

はじめてのAIプログラム学習キット5 Kaggle編

kaggle



スペクトラム・テクノロジー株式会社

<https://spectrum-tech.co.jp>

sales1@spectrum-tech.co.jp

学習キット 目次

	ページ
• 利用にあたっての注意事項	<u>3</u>
• kaggle	
1. 概要	<u>4</u>
2. 設定	
① 新規登録	<u>5</u>
② プロフィール・アカウント作成	<u>6</u>
3. メニュー説明	
① ホーム	<u>8</u>
② コンペ	<u>9</u>
③ データセット	<u>10</u>
④ コード	<u>11</u>
⑤ 議論	<u>12</u>
⑥ コース	
4. Code使用方法	
① Notebook作成	<u>13</u>
② Notebook使用例	<u>15</u>
5. Competition使用方法	<u>19</u>
6. dataset使用方法	<u>22</u>
7. Discussions使用方法	<u>24</u>
8. 事例	<u>25</u>
9. コンペに出展	<u>27</u>

利用にあたっての注意事項

① 記載内容

- Kaggleについて、弊社独自で翻訳、解釈しているため、間違いがあるかもしれません。正しい情報については、kaggle.comを参照ください。

② 免責事項

- 本マニュアルの操作により、重大な故障等が発生した場合においても、弊社は責任を負いません。内容を利用者が把握して操作するものとします。
- 本マニュアルの操作により、個人情報流出した場合、利用者の責によるものとします。

③ 問い合わせ

- 本マニュアルにおける、問い合わせについては、お受けできません。各利用者が調査して対応してください。但し、有料サポートを利用の場合は、この限りではありません。

④ 弊社AIプログラム学習キット5紹介

- https://spectrum-tech.co.jp/products/ai_tensorflow.html
- Pi版、グラフィック版など各種あります。事例は、その一部のものであります。本格的python, tensorflow, pytorchなど学習したい場合は、是非ご利用ください。

kaggle

1. 概要

① 概要

- Kaggle(カグル)は企業や研究者がデータを投稿し、世界中の統計家やデータ分析家がその最適モデルを競い合う、予測モデリング及び分析手法関連プラットフォーム及びその運営会社である。2017年googleが買収

② 利用方法

- Github: <https://github.com/Kaggle/kaggle-api>
- 公式サイトにアクセスして、Webで利用。使い方はjupyter notebookと同じ。また、GPU(時間制限あり)が準備されている。RTX2080よりは遅い。
- 英語環境のみ
- また、テストデータをダウンロードしてローカル環境でも動作

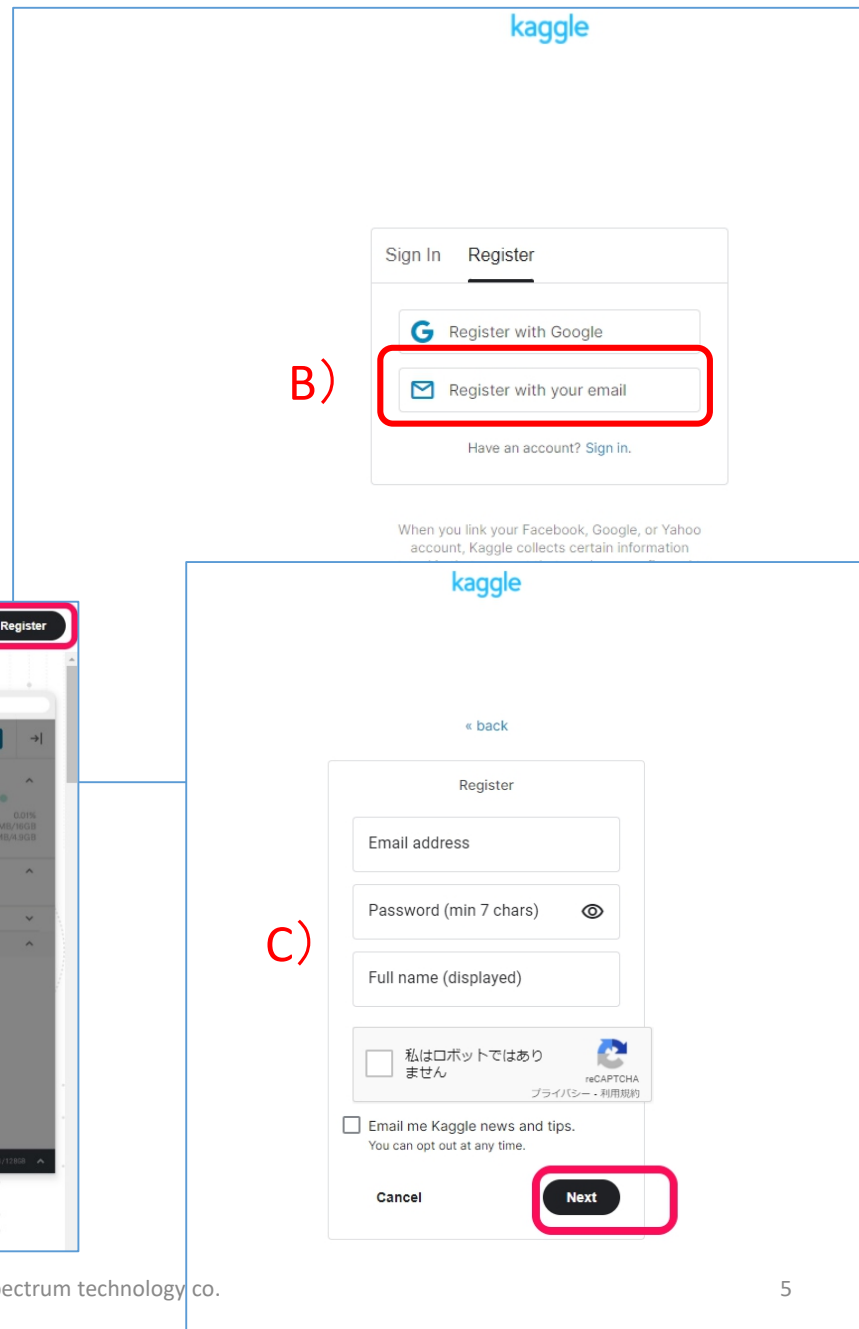
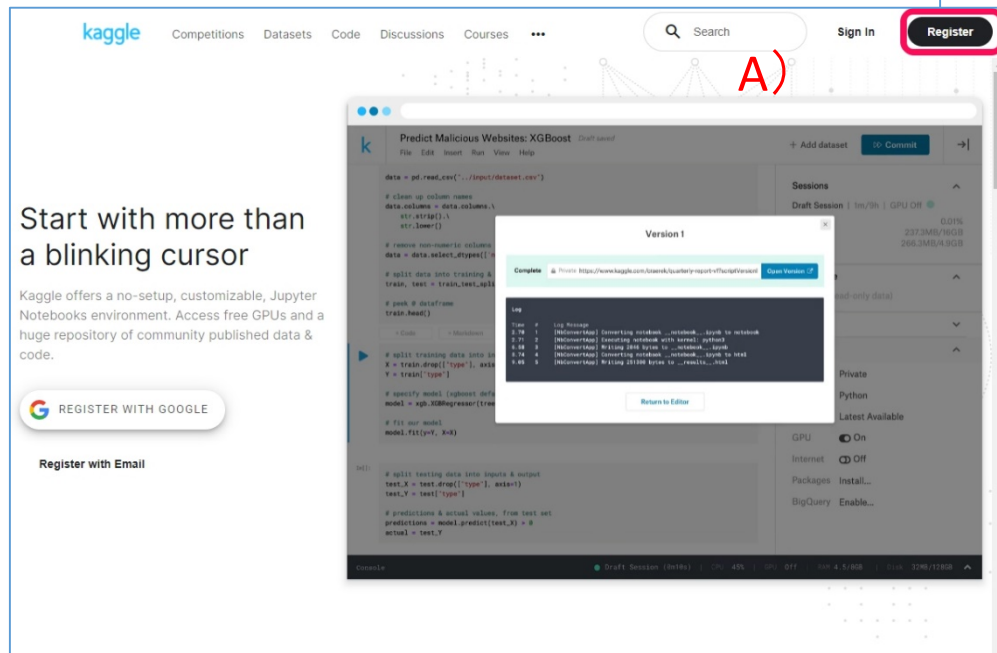
③ 必要なスキル

- Pythonのプログラムを作成できるスキル(中級)
- TensorflowなどのML(Machine learning)の基礎的なスキル(初級)
- 技術英語を読み書きできるスキル(翻訳ツールの活用)

kaggle 2. 設定

① 新規登録

- A) Register: アカウント登録
- B) Id: メール
- C) パスワード、ユーザ名
- D) メール認証、完了(写真省略)



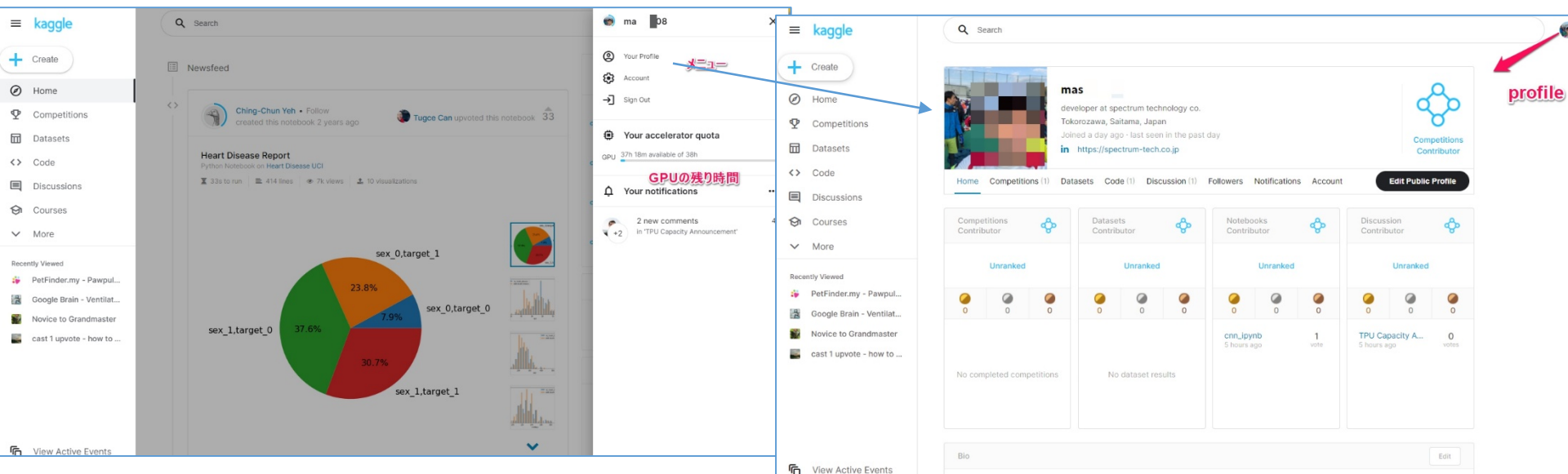
kaggle 2. 設定

② プロフィール・アカウント作成

A) プロフィールなどのメニュー

B) プロフィール

- 写真、会社、住所、ホームページなど公開してもよい情報を記入



kaggle

2. 設定

② プロフィール・アカウント作成

A) プロフィールなどのメニュー

B) プロフィール

- 写真、会社、住所、ホームページなど公開してもよい情報を記入

C) アカウント

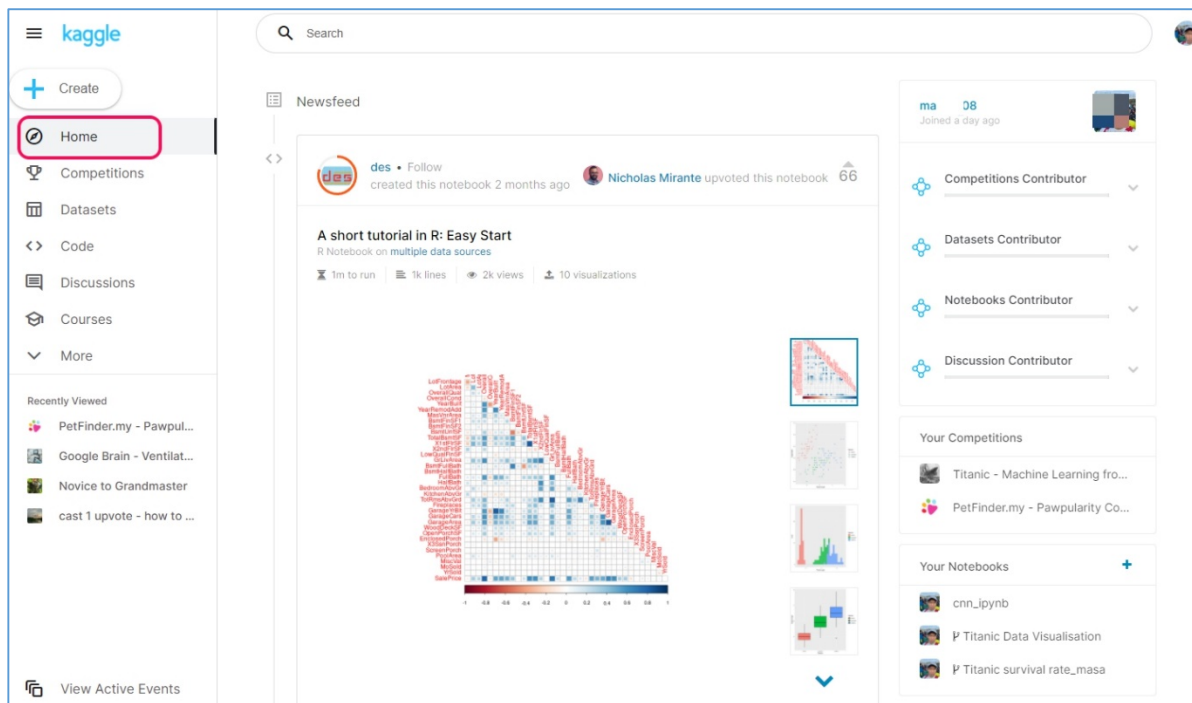
- GPUを利用するためのSMS認証があります。

The screenshot displays the Kaggle user interface for a user named 'ma 08'. The left sidebar contains a menu with options: 'Your Profile', 'Account', 'Sign Out', 'Your accelerator quota', and 'Your notifications'. A red label 'メニュー' (Menu) points to the 'Your Profile' option. The main content area shows the user's profile with a red label 'account' pointing to the 'Edit Public Profile' button. The 'Your accelerator quota' section shows 'GPU 37h 18m available of 38h' with a red label 'GPUの残り時間' (GPU remaining time). The 'Account' section shows 'Phone Verification Verified' with a red label 'gpu使用する場合にSMS認証があり。' (SMS authentication is required when using GPU). A red arrow points from the 'account' label to the 'Edit Public Profile' button.

kaggle

3. メニュー説明

- ① ホーム
 - 右側に個人のメニュー
- ② コンペ
- ③ データセット
- ④ Code
- ⑤ 議論
- ⑥ コース



kaggle

3. メニュー説明

① ホーム

② コンペ

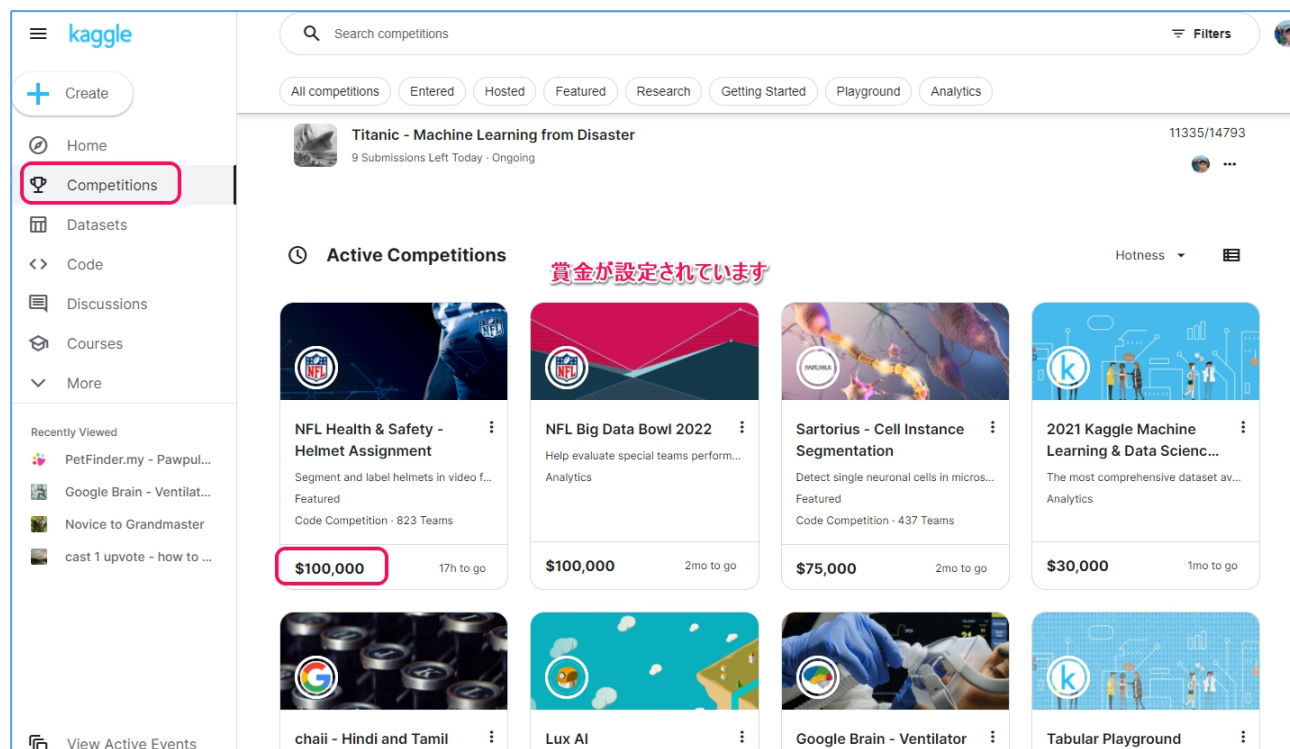
- 企業から賞金付きの課題が出されて、正確さなどでランク付けされます。モデルの作成がカギを握ります。

③ データセット

④ Code

⑤ 議論

⑥ コース



kaggle

3. メニュー説明

① ホーム

② コンペ

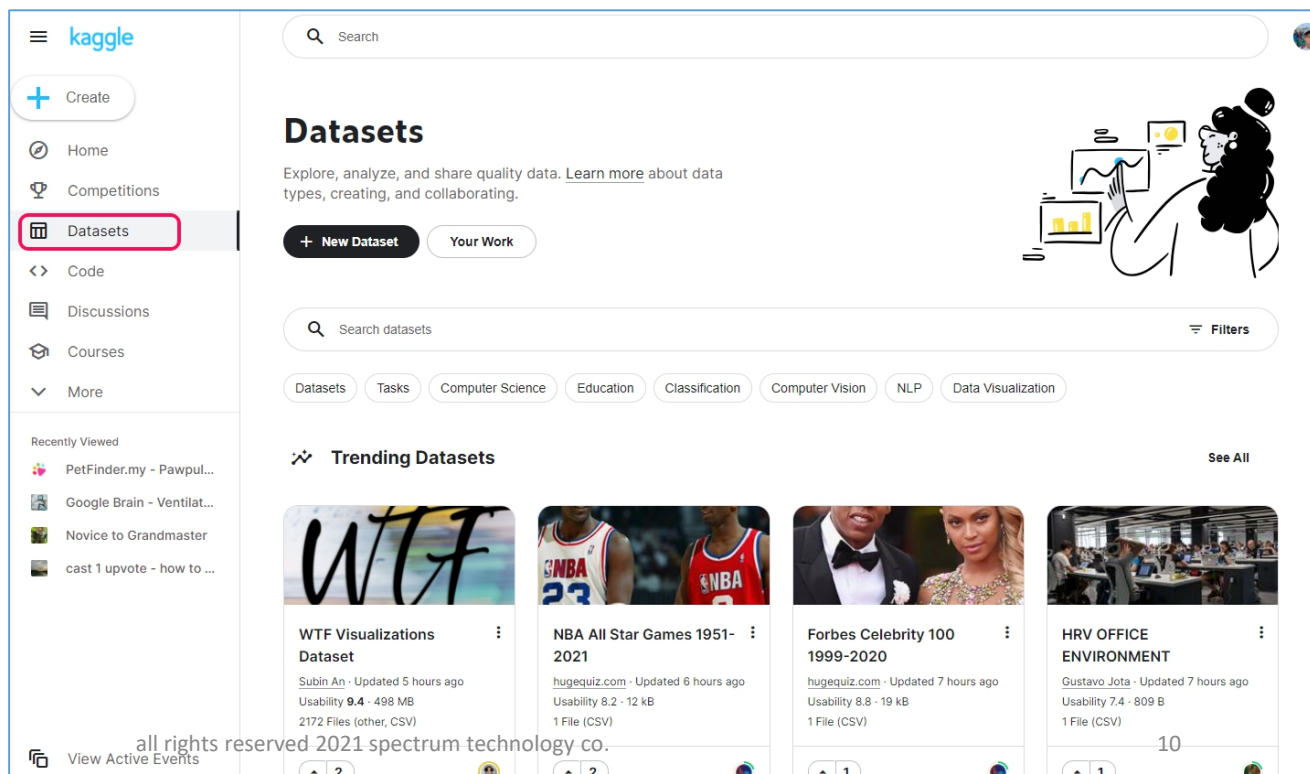
③ データセット

- 有名なmnistなどのデータが準備されています。個別にアップロードする必要なし

④ Code

⑤ 議論

⑥ コース



kaggle

3. メニュー説明

- ① ホーム
- ② コンペ
- ③ データセット
- ④ Code

- Jupyter notebookのようなものが準備されており、pythonのコードを入力します。
- また、ほとんどのモジュールはインストールされているのでpipなどの操作不要

- ⑤ 議論
- ⑥ コース

The screenshot displays the Kaggle website interface. On the left is a sidebar menu with the following items: 'Create' (with a plus icon), 'Home', 'Competitions', 'Datasets', 'Code' (highlighted with a red box and a code icon), 'Discussions', 'Courses', and 'More'. Below the menu is a 'Recently Viewed' section listing items like 'PetFinder.my - Pawpul...', 'Google Brain - Ventil...', 'Novice to Grandmaster', and 'cast 1 upvote - how to ...'. The main content area is titled 'Code' and includes the text 'Explore and run machine learning code with Kaggle Notebooks. Find help in the Documentation.' Below this are buttons for '+ New Notebook' and 'Your work'. A search bar for 'Search public notebooks' is present, followed by filter tags: 'All notebooks', 'Recently Viewed', 'Python', 'R', 'Beginner', 'NLP', 'Finance', 'Random Forest', 'GPU', 'TPU', and 'Competition notebook'. A 'Trending' section at the bottom shows various notebook thumbnails, including '[TPS-11] NN Baseline (Keras)', 'Understanding Neural Net', '[EDA + Versioning] Easy-to-use Dataset', and 'Titanic Prediction - ANN'. An illustration of a person sitting and working on a laptop is visible in the top right of the main content area.

kaggle

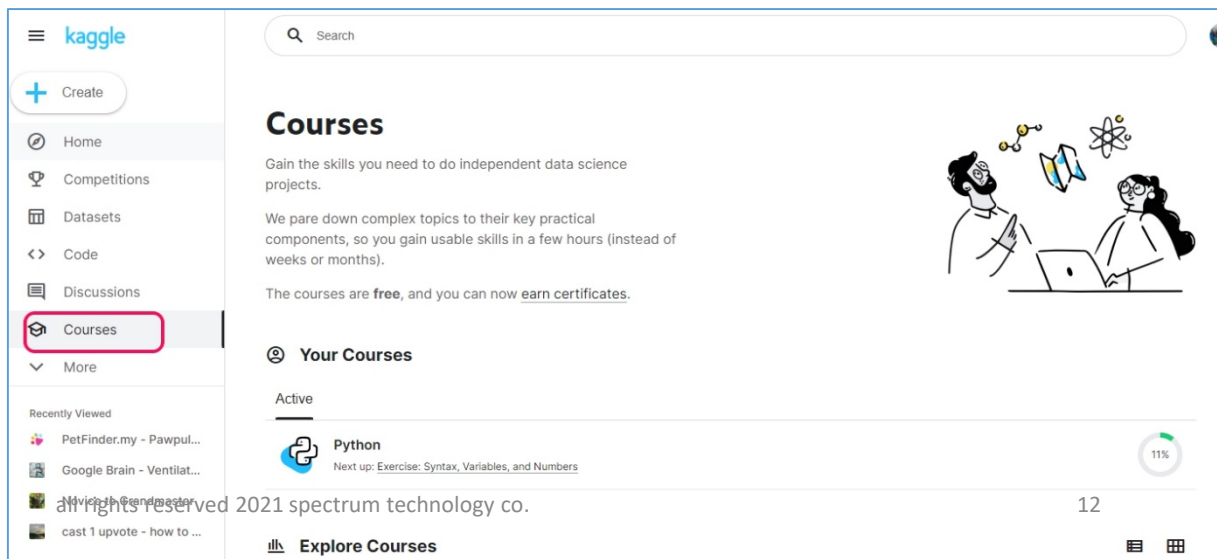
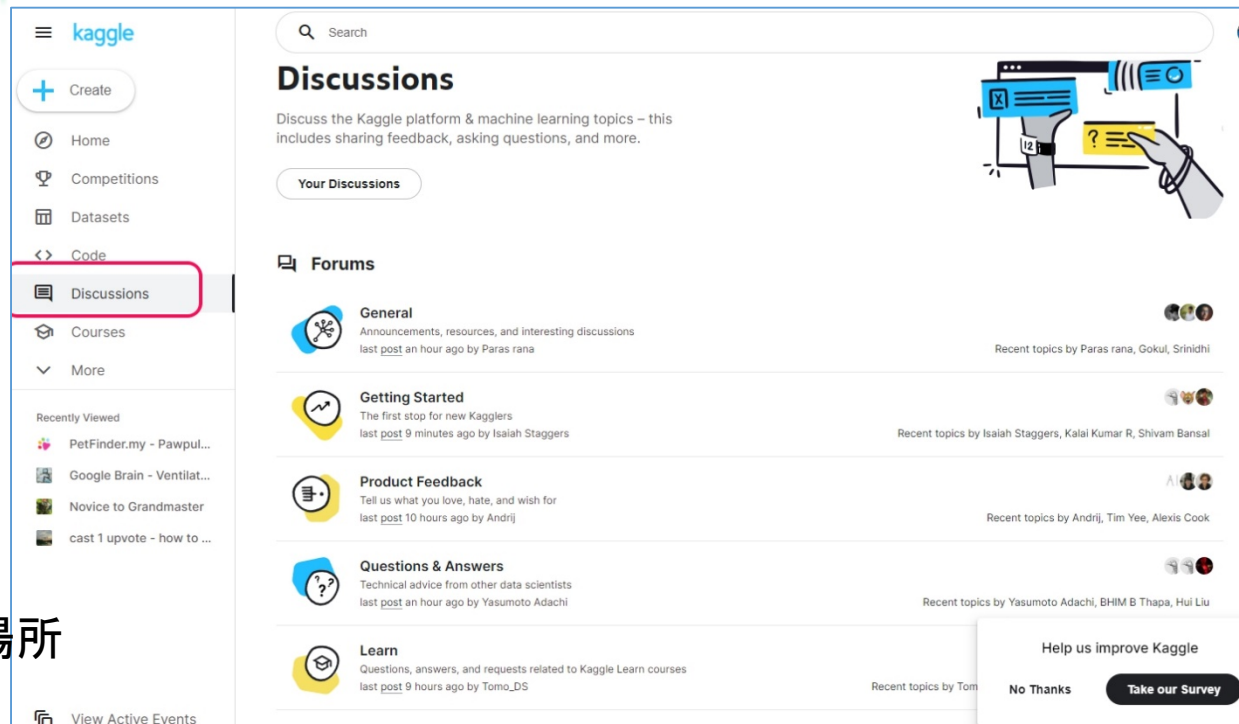
3. メニュー説明

- ① ホーム
- ② コンペ
- ③ データセット
- ④ Code
- ⑤ 議論
- ⑥ コース

• コミュニティのような場所

- ⑥ コース

• 学習コースが準備

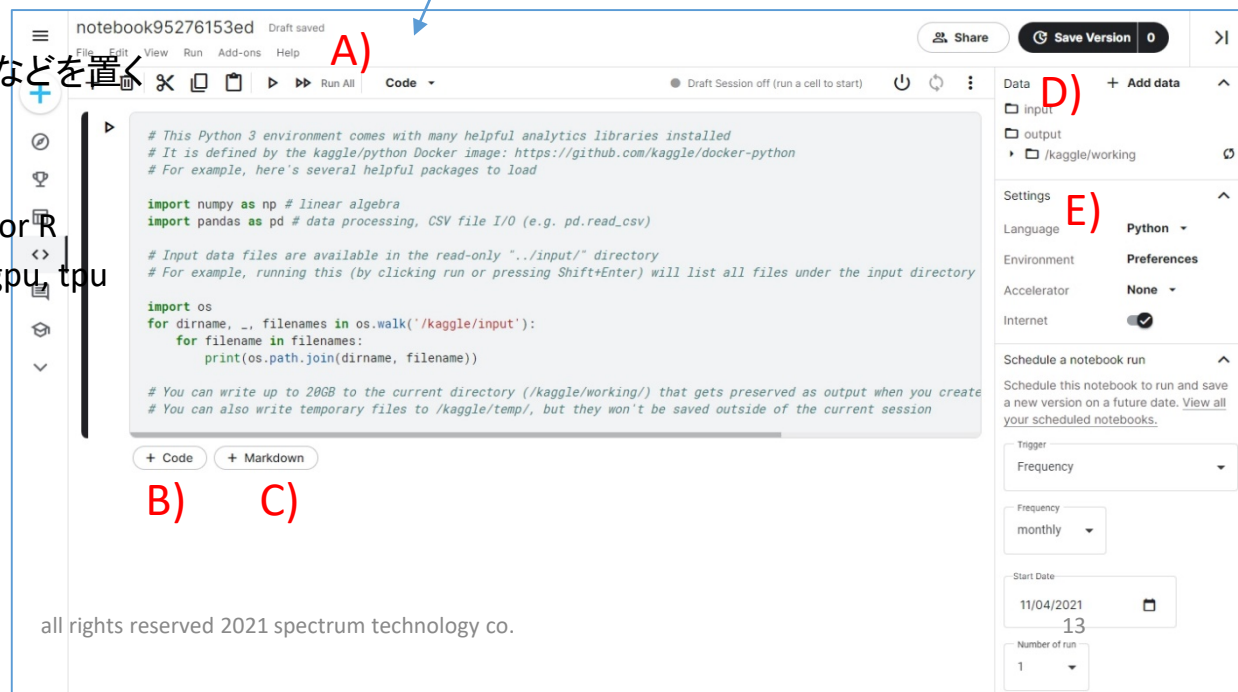
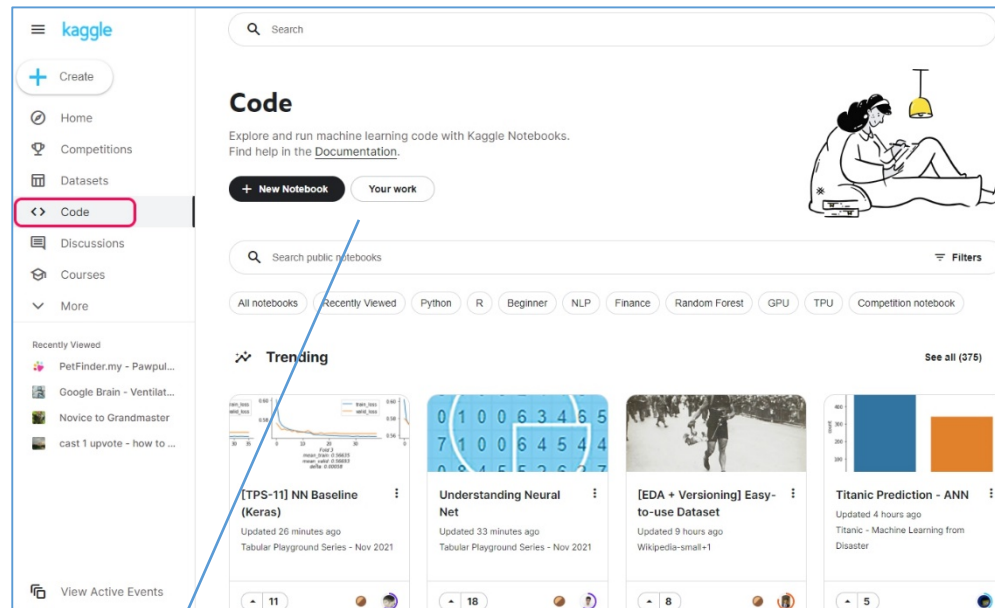


kaggle

4. Code使用方法

① Notebook作成

- 新規Notebook
 - Code > New notebook
- 使い方
 - A) Jupyter notebookと同じ
 - B) +code: pythonのプログラムを記入
 - C) +Markdown: コメント記入
 - D) Data
 - Input: 使用するcsvなどを置く
 - Output: 出力
 - E) Setting
 - Language: Python or R
 - Accelerator: None, gpu, tpu

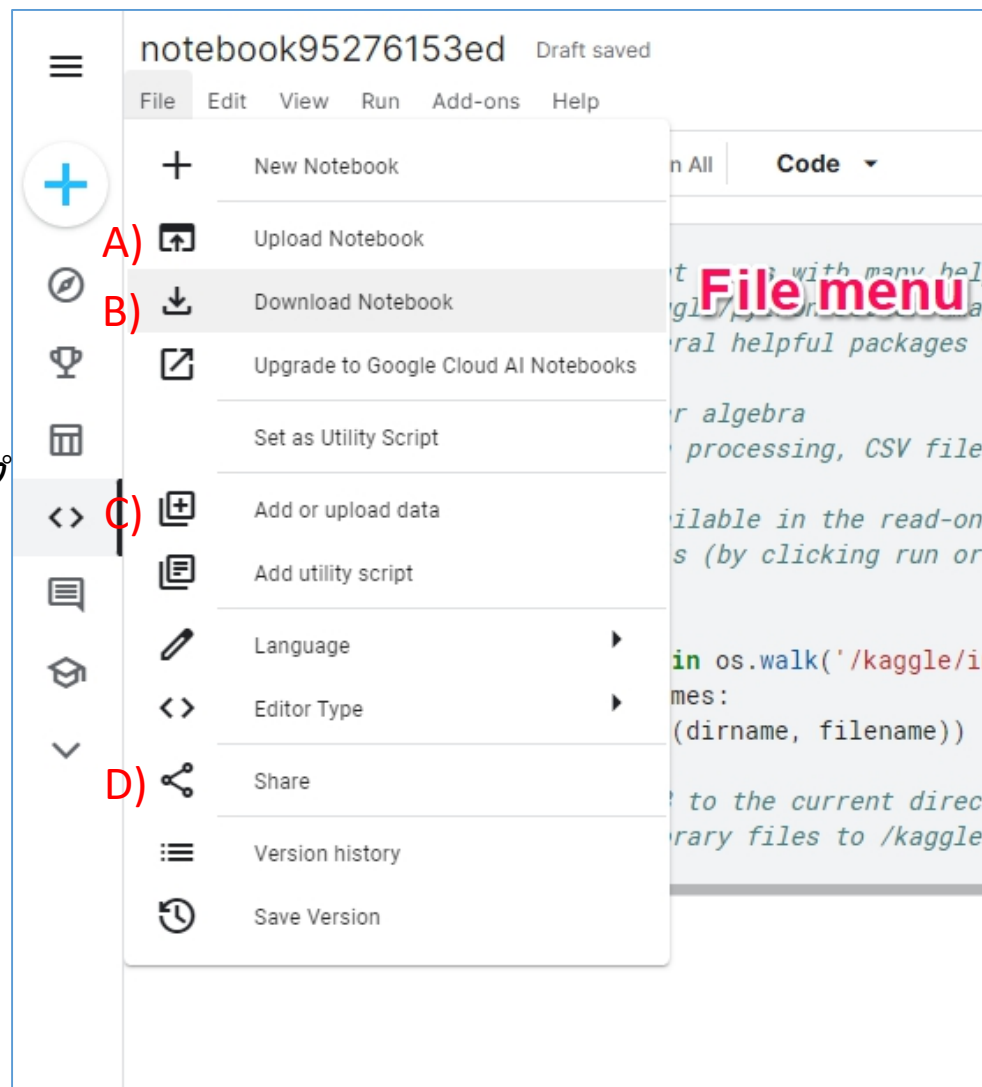


kaggle

4. Code使用方法

① Notebook作成

- 新規Notebook
- 使い方
- File > menu
 - A) Upload notebook: ローカルから
 - B) Download notebook: ローカルへ
 - C) Add or upload data: datasetのアップ
 - D) Share: Publicに公開できる



kaggle

4. Code使用方法

② Notebook使用例

- Courseからpython学習をコピーした使用例

A) Course > Python選択

B) Hello, Pythonを選択

The image displays two screenshots of the Kaggle website, illustrating the steps to access the 'Hello, Python' course.

Screenshot 1 (Left): The 'Courses' menu item in the left sidebar is highlighted with a red box and labeled 'A)'. The 'Python' course is highlighted in the 'Your Courses' section with a red box.

Screenshot 2 (Right): The 'Hello, Python' lesson is highlighted with a red box and labeled 'B)'. The lesson title is 'Hello, Python', and it shows a progress indicator of 50%.

The interface includes a search bar, a 'Create' button, and a sidebar with navigation links: Home, Competitions, Datasets, Code, Discussions, Courses, and More. The 'Courses' section lists several courses: 'Intro to Machine Learning', 'Pandas', 'Intermediate Machine Learning', and 'Data Visualization'.

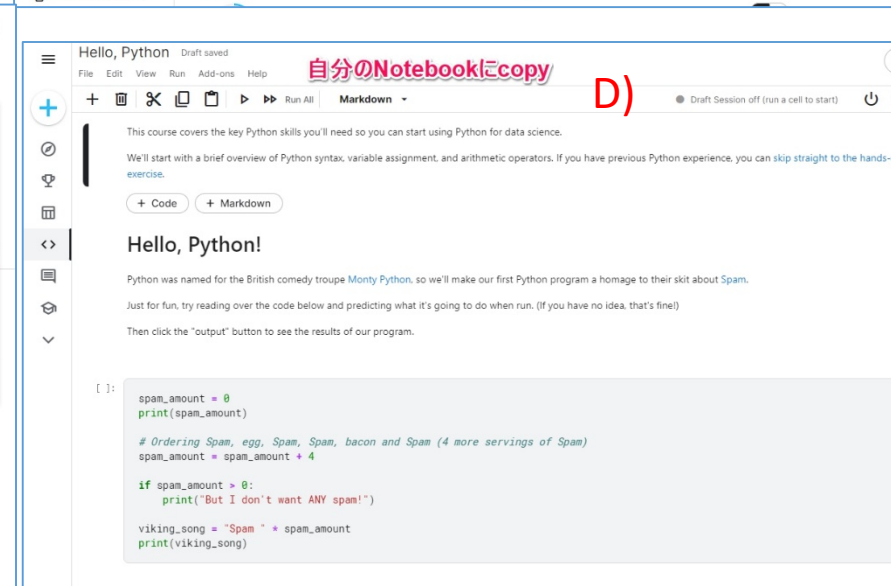
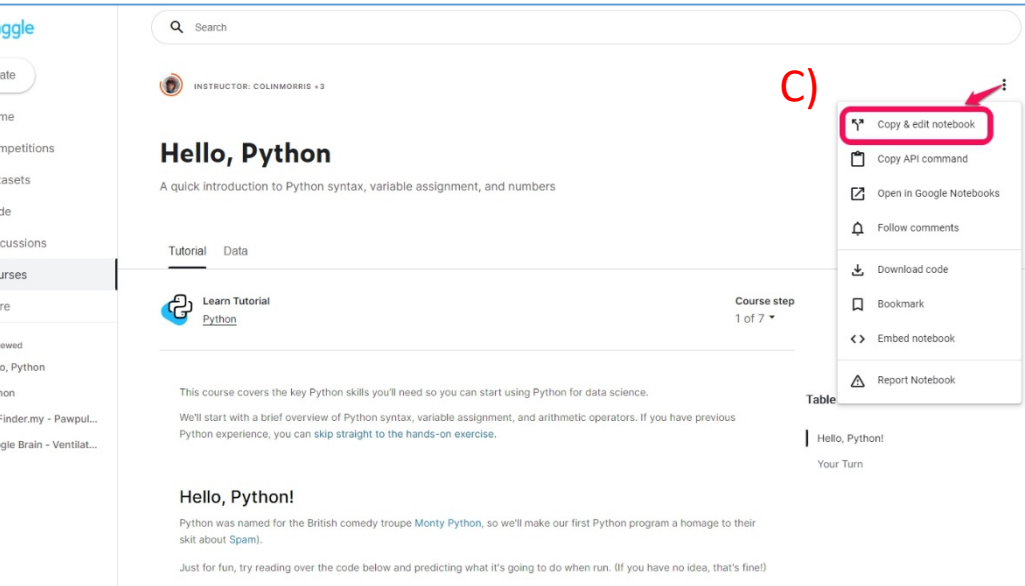
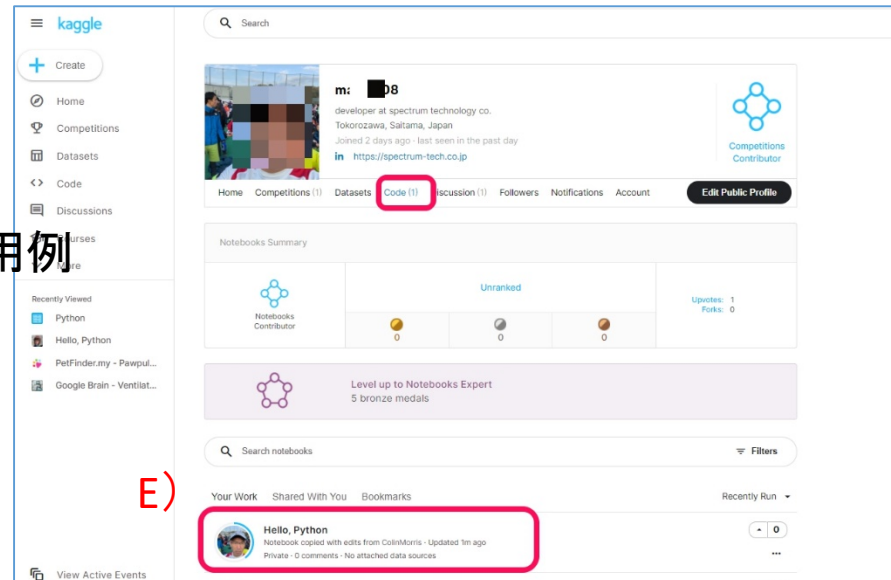
kaggle

4. Code使用方法

② Notebook使用例

- Courseからpython学習をコピーした使用例

- Course>Python選択
- Hello, Pythonを選択
- Copy notebook
- 自分のnotebookにコピーされます。
- Home>my page>codeにあることを確認



kaggle

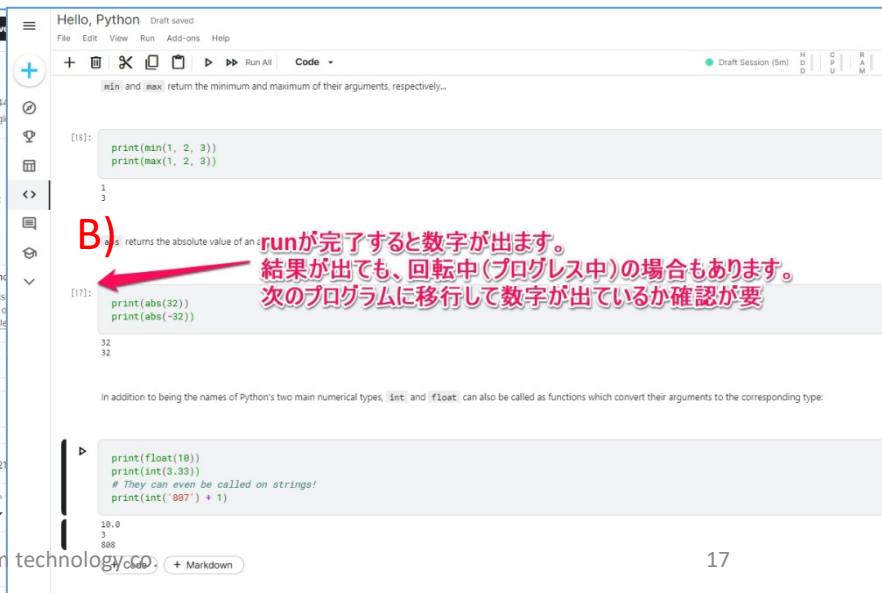
4. Code使用方法

② Notebook使用例

- Courseからpython学習をコピーした使用例
- Hello,Python使用例

A) Run: codeのある場所でRunを押すと実行。print文がある場合は出力される

B) 数字の確認: Runが完了すると数字がでます。プログレス中の場合もあるので注意。



kaggle

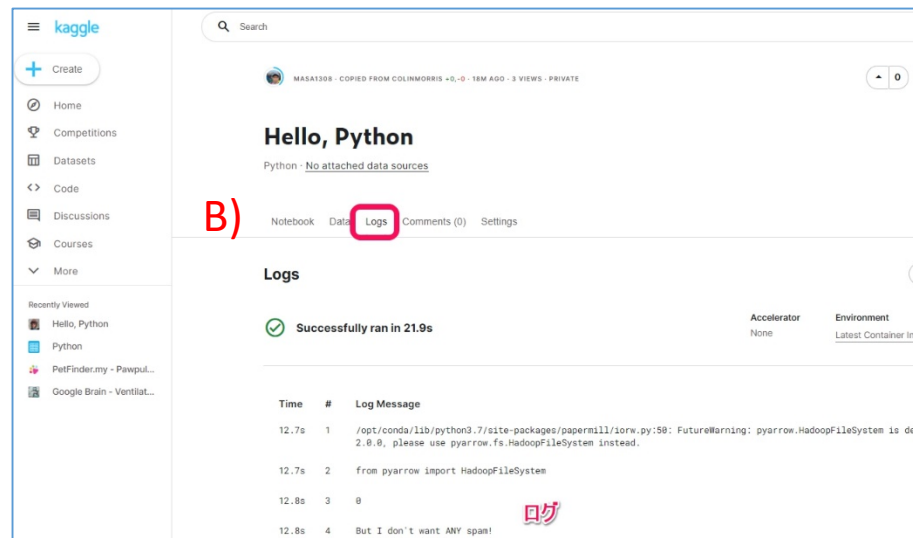
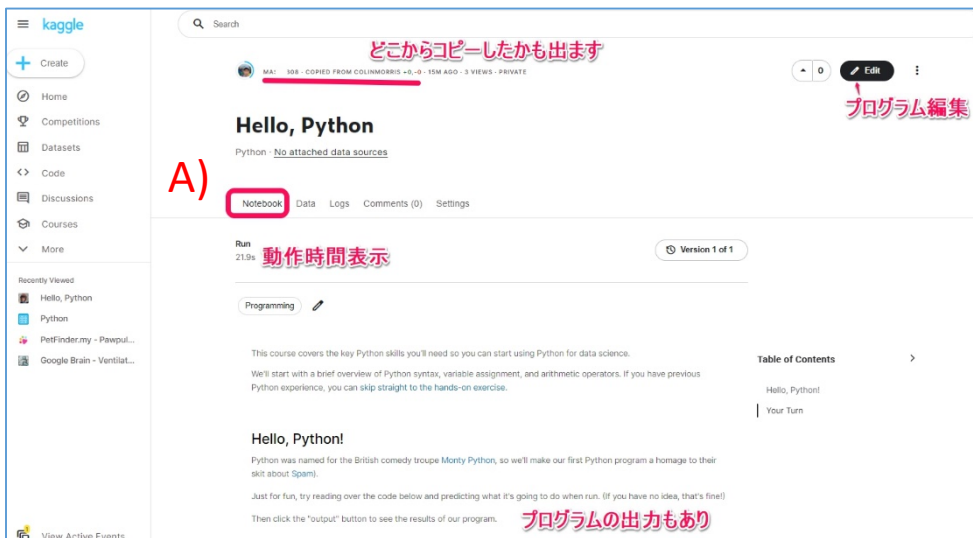
4. Code使用方法

② Notebook使用例

- Courseからpython学習をコピーした使用例
- Hello,Python使用例
 - Hello,pythonは全てのプログラムが完了すると自動保存。又は別のバージョンでSave version
 - My page>code: hello pythonを選択するとこれまでの状況が分かります

A) Notebook: 動作時間、実行結果

B) Logs: ログが表示



kaggle

5. Competition使用方法

① コンペ使用方法

A) 活動中で参加したいコンペの選択

B) コンペ内容確認: 概要をよく確認。賞金が設定しているのは、難しい課題です。初心者は、賞金のないものから

The screenshot displays the Kaggle website interface. On the left sidebar, the 'Competitions' tab is selected. The main content area shows 'Active Competitions'. The 'NFL Big Data Bowl 2022' competition is highlighted with a red box and labeled 'A)'. The 'Prize Money' of '\$100,000' is circled in red and labeled 'B)'. The 'Overview' tab is selected, showing the competition description. A blue arrow points from the 'NFL Big Data Bowl 2022' card to the 'Overview' tab.

Active Competitions

NFL Big Data Bowl 2022

Help evaluate special teams performance...

\$100,000 Prize Money

Overview

Description

Before National Football League (NFL) coaches celebrate a big W, they strategize ways to improve field position and score points. Both of these objectives receive significant contributions from special teams plays, which consist of punts, kickoffs, field goals and extra points. These play types take on important roles in a game's final score—so much so that coaches say they're a third of the game. Yet special teams remain an understudied part of American football, with an opportunity for data science to offer better ways to understand its impact.

The 2022 Big Data Bowl creates the opportunity for you (and the world!) to learn more about special teams play than ever before. We've provided the NFL's **Next Gen Stats** (NGS) tracking data from all 2018-2020 special teams plays. This data provides location information for each special teams player, wherever they are on the field, and includes their speed, acceleration, and direction. Additionally, and for the first time in Big Data Bowl history, participants can utilize scouting data from PFF, which supplements the tracking data with football specific metrics that coaches find critical to team success.

The NFL is America's most popular sports league. Founded in 1920, the organization behind American football has developed the model for the successful modern sports league. They're committed to advancing every aspect of the game, including the lesser researched special teams. In this competition, you'll quantify what happens on special teams plays. You might create a new special teams metric, quantify team or individual strategies, rank players, or even something we haven't considered.

With your creativity and analytical skills, the development of these new methods could lead to additional stats for special teams plays. If successful, your effort **may even be adopted by the NFL** for on air distribution, and you can watch future games knowing you had a hand in improving America's most popular sports league.

kaggle

5. Competition使用方法

① コンペ使用方法

- A) 活動中で参加したいコンペの選択
- B) コンペ内容確認
- C) データ: 提供されてるcsvデータなどを確認
- D) ルール: 賞金、チームの人数などのルール

Search

Analytics Competition

NFL Big Data Bowl 2022

Help evaluate special teams performance

The National Football League · 2 months to go

\$100,000 Prize Money

Overview **Data** Code Discussion Rules

Data Description

Summary of data

The 2022 Big Data Bowl data contains Next Gen Stats player tracking, play, game, player, and PFF scouting data for all 2018-2020 Special Teams plays. Here, you'll find a summary of each data set in the 2022 Data Bowl, a list of key variables to join on, and a description of each variable.

File descriptions

Game data: The `games.csv` contains the teams playing in each game. The key variable is `gameId`.

Play data: The `plays.csv` file contains play-level information from each game. The key variables are `gameId` and `playId`.

>_ `kaggle competitions download -c nfl-big-data-bowl-2022`

Data Explorer

5 GB

< **PFFScoutingData.csv** (2.06 MB)

↓

← 提供されているデータ

PFFScoutingData.csv
games.csv
players.csv
plays.csv
tracking2018.csv
tracking2019.csv

Analytics Competition

NFL Big Data Bowl 2022

Help evaluate special teams performance

The National Football League · 2 months to go

\$100,000 Prize Money

Overview Data Code Discussion **Rules**

By clicking on the "I understand and accept" button, you indicate that you agree to be bound with the rules outlined below.

I Understand and Accept

Rules

Updated: November 1, 2021. Section A.3 was removed. Section A.3 prohibited the use of external data. External data is allowed following the guidance in Section B.7.C.

One account per participant

You cannot sign up to Kaggle from multiple accounts and therefore you cannot submit from multiple accounts.

all rights reserved 2021 spectrum technology inc. No private sharing outside teams

kaggle

5. Competition使用方法

① コンペ使用方法

- A) 活動中で参加したいコンペの選択
- B) コンペ内容確認
- C) データ: 提供されてるcsvデータなどを確認
- D) ルール: 賞金、チームの人数などのルール
- E) コード: 参照になる他の人のNotebook

kaggle

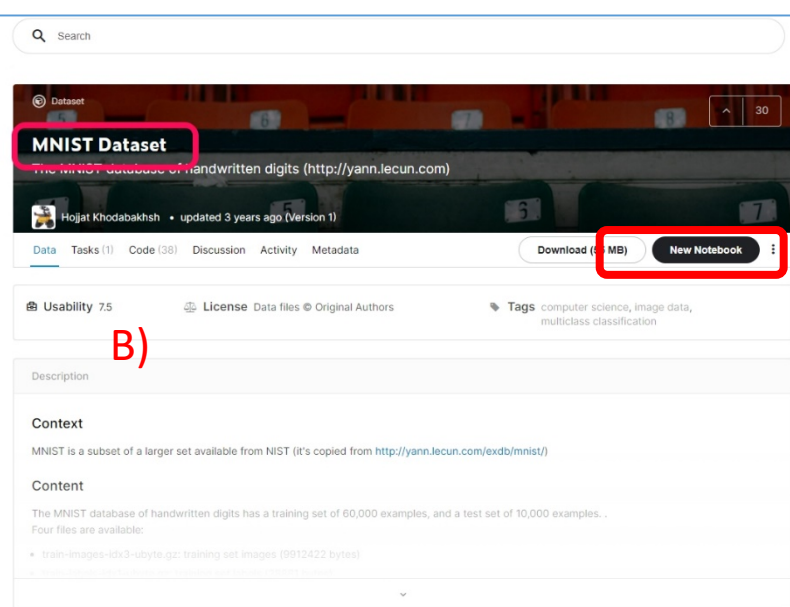
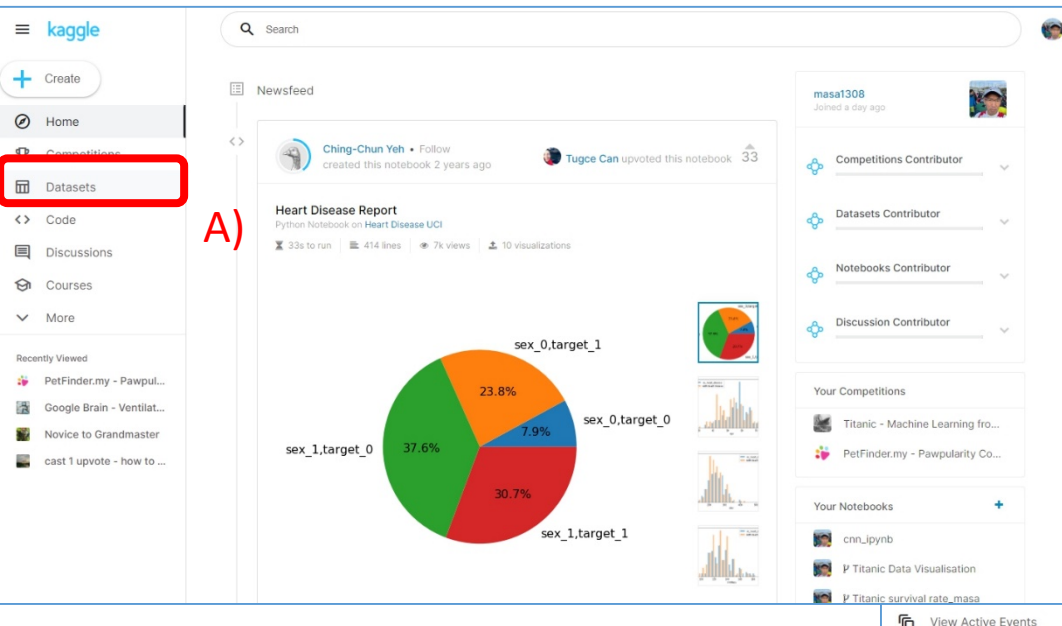
6. dataset使用方法

① dataset追加方法1

A) Dataset

B) Mnistを検索して、mnist datasetを選んで、notebook作成。

C) Notebookでコードを記入

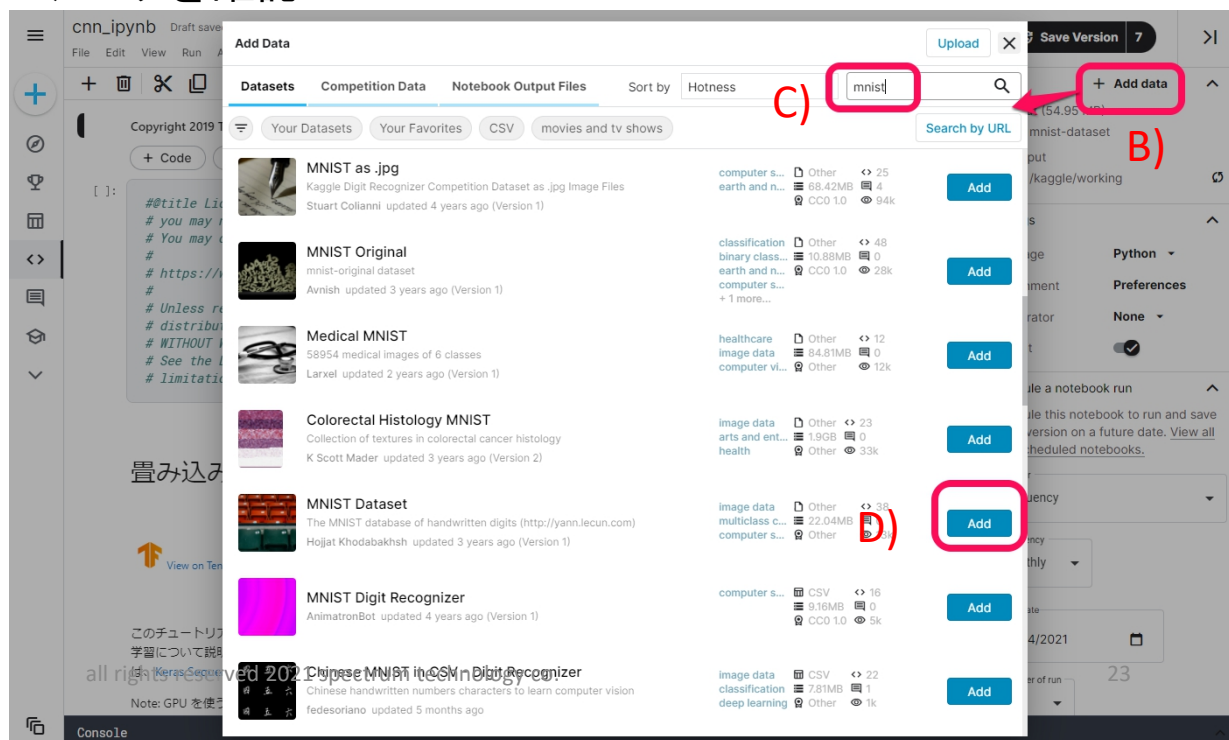


kaggle

6. dataset使用方法

② dataset追加方法2(一般的にはこちらか)

- A) Notebook作成またはアップロード
- B) Add data。
- C) mnistで検索
- D) Mnist datasetのAdd
- E) Notebookのinputにデータを確認



7. Discussions使用方法

① 議論使用方法

- A) Discussions > 操作方法、Q&Aの内容が記載
- B) 詳細: 個別の内容を見ると解決策があります。

The image shows a screenshot of the Kaggle website interface. On the left, the navigation menu is visible with the 'Discussions' link highlighted in a red box. The main content area is divided into two parts. The top part, labeled 'A)', shows the 'Discussions' section with a search bar and a list of discussion categories: General, Getting Started, Product Feedback, Questions & Answers, and Learn. A red annotation '操作方法、Q&Aなどのフォーラムあり' (There are forums for operation methods, Q&A, etc.) is placed over the 'Questions & Answers' category. The bottom part, labeled 'B)', shows a specific Q&A thread titled 'feature engineering: what is the logic?'. The thread includes a question by Valentina Teslenko and a comment section. A red annotation '質問事項と回答があります。困ったときは、searchで検索してください。' (There are questions and answers. When you are stuck, please search with search.) is placed over the thread title.

kaggle

8. 事例

[https://spectrum-](https://spectrum-tech.co.jp/products/ai_tensorflow.html)[tech.co.jp/products/ai_tensorflow.html](https://spectrum-tech.co.jp/products/ai_tensorflow.html)

弊社ホームページのAIプログラム学習キットも参照ください。

① CNN.ipynb

- 弊社で作成したtensorflow v2のcnn.ipynbを公開しています。
- <https://www.kaggle.com/masa1308/cnn-ipynb>
- Mnist dataset(手書き文字のベンチマーク用)を使って、Convolutional Neural Networksを学習するものです。
- Cpuとgpuの学習速度の違いを体感してください。本gpuはRTX2080より少し遅いくらいです。十分使えます。但し時間制限38hours/weekがあります。

The screenshot shows the Kaggle interface for a notebook titled 'cnn_ipynb'. The notebook is titled '畳み込みニューラルネットワーク (Convolutional Neural Networks)'. It includes a 'Table of Contents' section with a link to '畳み込みニューラルネットワーク...'. Below the title, there are four buttons: 'View on TensorFlow.org', 'Run in Google Colab', 'View source on GitHub', and 'Download notebook'. The main content area contains the following text:

このチュートリアルでは、MNIST の数の分類をするための、シンプルな畳み込みニューラルネットワーク (CNN: Convolutional Neural Network) の学習について説明します。このシンプルなネットワークは MNIST テストセットにおいて、99%以上の精度を達成します。このチュートリアルでは、[Keras Sequential API](#)を使用するため、ほんの数行のコードでモデルの作成と学習を行うことができます。

Note: GPU を使うことで CNN をより早く学習させることができます。もし、このノートブックを Colab で実行しているならば、編集 → ノートブックの設定 → ハードウェアアクセラレータ → GPU から無料の GPU を有効にすることができます。

TensorFlowのインポート

kaggle

8.事例

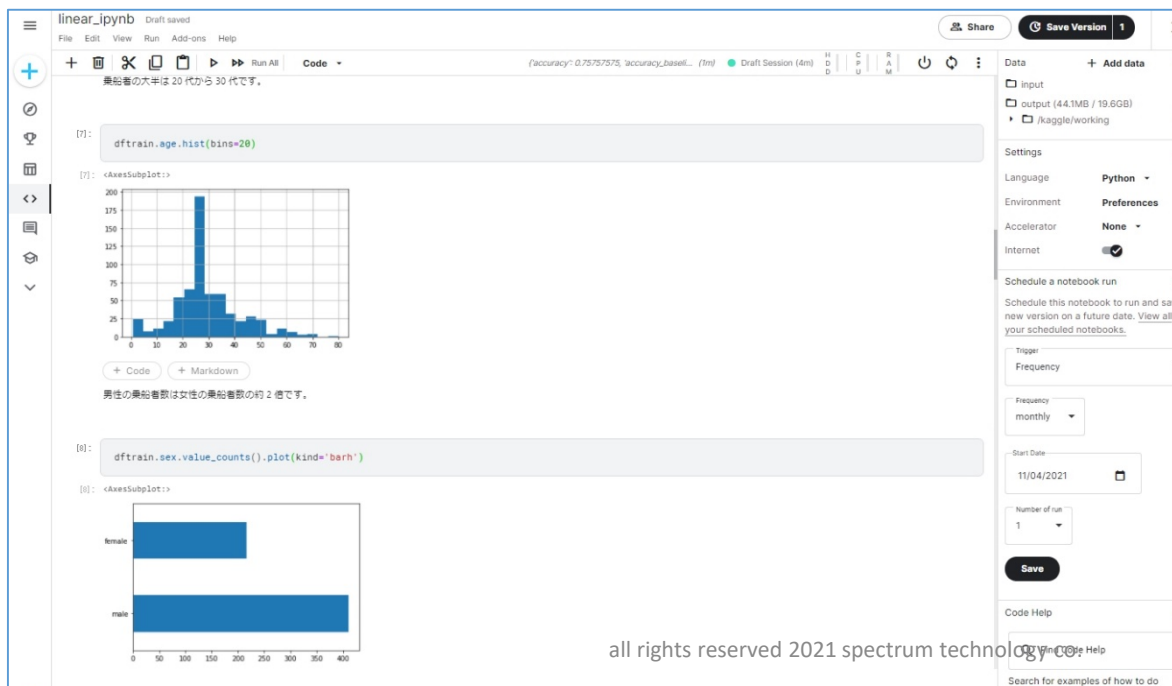
[https://spectrum-](https://spectrum-tech.co.jp/products/ai_tensorflow.html)

[tech.co.jp/products/ai_tensorflow.html](https://spectrum-tech.co.jp/products/ai_tensorflow.html)

弊社ホームページのAIプログラム学習キットも参照ください。

② linear.ipynb

- 弊社で作成したtensorflow v2のlinear.ipynbを公開しています。
- <https://www.kaggle.com/masa1308/linear-ipynb>
- titanic号の生存状況を予測モデルです。
- コンペにも同様なものがあります。以下も参照ください。
- <https://www.kaggle.com/c/titanic>

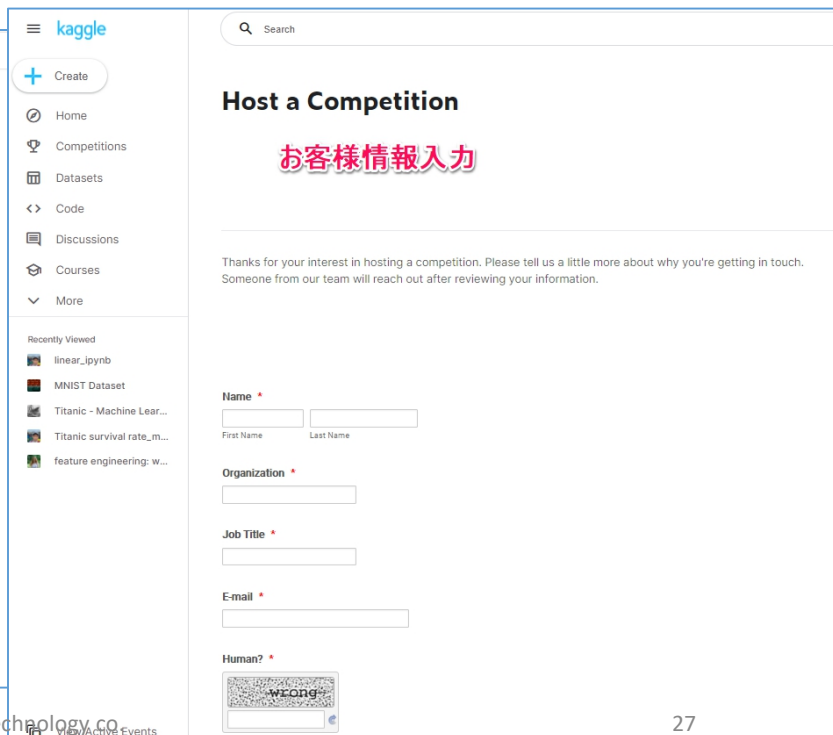
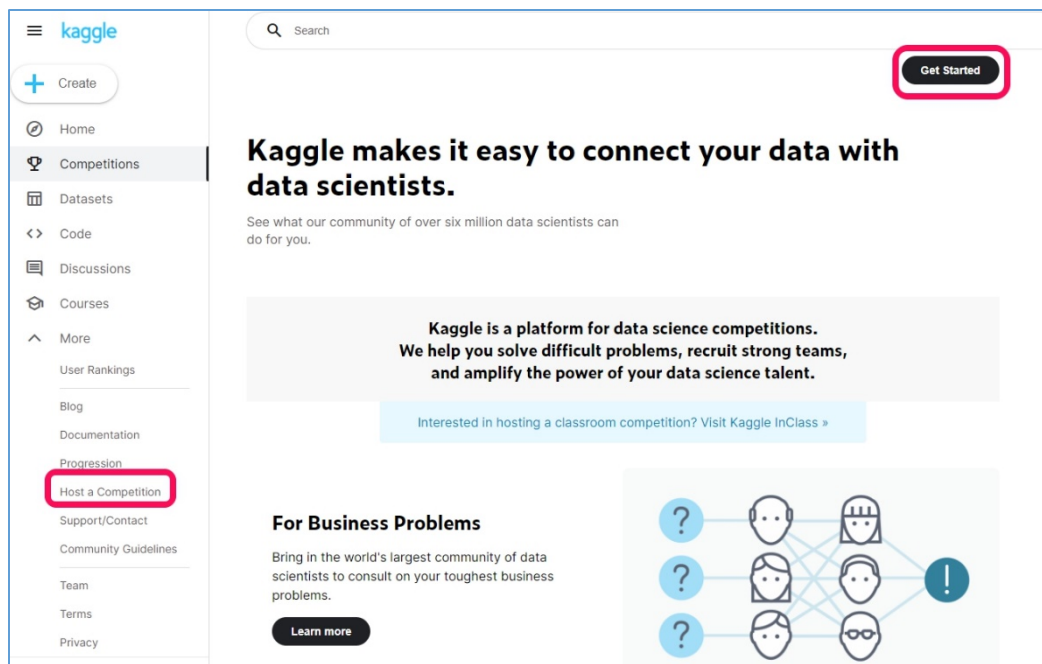


kaggle

9.コンペへ出展

① 出展に必要なこと

- スポンサーとしてコンペへの出展も可能です。
- プロフィール、解決したい内容を記述し、kaggleにコンタクトします。
- More > host a competition: get started
- プロフィール作成



kaggle

9.コンペへ出展

① 出展に必要なこと

- スポンサーとしてコンペへの出展も可能です。
- プロフィールを作成し、kaggleからのコンタクト待ち。
- More > host a competition: get started
- プロフィール作成

② コンペ内容作成

- 課題解決したい内容を作成し、必要なdataを添付し、賞金などを設定し提出
- 既にある課題内容をよく確認し、提出しましょう。

