

Инициализируем dvc.

DVC has enabled anonymous aggregate usage analytics.
Read the analytics documentation (and how to opt-out) here:
<https://dvc.org/doc/user-guide/analytics>

- Check out the documentation: <<https://dvc.org/doc>>
- Get help and share ideas: <<https://dvc.org/chat>>
- Star us on GitHub: <<https://github.com/iterative/dvc>>

```
root@Crash:/home/ubuntu_man/python_projects/mlops_practice# dvc remote list
mydisk  gdrive://1Gpbyy7YB2xUHxjnwH10YYP8WjgKXHY-D
root@Crash:/home/ubuntu_man/python_projects/mlops_practice#
```

```
root@Crash:/home/ubuntu_man/python_projects/mlops_practice# dvc push -r mydisk
0% Querying remote cache|0/1 [00:00<?, ?files/s]
/usr/local/lib/python3.8/dist-packages/oauth2client/_helpers.py:255: UserWarning: Cannot access /root/.cache/pydrive2fs/710796635688-iv
sgbgsb6uv1fap6635dhvuel09o66c.apps.googleusercontent.com/default.json: No such file or directory
  warnings.warn(_MISSING_FILE_MESSAGE.format(filename))
Your browser has been opened to visit:

https://accounts.google.com/o/oauth2/auth?client_id=710796635688-ivsgbgsb6uv1fap6635dhvuel09o66c.apps.googleusercontent.com&redirec
t_uri=http%3A%2F%2Flocalhost%3A8090%2F&scope=https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fdrive+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fd
rive.appdata&access_type=offline&response_type=code&approval_prompt=force

Authentication successful.
2 files pushed
root@Crash:/home/ubuntu_man/python_projects/mlops_practice#
```

Изменим данные у пятой строки датасета.

```
root@Crash: /home/ubuntu_man/python_projects/mlops_practice/lab4/datasets
GNU nano 4.8 titanic.csv
Изменён
,PassengerId,Survived,Pclass,Name,Sex,Age,SibSp,Parch,Ticket,Fare,Cabin,Embarked
0,1,0,3,"Braund, Mr. Owen Harris",male,22.0,1,0,A/5 21171,7.25,,S
1,2,1,1,"Cumings, Mrs. John Bradley (Florence Briggs Thayer)",female,38.0,1,0,PC 17599,71.2833,C85,C
2,3,1,3,"Heikkinen, Miss. Laina",female,26.0,0,0,STON/O2. 3101282,7.925,,S
3,4,1,1,"Futrelle, Mrs. Jacques Heath (Lily May Peel)",female,35.0,1,0,113803,53.1,C123,S
4,5,0,3,"Allen, Mr. William Henry",male,35.0,0,0,373450,8.05,,S
5,6,0,3,"Sandy, Ms. James",female,,0,0,330877,8.4583,,Q
6,7,0,1,"McCarthy, Mr. Timothy J",male,54.0,0,0,17463,51.8625,E46,S
7,8,0,3,"Palsson, Master. Gosta Leonard",male,2.0,3,1,349909,21.075,,S
8,9,1,3,"Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)",female,27.0,0,2,347742,11.1333,,S
9,10,1,2,"Nasser, Mrs. Nicholas (Adele Achem)",female,14.0,1,0,237736,30.0708,,C
10,11,1,3,"Sandstrom, Miss. Marguerite Rut",female,4.0,1,1,PP 9549,16.7,G6,S
11,12,1,1,"Bonnell, Miss. Elizabeth",female,58.0,0,0,113783,26.55,C103,S
12,13,0,3,"Saunderscock, Mr. William Henry",male,20.0,0,0,A/5 2151,8.05,,S
13,14,0,3,"Andersson, Mr. Anders Johan",male,39.0,1,5,347082,31.275,,S
14,15,0,3,"Vestrom, Miss. Hulda Amanda Adolfina",female,14.0,0,0,350406,7.8542,,S
15,16,1,2,"Hewlett, Mrs. (Mary D Kingcome) ",female,55.0,0,0,248706,16.0,,S
16,17,0,3,"Rice, Master. Eugene",male,2.0,4,1,382652,29.125,,Q
17,18,1,2,"Williams, Mr. Charles Eugene",male,,0,0,244373,13.0,,S
18,19,0,3,"Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)",female,31.0,1,0,345763,18.0,,S
```

Пушим в удаленный репозиторий.

```
root@Crash:/home/ubuntu_man/python_projects/mlops_practice/lab4# git commit -a -m "First change in dataset"
[main 4870eb9] First change in dataset
 1 file changed, 2 insertions(+), 2 deletions(-)
root@Crash:/home/ubuntu_man/python_projects/mlops_practice/lab4# dvc push -r mydisk
2 files pushed
root@Crash:/home/ubuntu_man/python_projects/mlops_practice/lab4#
```

С помощью скрипта `no_nan.py`, заменяем в признаке «Age» `nan` средним значением. Затем пушим в удаленный репозиторий.

```
root@Crash:/home/ubuntu_man/python_projects/mlops_practice/lab4# python3 no_nan.py
root@Crash:/home/ubuntu_man/python_projects/mlops_practice/lab4# git status
Текущая ветка: main
Ваша ветка опережает «origin/main» на 4 коммита.
(используйте «git push», чтобы опубликовать ваши локальные коммиты)

Неотслеживаемые файлы:
  (используйте «git add <файл>...», чтобы добавить в то, что будет включено в коммит)
    create_dataset.py
    no_nan.py

индекс пуст, но есть неотслеживаемые файлы
(используйте «git add», чтобы проиндексировать их)
root@Crash:/home/ubuntu_man/python_projects/mlops_practice/lab4# dvc add datasets
100% Adding... |██████████████████████████████████████████████████████████████| 1/1 [00:00, 23.79file/s]

To track the changes with git, run:

    git add datasets.dvc

To enable auto staging, run:

    dvc config core.autostage true
root@Crash:/home/ubuntu_man/python_projects/mlops_practice/lab4# git status
Текущая ветка: main
Ваша ветка опережает «origin/main» на 4 коммита.
(используйте «git push», чтобы опубликовать ваши локальные коммиты)

Изменения, которые не в индексе для коммита:
  (используйте «git add <файл>...», чтобы добавить файл в индекс)
  (используйте «git restore <файл>...», чтобы отменить изменения в рабочем каталоге)
    изменено:   datasets.dvc

Неотслеживаемые файлы:
  (используйте «git add <файл>...», чтобы добавить в то, что будет включено в коммит)
    create_dataset.py
    no_nan.py

индекс пуст (используйте «git add» и/или «git commit -a»)
root@Crash:/home/ubuntu_man/python_projects/mlops_practice/lab4# git commit -m "Second change in dataset"
[main 0ad789c] Second change in dataset
 1 file changed, 2 insertions(+), 2 deletions(-)
root@Crash:/home/ubuntu_man/python_projects/mlops_practice/lab4# dvc push -r mydisk
2 files pushed
```

С помощью скрипта `one_sex.py` создаем новый признак. Затем пушим в удаленный репозиторий.

Репозиторий на гугл диске:

<https://drive.google.com/drive/folders/1KGZ971jz0azcS9XZ-nUyZZrKswVmcRPU>