

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский университет
«Высшая школа экономики»**

**Отчет
по домашнему заданию
по курсу «Семинар наставника»**

Выполнил
Егоров Д.С.

Москва, 2023

Содержание

Введение	3
Методология	3
Результаты	4

Введение

В настоящей работе мы исследуем информацию о пользователях VK, полученную по api от социальной сети.

Методология

Набор данных основан на открытой информации, полученной от api vk при обращении с помощью `users.get`. Отметим, что в отличие от метода `users.search`, `users.get` можно вызвать с сервисным ключом доступа, что потенциально позволяет тестировать классы, связанные с получением данных от api на машинах, имеющих различные IP адреса (локально на рабочей станции через Jupiter Notebook, на мобильных телефонах, имеющих связь GSM и динамические IP-адреса, на удалённых серверах). Информация о пользователях охватывает никнеймы, имена, фамилии, дату рождения, статус аккаунта, родной город, пол, временную зону и некоторые другие параметры.

Для более равномерного распределения среди множества идентификаторов пользователей мы использовали функцию из стандартного класса `Random`, а затем для результата api проводили фильтрацию по требуемому городу.

Результаты

Данные по Красноярску

```
In 12 1 val filePath = "/Users/dmitrijegorow/Downloads/VK_API/vk_users_krsk.csv"
```

```
In 15 1 val df = DataFrame.readCSV(filePath)
      2 df
```

Out 15

user_id	first_name	last_name	nickname	city_id	city	country_id	country	home_town	year	sex	deactivated
31014586	Yana	Mikhaylova	null	73	Krasnoyarsk	1	Russia	null	1994	1	null
31014660	Tanya	Vepreva	null	73	Krasnoyarsk	null	null	null	1996	1	banned
31014661	Maria	Likhatskaya	null	73	Krasnoyarsk	null	null	null	null	1	banned
31014821	Vladislav	Shtuko	null	73	Krasnoyarsk	1	Russia	Красноярск	1993	2	null
31015070	Olga	Asharova	null	73	Krasnoyarsk	1	Russia	null	1970	1	null
31015372	Yana	Korolkova	null	73	Krasnoyarsk	1	Russia	null	1994	1	null
31015607	Armen	Gevorgyan	null	73	Krasnoyarsk	null	null	null	null	2	banned
31015622	Anyutka	Zolottse	null	73	Krasnoyarsk	1	Russia	Красноярск	null	1	null
31015937	Danila	Rus	null	73	Krasnoyarsk	null	null	null	1989	2	banned
31016009	Volodya	Kent	null	73	Krasnoyarsk	1	Russia	null	1991	2	null
31016139	Nasvay	Zelenenky	null	73	Krasnoyarsk	1	Russia	Ташкент	null	2	null
31016181	Zhenya	Aytova	null	73	Krasnoyarsk	1	Russia	Красноярск	null	1	null
31016250	Yanochka	Lukyanova	null	73	Krasnoyarsk	1	Russia	Krsk	null	1	null
31016415	Katerina	Gagarkina	null	73	Krasnoyarsk	1	Russia	красноярск	1984	1	null
31016439	Vitalya	Dubledenko	null	73	Krasnoyarsk	null	null	null	1993	2	banned
31016455	DELETED	null	null	73	Krasnoyarsk	null	null	null	null	1	deleted
31016547	Ruslan	Mironov	null	73	Krasnoyarsk	1	Russia	null	1984	2	null
31016756	Andrey	Iliq	null	73	Krasnoyarsk	1	Russia	null	null	2	null
31016822	Eldar	Korolev	null	73	Krasnoyarsk	null	null	null	null	2	banned
31016999	Sofya	Khamkheeva	null	73	Krasnoyarsk	1	Russia	null	null	1	null

... showing only top 20 of 2044 rows

DataFrame: rowsCount = 2044, columnsCount = 12

Рис. 1: Загрузка данных в Jupiter Notebook

```
In 16 1 df.first_name
```

Out 16

first_name
Yana
Tanya
Maria
Vladislav
Olga
Yana
Armen
Anyutka
Danila
Volodya
Nasvay
Zhenya
Yanochka
Katerina
Vitalya
DELETED
Ruslan
Andrey
Eldar
Sofya

... showing only top 20 of 2044 rows

DataColumn: name = "first_name", type = String, size = 2044

Рис. 2: Проверка имён

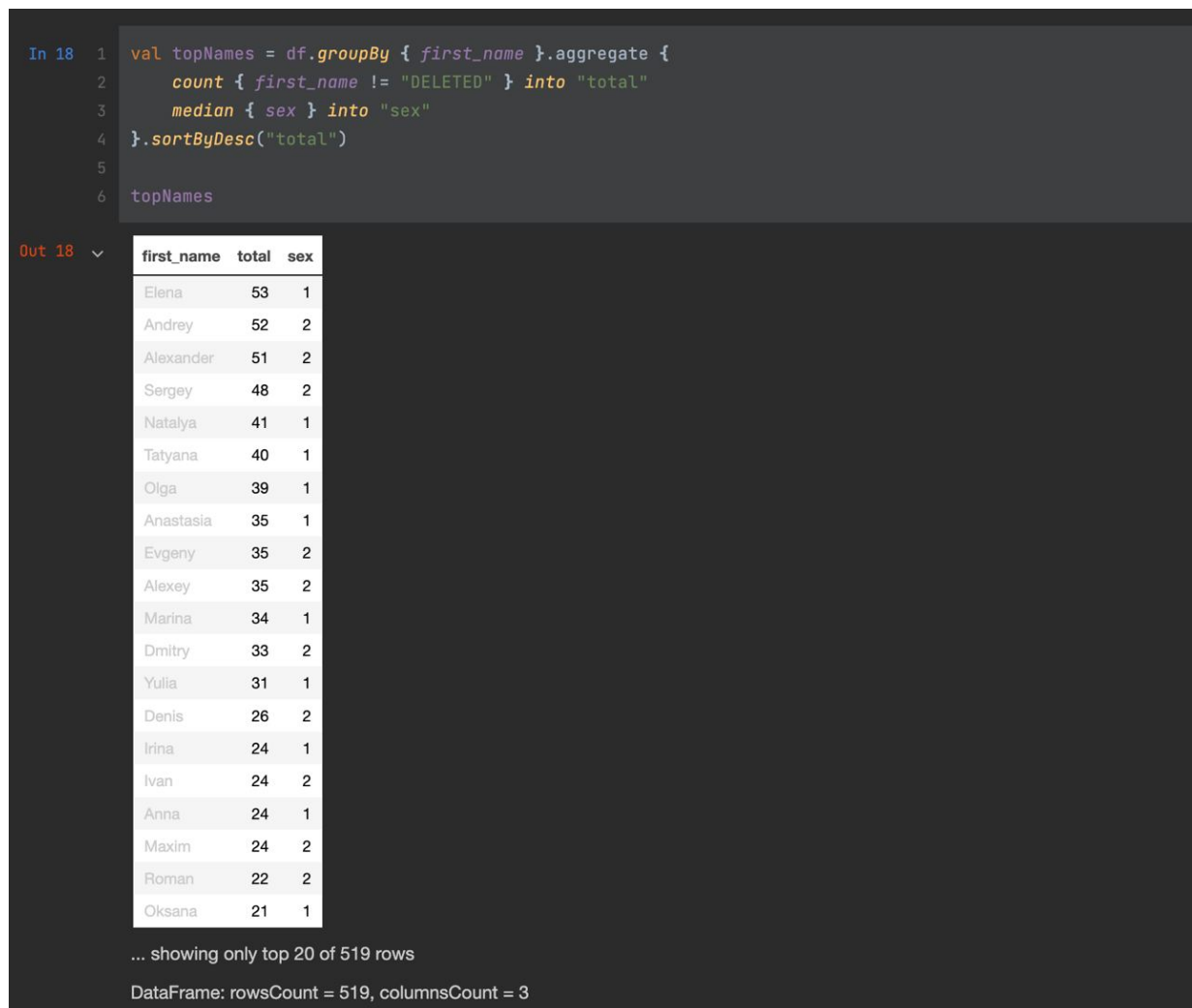


Рис. 3: Группировка по именам

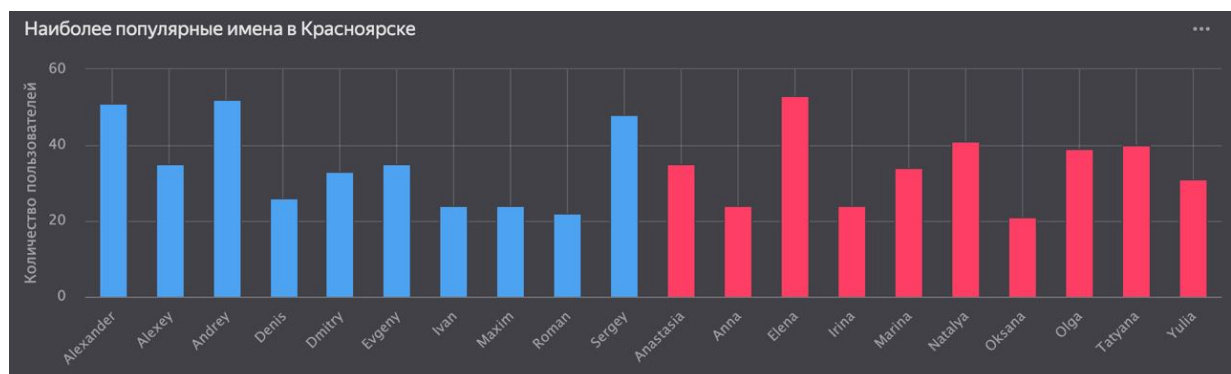


Рис. 4: Самые часто встречающиеся имена в Красноярске

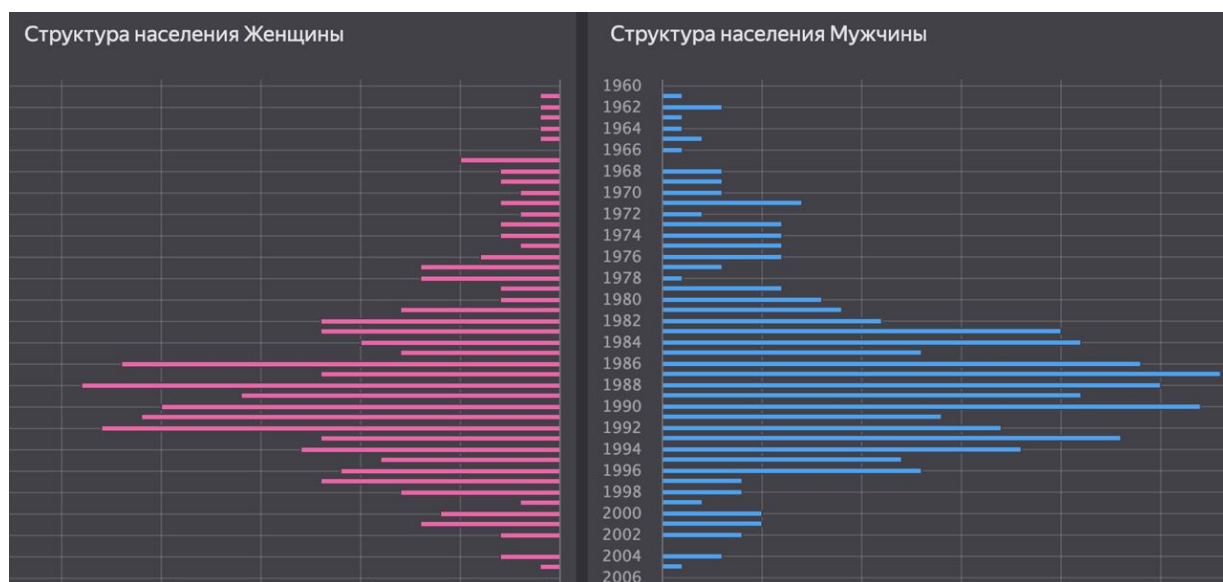


Рис. 5: Половозрастная структура из дашборда DataLens