

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

С. В. Рындина

**ИНТЕЛЛЕКТУАЛЬНЫЕ
ИНФОРМАЦИОННЫЕ СИСТЕМЫ
И ТЕХНОЛОГИИ: СИСТЕМЫ
BUSINESS INTELLIGENCE (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

Интеллектуальные информационные
системы и технологии: системы
Business Intelligence (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

Интеллектуальные информационные системы и технологии: системы Business Intelligence (Glarus BI): учеб.-метод. пособие / С. В. Рындина. – Пенза : Изд-во ПГУ, 2025. – 64 с.

Рассматриваются система Business Intelligence: Glarus BI и практические аспекты ее использования. Материал пособия соответствует программе дисциплины «Интеллектуальные информационные системы и технологии» и может быть использован при написании выпускной работы магистра.

Издание подготовлено на кафедре «Цифровая экономика» и предназначено для обучающихся по направлениям подготовки 38.04.01 «Экономика», 38.04.02 «Менеджмент», 38.04.03 «Управление персоналом», 38.04.08 «Финансы и кредит», также материал пособия может быть использован для знакомства с инфраструктурой обработки больших данных при изучении дисциплин «Основы анализа больших массивов данных» (направление подготовки 38.03.02 «Менеджмент»), «Бизнес-аналитика на основе больших данных», «Интеллектуальные информационные системы» (направление подготовки 38.03.05 «Бизнес-информатика»).

УДК 004.738.5(075)+339(075)

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
1. НАЧАЛО РАБОТЫ С GLARUS BI	5
1.1. Система Business Intelligence (Glarus BI)	5
1.2. Загрузка данных	7
1.3. Создание запросов.....	14
1.4. Создание дашборда	29
2. РАЗВЕДОЧНЫЙ АНАЛИЗ ДАННЫХ С GLARUS BI.....	39
2.1. Слияние данных	39
2.2. Объединение визуализаций.....	42
2.3. Разведочный анализ данных.....	48
Лабораторная работа. Создание отчета средствами GLARUS BI	61
СПИСОК ЛИТЕРАТУРЫ.....	62

ВВЕДЕНИЕ

В соответствии с учебным планом обучающиеся на первом курсе магистерских программ экономического профиля изучают дисциплину «Интеллектуальные информационные системы и технологии».

Целью учебно-методического пособия является знакомство студентов с одним из видов интеллектуальных информационных систем: систем бизнес-аналитики.

Рассмотрены вопросы, связанные с основным функционалом систем бизнес-аналитики: как создаются отчеты для поддержки принятия решений в бизнесе. На примере продукта Glarus BI от компании ООО «Гларус Диджитал» рассмотрены практические вопросы использования системы бизнес-аналитики: подключение источников данных, настройка статистических вычислений, построение отчетов в виде инструментальных панелей.

Задания для лабораторных работ имеют практическую направленность, позволяют познакомиться с основными компонентами BI-системы на конкретных бизнес-задачах. В процессе их выполнения формируются умения и навыки работы с системами бизнес-аналитики: подключение источников данных, создание запросов, использование статистических вычислений и методов визуализации для подготовки отчетов.

1. НАЧАЛО РАБОТЫ С GLARUS BI

1.1. 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 разработки 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.

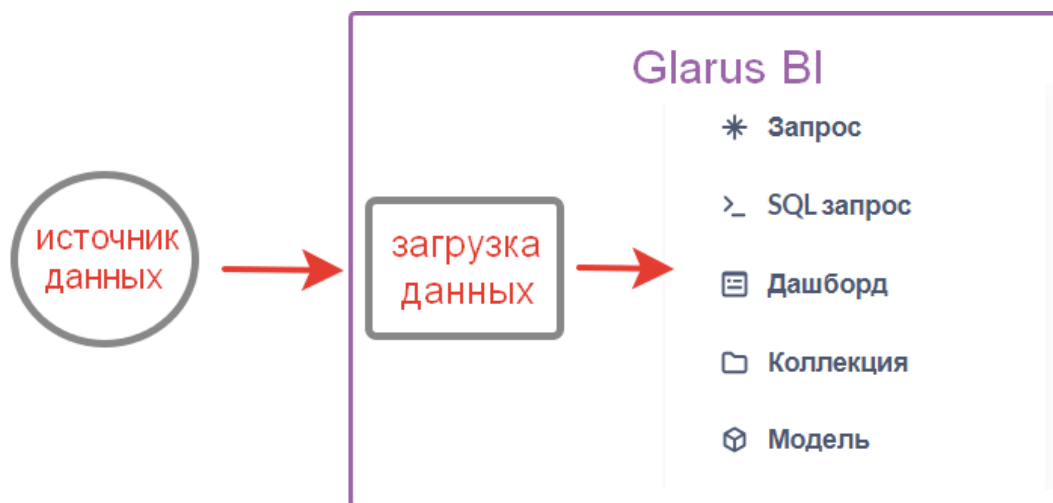


Рис. 1.1. Сущности Glarus BI

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 glarus-bi.online регистрационную форму.

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.

На рис. 1.2 показаны профиль пользователя и главное меню платформы.

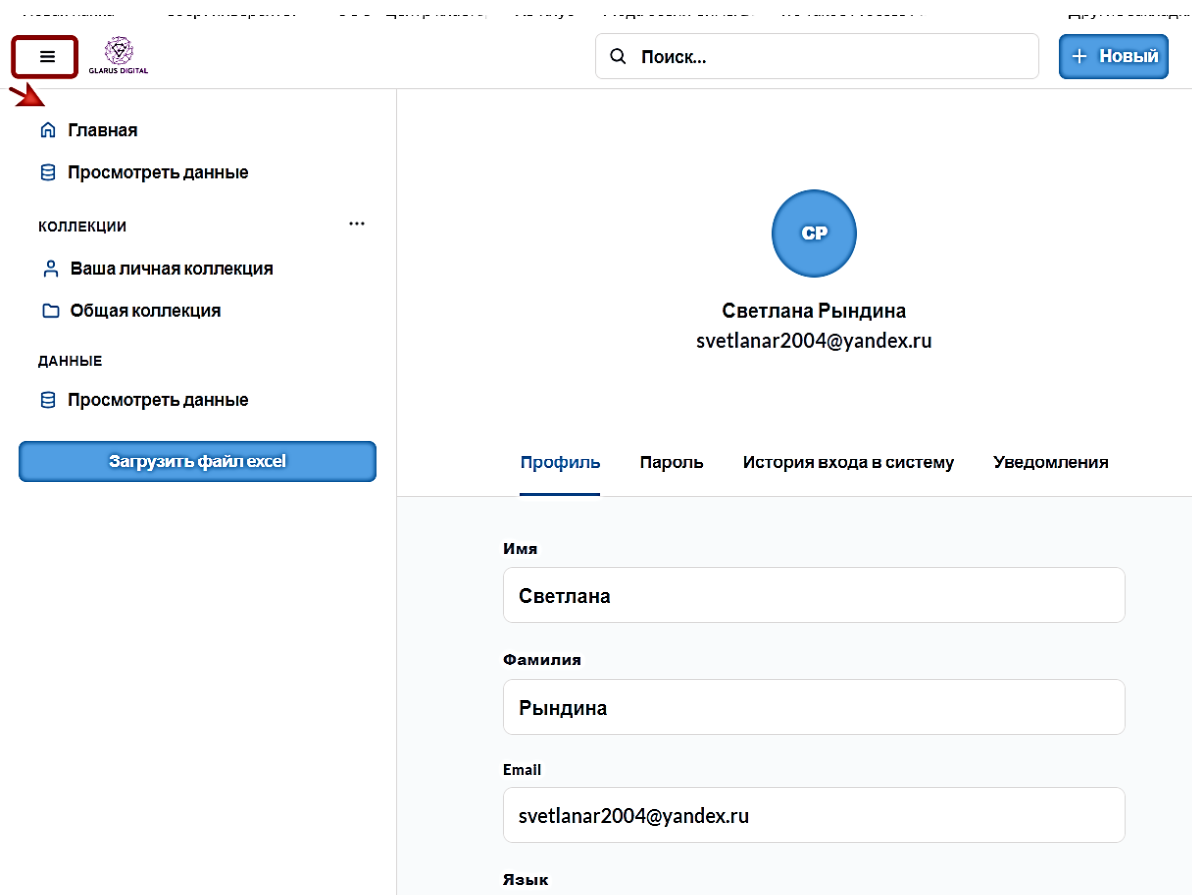


Рис. 1.2. Профиль пользователя и главное меню платформы

Выберем опцию Просмотреть данные. Изначально доступны только две дефолтные базы данных с таблицами Sample Database и Пример. После загрузки пользовательского файла с данными появляется и хранилище для файлов с данными пользователей ClickHouse_Test, в котором будут размещаться все загруженные файлы (рис. 1.3).

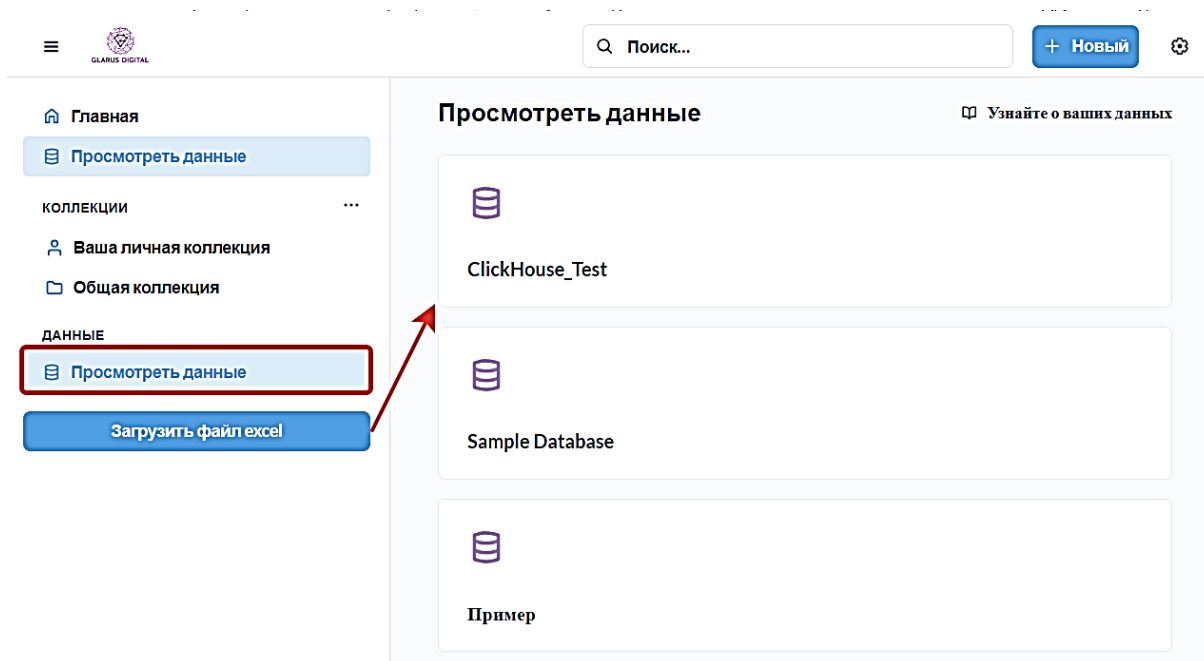


Рис. 1.3. Просмотр данных

Ваша личная коллекция хранит создаваемые сущности Glarus BI, доступные только самому пользователю. Общая коллекция позволяет настроить доступ к создаваемым сущностям и другим пользователям платформы.

1.2. Загрузка данных

Скачаем набор финансовых данных от компании Microsoft <https://go.microsoft.com/fwlink/?LinkID=521962> для выполнения тренировочных заданий. Загруженный на компьютер файл Financial Sample.xlsx имеет следующий состав полей:

- Segment – сегмент, к которому принадлежат товары;
- Country – страна, в которой происходит их реализация;
- Product – продукт;
- Discount Band – категория скидки (отсутствует/низкая/средняя/высокая);
- Units Sold – количество проданных единиц товара;
- Manufacturing Price – цена производства (себестоимость);

Sale Price – цена продажи;
Gross Sales – валовой объем продаж;
Discount – величина скидки;
Sales – продажа;
COGS – Cost of Goods Sold, себестоимость проданных товаров;
Profit – прибыль;
Date – дата продажи;
Month Number – номер месяца;
Month Name – название месяца;
Year – год.

Загрузим этот файл в Glarus BI (рис. 1.4).

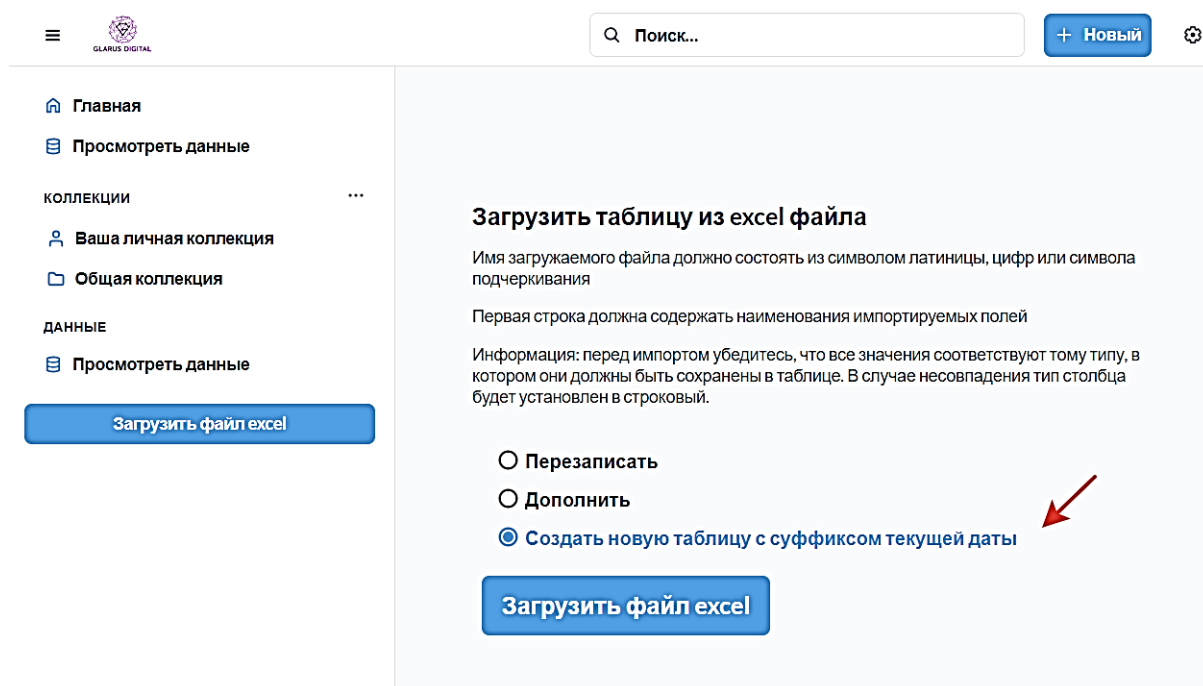


Рис. 1.4. Выбор файла для загрузки на платформу

После выбора файла в диалоговом окне проводника отобразится информация о выбранном файле и две кнопки, одна из которых позволяет изменить выбор файла, а вторая подтверждает загрузку (рис. 1.5).

Если все требования к файлу (содержимому, имени и расширению) соблюдены, то загрузка пройдет успешно (рис. 1.6).

Так как файл содержал только один безымянный лист с данными, то название загруженной таблицы в СУБД ClickHouse_Test будет составлено из имени файла Financial Sample и имени листа с данными Sheet1. При наведении курсора на имя таблицы становятся доступны опции сканирования, просмотра сведений о таблице и удаления (рис. 1.7).

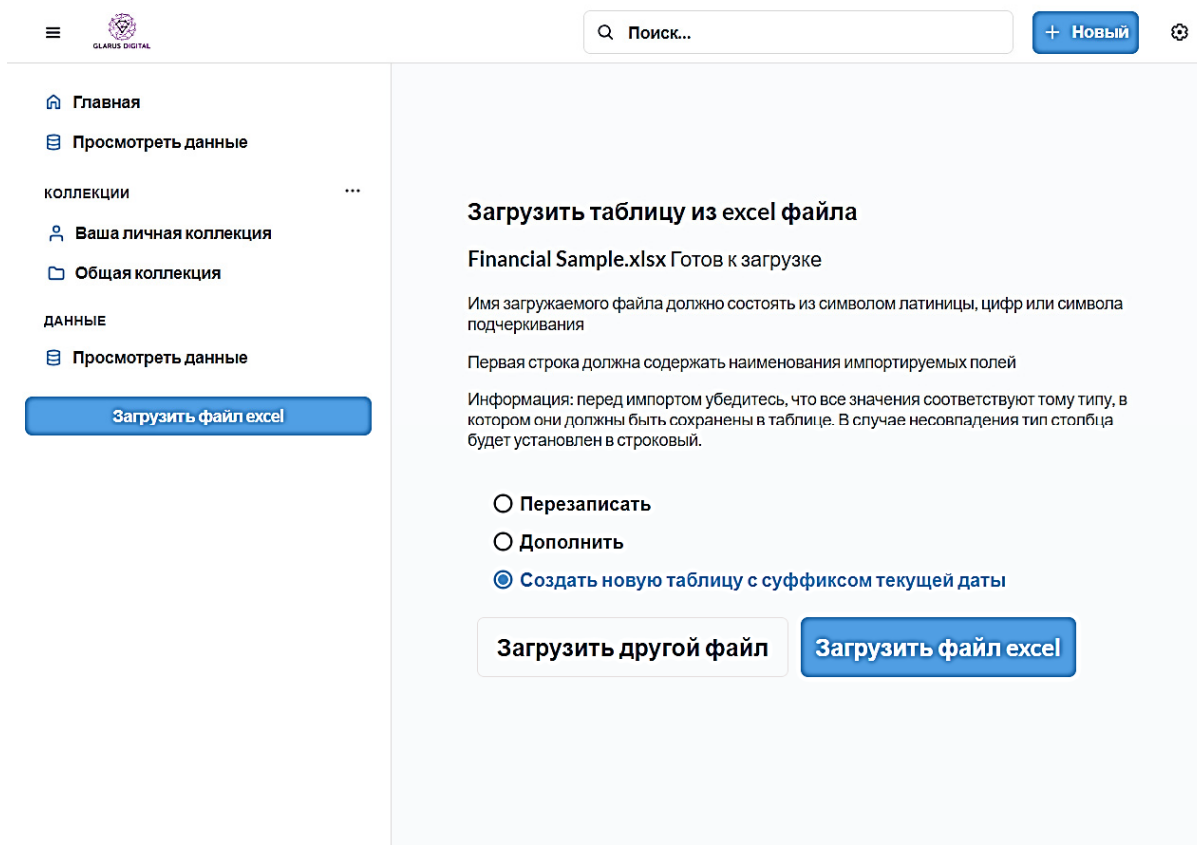


Рис. 1.5. Загрузка выбранного файла в Glarus BI

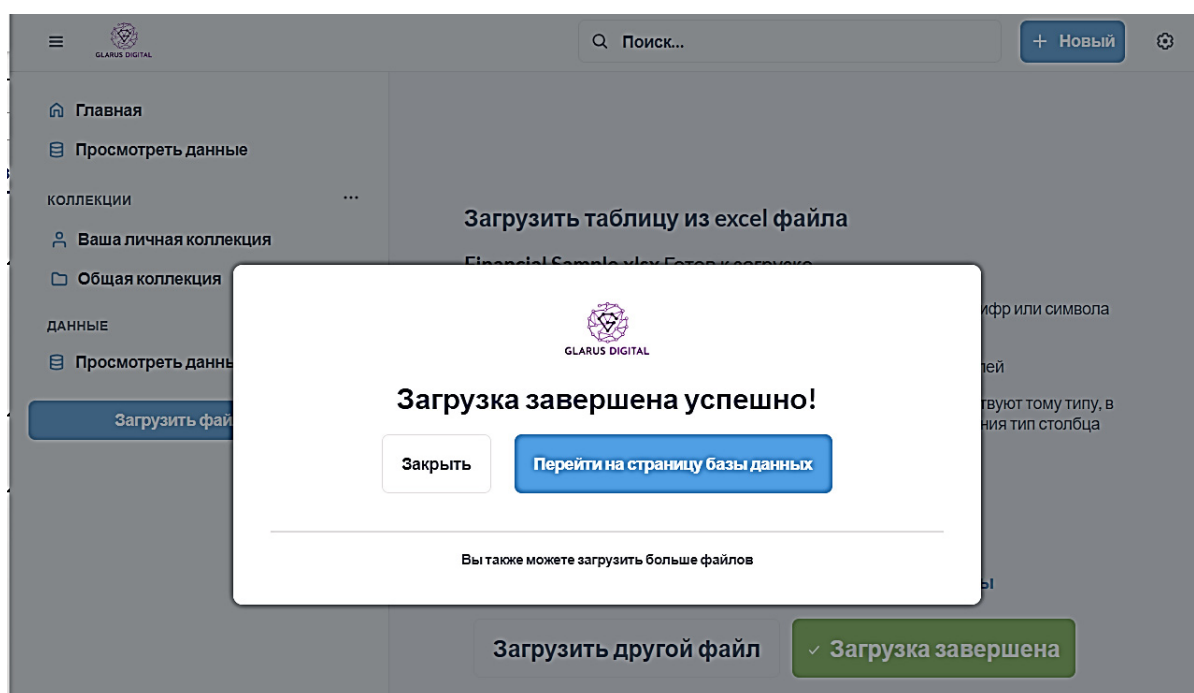


Рис. 1.6. Сообщение об успехе

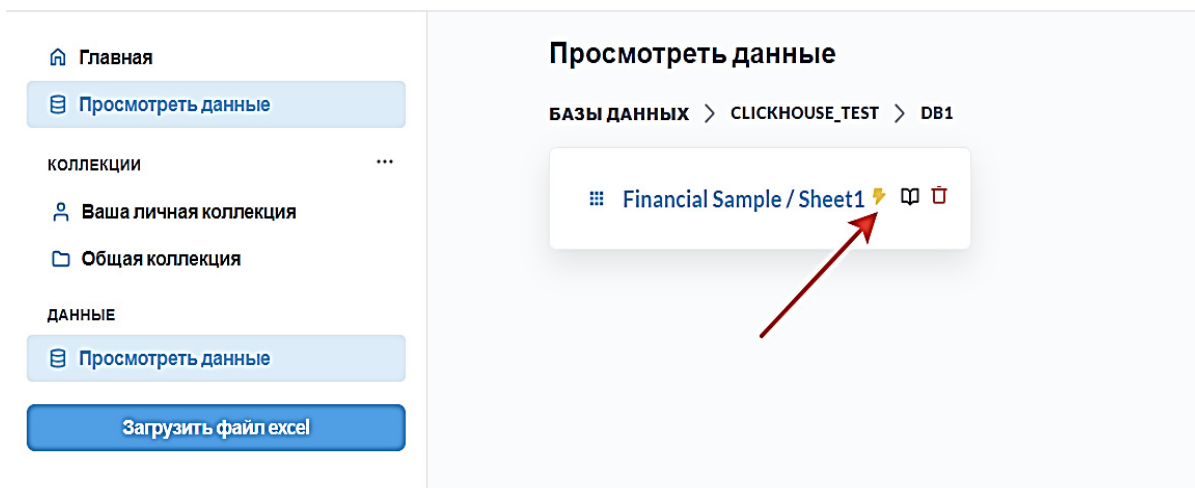
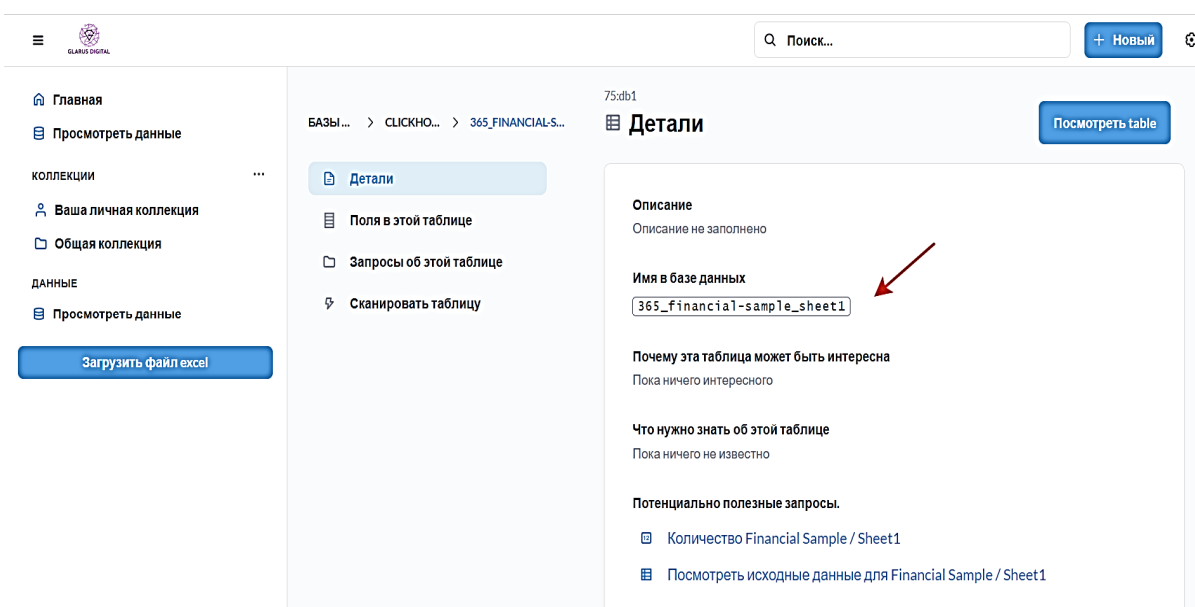


Рис. 1.7. Просмотр загруженных данных

Сканирование автоматически сгенерирует диаграммы на основе найденных типов данных в загруженной таблице. Это одна из максимально полезных функций для предварительного изучения данных и подсказок для формирования собственных пользовательских дашбордов с нужными визуализациями.

Но вначале посмотрим, верно ли загрузились данные, для чего выберем значок раскрытой книги  рядом с именем таблицы. Просмотрим сведения о таблице (рис. 1.8,а) и перейдем к проверке типов данных (рис. 1.8,б).



а)

Рис. 1.8. Сведения о таблице (а) и полях таблицы (б) (начало)

Детали				
Поля в этой таблице				
Запросы об этой таблице				
Сканировать таблицу				

Имя поля	Тип поля	Тип данных
Segment segment	T Категория	type/Text
Country country	📍 Страна	type/Text
Product product	T Категория	type/Text
Discount Band discount_band	T Категория	type/Text
Units Sold units_sold	# Нет типа поля	type/Float
Manufacturing Price manufacturing_price	T Категория	type/Integer
Sale Price sale_price	T Категория	type/Integer
Gross Sales gross_sales	# Нет типа поля	type/Float
Discounts discounts	# Скидка	type/Float
Sales _sales	# Нет типа поля	type/Float
Cogs cogs	# Нет типа поля	type/Float
Profit profit	# Нет типа поля	type/Float
Date date	📅 Нет типа поля	type/DateTime
Month Number month_number	# Количество	type/Integer
Month Name month_name	T Категория	type/Text
Year year	T Категория	type/Text

б)

Рис. 1.8. Окончание

Все поля имеют корректный тип данных:

- категориальные переменные Segment, Country, Product, Discount Band имеют тип данных Text – текстовый, причем для страны выбран особый тип поля – геолокация;

- числовые переменные Gross Sales, Discount, Sales, COGS, Profit имеют тип данных Float (вещественный, тип данных для десятичных дробей), также есть целочисленные числовые переменные – тип данных Integer, такие данные могут интерпретироваться как категориальные, если количество уникальных значений для таких числовых переменных невелико;

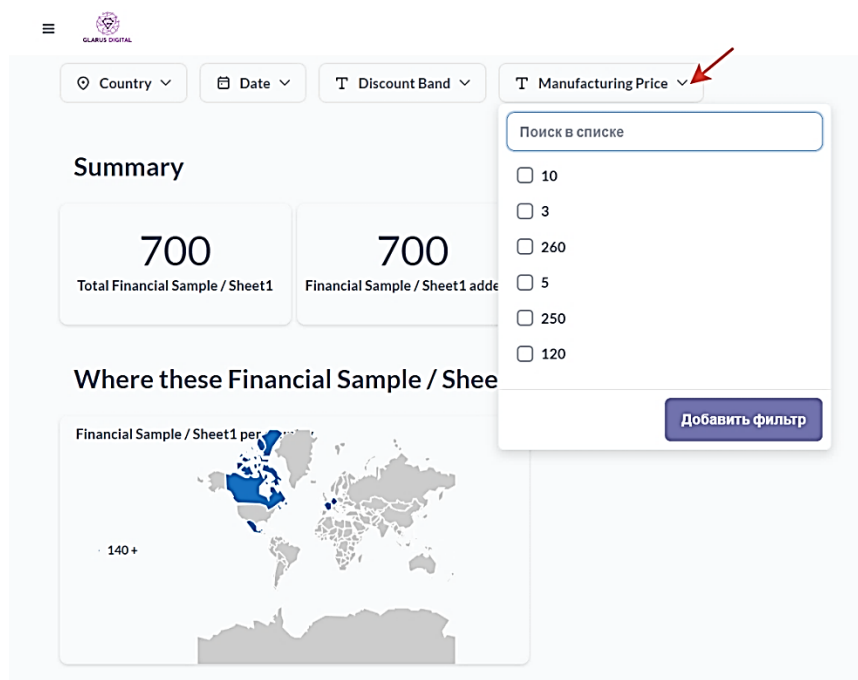
- переменная с типом DateTime – для даты.

Выполним сканирование загруженных данных. Как видим, эта опция доступна не только в виде значка рядом с именем (см. рис. 1.7), но и в текущем диалоговом окне.

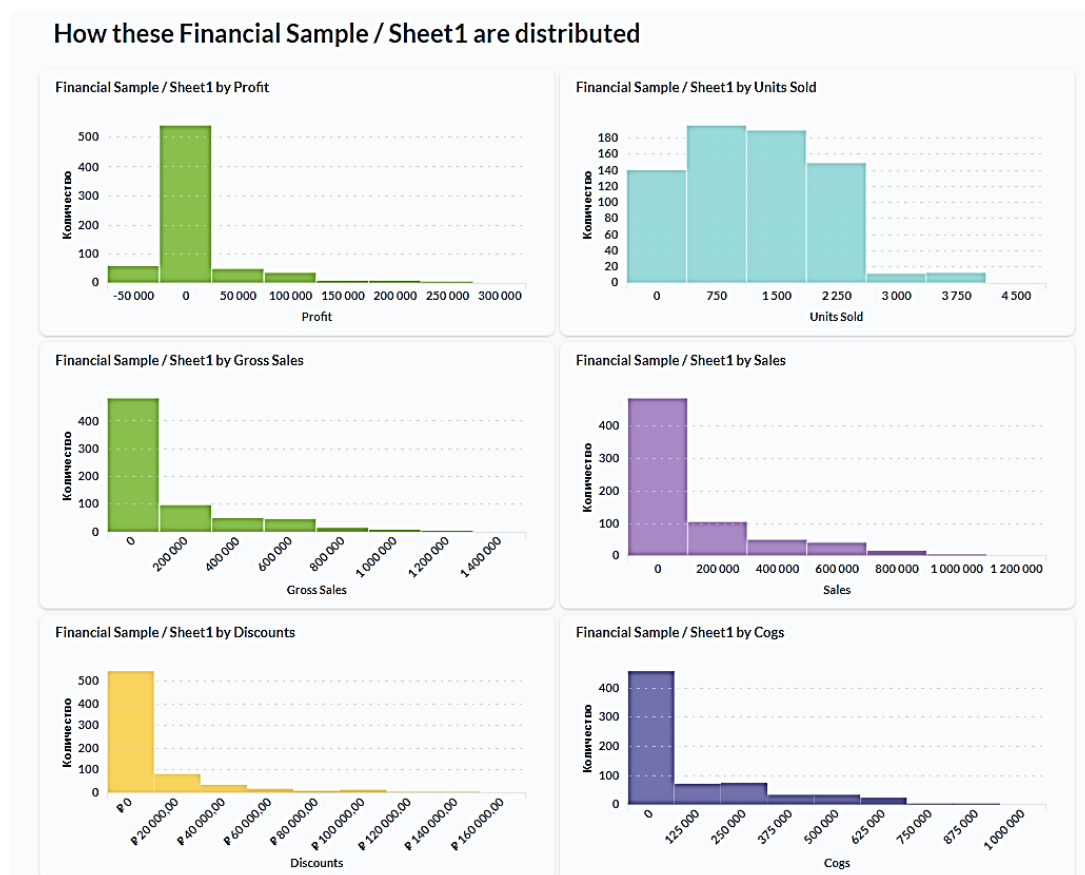
На просмотр выводится дашборд по всему массиву данных таблицы (рис. 1.9).

Верхняя строчка дашборда представляет собой группу фильтров по категориальным переменным. Один из фильтров Manufacturing Price раскрыт для настройки (рис. 1.9,а), как видим, уникальных значений для цены производства в этом наборе всего шесть. Далее идет краткая

информация о наборе данных (количество наблюдений и сведения об изменениях за последний месяц).



a)



б)

Рис. 1.9. Дашборд, автоматически сгенерированный платформой

Так как данные содержат геотеги (поле Country), то одна из визуализаций представлена в виде карты.

Часть визуализаций этого дашборда представлена на рис. 1.9,б.

Справа на построенном дашборде расположена панель навигации, с помощью которой есть возможность получить дополнительные дашборды, исследующие взаимосвязи отдельных полей с другими показателями набора данных (рис. 1.10).

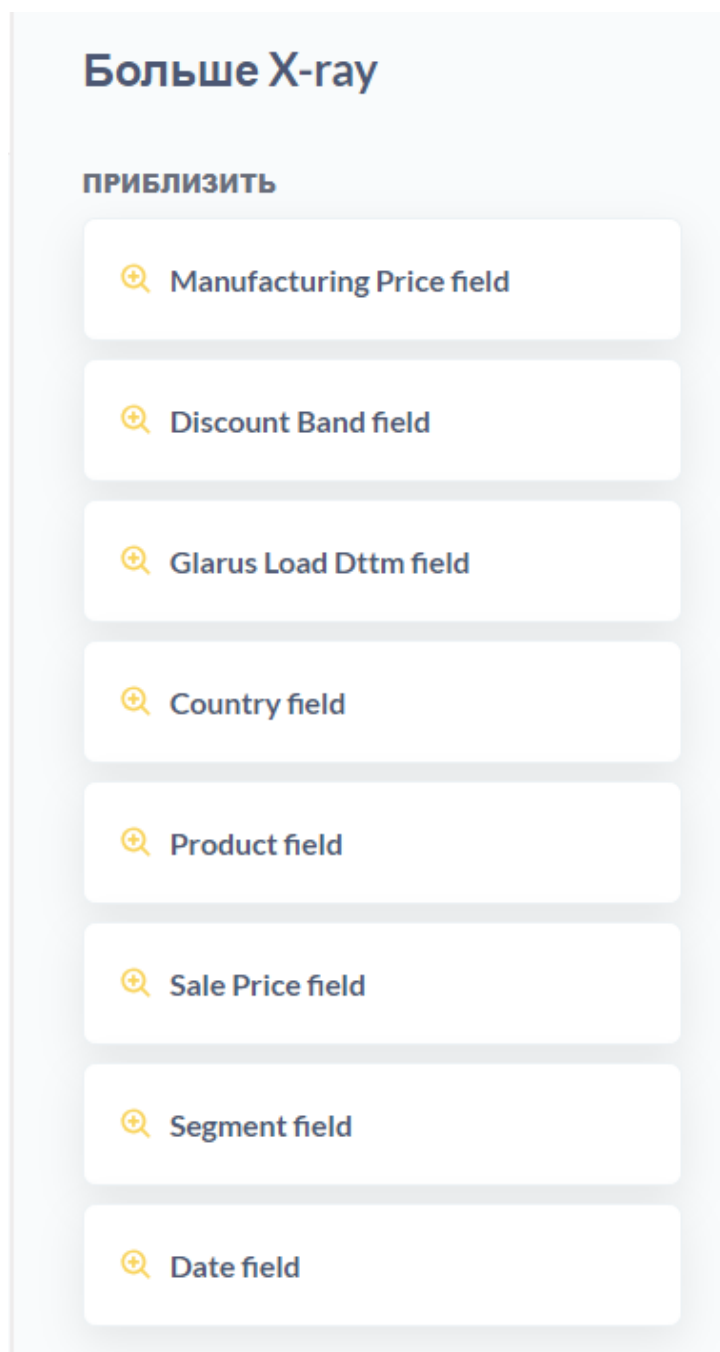


Рис. 1.10. Навигация по дашбордам, построенным на основе связей отдельных полей

Начало дашборда для поля Segment представлено на рис. 1.11.

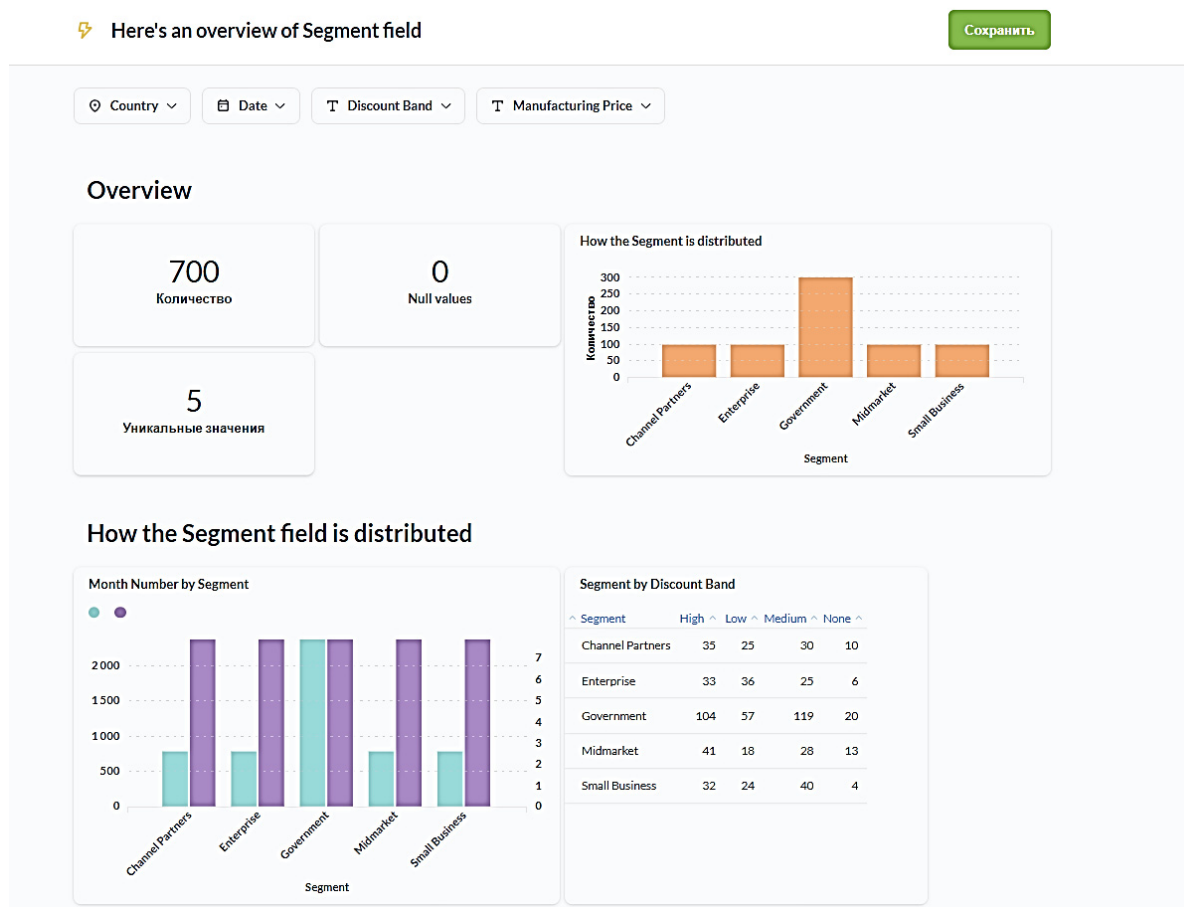


Рис. 1.11. Дашборд, построенный для поля Segment

Познакомившись с проведенным платформой разведочным анализом данных, т.е. построенными визуализациями на основе определенных типов данных, типов полей и выявленных взаимосвязей, перейдем к созданию собственных визуализаций.

1.3. Создание запросов

Пользовательские визуализации можно построить на основании всего набора данных, а можно по предварительно сделанному срезу. И то и другое реализуется с помощью запросов, которые служат основой визуализации.

Запрос, который загружает весь набор данных, – это просмотр содержимого таблицы. Он доступен и в диалоговом окне сведений о таблице – кнопка Посмотреть table (рис. 1.12), и с помощью главного меню – имя таблицы является ссылкой на просмотр ее содержимого (рис. 1.13). С помощью поисковой строки можно сразу получить доступ к нужной таблице в базе данных (рис. 1.14).

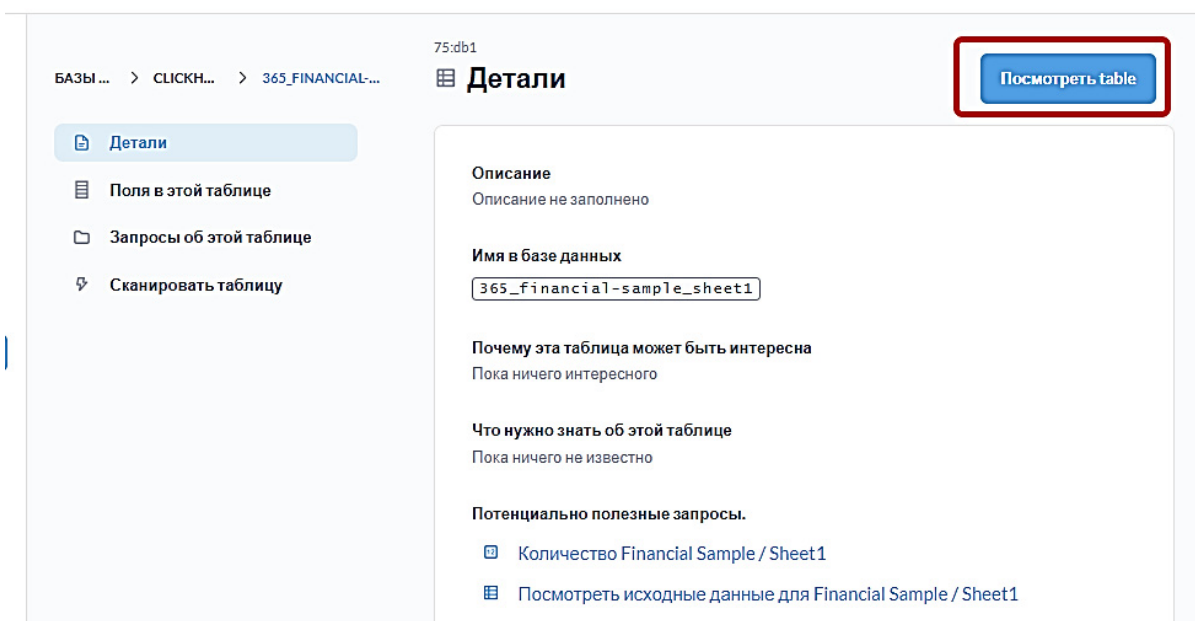


Рис. 1.12. Доступ к содержимому таблицы

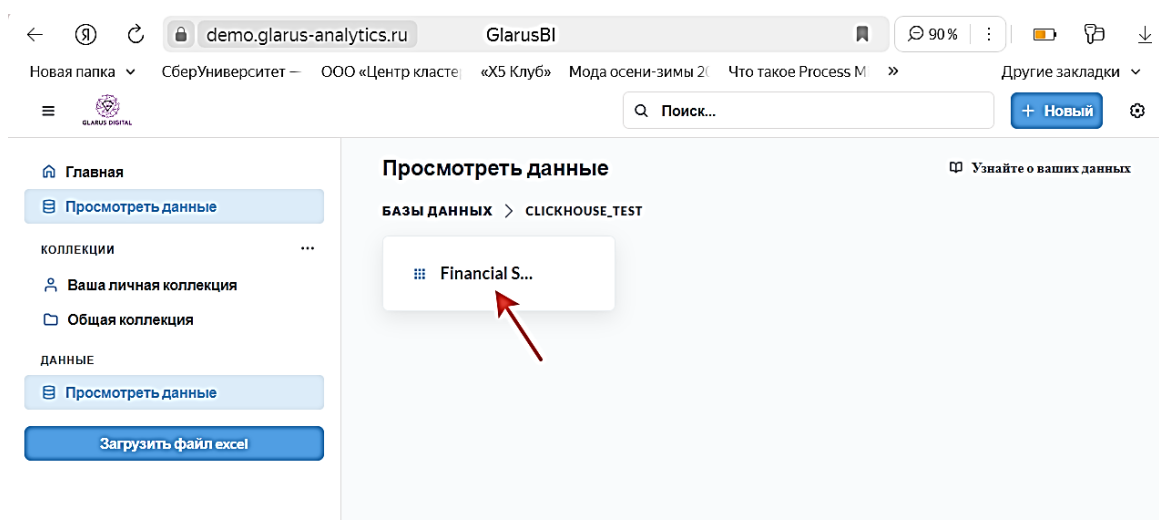


Рис. 1.13. Доступ к содержимому таблицы по ссылке, связанной с именем

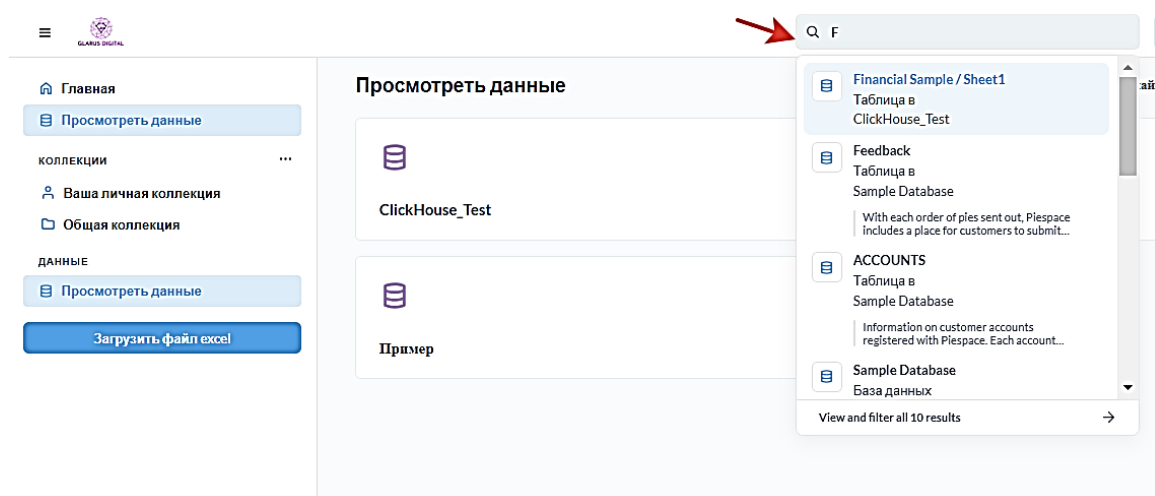


Рис. 1.14. Использование поисковой строки для доступа к данным

Для набора данных можно построить срез с помощью фильтра и сгруппировать данные (кнопка Суммировать), используя разные показатели для группировки (категориальные или даты/времени) и разные варианты агрегации на основе метрик (рис. 1.15). В качестве дефолтной используется метрика Количество записей, но ее можно изменить на более подходящую (рис. 1.16).

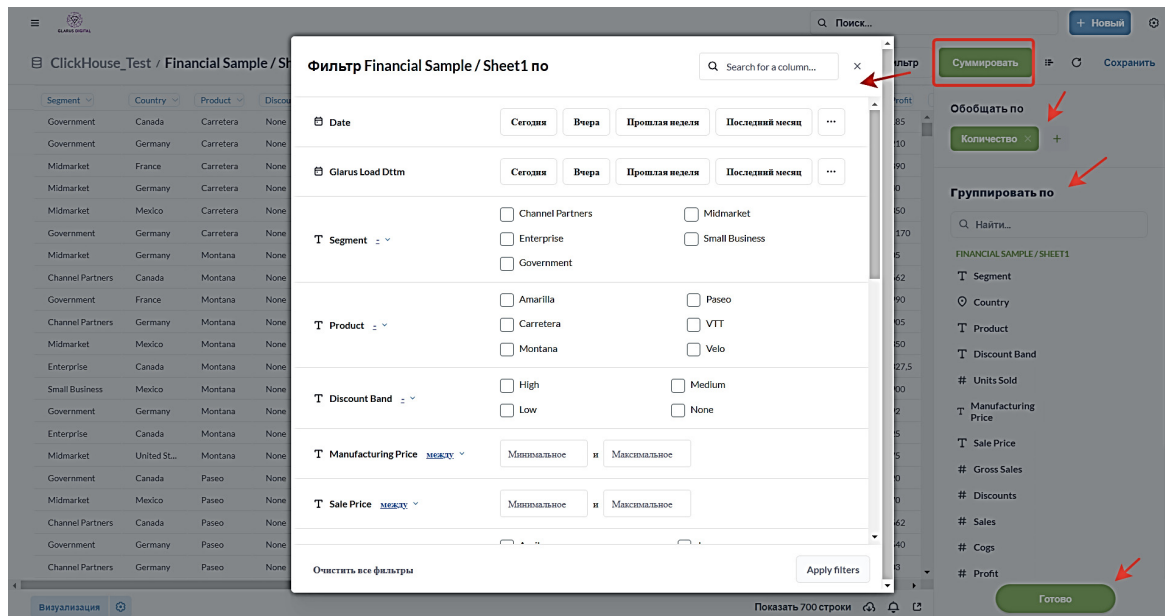


Рис. 1.15. Опции для построения среза данных

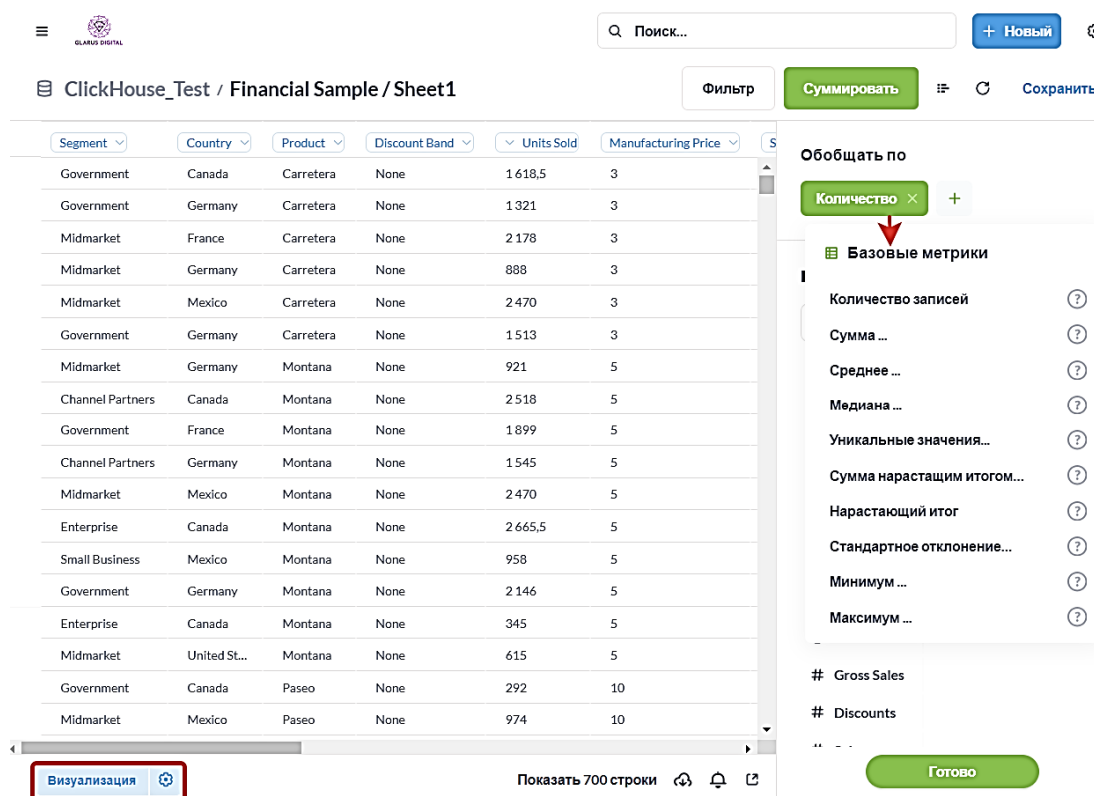


Рис. 1.16. Изменение метрики для агрегации количественных данных

Перейдем к построению визуализации (кнопка расположена в левом нижнем углу).

Список визуализаций, доступных для построения, появится справа (рис. 1.17). Сама таблица также является базовой визуализацией.

The interface shows a sidebar on the left with various visualization icons and their names in Russian: Таблица (Table), Гистограмма (Histogram), Линия (Line), Строка (Row), Область (Area), Комбо (Combo), Детали (Details), Карта (Map), Разброс (Scatter), and Водопад (Waterfall). Below these are 'ДРУГИЕ ГРАФИКИ' (Other Charts) including Пирог (Pie), Число (Number), Сводная таблица (Summary Table), Тренд (Trend), Спидометр (Speedometer), Прогресс (Progress), Воронка (Funnel), and Древоидная карта (Treemap).

At the bottom left is a blue button labeled 'Готово' (Done). At the bottom right is a blue button labeled 'Визуализация' (Visualization) with a settings icon.

On the right side, there is a table with columns 'Segment', 'Country', and 'Produ'. A red arrow points from the 'Визуализация' button to the table.

Segment	Country	Produ
Government	Canada	Carre
Government	Germany	Carre
Midmarket	France	Carre
Midmarket	Germany	Carre
Midmarket	Mexico	Carre
Government	Germany	Carre
Midmarket	Germany	Mont
Channel Partners	Canada	Mont
Government	France	Mont
Channel Partners	Germany	Mont
Midmarket	Mexico	Mont
Enterprise	Canada	Mont
Small Business	Mexico	Mont
Government	Germany	Mont
Enterprise	Canada	Mont
Midmarket	United St...	Mont
Government	Canada	Paseo
Midmarket	Mexico	Paseo
Channel Partners	Canada	Paseo
Government	Germany	Paseo
Channel Partners	Germany	Paseo

Рис. 1.17. Выбор визуализации для построения

Построим линейную диаграмму: ось X – Date, ось Y – Gross Sales. Для этого выберем в качестве визуализации линию и приступим к настройкам (рис. 1.18).

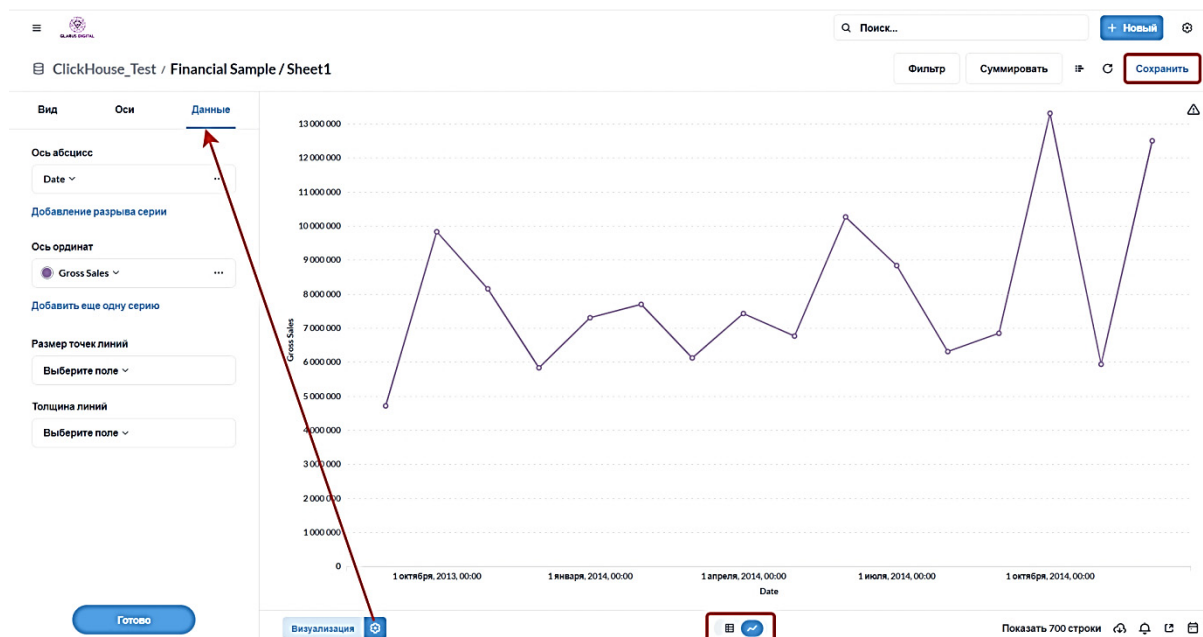


Рис. 1.18. Настройки для линейной диаграммы

Кнопка Готово переводит диаграмму в режим просмотра и убирает панель настроек. Внизу диаграммы доступны два режима для просмотра: исходных данных и визуализации. Сохраним построенную диаграмму (она сохраняется как запрос). Предложенное дефолтное название можно оставить без изменений, но добавить описание (рис. 1.19).

Сохранить новый запрос

Имя

Financial Sample / Sheet1 гистограмма

Описание

Линейная диаграмма по показателям Date и Gross Sales

В какую коллекцию следует это включить?

Светлана личная коллекция Рындина

Отмена

Сохранить

Рис. 1.19. Сохранение запроса

При успешном сохранении в диалоговом окне можно выбрать добавление сохраненного запроса на дашборд (рис. 1.20). Но так как ни одного дашборда еще не создано, то эта команда не сработает.

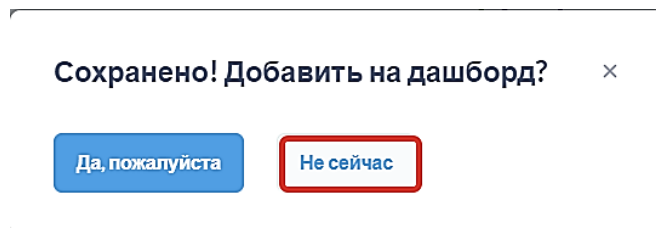


Рис. 1.20. Возможность добавления запроса на дашборд (при его наличии)

В правом верхнем углу кнопка Новый позволяет создать новый запрос (рис. 1.21).

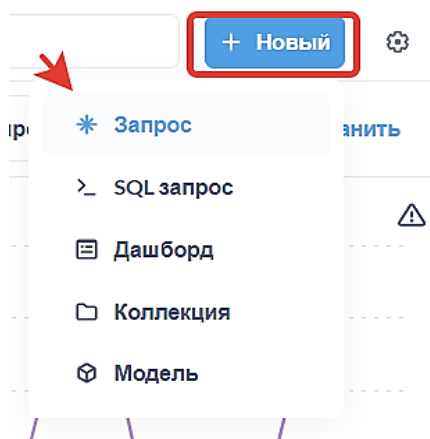


Рис. 1.21. Создание нового запроса

Для создания запроса необходимо выбрать начальные данные (рис. 1.22).

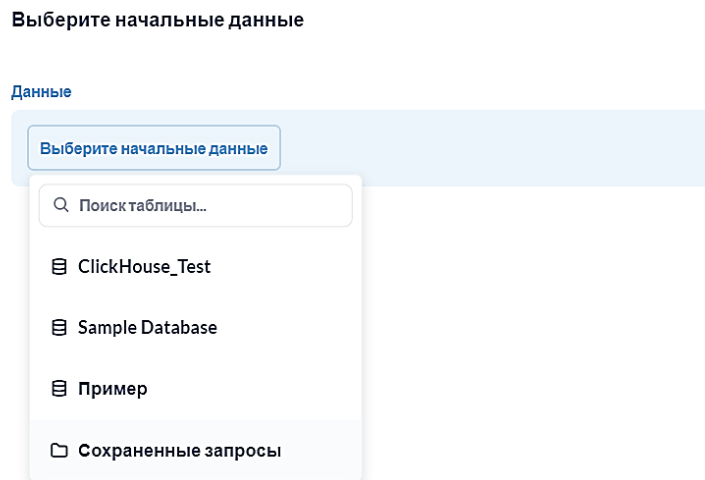


Рис. 1.22. Выбор данных для запроса

После выбора таблицы с данными приступим к настройкам запроса. Выберем метрики (рис. 1.23): сначала метрику Сумма для показателя Gross Sales, а затем добавим метрику Среднее (рис. 1.24) для того же самого показателя Gross Sales (рис. 1.25).

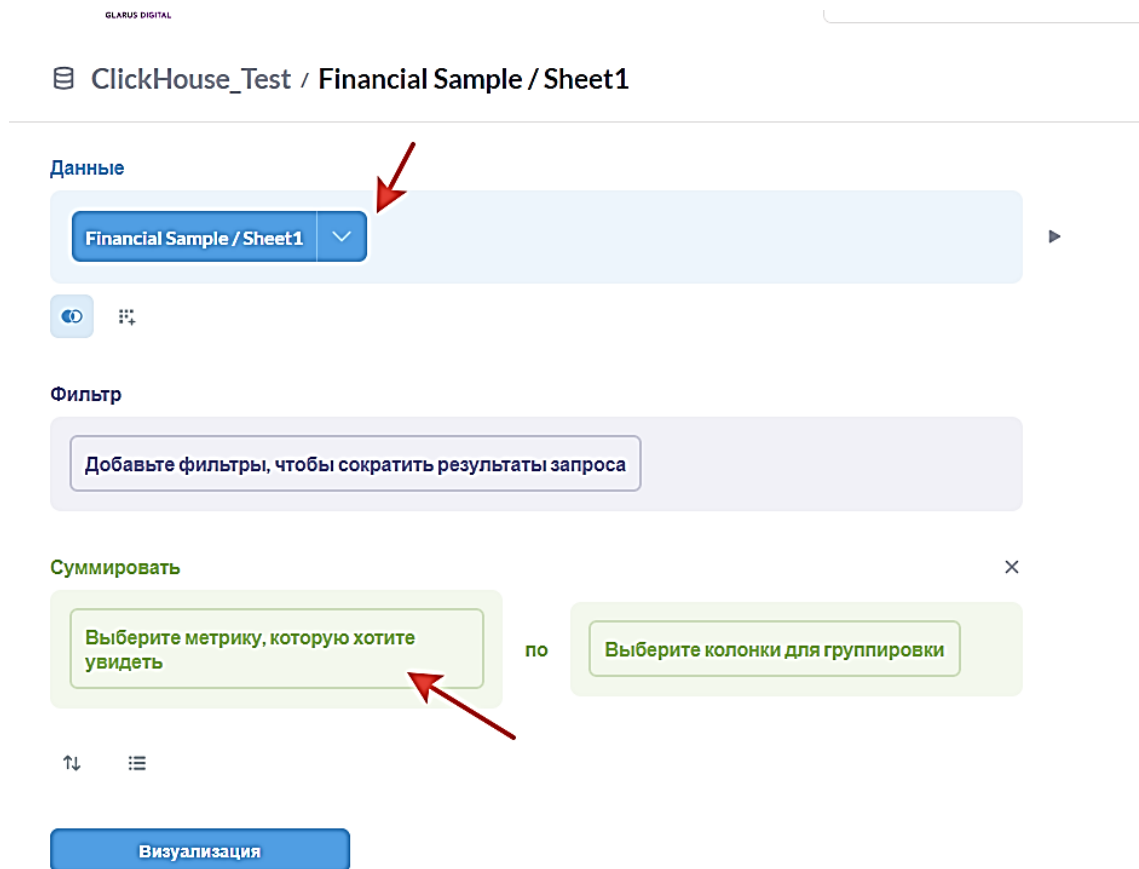


Рис. 1.23. Выбор метрики и показателя для агрегации

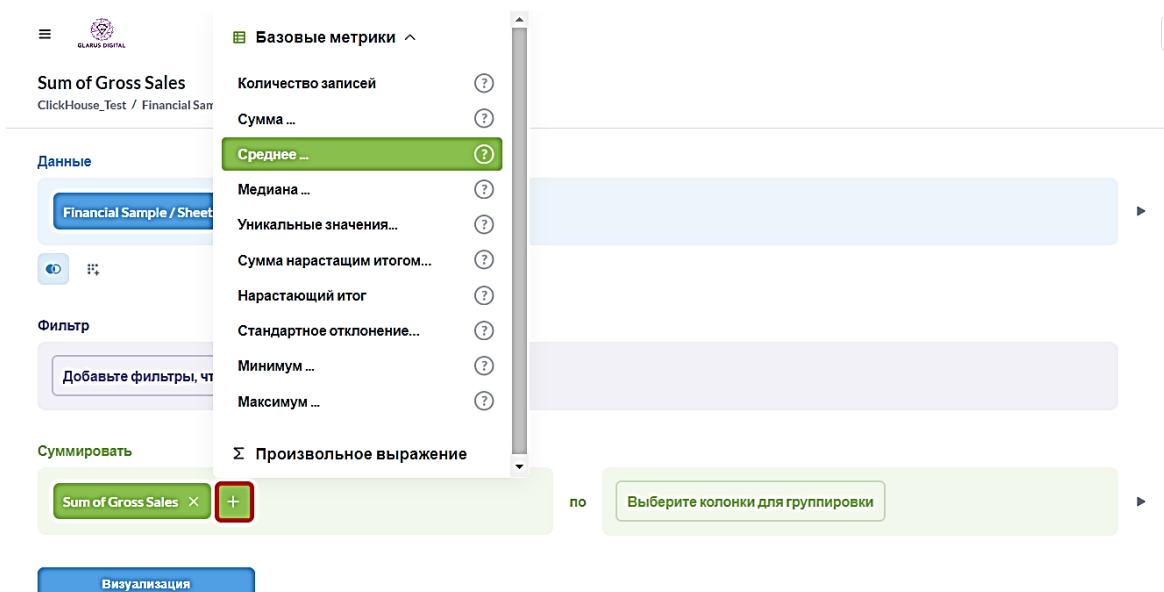


Рис. 1.24. Добавление новой метрики

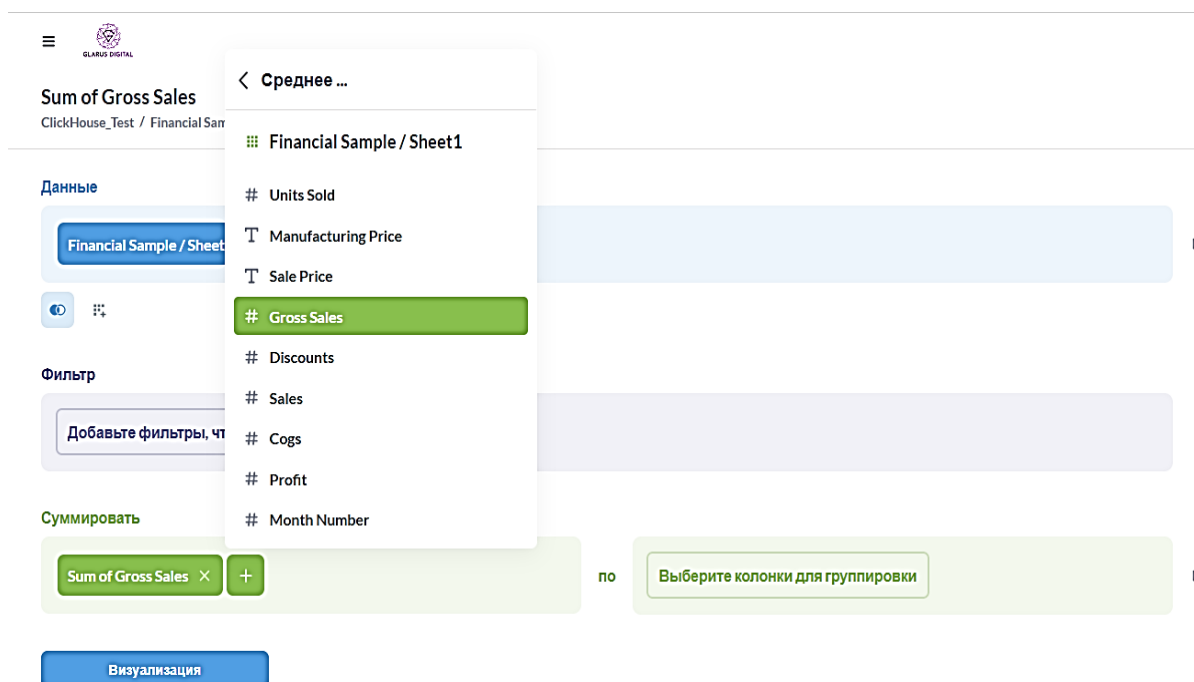


Рис. 1.25. Выбор показателя для агрегации

В качестве группирующего показателя выберем поле Date (рис. 1.26) и выведем на просмотр полученный запрос (рис. 1.27).

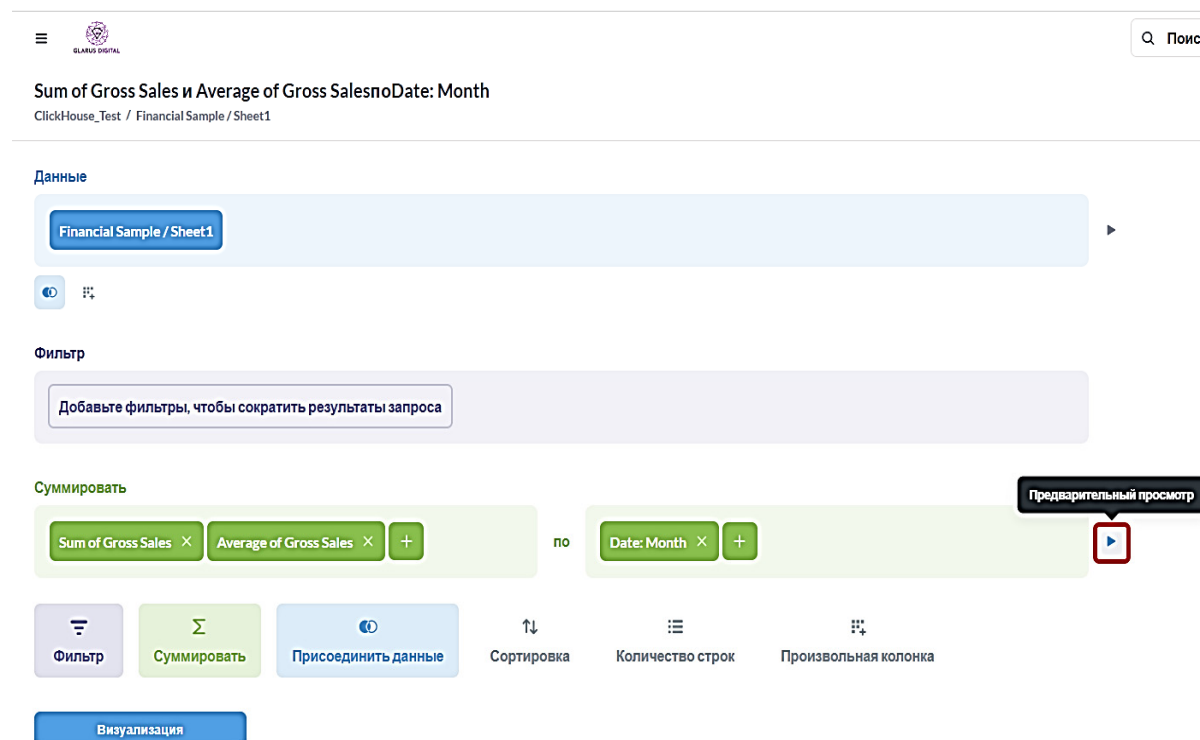


Рис. 1.26. Выбор показателя для группировки

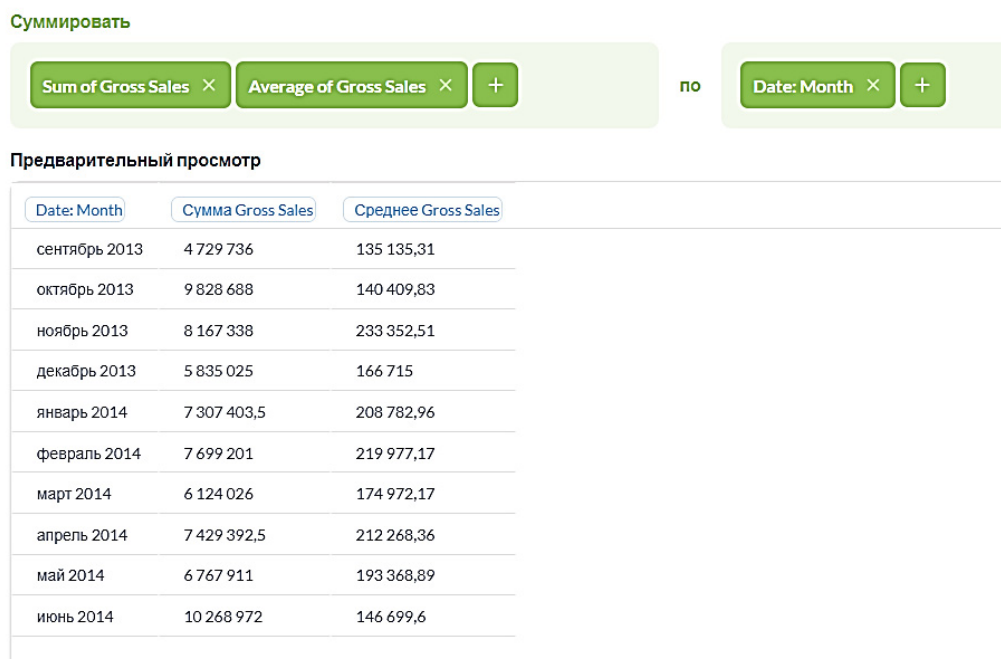


Рис. 1.27. Просмотр результатов настроек запроса

Построим визуализацию для этого запроса (рис. 1.28). На диаграмме две оси: слева – по метрике Сумма Gross Sales, а справа – по метрике Среднее Gross Sales. При необходимости можно вызвать настройки и поменять цвет линий или подписи осей и т.п. Также можно переключаться между режимами таблицы данных, соответствующей запросу, и его визуализацией.

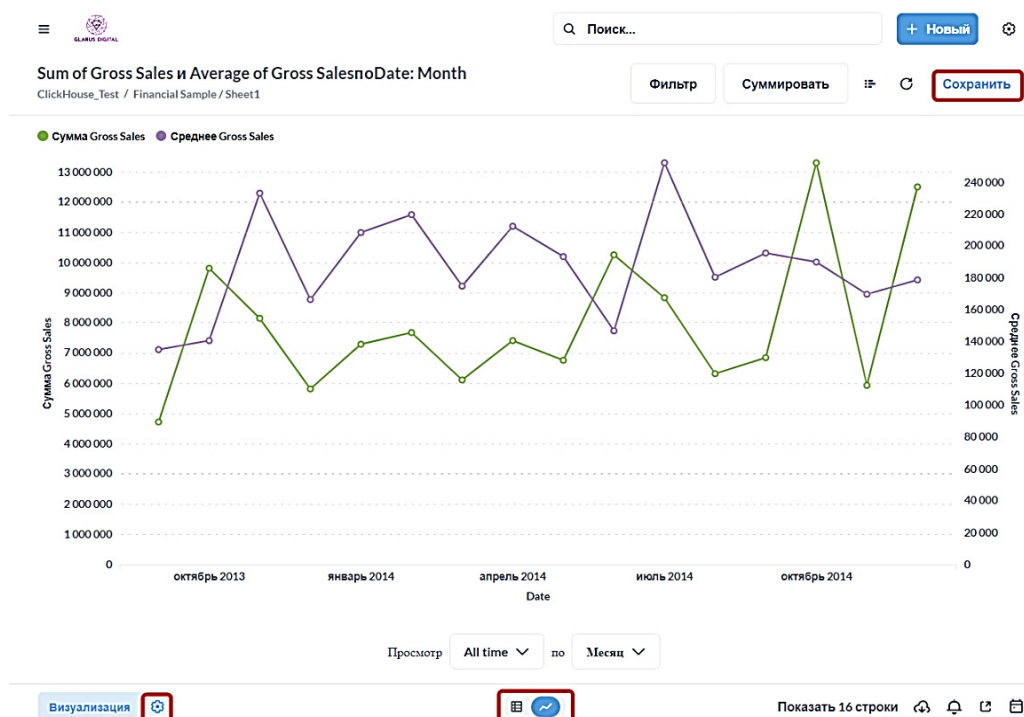


Рис. 1.28. Визуализация по запросу

При сохранении можно не изменять автоматически данное запросу название, добавив только описание (рис. 1.29).

Сохранить новый запрос

Имя

Financial Sample / Sheet1, Sum of Gross Sales и Average of Gross Sales, Group

Описание

Агрегация по дате среднего и суммарного значения показателя Gross Sales

В какую коллекцию следует это включить?

Светлана личная коллекция Рындина

Отмена Сохранить

Рис. 1.29. Сохранение запроса

Создадим аналогичный запрос, только вместо метрики Среднее для показателя Gross Sales второй метрикой выберем Сумма для показателя Discounts (рис. 1.30).

Sum of Gross Sales и Sum of DiscountsnoDate: Month

ClickHouse_Test / Financial Sample / Sheet1

Данные

Financial Sample / Sheet1

Фильтр

Добавьте фильтры, чтобы сократить результаты запроса

Суммировать

Sum of Gross Sales ×

Sum of Discounts × +

по

Date: Month × +

Визуализация

Фильтр Суммировать Присоединить данные Сортировка Количество строк Произвольная колонка

Рис. 1.30. Настройки для нового запроса

Построим визуализацию для этого запроса (рис. 1.31) и добавим фильтр по показателю Country, выбрав два значения: Germany и Canada (рис. 1.32).

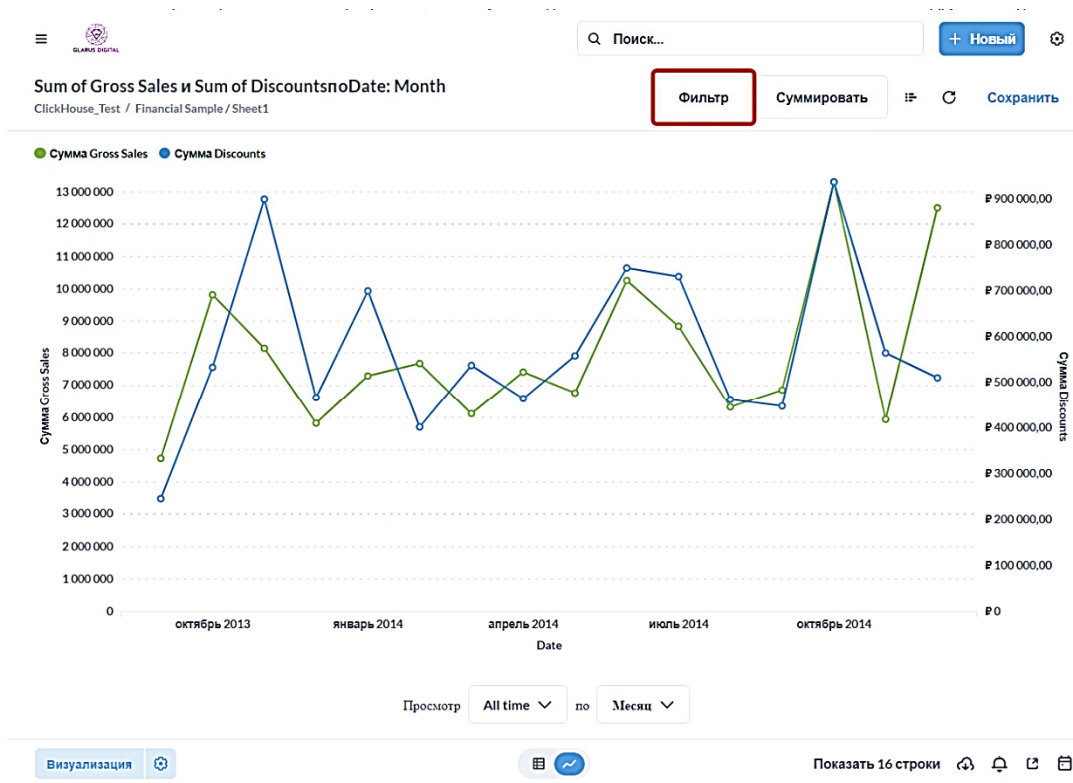


Рис. 1.31. Визуализация для запроса

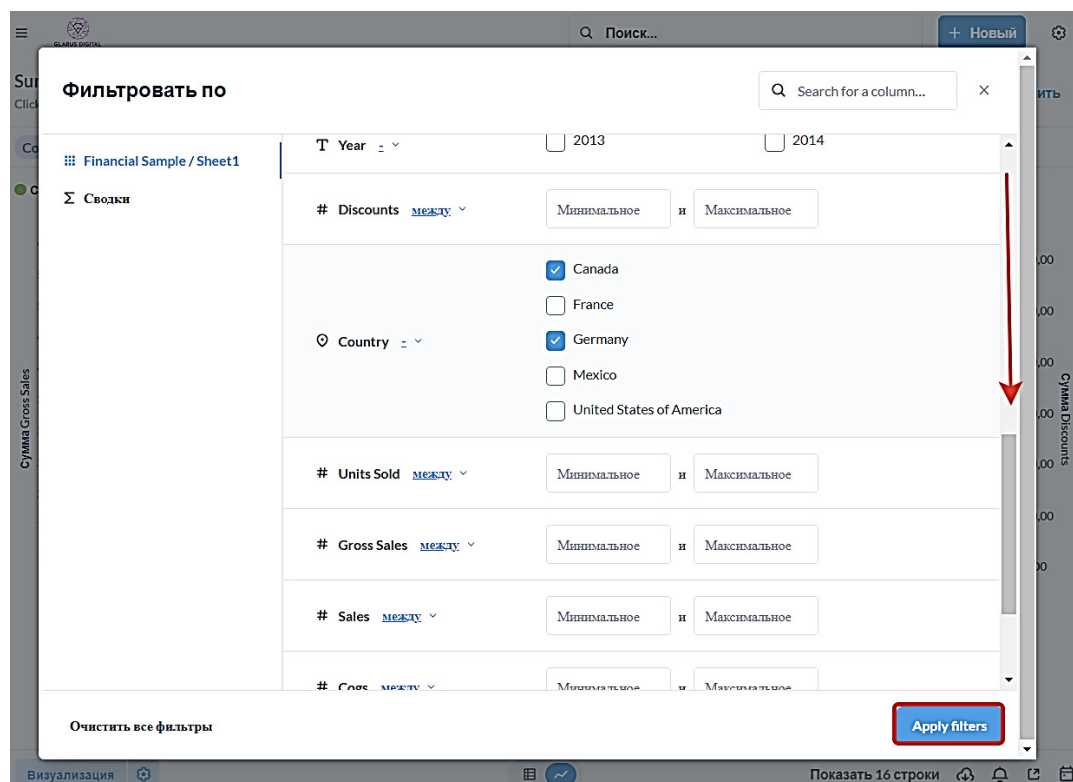


Рис. 1.32. Добавление фильтра

На визуализации отобразится применение фильтра (рис. 1.33). Сохраним получившийся запрос (рис. 1.34).

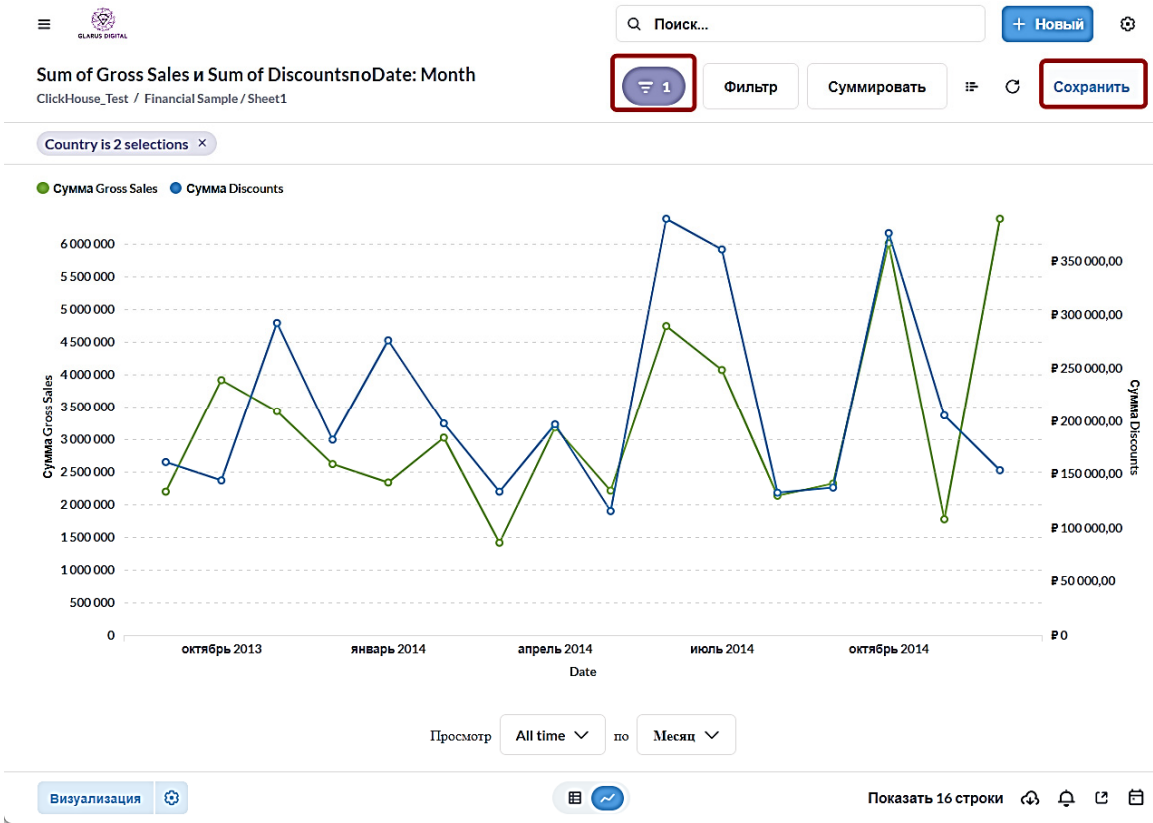


Рис. 1.33. Применение фильтра

Сохранить новый запрос

Имя

Financial Sample / Sheet1, Sum of Gross Sales и Sum of Discounts, Grouped by

Описание

Агрегация по дате двух показателей Gross Sales и Discount (метрика - сумма) с фильтром для двух стран (Германия и Канада)

В какую коллекцию следует это включить?

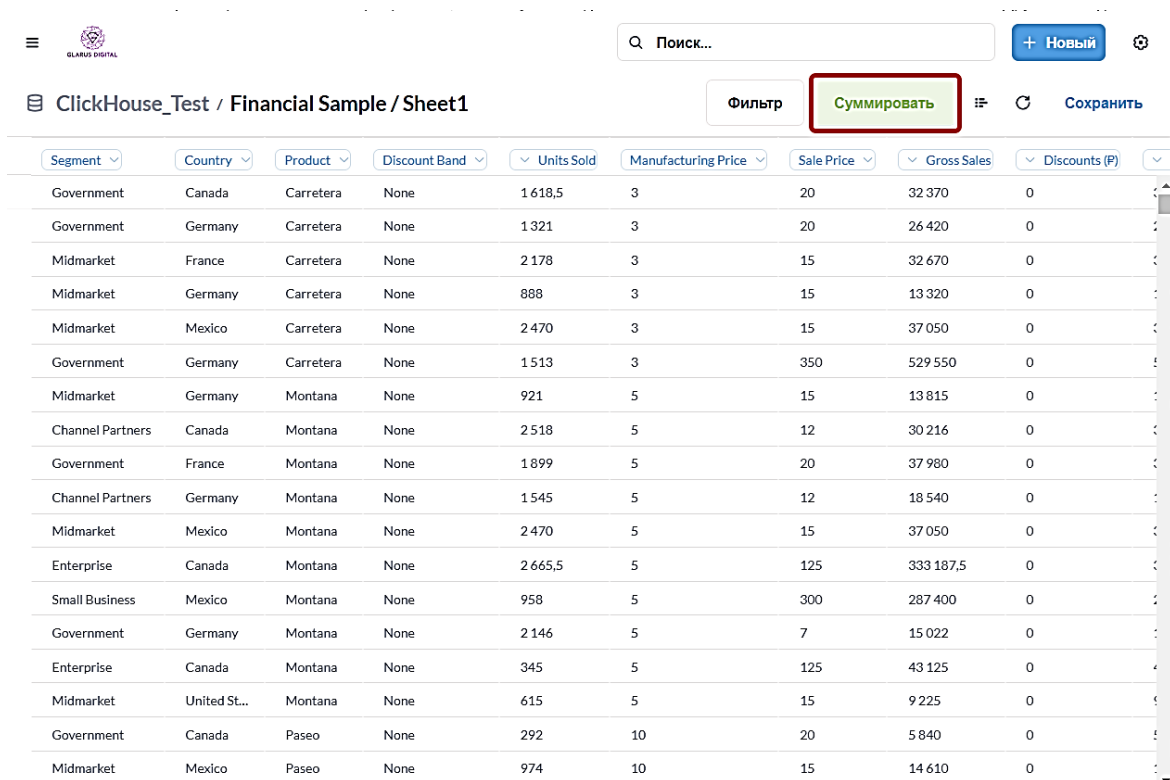
Светлана личная коллекция Рындина

Отмена

Сохранить

Рис. 1.34. Сохранение запроса

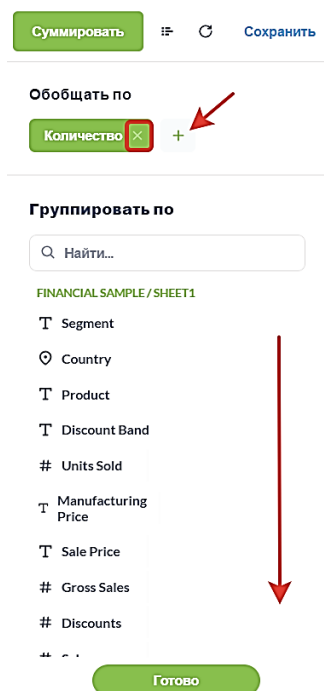
Создадим индикатор, который представляет собой суммарное значение показателя Sales. Откроем исходную таблицу данных (доступ к таблице через просмотр данных) и в верхнем правом углу воспользуемся кнопкой Суммировать (рис. 1.35).



Segment	Country	Product	Discount Band	Units Sold	Manufacturing Price	Sale Price	Gross Sales	Discounts (P)
Government	Canada	Carretera	None	1 618,5	3	20	32 370	0
Government	Germany	Carretera	None	1 321	3	20	26 420	0
Midmarket	France	Carretera	None	2 178	3	15	32 670	0
Midmarket	Germany	Carretera	None	888	3	15	13 320	0
Midmarket	Mexico	Carretera	None	2 470	3	15	37 050	0
Government	Germany	Carretera	None	1 513	3	350	529 550	0
Midmarket	Germany	Montana	None	921	5	15	13 815	0
Channel Partners	Canada	Montana	None	2 518	5	12	30 216	0
Government	France	Montana	None	1 899	5	20	37 980	0
Channel Partners	Germany	Montana	None	1 545	5	12	18 540	0
Midmarket	Mexico	Montana	None	2 470	5	15	37 050	0
Enterprise	Canada	Montana	None	2 665,5	5	125	333 187,5	0
Small Business	Mexico	Montana	None	958	5	300	287 400	0
Government	Germany	Montana	None	2 146	5	7	15 022	0
Enterprise	Canada	Montana	None	345	5	125	43 125	0
Midmarket	United St...	Montana	None	615	5	15	9 225	0
Government	Canada	Paseo	None	292	10	20	5 840	0
Midmarket	Mexico	Paseo	None	974	10	15	14 610	0

Рис. 1.35. Просмотр набора данных

Удалим метрику Количество записей и добавим метрику Сумма для показателя Sales (рис. 1.36). В результате получим значение итоговой суммы продаж (рис. 1.37). Сохраним полученный запрос.



Суммировать

Обобщать по

Количество

Группировать по

Найти...

FINANCIAL SAMPLE / SHEET1

- T Segment
- Country
- T Product
- T Discount Band
- # Units Sold
- T Manufacturing Price
- T Sale Price
- # Gross Sales
- # Discounts

Готово

Рис. 1.36. Настройка метрики и показателя для получения индикатора

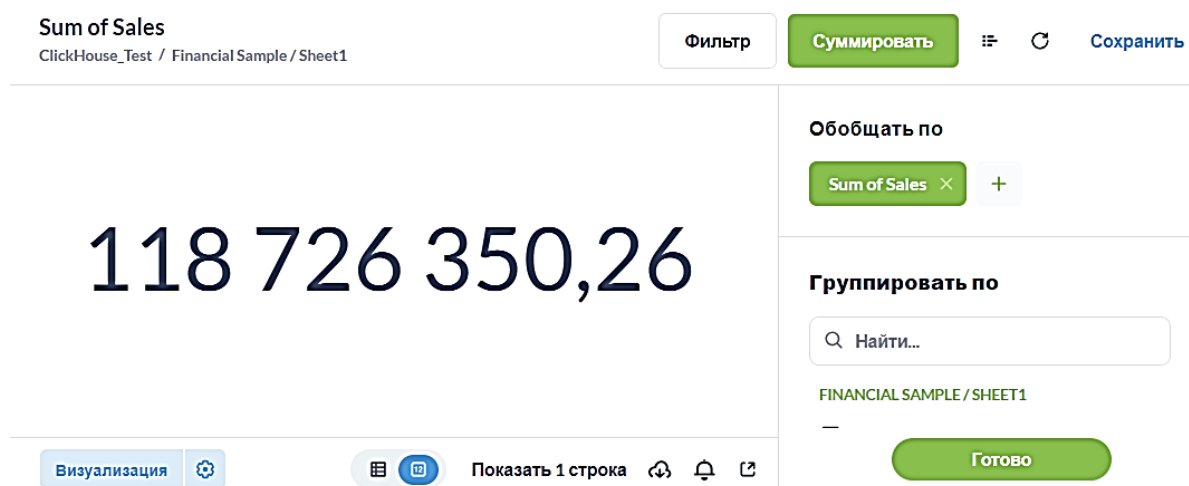


Рис. 1.37. Сумма продаж

Создадим сводную таблицу для исходного набора данных. Можно создать новый запрос, а можно перейти в таблицу через Просмотр данных и вызвать для нее редактор запросов (рис. 1.38).

ClickHouse_Test / Financial Sample / Sheet1

Фильтр Суммировать Сохранить

Поиск... + Новый

Segment	Country	Product	Discount Band	Units Sold	Manufacturing Price	Sale Price	Gross Sales
Government	Canada	Carretera	None	1 618,5	3	20	32 370
Government	Germany	Carretera	None	1 321	3	20	26 420
Midmarket	France	Carretera	None	2 178	3	15	32 670
Midmarket	Germany	Carretera	None	888	3	15	13 320
Midmarket	Mexico	Carretera	None	2 470	3	15	37 050
Government	Germany	Carretera	None	1 513	3	350	529 550
Midmarket	Germany	Montana	None	921	5	15	13 815

Визуализация

Показать 700 строки

Рис. 1.38. Переход к редактору запросов

По исходным данным через вызов списка визуализаций построить сводную таблицу невозможно. Она строится только для запросов, в которых настроена агрегация. Добавим метрики и числовые показатели, которые будут агрегироваться, а также показатели для группировки (рис. 1.39)

Данные

Financial Sample / Sheet1



Суммировать

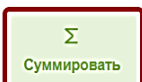
Sum of Units Sold × Sum of Gross Sales × Sum of Cogs ×
Sum of Discounts × Sum of Sales × Sum of Profit × +

по

Segment × Country × +



Фильтр



Суммировать



Присоединить данные



Сортировка



Количество строк



Произвольная колонка

Визуализация

Рис. 1.39. Настройки для сводной таблицы

Перейдем к визуализации, в которой с помощью настроек можно изменить порядок следования столбцов (рис. 1.40) и, например, настроить Условное форматирование для ячеек таблицы по некоторому правилу (рис. 1.41).

6 метрик по Segment и Country
ClickHouse_Test / Financial Sample / Sheet1

Фильтр Суммировать Сохранить

Столбцы Ячейки Цвета Вид Условное форматирование

Добавить или удалить столбцы

- Country
- Segment
- Сумма Sales
- Сумма Units Sold
- Сумма Gross Sales
- Сумма Cogs
- Сумма Discounts (P)
- Сумма Profit

Country	Segment	Сумма Sales	Сумма Units Sold	Сумма Gross Sales	Сумма Cogs	Сумма Discounts (P)	Сумма Profit
Canada	Channel Partners	491 164,14	44 062	528 744	132 186	37 579,5	396 158,5
France	Channel Partners	372 090,36	33 503	402 036	100 509	29 945,5	299 526,5
Germany	Channel Partners	336 425,88	29 689	356 268	89 067	19 842,5	267 200,88
Mexico	Channel Partners	234 379,08	21 163	253 956	63 489	19 576,5	190 970,58
United States of America	Channel Partners	366 534,18	32 846,5	394 158	98 539,5	27 623,5	295 618,13
Canada	Enterprise	3 967 491,25	34 075	4 259 375	4 089 000	291 88,5	170 375,25
France	Enterprise	3 890 890,63	33 222	4 152 750	3 986 640	261 85,5	166 109,63
Germany	Enterprise	4 086 826,25	34 902,5	4 362 812,5	4 188 300	275 98,5	174 512,25
Mexico	Enterprise	3 315 881,25	28 638	3 579 750	3 436 560	263 86,5	143 189,75
United States of America	Enterprise	4 350 605	37 714,5	4 714 312,5	4 525 740	363 70,5	188 572,5
Canada	Government	10 741 236,52	98 412	11 645 357	8 482 765	904 12,5	3 162 592,02
France	Government	12 127 782,72	104 724,5	12 916 348,5	9 417 867,5	788 56,5	3 498 481,22
Germany	Government	11 452 895,94	89 080	12 011 352	8 775 720	558 45,5	3 235 631,44
Mexico	Government	9 791 599,38	91 750	10 637 844	7 752 440	846 24,5	2 885 404,38
United States of America	Government	8 390 746,11	86 707	9 192 165	6 687 295	801 41,5	2 504 869,61
Canada	Midmarket	510 213,98	37 772,5	566 587,5	377 725	56 373,5	188 862,48

Готово Визуализация

Показать 25 строк

Рис. 1.40. Изменение порядка следования столбцов

6 метрик по Segment и Country

ClickHouse_Test / Financial Sample / Sheet1

Фильтр

Суммировать

Столбцы

Ячейки

Цвета

Вид

Условное форматирование

Какие колонки должны быть затронуты?

Сумма Sales

Объект форматирования

☒ Цвет фона
 ☐ Цвет текста
 ☐ Стил шрифта

Стил форматирования

☐ Один цвет
 ☒ Диапазон цветов

Цвета

Начать диапазон с

☒ Наименьшее значение в этом столбце
 ☐ Произвольное значение

Защитить диапазон

Готово

Country

Segment

Сумма Sales

Сумма Units Sold

Сумма Gross Sales

Canada	Channel Partners	491 164,14	44 062	528 744
France	Channel Partners	372 090,36	33 503	402 036
Germany	Channel Partners	336 425,88	29 689	356 268
Mexico	Channel Partners	234 379,08	21 163	253 956
United States of America	Channel Partners	366 534,18	32 846,5	394 158
Canada	Enterprise	3 967 491,25	34 075	4 259 375
France	Enterprise	3 890 890,63	33 222	4 152 750
Germany	Enterprise	4 086 826,25	34 902,5	4 362 812,5
Mexico	Enterprise	3 315 881,25	28 638	3 579 750
United States of America	Enterprise	4 350 605	37 714,5	4 714 312,5
Canada	Government	10 741 236,52	98 412	11 645 357
France	Government	12 127 782,72	104 724,5	12 916 348,5
Germany	Government	11 452 895,94	89 080	12 011 352
Mexico	Government	9 791 599,38	91 750	10 637 844
United States of America	Government	8 390 746,11	86 707	9 192 165
Canada	Midmarket	510 213,98	37 772,5	566 587,5
France	Midmarket	593 802,08	42 926	643 890
Germany	Midmarket	301 344,75	21 599	323 985
Mexico	Midmarket	511 136,4	36 059	540 885

Визуализация

Показать

Рис. 1.41. Настройка условного форматирования для ячеек таблицы

Построенных визуализаций вполне достаточно для формирования первого дашборда.

1.4. Создание дашборда

Glarus BI содержит следующие типы виджетов:

- запросы;
- фильтры;
- текстовый блок (текст или заголовок);
- ссылки;
- кнопки.

Виджет в виде текста позволяет разместить поясняющие надписи на дашборде. Заголовки помогают структурировать виджеты внутри дашборда, отделяя группы связанных виджетов визуально. Также для дашборда можно выбрать шаблоны разбиения на секции.

Запросы можно размещать, редактировать внутри дашборда и изменять размеры.

Ссылки позволяют получить доступ к другим связанным объектам, а кнопки – это виджеты, для которых можно настроить определенные действия, сделав дашборд еще более интерактивным.

Используем кнопку Новый, чтобы создать дашборд (рис. 1.42).

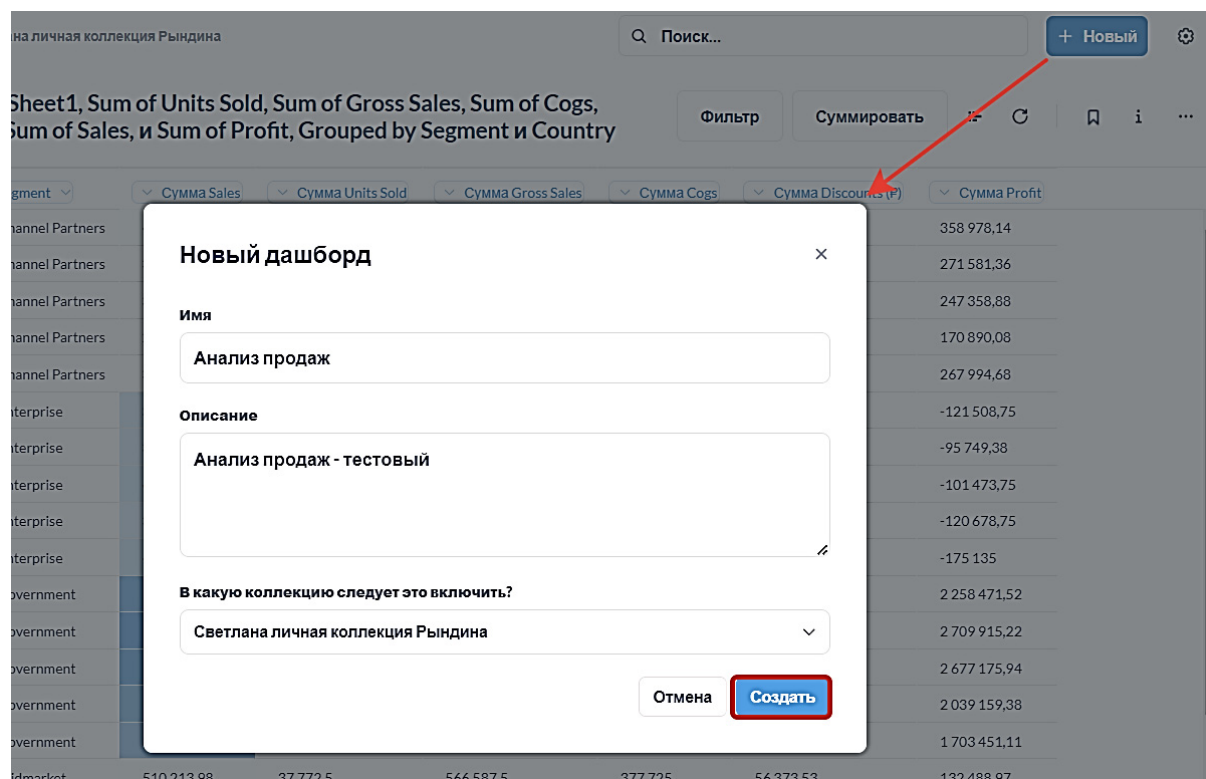


Рис. 1.42. Создание дашборда

Добавим на дашборд ранее сохраненные запросы (рис. 1.43).

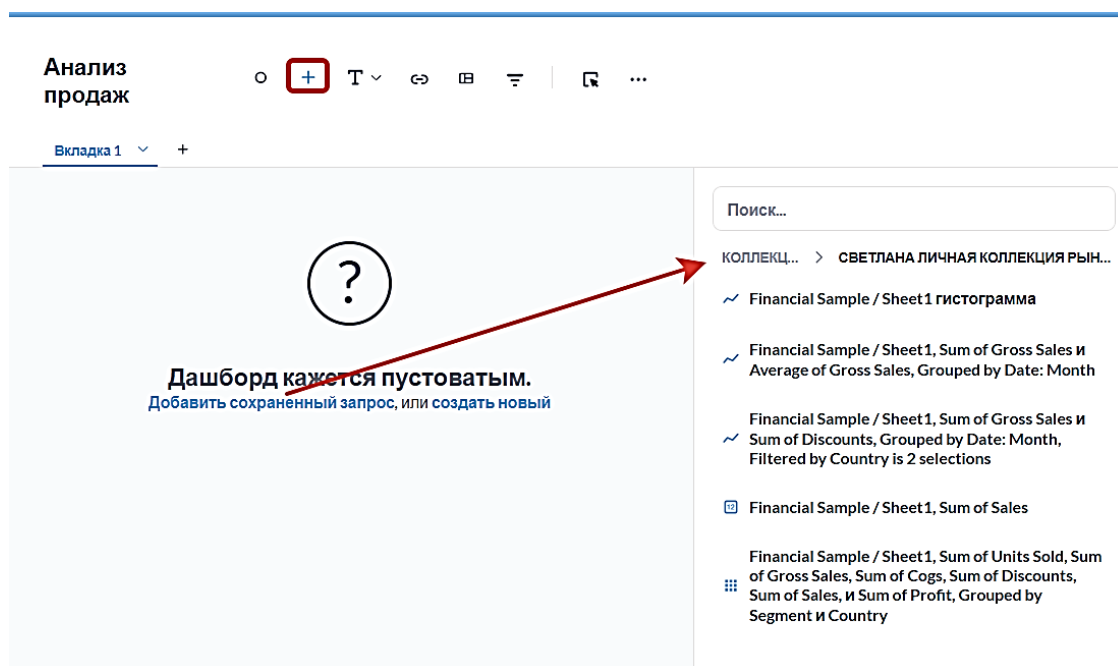


Рис. 1.43. Добавление созданных запросов на дашборд

Первым размещенным запросом на дашборде пусть станет Сводная таблица. Далее добавим фильтры (рис. 1.44) по показателям Country и Segment, которые используются в сводной таблице для группировки. Первым добавим фильтр по показателю Segment – это текст или категория.

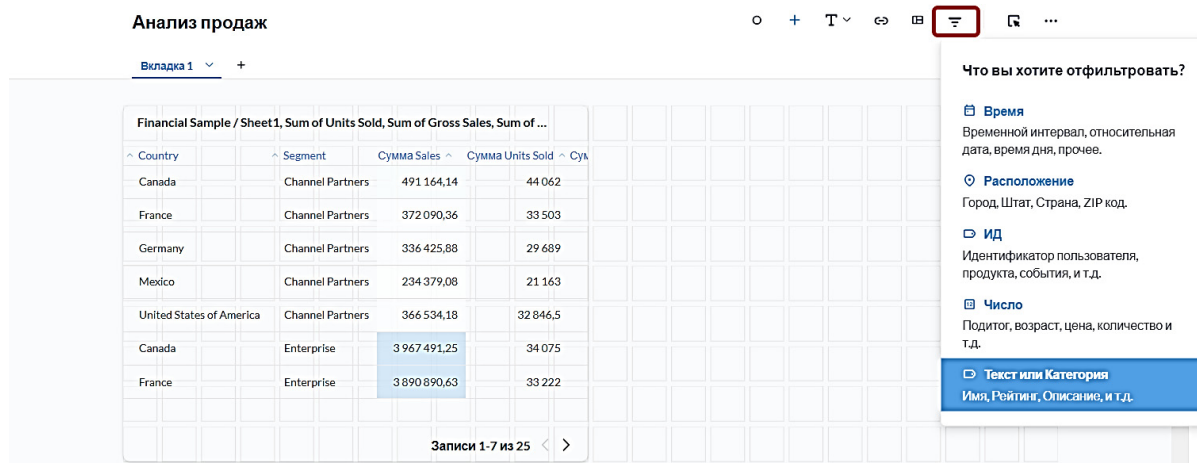


Рис. 1.44. Добавление фильтра

В качестве типа фильтра определим значения из колонки Segment (рис. 1.45).

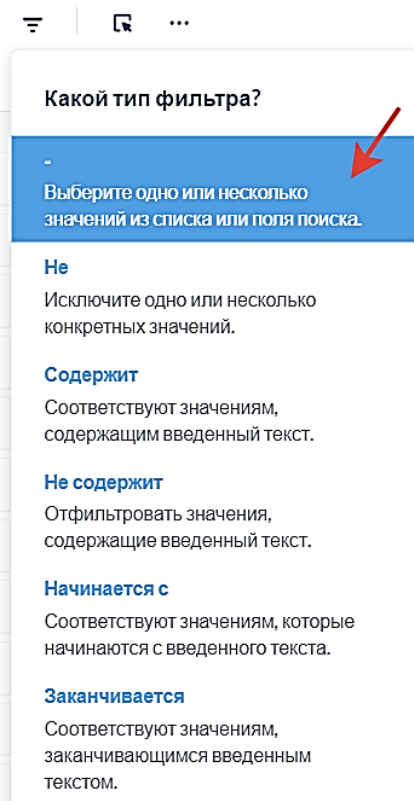


Рис. 1.45. Определение типа фильтра

В настройках фильтра (рис. 1.46) выберем столбец для фильтрации из списка (Segment) и такое же наименование внесем в качестве Ярлыка.

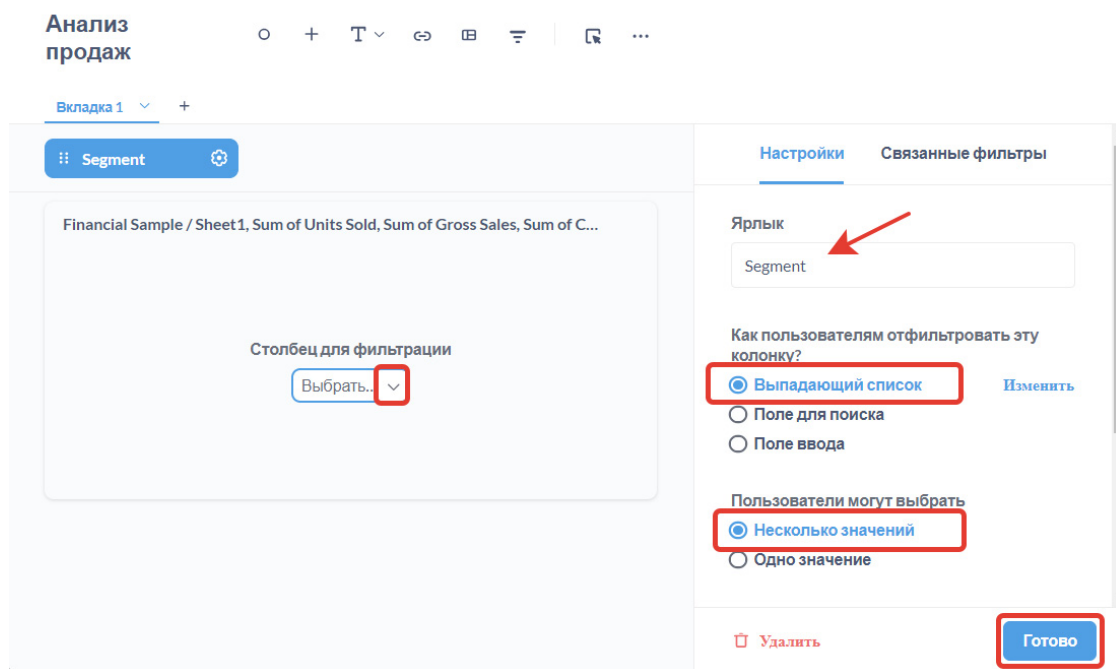


Рис. 1.46. Настройки фильтра

В режиме редактирования можно изменять как состав элементов дашборда, так и отдельные настройки этих элементов (рис. 1.47).

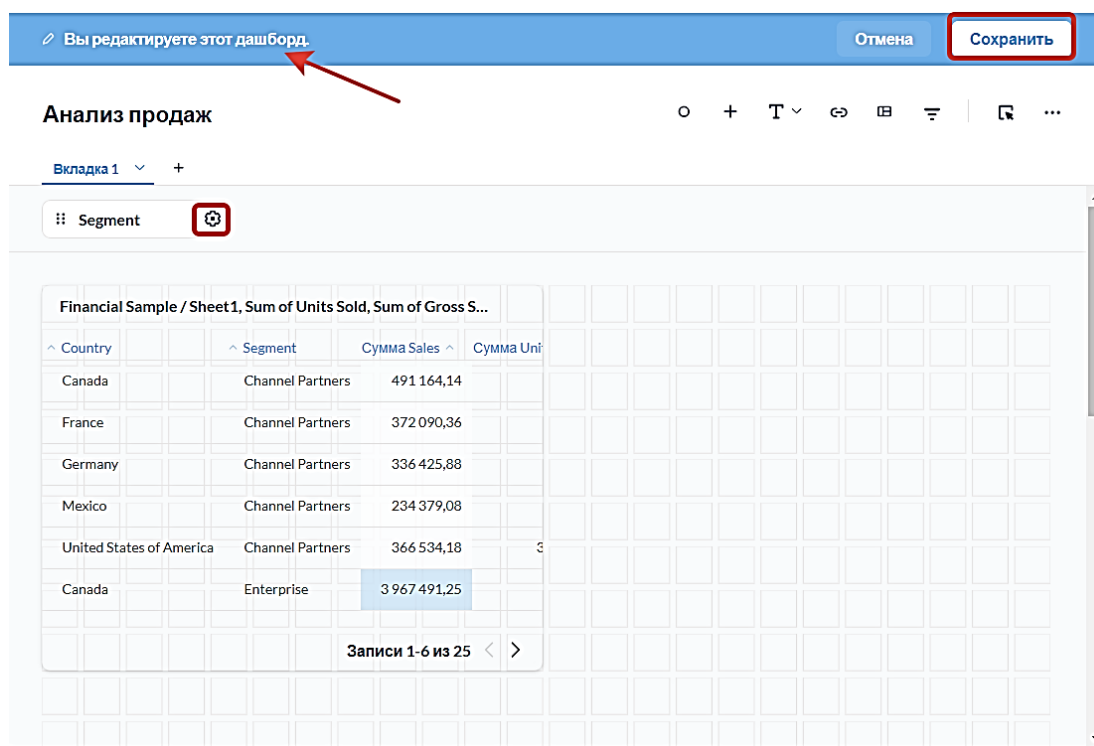


Рис. 1.47. Режим редактирования дашборда

Для перехода в режим просмотра дашборд нужно сохранить (рис. 1.48). Переход в режим редактирования осуществляется по пиктограмме  .

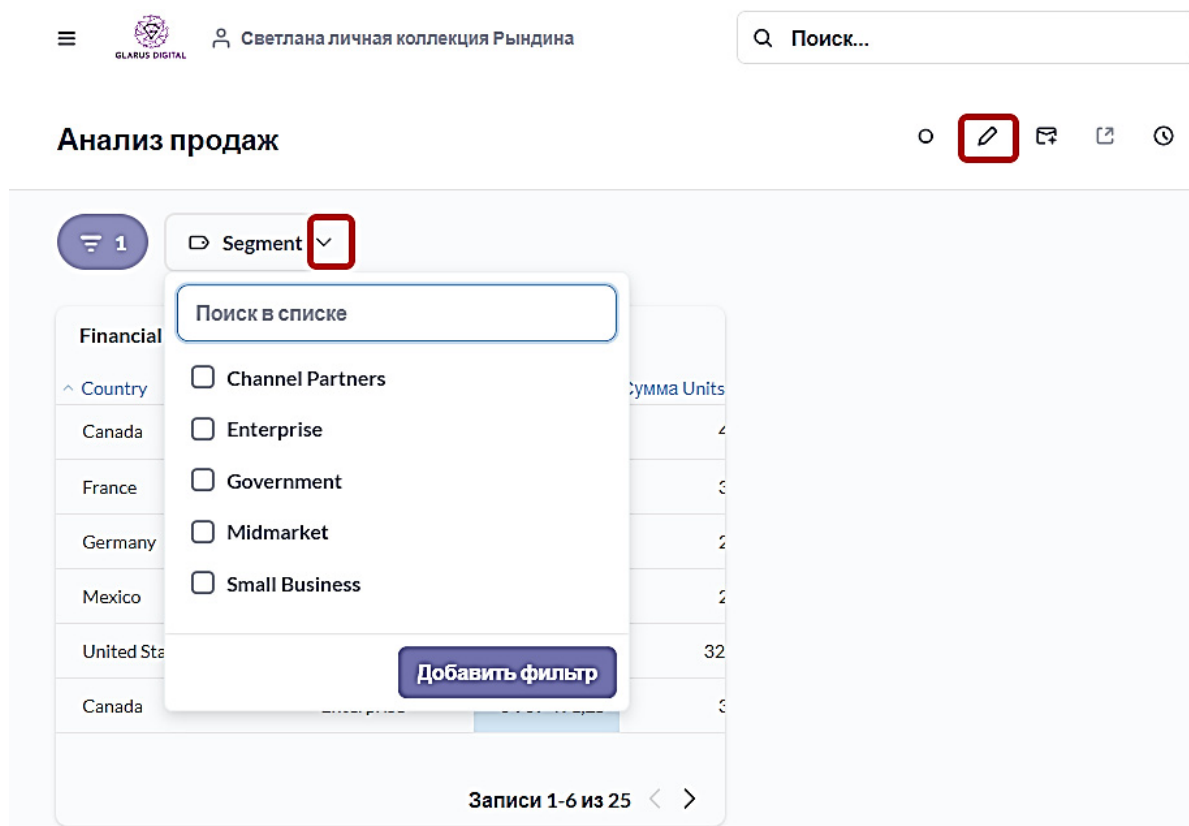


Рис. 1.48. Режим просмотра дашборда

Для добавления фильтра по показателю Country выберем расположение (см. рис. 1.44). Остальные настройки проводятся аналогично фильтру для показателя Segment. Также добавим еще два запроса на дашборд (рис. 1.49).

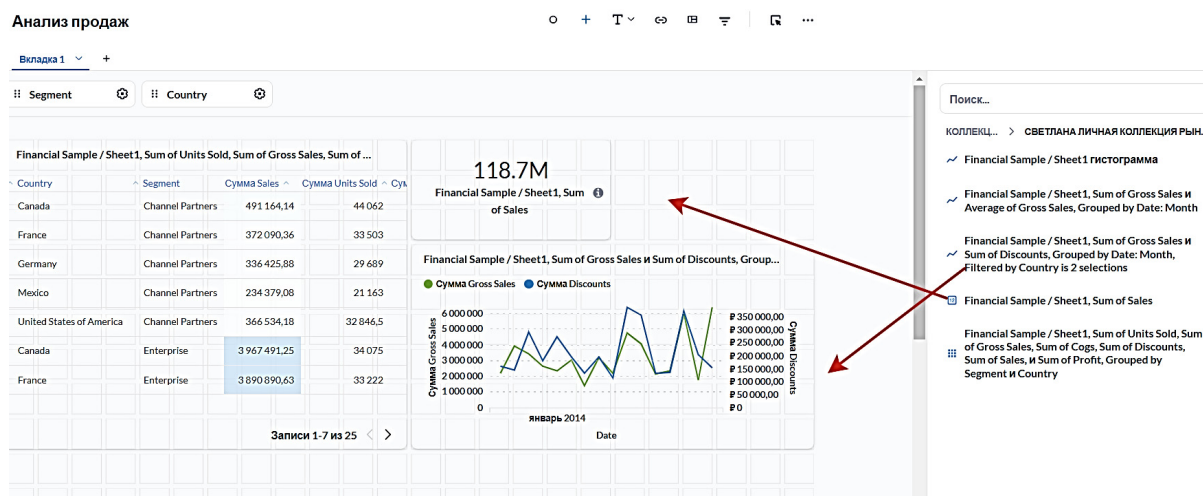


Рис. 1.49. Добавление на дашборд новых запросов

Переименуем названия запросов на дашборде, сделав их более информативными. Для этого воспользуемся значком  (рис. 1.50) в пиктографическом меню, расположенном в правом верхнем углу каждого активного запроса.

Анализ продаж

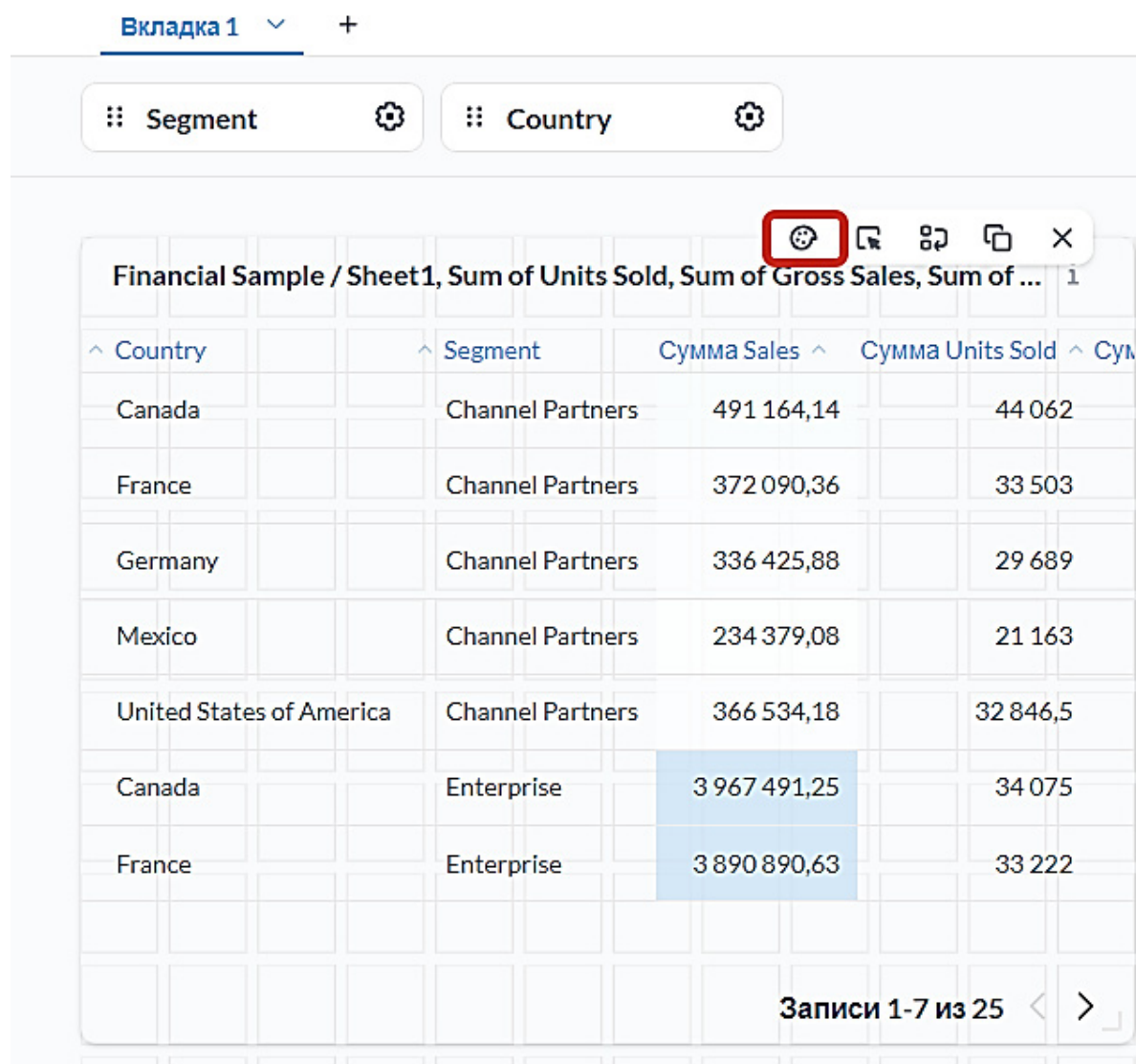


Рис. 1.50. Редактирование визуализации

Изменим заголовок (рис. 1.51) и сделаем аналогичные действия для остальных визуализаций на дашборде. Также поменяем фильтры местами. Сохраним измененный дашборд, он примет вид, показанный на рис. 1.52.

Карта

Вид

Фоновое изображение

Столбцы

Ячейки

Цвета

Условное форматирование

Цвет текста

Цвет фона карточки

Таблица с прокруткой

Заголовок

Сводная таблица по продажам в различных странах

Описание

Сводная таблица

Скрыть эту карту, если нет результатов

Financial Sample / Sheet1, Sum of Units Sold, Sum of Gross Sales, Sum of Cogs, Sum of Discounts, Sum of Sales, и Sum of Profit, Groupe...

Country	Segment	Сумма Sales	Сумма Units Sold	Сумма Gross Sales	Сумма Cogs	Сумма Discounts (P)	Сумма P
Canada	Channel Partners	491 164,14	44 062	528 744	132 186	37 579,86	358
France	Channel Partners	372 090,36	33 503	402 036	100 509	29 945,64	271
Germany	Channel Partners	336 425,88	29 689	356 268	89 067	19 842,12	247
Mexico	Channel Partners	234 379,08	21 163	253 956	63 489	19 576,92	170
United States of America	Channel Partners	366 534,18	32 846,5	394 158	98 539,5	27 623,82	267
Canada	Enterprise	3 967 491,25	34 075	4 259 375	4 089 000	291 883,75	-121
France	Enterprise	3 890 890,63	33 222	4 152 750	3 986 640	261 859,38	-95
Germany	Enterprise	4 086 826,25	34 902,5	4 362 812,5	4 188 300	275 986,25	-101
Mexico	Enterprise	3 315 881,25	28 638	3 579 750	3 436 560	263 868,75	-120
United States of America	Enterprise	4 350 605	37 714,5	4 714 312,5	4 525 740	363 707,50	-1
Canada	Government	10 741 236,52	98 412	11 645 357	8 482 765	904 120,48	2 258
France	Government	12 127 782,72	104 724,5	12 916 348,5	9 417 867,5	788 565,78	2 709
Germany	Government	11 452 895,94	89 080	12 011 352	8 775 720	558 456,06	2 677
Mexico	Government	9 791 599,38	91 750	10 637 844	7 752 440	846 244,62	2 039

Записи 1-14 из 25

Сбросить по умолчанию

Отмена

Готово

Рис. 1.51. Изменение заголовка

Анализ продаж

2

Country

Segment

Сводная таблица по продажам в различных странах

Country	Segment	Сумма Sales	Сумма Units Sold
Canada	Channel Partners	491 164,14	44 062
France	Channel Partners	372 090,36	33 503
Germany	Channel Partners	336 425,88	29 689
Mexico	Channel Partners	234 379,08	21 163
United States of America	Channel Partners	366 534,18	32 846,5
Canada	Enterprise	3 967 491,25	34 075
France	Enterprise	3 890 890,63	33 222

Записи 1-7 из 25

118 726 350,26

Продажи

Валовой объем продаж и скидки

Сумма Gross Sales

Сумма Discounts

Рис. 1.52. Сохраненный дашборд

Создадим новый запрос, в котором суммируем показатель Profit по показателю Segment (рис. 1.53). И воспользуемся сортировкой по показателю Sum of Profit (рис. 1.54). Для того чтобы сортировка изменила направление, нужно щелкнуть по левой стрелочке рядом с показателем сортировки.

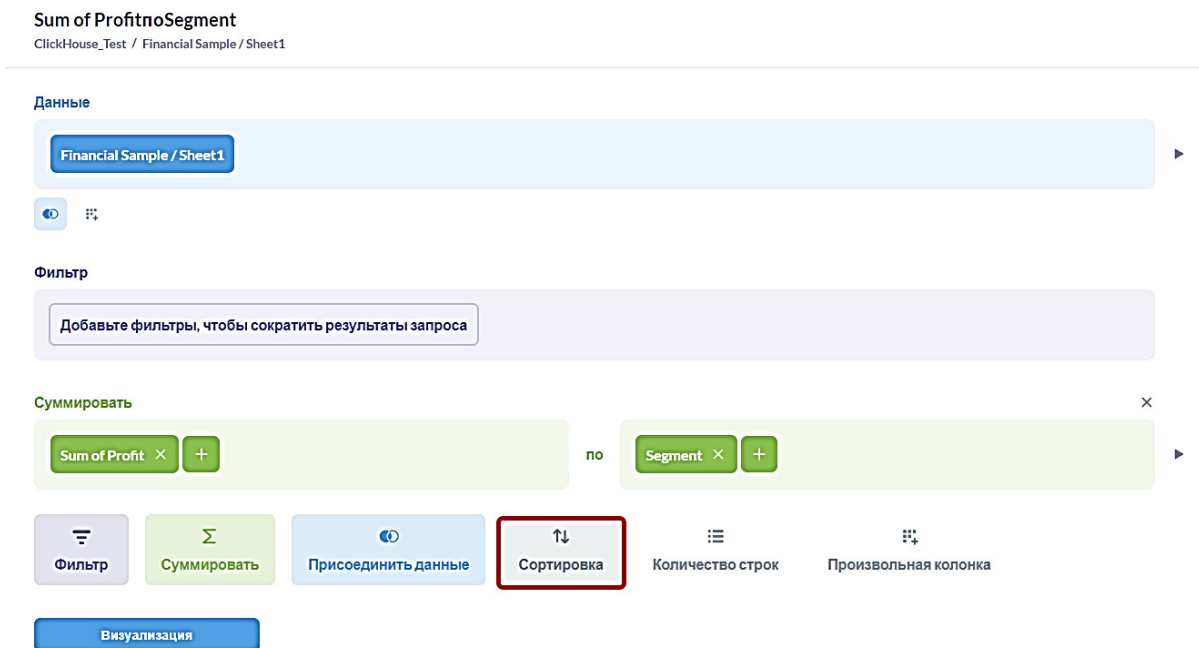


Рис. 1.53. Новый запрос

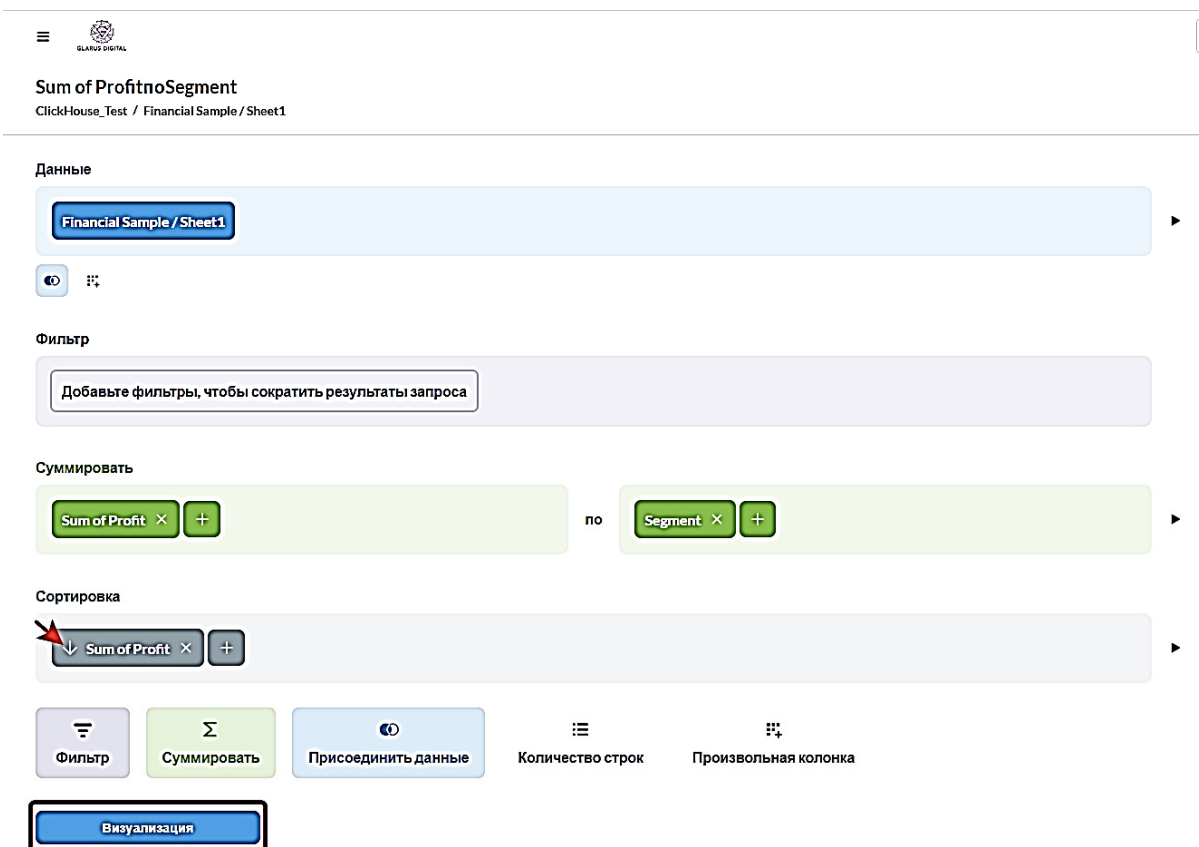


Рис. 1.54. Настройка сортировки

В качестве визуализации будет построена гистограмма.

При сохранении запроса выберем опцию Добавить на дашборд (рис. 1.55), а в качестве дашборда для размещения запроса будет выбран единственный доступный Анализ продаж (рис. 1.56).

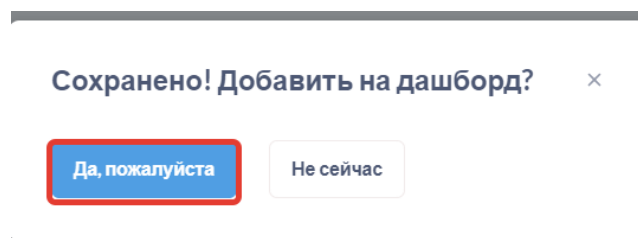


Рис. 1.55. Диалоговое окно завершения сохранения

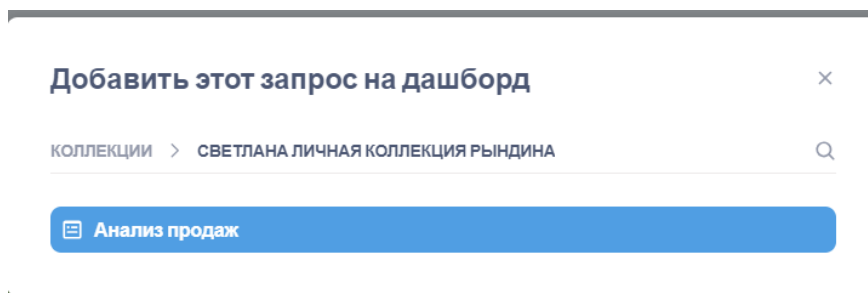


Рис. 1.56. Выбор дашборда для добавления запроса

В дашборде Анализ продаж скорректируем заголовок для только что добавленной визуализации и изменим цвет фона на серый (рис. 1.57), а также изменим расположение визуализаций (рис. 1.58).

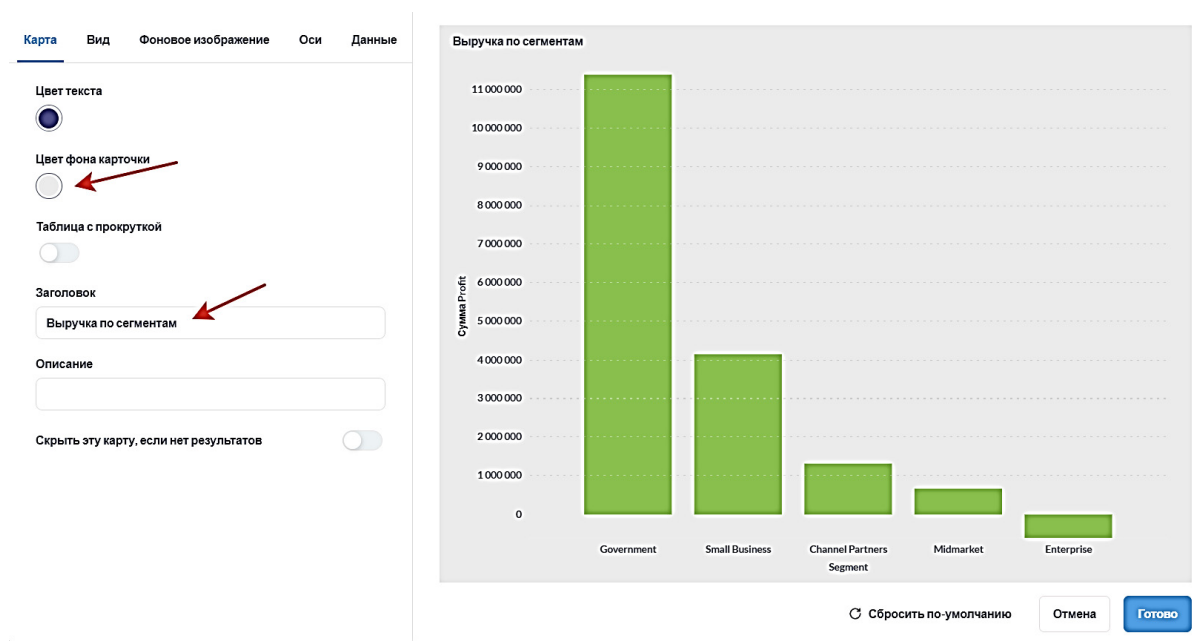


Рис. 1.57. Изменение визуализации

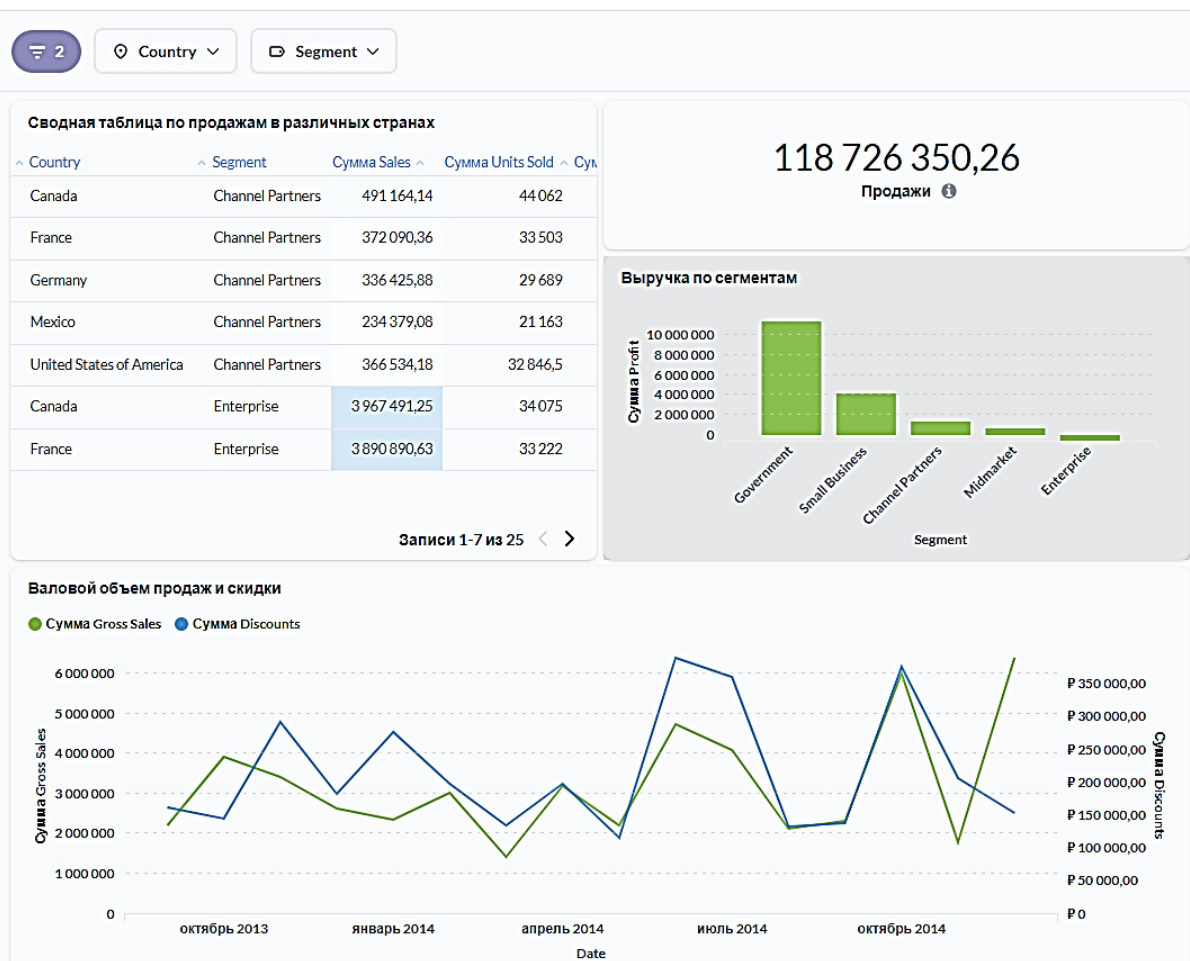


Рис. 1.58. Финальный вариант отчета

Первый базовый отчет по данным построен.

2. РАЗВЕДОЧНЫЙ АНАЛИЗ ДАННЫХ С GLARUS BI

Разведочный анализ данных (exploratory data analysis, EDA) – это предварительный анализ данных с целью выявления наиболее общих зависимостей, закономерностей и тенденций, например с помощью визуализаций.

В предыдущем разделе был построен базовый дашборд, включающий несколько вариантов визуализации по одному набору данных. Однако возможности Glarus BI для разведочного анализа данных существенно более широкие.

Рассмотрим построение более сложных, многокомпонентных визуализаций: совмещение данных из разных источников, объединение нескольких визуализаций, имеющих общее измерение, на одной диаграмме. Обогащение данных и сравнение изменения нескольких показателей по одному измерению с помощью одной визуализации позволяют осуществить более глубокое изучение имеющихся данных.

2.1. Слияние данных

Рассмотрим совмещение данных из разных источников. Создадим таблицу с двумя столбцами: Country и Population. Кроме MS Excel, можно воспользоваться, например, сервисом Google, но в качестве расширения файла при скачивании на компьютер выбрать .xlsx.

В первом столбце будут пять стран из файла Financial Sample.xlsx, а во втором столбце внесем информацию о численности населения этих стран на текущий момент (найдем эти данные с помощью поисковика). Имя файла – Population_Country.xlsx (рис. 2.1).

	A	B	C
1	Country	Population	
2	Canada	41 465 298	
3	Germany	84 358 845	
4	France	68 373 433	
5	United States of America	345 426 571	
6	Mexico	132 274 416	
7			

Рис. 2.1. Файл Population_Country

Обратите внимание, что в столбце Country все названия стран должны в точности совпадать по написанию с теми, что были в файле Financial Sample.xlsx. Если допустить орфографическую ошибку, использовать сокращение, то слияние произойдет с ошибкой.

Выполним загрузку файла в Glarus BI (рис. 2.2).

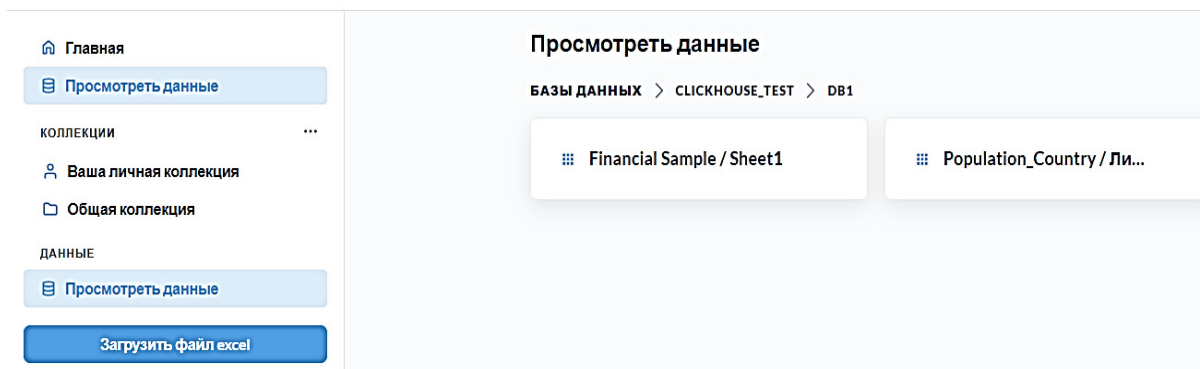


Рис. 2.2. Загрузка файла Population_Country.xlsx

Теперь составим запрос, объединяющий две этих таблицы. При выборе начальных данных должны отображаться оба набора (рис. 2.3), если это не так, нужно обновить страницу платформы.



Выберите начальные данные

Данные

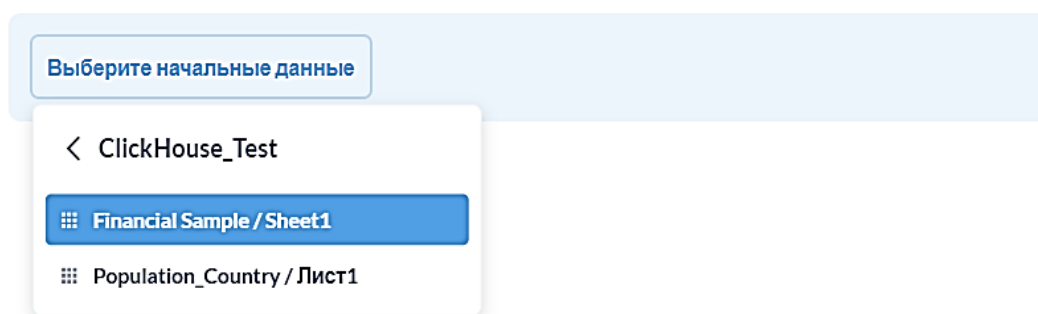


Рис. 2.3. Добавление первоначальных данных

Добавим второй набор (рис. 2.4). Тип объединения можно изменить, щелкнув по соответствующему значку. Обязательно нужно указать, между какими полями двух таблиц устанавливается связь при соединении данных. В этих полях должны быть данные одного типа с одними и теми же уникальными значениями.

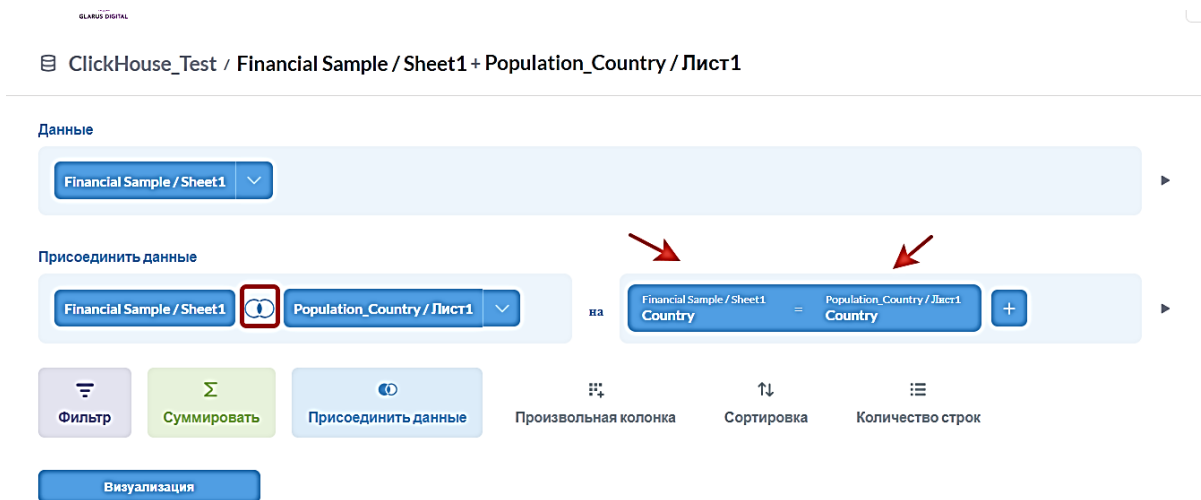


Рис. 2.4. Объединение двух наборов данных

При объединении данных в новой таблице будет продублирован показатель Country и для каждой строки в первой таблице будет добавлено значение в столбце Population, соответствующее численности населения. Поэтому сразу же настроим суммирование данных и найдем валовый объем продаж для каждой страны, чтобы сравнить его с численностью населения (рис. 2.5).

Сохраним этот запрос.

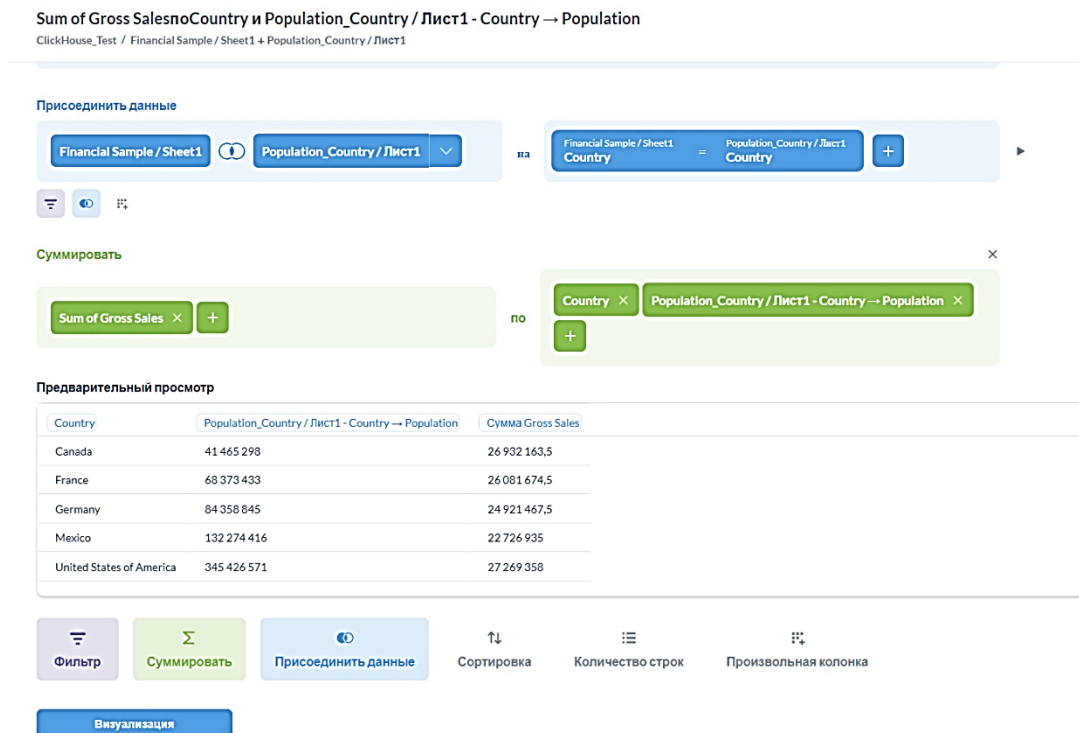


Рис. 2.5. Сводная таблица по объединенным данным двух наборов

2.2. Объединение визуализаций

Ранее были созданы визуализации (разд. 1.3), когда на одной диаграмме были представлены сразу несколько показателей (Gross Sales и Discounts) или несколько метрик для одного показателя (Сумма и Среднее для показателя Gross Sales) по одному группирующему измерению (Date:Month).

Но можно поступить иначе: объединить в одной визуализации два сохраненных запроса, имеющих общее измерение.

Создадим визуализации для объединения. Первая визуализация по данным Financial Sample: гистограмма по агрегированным данным (ось абсцисс – Sum of Gross Sales, ось ординат – Country). Настройки в редакторе запросов представлены на рис. 2.6, а визуализация – на рис. 2.7.

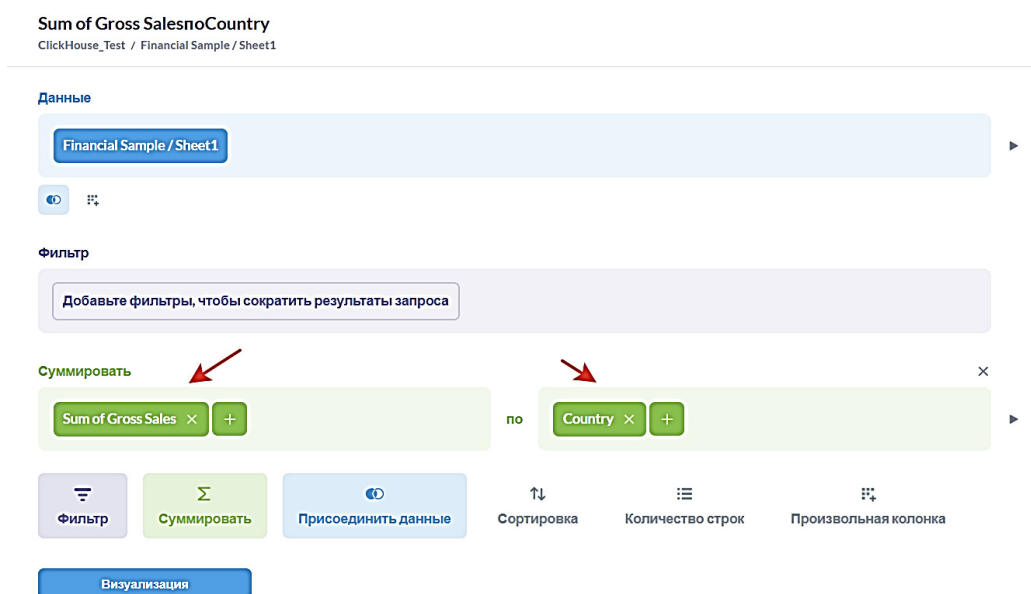


Рис. 2.6. Настройки агрегации данных

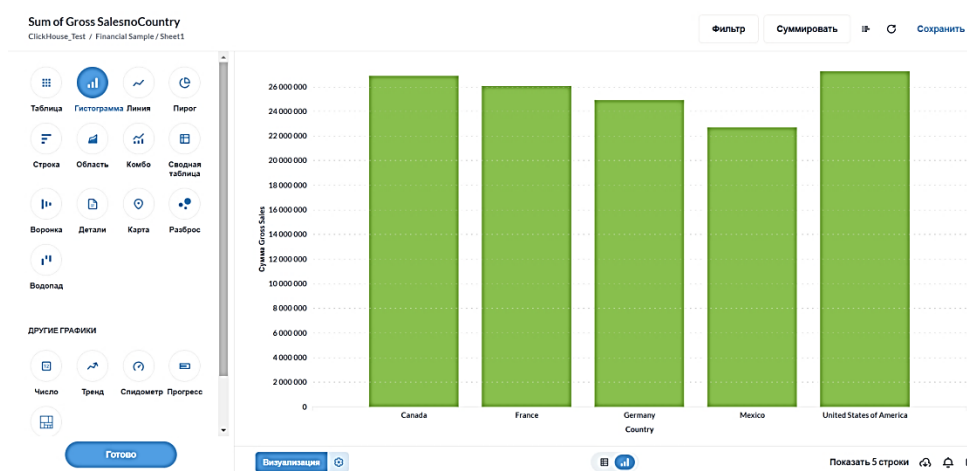


Рис. 2.7. Визуализация по запросу

Вторая визуализация будет построена по исходным данным Population: линейная диаграмма с осью абсцисс – Country, осью ординат – Population (рис. 2.8).

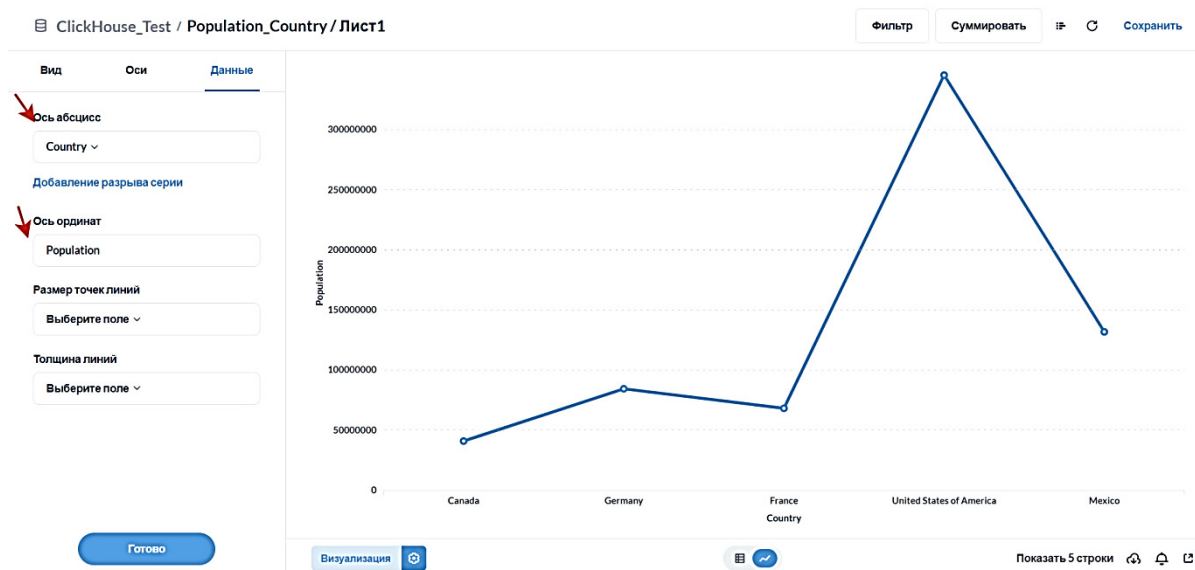


Рис. 2.8. Линейная диаграмма по исходным данным Population

Важно, что ось абсцисс и в текущей, и в предыдущей визуализации имеет одно и то же измерение Country, иначе совместить эти диаграммы в одной визуализации в дальнейшем не получится.

Теперь добавим обе диаграммы на дашборд Анализ данных. Для этого перейдем в открытом дашборде в режим редактирования и добавим новую вкладку (рис. 2.9). Добавление вкладок происходит с помощью кнопки «+», расположенной справа от уже имеющихся вкладок.

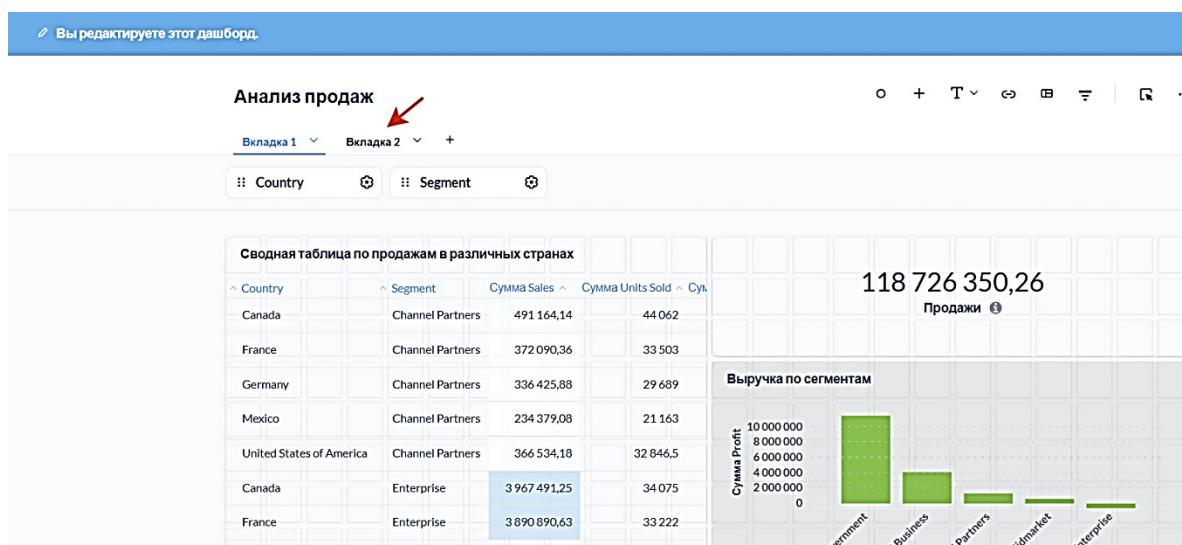


Рис. 2.9. Редактирование дашборда

В новой вкладке добавим два созданных дашборда, а также повторно первый дашборд, чтобы продемонстрировать объединение визуализаций (рис. 2.10). Добавление в уже имеющуюся на дашборде визуализацию новой диаграммы происходит с помощью кнопки **Добавить серии** .

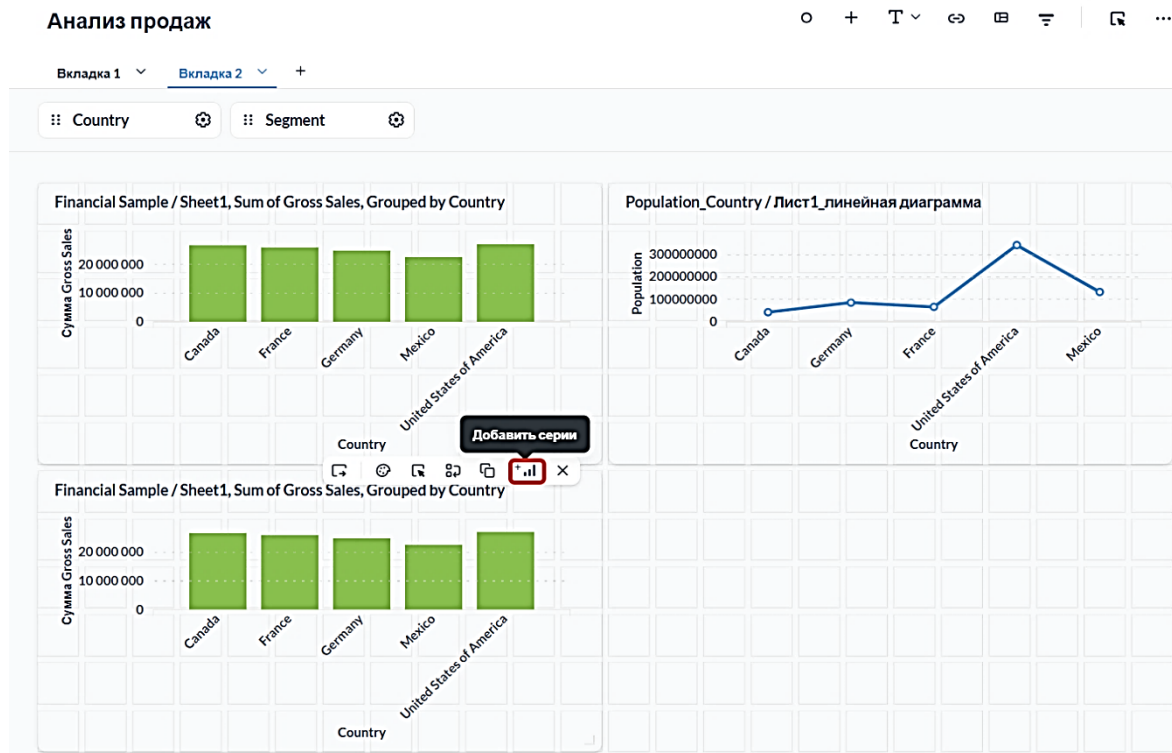


Рис. 2.10. Добавление серии

В окне редактирования визуализации появляются только те запросы, в которых измерение позволяет присоединить запрос к текущему (рис. 2.11).

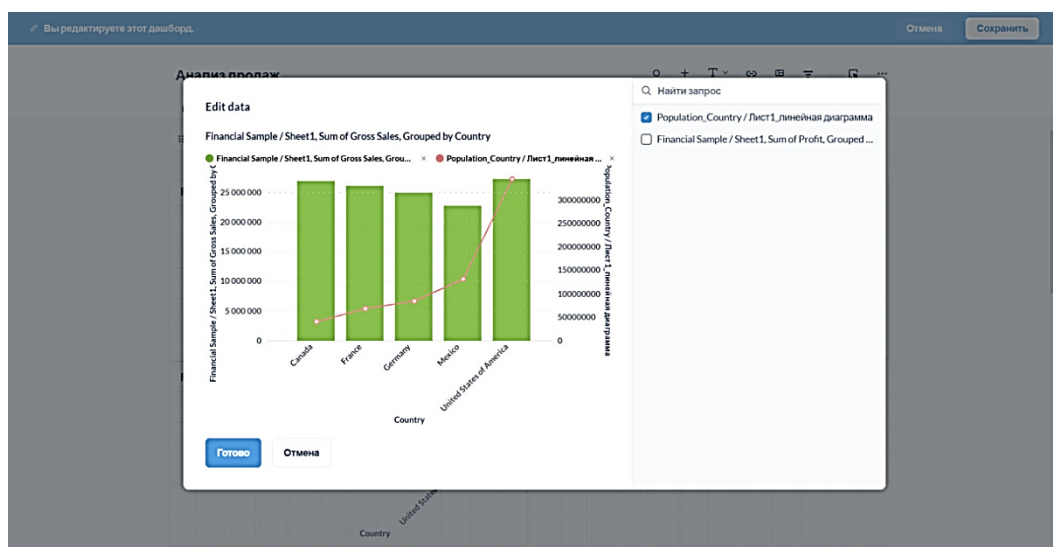


Рис. 2.11. Добавление запроса к текущему

В результате появится объединенная диаграмма, с помощью которой можно проанализировать, как соотносится численность населения с валовыми продажами продукции компании (рис. 2.12). Разумеется, два отдельных графика, даже расположенные рядом, сложнее соотносить по сравнению с визуализацией, объединяющей их на одной диаграмме.

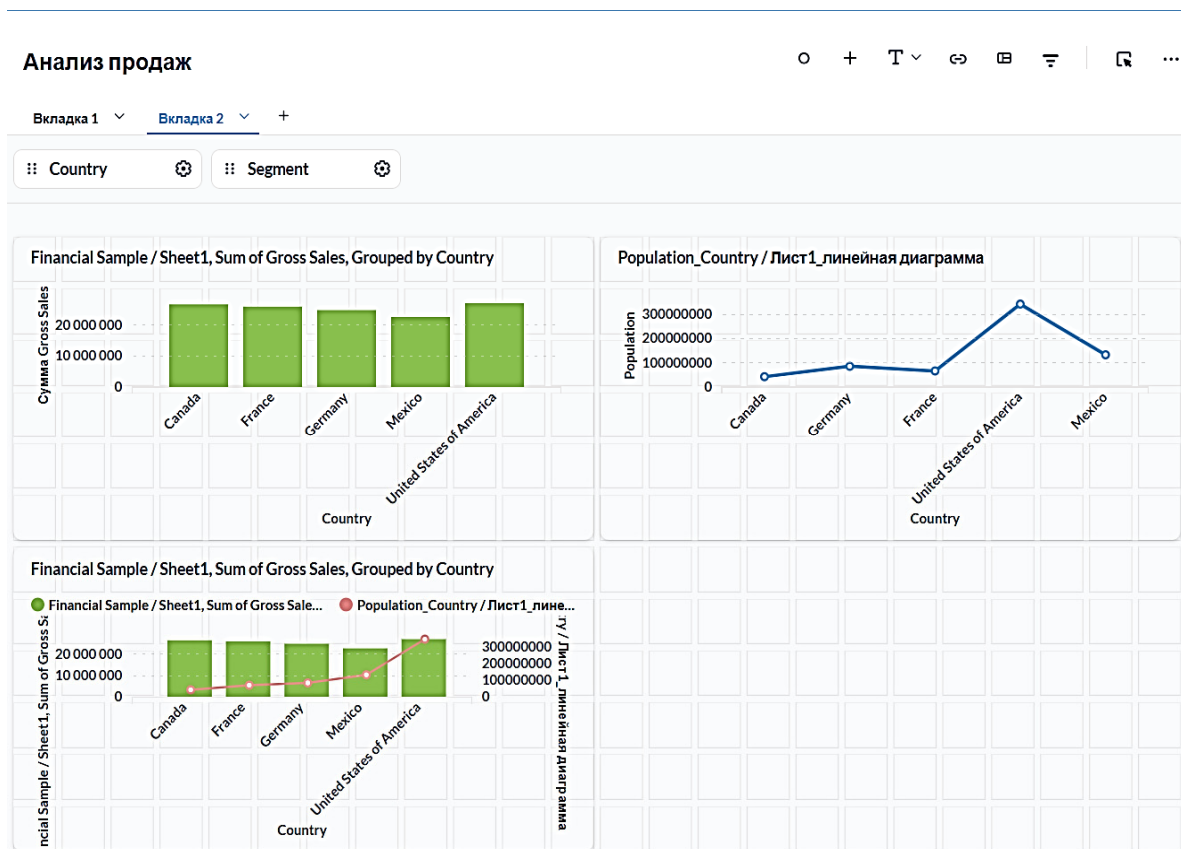


Рис. 2.12. Объединенная диаграмма

К тому же результату могли прийти с помощью запроса и настроек визуализации. Запрос для построения аналогичного графика имеет следующие настройки (рис. 2.13, 2.14).

В качестве визуализации выберем Комбо (рис. 2.15).

Настройки для визуализации показаны на рис. 2.16.

Так как диапазон значений показателей существенно различается, то изменим расположение одной из осей (рис. 2.17), переместив ее вправо (тогда у нее будут собственные настройки для диапазона). Для вызова диалогового окна редактирования одной из серий кликнем на три точки справа от имени показателя по оси ординат и переместимся на вкладку Стил

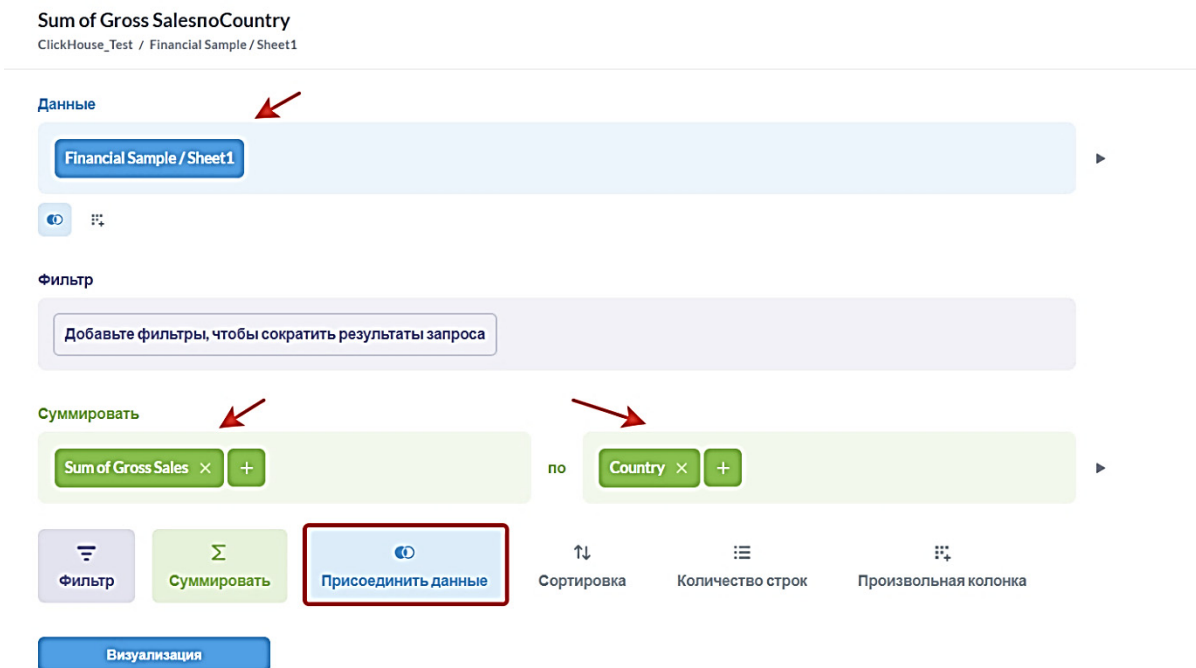


Рис. 2.13. Настройка агрегации

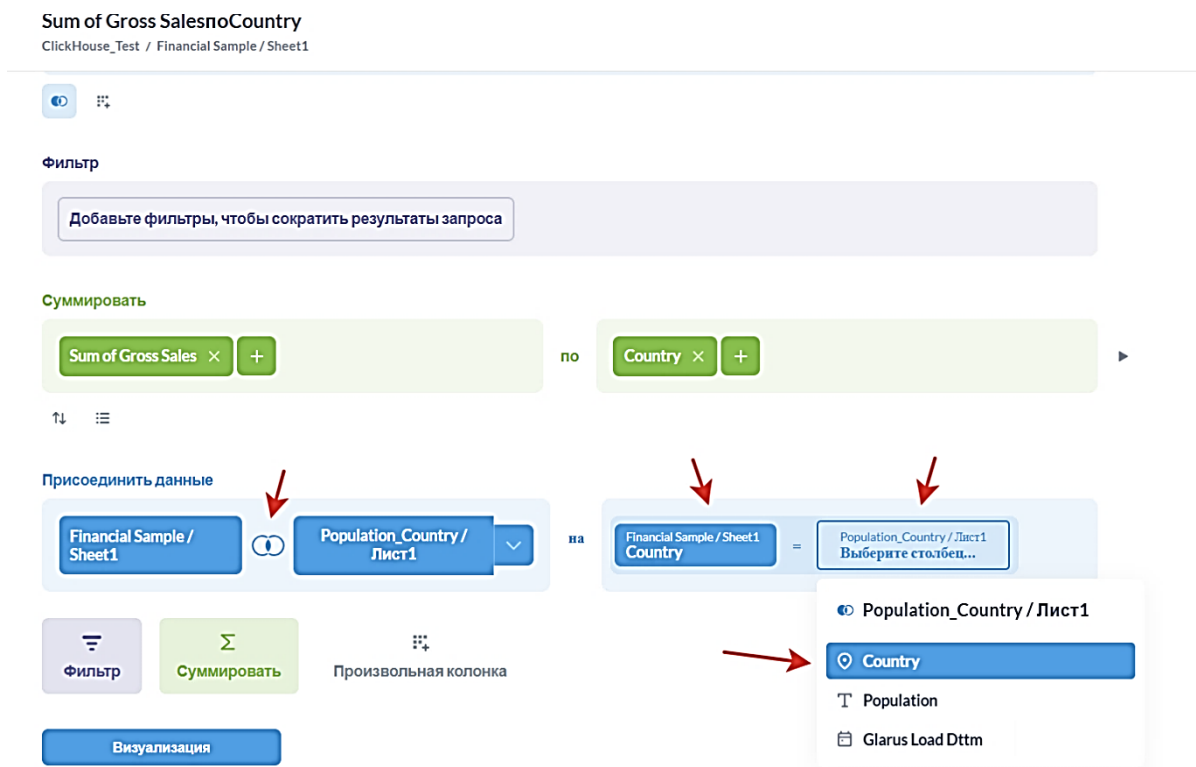


Рис. 2.14. Присоединение данных

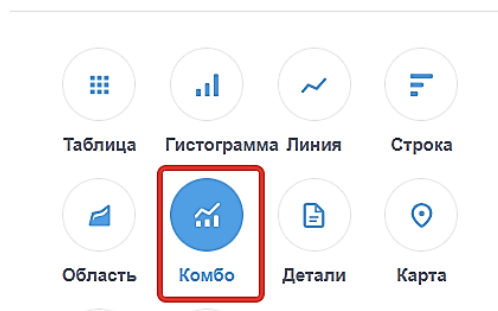


Рис. 2.15. Выбор типа визуализации

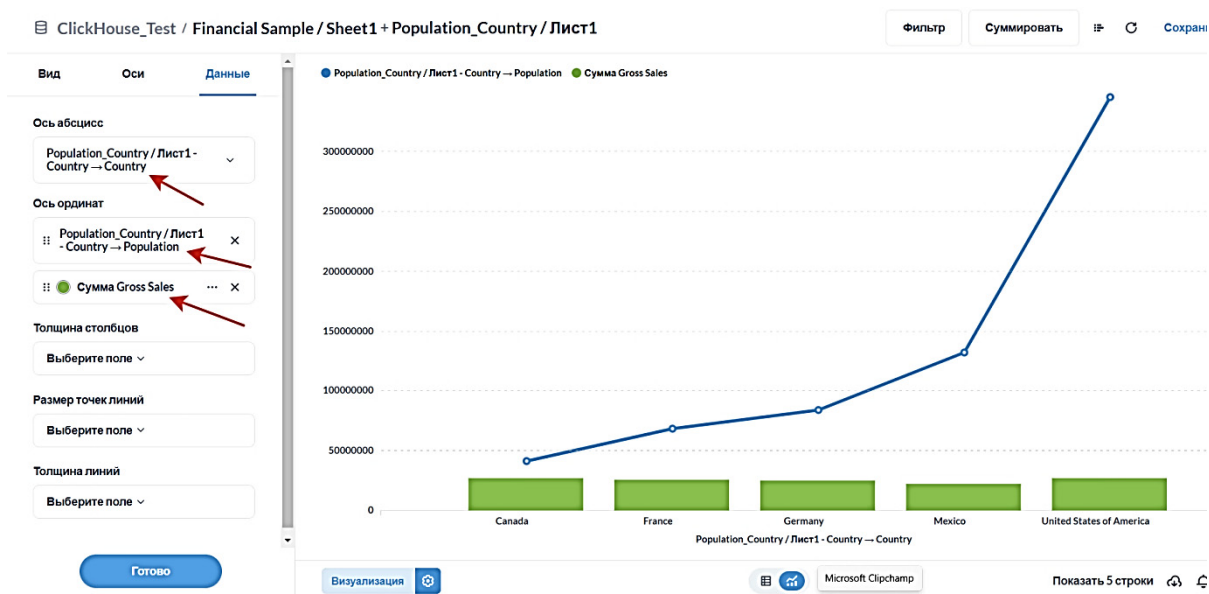


Рис. 2.16. Настройки визуализации



Рис. 2.17. Изменение расположения оси

В результате получим нужную визуализацию, повторяющую ранее построенную. Добавим ее на дашборд и проведем форматирование каждой из размещенных на Вкладке 2 визуализаций с помощью пиктограммы  (пиктографическое меню расположено справа в верхней строке каждой визуализации в активном состоянии). Изменим заголовки, а для объединенных диаграмм подписи легенды и подписи оси абсцисс, если это необходимо. Итоговый вид Вкладки 2 редактируемого дашборда показан на рис. 2.18.

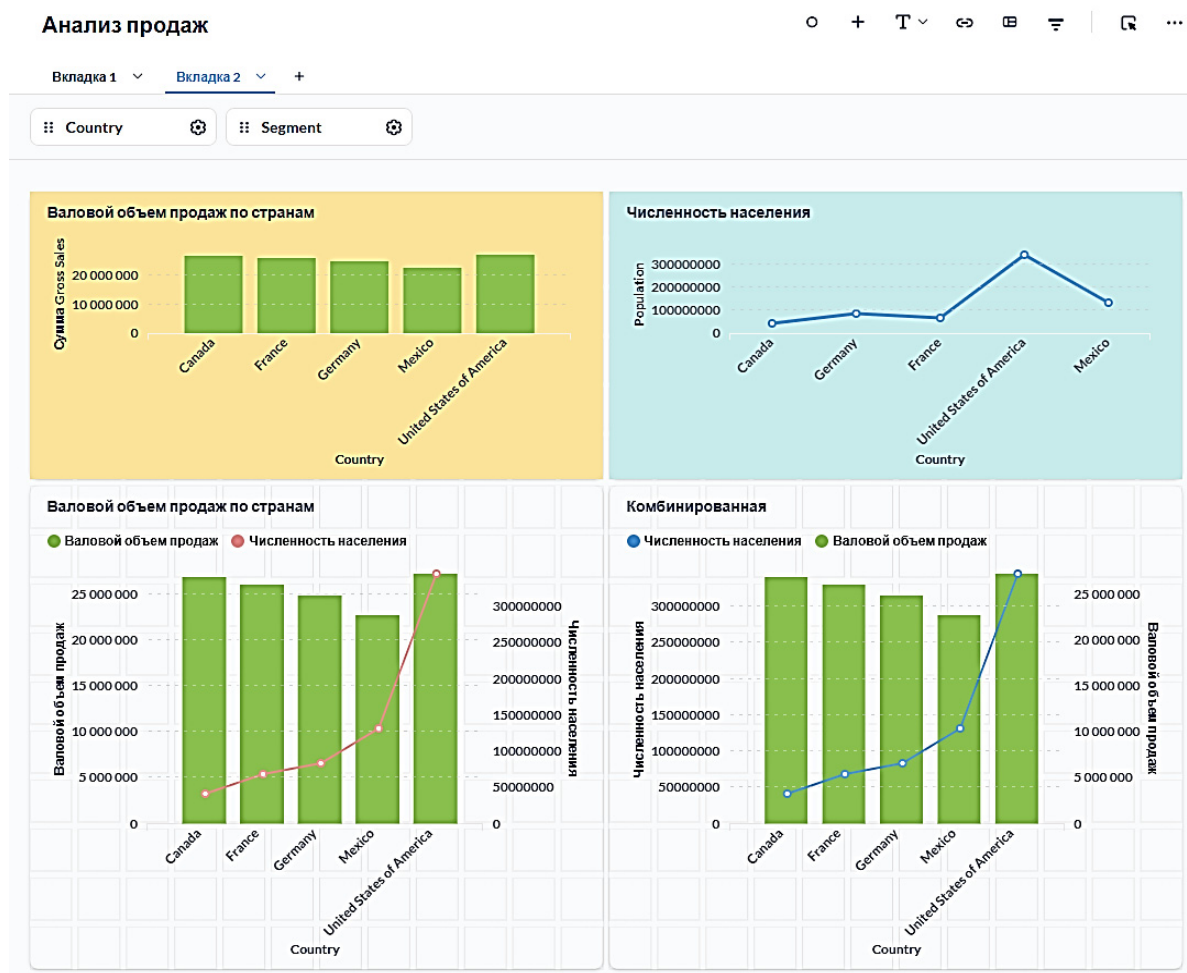


Рис. 2.18. Итоговый вид Вкладки 2

Можно поэкспериментировать и добавить на новую вкладку дашборда визуализации Карта, Область и т.п.

2.3. Разведочный анализ данных

Выполним разведочный анализ данных для набора данных insurance.csv (доступен по ссылке <https://www.kaggle.com/mirichoi0218/insurance>). На ресурсе kaggle.com для скачивания файла с данными необходимо зарегистрироваться.

Файл с именем insurance.csv представляет собой текстовый файл с данными, разделенными запятой. Такое представление данных называется плоскими таблицами. Можно просмотреть содержимое файла с помощью текстового редактора, не поддерживающего форматирование, например редактора Блокнот (рис. 2.19).

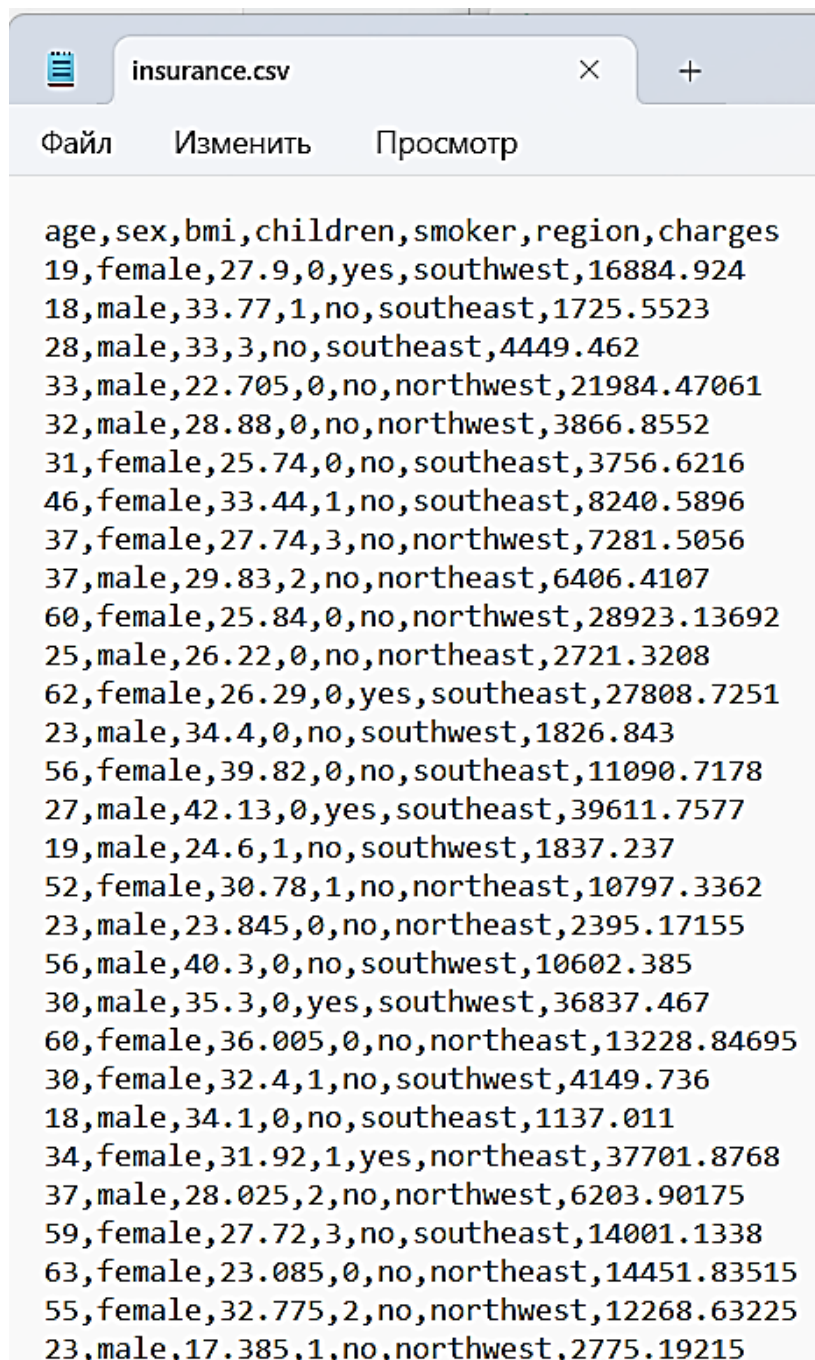


Рис. 2.19. Просмотр данных

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

Можно воспользоваться MS Excel, открыв новый файл и выполнив импорт данных из текстового файла (рис. 2.20).

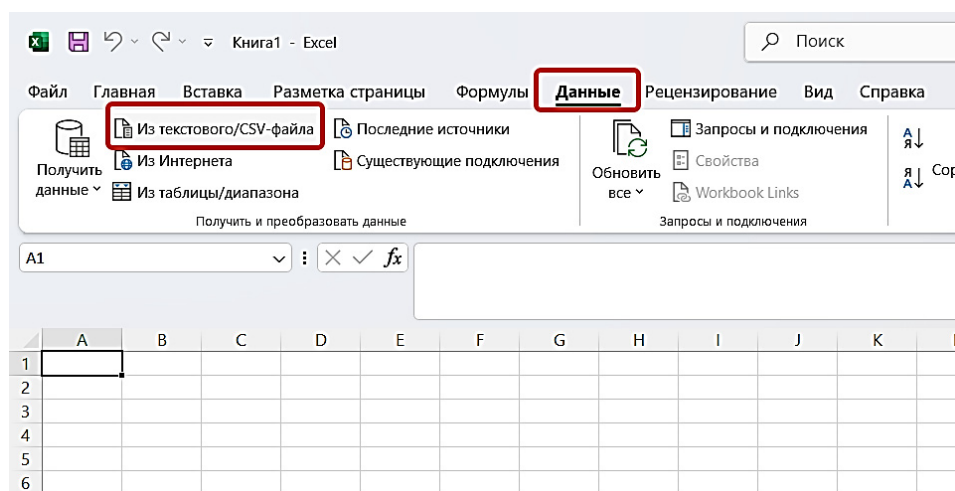


Рис. 2.20. Импорт данных из текстового файла

С помощью Мастера импорта нужно выполнить необходимые настройки и получить данные в таблице (рис. 2.21), сохранить файл в требуемом формате.

	A	B	C	D	E	F	G	H
1	age	sex	bmi	children	smoker	region	charges	
2	19	female	27.9	0	yes	southwest	16884.924	
3	18	male	33.77	1	no	southeast	1725.5523	
4	28	male	33	3	no	southeast	4449.462	
5	33	male	22.705	0	no	northwest	21984.47061	
6	32	male	28.88	0	no	northwest	3866.8552	
7	31	female	25.74	0	no	southeast	3756.6216	
8	46	female	33.44	1	no	southeast	8240.5896	
9	37	female	27.74	3	no	northwest	7281.5056	
10	37	male	29.83	2	no	northeast	6406.4107	
11	60	female	25.84	0	no	northwest	28923.13692	
12	25	male	26.22	0	no	northeast	2721.3208	
13	62	female	26.29	0	yes	southeast	27808.7251	
14	23	male	34.4	0	no	southwest	1826.843	
15	56	female	39.82	0	no	southeast	11090.7178	
16	27	male	42.13	0	yes	southeast	39611.7577	
17	19	male	24.6	1	no	southwest	1837.237	
18	52	female	30.78	1	no	northeast	10797.3362	
19	23	male	23.845	0	no	northeast	2395.17155	
20	56	male	40.3	0	no	southwest	10602.385	
21	30	male	35.3	0	yes	southwest	36837.467	
22	60	female	36.005	0	no	northeast	13228.84695	
23	30	female	32.4	1	no	southwest	4149.736	
24	18	male	34.1	0	no	southeast	1137.011	
25	34	female	31.92	1	yes	northeast	37701.8768	
26	37	male	28.025	2	no	northwest	6203.90175	
27	59	female	27.72	3	no	southeast	14001.1338	
28	63	female	23.085	0	no	northeast	14451.83515	
29	55	female	32.775	2	no	northwest	12268.63225	
30	23	male	17.385	1	no	northwest	2775.19215	
31	26	male	26.2	2	yes	southwest	38711.1	

Рис. 2.21. Импорт из текстового файла в книгу Excel

Можно воспользоваться онлайн-конвертором и преобразовать файл с расширением .csv в файл с расширением .xlsx (некоторые конверторы требовательны к размеру файла и не будут выполнять конвертацию для очень объемных наборов данных).

Загрузим файл insurance.xlsx на платформу. Просмотрим сведения о таблице, проконтролируем корректность распознавания типов данных (рис. 2.22).

БАЗЫ... > CLICKH... > 365_INSURANC... **Поля в insurance / Worksheet**

Детали

Поля в этой таблице

Запросы об этой таблице

Сканировать таблицу

Имя поля	Тип поля	Тип данных
Age age	# Нет типа поля	type/Integer
Sex sex	T Нет типа поля	type/Text
Bmi bmi	# Нет типа поля	type/Float
Children children	# Нет типа поля	type/Integer
Smoker smoker	T Нет типа поля	type/Text
Region region	T Нет типа поля	type/Text
Charges charges	# Нет типа поля	type/Float
Glarus Load Dttm glarus_load_dttm	📅 Нет типа поля	type/DateTime

Рис. 2.22. Просмотр типов полей данных для загруженного набора

Описание переменных набора:

- age: возраст основного бенефициара;
- sex: пол застрахованного;
- bmi: индекс массы тела;
- children: число детей, охваченных медицинским страхованием / число иждивенцев;
- smoker: курит ли застрахованный;
- region: жилой район получателя страховых выплат (север-восток, юго-восток, юго-запад, северо-запад);
- charges: индивидуальные медицинские расходы, оплачиваемые страховкой.

Разведочный анализ заключается в построении визуализаций, которые выявляют связи между данными.

Построим визуализацию, с помощью которой можно установить, есть ли связь между весом респондента и его медицинскими расходами.

Построим визуализацию для данных исходной таблицы Разброс (рис. 2.23).

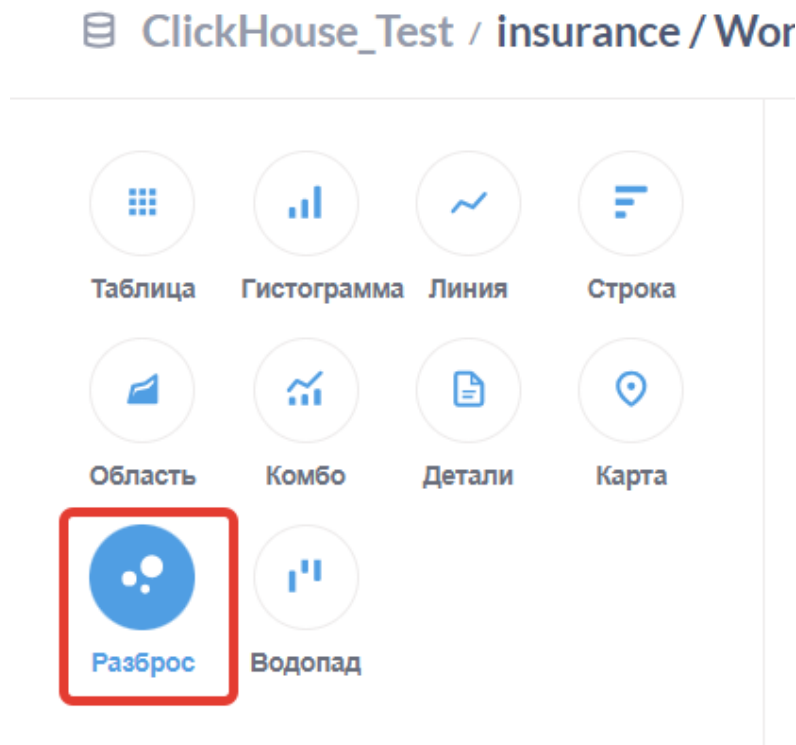


Рис. 2.23. Выбор типа визуализации

В настройках (рис. 2.24) выберем для оси абсцисс показатель bmi (индекс массы тела), а для оси ординат показатель charges (медицинские расходы). Построенную визуализацию можно сделать более информативной, добавив разрыв серии (рис. 2.25). Сохраним запрос.

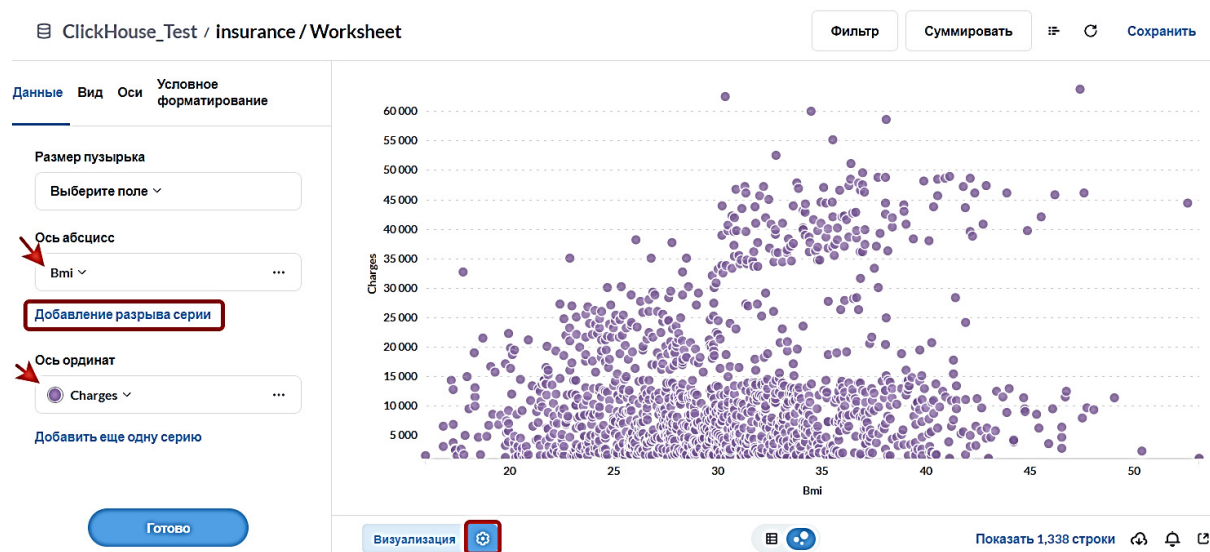


Рис. 2.24. Настройки для визуализации

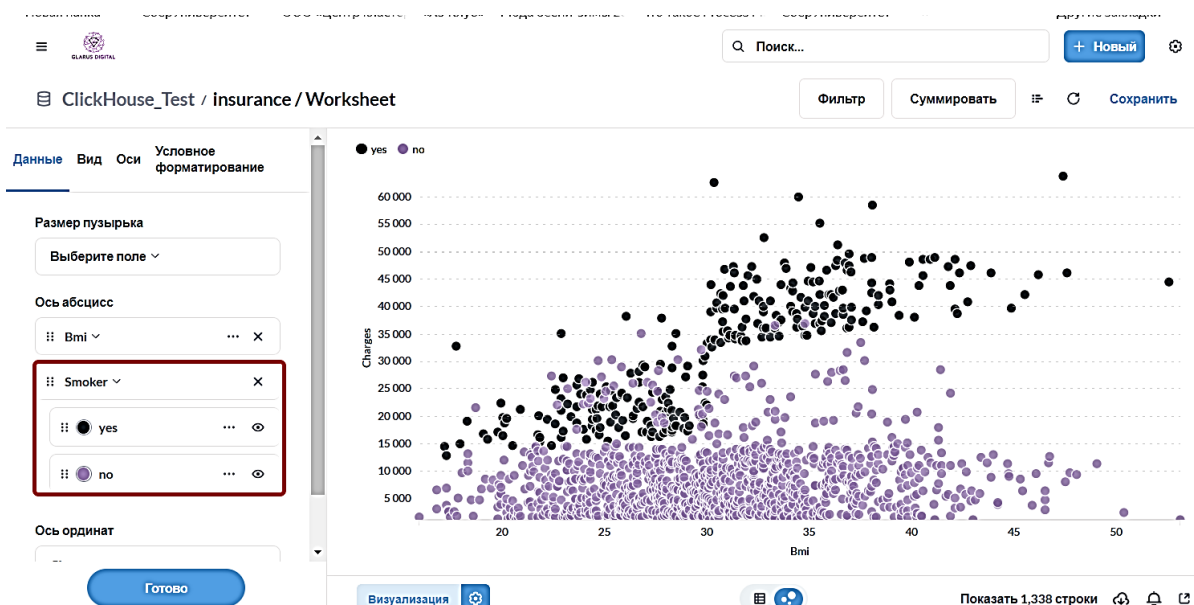


Рис. 2.25. Диаграмма рассеяния с выделением цветом тех, у кого есть вредная привычка

Как видим, расходы зависят и от лишнего веса, и от наличия вредной привычки. Не все обладатели лишнего веса имеют значительные медицинские расходы, однако с увеличением веса вероятность роста расходов повышается.

Можно проанализировать сбалансированность набора данных по показателю smoker. Для этого создадим запрос, в котором суммируем число наблюдений, соответствующее разным значениям показателя smoker (рис. 2.26).

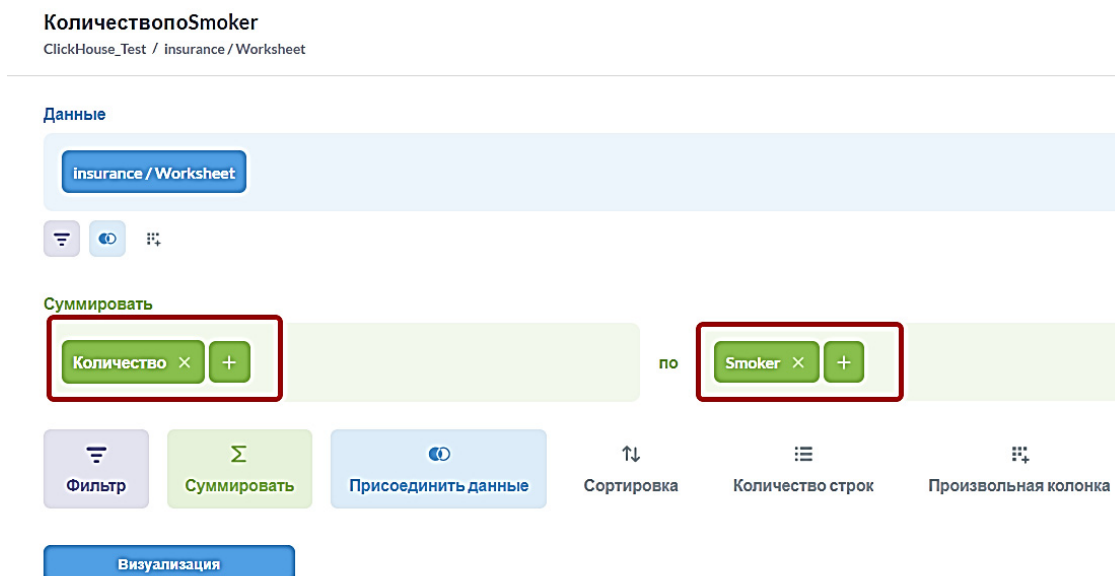


Рис. 2.26. Настройки запроса

В качестве визуализации выберем гистограмму (рис. 2.27) и сохраним запрос.

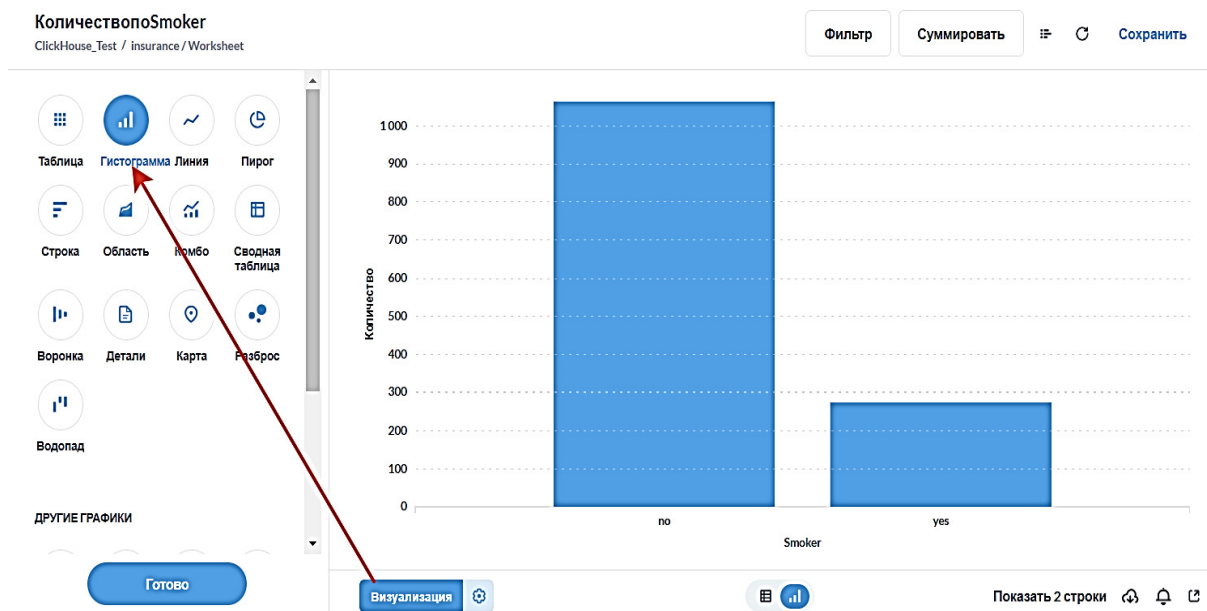


Рис. 2.27. Распространенность вредной привычки в наборе данных

Построим на основе этого запроса новый запрос, определив распространенность вредной привычки в зависимости от района. Для этого продублируем запрос (рис. 2.28, 2.29).

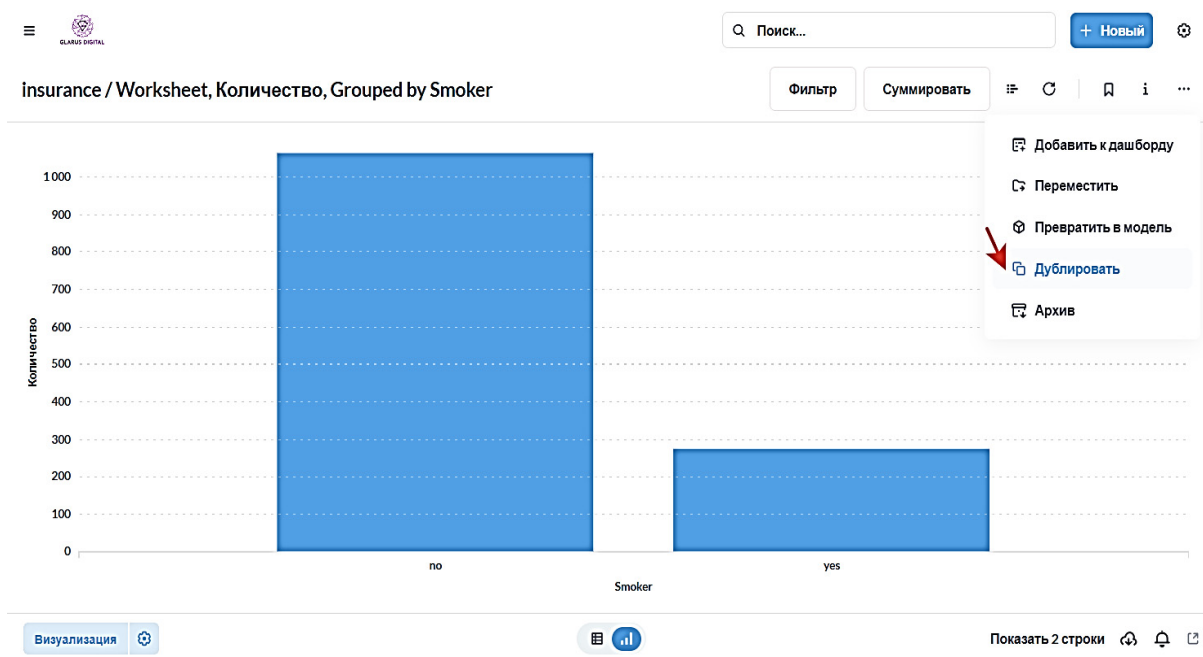


Рис. 2.28. Дублирование запроса

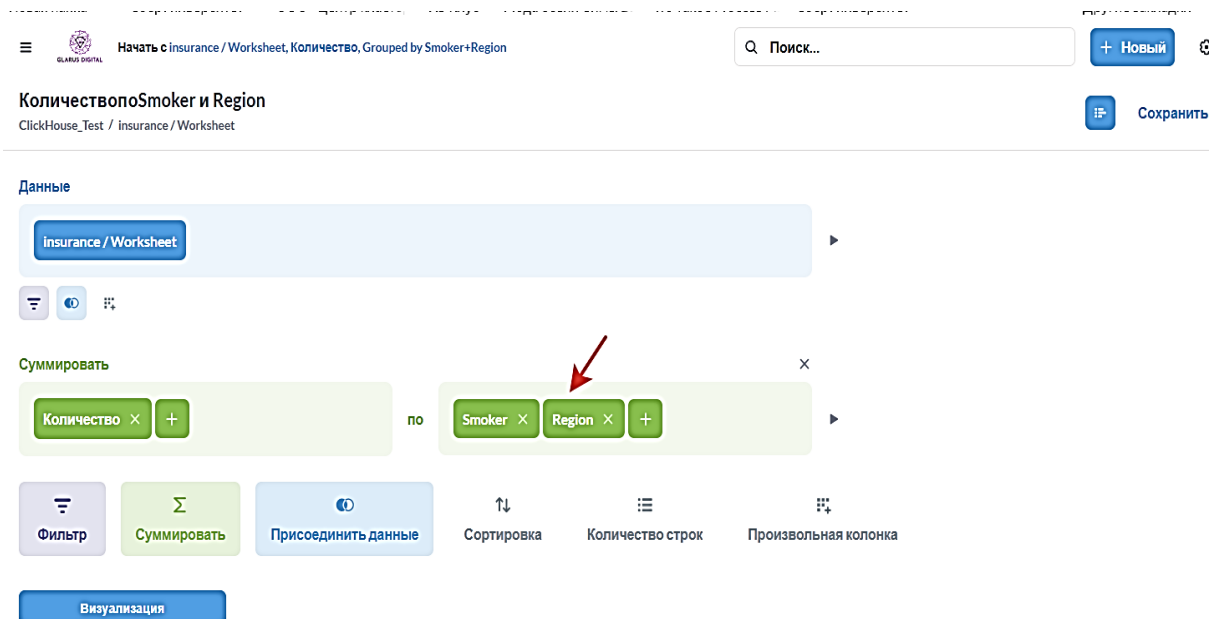


Рис. 2.31. Добавление показателя для группировки наблюдений

Перейдем к визуализации и в ее настройках изменим показатель для оси абсцисс и выберем опцию Добавление разрыва серии (рис. 2.32). В качестве показателя для разрыва серии выберем smoker, дополнительно скорректируем вид полученной визуализации (рис. 2.33). Сохраним запрос (рис. 2.34).

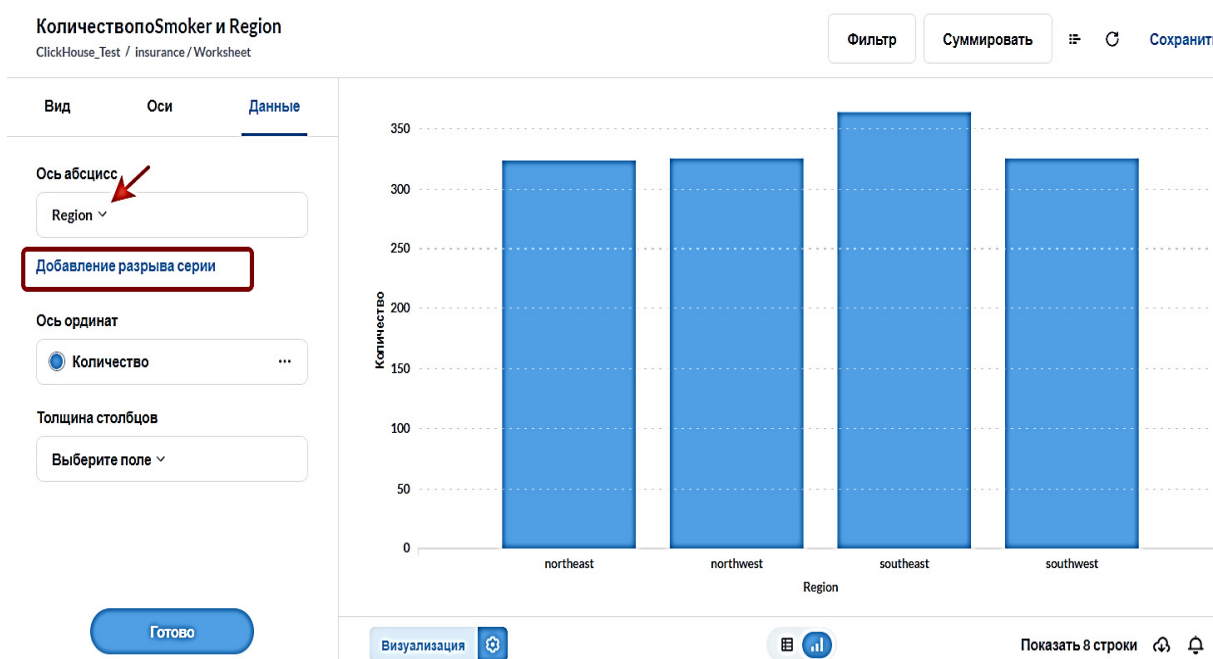


Рис. 2.32. Изменение настроек для визуализации

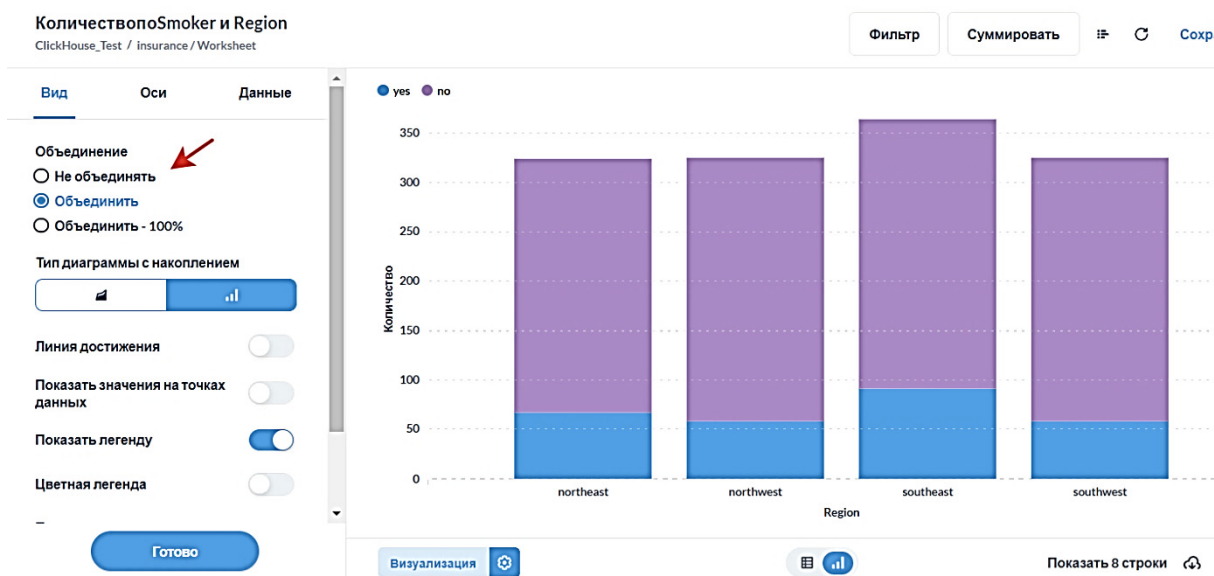


Рис. 2.33. Изменение настроек вида визуализации



Рис. 2.34. Итоговый вид визуализации

Проделаем аналогичные манипуляции с ранее построенной диаграммой рассеяния (см. рис. 2.25): перейдем к этому запросу, создадим его копию (рис. 2.35), поменяем показатель для оси абсцисс (вместо bmi выберем age). Сохраним запрос (рис. 2.36).

Разместим построенные визуализации на новом дашборде с именем Медицинское страхование (рис. 2.37) и добавим к нему фильтр по показателю sex (пол).

Создадим новый запрос по всем данным с визуализацией Таблица (рис. 2.38).

Дублировать "insurance / Worksheet Диаграмма рассеяния расходов от веса"

×

Имя

insurance / Worksheet Диаграмма рассеяния расходов от возраста

Описание

Это не обязательно, но может быть полезным

В какую коллекцию следует это включить?

Светлана личная коллекция Рындина

▼

Дублировать

Рис. 2.35. Изменение имени запроса для копии

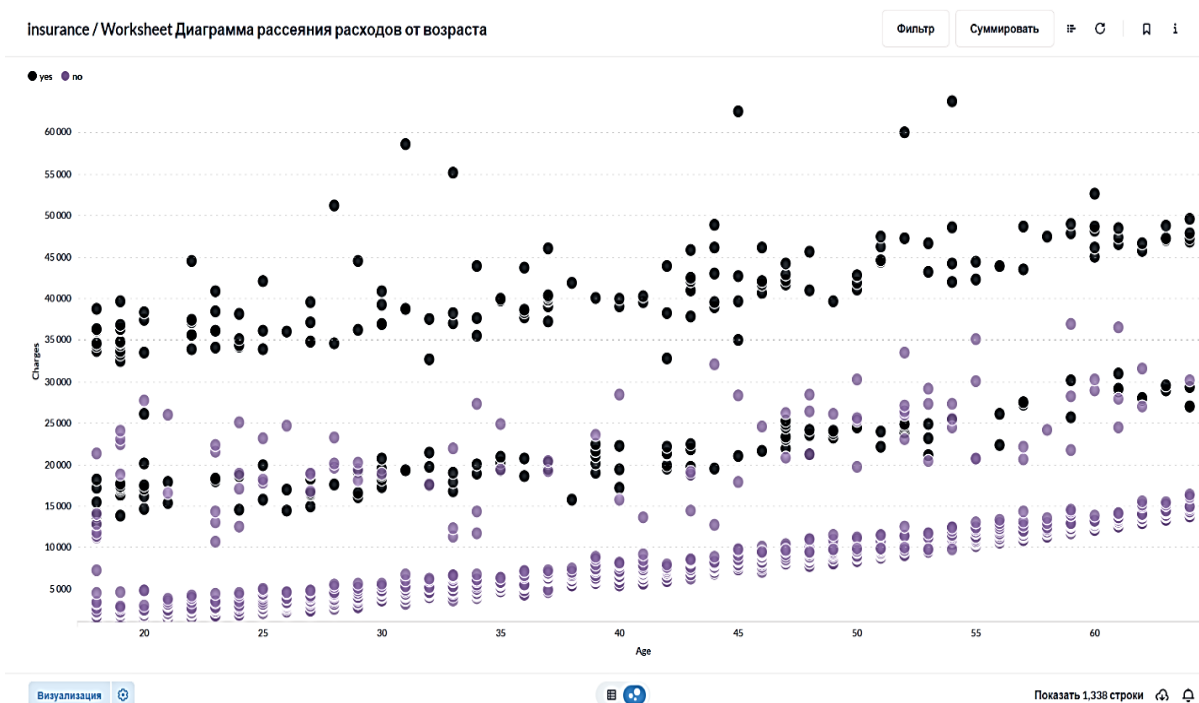


Рис. 2.36. Визуализация по измененному запросу



Рис. 2.37. Дашборд для разведочного анализа данных

insurance / Worksheet

Age	Sex	Bmi	Children	Smoker	Region	Charges	Glarus Load Dttm
19	female	27,9	0	yes	southwest	16 884,92	21 января, 2025, 20:15
18	male	33,77	1	no	southeast	1 725,55	21 января, 2025, 20:15
28	male	33	3	no	southeast	4 449,46	21 января, 2025, 20:15
33	male	22,71	0	no	northwest	21 984,47	21 января, 2025, 20:15
32	male	28,88	0	no	northwest	3 866,86	21 января, 2025, 20:15
31	female	25,74	0	no	southeast	3 756,62	21 января, 2025, 20:15
46	female	33,44	1	no	southeast	8 240,59	21 января, 2025, 20:15
37	female	27,74	3	no	northwest	7 281,51	21 января, 2025, 20:15
37	male	29,83	2	no	northeast	6 406,41	21 января, 2025, 20:15
60	female	25,84	0	no	northwest	28 923,14	21 января, 2025, 20:15

Рис. 2.38. Исходные данные набора

На созданный дашборд добавим ссылку (рис. 2.39) на источник данных на ресурсе (<https://www.kaggle.com/mirichoi0218/insurance>)

и на запрос с таблицей данных. Итоговый вид дашборда показан на рис. 2.40.



Рис. 2.39. Добавление ссылки

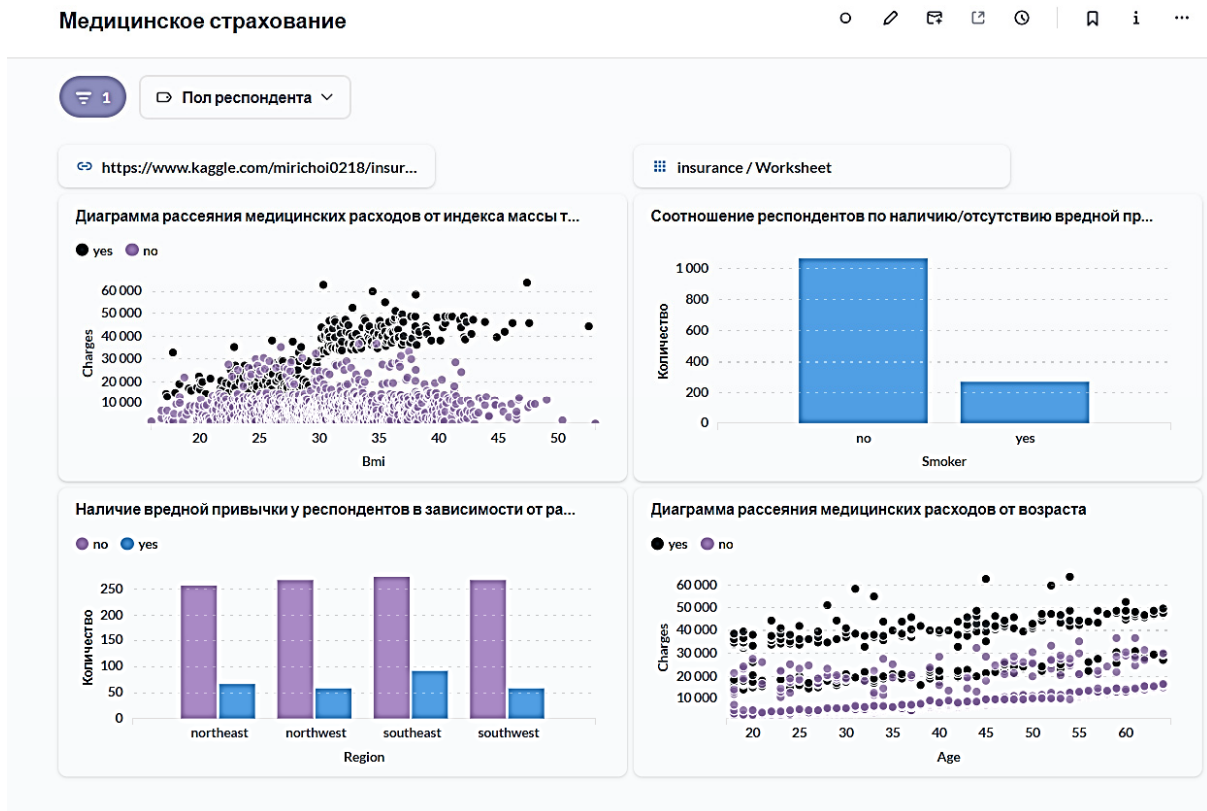


Рис. 2.40. Итоговый вид дашборда

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 вычислимыми полями. Получить навыки формирования отчета по данным, включающего несколько визуализаций, с пользовательскими настройками форматирования.

Необходимо:

1. Загрузить набор данных.
2. Создать запросы: различные виды диаграмм, сводную таблицу. Для сводной таблицы использовать условное форматирование. В одной из диаграмм использовать сортировку.
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 и задания

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

Дополнительные материалы

У автора пособия есть издания по другим BI-системам, с которыми можно ознакомиться для сравнения функционала этих систем от разных издателей [3, 4].

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).
3. Рындина С. В. Интеллектуальные информационные системы и технологии: системы Business Intelligence (Yandex DataLens) : учеб.-метод. пособие. Пенза : Изд-во ПГУ, 2023. 64 с.
4. Рындина С. В. Интеллектуальные информационные системы и технологии: системы Business Intelligence (ML-платформа Loginom) : учеб.-метод. пособие. Пенза : Изд-во ПГУ, 2022. 68 с.

Учебное издание

Рындина Светлана Валентиновна

**Интеллектуальные информационные
системы и технологии: системы
Business Intelligence (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. 3,72.
Тираж 6. Заказ № 96.

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

Вниманию авторов!

Издательство ПГУ выпускает учебную, научную и художественную литературу, презентационную и акцидентную продукцию, а также полноцветные юбилейные и мемориальные издания в соответствии с ГОСТ 7.60–2003.

Издательство ПГУ принимает к изданию рукописи, подготовленные с использованием текстового редактора Microsoft Word for Windows версий **2003 и выше**. Формат – А4, основной шрифт – Times New Roman, 14–16 pt через одинарный интервал (минимальный размер шрифта в таблицах и сносках – 12,5 pt). Тип файла в электронном виде – doc, docx.

Работа должна содержать индекс УДК, аннотацию.

Аннотация (ГОСТ 7.86–2003, ГОСТ 7.9–1995) включает характеристику основной темы, проблемы объекта, цели работы и ее результаты. В аннотации указывают, что нового несет в себе данный документ в сравнении с другими, родственными по тематике и целевому назначению. Аннотация может включать сведения о достоинствах произведения. Текст аннотации начинают фразой, в которой сформулирована главная тема документа. Заканчивается аннотация читательским адресом.

Рисунки и таблицы должны быть размещены в тексте после ссылки на них (растровые рисунки предоставляются в виде отдельных файлов в формате jpg, BMP с разрешением 300 dpi, векторные рисунки в формате Corel Draw с минимальной толщиной линии 0,75 pt. Рисунки должны быть доступны для правки!). Рисунки должны сопровождаться подрисуночными подписями, на все рисунки и таблицы в тексте должны быть ссылки.

Формулы в тексте выполняются только в редакторе формул **MathType версия 5.0** и выше. Символы греческого и русского алфавита должны быть набраны прямо, нежирно; латинского – курсивом, нежирно; обозначения векторов и матриц – прямо, жирно; цифры – прямо, нежирно. Наименования химических элементов набираются прямо, нежирно. Эти же требования необходимо соблюдать и в рисунках.

В списке литературы **нумерация источников** должна соответствовать очередности ссылок на них в тексте ([1], [2], ...). Номер источника указывается в квадратных скобках. Требования к оформлению списка литературы на русские и иностранные источники (ГОСТ Р 7.0.5–2008): для книг – фамилия и инициалы автора, название, город, издательство, год издания, том, количество страниц; для журнальных статей, сборников трудов – фамилия и инициалы автора, название статьи, полное название журнала или сборника, серия, год, том, номер, страницы; для материалов конференций – фамилия и инициалы автора, название статьи, название конференции, город, издательство, год, страницы.

К материалам **должна** прилагаться следующая информация: фамилия, имя, отчество, контактные телефоны.

Контакты Издательства ПГУ: т.: (8412) 66-60-49, 66-67-77, e-mail: iic@pnzgu.ru



