

# 전복 순살 무게 예측

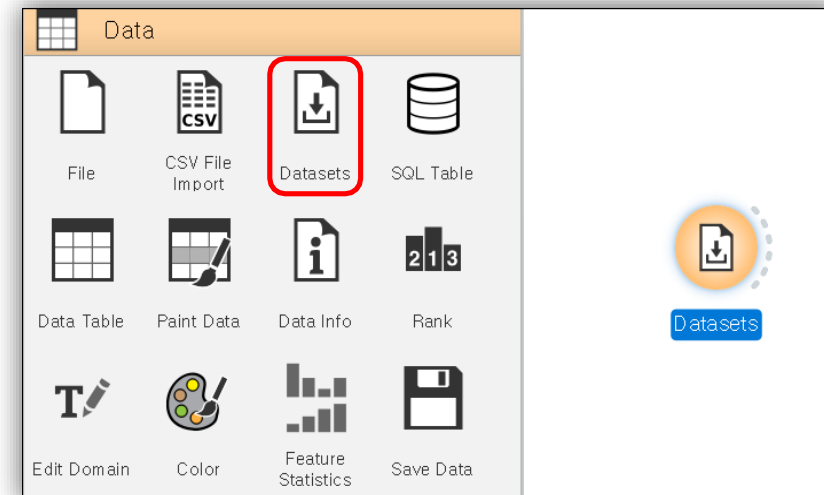


소프트웨어융합대학원  
진혜진

- 전복 데이터를 분석하고 전복의 순살 무게를 예측할 수 있는 인공지능 모델을 만들어보자.

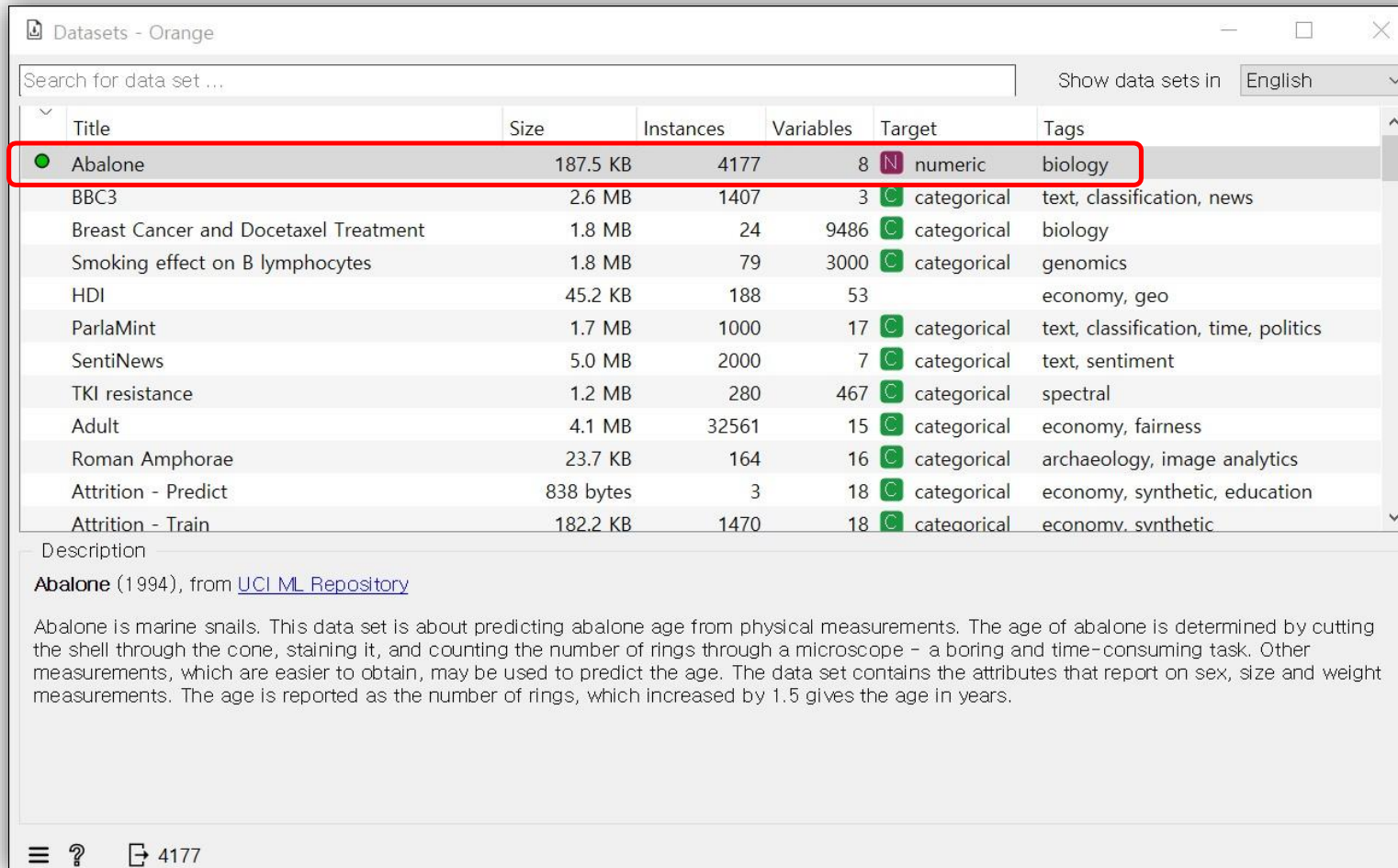
- 데이터 준비하기

- Orange3 데이터 다운로드
  - Data 카테고리 [Datasets] 위젯



## ■ Abalone(전복) 데이터

### ■ UCI Machine Learning Repository에서 제공하는 데이터



Datasets - Orange

Search for data set ... Show data sets in English

Title	Size	Instances	Variables	Target	Tags
Abalone	187.5 KB	4177	8	N numeric	biology
BBC3	2.6 MB	1407	3	C categorical	text, classification, news
Breast Cancer and Docetaxel Treatment	1.8 MB	24	9486	C categorical	biology
Smoking effect on B lymphocytes	1.8 MB	79	3000	C categorical	genomics
HDI	45.2 KB	188	53		economy, geo
ParlaMint	1.7 MB	1000	17	C categorical	text, classification, time, politics
SentiNews	5.0 MB	2000	7	C categorical	text, sentiment
TKI resistance	1.2 MB	280	467	C categorical	spectral
Adult	4.1 MB	32561	15	C categorical	economy, fairness
Roman Amphorae	23.7 KB	164	16	C categorical	archaeology, image analytics
Attrition - Predict	838 bytes	3	18	C categorical	economy, synthetic, education
Attrition - Train	182.2 KB	1470	18	C categorical	economy, synthetic

Description

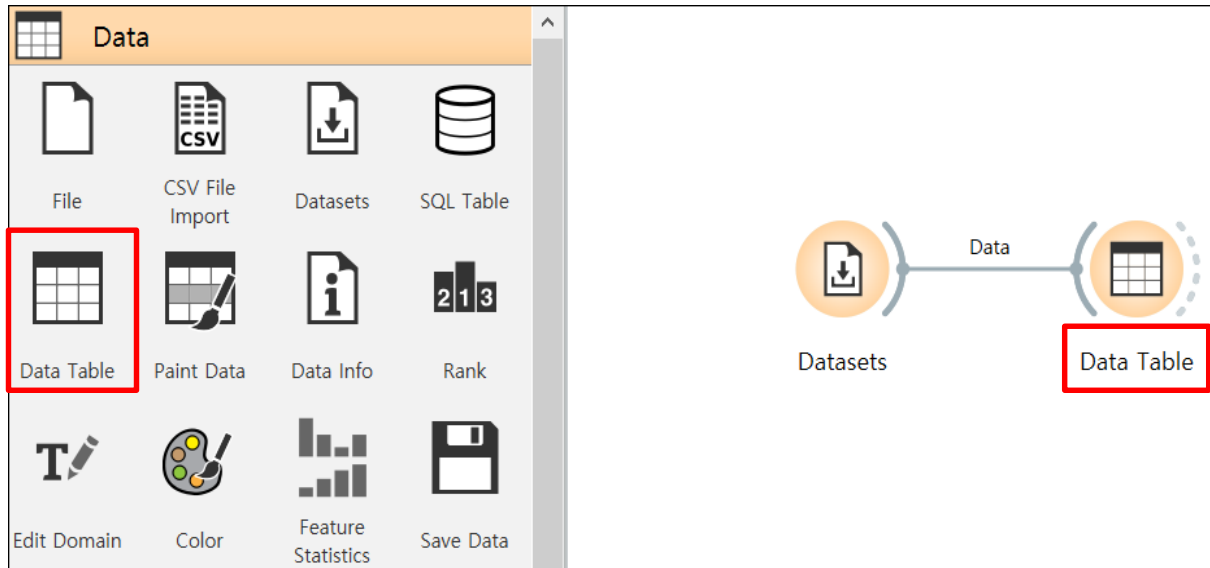
**Abalone** (1994), from [UCI ML Repository](https://archive.ics.uci.edu/ml/datasets/Abalone)

Abalone is marine snails. This data set is about predicting abalone age from physical measurements. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope - a boring and time-consuming task. Other measurements, which are easier to obtain, may be used to predict the age. The data set contains the attributes that report on sex, size and weight measurements. The age is reported as the number of rings, which increased by 1.5 gives the age in years.

4177

## ■ 데이터 불러오기

- Data 카테고리 [Data Table] 위젯



Data Table - Orange

**Info**

4177 instances (no missing data)  
8 features  
Numeric outcome  
No meta attributes.

**Variables**

☒ Show variable labels (if present)

☐ Visualize numeric values

☒ Color by instance classes

**Selection**

☒ Select full rows

Restore Original Order

☒ Send Automatically

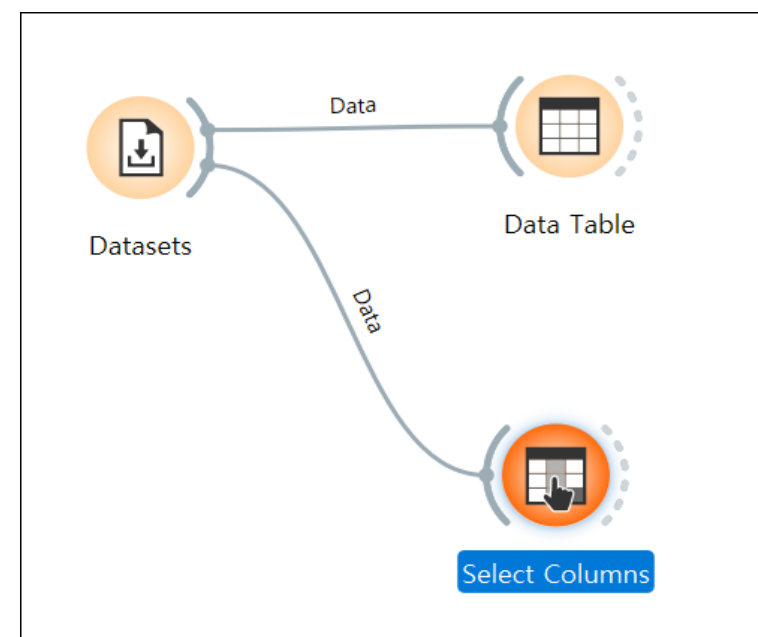
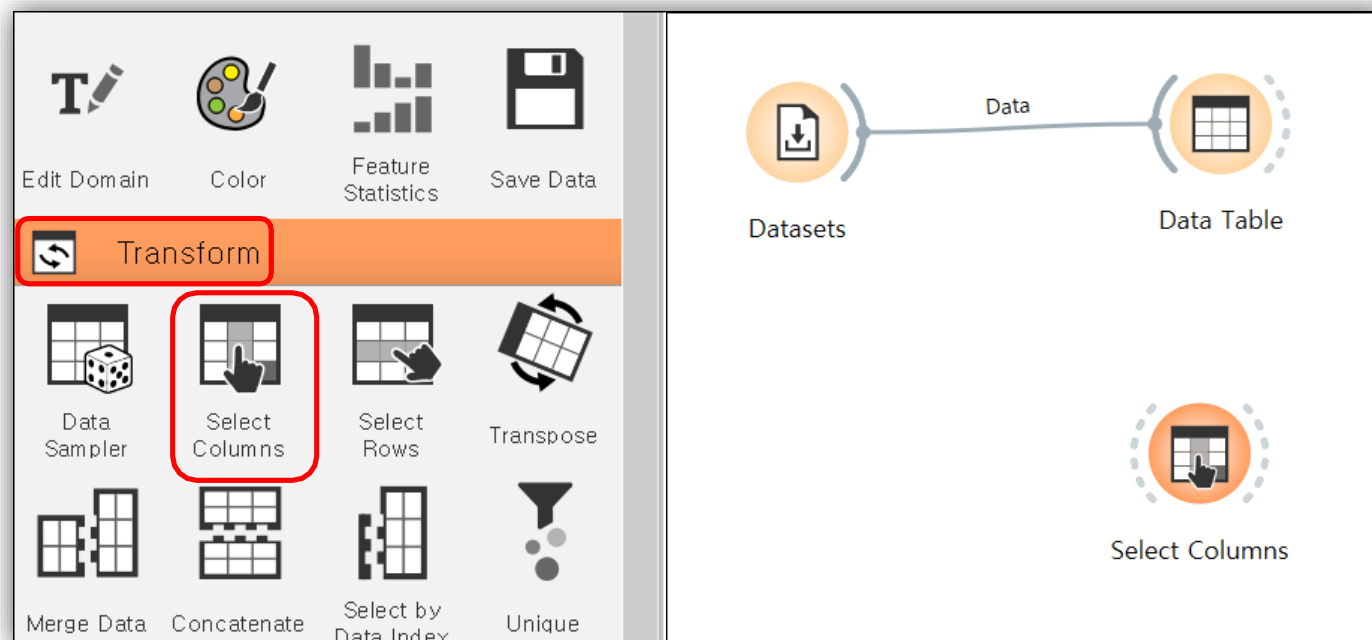
	Rings	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight
1	15	M	0.455	0.365	0.095	0.5140	0.2245	0.1
2	7	M	0.350	0.265	0.090	0.2255	0.0995	0.0
3	9	F	0.530	0.420	0.135	0.6770	0.2565	0.1
4	10	M	0.440	0.365	0.125	0.5160	0.2155	0.1
5	7	I	0.330	0.255	0.080	0.2050	0.0895	0.0
6	8	I	0.425	0.300	0.095	0.3515	0.1410	0.0
7	20	F	0.530	0.415	0.150	0.7775	0.2370	0.1
8	16	F	0.545	0.425	0.125	0.7680	0.2940	0.1
9	9	M	0.475	0.370	0.125	0.5095	0.2165	0.1
10	19	F	0.550	0.440	0.150	0.8945	0.3145	0.1
11	14	F	0.525	0.380	0.140	0.6065	0.1940	0.1
12	10	M	0.430	0.350	0.110	0.4060	0.1675	0.0
13	11	M	0.490	0.380	0.135	0.5415	0.2175	0.0
14	10	F	0.535	0.405	0.145	0.6845	0.2725	0.1
15	10	F	0.470	0.355	0.100	0.4755	0.1675	0.0
16	12	M	0.500	0.400	0.130	0.6645	0.2580	0.1
17	7	I	0.355	0.280	0.085	0.2905	0.0950	0.0
18	10	F	0.440	0.340	0.100	0.4510	0.1880	0.0

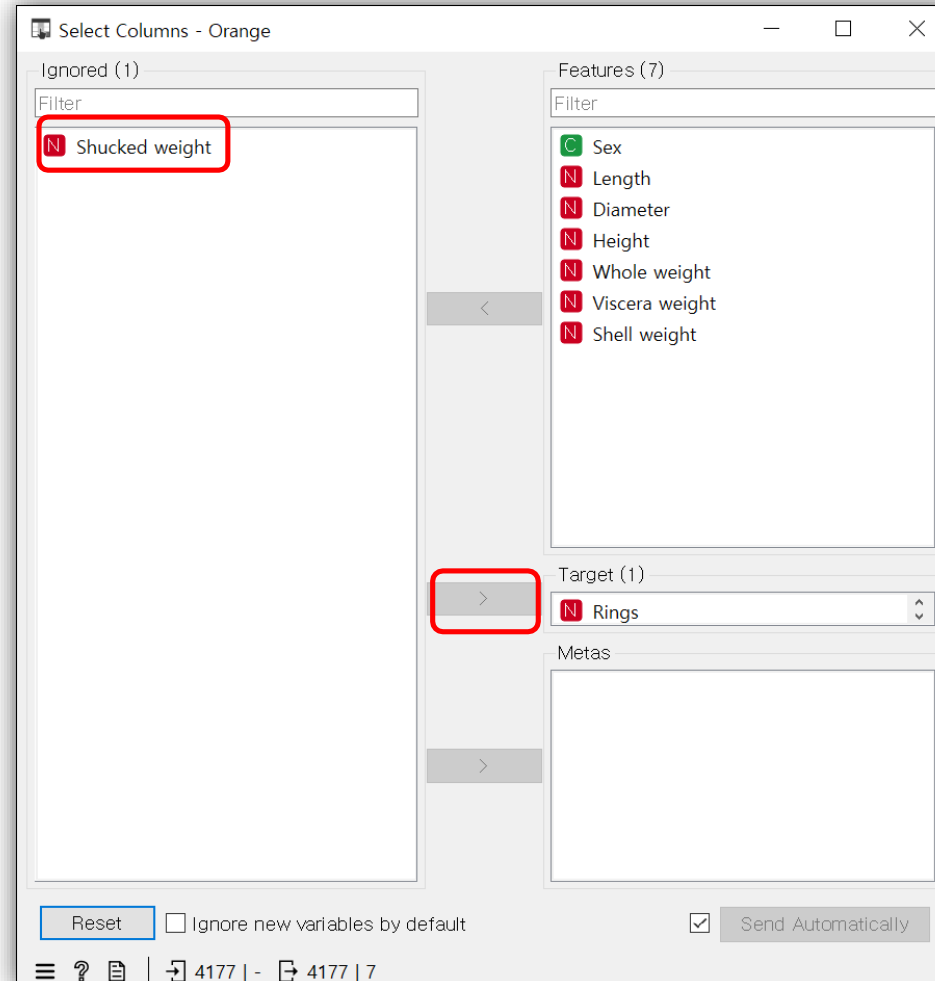
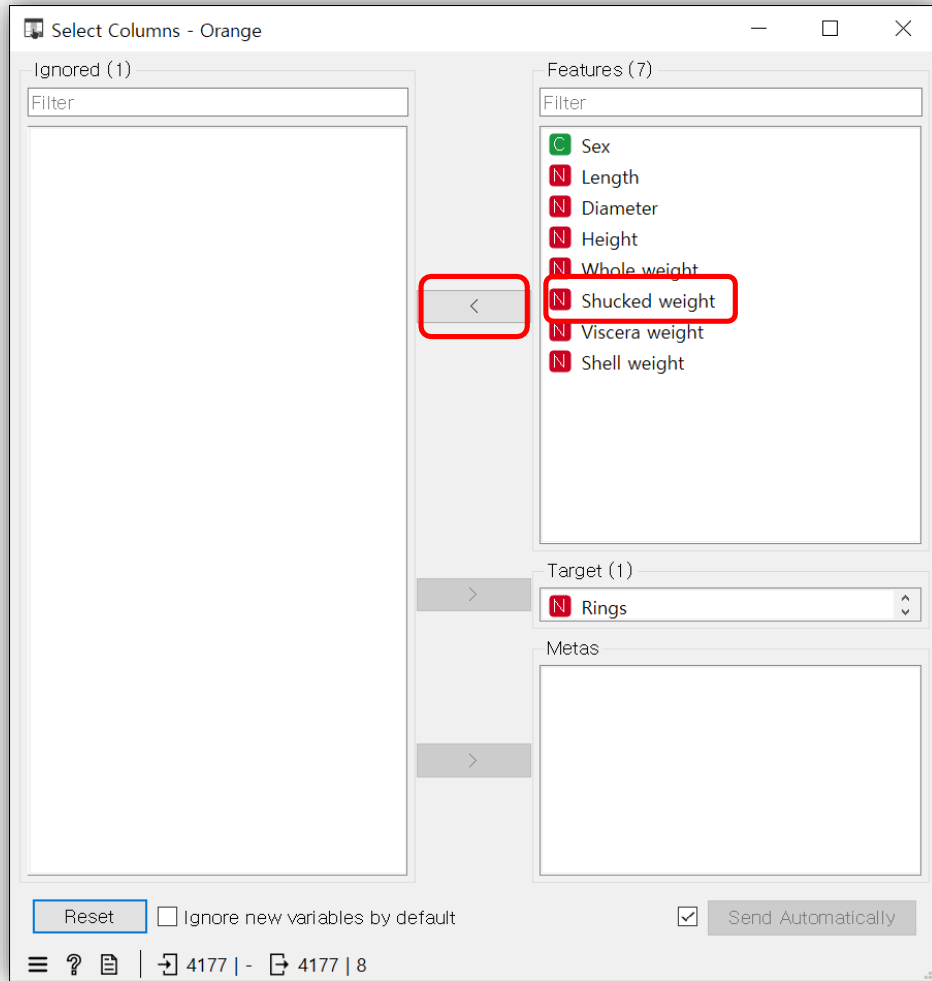
4177 | 4177 | 4177

- 데이터 속성 정보 확인하기
  - Rings : 나이테 개수
  - Sex : 성별
  - Length : 길이
  - Diameter : 직경
  - Height : 두께
  - Whole weight : 전체 무게
  - Shucked weight : 순살 무게
  - Viscera weight : 내장 무게
  - Shell weight : 껍질 무게

## ■ 데이터 전처리하기

### ■ Transform 카테고리 [Select Columns] 위젯







Select Columns - Orange

Ignored

Filter

Features (7)

Filter

- Sex
- Length
- Diameter
- Height
- Whole weight
- Viscera weight
- Shell weight

Target (2)

- Rings
- Shucked weight

Metas

Reset ☐ Ignore new variables by default ☒ Send Automatically

4177 | - 4177 | 7

Most widgets do not support multiple targets

Select Columns - Orange

Ignored

Filter

Features (7)

Filter

- Sex
- Length
- Diameter
- Height
- Whole weight
- Viscera weight
- Shell weight

Target (2)

- Rings
- Shucked weight

Metas

Reset ☐ Ignore new variables by default ☒ Send Automatically

4177 | - 4177 | 7

Most widgets do not support multiple targets

Select Columns - Orange

Ignored (1)

Filter

**N Rings**

Features (7)

Filter

- C Sex**
- N Length**
- N Diameter**
- N Height**
- N Whole weight**
- N Viscera weight**
- N Shell weight**

Target (1)

**N Shucked weight**

Metas

Reset ☐ Ignore new variables by default ☒ Send Automatically

4177 | - 4177 | 7

Select Columns - Orange

Ignored

Filter

Features (8)

Filter

- C Sex**
- N Length**
- N Diameter**
- N Height**
- N Whole weight**
- N Viscera weight**
- N Shell weight**
- N Rings**

Target (1)

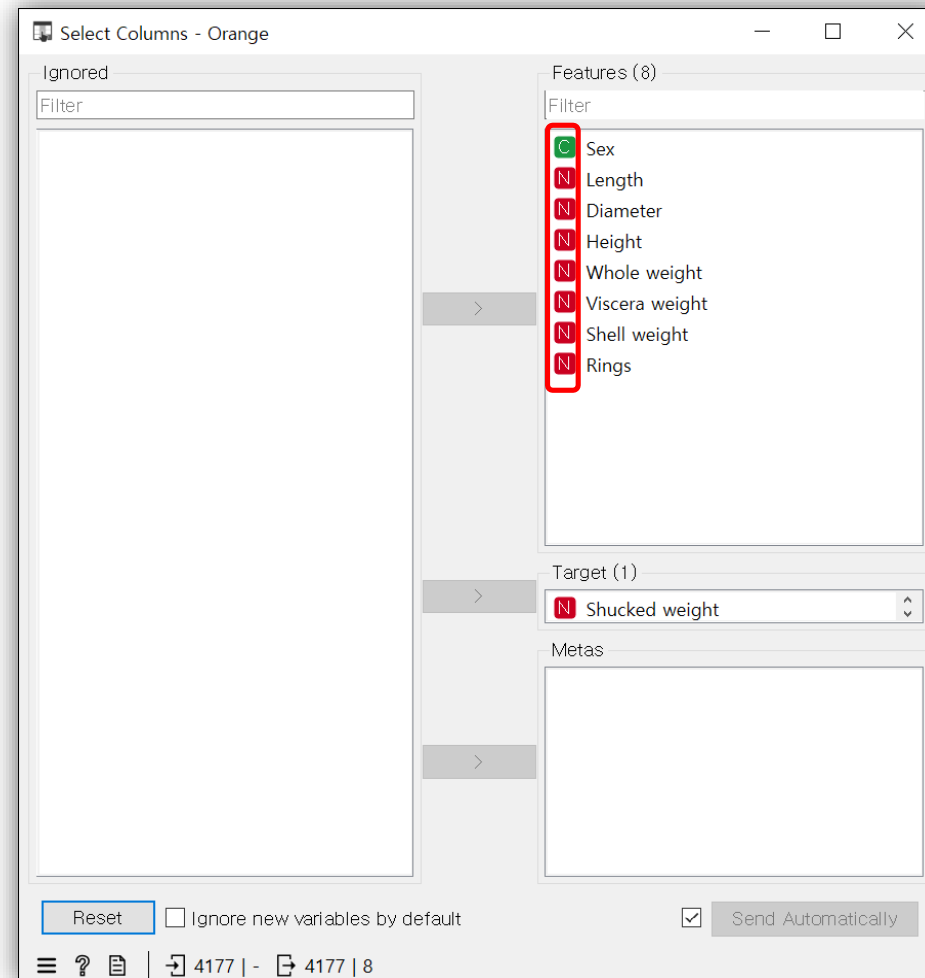
**N Shucked weight**

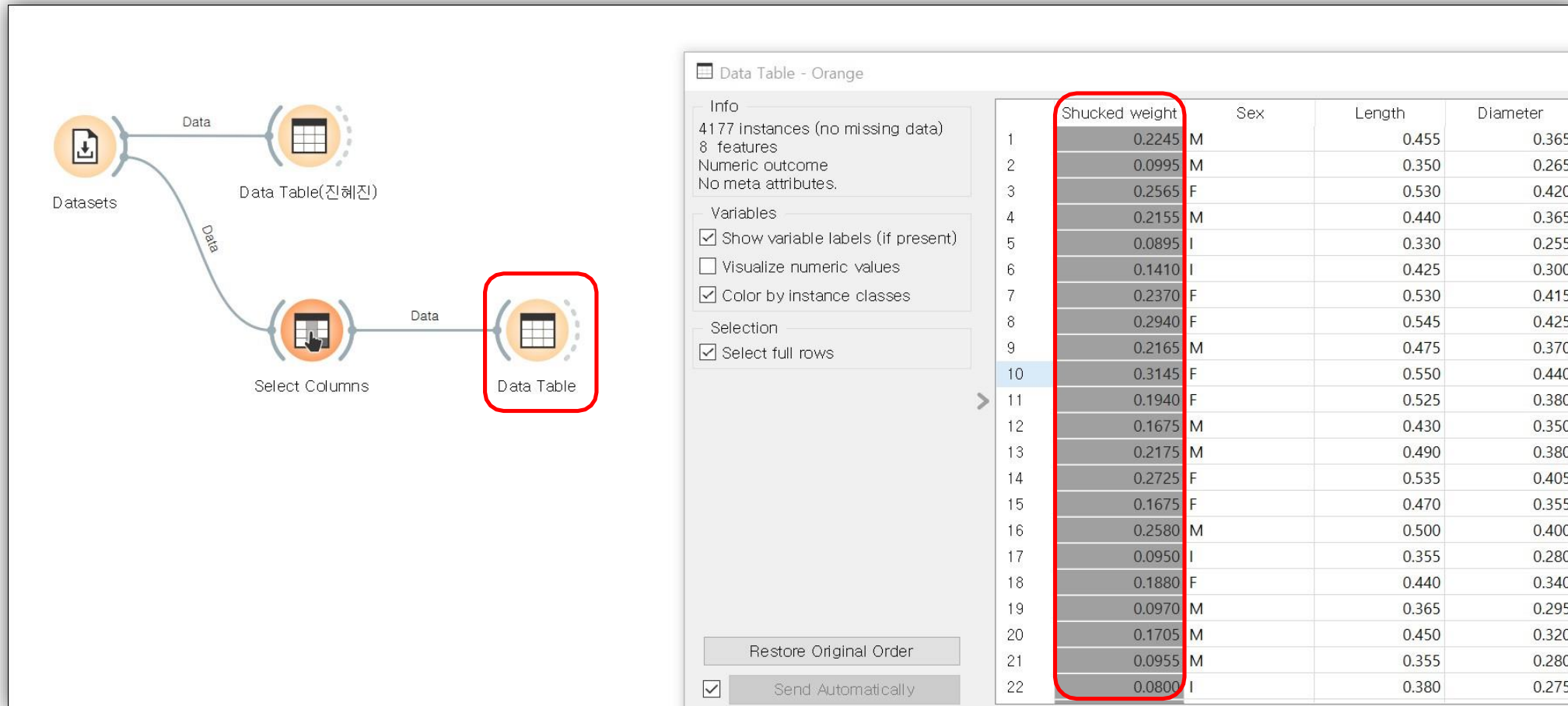
Metas

Reset ☐ Ignore new variables by default ☒ Send Automatically

4177 | - 4177 | 8

## ■ 데이터형식 알아보기





The image shows an Orange Data Mining workflow on the left and the 'Data Table - Orange' widget interface on the right.

**Workflow:**

- Datasets** (Data source icon) connects to **Data Table(진헤진)** (Data Table icon) via a **Data** link.
- Datasets** connects to **Select Columns** (Select Columns icon) via a **Data** link.
- Select Columns** connects to **Data Table** (Data Table icon) via a **Data** link.
- The **Data Table** widget is highlighted with a red box.

**Data Table - Orange Widget Interface:**

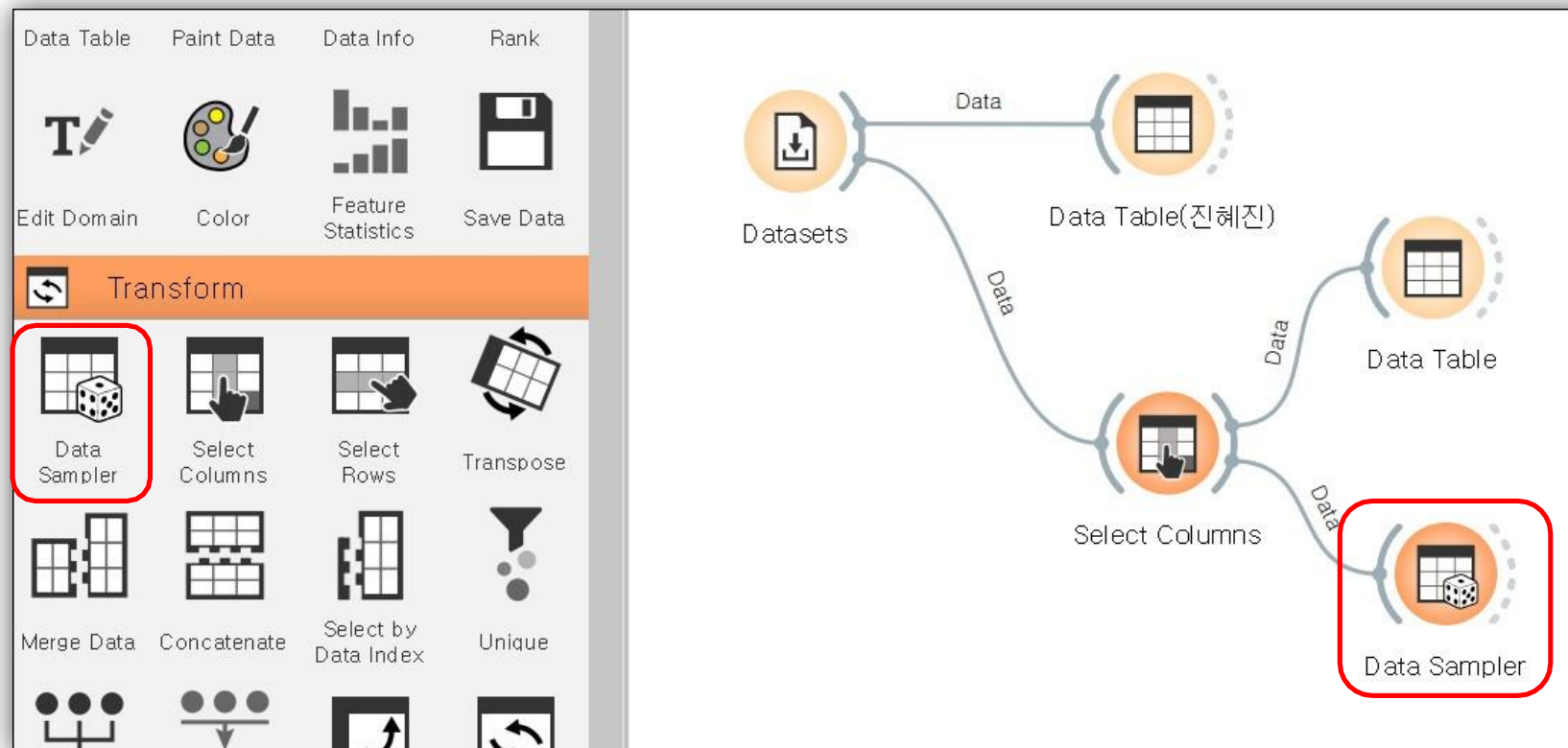
- Info:** 4177 instances (no missing data), 8 features, Numeric outcome, No meta attributes.
- Variables:**
  - ☒ Show variable labels (if present)
  - ☐ Visualize numeric values
  - ☒ Color by instance classes
- Selection:**
  - ☒ Select full rows
- Buttons:** Restore Original Order, ☒ Send Automatically

**Data Table View:**

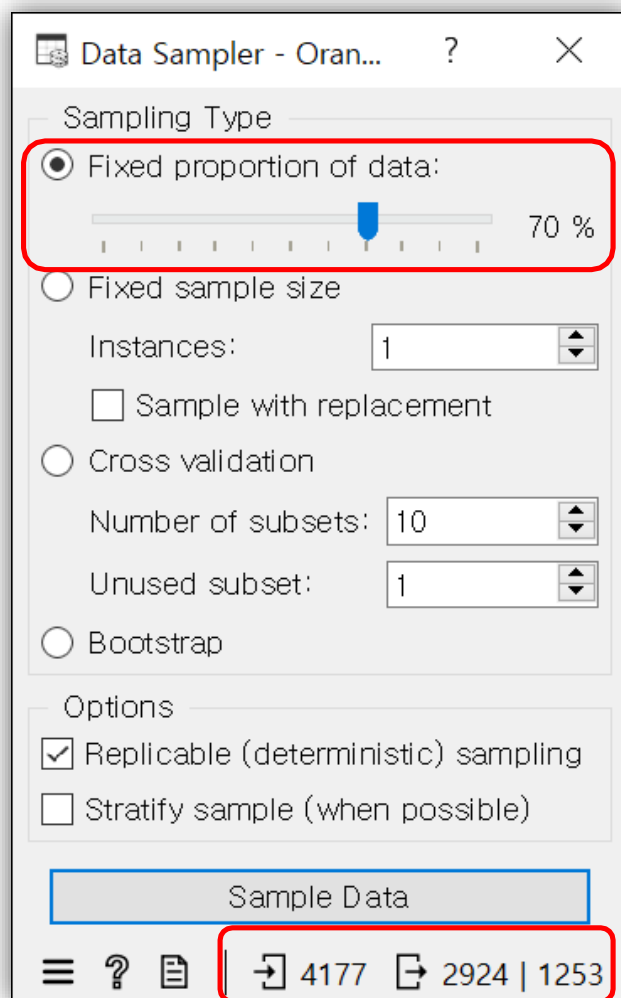
	Shucked weight	Sex	Length	Diameter
1	0.2245	M	0.455	0.365
2	0.0995	M	0.350	0.265
3	0.2565	F	0.530	0.420
4	0.2155	M	0.440	0.365
5	0.0895	I	0.330	0.255
6	0.1410	I	0.425	0.300
7	0.2370	F	0.530	0.415
8	0.2940	F	0.545	0.425
9	0.2165	M	0.475	0.370
10	0.3145	F	0.550	0.440
11	0.1940	F	0.525	0.380
12	0.1675	M	0.430	0.350
13	0.2175	M	0.490	0.380
14	0.2725	F	0.535	0.405
15	0.1675	F	0.470	0.355
16	0.2580	M	0.500	0.400
17	0.0950	I	0.355	0.280
18	0.1880	F	0.440	0.340
19	0.0970	M	0.365	0.295
20	0.1705	M	0.450	0.320
21	0.0955	M	0.355	0.280
22	0.0800	I	0.380	0.275

## ■ 모델 학습시키기

- 훈련데이터와 테스트데이터 나누기
- Transform 카테고리 [Data Sampler] 위젯



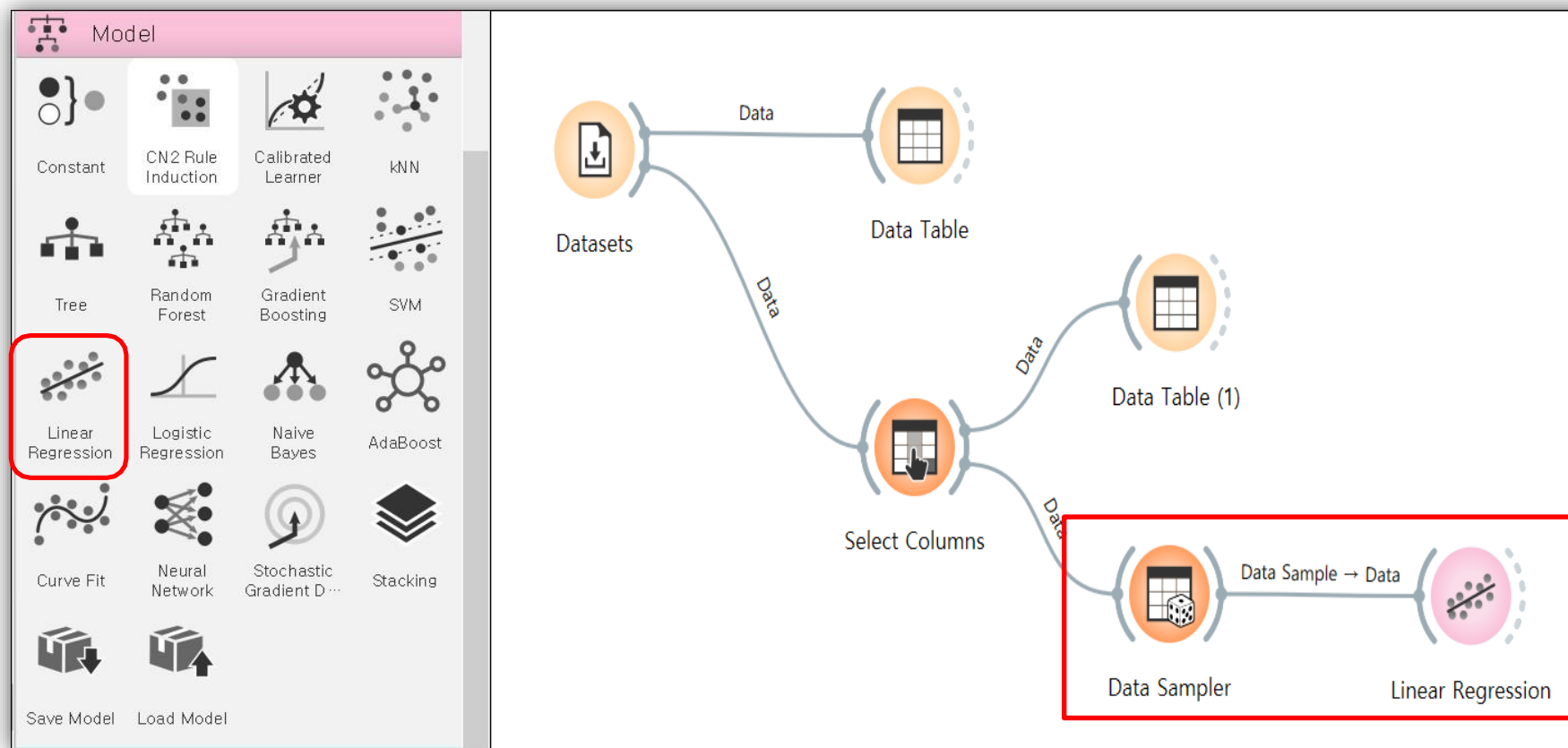
## ■ 훈련데이터와 테스트 데이터 비율 정하기



## ■ 모델 선택하기

### ■ 연속적인 값을 예측하는 모델

- Linear Regression 위젯



## ■ 학습시키기

Linear Regression - Orange

Name

Linear Regression

Parameters

☒ Fit intercept (unchecking it fixes it to zero)

Regularization

☒ No regularization

Regularization strength:

☐ Ridge regression (L2)

Alpha: 0.0001

☐ Lasso regression (L1)

Elastic net mixing:

L1

L2

0.50 : 0.50

☐ Elastic net regression

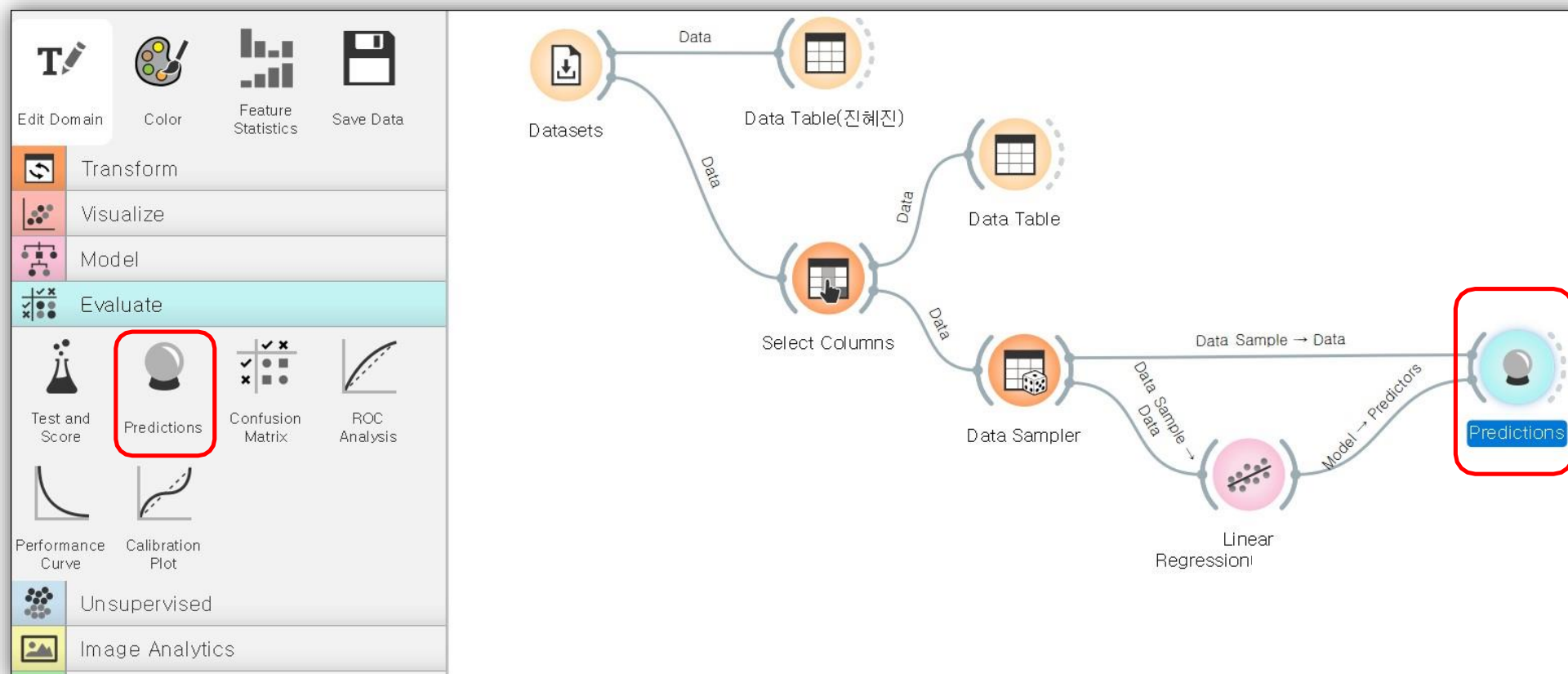
☒ Apply Automatically

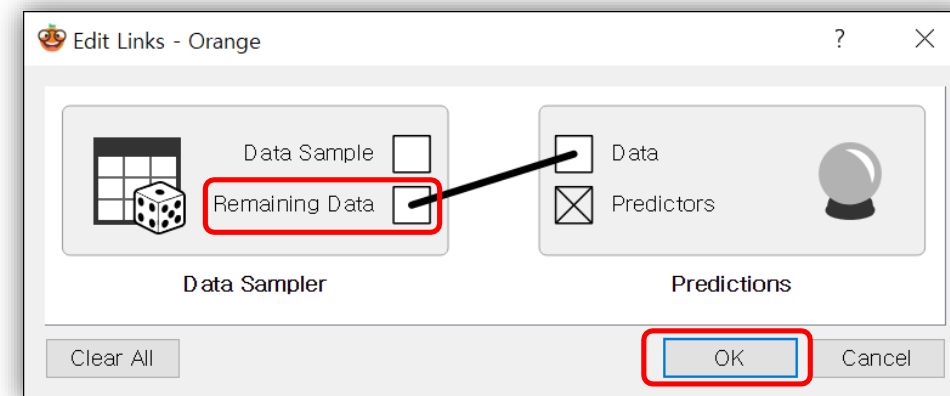
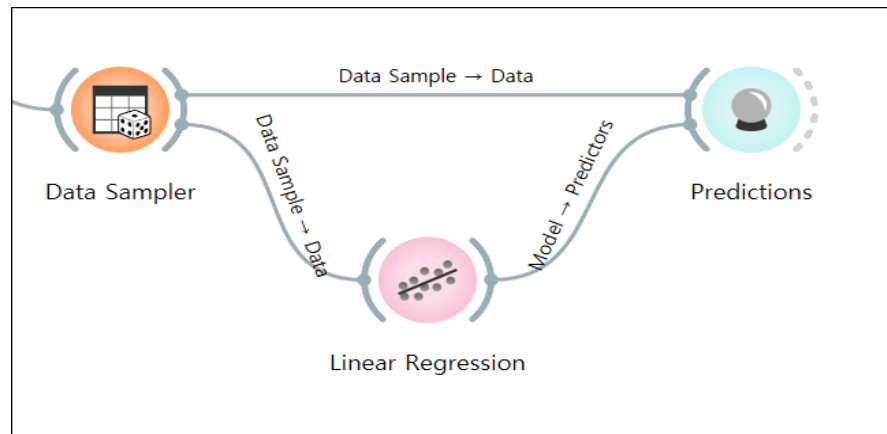
≡ ? 📄 | ↗ 2924 | - ↶ 11 | 🗒 📧



## ■ 모델 성능 확인하기

### ■ Evaluate 카테고리 Predictions 위젯





## ■ 성능 결과 확인하기

- 전복 데이터를 이용해서 얼마나 순살 무게를 정확히 예측했는지 확인한다.

