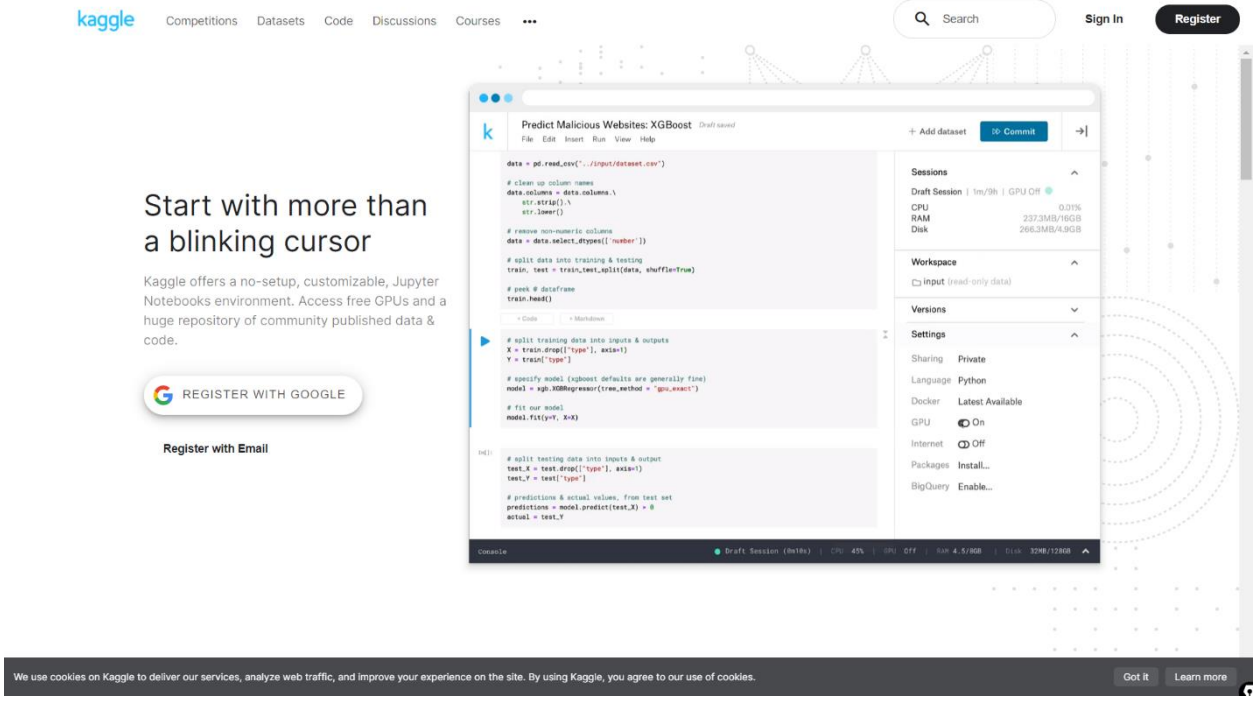
Первый взгляд и начало работы с Kaggle

Введение

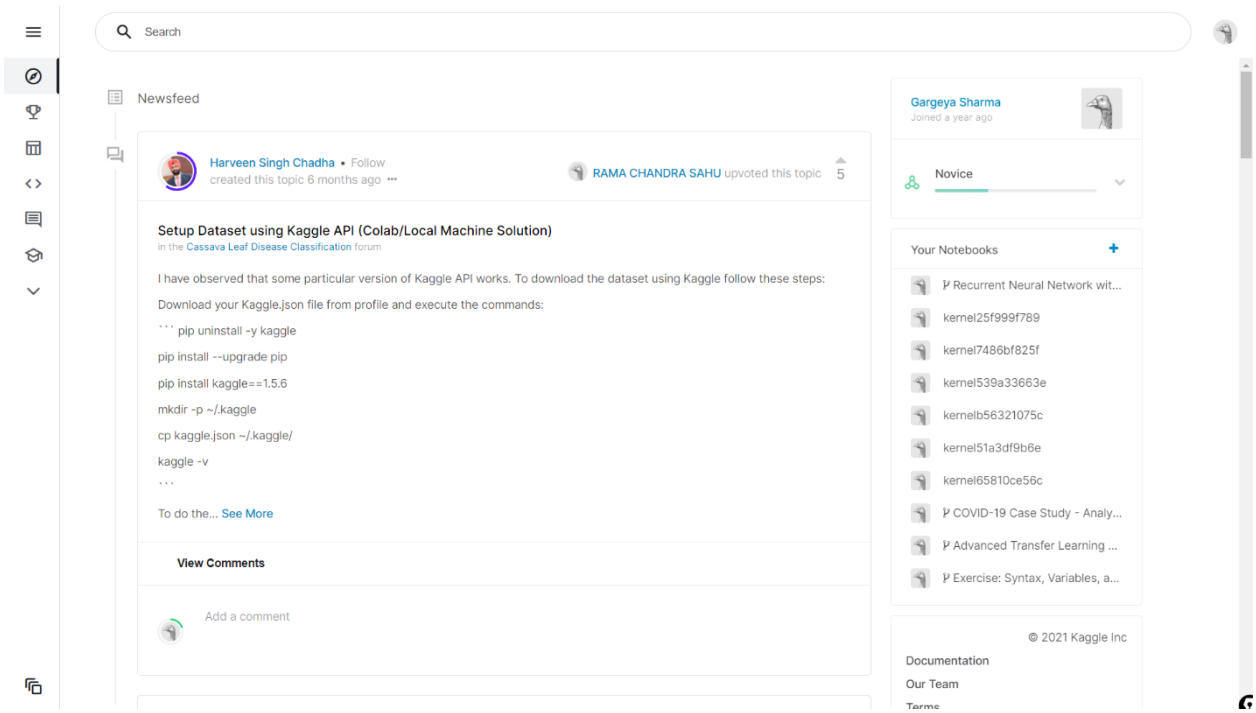
Сегодня в каждой карьере должно быть сообщество, группа людей, с которыми мы можем говорить о работе, ошибках, идеях и учиться. Kaggle - самое популярное и крупнейшее в мире сообщество специалистов по обработке и анализу данных. Наличие такого сообщества помогает нам чувствовать свою принадлежность, что является одним из важнейших чувств для нашего социального взаимодействия и здоровья.

В этой статье мы рассмотрим Kaggle как целое сообщество и Kaggle как платформу: все его различные инструменты, сервисы и ресурсы, доступные для нас, чтобы **изучать,** а также **практиковать науку о данных** .

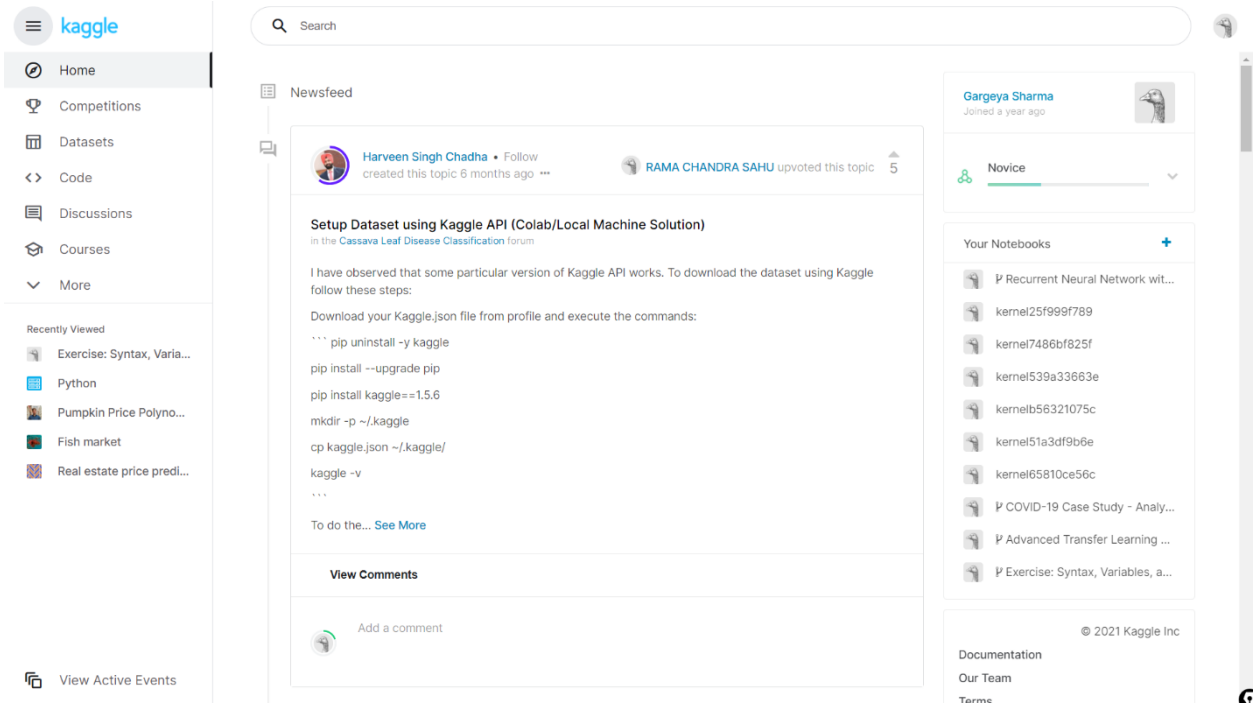
Давайте посмотрим на интерфейс, который мы получаем при первом посещении Kaggle.



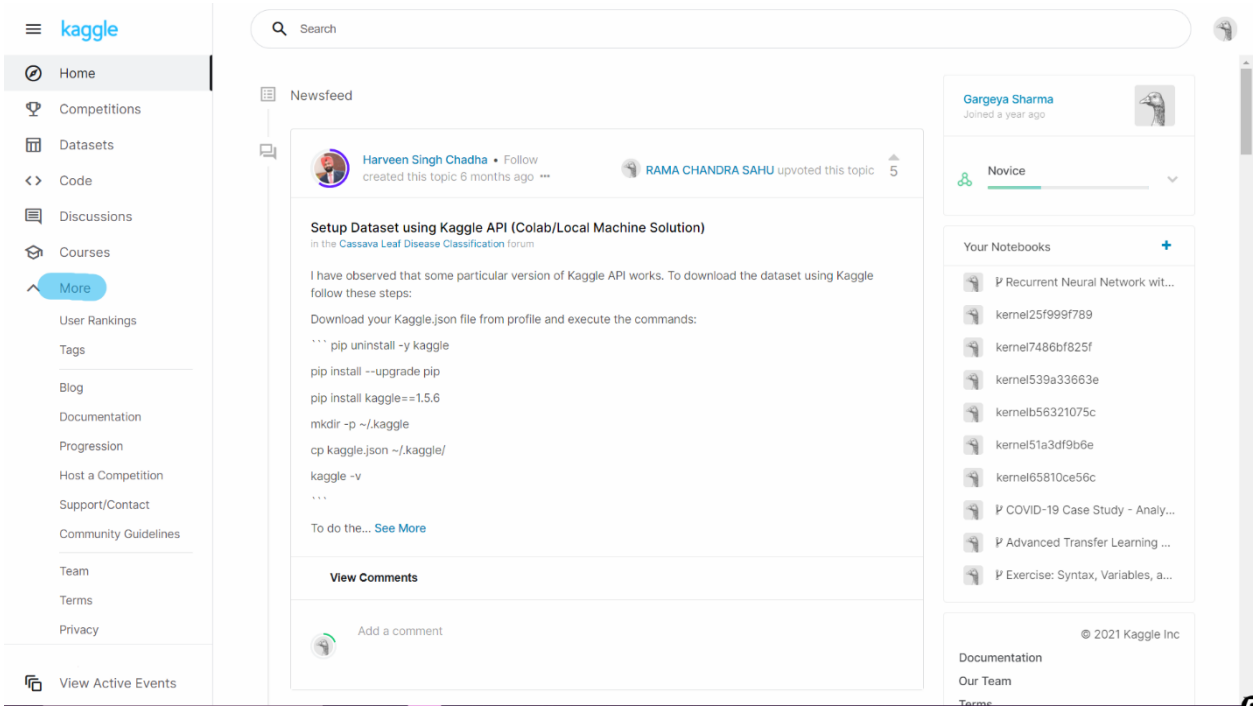
Прежде чем мы начнем использовать Kaggle, нам нужно создать учетную запись, а затем войти в систему, вы можете увидеть оба варианта в правом верхнем углу. Когда вы закончите с этим, это может выглядеть следующим образом:



Некоторые из видимых здесь вещей могут отличаться у вас, потому что интерфейс персонализирован в соответствии с тем, кто и как использовал Kaggle до сих пор с момента регистрации.



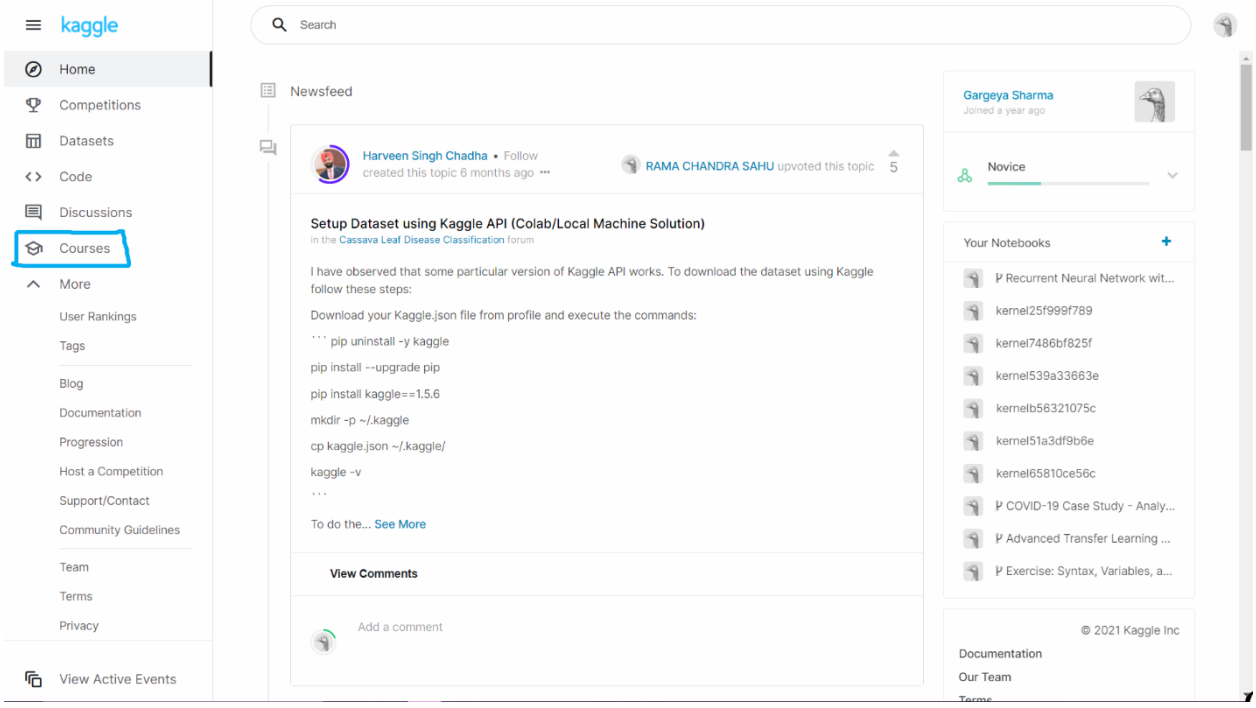
Как только нажимаешь «Home», это все, к чему я могу получить доступ из своей учетной записи Kaggle.

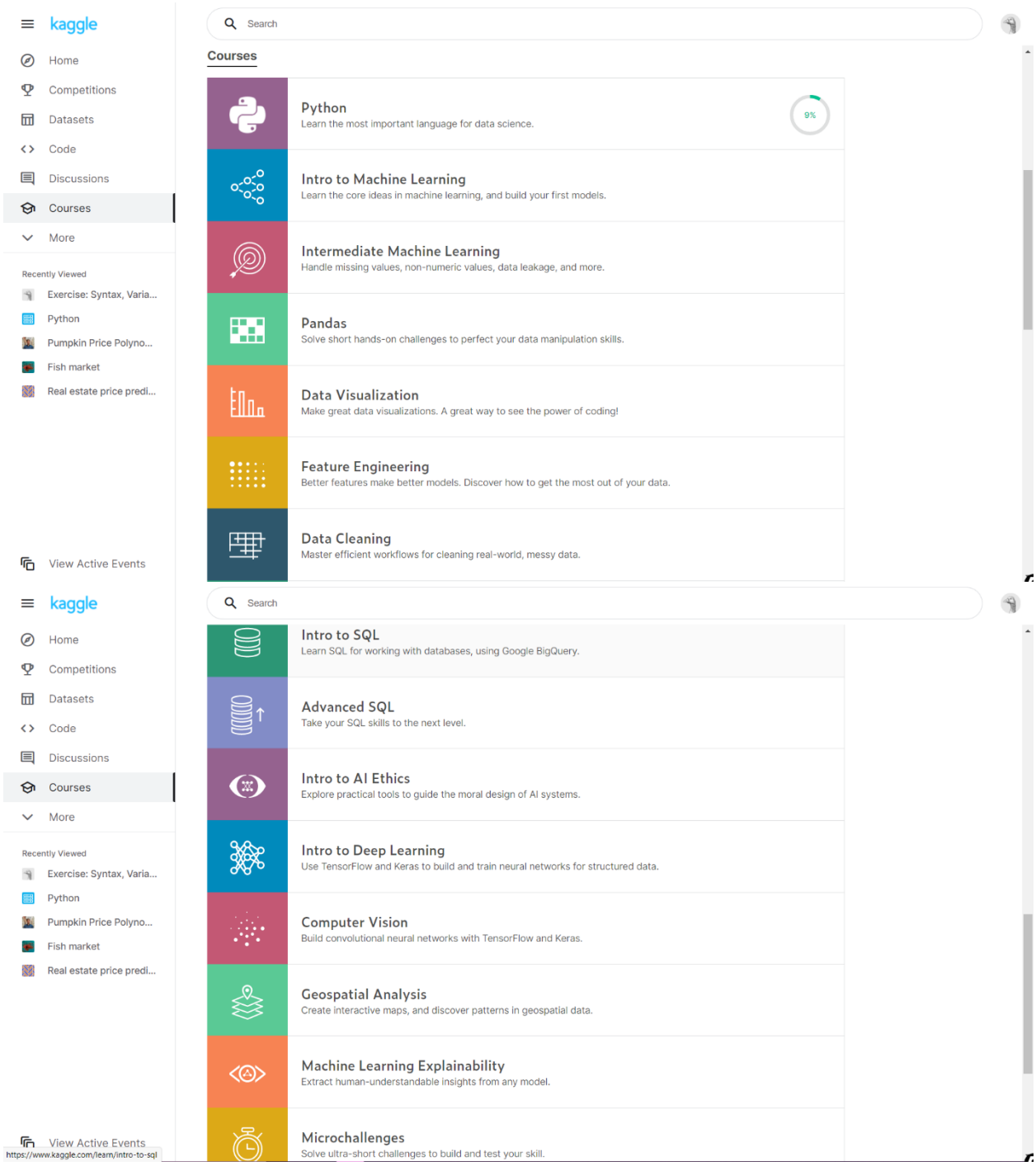


4 основные вещи, которые делают Kaggle **«САМЫМ ЛУЧШИМ».**

1. Доступны бесплатные курсы и бесплатные сертификаты.

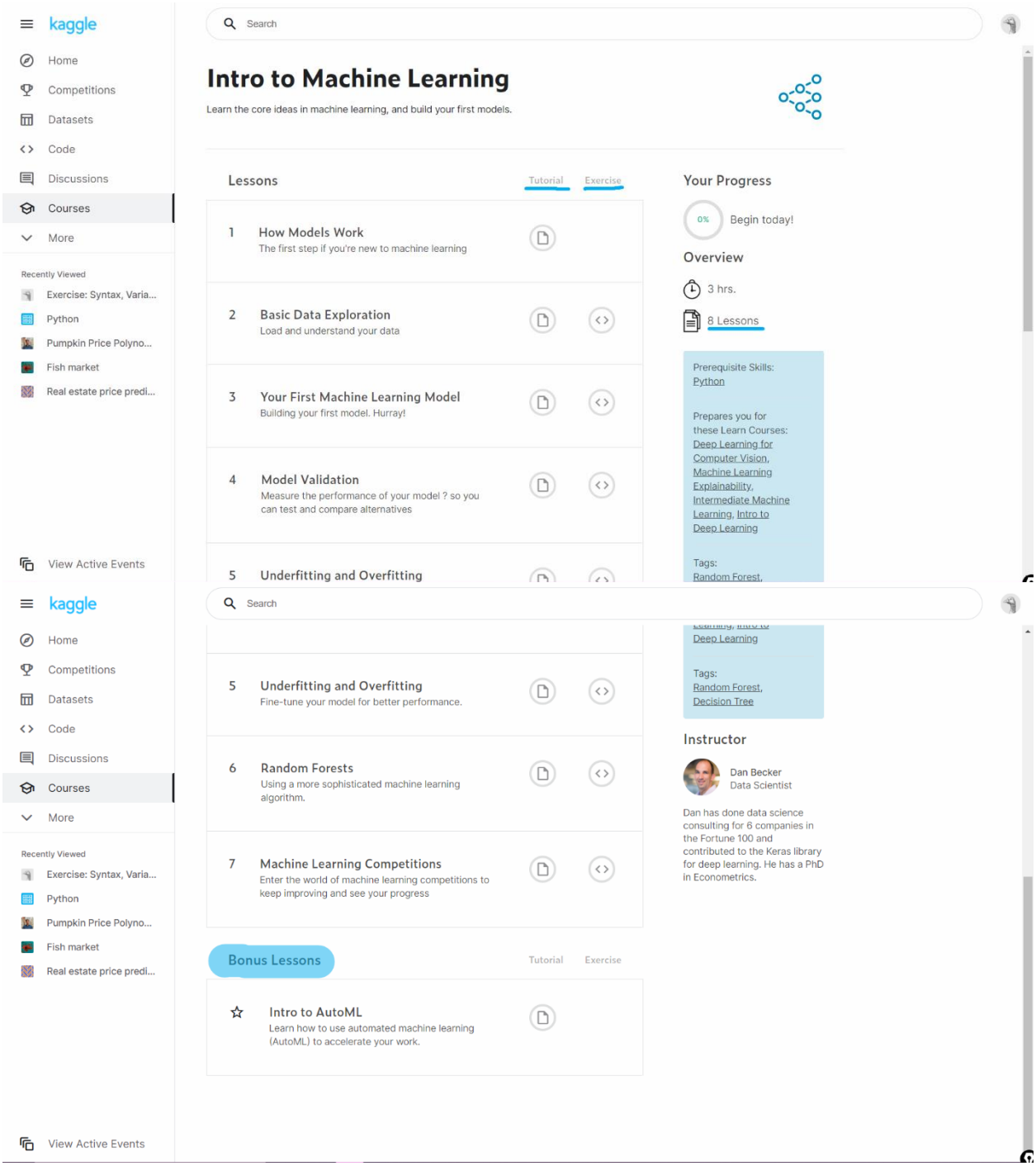
Существует так много курсов по разным областям машинного обучения и науки о данных. После каждого урока доступны не только курсы, но и практические тетради (тетради с упражнениями) для практического изучения темы. Для получения бесплатного сертификата Kaggle необходимо выполнить все задания и упражнения.





Есть еще несколько курсов, но с помощью этого я хотел показать вам, что на этих курсах существует такое разнообразие тем, что вам не нужно никуда идти, в любое время, чтобы почувствовать себя потерянным по какой-либо теме или проблеме, получить помощь отсюда.

Позвольте мне показать вам, как, как выглядят эти курсы на одном примере:

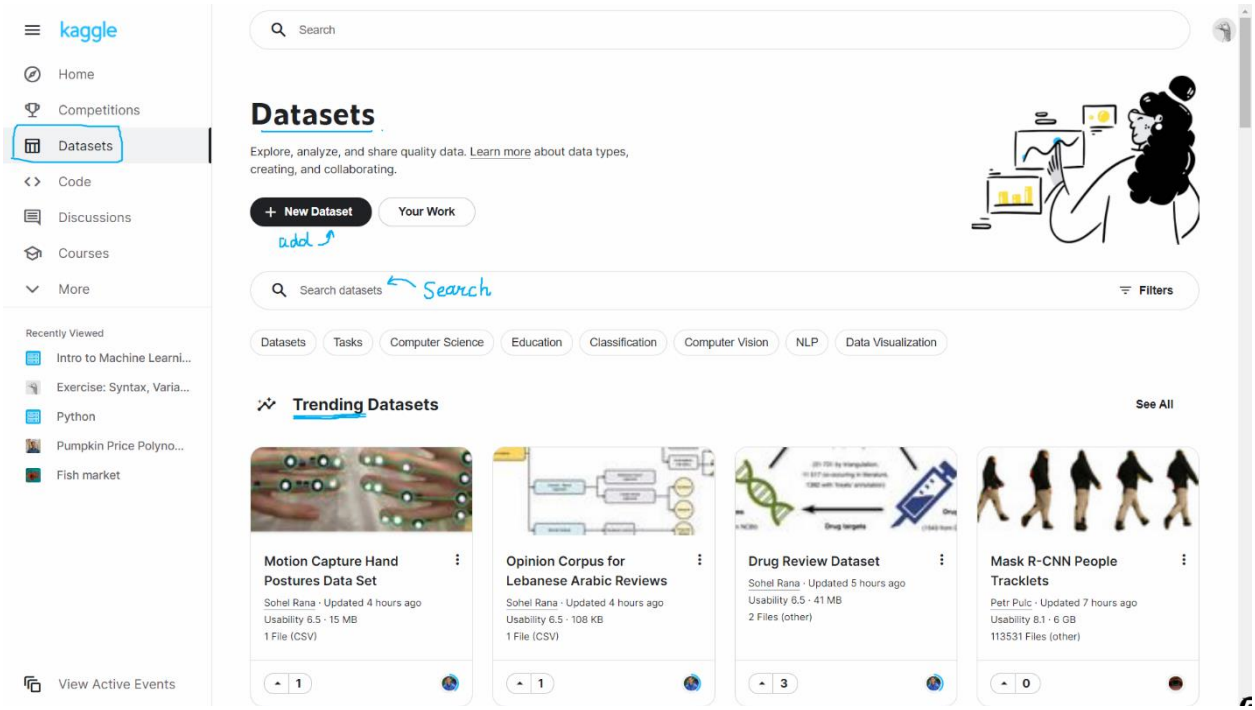


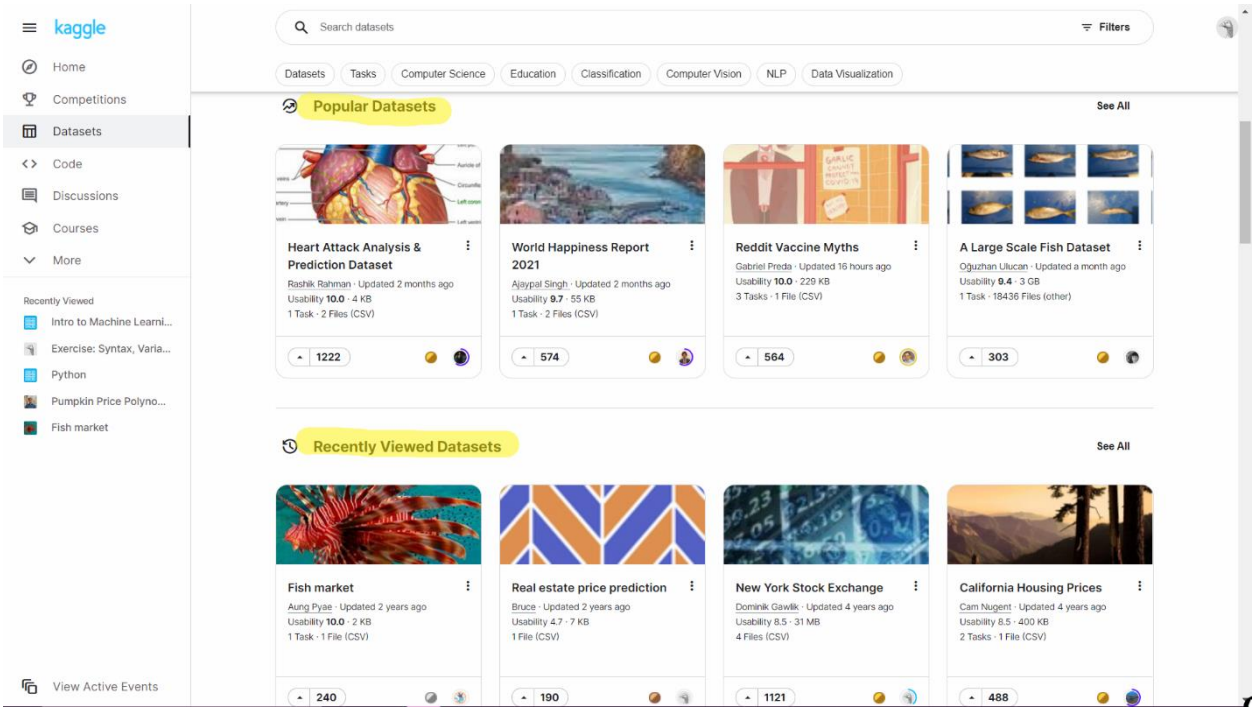
В конце каждого курса есть один дополнительный урок, который отличается по содержанию, но похож на вариант использования и понимание курса. В основном они включают какую-то известную или / или важную тему. Здесь у нас есть AutoML (от Google) для автоматизации машинного обучения.

2. Огромная коллекция общедоступных / предоставленных наборов данных для практики / работы над

Для решения любых задач в области науки о данных, машинного обучения или глубокого обучения нам нужны данные, и большую часть времени их большую часть времени. Вместо того, чтобы просматривать разные сайты для разных типов / размеров наборов данных, Kaggle предоставляет общее место для огромной коллекции всех этих наборов данных. Вы можете одним

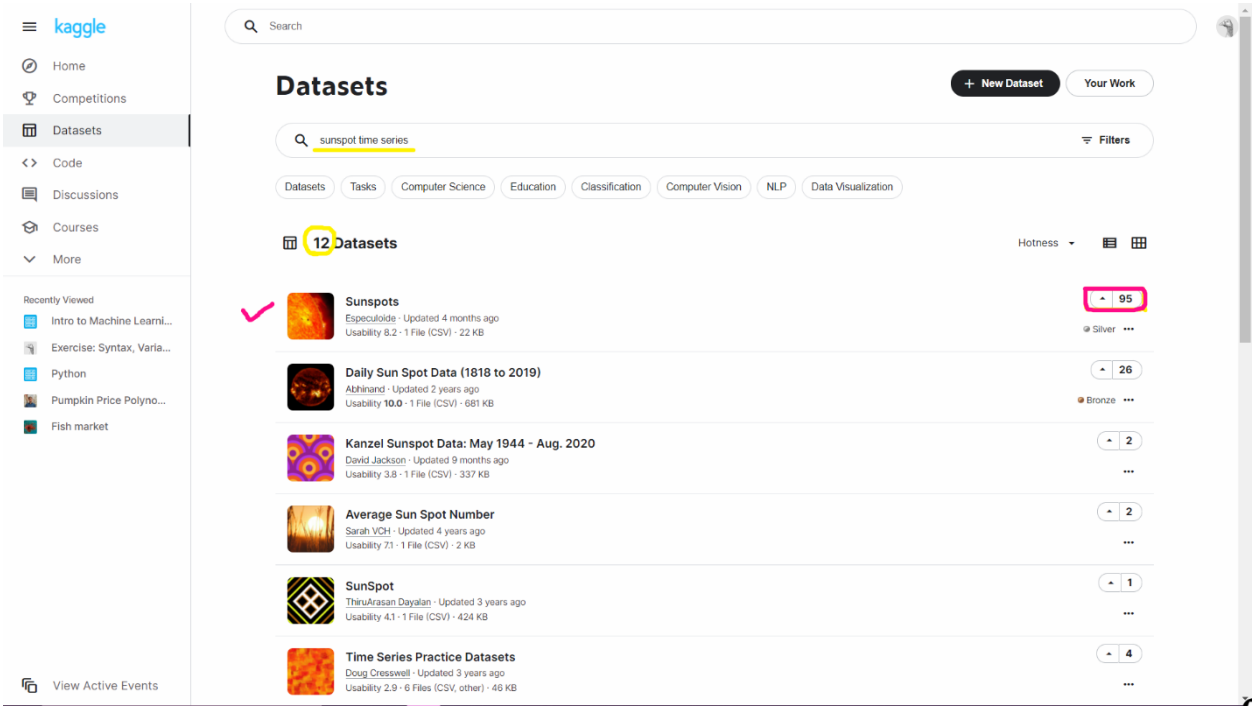
щелчком мыши отказаться от их использования. Они чрезвычайно просты в использовании.





Вот что вы увидите, щелкнув «Наборы данных» на панели навигации. Вы можете искать определенный набор данных, импортировать / вносить свой собственный набор данных в сообщество или изучать или начать работу над набором данных, показанным на этой странице. (Трендовые наборы данных, популярные наборы данных, недавно просмотренные наборы данных)

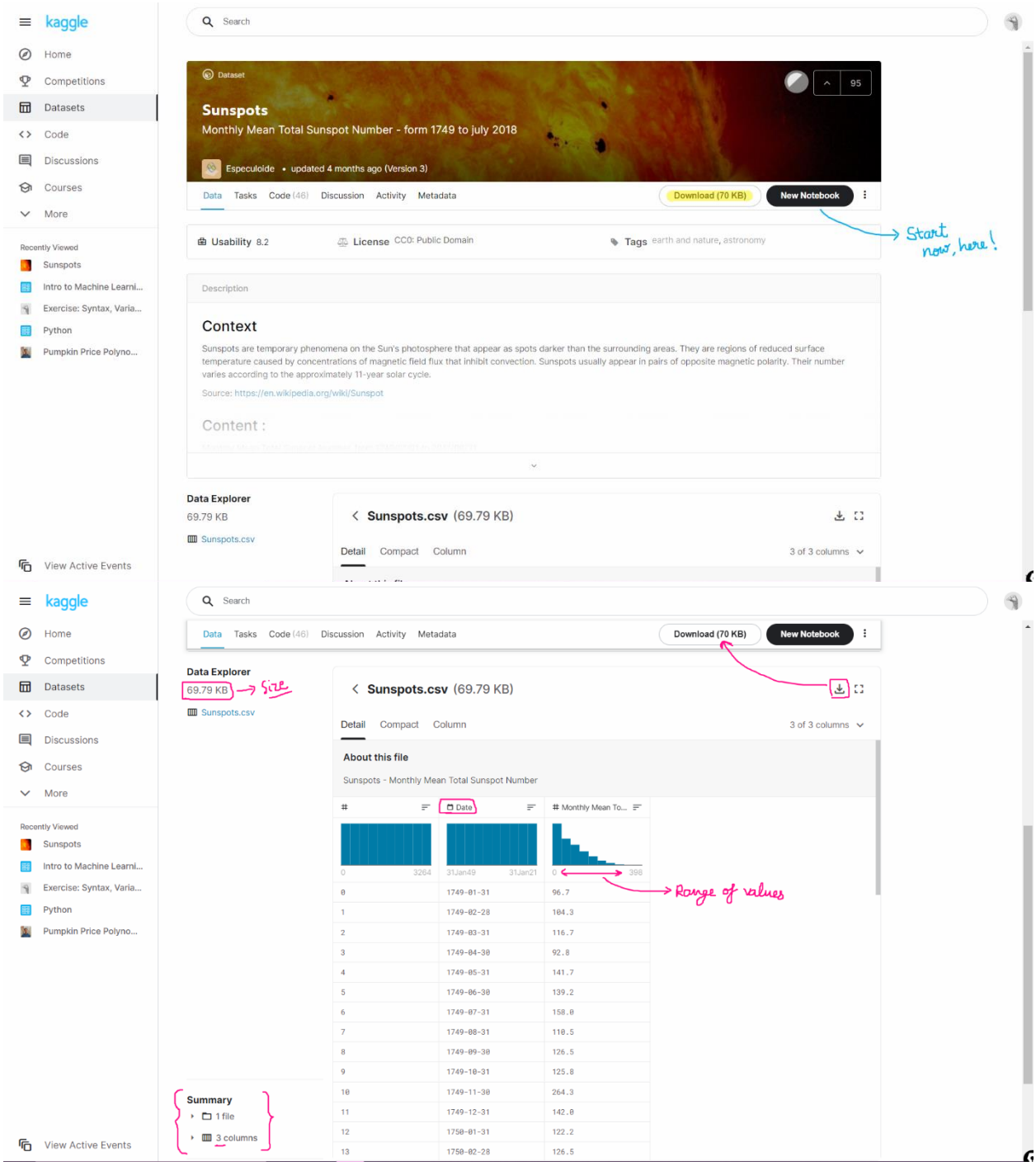
Для демонстрации я буду искать определенный набор данных («Набор данных солнечных пятен»). Посмотрим, как это выглядит.

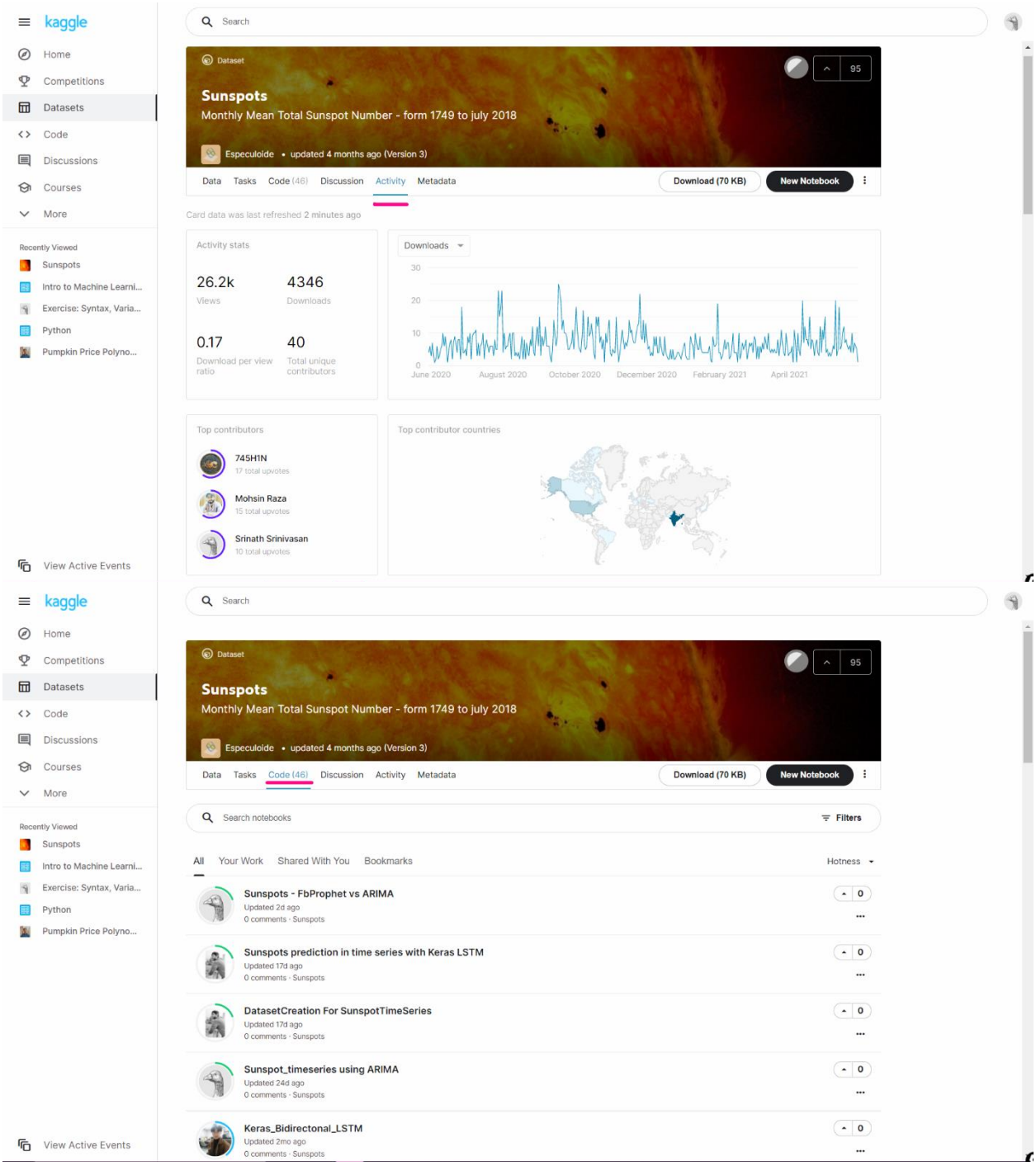


Число в красном выделении - это количество людей, проголосовавших за наиболее релевантный / понравившийся вариант. Давайте изучим и подробно рассмотрим этот набор данных.

Есть много вещей, которые мы можем использовать, чтобы узнать больше об этих данных и немедленно начать работу.

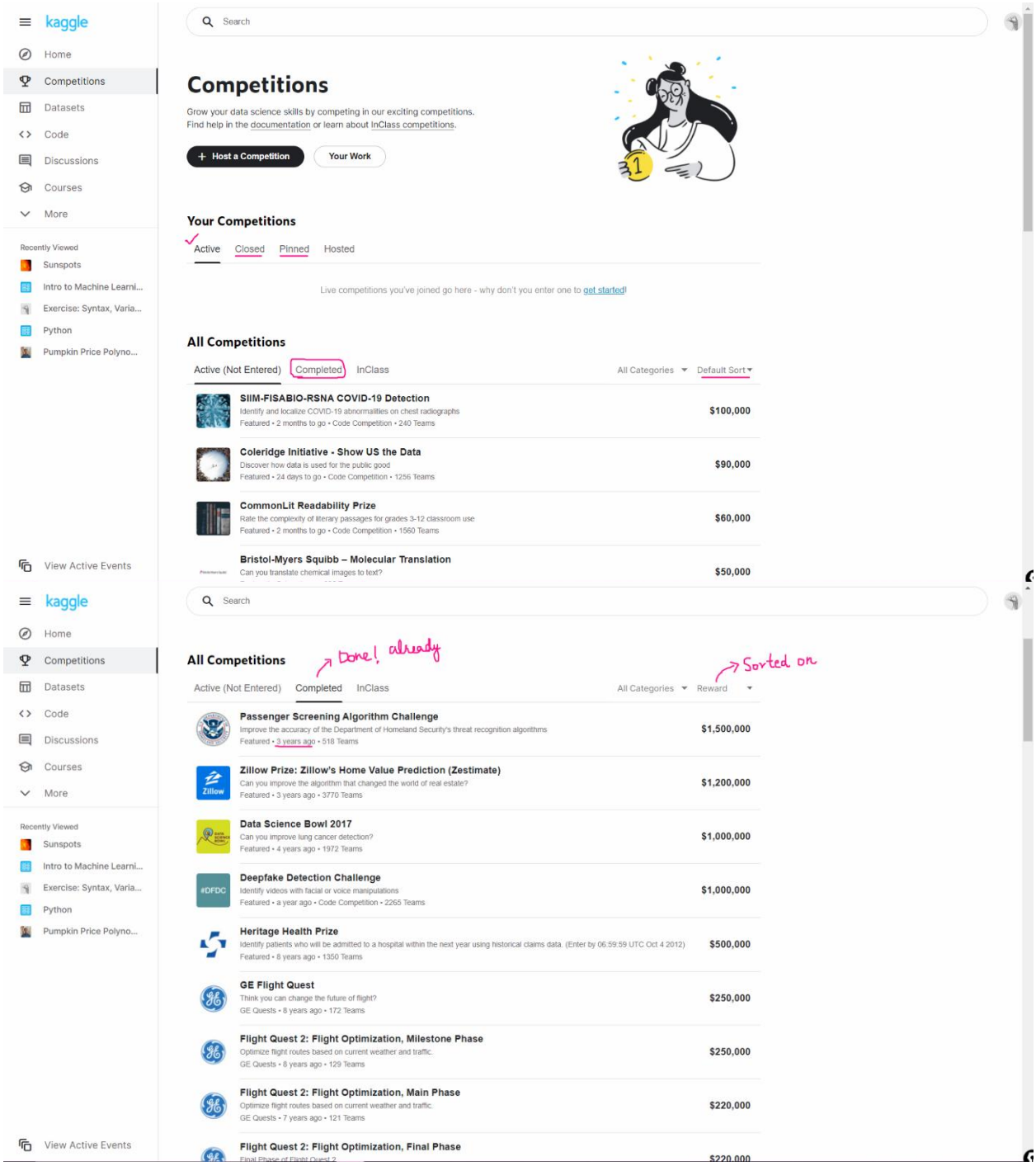
- Вы можете скачать набор данных,
- создайте новый блокнот Kaggle с уже загруженным в него набором данных.
- Некоторые подробности о столбцах внутри данных.
- Действия, связанные с этими данными.
- И последнее, но не менее важное: все записные книжки, созданные и опубликованные до этой даты, используют эти данные.


kaggle
Search
Home
Competitions
Datasets
Code
Discussions
Courses
More
Recently Viewed
Sunspots
Intro to Machine Learni...
Exercise: Syntax, Varia...
Python
Pumpkin Price Polyno...
View Active Events
Dataset
Sunspots
Monthly Mean Total Sunspot Number - form 1749 to july 2018
Especuloide • updated 4 months ago (Version 3)
Data Tasks Code (46) Discussion Activity Metadata
Download (70 KB)
New Notebook
Start now, here!
Usability 8.2
License CC0: Public Domain
Tags earth and nature, astronomy
Description
Context
Sunspots are temporary phenomena on the Sun's photosphere that appear as spots darker than the surrounding areas. They are regions of reduced surface temperature caused by concentrations of magnetic field flux that inhibit convection. Sunspots usually appear in pairs of opposite magnetic polarity. Their number varies according to the approximately 11-year solar cycle.
Source: https://en.wikipedia.org/wiki/Sunspot
Content :
Data Explorer
69.79 KB
Sunspots.csv
Sunspots.csv (69.79 KB)
Detail Compact Column
3 of 3 columns
Data Tasks Code (46) Discussion Activity Metadata
Download (70 KB)
New Notebook
Data Explorer
69.79 KB
size
Sunspots.csv
Sunspots.csv (69.79 KB)
Detail Compact Column
3 of 3 columns
About this file
Sunspots - Monthly Mean Total Sunspot Number
Date
Monthly Mean To...
Range of values
Summary
1 file
3 columns



3. Соревнования по науке о данных / машинному обучению / глубокому обучению

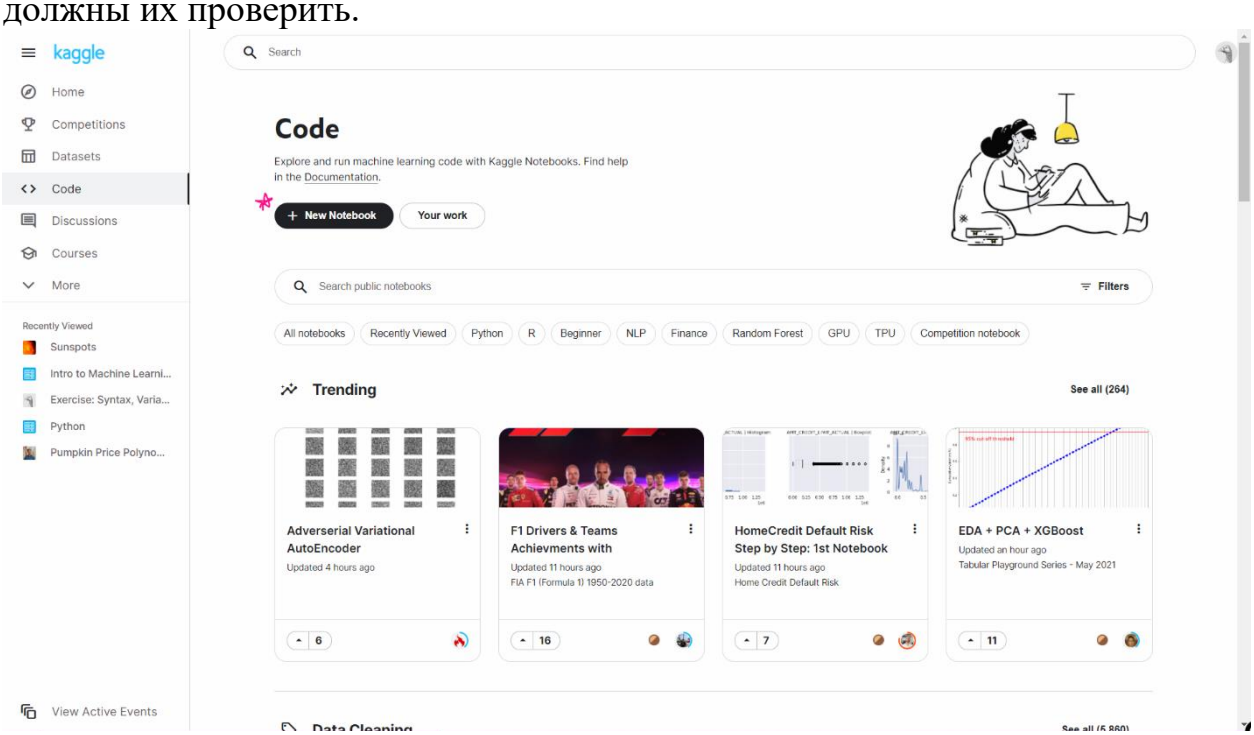
Хотя я не участвовал ни в одном из них, мне нравится, как мы решаем задачу в реальном времени вместе с сообществом Kaggle и выигрываем потрясающие денежные призы (если участвуем в этом конкретном соревновании). Я определенно хочу участвовать когда-нибудь в ближайшее время, надеюсь, изображения мотивируют вас. Необязательно, чтобы это могли делать только крупные компании или богатые предприятия. Вы тоже можете это сделать. Есть определенные протоколы, которым нужно следовать, и вуаля, у вас есть собственное соревнование, организованное вами.



4. Блокноты Kaggle (Код)

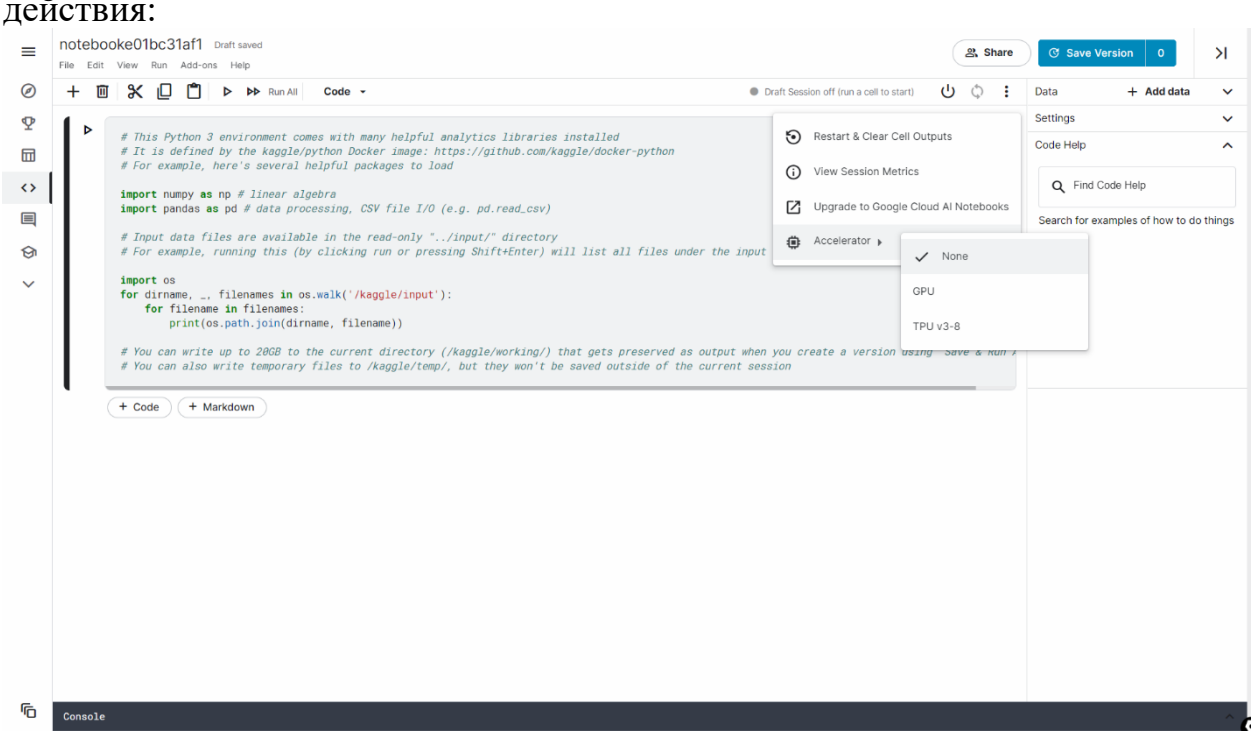
Для любой задачи, связанной с наукой о данных или информатикой, мы должны написать хоть какой-то код. Kaggle предоставляет нам свою собственную среду Notebook с определенным ограничением того, сколько мы можем хранить на них (в совокупности для каждой учетной записи), сколько часов доступного графического процессора и сколько часов доступно TPU. Они полностью интегрированы со всеми сервисами Kaggle и могут использоваться независимо, как любая другая среда для ноутбуков (Datalore, Google Colab, Jupyter и т.д.), что означает, что вы можете использовать их для своей практики, соревнований Kaggle, курсов Kaggle, анализа некоторых

Kaggle / или набор данных, не относящийся к Kaggle, и многое другое. Вы должны их проверить.

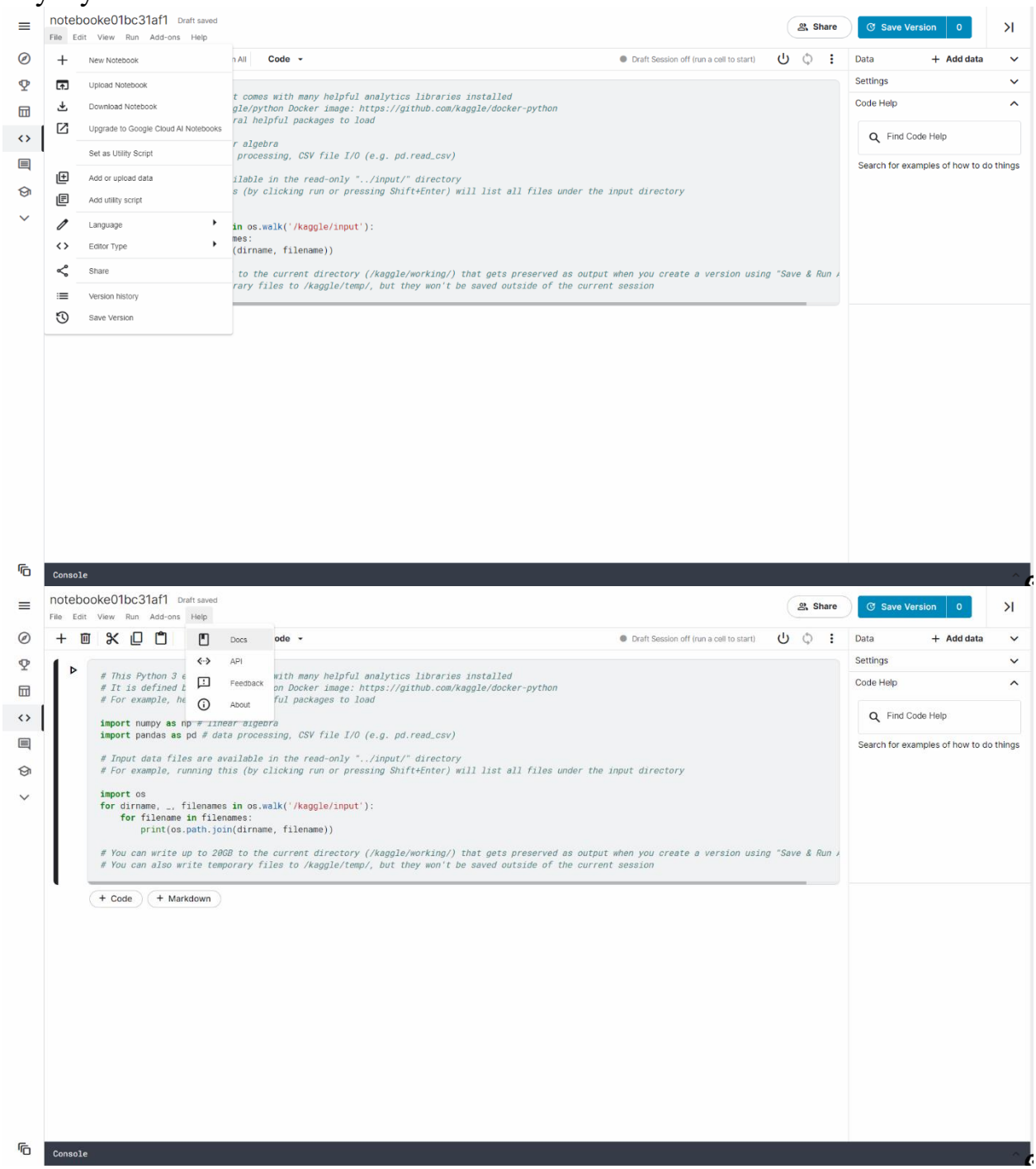


Нажав на эту черную кнопку, вы создаете свою записную книжку или открываете чужую записную книжку, которую хотите прочитать и изучить / сравнить. Все эти видимые записные книжки являются общедоступными, что означает, что ваши записные книжки не будут видны никому, если вы сами этого не сделаете.

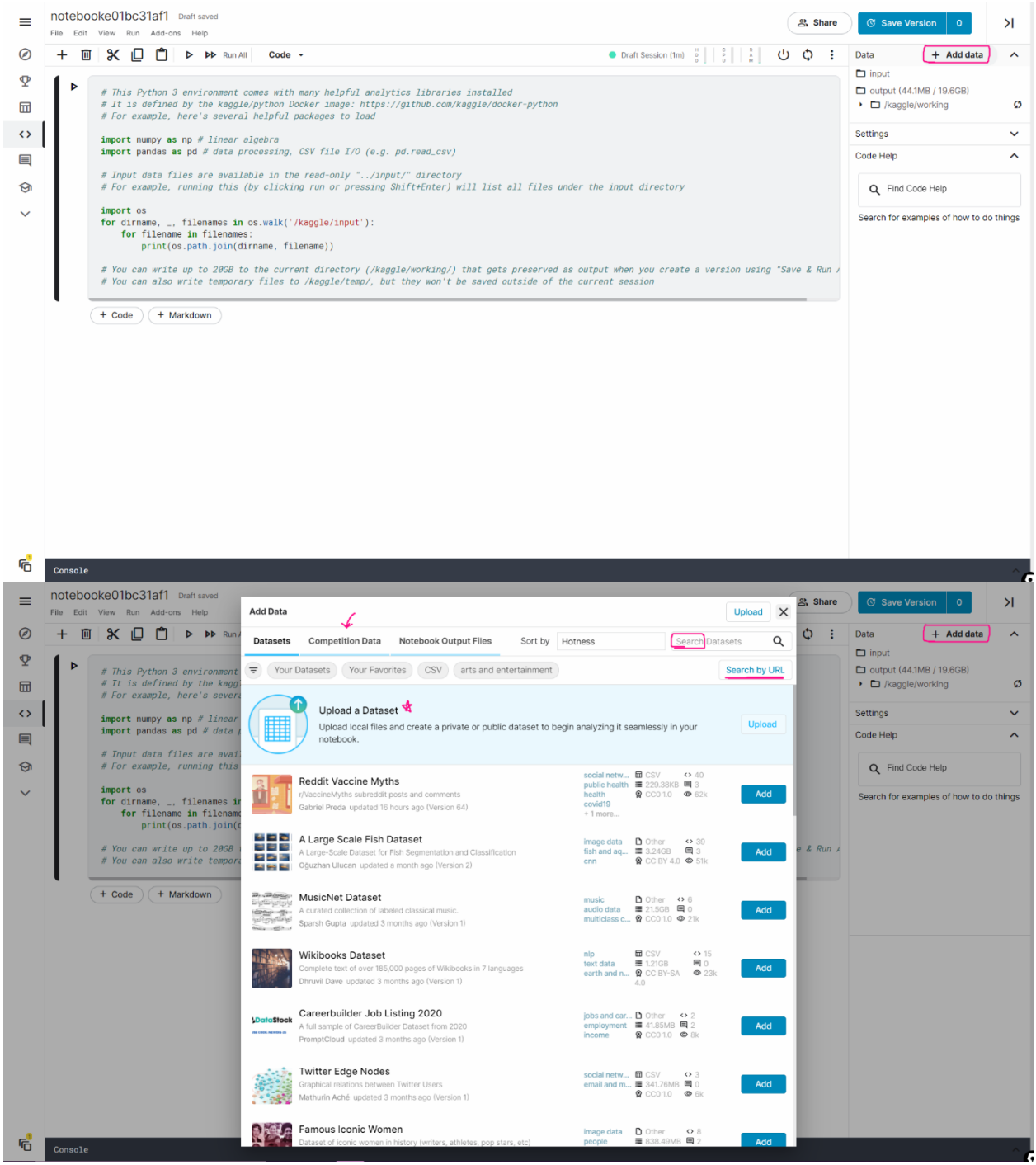
Чтобы переключиться с CPU на GPU или TPU, выполните следующие действия:



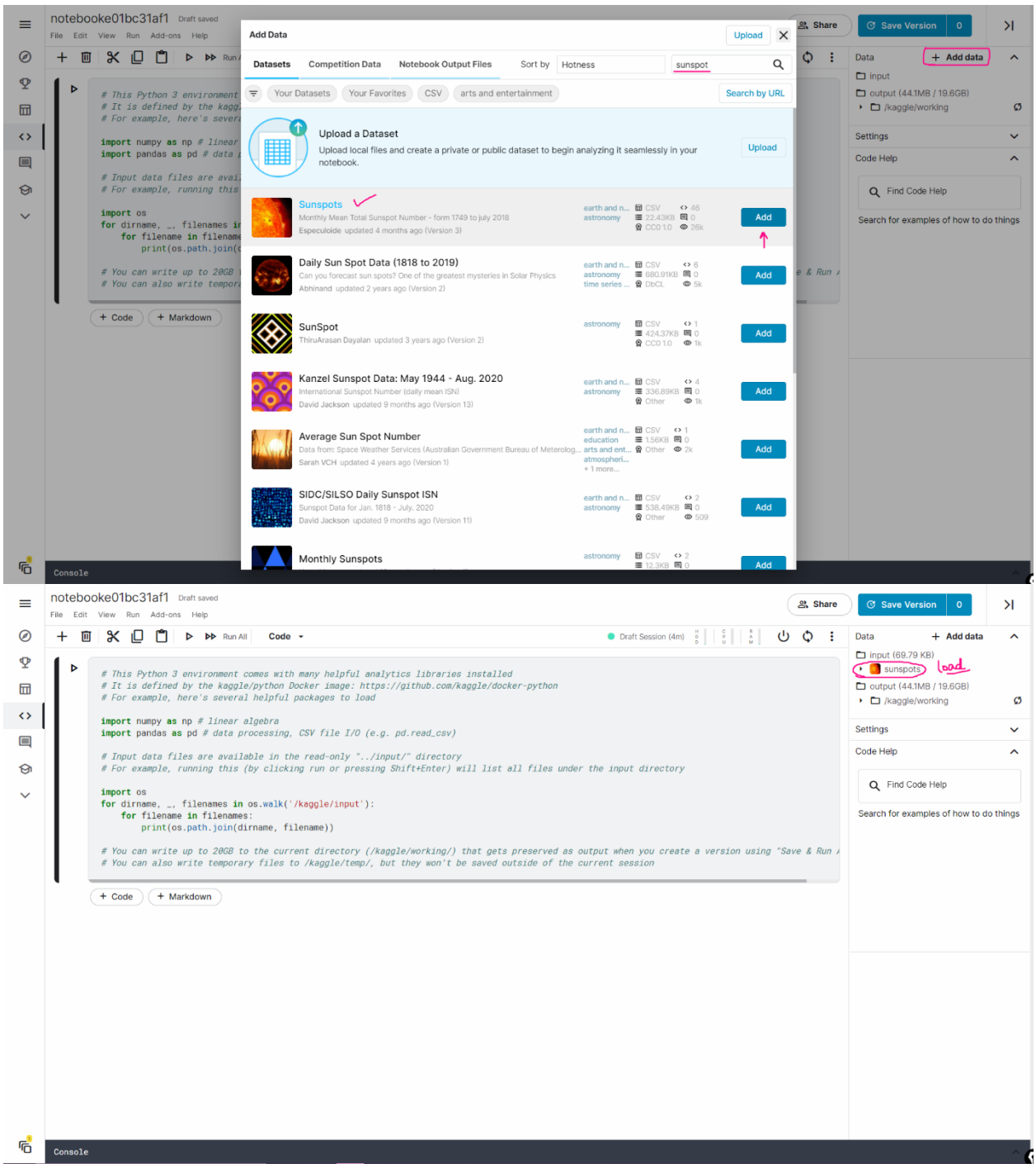
Это большинство функциональных возможностей, доступных вам в этом ноутбуке:



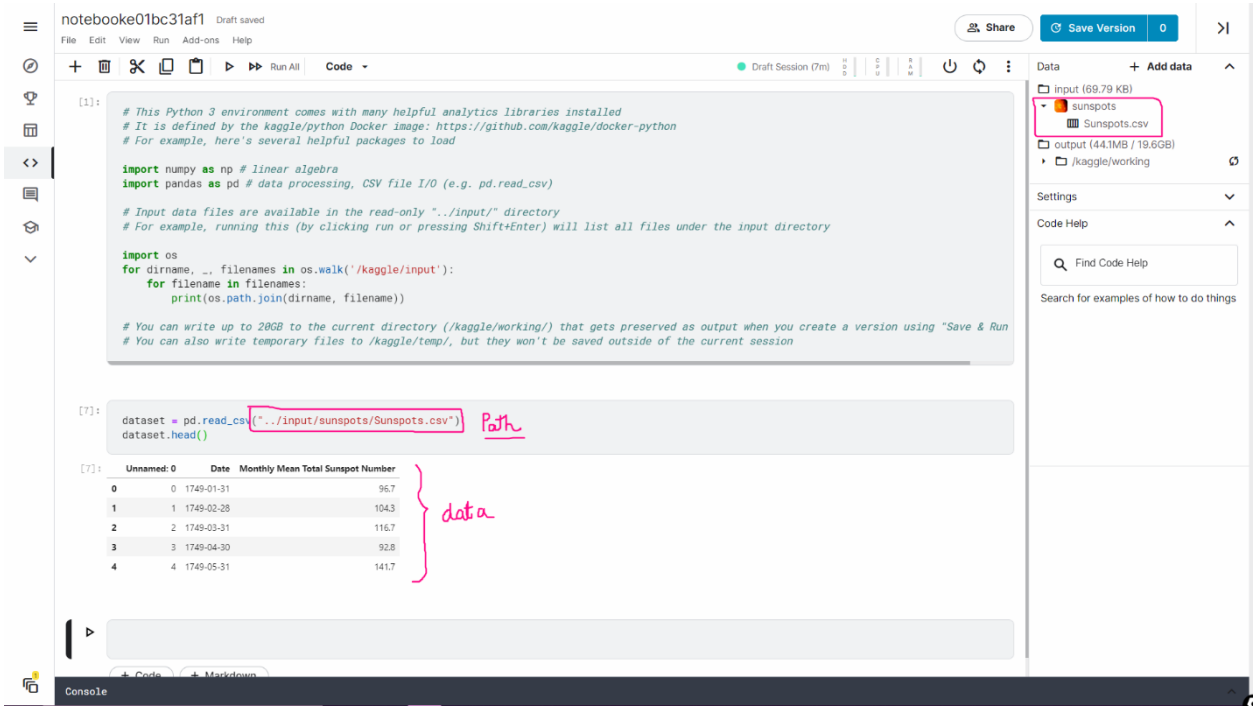
Давайте посмотрим, как использовать их с данными (импортированными / взятыми непосредственно из Kaggle / загруженными с URL-адреса и т.д.) и начнем работать над вашими задачами в области науки о данных.



Здесь покажем, как использовать набор данных «Солнечные пятна», который мы видели ранее. Начните с поиска.



Теперь данные успешно загружены. Выделение на изображении выше - это каталог, в котором он хранится. Давайте посмотрим на небольшой код *pandas*, чтобы узнать, как импортировать набор данных.



Последнее, что вы можете сделать после завершения своего проекта / работы, - это поделиться им с сообществом на Kaggle. Это важный шаг, потому что, делясь своими идеями, своей работой, мы расширяем возможности, доступные сообществу, и поддерживаем друг друга.

Слева от Большой синей кнопки в правом верхнем углу вы увидите кнопку **«Поделиться»** . Щелкните по нему и выберите «Публичный» в раскрывающемся меню.