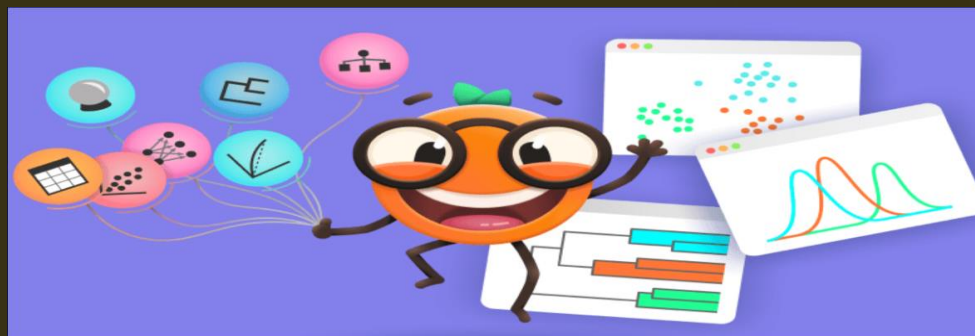


택배 배송 위치 군집화 k-Means



소프트웨어융합대학원
진혜진

■ 해결해야 할 문제

- 해를 거듭할수록 택배 수요는 폭발적으로 늘어나고 있고, 택배 비용 문제도 이슈가 되고 있다.
- 상황을 분석하기 위해 특정 지역의 택배 위치 데이터로 군집화하고, 패턴 등을 파악해 택배 배송 문제를 해결할 방법을 찾아보자.
- 인천 연안 지역 택배 위치 데이터로 군집화하여 택배 배송 문제를 해결할 인공지능 모델을 만들어보자.

■ 데이터 준비하기

■ 데이터

- Delivery.csv 파일 다운로드

- 371개의 인천 연안 지역 택배 배송 위치의 위도와 경도를 확인할 수 있다.

■ Delivery.csv

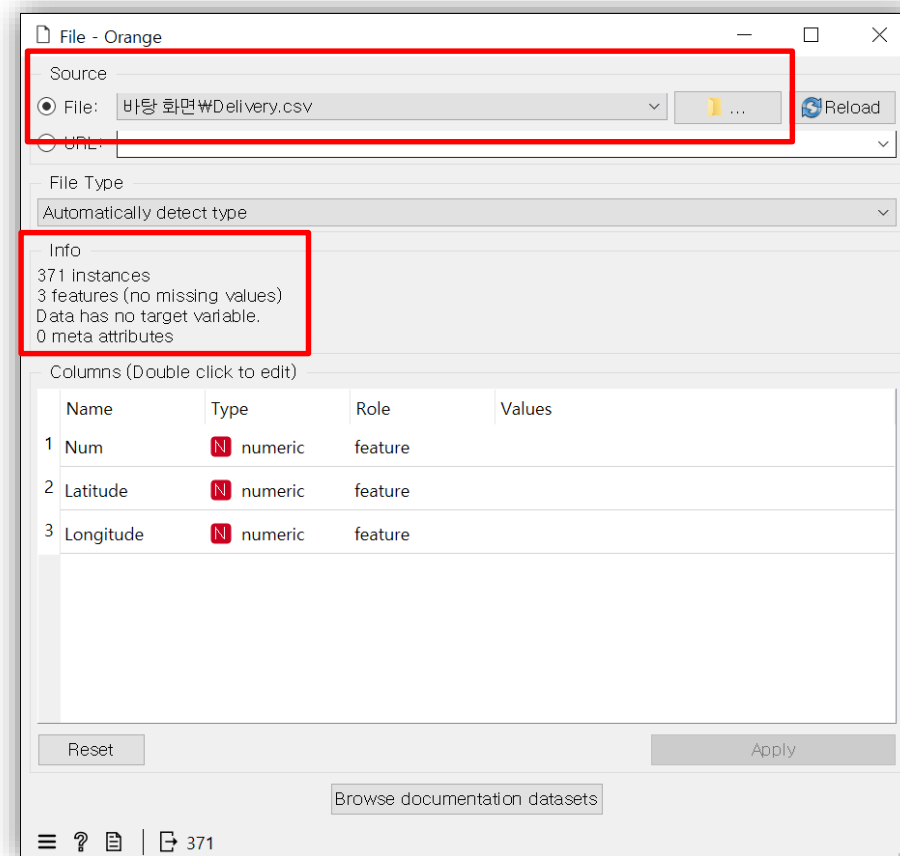
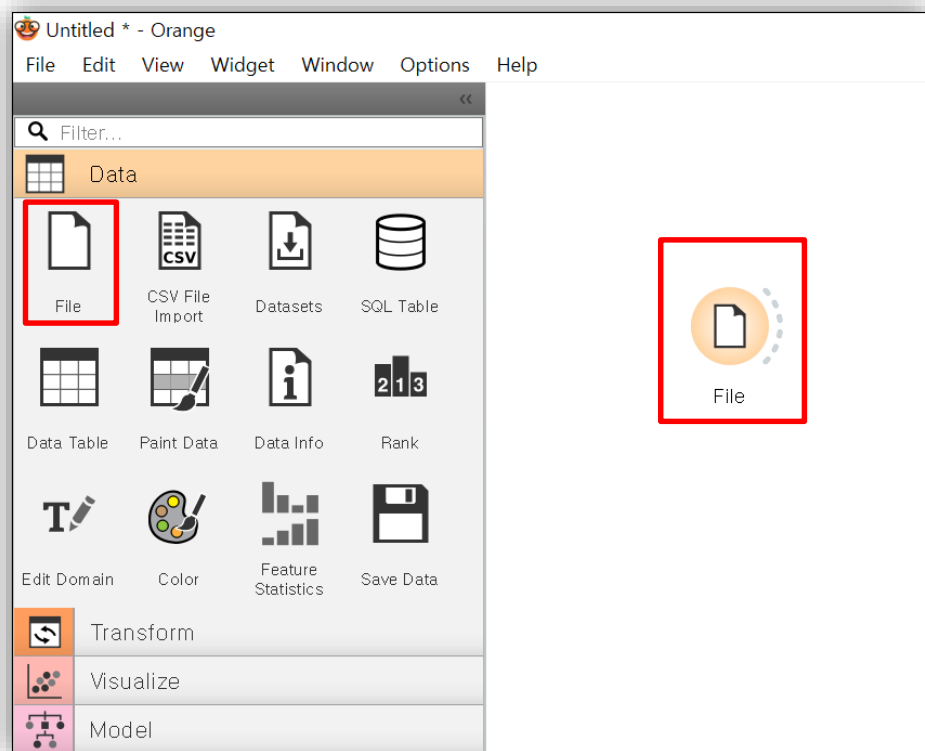
	A	B	C
1	Num	Latitude	Longitude
2	1	37.3368	126.7128
3	2	37.5013	126.7878
4	3	37.5225	126.7774
5	4	37.51118	126.7432
6	5	37.50878	126.7385
7	6	37.52849	126.7415
8	7	37.511	126.779
9	8	37.52944	126.7422
10	9	37.51636	126.7341
11	10	37.51334	126.7346
12	11	37.48605	126.8023
13	12	37.45911	126.7107
14	13	37.49984	126.7556
15	14	37.419	126.6966

358	357	37.52156	126.7184
359	358	37.53119	126.7222
360	359	37.49267	126.49
361	360	37.53246	126.6041
362	361	37.48232	126.6207
363	362	37.48406	126.6803
364	363	37.50697	126.7883
365	364	37.4856	126.6799
366	365	37.52679	126.6214
367	366	37.49304	126.6858
368	367	37.29901	126.833
369	368	37.52454	126.6223
370	369	37.49137	126.6781
371	370	37.52737	126.6235
372	371	37.45626	126.7052

■ 데이터 불러오기

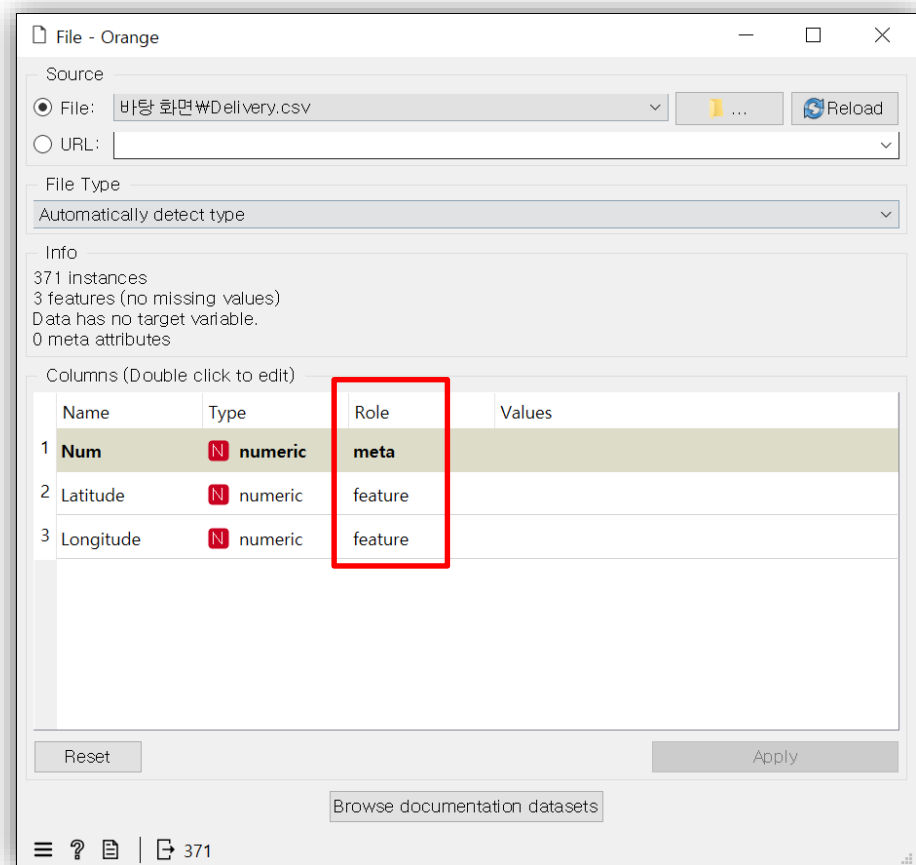
■ 택배 배송 위치 데이터 불러오기

- Data 카테고리에서 [File] 위젯을 캔버스로 가져온 후 더블 클릭하여 배송 위치 데이터인 Delivery.csv 파일을 불러온다.

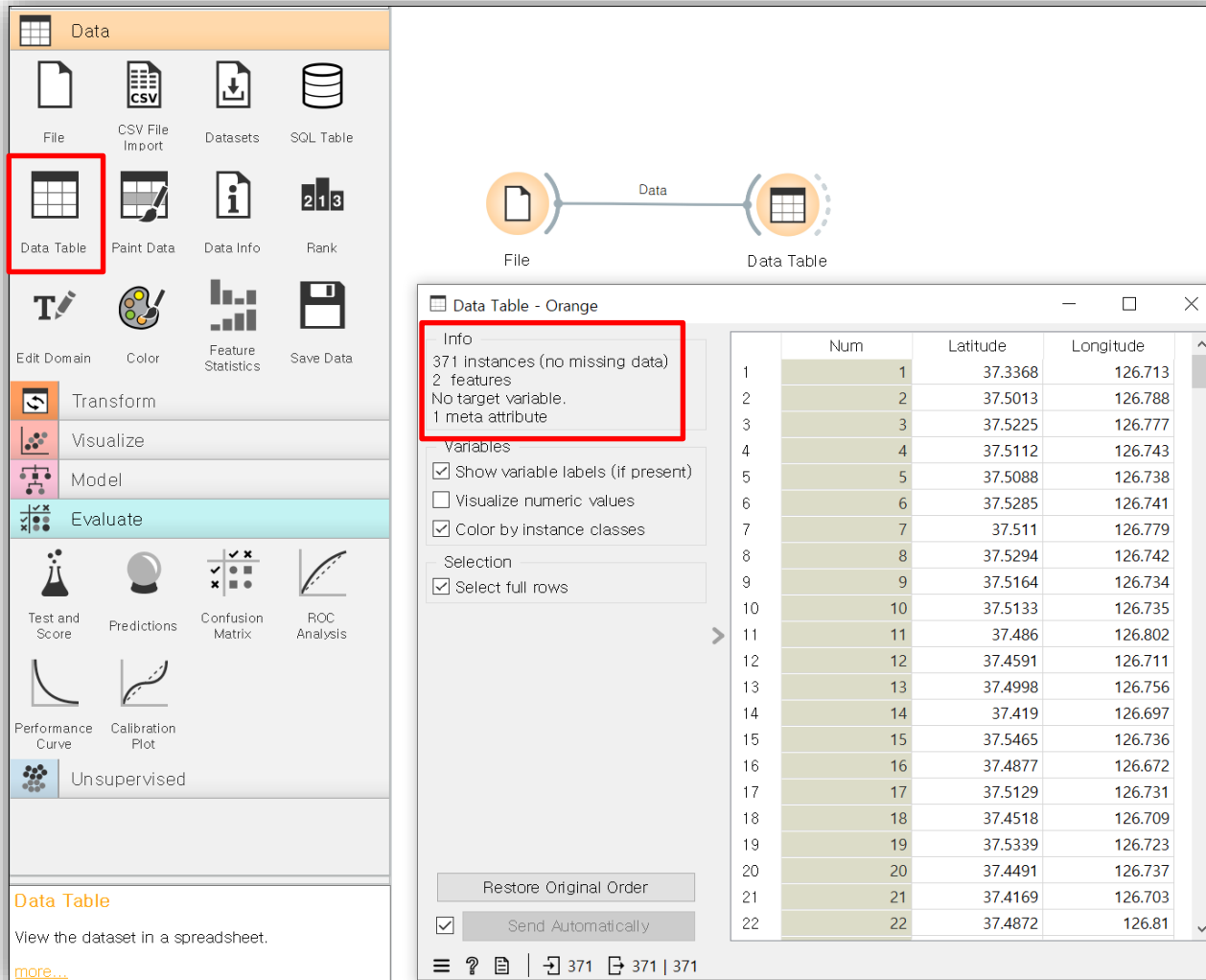


■ 속성 Role 변경하기

- 3개의 속성 중 일련번호(Num)은 참고만 하기 위해 meta로 설정하고, 나머지 속성은 feature로 설정한다.



- Data 카테고리에서 [Data Table] 위젯을 캔버스로 가져와서 [File] 위젯과 연결한 후 [Data Table] 위젯을 더블 클릭하여 데이터 정보를 살펴본다.



The screenshot shows the Orange3 software interface. On the left is the 'Data' widget palette, where the 'Data Table' widget is highlighted with a red box. The main canvas shows a workflow with a 'File' widget connected to a 'Data Table' widget. The 'Data Table' widget is double-clicked, opening its configuration window. In the 'Data Table - Orange' window, the 'Info' tab is selected, showing details about the dataset: 371 instances (no missing data), 2 features, no target variable, and 1 meta attribute. The 'Variables' section has checkboxes for 'Show variable labels (if present)', 'Visualize numeric values', and 'Color by instance classes'. The 'Selection' section has a checkbox for 'Select full rows'. The 'Data Table' widget is also visible in the bottom right corner of the main window, displaying a table of data.

	Num	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

■ 데이터 속성 정보 확인하기

- Num : 일련번호
- Latitude : 위도
- Longitude : 경도
- Latitude와 Longitude는 기계학습에 영향을 미치는 feature이다.

■ 군집화

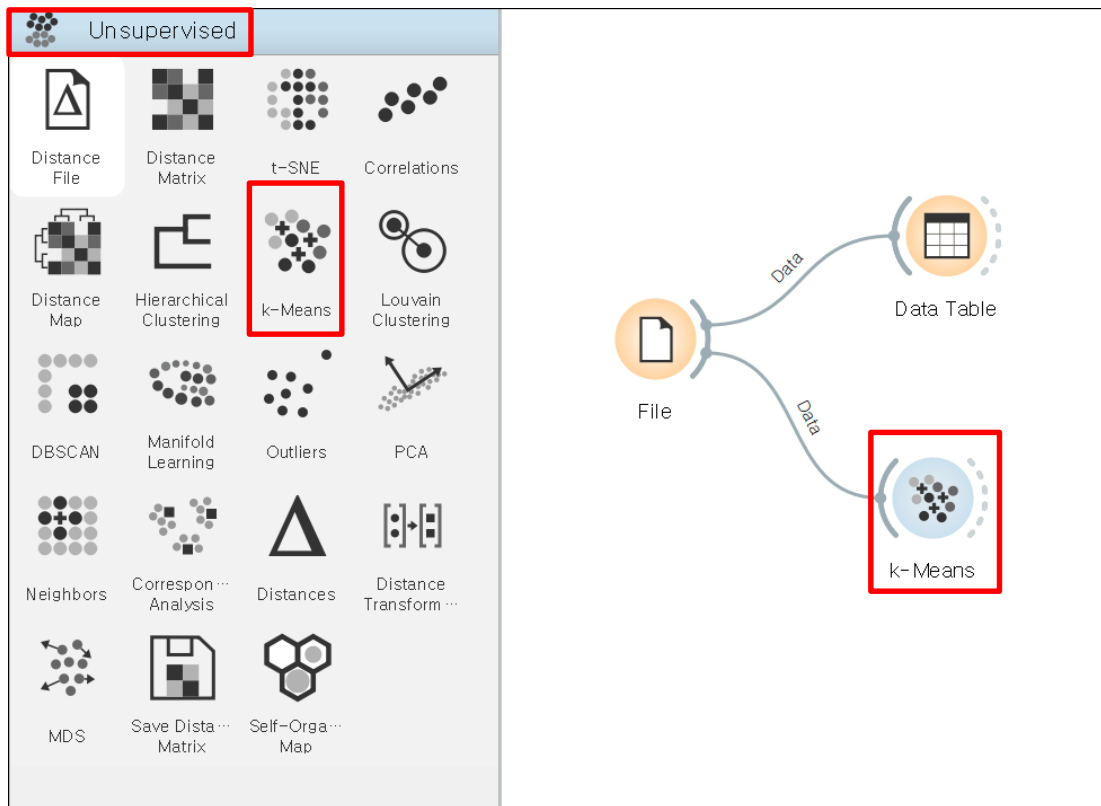
- 정답(레이블)이 없는 데이터로 학습을 하는 비지도 학습이다.
- 결과에 대한 사전 지식은 없지만, 해당 데이터를 통해 의미가 있는 결과를 얻고자 할때 사용하는 기계학습이다.

■ 학습 모델 선택하기

- 모델의 성능을 분석하고 비교하기 위해 k-Means로 군집화한다.

- 모델 선택하기

- Unsupervised(비지도 학습) 카테고리의 모델 중 [k-Means] 위젯을 캔버스로 가져와서 [File]위젯과 연결한다.

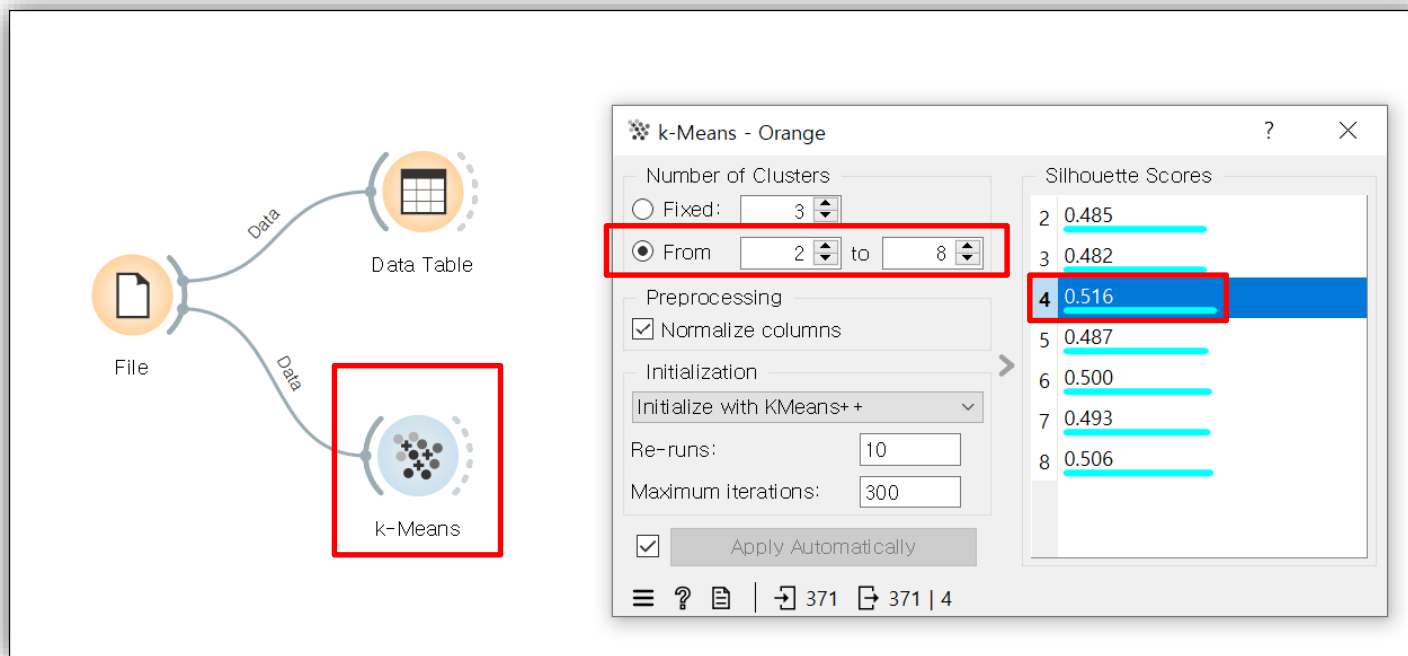


■ [k-Means] 위젯

- 위젯을 더블 클릭 한 후 대화 창이 나타나면 설정 변경을 통해 가장 적합한 군집 개수를 파악하여 군집화한다.

- Number of Clusters(군집 개수)

- » Fixed : 내가 원하는 군집의 개수를 설정한다.
- » From : 실루엣 점수(Silhouette Scores)를 보여 주는 범위를 설정한다.
- » Silhouette Scores : 해당 범위 내에서 가장 높은 점수의 군집 개수를 추천한다.



The image shows an Orange Data Mining workflow on the left and the k-Means widget settings on the right.

Workflow: A 'File' widget is connected to two 'Data' widgets. One 'Data' widget is connected to a 'Data Table' widget, and the other is connected to a 'k-Means' widget. The 'k-Means' widget is highlighted with a red box.

k-Means - Orange Settings:

- Number of Clusters:** ☒ From 2 to 8 (highlighted with a red box)
- Preprocessing:** ☒ Normalize columns
- Initialization:** Initialize with KMeans++
- Re-runs:** 10
- Maximum iterations:** 300
- ☒ Apply Automatically

Silhouette Scores:

Number of Clusters	Silhouette Score
2	0.485
3	0.482
4	0.516
5	0.487
6	0.500
7	0.493
8	0.506

The row for 4 clusters is highlighted with a red box, indicating it is the recommended number of clusters.

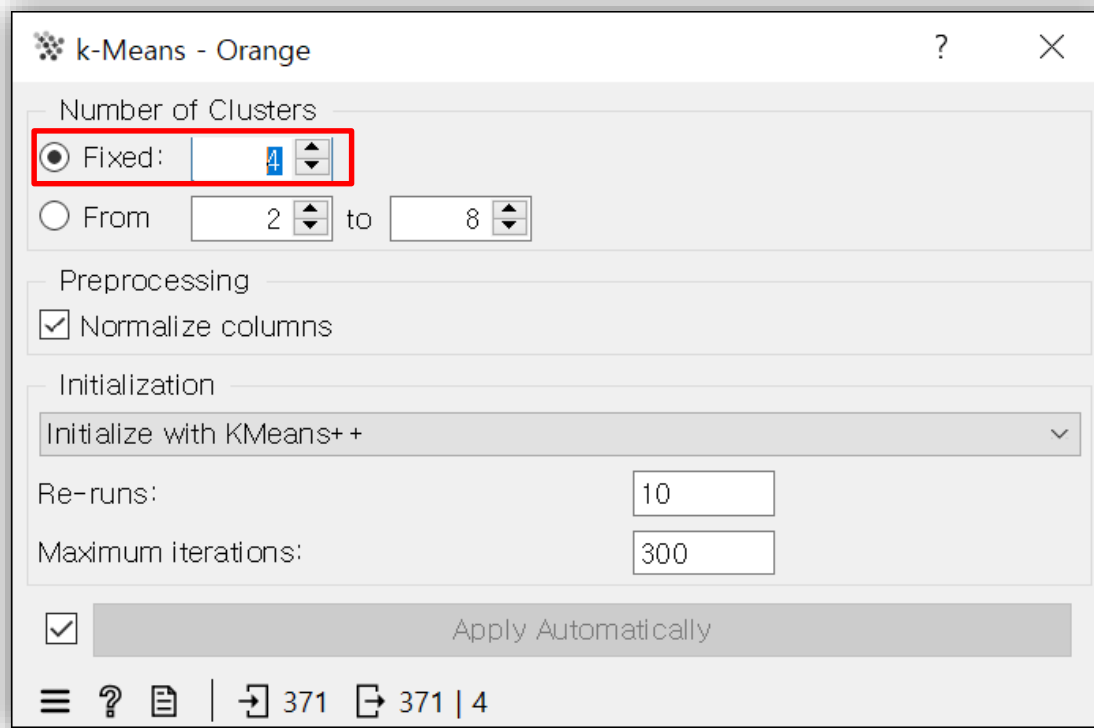
■ 모델 학습시키기

■ 군집 설정하기

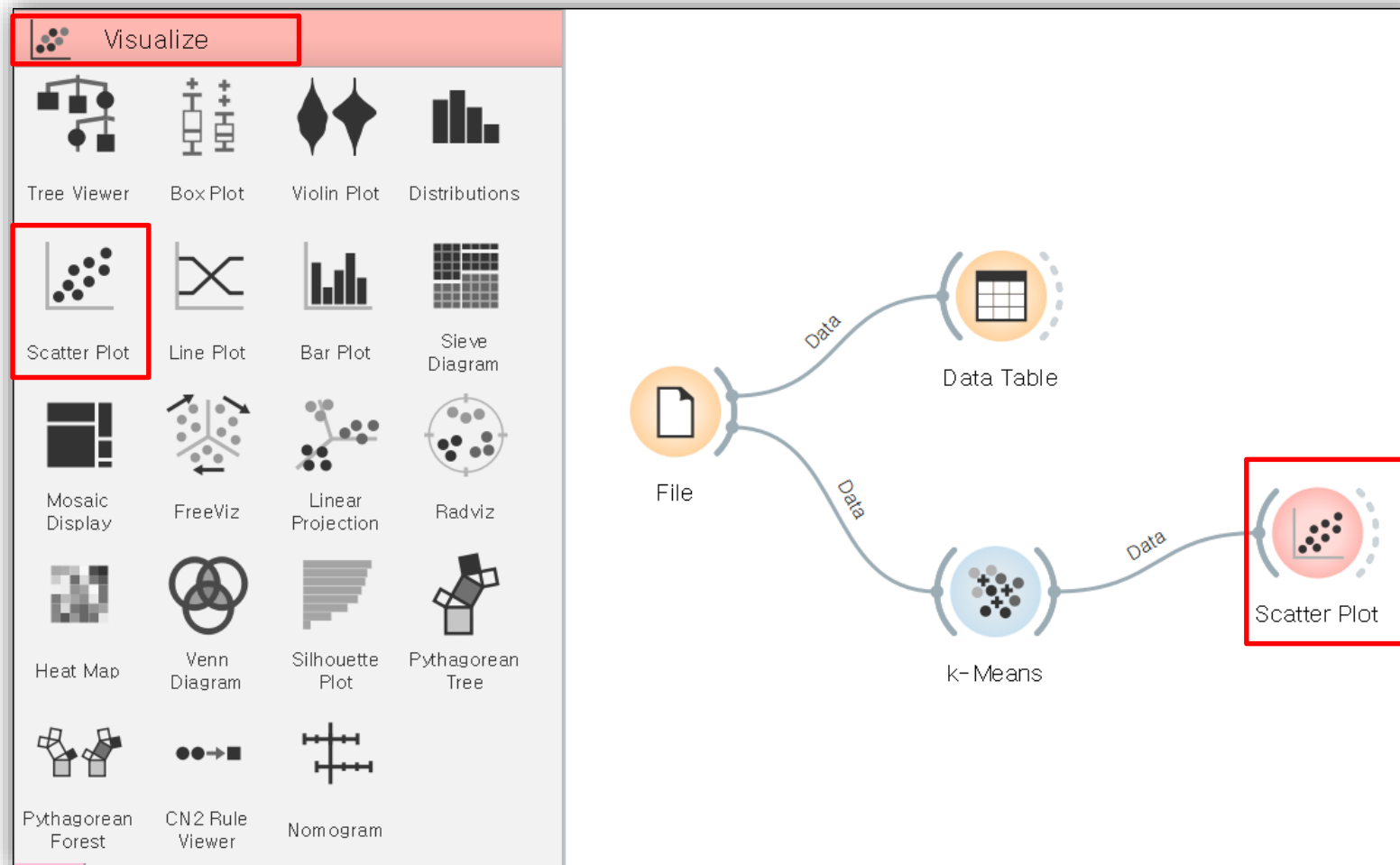
- 가장 높은 Silhouette Scores가 0.516이므로 k(Fixed)를 4로 설정한다.
- Fixed값으로 군집 개수를 설정하면 출력으로 설정한 개수만큼 군집이 생성된다.

■ 출력 확인하기

- 군집이 어떻게 이루어졌는지 확인하기 위해 Visualize 카테고리에서 [Scatter Plot] 위젯을 캔버스로 가져와서 [k-Means] 위젯과 연결한다.



- [Scatter Plot] 위젯 더블 클릭하여 시각화 결과를 확인해보자.



- **모델 성능 확인하기**
- **학습 결과 확인하기**
 - 군집화 그래프 확인하기

군집별로 색을 다르게 설정하면
군집이 어떻게 형성되었는지 더
쉽게 파악할 수 있다.



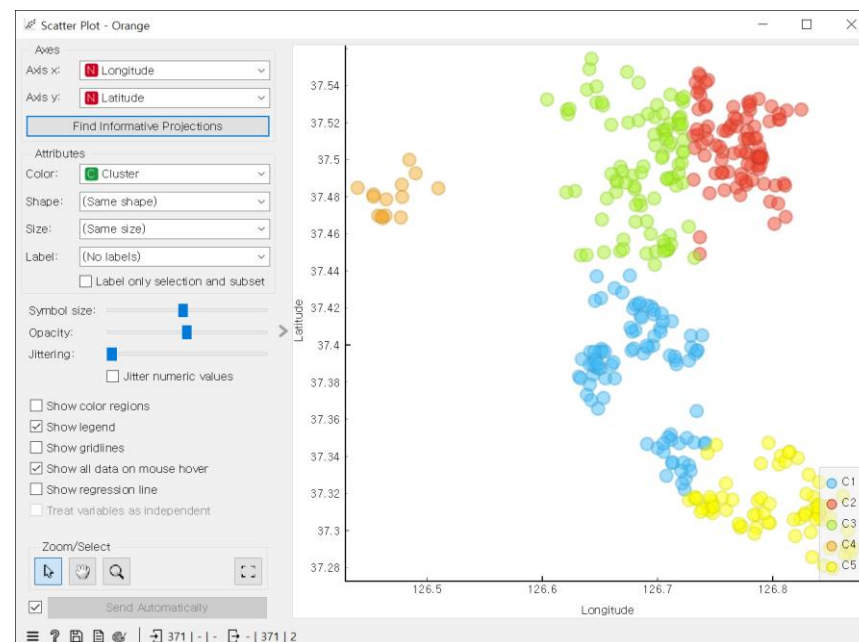
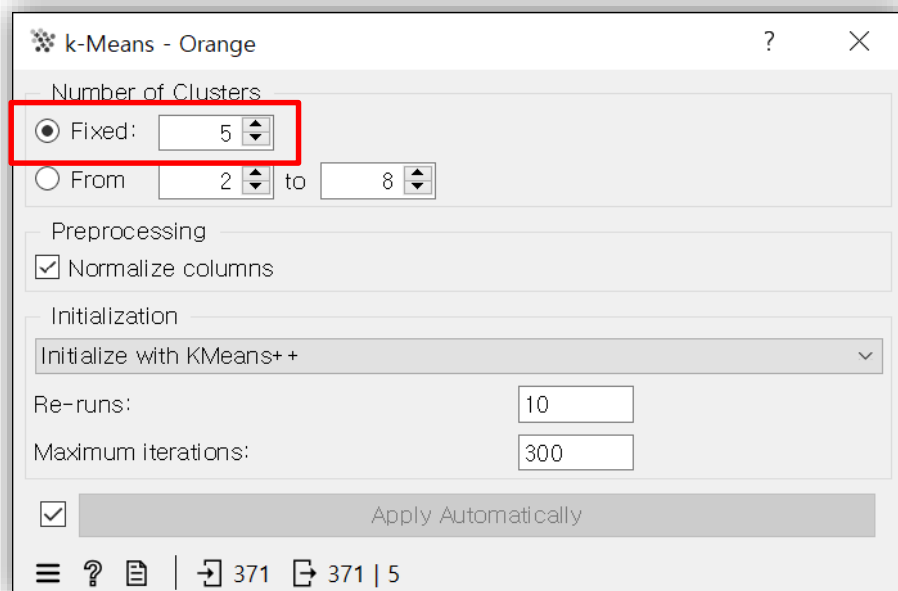
▪ x축, y축 설정 및 색 지정하기

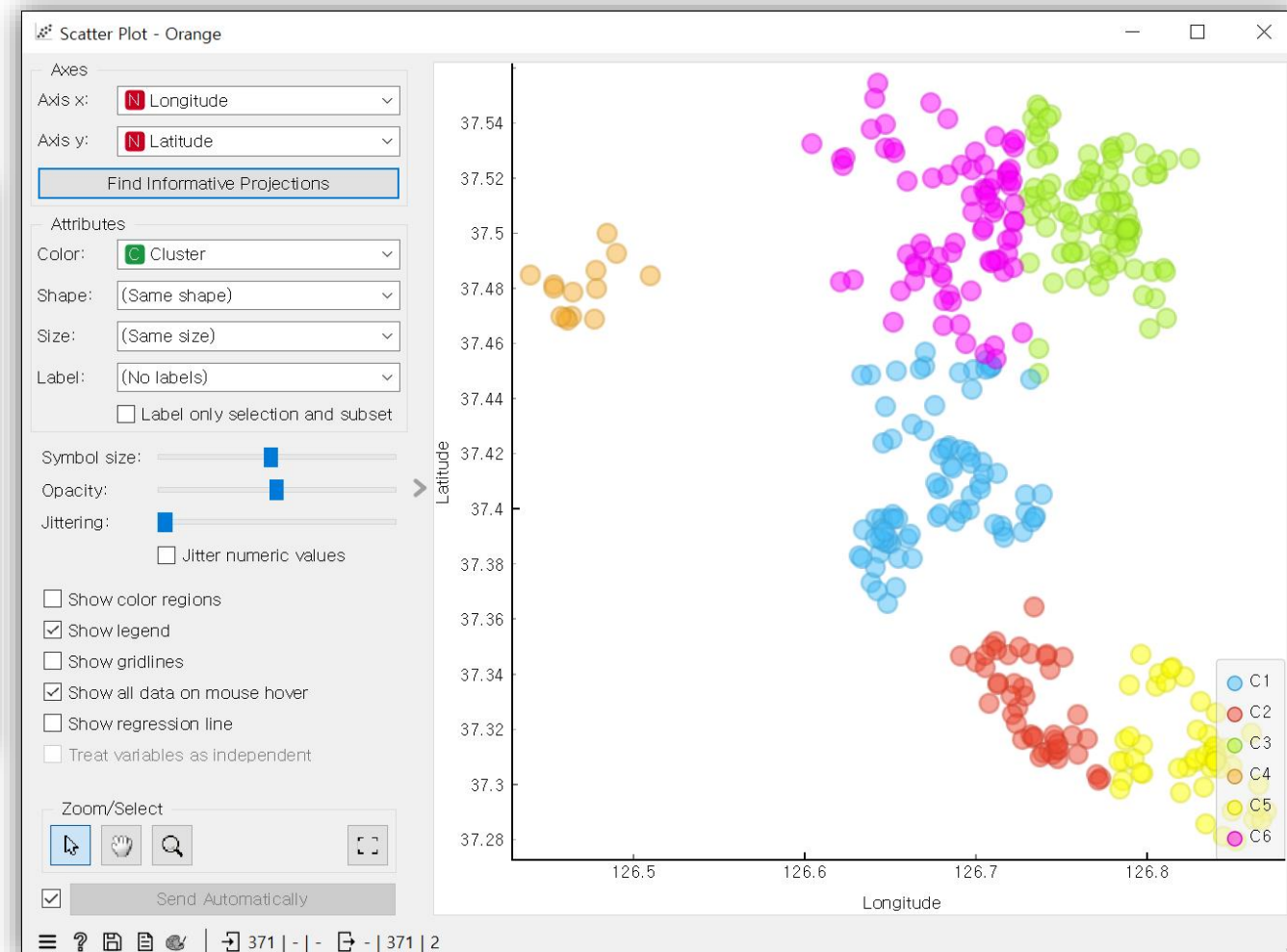
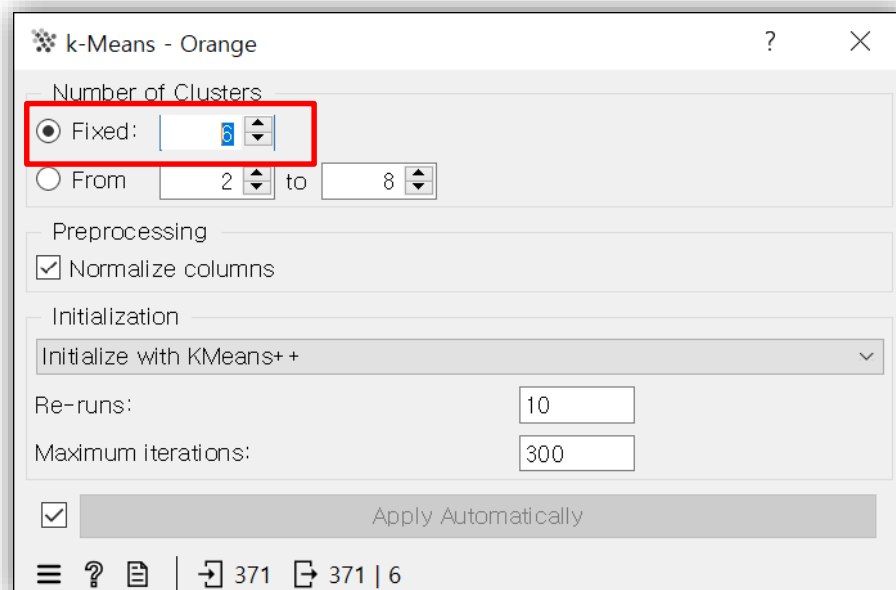
- x축을 경도, y축을 위도로 설정하고, Color는 군집(Cluster)로 설정하면
- 색이 서로 다른 4개의 군집이 잘 형성된 것을 확인할 수 있다.

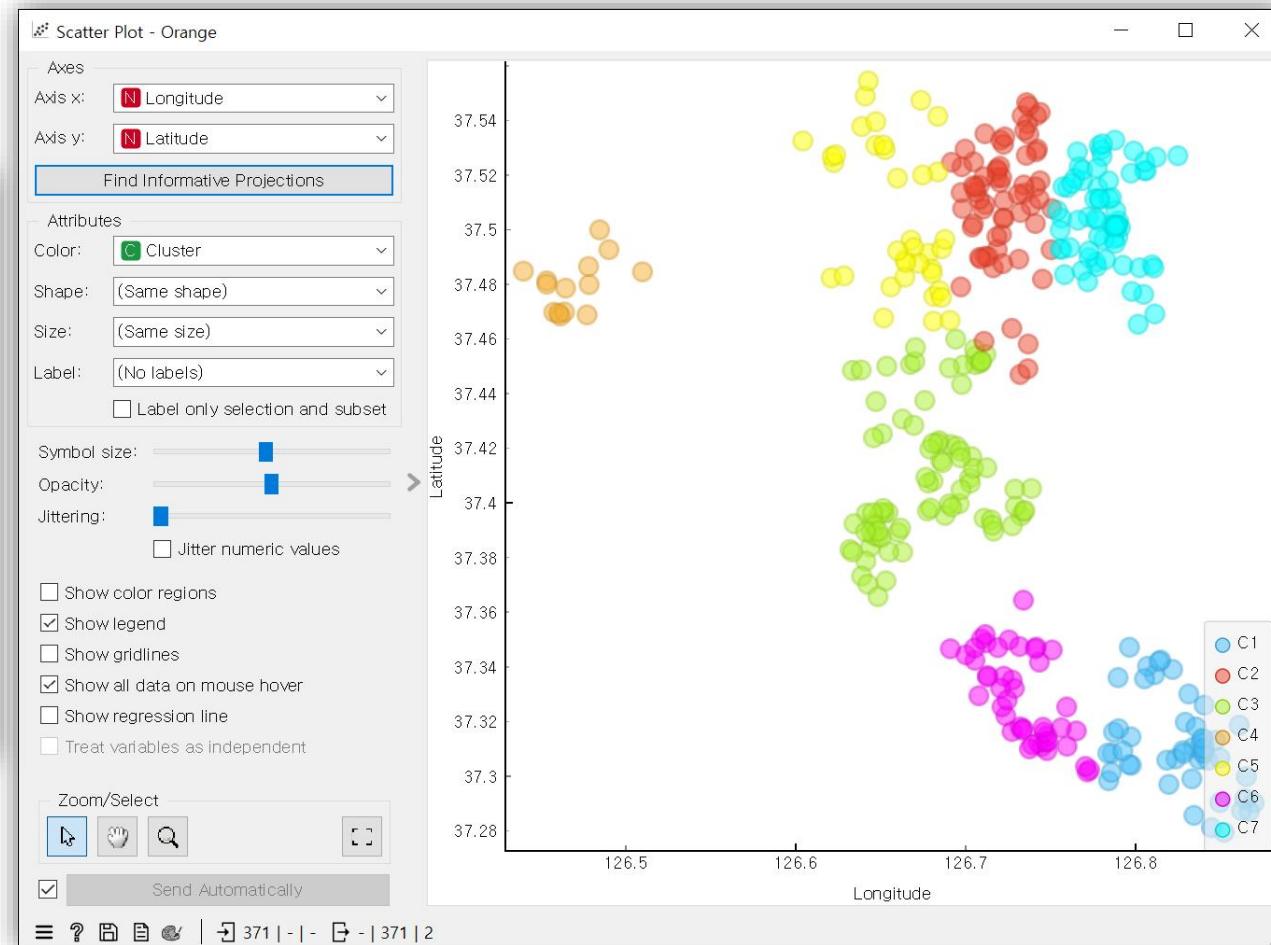
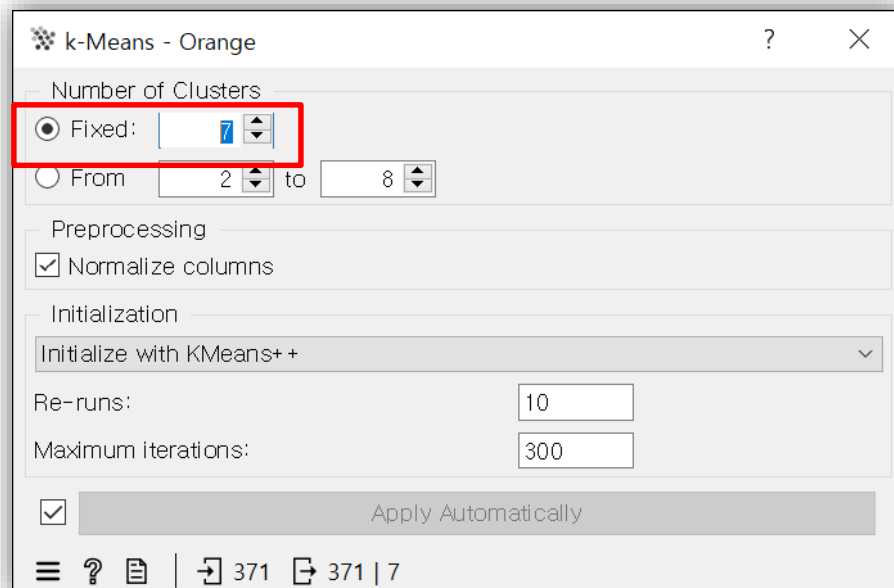


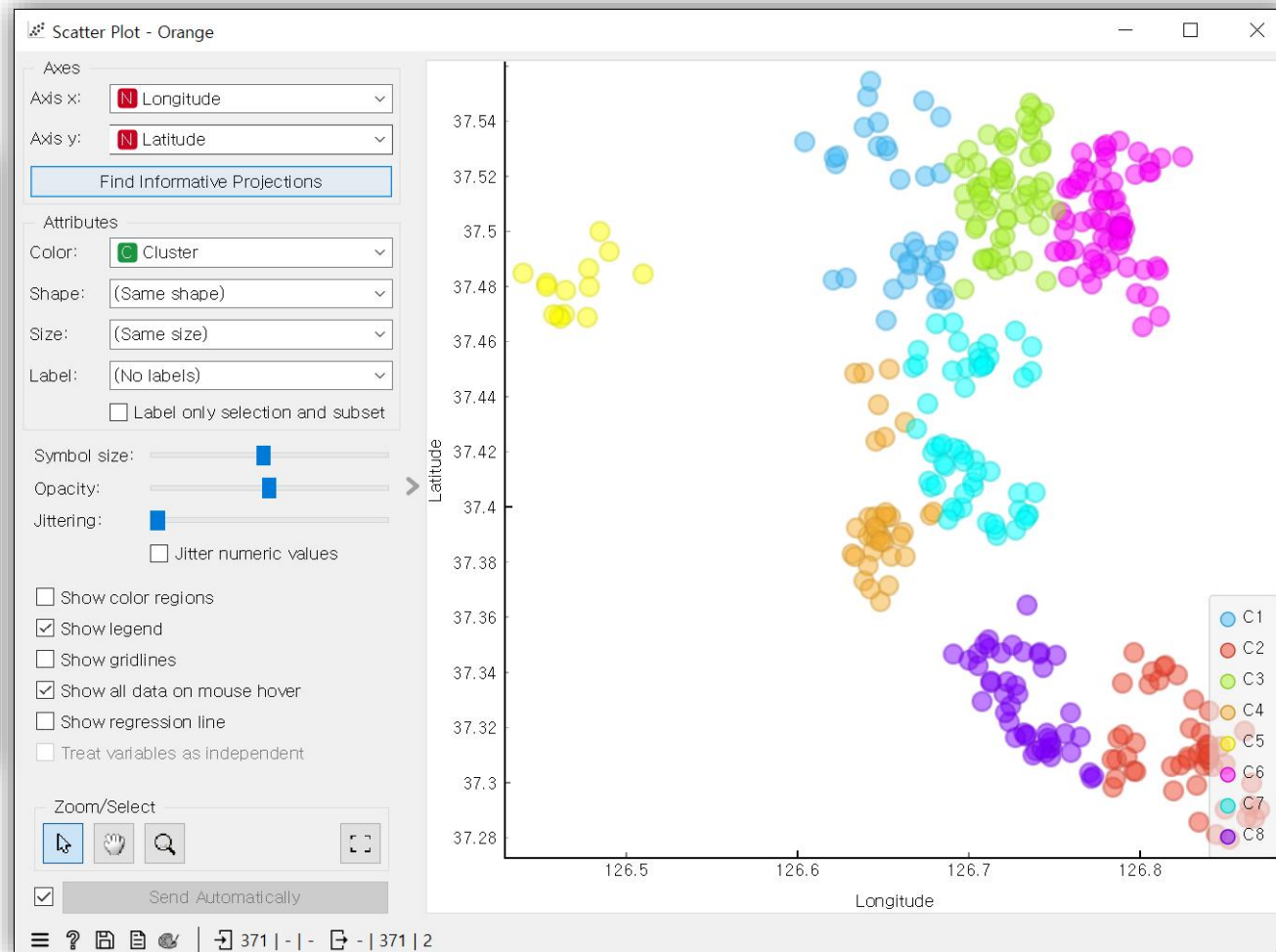
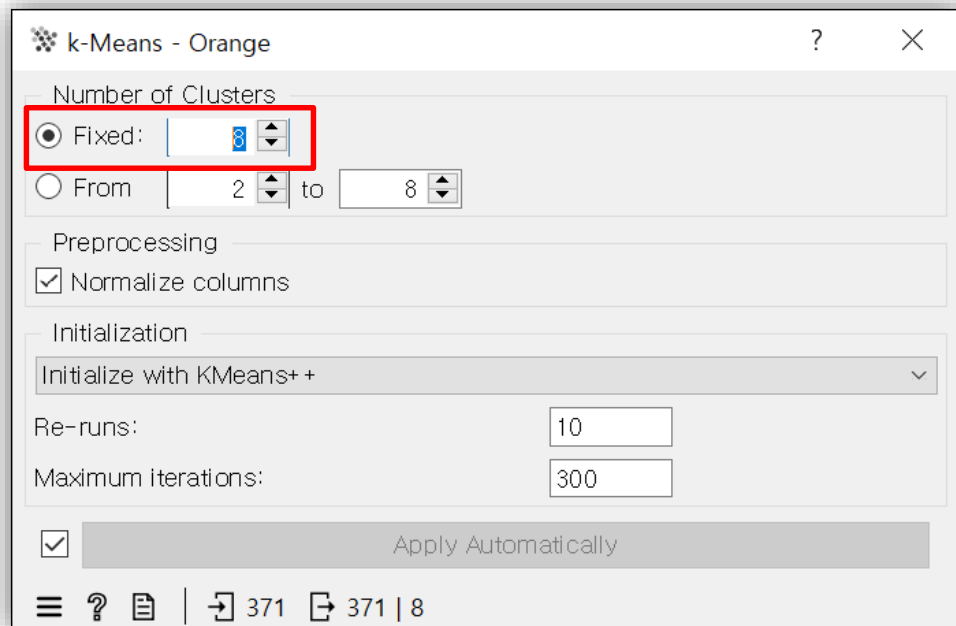
■ 성능 결과 확인하기

- 산점도로 군집이 어떻게 형성되었는지 알 수 있지만, 군집 개수를 늘려보거나 인천 연안 지도를 넣는 방법으로 군집화 상황을 더 자세히 알 수 있다.
- 군집 점 늘리기
 - Fixed의 값을 5, 6, 7, 8로 바꾸면서 택배 거점에 따른 군집 점을 형성해본다.

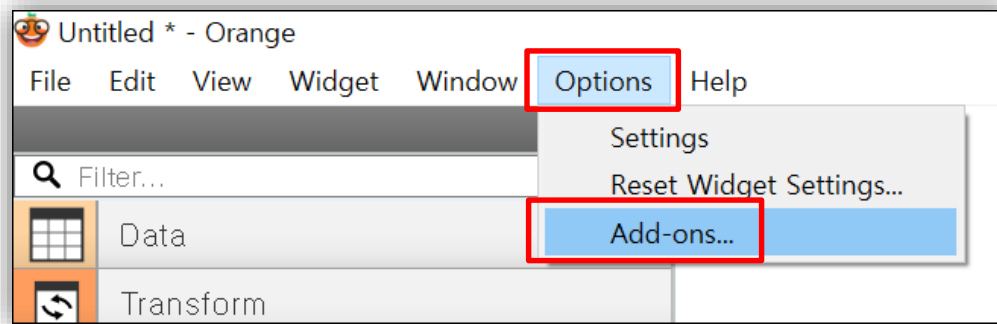




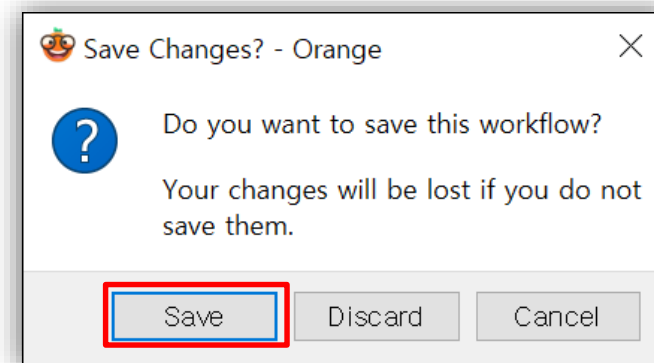
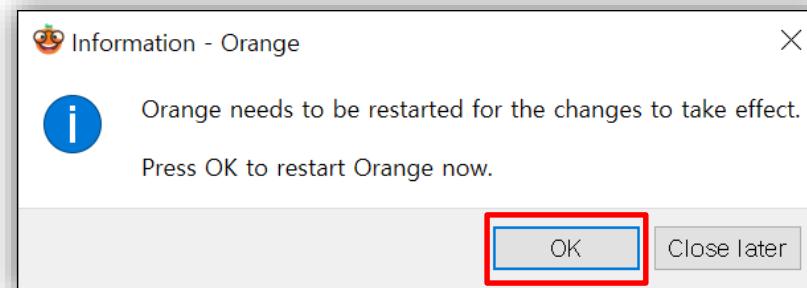
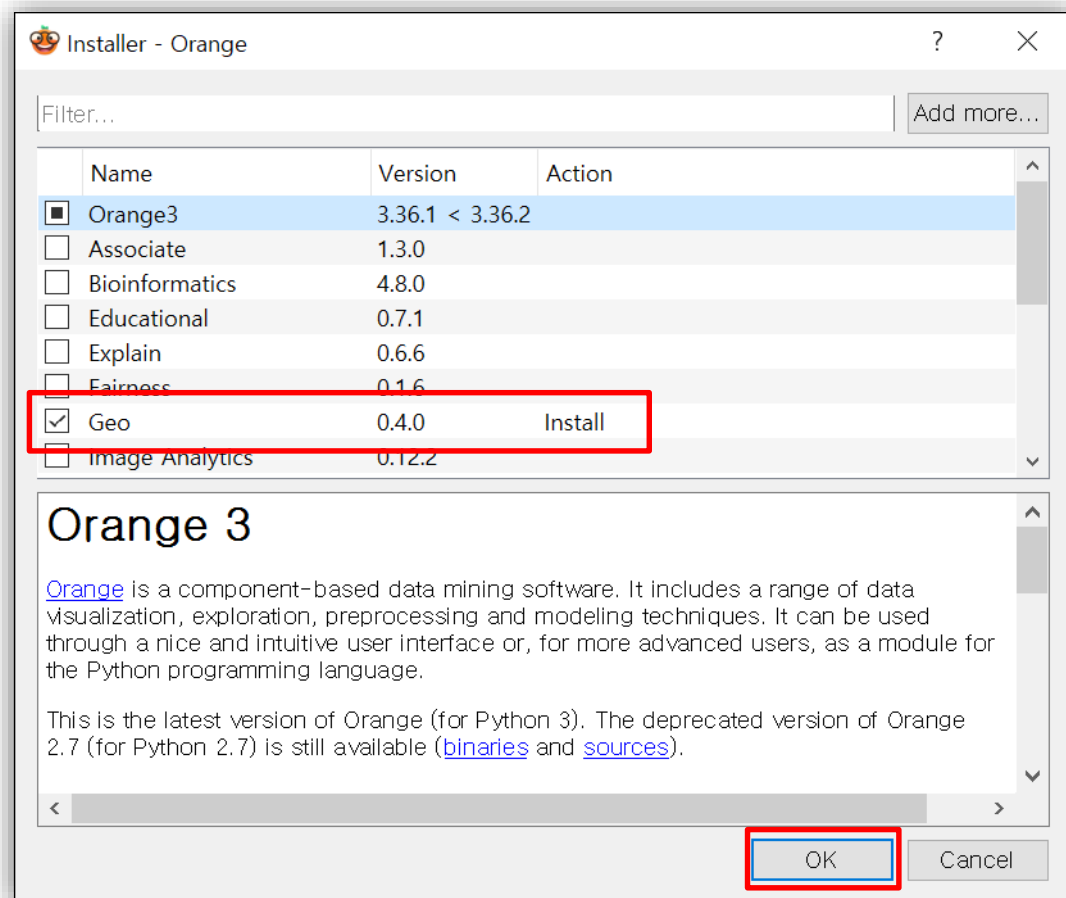




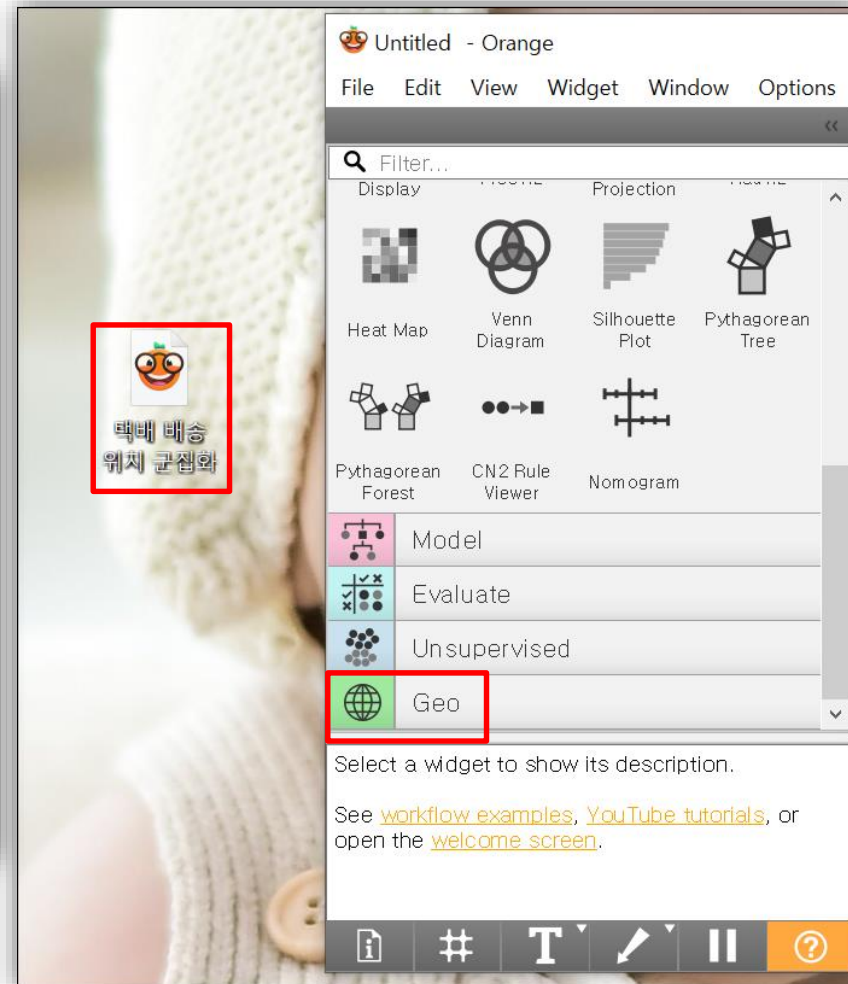
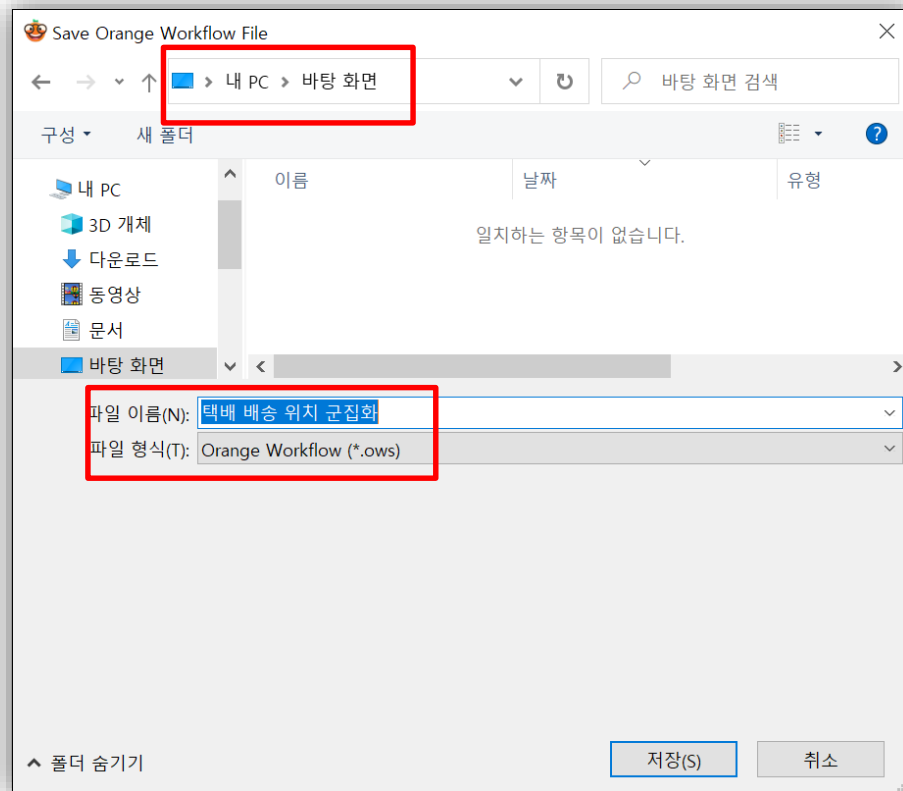
- 일반적으로 군집 점은 보통 4개 이상 존재하지만 군집점이 5~8개로 늘어날수록 더 작은 군집이 형성되면서 세분화되는 것을 확인할 수 있다.
- 인천 연안 지도로 보기
 - 인천 연안 지도를 삽입하여 어떤 지역이 군집화되었는지 살펴보자.
 - Options 메뉴에서 [Add-ons...]를 클릭하면 Installer 창이 나타난다.



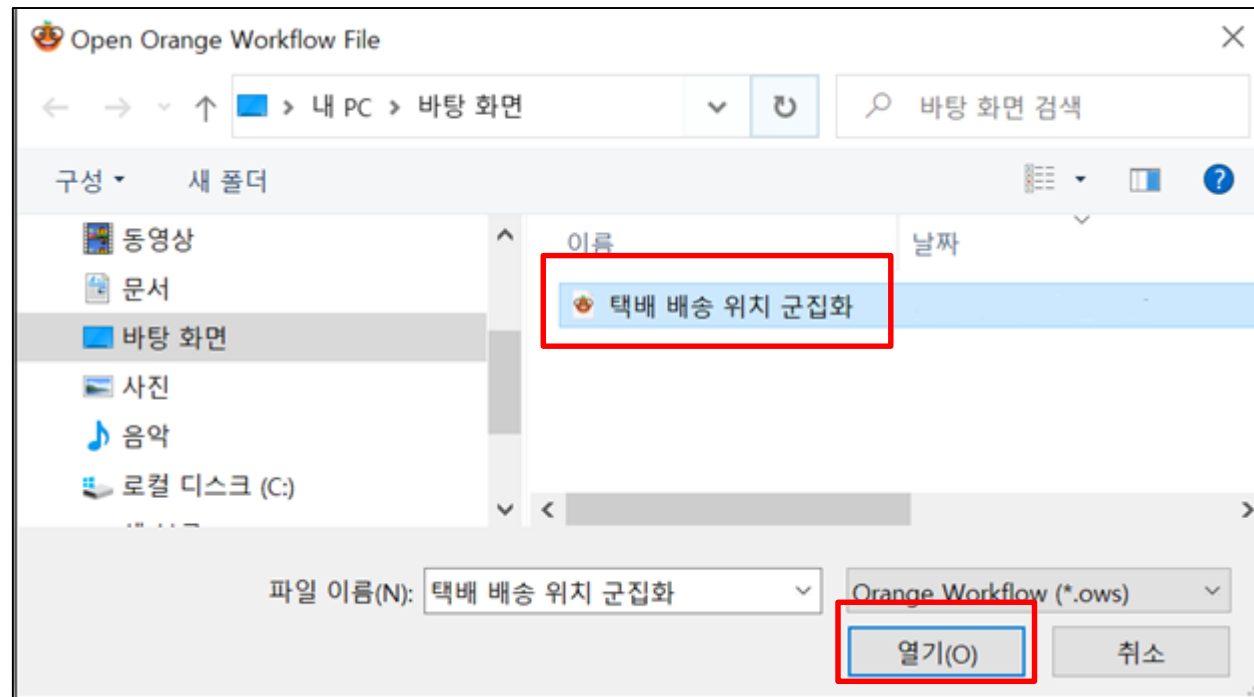
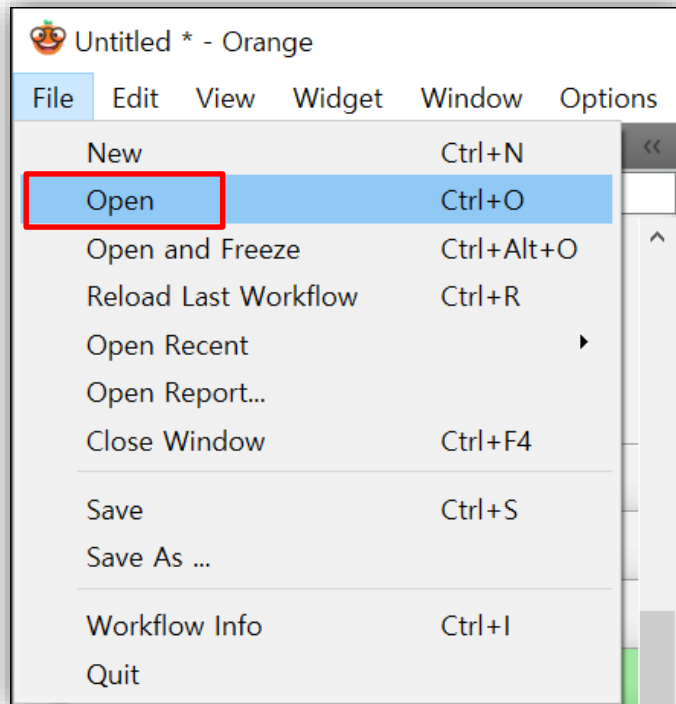
- Geo에 체크하고 OK를 클릭하면 Geo 카테고리가 추가로 설치된다.



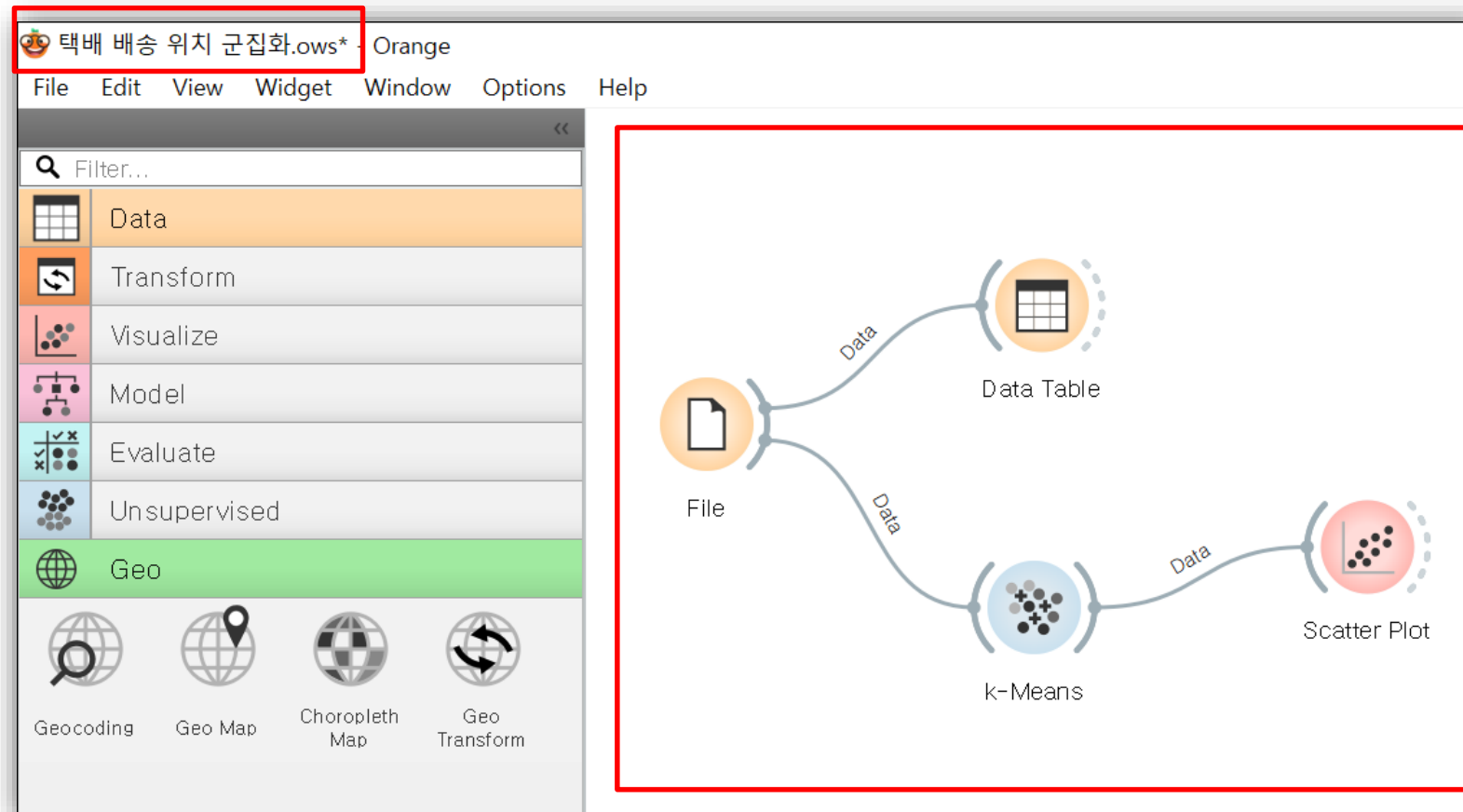
■ 파일 저장



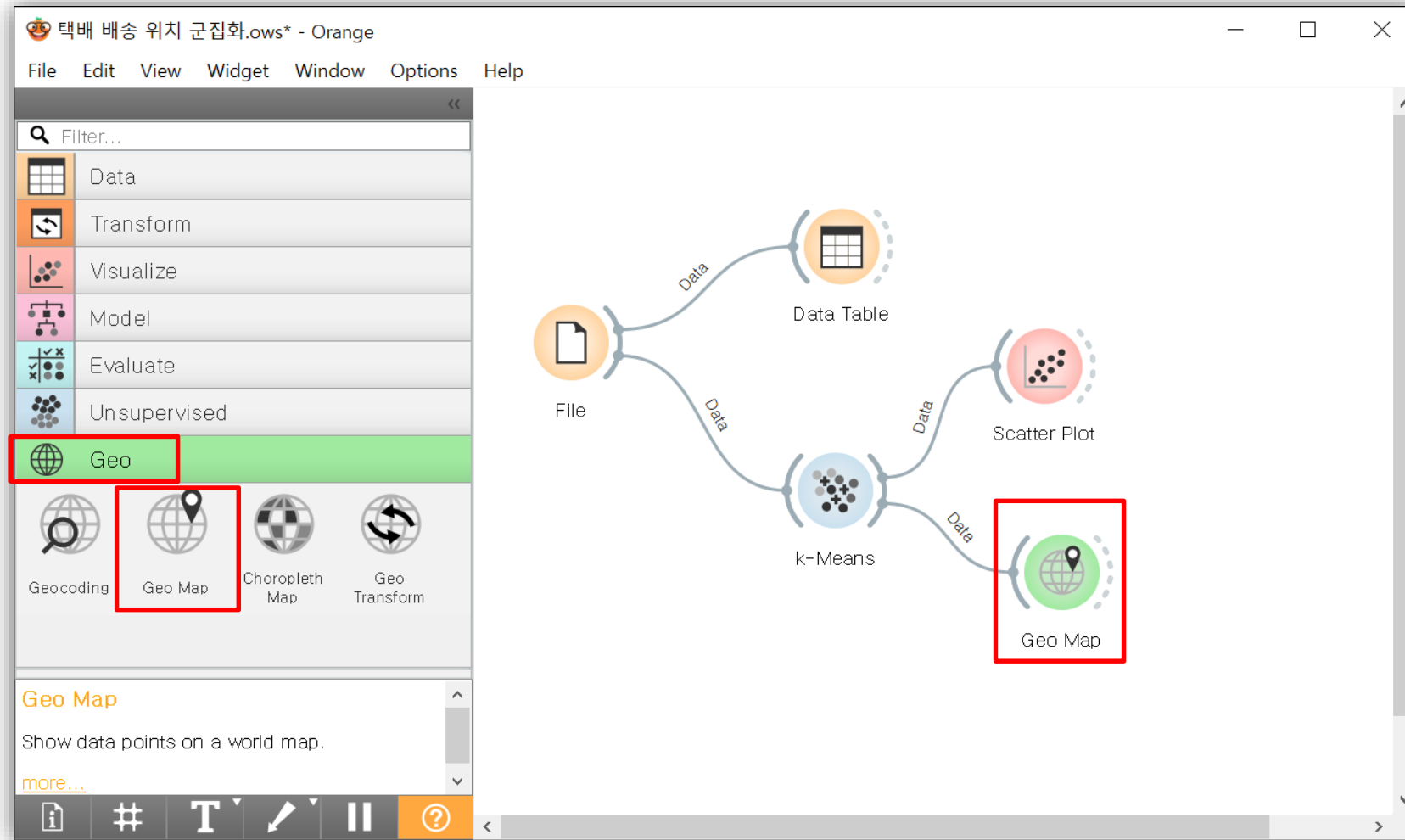
■ 저장된 Orange 파일 열기



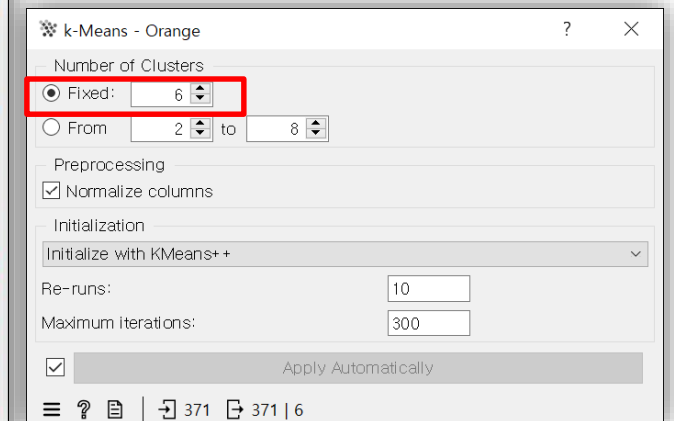
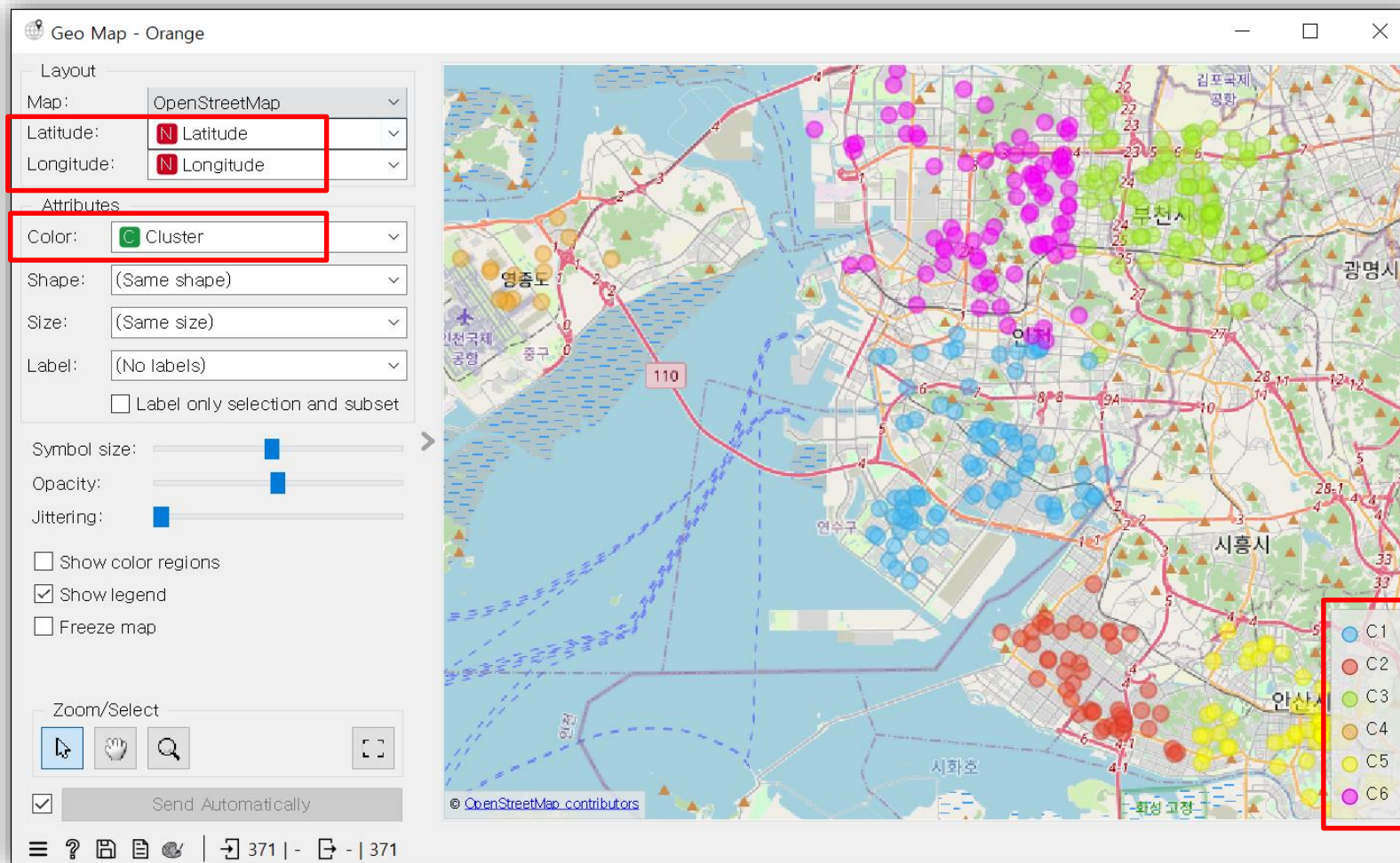
■ 저장된 Orange 파일 열기



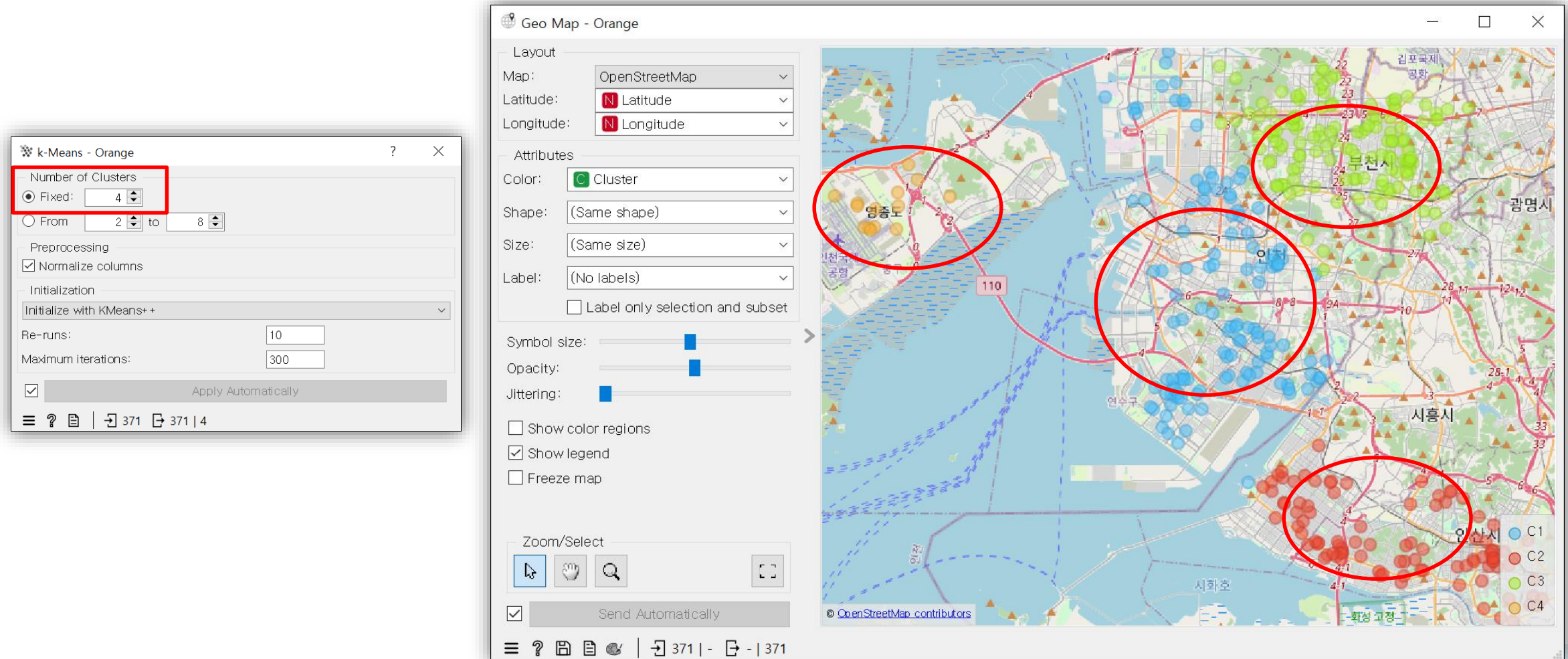
- [k-Means] 위젯에 Geo 카테고리의 [Geo Map] 위젯을 새롭게 연결하고 위도와 경도를 설정하면 Orange3 내에서도 군집화한 곳에 지도를 동시에 나타낼 수 있다.



- 색상은 군집(Cluster)로 설정하여 군집화된 곳이 어느 지역인지 쉽게 확인하도록 한다.
- 군집 점이 6일 때 지역을 여섯 군데로 군집화한 모습이다.

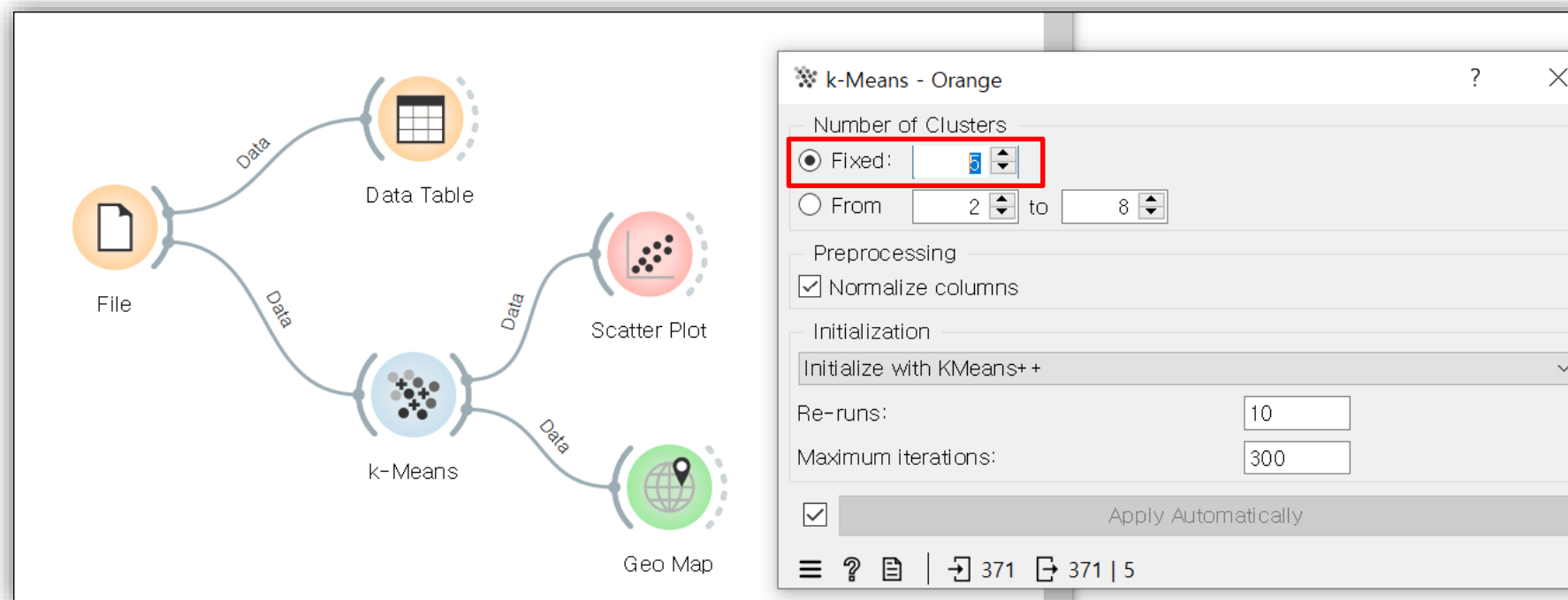


- [Geo Map] 위젯을 이용하면 k-Means 군집화와 인천 연안 지역 지도를 합한 지역별 군집화한 모습을 확인할 수 있다.
- 군집화가 영종도, 인천광역시, 부천시, 안산시로 네 군데 형성되었음을 볼 수 있다.



■ 군집 살펴보기

- [k-Means] 위젯을 더블 클릭하여 Fixed 값을 5, 6, 7, 8로 바꾸어 가면서 인천 연안 지역 지도와 택배 위치 데이터를 활용하여 군집 상황을 살펴보자.

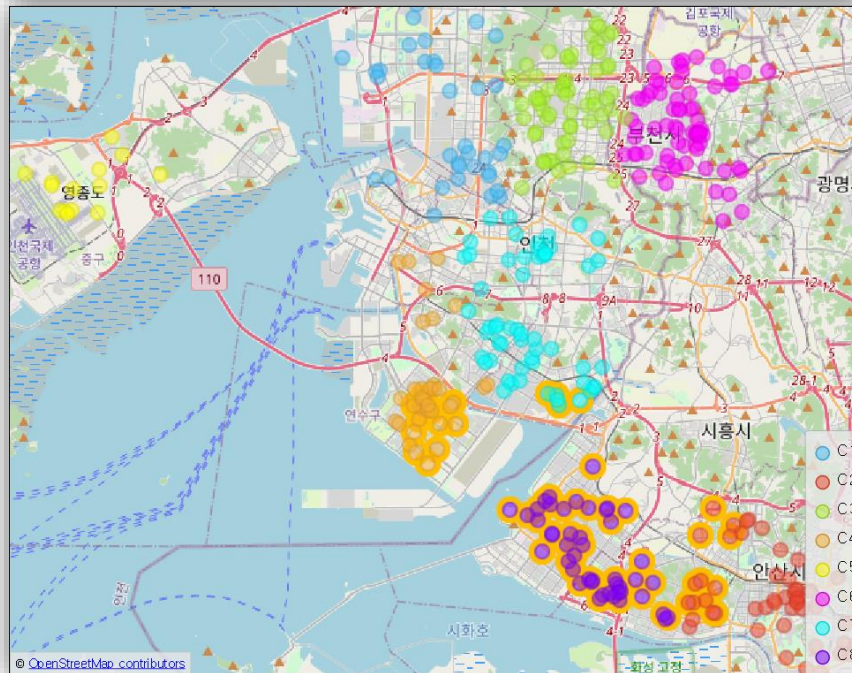
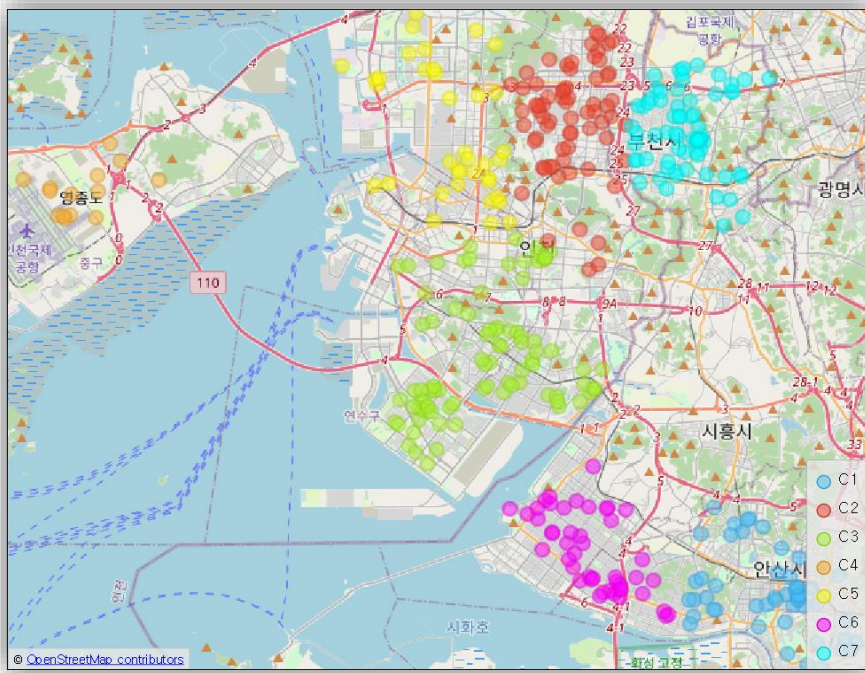
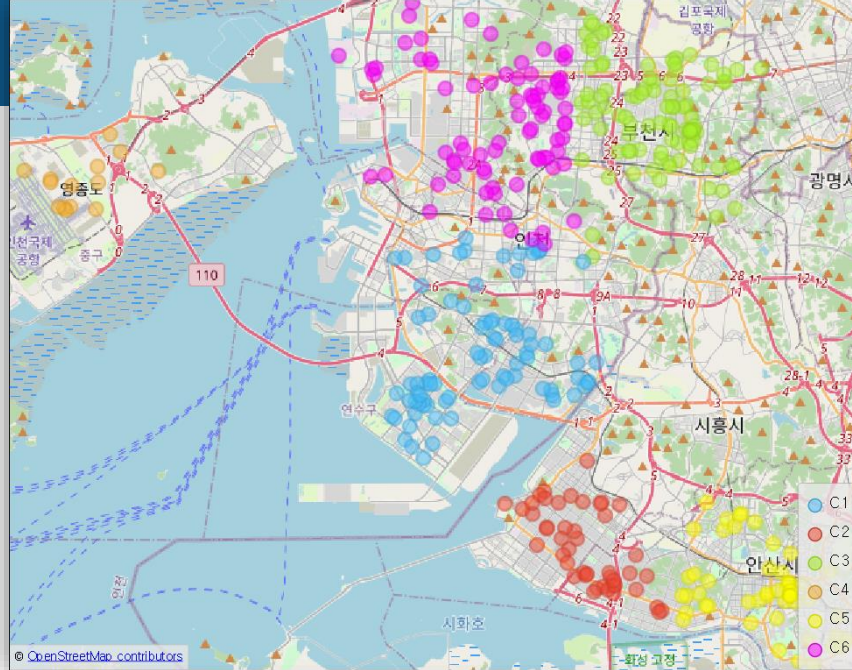
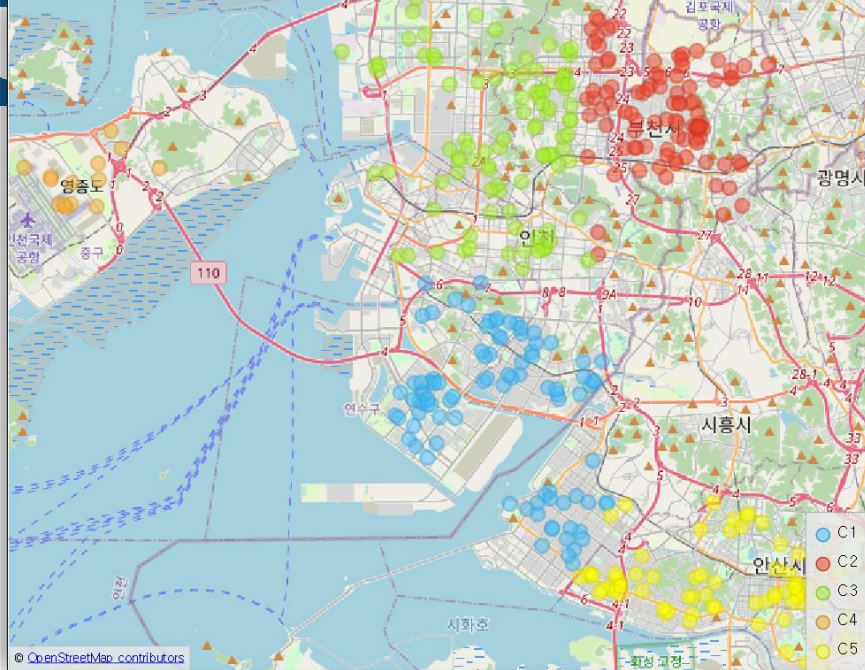


The image displays an Orange Data Mining workflow and the configuration for the k-Means widget.

Workflow: A 'File' widget connects to a 'Data Table' widget via a 'Data' link. The 'Data Table' widget connects to a 'k-Means' widget via a 'Data' link. The 'k-Means' widget connects to a 'Scatter Plot' widget via a 'Data' link. The 'k-Means' widget also connects to a 'Geo Map' widget via a 'Data' link.

k-Means - Orange Widget Settings:

- Number of Clusters:** ☒ Fixed: 5 (highlighted with a red box). ☐ From 2 to 8.
- Preprocessing:** ☒ Normalize columns.
- Initialization:** Initialize with KMeans++.
- Re-runs:** 10.
- Maximum iterations:** 300.
- ☒ Apply Automatically.



- 군집 상황을 분석하여 택배 집하장 선정 등에 적용할 수 있다.
- 인천광역시, 안산시 등의 행정 구역 기반의 택배 거점을 군집 점을 기반으로 변경한다면 비용을 절약할 뿐만 아니라 택배 집하장 선정 등을 최적화 할 수 있다.
- 군집화는 데이터에서 내가 알지 못했던 것을 발견하는 방법으로 사람이 파악하기 힘든 본질적인 문제나 숨겨진 특징 및 구조를 연구할 때 주로 사용한다.