

Master of Science Program in  
Artificial Intelligence for Business Analytics

# Data Science

## Week 3: Data Preparation

Eakasit Pacharawongsakda, Ph.D.

eakasit@datacubeth.ai

Co-Founder and Data Science Team Lead  
Cube Analytics Consulting Co., Ltd.



## Data Preparation

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL  
เนื่องจากเทคนิค Machine Learning แต่ละอย่างต้องการข้อมูลที่มีลักษณะ  
เฉพาะเจาะจงจำเป็นต้องมีการทำ Data Preparation
- Data Preparation เป็นกระบวนการที่สำคัญที่สุดในการทำโครงการทาง  
Machine Learning และก็เป็นขั้นตอนที่ใช้เวลาหนา  
บางครั้งเรียกว่า Data Preprocessing เป็นการแปลงข้อมูลให้เหมาะสมสำหรับ  
นำไปทำการวิเคราะห์

# Data Preparation

- สำหรับการศึกษาในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ขั้นตอนมาตรฐานในการทำ Data Preparation ได้แก่

- Data Cleaning:** Identifying and correcting mistakes or errors in the data.
- Data Transforms:** Changing the scale or distribution of variables.
- Feature Selection:** Identifying those input variables that are most relevant to the task.
- Feature Engineering:** Deriving new variables from available data.
- Dimensionality Reduction:** Creating compact projections of the data.

source: Data Preparation for Machine Learning

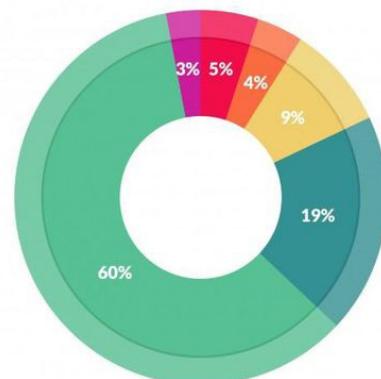
Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

3

# Data Preparation

- สำหรับการศึกษาในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- According to a survey in Forbes, data scientists **spend 80% of their time on data preparation:**



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

Source: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>

Copyright © 2025 Cube Analytics Consulting Co., Ltd.

<http://facebook.com/datacube.th>

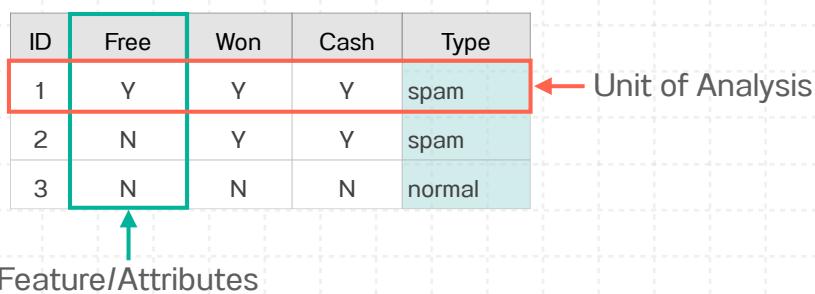
4

# Data Preparation for ML

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- Analytical Base Table (ABT)** คือ **ตารางที่รวมข้อมูล** ไว้เพื่อใช้ในการสร้างโมเดล ซึ่งประกอบด้วย

- Unit of Analysis** คือ **ระดับของข้อมูลที่จะทำการวิเคราะห์แสดงด้วย แถว (row)** เช่น e-mail แต่ละฉบับ ลูกค้า (customer) หรือ แทรนแซกชัน (transaction)
- Feature/Attribute** คือ **ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลแสดงด้วย คอลัมน์ (column)** เช่น คำที่เกิดขึ้นใน e-mail ความถี่ในการซื้อสินค้า (frequency) หรือ อายุ (age)



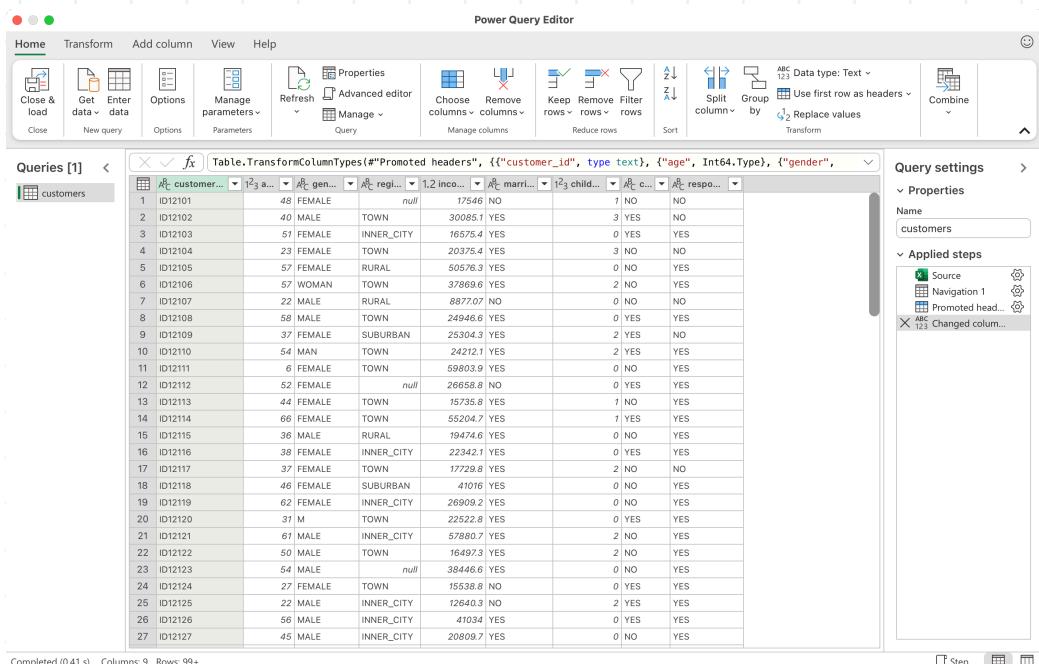
ID	Free	Won	Cash	Type
1	Y	Y	Y	spam
2	N	Y	Y	spam
3	N	N	N	normal

# Data Preparation with Power Query

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- PowerQuery เป็นซอฟต์แวร์ที่ใช้ในการจัดการและแก้ไขข้อมูลใน Excel และ Power BI

## Power BI



# Workshop 1

Customer Data Preparation with Power Query

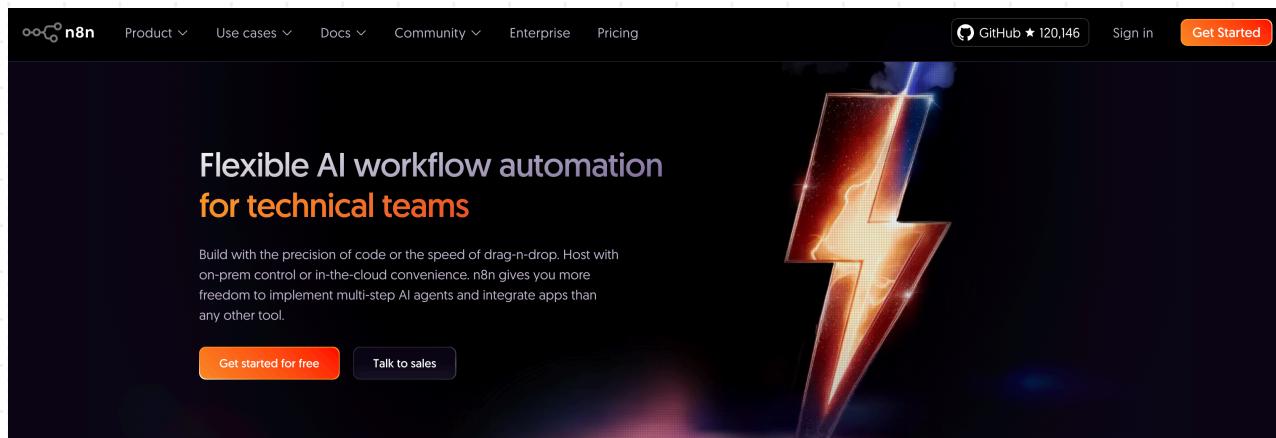
# Workshop 2

Transaction Data Preparation with Power Query

# Data Preparation with n8n

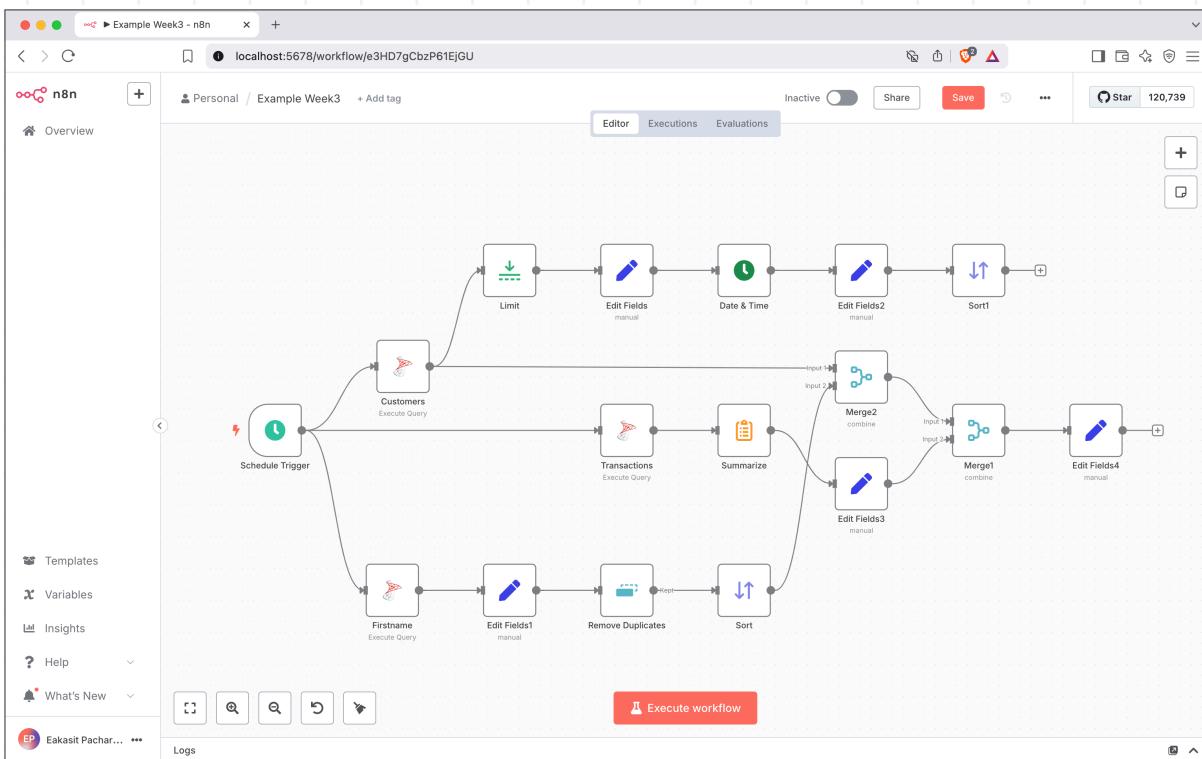
ສໍາເລັດການສູງຍາຍໃນຫຼັກສູດ M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- <https://n8n.io/>



# Data Preparation with n8n

ສໍາເລັດການສູງຍາຍໃນຫຼັກສູດ M. Sc. in Artificial Intelligence for Business Analytics, KMITL



# Data Preparation with n8n

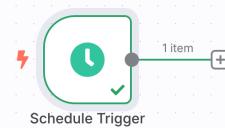
- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ตั้งเวลาให้ workflow ทำงาน

- อ่านข้อมูลจากฐานข้อมูล (database)
- Data Preparation
  - Filter
  - Limit
  - Select Fields
  - Edit Fields
  - Remove Duplicates
  - Sort
  - Summarize
  - Merge (Join)

## Schedule Trigger

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ใช้สำหรับการตั้งเวลาสำหรับการ execute workflow

- Trigger Interval:** ช่วงเวลาที่ต้องการตั้งเวลา
- Days/Weeks Between Triggers:** จำนวนวันหรือสัปดาห์ระหว่างการทำงาน
- Trigger on Weekdays:** กำหนดวันที่จะให้ execute
- Trigger at Hour:** ชั่วโมงที่จะให้เริ่มทำงาน
- Trigger at Minute:** นาทีที่จะให้เริ่มทำงาน



Schedule Trigger
Execute step

Parameters
Settings
Docs ↗

This workflow will run on the schedule you define here once you **activate** it.

For testing, you can also trigger it manually: by going back to the canvas and clicking 'execute workflow'

**Trigger Rules**

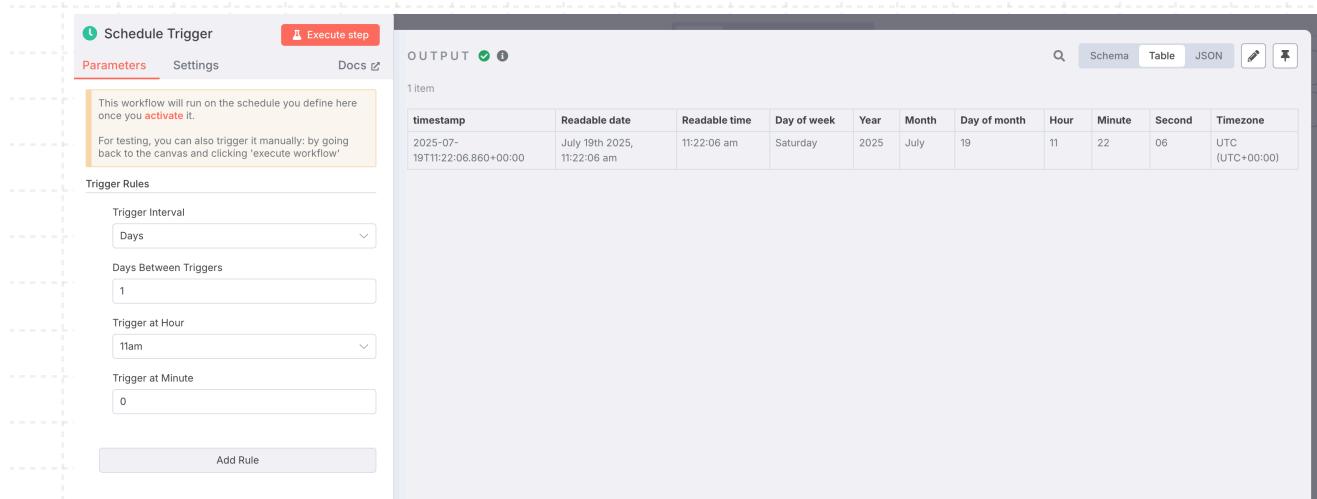
Trigger Interval	<input type="button" value="Days"/>
Days Between Triggers	<input type="text" value="1"/>
Trigger at Hour	<input type="button" value="Midnight"/>
Trigger at Minute	<input type="text" value="0"/>

# Example 1

## Set Schedule

# Example 1

- สำหรับการเรียนรู้ในหลักสูตร M.Sc. in Artificial Intelligence for Business Analytics, KMITL
- ตั้งเวลาให้ทำงานตอน 11:00 น.



**Schedule Trigger**

**Parameters** **Settings** **Docs** **Execute step**

This workflow will run on the schedule you define here once you **activate** it.

For testing, you can also trigger it manually: by going back to the canvas and clicking 'execute workflow'.

**Trigger Rules**

- Trigger Interval: Days
- Days Between Triggers: 1
- Trigger at Hour: 11am
- Trigger at Minute: 0

**OUTPUT** **1 item**

timestamp	Readable date	Readable time	Day of week	Year	Month	Day of month	Hour	Minute	Second	Timezone
2025-07-19T11:22:06.860+00:00	July 19th 2025, 11:22:06 am	Saturday	2025	July	19	11	22	06	UTC (UTC+00:00)	

# Data Preparation with n8n

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL  
• ตั้งเวลาให้ workflow ทำงาน

## • อ่านข้อมูลจากฐานข้อมูล (database)

### • Data Preparation

- Filter
- Limit
- Select Fields
- Edit Fields
- Remove Duplicates
- Sort
- Summarize
- Merge (Join)

# Connect to database

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL  
• สร้าง connection ไปยังฐานข้อมูล SQLServer ที่อยู่บน cloud

- ใส่รายละเอียดของ database server ดังนี้
  - **Server:** สำหรับใส่หมายเลข IP ของฐานข้อมูล
  - **Database:** สำหรับใส่ชื่อของฐานข้อมูล
  - **User:** สำหรับใส่ username เพื่อ connect กับฐานข้อมูล
  - **Password:** สำหรับใส่ password เพื่อ connect กับฐานข้อมูล
  - **Port:** สำหรับใส่พอร์ตของฐานข้อมูล
  - **TLS:** มีการใช้ Transport Layer Security (TLS) หรือไม่
  - **Ignore SSL Issues (Insecure):** ใช้การ connect ผ่าน SSL หรือไม่

# Microsoft SQL

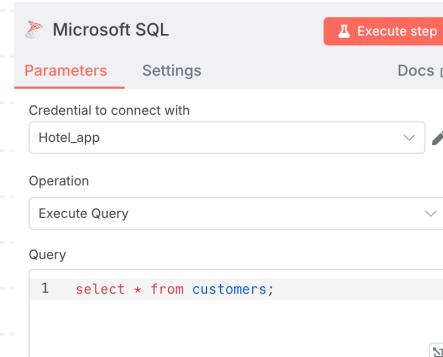
สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ใช้สำหรับการตั้งเวลาสำหรับการ execute workflow

- Credential to connect with:** เลือก credential ที่สร้างไว้เพื่อเชื่อมต่อกับฐานข้อมูล

- Operation:**

- Execute Query:** เขียนคำสั่ง SQL
- Insert:** เพิ่มแถวในตาราง
- Update:** แก้ไขข้อมูลในตาราง
- Delete:** ลบแถวในตาราง



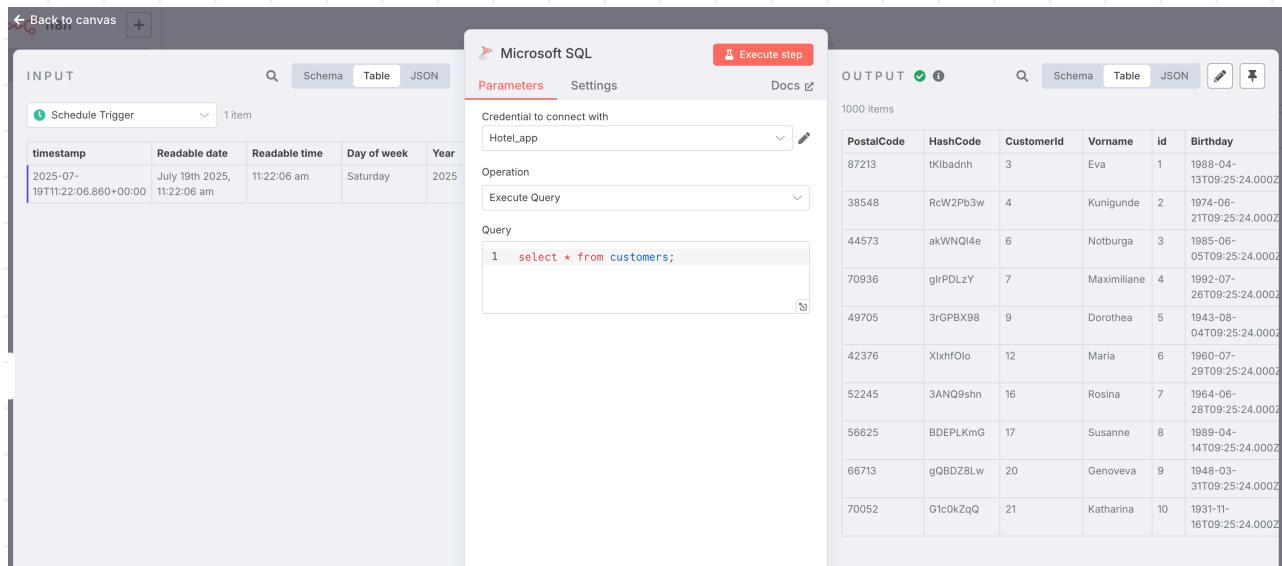
## Example 2

Connect to database

## Example 2.1

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ดึงข้อมูลจากตาราง Customers

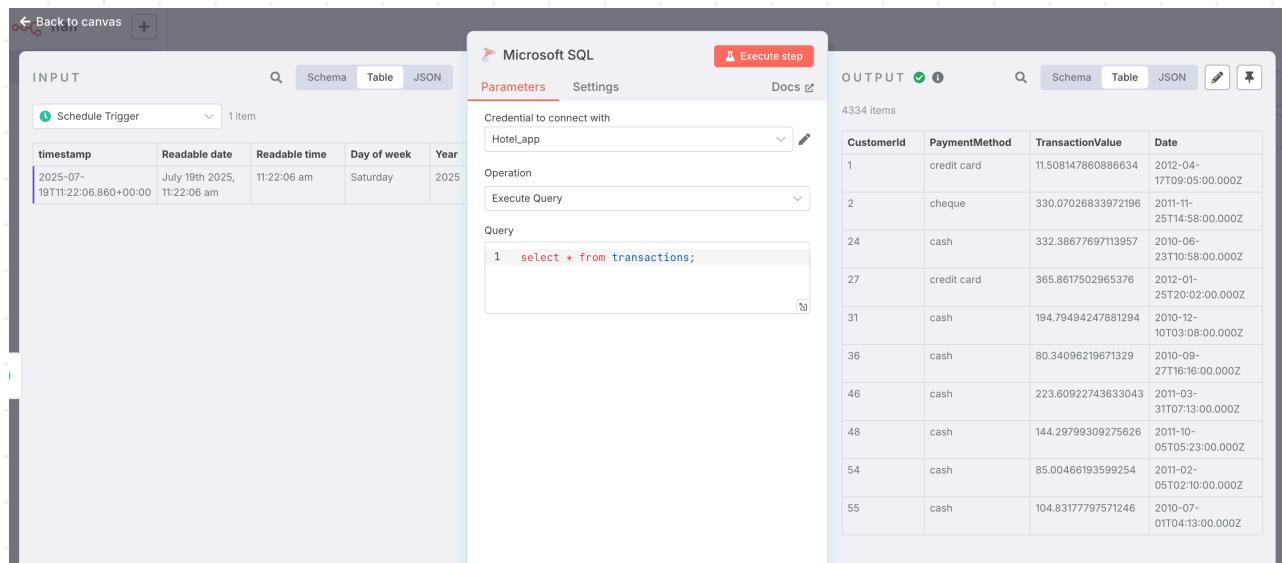


PostalCode	HashCode	CustomerId	Vorname	id	Birthday
87213	tkIbdnH	3	Eva	1	1988-04-13T09:25:24.000Z
38548	RcW2Pb3w	4	Kunigunde	2	1974-06-21T09:25:24.000Z
44573	akWNQ14e	6	Notburga	3	1985-06-05T09:25:24.000Z
70936	gIrpDLzY	7	Maximiliane	4	1992-07-26T09:25:24.000Z
49705	3rGPBX98	9	Dorothea	5	1943-08-04T09:25:24.000Z
42376	XlxhfOlo	12	Maria	6	1960-07-29T09:25:24.000Z
52245	3ANQ9shn	16	Rosina	7	1964-06-28T09:25:24.000Z
56625	BDEPLKmG	17	Susanne	8	1989-04-14T09:25:24.000Z
66713	gQBDZ8Lw	20	Genoveva	9	1948-03-31T09:25:24.000Z
70052	G1c0kZqQ	21	Katharina	10	1931-11-16T09:25:24.000Z

## Example 2.2

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ดึงข้อมูลจากตาราง Transactions

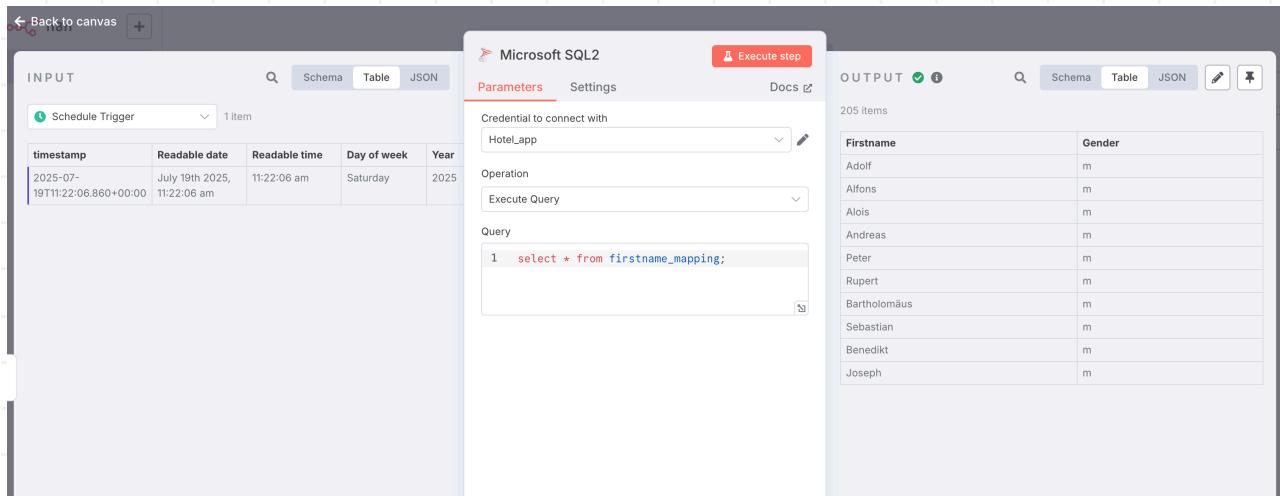


CustomerId	PaymentMethod	TransactionValue	Date
1	credit card	11.508147860886634	2012-04-17T09:05:00.000Z
2	cheque	330.07026833972196	2011-11-25T16:58:00.000Z
24	cash	332.38677697113957	2010-06-23T10:58:00.000Z
27	credit card	365.8617502965376	2012-01-25T20:02:00.000Z
31	cash	194.79494247881294	2010-12-10T03:08:00.000Z
36	cash	80.34096219671329	2010-09-27T16:16:00.000Z
46	cash	223.60922743633043	2011-03-31T07:13:00.000Z
48	cash	144.29799309275626	2011-10-05T05:23:00.000Z
54	cash	85.00466193599254	2011-02-05T02:10:00.000Z
55	cash	104.83177797571246	2010-07-01T04:13:00.000Z

## Example 2.3

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ดึงข้อมูลจากตาราง Firstname\_Mapping



The screenshot shows a data integration tool's workflow editor. It consists of three main panels:

- INPUT:** A table with one row containing timestamp, Readable date, Readable time, Day of week, and Year.
- Microsoft SQL:** A step configuration panel with the following settings:
  - Credential to connect with: Hotel\_App
  - Operation: Execute Query
  - Query: `1 select * from firstname_mapping;`
- OUTPUT:** A table showing the results of the query, with columns Firstname and Gender. The data includes:
 

Firstname	Gender
Adolf	m
Alfons	m
Alois	m
Andreas	m
Peter	m
Rupert	m
Bartholomäus	m
Sebastian	m
Benedikt	m
Joseph	m

## Data Preparation with n8n

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ตั้งเวลาให้ workflow ทำงาน
- อ่านข้อมูลจากฐานข้อมูล (database)
- Data Preparation**

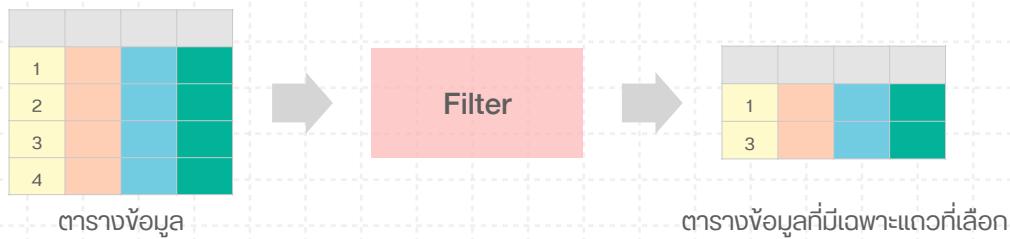
- Filter**

- Limit
- Select Fields
- Edit Fields
- Remove Duplicates
- Sort
- Summarize
- Merge (Join)

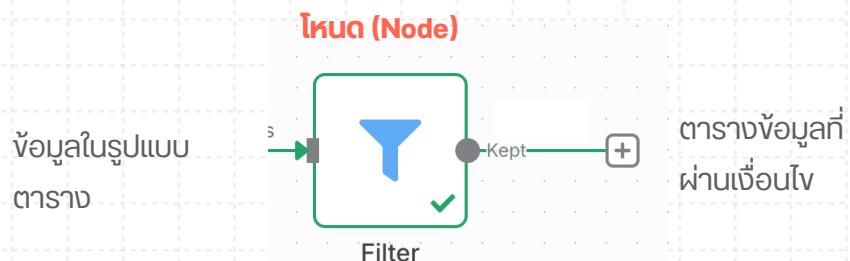
# Filter

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ใช้ในการเลือกตัวแอดวานซ์ในตารางข้อมูลตามเงื่อนไขที่ระบุไว้



- พบด้วยใช้ชื่อว่า **Filter**

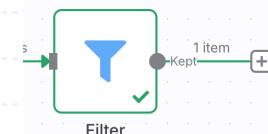
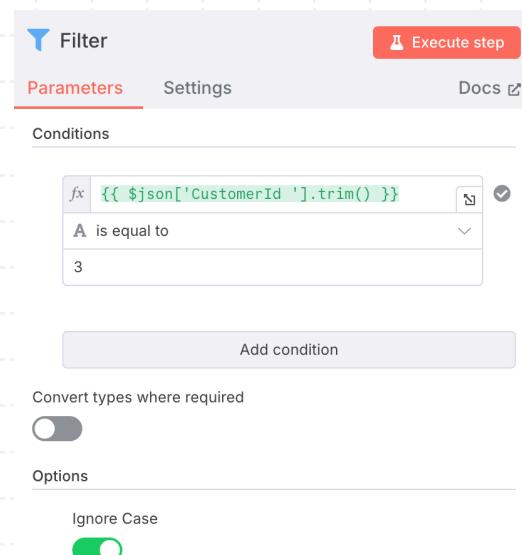


# Filter

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ใช้ในการเลือกตัวแอดวานซ์ในตารางข้อมูลตามเงื่อนไขที่ระบุไว้

- Conditions:** เงื่อนไขที่ต้องการ filter ข้อมูล
- Convert types where required:** เปลี่ยนประเภทข้อมูล
- Ignore Case:** ไม่ต้องพิจารณาอักษรตัวเล็ก  
หรือใหญ่

Filter

Parameters    Settings    Docs

Conditions

fx {{ \$json['CustomerId'] }}.trim()  
A is equal to  
3

Add condition

Convert types where required

Options

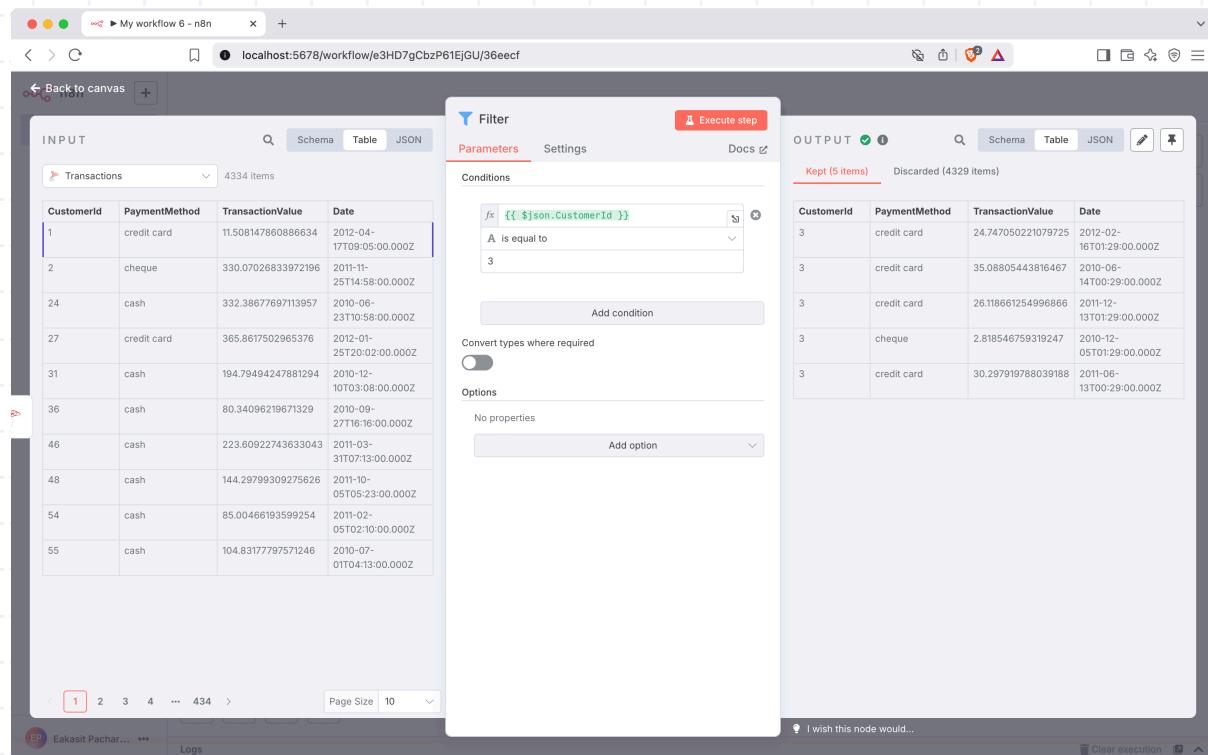
Ignore Case

# Example 3

Filter

## Example 3.1

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- แสดงเฉพาะข้อมูล transaction ของลูกค้าหมายเลข 3



The screenshot shows the CubeSpace interface with a 'Filter' step. The input table contains 4334 items with columns: CustomerId, PaymentMethod, TransactionValue, and Date. A specific row for CustomerId 1 is selected. The filter condition is set to '\$json.CustomerId' is equal to 3. The output table shows 5 kept items for CustomerId 3, with the same columns.

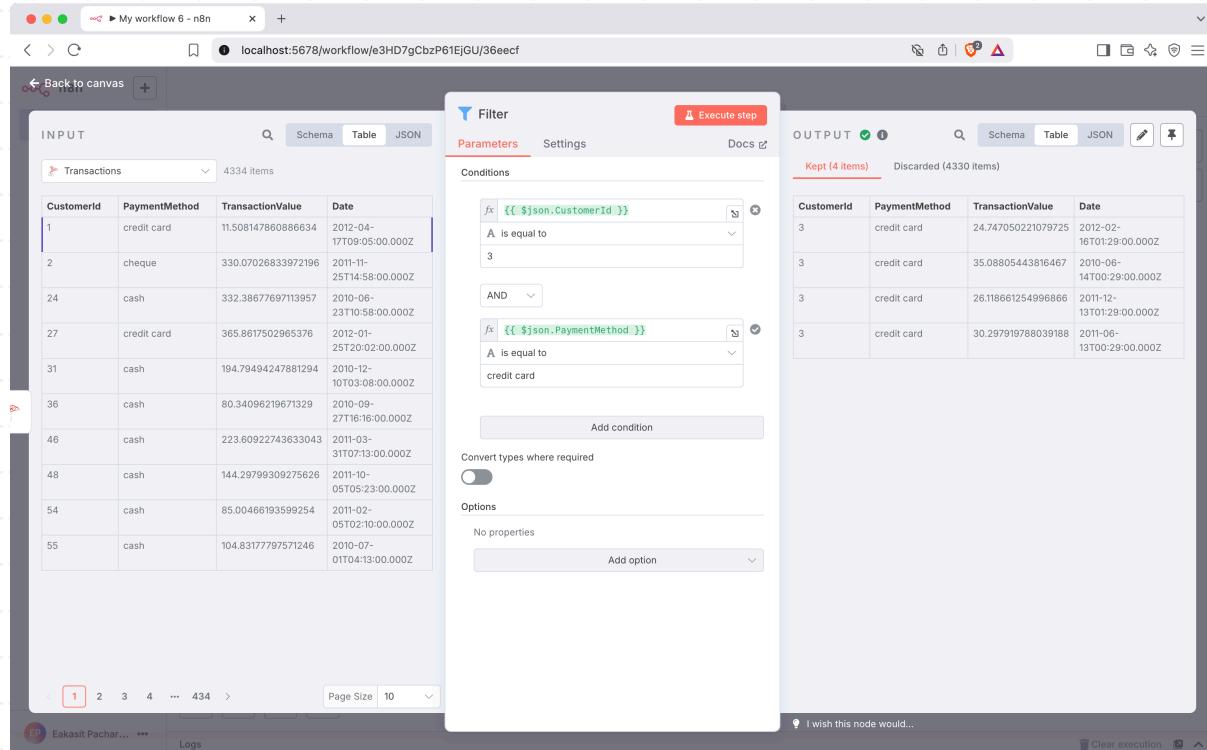
Customerid	PaymentMethod	TransactionValue	Date
1	credit card	11.508147860886634	2012-04-17T09:05:00.000Z
2	cheque	330.07026833972196	2011-11-25T14:58:00.000Z
24	cash	332.38677697113957	2010-06-23T10:58:00.000Z
27	credit card	365.8617502965376	2012-01-25T20:02:00.000Z
31	cash	194.79494247881294	2010-12-10T03:08:00.000Z
36	cash	80.34096219671329	2010-09-27T16:16:00.000Z
46	cash	223.60922743633043	2011-03-31T07:13:00.000Z
48	cash	144.29799309275626	2011-10-05T05:23:00.000Z
54	cash	85.00466193599254	2011-02-05T02:10:00.000Z
55	cash	104.83177797571246	2010-07-01T04:13:00.000Z

Customerid	PaymentMethod	TransactionValue	Date
3	credit card	24.747050221079725	2012-02-16T01:29:00.000Z
3	credit card	35.08805443816467	2010-06-14T00:29:00.000Z
3	credit card	26.118661254996866	2011-12-13T01:29:00.000Z
3	cheque	2.818546759319247	2010-12-05T01:29:00.000Z
3	credit card	30.297919788039188	2011-06-13T00:29:00.000Z

## Example 3.2

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- แสดงเฉพาะข้อมูล transaction ของลูกค้าหมายเลข 3 ที่ชำระเงินด้วยบัตรเครดิต



**INPUT**

CustomerId	PaymentMethod	TransactionValue	Date
1	credit card	11.508147860886634	2012-04-17T09:05:00.000Z
2	cheque	330.07026833972196	2011-11-25T14:58:00.000Z
24	cash	332.38677697113957	2010-06-23T10:58:00.000Z
27	credit card	365.861502965376	2012-01-25T20:02:00.000Z
31	cash	194.79494247881294	2010-12-10T03:08:00.000Z
36	cash	80.34096219671329	2010-09-27T16:16:00.000Z
46	cash	223.60922743633043	2011-03-31T07:13:00.000Z
48	cash	144.29799309275626	2011-10-05T05:23:00.000Z
54	cash	85.00466193599254	2011-02-05T02:10:00.000Z
55	cash	104.8317797571246	2010-07-01T04:13:00.000Z

**Filter**

Parameters

Conditions

```

fx: {{ $json.CustomerId }}  

A is equal to  

3  

AND  

fx: {{ $json.PaymentMethod }}  

A is equal to  

credit card

```

Add condition

Convert types where required

Options

No properties

Add option

**OUTPUT**

Kept (4 items) Discarded (4330 items)

CustomerId	PaymentMethod	TransactionValue	Date
3	credit card	24.747050221079725	2012-02-16T01:29:00.000Z
3	credit card	35.08805443816467	2010-06-14T00:29:00.000Z
3	credit card	26.118661254996866	2011-12-13T01:29:00.000Z
3	credit card	30.297919788039188	2011-06-13T00:29:00.000Z

## Data Preparation with n8n

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ตั้งเวลาให้ workflow ทำงาน

- อ่านข้อมูลจากฐานข้อมูล (database)

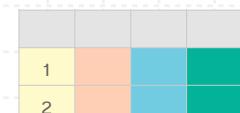
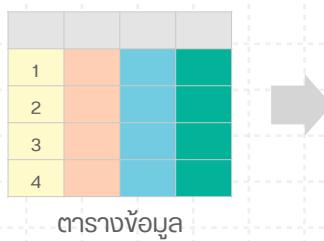
- Data Preparation

- Filter
- Limit**
- Select Fields
- Edit Fields
- Remove Duplicates
- Sort
- Summarize
- Merge (Join)

# Limit

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ใช้ในการเลือกແກ່ໃນຕາරັງຂ້ອມູລຕາມຈຳນວນແດວທີ່ຮັບໄວ



ຕາරັງຂ້ອມູລືກໍ່ແສດງ 2 ແດວແຮກ

- ໂຫດທີ່ໃຊ້ ຜ່ອວ່າ **Limit**

## ໂຫດ (Node)

ຂ້ອມູລໃນຮູບແບບ  
ຕາරັງ



ຕາරັງຂ້ອມູລືກໍ່  
ແສດງເດພາະແດວ  
ຕາມຈຳນວນທີ່ຮັບໄວ

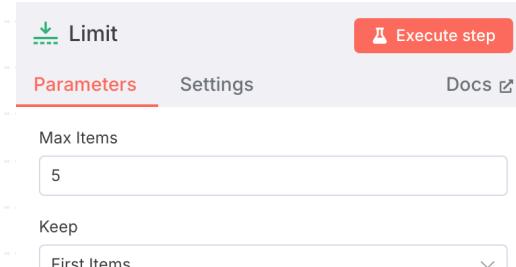
# Limit

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ใช้ในการเลือกແກ່ໃນຕາරັງຂ້ອມູລຕາມຈຳນວນແດວທີ່ຮັບໄວ

- Max Items:** ຈຳນວນແດວທີ່ຕ້ອງການເລືອກ

- Keep:** ເລືອກຈາກແດວເຮັ່ນຕັນຂອງຕາරັງຫຼືກ້າຍຕາරັງ

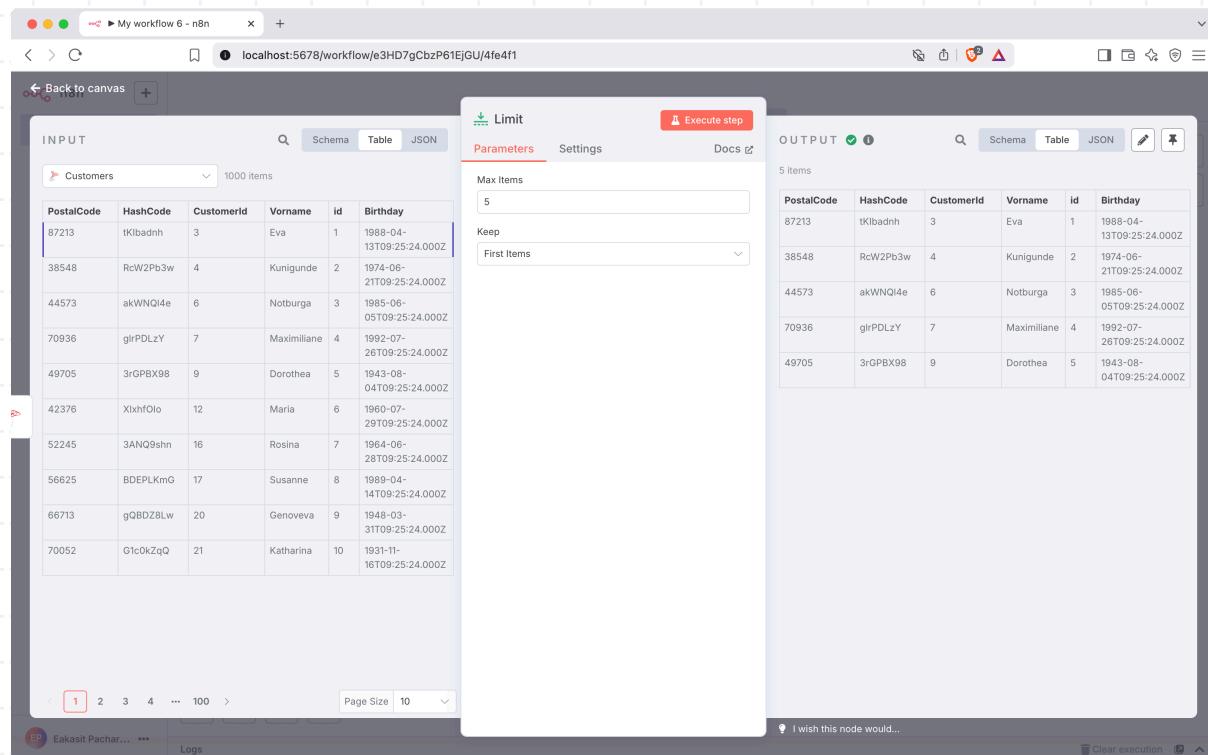


# Example 4

Limit

## Example 4

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- แสดงข้อมูลรายชื่อลูกค้า 5 แถวแรก



The screenshot shows a workflow step titled "Limit". The "Parameters" tab is selected, displaying "Max Items" set to 5 and "Keep" set to "First Items". The "INPUT" section shows a table of customer data with 1000 items. The "OUTPUT" section shows the result of the limit step, which contains 5 items from the input table.

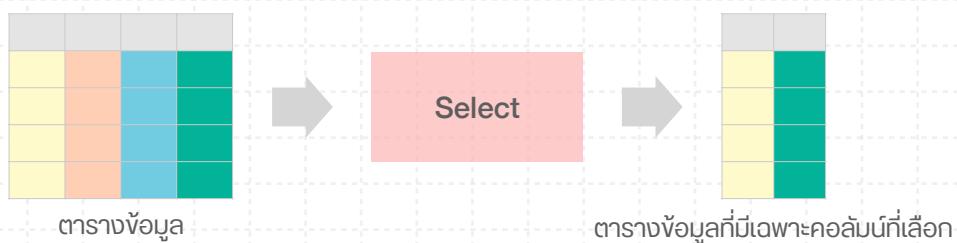
PostalCode	HashCode	Customerid	Vorname	Id	Birthday
87213	tKlbadnh	3	Eva	1	1988-04-13T09:25:24.000Z
38548	RcW2Pb3w	4	Kunigunde	2	1974-06-21T09:25:24.000Z
44573	akWNQI4e	6	Notburga	3	1985-06-05T09:25:24.000Z
70936	gIrPDLzY	7	Maximiliane	4	1992-07-26T09:25:24.000Z
49705	3rGPBX98	9	Dorothea	5	1943-08-04T09:25:24.000Z
42376	XixhfOlo	12	Maria	6	1960-07-29T09:25:24.000Z
52245	3ANQ9shn	16	Rosina	7	1964-06-28T09:25:24.000Z
56625	BDEPLKmg	17	Susanne	8	1989-04-14T09:25:24.000Z
66713	gQBDZ8Lw	20	Genoveva	9	1948-03-31T09:25:24.000Z
70052	G1c0kZqQ	21	Katharina	10	1931-11-16T09:25:24.000Z

# Data Preparation with n8n

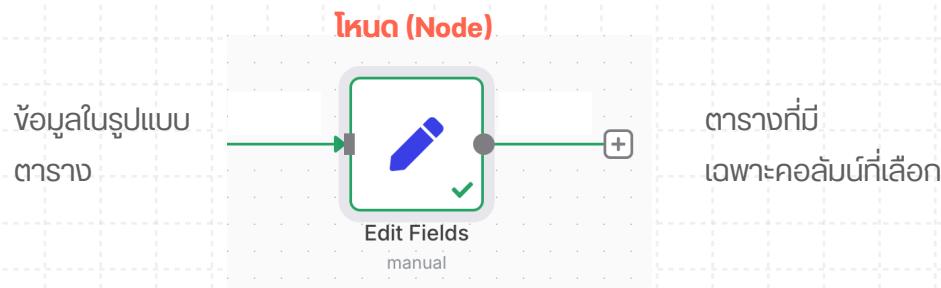
- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ตั้งเวลาให้ workflow ทำงาน
- อ่านข้อมูลจากฐานข้อมูล (database)
- Data Preparation
  - Filter
  - Limit
  - Select Fields**
  - Edit Fields
  - Remove Duplicates
  - Sort
  - Summarize
  - Merge (Join)

## Select

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ใช้ในการเลือกฟิลด์หรือคอลัมน์ในตารางข้อมูล



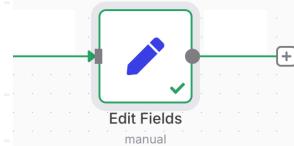
- ให้กดที่ใช้ชื่อว่า **Edit Fields (Set)**



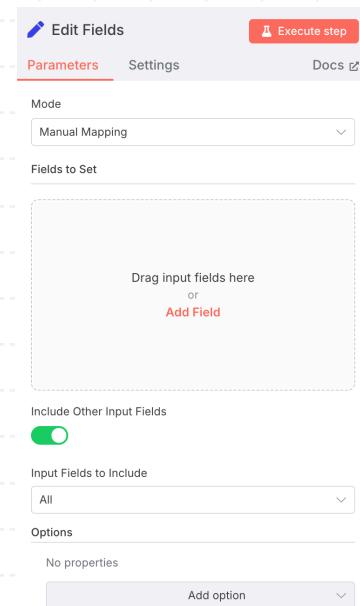
# Edit Fields

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ใช้ในการเลือกฟล็อตหรือคอลัมน์ในตารางข้อมูล



- Mode:** เลือกแบบกำหนดค่าเองหรือใช้ JSON
- Fields to Set:** เลือกคอลัมน์ที่ต้องการเปลี่ยนแปลงค่า เช่น เปลี่ยนชื่อคอลัมน์หรือคำนวนค่าใหม่
- Include Other Input Fields:** เลือกด้าต้องการให้แสดง คอลัมน์เดิมในตาราง
- Input Fields to Include:**
  - All:** แสดงทุกคอลัมน์ในตาราง
  - Selected:** แสดงเฉพาะคอลัมน์ที่เลือก
  - All Except:** แสดงทุกคอลัมน์ยกเว้นคอลัมน์ที่เลือก

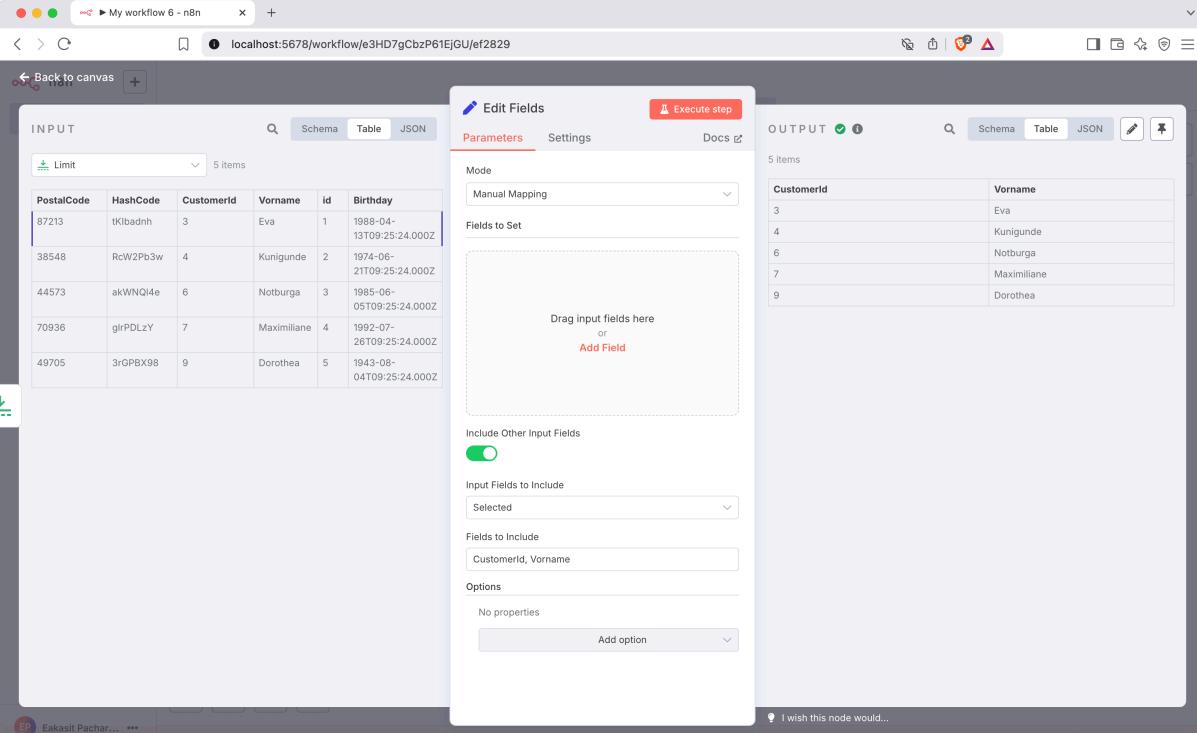


## Example 5

Select Fields

## Example 5.1

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- แสดงข้อมูลเฉพาะคอลัมน์ CustomerId และ Vorname

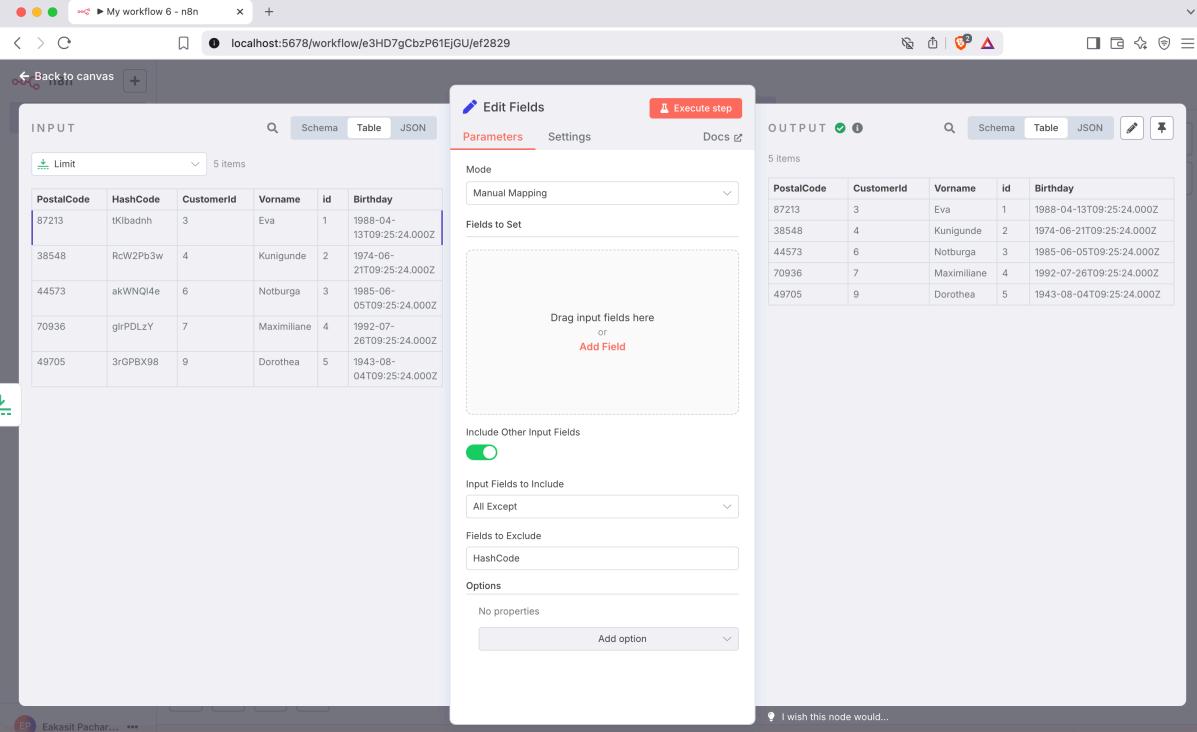


The screenshot shows the Datacube interface with the 'Edit Fields' node selected. The input table contains columns: PostalCode, HashCode, CustomerId, Vorname, Id, and Birthday. The output table shows only the CustomerId and Vorname columns. The 'Fields to Set' section is configured with 'Manual Mapping' mode, mapping 'CustomerId' and 'Vorname' from the input to the output. The 'Fields to Include' dropdown is set to 'Selected'. The 'Fields to Exclude' dropdown is empty.

CustomerId	Vorname
3	Eva
4	Kunigunde
6	Notburga
7	Maximiliane
9	Dorothea

## Example 5.2

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- แสดงข้อมูลทุกคอลัมน์ยกเว้นคอลัมน์HashCode



The screenshot shows the Datacube interface with the 'Edit Fields' node selected. The input table is the same as in Example 5.1. The output table shows all columns: PostalCode, CustomerId, Vorname, Id, and Birthday. The 'Fields to Set' section is configured with 'Manual Mapping' mode, mapping 'CustomerId' and 'Vorname' from the input to the output. The 'Fields to Include' dropdown is set to 'All Except', and the 'Fields to Exclude' dropdown is set to 'HashCode'.

PostalCode	CustomerId	Vorname	Id	Birthday
87213	3	Eva	1	1988-04-13T09:25:24.000Z
38548	4	Kunigunde	2	1974-06-21T09:25:24.000Z
44573	6	Notburga	3	1985-06-05T09:25:24.000Z
70936	7	Maximiliane	4	1992-07-26T09:25:24.000Z
49705	9	Dorothea	5	1943-08-04T09:25:24.000Z

# Data Preparation with n8n

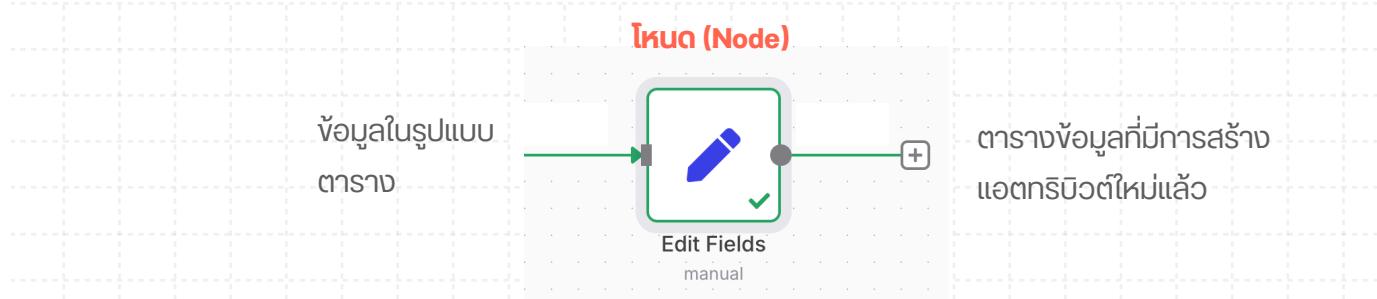
- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ตั้งเวลาให้ workflow ทำงาน
- อ่านข้อมูลจากฐานข้อมูล (database)
- Data Preparation
  - Filter
  - Limit
  - Select Fields
  - Edit Fields**
  - Remove Duplicates
  - Sort
  - Summarize
  - Merge (Join)

## Edit Fields

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ใช้ใน **สร้างฟลัตหรือคอลัมน์ใหม่** ในตารางข้อมูล



- ให้กดที่ใช้ชื่อว่า **Edit Fields (Set)**

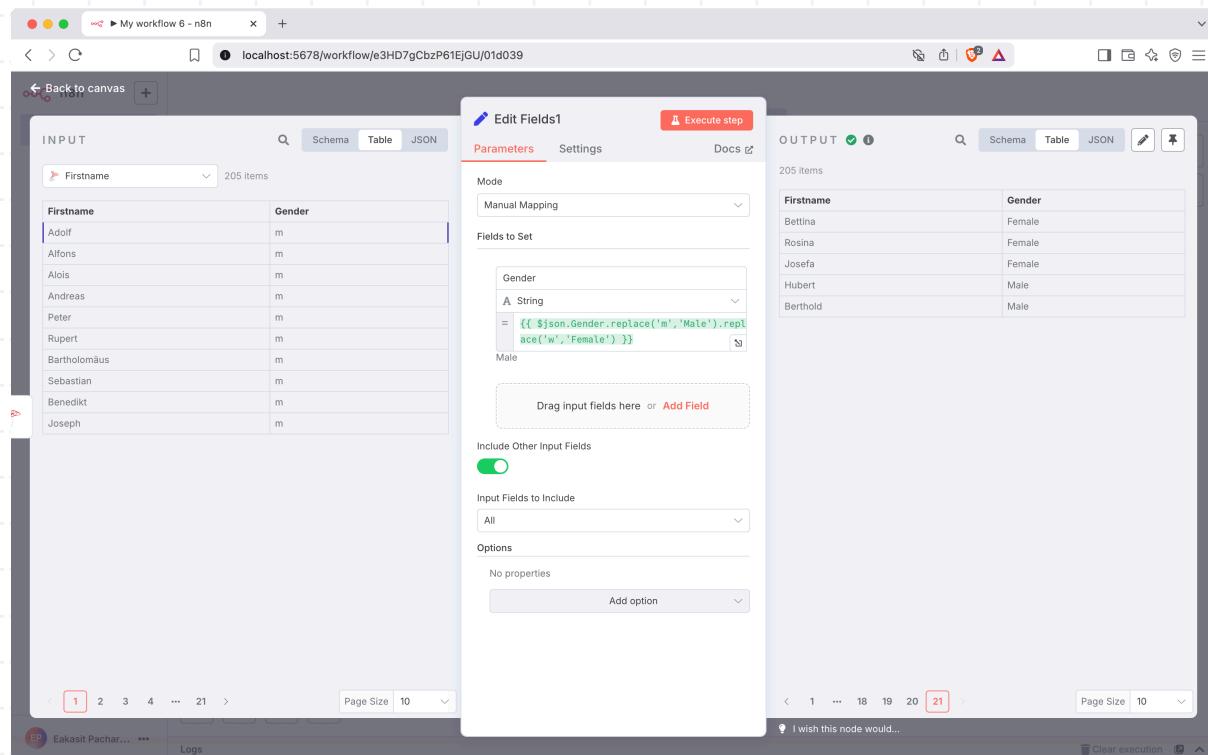


# Example 6

## Edit Fields

### Example 6.1

- สำหรับการสอนรายวิชาหลักสูตร M.Sc. in Artificial Intelligence for Business Analytics, KMITL
- Replace คำว่า 'm' ให้เป็น 'Male' และ 'w' ให้เป็น 'Female'



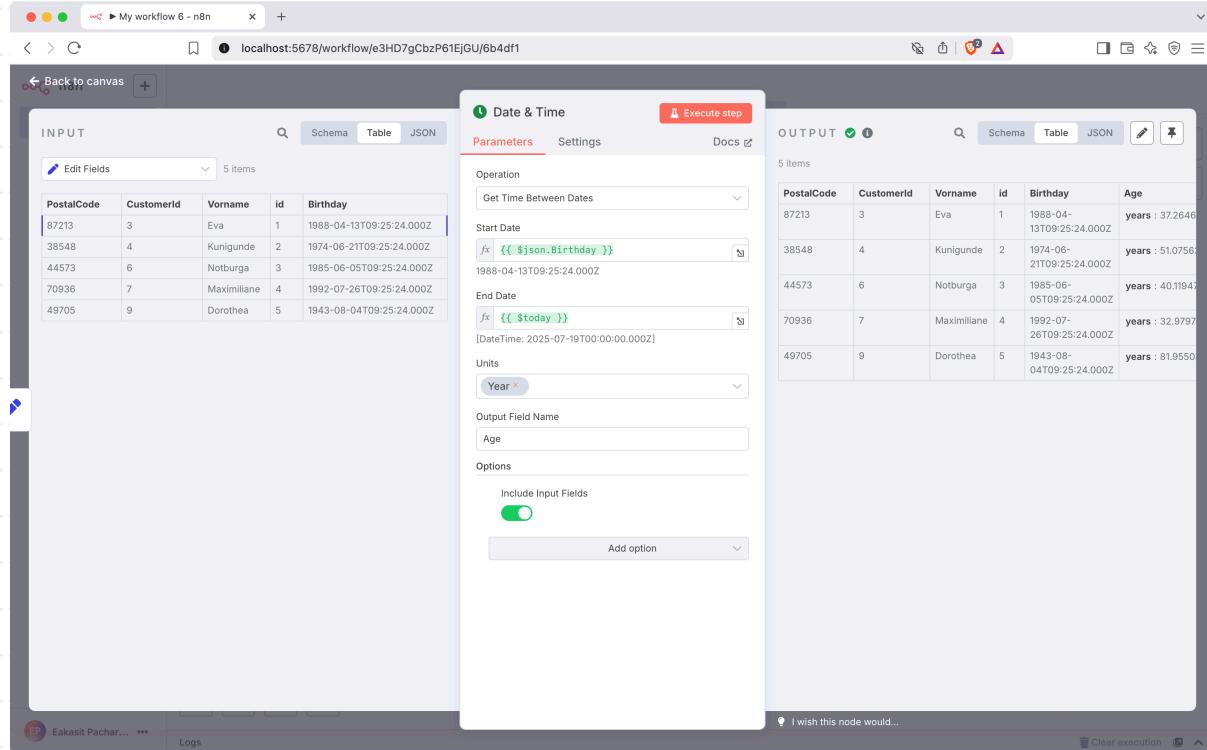
The screenshot shows the 'Edit Fields1' node configuration in the CubeSpace workflow editor. The 'Mode' is set to 'Manual Mapping'. The 'Fields to Set' section shows a mapping for the 'Gender' field from 'String' to 'Gender'. The formula is:

```
= {{ $json.Gender.replace('m','Male').replace('w','Female') }}
```

The 'Input' table (left) contains 205 items with columns 'Firstname' and 'Gender'. The 'Output' table (right) also has 205 items with columns 'Firstname' and 'Gender'. Both tables show the same data, with 'm' replaced by 'Male' and 'w' replaced by 'Female'.

## Example 6.2

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- สร้างคอลัมน์อายุจากวันเกิด



**INPUT**

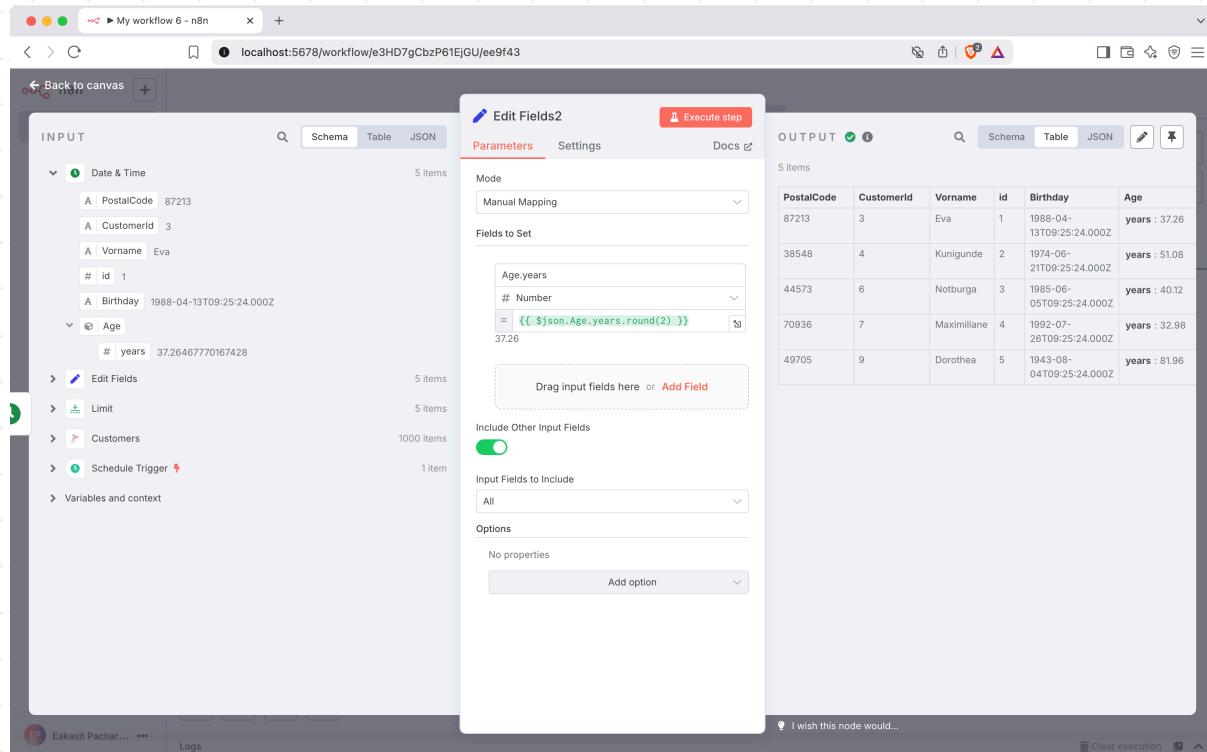
PostalCode	CustomerId	Vorname	id	Birthday
87213	3	Eva	1	1988-04-13T09:25:24.000Z
38548	4	Kunigunde	2	1974-06-21T09:25:24.000Z
44573	6	Notburga	3	1985-06-05T09:25:24.000Z
70936	7	Maximiliane	4	1992-07-26T09:25:24.000Z
49705	9	Dorothea	5	1943-08-04T09:25:24.000Z

**OUTPUT**

PostalCode	CustomerId	Vorname	id	Birthday	Age
87213	3	Eva	1	1988-04-13T09:25:24.000Z	years : 37.2646
38548	4	Kunigunde	2	1974-06-21T09:25:24.000Z	years : 51.0756
44573	6	Notburga	3	1985-06-05T09:25:24.000Z	years : 40.1194
70936	7	Maximiliane	4	1992-07-26T09:25:24.000Z	years : 32.9797
49705	9	Dorothea	5	1943-08-04T09:25:24.000Z	years : 81.9550

## Example 6.2

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- สร้างคอลัมน์อายุจากวันเกิด



**INPUT**

Mode	Manual Mapping
A PostalCode	87213
A CustomerId	3
A Vorname	Eva
# id	1
A Birthday	1988-04-13T09:25:24.000Z
# Age	# years : 37.26467770167428

**OUTPUT**

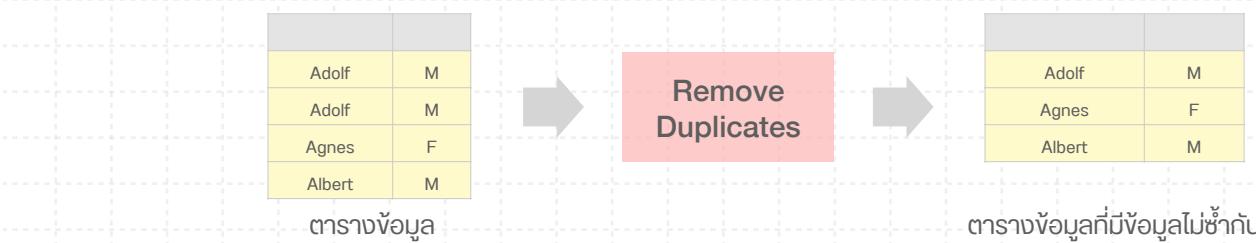
PostalCode	CustomerId	Vorname	id	Birthday	Age
87213	3	Eva	1	1988-04-13T09:25:24.000Z	years : 37.26
38548	4	Kunigunde	2	1974-06-21T09:25:24.000Z	years : 51.08
44573	6	Notburga	3	1985-06-05T09:25:24.000Z	years : 40.12
70936	7	Maximiliane	4	1992-07-26T09:25:24.000Z	years : 32.98
49705	9	Dorothea	5	1943-08-04T09:25:24.000Z	years : 81.96

# Data Preparation with n8n

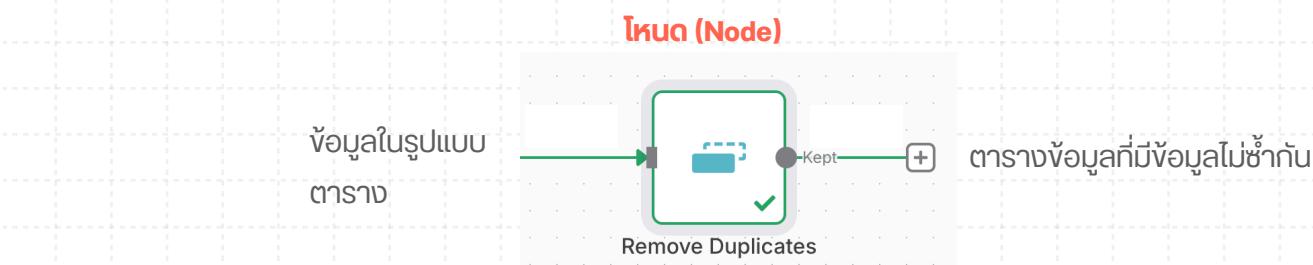
- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ตั้งเวลาให้ workflow ทำงาน
  - อ่านข้อมูลจากฐานข้อมูล (database)
  - Data Preparation
    - Filter
    - Limit
    - Select Fields
    - Edit Fields
    - **Remove Duplicates**
    - Sort
    - Summarize
    - Merge (Join)

## Remove Duplicates

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ใช้ในการลบเดอที่ซ้ำกันในตารางข้อมูล



- ให้กดที่ชื่อว่า **Remove Duplicates**



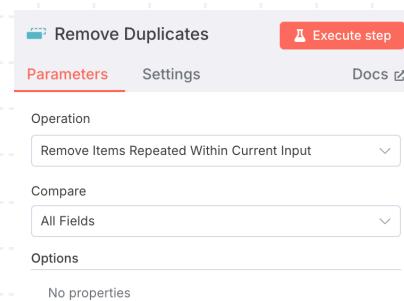
# Remove Duplicates

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ใช้ในการ **ลบเดวกี่ซ้ำกัน** ในตารางข้อมูล

- Operation:** เลือกรูปแบบการหา duplicate

- Compare:**

- All Fields:** พิจารณาทุกคอลัมน์ในตาราง
- All Fields Except:** พิจารณาทุกคอลัมน์ในตาราง **ยกเว้น** คอลัมน์ที่เลือก
- Selected Fields:** พิจารณาเฉพาะคอลัมน์ที่เลือก

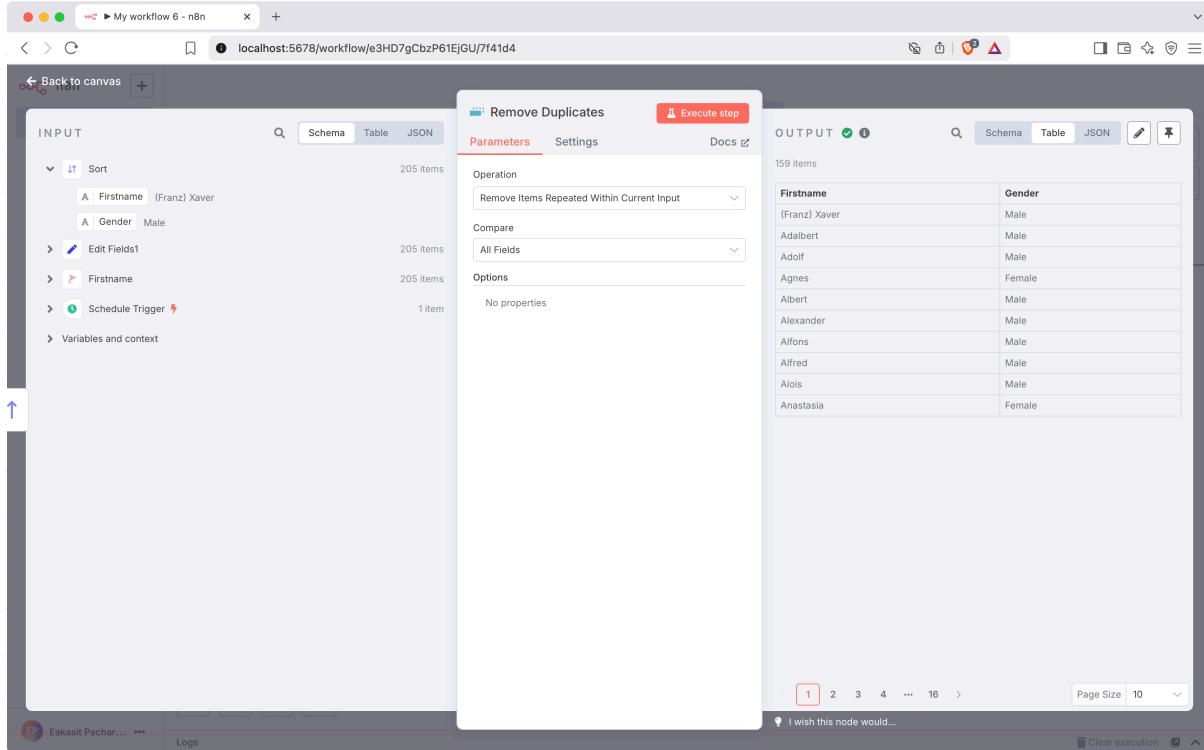



## Example 7

Remove Duplicates

# Example 7

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ลบข้อมูลซื้อสูกค้าที่เหมือนกันให้เหลือแค่ 1 แอด



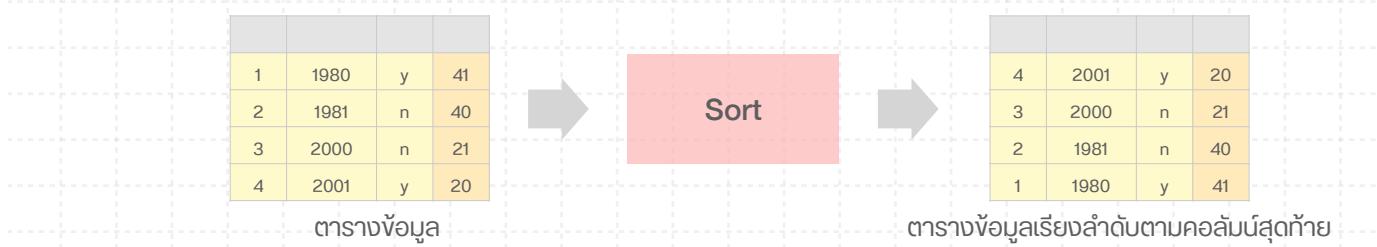
The screenshot shows the n8n workflow editor. On the left, the 'INPUT' panel lists various nodes: 'Sort', 'Edit Fields1', 'Firstname', 'Schedule Trigger', and 'Variables and context'. The 'Sort' node has two items: 'A Firstname (Franz) Xaver' and 'A Gender Male'. The 'Edit Fields1' node has one item: 'All Fields'. The 'Firstname' node has one item: 'Firstname'. The 'Schedule Trigger' node has one item: '1 item'. In the center, a 'Remove Duplicates' step is selected. The 'Parameters' tab is active, showing 'Operation: Remove Items Repeated Within Current Input', 'Compare: All Fields', and 'Options: No properties'. To the right, the 'OUTPUT' panel shows a table with 159 items. The columns are 'Firstname' and 'Gender'. The data includes names like Franz, Adalbert, Adolf, Agnes, Albert, Alexander, Alifons, Alfred, Alois, and Anastasia, with genders Male or Female. At the bottom, there is a page navigation bar with '1' highlighted.

# Data Preparation with n8n

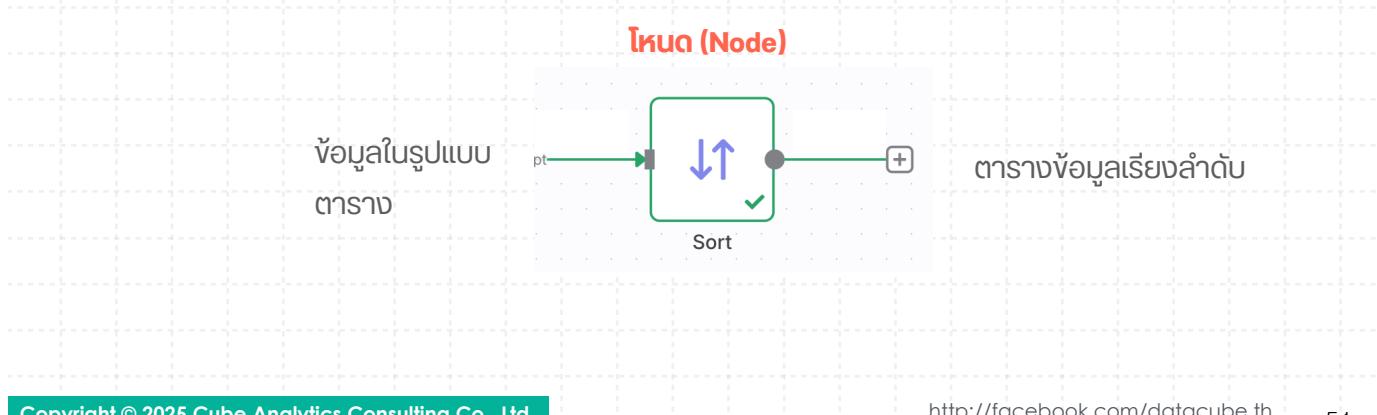
- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ตั้งเวลาให้ workflow ทำงาน
  - อ่านข้อมูลจากฐานข้อมูล (database)
  - Data Preparation
    - Filter
    - Limit
    - Select Fields
    - Edit Fields
    - Remove Duplicates
    - Sort**
    - Summarize
    - Merge (Join)

# Sort

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ใช้ในการ **เรียงลำดับข้อมูล** ในตาราง



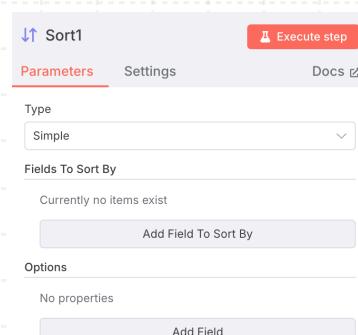
- พบด้วยว่า **Sort**



# Sort

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ใช้ในการ **เรียงลำดับข้อมูล** ในตาราง

- Type:** เลือกประเภทของการ Sort ถ้าเป็น Simple จะต้องเลือกคอลัมน์ที่ต้องการ Sort
- Fields to Sort By:** เลือกคอลัมน์ที่ต้องการ Sort

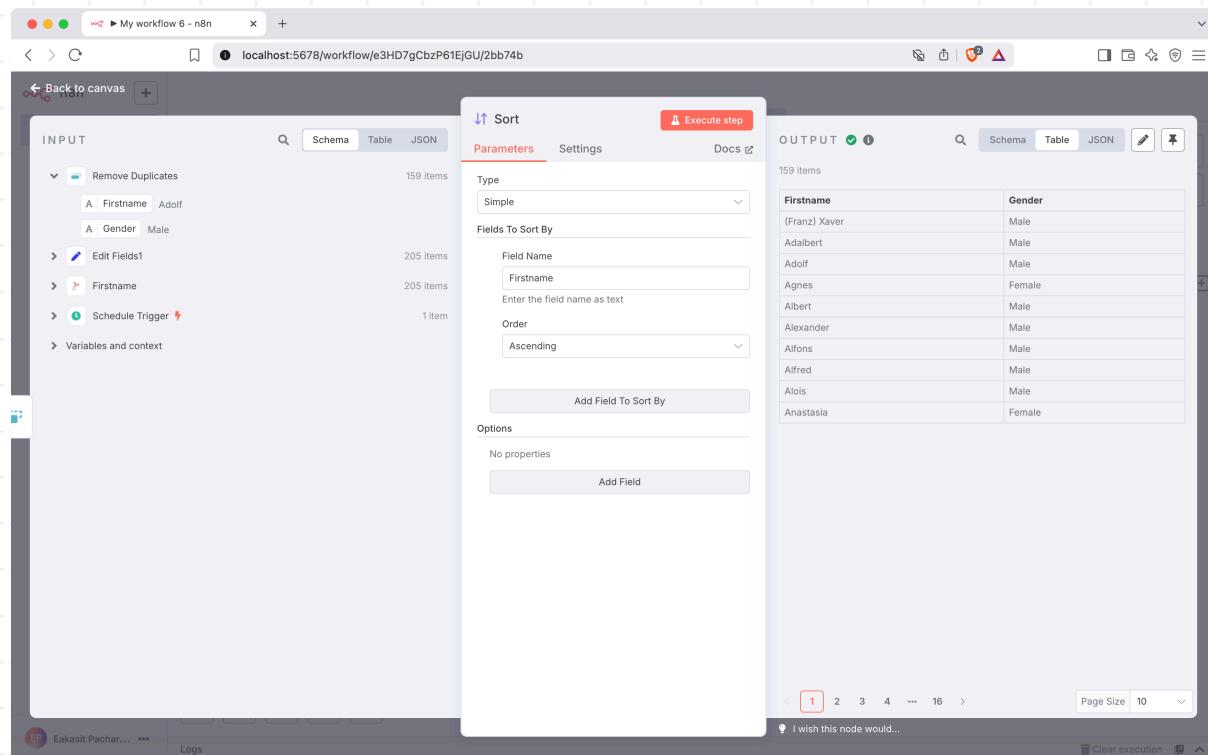



# Example 8

## Sort

### Example 8.1

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- เรียงลำดับตามชื่อ Firstname จากน้อยไปมาก (A to Z)

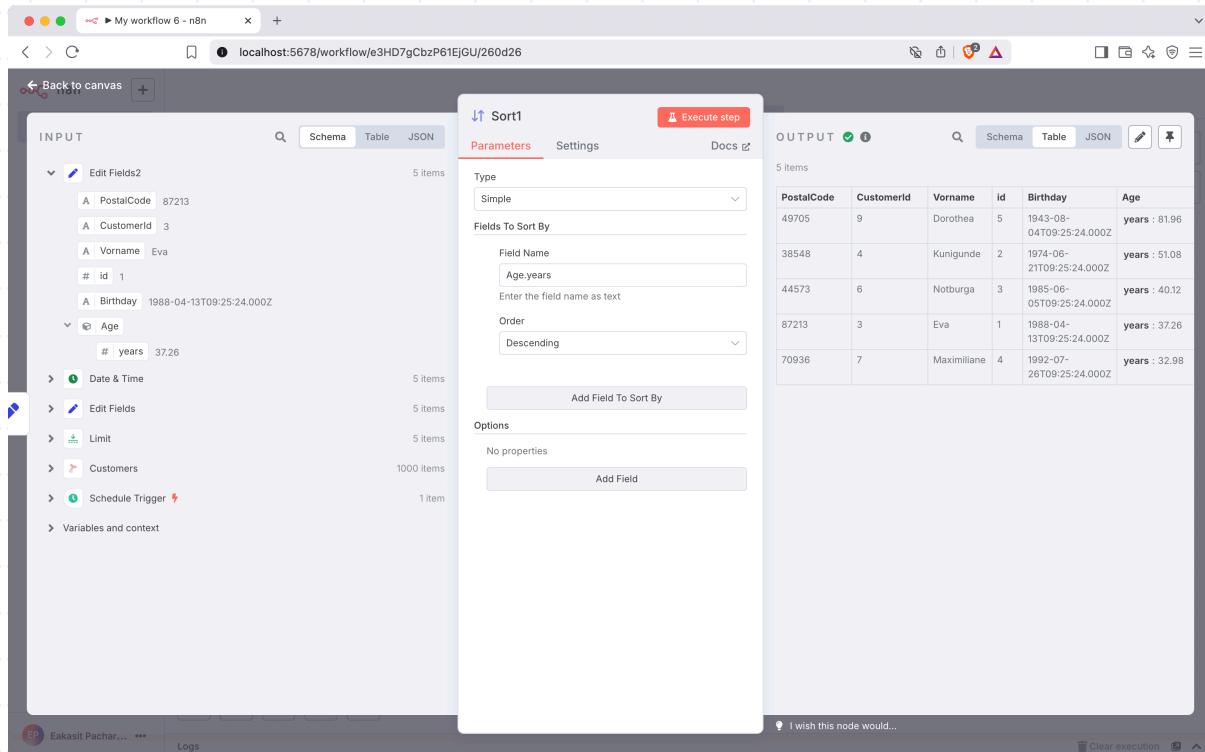


The screenshot shows a workflow titled "My workflow 6 - n8n" running on localhost:5678. The central node is a "Sort" step, which is currently active. The "INPUT" section shows a "Remove Duplicates" node with 159 items, and a "Firstname" node with 205 items. The "OUTPUT" section displays a table with 159 items, sorted by "Firstname" in ascending order. The table includes columns for "Firstname" and "Gender".

Firstname	Gender
(Franz) Xaver	Male
Adalbert	Male
Adolf	Male
Agnes	Female
Albert	Male
Alexander	Male
Alfons	Male
Alfred	Male
Alois	Male
Anastasia	Female

## Example 8.2

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- เรียงข้อมูลตามอายุจากมากไปน้อย



PostalCode	CustomerId	Vorname	id	Birthday	Age
49705	9	Dorothea	5	1943-08-04T09:25:24.000Z	years : 81.96
38548	4	Kunigunde	2	1974-06-21T09:25:24.000Z	years : 51.08
44573	6	Notburga	3	1985-06-05T09:25:24.000Z	years : 40.12
87213	3	Eva	1	1988-04-13T09:25:24.000Z	years : 37.26
70936	7	Maximiliane	4	1992-07-26T09:25:24.000Z	years : 32.98

## Data Preparation with n8n

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ตั้งเวลาให้ workflow ทำงาน

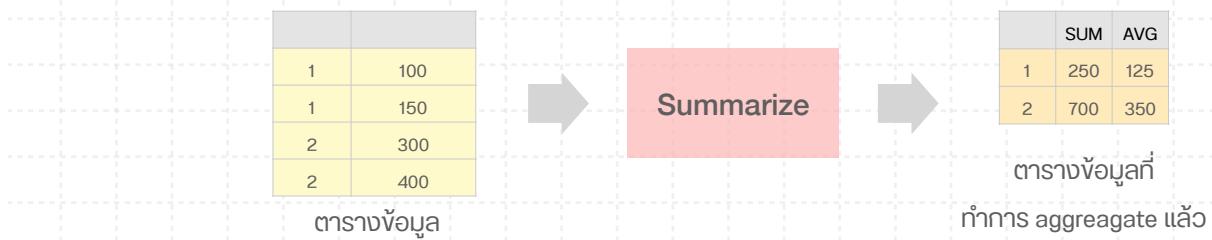
- อ่านข้อมูลจากฐานข้อมูล (database)

- Data Preparation

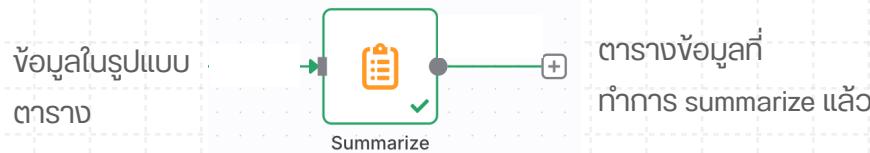
- Filter
- Limit
- Select Fields
- Edit Fields
- Remove Duplicates
- Sort
- Summarize**
- Merge (Join)

# Summarize

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ใช้ในการ group ข้อมูลในตารางข้อมูลและสร้างคอลัมน์ใหม่



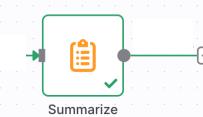
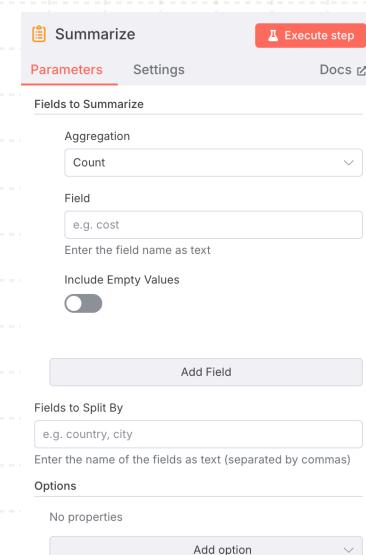
- พบด้วยว่า **Summarize**



# Summarize

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ใช้ในการ group ข้อมูลในตารางข้อมูลและสร้างคอลัมน์ใหม่

- Fields to Summarize:**
  - Aggregation:** ฟังก์ชันในการ summarize
  - Field:** คอลัมน์ที่ต้องการ summarize
- Fields to Split By:** เลือกคอลัมน์ที่ต้องการ Group

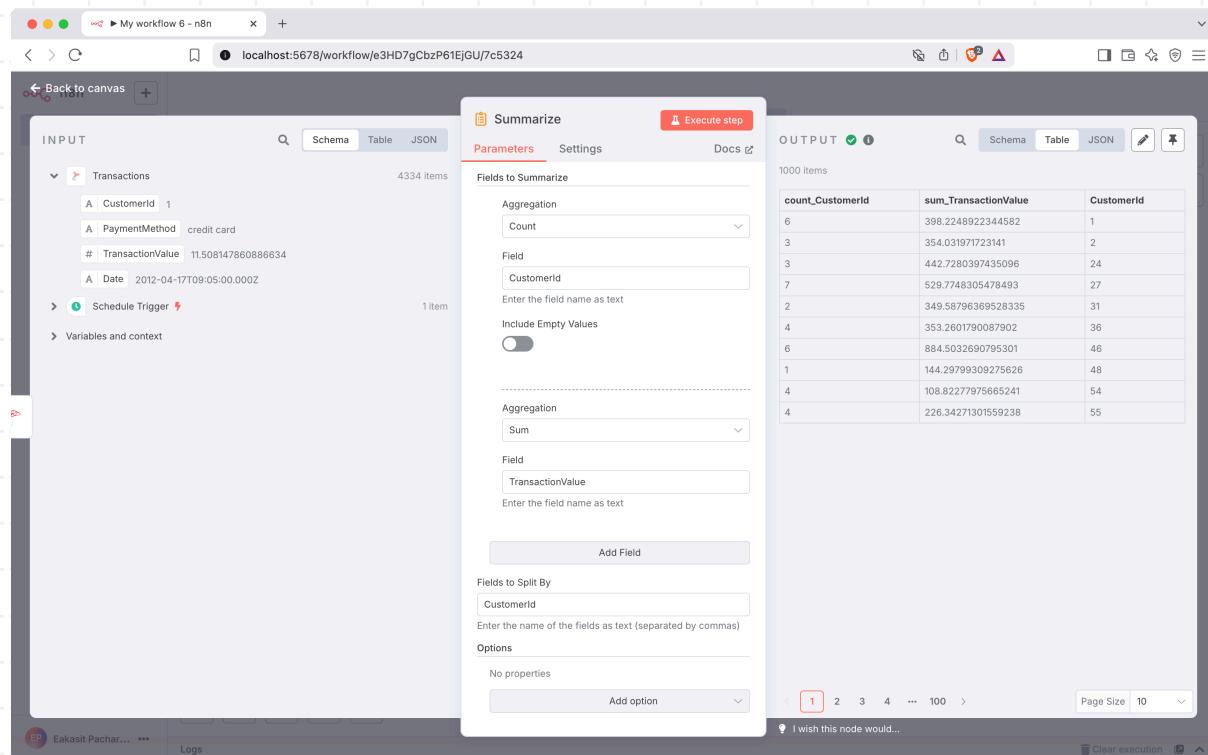
The screenshot shows the configuration for the "Summarize" step. At the top, there are tabs for "Parameters", "Settings", and "Docs". The "Parameters" tab is selected. It includes fields for "Fields to Summarize" (with "Aggregation" set to "Count" and "Field" set to "cost"), "Fields to Split By" (with "e.g. country, city" entered), and "Options" (with "No properties"). A red "Execute step" button is visible at the top right.

# Example 9

Summarize

## Example 9.1

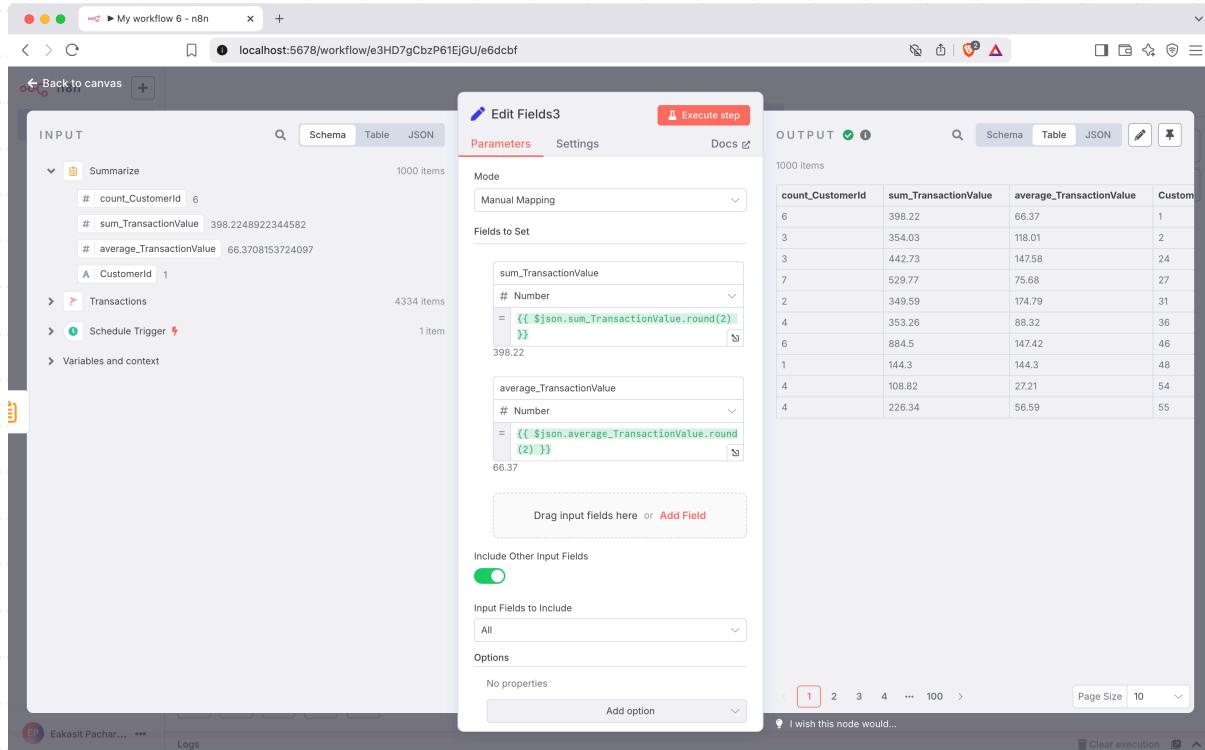
- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- สรุปจำนวนครั้งการใช้บริการ ยอดการใช้จ่ายโดยรวมและเฉลี่ย ของลูกค้าแต่ละคน



count_CustomerId	sum_TransactionValue	CustomerId
6	398.2248922344582	1
3	354.031971723141	2
3	442.7280397435096	24
7	529.748305478493	27
2	349.58796369528335	31
4	353.2601790087902	36
6	884.5032690795301	46
1	144.29799309275626	48
4	108.82277975665241	54
4	226.34271301559238	55

## Example 9.2

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- แสดงข้อมูลค่าใช้จ่ายเป็นตัวเลขที่มีศูนย์ 2 ตำแหน่ง



count_CustomerId	sum_TransactionValue	average_TransactionValue	Custom
6	398.22	66.37	1
3	354.03	118.01	2
3	442.73	147.58	24
7	529.77	75.68	27
2	349.59	174.79	31
4	353.26	88.32	36
6	884.5	147.42	46
1	144.3	144.3	48
4	108.82	27.21	54
4	226.34	56.59	55

## Data Preparation with n8n

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- ตั้งเวลาให้ workflow ทำงาน
  - อ่านข้อมูลจากฐานข้อมูล (database)
  - Data Preparation
    - Filter
    - Limit
    - Select Fields
    - Edit Fields
    - Remove Duplicates
    - Sort
    - Summarize
    - Merge (Join)**

# Merge (Join)

- สำหรับการบรรยายในหลักสูตร M.Sc. in Artificial Intelligence for Business Analytics, KMITL
- เชื่อมโยงข้อมูล 2 ตารางให้เป็นตารางเดียวกัน (join)

ID	Name
1	Ben
2	Eve
3	Maria
4	Sam
5	Nick

Name	Gender
Ben	Male
Eve	Female
Maria	Female
Nick	Male
Sam	Male

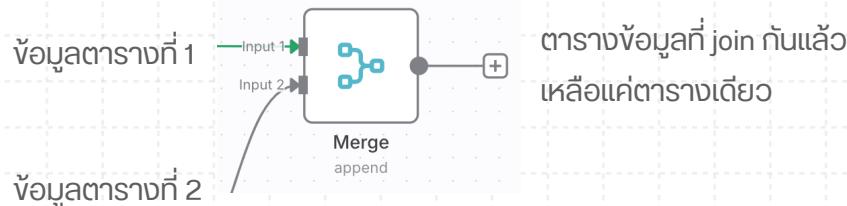


ID	Name	Gender
1	Ben	Male
2	Eve	Female
3	Maria	Female
4	Sam	Male
5	Nick	Male

ตารางข้อมูลลูกค้า      ตารางข้อมูลชื่อและเพศ

ตารางข้อมูลที่ join กันแล้ว

- พบดีที่ใช้ชื่อว่า **Merge**

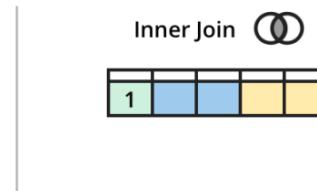


# Merge (Join)

- สำหรับการบรรยายในหลักสูตร M.Sc. in Artificial Intelligence for Business Analytics, KMITL
- ในการ Join มีหลาย Type ได้แก่

Table 1	
1	
2	

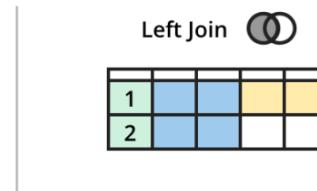
Table 2	
1	
3	
4	



Inner Join	
1	

Table 1	
1	
2	

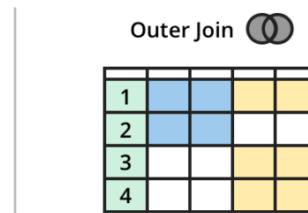
Table 2	
1	
3	
4	



Left Join	
1	
2	

Table 1	
1	
2	

Table 2	
1	
3	
4	



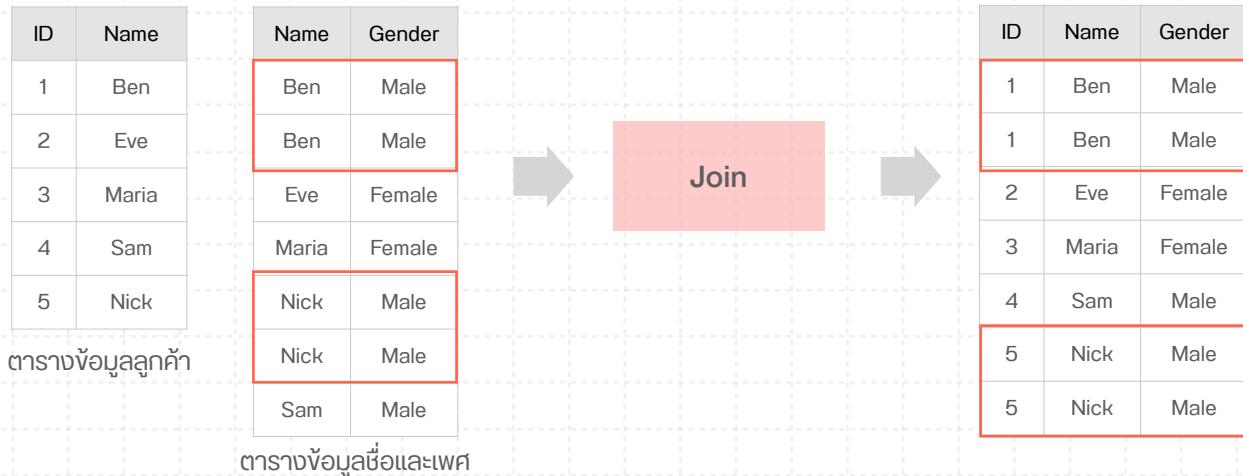
Outer Join	
1	
2	
3	
4	

# Merge (Join)

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- ข้อควรระวัง

- ด้วยข้อมูลที่ต้องการ Join มีมากกว่า 1 ແவງจะทำให้ข้อมูลหลังจากการ Join มีจำนวนมากขึ้น

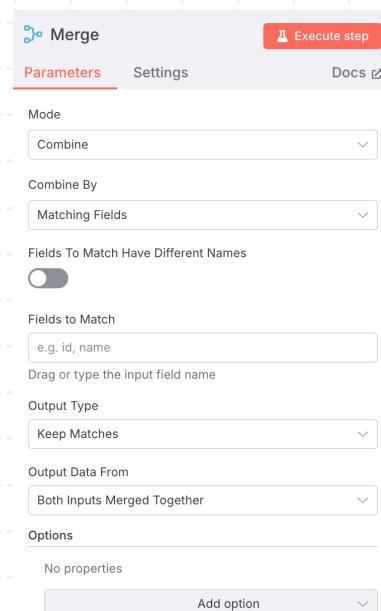
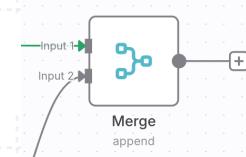


# Merge (Join)

สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL

- เชื่อมโยงข้อมูล 2 ตารางให้เป็นตารางเดียวกัน (join)

- Mode:** เลือกได้ว่าจะเป็น Append หรือ Join
- Combine By:** เลือกว่าจะ Join ด้วยชื่อคอลัมน์หรือลำดับ
- Fields To Match Have Different Names:** เลือกด้วยชื่อคอลัมน์ที่ต้องการ Join ไม่ตรงกับชื่อคอลัมน์ที่ต้องการ Join
- Fields to Match:** ระบุคอลัมน์ที่ใช้ในการ Join
- Output Type:**
  - Keep Matches:** แสดงเฉพาะข้อมูลที่มีกับ 2 ตาราง (Inner Join)
  - Keep Non-matches:** แสดงเฉพาะข้อมูลที่ไม่มีกับ 2 ตาราง
  - Keep Everything:** แสดงข้อมูลทั้งหมดจากทั้ง 2 ตาราง (Outer Join)



# Merge (Join)

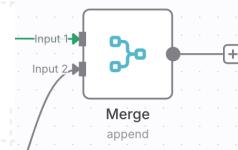
- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- เชื่อมโยงข้อมูล 2 ตาราง**ให้เป็นตารางเดียวกัน (join)

- **Output Type:**

- **Enrich Input 1:** แสดงข้อมูลจากตารางที่ 1 เป็นหลัก (Left Join)
- **Enrich Input 2:** แสดงข้อมูลจากตารางที่ 2 เป็นหลัก (Right Join)

- **Output Data From:**

- **Both Inputs Merge Together:** แสดงข้อมูลจากทั้ง 2 ตาราง
- **Input 1:** แสดงข้อมูลเฉพาะจากตารางที่ 1
- **Input 2:** แสดงข้อมูลเฉพาะจากตารางที่ 2



**Merge** Execute step

**Parameters** **Settings** **Docs**

**Mode**: Combine

**Combine By**: Matching Fields

**Fields To Match Have Different Names**:

**Fields to Match**: e.g. id, name

**Drag or type the input field name**

**Output Type**: Keep Matches

**Output Data From**: Both Inputs Merged Together

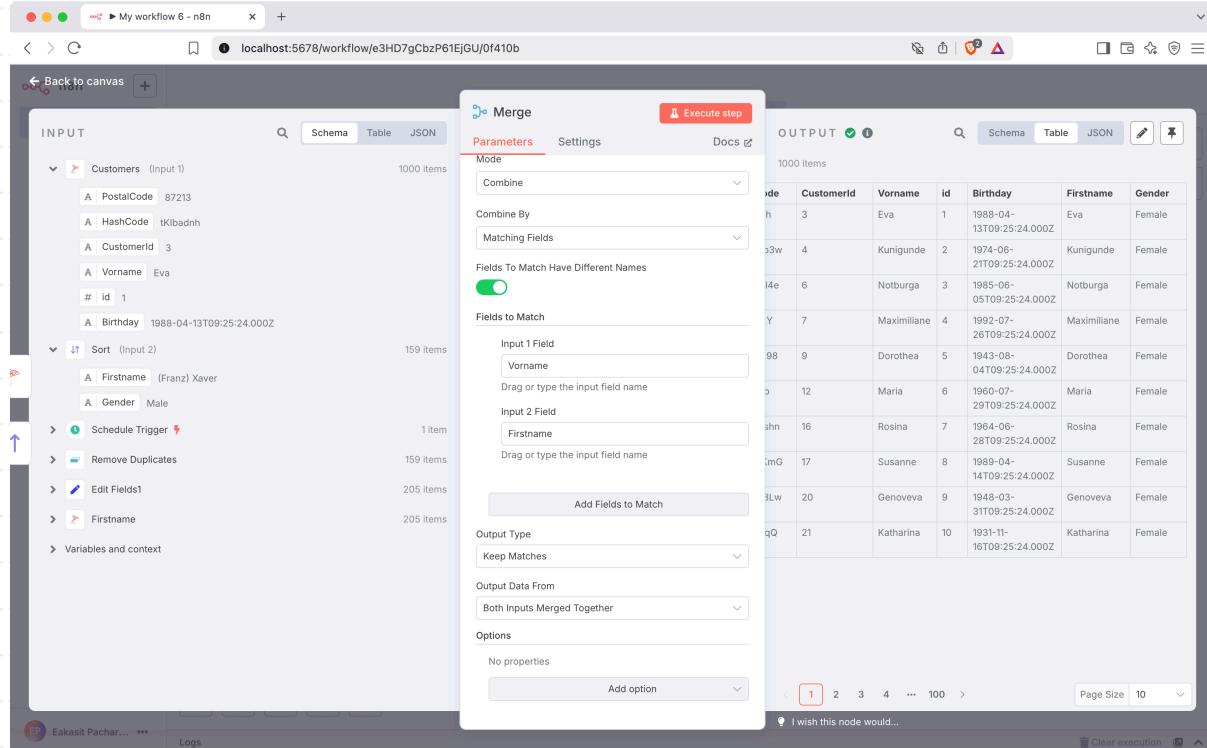
**Options**: No properties Add option

## Example 10

Marge (Join)

# Example 10

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- แสดงข้อมูลค่าใช้จ่ายและเพศ (gender) ของลูกค้าแต่ละคน



**INPUT**

- Customers (Input 1)
  - A PostalCode 87213
  - A HashCode tkIbadnh
  - A CustomerId 3
  - A Vorname Eva
  - # id 1
  - A Birthday 1988-04-13T09:25:24.000Z
- Sort (Input 2)
  - A Fisrtname (Franz) Xaver
  - A Gender Male
- Schedule Trigger
- Remove Duplicates
- Edit Fields1
- Fisrtname
- Variables and context

**Merge**

**Parameters**

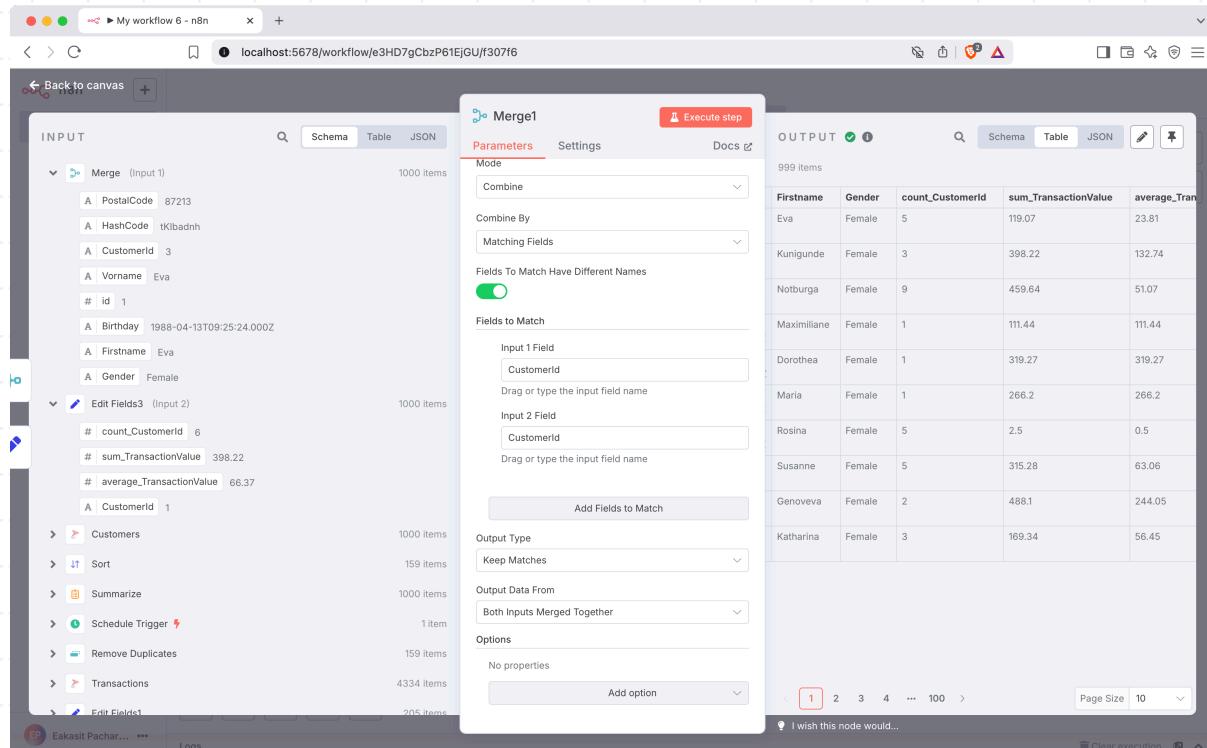
- Mode: Combine
- Combine By: Matching Fields
- Fields To Match Have Different Names: On
- Fields to Match:
  - Input 1 Field: Vorname
  - Input 2 Field: Fisrtname
- Add Fields to Match
- Output Type: Keep Matches
- Output Data From: Both Inputs Merged Together
- Options: No properties

**OUTPUT**

CustomerID	Vorname	id	Birthday	Fisrtname	Gender
3	Eva	1	1988-04-13T09:25:24.000Z	Eva	Female
4	Kunigunde	2	1974-06-21T09:25:24.000Z	Kunigunde	Female
6	Notburga	3	1995-06-05T09:25:24.000Z	Notburga	Female
7	Maximiliane	4	1992-07-26T09:25:24.000Z	Maximiliane	Female
9	Dorothea	5	1943-08-04T09:25:24.000Z	Dorothea	Female
12	Maria	6	1960-07-29T09:25:24.000Z	Maria	Female
16	Rosina	7	1964-06-28T09:25:24.000Z	Rosina	Female
17	Susanne	8	1989-04-14T09:25:24.000Z	Susanne	Female
20	Genoveva	9	1948-03-31T09:25:24.000Z	Genoveva	Female
21	Katharina	10	1931-11-16T09:25:24.000Z	Katharina	Female

# Example 10

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- แสดงข้อมูลค่าใช้จ่ายและเพศ (gender) ของลูกค้าแต่ละคน



**INPUT**

- Merge (Input 1)
  - A PostalCode 87213
  - A HashCode tkIbadnh
  - A CustomerId 3
  - A Vorname Eva
  - # id 1
  - A Birthday 1988-04-13T09:25:24.000Z
  - A Fisrtname Eva
  - A Gender Female
- Edit Fields3 (Input 2)
  - # count\_CustomerId 6
  - # sum\_TransactionValue 398.22
  - # average\_TransactionValue 66.37
  - A CustomerId 1
- Customers
- Sort
- Summarize
- Schedule Trigger
- Remove Duplicates
- Transactions
- Edit Field1
- Eakasit Pachar...
- Logs

**Merge1**

**Parameters**

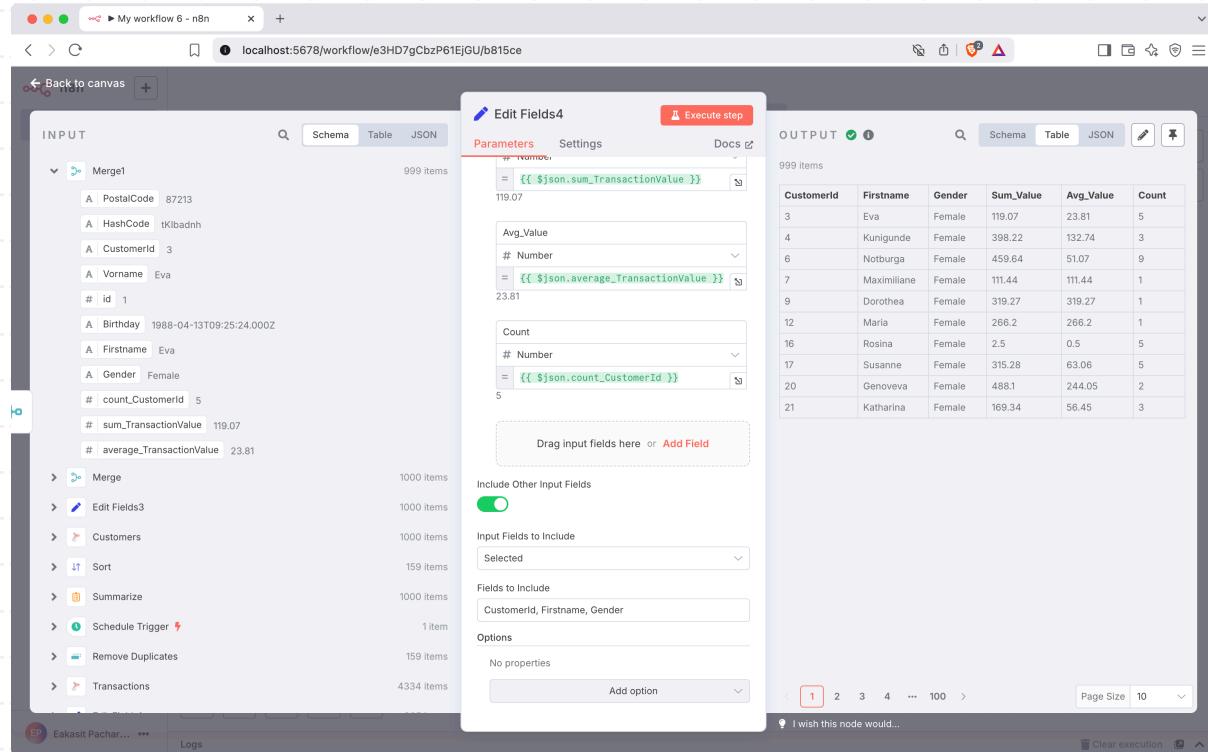
- Mode: Combine
- Combine By: Matching Fields
- Fields To Match Have Different Names: On
- Fields to Match:
  - Input 1 Field: CustomerId
  - Input 2 Field: CustomerId
- Add Fields to Match
- Output Type: Keep Matches
- Output Data From: Both Inputs Merged Together
- Options: No properties

**OUTPUT**

Fisrtname	Gender	count_CustomerId	sum_TransactionValue	average_TransactionValue
Eva	Female	5	119.07	23.81
Kunigunde	Female	3	398.22	132.74
Notburga	Female	9	459.64	51.07
Maximiliane	Female	1	111.44	111.44
Dorothea	Female	1	319.27	319.27
Maria	Female	1	266.2	266.2
Rosina	Female	5	2.5	0.5
Susanne	Female	5	315.28	63.06
Genoveva	Female	2	488.1	244.05
Katharina	Female	3	169.34	56.45

# Example 10

- สำหรับการบรรยายในหลักสูตร M. Sc. in Artificial Intelligence for Business Analytics, KMITL
- แสดงข้อมูลค่าใช้จ่ายและเพศ (gender) ของลูกค้าแต่ละคน



CustomerID	Firstname	Gender	Sum_Value	Avg_Value	Count
3	Eva	Female	119.07	23.81	5
4	Kunigunde	Female	398.22	132.74	3
6	Notburga	Female	459.64	51.07	9
7	Maximilliane	Female	111.44	111.44	1
9	Dorothea	Female	319.27	319.27	1
12	Maria	Female	266.2	266.2	1
16	Rosina	Female	2.5	0.5	5
17	Susanne	Female	315.28	63.06	5
20	Genoveva	Female	488.1	244.05	2
21	Katharina	Female	169.34	56.45	3