

DATA SCIENCE & INFORMATION TECHNOLOGY

KNOWLEDGE MANAGEMENT

การเตรียมข้อมูลด้วย Power Query

อ.นริศรา นาคเมธี

อาจารย์ผู้รับผิดชอบหลักสูตร
วท.บ.สาขาวิชาวิทยาการข้อมูลฯ
คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

Power Query คือ เครื่องมือช่วยในการแปลงข้อมูลและเตรียมข้อมูล ที่เป็น Graphic user interface ช่วยให้ผู้ใช้สามารถรับข้อมูลจากแหล่งที่มาข้อมูลได้หลากหลายแหล่ง และสามารถดำเนินการแยก แปลง และโหลดข้อมูล (ETL) เพื่อให้ข้อมูลพร้อมนำไปประมวลต่อได้ง่าย เนื่องจาก Power Query มีอยู่ในหลายผลิตภัณฑ์ของไมโครซอฟต์ เช่น Excel, Power BI

สำหรับในบทความนี้ ผู้เขียนขออธิบายการใช้งาน Power Query เพื่อนำข้อมูลเข้า Power BI แก้ไขชนิดข้อมูล เพิ่ม Column และแก้ไข format ข้อมูล

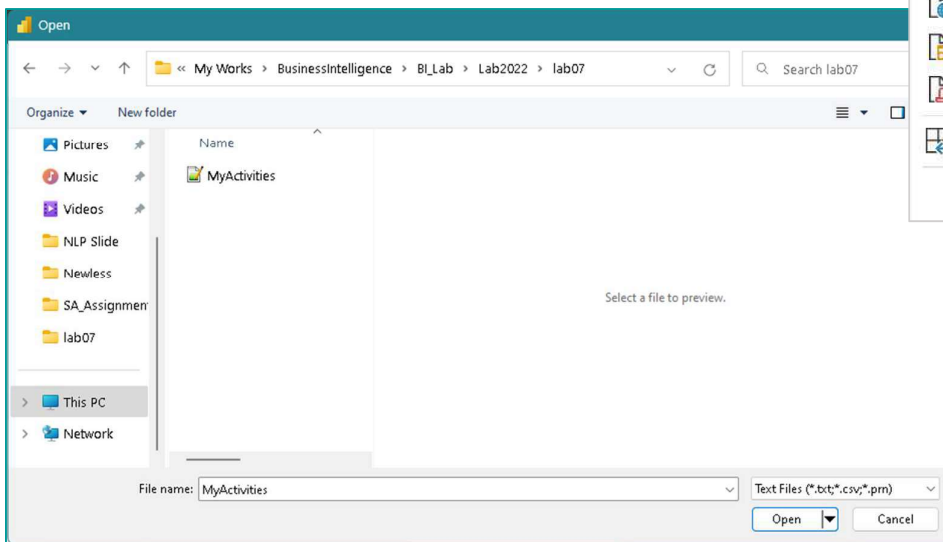
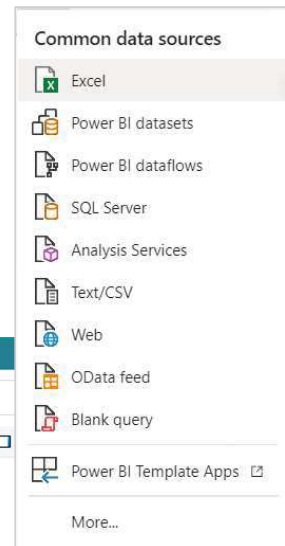
รายการ Software และไฟล์ข้อมูลที่ต้องใช้ในการฝึก

1. Microsoft Power BI Desktop
2. MyActivities.csv (สามารถ download ได้ที่ <https://shorturl.at/jkqIZ>)

เริ่มต้นใช้ Power Query โดยการเปิด Microsoft Power BI

1. การนำข้อมูลเข้า Microsoft Power BI

- 1.1 ไปที่แถบเมนู data แล้วเลือก Get data
- 1.2 จะปรากฏ เมนูย่อย Common data sources ขึ้นมาเพื่อให้เลือกประเภทของข้อมูลที่จะนำเข้า
- 1.3 เลือกประเภทของข้อมูลที่ต้องการนำเข้า จะปรากฏหน้าจอ ดังภาพที่ 1



ภาพที่ 1 หน้าต่างให้เลือกไฟล์ที่จะนำเข้าสู Power BI

- 1.4 เลือกไฟล์ข้อมูลที่ต้องการนำเข้า power BI จากนั้น เลือก Open

1.5 จะปรากฏหน้าจอ ดังภาพที่ 2

MyActivities.csv

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

Activity Type	Date	Distance	Calories	Time	Avg HR	Max HR	Avg Run Cadence	Max Run Cadence	Avg Speed
Cycling	1/7/2023 8:05:00 AM	1.6	72	0:15:55	113	140	0	0	6
Walking	1/7/2023 6:57:00 AM	1.83	85	0:26:39	95	106	85	228	14:34
Cycling	1/7/2023 6:39:00 AM	3.35	80	0:16:09	119	132	0	0	12.4
Walking	1/5/2023 6:25:00 AM	4.11	194	0:51:54	113	145	130	201	12:38
Walking	1/2/2023 6:57:00 AM	1.87	132	0:34:47	113	150	97	215	18:39
Cycling	12/25/2022 7:46:00 AM	3.21	73	0:33:10	85	115	0	0	5.8
Cycling	12/25/2022 6:32:00 AM	4.07	94	0:34:26	94	110	0	0	7.1
Walking	12/18/2022 6:34:00 AM	3.24	149	0:43:00	98	126	102	239	13:17
Walking	12/12/2022 6:37:00 AM	2.96	148	0:53:12	91	112	82	136	17:58
Walking	12/6/2022 5:22:00 PM	4.04	172	0:52:17	103	134	96	140	12:58
Running	12/3/2022 5:58:00 AM	6.01	301	1:14:08	126	154	128	185	12:20
Walking	12/2/2022 6:09:00 AM	3.06	153	0:41:54	109	143	114	179	13:40
Walking	12/1/2022 6:15:00 AM	5.22	267	1:07:09	118	154	112	241	12:51
Walking	11/29/2022 5:23:00 PM	2.37	147	0:55:59	91	121	46	141	13:54
Walking	11/27/2022 6:03:00 AM	6.62	336	1:19:20	125	160	114	167	12:12
Walking	11/19/2022 6:00:00 AM	7.15	389	1:18:37	136	167	126	185	12:20
Walking	11/18/2022 6:16:00 AM	3.14	185	0:47:44	114	141	98	140	12:12
Walking	11/17/2022 6:02:00 AM	5.08	317	1:02:00	137	166	120	202	12:12
Walking	11/16/2022 6:39:00 AM	3	177	0:36:10	130	152	129	191	12:03
Walking	11/12/2022 6:41:00 AM	4.21	231	0:55:53	117	161	120	179	13:17

Buttons: Extract Table Using Examples, Load, Transform Data, Cancel

ภาพที่ 2 หน้าต่างให้เลือกไฟล์ที่จะนำเข้าสู่ Power BI

1.6 กดปุ่ม Transform data จากนั้นจะปรากฏหน้าจอ Power Query ดังภาพที่ 3

Power Query Editor

Query: MyActivities














Formula Bar: = Table.TransformColumnTypes(#Promoted Headers,{{"Activity Type", type text}, {"Date", type datetime}, {"Time", type datetime}, {"Distance", type number}, {"Calories", Int64.Type}, {"Time", type text}, {"Avg HR", type number}, {"Max HR", type number}, {"Avg Run Cadence", type number}, {"Max Run Cadence", type number}, {"Avg Speed", type number}}})

Activity Type	Date	Distance	Calories	Time	Avg HR	Max HR	Avg Run Cadence	Max Run Cadence	Avg Speed
1	Cycling	1/7/2023 8:05:00 AM	1.6	72	0:15:55	113	140	0	6
2	Walking	1/7/2023 6:57:00 AM	1.83	85	0:26:39	95	106	85	228
3	Cycling	1/7/2023 6:39:00 AM	3.35	80	0:16:09	119	132	0	12.4
4	Walking	1/5/2023 6:25:00 AM	4.11	194	0:51:54	113	145	130	201
5	Walking	1/2/2023 6:57:00 AM	1.87	132	0:34:47	113	150	97	215
6	Cycling	12/25/2022 7:46:00 AM	3.21	73	0:33:10	85	115	0	5.8
7	Cycling	12/25/2022 6:32:00 AM	4.07	94	0:34:26	94	110	0	7.1
8	Walking	12/18/2022 6:34:00 AM	3.24	149	0:43:00	98	126	102	239
9	Walking	12/12/2022 6:37:00 AM	2.96	148	0:53:12	91	112	82	136
10	Walking	12/6/2022 5:22:00 PM	4.04	172	0:52:17	103	134	96	140
11	Running	12/3/2022 5:58:00 AM	6.01	301	1:14:08	126	154	128	185
12	Walking	12/2/2022 6:09:00 AM	3.06	153	0:41:54	109	143	114	179
13	Walking	12/1/2022 6:15:00 AM	5.22	267	1:07:09	118	154	112	241
14	Walking	11/29/2022 5:23:00 PM	2.37	147	0:55:59	91	121	46	141
15	Walking	11/27/2022 6:03:00 AM	6.62	336	1:19:20	125	160	114	167
16	Walking	11/19/2022 6:00:00 AM	7.15	389	1:18:37	136	167	126	185
17	Walking	11/18/2022 6:16:00 AM	3.14	185	0:47:44	114	141	98	140
18	Walking	11/17/2022 6:02:00 AM	5.08	317	1:02:00	137	166	120	202
19	Walking	11/16/2022 6:39:00 AM	3	177	0:36:10	130	152	129	191
20	Walking	11/12/2022 6:41:00 AM	4.21	231	0:55:53	117	161	120	179

11 COLUMNS, 140 ROWS | Column profiling based on top 1000 rows

ภาพที่ 3 หน้าต่าง Power Query ที่ปรากฏขึ้น หลังเลือก Transform ข้อมูล

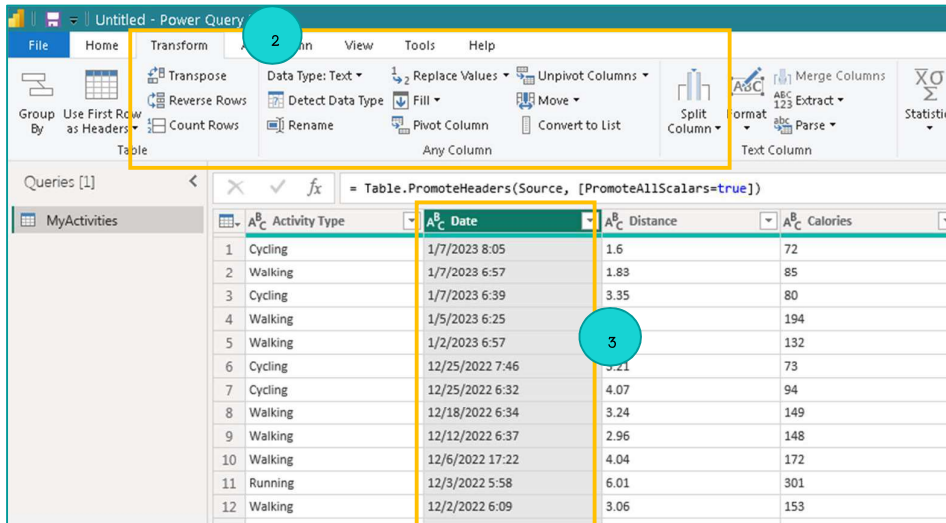
ชนิดข้อมูลใน Power BI

Data type	Icon	Description
Text		ข้อมูลที่เป็นตัวอักษร ข้อความ ตัวเลขและวันที่ที่นำเสนอในรูปแบบของ text format ความยาวสูงสุดของข้อความ คือ 268,435,456 Unicode characters หรือ 536,870,912 bytes.
True/False		ค่า Boolean (True/False)
Decimal number		ข้อมูลตัวเลขจำนวนทศนิยม 64-bit (eight-byte) ซึ่งสามารถเก็บตัวเลขได้ 15 หลัก
Fixed decimal number		ข้อมูลตัวเลขจำนวนทศนิยม ที่เป็น Currency type ที่จะมี , ขึ้นระหว่างหลักร้อยกับพัน หลักล้านกับหลักแสน ค่าตัวเลขที่เก็บได้คือ 922,337,203,685,477.5807 (positive or negative)
Whole number		ตัวเลขจำนวนเต็มขนาด 64-bit (eight-byte) เก็บตัวเลขได้สูงสุด 19 digits โดยมีค่าอยู่ในช่วง $-9,223,372,036,854,775,807$ ($-2^{63}+1$) ถึง $9,223,372,036,854,775,806$ ($2^{63}-2$)
Percentage		ตัวจำนวนทศนิยม ที่มี format ต้องมี % เนื่องจากเป็น percentage
Date/Time		ข้อมูลวันและเวลา โดยที่แสดงทั้งวันและเวลา โดยค่าของ Date/Time จะถูกเก็บในลักษณะ Decimal Number type ดังนั้นเราสามารถ convert ค่าได้ การเก็บเวลาจะถูกเก็บเป็นตัวเลขเศษส่วน 1/300 seconds (3.33 ms) และ Dates จะมีปี ค.ศ. ระหว่าง 1900 - 9999
Date		ข้อมูลวันและเวลา โดยที่แสดงแค่ทั้งวันที่ เดือน ปี
Time		ข้อมูลวันและเวลา โดยที่แสดงเพียงเวลา
Date/Time/Timezone		จะเก็บ UTC Date/Time พร้อมกับ time-zone
Duration		นำเสนอระยะเวลา โดยอยู่ในรูปแบบของจุดทศนิยม
Binary		ข้อมูล Binary data ถูกเก็บในรูปแบบ binary format.
Any		ข้อมูลประเภทอื่น ที่ไม่ทราบสามารถบอกชนิดได้อย่างชัดเจน



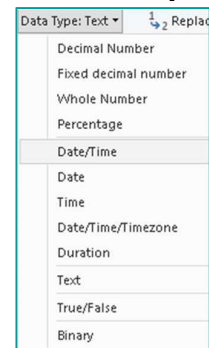
2. การเปลี่ยนชนิดข้อมูล

2.1 ไปที่เมนู Transform แล้วไปที่ Data Type

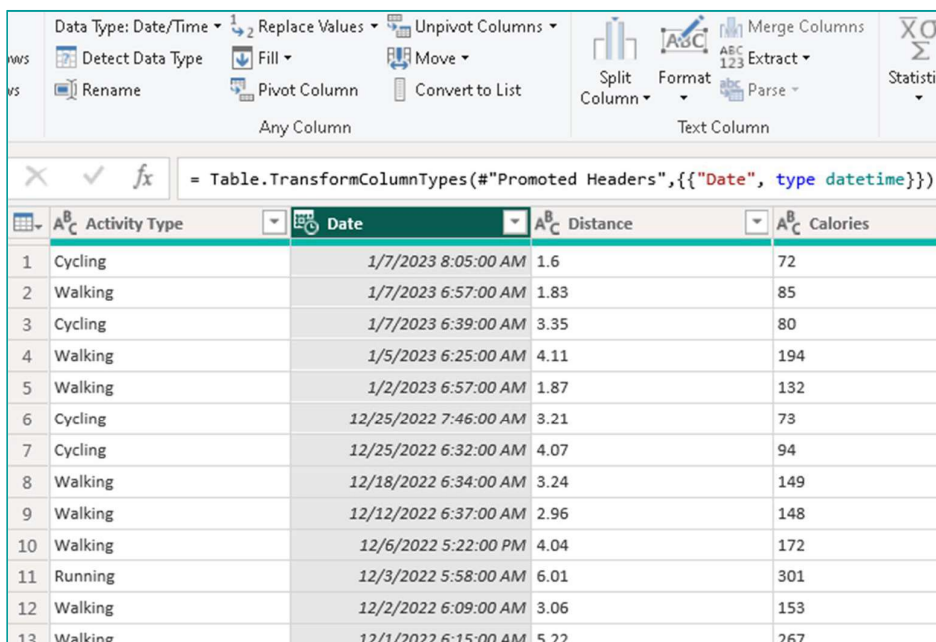


ภาพที่ 4 หน้าต่าง Power Query ที่ปรากฏขึ้น หลังเลือก Transform ข้อมูล

2.2 Click เลือก Data type จะปรากฏ menu list ขึ้นมา ดังภาพ



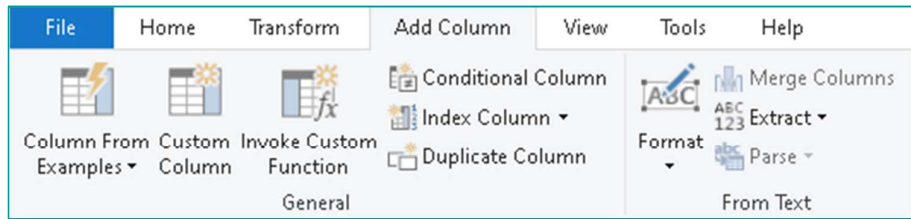
2.3 Click เลือก Date/Time ชนิดข้อมูลใน Column Date จะเปลี่ยนเป็นชนิด Date/Time



ภาพที่ 5 แสดงชนิดข้อมูลใน Column Date เปลี่ยนจาก Text เป็น Date/Time

3. การเพิ่ม Column ใหม่ ในกรณีที่เรต้องการข้อมูลเพื่อการวิเคราะห์เพิ่มเติม

3.1 เลือกเมนู Add Column



ภาพที่ 6 แสดงเมนู Add Column และ Submenu

3.2 เลือกรูป Column ที่ต้องการสร้าง

- Column From Examples คือ การสร้าง column ใหม่ พร้อมข้อมูลโดยดูจากข้อมูลตัวอย่าง
- Custom Column คือ การสร้าง column ใหม่ โดยที่เราสามารถกำหนดข้อมูลได้ตามต้องการจากสูตร
- Condition Column คือ การสร้าง column ใหม่ พร้อมข้อมูลตามเงื่อนไขที่เรากำหนด
- Index Column คือ การ Column ใหม่เพื่อเป็นดัชนี
- Duplicate Column คือ สำเนา Column ที่เรต้องการซึ่งเป็น Column ที่มีอยู่แล้วในตาราง มาเพิ่มอีก 1 Column โดย Column ที่เป็นสำเนาจะมีชนิดข้อมูลและข้อมูลเหมือนกับต้นฉบับ

3.2.1 การสร้าง Column From Examples

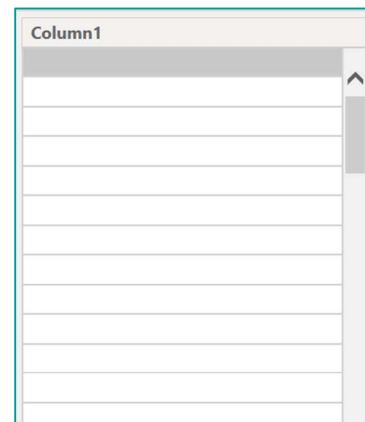
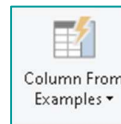
a) เลือกเมนู Add Column

b) เลือก Column From Examples

จะปรากฏ Column1 เปล่าๆ ขึ้นมา ดังภาพ

c) พิมพ์ตัวอย่างข้อมูลที่ยากให้สร้างขึ้นอัตโนมัติ

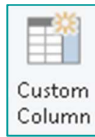
ลงที่แถวแรกของ Column1 ดังภาพ



Merged
Morning Cycling
Morning Walking
Morning Cycling
Morning Walking
Morning Walking
Morning Cycling
Morning Cycling
Morning Walking
Morning Walking
Morning Walking
Evening Walking
Morning Running
Morning Walking
Morning Walking
Evening Walking

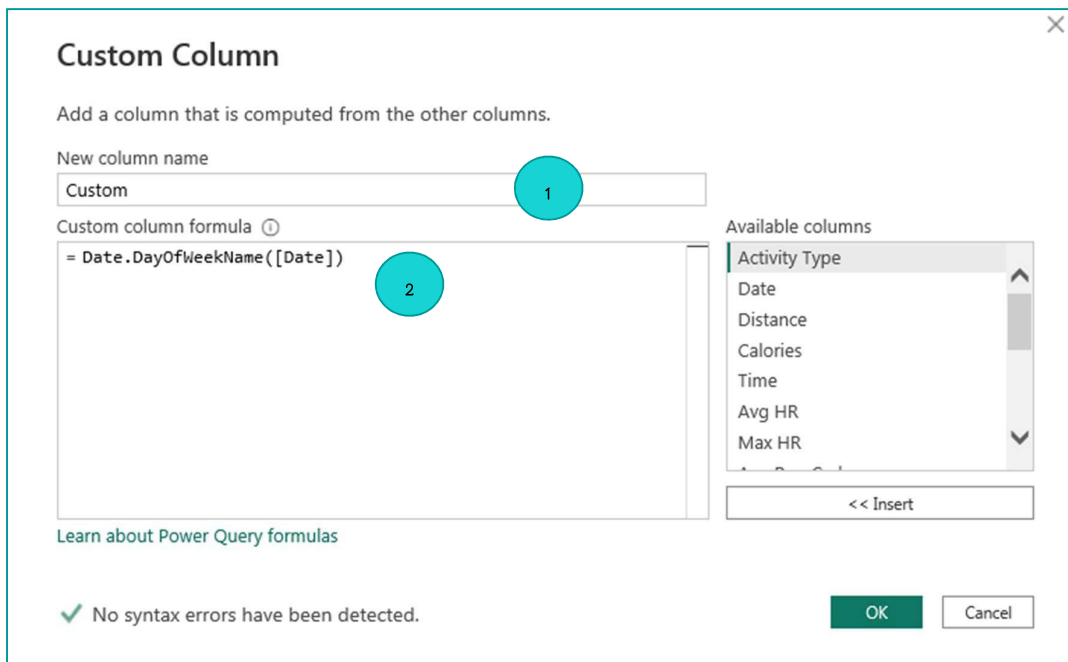
หมายเหตุ:

ตัวอย่างข้อมูลที่ยากให้สร้าง
ขึ้นอัตโนมัติ ควรเป็นข้อมูลที่เคยมีปรากฏอยู่แล้วใน Column อื่น



3.2.2 การสร้าง Custom Column

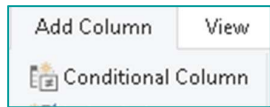
- เลือกเมนู Add Column
- เลือก Custom Column
- ระบุชื่อ Column ใหม่ที่ต้องการสร้างที่ช่องหมายเลข 1
- เขียน Function ลงใน Custom column formula ที่ช่องหมายเลข 2 เช่น ต้องการให้แสดงชื่อวัน



ภาพที่ 7 แสดงหน้าต่างการสร้าง Column ใหม่เพื่อแสดงข้อมูลตามต้องการ

Function ที่ทำงานเกี่ยวกับวันที่และเวลา

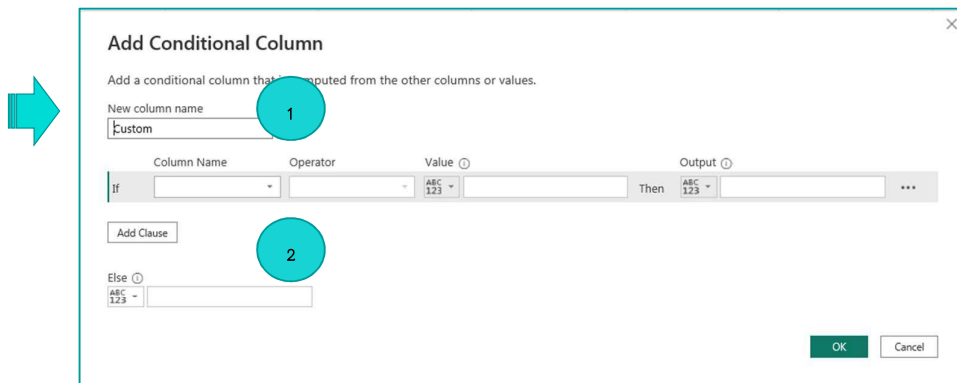
- แสดงวันที่อย่างเดียว Date.Day(ระบุข้อมูลวันเดือนปี) เช่น Date.Day("1/15/2023") จะได้วันที่ 15
- แสดงเดือนอย่างเดียว Date.Month(ระบุข้อมูลวันเดือนปี) เช่น Date.Month("1/15/2023") จะได้วันที่ 1
- แสดงปี Date.Year(ระบุข้อมูลวันเดือนปี) เช่น Date.Year("1/15/2023") จะได้ปี ค.ศ. 2023
- แสดงชื่อวันในสัปดาห์ Date.DayOfWeekName(ระบุข้อมูลวันเดือนปี) เช่น Date.DayOfWeekName("1/15/2023") จะได้ Sunday
- แสดงชื่อเดือน Date.MonthName(ระบุข้อมูลวันเดือนปี) เช่น Date.MonthName("1/15/2023") จะได้ January



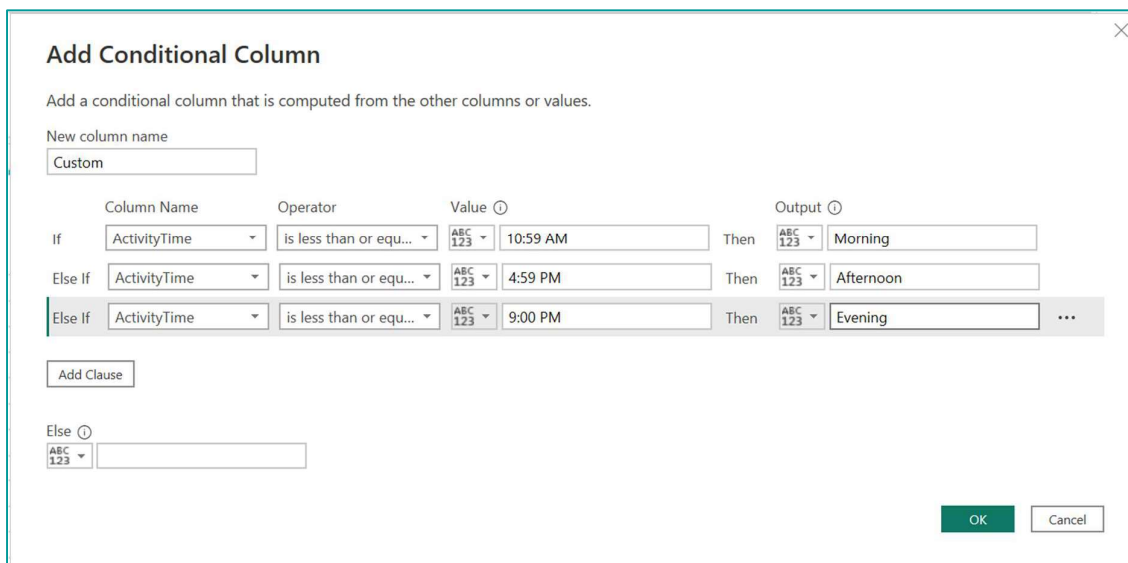
3.2.3 การสร้าง Conditional Column

- a) เลือกเมนู Add Column
- b) เลือก Conditional Column จากนั้นจะปรากฏหน้าจอ

ดังภาพ



- d) ตั้งชื่อ Column ที่หมายเลข 1
- e) กำหนดเงื่อนไขที่ต้องการสร้างข้อมูล ที่หมายเลข 2 โดยระบุชื่อ Column ที่จะใช้ข้อมูลใน Column นั้นมาเป็นข้อมูลตั้งต้นที่นำได้ตรวจสอบว่าตรงกับเงื่อนไขใหม่ ระบุ Operation ระบุ Value (ค่าที่ใช้เป็นเงื่อนไขตรวจสอบ) และระบุ Output ที่จะสร้างถ้าเงื่อนไขเป็นจริง ดังภาพที่ 8

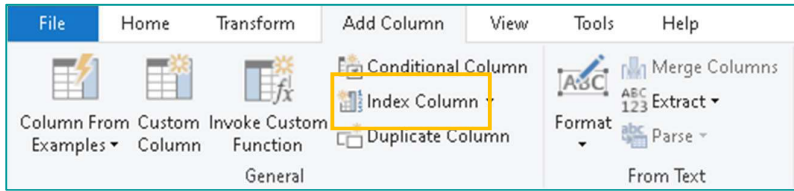


ภาพที่ 8 แสดงการระบุเงื่อนไขให้กับ Conditional Column ที่ต้องการสร้างขึ้นมา



3.2.4 การสร้าง Index Column

a) เลือกเมนู Add Column



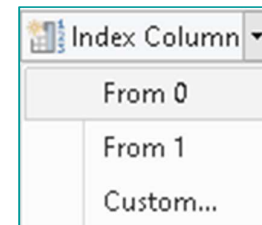
b) กดเลือก Index Column จากนั้นจะปรากฏ dropdowns list ดังภาพ



c) เลือกตัวเลขเริ่มต้นของ index ที่จะสร้างว่าจะเริ่มจาก 0

หรือ 1 หรือ เลขอื่นตามต้องการ

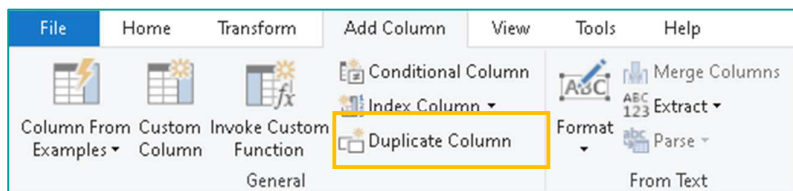
จากนั้นจะปรากฏ Column ใหม่ชื่อ Index ดังภาพ



1 ² ₃ Index
1
2
3
4
5
6
7

3.2.5 การสร้าง Duplicate Column

a) เลือกเมนู Add Column

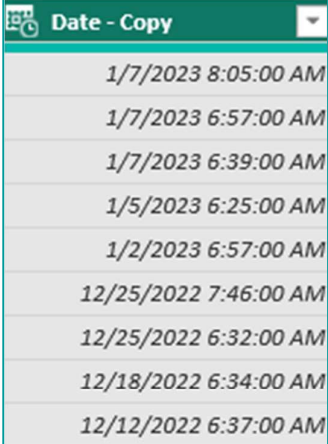


b) กดเลือก Column ต้นฉบับที่ต้องการสำเนา



Date
1/7/2023 8:05:00 AM
1/7/2023 6:57:00 AM
1/7/2023 6:39:00 AM
1/5/2023 6:25:00 AM
1/2/2023 6:57:00 AM
12/25/2022 7:46:00 AM
12/25/2022 6:32:00 AM
12/18/2022 6:34:00 AM
12/12/2022 6:37:00 AM

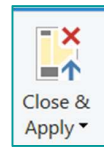
c) เลือก Duplicate Column



Date - Copy
1/7/2023 8:05:00 AM
1/7/2023 6:57:00 AM
1/7/2023 6:39:00 AM
1/5/2023 6:25:00 AM
1/2/2023 6:57:00 AM
12/25/2022 7:46:00 AM
12/25/2022 6:32:00 AM
12/18/2022 6:34:00 AM
12/12/2022 6:37:00 AM



4. บันทึกข้อมูลที่ Transform ให้เรียบร้อย โดย



การกด Close & Apply

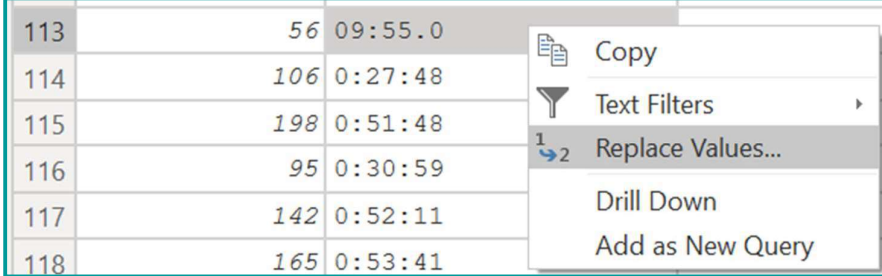
5. แก้ไขข้อมูลที่มี Format รูปแบบผิด ด้วย Power Query

5.2 ไปที่ Data tab จากนั้นไปที่ เมนู Home แล้วกดปุ่ม Transform data

5.3 เมื่อเข้ามาใน Power Query แล้ว เลือกตารางที่ต้องการแก้ไขข้อมูล Format

5.4 Click เลือก Column ที่ต้องการแก้ไข Format

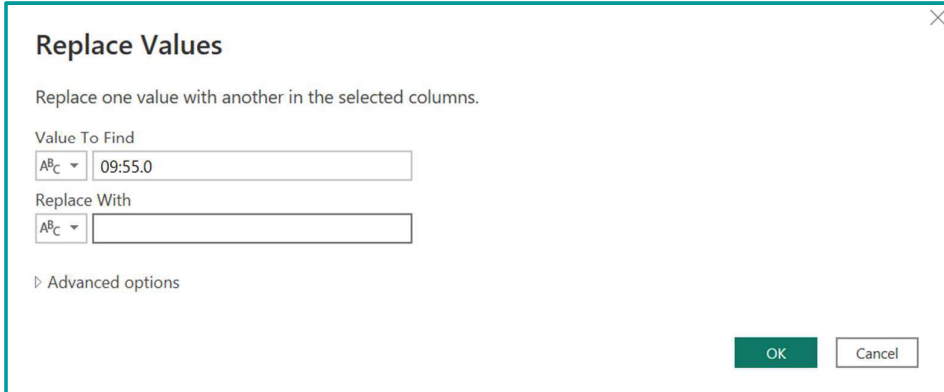
5.5 ไปที่แถวของข้อมูลที่มีรูปแบบผิดปกติ Click Mouse ขวา แล้วเลือก Replace value ดังภาพ



113	56	09:55.0
114	106	0:27:48
115	198	0:51:48
116	95	0:30:59
117	142	0:52:11
118	165	0:53:41

- Copy
- Text Filters
- Replace Values...
- Drill Down
- Add as New Query

5.6 จากนั้นจะปรากฏ pop window ขึ้นมาดังภาพ



Replace Values

Replace one value with another in the selected columns.

Value To Find
ABC 09:55.0

Replace With
ABC

Advanced options

OK Cancel

5.7 กรอกข้อมูลตามรูปแบบที่ต้องการในช่อง Replace with แล้วกด OK

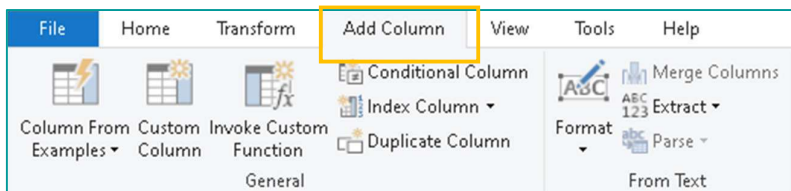
5.8 ไปที่ เมนูบาร์ด้านบน แล้วกดปุ่ม Save

6. การแปลงข้อมูลช่วงเวลาที่ถูกเก็บเป็นชนิดข้อมูล Text ให้เป็นชนิดข้อมูลที่สามารถนำไปคำนวณเวลารวมได้



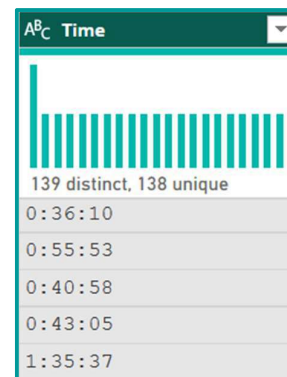
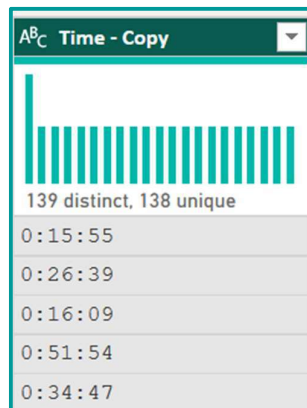
6.1 ทำการ Duplicate Column ที่เก็บข้อมูลช่วงเวลาเป็น Text ไว้ โดย

6.1.1 เลือกเมนู Add Column

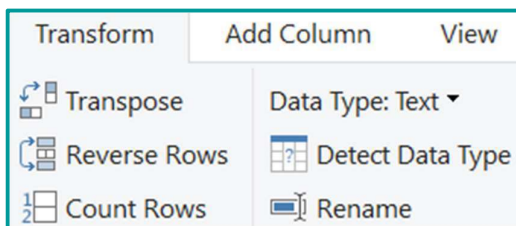


6.1.2 กดเลือก Column ต้นฉบับที่ต้องการสำเนา

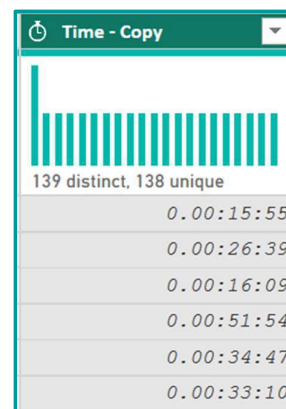
6.1.3 เลือก Duplicate Column

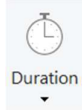


6.1.4 จากนั้นทำการเปลี่ยนชนิดข้อมูลของ Column copy โดยการไปที่ เมนู Transform และไปที่ Data Type



เปลี่ยน Data Type จาก Text ไปเป็น Duration

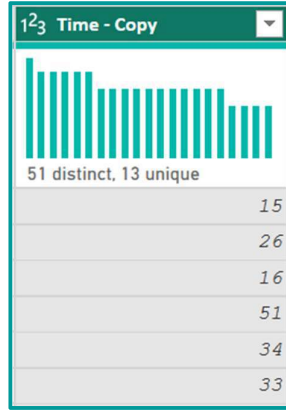
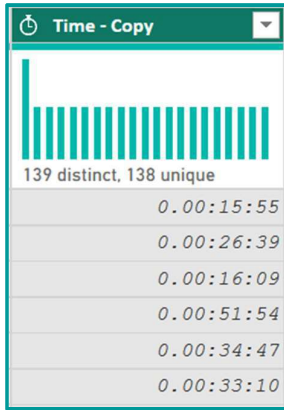




6.1.5 กดปุ่ม ที่เมนูบาร์ จะปรากฏ popup menu

6.1.6 เลือกผลลัพธ์การแปลงเป็นหน่วยเวลาที่ต้องการ

- Days
- Hours
- Minutes
- Seconds
- Total Years
- Total Days
- Total Hours
- Total Minutes
- Total Seconds



6.1.7 ไปที่ เมนูบาร์ด้านบน แล้วกดปุ่ม Save

6.1.8 กดปุ่ม Close & Apply



เมื่ออ่านมาถึงจุดนี้แล้ว จะเห็นได้ว่าการใช้งาน Power Query ในการเตรียมข้อมูลนั้นไม่ยากเลย
อยากให้ทุกคนลองทำกันดูนะคะ