



หนังสือเรียน

รายวิชาพื้นฐานวิทยาศาสตร์ และเทคโนโลยี

เทคโนโลยี (วิทยาการคำนวณ)

ชั้น

มัธยมศึกษาปีที่ ๕

ตามมาตรฐานการเรียนรู้และตัวชี้วัด

กลุ่มสาระการเรียนรู้วิทยาศาสตร์ และเทคโนโลยี (ฉบับปรับปรุง พ.ศ. ๒๕๖๐)

ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช ๒๕๕๑

จัดทำโดย

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ

พิมพ์ครั้งที่ ๔

ISBN 978-616-576-381-3

จำนวน ๑๒๐,๐๐๐ เล่ม พ.ศ. ๒๕๖๖

จัดจำหน่ายโดย

องค์การค้ำของ สกสศ. พิมพ์ที่โรงพิมพ์ สกสศ. ลาดพร้าว

๒๒๔๙ ถนนลาดพร้าว วังทองหลาง กรุงเทพมหานคร

❖ มีลิขสิทธิ์ตามพระราชบัญญัติ



ประกาศสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน
เรื่อง อนุญาตให้ใช้สื่อการเรียนรู้ในสถานศึกษา

ด้วยสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) ได้จัดทำหนังสือเรียนรายวิชา
พื้นฐานวิทยาศาสตร์และเทคโนโลยี เทคโนโลยี (วิทยาการคำนวณ) ชั้นมัธยมศึกษาปีที่ ๕ ตามมาตรฐาน
การเรียนรู้และตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี (ฉบับปรับปรุง พ.ศ. ๒๕๖๐)
ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช ๒๕๕๑ สำนักงานคณะกรรมการการศึกษา
ขั้นพื้นฐาน ได้พิจารณาแล้วอนุญาตให้ใช้ในสถานศึกษาได้

ประกาศ ณ วันที่ ๑๒ มกราคม พ.ศ. ๒๕๖๗

ว่าที่ร้อยตรี

(ธนุ วงษ์จินดา)

เลขาธิการคณะกรรมการการศึกษาขั้นพื้นฐาน

คำนำ

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) มีอำนาจหน้าที่ในการพัฒนาหลักสูตร วิธีการเรียนรู้ การประเมินผล การจัดทำหนังสือเรียน แบบฝึกหัด และสื่อการเรียนรู้ทุกประเภทที่ใช้ประกอบการเรียนรู้ในกลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยีของการจัดการศึกษาขั้นพื้นฐาน

หนังสือเรียนรายวิชาพื้นฐานวิทยาศาสตร์และเทคโนโลยี เทคโนโลยี (วิทยาการคำนวณ) ชั้นมัธยมศึกษาปีที่ ๕ นี้ สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) จัดทำขึ้นตามมาตรฐานการเรียนรู้และตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี (ฉบับปรับปรุง พ.ศ. ๒๕๖๐) ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช ๒๕๕๑ โดยมีเนื้อหาเกี่ยวกับวิทยาการข้อมูล การเก็บรวบรวมและสำรวจข้อมูล การวิเคราะห์ข้อมูล และการทำข้อมูลให้เป็นภาพ ซึ่งจะเป็นประโยชน์ต่อการนำไปใช้ในการดำรงชีวิตและรู้เท่าทันกับการเปลี่ยนแปลงของโลก รวมทั้งการพัฒนาความรู้และทักษะที่จำเป็นในยุคเศรษฐกิจดิจิทัล สอดคล้องกับการเรียนรู้ในศตวรรษที่ ๒๑ ได้เป็นอย่างดี

กระทรวงศึกษาธิการหวังเป็นอย่างยิ่งว่า หนังสือเรียนเล่มนี้จะเป็นประโยชน์ต่อการจัดการเรียนรู้และเป็นส่วนสำคัญในการพัฒนาคุณภาพและมาตรฐานการศึกษา กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี ขอขอบคุณสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี ตลอดจนบุคคลและหน่วยงานอื่น ๆ ที่มีส่วนเกี่ยวข้องในการจัดทำไว้ ณ โอกาสนี้

ว่าที่ร้อยตรี



(ธนู วงษ์จินดา)

เลขาธิการคณะกรรมการการศึกษาขั้นพื้นฐาน



คำชี้แจง

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) ได้จัดทำตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี (ฉบับปรับปรุง พ.ศ. ๒๕๖๐) ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช ๒๕๕๑ โดยมีจุดเน้นเพื่อต้องการพัฒนาผู้เรียนให้มีความรู้ความสามารถที่ทัดเทียมกับนานาชาติ ได้เรียนรู้วิทยาศาสตร์ที่เชื่อมโยงความรู้กับกระบวนการ ใช้กระบวนการสืบเสาะหาความรู้และแก้ปัญหาที่หลากหลาย มีการทำกิจกรรมด้วยการลงมือปฏิบัติเพื่อให้ผู้เรียนได้ใช้ทักษะกระบวนการทางวิทยาศาสตร์ และทักษะแห่งศตวรรษที่ ๒๑ ซึ่งในปีการศึกษา ๒๕๖๑ เป็นต้นไป โรงเรียนจะต้องใช้หลักสูตรกลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี (ฉบับปรับปรุง พ.ศ. ๒๕๖๐) โดยได้มีการย้ายสาระเทคโนโลยีออกจากกลุ่มสาระการเรียนรู้การงานอาชีพและเทคโนโลยี มาอยู่ที่กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี เนื่องจากความรู้ด้านเทคโนโลยี ทั้งการออกแบบและเทคโนโลยี และวิทยาการคำนวณ เป็นพื้นฐานที่สำคัญและเชื่อมโยงกับวิทยาศาสตร์ได้เป็นอย่างดี สสวท. จึงได้จัดทำหนังสือเรียนที่เป็นไปตามมาตรฐานหลักสูตรเพื่อให้โรงเรียนได้ใช้สำหรับจัดการเรียนการสอนในชั้นเรียน

หนังสือเรียนรายวิชาพื้นฐานวิทยาศาสตร์และเทคโนโลยี เทคโนโลยี (วิทยาการคำนวณ) ชั้นมัธยมศึกษาปีที่ ๕ นี้ จัดเนื้อหาตามพัฒนาการของผู้เรียน ให้ผู้เรียนได้เรียนรู้เกี่ยวกับวิทยาการข้อมูล การเก็บรวบรวมและสำรวจข้อมูล การวิเคราะห์ข้อมูล และการทำข้อมูลให้เป็นภาพ ซึ่งเป็นพื้นฐานสำคัญต่อการนำไปใช้ในชีวิตประจำวัน ผู้เรียนจะได้ทำกิจกรรมการเรียนรู้ การฝึกปฏิบัติ การตอบคำถาม การตรวจสอบความเข้าใจ และสรุปสิ่งที่ได้เรียนรู้ ตลอดจนมีกิจกรรมให้ผู้เรียนได้นำความรู้ที่เรียนในบทนั้น ๆ มาประยุกต์ใช้แก้ปัญหาอีกด้วย ในการจัดทำหนังสือเรียนเล่มนี้ ได้รับความร่วมมือเป็นอย่างดีจากผู้ทรงคุณวุฒิ นักวิชาการอิสระ คณาจารย์ทั้งหลาย รวมทั้งครูผู้สอน นักวิชาการ จากสถาบัน และสถานศึกษา ทั้งภาครัฐและเอกชน จึงขอขอบคุณไว้ ณ ที่นี้

สสวท. หวังเป็นอย่างยิ่งว่าหนังสือเรียนรายวิชาพื้นฐานวิทยาศาสตร์ เทคโนโลยี (วิทยาการคำนวณ) เล่มนี้ จะเป็นประโยชน์แก่ผู้เรียน และผู้ที่เกี่ยวข้องทุกฝ่าย ที่จะช่วยให้การจัดการศึกษาด้านวิทยาศาสตร์และเทคโนโลยี มีประสิทธิภาพและประสิทธิผล หากมีข้อเสนอแนะใดที่จะทำให้หนังสือเรียนเล่มนี้มีความสมบูรณ์ยิ่งขึ้น โปรดแจ้ง สสวท. ทราบด้วย จะขอบคุณยิ่ง



(รองศาสตราจารย์ธีระเดช เจียรสุขสกุล)

ผู้อำนวยการสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี

กระทรวงศึกษาธิการ



คำแนะนำการใช้หนังสือเรียน รายวิชาพื้นฐานวิทยาศาสตร์และเทคโนโลยี
เทคโนโลยี (วิทยาการคำนวณ)

หนังสือเล่มนี้ประกอบด้วย 4 บท ซึ่งแต่ละบทมีจุดประสงค์การเรียนรู้ เนื้อหา ตัวอย่างการนำไปใช้ กิจกรรม คำถามชวนคิด เกร็ดน่ารู้ สรุปท้ายบท กิจกรรมท้ายบท และแบบฝึกหัดท้ายบท ในการจัดกิจกรรมการเรียนรู้ให้มีประสิทธิภาพควรใช้คู่กับคู่มือครู รายวิชาพื้นฐานวิทยาศาสตร์และเทคโนโลยี เทคโนโลยี (วิทยาการคำนวณ) ชั้นมัธยมศึกษาปีที่ 5 สารการเรียนรู้เทคโนโลยี กลุ่มสาระการเรียนรู้วิทยาศาสตร์และเทคโนโลยี (ฉบับปรับปรุง พ.ศ. 2560) ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ซึ่งได้กำหนดตัวชี้วัด คือ รวบรวม วิเคราะห์ข้อมูล และใช้ความรู้ด้านวิทยาการคอมพิวเตอร์ สื่อดิจิทัล เทคโนโลยีสารสนเทศในการแก้ปัญหาหรือเพิ่มมูลค่าให้กับบริการหรือผลิตภัณฑ์ที่ใช้ในชีวิตจริงอย่างสร้างสรรค์

เนื้อหาและกิจกรรมในหนังสือเรียนนี้มีความสอดคล้องกับตัวชี้วัด โดยกำหนดเวลาในการเรียนรู้จำนวน 40 ชั่วโมง ผู้เรียนสามารถศึกษาแต่ละบทเรียนตามลำดับ ดังตารางต่อไปนี้

บทที่	เรื่อง	เวลาในการจัดกิจกรรม (ชั่วโมง)	บทเรียนที่ต้องศึกษา ก่อนหน้า
1	ข้อมูลมีคุณค่า	6	-
2	การเก็บรวบรวมและสำรวจข้อมูล	8	บทที่ 1
3	การวิเคราะห์ข้อมูล	12	บทที่ 1 และ 2
4	การทำข้อมูลให้เป็นภาพ และการสื่อสารด้วยข้อมูล	14	บทที่ 1, 2 และ 3

สัญลักษณ์ที่ควรรู้			
	จุดประสงค์ของบทเรียน สิ่งที่ต้องการให้เกิดกับผู้เรียน		ชวนคิด คำถามหรือกิจกรรมให้ลองคิดหรือปฏิบัติ
	ตัวอย่าง สิ่งที่ช่วยให้เข้าใจเนื้อหามากยิ่งขึ้น		กิจกรรม สิ่งที่ให้ปฏิบัติหลังจากมีความรู้แต่ละหัวข้อ
	เกร็ดน่ารู้ ความรู้ในเรื่องที่เกี่ยวข้อง		สรุปท้ายบท ข้อความสรุปเนื้อหาที่สำคัญของแต่ละบท
	กิจกรรมท้ายบท กิจกรรมที่ให้ปฏิบัติหลังจากเรียนรู้แต่ละบท		แบบฝึกหัดท้ายบท คำถามหรืองานเพื่อประเมินความรู้

สารบัญ	บทที่ 1 - 2	
บทที่	เนื้อหา	หน้า
1	1.1 ยุคของข้อมูลและสารสนเทศ 5 1.2 วิทยาการข้อมูล 10 1.3 กระบวนการวิทยาการข้อมูล 17 1.4 การคิดเชิงออกแบบสำหรับวิทยาการข้อมูล 20 กิจกรรมท้ายบท 27 แบบฝึกหัดท้ายบท 29	
		
ข้อมูลมีคุณค่า		
2	2.1 การเก็บรวบรวมข้อมูล 36 2.1.1 แหล่งข้อมูลทุติยภูมิ 37 2.1.2 ความเหมาะสมของแหล่งข้อมูล 41 2.2 การเตรียมข้อมูล 43 2.2.1 การทำความสะอาดข้อมูล 44 2.2.2 การแปลงข้อมูล 46 2.2.3 การเชื่อมโยงข้อมูล 53 2.3 การสำรวจข้อมูล 55 2.3.1 การสำรวจข้อมูลโดยใช้กราฟเส้น 56 2.3.2 การสำรวจข้อมูลโดยใช้ฮิสโทแกรม 56 2.3.3 การสำรวจข้อมูลโดยใช้แผนภาพการกระจาย 57 2.3.4 การสำรวจข้อมูลโดยใช้แผนภาพกล่อง 58 2.3.5 การสำรวจข้อมูลด้วยการเขียนโปรแกรม 59 2.4 ข้อมูลส่วนบุคคล 63 กิจกรรมท้ายบท 67 แบบฝึกหัดท้ายบท 67	
		
การเก็บรวบรวมและสำรวจข้อมูล		

สารบัญ	บทที่ 3 - 4	
บทที่	เนื้อหา	หน้า
3	3.1 การวิเคราะห์เชิงพรรณนา 80 3.1.1 การหาสัดส่วนหรือร้อยละ 80 3.1.2 การวัดค่ากลางของข้อมูล 81 3.1.3 การหาความสัมพันธ์ของชุดข้อมูล 84 3.2 การวิเคราะห์เชิงทำนาย 89 3.2.1 การทำนายเชิงตัวเลข 90 3.2.2 การทำนายเชิงหมวดหมู่ 98 กิจกรรมท้ายบท 108 แบบฝึกหัดท้ายบท 109	
		
การวิเคราะห์ข้อมูล		
4	4.1 การสื่อสารด้วยข้อมูล 115 4.2 การทำข้อมูลให้เป็นภาพ 118 4.2.1 แผนภูมิรูปร่างกลม 119 4.2.2 แผนภูมิแท่ง 120 4.2.3 กราฟเส้น 121 4.2.4 แผนภาพการกระจาย 122 4.2.5 การเลือกใช้แผนภาพ 122 4.3 การทำข้อมูลให้เป็นภาพอย่างเหมาะสม 123 4.4 การเล่าเรื่องราวจากข้อมูล 129 4.5 ข้อควรระวังในการนำเสนอข้อมูล 132 กิจกรรมท้ายบท 135 แบบฝึกหัดท้ายบท 136	
		
การทำข้อมูลให้เป็นภาพ และการสื่อสารด้วยข้อมูล		
ดัชนี 140		
บรรณานุกรม 142		
คณะผู้จัดทำ 144		





บทที่

1

ข้อมูลมีคุณค่า

- 1.1 ยุคของข้อมูลและสารสนเทศ
- 1.2 วิทยาการข้อมูล
- 1.3 กระบวนการวิทยาการข้อมูล
- 1.4 การคิดเชิงออกแบบสำหรับวิทยาการข้อมูล



จุดประสงค์ของบทเรียน
เมื่อเรียนจบบทนี้แล้วนักเรียนจะสามารถ

1. อธิบายกระบวนการวิทยาการข้อมูล
2. ตระหนักถึงความสำคัญของข้อมูล ประโยชน์หรือคุณค่าของข้อมูล
3. ประยุกต์การคิดเชิงออกแบบกับวิทยาการข้อมูลในการแก้ปัญหา

บทที่ 1

ข้อมูลมีคุณค่า



“ในยุคของข้อมูลและสารสนเทศ มีปริมาณข้อมูลเพิ่มมากขึ้นทุกวัน จากผู้ใช้ที่มีอยู่ทั่วโลกจำนวนมาก ทำให้ข้อมูลกลายเป็นสิ่งที่มีมูลค่ามหาศาล มีการใช้ศาสตร์ที่เรียกว่าวิทยาการข้อมูล (data science) ซึ่งมีความสำคัญและช่วยให้ผู้ใช้เข้าใจความหมายของข้อมูล และในขณะเดียวกันผู้ใช้จะได้รับความรู้จากข้อมูลที่ผ่านกระบวนการวิทยาการข้อมูลด้วย”



DATA ANALYTICS

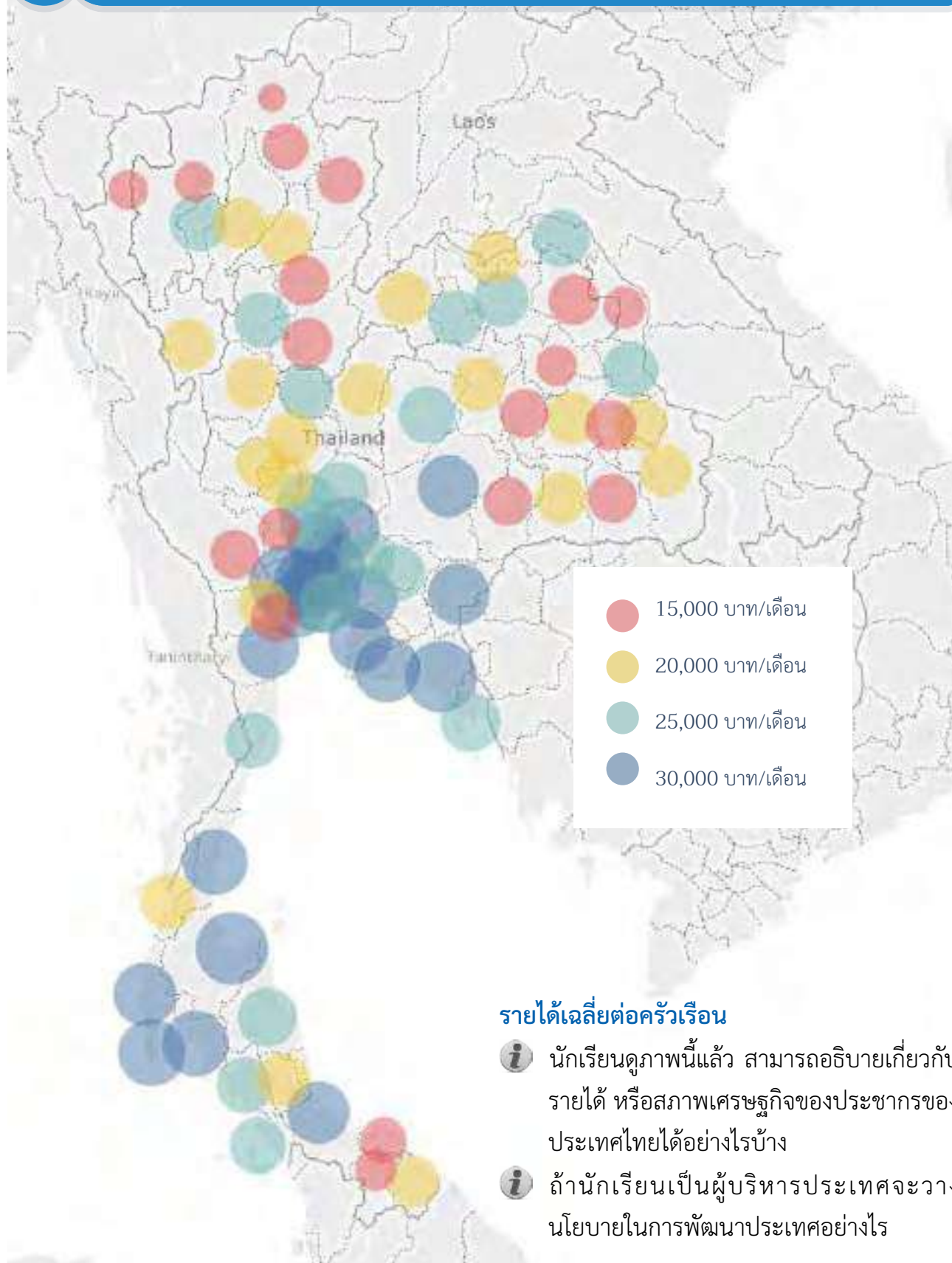


“องค์กรทั้งภาครัฐ และภาคธุรกิจ ล้วนแล้วแต่ใช้ประโยชน์จากข้อมูล เพื่อนำไปสู่การตัดสินใจในการแก้ปัญหาที่อาจไม่สามารถแก้ไขได้ในอดีต กระบวนการวิทยาการข้อมูลมีการนำข้อมูลที่เกี่ยวข้องมาวิเคราะห์ (data analytics) เพื่ออธิบาย ค้นหาคำตอบ หรือทำนายปรากฏการณ์ที่เกิดขึ้น การทำข้อมูลให้เป็นภาพ (data visualization) หรือการเล่าเรื่องราวที่เกิดจากข้อมูล (data story telling) ทำให้ผู้ใช้ได้รับความรู้และเข้าใจข้อมูลได้ง่าย ดังนั้น การเรียนรู้เกี่ยวกับการใช้ประโยชน์และการนำเสนอข้อมูลที่เกิดขึ้น จึงเป็นสิ่งสำคัญและจำเป็นสำหรับนักเรียนในยุคนี้”





ชวนคิด



1.1 ยุคของข้อมูลและสารสนเทศ (information age)

ในยุคของข้อมูลและสารสนเทศ ข้อมูลเป็นสิ่งที่มีความสำคัญอย่างยิ่ง และถูกนำมาใช้ประโยชน์ในด้านต่าง ๆ เช่น ด้านเศรษฐกิจ การศึกษา สาธารณสุข สิ่งแวดล้อม การเกษตร และการคมนาคม การจัดเก็บข้อมูลในรูปแบบเดิมทำให้การนำข้อมูลมาใช้ไม่สะดวก ไม่ทันกาล สูญหายง่าย แต่การจัดเก็บข้อมูลในรูปแบบดิจิทัล (digitization) และพัฒนาการของการสื่อสารบนอินเทอร์เน็ต ช่วยแก้ปัญหาเหล่านี้ ทำให้ผู้ใช้สามารถเข้าถึงข้อมูลและสารสนเทศได้ทุกที่ทุกเวลา ตัวอย่างเช่น การใช้แผนที่กระดาษในรูปแบบเดิม ดังรูป 1.1 จะแสดงตำแหน่งของสถานที่ต่าง ๆ แต่จะไม่แสดงข้อมูลการจราจรที่เป็นปัจจุบัน ทำให้ผู้ใช้ไม่สามารถวางแผนการเดินทางได้ แต่ระบบแผนที่นำทาง (Global Positioning System: GPS) ดังรูป 1.2 นอกจากจะแสดงตำแหน่งของสถานที่ต่าง ๆ แล้ว ยังมีข้อมูลเกี่ยวกับสภาพการจราจร ระยะเวลาที่ต้องใช้ในการเดินทางถึงจุดหมาย ช่วยทำให้ผู้ใช้วางแผนการเดินทางได้อย่างแม่นยำ เป็นการประหยัดทั้งเวลาและค่าใช้จ่าย



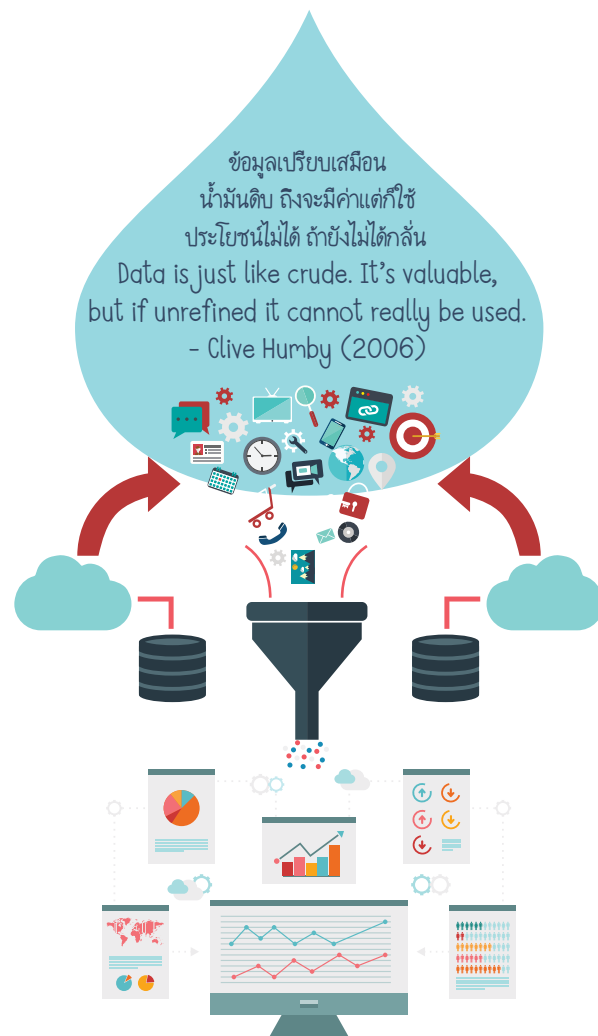
รูป 1.1 แผนที่กระดาษ



รูป 1.2 ระบบแผนที่นำทาง

ทั้งนี้ ในปัจจุบันเราไม่เพียงเป็นผู้ใช้ประโยชน์ข้อมูลดิจิทัลเท่านั้น แต่ยังเป็นหนึ่งในผู้ร่วมสร้างข้อมูลดิจิทัลด้วยเช่นกัน ตัวอย่างเช่น การอัปโหลดรูปภาพส่วนตัว การส่งอีเมลในแต่ละวัน การโพสต์ข้อความในสื่อสังคมออนไลน์ การส่งต่อข้อความ โดยการใช้อุปกรณ์อิเล็กทรอนิกส์ต่าง ๆ เช่น คอมพิวเตอร์ โน้ตบุ๊ก โทรศัพท์มือถือ หรือแม้กระทั่งเซ็นเซอร์ที่ติดกับอุปกรณ์ต่าง ๆ ซึ่งผู้ใช้คนอื่นสามารถนำข้อมูลดิจิทัลเหล่านี้ไปใช้ประโยชน์ต่อได้ อาจจัดได้ว่าข้อมูลเหล่านี้เป็นสินทรัพย์ (asset) ที่มีความสำคัญ แต่ถ้าข้อมูลที่มีอยู่ไม่ได้ถูกนำมาประมวลผลก็จะไม่เกิดคุณค่าใด ๆ ดังเช่นคำกล่าวที่ว่า “ข้อมูลนั้นมีค่าดั่งน้ำมันดิบ”

“ข้อมูลนั้นมีค่าดั่งน้ำมันดิบ”



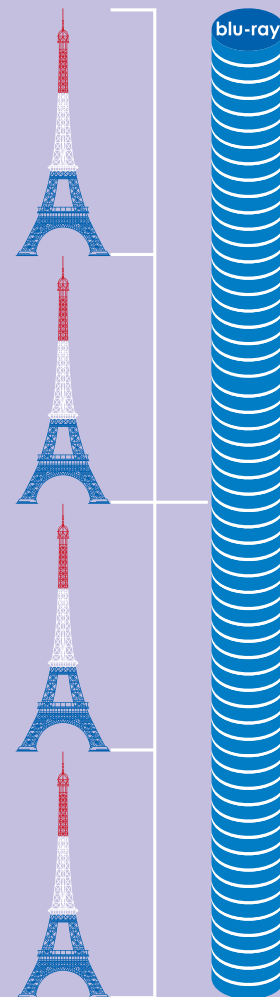
ชวนคิด

การที่เราแชร์ข้อมูลของเรา เช่น ภาพส่วนตัว อีเมล การโพสต์ข้อความ นักเรียนคิดว่าจะมีผู้นำข้อมูลของเราไปใช้ประโยชน์ต่ออย่างไรได้บ้าง



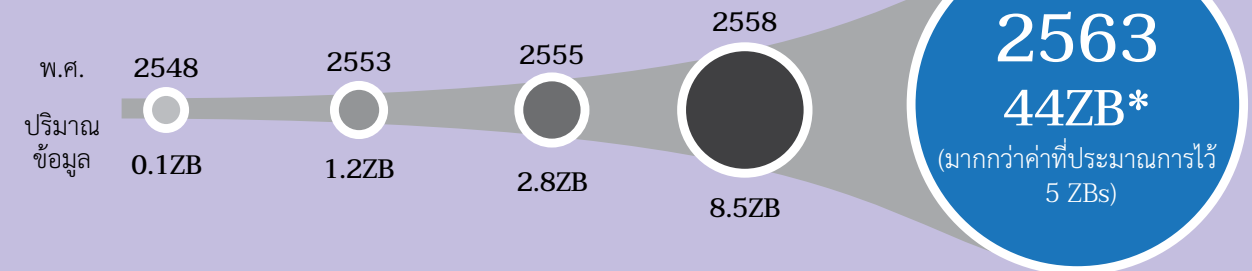
เกร็ดน่ารู้ : ข้อมูลมหาศาล

ในแต่ละวันประชากรกว่าร้อยล้านคนบนโลก ถ่ายภาพจากโทรศัพท์มือถือ ส่งอีเมล ส่งข้อความ สร้างวิดีโอ และข้อมูลที่เกี่ยวข้องกับการดำเนินธุรกิจ โดยปริมาณข้อมูลที่สร้างขึ้นในแต่ละวันนั้น คิดเป็น 2.5 เอกซะไบต์ (Exabyte-10¹⁸) ขนาดของข้อมูลนี้เทียบเท่ากับจำนวนเพลง 530 ล้านเพลง หรือไฟล์วิดีโอความละเอียดคมชัด (High Definition) ความยาว 90 ปี หรือถ้านำข้อมูลนี้บันทึกลงแผ่นบลูเรย์ (blu-ray) ต้องใช้แผ่นบลูเรย์ประมาณ 10 ล้านแผ่น ซึ่งถ้านำแผ่นทั้งหมดนี้มาเรียงซ้อนกันจะมีความสูงเท่ากับ 4 เท่าของหอไอเฟล



แผ่นบลูเรย์
ประมาณ
10 ล้านแผ่น

ด้วยปริมาณข้อมูลที่สร้างขึ้นในปัจจุบัน มีการคาดการณ์ว่าในปี พ.ศ. 2563 ปริมาณข้อมูลบนโลกจะมีขนาดประมาณ 44 เซตตะไบต์ (Zetabyte-10²¹)



บริษัทต่าง ๆ ได้มีการนำข้อมูลดิจิทัลมาใช้ประโยชน์ทำให้เกิดมูลค่ามหาศาล เช่น บริษัทที่ให้บริการจองโรงแรมที่พัก รถแท็กซี่ ขายสินค้าออนไลน์ และบริการสื่อสังคม (social media)

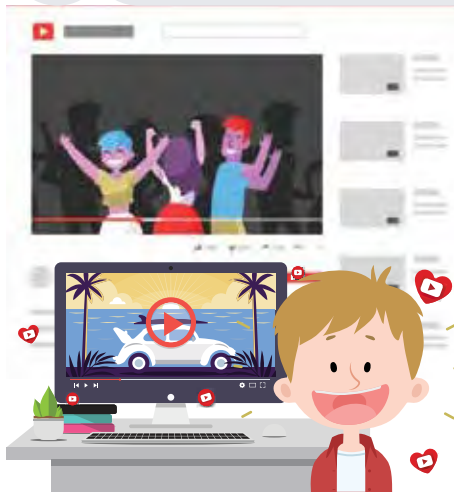
เฟซบุ๊ก (Facebook) เป็นบริษัทที่เปิดให้บริการสื่อสังคม ที่มีผู้ใช้หลายล้านคนทั่วโลกโดยไม่คิดค่าใช้จ่าย แต่สามารถสร้างรายได้จากการขายโฆษณาที่ตรงกับกลุ่มเป้าหมาย (user-targeted advertisements) ของสินค้าและบริการ โดยรวบรวมข้อมูลจากผู้ใช้งาน เช่น เพศ อายุ ที่อยู่ อาชีพ รวมถึงพฤติกรรมการใช้งานที่ผู้ใช้กระทำผ่านเฟซบุ๊ก ไม่ว่าจะเป็นการกดไลค์ (like) กดแชร์ (share) ข้อความ ภาพ หรือวิดีโอ เฟซบุ๊กจะนำข้อมูลเหล่านี้มาประมวลผลเป็นสารสนเทศที่บอกคุณลักษณะของผู้ใช้ และใช้สารสนเทศนี้ในการนำเสนอหรือโฆษณาสินค้าหรือสิ่งของที่ผู้ใช้สนใจ บริษัทที่เป็นเจ้าของสินค้าหรือผลิตภัณฑ์นั้นจะจ่ายค่าโฆษณาให้กับเฟซบุ๊ก เช่น บริษัทในกลุ่มธุรกิจรถยนต์ ธนาคาร อาหาร-เครื่องดื่ม โทรศัพท์มือถือ ซูเปอร์มาร์เก็ต และเกมออนไลน์



รูป 1.3 หน้าตัวอย่างโฆษณาบนเฟซบุ๊ก

จากที่มีการนำข้อมูลดิจิทัล ซึ่งมีอยู่มหาศาล ณ ขณะนี้ มาใช้ประโยชน์ในด้านต่าง ๆ ความรู้ด้านวิทยาการข้อมูลจึงมีบทบาทสำคัญ และอาชีพนักวิทยาศาสตร์ข้อมูลก็เป็นอาชีพที่น่าสนใจ และได้รับความนิยมเป็นอย่างมากในยุคของข้อมูลและสารสนเทศนี้

ชวนคิด



ยอรวมเวลาที่คนทั้งโลกดูยูทูป (YouTube) ใน 1 วัน = 1 พันล้านชั่วโมง ทุก ๆ 1 นาที จะมีการอัปโหลดวิดีโอ ความยาวรวม 300 ชั่วโมง (ข้อมูล ณ วันที่ 24 มกราคม 2561) ทำให้เราต้องใช้เวลาดูวิดีโอที่อัปโหลดเข้ามาไว้ในยูทูปทั้งปี จำนวนทั้งสิ้น 157 ล้านชั่วโมง

- ถ้าคนหนึ่งคนต้องการจะดูยูทูปแบบไม่หยุด (non-stop) ต้องใช้เวลาดูทั้งหมดกี่ปี
- ต้องใช้ความจุของหน่วยความจำในการจัดเก็บข้อมูลวิดีโอที่อัปโหลดทั้งปีจำนวนเท่าไร (ให้ใช้หน่วยที่ใหญ่ที่สุดเท่าที่รู้จัก)

เกร็ดน่ารู้ : นักวิทยาศาสตร์ข้อมูล (data scientist)

นักวิทยาศาสตร์ข้อมูล คือ ผู้ที่มีความสามารถในการค้นหา วิเคราะห์ ค้นพบสิ่งที่น่าสนใจ น่าสงสัย และเป็นประโยชน์ในโลกที่เต็มไปด้วยข้อมูลขนาดใหญ่จำนวนมหาศาล (Big data) นักวิทยาศาสตร์ข้อมูลมีหน้าที่ในการจัดการกับข้อมูลที่มีปริมาณมาก ยุ่งยาก และหลากหลาย โดยทำการวิเคราะห์ ค้นหา รูปแบบความสัมพันธ์ของข้อมูล และนำเสนอข้อมูลเหล่านั้นออกมาในรูปแบบที่ผู้อื่นเข้าใจง่าย ทำให้เกิดเป็นองค์ความรู้ และสามารถนำไปใช้ในการตัดสินใจได้



อาชีพนักวิทยาศาสตร์ข้อมูล เป็นอาชีพที่มีเสน่ห์และน่าสนใจที่สุดในศตวรรษที่ 21 บริษัทสตาร์ทอัพ บริษัทขนาดใหญ่ และองค์กรภาครัฐ ต่างมีความต้องการเป็นอย่างมาก แต่จำนวนนักวิทยาศาสตร์ข้อมูลที่มีอยู่ไม่เพียงพอความต้องการ อาชีพนี้สามารถทำรายได้สูงกว่าอาชีพอื่นที่มีลักษณะใกล้เคียงกัน คุณลักษณะของนักวิทยาศาสตร์ข้อมูลที่มีความสามารถ และประสบความสำเร็จในอาชีพ จะต้องเป็นผู้ที่มีความสามารถทั้งการเป็นนักสืบค้นและเจาะหาข้อมูล นักวิเคราะห์ นักสื่อสาร และเป็นผู้ให้คำปรึกษาที่เชื่อถือได้

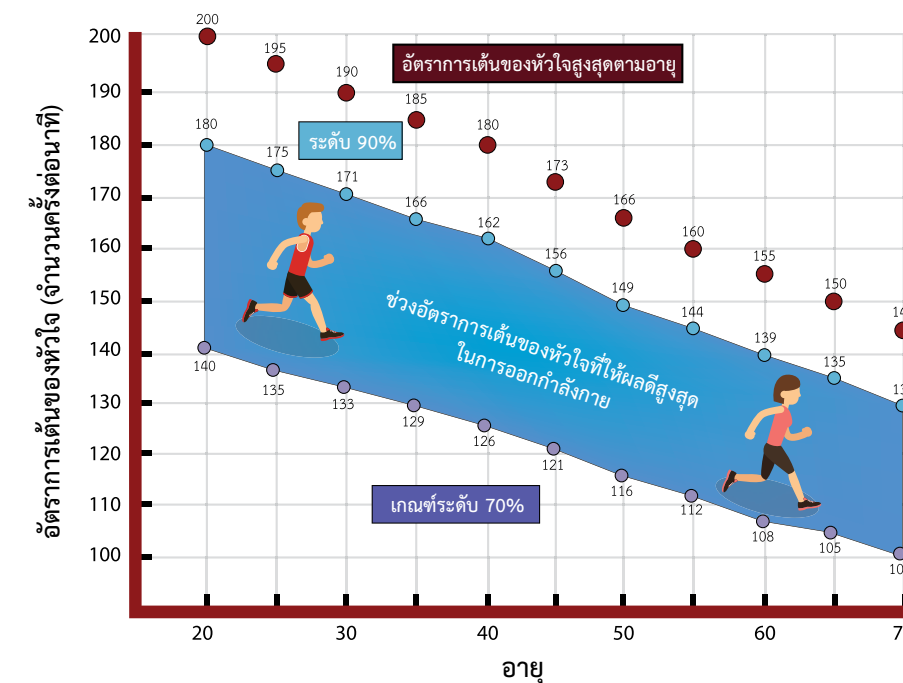
กิจกรรมที่ 1.1 □ ใคร ๆ ก็สามารถใช้ข้อมูลได้

ในชีวิตประจำวันเรามีการใช้ข้อมูลให้เป็นประโยชน์ต่อตนเองอยู่ตลอดเวลาให้อ่านสถานการณ์ที่เกี่ยวข้องกับการใช้ข้อมูลต่อไปนี้ และตอบคำถาม

“จาวามีปัญหาเรื่องความอ้วน และต้องการลดน้ำหนักโดยการวิ่งรอบสนามโรงเรียนทุกวัน แต่ปรากฏว่าน้ำหนักยังเท่าเดิม”

จาวาจึงค้นหาข้อมูล และพบข้อมูลดังกราฟ ซึ่งเป็นผลจากการศึกษาวิจัยและวิเคราะห์ข้อมูลทฤษฎีที่มีผู้จัดเก็บรวบรวมไว้โดยพบว่า การวิ่งที่ให้ประโยชน์สูงสุดต่อสุขภาพนั้น จะต้องวิ่งจนทำให้อัตราการเต้นของหัวใจ (Heart Rate) อยู่ระหว่าง 70 - 80% ของอัตราการเต้นของหัวใจสูงสุด (Maximum Heart Rate: MHR) และต้องวิ่งติดต่อกันอย่างน้อย 30 นาที

ให้นักเรียนนำข้อมูลดังกล่าวไปออกแบบโปรแกรมพัฒนาการวิ่งของจาวา เพื่อวางแผนการลดน้ำหนัก



*ผู้ออกกำลังกายเริ่มต้นไม่ควรเกินระดับ 80% ของ MHR

← กราฟแสดงช่วงของชีพจรที่ควรใช้ในการฝึกออกกำลังกายของกลุ่มอายุต่าง ๆ (ที่มา : <http://www.dms.moph.go.th/dmsweb/cpgcorner/heartex.pdf>)



ชวนคิด

ในด้านธุรกิจ ข้อมูลมีประโยชน์อย่างมากต่อการวางแผนการตลาด ตัวอย่างเช่น ร้านสะดวกซื้อ หรือห้างสรรพสินค้า ได้นำข้อมูลมาวิเคราะห์เพื่อจัดรายการส่งเสริมการขายสินค้า การจัดวางตำแหน่งสินค้าในร้าน และการเลือกสินค้าเพื่อนำมาจัดโปรโมชั่นลดราคา นักเรียนคิดว่าร้านสะดวกซื้อเหล่านี้เข้าถึงข้อมูลของกลุ่มลูกค้าและทราบความต้องการของลูกค้าได้อย่างไร



1.2 วิทยาการข้อมูล (data science)

ประเทศยักษ์ใหญ่อย่างสหรัฐอเมริกา ได้รายงานข้อมูลการประสบปัญหาการขาดแคลนผู้ที่มีทักษะการวิเคราะห์เชิงลึก และการจัดการข้อมูลจำนวนมาก เพื่อให้ได้ข้อมูลผลลัพธ์ที่มีความสำคัญ และเป็นประโยชน์ต่อการตัดสินใจของฝ่ายบริหารทั้งภาครัฐและเอกชน ซึ่งการทำงานในลักษณะนี้ จำเป็นต้องอาศัยความรู้และทักษะที่มีการผสมผสานหลายศาสตร์เข้าด้วยกันที่เรียกว่า “วิทยาการข้อมูล” ซึ่งเป็นการศึกษาถึงกระบวนการวิธีการ หรือเทคนิค ในการนำข้อมูลจำนวนมากมาประมวลผล เพื่อให้ได้องค์ความรู้ เข้าใจปรากฏการณ์หรือตีความ ทำนายหรือพยากรณ์ ค้นหารูปแบบหรือแนวโน้มจากข้อมูล และสามารถนำมาวิเคราะห์ต่อยอดเพื่อแนะนำทางเลือกที่เหมาะสม หรือใช้ในการตัดสินใจเพื่อประโยชน์สูงสุด

องค์ความรู้ของวิทยาการข้อมูล



ตัวอย่างของการใช้วิทยาการข้อมูลเพื่อประโยชน์ในด้านต่าง ๆ มีดังนี้



ตัวอย่างที่ 1.1 ข้อมูลช่วยอธิบายปรากฏการณ์ที่เกิดขึ้น

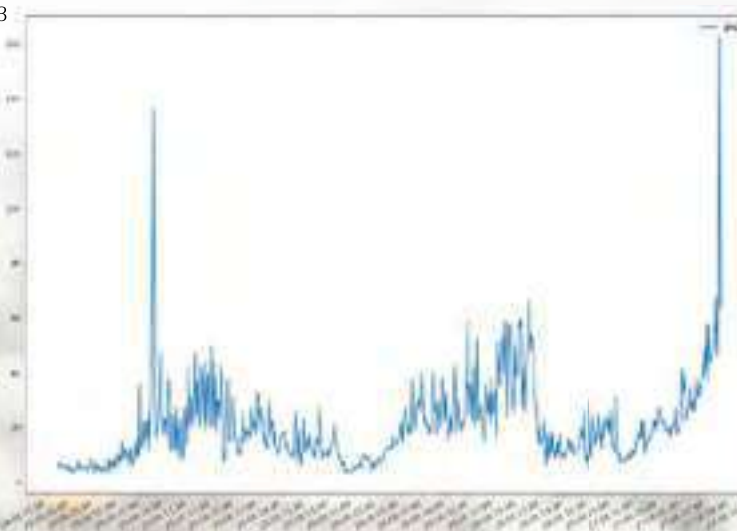


บริเวณที่อยู่อาศัยของ
นักเรียนมีปัญหาเรื่องฝุ่นละออง
หรือไม่ อย่างไร

ในบางภูมิภาคของประเทศมักพบปัญหาฝุ่นควันเป็นระยะ ๆ ซึ่งเมื่อนำเทคโนโลยีอินเทอร์เน็ตของสรรพสิ่ง (Internet of things: IoT) เข้ามาประยุกต์ใช้ร่วมกัน จะทำให้เราสามารถเก็บข้อมูลปริมาณฝุ่นในแต่ละช่วงเวลาได้อย่างต่อเนื่อง ตัวอย่างเช่น กราฟในรูป 1.4 แสดงถึงปริมาณฝุ่นเฉลี่ยในแต่ละช่วงเวลา บนถนนพระรามที่ 1 บริเวณแยกปทุมวัน ระหว่างวันที่ 25 - 27 มิถุนายน พ.ศ. 2561

นักเรียนจะเห็นว่า ปริมาณฝุ่นขนาด PM2.5 (Particulate Matter 2.5 คือฝุ่นละอองที่มีขนาดเล็กกว่า 2.5 ไมโครเมตร) มีค่าเกินมาตรฐาน คือมากกว่า $25 \mu\text{g}/\text{m}^3$ (ไมโครกรัมต่อลูกบาศก์เมตร) ในบางช่วง เช่น ในช่วงเวลาเช้าที่มีการจราจรหนาแน่น ค่าที่วัดได้สูงเกินกว่า $100 \mu\text{g}/\text{m}^3$

รูป 1.4 ปริมาณฝุ่นเฉลี่ยในแต่ละช่วงเวลา บนถนนพระรามที่ 1 บริเวณแยกปทุมวัน ระหว่างวันที่ 25 - 27 มิถุนายน พ.ศ. 2561



ตัวอย่างที่ 1.2 ข้อมูลช่วยเปลี่ยนมุมมองของสิ่งต่าง ๆ ต่อโลกได้

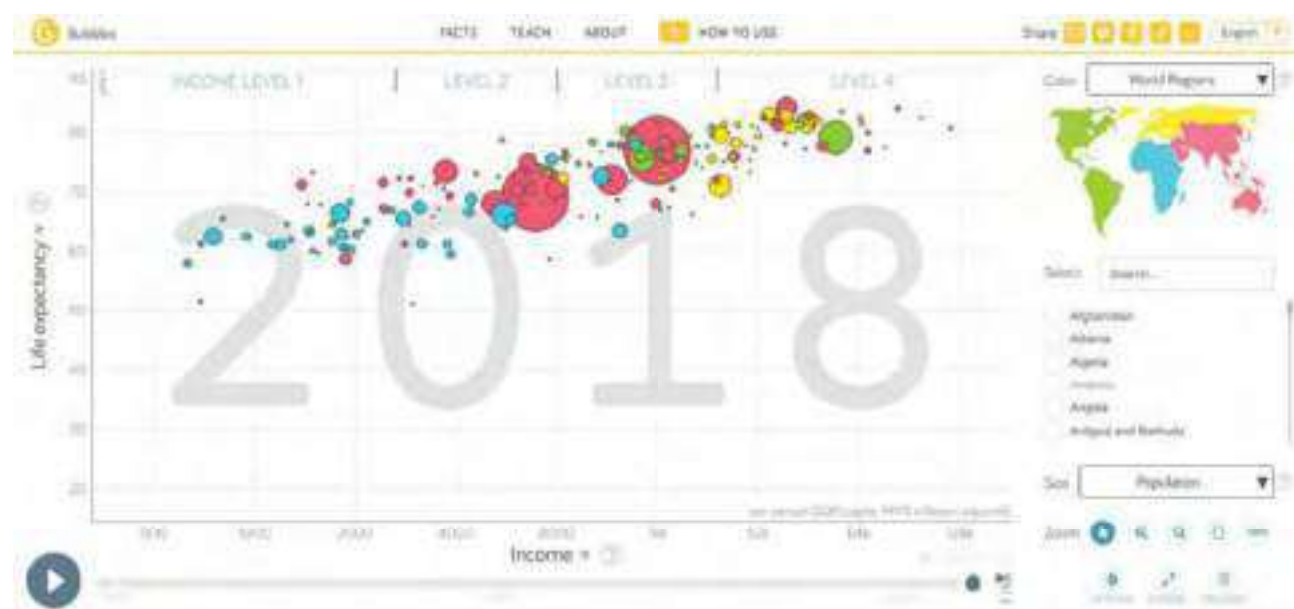


เราจะจัดกลุ่มประเทศต่าง ๆ
ในโลกใบนี้ได้กี่กลุ่ม อย่างไรบ้าง



นักเรียนอาจเคยได้ยินการเปรียบเทียบประเทศต่าง ๆ ในโลกว่า บางประเทศเป็นประเทศที่พัฒนาแล้ว และบางประเทศเป็นประเทศกำลังพัฒนา โดยมีความเชื่อว่าประเทศในกลุ่มที่พัฒนาแล้วน่าจะมีเศรษฐกิจ พึ่งพาการผลิตที่ทันสมัยกว่า มีรายได้และคุณภาพชีวิตที่ดีกว่า

นักเรียนอาจค้นหาข้อมูลของประเทศต่าง ๆ ที่เผยแพร่ได้ไม่ยาก ถ้าจะพิจารณาเกณฑ์ทั่วไปในการแบ่ง ประเทศเป็นสองกลุ่ม น่าจะเป็นเกณฑ์ด้านเศรษฐกิจ และเกณฑ์ด้านคุณภาพชีวิต นักเรียนสามารถนำข้อมูล ของประเทศต่าง ๆ มาวาดเป็นแผนภาพ โดยเกณฑ์ด้านเศรษฐกิจอาจแทนด้วยข้อมูลรายได้ และเกณฑ์ด้าน คุณภาพชีวิตแทนด้วยข้อมูลอายุขัยเฉลี่ย ดังรูป 1.5



แกน x แทนรายได้ของประชากร และแกน y แทนอายุขัยเฉลี่ย ขนาดของจุดแทนจำนวนประชากร ในที่นี้จุดแดงใหญ่สองจุด คือประเทศจีนและประเทศอินเดีย (จากเว็บไซต์ gapminder.org)



รูป 1.5 แสดงการกระจายของประเทศต่าง ๆ

แผนภาพนี้แสดงให้เห็นว่าความเชื่อที่ว่า ประเทศในโลกแบ่งเป็นสองกลุ่มอย่างชัดเจนนั้น ไม่จริง กล่าวคือประเทศต่าง ๆ มีระดับการพัฒนาที่แตกต่างกันจริง แต่ไม่ได้มีการกระจายที่แยกกลุ่ม ประเทศที่พัฒนาแล้วออกจากประเทศกลุ่มอื่น ๆ อย่างชัดเจน



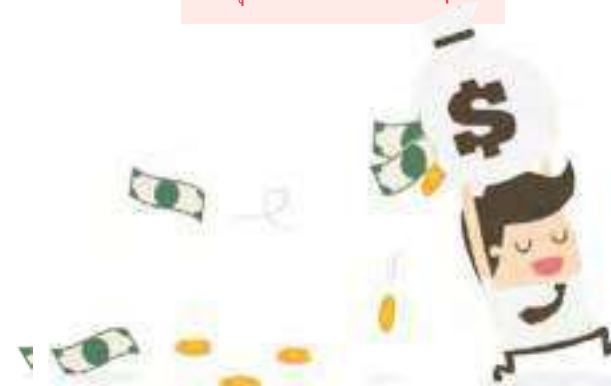
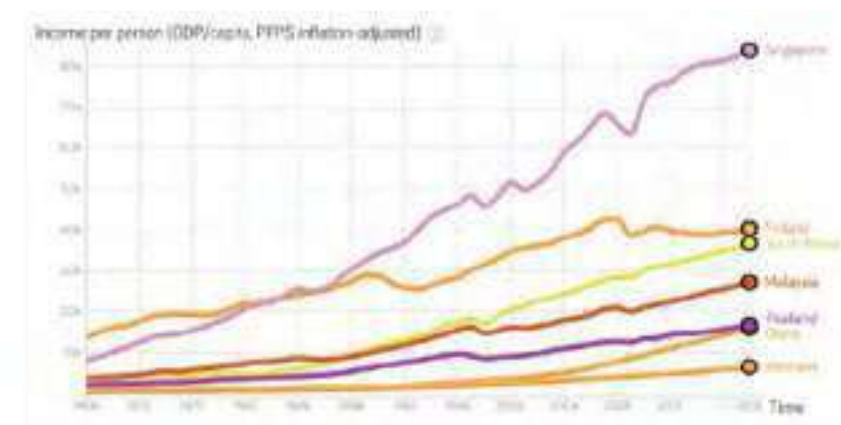
กิจกรรมที่ 1.2

ข้อมูลเปลี่ยนมุมมอง

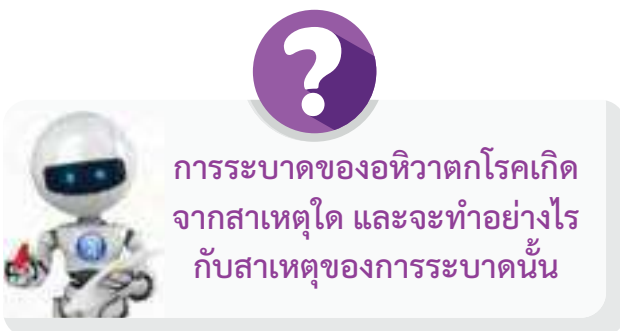


- 1 เกณฑ์ด้านคุณภาพชีวิตและเกณฑ์ด้านเศรษฐกิจที่ใช้ในตัวอย่างข้างต้นเป็นเพียงทางเลือกหนึ่งที่เป็นไปได้เท่านั้น ให้นักเรียนทดลองใช้เว็บไซต์ gapminder และเลือกข้อมูลอื่น ๆ แทนเกณฑ์ใดเกณฑ์หนึ่งหรือทั้งสองเกณฑ์ เช่น เลือกอัตราการตายของเด็ก (Child mortality) ในการวัดคุณภาพชีวิต และพิจารณาผลลัพธ์ของการกระจายที่ได้ นักเรียนจะพบช่องว่างระหว่างประเทศต่าง ๆ หรือไม่
- 2 ถ้าพิจารณาความเชื่ออื่น ๆ เช่น ความเชื่อเกี่ยวกับการพัฒนาประเทศ ข้อมูลก็อาจจะทำให้นักเรียนมีมุมมองต่อประเทศอื่น ๆ เปลี่ยนไป แผนภาพด้านล่างแสดงการเปลี่ยนแปลงของรายได้ในช่วง 50 ปีล่าสุด เทียบระหว่างประเทศไทยและประเทศอื่น ๆ นักเรียนทราบอะไรเพิ่มเติมจากแผนภาพนี้บ้าง

กราฟแสดงการเปลี่ยนแปลง
ของรายได้ในรอบ 50 ปี
ล่าสุดของประเทศต่าง ๆ



ตัวอย่างที่ 1.3 ข้อมูลช่วยในการตัดสินใจ



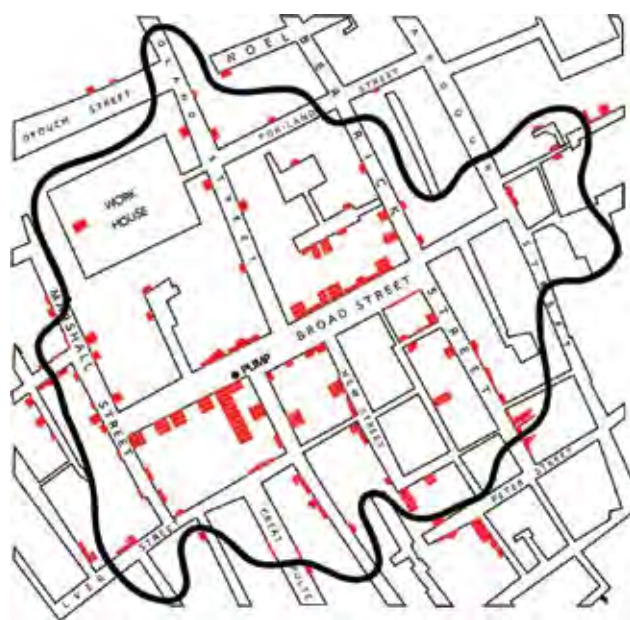
ในปี ค.ศ. 1854 (พ.ศ. 2397) ได้เกิดการระบาดของอหิวาตกโรคในบริเวณโกลเดนสแควร์ของกรุงลอนดอน มีผู้คนล้มตายเป็นจำนวนมากถึง 500 คนภายในเวลา 10 วัน ตั้งแต่วันที่ 31 สิงหาคม จนถึงวันที่ 9 กันยายน ค.ศ. 1854 ผังเมืองของกรุงลอนดอนแบ่งออกเป็นบล็อก ๆ และพบว่าการแพร่ระบาดของอหิวาตกโรค จะอยู่ภายในเฉพาะบางบล็อกของเมืองเท่านั้น จากการสืบหาข้อเท็จจริงของหมอจอห์น สโนว์ (Dr. John Snow) ซึ่งได้จัดทำแผนที่การกระจายของผู้ป่วย และตั้งสมมติฐานว่าการระบาดของอหิวาตกโรคอาจเกิดจากแหล่งน้ำ ซึ่งมีการปนเปื้อนของเชื้อโรค ในบริเวณโกลเดนสแควร์ ซึ่งมีแหล่งน้ำอยู่ 5 แหล่ง

ในการทดสอบสมมติฐานว่าการแพร่กระจายของอหิวาตกโรคเกิดจากการดื่มน้ำที่มีการปนเปื้อนของเชื้อโรคจากแหล่งใด เขาจึงสร้างแผนที่เพื่อแสดงข้อมูลจำนวนและตำแหน่งที่พบผู้ป่วยตายจากโรคด้วยแท่งสี่เหลี่ยม ดังรูป 1.6 เขาใช้เวลา 3 เดือนในการเก็บข้อมูล โดยการเดินสอบถามผู้อยู่อาศัยบริเวณนั้น สามารถเก็บข้อมูลผู้ป่วยตายได้ 616 คน บางคนไม่รู้ที่อยู่ชัดเจน บางคนป่วยตายนอกเมือง ดังนั้นแผนที่จึงมีแท่งสี่เหลี่ยมที่แสดงจำนวนผู้ป่วยตายเพียง 574 แท่ง จากแผนที่พบว่าจุดที่ผู้ป่วยตายจำนวนมากอยู่บริเวณถนนบรอดด์ ซึ่งอยู่ใกล้กับ

แหล่งน้ำที่คนในเมืองใช้มากที่สุด และคาดว่ามีการปนเปื้อนของเชื้อโรคที่เป็นสาเหตุของการแพร่ระบาดในครั้งนี้ เมื่อเทียบกับแหล่งน้ำอื่น ๆ ของเมือง จึงใช้เป็นข้อมูลในการตัดสินใจยกเลิกการใช้แหล่งน้ำนี้ จากนั้นพบว่า การแพร่ระบาดของโรคได้หยุดลง จึงน่าจะสรุปได้ว่าแหล่งน้ำนี้เป็นสาเหตุของการเกิดโรคระบาด การดำเนินการของหมอ สโนว์ แสดงให้เห็นถึงประโยชน์ของการใช้ข้อมูลในการตัดสินใจที่ถูกต้อง แม้ว่าในสมัยนั้นวิทยาการทางการแพทย์ยังไม่สามารถระบุได้ว่าโรคอหิวาตกโรคเป็นโรคติดต่อระบบทางเดินอาหาร ซึ่งเกิดจากการกินอาหารหรือดื่มน้ำที่มีเชื้อแบคทีเรียปนเปื้อนอยู่

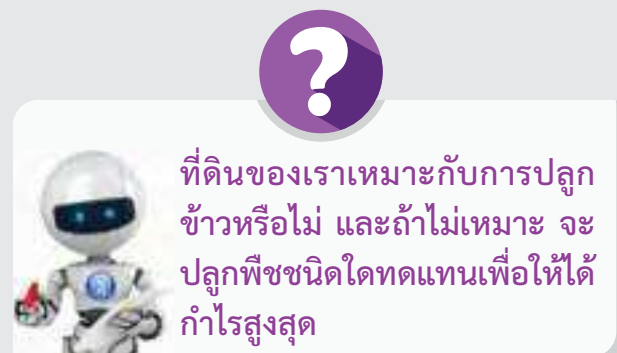


รูป 1.6 ข้อมูลผู้ป่วยที่ตายด้วยอหิวาตกโรคในปี ค.ศ. 1854 ณ กรุงลอนดอน ประเทศอังกฤษ



ที่มา : <https://datavizblog.com/2013/04/13/dataviz-history-the-ghost-map-myth-making-and-evolution-of-the-ghost-map/>

ตัวอย่างที่ 1.4 ข้อมูลช่วยเพิ่มมูลค่าให้กับบริการหรือผลิตภัณฑ์



ประเทศไทยใช้พื้นที่ปลูกข้าวมากถึง 71,104,553 ไร่ ซึ่งคิดเป็นร้อยละ 52.82 ของพื้นที่ที่ใช้ในการปลูกพืชเศรษฐกิจของประเทศทั้งหมด (ข้อมูล ณ วันที่ 24 มิถุนายน 2561) แต่ในจำนวนนี้ยังมีเกษตรกรที่ปลูกข้าวในพื้นที่ที่ไม่เหมาะสม ทั้งชนิดของดิน และการขนส่งไปยังแหล่งจำหน่ายสินค้า ดังนั้นเพื่อให้ได้ผลผลิตที่ดีในทุกพื้นที่ เกษตรกรควรปลูกพืชตามความเหมาะสมของสภาพดิน และวางแผนการปลูกพืชทดแทนเพื่อให้ได้ผลผลิตที่สอดคล้องกับความต้องการของตลาด การใช้ข้อมูลที่เกี่ยวข้องทั้งทางภูมิศาสตร์ แหล่งจำหน่ายสินค้า และการตลาด ล้วน

เป็นหัวใจสำคัญที่จะช่วยเหลือ และแก้ปัญหาให้กับเกษตรกรไทย สำหรับในประเทศไทยนั้น ข้อมูลดังกล่าวได้มีการจัดเก็บ และประมวลผลผ่านระบบแผนที่เกษตรเพื่อการบริหารจัดการเชิงรุกออนไลน์ (Agri-Map online) ดังรูป 1.7 เพื่อช่วยเหลือเกษตรกร และเจ้าหน้าที่ของรัฐในการบริหารจัดการพื้นที่เกษตรกรรมให้เกิดประโยชน์สูงสุด ตัวอย่างเช่น จังหวัดพิจิตร มีพื้นที่เหมาะสมในการปลูกข้าวนาปี คิดเป็นร้อยละ 91.90 ของพื้นที่เพาะปลูกทั้งหมด โดยมีพื้นที่ที่ไม่เหมาะสมในการเพาะปลูก คิดเป็นร้อยละ 8.10 ทางเลือกของเกษตรกรที่จะเปลี่ยนพื้นที่ที่ไม่เหมาะสมนี้ไปปลูกพืชชนิดอื่น ให้ใช้ข้อมูลสภาพดินที่เหมาะสมกับชนิดของพืชทดแทน รวมถึงข้อมูลที่ตั้งโรงงาน และแหล่งรับซื้อ ทั้งนี้สำนักงานเศรษฐกิจการเกษตร (สศก.) กระทรวงเกษตรและสหกรณ์ ได้เสนอแนะทางเลือกสำหรับเกษตรกร 2 ทางเลือก ดังนี้



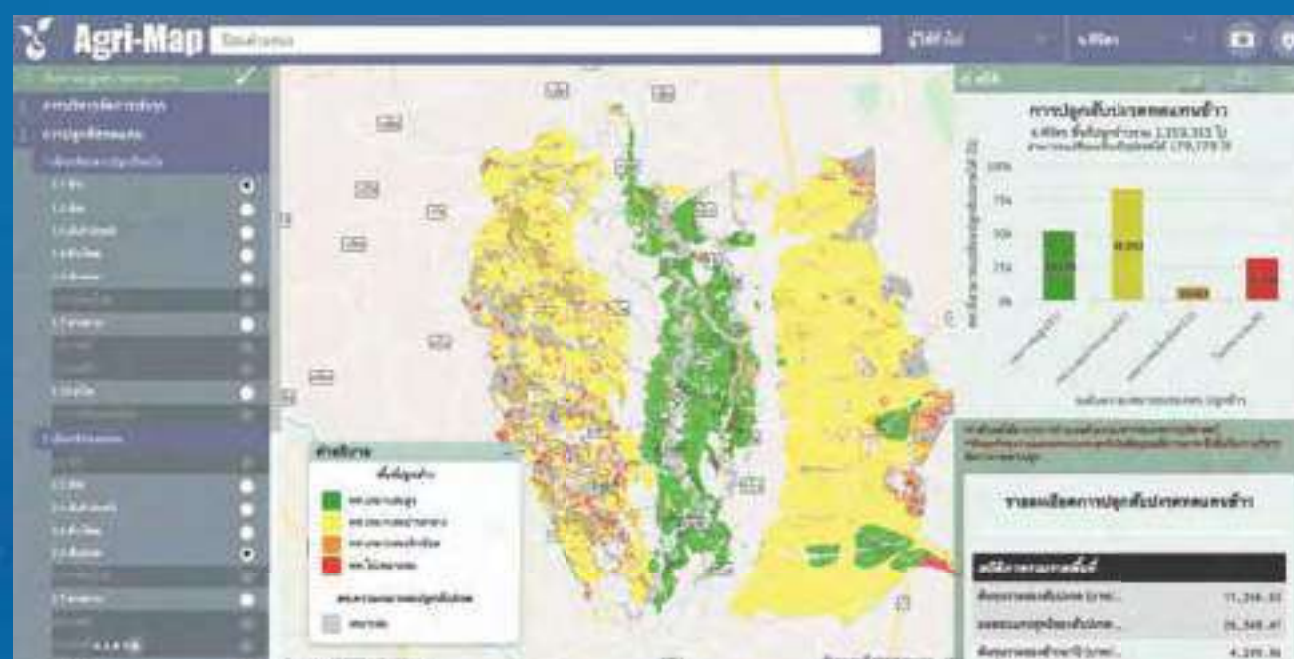


ทางเลือกที่ 1 ปรับเปลี่ยนพื้นที่ที่ไม่เหมาะสมสำหรับการปลูกข้าว ไปปลูกพืชที่เหมาะสมสำหรับพื้นที่นั้นแทน เช่น การปลูกอ้อยโรงงานทดแทน เพราะการปลูกข้าวในพื้นที่ที่ไม่เหมาะสมจะให้ผลตอบแทนสุทธิไร่ละ 1,002.76 บาท หากปรับเปลี่ยนไปปลูกอ้อยโรงงานจะได้ผลตอบแทนสุทธิไร่ละ 3,841.52 บาท



ทางเลือกที่ 2 ปรับเปลี่ยนพื้นที่ที่ไม่เหมาะสมสำหรับการปลูกข้าว ไปปลูกมะพร้าวทดแทน เพราะการปลูกข้าวในพื้นที่ที่ไม่เหมาะสมจะให้ผลตอบแทนสุทธิไร่ละ 1,002.76 บาท หากปรับเปลี่ยนไปปลูกมะพร้าวจะได้ผลตอบแทนสุทธิไร่ละ 5,393.21 บาท

(ข้อมูล ณ วันที่ 25 สิงหาคม 2560)



รูป 1.7 ระบบแผนที่เกษตรเพื่อการบริหารจัดการเชิงรุกออนไลน์



กิจกรรมที่ 1.3

ข้อมูลช่วยชุมชนและโลกอย่างไร



1

ทดลองเข้าเว็บไซต์ <http://agri-map-online.moac.go.th/loginAgri-Map> เพื่อสำรวจบริเวณพื้นที่ที่เป็นภูมิลำเนาของนักเรียน หรือเลือกพื้นที่ที่สนใจว่าเหมาะสำหรับทำการเกษตร เพาะปลูก หรือเลี้ยงสัตว์ชนิดใด พร้อมทั้งอธิบายเหตุผลประกอบ

2

ทดลองเข้าเว็บไซต์ <https://resourcetrade.earth/> เพื่อหาความสัมพันธ์ของการค้าระหว่างประเทศในโลก ให้ตั้งคำถามและหาคำตอบจากข้อมูลบนเว็บไซต์ และอธิบายให้เพื่อนฟังจากข้อมูลที่ค้นพบ

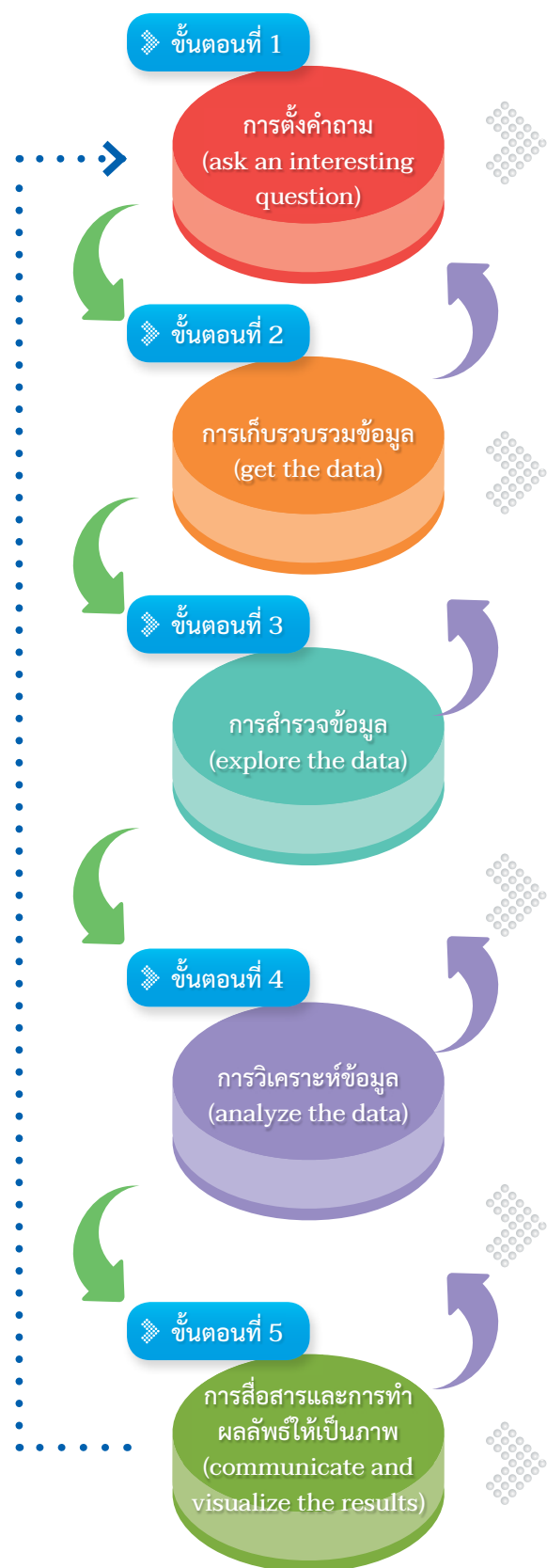


1.3 กระบวนการวิทยาการข้อมูล (data science process)

การเพิ่มมูลค่าให้กับผลิตภัณฑ์หรือบริการด้วยข้อมูลนั้น นอกจากจะต้องมีความเข้าใจเกี่ยวกับผลิตภัณฑ์หรือบริการแล้ว นักเรียนยังต้องเข้าใจกิจกรรมต่าง ๆ ที่เกี่ยวกับการจัดหาและประมวลผลข้อมูลอีกด้วย เนื่องจากกิจกรรมที่ต้องทำค่อนข้างหลากหลาย เพื่อให้ไม่ให้อึดอัดหรือพลาดประเด็นใดไป นักเรียนสามารถดำเนินการตามกระบวนการของวิทยาการข้อมูลที่ระบุขั้นตอนสำคัญต่าง ๆ ที่ประกอบด้วย การตั้งคำถาม การเก็บรวบรวมข้อมูล การสำรวจข้อมูล การวิเคราะห์ข้อมูล การสื่อสาร และการทำผลลัพธ์ให้เป็นภาพสำหรับผู้ใช้งานกลุ่มเป้าหมาย มีรายละเอียดดังรูป 1.8

DATA SCIENCE PROCESS





ตั้งคำถามที่ตนเองสนใจ

ต้องคำนึงถึงว่าจะเก็บข้อมูลเรื่องอะไร จากที่ไหน
จำนวนเท่าใด และความน่าเชื่อถือของแหล่งข้อมูล
ข้อมูลที่เก็บรวบรวมได้จะต้องทำการตรวจสอบ ขจัด
ข้อมูลที่ผิด หรือข้อมูลที่ไม่สมบูรณ์ เพื่อให้ได้ข้อมูล
นำเข้าที่ดี ไปสู่ผลลัพธ์ที่ดี เหมือนกับประโยคที่ว่า
“garbage in garbage out”

เป็นการทำความเข้าใจรูปแบบ และค่าของข้อมูล ใน
ขั้นตอนนี้เราจะต้องรวบรวมข้อมูล แล้วนำข้อมูลที่
รวบรวมมาพล็อต (plot) ทำให้เป็นภาพ (visualiza-
tions) หรือแผนภูมิ (charts) เพื่อให้มองเห็น
ความหมายที่ซ่อนเร้นอยู่ของข้อมูลผ่านกราฟ ซึ่งอาจ
พบความผิดปกติของข้อมูลได้

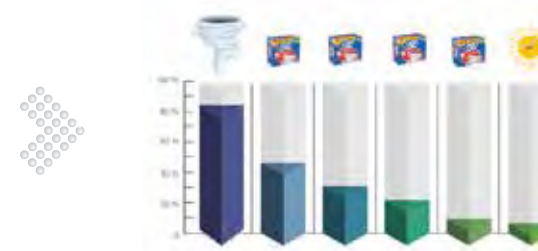
เพื่ออธิบายความหมาย ความสัมพันธ์ของข้อมูล และ
ทำนายเหตุการณ์ที่จะเกิดขึ้นในอนาคต

เป็นการสื่อสารผลลัพธ์ของข้อมูล โดยการถ่ายทอดเป็น
เรื่องราวหรือเป็นภาพให้ผู้อื่นเข้าใจว่าเราได้เรียนรู้อะไร
จากข้อมูล



รูป 1.8 กระบวนการวิทยาการข้อมูล

ที่มา : ปรับปรุงจากโมเดลของ Joe Blitzstein และ Hanspeter
Pfister



สำหรับขั้นตอนที่ 4 จะได้เรียนเพิ่มเติมในบทที่ 3 ต่อไป



สำหรับขั้นตอนที่ 5 จะได้เรียนเพิ่มเติมในบทที่ 4 ต่อไป

แต่ละขั้นตอนสามารถย้อนกลับไปได้เพื่อตรวจสอบกับคำถาม ทดลองทำ
แก้ไขใหม่ในทุก ๆ ขั้นตอนได้ สุดท้ายเมื่อได้ผลลัพธ์ของข้อมูลออกมาแล้ว
ควรย้อนกลับไปตรวจสอบกับคำถามที่ตั้งไว้อีกครั้ง

ห้างสรรพสินค้าแห่งหนึ่ง ต้องการ
วางแผนการสำรองป๊อปอาร์ตในปีถัดไป

ข้อมูลในอดีตแสดงให้เห็นว่าในช่วงฤดู
พายุเฮอริเคน ป๊อปอาร์ตมียอดขาย
สูงมาก

สำหรับขั้นตอนที่ 2
และ 3 นักเรียนจะ
ได้เรียนเพิ่มเติมใน
บทที่ 2 ต่อไป

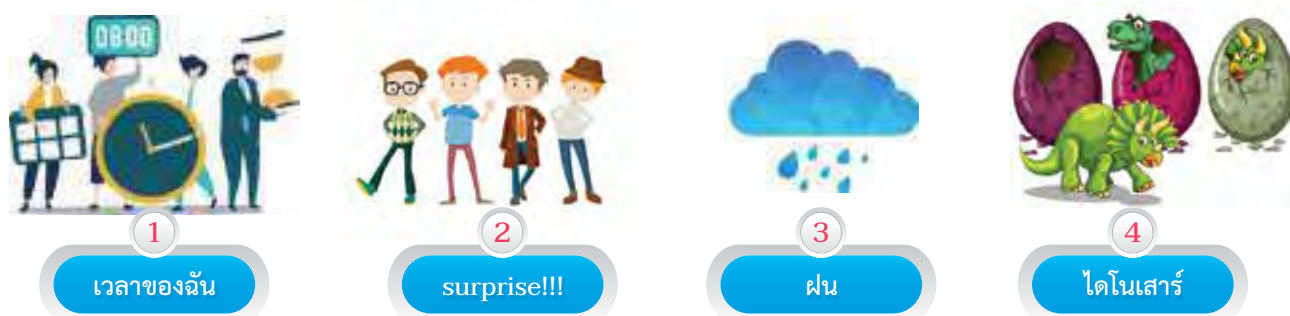
มียอดขายเพิ่มขึ้นเป็น 7 เท่าของช่วง
เวลาอื่น

ต้องการทำนายว่า ในอนาคต หากเข้าสู่
ช่วงฤดูที่มีพายุเฮอริเคน ควรเก็บสำรอง
ป๊อปอาร์ตจำนวนเท่าไร เพื่อให้เพียงพอ
กับความต้องการของลูกค้าในช่วงเวลา
นั้น

ตัวอย่างนี้แสดงให้เห็นถึงประโยชน์ของ
การวิเคราะห์เพื่อการทำนาย ที่แสดงถึง
ความสัมพันธ์กันระหว่างป๊อปอาร์ตกับ
ช่วงฤดูที่มีพายุเฮอริเคนซึ่งถ้าไม่มีข้อมูล
ประกอบ อาจมองว่าสองสิ่งนี้ไม่น่าจะ
สัมพันธ์กัน

กิจกรรมที่ 1.4 □ แนวคิดของวิทยาการข้อมูล

สร้างโมเดลของคำตอบ โดยใช้กระบวนการวิทยาการข้อมูลจากกิจกรรมใดกิจกรรมหนึ่งต่อไปนี้



1.4 การคิดเชิงออกแบบ (design thinking) สำหรับวิทยาการข้อมูล

การนำข้อมูลมาใช้เพื่อสื่อสารถึงแม้จะทำให้เข้าใจปัญหาหรือสถานการณ์มากยิ่งขึ้น แต่ถ้าไม่เข้าใจถึงความต้องการที่แท้จริงของผู้ใช้ ก็จะทำให้การนำเสนอข้อมูลผลลัพธ์ไม่ประสบผลสำเร็จ เนื่องจากผู้ใช้ไม่สามารถใช้ประโยชน์จากข้อมูลที่นำเสนอได้นั้นได้ เช่น ข้อมูลที่นำเสนอมีปริมาณมากหรือละเอียดเกินความต้องการ เมื่อผู้ใช้พิจารณาข้อมูลแล้วคิดว่าไม่จำเป็นสำหรับตนเอง ข้อบกพร่องนี้อาจทำให้การพัฒนาสินค้าหรือผลิตภัณฑ์นั้นไม่สามารถเข้าถึงความต้องการของผู้ใช้ได้อย่างแท้จริง

ตัวอย่างที่ 1.5

บริษัทสตาร์ทอัพต้องการพัฒนาแอปพลิเคชันสำหรับนักกีฬาอาชีพ โดยมีการติดตามข้อมูลการนอนหลับ และข้อมูลการทำงานของหัวใจ ได้แก่ อัตราการเต้นของหัวใจขณะหลับ (Sleeping HR) และอัตราการแปรผันของการเต้นของหัวใจ (Heart Rate Variability: HRV) ซึ่งแสดงเป็นกราฟดังรูป 1.9 โดยคาดหวังว่านักกีฬาจะใช้ข้อมูลที่แสดงด้วยกราฟนี้ เพื่อประกอบการตัดสินใจในการพัฒนาศักยภาพด้านกีฬาของตนเอง แต่พบว่านักกีฬาไม่สามารถใช้ข้อมูลเหล่านี้ในการตัดสินใจได้ จากการสอบถามนักกีฬาและโค้ชหลาย ๆ คน ที่ใช้แอปพลิเคชันนี้ สรุปว่าการแสดงข้อมูลให้เป็นภาพด้วยกราฟไม่ใช่ปัญหาของการใช้งาน เพียงแต่ไม่ตรงกับความต้องการ เพราะผู้ใช้ต้องการทราบว่าควรเข้านอนเวลาใด และตื่นนอนเวลาใด เพื่อให้มีการพักผ่อนที่เพียงพอต่อร่างกาย ผู้ใช้รู้สึกว่าการแสดงข้อมูลจากกราฟมีประโยชน์น้อยกว่าการทราบวิธี



การปฏิบัติที่ถูกต้อง ดังนั้นการปรับปรุงแอปพลิเคชันให้เพิ่มการแสดงระยะเวลาการนอนและคำแนะนำในการปฏิบัติตนให้เป็นนักกีฬาที่มีศักยภาพ ควรนำมาแสดงก่อนการแสดงผลข้อมูลอื่น ๆ ด้วยกราฟ ดังรูป 1.10 (ที่มา : Jon Wettersten และ Dean Malmgren, 2018)



ตัวอย่างนี้แสดงให้เห็นว่าการจะใช้วิทยาการข้อมูลให้เกิดประโยชน์สูงสุดต่อผู้ใช้งานนั้น ต้องทราบความต้องการที่ชัดเจนของผู้ใช้ด้วย แนวทางการทำงานในลักษณะนี้คล้ายคลึงกับการทำงานของนักออกแบบที่พยายามสร้างผลิตภัณฑ์ต้นแบบที่นอกจากจะเกิดประโยชน์ต่อผู้ใช้งานได้จริงแล้ว ยังต้องมีความสวยงามน่าใช้ไปพร้อม ๆ กันอีกด้วย ดังนั้น ในการแก้ปัญหาหรือพัฒนาผลิตภัณฑ์ด้วยข้อมูล นักเรียนอาจใช้แนวคิดเดียวกับนักออกแบบก็ได้ แนวคิดดังกล่าวเรียกรวม ๆ ว่า การคิดเชิงออกแบบ (design thinking) ลองพิจารณาสถานการณ์การทำงานของนักออกแบบ ดังต่อไปนี้



นกลีเกเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 5 ต้องการเก็บเงินด้วยตนเอง เพื่อใช้เรียนต่อ และซื้อของที่ต้องการ นกลีเกสามารถเขียนโปรแกรมได้ จึงคิดพัฒนาแอปพลิเคชันบันทึกรายรับรายจ่ายในครัวเรือนที่นอกจากจะนำมาใช้เองแล้ว ยังสามารถนำไปขายเพื่อสร้างรายได้เสริมให้กับครอบครัวได้ด้วย

ในการออกแบบแอปพลิเคชัน เด็กต้องทราบก่อนว่า ผู้ใช้อยากทราบข้อมูลอะไรบ้าง เธอจึงเริ่มสำรวจความต้องการของตนเองก่อน และเขียนรายการข้อมูลทั้งหมดที่ต้องการออกมา ถ้านักเด็กหยุดคิดและพัฒนาทันที แอปพลิเคชันที่ได้คงไม่ต่างจากแอปพลิเคชันอื่นทั่วไป



แต่นักเด็กไม่หยุดคิดแค่นี้ เธอคิดหาเหตุผลว่า ทำไมเธอถึงต้องการรายการข้อมูลเหล่านี้ เธอพบว่า มีเหตุผล 2 ข้อที่สำคัญ คือ 1) เธอต้องการติดตามรายรับรายจ่ายของตนเอง เพราะต้องการซื้อของสำคัญบางอย่าง และ 2) เธออยากทราบว่าเธอใช้จ่ายเงินซื้อของประเภทใดมากเกินไปหรือไม่ ด้วยเหตุนี้ เธอจึงออกแบบแอปพลิเคชันให้มีส่วนแสดงจำนวนเงินเก็บสะสมสำหรับซื้อของสำคัญ และอัตราส่วนของการใช้จ่ายแยกตามประเภท โดยการแสดงผลทั้งสองส่วนนี้ถูกออกแบบให้เรียกดูง่ายและสวยงาม



แต่ทั้งหมดนี้ เป็นความคิดของนักเด็กคนเดียว ซึ่งอาจไม่ตรงกับความต้องการของคนอื่น เธอจึงต้องสอบถามความต้องการของผู้ใช้คนอื่นเพิ่มเติม ซึ่งเป็นนักเรียนชั้นมัธยมศึกษา 5 คน ให้ทดลองใช้โปรแกรมตามรูปตัวอย่างหน้าจอที่เตรียมไว้ โดยยังไม่ได้พัฒนาแอปพลิเคชันขึ้นมา



จากตัวอย่างพบว่า การออกแบบทั่วไปนั้นใช้ได้ แต่การบันทึกรายการจ่ายโดยแยกเป็นประเภทแบบละเอียดนั้นซับซ้อนเกินไป เธอจึงนำข้อมูลที่ได้จากการทดลองมาปรับการออกแบบ ขณะนี้เธอเริ่มมั่นใจในทิศทางของแอปพลิเคชันแล้ว จึงเริ่มพัฒนา และเตรียมนำไปให้ผู้ใช้งานกลุ่มใหญ่ทดลองใช้ต่อไป



ชวนคิด

จากตัวอย่างกระบวนการออกแบบโปรแกรมของนักเด็กนักเรียน คิดว่าการทำงานของนักเด็กในขั้นตอนต่าง ๆ มีประโยชน์อย่างไร



จากตัวอย่าง นักเรียนจะพบว่าการออกแบบการแก้ปัญหาหรือพัฒนาผลิตภัณฑ์ด้วยข้อมูลของนักเรียนนั้น ถึงแม้จะใช้กระบวนการลองผิดลองถูก แต่ก็มีการใช้หลักการพื้นฐานของการคิดเชิงออกแบบ ดังนี้



การมองในมุมมองของผู้ใช้ การเอาใจเขามาใส่ใจเรา เพื่อสร้างความเข้าใจต่อผู้ใช้อย่างลึกซึ้ง สังเกตว่านักเรียนไม่ได้หยุดแค่รายการข้อมูลที่ต้องการทราบ แต่เธอยังพยายามคิดว่าเธอต้องการทราบข้อมูลเหล่านี้ไปทำไม การตั้งคำถามด้วย “ทำไม” เป็นขั้นตอนที่ดีในการสร้างความเข้าใจผู้ใช้ นอกจากนี้ นักเรียนยังสนใจวิธีการแสดงข้อมูลที่เข้าใจง่ายและประเด็นความสวยงาม



การลองผิดลองถูกและเรียนรู้ผ่านการทดลองกับกลุ่มผู้ใช้จริง นักเรียนแม้จะพยายามคิดว่าผู้ใช้ต้องการอะไร แต่ก็ไม่ได้ลืมที่จะนำสิ่งที่เธอคาดการณ์ไปทดสอบกับผู้ใช้จริง การสร้างแบบจำลองการใช้งานโปรแกรมด้วยภาพสำหรับสื่อสารหรือพูดคุยกับผู้ใช้เป็นวิธีที่ดีวิธีหนึ่งในการทำให้เห็นภาพสิ่งที่คิดนั้นได้ชัดเจน และช่วยให้นักเรียนได้รับข้อคิดเห็นที่ชัดเจนมากขึ้น



การทำซ้ำและปรับปรุง การออกแบบที่ดีมักผ่านการปรับปรุงและแก้ไขหลายครั้ง ข้อคิดเห็นจากผู้ใช้ เป็นสิ่งที่นักออกแบบคาดไม่ถึงมาก่อน การนำข้อมูลดังกล่าวมาปรับปรุงงานออกแบบจึงเป็นสิ่งที่สำคัญมาก และควรกล้าที่จะลองผิดลองถูก

สังเกตว่าการที่นักเรียนพยายามมองในมุมมองของผู้ใช้ ทำให้เธอมั่นใจว่าคำถามที่เธอต้องการทราบเกี่ยวกับข้อมูลรายรับรายจ่ายนั้นตรงกับผู้ใช้คนอื่นด้วยเช่นเดียวกัน คำถามนั้นคือ

- i เธอเก็บเงินที่จะซื้อของสำคัญที่ต้องการได้มากน้อยเพียงใด
- i เธอใช้จ่ายเงินไปกับสินค้าบางประเภทมากเกินไปหรือไม่

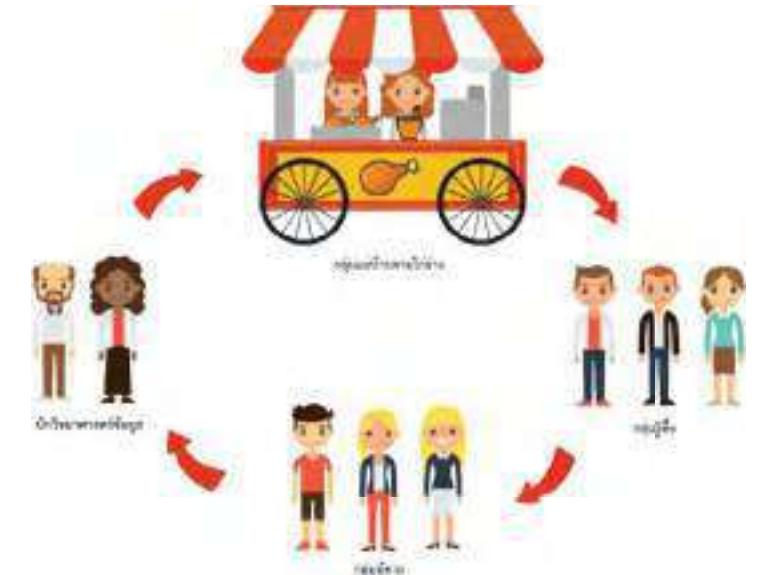
คำถามเหล่านี้เป็นตัวอย่างของจุดตั้งต้นในการเริ่มแก้ปัญหาหรือพัฒนาผลิตภัณฑ์ด้วยข้อมูล

A กิจกรรมที่ 1.5 □ คิดแบบนักออกแบบ

สถานการณ์ : ต้องการหาวิธีช่วยกลุ่มแม่บ้านในชุมชนขายไก่ย่าง ส้มตำ

แบ่งนักเรียนออกเป็นกลุ่ม โดยแต่ละกลุ่มประกอบด้วย ผู้ซื้อ ผู้ขาย และนักวิทยาศาสตร์ข้อมูล โดยให้กลุ่มนักวิทยาศาสตร์ข้อมูลเก็บข้อมูลความต้องการ และผลกระทบทั้งทางบวกและลบ จากกลุ่มผู้ซื้อ ผู้ขาย เพื่อตอบคำถามต่อไปนี้

- i ใครคือกลุ่มเป้าหมาย
- i ผลิตภัณฑ์เป็นที่ถูกใจผู้ซื้อหรือไม่ เพราะเหตุใด
- i จะมีแนวทางการทำให้ผลิตภัณฑ์นี้ขายดีขึ้นได้อย่างไร
- i ข้อมูลเพิ่มเติมมีอะไรบ้าง และหาได้จากที่ใด



A กิจกรรมที่ 1.6 □ โครงการวิทยาการข้อมูลของฉัน

ให้นักเรียนแต่ละกลุ่มเลือกประเด็นใดประเด็นหนึ่งต่อไปนี้

- i ปัญหาที่พบในท้องถิ่น ประเทศ หรือนานาชาติ
- i การเพิ่มมูลค่าให้กับบริการหรือผลิตภัณฑ์ในชุมชน

ให้สมาชิกในกลุ่มร่วมกันอภิปรายในประเด็นที่เลือก โดยพิจารณาจากมุมมองของผู้ที่เกี่ยวข้องหรือผู้ที่ได้รับผลกระทบทุกกลุ่ม เพื่อตัดสินใจเลือกหัวข้อโครงการและเขียนเค้าโครงโครงการ

