทาง ค่าโดยสารและเวลาเดินทางจากข้อมูล ละติจูด (latitude) ลองจิจูด (longitude) และเวลา (timestamp) เพื่อใช้ในการวิเคราะห์ข้อมูลต่อไป

### 5.1.2 การวิเคราะห์ข้อมูล

- หลังจากเตรียมข้อมูลเพื่อการศึกษาแล้ว ผู้จัดทำได้นำข้อมูลการรับ-ส่งผู้โดยสารมาทำการจัดกลุ่มข้อมูลด้วย ดีบีเอสแกน (DBSCAN Clustering) เพื่อหาจุดรับ-ส่งผู้โดยสารยอดนิยม ที่สามารถอ้างอิงถึงความหนาแน่นของผู้โดยสารในแต่ละจุดได้ ซึ่งวิเคราะห์ออกมาเป็น 300 จุดรับ-ส่งยอดนิยมในกรุงเทพมหานคร
- หลังจากได้จุดที่ต้องการศึกษาแล้ว ได้ทำการหาความสัมพันธ์ของข้อมูลการรับ-ส่งผู้โดยสารกับ จำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) เวลาเดินทาง (travel time) ระยะทาง (distance) รวมถึงความแตกต่างของข้อมูลในแต่ละชั่วโมงและในแต่ละวัน

### 5.1.3 ผลลัพธ์จากการศึกษา

- การจัดกลุ่มข้อมูล ได้ ข้อมูลจุดรับผู้โดยสาร (pickup points) และจุดส่งผู้โดยสาร (dropoff points) ตามความหนาแน่น ได้ผลลัพธ์เป็นจุดคลัสเตอร์ (clusterID) รับ-ส่งผู้โดยสารยอดนิยม
- การหาความสัมพันธ์ของข้อมูลระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอดนิยมกับเวลาเดินทาง (travel time) ได้ผลลัพธ์ว่า หากจำนวนการผ่านจุดรับผู้โดยสารมาก จะมีเวลาเดินทางที่มากขึ้นตามไปด้วย ซึ่งสามารถตอบสมมติฐานที่ว่าแท็กซี่มีการรับผู้โดยสาร ณ จุดที่มีความหนาแน่นของผู้โดยสารมาก จะมีประสิทธิภาพในการรับผู้โดยสารต่อวันมากกว่าได้
- การหาความสัมพันธ์ของข้อมูลระยะทาง (distance) และเวลาเดินทาง (traveltime) รวม ในแต่ละจุดคลัสเตอร์ (clusterID) สามารถแนะนำการเดินทางของรถแท็กซี่ได้ว่า จุดคลัสเตอร์ (clusterID) ไหนมีความหนาแน่นของการรับผู้โดยสาร (pickup) มากกว่าตามข้อมูลข้างต้น

## 5.2 ปัญหา อุปสรรค และแนวทางการแก้ปัญหา

ปัญหาที่เกิดขึ้นระหว่างการทำงานของโครงการ มีดังนี้

### 5.2.1 ไม่ทราบจำนวนผู้โดยสารที่รอ

เนื่องจากข้อมูลเป็นข้อมูลการรับส่งของรถแท็กซี่ทำให้ไม่สามารถทราบจำนวนผู้โดยสารที่รออยู่อย่างแน่ชัดได้ แนวทางการแก้ปัญหาอาจต้องเป็นการลงพื้นที่สำรวจตามจุดต่าง ๆ

### 5.2.2 ไม่มีข้อมูลการเดินทางของรถแท็กซี่ทั้งหมด

ข้อมูลที่เรา มี นั้นมีเพียงข้อมูลเปิดของมูลนิธิศูนย์ข้อมูลจราจรอัจฉริยะไทย (iTIC Open Data Archives) ซึ่งไม่ได้มีข้อมูลของรถแท็กซี่ทั้งหมด

### 5.2.3 กลุ่มข้อมูลที่วิเคราะห์ยังไม่สามารถแยกแยะทิศทางการวิ่งของรถแท็กซี่ได้

cluster ที่วิเคราะห์ออกมานั้นยังไม่สามารถแยกแยะทิศทางการวิ่งได้ อย่างเช่น ฟังโหน เข้าเมืองหรือฟังโหนออกเมือง หรือฟังโหนเป็นฟังโหนของถนน

## 5.3 แนวทางการพัฒนาต่อ

ในส่วนของแนวทางการพัฒนาโครงการต่อไป มีดังนี้

### 5.3.1 ลงพื้นที่สำรวจการรอรถแท็กซี่ตามตำแหน่งต่าง ๆ

5.3.2 หาข้อมูลการเดินทางของรถแท็กซี่ที่นอกเหนือจากข้อมูลเปิดของมูลนิธิศูนย์ข้อมูลจราจรอัจฉริยะไทย (iTIC Open Data Archives)

5.3.3 ค้นคว้ากระบวนการจัดกลุ่มข้อมูล (clustering algorithm) ที่แยกแยะกลุ่มรถแท็กซี่แต่ละฝั่งถนนได้

## บรรณานุกรม

Kong, et al. (2017). **Time-Location-Relationship-Combined-Service-Recommendation-Based-on-Taxi-Trajectory-Data**. [online]

Available : <https://ieeexplore.ieee.org/document/7880635>

Zheng, et al. (2017). **Mining-trip-attractive-areas-using-large-scale-taxi**. [online]

Available : <https://ieeexplore.ieee.org/document/8367413>

Kavepol Khunsri, et al. (2021). **A Big Data Analysis on Efficiency of Bangkok Taxi System**. [online]

Available : <https://ieeexplore.ieee.org/document/9454833>

Debomit Dey. (2022). **DBSCAN Clustering in ML | Density based clustering**. [online]

Available : <https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/>

Wikipedia. (2022). **Haversine Formula**. [online]

Available : [https://en.wikipedia.org/wiki/Haversine\\_formula](https://en.wikipedia.org/wiki/Haversine_formula)

OpenStreetMap. (2022). **แผนที่กรุงเทพมหานครจากข้อมูลพิกัดพื้นที่ (polygon)**. [online]

Available : <http://polygons.openstreetmap.fr/>

Narrative Knowledge Base. (2023). **What is a location polygon?**. [online]

Available : <https://kb.narrative.io/what-is-a-location-polygon>

iTIC Open Data Archives. (2023). **Historical raw vehicles and mobile probes data in Thailand.**  
[online]

Available : <https://itic.longdo.com/opendata/probe-data/>

กระทรวงคมนาคม. (2023). **ประกาศกระทรวงคมนาคม เรื่อง กำหนดอัตราค่าจ้างบรรทุกคนโดยสาร**  
[online]

Available : <http://taxi.ml.ac.th/2566.pdf>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ภาคผนวก



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## การวิเคราะห์ข้อมูลจากการเดินทางของรถแท็กซี่

### DATA ANALYTICS ON TAXI TRAJECTORIES

นายคาณ ชัยวงศ์ศรีอรุณ นายธนพล คำเพราะ ผศ.ดร.รัฐชัย ชาวอุทัย

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

#### 1. บทนำ

รูปแบบการเดินทางเป็นพฤติกรรมที่เกิดขึ้นจากการสังเคราะห์ปัจจัยต่าง ๆ ที่เกี่ยวข้อง ซึ่งการเข้าใจรูปแบบการเดินหน้านั้นเป็นสิ่งสำคัญอย่างมาก เนื่องจากมีการใช้งานอย่างแพร่หลายในหลากหลายรูปแบบ แต่เนื่องจากข้อจำกัดของเทคโนโลยีในการรับข้อมูล การศึกษาในหัวข้อนี้มักจะใช้ข้อมูลการสำรวจที่มีข้อจำกัดหลายประการ เช่น ความแม่นยำต่ำ และค่าใช้จ่ายสูง อย่างไรก็ตาม ด้วยการพัฒนาเทคโนโลยีการประมวลผลแบบระบุตำแหน่งอย่างรวดเร็ว การเก็บข้อมูลการเดินทางก็ง่ายขึ้นมากขึ้น แต่ก็ยังคงมีเรื่องของความเป็นส่วนตัวเป็นปัจจัยที่ต้องพิจารณา ดังนั้น การศึกษาในหัวข้อนี้มักจะอ้างอิงจากข้อมูลการขนส่งสาธารณะ เช่น ข้อมูลการเดินทางของรถแท็กซี่ ข้อมูลธุรกรรมสมาร์ตการ์ดสำหรับรถไฟฟ้าใต้ดิน และรถประจำทาง ในบทความนี้ งานวิจัยใช้ข้อมูลการเดินทางของรถแท็กซี่

การใช้แท็กซี่เป็นวิธีการเดินทางที่สำคัญอย่างหนึ่งในกรุงเทพมหานคร ซึ่งแตกต่างจากการใช้รถประจำทางหรือรถไฟที่มีสายทางและสถานีที่กำหนดไว้ล่วงหน้า แท็กซี่สามารถให้บริการเดินทางส่วนบุคคลได้อย่างสะดวกและสามารถกำหนดเส้นทางได้ตามความต้องการของผู้ใช้บริการเอง พิกัดและเวลาเดินทางของเส้นทางแท็กซี่ถูกบันทึกไว้ซึ่งสามารถใช้งานเพื่อประเมินประสิทธิภาพของผู้ขับและความต้องการของผู้โดยสารได้ คนขับแท็กซี่ที่มีความเข้าใจเชิงลึกในการจราจรในเมืองมักพิจารณาหลายปัจจัย เช่น เวลาเดินทางและระยะทางของถนน เพื่อเลือกเส้นทางที่เหมาะสมที่สุด ดังนั้นข้อมูลการเดินทางของแท็กซี่สามารถสะท้อนถึงสถานะการจราจรได้ดี ข้อมูลต้นทาง-ปลายทางที่ได้รับจากการใช้แท็กซี่มากมายยังสามารถสะท้อนถึงความต้องการในการเดินทางและแนวโน้มของผู้คนได้อย่างชัดเจน

การศึกษานี้มุ่งเน้นการวิเคราะห์รูปแบบการเดินทางของรถแท็กซี่ โดยเฉพาะอย่างยิ่งการแยกแยะการเดินทางของผู้โดยสารเพื่อค้นหาจุดที่

แท็กซี่ควรไปรอผู้โดยสารตามช่วงเวลาต่างๆ เพื่อช่วยเหลือแท็กซี่ที่มีรายได้น้อย แท็กซี่ที่ใหม่หรือออกจากเส้นทางที่คุ้นเคย ทำให้ไม่รู้ว่าควรจะไปรับผู้โดยสารที่ไหน

เราศึกษาข้อมูลเส้นทางรถแท็กซี่ในกรุงเทพมหานคร ในปีค.ศ. 2020 และพบพื้นที่ที่น่าสนใจสำหรับการเดินทาง โดยเราจะแนะนำจุดที่ควรไปรอผู้โดยสารตามช่วงเวลาต่างๆ พร้อมการพิสูจน์ว่าการจับรถแท็กซี่ผ่านจุดรับส่งยอดนิยมนจะมีโอกาสได้รับผู้โดยสารมากกว่า อย่างไรก็ตาม ข้อมูลต้นฉบับนั้นจะมีปัญหาอยู่อย่างเช่น มีสัญญาณรบกวน มีค่าที่หายไป มีความไม่สอดคล้องกัน และมีความซ้ำซ้อนของข้อมูล ซึ่งจะส่งผลกระทบต่อประสิทธิภาพของอัลกอริทึม ดังนั้น เราจำเป็นต้องล้างข้อมูลและประมวลผลข้อมูลล่วงหน้าก่อนการทำการค้นหาข้อมูล

ในบทความนี้ เราจะแนะนำอัลกอริทึมการจัดกลุ่มข้อมูลของการเรียงลำดับตามความหนาแน่นของจุด DBSCAN (Density-Based Spatial Clustering of Applications with Noise) ที่มีความสามารถในการค้นหาข้อมูลที่มีความหลากหลาย และยังสามารถระบุจุดที่ไม่ได้รวมอยู่ในกลุ่มหรือ Noise ได้ด้วย นอกจากนี้ เรายังจะแนะนำแนวคิดที่เกี่ยวข้องของอัลกอริทึมและวิธีการใช้งานอย่างถูกต้อง

#### 2. บทวรรณกรรม

ในบทนี้จะกล่าวถึงรายละเอียดของข้อมูลที่ใช้ในโครงการนี้ เช่น รายละเอียดหรือความหมายของข้อมูลรถ รายละเอียดของงานวิจัยที่นำมาใช้อ้างอิงและเปรียบเทียบผลการทดลอง รวมถึงเทคนิคที่ใช้ที่เราได้ศึกษาค้นคว้าเพื่อนำมาประกอบงานวิจัยนี้ ซึ่งสามารถดูรายละเอียดได้ ดังนี้

##### 2.1. งานที่เกี่ยวข้อง

**- Time-Location-Relationship Combined Service Recommendation Based on Taxi Trajectory Data**

ในงานวิจัยนี้ได้มีการเสนอคำแนะนำบริการแท็กซี่โดยใช้แบบจำลองความสัมพันธ์เชิงเวลากับตำแหน่ง (Time-Location-Relationship: TLR) เพื่อช่วยจัดการกับปัญหาการที่แท็กซี่นั้นไม่มีผู้โดยสารและความยากลำบากในการหารถแท็กซี่สำหรับผู้โดยสาร ด้วยการบูรณาการ ประมวลผล และวิเคราะห์ข้อมูลเส้นทางเดินรถของแท็กซี่ จะสามารถระบุความสัมพันธ์เชิงปริมาณระหว่างการขึ้นและลงของผู้โดยสารในเขตพื้นที่เมืองได้และสามารถรับรูปแบบการเคลื่อนย้ายของผู้คนได้อย่างมีประสิทธิภาพ เช่น สามารถอนุมานได้ว่าคนส่วนใหญ่มักใช้เวลาประมาณ 3 ชั่วโมงในการเยี่ยมชมสถานที่ทางประวัติศาสตร์ ในช่วงเวลา 11.00 - 14.00 นาฬิกา

**- Mining trip attractive areas using large-scale taxi**

เนื้อหาการวิจัยหลักของบทความนี้ประกอบด้วย 3 ด้าน ได้แก่ การประมวลผลข้อมูลสำหรับข้อมูลเส้นทางรถแท็กซี่ อัลกอริทึมการจัดกลุ่มสำหรับพื้นที่ที่น่าสนใจในการเดินทาง และการทดลองที่เกี่ยวข้อง ซึ่งส่วนหลักของงานวิจัยนี้คือการเสนออัลกอริทึม (Algorithm) กริดบีสแกน (GRIDBSCAN) ที่เหมาะสำหรับข้อมูลขนาดใหญ่ ซึ่งจะให้ผลออกมาคล้ายกับตัว ดีบีสแกน (DBSCAN) แต่จะมีระยะเวลาในการคำนวณนั้นน้อยกว่ามาก

**2.2. เทคนิคที่ใช้**

**- ดีบีสแกน (DBSCAN)**

ดیبีสแกน (DBSCAN) คือ การจัดกลุ่มเชิงพื้นที่ตามความหนาแน่นของข้อมูลพร้อมด้วยข้อมูลรบกวน (Density-Based Spatial Clustering of Applications with Noise : DBSCAN) เป็นอัลกอริทึม (Algorithm) ในการจัดกลุ่มข้อมูล (Clustering) โดยใช้การจัดกลุ่มจุดของข้อมูลที่มีความหนาแน่นและอยู่ใกล้กันอยู่ด้วยกัน พร้อมด้วยการแยกจุดที่มีความหนาแน่นน้อยออกไป โดยจะมีพารามิเตอร์ (Parameter) ที่สำคัญอยู่ด้วยกัน 2 ตัว คือค่า eps และ ค่า MinPts (Min Points)

**- สูตรของฮาเวอร์ไซน์ (Haversine Formula)**

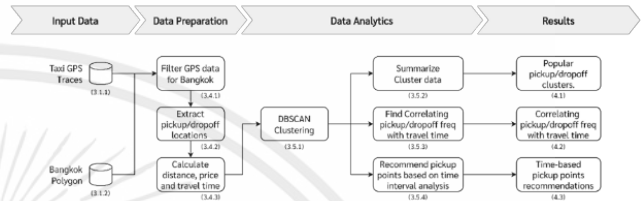
สูตรของฮาเวอร์ไซน์ (Haversine Formula) เป็นสูตรการคำนวณเพื่อใช้หาระยะทางที่สั้นที่สุดระหว่างจุดสองจุดบนพื้นผิวทรงกลมหรือ

พื้นผิวของโลก ซึ่งคำนวณผ่านค่าละติจูด (latitude) และลองจิจูด (longitude) ของจุดสองจุดที่แตกต่างกัน

**3. วิธีการวิเคราะห์ข้อมูล**

ในบทนี้จะเป็นส่วนของการอธิบายถึงขั้นตอนการพัฒนา ตั้งแต่ข้อมูลการเตรียมข้อมูล และการวิเคราะห์ข้อมูล

**3.1. ภาพรวมงานวิจัย**



รูปที่ 1 ภาพรวมการพัฒนางานวิจัย

ในส่วนของวิจัยของเรานั้นจะมีขั้นตอนในการพัฒนา 4 ส่วนด้วยกัน มีส่วนแรกคือการรวบรวมข้อมูลที่เกี่ยวข้องกับงานวิจัยของเรา ส่วนที่สองจะเป็นการเตรียมการข้อมูลให้พร้อมในการใช้งาน ส่วนที่สามจะเป็นการวิเคราะห์ข้อมูลเพื่อให้ได้ข้อมูลที่เราต้องการจะใช้ และส่วนสุดท้ายส่วนที่สี่จะเป็นผลลัพธ์สุดท้ายที่เราต้องการออกมา

**3.2. ข้อมูล**

ชุดข้อมูลการติดตาม GPS ของรถแท็กซี่ที่ใช้ในบทความนี้มาจากมูลนิธิศูนย์ข้อมูลจราจรอัจฉริยะไทย (iTIC Open Data Archives) เป็นข้อมูลของรถแท็กซี่ในประเทศไทย ในปี ค.ศ.2020 ซึ่งประกอบด้วยข้อมูลต่อไปนี้:

- 1) ข้อมูลรหัสประจำตัวของรถแต่ละคัน (VehicleID)
- 2) ข้อมูลแสดงสถานะของสัญญาณดาวเทียม (gpsvalid)
- 3) ละติจูด (latitude)
- 4) ลองจิจูด (longitude)
- 5) วันที่และเวลา(timestamp)
- 6) ความเร็วของรถแท็กซี่ (speed)
- 7) ทิศทางการเดินทางของรถยนต์(heading)
- 8) ค่าของไฟสัญญาณแสดงสถานะผู้โดยสาร (for\_hire\_light)
- 9) สถานะแสดงการทำงานของเครื่องยนต์ (engine\_acc)

### 3.3. การเตรียมข้อมูล

ในส่วนนี้จะเป็นการเตรียมข้อมูลเพื่อที่จะนำไปใช้ในงานวิจัยต่อไป

#### 3.3.1. เลือกพื้นที่เฉพาะกรุงเทพ

นำข้อมูลจีพีเอส (GPS) ของแท็กซี่และข้อมูลพื้นที่ (Polygon) ของกรุงเทพมหานคร มากรองข้อมูลโดยใช้พิกัดพื้นที่ (polygon) ของกรุงเทพมหานครมาใช้คัดกรองข้อมูลจีพีเอส (GPS) ของแท็กซี่ที่อยู่ในประเทศไทย ให้เหลือแค่ข้อมูลที่อยู่ในกรุงเทพมหานครเท่านั้น

#### 3.3.2. เตรียมข้อมูลในรูปการรับส่ง

กระบวนการในการหาจุดรับผู้โดยสาร (pickup points) และจุดส่งผู้โดยสาร (dropoff points) จากข้อมูลไฟแสดงสัญญาณรับผู้โดยสาร (for\_hire\_light) ของรถแท็กซี่ที่จะแตกต่างกันไปตามการเปลี่ยนแปลงจาก 0 เป็น 1 หรือ จาก 1 เป็น 0 โดยทั่วไปแล้ว ไฟแสดงสัญญาณรับผู้โดยสารในรถแท็กซี่จะถูกเปิดเมื่อรถพร้อมให้บริการ (หมายถึง ไม่มีผู้โดยสารอยู่ภายใน) หรือเมื่อมีผู้โดยสารอยู่ในรถแล้ว ซึ่งสามารถนำข้อมูลนี้มาคัดแยกข้อมูลการเดินทางได้

#### 3.3.3. คำนวณระยะทาง เวลา และค่าโดยสาร ในแต่ละการรับส่ง

##### - ระยะทาง (Distance)

ระยะทาง (distance) คิดจากระหว่างสองจุดบนพื้นผิวของโลกโดยใช้สูตรของฮาเวอร์ไซน์ (Haversine Formula) ซึ่งเป็นวิธีการคำนวณระยะทาง (distance) รอบวงกลมระหว่างสองจุดบนพื้นผิวของทรงกลม สูตรนี้นิยมใช้ในการนำทางและมีประโยชน์มากในการคำนวณระยะทาง (distance) ระหว่างสถานที่บนพื้นผิวของโลก สูตรนี้คำนึงถึงรัศมีของโลกและพิกัดละติจูดและลองจิจูดของสองจุด โดยใช้ฟังก์ชันตรีโกณมิติเพื่อคำนวณระยะทาง (distance) ระหว่างสองจุด โดย ระยะทาง (distance) ที่แสดงระยะทางระหว่างแต่ละคู่ของจุดในแต่ละแถวของข้อมูล ตั้งแต่จุดรับผู้โดยสาร (pickup points) ไปจนถึงจุดส่งผู้โดยสาร (dropoff points)

##### - เวลาเดินทาง (travel time)

เวลาเดินทาง (travel time) คือเวลาจากจุดรับผู้โดยสาร (pickup points) ไปยังจุดส่งผู้โดยสาร (dropoff points) โดยวิธีการจะคิดจากส่วนต่างระหว่างการประทับเวลา (timestamp) ของทั้งสองจุด

##### - ค่าโดยสาร (Price)

ค่าโดยสาร (price) จะคำนวณด้วยฟังก์ชัน ซึ่งเป็นการคำนวณค่าโดยสารค่าแท็กซี่ตามระยะทาง (distance) ที่เดินทางเป็นกิโลเมตร ค่า

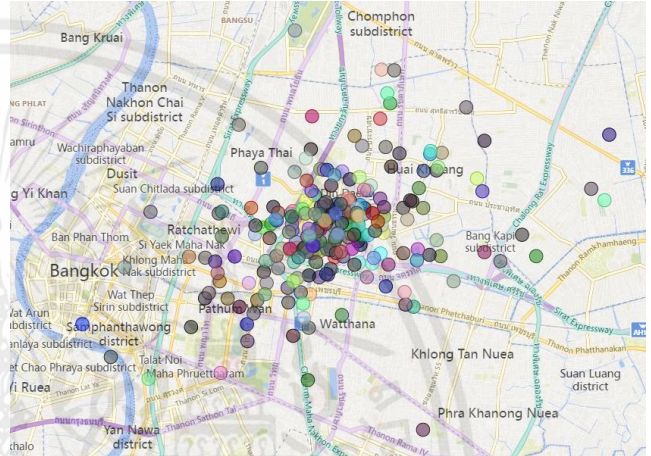
โดยสาร (price) จะคำนวณตามชุดของอัตราที่กำหนดไว้ล่วงหน้าสำหรับระยะทาง (distance) ต่างๆ

### 4. ผลการทดลอง

ในบทความนี้จะเป็นส่วนของผลการทดลองของงานวิจัย

#### 4.1. คัดสรรจัดรับส่งยอดนิยม

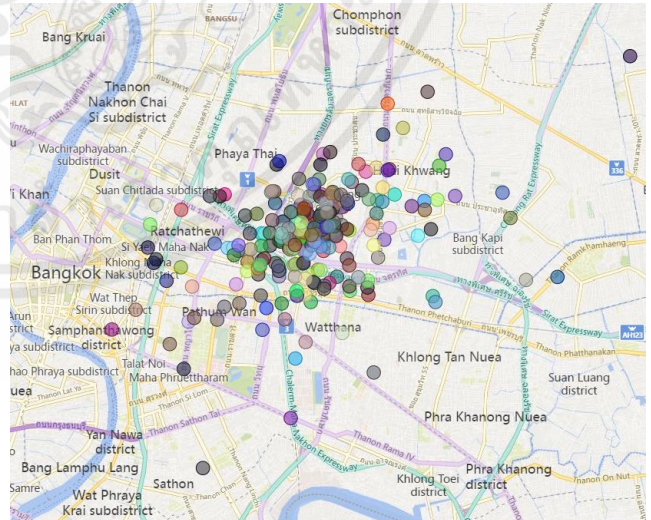
จัดกลุ่มข้อมูลจุดรับผู้โดยสาร (pickup points) ตามความหนาแน่น ได้ผลลัพธ์เป็นจุดคลัสเตอร์ (cluster) รับผู้โดยสารยอดนิยม ดังรูปที่ 2



รูปที่ 2 รูปแสดงจุด cluster ของจุดรับผู้โดยสาร (pickup points)

#### 4.2. การจัดกลุ่มข้อมูลจุดส่งผู้โดยสาร (dropoff points)

จัดกลุ่มข้อมูลจุดส่งผู้โดยสาร (dropoff points) ตามความหนาแน่น ได้ผลลัพธ์เป็นจุดคลัสเตอร์ (cluster) ส่งผู้โดยสารยอดนิยม ดังรูปที่ 3

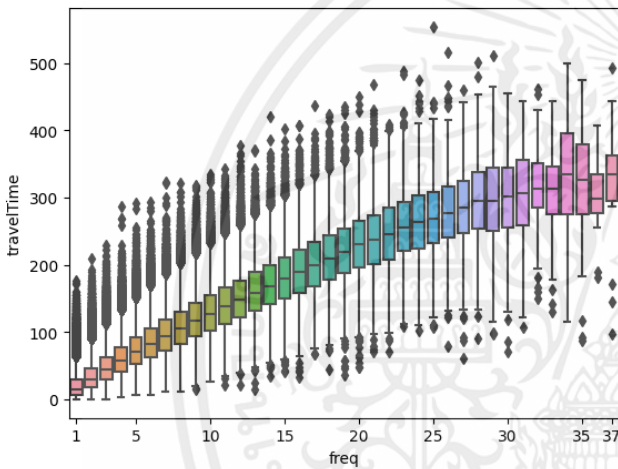


รูปที่ 3 รูปแสดงจุด cluster ของจุดส่งผู้โดยสาร (dropoff points)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**4.3. ความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับส่งยอดนิมกับปริมาณการรับผู้โดยสาร**

จากข้อมูลทั้งสองนี้ จะทำการหาความสัมพันธ์ของข้อมูล วิเคราะห์ความสัมพันธ์ของจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) และเวลาเดินทาง (travel time) โดยดูทิศทางของกราฟว่าเคลื่อนที่ไปทางไหน ถ้ากราฟเคลื่อนที่ขึ้น แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับเวลาเดินทาง (travel time) มีความสัมพันธ์กันเชื่อมโยงกันในทิศทางเดียวกัน แต่ถ้ากราฟเคลื่อนที่ลง แสดงว่าจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) กับเวลาเดินทาง (travel time) ไม่มีความสัมพันธ์กันหรือมีความสัมพันธ์ที่แตกต่างกันในทิศทางตรงกันข้าม



รูปที่ 4 รูปแสดงกราฟรูปแผนภูมิ (Box Plot) แสดงความสัมพันธ์ระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอดนิมกับเวลาเดินทาง (travel time)

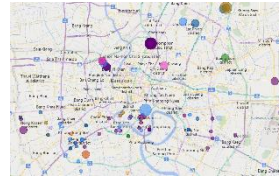
**4.4. ผลการแนะนำจุดรอรับผู้โดยสารตามช่วงเวลา**



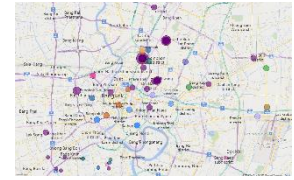
ก) การรับผู้โดยสารเวลา 06:00-12:00 ในวันธรรมดา



ข) การรับผู้โดยสารเวลา 12:00-16:00 ในวันธรรมดา



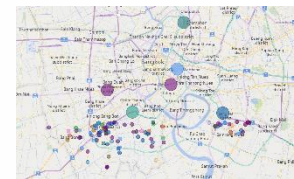
ค) การรับผู้โดยสารเวลา 16:00-20:00 ในวันธรรมดา



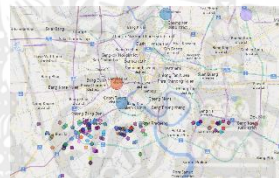
ง) การรับผู้โดยสารเวลา 20:00-06:00 ในวันธรรมดา



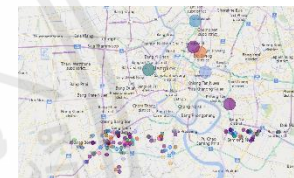
จ) การรับผู้โดยสารเวลา 06:00-12:00 ในวันสุดสัปดาห์



ฉ) การรับผู้โดยสารเวลา 12:00-16:00 ในวันสุดสัปดาห์



ช) การรับผู้โดยสารเวลา 16:00-20:00 ในวันสุดสัปดาห์



ซ) การรับผู้โดยสารเวลา 20:00-06:00 ในวันสุดสัปดาห์

**5. อภิปรายผลการทดลอง**

หากจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) มาก จะมีเวลาเดินทาง (travel time) มากขึ้นด้วย และสามารถสมมติฐานได้ว่าการรับผู้โดยสารที่จุดคนเยอะจะมีประสิทธิภาพมากกว่า นอกจากนี้ การหาความสัมพันธ์ของระยะทาง (distance) และเวลาเดินทาง (traveltime) ในแต่ละจุดคลัสเตอร์ (clusterID) สามารถแนะนำการเดินทางของรถแท็กซี่ได้ว่าจุดคลัสเตอร์ (clusterID) ไหนมีการรับผู้โดยสารมากกว่าจากการศึกษา

**6. บทสรุป**

สิ่งที่โครงการนี้ได้ทำการพัฒนาขึ้น แบ่งออกเป็น 3 ส่วน ดังนี้

**6.1. การเตรียมข้อมูล**

- ข้อมูลตั้งต้นที่ใช้ในการศึกษาเป็นข้อมูลเปิดของมูลนิธิศูนย์ข้อมูลจราจรอัจฉริยะไทย (iTIC Open Data Archives) ซึ่งเป็นข้อมูลการเดินทางของรถแท็กซี่ในประเทศไทยปี 2020

- จากการทำที่ผู้จัดทำได้ทำการคัดแยกข้อมูลให้อยู่ในกรุงเทพมหานครนั้นพบว่า จากข้อมูลจำนวน 1,222,450,816 แถว สามารถคัดแยกออกมาเหลือ 718,658,390 แถว ในขณะที่เดียวกัน จำนวนรถแท็กซี่ทั้งประเทศจาก 28,231 คัน คัดแยกได้เหลือ 27,510 คัน

- จากวัตถุประสงค์ที่ต้องการแนะนำจุดที่แท็กซี่ควรไปรอรับผู้โดยสารตามช่วงเวลาต่างๆ และพิสูจน์สมมติฐานที่ว่าแท็กซี่มีการรับผู้โดยสาร ณ จุดที่มีความหนาแน่นของผู้โดยสารมาก จะมีประสิทธิภาพในการรับผู้โดยสารต่อวันมากกว่า

- ผู้จัดทำจึงทำการคัดแยกข้อมูลรับ-ส่งผู้โดยสาร ด้วยข้อมูลสัญญาณไฟผู้โดยสารของรถแท็กซี่ที่มีการเปลี่ยนแปลง และคำนวณหาระยะทาง ค่าโดยสารและเวลาเดินทางจากข้อมูล ละติจูด (latitude) ลองจิจูด (longitude) และเวลา (timestamp) เพื่อใช้ในการวิเคราะห์ข้อมูลต่อไป

## 6.2. การวิเคราะห์ข้อมูล

- หลังจากเตรียมข้อมูลเพื่อการศึกษาแล้ว ผู้จัดทำได้นำข้อมูลการรับ-ส่งผู้โดยสารมาทำการจัดกลุ่มข้อมูลด้วย ดีบีเอสแกน (DBSCAN Clustering) เพื่อหาจุดรับ-ส่งผู้โดยสารยอดนิยม ที่สามารถอ้างอิงถึงความหนาแน่นของผู้โดยสารในแต่ละจุดได้ ซึ่งวิเคราะห์ออกมาเป็น 300 จุดรับ-ส่งยอดนิยมในกรุงเทพมหานคร

- หลังจากได้จุดที่ต้องการศึกษาแล้ว ได้ทำการหาความสัมพันธ์ของข้อมูลการรับ-ส่งผู้โดยสารกับ จำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) เวลาเดินทาง (travel time) ระยะทาง (distance) รวมถึงความแตกต่างของข้อมูลในแต่ละชั่วโมงและในแต่ละวัน

## 6.3. ผลลัพธ์จากการศึกษา

- การจัดกลุ่มข้อมูล ได้ ข้อมูลจุดรับผู้โดยสาร (pickup points) และจุดส่งผู้โดยสาร (dropoff points) ตามความหนาแน่น ได้ผลลัพธ์เป็นจุดคลัสเตอร์ (clusterID) รับ-ส่งผู้โดยสารยอดนิยม

- การหาความสัมพันธ์ของข้อมูลระหว่างจำนวนการผ่านจุดรับผู้โดยสาร (pickup frequency) ยอดนิยมกับเวลาเดินทาง (travel time) ได้ผลลัพธ์ว่า หากจำนวนการผ่านจุดรับผู้โดยสารมาก จะมีเวลาเดินทางที่มากขึ้นตามไปด้วย ซึ่งสามารถตอบสมมติฐานที่ว่าแท็กซี่มีการรับผู้โดยสาร ณ จุด

ที่มีความหนาแน่นของผู้โดยสารมาก จะมีประสิทธิภาพในการรับผู้โดยสารต่อวันมากกว่าได้

- การหาความสัมพันธ์ของข้อมูลระยะทาง (distance) และเวลาเดินทาง (traveltime) รวม ในแต่ละจุดคลัสเตอร์ (clusterID) สามารถแนะนำการเดินทางของรถแท็กซี่ได้ว่าจุดคลัสเตอร์ (clusterID) ไหนมีความหนาแน่นของการรับผู้โดยสาร (pickup) มากกว่าตามข้อมูลข้างต้น

## เอกสารอ้างอิง

Kong, et al. (2017). Time-Location-Relationship-Combined-Service-Recommendation-Based-on-Taxi-Trajectory-Data. [Online].

Available : <https://ieeexplore.ieee.org/document/7880635>

Zheng, et al. (2017). Mining-trip-attractive-areas-using-large-scale-taxi. [online]

Available : <https://ieeexplore.ieee.org/document/8367413>

Debommit Dey. (2022). DBSCAN Clustering in ML | Density based clustering. [online]

Available : <https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/>

Wikipedia. (2022). Haversine Formula. [online]

Available : [https://en.wikipedia.org/wiki/Haversine\\_formula](https://en.wikipedia.org/wiki/Haversine_formula)

OpenStreetMap. (2022). แผนที่กรุงเทพมหานครจากข้อมูลผู้กักพื้นที่ (polygon). [online]

Available : <http://polygons.openstreetmap.fr/>

Narrative Knowledge Base. (2023). What is a location polygon?. [online]

Available : <https://kb.narrative.io/what-is-a-location-polygon>

iTIC Open Data Archives. (2023). Historical raw vehicles and mobile probes data in Thailand. [online]

Available : <https://itic.longdo.com/opendata/probe-data/>



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้