

## Дашборд для АНО “РАК ПОБЕДИМ”.

### 1. Описание проекта

Мы аналитики АНО “РАК ПОБЕДИМ”. Перед нами стоит задача провести предобработку данных и построить дашборд.

Заказчик предоставил данные в двух файлах в формате Excel о подопечных фонда и о донорах фонда.

Данные о подопечных:

Google Sheets

[https://docs.google.com/spreadsheets/d/1\\_S4m-SK2juop8FkCMVCFIAJcUi2zFBj25VVOsUAaOT4/edit#gid=1721685457](https://docs.google.com/spreadsheets/d/1_S4m-SK2juop8FkCMVCFIAJcUi2zFBj25VVOsUAaOT4/edit#gid=1721685457)

Я.Диск

<https://disk.yandex.ru/i/wF4zNVkgsA1h0A>

Данные о донорах:

Я.Диск

[https://disk.yandex.ru/d/uenb\\_rcX8ECmcw](https://disk.yandex.ru/d/uenb_rcX8ECmcw)

Google Sheets

[https://docs.google.com/spreadsheets/d/1OfWqpwiyTnM9fTbrgxlkK\\_xXJL5803p-Y-FmECmOspU/edit#gid=42853625](https://docs.google.com/spreadsheets/d/1OfWqpwiyTnM9fTbrgxlkK_xXJL5803p-Y-FmECmOspU/edit#gid=42853625)

#### **Необходимо провести предобработку данных о подопечных фонда.**

Из-за несовершенства сбора данных информация в полях датасета представлена в разрозненном виде. Необходимо привести к единообразию написание значений в полях:

- 'Город'
- 'Сфера деятельности',
- 'Диагноз',
- 'Этап',
- 'Помощь'

Для выполнения этой задачи нужно использовать справочник по ссылке

<https://docs.google.com/spreadsheets/d/1ODxYYhWq1OMntpGo9gD8jYKJQV6IGfw0fAtkyhCr1Q0/edit#gid=822464959>

Устранить другие неточности в данных, провести анализ и дать рекомендации по сбору данных на будущее.

### 2. Предобработка файла с подопечными

Нам дан файл состоящий из 16 столбцов, 1568 строк. Пропуски есть почти в каждом столбце, полных дублей нет. Сразу удалим столбцы email, телефон, referer, Form name, дата заполнения анкеты (у нас есть дата отправки, они дублируют друг друга). Дата рождения и дата отправки анкеты - переведем в формат даты. Посчитаем возраст подопечных в отдельном столбце.

Так как файл небольшой (1500 строк), предобработку файлов было принято проводить в google sheets. Будем использовать регулярные выражения и скрипты.

### **2.1. Адрес. Оставим только город.**

В столбце 3 пропуска, удалим их, так как для дашборда мы должны отразить карту заболевших.

После применения формулы для сравнения 182 не найденных значений. Проблемы связаны с ошибками:

- в написании,
- маленькие города\поселки которых нет в словаре, их заменяем на административный центр
- другие страны, выделим их в отдельную группу

### **2.2. Сфера деятельности.**

192 пропуска-заменяем на другое, с остальным будем работать по той же формуле, но перед этим удалим из строк разную лишнюю информацию.

### **2.3. Диагноз.**

Из этого столбца выводим информацию о стадии заболевания, кто подал заявку (сам или третье лицо) и локализации рака.

### **2.4. Этап лечения, помощь, как узнали.**

Все эти столбцы так же приводим к нужному виду в соответствии со справочником.

После всех преобразований у нас получился файл:

<https://docs.google.com/spreadsheets/d/1X8hdBV5tLVrOGEK-3cBrYBHA110E71AJGuierjtMbSs/edit?usp=sharing>

Но так как мы использовали формулы, скрипты, он долго подгружается в Datalens, поэтому был создан второй файл, который содержит только значения из преобразованного файла. Этот файл загружен в Datalens.

<https://docs.google.com/spreadsheets/d/1tom-QPMDitHTNW1s1yKM6JSZFqnxWdS3LwVkc3kJQt4/edit?usp=sharing>

## **3. Предобработка файла с донорами.**

Файл состоит из 582 строк и 9 столбцов. Нужно было поменять формат данных для столбца с датой платежа. Остальные данные в предобработке не нуждаются.

## **4. Построение дашборда.**

Дашборд выполнен в Yandex Datalens. Разделен на 2 вкладки: подопечные и доноры. На вкладке с подопечными построена карта с подопечными, а также построены чарты с локализацией заболевания, стадией, этапом лечения, а так сфера деятельности подопечных, какая им требуется помощь и как они узнали о фонде, динамика заполнения анкет. Построен чарт с заболеваниями по возрастам и выведен топ-10 локализаций для каждого возраста. Дополнительно выведена

таблица с именинниками, чтобы можно было их поздравлять. На всю вкладку действует общий фильтр с датой, для удобства выгрузки отчета за определенный период.

На второй вкладке выведена информация по донорам. Общая информация по количеству доноров, по сумме пожертвований, динамика пожертвований, а также как платежи были проведены. Отдельно составлены таблицы с именами доноров, кто делает больше всего пожертвований и самые большие суммы, возможно таким донорам можно отправлять слова благодарности или какие-то другие знаки внимания.

## **5. Вывод и рекомендации.**

Для удобства анализа и визуализации данных, необходимо упростить анкету для подопечных, а именно применить, например, выпадающие списки для заполнения сферы деятельности, локализации заболевания, стадии, этапе лечения, городе проживания и другие.

Дашборд: <https://datalens.yandex/fdyn00s3nq663>