

ใบความรู้ที่ 3 เรื่อง การจัดเตรียมข้อมูล
หน่วยการเรียนรู้ที่ 2 เรื่อง การจัดการข้อมูล
แผนการจัดการเรียนรู้ที่ 10 เรื่อง การจัดเตรียมข้อมูล
รายวิชา วิทยาการคำนวณ รหัสวิชา ว23104 ภาคเรียนที่ 1 ชั้นมัธยมศึกษาปีที่ 3

หลังจากการเก็บรวบรวมข้อมูลเสร็จเรียบร้อยแล้ว ก่อนจะนำข้อมูลไปประมวลผลจำเป็นต้องมีการจัดเตรียมข้อมูลหรือทำความสะอาดข้อมูลเพื่อตรวจสอบข้อมูลก่อนว่ามีความถูกต้อง ครบถ้วน พร้อมทั้งจะนำไปประมวลผลด้วยมือหรือคอมพิวเตอร์หรือไม่ ถ้ามีข้อมูลผิดปกติก็จำเป็นต้องปรับปรุงแก้ไขให้ข้อมูลถูกต้องและสมบูรณ์ก่อน ลักษณะของความผิดปกติของข้อมูล เช่น ข้อมูลไม่สมบูรณ์ ข้อมูลไม่เป็นรูปแบบเดียวกัน ข้อมูลไม่ครบถ้วน ข้อมูลไม่ทันสมัย

ตัวอย่าง 3.1 การจัดเตรียมข้อมูล

หลังจากเก็บรวบรวมข้อมูลมาแล้ว ได้ข้อมูลดังนี้

ที่	ชื่อกลุ่ม	ระดับชั้น	ผักที่ปลูก	รายได้จากการจำหน่ายผักทั้งภาคเรียน
1	ผักวัยใส	ม.2	ผักกวางตุ้ง	200
2	คะน้ากรอบ	ม.2	คะน้า	250
3	คูหู่สู่สารพิษ	ม.1	ผักคะน้า	280
4	ผักบุ้งไฟแดง	ม.3	ผักบุ้ง	170
5	วัยรุ่นกินผัก	ม.2	ผักกวางตุ้ง	2.10
6	คะน้าของเรา	ม.3	ผักคะน้า	210
7	ผักปลอดสารจ๋า	ม.3	ผักกาด	180
8	ผักบุ้งทิพย์	ม.1	บุ้ง	200
9	ผักบุ้งที่รัก	ม.1	ผักบุ้ง	310
10	ถั่วไหม	ม.3	ถั่วฝักยาว	270
11	ของเผ็ด	ม.2	พริก	180
12	หอมไหมหอม	ม.1	ผักสด	210

ที่	ชื่อกลุ่ม	ระดับชั้น	ผักที่ปลูก	รายได้จากการจำหน่ายผักทั้งภาคเรียน
13	กวาดตุงปลอดสาร	ม.2	ผักกวาดตุง	250
14	ใบแมงลัก	ม.3	แมงลัก	230
15	คะน้าเพื่อนซี้	ม.2	ผักคะน้า	310
16	ผักสด	ม.1	หอม	260
17	ปลูกพริกกันเถอะ	ม.3	พริก	190
18	ใครไม่ลัก แมงลัก	ม.2	แมงลัก	3000
19	หอมจ๋า	ม.3	หอม	250
20	เรารักผักบุ้ง	ม.1	ผักบุ้ง	280
21	ของเผ็ด	ม.2	พริก	180

จากข้อมูลข้างต้น จะพบว่ามีข้อมูลบางชุดผิดปกติที่อาจส่งผลต่อการนำข้อมูลไปประมวลผล ดังนี้

ตำแหน่ง	ความผิดปกติ	ประเภทของความผิดปกติ			
		ความสมบูรณ์	รูปแบบเดียวกัน	ความครบถ้วน	ความทันสมัย
ลำดับที่ 2	คะน้า		✓		
ลำดับที่ 3	คู่มือสูตรพืช	✓			
ลำดับที่ 5	2.10	✓			
ลำดับที่ 8	บุ้ง	✓			
ลำดับที่ 12	ผักสด	✓			
ลำดับที่ 18	3000	✓			
ลำดับที่ 11 กับ 21	ข้อมูลซ้ำกัน	✓			

เมื่อพบความผิดพลาด เราต้องทำการแก้ไขข้อมูลให้ถูกต้องและเป็นรูปแบบเดียวกันเสียก่อน เช่น

- ลำดับที่ 2 ค่น้ำ เพื่อให้ข้อมูลลักษณะเดียวกันมีรูปแบบที่เหมือนกัน สามารถแก้ไขเป็น ผักค่น้ำ
- ลำดับที่ 3 ชื่อกลุ่มคู้หูสู่สารพิษ ควรแก้ไขเป็น คู้หูสู่สารพิษ
- ลำดับที่ 5 รายได้จากการจำหน่ายผักทั้งภาคเรียน ข้อมูลที่พบคือ 2.10 บาท ซึ่งไม่จะเป็นไปได้ โดยเมื่อเปรียบเทียบกับข้อมูลชุดอื่นแล้วรายได้จากการขายผักจะอยู่ในหลักร้อย ข้อมูลที่ถูกต้องควรจะเป็น 210 บาท
- ลำดับที่ 12 ระบุผักที่ปลูกมา คือ ผักสด ซึ่งสามารถคาดเดาได้ยากว่าเป็นผักชนิดใด ดังนั้นผู้เก็บข้อมูลจะควรแก้ไขข้อมูลเอง แต่ควรกลับไปสอบถามผู้ให้ข้อมูล เพื่อให้ได้ข้อมูลที่ถูกต้อง แต่หากไม่สามารถกลับไปถามได้ ให้ลบข้อมูลชุดนี้ทิ้งไป
- ลำดับที่ 18 รายได้จากการจำหน่ายผักทั้งภาคเรียน ข้อมูลที่พบคือ 3000 บาท ซึ่งเมื่อเปรียบเทียบกับข้อมูลของกลุ่มอื่น ๆ แล้วไม่น่าจะเป็นข้อมูลที่ถูกต้อง ควรจะเป็น 300 บาท
- ลำดับที่ 11 กับ 21 เมื่อตรวจสอบแล้วข้อมูลเหมือนกัน อาจเกิดจากการกรอกข้อมูลซ้ำ ดังนั้นควรลบข้อมูลออกไป 1 ชุด
- นอกจากนี้ข้อมูลผักที่ปลูกยังไม่มีความเป็นรูปแบบเดียวกัน จะเห็นได้จากการปลูกผักชนิดเดียวกันแต่เขียนต่างกัน เช่น บุ่ง กับ ผักบุ้ง ค่น้ำ กับผักค่น้ำ เราต้องจัดการให้เป็นรูปแบบเดียวกันเสียก่อน

จากข้อมูลที่ยังมีข้อผิดพลาดอยู่ สามารถทำความสะอาดข้อมูลเพื่อเตรียมพร้อมสำหรับการนำไปประมวลผลได้ดังนี้

ที่	ชื่อกลุ่ม	ระดับชั้น	ผักที่ปลูก	รายได้จากการจำหน่ายผักทั้งภาคเรียน
1	ผักวัยใส	ม.2	ผักกวางตุ้ง	200
2*	ค่น้ำกรอบ	ม.2	ผักค่น้ำ	250
3*	คู้หูสู่สารพิษ	ม.1	ผักค่น้ำ	280
4	ผักบุ้งไฟแดง	ม.3	ผักบุ้ง	170
5*	วัยรุ่นกินผัก	ม.2	ผักกวางตุ้ง	210
6	ค่น้ำของเรา	ม.3	ผักค่น้ำ	210
7	ผักปลอดสารจ๋า	ม.3	ผักกาด	180
8*	ผักบุ้งทิพย์	ม.1	ผักบุ้ง	200
9	ผักบุ้งที่รัก	ม.1	ผักบุ้ง	310
10	ถั่วไหม	ม.3	ถั่วฝักยาว	270

ที่	ชื่อกลุ่ม	ระดับชั้น	ผักที่ปลูก	รายได้จากการจำหน่ายผักทั้งภาคเรียน
11	ของเผ็ด	ม.2	พริก	180
12*	หอมไหมหอม	ม.1	หอม	210
13	กวาดั่งปลอดสาร	ม.2	ผักกวาดั่ง	250
14	ใบแมงลัก	ม.3	แมงลัก	230
15	คะน้าเพื่อนซี้	ม.2	ผักคะน้า	310
16	ผักสด	ม.1	หอม	260
17	ปลูกพริกกันเถอะ	ม.3	พริก	190
18*	ใครไม่ลัก แมงลัก	ม.2	แมงลัก	300
19	หอมจ้ำ	ม.3	หอม	250
20	เรารักผักบุ้ง	ม.1	ผักบุ้ง	280

*ข้อมูลที่ได้รับการปรับปรุงแก้ไข