

Google Colab: все, что вам нужно знать

Описание

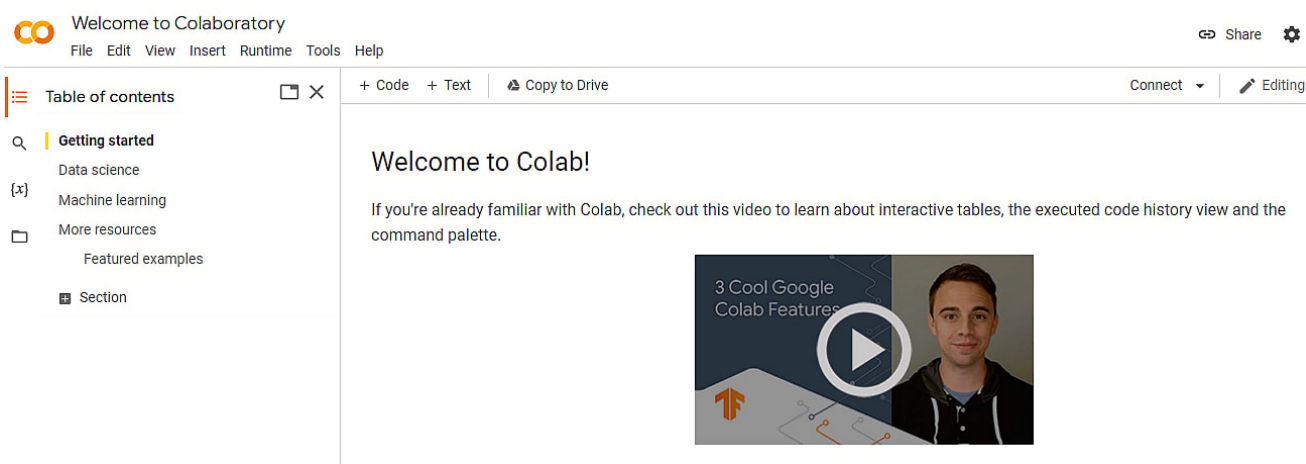
Google Colab делает науку о данных, глубокое обучение, нейронные сети и машинное обучение доступными для индивидуальных исследователей, которые не могут позволить себе дорогостоящую вычислительную инфраструктуру. Машинное обучение и наука о данных – две новые технологии, в которых хотят преуспеть все компьютерщики нового поколения. Существует множество онлайн-курсов, бесплатных лекций и онлайн-руководств по машинному обучению и науке о данных. Но практическая работа над проектами становится ограничением, поскольку для таких нагрузок нужны высокопроизводительные ПК. Ответ на этот вопрос – Google Colaboratory или, сокращенно, Colab. Продолжайте читать, чтобы получить окончательный обзор Google Colab.

Что такое Google Colab?

Google Colab – это Jupyter Notebook-подобный продукт от Google Research. Разработчик программ на Python может использовать этот блокнот для написания и выполнения произвольных программных кодов на Python, просто используя веб-браузер. В двух словах, Google Colab – это размещенная в облаке версия Jupyter Notebook. Чтобы использовать Colab, вам не нужно устанавливать среду выполнения или модернизировать компьютерное оборудование, чтобы соответствовать требованиям Python к интенсивной нагрузке на CPU/GPU. Кроме того, Colab предоставляет вам бесплатный доступ к вычислительной инфраструктуре, такой как хранилище, память, вычислительные мощности, графические процессоры (GPU) и тензорные процессоры (TPU). Google специально

разработала этот облачный инструмент для кодирования на Python, учитывая потребности программистов машинного обучения, аналитиков больших данных, ученых, изучающих ИИ, и тех, кто изучает Python. Самое приятное – это один блокнот для всех компонентов, необходимых для представления законченного проекта по машинному обучению или науке о данных руководителям или спонсорам программы. Например, ваш блокнот Colab может содержать исполняемые коды, живые коды Python, текст, HTML, LaTeX, изображения, визуализацию данных, диаграммы, графики, таблицы и многое другое.

Как работает Google Colab?



Google Colab – это просто онлайн-представление Jupyter Notebook. В то время как Jupyter Notebook требует установки на компьютер и может использовать только локальные ресурсы машины, Colab – это полноценное облачное приложение для кодирования на Python. Вы можете писать коды на языке Python с помощью Colab в браузерах Google Chrome или Mozilla Firefox. Вы также можете выполнять эти коды в браузере без использования среды выполнения или интерфейса командной строки. Кроме того, вы можете придать своему блокноту с проектами на Python профессиональный вид, добавив математические уравнения, графики, таблицы, изображения и другую графику. Кроме того, вы можете создавать визуализации данных на Python, а Colab будет отображать код в виде визуальных активов. Кроме того, Colab позволяет повторно использовать файлы Jupyter Notebook с GitHub. Кроме того, вы можете импортировать совместимые проекты машинного обучения и науки о данных из других источников. Colab эффективно обрабатывает импортированные активы, чтобы отобразить чистый и безошибочный Python-код.

Лучшие возможности Google Colab

Графические процессоры и TPU

Пользователи Free Colab получают бесплатный доступ к режимам работы GPU и TPU на срок до 12 часов. Время работы GPU оснащено процессором Intel Xeon CPU @2,20 ГГц, 13 ГБ оперативной памяти, ускорителем Tesla K80 и 12 ГБ GDDR5 VRAM. Время выполнения TPU состоит из процессора Intel Xeon CPU @2,30 ГГц, 13 ГБ оперативной памяти и облачного TPU с вычислительной мощностью 180 терафлопс. С Colab Pro или Pro+ вы сможете вводить в эксплуатацию больше CPU, TPU и GPU в течение более чем 12 часов.

Совместное использование ноутбука

Записная книжка с кодом Python никогда не была такой доступной до появления Colab. Теперь вы можете создавать ссылки для совместного доступа к файлам Colab, сохраненным на вашем Google Диске. Теперь поделитесь ссылкой с сотрудником, который хочет работать с вами. Кроме того, вы можете приглашать программистов к сотрудничеству с помощью электронной почты Google.

Установка специальной библиотеки

Colab позволяет устанавливать библиотеки, не относящиеся к Лаборатории (AWS S3, GCP, SQL, MySQL и т. д.), которые недоступны в фрагментах кода. Все, что вам нужно сделать, это добавить однострочный код со следующими кодовыми префиксами:

```
!pip install (??????: !pip install matplotlib-venn)
```

```
!apt-get install (??????: !apt-get -qq install -y libfluidsynth1)
```

Предустановленные библиотеки

Google Colab предлагает множество предустановленных библиотек, чтобы вы могли импортировать нужную библиотеку из фрагментов кода. К таким библиотекам относятся NumPy, Pandas, Matplotlib, PyTorch, TensorFlow, Keras и другие библиотеки ML.

Совместное кодирование



Совместное кодирование незаменимо для групповых проектов. Оно помогает команде завершить этапы работы раньше предполагаемого срока. Если вашей команде требуется совместная работа в режиме реального времени над проектами по ML и науке о данных, Google Collaborative – это именно то, что нужно. Просто

отправьте ссылку на редактируемый текст соавторам или пригласите их для группового кодирования. Весь блокнот Python автоматически обновляется по мере того, как команда кодирует, и вы получаете ощущение работы в общих листах или документах Google.

Облачное хранилище

Google Colab использует квоту вашего хранилища Google Drive для сохранения файлов. Таким образом, вы можете возобновить работу с любого компьютера, на котором есть доступ к вашему аккаунту Google Drive. Облачные хранилища также выполняют функцию резервного копирования данных на случай любых катастроф.

Интеграция с GitHub

Вы можете связать свой аккаунт GitHub с Google Colab, чтобы легко импортировать и экспортировать файлы кода. Для импорта вы можете нажать Ctrl+O и перейти на вкладку GitHub, чтобы получить файлы кода. Для отправки файлов на GitHub, наоборот, достаточно нажать на Save a copy to GitHub в меню File.

Несколько источников данных

Google Colaboratory поддерживает различные источники данных для ваших проектов по обучению ML и AI. Например, вы можете импортировать данные с локальной машины, подключить Google Drive к экземпляру Colab, получить удаленные данные и клонировать репозиторий GitHub в Colab.

Автоматический контроль версий

Как и в Google Sheets и Docs, в Google Colab также есть исчерпывающий трекер истории. Модуль отслеживает все изменения, внесенные с момента создания файла. Доступ к журналам можно получить из меню “Файл” и выбрать пункт “История изменений”.

Почему вы должны выбрать Google Colab?



- Google Colaboratory – это облачный инструмент. Вы можете начать создавать фантастические модели ML и data science, используя браузер Chrome.
- Colab является бесплатной системой с ограниченными ресурсами. Однако не стоит ожидать, что вы сможете хранить свои модели искусственного интеллекта или машинного обучения бесконечно долго на бесплатной инфраструктуре Colab.
- Если вы знаете, как работать с Jupyter, вам не нужно проходить кривую обучения в Google Colaboratory.
- Бесплатный доступ к графическим процессорам и TPU для работы с обширными моделями научного анализа данных и машинного обучения.
- Он поставляется с предустановленными и популярными библиотеками для работы с данными.
- Кодеры могут легко делиться блокнотом с коллегами для кодирования в режиме реального времени.
- Поскольку Google размещает ноутбук в Google Cloud, вам не нужно беспокоиться о контроле версий кодовых документов и их хранении.
- Легко интегрируется с GitHub.
- Вы можете обучать ИИ с помощью изображений.

- Вы также можете обучать модели на аудио и тексте.
- Исследователи также могут запускать программы TensorFlow в Colab.

Как использовать Google Colab



Вы можете использовать Google Colaboratory, если отвечаете следующим минимальным требованиям:

- Аккаунт Google, чтобы пользоваться всеми удобствами Colab.
- Компьютер, на котором может работать последняя версия браузера Google Chrome или Mozilla Firefox
- Google рекомендует Chrome для Colab.
- Примите правила и условия использования данных Google.

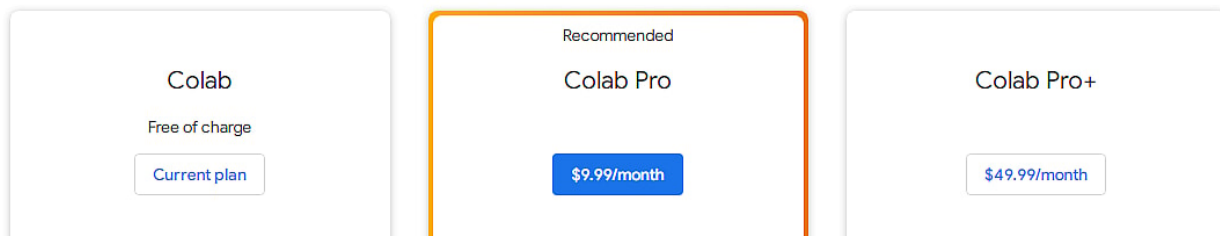
Получить доступ к Google Colaboratory можно с ее официального сайта. Colab бесплатна, однако ограниченное количество ресурсов не всегда гарантировано. Если вам нужна большая скорость и возможности обработки с гарантированными ресурсами, вы можете приобрести Colab Pro или Pro+.



Choose the Colab plan that's right for you

Whether you're a student, a hobbyist or a ML researcher, Colab has you covered

[Restrictions apply. Learn more here.](#)



Некоторые модели для науки о данных и машинного обучения, подходящие для Colab, вы можете найти в Google Seedbank.

Различия между Google Colab и Jupyter Notebook

Код Док Особенности	Лаборатория Google	Блокнот Jupyter
Мгновенный просмотр кодовых файлов	Да	Нет
Совместное использование кодовых документов	Да	Нет
Установленные библиотеки	Да	Нет
Облачный хостинг	Да	Нет
Синхронизация файлов	Да	Нет

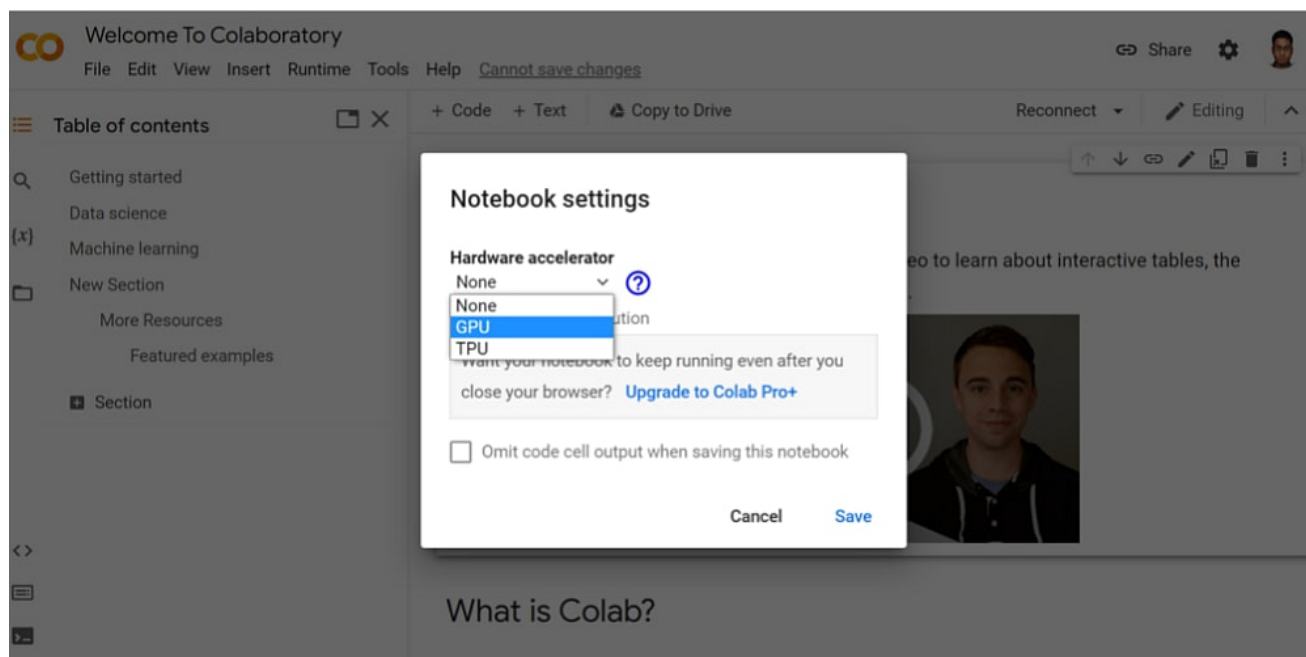
#1. Colab не требует установки программного обеспечения на локальной машине. Напротив, Jupyter Notebook требует установки программного обеспечения и ресурсов локальной машины для вычислений.

#2. Поскольку Colab работает в облаке, вы получаете автоматический контроль версий. Кроме того, Google Drive сохраняет блокнот Python автоматически. В отличие от Jupyter Notebook, вам нужно периодически сохранять блокнот и управлять контролем версий.

#3. Файлы Colab доступны на Google Диске для резервного копирования. С другой стороны, файлы Jupyter Notebook не резервируются автоматически.

#4. Вы можете отправить файлы Colab кому угодно, даже клиенту, который не является специалистом по исследованию данных. Они могут легко открыть документ в Google Colab и просмотреть содержимое. Получателю не нужно устанавливать программное обеспечение.

Напротив, получателю необходимо установить и запустить Jupyter Notebook, чтобы прочитать ваш проект. Таким образом, обмен этим файлом с клиентами, не занимающимися наукой о данных, становится сложной задачей.

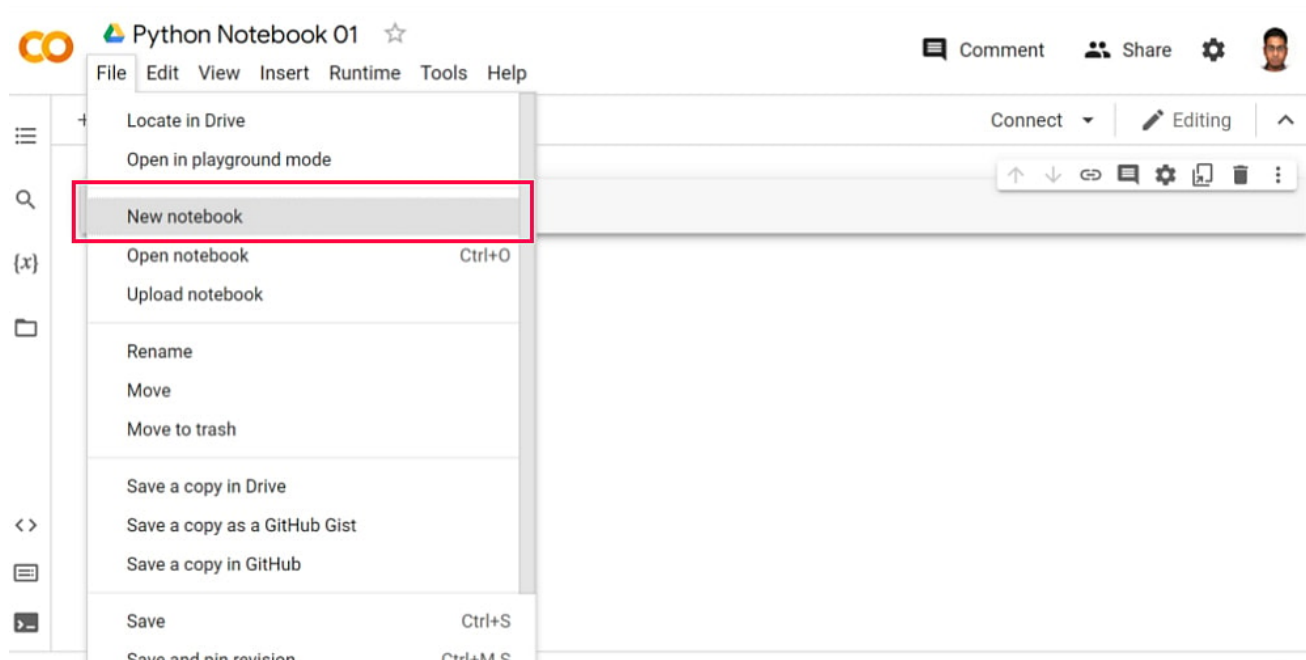


#5. Google Colaboratory поставляется с необходимыми библиотеками для проектов в области науки о данных и машинного обучения. Кроме того, она предоставляет вам определенное количество CPU, RAM, GPU и TPU в облаке. Таким образом, вы экономите время и деньги.

В отличие от этого, при работе с приложением Jupyter Notebook вам нужно найти и установить все библиотеки, необходимые для вашего проекта. Установка такого количества библиотек также потребляет ресурсы CPU, RAM и GPU локальной машины.

Выполнение общих задач в Google Colab

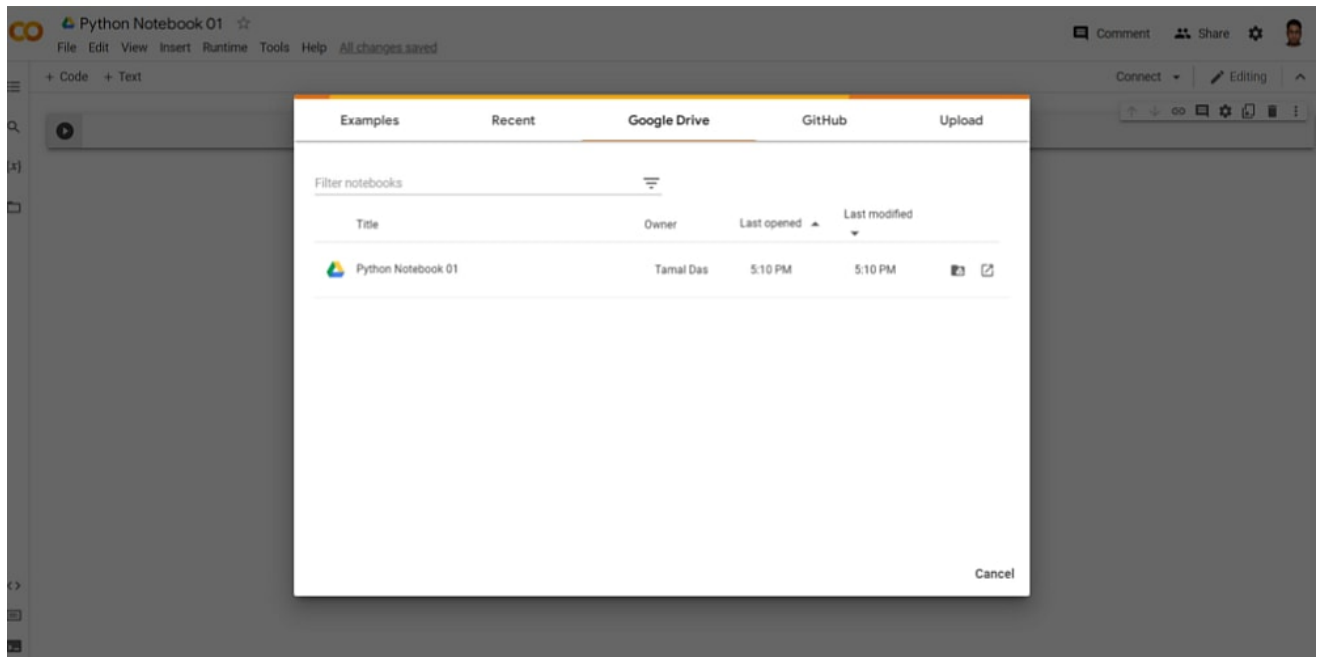
#1. Создайте блокнот



- Зайдите на портал Google Colab и увидите: “Добро пожаловать в Colab!”.
- В верхнем меню выберите пункт **Файл**.
- В контекстном меню Файл выберите **Новый блокнот**.
- Ваш новый блокнот Python готов. Вы можете переименовать файл блокнота.

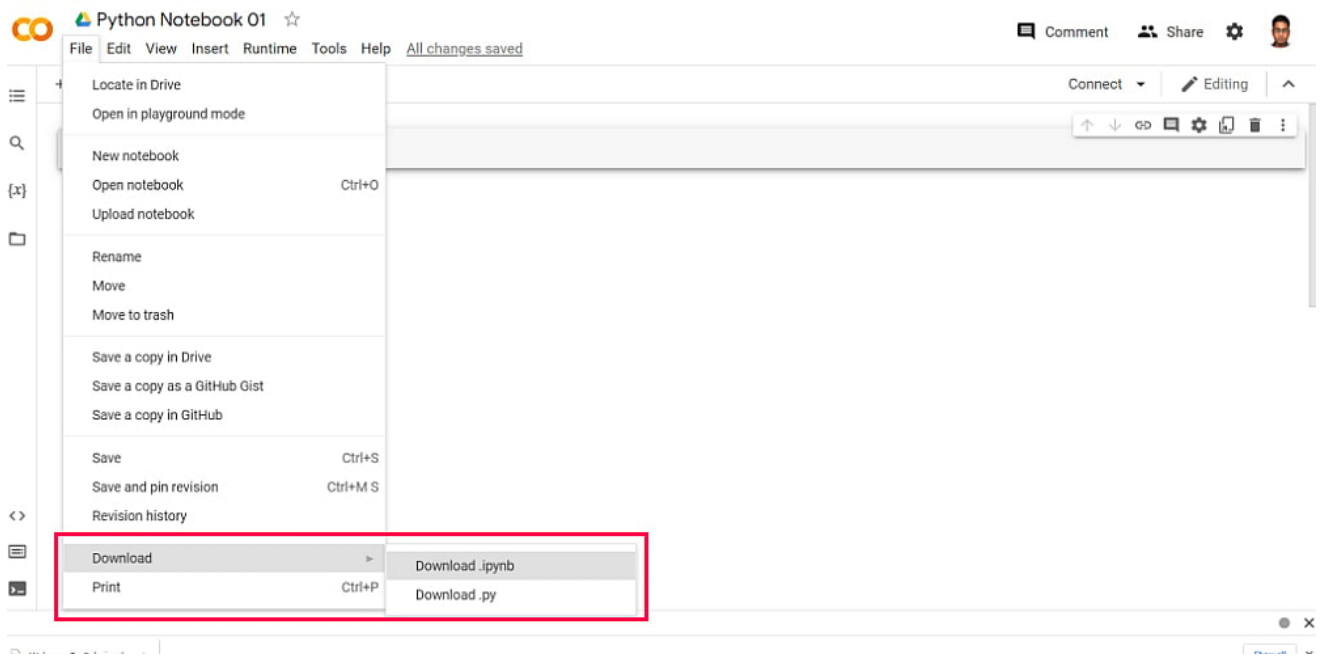
#2. Загрузка и выгрузка файлов

Вы можете загрузить локальные коды Python в Colab, выполнив следующие действия:



- В верхнем меню нажмите **Файл**.
- Откроется контекстное меню с множеством опций.
- Найдите **блокнот Upload** и щелкните на нем.
- Теперь вы увидите наложенную консоль с такими опциями, как Examples, Google Drive, GitHub и Upload.
- Перейдите на любую вкладку и выберите код содержимого, которое вы хотите загрузить.

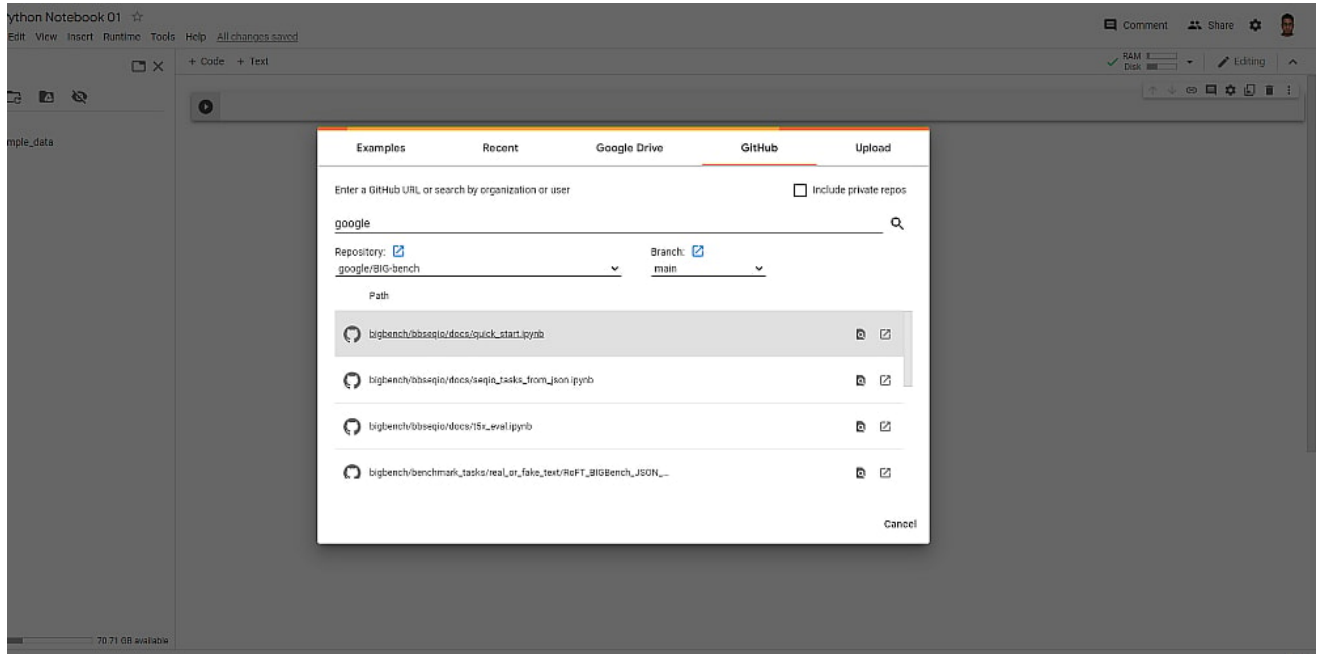
Загрузить готовый или начатый проект также очень просто. **Вот шаги:**



- Нажмите на меню " **Файл** " , расположенное в верхней строке меню.
- Наведите курсор на пункт " **Загрузить** ".
- Откроется контекстное меню с двумя вариантами формата загружаемых файлов: .ipynb и .py.
- Вы можете выбрать предпочтительный формат и загрузить файл.

#3. Доступ к GitHub

Получить доступ к GitHub в Colab проще простого. Вот что вы можете сделать:

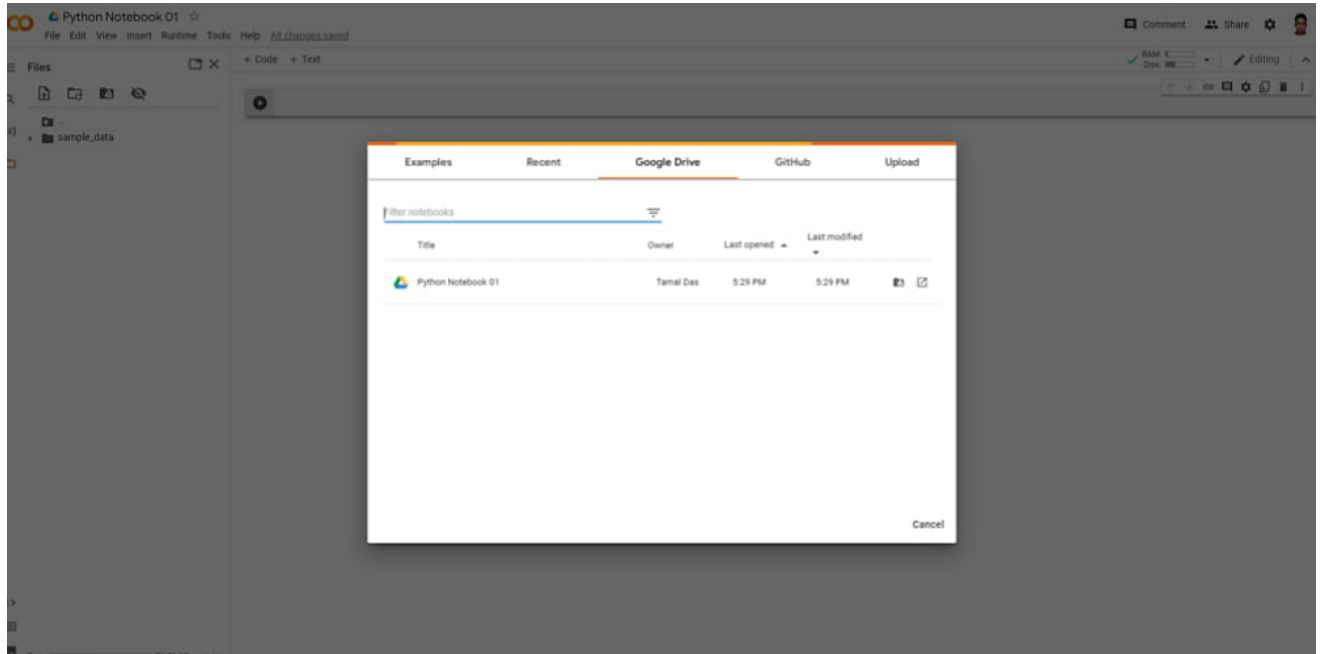


- Нажмите **Файл** в верхней строке меню.
- Выберите пункт **Загрузить блокнот** в контекстном меню.
- Откроется консоль с вкладкой **GitHub**.
- Кроме того, вы можете нажать **Ctrl+O**, чтобы получить доступ к той же консоли.
- Параметры поиска на GitHub: URL GitHub, имя пользователя и название организации.

#4. Доступ к локальным файлам

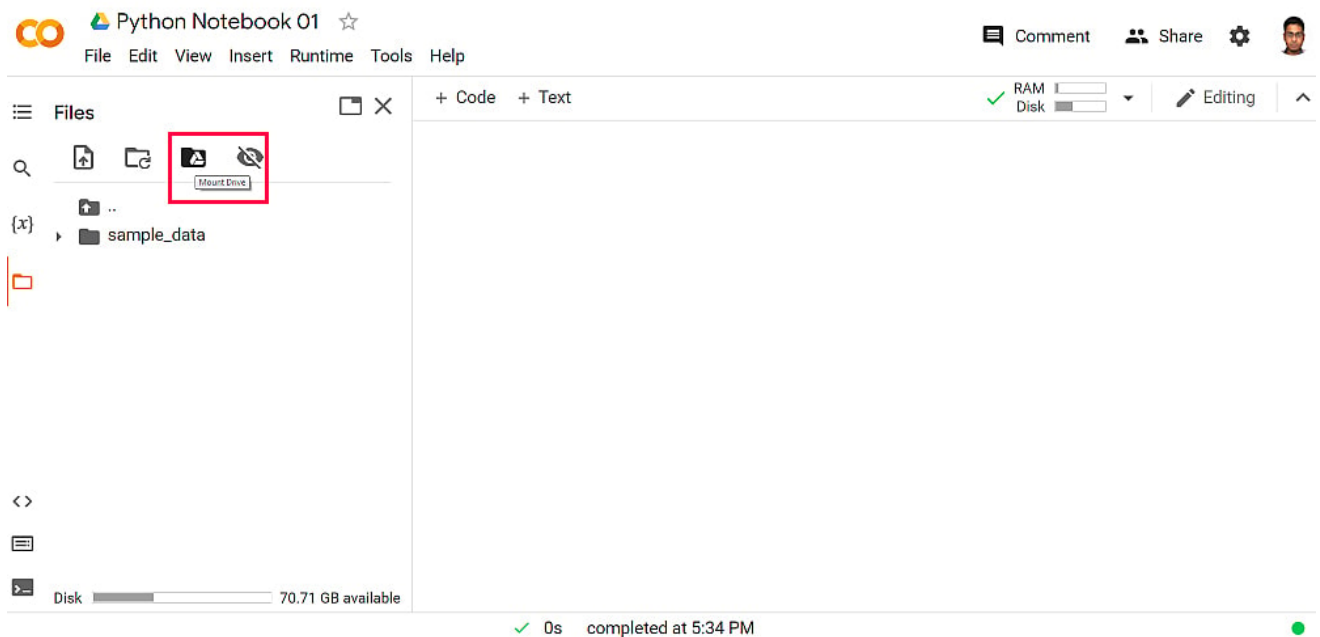
- Нажмите **Ctrl+O** на новом блокноте Colab.
- В появившейся консоли выберите вкладку **Upload (Загрузка)**.
- Нажмите кнопку **Choose File**, чтобы найти локальный файл, который вы хотите открыть в Colab.

#5. Доступ к Google Диску



- Нажмите **Файл** в верхнем меню.
- Вы можете выбрать **Открыть блокнот** или Загрузить блокнот.
- Появится консоль с вкладкой **Google Диск**.
- Нажмите на нее, чтобы получить доступ к файлам из Google Диска.

Если вы хотите подключить Google Drive к экземпляру Colab, выполните следующие действия:

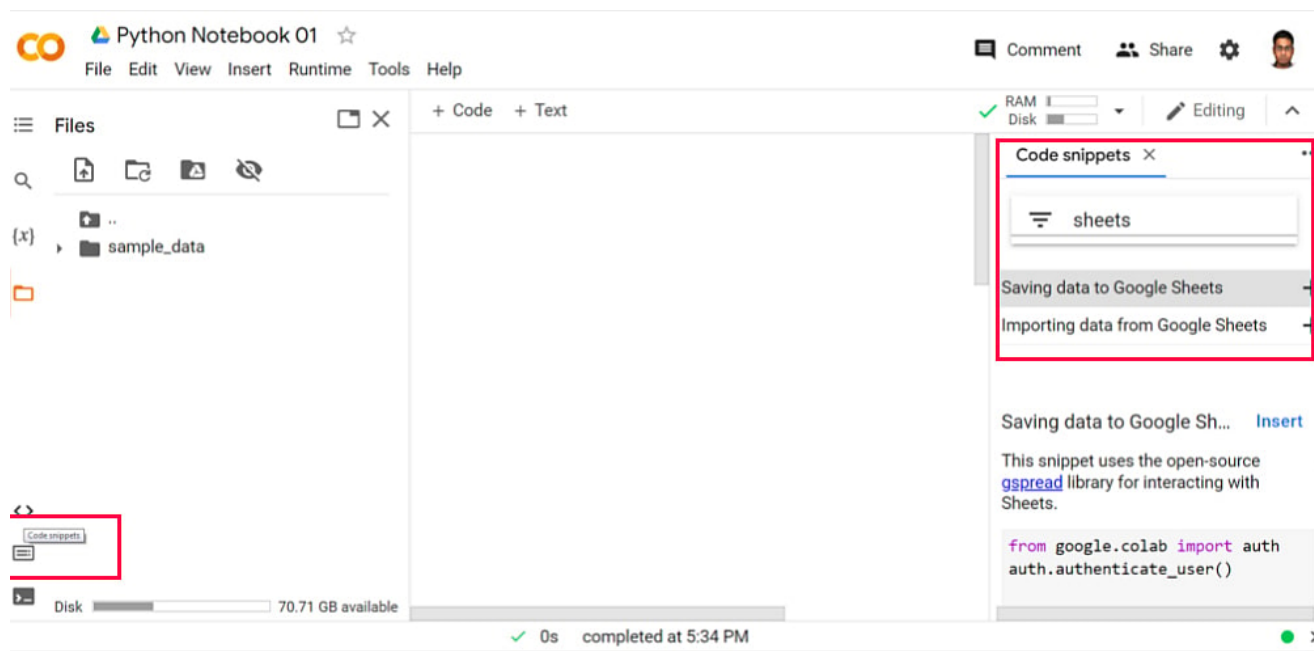


- Нажмите на **Файл**, расположенный на левой панели навигации.

- Выберите команду **Монтировать диск**.
- В появившемся уведомлении выберите **Подключиться к Google Дisku**.
- Google попросит вас выбрать аккаунт для авторизации.

#6. Сохранение в Google Sheets и импорт из них

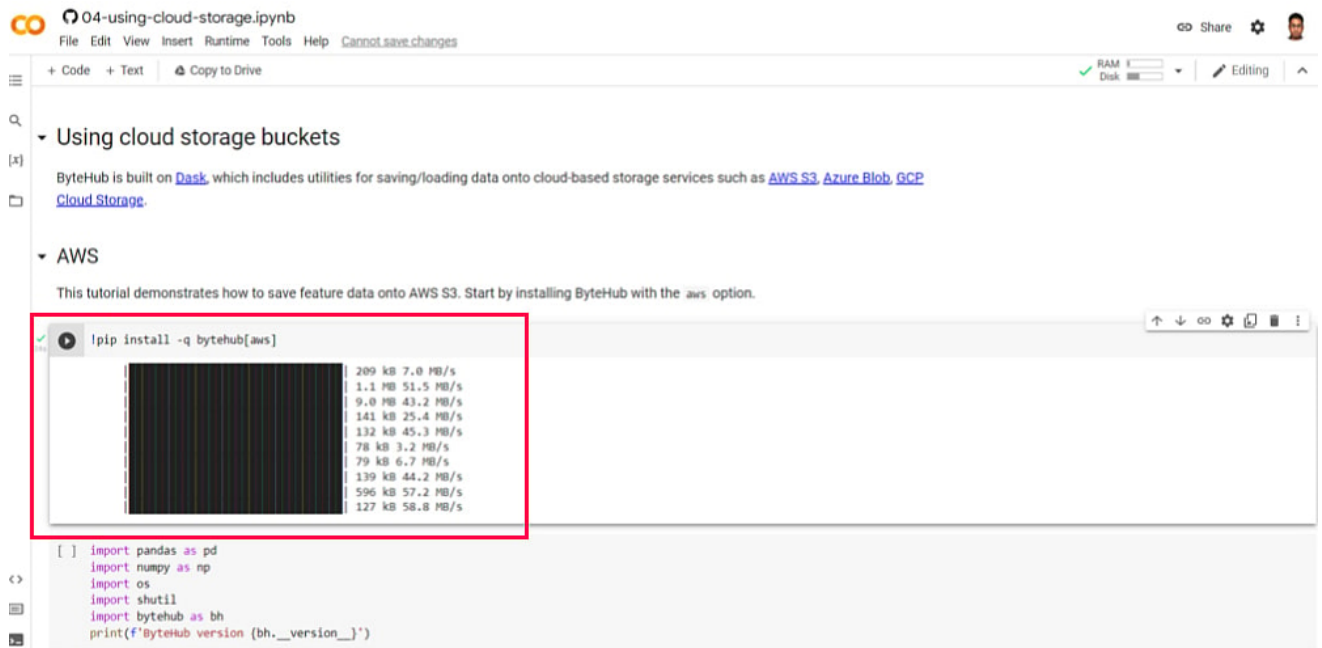
Вы можете легко сохранить данные из блокнота в файл Google Sheets для дальнейшей обработки. Для этого выполните следующие действия:



- Нажмите кнопку **Code Snippets** в левом нижнем углу.
- С правой стороны откроется панель навигации.
- Введите **Sheets** в фильтр, и вы найдете фрагменты кода **Сохранение данных** и **Импорт данных**.
- Дважды щелкните по заголовку, чтобы включить код в блокнот.

#7. Доступ к AWS S3

Вы можете получить доступ к файлам и кодовым ресурсам из облачных хранилищ, таких как AWS S3 и Azure Blob, с помощью облачных ведер хранения.



```
!pip install -q bytehub[aws]
```

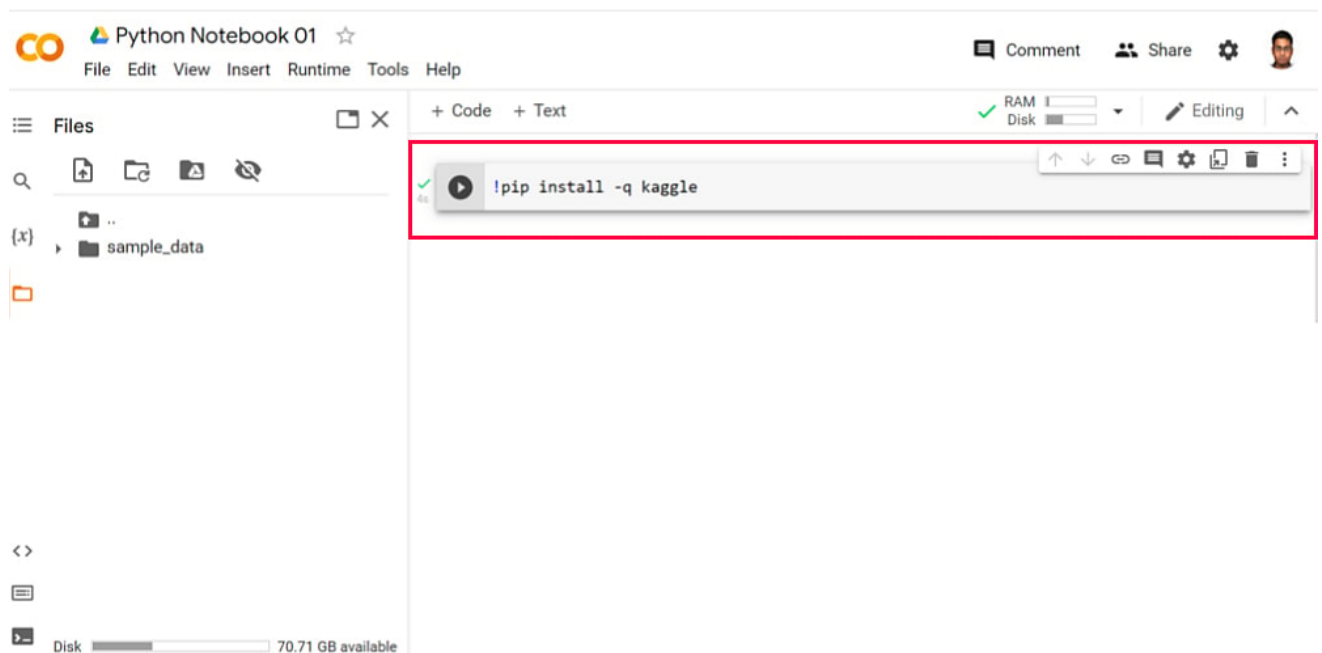
209 kB	7.0 MB/s
1.1 MB	51.5 MB/s
9.0 MB	43.2 MB/s
141 kB	25.4 MB/s
132 kB	45.3 MB/s
78 kB	3.2 MB/s
79 kB	6.7 MB/s
139 kB	44.2 MB/s
596 kB	57.2 MB/s
127 kB	58.8 MB/s

```
[ ] import pandas as pd
import numpy as np
import os
import shutil
import bytehub as bh
print(f'bytehub version {bh.__version__}')
```

Для этого необходимо установить ByteHub, который обладает функционалом для загрузки и сохранения данных в облачном хранилище. Вы можете выполнить следующий код:

```
!pip install -q bytehub[aws]
```

Доступ к наборам данных Kaggle



```
!pip install -q kaggle
```

- Зайдите в аккаунт Kaggle и нажмите на кнопку **Expire API Token** в разделе **API**

, чтобы удалить старые токены.

- Создайте **новый API-токен**, чтобы получить файл kaggle.json на локальном компьютере.
- Теперь используйте следующий код для установки Kaggle:

```
!pip install -q kaggle
```

- Теперь загрузите файл Kaggle.json в базу кода Python, следуя стандартной практике кодирования.

Заключительные слова

Теперь, когда вы подробно ознакомились с приложением Google Collaboratory, вы можете приступить к изучению, обучению или практике проектов по машинному обучению. Google Colab – это действительно удобное облачное приложение для тех, кто любит блокноты Jupyter.

Дата Создания

10.04.2024