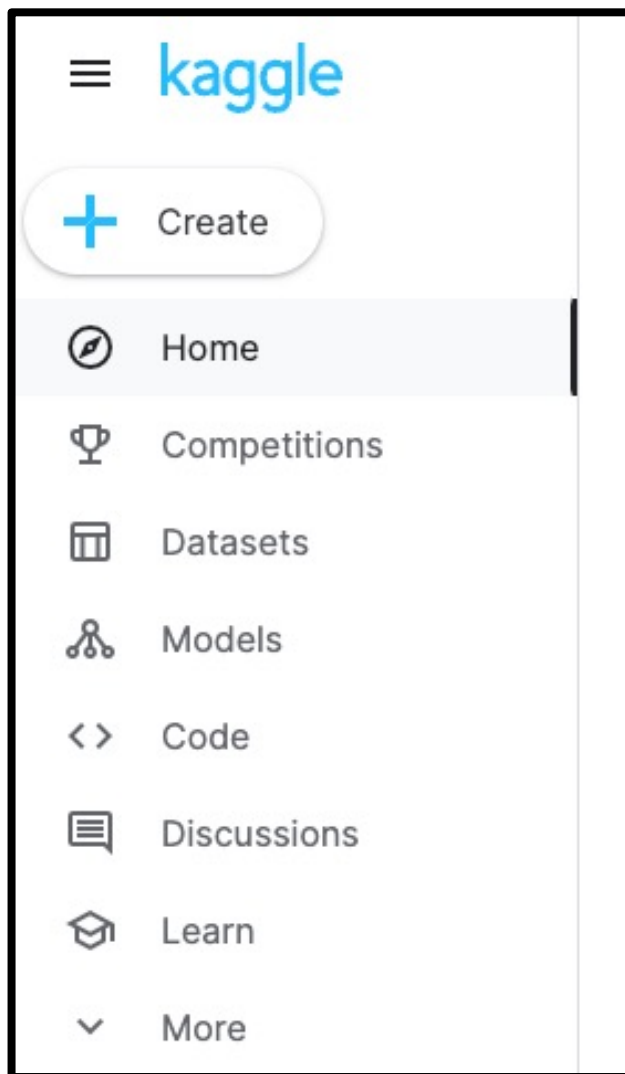


# Kaggleとは

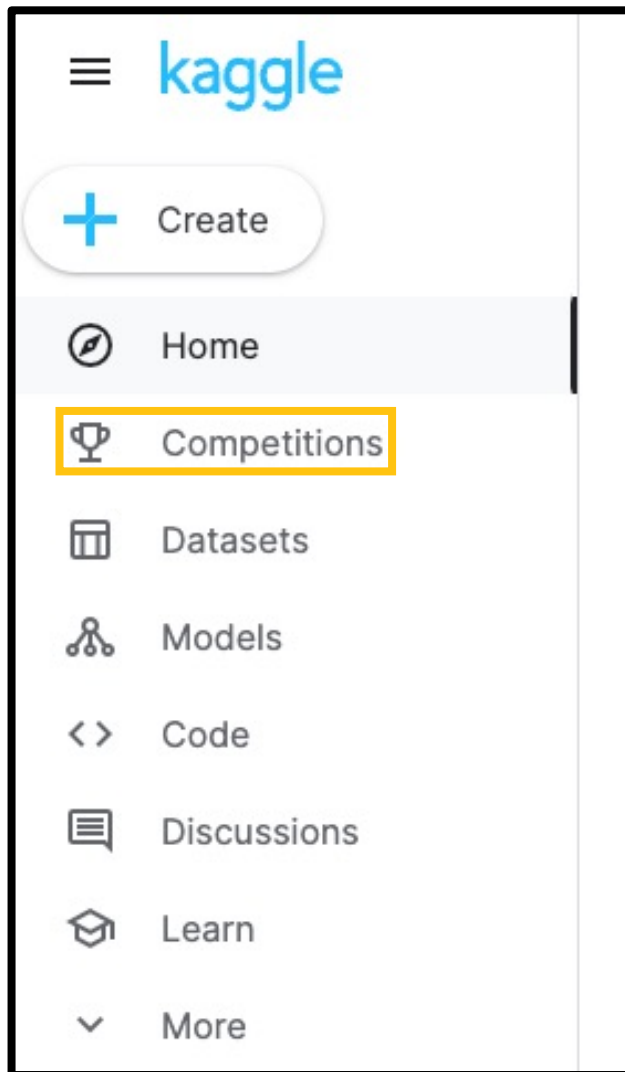


# データサイエンスと 機械学習の Webプラットフォーム

主に提供するサービスとして

- Competition
- Notebook
- Dataset
- Discussion

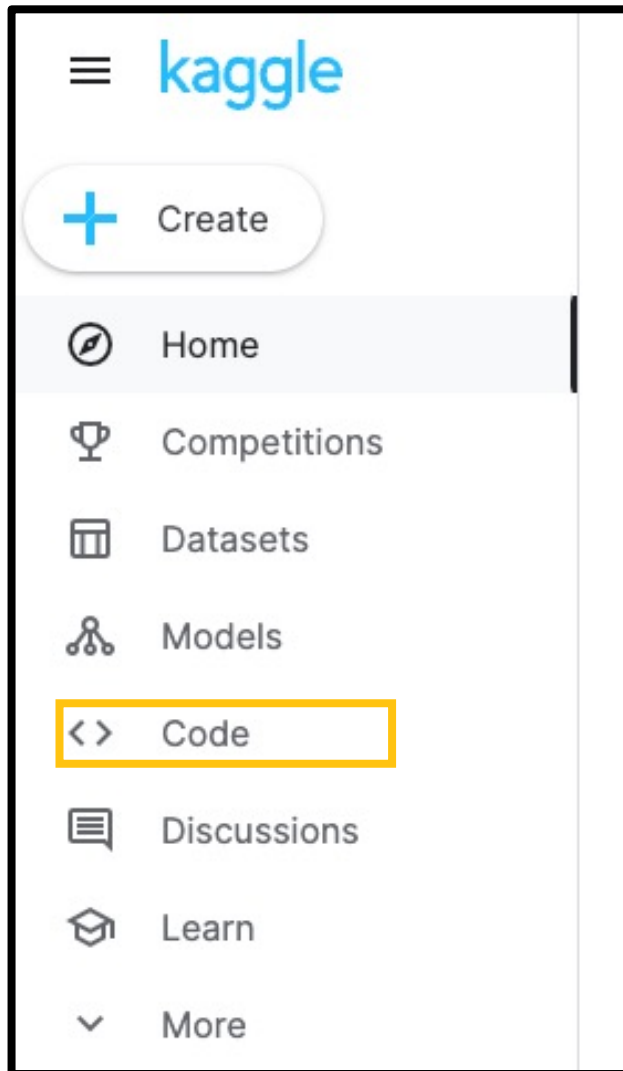
がある.



## Competition

企業や個人が主催するデータ分析系コンペティション.  
コンペのテーマに沿ってデータ分析・予測を行ない精度を競う.  
開催期間はコンペによって様々であり, 1週間から数ヶ月程度.

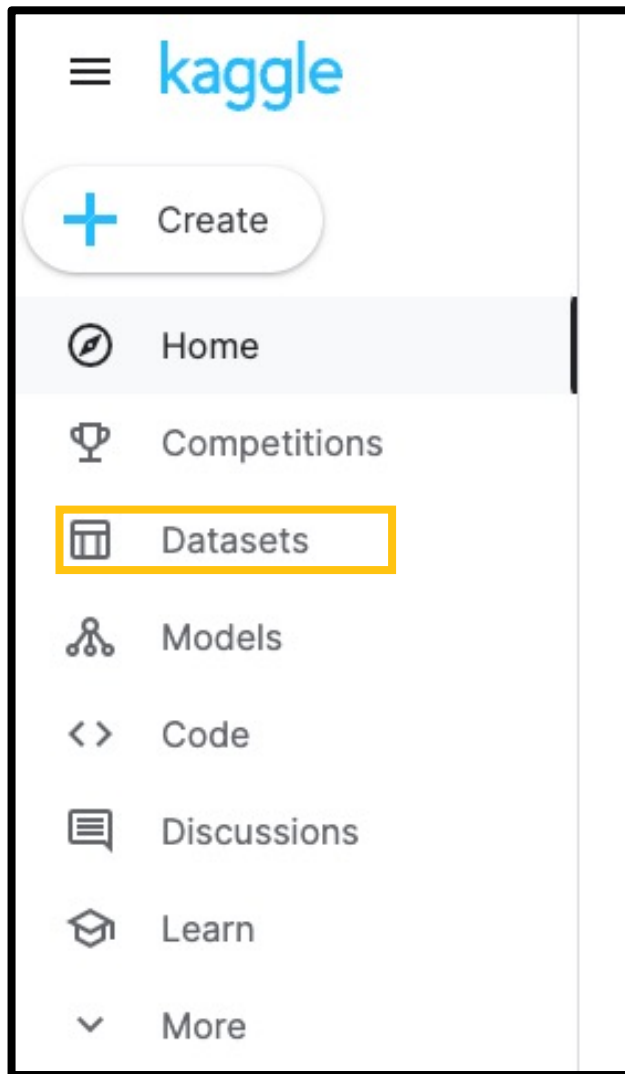
データが提供されるためデータ収集は必要無い場合が多い.  
上位成績者には賞金・賞品およびランキングポイントが貰える.



## Notebook (Code)

Kaggleサイト上でJupyterNotebookの実行環境を無料提供.

JupyterNotebookはWebブラウザでPythonの編集・実行ができる.  
KaggleNotebookではGPUも時間制限付きで利用できる.

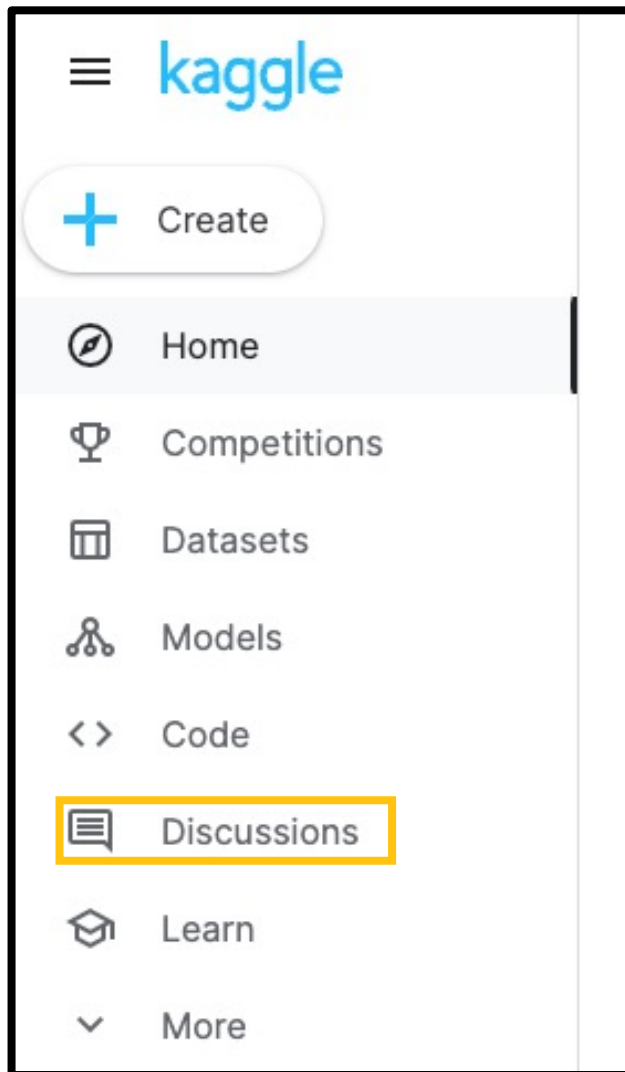


## Datasets

過去のコンペなどで使われたデータセットが提供されている。

テーブルデータだけでなく画像や自然言語系のデータも。

ライセンスについてはデータセットごとにバラバラだが、  
明記されているものがほとんどなので探す価値あり。



## Discussions

Kaggleユーザーの情報交換ためのフォーラム機能.

最前線の深層学習モデルについての議論は少なく,  
大抵は**コスト・リソースを考慮した現実的な手法**についての議論.  
→ 個人~小規模団体レベルの実践テクニックが得られる.

## 称号

以上の4つは成績やいいね数によってポイントが貰える.

それぞれのポイントが一定に達すると称号が得られ, 最高位はGrandmasterとなる.

Grandmasterは全体で493名おり, 日本では63名.

企業の事業内容によってはKaggleGrandmasterの数 = 企業価値.

以下, 内訳.

- Competition: 319
- Notebook: 135
- Discussion: 87
- Dataset: 53
- US: 96
- Japan: 63
- China: 41
- India: 35

# Kaggle会でやること（5月）



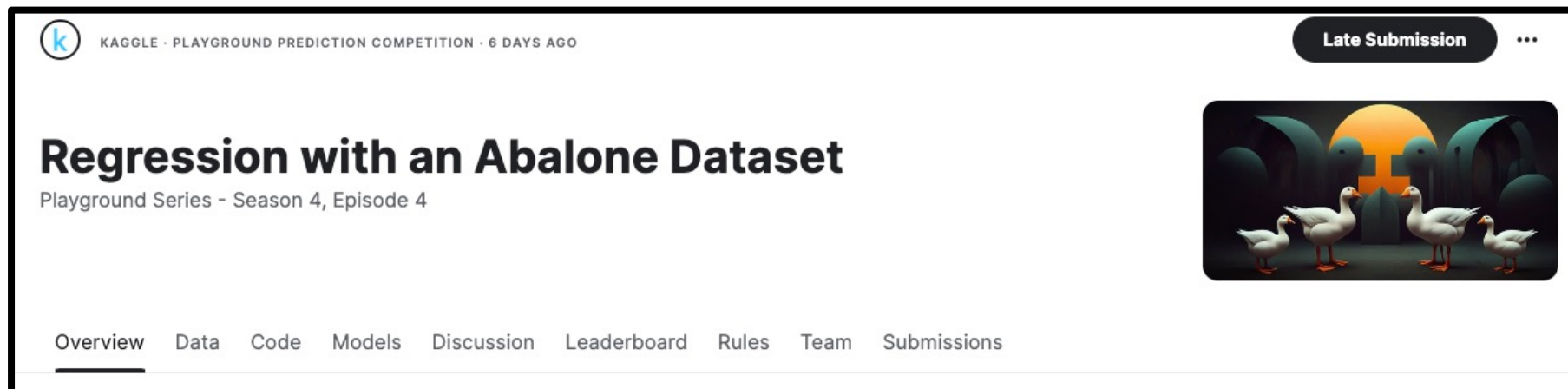
# PlayGround Series

Kaggle公式が開催するコンペ.

開催期間は1ヶ月で毎月初めから終わりまで.

予測タスクがほとんどでありデータ分析色が強い.

⇔ 検出タスクや分類タスクなど学習モデル色が強いもの.



## PlayGround Series

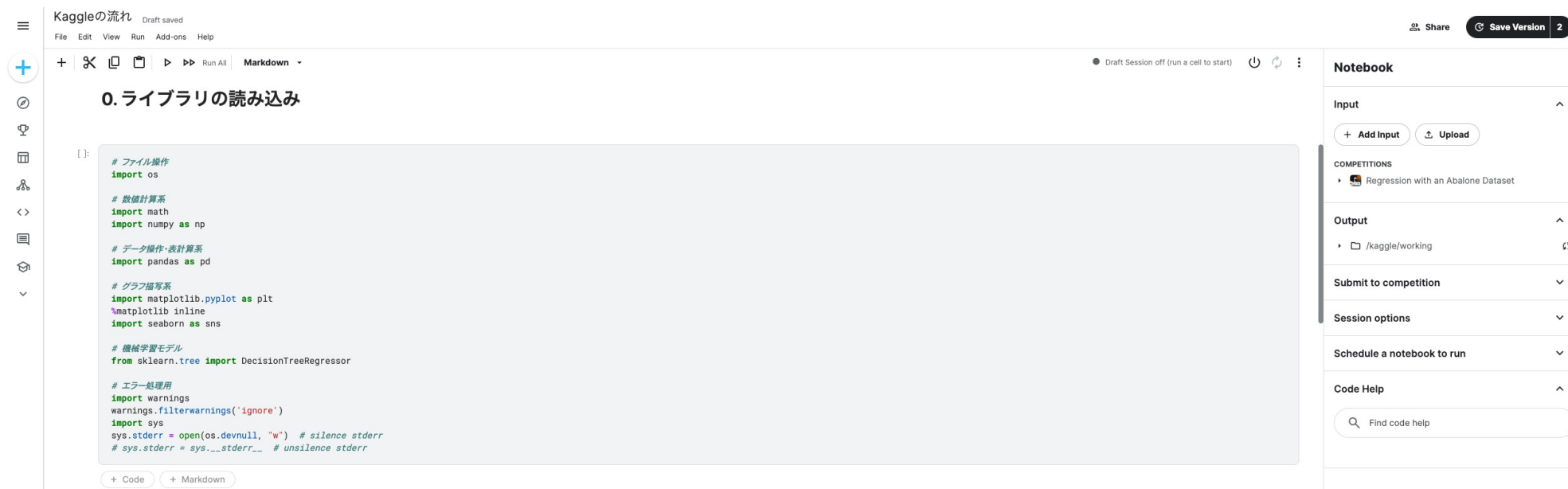
- Overview  
コンペの概要が確認できる。  
テーマや開催期間, タスク(評価方法)など.
- Data  
コンペで提供されているデータがダウンロードできる.
- Code  
他の参加者が公開したKaggleNotebookが閲覧できる。  
EDA-FEの方針や使用モデルなど説明付きで詳細に記載されているものもあり, 読むだけで勉強になる.
- Discussion  
参加者のフォーラム。  
コンペ終了後の方が盛り上がる。主にコンペ優勝・入賞者への質問など.
- Leaderboard & Submissions  
コンペのランキングがリアルタイムで確認できる。  
LeaderboardにはPublicとPrivateがあり, 開催期間中はPublicのスコアのみ確認でき, コンペ終了後にPrivateがオープンになり, 最終順位はPrivateのスコアで決定される。  
自分の提出一覧とスコアはSubmissionsから.

# Kaggle Notebook

JupyterNotebook + データの入出力機能 + ノートブックの共有機能

データの入出力：KaggleのDatasetsを直接利用可能 = データ分析をWeb上で完結可能.

ノートブックの共有：全体公開とURL共有が選択でき, Kaggleユーザーであれば共同編集が可能.



## 暫定では...

Googleが提供するサービスを使う予定.

- Google SpreadSheets
- Google Forms
- (Google Meets)

← KaggleがGoogle (Alphabet) のサービスであるため

**今週末のチーム分け発表の際に話す予定です**

# チュートリアル参照