datacube | CUBE Analytics CONSULTING | CubeSpace Learning • Working • Sharing

ตัวอย่างการใช้งาน RapidMiner TurboPrep

Eakasit Pacharawongsakda, Ph.D.
eakasit@datacubeth.ai
Co-Founder and Data Science Team Lead
Cube Analytics Consulting Co., Ltd.

About us

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- **ชื่อ:** เอกสิทธิ์ พัชรวงศ์ศักดา
- **การศึกษา:**
  - ปริญญาเอก วิทยาการคอมพิวเตอร์
    สถาบันเทคโนโลยีนานาชาติสิรินธร (SIIT) มหาวิทยาลัยธรรมศาสตร์
  - ปริญญาโท วิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเกษตรศาสตร์
  - ปริญญาตรี วิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเกษตรศาสตร์
    (เกียรตินิยมอันดับ 2)
- **ประสบการณ์**
  - Certified RapidMiner Expert and Ambassador
  - Research Collaboration with **Western Digital (Thailand)** เฟสที่ 1 ระยะเวลา 6 เดือน
  - ร่วมวิจัย โครงการสำรวจข้อมูลเพื่อการวิเคราะห์พฤติกรรมของนักท่องเที่ยวเชิงลึก ด้วยวิธีการทำเหมืองข้อมูล **การท่องเที่ยวแห่งประเทศไทย** (ททท)
  - วิทยากรอบรมการใช้งานซอฟต์แวร์ open source ทางด้าน data mining



CERTIFIED ANALYST | CERTIFIED EXPERT | PARTNER AMBASSADOR

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

About us

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

CERTIFIED ANALYST

RapidMiner **Analyst** Certification

This is to Certify that
Eakasit Pacharawongsakda

Successfully passed the examination for the Certified RapidMiner Analyst.

The RapidMiner Analyst certification level is designed for individuals who wish to demonstrate a fundamental understanding of how RapidMiner software works and is used.

Certified Analyst professionals will be able to prepare data and create predictive models in standard data environments typically found within most analyst positions.

The candidate has proven the ability to:

- Prepare data
- Evaluate the model's quality
- Deploy data mining models
- Build predictive models
- Score new data sets

With:

- RapidMiner Studio
- RapidMiner Server

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

About us

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

CERTIFIED EXPERT

rapidminer

Certification

This is to certify that

Eakasit Pacharawongsakda

successfully passed the examination for the

Certified RapidMiner Expert.

The RapidMiner Expert certification level is designed to demonstrate a complete understanding of how RapidMiner Studio and its ecosystem works and is used. Certified professionals will be able to use RapidMiner Studio to analyze Text and Web Information, deploy solutions to the RapidMiner Server, create reports and integrate with Hadoop platforms via RapidMiner Radoop. The candidate has proven the ability to,

- Perform Text and Web Mining
- Deploy and manage a RapidMiner Server
- Create and deploy RapidMiner WebApps
- Develop and integrate RapidMiner dashboard elements
- Connect RapidMiner Studio and Server to a Hadoop platform
- Perform machine learning tasks using the Radoop operators

with

- RapidMiner 8

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

About us

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

rapidminer

This is to certify that

Eakasit Pacharawongsakda

successfully passed the examination for

Applications & Use Cases Professional Certification

The Professional level is the first of two certification levels and is rewarded for solid experience in the topic. Some of the topics tested at this level are:

Introduction to Machine Learning and Data Science, **CRISP-DM and Variants**, **Use Cases for Machine Learning**, **Visualization** and **What to do with Models.**

This certificate can be verified using the open badge which was issued to the certificate holder.
The open badge can be uploaded and tested via this redirect: www.rapidminer.com/badgecheck

Knut Makowski
(Head of Education)

Peter Lee
(CEO RapidMiner)

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

CUBE Analytics CONSULTING

About us

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

rapidminer

This is to certify that

Eakasit Pacharawongsakda

successfully passed the examination for

Applications & Use Cases Master Certification

The Master level is the second level after Professional and stands for high proficiency in the topic. Some of the topics tested at this level are:

Running Processes, **Deploying Models**, **Model Management** and **Web Apps**.

This certificate can be verified using the open badge which was issued to the certificate holder.
The open badge can be uploaded and tested via this redirect: www.rapidminer.com/badgecheck

Knut Makowski
(Head of Education)

Peter Lee
(CEO RapidMiner)

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

About us

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

rapidminer

This is to certify that

Eakasit Pacharawongsakda

successfully passed the examination for

Data Engineering Professional Certification



The Professional level is the first of two certification levels and is rewarded for solid experience in the topic. Some of the topics tested at this level are:

Data Access, **Basic Transformations**, **Working with multiple Data Sets**, **Pivot Tables**, **Routines** and **Simple Text Processing**.

This certificate can be verified using the open badge which was issued to the certificate holder.
The open badge can be uploaded and tested via this redirect: www.rapidminer.com/badgecheck

Knut Makowski
(Head of Education)

Peter Lee
(CEO RapidMiner)

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

CUBE Analytics CONSULTING

About us

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

rapidminer

This is to certify that

Eakasit Pacharawongsakda

successfully passed the examination for

Data Engineering Master Certification



The Master level is the second level after Professional and stands for high proficiency in the topic. Some of the topics tested at this level are:

Loops and Branches, **Advanced Text Processing**, **Exception Handling**, **Logging**, **Data Cleansing**, **Regular Expressions**, **Macros**, **Web APIs**, and **Scripting**.

This certificate can be verified using the open badge which was issued to the certificate holder.
The open badge can be uploaded and tested via this redirect: www.rapidminer.com/badgecheck

Knut Makowski
(Head of Education)

Peter Lee
(CEO RapidMiner)

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

http://facebook.com/datacube.th

About us

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

rapidminer

This is to certify that

Eakasit Pacharawongsakda

successfully passed the examination for

Machine Learning Professional Certification

The Professional level is the first of two certification levels and is rewarded for solid experience in the topic. Some of the topics tested at this level are:

Classification, **Regression**, **Scoring**, **Hold-out Validation**, **Correlations**, **Feature Importance**, **Clustering** and **Association Rules**.



This certificate can be verified using the open badge which was issued to the certificate holder.
The open badge can be uploaded and tested via this redirect: www.rapidminer.com/badgecheck

Knut Makowski
(Head of Education)

Peter Lee
(CEO RapidMiner)

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

CUBE Analytics CONSULTING

About us

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

rapidminer

This is to certify that

Eakasit Pacharawongsakda

successfully passed the examination for

Machine Learning Master Certification

The Master level is the second level after Professional and stands for high proficiency in the topic. Some of the topics tested at this level are:

Complex Predictive Models, **Cross Validation and Correct Validation**, **Parameter Optimization**, **Model Selection**, **Feature Engineering**, **Time Series and Forecasting** and **Integrating R and Python models.**



This certificate can be verified using the open badge which was issued to the certificate holder.
The open badge can be uploaded and tested via this redirect: www.rapidminer.com/badgecheck

Knut Makowski
(Head of Education)

Peter Lee
(CEO RapidMiner)

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

About us

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

rapidminer

This is to certify that

Eakasit Pacharawongsakda

successfully passed the examination for

Platform Administration Master Certification

The Master level stands for high proficiency in the topic. Some of the topics tested at this level are:

Platform Overview, **RapidMiner Studio Installation**, **RapidMiner Server Administration**, **RapidMiner Radoop**, **RapidMiner Real-Time Scoring**, and **RapidMiner Marketplace**.

PLATFORM ADMINISTRATION MASTER

This certificate can be verified using the open badge which was issued to the certificate holder.
The open badge can be uploaded and tested via this redirect: www.rapidminer.com/badgecheck

Knut Makowski
(Head of Education)

Peter Lee
(CEO RapidMiner)

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

About Data Cube

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

Founded on

2013

Focused on training/implementation related to

Data Science / AI Machine Learning / Big Data

100+

public/in-house training classes

47,000+

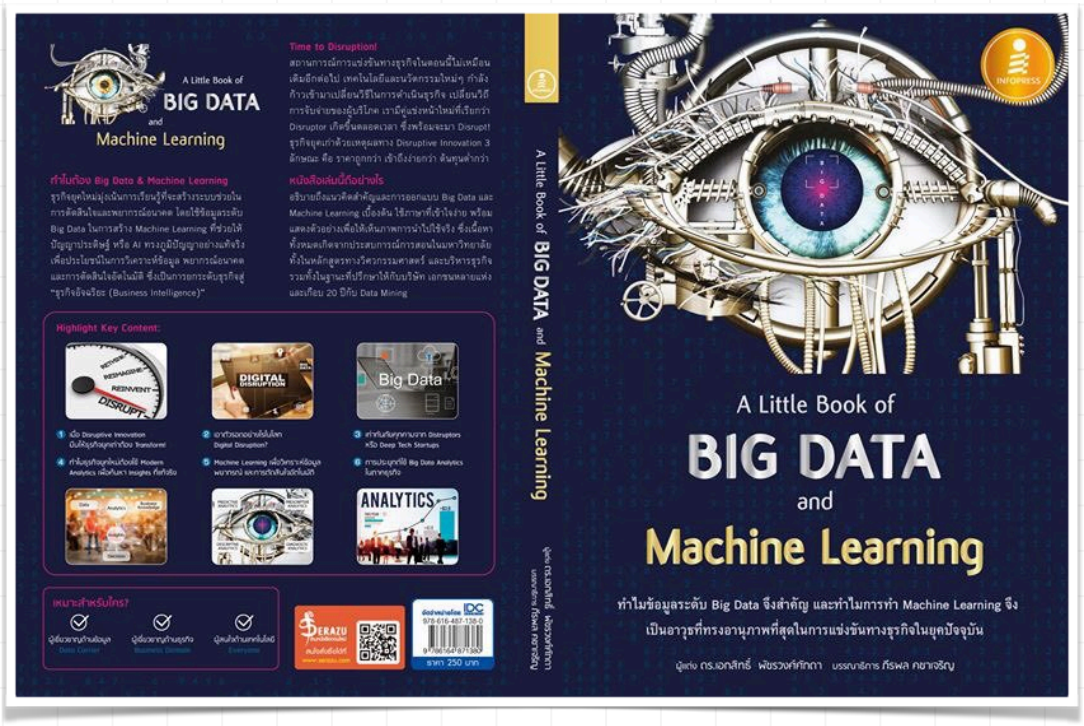
people like Facebook page

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

About us

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- หนังสือ Big Data และ Machine Learning ฉบับภาษาไทย



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

About us

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- หนังสือ Data Mining ฉบับภาษาไทย



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลแสดงในรูปแบบของตาราง
  - แถว เรียกว่า **ตัวอย่าง** (example)
  - คอลัมน์ เรียกว่า **แอตทริบิวต์** (attribute) ซึ่มีหน้าที่ (role) ที่ใช้งานบ่อย 3 แบบ
    - **ไอดี (ID)** เป็นแอตทริบิวต์ที่แสดงหมายเลขของข้อมูล หรือ primary key ในฐานข้อมูล
    - **แอตทริบิวต์ทั่วไป (attribute)** เป็นแอตทริบิวต์ปกติที่จะใช้ในการสร้างโมเดลหรือเรียกว่าเป็น ฟีเจอร์ (feature) หรือตัวแปรต้น (independent variable)
    - **ลาเบล (label)** เป็นแอตทริบิวต์ขนิดพิเศษที่มักจะใช้แสดงคำตอบของสิ่งที่เราต้องการจะสร้างโมเดลมาทำนาย หรือ เรียกว่า คลาส (class) หรือตัวแปรตาม (dependent

customer_id	age	gender	region	income	married	children	car	response
ID12101	48	FEMALE	INNER_CITY	17546.0	NO	1	NO	NO
ID12102	40	MALE	TOWN	30085.1	YES	3	YES	NO
ID12103	51	FEMALE	INNER_CITY	16575.4	YES	0	YES	YES
ID12104	23	FEMALE	TOWN	20375.4	YES	3	NO	NO

ID — attribute — label

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Value type

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ค่าของข้อมูลที่เก็บในแต่ละ**แอตทริบิวต์**
  - **Polynominal** ข้อมูลประเภท category (ข้อมูลที่ไม่ใช่ตัวเลข) มีค่ามากกว่า 2 ค่าขึ้นไป
  - **Binominal** ข้อมูลประเภท category (ข้อมูลที่ไม่ใช่ตัวเลข) มีค่าเพียง 2 ค่าเท่านั้น
  - **Numeric** หรือ **Integer** ข้อมูลประเภทตัวเลข
  - **Text** ข้อมูลประเภทข้อความ

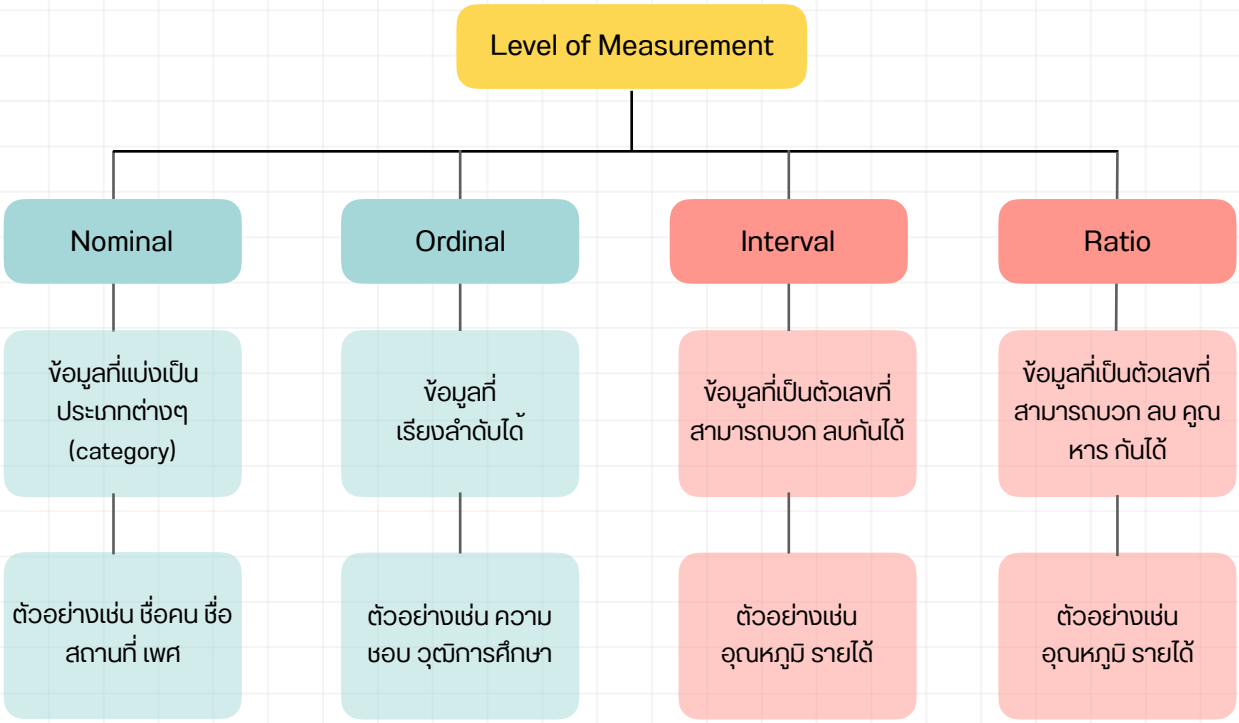
customer_id	age	gender	region	income	married	children	car	response
ID12101	48	FEMALE	INNER_CITY	17546.0	NO	1	NO	NO
ID12102	40	MALE	TOWN	30085.1	YES	3	YES	NO
ID12103	51	FEMALE	INNER_CITY	16575.4	YES	0	YES	YES
ID12104	23	FEMALE	TOWN	20375.4	YES	3	NO	NO

polynominal — numeric — binominal — numeric — binominal — numeric — binominal

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Level of Measurement

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

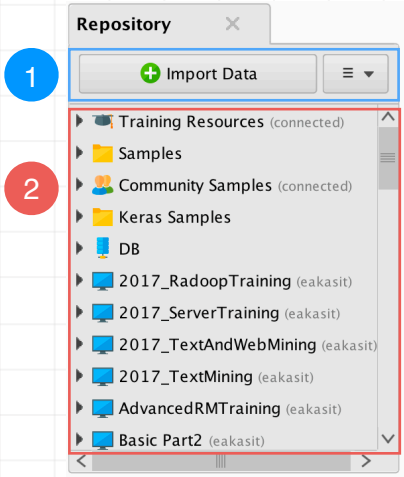


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data management

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- Repository
  - เป็นที่เก็บข้อมูลและ process เพื่อใช้งานใน RapidMiner Studio 9
  - ทำให้ไม่ต้องโหลดข้อมูลจากไฟล์ใหม่ทุกครั้ง
- องค์ประกอบในส่วน Repository
  - ส่วนที่ 1
    - สำหรับสร้าง Repository ใหม่
    - โหลดไฟล์ประเภทต่างๆ เข้าไปไว้ใน Repository
    - สร้างโฟลเดอร์ใหม่
  - ส่วนที่ 2
    - ข้อมูลและ process Sample ที่ RapidMiner Studio 9 เตรียมไว้ให้
    - ข้อมูลที่เก็บอยู่ในแต่ละ Repository

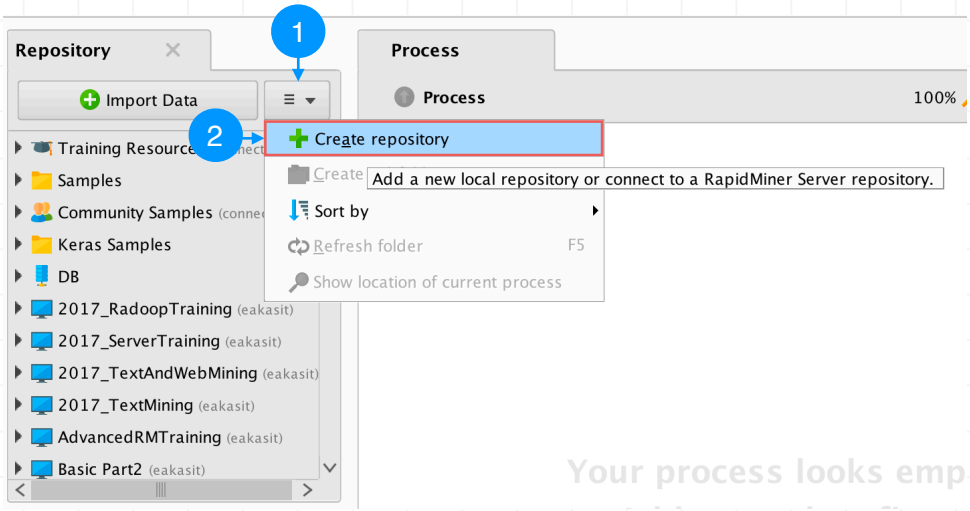


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data management

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- สร้าง Repository ใหม่
  - คลิกที่ ≡
  - หลังจากนั้นเลือกเมนู **Create repository**

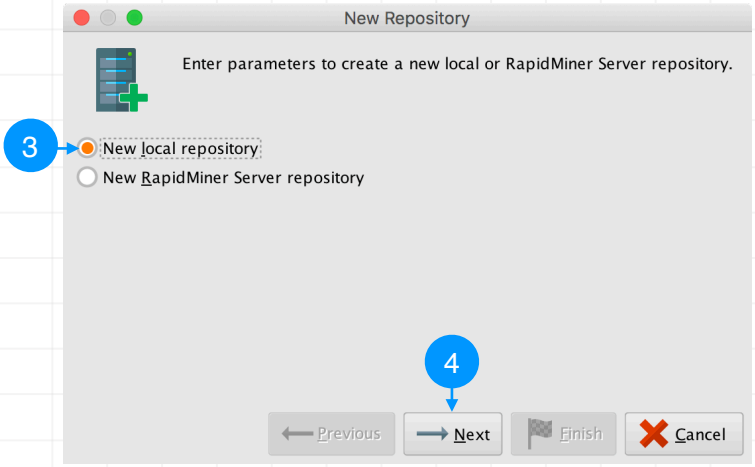


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data management

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- สร้าง Repository ใหม่ (ต่อ)
  - เลือก **New local repository**
  - กดปุ่ม **Next**

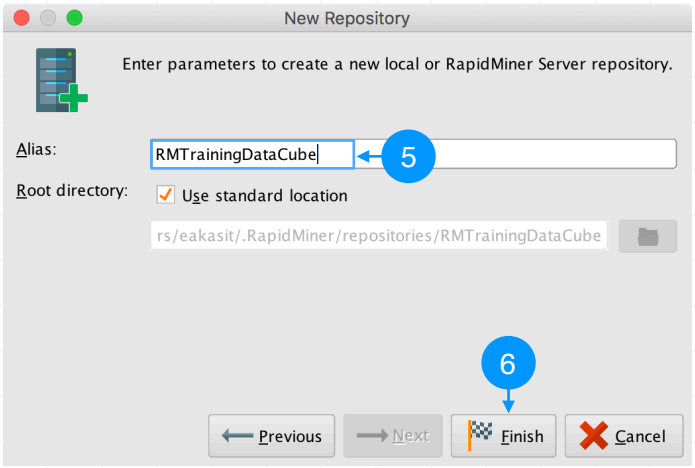


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data management

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- สร้าง Repository ใหม่ (ต่อ)
  - เปลี่ยนชื่อ **Alias** เป็น RMTrainingDataCube
  - กดปุ่ม **Finish**

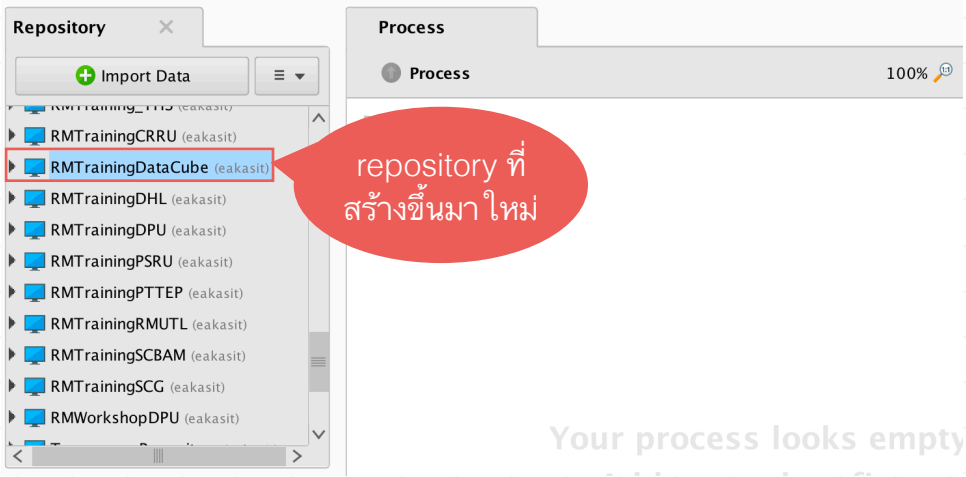


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data management

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- จะปรากฏ repository **RMTrainingDataCube** แสดงขึ้นมาในส่วนของ **Repository**

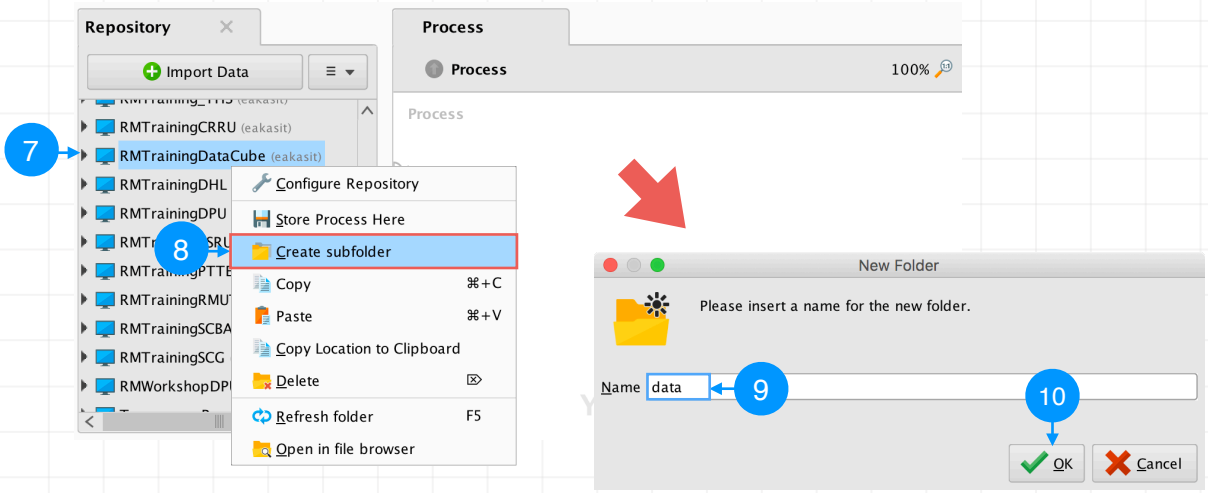


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data management

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- สร้างโฟลเดอร์ใน **RMTrainingDataCube**
  - data สำหรับเก็บข้อมูล

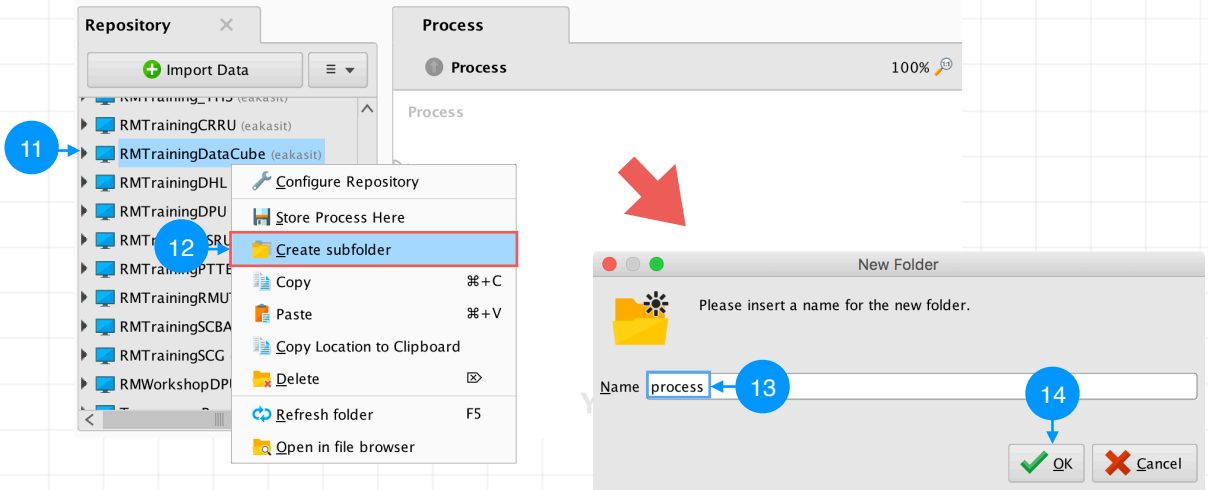


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data management

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- สร้างโฟลเดอร์ใน **RMTrainingDataCube**
  - data สำหรับเก็บข้อมูล
  - process สำหรับเก็บโพรเซสที่สร้างขึ้นมา

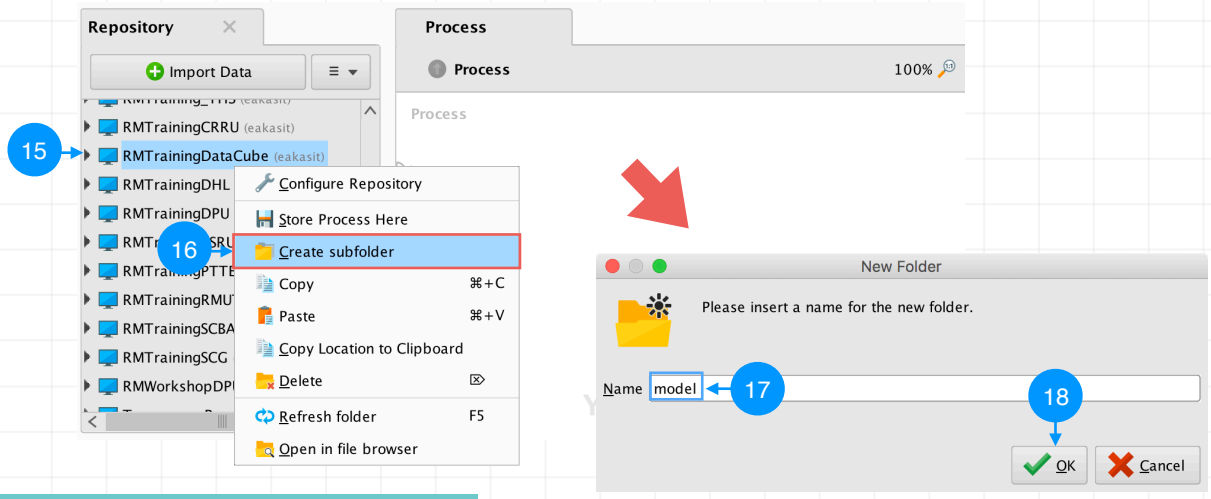


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data management

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- สร้างโฟลเดอร์ใน **RMTrainingDataCube**
  - data สำหรับเก็บข้อมูล
  - process สำหรับเก็บโพรเซสที่สร้างขึ้นมา
  - model สำหรับเก็บโมเดลที่สร้างขึ้นมา



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

CSV File

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ไฟล์ประเภท csv ย่อมาจาก Comma Separated Value
- ใช้เครื่องหมาย , (comma) คั่นระหว่างแอตทริบิวต์

customer_id	age	gender	region	income	married	children	car	response
ID12101	48	FEMALE	INNER_CITY	17546.0	NO	1	NO	NO
ID12102	40	MALE	TOWN	30085.1	YES	3	YES	NO
ID12103	51	FEMALE	INNER_CITY	16575.4	YES	0	YES	YES
ID12104	23	FEMALE	TOWN	20375.4	YES	3	NO	NO

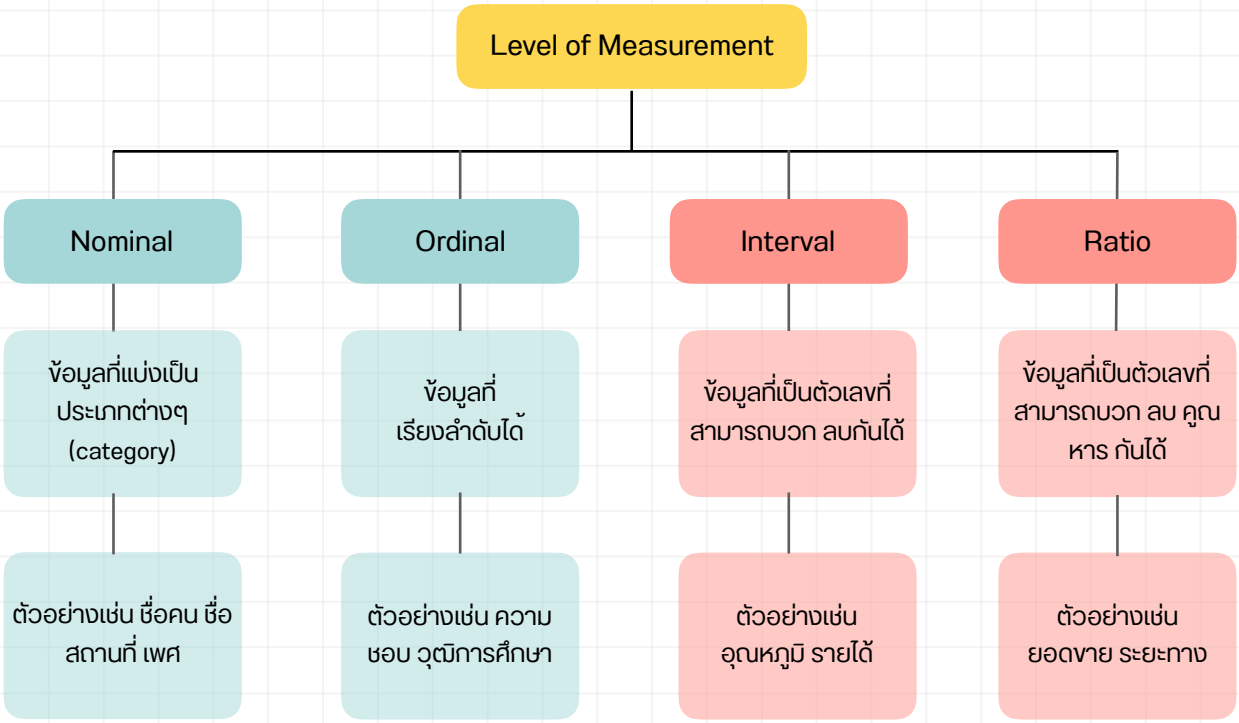
แถวแรกคือ header

```
customer_id,age,gender,region,income,married,children,car,response
ID12101,48,FEMALE,INNER_CITY,17546.0,NO,1,NO,NO
ID12102,40,MALE,TOWN,30085.1,YES,3,YES,NO
ID12103,51,FEMALE,INNER_CITY,16575.4,YES,0,YES,YES
ID12104,23,FEMALE,TOWN,20375.4,YES,3,NO,NO
```

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Level of Measurement

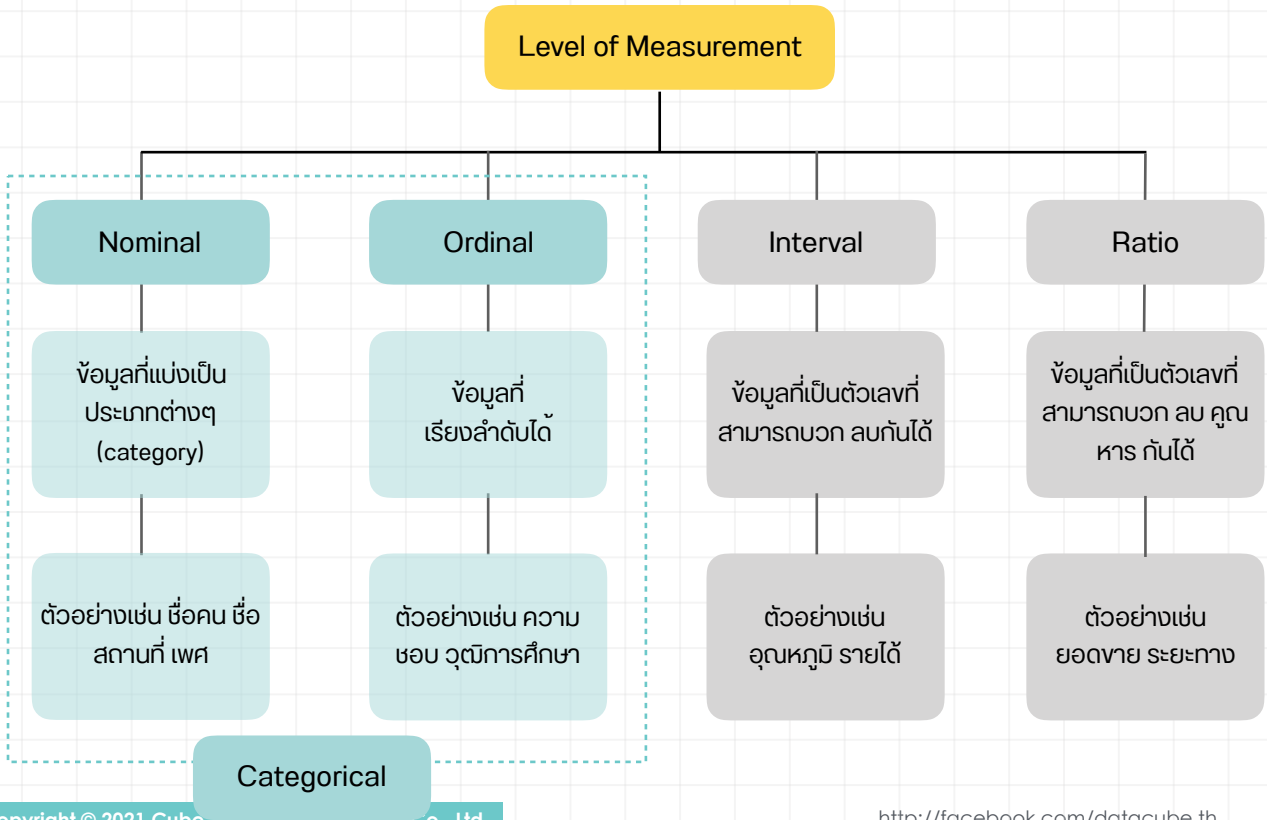
สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Level of Measurement

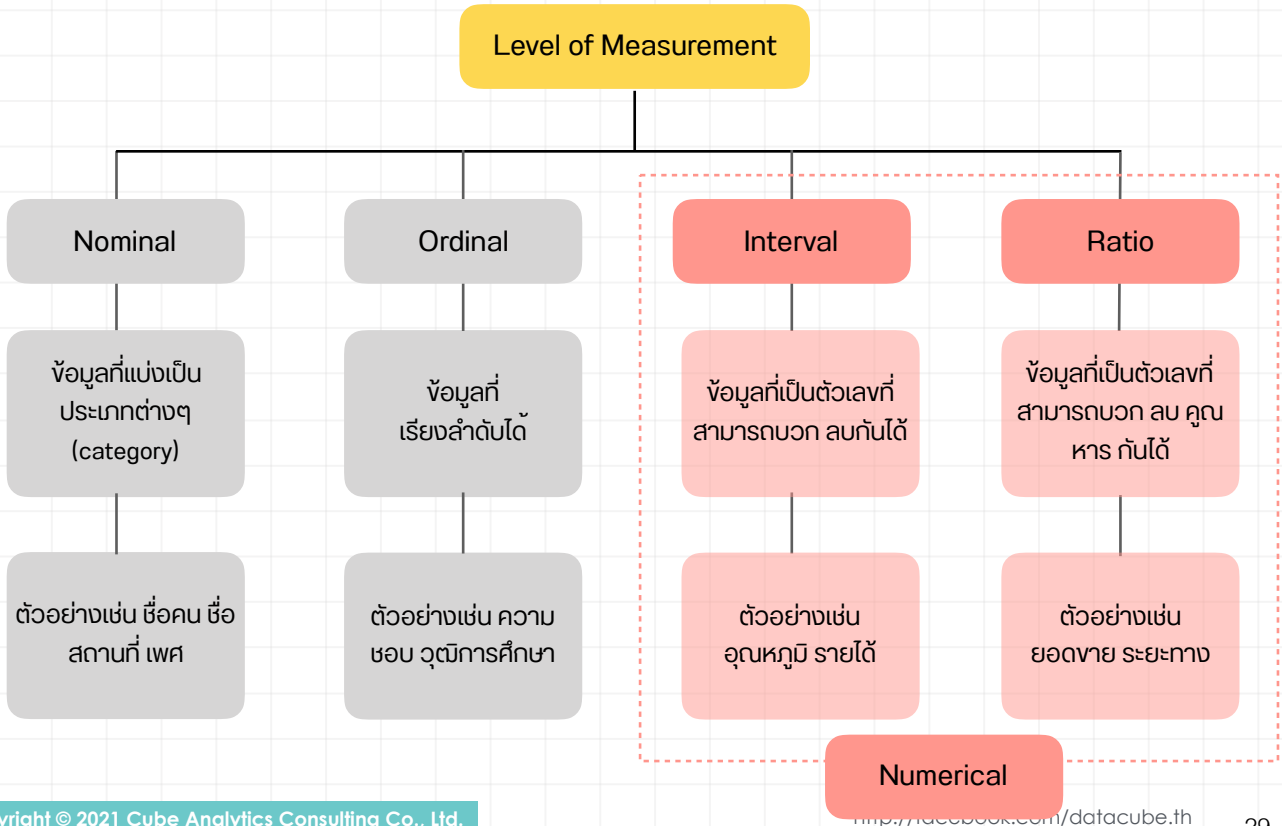
สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Level of Measurement

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

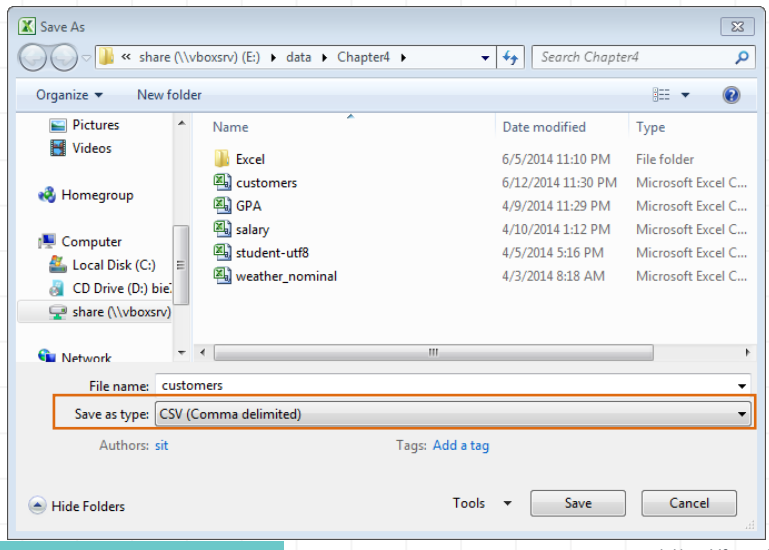


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

CSV File

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ไฟล์ csv สามารถ export ได้จาก Excel หรือ database ต่างๆ
- Export จาก Excel
  - เลือก **File** > **Save As** > **CSV (Comma delimited)**

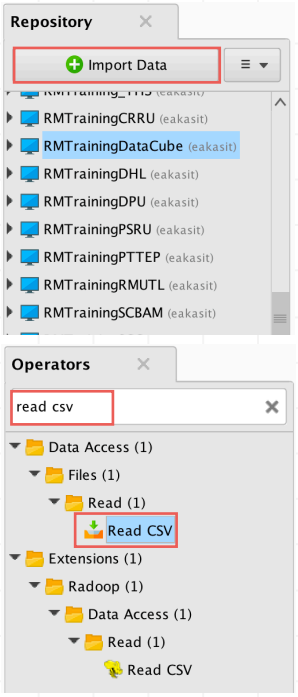


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Load CSV to RapidMiner

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- การโหลดไฟล์ csv เข้าไปใช้ใน RapidMiner Studio 9 ทำได้ 2 แบบ
  - **ใช้การ import ในส่วนของ Repositories**
    - โหลดมาเก็บไว้ใน Repository และใช้งานได้ตลอด
    - ถ้าข้อมูลในไฟล์ csv มีการเปลี่ยนแปลงจะไม่ update ต้องทำการโหลดใหม่
  - ใช้โอเปอเรเตอร์ **Read CSV**
    - โหลดเข้ามาใช้งานโดยการอ่านจากไฟล์ csv ทุกครั้ง เมื่อไฟล์ update ข้อมูลจะเปลี่ยนตาม



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

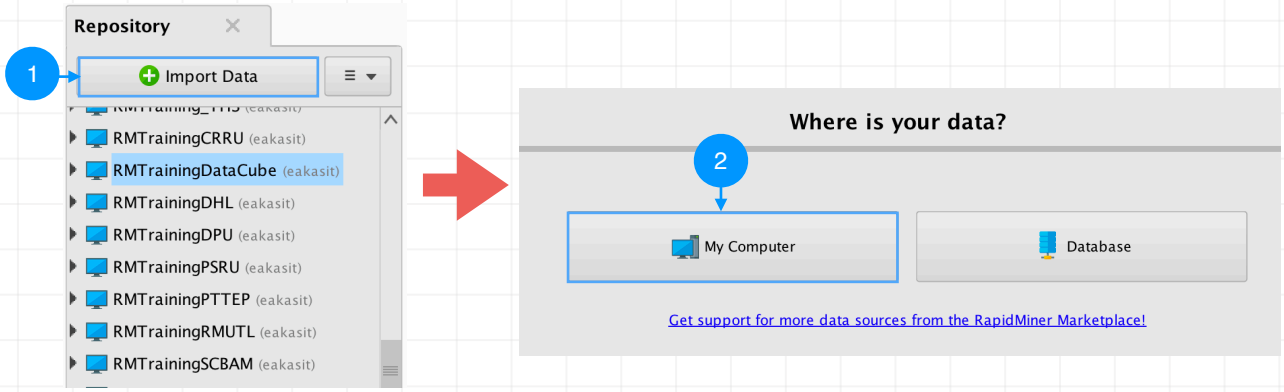
Import CSV

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ csv เข้าไปไว้ใน Repository
  - ในส่วน **Repositories** คลิกที่ปุ่ม



  - เลือกอ่านข้อมูลจาก Computer

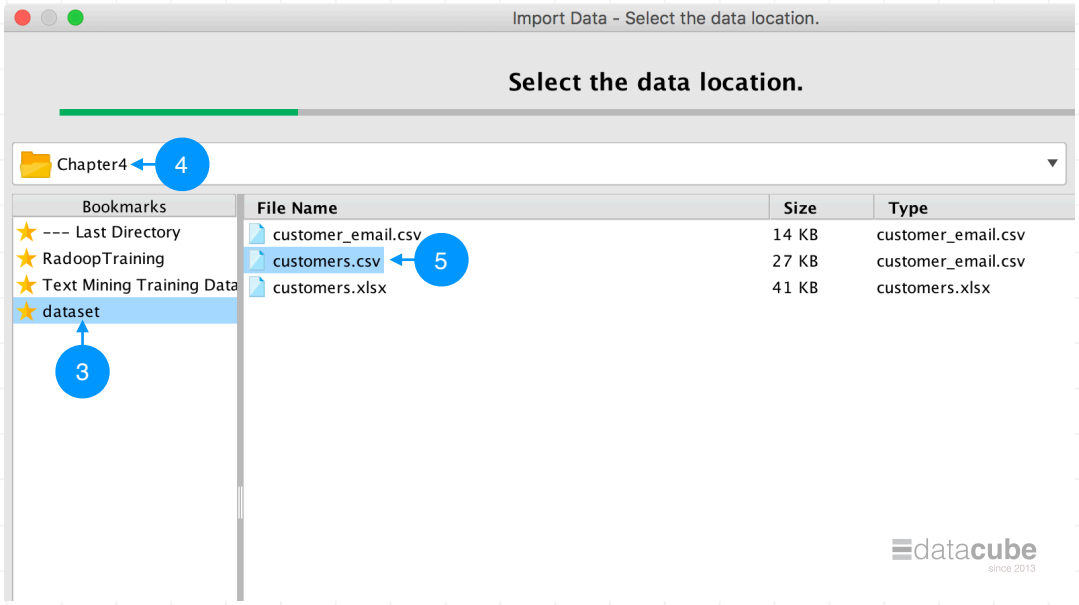


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import CSV

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ csv เข้าไปไว้ใน Repository
  - เลือกไฟล์ customers.csv จากโฟลเดอร์ dataset/Chapter4 และกด **Next**

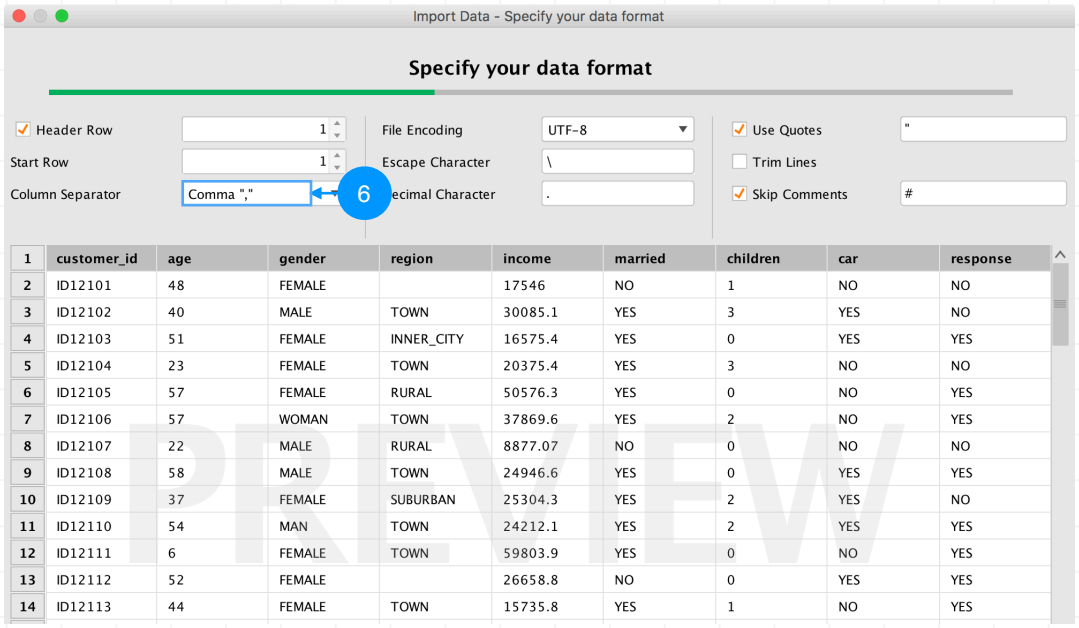


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import CSV

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ csv เข้าไปไว้ใน Repository
  - ในส่วนของ **Comma Separator** เลือก Comma “,” และกด **Next**



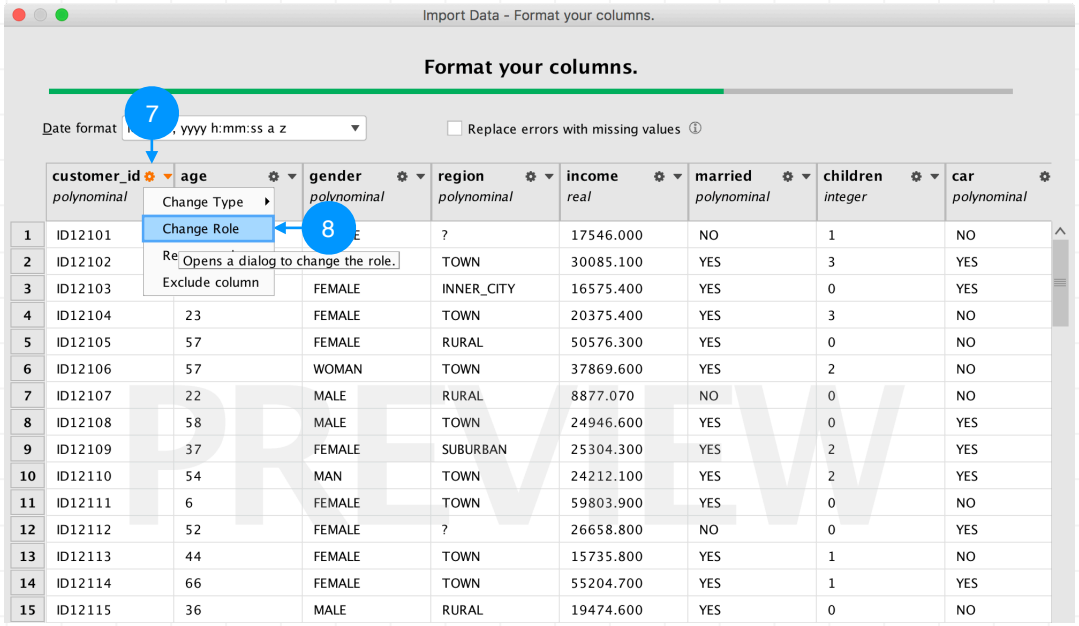
1	customer_id	age	gender	region	income	married	children	car	response
2	ID12101	48	FEMALE		17546	NO	1	NO	NO
3	ID12102	40	MALE	TOWN	30085.1	YES	3	YES	NO
4	ID12103	51	FEMALE	INNER_CITY	16575.4	YES	0	YES	YES
5	ID12104	23	FEMALE	TOWN	20375.4	YES	3	NO	NO
6	ID12105	57	FEMALE	RURAL	50576.3	YES	0	NO	YES
7	ID12106	57	WOMAN	TOWN	37869.6	YES	2	NO	YES
8	ID12107	22	MALE	RURAL	8877.07	NO	0	NO	NO
9	ID12108	58	MALE	TOWN	24946.6	YES	0	YES	YES
10	ID12109	37	FEMALE	SUBURBAN	25304.3	YES	2	YES	NO
11	ID12110	54	MAN	TOWN	24212.1	YES	2	YES	YES
12	ID12111	6	FEMALE	TOWN	59803.9	YES	0	NO	YES
13	ID12112	52	FEMALE		26658.8	NO	0	YES	YES
14	ID12113	44	FEMALE	TOWN	15735.8	YES	1	NO	YES

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import CSV

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ csv เข้าไปไว้ใน Repository
  - เปลี่ยนแอตทริบิวต์ **customer_id** ให้เป็นประเภท ID



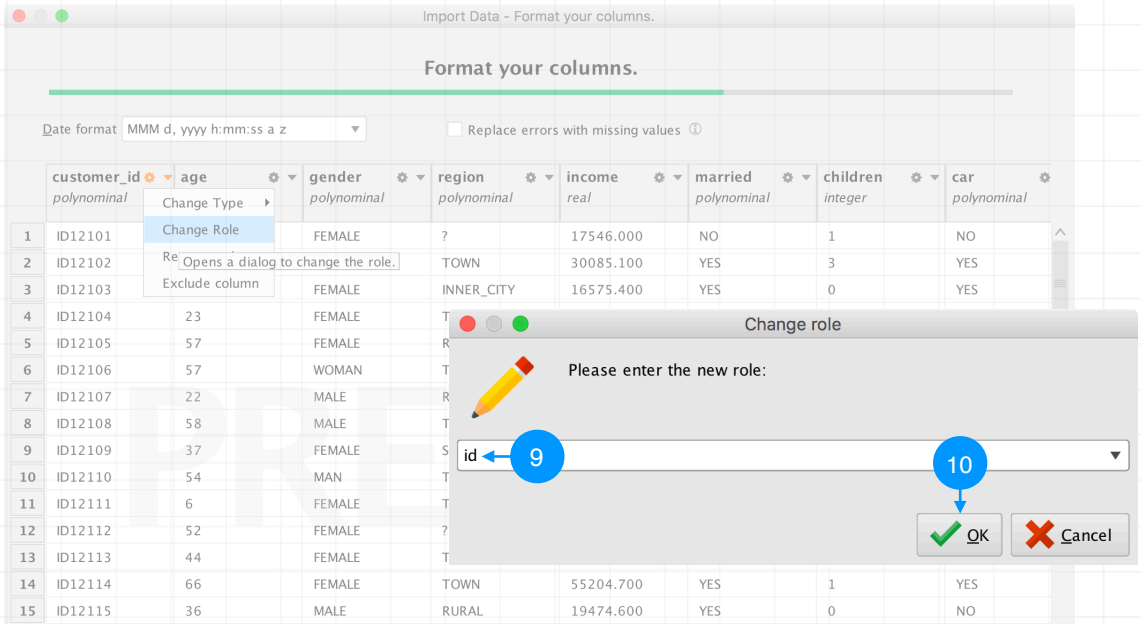
Copyright © 2021 Cube Analytics Consulting Co., Ltd.

CUBE Analytics CONSULTING

Import CSV

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ csv เข้าไปไว้ใน Repository
  - เปลี่ยนแอตทริบิวต์ **customer_id** ให้เป็นประเภท ID

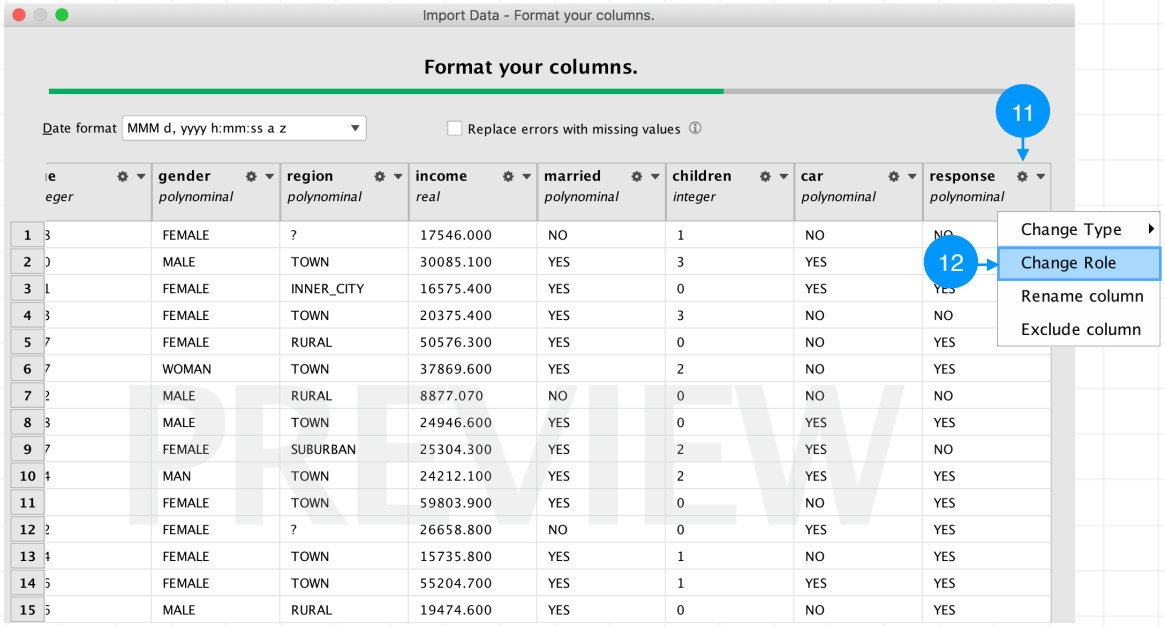


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import CSV

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ csv เข้าไปไว้ใน Repository
  - เปลี่ยนแอตทริบิวต์ **response** ให้เป็นประเภท label



	e integer	gender polynominal	region polynominal	income real	married polynominal	children integer	car polynominal	response polynominal
1	3	FEMALE	?	17546.000	NO	1	NO	NO
2	0	MALE	TOWN	30085.100	YES	3	YES	NO
3	1	FEMALE	INNER_CITY	16575.400	YES	0	YES	YES
4	3	FEMALE	TOWN	20375.400	YES	3	NO	NO
5	7	FEMALE	RURAL	50576.300	YES	0	NO	YES
6	7	WOMAN	TOWN	37869.600	YES	2	NO	YES
7	2	MALE	RURAL	8877.070	NO	0	NO	NO
8	3	MALE	TOWN	24946.600	YES	0	YES	YES
9	7	FEMALE	SUBURBAN	25304.300	YES	2	YES	NO
10	4	MAN	TOWN	24212.100	YES	2	YES	YES
11		FEMALE	TOWN	59803.900	YES	0	NO	YES
12	2	FEMALE	?	26658.800	NO	0	YES	YES
13	4	FEMALE	TOWN	15735.800	YES	1	NO	YES
14	5	FEMALE	TOWN	55204.700	YES	1	YES	YES
15	5	MALE	RURAL	19474.600	YES	0	NO	YES

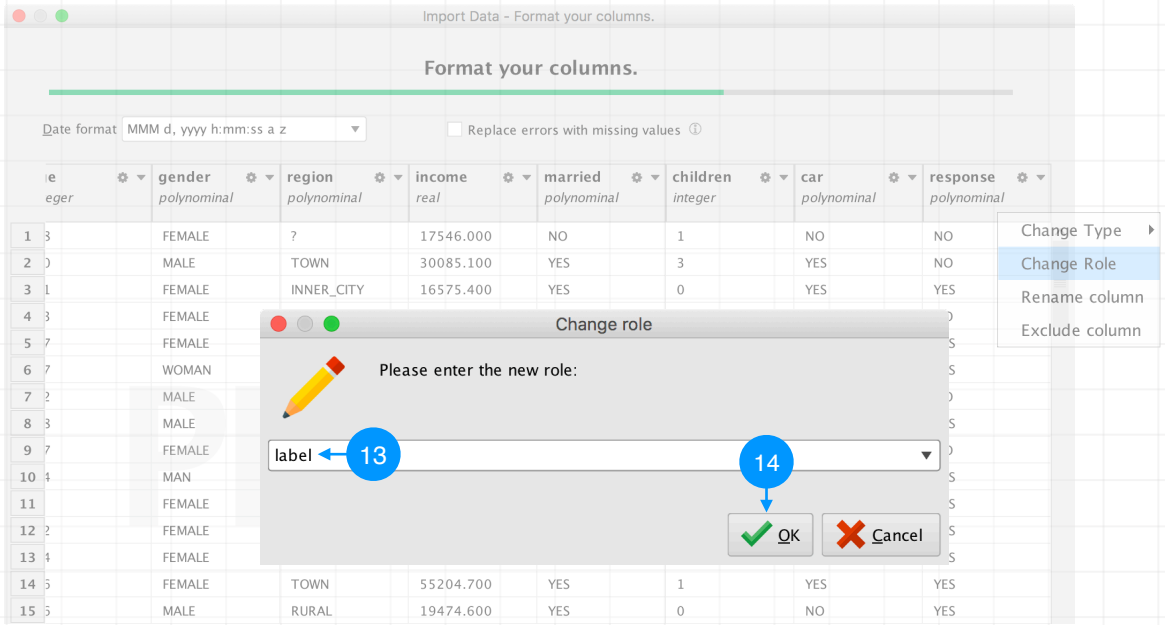
PREVIEW

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import CSV

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ csv เข้าไปไว้ใน Repository
  - เปลี่ยนแอตทริบิวต์ **response** ให้เป็นประเภท label



	e integer	gender polynominal	region polynominal	income real	married polynominal	children integer	car polynominal	response polynominal
1	3	FEMALE	?	17546.000	NO	1	NO	NO
2	0	MALE	TOWN	30085.100	YES	3	YES	NO
3	1	FEMALE	INNER_CITY	16575.400	YES	0	YES	YES
4	3	FEMALE						
5	7	FEMALE						
6	7	WOMAN						
7	2	MALE						
8	3	MALE						
9	7	FEMALE						
10	4	MAN						
11		FEMALE						
12	2	FEMALE						
13	4	FEMALE						
14	5	FEMALE	TOWN	55204.700	YES	1	YES	YES
15	5	MALE	RURAL	19474.600	YES	0	NO	YES

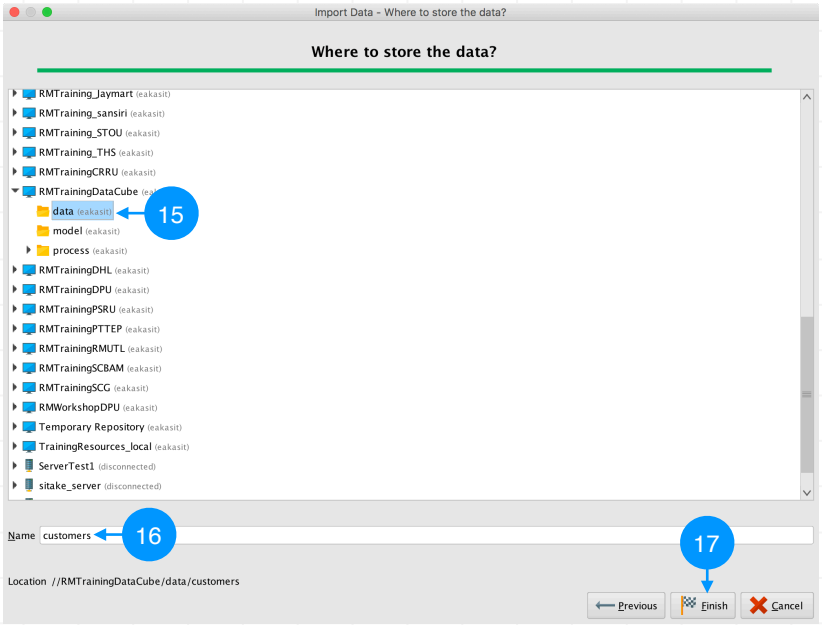
PREVIEW

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import CSV

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ csv เข้าไปไว้ใน Repository
  - save ชื่อว่า customers ไว้ที่โฟลเดอร์ **data** ภายใต้ **RMTrainingDataCube**

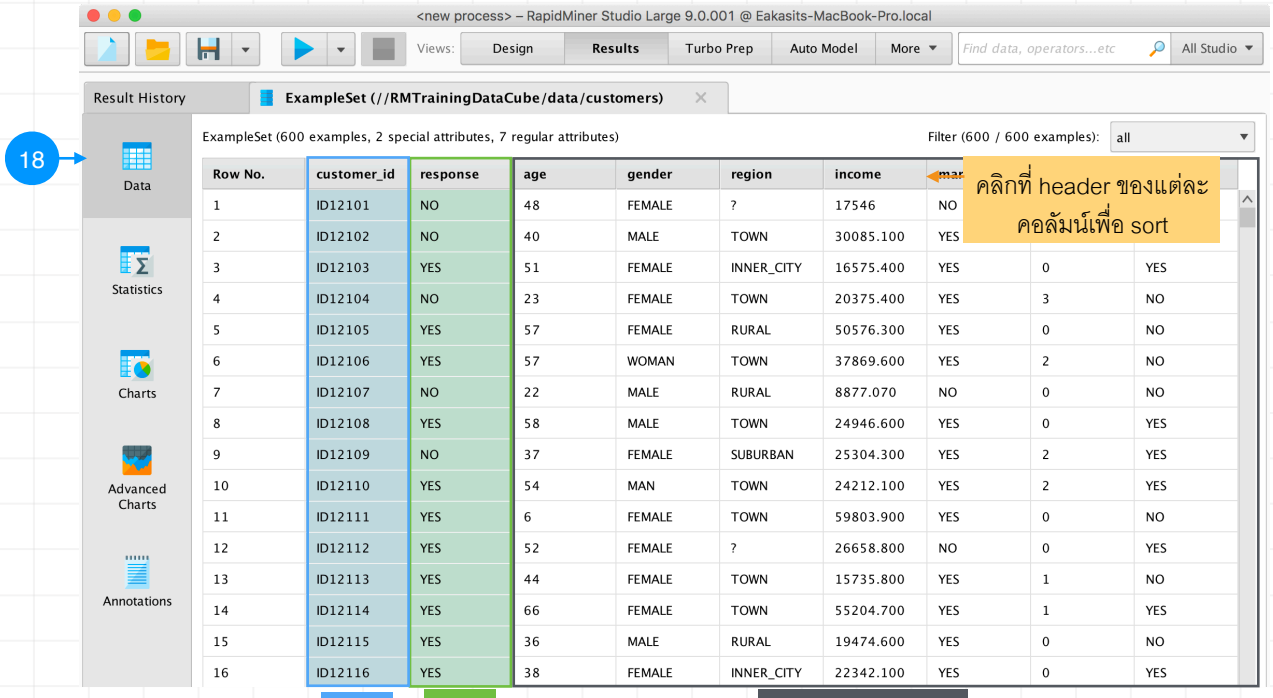


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import CSV

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลที่โหลดเข้าไปแสดงในรูปแบบของตาราง



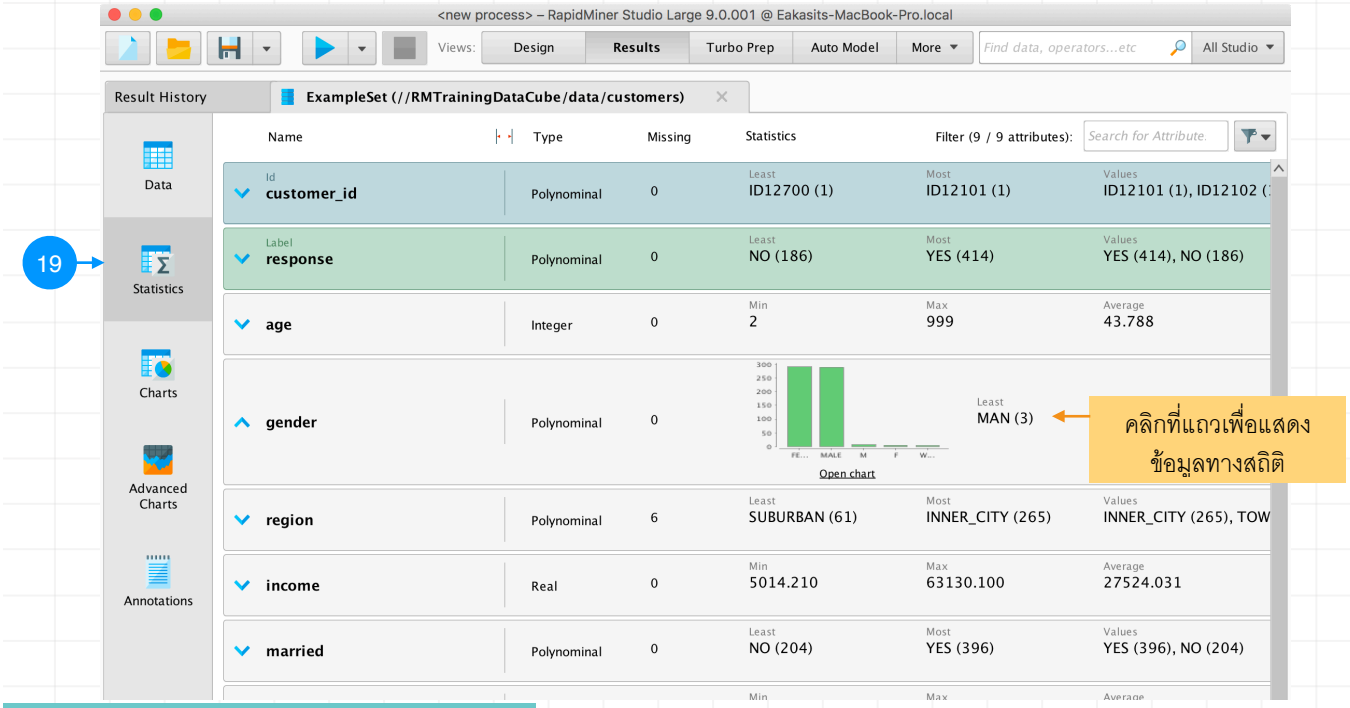
Row No.	customer_id	response	age	gender	region	income	mar		
1	ID12101	NO	48	FEMALE	?	17546	NO		
2	ID12102	NO	40	MALE	TOWN	30085.100	YES		
3	ID12103	YES	51	FEMALE	INNER_CITY	16575.400	YES	0	YES
4	ID12104	NO	23	FEMALE	TOWN	20375.400	YES	3	NO
5	ID12105	YES	57	FEMALE	RURAL	50576.300	YES	0	NO
6	ID12106	YES	57	WOMAN	TOWN	37869.600	YES	2	NO
7	ID12107	NO	22	MALE	RURAL	8877.070	NO	0	NO
8	ID12108	YES	58	MALE	TOWN	24946.600	YES	0	YES
9	ID12109	NO	37	FEMALE	SUBURBAN	25304.300	YES	2	YES
10	ID12110	YES	54	MAN	TOWN	24212.100	YES	2	YES
11	ID12111	YES	6	FEMALE	TOWN	59803.900	YES	0	NO
12	ID12112	YES	52	FEMALE	?	26658.800	NO	0	YES
13	ID12113	YES	44	FEMALE	TOWN	15735.800	YES	1	NO
14	ID12114	YES	66	FEMALE	TOWN	55204.700	YES	1	YES
15	ID12115	YES	36	MALE	RURAL	19474.600	YES	0	NO
16	ID12116	YES	38	FEMALE	INNER_CITY	22342.100	YES	0	YES

Copyright © 2021 Cube Analytics

Import CSV

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลที่โหลดเข้าไปแสดงในรูปแบบของค่าสถิติ

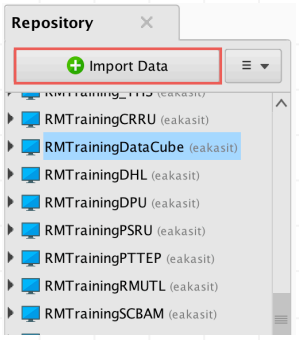


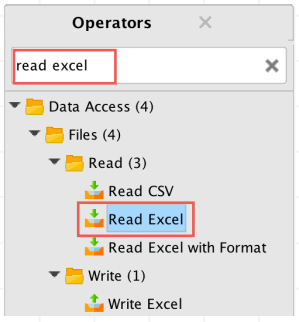
Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Load Excel to RapidMiner

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- การโหลดไฟล์ Excel เข้าไปใช้ใน RapidMiner Studio 9 ทำได้ 2 แบบ
  - **ใช้การ import ในส่วนของ Repositories**
    - โหลดมาเก็บไว้ใน Repository และใช้งานได้ตลอด
    - ถ้าข้อมูลในไฟล์ Excel มีการเปลี่ยนแปลงจะไม่ update ต้องทำการโหลดใหม่
  - ใช้โอเปอเรเตอร์ **Read Excel**
    - โหลดเข้ามาใช้งานโดยการอ่านจากไฟล์ Excel ทุกครั้ง เมื่อไฟล์ update ข้อมูลจะเปลี่ยนตาม





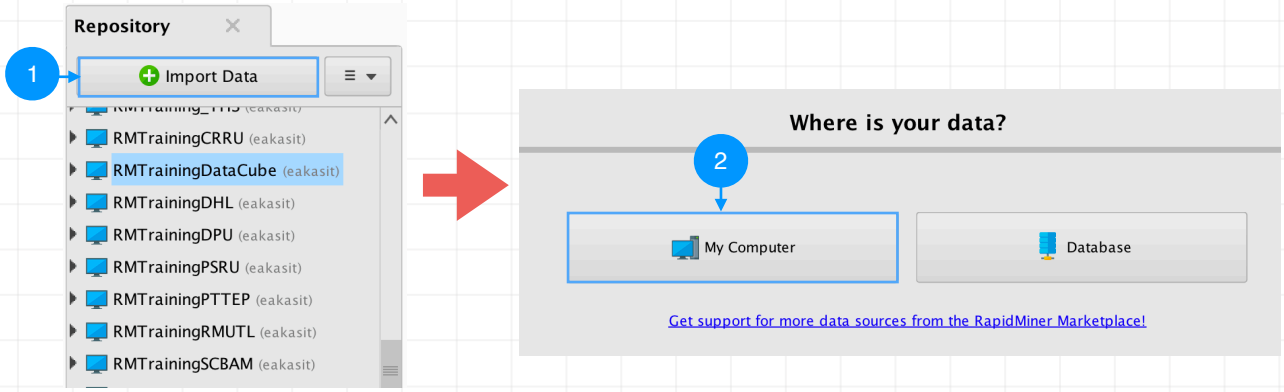
Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import Excel

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ Excel เข้าไปไว้ใน Repository
  - ในส่วน **Repositories** คลิกที่ปุ่ม
  - เลือกอ่านข้อมูลจาก Computer



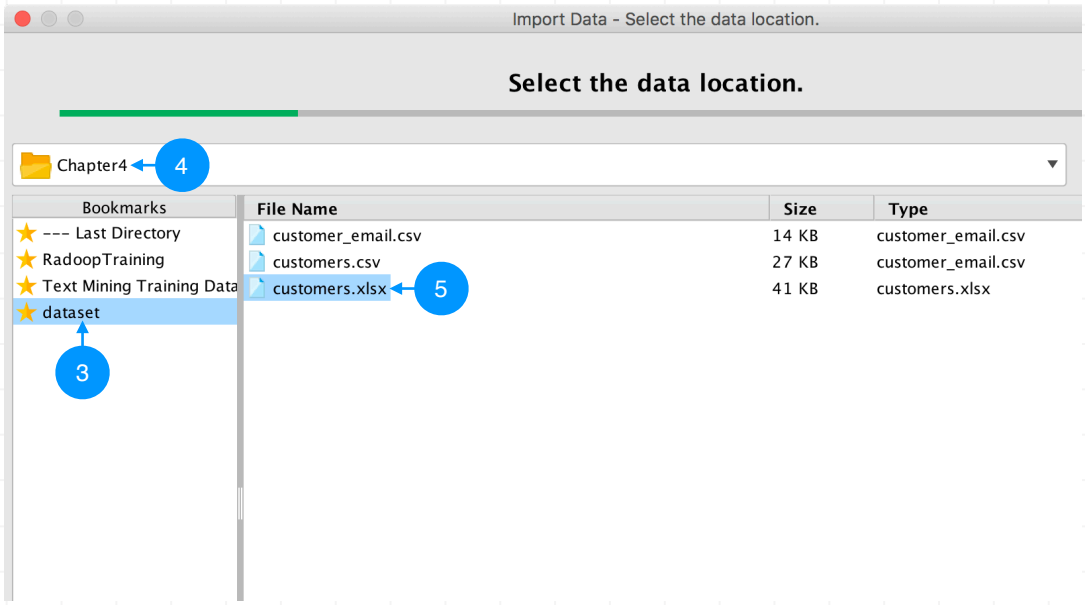


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import Excel

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ Excel เข้าไปไว้ใน Repository
  - เลือกไฟล์ customers.xlsx จากโฟลเดอร์ dataset/Chapter4

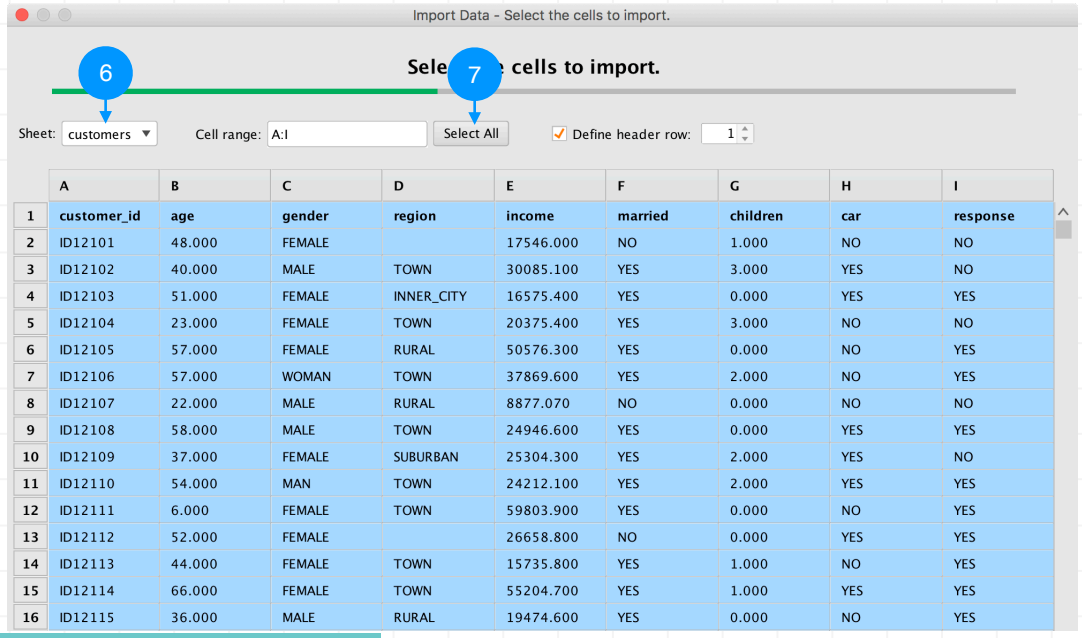


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import Excel

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ Excel เข้าไปไว้ใน Repository
  - เลือก Sheet ที่ต้องการ import จากไฟล์ Excel กดปุ่ม **Next**



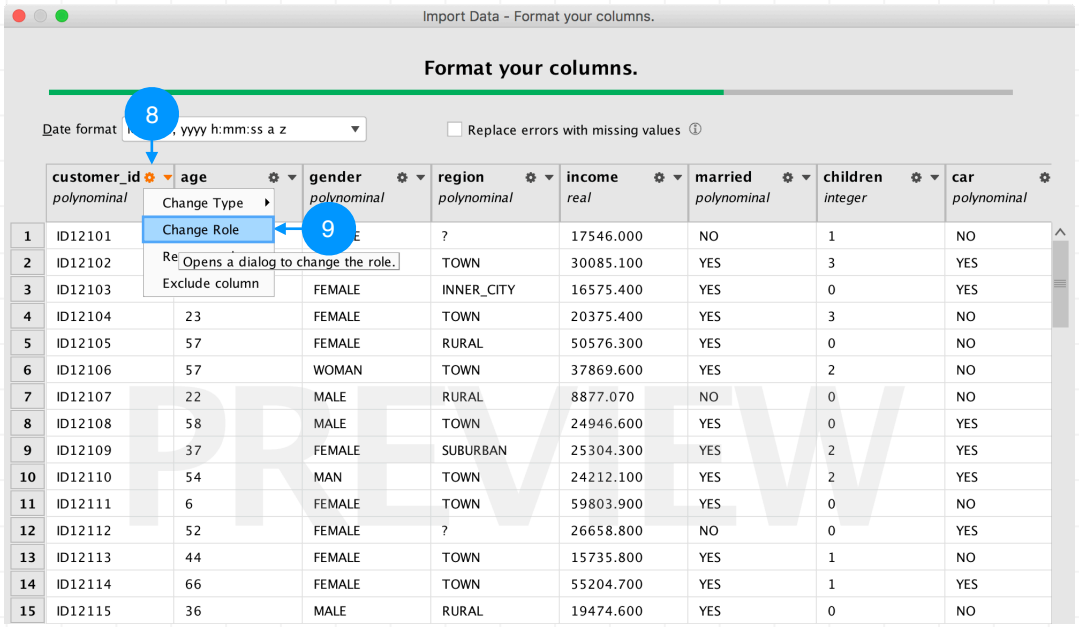
	A	B	C	D	E	F	G	H	I
1	customer_id	age	gender	region	income	married	children	car	response
2	ID12101	48.000	FEMALE		17546.000	NO	1.000	NO	NO
3	ID12102	40.000	MALE	TOWN	30085.100	YES	3.000	YES	NO
4	ID12103	51.000	FEMALE	INNER_CITY	16575.400	YES	0.000	YES	YES
5	ID12104	23.000	FEMALE	TOWN	20375.400	YES	3.000	NO	NO
6	ID12105	57.000	FEMALE	RURAL	50576.300	YES	0.000	NO	YES
7	ID12106	57.000	WOMAN	TOWN	37869.600	YES	2.000	NO	YES
8	ID12107	22.000	MALE	RURAL	8877.070	NO	0.000	NO	NO
9	ID12108	58.000	MALE	TOWN	24946.600	YES	0.000	YES	YES
10	ID12109	37.000	FEMALE	SUBURBAN	25304.300	YES	2.000	YES	NO
11	ID12110	54.000	MAN	TOWN	24212.100	YES	2.000	YES	YES
12	ID12111	6.000	FEMALE	TOWN	59803.900	YES	0.000	NO	YES
13	ID12112	52.000	FEMALE		26658.800	NO	0.000	YES	YES
14	ID12113	44.000	FEMALE	TOWN	15735.800	YES	1.000	NO	YES
15	ID12114	66.000	FEMALE	TOWN	55204.700	YES	1.000	YES	YES
16	ID12115	36.000	MALE	RURAL	19474.600	YES	0.000	NO	YES

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import Excel

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ Excel เข้าไปไว้ใน Repository
  - เปลี่ยนแอตทริบิวต์ **customer_id** ให้เป็นประเภท ID



	customer_id *polynominal*	age	gender *polynominal*	region *polynominal*	income *real*	married *polynominal*	children *integer*	car *polynominal*
1	ID12101		E	?	17546.000	NO	1	NO
2	ID12102			TOWN	30085.100	YES	3	YES
3	ID12103		FEMALE	INNER_CITY	16575.400	YES	0	YES
4	ID12104	23	FEMALE	TOWN	20375.400	YES	3	NO
5	ID12105	57	FEMALE	RURAL	50576.300	YES	0	NO
6	ID12106	57	WOMAN	TOWN	37869.600	YES	2	NO
7	ID12107	22	MALE	RURAL	8877.070	NO	0	NO
8	ID12108	58	MALE	TOWN	24946.600	YES	0	YES
9	ID12109	37	FEMALE	SUBURBAN	25304.300	YES	2	YES
10	ID12110	54	MAN	TOWN	24212.100	YES	2	YES
11	ID12111	6	FEMALE	TOWN	59803.900	YES	0	NO
12	ID12112	52	FEMALE	?	26658.800	NO	0	YES
13	ID12113	44	FEMALE	TOWN	15735.800	YES	1	NO
14	ID12114	66	FEMALE	TOWN	55204.700	YES	1	YES
15	ID12115	36	MALE	RURAL	19474.600	YES	0	NO

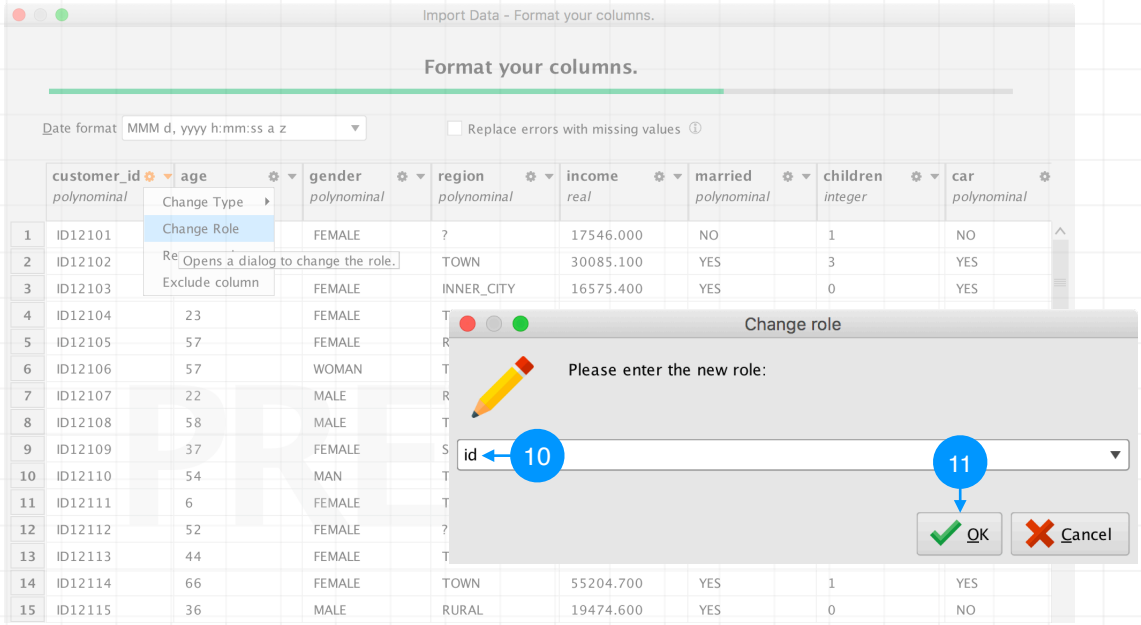
PREVIEW

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import Excel

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ Excel เข้าไปไว้ใน Repository
  - เปลี่ยนแอตทริบิวต์ **customer_id** ให้เป็นประเภท ID

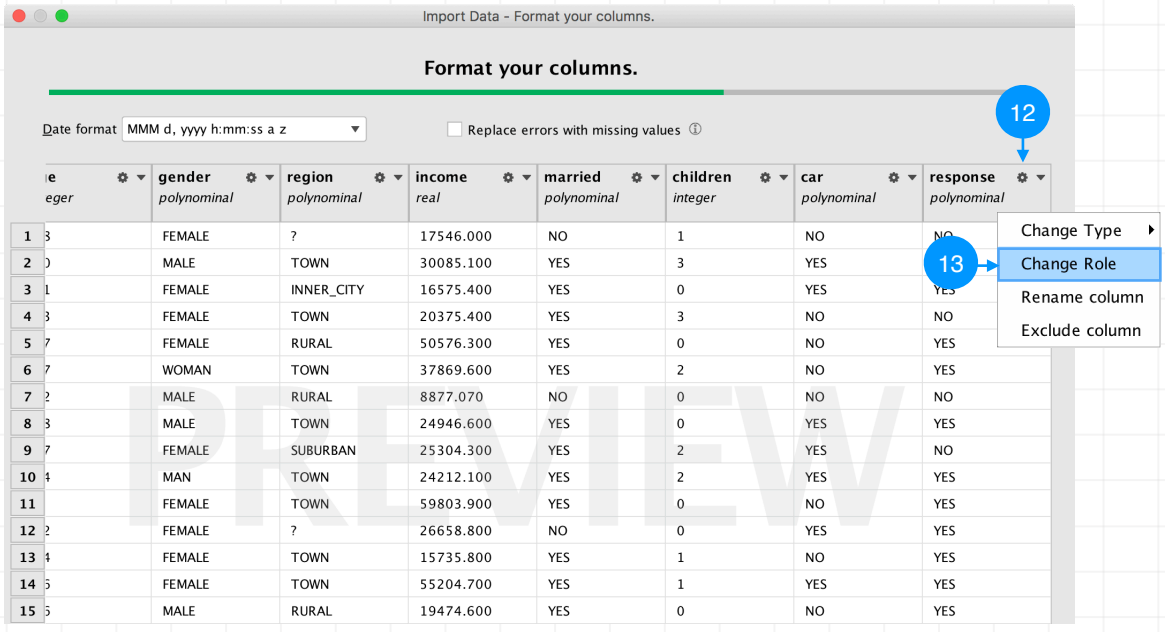


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import Excel

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ Excel เข้าไปไว้ใน Repository
  - เปลี่ยนแอตทริบิวต์ **response** ให้เป็นประเภท label

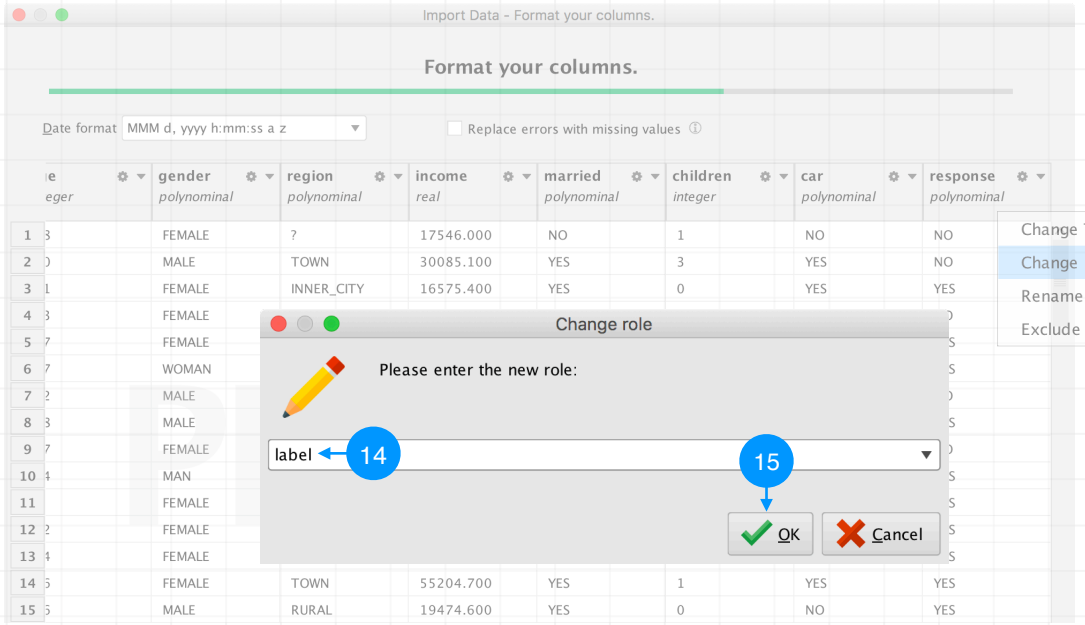


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import Excel

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ Excel เข้าไปไว้ใน Repository
  - เปลี่ยนแอตทริบิวต์ **response** ให้เป็นประเภท label

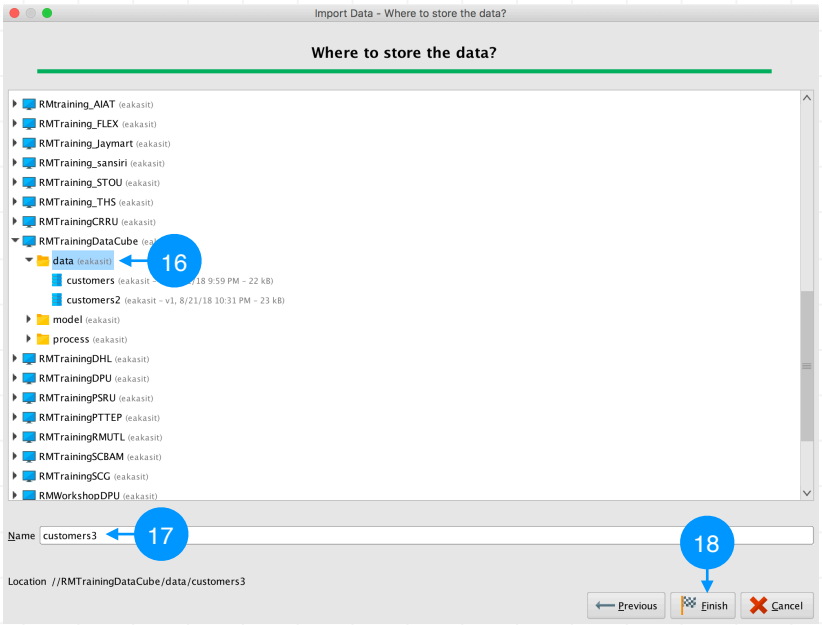


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import Excel

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- โหลดไฟล์ csv เข้าไปไว้ใน Repository
  - save ชื่อว่า customers3 ไว้ที่โฟลเดอร์ **data** ภายใต้ **RMTrainingDataCube**

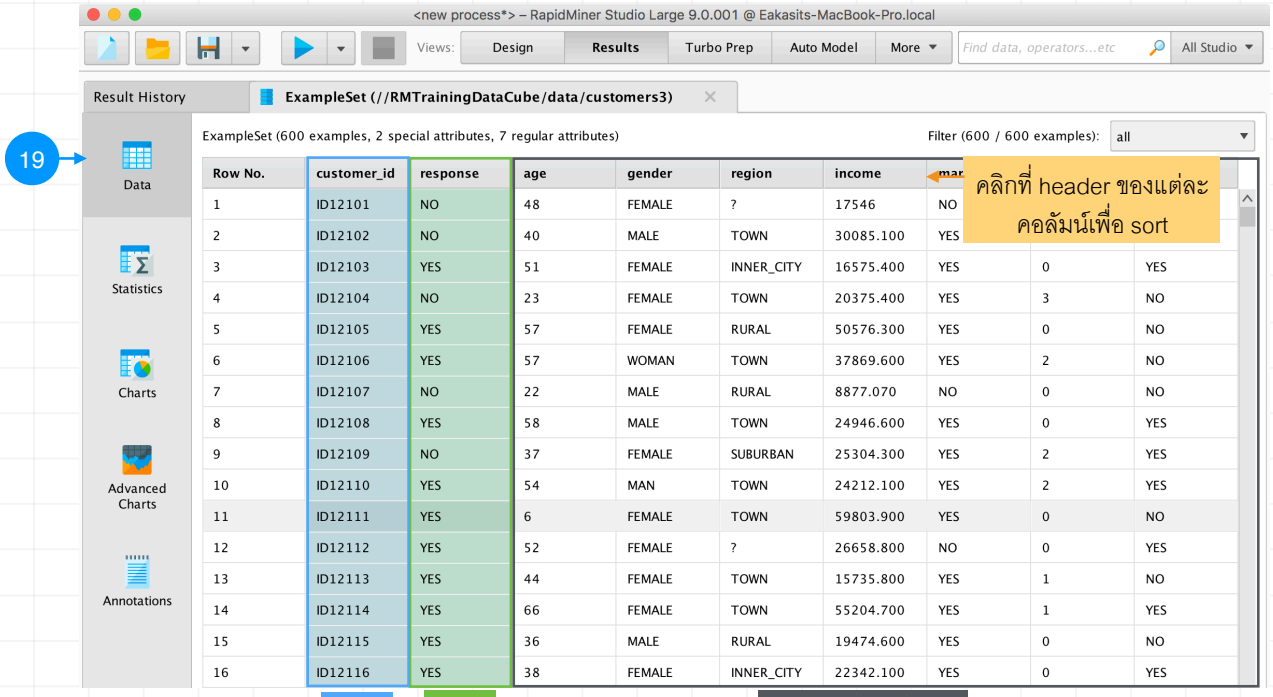


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import Excel

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลที่โหลดเข้าไปแสดงในรูปแบบของตาราง

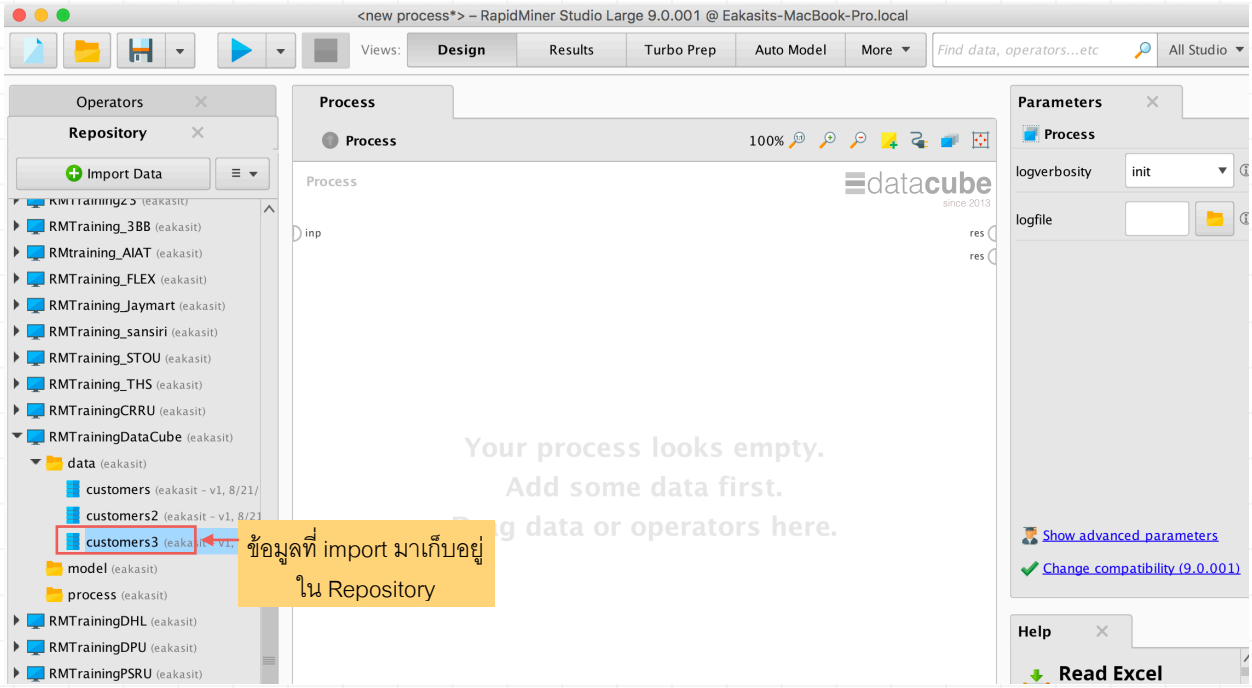


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Import Excel

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลที่โหลดได้จากไฟล์ Excel จะอยู่ใน **Repository**

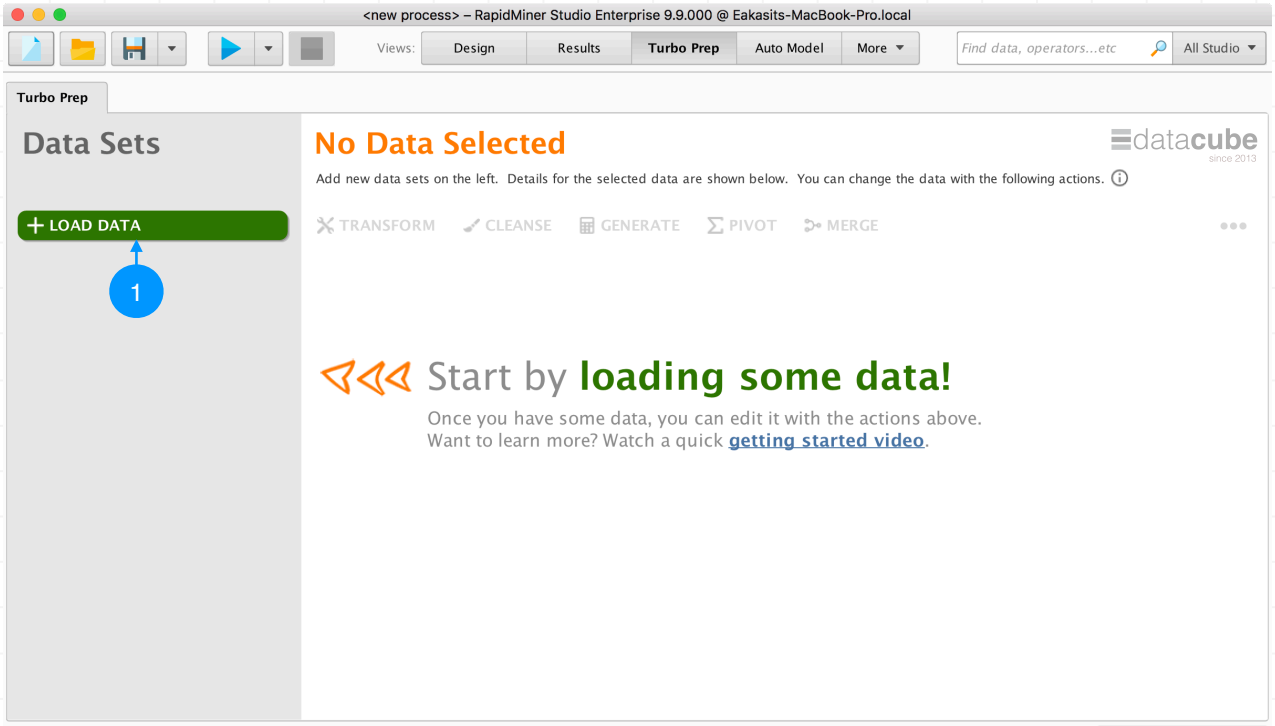


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Turbo Prep

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เป็น GUI ที่ช่วยให้การจัดการข้อมูลได้ง่ายขึ้น

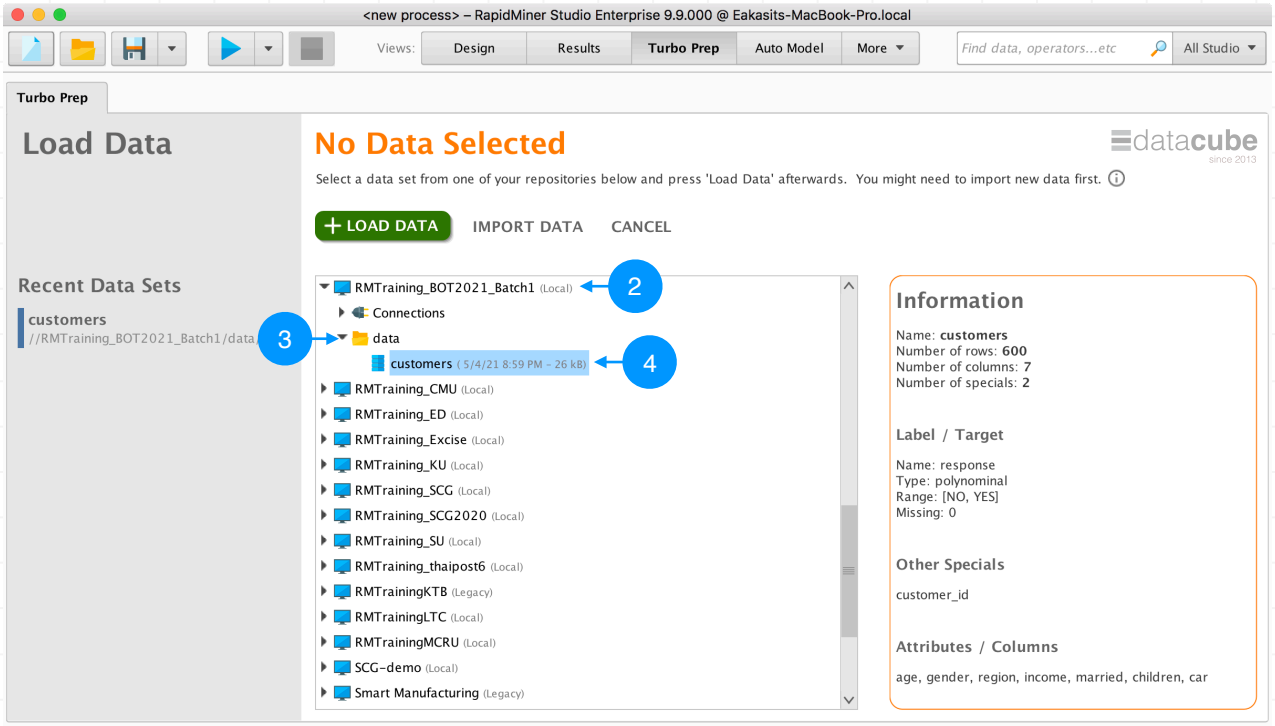


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Turbo Prep: Load Data

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกข้อมูล customers จาก repository



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data Preparation

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

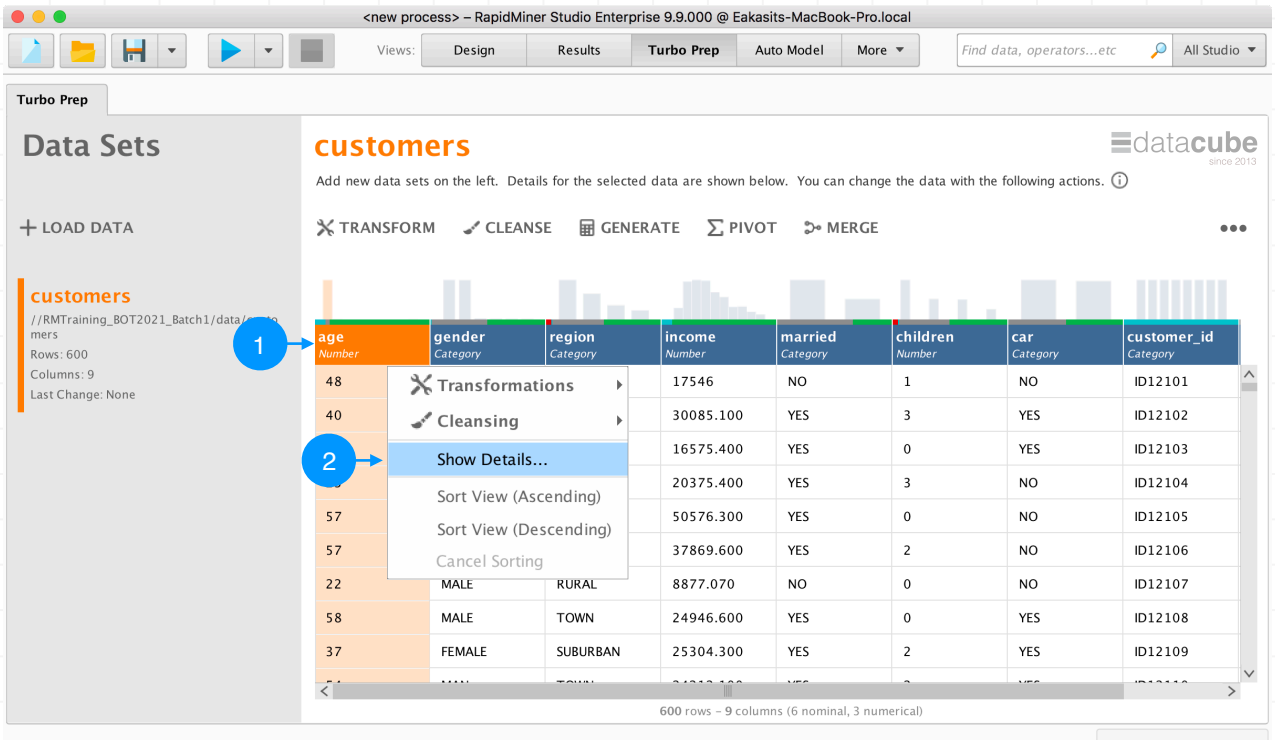
- การเตรียมข้อมูลด้วย Turbo Prep
  - **การดูรายละเอียดของข้อมูล (show detail)**
  - การเรียงลำดับข้อมูล (sort)
  - การแทนค่าข้อมูล (replace)
  - การแทนค่าว่าง (replace missing value)
  - การเลือกข้อมูลบางตัวอย่าง (filter examples)
  - การ join ข้อมูล 2 ชุดเข้าด้วยกัน
  - การ Export Data และ Process

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Show Details...

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Age** และคลิกขวา เลือกเมนู **Show Details...**

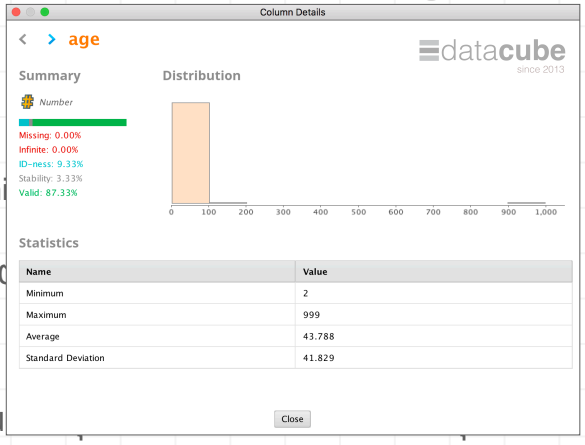


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Show Details...

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- รายละเอียดของแอตทริบิวต์ที่เลือก
  - **Missing** แสดงจำนวนรายการที่เป็นค่าว่าง
  - **Infinite** แสดงจำนวนรายการที่เป็นค่า infini
  - **ID-ness** แสดงสัดส่วนของจำนวนตัวเลขที่ข้อมูล
  - **Stability** แสดงสัดส่วนของจำนวนที่เกิดขึ้นข้อมูล
  - **Valid** จำนวนสัดส่วนของข้อมูลที่ไม่เป็น missing, infinity, id และ stable

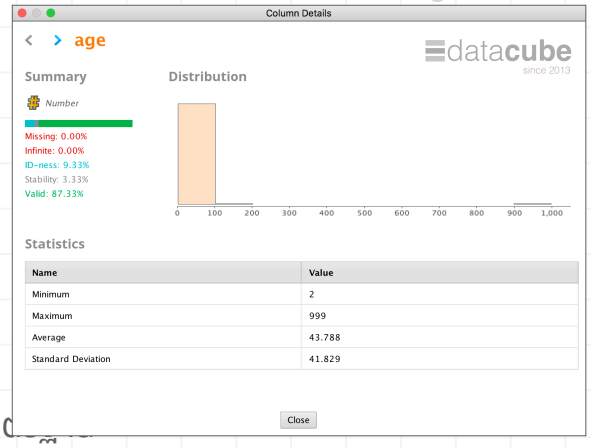


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Show Details...

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- รายละเอียดของแอตทริบิวต์ที่เลือก (ในกรณีที่เป็น Number)
  - **Minimum** แสดงค่าที่น้อยที่สุด
  - **Maximum** แสดงค่าที่มากที่สุด
  - **Average** แสดงค่าเฉลี่ย
  - **Standard Deviation** แสดงค่าเบี่ยงเบนมาต
  - **Distribution** แสดงการกระจายตัวของข้อมูล

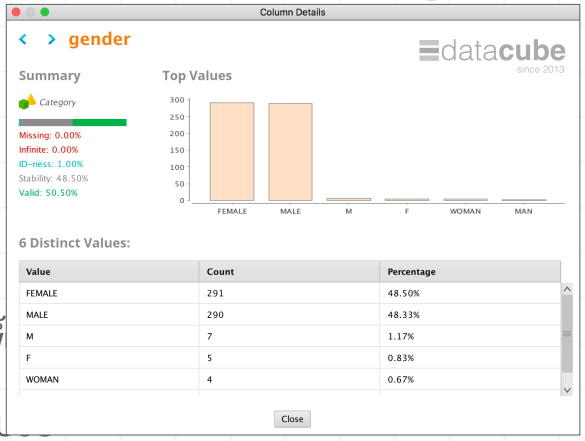


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Show Details...

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- รายละเอียดของแอตทริบิวต์ที่เลือก (ในกรณีที่เป็น Category)
  - **Value** แสดงค่าที่อยู่ในแอตทริบิวต์นั้น
  - **Count** แสดงจำนวนรายการที่เกิดขึ้น
  - **Percentage** แสดงจำนวน percent ที่เกิดขึ้
  - **Top Values** แสดงจำนวนรายการที่เกิดขึ้น



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data Preparation

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

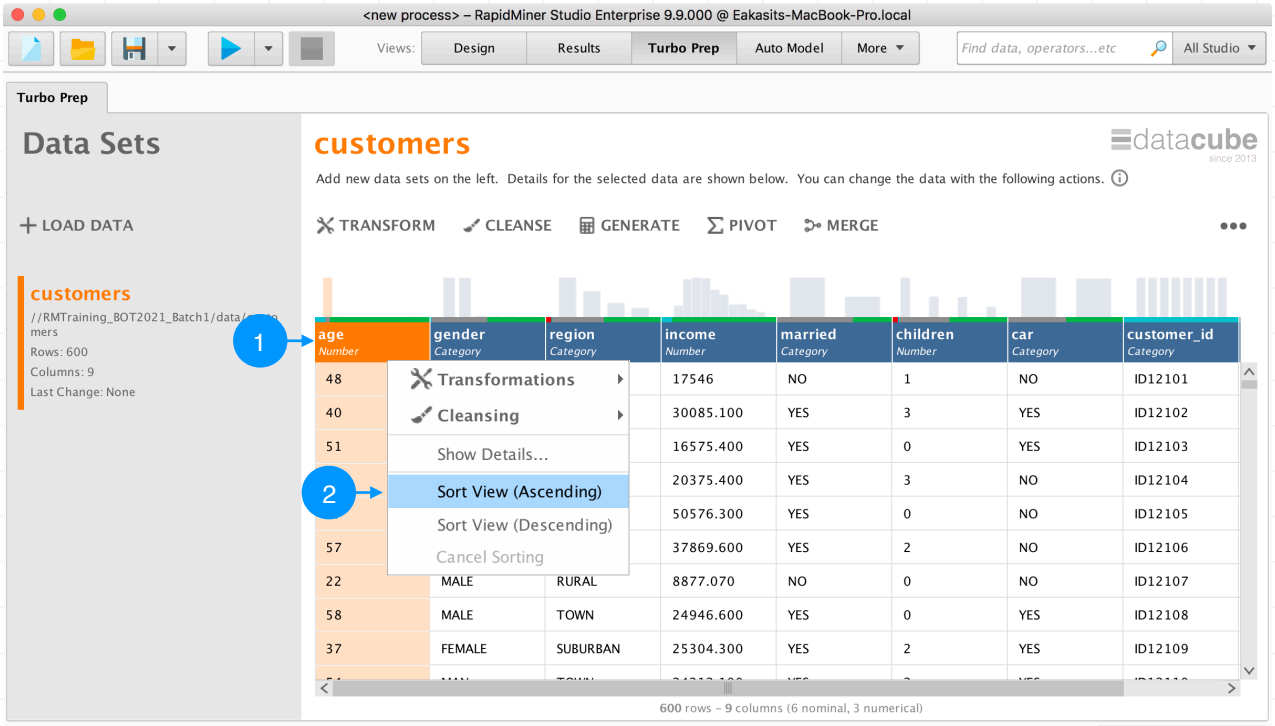
- การเตรียมข้อมูลด้วย Turbo Prep
  - การดูรายละเอียดของข้อมูล (show detail)
  - **การเรียงลำดับข้อมูล (sort)**
  - การแทนค่าข้อมูล (replace)
  - การแทนค่าว่าง (replace missing value)
  - การเลือกข้อมูลบางตัวอย่าง (filter examples)
  - การ join ข้อมูล 2 ชุดเข้าด้วยกัน
  - การ Export Data และ Process

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Sort View

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Age** และคลิกขวา เลือกเมนู **Sort View (Ascending)**

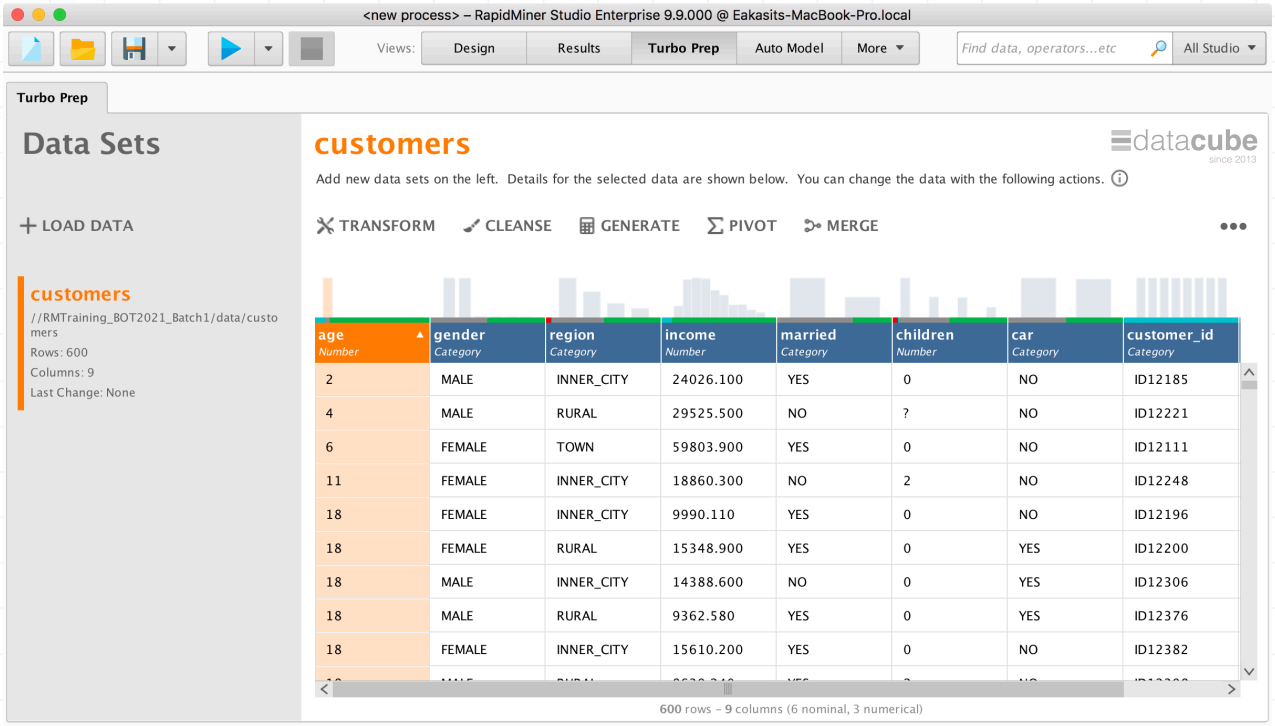


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Sort View

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลจะเรียงลำดับจาก**น้อยไปมาก** (Ascending)

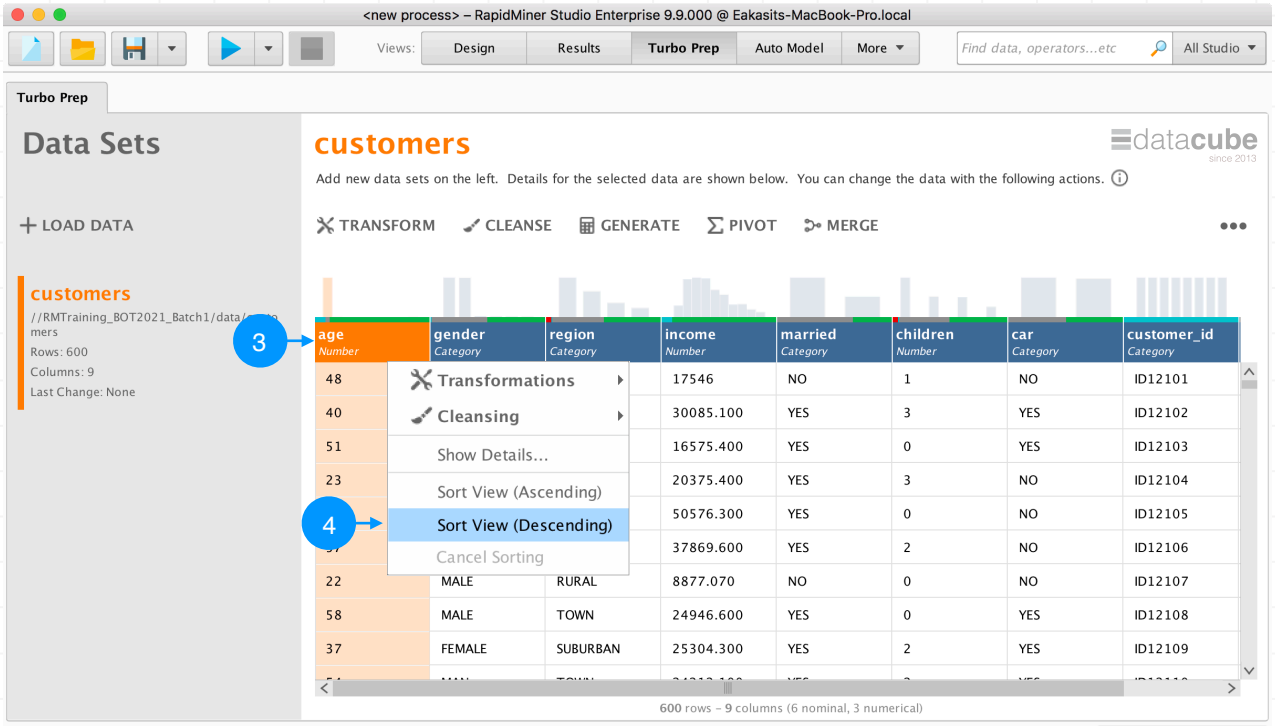


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Sort View

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Age** และคลิกขวา เลือกเมนู **Sort View (Descending)**

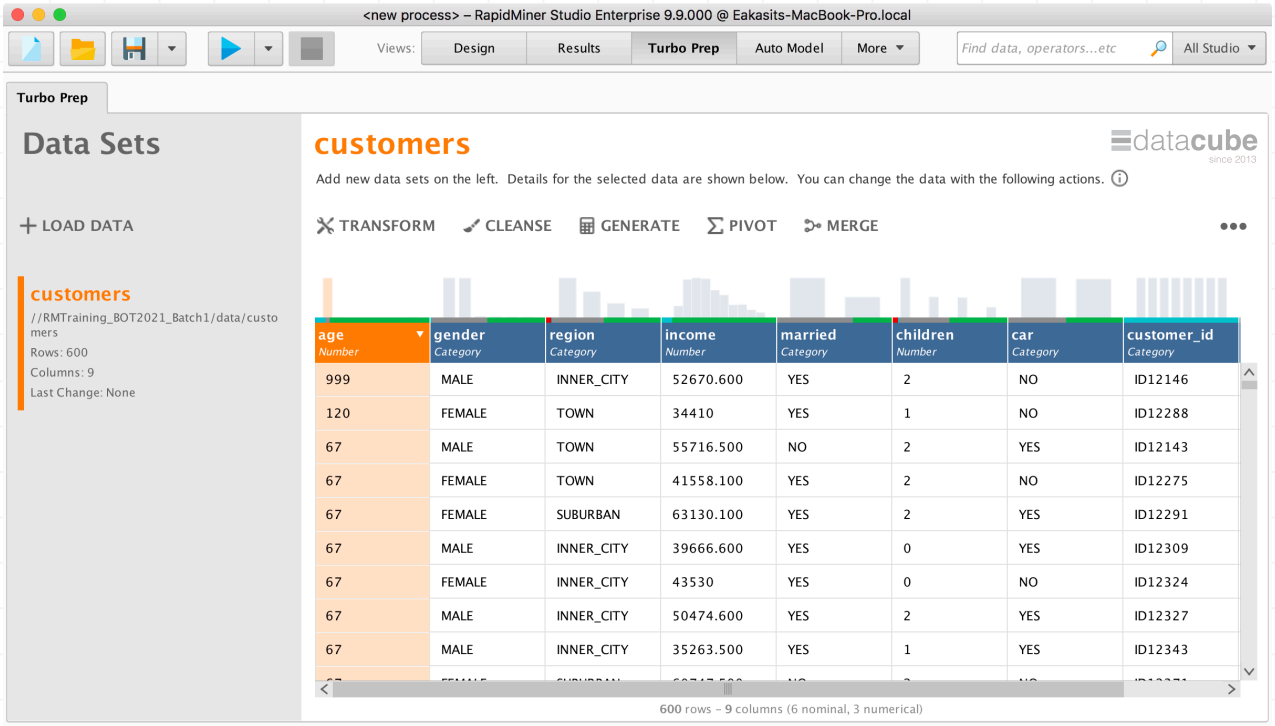


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Sort View

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลจะเรียงลำดับจาก**มากไปน้อย** (Descending)



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data Preparation

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

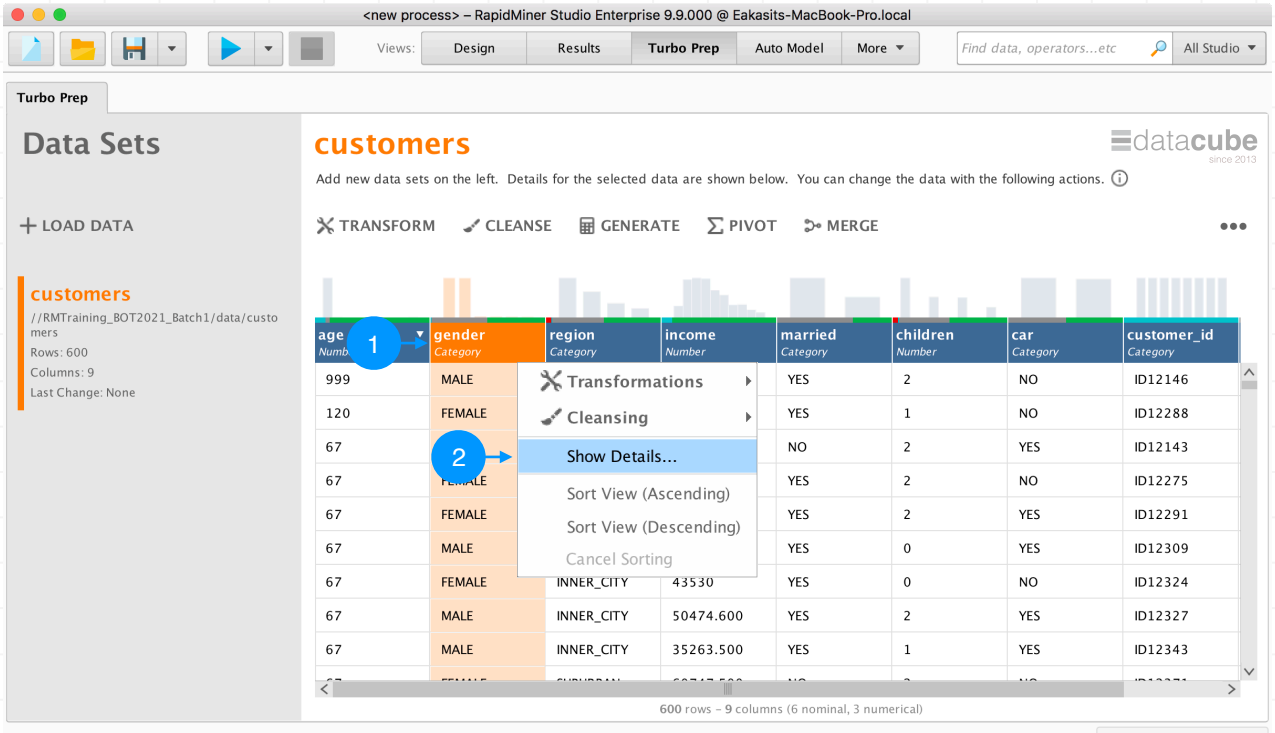
- การเตรียมข้อมูลด้วย Turbo Prep
  - การดูรายละเอียดของข้อมูล (show detail)
  - การเรียงลำดับข้อมูล (sort)
  - **การแทนค่าข้อมูล (replace)**
  - การแทนค่าว่าง (replace missing value)
  - การเลือกข้อมูลบางตัวอย่าง (filter examples)
  - การ join ข้อมูล 2 ชุดเข้าด้วยกัน
  - การ Export Data และ Process

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Gender** และคลิกขวา เลือกเมนู **Show Details...**

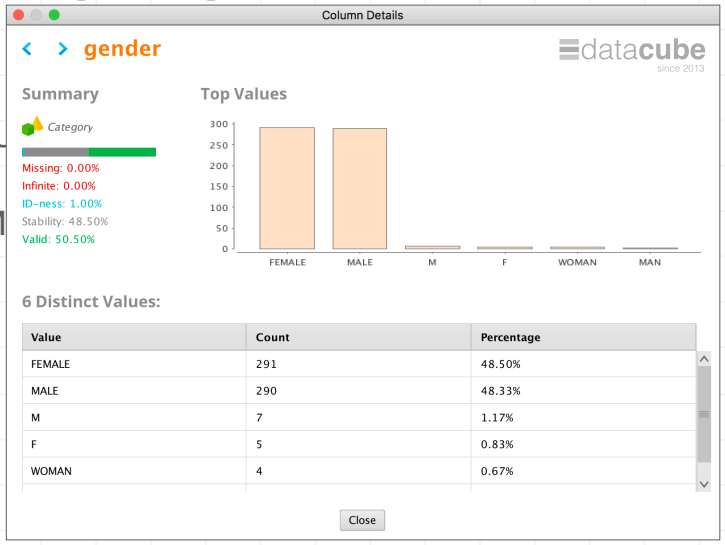


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลในแอตทริบิวต์ **Gender** มีค่าที่ไม่ถูกต้องอยู่ด้วยคือ **F, M, MAN** และ **WOMAN**
- ต้องเปลี่ยน F และ WOMAN ให้เป็น
- ต้องเปลี่ยน M และ MAN ให้เป็น M

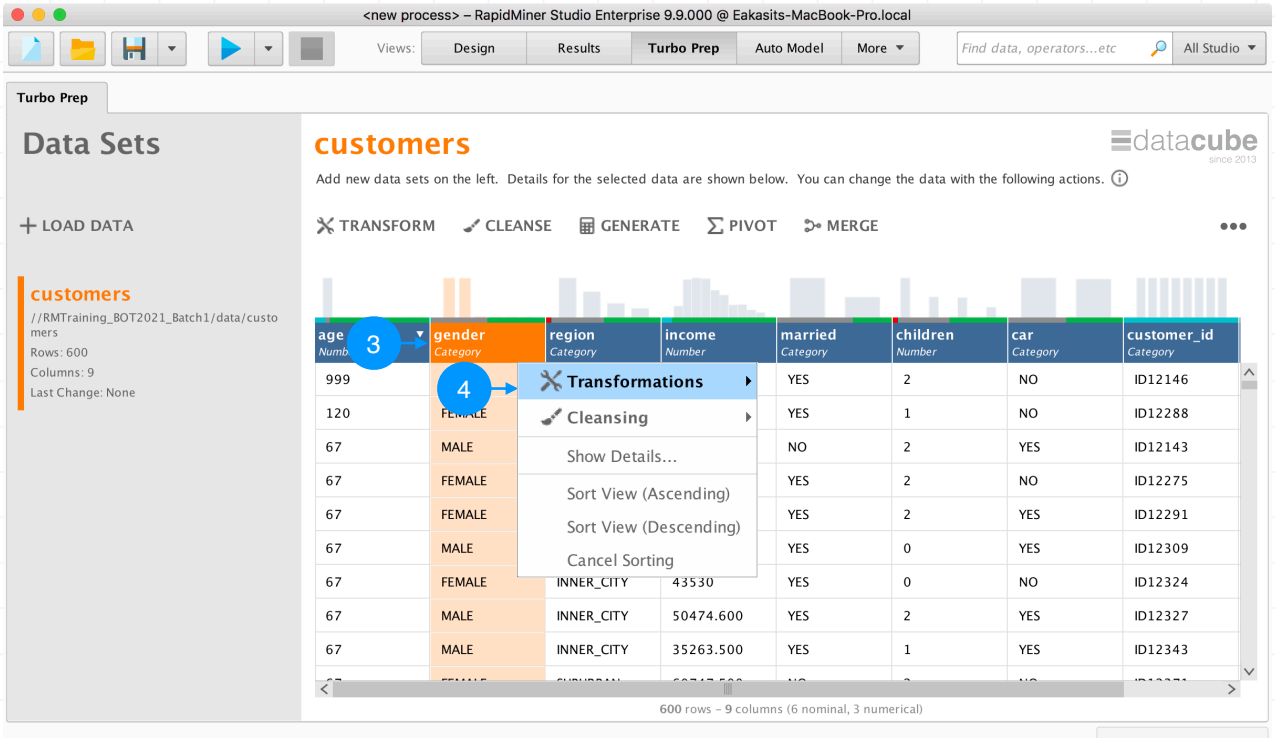


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Gender** และคลิกขวา เลือกเมนู **Transformations**

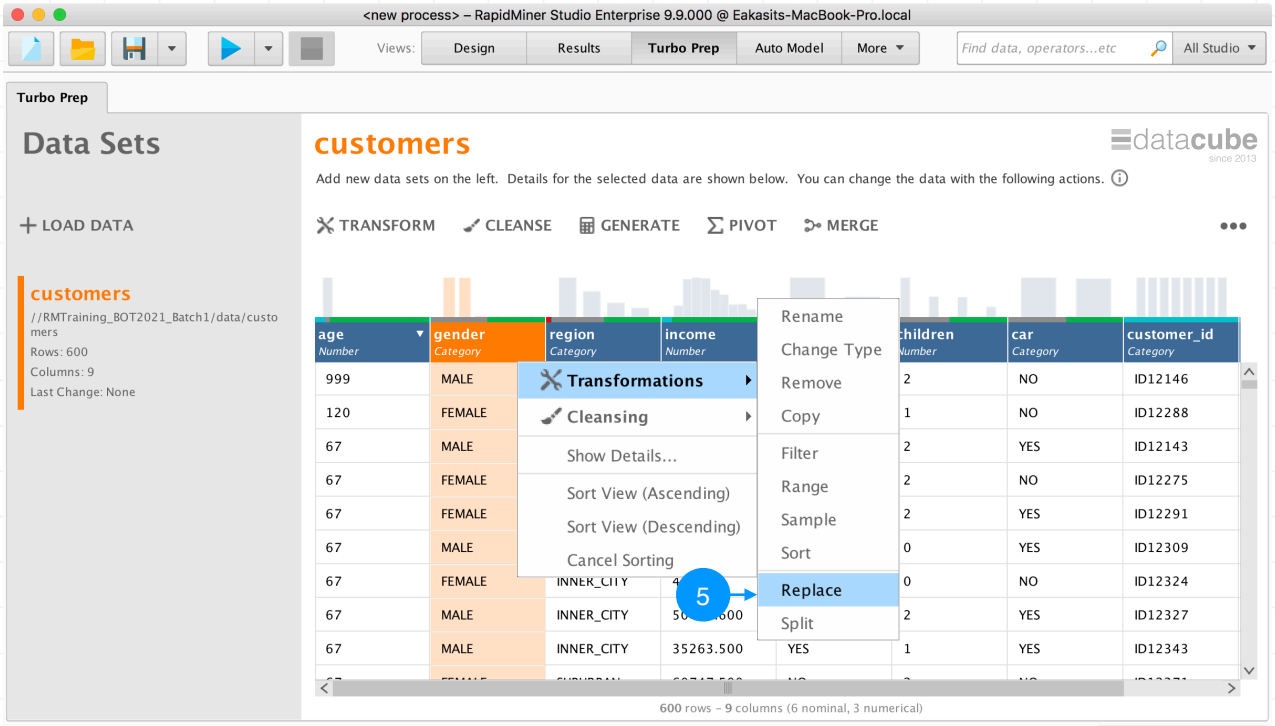


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกเมนู **Replace**

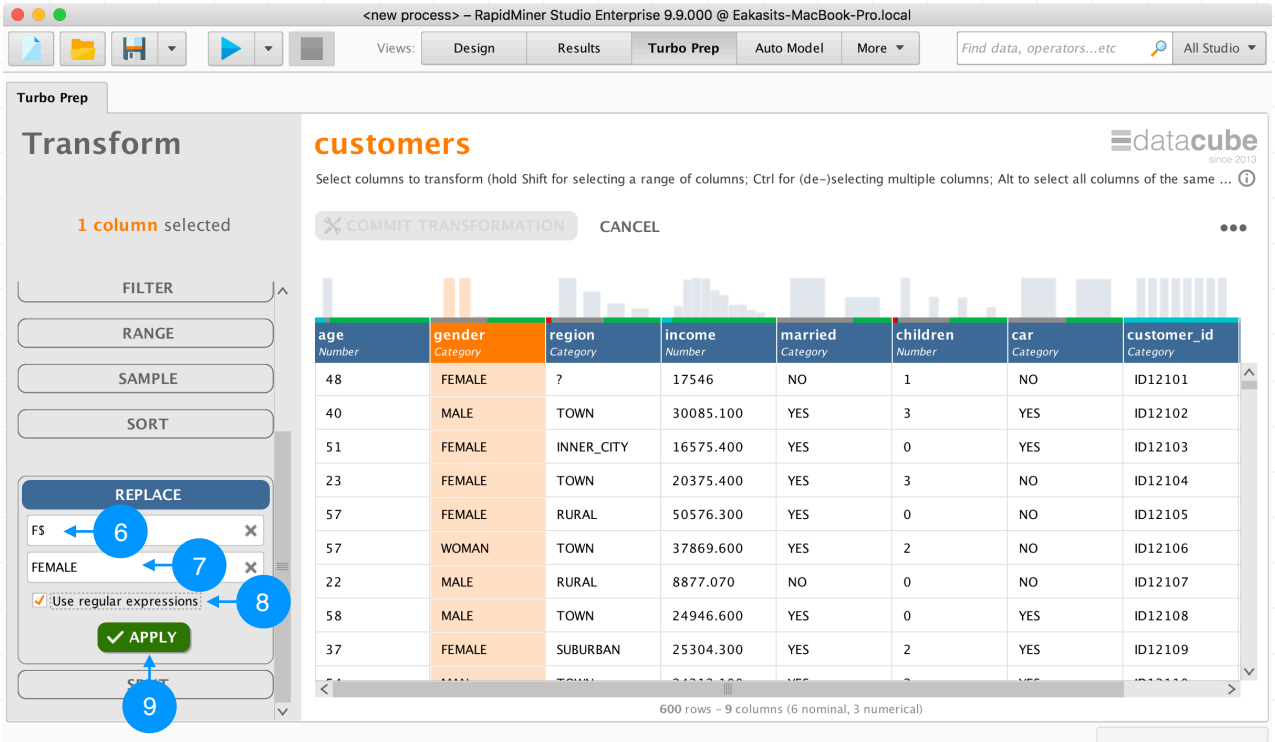


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เปลี่ยนค่าในส่วนของ Replace ด้านซ้ายมือ

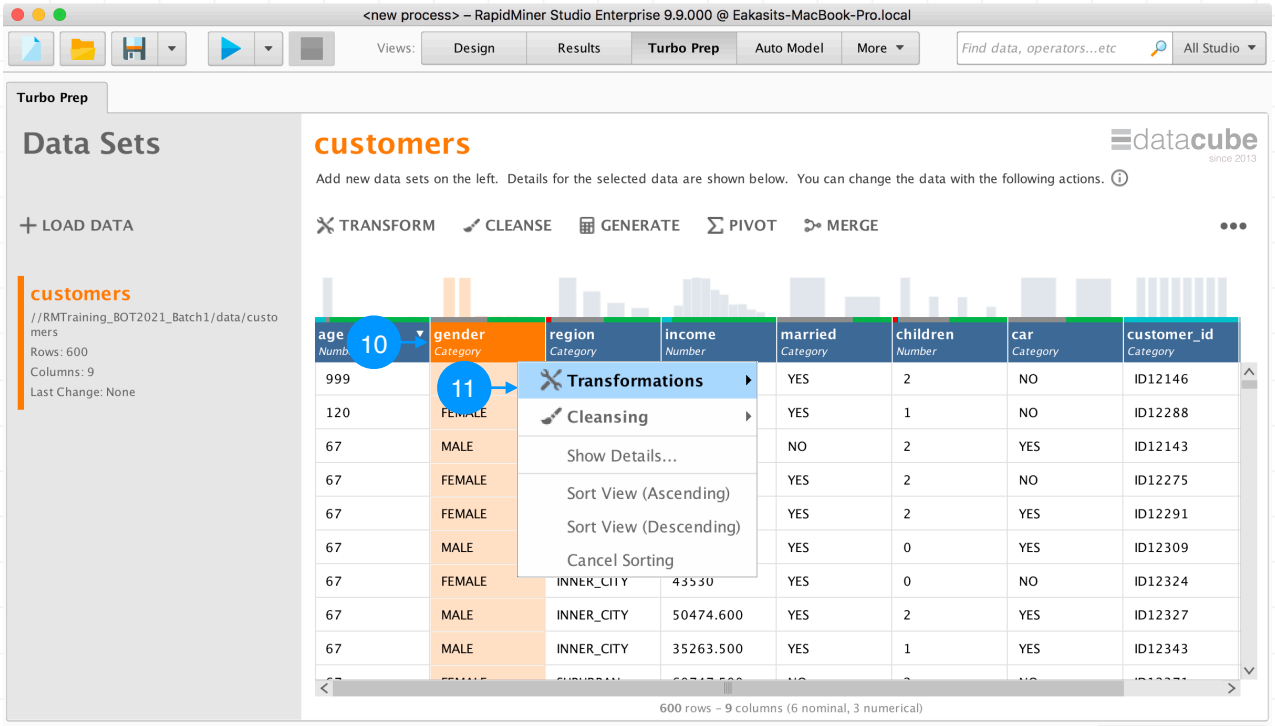


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Gender** และคลิกขวา เลือกเมนู **Transformations**

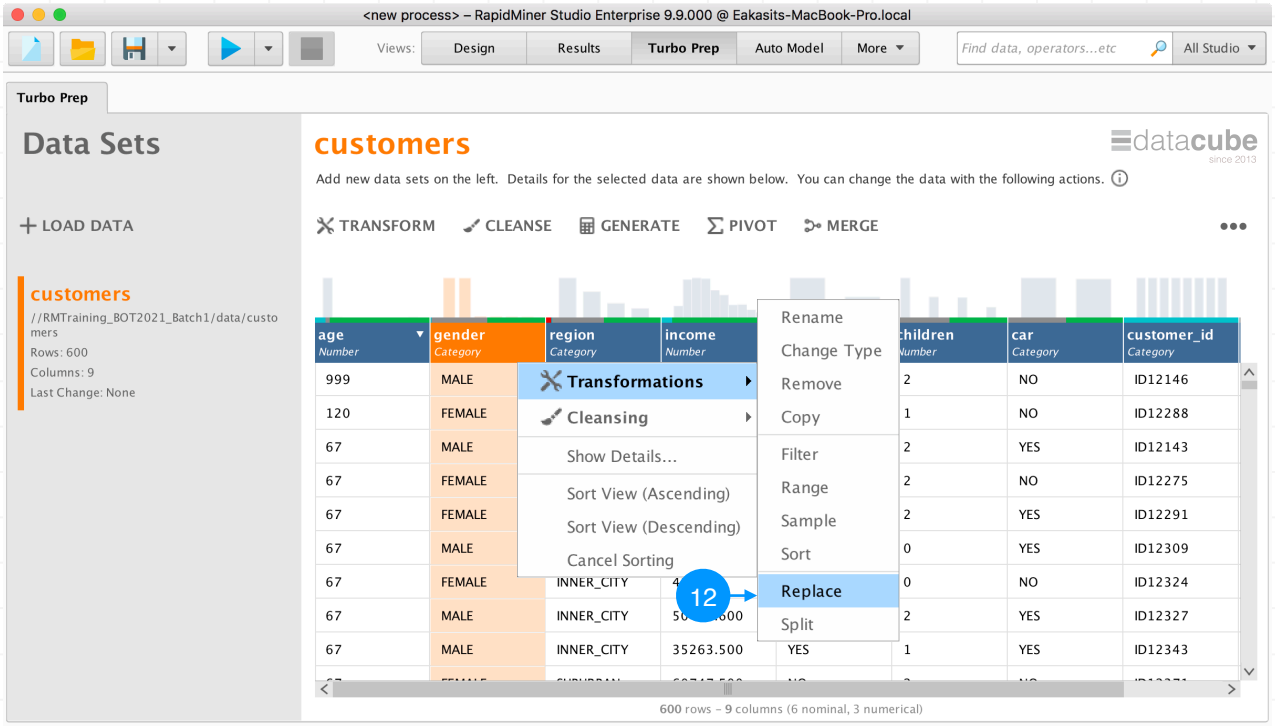


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกเมนู **Replace**

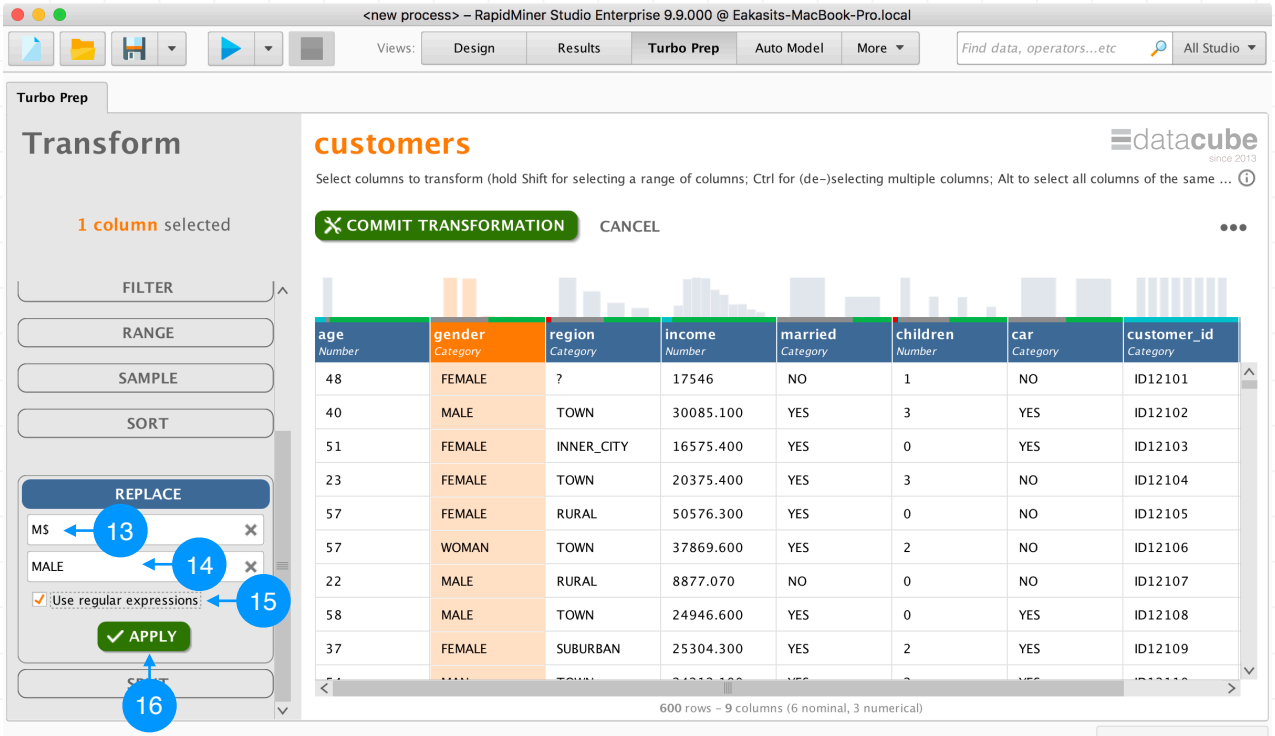


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เปลี่ยนค่าในส่วนของ Replace ด้านซ้ายมือ

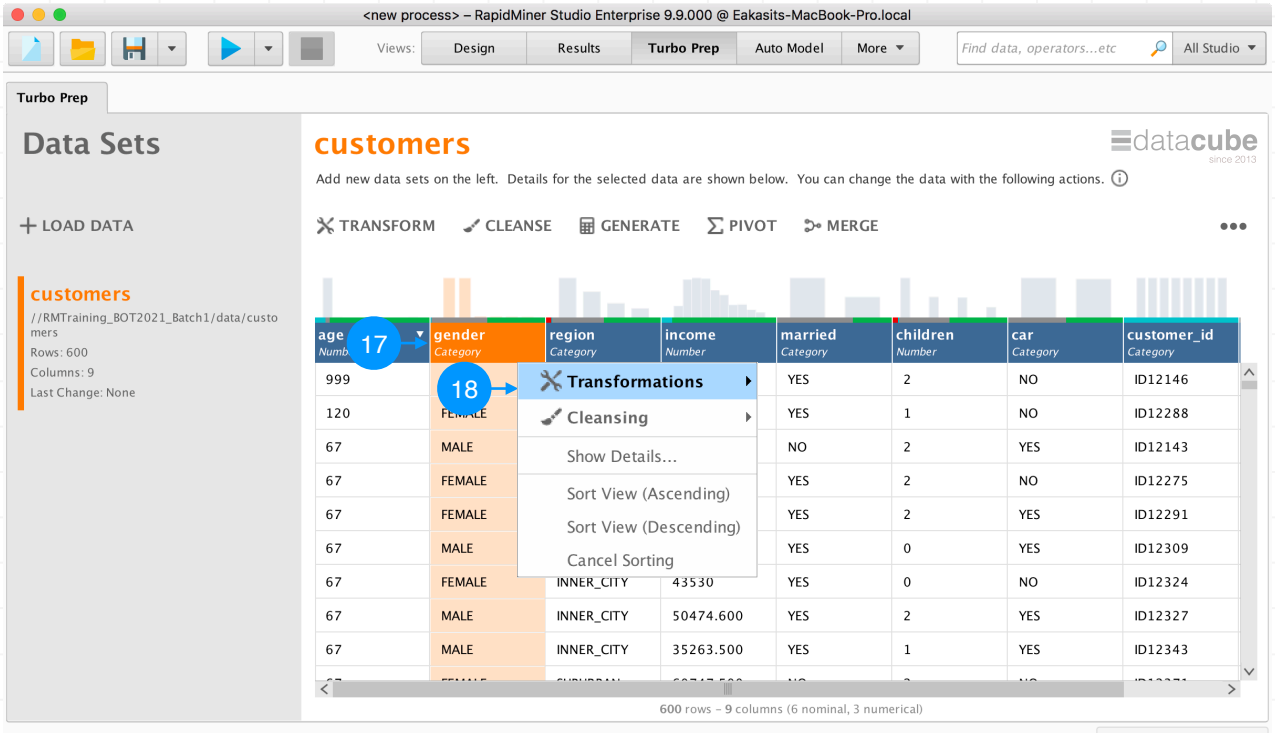


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Gender** และคลิกขวา เลือกเมนู **Transformations**

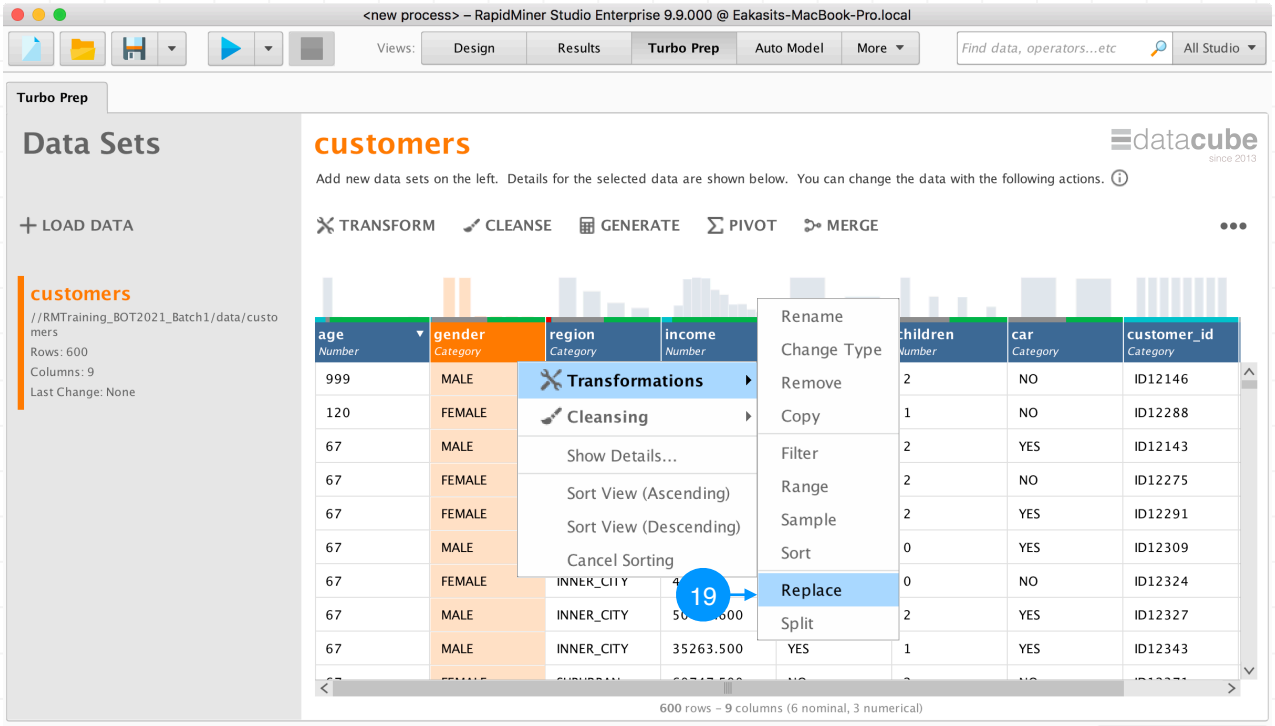


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกเมนู **Replace**

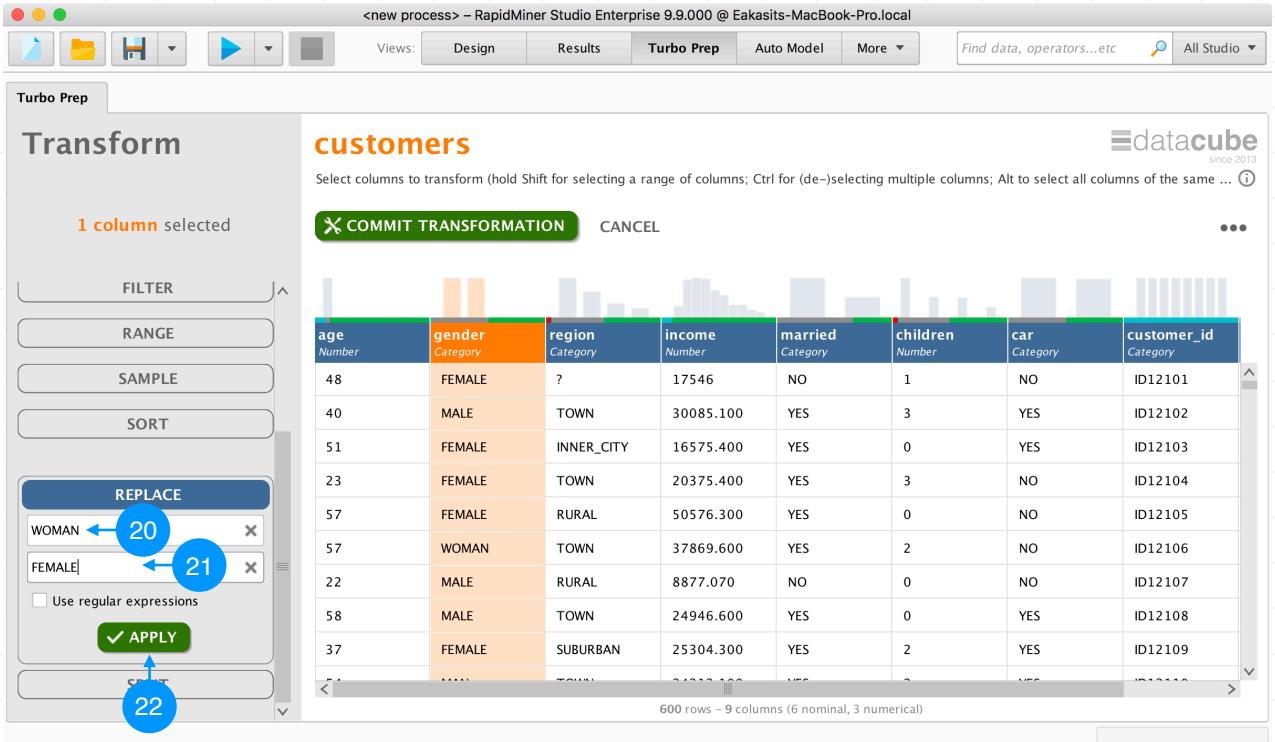


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เปลี่ยนค่าในส่วนของ Replace ด้านซ้ายมือ

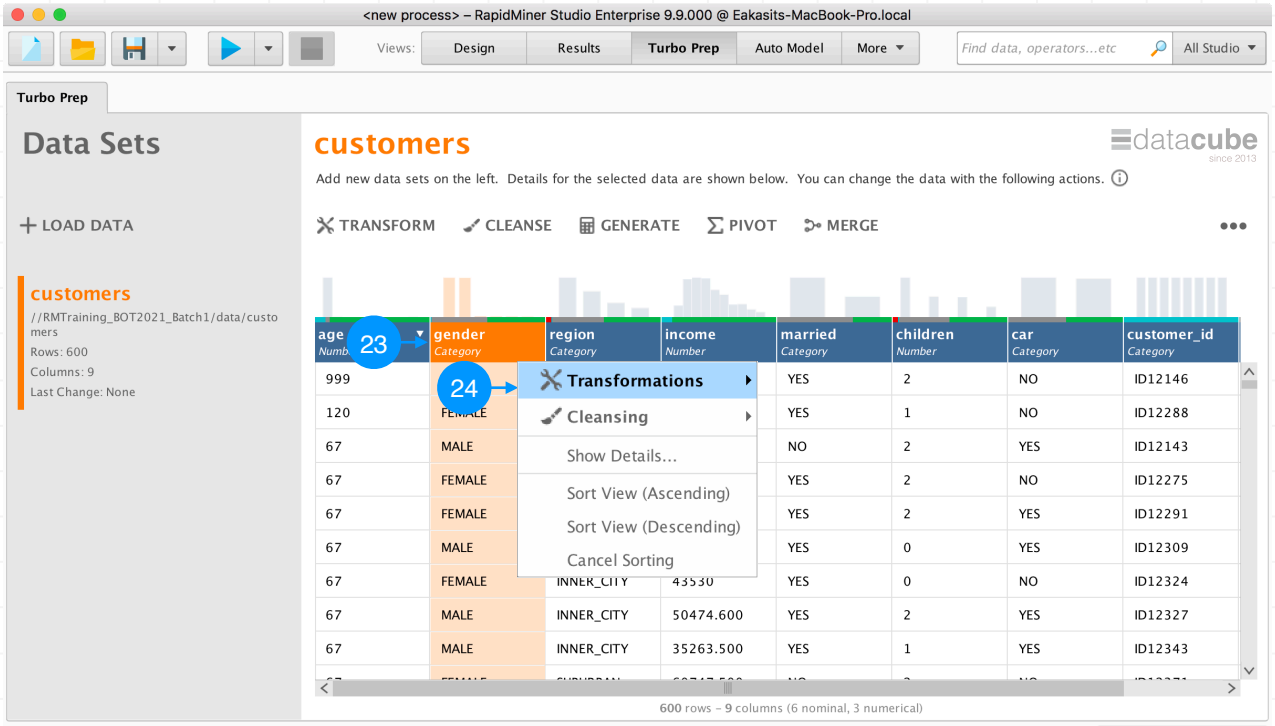


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Gender** และคลิกขวา เลือกเมนู **Transformations**

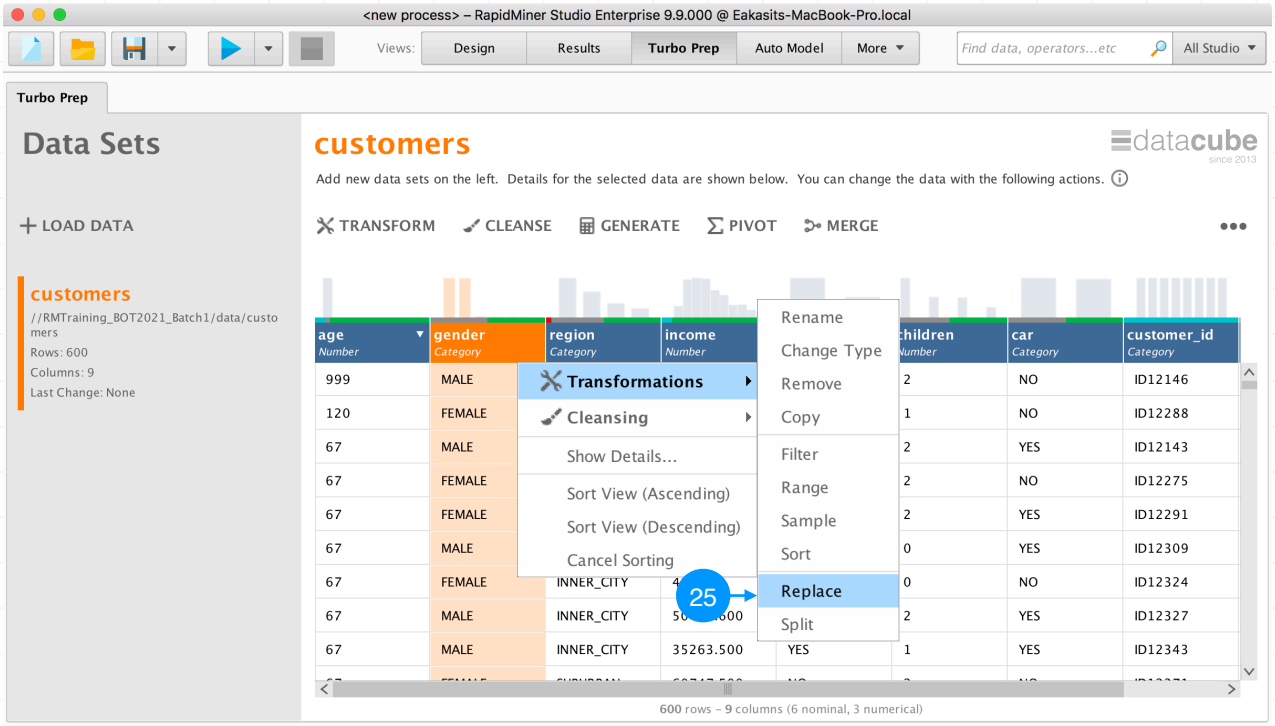


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกเมนู **Replace**

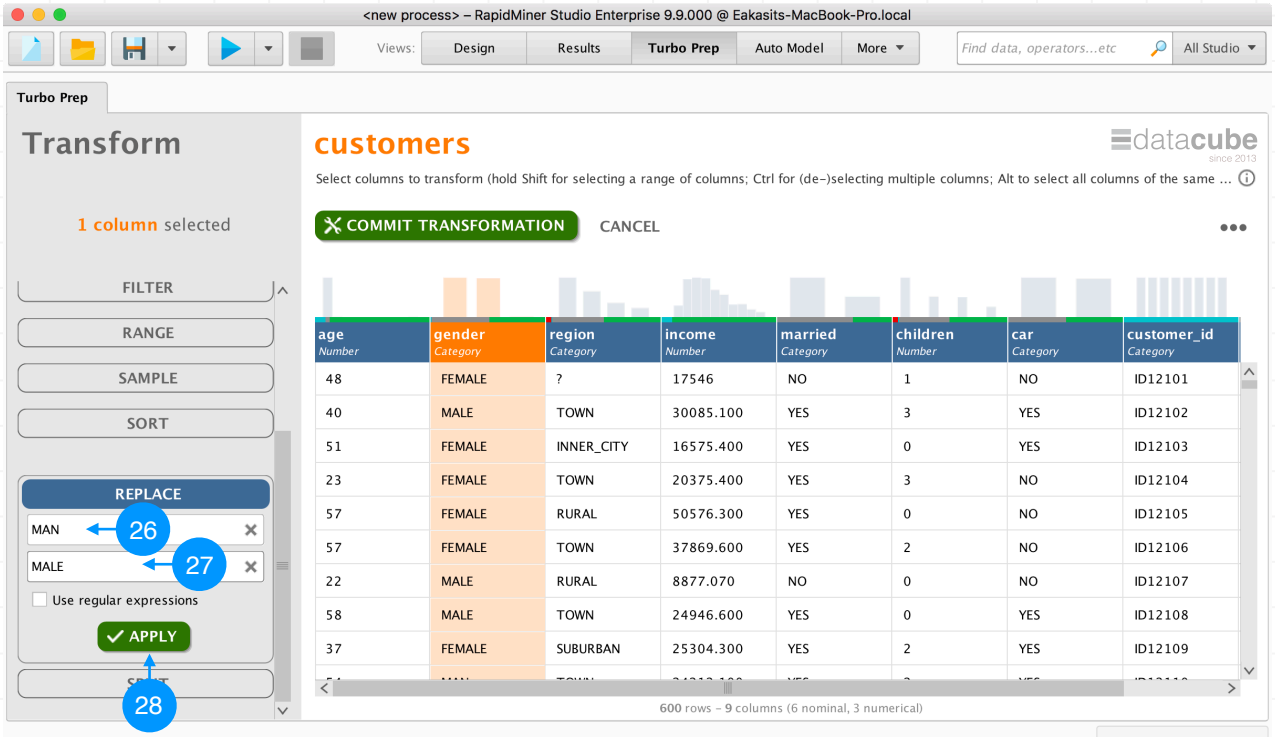


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เปลี่ยนค่าในส่วนของ Replace ด้านซ้ายมือ

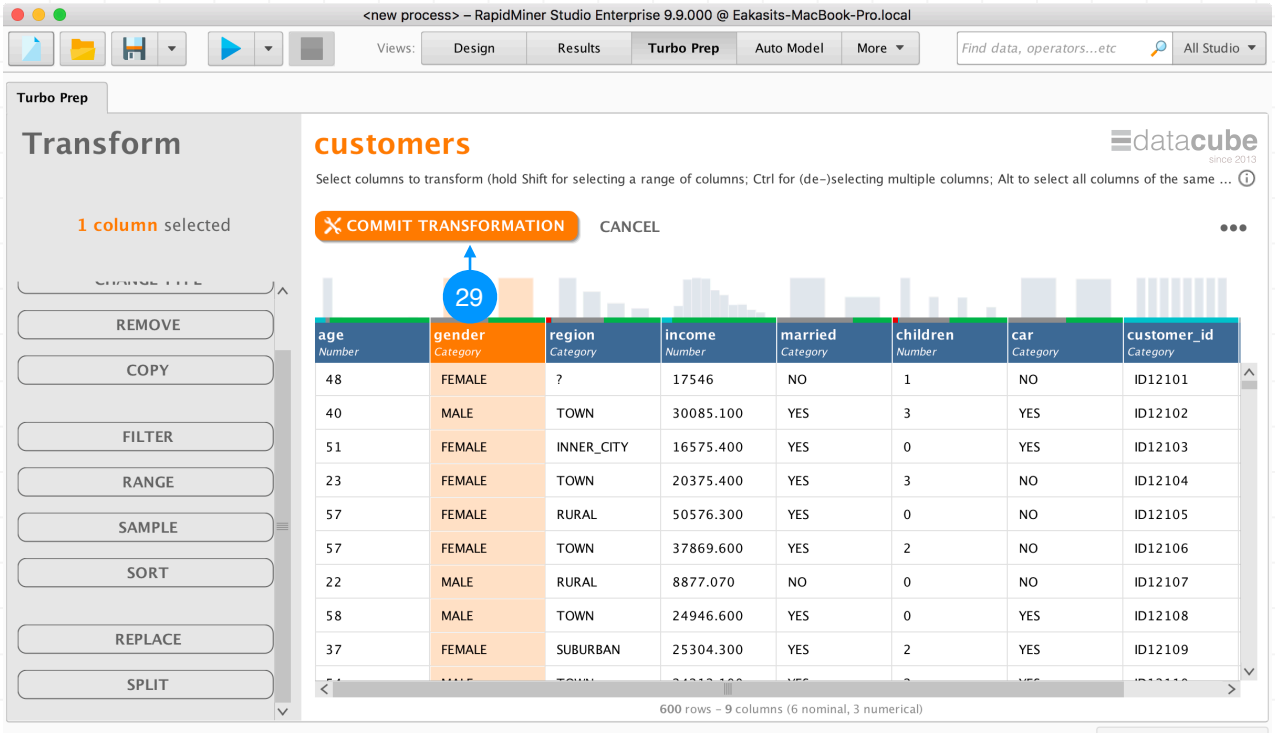


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- กดปุ่ม **COMMIT TRAMSFORMATION**

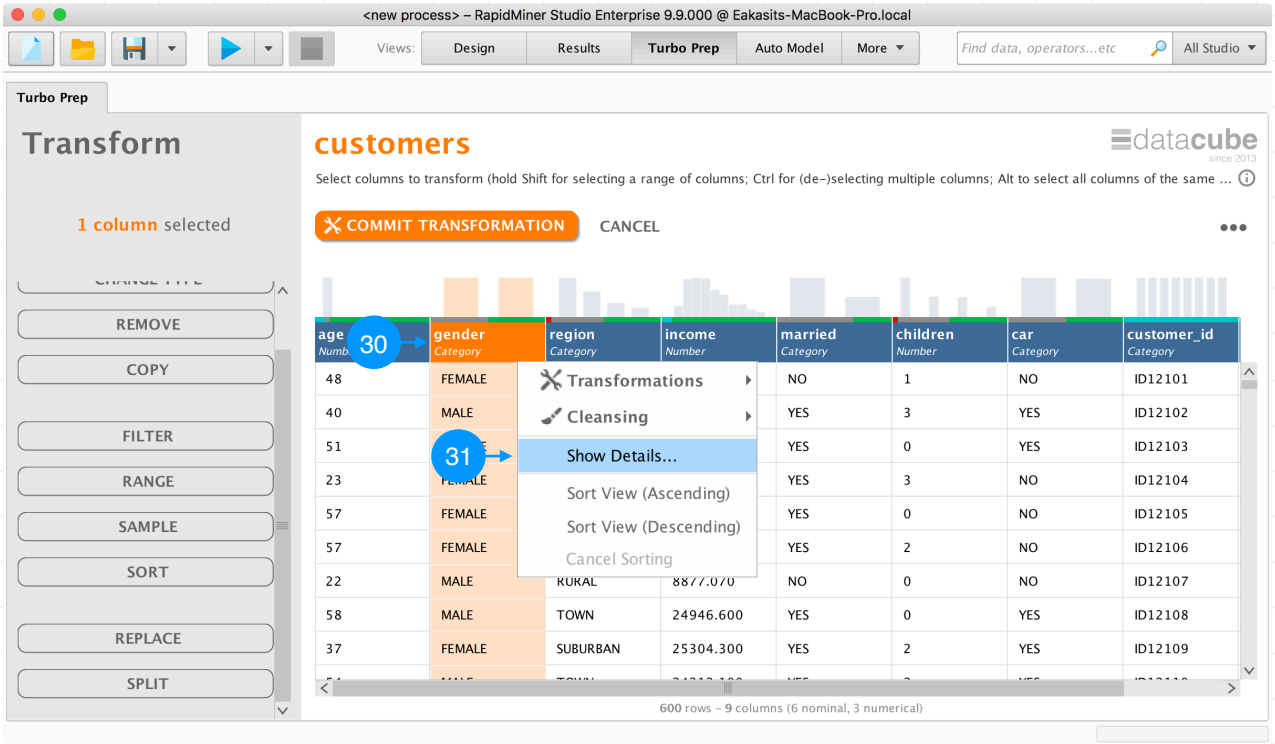


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Gender** และคลิกขวา เลือกเมนู **Show Details...**

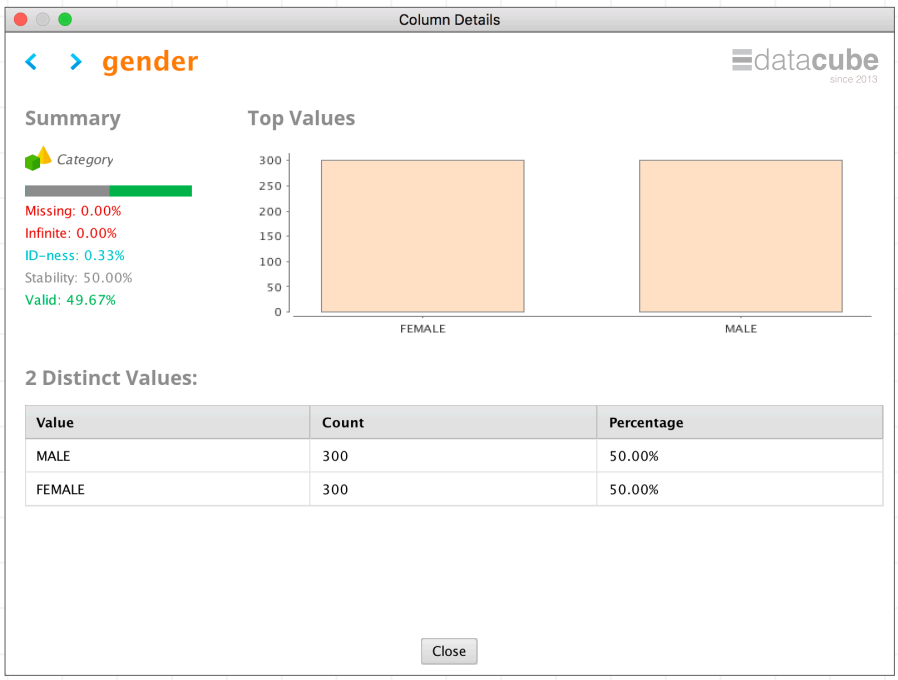


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลจะเหลือแค่ FEMALE และ MALE



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data Preparation

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- การเตรียมข้อมูลด้วย Turbo Prep
  - การดูรายละเอียดของข้อมูล (show detail)
  - การเรียงลำดับข้อมูล (sort)
  - การแทนค่าข้อมูล (replace)
  - **การแทนค่าว่าง (replace missing value)**
  - การเลือกข้อมูลบางตัวอย่าง (filter examples)
  - การ join ข้อมูล 2 ชุดเข้าด้วยกัน
  - การ Export Data และ Process

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace missing value

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลอาจจะขาดหายไป เนื่องจาก
  - ไม่มีข้อมูล
  - กรอกข้อมูลผิดพลาด

customer_id	response	age	gender	region	income	married	children	car
ID12101	NO	48	FEMALE		17546.000	NO	1	NO
ID12102	NO	40	MALE	TOWN	30085.100	YES	3	YES
ID12103	YES	51	FEMALE	INNER_CITY	16575.400	YES	0	YES
...	...	...	...		...	...	...	...
ID12110	YES	54	MAN	TOWN	24212.100	YES	2	YES
...	...	...	...		...	...	...	...
ID12120	YES	31	M	TOWN	22522.800	YES	0	YES
...	...	...	...		...	...	...	...
ID12138	NO	36	FEMALE	RURAL	13381.000	NO		YES

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace missing value

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

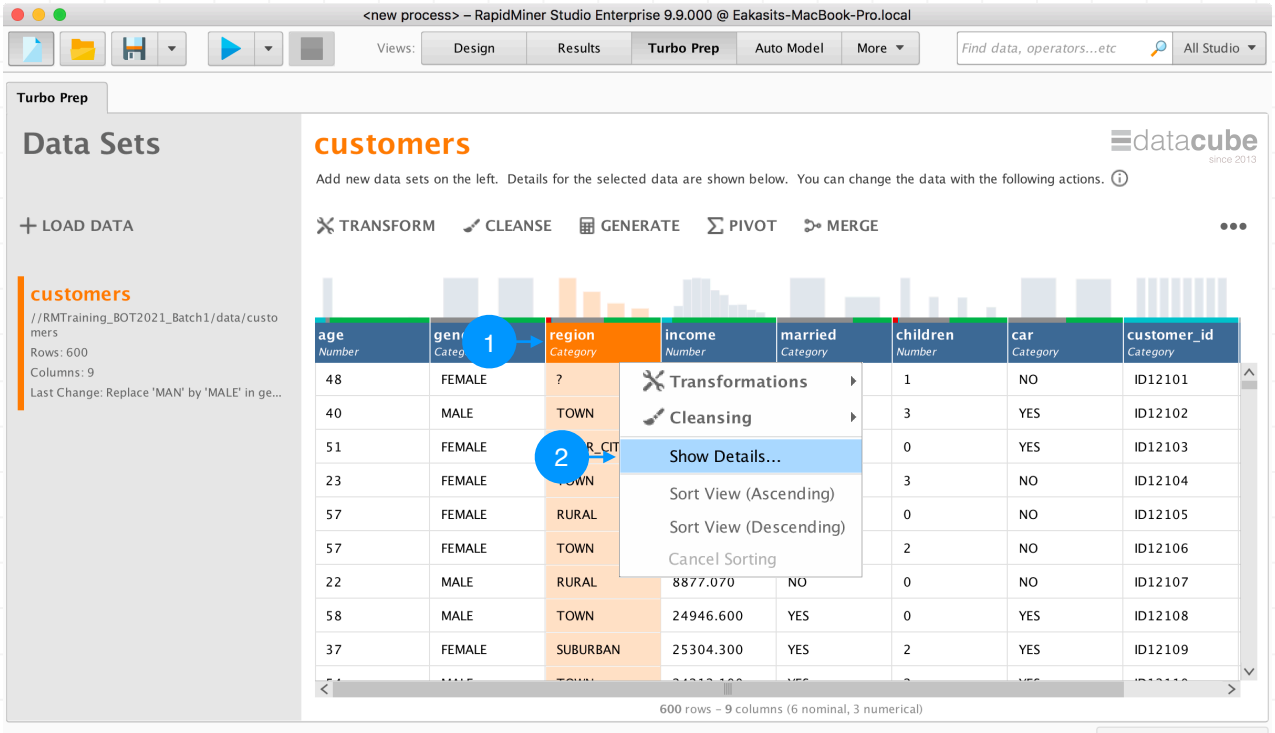
- การแทนที่ข้อมูลที่ขาดหายไป
  - แทนที่ด้วย**ค่าที่ระบุเอง** เช่น *ไม่ระบุ* หรือ *N/A*
  - แทนที่ด้วย**ค่าเฉลี่ย (mean)** ในกรณีแอตทริบิวต์เป็น**ตัวเลข (Number)**
  - แทนที่ด้วย**ค่าฐานนิยม (mode)** ในกรณีแอตทริบิวต์เป็น **Category**
  - แทนที่ด้วย**ค่า 0 (zero)**

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Region** และคลิกขวา เลือกเมนู **Show Details...**



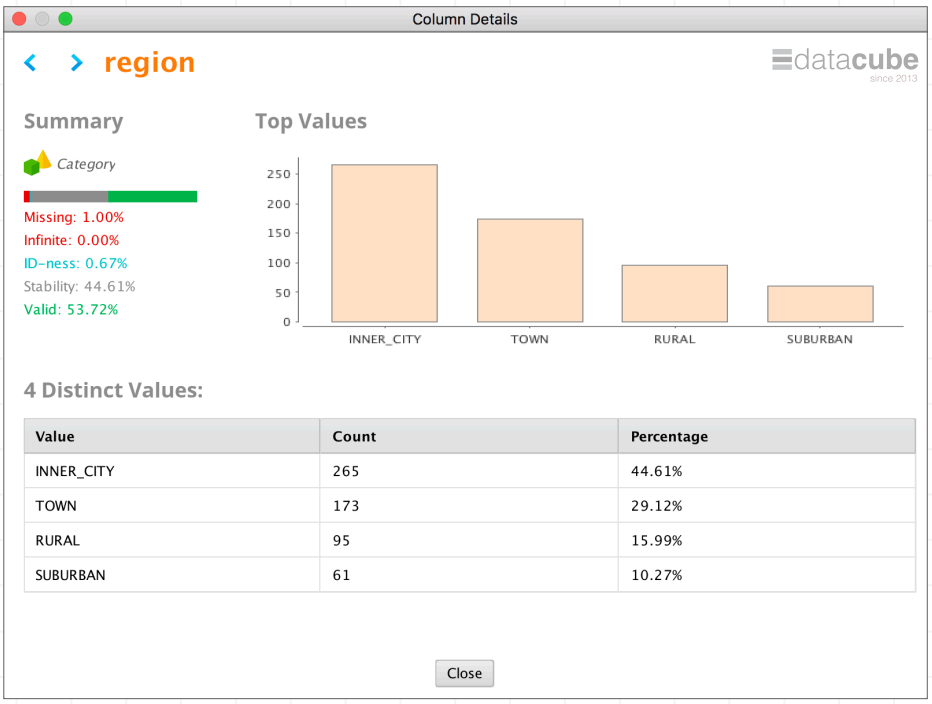
age Number	gen Categ	region Category	income Number	married Category	children Number	car Category	customer_id Category
48	FEMALE	?			1	NO	ID12101
40	MALE	TOWN			3	YES	ID12102
51	FEMALE	R_CIT			0	YES	ID12103
23	FEMALE	TOWN			3	NO	ID12104
57	FEMALE	RURAL			0	NO	ID12105
57	FEMALE	TOWN			2	NO	ID12106
22	MALE	RURAL	8877.070	NO	0	NO	ID12107
58	MALE	TOWN	24946.600	YES	0	YES	ID12108
37	FEMALE	SUBURBAN	25304.300	YES	2	YES	ID12109

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลในแอตทริบิวต์ **Region** มีค่าว่าง (missing value) อยู่ 1%

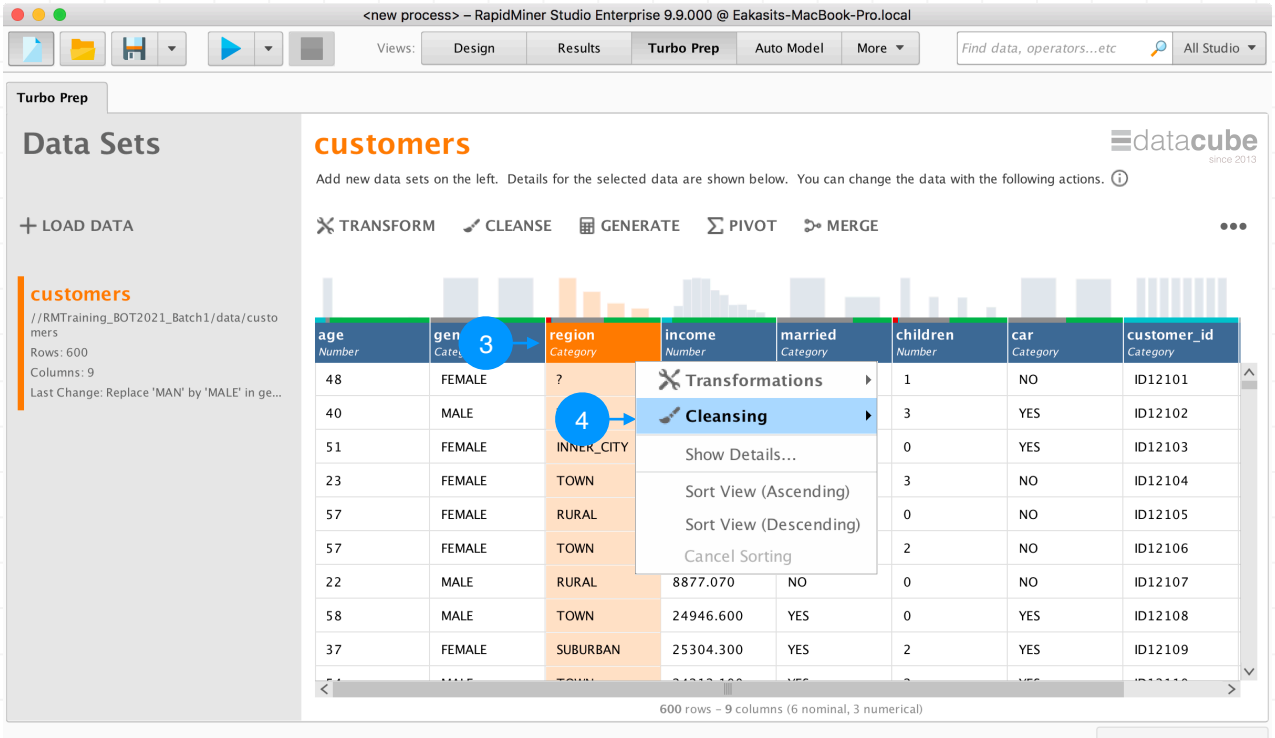


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Region** และคลิกขวา เลือกเมนู **Cleansing**



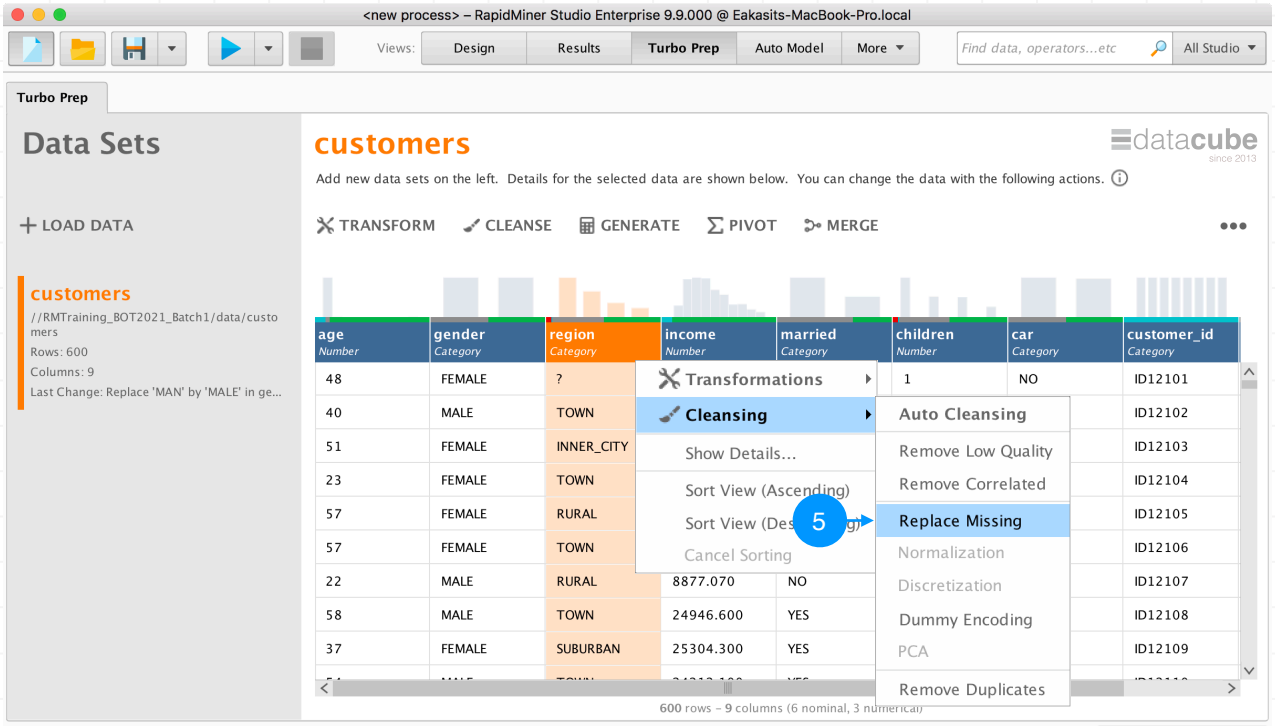
age Number	gen Category	region Category	income Number	married Category	children Number	car Category	customer_id Category
48	FEMALE	?			1	NO	ID12101
40	MALE				3	YES	ID12102
51	FEMALE	INNER_CITY			0	YES	ID12103
23	FEMALE	TOWN			3	NO	ID12104
57	FEMALE	RURAL			0	NO	ID12105
57	FEMALE	TOWN			2	NO	ID12106
22	MALE	RURAL	8877.070	NO	0	NO	ID12107
58	MALE	TOWN	24946.600	YES	0	YES	ID12108
37	FEMALE	SUBURBAN	25304.300	YES	2	YES	ID12109

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกเมนู **Replace Missing**

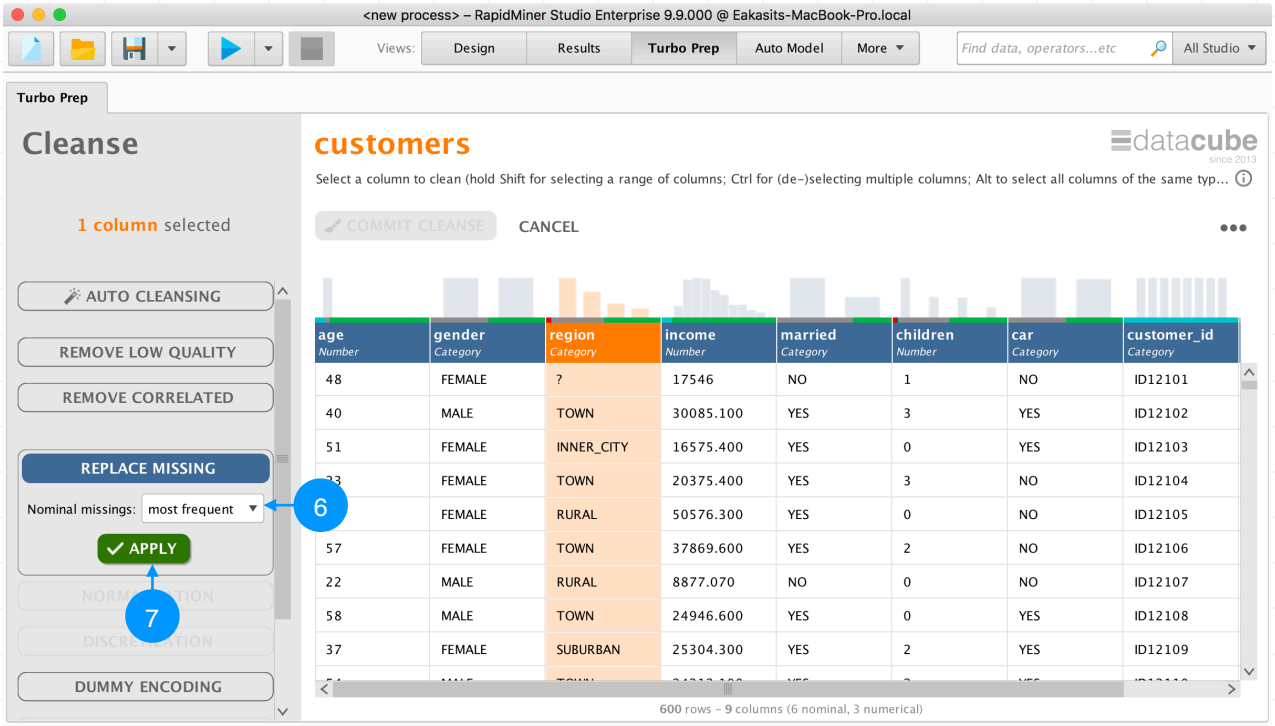


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เปลี่ยนค่าในส่วนของ Replace Missing ด้านซ้ายมือ

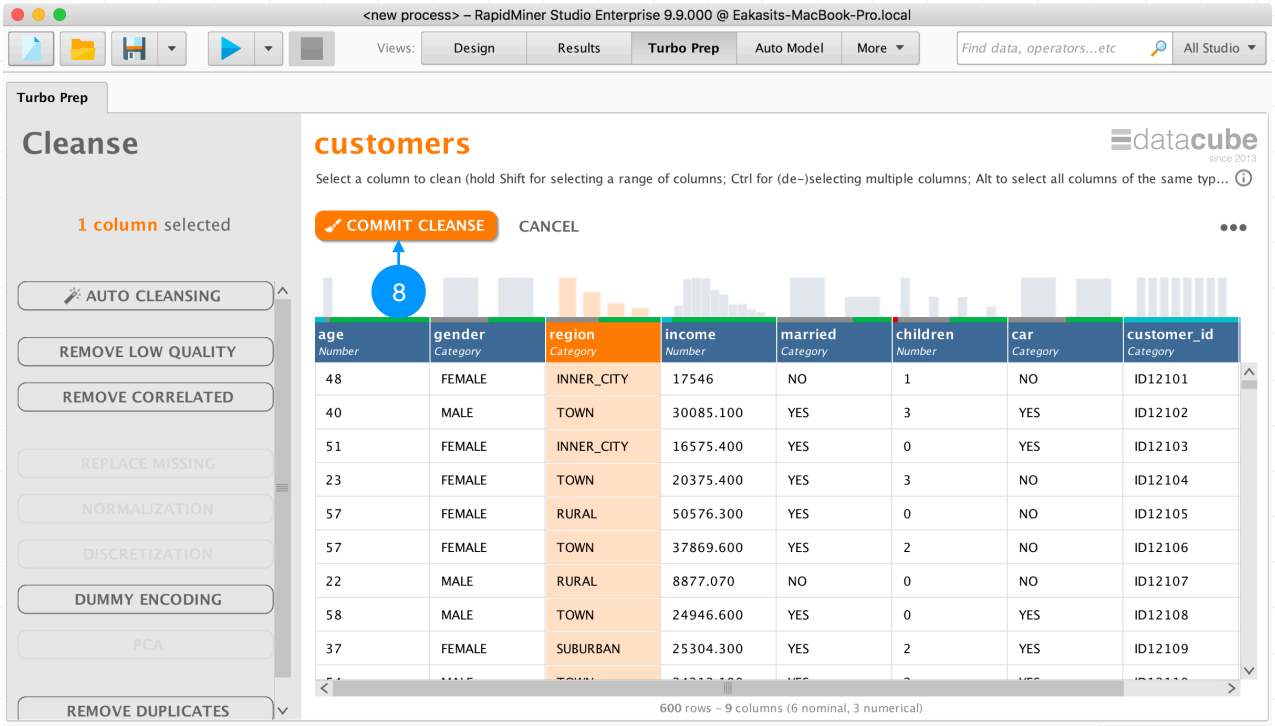


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- กดปุ่ม **COMMIT CLEANSE**

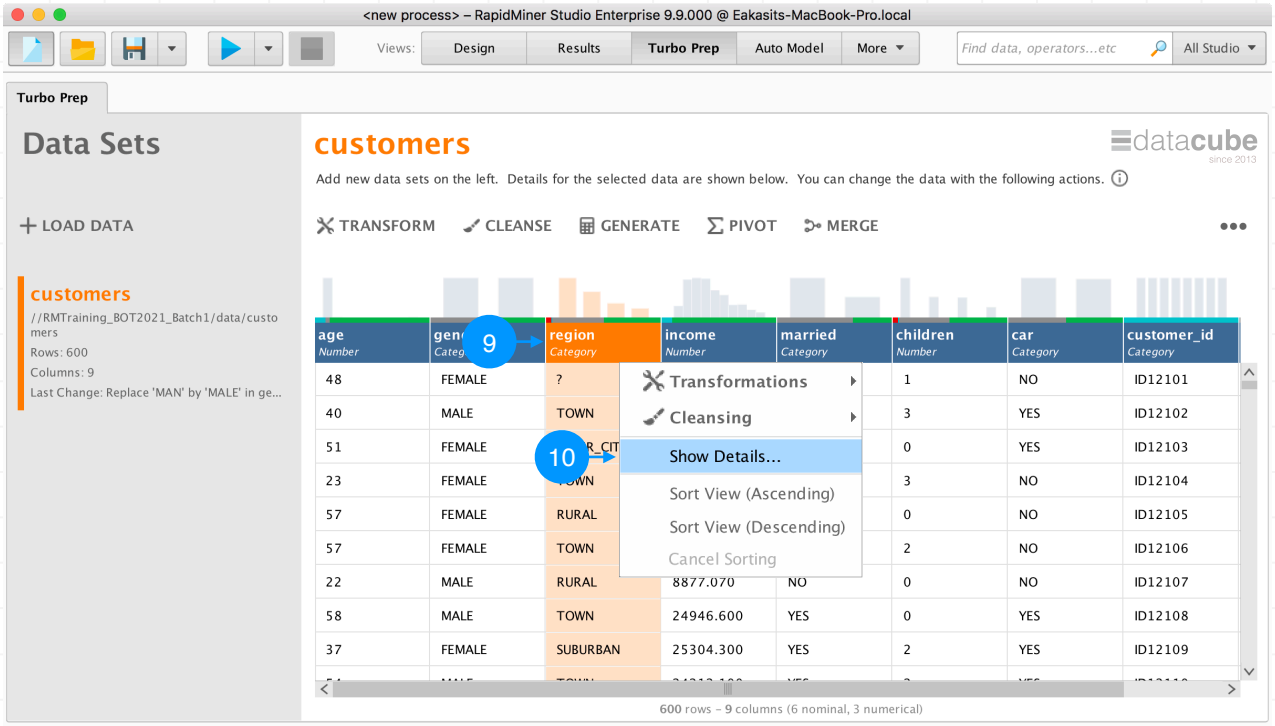


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Region** และคลิกขวา เลือกเมนู **Show Details...**

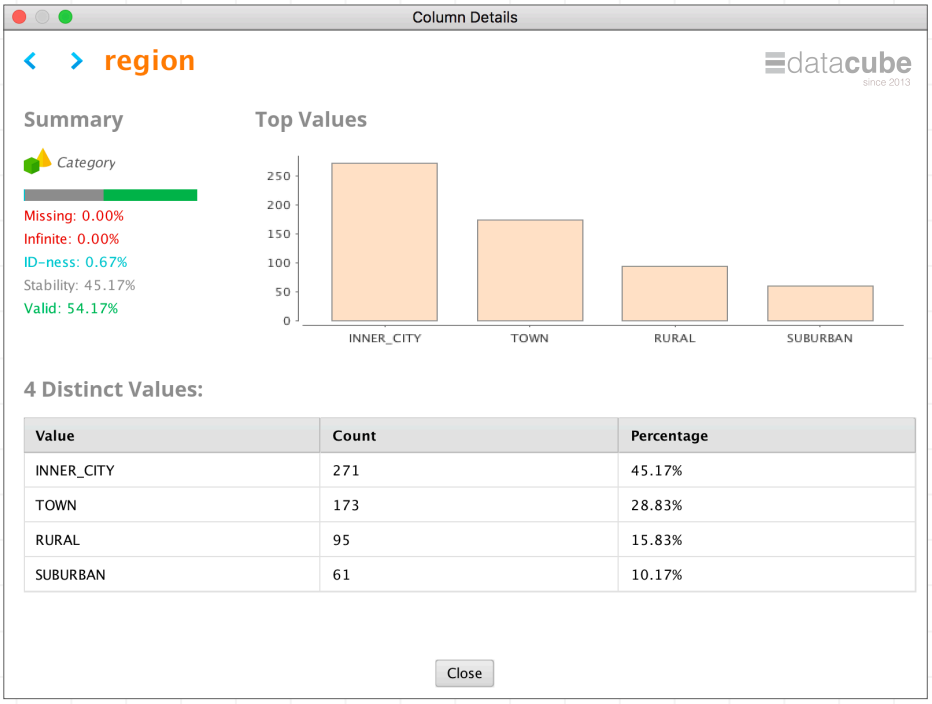


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลในแอตทริบิวต์ **Region** ไม่มีค่าว่าง (missing value) แล้ว

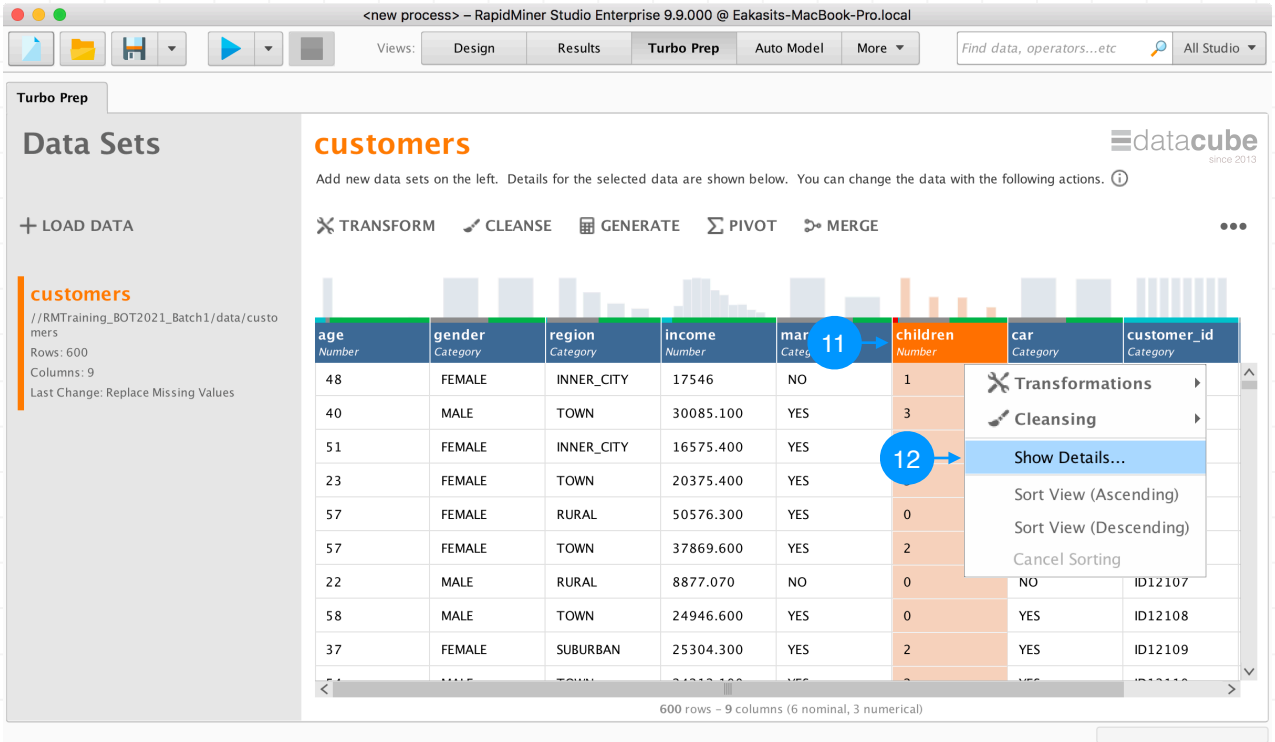


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Children** และคลิกขวา เลือกเมนู **Show Details...**

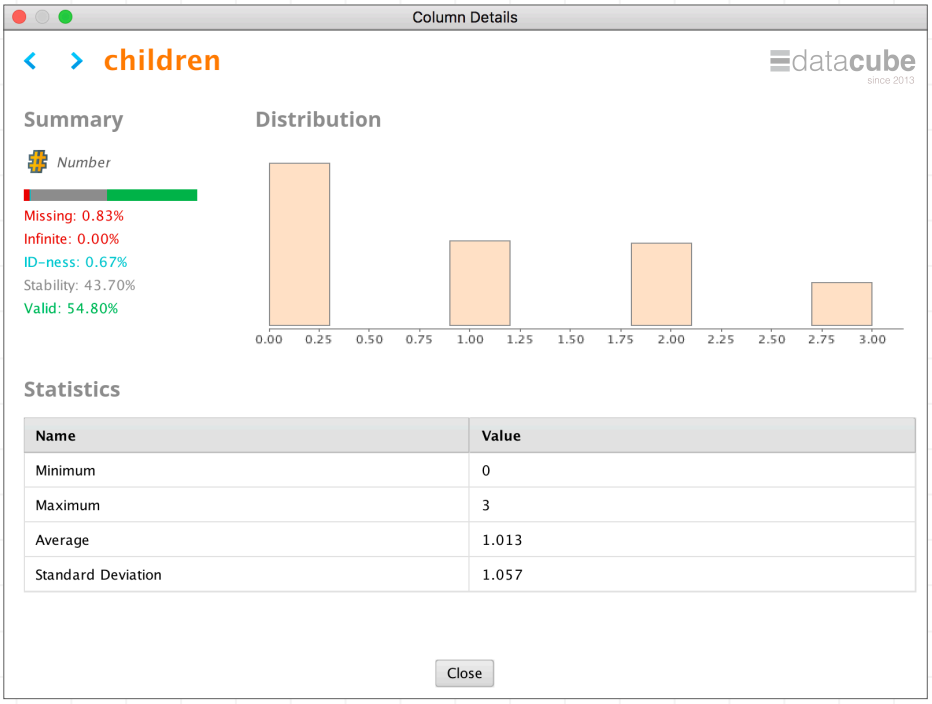


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลในแอตทริบิวต์ **Children** มีค่าว่าง (missing value) อยู่ 0.83%

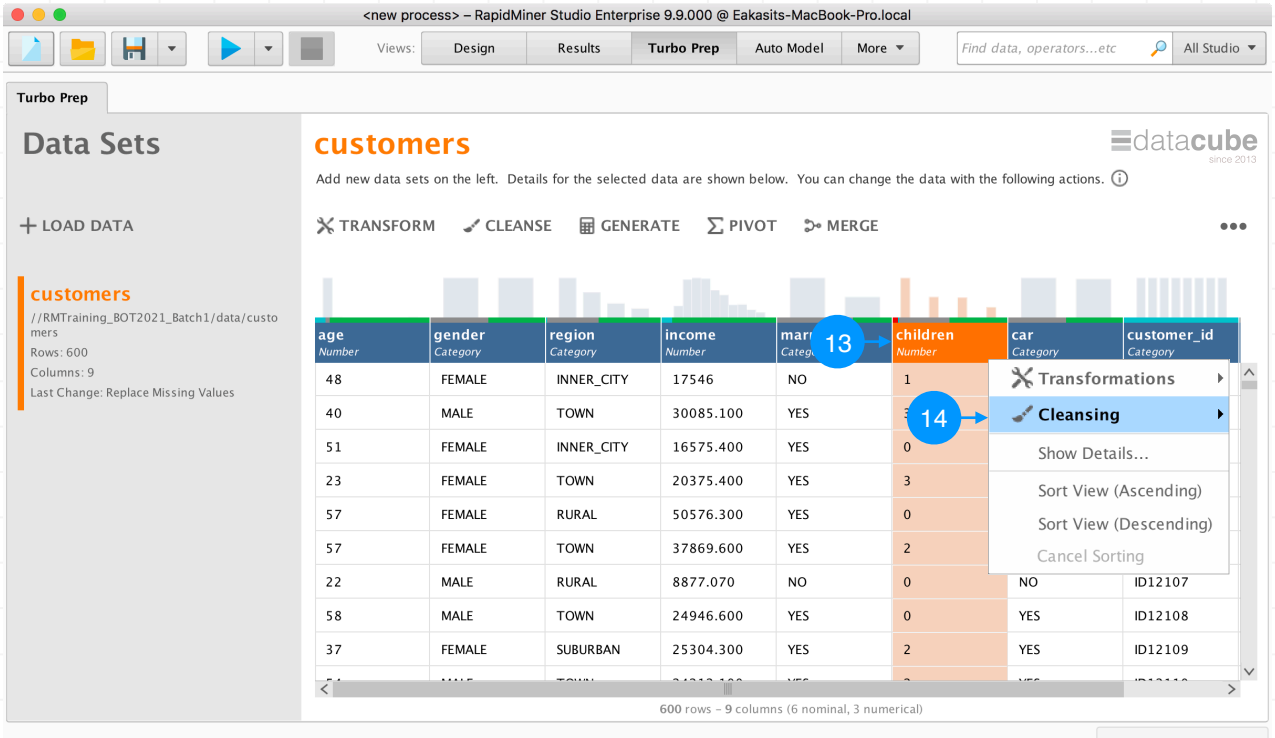


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Children** และคลิกขวา เลือกเมนู **Cleansing**

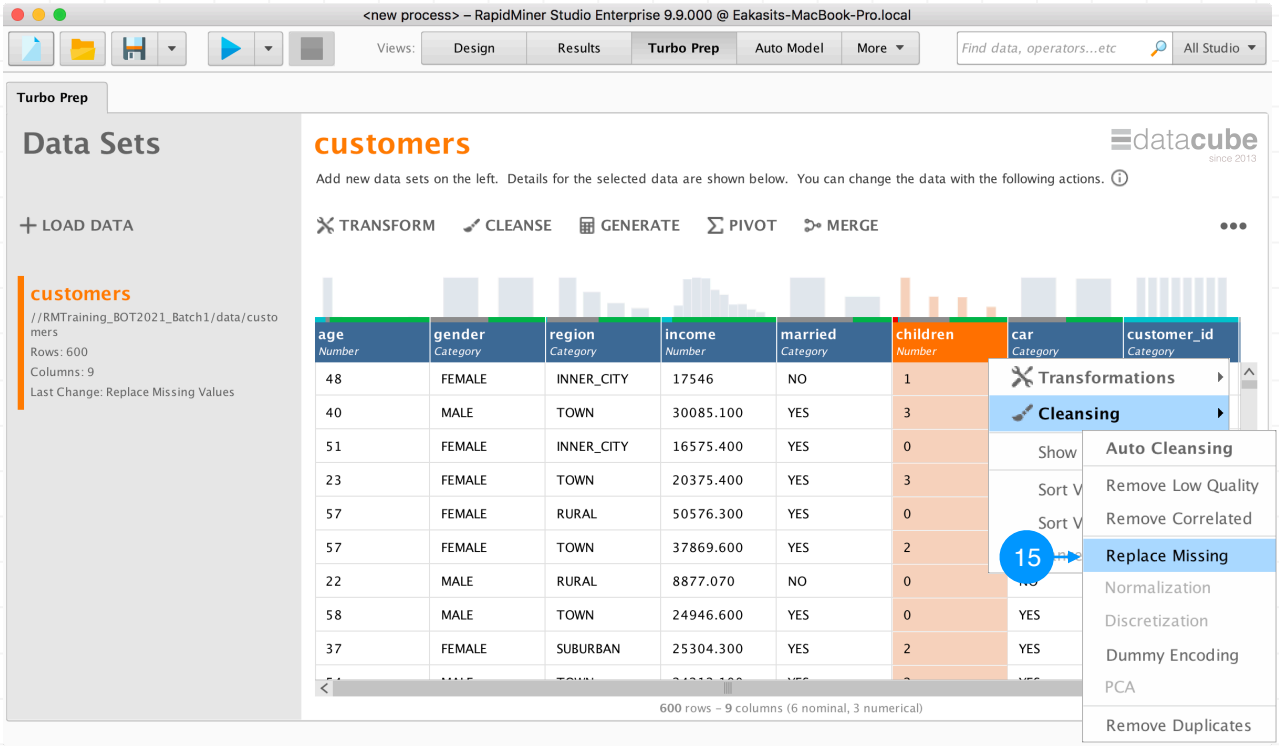


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกเมนู **Replace Missing**

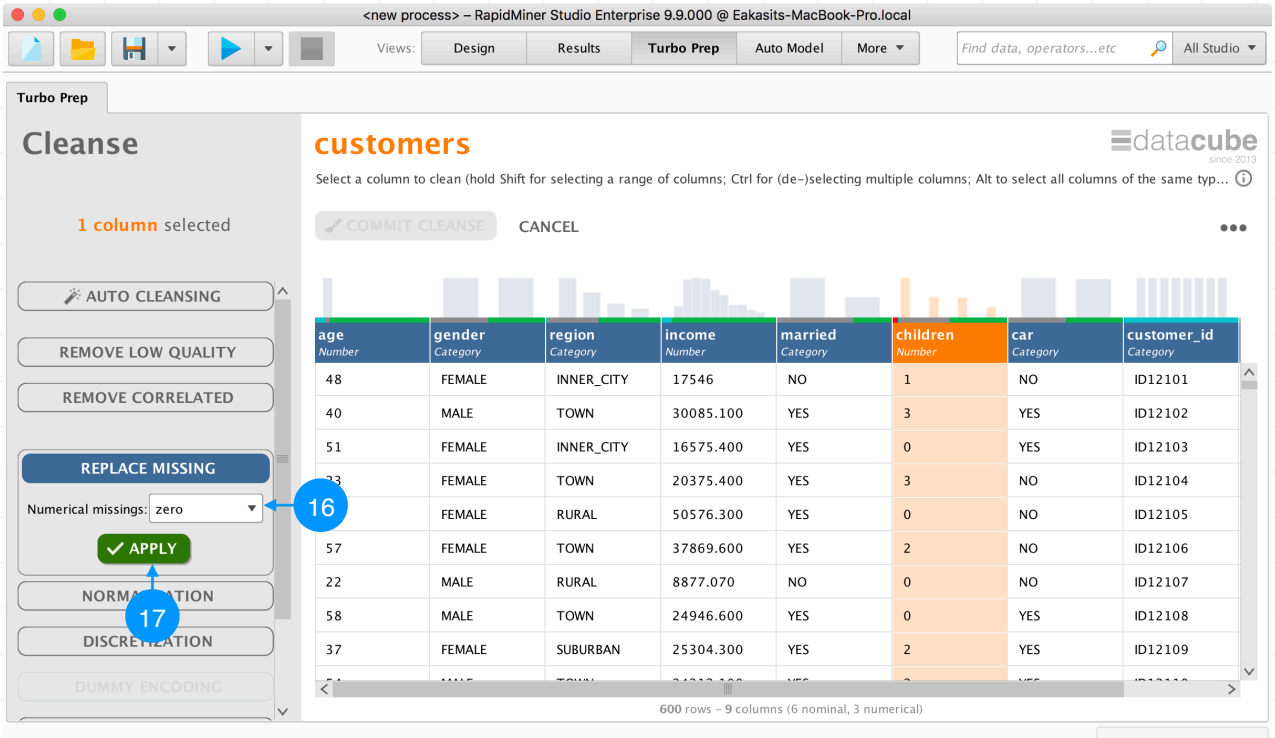


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เปลี่ยนค่าในส่วนของ Replace Missing ด้านซ้ายมือ

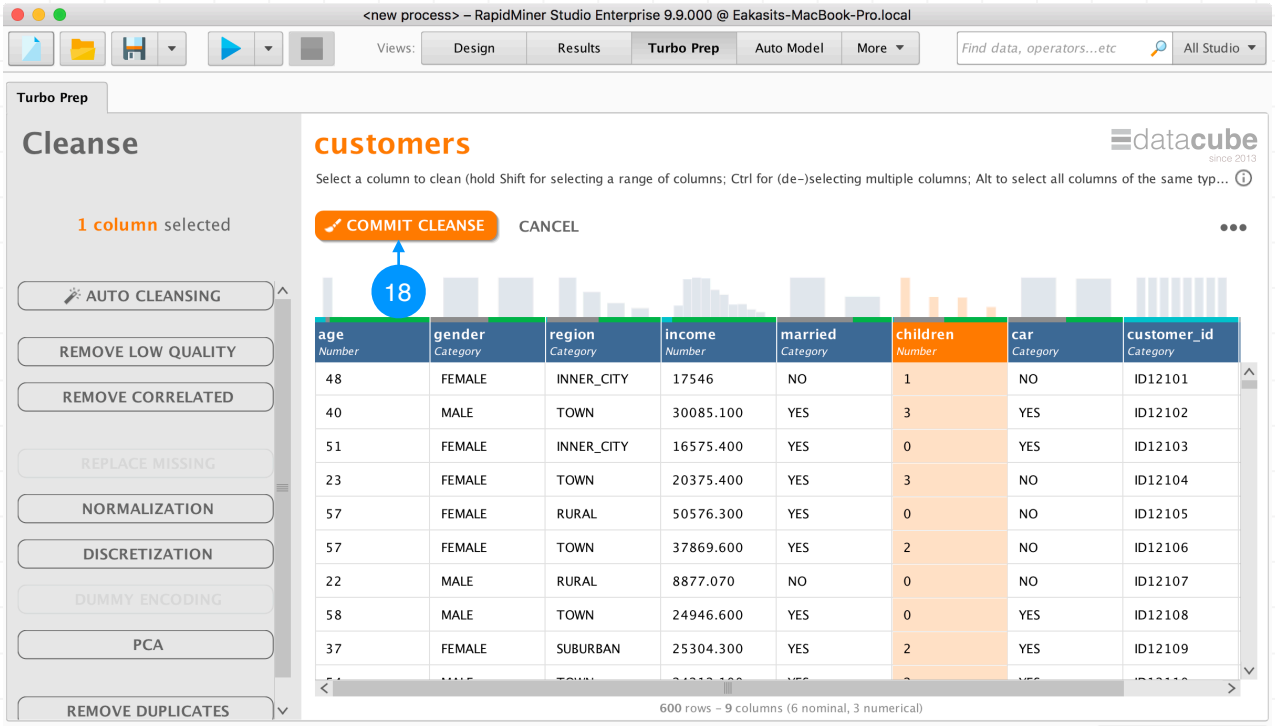


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- กดปุ่ม **COMMIT CLEANSE**

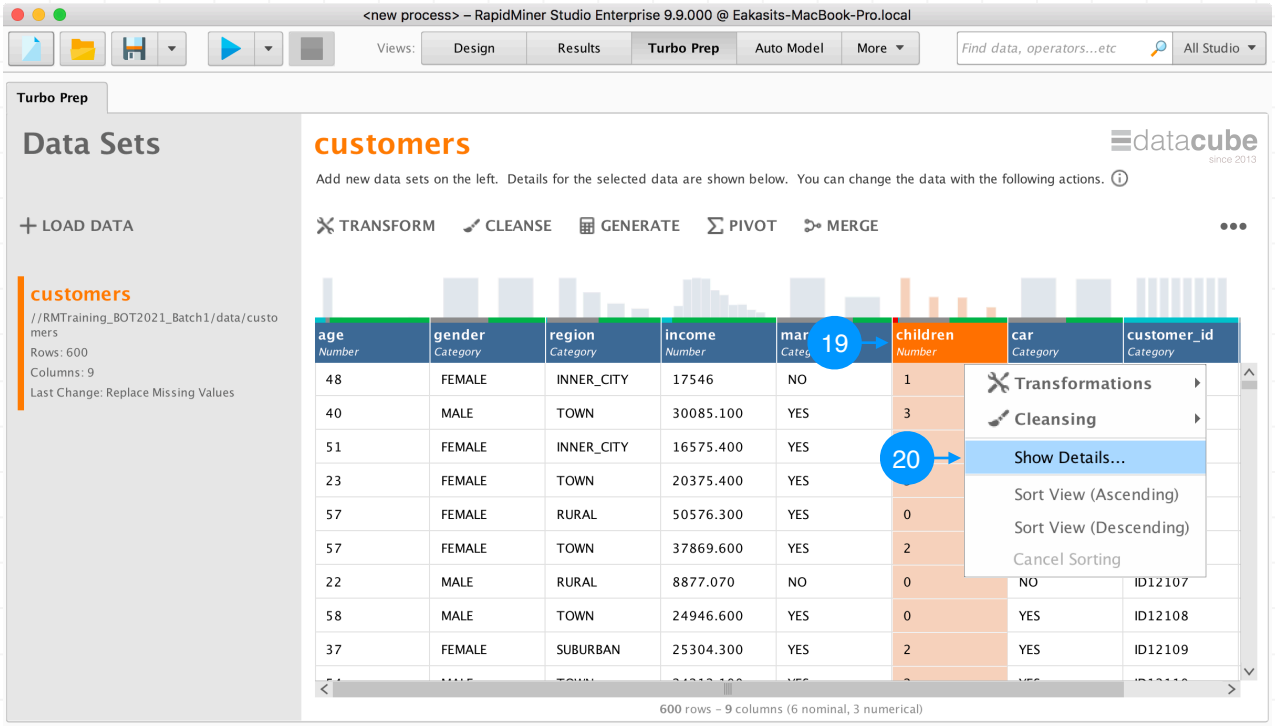


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Children** และคลิกขวา เลือกเมนู **Show Details...**

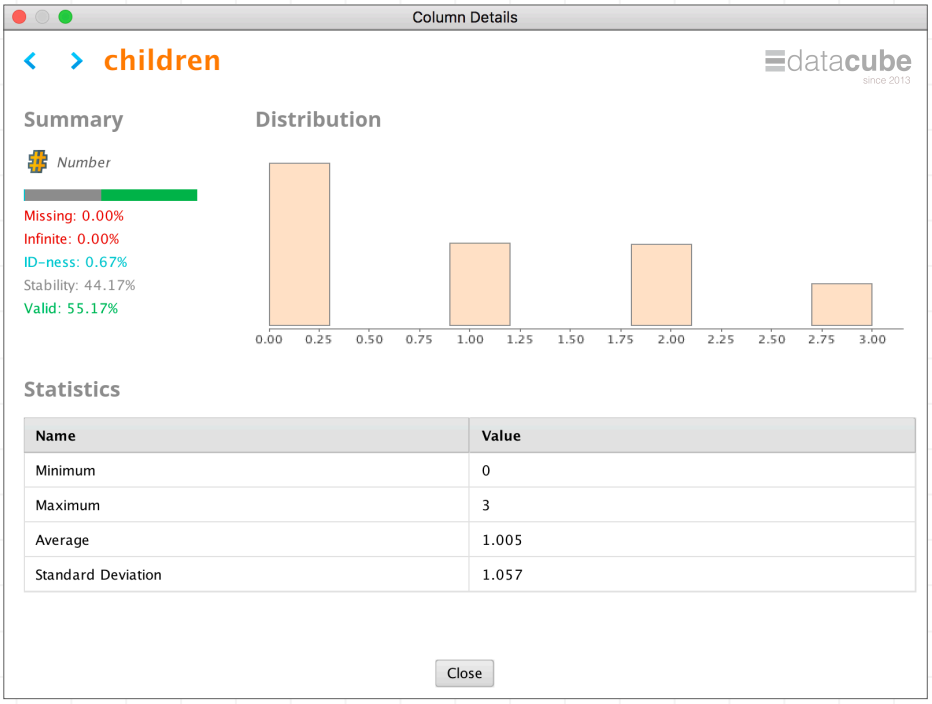


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Replace Missing Values

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- ข้อมูลในแอตทริบิวต์ **Children** ไม่มีค่าว่าง (missing value) แล้ว



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data Preparation

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

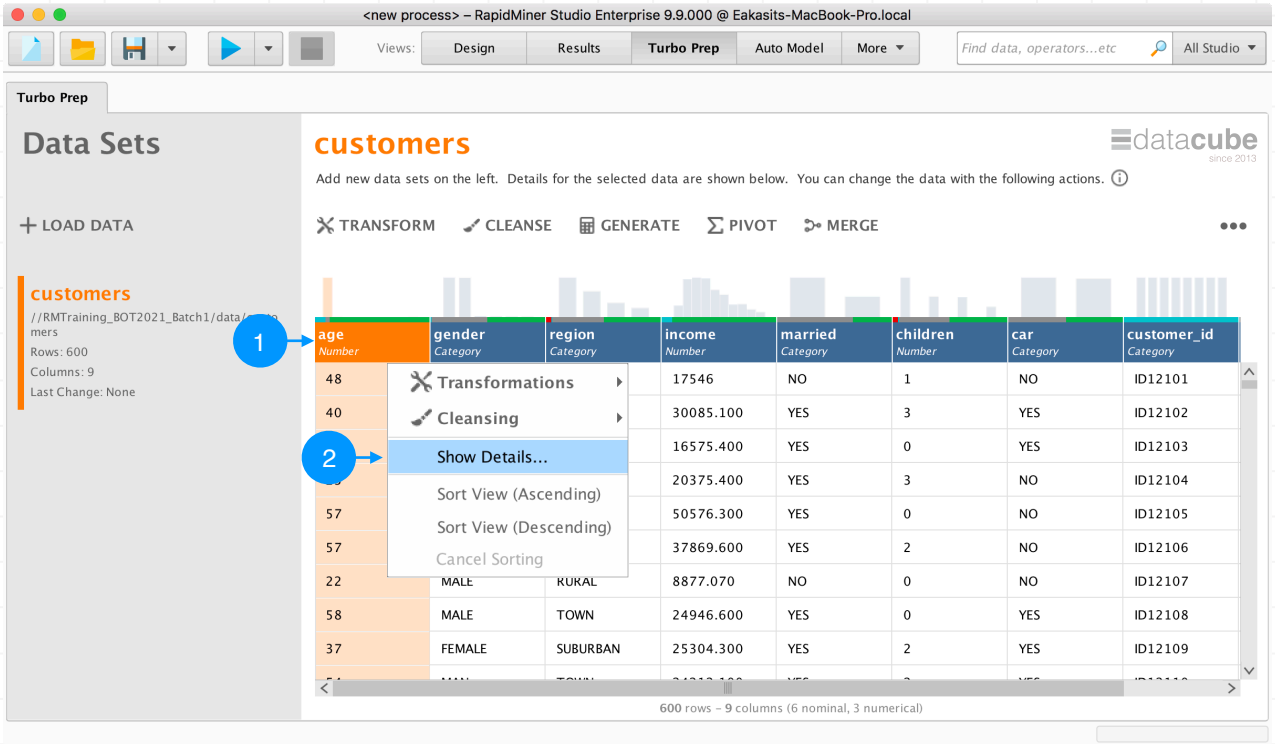
- การเตรียมข้อมูลด้วย Turbo Prep
  - การดูรายละเอียดของข้อมูล (show detail)
  - การเรียงลำดับข้อมูล (sort)
  - การแทนค่าข้อมูล (replace)
  - การแทนค่าว่าง (replace missing value)
  - **การเลือกข้อมูลบางตัวอย่าง (filter examples)**
  - การ join ข้อมูล 2 ชุดเข้าด้วยกัน
  - การ Export Data และ Process

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Filter Examples

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Age** และคลิกขวา เลือกเมนู **Show Details...**

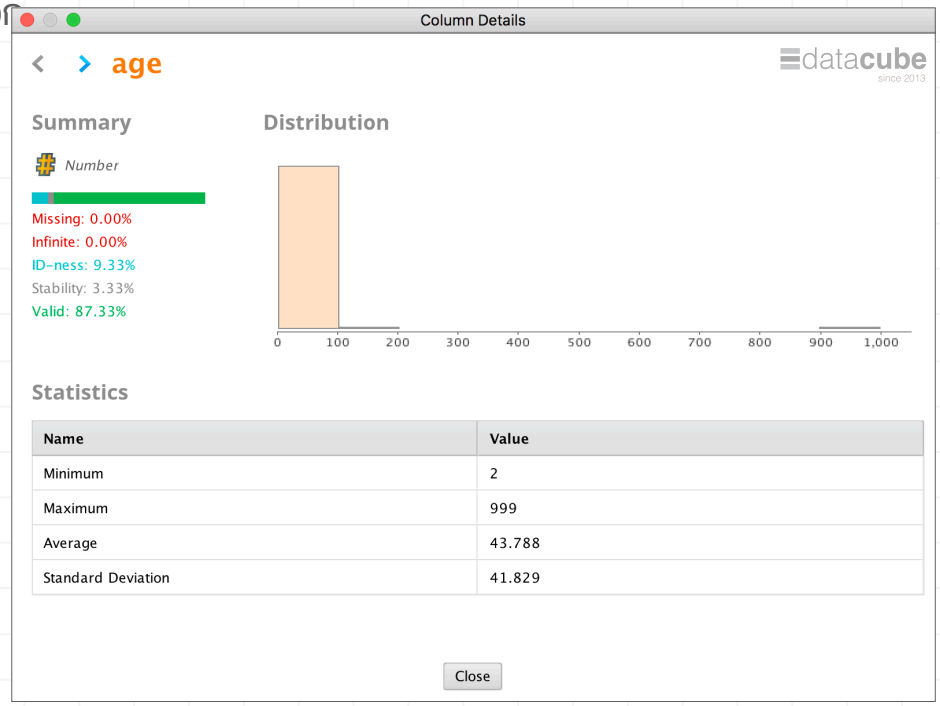


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Filter Examples

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- อายุมีค่าตั้งแต่ 2 - 999 ปี ถือว่าผิดปกติ ดังนั้นจะทำการลบข้อมูลที่ผิดปกติออก

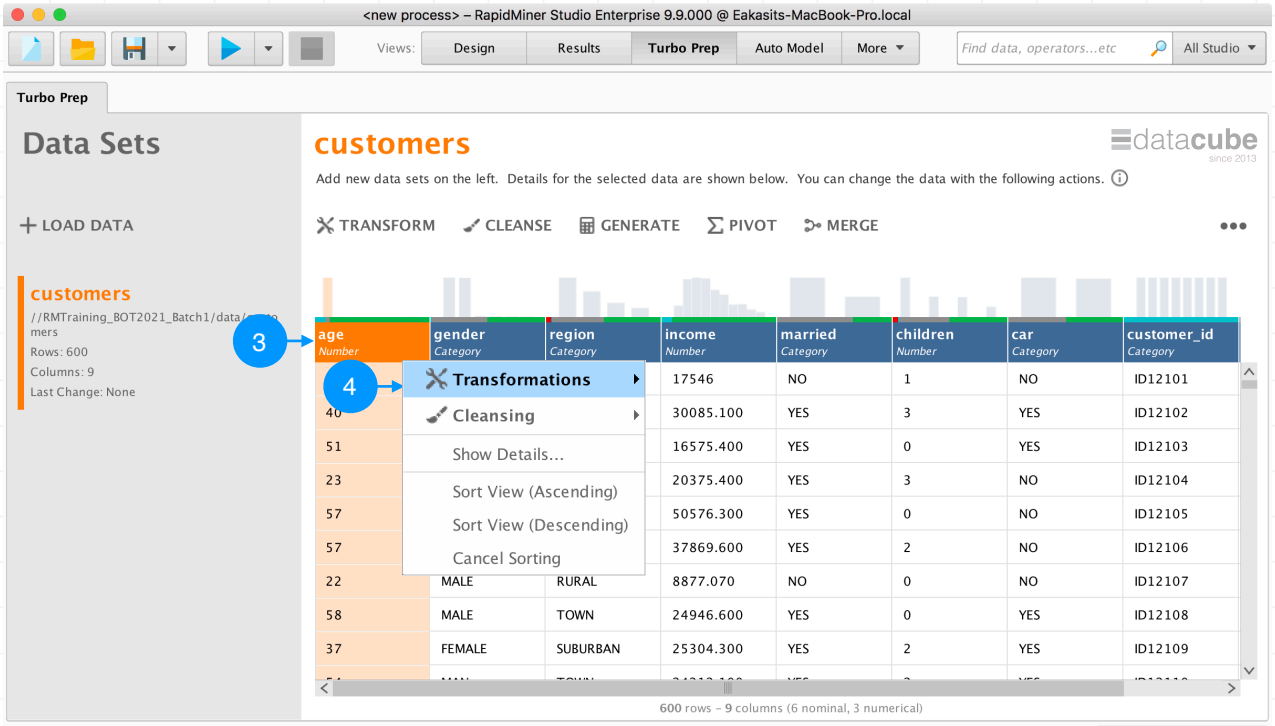


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Filter Examples

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Age** และคลิกขวา เลือกเมนู **Transformations**

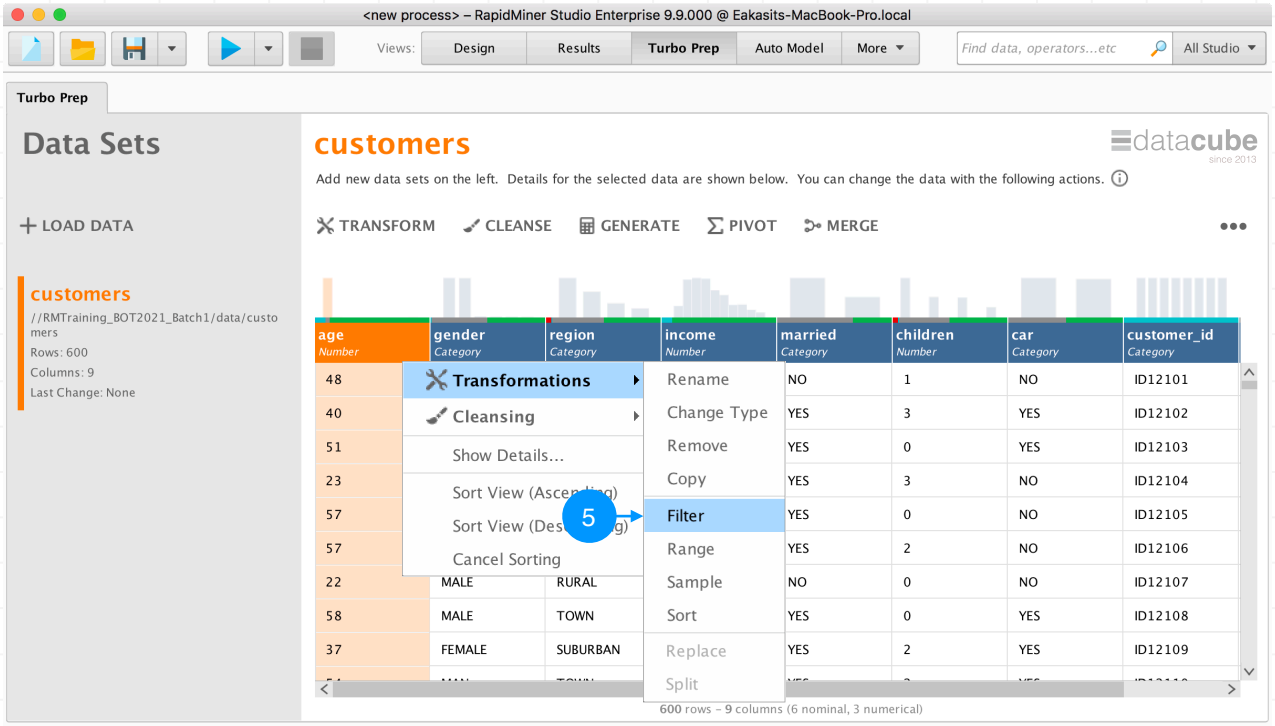


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Filter Examples

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกเมนู **Filter**

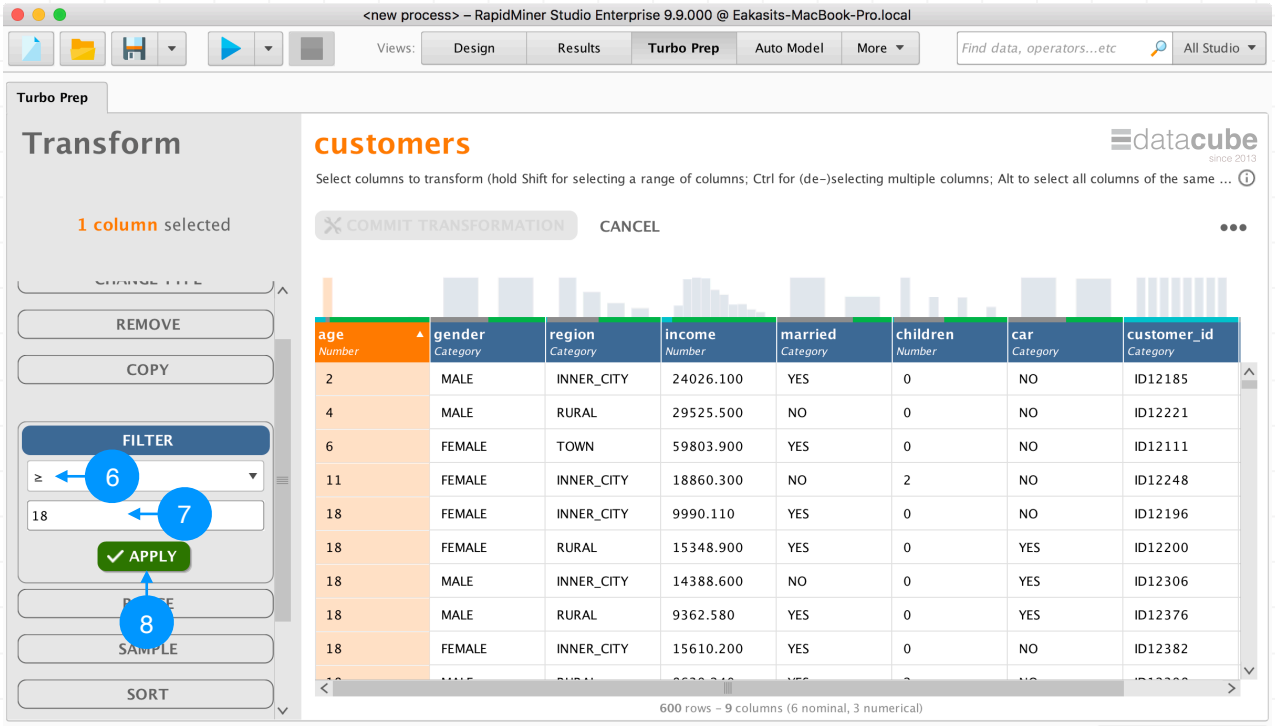


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Filter Examples

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เปลี่ยนค่าในส่วนของ Filter ด้านซ้ายมือ

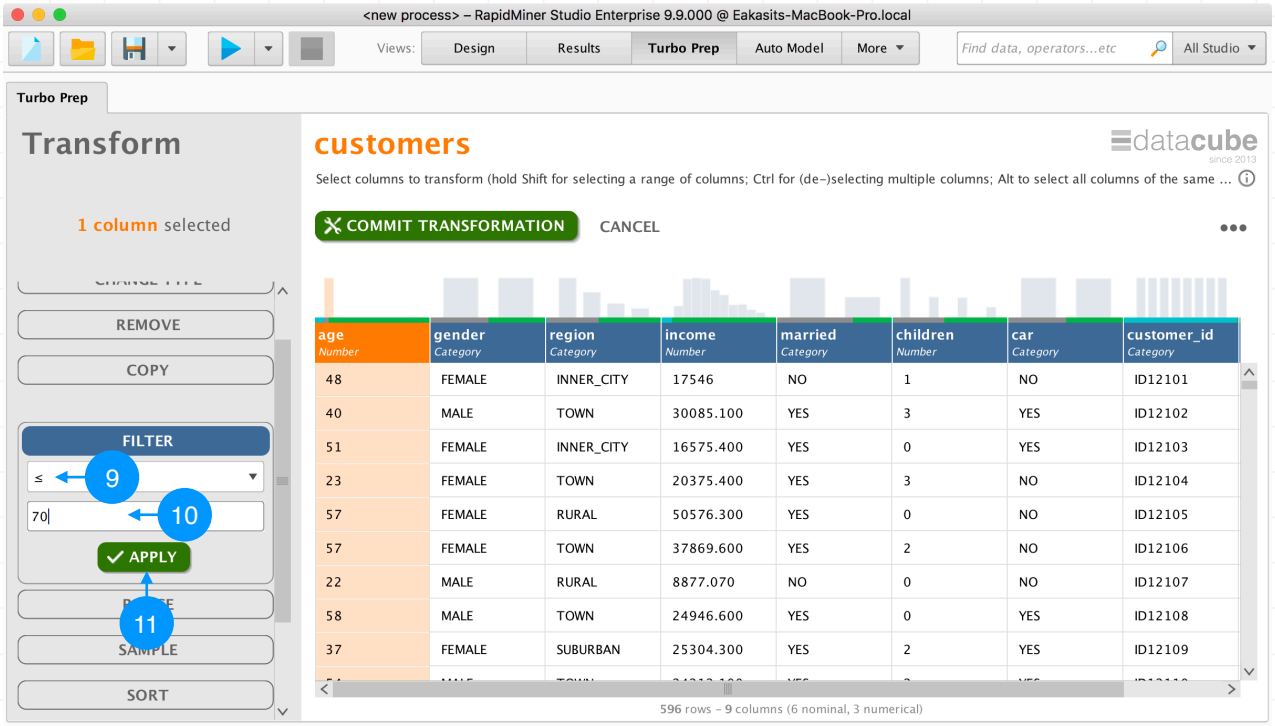


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Filter Examples

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เปลี่ยนค่าในส่วนของ Filter ด้านซ้ายมือ

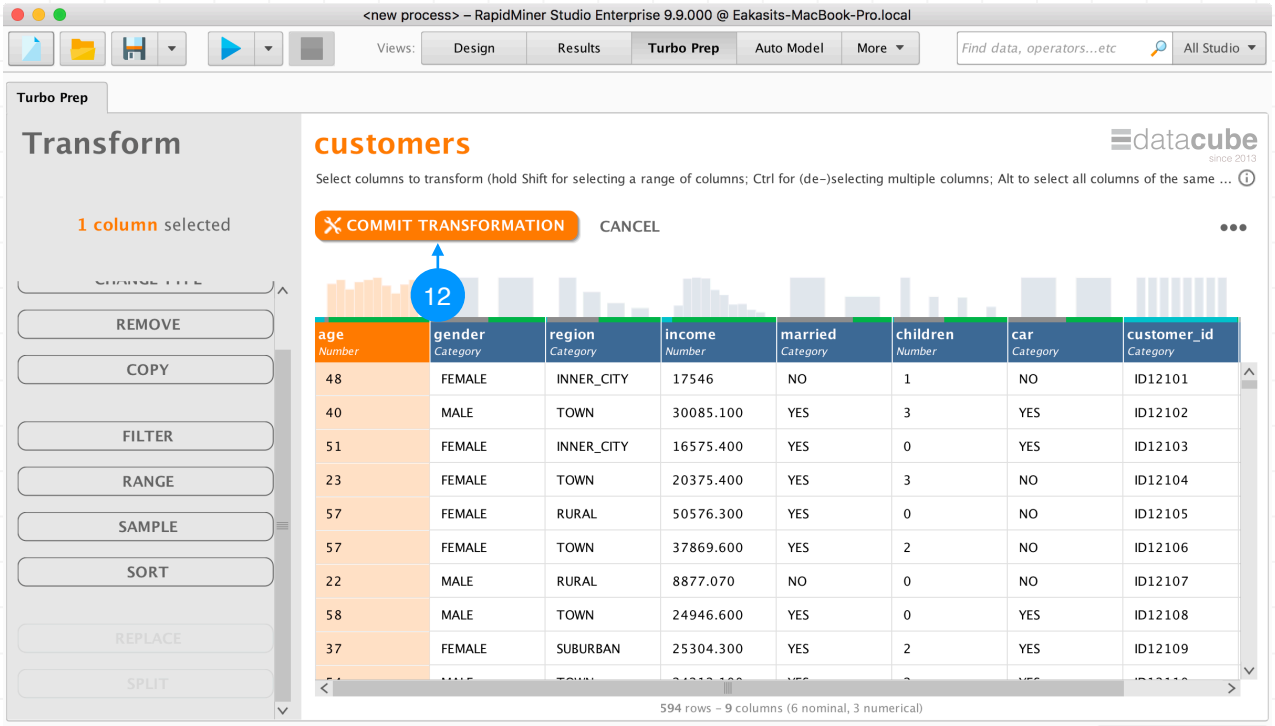


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Filter Examples

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- คลิกปุ่ม **COMMIT TRANSFORMATION**

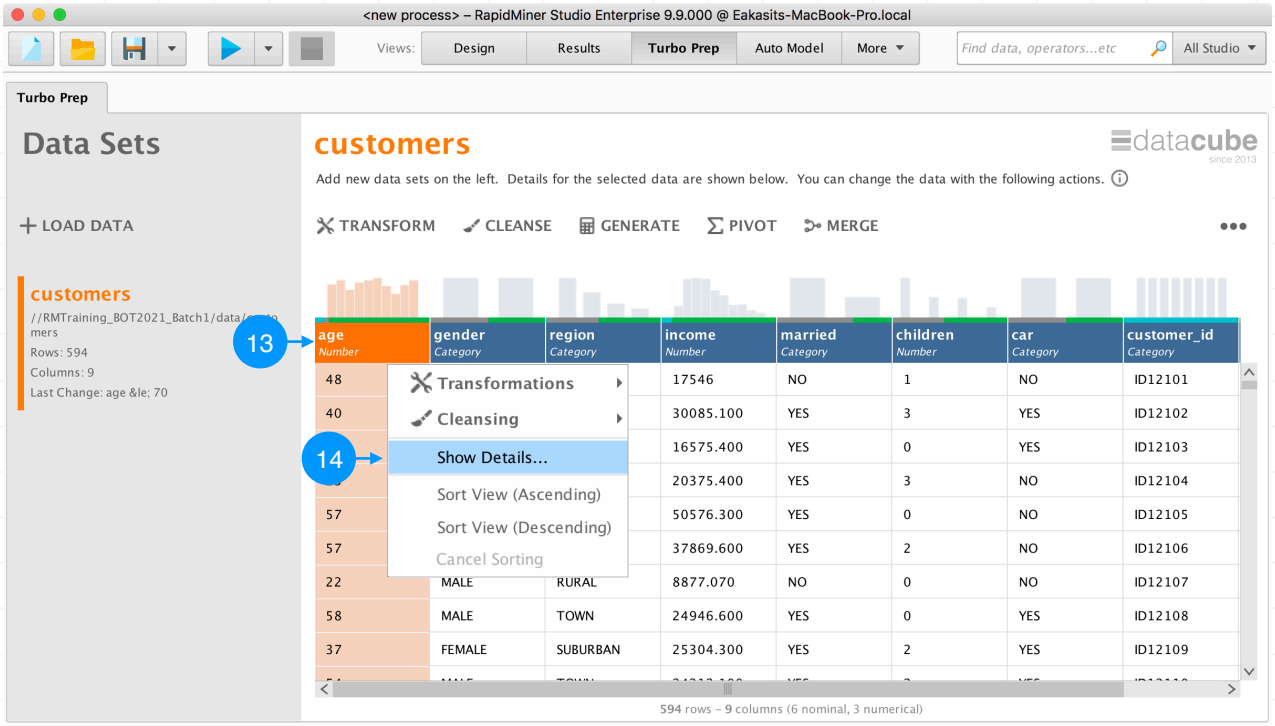


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Filter Examples

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกแอตทริบิวต์ **Age** และคลิกขวา เลือกเมนู **Show Details...**

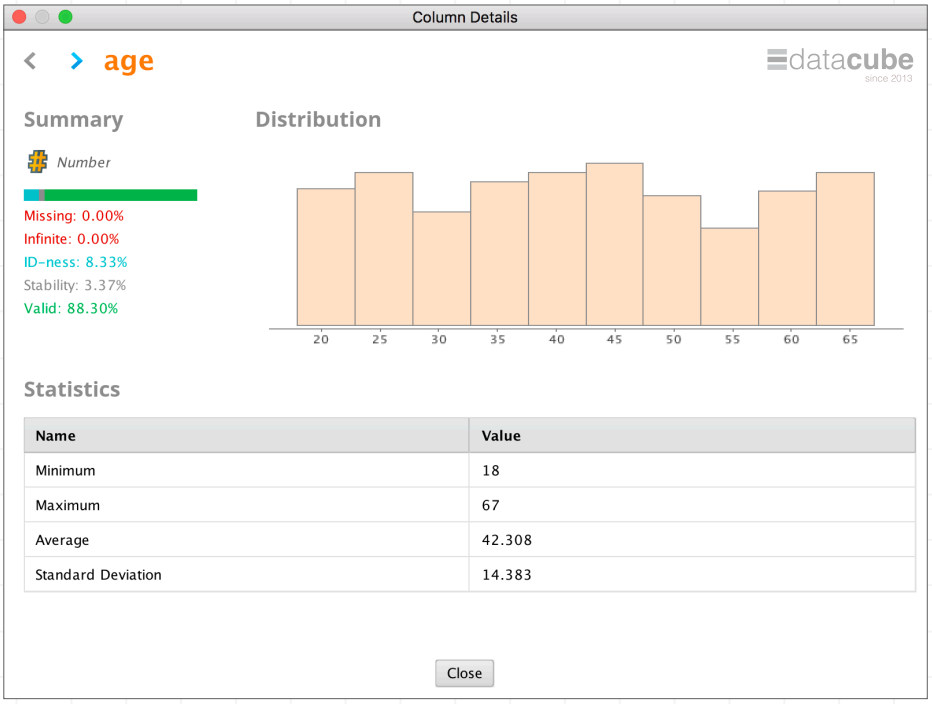


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Filter Examples

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- อายุมีค่าตั้งแต่ 18 - 67 ปี ถือว่าเป็นช่วงปกติ



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data Preparation

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

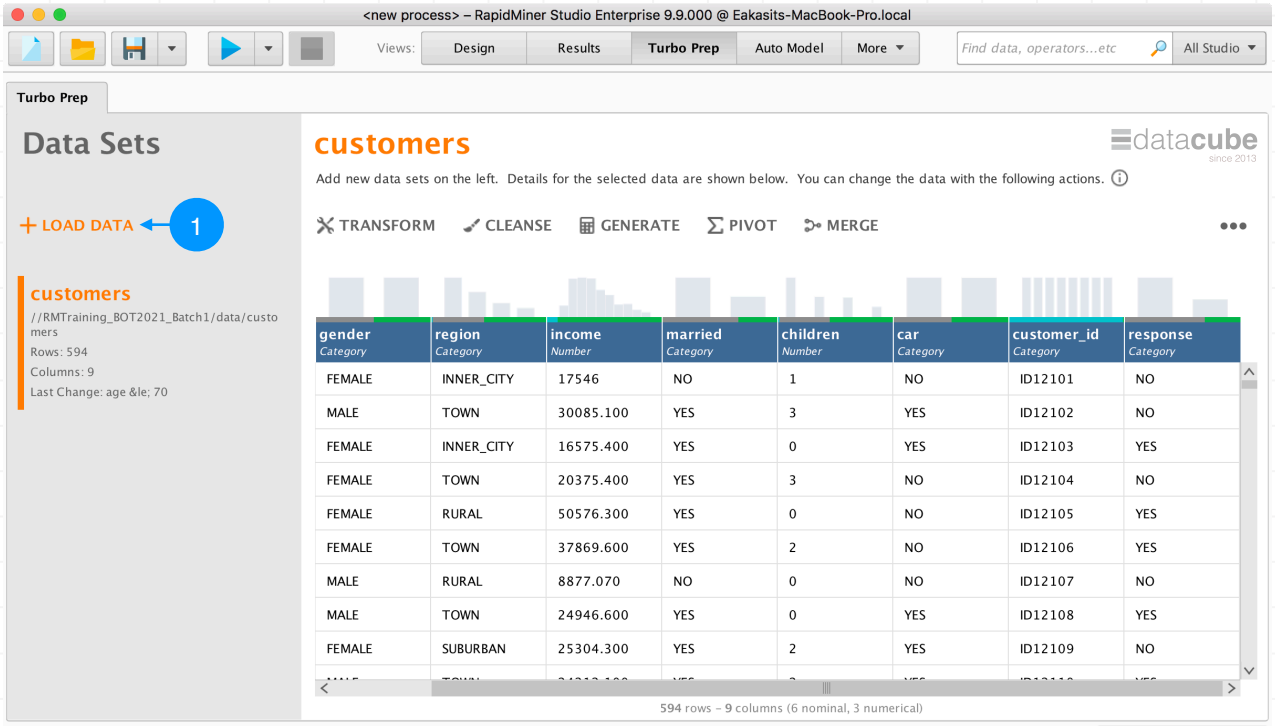
- การเตรียมข้อมูลด้วย Turbo Prep
  - การดูรายละเอียดของข้อมูล (show detail)
  - การเรียงลำดับข้อมูล (sort)
  - การแทนค่าข้อมูล (replace)
  - การแทนค่าว่าง (replace missing value)
  - การเลือกข้อมูลบางตัวอย่าง (filter examples)
  - **การ join ข้อมูล 2 ชุดเข้าด้วยกัน**
  - การ Export Data และ Process

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- คลิกที่เมนู **LOAD DATA**

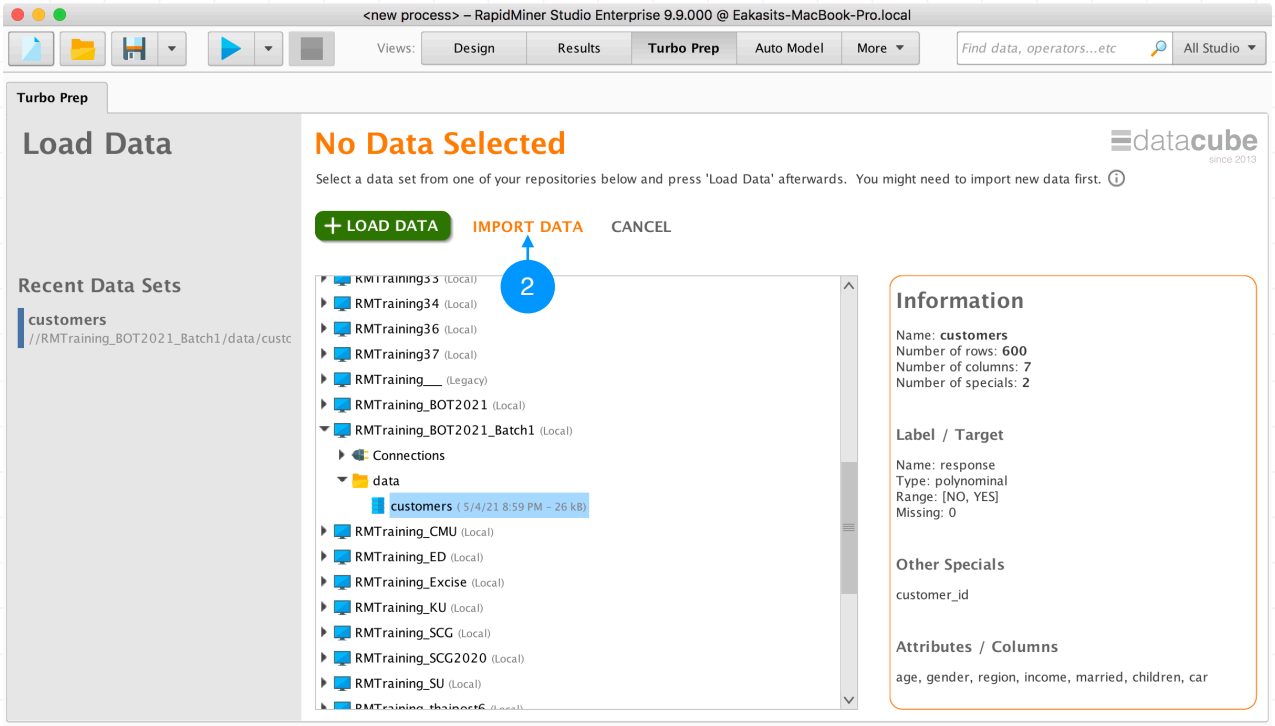


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- คลิกที่เมนู **IMPORT DATA** เพื่อโหลดข้อมูลอีกชุดหนึ่งได้แก่ customer email

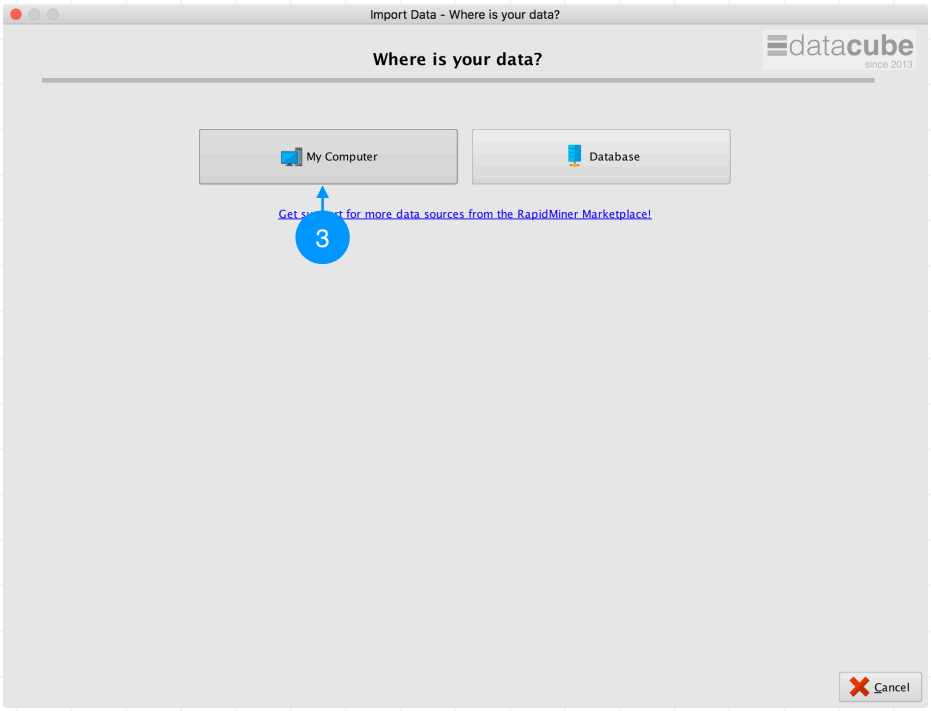


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- คลิกที่ปุ่ม **My Computer** เพื่อเลือกไฟล์ในเครื่องของเราเอง

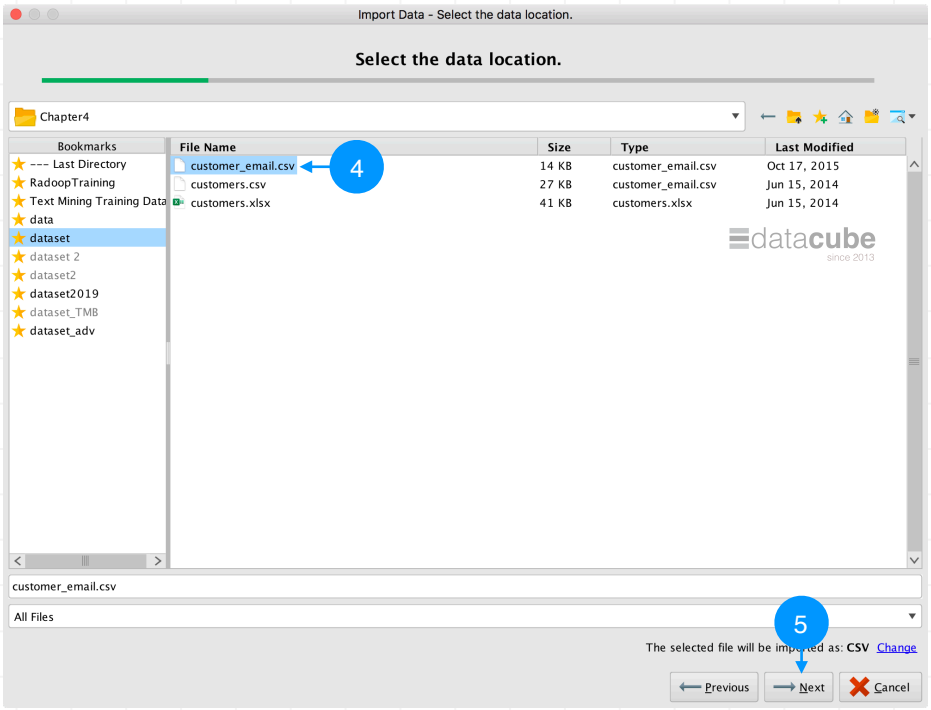


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกไฟล์ customer_email.csv

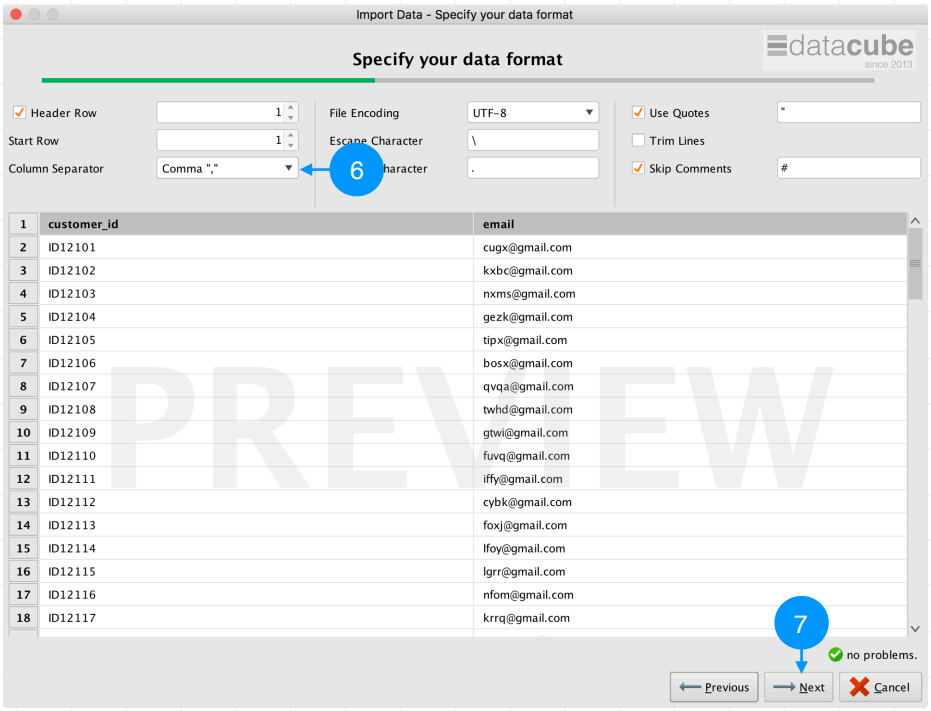


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือก Column Separator เป็น Comma “,”

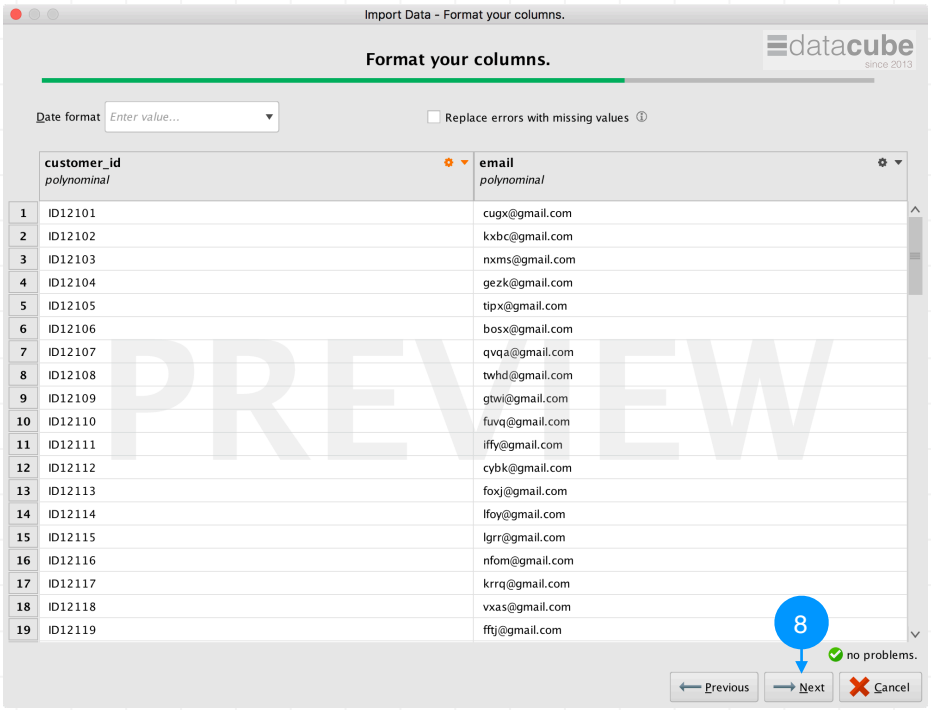


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- คลิกที่ปุ่ม Next

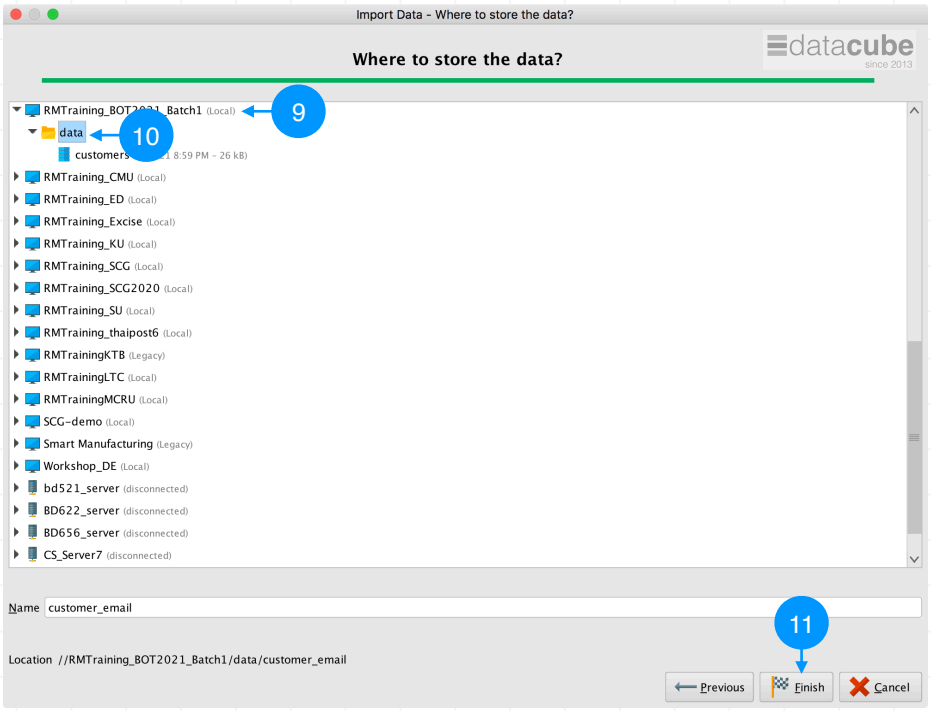


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- บันทึกข้อมูลใน repository ที่กำหนดและกดปุ่ม Finish

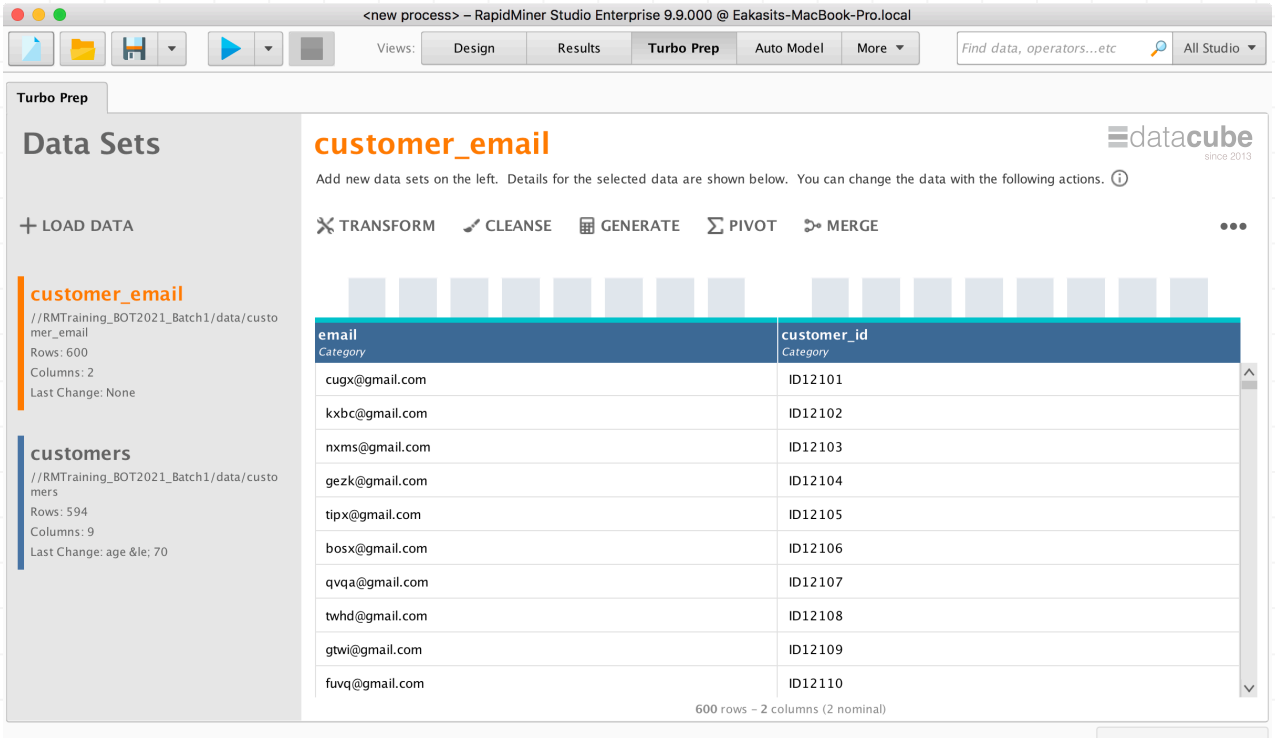


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- แสดงข้อมูล Email ของลูกค้าแต่ละราย

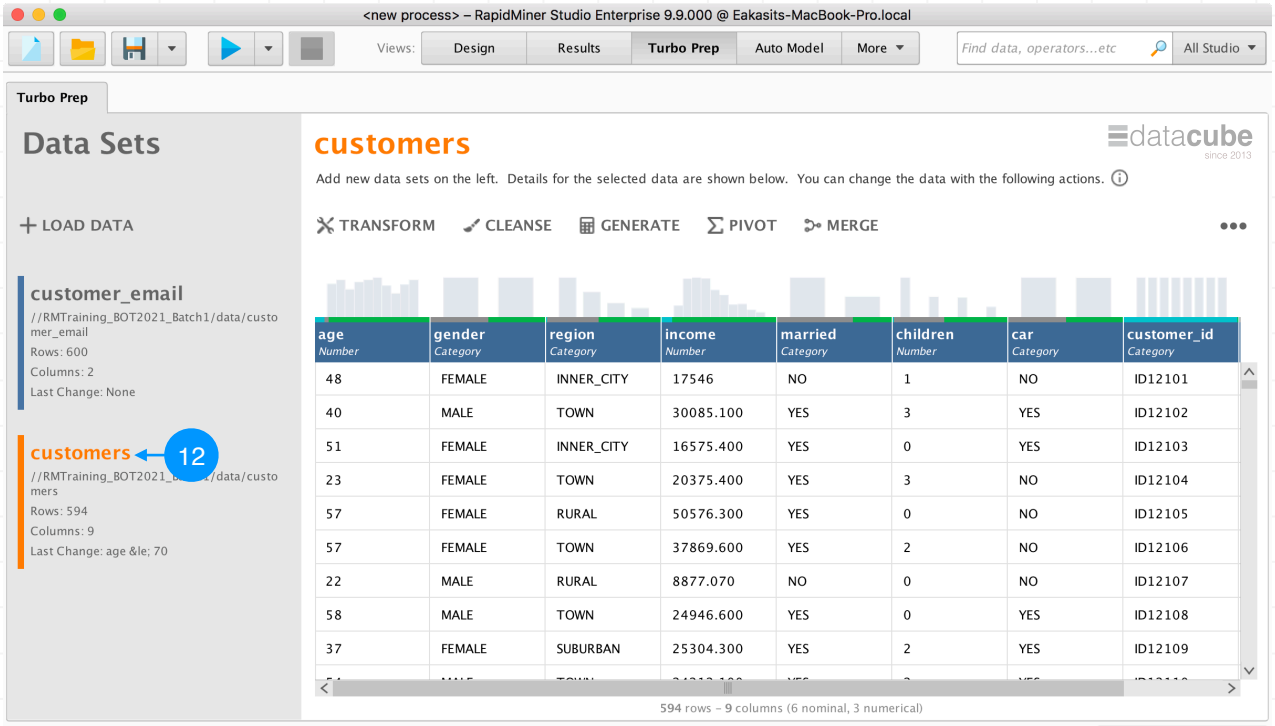


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกข้อมูล customers เพื่อเป็นข้อมูลหลัก

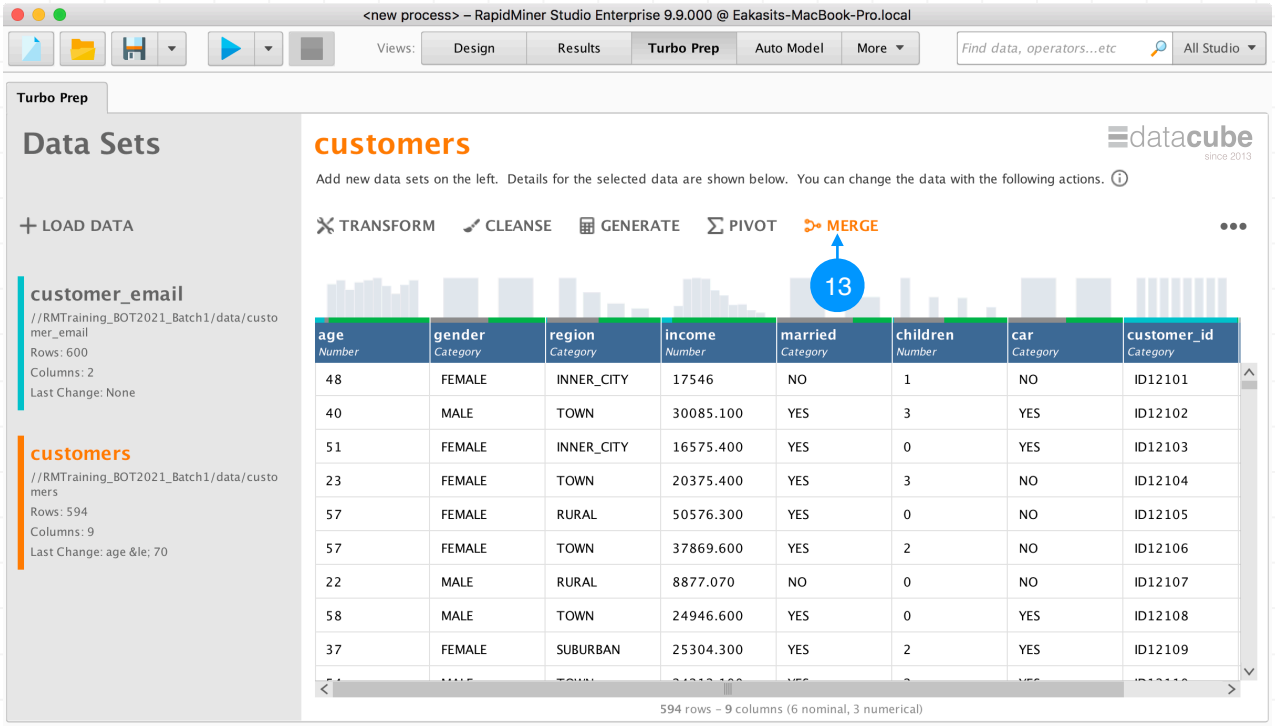


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกเมนู **MERGE** เพื่อทำการ Join ข้อมูล 2 ชุด

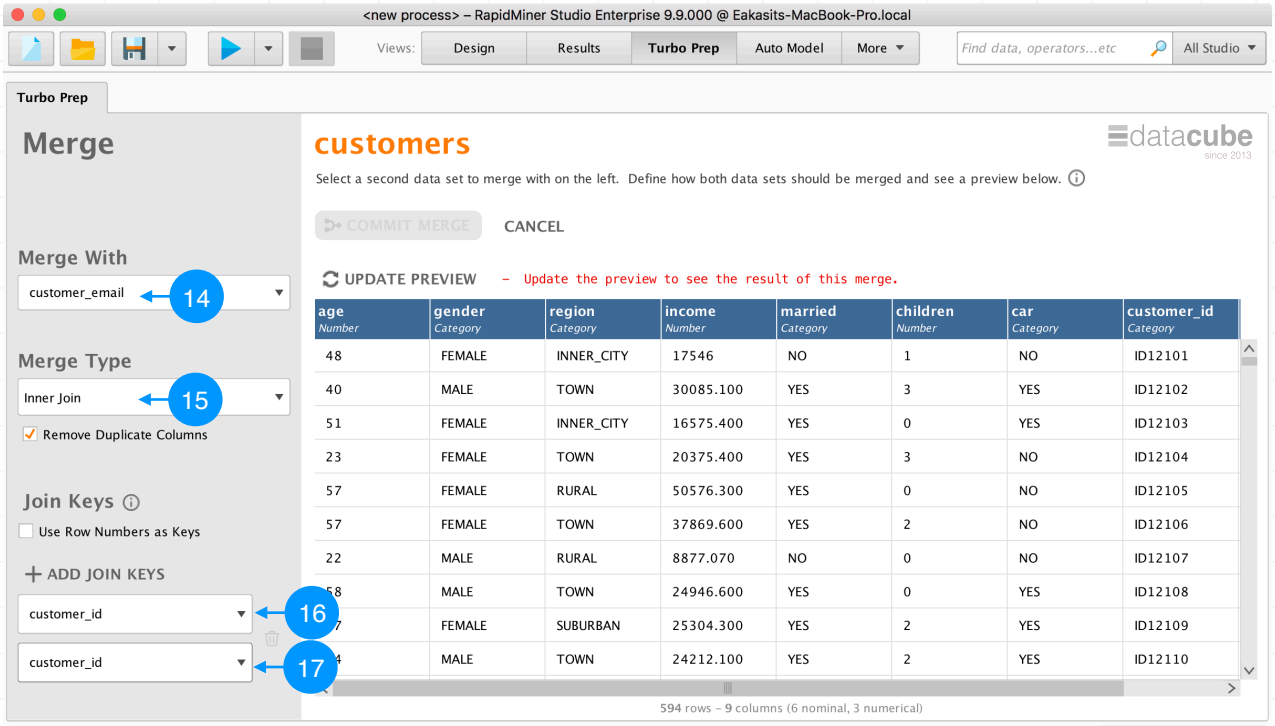


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เปลี่ยนค่าในส่วนของ Merge ด้านซ้ายมือ

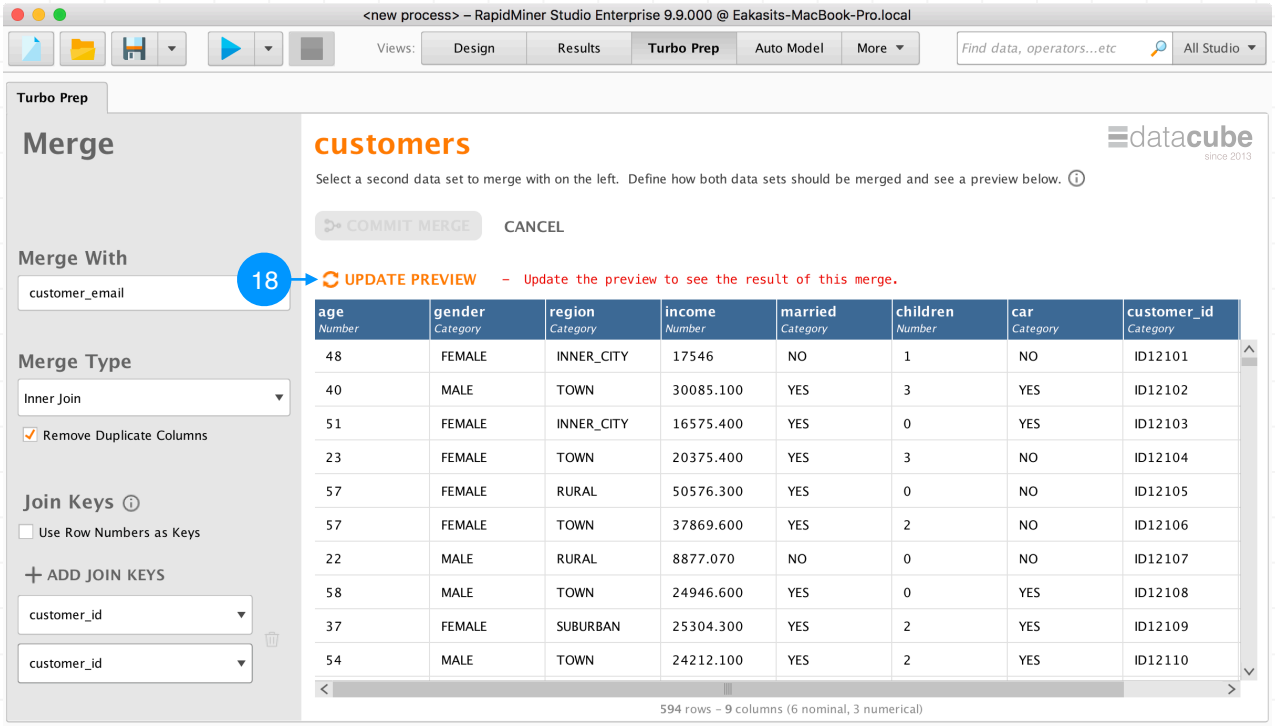


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกเมนู **UPDATE PREVIEW**

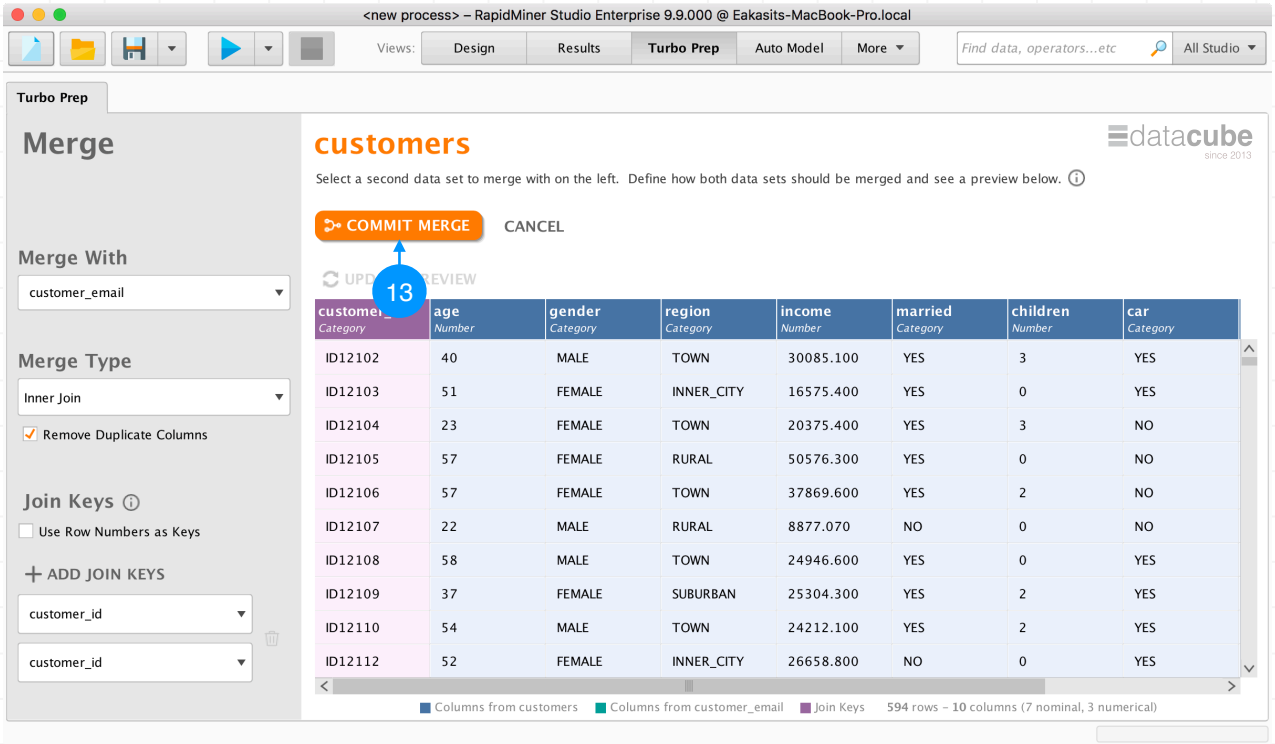


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Join

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- แสดงข้อมูลที่ Join กันได้แล้ว และ เลือกเมนู **COMMIT MERGE**



Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Data Preparation

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

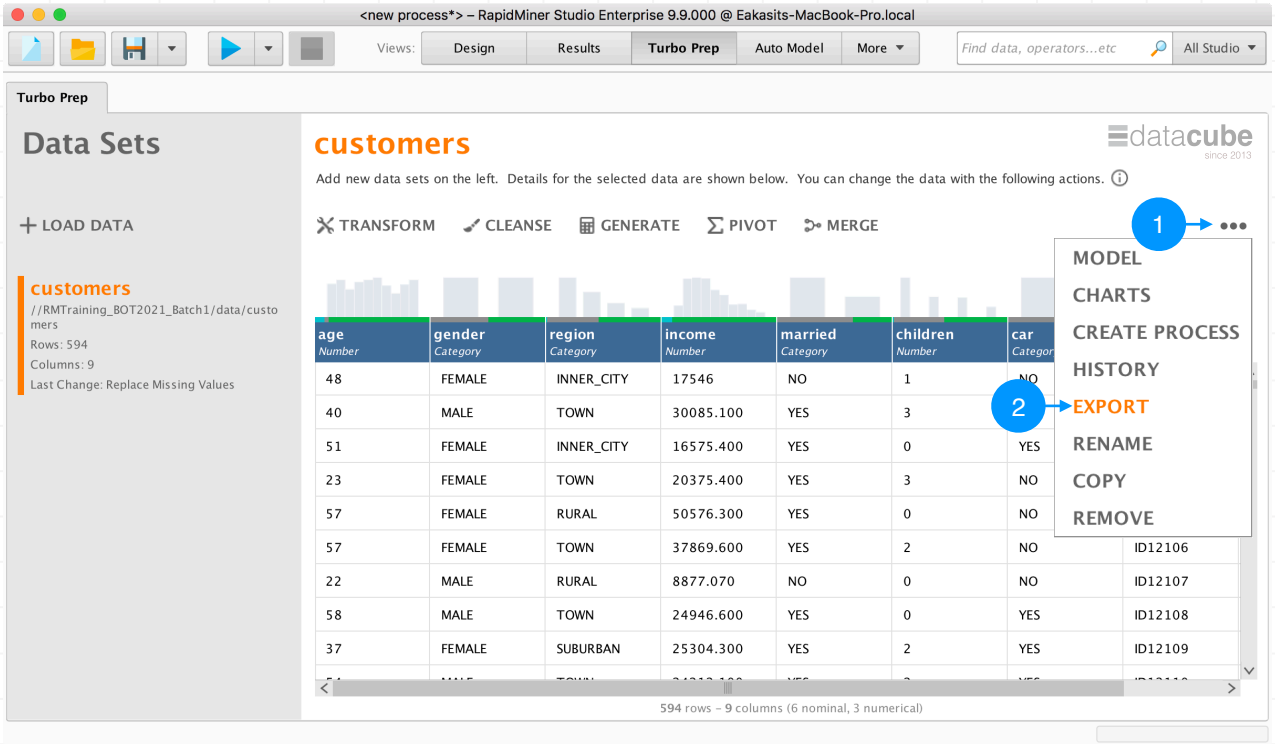
- การเตรียมข้อมูลด้วย Turbo Prep
  - การดูรายละเอียดของข้อมูล (show detail)
  - การเรียงลำดับข้อมูล (sort)
  - การแทนค่าข้อมูล (replace)
  - การแทนค่าว่าง (replace missing value)
  - การเลือกข้อมูลบางตัวอย่าง (filter examples)
  - การ join ข้อมูล 2 ชุดเข้าด้วยกัน
  - **การ Export Data และ Process**

Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Export Data

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกเมนู ... แล้วเลือกเมนู **Export**

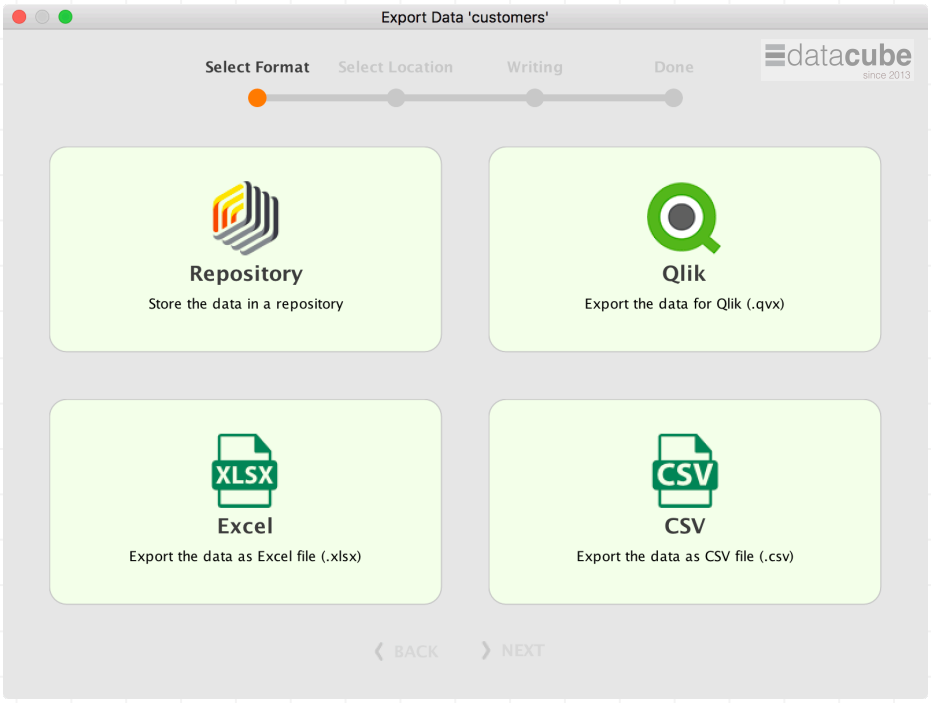


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Export Data

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- หน้าต่างการ Export จะปรากฏมาให้เลือก

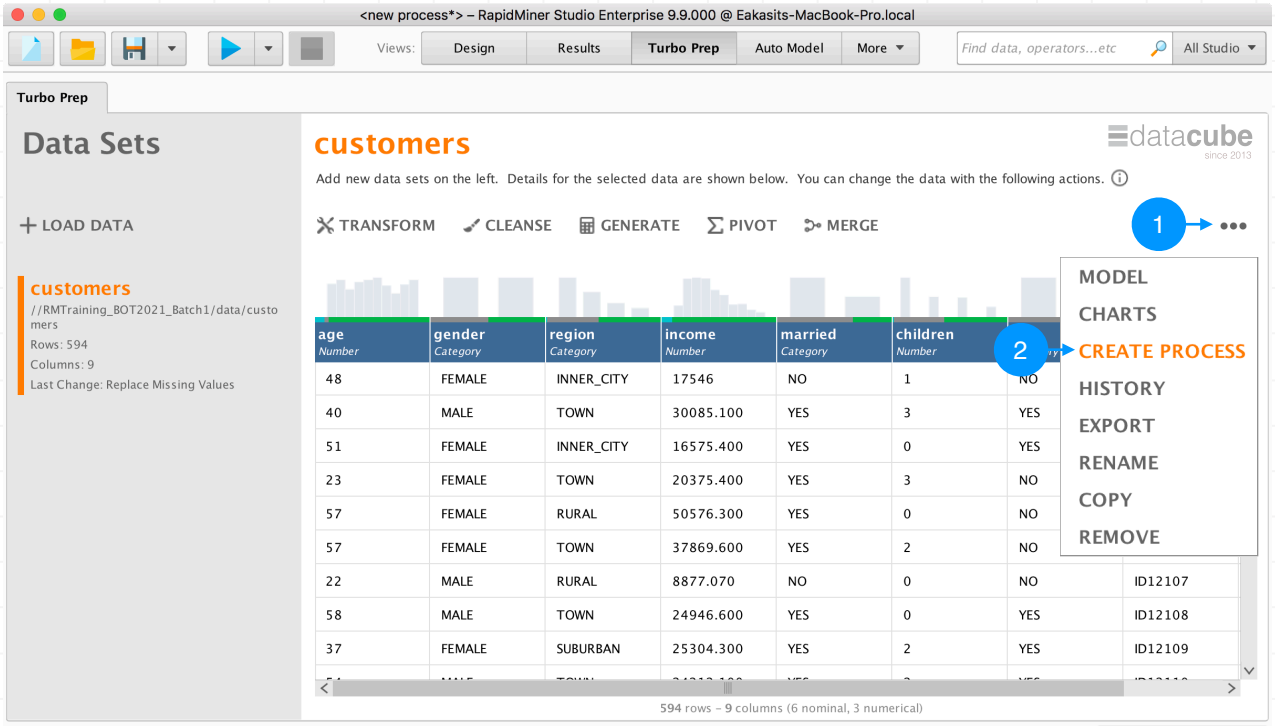


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Create Process

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- เลือกเมนู ... แล้วเลือกเมนู **Export**

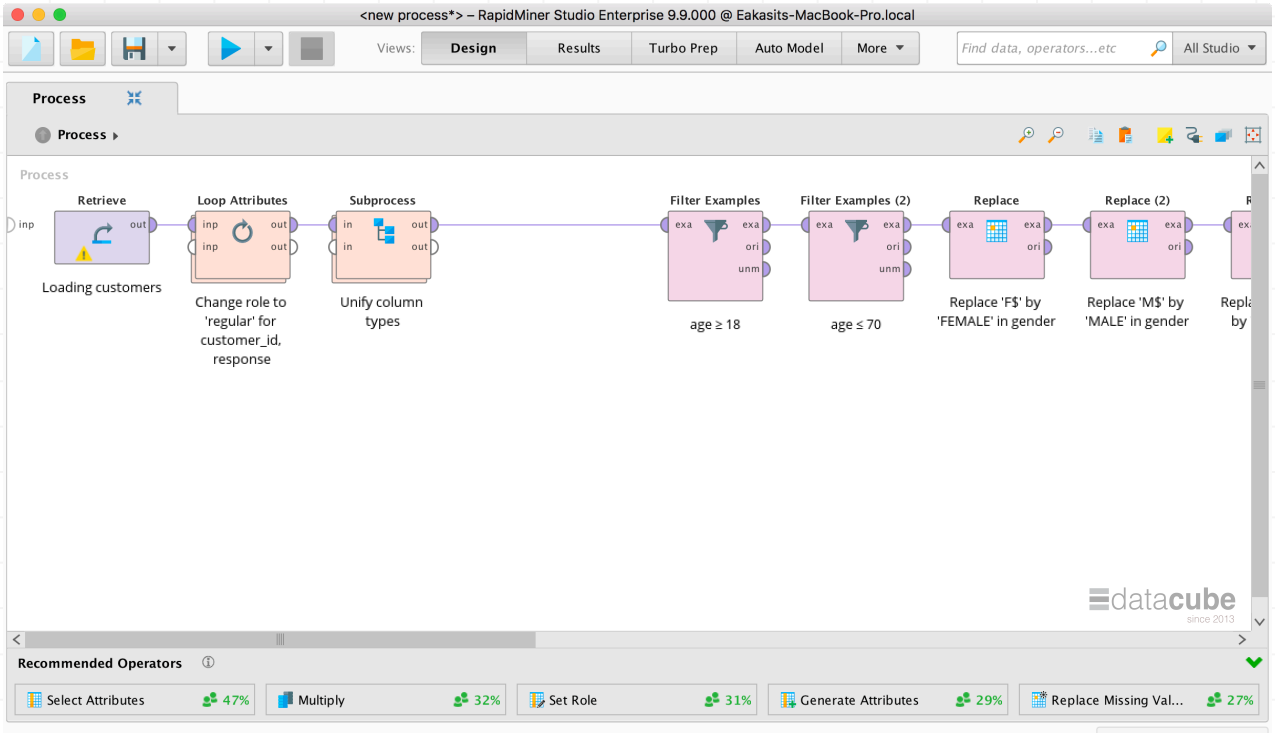


Copyright © 2021 Cube Analytics Consulting Co., Ltd.

Create Process

สำหรับการอบรมหลักสูตร Practical Data Science and Machine Learning

- จะสร้างโพรเซสขึ้นมาในหน้าแท็บ Design



Copyright © 2021 Cube Analytics Consulting Co., Ltd.