09.12.2022, 12:13 Способы загрузки внешних данных.ipynb - Colaboratory

Для згрузки данных выполним следующие действия:

1. Найдём айл в формате csv.

2.Скачаем этот файл на локальный компьютер

Опробуем различные варианты загрузки данного файла:

1. Загрузка файла через проводник

2. Загрузка файла с помощью filesмодуля

3. Чтение файла с Github

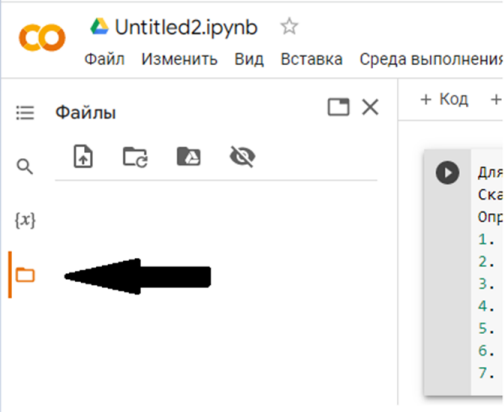
4. Клонирование репозитория Github

5. Загрузка файлов с помощью wgetкоманды Linux

6. Доступ к Google Диску путем его локального подключения

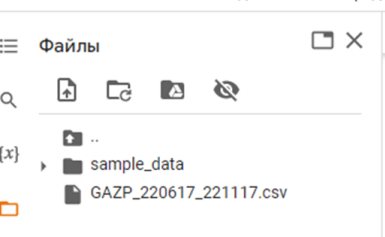
7. Загрузка наборов данных Kaggle

\**Загузка файла через проводник \** Вот что вам нужно сделать: Шаг 1. Щелкните значок « Файлы », чтобы открыть панель «Проводник файлов».



Шаг 2: Щелкните значок загрузки и выберите файлы, которые вы хотите загрузить, в диалоговом окне «Загрузка файла» (

https://colab.research.google.com/drive/1puIgTSZUJe8MYip\_4vATkpmWpan4dzK1#printMode=true 1/7

09.12.2022, 12:13 Способы загрузки внешних данных.ipynb - Colaboratory 

Шаг 3: После завершения загрузки вы можете прочитать файл, как обычно. Например,*pd.read\_csv('Salary\_Data.csv')*

https://colab.research.google.com/drive/1puIgTSZUJe8MYip\_4vATkpmWpan4dzK1#printMode=true 2/7

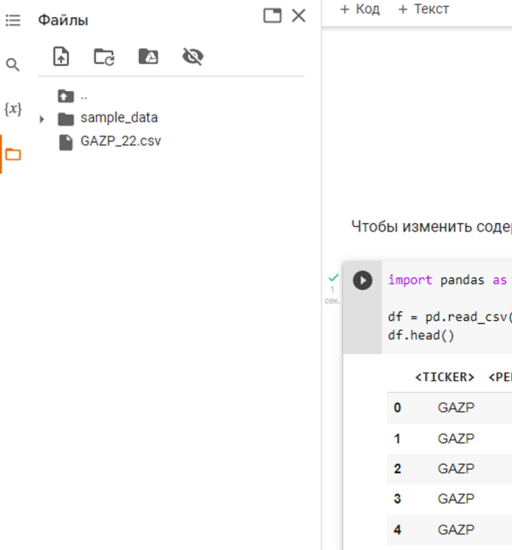
09.12.2022, 12:13 Способы загрузки внешних данных.ipynb - Colaboratory 

image.png

import pandas as pd

df = pd.read\_csv('GAZP\_22.csv') df.head()

https://colab.research.google.com/drive/1puIgTSZUJe8MYip\_4vATkpmWpan4dzK1#printMode=true 3/7

09.12.2022, 12:13 Способы загрузки внешних данных.ipynb - Colaboratory

**<TICKER> <PER> <DATE> <TIME> <OPEN> <HIGH> <LOW> <CLOSE> <VOL> 0** GAZP M 20220701 0 200.84 205.78 181.20 195.26 840519840

**1** GAZP M 20220801 0 194.50 275.96 172.29 254.90 814147380 **2. Загрузка файла с помощью модуля Colab** Вместо того, чтобы щелкать графический **2** GAZP M 20220901 0 258.00 259.85 189.42 217.70 1537265390

интерфейс, вы также можете использовать код Python для загрузки файлов. Вы можете импортировать filesмодуль из google.colab. Затем позвоните upload(), чтобы запустить

**3** GAZP M 20221001 0 219.00 220.55 137.10 170.27 807788870 диалоговое окно «Загрузка файлов» и выберите файл(ы), которые вы хотите загрузить **4** GAZP M 20221101 0 170.89 172.00 161.79 167.70 348461150

from google.colab import files

uploaded = files.upload()

Выбрать файлы Файл не выбран Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

Saving GAZP 220617 221117 csv to GAZP 220617 221117 csv

**3. Чтение файла с Github** Один из самых простых способов чтения данных — через Github. Нажмите на набор данных в репозитории Github, затем нажмите кнопку « Raw ».

import pandas as pddf

pd.read\_csv('https://raw.githubusercontent.com/ElizavetaMaklakova/GAZP\_22/main/GAZP\_220617\_

**<TICKER> <PER> <DATE> <TIME> <OPEN> <HIGH> <LOW> <CLOSE> <VOL> 0** GAZP M 20220701 0 200.84 205.78 181.20 195.26 840519840 **1** GAZP M 20220801 0 194.50 275.96 172.29 254.90 814147380 **2** GAZP M 20220901 0 258.00 259.85 189.42 217.70 1537265390 **3** GAZP M 20221001 0 219.00 220.55 137.10 170.27 807788870 **4** GAZP M 20221101 0 170.89 172.00 161.79 167.70 348461150 **5** GAZP M 20221201 0 168.00 168.76 162.70 163.16 62207020

**4. Клонирование репозитория Github** Вы также можете клонировать репозиторий Github в свою среду Colab так же, как и на свой локальный компьютер, используя файлы *git clone*.

!git clone https://github.com/ElizavetaMaklakova?tab=repositories

Cloning into 'ElizavetaMaklakova?tab=repositories'...

fatal: https://github.com/ElizavetaMaklakova?tab=repositories/info/refs not valid: is

**5. Загрузка файлов из Интернета с помощью *wget*команды Linux** Поскольку Google Colab позволяет вам делать все, что вы можете, в локально размещенном блокноте Jupyter, вы также можете использовать команды оболочки Linux, такие как *ls, dir, pwdи cdт*. д.,

https://colab.research.google.com/drive/1puIgTSZUJe8MYip\_4vATkpmWpan4dzK1#printMode=true 4/7

09.12.2022, 12:13 Способы загрузки внешних данных.ipynb - Colaboratory

используя !. Среди этих доступных команд Linux *wget*позволяет загружать файлы с использованием *протоколов HTTP , HTTPS и FTP \*. В простейшей форме, при использовании без каких-либо опций, \*wget*загрузит ресурс, указанный в URL-адресе, в текущий каталог, например:

**Переименуйте файл** Иногда вам может понадобиться сохранить загруженный файл под другим именем. Для этого просто передайте -*O*параметр, за которым следует новое имя:

!wget https://example.com/cats\_and\_dogs\_filtered.zipzip \ -O new\_cats\_and\_dogs\_filtered.zi

--2022-12-07 10:32:54-- https://example.com/cats\_and\_dogs\_filtered.zipzip Resolving example.com (example.com)... 93.184.216.34, 2606:2800:220:1:248:1893:25c8:1 Connecting to example.com (example.com)|93.184.216.34|:443... connected. HTTP request sent, awaiting response... 404 Not Found

2022-12-07 10:32:54 ERROR 404: Not Found.

--2022-12-07 10:32:54-- http://%20-o/

Resolving -o ( -o)... failed: Name or service not known.

wget: unable to resolve host address ‘ -o’

--2022-12-07 10:32:54-- http://new\_cats\_and\_dogs\_filtered.zip/

Resolving new\_cats\_and\_dogs\_filtered.zip (new\_cats\_and\_dogs\_filtered.zip)... failed: wget: unable to resolve host address ‘new\_cats\_and\_dogs\_filtered.zip’

**Сохранить файл в определенное место** По умолчанию *wget*файлы сохраняются в текущем рабочем каталоге. Чтобы сохранить файл в определенном месте, используйте - *P*опцию:

!wget https://example.com/cats\_and\_dogs\_filtered.zipzip \ -P /tmp/

--2022-12-07 10:33:20-- https://example.com/cats\_and\_dogs\_filtered.zipzip Resolving example.com (example.com)... 93.184.216.34, 2606:2800:220:1:248:1893:25c8:1 Connecting to example.com (example.com)|93.184.216.34|:443... connected. HTTP request sent, awaiting response... 404 Not Found

2022-12-07 10:33:20 ERROR 404: Not Found.

--2022-12-07 10:33:20-- http://%20/

Resolving ( )... failed: Name or service not known.

wget: unable to resolve host address ‘ ’

**Недействительный SSL-сертификат HTTPS** Если вы хотите загрузить файл по HTTPS с хоста с недействительным SSL-сертификатом, вы можете передать --no-check certificateпараметр:

!wget https://example.com/cats\_and\_dogs\_filtered.zip \ --no-check-certificate --2022-12-07 10:33:57-- https://example.com/cats\_and\_dogs\_filtered.zip

https://colab.research.google.com/drive/1puIgTSZUJe8MYip\_4vATkpmWpan4dzK1#printMode=true 5/7

09.12.2022, 12:13 Способы загрузки внешних данных.ipynb - Colaboratory

Resolving example.com (example.com)... 93.184.216.34, 2606:2800:220:1:248:1893:25c8:1 Connecting to example.com (example.com)|93.184.216.34|:443... connected. HTTP request sent, awaiting response... 404 Not Found

2022-12-07 10:33:57 ERROR 404: Not Found.

--2022-12-07 10:33:57-- http://%20--no-check-certificate/

Resolving --no-check-certificate ( --no-check-certificate)... failed: Name or servic wget: unable to resolve host address ‘ --no-check-certificate’

**Несколько файлов одновременно** Если вы хотите загрузить несколько файлов одновременно, используйте -iпараметр, за которым следует путь к файлу, содержащему список URL-адресов для загрузки. Каждый URL-адрес должен быть на отдельной строке. !wget -i dataset-urls.txt

!wget -i набор данных-urls.txt В следующем примере показан dataset-urls.txt : http://example-1.com/dataset.zip https://example-2.com/train.csv http://example 3.com/test.csv

**6. Доступ к Google Диску путем его локального подключения** Вы можете использовать driveмодуль google.colabдля подключения вашего Google Диска к Colab.

from google.colab import drivedrive.mount('/content/drive')

с диска импорта google.colabdrive.mount('/контент/диск') Выполнив приведенное выше заявление, вам будет предоставлена ссылка для аутентификации и текстовое поле для ввода кода авторизации. Выполнив приведенное выше заявление, вам будет предоставлена ссылка для аутентификации и текстовое поле для ввода кода авторизации. Щелкните ссылку аутентификации и следуйте инструкциям по созданию кода авторизации. Скопируйте отображаемый код и вставьте его в текстовое поле, как показано выше. После монтирования вы должны получить сообщение вроде: Mounted at /content/drive После этого вы сможете просматривать содержимое через «Проводник файлов» и читать данные, как обычно. Наконец, чтобы отключить ваш Google Диск:

drive.flush\_and\_unmount()

from google.colab import drive

drive.mount('/content/drive')

Mounted at /content/drive

https://colab.research.google.com/drive/1puIgTSZUJe8MYip\_4vATkpmWpan4dzK1#printMode=true 6/7

09.12.2022, 12:13 Способы загрузки внешних данных.ipynb - Colaboratory

Платные продукты Colab - Отменить подписку

https://colab.research.google.com/drive/1puIgTSZUJe8MYip\_4vATkpmWpan4dzK1#printMode=true 7/7