

ปฏิบัติการที่ 10

การนำเข้าข้อมูลและการเตรียมข้อมูลสำหรับการแสดงภาพข้อมูลด้วยเครื่องมือ Business Intelligence

วัตถุประสงค์

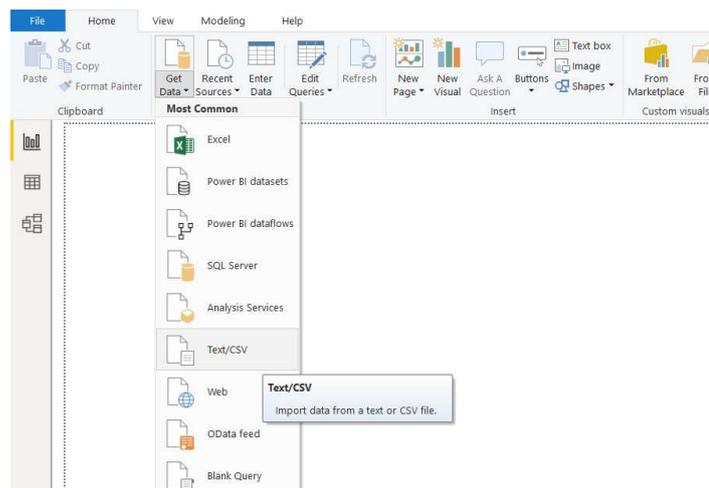
1. เพื่อให้สามารถนำข้อมูลเข้าสู่เครื่องมือ Business Intelligence ได้
2. เพื่อให้สามารถจัดเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการนำไปแสดงภาพข้อมูลด้วย เครื่องมือ Business Intelligence ได้

1. ชุดข้อมูลปฏิบัติการ

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

2. ขั้นตอนปฏิบัติการ

1. เปิดโปรแกรม Microsoft Power BI Desktop โดยเข้าใช้งานโปรแกรมด้วยบัญชีอีเมลนักศึกษา
2. ทำการนำเข้าข้อมูล โดยไปที่แท็บ HOME คลิก Get Data จากนั้นเลือกชนิดเพิ่มข้อมูลที่ต้องการ ในที่นี้เลือกชนิดเพิ่มข้อมูล Text/CSV



3. เลือกเพิ่มข้อมูล COVID-19_patient.csv จะปรากฏหน้าต่างดังรูป ทำการตรวจสอบข้อมูลแล้วคลิก Load ข้อมูลจะถูกนำเข้าสู่อุปกรณ์ Power BI

COVID-19_patient.csv

File Origin: 1252: Western European (Windows) | Delimiter: Comma | Data Type Detection: Based on first 200 rows

patient_id	sex	birth_year	country	region	disease	group	infection_reason	infection_order	infected_by
1	F	1984	China	filtered at airport	null		visit to Wuhan		1
2	M	1964	Korea	filtered at airport	null		visit to Wuhan		1
3	M	1966	Korea	capital area	null		visit to Wuhan		1
4	M	1964	Korea	capital area	null		visit to Wuhan		1
5	M	1987	Korea	capital area	null		visit to Wuhan		1
6	M	1964	Korea	capital area	null		contact with patient		2
7	M	1991	Korea	capital area	null		visit to Wuhan		1
8	F	1957	Korea	Jeollabuk-do	null		visit to Wuhan		1
9	F	1992	Korea	capital area	null		contact with patient		2
10	F	1966	Korea	capital area	null		contact with patient		3
11	M	1995	Korea	capital area	null		contact with patient		3
12	M	1971	China	capital area	null		contact with patient in Japan		2
13	M	1992	Korea	filtered at airport	null		residence in Wuhan		1
14	F	1980	China	capital area	null		contact with patient		3
15	M	1977	Korea	capital area	null		contact with patient		2
16	F	1977	Korea	Gwangju	null		visit to Thailand		1
17	M	1982	Korea	capital area	null		contact with patient in Singapore		2
18	F	1999	Korea	Gwangju	null		contact with patient		2
19	M	1983	Korea	capital area	null		contact with patient in Singapore		2
20	F	1978	Korea	capital area	null		contact with patient		3

Buttons: Load, Edit, Cancel

4. จากนั้นไปที่หน้าต่าง Data โดยคลิกที่  ทางด้านซ้ายของหน้าต่างโปรแกรม จะปรากฏตารางข้อมูล ดังรูป หน้าต่างนี้เป็นพื้นที่สำหรับจัดการข้อมูลที่ใช้นำเข้าสู่โปรแกรม จากหน้าต่าง Data จะมีแผงหน้าต่างย่อย Fields ทางด้านขวาของหน้าต่างโปรแกรมสำหรับเลือกตารางข้อมูลที่ต้องการและจัดการเขตข้อมูล (Field)

Untitled - Power BI Desktop

Clipboard: Paste, Copy, Format Painter

External data: Get Data, Recent Sources, Enter Data, Edit Queries, Refresh

Insert: New Page, New Visual, Ask A Question, Buttons, Text box, Image, Shapes

Custom visuals: From Marketplace, From File

Themes: Switch Theme

Relationships: Manage Relationships

Calculations: New Measure, New Column, New Quick Measure

Share: Publish

Fields

Search

COVID-19_patient

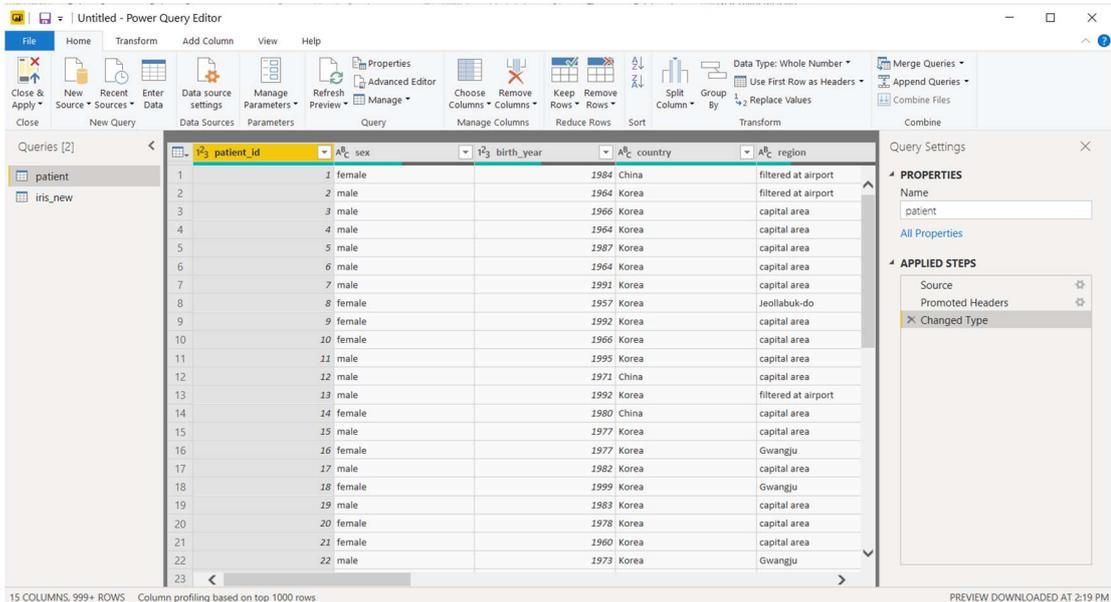
- birth_year
- confirmed_date
- contact_number
- country
- deceased_date
- disease
- group
- infected_by
- infection_order
- infection_reason
- patient_id
- region
- released_date

patient_id	sex	birth_year	country	region	disease	group	infection_reason	infection_order	infected_by
3151			Korea						
3152			Korea						
3153			Korea						
3154			Korea						
3155			Korea						
3156			Korea						
3157			Korea						
3158			Korea						
3159			Korea						
3160			Korea						
3161			Korea						
3162			Korea						
3163			Korea						
3164			Korea						
3165			Korea						
3166			Korea						
3167			Korea						
3168			Korea						

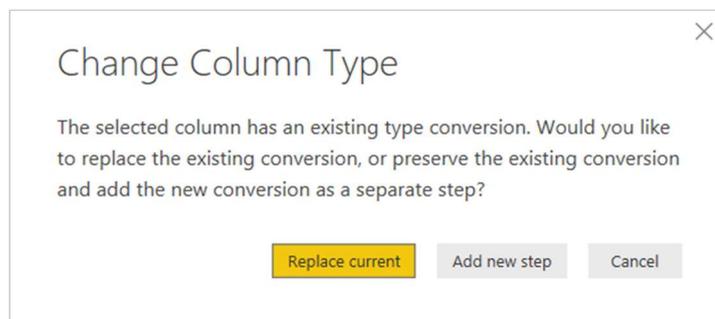
TABLE: COVID-19_patient (7,869 rows)

UPDATE AVAILABLE (CLICK TO DOWNLOAD)

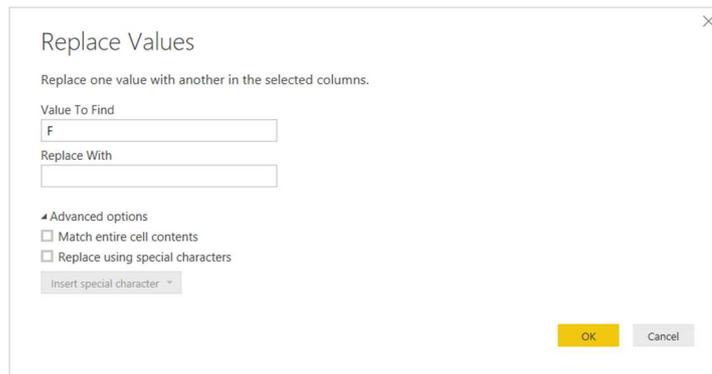
5. ต่อมาคลิก Edit Queries ที่แท็บ Home เพื่อแก้ไขข้อมูล จะปรากฏหน้าต่าง Power Query Editor ดังรูป



6. ในที่นี่จะเปลี่ยนชนิดข้อมูลของคอลัมน์ patient_id จากชนิด Whole Number เป็นชนิด Text โดยคลิกเลือก คอลัมน์ patient_id จากนั้นไปที่ส่วน Transform ในแท็บ Home คลิกที่ Data Type แล้วเลือก Text จะปรากฏหน้าต่างไดอะล็อก ดังรูป จากนั้นเลือก Replace current



7. ต่อมาเปลี่ยนข้อมูลของคอลัมน์ sex ที่ถูกเข้ารหัส โดยเปลี่ยนข้อมูล F เป็น female และ M เป็น male โดยคลิกเลือก คอลัมน์ sex จากนั้นไปที่ส่วน Transform ในแท็บ Home คลิกที่ Replace values จะปรากฏหน้าต่างไดอะล็อก ดังรูป



8. ทำการเปลี่ยนข้อมูลรหัส F เป็น female โดยในกล่องข้อความ Value to Find พิมพ์ F และกล่องข้อความ Replace With พิมพ์ female และทำเครื่องหมาย ✓ หน้าตัวเลือก Match entire cell contents แล้วคลิก OK จะพบว่าข้อมูลในคอลัมน์ sex มีการเปลี่ยนแปลงจากค่า F เป็น female
9. ทำการเปลี่ยนข้อมูลรหัส M เป็น male โดยใช้วิธีการเดียวกับข้อ 8
10. เมื่อทำการแก้ไขข้อมูลเสร็จแล้วคลิก Close & Apply ที่แท็บ Home เพื่อนำการเปลี่ยนแปลงดังกล่าวไปใช้กับข้อมูลในโปรแกรม Power BI และปิดหน้าต่าง Power Query Editor จะสังเกตพบว่าข้อมูลในตารางข้อมูลในโปรแกรม Power BI มีการเปลี่ยนแปลง
11. ต่อมาทำการสร้างคอลัมน์ใหม่สำหรับข้อมูลอายุ โดยจะใช้ข้อมูลจากคอลัมน์ birth_year เพื่อการคำนวณอายุ
12. คลิกที่ New column ในแท็บ Home จากนั้นสังเกตที่แถบสูตร (Formula bar) จะมีลักษณะดังรูป



13. ที่แถบสูตร ให้พิมพ์สูตรการคำนวณ ดังนี้
Age = YEAR(TODAY()) - 'COVID-19_patient'[birth_year]
14. จากนั้นกด ENTER จะพบว่ามีการเพิ่มคอลัมน์ชื่อ Age ปรากฏขึ้นในตารางข้อมูล พร้อมข้อมูลอายุซึ่งถูกคำนวณจากสูตรที่ให้
15. ทำการบันทึกงาน โดยคลิกที่แท็บ File จากนั้นคลิก Save ตั้งชื่อแฟ้มข้อมูลที่จะบันทึกเป็น Practice_10.pbix

3. แบบฝึกปฏิบัติการ

ให้นักศึกษาทำแบบฝึกปฏิบัติการ ตามลำดับขั้นตอนต่อไปนี้

1. ให้นักศึกษานำชุดข้อมูล 120 Years of Olympic History athletes and Results จากเพิ่มข้อมูล athlete_events.csv เข้าสู่โปรแกรม Microsoft Power BI
2. ทำการแทนที่ข้อมูลในเซลล์ข้อมูลของทุกคอลัมน์ที่มีค่า NA ด้วยค่า *null* เนื่องจาก Microsoft Power BI กำหนดให้ข้อมูลใดที่เป็นข้อมูลสูญหาย ต้องมีค่า *null*
3. แก้ไขชนิดข้อมูลของคอลัมน์ ID ให้เป็นชนิดข้อความ (Text)
4. แก้ไขชนิดข้อมูลของคอลัมน์ Age Height และ Weight ให้เป็นชนิดตัวเลข (Whole Number)
5. ทำการแปลงข้อมูลในคอลัมน์ Sex ที่ถูกเข้ารหัส โดยเปลี่ยนข้อมูล F เป็น Female และ M เป็น Male
6. สร้างคอลัมน์ใหม่สำหรับข้อมูล BMI โดย ค่า BMI คำนวณได้จากสูตร $BMI = \text{Weight(kg)}/[\text{Height(m)}]^2$
7. บันทึกงาน โดยตั้งชื่อไฟล์เป็น Lab_10_id.pbix และส่งเป็นการบ้าน

สิ่งที่ต้องส่งเป็นการบ้าน เพิ่มข้อมูล Power BI ที่บันทึกงานของนักศึกษา โดยตั้งชื่อไฟล์ในรูปแบบ Lab_10_id.pbix โดยแทน id ด้วยรหัสนักศึกษา ส่งผ่านเว็บไซต์ <http://hw.cs.science.cmu.ac.th>