

Day1. 강좌 소개 및 엑셀 데이터 분석

심선영 교수, 이주민 교수

강의 목표

- ❖ DA40 강좌의 흐름을 이해한다.
- ❖ 데이터 분석 방법론에 대해 이해할 수 있다.
- ❖ 캐글이 무엇인지 설명할 수 있다.
- ❖ 엑셀을 이용하여 데이터 분석을 할 수 있다.



경희사이버대학교

이 주 민

교수님

◆ 학 력

KAIST 전기 및 전자공학과 졸업
KAIST 전자전산학과 석사
KAIST 테크노경영대학원 경영공학 박사

◆ 경력

Deloitte Consulting, KAIST 지식경영센터 연구원, 행정안전부 경영평가 위원, 서울디지털재단 경영자문위원,
한국IT서비스학회 이사, 경영정보학회 이사, 한국지능정보시스템학회 이사, 공항공사 CS 자문위원 등

◆ 교육분야

빅데이터 마케팅, 의사결정을 위한 데이터 분석 ,
4차산업혁명과 스마트경영, 마케팅조사방법, 머신러닝,
인공지능 등

강의 목차

- ❖ 강좌소개
- ❖ 데이터 분석 개요
- ❖ 캐글 소개
- ❖ 엑셀 데이터 분석
- ❖ Wrap-up

강좌소개

DA40 Data Analysis Track

과정명	DA40 Data Analyst Fast Track	교육일정	8일(총 64시간) (8:00~17:00)
교육대상	Python/데이터 분석을 처음 접하는 연구원	강의형태	실시간 비대면(Webex)
대상특징	이과/공과 출신의 차량개발 연구 직종	실습도구	엑셀, 파이썬, 캐글
교육목표	데이터에 맞는 분석 가설을 세우고, 적용하여 분석 결과를 낼 수 있다. (Kaggle 초급 경진대회 참여 가능 수준)		

Day	강의 주제
1	강좌 소개 및 엑셀 데이터 분석
2	파이썬 기초 문법과 라이브러리
3	파이썬 시각화 및 머신러닝 입문
4	분류 알고리즘 이해 및 실습
5	캐글 데이터 분석 실습 1
6	캐글 데이터 분석 실습 2
7	회귀 알고리즘 이해 및 실습
8	캐글 데이터 분석 실습 3

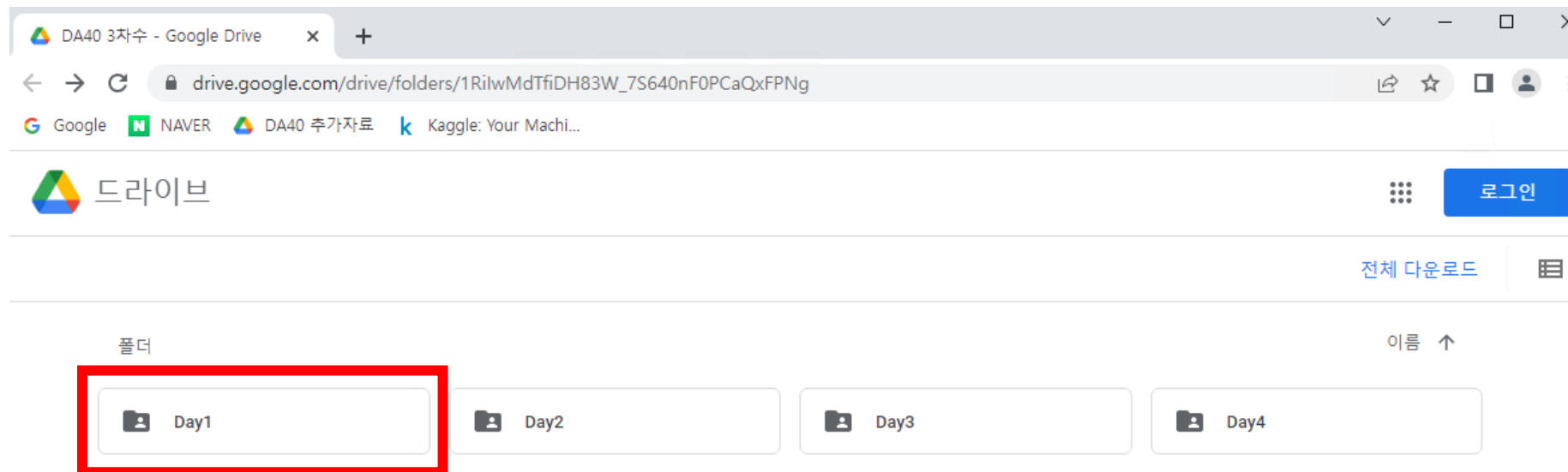
Day1. 강의 스케줄

목차	활동
강의 overview 데이터 분석 개요	강의
데이터 분석을 위한 엑셀 기능	엑셀 실습
엑셀 데이터 분석	엑셀 실습
Wrap-Up	학습 정리

수업 자료 다운받기!

■파일 다운로드

●바탕화면/



데이터 분석 개요

데이터 분석을 위한 데이터 과학 방법론



1. 목표 설정(Research Goal Setting)

- 데이터 분석의 첫 단계
- 연구 목표 설정
 - 무엇을 분석할까?
 - 결과는 어디에, 왜 필요한가?
 - 어떤 데이터가 필요한가?
 - 프로젝트 일정은 어떻게 되는 가 등
- 이해 당사자와 논의 필요

데이터 분석을 위한 데이터 과학 방법론

1. 목표 선정



2. 데이터수집



3. 데이터 가공



4. 모델구현



5. 모델 적용

2. 데이터 수집 (Data Retrieval)

- 필요 데이터 위치와 형태 확인
- 원시 데이터(raw data) 수집
 - 내부 데이터 베이스, 데이터 저장소
 - 외부 자료 등
 - 예) 정보 공유 정책에 따른 정부 기관 및 공적 조직에서 공유 데이터 증가 중
 - 예) 네이버, 트위터, 페이스북 등 서비스 데이터

사이트	설명
http://data.go.kr	한국 정부에서 제공하는 공공데이터
http://kostat.go.kr	한국 통계청에서 공개하는 데이터
http://opendata.hira.or.kr	한국 보건 의료 빅데이터 개방 시스템
http://www.localdata.kr	한국 지방행정 인허가 데이터
https://www.mcst.go.kr	한국 문화체육관광부 문화 데이터
http://data.seoul.go.kr	서울시 열린데이터 광장
https://data.gg.go.kr	경기도 공공데이터 개방 포털
http://data.gov	미국 정부의 공공데이터
http://data.worldbank.org	세계 은행에서 제공하는 개방 데이터
http://open.fda.gov	미국 식약청의 개방 데이터

데이터 분석을 위한 데이터 과학 방법론

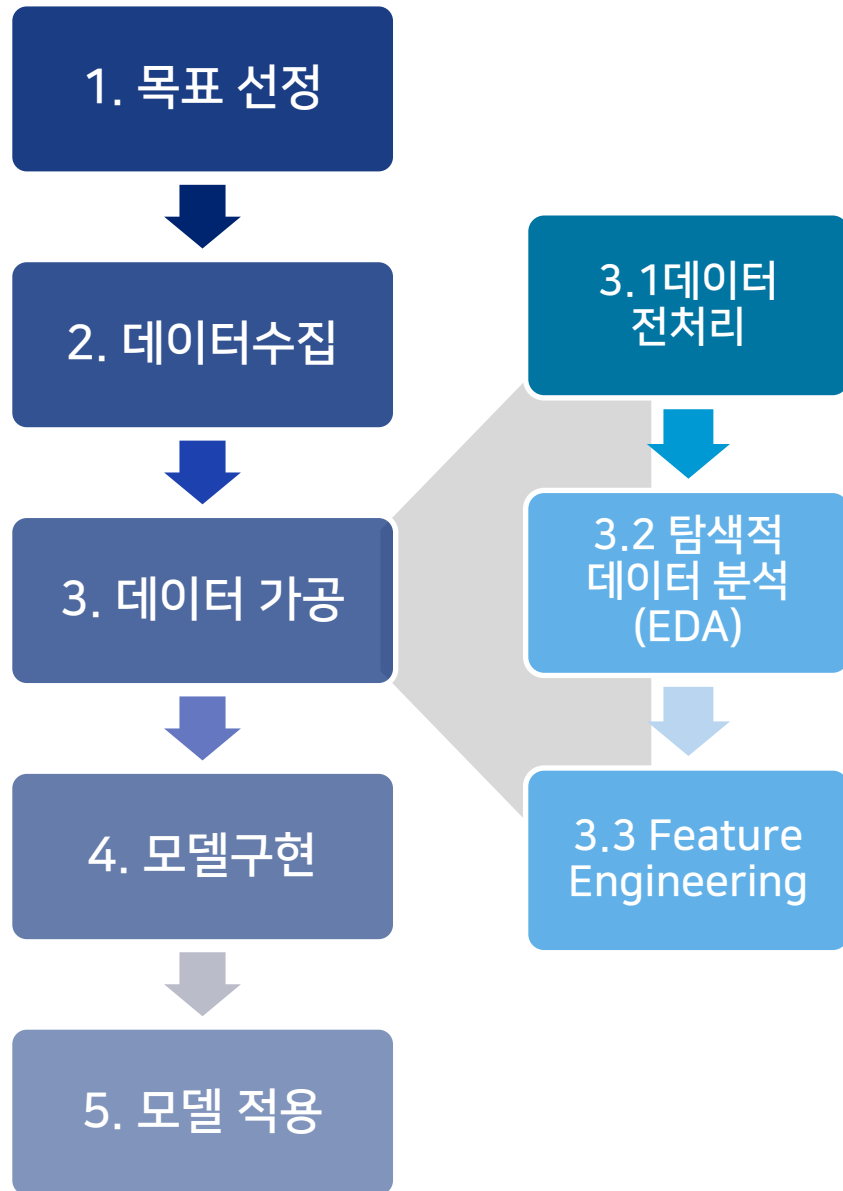


2. 데이터 수집 (Data Retrieval)

[표] 다양한 데이터 수집 기술

수집기술	설명	수집데이터
크롤링	SNS, 뉴스 웹 정보처럼 인터넷에서 제공하는 데이터를 수집 가능	웹주출 데이터
FTP	<ul style="list-style-type: none"> TCP/IP 프로토콜을 활용하는 인터넷 서버에서 각종 파일 송수신 보안을 강화를 위해 SFTP 사용 고려 서버 간 연동시에는 전용 네트워크 구축을 고려 	파일
OpenAPI	<ul style="list-style-type: none"> 서비스 데이터 등을 어디서나 쉽게 이용하도록 개방된 API로 데이터 수집 방식을 제공 다양한 애플리케이션을 개발할 수 있도록 개발자와 소비자에게 공개 	실시간 수집 데이터
RSS	<ul style="list-style-type: none"> 웹 기반의 최신 정보를 공유하기 위한 XML 기반의 콘텐츠 배급 프로토콜이다. 	XML 기반 웹콘텐츠
스트리밍	<ul style="list-style-type: none"> 인터넷에서 실시간으로 음성/오디오/비디오 데이터를 수집기술 	음성/오디오/비디오 실시간 데이터
로그 수집기	<ul style="list-style-type: none"> 웹 서버 로그, 웹 로그, 트랜잭션 로그, 클릭 로그, DB 로그 등 각종 로그 데이터를 수집하는 오픈 소스 기술 Chukwa, Flume, Scribe 등이 있다. 	로그
RDB 수집기	<ul style="list-style-type: none"> RDB에서 정형 데이터를 수집한 뒤 HDFS (하둡 분산 파일 시스템)나 HBase와 같은 NoSQL에 저장하는 오픈 소스 기술 Sqoop, Direct JDBC/ODBC 등 	RDB 기반 데이터

데이터 분석을 위한 데이터 과학 방법론



3.1 데이터 전처리 (Data Preparation)

- 분석 가능한 데이터 형태로 준비하는 작업
- 수집한 원시 데이터 품질 높이기 작업
- 오류를 여과(filtering)하거나 수정하고 정제
- 데이터 통합 및 형태 변환

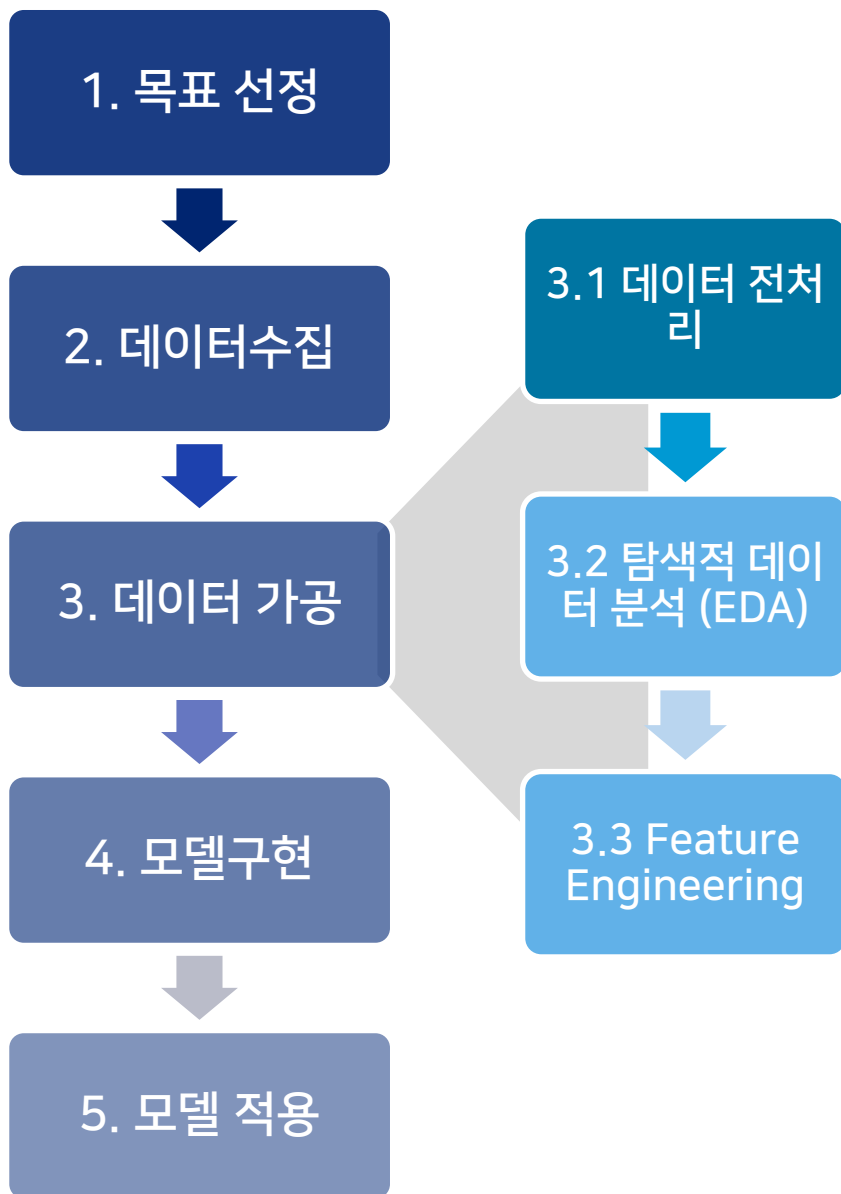
3.2 탐색적 데이터 분석 (EDA)

- 데이터 이해를 위해 수행
- 기술통계 (Descriptive statistics)
- 데이터 시각화(Data Visualization)

3.3 피쳐 엔지니어링(Feature Engineering)

- Feature extraction (특징량 추출)
- Feature selection (특징량 선택)
- 인코딩(Encoding)

데이터 분석을 위한 데이터 과학 방법론



[표] 데이터 준비에 필요한 작업

종류	설명
데이터 필터링	• 오류 발견, 보정, 삭제, 중복성 확인 등 과정을 통해 데이터 품질 향상
데이터 정제	• 결측치 보완, 이상치 제거, 데이터 불일치성 교정 등
데이터 통합	• 데이터 분석이 용이하도록 유사데이터 및 연계가 필요한 데이터를 통합
데이터 축소	• 분석 시간 단축을 위해 분석에 사용하지 않는 항목 제거
데이터 변환	• 데이터 분석에 용이한 형태로 데이터 유형을 변환

데이터 분석을 위한 데이터 과학 방법론

1. 목표 선정

2. 데이터수집

3. 데이터 가공

4. 모델구현

5. 모델 적용

4. 모델 구현

- 변수를 선택하여 모델 구성 및 실행/평가를 반복 수행하여 문제 해결모델을 완성한다.
- 데이터 세트 구분, 모델 생성 및 학습, 모델 평가를 포함
- 분석 데이터의 특성과 목적에 맞게 모델 유형을 선택한다.

[표] 데이터 분석 종류 (p.72)

통계분석 모델
(상관분석, 회귀분석,
분산분석 등)

데이터 마이닝 모델
(예측, 분류, 군집화,
패턴분석, 순차패턴
분석 등)

딥러닝 모델
(CNN, RNN 등)

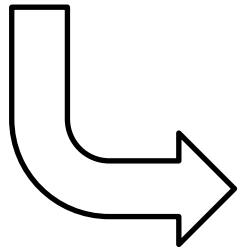
텍스트 마이닝 모델

소셜네트워크 분석
모델

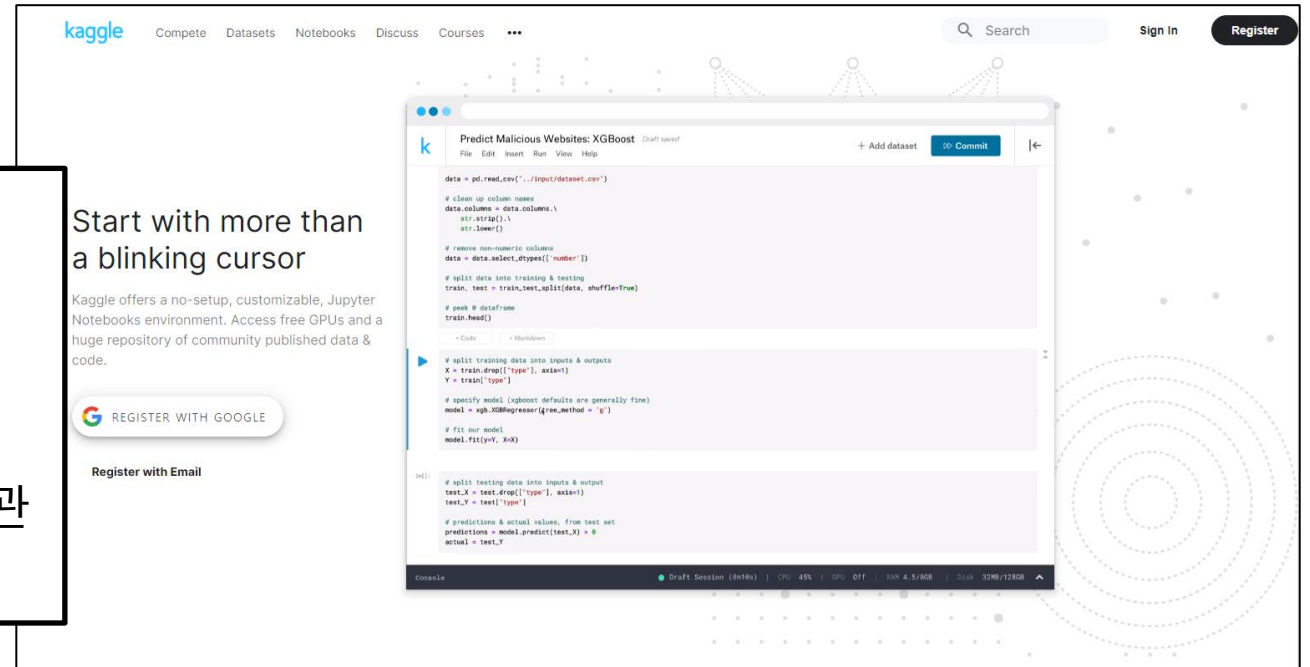
캐글 소개

캐글이란 kaggle

- <https://www.kaggle.com/>
- 데이터 분석에 관한 온라인 플랫폼
- 다양한 공개 데이터
- 데이터 분석 경진대회 (Competition) 상시 개최 (상금!!)
- 전세계의 데이터 과학자 & 인공지능 시스템 개발자가 관심을 가지고 있는 웹사이트

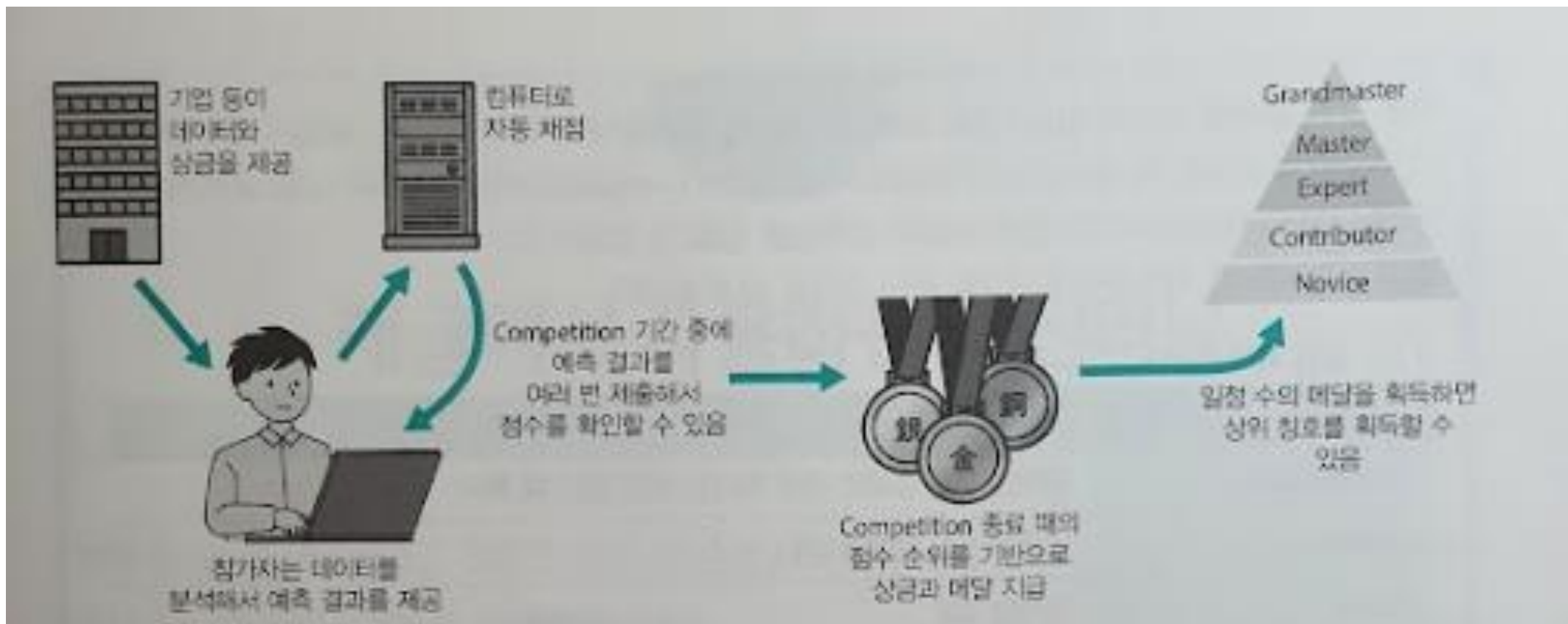


- 캐글에 올라온 데이터를 직접 분석해 보자!



캐글소개

캐글이란 kaggle

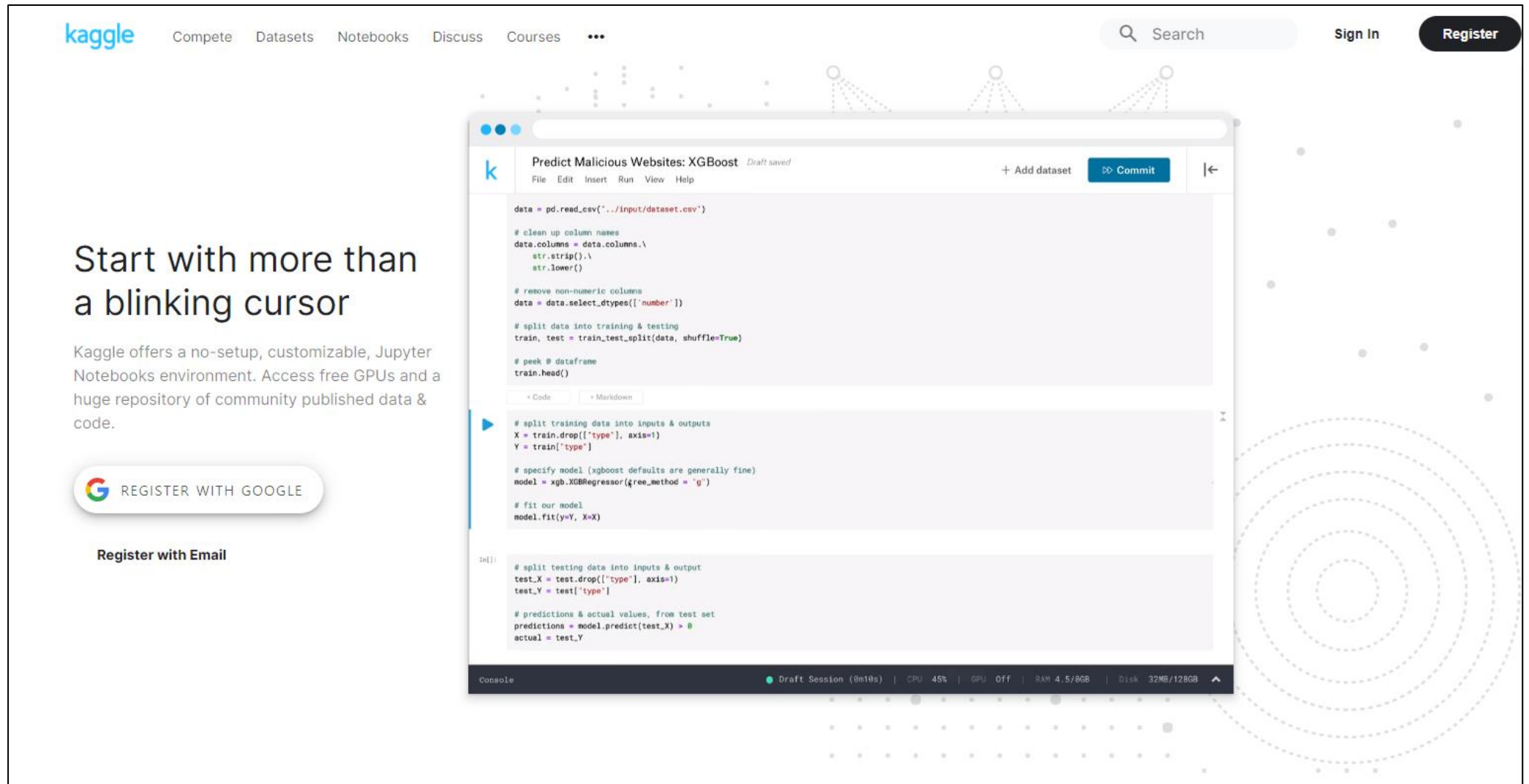


[그림] Kaggle Competition 개요

캐글소개

캐글이란 kaggle

캐글 사이트 들어가기
<https://www.kaggle.com/>



The image displays the Kaggle website interface. At the top, the Kaggle logo is on the left, and navigation links for 'Compete', 'Datasets', 'Notebooks', 'Discuss', and 'Courses' are in the center. On the right, there are links for 'Search', 'Sign In', and a 'Register' button. The main content area features the text 'Start with more than a blinking cursor' and a description: 'Kaggle offers a no-setup, customizable, Jupyter Notebooks environment. Access free GPUs and a huge repository of community published data & code.' Below this, there are two registration options: 'REGISTER WITH GOOGLE' (with a Google logo) and 'Register with Email'. A large, semi-transparent window in the center shows a Jupyter Notebook titled 'Predict Malicious Websites: XGBoost'. The notebook code includes data loading, cleaning, splitting, and model training using XGBoost. The console at the bottom shows the status of the 'Draft Session'.

kaggle Compete Datasets Notebooks Discuss Courses ... Search Sign In Register

Start with more than a blinking cursor

Kaggle offers a no-setup, customizable, Jupyter Notebooks environment. Access free GPUs and a huge repository of community published data & code.

REGISTER WITH GOOGLE

Register with Email

Predict Malicious Websites: XGBoost Draft saved

```
data = pd.read_csv('../input/dataset.csv')

# clean up column names
data.columns = data.columns.\
    str.strip().\
    str.lower()

# remove non-numeric columns
data = data.select_dtypes(['number'])

# split data into training & testing
train, test = train_test_split(data, shuffle=True)

# peek @ dataframe
train.head()

# split training data into inputs & outputs
X = train.drop(['type'], axis=1)
Y = train['type']

# specify model (xgboost defaults are generally fine)
model = xgb.XGBRegressor(tree_method = 'g')

# fit our model
model.fit(y=Y, X=X)

# split testing data into inputs & output
test_X = test.drop(['type'], axis=1)
test_Y = test['type']

# predictions & actual values, from test set
predictions = model.predict(test_X) > 0
actual = test_Y
```

Console Draft Session (9m10s) CPU 45% GPU Off RAM 4.5/8GB Disk 32MB/128GB

캐글 경진대회 실습

❖ 수업시간에 참여할 캐글 경진대회들

■ Titanic - Machine Learning from Disaster(타이타닉 생존자 예측)

- <https://www.kaggle.com/competitions/titanic>

■ Bike Sharing Demand Competition (자전거 공유)

- <https://www.kaggle.com/competitions/bike-sharing-demand>

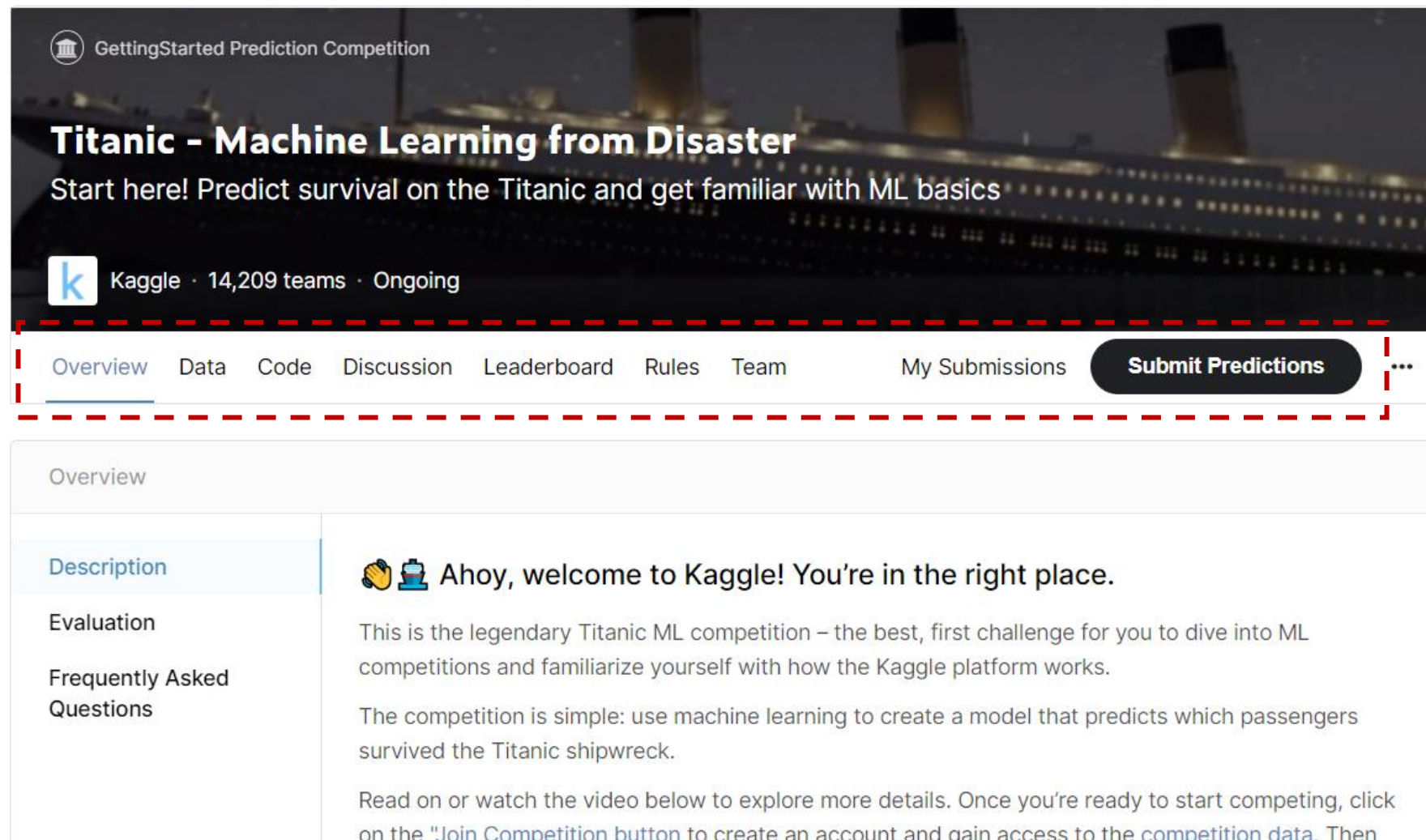
사례) Titanic-Machine Learning from Disaster Competition

- ❖ 캐글 home에서 “Titanic”으로 검색
- ❖ 검색 결과 중 “Competition”으로 표시된 “Titanic-Machine Learning from Disaster”선택

The screenshot shows the Kaggle search interface for the term 'titanic'. At the top, there's a search bar with 'titanic' entered. Below it, a summary bar shows the search scope: 'Searching for titanic within'. This is followed by a row of filters: Comments (64,508), Notebooks (51,110), Topics (5,380), Datasets (688), Competitions (137), and Users (86). Below these, there are more filters: Blogs (3) and Tutorials (1). On the left side, there's a 'Filter by' section with two categories: 'Date' and 'Viewed By You'. Under 'Date', there are three radio buttons: 'Last 90 days' (3,019 results), 'Last week' (249 results), and 'Today' (32 results). Under 'Viewed By You', there are two radio buttons: 'Viewed' (10 results) and 'Not Viewed' (121,903 results). On the right side, the search results are displayed. The first result is highlighted with a red dashed box. It is a 'Competition' titled 'Titanic - Machine Learning from Disaster'. It includes a small image of the Titanic ship, the text 'Getting Started', 'Ongoing • 14209 teams', and a link to the competition page. Below it, another result is partially visible, titled 'Spaceship Titanic'.

사례) Titanic-Machine Learning from Disaster Competition

❖ <https://www.kaggle.com/competitions/titanic>



The screenshot shows the Kaggle competition page for 'Titanic - Machine Learning from Disaster'. The header features a dark background with a ship at night and the text 'GettingStarted Prediction Competition', 'Titanic - Machine Learning from Disaster', and 'Start here! Predict survival on the Titanic and get familiar with ML basics'. Below this, it says 'Kaggle · 14,209 teams · Ongoing'. A navigation bar includes links for Overview, Data, Code, Discussion, Leaderboard, Rules, Team, My Submissions, and a prominent 'Submit Predictions' button. The main content area is titled 'Overview' and contains a 'Description' section with a welcome message and instructions on how to participate in the competition.

GettingStarted Prediction Competition

Titanic - Machine Learning from Disaster

Start here! Predict survival on the Titanic and get familiar with ML basics

Kaggle · 14,209 teams · Ongoing

Overview Data Code Discussion Leaderboard Rules Team My Submissions **Submit Predictions** ...

Overview

Description

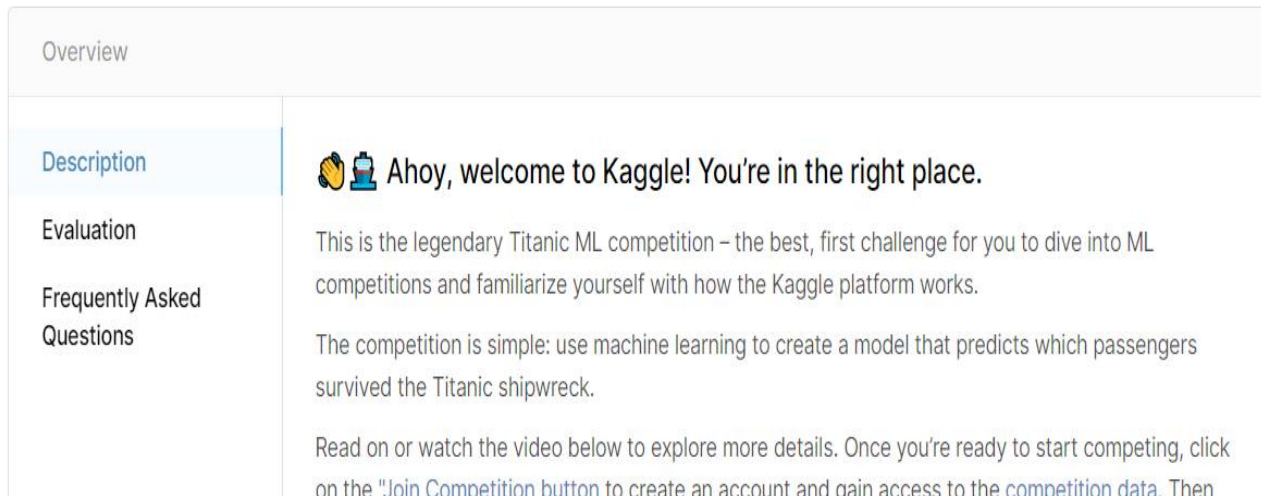
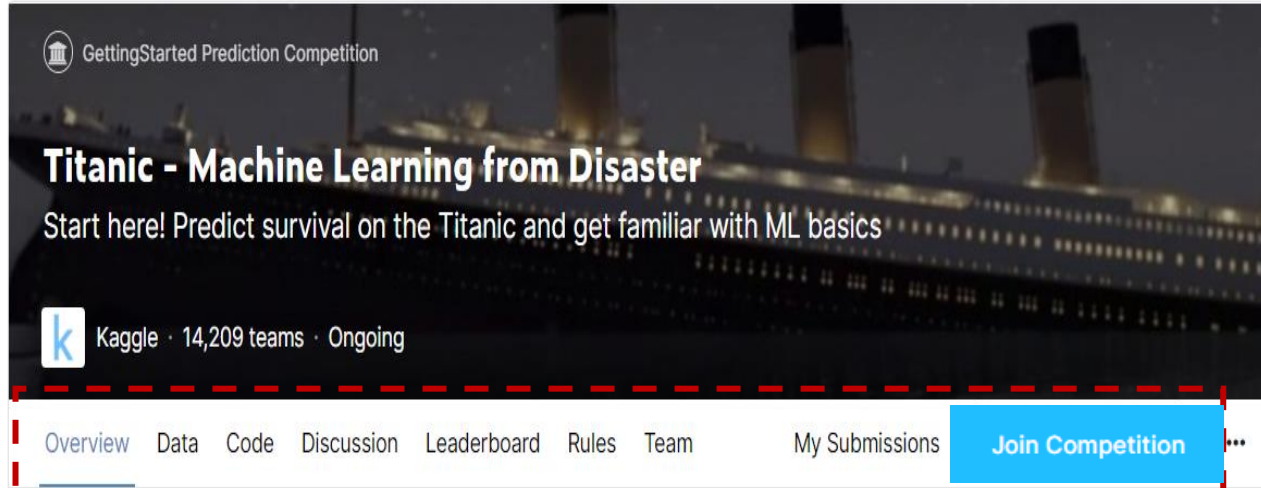
👋🏡 Ahoy, welcome to Kaggle! You're in the right place.

This is the legendary Titanic ML competition – the best, first challenge for you to dive into ML competitions and familiarize yourself with how the Kaggle platform works.

The competition is simple: use machine learning to create a model that predicts which passengers survived the Titanic shipwreck.

Read on or watch the video below to explore more details. Once you're ready to start competing, click on the "Join Competition button" to create an account and gain access to the competition data. Then

Titanic-Machine Learning from Disaster Competition



대회 참가

- 대회 메뉴에서 Titanic으로 검색
- [Join Competition](#) 을 눌러 대회 참가
- 관련 규정 등 동의 필요

메뉴 설명

메뉴	설명
Overview	경진대회에 대한 일반적인 설명
Data	경진대회에서 사용되는 데이터셋에 대한 설명, 데이터 다운로드
Code	경진대회와 관련된 각종 정보, 공개된 코드
Discussion	경진대회와 관련된 토론
Leaderboard	순위표
Rules	대회 규칙
Team	팀 생성 및 관리

사례) Titanic-Machine Learning from Disaster Competition

❖ <https://www.kaggle.com/competitions/titanic>



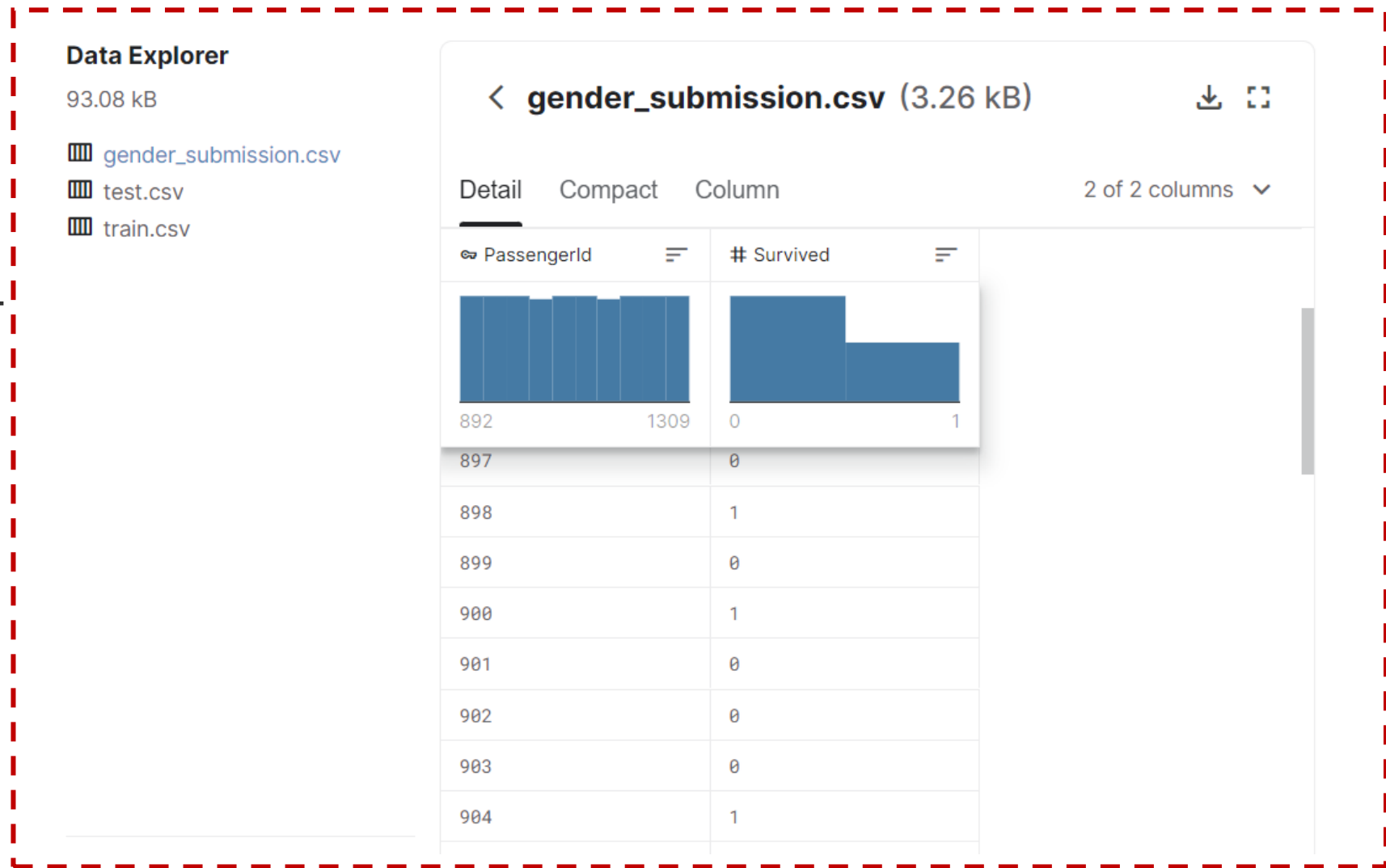
Data Dictionary

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

사례) Titanic-Machine Learning from Disaster Competition

❖ 데이터 파일 확인

- 총 3개의 CSV
- 필요 시 다운로드
- (캐글 노트북 작업을 위해서는)
- 다운받지 않아도 됨)



사례) Titanic-Machine Learning from Disaster Competition

Data Description

Overview

The data has been split into two groups:

- training set (train.csv)
- test set (test.csv)

The training set should be used to build your machine learning models. For the training set, we provide the outcome (also known as the “ground truth”) for each passenger. Your model will be based on “features” like passengers’ gender and class. You can also use [feature engineering](#) to create new features.

The test set should be used to see how well your model performs on unseen data. For the test set, we do not provide the ground

Data (34 KB)

[API](#) [kaggle competitions download -c titanic](#) [?](#) [Download All](#) [✕](#)

Data Sources	About this file	Columns
<div><div>gender_submissio...</div><div>2 columns</div></div>	An example of what a submission file should look like.	<div>PassengerId</div>
<div><div>test.csv</div><div>11 columns</div></div>	<i>These predictions assume only female passengers survive.</i>	<div>Survived</div>
<div><div>train.csv</div><div>12 columns</div></div>		

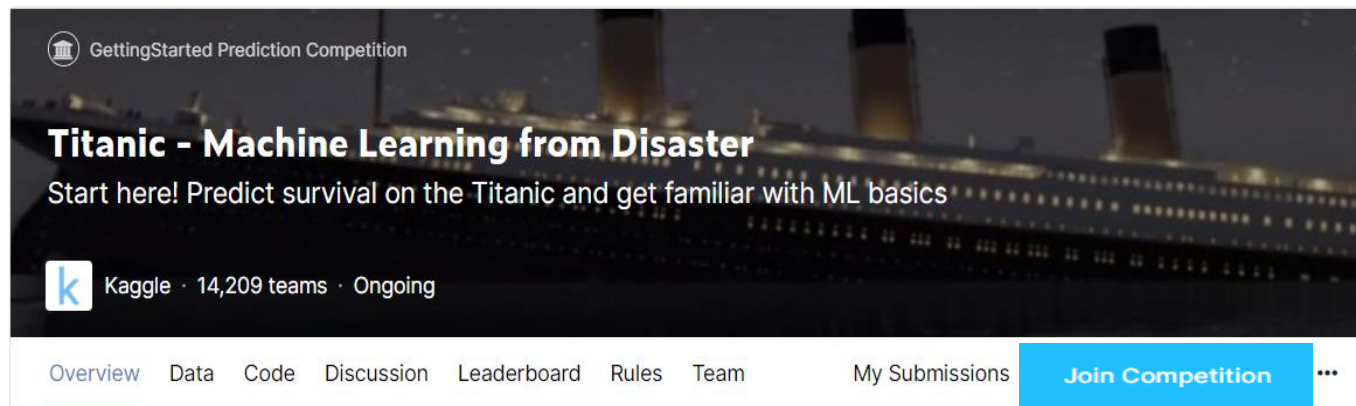
데이터 다운로드

- Download All 버튼을 눌러 데이터 다운로드

데이터 설명

- train.csv
 - 학습에 사용되는 데이터
- test.csv
 - 테스트에 사용되는 데이터
- gender_submission
 - 테스트 후 예측결과를 만들어 제출할 때 참고할 샘플 데이터

Titanic-Machine Learning from Disaster Competition



Overview

Description

Evaluation

Frequently Questions

Public Leaderboard

Private Leaderboard

This leaderboard is calculated with approximately 50% of the test data.

The final results will be based on the other 50%, so the final standings may be different.

Raw Data

Refresh

#	Team Name	Notebook	Team Members	Score	Entries	Last
1	YUMakaggle		<div><div></div><div></div></div>	1.00000	1	2mo
2	JacopJosef		<div><div></div><div></div></div>	1.00000	1	2mo
3	PolarodAngelos		<div><div></div><div></div></div>	1.00000	1	2mo
4	JeremyBush		<div><div></div><div></div></div>	1.00000	1	2mo
5	KayenosSchumer		<div><div></div><div></div></div>	1.00000	1	2mo
6	Brandon Pollard		<div><div></div><div></div></div>	1.00000	1	2mo
7	Daniella Hyde		<div><div></div><div></div></div>	1.00000	1	2mo
8	Gloria Mcgrath		<div><div></div><div></div></div>	1.00000	1	2mo

리더보드

- 결과를 제출한 팀들의 순위 확인
- 자신의 순위 확인 가능
- 점수는 예측 정확도 (1: 100%, 0: 0%)
- 정답을 모르는 상태에서는 80% 예측도 어려움

엑셀 데이터 분석

데이터 파일 살펴보기

❖ 파일:

- Day1_엑셀기초 실습_학생용.xlsx

수식

❖ 수식작성방법

수식의 구조 = (피연산자) (연산자) (피연산자)

The image shows an Excel spreadsheet with columns A, B, C, and D, and rows 5, 6, 7, and 8. In row 5, cell A5 contains the text "[B7]에 100 더하기 1000을 해 본다." In row 6, cell B6 is selected. A red box labeled "1" highlights the formula "=100+1000" entered directly into cell B6. Below this box, the text "셀에 직접 입력" (Direct input to cell) is written in red. In row 7, the formula bar is active, showing the formula "=100+1000". A red box labeled "2" highlights the formula in the bar. To the right of the formula bar, the text "수식 입력줄에 입력" (Input to formula bar) is written in red.

- 피연산자: 숫자, 셀 주소, 함수, 수식 등의 연산할 대상을 의미

수식

❖ 연산자

연산자 유형	연산자	기능
산술 연산자	+	더하기
	-	빼기
	*	곱하기
	/	나누기
비교 연산자	=	같다
	<>	다르다
	>	보다 크다
	<	보다 작다
	<=	보다 크거나 같다
	>=	보다 작거나 같다
문자열 연산자	&	문자열 연결

셀 참조 방식

상대참조

- ❖ 수식 복사 결과, 참조하는 셀이 상대적 위치에 따라 자동으로 바뀜
- ❖ 셀의 일반 주소 형태: A(열)1(행)

절대참조

- ❖ 셀 주소에서 행 번호와 열 번호 앞 모두 '\$' 표시
- ❖ 수식을 복사하여도 고정된 절대참조는 변하지 않음.
 - \$A(열)\$1(행)

혼합참조

- ❖ 행 번호와 열 번호 중 하나만 '\$'를 붙여 절대참조 형태 생성
- ❖ 열 고정 혼합참조 : \$A(열)1(행)
- ❖ 행 고정 혼합참조 : A(열)\$1(행)

셀 참조 방식

❖ 셀 참조 변환 키

- F4 단축키 이용
- [상대참조]->[절대참조]->[행고정]->[열고정]->[상대참조] 로 순환

[A1] -> [\$A\$1]->[A\$1]->[\$A1]->[A1]

엑셀의 함수

❖ 함수 vs. 수식

	A	B	C	D	E	F
1		국어	영어	수학	평균	분산
2	학생1	80	60	90	$= (B2 + C2 + D2) / 3$	$= ((B2 - \$E\$2)^2 + (C2 - \$E\$2)^2 + (D2 - \$E\$2)^2) / 3$

	A	B	C	D	E	F
1		국어	영어	수학	평균	분산
2	학생1	80	60	90	$= \text{AVERAGE}(B2:D2)$	$= \text{VAR.P}(B2:D2)$

엑셀의 함수

❖ 함수의 형식

⊕ = 함수명 (인수1, 인수2, .. ,)

⊕ 예시
= SUM(A1, A2, A3, B1:B3, 500)
= AVERAGE(C1:C3)

❖
=SUM(A1, A2, A3, B1:B3, 500, AVERAGE(C1:C3))
=IF(AVERAGE(A1:A3)>70, "우수", "노력")

엑셀의 함수

❖ 중첩함수

- ❗ 함수 안에 또 다른 함수가 인수로 사용된 함수
- ❗ 다른 함수의 결과값을 인수로 사용

- = SUM(A1, A2, A3, B1:B3, 500, AVERAGE(C1:C3))
- = IF(AVERAGE(A1:A3)>70, "우수", "노력")
- = IF(A>90, "매우 우수",
IF(A>80, "우수", IF(A>70, "보통", IF(A>60, "노력"))))

엑셀의 함수

❖ 함수의 종류

범주	기능
재무	재무 및 회계 관련 함수
날짜 및 시간	날짜나 시간을 계산하는 함수
수학/삼각	수학적 계산이나 삼각 함수를 계산하는 함수
통계	통계값을 계산하는 함수
찾기/ 참조영역	셀을 참조하거나 행 또는 열과 같은 위치를 찾기 위한 함수
데이터 베이스	데이터 목록에서 조건에 맞는 데이터를 검색하고 추출하기 위한 함수

엑셀의 함수

❖ 함수의 종류

범주	기능
텍스트	텍스트를 조작하는 함수
논리	참 또는 거짓과 같은 논리값을 판별하거나 여러 개의 조건을 검사하는 함수
정보	셀에 저장된 데이터의 종류를 판단하거나 정보를 표시하는 함수
공학	공학용 계산을 위한 함수
큐브	SQL 서버의 데이터를 다루기 위해 OLAP을 지원하는 함수
호환성	엑셀 2010 이상의 버전에서 이름이 바뀐 기존 버전의 함수를 사용 가능하도록 지원하는 함수

엑셀의 함수

❖ 함수 사용 방법

■ 수식 입력줄

- '='을 입력 후 원하는 함수 이름의 일부만 입력하면 '수식 자동완성' 기능에 의해 입력 문자와 일치하는 함수 목록이 표시

The screenshot shows the Excel interface with the formula bar at the top. The formula bar contains '=av'. Below the formula bar, a dropdown menu is open, displaying a list of functions that start with 'av': AVEDEV, AVERAGE, AVERAGEA, AVERAGEIF, and AVERAGEIFS. The first function, AVEDEV, is highlighted in blue. To the right of the dropdown menu, a tooltip text reads: '데이터 요소의 절대 편차의 평균을 구합니다.'

	A	B	C	D	E
1		국어	영어	수학	평균
2	학생1	80	60	90	=av
3	학생2	50	100	75	
4					
5					

엑셀의 함수

❖ 함수 사용 방법

■ 함수 라이브러리

- 수식 탭의 함수라이브러리 그룹에서 원하는 함수 범주와 해당 함수를 선택

The screenshot shows the Excel ribbon with the '수식' (Formulas) tab selected. The '함수 라이브러리' (Function Library) group is visible, and the '기타 함수' (More Functions) button is highlighted. A dropdown menu is open, showing a list of functions. The 'AVERAGE' function is highlighted in the list.

	A	B	C	D	E	F
1		국어	영어	수학	평균	
2	학생1	80	60	90		
3	학생2	50	100	75		

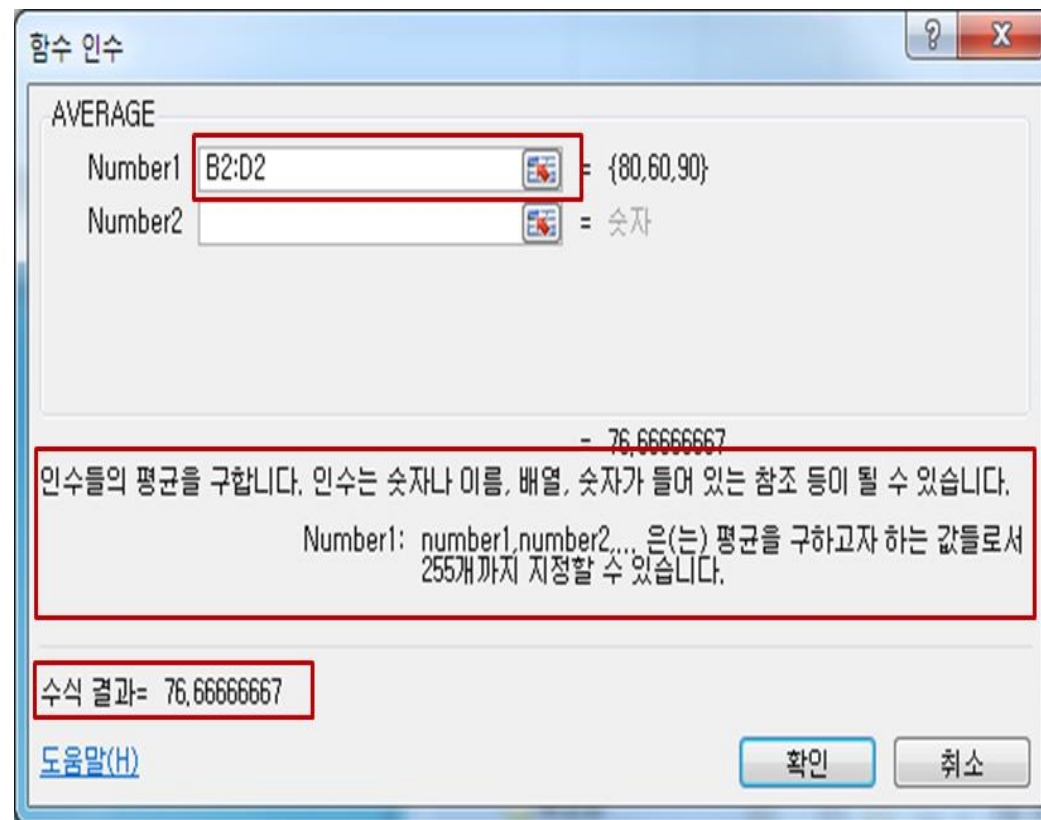
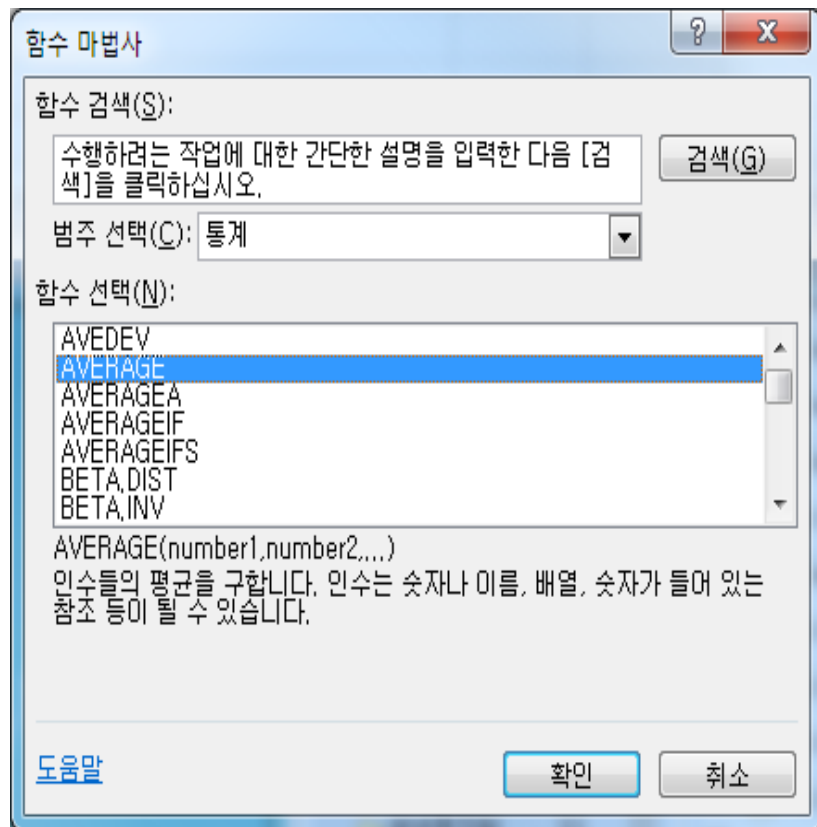
엑셀의 함수

❖ 함수 사용 방법

■ 함수 마법사

- 수식 입력줄 앞이나 [수식] - [함수라이브러리]의 가장 앞부분에 있는 ' f_x 함수삽입' 아이콘을 클릭하여 입력

실습 Go!



엑셀 함수

❖ 논리함수

- ⊕ 참 또는 거짓과 같은 논리값을 판별하는 함수
- ⊕ 여러 개의 조건을 검사하는 함수

IF()

NOT ()

AND(), OR ()

IFERROR ()

엑셀 함수

❖ 논리함수

■ if 문

= IF (logical_test, value_if_true, [value_if_false])

• 예시) 직원 평가

	A	B	C
1	직원 평가표		
2	성명	Point	결과
3	홍길동	85	우수
4	김철수	72	노력
5	이민우	63	노력
6	박영훈	75	우수
7	최민준	69	노력
8			

	True	False
<i>f_x</i>	=if(B3>=75,"우수","노력")	

평가 Point가 75점 이상인 직원은 '우수',
이하는 '노력'이라는 문구 추가의 예제

실습 Go!

엑셀 함수

❖ 논리함수

= AND (logical1, [logical2], ...)

설명

인수가 모두 참(TRUE)일 경우에만 참(TRUE)을 반환한다.

인수

- logical1: 참 또는 거짓을 판단할 수 있는 첫 번째 조건 필수 요소
- [logical2]: 참 또는 거짓을 판단할 수 있는 두 번째 조건 추가 요소

= OR (logical1, [logical2], ...)

설명

인수가 하나라도 참(TRUE)이면 참(TRUE)을 반환한다.

인수

- logical1: 참 또는 거짓을 판단할 수 있는 첫 번째 조건 필수 요소
- [logical2]: 참 또는 거짓을 판단할 수 있는 두 번째 조건 추가 요소

실습 Go!

엑셀 함수

❖ 논리함수

= IFERROR (value, value_if_error)


설명

수식이 오류이면 사용자가 지정한 값을 반환하고, 그렇지 않으면 수식 결과를 반환한다.

인수

- value: 오류를 검사할 인수
- value_if_error: 수식에서 오류가 발생할 경우에 반환할 값

	A	B	C	D	E	J
7	제품번호	용기	포장 색깔	주문 수량	재고 수량	추가입고수량
8	1	유리	브라운	146	모름	파악안됨
9	2	비닐	화이트	572	423	149
10	3	플라스틱	화이트	638	500	138
11	4	유리	화이트	155	100	55
12	5	비닐	브라운	229	모름	파악안됨

 =IFERROR(D8-E8,"파악안됨")

오류를 검사할 수식

오류 발생 시 반환할 값

실습 Go!

엑셀의 함수

❖ 통계함수

⊕ 자료의 수집, 분류, 정리, 요약 및 추론을 위해 사용하는 함수

RANK()

COUNTIF(), COUNTIFS()

COUNT(), COUNTA(),
COUNTBLANK()

AVERAGEIF(),
AVERAGEIFS()

엑셀의 함수

❖ 통계함수

= RANK (number, ref, [order])

인수

- number: 순위를 구하려는 수
- Ref: 수 목록
- [order]: 순위를 결정하는 방법
 - 0 이거나 생략된 경우: 내림차 순으로 순위 부여
 - 1: 오름차 순으로 순위 부여

함수 범주

정확도 향상/이름 변경

호환성 함수

통계

RANK.EQ

RANK.AVG

RANK

엑셀 함수

❖ 통계함수

■ count 함수들

= COUNT (value1, [value2], ..)
= COUNTA (value1, [value2], ..)
= COUNTBLANK (value1, [value2], ..)

- COUNT() : 범위나 숫자배열에 포함된 숫자의 개수를 센다.
- COUNTA() : 범위에서 비어있지 않은 셀의 개수를 센다.
- COUNTBLANK() : 범위에서 빈 셀의 개수를 센다.

엑셀 함수

❖ 통계함수

■ count 함수들

= COUNT (value1, [value2], ..)
 = COUNTA (value1, [value2], ..)
 = COUNTBLANK (value1, [value2], ..)

	A	B	C	D	E	F
1	테스트용 데이터			Count예제		
2	12022-01-01	5300		COUNT	4	=COUNT(A2:B5) 1345
3	2월요일	6		COUNTA	7	=COUNTA(A2:B5) 1234578
4	30	7#N/A		COUNTBLANK	1	=COUNTBLANK(A2:B5) 6
5	4300	8#DIV/0!		COUNTIF	2	=COUNTIF(A2:B5, "300") 45

통계자료로 처리할 영역(A2:B5)

COUNTIF할 조건

※ 날짜 데이터는 엑셀에서 수치 자료로 판단됨을 주의!

예제2-8
실습 Go!

엑셀 함수

❖ 통계함수

= COUNTIF (range, criteria)

설명

지정한 조건에 맞는 값의 개수를 알려준다.

인수

- range: 조건을 적용 할 셀 범위
- criteria: 조건을 정의하는 부분

❖ 통계함수

= COUNTIFS(criteria_range1, criteria1, [criteria_range2, criteria2], ...)

조건 범위 1		조건 범위 2		
날짜	제품	사이즈	주문수량	
10월 01일	자켓	S	46	
10월 02일	티셔츠	M	72	
10월 03일	자켓	M	38	
10월 04일	티셔츠	S	55	
10월 05일	티셔츠	M	29	
10월 06일	자켓	M	33	
10월 07일	자켓	S	77	
10월 08일	티셔츠	S	32	
제품	사이즈	평균 주문수량	주문건수	
자켓	S	61.5	(\$9,812)	
자켓	M	35.5		
조건 1	조건 2	43.5		
		50.5		

[illegible]

엑셀 함수

❖ 통계함수

=AVERAGEIF (range, criteria, [average_range])
=AVERAGEIFS (average_range, criteria_range1, criteria1,
[criteria_range2, criteria2], ...)

설명

- AVERAGEIF () : 범위에서 지정한 조건을 만족하는 모든 셀의 평균(산술 평균)을 구한다.
- AVERAGEIFS () : 여러 조건을 만족하는 모든 셀의 평균을 구한다.

인수

- range: 지정된 조건을 적용할 범위
- criteria: 지정된 조건
- [average_range]: 평균을 계산할 셀 범위로 생략할 수 있다. 생략시 range가 사용된다.

엑셀 함수

❖ 통계함수

=AVERAGEIF (range, criteria, [average_range])
 =AVERAGEIFS (average_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)

날짜	제품	사이즈	주문수량
10월 01일	자켓	S	46
10월 02일	티셔츠	M	72
10월 03일	자켓	M	38
10월 04일	티셔츠	S	55
10월 05일	티셔츠	M	29
10월 06일	자켓	M	33
10월 07일	자켓	S	77
10월 08일	티셔츠	S	32

제품	사이즈	평균 주문수량
자켓	S	59.12)
자켓	M	35.5
조건 1	조건 2	43.5
		50.5

조건 범위 1 조건 범위 2

평균 범위

VERAVERAGEIFS(\$D\$2:\$D\$9,\$B\$2:\$B\$9,A12,\$C\$2:\$C\$9,B12)

함수 인수

AVERAGEIFS

Average_range \$D\$2:\$D\$9 = {46;72;38;55;29;33;77;32}

Criteria_range1 \$B\$2:\$B\$9 = {"자켓";"티셔츠";"자켓";"티셔츠";"티셔츠";"자켓";"자켓";"티셔츠"}

Criteria1 A12 = "자켓"

Criteria_range2 \$C\$2:\$C\$9 = {"S";"M";"M";"S";"M";"M";"S";"S"}

Criteria2 B12 = "S"

= 61.5

주어진 조건에 따라 지정되는 셀의 평균(산술 평균)을 구합니다.

Criteria2: 음(는) 숫자, 식 또는 텍스트 형식의 조건으로 평균을 구할 셀을 정의합니다.

수식 결과= 61.5

도움말(H)

확인 취소

엑셀 함수

❖ 수학·삼각함수

= SUMIF(range, criteria, [sum_range])

설명

지정한 조건에 맞는 범위의 값을 더한다.

인수

- range: 조건을 적용 할 셀 범위
- criteria: 조건을 정의하는 부분으로 숫자, 식, 셀참조, 텍스트, 함수 등이 가능
- sum_range: 실제 값을 더할 셀 범위 (합계를 구할 부분)

엑셀 함수

❖ 수학·삼각함수

= SUMIFS (sum_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)

설명

다중 조건에 대한 합계를 구함

인수

- criteria_range: 조건을 적용 할 셀 범위
- criteria: 조건을 정의하는 부분으로 숫자, 식, 셀참조, 텍스트, 함수 등이 가능
- sum_range: 실제 값을 더할 셀 범위 (합계를 구할 부분)

= SUMIF(range, criteria, [sum_range])

The figure shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G
1	영업점별 판매현황						
2	영업점	지역	수탁금액		지역별	합계	
3	영업부	종구	1,250		종구	2839	
4	서소문	종구	633		강서구	1057	
5	퇴계로	종구	956				
6	화곡동	강서구	568				
7	김포	강서구	489				
8							

The formula bar shows: `=SUMIF(B3:B7, E3, C3:C7)`

Annotations:

- 비교할 영역 (Comparison Area):** Points to the range B3:B7 (지역 column).
- 합계할 영역 (Summation Area):** Points to the range C3:C7 (수탁금액 column).
- 비교할 값 (Comparison Value):** Points to the cell E3 (종구).

❖ 수학.삼각함수

= SUMIFS (sum_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)

SUMIFS
=SUMIFS(E2:E16,B2:B16,"자켓",C2:C16,"흰색")

	A	B	C	D	E	F	G	H	I	J	K	L
1	입고일	분류	색	사이즈	수량		흰색 자켓 수량 합계					
2	05월 02일	자켓	흰색	3	54		=SUMIFS(E2:E16,B2:B16,"자켓",C2:C16,"흰색")					
3	05월 03일	스커트	검정	5	69							
4	05월 04일	바지	흰색	7	37							
5	05월 06일	자켓	검정	7	55							
6	05월 07일	스커트	흰색	11	30							
7	05월 11일	자켓	흰색	11	64							
8	05월 12일	바지	흰색	13	47							
9	05월 13일	자켓	검정	13	64							
10	05월 17일	바지	흰색	11	62							
11	05월 20일	스커트	흰색	13	76							
12	05월 21일	자켓	검정	7	82							
13	05월 22일	스커트	흰색	5	65							
14	05월 23일	바지	검정	3	35							
15	05월 24일	자켓	흰색	3	96							
16	05월 25일	자켓	검정	7	56							

함수 인수

SUMIFS

Sum_range	E2:E16	=	{54;69;37;55;30;64;47;64;62;76;82}
Criteria_range1	B2:B16	=	{ "자켓"; "스커트"; "바지"; "자켓"; "스커트"; "바지"; "자켓"; "스커트"; "바지"; "자켓"; "스커트"; "바지"; "자켓"; "스커트"; "바지" }
Criteria1	"자켓"	=	"자켓"
Criteria_range2	C2:C16	=	{ "흰색"; "검정"; "흰색"; "검정"; "흰색"; "검정"; "흰색"; "검정"; "흰색"; "검정"; "흰색"; "검정"; "흰색"; "검정"; "흰색" }
Criteria2	"흰색"	=	"흰색"

= 214

주어진 조건에 따라 지정되는 셀을 더합니다.

Criteria2: 은(는) 숫자, 식 또는 텍스트 형식으로 된 조건으로 더할 셀을 정의합니다.

수식 결과= 214

[도움말\(H\)](#)

확인
취소

엑셀 함수

❖ 찾기·참조 함수

= VLOOKUP (lookup_value, table_array, col_index_num, [range_lookup])

설명

세로 (vertical) 방향의 데이터 목록에서 원하는 값을 찾는다.

인수

- lookup_value: 테이블의 첫 번째 행 또는 열에서 찾으려는 값
- table_array: 데이터를 찾을 테이블
- col_index_num: 테이블에서 가져올 데이터가 있는 열 번호
- [range_lookup]: (생략 가능): 정확히 일치하는 값을 찾을 것인지, 수치 범위 내에서 근사값을 찾을 것인지에 대한 선택 사항. 1 또는 생략 시 근사값을 찾고, 0일 때는 정확하게 일치하는 값을 찾는다.

엑셀 함수

❖ 찾기·참조 함수

= HLOOKUP (lookup_value, table_array, row_index_num, [range_lookup])

설명

가로 (horizontal) 방향의 데이터 목록에서 원하는 값을 찾는다.

인수

- lookup_value: 테이블의 첫 번째 행 또는 열에서 찾으려는 값
- table_array: 데이터를 찾을 테이블
- row_index_num: 테이블에서 가져올 데이터가 있는 행 번호
- [range_lookup]: (생략 가능): 정확히 일치하는 값을 찾을 것인지, 수치 범위 내에서 근사값을 찾을 것인지에 대한 선택 사항. 1 또는 생략 시 근사값을 찾고, 0일 때는 정확하게 일치하는 값을 찾는다.

실습 Go!

엑셀 피벗테이블 & 차트

❖ 피벗테이블(Pivot Table)

- 피벗: 회전축, 중심점
- 데이터베이스를 요약 분석할 수 있도록 도와주는 기능
- 중심이 되는 축(기준)을 중심으로 데이터를 회전하여 사용자가 필요한 관점으로 정보를 정리

엑셀 피벗테이블 & 차트

❖ 피벗테이블(Pivot Table)

거래처의 제품별 수량

	A	B
3	행 레이블	합계 : 수량
4	A마트	57
5	통조림	10
6	차	23
7	음료	21
8	냉동식품	3
9	개나리 마트	39
10	유제품	19
11	라면	5
12	냉동식품	15
13	경호 다방	76
14	음료	37
15	양념	33
16	과자	6
17	기다려 마트	58
18	통조림	30
19	라면	9
20	냉동식품	10
21	과자	9
22	나다 마트	92
23	잼	10
24	유제품	33
25	양념	23
26	라면	26

거래데이터(원본)

	A	B	C	D
1	거래처명	제품	단가	수량
2	A마트	차	3000	23
3	A마트	냉동식품	6000	3
4	A마트	음료	700	21
5	A마트	통조림	2500	10
6	개나리 마트	라면	500	5
7	개나리 마트	유제품	700	11
8	개나리 마트	유제품	700	8
9	개나리 마트	냉동식품	6000	9
10	개나리 마트	냉동식품	6000	6
11	경호 다방	양념	3000	33
12	경호 다방	음료	700	37
13	경호 다방	과자	800	6
14	기다려 마트	냉동식품	6000	8
15	기다려 마트	통조림	2500	30
16	기다려 마트	라면	500	9
17	기다려 마트	과자	800	9
18	기다려 마트	냉동식품	6000	2
19	나다 마트	양념	3000	23
20	나다 마트	유제품	700	33
21	나다 마트	라면	500	26

제품별 수량

	A	B
3	행 레이블	합계 : 수량
4	통조림	180
5	차	152
6	잼	120
7	음료	109
8	유제품	314
9	양념	158
10	라면	153
11	냉동식품	320
12	과자	197
13	총합계	1703

제품의 거래처별 수량

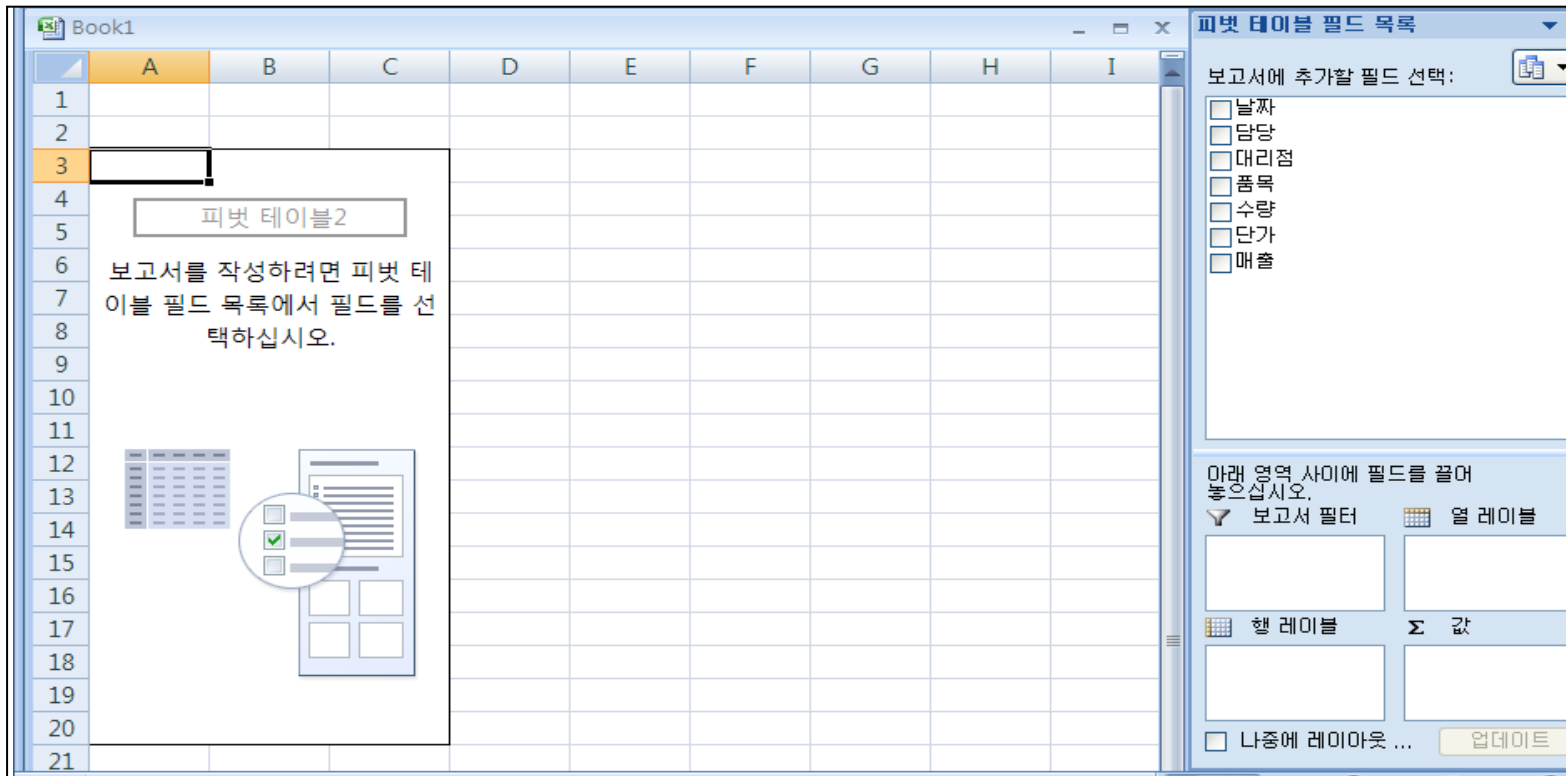
	A	B
3	행 레이블	합계 : 수량
4	통조림	180
5	A마트	10
6	기다려 마트	30
7	도도 마트	34
8	마성 백화점	40
9	왔다 마트	23
10	유대리 마트	22
11	진달래 마트	21
12	차	152
13	A마트	23
14	너도 마트	24
15	용산1가게	71
16	하나다마트	34
17	잼	120
18	나다 마트	10
19	마성 백화점	22
20	진달래 마트	35
21	풀 마트	21
22	하나다마트	32
23	음료	109
24	A마트	21
25	경호 다방	37
26	도도 마트	7
27	오마이 백화점	23

엑셀 피벗테이블 & 차트

❖ 피벗테이블(Pivot Table)

■ Step 1. [삽입]탭-[표]그룹-[피벗 테이블]

■ Step 2. 워크시트 왼쪽에는 피벗테이블 영역이 생기고, 오른쪽에는 피벗 테이블 필드 목록이 표시



엑셀 피벗테이블 & 차트

❖ 피벗테이블(Pivot Table)

■ Step 3. 필드 레이아웃 지정하기

	A	B	C	D	E	F	G	H
2								
3	합계 : 매출액	열 레이블						
4	행 레이블	광주	대구	대전	부산	서울	제주	총합계
5	과자	2646		11438	9522	12230		35836
6	냉동식품			4905	44156	50726	36752	136539
7	라면	31128	49145	8747			2181	91201
8	맥주				1427	49509	3344	54280
9	유제품		5302	50984	685			56971
10	음료				12965	19308	9624	41897
11	통조림	45885		63602	1274	68579	17543	196883
12	총합계	79659	54447	139676	70029	200352	69444	613607
13								
14								
15								
16								
17								
18								
19								

피벗 테이블 필드

보고서에 추가할 필드 선택:

- ☐ 날짜
- ☒ 제품
- ☒ 매출액
- ☒ 대리점

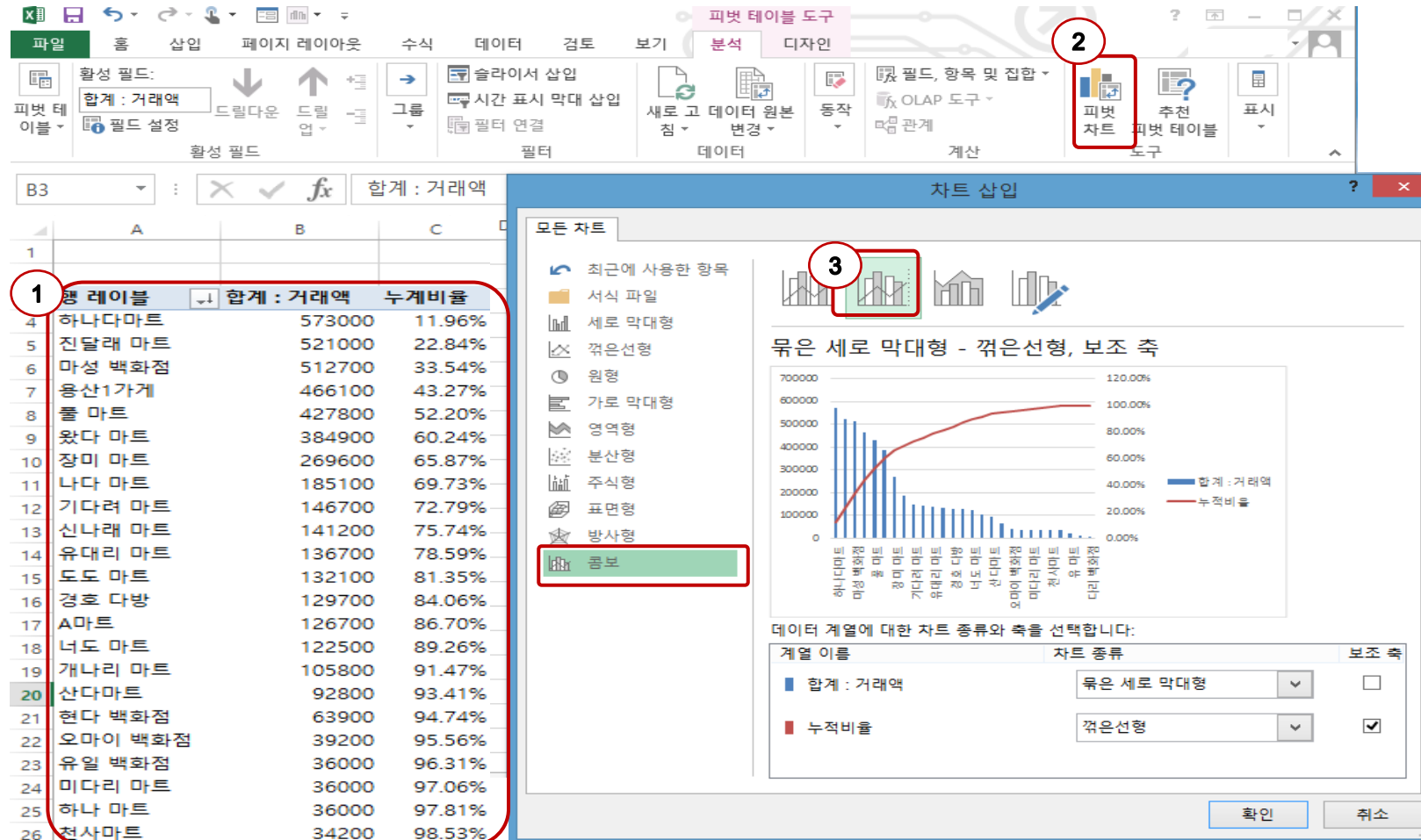
추가 테이블...

아래 영역 사이에 필드를 끌어 놓으십시오.

필터:
 열: 대리점

행: 제품
 값: 합계 : 매출액

■ 피벗테이블로 정리된 데이터를 차트로 나타낸 것



엑셀서식



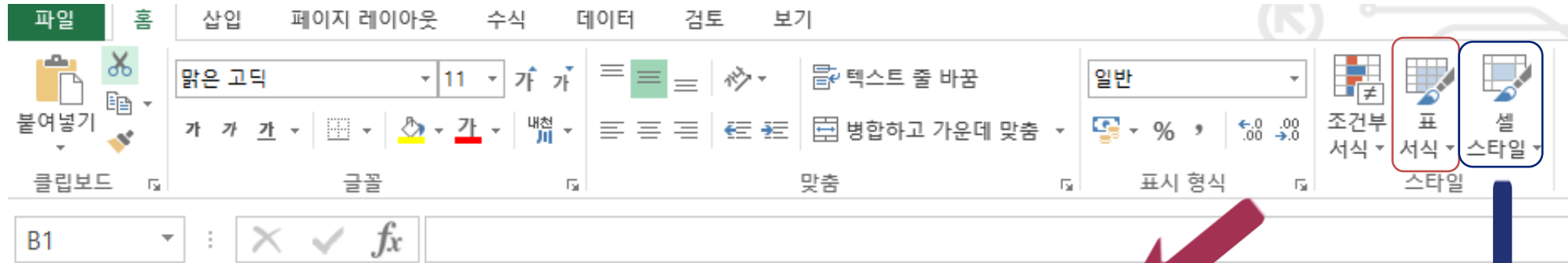
자동서식



조건부 서식

엑셀서식

❖ 자동 서식



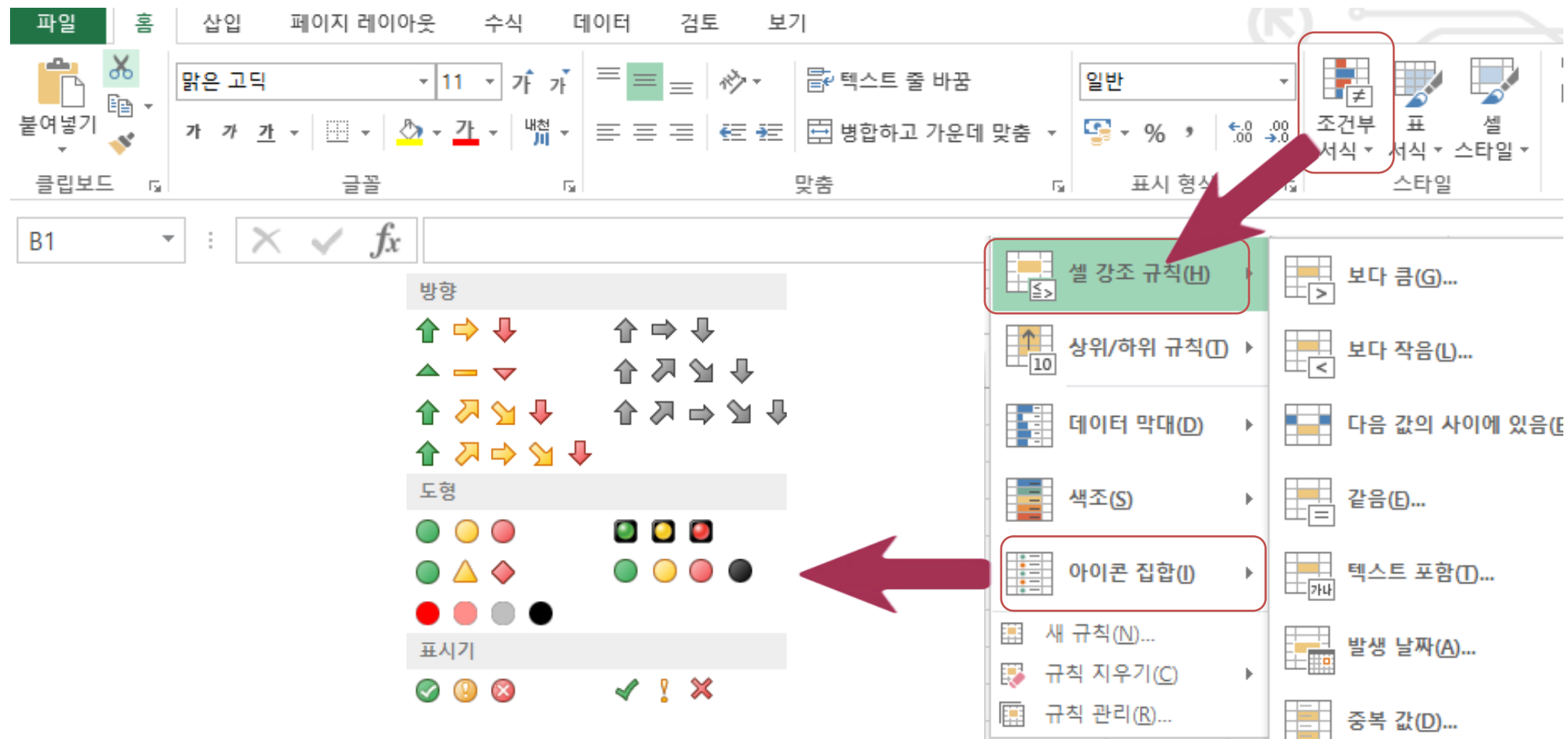
! 표 서식

! 셀 스타일



엑셀서식

❖ 조건부서식



타이타닉 데이터 파일 살펴보기

데이터 확인

- 각 항목의 의미 확인

	A	B	C	D	E	F	G	H	I	J	K
1	Passenger	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
2	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25	
3	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85
4	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2.	7.925	
5	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123
6	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05	S
7	6	0	3	Moran, Mr. James	male		0	0	330877	8.4583	Q
8	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46
9	8	0	3	Palsson, Master. Gosta Leonard	male	2	3				
10	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0				
11	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1				
12	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1				
13	12	1	1	Bonnell, Miss. Elizabeth	female	58	0				
14	13	0	3	Saunderscock, Mr. William Henry	male	20	0				
15	14	0	3	Andersson, Mr. Anders Johan	male	39	1				

이름	내용
PassengerId	승객 ID
Survived	생존여부(0: 사망, 1: 생존)
Pclass	탑승권 종류 (1: 1등석, 2: 2등석, 3: 3등석)
Name	이름
Sex	성별
Age	나이
SibSp	형제, 배우자를 합한 수
Parch	부모, 자녀를 합한 수
Ticket	탑승권 번호
Fare	탑승권 가격
Cabin	캐빈 번호
Embarked	탑승지 (C: 세르부르, Q: 퀸즈타운, S: 사우스햄프턴)

척도

명목척도

서열척도

등간척도

비율척도

척도

명목척도란?

- 관찰대상의 속성에 따라 관찰대상을 상호 배타적이고 포괄적인 범주로 구분하는 수치를 부여하는 도구
- 숫자에 일종의 상징적 의미를 부여하여 분류

척도



사례

Ex

성별	①남 ②여
직업	①전문직 ②사무직 ③노무직 ④주부 ⑤학생 ...
구매여부	①구매자 ②비구매자
지역	①서울 ②부산 ...
구매브랜드	①삼성 ②LG ③대우 ...

척도



식료품 쇼핑행사 조사

Ex

- 쇼핑장소 : ①~⑦까지 숫자 지정
 - ① 시내 주요지역의 백화점
 - ② 쇼핑센터
 - ③ E마트 등의 대형할인매장
 - ④ 주거지 주변의 일반 슈퍼
 - ⑤ 재래시장
 - ⑥ 인터넷 쇼핑
 - ⑦ 기타
- **Q) 1번 선택 - 우월?**

척도



박찬호 선수와 박지성 선수의 백넘버

Ex



박찬호 61번

-



박지성 13번

= 48

48은
아무런 의미가 없음

<출처> https://cdnimage.dailian.co.kr/news/201407/news_1405495936_448358_m_1.jpg (21.08.26)
<http://mimg.segye.com/content/image/2016/02/12/20160212001560.jpg> (21.08.26)

척도

서열척도란?

- 관찰대상이 가지고 있는 속성 크기에 따라 관찰대상의 순위를 나타내는 수치를 부여하는 측정도구
 - 상대적인 순위만 구분할 뿐, 서열 간의 차이는 중요하지 않음
 - 측정값들의 대표치: 중위수 (median)

척도

❖ 서열척도



사례 1. 도쿄 올림픽 양궁 순위

Ex



- ▶ 순위만 측정
- ▶ 등수 간의 차이는 알수 없음.

<출처> https://news.imaeil.com/inc/photos/2021/07/25/2021072517243425184_l.jpg (21.08.26)

척도

❖ 서열척도



사례 2. 직장 선택의 우선순위

Ex

- 귀하께서 직장을 선택하실 때 중요하다고 생각하는 직장의 특성을 보기에서 골라 순서대로 말씀해 주십시오.

첫째 () 둘째 () 셋째 () 넷째 ()

보기

① 안정성

② 발전성

③ 수익성

④ 장래성

척도

❖ 서열척도



사례 3. 백화점 선호도

Ex

명목척도(Nominal)		서열척도(Ordinal)
No	백화점	선호도순위(Ranking)
1	서울백화점	7
2	소공동백화점	2
3	퇴계로백화점	8
4	청량리백화점	3
5	영등포백화점	1
6	신촌백화점	5
7	강남백화점	9
8	반포백화점	6
9	역삼백화점	4
10	목동백화점	10

척도

등간척도란?

- 관찰대상이 가지고 있는 속성 크기의 **순서** 뿐 아니라 **상대적인 차이**도 고려하여 측정하는 도구
- 관찰대상이 가지고 있는 속성 크기의 양적 차이를 측정하기 위해서 등간격으로 분할한 척도를 사용함

척도

❖ 등간척도(interval scale)

마케팅 조사의 예

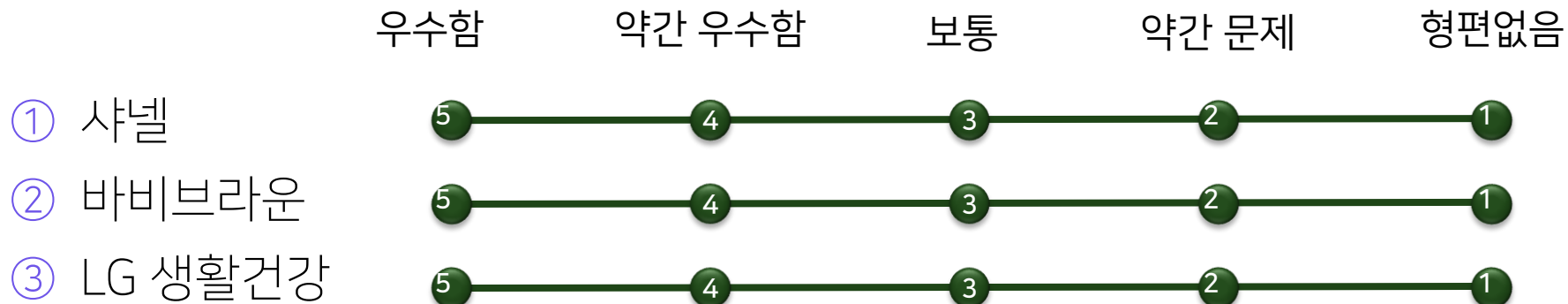
- 브랜드에 대한 태도, 제품속성별 소비자 지각(예: 품질수준), 구매 의도, 만족도



등간척도 사례 1

Ex

- 귀하는 다음의 화장품 브랜드들에 대해 어떻게 평가하는지 적절한 위치에 표시해 주세요.



척도



등간척도 사례 2

Ex

- 귀하께서 스마트폰을 사용하시면서 느끼시는 혜택에 관한 질문입니다.
1점 부터 5점 중에서 각 문항에 동의하시는 정도에 따라 점수를 주시면 됩니다.

	매우 중요하지 않다		보통이다		매우 중요하다
① 업무(일/학업)진행속도를 향상시킨다.	1		3		5
② 업무(일/학업)성과를 향상시킨다.	1		3		5
③ 업무(일/학업)의 생산성을 향상시킨다.	1		3		5
④ 업무(일/학업)의 효율성을 향상시킨다.	1		3		5
⑤ 업무(일/학업)에 유용하다.	1		3		5

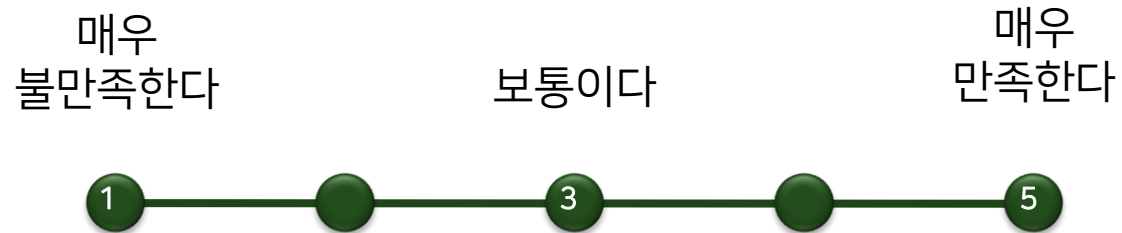
척도



등간척도 사례 3

Ex

- 귀하께서는 현재의 모바일 뱅킹 서비스에 대해서 어느 정도 만족하십니까? ()



척도

비율척도란?

- 등간척도에 비율의 개념이 추가된 척도
- 절대적 기준값이 존재하는 척도

Ex

— 몸무게, 키, 해수면을 기준으로 하는 산의 높이 등

척도

Ex

- 제품가격, 소득, 매출액, 구매 횟수, 비용, 시장 점유율 등



비율척도 사례 2

Ex

- 패밀리 레스토랑 이용에 관한 일반문항입니다.
 - (1) 귀하께서 지난 3개월간 월 평균 몇 번이나 패밀리 레스토랑(TGI, 베니건스, 마르쉐, 아웃백스테이크, 시즐러, VIPS, 코코스)을 이용하셨습니까?
한 달에 ()번
 - (2) 위 패밀리 레스토랑을 이용하실 때 1인당 기준으로
어느 정도의 비용을 지출하십니까?
1인당 ()만원

척도

명목척도		서열척도		등간척도	비율척도
분류	백화점	순위	선호점수	선호도 평가(7점척도 이용)	지난 3개월간 구입액
1	뉴코아	6	65	4	0
2	애경	7	88	3	0
3	롯데	2	20	7	30만 원
4	갤러리아	4	35	5	10만 원
5	신세계	1	10	7	35만 원
6	NC	5	50	5	5만 원
7	현대	3	27	6	20만 원

<출처> 마케팅조사원론, 안광호, 임병호, 창명

척도 정리

명목척도

- 관찰대상의 관심속성을 측정하여 그 값을 범주로 나타냄
- 모든 연산이 불가능함

Ex

— 남/녀 구분

서열척도

- 관찰대상의 관심속성을 측정하여 그 값을 순위로 나타냄
- 모든 연산이 불가능함

Ex

— 1등, 2등, 3등

등간척도

- 관찰대상의 속성값을 상대적 크기로 나타냄
- 사칙연산 중에 가감(+, -)만이 가능함

Ex

— 온도, 브랜드에 대한 태도, 제품속성별 소비자 지각, 구매 의도, 만족도

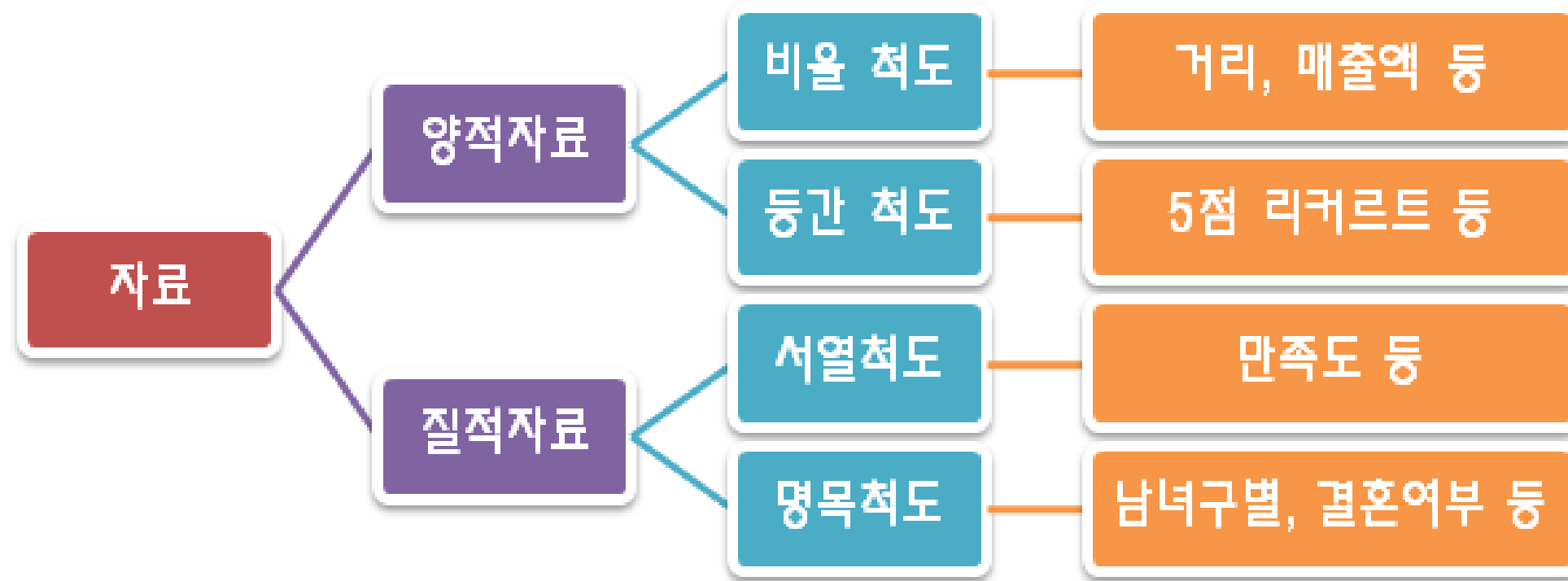
비율척도

- 절대적 기준인 영점이 존재하고
모든 사칙연산 (+, -, ×, ÷)이 가능함

Ex

— 키, 몸무게, 매출액, 횟수, 나이

자료(데이터)에 따른 척도의 종류



타이타닉 데이터 파일 살펴보기

❖ 데이터 파일 확인

■파일:

- Day1_train_학생용.xlsx

데이터 분석 프로세스별 실습

❖ 틀 고정 & 데이터 필터

- 보기>창>틀고정>첫행고정
- 데이터>정렬 및 필터>필터

❖ data set overview - 데이터 형태 파악

- 데이터 필터기능 이용- 필드값 파악
- 전체 데이터 수 및 변수 파악

엑셀 함수 및 기능

- count(): 숫자가 들어있는 셀의 개수
- counta(): 비어있지 않는 셀의 개수
- countblank() : 빈 칸 셀의 개수
- countif: 조건에 맞는 셀의 개수

데이터 분석 프로세스별 실습

❖ 데이터 전처리

- 결측치 파악 : countblank()
- 결측치 처리: mean, countif , if
 - 결측값 처리방법
 - ✓ 단순대치법(Single Imputation)
 - » 레코드 삭제
 - » 평균대치, 회귀대체 등
 - ✓ 다중대치법: 단순대치법을 여러 번하여 가상적 완전 자료를 만드는 방법
 - » Mice, Amelia , MissForest, Hmisc, Mi 등

엑셀 함수 및 기능

- count(): 숫자가 들어있는 셀의 개수
- counta(): 비어있지 않는 셀의 개수
- countblank() : 빈 칸 셀의 개수
- countif: 조건에 맞는 셀의 개수
- mean: 평균
- If: 조건문
- 피벗테이블, 피벗차트

데이터 분석 프로세스별 실습

- ❖ EDA - 피벗 테이블 및 피벗 차트 그리기
 - 생존 비율 파악하기 (값 필드 설정 – 값표시형식 – 열 합계 비율)
 - 여성 vs. 남성 생존 비율 파악하기
 - Pclass 등급별 생존률 파악하기
 - Pclass등급별 성별 생존률 파악하기
 - SibSp,Parch 등 생존률 파악하기
 - 등

데이터 분석 프로세스별 실습

❖ 가설 세우기

- 가설1) 남자이면 사망, 여자이면 생존
- 가설2) Pclass가 1등급이면 생존, 2등급이면 사망, 3등급이면 사망
- 가설3) 1등급&2등급인 여자는 생존, 3등급 여자는 사망, 모든 남자: 사망

❖ 예측하기

- If, AND, OR 등 사용

❖ 모델 예측력 파악

- test data의 정답률 파악

Competition 예측 시도

시도	시각화 분석	가설	참고
1	죽은사람: 62% 살은 사람: 38%	다 죽는다	All '0'
2	남자가 더 많이 죽었음.	남자이면 사망, 여자이면 생존	=if(Sex=0,0,1)
3	등급이 높은 곳은 생존률이 높다.	Pclass가 1등급이면 생존, 2등급이면 사망, 3등급이면 사망	=if(Pclass=1,1,0)
4	1등급: 여자생존률 매우 높음 2등급: 여자 생존률 비슷 3등급: 여자 생존률 낮음 남자: 전체적으로 생존률 낮음	남자는 모두 죽는다. 단, 여자는 1,2등급 경우 산다.	=if(Sex=0, 0, if(OR(Pclass=1,Pclass=2),1,0)

Wrap-Up

❖ Day 1 Wrap-up 설문

- <https://forms.gle/egHEEJdorWtAr5Wb8>

Day1. 교육 사전 설문조사

<https://forms.gle/fPHUUVRFY1vBq2EP8>