ข้อมูลมหัศจรรย์และถิ่นที่อยูู

สนุก กับข้อมูล เรียนรู้แนวคิดพื้นฐาน

ทดลองตั้งโจทย์จากข้อมูล ที่หาได้

สัมผัสความสาหัสในการนําาข้อมูล ไปใช้

จิตร์ท ัศน์ ฝักเจริญ ผล – ภาควิชาวิศวกรรมคอมพิวเตอร์ม.เกษตรศาสตร์

Image by Gerd Altmann From https://pixabay.com/illustrations/web-connection-data-wave-binary-4809584/

กิจกรรม 1

● แบูง กลูุม กลูุม ละ 10 คน

● เก็บข้อมูล ของสมาชิกในกลูุม :

– สูว นสูง

– ขนาดรอบหัว

– ขนาดรอบข้อมือ

– ขนาดเท้า

– อาชีพ

ข้อ มูล แบบมีโครงสร้าง

● รูป แบบของ “ข้อมูล ” ที่เ ราหาได้ม ักอยููใ นรูป ของ

ตารางชื่อ อายุ

ตําาแหน่ง เงินเดือน (ปี) (บาท)

อายุง าน

(ปี)

สมดี 40 ออกแบบ 35,000 4

สมใจ 25 ขนสินค้า 40,000 2

สมควร 37 ขายหน้าร้าน 22,000 3

ศัพท์แสง (หลายชื่อเรียก)

ชื่อ อายุ

ตําาแหน่ง (ปี) เงินเดือน

(บาท)

คุณ ลักษณะ (attribute, feature)

ตัวแปร (variable)

จําานวนคุณ ลักษณะ = มิต ิ

อายุง าน

(ปี)

สมดี 40 ออกแบบ 35,000 4

สมใจ 25 ขนสินค้า 40,000 2

ขายหน้า สมควร 37 ร้า น 22,000 3

ตัวอย่า ง (example)

กรณี(case)

เรคอร์ด

Source: http://statbbi.nso.go.th/staticreport/Page/sector/TH/report/sector\_08\_11\_TH\_.xlsx From: http://statbbi.nso.go.th/staticreport/page/sector/th/08.aspx

ตัวอยูา งข้อมูล เชิงสถิติ(1)

Source: http://statbbi.nso.go.th/staticreport/Page/sector/TH/report/sector\_08\_5\_TH\_.xlsx

ตัวอยูา ง (2)

ที่มาของข้อมูล

● สอบถาม/เก็บจากกลูุม เป้าหมาย

– เก็บ ทั้งหมด

– สูุม เก็บบางสูว น (sample)

● ความคลาดเคลื่อน

● อคติและความลําาเอียง (bias)

– ข้อมูลตกหลูน

ข้อมูล บอกอะไรกับเรา

● ลัก ษณะการกระจาย

ข้อมูล บอกอะไรกับเรา

● “คูา กลาง” (1)

ข้อมูล บอกอะไรกับเรา

● “คูา กลาง” (2)

ข้อควรระวัง - คูา กลาง

?? เป็นค่า เฉลี่ยของอะไร ??

ค่า เฉลี่ย = 23,542.27 บาท

ข้อมูล บอกอะไรกับเรา

● กลูุม ก้อน (ผูานทางลักษณะการกระจาย) [1 มิติ]

ข้อมูล บอกอะไรกับเรา

● กลูุม ก้อน (ผูานทางลักษณะการกระจาย) [1 มิติ]

ข้อมูล บอกอะไรกับเรา

● กลูุม ก้อน (ผูานทางลักษณะการกระจาย) [2 มิติ]

Source: http://statbbi.nso.go.th/staticreport/Page/sector/TH/report/sector\_03\_20\_TH\_.xlsx From: http://statbbi.nso.go.th/staticreport/page/sector/th/03.aspx

ข้อมูล บอกอะไรกับเรา

● แนวโน้ม

ข้อมูล บอกอะไรกับเรา

● ความสัมพันธ์

ข้อมูล บอกอะไรกับเรา

● การทําา นาย (prediction)

CRAAP test

● Currency

● Relevance

● Authority

● Accuracy

● Purpose

Source: https://en.wikipedia.org/wiki/CRAAP\_test CRAAP test was developed by Sarah Blakeslee at California State University, Chico.

การเชื่อมโยงข้อมูล

ข้อมูลอาจไมูบ อกอะไรกับเรา

● (แตูเราเข้าใจผิด – เรียนเพิ่มวิชา STAT101)

ความไมูป กติของข้อมูล

● ข้อมูล หาย

● ข้อมูล ผิด

● ข้อมูล ขอบ (outliers)

Bias ในข้อมูล

● Selection bias

● Exclusion bias

● Observer bias

กิจกรรม 2

● จากกลูุม เดิม แบูง กลูุม ยูอ ย

– plot จุด ตัวอยูา งลงบนกราฟโดยเลือก attribute สองคูา

– ทําา หลาย ๆ แบบ (กลูุม ยูอ ยละแบบ)

– พิจารณาแนวโน้มความสัมพันธ์ระหวูา ง attribute

กระบวนการข้อมูล

● ตั้งคําา ถาม

● เก็บข้อมูล / เก็บกวาดข้อมูล

● ศึกษาข้อมูล / วิเคราะห์ข ้อมูล

● นําาเสนอ

● (กลับไปข้อแรก)

ประเภทของคุณ สมบัติ(หรือตัวแปร)

● คูา ตัวเลขแบบตูอ เนื่อง

● คูาตัวเลขแบบไมูต ูอ เนื่อง

● ประเภท

– ประเภทที่มีประเภทยูอ ย

● ชูว ง

Source: http://statbbi.nso.go.th/staticreport/Page/sector/TH/report/sector\_05\_15\_TH\_.xlsx From:http://statbbi.nso.go.th/staticreport/page/sector/th/05.aspx

ข้อมูล แบบประเภท (ตัวอยูา ง)

แหลูง ข้อมูลสาธารณะ (ในประเทศ)

● สําานักงานสถิติแหูง ชาติ

● Government Open Data (https://data.go.th)

● เฉพาะด้าน เชูน TPMAP (https://www.tpmap.in.th)

● หนูวยงานตูา งมักมีหลายระบบ ต้องลองค้น หา

– กระทรวงพาณิช ย์(https://data.moc.go.th/ มีapi https://data.moc.go.th/developer )

– กระทรวงศึกษาธิการ (https://ods.moe.go.th:8888/) + API?

– ศูน ย์ข้อมูลอุบ ัติเหตุ(http://www.thairsc.com/)

– กรมประมง (หนังสือสถิติการประมง https://www.fisheries.go.th/strategy-stat/index.php/document-public)

– กระทรวงสาธารณสุข (สถิติสาธารณสุข / มีห ลายระบบ)

– กระทรวงคมนาคม (http://www.motoc.mot.go.th/stat/)

– กระทรวงการทูอ งเที่ยว (https://www.mots.go.th/more\_news\_new.php?cid=411)

– สถิติข้อมูล เกี่ยวกับการจัดการนํา้า

แหลูง ข้อมูล (นานาชาติ)

● WHO (

https://www.who.int/healthinfo/statistics/en/)

● World bank (https://data.worldbank.org/)

● PISA (https://www.oecd.org/pisa/data/) มีdata

explorer

ข้อมูล จาก API

● การอูา นข้อมูล

– curl

– เขีย นโปรแกรม (ด้วยภาษาเชูน python, ruby, etc..)

กิจกรรม 3

● แบูง กลูุม ใหมูต ามความสนใจ

– เลือกแหลูง ข้อมูล 2 – 3 ที่

– สําารวจข้อมูล

– จากแหลูง ข้อมูลดังกลูา ว

● สามารถนําามาศึกษาความสัมพันธ์ของอะไรได้บ้าง?

● อาจจะนําามาตอบคําา ถามอะไรได้บ ้าง?