

# Kaggle 사이트

캐글(Kaggle)은 2010년 설립된 예측모델 및 분석 대회 플랫폼이며, 기업 및 단체에서 과제와 함께 데이터를 공유하면, 데이터 과학자들이 이를 해결하는 모델을 개발하기 위해 경쟁하는 구조임.

캐글은 2017년 3월 구글에 인수되었으며, 상금을 놓고 경쟁하는 대회이지만 데이터분석/딥러닝 학습을 위해 필요한 데이터를 구할수 있는 좋은 사이트임

[데이터 전처리 예제를 위한 Titanic 데이터 분석]

[www.kaggle.com/c/titanic](http://www.kaggle.com/c/titanic) 접속하여 총 3개의 데이터 csv 파일을 다운로드 함  
(train.csv, test.csv, gender\_submission.csv)

train.csv 일부

	A	B	C	D	E	F	G	H	I	J	K	L
1	Passenger	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	1	0	3	Braund, M	male	22	1	0	A/5 21171	7.25		S
3	2	1	1	Cumings, female	female	38	1	0	PC 17599	71.2833	C85	C
4	3	1	3	Heikkinen	female	26	0	0	STON/O2.	7.925		S
5	4	1	1	Futrelle, M	female	35	1	0	113803	53.1	C123	S
6	5	0	3	Allen, Mr.	male	35	0	0	373450	8.05		S
7	6	0	3	Moran, M	male		0	0	330877	8.4583		Q
8	7	0	1	McCarthy, male	male	54	0	0	17463	51.8625	E46	S
9	8	0	3	Palsson, M	male	2	3	1	349909	21.075		S
10	9	1	3	Johnson, M	female	27	0	2	347742	11.1333		S
11	10	1	2	Nasser, M	female	14	1	0	237736	30.0708		C
12	11	1	3	Sandstrom	female	4	1	1	PP 9549	16.7	G6	S
13	12	1	1	Bonnell, M	female	58	0	0	113783	26.55	C103	S
14	13	0	3	Saunders	male	20	0	0	A/5. 2151	8.05		S
15	14	0	3	Andersson	male	39	1	5	347082	31.275		S