

การวิเคราะห์รูปแบบการจรรยาบรรณบนถนนโดยใช้ข้อมูลเชิงพื้นที่จากซีดีอาร์และจีพีเอส

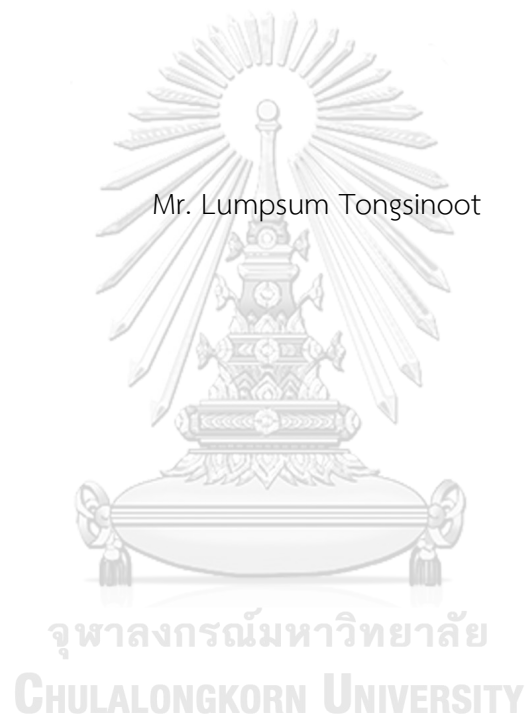


บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)  
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)  
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต  
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์  
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย  
ปีการศึกษา 2560  
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

ANALYSIS OF ROAD TRAFFIC PATTERNS USING CDR AND GPS



A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Engineering Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2017

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การวิเคราะห์รูปแบบการจลาจรบนถนนโดยใช้ข้อมูลเชิงพื้นที่จากซีดีอาร์และจีพีเอส
โดย	นายลำซำ ทองสีนุช
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร. วีระ เหมเมืองสิน

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโทบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์  
(รองศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร. เกริก ภิรมย์โสภา)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก  
(ผู้ช่วยศาสตราจารย์ ดร. วีระ เหมเมืองสิน)

..... กรรมการ  
(รองศาสตราจารย์ ดร. กุลธิดา โรจน์วิบูลย์ชัย)

..... กรรมการภายนอกมหาวิทยาลัย  
(รองศาสตราจารย์ ดร. วรเศรษฐ สุวรรณิก)

ล่าซ่า ทองสินุช : การวิเคราะห์รูปแบบการจราจรบนถนนโดยใช้ข้อมูลเชิงพื้นที่จากซีดีอาร์ และจีพีเอส (ANALYSIS OF ROAD TRAFFIC PATTERNS USING CDR AND GPS) อ.ที่  
 ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร. วีระ เหมืองสิน, 94 หน้า.

กรุงเทพมหานครเป็นเมืองหนึ่งที่มีความหนาแน่นของประชากรสูง การทราบถึงปริมาณความต้องการในการเดินทางระหว่างพื้นที่และเส้นทางที่ใช้ในการเดินทางจะเป็นเครื่องมือที่สำคัญที่ใช้ในการวางแผนระบบขนส่งภายในเมือง ในปัจจุบันมีอุปกรณ์ที่เก็บข้อมูลตำแหน่งของผู้ใช้อยู่มากมาย เช่น โปรแกรมบนโทรศัพท์มือถือ, จีพีเอสที่ติดบนรถแท็กซี่ (GPS), บันทึกการใช้งานเครือข่ายไวไฟ (Wi-Fi Log) และบันทึกรายละเอียดการใช้งานโทรศัพท์ (CDR) โดยข้อมูลที่นักวิจัยนิยมใช้ในการหาปริมาณความต้องการในการเดินทางระหว่างพื้นที่คือข้อมูลจากจีพีเอสที่ติดบนรถแท็กซี่และบันทึกรายละเอียดการใช้งานโทรศัพท์ ซึ่งข้อมูลจากแต่ละแหล่งมีข้อดีข้อเสียแตกต่างกันโดยข้อมูลจีพีเอสเป็นข้อมูลที่รายงานตำแหน่งของผู้ใช้งานอย่างแม่นยำและมีรอบการส่งตำแหน่งที่แน่นอนแต่ข้อมูลมีขนาดเล็กไม่สามารถใช้เป็นตัวแทนประชากรทั้งเมืองได้ แต่ข้อมูลจากซีดีอาร์เป็นข้อมูลจากกลุ่มผู้ใช้งานขนาดใหญ่แต่ตำแหน่งที่ได้เป็นตำแหน่งของเสากระจายสัญญาณเท่านั้น ผู้วิจัยจึงนำข้อดีของข้อมูลจากทั้งสองแหล่งมาใช้ในการสร้างเครื่องมือ โดยวิทยานิพนธ์ฉบับนี้เสนอวิธีการในการสร้างเครื่องมือสำหรับวิเคราะห์การเคลื่อนที่ของประชากรระหว่างบ้านและสถานที่ทำงาน โดยใช้ข้อมูลจากบันทึกรายละเอียดการใช้งานโทรศัพท์ในการหาบ้านและสถานที่ทำงานของประชากรทำให้ทราบถึงปริมาณความต้องการในการเดินทางระหว่างพื้นที่ และใช้ข้อมูลจีพีเอสจากรถแท็กซี่ในการหาเส้นทางบนถนนที่เป็นที่นิยมในการเดินทางระหว่างพื้นที่ และผู้วิจัยได้วิเคราะห์ในหลาย ๆ พื้นที่ยกตัวอย่างเช่น ประชากรที่ทำงานในพื้นที่สีลมและช่องนนทรี ,การเดินทางระหว่างสีลมกับอนุสาวรีย์ชัยสมรภูมิ และ การเดินทางจากฝั่งธนบุรีไปฝั่งพระนคร

ภาควิชา วิศวกรรมคอมพิวเตอร์

ลายมือชื่อนิสิต .....

สาขาวิชา วิศวกรรมคอมพิวเตอร์

ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ปีการศึกษา 2560



# # 5770560621 : MAJOR COMPUTER ENGINEERING

KEYWORDS: GPS / CDR

LUMPSUM TONGSINOOT: ANALYSIS OF ROAD TRAFFIC PATTERNS USING CDR AND GPS. ADVISOR: ASST. PROF. VEERA MUANGSIN, Ph.D., 94 pp.

Bangkok is a city with high population density. To acknowledge travel demand quantity inter-area and road route selected. That information is a necessary to plan transportation system in the city. At present, there are several devices to record data of users' locations such as mobile applications, taxi GPS, Wi-Fi log, and CDR. The popular data source among researchers to explore urban mobility pattern that from taxi GPS and CDR. The data from each source possesses different advantages as well as disadvantages. GPS data reports taxis' locations with high accurately and fixed data delivery period. However, on account of small size of the data, it cannot represent the population of the city. Regarding CDR data, it is from the large group of users but with received locations of cell towers. Thus, the researcher applied the advantages of the data from both source for designing a tool. This thesis propose the method of an analysis tool for population mobility between their homes and workplaces. The CDR data was used to find homes and workplaces so as to acknowledge Transportation demand quantity of inter-area. GPS data from taxis was also exploited to search for famous inter-area driving routes. The researcher revealed the examples of the analysis of population working in Silom and Chong Nonsi areas , choosing a route from Silom to Victory Monument and choosing a bridge to travel from Thonburi side to Phra Nakhon side.

Department: Computer Engineering      Student's Signature .....

Field of Study: Computer Engineering      Advisor's Signature .....

Academic Year: 2017

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความช่วยเหลือบุคคลหลายท่าน ท่านแรกขอขอบพระคุณผู้ช่วยศาสตราจารย์ ดร. วีระ เหมืองสิน อาจารย์ที่ปรึกษาวิทยานิพนธ์ที่กรุณาเสียเวลาให้คำปรึกษาตลอดทุกขั้นตอนของการทำวิจัย ให้ความช่วยเหลือในการแก้ปัญหาต่าง ๆ ที่เกิดขึ้นระหว่างทำวิจัย รวมถึงให้คำแนะนำและแนวคิดที่มีประโยชน์ในการทำวิจัย

ขอขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่าน ผู้ช่วยศาสตราจารย์ ดร. เกริก ภิรมย์โสภา ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร. กุลธิดา โรจนวิบูลย์ชัย และรองศาสตราจารย์ ดร. วรเศรษฐ สุวรรณิก กรรมการ สอบวิทยานิพนธ์ ที่ให้ความกรุณาตรวจสอบและให้คำแนะนำที่มีประโยชน์ในการปรับปรุงและแก้ไขวิทยานิพนธ์ให้มีความถูกต้องและสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณมูลนิธิศูนย์ข้อมูลจราจรอัจฉริยะไทย และบริษัท แอดวานซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน) ที่กรุณาแบ่งทรัพยากร สิทธิในการเข้าถึงข้อมูลเพื่อศึกษาวิจัย และให้คำแนะนำที่มีประโยชน์ในการใช้ข้อมูลอย่างถูกต้องและมีประสิทธิภาพจนทำให้งานวิจัยนี้สำเร็จลุล่วงได้

ขอขอบพระคุณภาควิชาวิศวกรรมคอมพิวเตอร์ที่ให้ทุนการศึกษาอัจฉริยะคืนรังเพื่อสนับสนุนค่าใช้จ่ายในการศึกษาเล่าเรียนรวมถึงสถานที่และทรัพยากรที่จำเป็นต่อการทำวิจัยทำให้การวิจัยสำเร็จและลุล่วงด้วยดี

ขอขอบพระคุณ เพื่อน ๆ พี่ ๆ ในภาควิชาวิศวกรรมคอมพิวเตอร์ที่ให้คำปรึกษาและคำแนะนำทั้งเรื่องงานวิจัยและการใช้ชีวิต

ขอขอบพระคุณบิดา มารดา และสมาชิกในครอบครัว ที่ได้ให้ความรัก กำลังใจ และให้การสนับสนุนข้าพเจ้าเสมอมา และขอขอบพระคุณคณาจารย์ทุกท่านที่ประสิทธิ์ประสาทวิชาให้แก่ข้าพเจ้า

สุดท้ายนี้ข้าพเจ้าขอขอบพระคุณทุกท่านที่มีส่วนช่วยเหลือจนสามารถทำวิทยานิพนธ์ฉบับนี้สำเร็จ ข้าพเจ้าหวังว่าวิทยานิพนธ์นี้จะมีประโยชน์ต่อผู้ที่สนใจ

## สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฎ
สารบัญรูปภาพ.....	ฏ
บทที่ 1 บทนำ.....	1
1.1. ที่มาและความสำคัญ.....	1
1.2. วัตถุประสงค์.....	4
1.3. ขอบเขตของการวิจัย.....	4
1.4. ขั้นตอนการดำเนินงาน.....	4
1.5. ประโยชน์ที่คาดว่าจะได้รับ.....	5
1.6. งานวิจัยที่ได้รับการตีพิมพ์.....	5
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	6
2.1. แนวทางในการวิจัย.....	6
2.1.1. พลวัตทางสังคม (Social Dynamics).....	6
2.1.2. พลวัตการจราจร (Traffic Dynamics).....	6
2.1.3. พลศาสตร์การดำเนินงาน (Operational Dynamics).....	6
2.2. ข้อมูลที่นำมาใช้ในการวิเคราะห์.....	7
2.2.1. ข้อมูลจากเครื่องรับจีพีเอสที่ติดบนยานพาหนะ.....	7
2.2.2. ข้อมูลจากโทรศัพท์เคลื่อนที่.....	8
2.3. การหาตำแหน่งบ้านและสถานที่ทำงาน.....	9

2.4. การแปลงพิกัดจีพีเอสเป็นพิกัดบนถนน .....	10
2.4.1. วิธีการคำนวณน้ำหนักส่วนเพิ่ม (Incremental Weight Method) .....	10
2.4.2. วิธีการคำนวณน้ำหนักทั้งหมด (Global Weight Method).....	11
2.4.3. วิธีการเปรียบเทียบเส้นโค้ง (Global Geometrical Method).....	11
2.5. การจัดกลุ่มเส้นทางที่ใช้ในการเดินทาง .....	11
2.5.1. การจัดกลุ่มโดยอาศัยลำดับของกริดที่ผ่าน .....	12
2.5.2. การจัดกลุ่มโดยอาศัยแนววิถีจากลำดับจีพีเอส .....	12
2.5.3. การจัดกลุ่มโดยอาศัยลำดับของช่วงถนน .....	12
2.6. แผนภาพโวโรนอย (Voronoi Diagram).....	13
2.7. การจัดกลุ่มแบบลำดับชั้น (Hierarchical Clustering).....	14
2.8. Haversine Formula.....	16
2.9. Levenshtein Distance .....	16
บทที่ 3 แนวคิดและวิธีการดำเนินงาน.....	18
3.1. ข้อมูลตั้งต้น .....	20
3.1.1. ข้อมูลจีพีเอส .....	20
3.1.2. ข้อมูลซีดีอาร์ .....	21
3.1.3. ข้อมูลจี-ซีดีอาร์.....	21
3.1.4. ข้อมูลถนน.....	22
3.2. การสำรวจข้อมูล.....	23
3.2.1. ข้อมูลซีดีอาร์และจี-ซีดีอาร์ .....	23
3.2.2. ข้อมูลจีพีเอส .....	25
3.3. การหาพื้นที่ให้บริการของเสากระจายสัญญาณ .....	28
3.4. การแบ่งพื้นที่ที่ใช้ศึกษา .....	30

3.5. การแปลงพื้นที่ให้บริการของเสากระจายสัญญาณให้อยู่ในกริด .....	31
3.6. การหาสถานที่ทำงานของผู้ใช้ .....	33
3.7. การหาบ้านของผู้ใช้.....	35
3.8. การกรองข้อมูลจีพีเอส .....	37
3.9. การสกัดข้อมูลการเดินทางที่ใช้ในการส่งผู้โดยสาร .....	37
3.10. การแปลงข้อมูลตำแหน่งจีพีเอสเป็นเส้นทางบนถนน .....	38
3.10.1. การหากลุ่มถนนที่มีความน่าจะเป็นสูงที่รถแท็กซี่อยู่จากพิกัดจีพีเอส .....	41
3.10.2. การเลือกถนนที่มีความน่าจะเป็นที่รถแท็กซี่อยู่ .....	41
3.11. การจัดกลุ่มรูปแบบเส้นทางระหว่างพื้นที่ .....	45
บทที่ 4 การทดลองและผลการทดลอง .....	49
4.1. เครื่องมือที่ใช้ในการพัฒนา .....	49
4.2. ผลการทดลอง.....	50
4.2.1. ผลการทดลองการหาตำแหน่งบ้านและที่ทำงาน .....	50
4.2.2. ผลการทดลองการแปลงพิกัดจีพีเอสเป็นพิกัดบนถนน .....	52
4.2.2.1. ความผิดพลาดที่ได้จากข้อมูลถนน.....	52
4.2.2.2. ความผิดพลาดจากวิธีการแปลงพิกัดจีพีเอสเป็นพิกัดบนถนน .....	54
4.2.3. ผลการทดลองการจัดกลุ่มเส้นทางที่ใช้ส่งผู้โดยสารระหว่างพื้นที่.....	56
4.3. เครื่องมือวิเคราะห์ความต้องการและเส้นทางที่นิยมใช้ในการเดินทาง .....	60
4.3.1. ปริมาณความต้องการในการเดินทางจากตำแหน่งบ้านและสถานที่ทำงาน .....	60
4.3.2. เส้นทางที่นิยมในการเดินทางระหว่างพื้นที่ตั้งต้นไปยังพื้นที่ปลายทาง.....	63
4.4. วิเคราะห์จากการหาตำแหน่งบ้านและที่ทำงาน.....	67
4.4.1. การกระจายตัวของตำแหน่งบ้านและที่ทำงาน .....	67
4.4.2. ระยะการกระจัดจากบ้านและที่ทำงาน .....	69

4.4.3. เวลาที่อยู่บ้านและที่ทำงาน.....	71
4.5. วิเคราะห์จากผลการแยกเส้นทางที่ใช้ส่งผู้โดยสาร.....	72
4.6. กรณีศึกษาที่ 1 วิเคราะห์ความต้องการและรูปแบบการเดินทางระหว่างพื้นที่.....	75
4.7. กรณีศึกษาที่ 2 ศึกษาเส้นทางที่ถูกเลือกในการเดินทางระหว่างพื้นที่.....	80
4.8. กรณีศึกษาที่ 3 ศึกษาการเดินทางจากฝั่งธนบุรีไปฝั่งพระนคร.....	83
บทที่ 5 บทสรุปและแนวทางในการพัฒนาต่อ.....	86
5.1. บทสรุป.....	86
5.2. ข้อเสนอแนะและแนวทางในการพัฒนาต่อ.....	88
รายการอ้างอิง.....	89
ประวัติผู้เขียนวิทยานิพนธ์.....	94



## สารบัญตาราง

ตารางที่ 1 เปรียบเทียบลักษณะของข้อมูลซีดีอาร์กับจีพีเอสที่ติดอยู่บนรถ .....	8
ตารางที่ 2 ตัวอย่างข้อมูลดิบของจีพีเอสที่ติดอยู่ในรถแท็กซี่ .....	20
ตารางที่ 3 ตัวอย่างข้อมูลดิบของซีดีอาร์ .....	21
ตารางที่ 4 ตัวอย่างข้อมูลดิบของจี-ซีดีอาร์ .....	22
ตารางที่ 5 เปรียบเทียบผลการหาบ้านของผู้ใช้งานระหว่างวิธีการที่ทำการกรองผู้ใช้งานอยู่กับที่ กับวิธีการหาชั่วโมงที่ผู้ใช้งานนอน.....	51
ตารางที่ 6 ทิศทางและจำนวนคนที่เดินทางมาในพื้นที่ที่ศึกษา.....	77
ตารางที่ 7 สะพานที่ถูกใช้ในการเดินทางจากฝั่งธนบุรีไปยังฝั่งพระนครมากที่สุด 10 อันดับแรก	84

## สารบัญรูปภาพ

ภาพที่ 1 การแปลง mobility graph ให้อยู่ในรูปแบบ OD matrix .....	2
ภาพที่ 2 เปรียบเทียบลักษณะข้อมูลที่ได้จากข้อมูลพิกัดของเครื่องมือที่เคลื่อนที่และข้อมูลจากรายการอุปกรณ์ที่ถูกตรวจจับโดยเครื่องตรวจจับ .....	3
ภาพที่ 3 แผนภาพไวโรนอยของจุดกำหนดที่เป็นสีแดง .....	14
ภาพที่ 4 ขั้นตอนของ Levenshtein distance.....	17
ภาพที่ 5 กระบวนการพัฒนาเครื่องมือสำหรับวิเคราะห์การเคลื่อนที่ของประชากร .....	19
ภาพที่ 6 ตัวอย่างเวกเตอร์ถนนบริเวณแยกปทุมวัน .....	23
ภาพที่ 7 กราฟแสดงค่าของสถิติความถี่จำนวนคนที่ใช้งานในแต่ละชั่วโมงเฉลี่ยต่อวัน .....	24
ภาพที่ 8 กราฟแสดงค่าของสถิติความถี่การเคลื่อนที่และอยู่กับที่ของผู้ใช้จาก ข้อมูล 2 การใช้งานที่ติดกัน .....	24
ภาพที่ 9 กราฟแสดงค่าของสถิติความถี่จำนวนเสาเฉลี่ยที่ผู้ใช้ใช้งานในแต่ละวัน .....	25
ภาพที่ 10 ตัวอย่างการเดินทางเพื่อส่งผู้โดยสารจากพิกัดสี่เขียวไปยังพิกัดสีน้ำเงินที่ผิดปกติ .....	26
ภาพที่ 11 ความหนาแน่นจำนวนพิกัดจีพีเอสในกริดขนาด $500 \times 500$ เมตร.....	27
ภาพที่ 12 ตำแหน่งของเสากระจายสัญญาณในพื้นที่ศึกษา.....	28
ภาพที่ 13 ผลจากกระบวนการหาพื้นที่ให้บริการของเสากระจายสัญญาณ .....	29
ภาพที่ 14 กระบวนการในการหาพื้นที่ให้บริการของเสากระจายสัญญาณ .....	29
ภาพที่ 15 ความแตกต่างของขนาดพื้นที่ให้บริการระหว่างใจกลางเมืองกับชานเมือง .....	30
ภาพที่ 16 ตารางตารางขนาด $500 \times 500$ ตารางเมตร ของพื้นที่จังหวัดกรุงเทพมหานครและสนามบินสุวรรณภูมิ)ตำบลราชาเทวะและหนองปรือ อำเภอบางพลี จังหวัดสมุทรปราการ(.....	31
ภาพที่ 17 รูปแบบความสัมพันธ์ระหว่างไวโรนอยกับกริด .....	32
ภาพที่ 18 กระบวนการในการหาที่ทำงานของของผู้ใช้งาน.....	34



ภาพที่ 19 กราฟการกระจายตัวของจำนวนวันที่โทรในช่วงโมงต่าง ๆ ของตัวอย่างผู้ใช้งานหนึ่งคน.....	35
ภาพที่ 20 กระบวนการในการหาบ้านของของผู้ใช้งาน .....	37
ภาพที่ 21 ตัวอย่างข้อมูลดิบที่ได้รับจากระบบที่ที่มีความคลาดเคลื่อนจากพิกัดถนน.....	39
ภาพที่ 22 ปัญหาที่เกิดจากข้อผิดพลาดของตำแหน่งจีพีเอสที่ต้องแก้ไข.....	40
ภาพที่ 23 วิธีการเลือกถนนที่มีความน่าจะเป็นสูงที่รถจะอยู่จากพิกัดจีพีเอส .....	42
ภาพที่ 24 วิธีการหาเส้นทางบนถนนทั้งหมดระหว่างช่วงถนนที่มีความน่าจะเป็นสูง .....	44
ภาพที่ 25 ตัวอย่างการเลือกเส้นทางที่จะนำไปจัดกลุ่ม .....	46
ภาพที่ 26 ตัวอย่างของเหตุการณ์ที่ทำให้ลำดับของถนนมีความคลาดเคลื่อน.....	47
ภาพที่ 27 เครื่องมือที่ใช้ในการพัฒนาระบบและแสดงการไหลของข้อมูลระหว่างเครื่องมือ .....	49
ภาพที่ 28 ตัวอย่างข้อผิดพลาดเส้นทางที่ไม่สามารถใช้ได้จริง .....	53
ภาพที่ 29 ตัวอย่างข้อผิดพลาดจากไม่มีเส้นเชื่อมในกรณีที่ถนนหลักไม่มีเกาะกลางถนน.....	54
ภาพที่ 30 ข้อผิดพลาดที่ได้จากการใช้เส้นทางที่สั้นที่สุดในการหาเส้นทางที่ดีที่สุด.....	55
ภาพที่ 31 ข้อผิดพลาดที่เกิดจากจำนวนเลนถนน .....	56
ภาพที่ 32 การทดสอบการจัดกลุ่มในระยะใกล้.....	57
ภาพที่ 33 การทดสอบการจัดกลุ่มในระยะไกล.....	58
ภาพที่ 34 ตัวอย่างเส้นทางที่มีความคล้ายกันที่% 79 .....	59
ภาพที่ 35 สถาปัตยกรรมระบบวิเคราะห์ความต้องการและเส้นทางที่นิยม .....	60
ภาพที่ 36 เครื่องมือศึกษาปริมาณความต้องการในการเดินทางจากตำแหน่งบ้านและสถานที่ทำงาน.....	61
ภาพที่ 37 กล้องแสดงแผนที่แบบโต้ตอบความหนาแน่นของตำแหน่งบ้านและที่ทำงาน.....	61
ภาพที่ 38 กล้องแสดงแผนที่แบบโต้ตอบพื้นที่ที่ถูกแบ่งออกเป็น 16 กลุ่ม.....	62
ภาพที่ 39 ตารางแสดงจำนวนคนที่เดินทางจากกลุ่มต่าง ๆ.....	62
ภาพที่ 40 เครื่องมือศึกษาเส้นทางที่นิยมในการระหว่างพื้นที่ตั้งต้นไปยังพื้นที่ปลายทาง .....	63

ภาพที่ 41	กล่องแสดงแผนที่แบบโต้ตอบสำหรับศึกษาเส้นทางที่นิยมใช้ในการเดินทาง.....	64
ภาพที่ 42	กล่องตารางสรุปข้อมูลแต่ละกลุ่ม.....	65
ภาพที่ 43	กล่องแสดงจำนวนการเดินทางที่เลือกใช้เส้นทางในแต่ละกลุ่มในช่วงเวลาต่าง ๆ.....	65
ภาพที่ 44	สัดส่วนการเลือกใช้กลุ่มเส้นทางยอดนิยม .....	66
ภาพที่ 45	แผนภาพกล่องเวลาที่ใช้ในการเดินทางของกลุ่มเส้นทางที่นิยม .....	66
ภาพที่ 46	เวลาเฉลี่ยที่ใช้ในการเดินทางของแต่ละกลุ่มในช่วงเวลาต่าง ๆ.....	67
ภาพที่ 47	การกระจายตัวของบ้านในพื้นที่กรุงเทพมหานคร.....	68
ภาพที่ 48	การกระจายตัวของสถานที่ทำงานในพื้นที่กรุงเทพมหานคร .....	69
ภาพที่ 49	การกระจายตัวที่มากที่สุด 20 อันดับ ของสถานที่ทำงานในพื้นที่กรุงเทพมหานคร ...	69
ภาพที่ 50	ระยะห่างระหว่างบ้านกับที่ทำงาน .....	70
ภาพที่ 51	ตำแหน่งของที่ทำงานของผู้ใช้ที่เดินทางมากกว่า 3 กิโลเมตรจากบ้าน.....	70
ภาพที่ 52	ตำแหน่งของบ้านของผู้ใช้ที่เดินทางมากกว่า 3 กิโลเมตรเพื่อไปทำงาน .....	71
ภาพที่ 53	การใช้งานจากตำแหน่งที่เป็นบ้านและที่ทำงาน .....	72
ภาพที่ 54	ความถี่ในการรับผู้โดยสารในเวลาต่าง ๆ .....	73
ภาพที่ 55	ความถี่ของระยะทางที่ส่งผู้โดยสาร .....	74
ภาพที่ 56	ระยะทางเฉลี่ยในช่วงเวลาต่าง ๆ ของเส้นทางที่ส่งผู้โดยสารภายในกรุงเทพฯ.....	74
ภาพที่ 57	กราฟเปรียบเทียบระยะทางเฉลี่ยในช่วงเวลาต่าง ๆ ของเส้นทางที่ส่งผู้โดยสารระหว่าง ในเมืองกับชานเมือง .....	75
ภาพที่ 58	พื้นที่ที่ใช้ในการศึกษาที่ 1 .....	75
ภาพที่ 59	กริดที่เป็นตำแหน่งบ้านมาทำงานในพื้นที่ศึกษา .....	76
ภาพที่ 60	กลุ่มที่แบ่งพื้นที่โดยอาศัยระยะห่างและทิศทางจากพื้นที่ศึกษา .....	77
ภาพที่ 61	ถนนที่นิยมใช้ในการเดินทางโดยเส้นทางที่นิยม 10 ลำดับแรก.....	79
ภาพที่ 62	เส้นทางที่นิยมใช้งานการเดินทางมากที่สุด 3 ลำดับ .....	79
ภาพที่ 63	ถนน 8 เส้นทางที่ถูกใช้มากที่สุดเพื่อเดินทางจากบริเวณถนนสีลมไปบริเวณ .....	81

ภาพที่ 64 เส้นทางที่ถูกใช้มากที่สุด 3 ลำดับเพื่อเดินทางจากบริเวณถนนสีลมไปบริเวณ  
อนุสาวรีย์ชัยสมรภูมิ โดยเรียงจากซ้ายไปขวา.....81

ภาพที่ 65 การเลือกใช้เส้นทาง 8 อันดับแรกเพื่อเดินทางจากบริเวณถนนสีลมไปบริเวณ.....82

ภาพที่ 66 ความเร็วเฉลี่ยในช่วงเวลาต่าง ๆ ของเส้นทางที่ถูกใช้มากที่สุด 8 อันดับเพื่อเดินทาง  
จากบริเวณถนนสีลมไปบริเวณอนุสาวรีย์ชัยสมรภูมิ.....82

ภาพที่ 67 ความหนาแน่นของคนที่มีบ้านในฝั่งธนบุรีและมีที่ทำงานที่ฝั่งพระนคร.....83

ภาพที่ 68 ความหนาแน่นของจุดเริ่มต้นและปลายทางของการเดินทางผ่านสะพาน ที่มีการใช้งาน  
มากที่สุด 3 อันดับ.....85



## บทที่ 1

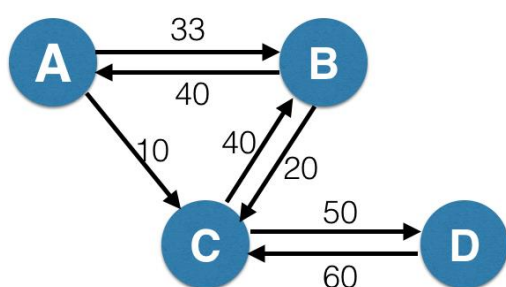
### บทนำ

#### 1.1. ที่มาและความสำคัญ

กรุงเทพมหานครเป็นศูนย์กลางทางเศรษฐกิจและการท่องเที่ยว ในปี 2558 จากรายงานสถิติจำนวนประชากร สำนักบริหารการทะเบียน กรมการปกครอง กรุงเทพมหานครมีประชากร 5,696,409 คน [1] ความหนาแน่นต่อพื้นที่ 3,631.20 คน/ตร.กม. ทำให้เกิดความต้องการในการเดินทางเป็นปริมาณมาก แต่การเดินทางระบบรางและทางเรือยังไม่ครอบคลุม ทำให้การเดินทางส่วนใหญ่เป็นทางถนน โดยจำนวนรถที่จดทะเบียนในกรุงเทพมหานครในปี 2558 รวมทุกประเภท 9,018,594 คัน [2] ส่งผลให้เกิดปัญหาการจราจรติดขัดในช่วงเวลาเร่งด่วน สถิติจราจรปี 2558 ความเร็วเฉลี่ยในชั่วโมงเร่งด่วนพื้นที่กรุงเทพมหานครชั้นนอก 31.15 กม./ชม., พื้นที่กรุงเทพชั้นกลาง 26.00 กม./ชม. และพื้นที่กรุงเทพชั้นใน 15.65 กม./ชม. [2] การแก้ปัญหาการจราจรจำเป็นต้องทราบปริมาณความต้องการในการเดินทางจากพื้นที่หนึ่งไปอีกพื้นที่หนึ่งเพื่อใช้เครื่องมือในการวางแผนระบบขนส่งภายในเมืองเพื่อสามารถรองรับความต้องการของประชากรภายในเมืองได้อย่างมีประสิทธิภาพ

โดยจากงานวิจัยก่อนหน้านี้พบว่าการเดินทางของคนแต่ละคนมักจะมีรูปแบบการเดินทางซ้ำ ๆ กันในแต่ละช่วงเวลาเช่น เดินทางไปทำงานในตอนเช้ากลับบ้านในตอนเย็น โดยใช้เส้นทางเหมือนกันเกือบทุกวัน [3], [4] เมื่อนำการเดินทางของแต่ละคนมารวมกันจะทำให้ทราบถึงความต้องการในการเดินทางโดยรวม ซึ่งสามารถแทนได้ด้วย OD matrix

OD matrix (Origin-Destination Matrix) คือตาราง 2 มิติ ที่แสดงถึงปริมาณการเดินทางจากบริเวณหนึ่งไปยังบริเวณหนึ่งโดยสร้างจากผลรวมของข้อมูลจุดเริ่มต้นและจุดหมายปลายทางของแต่ละการเดินทาง จากภาพที่ 1 แสดงตัวอย่างของ OD matrix จาก mobility graph ด้านซ้าย ในแต่ละช่องในตารางแสดงปริมาณการเดินทางจากพื้นที่ของแถวนั้นไปยังพื้นที่ของคอลัมน์นั้นเช่น จากรูปมีปริมาณการเดินทางจากพื้นที่ A ไปพื้นที่ C เท่ากับ 10 เที่ยว โดยสามารถเพิ่มมิติเพื่อให้วิเคราะห์ได้ละเอียดขึ้น เช่น เพิ่มมิติเวลาที่เริ่มออกเดินทางทำให้ทราบถึงปริมาณความต้องการในแต่ละช่วงเวลา ซึ่งสามารถนำไปใช้ในการปรับปรุงระบบขนส่งที่มีอยู่เช่น กำหนดเส้นทางและความถี่ของรถประจำทางในแต่ละช่วงเวลา หรือใช้วางแผนเพื่อสร้างระบบขนส่งในอนาคต เช่น ใช้เพื่อออกแบบเส้นทางของการขนส่งระบบราง โดยอาศัยปริมาณความต้องการในการเดินทางระหว่างพื้นที่ตามแนว



a) mobility graph

	A	B	C	D
A	0	33	10	0
B	40	0	20	0
C	0	40	0	50
D	0	0	60	0

(b) OD matrix

ภาพที่ 1 การแปลง mobility graph ให้อยู่ในรูปแบบ OD matrix

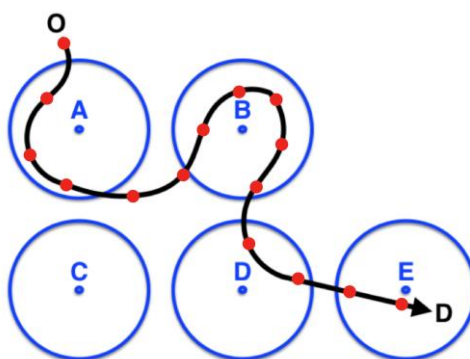
การสร้าง OD matrix จำเป็นต้องใช้เครื่องมือในการเก็บข้อมูลการเคลื่อนที่ของแต่ละบุคคลในรูปแบบ ตำแหน่งและเวลา ในปัจจุบันมีเครื่องมือที่ใช้ระบุตำแหน่งของประชากรรายบุคคลได้ เช่น GPS, Wi-Fi , เสากกระจายสัญญาณโทรศัพท์มือถือ โดยลักษณะของข้อมูลตำแหน่งมีอยู่ 2 แบบคือ พิกัดของเครื่องมือที่เคลื่อนที่ โดยเครื่องมือจะทำการคำนวณตำแหน่งของตัวเองและเก็บตำแหน่งขณะที่เครื่องมือเคลื่อนที่ไปที่ต่าง ๆ เช่น ข้อมูลจีพีเอสที่เก็บโดยเครื่องรับจีพีเอสภายในรถ และอีกแบบคือรายการอุปกรณ์ที่เข้ามาใกล้เครื่องตรวจจับที่อยู่กับที่ โดยอุปกรณ์จะถูกตรวจจับหรือติดต่อยังเครื่องตรวจจับที่อยู่ใกล้ และถูกเครื่องตรวจจับบันทึกว่าพบอุปกรณ์ที่ถูกตรวจจับหรือติดต่อก่อนที่เครื่องตรวจจับในเวลานั้น โดยตำแหน่งที่ได้จะเป็นตำแหน่งของเครื่องตรวจจับ ตัวอย่างข้อมูลเช่น บันทึกรายละเอียดการใช้งานโทรศัพท์ (Call Detail Record : CDR), ข้อมูลการใช้งานไวไฟ (Wi-Fi log) ภาพที่ 2 แสดงการเปรียบเทียบลักษณะข้อมูลทั้งสองแบบ โดยภาพ (a) เป็นตัวอย่างเส้นทางการเดินทางจากจุด O ไปยังจุด D โดยจุดสีแดงแสดงพิกัดจีพีเอสและวงกลมสีน้ำเงินแสดง เครื่องตรวจจับ A, B, D และ E ที่เดินทางผ่าน ภาพ (b) ข้อมูลพิกัดของเครื่องมือที่เคลื่อนที่ โดยข้อมูลที่ได้จะเป็นพิกัดที่อยู่บนเส้นทางการเดินทาง ภาพ (c) รายการอุปกรณ์ที่เข้ามาใกล้เครื่องตรวจจับที่อยู่กับที่ ข้อมูลที่ได้จะเป็นพิกัดของเครื่องตรวจจับที่ตรวจพบอุปกรณ์ โดยข้อมูลแต่ละแหล่งมีข้อดีข้อเสียต่างกัน

- ด้านความแม่นยำของการระบุตำแหน่งของข้อมูลซีดีอาร์บอกได้เพียงผู้ใช้นั้นใช้บริการที่เสากกระจายสัญญาณใดและประมาณตำแหน่งของผู้ใช้จากพื้นที่ให้บริการของเสากกระจายสัญญาณต้นนั้น ส่วนข้อมูลที่ได้จากจีพีเอสสามารถระบุตำแหน่งแม่นยำในระดับเมตร
- ด้านขนาดกลุ่มตัวอย่างข้อมูลซีดีอาร์มีปริมาณผู้ใช้งานมากกว่าข้อมูลจากจีพีเอส ด้านความถี่และความสม่ำเสมอของข้อมูลซีดีอาร์เกิดข้อมูลเฉพาะเวลาที่มีการใช้โทรศัพท์จึงมี

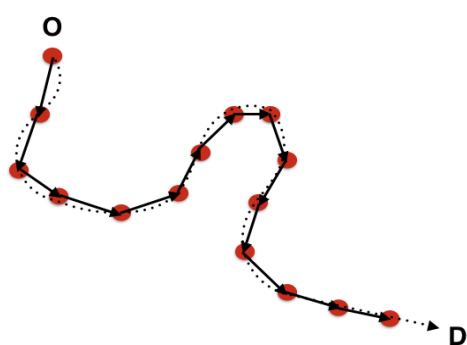
ความถี่และความสม่ำเสมอแตกต่างกันในแต่ละคน จีพีเอสมีความถี่และความสม่ำเสมอสูง แต่มีข้อมูลเฉพาะเมื่อเดินทางด้วยรถแท็กซี่

- ด้านพื้นที่ที่ครอบคลุมซีตีมาร์เก็บข้อมูลผู้ใช้งานจากทุกพื้นที่ แต่จีพีเอสมีเฉพาะบนถนน
- ข้อมูลซีตีมาร์สามารถศึกษาแต่ละเบอร์เป็นช่วงเวลานาน ๆ แต่จีพีเอสในรถแท็กซี่ไม่มีข้อมูลผู้โดยสารจึงไม่สามารถรวบรวมข้อมูลการเดินทางหลายๆครั้งของผู้โดยสารแต่ละคนได้ แต่ถ้าเป็นจีพีเอสจากรถแท็กซี่แบบ Uber หรือรถส่วนบุคคลจึงสามารถทำได้

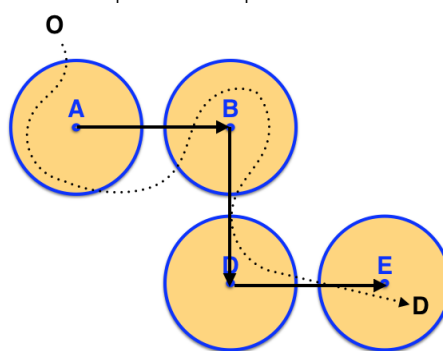
การสร้าง OD matrix ในปัจจุบันมักใช้ข้อมูลจากจีพีเอสหรือซีตีมาร์เพียงอย่างเดียวอย่างหนึ่ง ซึ่งข้อมูลจากแต่ละแหล่งมีข้อดีข้อเสียแตกต่างกัน ผู้วิจัยจึงมีแนวคิดในการใช้ข้อมูลจากทั้งจีพีเอสและซีตีมาร์มาสร้าง OD matrix ซึ่งผลที่ได้จะทำให้ OD matrix สามารถมองได้สองระดับคือสามารถทราบถึงความต้องการในการเดินทางทั้งหมดในแต่ละพื้นที่โดยอาศัยข้อดีจากจำนวนผู้ใช้งานที่มากของซีตีมาร์และสามารถเห็นถึงเส้นทางที่นิยมใช้ในการเดินทางระหว่างพื้นที่โดยอาศัยข้อดีจากความแม่นยำและความต่อเนื่องของข้อมูลจีพีเอส



(a) เส้นทางการเดินทางจากจุด O ไปยังจุด D



(b) ข้อมูลพิกัดของเครื่องมือที่เคลื่อนที่



(c) ข้อมูลจากรายการอุปกรณ์ที่ถูกตรวจจับโดยเครื่องตรวจจับ

ภาพที่ 2 เปรียบเทียบลักษณะข้อมูลที่ได้จากข้อมูลพิกัดของเครื่องมือที่เคลื่อนที่และข้อมูลจากรายการอุปกรณ์ที่ถูกตรวจจับโดยเครื่องตรวจจับ

## 1.2. วัตถุประสงค์

เพื่อสร้างเครื่องมือสำหรับวิเคราะห์การเคลื่อนที่ของประชากรระหว่างสถานที่ที่เดินทางไป เป็นประจำจากข้อมูลจีพีเอสของรถแท็กซี่และซีดีอาร์โดยสามารถวิเคราะห์ได้ทั้งปริมาณการเดินทาง ระหว่างเขต เส้นทางที่ใช้ในการเดินทางบนถนนระหว่างเขต และพื้นที่ให้บริการของถนนแต่ละช่วง

## 1.3. ขอบเขตของการวิจัย

- งานวิจัยนี้ศึกษาการเดินทางระหว่างพื้นที่สำคัญของแต่ละบุคคลที่มีการเดินทางที่เป็นรูปแบบที่แน่นอน
- งานวิจัยนี้จะทำการศึกษารูปแบบการเคลื่อนที่ของประชากรในเขตกรุงเทพมหานครจากข้อมูลซีดีอาร์ และจีพีเอสจากรถแท็กซี่ในเดือนสิงหาคม กันยายน และตุลาคม 2559

## 1.4. ขั้นตอนการดำเนินงาน

1. ศึกษางานวิจัยที่เกี่ยวข้องกับการวิเคราะห์การเคลื่อนที่ของประชากร
2. ศึกษาข้อมูลดิบที่ได้รับมาเบื้องต้น
3. กรองข้อมูลที่ใช้ไม่ได้
4. ออกแบบขั้นตอนวิธีการในการหาบ้านและที่ทำงานจากข้อมูลซีดีอาร์
5. พัฒนาและปรับปรุงขั้นตอนวิธีการในการหาบ้านและที่ทำงานให้มีความแม่นยำมากขึ้น
6. สร้าง OD matrix จากตำแหน่งบ้านและที่ทำงาน
7. กรองข้อมูลการเดินทางที่ใช้ส่งผู้โดยสาร
8. ออกแบบขั้นตอนวิธีการในการแปลงพิกัดจีพีเอสเป็นถนนที่ใช้ในการเดินทาง
9. พัฒนาและปรับปรุงขั้นตอนวิธีการในการแปลงพิกัดจีพีเอสเป็นถนนที่ใช้ในการเดินทาง
10. ออกแบบขั้นตอนวิธีการในการจัดกลุ่มการเดินทาง
11. พัฒนาและปรับปรุงขั้นตอนวิธีการการจัดกลุ่มการเดินทาง
12. สร้างส่วนติดต่อผู้ใช้งาน
13. ทดลองวิเคราะห์ข้อมูลในหลาย ๆ พื้นที่
14. สรุปผลการทดลองและจัดทำวิทยานิพนธ์

### 1.5. ประโยชน์ที่คาดว่าจะได้รับ

- ได้กระบวนการที่ใช้ในการสร้างแบบจำลองการเคลื่อนที่ของประชากรที่เหมาะสมกับข้อมูลที่อยู่ในประเทศไทย
- แบบจำลองที่ได้สามารถใช้เป็นเครื่องมือในการวางแผนระบบขนส่งในกรุงเทพมหานคร

### 1.6. งานวิจัยที่ได้รับการตีพิมพ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้นำเสนอในการประชุมวิชาการดังนี้

Lumpsum Tongsinoot and Veera Muangsin, "Exploring Home and Work Locations in a City from Mobile Phone Data", The 15<sup>th</sup> IEEE International Conference on Smart City (SmartCity-2017), Bangkok, Thailand, Dec 18-20, 2017.





## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

#### 2.1. แนวทางในการวิจัย

ในปัจจุบันงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์การเคลื่อนที่ของประชากรมีความหลากหลายขึ้นอยู่กับประเด็นที่ต้องการวิเคราะห์และข้อมูลที่นำมาใช้วิเคราะห์ จากวิจัย [5] ได้สำรวจงานวิจัยที่วิเคราะห์จากข้อมูลการเคลื่อนที่ที่สามารถแบ่งออกได้ดังนี้

##### 2.1.1. พลวัตทางสังคม (Social Dynamics)

พลวัตทางสังคมเป็นการวิเคราะห์พฤติกรรมการเดินทางของประชากรภายในเมือง เช่น

- หาสถานที่สำคัญของแต่ละบุคคล [4], [12], [13]
- สถานที่สำคัญของเมืองที่มีการเดินทางไปมากหรือจัดกลุ่มพื้นที่ [6], [7], [9]
- พฤติกรรมการเดินทางของประชากรระหว่างพื้นที่ [7], [8], [9]

##### 2.1.2. พลวัตการจราจร (Traffic Dynamics)

Traffic dynamics เป็นการวิเคราะห์สภาพจราจรบนถนนโดยอาศัยข้อมูลการเดินทางของยานพาหนะในแต่ละช่วงเวลา ผลที่ได้มักอยู่ในรูปแบบของสภาพการจราจรบนถนนเส้นต่าง ๆ คำนวณระยะเวลาในการเดินทางบนถนน [31] หรือพยากรณ์สภาพจราจร

##### 2.1.3. พลศาสตร์การดำเนินงาน (Operational Dynamics)

เป็นการวิเคราะห์พฤติกรรมอื่น ๆ เช่น วิเคราะห์พฤติกรรมการรับส่งหรือหาลูกค้าของรถแท็กซี่ การเลือกเส้นทางในการเดินทาง [10], [6] ตรวจสอบพฤติกรรมผิดปกติบนท้องถนนเช่น ตรวจสอบอุบัติเหตุบนถนน

งานวิจัยนี้จัดอยู่ในกลุ่มพลวัตทางสังคมเนื่องจากการวิเคราะห์พฤติกรรมการเดินทางของประชากรภายในเมืองเป็นพื้นฐานที่สามารถต่อยอดได้อีกมากมาย โดยจะทำการสกัดสถานที่ที่แต่ละคนเดินทางไปเป็นประจำในช่วงเวลาต่าง ๆ และจะทำการวิเคราะห์รูปแบบการเดินทางระหว่างสถานที่ที่สำคัญ

## 2.2. ข้อมูลที่นำมาใช้ในการวิเคราะห์

### 2.2.1. ข้อมูลจากเครื่องรับจีพีเอสที่ติดบนยานพาหนะ

ปัจจุบันมียานพาหนะจำนวนหนึ่งติดตั้งเครื่องรับจีพีเอส ไม่ว่าจะเป็นรถยนต์ส่วนบุคคล, รถประจำทาง หรือรถแท็กซี่ โดยเครื่องรับจีพีเอสจะเก็บตำแหน่งของรถพร้อมเวลาระบบกำหนดตำแหน่งบนโลก (GPS: Global Positioning System) คือ ระบบที่ใช้ดาวเทียมเพื่อบอกตำแหน่งบนหรือใกล้พื้นผิวโลก ทำงานโดยเครื่องรับจีพีเอสรับตำแหน่งของดาวเทียมและเวลาที่ส่งจากดาวเทียม ตำแหน่งของเครื่องรับจีพีเอสคำนวณจากระยะทางระหว่างเครื่องรับจีพีเอสกับดาวเทียมแต่ละดวง โดยคำนวณจากระยะเวลาในการส่งจากดาวเทียมมายังเครื่องรับ เพื่อให้เกิดความแม่นยำในการระบุตำแหน่งของเครื่องรับจีพีเอสต้องได้รับข้อมูลจากดาวเทียมอย่างน้อย 4 ดวง โดย 3 ดวงใช้ในการระบุตำแหน่งบนพื้นผิวโลก อีก 1 ดวงใช้ในการกำจัดค่าผิดพลาด

ปัจจัยที่มีผลต่อความแม่นยำของการคำนวณตำแหน่งของเครื่องรับจีพีเอส

- ค่าความผิดพลาดจากดาวเทียม เนื่องจากการหาระยะห่างระหว่างเครื่องรับจีพีเอสกับดาวเทียม จำเป็นต้องอาศัยเวลาและตำแหน่งของดาวเทียม ดังนั้นถ้าเวลาและตำแหน่งของดาวเทียมคลาดเคลื่อนจากความเป็นจริง ผลการคำนวณที่ได้จะได้ตำแหน่งที่ผิดเพี้ยนไป
- เสารับสัญญาณ เนื่องจากต้องรับสัญญาณจากดาวเทียม เมื่อสภาพอากาศไม่ดี, สภาพแวดล้อมภายในเมืองหรือใต้ต้นไม้สัญญาณที่ได้รับจะอ่อนลง เสารับสัญญาณที่รับสัญญาณอ่อนได้ดีจะสามารถรับสัญญาณจากดาวเทียมได้มากกว่าในสภาพแวดล้อมไม่ดี
- ค่าความผิดพลาดจากการรับสัญญาณสะท้อน เนื่องจากการหาระยะห่างระหว่างเครื่องรับจีพีเอสกับดาวเทียมจำเป็นต้องได้รับสัญญาณเป็นแนวเส้นตรงจากดาวเทียม แต่ในความเป็นจริงในบางพื้นที่ได้รับสัญญาณโดยการสะท้อนจากสภาพแวดล้อมรอบข้างเช่น ตึกสูงหรือภูเขา ทำให้การคำนวณผิดพลาดได้ ข้อดีของข้อมูลจากยานพาหนะที่ติดตั้งเครื่องรับจีพีเอสคือได้ตำแหน่งของ

ยานพาหนะที่มีความแม่นยำสูง และมีการส่งข้อมูลมาตลอดเวลาที่เปิดเครื่องรับจีพีเอสใน ความถี่ของข้อมูลที่แน่นอนทำให้ทราบถึงเส้นทางที่ใช้ในการเดินทาง แต่มีข้อเสียตรงที่ข้อมูลที่ได้รับทั้งหมดอยู่บนถนนเท่านั้น, จำนวนรถน้อย และทราบเฉพาะตำแหน่งของยานพาหนะเท่านั้นไม่ทราบตำแหน่งที่ผู้โดยสารต้องการไปจริง ๆ

## 2.2.2. ข้อมูลจากโทรศัพท์เคลื่อนที่

ข้อมูลจากโทรศัพท์เคลื่อนที่ (CDR : Call Detail Record) เป็นข้อมูลที่ได้มาจากอุปกรณ์กระจายสัญญาณโทรศัพท์ โดยอุปกรณ์กระจายสัญญาณโทรศัพท์จะเก็บข้อมูลทุกครั้งที่โทรศัพท์เคลื่อนที่ทำการติดต่อกับอุปกรณ์กระจายสัญญาณโทรศัพท์เพื่อโทรไปหาบุคคลอื่น, รับโทรศัพท์, รับส่งข้อความสั้น (SMS) และใช้งานอินเทอร์เน็ต ข้อมูลการใช้งานทั้งหมดถูกรวบรวมโดยผู้ให้บริการเครือข่ายเพื่อนำมาคิดค่าบริการ ประกอบด้วย หมายเลขโทรศัพท์, วันเวลาที่ทำการโทร, เวลาที่ใช้ในการโทรทั้งหมด และตำแหน่งของเสากระจายสัญญาณโทรศัพท์ที่ติดต่อ

ข้อดีของข้อมูลซีดีอาร์คือได้ขนาดกลุ่มตัวอย่างขนาดใหญ่และครอบคลุมพื้นที่ขนาดใหญ่ แต่ข้อเสียคือได้ตำแหน่งที่มีความแม่นยำต่ำ และมีความถี่ของตำแหน่งขึ้นอยู่กับการใช้งาน โดยตำแหน่งที่ได้ขึ้นอยู่กับความหนาแน่นของเสากระจายสัญญาณโทรศัพท์ซึ่งจะขึ้นอยู่กับปริมาณการใช้งานของลูกค้าในพื้นที่นั้น ทำให้มีความหนาแน่นของเสากระจายสัญญาณโทรศัพท์สูงในพื้นที่ที่มีการใช้งานมาก และหนาแน่นของเสากระจายสัญญาณโทรศัพท์ต่ำในพื้นที่ที่มีการใช้งานน้อย

เนื่องจากข้อมูลจากแหล่งต่างกัันมีเอกลักษณ์ที่ [11] ได้ทำการเปรียบเทียบลักษณะของข้อมูลจากแต่ละแหล่ง ตารางที่ 1 เป็นตารางเปรียบเทียบลักษณะของข้อมูลซีดีอาร์และเครื่องรับจีพีเอสที่ติดอยู่บนรถ โดยข้อมูลจากเครื่องรับจีพีเอสที่ติดอยู่บนรถได้เปรียบตรงที่ให้ข้อมูลที่แม่นยำกว่า, มีความถี่สูง และบอกข้อมูลเพิ่มเติมมากกว่า แต่ข้อมูลซีดีอาร์มีจุดเด่นอยู่ที่มีกลุ่มตัวอย่างขนาดใหญ่และข้อมูลไม่ได้เฉพาะบนถนน

ตารางที่ 1 เปรียบเทียบลักษณะของข้อมูลซีดีอาร์กับจีพีเอสที่ติดอยู่บนรถ

ข้อมูลตั้งต้น	ซีดีอาร์	จีพีเอสที่ติดอยู่บนรถ
ความแม่นยำของตำแหน่ง	ระดับเมตร – กิโลเมตร ขึ้นอยู่กับความหนาแน่นของเสา	ระดับเมตร
ตำแหน่งที่บอก	ตำแหน่งของเสา	ตำแหน่งบนถนน
ความเร็ว	ไม่มี	มี
ทิศทาง	ไม่มี	มี
ความถี่ในการส่งข้อมูล	ต่ำและไม่แน่นอน	ความถี่สูงและแน่นอน
ขนาดกลุ่มตัวอย่าง	ใหญ่	เล็ก
ข้อมูลเส้นทาง	ไม่มี	มี

ในปัจจุบันมีงานวิจัยไม่มากนักที่ใช้ข้อมูลจากแหล่งข้อมูลมากกว่าหนึ่งแหล่ง โดยงานวิจัยที่น่าสนใจเป็นของ Desheng et. Al [8] ผู้วิจัยได้ใช้ข้อมูลจากซีทีอาร์และข้อมูลจากระบบขนส่งสาธารณะ(รถแท็กซี่, รถไฟฟ้าและรถประจำทาง) มาสร้าง mobility graph ภายในเมือง เพื่อศึกษาการเคลื่อนที่ของประชากรในแต่ละพื้นที่ โดยผู้วิจัยแสดงให้เห็นว่าข้อมูลจากแต่ละแหล่งมีเอกลักษณ์เฉพาะตัวเช่น ระยะทางในแต่ละการเดินทางที่ต่างกัน และระยะเวลาในการรวบรวมข้อมูลมีผลต่อผลลัพธ์ของ mobility graph ถ้ามีเก็บข้อมูลน้อยเกินไปข้อมูลแต่ละแหล่งจะต่างกันมาก แต่ถ้ารวมข้อมูลหลาย ๆ วันความต่างของข้อมูลจะลดลงจนสามารถใช้แทนกันได้ โดยในงานวิจัยได้ใช้ข้อมูลจากซีทีอาร์เป็นหลักแต่ทำการพิจารณาให้ข้อมูลนั้นครอบคลุมข้อมูลจากแหล่งอื่นโดยใช้การรวมข้อมูลของหลายวัน

ในงานวิจัยนี้จะทำการวิเคราะห์โดยใช้ข้อมูลซีทีอาร์และข้อมูลจีพีเอสของรถแท็กซี่ โดยข้อดีของข้อมูลซีทีอาร์ที่มีกลุ่มตัวอย่างขนาดใหญ่ในการหาปริมาณความต้องการในการเดินทางระหว่างพื้นที่ และใช้ข้อดีจากข้อมูลจีพีเอสของรถแท็กซี่ที่สามารถศึกษาเส้นทางในการเดินทางเพื่อนำมาใช้ในการหาเส้นทางที่นิยมในการเดินทาง

### 2.3. การหาตำแหน่งบ้านและสถานที่ทำงาน

งานวิจัยที่เกี่ยวกับการหาตำแหน่งบ้านและสถานที่ทำงานจากซีทีอาร์ส่วนใหญ่จะเลือกตำแหน่งบ้านและสถานที่ทำงานจากบริเวณที่มีการใช้งานเป็นจำนวนวันสูงสุดในช่วงเวลาที่ผู้วิจัยคาดว่าผู้ใช้งานจะอยู่ที่บ้านหรือสถานที่ทำงาน [12], [13], [14], [15] ซึ่งการเลือกตำแหน่งบ้านและสถานที่ทำงานด้วยวิธีนี้มีข้อผิดพลาดในบางสถานการณ์เช่น ผู้ใช้งานคนนั้นเดินทางไปรับส่งบุคคลอื่นเป็นประจำในช่วงเวลาทำงานซึ่งมีโอกาสที่ผู้ใช้งานจะโทรติดต่อกับบุคคลนั้นส่งผลให้มีข้อมูลการใช้งานในสถานที่นั้นเป็นจำนวนวันที่สูงตามไปด้วยทำให้มีโอกาสที่สถานที่นั้นจะถูกเลือกเป็นตำแหน่งที่ทำงาน หรือสถานการณ์ที่ผู้ใช้งานไปพักผ่อนสังสรรค์กับเพื่อนหลังเลิกงานซึ่งมีโอกาสสูงที่จะถูกเลือกเป็นตำแหน่งบ้าน

นักวิจัยกลุ่มหนึ่งปรับปรุงวิธีในการหาตำแหน่งโดยทำการกรองข้อมูลการใช้งานจากซีทีอาร์ให้เหลือเฉพาะข้อมูลในช่วงที่ผู้ใช้งานอยู่กับที่โดยพิจารณาจากข้อมูลซีทีอาร์ในช่วงเวลาที่ใกล้เคียงกันจะต้องอยู่ในตำแหน่งเดียวกัน [16], [17], [18] แต่วิธีการนี้แก้ปัญหาได้เฉพาะกรณีการใช้งานระหว่างการเดินทาง แต่ไม่สามารถแก้ปัญหาผู้ใช้งานไปอยู่ในสถานที่นั้นเป็นช่วงเวลาสั้น ๆ เช่น กรณีไปพักผ่อนสังสรรค์

งานวิจัยอีกกลุ่มหนึ่งพิจารณาจำนวนวันที่อยู่ในสถานที่นั้นและเวลาที่อยู่ในสถานที่นั้นเช่นงานของ S. Isaacman et al. [16] ที่มีเงื่อนไขต้องอยู่ในสถานที่นั้นเป็นจำนวนชั่วโมงมากกว่า

ครั้งหนึ่งของช่วงเวลาที่ถูกวิจัยคาดว่าผู้ใช้งานจะอยู่ที่บ้านหรือสถานที่ทำงาน แต่วิธีนี้มีปัญหาเกี่ยวกับการหาบ้านเนื่องจากแต่ละคนใช้เวลาอยู่ที่บ้านคนละช่วงเวลา ดังนั้นจะทำให้มีปัญหาในการเข้าเงื่อนไขจำนวนชั่วโมงที่จะต้องมากกว่าครั้งหนึ่งของช่วงเวลาที่ถูกวิจัยคาดว่าผู้ใช้งานจะอยู่ที่บ้านถ้าช่วงเวลาที่พักอยู่ที่บ้านจริงกับช่วงเวลาที่ถูกวิจัยกำหนดไม่ตรงกัน และจากการทดลองพบว่าข้อมูลซีดีอาร์ของผู้ใช้แต่ละคนมีเพียง 9.72 ครั้งต่อวันซึ่งส่วนใหญ่จะเป็นเวลากลางวันทำให้วิธีในการกรองซีดีอาร์ให้เหลือเฉพาะข้อมูลในช่วงที่ผู้ใช้งานอยู่กับที่แทบจะเป็นไปไม่ได้

ผู้วิจัยปรับปรุงวิธีการในการหาบ้านและสถานที่ทำงานใหม่โดยการหาบ้านจะหาจากตำแหน่งใช้งานบ่อยที่สุดจากการใช้งานครั้งสุดท้ายก่อนเข้านอนและการใช้งานครั้งแรกหลังจากตื่นนอน และหาสถานที่ทำงานโดยการพิจารณาตำแหน่งที่มีการใช้งานเป็นจำนวนวันมากและเป็นตำแหน่งที่มีการใช้งานสูงอันดับต้น ๆ ในแต่ละชั่วโมงของช่วงเวลาทำงาน

## 2.4. การแปลงพิกัดจีพีเอสเป็นพิกัดบนถนน

พิกัดจีพีเอสที่ได้จากรถแท็กซี่เป็นตำแหน่งที่ได้จากการคำนวณของเครื่องจีพีเอสทำให้ค่าที่ได้ไม่ใช่ตำแหน่งบนถนนที่แท็กซี่นั้นอยู่จริง โดยความคลาดเคลื่อนจากถนนขึ้นอยู่กับสภาพแวดล้อมของพื้นที่รอบรถแท็กซี่ ดังนั้นการที่จะทราบชื่อถนนที่รถแท็กซี่ใช้จำเป็นต้องทำการแปลงพิกัดจีพีเอสให้เป็นตำแหน่งของถนน ซึ่งจากการสำรวจงานวิจัยที่เกี่ยวกับการแปลงพิกัดจีพีเอสเป็นพิกัดถนนสามารถแบ่งได้เป็น 3 กลุ่มดังนี้

### 2.4.1. วิธีการคำนวณน้ำหนักส่วนเพิ่ม (Incremental Weight Method)

การแปลงพิกัดจีพีเอสเป็นพิกัดบนถนนโดยวิธีการคำนวณน้ำหนักส่วนเพิ่ม [19], [20], [21], [22] เป็นการคำนวณหาหน้าหนักที่มากที่สุดโดยอาศัยพิกัดปัจจุบัน พิกัดบนถนนที่ผ่านการแปลงพิกัดจีพีเอสเป็นพิกัดบนถนนของพิกัดจีพีเอสก่อนหน้า และถนนที่ใกล้กับพิกัดปัจจุบัน โดยการคำนวณส่วนใหญ่จะอาศัยระยะห่างที่น้อยที่สุดระหว่างพิกัดจีพีเอสกับถนน, ทิศทางระหว่างทิศหัวรถกับทิศทางถนน, ทิศทางระหว่างจุดจีพีเอสก่อนหน้า, ระยะทางที่สั้นที่สุดจากถนนที่ผ่านการแปลงพิกัดจีพีเอสเป็นพิกัดบนถนนของพิกัดจีพีเอสก่อนหน้า และความเร็วที่ต้องใช้ในการเดินทางจากตำแหน่งถนนก่อนหน้ามายังตำแหน่งถนนปัจจุบันเป็นต้น โดยการแปลงพิกัดจีพีเอสเป็นพิกัดบนถนนด้วยวิธีนี้พบว่ามีข้อผิดพลาดที่สูงกว่าวิธีอื่น ซึ่งมักจะถูกนำมาใช้ในเครื่องนำทางระบบจีพีเอสเนื่องจากทราบเฉพาะพิกัดปัจจุบันและพิกัดก่อนหน้า

### 2.4.2. วิธีการคำนวณน้ำหนักทั้งหมด (Global Weight Method)

การแปลงพิกัดจีพีเอสเป็นพิกัดบนถนนโดยการคำนวณน้ำหนักทั้งหมด [23], [24], [25], [26] เป็นการคำนวณเหมือนกับวิธีการคำนวณน้ำหนักส่วนเพิ่ม แต่จะอาศัยข้อมูลตำแหน่งจากทุกจุดมาคำนวณหาเส้นทางที่มีน้ำหนักมากที่สุด ซึ่งโดยหลักการคือเพิ่มความแม่นยำของคำนวณน้ำหนักที่โดยอาศัยพิกัดปัจจุบัน, พิกัดก่อนหน้า และพิกัดจีพีเอสในอนาคต เนื่องจากผลที่ได้จากวิธีการคำนวณน้ำหนักส่วนเพิ่มในบางพื้นที่อาจมีถนนที่เป็นไปได้มากกว่าหนึ่งเส้นซึ่งวิธีคำนวณโดยอาศัยพิกัดปัจจุบันและพิกัดก่อนหน้าจะเลือกถนนที่มีน้ำหนักมากที่สุด ณ เวลานั้นและไม่สามารถกลับมาเปลี่ยนถนนได้ แต่ถ้าพิจารณาจากพิกัดจีพีเอสถัดไปจากตำแหน่งนั้นพบว่าถนนที่ถูกเลือกไปนั้นอาจจะไม่ใช่ถนนที่ถูกต้อง ซึ่งวิธีการคำนวณน้ำหนักทั้งหมดจะแก้ปัญหานี้ได้ แต่วิธีการคำนวณน้ำหนักทั้งหมดมีปัญหาตรงที่จำเป็นต้องทราบพิกัดทั้งหมดของการเดินทางทั้งหมด

### 2.4.3. วิธีการเปรียบเทียบเส้นโค้ง (Global Geometrical Method)

วิธีการแปลงพิกัดจีพีเอสเป็นพิกัดบนถนนที่กล่าวมาแล้วทั้ง 2 วิธีนั้นเป็นการคำนวณน้ำหนักโดยอาศัยระยะทางหรือทิศทางแต่ไม่ได้สนใจรูปร่างของถนนว่าถนนโค้งในทางใด [27], [28] โดยวิธีการเปรียบเทียบเส้นโค้งถนนกับเส้นโค้งที่ได้จากลำดับพิกัดจีพีเอสมีพื้นฐานจากการหาระยะห่างระหว่างเส้นโค้งของถนนกับเส้นโค้งจากลำดับพิกัดจีพีเอส โดยวิธีที่ใช้มากที่สุดคือการหาระยะห่างที่น้อยที่สุดโดยวิธีหาระยะห่างฟรีเชท (Fréchet distance) ซึ่งเป็นวิธีการหาความคล้ายกันของเส้นโค้ง

โดยในงานวิจัยนี้เลือกใช้การคำนวณน้ำหนักที่มากที่สุดโดยอาศัยตำแหน่งพิกัดในอดีต ปัจจุบัน และอนาคต ของ Jian Huang [23] ที่มีความแม่นยำที่ 95.4 %

## 2.5. การจัดกลุ่มเส้นทางที่ใช้ในการเดินทาง

การหาเส้นทางที่นิยมจากเส้นทางทั้งหมดที่ใช้ในการเดินทางระหว่างพื้นที่ ส่วนใหญ่เส้นทางแต่ละกลุ่มจะไม่เหมือนกันทั้งหมดมีความแตกต่างบ้างในบางช่วงแต่เส้นทางส่วนใหญ่เป็นเส้นทางเดียวกัน ดังนั้นจำเป็นต้องทำการจัดกลุ่มเส้นทางที่มีเส้นทางส่วนใหญ่เหมือนกัน โดยการจัดกลุ่มเส้นทางสามารถแยกได้เป็น 3 กลุ่มคือ

### 2.5.1. การจัดกลุ่มโดยอาศัยลำดับของกริดที่ผ่าน

การจัดกลุ่มแบบกริดนี้จะทำการแปลงแนววิถีจากลำดับพิกัดจีพีเอสให้อยู่ในรูปแบบของลำดับกริดที่ผ่านก่อนที่จะนำมาจัดกลุ่มเพื่อให้ง่ายต่อการนำไปจัดกลุ่มและสามารถใช้วิธีการในการหาความคล้ายกันของลำดับกริดไม่ว่าจะเป็นลำดับย่อยร่วมยาวสุด (Longest Common Sub-Sequence) หรือ ระยะทางเลเวนชเตย์น (Levenshtein distance) ซึ่งมีข้อดีที่สามารถจัดกลุ่มเส้นทางใดและเป็นวิธีที่ไม่ซับซ้อน แต่การจัดกลุ่มโดยใช้วิธีนี้ความละเอียดของเส้นทางจะอยู่ที่ขนาดของกริด ถ้าต้องการความละเอียดที่สูงกริดจะต้องมีขนาดเล็กซึ่งจะทำให้ได้ลำดับกริดที่ยาวขึ้นส่งผลให้ความเร็วในการจัดกลุ่มมากขึ้น ตัวอย่างงานวิจัยที่ใช้วิธีการจัดกลุ่มแบบนี้คือ M. Gao et al. [29] ได้ทำการแปลงลำดับพิกัดจีพีเอสเป็นลำดับกริดและใช้ลำดับย่อยร่วมยาวสุดในการหาเส้นทางที่เหมือนกัน,

N. Nandan [30] เสนอวิธีในการวัดความคล้ายของข้อมูลจีพีเอสจากโทรศัพท์มือถือ โดยทำการแปลงข้อมูลจีพีเอสให้อยู่ในรูปกริดและพัฒนาวิธีในการหาความคล้ายจากลำดับกริดที่มีข้อมูลและเวลาที่อยู่ในละกริด โดยอาศัยพื้นฐานจากลำดับย่อยร่วมยาวสุด, M. Lu et al. [31] ทำการแปลงลำดับพิกัดจีพีเอสเป็นลำดับกริดและใช้การจัดกลุ่มแบบลำดับชั้น โดยคำนวณความคล้ายด้วยระยะทางเลเวนชเตย์น

### 2.5.2. การจัดกลุ่มโดยอาศัยแนววิถีจากลำดับจีพีเอส

การจัดกลุ่มโดยอาศัยแนววิถีจากลำดับจีพีเอส [32], [33] เป็นการจัดกลุ่มโดยเปรียบเทียบความคล้ายในรูปร่างของแนววิถีจากเส้นเชื่อมระหว่างลำดับพิกัดจีพีเอสโดยตรง ซึ่งการจัดกลุ่มเส้นทางด้วยวิธีนี้มีข้อดีที่มีความละเอียดของเส้นทางที่สูงและสามารถนำเส้นทางอะไรก็ได้มาจัดกลุ่มและสร้างตัวแทนของเส้นทางในแต่ละกลุ่ม ตัวอย่างที่นำไปประยุกต์เช่นหาเส้นทางอพยพของสัตว์, เส้นทางเดินเรือ, เส้นทางที่คนเดินภายในสถานที่ต่างๆ เป็นต้น

### 2.5.3. การจัดกลุ่มโดยอาศัยลำดับของช่วงถนน

การจัดกลุ่มโดยอาศัยลำดับของช่วงถนนที่ผ่านระหว่างการเดินทาง [34], [35], [36] เป็นการจัดกลุ่มโดยการหาความคล้ายกันของลำดับของช่วงถนนที่แต่ละเส้นทางใช้ในการเดินทาง โดยวิธีวัดความคล้ายสามารถใช้ได้เกี่ยวกับการจัดกลุ่มโดยอาศัยลำดับของกริดที่ผ่านระหว่างการเดินทาง วิธีการจัดกลุ่มโดยอาศัยลำดับของช่วงถนนที่ผ่านระหว่างการเดินทางมีข้อที่ได้เปรียบตรงที่ทราบถนนที่ใช้ซึ่งทำให้สามารถนำข้อมูลนี้ไปต่อยอดได้อีกหลาย

ด้านเช่นหาความหนาแน่นของรถในชั่วโมงต่าง ๆ และหาความเร็วเฉลี่ยบนถนน แต่มีข้อเสียที่จะต้องทำการแปลงข้อมูลพิกัดจีพีเอสเป็นลำดับของช่วงถนนที่ผ่านซึ่งมีความซับซ้อนสูงกว่าวิธีการจัดกลุ่มโดยอาศัยลำดับของกริดที่ผ่านระหว่างการเดินทาง งานวิจัยนี้ได้เลือกใช้การจัดกลุ่มโดยอาศัยลำดับของช่วงถนนที่ผ่านระหว่างการเดินทาง เนื่องจากทำให้ทราบถึงถนนที่ใช้ในแต่ละเส้นทาง โดยหาความคล้ายด้วยระยะทางเลเวนชเตย์น เพราะทราบถึงความแตกต่างระหว่างสองเส้นทาง ซึ่งสามารถใช้ในการพิจารณารวมกลุ่มในกรณีที่เป็นทางขนานกันได้

## 2.6. แผนภาพโวโรนอย (Voronoi Diagram)

เนื่องมาจากความหนาแน่นของเสากระจายสัญญาณโทรศัพท์ในแต่ละพื้นที่ไม่เท่ากัน และทิศทางให้การให้บริการต่างกัน งานวิจัยนี้้นำแผนภาพโวโรนอยมาใช้ในการหาพื้นที่ที่เสากระจายสัญญาณโทรศัพท์แต่ละเสา

แผนภาพโวโรนอยเป็นวิธีการสร้างรูปหลายเหลี่ยมล้อมรอบจุดต้นกำเนิดเพื่อแบ่งพื้นที่ออกเป็นพื้นที่ย่อย โดยพื้นที่ย่อยรอบจุดกำเนิดแสดงถึงพื้นที่อิทธิพลของจุดกำเนิดนั้น [37] ทุกตำแหน่งภายในรูปหลายเหลี่ยมจะอยู่ใกล้จุดกำเนิดภายในรูปหลายเหลี่ยมนั้นมากกว่าจุดกำเนิดอื่น จากภาพที่ 3 เป็นรูปที่แสดงแผนภาพโวโรนอยของจุดกำหนดที่เป็นสีแดง โดยจุดใด ๆ ในพื้นที่สีเขียวจะอยู่ใกล้กับจุดสีที่อยู่ภายในสีเขียวมากกว่าจุดสีแดงนอกจุดเขียว

กำหนดให้  $P$  เป็นเซตของจุดกำเนิด

$d(x,y)$  เป็นระยะทางที่สั้นที่สุดจากจุด  $x$  ไปยังจุด  $y$

$\overline{xy}$  เป็นส่วนของเส้นจาก  $x$  ไปยัง  $y$

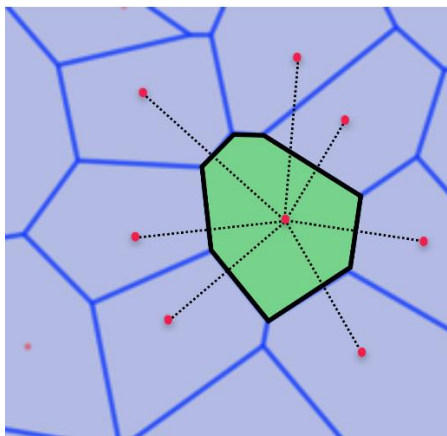
พื้นที่โวโรนอยของจุด  $p$  กำหนดโดย

$$VR(p) = \bigcap_{p,q \in P, q \neq p} \{x | d(p,x) \leq d(q,x)\}$$

แผนภาพโวโรนอยของเซตต้นกำเนิด  $P$  กำหนดโดย

$$V(P) = \bigcup_{p,q \in P, q \neq p} \overline{VR(p)} \cap \overline{VR(q)}$$





ภาพที่ 3 แผนภาพโวโรนอยของจุดกำหนดที่เป็นสีแดง

## 2.7. การจัดกลุ่มแบบลำดับชั้น (Hierarchical Clustering)

การจัดกลุ่มแบบลำดับชั้นเป็นการจัดกลุ่มข้อมูลรูปแบบหนึ่งที่ใช้ระยะห่างหรือความคล้ายของข้อมูลแต่ละคู่เป็นข้อมูลตั้งต้น โดยผลที่ได้จะเป็นกลุ่มที่จำนวนกลุ่มขึ้นอยู่กับความคล้ายที่ผู้ใช้เลือกเพื่อทำการแยกกลุ่ม ซึ่งสามารถพิจารณาจำนวนกลุ่มได้จากลำดับชั้นของความคล้ายของข้อมูล [38] ซึ่งเป็นวิธีที่มีข้อดีคือไม่จำเป็นต้องใส่จำนวนกลุ่มและสามารถศึกษาระยะห่างหรือความคล้ายของกลุ่มภายหลังเพื่อใช้ตัดสินใจในการหาจำนวนกลุ่มที่เหมาะสม เทคนิคการจัดกลุ่มแบบลำดับชั้นสามารถแบ่งออกได้เป็น 2 กลุ่มย่อย คือ

- Agglomerative Hierarchical Cluster

Agglomerative Hierarchical Cluster เป็นการจัดกลุ่มโดยเริ่มแบ่งกลุ่มออกเป็น  $n$  กลุ่มตามจำนวนข้อมูลแล้วรวมกลุ่มที่คล้ายกันเข้าด้วยกันจนเหลือเพียงกลุ่มเดียว

- 1) กำหนดให้หนึ่งอ็อบเจกต์เป็นหนึ่งกลุ่ม
- 2) ทำการรวมกลุ่มที่คล้ายกันมากที่สุดจากกลุ่มที่มีในปัจจุบัน
- 3) ทำกระบวนการที่ 2 ซ้ำจนกว่าจะเหลือเพียงกลุ่มเดียว

- Divisive Hierarchical Cluster

Divisive Hierarchical Cluster เป็นการจัดกลุ่มโดยเริ่มจากรวมข้อมูลทั้งหมดเป็นกลุ่มเดียว แล้วแบ่งกลุ่มเป็นสองกลุ่มที่มีความคล้ายกันน้อยที่สุด จนแต่ละกลุ่มมีข้อมูลเพียงตัวเดียว

- 1) กำหนดให้ทุกอ็อบเจกต์เป็นหนึ่งกลุ่ม
- 2) แบ่งกลุ่มที่มีความคล้ายกันน้อยที่สุดออกเป็นสองกลุ่ม
- 3) ทำกระบวนการที่สองซ้ำจนกว่าจะได้หนึ่งอ็อบเจกต์เป็นหนึ่งกลุ่ม

เงื่อนไขในการวัดความคล้าย

- UPGMA (Unweighted Pair-Group Method Using Arithmetic Average) วัดความคล้ายของกลุ่มสองกลุ่มจากผลเฉลี่ยของระยะห่างของทุกคู่ข้อมูลระหว่างสองกลุ่ม

$$\frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a, b)$$

โดย A,B เป็นกลุ่มที่พิจารณาเงื่อนไขการรวมกลุ่ม โดย  $A \neq B$

- Single Linkage วัดความคล้ายของกลุ่มสองกลุ่มจากรยะห่างจากรยะห่างที่น้อยที่สุดของข้อมูลระหว่างสองกลุ่ม

$$\min \{d(a, b) : a \in A, b \in B\}$$

โดย A,B เป็นกลุ่มที่พิจารณาเงื่อนไขการรวมกลุ่ม โดย  $A \neq B$

- Complete Linkage วัดความคล้ายของกลุ่มสองกลุ่มจากรยะห่างจากรยะห่างที่มากที่สุดของข้อมูลระหว่างสองกลุ่ม

$$\max \{d(a, b) : a \in A, b \in B\}$$

โดย A,B เป็นกลุ่มที่พิจารณาเงื่อนไขการรวมกลุ่ม โดย  $A \neq B$

- Median วัดความคล้ายของกลุ่มสองกลุ่มจากรยะห่างจากค่ามัธยฐานของข้อมูลแต่ละกลุ่ม

- Centroid วัดความคล้ายของกลุ่มสองกลุ่มจากรยะห่างจากค่าเซนทรอยด์ของข้อมูลแต่ละกลุ่ม

เนื่องจากผลจากการจัดกลุ่มโดยการจัดกลุ่มแบบลำดับชั้นสามารถนิยามความห่างของกลุ่มเสาสัญญาณแต่ละกลุ่มได้ งานวิจัยนี้จึงนำการจัดกลุ่มแบบลำดับชั้นมาใช้ในการจัดกลุ่มของเสาสัญญาณ

## 2.8. Haversine Formula

ในขั้นตอนการจัดกลุ่มโดยการจัดกลุ่มแบบลำดับขั้นจำเป็นต้องนิยามค่าความห่างระหว่างข้อมูลก่อน ซึ่งงานวิจัยนี้้นำการจัดกลุ่มแบบลำดับขั้นมาจัดกลุ่มเสากกระจายสัญญาณ ดังนั้นต้องหาระยะห่างบนพื้นผิวโลกระหว่างเสากกระจายสัญญาณ

Haversine formula เป็นการหาระยะทางที่สั้นที่สุดระหว่างจุดสองจุดบนพื้นผิวทรงกลม วัดไปตามพื้นผิวทรงกลม โดยอาศัยค่าละติจูดและลองจิจูดของจุดสองจุด [39]

$$d = 2R \operatorname{atan}(\sqrt{a}, \sqrt{1-a})$$

$$a = \sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1) \times \cos(\varphi_2) \times \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)$$

โดย  $\varphi$  คือค่าละติจูดในหน่วยเรเดียน

$\lambda$  คือค่าลองจิจูดในหน่วยเรเดียน

R คือรัศมีของโลก = 6,371 กิโลเมตร

## 2.9. Levenshtein Distance

ขั้นตอนการจัดกลุ่มเส้นทางที่นิยมในการเดินทางระหว่างพื้นที่ผู้วิจัยเลือกใช้การเปรียบเทียบลำดับของช่วงถนนที่เดินทาง ดังนั้นจำเป็นต้องหาวิธีในการนิยามค่าความห่างระหว่างลำดับของช่วงถนนที่เดินทาง โดยวิธีหนึ่งที่เป็นที่นิยมคือใช้วิธีการ Levenshtein distance

Levenshtein distance เป็นวิธีในการวัดค่าความแตกต่างของ 2 สายอักขระ โดยผลที่ได้เป็นจำนวนครั้งที่น้อยที่สุดที่ทำการแก้ไข โดยการเพิ่ม, ลบ หรือแทนที่อักขระ ภายใน 2 สายอักขระ [40] ตัวอย่างสายอักขระ GGTA กับ GTC ที่มีค่า Levenshtein distance เป็น 3 คือ

1. เพิ่ม G ไปหน้าสุดของสายอักขระที่ 2  $\_GTC \rightarrow GGTC$  (GGTA | GGTC)
2. เปลี่ยนจากตัว A เป็นตัว C ในสายอักขระที่ 1  $GGTAA \rightarrow GGTC A$  (GGTA | GGTC)
3. ลบตัว A ในสายอักขระที่ 1 ออก  $GGTCA \rightarrow GGTC$  (GGTC | GGTC)

โดยขั้นตอนของ Levenshtein distance สามารถสรุปได้ตามภาพที่ 4

```

Input:  $S_1, S_2$  are sequence of words
Output: minimum number of edits

 $D(0,0) = 0$ 
for (i = 1 to  $|S_1|$ )
     $D(i,0) = i$ 
for (j = 1 to  $|S_2|$ )
     $D(0,j) = j$ 

for (i = 1 to  $|S_1|$ )
    for (j = 1 to  $|S_2|$ )
         $D(i,j) = \min \begin{cases} D(i-1,j) + 1 \\ D(i,j-1) + 1 \\ D(i-1,j-1) + (\text{if } (s_1[i] = s_2[j]) \text{ then } 0 \text{ else } 1) \end{cases}$ 
return( $D(|S_1|, |S_2|)$ )

```

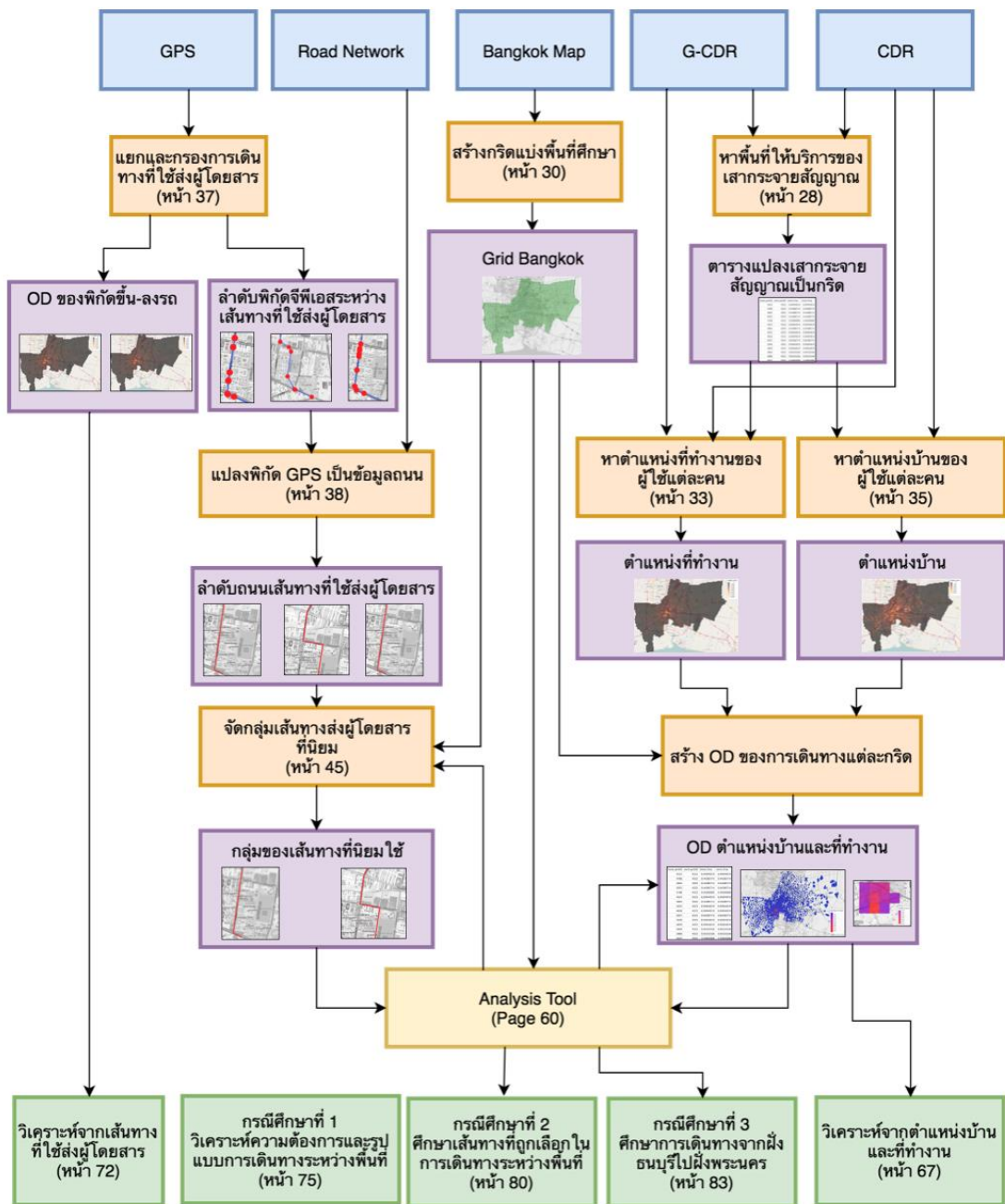
ภาพที่ 4 ขั้นตอนของ Levenshtein distance

ในงานวิจัยนี้ได้เปลี่ยนจากการวัดค่าความแตกต่างของ 2 สายอักขระ เป็นการวัดค่าความแตกต่างของ 2 สายลำดับถนน โดยการแก้ไขจะเป็นเพิ่ม ลบ หรือแทนที่ ช่วงของถนนที่อยู่ในรูปแบบรหัสช่วงถนน

### บทที่ 3

#### แนวคิดและวิธีการดำเนินงาน

งานวิจัยชิ้นนี้มีแนวคิดที่จะสร้างเครื่องมือสำหรับวิเคราะห์การเคลื่อนที่ของประชากรระหว่างพื้นที่ที่คนเดินทางไปเป็นประจำ ใช้ข้อมูลตั้งต้นจากข้อมูลจีพีเอสของรถแท็กซี่และข้อมูลซีดีอาร์ ซึ่งงานวิจัยนี้อาศัยความแตกต่างของลักษณะของข้อมูลจากแต่ละแหล่ง โดยจะอาศัยข้อได้เปรียบด้านขนาดประชากรขนาดใหญ่ของซีดีอาร์นำมาสกัดพื้นที่สำคัญของแต่ละบุคคลซึ่งประกอบด้วยบ้านและสถานที่ทำงานที่แต่ละคนเดินทางไปเป็นประจำจากการใช้งานโทรศัพท์เคลื่อนที่ ผลที่ได้ทำให้เห็นพื้นที่ภายในเมืองที่คนเดินทางไปมากในแต่ละช่วงเวลาทำให้ทราบถึงปริมาณความต้องการในการเดินทางระหว่างพื้นที่ และใช้ข้อมูลตั้งต้นจากจีพีเอสในการวิเคราะห์เส้นทางที่นิยมใช้ในการเดินทางระหว่างพื้นที่ งานวิจัยนี้ได้ออกแบบกระบวนการในการวิจัยเพื่อพัฒนาเครื่องมือสำหรับวิเคราะห์การเคลื่อนที่ของประชากรระหว่างพื้นที่ที่คนเดินทางไปเป็นประจำตามที่แสดงในภาพที่ 5 โดยสีน้ำเงินแสดงข้อมูลตั้งต้นที่ใช้ในการวิจัย สีส้มเป็นขั้นตอนวิธีในการแปลงหรือสกัดข้อมูลเพื่อนำไปใช้ในการวิเคราะห์ สีม่วงแสดงผลลัพธ์ที่ได้จากขั้นตอนวิธีต่าง ๆ สีเหลืองแสดงเครื่องมือที่ใช้ในการวิเคราะห์การเคลื่อนที่ของประชากรระหว่างบ้านและที่ทำงาน และสีเขียวแสดงการวิเคราะห์ข้อมูลที่ผ่านการดำเนินการในขั้นตอนวิธีต่าง ๆ และกรณีศึกษาที่วิเคราะห์จากเครื่องมือที่สร้างขึ้น



ภาพที่ 5 กระบวนการพัฒนาเครื่องมือสำหรับวิเคราะห์การเคลื่อนที่ของประชากร

### 3.1. ข้อมูลตั้งต้น

การวิเคราะห์การเคลื่อนที่ของประชากรข้อมูลตั้งต้นถือเป็นสิ่งที่มีความสำคัญอย่างมาก โดยในงานวิจัยนี้ใช้ข้อมูลจาก 4 แหล่งดังนี้

#### 3.1.1. ข้อมูลจีพีเอส

ข้อมูลจีพีเอสจากรถแท็กซี่ที่ให้บริการอยู่ในกรุงเทพฯที่รวบรวมโดยมูลนิธิศูนย์ข้อมูลจราจรอัจฉริยะไทย ในช่วงเวลา 1 กุมภาพันธ์ 2559 ถึง 31 ตุลาคม 2559 ข้อมูลที่ได้รับประกอบไปด้วย

- vehicle\_id คือรหัสประจำตัวของรถแท็กซี่
- valid\_flag คือตัวบ่งชี้ค่าเครื่องรับจีพีเอสเห็นดาวเทียมครบตามมาตรฐานหรือไม่ ซึ่งจะมีผลต่อความถูกต้องของตำแหน่ง
- latitude คือตำแหน่งละติจูดของรถแท็กซี่ที่อยู่
- longitude คือตำแหน่งลองจิจูดที่รถแท็กซี่อยู่
- speed คือความเร็วในหน่วย กม./ชม.
- timestamp คือวันและเวลาของข้อมูลที่ส่งมา
- passenger\_flag คือตัวบ่งชี้สถานะไฟแสดงรถว่างที่ติดอยู่บนหน้ารถ ดังนั้น passenger\_flag เป็น 1 แสดงว่ารถว่าง, เป็น 0 แสดงว่ามีผู้โดยสาร
- engine\_flag เป็นตัวบ่งชี้สถานะเครื่องยนต์ โดยเป็น 1 คือเครื่องยนต์ติด เป็น 0 คือดับเครื่องยนต์

ความถี่ในการส่งข้อมูลขึ้นอยู่กับสถานะของเครื่องยนต์ โดยถ้า engine\_flag เป็น 1 จะส่งข้อมูลทุก 1 นาที ถ้า engine\_flag เป็น 0 จะส่งข้อมูลทุก 3 นาที

vehicle_id	valid_flag	heading	timestamp	latitude	longitude	speed	passenger_flag	engine_flag
AK2756	1	123	2016-02-01 00:30:58	13.85271	100.6970	0	1	0
AK2756	1	123	2016-02-01 00:33:58	13.85271	100.6970	0	1	0
AK2756	1	123	2016-02-01 00:36:58	13.85271	100.6970	0	1	0
AK2756	1	123	2016-02-01 00:39:58	13.85271	100.6970	2	1	1
AK2756	1	273	2016-02-01 00:40:58	13.85282	100.6941	39	0	1
AK2756	1	156	2016-02-01 00:41:58	13.84758	100.6956	58	0	1
AK2756	1	176	2016-02-01 00:43:58	13.84006	100.6966	32	0	1
AK2756	1	352	2016-02-01 00:44:58	13.83997	100.6964	20	0	1
AK2756	1	245	2016-02-01 00:45:58	13.83835	100.6918	32	1	1

ตารางที่ 2 ตัวอย่างข้อมูลดิบของจีพีเอสที่ติดอยู่ในรถแท็กซี่

### 3.1.2. ข้อมูลซีดีอาร์

ข้อมูลซีดีอาร์เป็นข้อมูลการใช้งานโทรศัพท์ที่ถูกรวบรวมโดยผู้ให้บริการเครือข่ายโทรศัพท์มือถือรายใหญ่ของประเทศไทยในช่วงเวลา 1 สิงหาคม 2559 ถึง 31 ตุลาคม 2559 งานวิจัยนี้สนใจข้อมูลซีดีอาร์ของผู้ใช้ในเขตกรุงเทพมหานคร ประกอบด้วย

- mobile\_no คือเบอร์โทรศัพท์ที่ถูกเข้ารหัส ที่ติดต่อกับเสากระจายสัญญาณ
- date\_time คือเวลาที่ใช้งาน
- call\_type คือประเภทการใช้งาน มี Incoming/Outgoing Voice และ Incoming/Outgoing SMS
- duration ระยะเวลาที่ใช้
- system\_network คือระบบและคลื่นที่ใช้งาน
- latitude คือตำแหน่งละติจูดของเสากระจายสัญญาณที่ติดต่อ
- longitude คือตำแหน่งลองจิจูดของเสากระจายสัญญาณที่ติดต่อ

mobile_no	date_time	call_type	duration	system_network	latitude	longitude
NNNNNNNNNN	2015-09-01 07:01:33	Incoming_SMS	1	2G_900	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01 07:01:43	Outgoing_Voice	15	4G_2100	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01 07:05:51	Incoming_Voice	6	4G_2100	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01 07:14:01	Incoming_Voice	75	3G_1800	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01 07:20:31	Incoming_Voice	13	3G_1800	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01 07:22:51	Incoming_SMS	1	2G_900	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01 07:46:12	Outgoing_Voice	42	4G_2100	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01 07:54:53	Incoming_SMS	1	2G_900	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01 07:55:01	Outgoing_Voice	33	3G_1800	XX.XXXXXX	YYY.YYYYYY

ตารางที่ 3 ตัวอย่างข้อมูลดิบของซีดีอาร์

### 3.1.3. ข้อมูลจี-ซีดีอาร์

ข้อมูลจี-ซีดีอาร์ (GGSN Charging Detail Record) คือข้อมูลเกี่ยวกับการใช้อินเทอร์เน็ตที่ถูกรวบรวมโดย Gateway GPRS Support Node (GGSN) ข้อมูลจะเป็นข้อมูลที่ผ่านการสรุปรายวันของผู้ใช้แต่ละคนในแต่ละเสากระจายสัญญาณ ในช่วงเวลา 01 สิงหาคม 2559 ถึง 31 ตุลาคม 2559 ของผู้ใช้ในเขตกรุงเทพมหานคร ประกอบด้วย



- mobile\_no คือเบอร์โทรศัพท์ที่ถูกเข้ารหัส
- Date วันที่ทำการใช้งาน
- upload volume คือผลรวมปริมาณการใช้ข้อมูลฝั่งบรรจุขึ้นตลอดวัน มีหน่วยเป็น KB
- download volume คือผลรวมปริมาณการใช้ข้อมูลฝั่งบรรจุลงตลอดวัน มีหน่วยเป็น KB
- latitude คือตำแหน่งละติจูดของเสากระจายสัญญาณที่ติดต่อ
- longitude คือตำแหน่งลองจิจูดของเสากระจายสัญญาณที่ติดต่อ

mobile_no	date	upload_volume	download_volume	System_Network	Latitude	Longitude
NNNNNNNNNN	2015-09-01	234	2352	3G_2100	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01	10	94	4G_2100	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01	1256	345643	4G_2100	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01	34	125	3G_1800	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01	5	8	4G_2100	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01	32	68	3G_1800	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-01	6451	423	3G_2100	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-02	123	9745	4G_2100	XX.XXXXXX	YYY.YYYYYY
NNNNNNNNNN	2015-09-02	342	45456	3G_2100	XX.XXXXXX	YYY.YYYYYY

ตารางที่ 4 ตัวอย่างข้อมูลดิบของจี-ซีดีอาร์

### 3.1.4. ข้อมูลถนน

ข้อมูลถนนในรูปแบบ shape file ประกอบด้วยข้อมูล 2 ส่วนคือเวกเตอร์ของถนนอ้างอิงโดยลำดับพิกัดละติจูดและลองจิจูดโดยแต่ละเส้นจะแสดงถึงการเชื่อมต่อกันของถนนจากเส้นหนึ่งไปยังอีกเส้นหนึ่ง ตัวอย่างจากภาพที่ 5 แสดงตัวอย่างเวกเตอร์ถนนบริเวณแยกปทุมวันโดยจุดสีแดงเป็นพิกัดเริ่มต้นและพิกัดสิ้นสุดของแต่ละเวกเตอร์ จุดสีน้ำเงินเป็นพิกัดระหว่างเวกเตอร์เพื่อเปลี่ยนทิศทางของถนน พิกัดเริ่มต้นและพิกัดสิ้นสุดเป็นตำแหน่งเดียวกันแสดงว่าเวกเตอร์ทั้งสองเชื่อมต่อกัน ซึ่งในรูปข้างต้นมีทั้งหมด 27 เส้นเพื่อใช้อธิบายการเชื่อมต่อของถนนบริเวณแยกปทุมวัน ข้อมูลอีกส่วนเป็นข้อมูลที่ใช้อธิบายเวกเตอร์ถนนประกอบด้วย

- ROADKEY คือรหัสถนน
- NAME\_T คือชื่อถนนภาษาไทย
- NAME\_E คือชื่อถนนภาษาอังกฤษ
- LANE คือจำนวนช่องทางเดินรถ
- ONEWAY คือตัวบ่งชี้ว่าถนนนั้นเป็นถนนที่แล่นรถทางเดียวหรือไม่ โดยจะเป็น
- FT เมื่อถนนส่วนนั้นเป็นถนนที่แล่นรถทางเดียว



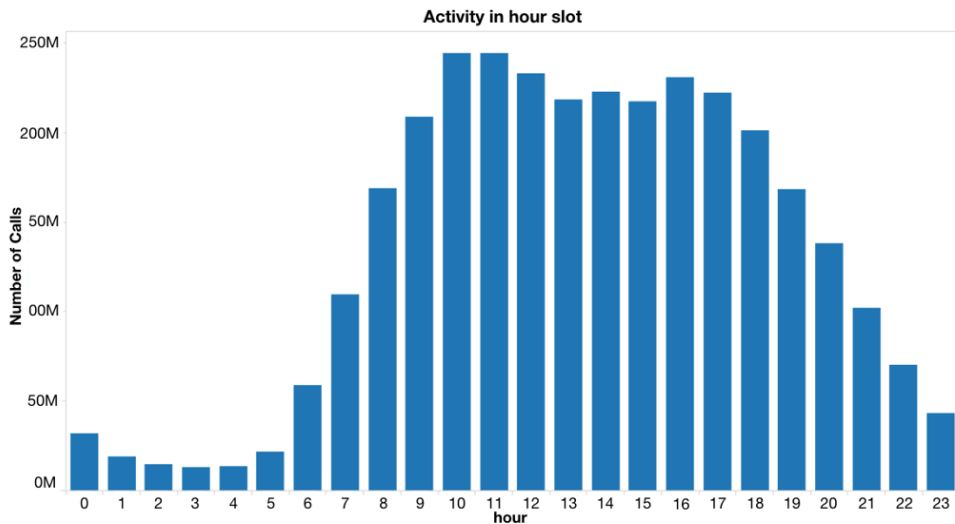
ภาพที่ 6 ตัวอย่างเวกเตอร์ถนนบริเวณแยกปทุมวัน

## 3.2. การสำรวจข้อมูล

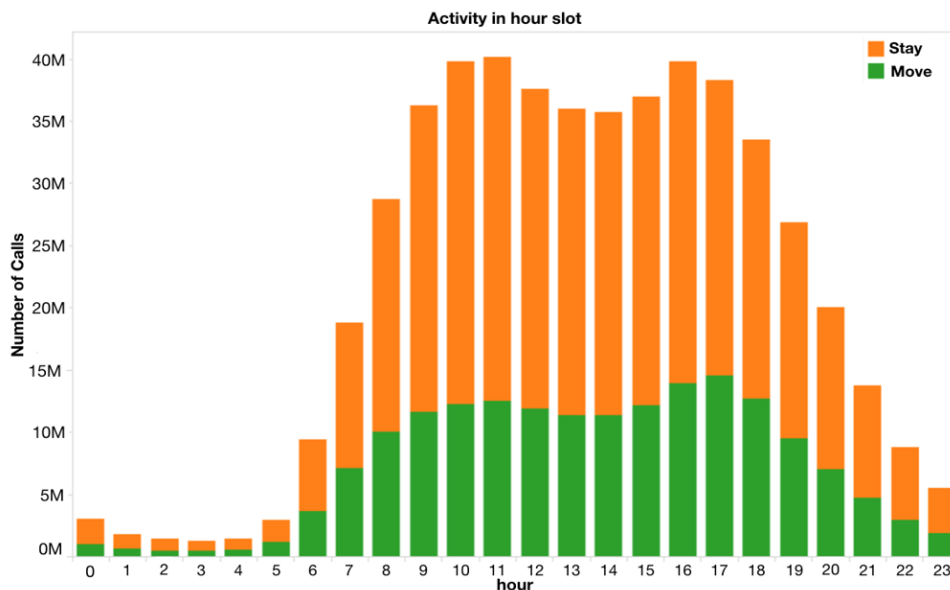
### 3.2.1. ข้อมูลซีดีอาร์และจี-ซีดีอาร์

จากข้อมูลซีดีอาร์พบว่า มีผู้ใช้งานทั้งหมด 4,340,833 คน มีการใช้งานเฉลี่ยต่อคน อยู่ที่ 8.04 ครั้งต่อวัน เมื่อพิจารณาจำนวนการใช้งานในแต่ละช่วงเวลา (ภาพที่ 7), พบว่า จำนวนการใช้งานมากในช่วงกลางวันโดยจะมีการใช้งานที่พุ่งสูงขึ้นตั้งแต่ 6 โมงเช้า จนไปสูงที่สุดในชั่วโมงที่ 10 และมีการใช้งานที่สูงต่อเนื่องไปจนถึงชั่วโมงที่ 17 ก่อนที่การใช้งานจะลดลงอย่างต่อเนื่อง โดยช่วงเวลา 01:00 – 04:00 น. เป็นช่วงที่มีการใช้งานน้อยที่สุดในช่วงเวลา 09:00 - 16:00 น. มีการโทรเฉลี่ยต่อคนอยู่ที่ 4.60 ครั้งต่อวัน ช่วงเวลา 19:00 – 05:00 น. มีการโทรเฉลี่ยต่อคนอยู่ที่ 1.61 ครั้งต่อวัน เมื่อพิจารณาในมุมมองการเคลื่อนที่ของผู้ใช้จากข้อมูลซีดีอาร์ 2 บันทึกที่ติดกันตามวิธีการจากงานวิจัย [17] พบว่ามีสัดส่วนคนที่อยู่กับที่กับคนที่กำลังเดินทางมีสัดส่วนใกล้เคียงกันตลอด 24 ชั่วโมง (ภาพที่ 8) ทำให้ทราบถึงข้อมูลซีดีอาร์ในช่วงเวลากลางคืนเป็นข้อมูลที่รวมระหว่างคนที่กำลังมีกิจกรรมภายนอก

บ้านตอนกลางคืนกับคนที่ใช้โทรศัพท์จากที่บ้าน ทำให้มีความท้าทายในการหาตำแหน่งบ้าน จากข้อมูลซีดีอาร์ที่มีน้อยในช่วงเวลากลางคืนซึ่งมีข้อมูลที่ปะปนกันระหว่างคนที่มิกิจกรรม ตอนกลางคืนกับการใช้งานจากที่บ้าน



ภาพที่ 7 กราฟแสดงค่าของสถิติความถี่จำนวนคนที่ใช้งานในแต่ละชั่วโมงเฉลี่ยต่อวัน

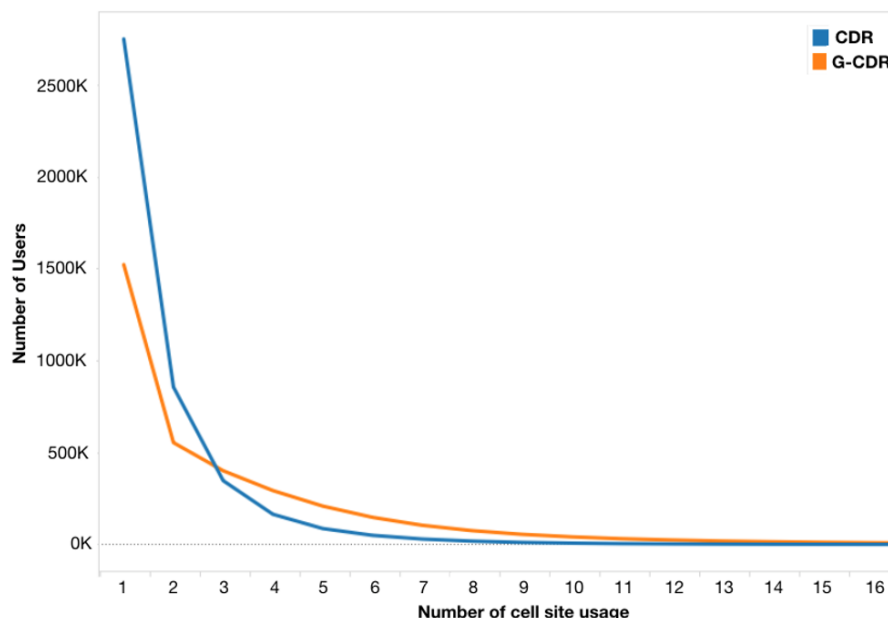


ภาพที่ 8 กราฟแสดงค่าของสถิติความถี่การเคลื่อนที่และอยู่กับที่ของผู้ใช้งาน

#### ข้อมูล 2 การใช้งานที่ติดกัน

พบว่าจากผู้ใช้งานทั้งหมด 4,340,833 คน มีผู้ใช้งานการใช้งานอินเทอร์เน็ตทั้งหมด 3,564,353 คน โดยมีการติดต่อกับเสากระจายสัญญาณเฉลี่ย 29.83 ครั้งต่อคนต่อวัน ซึ่งมีการใช้งานเฉลี่ยต่อวันมากกว่าข้อมูลซีดีอาร์ที่เป็นข้อมูลการโทรและ SMS แต่ข้อมูลจี-ซีดีอาร์ถูกรวบรวมเป็นข้อมูลสรุปการใช้งานรายวันของแต่ละคนในแต่ละเสากระจายสัญญาณโดยรวม

ค่าอัปโหลดและดาวน์โหลดของแต่ละวัน จากกราฟแสดงค่าของสถิติความถี่จำนวนเสาเฉลี่ยที่ผู้ใช้งานในแต่ละวัน (ภาพที่ 9) พบว่าจำนวนเสาที่ใช้งานจากข้อมูลจี-ซีดีอาร์สูงกว่าข้อมูลซีดีอาร์

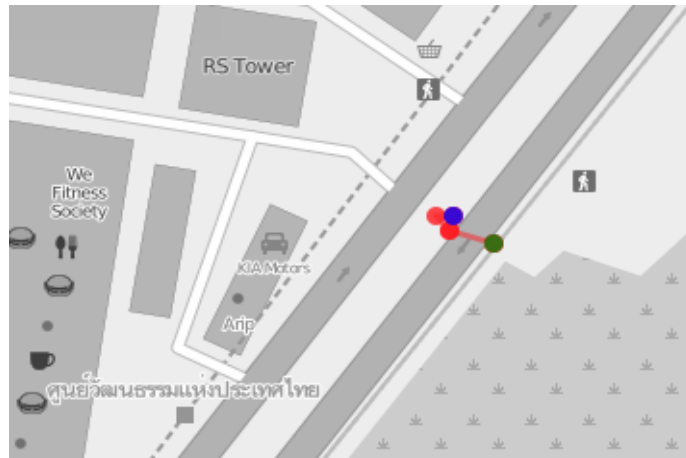


ภาพที่ 9 กราฟแสดงค่าของสถิติความถี่จำนวนเสาเฉลี่ยที่ผู้ใช้ใช้งานในแต่ละวัน

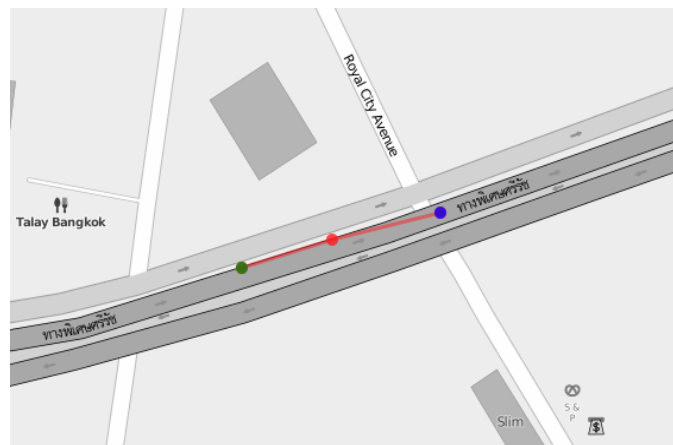
### 3.2.2. ข้อมูลจีพีเอส

ข้อมูลจีพีเอสที่ได้รับจากมูลนิธิศูนย์ข้อมูลจราจรอัจฉริยะไทย ตั้งแต่วันที่ 1 กุมภาพันธ์ 2559 ถึง 31 ตุลาคม 2559 ของแท็กซี่ทั้งหมด 3,965 คัน โดยเมื่อทำการกรองให้เหลือเฉพาะรถที่มีการให้บริการในกรุงเทพและกรองข้อมูลที่มีความผิดปกติออกเหลือ 2,386 คัน มีข้อมูลที่ให้บริการในกรุงเทพทั้งสิ้น 340,628,394 พิกัด

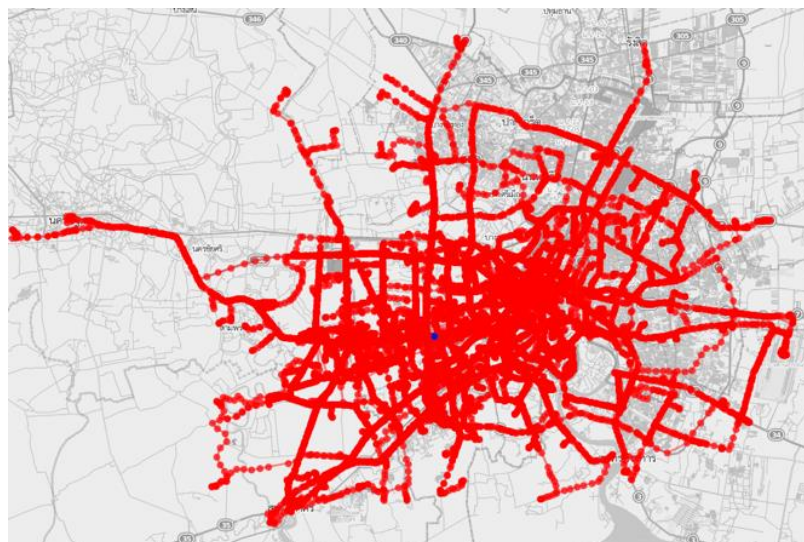
เมื่อพิจารณาการเดินทางเฉพาะการเดินทางที่ใช้ส่งผู้โดยสารโดยการพิจารณาจากการเปลี่ยนตัวบ่งชี้สถานะไฟแสดงรถว่างที่ติดอยู่หน้ารถ (passenger\_flag) โดยเริ่มการเดินทางเมื่อตัวบ่งชี้สถานะไฟแสดงรถว่างที่ติดหน้ารถเปลี่ยนจาก 1 เป็น 0 และสิ้นสุดการเดินทางส่งผู้โดยสารเมื่อสถานะไฟแสดงรถว่างที่ติดหน้ารถเปลี่ยนจาก 0 เป็น 1 พบว่าสามารถแยก 3,715,275 เที่ยวที่มีการเดินทางส่งผู้โดยสารภายในกรุงเทพฯ แต่จากการตรวจสอบของการเดินทางส่งผู้โดยสารที่แยกได้พบว่ามีข้อผิดพลาดอยู่หลายจุดเช่น การเดินทางส่งโดยสารที่ใช้ระยะทางสั้นเกินหรือยาวผิดปกติ, มีความผิดพลาดของตำแหน่ง, อยู่ตำแหน่งเดิมเป็นเวลานานผิดปกติ หรือมีการดับเครื่องย่นระหว่างส่งผู้โดยสาร โดยแสดงตัวอย่างบางกรณีที่พบในภาพที่ 10 ดังนั้นจำเป็นต้องมีกระบวนการกรองข้อมูลการเดินทางที่เสียออก



(a) ตัวอย่างการเดินทางที่ขึ้นและลงในบริเวณเดียวกัน

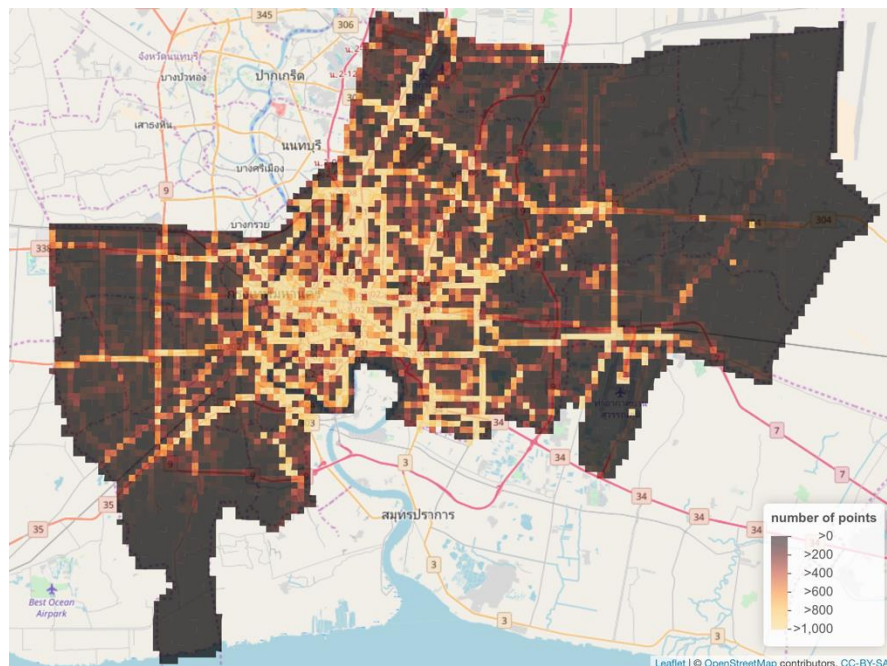


(b) ตัวอย่างการเดินทางที่มีระยะทางรวมสั้นมาก



(c) ตัวอย่างการเดินทางที่มีระยะทางยาวและกระจายตัวผิดปกติโดยจุดขึ้นลงรถอยู่ที่บริเวณเดียวกัน  
ภาพที่ 10 ตัวอย่างการเดินทางเพื่อส่งผู้โดยสารจากพิกัดสี่เหลี่ยมไปยังพิกัดสี่เหลี่ยมที่ผิดปกติ

สิ่งที่สำคัญของข้อมูลจีพีเอสที่มีความจำเป็นต่อการนำไปใช้ในการหาเส้นทางที่นิยมในการเดินทางระหว่างพื้นที่คือการกระจายตัวของรถแท็กซี่ ภาพที่ 11 แสดงความหนาแน่นจำนวนพิกัดจีพีเอสของข้อมูลในกริดขนาด  $500 \times 500$  เมตร ที่อยู่ในพื้นที่ศึกษาจากข้อมูล 1 สัปดาห์ระหว่าง 1 สิงหาคม 2559 ถึง 7 สิงหาคม 2559 พบว่ามีความหนาแน่นสูงบริเวณกรุงเทพมหานครชั้นในและบริเวณถนนเส้นหลักของเขตชานเมือง ซึ่งจากข้อมูลนี้เป็นปัญหาที่ทำให้ต้องออกแบบให้สามารถกำหนดพื้นที่ที่ใช้ศึกษาให้มีขนาดใหญ่ในเขตชานเมืองและสามารถกำหนดพื้นที่ที่ใช้ศึกษาให้มีขนาดเล็กในเขตตัวเมือง

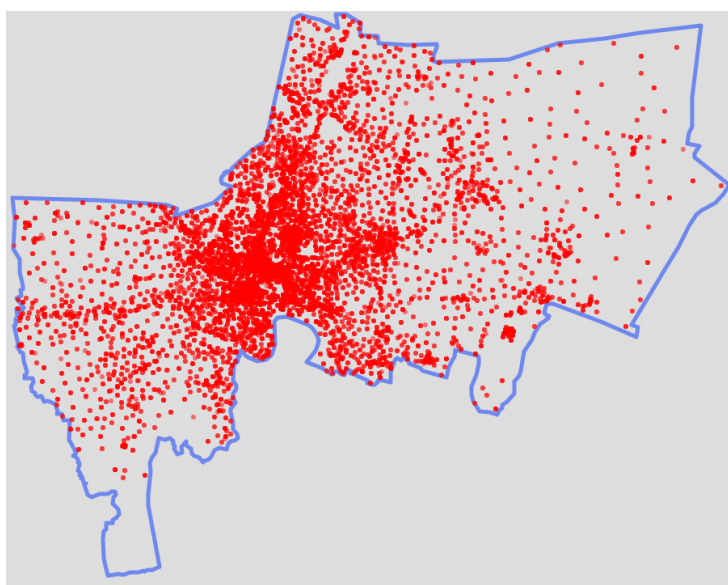


ภาพที่ 11 ความหนาแน่นจำนวนพิกัดจีพีเอสในกริดขนาด  $500 \times 500$  เมตร



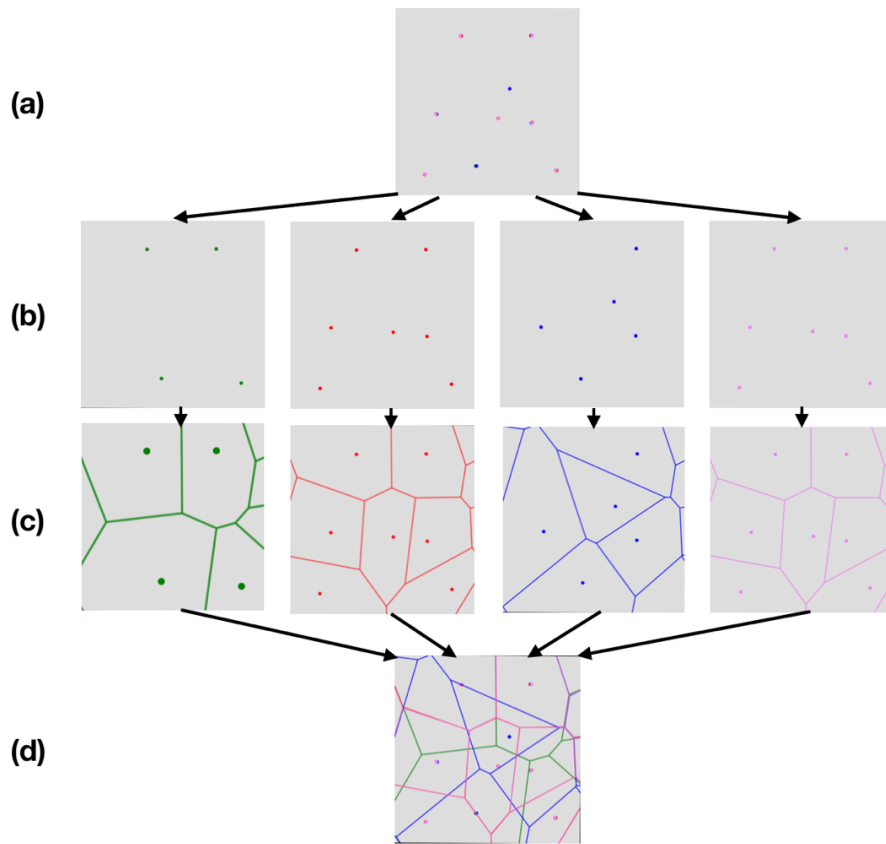
### 3.3. การหาพื้นที่ให้บริการของเสากกระจายสัญญาณ

ตำแหน่งที่ได้มาจากจี-ซีดีอาร์และซีดีอาร์ทราบเพียงตำแหน่งของเสากกระจายสัญญาณที่ใช้ ไม่ทราบถึงพื้นที่ให้บริการของแต่ละเสากกระจายสัญญาณ โดยพื้นที่ให้บริการของเสากกระจายสัญญาณไม่เท่ากันขึ้นอยู่กับความหนาแน่นของการใช้งานในแต่ละพื้นที่ ภาพที่ 12 แสดงตำแหน่งเสากกระจายสัญญาณในพื้นที่กรุงเทพมหานครพบว่าเสากกระจายสัญญาณจะมีความหนาแน่นสูงในเขตเมืองและมีความหนาแน่นต่ำในเขตชานเมือง การหาพื้นที่ให้บริการของเสากกระจายสัญญาณสามารถหาจากแผนภาพโวโรนอย (Voronoi) ของเสากกระจายสัญญาณ

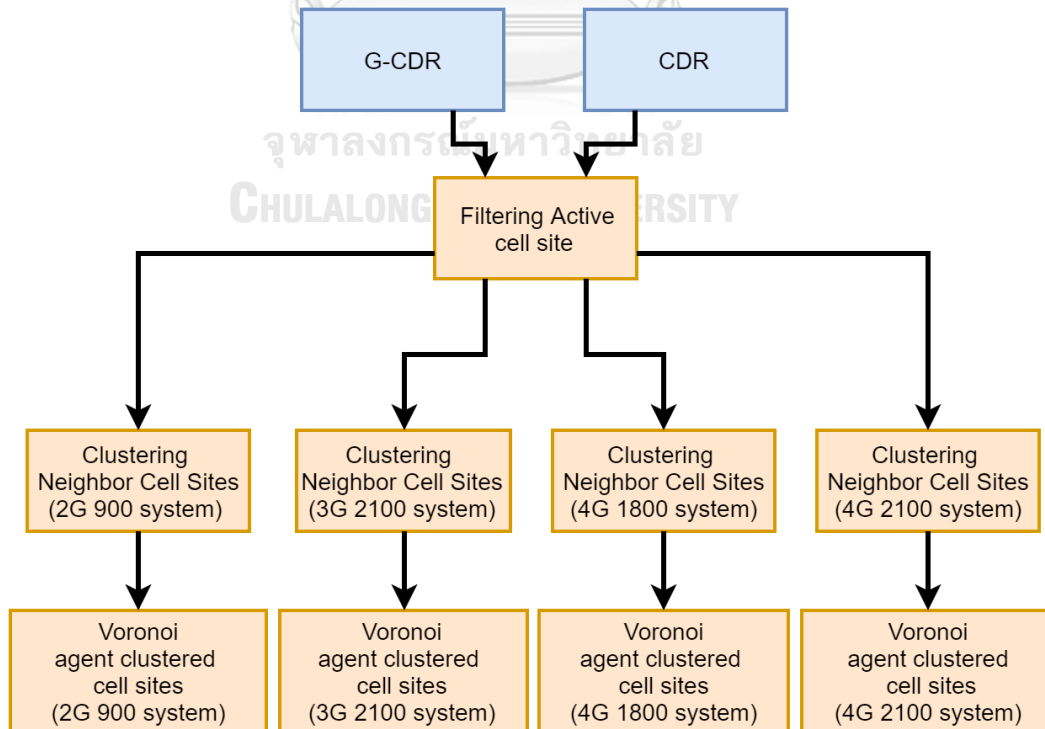


ภาพที่ 12 ตำแหน่งของเสากกระจายสัญญาณในพื้นที่ศึกษา

แผนภาพโวโรนอยเป็นวิธีการสร้างรูปหลายเหลี่ยมล้อมรอบจุดต้นกำเนิดเพื่อแบ่งพื้นที่ออกเป็นพื้นที่ย่อย โดยพื้นที่ย่อยรอบจุดกำเนิดแสดงถึงพื้นที่อิทธิพลของจุดกำเนิดนั้น ทุกตำแหน่งภายในรูปหลายเหลี่ยมจะอยู่ใกล้จุดกำเนิดภายในรูปหลายเหลี่ยมนั้นมากกว่าจุดกำเนิดอื่น จากข้อมูลการใช้งานพบว่าการแบ่งช่วงคลื่นที่ให้บริการในพื้นที่เดียวกันอยู่หลายช่วงคลื่นคือ คลื่น 900 MHz สำหรับ 2G, คลื่น 1800 MHz สำหรับ 3G, คลื่น 2100 MHz สำหรับ 3G และคลื่น 2100 MHz สำหรับ 4G ซึ่งตำแหน่งของเสากกระจายสัญญาณของแต่ละคลื่นไม่จำเป็นต้องอยู่ที่ตำแหน่งเดียวกัน การหาพื้นที่ให้บริการจำเป็นต้องหาแผนภาพโวโรนอยของแต่ละคลื่นความถี่ตามภาพที่ 13 แสดงกระบวนการในการหาพื้นที่ให้บริการของเสากกระจายสัญญาณโดย (a) แสดงตำแหน่งของเสากกระจายสัญญาณ (b) แสดงตำแหน่งของเสากกระจายสัญญาณแยกตามคลื่นความถี่, (c) แสดงรูปหลายเหลี่ยมโวโรนอยเสากกระจายสัญญาณแยกตามคลื่นความถี่ และ(d) แสดงรูปหลายเหลี่ยมโวโรนอยเสากกระจายสัญญาณทั้งหมดในพื้นที่ ซึ่งผลที่ได้สามารถสรุปกระบวนการได้ตามภาพที่ 14



ภาพที่ 13 ผลจากกระบวนการหาพื้นที่ให้บริการของเสากระจายสัญญาณ

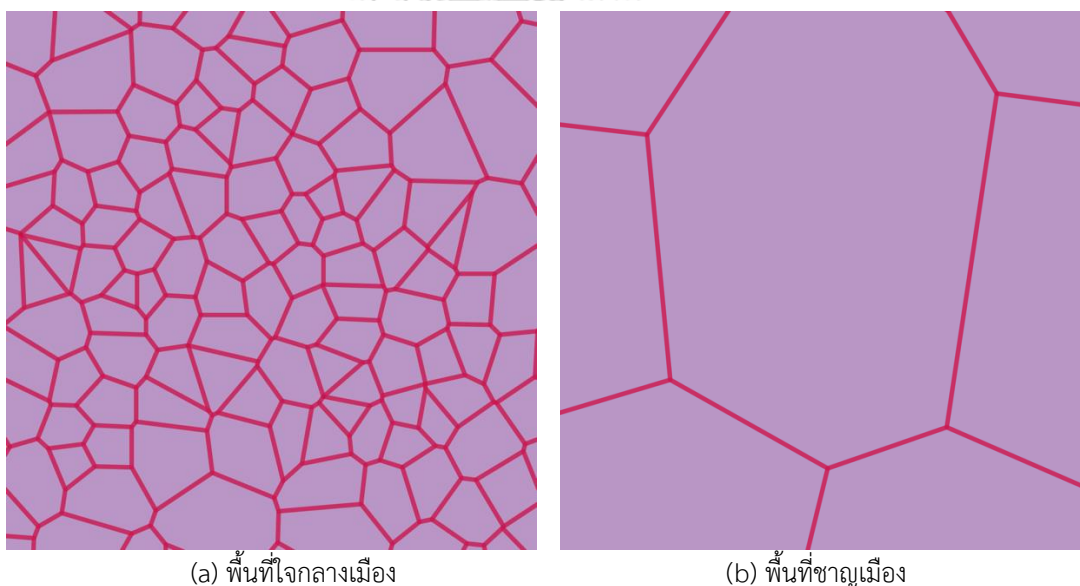


ภาพที่ 14 กระบวนการในการหาพื้นที่ให้บริการของเสากระจายสัญญาณ

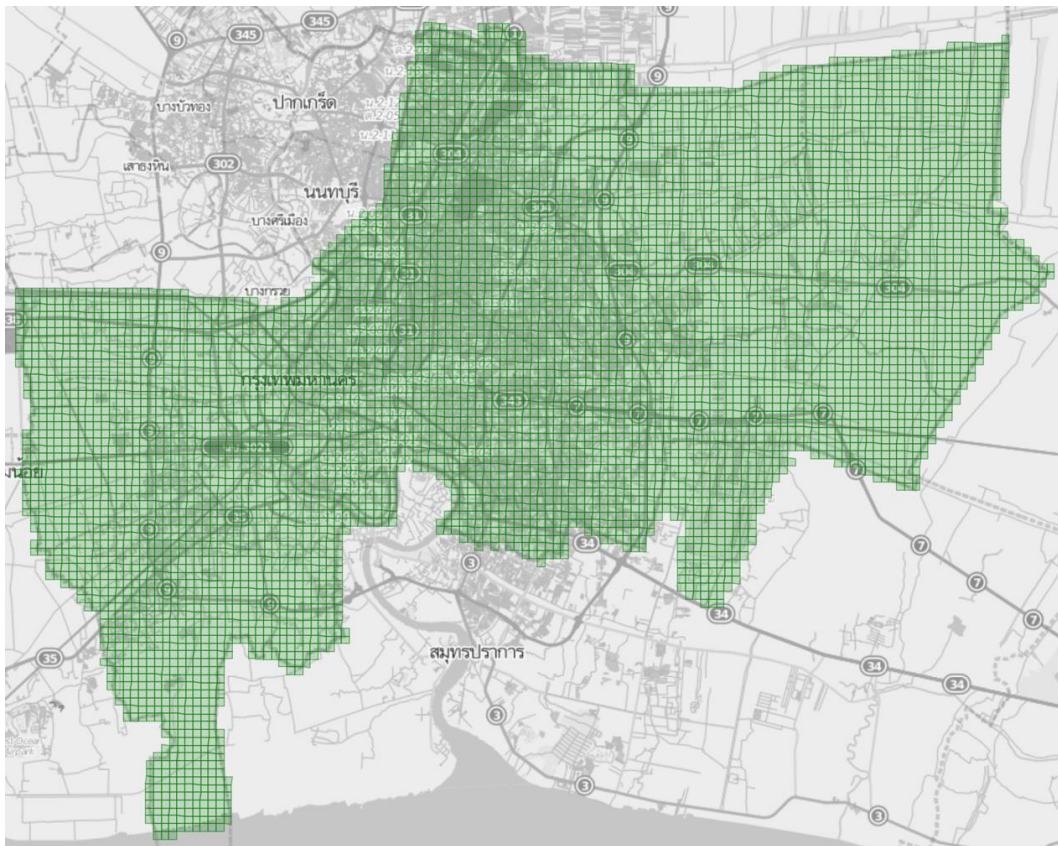


### 3.4. การแบ่งพื้นที่ที่ใช้ศึกษา

รูปหลายเหลี่ยมโวลโรนอยที่แสดงถึงพื้นที่ให้บริการของแต่ละเสากระจายสัญญาณมีขนาดไม่เท่ากันขึ้นอยู่กับความหนาแน่นของพื้นที่ให้บริการ ทำให้ขนาดของรูปหลายเหลี่ยมโวลโรนอยแตกต่างกันมากระหว่างเขตเมืองกับชานเมืองโดยภาพที่ 15(a) เป็นพื้นที่ใจกลางเมือง รูปที่ 15(b) เป็นพื้นที่ชานเมืองมีขนาด  $4 \times 4$  ตารางกิโลเมตรเท่ากัน พบว่าพื้นที่ใจกลางเมืองรูปหลายเหลี่ยมโวลโรนอยมีขนาดเล็กมาก แต่รูปหลายเหลี่ยมโวลโรนอยของพื้นที่ชานเมืองมีขนาดใหญ่มาก เนื่องจากมีการหาโวลโรนอยแยกตามคลื่นที่ปล่อยทำให้ถ้าจะศึกษาในระดับของโวลโรนอยจำเป็นต้องหาพื้นที่ทับซ้อนของแต่ละโวลโรนอย ซึ่งรูปร่างหลายเหลี่ยมมีขนาดเล็กมากไปจนถึงใหญ่มากทำให้นำไปใช้งานต่อได้ยาก ดังนั้นผู้วิจัยจึงทำการแปลงแผนภาพโวลโรนอยให้อยู่ในรูปแบบของตารางกริดที่นำไปใช้งานได้ง่ายกว่า โดยทำการแบ่งพื้นที่ศึกษาออกเป็นกริดขนาด  $500 \text{ m} \times 500 \text{ m}$  ซึ่งเป็นค่าที่มากกว่าค่าเฉลี่ยระยะห่างระหว่างเสากระจายสัญญาณที่ 328.05 เมตร และเสากระจายสัญญาณ 80 % มีระยะห่างระหว่างเสากระจายสัญญาณน้อยกว่า 500 เมตร โดยพื้นที่ที่ศึกษาประกอบด้วยจังหวัดกรุงเทพมหานครและสนามบินสุวรรณภูมิ (ตำบลราชาเทวะและหนองปรือ อำเภอบางพลี จังหวัดสมุทรปราการ) โดยแบ่งพื้นที่เป็น 6,811 กริด ตามภาพที่ 16



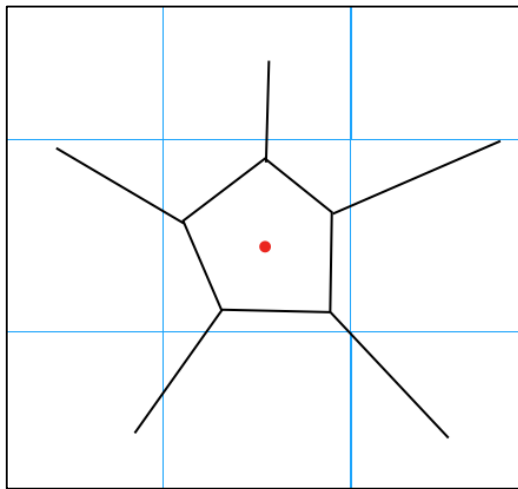
ภาพที่ 15 ความแตกต่างของขนาดพื้นที่ให้บริการระหว่างใจกลางเมืองกับชานเมือง



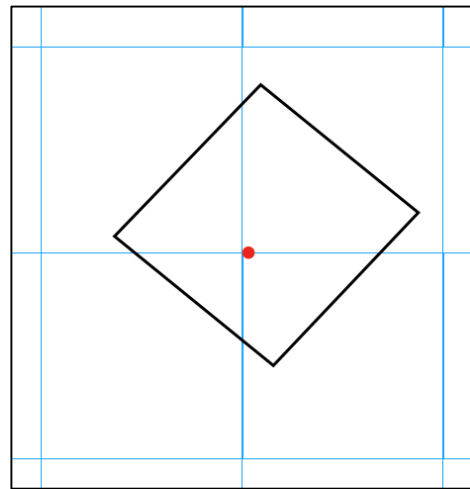
ภาพที่ 16 ตารางตารางขนาด 500 X 500 ตารางเมตร ของพื้นที่จังหวัดกรุงเทพมหานครและ  
 สนามป็นสุวรรณภูมิ(ตำบลราชาเทวะและหนองปรือ อำเภอบางพลี จังหวัดสมุทรปราการ)

### 3.5. การแปลงพื้นที่ให้บริการของเสากระจายสัญญาณให้อยู่ในกริด

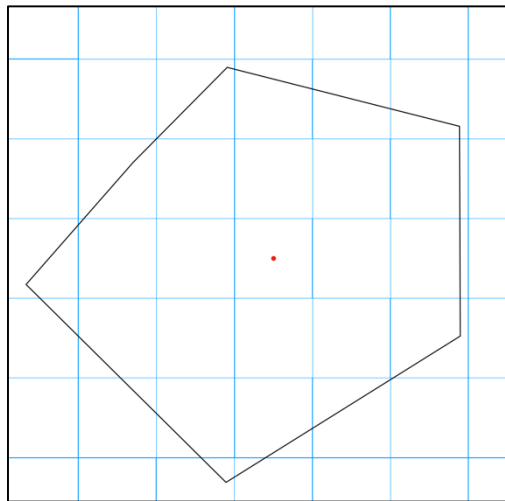
การแปลงพื้นที่ให้บริการของเสากระจายสัญญาณให้อยู่ในกริดเป็นการหาความน่าจะเป็นที่  
 โวลูนอยอยู่ในกริดต่าง เนื่องจากขนาดของรูปหลายเหลี่ยมโวลูนอยแตกต่างกันมากระหว่างเขต  
 เมืองกับชานเมืองทำให้การแปลงต้องพิจารณาความแตกต่างระหว่างพื้นที่ทั้งสอง โดยรูปแบบ  
 ความสัมพันธ์ระหว่างโวลูนอยกับกริดสามารถแบ่งออกได้เป็น 3 แบบตามภาพที่ 17 คือโวลูนอยมี  
 ขนาดเล็กอยู่ในกริดใดกริดหนึ่ง, แบบที่โวลูนอยมีขนาดเล็กอยู่ระหว่างกริด โดย 2 แบบแรกจะ  
 พบได้ในเขตที่มีการใช้งานหนาแน่นสูง และแบบที่โวลูนอยมีขนาดใหญ่ครอบคลุมเต็มกริดพบได้ใน  
 เขตที่มีการใช้งานหนาแน่นต่ำ



(a) โวโรนอยมีขนาดเล็กอยู่ภายในกริดใดกริดหนึ่ง



(b) โวโรนอยมีขนาดเล็กอยู่ระหว่างกริด



(c) โวโรนอยมีขนาดใหญ่ครอบคลุมเต็มกริด

### ภาพที่ 17 รูปแบบความสัมพันธ์ระหว่างโวโรนอยกับกริด

จากรูปแบบความสัมพันธ์การแปลงพื้นที่ให้บริการของเสากระจายสัญญาณให้อยู่ในกริดสามารถ 2 วิธีคือใช้อัตราส่วนระหว่างพื้นที่ของโวโรนอยภายในกริดกับพื้นที่โวโรนอยทั้งหมดวิธีนี้เหมาะสมกับโวโรนอยรูปแบบที่ 1 และ 2 แต่จะทำให้รูปแบบที่โวโรนอยมีขนาดใหญ่ครอบคลุมเต็มกริดมีค่าน้อยมากเมื่อเทียบกับโวโรนอยมีขนาดเล็ก ซึ่งจะมีผลต่อค่าน้ำหนักที่ใช้หาสถานที่สำคัญทำให้ไม่สามารถหาบ้านในเขตนอกเมืองได้ อีกวิธีการหนึ่งคือใช้อัตราส่วนระหว่างพื้นที่ของโวโรนอยภายในกริดกับพื้นที่กริด วิธีการนี้เหมาะสมกับโวโรนอยรูปแบบที่ 3 แต่จะทำให้รูปแบบที่โวโรนอยมีขนาดเล็กมีค่าน้อยกว่า 1 ถึงแม้ว่าโวโรนอยนั้นจะอยู่ในกริดทั้งหมด ดังนั้นผู้วิจัยจึงสร้างสมการ (1) โดยจะใช้พื้นที่โวโรนอยทั้งหมดหรือพื้นที่กริดขึ้นอยู่กับขนาดของโวโรนอย

$$pvg(v_i, g_j) = \frac{\text{overlap\_area}(v_i, g_j)}{\min(\text{area}(v_i), \text{area}(g_j))} \quad (1)$$

โดย  $pvg(v_i, g_j)$  คือความน่าจะเป็นที่โวลูนอยที่  $i$  จะอยู่ในกริด  $j$ ,  $\text{overlap\_area}(v_i, g_j)$  คือพื้นที่ของโวลูนอยที่  $i$  อยู่ในกริด  $j$ ,  $\text{area}(v_i)$  คือพื้นที่ทั้งหมดของโวลูนอยที่  $i$ ,  $\text{area}(g_j)$  คือพื้นที่กริด  $j$

### 3.6. การหาสถานที่ทำงานของผู้ใช้

การหาตำแหน่งที่ทำงานในงานนี้เป็นการหาที่ทำงานของคนทั่วไปที่ทำงานในสถานที่แน่นอนและมีเวลาเข้าออกที่ทำงานชัดเจน โดยหาจากกริดที่ผู้ใช้งานไปอยู่ในเป็นจำนวนวันมากที่สุด ในช่วงเวลาที่คนทั่วไป คือ 10:00 – 15:00 โดยความแม่นยำจะขึ้นอยู่กับจำนวนวันที่พบ ดังนั้น เพื่อให้ได้ความแม่นยำที่มากที่สุดจึงใช้ข้อมูลทั้งหมดที่สามารถหาตำแหน่งของผู้ใช้งานได้ คือข้อมูลซีดีอาร์และจี-ซีดีอาร์ แต่ข้อมูลจี-ซีดีอาร์เป็นข้อมูลที่ถูกรวมเป็นการใช้งานรายวันทำให้ไม่สามารถเลือกเฉพาะช่วงเวลาทำงานได้ แต่จากภาพที่ 7 การใช้งานที่เกิดในช่วงเวลาทำงานมีมากกว่าช่วงเวลาอื่นมากทำให้ข้อมูลที่สรุปในรูปแบบรายวันบริเวณสถานที่ทำงานมีการใช้งานที่สูงตามไปด้วย ดังนั้นจะทำการรวมซีดีอาร์กับจี-ซีดีอาร์กลายเป็นการใช้งานในแต่ละวันและทำการหากริดที่มีค่าความน่าจะเป็นที่กริดจะเป็นสถานที่สำคัญ  $\text{days}()$  มากที่สุด 20 อันดับได้เป็นกริดสำคัญของแต่ละคน

$$\text{days}(g_j) = \sum_{v_i \in V_{g_j}} (\text{distinct\_days}(v_i) \times pvg(v_i, g_j)) \quad (2)$$

โดย  $\text{days}(g_j)$  เป็นความน่าจะเป็นที่กริดที่  $j$  เป็นกริดที่มีสถานที่สำคัญ,  $V_{g_j}$  เป็นโวลูนอยทั้งหมดที่มีส่วนใดส่วนหนึ่งอยู่ในกริด  $j$ ,  $\text{distinct\_days}(v_i)$  เป็นจำนวนวันทั้งหมดที่อยู่ในโวลูนอยที่  $i$

กริดสำคัญของแต่ละคนบอกได้แค่กริดที่ผู้ใช้งานอยู่มากเป็นข้อมูลไม่มีเวลากำกับทำให้บางกริดไม่ใช่สถานที่ทำงาน ผู้วิจัยใช้ข้อมูลจากซีดีอาร์เพื่อทำการกรองกริดสำคัญของแต่ละคนให้เหลือแต่สถานที่ทำงานจากการพิจารณาสถานที่ที่เดินทางในช่วงเวลาทำงานและมีการใช้งานในสถานที่นั้นสูงตลอดช่วงเวลางานเพื่อกรองสถานที่ไปอยู่ในเวลางานในช่วงเวลาสั้น ๆ และทำการนับจำนวนวันที่ใช้งานในแต่ละกริดแต่ละชั่วโมงเพื่อดูความน่าจะเป็นที่จะพบผู้ใช้งานในแต่ละกริดในชั่วโมงต่าง ๆ จากนั้นทำการเลือกกริดที่มีจำนวนวันมากที่สุด 4 อันดับในแต่ละชั่วโมงของผู้ใช้แต่ละคนเป็นกริดสำคัญในแต่ละชั่วโมงของแต่ละคน (hours) และทำการคำนวณค่าที่มาจากจำนวนวัน

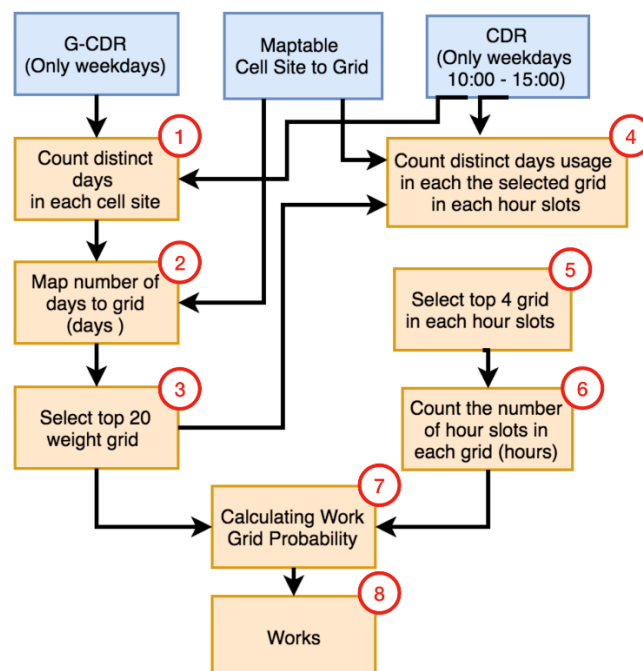
ของกริดสำคัญของแต่ละคนและจำนวนชั่วโมงที่กริดนั้นเป็นกริดสำคัญในแต่ละชั่วโมงของแต่ละคน ตามสมการที่ (3)

$$work(g_i) = days(g_i) \times hours(g_i) \quad (3)$$

โดย  $work(g_i)$  เป็นผลคูณจำนวนวันและจำนวนชั่วโมง,  $hours(g_i)$  คือจำนวนชั่วโมงที่กริดที่  $i$  เป็นกริดที่ติด 4 กริดที่มีการใช้งานมากที่สุดในแต่ละชั่วโมง โดยกริดที่มีค่า  $work(g_i)$  มากที่สุดจะถูกเลือกเป็นสถานที่ทำงานของแต่ละคน

โดยกระบวนการหาที่ทำงานของผู้ใช้งานทั้งหมดสรุปได้ตามภาพที่ 18 คือ

- 1) นับจำนวนวันที่มีการใช้งานในแต่ละเสากระจายสัญญาณจากจี-ซีดีอาร์และซีดีอาร์
- 2) หาความน่าจะเป็นที่แต่ละกริดจะเป็นกริดสำคัญของแต่ละคนโดยใช้สมการ  $days$
- 3) เลือก 20 กริดของแต่ละคนที่มีความน่าจะเป็นกริดสำคัญมากที่สุด
- 4) คำนวณจำนวนวันที่มีการใช้งานในชั่วโมงต่าง ๆ ในกริดที่เป็นกริดสำคัญจาก 3) ของแต่ละคน
- 5) เลือกกริด 4 อันดับที่มีค่ามากที่สุดจากการคำนวณใน 4) ของแต่ละชั่วโมงแต่ละคน
- 6) นับจำนวนชั่วโมงที่กริดนั้นเป็นกริดที่ติด 4 กริดที่มีการใช้งานมากที่สุดในแต่ละชั่วโมง
- 7) คำนวณความน่าจะเป็นที่กริดต่าง ๆ จะเป็นที่ทำงานโดยใช้สมการ  $work$
- 8) เลือกกริดที่มีความน่าจะเป็นมากที่สุดที่จะเป็นที่ทำงาน



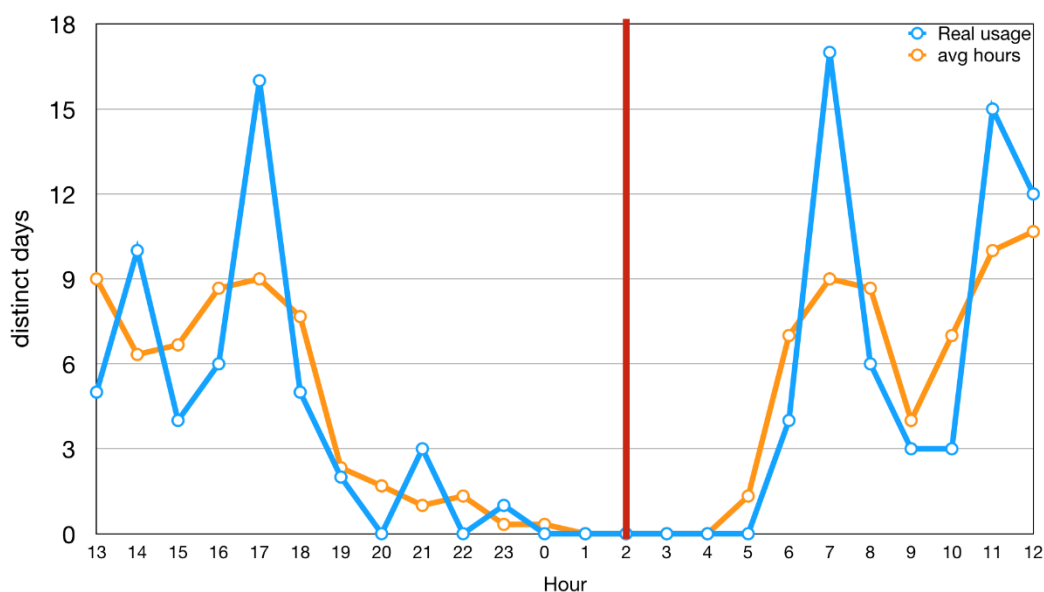
ภาพที่ 18 กระบวนการในการหาที่ทำงานของผู้ใช้งาน



### 3.7. การหาบ้านของผู้ใช้

การหาบ้านของผู้ใช้งานมีความท้าทายเพราะมีการใช้งานน้อยในช่วงเวลากลางคืนทำให้ไม่สามารถใช้ข้อมูลจากจี-ซีดีอาร์เข้ามาช่วยในการหาสถานที่สำคัญได้เนื่องจากจะถูกข้อมูลที่เกิดจากช่วงเวลากลางวันกลบหมด หรือได้สถานที่ที่ผู้ใช้ไปเที่ยวในช่วงกลางคืนแทน ผู้วิจัยจึงใช้วิธีในการหาช่วงเวลาที่ผู้ใช้แต่ละคนนอนและนำข้อมูลการใช้งานครั้งสุดท้ายก่อนเข้านอนและการใช้งานครั้งแรกสุดหลังตื่นนอน ซึ่งมีความน่าจะเป็นที่ผู้ใช้งานจะใช้จากที่บ้านสูงกว่าการใช้งานในช่วงเวลาอื่น

การหาเวลานอนหาจากช่วงเวลาที่มีการใช้งานน้อยสุดโดยจะดูการกระจายของการใช้งานตลอด 24 ชั่วโมง โดยขั้นแรกจะทำการนับจำนวนวันที่มีการใช้งานจากซีดีอาร์ในชั่วโมงต่าง ๆ และทำการปรับเรียบเส้นโค้งโดยการเฉลี่ยจำนวนวันที่มีการใช้งานในชั่วโมงนั้นกับชั่วโมงที่ติดกัน  $\pm 1$  ชั่วโมง และเลือกชั่วโมงที่มีการใช้งานน้อยที่สุดเป็นชั่วโมงที่ผู้ใช้คนนั้นนอน แต่ถ้ามีหลายเวลาที่มีการใช้งานน้อยที่สุด จะเลือกชั่วโมงในตำแหน่งควอไทล์ ที่ 20 โดยเริ่มจาก 21.00 น. เป็นต้นไป เนื่องจากเวลา 21.00 น. เป็นเวลาที่ผู้คนส่วนใหญ่เริ่มเข้านอน การใช้ตำแหน่งควอไทล์ที่ 20 เพื่อให้มั่นใจว่าเป็นเวลานอนประจำ และอยู่ในช่วงรูปร่างยูแรกในกรณีมีช่วงรูปร่างยู จากกราฟการกระจายตัวของจำนวนวันที่โทรในชั่วโมงต่าง ๆ ของตัวอย่างผู้ใช้งานหนึ่งคน (ภาพที่ 19) ชั่วโมงที่มันในว่าผู้ใช้คนนั้นนอนคือในช่วง 02:00-02:59 น. เนื่องจากค่าเฉลี่ยจำนวนวันที่มีการใช้งานในชั่วโมงนั้นกับชั่วโมงที่ติดกันที่น้อยที่สุดคือ 0 แต่ชั่วโมงที่มีจำนวนวันที่มีการใช้งานในชั่วโมงนั้นกับชั่วโมงที่ติดกันเท่ากับ 0 ทั้งหมด 4 ชั่วโมง โดยตำแหน่งควอไทล์ ที่ 20 ที่เริ่มจาก 3 ทุ่มเป็นต้นไปเป็นชั่วโมงที่ 2



ภาพที่ 19 กราฟการกระจายตัวของจำนวนวันที่โทรในชั่วโมงต่าง ๆ ของตัวอย่างผู้ใช้งานหนึ่งคน

เมื่อได้ชั่วโมงที่มีความมั่นใจสูงที่ผู้ใช้งานคนนั้นนอนจะทำการกรองข้อมูลซีดีอาร์ให้เหลือเฉพาะข้อมูลการใช้งานครั้งสุดท้ายก่อนเข้านอนและการใช้งานครั้งแรกสุดตอนตื่นนอน และทำการกรองข้อมูลอีกรอบให้อยู่ในช่วงเวลา  $\pm 4$  ชั่วโมงจากเวลานอนของแต่ละคนซึ่งเป็นช่วงเวลาที่คาดว่าจะอยู่ที่บ้านเพื่อเพิ่มความแม่นยำว่าการโทรครั้งนั้นน่าจะเป็นการโทรจากที่บ้าน

$$LCB = \{x \mid x \in CDR \cap \text{time}(x) < \text{sleeptime} \cap \min(\text{time}(x) - \text{sleeptime})\}$$

$$FCA = \{x \mid x \in CDR \cap \text{time}(x) > \text{sleeptime} \cap \min(\text{sleeptime} - \text{time}(x))\}$$

โดย  $LCB$  เป็นการโทรครั้งสุดท้ายก่อนเข้านอนของผู้ใช้,  $FCA$  เป็นการโทรครั้งแรกหลังตื่นนอนของผู้ใช้,  $\text{time}(x)$  เป็นเวลาที่มีการใช้งาน,  $\text{sleeptime}$  เป็นชั่วโมงที่มั่นใจว่าผู้ใช้คนนั้นนอน

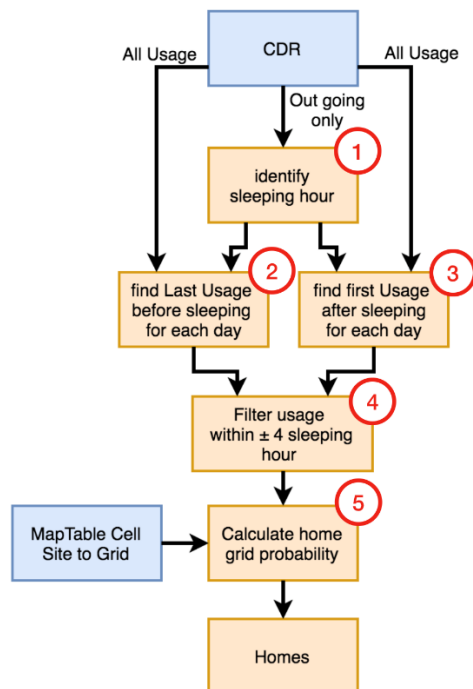
เมื่อได้ข้อมูลซีดีอาร์ที่กรองเหลือเฉพาะข้อมูลที่คาดว่าจะอยู่ที่บ้านแล้วจะทำการคำนวณค่าของแต่ละกริด โดยคำนวณจากความถี่ที่มีการใช้งานในแต่ละกริดซึ่งค่าความถี่สูงสุดที่เป็นไปได้คือ 2 เท่าของจำนวนวันทั้งหมดเนื่องจากกรองเหลือเฉพาะข้อมูลการโทรครั้งแรกและครั้งสุดท้ายของแต่ละวันแล้ว โดยกริดที่เป็นที่เป็นตำแหน่งบ้านต้องมีค่ามากกว่า 10 % ของวันทั้งหมด

$$\text{home}(m_k) = \max_{g_i \in G} \left( \left( \text{count}(LCB(g_i, m_k)) + \text{count}(FCA(g_i, m_k)) \right) \times \text{pvg}(v_i, g_j) \right)$$

โดย  $\text{home}(m_k)$  เป็นกริดที่มีความน่าจะเป็นที่บ้านของผู้ใช้  $m_k$ ,  $LCB(g_i, m_k)$  โทรครั้งสุดท้ายก่อนเข้านอนของผู้ใช้  $m_k$  ที่มีไวโรนอยบางส่วนอยู่ในกริดที่  $i$ ,  $FCA(g_i, m_k)$  เป็นการโทรครั้งแรกหลังตื่นนอนของผู้ใช้  $m_k$  ที่มีไวโรนอยบางส่วนอยู่ในกริดที่  $i$

กระบวนการการหาบ้านของผู้ใช้งานจากภาพที่ 20 มีทั้งหมด 5 ขั้นตอนหลักคือ

- 1) หาชั่วโมงที่ผู้ใช้แต่ละคนนอน
- 2) หากการใช้งานครั้งสุดท้ายก่อนชั่วโมงที่ผู้ใช้นอน ตามสมการ  $LCB$
- 3) หากการใช้งานครั้งแรกหลังชั่วโมงที่ผู้ใช้นอน ตามสมการ  $FCA$
- 4) กรองข้อมูลจาก 2) และ 3) ให้เหลือเฉพาะข้อมูลในช่วง  $\pm 4$  จากชั่วโมงที่ผู้ใช้แต่ละคนนอน
- 5) คำนวณความน่าจะเป็นที่กริดต่าง ๆ จะเป็นบ้านโดยใช้สมการ  $\text{home}$



ภาพที่ 20 กระบวนการในการหาบ้านของของผู้ใช้งาน

### 3.8. การกรองข้อมูลจีพีเอส

ข้อมูลจีพีเอสถูกส่งมาจากรถแท็กซี่ พบว่าข้อมูลของรถบางคันมีความผิดปกติเช่นไม่มีการเปลี่ยนแปลงของสถานะเครื่องยนต์หรือสถานะของไฟแสดงรถว่างหน้ารถ ซึ่งเกิดจากสายไฟที่ใช้เก็บข้อมูลนั้นหลุดออก ดังนั้นในขั้นตอนนี้จะทำการกรองข้อมูลของรถคันที่ไม่มีการเปลี่ยนค่าสถานะเครื่องยนต์, สถานะว่างหน้ารถ, ความเร็ว, ละติจูด และลองจิจูด

### 3.9. การสกัดข้อมูลการเดินทางที่ใช้ในการส่งผู้โดยสาร

ข้อมูลที่ผ่านการกรองแล้วจะเป็นข้อมูลต่อเนื่องกันที่เปลี่ยนแปลงตามเวลาของรถแท็กซี่แต่ละคัน ในการสกัดการเดินทางที่ใช้ในการส่งลูกค้าจะอาศัยการเปลี่ยนแปลงไฟแสดงรถว่างที่ติดอยู่หน้ารถ โดยเมื่อผู้โดยสารขึ้นรถแท็กซี่คนขึ้นรถแท็กซี่จะทำการกดมิเตอร์ไฟแสดงรถว่างที่ติดอยู่หน้ารถจะดับลง (passenger\_flag = 0) เมื่อถึงที่หมายคนขึ้นรถแท็กซี่จะทำการกดมิเตอร์ไฟแสดงรถว่างให้ติดขึ้น (passenger\_flag = 1)

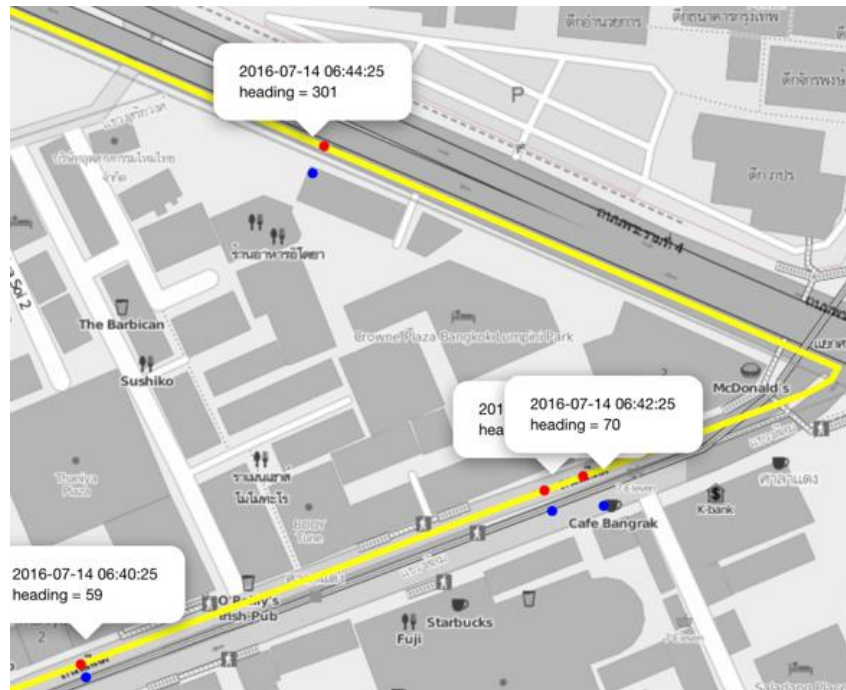


ข้อมูลการเดินทางที่ได้จากการสกัดการเดินทางที่ใช้ส่งลูกค้าพบว่าความผิดพลาดอยู่ ดังนั้นจำเป็นที่จะต้องทำการกรองข้อมูลการเดินทางที่ใช้ส่งผู้โดยสารแต่ละคนโดยมีเงื่อนไขดังนี้

- ตลอดการเดินทางสถานะเครื่องยนต์ต้องเป็น 1 ตลอดเวลา
- ข้อมูลการเดินทางจะต้องไม่มีข้อมูลที่สูญหาย โดยพิจารณาจากข้อมูลที่ติดกัน จะต้องใช้เวลาห่างกันไม่เกิน 3 นาที
- ข้อมูลการเดินทางนั้นต้องมีระยะห่างระหว่างจุดเริ่มต้นและจุดสิ้นสุดมากกว่า 1 กิโลเมตร
- ข้อมูลการเดินทางจะต้องมีข้อมูลมากกว่า 3 บันทึก
- ข้อมูลการเดินทางที่ได้จะต้องไม่มีข้อมูลพิกัดที่ผิดพลาด โดยพิจารณาจากความเร็วที่คำนวณจากระยะทางและเวลาของข้อมูลเวลาที่ติดกันจะต้องไม่มากกว่า 150 กิโลเมตรต่อชั่วโมง
- เวลารวมตลอดทั้งการเดินทางต้องไม่มากกว่า 6 ชั่วโมงเพื่อทำการกรองข้อมูลการเดินทางที่เดินทางไปต่างจังหวัดหรือเกิดข้อผิดพลาดของสัญญาณแสดงสถานะผู้โดยสาร

### 3.10. การแปลงข้อมูลตำแหน่งจีพีเอสเป็นเส้นทางบนถนน

ข้อมูลพิกัดที่ได้จากรถแท็กซี่เป็นพิกัดที่ได้จากการเครื่องรับจีพีเอสเพียงพิกัดเดียว ณ เวลาที่ทำการส่งข้อมูลไปที่ศูนย์ทำให้และข้อมูลที่ได้มีข้อผิดพลาดขึ้นอยู่กับสภาพแวดล้อม ภาพที่ 21 แสดงตัวอย่างข้อมูลดิบที่ได้รับจากรถแท็กซี่เป็นจุดสีน้ำเงินซึ่งมีความคลาดเคลื่อนจากพิกัดถนน โดยพิกัดจีพีเอสที่ถูกต้องใกล้พิกัดที่ได้รับจากรถแท็กซี่มากที่สุดคือจุดสีแดง แต่จุดสีแดงไม่เพียงพอต่อการจัดกลุ่มเส้นทางเนื่องจากข้อมูลมีค่าการส่งที่ 1 นาที ทำให้ในเส้นทางที่แท็กซี่สามารถทำความเร็วข้อมูลที่ได้รับอาจเกิดการกระโดดในสะพานหรือถนนสายอื่นๆ ได้ ดังนั้นจำเป็นที่จะต้องหาลำดับของถนนที่เดินทางระหว่างจุดสีแดงซึ่งแสดงเป็นเส้นสีเหลือง



ภาพที่ 21 ตัวอย่างข้อมูลดิบที่ได้รับจากรถแท็กซี่ที่มีความคลาดเคลื่อนจากพิกัดถนน

การแปลงข้อมูลพิกัดจีพีเอสให้เป็นเส้นทางบนถนนที่ใช้ในการส่งผู้โดยสารจากข้อมูลแต่ละจุดจำเป็นต้องแก้ปัญหาต่อไปนี้

- พิกัดจีพีเอสที่ผิดพลาดไปอยู่บนถนนเส้นอื่น ตัวอย่างภาพที่ 22 (a) เป็นการเดินทางจากแยกปทุมวันไปแยกพญาไทโดยใช้เส้นทางถนนพญาไท พบว่ามีพิกัดจีพีเอสกลุ่มหนึ่งอยู่บนถนนที่เข้าไปยังซอยเพชรบุรี 16 แต่เมื่อพิจารณาจากทิศทางของหัวรถแท็กซี่พบว่าขนานกับถนนพญาไท และพิกัดจีพีเอสต่อไปจากกลุ่มนั้นพบว่าอยู่บนถนนพญาไท ดังนั้นกลุ่มของจุดนั้นมีความน่าจะเป็นสูงที่เกิดจากความคลาดเคลื่อนของพิกัดจีพีเอสขณะติดไฟแดงที่แยกราชเทวี
- พิกัดจีพีเอสที่แท้จริงจะอยู่บนถนนเส้นใดจากถนนที่ขนานมากกว่าหนึ่งเส้น ตัวอย่างภาพที่ 22 (b) เป็นการเดินทางจากสีลมไปถนนพญาไทผ่านถนนพระราม 4 แต่ที่ถนนพระราม 4 มีสะพานไทยญี่ปุ่นขนานตามแนวถนนพระราม 4 ดังนั้นจำเป็นต้องเลือกว่าพิกัดจีพีเอสนั้นจะอยู่บนถนนเส้นใด แต่เมื่อพิจารณาจากพิกัดจีพีเอสก่อนหน้าอยู่บริเวณถนนสีลม ซึ่งการเดินทางจากพิกัดตรงนั้นสามารถเดินทางไปยังถนนพระราม 4 ได้เลย แต่ถ้าเดินทางไปสะพานไทยญี่ปุ่นจำเป็นต้องอ้อมไปขึ้นสะพานไทยญี่ปุ่น



(a) ตัวอย่างพิกัดจีพีเอสที่ผิดพลาดไปอยู่บนถนนเส้นอื่น



(b) ถนนที่ขนานกันในบริเวณพิกัดจีพีเอส

ภาพที่ 22 ปัญหาที่เกิดจากข้อผิดพลาดของตำแหน่งจีพีเอสที่ต้องแก้ไข

จากปัญหาข้างต้นทำให้ไม่สามารถทำการแปลงพิกัดจีพีเอสแต่ละจุดไปยังถนนได้โดยตรง ต้องพิจารณาโดยอาศัยทิศทางของหัวรถแท็กซี่, ทิศทางของช่วงถนน, พิกัดจีพีเอสก่อนหน้า, พิกัดจีพีเอสถัดไปอีกจุดหนึ่ง และการเชื่อมต่อกันของถนน ดังนั้นผู้วิจัยได้แบ่งขั้นตอนเป็น 2 ส่วนหลัก ๆ ได้แก่ขั้นตอนการหากลุ่มถนนที่มีความน่าจะเป็นสูงที่รถแท็กซี่อยู่จากพิกัดจีพีเอส และขั้นตอนการหาเส้นทางบนถนนทั้งหมดระหว่างช่วงถนนและเลือกถนนที่มีความน่าจะเป็นที่รถแท็กซี่อยู่

### 3.10.1. การหากรุ่นถนนที่มีความน่าจะเป็นสูงที่รถแท็กซี่ที่อยู่จากพิกัดจีพีเอส

ขั้นตอนการหากรุ่นถนนที่มีความน่าจะเป็นสูงที่รถแท็กซี่ที่อยู่จากพิกัดจีพีเอส เป็นการกรองช่วงถนนให้เหลือเฉพาะถนนที่มีความน่าจะเป็นสูง โดยอาศัยพิกัดจีพีเอสของจุดที่จะทำการแปลงเป็นถนน, ทิศทางของหัวรถแท็กซี่ ณ เวลานั้น และทิศทางของช่วงถนน โดย

- ขั้นแรกจะหาถนนที่ห่างจากพิกัดจีพีเอสไม่เกิน 150 เมตร และคำนวณ

$$DistFact = Haversine\_dist(taxiPoint, road)/150$$

- พิจารณาทิศทางหัวรถแท็กซี่กับทิศทางของถนน โดยใช้สมการ HeadFact ซึ่งค่าที่ได้ต้องมากกว่า 0.25881905 ซึ่งเป็นการกรองให้เหลือเฉพาะถนนที่ทำมุมไม่เกิน  $75^{\circ}$  จากทิศทางของหัวรถ

$$HeadFact = \begin{cases} \cos(HeadingofTaxi - DirectionofRoad), & \text{Oneway} = \text{TRUE} \\ \text{abs}(\cos(HeadingofTaxi - DirectionofRoad)), & \text{Oneway} = \text{FALSE} \end{cases}$$

โดย HeadingofTaxi คือทิศทางของหัวรถแท็กซี่, DirectionRoad คือทิศทางของถนนช่วงนั้น และ Oneway เป็นตัวบ่งชี้ถนนนั้นแล่นรถทางเดียวหรือไม่

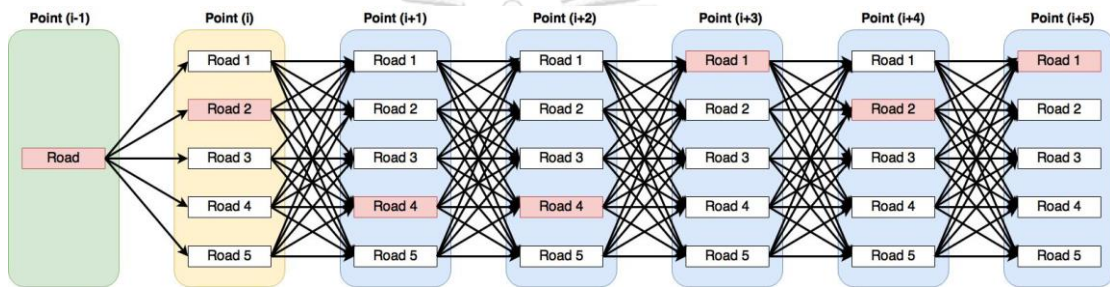
- ทำการเลือกถนนที่มีความน่าจะเป็นที่รถแท็กซี่อยู่ 5 อันดับแรกจากสมการ

$$RoadProb = (1 + DistFact) \times (1 + HeadFact)$$

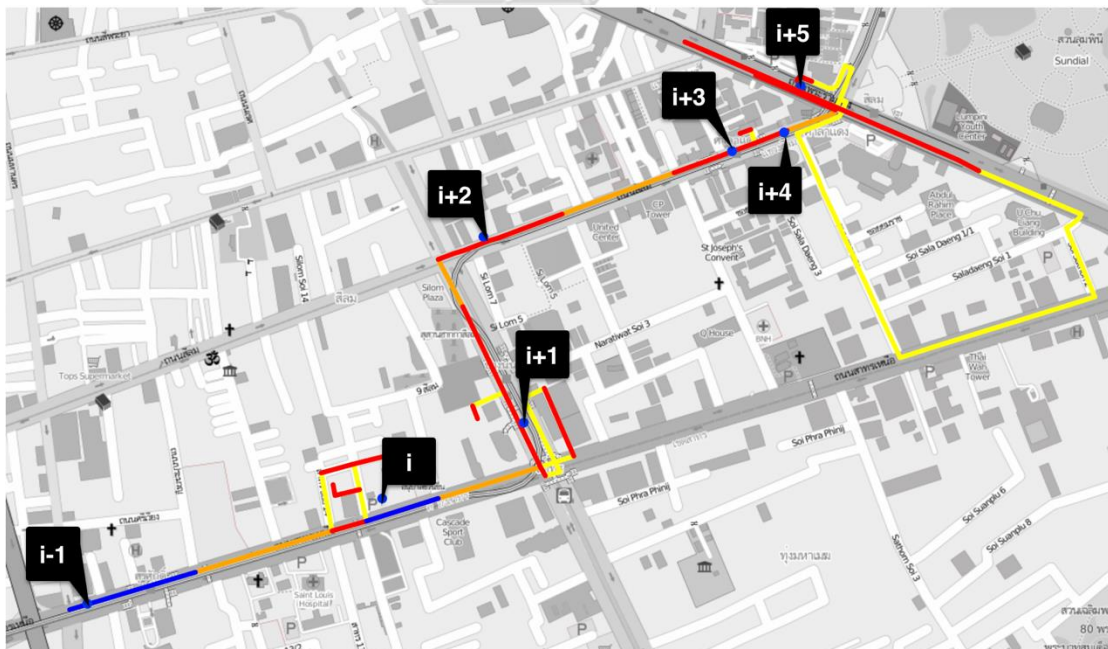
### 3.10.2. การเลือกถนนที่มีความน่าจะเป็นที่รถแท็กซี่อยู่

ขั้นตอนต่อมาเป็นกระบวนการในการเลือกถนนจาก 1 ใน 5 ที่ได้จากการหาถนนที่มีความน่าจะเป็นที่รถแท็กซี่อยู่ 5 อันดับแรกของแต่ละตำแหน่งจีพีเอส โดยพิจารณาที่ละ 6 ตำแหน่งจีพีเอสตามภาพที่ 23 (a) เมื่อต้องการเลือกถนนของตำแหน่งจีพีเอสที่  $i$  จะใช้ถนนที่ทำการเลือกแล้วของพิกัดที่  $(i-1)$  และถนนที่ผ่านการกรองของพิกัดที่  $(i+1)$ ,  $(i+2)$ ,  $(i+3)$ ,  $(i+4)$  และ  $(i+5)$  โดยจะทำการหาเส้นทางที่สั้นที่สุดจากถนนที่ถูกเลือกของพิกัดจีพีเอสที่  $(i-1)$  ผ่านถนนที่ผ่านการกรองของพิกัดจีพีเอสที่  $i$ ,  $(i+1)$ ,  $(i+2)$ ,  $(i+3)$ ,  $(i+4)$  และ  $(i+5)$  ตามลำดับ โดยรูปแบบที่เป็นไปได้มากที่สุดจะมี 15,625 เส้นทางที่ใช้เดินทางจากพิกัดที่  $(i-1)$  ไปยังพิกัดที่  $(i+5)$  และเลือกเส้นทางจากถนน  $(i-1)$  ไปยังถนนในพิกัดที่  $i$  จากเส้นทางที่ใช้ระยะทางตามถนนที่สั้นที่สุดจากพิกัดที่  $(i-1)$  ไปยังพิกัดที่  $(i+5)$  ก่อนที่จะเลื่อนพิกัดเพื่อเลือกถนนของพิกัดจีพีเอสที่  $(i+1)$  สาเหตุที่พิจารณาจนถึงพิกัดที่  $(i+5)$  เนื่องจากต้องการความแม่นยำที่มากขึ้นในกรณีที่มีเส้นถนนที่ขนานกันโดยใช้พิกัด 5 พิกัดดูความน่าจะเป็นที่อยู่บน

ถนนเส้นใดเส้นหนึ่งระหว่างถนนที่ขนานกัน ภาพที่ 23 (b) แสดงตัวอย่างการเลือกถนนที่มีความน่าจะเป็นสูงที่รถจะอยู่จากพิกัดจีพีเอสที่  $i$  จากข้อมูลจริง โดยพิกัดสีน้ำเงินเป็นพิกัดจีพีเอสจริงที่ได้จากรถแท็กซี่ ช่วงถนนสีแดงแสดงถนนที่มีความน่าจะเป็นสูงที่รถแท็กซี่จะอยู่จากกระบวนการหากลุ่มถนนที่มีความน่าจะเป็นสูงที่รถแท็กซี่อยู่จากพิกัดจีพีเอสซึ่งมีได้มากที่สุด 5 ช่วงถนน เส้นสีเหลืองและส้มแสดงเส้นทางทั้งหมดที่สั้นที่สุดจากถนนสีน้ำเงินในพิกัดที่  $(i-1)$  ไปยังถนนที่มีความน่าจะเป็นสูงที่รถแท็กซี่จะอยู่ของพิกัดที่  $(i+5)$  โดยผ่านถนนที่มีความน่าจะเป็นสูงที่รถแท็กซี่จะอยู่ของพิกัดที่  $i, (i+1), (i+2), (i+3)$  และ  $(i+4)$  จากเส้นทางสีส้มที่เป็นเส้นทางที่มีระยะทางรวมที่สั้นที่สุดทำให้ทราบถนนที่มีความน่าจะเป็นสูงที่พิกัดที่  $i$  อยู่คือเส้นสีน้ำเงินใกล้พิกัดที่  $i$  และเส้นทางที่ใช้เดินทางจากพิกัดที่  $i-1$  ไปพิกัดที่  $i$  คือเส้นสีส้มระหว่างเส้นสีน้ำเงินทั้งสอง



(a) แสดงวิธีการเลือกถนนที่มีความน่าจะเป็นสูงที่รถจะอยู่จากพิกัดจีพีเอสที่  $i$



(b) ตัวอย่างการเลือกถนนที่มีความน่าจะเป็นสูงที่รถจะอยู่จากพิกัดจีพีเอสที่  $i$  จากข้อมูลจริง ภาพที่ 23 วิธีการเลือกถนนที่มีความน่าจะเป็นสูงที่รถจะอยู่จากพิกัดจีพีเอส

การหาเส้นทางที่สั้นที่สุดจำเป็นต้องสร้างกราฟโดยอาศัยการเชื่อมต่อกันของถนน โดยให้จุดที่เชื่อมต่อระหว่างถนนเป็นจุดยอด ถนนเป็นเส้นเชื่อม โดยแต่ละเส้นเชื่อมอาจเป็นทั้งที่มีทิศทางเดียวหรือสองทิศทางขึ้นอยู่กับถนนนั้นเป็นถนนที่แล่นรถทางเดียวหรือไม่ เนื่องจากกราฟนี้จะถูกนำไปใช้เพื่อหาเส้นทางที่สั้นที่สุด ซึ่งถ้าใช้ความยาวของถนนเป็นค่าน้ำหนักจะทำให้ผลที่ได้จากการหาเส้นทางที่สั้นที่สุดอาจหลุดออกจากเส้นทางหลักและเข้าซอยที่มีขนาดเล็กเพื่อลัดเส้นทางได้ ซึ่งโดยปกติจะใช้ระดับความสำคัญของถนนเป็นตัวช่วยในการเลือกเส้นทาง แต่ข้อมูลถนนที่ได้รับมาไม่มีข้อมูลระดับความสำคัญของถนน และจากงานวิจัย [31] ได้ศึกษาพฤติกรรมการเลือกใช้นถนนพบว่าในเส้นทางที่มีระยะทางใกล้เคียงกัน ผู้ใช้มักจะเลือกใช้นถนนที่มีจำนวนเลนมากกว่า ดังนั้นค่าน้ำหนักของถนนคิดตามสมการที่ roadW

$$roadW = \begin{cases} Length * (1.7 - (0.35 * 3)), Oneway = TRUE, Lane \geq 3 \\ Length * (1.7 - (0.35 * Lane)), Oneway = TRUE, Lane < 3 \\ Length * \left(1.7 - \left(0.35 * \frac{Lane}{2}\right)\right), Oneway = FALSE \end{cases}$$

โดย roadW คือน้ำหนักของเส้นเชื่อม, Length คือความยาวของถนนช่วงนั้น, Lane เป็นจำนวนเลนของถนนช่วงนั้น และ Oneway เป็นตัวบ่งชี้ถนนนั้นแล่นรถทางเดียว

ผลที่ได้จากการแปลงข้อมูลตำแหน่งจีพีเอสเป็นเส้นทางบนถนน คือ เส้นทางบนถนนที่ใช้เดินทางเพื่อส่งผู้โดยสาร, ระยะทางรวมตลอดการเดินทาง และความเร็วเฉลี่ยที่ได้จากการคำนวณโดยใช้ระยะทางและเวลาระหว่างคู่พิกัด โดยวิธีการทั้งหมดถูกแสดงในภาพที่

```

Input: Roads = (Vertex, Edge), Edge = (direction, oneway, lane, length)
       Points = (latitude, longitude, Heading) , that ordered by time
Output: RoadPath are road list

Edge(Weight) = 
$$\begin{cases} Length * (1.7 - (0.35 * 3)), Oneway = TRUE, Lane \geq 3 \\ Length * (1.7 - (0.35 * Lane)), Oneway = TRUE, Lane < 3 \\ Length * \left(1.7 - \left(0.35 * \frac{Lane}{2}\right)\right), Oneway = FALSE \end{cases}$$


for (i = 1 to length(Points))
  NearRoads = filter road with in 150 meter from Pointsi
  for (j = 1 to length(NearRoads))
    DistFact = 
$$\frac{\text{Haversine distance}(\text{Points}_i, \text{NearRoads}_j)}{150}$$

    if ( NearRoadsj is oneway)
      HeadFact = cos(Pointsi(Heading) - NearRoadsj(direction))
    else
      HeadFact = abs(cos(Pointsi(Heading) - NearRoadsj(direction)))
    NearRoads_weightj = (1 + DistFact) × (1 + HeadFact)
  end
  Roadi = select Roads with top 5 NearRoads_weight
end
for (i = 1 to length(Roads) - 1 )
  for ( roadO in Roadsi )
    for ( roadD in Roads(i+1) )
      RoadPathi(roadO , roadD ) = shortedPath(roadO , roadD )
    end
  end
end
AllPath = InnerJoin(RoadPathi, RoadPathi+1, RoadPathi+2,
                   RoadPathi+3, RoadPathi+4, RoadPathi+5)
TripPath = Path (RoadPathi to RoadPathi+1) ∈ min_distance(AllPath )
for (i = 3 to length(Roads) - 5 )
  AllPath = InnerJoin(RoadPathi(roadO = tail(TripPath)), RoadPathi+1,
                    RoadPathi+2, RoadPathi+3, RoadPathi+4, RoadPathi+5)
  minPath = Path (RoadPathi to RoadPathi+1) ∈ min_distance(AllPath )
  TripPath = Concatenate(TripPath, minPath)
End
minPath = Path (RoadPathi to RoadPathi+5) ∈ min_distance(AllPath )
TripPath = Concatenate(TripPath, minPath)

```

ภาพที่ 24 วิธีการหาเส้นทางบนถนนทั้งหมดระหว่างช่วงถนนที่มีความน่าจะเป็นสูง



### 3.11. การจัดกลุ่มรูปแบบเส้นทางระหว่างพื้นที่

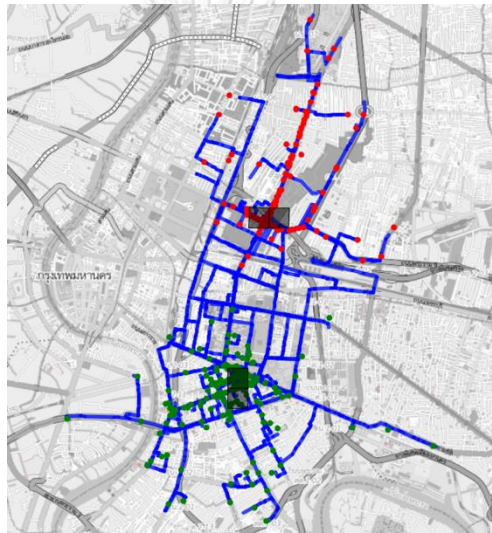
ขั้นตอนการจัดกลุ่มรูปแบบเส้นทางระหว่างพื้นที่เป็นขั้นตอนในการจัดกลุ่มเส้นทางที่มีความคล้ายกันของอันดับถนนที่ใช้เดินทางจากพื้นที่หนึ่งไปอีกพื้นที่หนึ่ง พื้นที่ที่ใช้เป็นตำแหน่งพื้นที่เริ่มต้นและพื้นที่ปลายทางที่ใช้ศึกษาจะอยู่ในรูปแบบกลุ่มของกริดพื้นที่เริ่มต้นและกลุ่มของกริดพื้นที่ปลายทาง ตัวอย่างตามภาพที่ 25 (a) เส้นทางที่ใช้เป็นส่วนใดส่วนหนึ่งของเส้นทางที่ใช้ส่งผู้โดยสารที่เดินทางออกจากกลุ่มของกริดพื้นที่เริ่มต้นและเดินทางไปยังกลุ่มของกริดพื้นที่ปลายทาง ภาพที่ 25 (b) เส้นทางที่จะถูกจัดกลุ่มเป็นช่วงของถนนสุดท้ายที่อยู่ในกลุ่มของกริดพื้นที่เริ่มต้นจนถึงช่วงถนนแรกที่อยู่ในกลุ่มของกริดพื้นที่ปลายทาง และภาพที่ 25 (c) เส้นทางที่จะถูกนำไปจัดกลุ่มเป็นลำดับของรหัสถนนตั้งแต่ถนนสุดท้ายที่ออกจากกลุ่มของกริดเริ่มต้นจนถึงถนนแรกที่เข้าไปในกลุ่มของกริดพื้นที่ปลายทาง



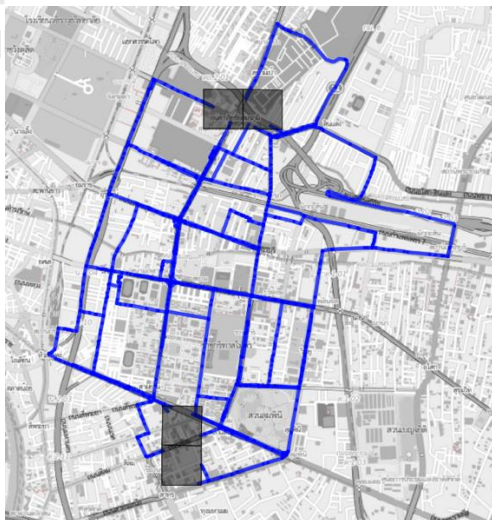




a) แสดงกลุ่มของกริดพื้นที่เริ่มต้นและกลุ่มของกริดพื้นที่ปลายทาง



b) แสดงเส้นทางที่ใช้ส่งผู้โดยสารที่ถูกนำมาใช้ในการหาเส้นทางจากพื้นที่ทั้งสอง  
จุดสีเขียวคือจุดที่ผู้โดยสารขึ้นรถแท็กซี่และจุดสีแดงคือจุดที่ผู้โดยสารออกจากรถแท็กซี่



c) แสดงเส้นทางที่จะถูกนำไปจัดกลุ่ม

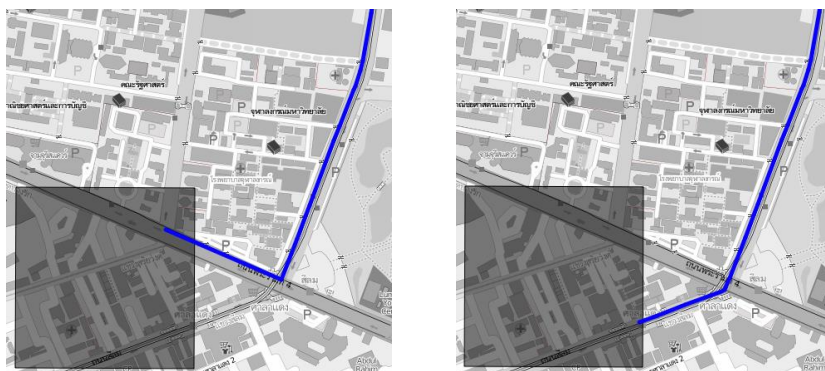
ภาพที่ 25 ตัวอย่างการเลือกเส้นทางที่จะนำไปจัดกลุ่ม

วิธีที่นำมาใช้ในการจัดกลุ่มไม่สามารถจัดกลุ่มโดยให้ลำดับของรหัสถนนที่เหมือนกันทั้งหมดระหว่างคู่ลำดับของรหัสถนนได้ เนื่องจากถึงแม้จะสามารถหาลำดับของช่วงถนนทุกช่วงที่มีความน่าจะเป็นสูงสุดที่รถแท็กซี่จะผ่าน แต่จากการทดลองพบว่ามีข้อผิดพลาดเล็กน้อยอยู่บ้างทำให้คู่ลำดับของรหัสถนนไม่เหมือนกัน 100 เปอร์เซ็นต์ เช่น ภาพที่ 26 (a) แสดงตัวอย่างผลที่เกิดจากความคลาดเคลื่อนของพิกัดจีพีเอสที่อยู่ใกล้กับส่วนของถนนที่เป็นสีแดงทำให้ระบบทำการแปลงพิกัดลงบนถนนสีแดงทำให้ลำดับถนนของทั้งสองมีความคลาดเคลื่อน และบางช่วงมีการใช้ถนนแตกต่างกับเล็กน้อยแต่ยังใช้ถนนหลักเส้นเดียวกัน (ภาพที่ 26 (b)) ทำให้ถ้าจัดกลุ่มโดยแต่ละกลุ่มต้องมีลำดับของรหัสถนนที่เหมือนกันทั้งหมดลำดับของถนนที่ผิดพลาดเล็กน้อยเหล่านี้ถูกจัดอยู่คนละกลุ่ม



ลำดับถนนฝั่งขวา	42063-31416-96480-64893-21426-83720-32916-58048
ลำดับถนนฝั่งซ้าย	42063-31416-96480-40450-40451-64893-21426-83720-32916-58048

a) แสดงตัวอย่างผลที่เกิดจากความคลาดเคลื่อนของพิกัดจีพีเอสที่อยู่ใกล้กับส่วนของถนนที่เป็นสีแดง



ลำดับถนนฝั่งขวา	442063-6546-290330-67005-19824-468578-234676-208451-...
ลำดับถนนฝั่งซ้าย	546051-290330-67005-19824-468578-234676-208451-...

b) แสดงตัวอย่างถนนที่ออกมาจากกริดที่ใช้ศึกษาอยู่บนถนนคนละเส้นกันแต่มาใช้ถนนร่วมกันภายหลัง

ภาพที่ 26 ตัวอย่างของเหตุการณ์ที่ทำให้ลำดับของถนนมีความคลาดเคลื่อน

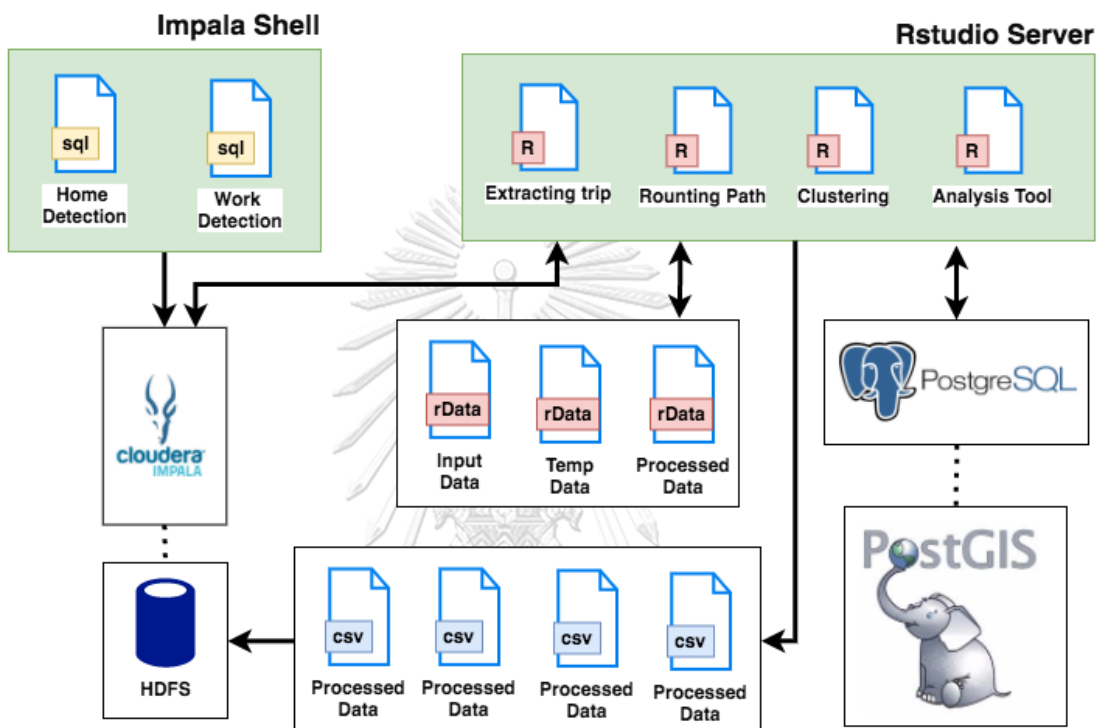
วิธีที่ถูกนำมาเพื่อแก้ปัญหาข้างต้นคือจะวัดความคล้ายของลำดับถนแต่ละชุด และใช้ Hierarchical Cluster รวมชุดของถนที่คล้ายกันเข้าเป็นกลุ่มเดียวกัน ซึ่งวิธีที่ใช้ในการวัดความคล้ายกันของลำดับถนในแต่ละเส้นทางจะใช้ Levenshtein distance เพื่อหาจำนวนครั้งที่ย่อยที่สุดที่ทำการเพิ่ม, ลบ หรือเปลี่ยนรหัสถนในคู่ลำดับของรหัสถนที่ต้องการเปรียบเทียบเพื่อให้ได้ลำดับของรหัสถนที่เหมือนกัน ทำให้ทราบถึงความคล้ายกันของคู่ลำดับรหัสถนที่เปรียบเทียบ แต่ในการทำ Hierarchical Cluster ไม่สามารถใช้ผลความคล้ายกันของคู่ลำดับรหัสถนโดยวิธี Levenshtein distance ได้โดยตรงกลุ่มของกริดพื้นที่เริ่มต้นที่อยู่ใกล้กับกลุ่มของกริดพื้นที่ปลายทางพบความคลาดเคลื่อนที่น้อยกว่า กริดพื้นที่เริ่มต้นกับกลุ่มของกริดพื้นที่ปลายทางที่มีระยะห่างมากกว่า ดังนั้นความคล้ายกันที่จะนำไปจัดกลุ่มโดยวิธี Hierarchical Cluster จะเป็นร้อยละของความคล้ายกันของคู่ลำดับรหัสถนที่เปรียบเทียบซึ่งผู้ศึกษาสามารถเลือกร้อยละของความคล้ายกันในการจัดกลุ่มได้



## บทที่ 4

### การทดลองและผลการทดลอง

#### 4.1. เครื่องมือที่ใช้ในการพัฒนา



ภาพที่ 27 เครื่องมือที่ใช้ในการพัฒนาระบบและแสดงการไหลของข้อมูลระหว่างเครื่องมือ

การพัฒนาโปรแกรมในงานวิจัยนี้พัฒนาเพื่อให้อาจรองรับข้อมูลขนาดใหญ่โดยอาศัยคอมพิวเตอร์หลายเครื่องช่วยกันประมวลผลโดยระบบพัฒนาโดยอาศัย 2 ภาษา คือ Impala SQL และภาษา R สาเหตุที่เลือกภาษา R เนื่องจากมีเครื่องมือที่ใช้จัดการกับข้อมูลเชิงพื้นที่จำนวนมาก และสามารถติดต่อฐานข้อมูลได้หลายรูปแบบ นอกจากนี้สามารถสร้างหน้าต่างติดต่อผู้ใช้ได้โดยตรง โดยภายในระบบข้อมูลจะถูกจัดเก็บอยู่ใน 4 รูปแบบเพื่อความเหมาะสมและความเร็วในการทำงาน คือ

- Impala เป็นโอเพ่นซอร์สระบบประมวลผลแบบขนานเพื่อจัดการกับข้อมูลขนาดใหญ่ โดยข้อมูลจะถูกเก็บในรูปแบบ HDFS ซึ่งข้อมูลจะถูกแบ่งและกระจายอยู่ในคลัสเตอร์ ซึ่ง Impala มีลักษณะเด่นคือใช้เวลาในการจัดการข้อมูลต่ำเมื่อเทียบกับเครื่องมือที่ใช้ในการจัดการข้อมูลขนาดใหญ่อื่น ๆ โดยระบบจะใช้เป็นที่จัดเก็บข้อมูลหลัก

- PostgreSQL เป็นระบบการจัดการฐานข้อมูลเชิงวัตถุ-สัมพันธ์ สามารถติดตั้งโปรแกรมเสริม PostGIS เพื่อทำการเก็บและค้นหาข้อมูลเชิงพื้นที่ได้ โดยในระบบนี้ใช้ในการเก็บข้อมูลถนน
- rData เป็นอาร์อีอบเจกต์ที่ใช้เก็บข้อมูลในภาษา R ซึ่งใช้ในเก็บข้อมูลตั้งต้น, ข้อมูลชั่วคราวและเก็บผลลัพธ์ สาเหตุที่ใช้อาร์อีอบเจกต์เพื่อทำการแบ่งข้อมูลตั้งต้นสำหรับการประมวลผลโดยอาศัยคอมพิวเตอร์หลายเครื่องโดยไม่จำเป็นต้องแก้ไขโปรแกรมและลดภาระในการประมวลผลจากการดึงข้อมูลจาก Impala
- csv (comma-separated values) เป็นการเก็บข้อมูลในรูปแบบตาราง โดยในแต่ละบรรทัดจะหมายถึงข้อมูลหนึ่งแถว ในแต่ละแถวข้อมูลจะถูกขึ้นด้วยเครื่องหมายจุลภาคแสดงถึงข้อมูลแต่ละคอลัมน์ โดยจะถูกนำมาใช้เมื่อต้องการนำข้อมูลเข้า HDFS ซึ่งวิธีการนำเพิ่มข้อมูล csv ไปวางบน HDFS โดยตรงเป็นวิธีที่เร็วกว่าทำการเพิ่มข้อมูลโดยอาศัยคำสั่ง SQL

## 4.2. ผลการทดลอง

### 4.2.1. ผลการทดลองการหาตำแหน่งบ้านและที่ทำงาน

จากจำนวนผู้ใช้งานทั้งหมด 4,340,833 คน จากซีดีอาร์และจี-ซีดีอาร์ พบว่าสามารถหาที่ทำงานได้ 3,320,135 คน และหาบ้านได้ 1,408,188 คน สาเหตุที่สามารถสกัดหาบ้านได้น้อยกว่าที่ทำงานเนื่องจากมีคนทำงานในกรุงเทพฯ แต่มีบ้านอยู่นอกกรุงเทพฯ หรือไม่ สามารถสกัดบ้านได้

การตรวจสอบความถูกต้องของตำแหน่งสามารถตรวจสอบได้แต่ตำแหน่งบ้านเท่านั้น เนื่องจากไม่มีการเก็บตำแหน่งที่ทำงานของลูกค้า แต่สามารถพิจารณาจากการกระจายตัวของสถานที่ทำงานที่จะพุดถึงในหัวข้อการกระจายตัวของตำแหน่งบ้านและที่ทำงาน ส่วนการตรวจสอบตำแหน่งบ้านจะทำการตรวจสอบจากลูกค้าระบบรายเดือนที่เก็บอยู่เพื่อส่งใบแจ้งค่าใช้จ่ายและใบเสร็จรับเงินทั้งหมด 1,391,575 คน โดยข้อมูลที่ใช้เป็นข้อมูลที่มีความละเอียดระดับตำบลเพื่อเป็นการป้องกันความเป็นส่วนตัวของลูกค้า โดยการตรวจสอบว่ากริดที่เป็นบ้านอยู่ในตำบลเดียวกับตำบลในที่อยู่เพื่อส่งใบแจ้งค่าใช้จ่ายหรือไม่

การเปรียบเทียบความถูกต้องของตำแหน่งบ้านและที่ทำงานจะทำการเปรียบเทียบกับวิธีการที่ทำการกรองให้เหลือแต่ข้อมูลที่อยู่กับที่ก่อนที่จะหาตำแหน่งบ้าน [18] โดยกำหนดค่าเวลาที่น้อยที่สุดที่จาก 2 การใช้งานที่ติดกันที่ 15 นาที, กำหนดค่าเวลาที่มากที่สุด

จาก 2 การใช้งานที่ติดกันที่ 3 ชั่วโมง และระยะห่างระหว่างตำแหน่งของผู้ใช้จาก 2 การใช้งานที่ติดกันที่ 1 กิโลเมตร

ผลการเปรียบเทียบพบว่าวิธีการที่ทำการกรองให้เหลือแต่ข้อมูลที่ใช้อยู่กับที่สามารถหาตำแหน่งบ้านได้ทั้งหมด 999,180 คน เป็นลูกค้ารายเดือน 244,613 คน คิดเป็น 17.57 % ของลูกค้าระบบรายเดือน วิธีในงานวิจัยนี้สามารถหาตำแหน่งบ้านได้ 1,408,188 คน เป็นลูกค้ารายเดือน 481,056 คน โดยทั้ง 2 วิธีมีความถูกต้องใกล้เคียงกันที่ 69.02 % และ 66.95 % สาเหตุที่วิธีการที่ทำการกรองให้เหลือแต่ข้อมูลที่ใช้อยู่กับที่สามารถหาตำแหน่งบ้านได้น้อยกว่าเนื่องจากปัจจุบันมีการใช้งานโทรศัพท์น้อย โดยเฉพาะในช่วงเวลากลางคืนซึ่งเป็นช่วงเวลาที่ใช้ในการหาตำแหน่งบ้าน ทำให้การหาตำแหน่งที่ผู้ใช้ที่อยู่กับที่จากการใช้งานโทรศัพท์ 2 ครั้งที่ติดกันมีโอกาสน้อยที่จะอยู่ในเงื่อนไขกรอบเวลา 15 นาที ถึง 3 ชั่วโมง

ตารางที่ 5 เปรียบเทียบผลการหาบ้านของผู้ใช้งานระหว่างวิธีการที่ทำการกรองผู้ใช้งานอยู่กับที่กับวิธีการหาชั่วโมงที่ผู้ใช้งานนอน

กระบวนการ	จำนวนผู้ใช้ที่ตรวจพบบ้าน	อัตราร้อยละผู้ใช้ที่ตรวจพบบ้าน	จำนวนผู้ใช้ระบบรายเดือนที่ตรวจพบบ้าน	อัตราร้อยละผู้ใช้ระบบรายเดือนที่ตรวจพบบ้าน	ความถูกต้อง	อัตราร้อยละความถูกต้อง
วิธีการกรองผู้ใช้งานอยู่กับที่	999,180	23.02 %	244,613	17.57 %	163,769	66.95 %
วิธีการหาชั่วโมงที่ผู้ใช้งานนอน	1,408,188	32.44 %	481,056	34.57 %	332,007	69.02 %

เมื่อทราบตำแหน่งบ้านและที่ทำงานทำให้ทราบถึงความหนาแน่นของตำแหน่งบ้านและที่ทำงานในแต่ละพื้นที่ รวมถึงระยะทางระหว่างบ้านและที่ทำงานซึ่งเป็นความต้องการในการเดินทางหลักของประชากรในเมือง สามารถใช้ในการตัดสินใจสร้างระบบขนส่งสาธารณะเพื่อรองรับความต้องการในการเดินทางกลุ่มนี้จะทำให้ปริมาณการใช้รถส่วนบุคคลลดลงส่งผลให้ปัญหาการจราจรติดขัดในชั่วโมงเริ่มดาวน์ลดลง โดยผู้วิจัยได้ทำการวิเคราะห์ข้อมูลจากตำแหน่งของบ้านและที่ทำงานของประชากรที่หาได้ทั้งหมด 1,408,188 คน ในบทที่ 4.4 หน้า 67

#### 4.2.2. ผลการทดลองการแปลงพิกัดจีพีเอสเป็นพิกัดบนถนน

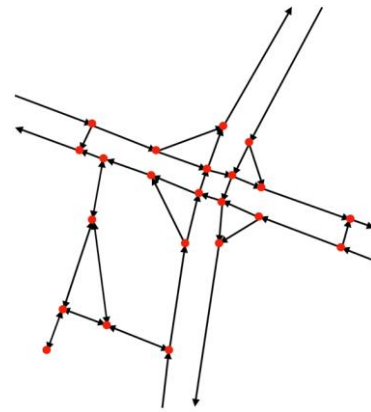
การแปลงพิกัดจีพีเอสเป็นพิกัดบนถนนเป็นการพัฒนาต่อยอดจาก [23] ที่มีความแม่นยำมากกว่า 95 % ที่รอบการส่งข้อมูล 1 นาที จากการสุ่มตรวจผลที่ได้จากการแปลงพิกัดจีพีเอสเป็นพิกัดบนถนนพบว่ามีความผิดพลาดซึ่งสามารถแยกได้เป็น 2 ปัจจัยคือความผิดพลาดจากข้อมูลถนนและความผิดพลาดจากวิธีในการแปลงพิกัดจีพีเอสเป็นพิกัดบนถนน โดย

##### 4.2.2.1. ความผิดพลาดที่ได้จากข้อมูลถนน

การแปลงพิกัดจีพีเอสเป็นพิกัดบนถนนอาศัยการหาเส้นทางที่สั้นที่สุดจากถนนที่ใกล้กับพิกัดจีพีเอสที่ติดกัน ดังนั้นจำเป็นต้องมีการสร้างกราฟจากข้อมูลถนน โดยให้จุดที่เชื่อมต่อระหว่างถนนเป็นจุดยอดและถนนเป็นเส้นเชื่อม โดยแต่ละเส้นเชื่อมอาจเป็นทั้งที่มีทิศทางเดียวหรือสองทิศทางขึ้นอยู่กับถนนนั้นเป็นถนนที่เล่นรถทางเดียวหรือไม่ ซึ่งจากการทดลองพบว่าเมื่อสร้างเป็นกราฟแล้วในบางเส้นทางสามารถผ่านได้ซึ่งในความเป็นจริงไม่สามารถผ่านได้ ตัวอย่างแสดงในภาพที่ 28 โดยเป็นข้อมูลถนนบริเวณแยกราชเทวี (ภาพที่ 28 (a)) สามารถสร้างเป็นกราฟการเชื่อมต่อของถนนได้ตามภาพภาพที่ 28 (b) โดยจุดสีแดงแสดงถึงจุดที่เป็นบริเวณเชื่อมต่อของช่วงถนนเส้นสีดำเส้นเชื่อมที่มีทิศทางแสดงถึงทิศทางที่รถสามารถผ่านได้ตามทิศทางของหัวลูกศร โดยเส้นทางจากถนนพญาไทเลี้ยวขวาไปถนนเพชรบุรีตามภาพที่ 28 (c) ซึ่งในความเป็นจริงไม่สามารถเลี้ยวขวาตามเส้นทางดังกล่าวได้จำเป็นต้องอ้อมไปถนนเส้นอื่น



(a) ข้อมูลเวกเตอร์ถนนแยกราชเทวี



(b) กราฟแสดงทิศทางที่รถสามารถผ่านได้

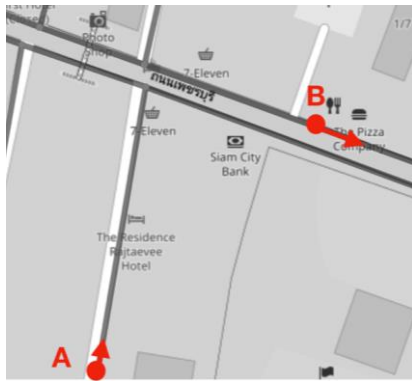


(c) เส้นทางที่กราฟแสดงเป็นเส้นทางจากถนนพญาไทเลี้ยวขวาไปถนนเพชรบุรี

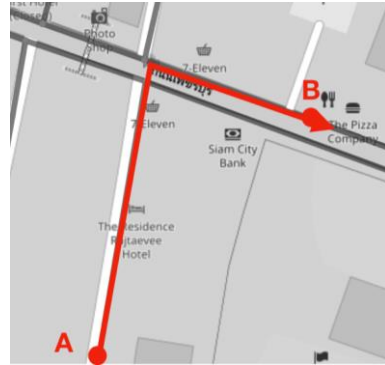
ภาพที่ 28 ตัวอย่างข้อผิดพลาดเส้นทางที่ไม่สามารถใช้ได้จริง

ข้อมูลถนนที่ได้เป็นข้อมูลมีตัวบ่งชี้ว่าถนนนั้นเป็นถนนที่เล่นทางเดียวหรือไม่ ดังนั้นข้อมูลถนนที่มีตัวบ่งชี้ว่าถนนนั้นเป็นถนนที่เล่นทางเดียวที่ขนานกัน 2 เส้น ในทิศทางตรงกันข้ามกันจะสันนิษฐานได้ว่ามีเกาะกลางถนนกั้นระหว่างถนน 2 เส้น แต่จากการตรวจสอบพบว่าในถนนบางเส้นไม่มีเกาะกลางถนนกั้นระหว่างถนนทำให้ในความเป็นจริงสามารถข้ามถนนเพื่อเข้าซอยได้หรือออกจากซอยเพื่อเข้าถนนหลักได้ทั้งสองทิศทาง แต่เนื่องจากข้อมูลถนนทำการแยกเป็นถนนคนละเส้นกันทำให้จำเป็นต้องไปหาเส้นทางเพื่อทำการกลับรถ ซึ่งแสดงตัวอย่างในภาพที่ 29 เป็นการเดินทางจากซอยเพชรบุรี 20 เพื่อเดินทางไปแยกประตูน้ำ โดยเส้นทางที่สั้นที่สุดแสดงในภาพที่ 29 (b) คือเลี้ยวขวาที่ถนนเพชรบุรี แต่เนื่องจากข้อมูลถนนเพชรบุรีพบว่าไม่มีเส้นเชื่อมจากช่วงถนนจากซอยเพชรบุรี 20 ไปยังถนนเพชรบุรีในฝั่งที่เดินทางไปแยกประตูน้ำ ทำให้ต้องเลี้ยวซ้ายเพื่อไปกลับรถที่แยกราชเทวี ตามภาพที่ 29 (c)

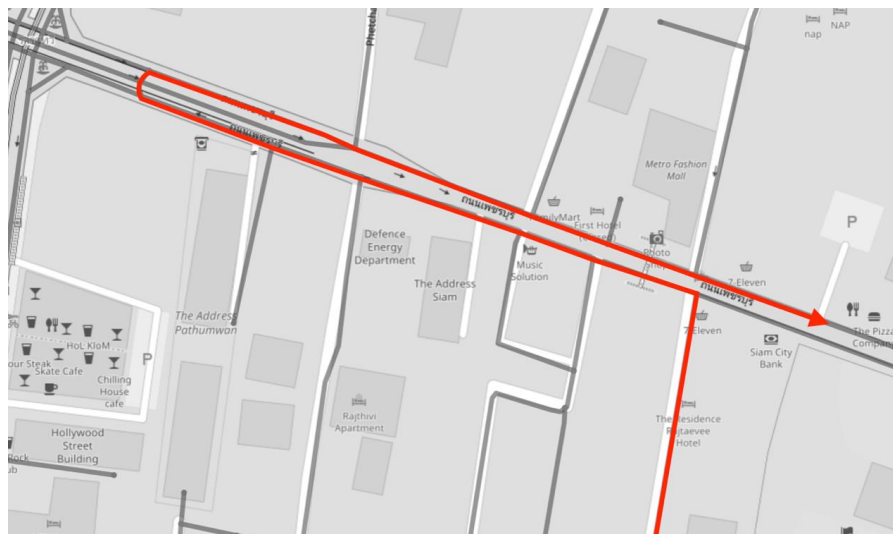




(a) พิกัดจากจุด A ไป B



(b) เส้นทางที่ดีที่สุดที่สามารถเดินทางไปได้

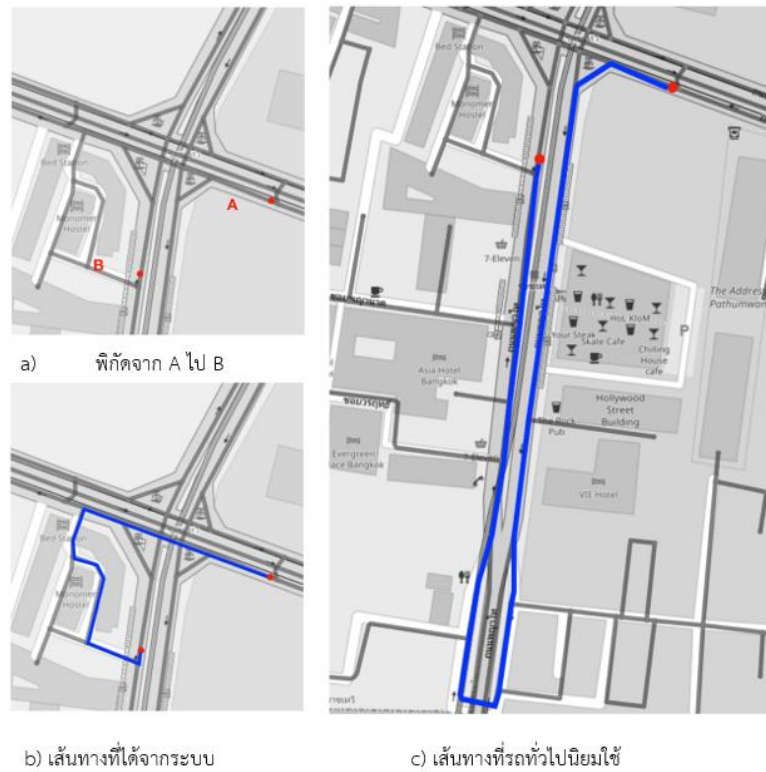


(c) เส้นทางที่ได้จากระบบ

ภาพที่ 29 ตัวอย่างข้อผิดพลาดจากไม่มีเส้นเชื่อมในกรณีถนนหลักไม่มีเกาะกลางถนน

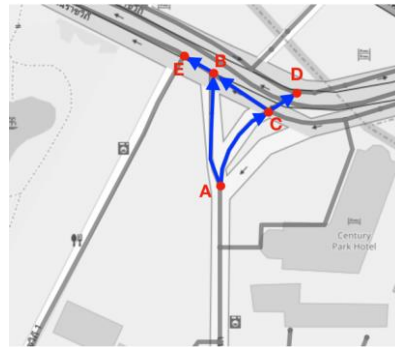
#### 4.2.2.2. ความผิดพลาดจากวิธีการแปลงพิกัดจีพีเอสเป็นพิกัดบนถนน

การแปลงพิกัดจีพีเอสเป็นพิกัดบนถนนอาศัยการหาเส้นทางที่สั้นที่สุดจากถนนที่ใกล้กับพิกัดจีพีเอสที่ติดกัน ดังนั้นเส้นทางที่ได้จะพยายามหาทางที่สั้นที่สุดซึ่งในบางครั้งจะได้เส้นทางที่เข้าไปที่ข้อย่อยเล็ก ๆ ที่รถทั่วไปไม่นิยมใช้ แสดงในภาพที่ 30 แสดงข้อผิดพลาดบริเวณแยกราชเทวีที่ระบบให้เส้นทางที่เข้าไปในซอยซึ่งเป็นเส้นทางที่สามารถเดินทางได้จริงแต่ไม่เป็นที่นิยม

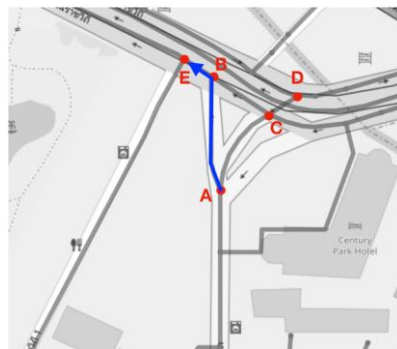


ภาพที่ 30 ข้อผิดพลาดที่ได้จากการใช้เส้นทางที่สั้นที่สุดในการหาเส้นทางที่ดีที่สุด

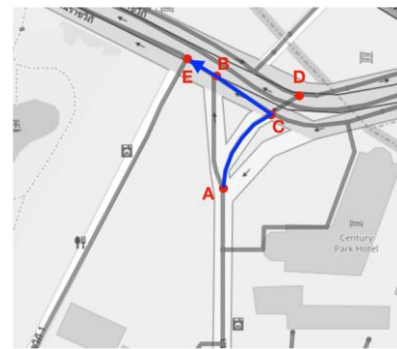
อีกปัญหาหนึ่งคือการใช้จำนวนเลนแทนลำดับความสำคัญของถนนซึ่งจากการทดลองจะมีปัญหาในทางแยกที่มีเลนสำหรับเลี้ยวขวาโดยเฉพาะ ซึ่งปกติมักจะมีจำนวนน้อยกว่าเส้นทางหลัก ซึ่งแสดงในภาพที่ 31 โดยภาพที่ 31 (a) แสดงกราฟบริเวณถนนราชปรารภตัดกับถนนราชวิถี โดยถนนสำหรับที่ใช้เลี้ยวโดยเฉพาะจาก A ไป B มีขนาด 1 เลน แต่ถนนที่ใช้เดินทางจาก A ไป C มีขนาด 4 เลน ส่งผลให้เส้นทางที่ได้จากระบบเป็นเส้นทางจาก A ไป C ไป B ไป E แทนที่จะเป็นเส้นทางจาก A ไป B ไป E



(a) กราฟแสดงเส้นทางที่รถสามารถเดินทางได้



(b) แสดงเส้นทางที่ควรจะเป็น



(a) เส้นทางที่เกิดความผิดพลาดจากขนาดเลน

### ภาพที่ 31 ข้อผิดพลาดที่เกิดจากจำนวนเลนถนน

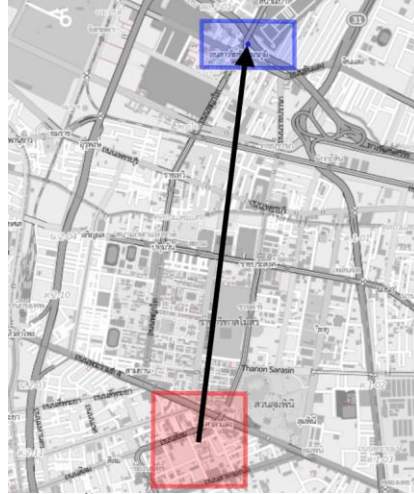
จากการทดลองพบว่าความผิดพลาดที่เกิดขึ้นเกือบทั้งหมดเกิดขึ้นบริเวณแยกต่าง ๆ โดยถนนหลักที่รถแท็กซี่ใช้ยังมีความถูกต้องอยู่ ซึ่งในงานวิจัยนี้ได้เพิ่มกระบวนการกรองผลที่ได้จากการแปลงให้มีความถูกต้องมากขึ้นโดยทำการพิจารณาความเร็วที่รถต้องใช้ในการเดินทางบนถนนที่ได้จากการแปลงต้องใช้ความเร็วไม่เกิน 150 กิโลเมตรต่อชั่วโมง

#### 4.2.3. ผลการทดลองการจัดกลุ่มเส้นทางที่ใช้ส่งผู้โดยสารระหว่างพื้นที่

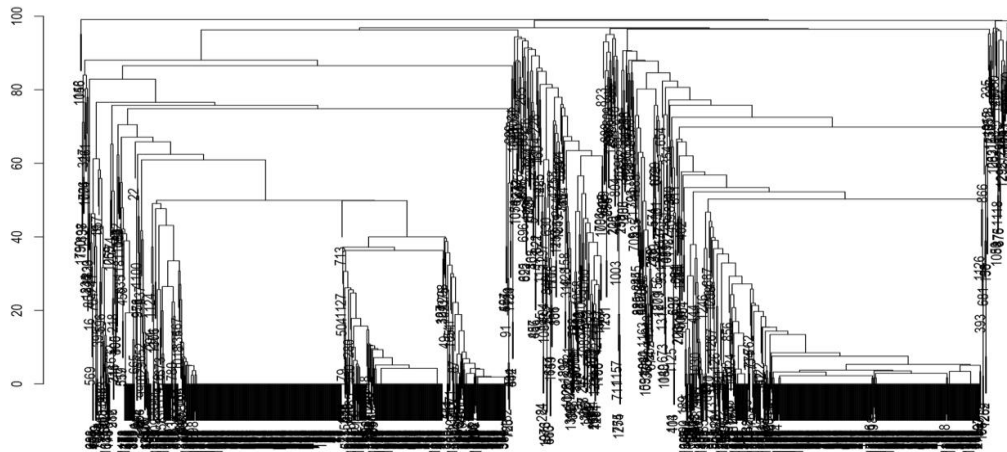
การทดสอบการจัดกลุ่มเส้นทางที่ใช้ส่งผู้โดยสารเป็นการสุ่มตรวจสอบผลการจัดกลุ่มโดยความถูกต้องพิจารณาจากถนนเส้นหลักที่ใช้ถูกจัดอยู่ในกลุ่มเดียวกันหรือไม่ โดยทดสอบ 2 แบบตามระยะทางระหว่างพื้นที่ที่เป็นพื้นที่เริ่มต้นไปยังพื้นที่ปลายทาง

การทดสอบแรกเป็นการทดสอบการจัดกลุ่มในระยะใกล้โดยใช้พื้นที่ทดสอบเป็นถนนสี่เลนและปลายทางเป็นอนุสาวรีย์ชัยสมรภูมิ (ภาพที่ 32 (a)) ระยะทาง 4.28 กิโลเมตร มีทั้งหมด 1,311 การเดินทาง โดยความคล้ายของแต่ละเส้นทางแสดงโดยเดนโดรแกรมของสัดส่วนจำนวนครั้งในการแก้ไขถนนต่อจำนวนถนนทั้งหมด (ภาพที่ 34 (b)) พบว่าเส้นทางจำนวนมากที่ความคล้ายกันเกือบ 100 % โดยจัดกลุ่มเส้นทางที่คล้ายกันมากกว่า 80 % ได้

ทั้งหมด 372 กลุ่ม มี 7 กลุ่มที่มีสมาชิกมากกว่า 1 % ของเส้นทางทั้งหมด ซึ่งทั้ง 7 กลุ่มมีสมาชิกรวมคิดเป็น 60 % ของการเดินทางทั้งหมด



(a) พื้นที่ที่ใช้ทดสอบจัดกลุ่มระยะใกล้



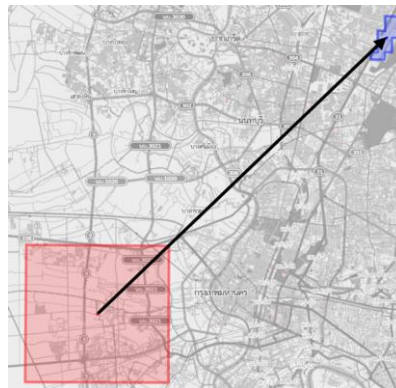
(b) เคนโดแกรมการจัดกลุ่มแบบลำดับขั้น



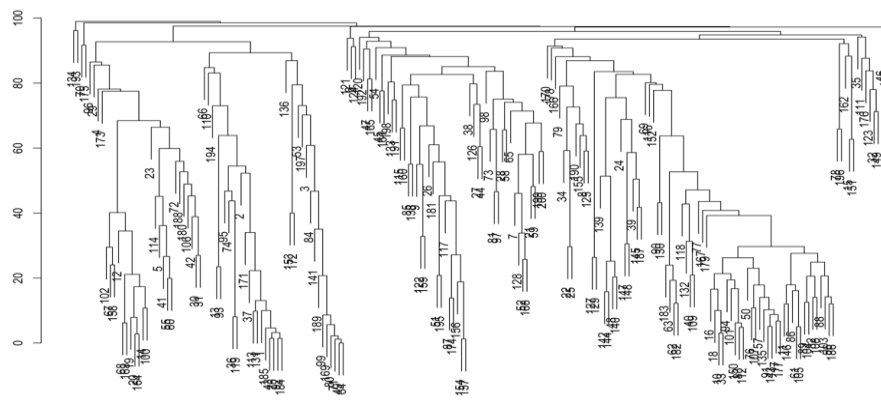
(c) ผลการจัดกลุ่มเส้นทางที่ถูกใช้สูงสุด 3 อันดับ เรียงจากซ้ายไปขวา

ภาพที่ 32 การทดสอบการจัดกลุ่มในระยะใกล้

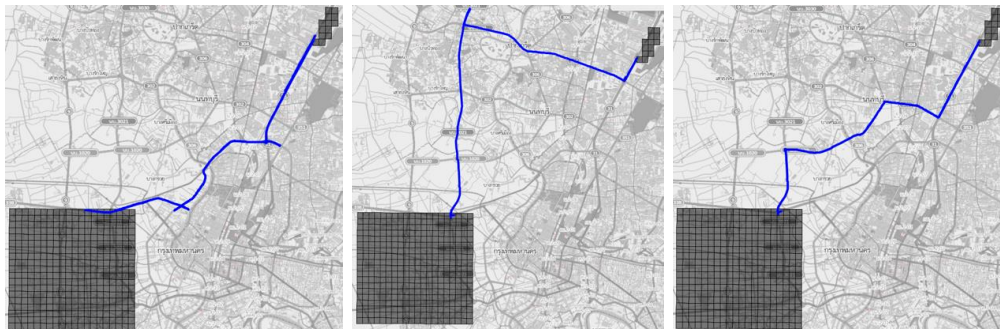
การทดสอบที่สองเป็นการทดสอบการจัดกลุ่มระยะไกลจากพื้นที่ในฝั่งธนบุรีไปยังท่าอากาศยานนานาชาติดอนเมือง (ภาพที่ 33 a) ระยะทาง 27.87 กิโลเมตร มีทั้งหมด 200 การเดินทาง โดยความคล้ายของแต่ละเส้นทางแสดงโดยเดนโตรแกรมสัดส่วนจำนวนครั้งในการแก้ไขถนนต่อจำนวนถนนทั้งหมด (ภาพที่ 33 b) พบว่าเปอร์เซ็นต์ความคล้ายลดลงกว่าการทดลองระยะใกล้ โดยจัดกลุ่มเส้นทางที่คล้ายกันมากกว่า 80 % มี 10 กลุ่มที่มีสมาชิกมากกว่า 1 % ของเส้นทางทั้งหมด โดยทั้ง 10 กลุ่มมีสมาชิกรวมคิดเป็น 34 % ของการเดินทางทั้งหมด



(a) พื้นที่ที่ใช้ทดสอบจัดกลุ่มระยะไกล



(b) เดนโตรแกรมการจัดกลุ่มแบบลำดับชั้น



(c) ผลการจัดกลุ่มเส้นทางที่ถูกใช้สูงสุด 3 อันดับ เรียงจากซ้ายไปขวา

ภาพที่ 33 การทดสอบการจัดกลุ่มในระยะไกล



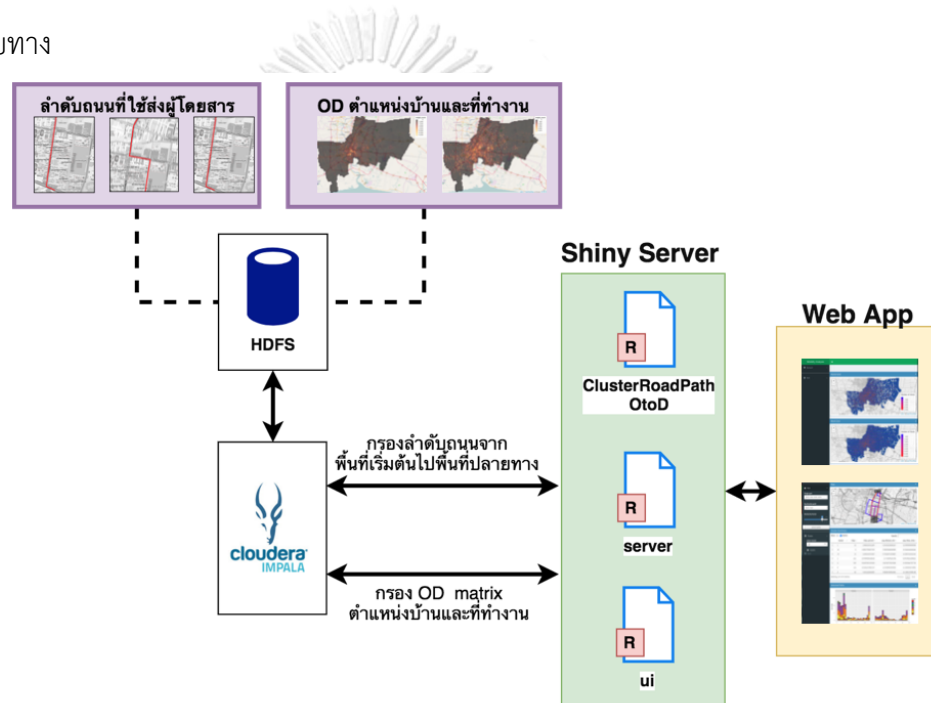
การจัดกลุ่มเส้นทางที่ใช้ส่งผู้โดยสารเป็นการจัดกลุ่มโดยอาศัยการเปรียบเทียบความคล้ายกันจากลำดับช่วงถนนที่รถแท็กซี่ผ่านโดยใช้วิธี Levenshtein distance และคำนวณเป็นร้อยละของความคล้ายระหว่างคู่เส้นทางที่ใช้ส่งผู้โดยสาร ซึ่งจากการทดลองพบว่าร้อยละของความคล้ายระหว่างคู่เส้นทางที่ใช้ส่งผู้โดยสารขึ้นอยู่กับมีเส้นทางที่ใช้เดินทางได้มากน้อยเพียงใด ถ้ามีเส้นทางที่หลากหลายจะส่งผลให้ร้อยละของความคล้ายระหว่างคู่เส้นทางที่ใช้ส่งผู้โดยสารลดลง ซึ่งแสดงตัวอย่างจากภาพที่ 32(b) และ 33(b) พบว่าแผนผังการจราจรจัดกลุ่มแบบลำดับขั้นทั้งสองมีความต่างกันโดยภาพที่ 32(b) แต่ละเส้นทางเดียวกันที่ความคล้ายที่สูงมากจนเหมือนเกือบ 100 % ทำให้สามารถใช้ค่าที่สูงในการจัดกลุ่มได้ ส่วนภาพที่ 33(b) มีความคล้ายกันน้อยกว่าทำให้ต้องใช้ค่าความคล้ายที่น้อยกว่าเพื่อทำการจัดกลุ่มเพราะถ้าใช้ค่าที่สูงจะทำให้เส้นทางจะทำให้การเดินทางที่มีช่วงที่เดินทางคล้ายกันแยกออกจากกัน และค่าความคล้ายที่ใช้จัดกลุ่มนี้จะส่งผลต่อความละเอียดในการจัดกลุ่มขึ้นซึ่งแสดงในภาพที่ 34 เป็นการเดินทางจากฝั่งธนบุรีไปสยามสแควร์ที่มีเส้นทางที่ในช่วงแรกและช่วงท้ายเหมือนกันแต่ต่างกันตรงเส้นทางที่ใช้ในช่วงกลาง โดยมีความคล้ายกันที่ 79 % ซึ่งความละเอียดในการจัดกลุ่มขึ้นอยู่กับความต้องการของผู้ใช้แต่ละคน ดังนั้นในการจัดกลุ่มเส้นทางที่ใช้ส่งผู้โดยสารจะให้ผู้ใช้งานเป็นผู้กำหนดค่าร้อยละของความคล้ายระหว่างคู่เส้นทางที่ใช้ส่งผู้โดยสารตามระดับความคล้ายที่ผู้ใช้งานต้องการ



ภาพที่ 34 ตัวอย่างเส้นทางที่มีความคล้ายกันที่ 79 %

#### 4.3. เครื่องมือวิเคราะห์ความต้องการและเส้นทางที่นิยมใช้ในการเดินทาง

เครื่องมือวิเคราะห์ความต้องการและเส้นทางที่นิยมใช้ในการเดินทางพัฒนาด้วยภาษา R โดยอาศัย Shiny package สร้างเว็บแอปพลิเคชันเพื่อให้ผู้ใช้สามารถเลือกพื้นที่ที่ต้องการศึกษา และแสดงผลในแผนที่แบบโต้ตอบพร้อมค่าทางสถิติ ภาพที่ 35 แสดงสถาปัตยกรรมระบบวิเคราะห์ความต้องการและเส้นทางที่นิยม โดยข้อมูลตั้งต้นจะถูกเก็บใน HDFS และเรียกใช้ผ่าน Impala ซึ่งประกอบด้วยข้อมูล OD-matrix ระหว่างกริดบ้านและที่ทำงานและลำดับถนนที่ใช้ในการส่งผู้โดยสาร ระบบได้แบ่งเป็นสองส่วนคือส่วนศึกษาปริมาณความต้องการในการเดินทางจากตำแหน่งบ้านและสถานที่ทำงาน และส่วนที่สองคือเส้นทางที่นิยมในการเดินทางระหว่างพื้นที่ตั้งต้นไปยังพื้นที่ปลายทาง

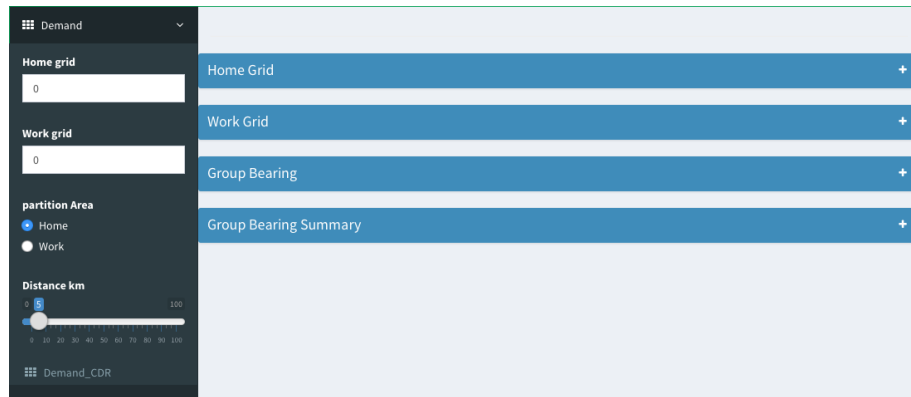


ภาพที่ 35 สถาปัตยกรรมระบบวิเคราะห์ความต้องการและเส้นทางที่นิยม

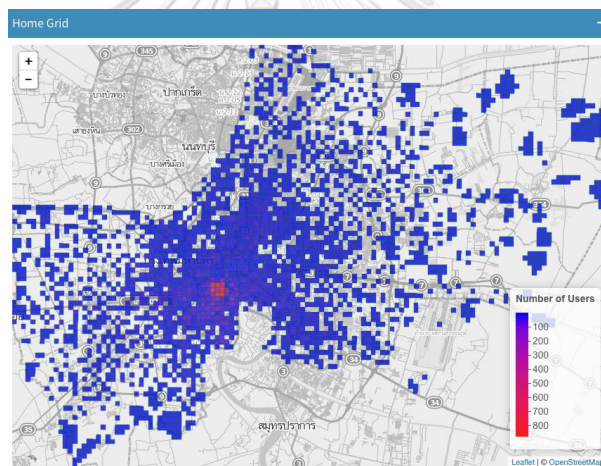
##### 4.3.1. ปริมาณความต้องการในการเดินทางจากตำแหน่งบ้านและสถานที่ทำงาน

ส่วนศึกษาปริมาณความต้องการในการเดินทางจากตำแหน่งบ้านและสถานที่ทำงานแสดงในภาพที่ 36 โดยจะมีพื้นที่ให้ผู้ใช้งานใส่กลุ่มของกริดที่เป็นตำแหน่งบ้านและที่ทำงานเพื่อทำการกรองข้อมูลตำแหน่งบ้านและที่ทำงานให้เหลือเฉพาะในพื้นที่ที่ต้องการศึกษา แต่ถ้าไม่ต้องการกรองให้ใส่ค่าเป็น 0 เมื่อทำการเลือกกลุ่มของกริดที่ต้องการศึกษาเรียบร้อยแล้วระบบจะทำการแสดงความหนาแน่นของตำแหน่งบ้านและที่ทำงานจากข้อมูลที่ผ่านการกรองในรูปแบบแผนที่แบบโต้ตอบ โดยจะแสดงความหนาแน่นของตำแหน่งบ้านในแต่ละกริดตามภาพที่ 37(a) ในกล่อง Home Grid และแสดงความหนาแน่นของตำแหน่งที่

ทำงานในแต่ละกริดตามภาพที่ 37(a) ในกล่อง Work Grid โดยจากตัวอย่างภาพที่ 37 (a) และ (b) เป็นการเลือกพื้นที่บ้านเป็น 0 ซึ่งหมายถึงพื้นที่ทุกกริด และเลือกพื้นที่บ้านเป็นกริดบริเวณสีลมและช่องนนทรี



ภาพที่ 36 เครื่องมือศึกษาปริมาณความต้องการในการเดินทางจากตำแหน่งบ้านและสถานที่ทำงาน



(a) กล่องแสดงแผนที่แบบโต้ตอบความหนาแน่นของตำแหน่งบ้าน

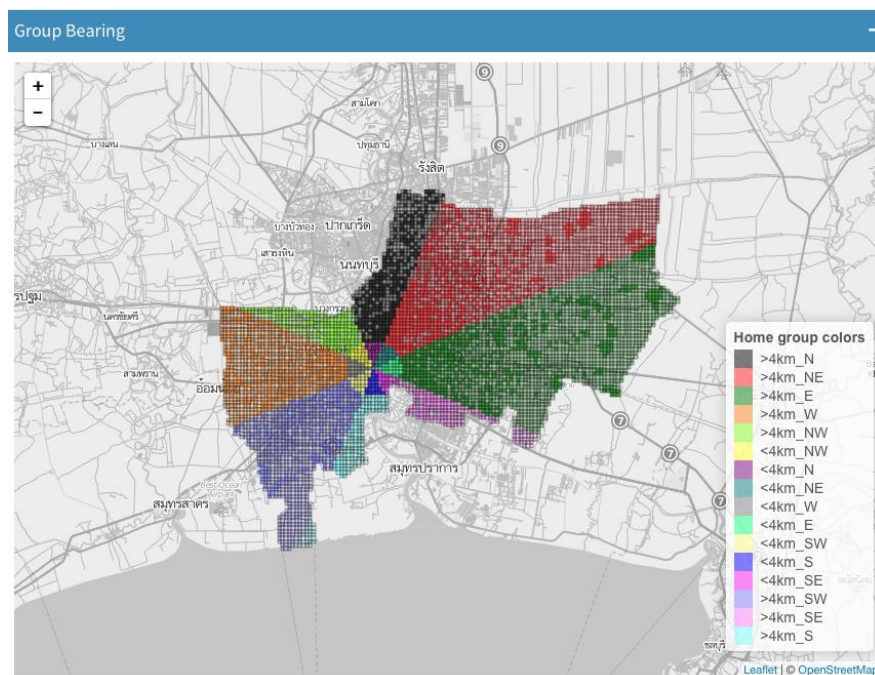


(b) กล่องแสดงแผนที่แบบโต้ตอบความหนาแน่นของตำแหน่งที่ทำงาน

ภาพที่ 37 กล่องแสดงแผนที่แบบโต้ตอบความหนาแน่นของตำแหน่งบ้านและที่ทำงาน



นอกจากนี้ระบบสามารถแบ่งพื้นที่ออกเป็น 16 กลุ่มตามทิศทางระหว่างบ้านกับที่ทำงาน 8 ทิศ และตามระยะทางจากบ้านกับที่ทำงานเพื่อศึกษาปริมาณจำนวนคนที่เดินทางมาจากกลุ่มต่าง ๆ โดยให้ผู้ใช้ปรับค่า Partition Area เพื่อเลือกระยะทางจากบ้านกับที่ทำงานเพื่อใช้ในการแยกตำแหน่งบ้านที่ใกล้และไกลออกจากกัน ผลที่ได้จะแสดงพื้นที่ที่ถูกแบ่งในภาพที่ 38 ในกล่อง Group Bearing และจำนวนคนในแต่ละกลุ่มจากภาพที่ 39 ในกล่อง Group Bearing Summary



ภาพที่ 38 กล่องแสดงแผนที่แบบโต้ตอบพื้นที่ที่ถูกแบ่งออกเป็น 16 กลุ่ม

Group Bearing Summary

Show 10 entries Search:

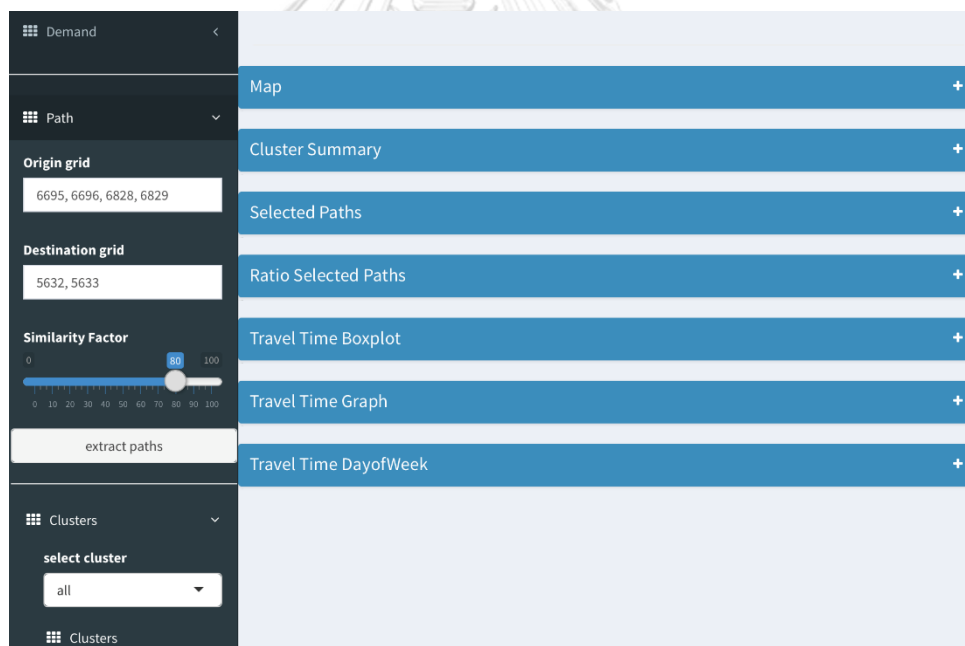
	Displacement_Patition	bearing_name	freq_Patition
1	All	All	20406.5
2119	>4km	N	1249
2725	>4km	NE	2306.65274725275
415	>4km	E	2130.84725274725
6791	>4km	W	2002.55555555556
4723	>4km	NW	623.944444444444
4498	<4km	NW	874.5
2107	<4km	N	950
2638	<4km	NE	798.5
6029	<4km	W	1281

Showing 1 to 10 of 17 entries Previous 1 2 Next

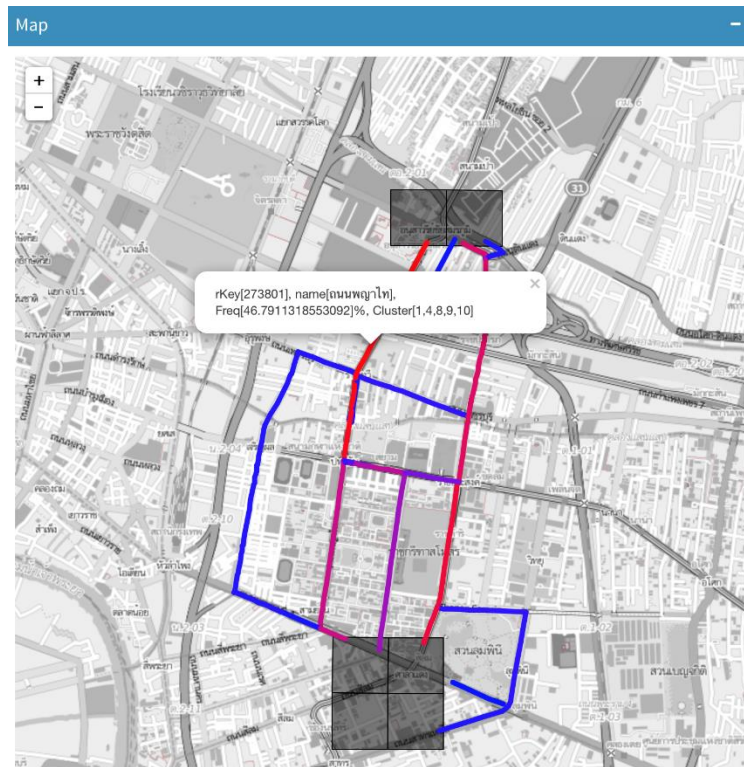
ภาพที่ 39 ตารางแสดงจำนวนคนที่เดินทางจากกลุ่มต่าง ๆ

#### 4.3.2. เส้นทางที่นิยมในการเดินทางระหว่างพื้นที่ตั้งต้นไปยังพื้นที่ปลายทาง

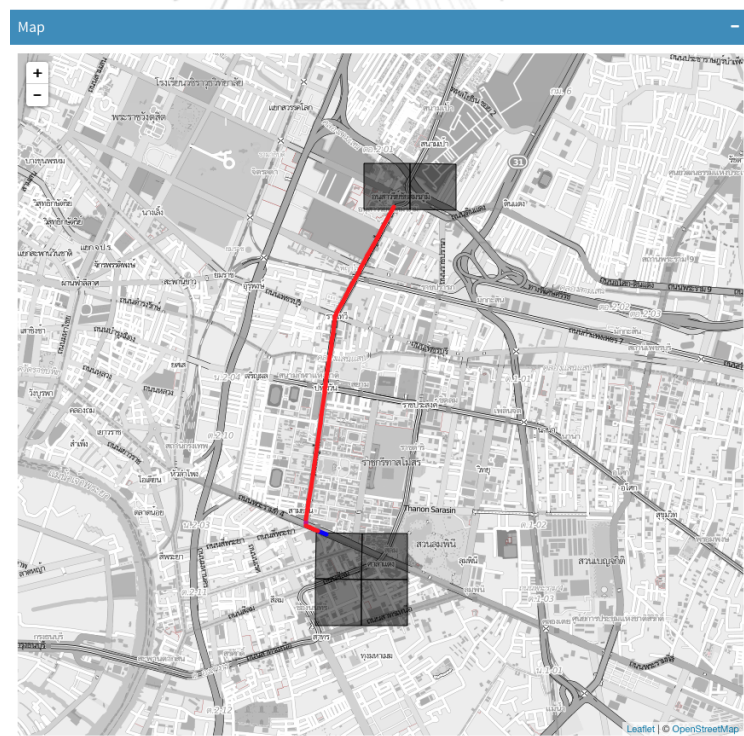
ส่วนศึกษาเส้นทางที่นิยมในการเดินทางระหว่างพื้นที่ตั้งต้นไปยังพื้นที่ปลายทาง แสดงในภาพที่ 40 โดยมีพื้นที่ให้ผู้ใช้กรอกกลุ่มของกริดที่เป็นพื้นที่เริ่มต้นของการเดินทาง และกลุ่มของกริดที่เป็นพื้นที่ปลายทาง พร้อมกับร้อยละของความคล้ายที่นำมาใช้ในการจัดกลุ่ม (Similarity factor) หลังจากกำหนดค่าทั้งหมดเรียบร้อยแล้วทำการจัดกลุ่มโดยการกดปุ่ม extract paths ระบบจะทำการกรองลำดับของช่วงถนนที่ใช้ส่งผู้โดยสารให้เหลือลำดับของช่วงถนนตั้งแต่ออกจากกลุ่มของกริดที่เป็นจุดเริ่มต้นจนถึงช่วงถนนแรกที่เข้ากลุ่มของกริดปลายทางและทำการจัดกลุ่มถนนที่เป็นที่นิยมตามความคล้ายของช่วงถนนตามที่ใช้ตั้งไว้ที่ Similarity factor ผลที่ได้จะแสดงตามภาพที่ 41(a) ในกล่อง Map ที่เป็นช่วงถนนที่เส้นทางที่นิยมใช้โดยสามารถกดเข้าไปดูสัดส่วนของการเดินทางที่ใช้ถนนเส้นนั้นกับการเดินทางระหว่างกลุ่มกริดที่ศึกษาทั้งหมดพร้อมกับแสดงรหัสกลุ่มที่ใช้ถนนเส้นนั้น โดยผู้ใช้สามารถเลือกดูเส้นทางเฉพาะกลุ่มได้โดยผลจะแสดงเฉพาะเส้นทางในกลุ่มนั้นตามภาพที่ 41(b)



ภาพที่ 40 เครื่องมือศึกษาเส้นทางที่นิยมในการระหว่างพื้นที่ตั้งต้นไปยังพื้นที่ปลายทาง



(a) กล่องแสดงแผนที่แบบโต้ตอบถนนที่นิยมใช้ในการเดินทาง



(b) กล่องแสดงแผนที่แบบโต้ตอบแสดงเส้นทางที่นิยมเพียงกลุ่มเดียว

ภาพที่ 41 กล่องแสดงแผนที่แบบโต้ตอบสำหรับศึกษาเส้นทางที่นิยมใช้ในการเดินทาง

นอกจากนี้ระบบจะแสดงค่าทางสถิติของแต่ละกลุ่มประกอบด้วย

- ตารางสรุปข้อมูลแต่ละกลุ่ม (Cluster Summary)

ตารางสรุปข้อมูลแต่ละกลุ่มประกอบไปด้วย รหัสกลุ่มเส้นทางเพื่อใช้อ้างอิงกับกราฟอื่น, จำนวนการเดินทางที่เลือกใช้กลุ่มเส้นทาง, ร้อยละของการเลือกเส้นทาง, ระยะทางเฉลี่ย และเวลาเฉลี่ยที่ใช้ในการเดินทาง

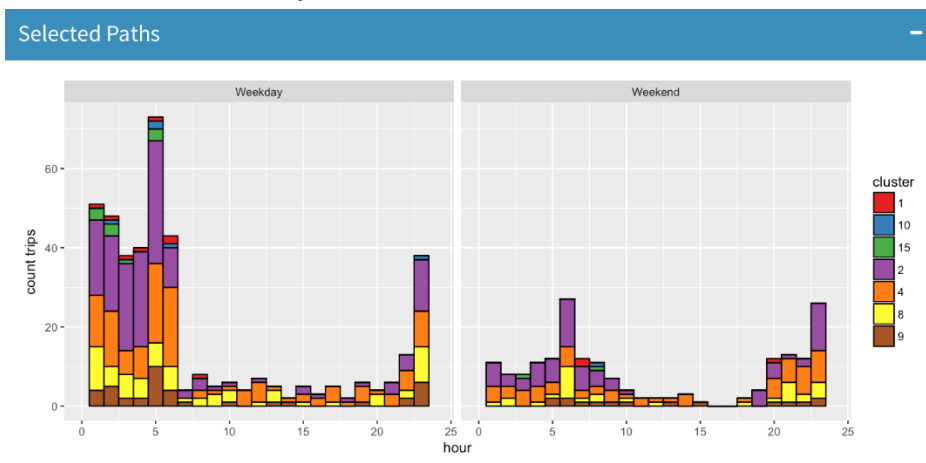
Cluster Summary					
Show 10 entries		Search:			
cluster	trips	trips_percent	avg_distance_km	avg_times_mins	
1	12	1.4002333722287	5.01532516606327	13.59444444444444	
2	9	1.05017502917153	5.39232046426092	11.04444444444444	
3	12	1.4002333722287	5.79182972205887	11.00972222222222	
4	252	29.4049008168028	3.7776994255798	9.64146825396825	
5	203	23.6872812135356	3.81520775912006	11.8259441707718	
6	114	13.3022170361727	4.27284902353564	11.4340643274854	
7	63	7.3512252042007	4.60653736652562	12.1566137566138	

Showing 1 to 7 of 7 entries Previous  Next

ภาพที่ 42 กล้องตารางสรุปข้อมูลแต่ละกลุ่ม

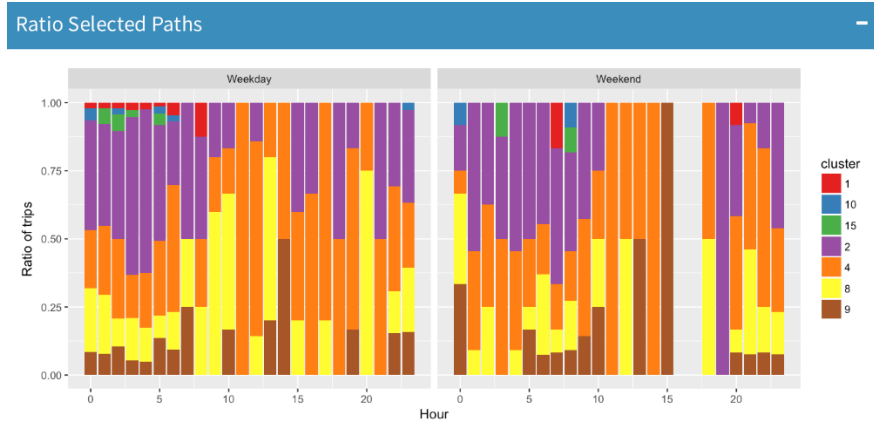
- จำนวนการเดินทางที่เลือกใช้เส้นทางในแต่ละกลุ่มในช่วงเวลาต่าง ๆ (Selected Paths)

กล่องแสดงจำนวนการเดินทางที่เลือกใช้เส้นทางในแต่ละกลุ่มในช่วงเวลาต่าง ๆ เป็นกราฟแบบเรียงทับซ้อนของจำนวนการเดินทางที่เลือกใช้กลุ่มการเดินทางต่าง ๆ ในแต่ละชั่วโมง ทำให้ทราบถึงจำนวนการเดินทางด้วยรถแท็กซี่ระหว่างพื้นที่เริ่มต้นไปพื้นที่ปลายทางและเส้นทางที่ถูกเลือกในแต่ละช่วงเวลา



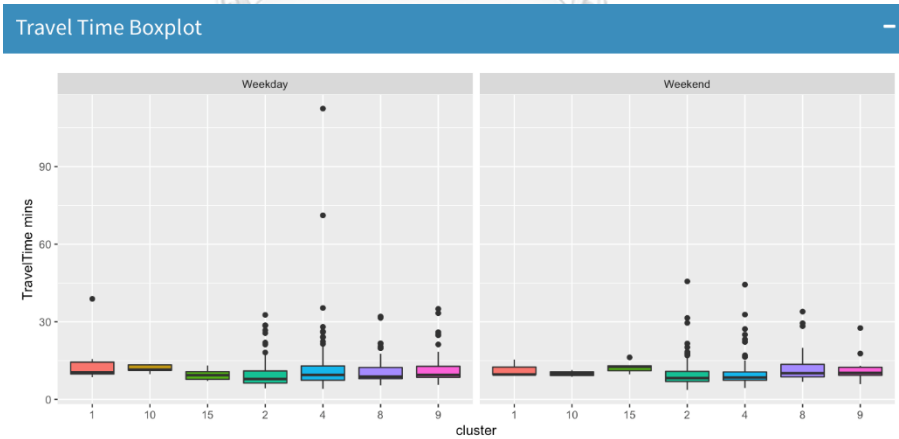
ภาพที่ 43 กล่องแสดงจำนวนการเดินทางที่เลือกใช้เส้นทางในแต่ละกลุ่มในช่วงเวลาต่าง ๆ

- สัดส่วนการเลือกใช้กลุ่มเส้นทางยอดนิยม (Ratio Selected Paths)  
 กล่องแสดงสัดส่วนการเลือกใช้กลุ่มเส้นทางจากกลุ่มเส้นทางที่นิยมใช้ทั้งหมด ทำให้ทราบถึงการเปลี่ยนแปลงเส้นทางในแต่ละช่วงเวลา



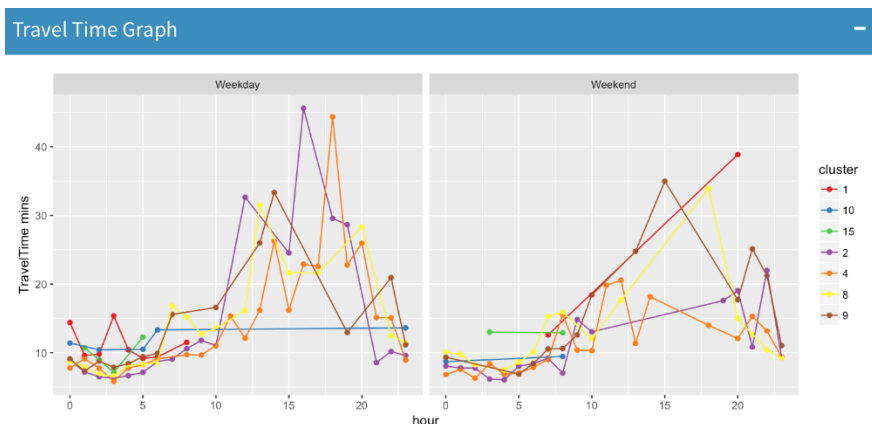
ภาพที่ 44 สัดส่วนการเลือกใช้กลุ่มเส้นทางยอดนิยม

- แผนภาพกล่องเวลาที่ใช้ในการเดินทางของกลุ่มเส้นทางที่นิยม (TravelTime boxplot)  
 ส่วนแสดงแผนภาพกล่องเวลาที่ใช้ในการเดินทางของกลุ่มเส้นทางที่นิยมทำให้ทราบถึงค่าทางสถิติของเวลาที่ใช้ในการเดินทางของแต่ละกลุ่ม



ภาพที่ 45 แผนภาพกล่องเวลาที่ใช้ในการเดินทางของกลุ่มเส้นทางที่นิยม

- เวลาเฉลี่ยที่ใช้ในการเดินทางของแต่ละกลุ่มในชั่วโมงต่าง ๆ (TravelTime graph)  
เวลาเฉลี่ยที่ใช้ในการเดินทางของแต่ละกลุ่มในชั่วโมงต่าง ๆ เป็นกราฟที่ทำให้เห็นการเปลี่ยนแปลงของเวลาในการเดินทางของแต่ละกลุ่มเส้นทาง ทำให้สามารถเปรียบเทียบเวลาที่ใช้ในการเดินทางระหว่างกลุ่มได้ ซึ่งเป็นสาเหตุหนึ่งที่คนขับแท็กซี่ใช้ในการเลือกเส้นทาง



ภาพที่ 46 เวลาเฉลี่ยที่ใช้ในการเดินทางของแต่ละกลุ่มในชั่วโมงต่าง ๆ

#### 4.4. วิเคราะห์จากการหาตำแหน่งบ้านและที่ทำงาน

ข้อมูลซีดีอาร์ที่ผ่านกระบวนการในการหาบ้านและที่ทำงานแล้วถือเป็นข้อมูลหนึ่งที่สามารถนำมาใช้ประโยชน์ในการศึกษาการกระจายตัวของประชากรในพื้นที่ศึกษารวมถึงศึกษาพฤติกรรมของผู้ใช้งานในมุมมองของระยะทางและช่วงเวลาที่อยู่ในสถานที่ต่าง ๆ

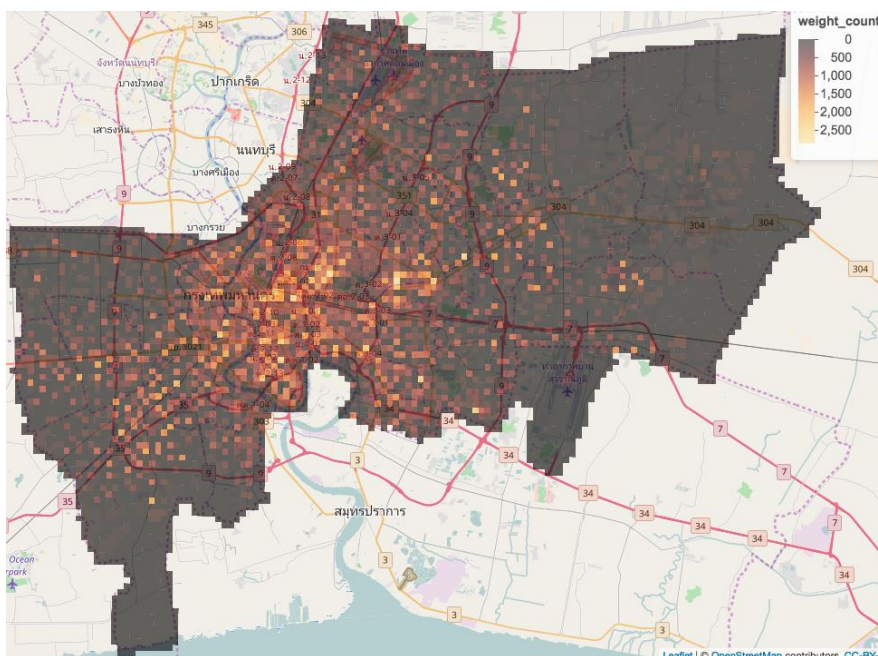
##### 4.4.1. การกระจายตัวของตำแหน่งบ้านและที่ทำงาน

จากข้อมูลบ้านและที่ทำงานของผู้ใช้งานแต่ละคนสามารถนำมาพิจารณาในเชิงพื้นที่ โดยทำการนับจำนวนคนที่อยู่ในแต่ละกริดทำให้เห็นถึงความหนาแน่นและการกระจายตัวของประชากรในพื้นที่ที่ศึกษา ซึ่งเป็นข้อมูลที่สำคัญที่สามารถนำไปใช้ในการจัดสรรสิ่งอำนวยความสะดวกให้เพียงพอต่อความต้องการในพื้นที่อย่างแท้จริงได้

การกระจายตัวของบ้านในพื้นที่กรุงเทพมหานคร (ภาพที่ 47) พบว่าจาก 1,408,188 คน ที่สามารถหาตำแหน่งบ้านได้ มีการกระจายตัวเป็นวงกว้างจากกรุงเทพฯ ขึ้นในมายังกรุงเทพฯ ขึ้นนอก โดยด้านตะวันตกของกรุงเทพฯจะมีบ้านกระจายอยู่มากกว่าด้านตะวันออก ทางด้านตะวันออกจะมีกลุ่มพื้นที่ที่มีความหนาแน่นสูงซึ่งเป็นพื้นที่ใกล้สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังและนิคมอุตสาหกรรมลาดกระบัง พื้นที่ที่มีความหนาแน่นต่อพื้นที่สูงคือบริเวณกรุงเทพฯ ขึ้นในโดยพื้นที่ที่มีความหนาแน่นสูงสุดมี

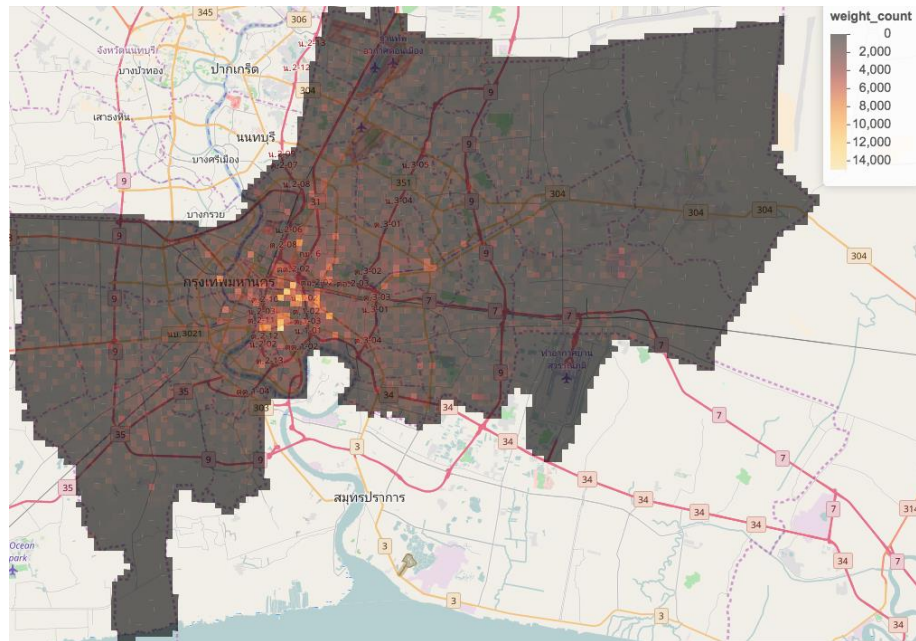


ผู้ใช้งาน 11,396 คนต่อตารางกิโลเมตร ซึ่งเมื่อเปรียบเทียบกับประชากรกรุงเทพฯ ทั้งหมด พื้นที่ที่มีความหนาแน่นสูงสุดมีประชากร 46,100 คนต่อตารางกิโลเมตร

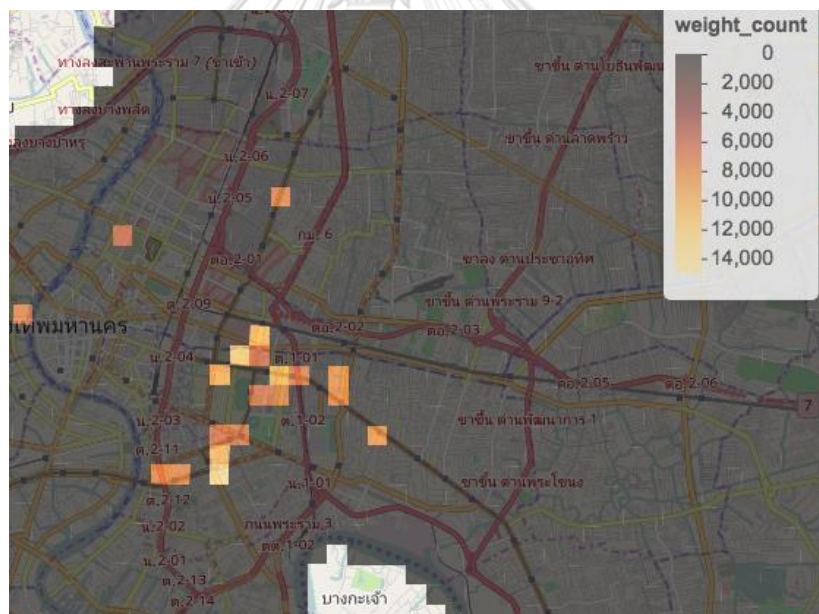


ภาพที่ 47 การกระจายตัวของบ้านในพื้นที่กรุงเทพมหานคร

การกระจายตัวของที่ทำงานในพื้นที่กรุงเทพมหานคร (ภาพที่ 48) จาก 3,320,135 คน ที่สามารถหาที่ทำงานได้ พบว่ามีความหนาแน่นสูงกระจุกตัวอยู่ในเขตกรุงเทพฯ ชั้นใน พื้นที่ที่มีความหนาแน่นสูงสุดมีผู้ใช้งาน 59,766 คนต่อตารางกิโลเมตร คิดความหนาแน่นของประชากร 140,912.44 คนต่อตารางกิโลเมตร เมื่อพิจารณา 20 พื้นที่ที่มีความหนาแน่นสูงสุด (ภาพที่ 49) พบว่ามีความหนาแน่นผู้ใช้งานเฉลี่ย 34,868 คนต่อตารางกิโลเมตร คิดความหนาแน่นของประชากร 82,209.52 คนต่อตารางกิโลเมตร ซึ่งเกือบทั้งหมดเป็นพื้นที่ย่านธุรกิจ เช่น สีลม, ชิดลม, ชองนนทบุรี, นานา, อโศก และพร้อมพงษ์ โดย 18 พื้นที่ที่มีความหนาแน่นสูงสุดอยู่ในเส้นทางรถไฟฟ้า



ภาพที่ 48 การกระจายตัวของสถานที่ทำงานในพื้นที่กรุงเทพมหานคร

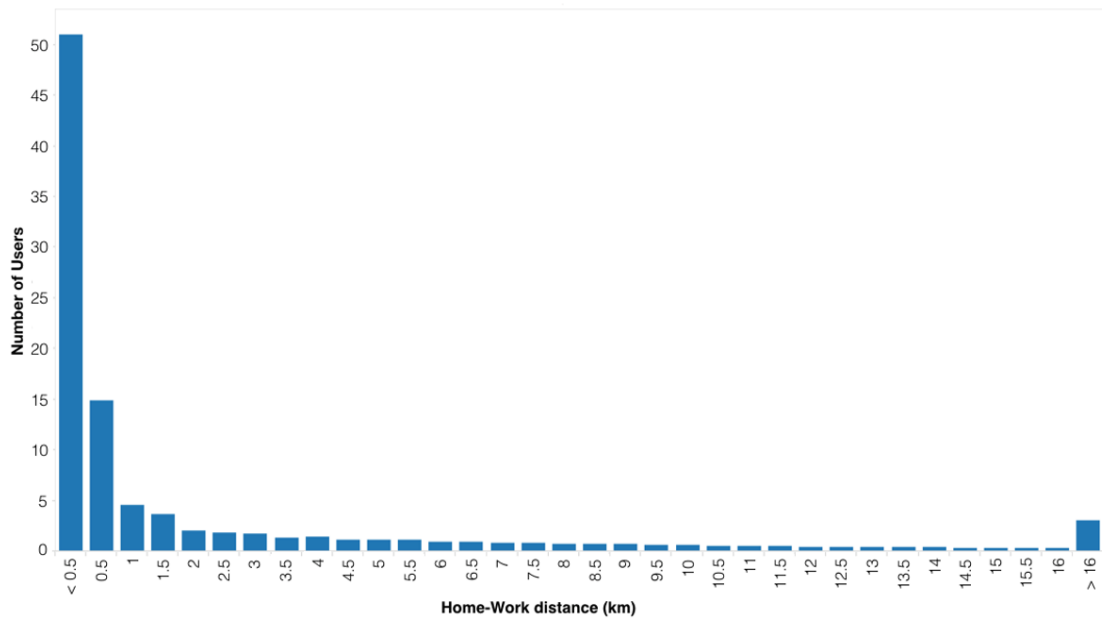


ภาพที่ 49 การกระจายตัวที่มากที่สุด 20 อันดับ ของสถานที่ทำงานในพื้นที่กรุงเทพมหานคร

#### 4.4.2. ระยะการกระจัดจากบ้านและที่ทำงาน

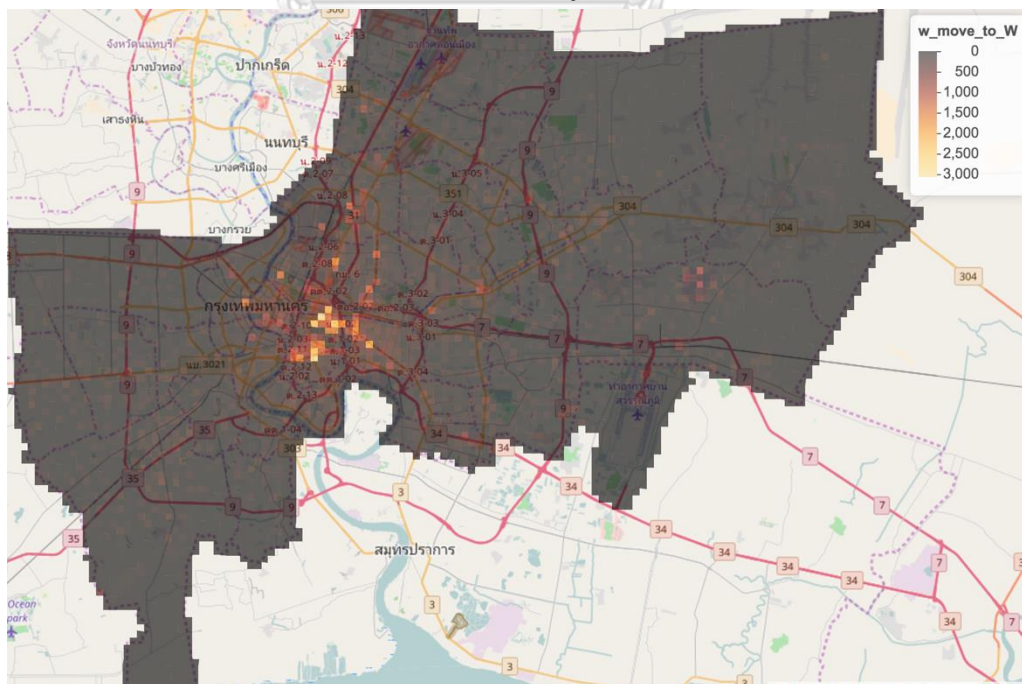
เมื่อทราบที่ทำงานและบ้านของคนที่อยู่ในพื้นที่ศึกษาทำให้สามารถพิจารณาในมุมมองของความต้องการในการเดินทางของคนในพื้นที่ได้ โดยความต้องการในการเดินทางพื้นฐานคือความต้องการในการเดินทางระหว่างบ้านและที่ทำงาน เมื่อพิจารณาระยะห่างระหว่างบ้านกับที่ทำงาน(ภาพที่ 50) พบว่าคนส่วนใหญ่เลือกที่จะอยู่อาศัยในบริเวณใกล้เคียงกับที่ทำงาน โดย 80 % ของผู้ใช้งานทั้งหมดมีบ้านอยู่ในรัศมี 3.5 กิโลเมตรจากบ้าน



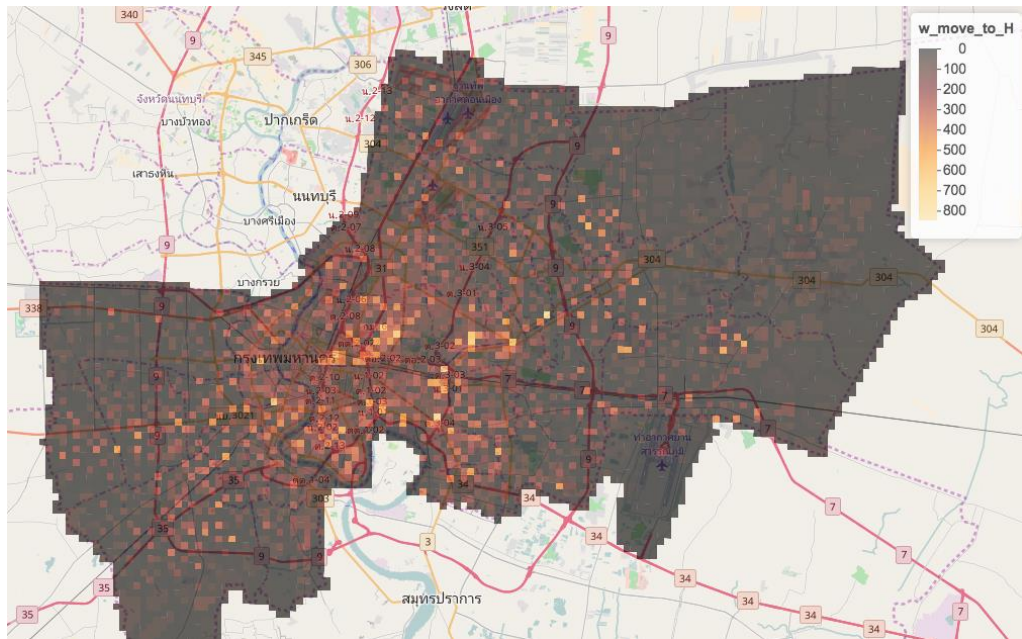


ภาพที่ 50 ระยะห่างระหว่างบ้านกับที่ทำงาน

พบว่า มี 22.12% ที่มีระยะห่างระหว่างบ้านกับสถานที่ทำงานมากกว่า 3 กิโลเมตร ซึ่งเป็นกลุ่มที่เป็นความต้องการในการเดินทางระหว่างบ้านและที่ทำงานในกรุงเทพฯ โดยพิจารณาเฉพาะคนกลุ่มนี้พบว่า ตำแหน่งที่ทำงาน (ภาพที่ 51) ที่ผู้ใช้งานเดินทางมาส่วนใหญ่อยู่ในกรุงเทพฯ ชั้นในสูงมากในบริเวณ สยามสแควร์, ชิดลม, อโศก, พร้อมพงษ์ และ ชองนนทบุรี โดยเดินทางมาจากบ้านที่กระจายอยู่ทั่วกรุงเทพฯ (ภาพที่ 52)



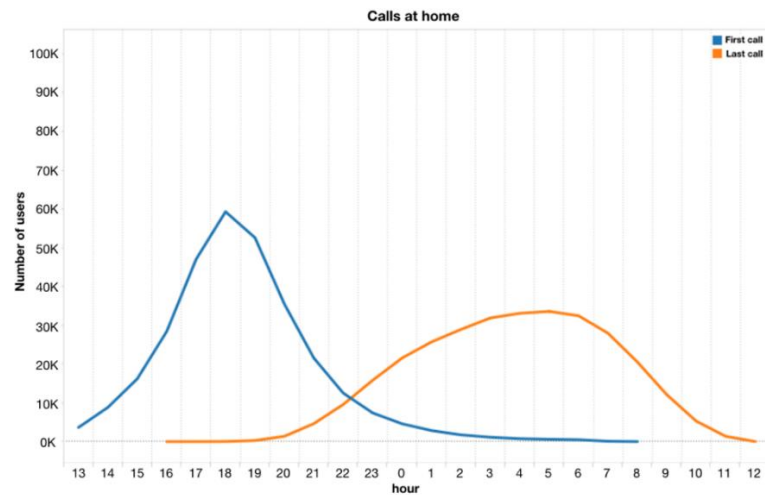
ภาพที่ 51 ตำแหน่งของที่ทำงานของผู้ใช้ที่เดินทางมากกว่า 3 กิโลเมตรจากบ้าน



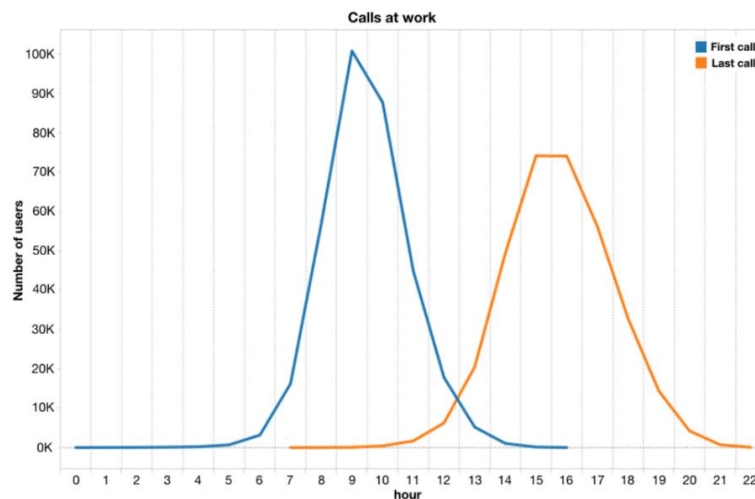
ภาพที่ 52 ตำแหน่งของบ้านของผู้ใช้ที่เดินทางมากกว่า 3 กิโลเมตรเพื่อไปทำงาน

#### 4.4.3. เวลาที่อยู่บ้านและที่ทำงาน

พิจารณาการใช้งานจากตำแหน่งที่เป็นบ้าน (ภาพที่ 53 a) และที่ทำงาน (ภาพที่ 53 b) โดยเส้นสีน้ำเงินคือเวลาเฉลี่ยที่โทรครั้งแรกจากตำแหน่งบ้านหรือที่ทำงาน เส้นสีส้มคือเวลาเฉลี่ยการโทรครั้งสุดท้ายจากที่ตำแหน่งบ้านหรือที่ทำงานของแต่วัน โดยการใช้งานจากที่ทำงานพบว่าผู้ใช้งานเริ่มมีการใช้งานแรกสูงในช่วง 07.00-12.00 น. โดยสูงสุดในชั่วโมงที่ 9.00 น. มีการโทรครั้งสุดท้ายในช่วง 13.00-19.00 น. โดยสูงสุดในชั่วโมงที่ 13 และ 14 ส่วนการใช้งานในบริเวณที่เป็นบ้านพบว่ามีการใช้งานครั้งแรกสูงในช่วง 16.00 – 20.00 น. โดยมีจุดสูงสุดที่ต่ำกว่าและกระจายตัวที่กว้างกว่าการใช้งานแรกในบริเวณที่ทำงานแสดงให้เห็นถึงมีคนกลุ่มหนึ่งที่ไม่ได้กลับบ้านทันทีหลังจากเลิกงาน ส่วนการโทรครั้งสุดท้ายในบริเวณที่เป็นบ้านมีการกระจายที่กว้างมาก



(a) การใช้งานจากตำแหน่งที่เป็นบ้าน

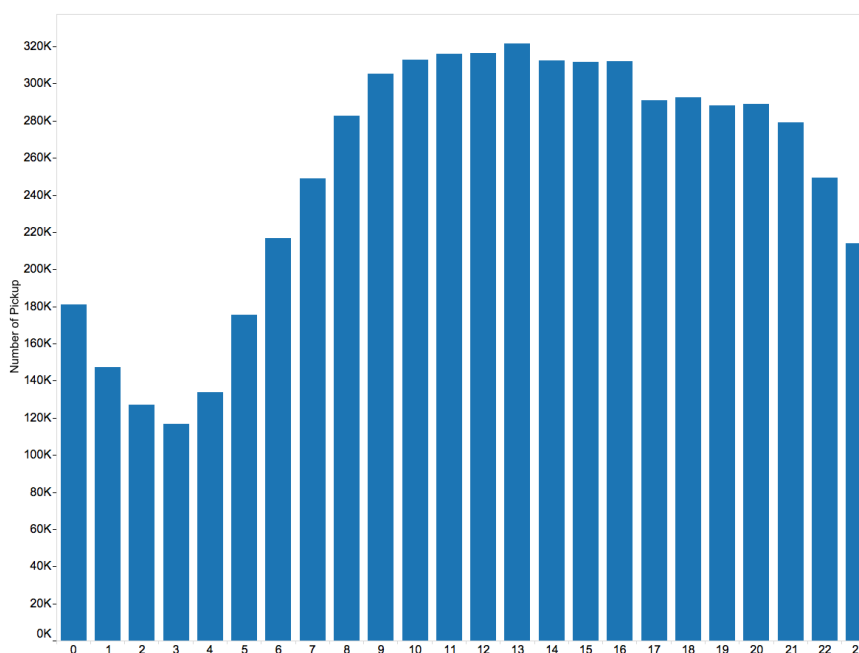


(b) การใช้งานจากตำแหน่งที่เป็นที่ทำงาน

ภาพที่ 53 การใช้งานจากตำแหน่งที่เป็นบ้านและที่ทำงาน

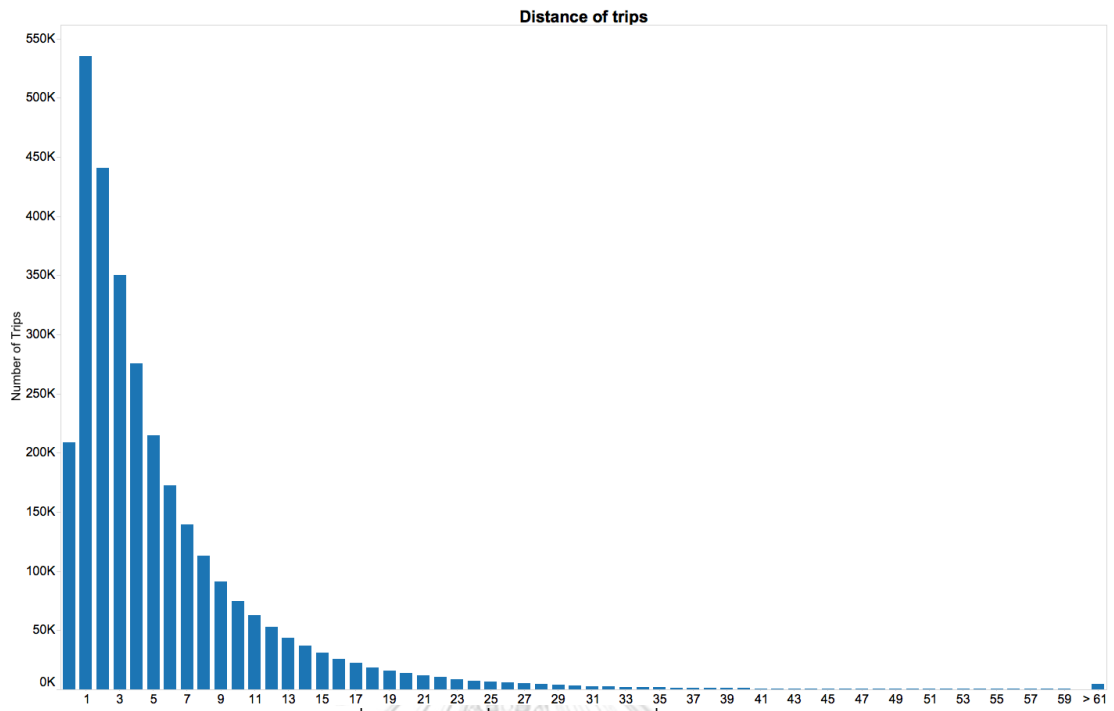
#### 4.5. วิเคราะห์จากผลการแยกเส้นทางที่ใช้ส่งผู้โดยสาร

ข้อมูลแท็กซี่ที่ผ่านการกรองการเดินทางส่งผู้โดยสารที่มีความผิดปกติออกแล้วพบว่าเหลือการเดินทางที่ใช้ส่งผู้โดยสาร 3,041,793 เที่ยว โดยความถี่ในการรับผู้โดยสารในเวลาต่าง ๆ (ภาพที่ 54) พบว่าในช่วงที่ 3 เป็นช่วงเวลาที่มีการใช้งานต่ำที่สุด และสูงขึ้นอย่างรวดเร็วในช่วง 05.00 – 08.59 น. ช่วงเวลา 09.00 - 16.59 น. เป็นช่วงที่มีการใช้งานมากที่สุดโดยมีการใช้งานมากที่สุดในชั่วโมงที่ 13 หลังจาก 17.00 น. มีการใช้งานลดลงเล็กน้อยและคงที่จนถึงชั่วโมงที่ 21 หลังจากนั้นการให้บริการลดลงจนถึงจุดต่ำสุด

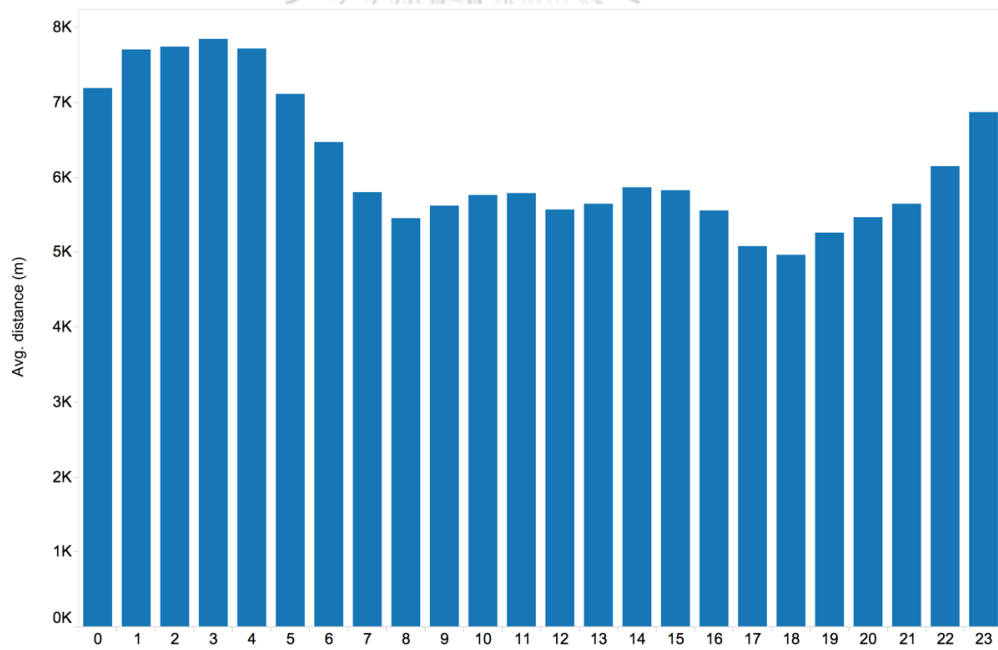


ภาพที่ 54 ความถี่ในการรับผู้โดยสารในเวลาต่าง ๆ

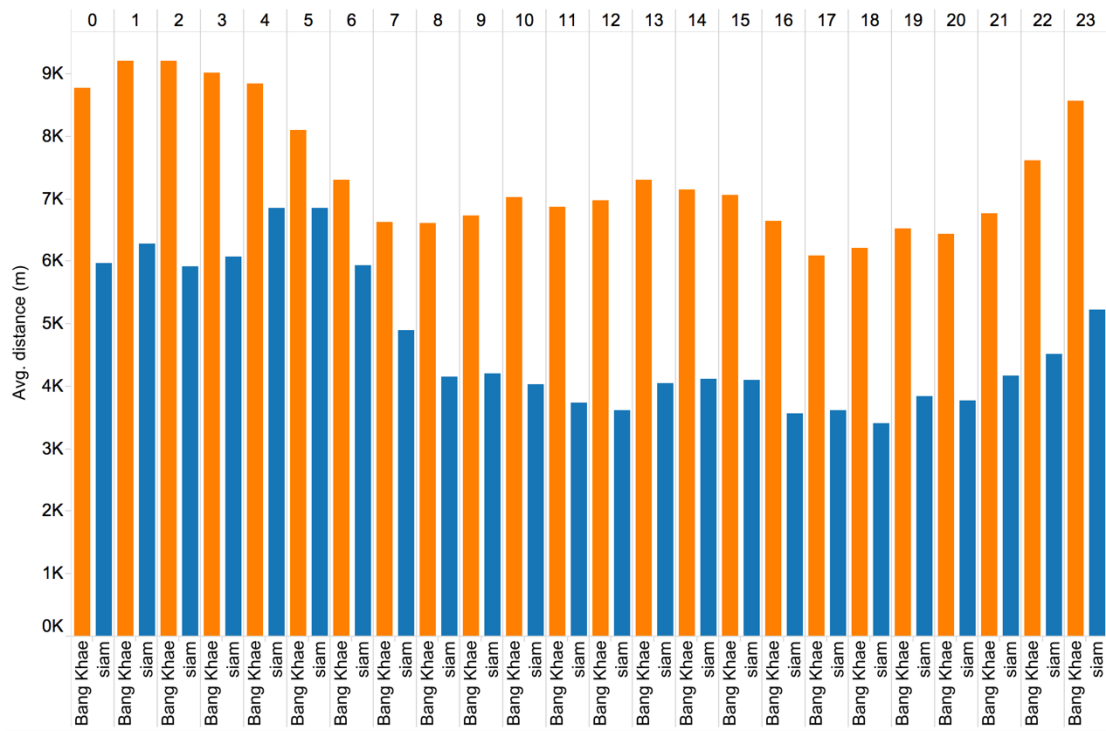
มุมมองของระยะทางที่ใช้ส่งผู้โดยสาร พบว่าความถี่ของระยะทางที่ใช้ส่งผู้โดยสาร ภาพที่ 55 พบว่าระยะทางที่ใช้ส่งผู้โดยสารมากอยู่ในช่วง 1-5 กิโลเมตร ก่อนที่จะลดลงอย่างต่อเนื่อง โดย 80 % ของข้อมูลใช้ระยะทางในการเดินทาง 8.855 กิโลเมตร เมื่อพิจารณาระยะทางที่ใช้ส่งผู้โดยสารพบว่าในช่วงเวลาต่าง ๆ ระยะทางที่รถแท็กซี่ใช้ส่งผู้โดยสารไม่เท่ากันซึ่งแสดงในภาพที่ 56 แสดงระยะทางเฉลี่ยในชั่วโมงต่าง ๆ ของเส้นทางที่ส่งผู้โดยสารภายในกรุงเทพฯ พบว่าในช่วงเวลา 23.00 – 05.00 น. เป็นช่วงที่มีระยะทางที่ใช้ส่งผู้โดยสารที่สูงโดยระยะทางเฉลี่ยที่สูงที่สุดอยู่ที่ ชั่วโมงที่ 3 ที่มีระยะทางเฉลี่ย 7.85 กิโลเมตร หลังจากนั้นลดลงมาจนถึงชั่วโมงที่ 8 และระยะทางเฉลี่ยอยู่ในกรอบ 4.7 – 5.7 กิโลเมตรจนถึงชั่วโมงที่ 21 และสูงขึ้นจนถึงจุดสูงสุด เมื่อทำการเปรียบเทียบความเปลี่ยนแปลงของระยะทางที่ใช้ส่งผู้โดยสารในชั่วโมงต่าง ๆ พบว่ามีความแตกต่างกันในแต่ละพื้นที่ ตัวอย่างแสดงในภาพที่ 57 กราฟสี่เหลี่ยมเป็นข้อมูลรถแท็กซี่ที่รับผู้โดยสารในเขตบางแค สีน้ำเงินเป็นข้อมูลรถแท็กซี่ที่ให้บริการบริเวณสยาม พบว่ารถแท็กซี่ที่ให้บริการในเขตบางแคมีระยะทางมากกว่ารถแท็กซี่ที่ให้บริการบริเวณสยาม และการเปลี่ยนแปลงของระยะทางเฉลี่ยในชั่วโมงที่มีระยะทางเฉลี่ยมากที่สุดกับน้อยที่สุดของรถแท็กซี่ที่ให้บริการบริเวณสยามสแควร์อยู่ที่ 50.32 % ซึ่งมากกว่ารถแท็กซี่ที่ให้บริการในเขตบางแคที่ 33.82 %



ภาพที่ 55 ความถี่ของระยะทางที่ใช้ส่งผู้โดยสาร



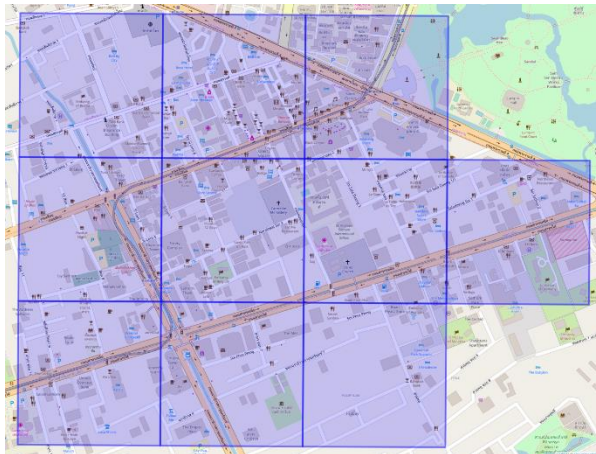
ภาพที่ 56 ระยะทางเฉลี่ยในชั่วโมงต่าง ๆ ของเส้นทางที่ส่งผู้โดยสารภายในกรุงเทพฯ



ภาพที่ 57 กราฟเปรียบเทียบระยะทางเฉลี่ยในชั่วโมงต่าง ๆ ของเส้นทางที่ส่งผู้โดยสารระหว่างในเมืองกับชานเมือง

#### 4.6. กรณีศึกษาที่ 1 วิเคราะห์ความต้องการและรูปแบบการเดินทางระหว่างพื้นที่

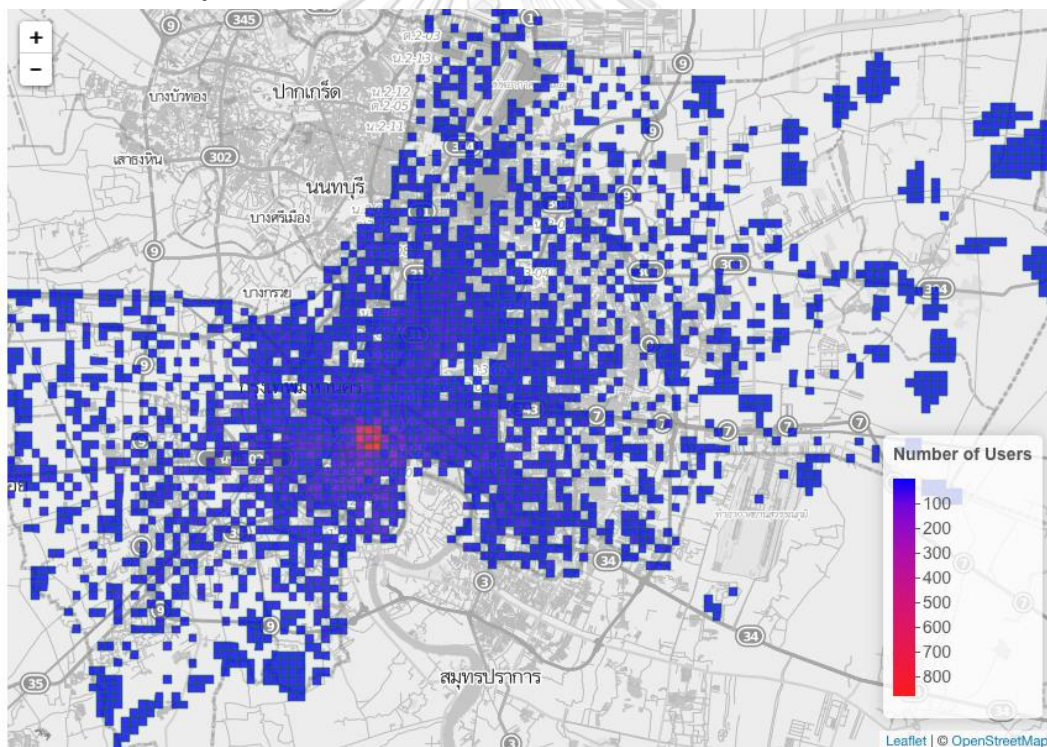
ในกรณีศึกษานี้เป็นการศึกษาจากคนที่มีการทำงานอยู่ในบริเวณถนนสีลมและชองนนทรีซึ่งแสดงภายในกริดสีน้ำเงินตามภาพที่ 58 เป็นพื้นที่ทั้งหมด 2.5 ตารางกิโลเมตร โดยที่เลือกพื้นที่นี้เป็นกรณีศึกษาเนื่องจากเป็นพื้นที่ที่ตั้งของสำนักงานขนาดใหญ่และเป็นหนึ่งในพื้นที่ที่เป็นสถานที่ที่คนเดินทางมาทำงานมาก



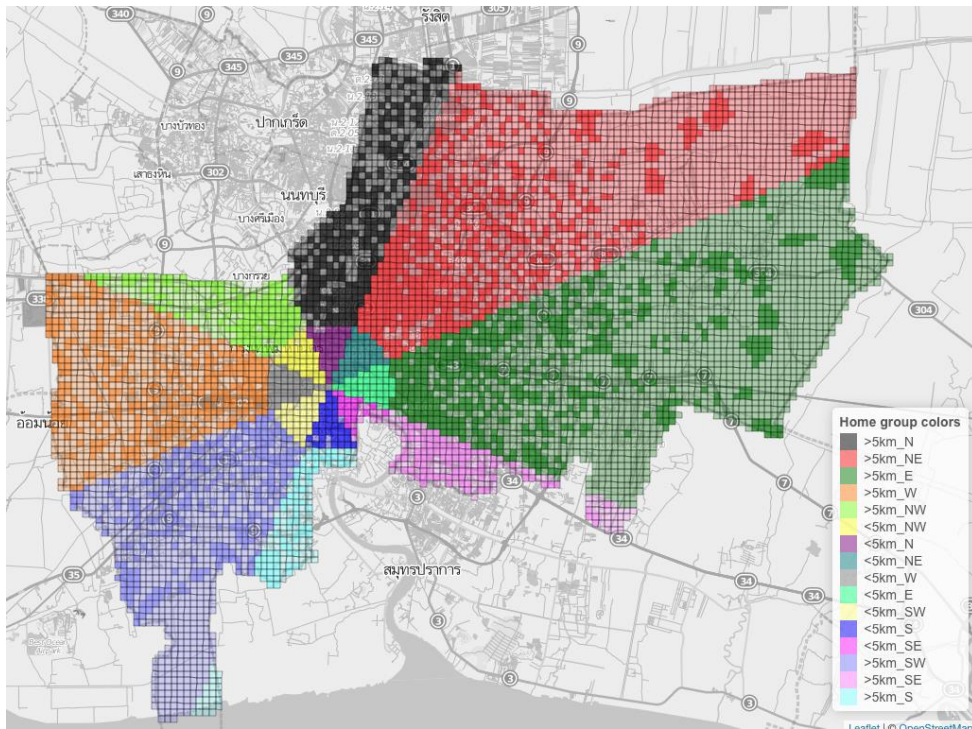
ภาพที่ 58 พื้นที่ที่ใช้ในการศึกษาที่ 1



ระบบสามารถศึกษาในมุมมองจากตำแหน่งบ้านและที่ทำงานพบว่าคนเดินทางมาทำงานในบริเวณนี้จากทั่วพื้นที่กรุงเทพมหานครตามภาพที่ 59 โดยมีระดับสีจากสีน้ำเงินจนถึงสีแดงแสดงปริมาณจำนวนคนที่มีบ้านในตำแหน่งนั้น ซึ่งพบว่าตำแหน่งบ้านที่มีความหนาแน่นสูงจะอยู่ในบริเวณใกล้เคียงกับพื้นที่ศึกษาโดยมีบ้าน 36.11 % ของคนที่มีทำงานในพื้นที่ศึกษาอยู่ในรัศมี 5 กิโลเมตร แสดงให้เห็นถึงคนส่วนใหญ่มักจะหาที่พักใกล้กับสถานที่ทำงาน นอกจากนี้ระบบสามารถศึกษาทิศทางระหว่างตำแหน่งบ้านและตำแหน่งสถานที่ทำงาน โดยแบ่งระยะห่างจากตำแหน่งบ้านเพื่อทำการแยกบ้านที่อยู่ใกล้สถานที่ทำงานออกจากบ้านที่ไกลจากสถานที่ทำงานและแบ่งทิศทางระหว่างตำแหน่งบ้านและตำแหน่งสถานที่ทำงานออกเป็น 8 ทิศได้ทั้งหมด 16 กลุ่มซึ่งแสดงอยู่ในภาพที่ 60 โดยเฉดสีแสดงถึงกลุ่มที่ถูกแบ่งและสีที่เข้มกว่าในเฉดสีเดียวกันแสดงถึงกริดที่มีคนเดินทางมาพื้นที่ศึกษา ตารางที่ 6 แสดงปริมาณคนที่มีบ้านอยู่ในอยู่ในกลุ่มต่าง ๆ โดยในรัศมี 5 กิโลเมตรจากที่ทำงานคนมีบ้านอยู่ทางทิศใต้และตะวันตกเฉียงใต้มากที่สุด บ้านที่มีระยะมากกว่า 5 กิโลเมตรจากที่ทำงานส่วนใหญ่อยู่ในทิศตะวันออกเฉียงและทิศเหนือมากกว่าทิศอื่น ๆ



ภาพที่ 59 กริดที่เป็นตำแหน่งบ้านมาทำงานในพื้นที่ศึกษา



ภาพที่ 60 กลุ่มที่แบ่งพื้นที่โดยอาศัยระยะห่างและทิศทางจากพื้นที่ศึกษา

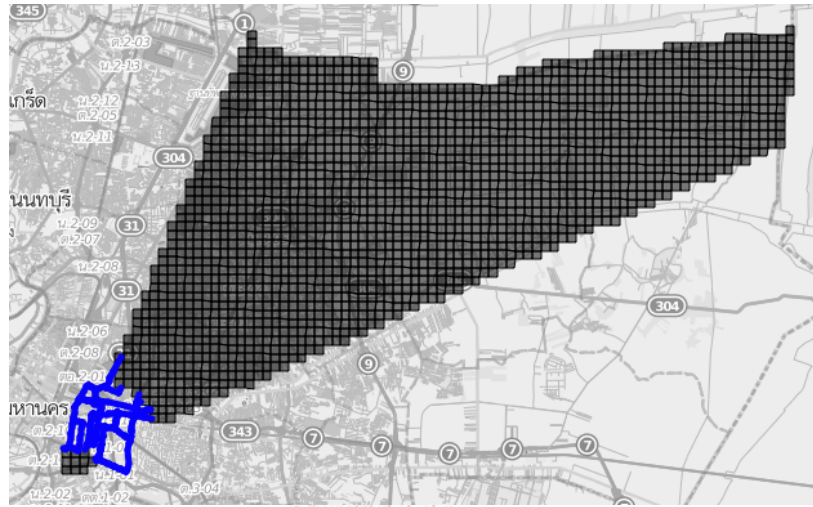
ตารางที่ 6 ทิศทางและจำนวนคนที่เดินทางมาในพื้นที่ที่ศึกษา

ระยะการกระจัดจาก ศูนย์กลาง	ทิศจากจุดศูนย์กลาง	จำนวนคน
< 5 km	N	340.0000
	NE	397.0000
	E	865.1667
	SE	542.0000
	S	1148.0000
	SW	1006.5000
	W	714.0000
	NW	417.5000

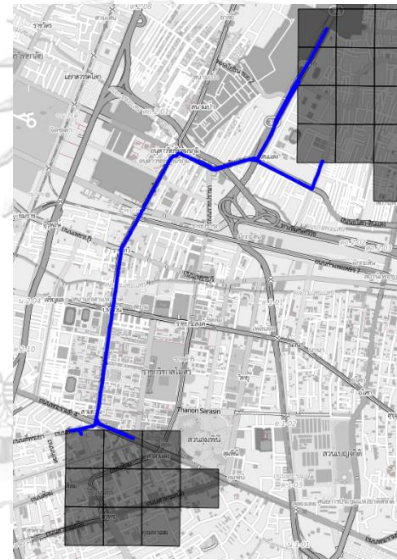


ระยะการกระจัดจาก ศูนย์กลาง	ทิศจากจุดศูนย์กลาง	จำนวนคน
> 5 km	N	1159.0000
	NE	2173.1527
	E	1895.3473
	SE	515.5000
	S	424.7143
	SW	1100.6190
	W	1779.5556
	NW	558.4444

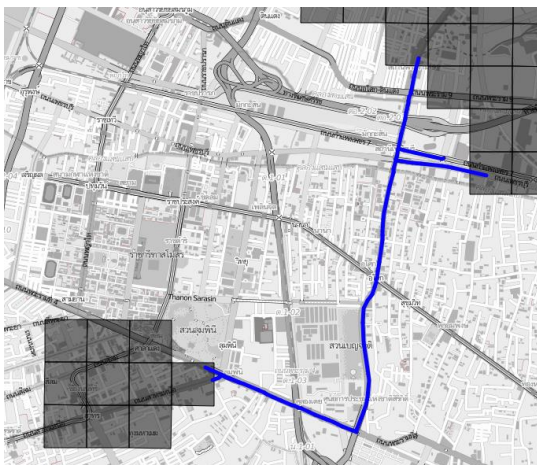
ระบบนั้นนอกจากจะศึกษาตำแหน่งและปริมาณระหว่างบ้านและสถานที่ทำงานแล้วยังสามารถศึกษาเส้นทางบนถนนที่นิยมใช้ในการเดินทางระหว่างพื้นที่ โดยจะสาธิตเส้นทางที่นิยมเดินทางจากพื้นที่ที่บ้านระยะห่างจากที่ทำงานมากกว่า 5 กิโลเมตรและมีอยู่ทางทิศตะวันออกเฉียงเหนือ ซึ่งเป็นกลุ่มที่มีคนเดินทางมาในพื้นที่ศึกษามากแสดงเส้นทางทั้งหมดที่ถูกใช้งานในการเดินทางระหว่างพื้นที่แสดงในภาพที่ 61 เส้นทางที่นิยมใช้ในการเดินทาง 3 อันดับแรกแสดงในภาพที่ 62 โดยเส้นทางที่ถูกใช้มากที่สุดคิดเป็น 6.72 % ของการเดินทางระหว่างพื้นที่ทั้งหมด โดยใช้ถนนดินแดง และถนนพญาไท ระยะทาง 16.09 กิโลเมตร, เส้นทางที่ถูกใช้เป็นอันดับสองคิดเป็น 4.27 % ของการเดินทางระหว่างพื้นที่ทั้งหมด ใช้ถนนรัชดาภิเษก และถนนพระราม 4 ระยะทาง 18.17 กิโลเมตร และเส้นทางที่ถูกเลือกใช้งานเป็นอันดับ 3 คิดเป็น 2.90 % ของการเดินทางระหว่างพื้นที่ทั้งหมด โดยใช้ถนนรัชดาภิเษก, ถนนสุขุมวิท และถนนวิฑู ระยะทาง 18.71 กิโลเมตร



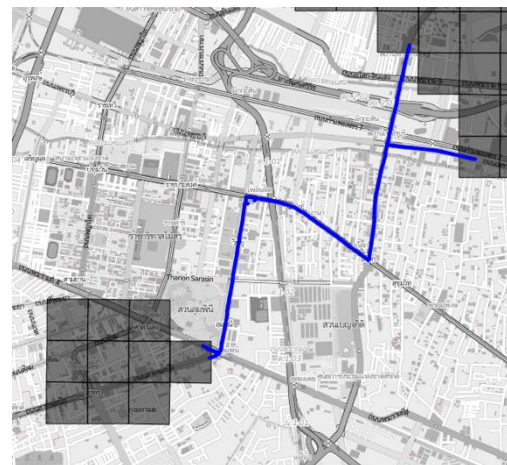
ภาพที่ 61 ถนนที่นิยมใช้ในการเดินทางโดยเส้นทางที่นิยม 10 ลำดับแรก



(a) เส้นทางที่นิยมใช้งานการเดินทางเป็นลำดับ 1



(b) เส้นทางที่นิยมใช้งานการเดินทางเป็นลำดับ 2

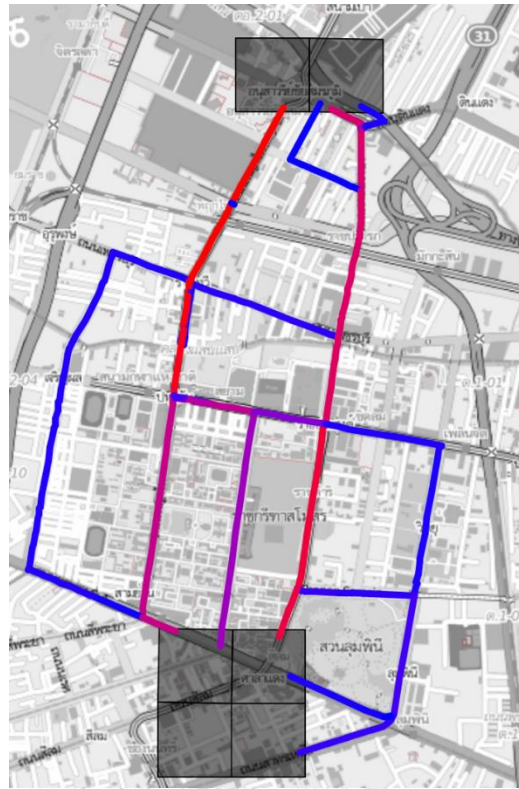


(c) เส้นทางที่นิยมใช้งานการเดินทางเป็นลำดับ 3

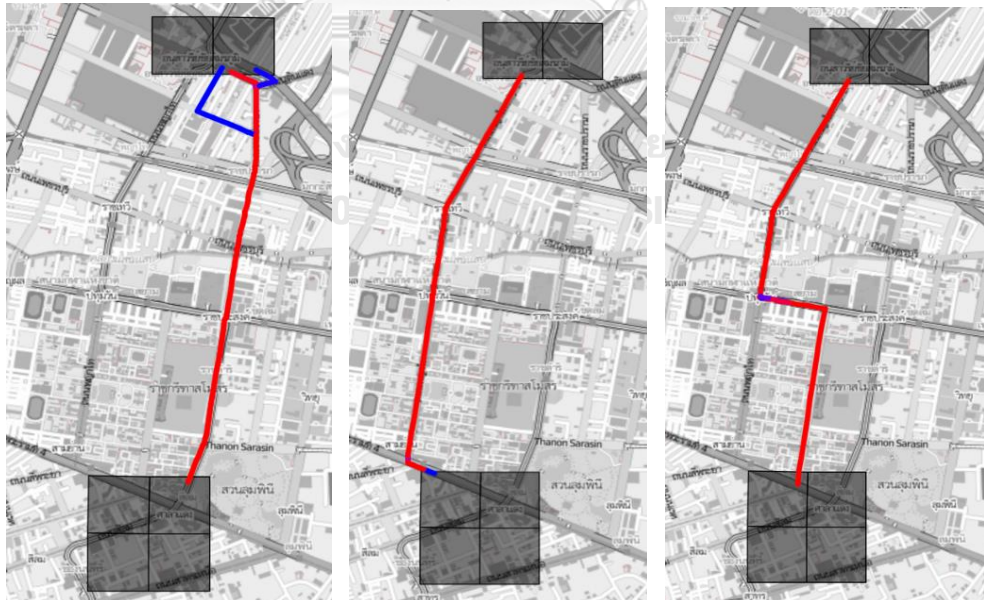
ภาพที่ 62 เส้นทางที่นิยมใช้งานการเดินทางมากที่สุด 3 ลำดับ

#### 4.7. กรณีศึกษาที่ 2 ศึกษาเส้นทางที่ถูกเลือกในการเดินทางระหว่างพื้นที่

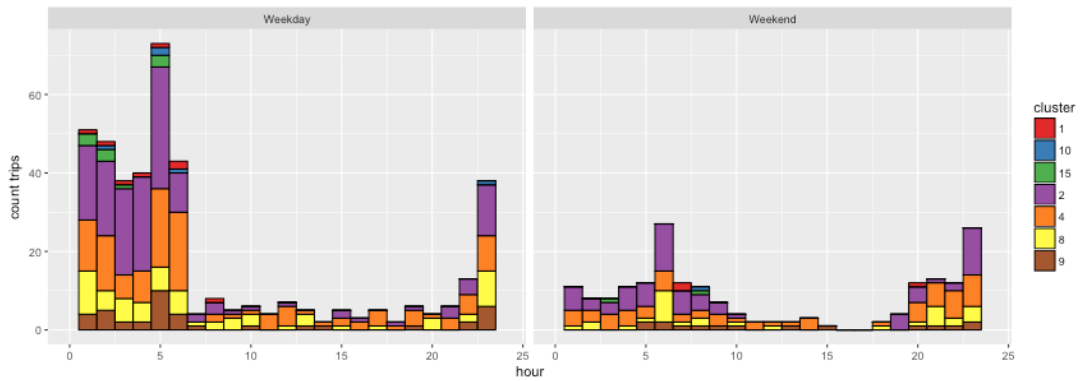
ในกรณีศึกษานี้เป็นการศึกษาเส้นทางที่แท็กซี่เลือกเดินทางจากบริเวณถนนสีลมไปบริเวณอนุสาวรีย์ชัยสมรภูมิระยะทาง 4.28 กิโลเมตร สาเหตุที่เลือกการเดินทางระหว่าง 2 พื้นที่นี้ เนื่องจากเป็นพื้นที่ที่มีความต้องการเดินทางผ่านเป็นจำนวนมาก พบว่ามี 8 เส้นทางที่ถูกเลือกใช้สูงกว่า 1 % ของจำนวนเส้นทางทั้งหมด ภาพที่ 63 แสดงถนนที่ใช้จาก 8 เส้นทางข้างต้น โดยระดับสีจากสีน้ำเงินไปจนถึงสีแดงแสดงถึงปริมาณการเดินทางผ่านถนนเส้นนั้น โดยเส้นทางที่ถูกใช้มากที่สุด 3 อันดับแรกแสดงในภาพที่ 64 โดยเส้นทางที่ถูกใช้มากที่สุดคือเส้นทางที่ 2 คิดเป็น 29.75 % ของการเดินทางระหว่างพื้นที่ทั้งหมด โดยใช้ถนนราชดำเนิน, ถนนราชปรารภและถนนราชวิถี รวมระยะทาง 3.78 กิโลเมตร, เส้นทางที่ถูกใช้เป็นอันดับสองคือเส้นทางที่ 4 คิดเป็น 23.80 % ของการเดินทางระหว่างพื้นที่ทั้งหมด ใช้ถนนพระราม 4 และถนนพญาไท รวมระยะทาง 3.81 กิโลเมตร และเส้นทางที่ถูกเลือกใช้งานเป็นอันดับ 3 คือเส้นทางที่ 8 คิดเป็น 13.41 % ของการเดินทางระหว่างพื้นที่ทั้งหมด โดยใช้ถนนอังรีดูนังต์, ถนนพระราม 1 และถนนพญาไท รวมระยะทาง 4.28 กิโลเมตร ภาพที่ 65 แสดงเส้นทางที่ถูกเลือกในเวลาต่าง ๆ ของวันพบว่าช่วงเวลา 00.01 – 06.00 น. เส้นทางที่ 2 จะถูกเลือกใช้งานมากที่สุดก่อนที่จะถูกใช้น้อยลงในช่วงกลางวัน และกลับมาเพิ่มสูงหลังจาก 21.00 น. ตรงกันข้ามกับเส้นทางที่ 4 ที่ถูกใช้งานน้อยกว่าเส้นทางที่ 2 ในช่วง 00.01 – 05.00 น. ก่อนที่จะถูกใช้งานเป็นสัดส่วนมากขึ้นเมื่อเทียบกับเส้นทางอื่นในช่วงเวลากลางวันและค่ำ ภาพที่ 66 แสดงเวลาในการเดินทางเฉลี่ยในชั่วโมงต่าง ๆ โดยช่วงเวลา 23.00 - 10.00 น. เป็นช่วงเวลาที่มียุทธศาสตร์ในการเดินทางที่ใกล้เคียงกันก่อนที่จะใช้เวลาเพิ่มขึ้นจนสูงในช่วง 16.00 – 18.00 น.



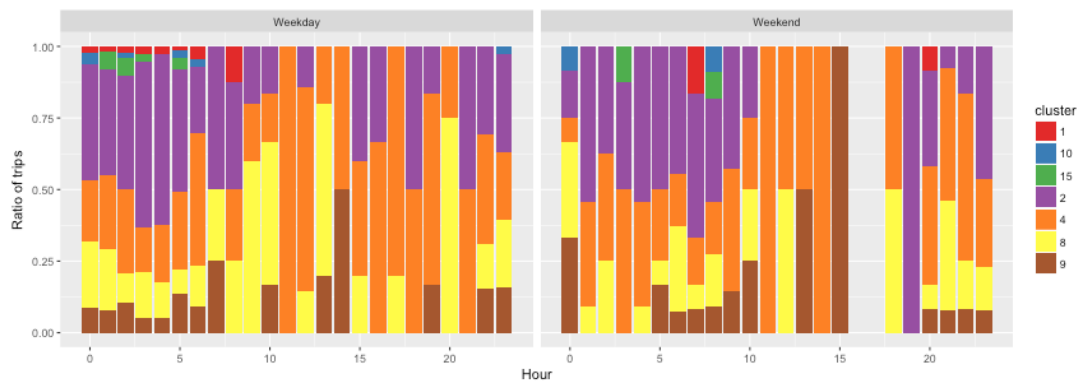
ภาพที่ 63 ถนน 8 เส้นทางที่ถูกใช้มากที่สุดเพื่อเดินทางจากบริเวณถนนสีลมไปบริเวณ  
อนุสาวรีย์ชัยสมรภูมิ



ภาพที่ 64 เส้นทางที่ถูกใช้มากที่สุด 3 ลำดับเพื่อเดินทางจากบริเวณถนนสีลมไปบริเวณ  
อนุสาวรีย์ชัยสมรภูมิ โดยเรียงจากซ้ายไปขวา

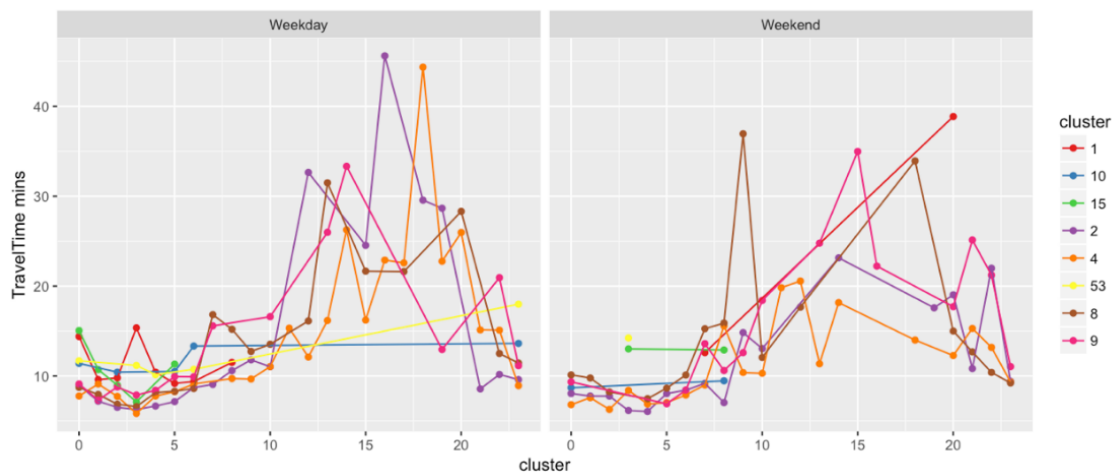


(a) จำนวนรถที่เลือกใช้เส้นทางเส้นทาง 8 อันดับแรกในช่วงเวลาต่าง ๆ



(b) สัดส่วนการเลือกใช้เส้นทางในช่วงเวลาต่าง ๆ

ภาพที่ 65 การเลือกใช้เส้นทาง 8 อันดับแรกเพื่อเดินทางจากบริเวณถนนสีลมไปบริเวณ  
อนุสาวรีย์ชัยสมรภูมิในช่วงเวลาต่าง ๆ

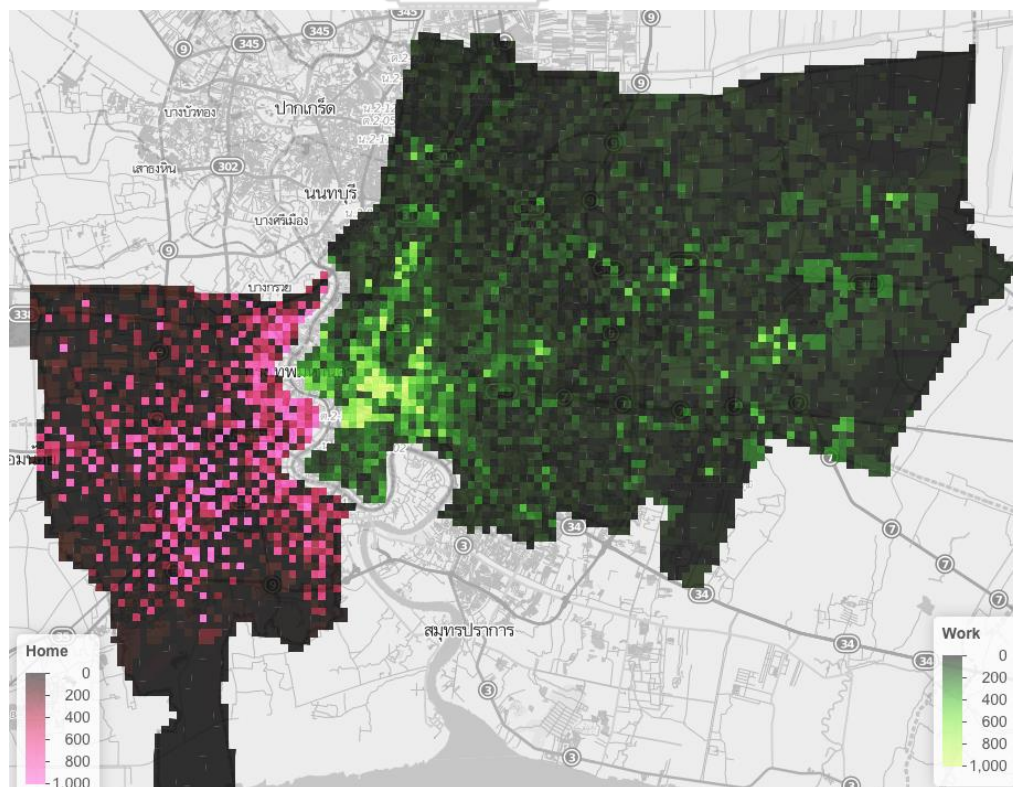


ภาพที่ 66 ความเร็วเฉลี่ยในช่วงเวลาต่าง ๆ ของเส้นทางที่ถูกใช้มากที่สุด 8 อันดับเพื่อเดินทางจาก  
บริเวณถนนสีลมไปบริเวณอนุสาวรีย์ชัยสมรภูมิ



#### 4.8. กรณีศึกษาที่ 3 ศึกษาการเดินทางจากฝั่งธนบุรีไปฝั่งพระนคร

กรุงเทพมหานครเป็นเมืองที่มีแม่น้ำเจ้าพระยาไหลผ่านแยกพื้นที่ออกเป็นสองส่วน การเดินทางข้ามระหว่าง 2 พื้นที่โดยรถจำเป็นต้องใช้สะพานข้ามแม่น้ำเจ้าพระยา ดังนั้นการศึกษาการเลือกใช้สะพานจึงมีความสำคัญในการออกแบบการขนส่งสาธารณะเพื่อแก้ปัญหาการติดที่เกิดจากความต้องการในการเดินทางระหว่างสองฝั่งแม่น้ำ พิจารณาจากคนที่มีบ้านอยู่ที่ฝั่งธนบุรีและมีสถานที่ทำงานอยู่ในฝั่งพระนครพบว่ามี 35,964 คน คิดเป็นประชากร 84,793.62 คน จากภาพที่ 67 แสดงความหนาแน่นของตำแหน่งบ้าน(เขตสีชมพู)และความหนาแน่นของตำแหน่งที่ทำงาน(เขตสีเขียว) พบว่ามีความหนาแน่นสูงบริเวณใกล้กับแม่น้ำเจ้าพระยา และสถานที่ทำงานมีความหนาแน่นสูงที่ถนนสารเหนือ, สีลม, ราชประสงค์, นานา และโอศก จากข้อมูลที่ทำกรแปลงจุดจีพีเอสเป็นถนนทำให้ทราบถึงสะพานที่ถูกเลือกใช้ในการข้ามแม่น้ำเจ้าพระยา โดยการเดินทางจากฝั่งธนบุรีไปฝั่งพระนครมีจำนวนการเดินทางทั้งหมด 80,847 การเดินทาง โดยสะพานที่ถูกเลือกใช้มากที่สุด 10 อันดับแรกแสดงในตารางที่ 7 และภาพที่ 68 แสดงจุดที่ผู้โดยสารขึ้นและลงรถของสะพานที่ถูกใช้มากที่สุด 3 อันดับแรกคือสะพานสะพานกรุงเทพ, สะพานสมเด็จพระปิ่นเกล้า และสะพานสมเด็จพระเจ้าตากสินมหาราช โดยแต่ละสะพานจะมีความหนาแน่นของตำแหน่งขึ้นและลงที่แตกต่างกัน



ภาพที่ 67 ความหนาแน่นของคนที่มีบ้านในฝั่งธนบุรีและมีที่ทำงานที่ฝั่งพระนคร

ตารางที่ 7 สะพานที่ถูกใช้ในการเดินทางจากฝั่งธนบุรีไปยังฝั่งพระนครมากที่สุด 10 อันดับแรก

สะพาน	จำนวนการเดินทางผ่านสะพาน
สะพานกรุงเทพ	12,838
สะพานสมเด็จพระปิ่นเกล้า	11,330
สะพานสมเด็จพระเจ้าตากสินมหาราช	11,235
สะพานสมเด็จพระปกเกล้า	7,364
สะพานพระราม 3	7,115
สะพานพระราม 7	6,869
สะพานภูมิพล 1	6,201
สะพานพระพุทธยอดฟ้า	5,448
สะพานพระราม 8	5,347
สะพานกรุงธน	5,126

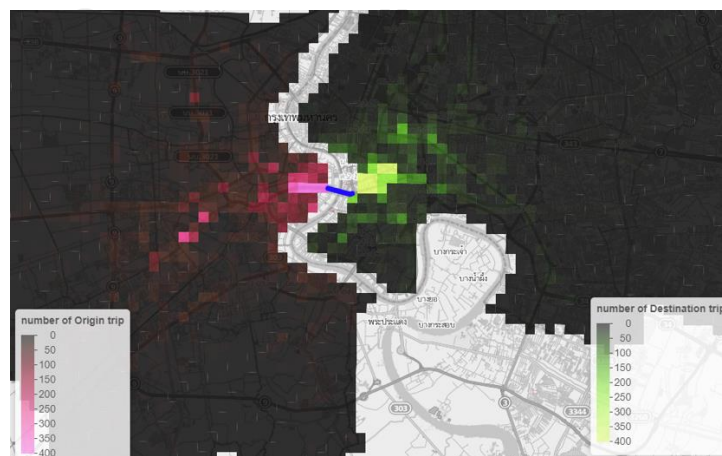
การเลือกใช้สะพานของรถแท็กซี่พิจารณาจากการเปรียบเทียบระยะห่างระหว่างจุดที่ผู้โดยสารขึ้นรถกับสะพานต่าง ๆ และจากสะพานต่าง ๆ ไปยังจุดหมายปลายทาง พบว่ามี 69.90 % เลือกใช้สะพานที่ทำให้ระยะทางรวมทั้งการเดินทางสั้นที่สุด, 43.77 % เลือกใช้สะพานที่อยู่ใกล้กับจุดที่ผู้โดยสารขึ้นรถมากที่สุด และมีเพียง 3.06 % ที่เลือกใช้สะพานที่อยู่ใกล้กับจุดที่ผู้โดยสารลงจากรถแท็กซี่



(a) สะพานกรุงเทพ



(b) สะพานสมเด็จพระปิ่นเกล้า



(c) สะพานสมเด็จพระเจ้าตากสินมหาราช

ภาพที่ 68 ความหนาแน่นของจุดเริ่มต้นและปลายทางของการเดินทางผ่านสะพาน  
ที่มีการใช้งานมากที่สุด 3 อันดับ



## บทที่ 5

### บทสรุปและแนวทางในการพัฒนาต่อ

#### 5.1. บทสรุป

งานวิจัยนี้นำข้อมูลซีดีอาร์และจีพีเอสมาสร้างเครื่องมือสำหรับใช้ในการวิเคราะห์ความต้องการและเส้นทางบนถนนที่ใช้ในการเดินทางระหว่างพื้นที่ โดยอาศัยประชากรขนาดใหญ่และสามารถวิเคราะห์พฤติกรรมระยะยาวเป็นรายบุคคลจากข้อมูลซีดีอาร์มาใช้ในการหาตำแหน่งบ้านและที่ทำงาน อาศัยความแม่นยำของตำแหน่งและรอบการส่งข้อมูลที่คงที่ที่จากข้อมูลจีพีเอสมาใช้ในการหาเส้นทางบนถนนที่นิยม โดยในงานวิจัยนี้ผู้วิจัยได้พัฒนาวิธีการและโปรแกรมประยุกต์

ข้อมูลซีดีอาร์ทราบเพียงตำแหน่งของเสากระจายสัญญาณ ซึ่งพื้นที่ให้บริการของแต่ละเสาไม่เท่ากันขึ้นกับปริมาณการใช้งาน ดังนั้นผู้วิจัยพัฒนาวิธีการในการแปลงพื้นที่ให้บริการเสากระจายสัญญาณเป็นกริดโดยไม่ส่งผลกระทบต่อความแม่นยำในการหาตำแหน่งบ้านและที่ทำงานระหว่างการใช้งานในเมืองที่เสากระจายสัญญาณอยู่ใกล้กันกับขานเมืองที่เสาอยู่ห่างกัน

การหาปริมาณความต้องการในการเดินทางระหว่างบ้านและสถานที่ทำงาน ผู้วิจัยจำเป็นต้องออกแบบวิธีการตำแหน่งบ้านจากข้อมูลซีดีอาร์ที่มีการใช้งานน้อย โดยใช้การใช้งานครั้งสุดท้ายก่อนเข้านอนและการใช้งานครั้งแรกหลังตื่นนอนซึ่งเป็นการใช้งานที่มีความน่าจะเป็นมากที่สุดที่การใช้งานนั้นจะเป็นการใช้งานจากที่บ้าน โดยผู้วิจัยได้ทดสอบความถูกต้องของตำแหน่งบ้านกับข้อมูลลูกค้ารายเดือนที่บริษัททราบที่อยู่โดยเปรียบเทียบแวงที่ได้จากการหาตำแหน่งบ้านกับแวงที่บริษัทต้องส่งเอกสารให้กับลูกค้าพบว่ามีความถูกต้องอยู่ที่ 69.02 %

การหาเส้นทางบนถนนที่นิยมใช้ในการเดินทางระหว่างพื้นที่พบว่าปัญหาการส่งข้อมูลที่ 1 นาที ทำให้ข้อมูลในการใช้เส้นทางในบางช่วงหายไปโดยเฉพาะในบริเวณที่รถสามารถทำความเร็วหรือเป็นบริเวณที่ช่วงถนนสั้น ซึ่งผู้วิจัยพบว่ามีกริดกระโดดข้ามถนนหลายช่วงทำให้ในการจัดกลุ่มเส้นทางบนถนนที่นิยมจำเป็นต้องหาลำดับของช่วงถนนทั้งหมดที่เดินทางผ่าน ผู้วิจัยได้พัฒนาและปรับปรุงวิธีการในการแปลงข้อมูลพิกัดจีพีเอสเป็นพิกัดถนนให้ส่งผลลำดับถนนที่ผ่านออกมาด้วย

ผู้วิจัยได้พัฒนาโปรแกรมประยุกต์ในรูปแบบเว็บแอปพลิเคชันที่ผู้ใช้งานสามารถเลือกกลุ่มของกริดที่ต้องการศึกษาตำแหน่งของบ้านและที่ทำงานทำให้ทราบถึงความต้องการระหว่างบ้านและที่ทำงาน และสามารถเลือกกลุ่มของกริดเริ่มต้นและกลุ่มของกริดปลายทางเพื่อศึกษาเส้นทางที่นิยมใช้ในการเดินทาง

เมื่อพัฒนาระบบเสร็จสมบูรณ์แล้วผู้วิจัยได้นำข้อมูลภายในระบบมาวิเคราะห์พฤติกรรม

พบว่า

- ประชากรในกรุงเทพฯ ส่วนใหญ่เลือกที่จะอยู่อาศัยใกล้กับที่ทำงาน โดย 80 % ประชากรมีบ้านในรัศมี 3.5 กิโลเมตรจากที่ทำงาน
- การเดินทางด้วยรถแท็กซี่ส่วนใหญ่มักจะเดินทางในระยะสั้น ๆ มีการใช้งานที่สูงสุด ในช่วงในช่วง 1 – 2 กิโลเมตร โดย 80 % ของการเดินทางด้วยรถแท็กซี่ทั้งหมดมี ระยะทางน้อยกว่า 8.855 กิโลเมตร
- การเดินทางด้วยแท็กซี่ในเขตชานเมืองมีระยะทางที่สูงกว่าการเดินทางในเขตเมือง โดยในช่วงเวลากลางดึกการเดินทางด้วยรถแท็กซี่จะเดินทางไกลกว่าในช่วงเวลา กลางวัน โดยการเปลี่ยนแปลงของระยะทางระหว่างระยะทางเฉลี่ยในชั่วโมงที่ สูงสุดกับชั่วโมงเฉลี่ยระหว่างระยะทางที่ต่ำที่สุดในเขตเมืองสูงกว่าเขตชานเมือง
- ความต้องการในการเดินทางจากฝั่งธนบุรีมายังฝั่งพระนครสูงในพื้นที่ใกล้กับแม่น้ำ เข้าพระยาเพื่อไปทำงานฝั่งพระนครซึ่งไม่ไกลจากแม่น้ำเจ้าพระยามากนัก
- การเลือกสะพานที่ใช้งานการข้ามแม่น้ำเจ้าพระยาจากฝั่งธนบุรีไปยังฝั่งพระนคร พบว่า 69.90 % เลือกใช้สะพานที่ทำให้ระยะทางรวมทั้งการเดินทางสั้นที่สุด 43.77 % เลือกใช้สะพานที่อยู่ใกล้กับจุดที่ ผู้โดยสารขึ้นรถมากที่สุด และมีเพียง 3.06 % ที่เลือกใช้สะพานที่อยู่ใกล้กับจุดที่ผู้โดยสารลงจากรถแท็กซี่

ผลจากการวิเคราะห์ทำให้ทราบว่า การเดินทางส่วนใหญ่เป็นการเดินทางในระยะสั้น ๆ ไม่ ว่าจะเป็นจากตำแหน่งบ้านและที่ทำงานหรือความต้องการในการเดินทางข้ามแม่น้ำ นอกจากนี้คน กรุงเทพฯ เป็นคนที่ไม่ยอมเดินถึงแม้การเดินทางระยะทางสั้น ๆ คนกรุงเทพฯ เลือกที่จะเดินทาง ด้วยรถ ดังนั้นจึงเป็นสาเหตุหนึ่งที่คนกรุงเทพฯ ใช้รถส่วนตัวในการเดินทาง ดังนั้นถ้าออกแบบ ระบบขนส่งที่ทำให้คนกรุงเทพฯ เดินน้อยที่สุดและระยะเวลาในการรอไม่มาก เช่นมีเรือข้ามฟาก และรถรับผู้โดยสารรับส่งเป็นระยะทางสั้นๆ โดยเฉพาะในพื้นที่ชองนนทบุรีและสีลมที่มีคนจำนวน มากอยู่ฝั่งธนบุรีแต่มีที่ทำงานที่ฝั่งพระนคร

## 5.2. ข้อเสนอแนะและแนวทางในการพัฒนาต่อ

จากการทดลองการหาตำแหน่งบ้านและสถานที่ทำงานพบว่าระบบสามารถหาที่ทำงานของผู้ใช้งานกรุงเทพได้แต่ไม่สามารถหาบ้านของผู้ใช้งานได้ซึ่งสาเหตุหนึ่งคือทำการหาบ้านเพราะในพื้นที่กรุงเทพฯ เท่านั้น ซึ่งในความเป็นจริงพบว่ามีผู้ใช้งานจำนวนมากที่พักอยู่ในพื้นที่ปริมณฑล ดังนั้นการทราบความต้องการที่แท้จริงจำเป็นต้องขยายพื้นที่ในการศึกษาเป็นพื้นที่กรุงเทพฯ และปริมณฑล นอกจากนี้พบว่าเส้นทางที่นิยมใช้จากรถแท็กซี่ในการเดินทางบางพื้นที่ยังมีปริมาณน้อยทำให้ไม่สามารถสรุปสาเหตุของการเลือกเส้นทางหรือการเปลี่ยนเส้นทางในบางช่วงเวลาได้ ดังนั้นเพื่อที่จะสามารถวิเคราะห์ความแม่นยำจำเป็นต้องเพิ่มปริมาณรถแท็กซี่ให้มากขึ้น และเพิ่มข้อมูลอื่น ๆ เช่นจำนวนไฟจราจร เพื่อสามารถวิเคราะห์ถึงการเลือกเส้นทางได้มากขึ้น

จากการวิเคราะห์ข้อมูลทำให้เห็นถึงเป็นเป็นไปได้ในการศึกษาเช่น จากข้อมูลจีพีเอสพบว่า การเดินทางจำนวนมากเดินทางในระยะสั้นมาก เช่นการเดินทางจากสนามบินไปบริเวณใกล้เคียงที่เป็นพื้นที่ที่ไม่มีสิ่งปลูกสร้างใด ๆ ซึ่งผู้วิจัยได้ตั้งสมมติฐานว่าการเดินทางเส้นทางนั้นเป็นการเดินทางแบบเหมาจ่ายจึงทำการปิดมิเตอร์ในบริเวณดังกล่าว หรือพบเส้นทางที่ซับซ้อนเพื่อส่งผู้โดยสารซึ่งสามารถสังเกตได้จากกระยะทางที่ยาวผิดปกติจากการเดินทางระหว่างพื้นที่โดยรถคันอื่น ทำให้เห็นถึงความเป็นไปได้ที่จะทำระบบประเมินพฤติกรรมของรถแท็กซี่ในการรับส่งผู้โดยสาร

## รายการอ้างอิง

1. ข้อมูลสถิติกรุงเทพมหานคร. 2016 10 october 2016]; Available from: <http://www.bangkok.go.th/info/>.
2. ปริมาณจราจรและอัตราความเร็วในการเดินทาง, สำนักการจราจรและขนส่ง กรุงเทพมหานคร, Editor. 2015, สถิติจราจร ปี 2558: กรุงเทพมหานคร.
3. Zheng, Y., et al., *Urban computing with taxicabs*, in *Proceedings of the 13th international conference on Ubiquitous computing*. 2011, ACM: Beijing, China. p. 89-98.
4. Schneider, C.M., et al., *Unravelling daily human mobility motifs*. *J R Soc Interface*, 2013. 10(84): p. 20130246-20130246.
5. Castro, P.S., et al., *From taxi GPS traces to social and community dynamics: A survey*. *ACM Comput. Surv.*, 2013. 46(2): p. 1-34.
6. Zhang, D., et al., *Understanding Taxi Service Strategies From Taxi GPS Traces*. *IEEE Trans. Intelligent Transportation Systems*, 2015. 16(1): p. 123-135.
7. Mao, F., M. Ji, and T. Liu, *Mining spatiotemporal patterns of urban dwellers from taxi trajectory data*. *Frontiers of Earth Science*, 2016. 10(2): p. 205-221.
8. Zhang, D., et al., *Exploring human mobility with multi-source data at extremely large metropolitan scales*, in *MobiCom*, S.-J. Lee, A. Sabharwal, and P. Sinha, Editors. 2014, ACM. p. 201-212.
9. Liu, X., et al., *Revealing daily travel patterns and city structure with taxi trip data*. *CoRR*, 2013. abs/1310.6592.
10. Chen, C., et al., *B-Planner: Night bus route planning using large-scale taxi GPS traces*, in *PerCom*. 2013, IEEE Computer Society. p. 225-233.
11. Lin, M. and W.-J. Hsu, *Mining GPS data for mobility patterns: A survey*. *Pervasive and Mobile Computing*, 2014. 12: p. 1-16.
12. Çolak, S., et al., *Analyzing Cell Phone Location Data for Urban Travel*. *Transportation Research Record: Journal of the Transportation Research Board*, 2015. 2526: p. 126-135.

13. Wang, Z.D.a.L.L.a.S., *MobilePulse: Dynamic profiling of land use pattern and OD matrix estimation from 10 million individual cell phone records in Shanghai*, in *19th International Conference on Geoinformatics 2011*. 2011, IEEE: Shanghai, China. p. 1-6.
14. Pozdnoukhov, M.Y.a.M.S.a.S.F.a.J.F.P.a.A., *A Generative Model of Urban Activities from Cellular Data*. *IEEE Transactions on Intelligent Transportation Systems*, 2017. PP(99): p. 1-15.
15. Pozdnoukhov, D.Z.J.C.S.F.D.T.A., *Connected Population Synthesis for Urban Simulation*. 2017.
16. Isaacman, S., et al., *Identifying Important Places in People's Lives from Cellular Network Data*, in *Pervasive Computing*, K. Lyons, J. Hightower, and E. Huang, Editors. 2011, Springer Berlin / Heidelberg. p. 133-151.
17. Phithakkitnukoon, S., et al., *Activity-Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data*, in *HBU*, A.A. Salah, et al., Editors. 2010, Springer. p. 14-25.
18. Yang, P., et al., *Identifying Significant Places Using Multi-day Call Detail Records*, in *ICTAI*. 2014, IEEE Computer Society. p. 360-366.
19. Kang, J.-s.Y.a.S.-p., *The map matching algorithm of GPS data with relatively long polling time intervals*. *Journal of Eastern Asia Society for Transportation Studies*, 2005: p. 2561-2573.
20. Griffin, T., Y. Huang, and S. Seals, *Routing-based map matching for extracting routes from GPS trajectories*, in *Proceedings of the 2nd International Conference on Computing for Geospatial Research & Applications*. 2011, ACM: Washington, DC, USA. p. 1-6.
21. Velaga, N.R., M.A. Quddus, and A.L. Bristow, *Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems*. *Transportation Research Part C: Emerging Technologies*, 2009. 17(6): p. 672-683.
22. Quddus, Y.Z.a.M.A., *Weight-Based Shortest-Path Aided Map-Matching Algorithm for Low-Frequency Positioning Data*, in *Transportation Research Board 90th Annual Meeting*. 2011: Washington, DC USA.

23. Huang, J., et al., *Parallel Map Matching on Massive Vehicle GPS Data Using MapReduce*, in *HPCC/EUC*. 2013, IEEE. p. 1498-1503.
24. Marchal, F., J.K. Hackney, and K.W. Axhausen, *Efficient map-matching of large GPS data sets. Tests on a speed monitoring experiment in Zürich*. *Transportation Research Record*, 2005. 1935: p. 100.
25. Hummel, O.P.a.B., *A statistical approach to map matching using road network geometry, topology and vehicular motion constraints*, in *2008 11th International IEEE Conference on Intelligent Transportation Systems*. 2008: Beijing, China. p. 862-867.
26. Newson, P. and J. Krumm, *Hidden Markov map matching through noise and sparseness*, in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2009, ACM: Seattle, Washington. p. 336-343.
27. White, C.E., D. Bernstein, and A.L. Kornhauser, *Some map matching algorithms for personal navigation assistants*. *Transportation Research Part C: Emerging Technologies*, 2000. 8(1): p. 91-108.
28. H. Wei, Y.W., G. Forman and Y. Zhu, *Map Matching by Fréchet Distance and Global Weight Optimization*. 2013.
29. Gao, M., et al., *Analysis of Travel Time Patterns in Urban Using Taxi GPS Data*, in *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCoM), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*. 2013, IEEE. p. 512-517.
30. Nandan, N., *A Grid-Based Approach for Similarity Mining of Massive Geospatial Trajectories*, in *Computer and Information Technology (CIT), 2014 IEEE International Conference on*. 2014, IEEE Computer Society. p. 765-768.
31. Lu, M., et al., *Visual Analysis of Multiple Route Choices Based on General GPS Trajectories*. *IEEE Trans. Big Data*, 2017. 3(2): p. 234-247.
32. Yanagisawa, Y., J.-i. Akahani, and T. Satoh, *Shape-Based Similarity Query for Trajectory of Mobile Objects*, in *Mobile Data Management: 4th International Conference, MDM 2003 Melbourne, Australia, January 21–24, 2003*

- Proceedings*, M.-S. Chen, et al., Editors. 2003, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 63-77.
33. Hung, C.-C., W.-C. Peng, and W.-C. Lee, *Clustering and aggregating clues of trajectories for mining trajectory patterns and routes*. *The VLDB Journal*, 2015. 24(2): p. 169-192.
  34. Li, X., et al., *Traffic Density-Based Discovery of Hot Routes in Road Networks*, in *Advances in Spatial and Temporal Databases: 10th International Symposium, SSTD 2007, Boston, MA, USA, July 16-18, 2007. Proceedings*, D. Papadias, D. Zhang, and G. Kollios, Editors. 2007, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 441-459.
  35. Won, J.-I., et al., *Trajectory clustering in road network environment*, in *Computational Intelligence and Data Mining, 2009. CIDM '09*. 2009, IEEE. p. 299-305.
  36. Han, B., L. Liu, and E. Omiecinski, *NEAT: Road Network Aware Trajectory Clustering*, in *Proceedings of the 2012 IEEE 32nd International Conference on Distributed Computing Systems*. 2012, IEEE Computer Society. p. 142-151.
  37. Atsuyuki Okabe, B.B., Kokichi Sugihara, Sung Nok Chiu, D. G. Kendall, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, Second Edition*. 2008.
  38. Sayad, D.S. *Hierarchical Clustering*. 2016 [cited 2016 10 october 2016]; Available from: [http://www.saedsayad.com/clustering\\_hierarchical.htm](http://www.saedsayad.com/clustering_hierarchical.htm).
  39. Ltd, M.T. *Calculate distance, bearing and more between Latitude/Longitude points*. 2016 [cited 2016 10 october 2016]; Available from: <http://www.movable-type.co.uk/scripts/latlong.html>.
  40. Navarro, G., *A guided tour to approximate string matching*. *ACM Comput. Surv.*, 2001. 33(1): p. 31-88.





### ประวัติผู้เขียนวิทยานิพนธ์

นายลำชำ ทองสีนุช เกิดวันที่ 23 กรกฎาคม พ.ศ. 2534 ที่จังหวัดปัตตานี สำเร็จการศึกษาหลักสูตรวิศวกรรมศาสตรบัณฑิต จากภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และเข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ ณ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2557

