

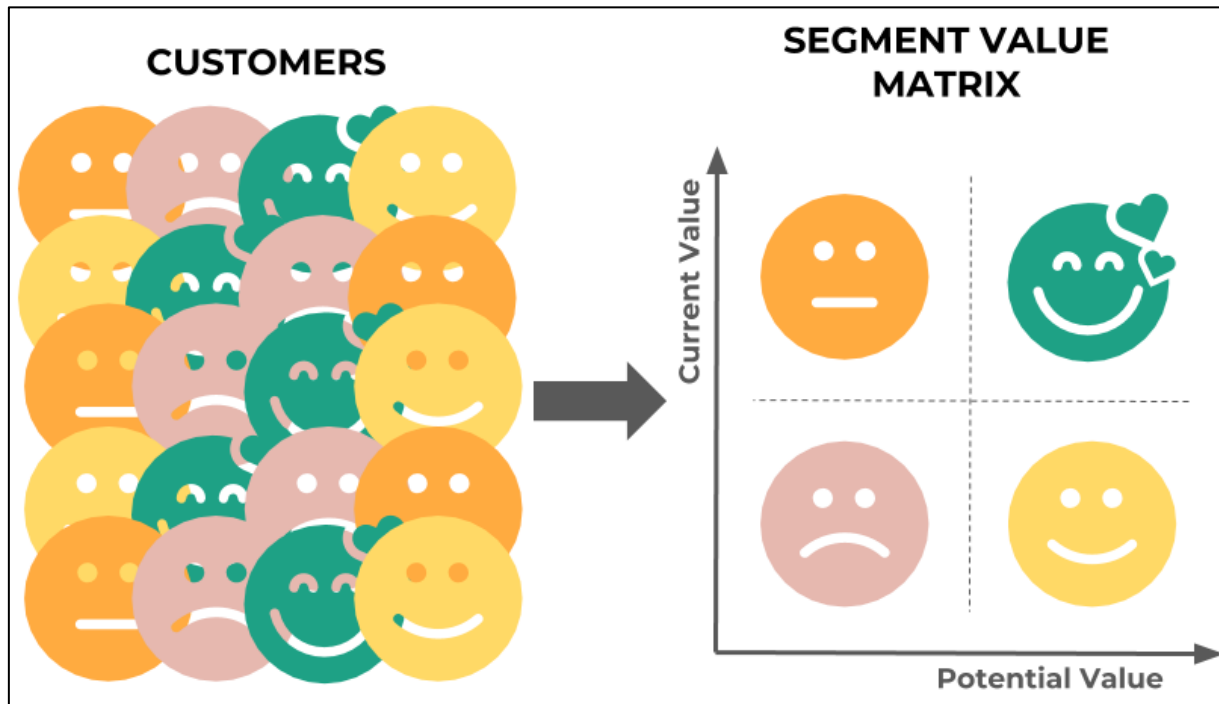
# MS Azure ML Studio

도매 고객 분류

```
#selection at the end --add back the deselected mirror modifier object
mirror_ob.select=1
modifier_ob.select=1
bpy.context.scene.objects.active = modifier_ob
print("Selected" + str(modifier_ob)) # modifier ob is the active ob
```

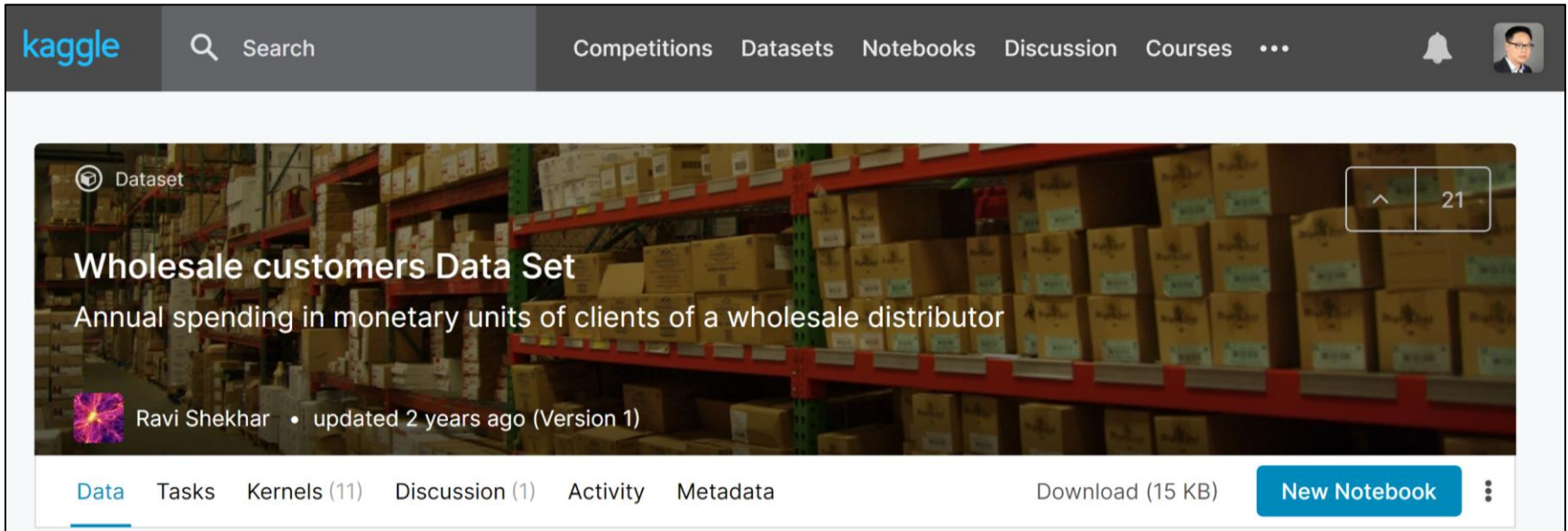
# 도매 고객 분류

- 특정 기준에 따라 **고객군을 분석해서 분류**함
- 마케팅 전략의 필수 요소
- 홈쇼핑 구매 패턴 분석, 임상실험 환자군 구별 등에 사용 가능



# 분석할 데이터

- UCI Wholesale Customer Dataset
- Kaggle에서 취득
- 8개 항목, 440개의 데이터로 구성됨



The screenshot displays the Kaggle interface for the 'Wholesale customers Data Set'. The header includes the Kaggle logo, a search bar, and navigation links for Competitions, Datasets, Notebooks, Discussion, Courses, and a menu icon. A user profile icon is visible in the top right. The main content area features a large image of a warehouse with the dataset title 'Wholesale customers Data Set' and a subtitle 'Annual spending in monetary units of clients of a wholesale distributor'. Below this, the creator's name 'Ravi Shekhar' and the update information 'updated 2 years ago (Version 1)' are shown. At the bottom, there are tabs for 'Data', 'Tasks', 'Kernels (11)', 'Discussion (1)', 'Activity', and 'Metadata'. On the right side of the bottom bar, there is a 'Download (15 KB)' link and a 'New Notebook' button.



# 분석할 데이터

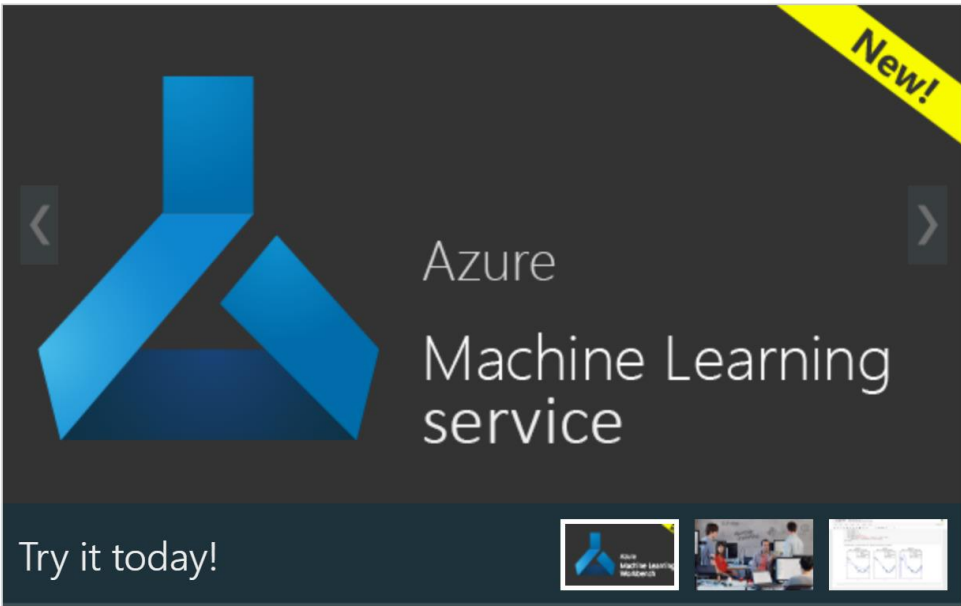
Column	Mean
Channel	구매처 1: 호텔, 레스토랑, 카페 2: 소매
Region	지역 1: 리스본(수도) 2: 포르투 3: 기타
Fresh	신선 식품(채소)
Milk	유제품
Grocery	식료품 및 잡화
Frozen	냉동 식품
Detergent	세제
Delicatessen	조제 식품 (육류, 치즈, 튀김, 수입 식품 등)



# 사용할 도구

## ▪ Microsoft Azure Machine Learning Studio

Microsoft Azure Machine Learning Studio (classic)



New!


Azure  
Machine Learning  
service

Try it today!

Welcome to Azure Machine Learning Studio (classic)

Try it for free

No [Azure subscription](#)? No credit card? No problem! Choose anonymous Guest Access, or sign in with your work or school account, or a Microsoft account.

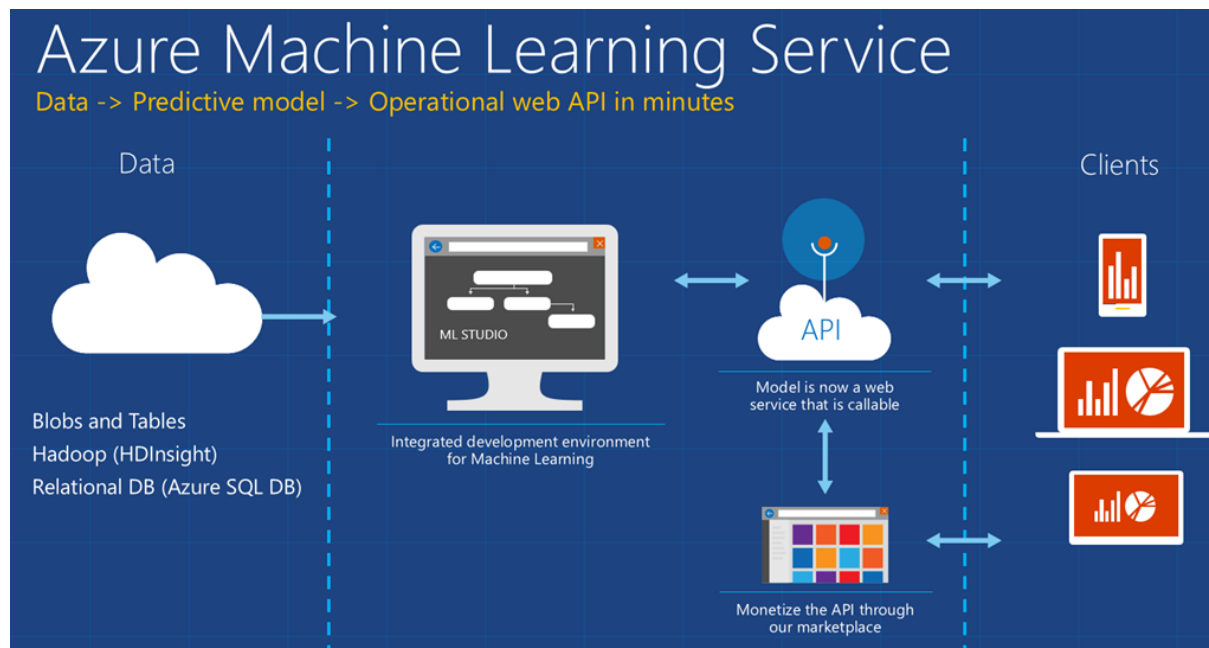
[Sign In](#) 

Not an Azure ML Studio (classic) user?  
[Sign up here](#)

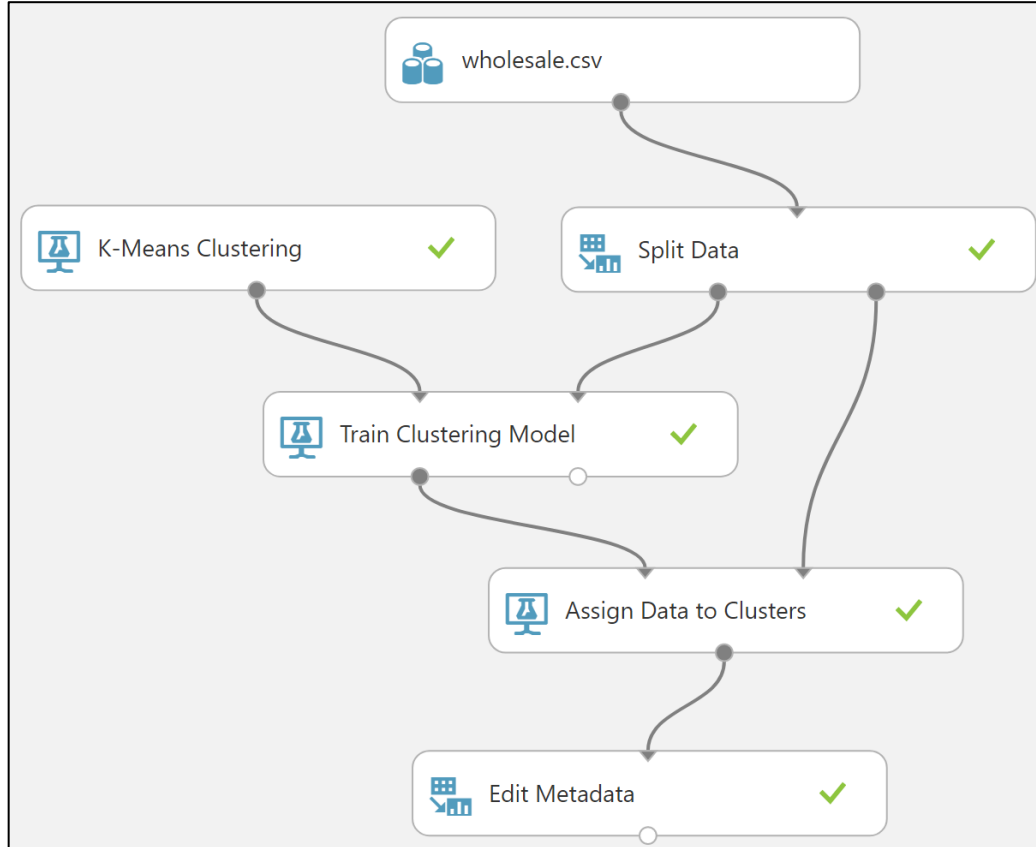
[Pricing & FAQ](#)

# 데이터 분석 절차

- **데이터 취득**: raw data를 모델에 추가하는 절차
- **데이터 준비**: 데이터를 학습에 용이하게 조작하는 절차
- **학습 및 평가**: 실질적으로 학습을 시도하고 결과를 확인하는 절차
- **배포**: 학습된 모델을 이용할 수 있도록 웹에 공개하는 절차



# 모델 개요



# 데이터 취득



1. Kaggle을 통해 **raw data 파일 획득**
2. Azure ML Studio에 접속해서 **새 dataset 등록**
3. **새 experiments 생성**
4. My datasets 항목에서 **등록된 dataset을 모듈로 추가**

wholesale > wholesale.csv > dataset

rows  
440

columns  
8

view as



Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen
2	3	12669	9656	7561	214	2674	1338
2	3	7057	9810	9568	1762	3293	1776
2	3	6353	8808	7684	2405	3516	7844



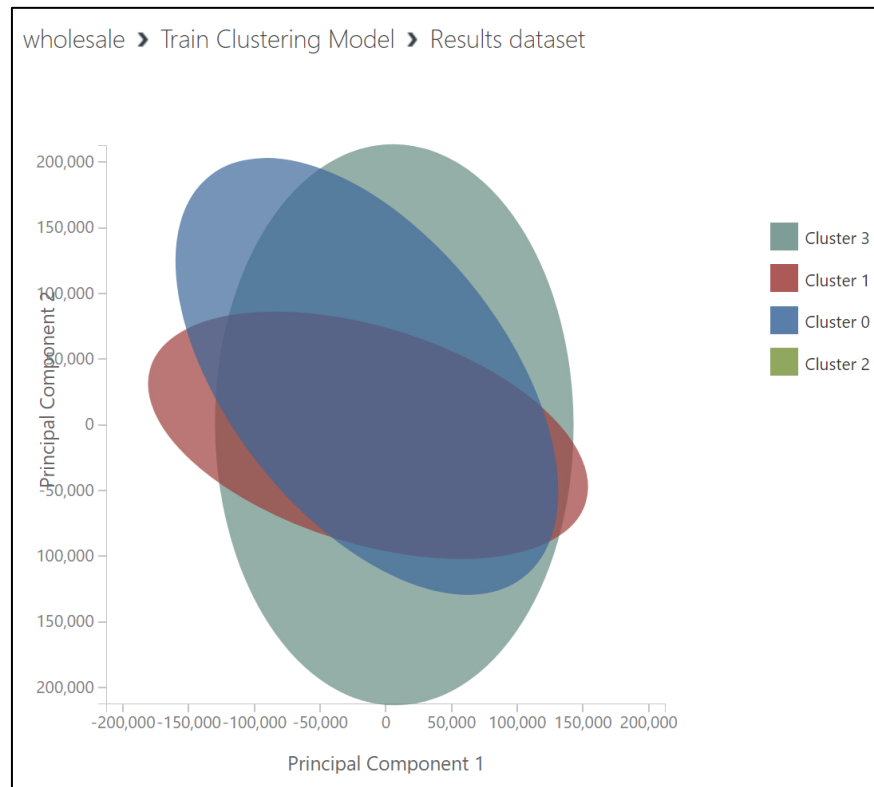
# 데이터 준비

1. **Split data** 모듈 추가
2. Dataset 모듈과 연결
3. 데이터를 **train set과 test set으로 분리**(70% : 30%)

rows		columns							
308		8							
		Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen
view as									
		2	3	2343	7845	11874	52	4196	1697
		2	3	31714	12319	11757	287	3881	2931
		1	3	34454	7435	8469	2540	1711	2893
		2	3	5417	9933	10487	38	7572	1282

# 학습 및 평가

1. **K-Means Clustering** 모듈을 추가하고 **centroid를 4개로 설정**
2. **Train Clustering Model** 모듈을 추가하고 **column selector를 통해 모든 column을 선택**



# 학습의 핵심

- ① 준비한 데이터를,
- ② K-Means Clustering 알고리즘을 적용해서
- ③ 학습하여 데이터 간의 연관성과 경향성을 알아냄

① Wholesale customer dataset + ② K-Means Clustering algorithm



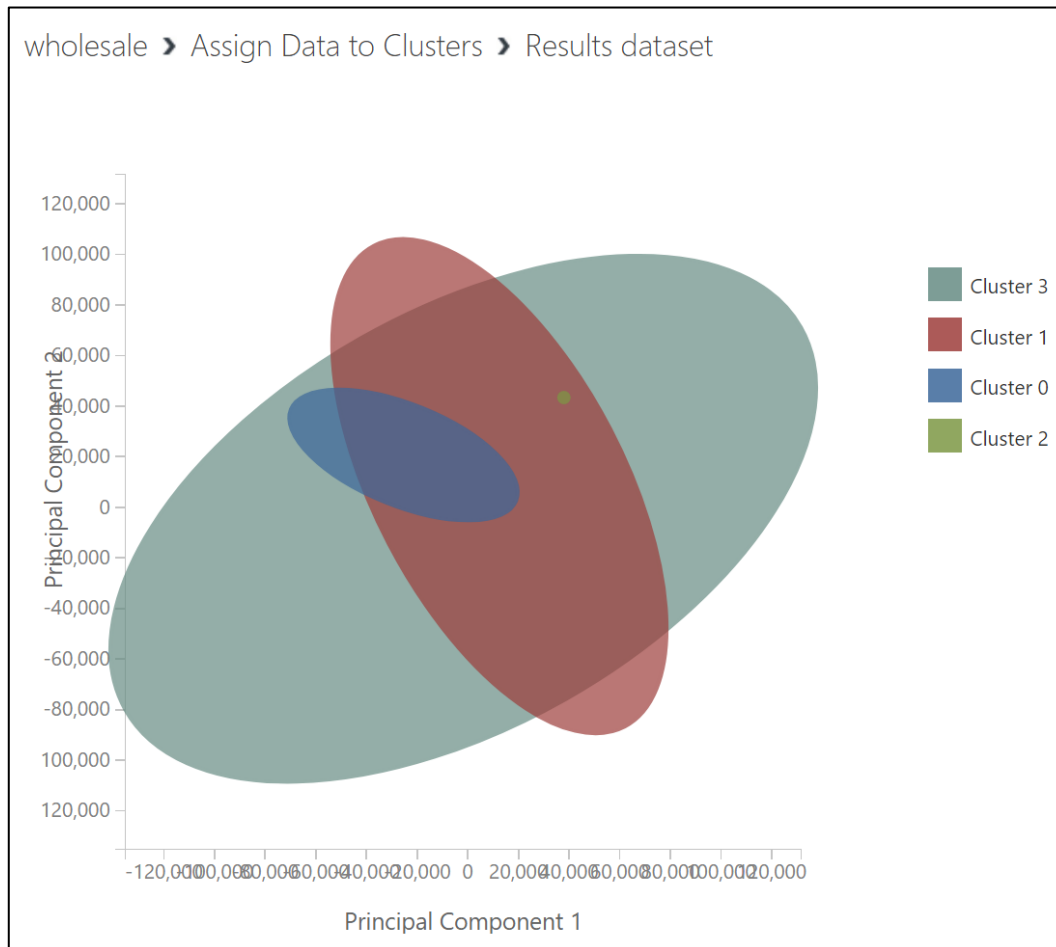
③ Train Clustering Model

# K-Means Clustering 알고리즘

- 대표적인 군집화 알고리즘
- 데이터로부터 임의의 집합 K개를 추출하여 구심점으로 설정하고, **집합 내 데이터와 구심점 간의 거리의 제곱합(=분산)을 최소화** 하는 집합을 찾는 게 목표
- Azure ML Studio에선 결과를 **주성분 분석(PCA)**의 형태로 확인 가능

# 학습 및 평가






## 3. Assign Data to Clusters 모듈을 추가하고 trained model과 test dataset을 연결



# 학습 및 평가

4. **Edit metadata** 모듈을 추가하고 results dataset 연결
5. Column selector를 통해 **판정 결과 열과 구심점 간의 거리 열을 선택**

adata > Results dataset

Assignments	DistancesToClusterCenter no.0	DistancesToClusterCenter no.1	DistancesToClusterCenter no.2	DistancesToClusterCenter no.3
				
3	42686.537233	17455.336197	85650.116575	6031.839839
1	47325.084212	10466.323546	79413.857136	13290.524456
3	39060.400325	26821.28152	93931.016934	7066.602781
3	46197.938431	19056.233533	88046.242051	5391.049559
3	35085.207993	27865.89468	93170.071854	10683.363714
3	50147.807657	27899.414349	97007.729199	8855.377175
3	47892.703603	21341.930094	90491.283571	5983.214468



# 배포

1. SET UP WEB SERVICE 버튼 클릭
2. DEPLOY WEB SERVICE 버튼 클릭
3. Test 텍스트를 클릭해서 테스트 실행

