



G

I

S

T

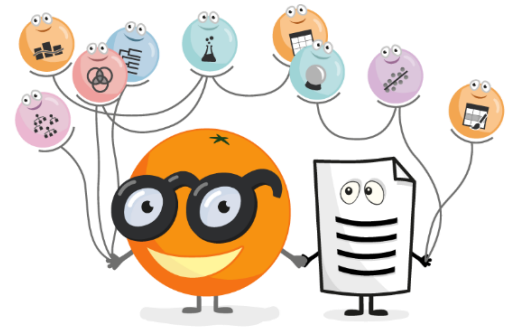


# 창의자율프로젝트

# REFERENCE SITE

<https://orangedatamining.com/>

<https://orange3.readthedocs.io/projects/orange-visual-programming/en/latest/widgets/model/neuralnetwork.html>



# REFERENCE SITE

## Orange

<b>Developer(s)</b>	University of Ljubljana
<b>Initial release</b>	10 October 1996; 26 years ago <sup>[1]</sup>
<b>Stable release</b>	3.34.0 <sup>[2]</sup> / 5 December 2022; 3 months ago
<b>Repository</b>	<a href="#">Orange Repository</a> 
<b>Written in</b>	Python, Cython, C++, C
<b>Operating system</b>	Cross-platform
<b>Type</b>	Machine learning, Data mining, Data visualization, Data analysis
<b>License</b>	GPLv3 or later <sup>[3][4]</sup>
<b>Website</b>	<a href="#">orangedatamining.com</a>  

## Visual Programming Front-End 기능 제공

- Machine Learning
- Data Mining
- Data Visualization
- Data Analysis

# CONTENTS

Orange Visual Programming

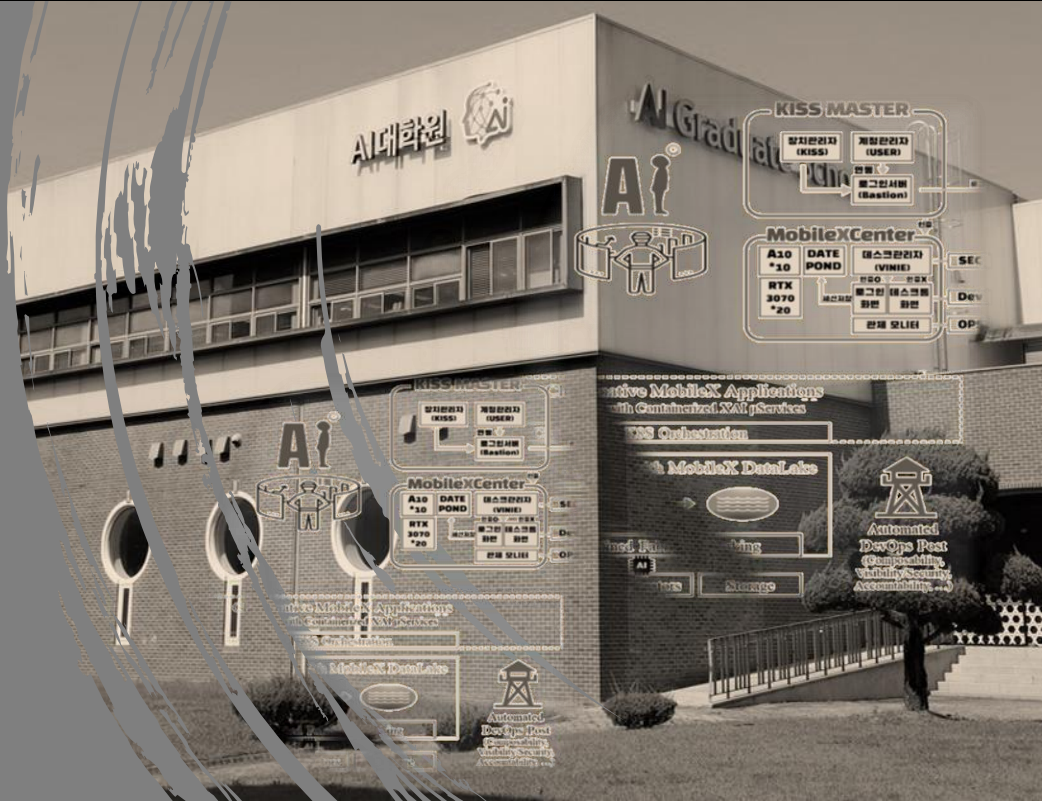
1. Abalone flesh weight prediction using linear regression

2. Star terror prevention using logistic regression

3. Courier delivery location clustering using K-Means



# 1. Abalone flesh weight prediction using Linear Regression



## Problem

- 전복의 나이테, 성별, 길이, 직경, 두께, 전체 무게, 내장 무게, 껍질 무게에 해당하는 총 8가지 데이터를 입력하면 AI가 전복의 순살(flesh) 무게를 예측할 수 있을까?

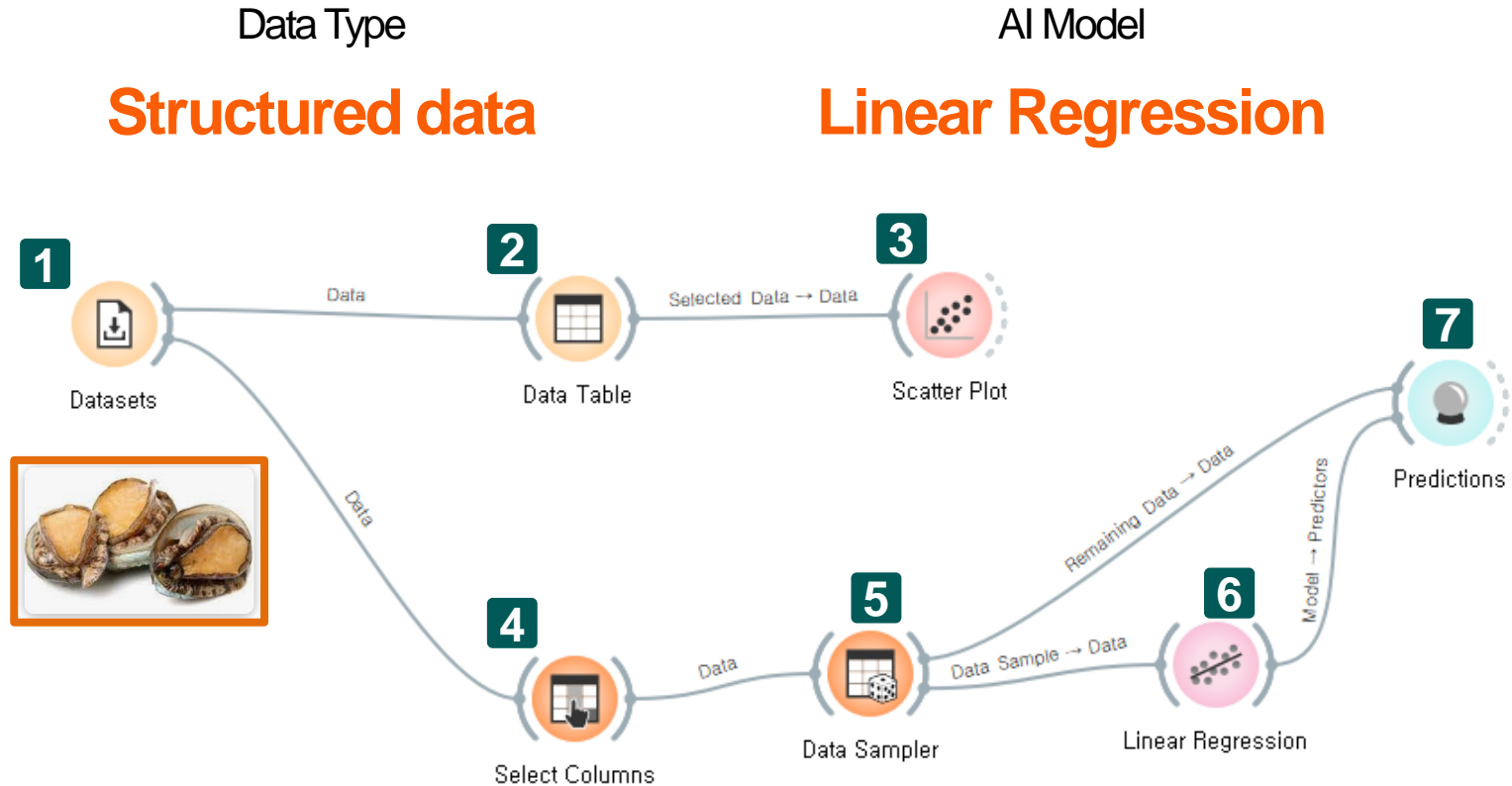


male



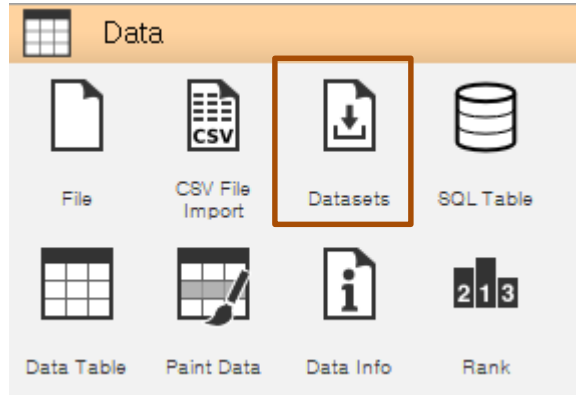
female

# 1. Abalone flesh weight prediction using Linear Regression



# 1. Abalone flesh weight prediction using Linear Regression

## 1 Datasets



The screenshot shows the 'Datasets - Orange' window. The 'Abalone' dataset is selected and highlighted in blue. Below the table, the 'Description' section is expanded, showing the source and details of the dataset.

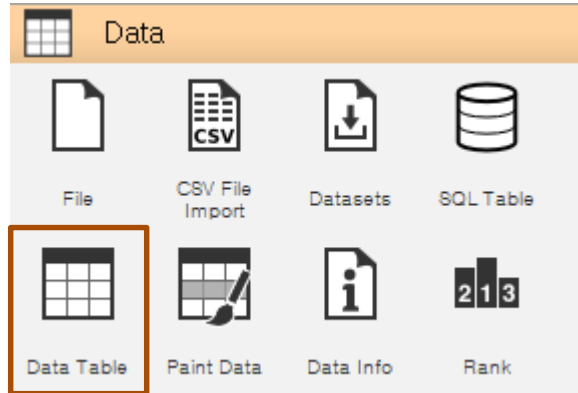
Title	Size	Instances	Variables	Target	Tags
Abalone	187.5 KB	4177	8	numeric	biology
Breast Cancer and Docetaxel Treatment	1.8 MB	24	9486	categorical	biology
Smoking effect on B lymphocytes	1.8 MB	79	3000	categorical	genomics
HDI	45.2 KB	188	52		economy, geo
SentiNews	5.0 MB	2000	7	categorical	text, sentiment
TKI resistance	1.2 MB	280	467	categorical	spectral
Adult	4.1 MB	32561	15	categorical	economy
Roman Amphorae	23.7 KB	164	16	categorical	archaeology, image analytics
Attrition - Predict	838 bytes	3	18	categorical	economy, synthetic, education
Attrition - Train	182.2 KB	1470	18	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.



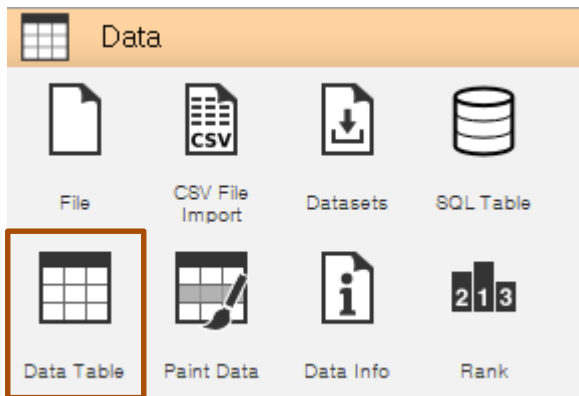
# 1. Abalone flesh weight prediction using Linear Regression

## 2 Data Table



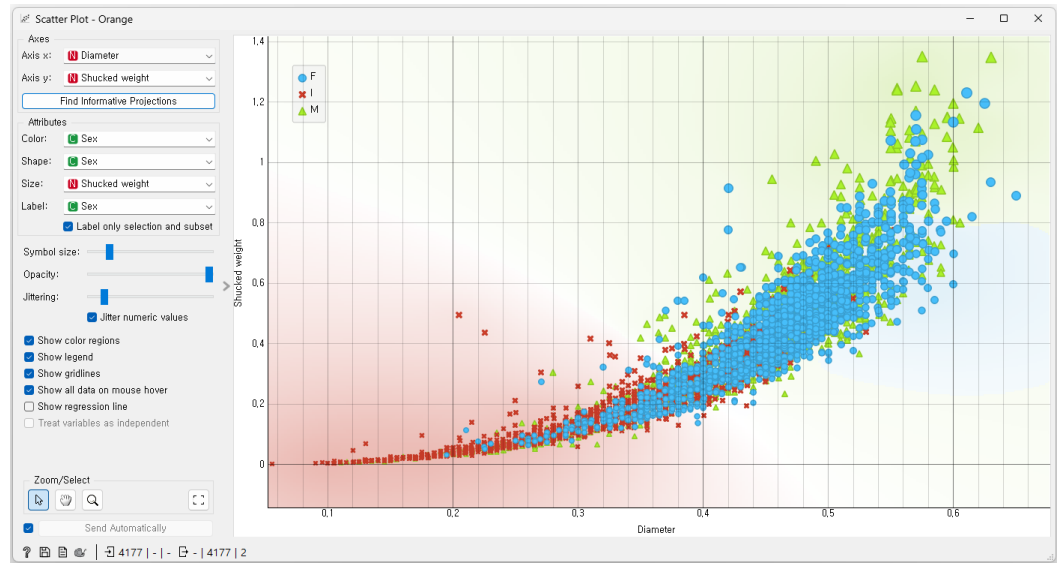
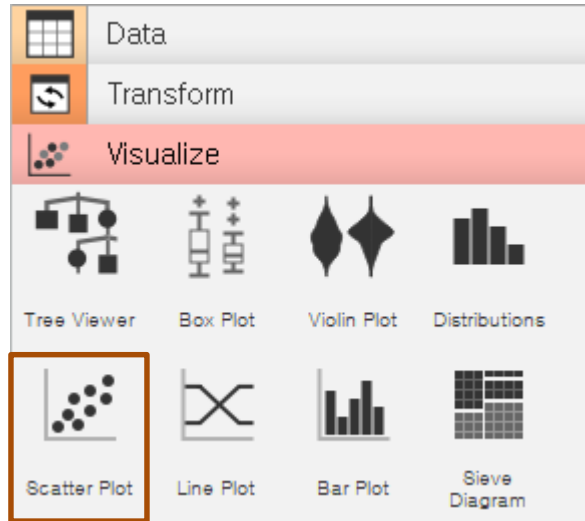
	Rings	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight
1	15	M	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.1500
2	7	M	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.0700
3	9	F	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.2100
4	10	M	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.1550
5	7	I	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.0550
6	8	I	0.425	0.300	0.095	0.3515	0.1410	0.0775	0.1200
7	20	F	0.530	0.415	0.150	0.7775	0.2370	0.1415	0.3300
8	16	F	0.545	0.425	0.125	0.7680	0.2940	0.1495	0.2600
9	9	M	0.475	0.370	0.125	0.5095	0.2165	0.1125	0.1650
10	19	F	0.550	0.440	0.150	0.8945	0.3145	0.1510	0.3200
11	14	F	0.525	0.380	0.140	0.6065	0.1940	0.1475	0.2100
12	10	M	0.430	0.350	0.110	0.4060	0.1675	0.0810	0.1350
4168	9	M	0.500	0.380	0.125	0.5770	0.2690	0.1265	0.1535
4169	8	F	0.515	0.400	0.125	0.6150	0.2865	0.1230	0.1765
4170	10	M	0.520	0.385	0.165	0.7910	0.3750	0.1800	0.1815
4171	10	M	0.550	0.430	0.130	0.8395	0.3155	0.1955	0.2405
4172	8	M	0.560	0.430	0.155	0.8675	0.4000	0.1720	0.2290
4173	11	F	0.565	0.450	0.165	0.8870	0.3700	0.2390	0.2490
4174	10	M	0.590	0.440	0.135	0.9660	0.4390	0.2145	0.2605
4175	9	M	0.600	0.475	0.205	1.1760	0.5255	0.2875	0.3080
4176	10	F	0.625	0.485	0.150	1.0945	0.5310	0.2610	0.2960
4177	12	M	0.710	0.555	0.195	1.9485	0.9455	0.3765	0.4950

## 2 Data Table

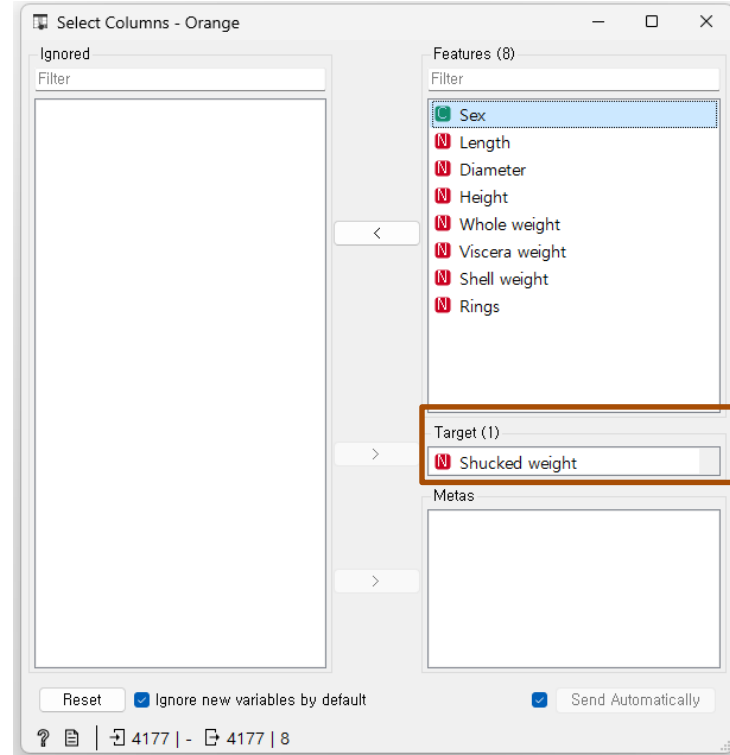
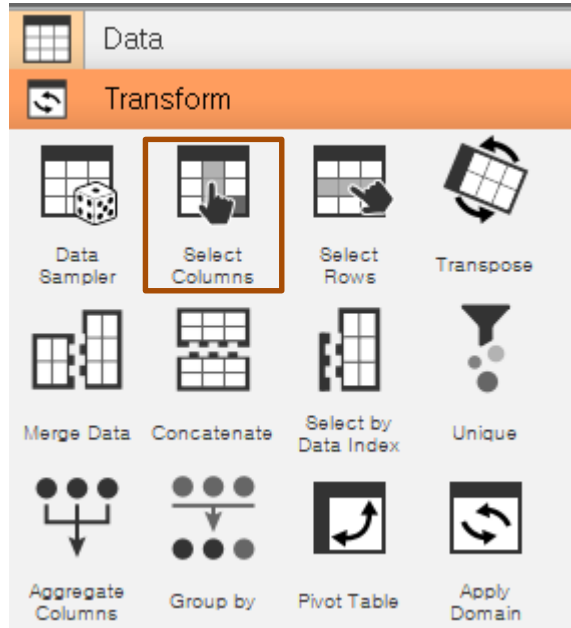


Attribute name	Attribute
Rings	나이테: 연도를 나타냄.
Sex	성별: M(수), F(암), I(유아)
Length	길이: 최장 껍질 측정(mm)
Diameter	직경: 길이에 수직(mm)
Height	두께: 껍질과 살 포함(mm)
Whole weight	전체 무게: 그램 단위(g)
<b>Shucked weight</b>	<b>순살 무게: 그램 단위(g)</b>
Viscera weight	내장 무게: 피를 뺀 후 장 무게(g)
Shell weight	껍질 무게: 건조 후(g)

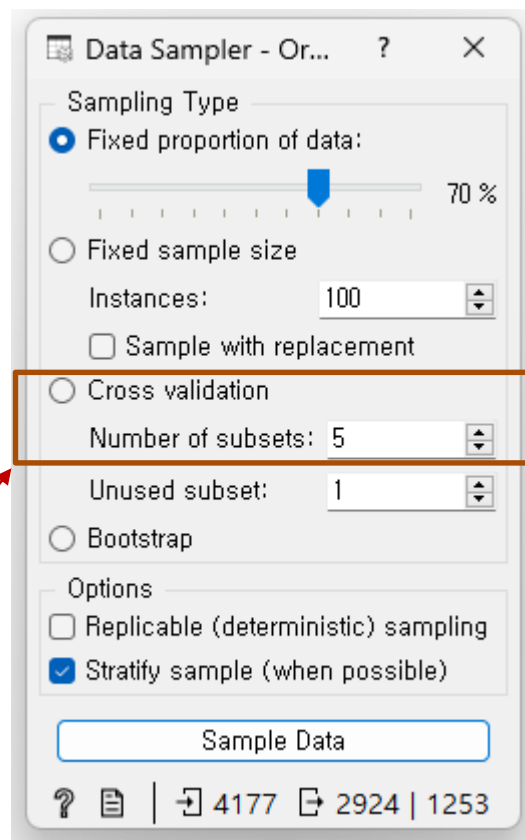
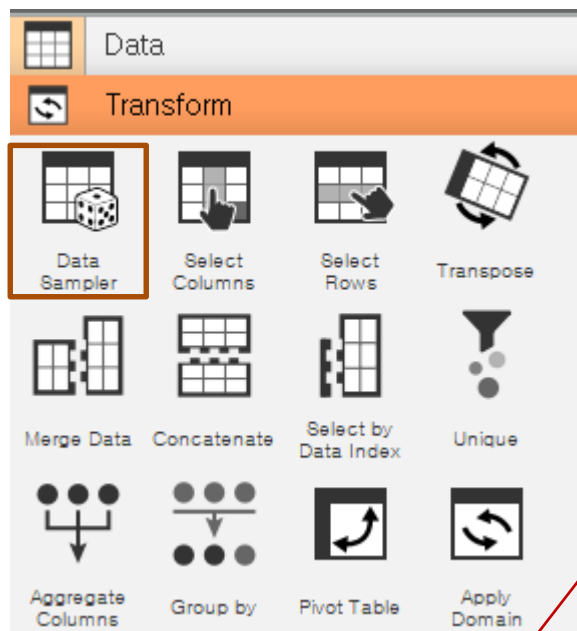
## 3 Scatter Plot



## 4 Select Columns

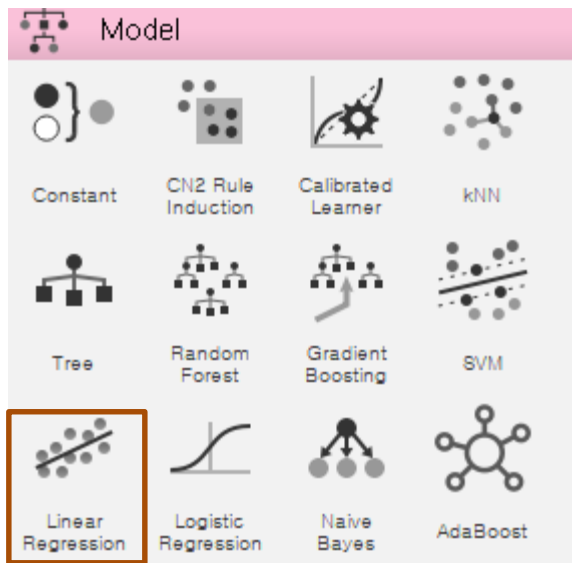


## 5 Data Sampler

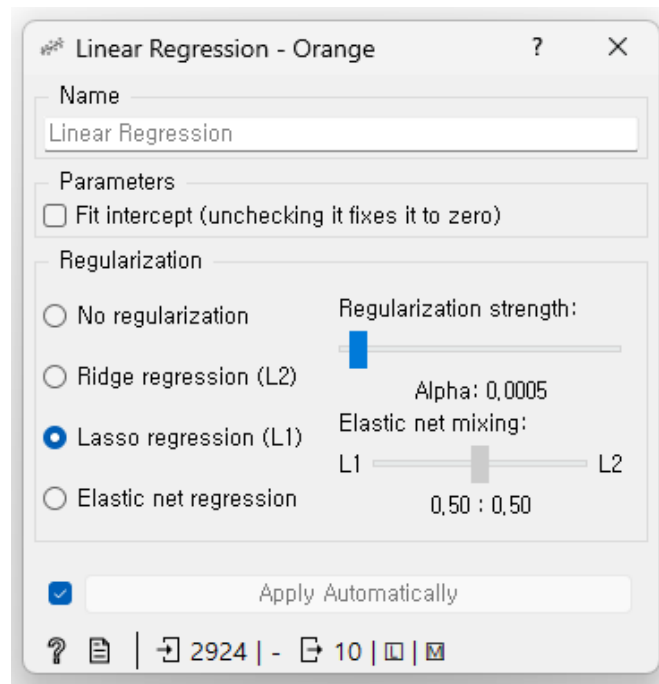


- **Cross validation:** Stratified K-fold Cross Validation

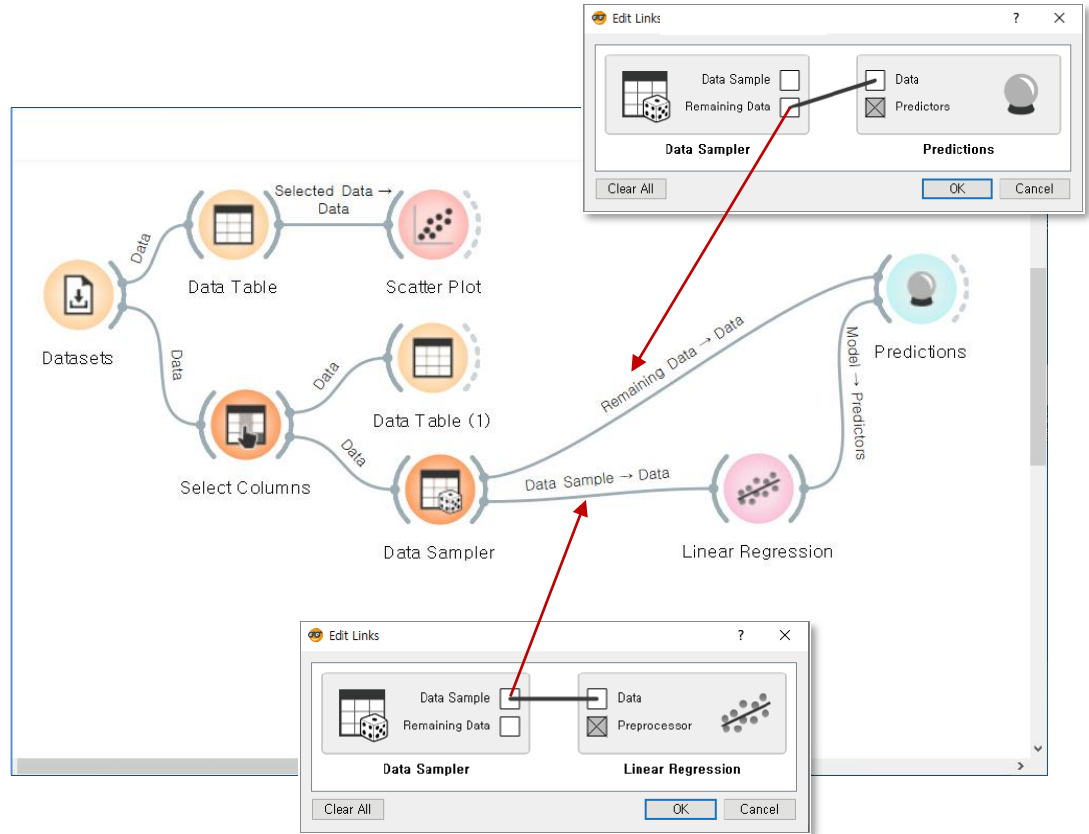
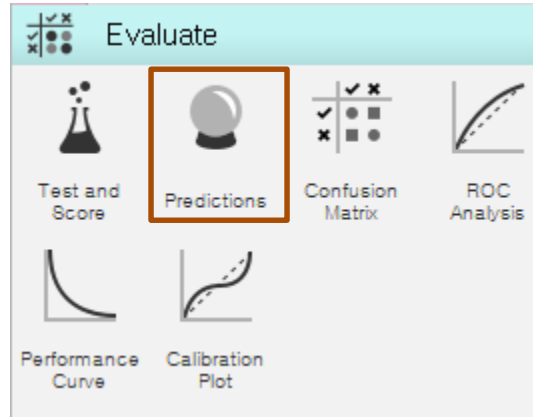
## 6 Linear Regression



- **Regularization**
  - Ridge regression(L2): 각 계수의 제곱을 더하는 방식
  - Lasso regression(L1): 각 계수의 절댓값을 더하는 방식
  - Elastic net regression: L2와 L1 방식을 절충한 것



## 7 Predictions



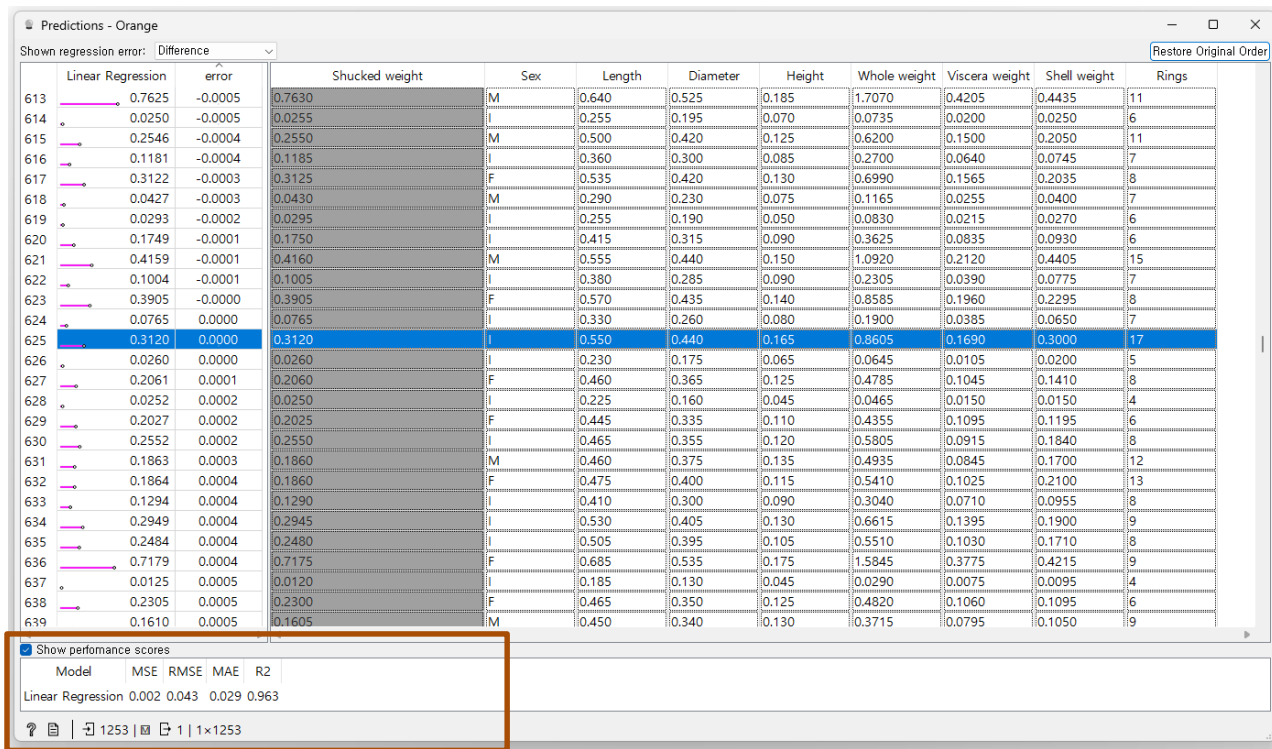
# 1. Abalone flesh weight prediction using Linear Regression

## 7 Predictions

### • Performance Evaluation

- MSE(Mean Squared Error)
- RMSE(Root Mean Squared Error)
- MAE(Mean Absolute Error)
- R2(R Squared)

MSE, RMSE, MAE는 0에 가까울수록, R2는 1에 가까울수록 정확도가 높음





### Conclusion

- 전복의 나이테, 성별, 길이, 직경, 두께, 전체 무게, 내장 무게, 껍질 무게에 해당하는 총 8개의 변수와 전복 순살(flesh)의 상관 관계를 분석함
- Linear regression을 이용하여 8가지 변수 값에 따른 전복의 순살 무게를 예측함

## 2. Star Terror Prevention Using Logistic Regression



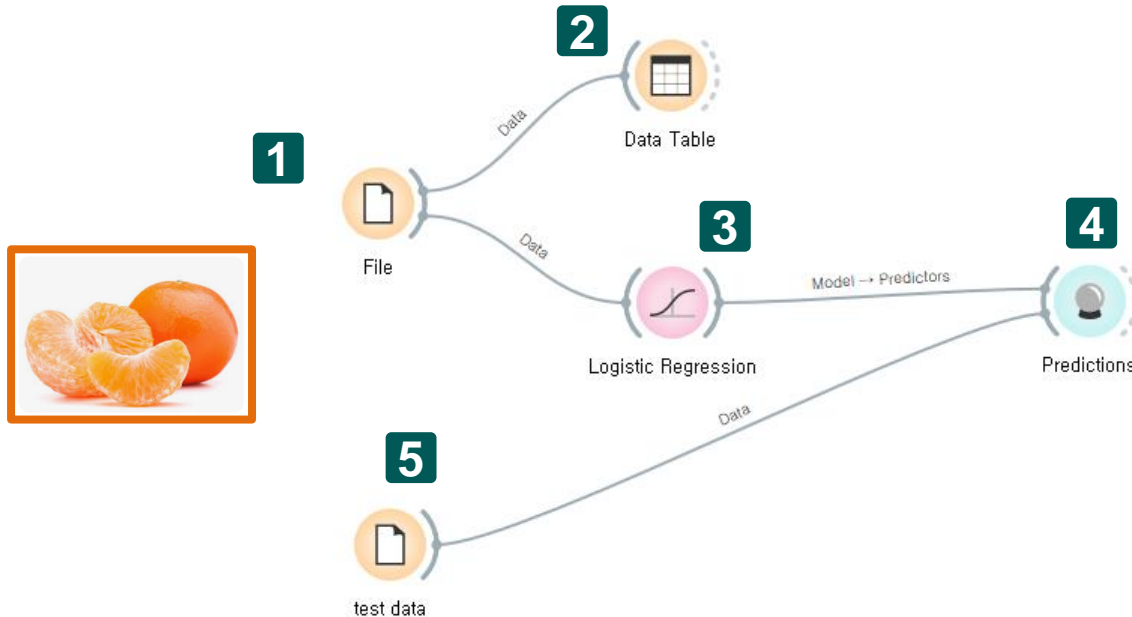
## 19

Data Type

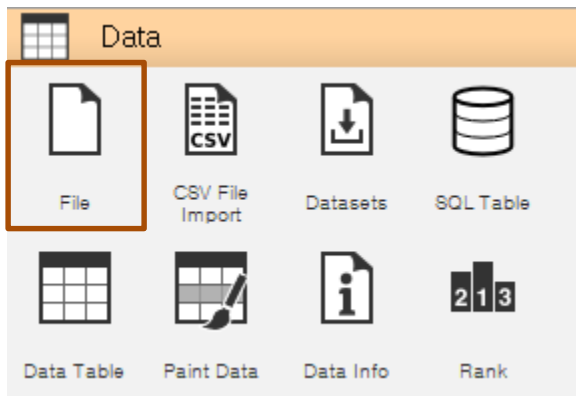
Structured data

AI Model

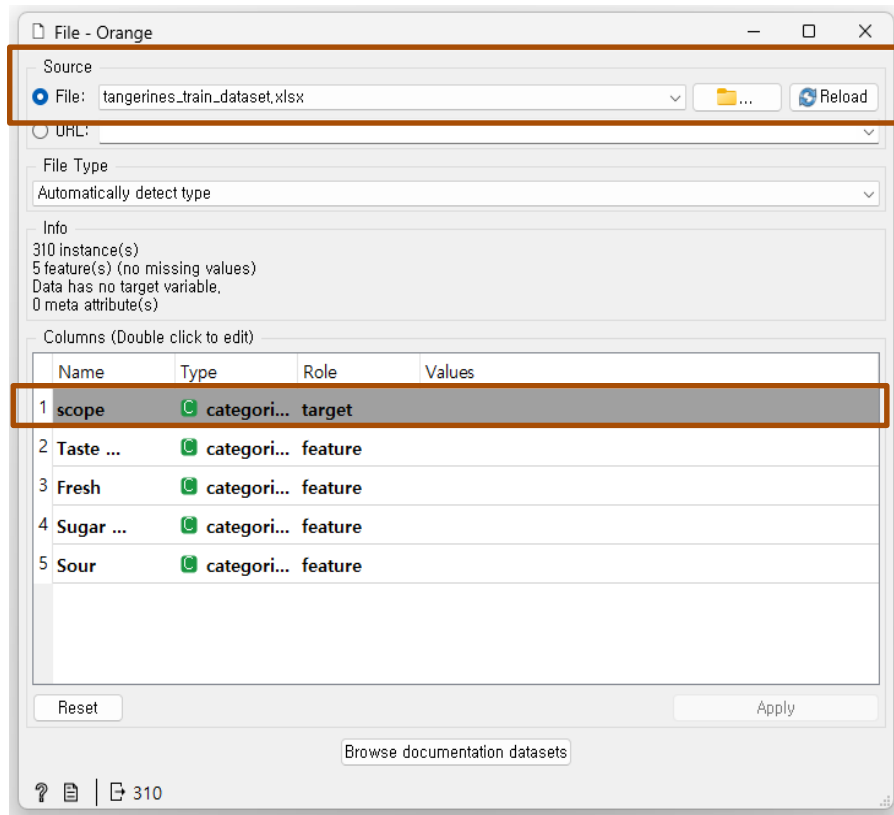
Multinomial Logistic Regression



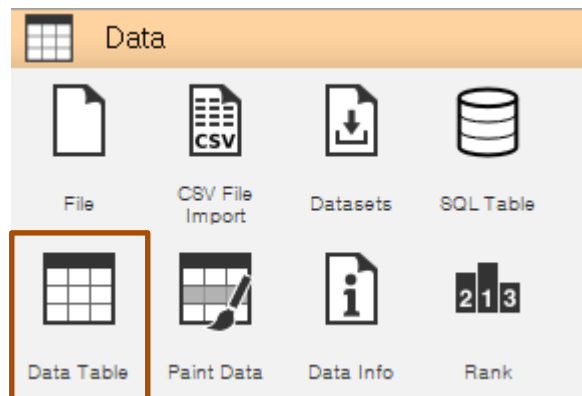
### 1 Datasets



- tangerines\_train\_dataset.xlsx  
(310 instances)
- tangerines\_test\_dataset.xlsx  
(15 instances)



## 2 Data Table



Data Table

Info  
310 instances (no missing data)  
5 features  
No target variable.  
No meta attributes

Variables  
☒ Show variable labels (if present)  
☐ Visualize numeric values  
☒ Color by instance classes

Selection  
☒ Select full rows

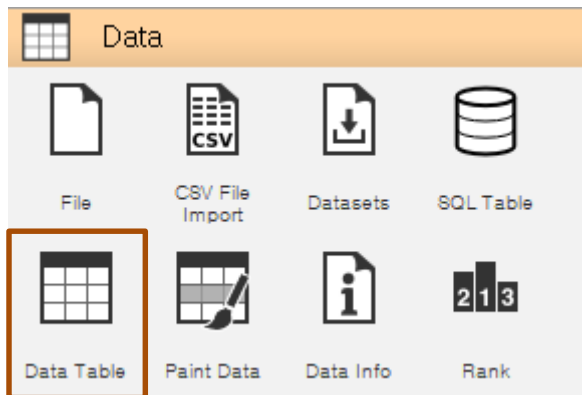
Restore Original Order

☒ Send Automatically

	scope	Taste satisfaction	Fresh	Sugar content	Sour
1	1	3	3	2	3
2	1	3	2	3	3
3	1	2	3	3	3
4	1	3	3	3	3
5	1	3	3	3	3
6	1	3	3	3	3
7	1	3	3	3	3
8	1	3	3	2	3
9	1	3	3	3	2
10	1	3	2	3	3
11	1	3	3	2	3
12	1	3	2	3	3
13	1	3	3	3	3
14	1	2	3	3	3
15	1	3	3	3	3
16	1	3	3	3	3
17	1	3	3	3	3
18	1	3	3	3	3
19	1	3	3	3	3
20	1	3	3	3	3
21	1	3	3	3	1
22	1	3	2	3	1
23	1	3	2	3	1
24	1	3	3	3	1
25	1	3	3	3	1

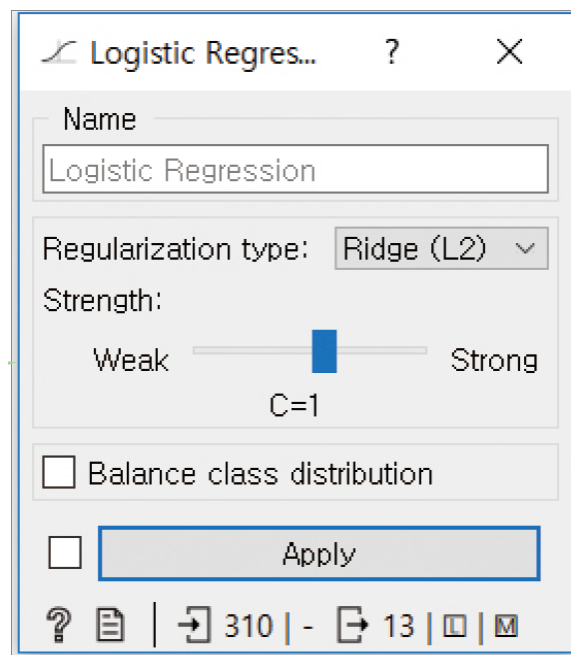
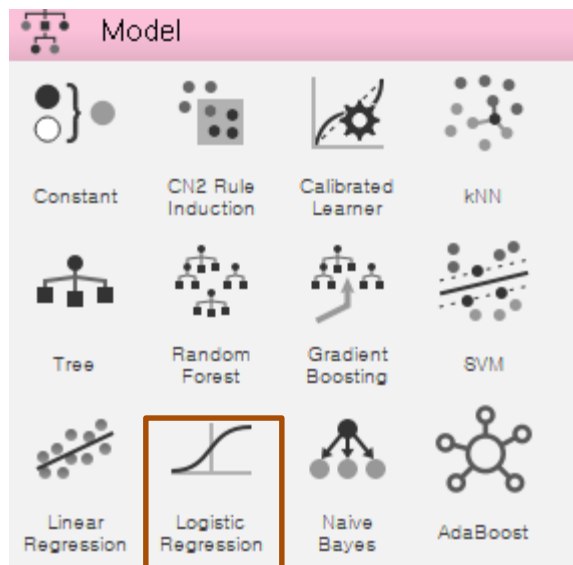
310 | 310 | 310

### 2 Data Table



Attribute name	Attribute
scope	종합 평점 (1~5: 점수가 높을수록 좋은 상품)
Taste satisfaction	맛 만족도 (1: 예상보다 맛있어요, 2: 괜찮아요, 3: 예상보다 맛이 없어요.)
Fresh	싱싱함 (1: 예상보다 싱싱해요, 2: 보통이에요, 3: 예상보다 싱싱하지 않아요.)
Sugar content	당도 (1: 아주 달콤해요, 2: 적당히 달아요, 3: 달지 않아요.)
Sour	새콤함 (1: 많이 새콤해요, 2: 적당히 새콤해요, 3: 새콤하지 않아요.)

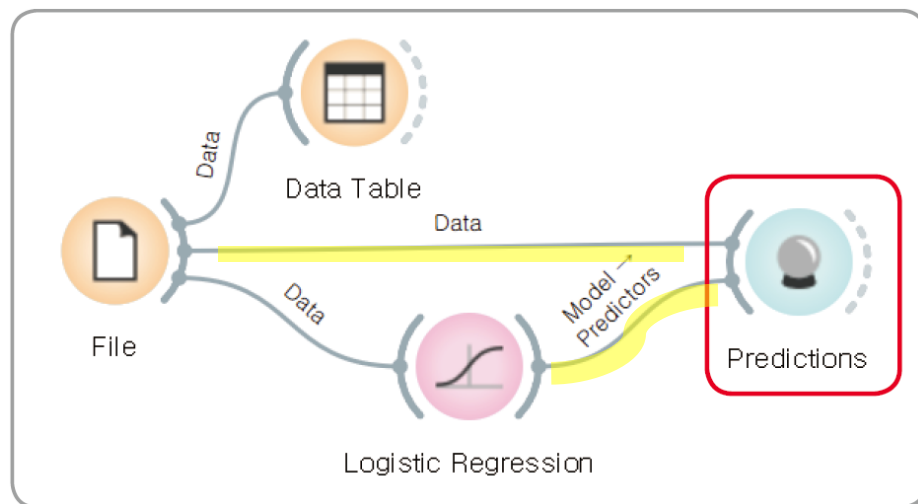
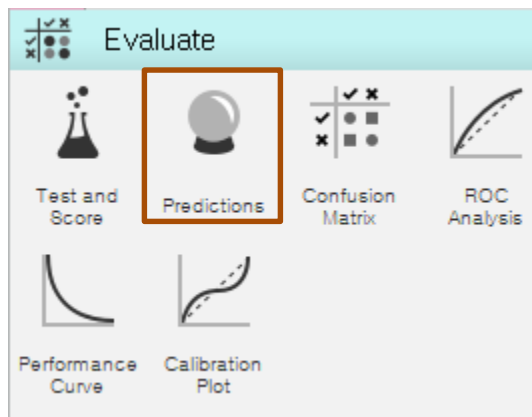
### 3 Logistic Regression



- **Regularization**
  - Ridge: 분류를 위한 식의 가중치 제공의 합
  - Lasso: 분류를 위한 식의 가중치 절댓값의 합
  - Weak Strong: 데이터를 분류할 때의 강도



### 4 Predictions



## 4 Predictions

Predictions - Orange

Show probabilities for: **Classes in data** Restore Original Order

	Logistic Regression	scope	taste satisfaction	Fresh	Sugar content	Sour
1	0.39 : 0.36 : 0.25 : 0.00 : 0.00 → 1	1	3	3	3	3
2	0.34 : 0.37 : 0.29 : 0.00 : 0.00 → 2	1	3	2	3	3
3	0.45 : 0.14 : 0.41 : 0.00 : 0.00 → 1	1	3	3	3	3
4	0.75 : 0.17 : 0.08 : 0.00 : 0.00 → 1	1	3	3	3	3
5	0.75 : 0.17 : 0.08 : 0.00 : 0.00 → 1	1	3	3	3	3
6	0.75 : 0.17 : 0.08 : 0.00 : 0.00 → 1	1	3	3	3	3
7	0.75 : 0.17 : 0.08 : 0.00 : 0.00 → 1	1	3	3	3	3
8	0.39 : 0.36 : 0.25 : 0.00 : 0.00 → 1	1	3	2	3	3
9	0.38 : 0.24 : 0.37 : 0.00 : 0.00 → 1	1	3	3	3	2
10	0.34 : 0.37 : 0.29 : 0.00 : 0.00 → 2	1	2	3	3	3
11	0.39 : 0.36 : 0.25 : 0.00 : 0.00 → 1	1	3	2	3	3
12	0.34 : 0.37 : 0.29 : 0.00 : 0.00 → 2	1	2	3	3	3
13	0.75 : 0.17 : 0.08 : 0.00 : 0.00 → 1	1	3	3	3	3
14	0.45 : 0.14 : 0.41 : 0.00 : 0.00 → 1	1	3	3	3	3
15	0.75 : 0.17 : 0.08 : 0.00 : 0.00 → 1	1	3	3	3	3
16	0.75 : 0.17 : 0.08 : 0.00 : 0.00 → 1	1	3	3	3	3
17	0.75 : 0.17 : 0.08 : 0.00 : 0.00 → 1	1	3	3	3	3
18	0.75 : 0.17 : 0.08 : 0.00 : 0.00 → 1	1	3	3	3	3
19	0.75 : 0.17 : 0.08 : 0.00 : 0.00 → 1	1	3	3	3	3
20	0.75 : 0.17 : 0.08 : 0.00 : 0.00 → 1	1	3	3	3	3

☒ Show performance scores Target class: (Average over classes)

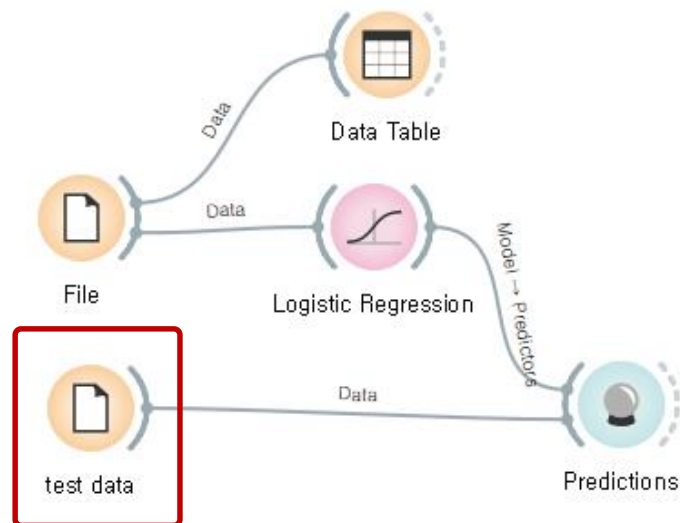
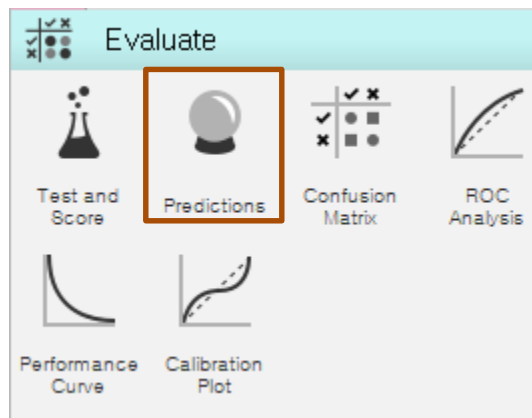
Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.952	0.819	0.811	0.825	0.819

Logistic Regression의 분류가 scope와 유사

모델 성능 지표 AUC가 0.952로 매우 높은 정확도를 보여줌(AUC: 0~1)

27

### 4 Predictions



## 4 Predictions

Predictions - Orange

Show probabilities for: Classes in data ☒ Show classification errors Restore Original Order

	Logistic Regression	error	scope	aste satisfactio	Fresh	Sugar content	Sour
1	0.00 : 0.01 : 0.25 : 0.73 : 0.01 → 4	0.999	1	2	2	2	2
4	0.00 : 0.00 : 0.00 : 0.05 : 0.94 → 5	0.997	2	1	1	1	2
5	0.00 : 0.00 : 0.02 : 0.14 : 0.84 → 5	0.997	2	1	1	2	2
6	0.00 : 0.22 : 0.07 : 0.19 : 0.51 → 5	0.781	2	2	1	1	1
9	0.00 : 0.00 : 0.00 : 0.05 : 0.94 → 5	0.997	3	1	1	1	2
10	0.75 : 0.17 : 0.08 : 0.00 : 0.00 → 1	1.000	4	3	3	3	3
13	0.02 : 0.11 : 0.70 : 0.17 : 0.01 → 3	0.988	5	2	2	2	3
14	0.01 : 0.02 : 0.04 : 0.03 : 0.89 → 5	0.109	5	1	1	2	3
15	0.00 : 0.02 : 0.19 : 0.66 : 0.14 → 4	0.862	5	2	1	2	2

☒ Show performance scores Target class: (Average over classes)

Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.367	0.067	0.040	0.029	0.067

15 | 15 | 1x15

1번 사례는 맛 만족도, 싱싱함, 당도, 새콤함이 모두 괜찮았는데 별 1개, **Logistic Regression 예측 평점 4**

6번 사례는 맛 만족도 보통, 나머지 아주 괜찮았지만, 실제 검질 까기 귀찮아서 별 2개, **Logistic Regression 예측 평점 5**

10번 사례는 전체적으로 상품이 불만족스럽지만, 낮은 별점을 주기 곤란하여 별 4개, **Logistic Regression 예측 평점 1**

## 딥러닝 모델 평가 지표

### Confusion Matrix

		예측값	
		Positive	Negative
실제값	True	TP	FN
	False	FP	TN

Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.952	0.819	0.811	0.825	0.819

? | 310 | 310 | 1×310

train data

Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.367	0.067	0.040	0.029	0.067

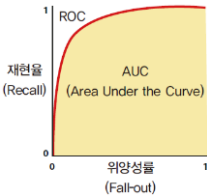
? | 15 | 15 | 1×15

test data

Positive: 1로 예측, Negative: 0으로 예측

- TP(True Positive): 실제값이 True인 것을 Positive라고 예측
- TN(True Negative): 실제값이 False인 것을 Negative라고 예측
- FP(False Positive): 실제값이 False인 것을 Positive라고 예측
- FN(False Negative): 실제값이 True인 것을 Negative라고 예측

## 딥러닝 모델 평가 지표

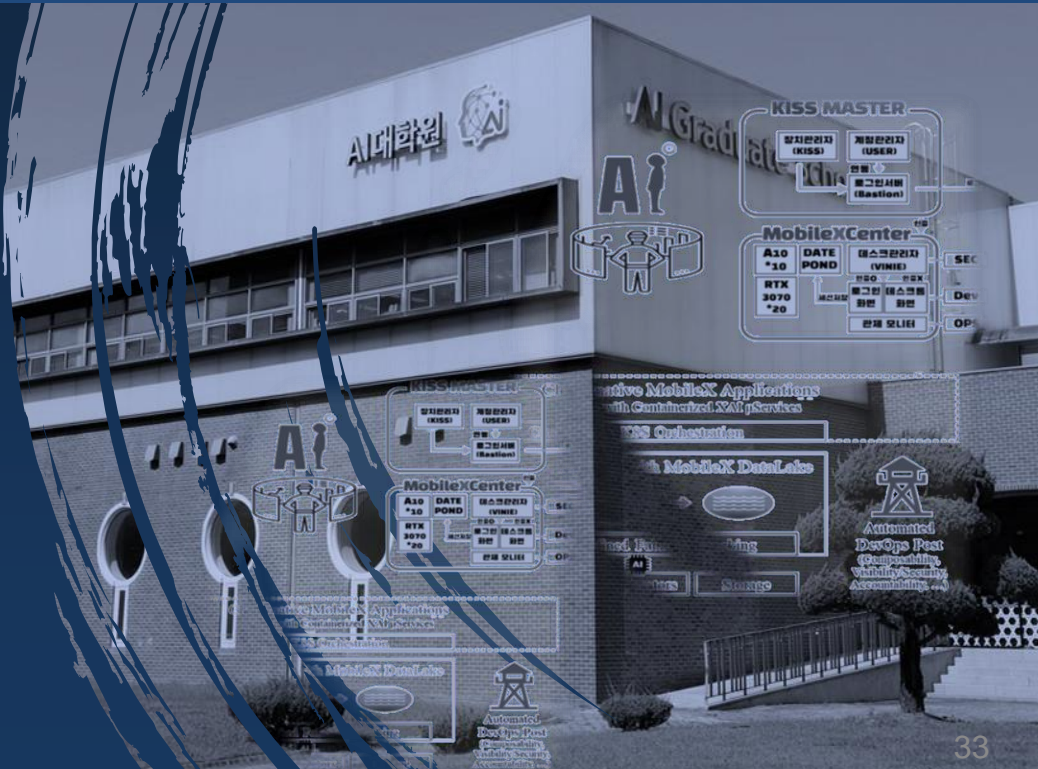
지표	의미	그래프또는식
AUC (Area Under the ROC Curve)	재현율(Recall, 실제 True인 것 중에서 모델이 True라고 분류한 것)과 위양성률(Fall-out, 실제 False인 것 중에서 모델이 True라고 분류한 것의 비율 관계를 나타낸 ROC(Receiver Operating Characteristic) 그래프의 아래쪽 면적	
분류 정확도 (CA)	모델이 입력된 데이터에 대해 얼마나 정확하게 분류하는지를 나타내는 값 (1에 가까울수록 정확도가 높음)	$\frac{TP + TN}{TP + FP + TN + FN}$
정밀도 (Precision)	모델이 True라고 분류한 것 중에서 실제 True인 것의 비율	$\frac{TP}{TP + FP}$
재현율 (Recall)	실제 True인 것 중에서 모델이 True라고 분류한 것의 비율	$\frac{TP}{TP + FN}$
F1	정밀도와 재현율, 두 값의 조화 평균으로 하나의 수치로 나타낸 지표	

### Conclusion

- 맛 만족도, 싱싱함, 당도, 새콤함 등의 변수가 모두 좋더라도 별 개수가 작게 나올 수 있음이 분석되었음
- 사용자의 별 평가 개수가 전체적인 맛 평가와 유사하다고 할 수 없음
- Logistic Regression은 맛 만족도, 싱싱함, 당도, 새콤함 등의 변수를 기반으로 전체적인 맛 평가 결과를 신뢰할 수 있게 예측함

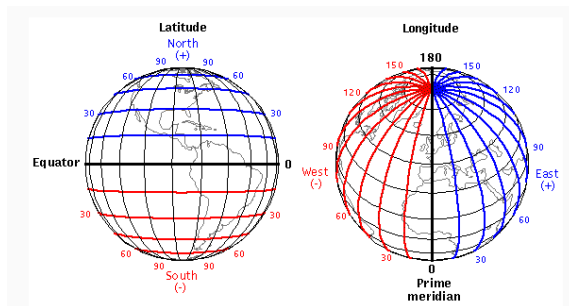


### 3. Courier delivery location clustering using k-Means



## Problem

- 주소지 중심으로 택배를 분류하면, 인근 거리임에도 행정 구역상 주소지가 다를 경우 택배 배달원의 배달 업무가 비효율적으로 수행될 수 있음
- AI가 택배 배달원의 효율적인 배달 업무를 위해서 주소지 중심으로 택배를 분류하지 않고, 인근 거리 위주로 택배 배달 물품을 분류할 수 있을까?



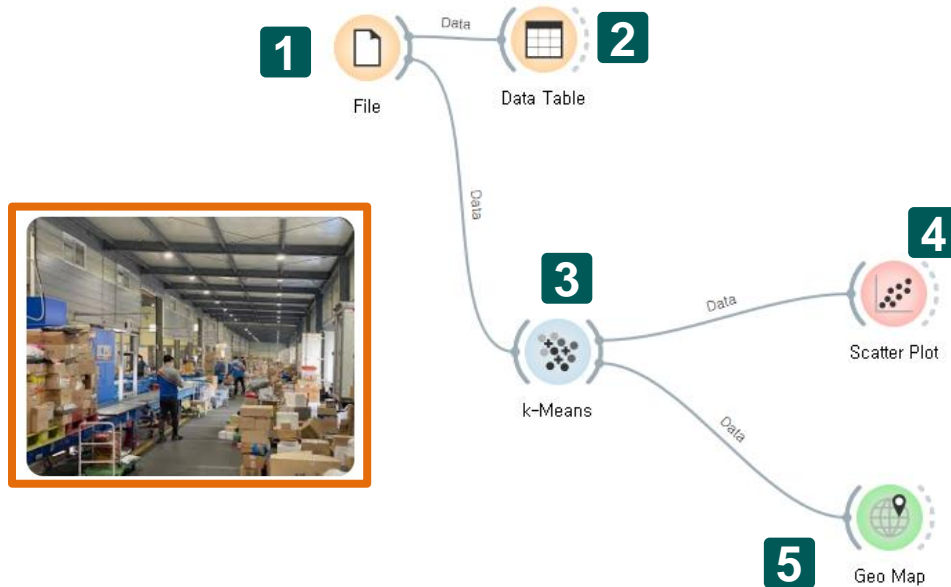
### 3. Courier delivery location clustering using k-Means

Data Type

**Structured data**

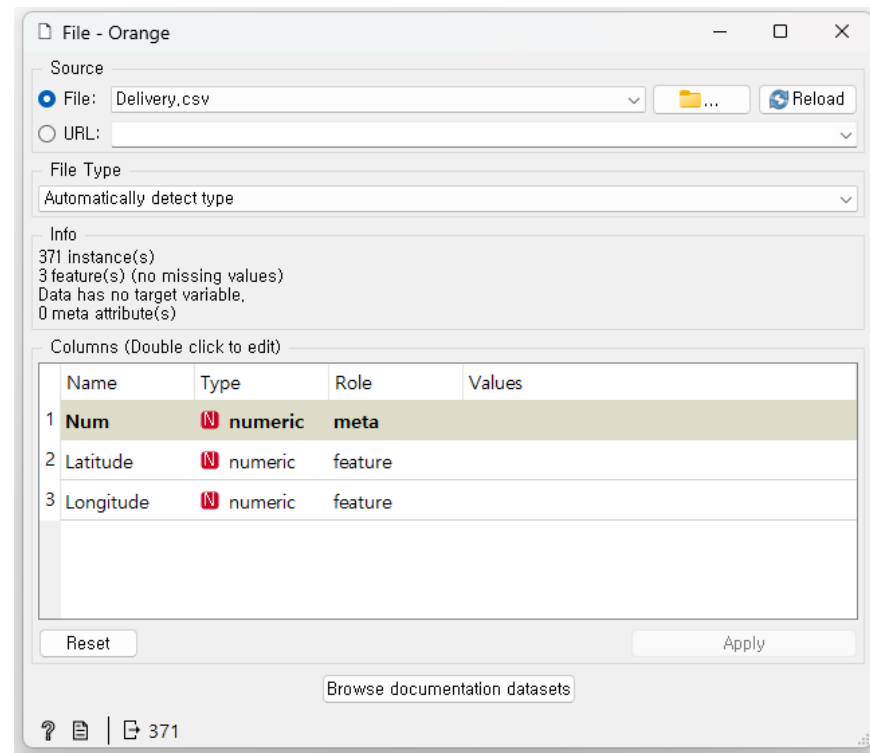
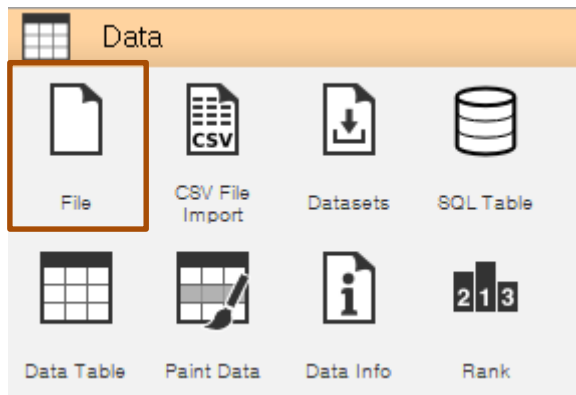
AI Model

**K-Means**

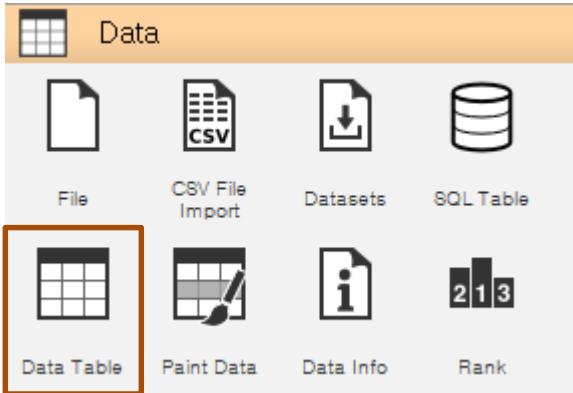


### 3. Courier delivery location clustering using k-Means

#### 1 File



## 2 Data Table



Data Table - Orange

Info  
371 instances (no missing data)  
2 features  
No target variable,  
1 meta attribute

Variables  
☒ Show variable labels (if present)  
☒ Visualize numeric values  
☒ Color by instance classes

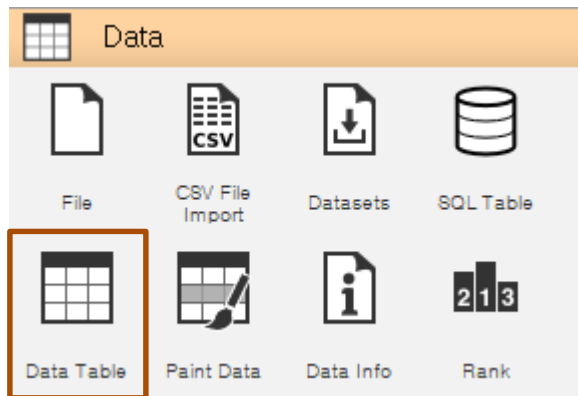
Selection  
☐ Select full rows

Restore Original Order  
☒ Send Automatically

	Núm	Latitude	Longitude
1	1	37.3368	126.713
2	2	37.5013	126.788
3	3	37.5225	126.777
4	4	37.5112	126.743
5	5	37.5088	126.738
6	6	37.5285	126.741
7	7	37.511	126.779
8	8	37.5294	126.742
9	9	37.5164	126.734
10	10	37.5133	126.735
11	11	37.486	126.802
12	12	37.4591	126.711
13	13	37.4998	126.756
14	14	37.419	126.697
15	15	37.5465	126.736
16	16	37.4877	126.672
17	17	37.5129	126.731
18	18	37.4518	126.709
19	19	37.5339	126.723
20	20	37.4491	126.737
21	21	37.4169	126.703
22	22	37.4872	126.81
23	23	37.493	126.756

? | 371 | 371 | 371

## 2 Data Table

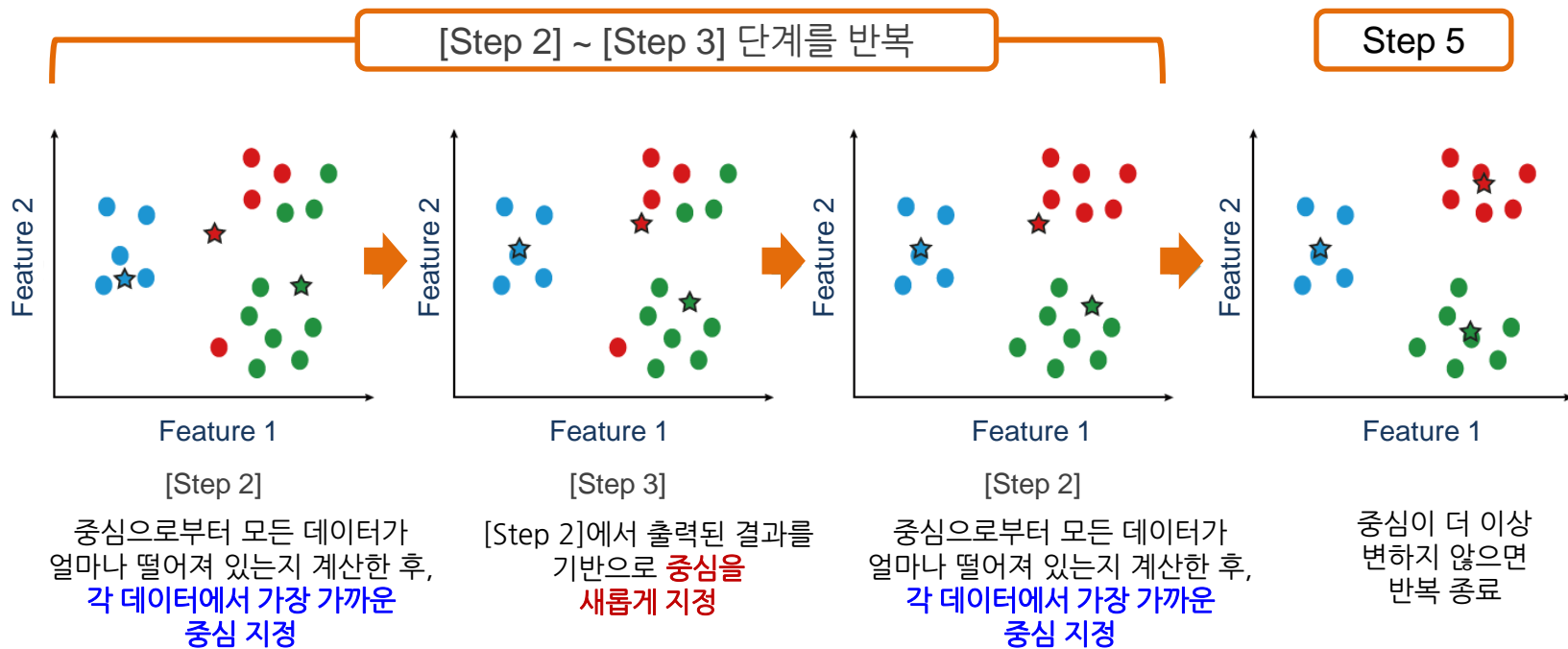


Attribute name	Attribute
Num	일련번호
Latitude	위도
Longitude	경도

#### Clustering process of k-Means

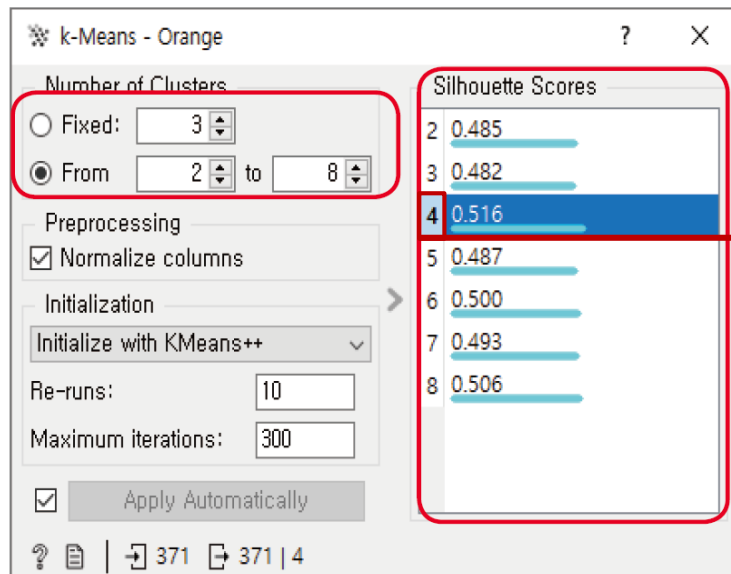
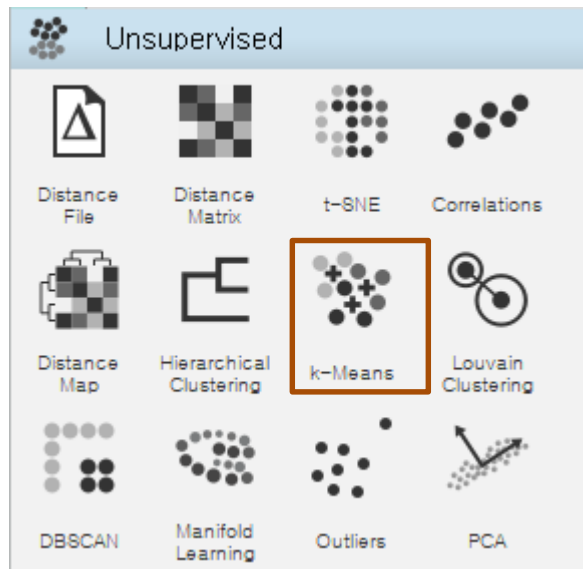


#### Clustering process of k-Means





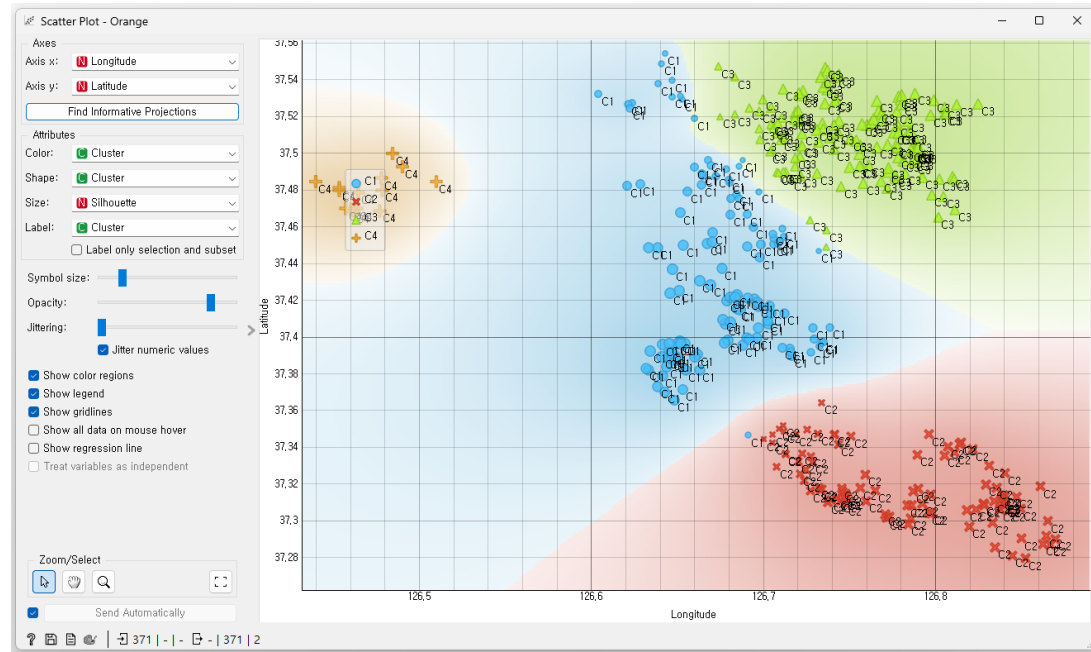
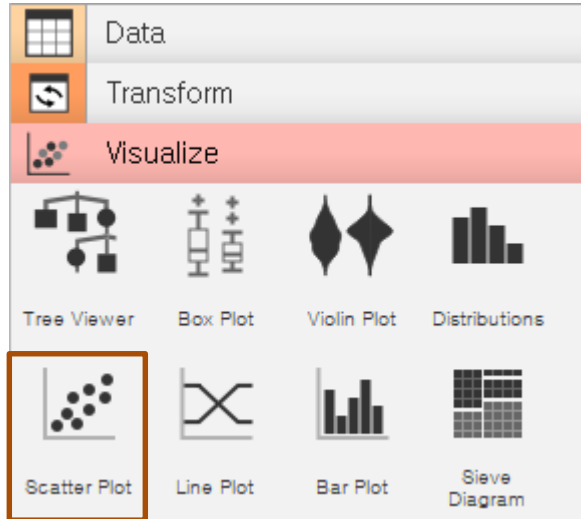
## 3 K-Means



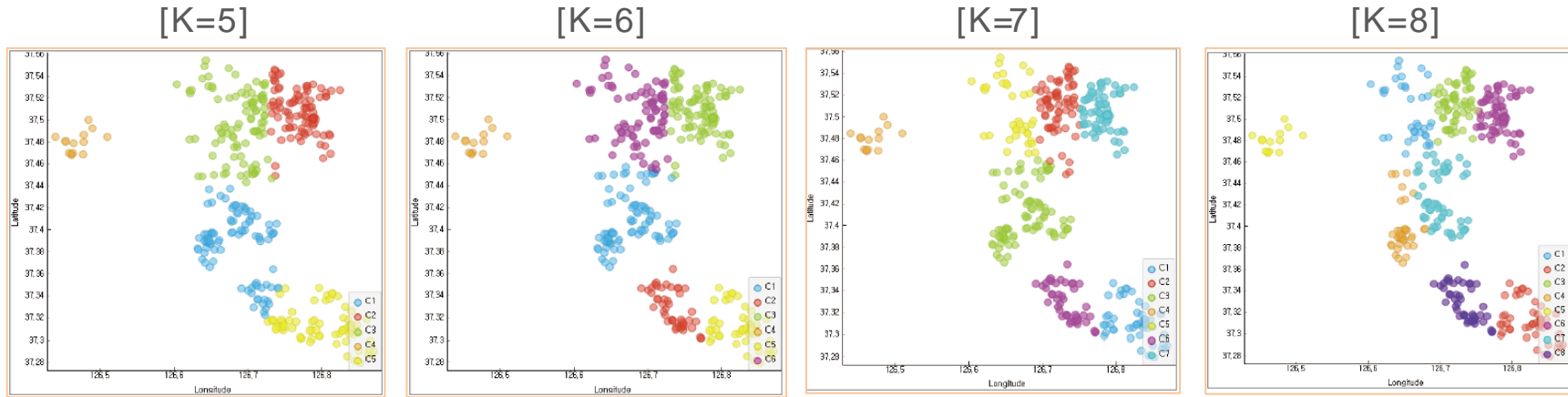
- **Number of Clusters**

- Fixed: 원하는 cluster의 개수를 설정
- From: **Silhouette Scores**를 보여 주는 범위를 설정  
[Silhouette Scores] 해당 범위 내에서 가장 높은 점수의 cluster 개수를 추천

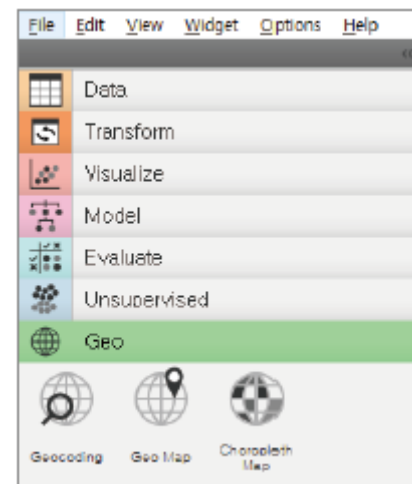
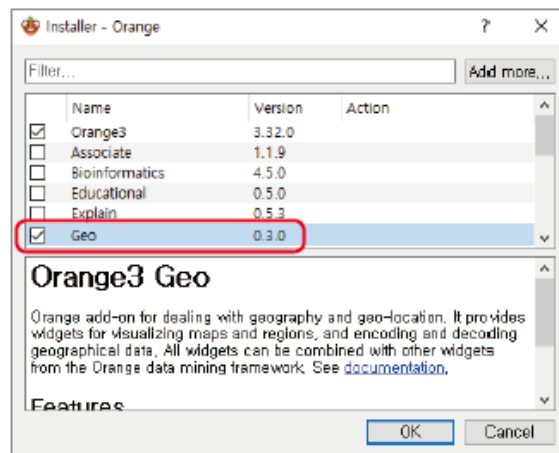
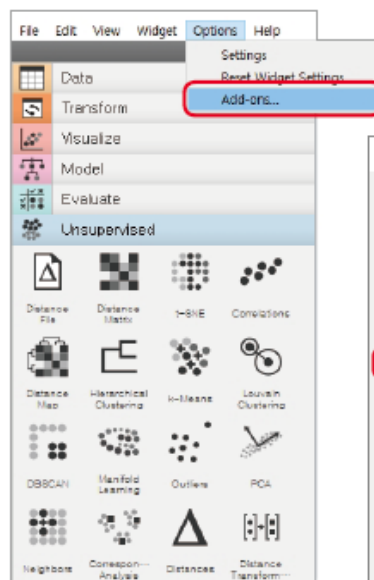
#### 4 Scatter Plot



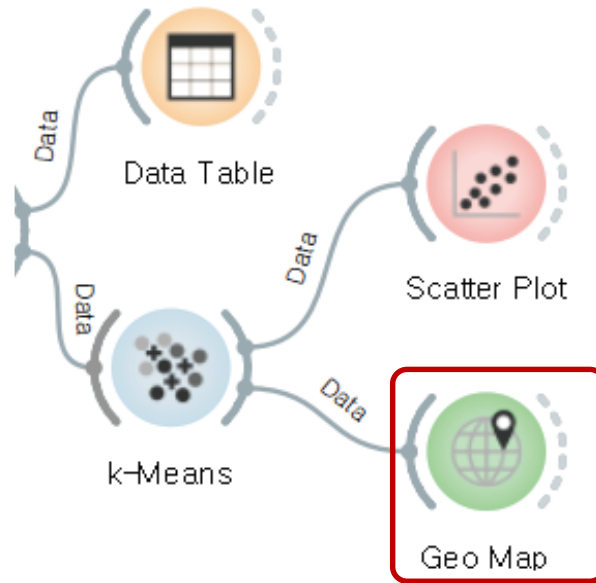
#### 4 Scatter Plot



#### 5 Geo Map

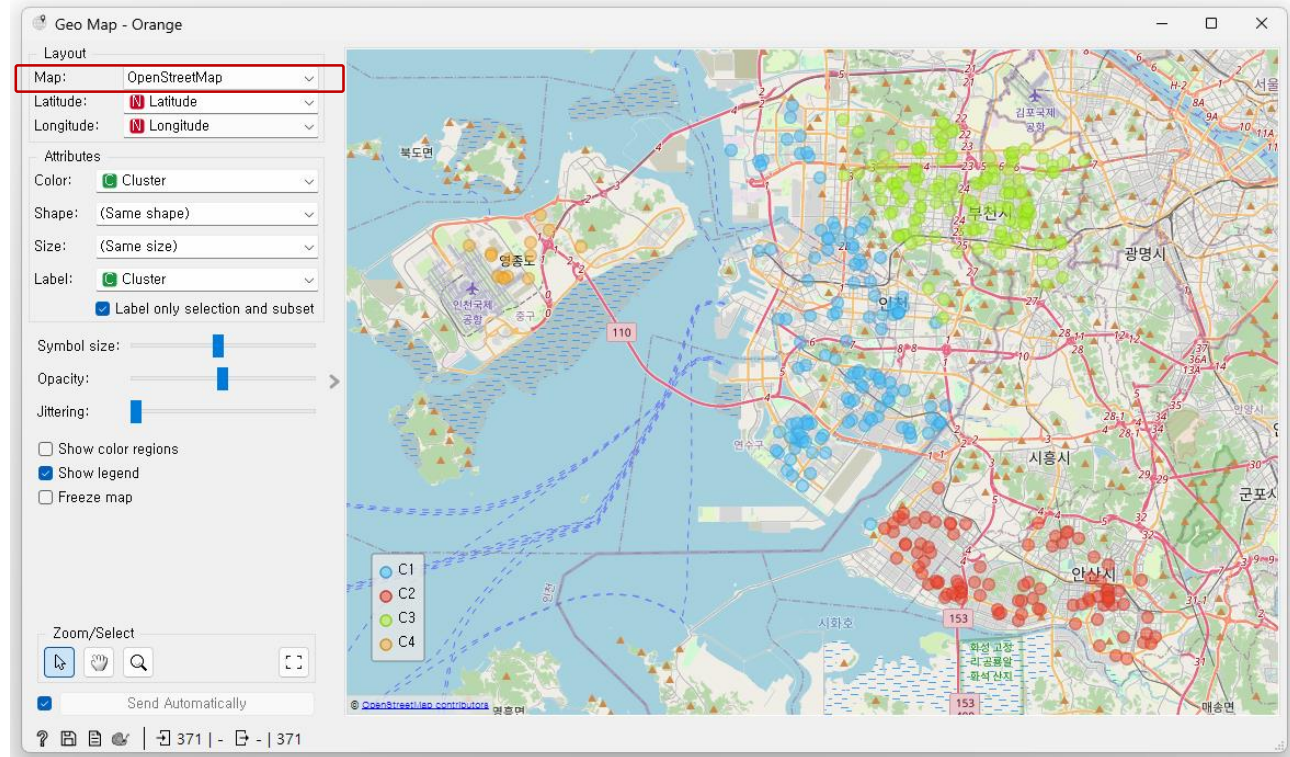


#### 5 Geo Map



## 5 Geo Map

- Map
  - OpenStreetMap
  - Black and white
  - Topographic
  - Satellite
- Print
- Dark



## Conclusion

- 인천광역시, 안산시 등의 행정 구역 기반의 택배 거점을 클러스터 기반으로 변경한다면, 비용을 절약할 뿐만 아니라 택배 집하장 선정 등을 최적화할 수 있을 것으로 기대함

An architectural rendering of a modern, multi-story building with a mix of brick and light-colored panels. The building features large windows and a flat roof. In the foreground, there is a paved plaza with several people walking and a few cars parked. The scene is set during sunset, with a warm orange glow. A large, semi-transparent circle with a dashed border is centered over the image, containing the title text. In the bottom left corner, there is a stylized graphic of two parallel diagonal lines.

# **2-Hour Group Project**



An architectural rendering of a modern, multi-story building with a mix of brick and light-colored panels. The scene is set at sunset, with a warm orange glow. A large, semi-transparent circle with a dashed border is centered over the building. Inside the circle, the text "Thank you" is written in a large, bold, black font. Below it, in a smaller black font, are the Korean words "광주과학기술원" (Gwangju Institute of Science and Technology) and "AI대학원" (AI Graduate School). The building has a prominent horizontal band of windows and a series of circular decorative elements on the ground floor. People are shown walking on the sidewalk and standing near the entrance, and several cars are parked on the right. The sky is a deep orange with some clouds.

# Thank you

광주과학기술원  
AI대학원