



Introduction to Kaggle



**Great discoveries
and improvements invariably
involve the **cooperation**
of many minds.**

— *Alexander Graham Bell*



Hello!

I am Agil Haykal



I am a Data expert with extensive experience in multiple industries such as marketplace, insurance, banking, general taxation, consulting, and training.

In total, I trained more than 300 data scientists, engineers, and analysts.



Table of Content

What will We Learn Today?

1. Why do we need kaggle?
2. History of kaggle
3. What is kaggle?
4. kaggle services
5. Kaggle rankings and progression system
6. Kaggle competition
7. Kaggle notebook vs Google Colab
8. Tips on using kaggle services

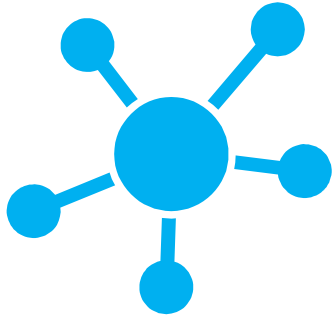


kaggle™

kaggle is the **world's largest data science community** with powerful tools and resources to help you achieve your data science goals.



Why do we need Kaggle?



The biggest data science community



Provide data science projects and competitions



Platform discussion from various backgrounds and industries



History of Kaggle

- Founded in April, 2010 by **Anthony Goldbloom** and **Ben Hamner**
- In the early establishment, just used as the **platform for data science competitions**
- On 8 March 2017, **Google** announced that they were acquiring Kaggle.
- In June 2017, Kaggle announced that it passed **1 million registered users**, or kagglers. The community spans **194 countries**. It is a diverse community, ranging from those just starting out and many of the world's best-known researchers.

Anthony Goldbloom



Ben Hamner

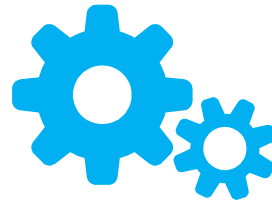


What is Kaggle?

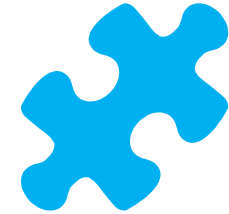
kaggle, a subsidiary of Google LLC, is an online community of data scientist and machine learning practitioners. It allows users to:



Find and publish
datasets



Explore and build
models



Work
collaboratively



Join competitions



Participate in courses



Find & Publish Datasets

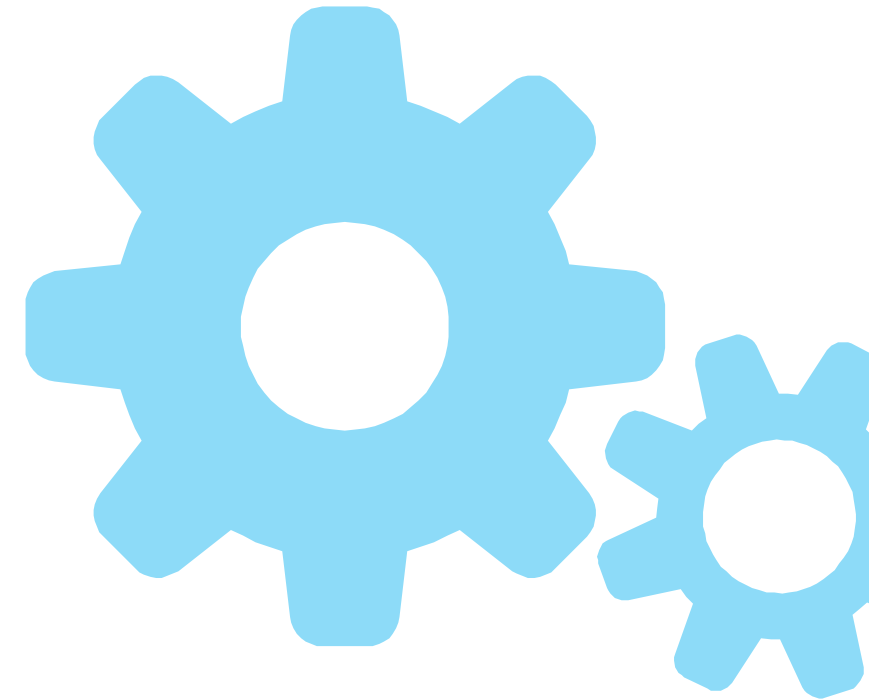
- Kagglers **share datasets** with each others
- Published datasets are on **various kinds and industries**





Kaggle Kernels

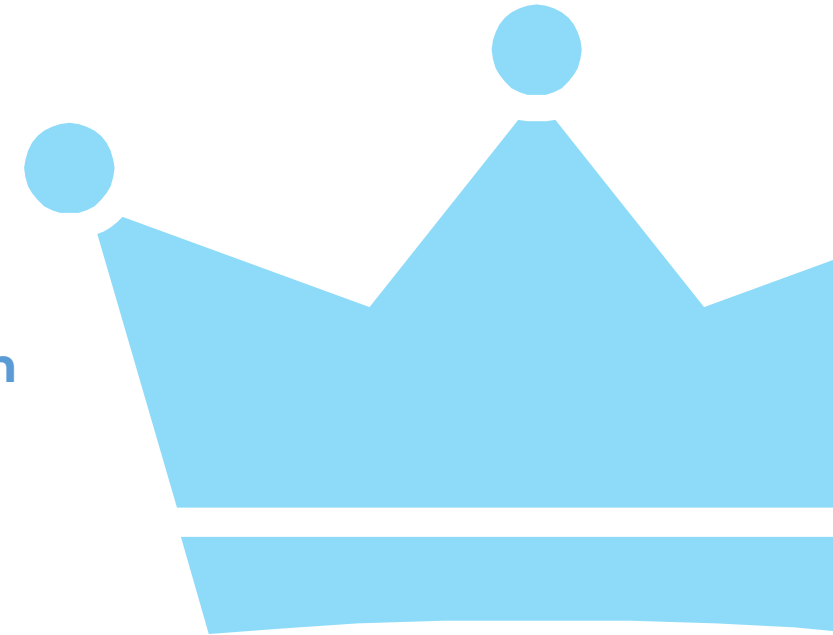
- A **cloud-based workbench** for data science and machine learning
- Allows data scientists to **share code and analysis** in Python, R and R Markdown
- This service provide Kagglers to **explore and build models in collaborative way**





Join Competitions

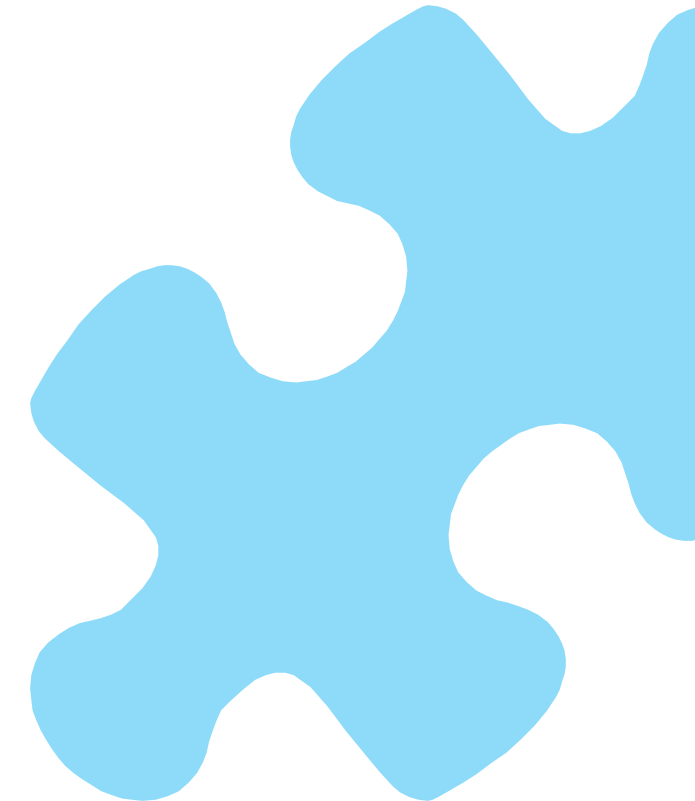
- This is kaggle's **first product**
- Companies **post problems** and machine learners compete to build the best algorithm, typically with **cash prizes**
- Companies can get the **best model by spending enough money and fast duration**
- Novice Kagglers can use this service as the medium for **data science practical learning**





Discussion Communities

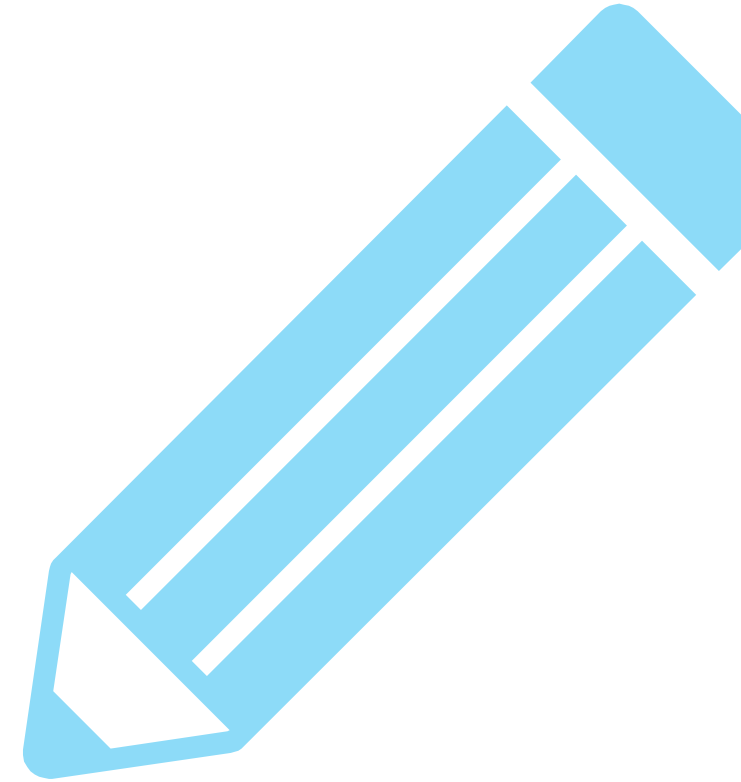
- Communities are a new place on Kaggle to **discover content and engage in discussion** around topics
- Communities including beginner users to TensorFlow, Pytorch topics





Participate in Courses

- Additional contents to **learn machine learning** and **AI**
- The courses are **free** and provide **certificates** to be earned
- Including Python, Machine Learning, Data Cleaning, Data Visualization, Pandas, SQL, Computer Vision, NLP, Geospatial Analysis, etc.





Kaggle Rankings and Progression System

User rankings are separately displayed within competitions, datasets, notebooks, and discussion


Competitions

Datasets


Notebooks

Discussion


Learn more about rankings >




234
Grandmasters




1,635
Masters






































7,015
Experts



63,435
Contributors



91,751
Novices

Rank	Tier	User		Medals	Points
1		 Psi	joined 10 years ago	 22  6  0	233,465
2		 Dieter	joined 4 years ago	 23  8  3	227,316
3		 Guanshuo Xu	joined 6 years ago	 21  18  2	200,284
4		 bestfitting	joined 5 years ago	 31  10  1	190,320
5		 llu	joined 3 years ago	 11  4  3	144,299
6		 Μαριος Μιχαηλιδης KazAnova	joined 8 years ago	 41  57  42	136,376
7		 DungNB	joined 4 years ago	 6  11  4	123,055



Kaggle Rankings for Personal Branding

People can use it as personal branding at LinkedIn or Resume

People



Siti Khotijah • 2nd

Artificial Intelligence and Data Science Enthusiast | Machine Learning | Kaggle Master Bandung

Current: **Kaggle** Competitions Expert at **Kaggle**

👤 Wiama Daya, Fajar Jaman, and 30 other shared connections



Ashok Lathwal • 2nd

Data Scientist @ Cisco | Kaggle Master | Google Cloud & AWS ML Certified Bangalore Urban

Current: **Kaggle Master** at **Kaggle**

👤 Lokesh Nandula, PMP, JP Bray, and 19 other shared connections



Andika Rachman • 2nd


Machine Learning Engineer | Kaggle Master Indonesia

👤 Tommy Martin, Kevin Aluwi, and 86 other shared connections




Kaggle Competition


Challenges are not just published with the prize. They are so many challenges with **swag or new knowledge**

 Featured Code Competition

NFL Health & Safety - Helmet Assignment

Segment and label helmets in video footage

 The National Football League · 563 teams · a month to go (a month to go until merger deadline)



\$100,000

Prize Money

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#)

[Join Competition](#) ...

Overview

Description

Evaluation

Timeline

Prizes

Code Requirements

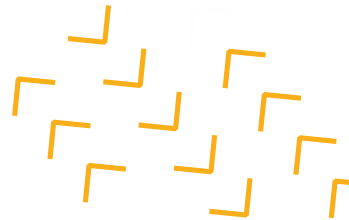
The National Football League (NFL) and Amazon Web Services (AWS) are teaming up to develop the best sports injury surveillance and mitigation program. In previous competitions, Kaggle has helped detect helmet impacts. As a next step, the NFL wants to assign specific players to each helmet, which would help accurately identify each player's "exposures" throughout a football play.





Impact of Kaggle Competitions

- Improving gesture recognition for Microsoft Kinect
- Improving the search for the Higgs boson at CERN
- COVID-19 case visualization in India
- COVID-19 journal texts exploration
- Shopee – Price Match Guarantee
- Shopee Code League
- BRI Data Hackathon



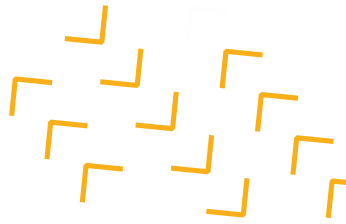


Kaggle notebook vs Google Colab

kaggle™

colab

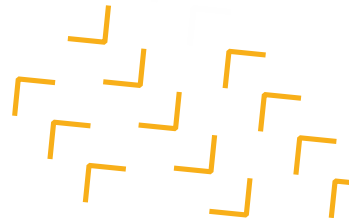
Both owned by Google

- Kaggle connected their dataset
 - CPU Memory 16GB
 - GPU RAM 17GB
 - Max Session 9 hrs
 - Kaggle Kernels often seem a little laggy.
- colab connected to gdrive
 - CPU Memory 14GB
 - GPU RAM 12GB
 - Max Session 12 hrs
 - Authenticate every session
 - Can't unzip files in Drive easily
- 

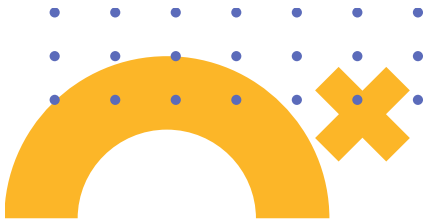


Tips on Using Kaggle Services

- Learn data science basic skills through the courses
- Explore various datasets in any types and business cases
- Understand and dive in the public notebooks in the framework of data science methodology
- Be curious! Join the provided discussion forums as you are interest in
- Participate in competitions starting from the easiest challenge
- Do not be inferior and feel overwhelmed!



Let's Explore our Dataset!



Let's Play

Kahoot!

www.kahoot.it



**Thank
YOU**

