

01 Data ?!

Steps



First-Understand the Business Domain



Second-Understand the Business Problem



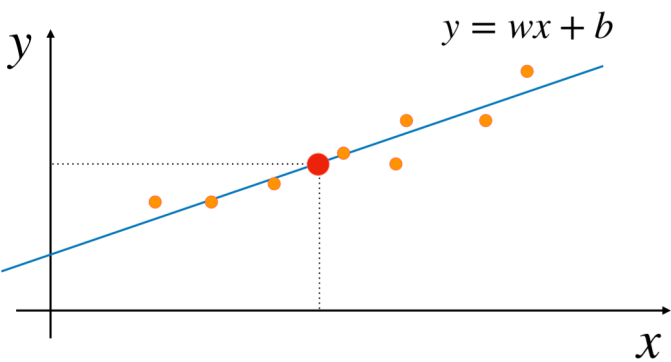
Third- What is the Right Data, Right Column and Right Algorithm



Last-Combine Knowledge With Machine Learning

Data

근속연수	연봉
1.5	3,100
2.5	3,900
4.2	4,300
5.1	4,900
6.7	5,400
8.3	6,700
9.5	9,200
13	12,900

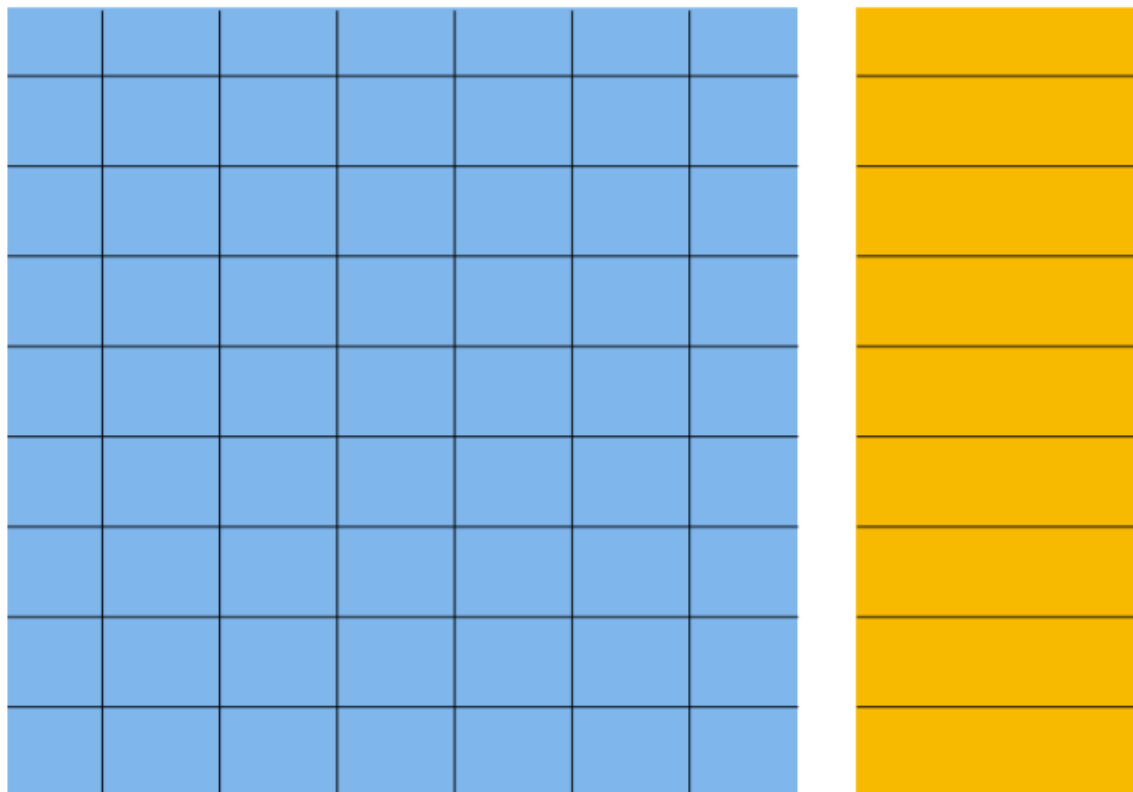


Feature

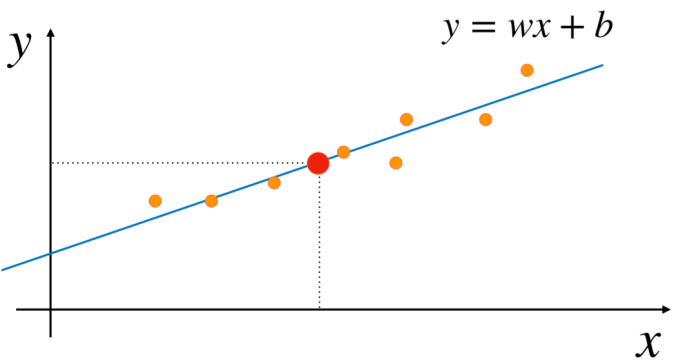
Label

이름	나이	성별	선실 등급	티켓번호	티켓 요금	부모 자식	형제 자매	키	출항지	생존 여부
Rose	25	여	A	EA-1039	300	2명	1명	167	런던	Y
Jack	20	남	C	GE-3059	29	-	-	178	도버	N
Mark	57	남	B	BA-2031	89	4명	3명	167	뉴포트	N
Andy	48	남	B	NN-3928	102	5명	7명	182	런던	Y

Hello AI

$$X = \text{Feature}$$
$$\mathcal{Y} = \text{Label}$$


근속연수(x)	연봉(y)
1.5	3,100
2.5	3,900
4.2	4,300
5.1	4,900
6.7	5,400
8.3	6,700
9.5	9,200
13	12,900



Hello AI

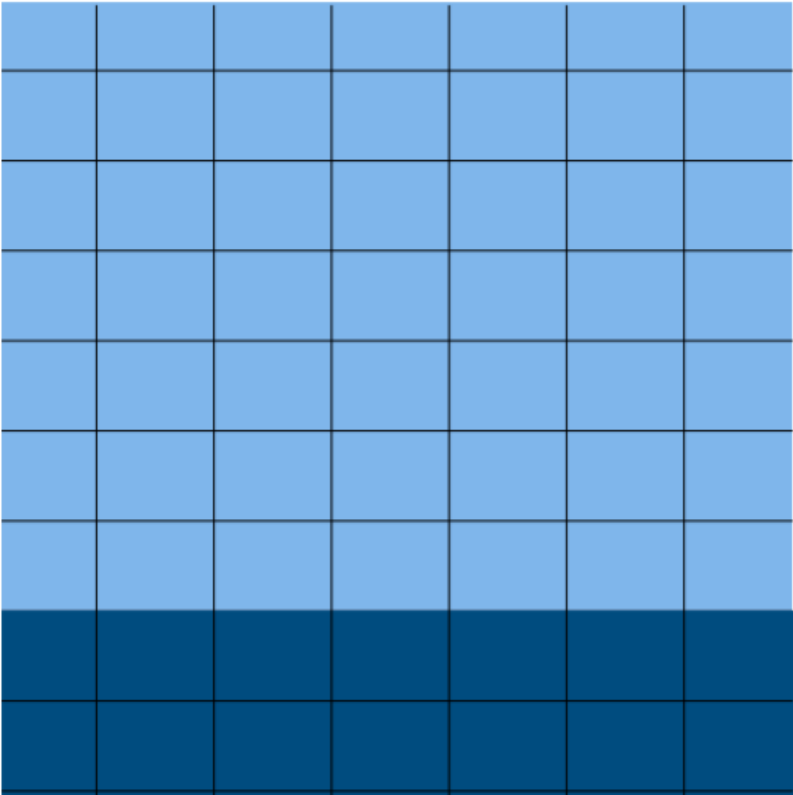
$$X = \text{Feature}$$
$$\mathcal{Y} = \text{Label}$$

The diagram illustrates the number 70 as a collection of 10 groups of 7. On the left, there is a 10x7 grid of blue squares, totaling 70 squares. On the right, there is a single vertical column of 10 yellow squares, also totaling 10 squares. This visual representation shows that 70 can be thought of as 10 groups of 7, which is the same as 10 groups of 10, leading to the multiplication fact $10 \times 10 = 100$.

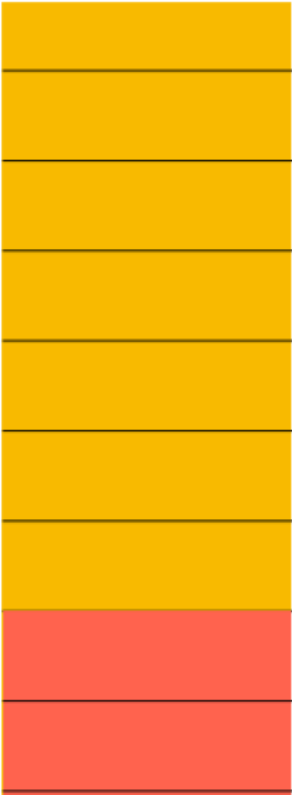
\mathcal{X} = Feature

\mathcal{Y} = Label

X_{train}



X_{test}



Y_{train}

Y_{test}

Random Split

 $\mathcal{X} = \text{Feature}$ $\mathcal{Y} = \text{Label}$

True
True
True
True
True
False
False
False
False

<i>X_train</i>							True	<i>Y_train</i>
							True	
							True	
							True	
							True	
							False	
							False	
<i>X_test</i>							False	<i>Y_test</i>
							False	

Random Split

X_{train}							True	Y_{train}
							True	
							True	
							True	
							True	
X_{test}							False	Y_{test}
							False	
							False	
							False	
							False	

02 Open Data

kaggleTM



Competitions

Documentation

InClass

General

InClass

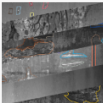
Sort by

Grouped

All Categories


Search competitions

19 Active Competitions




Severstal: Steel Defect Detection
Can you detect and classify defects in steel?
Featured · Code Competition · a month to go · manufacturing, image data

\$120,000
1,665 teams




Lyft 3D Object Detection for Autonomous Vehicles
Can you advance the state of the art in 3D object detection?
Featured · 2 months to go · image data, object detection

\$25,000
123 teams



RSNA Intracranial Hemorrhage Detection
Identify acute intracranial hemorrhage and its subtypes
Featured · 2 months to go · health foundations and medical research, image data

\$25,000
399 teams



The 3rd YouTube-8M Video Understanding Challenge
Temporal localization of topics within video
Research · 16 days to go · video data, object detection

\$25,000
263 teams

국가통계포털이란?

- 국가통계포털(KOSIS, Korean Statistical Information Service)은 국내·국제·북한의 주요 통계를 한 곳에 모아 이용자가 원하는 통계를 한 번에 찾을 수 있도록 통계청이 제공하는 One-Stop 통계 서비스입니다.
- 현재 300여 개 기관이 작성하는 경제·사회·환경에 관한 1,000여 종의 국가 승인통계를 수록하고 있으며, 국제금융·경제에 관한 IMF, Worldbank, OECD 등의 최신 통계도 제공하고 있습니다.
- 쉽고 편리한 검색기능, 일반인들도 쉽게 이해할 수 있는 다양한 콘텐츠 및 통계설명자료 서비스를 통해 이용자가 원하는 통계자료를 쉽고 빠르고 정확하게 찾아보실 수 있습니다.

국가통계포털에서 제공하는 데이터

■ KOSIS가 제공하는 주요 서비스



출처 <https://kosis.kr/serviceInfo/kosisIntroduce.do>

- 국가통계포털에서는 카테고리별로 다양한 데이터를 제공하고 있고 비교적 업데이트 주기도 빠른 편이기 때문에 실질적인 데이터를 구하는데 도움이 된다.

차트 분석

출생아수, 합계출산율, 자연증가 등

통계설명자료 | 온라인간행물 | 보도자료 | 관련통계표

출처 : 통계청, 「인구동향조사」 [더보기](#)

자료갱신일: 2021-09-28 / 수록기간: 년 1970 ~ 2020 / 자료문의처 : 042-481-2258 / 기능문의: KOSIS Q&A 게시판

일괄설정 + 항목[1/1] 기본항목별[6/6] 시점[51/51]

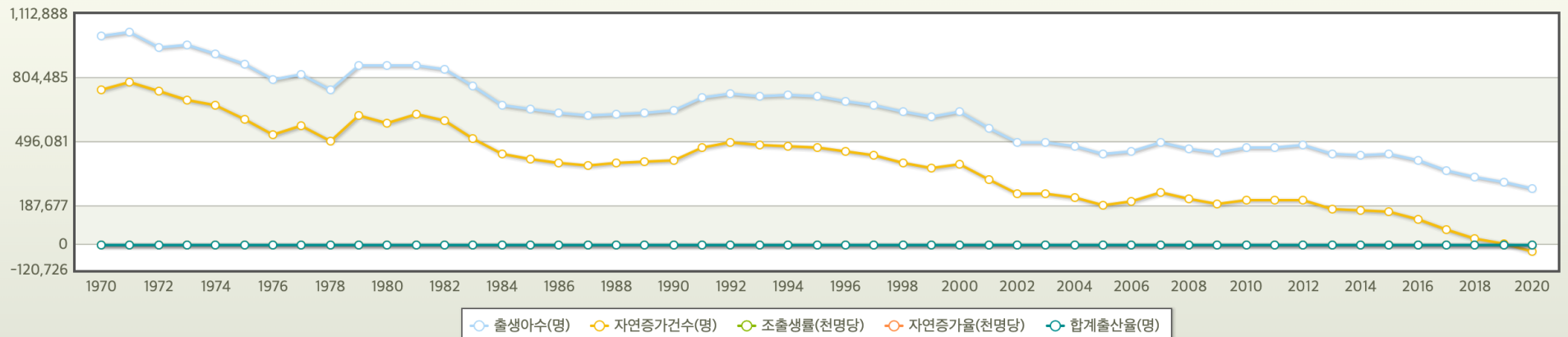
차트조회

새창보기 주석정보 주주정보 행렬전환 분석 차트 부가기능설정 스크랩 다운로드 인쇄 도움말



차트저장 차트인쇄 5 닫기

출생아수, 합계출산율, 자연증가 등



※분류명 : 기본항목별

* 차트의 범례는 최대 20개까지 제공되며, 항목과 분류값을 재설정하여 조회할 수 있습니다.

* 차트 기능은 구글 크롬 및 마이크로소프트 익스플로러11에 최적화 되어 있습니다.

감사합니다.