

Инструкция по настройке сервиса ColabResearch.

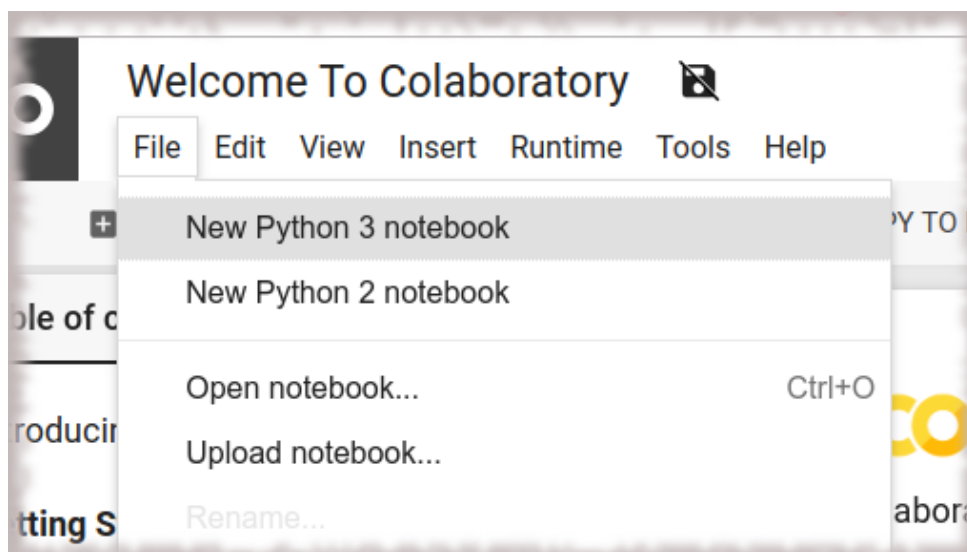
Возможно, у некоторых из вас не очень крутые компьютеры или вы не совсем дружите с **Linux**, а последняя неделя обязательно потребует от вас настройки фреймворка **TensorFlow**, заставить работать который в операционной системе **Windows**, не так-то просто. Не стоит волноваться, корпорация **Google** спешит на помощь. В этой инструкции мы расскажем, как настроить работу с бесплатным сервисом **ColabResearch**.

Прежде всего вам понадобится аккаунт в **Google** (если у вас пока нет аккаунта, создайте его на странице <https://accounts.google.com/signup> [1], заполните регистрационные данные и войдите в аккаунт. После регистрации вы получите доступ ко многим сервисам **Google**, в том числе и к **GoogleDisk** [2] и **ColabResearch** [3], которые вам понадобятся для работы с материалами курса.

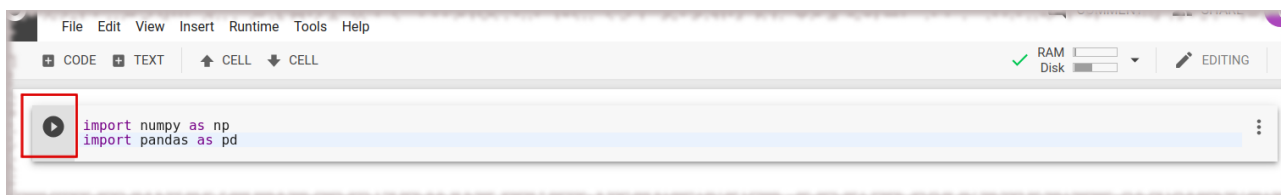
Чтобы начать пользоваться сервисом **ColabResearch**, наберите в адресной строке вашего браузера (предпочтительнее **GoogleChrome**) <https://colab.research.google.com>. Автоматически вы будете перенаправлены на страницу <https://colab.research.google.com/notebooks/welcome.ipynb>, на которой вы можете ознакомиться с интерфейсом сервиса. Внешний вид сервиса на первый взгляд отличается от тех Jupyter-ноутбуков, которые вы запускали локально. Однако вы легко освоите данный интерфейс за несколько минут.


*(Примечание: Если при работе с сервисом вы получаете уведомление об ошибке во всплывающем окне - «**Error. Could not access the resources needed to display output. This is probably because third-party cookies are not allowed by your browser**», измените настройки вашего браузера на странице — и отключите блокировку сторонних файлов cookie в настройках вашего браузера. Для **GoogleChrome**: введите в адресной строке - **chrome://settings/content/cookies** и сдвиньте ползунок влево напротив «**Блокировать сторонние файлы cookie**».)*

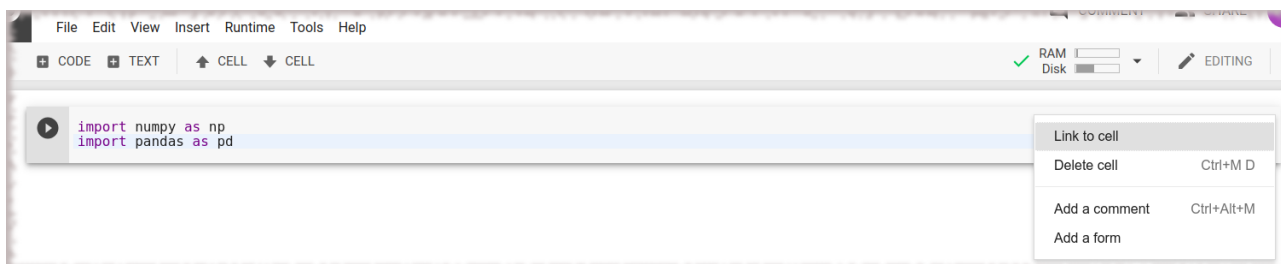
Работа с сервисом начинается с создания нового Jupyter-ноутбука. Зайдите в меню **File** и выберите пункт - «**New Python notebook**» с нужной нам версией **Python**.



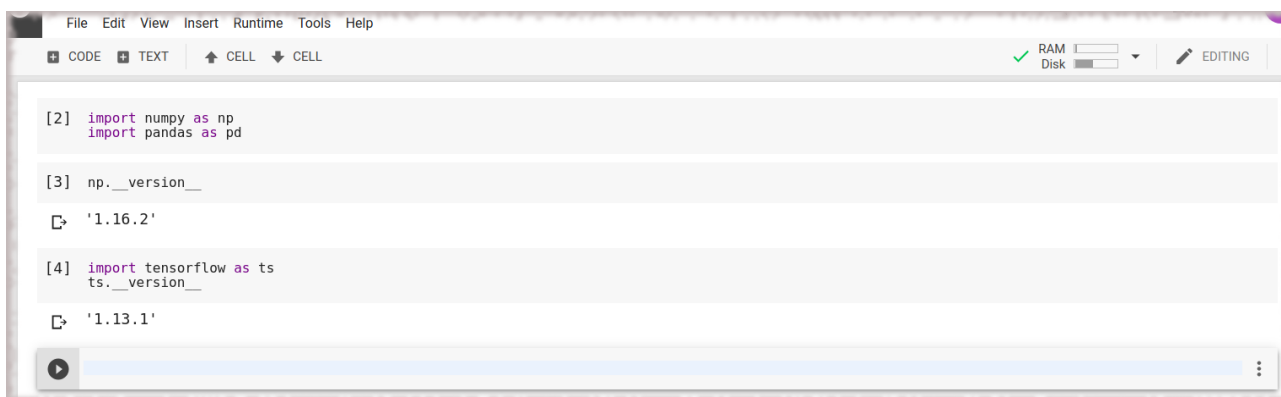
Запустить содержимое ячейки на выполнение можно нажав на треугольник, расположенный справа или комбинацией клавиш **<Ctrl>+<Enter>**. Комбинация **<Shift>+<Enter>** не только запустит ячейку на выполнение, но и добавит новую ячейку.



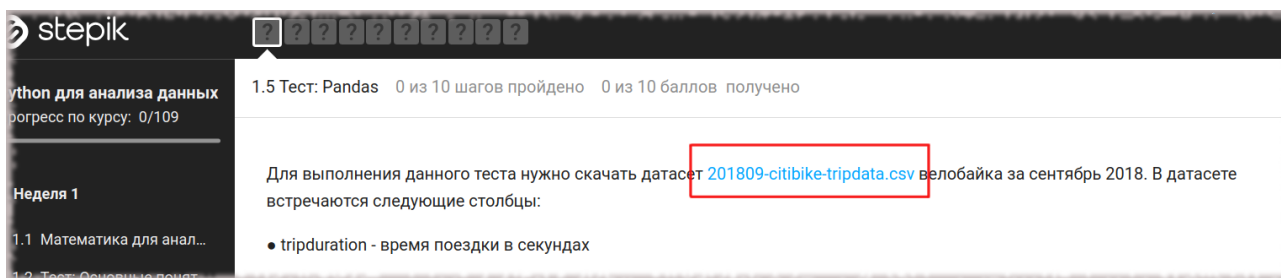
Клик на значке  расположенном в правой части каждой ячейки, открывает меню управления ячейкой. Возможные действия: удалить ячейку, создать ссылку на нее или добавить комментарий.



В качестве примера, убедимся, что библиотеки **NumPy** и **TensorFlow** установлены и узнаем их версии.



Как правило, работать в сервисе **ColabResearch** вы будете со своими файлами данных. Давайте рассмотрим, как можно загрузить их. Для демонстрации загрузки возьмем файл с данными из теста по **Pandas** нашего курса - <https://stepik.org/lesson/198805/step/1?unit=172977>

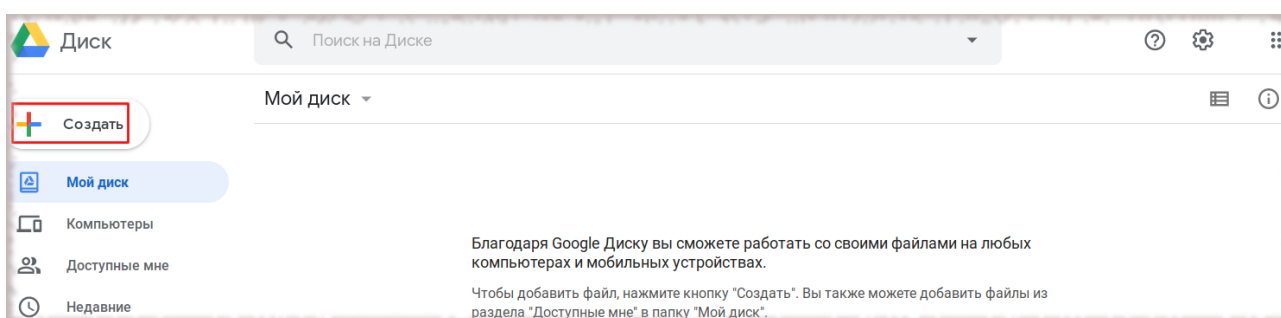


Файл с данными **201809-citibike-tripdata.csv** — в несжатом виде занимает больше 300Мб, и при обработке этого файла на компьютерах с объёмом оперативной памяти 2-4 Гб возможно завершение выполнения с ошибкой (недостаточно памяти). Скачаем файл себе на компьютер по ссылке и распакуем его.

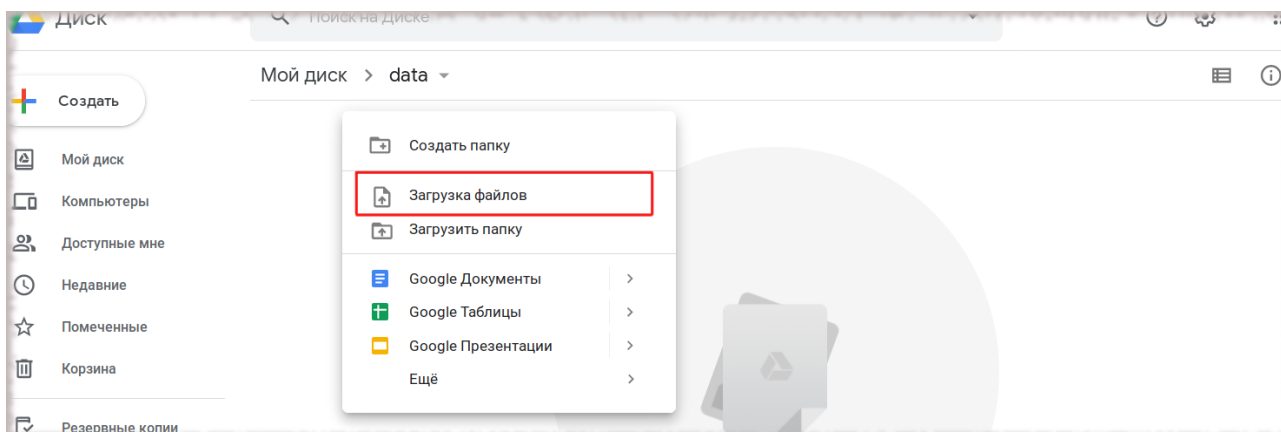
*(Примечание: **Pandas** способен работать и с файлами в формате **zip**, однако, здесь мы разбираем общий случай, когда файлы данных предоставлены нам в формате **csv**).*

Теперь выполним следующие шаги:

1. Загрузим файл с данными на **GoogleDisk**, для этого перейдем в корневой каталог нашего диска, по адресу - <https://drive.google.com/drive/my-drive>
2. Создадим в корневом каталоге папку **data**.



В открывшемся меню выбираем пункт «**Папку**» и вводим имя — **data**. После появления отображения папки в диске, перейдем в нее. Клик правой клавиши мыши в любом пустом месте, откроет контекстное меню



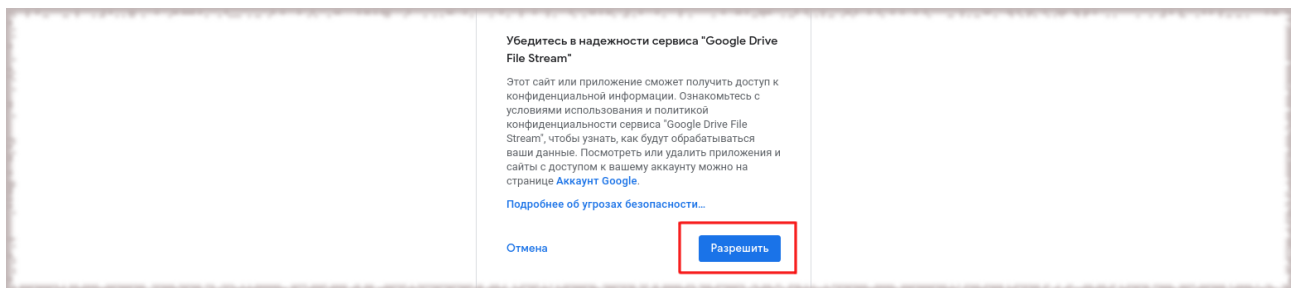
Выберем пункт «**Загрузка файлов**» и укажем расположение файла **201809-citibike-tripdata.csv** на нашем компьютере.

3. Создадим новый Jupyter-ноутбук и в первой ячейке введем следующий код:

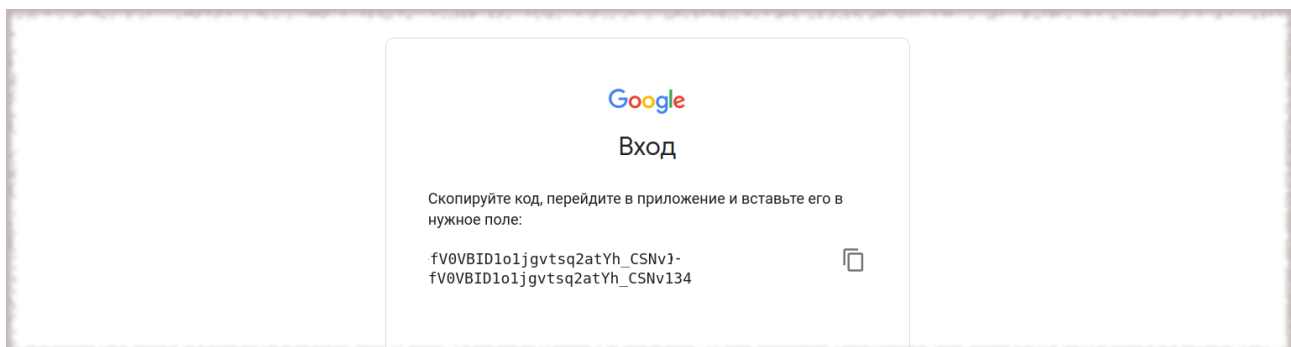
```
from google.colab import drive
drive.mount('/gdrive' )
```

Запустим ячейку на выполнение. Появится запрос на авторизацию, мы должны ввести проверочный код, разрешающий доступ к нашему **GoogleDisk**.

4. Переходим по ссылке, указанной в сообщении, выбираем аккаунт **Google** и нажимаем кнопку «Разрешить»



5. **Google** сообщит код доступа.



Скопируем проверочный код, появившийся на новой странице, и вставим его в поле ввода в Jupyter-ноутбуке, после чего нажмём <Enter>.



После получения сообщения «**Mounted at /gdrive**» мы уже можем обращаться к нашим данным из Jupyter-ноутбука, однако, здесь важно указать правильный путь до файла с данными. В данном пример он будет таким: **/gdrive/My Drive/data/201809-citibike-tripdata.csv** . К примеру, загрузка данных в **Pandas** будет выглядеть следующим образом:

```
import pandas as pd
data = pd.read_csv("/gdrive/My Drive/data/201809-citibike-tripdata.csv")
```



```
[5] from google.colab import drive
drive.mount('/gdrive')

Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6qk8qdgf4n4g3pfee6491hc0brc4i.app:
Enter your authorization code:
.....
Mounted at /gdrive

import pandas as pd
data = pd.read_csv('/gdrive/My Drive/data/201809-citibike-tripdata.csv')
```

Теперь мы можем работать с данными в текущем сеансе. В случае начала нового сеанса необходимо повторить все шаги, начиная с шага 3. Важно заметить, что данные заново загружать не нужно – они будут постоянно храниться на **GoogleDisk**.

- [1] справка **Google** - <https://support.google.com/accounts/answer/27441?hl=ru>
- [2] справка **GoogleDisk** - <https://support.google.com/drive#topic=14940>
- [3] что такое **ColabResearch** - <https://research.google.com/colaboratory/faq.html>