บทที่ 3
วิธีการดำเนินโครงการ

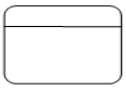
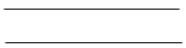
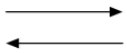
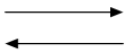
ผู้วิเคราะห์ได้วิเคราะห์ข้อมูลปริมาณการใช้เน็ตประชารัฐ เพื่อเผยแพร่ข้อมูลบนเว็บไซต์ และการจัดการข้อมูลที่มีอยู่ให้เกิดประโยชน์ และได้สารสนเทศที่สอดคล้องกับข้อมูล ผู้วิเคราะห์ได้วิเคราะห์ข้อมูล และออกแบบฐานข้อมูลโดยใช้เครื่องมือที่ประกอบไปด้วยแผนภาพกระแสข้อมูล (Data Flow Diagram) กระบวนการทำนอร์มัลไลเซชัน (Normalization) เพื่อลดการซ้ำซ้อนของข้อมูล และกระบวนในการพัฒนาฐานข้อมูลโดยใช้กระบวนการ CRISP-DM 6ขั้นตอน ทำการวิเคราะห์ข้อมูลแบบอนุกรมเวลาโดยใช้โปรแกรม tableau public โปรแกรม Minitab เพื่อเปรียบเทียบค่าด้วยวิธี moving average ในการวิเคราะห์ข้อมูล ซึ่งในบทนี้จะแสดงถึงวิธีในการดำเนินงาน ดังนี้

3.1 แผนภาพกระแสข้อมูล Data Flow Diagram

แผนภาพกระแสข้อมูล (DFD) ย่อมาจาก Data Flow Diagram เป็นเครื่องมือเชิงโครงสร้างที่ใช้บรรยายภาพรวมของระบบโดยแสดงขั้นตอนการทำงานของระบบหรือโพรเซส(process) ระบุแหล่งกำเนิดของข้อมูล การไหลของข้อมูล ปลายทางข้อมูล การเก็บข้อมูลและการประมวลผลข้อมูล โดยจะช่วยแสดงแผนภาพ ว่าข้อมูลมาจากไหน จะไปไหน เก็บข้อมูลไว้ที่ไหน มีอะไรเกิดขึ้นกับข้อมูลระหว่างทางเรียกว่าแผนภาพกระแสข้อมูลหรือ แผนภาพแสดงความเคลื่อนไหวของข้อมูล

สัญลักษณ์ที่ใช้ในการเขียนแผนภาพกระแสข้อมูลนั้น ประกอบด้วย 4 สัญลักษณ์ คือ การประมวลผล (Process) เส้นทางการไหลของข้อมูล (Data Flows) ตัวแทนข้อมูล (External Agents) และแหล่งจัดเก็บข้อมูล (Data Store) โดยได้มีการศึกษาคิดค้นพัฒนาวิธีการอยู่หลายแบบ แต่ที่เป็นมาตรฐานมี 2 กลุ่ม ได้แก่ ชุดสัญลักษณ์มาตรฐานที่พัฒนาโดย Gane and Sarson (1979) และชุดสัญลักษณ์มาตรฐานที่พัฒนาโดย DeMarco and Yourdon (DeMarco, 1979; Yourdon and Constantine,1979) โดยมีสัญลักษณ์ ดังตารางที่ 3.1

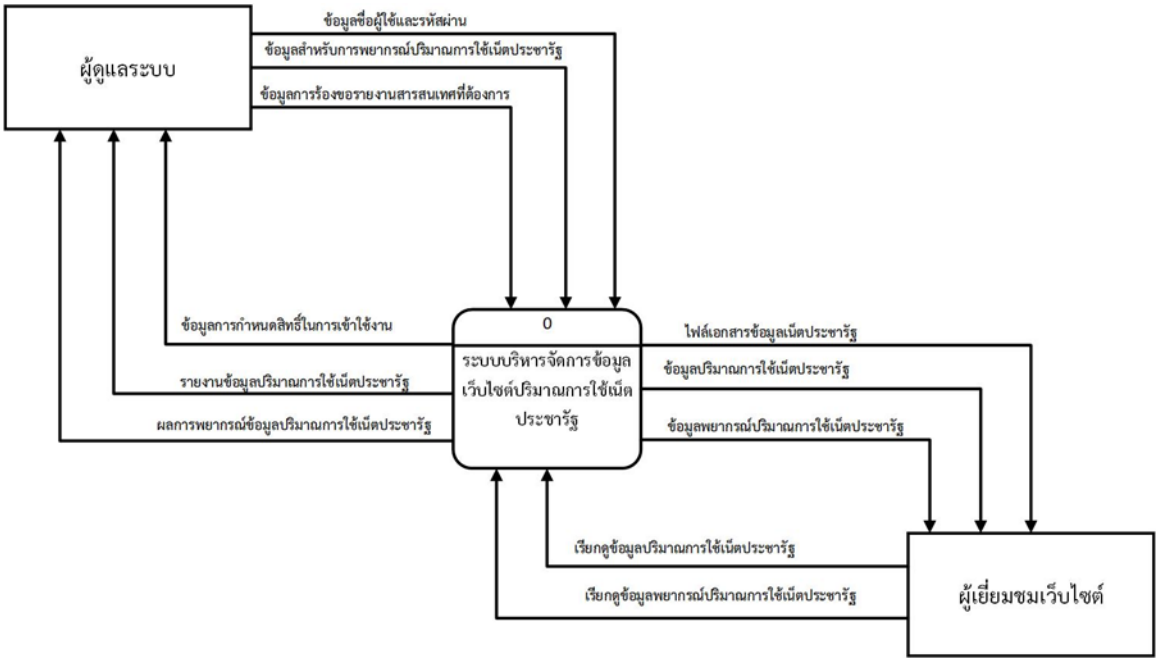
ตารางที่ 3.1 สัญลักษณ์ที่ใช้ในการเขียนแผนการไหลของข้อมูล

DeMarco & Yourdon	Gane & Sarson	ความหมาย
		Process : ขั้นตอนการทำงานภายในระบบ
		Data Store : แหล่งข้อมูลสามารถเป็นได้ทั้งไฟล์ข้อมูลและฐานข้อมูล (File or Database)
		External Agent : ปัจจัยหรือสภาพแวดล้อมที่มีผลกระทบต่อระบบ
		Data Flow : เส้นทางการไหลของข้อมูล แสดงทิศทางของข้อมูลจากขั้นตอนการทำงานหนึ่งไปยังอีกขั้นตอนหนึ่ง

ที่มา : https://sites.google.com/site/krittiyaporn8345/4-2-dfd (2551)

3.1.1 แผนภาพบริบท Context Diagram

แผนภาพบริบท (Context Diagram) จะแสดงภาพโดยรวมระบบของการพัฒนาระบบบริหารจัดการฐานข้อมูลปริมาณการใช้เน็ตประชารัฐ ดังภาพที่ 3.1



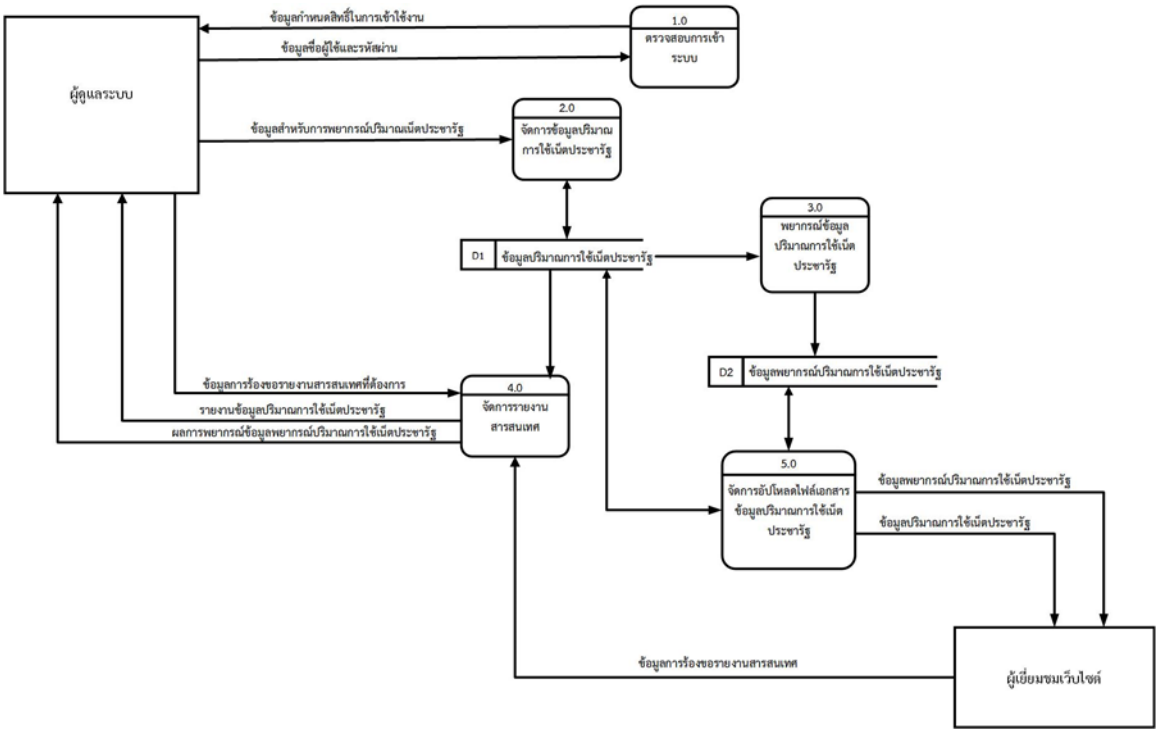
ภาพที่ 3.1 แผนภาพบริบท Context Diagram

จากรูปภาพที่ 3.1 เป็นแผนภาพบริบทระบบของระบบบริหารจัดการฐานข้อมูลปริมาณการใช้เน็ตประชารัฐโดยสามารถแบ่งผู้ใช้ออกเป็น 2 ประเภทดังนี้

1) **ผู้ดูแลระบบ (System Administrator)** สามารถลงชื่อเข้าใช้และรหัสผ่านได้ สามารถเรียกดูข้อมูลพยากรณ์ปริมาณการใช้เน็ตประชารัฐได้ สามารถเรียกดูข้อมูลปริมาณการใช้เน็ตประชารัฐได้ และสามารถจัดการข้อมูลปริมาณการใช้เน็ตประชารัฐได้

2) **ผู้เยี่ยมชมเว็บไซต์ (Website visitor)** สามารถเรียกดูข้อมูลการพยากรณ์ปริมาณการใช้เน็ตประชารัฐได้ สามารถเรียกดูข้อมูลปริมาณการใช้เน็ตประชารัฐได้ และสามารถดาวน์โหลดเอกสารบนเว็บไซต์ได้

3.1.2 แผนภาพกระแสข้อมูลระดับที่ 0 (Data Flow Diagram Level 0)



ภาพที่ 3.2 แผนภาพกระแสข้อมูลระดับที่ 0 (Data Flow Diagram Level 0)

จากภาพที่ 3.2 แสดงแผนภาพกระแสข้อมูลระดับที่ 0 (Data Flow Diagram Level 0) ระบบบริหารฐานข้อมูลปริมาณการใช้เน็ตประชารัฐ ประกอบด้วย 5 กระบวนการหลัก ดังนี้

ตารางที่ 3.2 แสดงคำอธิบายกระบวนการตรวจสอบการเข้าระบบ

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลปริมาณการใช้เน็ตประชารัฐ
DFD Number	1
Process Name	ตรวจสอบการเข้าระบบ
Input Data Flow	ชื่อผู้ใช้และรหัสผ่าน ข้อมูลสิทธิ์การเข้าใช้ระบบ ข้อมูลผู้ดูแลระบบ
Output Data Flow	สิทธิ์ในการเข้าถึงระบบ สิทธิ์การเข้าใช้ระบบ ข้อมูลผู้ดูแลระบบ
Data Store Used	ข้อมูลผู้ใช้และรหัสผ่าน ข้อมูลผู้ดูแลระบบ
Description	เป็นกระบวนการสำหรับตรวจสอบ และกำหนดสิทธิ์ในการเข้าใช้ระบบ โดยระบบจะทำการตรวจสอบชื่อผู้ใช้และรหัสผ่าน ผู้มีสิทธิ์เข้าใช้ระบบโดยมีชื่อผู้ใช้ระบบได้แก่ ผู้ดูแลระบบ

ตารางที่ 3.3 แสดงคำอธิบายกระบวนการจัดการข้อมูลปริมาณการใช้เน็ตประชารัฐ

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลปริมาณการใช้เน็ตประชารัฐ
DFD Number	2
Process Name	จัดการข้อมูลปริมาณการใช้เน็ตประชารัฐ
Input Data Flow	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Output Data Flow	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Data Store Used	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Description	เป็นกระบวนการสำหรับจัดการข้อมูลปริมาณการใช้เน็ตประชารัฐได้แก่ การแก้ไขข้อมูล เพิ่มข้อมูล ลบข้อมูล ผู้มีสิทธิ์ใช้งานกระบวนการนี้ ได้แก่ ผู้ดูแลระบบ และออกสารสนเทศข้อมูลปริมาณการใช้เน็ตประชารัฐให้กับผู้เยี่ยมชมเว็บไซต์

ตารางที่ 3.4 แสดงคำอธิบายกระบวนการพยากรณ์ข้อมูลปริมาณการใช้เน็ตประชารัฐ

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลปริมาณการใช้เน็ตประชารัฐ
DFD Number	3
Process Name	การพยากรณ์ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Input Data Flow	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Output Data Flow	ข้อมูลสำหรับการพยากรณ์ปริมาณการใช้เน็ตประชารัฐ
Data Store Used	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Description	เป็นกระบวนการพยากรณ์ข้อมูลปริมาณการใช้เน็ตประชารัฐ เพื่อเผยแผ่ข้อมูลออกสู่เว็บไซต์

ตารางที่ 3.5 แสดงคำอธิบายกระบวนการจัดการอัปโหลดไฟล์ข้อมูลปริมาณการใช้เน็ตประชารัฐ

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลปริมาณการใช้เน็ตประชารัฐ
DFD Number	4
Process Name	จัดการอัปโหลดไฟล์เอกสารข้อมูลปริมาณการใช้เน็ตประชารัฐ
Input Data Flow	ข้อมูลปริมาณการใช้เน็ตประชารัฐและข้อมูลสำหรับการพยากรณ์ปริมาณการใช้เน็ตประชารัฐ
Output Data Flow	ข้อมูลปริมาณการใช้เน็ตประชารัฐเป็นไฟล์ Excel และข้อมูลเอกสารที่เกี่ยวข้องที่เป็นไฟล์ Excel
Data Store Used	ข้อมูลปริมาณการใช้เน็ตประชารัฐและข้อมูลสำหรับการพยากรณ์ปริมาณการใช้เน็ตประชารัฐ
Description	เป็นกระบวนการสำหรับจัดการดาวน์โหลดไฟล์เอกสารต่าง ๆเช่น ข้อมูลปริมาณการใช้เน็ตประชารัฐและข้อมูลสำหรับการพยากรณ์ปริมาณการใช้เน็ตประชารัฐที่ผู้เยี่ยมชมเว็บไซต์สามารถดาวน์โหลดได้

ตารางที่ 3.6 แสดงคำอธิบายกระบวนการจัดการรายงานสารสนเทศ

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลปริมาณการใช้เน็ตประชารัฐ
DFD Number	5
Process Name	จัดการรายงานสารสนเทศ
Input Data Flow	ข้อมูลปริมาณการใช้เน็ตประชารัฐและข้อมูลสำหรับการพยากรณ์ปริมาณการใช้เน็ตประชารัฐ
Output Data Flow	รายงานข้อมูลปริมาณการใช้เน็ตประชารัฐและผลการพยากรณ์ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Data Store Used	ข้อมูลปริมาณการใช้เน็ตประชารัฐและข้อมูลสำหรับการพยากรณ์ปริมาณการใช้เน็ตประชารัฐ
Description	เป็นกระบวนการสำหรับออกรายงานสารสนเทศ โดยผู้มีสิทธิ์เข้าใช้ระบบได้แก่ ผู้ดูแลระบบ

จากแผนภาพกระแสข้อมูลระดับที่ 0 กระบวนการที่ 2 สามารถแยกย่อยเป็นกระบวนการย่อยระดับที่ 1 ได้ดังนี้

แผนภาพกระแสข้อมูล ระดับที่ 1 กระบวนการที่ 2.1, 2.2, 2.3 จัดการข้อมูลปริมาณการใช้เน็ตประชารัฐ

ภาพที่ 3.3 แผนภาพกระแสข้อมูล ระดับที่ 1 กระบวนการที่ 2.1, 2.2, 2.3 จัดการข้อมูลปริมาณการใช้เน็ตประชารัฐ

ตารางที่ 3.7 แสดงคำอธิบายกระบวนการที่ 2.1 ปรับปรุงข้อมูลปริมาณการใช้เน็ตประชารัฐ

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลปริมาณการใช้เน็ตประชารัฐ
DFD Number	2.1
Process Name	ปรับปรุงข้อมูลปริมาณการใช้เน็ตประชารัฐ
Input Data Flow	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Output Data Flow	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Data Store Used	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Description	เป็นกระบวนการสำหรับการปรับปรุงข้อมูลปริมาณการใช้เน็ตประชารัฐ โดยผู้มีสิทธิ์เข้าใช้ระบบ ได้แก่ ผู้ดูแลระบบ

ตารางที่ 3.8 แสดงคำอธิบายกระบวนการที่ 2.2 แก้ไข เพิ่ม ลบข้อมูลปริมาณการใช้เน็ตประชารัฐ

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลปริมาณการใช้เน็ตประชารัฐ
DFD Number	2.2
Process Name	แก้ไข เพิ่ม ลบข้อมูลปริมาณการใช้เน็ตประชารัฐ
Input Data Flow	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Output Data Flow	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Data Store Used	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Description	เป็นกระบวนการสำหรับการแก้ไข เพิ่ม ลบข้อมูลปริมาณการใช้เน็ตประชารัฐ โดยผู้มีสิทธิ์เข้าใช้ระบบ ได้แก่ ผู้ดูแลระบบ

ตารางที่ 3.9 แสดงคำอธิบายกระบวนการที่ 2.3 บันทึกข้อมูลปริมาณการใช้เน็ตประชารัฐ

Process Description	
System	ระบบบริหารจัดการฐานข้อมูลปริมาณการใช้เน็ตประชารัฐ
DFD Number	2.3
Process Name	บันทึกข้อมูลปริมาณการใช้เน็ตประชารัฐ
Input Data Flow	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Output Data Flow	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Data Store Used	ข้อมูลปริมาณการใช้เน็ตประชารัฐ
Description	เป็นกระบวนการสำหรับการบันทึกข้อมูลปริมาณการใช้เน็ตประชารัฐ เป็นไฟล์เอ็กซ์เซลล์ โดยผู้มีสิทธิ์เข้าใช้ระบบ ได้แก่ ผู้ดูแลระบบ

3.2 กระบวนการทำนอร์มัลไลเซชัน (Normalization)

การทำนอร์มัลไลเซชัน (normalization) เป็นสิ่งที่สำคัญและจำเป็นที่สุดสำหรับการออกแบบฐานข้อมูลแต่ไม่ได้หมายความว่าจะต้องทำการนอร์มัลไลเซขันจนถึงระดับสูงสุดทุกครั้งจึงจะถือว่าเป็นสิ่งที่ดีที่สุด โดยทั่วไปการแสดงผลข้อมูลจากตารางที่อยู่ในนอร์มัลฟอร์มระดับสูงขึ้นเรื่อย ๆจะมีการเชื่อมต่อตารางเป็นจำนวนมากขึ้นตามไปด้วย ทำให้การแสดงผลและการโต้ตอบระหว่างระบบฐานข้อมูลกับผู้ใช้กระทำได้ช้า การออกแบบฐานข้อมูลที่ดีจึงควรพิจารณาถึงความต้องการของผู้ใช้และต้องสามารถตอบสนองกับการใช้งานได้อย่างรวดเร็ว

ในทางทฤษฎีการทำนอร์มัลไลเซชันจะทำได้ระดับสูงสุดคือ นอร์มัลฟอร์มที่ 5 แต่ในทางปฏิบัติแล้วในบางกรณีอาจทำไม่ถึงนอร์มัลฟอร์มที่ 5 ส่วนใหญ่พบว่าการทำถึงระดับที่ 3 ก็เพียงพอที่จะแก้ปัญหาความผิดปกติและความซ้ำซ้อนของข้อมูลได้ ซึ่งประกอบด้วยนอร์มัลฟอร์มแบบต่าง ๆ ดังต่อไปนี้

- นอร์มัลฟอร์มที่ 1 (First Normal Form : 1NF)
- นอร์มัลฟอร์มที่ 2 (Second Normal Form : 2NF)
- นอร์มัลฟอร์มที่ 3 (Third Normal Form : 3NF)
- บอยซ์คอดด์นอร์มัลฟอร์ม (Boyce–Codd Normal Form : BCNF)
- นอร์มัลฟอร์มที่ 4 (Fourth Normal Form : 4NF)

- นอร์มัลฟอร์มที่ 5 (Fifth Normal Form : 5NF)

3.2.1 กระบวนการทำนอร์มัลฟอร์มที่ 1 (First Normal Form: 1NF)

เป็นระดับที่ใช้สำหรับการปรับโครงสร้างของข้อมูลของรีเลชัน ค่าของแอททริบิวท์ต่าง ๆ ในแต่ละทัพเพิล จะต้องมีค่าเพียงค่าเดียว และรีเลชันนั้นจะต้องไม่มีกลุ่มข้อมูลซ้ำ

ตารางที่ 3.10 ตารางข้อมูลก่อนการทำนอร์มัลฟอร์มที่ 1 (First Normal Form : 1NF)

year	month	circuit_no	region	village	subdistrict	district	province	ap_name	ip_address	upload	download	num_users	num_devices
2563	มกราคม	5641d9157	ภาคกลาง	หมู่ 4 ตอนดาล	หนองน้อย	วิเศษชัย	ชัยนาท	CNT-SR711056	172.18.22.122	10369.19777	251139.0968	868	87
2563	มกราคม	5641d9164	ภาคกลาง	หมู่ 8 ทุ่งพง	หนองน้อย	วิเศษชัย	ชัยนาท	CNT-SR770153	172.18.22.178	7090.780382	152578.4056	426	32
2563	มกราคม	5641d9162	ภาคกลาง	หมู่ 6 ตอนแค	หนองน้อย	วิเศษชัย	ชัยนาท	CNT-SR769899	172.18.22.138	2282.044633	49073.33508	497	118
2563	มกราคม	5641d9161	ภาคกลาง	หมู่ 1 วังจิก	สามง่ามท่าโบสถ์	พันดา	ชัยนาท	CNT-SR767794	172.18.22.130	176.0930824	2553.807073	117	38
2563	กุมภาพันธ์	3527d9027	ภาคกลาง	หมู่ 5 บ้านขนอน	หนองน้ำส้ม	อุทัย	พระนครศรีอยุธยา	AYA-SR722538	172.16.28.50	3504.387402	91949.0389	1010	247
2563	กุมภาพันธ์	3527d9028	ภาคกลาง	หมู่ 6 บ้านขนอน	หนองน้ำส้ม	อุทัย	พระนครศรีอยุธยา	AYA-SR722539	172.16.28.58	11138.88543	207505.4985	595	74
2563	กุมภาพันธ์	3527d9026	ภาคกลาง	หมู่ 1 บ้านคู่มักดบ	หนองน้ำส้ม	อุทัย	พระนครศรีอยุธยา	AYA-SR722536	172.16.28.34	11727.34173	211275.6395	978	114
2563	กุมภาพันธ์	3527d9029	ภาคกลาง	หมู่ 3 บ้านวังลึก	หนองน้ำส้ม	อุทัย	พระนครศรีอยุธยา	AYA-SR771145	172.16.28.42	902.6290359	12071.51737	390	104
2563	กุมภาพันธ์	3527d9024	ภาคกลาง	หมู่ 4 บ้านชื่อทราย	หนองน้ำส้ม	อุทัย	พระนครศรีอยุธยา	AYA-SR683957	172.16.28.26	8194.036235	201007.6496	832	65
2563	กุมภาพันธ์	3527d9019	ภาคกลาง	หมู่ 13 บัวชม	ลำตาเสา	วังน้อย	พระนครศรีอยุธยา	AYA-SR770227	172.16.25.18	16251.17405	379775.2291	1561	146
2563	กุมภาพันธ์	3527d9011	ภาคกลาง	หมู่ 2 คลอง 10	ข้าวงาม	วังน้อย	พระนครศรีอยุธยา	AYA-SR767494	172.16.25.90	14634.72507	290874.2777	1359	163
2563	กุมภาพันธ์	3536d9010	ภาคกลาง	หมู่ 6 โรงเจ้า	สามเรือน	บางปะอิน	พระนครศรีอยุธยา	AYA-SR722045	172.16.28.10	20545.44805	265707.4738	2078	208
2563	กุมภาพันธ์	3531d9041	ภาคกลาง	หมู่ 2 บ้านพรานนก	โพสาวหาญ	อุทัย	พระนครศรีอยุธยา	AYA-SR683952	172.16.28.2	9895.697826	131157.9088	2039	170
2563	กุมภาพันธ์	3531d9008	ภาคกลาง	หมู่ 1 บ้านช่อง	ระโสม	ภาชี	พระนครศรีอยุธยา	AYA-SR683768	172.16.21.170	13773.94156	305100.8208	1333	112
2563	กุมภาพันธ์	3527d9004	ภาคกลาง	หมู่ 7 คลองสี่	วังน้อย	วังน้อย	พระนครศรีอยุธยา	AYA-SR683856	172.16.25.114	1503.737324	35547.06071	391	116
2563	กุมภาพันธ์	3527d9012	ภาคกลาง	หมู่ 5 คลอง 5	วังน้อย	วังน้อย	พระนครศรีอยุธยา	AYA-SR768599	172.16.25.130	2914.290437	58619.86517	716	160
2563	กุมภาพันธ์	3537d9089	ภาคกลาง	หมู่ 3 สง่างาม	กกแก้วบูรพา	บางไทร	พระนครศรีอยุธยา	AYA-SR770156	172.16.22.170	15836.29896	338085.2851	3874	416

จากตารางที่ 3.10 จะปรับค่าให้เป็น 1NF โดยมีการทำ multi-value ให้กลายเป็น single value โดยเติมค่าของข้อมูลใน attribute ให้เต็ม โดยกำหนดตาราง pracharat (year, month, circuit_no, region, village, subdistrict, district, province, ip_address, ap_name, upload, download, num_users, num_devices) ซึ่งทุก ๆ field ในแต่ละ record จะเป็น single value นั่นคือ ในตารางหนึ่ง ๆ จะไม่มีค่าของกลุ่มข้อมูลที่ไม่ซ้ำกัน (Repeating Group)

สรุป 1NF

FD1 : pracharat(circuit_no, ip_address → year, month, region, village, subdistrict, district, province, ap_name, upload, download, num_users, num_devices)

3.2.2 กระบวนการทำนอร์มัลฟอร์มที่ 2 (Second Normal Form: 2NF)

เป็นการแก้ไขปัญหาความซ้ำซ้อนของข้อมูล ที่ปรากฏในรีเลชันที่ผ่านการนอร์มัลไลซ์ ระดับที่ 1 โดยรูปแบบนอร์มัลฟอร์มระดับที่ 2 จะต้องมีคุณสมบัติดังนี้คือ รีเลชันใด ๆ จะมีคุณสมบัติอยู่ในรูปแบบนอร์มัลฟอร์มระดับที่ 2 ก็ต่อเมื่อ

1. จะต้องเป็นรีเลชั่นที่มีคุณสมบัติเป็น 1NF
2. แอททรีบิวท์ที่ไม่ใช่คีย์จะต้องขึ้นกับคีย์หลักอย่างแท้จริง

จากตารางที่ 3.10 จะพบว่าในรีเลชันการใช้เน็ตประชารัฐจะมี circuit_no และ ip_address ประกอบกันเป็นคีย์หลัก จากการพิจารณาพบว่า จะไม่ตรงตามนิยามของรูปแบบนอร์มัลฟอร์มระดับที่ 2 เพราะแอตทริบิวต์ ap_name, upload, download, num_users และ num_devices จะมีค่าเปลี่ยนแปลงตามแอตทริบิวต์ circuit_no และ ip_address จึงต้องทำการแยกรีเลชันออกเป็น 3 รีเล

ชัน โดยให้ตารางที่ 1 ชื่อรีเลชันregions, ตารางที่ 2 ชื่อรีเลชันip และตารางที่ 3 ชื่อรีเลชันpracharat เพื่อให้ตารางข้างต้นมีคุณสมบัติเป็นนอร์มัลฟอร์มที่ 2 จะต้องทำการแตกตารางออกมา ตามความสัมพันธ์ของฟังก์ชันการขึ้นต่อกัน เป็น 3ตาราง ดังนี้

ตารางที่ 3.11 แสดงข้อมูลตาราง pracharat ที่อยู่ในรูปของนอร์มัลฟอร์มที่ 2 (Second Normal Form : 2NF)

Pracharat_id	CIRCUIT_NO	IP_ADDRESS	UPLOAD	DOWNLOAD	NUM_USERS	NUM_DEVICES	YEAR	MONTH
P1	7561d9000	172.18.168.2	5779.00	139923.47	1,060	103	2563	กุมภาพันธ์
P2	4485d9000	172.16.140.2	14293.02	262269.53	908	76	2563	กุมภาพันธ์
P3	4485d9001	172.16.140.10	6470.98	102852.57	761	94	2563	กุมภาพันธ์
P4	4486d9000	172.16.140.50	7545.41	159559.70	970	268	2563	กุมภาพันธ์
P5	4486d9001	172.16.140.58	4082.65	71836.22	1,108	349	2563	กุมภาพันธ์
P6	4486d9002	172.16.140.66	20992.69	333250.45	980	76	2563	กุมภาพันธ์
P7	4486d9003	172.16.140.74	10364.73	265594.44	1,419	96	2563	กุมภาพันธ์
P8	4481d9000	172.16.140.18	9515.22	170766.26	996	138	2563	กุมภาพันธ์
P9	4481d9001	172.16.140.26	9557.97	109401.15	969	253	2563	กุมภาพันธ์
P10	4481d9002	172.16.140.34	3451.59	66162.68	698	113	2563	กุมภาพันธ์
P11	4481d9003	172.16.140.42	17685.41	381040.02	1,489	97	2563	กุมภาพันธ์
P12	5334d9000	172.18.36.2	15797.06	270257.81	1,401	99	2563	กุมภาพันธ์
P13	5334d9001	172.18.36.10	7291.10	112815.44	613	92	2563	กุมภาพันธ์

FD1 : pracharat(circuit_no, ip_address → year, month, upload, download, num_users, num_devices)

ตารางที่ 3.12 แสดงข้อมูลตาราง regions ที่อยู่ในรูปของนอร์มัลฟอร์มที่ 2 (Second Normal Form : 2NF)

circuit_no	region	village	subdistrict	district	province
5641d9157	ภาคกลาง	หมู่ 4 ดอนตาล	หนองน้อย	วัดสิงห์	ชัยนาท
5641d9164	ภาคกลาง	หมู่ 8 ทุ่งพงษ์	หนองน้อย	วัดสิงห์	ชัยนาท
5641d9162	ภาคกลาง	หมู่ 6 ดอนแต้	หนองน้อย	วัดสิงห์	ชัยนาท
5641d9161	ภาคกลาง	หมู่ 1 วังจิก	สามง่ามท่าโบสถ์	หันคา	ชัยนาท
3527d9027	ภาคกลาง	หมู่ 5 บ้านขนอน	หนองน้ำส้ม	อุทัย	พระนครศรีอยุธย
3527d9028	ภาคกลาง	หมู่ 6 บ้านขนอน	หนองน้ำส้ม	อุทัย	พระนครศรีอยุธย
3527d9026	ภาคกลาง	หมู่ 1 บ้านคู่ผักตบ	หนองน้ำส้ม	อุทัย	พระนครศรีอยุธย
3527d9029	ภาคกลาง	หมู่ 3 บ้านวังลึก	หนองน้ำส้ม	อุทัย	พระนครศรีอยุธย
3527d9024	ภาคกลาง	หมู่ 4 บ้านขื่อทราย	หนองน้ำส้ม	อุทัย	พระนครศรีอยุธย
3527d9019	ภาคกลาง	หมู่ 13 บัวชม	ลำตาเสา	วังน้อย	พระนครศรีอยุธย
3527d9011	ภาคกลาง	หมู่ 2 คลอง 10	ข้าวงาม	วังน้อย	พระนครศรีอยุธย
3536d9010	ภาคกลาง	หมู่ 6 โรงเจ้า	สามเรือน	บางปะอิน	พระนครศรีอยุธย
3531d9041	ภาคกลาง	หมู่ 2 บ้านพรานนก	โพสาวหาญ	อุทัย	พระนครศรีอยุธย

FD2 : regions(circuit_no → region, village, subdistrict, district, province)

ตารางที่ 3.13 แสดงข้อมูลตาราง ip ที่อยู่ในรูปของนอร์มัลฟอร์มที่ 2 (Second Normal Form : 2NF)

ip_address	ap_name
172.18.22.122	CNT-SR711056
172.18.22.178	CNT-SR770153
172.18.22.138	CNT-SR769899
172.18.22.130	CNT-SR767794
172.16.28.50	AYA-SR722538
172.16.28.58	AYA-SR722539
172.16.28.34	AYA-SR722536
172.16.28.42	AYA-SR771145
172.16.28.26	AYA-SR683957
172.16.25.18	AYA-SR770227
172.16.25.90	AYA-SR767494
172.16.28.10	AYA-SR722045
172.16.28.2	AYA-SR683952
172.16.21.170	AYA-SR683768

FD3 : ip(ip_address → ap_name)

จากตารางที่ 3.13 จะเห็นได้ว่า Attribute ทุกตัวไม่มี Transitive Dependency จึงอยู่ในรูปแบบของ 3NF แล้ว และไม่มี Attribute อื่นใน Relation ที่สามารถระบุค่าของ Attribute ที่เป็นคีย์หลักหรือส่วนใดของคีย์หลักในกรณีที่คีย์หลักเป็นคีย์ผสม ตารางดังกล่าวจึงอยู่ในรูปแบบของ BCNF แล้ว

จากข้อมูลของปริมาณการใช้เน็ตประชารัฐที่ได้นำมาผ่านกระบวนกานอร์มัลไลเซชันทั้ง 2 ขั้นตอนก็เพียงพอต่อการใช้งานและตอบสนองความต้องการของผู้ใช้งานแล้ว

สรุปการทำนอร์มัลไลเซชัน (Normalization)

1NF

FD1 : pracharat(circuit_no, ip_address → year, month, region, village, subdistrict, district, province, ap_name, upload, download, num_users, num_devices)

2NF

FD1 : pracharat(circuit_no, ip_address → year, month, upload, download, num_users, num_devices)

FD2 : regions(circuit_no → region, village, subdistrict, district, province)

FD3 : ip(ip_address → ap_name)

3.3 ความสัมพันธ์ของฐานข้อมูล (ER-Diagram)

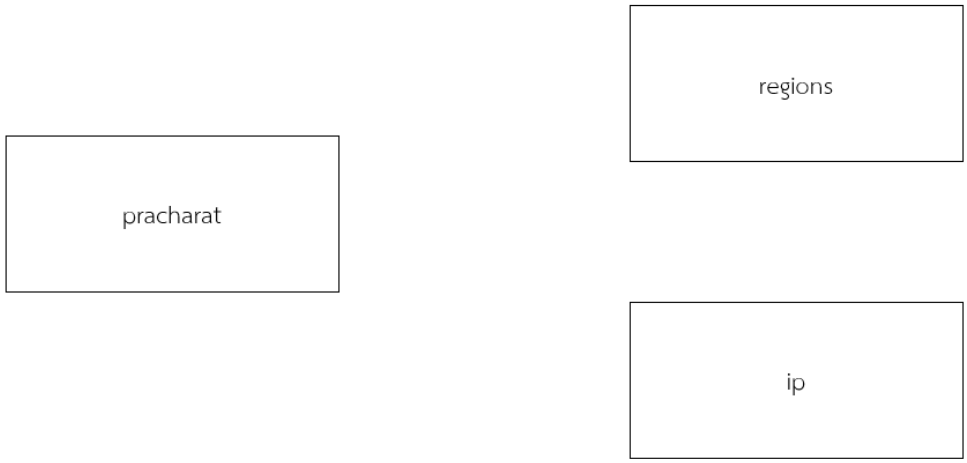
ER Diagram คือ แบบจำลองที่ใช้อธิบายโครงสร้างของฐานข้อมูลซึ่งเขียนออกมาในลักษณะของรูปภาพ การอธิบายโครงสร้างและความสัมพันธ์ของข้อมูล (Relationship) ประกอบด้วย

1. เอนทิตี้ (Entity) เป็นวัตถุ หรือสิ่งของที่เราสนใจในระบบงานนั้น ๆ
2. แอททริบิว (Attribute) เป็นคุณสมบัติของวัตถุที่เราสนใจ
3. ความสัมพันธ์ (Relationship) คือ ความสัมพันธ์ระหว่างเอนทิตี้

ER Diagram มีความสำคัญต่อการพัฒนาระบบงานฐานข้อมูล Application ต่าง ๆที่ต้องการการเก็บข้อมูลอย่างมีระบบมีโครงสร้าง ดังนั้น ER Diagram จึงใช้เพื่อเป็นเอกสารในการสื่อสารระหว่าง นักออกแบบระบบ และนักพัฒนาระบบเพื่อให้สื่อสารอย่างตรงกัน และเป็นสากล โดยมีขั้นตอนในการพัฒนาดังนี้

3.3.1 การกำหนดเอนทิตี้

การกำหนดเอนทิตี้เป็นการกำหนดสิ่งที่สนใจและต้องการจัดเก็บข้อมูลที่มีในระบบงาน โดยดูจากลักษณะหน้าที่ของระบบงานว่ามีรายละเอียดการทำงานอย่างไรผู้วิเคราะห์ได้กำหนดเอนทิตี้ของฐานข้อมูลการใช้เน็ตประชารัฐ ประกอบด้วย เอนทิตี้ pracharat, เอนทิตี้ circuit, เอนทิตี้ ip และเอนทิตี้ volume



ภาพที่ 3.4 การกำหนดเอนทิตี้ของฐานข้อมูลการใช้เน็ตประชารัฐ

3.3.2 การกำหนดแอตทริบิวต์ของเอนทิตี้

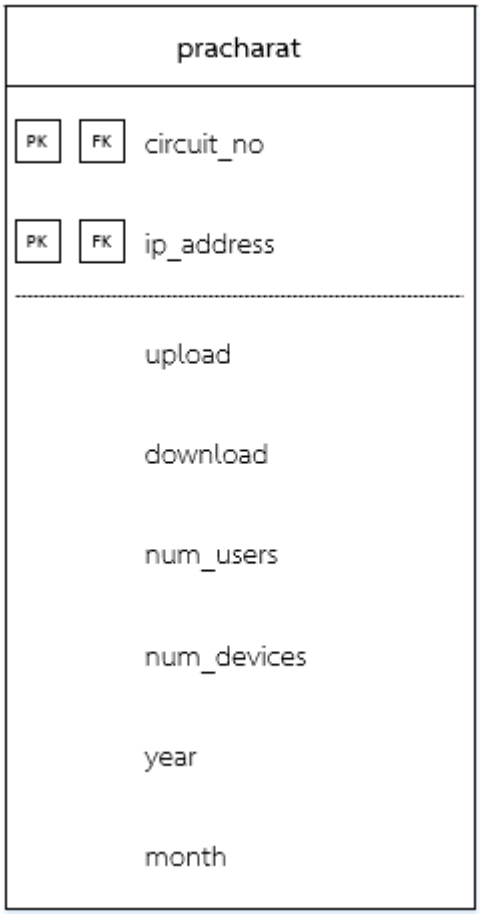
การกำหนดแอตทริบิวต์ของเอนทิตี้เป็นการกำหนดคุณลักษณะต่าง ๆของแต่ละเอนทิตี้รวมทั้งพิจารณาแอตทริบิวต์ที่จะทำหน้าที่เป็นคีย์หลักของเอนทิตี้ด้วย ประกอบไปด้วย

เอนทิตี้ pracharat(upload, download, num_users, num_devices, year, month)

เอนทิตี้ circuit(circuit_no, region, village, subdistrict, district, province)

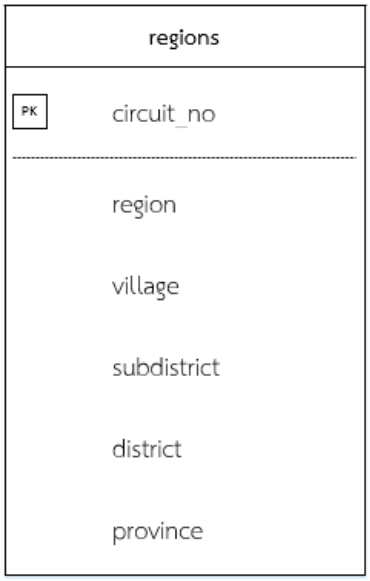
เอนทิตี้ ip (ip_address, ap_name)

ภาพที่ 3.5 เป็นการกำหนดแอตทริบิวต์ของเอนทิตี้ pracharat ประกอบด้วย แอตทริบิวต์category_name, แอตทริบิวต์continent_name โดยมีคีย์หลักคือ Pracharat_id และคีย์นอก (Foreign Key) คือแอตทริบิวต์ pracharat_id



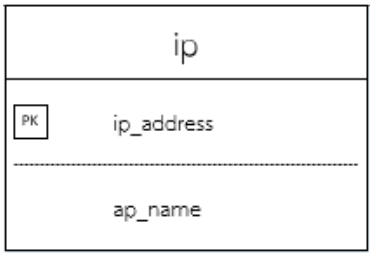
ภาพที่ 3.5 การกำหนดแอตทริบิวต์ของเอนทิตี้ pracharat

ภาพที่ 3.6 เป็นการกำหนดแอตทริบิวต์ของเอนทิตี้ regions ประกอบด้วยแอตทริบิวต์ circuit_no, แอตทริบิวต์ region, แอตทริบิวต์ village, แอตทริบิวต์ subdistrict, แอตทริบิวต์ district, แอตทริบิวต์ province โดยมีคีย์หลัก (Primary key) คือแอตทริบิวต์ circuit_no



ภาพที่ 3.6 การกำหนดแอตทริบิวต์ของเอนทิตี้ regions

ภาพที่ 3.7 เป็นการกำหนดแอตทริบิวต์ของเอนทิตี้ ip ประกอบด้วยแอตทริบิวต์ ip_addressและแอตทริบิวต์ ap_name โดยมีคีย์หลัก (Primary key) คือแอตทริบิวต์ ip_address



ภาพที่ 3.7 การกำหนดแอตทริบิวต์ของเอนทิตี้ ip

3.3.3 การกำหนดความสัมพันธ์ระหว่างเอนทิตี้ (Relationships)

การสร้างความสัมพันธ์ระหว่างเอนทิตี้ในฐานข้อมูลเชิงสัมพันธ์กระทำได้ โดยการกำหนดให้เอนทิตี้ที่มีความสัมพันธ์กันมีแอททริบิวต์ที่เหมือนกัน และใช้ค่าของแอททริบิวต์ที่เหมือนกันเป็นตัวระบุข้อมูลในเอนทิตี้ที่มีความสัมพันธ์กัน ความสัมพันธ์ระหว่างเอนทิตี้ แบ่งออกเป็น 3 ประเภท ดังนี้

1. ความสัมพันธ์แบบหนึ่งต่อหนึ่ง (One-to-one Relationship)

ความสัมพันธ์แบบหนึ่งต่อหนึ่งเป็นการแสดงความสัมพันธ์ของข้อมูลในเอนทิตี้หนึ่งว่ามีความสัมพันธ์กับข้อมูลของอีกเอนทิตี้หนึ่ง ในลักษณะที่เป็นหนึ่งต่อหนึ่ง เช่น ประชาชน 1 คน จะต้องมีหมายเลขรหัสประจำตัวประชาชน 1 หมายเลข ซึ่งไม่ซ้ำกัน

2. ความสัมพันธ์แบบหนึ่งต่อกลุ่ม (One-to-many Relationship)

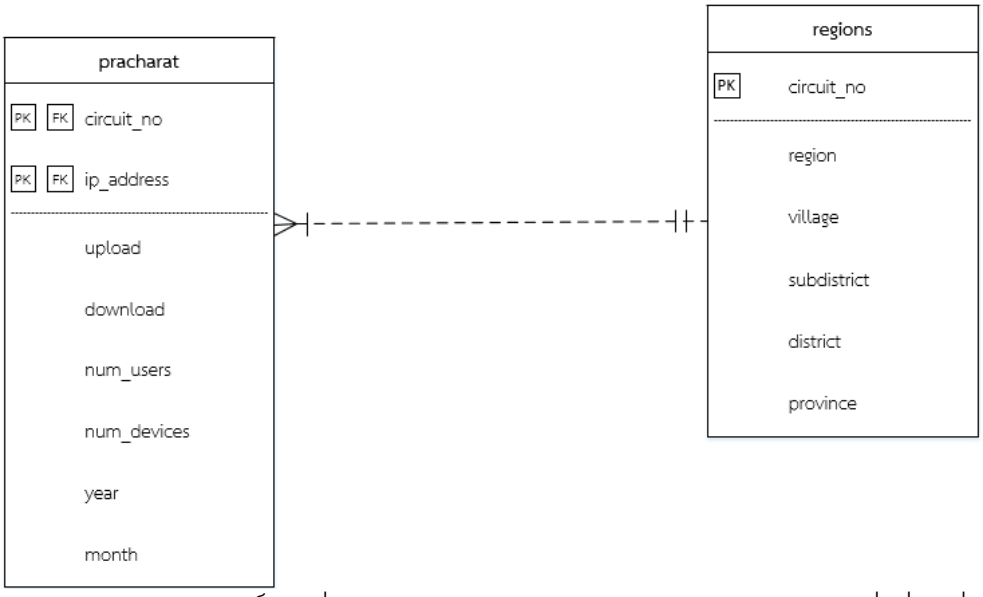
ความสัมพันธ์แบบหนึ่งต่อกลุ่ม เป็นการแสดงความสัมพันธ์ของข้อมูลของเอนทิตี้หนึ่ง ว่ามีความสัมพันธ์กับข้อมูลหลายข้อมูลกับอีกเอนทิตี้หนึ่ง เช่น แผนกแต่ละแผนกจะประกอบไปด้วยพนักงานที่สังกัดอยู่ในแผนกหลายคน

3. ความสัมพันธ์แบบกลุ่มต่อกลุ่ม (Many-to-many Relationship)

ความสัมพันธ์แบบกลุ่มต่อกลุ่ม เป็นการแสดงความสัมพันธ์ของข้อมูลของ 2 เอนทิตี้ ในลักษณะแบบกลุ่มต่อกลุ่ม เช่น นักศึกษาหลายคน อาจเรียนอยู่ในหลายหลักสูตร

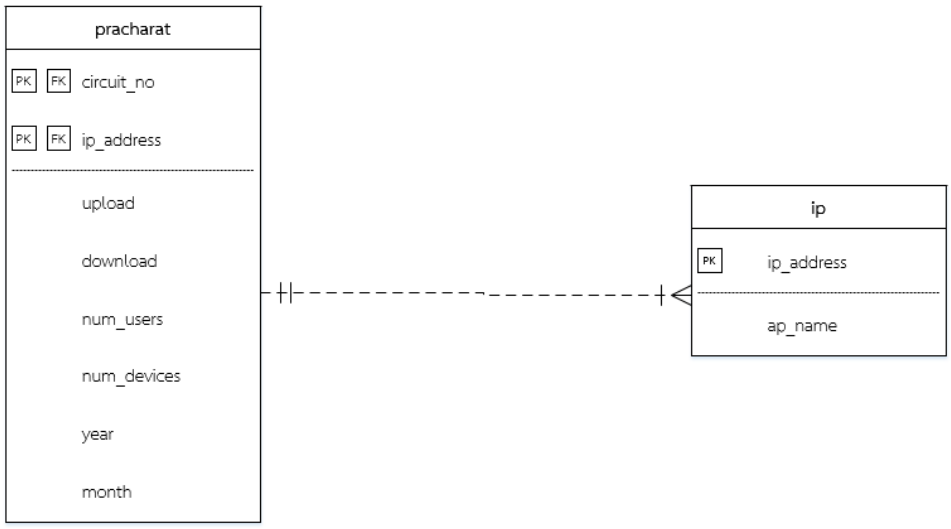
จากการศึกษาผู้วิเคราะห์ได้กำหนดความสัมพันธ์ระหว่างเอนทิตี้ของทั้ง 3 ตาราง
ประกอบไปด้วยตารางที่ 1 รีเลชัน pracharat, ตารางที่ 2 รีเลชันip, และตารางที่ 3 รีเลชัน regions มีความสัมพันธ์ระหว่างเอนทิตี้ดังนี้

1. ตาราง pracharat มีความสัมพันธ์ระหว่างเอนทิตี้กับตาราง regions แบบหนึ่งต่อกลุ่ม (1 : N) หมายความว่า ปริมาณการใช้เน็ต 1 ปริมาณสามารถใช้ได้หลายพื้นที่



ภาพที่ 3.8 แสดงความสัมพันธ์ระหว่างตาราง pracharat กับตาราง regions แบบหนึ่งต่อกลุ่ม (1 : N)

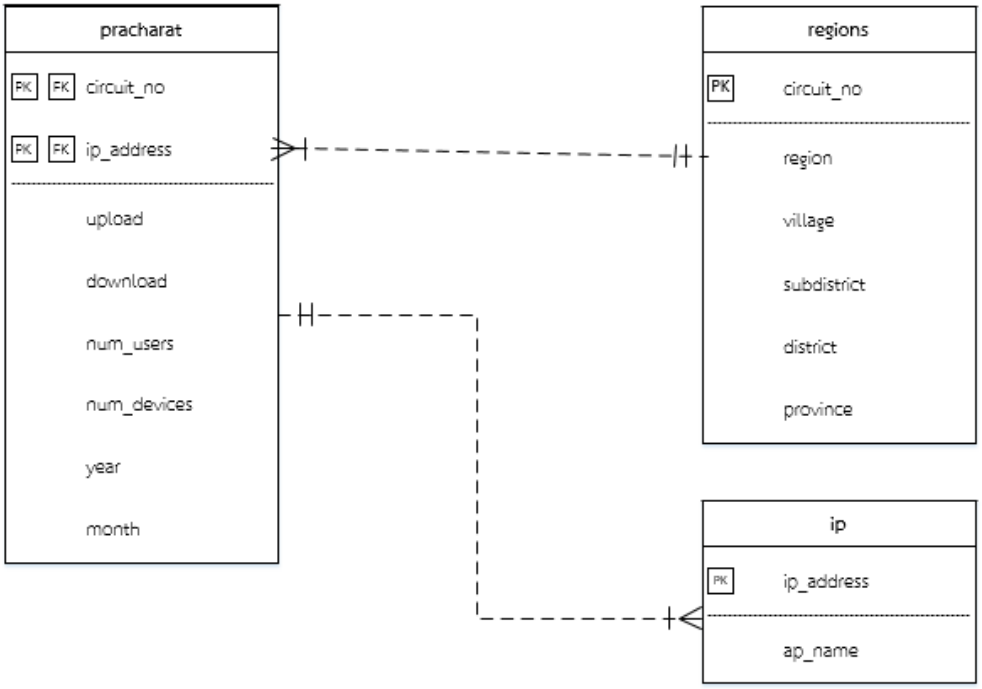
2.ตาราง pracharat มีความสัมพันธ์ระหว่างเอนทิตี้กับตาราง ip แบบหนึ่งต่อกลุ่ม (1 : N) หมายความว่า อุปกรณ์กระจายสัญญาณ 1 ตัว สามารถใช้ปริมาณได้หลายปริมาณ



ภาพที่ 3.9 แสดงความสัมพันธ์ระหว่างตาราง pracharat กับตาราง ip แบบหนึ่งต่อกลุ่ม (1 : N)

3.3.4 การเขียนเป็นแบบจำลองแผนภาพ Entity Relationship Diagram (ERD)

การนำรายละเอียดการออกแบบทั้งหมดในแต่ละขั้นตอน นำมาวาดแบบจำลองแผนภาพ E-R Diagram กำหนดให้ความสัมพันธ์ระหว่างเอนทิตี้ ดังต่อไปนี้



ภาพที่ 3.10 ความสัมพันธ์ของฐานข้อมูลแบบ Crow's Foot Model

จากความสัมพันธ์ของฐานข้อมูลแบบ Crow's Foot Model จะได้ตารางทั้งหมด 3 ตาราง ได้แก่ ตาราง pracharat ตารางregions และตาราง ip แต่ละตารางเก็บข้อมูลคือตาราง pracharat เก็บ circuit_no, ip_address, year, month, upload, download, num_users, num_devices ส่วนตาราง regions เก็บ circuit_no, region, village, subdistrict, district, province และตาราง ip เก็บ ip_address กับ ap_name

3.4 พจนานุกรมข้อมูล (Data Dictionary)

จากการออกแบบฐานข้อมูลซึ่งมีการจัดการระบบฐานข้อมูลให้กับระบบที่ประกอบไปด้วยตารางข้อมูลต่าง ๆ ดังตารางที่ 3.14

ตารางที่ 3.14 แสดงชื่อตารางทั้งหมดของระบบฐานข้อมูล

ลำดับ	ชื่อตาราง	ประเภท	รายละเอียด
1	Pracharat	Transaction	เก็บข้อมูลปริมาณการใช้เน็ตประชารัฐ
2	regions	Reference	เก็บข้อมูลพื้นที่ เช่น หมู่บ้าน อำเภอ จังหวัด
3	ip	Reference	เก็บข้อมูลเกี่ยวกับอุปกรณ์กระจายสัญญาณ

คำอธิบาย ประเภทของตาราง ได้แก่

Transaction หมายถึง ตารางที่มีการเปลี่ยนแปลงของข้อมูล

Reference หมายถึง ตารางที่มีการอ้างอิงถึงข้อมูล

จากตารางที่ 3.14 สามารถแสดงรายละเอียดของแต่ละตารางได้ดังนี้

ตารางที่ 3.15 แสดงรายละเอียดของตาราง pracharat

ชื่อตาราง : pracharat

ประเภทตาราง : Transaction

คำอธิบาย : เก็บข้อมูลปริมาณการใช้เน็ตประชารัฐ

คีย์หลัก : circuit_no และ ip_address

คีย์รอง : circuit_no และ ip_address

เขตข้อมูล	ชนิดและขนาด	ความหมาย	ตัวอย่าง
pracharat_id	INT(5)	รหัสประชารัฐ	P1
circuit_no	INT(10)	เลขหมายวงจร	7561d9000
ip_address	INT(12)	หมายเลข IP	172.18.168.2
upload	DECIMAL(15,2)	ปริมาณการนำข้อมูลเข้า (เมกะไบต์)	5,779.00
download	DECIMAL(15,2)	ปริมาณการนำข้อมูลออก (เมกะไบต์)	139,923.47
num_users	NUMBER	จำนวนผู้ใช้งาน (ครั้ง)	1,145

ตารางที่ 3.15 แสดงรายละเอียดของตาราง pracharat (ต่อ)

ชื่อตาราง : pracharat ประเภทตาราง : Transaction คำอธิบาย : เก็บข้อมูลปริมาณการใช้เน็ตประชารัฐ คีย์หลัก : circuit_no และ ip_address คีย์รอง : circuit_no และ ip_address			
เขตข้อมูล	**ชนิดและขนาด**	**ความหมาย**	**ตัวอย่าง**
num_devices	NUMBER	จำนวนอุปกรณ์ (เครื่อง)	143
year	VARCHAR(4)	ปี	2563
month	VARCHAR(50)	เดือน	มกราคม

ตารางที่ 3.16 แสดงรายละเอียดของตาราง regions

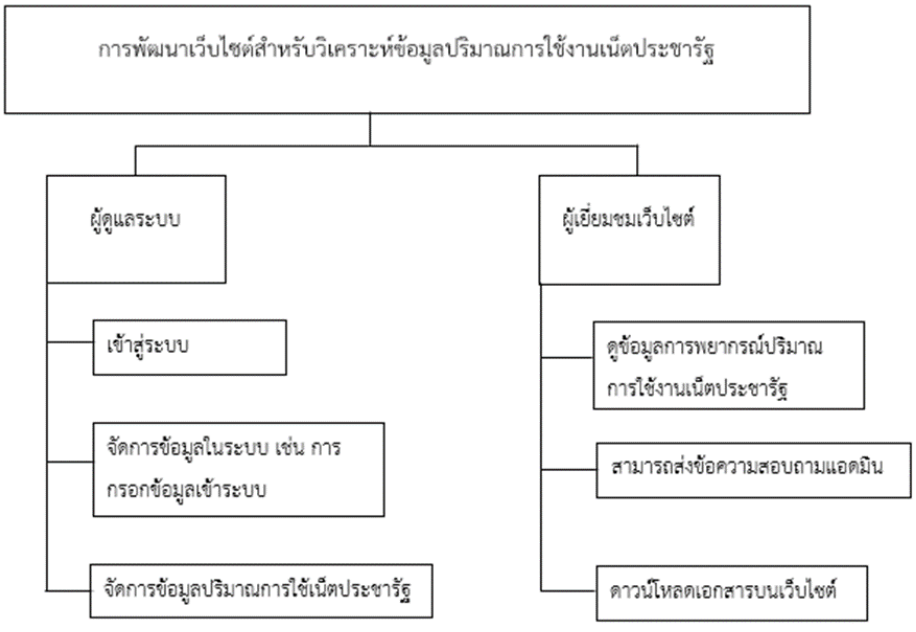
ชื่อตาราง : regions ประเภทตาราง : Reference คำอธิบาย : เก็บข้อมูลพื้นที่ เช่น หมู่บ้าน อำเภอ จังหวัด คีย์หลัก : circuit_no คีย์รอง : –			
เขตข้อมูล	**ชนิดและขนาด**	**ความหมาย**	**ตัวอย่าง**
circuit_no	INT(10)	เลขหมายวงจร	7561d9000
region	VARCHAR(150)	ภูมิภาค	ใต้
village	VARCHAR(150)	ชื่อหมู่บ้าน	หมู่ 9 หน้าค่าย
subdistrict	VARCHAR(150)	ตำบล	คลองท่อมใต้
district	VARCHAR(150)	อำเภอ	คลองท่อม
province	VARCHAR(150)	จังหวัด	กระบี่

ตารางที่ 3.17 แสดงรายละเอียดของตาราง ip

ชื่อตาราง : ip ประเภทตาราง : Reference คำอธิบาย : เก็บข้อมูลเกี่ยวกับอุปกรณ์กระจายสัญญาณ คีย์หลัก : ip_address คีย์รอง : –			
เขตข้อมูล	**ชนิดและขนาด**	**ความหมาย**	**ตัวอย่าง**
ip_address	INT(14)	หมายเลข IP	172.18.168.170
ap_name	VARCHAR(50)	ชื่ออุปกรณ์กระจายสัญญาณ Wireless	KBI–SR664686

3.5 โครงสร้างระบบ (System Structure)

ลักษณะภายในของโครงสร้างระบบแสดงถึงโครงสร้างการใช้งานในระบบฐานข้อมูลปริมาณการใช้เน็ตประชารัฐ ส่วนที่เป็นส่วนสำคัญโดยตรง ได้แก่ ผู้ดูแลระบบ และผู้เยี่ยมชมเว็บไซต์ ดังภาพที่ 3.11



ภาพที่ 3.11 แผนผังโครงสร้างการใช้งานในระบบฐานข้อมูลการใช้เน็ตประชารัฐ

จากภาพที่ 3.11 เพื่อความสะดวกในการออกแบบฐานข้อมูลผู้วิเคราะห์จึงจัดแบ่งระบบออกเป็นส่วนย่อย ๆ มีทั้งหมด 2 ส่วนด้วยกัน และให้แต่ละส่วนมีหน้าที่รับผิดชอบการทำงานในแต่ละด้าน โดยไม่คาบเกี่ยวกันแต่สัมพันธ์กัน ส่วนแรกคือส่วนของผู้ดูแลระบบ จะเป็นส่วนที่สามารถจัดการข้อมูลทั้งหมดในระบบได้ ส่วนที่สองคือส่วนผู้เยี่ยมชมเว็บไซต์ เป็นส่วนที่บุคคลภายนอกสามารถเข้ามาเยี่ยมชมข้อมูลบนเว็บไซต์ได้และสามารถดาวน์โหลดข้อมูลบนเว็บไซต์ได้

3.6 กระบวนการ CRISP-DM

กระบวนการ CRISP-DM นี้จะประกอบด้วย 6 ขั้นตอน แต่ละขั้นตอนจะเป็นขั้นตอนที่ต่อเนื่องกันนั่นคือขั้นตอนถัดไปจะรอผลลัพธ์จากขั้นตอนก่อนหน้าซึ่งแสดงด้วยลูกศรที่เชื่อมระหว่างกล่องสี่เหลี่ยมแต่ละกล่อง ตัวอย่างเช่นเมื่อได้ผลลัพธ์จากขั้นตอนการเตรียมข้อมูล (Data Preparation) แล้วจะนำไปสร้างโมเดลจำแนกประเภทข้อมูลในขั้น Modeling และหลังจากนั้นอาจจะย้อนกลับมาเปลี่ยนแปลงข้อมูลให้ถูกต้องมากขึ้นเพื่อหวังว่าจะโมเดลที่ให้ความถูกต้องมากขึ้น เป็นต้น

ภาพที่ 3.12 แสดงกระบวนการ CRISP-DM

ที่มา : https://datacubeth.ai/crisp-dm/

โดยกระบวนการ CRISP-DM จะประกอบไปด้วย 6 ขั้นตอน ดังนี้

1. Business Understanding

ขั้นตอนแรกมุ่งไปที่การทำความเข้าใจธุรกิจ ปัญหาและวัตถุประสงค์ของโครงการจากมุมมองทางธุรกิจ จากนั้นแปลงปัญหาให้อยู่ในรูปของโจทย์สำหรับการวิเคราะห์ข้อมูล และวางแผนกาดำเนินงานเบื้องต้น

2. Data Understanding

ขั้นตอนนี้เริ่มต้นด้วยการรวบรวมข้อมูล จากนั้นทำความเข้าใจ ตรวจสอบคุณภาพ และเลือกข้อมูลที่เก็บรวบรวมมาว่าจะใช้ข้อมูลใดบ้างในการวิเคราะห์และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์

3. Data Preparation

ขั้นตอนนี้เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา (raw data) ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง (data cleaning) เช่น การแปลงข้อมูลให้อยู่ในช่วง (scale) เดียวกัน หรือการเติมข้อมูลที่ขาดหายไป เป็นต้น โดยขั้นตอนนี้จะเป็นขั้นตอนที่ใช้เวลามากที่สุดของกระบวนการ CRISP-DM

4. Modeling

ขั้นตอนนี้จะเป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคทางดาต้า ไมน์นิง ที่ได้แนะนำไปแล้ว เช่น การจำแนกประเภทข้อมูล หรือ การแบ่งกลุ่มข้อมูล ซึ่งในขั้นตอนนี้หลายเทคนิคจะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด ดังนั้นในบางครั้งอาจจะต้องมีการย้อนกลับไปที่ขั้นตอนที่ 3 Data Preparation เพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิคด้วย ตัวอย่างเทคนิคในการวิเคราะห์ข้อมูลต่าง ๆ

5. Evaluation

การวัดประสิทธิภาพของโมเดลที่ได้จากขั้นตอนที่ 4 เพื่อวัดว่าโมเดลมีประสิทธิภาพเพียงพอต่อการนำไปใช้งานแล้วหรือไม่ ซึ่งโมเดลแต่ละประเภทมีการวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรก หรือ มีความน่าเชื่อถือมากน้อยเพียงใด ซึ่งอาจจะย้อนกลับไปยังขั้นตอนก่อนหน้าเพื่อเปลี่ยนแปลงแก้ไขเพื่อให้ได้ผลลัพธ์ตามที่ต้องการได้

6. Deployment

ความรู้หรือข้อมูลที่ได้รับจากกระบวนการวิเคราะห์จำเป็นต้องนำเสนอในลักษณะที่ผู้มีส่วนได้ส่วนเสียสามารถนำไปใช้ได้เมื่อพวกเขาต้องการ ในขั้นตอนการปรับใช้ต้องมีการสร้างแผนสำหรับการปรับใช้การบำรุงรักษาและการตรวจสอบสำหรับการนำไปใช้งานและการสนับสนุนในอนาคตด้วย จากมุมมองของโครงการรายงานขั้นสุดท้ายของโครงการจำเป็นต้องสรุปประสบการณ์ของโครงการและทบทวนโครงการเพื่อดูสิ่งที่ต้องปรับปรุงบทเรียนที่สร้างขึ้นการนำผลองค์ความรู้ที่ได้จากการหาผลลัพธ์ด้วยเทคนิค Data Mining ไปใช้ ประโยชน์ต่อในองค์กร หรือบริษัท

ขั้นตอนในการดำเนินการวิเคราะห์ข้อมูล ผู้วิเคราะห์ได้เลือกวิธีการวิเคราะห์ข้อมูลโดยใช้โมเดล Time Series เพื่อต้องการพยากรณ์ข้อมูลล่วงหน้าของแต่ละเดือนข้อมูลปริมาณการใช้เน็ตประชารัฐ ที่จะมีแนวโน้มเพิ่มขึ้นหรือลดลง ข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่เปลี่ยนแปลงไปตามเดือนที่มีการใช้เน็ตประชารัฐ มีการจัดเก็บข้อมูลเป็นช่วงเวลามีการเพิ่มขึ้นหรือลดลงอย่างต่อเนื่อง สามารถนำข้อมูลมาพยากรณ์หรือทำนายข้อมูลที่จะเกิดขึ้นในอนาคตได้ จากการที่ผู้วิเคราะห์ได้ศึกษาวรรณกรรมที่เกี่ยวข้องงานวิจัยที่พบส่วนใหญ่จะใช้โมเดล Time Series ในการวิเคราะห์ข้อมูล

3.1.1 การทำความเข้าใจธุรกิจ (Business Understanding)

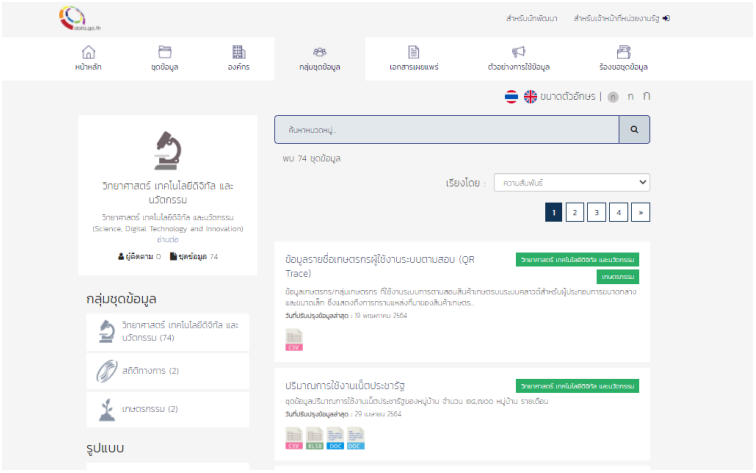
ผู้วิเคราะห์ทำความเข้าใจกระบวนการทางข้อมูล และรับฟังปัญหารวมถึงการมองหาปัญหาจากเรื่องต่าง ๆ เพื่อที่นำผลการวิเคราะห์ข้อมูลไปใช้ให้เกิดประโยชน์ โดยเรียงลำดับความสำคัญ และกำหนดวัตถุประสงค์ที่จะนำไปสู่รูปแบบการวิเคราะห์ข้อมูล ให้อยู่ในรูปโจทย์ของการวิเคราะห์ฐานข้อมูล และวางแผนในการดำเนินการโดยการวิเคราะห์ข้อมูลรูปแบบความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลของข้อมูลปริมาณขยะมูลฝอย ในประเทศไทย เพื่อที่จะนำผลการวิเคราะห์ข้อมูลไปใช้ประโยชน์ตามความต้องการ และตามวัตถุประสงค์ที่จะนำไปสู่รูปแบบการวิเคราะห์ข้อมูล

3.1.2 การทำความเข้าใจข้อมูล (Data Understanding)

ผู้วิเคราะห์ข้อมูลทำการจัดเก็บรวบรวมข้อมูล และพิจารณาตรวจสอบข้อมูลที่ได้รับดูความถูกต้อง มีความน่าเชื่อถือของข้อมูล โดยพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์ ให้สอดคล้องกับวัตถุประสงค์ที่กำหนดไว้ ผู้วิเคราะห์ได้สำรวจภาพรวมก่อนนำข้อมูลมาใช้ในการวิเคราะห์ข้อมูล จากเว็บไซต์ https://data.go.thซึ่งเป็น

ศูนย์กลางการให้บริการข้อมูลเปิดภาครัฐ (Open Government Data) ภายใต้ชื่อdata.go.th ที่สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) ผู้วิเคราะห์ข้อมูลจึงได้ดำเนินการตามขั้นตอนดังนี้

1.) ผู้วิเคราะห์รวบรวมข้อมูลจาก จากเว็บไซต์ https://data.go.th ผู้วิเคราะห์ทำการเลือกกลุ่มชุดข้อมูลวิทยาศาสตร์ เทคโนโลยีดิจิทัล และนวัตกรรมจากนั้นเลือกหัวข้อข้อมูลปริมาณการใช้เน็ตประชารัฐ



ภาพที่ 3.13 แสดงขั้นตอนการหาข้อมูลจากเว็บไซต์

ที่มา : data.go.th

2.) ผู้วิเคราะห์ทำการจัดเก็บ และรวบรวมข้อมูลทั้งหมดของข้อมูลปริมาณการใช้เน็ตประชารัฐในปี 2563 ตั้งแต่เดือนมกราคม – ธันวาคม

ภาพที่ 3.14 แสดงขั้นตอนการดาวน์โหลดข้อมูล

3.) ตรวจสอบความถูกต้อง และความน่าเชื่อถือของข้อมูล ซึ่ง 1 ไฟล์จประกอบด้วย 14แอตทริบิวท์คือ year, month, circuit_no, region, village, subdistrict, district, province, ip_address, ap_name, upload, download, num_users, num_devices

YEAR	MONTH	CIRCUIT_NO	REGION	VILLAGE	SUBDISTRICT	DISTRICT	PROVINCE	AP_NAME	IP_ADDRESS	UPLOAD	DOWNLOAD	NUM_USERS	NUM_DEVICES
2563	มีนาคม	7561d9039	ใต้	หมู่ 9 หน้าค่าย	คลองพ่อแก้	คลองพ่อแก้	กระบี่	KBI-SR669686	172.18.168.170	2322.88	52839.10	132	36
2563	มีนาคม	7561d9071	ใต้	หมู่ 9 ไสยง	ดินอุ่น	เขาพนม	กระบี่	KBI-SR707339	172.18.169.18	3241.57	64914.77	297	51
2563	มีนาคม	7561d9026	ใต้	หมู่ 9 ไร่น้ำเย็น	พรุเตียว	เขาพนม	กระบี่	KBI-SR669649	172.18.168.242	577.67	16220.12	61	13
2563	มีนาคม	7561d9022	ใต้	หมู่ 9 ปากบางหว้าน	เขาดิน	เขาพนม	กระบี่	KBI-SR669636	172.18.169.146	866.09	28080.95	81	14
2563	มีนาคม	7561d9043	ใต้	หมู่ 9 ทุ่งควนหิน	ห้วยน้ำขาว	คลองพ่อแก้	กระบี่	KBI-SR669693	172.18.169.74	1489.67	26005.49	250	82
2563	มีนาคม	7561d9062	ใต้	หมู่ 9 ท่ายาง	คลองขนาน	เหนือคลอง	กระบี่	KBI-SR669802	172.18.170.50	2284.20	35877.69	159	69
2563	มีนาคม	7561d9002	ใต้	หมู่ 8 สะพ้อนกลวง	ดินอุ่น	เขาพนม	กระบี่	KBI-SR669643	172.18.168.202	410.87	12337.82	72	17
2563	มีนาคม	7561d9042	ใต้	หมู่ 8 บ้านควน	ห้วยน้ำขาว	คลองพ่อแก้	กระบี่	KBI-SR669692	172.18.169.66	2363.48	55677.63	286	57
2563	มีนาคม	7568d9045	ใต้	หมู่ 8 บางใหญ่	คีรีวง	ปลายพระยา	กระบี่	KBI-SR669729	172.18.169.154	549.60	18417.80	61	10
2563	มีนาคม	7561d9038	ใต้	หมู่ 8 บางตรามใต้	คลองพ่อแก้เหนือ	คลองพ่อแก้	กระบี่	KBI-SR669681	172.18.168.106	522.87	15920.56	40	26
2563	มีนาคม	7561d9058	ใต้	หมู่ 8 ไนทับ	ห้วยยูง	เหนือคลอง	กระบี่	KBI-SR669783	172.18.170.26	860.88	25602.59	104	31
2563	มีนาคม	7561d9034	ใต้	หมู่ 8 คลองพะยวง	พรุดินนา	คลองพ่อแก้	กระบี่	KBI-SR669671	172.18.169.122	2525.78	70981.29	205	74
2563	มีนาคม	7561d9089	ใต้	หมู่ 8 เขาวิทยลัด	หน้าเขา	เขาพนม	กระบี่	KBI-SR772202	172.18.170.202	1389.35	35615.94	150	59
2563	มีนาคม	7561d9025	ใต้	หมู่ 8 เขาประ	พรุเตียว	เขาพนม	กระบี่	KBI-SR669648	172.18.168.234	388.19	8518.63	23	11
2563	มีนาคม	7561d9055	ใต้	หมู่ 7 ไสสด	ดินอุดม	ลำทับ	กระบี่	KBI-SR669770	172.18.168.138	2613.97	48362.57	163	37

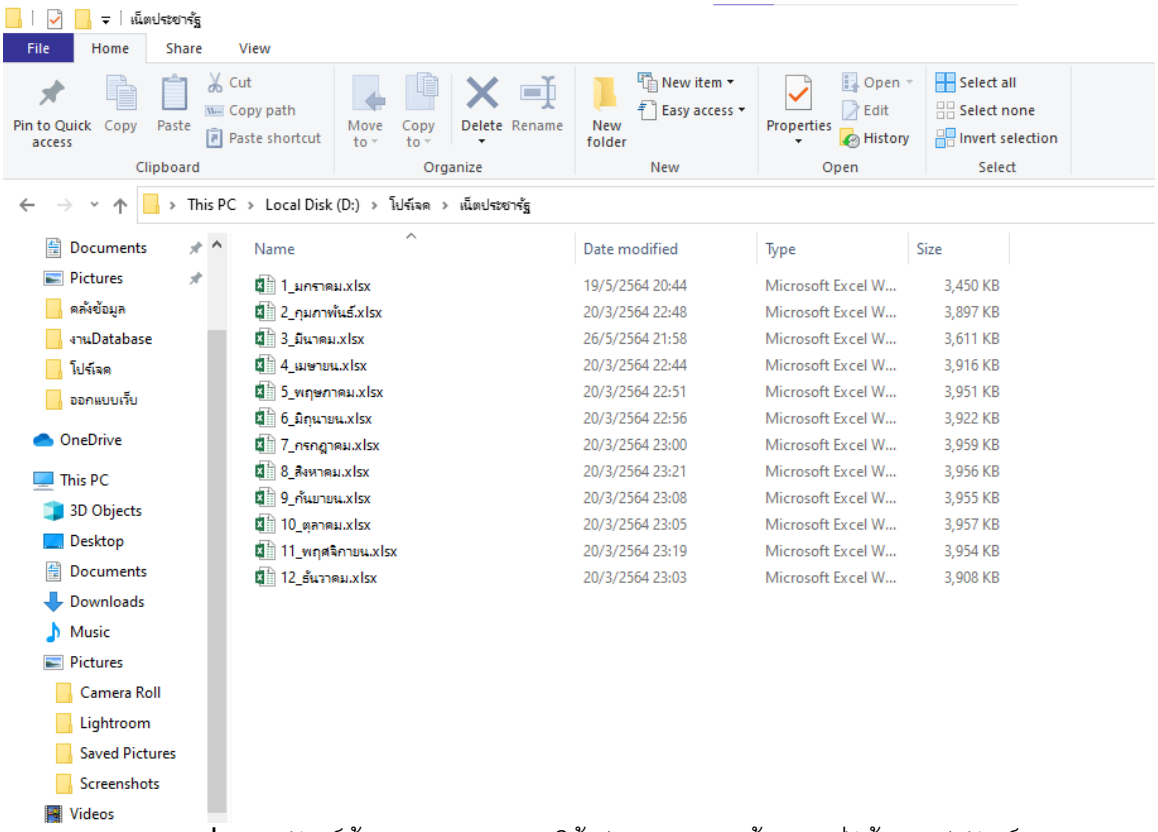
ภาพที่ 3.15 ไฟล์ข้อมูลการบันทึกปริมาณการใช้เน็ตประชารัฐรายเดือน

ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมา ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) มักใช้เวลาค่อนข้างมากระบบการรับข้อมูลป้อนเข้าสู่ระบบในปัจจุบันจะลดการคีย์ข้อมูลจากคนให้น้อยที่สุด แต่จะใช้วิธีการสแกน การติ๊กเลือก เพื่อลดความผิดพลาดให้น้อยที่สุด เพราะขั้นตอนใช้เวลามาก การลดข้อผิดพลาดของข้อมูลได้มากเท่าใด ก็จะยิ่งมีประสิทธิภาพมากขึ้นเท่านั้น

3.1.3 การเตรียมข้อมูล (Data Preparation)

ผู้วิเคราะห์ข้อมูลทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวม และทำการจัดการข้อมูลที่ได้ให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ เพื่อลดความผิดพลาดการทำข้อมูล ให้ถูกต้อง โดยใช้กระบวนการ data cleaning เพื่อให้ได้ข้อมูลอยู่ในรูปแบบและมาตรฐานเดียวกัน และกำจัดข้อมูลเสียออก ผู้วิเคราะห์ข้อมูลจึงได้ดำเนินการตามขั้นตอนดังนี้

1.) การปรับมาตรฐาน (Standardizing) ผู้วิเคราะห์ข้อมูลทำการรวบแปลงไฟล์ข้อมูลให้เป็นไฟล์เดียวกันอยู่ในรูปแบบเดียวกันพร้อมทั้งสร้างแอททริบิวต์เพื่อกำหนดคีย์หลักของแต่ละไฟล์ขึ้นมาเพื่อเป็นตัวที่ใช้ระบุความสัมพันธ์ของข้อมูลที่จะแสดงปริมาณการใช้เน็ตประชารัฐ



ภาพที่ 3.16 ไฟล์ข้อมูลปริมาณการใช้เน็ตประชารัฐทั้งหมดที่ได้จากเว็บไซต์

ผู้วิเคราะห์ได้ทำการจัดการข้อมูลให้อยู่ในรูปแบบเดียวกัน โดยการรวบรวมข้อมูลจากแต่ละหมู่บ้านให้เป็นรายเดือน เพื่อให้ข้อมูลรายเดือนไปพยากรณ์ตามวัตถุประสงค์ที่กำหนดไว้ข้างต้น

เดือน	ค่าปริมาณการนำข้อมูลเข้า
มกราคม	416,199,895.26
กุมภาพันธ์	246,615,350.16
มีนาคม	249,902,165.12
เมษายน	246,650,401.64
พฤษภาคม	251,511,850.08
มิถุนายน	298,773,716.62
กรกฎาคม	324,707,757.20
สิงหาคม	315,575,334.24
กันยายน	300,869,388.04
ตุลาคม	291,106,942.02
พฤศจิกายน	303,540,472.40
ธันวาคม	325,759,936.33

ภาพที่ 3.17 ค่าปริมาณการนำข้อมูลเข้า

เดือน	ค่าปริมาณการนำข้อมูลออก
มกราคม	6,313,602,697.54
กุมภาพันธ์	4,914,796,724.42
มีนาคม	4,926,267,733.02
เมษายน	4,956,912,229.44
พฤษภาคม	5,031,792,836.91
มิถุนายน	5,859,709,399.02
กรกฎาคม	6,165,636,987.63
สิงหาคม	6,109,337,267.87
กันยายน	5,764,394,413.90
ตุลาคม	5,509,308,458.35
พฤศจิกายน	5,702,892,020.92
ธันวาคม	5,578,847,057.00

ภาพที่ 3.18 ค่าปริมาณการนำข้อมูลออก

2.) การจัดหมวดหมู่ (Transfrom) ผู้วิเคราะห์นำ dataset ที่ผ่านการรวบรวมรายเดือนมาจัดหมวดหมู่เพื่อความถูกต้อง ผู้วิเคราะห์ดำเนินการจัดกลุ่มข้อมูลแบ่งตามภูมิภาค ดังนี้

2.1 การจัดกลุ่มตามภูมิภาค

REGION	VILLAGE	SUBDISTRICT	DISTRICT	PROVINCE
ภาคใต้	หมู่ 2 คลองหมาก	เกาะลันตาน้อย	เกาะลันตา	กระบี่
ภาคตะวันออกเฉียงเหนือ	หมู่ 13 บ้านมะเกลือ	บ้านกอก	จัตุรัส	ชัยภูมิ
ภาคตะวันออกเฉียงเหนือ	หมู่ 17 บ้านโนนทอง	บ้านกอก	จัตุรัส	ชัยภูมิ
ภาคตะวันออกเฉียงเหนือ	หมู่ 1 เพชร	บ้านเพชร	ภูเขียว	ชัยภูมิ
ภาคตะวันออกเฉียงเหนือ	หมู่ 8 เพชรใต้	บ้านเพชร	ภูเขียว	ชัยภูมิ
ภาคตะวันออกเฉียงเหนือ	หมู่ 9 กุดตุม	บ้านเพชร	ภูเขียว	ชัยภูมิ
ภาคตะวันออกเฉียงเหนือ	หมู่ 16 โคกสะอาด	โคกสะอาด	ภูเขียว	ชัยภูมิ
ภาคตะวันออกเฉียงเหนือ	หมู่ 3 นางแม้ง	โพนทอง	เมืองชัยภูมิ	ชัยภูมิ
ภาคตะวันออกเฉียงเหนือ	หมู่ 4 หนองหญ้ารังกา	โพนทอง	เมืองชัยภูมิ	ชัยภูมิ
ภาคตะวันออกเฉียงเหนือ	หมู่ 5 หนองคอนไทย	กุดตุ้ม	เมืองชัยภูมิ	ชัยภูมิ
ภาคตะวันออกเฉียงเหนือ	หมู่ 14 หนองคอนไทย	กุดตุ้ม	เมืองชัยภูมิ	ชัยภูมิ
ภาคเหนือ	หมู่ 8 อมเม็ง	กองแขก	แม่แจ่ม	เชียงใหม่
ภาคเหนือ	หมู่ 4 นาฟ่อน	บ่อหลวง	ฮอด	เชียงใหม่
ภาคเหนือ	หมู่ 12 นางยางคิน	กองแขก	แม่แจ่ม	เชียงใหม่
ภาคเหนือ	หมู่ 7 สันป่าตึง	แม่ระมาด	แม่ระมาด	ตาก
ภาคใต้	หมู่ 2 ควนแร่	นาไม้ไผ่	ทุ่งสง	นครศรีธรรมราช
ภาคเหนือ	หมู่ 8 จุนพัฒนา	จุน	จุน	พะเยา
ภาคเหนือ	หมู่ 13 ค้างทงษ์ใหม่	จุน	จุน	พะเยา

ภาพที่ 3.19 ข้อมูลในแต่ละไฟล์ที่ต้องการจัดกลุ่มให้รวมอยู่ในภาคเดียวกัน

2.2 นำข้อมูลที่ผ่านการรวบรวมเป็นรายเดือนมาไว้ในไฟล์ภูมิภาคเดียวกัน

REGION	VILLAGE	SUBDISTRICT	DISTRICT	PROVINCE
ใต้	หมู่ 9 หน้าค่าย	คลองท่อมใต้	คลองท่อม	กระบี่
ใต้	หมู่ 9 ไสยง	สินปุน	เขาพนม	กระบี่
ใต้	หมู่ 9 วังน้ำเย็น	พรุเตียว	เขาพนม	กระบี่
ใต้	หมู่ 9 ป่ากบางสร้าน	เขาดิน	เขาพนม	กระบี่
ใต้	หมู่ 9 ทุ่งควนหิน	ห้วยน้ำขาว	คลองท่อม	กระบี่
ใต้	หมู่ 9 ท่ายาง	คลองขนาน	เหนือคลอง	กระบี่
ใต้	หมู่ 8 สะท้อนกลวง	สินปุน	เขาพนม	กระบี่
ใต้	หมู่ 8 บ้านควน	ห้วยน้ำขาว	คลองท่อม	กระบี่
ใต้	หมู่ 8 บางใหญ่	คีรีวง	ปลายพระยา	กระบี่
ใต้	หมู่ 8 บางตรามใต้	คลองท่อมเหนือ	คลองท่อม	กระบี่
ใต้	หมู่ 8 ในทับ	ห้วยยูง	เหนือคลอง	กระบี่
ใต้	หมู่ 8 คลองพะเนาว	พรุดินนา	คลองท่อม	กระบี่
ใต้	หมู่ 8 เขาวังพลัด	หน้าเขา	เขาพนม	กระบี่
ใต้	หมู่ 8 เขาประ	พรุเตียว	เขาพนม	กระบี่
ใต้	หมู่ 7 ไสสด	ดินอุดม	ลำทับ	กระบี่
ใต้	หมู่ 7 ย่านอุดม	ลำทับ	ลำทับ	กระบี่
ใต้	หมู่ 7 พรุพี	ทรายขาว	คลองท่อม	กระบี่

ภาพที่ 3.20 แสดงการจัดกลุ่มภาคใต้

3.1.4 การสร้างแบบจำลอง (Modeling)

ผู้วิเคราะห์ข้อมูลได้วิเคราะห์ด้วยเทคนิคทางสถิติ ทำการเรียนรู้จากข้อมูลเดิมของการใช้เน็ตประชารัฐเพื่อให้ได้คำตอบที่ดีที่สุด การเปรียบเทียบความคลาดเคลื่อนของวิธีการพยากรณ์ โดยนำโมเดล Time-Series Forecasting มาช่วยในการหาค่าความคลาดเคลื่อนของการพยากรณ์ในวิธีต่าง Moving Average Method โดยนำมาเปรียบเทียบกับโปรแกรมว่าโปรแกรมใดให้ค่าความถูกต้องแม่นยำมากที่สุดนำมาใช้ในการพยากรณ์ข้อมูลการใช้เน็ตประชารัฐ การพยากรณ์ในรูปแบบคณิตศาสตร์ใช้รูปแบบสมการทางคณิตศาสตร์ ผู้วิเคราะห์ข้อมูลได้เลือกใช้เทคนิครูปแบบอนุกรมเวลา Time series Models ด้วย ชุดข้อมูลปริมาณการใช้เน็ตประชารัฐในแต่ละเดือน โดยมีขั้นตอนการดำเนินการดังนี้

$$MA_n = \frac{\sum A_t}{n}$$

เมื่อกำหนด t = ช่วงเวลา

n = จำนวนข้อมูลที่จะคำนวณค่าเฉลี่ยเคลื่อนที่

A_n = ข้อมูลจริง ณ ช่วงเวลา t

MA_n = ค่าพยากรณ์วิธีค่าเฉลี่ยเคลื่อนที่ n ช่วงเวลา

month	region	province	upload	download
มีนาคม	ภาคกลาง	สมุทรสาคร	15891.43	338042.86
มีนาคม	ภาคกลาง	ชัยนาท	24945.85	379759.95
มีนาคม	ภาคกลาง	ชัยนาท	9795.69	43250.40
มีนาคม	ภาคกลาง	ชัยนาท	749.80	23474.40
มีนาคม	ภาคกลาง	ชัยนาท	8949.73	136736.25
มีนาคม	ภาคกลาง	ชัยนาท	29920.32	465428.02
มีนาคม	ภาคกลาง	ชัยนาท	2237.95	46943.98
มีนาคม	ภาคกลาง	ชัยนาท	4441.76	129851.15
มีนาคม	ภาคกลาง	ชัยนาท	15509.22	340718.16
มีนาคม	ภาคกลาง	ชัยนาท	4382.67	129552.31
มีนาคม	ภาคกลาง	ชัยนาท	13760.55	352962.03
มีนาคม	ภาคกลาง	ชัยนาท	17455.50	420001.80
มีนาคม	ภาคกลาง	ชัยนาท	14175.43	181067.38

ภาพที่ 3.21 ข้อมูลปริมาณเน็ตประชารัฐ

จากที่รูปภาพที่ 3.19 ประกอบไปด้วย 5 แอตทริบิวต์ คือ

1.) Month เก็บข้อมูลเดือน

2.) Region เก็บข้อมูลภูมิภาค

3.) Province เก็บข้อมูลจังหวัด

4.) Upload เก็บปริมาณนำข้อมูลเข้า

5.) Download เก็บปริมาณนำข้อมูลออก

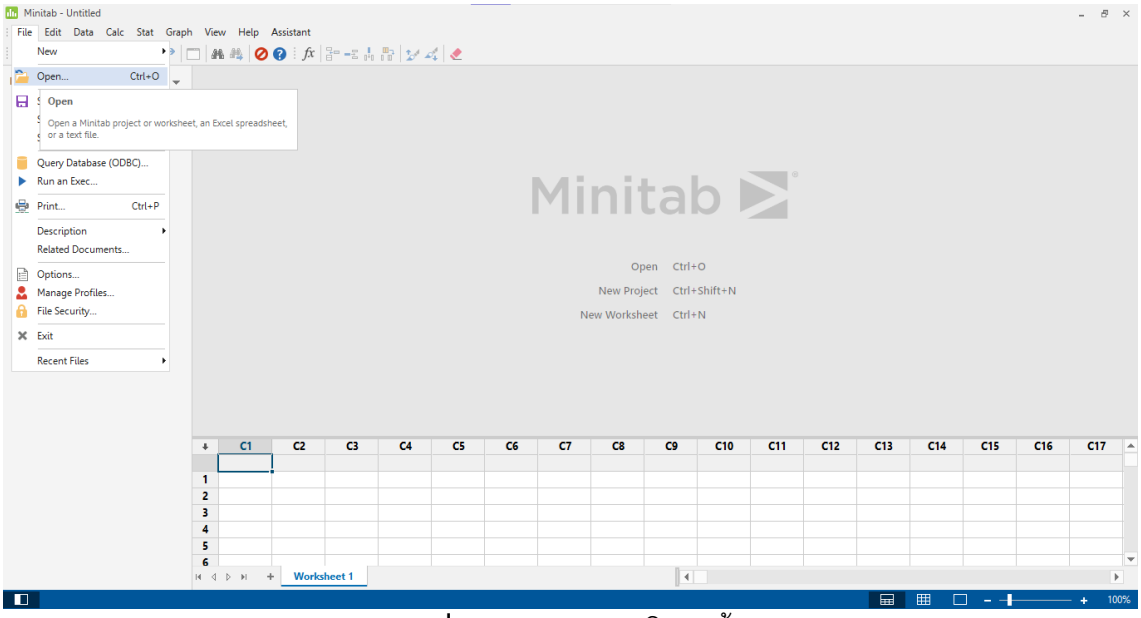
3.1.5 การประเมินประสิทธิภาพ (Evaluation)

เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไปใช้งาน ผู้วิเคราะห์ข้อมูลวัดประสิทธิภาพของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรกว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใดโดยเปรียบเทียบนำผลที่ได้สามารถวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ โดยผู้วิเคราะห์ข้อมูลได้เปรียบเทียบนำผลที่ได้ ดังนี้

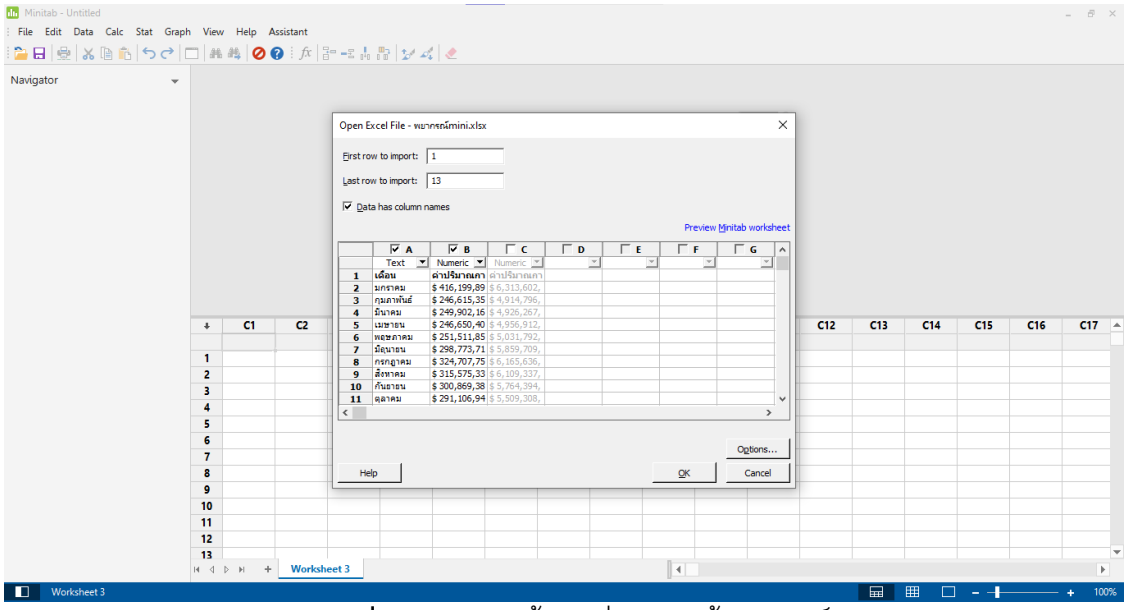
การสร้างโมเดล Time Series Models จะทำการเลือกแอททริบิวต์ข้อมูลเชิงปริมาณที่จัดเก็บไว้ มาคำนวณหาค่าความเป็นไปได้เชิงพยากรณ์ โดยการคำนวณจากทางคณิตศาสตร์ ดังนี้

1) การใช้เทคนิคการวิเคราะห์อนุกรมเวลา ด้วยวิธี Moving Average โดยผลลัพธ์จากการพยากรณ์ผู้วิเคราะห์ข้อมูลได้ข้อมูลที่ผ่านการวิเคราะห์ข้อมูล จากการสร้างโมเดล Time Series Models จึงนำข้อมูลดังกล่าว มาทดลองกับโปรแกรม Minitab ซึ่งมีขั้นตอนการทำงานดังนี้

ขั้นตอนที่ 1 ทำการ Upload โดยเลือกคำสั่ง File → open → เลือกไฟล์ข้อมูลที่เตรียมไว้

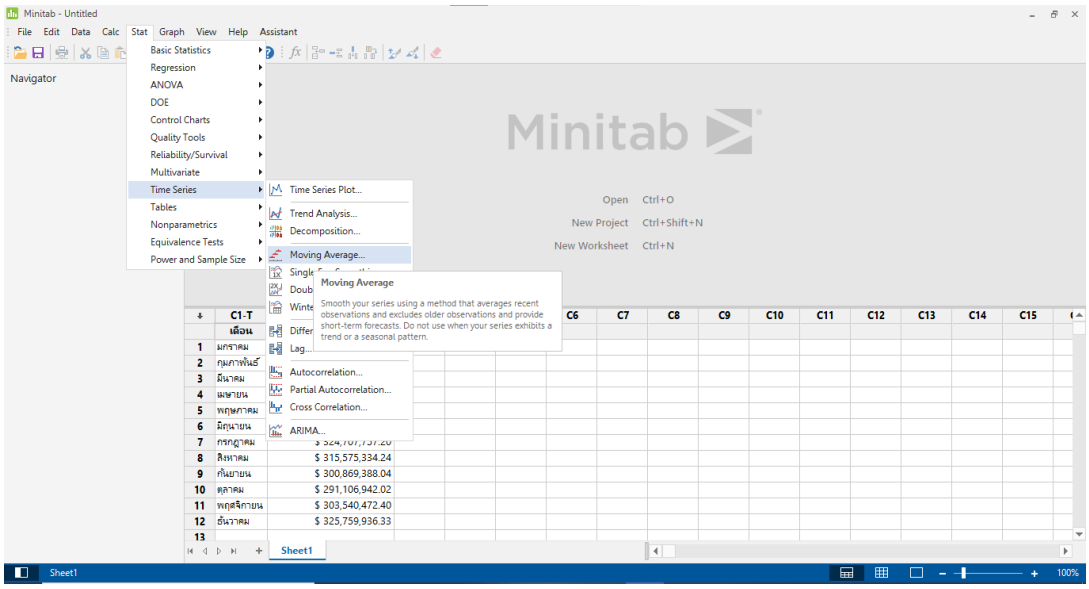


ภาพที่ 3.22 แสดงการโหลดข้อมูล



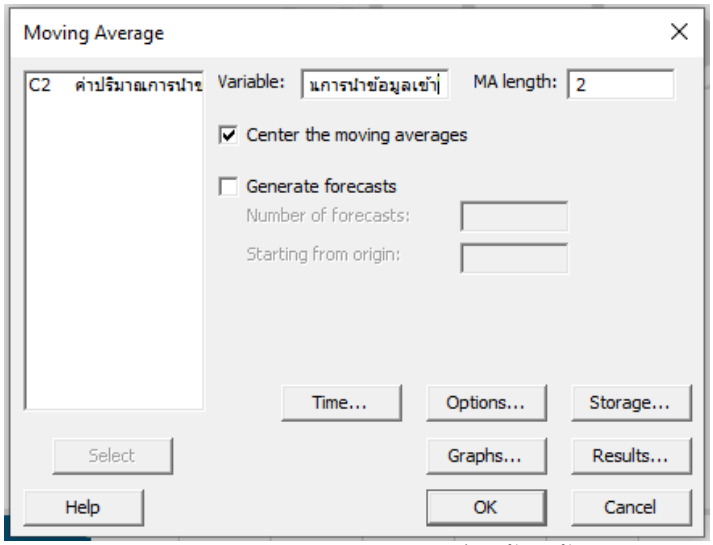
ภาพที่ 3.23 แสดงข้อมูลที่เตรียมไว้พยากรณ์

ขั้นตอนที่ 2 ทำการเลือกรูปแบบวิธีที่ต้องการจะนำข้อมูลไปทำการวิเคราะห์ (Moving Average) เลือกคำสั่ง Stat → Time Series → Moving Average



ภาพที่ 3.24 แสดงรูปแบบวิธีที่ต้องการจะนำข้อมูลไปทำการวิเคราะห์

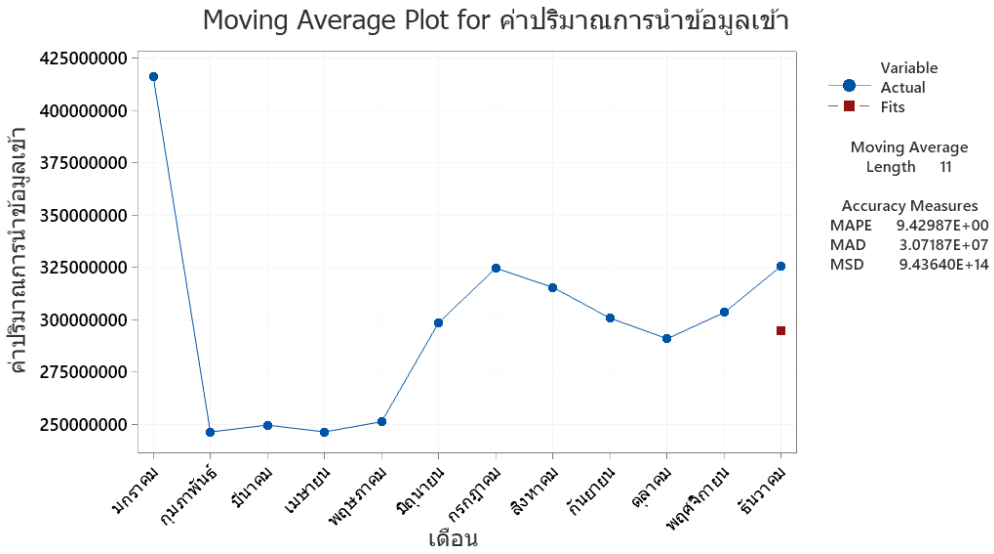
ขั้นตอนที่ 3 ทำการกำหนดค่าให้กับข้อมูลก่อนนำข้อมูลเข้าทำการพยากรณ์ โดยการ แทนค่า Variable = ค่าปริมาณการนำข้อมูลเข้า (ตัวแปรที่ต้องการนำมาคำนวณ) และ MA length=2 (ค่าพยากรณ์ล่วงหน้า) แล้วคลิกตกลง



ภาพที่ 3.25 แสดงการกำหนดค่าให้กับข้อมูล

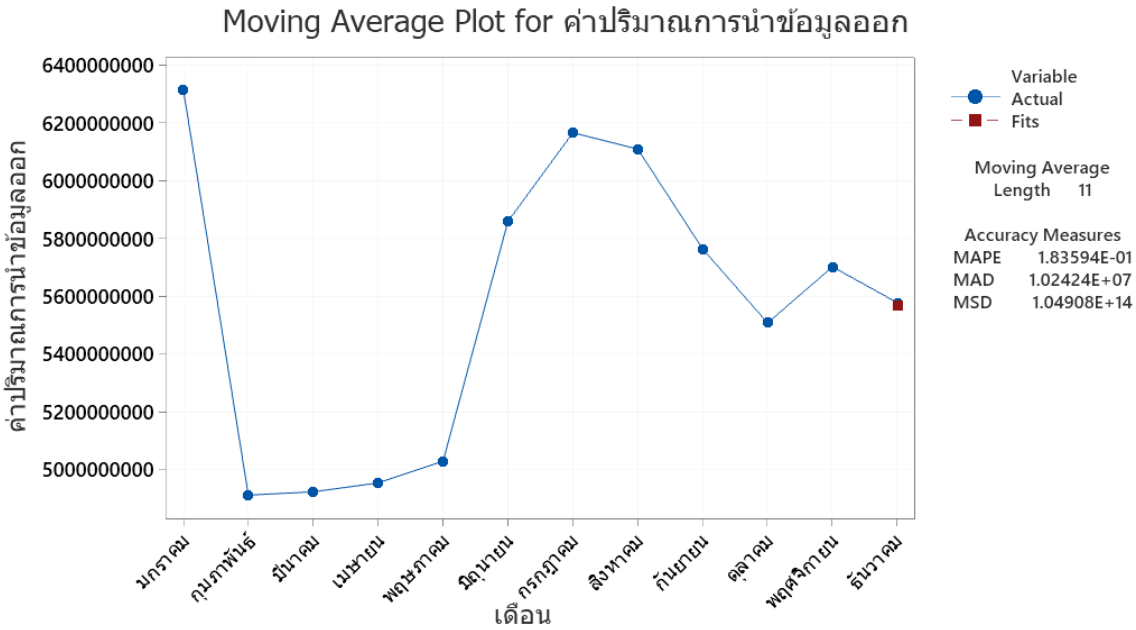
ขั้นตอนที่ 4 การพยากรณ์แล้วทำการแทนค่าสมการตามสูตรอนุกรมเวลาวิธีการค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก แล้วได้ผลลัพธ์ดังภาพ

4.1 ปริมาณการนำข้อมูลเข้า



ภาพที่ 3.26 ตัวอย่างกราฟแสดงผลการพยากรณ์ของปริมาณการนำข้อมูลเข้า

4.2 ปริมาณการนำข้อมูลออก



ภาพที่ 3.27 กราฟแสดงผลการพยากรณ์ของปริมาณการนำข้อมูลออก

จากค่าปริมาณการนำข้อมูลเข้าภาพที่ 3.25 และค่าปริมาณการนำข้อมูลออกภาพที่ 3.26,

เส้นสีแดง หมายถึงค่าพยากรณ์

เส้นสีน้ำเงิน หมายถึงค่าจริง

เกณฑ์การเปรียบเทียบประสิทธิภาพในการพยากรณ์

ในการศึกษาครั้งนี้ผู้วิเคราะห์ ได้เปรียบเทียบการพยากรณ์โดย พิจารณาจากค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute PercentageError, MAPE) เป็นวิธีวัดความแม่นยำโดยคำนวณเปอร์เซ็นต์ความผิดพลาดในการพยากรณ์

$$MAPE = \frac{\sum |\text{ค่าเกิดขึ้นจริง} - \text{ค่าพยากรณ์}| \div \text{ค่าเกิดขึ้นจริง}}{\text{n}} \times 100$$

การพยากรณ์โดยการวิเคราะห์อนุกรมเวลาของการทำเหมืองข้อมูล ผ่านโปรแกรมโดยผลลัพธ์ที่ได้จากการพยากรณ์ล่วงหน้าและหลักเกณฑ์ในการเลือกค่าตรวจสอบนั้นมีหลักเกณฑ์ปฏิบัติคือ ค่า MAPE ส่วนใหญ่แล้วใช้เปรียบเทียบความแม่นยำของค่าพยากรณ์ของข้อมูลทางธุรกิจ ทั้งนี้ค่า MAPE เป็นค่าวัดเปรียบเทียบดังนั้นค่า MAPE จะนิยมกว่าค่า MADซึ่งจากการศึกษาพบว่าถ้าค่า MAPE ที่สามารถเชื่อถือได้

ตารางที่ 3.18 ตารางเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE)

ค่า MAPE	**ความน่าเชื่อถือ**
น้อยกว่า 10%	จัดว่าการพยากรณ์ค่อนข้างแม่นยำ
10% ถึง 20%	จัดว่าการพยากรณ์ใช้ได้ดี
20% ถึง 50%	จัดว่าการพยากรณ์พอใช้
มากกว่า 50%	จัดว่าการพยากรณ์ไม่แม่นยำ

ตัวอย่าง ผู้วิเคราะห์ใช้แกรม Excel สร้างตารางและคอลัมน์ นำข้อมูลจริงและค่าพยากรณ์ล่วงหน้า 3เดือน มาคำนวณหาค่า MADก่อนและคำนวณหาค่า MAPE เพื่อหาวิธีที่มีความแม่นยำมากที่สุด สำหรับการพยากรณ์ล่วงหน้า ดังนี้

สูตรในการหาค่า MAD คือ

$$MAD = \frac{\sum |\text{ค่าเกิดขึ้นจริง} - \text{ค่าพยากรณ์}|}{n}$$

	ค่าปริมาณการนำข้อมูลเข้า	ค่าพยากรณ์ล่วงหน้า 3 เดือน	MAD	MAPE
มกราคม	416,199,895.26			
กุมภาพันธ์	246,615,350.16			
มีนาคม	249,902,165.12			
เมษายน	246,650,401.64	304,239,136.85	57,588,735.21	23.3%
พฤษภาคม	251,511,850.08	247,722,638.97	3,789,211.11	1.5%
มิถุนายน	298,773,716.62	249,354,805.61	49,418,911.01	16.5%
กรกฎาคม	324,707,757.20	265,645,322.78	59,062,434.42	18.2%
สิงหาคม	315,575,334.24	291,664,441.30	23,910,892.94	7.6%
กันยายน	300,869,388.04	313,018,936.02	12,149,547.98	4.0%
ตุลาคม	291,106,942.02	313,717,493.16	22,610,551.14	7.8%
พฤศจิกายน	303,540,472.40	302,517,221.43	1,023,250.97	0.3%
ธันวาคม	325,759,936.33	298,505,600.82	27,254,335.51	8.4%
			ค่าMAPE	9.7%

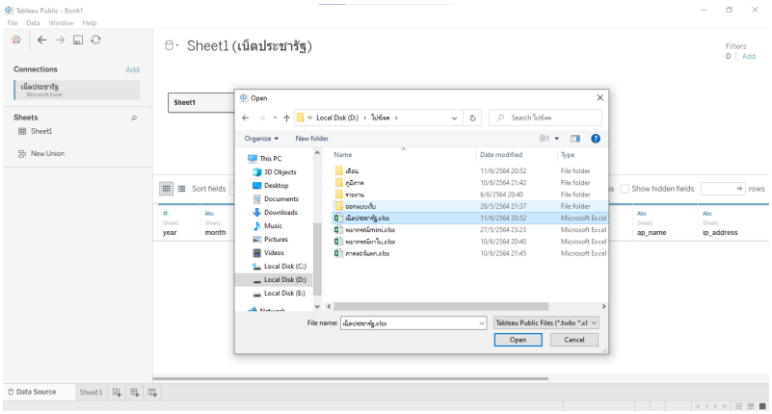
ภาพที่ 3.28 ยกตัวอย่างค่าพยากรณ์ล่วงหน้า 3เดือน ในการหาค่าMAPE

3.1.6 การนำผลลัพธ์ไปใช้งาน (Deployment)

ผู้วิเคราะห์ข้อมูลนำผลองค์ความรู้ที่ได้เหล่านี้ไปนำเสนอข้อมูลแบบ Visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public โดยดำเนินการตามขั้นตอน ดังนี้

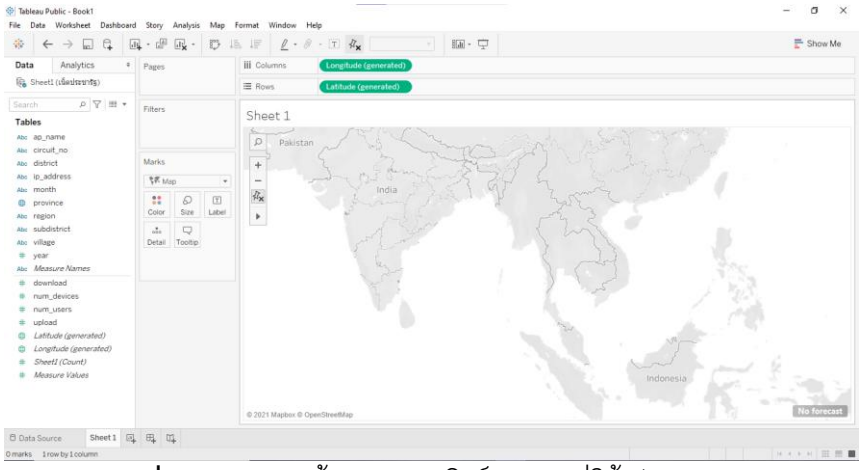
1. การนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public

1.1) ทำการ Import ข้อมูลปริมาณการใช้เน็ตประชารัฐเข้าโปรแกรม Tableau Public



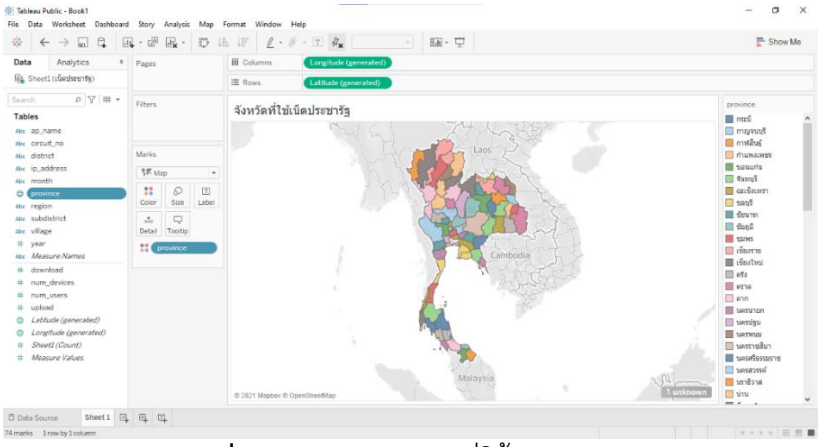
ภาพที่ 3.29 Import ข้อมูลปริมาณการใช้เน็ตประชารัฐเพื่อหาผลลัพธ์

1.2) กำหนด Columns เป็น Longitude กำหนด Rows เป็น Latitude



ภาพที่ 3.30 แสดงขั้นตอนการโชว์จังหวัดที่ใช้เน็ตประชารัฐบน Map

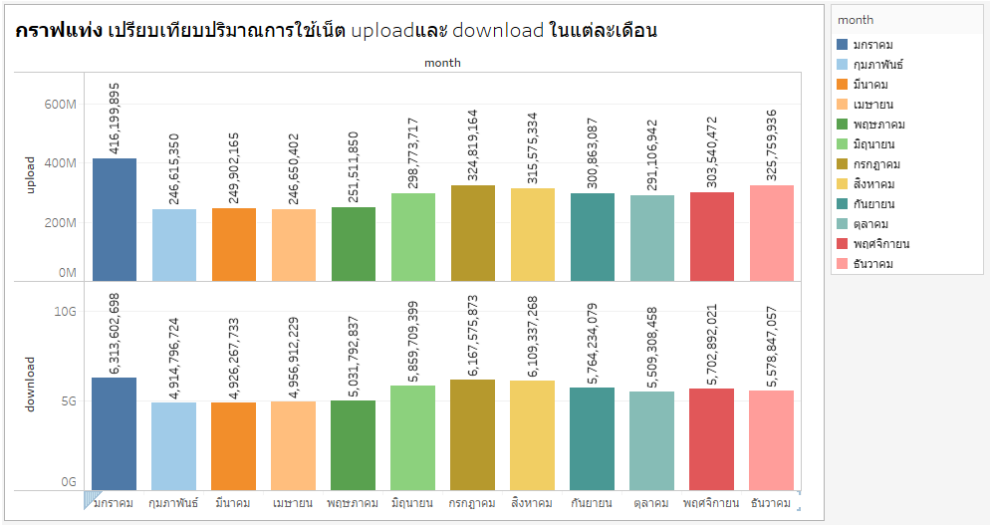
1.3) กำหนด province ใส่ใน Color



ภาพที่ 3.31 แสดงจังหวัดที่ใช้เน็ตประชารัฐ

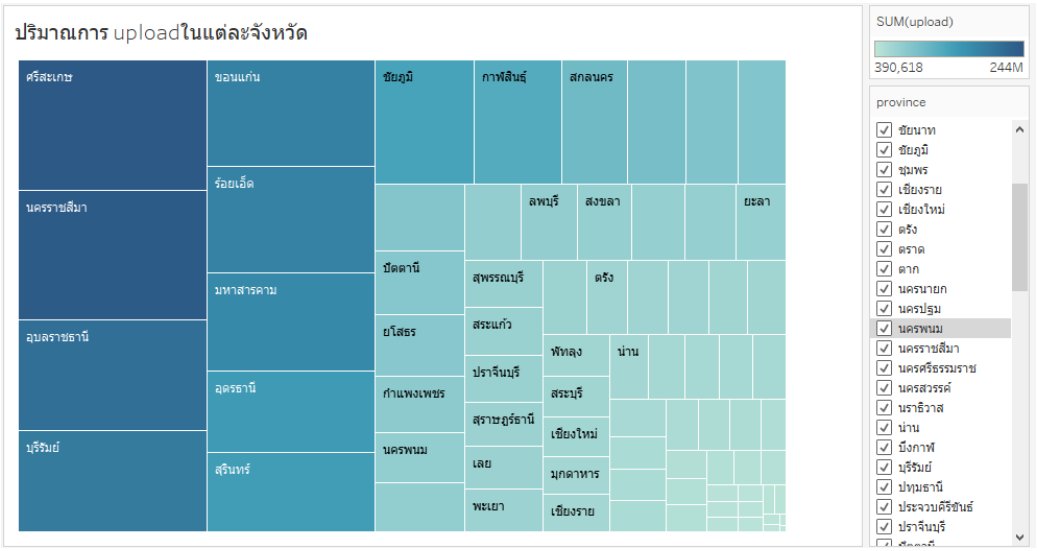
2. นำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้การแสดงข้อมูลโดยใช้ highcharts.js แสดงข้อมูลแบบแดชบอร์ด โดยดำเนินการตามขั้นตอนดังนี้

2.1) กราฟแท่งเปรียบเทียบปริมาณการใช้เน็ต uploadและ download ในแต่ละเดือน



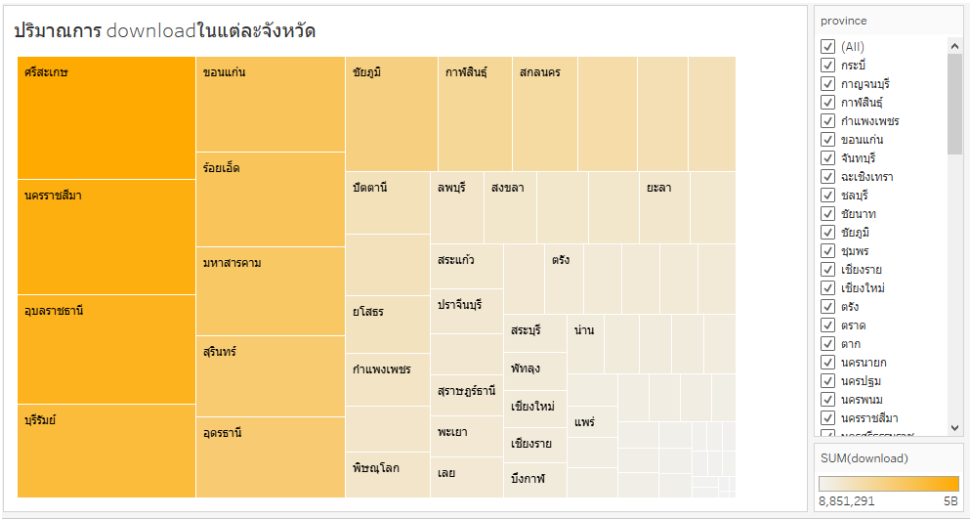
ภาพที่ 3.32 แสดงการเปรียบเทียบปริมาณการใช้เน็ต uploadและ download ในแต่ละเดือน

2.2) แผนที่กราฟ แสดงปริมาณการ uploadในแต่ละจังหวัด



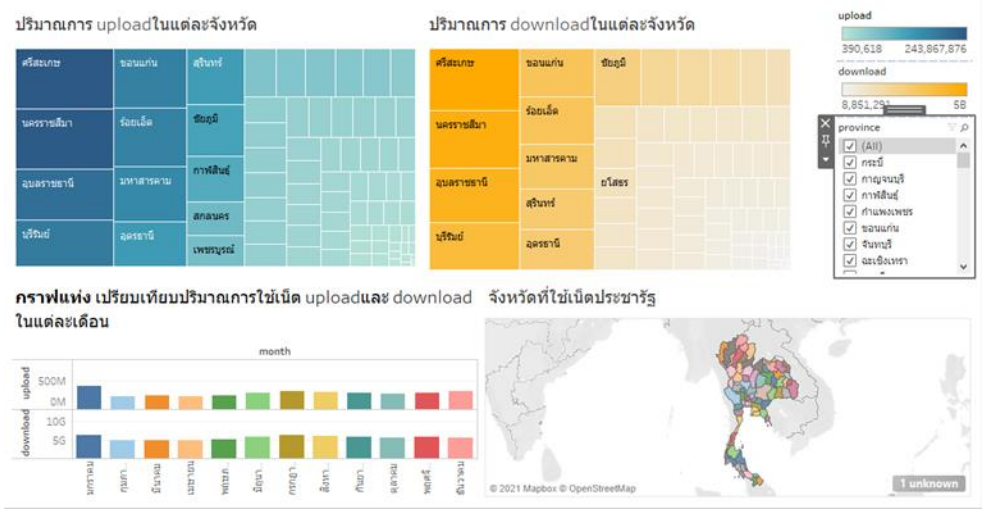
ภาพที่ 3.33 แสดงปริมาณการ uploadในแต่ละจังหวัด

2.3) แผนที่กราฟ แสดงปริมาณการ downloadในแต่ละจังหวัด



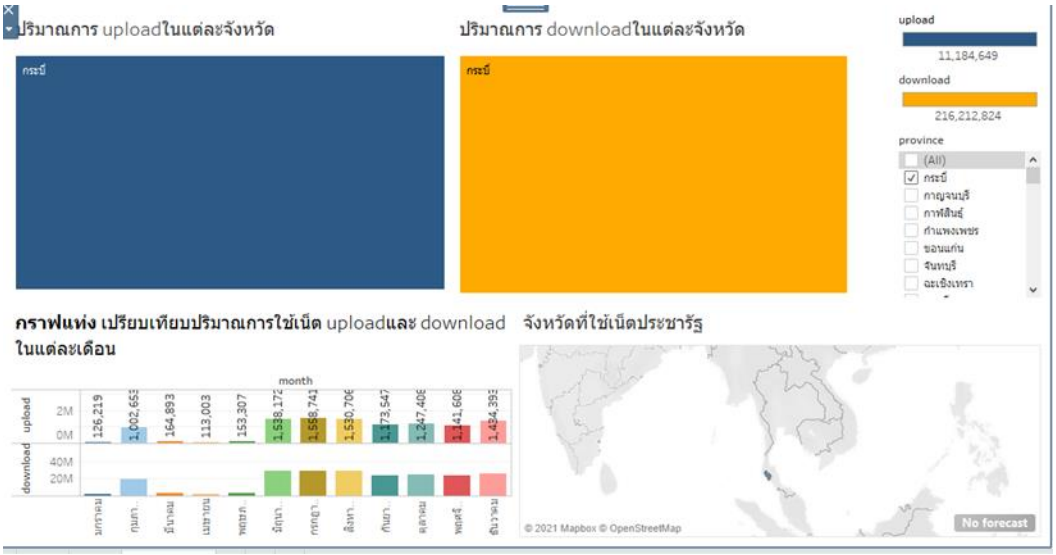
ภาพที่ 3.34 แสดงปริมาณการ downloadในแต่ละจังหวัด

2.4) การแสดงหน้าแดชบอร์ดบนโปรแกรม Tableau Public

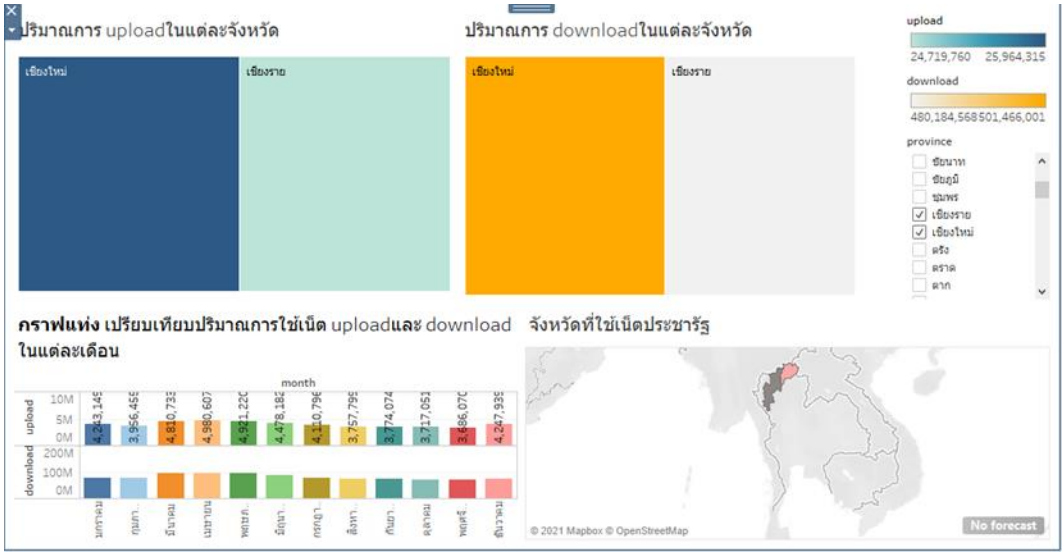


ภาพที่ 3.35 แสดงผลข้อมูลปริมาณการใช้เน็ตประชารัฐ

ตัวอย่าง การแสดงหน้าแดชบอร์ดบนโปรแกรม Tableau Public



ภาพที่3.36 แสดงข้อมูลของจังหวัดกระบี่



ภาพที่3.37 แสดงข้อมูลของจังหวัดเชียงใหม่และเชียงราย

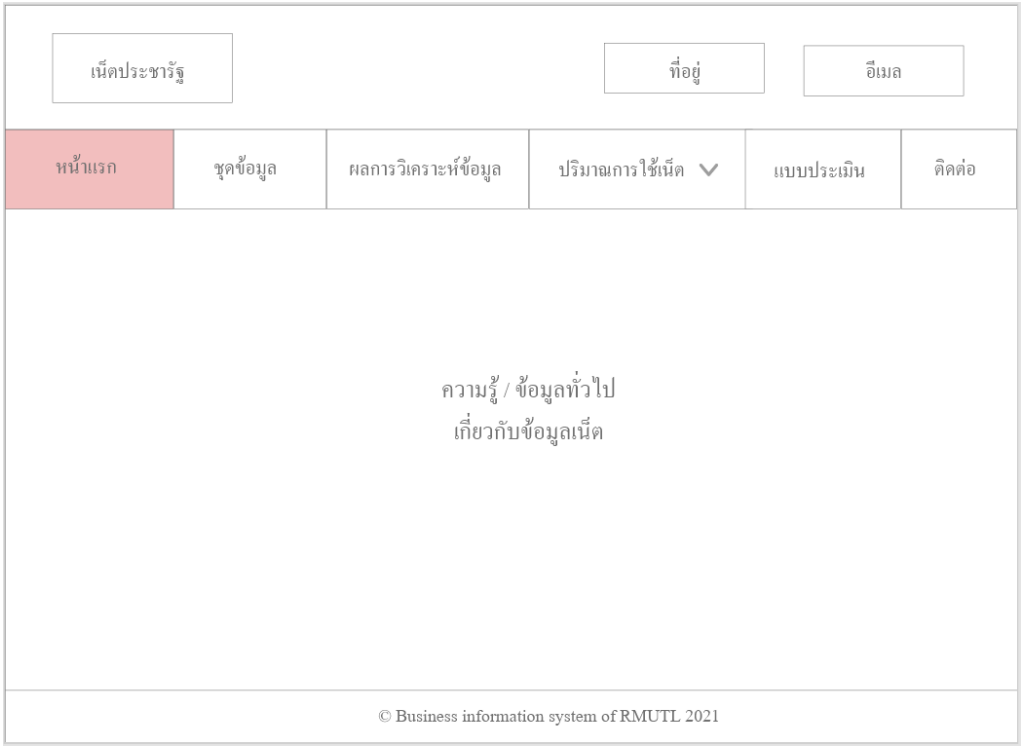
3.7 การออกแบบและพัฒนาเว็บไซต์

การวางแผนการจัดลำดับ เนื้อหาสาระของเว็บไซต์ ออกเป็นหมวดหมู่ เพื่อจัดทำเป็นโครงสร้างในการจัดวางหน้าเว็บเพจทั้งหมด ทำให้เห็นโครงสร้างทั้งหมดของเว็บไซต์และการออกแบบโครงสร้างหรือจัดระเบียบของข้อมูลที่ชัดเจน แยกย่อยเนื้อหาออกเป็นหาออกเป็นส่วนต่าง ๆ ที่สัมพันธ์และให้มีอยู่ในมาตรฐานเดียวกัน จะช่วยให้นำมาใช้งานและง่าย ต่อการเข้าอ่านเนื้อหาของผู้ใช้เว็บไซต์

3.7.1 การออกแบบ Wireframe หน้าจอเว็บไซต์

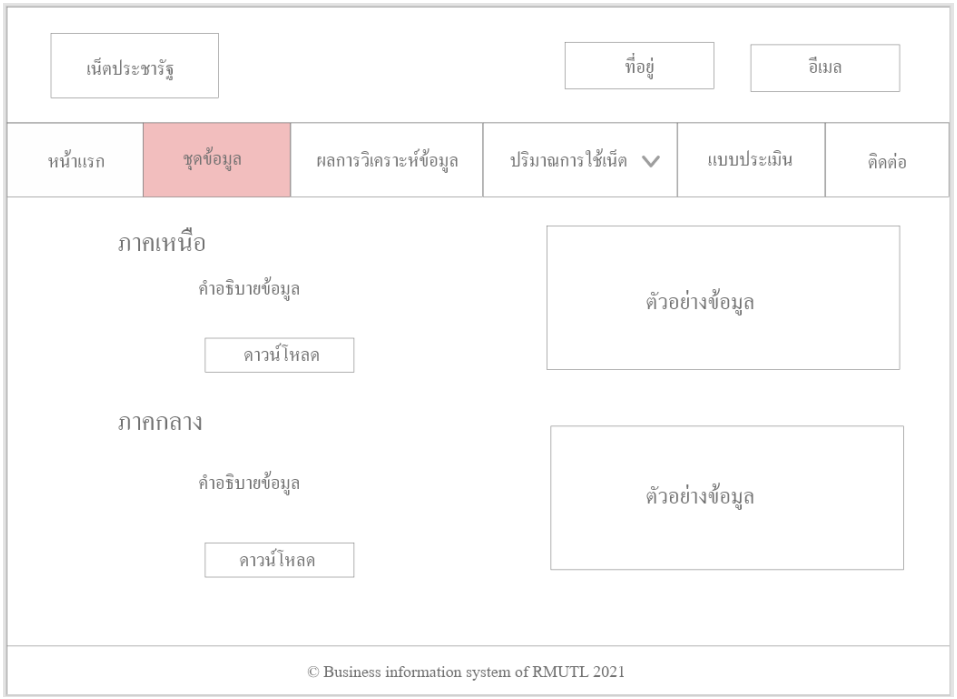
- หน้าเว็บสำหรับผู้ใช้งาน

1.) หน้าแรกของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บไซต์



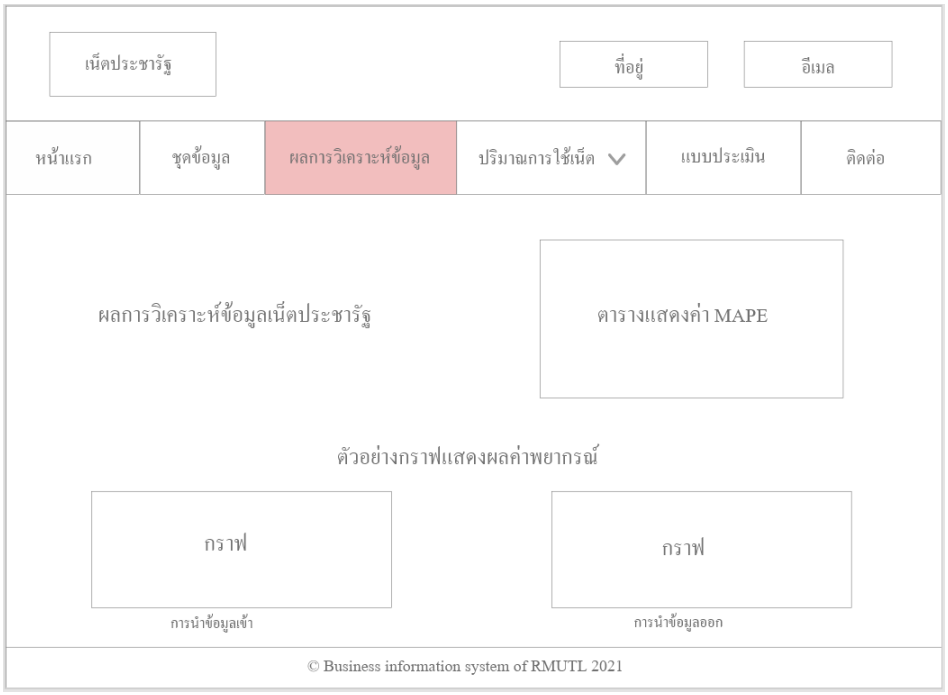
ภาพที่ 3.38 หน้าแรกของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บไซต์

2.) หน้าแสดงชุดข้อมูล



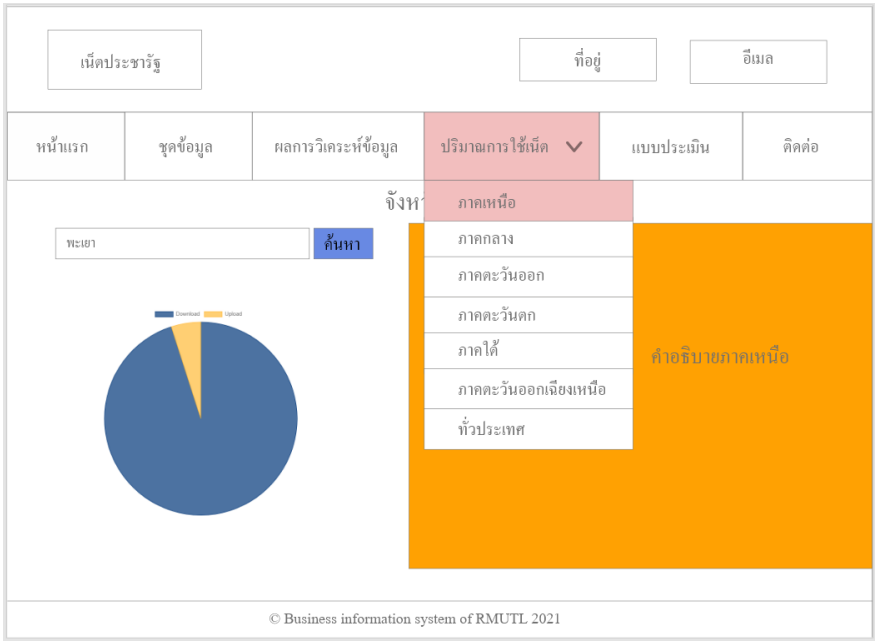
ภาพที่ 3.39 หน้าแสดงชุดข้อมูล

3.) หน้าแสดงผลการวิเคราะห์ข้อมูล



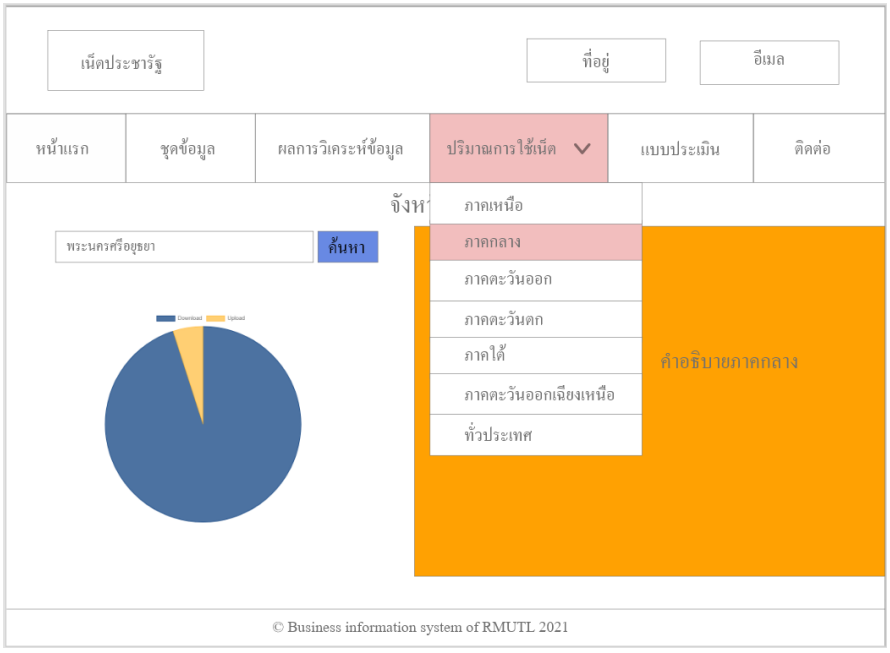
ภาพที่ 3.40 หน้าแสดงผลการวิเคราะห์ข้อมูล

4.) หน้าแสดงปริมาณการใช้เน็ตประชารัฐของภาคเหนือ



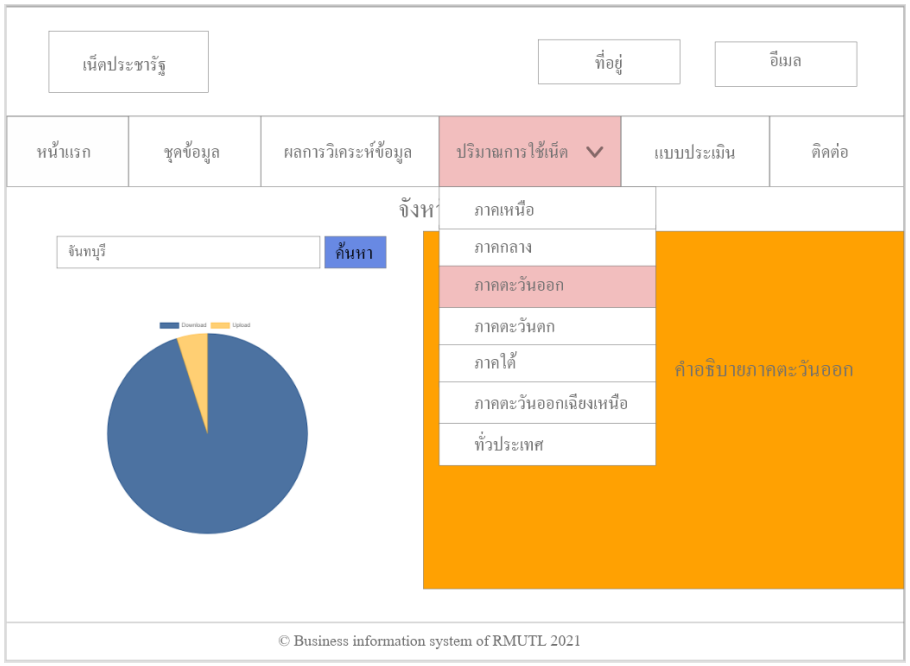
ภาพที่ 3.41 หน้าแสดงปริมาณการใช้เน็ตประชารัฐของภาคเหนือ

5.) หน้าแสดงปริมาณการใช้เน็ตประชารัฐของภาคกลาง



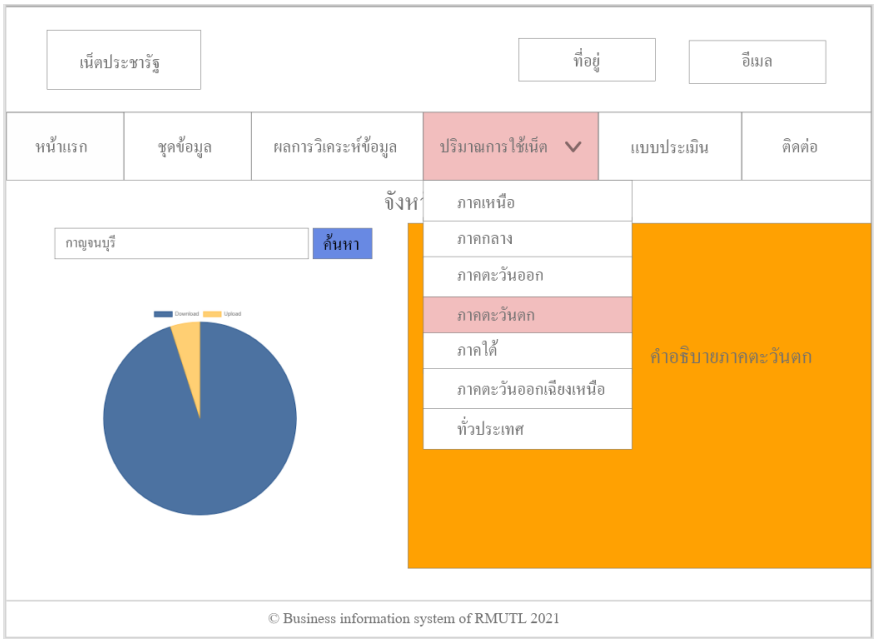
ภาพที่ 3.42 หน้าแสดงปริมาณการใช้เน็ตประชารัฐของภาคกลาง

6.) หน้าแสดงปริมาณการใช้เน็ตประชารัฐของภาคตะวันออก



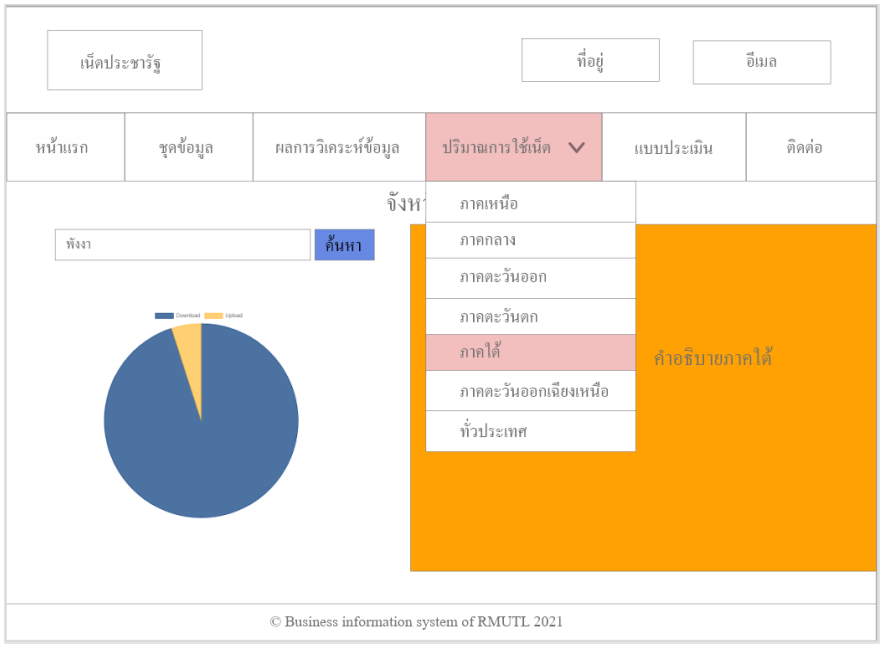
ภาพที่ 3.43 หน้าแสดงปริมาณการใช้เน็ตประชารัฐของภาคตะวันออก

7.) หน้าแสดงปริมาณการใช้เน็ตประชารัฐของภาคตะวันตก



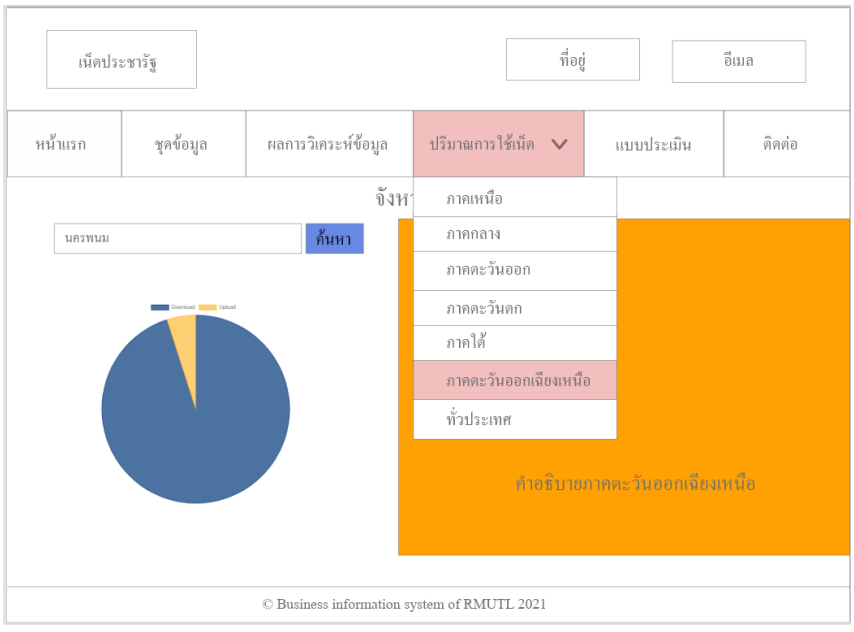
ภาพที่ 3.44 หน้าแสดงปริมาณการใช้เน็ตประชารัฐของภาคตะวันตก

8.) หน้าแสดงปริมาณการใช้เน็ตประชารัฐของภาคใต้



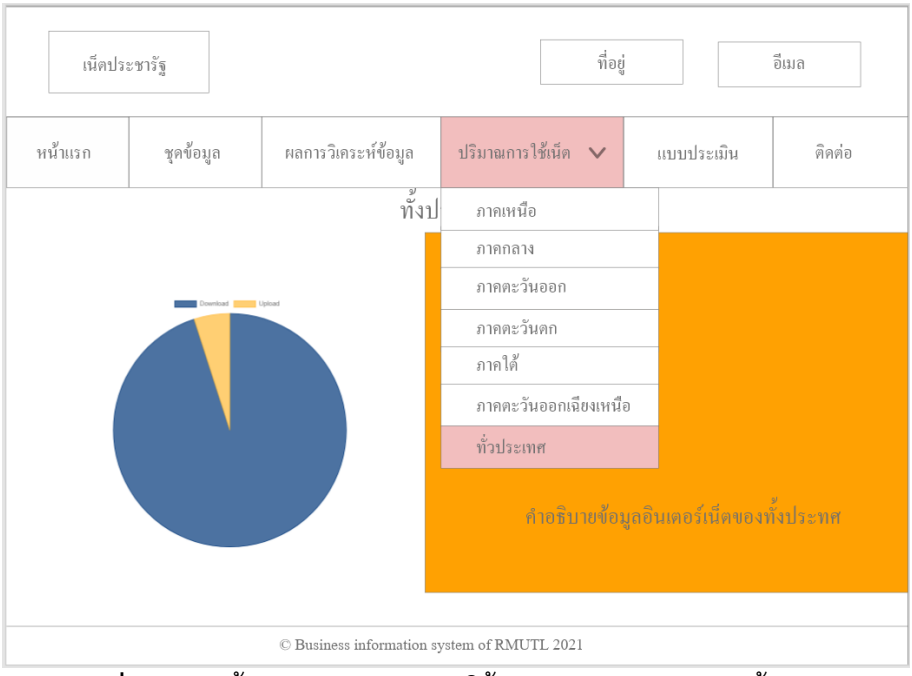
ภาพที่ 3.45 หน้าแสดงปริมาณการใช้เน็ตประชารัฐของภาคใต้

9.) หน้าแสดงปริมาณการใช้เน็ตประชารัฐของภาคตะวันออกเฉียงเหนือ



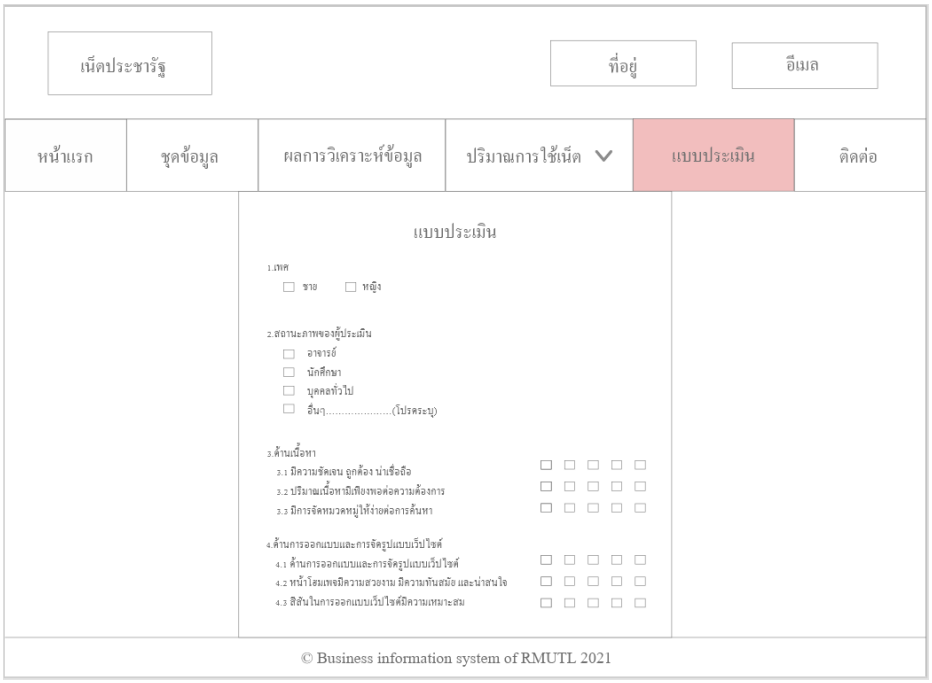
ภาพที่ 3.46 หน้าแสดงปริมาณการใช้เน็ตประชารัฐของภาคตะวันออกเฉียงเหนือ

10.) หน้าแสดงปริมาณการใช้งานเน็ตประชารัฐของทั้งประทศ



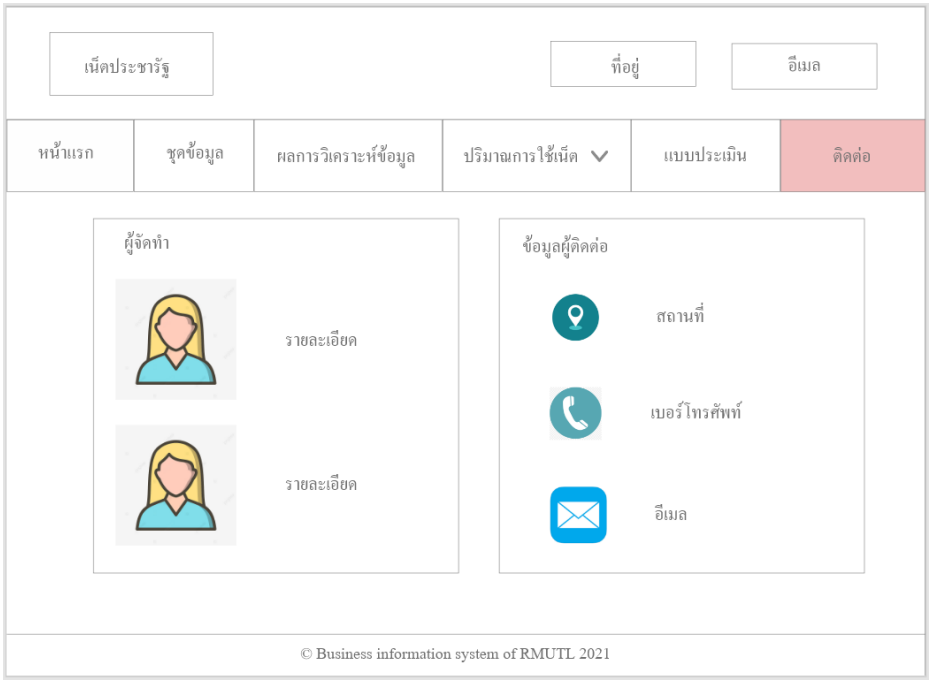
ภาพที่ 3.47 หน้าแสดงปริมาณการใช้งานเน็ตประชารัฐของทั้งประทศ

11.) หน้าแสดงแบบประเมิน



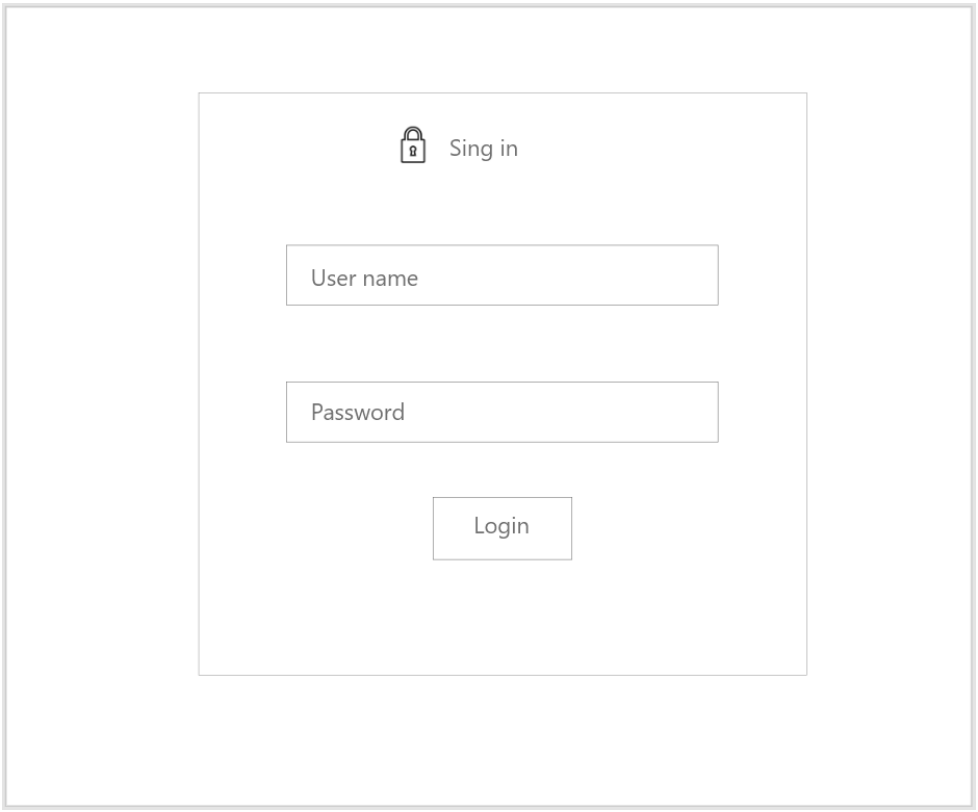
ภาพที่ 3.48 หน้าแสดงแบบประเมิน

12.) หน้าแสดงข้อมูลผู้จัดทำ



ภาพที่ 3.49 หน้าแสดงข้อมูลผู้จัดทำ

- หน้าเว็บสำหรับแอดมิน
  ส่วนเข้าสู่ระบบของแอดมิน ประกอบไปด้วยรายละเอียด ดังนี้
  1.) รูปโปรไฟล์แอดมิน
  2.) ช่องกรอกชื่อของแอดมิน
  3.) ชื่อกรอกรหัสของแอดมิน
  4.) ปุ่มกดล็อคอิน
  5.) ลิงค์การลืมรหัสผ่านและสร้างบัญชีใหม่



ภาพที่ 3.50 ส่วนเข้าสู่ระบบของแอดมิน

ส่วนที่แอดมินสามารถจัดการในเว็บไซต์ได้ ประกอบไปด้วยรายละเอียด ดังนี้

1.) ช่องชื่อของแอดมินที่เข้าใช้งาน

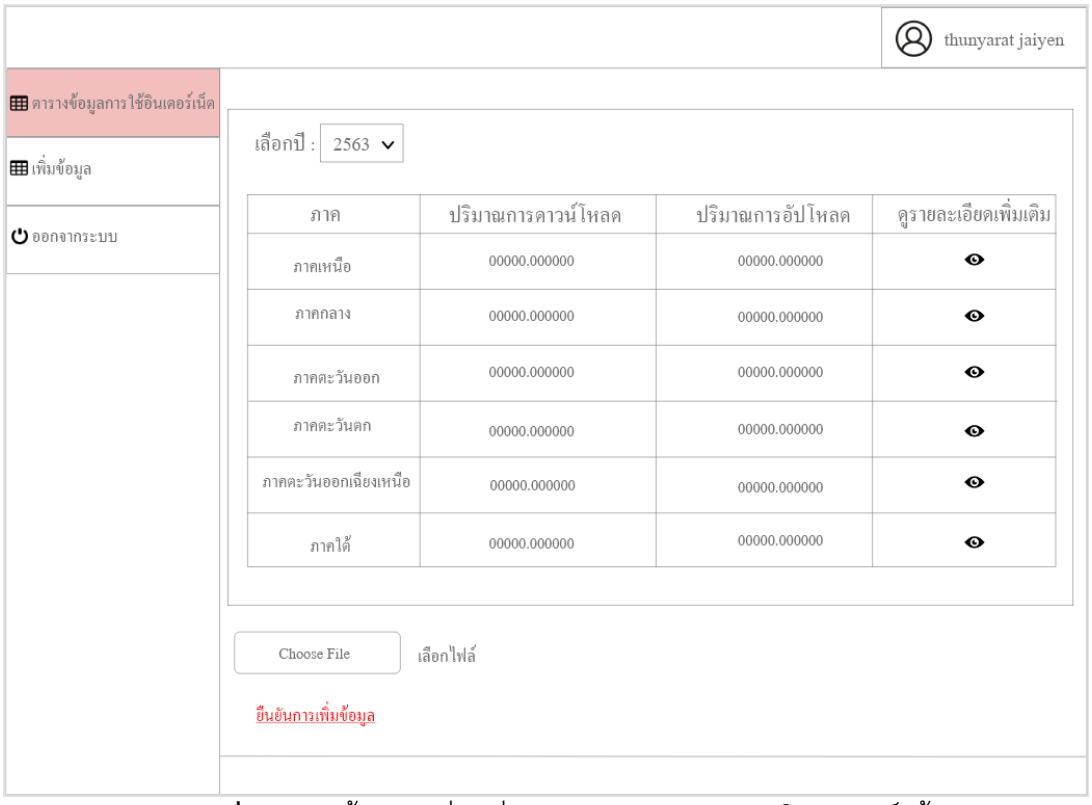
2.) ปุ่มกดออกจากระบบ

3.) ช่องชื่อไฟล์

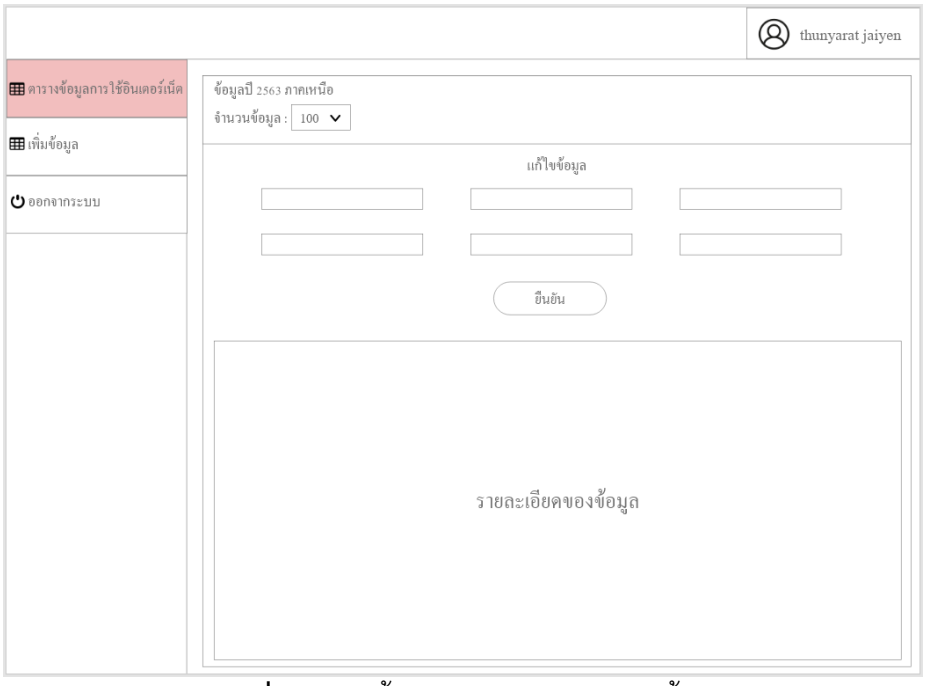
4.) ปุ่มกดอัปโหลดไฟล์

5.) ปุ่มกดบันทึกและยกเลิก

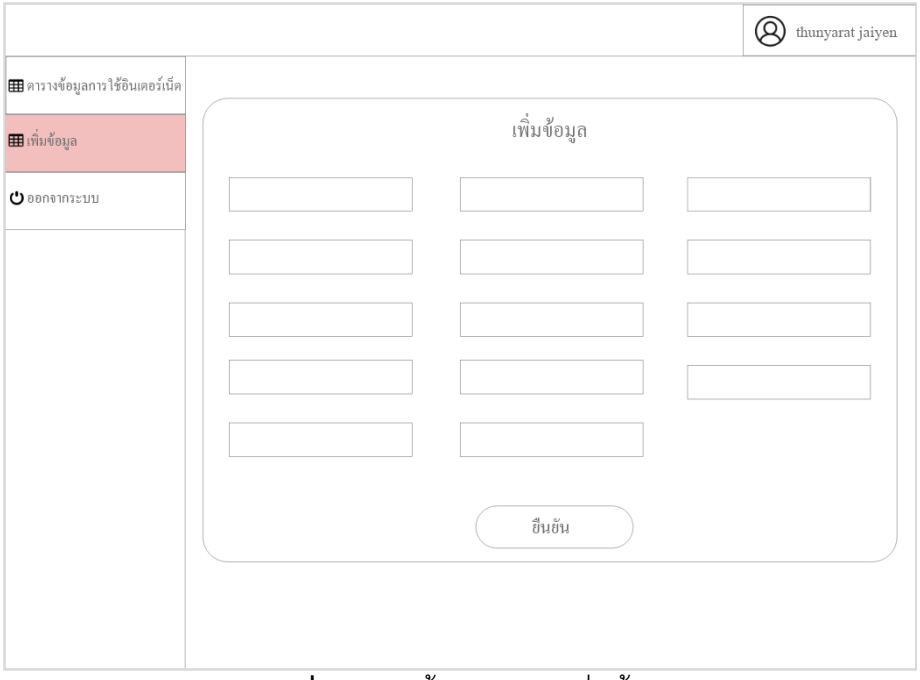
6.) ตารางข้อมูลของไฟล์ สามารถแก้ไขและลบไฟล์นั้นได้



ภาพที่ 3.51 หน้าแสดงส่วนที่แอดมินสามารถจัดการในเว็บไซต์ได้



ภาพที่ 3.52 หน้าแสดงรายละเอียดของข้อมูล



ภาพที่ 3.53 หน้าแสดงการเพิ่มข้อมูล

3.8 บทสรุป

จากขั้นตอนการดำเนินงานผู้วิเคราะห์ได้แสดงกระบวนการวิเคราะห์การพยากรณ์ข้อมูลปริมาณการใช้เน็ตประชารัฐ ด้วยกระบวนการ Data Flow Diagram กระบวนการทำนอร์มัลไลเซชัน (Normalization) และวิเคราะห์ข้อมูล CRISP-DM มาใช้ในการรวบรวมข้อมูลวิเคราะห์ข้อมูลที่เกี่ยวข้องกับงาน เพื่อให้ได้สารสนเทศของงานอย่างเพียงพอในการนำไปใช้ประโยชน์ โดยผู้วิเคราะห์ได้วิเคราะห์ข้อมูลเป็นหลักสำคัญที่จะช่วยให้เข้าใจในงานแต่ละส่วนจากการวิเคราะห์ข้อมูลด้วยกระบวนการวิเคราะห์ข้อมูล CRISP-DM รวมถึงการสร้างโมเดล Time-Series Forecasting โดยเลือกใช้โปรแกรม Tableau Public ในการพยากรณ์ล่วงหน้าตั้งแต่ 2เดือน ไปจนถึงพยากรณ์ล่วงหน้า 11เดือน สำหรับการวิเคราะห์ปริมาณการใช้เน็ตประชารัฐของปี 2563 และนำผลการพยากรณ์ที่ได้ไปประเมินประสิทธิภาพของโมเดล ด้วยวิธีการคำนวณหาค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) ที่ต่ำที่สุด ผู้วิเคราะห์ได้นำข้อมูลสารสนเทศมาทำการแสดงผลและเผยแพร่ข้อมูลสารสนเทศผ่านเว็บแอปพลิเคชันจะพัฒนาโดยใช้ภาษา PHP, JavaScript และชุดคำสั่ง CSS3 เพื่อนำเข้าวิเคราะห์ข้อมูลด้วยโมเดลที่เลือกใช้และผ่านการทดสอบประสิทธิภาพของการวิเคราะห์ข้อมูลแล้ว และนำเสนอข้อมูลภาพด้วยเฟรมเวิร์คการแสดงข้อมูลภาพชื่อ highcharts.js เพื่อสร้างกราฟและ dashboard สำหรับแสดงผลข้อมูลภาพจากข้อมูลที่ผ่านการวิเคราะห์แล้วภายในเว็บแอปพลิเคชันที่พัฒนาขึ้น