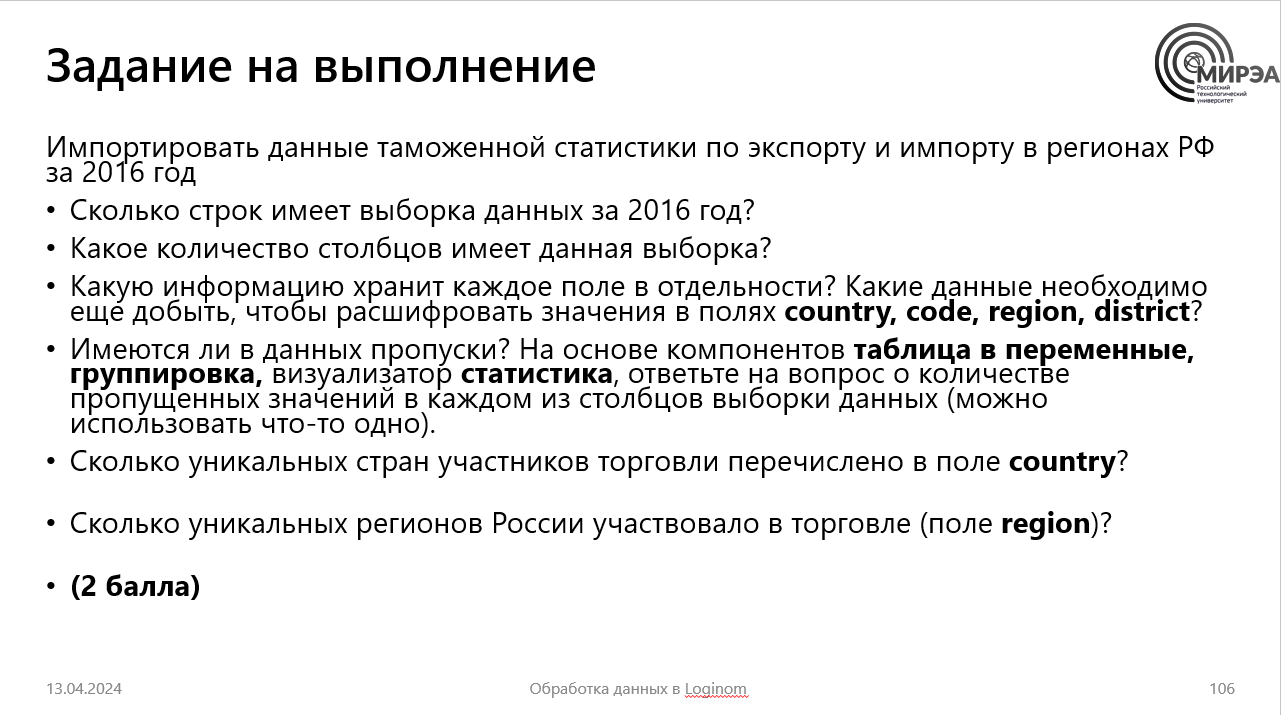
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  | | МИНОБРНАУКИ РОССИИ | | | | Федеральное государственное бюджетное образовательное учреждение  высшего образования  **«МИРЭА** – **Российский технологический университет»**  **РТУ МИРЭА** | | | | |
| Институт информационных технологий (ИТ) | |
| Кафедра прикладной математики | |

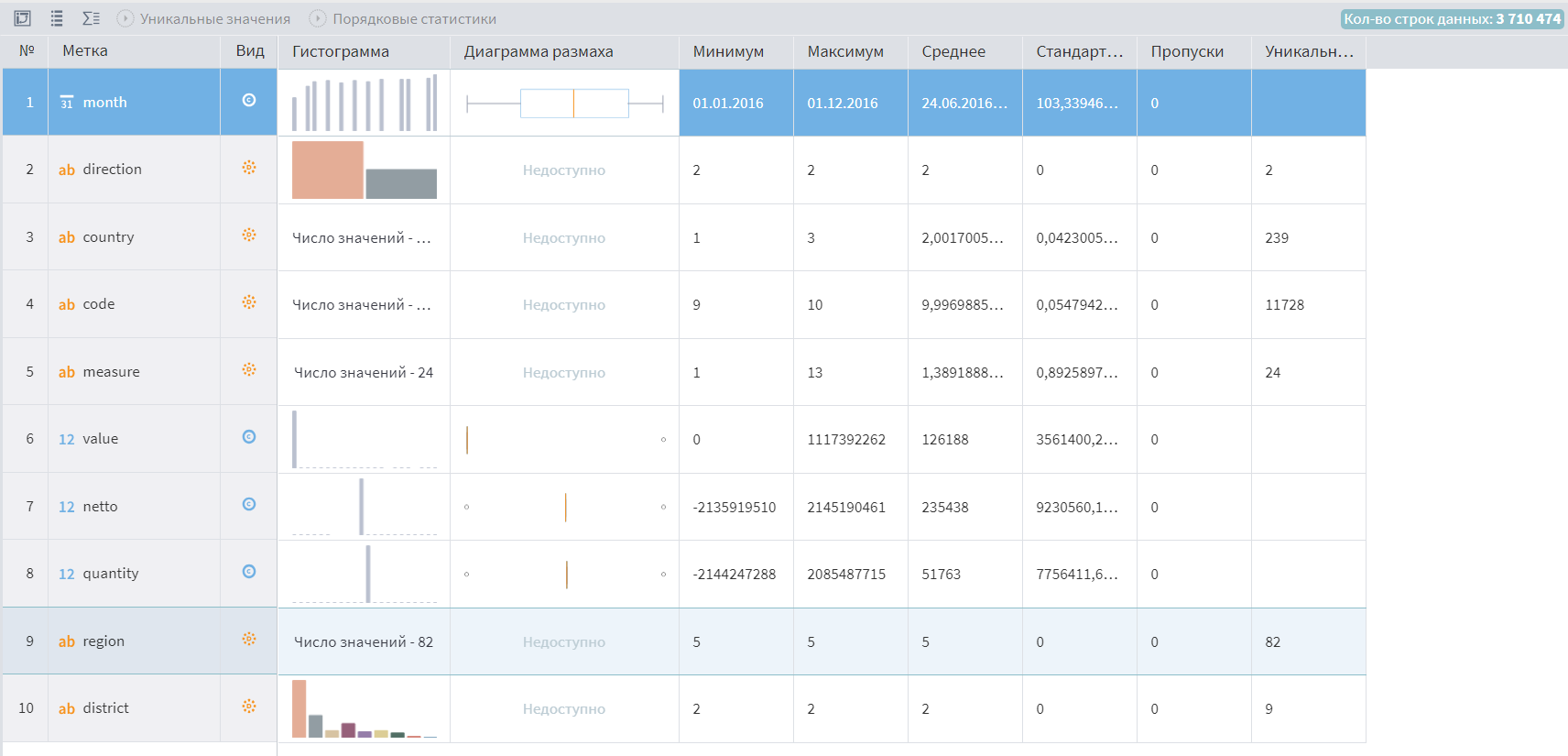
|  |  |
| --- | --- |
| **ОТЧЁТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ № 2** | |
| **по дисциплине «Большие данные»** | |
|  | |
| Выполнил студент группы КМБО-03-22 | Лыков Д.С. |
| Проверил ассистент кафедры ПМ ИИТ | Высоцкая А.А. |

Москва 2024

**Задание 1.**

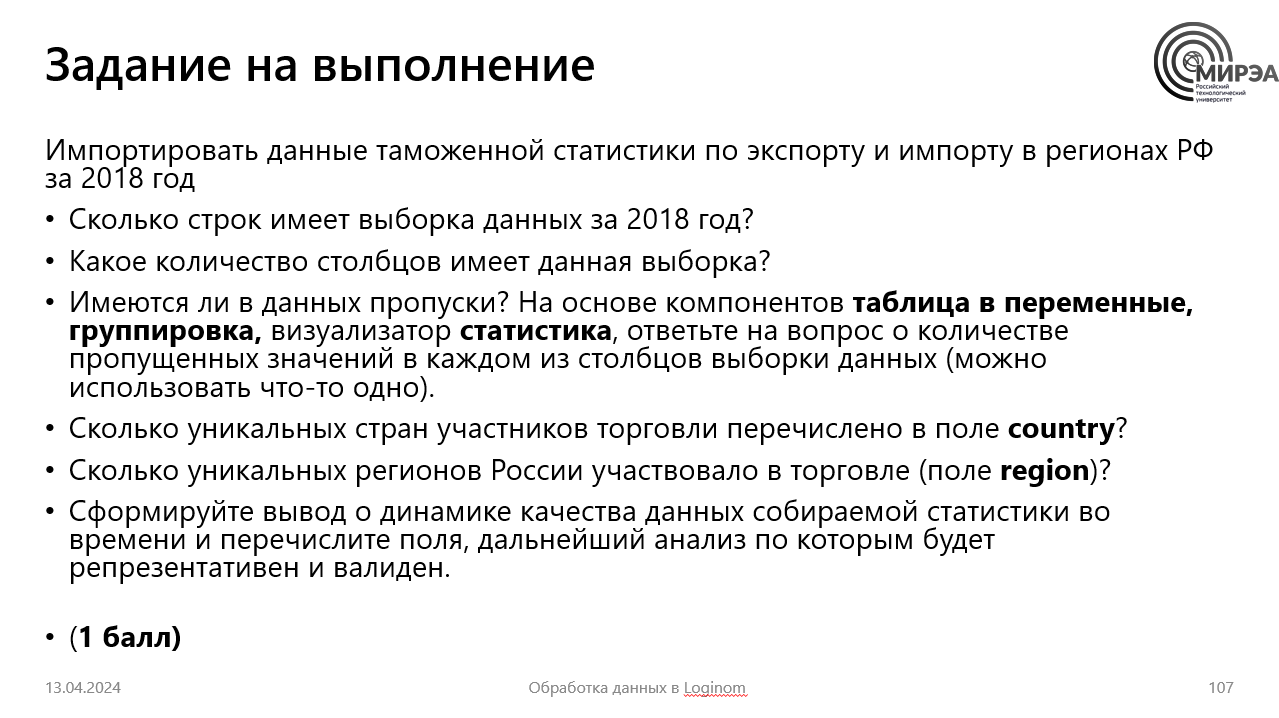


**Ответ.**

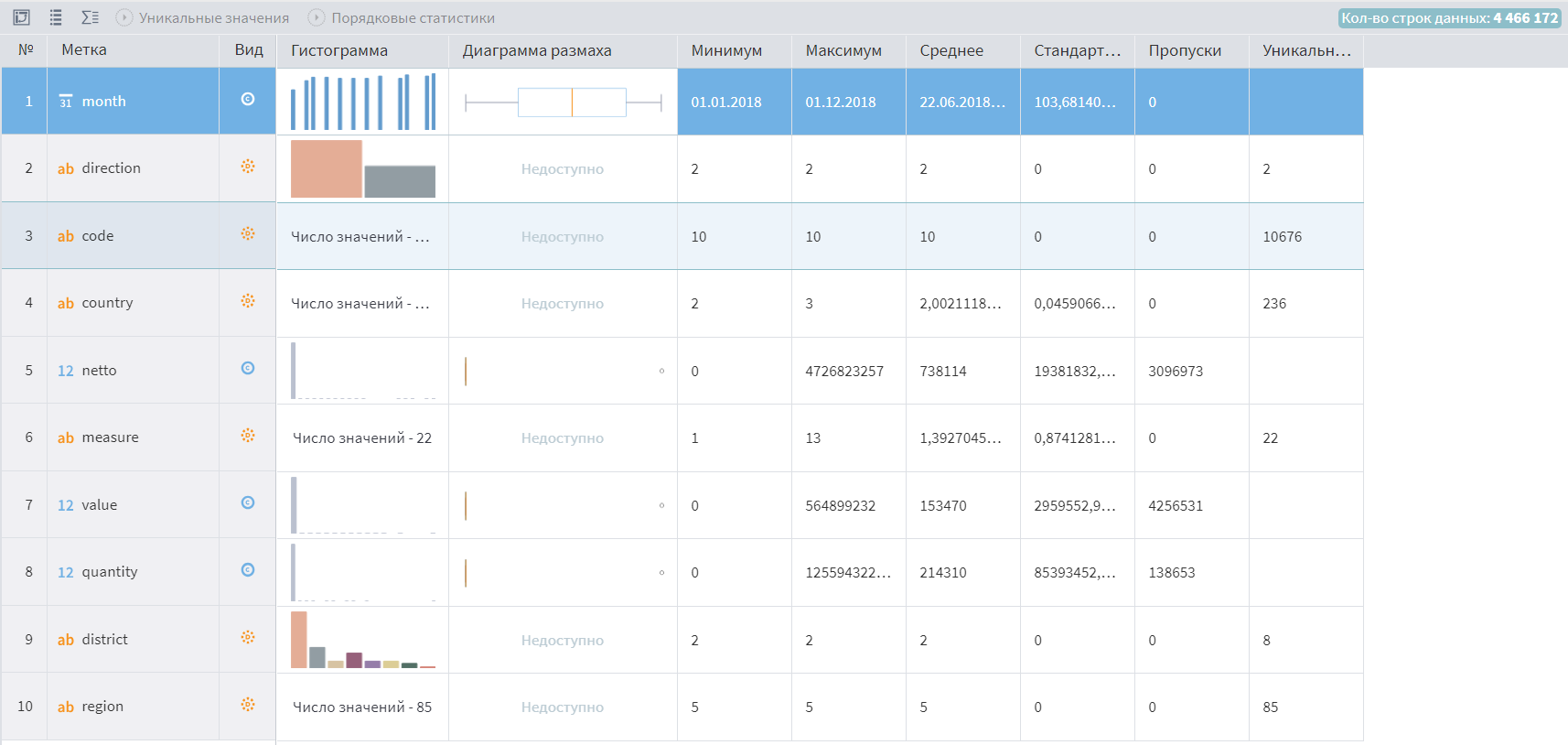


* *Сколько строк имеет выборка данных за 2016 год?* - 3 710 474
* *Какое количество столбцов имеет данная выборка?* - 10
* *Какую информацию хранит каждое поле в отдельности? Какие данные необходимо еще добыть, чтобы расшифровать значения в полях* ***country, code, region, district****?* - Месяц, направление, страну, код, единицу измерения, количество, массу нетто, количество, регион, федеральный округ.
* Данные не имеют пропусков.
* 239 уникальных стран.
* 82 уникальных региона России.

**Задание 2.**

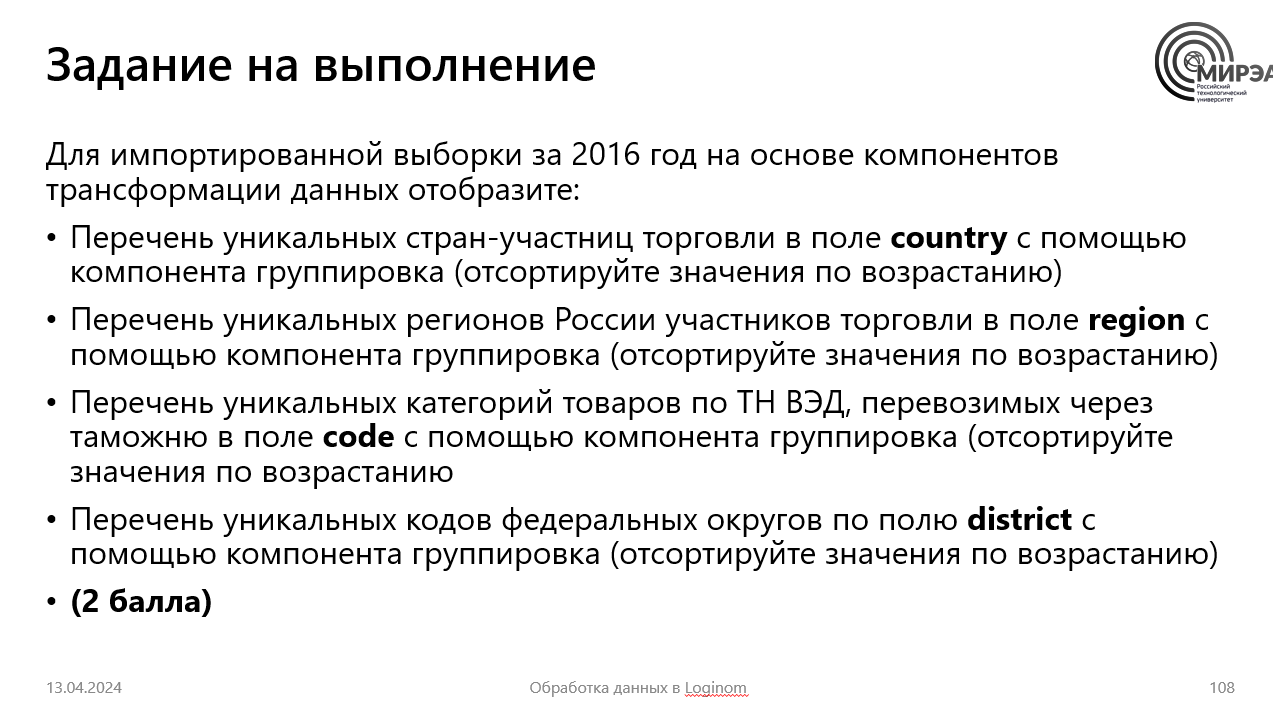


**Ответ.**

****

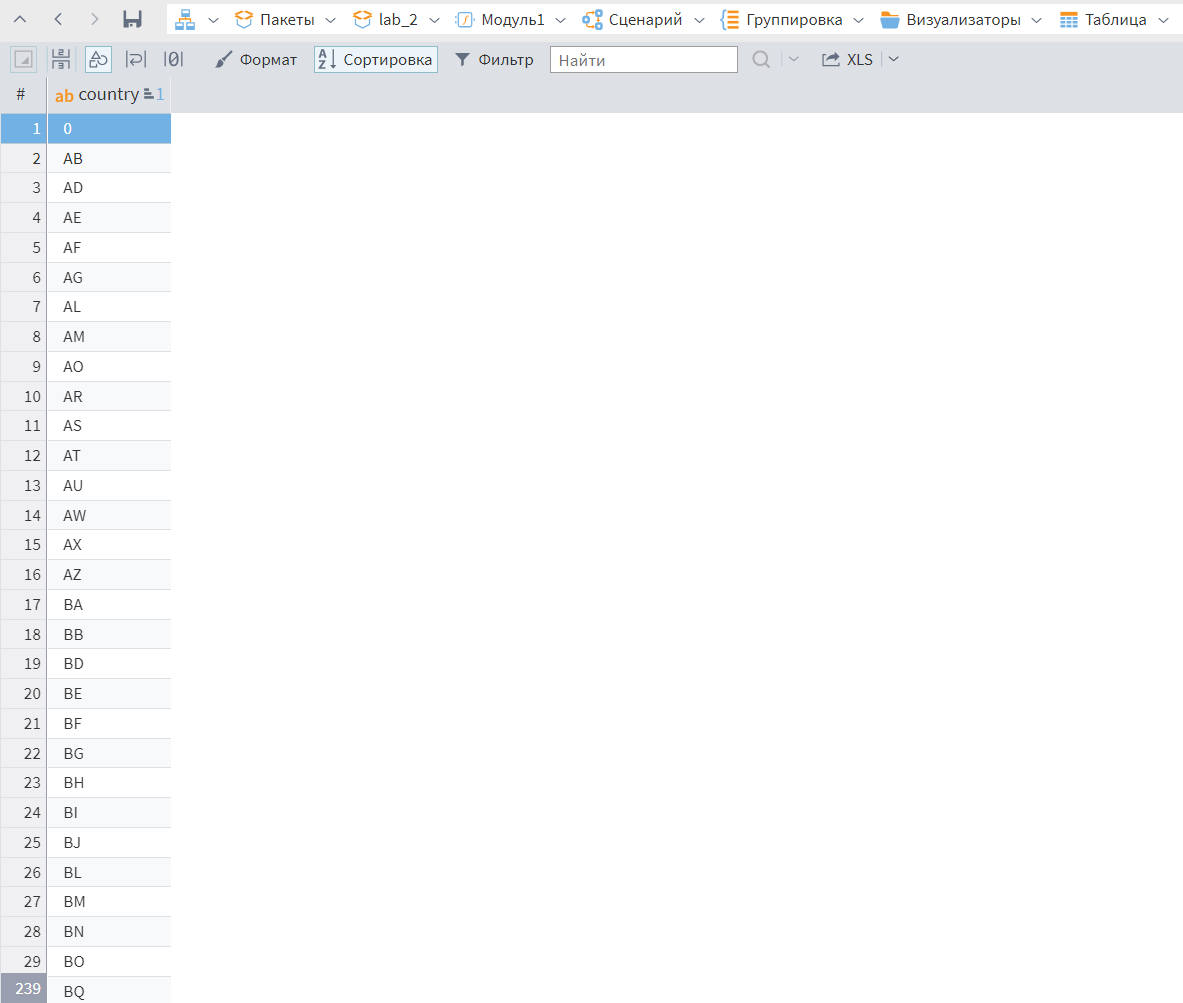
* 4 446 172 – строк за 2018 год
* 10 столбцов
* Пропусков: Value – 4 256 531, Netto – 3 096 973, Quantity – 138 653
* Уникальных стран – 236
* Уникальных регионов – 85
* Появилось ухудшение качества данных. Для анализа следует использовать только те поля, где нет пропущенных значений. Возможно, будет полезно использование полей с пропусками, но только при условии, что эти пропуски не зависят от других полей.

**Задание 3.**

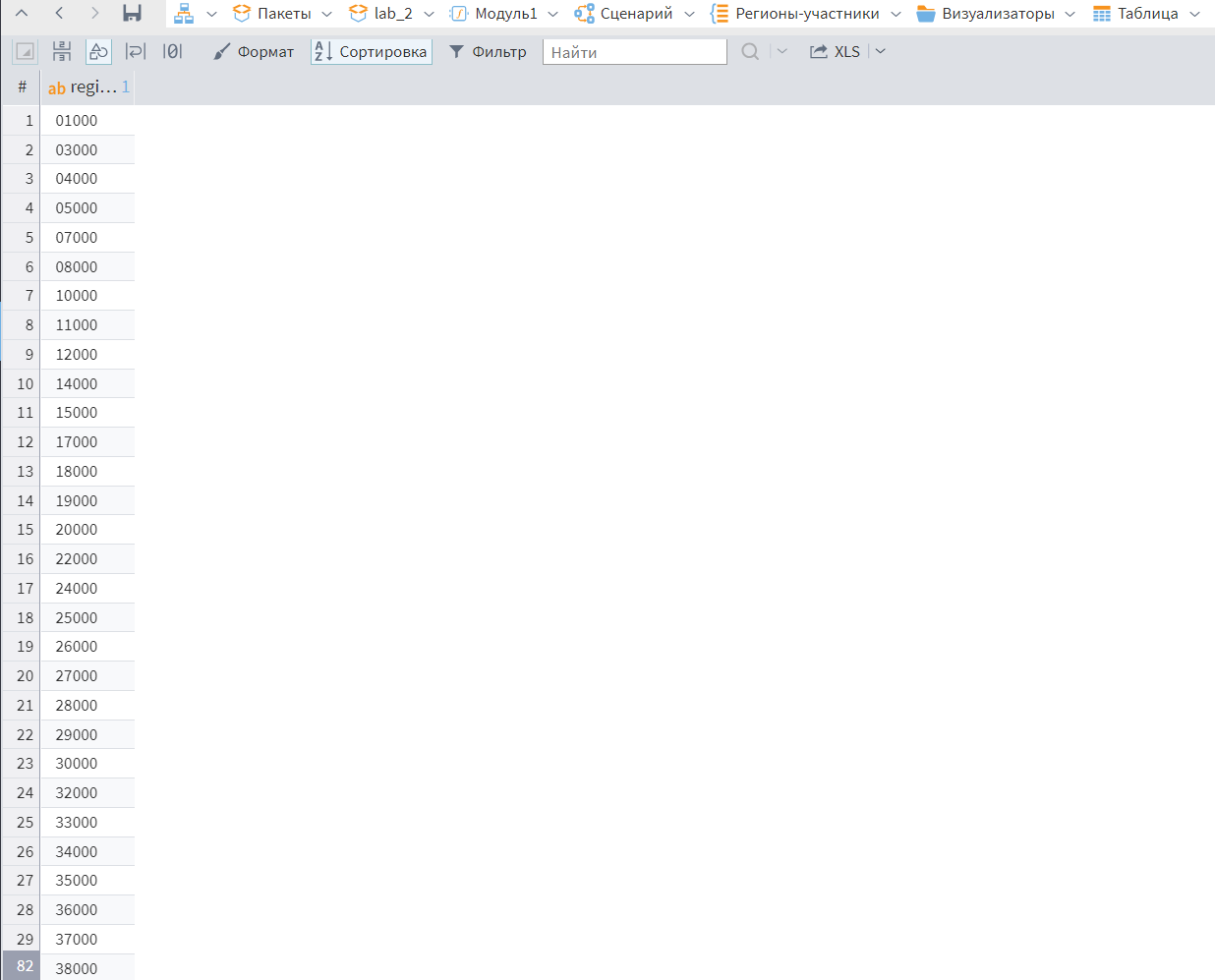
****

**Ответ.**

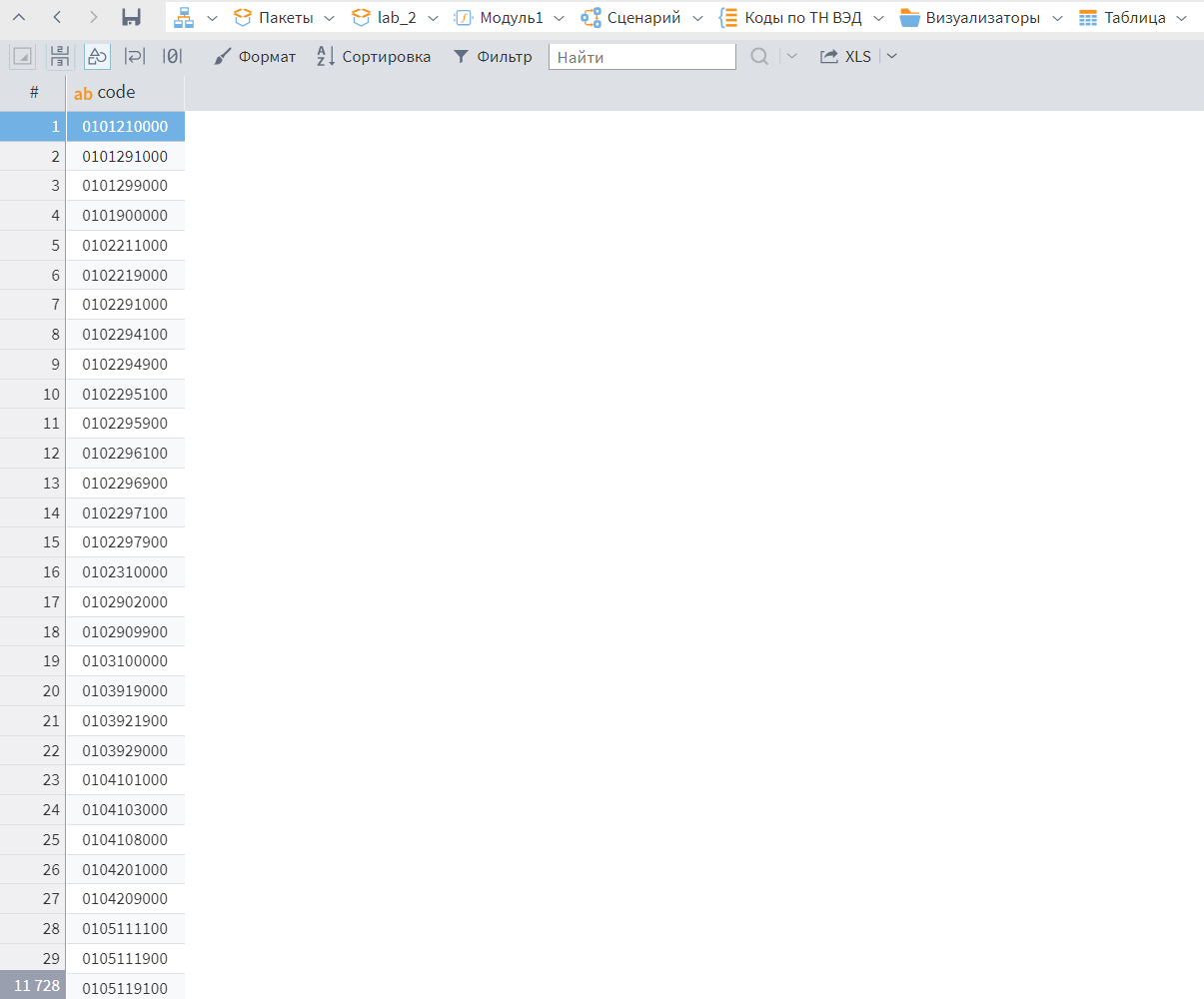
Перечень уникальных стран-участниц торговли, отсортированных по возрастанию.

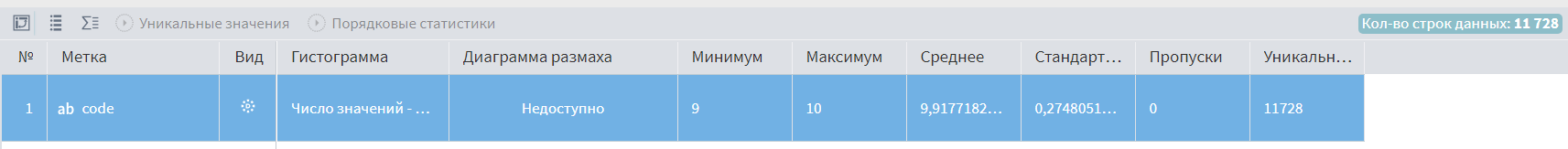


Перечень уникальных регионов России участников торговли, отсортированных по возрастанию.



Перечень уникальных кодов ТН ВЭД, отсортированных по возрастанию.





Перечень уникальных кодов федеральных кодов, отсортированных по возрастанию.

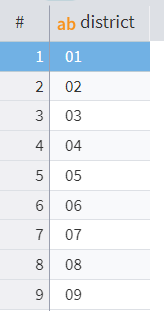
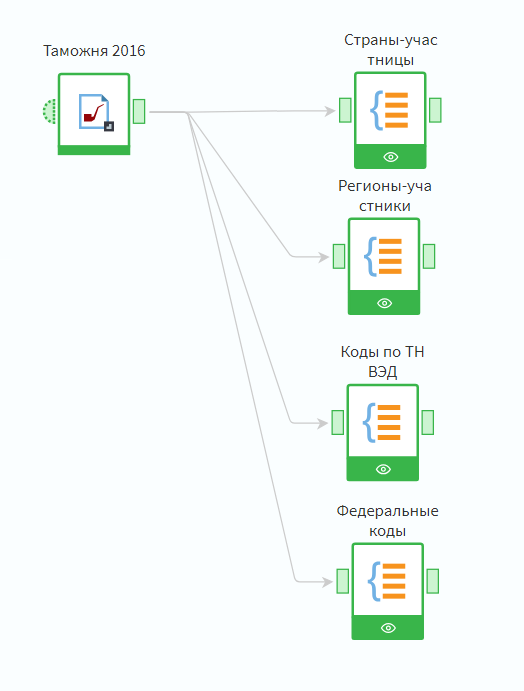
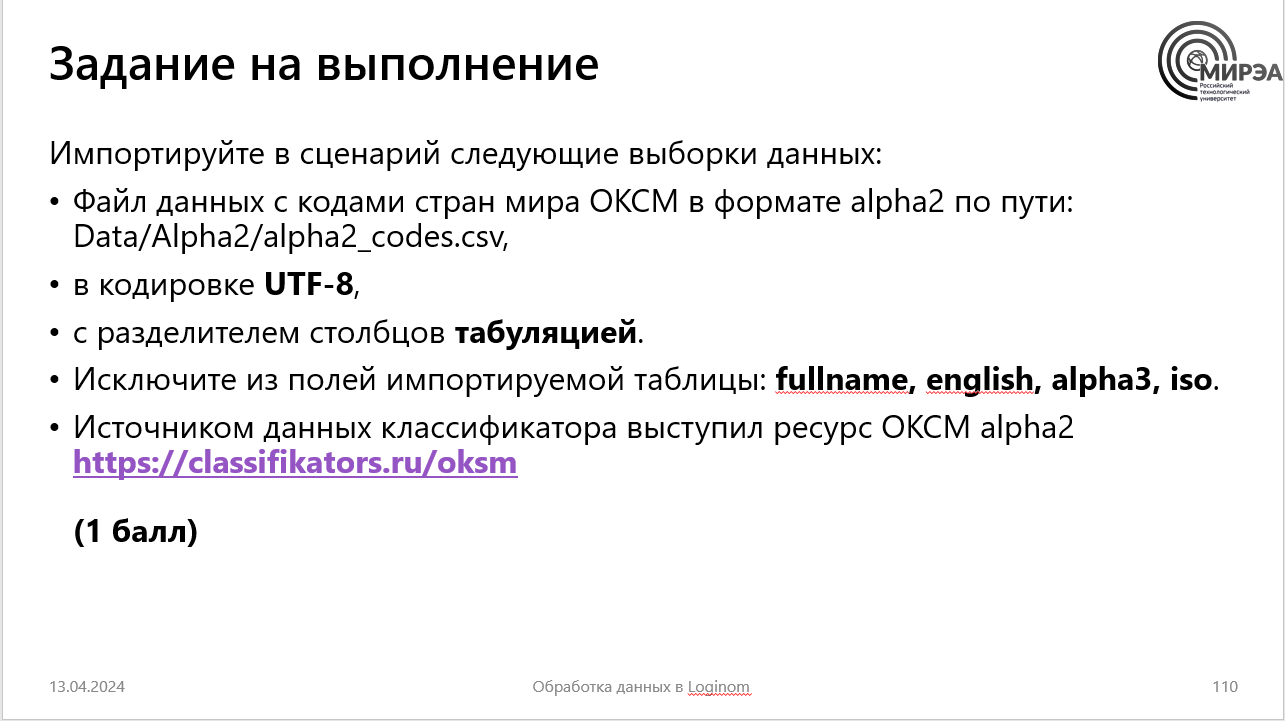


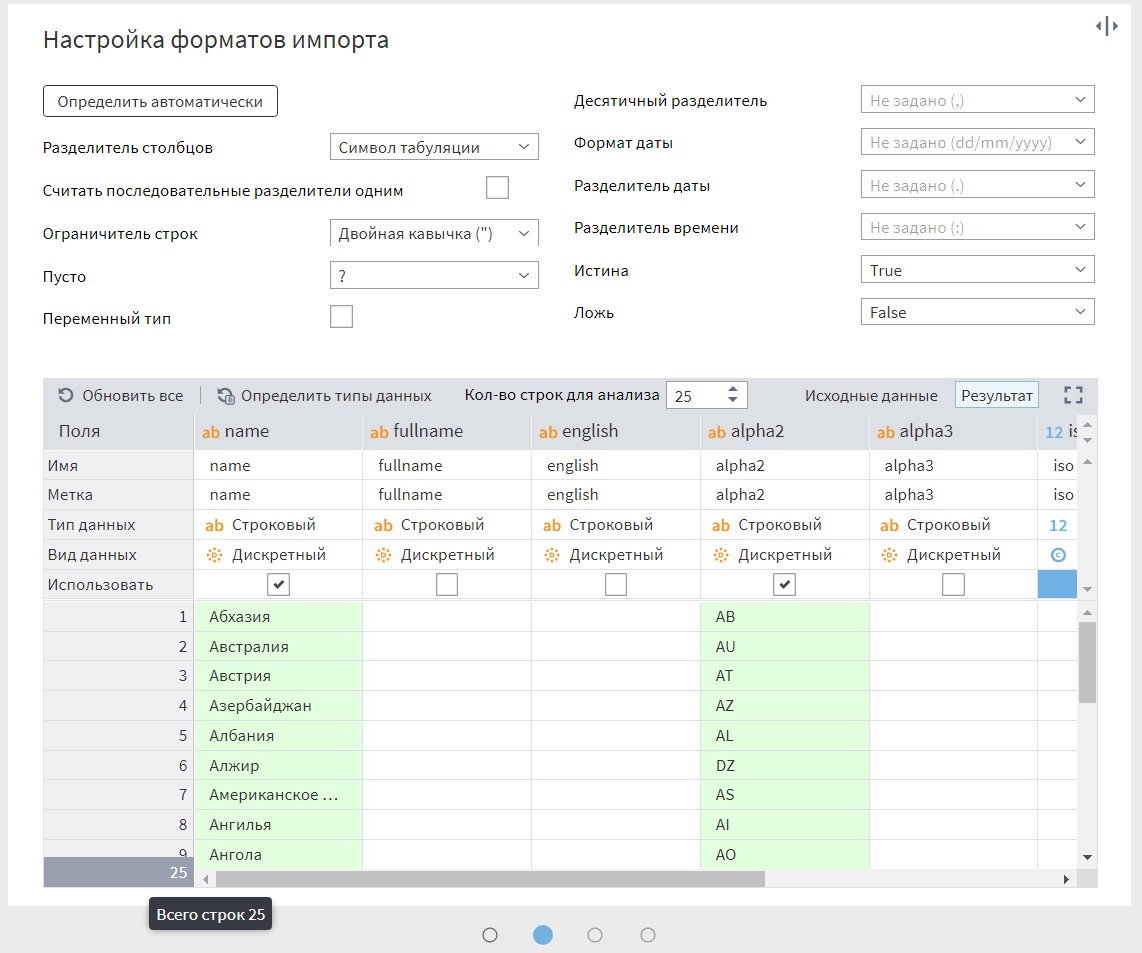
Схема.

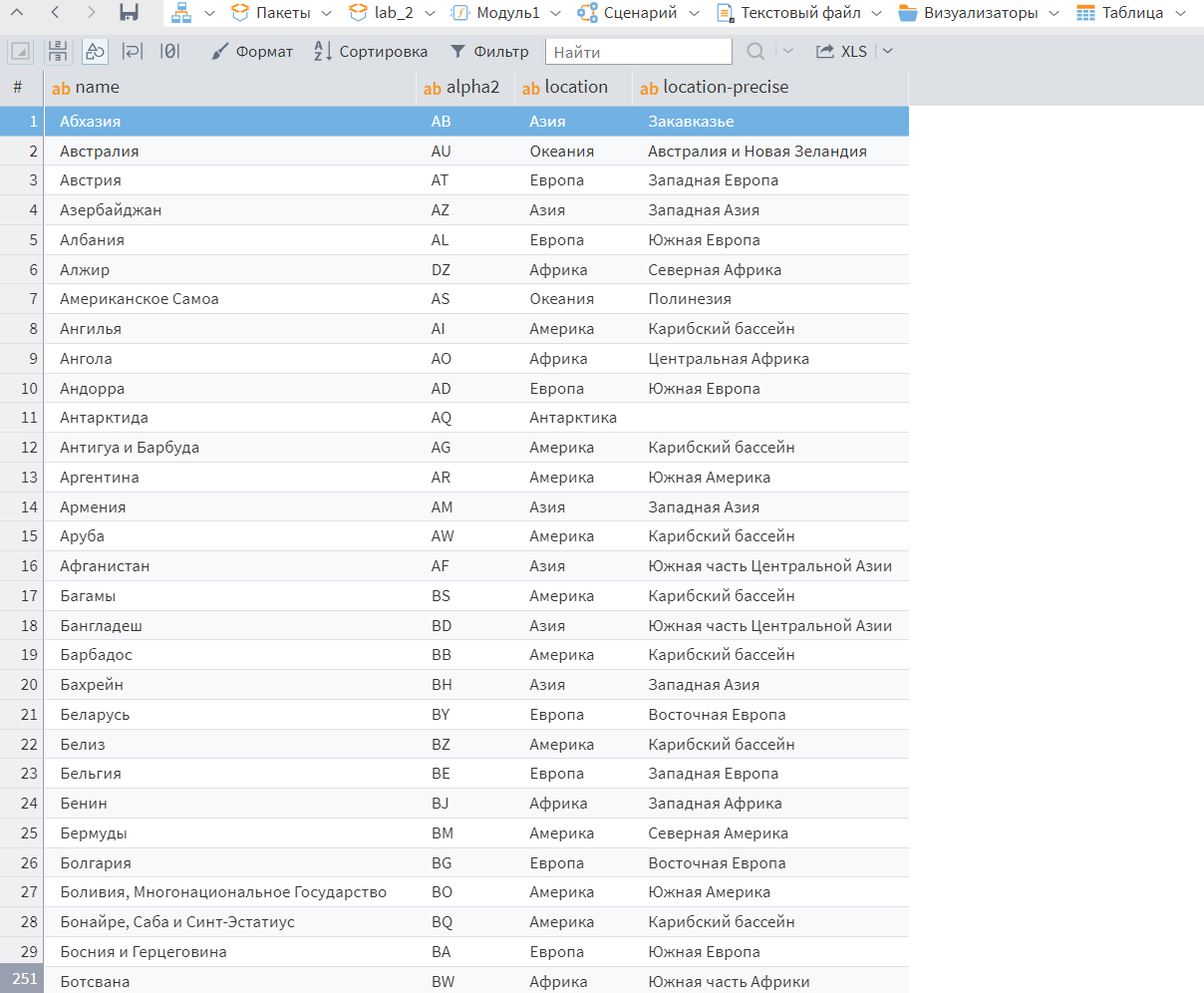


**Задание 4.**

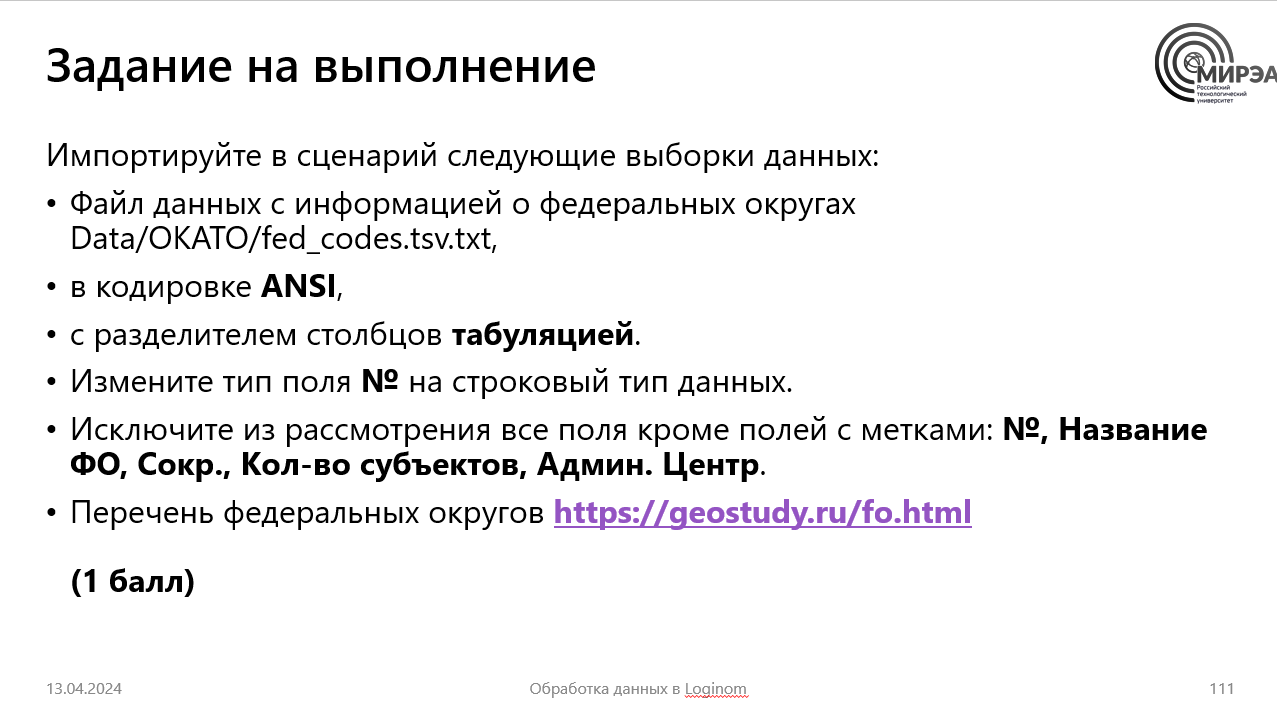


**Ответ.**

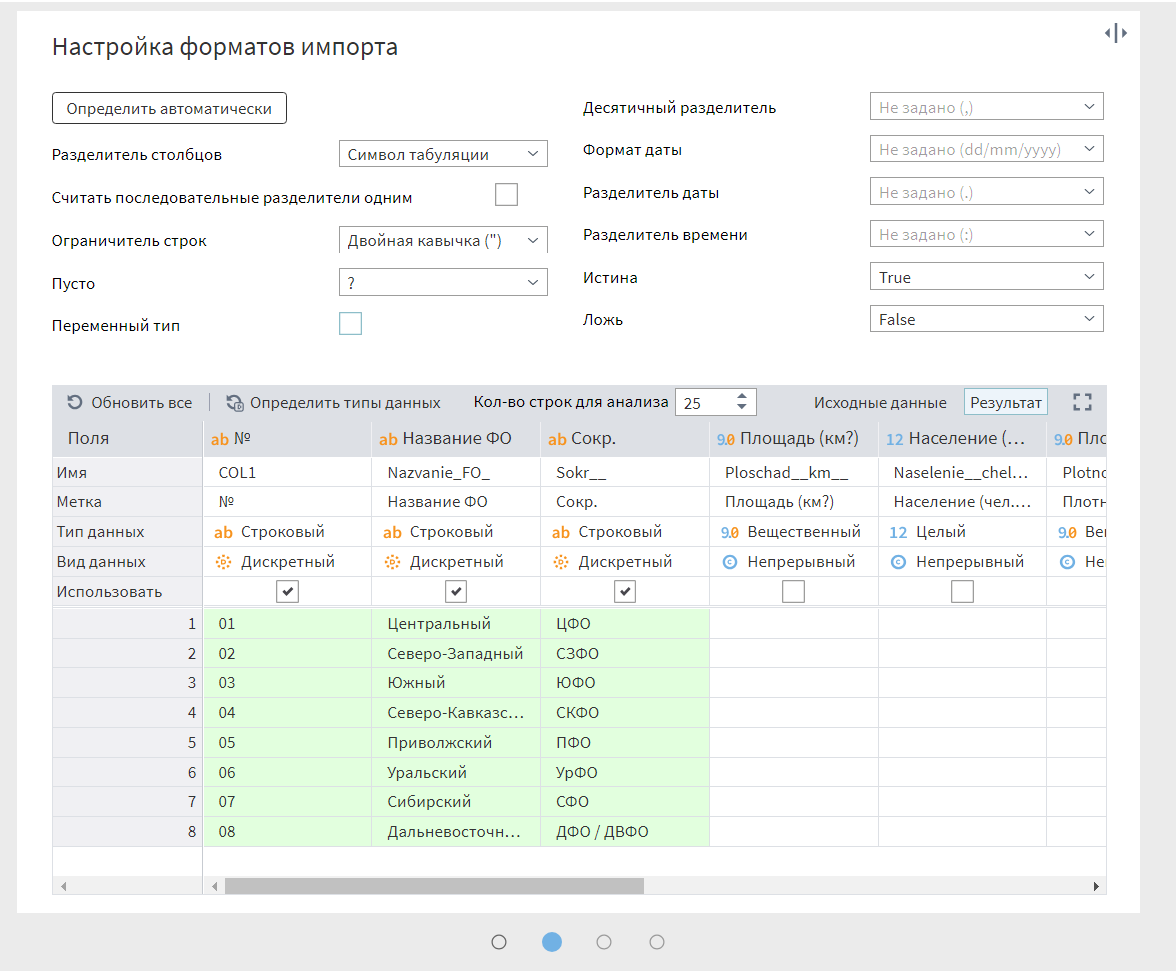
****

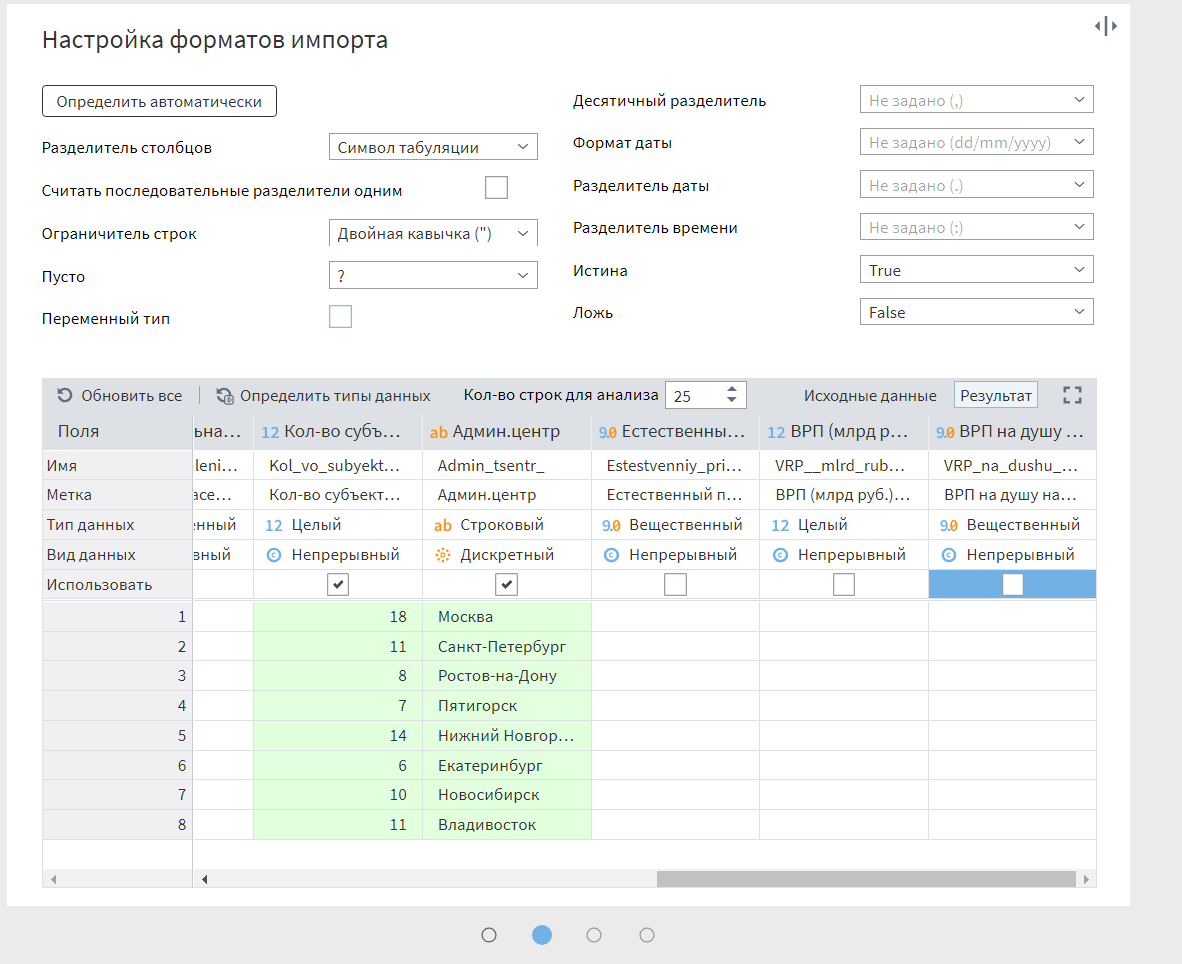
****

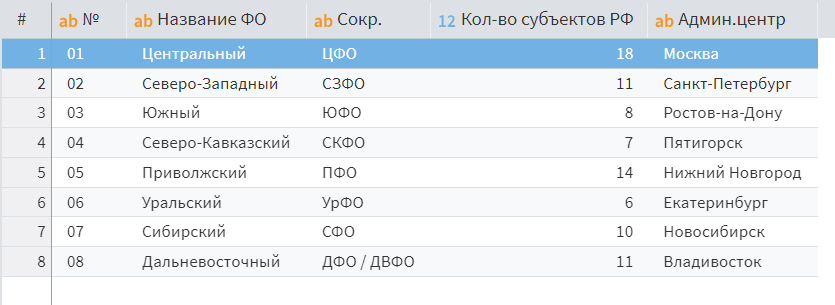
**Задание 5.**

****

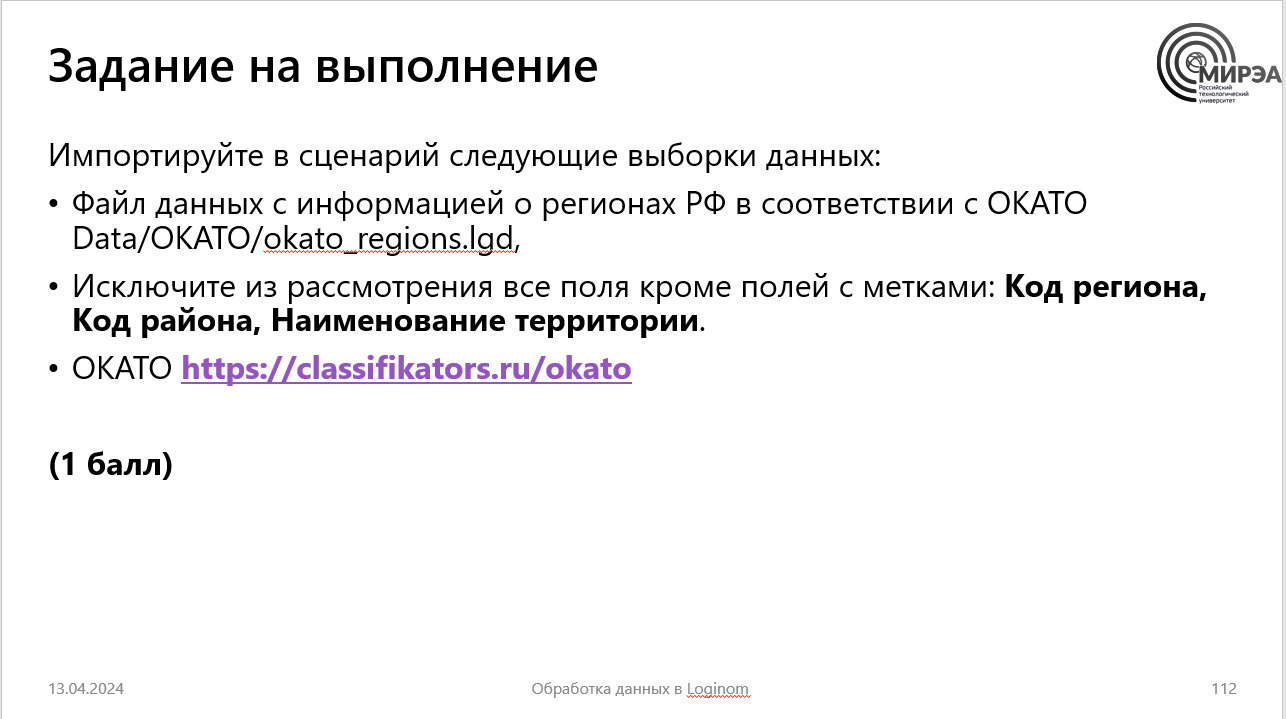
**Ответ.**

****

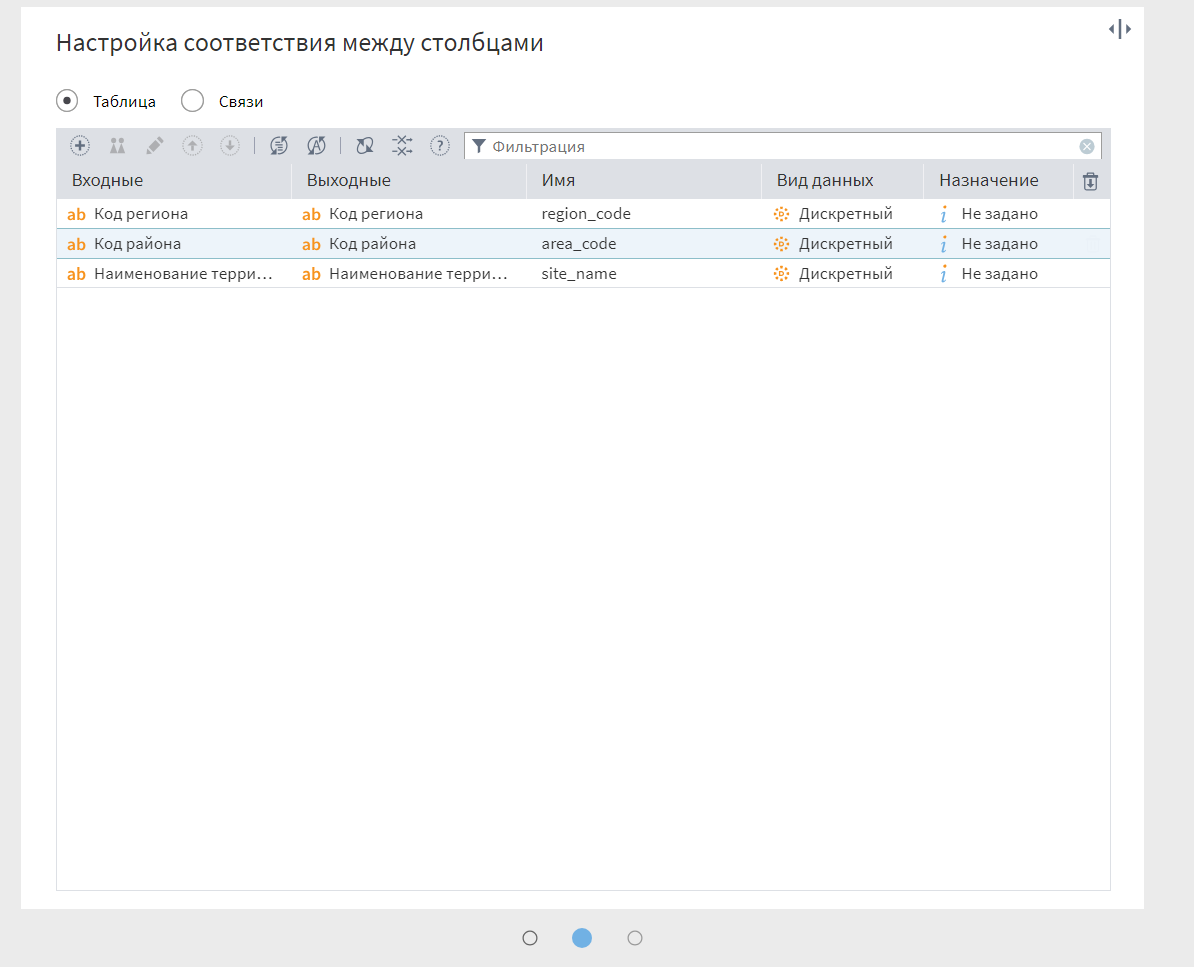
****

****

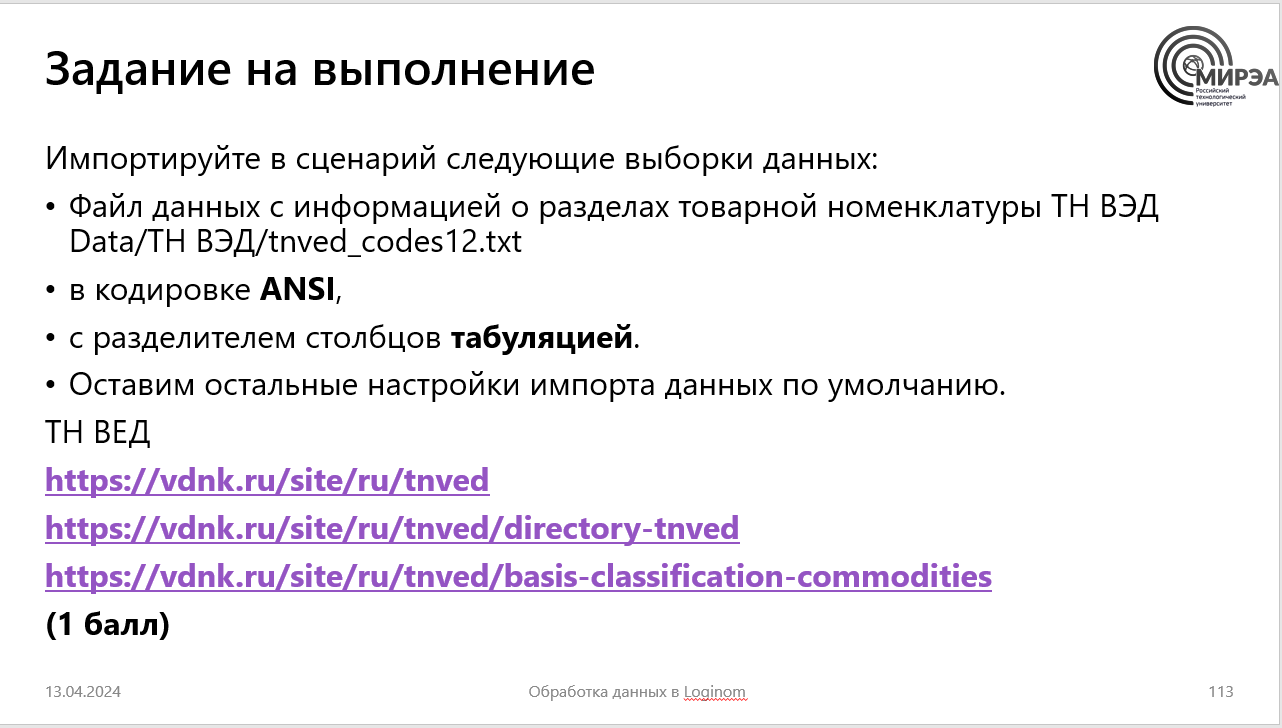
**Задание 6.**

****

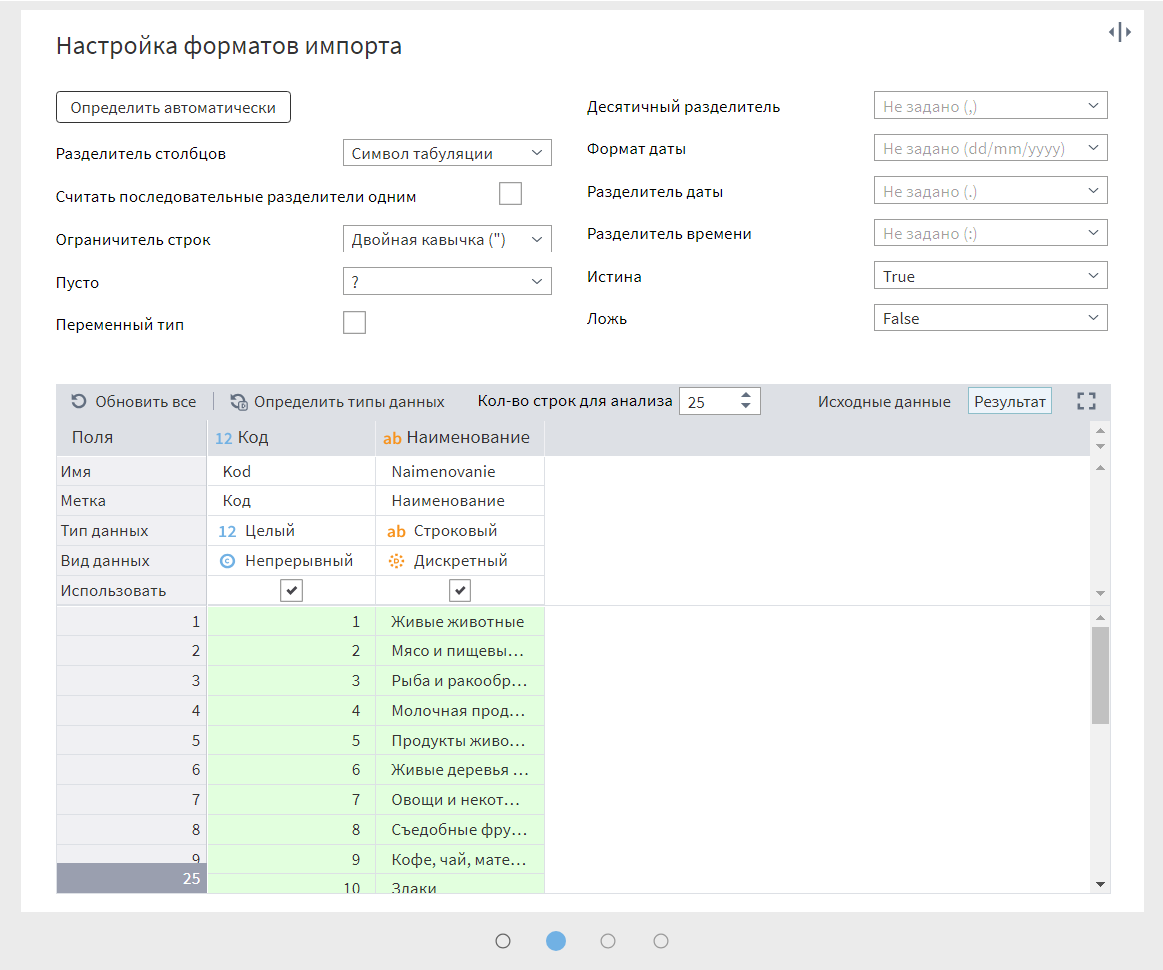
**Ответ.**

****

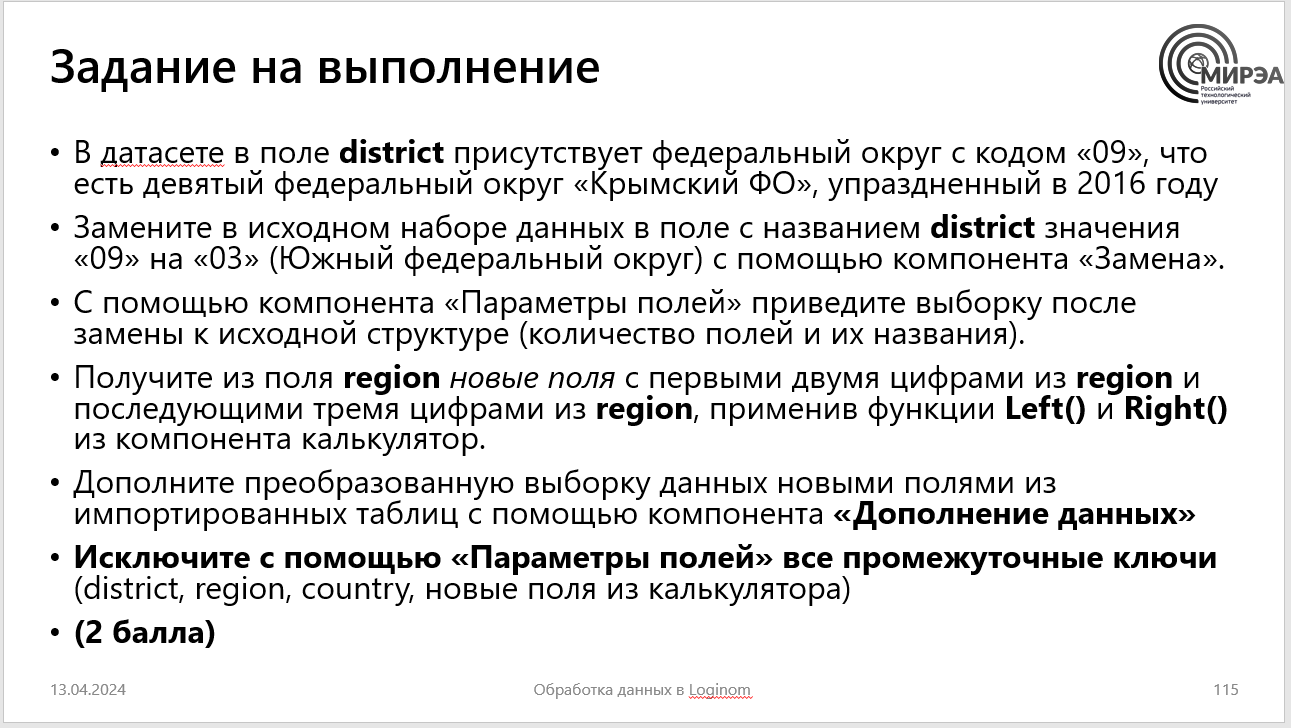
**Задание 7.**

****

**Ответ.**

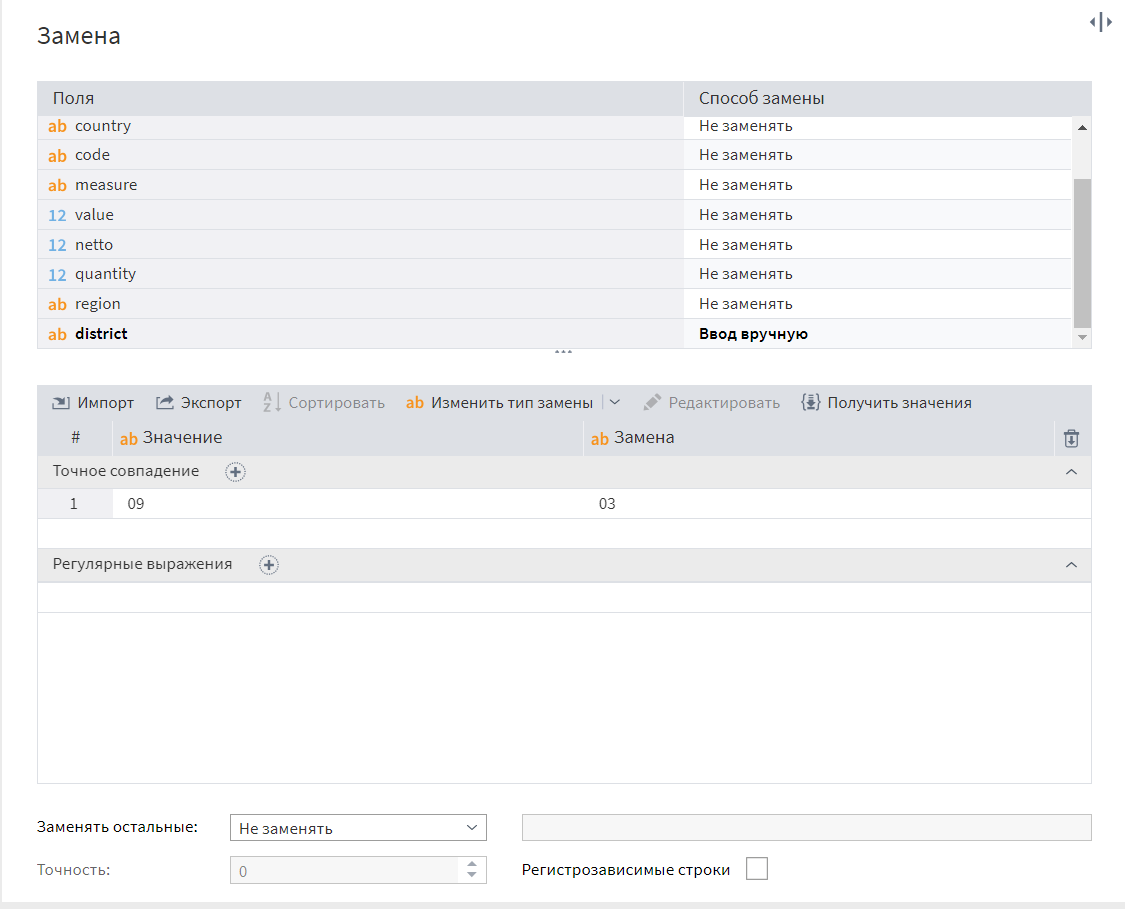
****

**Задание 8.**

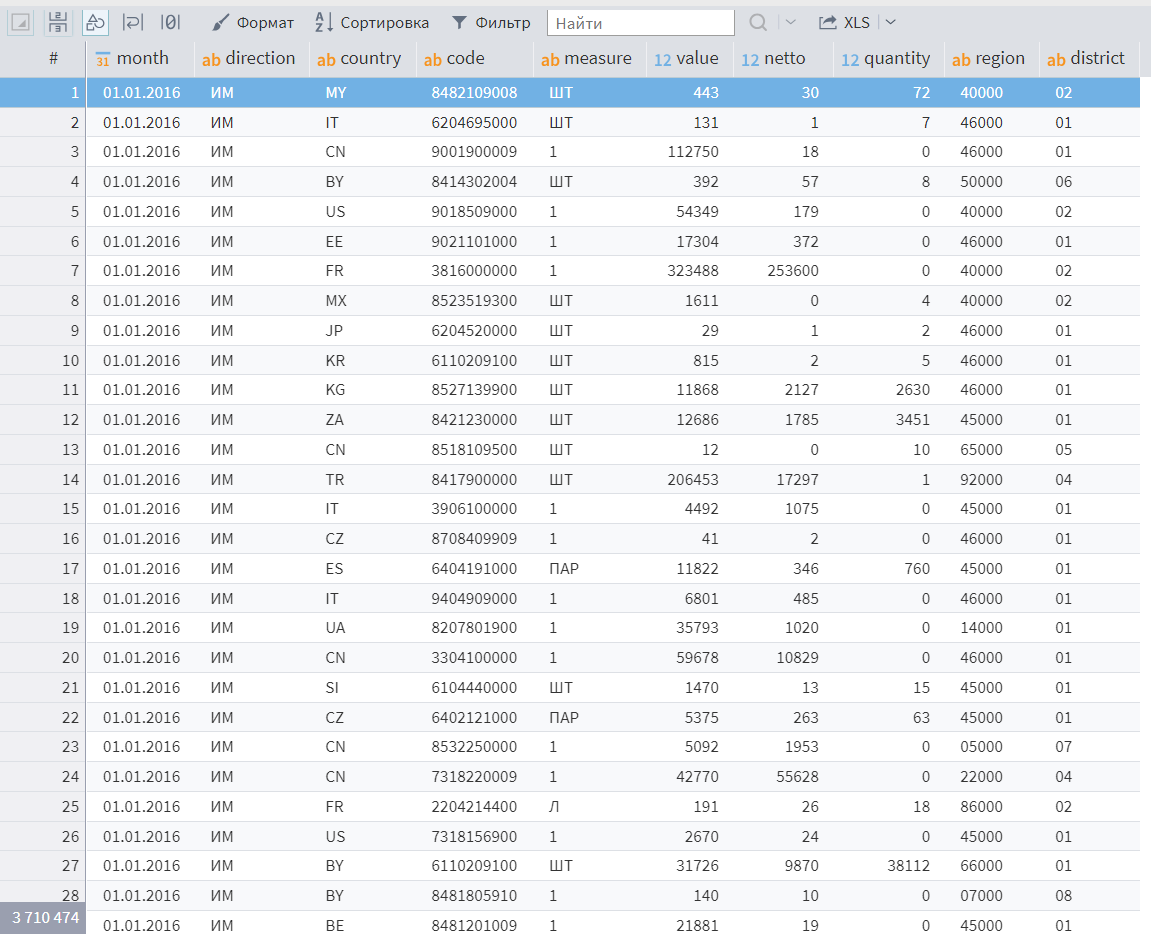
****

**Ответ.**

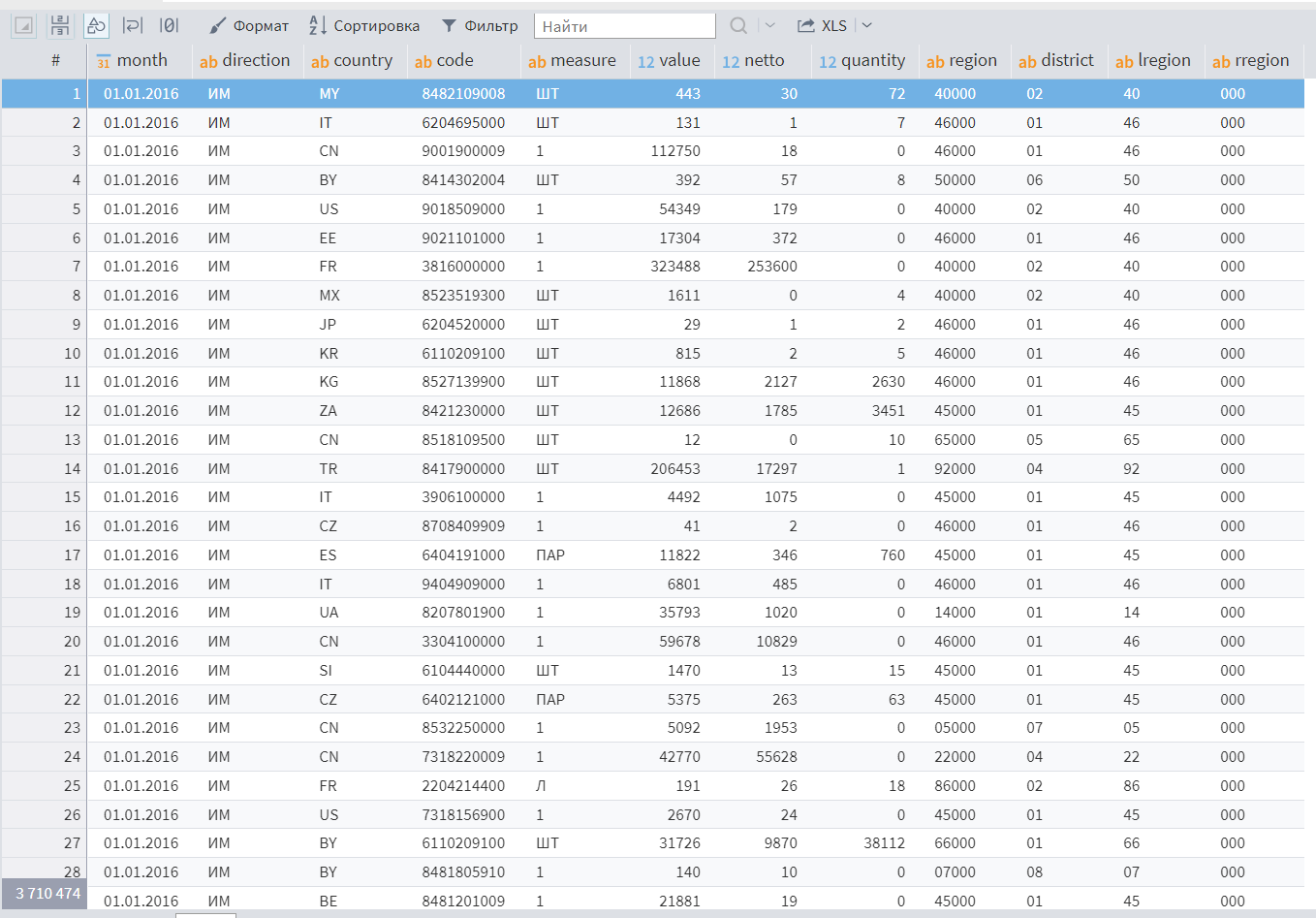
Замена «09» на «03» в поле **distinct**.

****

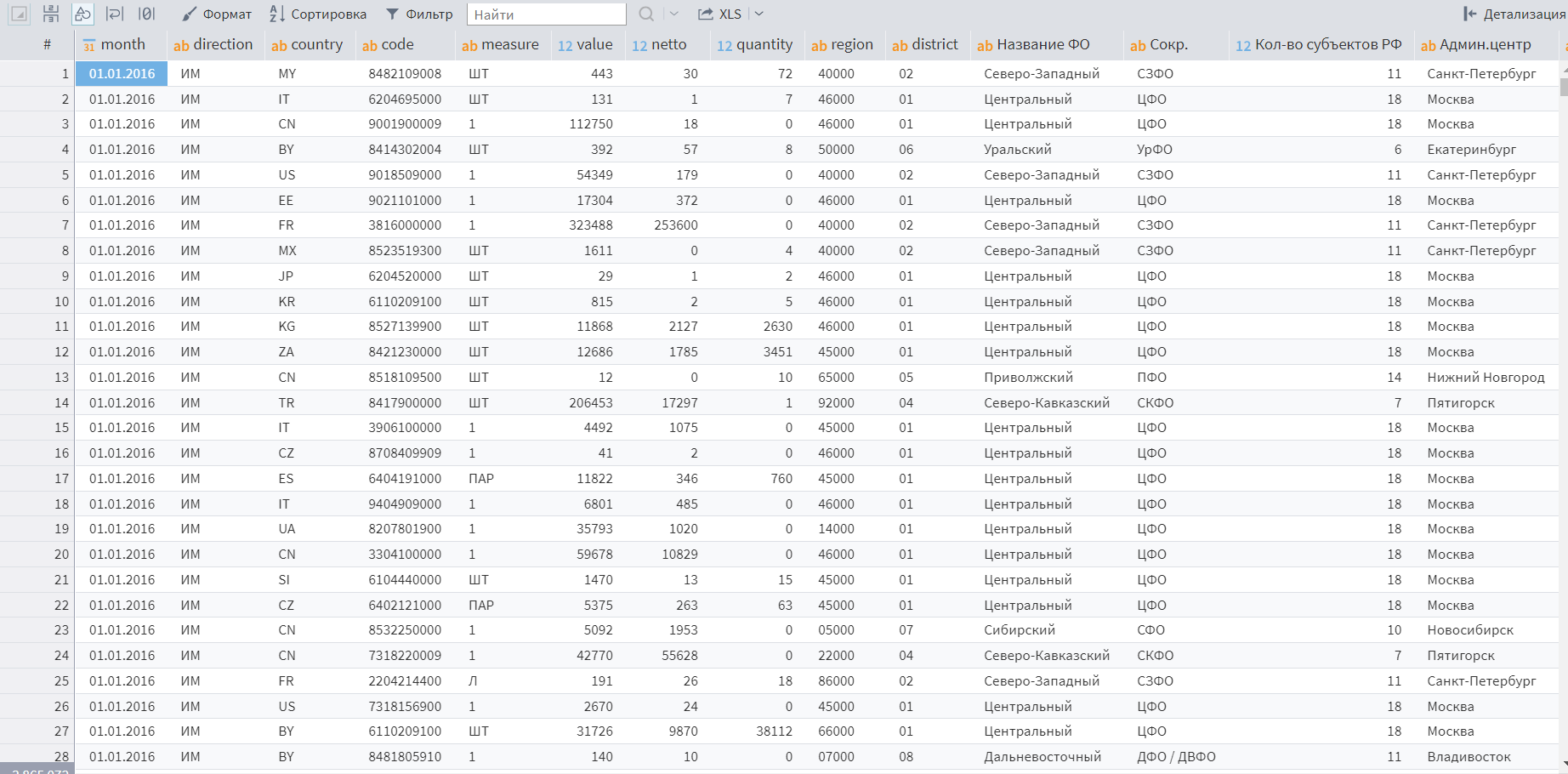
Возвращение к первозданному виду с помощью «параметра полей».



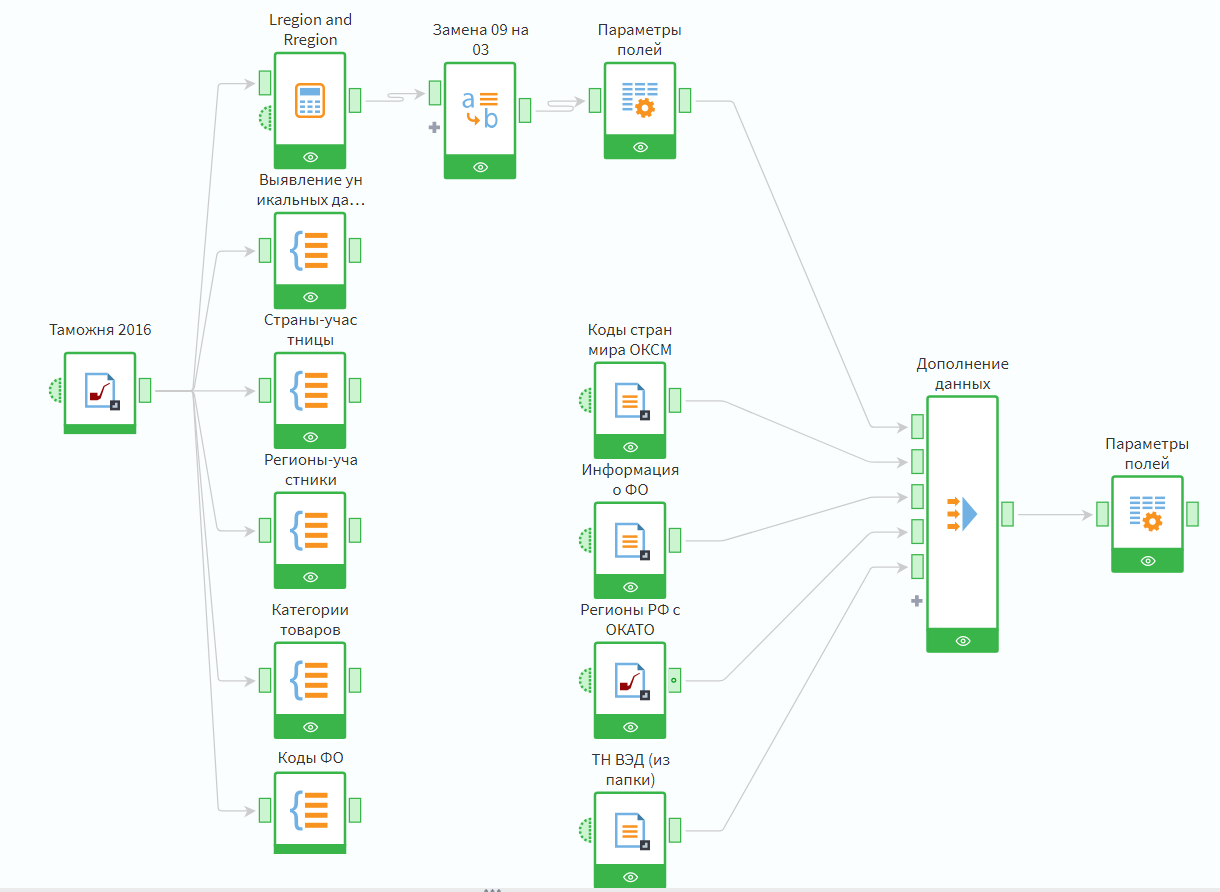
Добавление lregion и rregion, как новых полей, содержащих результаты использования функций Left и Right в region.



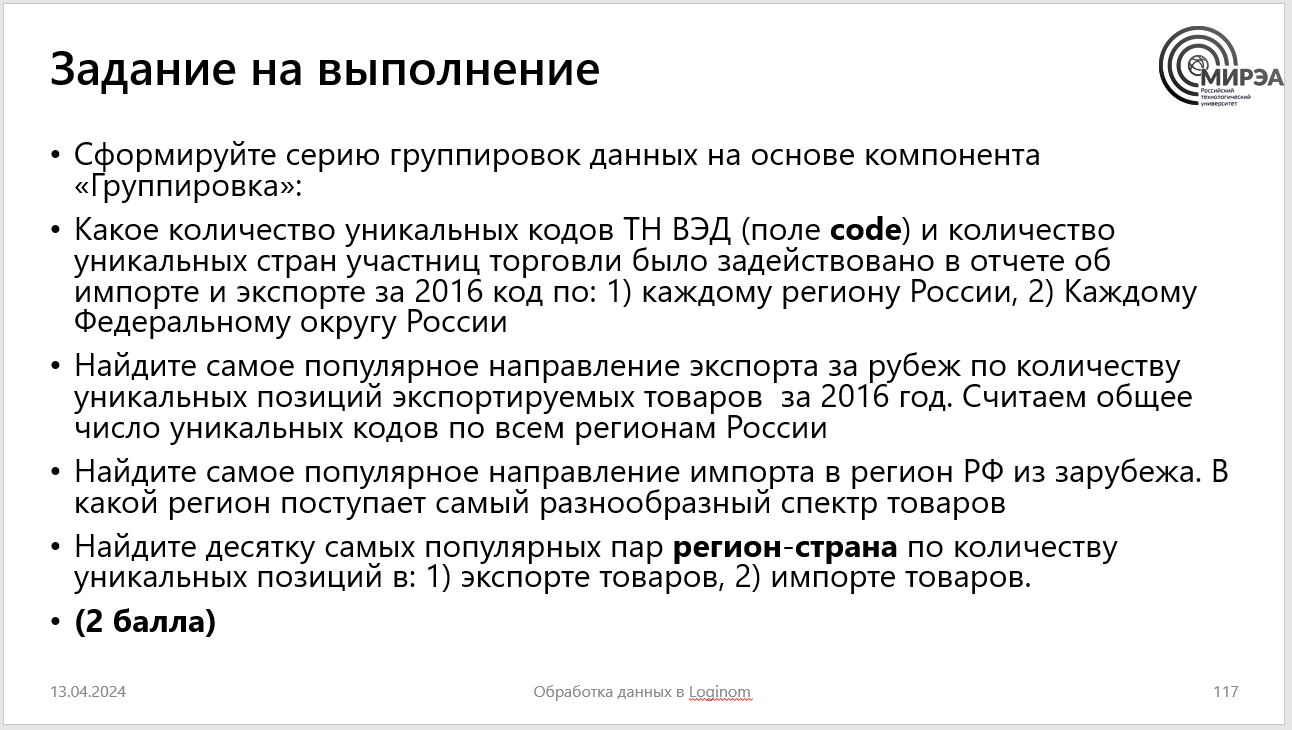
Слияние таблицы и всех импортированных таблиц. Исключение с помощью «Параметры полей» всех промежуточных ключей. Конечная таблица.



Итоговая схема.

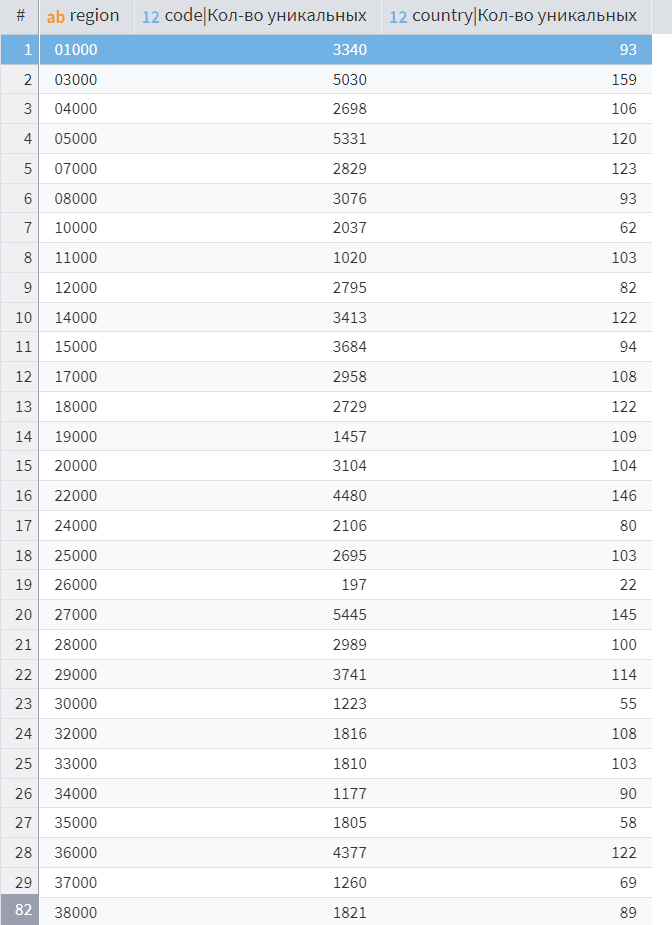


**Задание 9.**

****

**Ответ.**

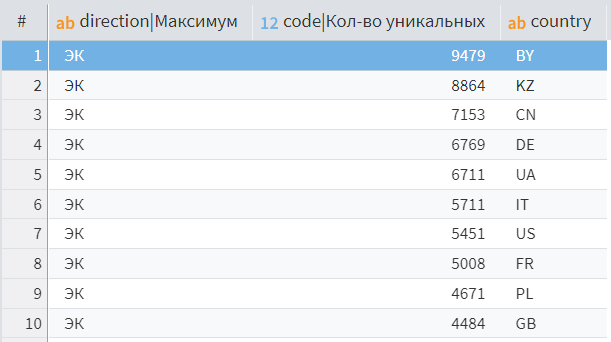
Кол-во уникальных кодов ТН ВЭД и количество уникальных стран участниц торговли по каждому региону России.



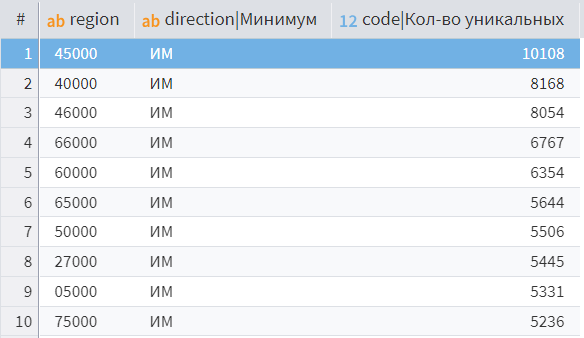
Кол-во уникальных кодов ТН ВЭД и количество уникальных стран участниц торговли по ФО России.

****

Самое популярное направление экспорта (по уникальным значениям кодов товара).

****

Самое популярное направление импорта (по уникальным значениям кодов товара). В какой регион поступает самый разнообразный спектр товаров?



Найдите десятку самых популярных пар **регион**-**страна** по количеству уникальных позиций в:

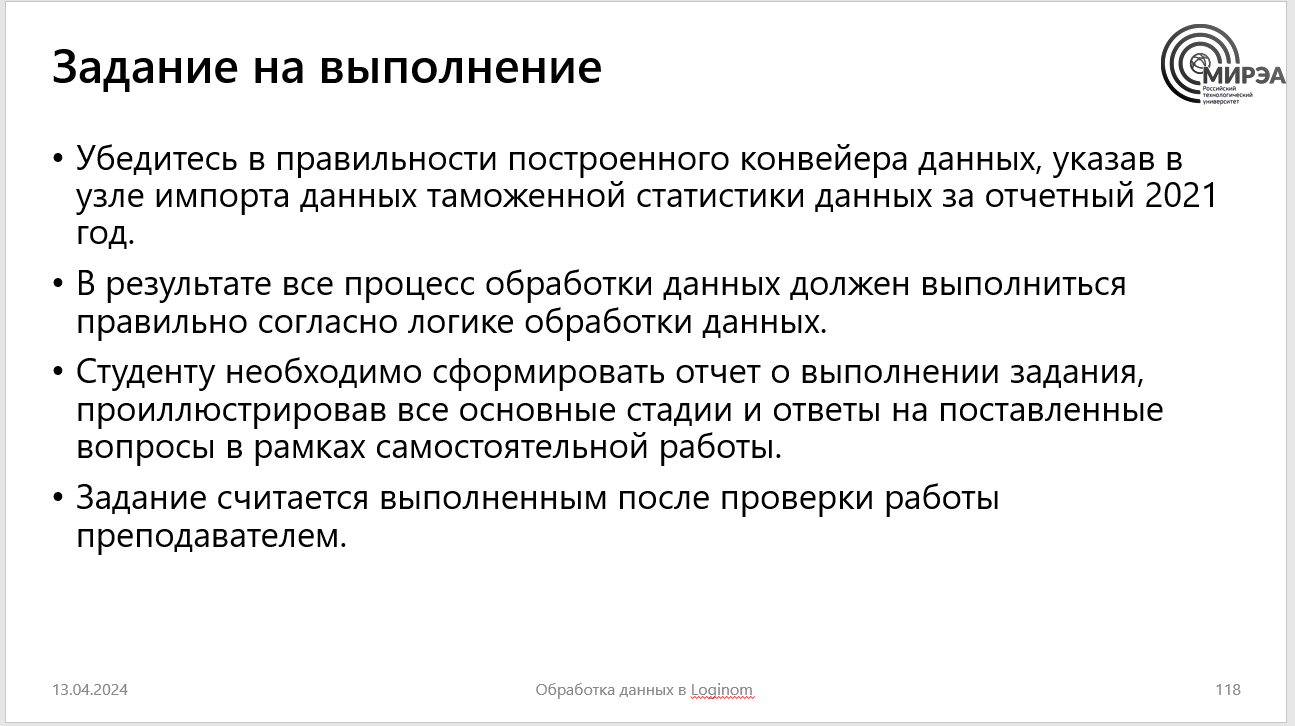
1) экспорте товаров



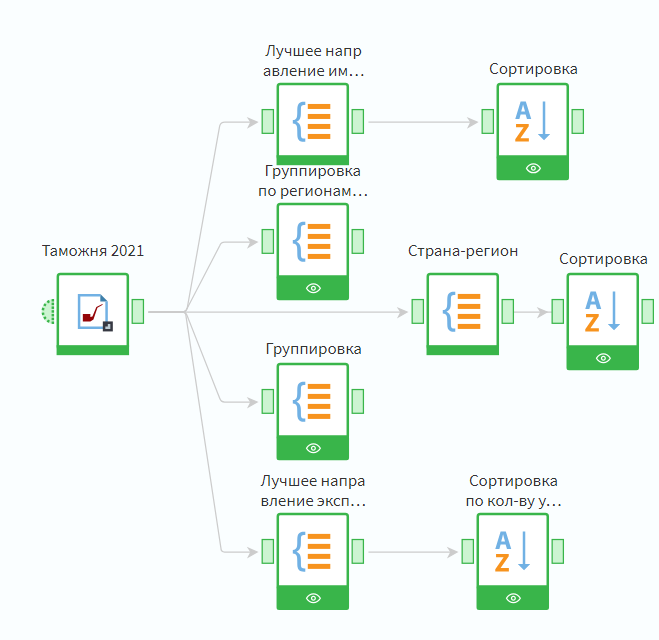
2) импорте товаров



**Задание 10.**

****

**Ответ.**

****