

# Отчет по работе с API VK и анализу данных

Бирючева Александра

4 мая 2024 г.

## 1 Постановка задачи

Выгрузить датасет из vk, провести анализ полученных данных и подготовить визуализацию о полученных результатах. Задача - найти топ женских и мужских имен в определенном населенном пункте, на основе людей, зарегистрированных в vk, а также провести анализ их возраста.

## 2 Что сделано

С помощью API VK были выгружены данные о людях, проживающих в поселке Бор, затем данные загрузили на Datalens, провели их начальную обработку, и использовали их для создания диаграмм, отображающих необходимую нам информацию.

## 3 Используемые технологии

Jupyter Notebook, VK API, Pandas, Datalens

## 4 Схема работы вашего приложения

После подключения vk token, с помощью vk api и users.search, загружаем данные. При загрузке 1000 человек, мы получаем меньше людей, так как есть удаленные аккаунты, данные о которых мы не можем выгрузить. Поэтому выгружаем данные о 500 женщин и 500 мужчин, затем склеиваем данные. Полученные данные загружаем в Datalens и создаем таблицы с необходимой нам информацией.

## 5 Выводы по полученным результатам

Топ 3 женских имен в поселке Бор: Ирина, Дарья, Юлия. Топ 3 мужских: Сергей, Николай и Саня (Александр). Самый частый год рождения людей, у которых отмечен данный населенный пункт в качестве города проживания, это 2003-2004. Также есть часть людей (3 мужчины, 1 женщина), которые отметили 1900 год как год рождения.

## 6 Обнаруженные ошибки

Кроме вышеупомянутой проблемы с выгрузкой ровно 1000 человек, была также найдена проблема с годами рождения, так как некоторые пользователи не отмечали день рождения, либо же был скрыт год, что мешало представлению половозрастной статистики. Поэтому данные о пользователях с "некорректным" годом рождения игнорировались при создании чарта, где использовался возраст людей. Также при попытке выгрузки данных из vk с помощью json и записи данных прямо в файл возникла ошибка с кодировкой - некоторые пользователи использовали символы, не входящие в utf-8, и соответственно при попытке записи этих данных мы получали ошибку. В последствии для решения данной проблемы был использован pandas для выгрузки данных.