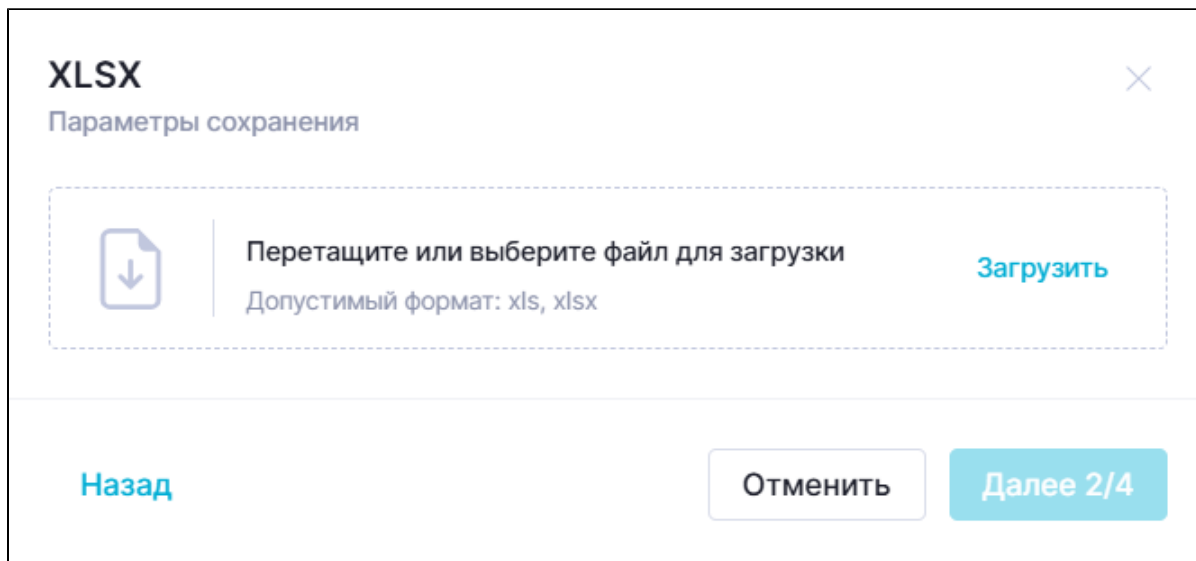


Источники на основе файлов

После выбора в качестве источника файла типа:

- XLSX
- CSV или TXT
- XML
- JSON

и нажатия кнопки "Далее" открывается окно для загрузки файла. Для загрузки файла следует выбрать нужный файл, нажав предварительно кнопку "Загрузить" или перетащив файл в окно, и нажать кнопку "Далее".




Загрузка файла

Если в качестве источника выбран файл CSV, то в окне загрузки файла следует указать разделитель, а также используемую в файле кодировку.

CSV

Параметры сохранения



Перетащите или выберите файл для загрузки
Допустимый формат: csv, txt

[Загрузить](#)

Разделитель * ?

;

Кодировка * ?

utf-8

[Назад](#) [Отменить](#) [Далее 2/4](#)

Загрузка файла CSV

Доступны следующие виды кодировки файла для загрузки:

- windows-1251;
- utf-8;
- utf-16;
- koi8_r;
- latin_1;
- macyrillic.

Структуры данных файлов для создания датасетов

XLSX

Файл для создания датасета из Excel должен иметь следующий формат:

	A	B	C	D	E	F	G
1	Макрорегион	Страна	Город	Экспорт, млн.руб	Импорт, млн.руб	Торговые санкции	Дата введения
2	Азия	Индонезия	Денпасар	3744	14288	false	null
3	Азия	Индонезия	Джакарта	3808	10800	false	null
4	Азия	Индонезия	Маланг	5520	9024	false	null
5	Азия	Китай	Гуанчжоу	1920	3792	false	null
6	Азия	Китай	Пекин	9072	5520	false	null
7	Азия	Китай	Шанхай	3680	15088	false	null
8	Азия	Малайзия	Кота-Бару	9072	13952	false	null
9	Азия	Малайзия	Куала-Лумпур	15584	5520	false	null
10	Азия	Малайзия	Малакка	14912	10784	false	null
11	Азия	Южная Корея	Бусан	3744	14288	true	25.03.25
12	Азия	Южная Корея	Кёнджу	14736	5248	true	25.03.25
13	Азия	Южная Корея	Сеул	2848	15792	true	25.03.25
14	Азия	Япония	Киото	6496	7312	true	18.04.25
15	Азия	Япония	Осака	12880	13952	true	18.04.25
16	Азия	Япония	Токио	12576	14512	true	18.04.25
17	Африка	Египет	Гиза	5584	7664	false	null
18	Африка	Египет	Каир	8112	2064	false	null
19	Африка	Египет	Луксор	2064	5568	false	null
20	Африка	Нигерия	Абуджа	192	11872	false	null
21	Африка	Нигерия	Ибадан	8448	15296	false	null
22	Африка	Нигерия	Лагос	12540	14000	false	null
23	Африка	Тунис	Кайруан	9488	15264	false	null
24	Африка	Тунис	Монастир	2784	13344	false	null
25	Африка	Тунис	Тунис	5520	9104	false	null
26	Африка	Эфиопия	Аддис-Абеба	14320	10784	false	null
27	Африка	Эфиопия	Гондэр	7872	5552	false	null
28	Африка	Эфиопия	Харэр	2944	15296	false	null

CSV

Файл для создания датасета из CSV должен иметь следующий формат (разделитель может быть иным):

Файл Изменить Просмотр

```
Макрорегион;Страна;Город;Экспорт, млн.руб;Импорт, млн,руб;Торговые санкции;Дата введения;  
Азия;Индонезия;Денпасар;3744;14288;false;null;  
Азия;Индонезия;Джакарта;3808;10800;false;null;  
Азия;Индонезия;Маланг;5520;9024;false;null;  
Азия;Китай;Гуанчжоу;1920;3792;false;null;  
Азия;Китай;Пекин;9072;5520;false;null;  
Азия;Китай;Шанхай;3680;15088;false;null;  
Азия;Малайзия;Кота-Бару;9072;13952;false;null;  
Азия;Малайзия;Куала-Лумпур;15584;5520;false;null;  
Азия;Малайзия;Малакка;14912;10784;false;null;  
Азия;Южная Корея;Бусан;3744;14288;true;25.03.25;  
Азия;Южная Корея;Кёнджу;14736;5248;true;25.03.25;  
Азия;Южная Корея;Сеул;2848;15792;true;25.03.25;  
Азия;Япония;Киото;6496;7312;true;18.04.25;  
Азия;Япония;Осака;12880;13952;true;18.04.25;  
Азия;Япония;Токио;12576;14512;true;18.04.25;  
Африка;Египет;Гиза;5584;7664;false;null;  
Африка;Египет;Каир;8112;2064;false;null;  
Африка;Египет;Луксор;2064;5568;false;null;  
Африка;Нигерия;Абуджа;192;11872;false;null;  
Африка;Нигерия;Ибадан;8448;15296;false;null;  
Африка;Нигерия;Лагос;12540;14000;false;null;  
Африка;Тунис;Кайруан;9488;15264;false;null;  
Африка;Тунис;Монастир;2784;13344;false;null;  
Африка;Тунис;Тунис;5520;9104;false;null;  
Африка;Эфиопия;Аддис-Абеба;14320;10784;false;null;  
Африка;Эфиопия;Гондэр;7872;5552;false;null;  
Африка;Эфиопия;Харэр;2944;15296;false;null;
```

XML

Файл для создания датасета из XML должен иметь следующий формат:

```
1 <root>
2   <columns>
3     <item>Макрорегион</item>
4     <item>Страна</item>
5     <item>Город</item>
6     <item>Экспорт, млн.руб</item>
7     <item>Импорт, млн,руб</item>
8     <item>Торговые санкции</item>
9     <item>Дата введения</item>
10  </columns>
11  <field_types>
12    <item>string</item>
13    <item>string</item>
14    <item>string</item>
15    <item>int64</item>
16    <item>int64</item>
17    <item>string</item>
18    <item>time.time</item>
19  <items>
20    <item>
21      <field>Азия</field>
22      <field>Индонезия</field>
23      <field>Денпасар</field>
24      <field>3744</field>
25      <field>14288</field>
26      <field>>false</field>
27      <field>nil</field>
28    </item>
29
30    ...
31
32  </items>
33
34 </root>
```

JSON

Файл для создания датасета из JSON должен иметь следующий формат:

```

1  {
2      "columns": [
3          "Макрорегион",
4          "Страна",
5          "Город",
6          "Экспорт, млн.руб",
7          "Импорт, млн.руб",
8          "Торговые санкции",
9          "Дата введения"
10     ],
11     "field_types": [
12         "string",
13         "string",
14         "string",
15         "int64",
16         "int64",
17         "string",
18         "date.time"
19     ],
20     "items": [
21         {
22             "Макрорегион": "Азия",
23             "Страна": "Индонезия",
24             "Город": "Денпасар",
25             "Экспорт, млн.руб": 3744,
26             "Импорт, млн.руб": 14288,
27             "Торговые санкции": false,
28             "Дата введения": "nil"
29         }
30     ]
31 }

```

Правила распознавания данных для заполнения колонок

1. Колонки определяются по заполненным ячейкам первой строки, а также включают в себя пустые столбцы, если они находятся среди заполненных, в случае файлов XLSX и CSV. В случае файлов JSON и XML, колонки формируются из перечисления в блоке columns.
2. За название колонки берется значение ячейки первой строки, или порядковый номер столбца, когда ячейка пуста, в случае файлов XLSX и CSV. В случае файлов JSON и XML, названия колонок указаны в перечислении columns.
3. В случае файлов XLSX и CSV тип данных колонки определяет первая, следующая после названия, заполненная ячейка. В случае обнаружения в ячейке, следующей за ячейкой, определяющей тип данных, типа данных, отличного от определенного, тип данных всей колонки определяется как "Строка". В случае файлов JSON и XML, типы данных колонок берутся из перечисления в блоке field_types.
4. В случае файлов XLSX и CSV колонка, содержащая только значения 1 и 0, НЕ определяется как содержащая тип данных "Булево".
5. В случае файлов XLSX и CSV тип данных "Дата" определяется, только когда дата находится в одном из форматов:

dd.mm.yyyy
dd.mm.yyyy H
dd.mm.yyyy H:M
dd.mm.yyyy H:M:S
dd.mm.yy
dd.mm.yy H
dd.mm.yy H:M
dd.mm.yy H:M:S
dd-mm-yyyy
dd-mm-yyyy H
dd-mm-yyyy H:M
dd-mm-yyyy H:M:S
dd-mm-yy
dd-mm-yy H
dd-mm-yy H:M
dd-mm-yy H:M:S
dd/mm/yyyy
dd/mm/yyyy H
dd/mm/yyyy H:M
dd/mm/yyyy H:M:S
dd/mm/yy
dd/mm/yy H
dd/mm/yy H:M
dd/mm/yy H:M:S
yyyy.mm.dd
yyyy.mm.dd H
yyyy.mm.dd H:M
yyyy.mm.dd H:M:S
yyyy-mm-dd
yyyy-mm-dd H
yyyy-mm-dd H:M
yyyy-mm-dd H:M:S
yyyy/mm/dd
yyyy/mm/dd H
yyyy/mm/dd H:M
yyyy/mm/dd H:M:S



Тип данных "Дата" автоматически приводится к виду yyyy-mm-dd HH:MM:SS, который потом может быть изменен в настройках датасета.

Для перехода на следующий этап необходимо нажать кнопку "Далее 2/4".