

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

С. В. Рындина

ОСНОВЫ АНАЛИЗА БОЛЬШИХ МАССИВОВ ДАННЫХ (GLARUS BI)

Учебно-методическое пособие

ПЕНЗА 2025

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Пензенский государственный университет» (ПГУ)

С. В. Рындина

Основы анализа больших массивов данных (Glarus BI)

Учебно-методическое пособие

Пенза
Издательство ПГУ
2025

УДК 004.738.5(075)+339(075)

P95

Р е ц е н з е н т

кандидат экономических наук, доцент,
руководитель сектора методологии учета и оформления
банковских операций ПАО Банк «Кузнецкий»

О. Б. Волошина

Рындина, Светлана Валентиновна.

P95

Основы анализа больших массивов данных (Glarus BI) : учеб.-метод. пособие / С. В. Рындина. – Пенза : Изд-во ПГУ, 2025. – 44 с.

Рассматриваются платформа Glarus BI и практические аспекты ее использования для анализа больших массивов данных. Материал соответствует программе дисциплины «Основы анализа больших массивов данных» и может быть использован при написании выпускной работы бакалавра.

Издание подготовлено на кафедре «Цифровая экономика» ПГУ и предназначено для обучающихся по направлению подготовки 38.03.02 «Менеджмент».

УДК 004.738.5(075)+339(075)

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
1. СИСТЕМА BUSINESS INTELLIGENCE – GLARUS BI	5
2. ИМПОРТ ДАННЫХ. ТИПЫ ДАННЫХ.....	7
3. ПОСТРОЕНИЕ ЗАПРОСОВ	12
4. ОБЪЕДИНЕНИЕ ДАННЫХ.....	24
5. ПОСТРОЕНИЕ ДАШБОРДОВ	29
6. ОБЪЕДИНЕНИЕ ВИЗУАЛИЗАЦИЙ.....	36
Лабораторная работа. Создание отчета средствами Glarus BI.....	40
СПИСОК ЛИТЕРАТУРЫ.....	43

ВВЕДЕНИЕ

В соответствии с учебным планом обучающиеся на втором курсе направления подготовки «Менеджмент» изучают дисциплину «Основы анализа больших массивов данных».

Целью учебно-методического пособия является знакомство студентов с анализом массивов данных на основе использования платформы Glarus BI от компании ООО «Гларус Диджитал».

На примере Glarus BI рассмотрены практические вопросы анализа массивов данных: импорт данных, группировка, слияние и фильтрация данных, а также построение различных визуализаций.

Задания для лабораторных работ имеют практическую направленность, позволяют познакомиться с основными компонентами Glarus BI на конкретных бизнес-задачах. В процессе выполнения лабораторной работы формируются умения и навыки работы с большими массивами данных: подключение источников данных, использование статистических вычислений и методов визуализации для подготовки отчетов.

1. СИСТЕМА BUSINESS INTELLIGENCE – GLARUS BI

Glarus BI – это российская BI-платформа уровня Enterprise для бизнес-аналитики [1]. Сервис позволяет подключаться к различным источникам данных, строить визуализации, собирать дашборды и делиться полученными результатами. За основу разработчики платформы взяли open source BI-систему Metabase. Glarus BI включена в Реестр программного обеспечения Минцифры (Запись в реестре от 30.01.2023 № 16417).

Glarus BI состоит из нескольких сущностей, которые обеспечивают полный цикл работы с данными (рис. 1.1):

- запрос (SQL-запрос);
- дашборд (dashboard, приборная панель) – набор визуализаций по запросам, фильтров, текстовых блоков, ссылок и других элементов;
- коллекция;
- модель.

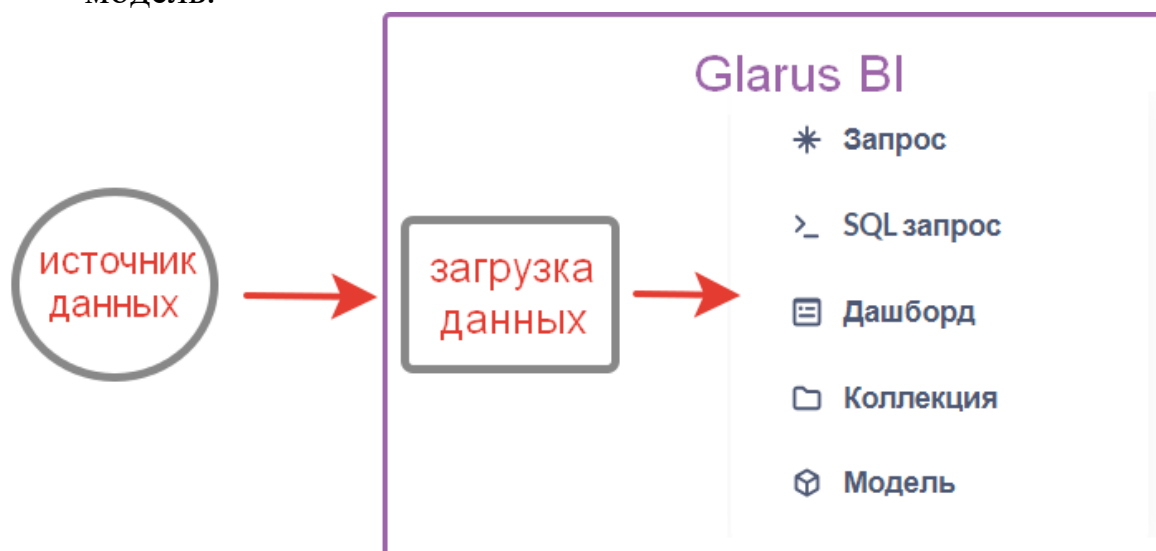


Рис. 1. Сущности Glarus BI

Коллекция – папка, в которой хранятся запросы, дашборды, модели. Модель – это набор данных или запрос, который можно снабдить метаданными (т.е. описанием отдельных показателей). В настоящем пособии модели рассматриваться не будут.

Платформу можно установить на хостинге, принадлежащем организации, или сразу развернуть облачную виртуальную машину с Glarus BI у таких облачных провайдеров, как VK cloud, MWS или T1 cloud, но в пособии будет рассматриваться только вариант работы в облаке <https://demo.glarus-analytics.ru>, доступ к которому можно получить автоматически, заполнив на сайте <https://glarus-bi.online> регистрационную форму, которая открывается щелчком на кнопке (рис. 2).

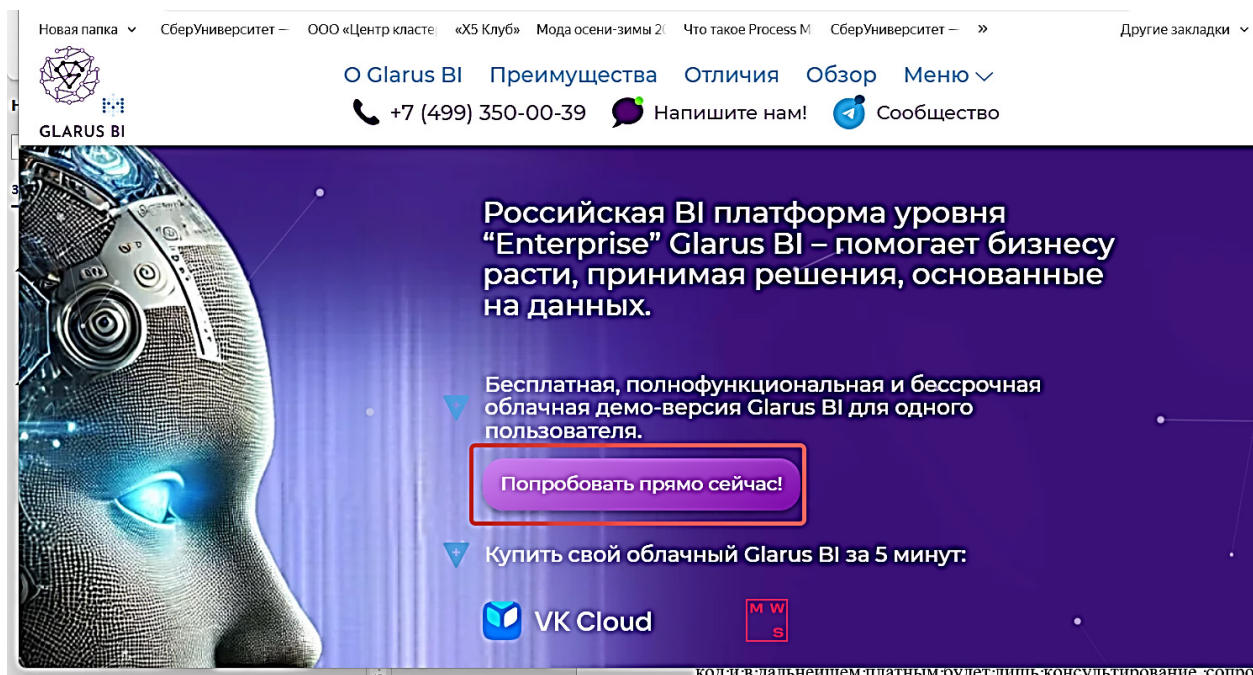


Рис. 2. Облачная демоверсия для одного пользователя

Пользовательский доступ для тестирования возможностей платформы по работе с данными предоставляется бесплатно. Роль пользователя не единственная из возможных ролей на платформе. Так, роль администратора позволяет подключать новых пользователей, осуществлять разграничение прав доступа пользователям и распределять ресурсы. Доступ к платформе в роли администратора возможен при коммерческом использовании платформы и по академической программе для учебных организаций.

В коммерческих целях выгоднее произвести установку на собственном хостинге, тем более что компания, разработавшая систему, передает исходный код и в дальнейшем платным будет лишь консультирование, сопровождение и обучение при запросе на эти услуги со стороны заказчика.

2. ИМПОРТ ДАННЫХ. ТИПЫ ДАННЫХ

Любой проект по анализу данных начинается с данных. У данных есть тип и структура.

В демоверсии Glarus BI требуется определенное расширение файла (.xlsx).

Наборы данных, которые будут использоваться:

– данные о транзакциях (https://www.kaggle.com/datasets/regivm/retailtransactiondata?select=Retail_Data_Transactions.csv);

– данные об отклике тех же клиентов на рекламные взаимодействия (https://www.kaggle.com/datasets/regivm/retailtransactiondata?select=Retail_Data_Response.csv).

Переведем эти наборы в нужный формат файла с помощью MS Excel. Сначала импортируем в созданный файл в MS Excel данные из Retail_Data_Transactions.csv (рис. 3), изменив представление для даты, если потребуется (рис. 4).

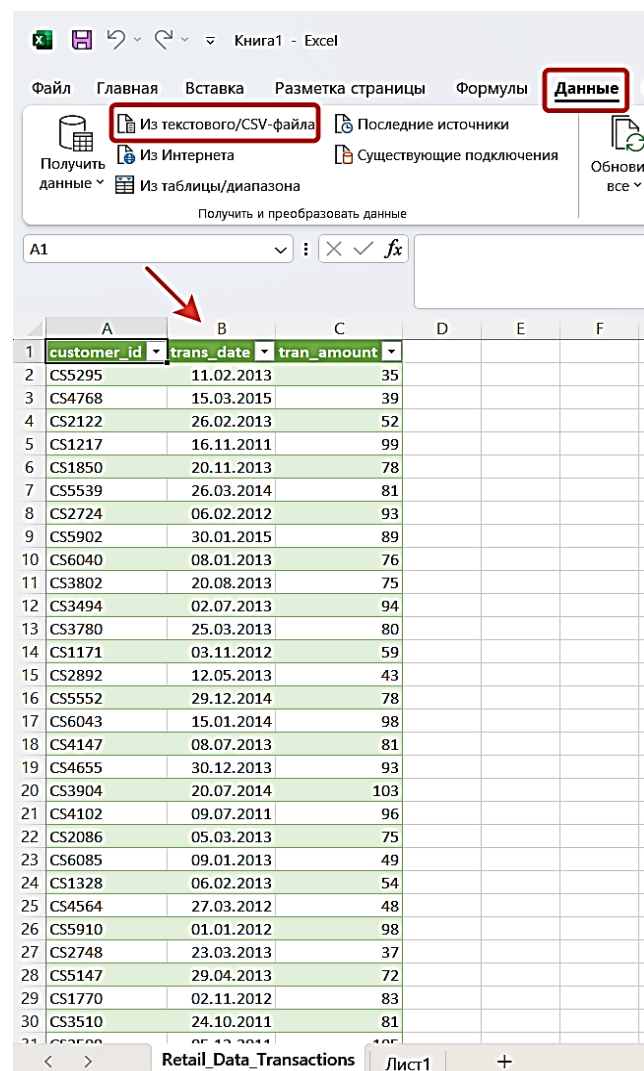


Рис. 3. Импорт данных из текстового файла

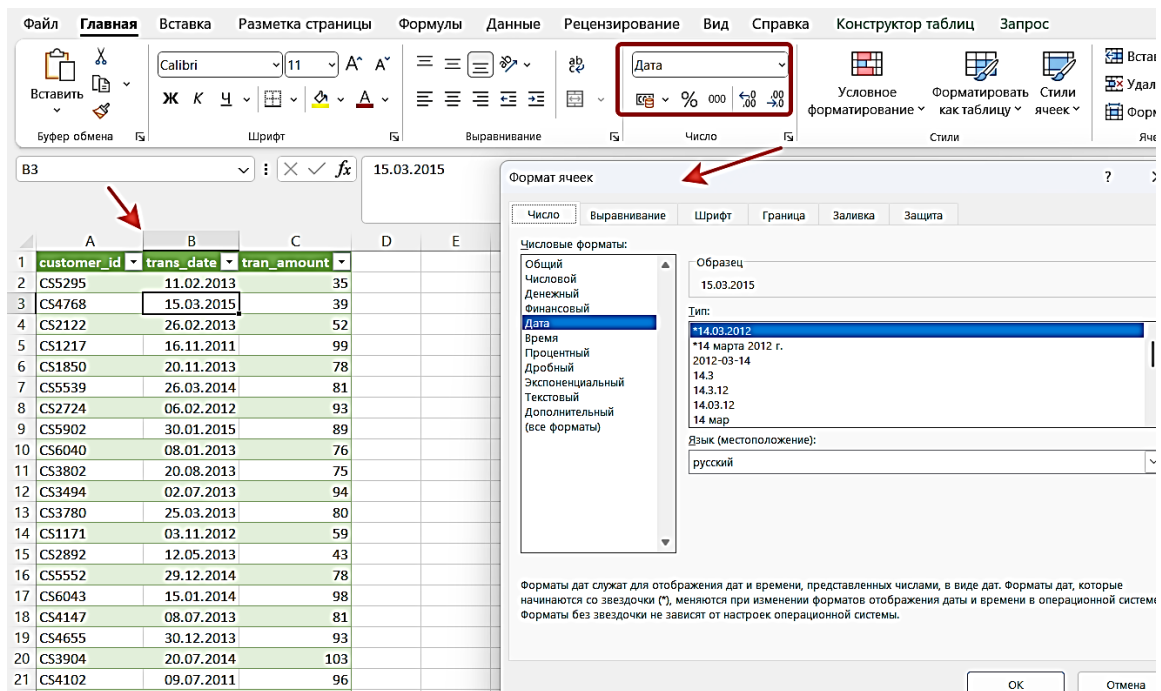


Рис. 4. Форматирование даты

Аналогично импортируем в тот же файл данные из Retail_Data_Response.csv и удалим все остальные листы (рис. 5).

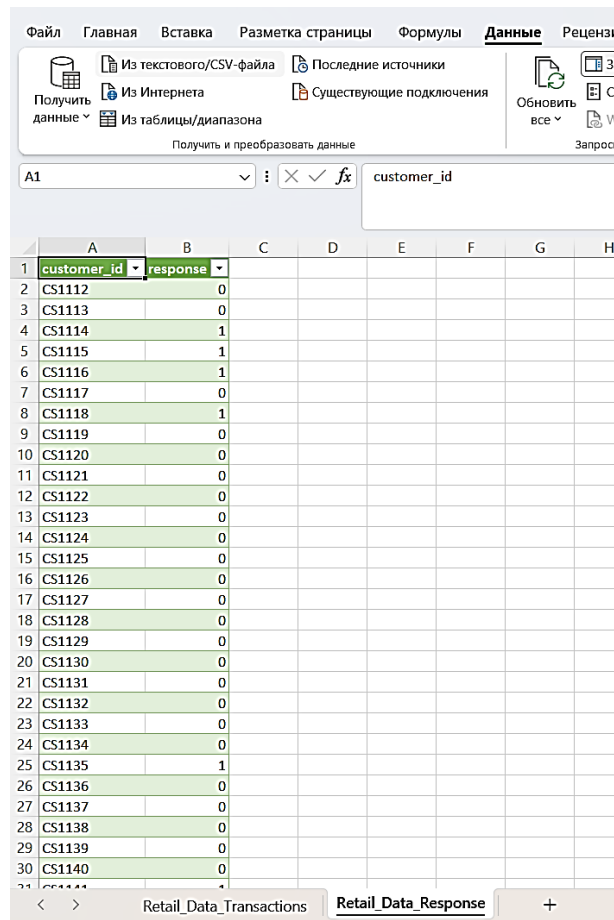


Рис. 5. Импорт данных из второго файла

Сохраним файл с именем retail.xlsx. И загрузим этот файл на платформу (рис. 6).

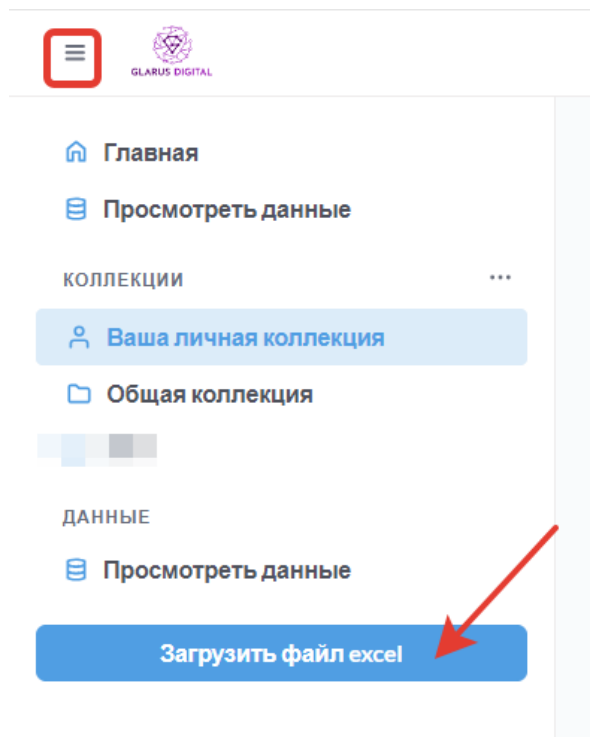


Рис. 6. Загрузка файлов на платформу

Загруженный файл будет размещен в СУБД ClickHouse_Test, к которой можно получить доступ, выбрав опцию Просмотреть данные (рис. 7).

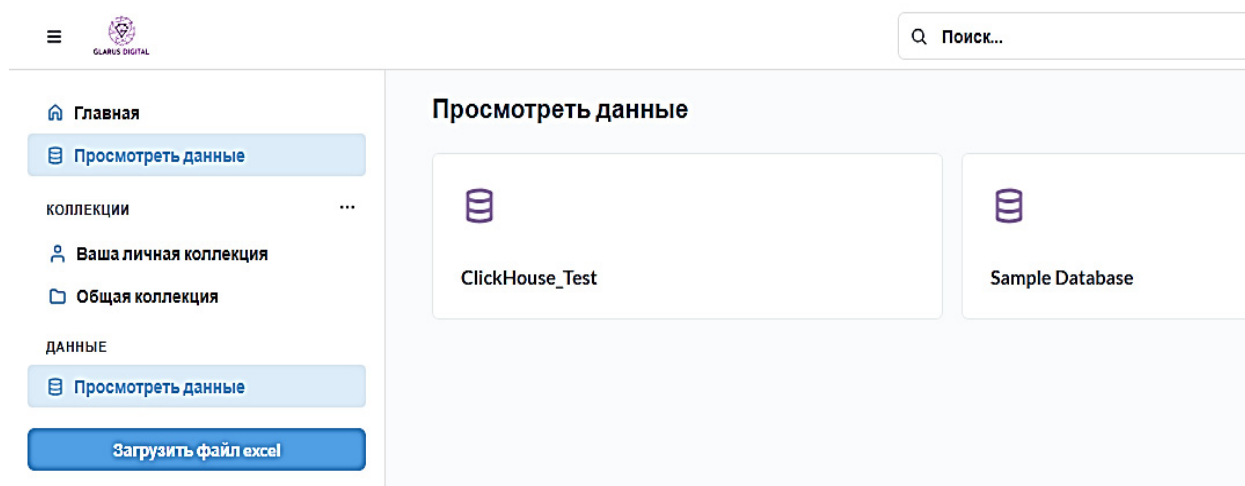


Рис. 7. Просмотр данных

Название таблицы с данными в базе данных составлено из имени файла retail и имени листа с данными Retail_Data_Response или Retail_Data_transactions (рис. 8).

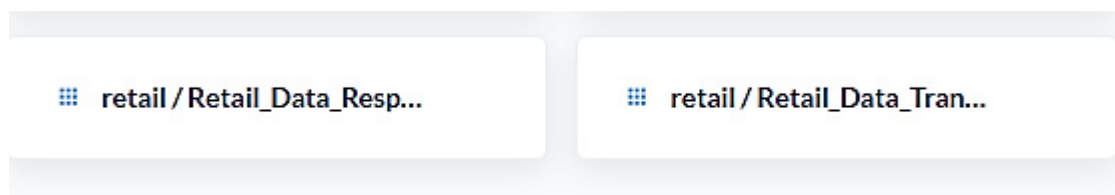


Рис. 8. Таблицы с данными в базе данных ClickHouse_Test

При наведении курсора на имя таблицы становится доступно пиктографическое меню, предоставляющее опции сканирования, просмотра сведений о таблице и удаления (рис. 9).

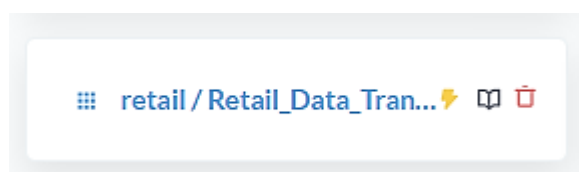


Рис. 9. Пиктографическое меню для действий с таблицей

Просмотрим сведения о таблице, проконтролируем корректность распознавания типов данных (рис. 10).

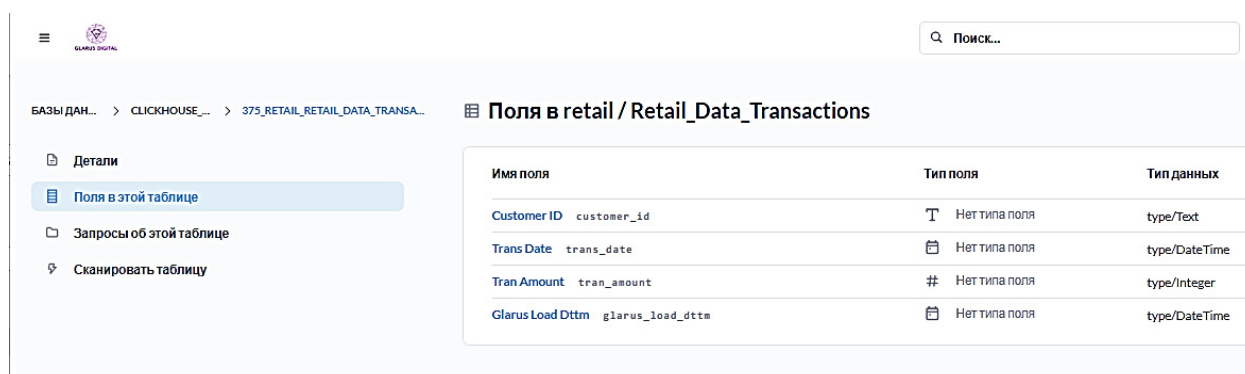


Рис. 10. Сведения о наборе данных Retail_Data_Transactions

Типы данных:

- числовые данные, имеют маркер #:
 - integer – целочисленные данные;
 - float – вещественные (десятичные дроби);
- категориальные данные, имеют маркер T:
 - text – строковые данные;
- данные даты и времени, имеют маркер 📅:
 - DateTime – временные данные.

Таблица Retail_Data_Transactions содержит следующие данные:

- customer_id – идентификатор клиента (текстовая/символьная переменная);
- trans_date – дата транзакции;
- tran_amount – сумма покупки.

Поле Glarus Load dtm с типом дата/время в таблицу добавлено платформой (это момент загрузки данных).

Таблица Retail_Data_Response содержит следующие данные:

- customer_id – идентификатор клиента (текстовая/символьная переменная);
- response – наличие/отсутствие отклика на рекламную кампанию (бинарная переменная отклика).

3. ПОСТРОЕНИЕ ЗАПРОСОВ

Построим визуализацию, с помощью которой можно сравнить ежемесячные продажи.

Создадим запрос на основе данных, воспользовавшись кнопкой добавления новой сущности (рис. 11).

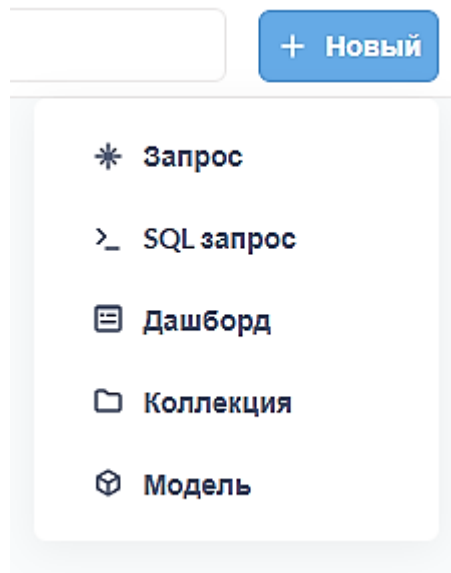


Рис. 11. Создание запроса

Выберем начальные данные Retail_Data_Transactions (рис. 12).

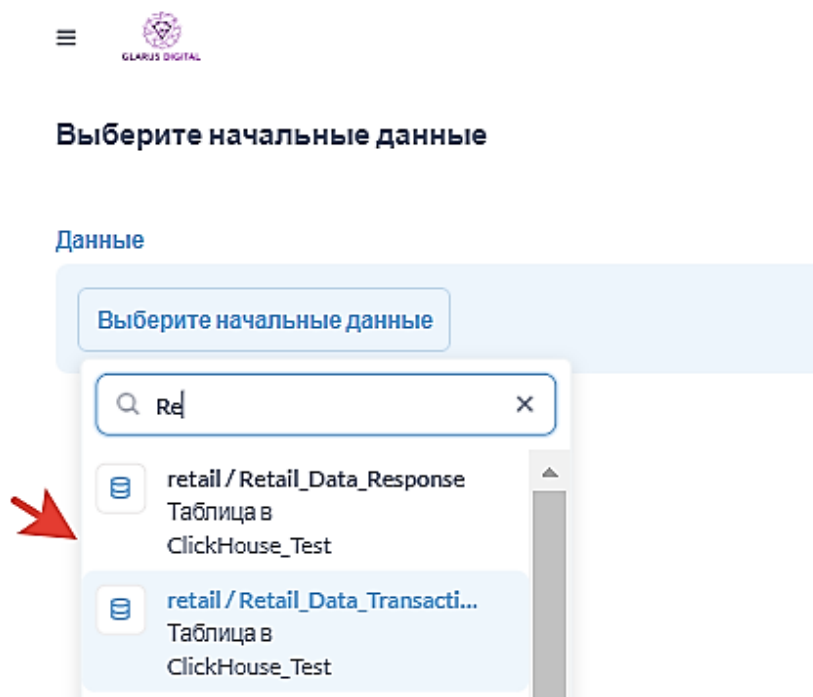


Рис. 12. Выбор начальных данных

Их можно просмотреть с помощью стрелки справа в поле имени набора (рис. 13), она же появится, если добавить условие фильтрации или настройки для агрегации (суммирования). К данным можно присоединить дополнительный набор (кнопка ).

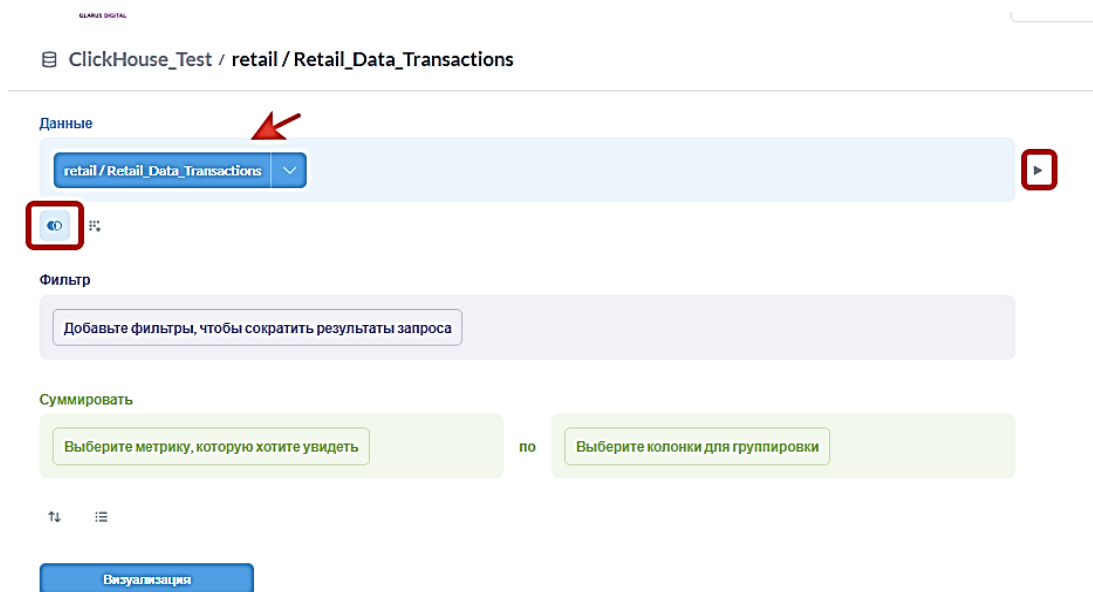


Рис. 13. Редактор запросов

Выберем метрику (сумма) и показатель (tran_amount) для агрегации (рис. 14, 15), а также группирующий показатель (trans_date) с настройкой Месяц года (рис. 16).

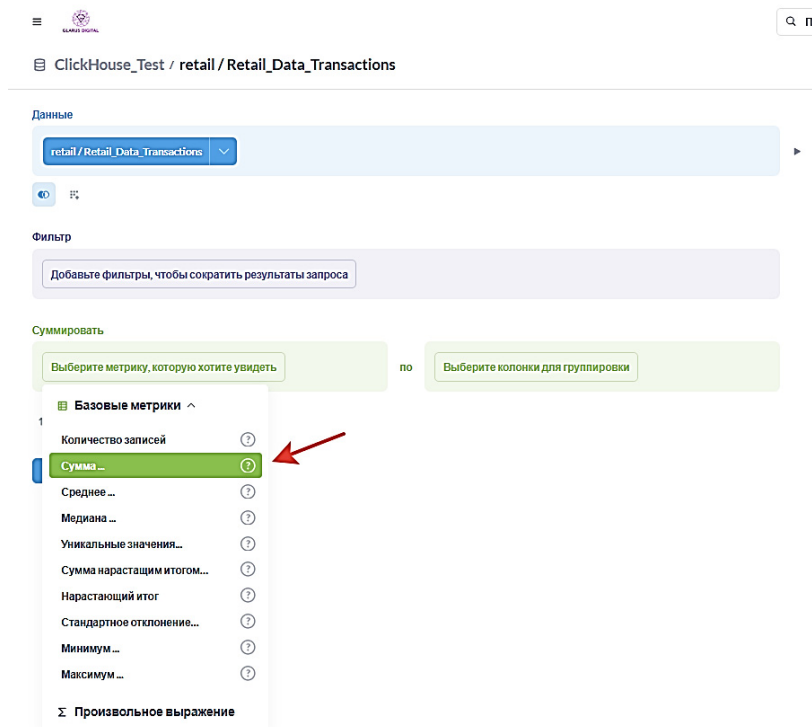


Рис. 14. Выбор метрики

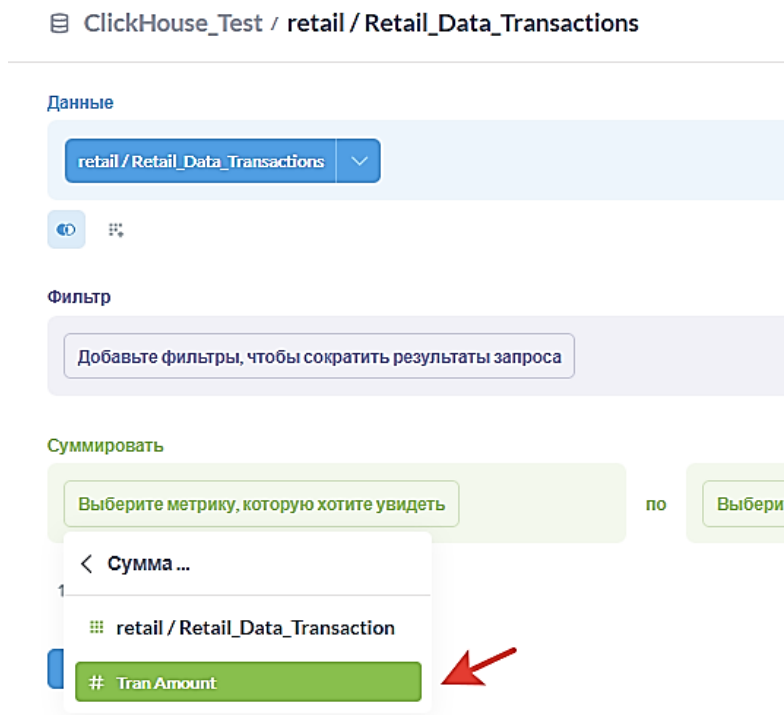


Рис. 15. Выбор показателя для агрегации

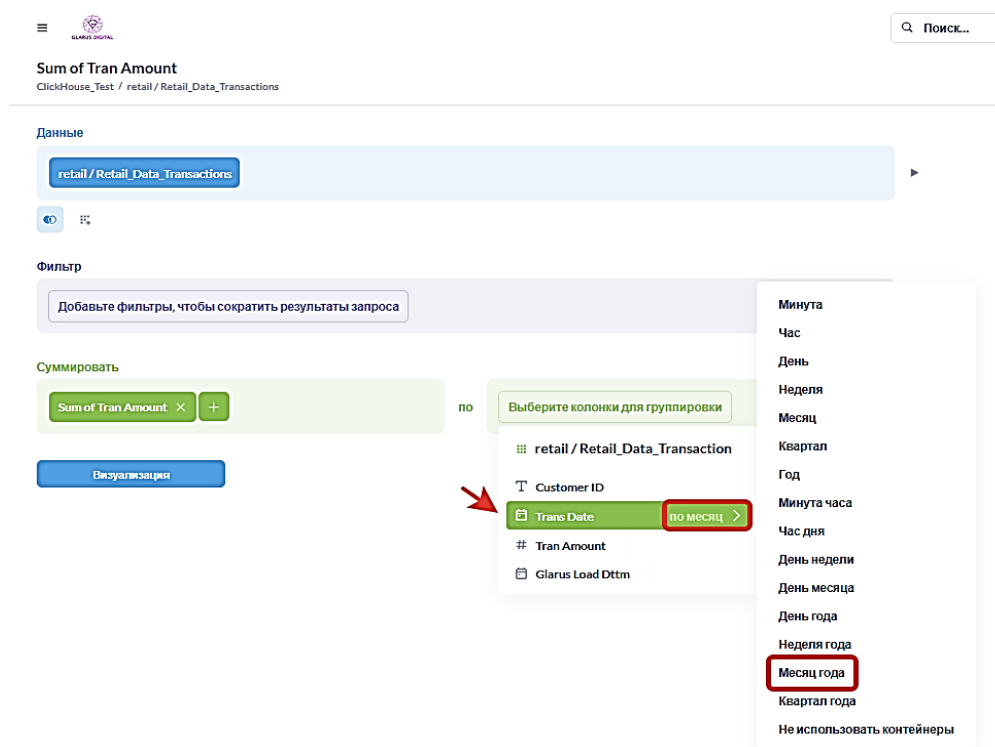


Рис. 16. Выбор показателя для группировки

Построим визуализацию для запроса, нажав на соответствующую кнопку слева внизу (рис. 17). В визуализации можно переключаться между представлением Таблица и Диаграмма. А для показателя Trans Data, который использовался для группировки, возможно изменить настройки

непосредственно в диаграмме. Переключимся с варианта Месяц года на Месяц (рис. 18). Тип подходящей визуализации стал иным, количество строк данных в запросе также изменилось (с 12 на 47).

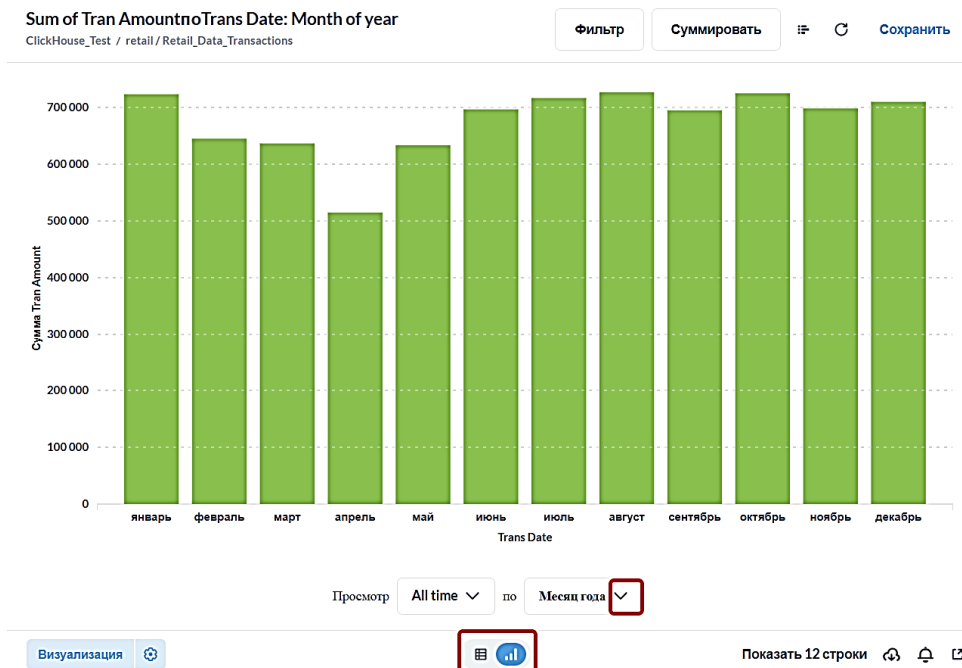


Рис. 17. Визуализация

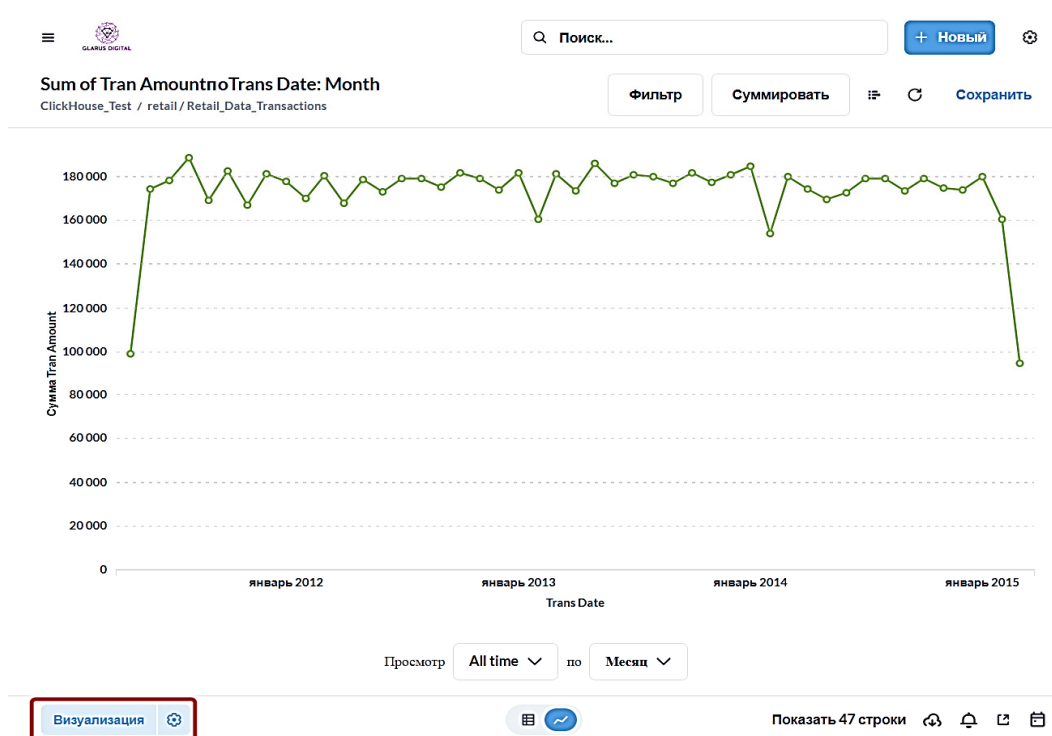


Рис. 18. Изменение запроса/визуализации при смене гранулярности данных

С помощью кнопок **Визуализация** и  можно изменить тип визуализации (рис. 19) и провести форматирование визуализации (рис. 20).

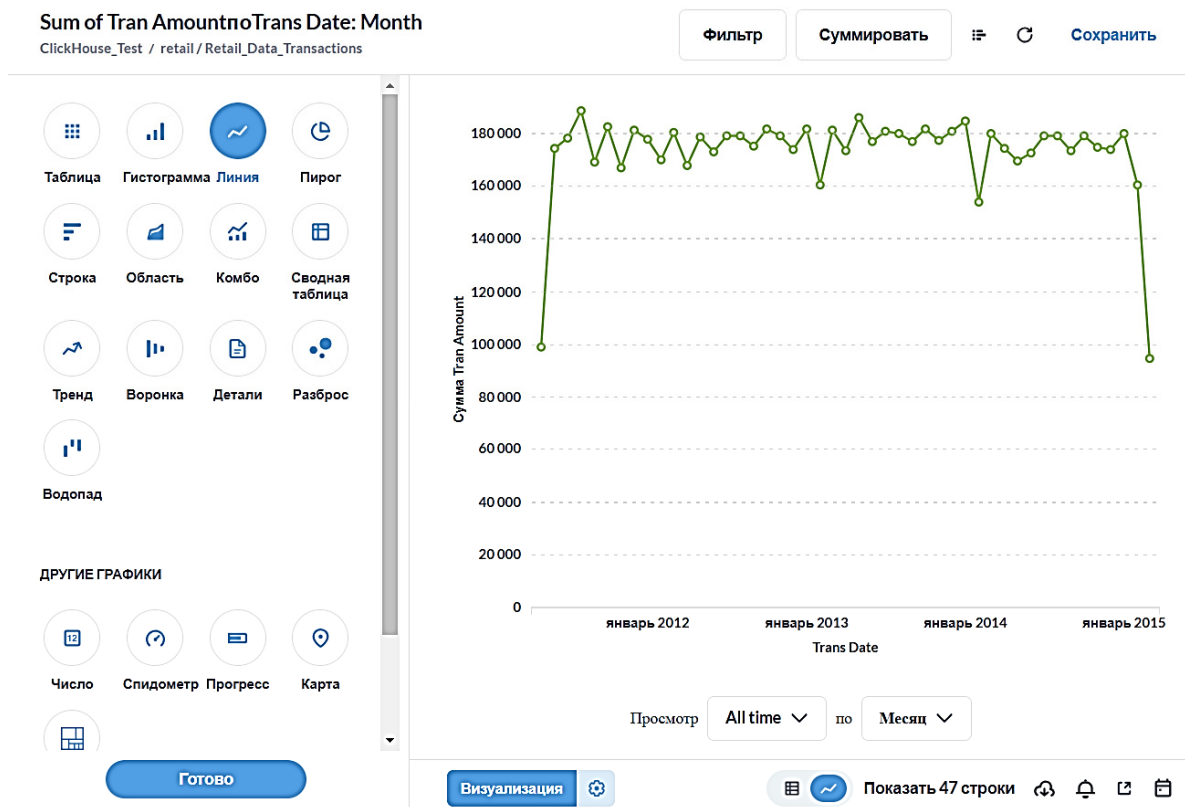


Рис. 19. Выбор типа визуализации

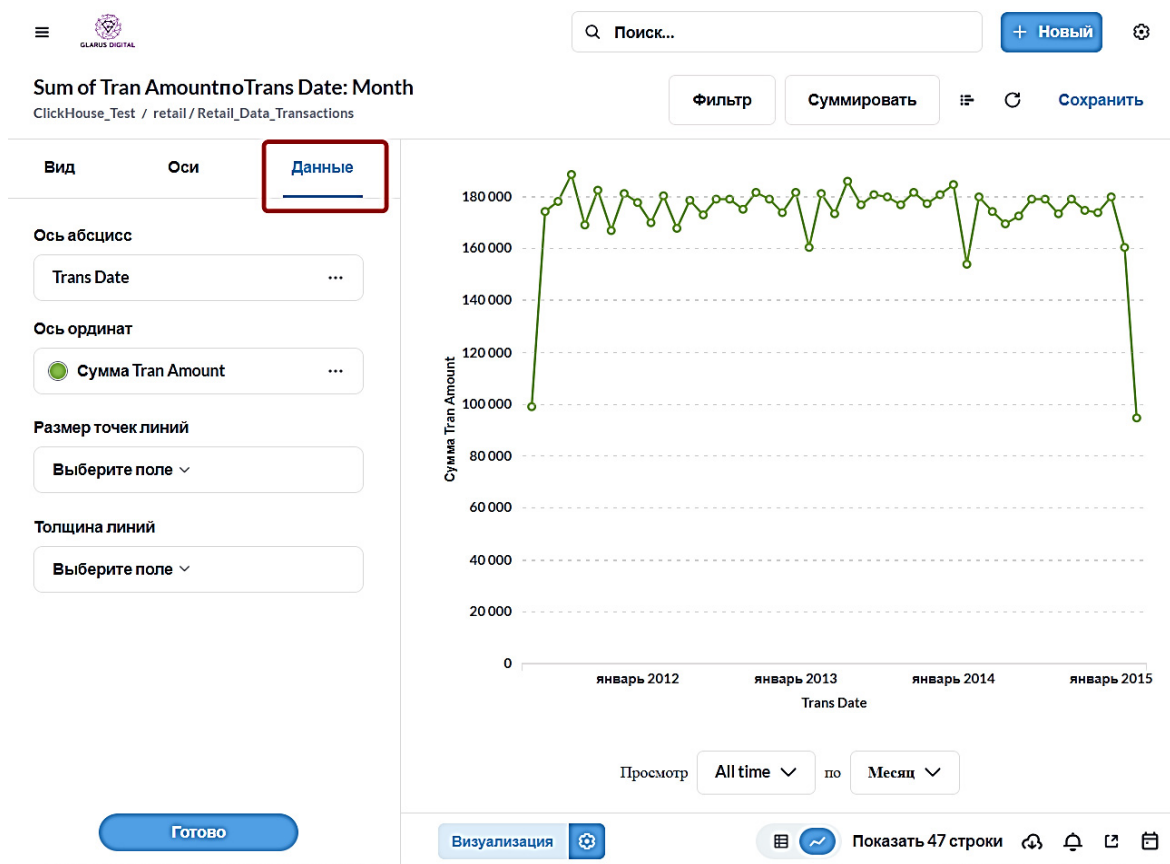


Рис. 20. Форматирование визуализации

Сохраним запрос (рис. 21).

The dialog box is titled "Сохранить новый запрос" (Save new request) and contains the following fields:

- Имя** (Name): A text input field containing "retail / Retail_Data_Transactions Продажи клиентам по месяцам".
- Описание** (Description): A text area containing the placeholder text "Это не обязательно, но может быть полезным".
- В какую коллекцию следует это включить?** (Which collection should this be included in?): A dropdown menu showing "Светлана личная коллекция Рындина".
- Buttons:** "Отмена" (Cancel) and "Сохранить" (Save).

Рис. 21. Сохранение запроса

Воспользуемся дополнительными опциями меню (рис. 22) и продублируем запрос с новым именем (рис. 23).

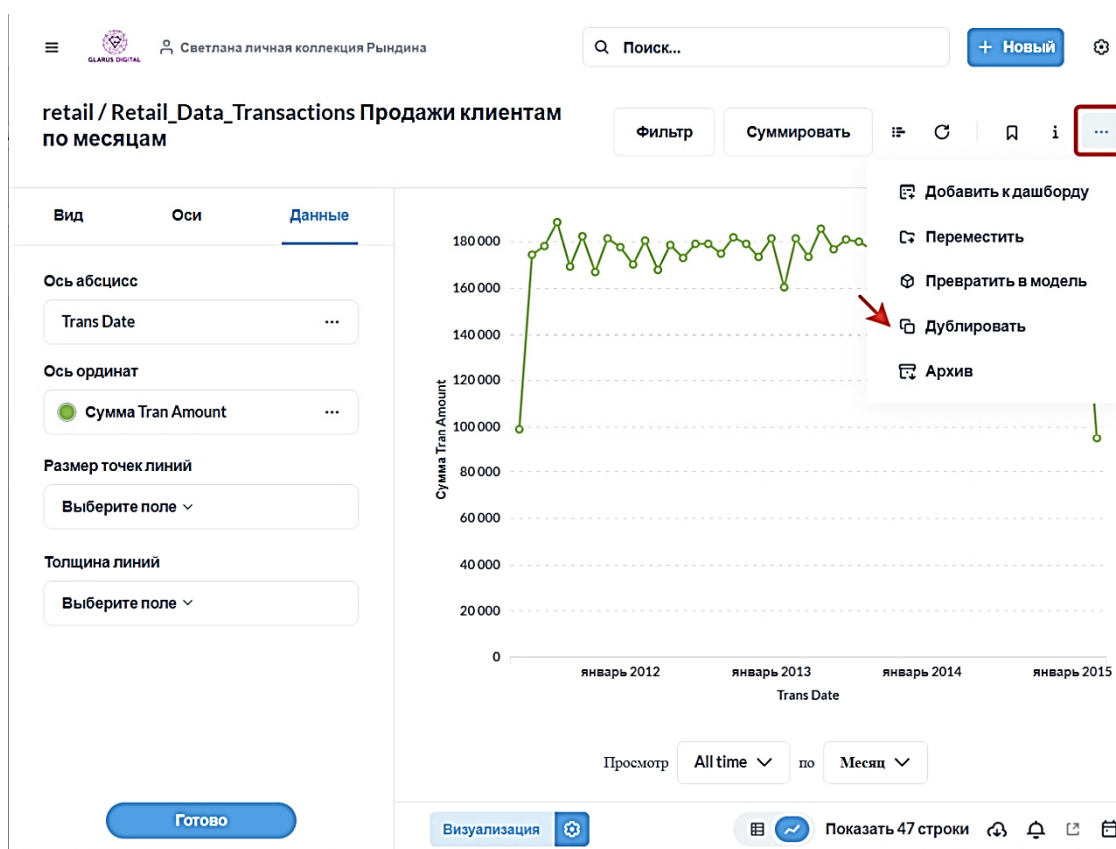


Рис. 22. Дублирование запроса

Построим запрос по клиентам, которым реализовано товаров на самую большую сумму. Создадим дубль открытого запроса Продажи клиентам по дням недели и сохраним его под именем Самые активные клиенты (рис. 25).

Дублировать "retail / Retail_Data_Transactions Продажи клиентам по дням недели"

Имя

retail / Retail_Data_Transactions Самые активные клиенты

Описание

Это не обязательно, но может быть полезным

В какую коллекцию следует это включить?

Светлана личная коллекция Рындина

Дублировать

Рис. 25. Создание копии запроса

Перейдем в редактор запросов по кнопке  и изменим показатель группировки на customer_id (рис. 26).

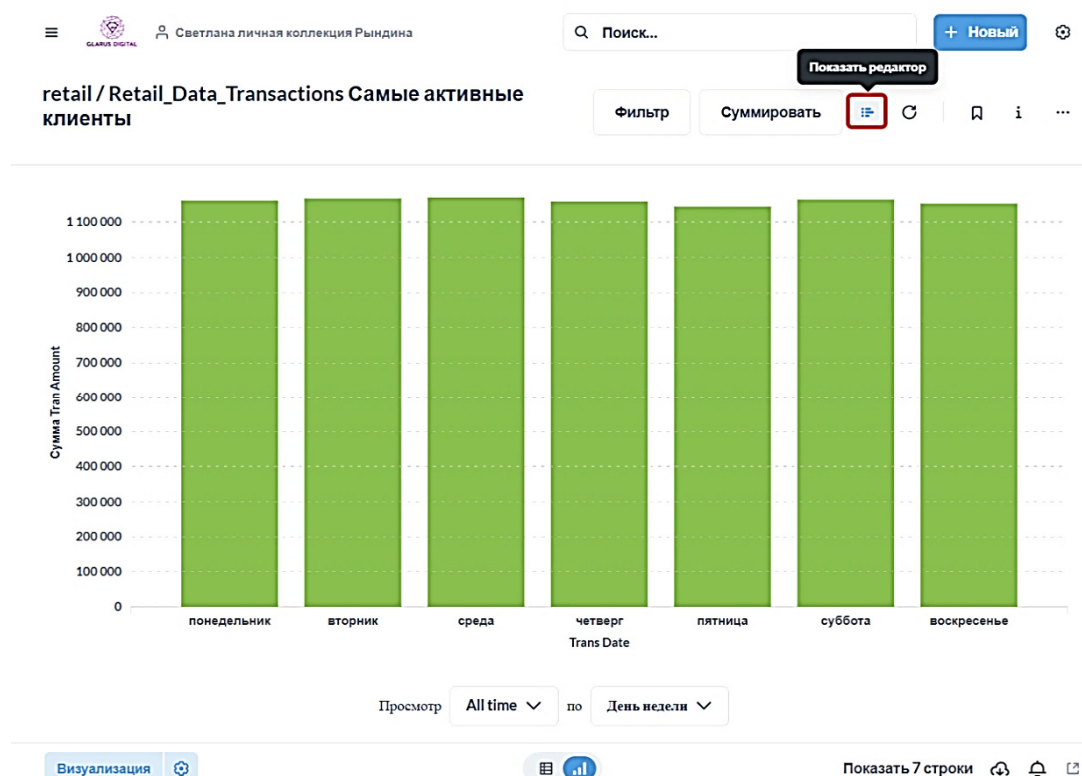


Рис. 26. Переход в редактор запросов

Добавим сортировку (по убыванию) по показателю Sum of tran_amount, изменив ее направление щелчком на стрелке рядом с именем показателя сортировки (рис. 27).

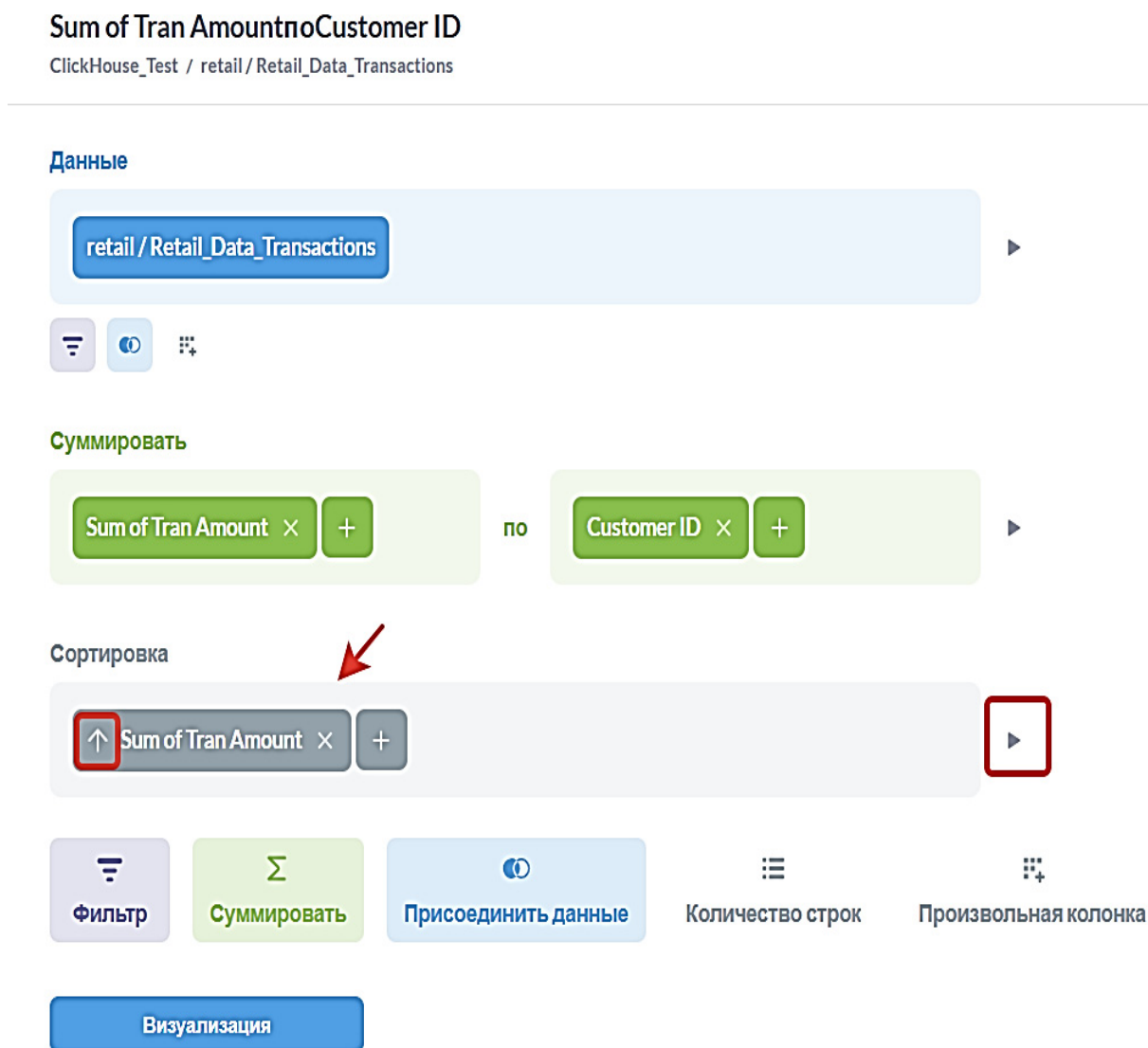
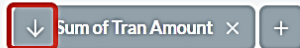


Рис. 27. Добавление сортировки к запросу

Раскроем таблицу отсортированного запроса (рис. 28) и выберем значение для границы активности покупок, пусть это будет сумма 2300. Добавим фильтр для показателя Sum of tran_amount с этим значением (рис. 29).

Сохраним настроенный запрос.

Сортировка



Предварительный просмотр

Customer ID	Сумма Tran Amount
CS4424	2 933
CS4320	2 647
CS5752	2 612
CS4660	2 527
CS3799	2 513
CS5109	2 506
CS4074	2 462
CS3805	2 453
CS4608	2 449
CS5555	2 439

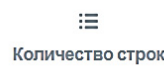


Рис. 28. Просмотр результатов сортировки

Фильтр

Добавьте фильтры, чтобы сократить результаты запроса

< Sum of Tran A...

Больше или равно

▼

Добавить фильтр

Калькулятор

Рис. 29. Добавление фильтра

Построим запрос, в котором будут присутствовать клиенты, совершившие наибольшее число транзакций за период. Для этого создадим новый запрос, добавим к нему исходные данные. Метрикой для агрегации выберем количество — эта та метрика, для которой не нужен собственный показатель агрегации, в роли него будет выступать показатель для группировки `customer_id`. Также добавим в настройки запроса сортировку по убыванию для агрегирующего показателя Количество (рис. 30).

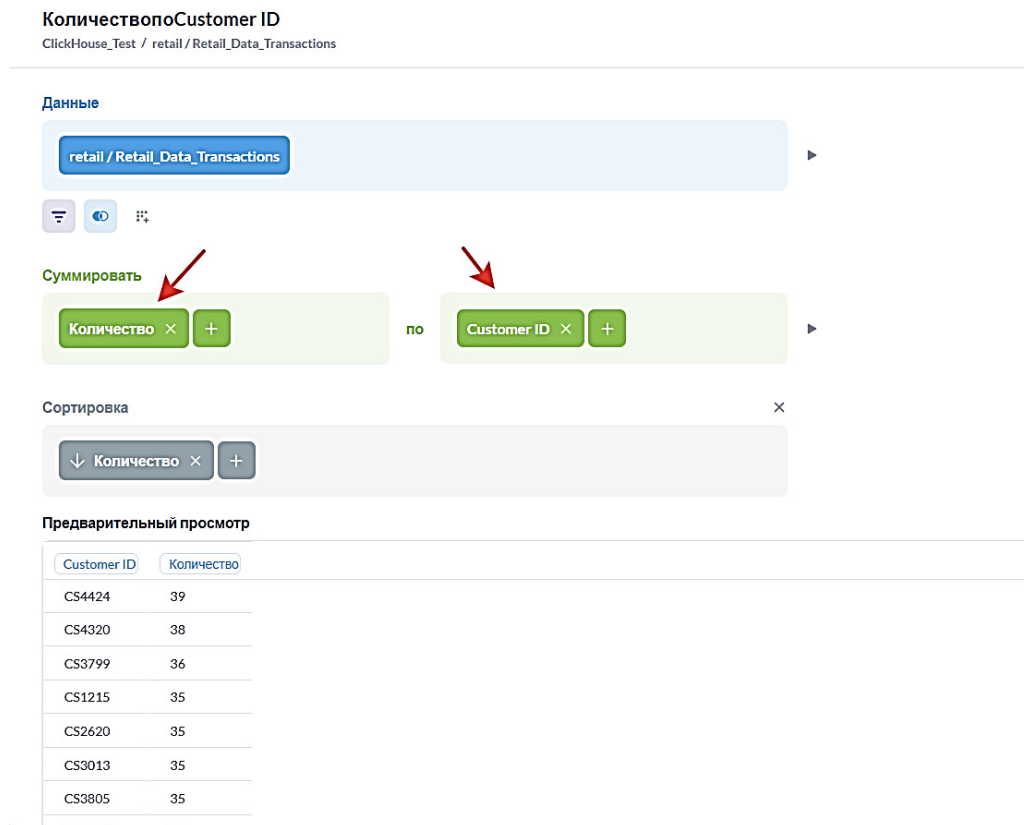


Рис. 30. Настройки для запроса

Как видим, для фильтра по показателю Количество вполне можно выбрать значение 34 (рис. 31). Сохраним запрос (рис. 32).

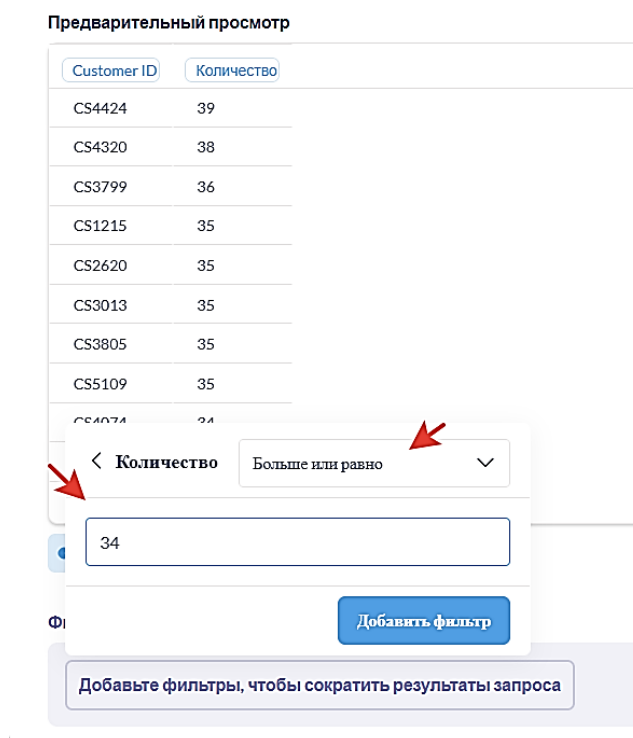


Рис. 31. Добавление фильтра

Сохранить новый запрос

×

Имя

retail / Retail Data Transactions Клиенты, часто совершающие покупки

Описание

Это не обязательно, но может быть полезным

В какую коллекцию следует это включить?

Светлана личная коллекция Рындина

Отмена

Сохранить

Рис. 32. Сохранение запроса

Размещением визуализаций на дашборде займемся чуть позже, а пока научимся объединять данные.

4. ОБЪЕДИНЕНИЕ ДАННЫХ

В процессе анализа данных может возникнуть необходимость объединять результаты запросов или обогащать данные, соединяя массивы данных из различных источников.

Создадим запрос, который сравнивает общую сумму продаж клиентам, откликнувшимся на рекламу, с общей суммой продаж тем, кто рекламу проигнорировал.

Для этого к данным первой таблицы `Retail_Data_Transactions` присоединим данные второй таблицы `Retail_Data_Response`, изменив тип соединения, кликнув по значку соединения и выбрав внутреннее (рис. 33), и выберем поля этих таблиц (`Customer ID`), по которым объединение произойдет. Важно, чтобы поля, по которым происходит соединение таблиц, имели один и тот же тип данных и содержали одинаковые значения. Если рядом с таблицей запроса появляется стрелка, раскрывающая список полей, то в ней можно настроить поля для отображения (рис. 34).

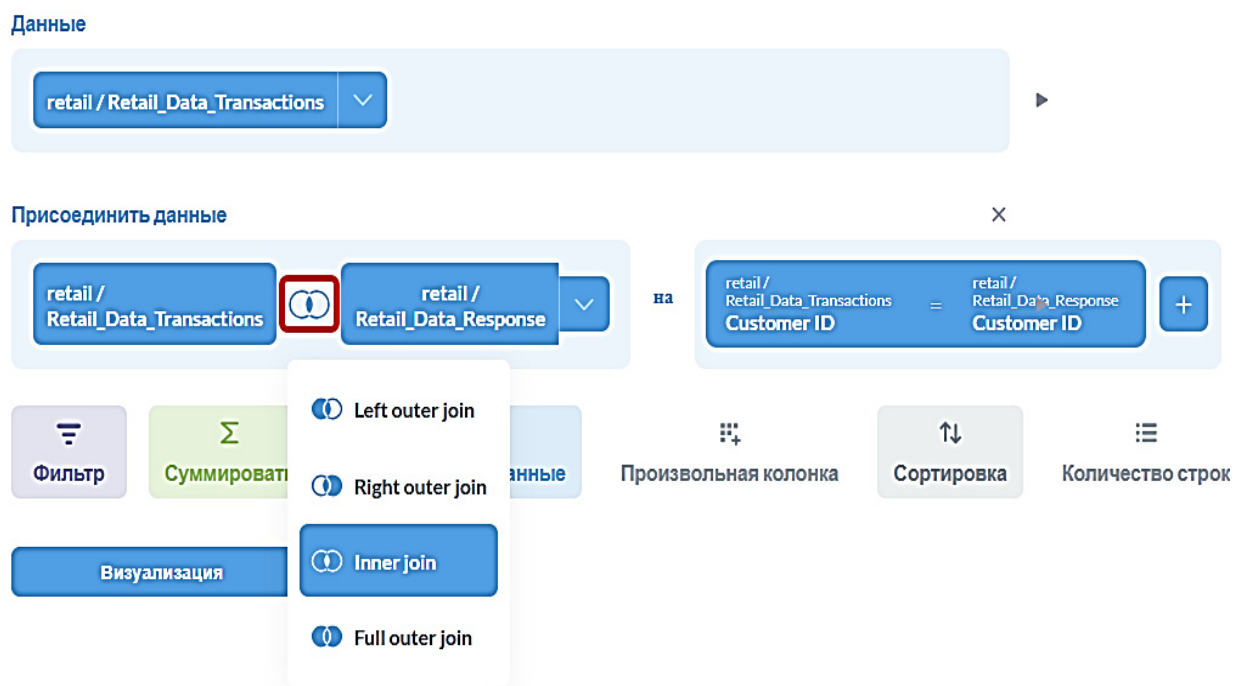


Рис. 33. Изменение типа соединения

Агрегацию проведем по полю `tran_amount` первой таблицы `Retail_Data_Transactions`, а группировать будем по переменной `response` второй таблицы `Retail_Data_Response` (рис. 35).

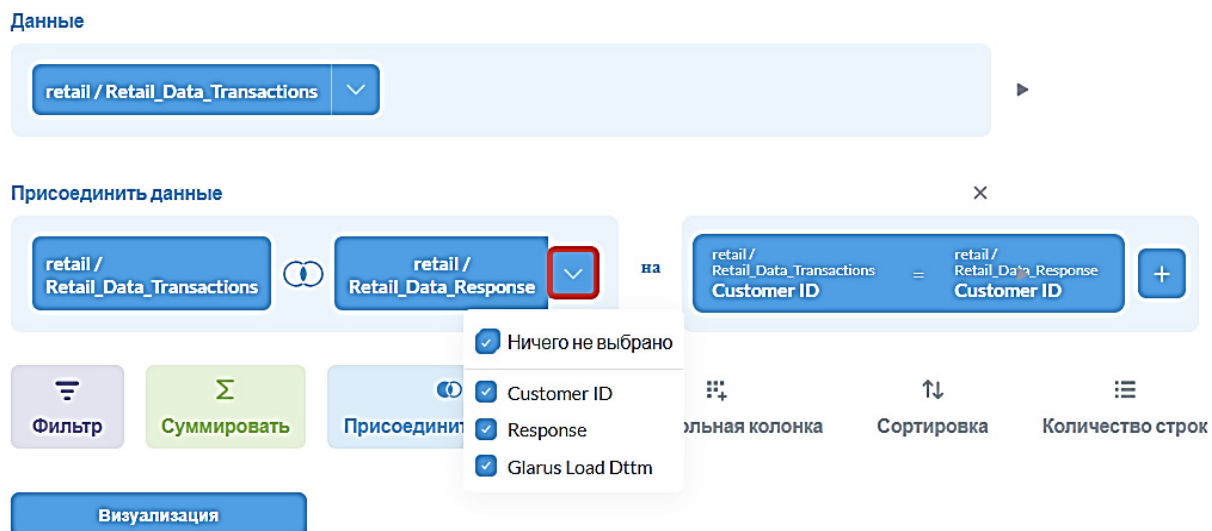


Рис. 34. Выбор полей

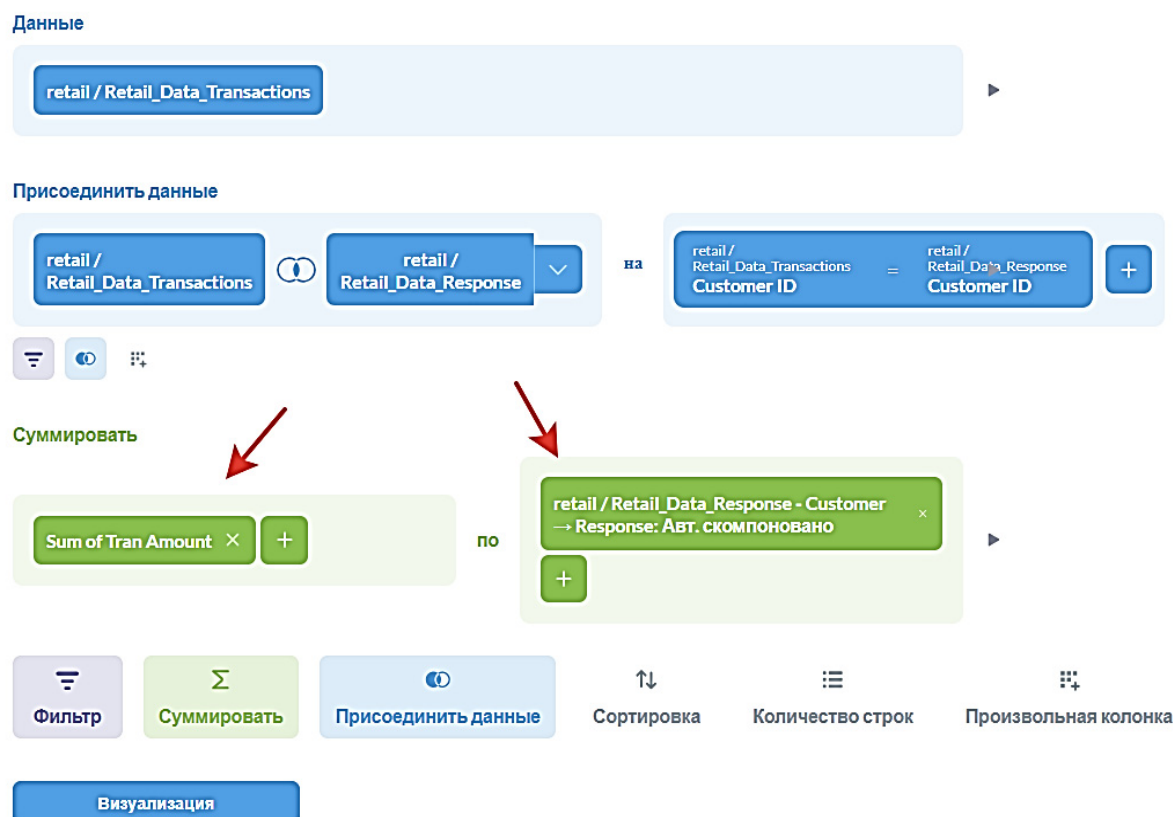


Рис. 35. Настройки для запроса по данным двух объединенных таблиц

Построим визуализацию (гистограмма) и сохраним запрос (рис. 36).

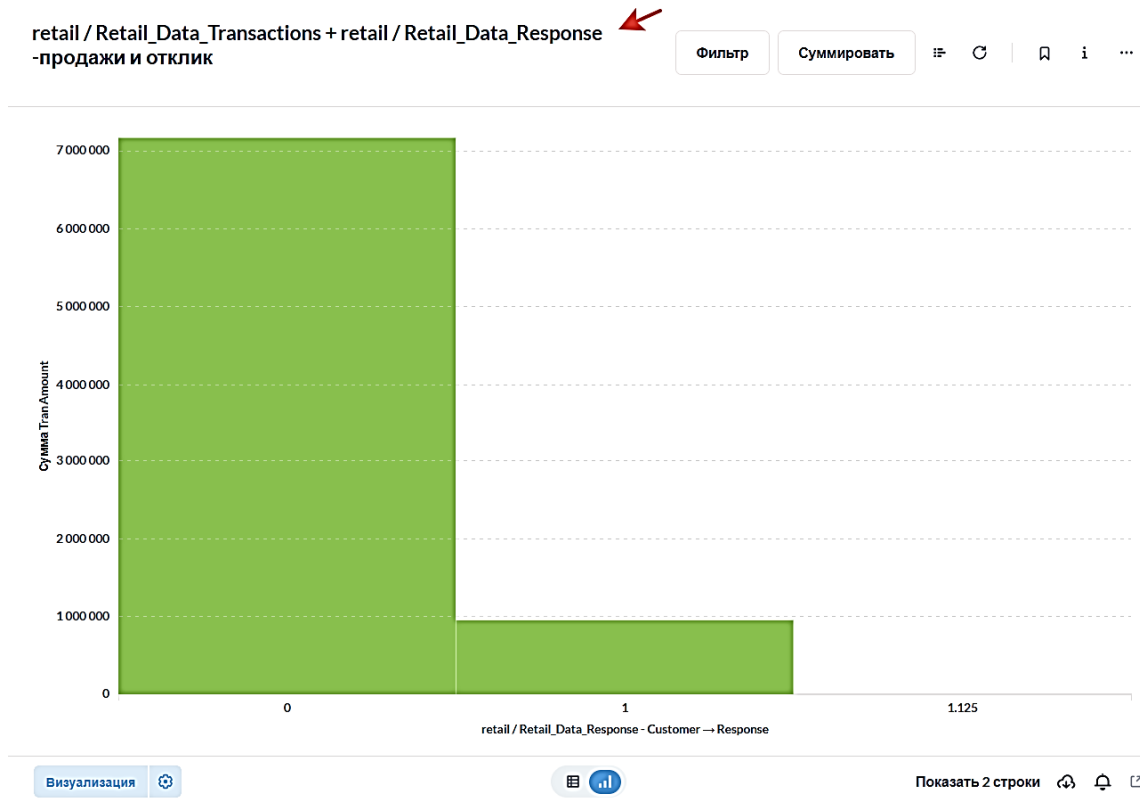


Рис. 36. Сохранение запроса

В новом запросе объединим данные двух ранее построенных запросов: по активным клиентам (с самой большой общей суммой покупок) и по клиентам, совершившим самое большое число транзакций (рис. 37). Поле `customer_id` удалим из числа отображаемых полей.

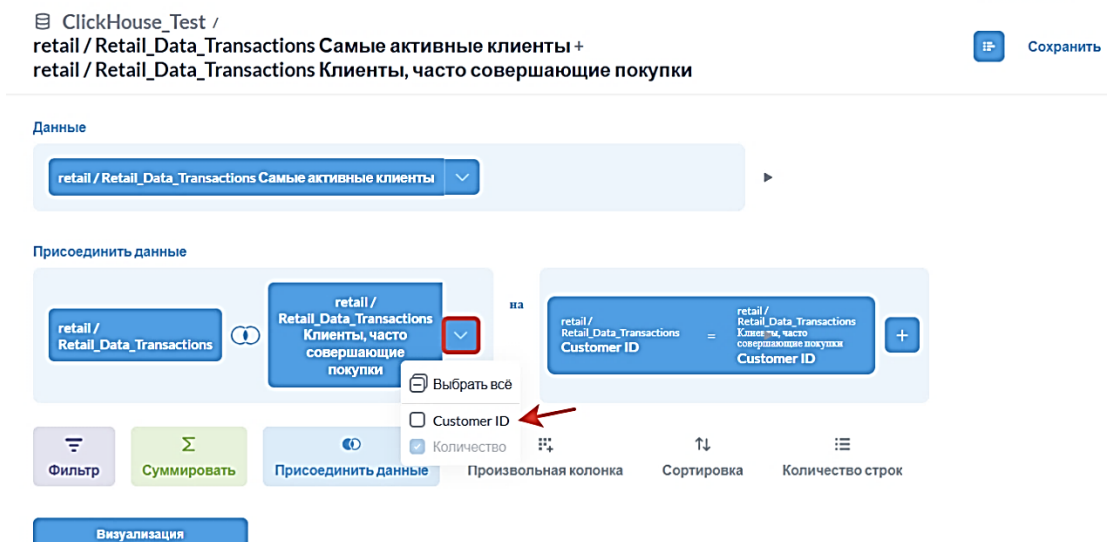


Рис. 37. Объединение данных двух запросов

Построим визуализацию: автоматически создается визуализация Таблица с отображением объединенных данных (рис. 38).

The screenshot shows the ClickHouse Test web interface. At the top, there is a search bar with the text "Поиск...". Below it, the breadcrumb path is "ClickHouse_Test / retail / Retail_Data_Transactions Самые активные клиенты + retail / Retail_Data_Transactions Клиенты, часто совершающие покупки". A "Фильтр" button is located to the right of the breadcrumb. The table has three columns: "Customer ID", "Сумма Tran Amount", and "retail / Retail_Data_Transactions Клиенты, _dd3670ac → Count". The table contains 10 rows of data. At the bottom, there is a "Визуализация" button and a settings icon.

Customer ID	Сумма Tran Amount	retail / Retail_Data_Transactions Клиенты, _dd3670ac → Count
CS4424	2 933	39
CS4320	2 647	38
CS3799	2 513	36
CS5109	2 506	35
CS4074	2 462	34
CS3805	2 453	35
CS1215	2 362	35
CS2620	2 360	35
CS4998	2 358	34

Рис. 38. Визуализация для объединенных запросов

Перейдем в настройки визуализации. На вкладке Столбцы есть возможность управлять отображением значений полей в таблице (рис. 39). Сохраним новый запрос (рис. 40).

ClickHouse_Test / retail / Retail_Data_Transactions Самые активные клиенты + retail / Retail_Data_Transactions Клиенты, часто совершающие покупки

Фильтр Суммировать Сохранить

Столбцы Ячейки Цвета Вид Условное форматирование

Сводная таблица ☐

Добавить или удалить столбцы

- T Customer ID ...
- # Сумма Tran Amount ...
- * retail / Retail_Data_Transactions Клиенты, dd3670ac → Count ...

Customer ID	Сумма Tran Amount	retail / Retail_Data_Transactions Клиенты, dd3670ac → Count
CS4424	2 933	39
CS4320	2 647	38
CS3799	2 513	36
CS5109	2 506	35
CS4074	2 462	34
CS3805	2 453	35
CS1215	2 362	35
CS2620	2 360	35
CS4998	2 358	34

Готово Визуализация

Показать 9 строк

Рис. 39. Управление отображением показателей в таблице

Сохранить новый запрос

Имя

a Transactions Самые активные клиенты, часто совершающие покупки

Описание

Это не обязательно, но может быть полезным

В какую коллекцию следует это включить?

Светлана личная коллекция Рындина

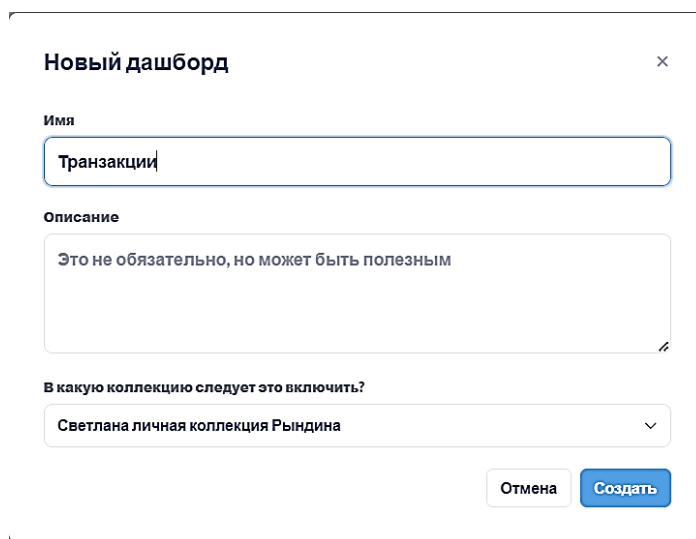
Отмена Сохранить

Рис. 40. Сохранение запроса

Перейдем к созданию информационных панелей (дашбордов) по построенным запросам.

5. ПОСТРОЕНИЕ ДАШБОРДОВ

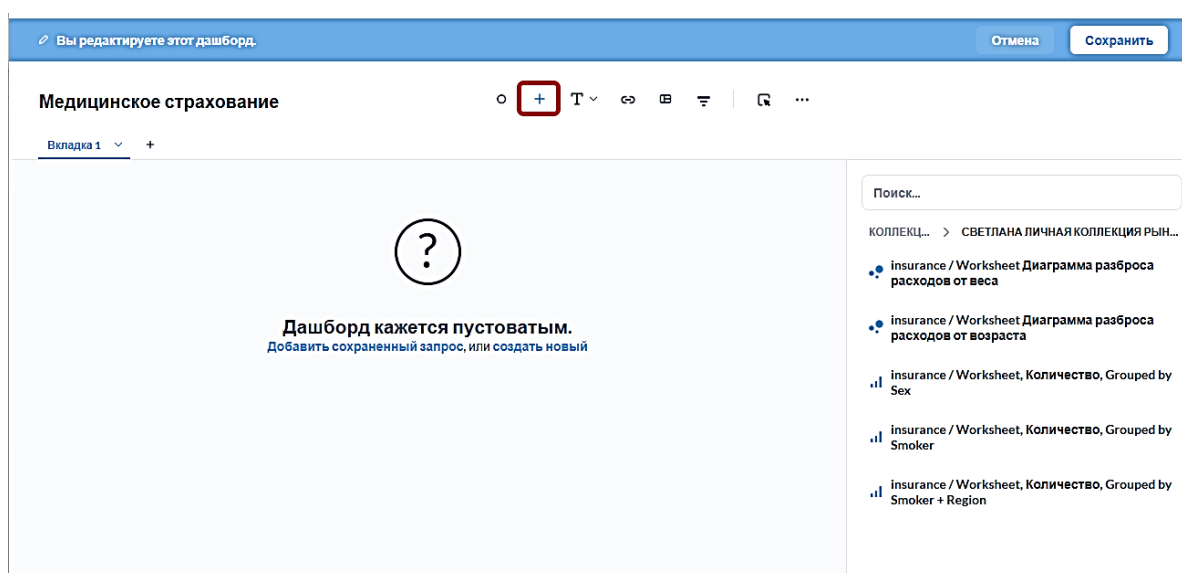
Создадим дашборд с именем Транзакции (рис. 41). Воспользуемся кнопкой в правом верхнем углу (см. рис. 11).



Диалоговое окно «Новый дашборд» с заголовком и кнопкой закрытия (X). Оно содержит три поля: текстовое поле «Имя» со значением «Транзакции», текстовое поле «Описание» со подсказкой «Это не обязательно, но может быть полезным», и выпадающий список «В какую коллекцию следует это включить?» со значением «Светлана личная коллекция Рындина». В нижнем правом углу расположены кнопки «Отмена» и «Создать».

Рис. 41. Создание дашборда

Справа вверху в режиме редактирования дашборда имеется пиктографическое меню. Кнопка «+» открывает справа панель с запросами, которые можно добавить на дашборд (рис. 42). Чтобы ограничить число показываемых запросов, можно набрать ключевое слово (например, общее для имен запросов или часть названия таблицы с данными, по которой строился запрос).



Интерфейс редактирования дашборда. Вверху — заголовок «Медицинское страхование» и панель инструментов с кнопкой «+» (выделена красным квадратом). В центре — сообщение «Дашборд кажется пустоватым. Добавить сохраненный запрос, или создать новый». Справа — панель «Поиск...» с фильтрами и списком запросов:

- КОЛЛЕКЦИ... > СВЕТЛАНА ЛИЧНАЯ КОЛЛЕКЦИЯ РЫН...
- insurance / Worksheet Диаграмма разброса расходов от веса
- insurance / Worksheet Диаграмма разброса расходов от возраста
- insurance / Worksheet, Количество, Grouped by Sex
- insurance / Worksheet, Количество, Grouped by Smoker
- insurance / Worksheet, Количество, Grouped by Smoker + Region

Рис. 42. Панель с перечнем созданных запросов

Также на дашборд можно добавить текст или заголовок (рис. 43), чтобы сделать его более информативным. Дашборд в настоящий момент

находится в режиме редактирования, а после сохранения перейдет в режим просмотра.

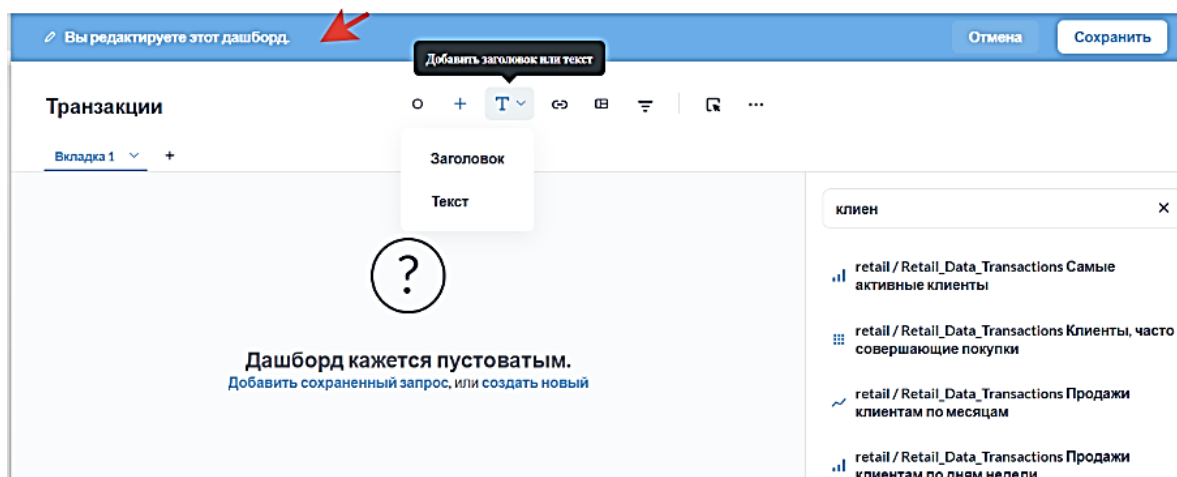


Рис. 43. Добавление описания и заголовков на дашборд

Разместим построенные визуализации на дашборде. Отформатируем представление визуализаций. В активной диаграмме справа сверху появляется пиктографическое меню, значок  которого позволяет изменить форматирование визуальных элементов диаграммы (рис. 44).

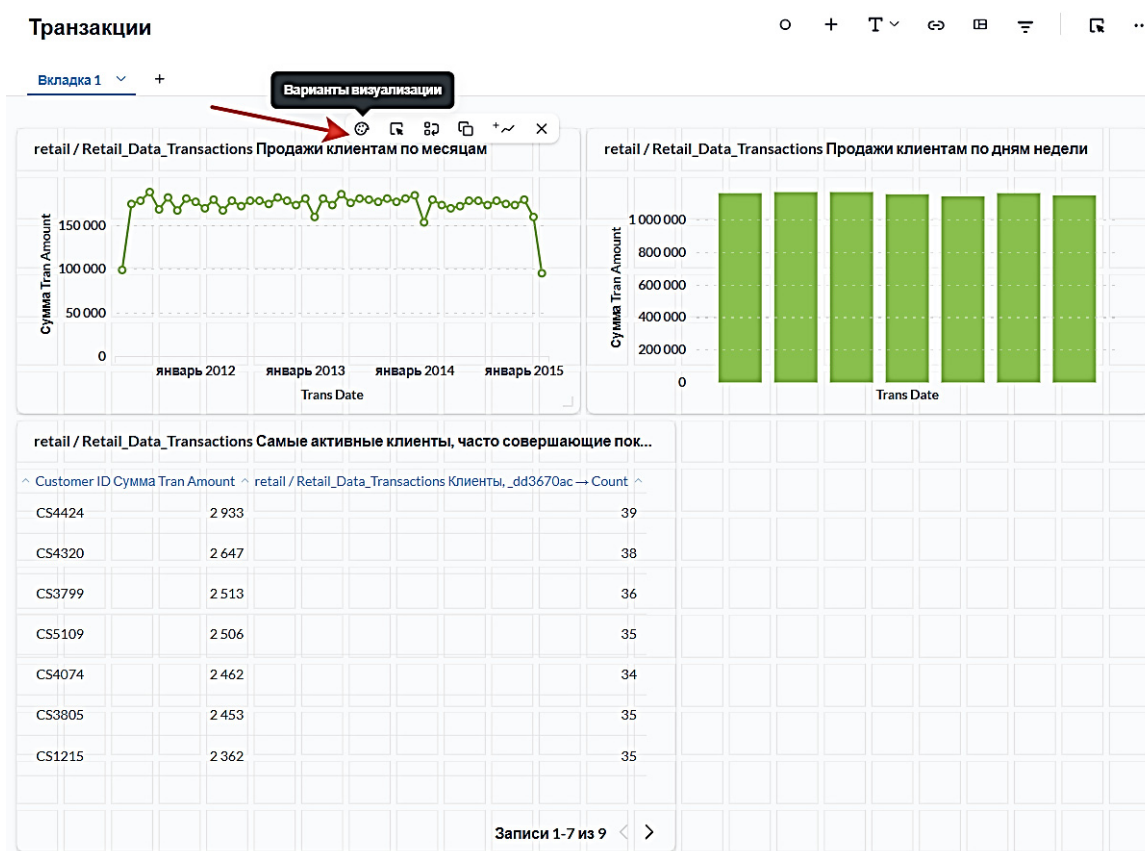


Рис. 44. Переход к форматированию диаграммы

Для первой визуализации изменим заголовок и фон (рис. 45).

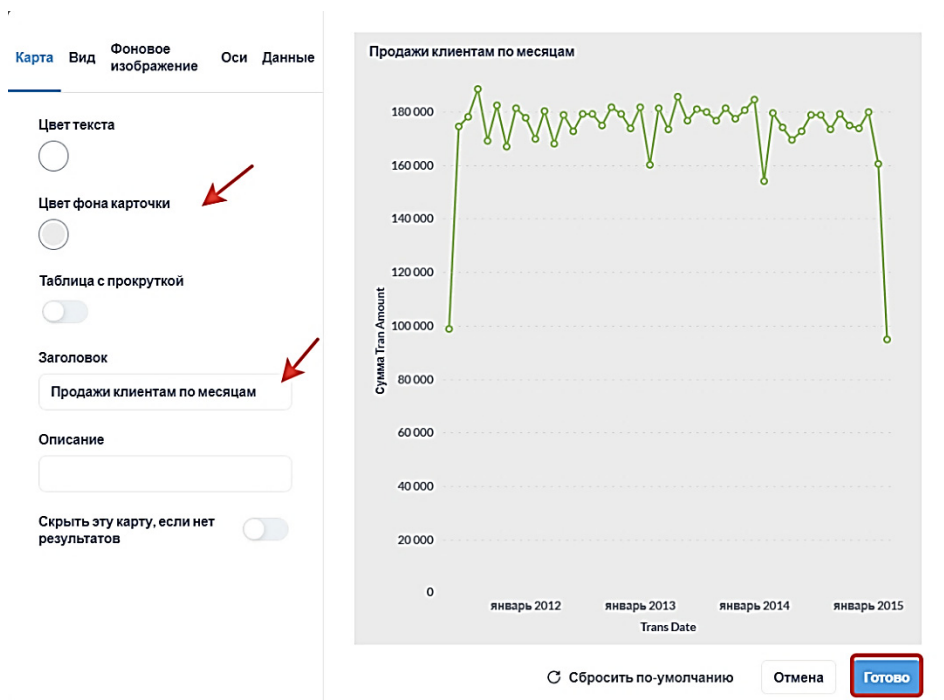


Рис. 45. Изменение заголовка и фона визуализации

Для второй визуализации также скорректируем заголовок, а в качестве фона выберем изображение (рис. 46). В качестве фонового изображения можно выбрать любое подходящее (поиск в Интернете).

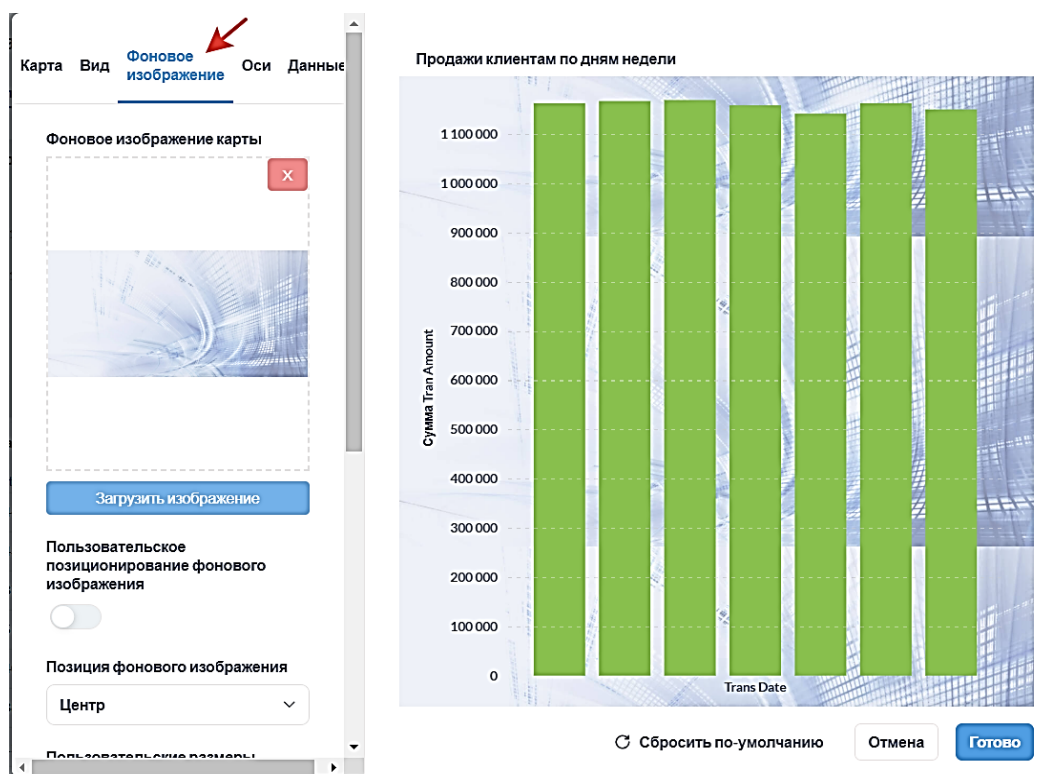


Рис. 46. Использование в качестве фона изображения

В визуализации по самым активным клиентам, часто совершающим транзакции, слишком длинное имя столбца нуждается в коррекции (рис. 47).

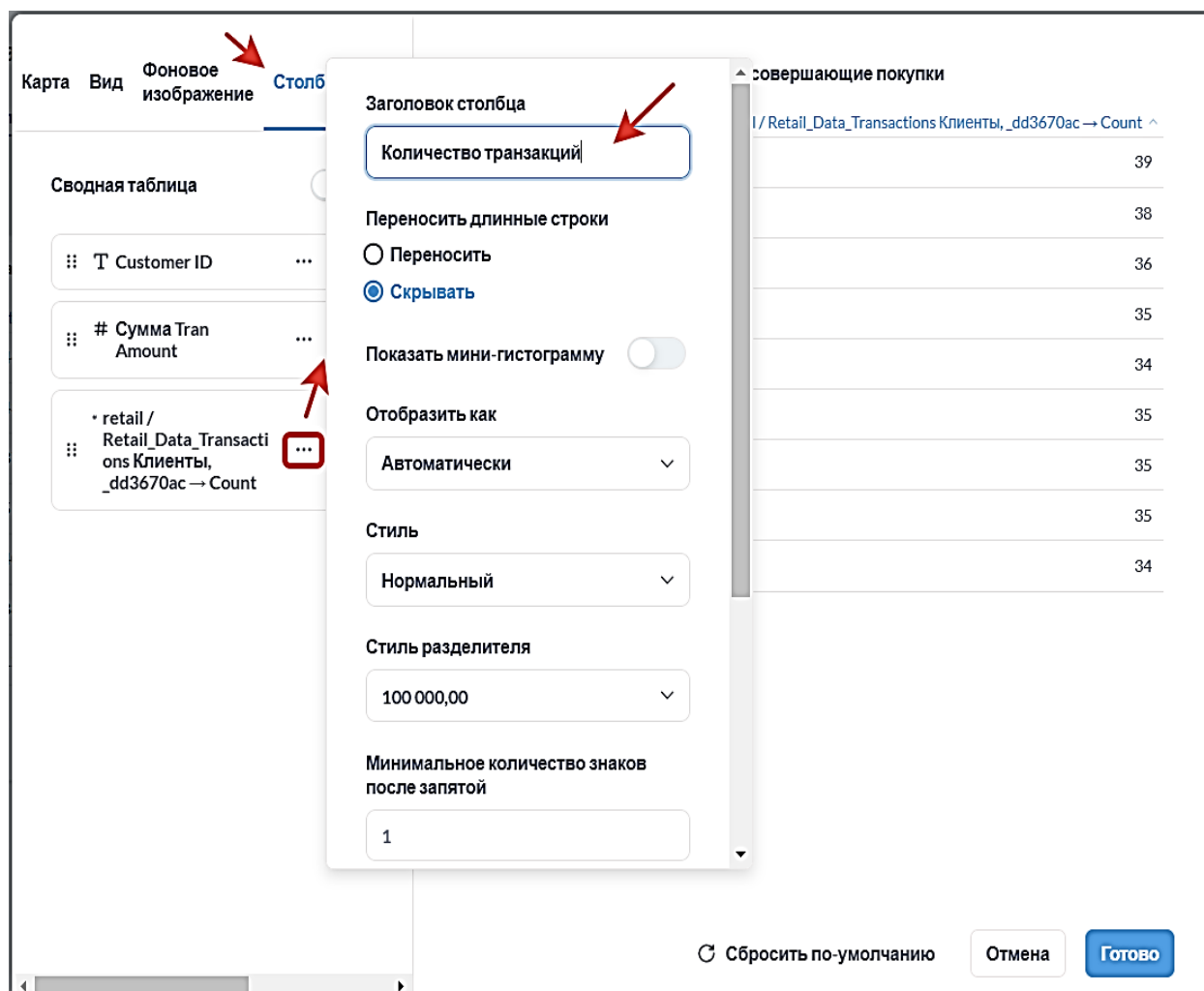


Рис. 47. Изменение имени столбца

После завершения форматирования визуализаций добавим фильтр (рис. 48).

Выберем для настройки фильтра показатель `trans_date` (рис. 49–51). В качестве типа фильтра выбирается Диапазон дат, чтобы иметь возможность рассматривать различные варианты диапазонов значений показателя `trans_date`. При выборе показателя для фильтрации он добавляется на каждую диаграмму, для которой эта фильтрация будет применяться. Также можно изменить ярлык (подпись) для фильтра.

В режиме редактирования всегда можно вернуться к настройкам фильтра (рис. 52).

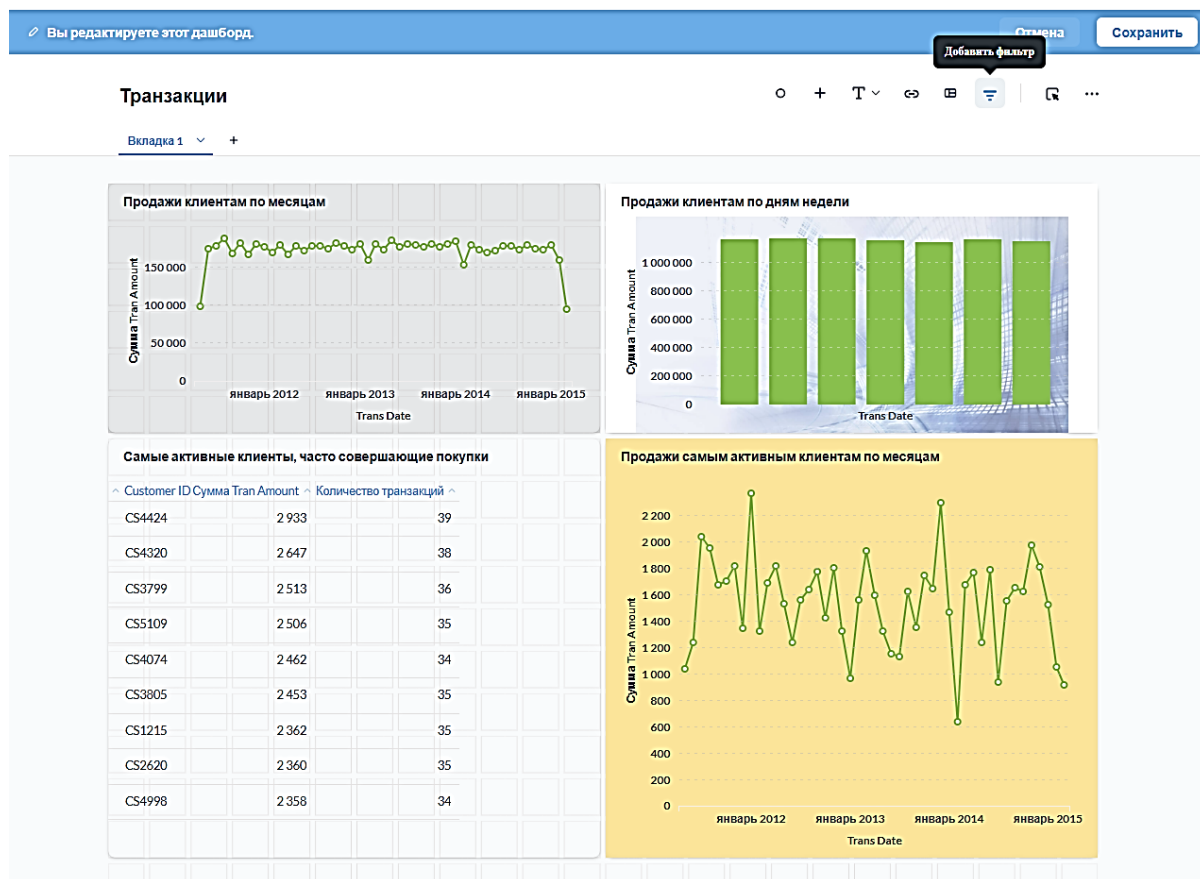


Рис. 48. Добавление фильтра

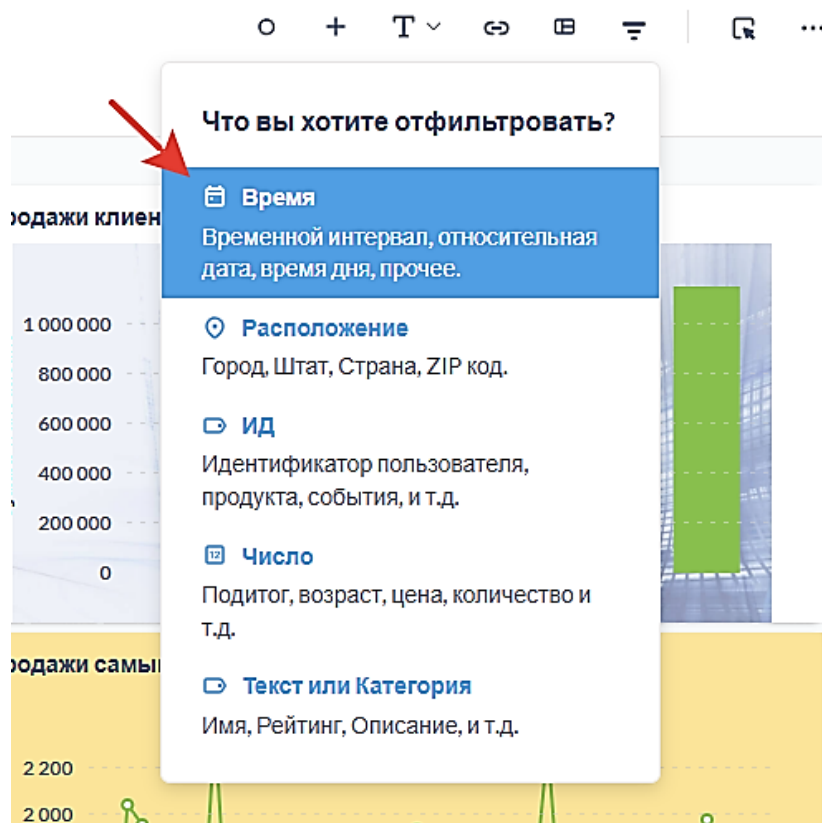


Рис. 49. Выбор типа показателя (поля) для настройки фильтра

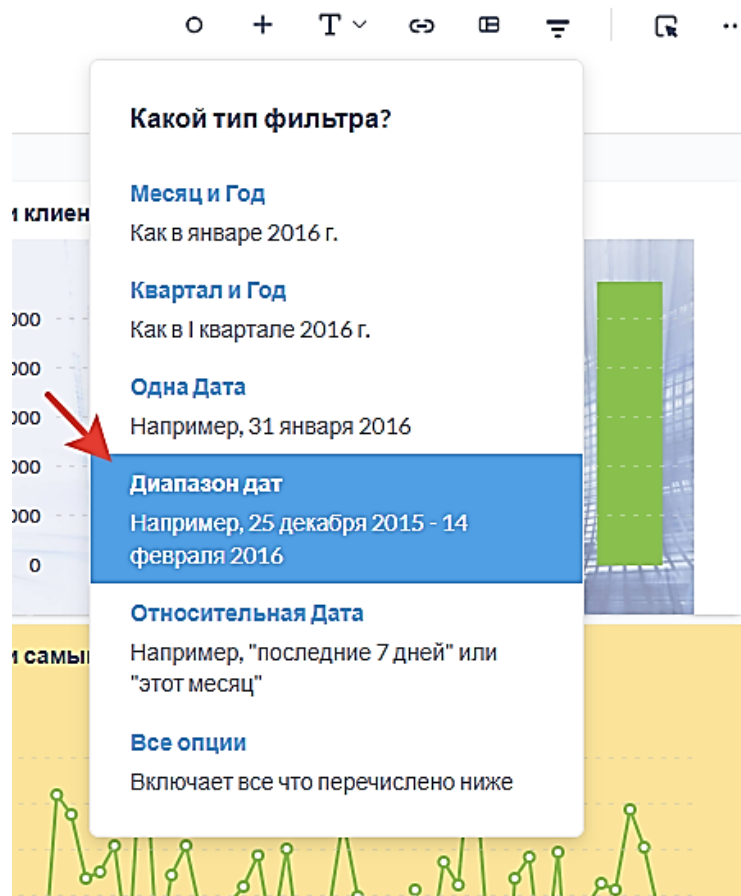


Рис. 50. Выбор типа фильтра

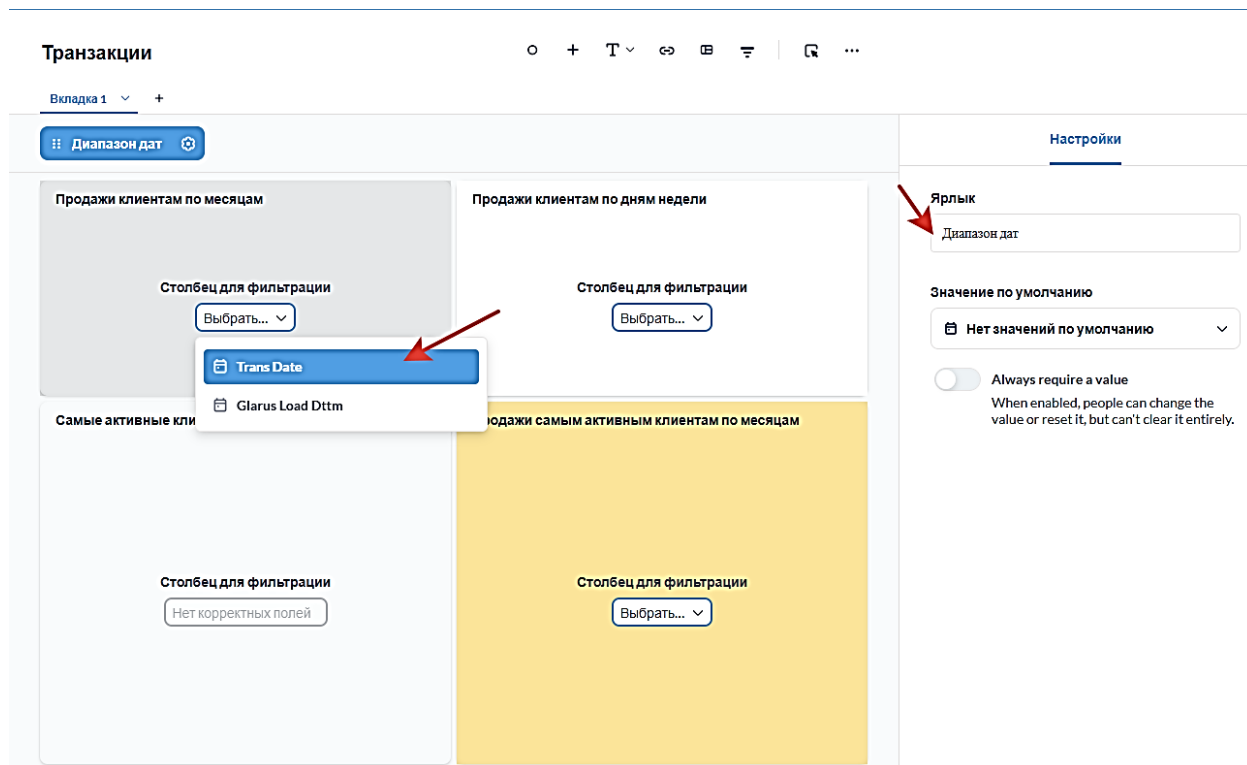


Рис. 51. Выбор показателя для фильтрации

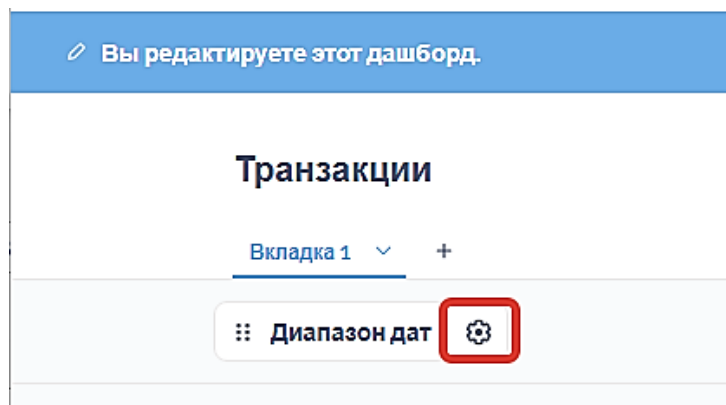


Рис. 52. Изменение настроек фильтра (при необходимости)

При сохранении дашборда автоматически перейдем в режим просмотра (рис. 53). Обрато в режим редактирования возвращает кнопка пиктографического меню . Внесем значения в фильтр по показателю trans_date.

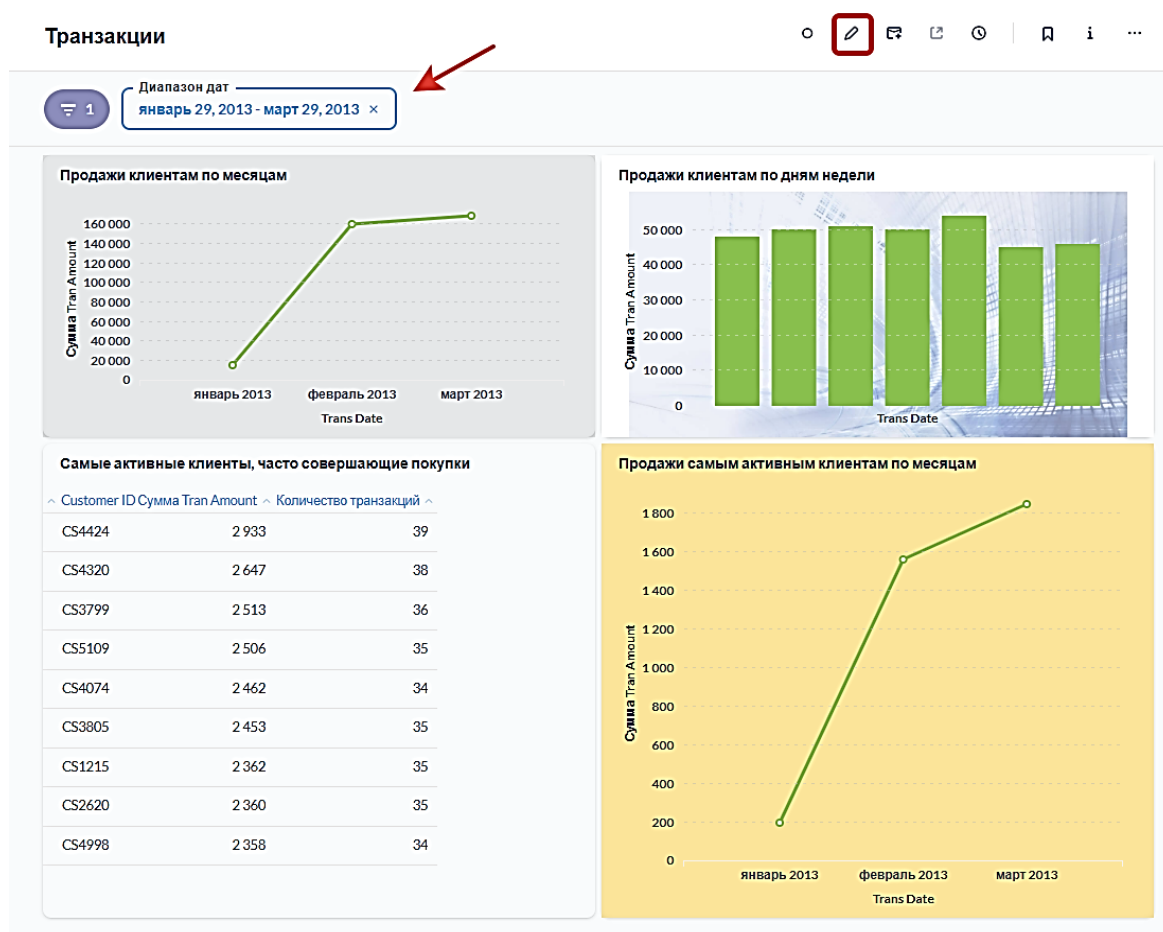


Рис. 53. Дашборд в режиме просмотра с настроенным фильтром

Добавим на дашборд вторую вкладку и поработаем с объединением визуализаций.

6. ОБЪЕДИНЕНИЕ ВИЗУАЛИЗАЦИЙ

Построим гистограмму по числу клиентов, сгруппированному в зависимости от отклика на рекламу (рис. 54, 55).

Количество по Response: Авт. скомпоновано
ClickHouse_Test / retail / Retail_Data_Response

Данные

retail/Retail_Data_Response

Фильтр

Добавьте фильтры, чтобы сократить результаты запроса

Суммировать

Количество × + по Response: Авт. скомпоновано × +

Фильтр Σ Присоединить данные Сортировка Количество строк Произвольная колонка

Визуализация

Рис. 54. Настройки запроса

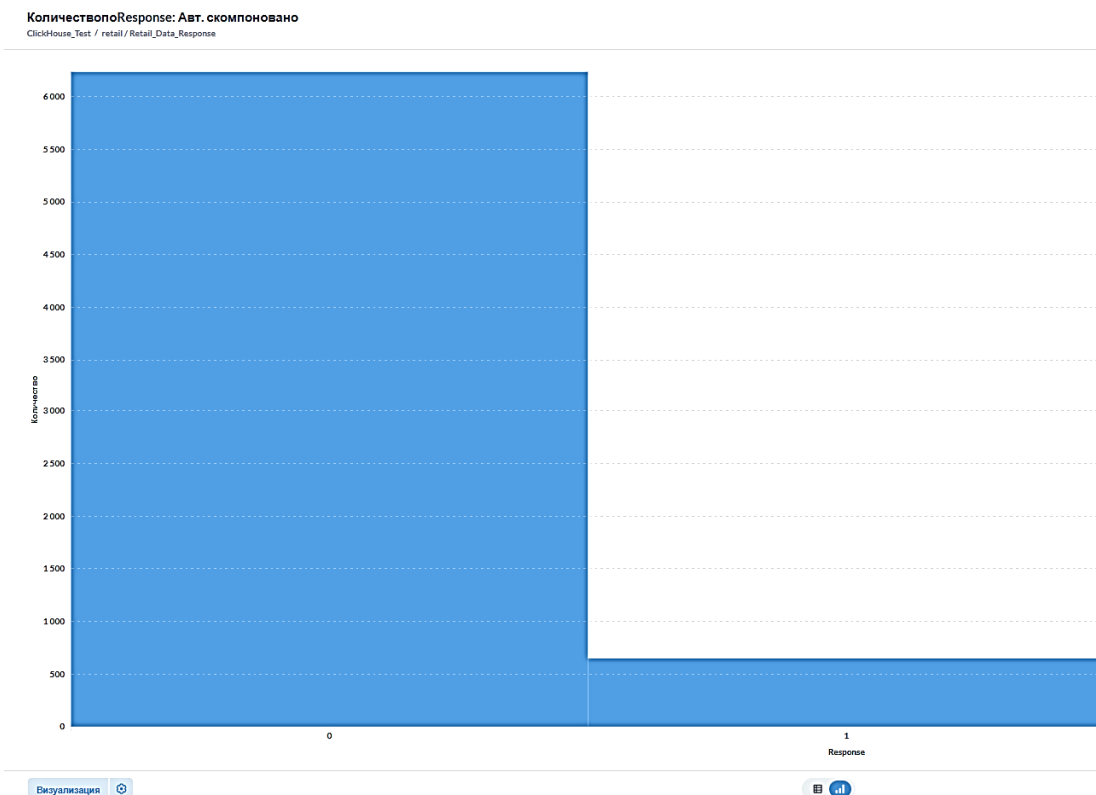


Рис. 55. Построенная визуализация

Разместим на дашборде гистограмму, построенную по количеству наблюдений, соответствующих каждому из значений переменной response (рис. 56). И в пиктографическом меню этой визуализации выберем значок  (Добавить серии), который позволяет дополнить диаграмму другой визуализацией, если в объединяемых визуализациях измерения оси абсцисс совпадают.

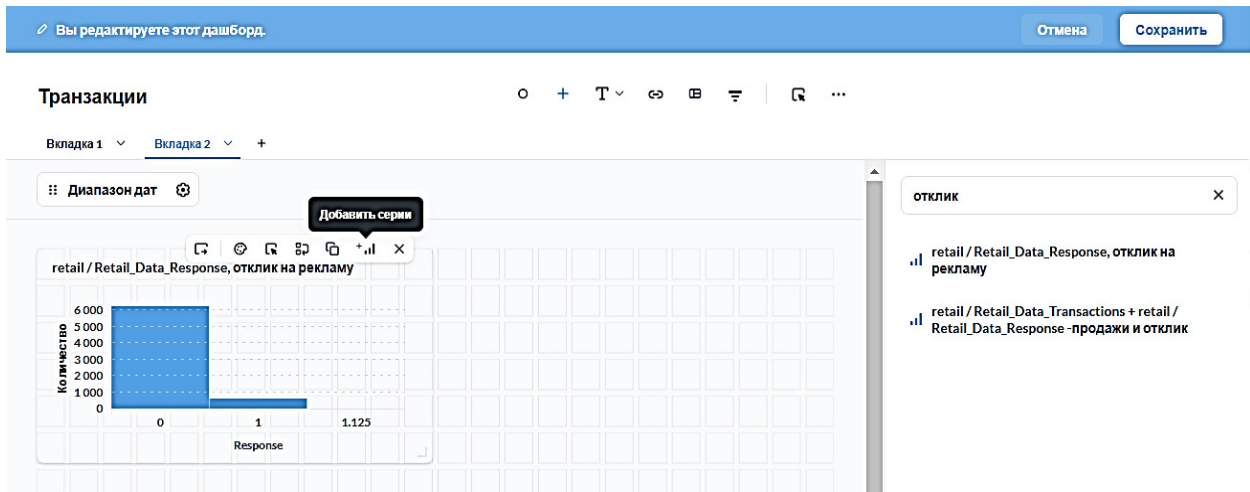


Рис. 56. Добавление визуализации на Вкладку 2 дашборда Продажи

В диалоговом окне для опции Добавить серии будут отображаться те диаграммы, которые можно объединить с текущей (рис. 57).

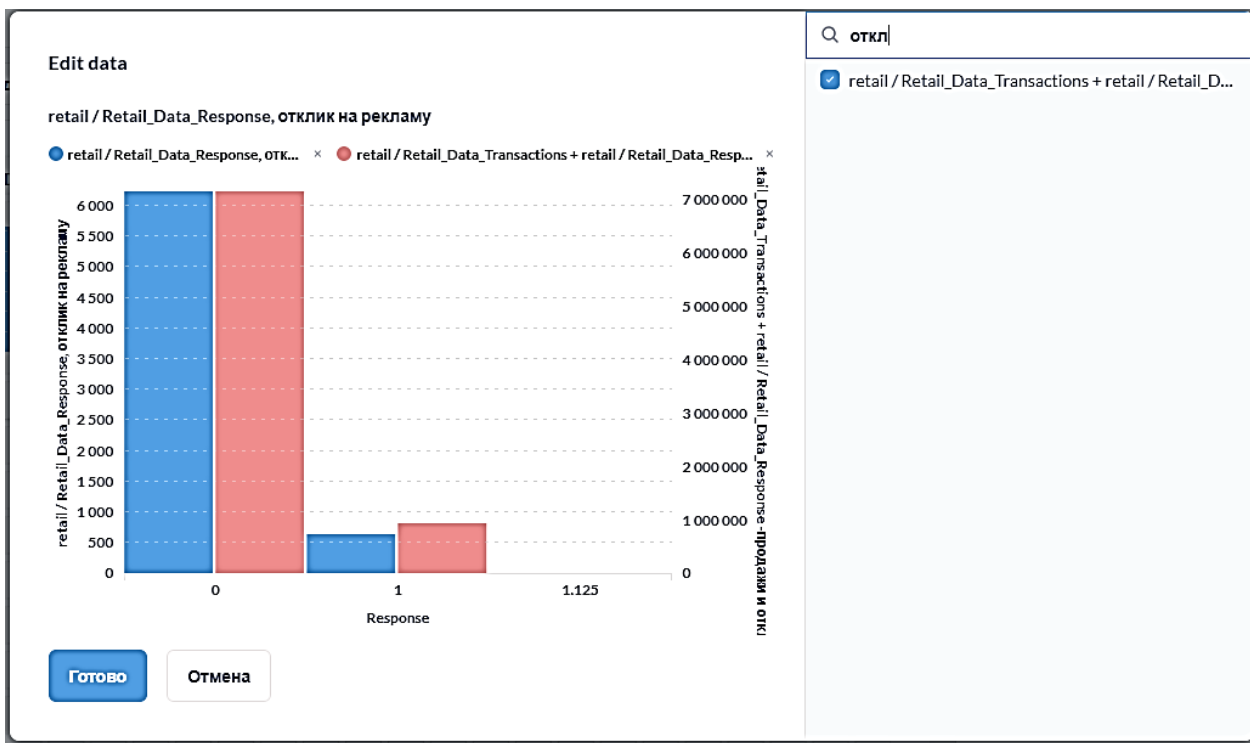


Рис. 57. Добавление серии на диаграмму

Скорректируем настройку объединенных диаграмм (рис. 58). Как видим, откликнувшиеся на рекламу покупают больше, просто в общем числе клиентов их доля невелика.

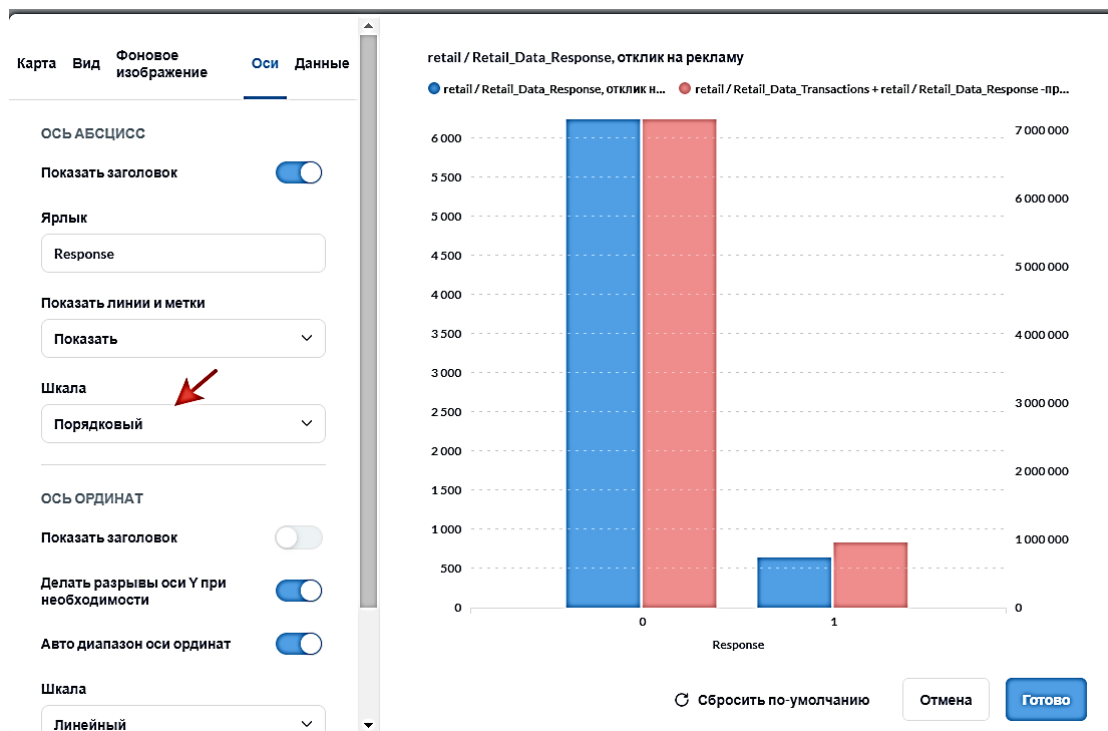


Рис. 58. Форматирование объединенной диаграммы

Изменим заголовок диаграммы, а на вкладке Вид скорректируем подписи легенды (рис. 59).

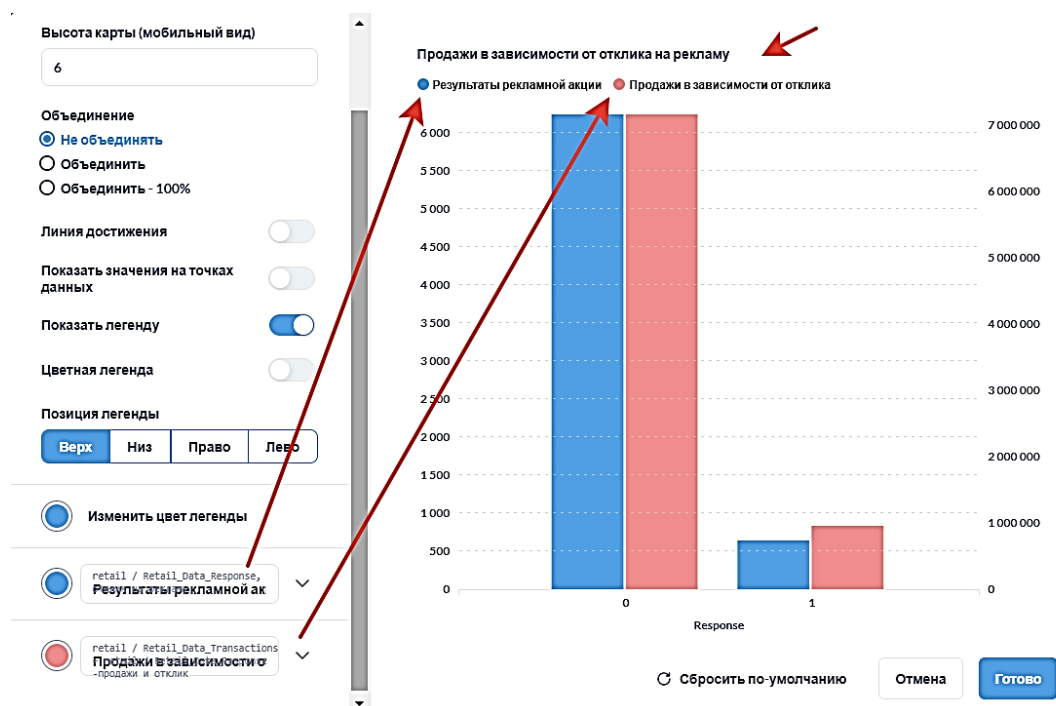


Рис. 59. Изменение легенды

После сохранения дашборд примет вид как на рис. 60. Готовый дашборд можно скачать в формате .pdf. Каждая вкладка будет скачиваться отдельной командой в отдельный файл. Но все файлы можно объединить в онлайн-редакторе.

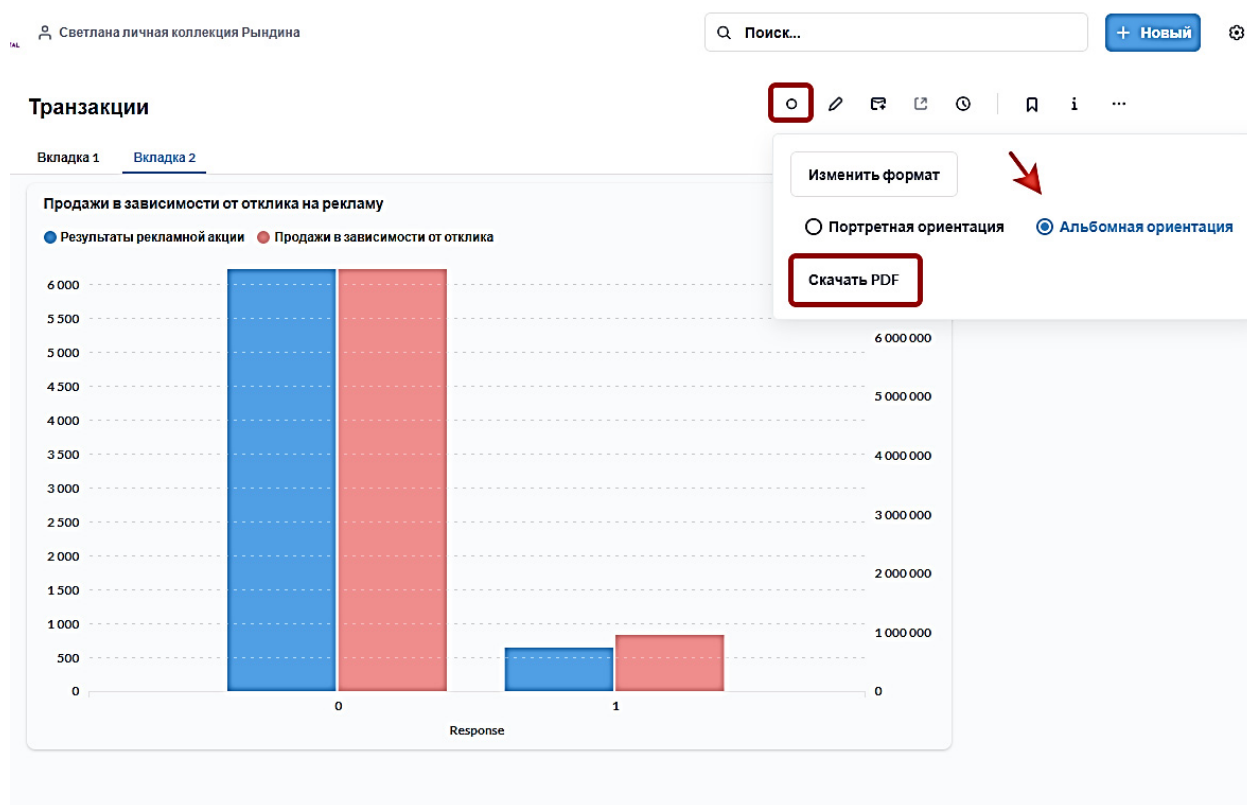


Рис. 60. Скачивание вкладки отчета в формате .pdf

Для дальнейшего совершенствования навыков работы на платформе Glarus BI можно воспользоваться справочной документацией [2].

Лабораторная работа

СОЗДАНИЕ ОТЧЕТА СРЕДСТВАМИ GLARUS BI

Цель работы: для набора данных создать отчет, используя возможности Glarus BI.

Формируемые знания, умения и навыки: научиться загружать данные, работать с вычисляемыми полями. Получить навыки формирования отчета по данным, включающего несколько визуализаций, с пользовательскими настройками форматирования.

Необходимо: по ссылке <https://www.kaggle.com/datasets/marian447/retail-store-sales-transactions?resource=download> загрузить набор данных scanner_data.csv. Перевести данные в формат .xlsx (использовать MS Excel). Удалить первый столбец с номерами строк и дополнительный лист (рис. 61). При ошибочном распознавании данных столбца sales_amount как текста нужно выделить столбец и выбрать опцию Текст по столбцам (рис. 62), два раза нажать в кнопку Далее в Мастере преобразований, а затем кликнуть Подробнее и заменить запятую на точку (рис. 63). Сохранить файл под именем scanner.xlsx (рис. 64).

Column1	Date	Customer_ID	Transaction_ID	SKU_Category	SKU	Quantity	Sales_Amount
1	02.01.2016	2547	1 X52	OEM7L	1	3.13	
2	02.01.2016	822	2 2ML	68BRQ	1	5.46	
3	02.01.2016	3686	3 OH2	CZUZX	1	6.35	
4	02.01.2016	3719	4 OH2	549KK	1	5.59	
5	02.01.2016	9200	5 OH2	K8ENN	1	6.88	
6	02.01.2016	5010	6 JPI	GVBRC	1	10.77	
7	02.01.2016	1666	7 XG4	AHAE7	1	3.65	
8	02.01.2016	1666	7 FEW	AHZNS	1	8.21	
9	02.01.2016	1253	8 OH2	9STQJ	1	8.25	
10	02.01.2016	5541	9 NSF	7IE9S	1	8.18	
11	02.01.2016	5541	9 H8O	M15RG	1	6.35	
12	02.01.2016	7548	10 N8U	UNJKW	1	2.11	
13	02.01.2016	7548	10 JRS	KWDJZ	1	5.38	
14	02.01.2016	6865	11 I4Y	8ETY5	1	8.04	
15	02.01.2016	6044	12 N8U	EMJ1S	1	3.62	
16	02.01.2016	6044	12 X52	CRK31	1	3.96	
17	02.01.2016	592	13 YMJ	YJTMP	1	11.65	
18	02.01.2016	592	13 P42	B2IW9	1	9	
19	02.01.2016	6450	14 N8U	4A2M0	1	7.02	
20	02.01.2016	3874	15 FEW	F79YP	1	16.91	
21	02.01.2016	3874	15 SFC	26XOE	2	4.75	
22	02.01.2016	5599	16 7C6	OZEQ8	1	7.85	
23	02.01.2016	5599	16 OQA	DWJIC4	1	12.87	
24	02.01.2016	9223	17 TW8	E63V3	1	1.62	
25	02.01.2016	9223	17 Q4N	EGB8E	1	10.5	
26	02.01.2016	4921	18 H15	NEYQ4	1	3.75	
27	02.01.2016	6294	19 NSF	4PLF3	1	2.91	
28	02.01.2016	6294	19 F9B	W9RKF	1	7.9	
29	02.01.2016	6294	19 C8Z	S20UE	1	9.06	
30	02.01.2016	6294	19 C8Z	S20UE	1	9.06	

Рис. 61. Вид таблицы после импорта данных из текстового файла

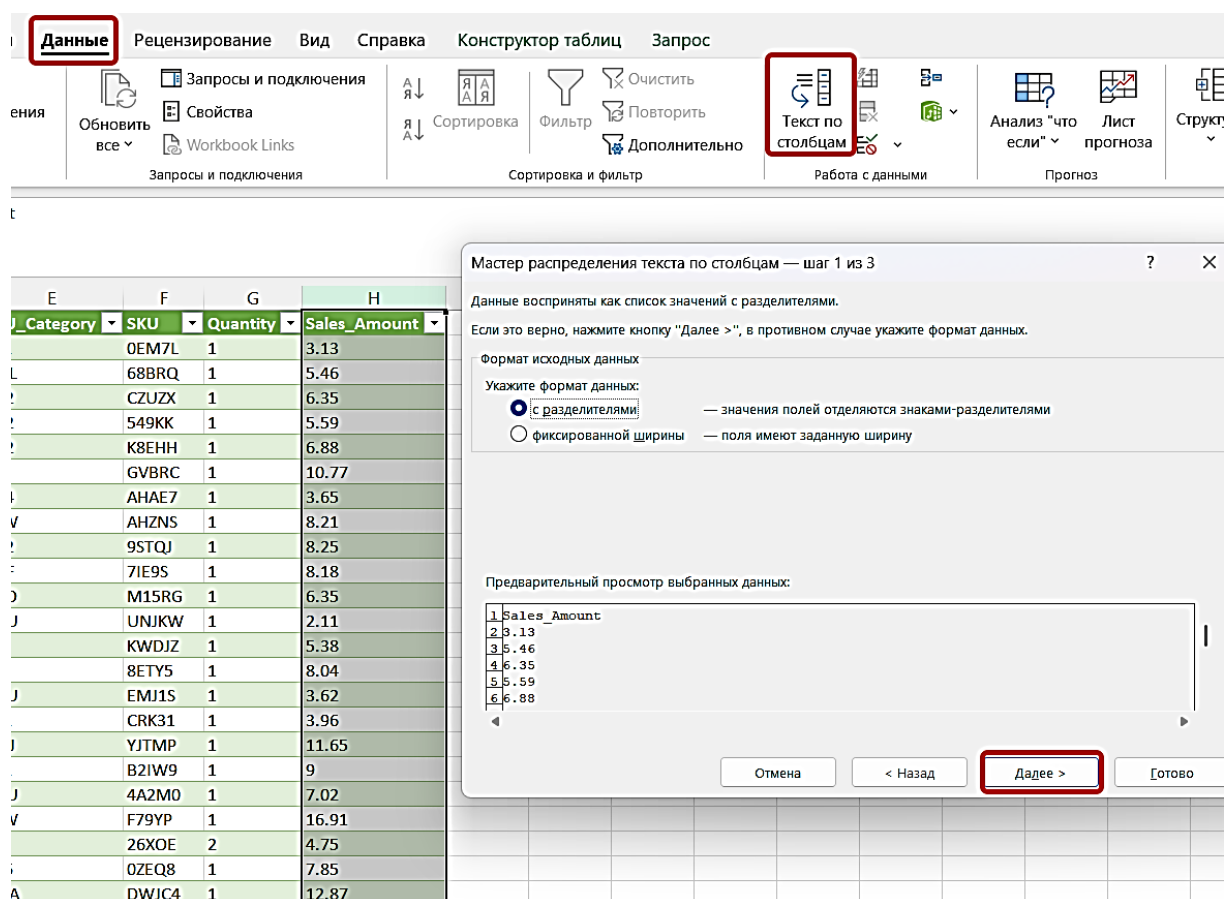


Рис. 62. Выбор опции Текст по столбцам для показателя Sales_amount

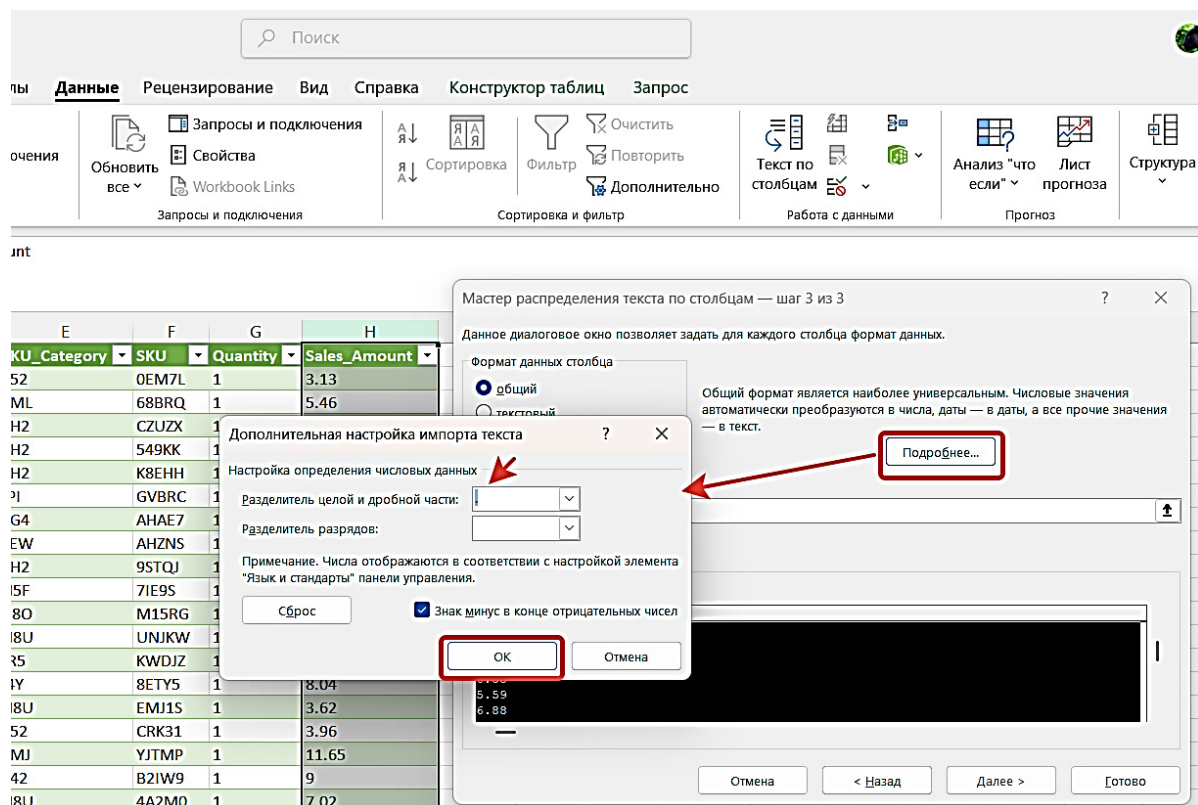


Рис. 63. Выбор дополнительных настроек на третьем шаге

	A	B	C	D	E	F	G
1	Date	Customer_ID	Transaction_ID	SKU_Category	SKU	Quantity	Sales_Amount
2	02.01.2016	2547	1 X52	OEM7L	1		3,13
3	02.01.2016	822	2 2ML	68BRQ	1		5,46
4	02.01.2016	3686	3 OH2	CZUZX	1		6,35
5	02.01.2016	3719	4 OH2	549KK	1		5,59
6	02.01.2016	9200	5 OH2	K8EHH	1		6,88
7	02.01.2016	5010	6 JPI	GVBRQ	1		10,77
8	02.01.2016	1666	7 XG4	AHAE7	1		3,65
9	02.01.2016	1666	7 FEW	AHZNS	1		8,21
10	02.01.2016	1253	8 OH2	9STQJ	1		8,25
11	02.01.2016	5541	9 N5F	7IE9S	1		8,18
12	02.01.2016	5541	9 H8O	M15RG	1		6,35
13	02.01.2016	7548	10 N8U	UNJKW	1		2,11
14	02.01.2016	7548	10 JR5	KWDJZ	1		5,38
15	02.01.2016	6865	11 I4Y	8ETY5	1		8,04
16	02.01.2016	6044	12 N8U	EMJ1S	1		3,62
17	02.01.2016	6044	12 X52	CRK31	1		3,96

< > scanner_data +

Рис. 64. Вид таблицы с данными перед сохранением

Дальнейший план работы:

1. Загрузить файл scanner.xlsx на платформу Glarus BI.
2. Построить различные виды запросов.
3. Создать дашборд, на который добавить заголовки и поясняющий текст. Разместить на дашборде построенные запросы, при необходимости скорректировать визуализации (цвет фона, заголовки для визуализаций), добавить на дашборд фильтры.

Контрольные вопросы

1. Как загрузить данные в Glarus BI? Как просмотреть корректность определения типов данных для полей?
2. Какие визуализации включены в дашборд? Какую бизнес-проблему описывает дашборд? Какие элементы форматирования использовались для повышения наглядности представления данных дашборда?
3. Как создать сводную таблицу?
4. Как создать фильтр?
5. Как скачать дашборд в виде файла с расширением .pdf?

СПИСОК ЛИТЕРАТУРЫ

1. Glarus BI : офиц. сайт. URL: <https://glarus-bi.ru/> (дата обращения: 10.01.2025).

2. Документация Glarus BI : офиц. сайт. URL: <https://glarus-bi.ru/documentation/index.html> (дата обращения: 10.01.2025).

Учебное издание

Рындина Светлана Валентиновна

**ОСНОВЫ анализа больших
массивов данных (Glarus BI)**

Редактор *А. Г. Темникова*
Технический редактор *Н. В. Иванова*
Компьютерная верстка *Н. В. Ивановой*

Подписано в печать 25.04.2025.
Формат 60×84¹/₁₆. Усл. печ. л. 2,56.
Тираж 6. Заказ № 102.

Издательство ПГУ
440026, г. Пенза, ул. Красная, 40.
Тел.: (8412) 66-60-49, 66-67-77; e-mail: iic@pnzgu.ru

