บทที่ 3
วิธีการดำเนินโครงงาน

ผู้จัดทำได้วิเคราะห์ความสัมพันธ์ของแหล่งข้อมูลและการจัดการข้อมูลที่มีอยู่ให้เกิดประโยชน์และได้สารสนเทศที่สอดคล้องกับข้อมูลผู้จัดทำได้ศึกษาวิเคราะห์ข้อมูลเชิงสัมพันธ์เชิงสาเหตุและการจัดการข้อมูลได้อยู่ในรูปแบบที่ถูกต้องโดยได้วิเคราะห์และออกแบบข้อมูลใหม่โดยใช้เครื่องมือประกอบไปด้วย กระบวนการในการพัฒนาฐานข้อมูล CRISP-DM สร้างแบบฟอร์มข้อมูลและปรับโครงสร้างข้อมูลใหม่ เทคนิควิธีทางเหมืองข้อมูล วิเคราะห์ความสัมพันธ์ข้อมูล (Visual Analytics tableau) ในการวิเคราะห์ข้อมูลซึ่งในบทนี้จะแสดงถึงวิธีในการดำเนินงาน ดังนี้

3.1 แผนภาพกระแสข้อมูล Data Flow Diagram

แผนภาพกระแสข้อมูล Data Flow Diagram คือ แบบจำลองกระบวนการ (Process Model) เป็นประเภทหนึ่ง มักนำมาใช้ในการวิเคราะห์และออกแบบระบบเชิงโครงสร้าง โดยแสดงความสัมพันธ์ระหว่างกระบวนการกับข้อมูล ซึ่งเป็นการแสดงการเคลื่อนที่ของข้อมูลจากกระบวนการหนึ่งไปยังกระบวนการหนึ่งโดยไม่อิงเทคโนโลยีการจัดเก็บข้อมูล เช่น สื่อบันทึกข้อมูล ฯลฯ

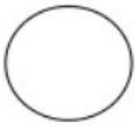
สัญลักษณ์ที่ใช้เป็นมาตรฐานในการแสดงแผนภาพกระแสข้อมูลมีหลายชนิด แต่ในที่นี้จะแสดงให้เห็นเพียง 2 ชนิด ได้แก่ ชุดสัญลักษณ์มาตรฐานที่พัฒนาโดย Gane and Sarson (1979) และชุดสัญลักษณ์มาตรฐานที่พัฒนาโดย DeMarco and Yourdon (DeMarco, 1979); Yourdon and Constantine, 1979) โดยมีสัญลักษณ์ดังต่อไปนี้

3.1.1 สัญลักษณ์ที่ใช้ในการวาดแผนภาพกระแสข้อมูล

สัญลักษณ์ที่นิมใช้ในการวาดแผนภาพกระแสข้อมูลมี 2 แบบ คือ

1) Gane and Sarson
2) DeMarco and Yourdon

ตารางที่ 3.1 สัญลักษณ์การไหลของข้อมูล แบบ Gane and Sarson

สัญลักษณ์	คำอธิบาย
	กระบวนการ (Process)
	การเคลื่อนที่ของข้อมูล (Data Flow)
	หน่วยภายนอก (External Entity)
	แฟ้มข้อมูล (Data Store/File)

ที่มา : https://www.gotoknow.org (2560)

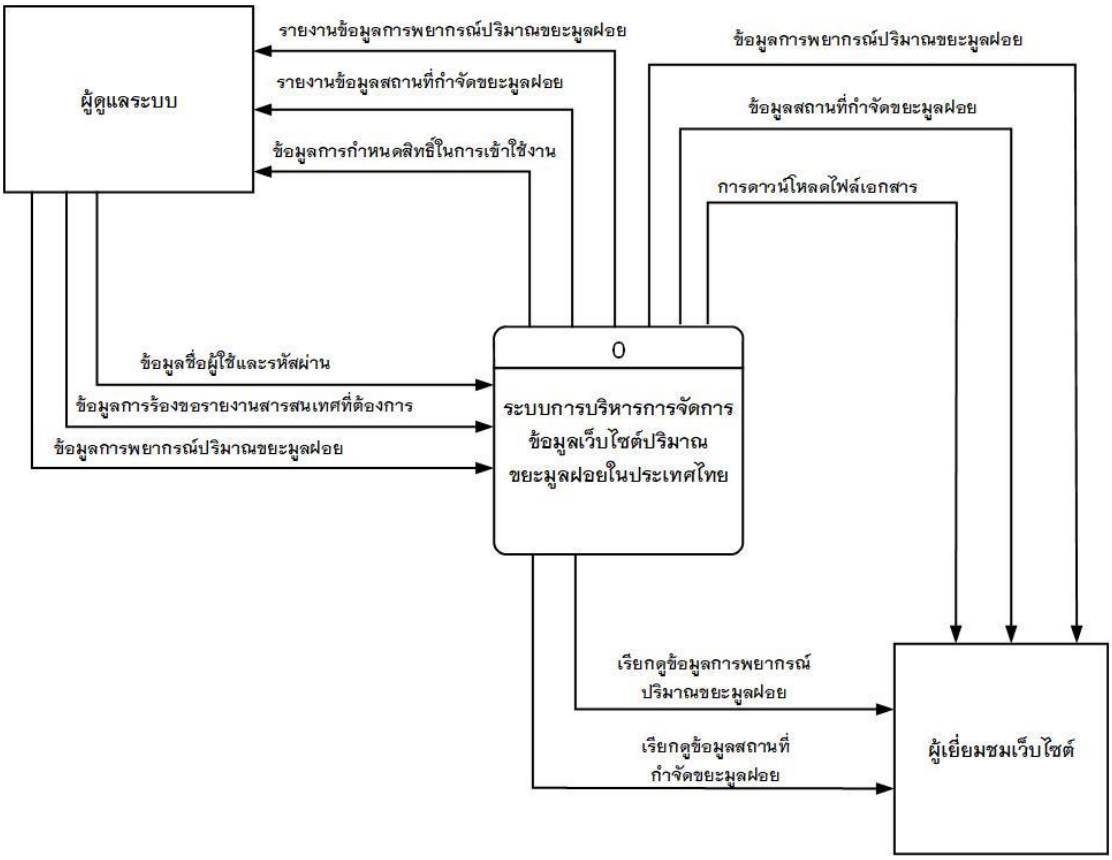
ตารางที่ 3.2 สัญลักษณ์การไหลของข้อมูล แบบ Yourdon

สัญลักษณ์	คำอธิบาย
	กระบวนการ (Process)
	การเคลื่อนที่ของข้อมูล (Data Flow)
	หน่วยภายนอก (External Entity)
	แฟ้มข้อมูล (Data Store/File)

ที่มา : https://www.gotoknow.org (2560)

3.1.2 แผนภาพบริบท Context Diagram

แผนภาพบริบท (Context Diagram) จะแสดงภาพโดยรวมระบบของการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย ดังภาพ 3.1

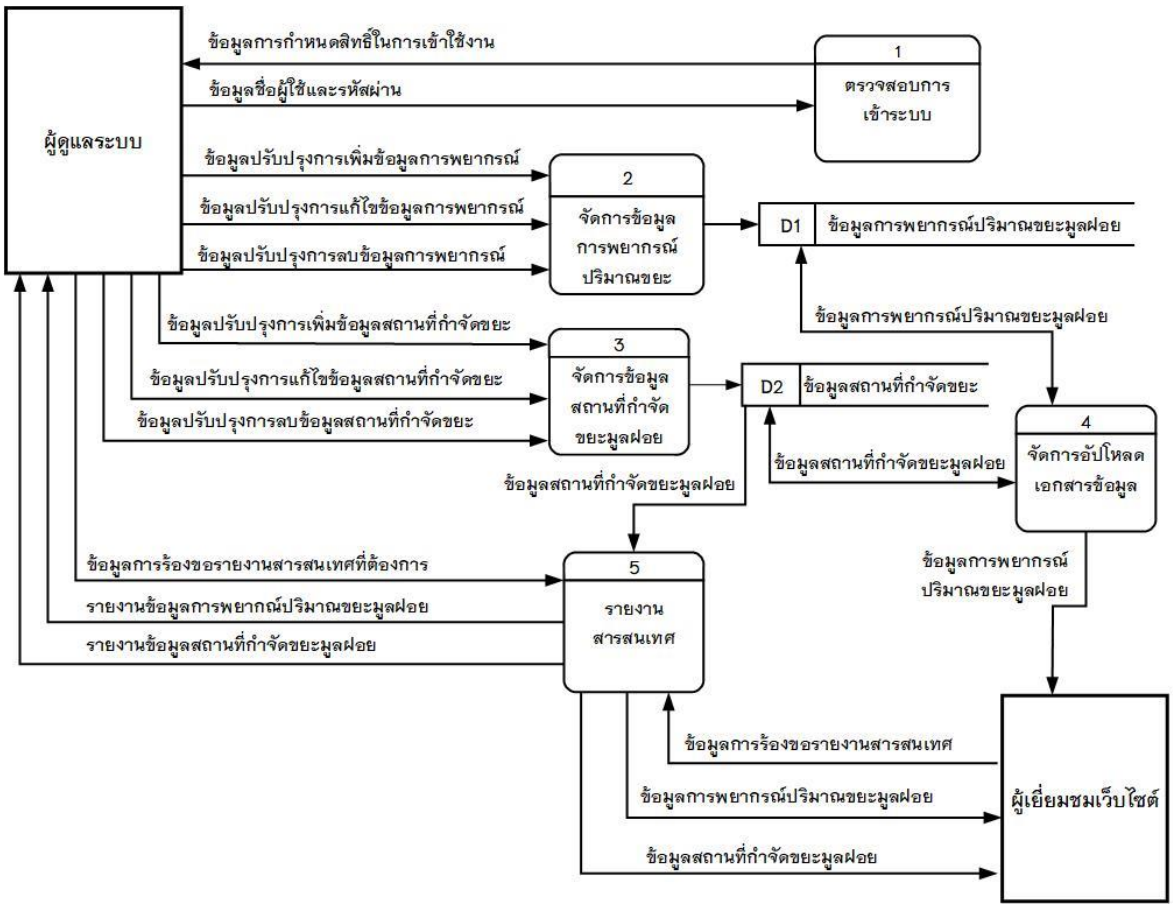


ภาพที่ 3.1 แผนภาพบริบท Context Diagram

จากภาพที่ 3.2 เป็นแผนผังระบบของการพัฒนาระบบบริหารจัดการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย
โดยสามารถแบ่งผู้ใช้ออกเป็น 2 ประเภทดังนี้

1) **ผู้ดูแลระบบ (System Administrator)** สามารถล็อกอินได้ สามารถออกจากระบบได้ สามารถจัดการข้อมูลผู้ใช้ระบบทั้งหมด

2) **ผู้ใช้ทั่วไป (User)** สามารถดูแหล่งการเรียนรู้ได้ สามารถดูการพยากรณ์ของปริมาณขยะ สามารถดูสถานที่กำจัดขยะ



ภาพที่ 3.2 แสดงแผนภาพกระแสข้อมูลระดับที่ 0 (Data Flow Diagram Level 0)

จากภาพที่ 3.2 แสดงแผนภาพกระแสข้อมูลระดับที่ 0 (Data Flow Diagram Level 0) ระบบบริหารฐานข้อมูลปริมาณขยะมูลฝอยในระเทศไทยและสถานที่กำจัดปริมาณขยะ ประกอบด้วย 5 กระบวนการหลักดังนี้

ตารางที่ 3.3 แสดงคำอธิบายกระบวนการ ตรวจสอบระบบ

Process Description	
System	ระบบบริหารจัดการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย
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 และกำหนดสิทธิ์ในการเข้าใจระบบ ผู้ใช้ถึงจะมีสิทธิ์เข้าใช้ระบบโดยมีผู้ใช้ระบบได้แก่ ผู้ดูแลระบบ

ตารางที่ 3.4 แสดงคำอธิบายกระบวนการ การจัดการข้อมูลการพยากรณ์ปริมาณขยะ

Process Description	
System	ระบบบริหารจัดการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย
DFD Number	2
Process Name	จัดการข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย
Input Data Flow	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย
Output Data Flow	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย
Data Store Used	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย
Description	เป็นกระบวนการสำหรับจัดการข้อมูลเกี่ยวกับการพยากรณ์ปริมาณขยะในประเทศไทย โดยผู้ใช้ระบบในส่วนนี้ได้แก่ การเพิ่ม ลบ แก้ไข ผู้มีสิทธิ์ใช้งานกระบวนการนี้ ได้แก่ ผู้ดูแลระบบ และออกสารสนเทศข้อมูลการพยากรณ์ปริมาณขยะมูลในประเทศไทย

ตารางที่ 3.5 แสดงคำอธิบายกระบวนการ จัดการข้อมูลสถานที่กำจัดขยะมูลฝอย ในประเทศไทย

Process Description	
System	ระบบบริหารจัดการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย
DFD Number	3
Process Name	จัดการข้อมูลการสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Input Data Flow	ข้อมูลการสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Output Data Flow	ข้อมูลการสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Data Store Used	ข้อมูลการสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Description	เป็นกระบวนการสำหรับการแสดงสถานที่กำจัดในแต่ละจังหวัด โดยผู้ใช้ระบบในส่วนนี้ได้แก่ การเพิ่ม ลบ แก้ไข ผู้มีสิทธิ์ใช้งานกระบวนการนี้ ได้แก่ ผู้ดูแลระบบ และออกสารสนเทศข้อมูลสถานที่กำจัดขยะมูลในประเทศไทย

ตารางที่ 3.6 แสดงคำอธิบายกระบวนการจัดการอัปโหลดเอกสารข้อมูลต่างๆ

Process Description	
System	ระบบบริหารจัดการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย
DFD Number	4
Process Name	จัดการอัปโหลดเอกสารข้อมูลต่างๆ
Input Data Flow	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Output Data Flow	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย และข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย เป็นไฟล์ Excel หรือ PDF และเอกสารที่เกี่ยวข้องที่เป็นไฟล์ Excel หรือ PDF
Data Store Used	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย และข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย

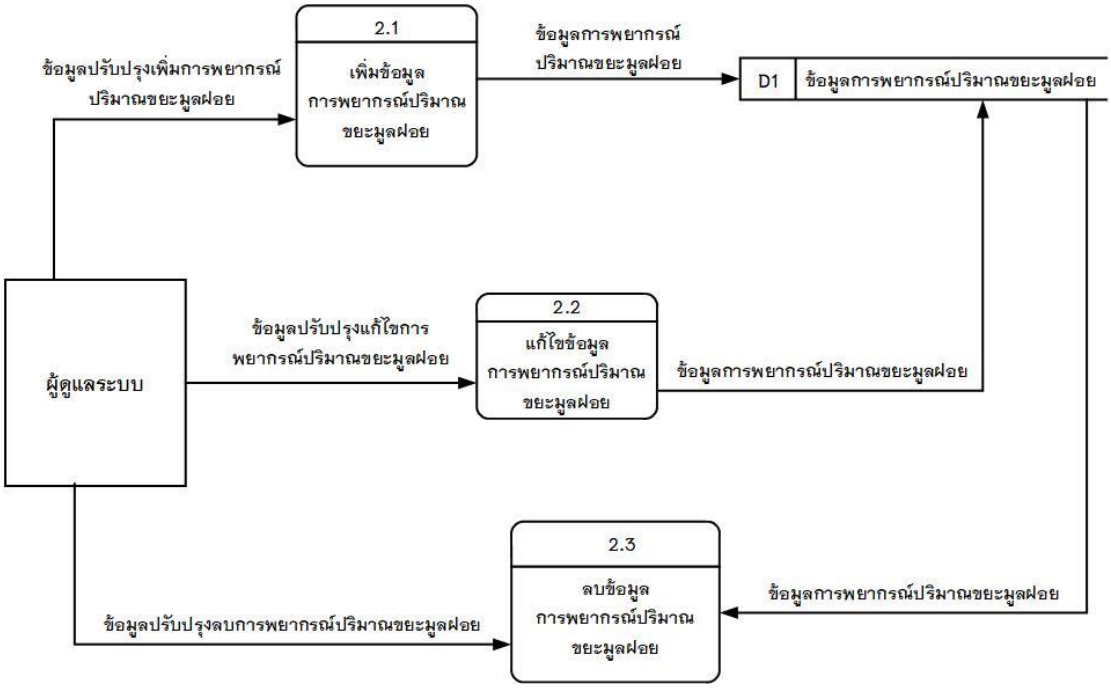
Description	เป็นกระบวนการสำหรับจัดการดาวน์โหลดเอกสารต่าง ๆเช่นข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย และสถานที่กำจัดขยะมูลฝอยในประเทศไทยที่ผู้เยี่ยมชมเว็บไซต์สามารถดาวน์โหลดได้

ตารางที่ 3.7 แสดงคำอธิบายกระบวนการจัดการรายงานสารสนเทศ

Process Description	
System	ระบบบริหารจัดการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย
DFD Number	5
Process Name	จัดการรายงานสารสนเทศ
Input Data Flow	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย,ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Output Data Flow	รายงานข้อมูลการพยากรณ์ขยะมูลฝอยในประเทศไทย และรายงานข้อมูลสถานที่กำจัดมูลฝอยในประเทศไทย
Data Store Used	ข้อมูลปริมาณการพยากรณ์ปริมาณขยะ, ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Description	เป็นกระบวนการสำหรับออกรายงานสารสนเทศ โดยผู้มีสิทธิ์เข้าใช้ระบบได้แก่ ผู้ดูแลระบบ

จากแผนภาพกระแสข้อมูล ระดับที่ 0 สามารถแยกย่อยเป็นกระบวนการย่อยระดับที่ 1 ได้ดังนี้

แผนภาพกระแสข้อมูล ระดับที่ 1 กระบวนการที่ 2 จัดการข้อมูล ประเด็นที่ 1 ปริมาณขยะในประเทศไทย



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

ตารางที่ 3.8 แสดงคำอธิบายกระบวนการที่ 2.1 เพิ่มข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย

Process Description	
System	ระบบบริหารจัดการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย
DFD Number	2.1
Process Name	ระบบการเพิ่มข้อมูล
Input Data Flow	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย
Output Data Flow	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย

Data Store Used	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย
Description	เป็นกระบวนการสำหรับการเพิ่มข้อมูลในการพยากรณ์ขยะมูลฝอย โดยผู้ใช้ระบบในส่วนนี้ได้แก่ ผู้ดูแลระบบ

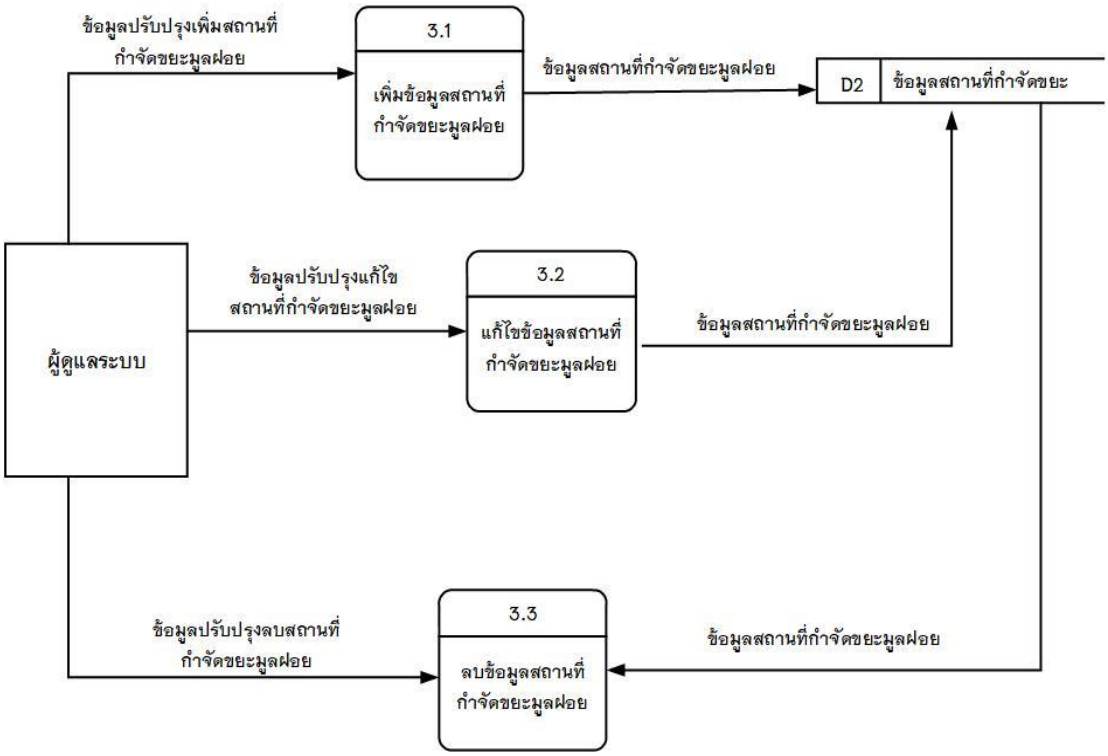
ตารางที่ 3.9 แสดงคำอธิบายกระบวนการที่ 2.2 แก้ไขข้อมูลการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย

Process Description	
System	ระบบบริหารจัดการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย
DFD Number	2.2
Process Name	ระบบการเพิ่มข้อมูล
Input Data Flow	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอย
Output Data Flow	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอย
Data Store Used	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอย
Description	เป็นกระบวนการสำหรับการเพิ่มข้อมูลในการพยากรณ์ขยะมูลฝอย โดยผู้ใช้ระบบในส่วนนี้ได้แก่ ผู้ดูแลระบบ

ตารางที่ 3.10 แสดงคำอธิบายกระบวนการที่ 2.3 ลบข้อมูลการพยากรณ์ปริมาณมูลฝอยในประเทศไทย

Process Description	
System	ระบบบริหารจัดการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย
DFD Number	2.3
Process Name	ระบบการลบข้อมูล
Input Data Flow	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอย
Output Data Flow	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอย
Data Store Used	ข้อมูลการพยากรณ์ปริมาณขยะมูลฝอย
Description	เป็นกระบวนการสำหรับการลบข้อมูลในการพยากรณ์ขยะมูลฝอย โดยผู้ใช้ระบบในส่วนนี้ได้แก่ ผู้ดูแลระบบ

แผนภาพกระแสข้อมูล ระดับที่ 1 กระบวนการที่ 2 จัดการข้อมูล ประเด็นที่ 2 สถานที่กำจัดขยะในประเทศไทย



ภาพที่ 3.4 แผนภาพกระแสข้อมูล ระดับที่ 1 กระบวนการที่ 3.1, 3.2, 3.3 จัดการข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย

ตารางที่ 3.11 แสดงคำอธิบายกระบวนการที่ 3.1 เพิ่มข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย

Process Description	
System	ระบบบริหารจัดการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย
DFD Number	3.1
Process Name	การเพิ่มข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Input Data Flow	ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Output Data Flow	ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย

Data Store Used	ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Description	เป็นกระบวนการสำหรับการเพิ่มข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย โดยผู้ใช้ระบบในส่วนนี้ได้แก่ ผู้ดูแลระบบ

ตารางที่ 3.12 แสดงคำอธิบายกระบวนการที่ 3.2 แก้ไขสถานที่กำจัดขยะมูลฝอยในประเทศไทย

Process Description	
System	ระบบบริหารจัดการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย
DFD Number	3.2
Process Name	แก้ไขข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Input Data Flow	ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Output Data Flow	ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Data Store Used	ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Description	เป็นกระบวนการสำหรับการเพิ่มข้อมูลข้อมูลสถานที่กำจัดขยะในประเทศไทย โดยผู้ใช้ระบบในส่วนนี้ได้แก่ ผู้ดูแลระบบ

ตารางที่ 3.13 แสดงคำอธิบายกระบวนการที่ 3.3 ลบข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย

Process Description	
System	ระบบบริหารจัดการพัฒนาเว็บแอปพลิเคชันเพื่อการวิเคราะห์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย
DFD Number	3.3
Process Name	ลบข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Input Data Flow	ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Output Data Flow	ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Data Store Used	ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทย
Description	เป็นกระบวนการสำหรับการลบข้อมูลในสถานที่กำจัดขยะมูลฝอยในประเทศไทย โดยผู้ใช้ระบบในส่วนนี้ได้แก่ ผู้ดูแลระบบ

3.2 กระบวนการทำนอร์มัลไลเซชัน (Normalization)

การทำนอร์มัลไลเซชัน เป็นวิธีการในการกำหนดแอตทริบิวต์ให้กับแต่ละเอนทิตี เพื่อให้ได้โครงสร้างของตารางที่ดี สามารถควบคุมความซ้ำซ้อนของข้อมูลหลีกเลี่ยงความผิดปกติของข้อมูล โดยทั่วไปผลลัพธ์ของการนอร์มัลไลเซชัน จะได้ตารางที่มีโครงสร้างซับซ้อนน้อยลง แต่จำนวนของตารางจะมากขึ้น

การทำนอร์มัลไลเซชัน จะประกอบด้วยนอร์มัลฟอร์ม (Normal Form) แบบต่าง ๆ ที่มีเงื่อนไขของการทำให้อยู่ในรูปของนอร์มัลฟอร์มที่แตกต่างกันไป ขึ้นอยู่กับผู้ออกแบบฐานข้อมูลว่า ต้องการลดความซ้ำซ้อนในฐานข้อมูลให้อยู่ในระดับใด 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)

กระบวนการทำนอร์มัลฟอร์มที่ 1 (First Normal Form : 1NF)

ค่าของแอททริบิวท์ต่าง ๆ ในแต่ละจะต้องมีค่าเพียงค่าเดียว และรีเลชันนั้นจะต้องไม่มีกลุ่มข้อมูลซ้ำ

ตารางที่ 3.14 ตารางข้อมูลก่อนที่จะทำการนอร์มัลฟอร์มที่ 1 ในประเด็นที่ 1

Province_id	Province_name	region_id	region	Year	Category_id	Category_name	garbage
1001	กำแพงเพชร	P01	ภาคเหนือ	2013	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	31,996.10
1002	เชียงราย	P01	ภาคเหนือ	2013	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	108,500.55
1003	เชียงใหม่	P01	ภาคเหนือ	2013	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	253,864.07
1004	ตาก	P01	ภาคเหนือ	2013	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	45,689.75

จากตารางที่ 3.14 อยู่ในรูปแบบ Unnormalized Form จะต้องทำให้อยู่ในรูปของ 1NF โดยสามารถทำให้อยู่ในรูปแบบ 1NF ได้

ตารางข้อมูลที่อยู่ในรูปของนอร์มัลฟอร์มที่ 1 ในประเด็นที่ 1 ปริมาณขยะมูลฝอยในประเทศไทย (First Normal Form : 1NF)

ตารางที่ 3.15 ตารางข้อมูลที่อยู่ในรูปของนอร์มัลฟอร์มที่ 1 ในประเด็นที่ 1

Province_id	Province_name	region_id	region	Year	Category_id	Category_name	garbage
1001	กำแพงเพชร	P01	ภาคเหนือ	2013	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	31,996.10
1001	กำแพงเพชร	P01	ภาคเหนือ	2013	002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)	127,181.69
1001	กำแพงเพชร	P01	ภาคเหนือ	2013	003	ปริมาณขยะมูลฝอยที่ นำไปใช้ประโยชน์ (ตัน)	63,282.05
1002	เชียงราย	P01	ภาคเหนือ	2013	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	108,500.55
1002	เชียงราย	P01	ภาคเหนือ	2013	002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)	240,928.64
1002	เชียงราย	P01	ภาคเหนือ	2013	003	ปริมาณขยะมูลฝอยที่ นำไปใช้ประโยชน์ (ตัน)	185,718.77
1003	เชียงใหม่	P01	ภาคเหนือ	2013	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	253,864.07
1003	เชียงใหม่	P01	ภาคเหนือ	2013	002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)	144,595.79
1003	เชียงใหม่	P01	ภาคเหนือ	2013	003	ปริมาณขยะมูลฝอยที่ นำไปใช้ประโยชน์ (ตัน)	228,944.71
1004	ตาก	P01	ภาคเหนือ	2013	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	45,689.75
1004	ตาก	P01	ภาคเหนือ	2013	002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)	142,227.62
1004	ตาก	P01	ภาคเหนือ	2013	003	ปริมาณขยะมูลฝอยที่ นำไปใช้ประโยชน์ (ตัน)	66,537.87

จากตารางที่ 3.15 จะปรับค่าให้เป็น 1NF โดยมีการทำ multi-value ให้กลายเป็น single value โดยเติมค่าของข้อมูลใน attribute ให้เต็ม โดยกำหนดตาราง Province (Province_id, Province_name, region, region_name, Year, Category_id, Category_name, garbage และในค่าข้อมูลของแต่ละ tuple จะไม่มีค่าที่ซ้ำกัน

กระบวนการทำนอร์มัลฟอร์มที่ 2 (Second Normal Form : 2NF)

รีเลชั่นใดๆ จะอยู่ในรูปแบบบรรทัดฐานระดับที่ 2 ก็ต่อเมื่อ รีเลชั่นนั้น ๆ อยู่ในรูปแบบบรรทัดฐานระดับที่ 1 และแอททริบิวต์ทุกตัวที่ไม่ได้เป็นคีย์หลัก จะต้องมีความสัมพันธ์ระหว่างค่าของแอททริบิวต์แบบ ฟังก์ชั่นกับคีย์หลัก (Fully Functional Dependency)

รีเลชั่นที่มีคุณสมบัติเป็นนอร์มัลฟอร์มระดับที่ 2 (Second Normal Form: 2NF)

1) จะต้องเป็นรีเลชั่นที่มีคุณสมบัติเป็น 1NF
2) แอททรีบิวท์ที่ไม่ใช่คีย์จะต้องขึ้นกับคีย์หลักอย่างแท้จริง

ตารางข้อมูลที่อยู่ในรูปของนอร์มัลฟอร์มที่ 2 ในประเด็นที่ 1 (Second Normal Form : 2NF)

ตารางที่ 3.16 ตารางข้อมูลที่อยู่ในรูปของนอร์มัลฟอร์มที่ 2 ในประเด็นที่ 1

Province_id	Province_name
1001	กำแพงเพชร
1001	กำแพงเพชร
1001	กำแพงเพชร
1002	เชียงราย
1002	เชียงราย
1002	เชียงราย
1003	เชียงใหม่
1003	เชียงใหม่
1003	เชียงใหม่
1004	ตาก
1004	ตาก
1004	ตาก

จากตารางที่ 3.16 จะมีการตัด Attribute ในตารางเดิม เมื่อนำไปใส่ตารางใหม่แล้วให้ตัดออกจากตารางเดิม ยกเว้น Attribute ที่เป็น PK ของตารางใหม่ ให้คงไว้ในตารางเดิม เพื่อใช้เป็น FK ระหว่างตารางเดิม โดยมีการกำหนดตารางใหม่ Province (Province_id, Province_name)

ตารางที่ 3.17 ตาราง region ที่อยู่ในรูปของนอร์มัลฟอร์มที่ 2 (Normal Form : 2NF)

region_id	ragion_name
P01	ภาคเหนือ
P01	ภาคเหนือ
P01	ภาคเหนือ
P01	ภาคเหนือ
P01	ภาคเหนือ
P01	ภาคเหนือ
P01	ภาคเหนือ
P01	ภาคเหนือ
P01	ภาคเหนือ
P01	ภาคเหนือ
P01	ภาคเหนือ
P01	ภาคเหนือ

ตารางที่ 3.18 ตาราง Category ที่อยู่ในรูปของนอร์มัลฟอร์มที่ 2 (Normal Form : 2NF)

Category_ID	Category
001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)
002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)
003	ปริมาณขยะมูลฝอยที่ นำไปใช้ประโยชน์ (ตัน)
001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)
002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)
003	ปริมาณขยะมูลฝอยที่ นำไปใช้ประโยชน์ (ตัน)
001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)
002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)
003	ปริมาณขยะมูลฝอยที่ นำไปใช้ประโยชน์ (ตัน)
001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)
002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)
003	ปริมาณขยะมูลฝอยที่ นำไปใช้ประโยชน์ (ตัน)

ตารางที่ 3.19 ตาราง Garbage_dumping ที่สมบูรณ์ในประเด็นที่ 1

Garbage_Dumping_id	Province_id	region_id	Category_id	Year	garbage
amg0001	1001	P01	001	2013	31,996.10
amg0002	1001	P01	002	2013	127,181.69
amg0003	1001	P01	003	2013	63,282.05
amg0004	1002	P01	001	2013	108,500.55
amg0005	1002	P01	002	2013	240,928.64
amg0006	1002	P01	003	2013	185,718.77
amg0007	1003	P01	001	2013	253,864.07
amg0008	1003	P01	002	2013	144,595.79
amg0009	1003	P01	003	2013	228,944.71
amg0010	1004	P01	001	2013	45,689.75
amg0011	1004	P01	002	2013	142,227.62
amg0012	1004	P01	003	2013	66,537.87

กระบวนการทำนอร์มัลฟอร์มที่ 1 ในประเด็นที่ 2 สถานที่กำจัดขยะ (First Normal Form : 1NF)

ค่าของแอททริบิวท์ต่าง ๆ ในแต่ละทัพเพิลจะต้องมีค่าเพียงค่าเดียว และรีเลชันนั้นจะต้องไม่มีกลุ่มข้อมูลซ้ำ

ตารางข้อมูลก่อนที่จะทำการนอร์มัลฟอร์มที่ 1 ในประเด็นที่ 2 (First Normal Form : 1NF)

ตารางที่ 3.20 ตารางข้อมูลก่อนที่จะทำการนอร์มัลฟอร์มที่ 1 ในประเด็นที่ 2 สถานที่กำจัดขยะในประเทศไทย (First Normal Form : 1NF)

Province_id	Province_name	District_id	District_name	Sub-district_id	Sub-district_name	area size	Category_ID	Category_name	volume	year
1001	กำแพงเพชร	62170	-	S01	-	-	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	-	2013
							002			
1002	เชียงราย	57230	เทิง	S02	ปล้อง	10.5	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	383.7	2013
							002			
1003	เชียงใหม่	50130	แม่ออน	S03	ออนเหนือ	5	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	845.51	2013
							002			
1004	ตาก	63120	บ้านตาก	S04	เกาะตะเภา	10	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	15.00	2013
							002			

จากตารางที่ 3.20 อยู่ในรูปแบบ Unnormalized Form จะต้องทำให้อยู่ในรูปของ 1NF โดยสามารถทำให้อยู่ในรูปแบบ 1NF ได้

ตารางที่ 3.21 ตารางที่อยู่ในรูปของนอร์มัลฟอร์มที่ 1 ในประเด็นที่ 2

Province_id	Province_name	District_id	District_name	Sub-district_id	Sub-district_name	area size	Category_id	Category_name	volume	year
1001	กำแพงเพชร	-	-	-	-	-	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	-	2013
1001	กำแพงเพชร	62170	ลานกระบือ	S01	ลานกระบือ	13	002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)	93.14	2013
1002	เชียงราย	57230	เทิง	S02	ปล้อง	10.5	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	383.7	2013
1002	เชียงราย	57150	เชียงแสน	S03	เวียง	50	002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)	209.26	2013
1003	เชียงใหม่	50130	แม่ออน	S04	ออนเหนือ	5	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	845.51	2013
1003	เชียงใหม่	50150	แม่แดง	S05	แม่หอพระ	2	002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)	272.27	2013
1004	ตาก	63120	บ้านตาก	S06	เกาะตะเภา	10	001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	15.00	2013
1004	ตาก	63160	พบพระ	S07	ช่องแคบ	5	002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)	331.3	2013

จากตารางที่ 3.21 จะปรับค่าให้เป็น 1NF โดยมีการทำ multi-value ให้กลายเป็น single value โดยเติมค่าของข้อมูลใน attribute ให้เต็ม โดยกำหนดตาราง Province (Province_id, Province_name, District_id, District_name, Sub-district_id, Sub-district_name, area size,

Category_id, Category_name, Volume, Year และในค่าข้อมูลของแต่ละ tuple จะไม่มีค่าที่ซ้ำกัน

จากการพิจารณารีเลชันสถานที่กำจัดขยะ จะพบว่ามีฟังก์ชันการขึ้นต่อกัน ดังนี้ จะพบว่าในรีเลชันสถานที่กำจัดขยะจะมี Province_id, District_id, Sub-district_id, Category_id, ประกอบกันเป็นคีย์หลัก และจาการพิจารณาพบว่า จะไม่ตรงตามนิยามของรูปแบบนอมัลระดับที่ 2 เพราะแอตทริบิวต์ aera size, Volum, Year จะมีค่าเปลี่ยนแปลงตามแอตทริบิวส์ Province_id, District_id, Sub-district_id, Category_id ดังนั้นจึงต้องทำการแยกรีเลชันออกเป็น 5 รีเลชัน ดังนี้

ตารางที่ 3.22 ตาราง Province ที่อยู่ในรูปของนอร์มัลฟอร์มที่ 2 (Normal Form : 2NF)

Province_id	Province_name
1001	กำแพงเพชร
1001	กำแพงเพชร
1002	เชียงราย
1002	เชียงราย
1003	เชียงใหม่
1003	เชียงใหม่
1004	ตาก
1004	ตาก

ตารางที่ 3.23 ตาราง District ที่อยู่ในรูปของนอร์มัลฟอร์มที่ 2 (Normal Form : 2NF)

District_id	District_name
–	–
62170	สานกระบือ
57230	เทิง
57150	เชียงแสน
50130	แม่ออน
50150	แม่แตง
63120	บ้านตาก
63160	พบพระ

ตารางที่ 3.24 ตาราง Sub-district ที่อยู่ในรูปของนอร์มัลฟอร์มที่ 2 (Normal Form : 2NF)

Sub-district_id	Sub-district
-	-
S01	ลานกระบือ
S02	ปล้อง
S03	เวียง
S04	ออนเหนือ
S05	แม่หอพระ
S06	เกาะตะเภา
S07	ช่องแคบ

ตารางที่ 3.25 ตารางข้อมูล Category ที่อยู่ในรูปของนอร์มัลฟอร์มที่ 2 (Normal Form: 2NF)

Category_id	category_name
001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)
002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)
001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)
002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)
001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)
002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)
001	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)
002	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)

ตารางที่ 3.26 ตารางข้อมูล Garbage_location ที่สมบูรณ์ในประเด็นที่ 2

Garbage_location_id	Province_id	District_id	Sub-district_id	Category_id	area size	volume	Year
log001	1001	-	-	001	-	-	2013
log002	1001	62170	S01	002	13	93.14	2013
log003	1002	57230	S02	001	10.5	383.7	2013
log004	1002	57150	S03	002	50	209.26	2013
log005	1003	50130	S04	001	5	845.51	2013
log006	1003	50150	S05	002	2	272.27	2013
log007	1004	63120	S06	001	10	15.00	2013
log008	1004	63160	S07	002	5	331.3	2013

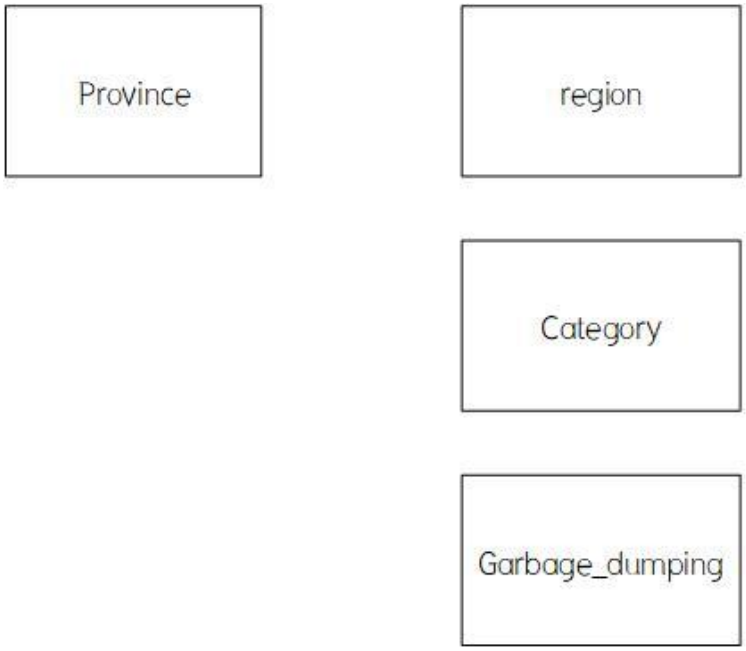
จากการที่ผู้วิเคราะห์ได้ออกแบบเป็นฐานข้อมูลไว้ มาปรับปรุงโดยใช้วิธีการทำนอร์มัลไลเซชัน เพื่อปรับเปลี่ยนโครงสร้างของรีเลชันต่าง ๆ ทำให้สามารถแก้ปัญหาลดความซ้ำซ้อนของข้อมูลในลักษณะต่าง ๆ และทำให้รีเลชันมีโครงสร้างที่เหมาะสมในการใช้งานต่อไป โดยผู้วิเคราะห์ได้ทำนอร์มัลไลเซชันได้ 2 และระดับขั้นตอนโดยประเด็นที่ 1 จะได้ตารางทั้งหมด 4 ตารางประกอบไปด้วยตารางที่ 1 ชื่อรีเลชัน Province ตารางที่ 2 ชื่อรีเลชัน reion ตารางที่ 3 ชื่อรีเลชัน Category ตารางที่ 4 ชื่อรีเลชัน Garbage_Dumping และในระดับขั้นตอนโดยประเด็นที่ 2 จะได้ตารางทั้งหมด 5 ตารางประกอบไปด้วยตารางที่ 1 ชื่อรีเลชัน Province ตารางที่ 2 ชื่อรีเลชัน District ตารางที่ 3 ชื่อรีเลชัน Sub-district_id ตารางที่ 4 ชื่อรีเลชัน Category และตารางที่ 5 ชื่อรีเลชัน Garbage_location จากข้อมูลของปริมาณขยะมูลฝอยในประเทศไทยและสถานที่กำจัดขยะในประเทศไทยที่ได้นำมาผ่านกระบวนการนอร์มัลไลเซชันทั้ง 2 ขั้นตอนก็เพียงพอต่อการใช้งานและตอบสนองความต้องการของผู้ใช้งานแล้ว

3.3 ความสัมพันธ์ของฐานข้อมูล (ER-Diagram)

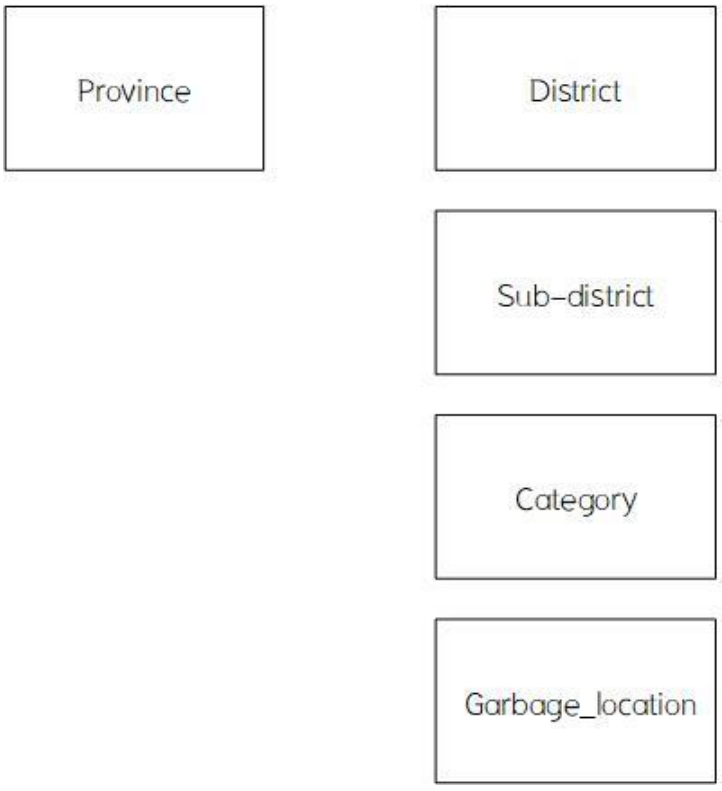
การออกแบบฐานข้อมูลโดยใช้แบบจำลองความสัมพันธ์ระหว่างข้อมูล (ER Model) มีขั้นตอนที่เกี่ยวข้องหลายขั้นตอน และต้องใช้ความรอบคอบในการออกแบบเพื่อให้ได้ ER-Diagram ที่ถูกต้องเหมาะสมกับระบบงาน ในขั้นตอนการออกแบบฐานข้อมูลนี้ใช้ตัวอย่างฐานข้อมูลลงทะเบียนสามารถนำมาสร้างแบบจำลองข้อมูลด้วย ER Model โดยมีขั้นตอนในการพัฒนาดังนี้

3.3.1 การกำหนดเอนทิตี้

การกำหนดเอนทิตี้เป็นการกำหนดสิ่งที่สนใจและต้องการจัดเก็บข้อมูลที่มีในระบบงาน โดยดูจากลักษณะหน้าที่ของระบบงานว่ามีรายละเอียดการทำงานอย่างไร โดยจะต้องพิจารณาด้วยว่าเป็นเอนทิตี้ประเภทใด เช่น เอนทิตี้แบบปกติ (Regular Entity) หรือเอนทิตี้แบบอ่อนแอ (Weak Entity) เช่น ฐานข้อมูลของปริมาณขยะในประเด็นที่ 1 ประกอบด้วย เอนทิตี้ Province เอนทิตี้ region เอนทิตี้ category เอนทิตี้ Garbage_dumping และฐานข้อมูลของสถานที่กำจัดขยะในประเด็นที่ 2 ประกอบด้วย เอนทิตี้ Province เอนทิตี้ District เอนทิตี้ Sub-district เอนทิตี้ Category เอนทิตี้ Garbage_location



ภาพที่ 3.5 การกำหนดเอนทิตี้ในประเด็นที่ 1 ปริมาณขยะมูลฝอยในประเทศไทย



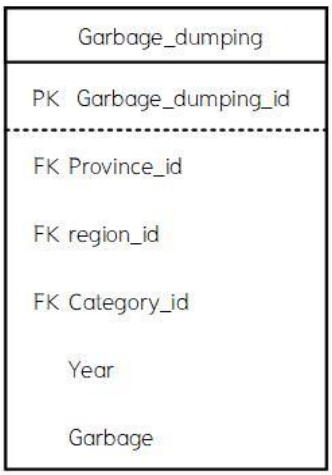
ภาพที่ 3.6 การกำหนดเอนทิตี้ในประเด็นที่ 2 สถานที่กำจัดขยะมูลฝอยในประเทศไทย

3.3.2 การกำหนดแอตทริบิวต์ของเอนทิตี

การกำหนดแอตทริบิวต์ของเอนทิตีเป็นการกำหนดคุณลักษณะต่าง ๆ ของแต่ละเอนทิตีรวมทั้งพิจารณาแอตทริบิวต์ที่จะทำหน้าที่เป็นคีย์หลักของเอนทิตีด้วย

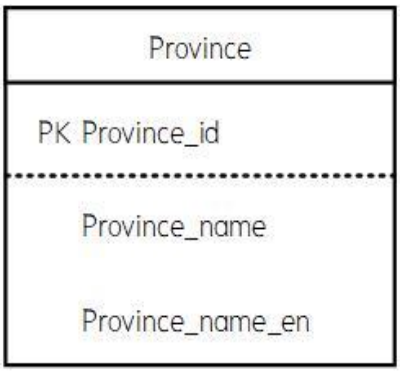
ดังตัวอย่างประเด็นที่ 1 ปริมาณขยะ ประกอบด้วย

1) เอนทิตี้ Garbage_dumping ประกอบด้วย Garbage_dumping_id, Province_id, region_id, Category_id, Year, Garbage



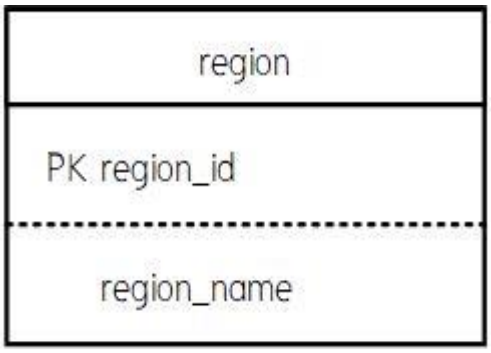
ภาพที่ 3.7 การกำหนดแอตทริบิวต์ Garbage_dumping ในประเด็นที่ 1 ปริมาณขยะมูลฝอยในประเทศไทย

2) เอนทิตี้ Province ประกอบด้วยProvince_id,Province_name, Province_name_en



ภาพที่ 3.8 การกำหนดแอตทริบิวต์ Province ในประเด็นที่ 1 ปริมาณขยะมูลฝอยในประเทศไทย

3) เอนทิตี้ region ประกอบด้วย region_id, region_name



ภาพที่ 3.9 การกำหนดแอตทริบิวต์ region ในประเด็นที่ 1 ปริมาณขยะมูลฝอยในประเทศไทย

4) เอนทิตี้ Category ประกอบด้วย Category_id, Category_name

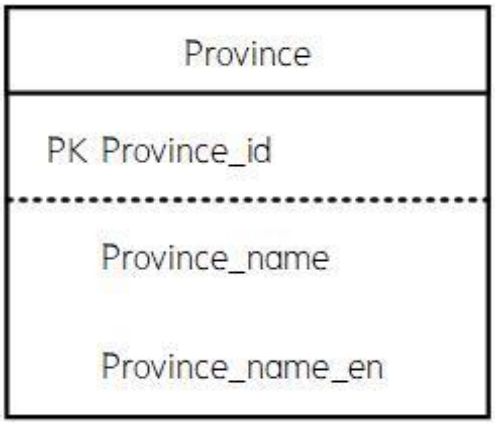
ภาพที่ 3.10 การกำหนดแอตทริบิวต์ Category ในประเด็นที่ 1 ปริมาณขยะมูลฝอย ในประเทศไทย

ดังตัวอย่างประเด็นที่ 2 สถานที่กำจัดขยะ ประกอบด้วย

1) เอนทิตี้ Garbage_location ประกอบด้วย Garbage_location_id, Province_id, District_id, Sub-distrct_id, Category_id, Area size, Volume, year

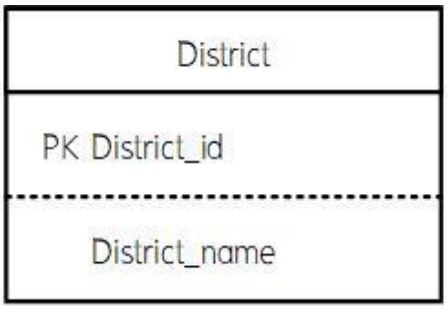
ภาพที่ 3.11 การกำหนดแอตทริบิวต์ Garbage_location ในประเด็นที่ 2
สถานที่กำจัดขยะมูลฝอยในประเทศไทย

2) เอนทิตี้ Province ประกอบด้วย Province_id, Province_name, Province_name_en



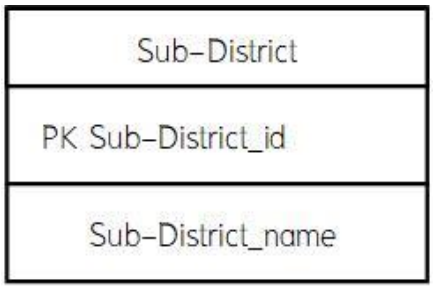
ภาพที่ 3.12 การกำหนดแอตทริบิวต์ Province ในประเด็นที่ 2
สถานที่กำจัดขยะมูลฝอยในประเทศไทย

3) เอนทิตี District ประกอบด้วย District_id, District_name



ภาพที่ 3.13 การกำหนดแอตทริบิวต์ District ในประเด็นที่ 2
สถานที่กำจัดขยะมูลฝอยในประเทศไทย

4) เอนทิตี Sub–District ประกอบด้วย Sub–District_id, Sub–District _Name



ภาพที่ 3.14 การกำหนดแอตทริบิวต์ Sub–District ในประเด็นที่ 2
สถานที่กำจัดขยะมูลฝอยในประเทศไทย

5) เอนทิตี Category ประกอบด้วย Category_id, Category_name

ภาพที่ 3.15 การกำหนดแอตทริบิวต์ Category ในประเด็นที่ 2
สถานที่กำจัดขยะมูลฝอยในประเทศไทย

3.3.3 การกำหนดความสัมพันธ์ระหว่างเอนทิตี้ (Relationships)

การกำหนดแอตทริบิวต์ของเอนทิตี้เป็นการกำหนดคุณลักษณะต่าง ๆ ของแต่ละเอนทิตี้รวมทั้งพิจารณาแอตทริบิวต์ที่จะทำหน้าที่เป็นคีย์หลักของเอนทิตี้ด้วย โดยแบ่งหลักได้อยู่ 3 ประเภท

1) ความสัมพันธ์แบบหนึ่งต่อหนึ่ง (One-to-One Relationship)

เป็นการแสดงความสัมพันธ์ของข้อมูลเอนทิตี้หนึ่งว่ามีความสัมพันธ์กับข้อมูล อย่างมากหนึ่งข้อมูลกับอีกเอนทิตี้หนึ่งในลักษณะที่เป็นหนึ่งต่อหนึ่ง

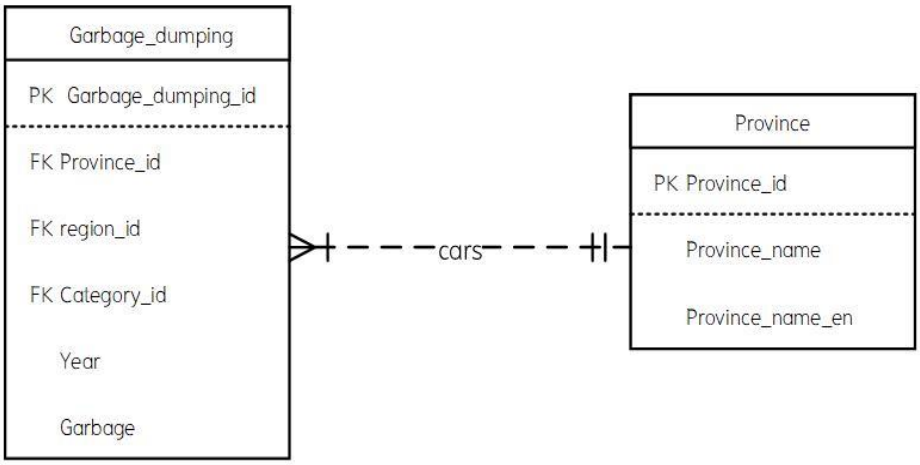
2) ความสัมพันธ์แบบหนึ่งต่อกลุ่ม (One-to-Many Relationship)

เป็นการแสดงความสัมพันธ์ระหว่างข้อมูลของเอนทิตี้หนึ่งว่ามีความสัมพันธ์กับข้อมูลของอีกเอนทิตี้หนึ่ง

3) ความสัมพันธ์แบบกลุ่มต่อกลุ่ม (Many-to-Many Relationship)

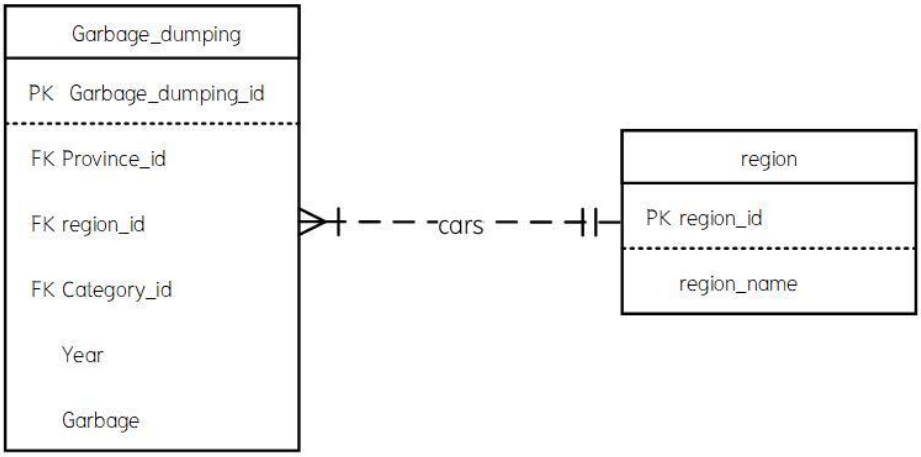
เป็นการแสดงความสัมพันธ์ของข้อมูลระหว่างเอนทิตี้ในลักษณะกลุ่มต่อกลุ่ม เช่น ความสัมพันธ์ของ แผนกกับพนักงาน

3.1) ตาราง Garbage Dumping ในประเด็นที่ 1 ปริมาณขยะมูลฝอยในประเทศไทยที่มีความสัมพันธ์ระหว่างเอนทิตี้กับตาราง Province แบบหนึ่งต่อกลุ่ม (1 : N) หมายความว่า ปริมาณขยะ 1 ปริมาณสามารถมีได้หลายจังหวัด



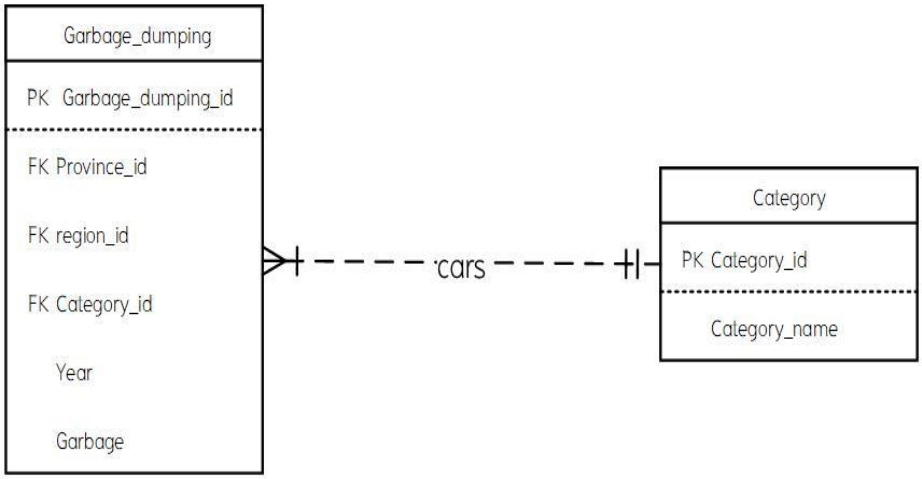
ภาพที่ 3.16 แสดงความสัมพันธ์ระหว่างตาราง Garbage_dumping กับตาราง Province แบบหนึ่งต่อกลุ่ม (1 : N)

3.2) ตาราง Garbage_dumping ในประเด็นที่ 1 ปริมาณขยะมูลฝอยในประเทศไทยที่มีความสัมพันธ์ระหว่างเอนทิตี้กับตาราง region แบบหนึ่งต่อกลุ่ม (1 : N) หมายความว่า ปริมาณขยะ 1 ปริมาณสามารถมีได้หลายๆภาค



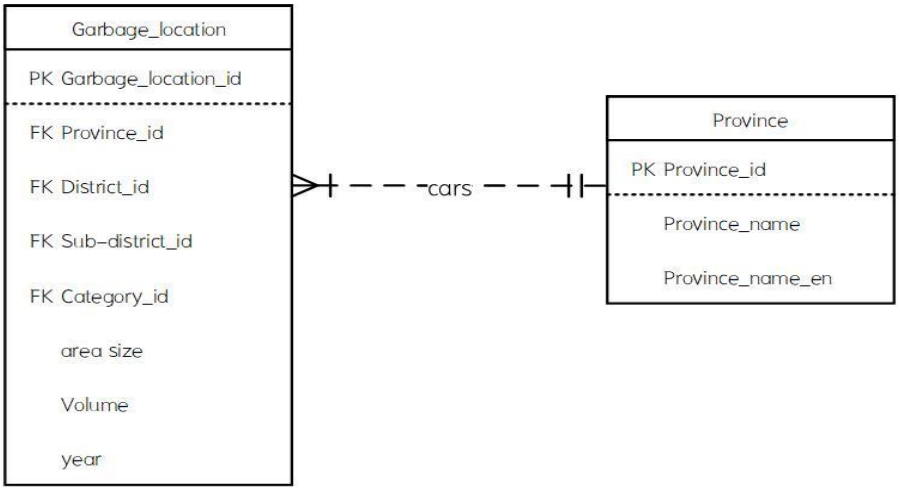
ภาพที่ 3.17 แสดงความสัมพันธ์ระหว่างตาราง Garbage_dumping กับตาราง region แบบหนึ่งต่อกลุ่ม (1 : N)

3.3) ตาราง Garbage_dumping ในประเด็นที่ 1 ปริมาณขยะมูลฝอยในประเทศไทยที่มีความสัมพันธ์ระหว่างเอนทิตี้กับตาราง Category แบบหนึ่งต่อกลุ่ม (1 : N) หมายความว่า ปริมาณขยะ 1 ปริมาณสามารถมีประเภทในการกำจัดได้ขยะได้หลายประเภท



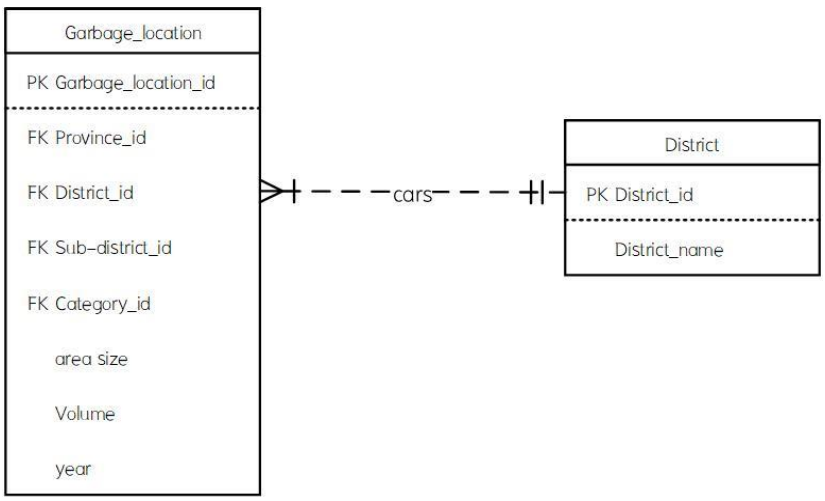
ภาพที่ 3.18 แสดงความสัมพันธ์ระหว่างตาราง Garbage_dumping กับตาราง Category แบบหนึ่งต่อกลุ่ม (1 : N)

3.4) ตาราง Garbage_location ในประเด็นที่ 2 มีความสัมพันธ์ระหว่างเอนทิตี้กับตาราง Province แบบหนึ่งต่อกลุ่ม (1 : N) หมายความว่า สถานที่กำจัด 1 สถานที่สามารถมีจังหวัดได้หลายจังหวัด



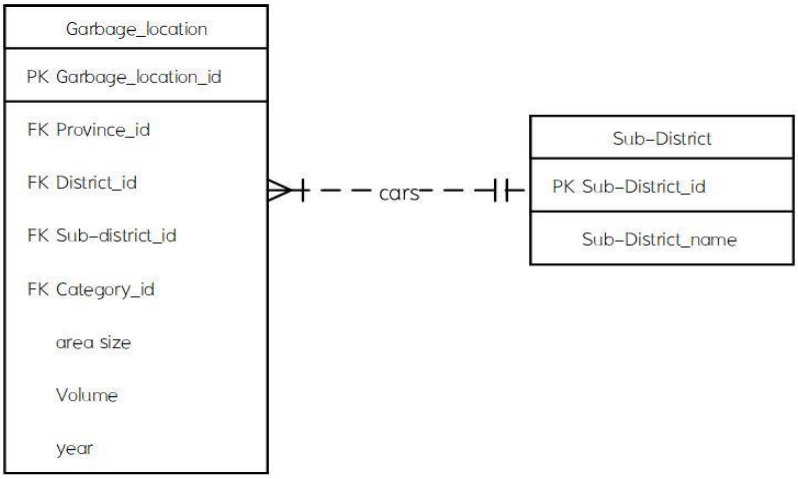
ภาพที่ 3.19 แสดงความสัมพันธ์ระหว่างตาราง Garbage_location กับตาราง Province แบบหนึ่งต่อกลุ่ม (1 : N)

3.5) ตาราง Garbage_location ในประเด็นที่ 2 มีความสัมพันธ์ระหว่างเอนทิตี้กับตาราง District แบบหนึ่งต่อกลุ่ม (1 : N) หมายความว่า สถานที่กำจัด 1 สถานที่กำจัดสามารถมีได้หลายอำเภอ



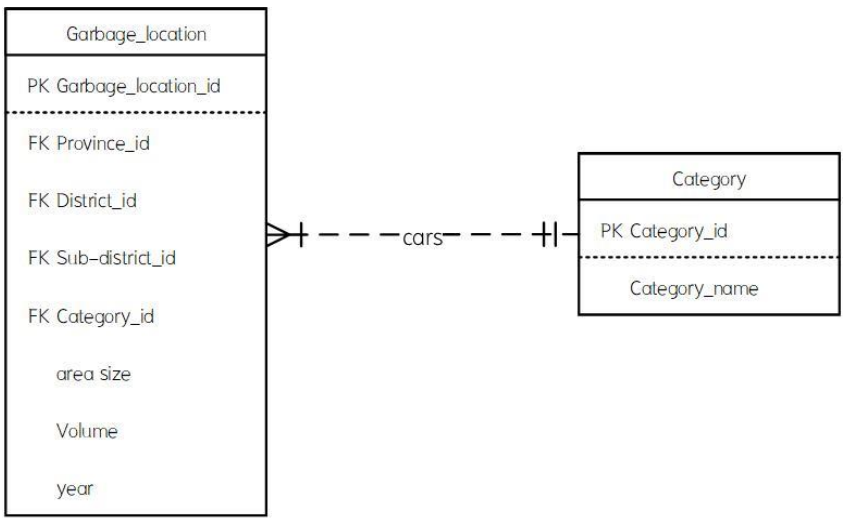
ภาพที่ 3.20 แสดงความสัมพันธ์ระหว่างตาราง Garbage_location กับตาราง District แบบหนึ่งต่อกลุ่ม (1 : N)

3.6) ตาราง Garbage_location ในประเด็นที่ 2 มีความสัมพันธ์ระหว่างเอนทิตี้กับตาราง Sub-District แบบหนึ่งต่อกลุ่ม (1 : N) หมายความว่า สถานที่กำจัด 1 สถานที่กำจัดสามารถมีได้หลายตำบล



ภาพที่ 3.21 แสดงความสัมพันธ์ระหว่างตาราง Garbage_location กับตาราง Sub-District แบบหนึ่งต่อกลุ่ม (1 : N)

5.) ตาราง Garbage location ในประเด็นที่ 2 มีความสัมพันธ์ระหว่างเอนทิตี้กับตาราง Category แบบหนึ่งต่อกลุ่ม (1 : N) หมายความว่า สถานที่กำจัด 1 สถานที่กำจัดสามารถมีประเภทได้หลายประเภท

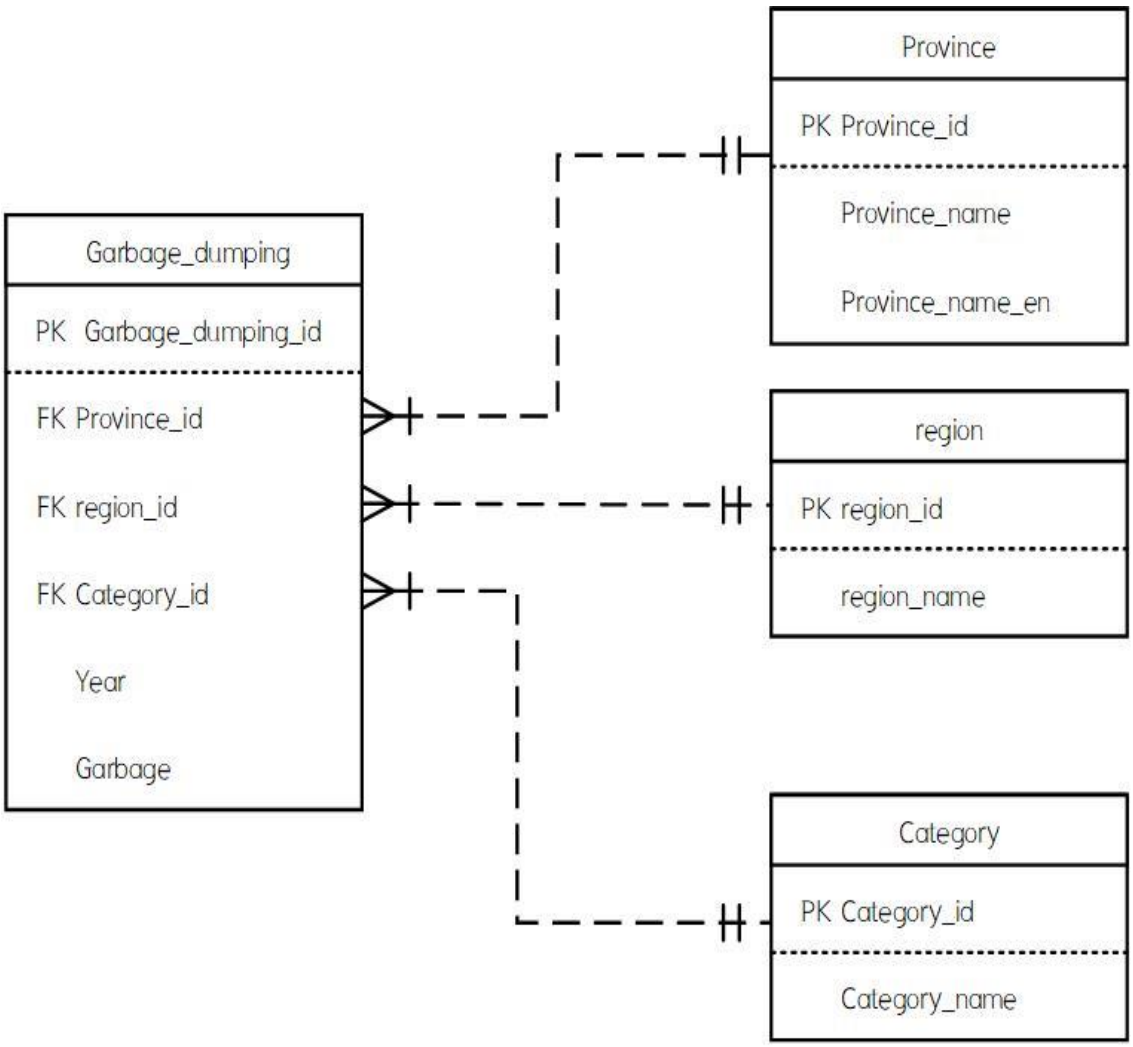


ภาพที่ 3.22 แสดงความสัมพันธ์ระหว่างตาราง Garbage location กับตาราง Category แบบหนึ่งต่อกลุ่ม (1 : N)

3.3.4 การเขียนเป็นแบบจำลองแผนภาพ Entity Relationship Diagram (ERD)

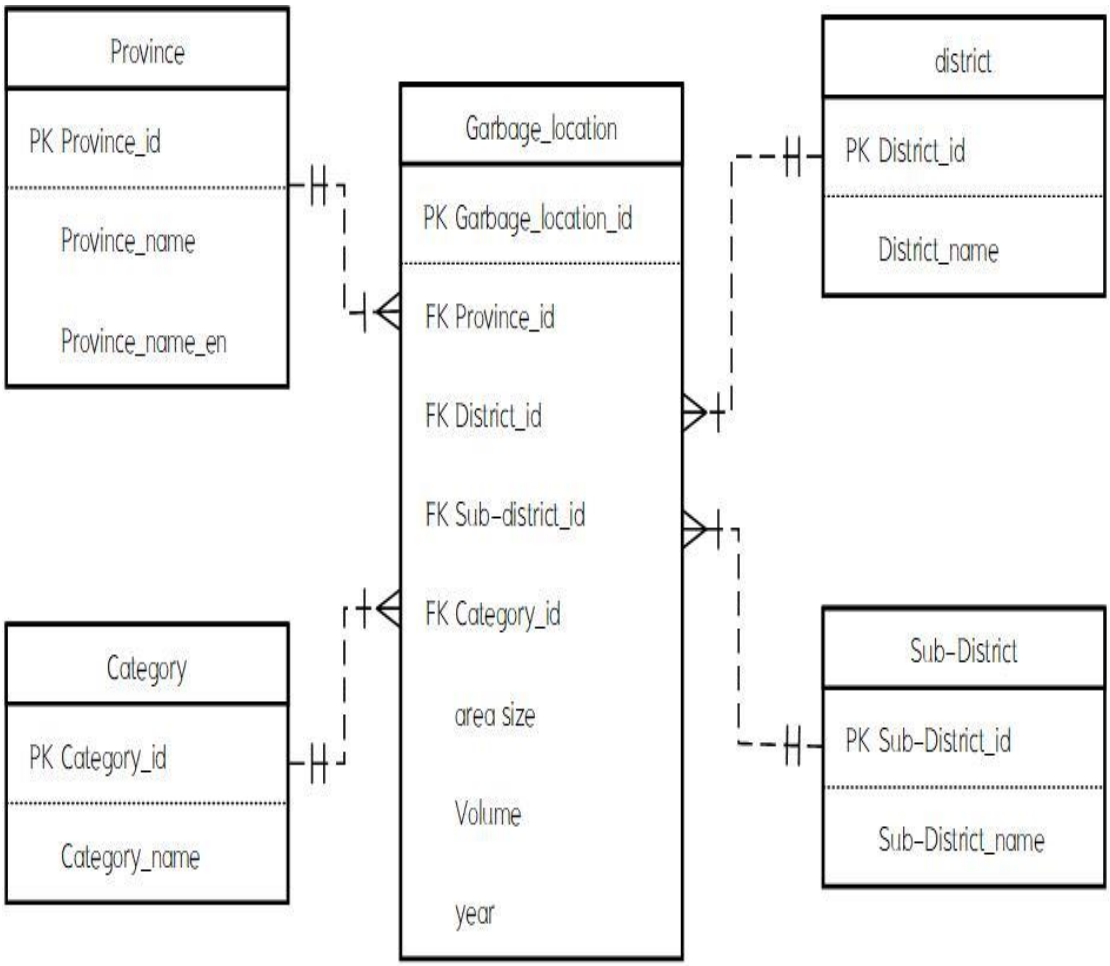
การนำรายละเอียดการออกแบบทั้งหมดในแต่ละขั้นตอน นำมาวาดแบบจำลองแผนภาพ E-R Diagram กำหนดให้ความสัมพันธ์ระหว่างเอนทิตี้ ดังต่อไปนี้

1) การเขียนเป็นแบบจำลองแผนภาพ ในประเด็นที่ 1 Entity Relationship Diagram



ภาพที่ 3.23 แสดงความสัมพันธ์ของฐานข้อมูลแบบ Crow's Foot Model ในประเด็นที่ 1 ปริมาณขยะมูลฝอยในประเทศไทย

2) การเขียนเป็นแบบจำลองแผนภาพ ในประเด็นที่ 2 Entity Relationship Diagram



ภาพที่ 3.24 แสดงความสัมพันธ์ของฐานข้อมูลแบบ Crow's Foot Model ในประเด็นที่ 2 สถานที่กำจัดขยะมูลฝอยในประเทศไทย

3.4 พจนานุกรมข้อมูล (Data Dictionary)

จากการออกแบบฐานข้อมูลซึ่งมีการจัดการระบบฐานข้อมูลให้กับระบบที่ประกอบไปด้วยตารางข้อมูลต่าง ๆ ดังตารางที่ 3.27

ตารางที่ 3.27 แสดงชื่อตารางทั้งหมดของระบบฐานข้อมูล ในประเด็นที่ 1 ปริมาณขยะมูลฝอย

ลำดับ	ชื่อตาราง	ประเภท	รายละเอียด
1	Garbage_dumping	Transaction	เก็บข้อมูลปริมาณขยะ
2	Garbage_location	Transaction	เก็บข้อมูลสถานที่กำจัดขยะ
3	Province	Reference	เก็บข้อมูลจังหวัด
4	region	Reference	เก็บข้อมูลภาค
5	Category	Reference	เก็บข้อมูลประเภทการกำจัดขยะ
6	District	Reference	เก็บข้อมูลอำเภอ
7	Sub-district	Reference	เก็บข้อมูลตำบล

คำอธิบาย ประเภทของตาราง ได้แก่

Transaction หมายถึง ตารางที่มีการเปลี่ยนแปลงของข้อมูล

Reference หมายถึง ตารางที่มีการอ้างอิงถึงข้อมูล

จากตารางที่ 3.28 สามารถแสดงรายละเอียดของแต่ละตารางได้ดังนี้

ตารางที่ 3.28 แสดงรายละเอียดของตาราง Garbage_dumping

ชื่อตาราง : Garbage_dumping
ประเภทตาราง : Transaction
คำอธิบาย : เก็บข้อมูลปริมาณขยะ
คีย์หลัก : Garbage_dumping_id
คีย์รอง : Province_id, region_id, Category_id

เขตข้อมูล	ชนิดและขนาด	ความหมาย	ตัวอย่าง
Garbage_dumping_id	int (11)	รหัสปริมาณขยะ	amg0001
Province_id	Int (11)	รหัสจังหวัด	1001
region_id	VARCHAR (5)	รหัสภาค	P01

Category_id	Int (11)	รหัสประเภทการกำจัดขยะ	001
Year	Year (4)	ปี	2013
garbage	Double	ปริมาณขยะ	31,996.10

ตารางที่ 3.29 แสดงรายละเอียดของตาราง Garbage_location

ชื่อตาราง : Garbage_location
ประเภทตาราง : Transaction
คำอธิบาย : เก็บข้อมูลสถานที่กำจัดขยะ
คีย์หลัก : Garbage_location_id
คีย์รอง : Province_id, District_id, Sub-district_id, Category_id

เขตข้อมูล	ชนิดและขนาด	ความหมาย	ตัวอย่าง
Garbage_location_id	int (11)	รหัสปริมาณขยะ	amg0001
Province_id	int (11)	รหัสจังหวัด	1001
District_id	int (11)	รหัสอำเภอ	001
Sub-district_id	int (11)	รหัสตำบล	2013
Category_id	int (11)	รหัสประเภทการกำจัดขยะ	31,996.10
Area size	Double	ขนาดพื้นที่	13
Volume	Double	ปริมาณขยะ	93.14
Year	Year (4)	ปี	2013

ตารางที่ 3.30 แสดงรายละเอียดของตาราง Province

ชื่อตาราง : Province
ประเภทตาราง : Reference
คำอธิบาย : เก็บข้อมูลจังหวัด
คีย์หลัก : Province_id
คีย์รอง : -

เขตข้อมูล	ชนิดและขนาด	ความหมาย	ตัวอย่าง
Province_id	int (10)	รหัสจังหวัด	1001

Province_name	VARCHAR (100)	ชื่อจังหวัด	กำแพงเพชร
Province_name_en	VARCHAR (100)	ชื่อจังหวัด ภาษาอังกฤษ	Chiang Mai

ตารางที่ 3.31 แสดงรายละเอียดของตาราง District

ชื่อตาราง : District ประเภทตาราง : Reference คำอธิบาย : เก็บข้อมูลประเภทการกำจัดขยะ คีย์หลัก : District_id คีย์รอง : –			
เขตข้อมูล	**ชนิดและขนาด**	**ความหมาย**	**ตัวอย่าง**
District_id	int (11)	รหัสอำเภอ	62170
District_Name	VARCHAR (100)	ชื่ออำเภอ	ลานกระบือ

ตารางที่ 3.32 แสดงรายละเอียดของตาราง Sub-district

ชื่อตาราง : Sub-district ประเภทตาราง : Reference คำอธิบาย : เก็บข้อมูลประเภทการกำจัดขยะ คีย์หลัก : Sub-district_id คีย์รอง : –			
เขตข้อมูล	**ชนิดและขนาด**	**ความหมาย**	**ตัวอย่าง**
Sub-district_id	int (11)	รหัสตำบล	S01
Sub-district_Name	VARCHAR (100)	ชื่อตำบล	ลานกระบือ

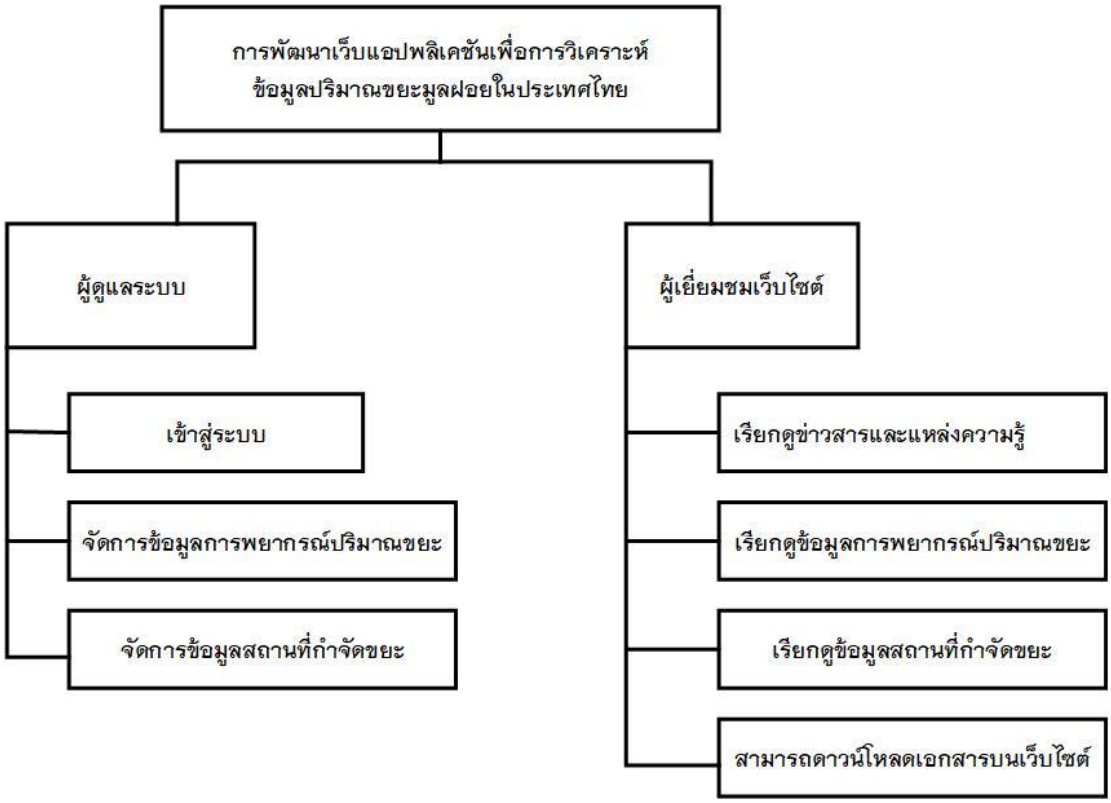
ตารางที่ 3.33 แสดงรายละเอียดของตาราง Category

ชื่อตาราง : Category ประเภทตาราง : Reference คำอธิบาย : เก็บข้อมูลประเภทการกำจัดขยะ คีย์หลัก : Category_id คีย์รอง : –			
เขตข้อมูล	**ชนิดและขนาด**	**ความหมาย**	**ตัวอย่าง**
Category_id	int (5)	รหัสประเภทการกำจัดขยะ	001
Category_Name	VARCHAR (100)	ชื่อประเภทการกำจัดขยะ	ปริมาณขยะมูลฝอยที่กำจัดถูกต้อง (ตัน)

ตารางที่ 3.33 แสดงรายละเอียดของตาราง Category

3.6 โครงสร้างระบบ (System Structure)

ลักษณะภายในของโครงสร้างระบบแสดงถึงโครงสร้างการใช้งานในระบบฐานข้อมูลปริมาณขยะมูลฝอยในประเทศไทย ส่วนที่เป็นส่วนสำคัญโดยตรง ได้แก่ ผู้ดูแลระบบและผู้เยี่ยมชมเว็บไซต์ ดังภาพที่ 3.25



ภาพที่ 3.25 แผนผังโครงสร้างการใช้งานในระบบฐานข้อมูลปริมาณขยะและสถานที่กำจัดขยะ

จากภาพที่ 3.26 เพื่อความสะดวกในการออกแบบฐานข้อมูลผู้วิเคราะห์จึงจัดแบ่งระบบออกเป็นส่วนย่อย ๆ มีทั้งหมด 2 ส่วนด้วยกัน และให้แต่ละส่วนมีหน้าที่รับผิดชอบการทำงานใน แต่ละด้าน โดยไม่คาบเกี่ยวกันแต่สัมพันธ์กัน ส่วนแรก คือส่วนของผู้ดูแลระบบ จะเป็นส่วนที่ สามารถจัดการข้อมูลทั้งหมดในระบบได้และสามารถอัปโหลดข้อมูลบนเว็บไซต์ ส่วนที่สอง คือส่วนผู้เยี่ยมชมเว็บไซต์ เป็นส่วนที่บุคคลภายนอกสามารถเข้ามาเยี่ยมชมข้อมูลบนเว็บไซต์ได้และสามารถดาวน์โหลดข้อมูลบนเว็บไซต์ได้

3.6 กระบวนการ CRISP–DM (Cross Industry Standard Process for Data Mining)

CRISP–DM เป็นกระบวนการหลักในการจัดทำเหมืองข้อมูลเพื่อการวิเคราะห์และใช้ประโยชน์ในทางธุรกิจกระบวนวิเคราะห์ข้อมูลด้วย CRISP–DM หรือ (Cross Industry Standard Process for Data Mining) ประกอบด้วย 6 ขั้นตอน แต่ละขั้นตอนในรูปจะเป็นขั้นตอนที่ต่อเนื่องกันนั่นคือขั้นตอนถัดไปจะรอผลลัพธ์จากขั้นตอนก่อนหน้าซึ่งแสดงด้วยลูกศรที่เชื่อมระหว่างกล่องสี่เหลี่ยมแต่ละกล่อง ตัวอย่างเช่นเมื่อได้ผลลัพธ์จากขั้นตอนการเตรียมข้อมูล (Data Preparation) แล้วจะนำไปสร้างโมเดลจำแนกประเภทข้อมูลในชั้น Modeling และหลังจากนั้นอาจจะย้อนกลับมาเปลี่ยนแปลงข้อมูลให้ถูกต้องมากขึ้นเพื่อหวังว่าโมเดลที่ให้ความถูกต้องมากขึ้นก็ได้ เป็นต้น

ภาพที่ 3.26 แสดงกระบวนการ CRISP–DM

ที่มา : dataminingtrend.com

ในกระบวนการนี้ประกอบด้วย 6 ขั้นตอน

1) Business Understanding เป็นขั้นตอนในกระบวนการ เป็นการแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูล Data Mining พร้อมทั้งวางแผนในการดำเนินการ

2) Data Understanding เริ่มจากการเก็บรวบรวมข้อมูล หลังจากนั้นก็เป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมา เพื่อดูความถูกต้อง และพิจารณาว่าใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์

3) Data Preparation เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมาให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง เช่น แปลงข้อมูลให้อยู่ในช่วงเดียวกัน หรือการเติมข้อมูลข้อมูลที่ขาดหายไป เป็นขั้นตอนที่ใช้เวลามากที่สุดของกระบวนการ CRISP-DM

4) Modeling เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคทาง Data Mining ที่ได้แนะนำไปแล้ว เช่นการจำแนกประเภทข้อมูล หรือการแบ่งกลุ่มข้อมูล

5) Evaluation ได้ผลการวิเคราะห์ข้อมูลด้วยเทคนิคทาง Data Mining แล้ว แต่ก่อนที่จะนำผลลัพธ์ที่ได้ไปใช้งานต้องมีการวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ ที่ได้ตั้งไว้หรือไม่ มีความน่าเชื่อถือเพียงใด

6) Deployment มีการนำความรู้ที่ได้จากการได้ผลลัพธ์ด้วยเทคนิค Data Mining ไปใช้ประโยชน์ต่อในองค์กร

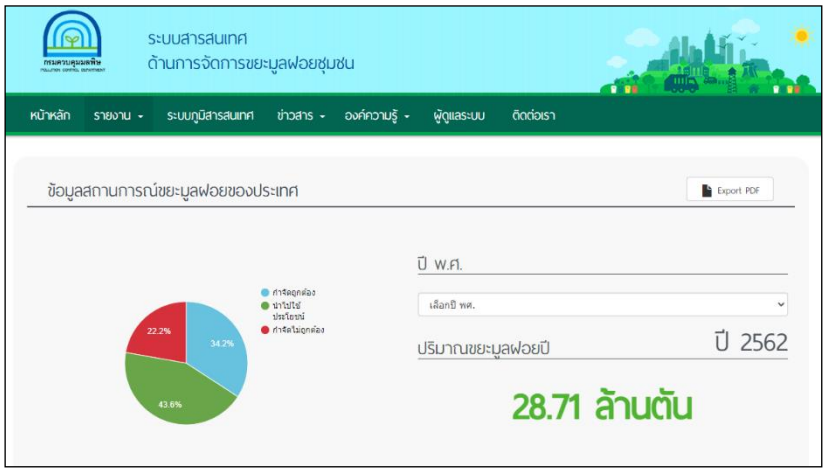
3.1.1 ความเข้าใจในธุรกิจ (Business Understanding)

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://thaimsw.pcd.go.th ซึ่งเป็นระบบสารสนเทศของกรมควบคุมมลพิษ ผู้วิเคราะห์ข้อมูลจึงได้ ดำเนินการตามขั้นตอนดังนี้

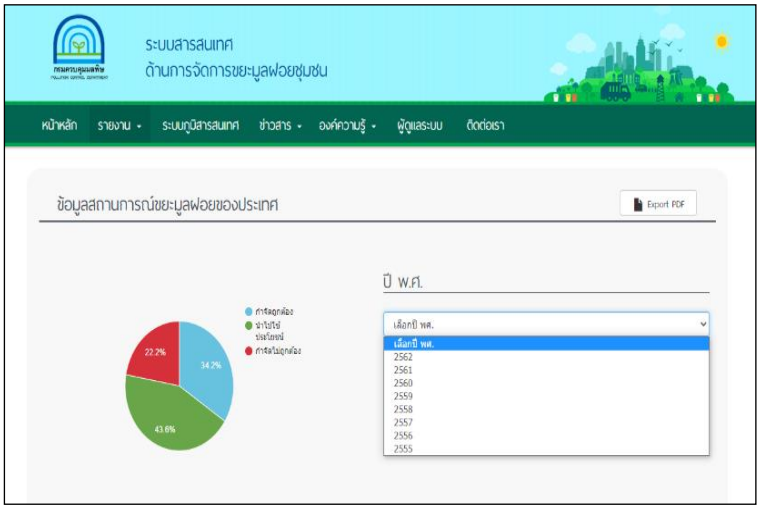
1) ผู้วิเคราะห์รวบรวมข้อมูลจาก จากเว็บไซต์ https://thaimsw.pcd.go.th ผู้วิเคราะห์ทำการเลือกที่รายงาน ข้อมูลสถานการ์ณขยะมูลฝอยของประเทศ



ภาพที่ 3.27 แสดงขั้นตอนการหาข้อมูลจากเว็บไซต์

ที่มา : https://thaimsw.pcd.go.th

2) ผู้วิเคราะห์ทำการจัดเก็บ และรวบรวมข้อมูลทั้งหมด ของข้อมูลปริมาณขยะมูลฝอยในช่วงปี พ.ศ. 2556 – 2562



ภาพที่ 3.28 แสดงขั้นตอนการดาวน์โหลดข้อมูล

ที่มา : https://thaimsw.pcd.go.th

3) ตรวจสอบความถูกต้อง และความน่าเชื่อถือของข้อมูล ซึ่ง 1 ไฟล์ จะประกอบด้วย 5 แอตทริบิวท์ คือ จังหวัด, ปริมาณขยะมูลฝอยที่เกิดขึ้น, ปริมาณขยะมูลฝอยที่นำไปใช้ประโยชน์, ปริมาณขยะมูลฝอยที่กำจัดถูกต้อง, ปริมาณขยะมูลฝอยที่กำจัดไม่ถูกต้อง

จังหวัด	ปริมาณขยะมูลฝอยที่ เกิดขึ้น (ตัน)	ปริมาณขยะมูลฝอยที่ นำไปใช้ไประโยชน์ (ตัน)	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)
กระบี่	200,593.05	47,628.85	50,534.25	102,429.95
กรุงเทพมหานคร	7,446,000.00	478,150.00	3,244,850.00	3,723,000.00
กาญจนบุรี	250,495.85	7,921.96	8,068.69	234,505.20
กาฬสินธุ์	348,216.45	126,883.49	30,638.10	190,694.86
กำแพงเพชร	256,739.86	55,247.64	89,498.00	111,994.22
ขอนแก่น	682,514.91	182,613.53	49,869.95	450,031.43
จันทบุรี	220,843.93	21,520.51	78,777.95	120,545.46
ฉะเชิงเทรา	314,406.09	35,349.63	6,424.00	272,632.46
ชลบุรี	907,884.66	79,607.49	397,120.00	431,157.17
ชัยนาท	129,691.80	54,987.25	27,353.10	47,351.45
ชัยภูมิ	396,644.01	108,705.68	3,650.00	284,288.32
ชุมพร	133,221.35	3,051.40	10,658.00	119,511.95
ตรัง	242,670.25	88,476.00	7,730.70	146,463.55
ตราด	97,715.20	18,583.80	30,039.50	49,091.90
ตาก	233,519.00	48,384.73	10,220.00	174,914.27
นครนายก	93,570.36	20,221.15	34,025.30	39,323.92

ภาพที่ 3.29 ไฟล์ข้อมูลการบันทึกปริมาณขยะมูลฝอยรายปี

ที่มา : https://thaimsw.pcd.go.th

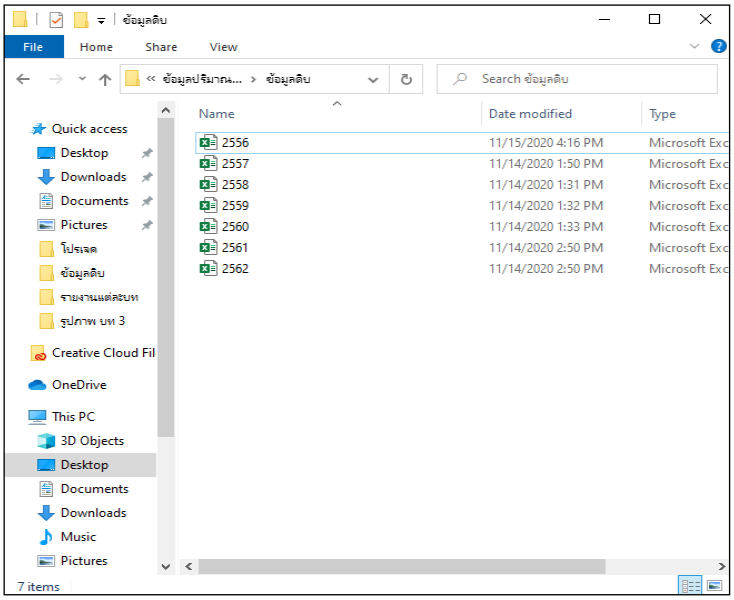
ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมา ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไปวิเคราะห์ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) มักใช้เวลาค่อนข้างมาก ระบบการรับข้อมูลป้อนเข้าสู่ระบบในปัจจุบันจะลดการคีย์ข้อมูลจากคนให้น้อยที่สุด แต่จะใช้วิธีการสแกน การติ๊กเลือก เพื่อลดความผิดพลาดให้น้อยที่สุด เพราะขั้นตอนใช้เวลามากการลดข้อผิดพลาดของข้อมูลได้มากเท่าใด ก็จะยิ่งมีประสิทธิภาพมากขึ้นเท่านั้น

3.1.3 การเตรียมข้อมูล (Data Preparation)

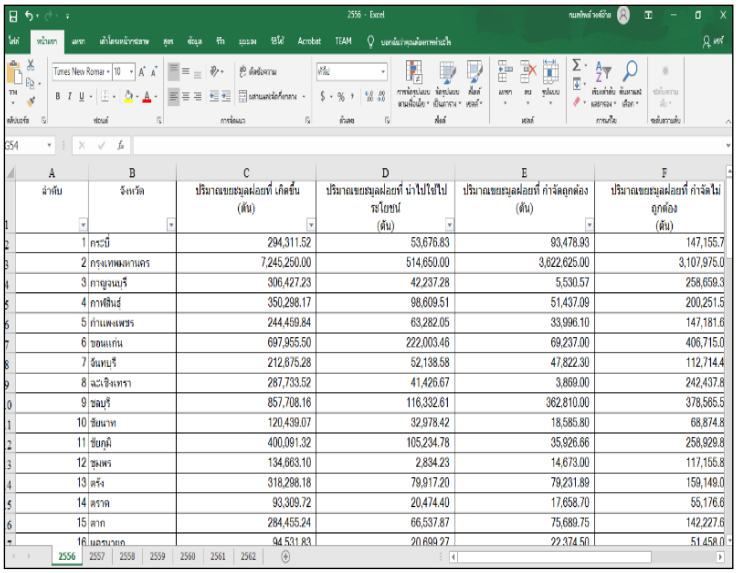
ผู้วิเคราะห์ข้อมูลทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวม และทำการจัดการข้อมูลที่ได้ให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ เพื่อลดความผิดพลาดการทำข้อมูล

ให้ถูกต้อง โดยใช้กระบวนการ data cleaning เพื่อให้ได้ข้อมูลอยู่ในรูปแบบและมาตรฐานเดียวกัน และกำจัดข้อมูลเสียออก ผู้วิเคราะห์ข้อมูลจึงได้ดำเนินการตามขั้นตอนดังนี้

1) การปรับมาตรฐาน (Standardizing) ผู้วิเคราะห์ข้อมูลทำการรวบแปลงไฟล์ข้อมูลให้เป็นไฟล์เดียวกันอยู่ในรูปแบบเดียวกันพร้อมทั้งสร้างแอททริบิวต์เพื่อกำหนดคีย์หลักของแต่ละไฟล์ขึ้นมาเพื่อเป็นตัวที่ใช้ระบุความสัมพันธ์ของข้อมูลที่จะแสดงปริมาณขยะมูลฝอยในประเทศไทย



ภาพที่ 3.30 ไฟล์ข้อมูลปริมาณขยะมูลฝอยทั้งหมดที่ได้จากเว็บไซต์

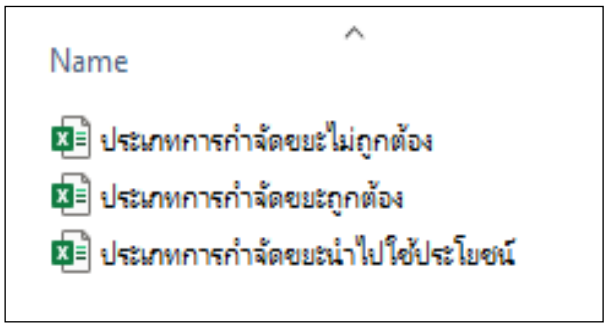


ลำดับ	จังหวัด	ปริมาณขยะมูลฝอยที่ เกิดขึ้น (ตัน)	ปริมาณขยะมูลฝอยที่ นำไปใช้ไป รอโยชน์ (ตัน)	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)
1	กระบี่	294,311.52	53,676.83	93,478.93	147,155.7
2	กรุงเทพมหานคร	7,245,250.00	514,650.00	3,622,625.00	3,107,975.0
3	กาญจนบุรี	306,427.23	42,237.28	5,530.57	258,659.3
4	กาฬสินธุ์	350,298.17	98,609.51	51,437.09	200,251.5
5	กำแพงเพชร	244,459.84	63,282.05	33,996.10	147,181.6
6	ขอนแก่น	697,955.50	222,003.46	69,237.00	406,715.0
7	จันทบุรี	212,675.28	52,138.58	47,822.30	112,714.4
8	ฉะเชิงเทรา	287,733.52	41,426.67	3,869.00	242,437.8
9	ชลบุรี	857,708.16	116,332.61	362,810.00	378,565.5
10	ชัยนาท	120,439.07	32,978.42	18,585.80	68,874.8
11	ชัยภูมิ	400,091.32	105,234.78	35,926.66	258,929.8
12	ชุมพร	134,663.10	2,834.23	14,673.00	117,155.8
13	เชียงราย	318,298.18	79,917.20	79,231.89	159,149.0
14	ตราด	93,309.72	20,474.40	17,658.70	55,176.6
15	ตาก	284,455.24	66,537.87	75,689.75	142,227.6
16	นครนายก	94,531.83	20,699.27	22,374.50	51,458.0

ภาพที่ 3.31 ข้อมูลที่ปรับมาตรฐานรวมให้เป็นรายปีในแต่ละจังหวัดในประเทศไทย

2) การจัดหมวดหมู่ (Transfrom) ผู้วิเคราะห์นำ dataset ที่ผ่านการรวบรวมรายปีมาจัดหมวดหมู่เพื่อความถูกต้อง ผู้วิเคราะห์ดำเนินการจัดกลุ่มข้อมูลแบ่งตามภูมิภาค โดยการสร้างแอตทริบิวท์เพิ่มขึ้น 1 ตัว คือ แอตทริบิวท์ region (ภูมิภาค) เพื่อแบ่งหมวดหมู่และแอตทริบิวท์ที่กำหนดให้ Province_name เป็นคีย์หลักเพื่อเป็นตัวที่ใช้ระบุความสัมพันธ์ของข้อมูลที่จะแสดงข้อมูลภาคต่าง ๆ ในแต่ละปี เตรียมความพร้อมการนำข้อมูลไปแสดงบนโปรแกรม Minitab และโปรแกรม Tableau Public โดยดำเนินตามขั้นตอนดังนี้

2.1) การจัดกลุ่มตามจังหวัดโดยแบ่งเป็นประเภทของการกำจัดขยะ



ภาพที่ 3.32 ไฟล์ข้อมูลที่ต้องการจัดกลุ่มให้รวมอยู่ในแต่ละประเภทของการกำจัดขยะมูลฝอย

2.2) นำข้อมูลที่ผ่านการรวบรวมเป็นรายจังหวัดที่แบ่งเป็นตามประเภท

	C1-T	C2-T	C3-T	C4-T	C5
	Province_Name	Ragion	Year	Category	garbage
1	กำแพงเพชร	ภาคเหนือ	2013	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	43996
2	กำแพงเพชร	ภาคเหนือ	2014	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	62123
3	กำแพงเพชร	ภาคเหนือ	2015	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	77498
4	กำแพงเพชร	ภาคเหนือ	2016	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	55159
5	กำแพงเพชร	ภาคเหนือ	2017	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	53151
6	กำแพงเพชร	ภาคเหนือ	2018	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	52239
7	กำแพงเพชร	ภาคเหนือ	2019	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	57772

Sheet1

ภาพที่ 3.33 แสดงการจัดกลุ่มรายประเภทการกำจัดขยะมูลฝอย

จากรูปภาพที่ 3.26 ประกอบไปด้วย 5 แอตทริบิวท์ คือ

1) Province_name เก็บข้อมูลชื่อจังหวัด

2) region เก็บข้อมูลภาค

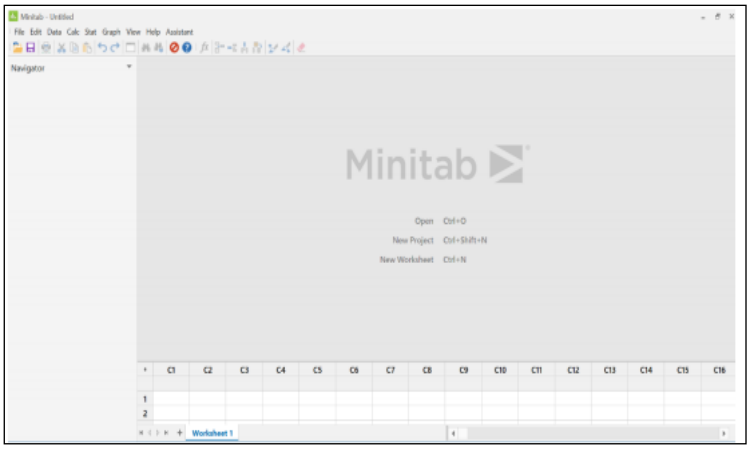
3) Year เก็บข้อมูลรายปี

4) Category เก็บข้อมูลประเภทการกำจัดขยะ

5) garbage เก็บข้อมูลปริมาณขยะ

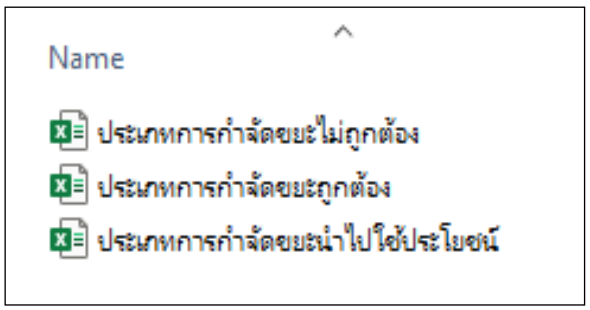
3) การอัปโหลดข้อมูล (Loading) โดยใช้ Minitab ดำเนินกระบวนการ การโหลดข้อมูลเข้าระบบฐานข้อมูลของ Minitab เพื่อเป็นการพยากรณ์ของข้อมูล โดยมีขั้นตอนดำเนินการดังนี้

3.1) โหลดข้อมูลที่ได้เตรียมไว้แล้วลงโปรแกรม Minitab



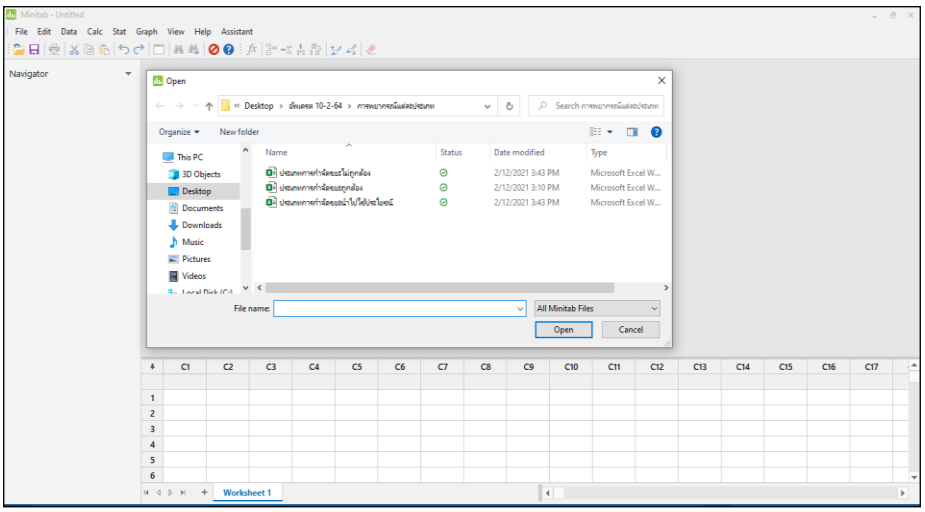
ภาพที่ 3.34 หน้าต่างพื้นที่การทำงานของโปรแกรม Minitab

3.2) จัดเตรียมข้อมูลที่นำมาใช้ในการพยากรณ์



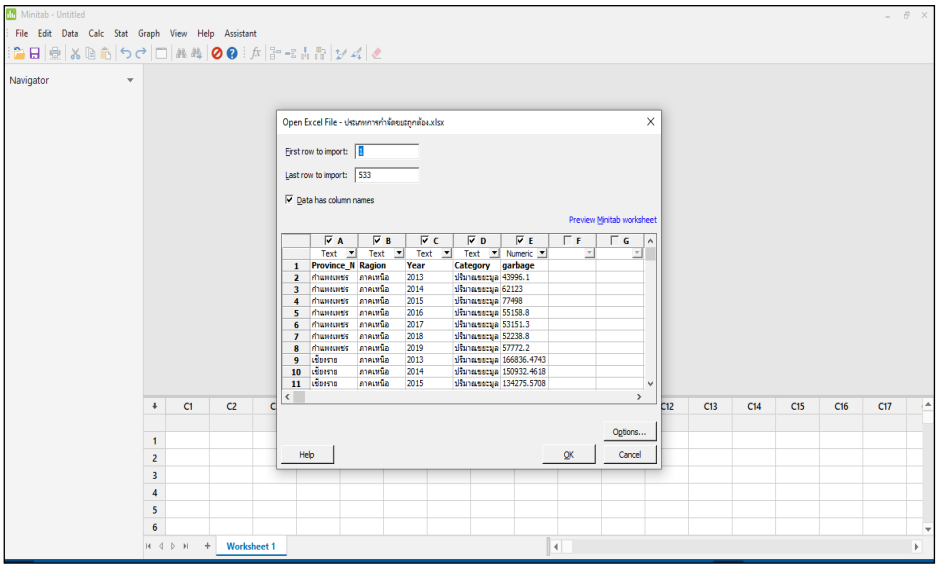
ภาพที่ 3.35 ข้อมูลที่นำมาใช้ในการพยากรณ์แต่ละประเภท

3.3) ใช้ฟังก์ชัน File open เพื่อทำการนำข้อมูลเข้าฐานข้อมูลรูปแบบของ Microsoft Excel และเลือกข้อมูลที่ผ่านการจัดกลุ่มข้อมูลแล้ว และนำข้อมูลเข้าโปรแกรม



ภาพที่ 3.36 เลือกไฟล์ที่ต้องการนำข้อมูลมาพยากรณ์

3.4) ทำการ Import ข้อมูลทั้งหมดของแต่ละภาคเข้าโปรแกรมเพื่อทำการหาค่าผลลัพธ์ของการพยากรณ์



ภาพที่ 3.37 Import ข้อมูลเพื่อหาค่าการพยากรณ์

3.5) ข้อมูลที่ถูก Import ลงโปรแกรมจะมีทั้งหมด 5 แอตทริบิวท์ ได้แก่ Province_Name, Region, year, Category และจะนำแอตทริบิวท์ garbage มาคำนวณหาค่าพยากรณ์ด้วยวิธีต่างๆ

	C1-T	C2-T	C3-T	C4-T	C5
	Province_Name	Ragion	Year	Category	garbage
1	กำแพงเพชร	ภาคเหนือ	2013	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	43996
2	กำแพงเพชร	ภาคเหนือ	2014	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	62123
3	กำแพงเพชร	ภาคเหนือ	2015	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	77498
4	กำแพงเพชร	ภาคเหนือ	2016	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	55159
5	กำแพงเพชร	ภาคเหนือ	2017	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	53151
6	กำแพงเพชร	ภาคเหนือ	2018	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	52239
7	กำแพงเพชร	ภาคเหนือ	2019	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	57772

Sheet1

ภาพที่ 3.38 แอตทริบิวท์ที่นำมาคำนวณหาค่าพยากรณ์

3.1.4 การสร้างแบบจำลอง (Modeling)

ผู้วิเคราะห์ข้อมูลจะวิเคราะห์ด้วยเทคนิคทางสถิติ ทำการเรียนรู้จากข้อมูลเดิมของปริมาณขยะมูลฝอย การเปรียบเทียบความคลาดเคลื่อนของวิธีการพยากรณ์ โดยนำโมเดล Time series analysis มาช่วยในการหาค่าความคลาดเคลื่อนของการพยากรณ์ในวิธีต่าง ๆ โดยมีขั้นตอนการดำเนินการดังนี้

ในการพยากรณ์โดยใช้รูปแบบอนุกรมเวลาจะเป็นการพยากรณ์ข้อมูลในอนาคตจากข้อมูลในอดีตเท่านั้น ตัวแปรอื่น ๆ จะไม่นำมาพิจารณา ผู้วิเคราะห์ข้อมูลได้เลือกข้อมูลในรูปแบบมีแนวโน้ม (Trend)

	C1-T	C2-T	C3-T	C4-T	C5
	Province_Name	Ragion	Year	Category	garbage
1	กำแพงเพชร	ภาคเหนือ	2013	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	43996
2	กำแพงเพชร	ภาคเหนือ	2014	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	62123
3	กำแพงเพชร	ภาคเหนือ	2015	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	77498
4	กำแพงเพชร	ภาคเหนือ	2016	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	55159
5	กำแพงเพชร	ภาคเหนือ	2017	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	53151
6	กำแพงเพชร	ภาคเหนือ	2018	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	52239
7	กำแพงเพชร	ภาคเหนือ	2019	ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	57772

Sheet1

ภาพที่ 3.39 ตารางข้อมูลปริมาณขยะมูลฝอยโดยแบ่งตามจังหวัดและประเภทการกำจัดขยะ

การสร้างโมเดล Time Series Models จะทำการเลือกแอททริบิวต์ข้อมูลเชิงปริมาณที่จัดเก็บไว้ มาคำนวณหาค่าความเป็นไปได้เชิงพยากรณ์ โดยการคำนวณจากทางคณิตศาสตร์ดังนี้

ขั้นตอนที่ 1 คำนวณหาค่า Smoothing Factor จากสมการ

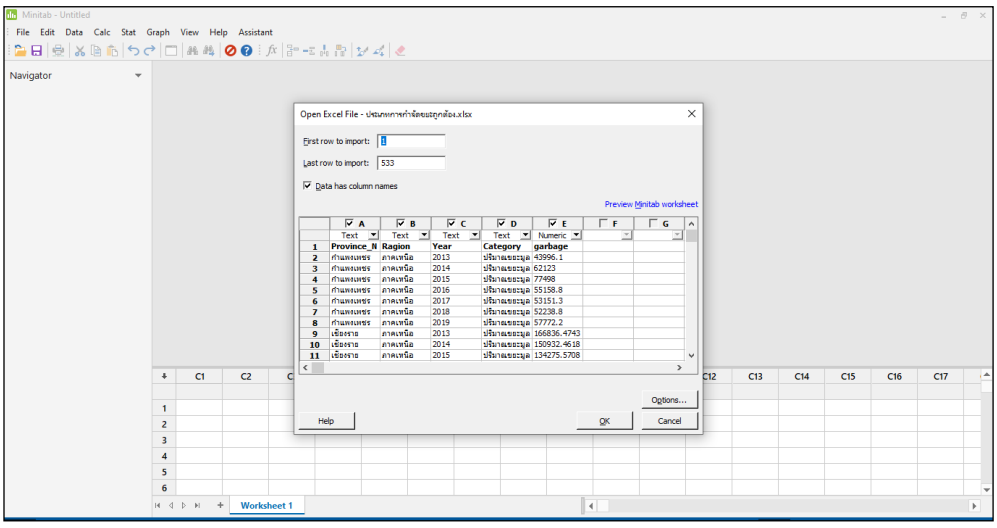
$$2 / (n + 1)$$

ขั้นตอนที่ 2 คำนวณ EMA

$$EMA_n = aP_n + EMA_{n-1}(1 - a)$$

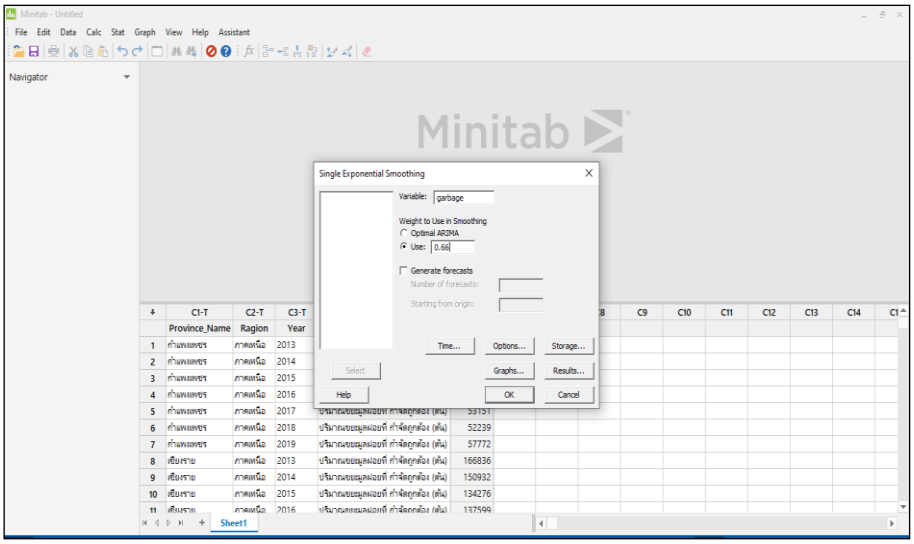
1) การใช้เทคนิคการวิเคราะห์อนุกรมเวลา ด้วยวิธีการพยากรณ์ค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก โดยผลลัพธ์จากการพยากรณ์ผู้วิเคราะห์ข้อมูลได้ข้อมูลที่ผ่านการวิเคราะห์ข้อมูลจากการสร้างโมเดล Time Series Models จึงนำข้อมูลดังกล่าว มาทดลองกับโปรแกรม Minitab ซึ่งมีขั้นตอนการทำงานดังนี้

ขั้นตอนที่ 1 ใช้โปรแกรม Minitab เลือกคำสั่ง file –> open –>เลือกไฟล์ที่จะนำเข้า



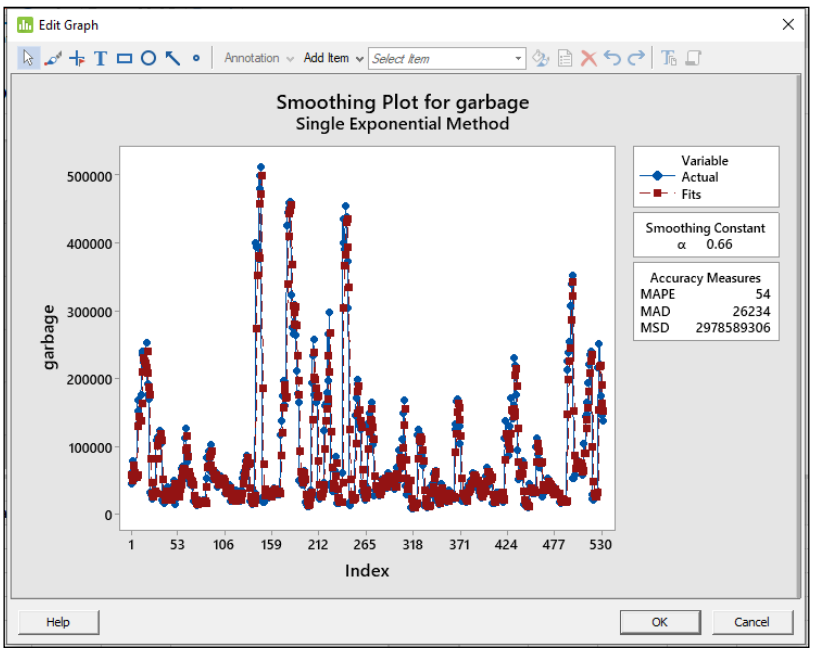
ภาพที่ 3.40 แสดงขั้นตอนการนำข้อมูลเข้าโปรแกรม Minitab

ขั้นตอนที่ 2 คลิกเลือกคำสั่งที่แถบเมนู stat –> time series –> single Exponential Smoothing –> ok



ภาพที่ 3.41 แสดงขั้นตอนการกำหนดรูปแบบการพยากรณ์วิธี Single EMA

ขั้นตอนที่ 3 ผลลัพธ์ที่แสดงการพยากรณ์แล้วทำการแทนค่าสมการตามสูตรอนุกรมเวลาวิธีการ Single Exponential Moving Average ในการพยากรณ์จะมีความคลาดเคลื่อน MAPE 54% ในภาพของแต่ละประเภทการกำจัดขยะมูลฝอย



ภาพที่ 3.42 แสดงผลลัพธ์กราฟวิธี Single EMA

การพยากรณ์โดยการวิเคราะห์อนุกรมเวลาของการทำเหมืองข้อมูล ผ่านโปรแกรม โดยผลลัพธ์ที่ได้จากการพยากรณ์ล่วงหน้าและหลักเกณฑ์ในการเลือกค่าตรวจสอบนั้น มีหลักเกณฑ์ปฏิบัติคือ ค่า MAPE ส่วนใหญ่แล้วใช้เปรียบเทียบความแม่นยำของค่าพยากรณ์ของข้อมูลทางธุรกิจ ทั้งนี้ค่า MAPE เป็นค่าวัดเปรียบเทียบดังนั้นค่า MAPE จะนิยมกว่าค่า MAD ซึ่งจากการศึกษาพบว่าถ้าค่า MAPE ที่สามารถเชื่อถือได้ดังตารางที่ 3.20

ตารางที่ 3.23 ตารางเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE)

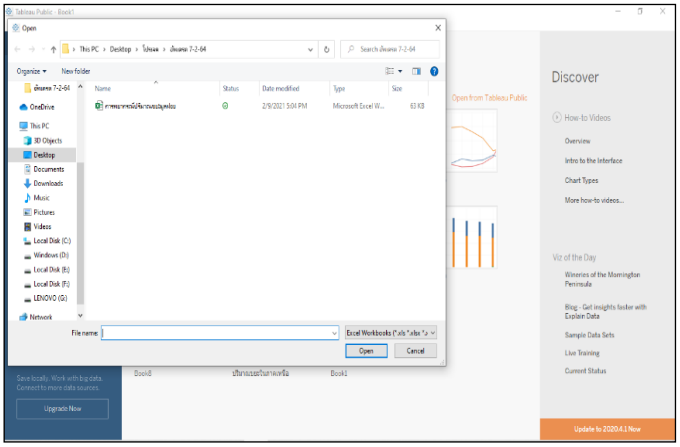
ค่า MAPE	**ความน่าเชื่อถือ**
น้อยกว่า 10%	จัดว่าการพยากรณ์ค่อนข้างแม่นยำ
10% ถึง 20%	จัดว่าการพยากรณ์ใช้ได้ดี
20% ถึง 50%	จัดว่าการพยากรณ์พอใช้
มากกว่า 50%	จัดว่าการพยากรณ์ไม่แม่นยำ

ตารางที่ 3.34 ตารางเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE)

ที่มา : รศ.สุพรรณี อิ้งปัญสัตวงศ์ (2563) , 12 ธันวาคม 2563.

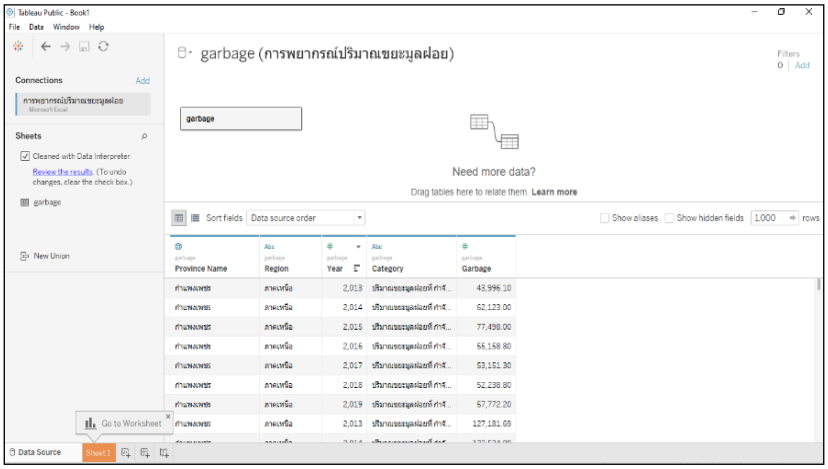
2) การพยากรณ์โดยการวิเคราะห์อนุกรมเวลาของข้อมูล (Data mining time series analysis) ผ่านโปรแกรม Tableau Public โดยผลลัพธ์จากการพยากรณ์ล่วงหน้า ผู้วิเคราะห์ข้อมูลได้ข้อมูลที่ผ่านการวิเคราะห์ข้อมูลจึงนำข้อมูลดังกล่าวมาทดสอบกับโปรแกรม Tableau Public ซึ่งมีขั้นตอนการทำงานดังนี้

2.1) หน้าต่างที่ทำการเชื่อมต่อฐานข้อมูลไฟล์ Excel เพื่อนำมาวิเคราะห์



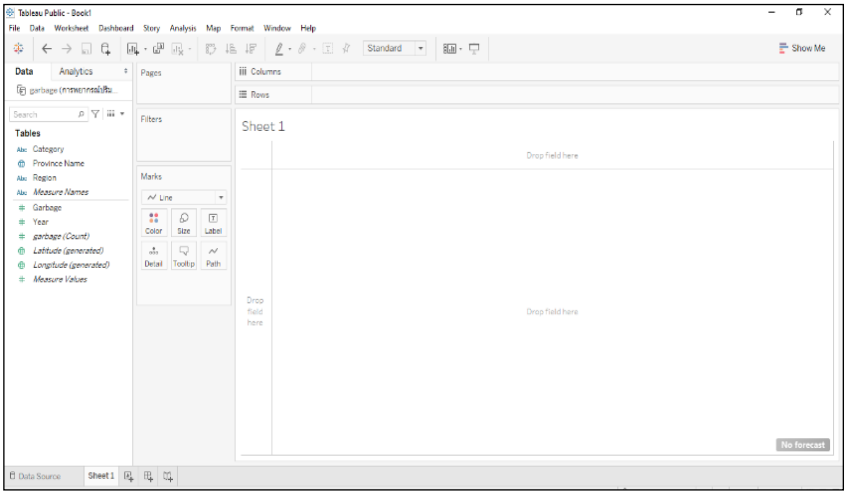
ภาพที่ 3.43 หน้าต่างที่ทำการเชื่อมต่อฐานข้อมูลไฟล์ Excel เพื่อนำมาวิเคราะห์

2.2) ทำการใช้เครื่องมือทำเหมืองข้อมูล (Data Mining) ของโปแกรม Tableau Public ในการพยากรณ์ข้อมูล



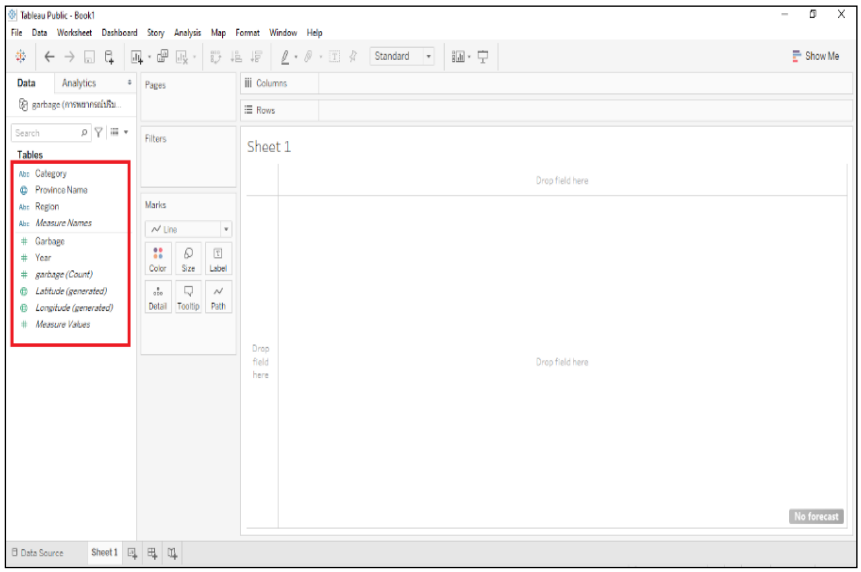
ภาพที่ 3.44 Import แสดงผลการโหลดฐานข้อมูลเข้าสู่โปรแกรม Tableau Public

2.3) หน้าต่างพื้นที่การทำงานของโปรแกรม Tableau Public



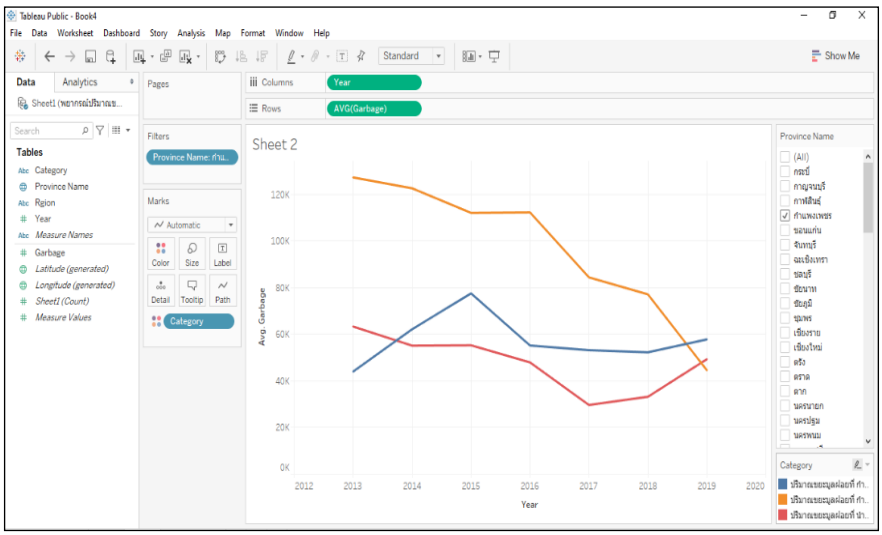
ภาพที่ 3.45 หน้าต่างพื้นที่ทำงานของโปรแกรม tableau

2.4) หน้าต่างที่แสดงข้อมูลที่ Import ข้อมูลเข้ามาแล้วจะมีข้อมูลอยู่แถบด้านซ้ายมือ



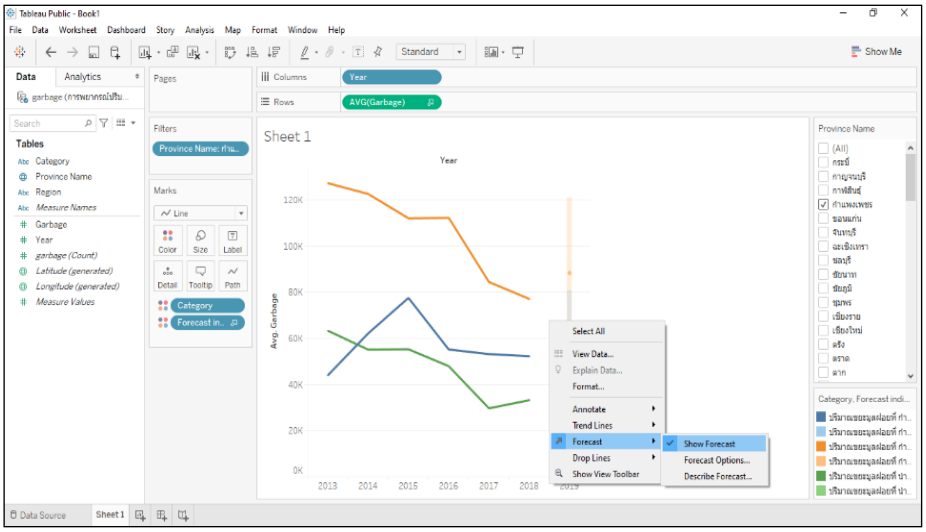
ภาพที่ 3.46 หน้าต่างที่แสดงข้อมูลที่ Import ข้อมูล

2.5) การเตรียมกราฟข้อมูลก่อนการวิเคราะห์ข้อมูล



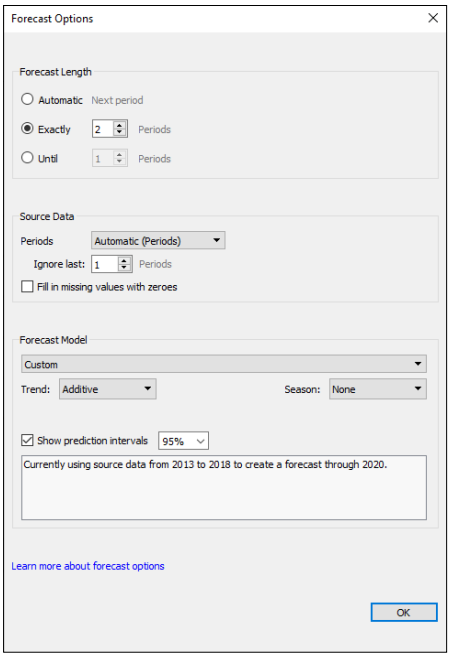
ภาพที่ 3.47 แสดงการเตรียมกราฟข้อมูลก่อนการวิเคราะห์ข้อมูล

2.6) เครื่องมือในการวิเคราะห์ข้อมูลของโปรแกรม Tableau Public



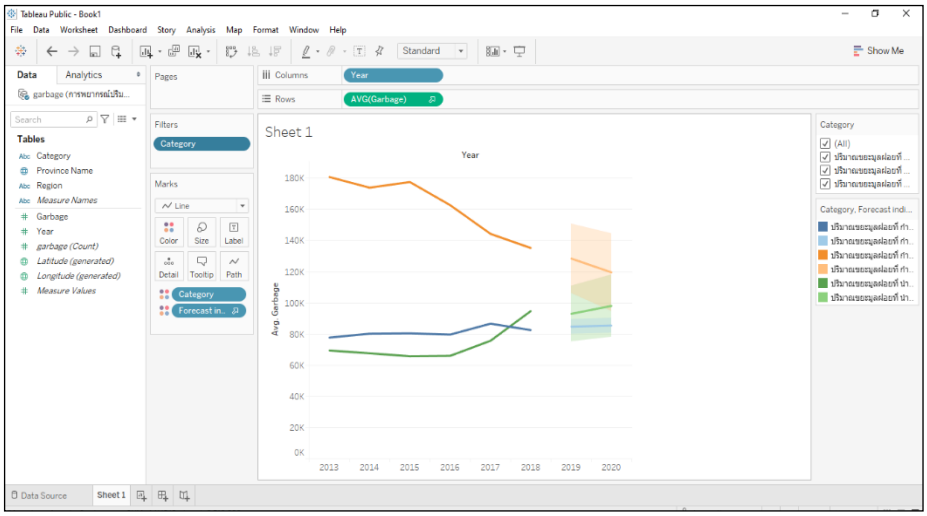
ภาพที่ 3.48 การแสดงเครื่องมือในการวิเคราะห์

2.7) การใช้เครื่องมือทำเหมืองข้อมูล (Data Mining) ของโปแกรม Tableau Public เพื่อพยากรณ์ข้อมูลปริมาณขยะมูลฝอยในช่วงเวลา 1 ปีถัดไปโดยใช้ Exactly ในการกำหนดค่าการพยากรณ์



ภาพที่ 3.49 การแสดงเครื่องมือในการพยากรณ์

2.8) ผลลัพธ์ในการพยากรณ์ช่วงเวลา 2 ปีถัดไป



ภาพที่ 3.50 การแสดงผลลัพธ์ในการพยากรณ์ช่วงเวลา 2 ปีถัดไป

2.9) ผลการคำนวณค่าความคลาดเคลื่อน (MAPE)

Describe Forecast

Summary | Models

All forecasts were computed using exponential smoothing.

Avg. Garbage

Color	Model			Quality Metrics					Smoothing Coefficients		
Category	Level	Trend	Season	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
ปริมาณขยะมูลฝอยที่ นำไปใช้ประโยชน์ (ตัน)	Additive	Additive	None	9,095	7,814	1.19	10.7%	119	0.500	0.000	0.000
ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)	Additive	Additive	None	11,351	9,904	0.94	5.9%	122	0.500	0.013	0.000
ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	Additive	Additive	None	2,405	2,207	0.75	2.7%	103	0.072	0.500	0.000

ภาพที่ 3.51 ผลการคำนวณค่าความคลาดเคลื่อน (MAPE)

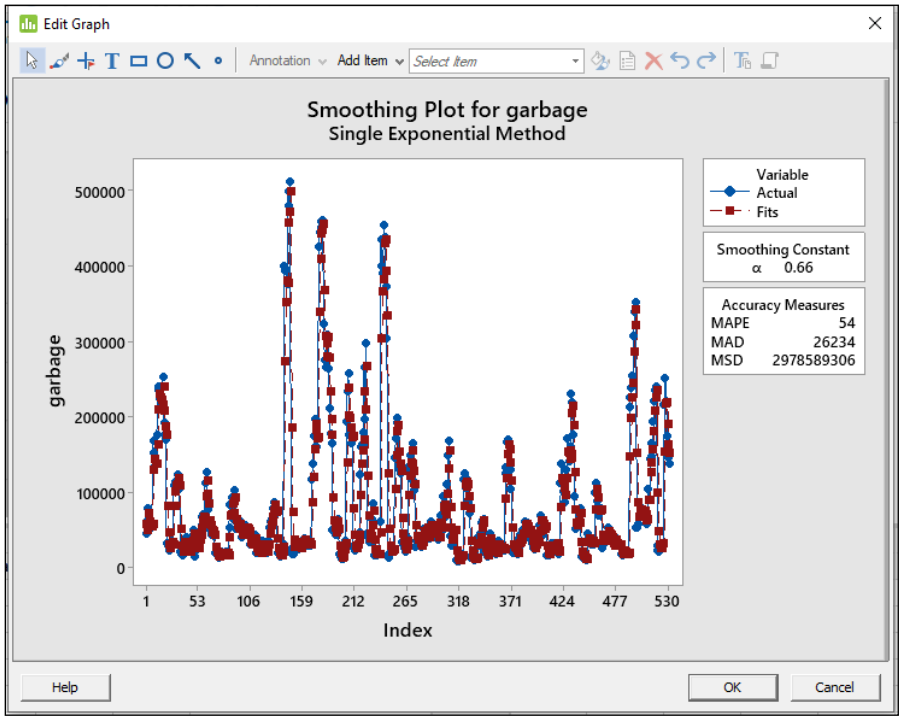
3.1.5 การประเมินประสิทธิภาพ (Evaluation)

เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไปใช้งานผู้วิเคราะห์ข้อมูลวัดประสิทธิภาพของผลลัพธ์ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใด โดยทำการทดลองโมเดลเทคนิค Time Series Models ด้วยการ

วิเคราะห์อนุกรมเวลาของข้อมูล (Data mining time series analysis) ผ่าน 2 โปรแกรม มีค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ยดังนี้

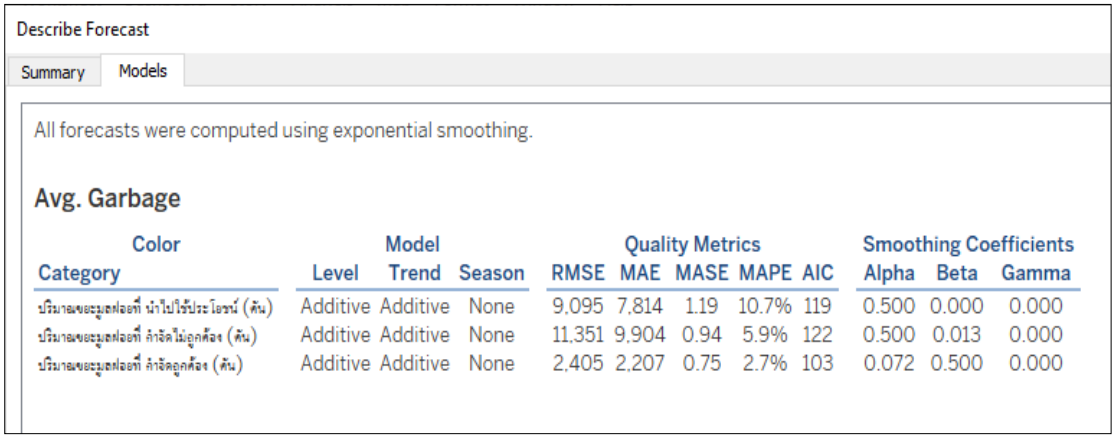
1) การวัดประสิทธิภาพของผลลัพธ์ด้วยโปรแกรม Minitab

ผลการตรวจสอบค่าเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์เฉลี่ย (Mean Absolute Percentage Error : MAPE) จากการทดลองโมเดลเทคนิค Time Series Models ด้วยการวิเคราะห์อนุกรมเวลาของข้อมูล (Data mining time series analysis) ผ่านโปรแกรม Minitab ได้ผลลัพธ์ถึง 54%



ภาพที่ 3.52 ผลลัพธ์การวัดประสิทธิภาพของผลลัพธ์โปรแกรม Minitab

2) การวัดประสิทธิภาพของผลลัพธ์ด้วยโปรแกรม Tableau Public



Describe Forecast

Summary | Models

All forecasts were computed using exponential smoothing.

Avg. Garbage

Color	Model			Quality Metrics					Smoothing Coefficients		
Category	Level	Trend	Season	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
ปริมาณขยะมูลฝอยที่ นำไปใช้ประโยชน์ (ตัน)	Additive	Additive	None	9,095	7,814	1.19	10.7%	119	0.500	0.000	0.000
ปริมาณขยะมูลฝอยที่ กำจัดไม่ถูกต้อง (ตัน)	Additive	Additive	None	11,351	9,904	0.94	5.9%	122	0.500	0.013	0.000
ปริมาณขยะมูลฝอยที่ กำจัดถูกต้อง (ตัน)	Additive	Additive	None	2,405	2,207	0.75	2.7%	103	0.072	0.500	0.000

ภาพที่ 3.53 ผลลัพธ์การวัดประสิทธิภาพของผลลัพธ์ด้วยโปรแกรม Tableau Public

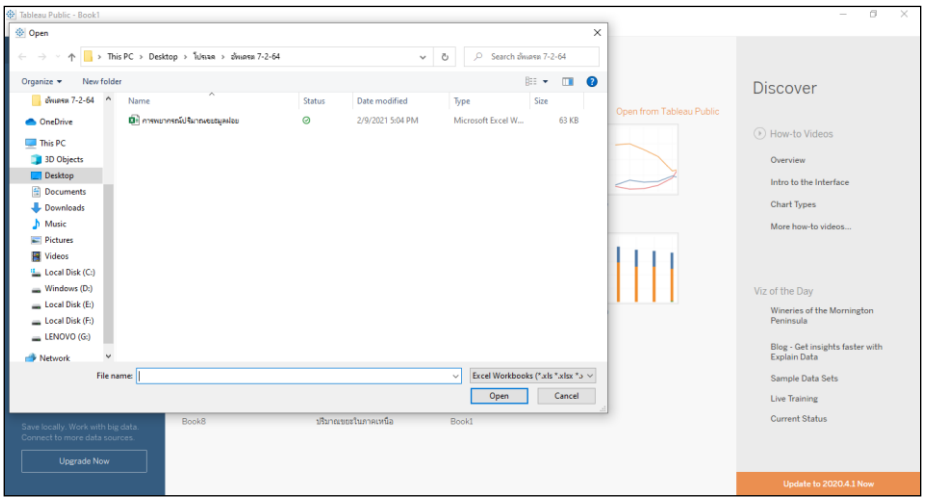
ผลการตรวจสอบค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error : MAPE) จากการทดลองโมเดลเทคนิค Time Series Models ด้วยการวิเคราะห์อนุกรมเวลาของข้อมูล (Data mining time series analysis) ผ่านโปรแกรม Tableau Public ได้ผลลัพธ์ต่ำสุดที่แบ่งเป็นประเภท อยู่ 3 ประเภท 1.) ปริมาณขยะมูลฝอยที่นำไปใช้ประโยชน์ 10.7% 2.) ปริมาณขยะมูลฝอยที่กำจัดไม่ถูกต้อง 5.9% 3.) ปริมาณขยะมูลฝอยที่กำจัดถูกต้อง 2.7%

จากการทดสอบประสิทธิภาพโมเดล และการประเมินค่าความคลาดเคลื่อน MAPE ที่ค่าคงที่เกิดยิ่งน้อยยิ่งเป็นผลดีแสดงว่าค่าพยากรณ์ออกมานั้นมีความคลาดเคลื่อนจากค่าจริงเพียงใด ผู้วิเคราะห์จึงเลือกใช้โมเดลการวิเคราะห์อนุกรมเวลาของข้อมูล (Data mining time series analysis) ผ่านโปรแกรม Tableau Public ให้ผลลัพธ์ที่น่าพึงพอใจอยู่ในระดับใช้ได้ดี สามารถนำโมเดลไปใช้งานได้ ผู้วิเคราะห์จึงเลือกใช้โมเดลอนุกรมเวลาของ Tableau Public ในการพยากรณ์ปริมาณขยะมูลฝอยในประเทศไทย เพื่อประโยชน์ในการตัดสินให้สอดคล้องกับความต้องการต่อไป

3.1.6 การนำผลลัพธ์ไปใช้งาน (Deployment)

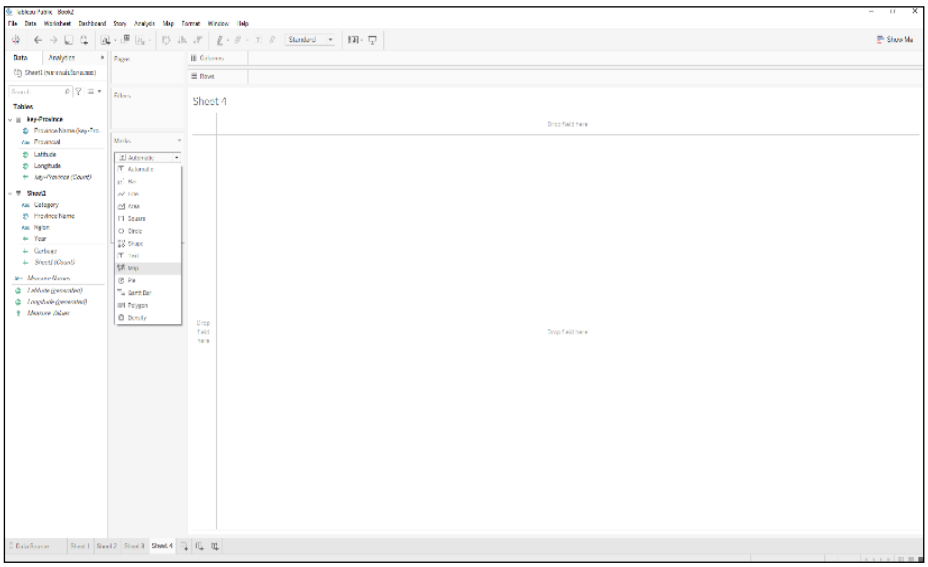
ผู้วิเคราะห์ข้อมูลนำผลองค์ความรู้ที่ได้เหล่านี้ไปนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public โดยดำเนินการตามขั้นตอนดังนี้

1) ทำการ Import ข้อมูลปริมาณขยะมูลฝอยในประเทศไทยเข้าโปรแกรม Tableau Public



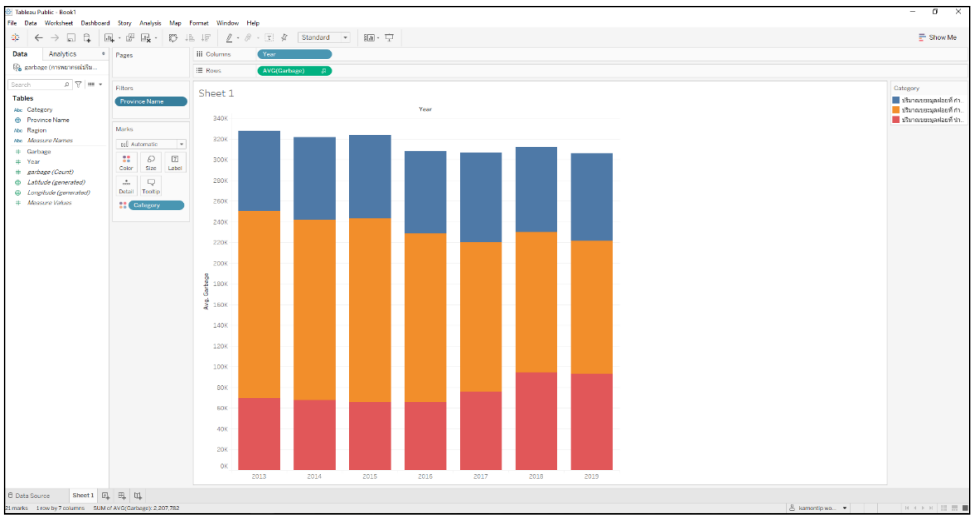
ภาพที่ 3.54 Import ข้อมูลปริมาณขยะมูลฝอยในประเทศไทยเพื่อหาผลลัพธ์

2) เข้าหน้า Dash Board การทำงานข้อมูลที่ Import จะอยู่แถบด้านซ้าย



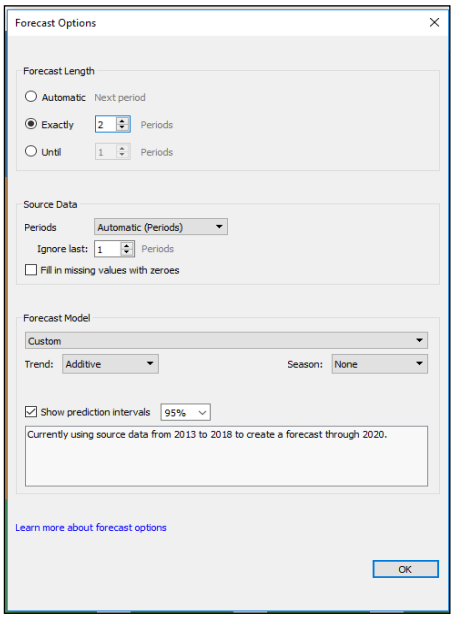
ภาพที่ 3.55 แสดงขั้นตอนการปรับ Mark ให้เป็น Bar

3) กำหนด Columns เป็น year กำหนด Rows เป็น AVG(Garbage), กำหนด category_name ใส่ใน Color และกำหนด Province ใส่ใน filters

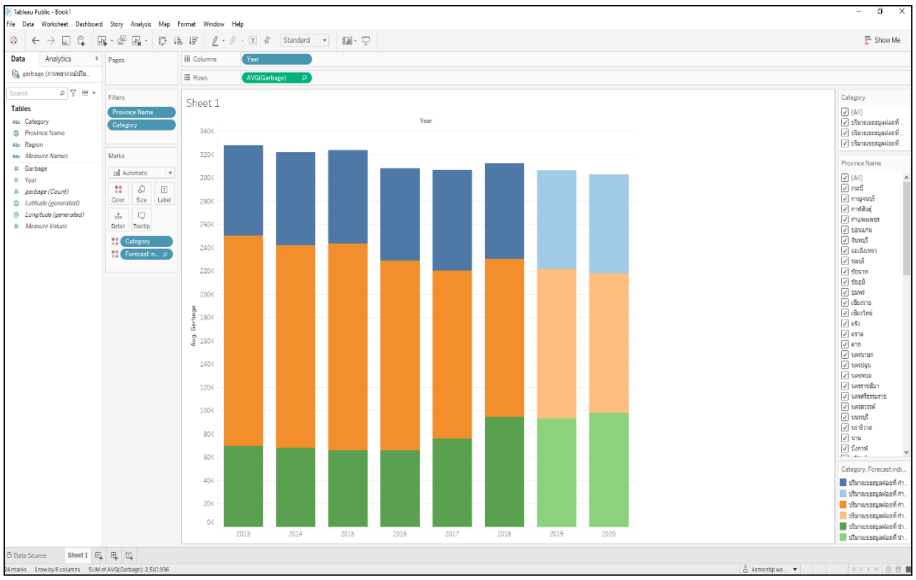


ภาพที่ 3.56 แสดงขั้นตอนการแสดง category บน Map

4) การกำหนดค่าการพยากรณ์โดยรูปแบบกราฟแท่ง

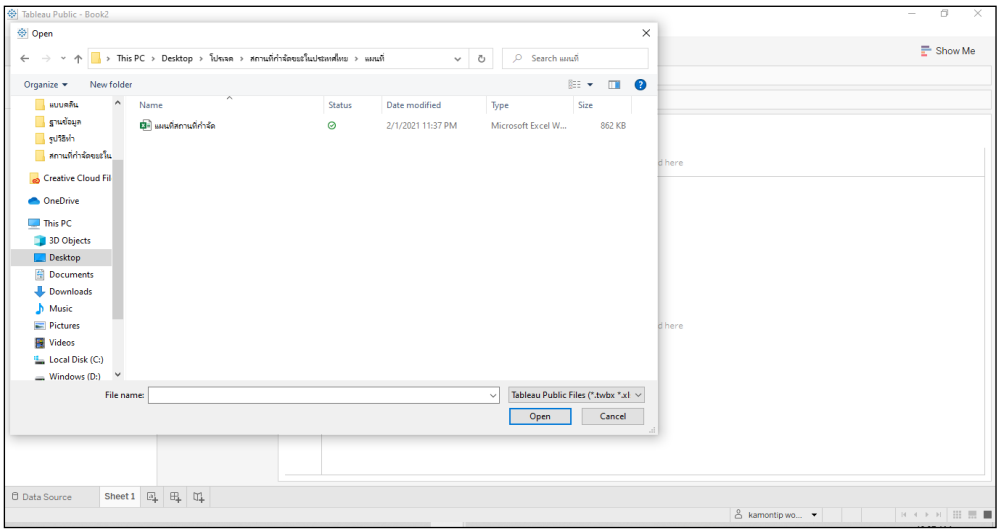


ภาพที่ 3.57 แสดงขั้นตอนการกำหนดค่าการพยากรณ์ในรูปแบบกราฟแท่ง



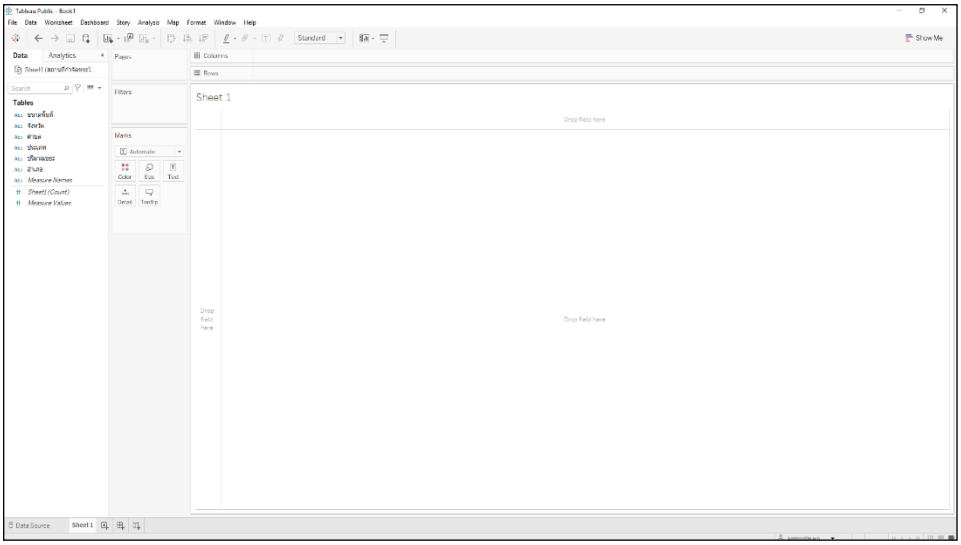
ภาพที่ 3.58 แสดงผลลัพธ์การพยากรณ์เป็นสีและรายละเอียดต่างๆ

5) หน้า Dash Board ในประเด็นที่ 2 สถานที่กำจัดขยะ การทำงานข้อมูลที่ Import จะอยู่แถบด้านซ้าย



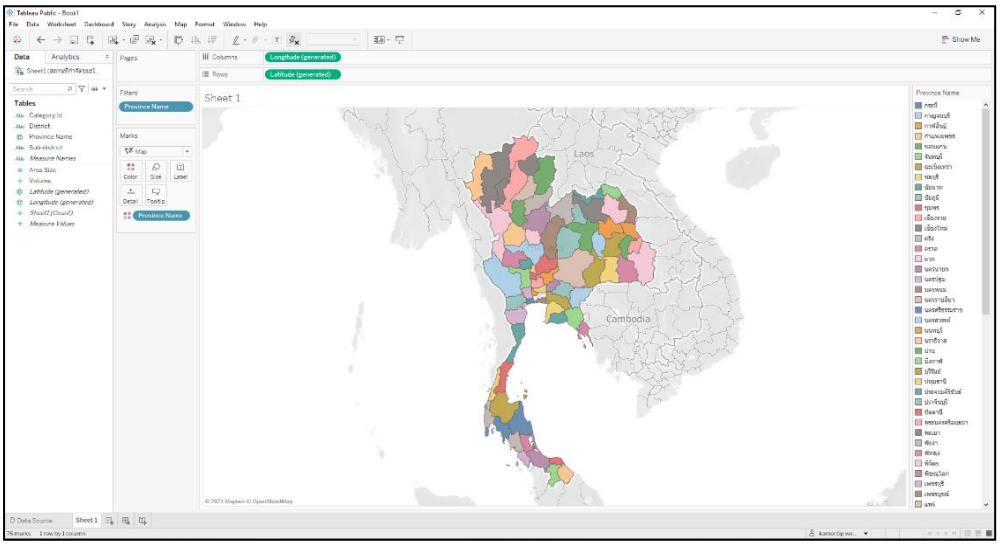
ภาพที่ 3.59 Import ข้อมูลสถานที่กำจัดขยะมูลฝอยในประเทศไทยเพื่อหาผลลัพธ์

6) เข้าหน้า Dash Board การทำงานข้อมูลที่ Import จะอยู่แถบด้านซ้าย

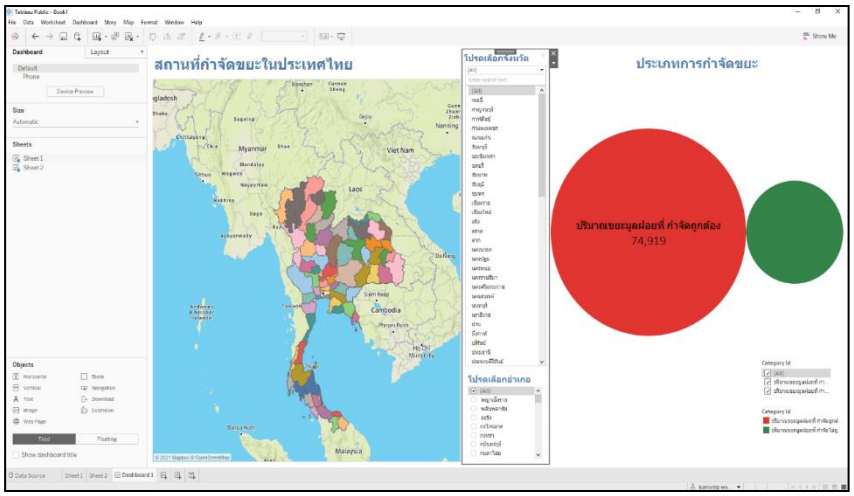


ภาพที่ 3.60 แสดงขั้นตอนการปรับ Mark ให้เป็น Map

7) กำหนด Columns เป็น Longitude (generated) กำหนด Rows เป็น Latitude (generated) และกำหนด Province Name ใส่ใน Color, กำหนด District ใส่ใน Detail และกำหนด SUM(Area Size)ใส่ใน Detail และกำหนด Province Name และกำหนด District ใส่ใน Filters



ภาพที่ 3.61 แสดงขั้นตอนการแสดงสถานที่กำจัดขยะในประเทศไทย บน Map



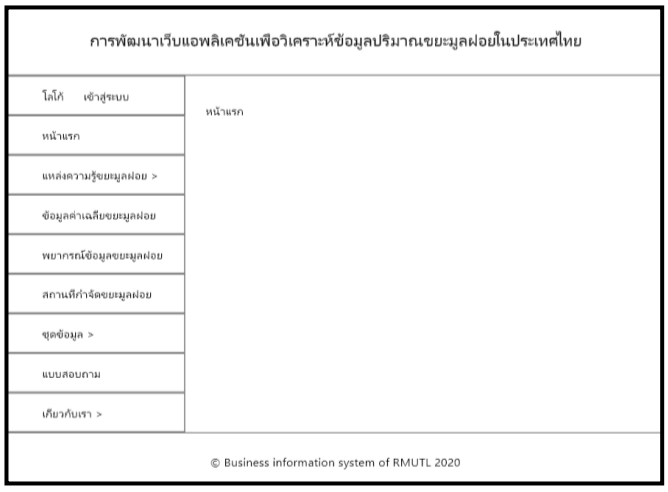
ภาพที่ 3.62 แสดงผลลัพธ์สถานที่กำจัดขยะมูลฝอยในประเทศไทย

3.6 การออกแบบและพัฒนาเว็บไซต์

การวางแผนการจัดลำดับ เนื้อหาสาระของเว็บไซต์ ออกเป็นหมวดหมู่ เพื่อจัดทำเป็นโครงสร้างในการจัดวางหน้าเว็บเพจทั้งหมด ทำให้เห็นโครงสร้างทั้งหมดของเว็บไซต์และการออกแบบโครงสร้างหรือจัดระเบียบของข้อมูลที่ชัดเจน แยกย่อยเนื้อหาออกเป็นหาออกเป็นส่วนต่าง ๆ ที่สัมพันธ์และให้มีอยู่ในมาตรฐานเดียวกัน จะช่วยให้นำมาใช้งานและง่าย ต่อการเข้าอ่านเนื้อหาของผู้ใช้เว็บไซต์

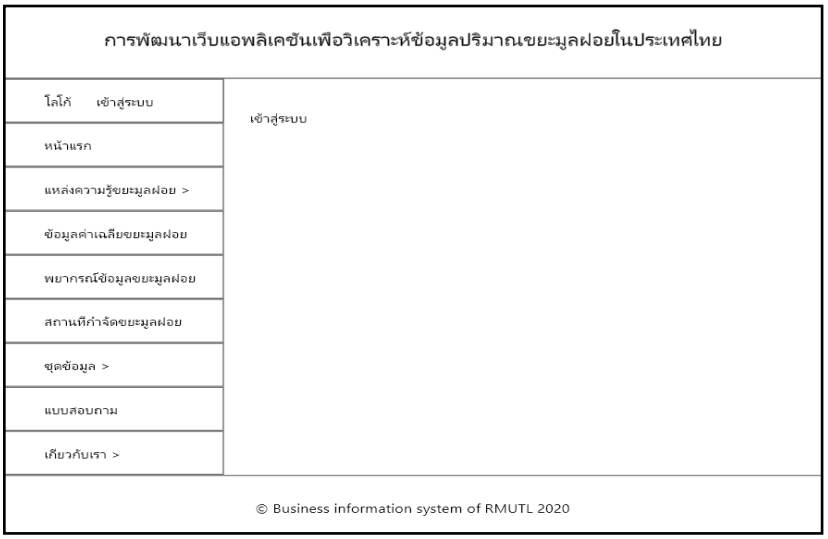
3.6.1 การออกแบบ Wireframe หน้าจอเว็บไซต์

1) หน้าแรกของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บไซต์



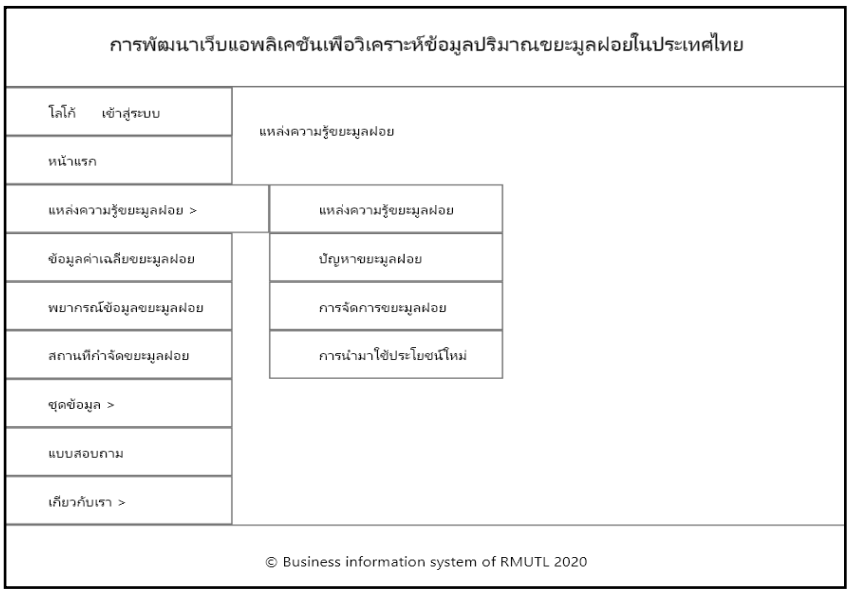
ภาพที่ 3.63 หน้าแรกของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บไซต์

2) หน้าเข้าสู่ระบบของผู้ดูแลระบบ



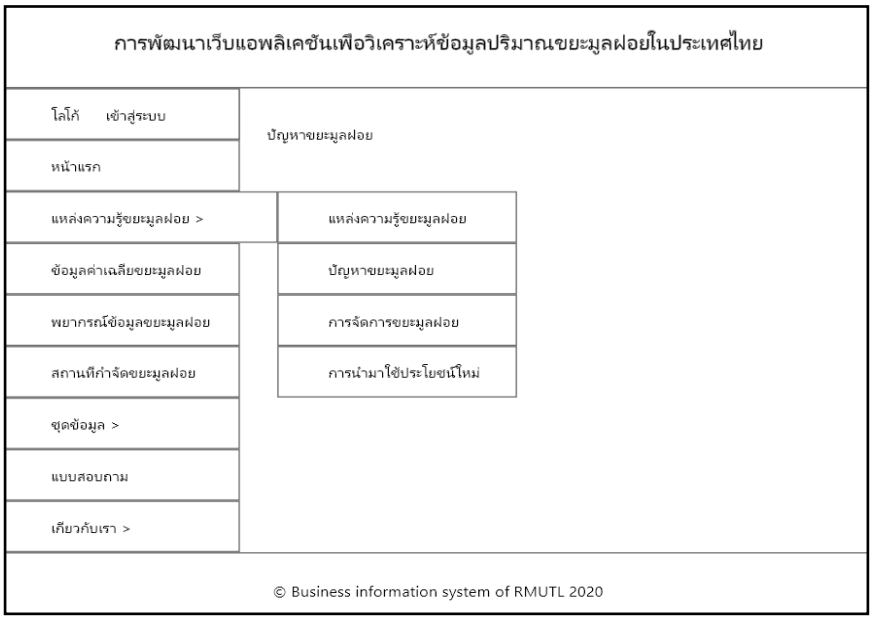
ภาพที่ 3.64 หน้าเข้าสู่ระบบของผู้ดูแลระบบ

3) หน้าแหล่งความรู้ขยะมูลฝอย



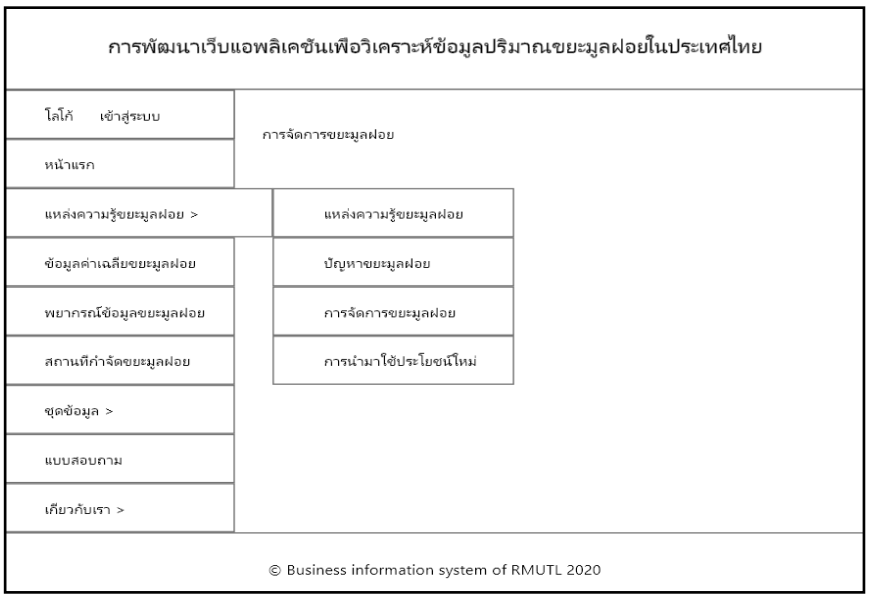
ภาพที่ 3.65 หน้าแหล่งความรู้ขยะมูลฝอย

4) หน้าความรู้เกี่ยวกับปัญหาขยะมูลฝอย



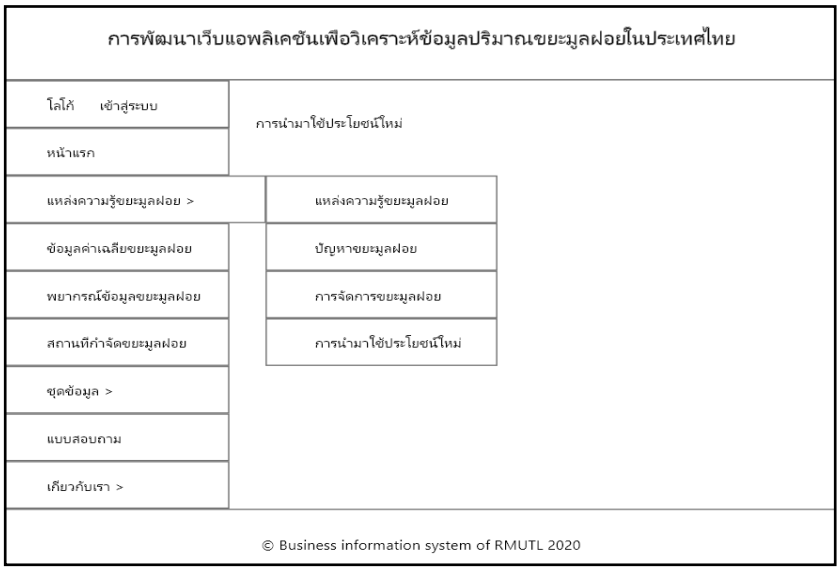
ภาพที่ 3.66 หน้าความรู้เกี่ยวกับปัญหาขยะมูลฝอย

5.) หน้าการจัดการขยะมูลฝอย



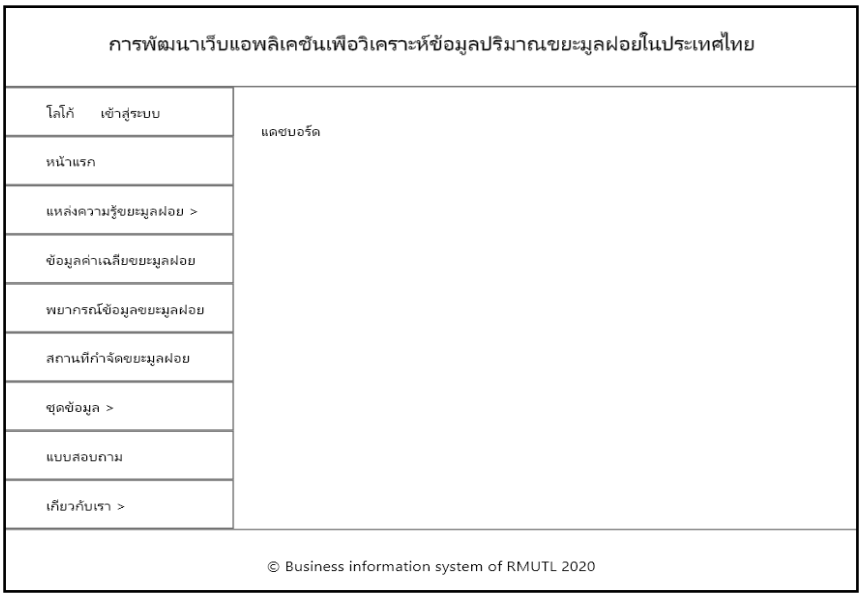
ภาพที่ 3.67 หน้าการจัดการขยะมูลฝอย

6) หน้าการนำขยะมาใช้ประโยชน์



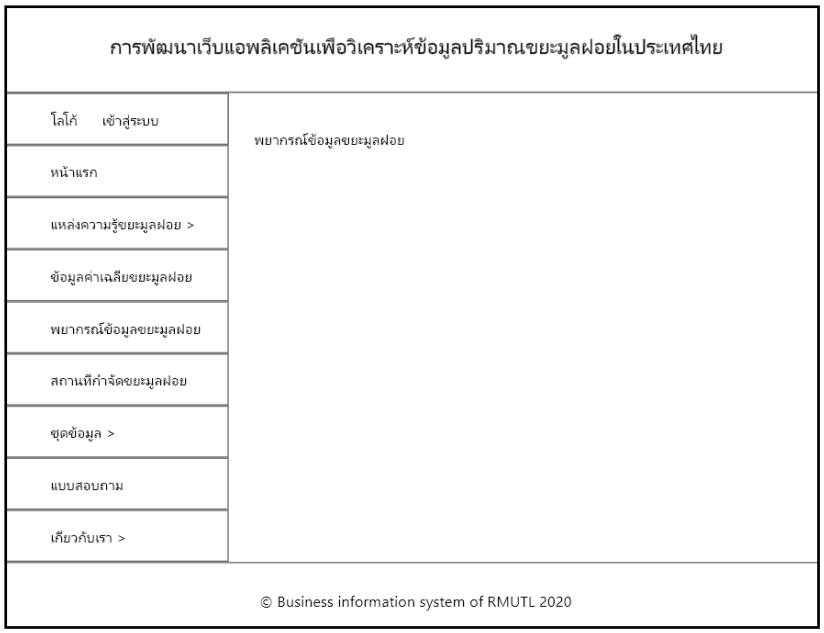
ภาพที่ 3.68 หน้าการนำขยะมาใช้ประโยชน์

7) หน้าการแสดงแดชบอร์ด



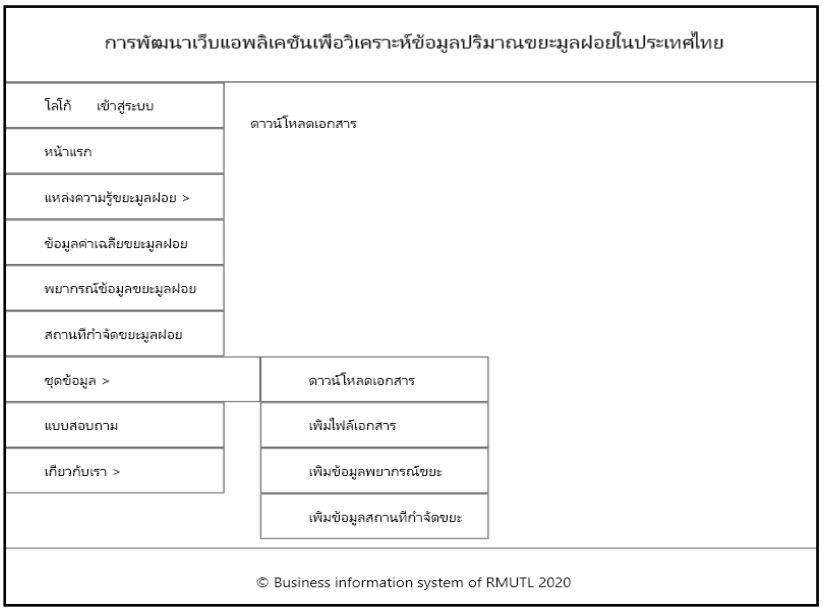
ภาพที่ 3.69 หน้าการแสดงแดชบอร์ด

8) หน้าการแสดงพยากรณ์ปริมาณขยะมูลฝอย



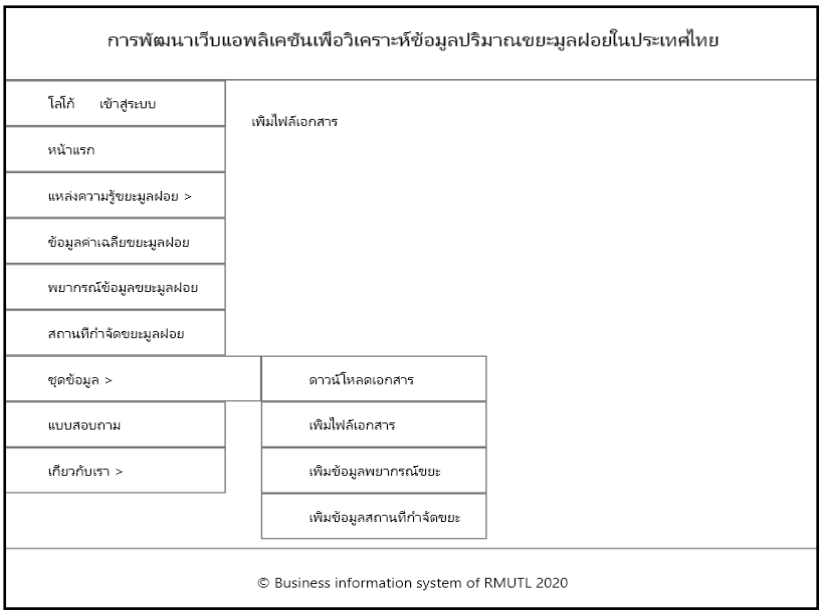
ภาพที่ 3.70 หน้าการแสดงพยากรณ์ปริมาณขยะมูลฝอย

9) หน้าการแสดงดาวน์โหลดเอกสารต่างๆ



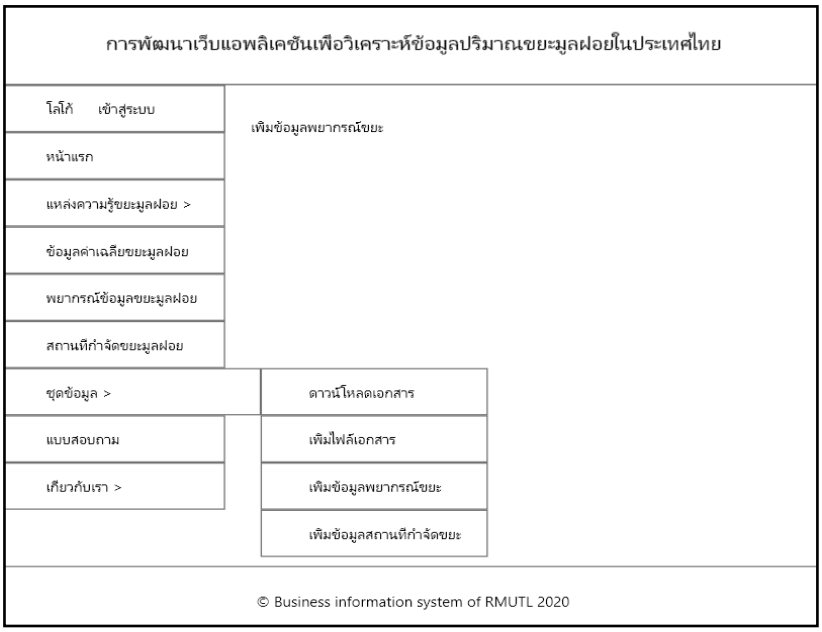
ภาพที่ 3.71 หน้าดาวน์โหลดเอกสารต่างๆ

10) หน้าเพิ่มไฟล์เอกสาร



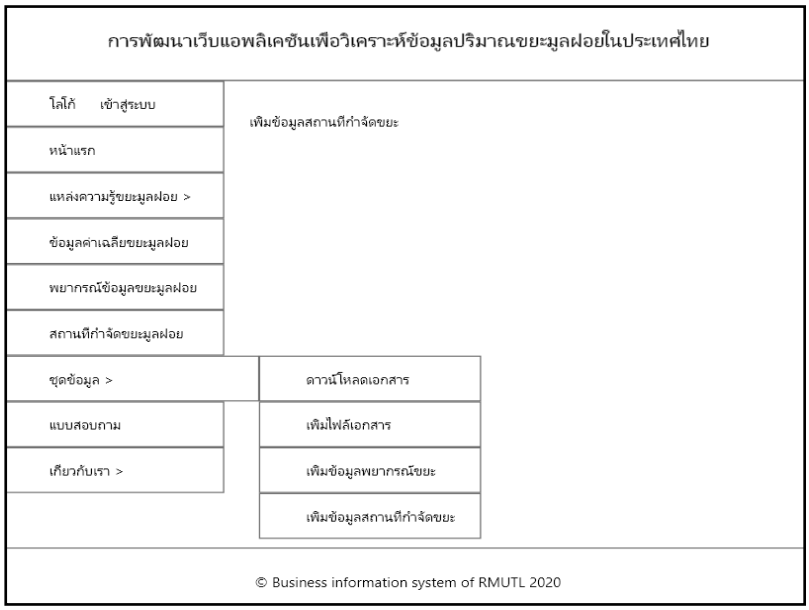
ภาพที่ 3.72 หน้าเพิ่มไฟล์เอกสาร

11) หน้าการเพิ่มข้อมูลการพยากรณ์ขยะมูลฝอย



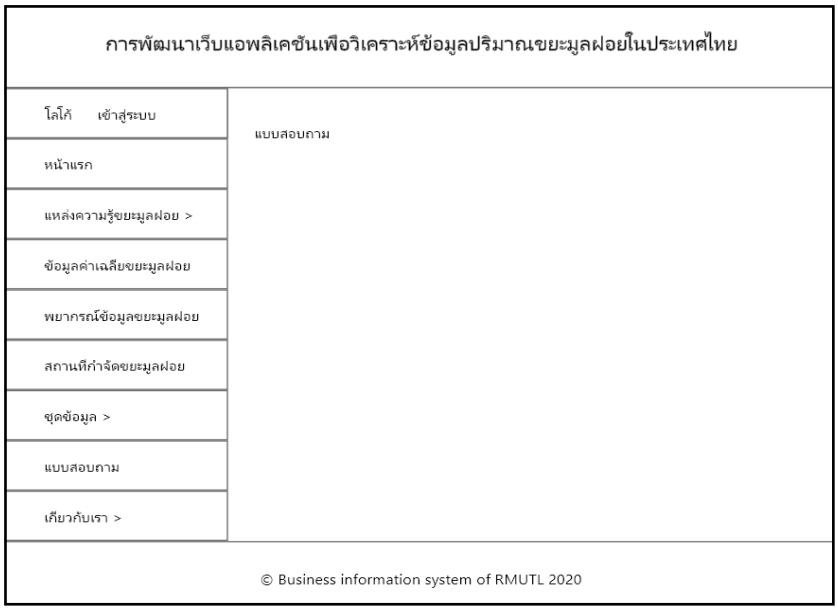
ภาพที่ 3.73 หน้าเพิ่มข้อมูลการพยากรณ์ขยะมูลฝอย

12) หน้าการเพิ่มข้อมูลสถานที่กำจัดขยะมูลฝอย



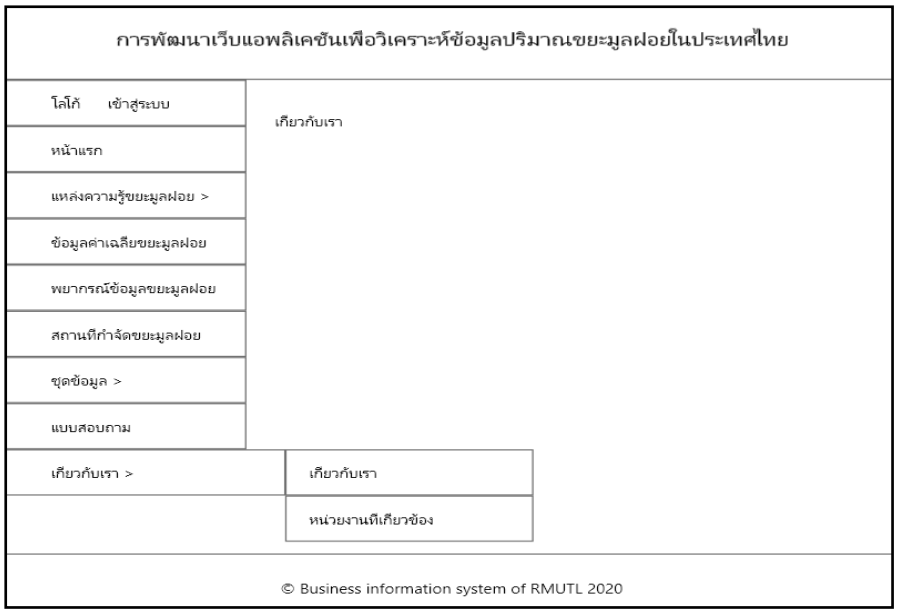
ภาพที่ 3.74 หน้าการเพิ่มสถานที่กำจัดขยะมูลฝอย

13) หน้าการแสดงแบบสอบถาม



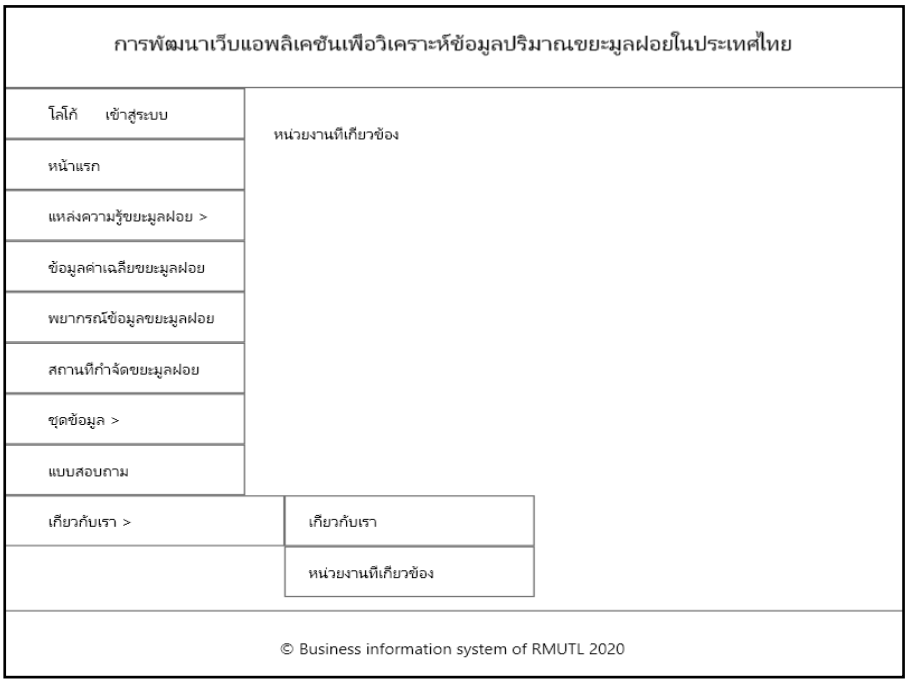
ภาพที่ 3.75 หน้าการแสดงแบบสอบถาม

14) หน้าการแสดงเกี่ยวกับเรา



ภาพที่ 3.76 หน้าการแสดงเกี่ยวกับเรา

15) หน้าการแสดงหน่วยงานที่เกี่ยวข้อง



ภาพที่ 3.77 หน้าการแสดงหน่วยงานที่เกี่ยวข้อง

3.7 บทสรุป

จากขั้นตอนการดำเนินงานผู้วิเคราะห์ได้แสดงกระบวนการวิเคราะห์การพยากรณ์ข้อมูลปริมาณขยะมูลฝอยในประเทศไทย ด้วยกระบวนการ Data Flow Diagram กระบวนการทำนอร์มัลไลเซชัน (Normalization) และวิเคราะห์ข้อมูล CRISP–DM มาใช้ในการรวบรวมข้อมูลวิเคราะห์ข้อมูลที่เกี่ยวข้องกับงาน เพื่อให้ได้สารสนเทศของงานอย่างเพียงพอในการนำไปใช้ประโยชน์ โดยผู้วิเคราะห์ได้วิเคราะห์ข้อมูลเป็นหลักสำคัญที่จะช่วยให้เข้าใจในงานแต่ละส่วน จากการวิเคราะห์ข้อมูลด้วยกระบวนการวิเคราะห์ข้อมูล CRISP–DM รวมถึงการสร้างโมเดล Time series analysis โดยเลือกใช้โปรแกรม Tableau Public ในการพยากรณ์ค่าปริมาณการส่งออกข้าวหอมมะลิไทยล่วงหน้าของปีถัดไปในช่วงเวลาเดียวกัน และนำผลการพยากรณ์ที่ได้ไปประเมินประสิทธิภาพของโมเดล ด้วยวิธีการคำนวณหาค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) ที่ต่ำที่สุด ผู้วิเคราะห์ได้นำข้อมูลสารสนเทศมาทำการแสดงผลและเผยแพร่ข้อมูลสารสนเทศผ่านเว็บแอปพลิเคชันจะพัฒนาโดยใช้ภาษา HTML, PHP, JavaScript และชุดคำสั่ง CSS3 เพื่อนำเข้าวิเคราะห์ข้อมูลด้วยโมเดลที่เลือกใช้และผ่านการทดสอบประสิทธิภาพของการวิเคราะห์ข้อมูลแล้ว และนำเสนอข้อมูลภาพด้วยเฟรมเวิร์คการแสดงข้อมูลภาพชื่อ highcharts.js เพื่อสร้างกราฟและ dashboard สำหรับแสดงผลข้อมูลภาพจากข้อมูลที่ผ่านการวิเคราะห์แล้วภายในเว็บแอปพลิเคชันที่พัฒนาขึ้น