

Ход обработки:

У нас 1569 строк в таблице с подопечными, поэтому предобработка данных проведена в Google Sheets.

- 1) Проверили столбец ФИО на дубликаты.
- 2) Посчитали возраст подопечных по формуле =ЕСЛИ(ЕПУСТО(D2); ;ДОЛЯГОДА(D2;СЕГОДНЯ());1)).
- 3) Удалены значения "с.", "г." и другие аналогичные из столбца "Город".
- 4) Извлекли первое слово – название населенного пункта формулой = ИНДЕКС(SPLIT(G2; " "); 1).
- 5) Визуально проверили и отредактировали столбец "Город".
- 6) Заполнили регион функцией ВПР из справочника.
- 7) Для категоризации столбцов "Сфера деятельности" и "Диагноз" создали столбцы с категориями из справочника и использовали формулу проверки принадлежности к определенной категории и в отдельном столбце выводим категорию.
- 8) Для извлечения стадии заболевания использовали формулу с регулярным выражением, вручную римские цифры заменили на арабские.
- 9) Столбцы "Заявитель", "Этап лечения", "Помощь", "Как узнали" обработаны методом фильтрации по совпадению значений из справочника.
- 10) Создали новый лист, на который перенесли только обработанные данные, чтобы не загружать датасет при импорте в DataLens.

Таким образом, данные были предобработаны и готовы к использованию в дальнейшем импорте в другие инструменты.