

ปฏิบัติการที่ 11

การแสดงผลข้อมูลสำหรับข้อมูลเชิงตัวเลขและข้อมูลจำแนกประเภท

วัตถุประสงค์

1. เพื่อให้สามารถนำเสนอข้อมูลเชิงตัวเลขโดยการสร้างการแสดงผลด้วยฮิสโตแกรม แผนภาพการกระจาย และแผนภูมิรูปไวโอลิน ได้
2. เพื่อให้สามารถนำเสนอข้อมูลจำแนกประเภทโดยการสร้างการแสดงผลด้วยแผนภูมิแท่ง และแผนภูมิต้นไม้ ได้เพื่อให้สามารถสร้างเครื่องมือกรองข้อมูลเบื้องต้นสำหรับการเลือกข้อมูลเพื่อแสดงผลได้


1. ชุดข้อมูลปฏิบัติการ

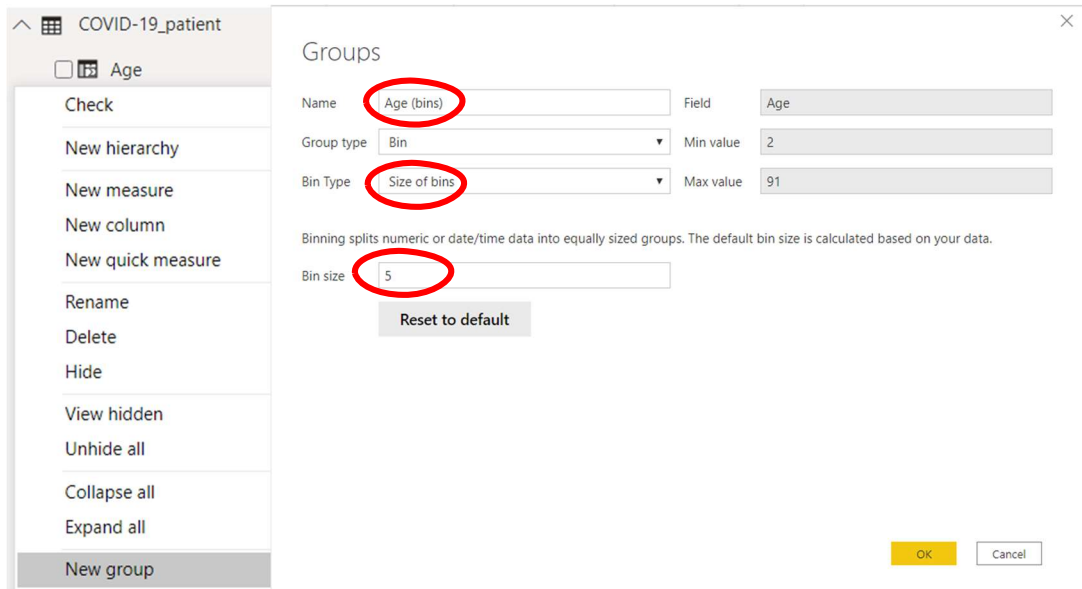
- ชุดข้อมูล COVID-19 Patient (สำหรับการสาธิตและฝึกปฏิบัติการ)
- ชุดข้อมูล 120 Years of Olympic History athletes and Results (สำหรับการสาธิตและฝึกปฏิบัติการ)


2. ขั้นตอนปฏิบัติการ

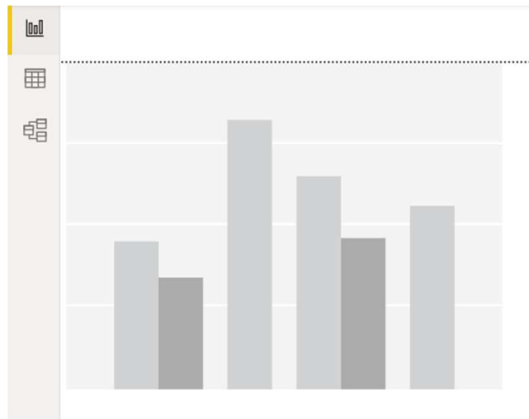
1. ดาวน์โหลดเพิ่มข้อมูล Lab_11.pbix จากเว็บไซต์กระบวนวิชา
2. เปิดเพิ่มข้อมูล Lab_11.pbix โดยโปรแกรม Microsoft Power BI
3. เพิ่มข้อมูล Lab_11.pbix ประกอบด้วยข้อมูล 2 ตารางข้อมูล คือ COVID-19_patient และ athlete_events


athlete_events	COVID-19_patient
Σ Age	Age
■ BMI	Σ birth_year
City	▶ ■ confirmed_date
Event	Σ contact_number
Games	country
Σ Height	▶ ■ deceased_date
ID	Σ disease
Medal	group
Name	Σ infected_by
NOC	Σ infection_order
Season	infection_reason
Sex	patient_id
Sport	region
Team	▶ ■ released_date
Σ Weight	sex
Σ Year	state

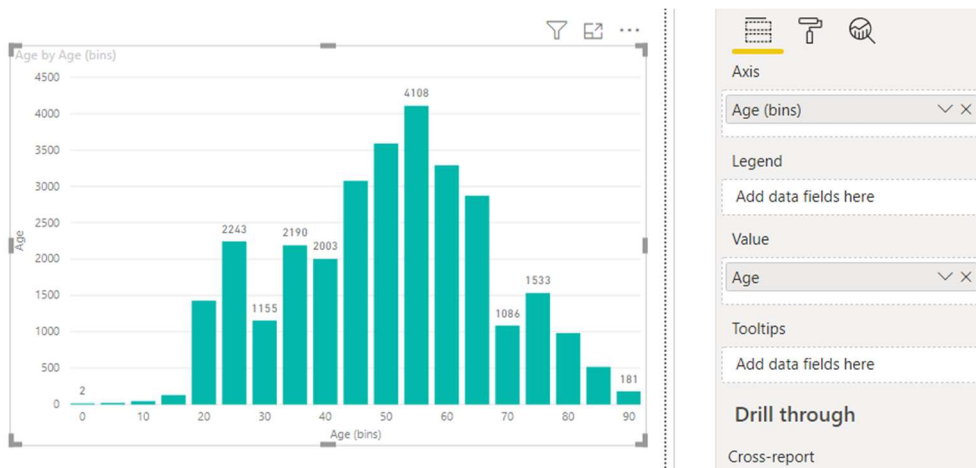
4. ไปที่หน้าที่ Report โดยคลิกที่  ทางด้านซ้ายของหน้าต่างโปรแกรม จะปรากฏพื้นที่ว่างเปล่า ดังรูป หน้าต่างนี้เป็นพื้นที่สำหรับสร้างการแสดงผลข้อมูล จากหน้าต่าง Report จะมีแผงหน้าต่างย่อย Fields Visualization และ Filters ทางด้านขวาของหน้าต่างโปรแกรม ซึ่งจะเป็นเครื่องมือช่วยในการสร้างการแสดงผลข้อมูล
5. หน้าด้านล่างหน้าต่าง Report จะปรากฏแท็บ Page ซึ่ง นักศึกษาสามารถสร้างเพจสำหรับรายงานได้จากส่วนนี้ ให้นักศึกษาเปลี่ยนชื่อ Page 1 โดยคลิกขวาที่แท็บ Page 1 เลือก Rename Page จากนั้นเปลี่ยนชื่อเป็น Covid-19 Data
6. การสร้างฮิสโตแกรม (Histogram) ของข้อมูลตัวเลข ในที่นี้จะสร้างฮิสโตแกรมของข้อมูลอายุ (Age) จากตารางข้อมูล COVID-19_patient ซึ่งเป็นข้อมูลเชิงตัวเลข จะต้องแบ่งข้อมูลออกเป็นกลุ่ม (Bin) โดยการคลิกขวาที่ฟิลด์ Age ที่หน้าต่าง Field แล้วเลือก New Group จะปรากฏหน้าต่าง Groups






7. ในที่นี้จะแบ่งข้อมูลอายุออกเป็นกลุ่ม (Bin) โดยการกำหนดให้แต่ละกลุ่มมีขนาดช่วงอายุ 5 ปี ช่วงอายุ โดยเลือก Bin Type เป็น Size of bins และกำหนด Bin size เท่ากับ 5 แล้วคลิก OK กลุ่มของอายุจะถูกสร้างเป็นฟิลด์ใหม่ชื่อ Age (bins)
8. ที่หน้าต่างย่อย Visualization คลิกปุ่ม  Clustered Column Chart จะปรากฏการแสดงผลที่พื้นที่รายงาน ดังรูป



9. ที่หน้าต่าง Visualization คลิกที่ปุ่ม  จะปรากฏการแสดงผลดังรูปด้านล่าง ให้ลาก ฟิลด์ Age (bins) จากหน้าต่าง Fields มาวางที่ช่อง Axis และลาก ฟิลด์ Age มาวางที่ช่อง Value โปรแกรมจะสร้างฮิสโตแกรมโดยนับจำนวนของข้อมูลในฟิลด์ที่ปรากฏในช่อง Value แยกตามข้อมูลในฟิลด์ที่ปรากฏในช่อง Axis มาสร้างเป็นกราฟแต่ละแห่ง ดังรูป




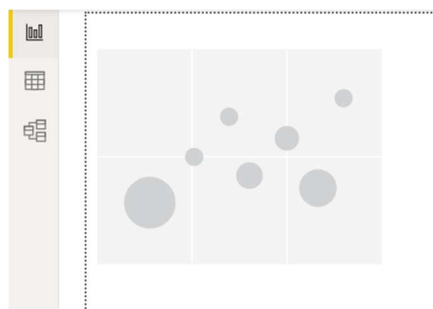
10. สามารถปรับแต่งกราฟ โดยคลิกที่ปุ่ม  จะปรากฏการแสดงผลดังรูปด้านล่าง นักศึกษาสามารถปรับแต่งส่วนประกอบต่างๆ ของกราฟได้จากส่วนนี้
11. การสร้างแผนภูมิแท่งแนวตั้ง (Column Chart) ของข้อมูลจำแนกประเภท ในที่นี้จะสร้างแผนภูมิแท่งแสดงจำนวนผู้ป่วยแต่ละเพศ (Sex) แยกตามสภาพของผู้ป่วย (State) จากตารางข้อมูล COVID-19_patient ที่หน้าต่างย่อย Visualization คลิกปุ่ม  Clustered Column Chart
12. คลิกที่ปุ่ม  จะปรากฏการแสดงผลดังรูปด้านล่าง ให้ลาก ฟิลด์ sex จากหน้าต่าง Fields มาวางที่ช่อง Axis และช่อง Value และลากฟิลด์ state มาวางที่ช่อง Legend โปรแกรมจะสร้างแผนภูมิแท่ง โดยนับจำนวนของข้อมูลในฟิลด์ที่ปรากฏในช่อง Value แยกเป็นกลุ่มตามข้อมูลใน


ฟิลด์ที่ปรากฏในช่อง Axis โดยแต่ละกลุ่มจะมีแท่งกราฟย่อยแบ่งตามข้อมูลในฟิลด์ ปรากฏในช่อง Legend ดังรูป

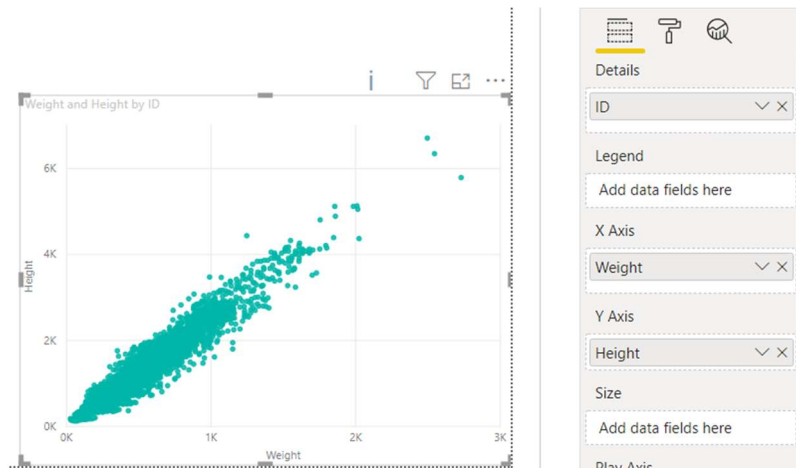



หมายเหตุ ยังมีวิธีการสร้างแผนภูมิแท่งอีกหลายรูปแบบ เช่น แผนภูมิแท่งแนวนอน (Clustered Bar Chart) แผนภูมิแท่งแนวตั้งแบบกองซ้อน (Stacked Column Chart) และแผนภูมิแท่งแนวนอนแบบกองซ้อน (Stacked Bar Chart) ซึ่งสามารถสร้างได้โดยประยุกต์ใช้วิธีการแบบเดียวกับที่กล่าวข้างต้น

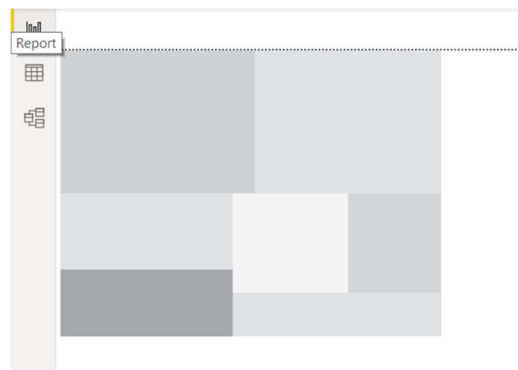
13. การสร้างแผนภาพการกระจาย (Scatter Plot) ของข้อมูลตัวเลข ในที่นี่จะสร้างกราฟการกระจายของข้อมูลน้ำหนัก (Weight) และส่วนสูง (Height) จากตารางข้อมูล athlete_events ให้นักศึกษาสร้างรายงานใหม่ โดยคลิกที่ปุ่ม + (New page) บริเวณแท็บ Page ด้านล่างของหน้าต่าง Report จากนั้นเปลี่ยนชื่อเป็น Athlete Events
14. ไปยังเพจ Athlete Events แล้วคลิกปุ่ม  Scatter Chart จากหน้าต่าง Visualization จะปรากฏการแสดงผลที่พื้นที่รายงาน ดังรูป




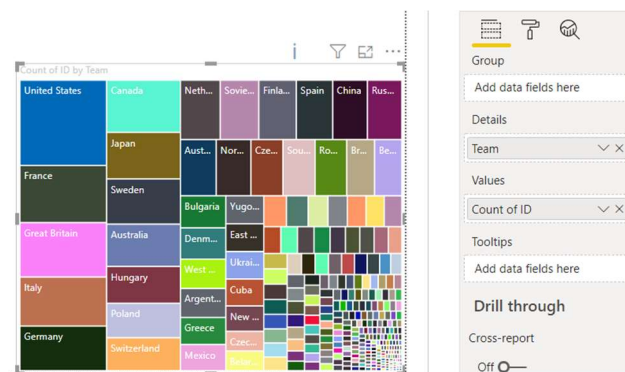
15. คลิกที่ปุ่ม  จะปรากฏการแสดงผลดังรูปด้านล่าง ให้ลาก ฟิลด์ Weight จากหน้าต่าง Fields มาวางที่ช่อง X Axis ลากฟิลด์ Height มาวางที่ช่อง Y Axis และลากฟิลด์ ID มาวางที่ช่อง Details โปรแกรมจะสร้างแผนภาพการกระจายโดยนำข้อมูลแต่ละตัว (ระเบียนข้อมูล/แถว) ที่มีค่าของฟิลด์ที่ปรากฏในช่อง Details ไม่ซ้ำกัน มาวางเป็นจุดบนกราฟโดยแกน x และ y ของกราฟแทนข้อมูลในฟิลด์ที่ปรากฏในช่อง X Axis และ Y Axis ตามลำดับ ดังรูป




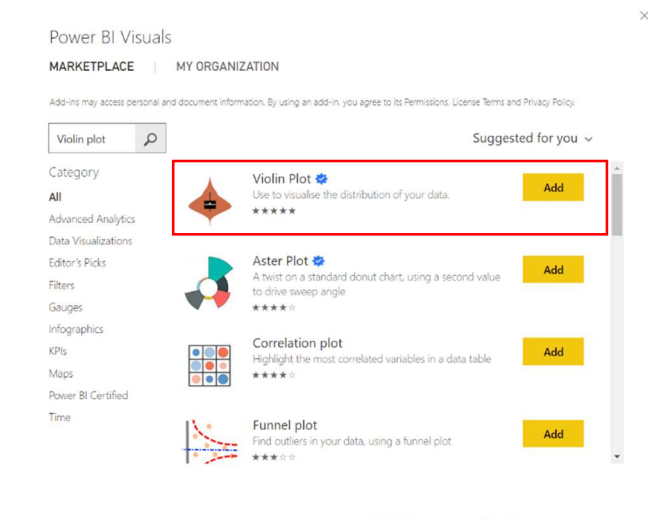
16. การสร้างแผนภูมิต้นไม้ (Tree Map) ของข้อมูลจำแนกประเภท ในที่นี้จะสร้างแผนภูมิต้นไม้ของข้อมูลทีม (Team) จากตารางข้อมูล athlete_events ที่หน้าต่างย่อย Visualization คลิกปุ่ม  Treemap จะปรากฏการแสดงผลที่พื้นที่รายงาน ดังรูป




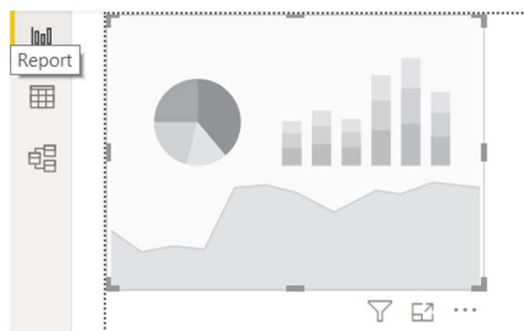
17. คลิกที่ปุ่ม  จะปรากฏการแสดงผลดังรูปด้านล่าง ให้ลาก ฟิลด์ ID จากหน้าต่าง Fields มาวางที่ช่อง Values และลากฟิลด์ Team มาวางที่ช่อง Details โปรแกรมจะสร้างแผนภูมิต้นไม้ โดยแผนภูมิจะแบ่งพื้นที่ออกเป็นสี่เหลี่ยมขนาดเล็ก แต่ละรูปสี่เหลี่ยมแทนข้อมูลจำแนกประเภทตามฟิลด์ที่ปรากฏในช่อง Details ขนาดของสี่เหลี่ยมแสดงสัดส่วนของจำนวนข้อมูลที่ค่าในฟิลด์ที่ปรากฏในช่อง Values ไม่ซ้ำกันจำนวนข้อมูลทั้งหมด ดังรูป




18. การสร้างแผนภูมิรูปไวโอลิน (Violin Plot) เนื่องจากโปรแกรม Power BI ไม่ได้ติดตั้งเครื่องมือสำหรับสร้างแผนภูมิรูปไวโอลินมาพร้อมตั้งแต่ต้น จึงต้องนำเข้าเครื่องมือดังกล่าว โดยที่หน้าต่าง Visualization คลิกที่ปุ่ม  แล้วเลือก Import from AppSource จะปรากฏหน้าต่าง Power BI Visuals ดังรูป

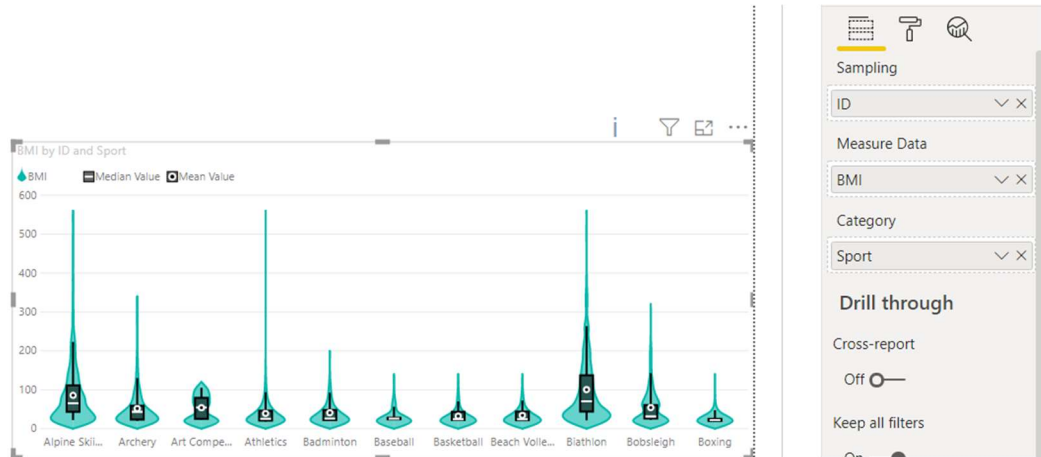



19. ที่ช่อง Search พิมพ์คำค้น Violin plot แล้วคลิก Search จะปรากฏรายการผลการค้นหา ให้เลือก Violin plot โดยคลิก Add หลังชื่อรายการ โปรแกรมจะนำเข้าเครื่องมือนั้นเข้าสู่โปรแกรม Power
20. ในที่นี้จะสร้างแผนภูมิรูปไวโอลิน (Violin Plot) ของข้อมูลดัชนีมวลกาย (BMI) จากตารางข้อมูล athlete_events โดยจะแสดงให้เห็นถึงการกระจายของข้อมูลดัชนีมวลกายของนักกีฬาในแต่ละชนิดกีฬา (Sport) ที่หน้าต่างย่อย Visualization คลิกปุ่ม  Violin Plot จะปรากฏการแสดงผลที่พื้นที่รายงาน ดังรูป

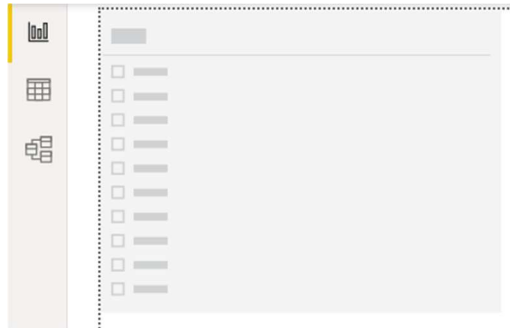



21. คลิกที่ปุ่ม  จะปรากฏการแสดงผลดังรูปด้านล่าง ให้ลากฟิลด์ ID จากหน้าต่าง Fields มาวางที่ช่อง Sampling ลากฟิลด์ BMI มาวางที่ช่อง Measure Data และลากฟิลด์ Sport มาวางที่ช่อง Category โปรแกรมจะสร้างแผนภูมิรูปไวโอลินแสดงลักษณะการกระจายของข้อมูลใน

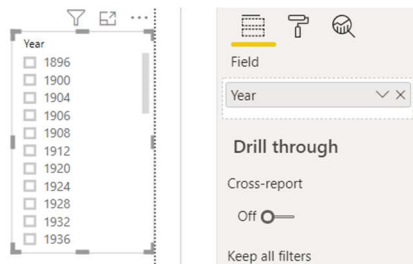
ฟิลด์ที่ปรากฏในช่อง Measure Data โดยถือเอาข้อมูลในฟิลด์ที่ปรากฏในช่อง Sampling ที่ไม่ซ้ำกันเป็นกลุ่มตัวอย่าง และแบ่งแยกข้อมูลออกเป็นกลุ่มๆ ตามข้อมูลในฟิลด์ที่ปรากฏในช่อง Category



22. จากการสร้างภาพแสดงข้อมูลบนตารางข้อมูล athlete_events ในข้างต้น จะเห็นว่าเป็นการนำข้อมูลนักกีฬาโอลิมปิก ทุกๆ ปีมาแสดง นักศึกษาสามารถสร้างเครื่องมือที่ช่วยกรองข้อมูลมาเฉพาะปีที่ต้องการได้ โดยที่หน้าต่าง Visualization คลิกปุ่ม  Slicer จะปรากฏการแสดงผลที่พื้นที่รายงาน ดังรูป



23. คลิกที่ปุ่ม  จะปรากฏการแสดงผลดังรูปด้านล่าง ให้ลาก ฟิลด์ Year จากหน้าต่าง Fields มาวางที่ช่อง Values โปรแกรมจะสร้างรายการช่องทำเครื่องหมาย (Checkbox) ของปีที่มีการจัดการแข่งขันโอลิมปิก ดังรูป เมื่อนักศึกษาคลิกทำเครื่องหมายหน้าปีที่ต้องการ แผนภาพอื่นๆ จะเปลี่ยนแปลงการแสดงผล โดยนำข้อมูลเฉพาะของปีนั้นมาสร้างแผนภาพ



หมายเหตุ ยังมีวิธีการนำเสนอข้อมูลด้วยการมองภาพข้อมูลอีกมากมาย และแต่ละวิธียังมีรายละเอียดทางเทคนิคในการสร้างแผนภาพให้น่าสนใจอีกมาก นักศึกษาสามารถศึกษาเพิ่มเติมได้จาก

https://www.youtube.com/watch?v=bzF5kN2BO60&list=PLzN99cpDw6oA_08IRfOLohNWtwO3AjyJn

3. แบบฝึกปฏิบัติการ

ให้นักศึกษาทำตามขั้นตอนปฏิบัติการ และบันทึกงาน โดยตั้งชื่อไฟล์เป็น Lab_11_id.pbix และส่งเป็นการบ้าน

สิ่งที่ต้องส่งเป็นการบ้าน เพิ่มข้อมูล Power BI ที่บันทึกงานของนักศึกษา โดยตั้งชื่อไฟล์ในรูปแบบ Lab_11_id.pbix โดยแทน id ด้วยรหัสนักศึกษา ส่งผ่านเว็บไซต์ <http://hw.cs.science.cmu.ac.th>