

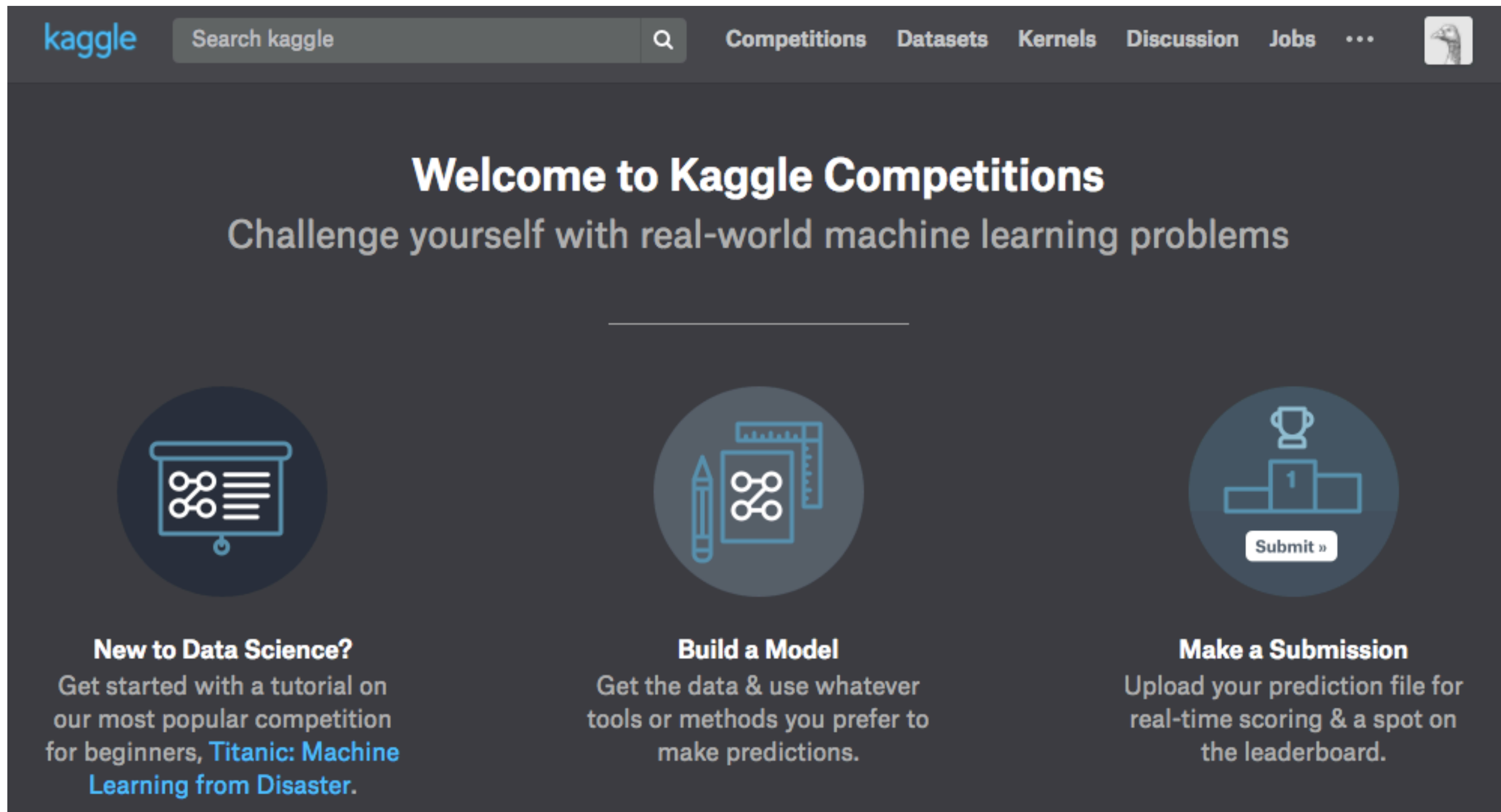
Kaggle 介紹

Introduction to Kaggle



Kaggle 簡介

- Kaggle是目前全球最大的資料科學競賽平台，除了有豐富的資料集外，還有完整的評分與排名競賽系統，提供企業、研究單位等以群眾外包(crowdsourcing)的模式解決建模預測問題。

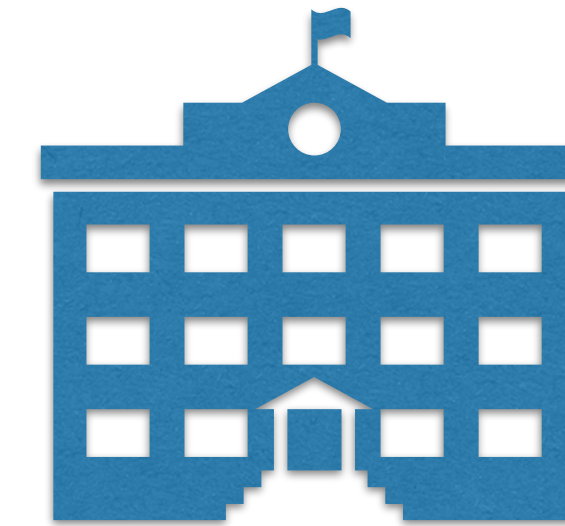


Why it works?



資料科學人才：

擁有技術，但缺乏數據與實戰機會。



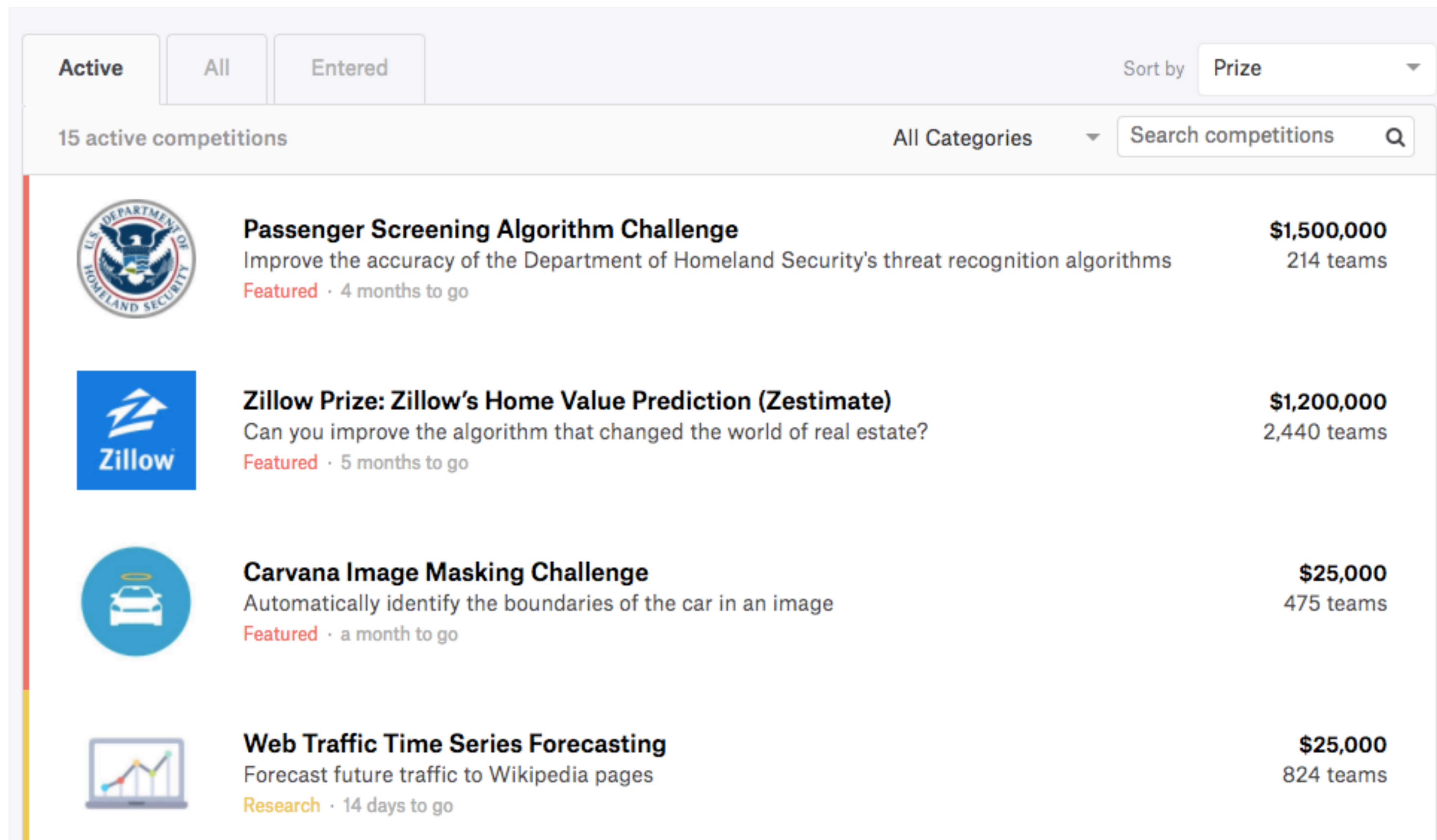
企業/研究單位：

擁有數據，但若僅依賴內部研發人員分析建模耗費時間且人力成本高。





- 天下沒有白吃的午餐(No free lunch) 定理
 - 沒有一種建模方法適用於所有問題
 - 問題可能的策略和解法很多

Competitions

- 競賽資訊



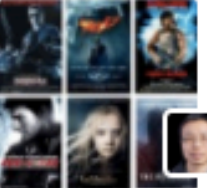




The screenshot displays a web interface for a competition platform. At the top, there are tabs for 'Active', 'All', and 'Entered', with 'Active' selected. To the right, a 'Sort by' dropdown menu is set to 'Prize'. Below the tabs, it indicates '15 active competitions'. A search bar labeled 'Search competitions' is also present. The main content area lists four competitions, each with a logo, title, description, status, and prize information.

Logo	Competition Name	Description	Status & Time	Prize	Teams
	Passenger Screening Algorithm Challenge	Improve the accuracy of the Department of Homeland Security's threat recognition algorithms	Featured · 4 months to go	\$1,500,000	214 teams
	Zillow Prize: Zillow's Home Value Prediction (Zestimate)	Can you improve the algorithm that changed the world of real estate?	Featured · 5 months to go	\$1,200,000	2,440 teams
	Carvana Image Masking Challenge	Automatically identify the boundaries of the car in an image	Featured · a month to go	\$25,000	475 teams
	Web Traffic Time Series Forecasting	Forecast future traffic to Wikipedia pages	Research · 14 days to go	\$25,000	824 teams




Datasets

- 可用資料集

974 featured datasets		Sort by	Most Votes
Featured	All	Mine	Upvoted
Search datasets			
696		IMDB 5000 Movie Dataset 5000+ movie data scraped from IMDB website chuansun76 · updated a year ago · film	43,286 downloads 79 comments
610		European Soccer Database 25k+ matches, players & teams attributes for European Professional Football Hugo Mathien · updated 10 months ago · association football, europe	31,873 downloads 94 comments
605		Credit Card Fraud Detection Anonymized credit card transactions labeled as fraudulent or genuine Andrea · updated 10 months ago · crime, finance	31,010 downloads 64 comments
531		Human Resources Analytics Why are our best and most experienced employees leaving prematurely? ludoben · updated 9 months ago · employment	29,523 downloads 90 comments
373		Iris Species Classify iris plants into three species in this classic dataset UCI Machine Learning · updated a year ago · botany	16,006 downloads 89 comments



Join a competition



House Prices: Advanced Regression Techniques

Predict sales prices and practice feature engineering, RFs, and gradient boosting

1,617 teams · 3 years to go

[Overview](#) [Data](#) [Kernels](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Join Competition](#)


Overview

[Description](#)
[Evaluation](#)
[Frequently Asked Questions](#)
[Tutorials](#)

Start here if...

You have some experience with R or Python and machine learning basics. This is a perfect competition for data science students who have completed an online course in machine learning and are looking to expand their skill set before trying a featured competition.


Competition Description



Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this playground competition's dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket



Kernels



House Prices: Advanced Regression Techniques



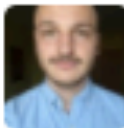





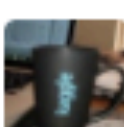

Predict sales prices and practice feature engineering, RFs, and gradient boosting

1,617 teams · 3 years to go

[Overview](#) [Data](#) [Kernels](#) [Discussion](#) [Leaderboard](#) [Rules](#) [New Kernel](#)

Sort by Most Votes

[All](#) [Mine](#) All Languages All Output Types

486		 Comprehensive data exploration with Python run 3 days ago by Pedro Marcelino	Py	190
426		 Regularized Linear Models run 4 months ago by Alexandru Papiu	Py	170
164		 Detailed Exploratory Data Analysis using R run 8 months ago by AiO (+334 / -613)	R	70
128		 Detailed Data Exploration in Python run 4 months ago by Angela	Py	46
120		 Ensemble Model: Stacked Model Example run a year ago by JMT5802	R	60

- 實作程式分享討論

Notes

建議初學者可先在此多觀摩他人 Jupyter Notebook，學習他人如何實作解決問題。