

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. Аналитические
платформы для работы с данными

Учебное пособие

Пенза
Издательство ПГУ
2025

УДК 004.6 (075)

P95

Р е ц е н з е н т ы :

доктор технических наук, доцент,
заведующий кафедрой «Информационно-вычислительные системы»
Пензенского государственного университета

А. В. Кузьмин;

кандидат экономических наук, доцент,
руководитель сектора методологии учета и оформления
банковских операций ПАО Банк «Кузнецкий» (г. Пенза)

О. Б. Волошина

Рындина, Светлана Валентиновна.

P95

Интеллектуальные информационные системы и технологии. Аналитические платформы для работы с данными : учеб. пособие / С. В. Рындина. – Пенза : Изд-во ПГУ, 2025. – 108 с.

ISBN 978-5-908089-14-2

Рассмотрены инструменты для работы с данными: BI-системы, ML-платформа Loginom.

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» (направление подготовки 38.03.05 «Бизнес-информатика»), «Интеллектуальные информационные системы» (направление подготовки 38.03.05 «Бизнес-информатика»).

УДК 004.6 (075)

ISBN 978-5-908089-14-2

© Пензенский государственный
университет, 2025

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
1. АНАЛИТИКА ДАННЫХ И ЕЕ ИНСТРУМЕНТЫ.....	5
1.1. Виды аналитики данных	5
1.2. Ключевые аспекты организации аналитических процессов на предприятии.....	10
1.3. Наборы данных для тестирования возможностей аналитических инструментов	18
2. BI-СИСТЕМЫ.....	26
2.1. Аналитика с Yandex DataLens	29
2.2. Аналитика с Visiology	48
2.3. Аналитика с Glarus BI	67
3. ML-ПЛАТФОРМА LOGINOM	90
СПИСОК ЛИТЕРАТУРЫ.....	106

ВВЕДЕНИЕ

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

В пособии рассмотрены цифровые сервисы для аналитики данных, включающие инструменты для хранения данных, для преобразования и визуализации данных, для построения информационных панелей (дашбордов) и для извлечения ценности из данных.

В пособии рассмотрены основные виды аналитики данных, особенности организации аналитических процессов в компаниях, требования к программному обеспечению, компетенциям персонала, инфраструктуре.

В фокусе внимания в пособии отечественные решения для аналитики и интеллектуального анализа данных.

1. АНАЛИТИКА ДАННЫХ И ЕЕ ИНСТРУМЕНТЫ

1.1. Виды аналитики данных

Аналитика данных – это процесс сбора, обработки, анализа и интерпретации данных для выявления закономерностей, принятия решений и улучшения бизнес-результатов.

Современная аналитика все чаще включает интеллектуальные технологии:

- искусственный интеллект (ИИ), Artificial Intelligence;
- машинное обучение, Machine Learning;
- обработку естественного языка, Natural Language Processing.

Эти технологии позволяют переходить от описания прошлого к прогнозированию будущего и автоматизации решений.

Существует четыре основных вида аналитики, выделяемых по принципу «лестницы зрелости» (табл. 1.1).

Таблица 1.1

Основные виды аналитики

Уровень	Вид аналитики	Основной бизнес-вопрос, который такая аналитика обслуживает
1	Описательная (Descriptive)	Что произошло?
2	Диагностическая (Diagnostic)	Почему произошло?
3	Прогнозная (Predictive)	Что может произойти?
4	Прескриптивная (Prescriptive)	Что делать?

Цель описательной аналитики – понять прошлое. Типичные метрики, которые находятся в центре внимания описательной аналитики: оборот за месяц, количество новых клиентов, конверсия на сайте, отчеты по продажам.

Инструменты, способные обеспечить такой уровень аналитики: BI-системы (например, отечественные решения Yandex DataLens, GlarusBI, Visiology), табличные процессоры (Excel, Google Sheets), системы отчетности (1C)). Для извлечения данных из первоисточников, в роли которых выступают хранилища данных и/или базы данных приложений, систем, сервисов и т.п., используется SQL (Structured Query Language, язык структурированных запросов).

BI-системы – это автоматизированные аналитические решения, которые собирают, обрабатывают и визуализируют большие объемы данных из разных источников.

Интеллектуальные инструменты для описательной аналитики:

- автоматическая генерация отчетов;
- построение графиков по данным;
- подсказки и визуализации для бизнес-вопросов на естественном языке.

Такие инструменты в настоящий момент активно внедряются в популярные BI-решения, в том числе и от российских разработчиков (подробнее в разделах, посвященных соответствующим продуктам).

К преимуществам описательной аналитики относится то, что она подходит для оперативной отчетности и мониторинга, проста для понимания и имеет очень низкий порог для интеграции в процессы (минимальные требования к компетенциям персонала или наличие недорогих, легко развертываемых на любой инфраструктуре или даже арендуемых облачных решений).

Однако с описательной аналитикой доступен только ретроспективный анализ, она предоставляет недостаточно информации для объяснения причин появления тех или иных значений метрик/показателей, на ее основе сложно формулировать обоснованные рекомендации для бизнеса.

Диагностическая аналитика позволяет глубже проникнуть в суть прошлой бизнес-деятельности и не только оперирует историческими данными, фиксируя агрегированные значения за прошедшие периоды, но и способна предоставить информацию о причинах таких значений показателей, т.е. поясняет, почему что-то произошло.

Типичные вопросы, на которые способна ответить диагностическая аналитика:

- Почему продажи упали в июле?
- Какие факторы повлияли на отток клиентов?
- Где произошел сбой в логистике?

Инструменты: BI-системы с drill-down-анализом (аналитический подход, который позволяет исследовать данные на различных уровнях детализации для более полного понимания их структуры и взаимосвязей), SQL и Python для анализа корреляций, платформы (например, Yandex DataLens).

Интеллектуальные инструменты:

- автоматический поиск причин изменений;

- автоматическая кластеризация аномалий;
- NLP для анализа текстовых данных (например, анализ отзывов клиентов).

Диагностическая аналитика углубляет понимание бизнес-процессов, помогает строить причинно-следственные связи для бизнес-событий, может выявлять неочевидные зависимости. Но она не всегда автоматизируема, хотя разнообразные интеллектуальные инструменты, которые все чаще появляются для решения диагностических задач, кардинально снижают долю ручных операций в процессе, также диагностическая аналитика требовательна к качеству исходных данных.

Прогнозная аналитика способна прогнозировать/предсказывать, что может произойти в будущем, например, построить прогноз спроса на товар, оценить вероятность оттока клиентов или спрогнозировать вероятность отказа оборудования, чтобы своевременно выполнить технические обслуживание/ремонт/замену и не останавливать внезапно и на продолжительный срок производственный процесс из-за аварии. Такая аналитика требует более серьезного инструментария: это прикладное программирование на языках R и Python. Каждый из этих языков программирования имеет хорошо проработанную инфраструктуру для аналитики. На языке R существует большое число пакетов, предоставляющих функции/методы/классы для обработки данных, на языке Python еще более обширный перечень библиотек (Pandas, Scikit-learn и т.д.) для построения в том числе приложений с использованием технологий ИИ: ВІ-системы, поддерживающие запуск скриптов для реализации дополнительных методов работы с данными, машинного обучения и т.п., или ВІ с МL-вставками, т.е. включающими специальные модули для предоставления доступа к методам машинного обучения; МL-платформы, например, российская аналитическая платформа Loginom); облачные МL-сервисы, среди отечественных разработок можно выделить Yandex DataSphere – сервис для полного цикла машинного обучения (Yandex Cloud), МL Space (SberCloud), Cloud МL Platform (VK Cloud Solutions).

Интеллектуальные инструменты:

- автоматическое машинное обучение (AutoML), выбор модели, обучение, валидация без глубоких знаний;
- интеграция МL-моделей в ВІ.

Преимущества прогнозной аналитики заключаются в предоставлении возможностей действовать проактивно, контролировать и сокращать риски, увереннее смотреть в будущее, снижая его неопределенность.

Однако для прогнозной аналитики критически важным является наличие качественных исторических данных, при этом модели должны периодически тестироваться на актуальность, так как быстро устаревают, для их обновления требуются свежие качественные данные, интерпретация результатов – сложный процесс, требующий привлечения специалистов по данным, равно как без них не обойтись и на других этапах реализации прогнозной аналитики. Повысить квалификацию неспециалистов до нужного уровня скорее всего не получится, и нужно в обязательном порядке опереться на профессиональных аналитиков, инженеров данных и т.п.

Прескриптивная аналитика помогает выработать программу действий для достижения бизнес-целей, например, как оптимально распределить бюджет на рекламу, как минимизировать простои на производстве, какое предложение сделать клиенту, чтобы он пересмотрел свое решение и остался. Самые продвинутые инструменты требуются именно для такой аналитики: оптимизационные решатели (комбинация прогнозов и алгоритмов принятия решений), встроенные прескриптивные модули в ERP-системах, кастомные системы на Python.

Интеллектуальные инструменты:

- гибридные AI-системы;
- Digital Twins – цифровые двойники процессов для тестирования решений.

Преимущества прескриптивной аналитики в том, что она максимально близка к автоматизации решений, учитывает множество факторов и ограничений, позволяет моделировать различные сценарии развития бизнеса на основе данных с целью последующей оптимизации, т.е. выбора наилучшей последовательности решений для реализации самого успешного сценария будущего.

Прескриптивная аналитика сложна в реализации, требует больших массивов данных, мощных вычислений и экспертизы.

Как правило, редко используется в чистом виде, а чаще входит в состав ERP, SCM, CRM и имеет высокую стоимость внедрения (обычно поставляется как дополнительный компонент/модуль или слой в системе с базовым функционалом).

Все перечисленные системы ERP, SCM, CRM используются для обслуживания и автоматизации бизнес-процессов:

ERP, Enterprise Resource Planning (планирование ресурсов предприятия) – это системы для управления бизнес-процессами на предприятии. Отечественные разработки таких систем – 1С: ERP, Галактика ERP, Парус, Битрикс24.

SCM, Supply Chain Management (управление цепочками поставок) – логистическое программное решение, которое управляет потоком товаров, данных и финансов по мере перемещения продукта или услуги из пункта отправления в конечный пункт назначения. Отечественные разработки таких систем – RS.SCM, «Умная Логистика Карго», Lean ERP SCMo, Visary SCM, Global-SCM, AXELOT SCM.

CRM, Customer Relationship Management (система управления взаимоотношениями с клиентами) – это программное обеспечение, которое автоматизирует процессы взаимодействия с заказчиками, помогает управлять продажами, маркетингом и сервисом. Отечественные разработки таких систем – 1С:CRM, Битрикс24, Мегалан.

Существуют также комплексные решения, чаще платформенные, которые объединяют представленные системы в одну экосистему, позволяющую реализовать все указанные группы процессов.

Подведем итоги.

Описательная статистика как самая простая и базовая реализуется большинством компаний и даже в случае соло-бизнеса, например самозанятыми и индивидуальными предпринимателями. Для нее характерны простота, очевидная полезность и применимость результатов в виде базовой отчетности, как минимум пригодной основы для различных внешних инстанций.

Диагностическая аналитика дает более глубокое понимание данных бизнеса, но и требует большей вовлеченности в процессы, а также связана с освоением более сложных процедур обработки данных.

Прогнозная и перспективная выполняются специалистами в области данных с привлечением более качественных и мощных инструментов аналитики, работа с которыми требует профессиональных знаний и умений. Сложные, редкие и дорогие инструменты имеют большие перспективы быть полезными и отработать затраты на свое внедрение и использование там, где бизнес масштабен, а цена ошибок велика. В небольшом бизнесе, который не связан напрямую с оказанием услуг по аналитике, т.е. работа с данными не входит в ключевые компетенции бизнеса, более

простые и недорогие инструменты, low-code и zero-code (малокодовые и бескодовые) решения, которые могут освоить и непрофессионалы, принесут достаточно ценной информации для управления бизнесом, и вложения в такую аналитику точно окупятся.

1.2. Ключевые аспекты организации аналитических процессов на предприятии

Одним из самых важных аспектов аналитической культуры является понимание того, какие основные системы формируют аналитический ландшафт предприятия, их функционал и возможности в аналитических процессах.

Рассмотрим основные различия между BI-, ML-, ERP-системами, которые обслуживают три разных, но часто взаимодополняющих направления в цифровой трансформации бизнеса.

1. ERP-система – это один из основных поставщиков первичных данных. Ее основные функциональные возможности:

- используется для управления всеми ключевыми бизнес-процессами компании;
- объединяет данные из различных отделов: финансы, логистика, закупки, производство, HR, продажи и т.д.;
- автоматизирует рутинные операции: учет, отчетность, управление запасами, заказами, зарплатой и т.п.

Цель внедрения ERP на предприятии – оптимизация внутренних процессов, повышение эффективности и снижение издержек.

2. BI-система отвечает за базовую бизнес-аналитику. Она тоже продуцирует данные, но уже производные, которые помогают в управлении предприятием и в формировании отчетности. Основным ресурсом для работы такой системы являются данные различных бизнес-процессов. Чем большую гранулярность имеют данные, тем больше полезной информации можно из них получить.

Гранулярность данных – это уровень детализации информации в базе данных или аналитической системе.

Например, продажи в розничной сети могут быть описаны с помощью данных следующей гранулярности:

- высокая гранулярность соответствует данным, когда в базу передается каждая транзакция с указанием времени, кассы, товара, цены, скидки, способа оплаты;
- средняя гранулярность соответствует данным ежедневных продаж по категориям товаров в каждом магазине;

– низкая гранулярность соответствует максимально агрегированным/общим данным о месячной выручке по регионам (или по отдельным магазинам, если сеть локализована в одном регионе).

Чем выше гранулярность, тем более детализированы данные. Чем ниже – тем более они агрегированы и обобщены.

BI-системы используются для сбора, интеграции, анализа и визуализации бизнес-данных с целью поддержки принятия решений.

Основные функции BI-систем:

- сбор и интеграция данных из разных источников (например, ERP-систем);
- визуализация данных (графики, диаграммы, таблицы);
- построение дашбордов и отчетов (часто на основе визуализаций с использованием специальных приемов объединения и структуризации визуальной информации);
- аналитика в реальном времени и за исторический период;
- OLAP-анализ (Online Analytical Processing) – многомерный анализ данных;
- поддержка ad-hoc-запросов (анализ по требованию);
- регламентированная отчетность (ежедневные, ежемесячные отчеты);
- фокус на описательной аналитике (descriptive analytics) – анализ прошлых и текущих данных.

Цель внедрения BI-систем на предприятии – понимание происходящих процессов в бизнесе для повышения управляемости.

3. ML-платформы позволяют строить модели, которые обучаются на данных и делают прогнозы или принимают решения без явного программирования.

Основные функции ML-платформ:

- обучение моделей на исторических данных;
- прогнозирование (спрос, отток клиентов, мошенничество), классификация и кластеризация;
- ассоциативные правила и персонализированные рекомендации;
- автоматизация процессов (например, обработка текстов);
- интеграция обученных моделей в бизнес-процессы;
- фокус на предсказательной и прескриптивной аналитике (predictive и prescriptive analytics).

Цель внедрения ML-платформ на предприятии – снизить неопределенность исполнения бизнес-процессов за счет предсказания

событий/исходов различных вариантов управления и автоматизировать действия на основе паттернов в данных.

Основные аспекты рассмотренных систем объединены в табл. 1.2.

Таблица 1.2

Сравнение BI-, ML-, ERP-систем

Аспект сравнения	Вид системы		
	ERP-система	BI-система	ML-платформа
Основная цель	Управление операциями	Анализ данных и отчетность	Прогнозирование и автоматизация
Тип данных	Операционные (транзакции)	Агрегированные и исторические	Структурированные и неструктурированные
Фокус	Процессы и автоматизация	Анализ прошлой и текущей ситуации в бизнесе	Прогноз будущего развития бизнес-деятельности и возможных вариантов его достижения
Временной аспект	Реальное время / текущие данные	Прошлые и текущие данные	Будущее (прогнозы)
Пользователи	Менеджеры, операторы, бухгалтеры	Аналитики, руководители	Команды специалистов по данным, продукты с ИИ
Выходные данные	Документы, отчеты, процессы	Дашборды, графики, отчеты	Модели, прогнозы, рекомендации
Пример использования	Учет товара на складе	Анализ прибыли по регионам	Прогноз спроса на следующий квартал

Рассмотрим, как эти системы работают в связке:

1. ERP генерирует операционные данные (продажи, закупки, движение товаров).

Пример: ERP фиксирует выполненные заказы.

2. BI загружает данные из ERP-системы, агрегирует и показывает аналитику (например, дашборд по выручке).

Пример: BI строит отчет «Динамика продаж за последние два квартала», по которому можно сделать вывод, что продажи в последнем квартале выросли на 15 % по отношению к предыдущему кварталу.

3. ML использует исторические данные из BI- и ERP-систем, чтобы предсказать будущее (например, спрос на товары), и может автоматизировать действия (например, запуск закупки в ERP).

Пример: ML анализирует историю и выдает прогноз: «С вероятностью 85 % спрос вырастет на 20 % в следующем квартале – рекомендуем увеличить закупки».

На предприятии все три компонента важны:

- ERP обеспечивает стабильность и прозрачность операций;
- BI дает понимание текущего положения;
- ML добавляет интеллект и способность к предвосхищению.

Современные платформы все чаще объединяют BI и интеллектуальную аналитику, будущее за «умными дашбордами», которые не просто показывают данные, а «предлагают действия». Грань между BI-системами и ML-платформами за счет внедрения интеллектуальных инструментов будет размываться, а возможности BI-систем по выполнению продвинутой аналитики будут расти и развиваться.

Еще одним важным аспектом в формировании и развитии аналитической культуры является понимание того, как это развитие можно организовать. Для этого необходимо выяснить текущий уровень и текущее состояние аналитики в компании и определить желаемый уровень, а также ближайший к текущему, т.е. потенциально достижимый уровень развития, выявить, чем они различаются, и наметить план действий для ликвидации обнаруженных разрывов.

Уровни аналитической зрелости отражают этапы эволюции процесса принятия решений в компаниях и способность эффективно применять аналитику для достижения стратегических целей.

Традиционно выделяют пять уровней аналитической зрелости. Рассмотрим каждый из них с различных точек зрения. Начнем с характеристики аналитических процессов и управления на их основе.

1. Начальный уровень («реактивный»):

- описание проблемы после ее возникновения;
- применение простого анализа и базовых отчетов;
- использование устаревших данных и слабых методов анализа;
- решения принимаются на основе интуиции и опыта руководителя.

2. Адекватный уровень («рутинный»):

- регулярный сбор и систематический анализ данных;
- появление автоматизированных отчетов и планов;
- используются стандартные метрики и KPI (Key Performance Indicators, ключевые показатели эффективности, которые отражают результаты работы компании, подразделения или отдельных сотрудников);

– улучшается понимание текущих ситуаций, хотя большинство действий остаются реактивными.

3. Средний уровень («прогностический»):

– организация переходит к активному использованию прогностической аналитики;

– применяются статистические методы и модели предсказательных сценариев;

– прогнозируются возможные события и будущие риски;

– меньше зависимы от исторических данных, больше внимания уделяется будущим сценариям.

4. Зрелый уровень («инновационный»):

– активно применяется аналитика высокого уровня, включая предсказательную и прескриптивную аналитику;

– большое внимание уделяется инновационным методикам и экспериментальным проектам;

– постоянно тестируются гипотезы и внедряются новые подходы;

– команды становятся самостоятельными центрами компетенций в аналитике.

5. Оптимальный уровень («преобразующий»):

– процессы анализа данных интегрированы во все аспекты деятельности компании;

– информация превращается в знания и стратегии, влияющие на стратегию компании;

– организации применяют самые современные технологии и методики, такие как AI и ML;

– культура аналитики становится частью корпоративной культуры.

Определив свое текущее положение и направления дальнейшего развития, организация приступает к трансформации аналитических подходов и формированию необходимых компетенций у сотрудников.

На начальном уровне («реактивном») команда должна уметь решать базовые задачи по обработке данных и подготовке элементарных отчетов. Требования к компетенциям сотрудников, участвующих в аналитических процессах на этом уровне:

– сбор данных вручную и ввод в простую форму отчета;

– первичный анализ числовой информации;

– навыки составления отчетов на основе подготовленных данных;

- умение работать в табличных процессорах (например, MS Excel);

- понимание принципов простой статистики и построения сводных таблиц.

На следующем уровне («рутинном») возникает необходимость организовать регулярный сбор и обработку данных, сотрудники, участвующие в аналитических процессах, осваивают создание регулярных отчетов и применение стандартных методик анализа.

Базовый набор необходимых компетенций включает:

- автоматизацию сбора и хранения данных;
- владение базовыми методами анализа данных (средние значения, медианы, дисперсии и т.п.);
- работу с OLAP-технологиями и многомерными моделями данных;
- формулировку простых гипотез и проверку их достоверности;
- использование специализированных инструментов для построения отчетов (BI-систем).

На среднем уровне («прогностическом») появляется потребность в применении прогностических методов и создании прогнозов, а ключевые способности, которые должны демонстрировать сотрудники в аналитических процессах, заключаются в умении обрабатывать и интерпретировать большие объемы данных.

Необходимые умения включают:

- моделирование временных рядов и работу с историческими данными;
- статистический анализ данных (регрессии, корреляционные исследования, кластерный анализ);
- основы Machine Learning (предсказательные модели, классификаторы);
- глубокое знание предметной области и умение сформулировать правильную постановку задачи.

На четвертом уровне («инновационном») сотрудники сталкиваются с необходимостью постоянного тестирования гипотез и экспериментов с новыми аналитическими подходами, эффективность аналитической работы определяется способностью находить нестандартные решения и извлекать максимальную пользу из данных.

Необходимые компетенции включают:

- углубленное владение статистическим анализом и технологиями искусственного интеллекта;

- программирование на языках R/Python для работы с данными и разработки моделей;
- экспертизу в современных технологиях работы с большими данными (облачные вычисления, распределенные вычисления);
- понимание современных тенденций в области анализа данных и машинного обучения;
- способность ставить долгосрочные цели и вести проекты трансформации бизнеса на основе данных.

Аналитика на оптимальном уровне («преобразующем») полностью интегрируется в процессы принятия решений и влияет на общую стратегию компании, характеризуется глубоким пониманием роли данных в достижении конкурентных преимуществ сотрудниками.

Необходимые компетенции на этом уровне:

- стратегическое мышление и глубокое понимание бизнес-задач;
- управление проектами преобразования и трансформации бизнеса;
- передача накопленного опыта сотрудникам компании и вовлечение всей организации в процесс аналитики;
- проведение исследований и разработок в области инновационных технологий и методов анализа;
- участие в международных конференциях и публикация научных статей по вопросам анализа данных.

Переход от одного этапа к другому сопровождается развитием профессиональных навыков членов команды и ростом общего уровня компетентности.

Инфраструктура, инструменты и технологии обеспечивают эффективную работу аналитических процессов, на каждом уровне аналитической зрелости возникают повышающиеся требования к инфраструктуре, позволяющие бесперебойно, надежно и безопасно эксплуатировать настроенные аналитические процессы.

Каждому уровню аналитической зрелости соответствуют свои требования к инфраструктуре.

1. Требования к инфраструктуре на начальном уровне:

- простейшие серверы и рабочие станции для ввода и хранения данных;
- пространства для хранения файлов (локально или в облаке).

Примеры используемых решений:

- локальные файлы Excel для сбора и первичной обработки данных;
- веб-интерфейсы для экспорта данных из учетных систем (ERP, CRM).

2. Требования к инфраструктуре на уровне «рутинный»:

- серверы с достаточным объемом памяти и процессорными ресурсами для обработки регулярных запросов;
- хранилища данных (Data Warehouse);
- локальные сети и каналы связи для синхронизации данных.

Примеры используемых решений:

- создание регулярных ежедневных отчетов на основе собранных данных;
- использование SQL-запросов для извлечения данных из единой базы.

3. Требования к инфраструктуре на уровне «прогностический»:

- мощные серверы с высокой производительностью и емкостью дисков;
- облачные ресурсы для быстрого масштабирования (например, Yandex Cloud);
- большие хранилища данных и механизмы распределенного хранения (Big Data технологии).

Примеры используемых решений:

- импорт данных из различных источников (лог-файлы, социальные сети, IoT-устройства);
- построение прогнозных моделей для будущего спроса или риска дефолта.

4. Требования к инфраструктуре на уровне «инновационный»:

- масштабируемая высокопроизводительная инфраструктура с возможностью быстрого горизонтального масштабирования;
- выделенные GPU-ресурсы (Graphics Processing Unit, вычислительные мощности, которые используются для решения задач, требующих параллельной обработки данных, с помощью графических процессоров) для работы с Deep Learning (глубокое обучение, вид машинного обучения, который использует многослойные нейронные сети для анализа данных и принятия решений) алгоритмами;
- современные виртуальные среды для разработчиков и исследователей (Docker, Kubernetes).

Примеры используемых решений:

- реализация рекомендательных систем на основе машинного обучения;

- создание собственных библиотек для специфичных аналитических задач.

5. Требования к инфраструктуре на уровне «преобразующий»:

- максимально производительные облачные ресурсы и гибридные архитектуры;

- центры обработки данных с собственной системой резервирования и дублирования.

Примеры используемых решений:

- использование блокчейна для подтверждения происхождения данных.

Таким образом, по мере повышения уровня аналитической зрелости компании усложняются и усиливаются технологические требования к инфраструктуре и используемым инструментам. Чтобы поддерживать высокий уровень аналитической зрелости, необходимо постоянно обновлять инфраструктуру и внедрять новейшие технологии.

1.3. Наборы данных для тестирования возможностей аналитических инструментов

В реальном бизнесе есть собственные данные в цифровом формате, даже если бизнес небольшой и управляет им самозанятый, оплата налогов через приложение позволяет подводить итоги по заказам/доходам и т.п.

Но при отсутствии возможности использовать для знакомства с различными аналитическими инструментами собственные данные можно воспользоваться платформой Kaggle [1].

Платформа предоставляет пользователям доступ:

- к наборам данных для бесплатной загрузки и тестирования навыков моделирования. Это могут быть данные о заработной плате экспертов из разных индустрий, рейтинг видов спорта и т.п.;

- облачным инструментам для обработки данных и машинного обучения, например Kaggle Kernels (Notebooks). Поддерживают различные языки программирования (Python, R и др.), есть доступ к GPU и TPU для ускорения вычислений;

- встроенным библиотекам и инструментам для анализа данных и машинного обучения: Pandas, NumPy, Scikit-learn, TensorFlow и др.

Для просмотра содержимого страниц платформы регистрироваться необязательно. Но для скачивания файлов с данными регистрация необходима.

Будем использовать синтетический набор данных, созданный специально для тех, кто изучает анализ данных, бизнес-аналитику и науку о данных, также он полезен и при изучении возможностей/функционала аналитических инструментов.

Оригинальное название набора – SaaS Subscription & Churn Analytics Dataset (доступен по ссылке <https://www.kaggle.com/datasets/rivalytics/saas-subscription-and-churn-analytics-dataset>).

Выполним перевод страницы с описанием набора в браузере (рис. 1.1).

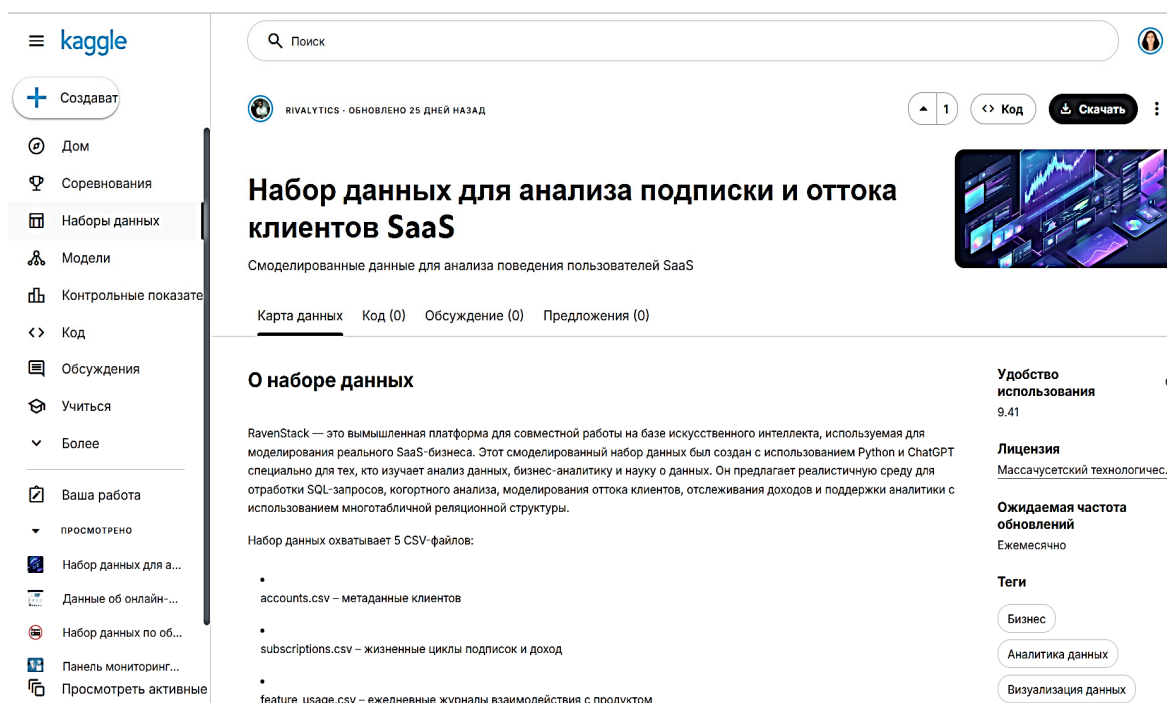


Рис. 1.1. Описание набора данных

В этот набор входит несколько файлов, доступ к которым возможен в обозревателе данных справа ниже на странице (рис. 1.2).

Можно скачать весь набор с помощью соответствующей кнопки рядом с названием набора, а можно скачать только тот файл, который открыт в обозревателе данных. Скачаем все файлы, они понадобятся в дальнейшем при рассмотрении VI-систем и других аналитических решений.

Карта данных Код (0) Обсуждение (0) Предложения (0)

README.md (7,81 кБ)

Деталь Компактный Столбец 10 из 10 столбцов

Об этом файле Предложить правки

Таблица «Учётные записи» содержит метаданные о каждом клиенте (компании), использующем платформу RavenStack. Сюда входят отрасль, страна, дата регистрации, источник рекомендации, начальный уровень тарифного плана, количество мест, статус пробной версии и информация о том, ушёл ли клиент в итоге. Эта таблица является родительской для большинства других наборов данных.

А идентификатор...	А Имя учетной зап...	А промышленность	А страна	А дата_регистрации	А ref
Нить	Нить	Нить	Нить	Дата	Нить
500 уникальные ценности	500 уникальные ценности	DevTools 23% FinTech 22% Другое (275) 55%	НАС 58% Великобритания 12% Другое (151) 30%	2023-01-02 31 декабря 2024 г.	орган
A-2e4581	Компания_0	EdTech	НАС	2024-10-16	партн
A-43a9e3	Компания_1	FinTech	В	2023-08-17	друго
A-0a282f	Компания_2	DevTools	НАС	2024-08-27	орган
A-1f0ak7	Компания_3	HealthTech	Великобритания	2023-08-27	друго
A-ce558d	Компания_4	HealthTech	НАС	2024-10-27	событ

Обозреватель данных
Версия 1 (2,07 МБ)

- README.md
- ravenstack_accounts.csv
- ravenstack_churn_events.csv
- ravenstack_feature_usage.csv
- ravenstack_subscriptions.csv
- ravenstack_support_tickets.csv

Краткое содержание

- 6 файлов
- 50 столбцов

Рис. 1.2. Содержимое файла ravenstack_accounts.csv

Если файлы были скачаны как архив, то их необходимо распаковать. Лучше организовать папку, в которую нужно разместить все пять файлов набора, извлеченных из архива (при необходимости).

В файле README.md есть описание полей каждого файла. Хорошим тоном для аналитики является оформление вводной части отчета, содержащей метаданные набора/датасета/таблицы по схеме: имя поля, тип данных и описание/детализация содержимого поля. Тип данных будет уточнен при обращении к данным в каждом аналитическом продукте отдельно.

Приведем такое описание для каждого из файлов набора.

В файле ravenstack_accounts.csv представлены следующие данные:

- account_id – идентификатор клиента (первичный ключ);
- account_name – вымышленное название компании;
- industry – отрасль (DevTools – технологии разработки, EdTech – технологии образования, FinTech – технологии финансов, Cybersecurity – кибербезопасность, HealthTech – технологии здравоохранения);
- country – код страны ISO-2 (UK – Великобритания, CA – Канада, IN – Индия, AU – Австралия, FR – Франция, DE – Германия, US – США);
- signup_date – дата создания учетной записи;
- referral_source – источник трафика, который приводит посетителей на сайт (organic, ads, event, partner, other);

- plan_tier – начальный план (Basic, Pro, Enterprise);
- seats – количество лицензированных пользователей;
- is_trial – в настоящее время пробует продукт, т.е. еще не стал полноценным клиентом;
- churn_flag – True, если клиент в настоящий момент перестал пользоваться сервисом.

В файле ravenstack_churn_events.csv представлены следующие данные (рис. 1.3):

- churn_event_id – уникальное событие ухода клиента (первичный ключ);
- account_id – ссылка на идентификатор клиента (внешний ключ);
- churn_date – когда аккаунт перестал быть действующим (клиент ушел из сервиса);
- reason_code – причина отключения аккаунта (pricing, support, features и т.д.);
- refund_amount_usd – сумма к возврату;
- preceding_upgrade_flag – True, если обновление было выполнено в течение 90 дней;
- preceding_downgrade_flag – True, если понижение рейтинга произошло в течение 90 дней;
- is_reactivation – True, если клиент вернулся после отключения от сервиса;
- feedback_text – дополнительный комментарий клиента.

Набор данных для анализа подписки и оттока клиентов SaaS

Карта данных Код (0) Обсуждение (0) Предложения (0)

README.md (7,81 КБ)

Деталь Компактный Столбец 9 из 9 столбцов

Об этом файле

Предложить правки

Таблица событий оттока фиксирует результаты оттока для аккаунтов, прекративших обслуживание. Включает дату оттока, сумму возврата (если таковая имеется), причину оттока, признак повторной активации и отзыв в свободной форме. Также отслеживается, был ли отток связан с недавним повышением или понижением тарифного плана, что помогает выявить причины оттока.

churn_event_id	идентификатор...	дата_оттока	код_причины	сумма_возврата...	pre
Нить	Нить	Дата	Нить	Плавать	Буле
600 уникальные ценности	352 уникальные ценности	2023-01-25 31 декабря 2024 г.	функции 19% поддерживать 17% Другое (382) 64%	0 393	
C-816288	Кабина A-c37	2024-10-27	ценообразование	4.83	ЛОЖЬ
C-5a81e7	A-374969	2024-06-25	поддерживать	96.45	ИСТИНА
C-a174be	A-607346	2024-11-12	бюджет	0.0	ЛОЖЬ
C-accb39	A-1e50e0	01.11.2023	бюджет	54.94	ЛОЖЬ
C-924889	A-956988	2024-12-30	неизвестный	0.0	ЛОЖЬ

Версия 1 (2,07 МБ)

- README.md
- ravenstack_accounts.csv
- ravenstack_churn_events.csv
- ravenstack_feature_usage.csv
- ravenstack_subscriptions.csv
- ravenstack_support_tickets.csv

Краткое содержание

- 6 файлов
- 50 столбцов

Рис. 1.3. Содержимое файла ravenstack_churn_events.csv

В файле `ravenstack_subscriptions.csv` представлены следующие данные (рис. 1.4):

- `subscription_id` – идентификатор подписки (первичный ключ);
- `account_id` – ссылка на идентификатор клиента (внешний ключ);
- `start_date` – дата начала подписки;
- `end_date` – дата завершения подписки (может быть пустым для активных планов);
- `plan_tier` – план на момент выставления счета;
- `seats` – количество лицензий;
- `mrr_amount` – ежемесячный доход;
- `arr_amount` – годовой доход;
- `is_trial` – `True`, если клиент в настоящее время пробует продукт, т.е. еще не стал полноценным клиентом;
- `upgrade_flag` – `True`, если план обновлен в середине цикла;
- `downgrade_flag` – `True`, если план был понижен в середине цикла;
- `churn_flag` – `True`, если клиент в настоящий момент перестал пользоваться сервисом;
- `billing_frequency` – частота выставления счетов (месяц или год);
- `auto_renew_flag` – `True`, если автоматическое обновление.

Набор данных для анализа подписки и оттока клиентов SaaS

Карта данных

Код (0)

Обсуждение (0)

Предложения (0)

1

Код

Скачать

README.md (7,81 кБ)

Деталь

Компактный

Столбец

10 из 14 столбцов

Предложить правки

Об этом файле

Таблица подписок отслеживает историю каждой подписки, привязанной к учётной записи. Включает даты начала и окончания, количество рабочих мест, частоту выставления счетов, размер дохода (MRR, ARR), статус пробной версии, а также информацию о том, повысил ли клиент тарифный план, понизил его или отказался от него в течение этого периода.

Идентификатор...	Идентификатор...	Дата начала	Конечная дата	Уровень плана	# сид
Нить	Нить	Дата	Дата	Нить	Целое
5000 уникальные ценности	500 уникальные ценности			Предприятие 34%	1
S-8сес59	A-3с1а3f	2023-12-23	2024-04-12	Предприятие	14
C-0ф6ф44	A-969фe9	2024-06-11		Про	17
C-51с0d1	A-659280	2024-11-25		Предприятие	62
S-f81687	A-e7a1e2	2024-11-23	2024-12-13	Предприятие	5
S-сff5a2	A-6a6516	2024-01-10		Предприятие	27
C-469613	A-fa2041	2024-08-13		Про	15

Обозреватель данных

Версия 1 (2,07 МБ)

README.md

ravenstack_accounts.csv

ravenstack_churn_events.csv

ravenstack_feature_usage.csv

ravenstack_subscriptions.csv

ravenstack_support_tickets.csv

Краткое содержание

6 файлов

50 столбцов

Рис. 1.4. Содержимое файла `ravenstack_subscriptions.csv`

В файле `ravenstack_feature_usage.csv` представлены следующие данные (рис. 1.5):

- `usage_id` – идентификатор уникального события использования (первичный ключ);
- `subscription_id` – ссылка на идентификатор подписки (внешний ключ);
- `usage_date` – дата использования;
- `feature_name` – одна из 40 SaaS-функций;
- `usage_count` – частота событий;
- `usage_duration_secs` – затраченное время;
- `error_count` – зарегистрированные ошибки;
- `is_beta_feature` – для функции, которая находится в стадии тестируемой, значение `True`.

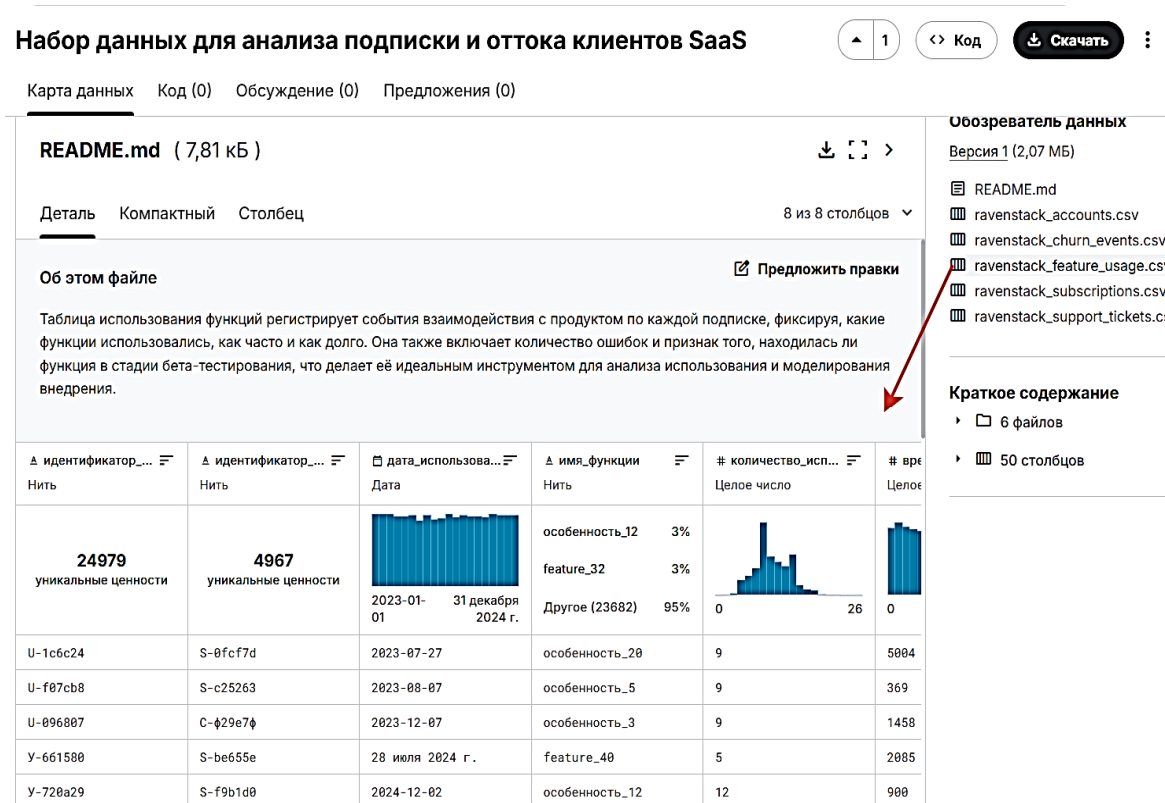


Рис. 1.5. Содержимое файла `ravenstack_feature_usage.csv`

В файле `ravenstack_support_tickets.csv` представлены следующие данные (рис. 1.6):

- `ticket_id` – идентификатор тикета (ticket/талон, зафиксированный запрос в службу технической поддержки) (первичный ключ);

- account_id – ссылка на идентификатор клиента (внешний ключ);
- shipped_at – момент открытия запроса в поддержку;
- closed_at – момент закрытия запроса в поддержку;
- resolution_time_hours – продолжительность работы с запросом;
- priority – приоритет запроса (low, medium, high, urgent);
- first_response_time_minutes – продолжительность ожидания до первого ответа на запрос;
- satisfaction_score – показатель удовлетворенности клиента от 1 до 5, если null, то клиент не оценил качество работы с его запросом;
- escalation_flag – True, если эскалация, т.е. привлечение дополнительных сотрудников и/или дополнительных ресурсов к работе над запросом, с которым возникли трудности на текущем уровне поддержки.

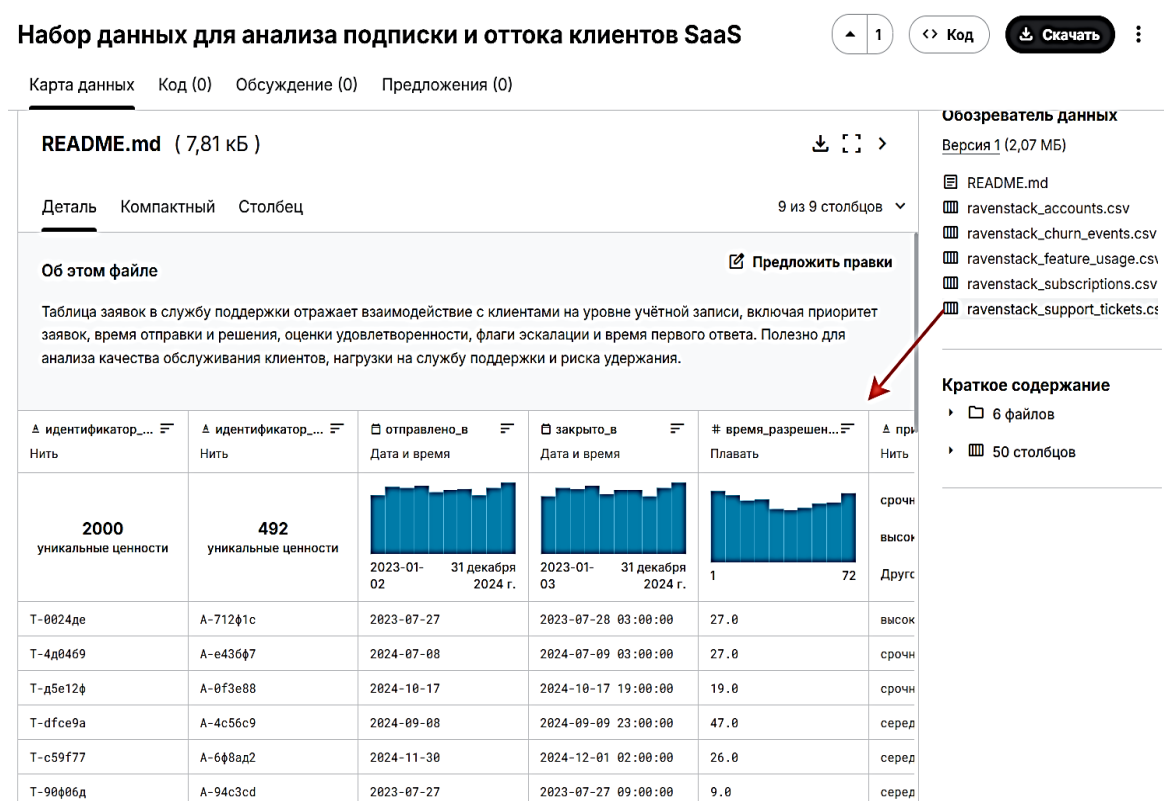


Рис. 1.6. Содержимое файла ravenstack_support_tickets.csv

Для соединения файлов необходимо использовать схему (табл. 1.3).

Таблица 1.3

Соединение наборов данных

Файл с таблицей	Присоединяемый файл с таблицей	Поля для соединения
ravenstack_accounts.csv	ravenstack_churn_events.csv	account_id (первичный ключ ravenstack_accounts.csv) account_id (внешний ключ ravenstack_churn_events.csv)
ravenstack_accounts.csv	ravenstack_subscriptions.csv	account_id (первичный ключ ravenstack_accounts.csv) account_id (внешний ключ ravenstack_subscriptions.csv)
ravenstack_accounts.csv	ravenstack_support_tickets.csv	account_id (первичный ключ ravenstack_accounts.csv) account_id (внешний ключ ravenstack_subscriptions.csv)
ravenstack_subscriptions.csv	ravenstack_feature_usage.csv	subscription_id (первичный ключ ravenstack_subscriptions.csv) subscription_id (внешний ключ ravenstack_feature_usage.csv)

При необходимости можно подобрать на ресурсе наборы данных, отвечающие определенному направлению бизнес-деятельности: логистике, продажам и т.п.

2. BI-СИСТЕМЫ

В этом разделе рассмотрим несколько отечественных решений: Yandex DataLens, GlarusBI и Visiology.

Yandex DataLens – это облачное решение от Яндекса, предназначенное для быстрой визуализации данных и подготовки отчетности. Его ключевая задача – упростить работу с данными для бизнеса, помогая специалистам быстрее получать инсайты и повышать качество принимаемых решений.

Сервис Yandex DataLens обладает рядом особенностей, делающих его удобным инструментом для компаний, которые только начинают настраивать аналитические процессы:

- простота использования. Интерфейс сервиса понятен даже тем, кто ранее не работал с инструментами бизнес-аналитики;
- высокая производительность. Сервис быстро обрабатывает большие объемы данных и строит графики и таблицы;
- автоматическое обновление данных. Возможна настройка автоматического подключения к внешним источникам данных;
- безопасность. Данные хранятся на надежных серверах Яндекса с высоким уровнем шифрования и защитой от несанкционированного доступа;
- масштабируемость. Тарифные планы гибкие и позволяют увеличивать число пользователей.

Однако сервис имеет определенные ограничения:

- автономная работа не поддерживается, необходим доступ к Интернету для работы с сервисом;
- пока в сервисе небольшое количество встроенных алгоритмов машинного обучения.

Visiology – инструмент визуальной аналитики и представления данных, разработанный одноименной российской компанией «Визиолоджи». Платформа разрабатывалась с учетом потребностей бизнеса и госсектора и широко в них представлена. Основное внимание уделяется визуальному отображению данных и удобной навигации по ним. Также в платформе имеется собственный аналитический движок, что ускоряет обработку больших массивов данных.

Сильные стороны BI-системы Visiology:

- интерактивные дашборды. Пользователи могут взаимодействовать с графиками и картами прямо на экране (в браузере),

фильтруя данные, увеличивая интересующие зоны и создавая собственные виджеты;

- комплексный подход. Решение охватывает весь цикл работы с данными: сбор, хранение, обработку и представление информации;

- высокие стандарты информационной безопасности. Сервис разработан с учетом строгих требований государственных структур и промышленных предприятий к защите данных;

- развертывание на инфраструктуре заказчика. Доступ в Интернет необязателен;

- использование языка запросов DAX для расчетов в платформе;

- встроенный ETL-инструмент, который позволяет обрабатывать и подготавливать данные для загрузки на платформу;

- возможность кастомизации виджетов и создание различных сценариев использования дашбордов с помощью Java Script;

- хранение и расчет данных в собственном хранилище, что обеспечивает более быструю загрузку данных для пользователей и быструю обработку при взаимодействии пользователя с дашбордом.

К ограничениям Visiology можно отнести:

- более сложный в освоении интерфейс по сравнению с DataLens;

- меньшую эффективность при работе с динамическими данными в режиме реального времени.

GlarusBI – отечественная платформа для комплексного анализа данных и визуализации. Создана российской компанией IBS Group, широко используется в финансовой сфере, телекоммуникациях и промышленном секторе.

Сильные стороны GlarusBI:

- универсальность и широкий набор инструментов. Доступны удобные инструменты для ETL-процесса (Extract, Transform, Load), консолидации данных и анализа многомерных кубов;

- есть возможность расширения функционала через написание скриптов на Python, что добавляет гибкости и открывает новые перспективы для анализа данных;

- система способна интегрироваться с различными СУБД и файлами определенных форматов (.xls, .csv, .json);

– возможно как развертывание на собственной инфраструктуре компании, так и работа через веб-интерфейс.

К ограничениям стоит отнести несколько более высокий порог вхождения, чем в случае с DataLens, система больше подходит для опытных аналитиков и разработчиков, хотя не исключен вариант использования в качестве первой системы, с которой компания начнет путь в аналитике.

Основные особенности продуктов представим в табл. 2.1.

Таблица 2.1

**Функциональные возможности Yandex DataLens,
Visiology и GlarusBI**

Аспект сравнения	BI-система		
	Yandex DataLens	Visiology	GlarusBI
Назначение	Быстрая визуализация данных	Интерактивные дашборды, работа с большими объемами данных	Универсальная BI-платформа
Подключение к источникам данных	Да, стандартно	Да, через плагины и стандартно	Да, встроено
Автоматическое обновление данных	Да	Да	Да
Безопасность данных	Высокий уровень	Высокий уровень	Высокий уровень
Расширяемость	Скрипты, шаблоны	Плагины, скрипты Java Script	Через Python API
Машинное обучение	Базовые алгоритмы	Нет	Через Python

Каждый из рассмотренных инструментов решает разные задачи и предназначен для разной аудитории:

- простота и быстрота запуска проекта – Yandex DataLens;
- нужны мощные интерактивные дашборды и удобное восприятие информации руководителями и сотрудниками компании – Visiology;

– требуются гибкость и глубокие возможности для детального анализа данных – GlarusBI.

А теперь рассмотрим особенности работы в каждом продукте.

2.1. Аналитика с Yandex DataLens

Для начала работы в сервисе Yandex DataLens необходимо создать аккаунт в почте @yandex.ru (регистрация в любом сервисе Яндекса приводит к созданию Yandex ID, позволяющего пользоваться всеми сервисами экосистемы).

Yandex DataLens (далее просто DataLens) позиционируется как корпоративный инструмент визуализации данных, который эффективнее всего обслуживает данные в виде таблиц, заранее подготовленных в какой-либо базе данных.

Ограничения функционала DataLens:

- не подходит для формирования строгих отчетных форм (финансовой, бухгалтерской, налоговой отчетности);
- нет гибких возможностей работы со сложной инфографикой (способной представить информацию в виде последовательных/иерархических графических структур с текстовыми пояснениями);
- нет возможности хранить данные;
- не получится создать ETL-процедуры (Extract, Transform, Load), процесс, с помощью которого данные из разных источников приводят к единому виду и собирают в одном месте;
- нет API для выгрузки данных.

DataLens – это часть экосистемы Yandex Cloud, в которой отдельные сервисы отвечают за перечисленный функционал, отсутствующий в DataLens.

Так, для ETL-пайплайнов предназначен Airflow, а трансформация и перемещение данных возможны с помощью Data Transfer. Углубленная аналитика и методы машинного обучения поддерживаются специальной платформой DataSphere.

Можно рассматривать Yandex Cloud как конструктор, который позволяет создать нужную для аналитики конфигурацию, подключив необходимые сервисы и платформы и оплатив только то, что действительно необходимо (рис. 2.1).

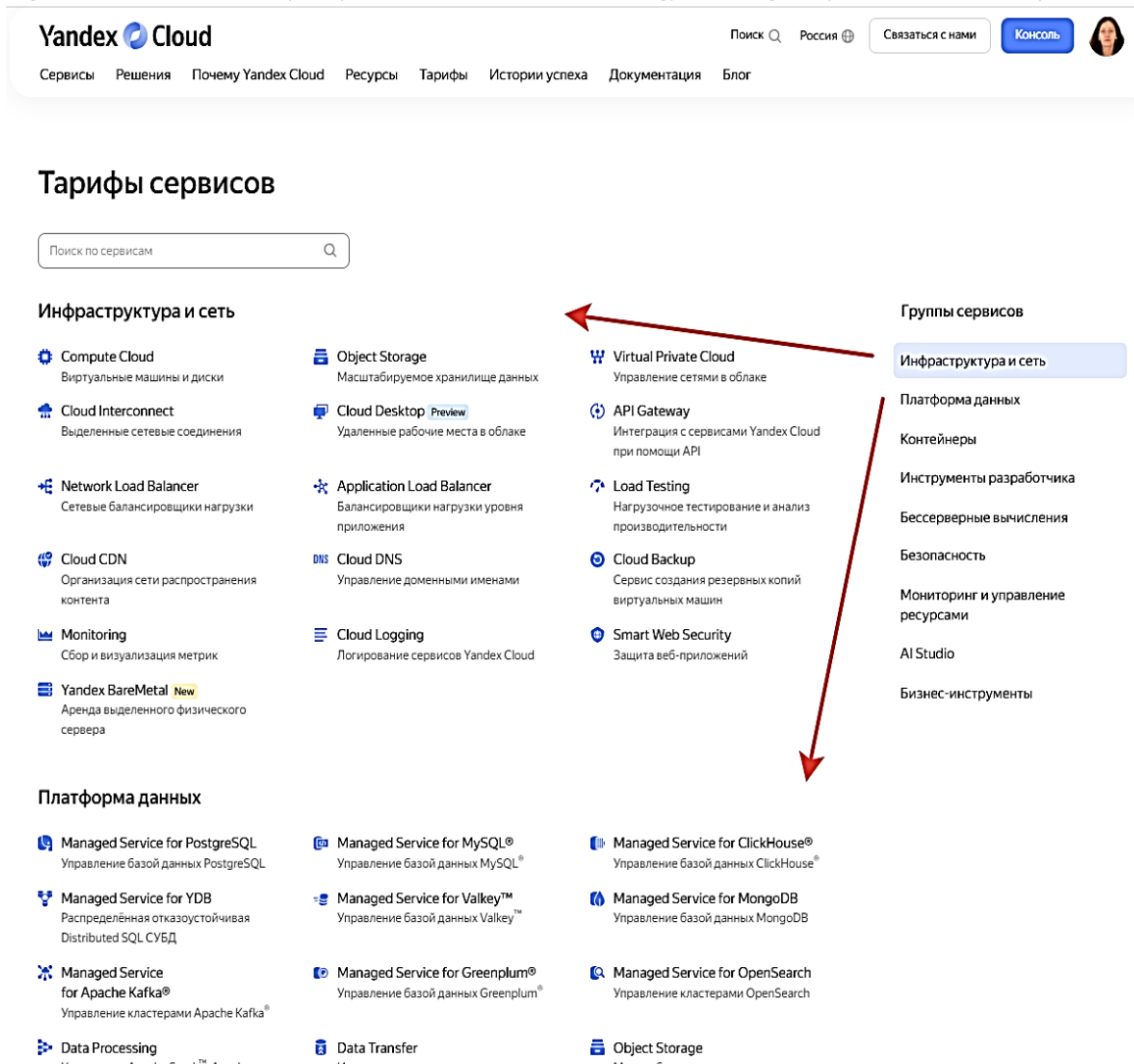


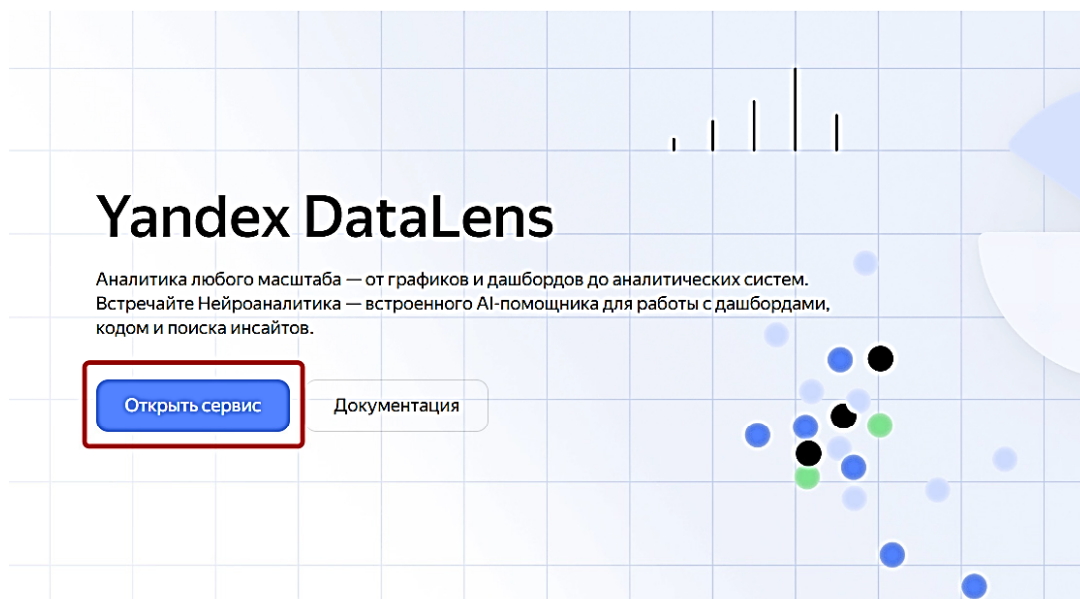
Рис. 2.1. Панель сервисов Yandex Cloud

Тарифные планы DataLens:

- Community подходит для небольших команд и некоммерческих проектов (не тарифицируется, т.е. доступ без оплаты);
- Business подходит для корпоративных внедрений и решения бизнес-задач (ежемесячная подписка на одного активного пользователя составляла 990 рублей летом 2025 г.).

Для учебных целей возможностей доступа без оплаты в варианте Community вполне достаточно.

Для работы в сервисе по адресу <https://datalens.yandex.cloud/> (рис. 2.2) необходимо авторизоваться с аккаунтом @yandex.ru (рис. 2.3).



Преимущества



Нейроаналитик SOON

AI-агент повышает эффективность команды, сокращает рутину и ищет скрытые инсайты.



Быстрый старт

Активируйте сервис, подключите к своему источнику и постройте первый дашборд сегодня.

Рис. 2.2. Вход в сервис DataLens

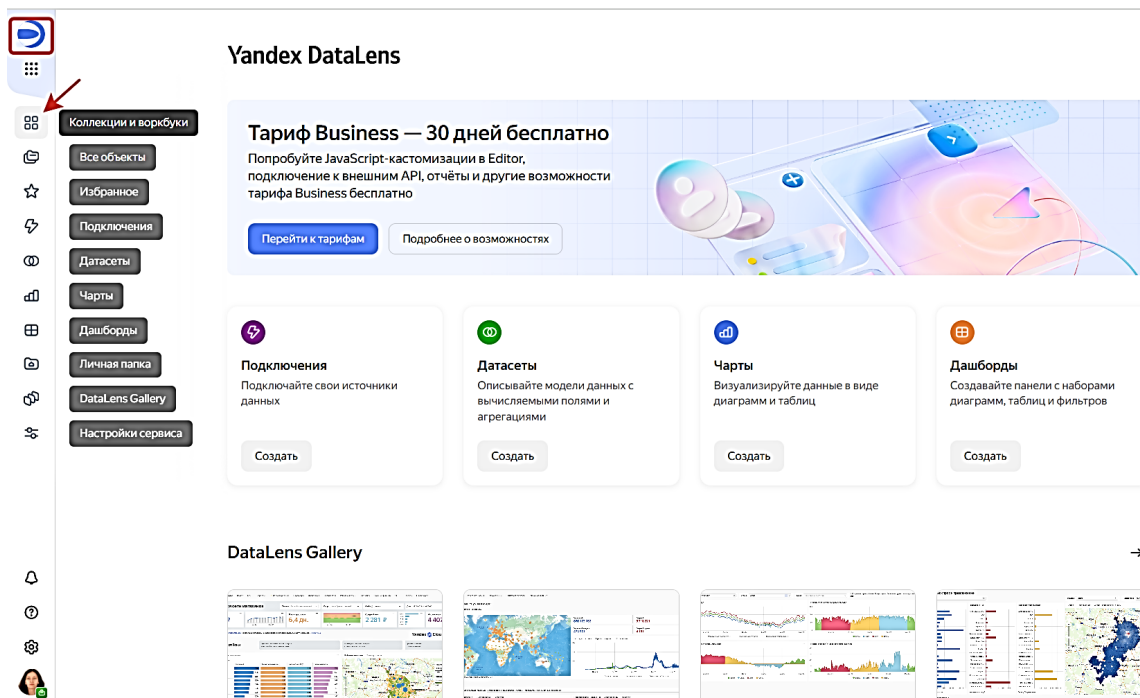


Рис. 2.3. Домашняя страница сервиса после авторизации

Крайняя левая кнопка на вертикальной панели навигации отвечает за переход на домашнюю страницу. При наведении на пиктограмму на левой панели появляются подписи ко всем пиктограммам, а та, на которую направлен курсор, выделяется цветом.

Как видим, на главной странице доступны все основные сущности, которые обеспечивают полный цикл работы с данными в DataLens (рис. 2.4):

- подключение – набор параметров для доступа к источнику данных;
- датасет (dataset, набор данных) – описание набора данных из источника;
- чарт (chart, диаграмма) – визуализация данных из датасета в виде таблиц, диаграмм и карт;
- дашборд (dashboard, приборная панель) – набор чартов, селекторов для фильтрации данных и текстовых блоков.

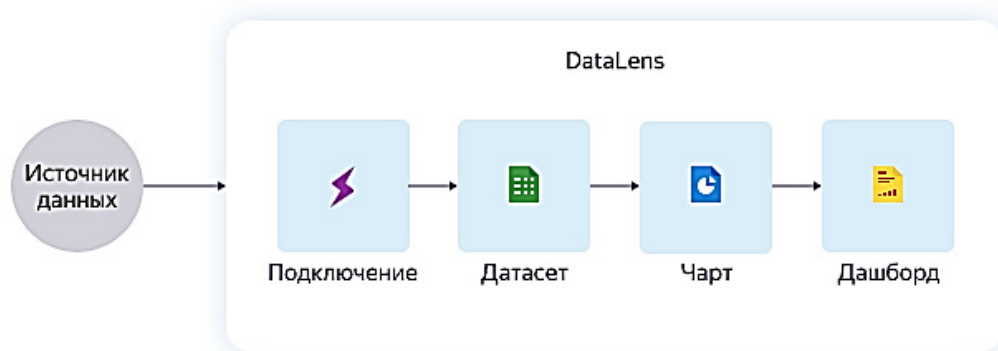


Рис. 2.4. Сущности DataLens

Объекты в DataLens можно размещать в специальных контейнерах. В воркбуках можно размещать объекты, а в коллекции можно сохранять воркбуки и другие коллекции.

Начнем работу в DataLens с создания воркбука с именем Анализ клиентов RavenStack (рис. 2.5–2.7).

На странице созданного воркбука (рис. 2.7) доступны для создания объекты, которые будут сохранены в него. Для выполнения операций с самим воркбуком, а не с размещенными в нем объектами необходимо перейти с помощью левой панели навигации на страницу Коллекции и воркбуки и раскрыть дополнительное меню рядом с именем воркбука, к которому нужно применить какую-либо операцию, например удаление (рис. 2.8).

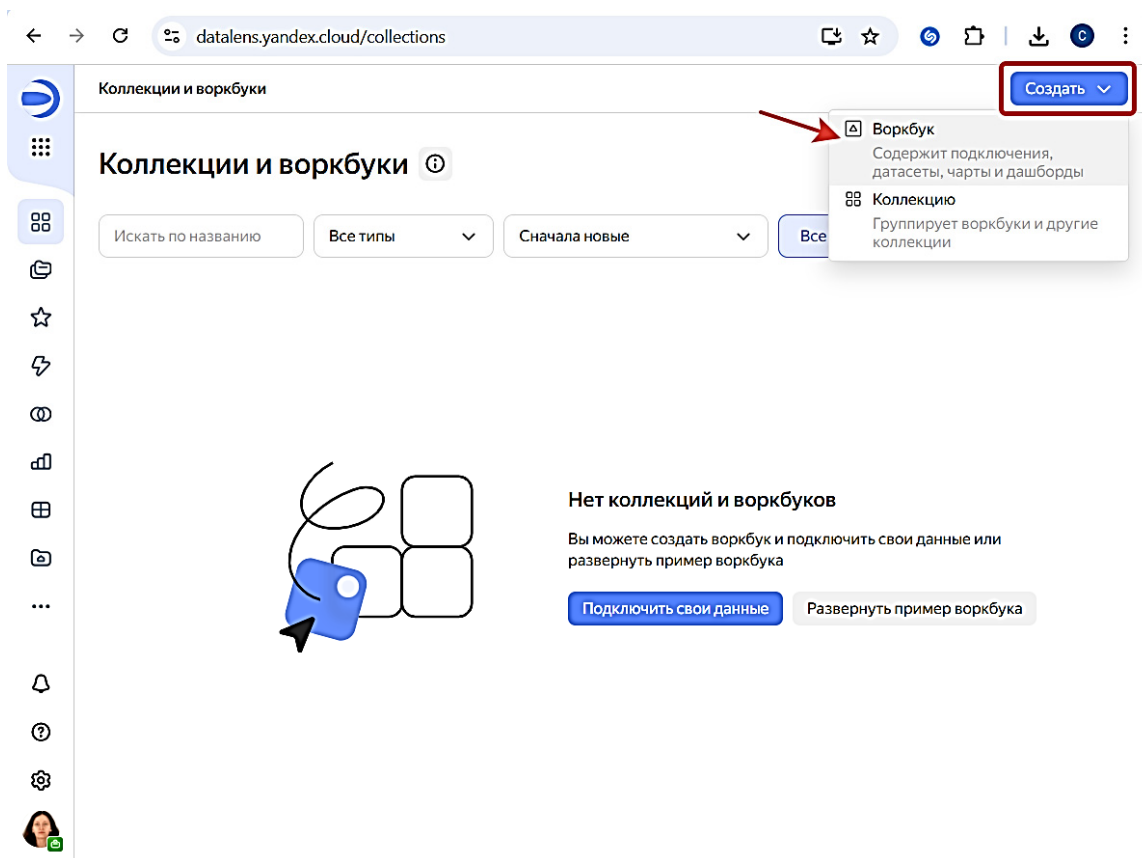


Рис. 2.5. Создание воркбука на странице Коллекции и воркбуки

The screenshot shows the 'Создать воркбук' (Create Workbook) dialog box. It has a title bar with a close button (X). The 'Название' (Name) field contains the text 'Анализ клиентов RavenStack'. The 'Описание' (Description) field contains the text 'Для учебных целей'. Below the description field, there is a section 'Импорт из файла' (Import from file) with a question mark icon and a button 'Выбрать файл' (Choose file). At the bottom right, there are two buttons: 'Отменить' (Cancel) and 'Создать' (Create). The 'Создать' button is highlighted with a red rectangle.

Рис. 2.6. Настройки для воркбука (название, описание)

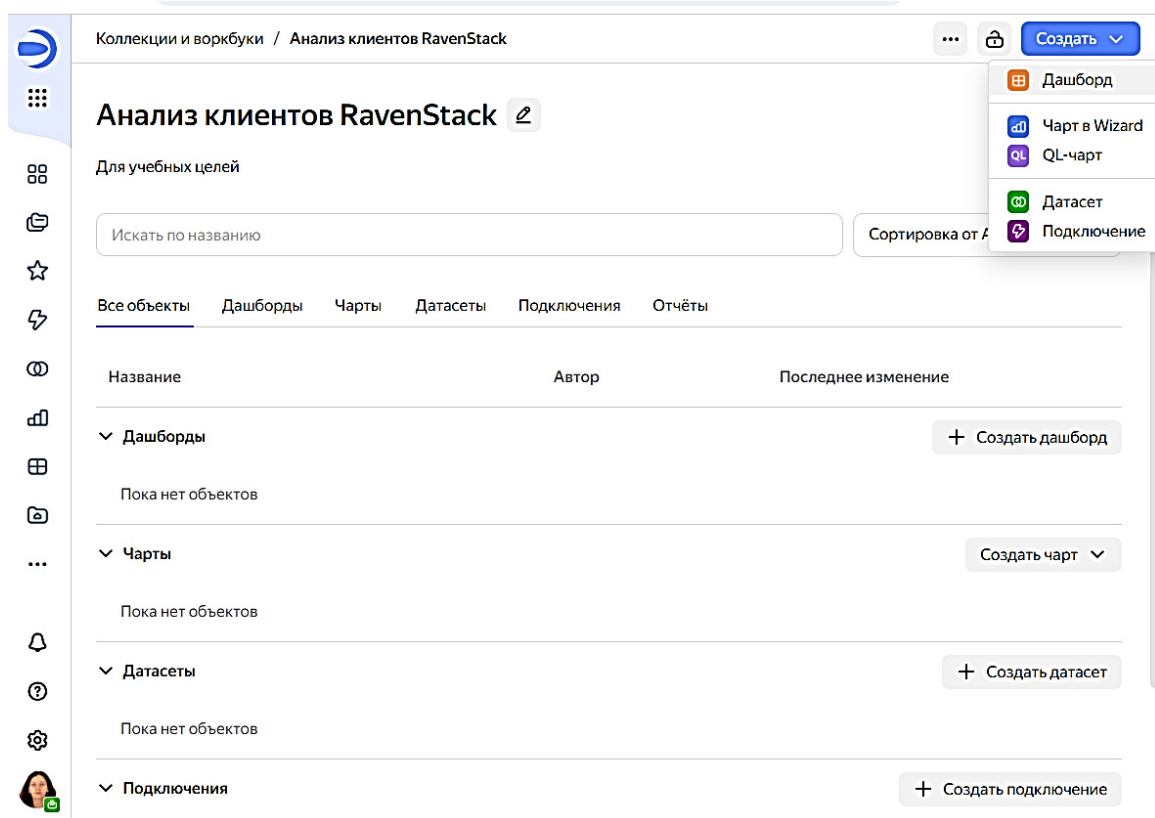


Рис. 2.7. Переход на страницу созданного воркбука

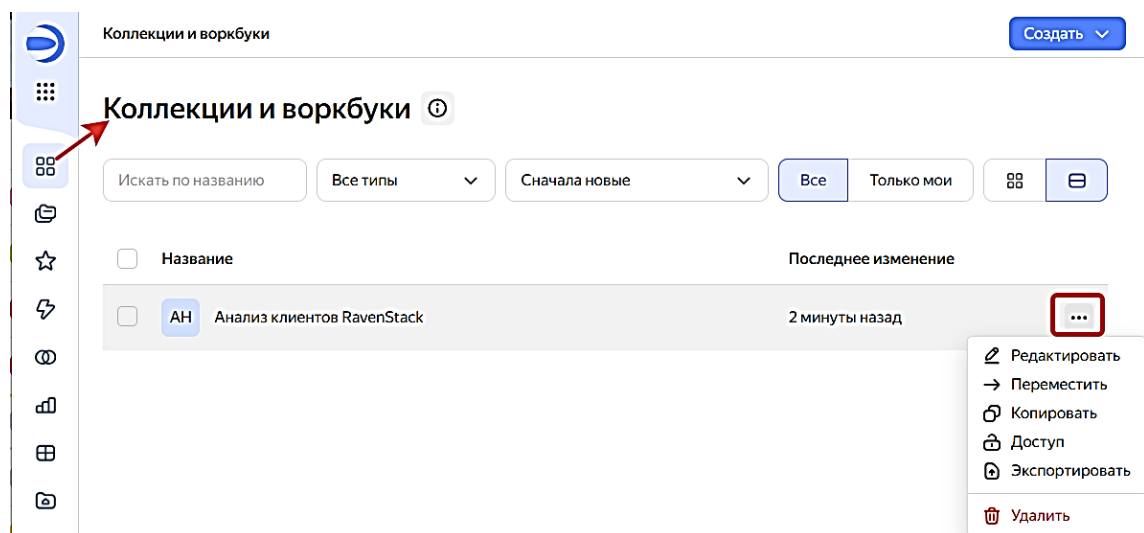


Рис. 2.8. Коллекции и воркбуки

В воркбуке Анализ клиентов RavenStack будем сохранять объекты, создаваемые для набора данных с kaggle.com (см. раздел 1.3).

На основе файлов набора будем создавать подключения. Обратите внимание, что копирование воркбуков с подключением к файлам не поддерживается.

Создавать подключение, обращаясь к файлу в архиве, нельзя, он должен быть предварительно распакован.

На странице воркбука Анализ клиентов RavenStack каким-либо способом приступим к созданию подключения. В настройках нового подключения выберем источник данных (рис. 2.9). Очевидно, что это файл, который нужно выбрать на своем диске для загрузки на сервер Yandex DataLens (рис. 2.10).

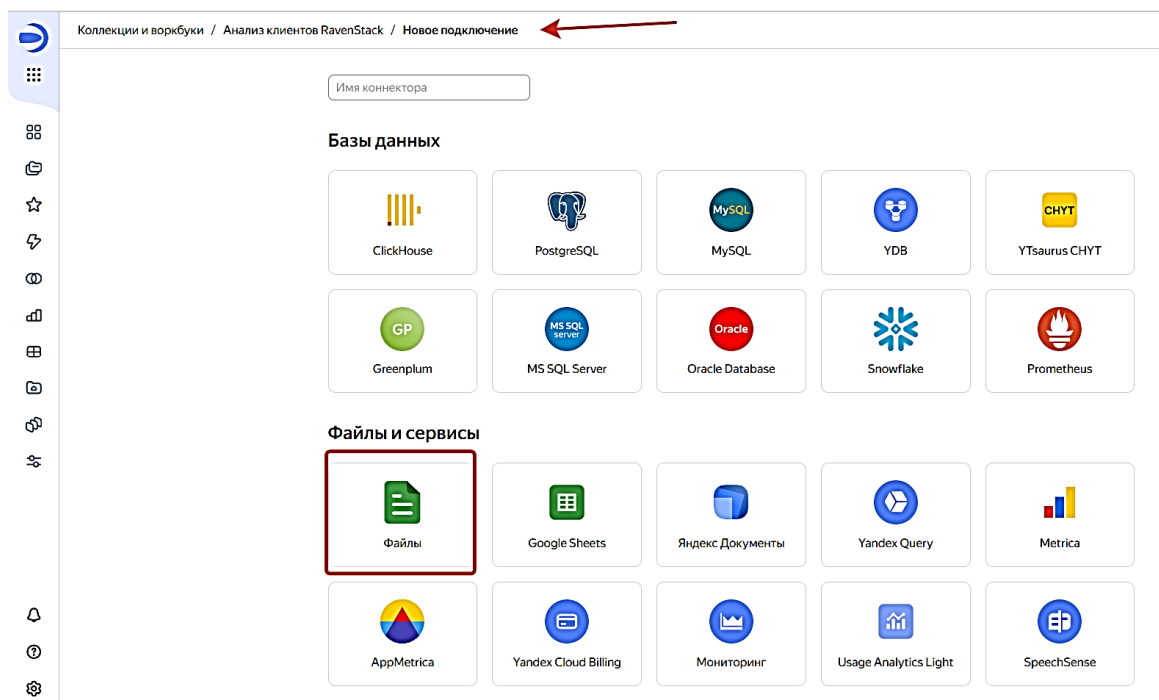


Рис. 2.9. Выбор источника данных для подключения

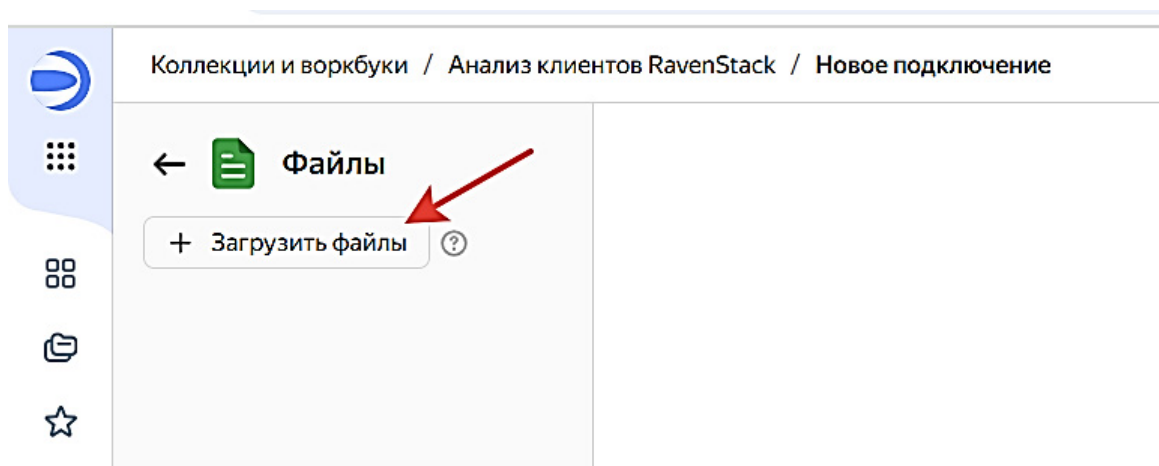


Рис. 2.10. Загрузка файла на сервер

Загрузим файл ravenstack_accounts.csv. После загрузки появится информация о новом подключении: кодировка и разделитель

данных, так как данные загружаются в формате csv – это плоские таблицы с разделителями данных в строке, которые являются текстовыми файлами. Также на предварительный просмотр будет выведено начало таблицы (рис. 2.11). Рядом с именем поля отобразится тип данных: текстовый, например `account_id`, числовой (`seats`) и дата/время (`signup_date`).

account_id	account_name	industry	country	signup_date	referral_source	plan_tier	seats	is_trial	churn_flag
A-2e4581	Company_0	EdTech	US	2024-10-16	partner	Basic	9	False	False
A-43a9e3	Company_1	FinTech	IN	2023-08-17	other	Basic	18	False	True
A-0a282f	Company_2	DevTools	US	2024-08-27	organic	Basic	1	False	False

Рис. 2.11. Предварительный просмотр содержимого загруженного файла

Настроенное подключение необходимо сохранить. Дадим ему имя Аккаунты, которое потом можно изменить, если в этом возникнет необходимость (рис. 2.12).

Создать подключение

Название: Аккаунты

Отменить Создать

Рис. 2.12. Сохранение настроенного подключения

После сохранения сервис перемещает фокус активности на страницу этого объекта (рис. 2.13) и можно создать связанный объект – следующий в последовательности процесса аналитики (см. рис. 2.4).

Перейдем на вкладку Поля и познакомимся с доступными настройками. Поле, которое не планируется использовать, можно скрыть. Также для каждого поля доступны операции в дополнительном меню, которые позволяют дублировать, редактировать поле и даже удалить. При этом в новом датасете, основанном на том же подключении, удаленное поле опять будет доступно для манипуляций. Каждый датасет, основанный на одном и том же подключении, может иметь свои собственные настройки, которые не влияют ни на подключение, ни на другие датасеты.

Создадим датасет с таким же именем Аккаунты. Объекты одного и того же воркбука, принадлежащие к одному типу, не могут иметь одинаковые названия. Но так как подключение и датасет — это объекты разного типа, то дать им одно и то же название допустимо. На странице датасета доступно создание нового объекта — чарта. Создадим чарт и приступим к его настройке (рис. 2.15).

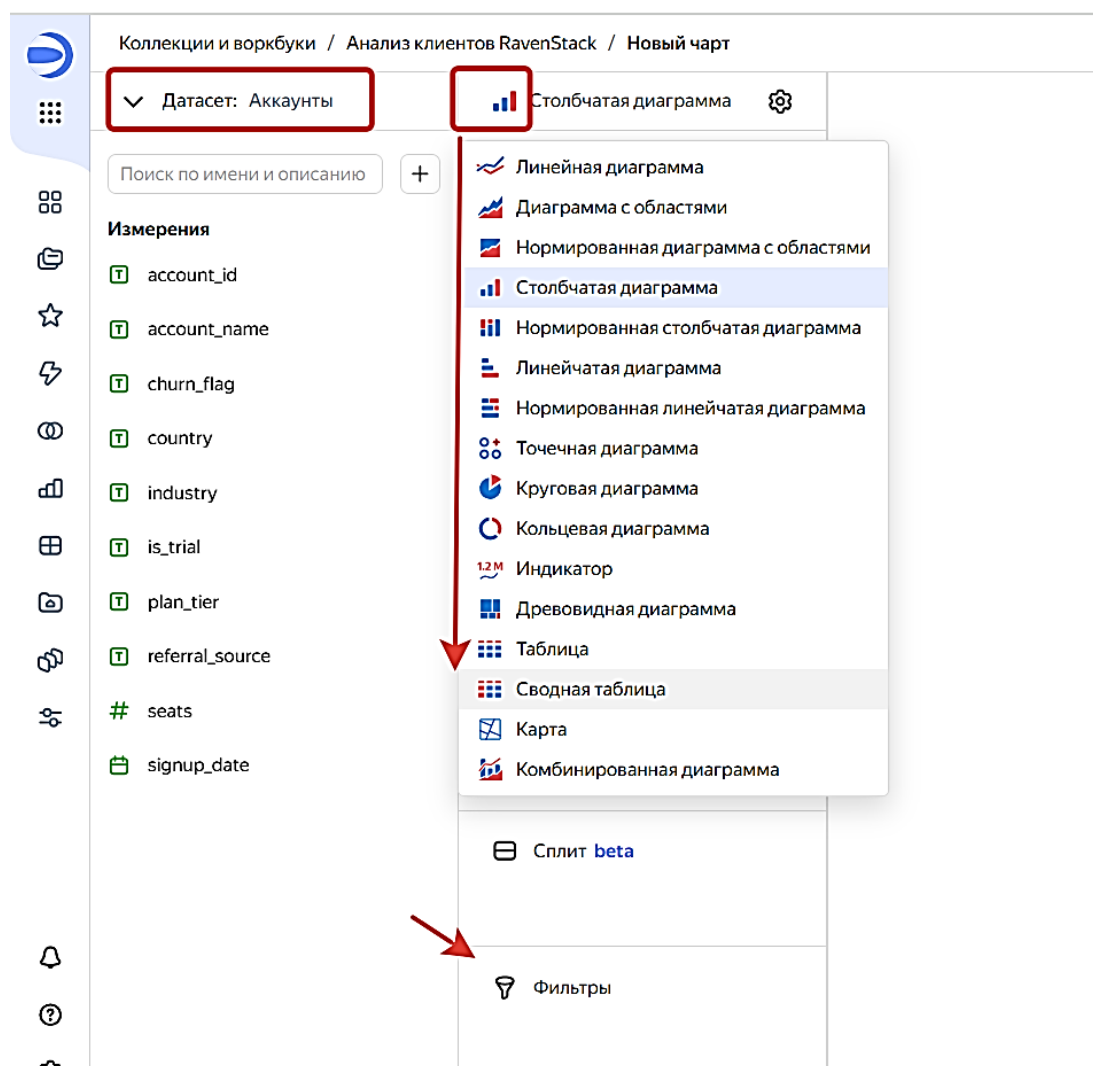


Рис. 2.15. Настройки для чарта

Так как чарт создан со страницы датасета, то выбор датасета уже осуществлен, но его можно изменить по стрелке слева от названия. Так как пока в наличии единственный датасет, эта опция бесполезна. Ниже располагаются поля датасета, которые можно использовать при построении чарта. По умолчанию предлагается вид чарта – столбчатая диаграмма, изменим вид на сводную таблицу. Настройки для сводной таблицы представлены на рис. 2.16.

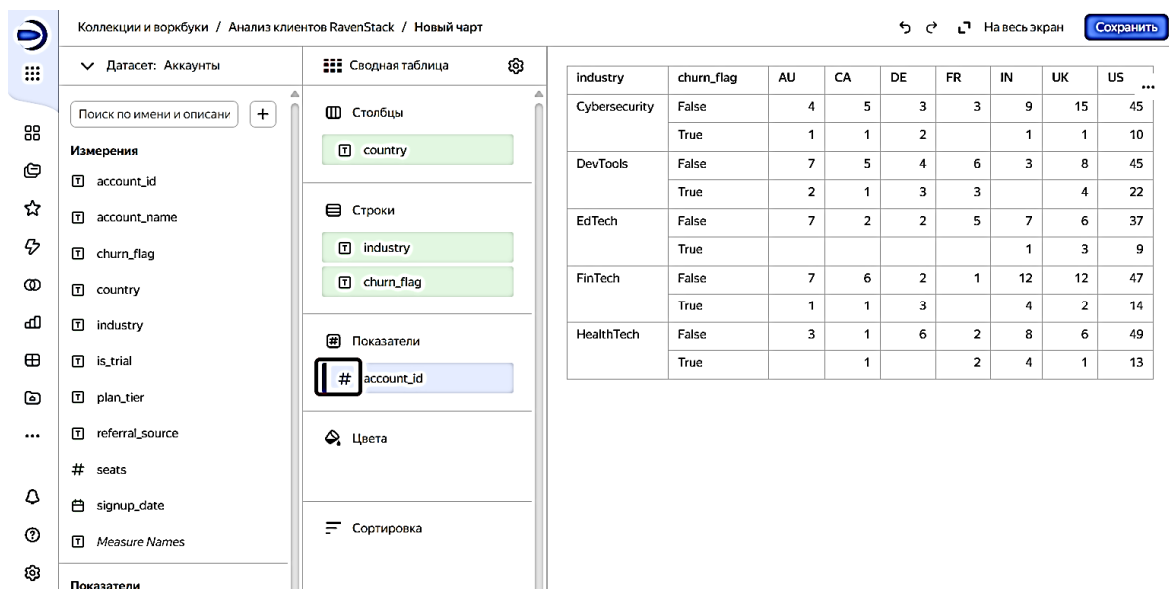


Рис. 2.16. Сводная таблица по числу настоящих и прекративших сотрудничество клиентов в разрезе стран и индустрий

Обратите внимание, что текстовый показатель `account_id` преобразован в количественную метрику и показывает число уникальных значений. Показатель в сводной таблице может быть только числовым, поэтому сервис подбирает подходящую метрику для соответствующего поля.

В то же время в качестве полей для столбцов и для строк сводной таблицы выбираются категориальные (текстовые) поля с небольшим числом уникальных значений.

Если выбрать в качестве столбца или строки сводной таблицы поле `account_id`, получим разреженную таблицу с числом столбцов или строк, равным числу наблюдений (для используемых данных это будет 500).

В DataLens есть два вида применения условий к данным: фильтры действуют на уровне чарта и на постоянной основе, а селекторы (с ними будем иметь дело позднее) настраиваются для дашбордов, являются динамическими, т.е. их условия можно изменять

в процессе просмотра, и могут быть применены сразу к нескольким чартам, размещенным на дашборде.

Сохраним сводную таблицу с названием Отток клиентов по странам и отраслям (рис. 2.17). Если сервис отказывается сохранять объект с введенным названием, то причиной может быть либо совпадение с уже имеющимся в воркбуке аналогичным объектом с тем же названием (при нестабильном Интернете нужно проверять, не был ли уже создан объект в момент якобы пропавшего соединения с сервисом), либо использование запрещенного символа, например «/».

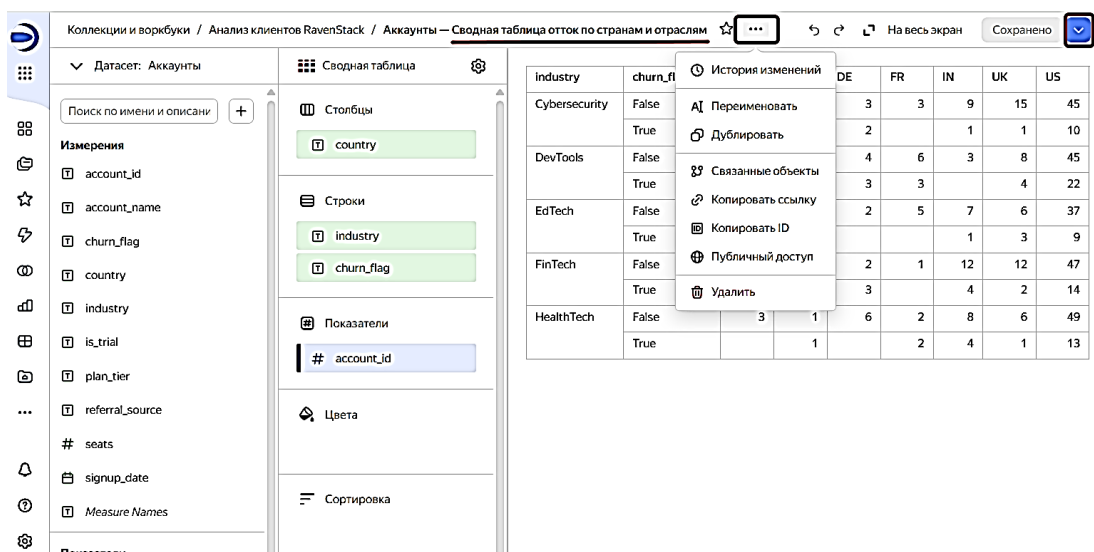


Рис. 2.17. Страница чарта

При внесении изменений в чарт станет активной кнопка Сохранить, также можно сохранить открытый чарт как копию или черновик.

Создадим новый чарт. Это можно сделать различными способами: с главной страницы, со страницы воркбука, со страницы датасета. Для последнего варианта операция Создать чарт будет выполнена с подстановкой текущего датасета. Настройки для нового датасета представлены на рис. 2.18 и 2.19.

Ось X – показатель signup_date, для которого настроена группировка по месяцам. Настройки для показателей производятся в диалоговом окне, которое вызывается кликом по маркеру типа данных рядом с именем показателя. Ось Y – account_id, переводится в числовой формат подсчетом числа уникальных значений. Также добавлена подсветка цветом, сколько из зарегистрировавшихся в соответствующем месяце в настоящий момент уже перестали быть клиентами.

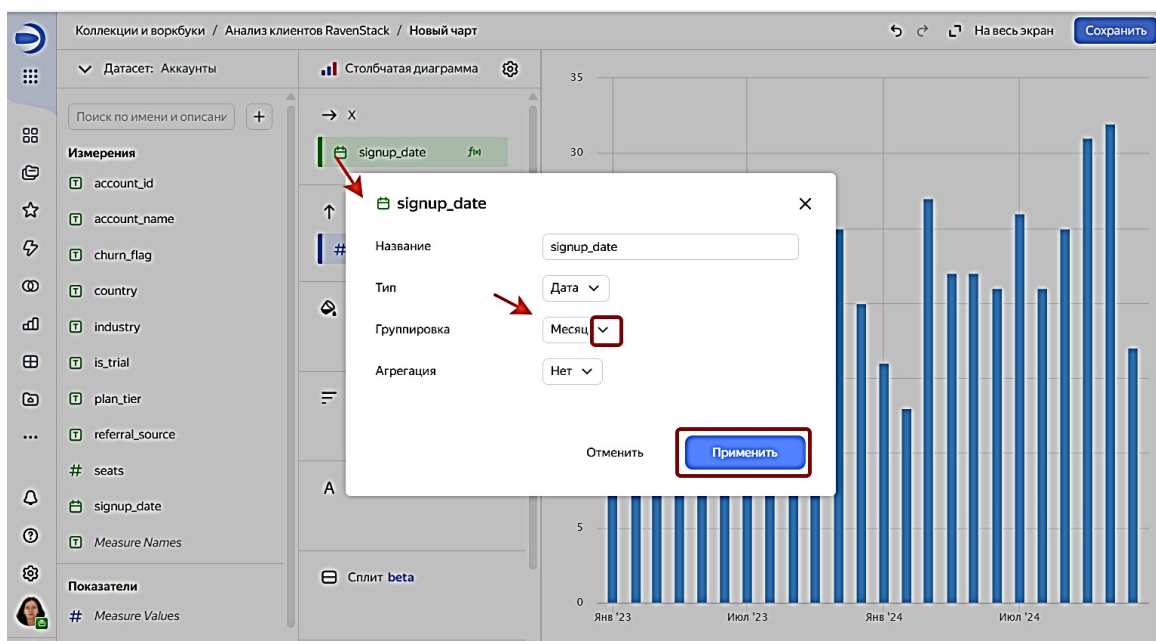


Рис. 2.18. Настройки для показателя signup_date

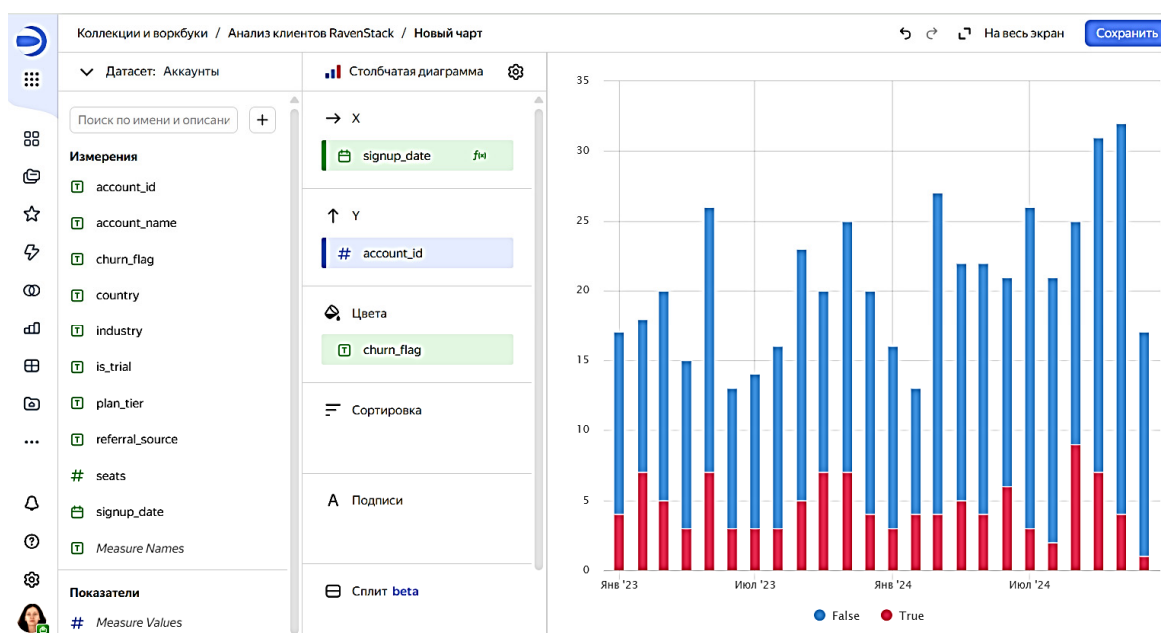


Рис. 2.19. Подключение цветового разделения столбцов на основе значений показателя churn_flag

Сохраним новый чарт с названием Аккаунты – Столбчатая диаграмма подключений новых клиентов по месяцам.

Создадим два новых чарта, используем вид Индикатор. Для первого чарта добавим в настройки фильтра показатель churn_flag (рис. 2.20) и выберем значение False (рис. 2.21). Сохраним настроенный чарт с названием Аккаунты – Индикатор количества действующих лицензий.

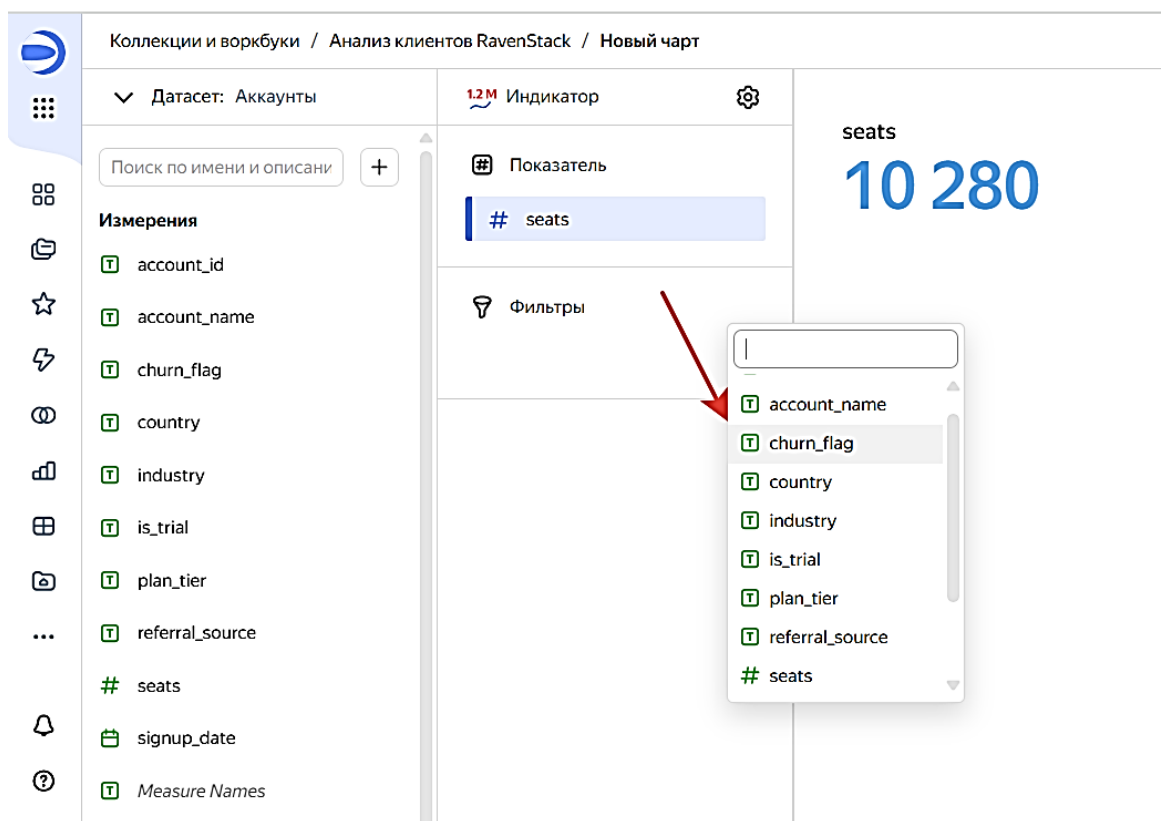


Рис. 2.20. Добавление показателя в фильтр

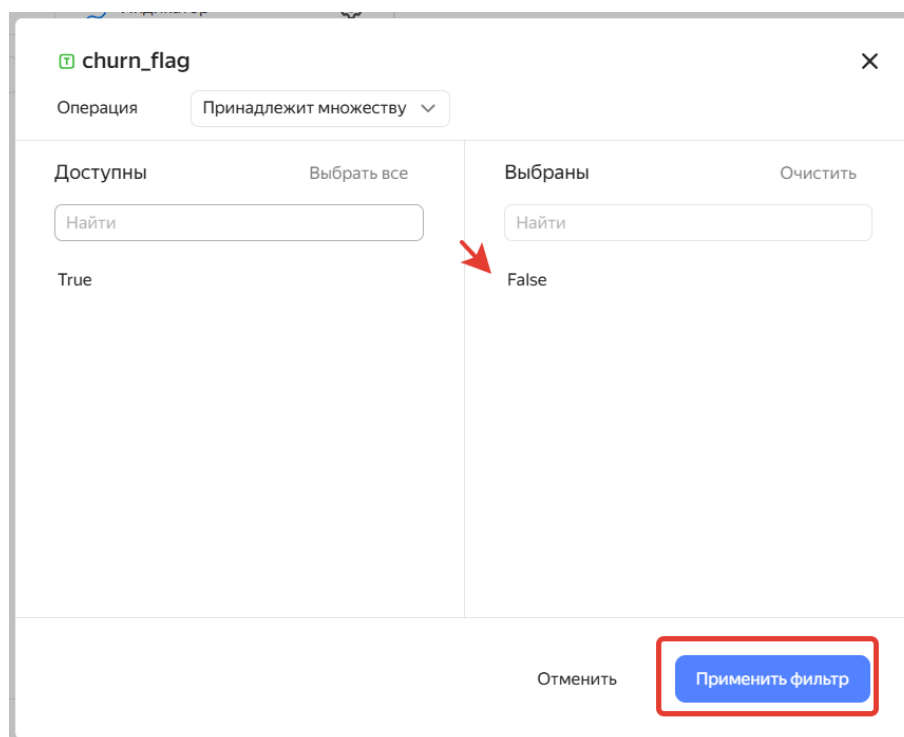


Рис. 2.21. Настройки для показателя churn_flag, используемого в фильтре

Так как второй чарт будет дублем только что созданного, только с измененным значением для фильтра, создадим копию, дав ей название Аккаунты – Индикатор количества отозванных лицензий (рис. 2.22).

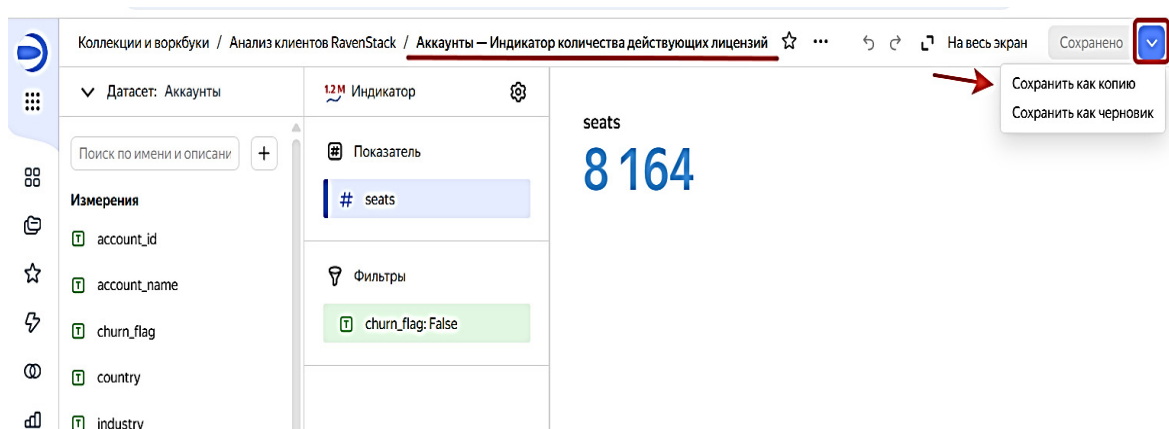


Рис. 2.22. Создание копии чарта

Изменим настройки фильтра и обязательно сохраним изменения в чарте (рис. 2.23).

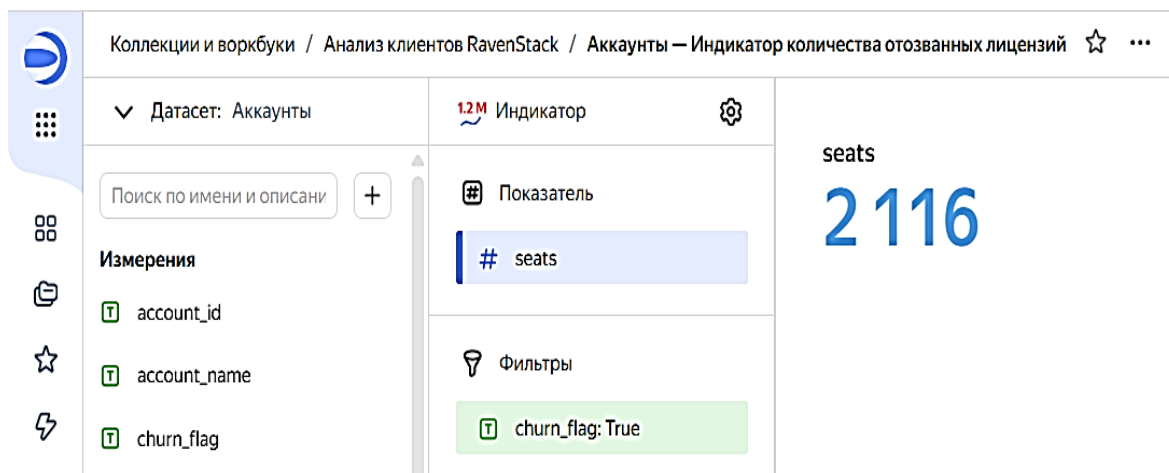


Рис. 2.23. Чарт Индикатор количества отозванных лицензий

Перейдем в воркбук Анализ клиентов RavenStack и создадим дашборд, на котором будем размещать созданные чарты (рис. 2.24, 2.25).

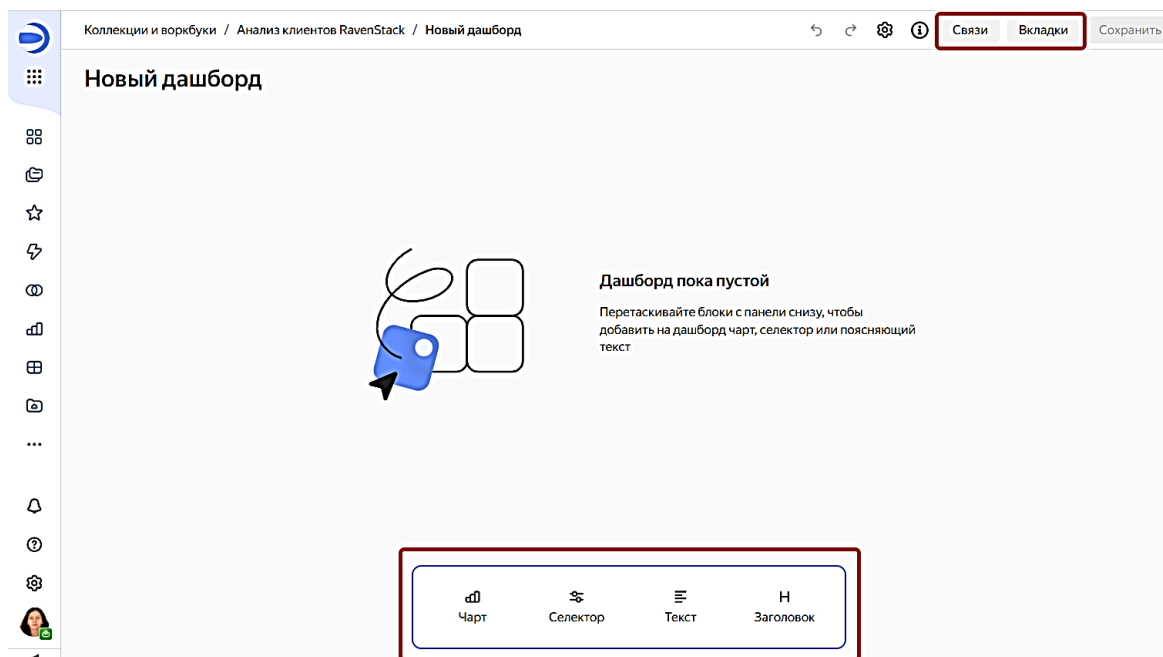


Рис. 2.24. Создание нового дашборда

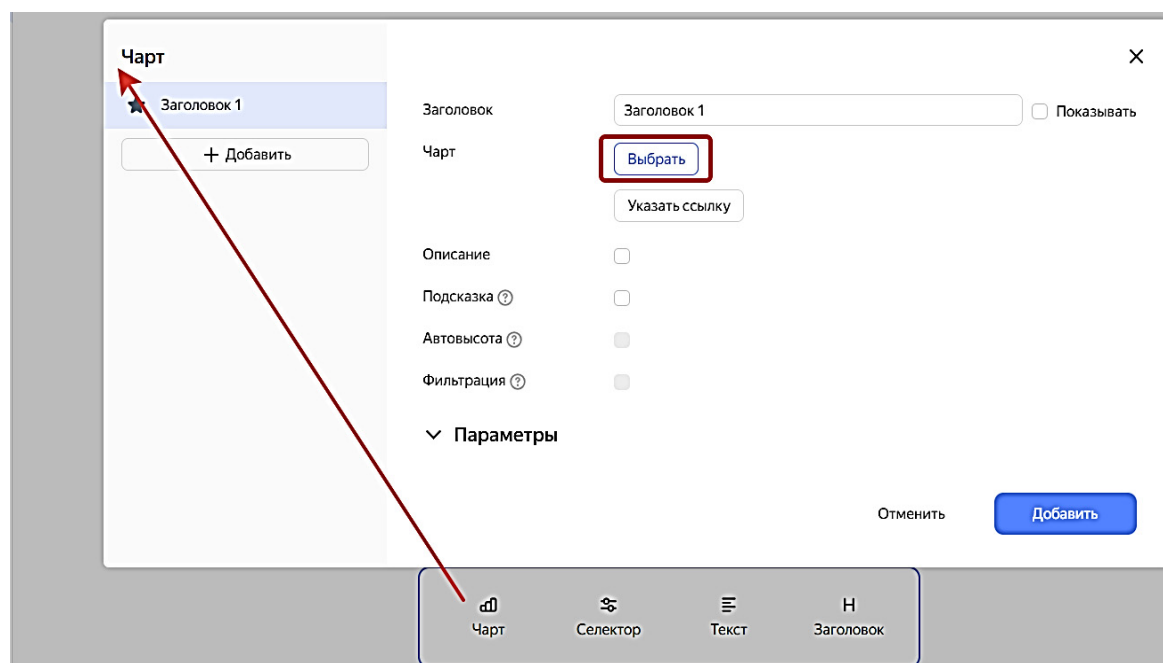


Рис. 2.25. Размещение на дашборде ранее созданных чартов

После размещения объектов на дашборде его необходимо сохранить (рис. 2.26) и делать это в дальнейшем после каждого редактирования.

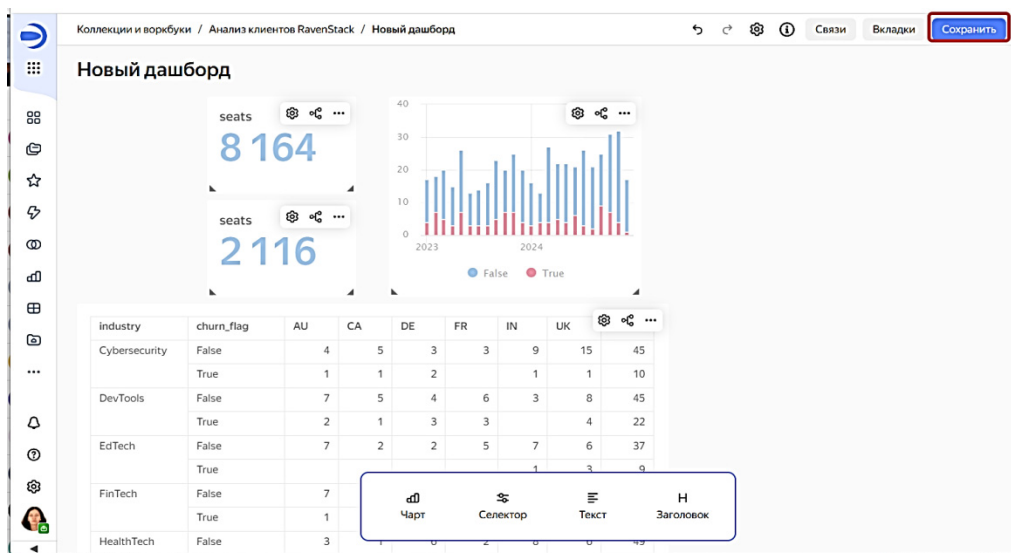


Рис. 2.26. Сохранение дашборда под именем Клиенты RavenStack

С помощью кнопок Редактировать/Сохранить происходит смена режимов редактирования и просмотра. В режиме редактирования меняются настройки, добавляются объекты, в режиме просмотра выбираются значения селекторов и знакомятся с визуальным представлением данных.

Для каждого объекта на дашборде присутствует доступ к его индивидуальным настройкам, которые происходят в этом объекте. Перейдем к настройкам первого из двух индикаторов (рис. 2.27) и изменим подпись, не забудем нажать кнопку Сохранить в этом чарте, чтобы изменения были приняты.

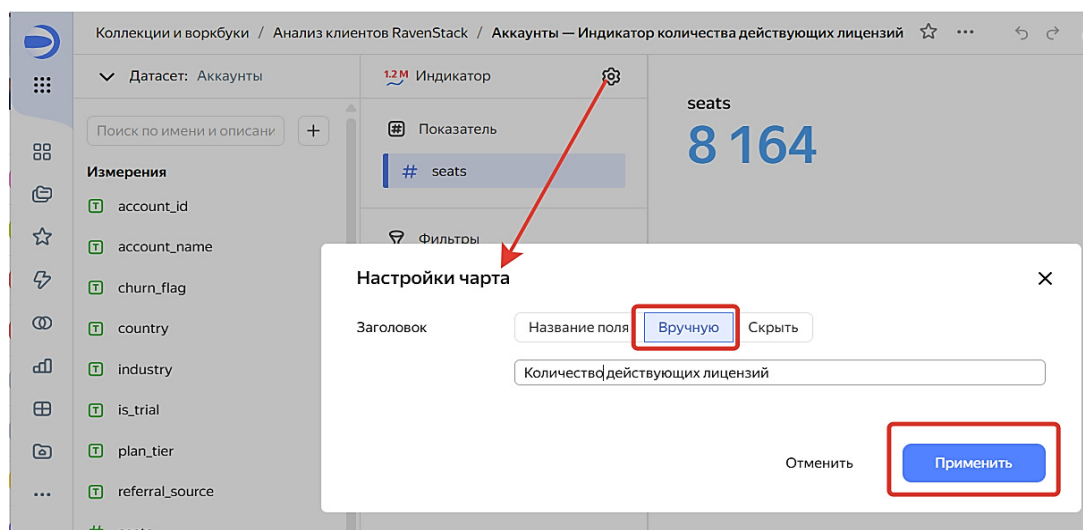


Рис. 2.27. Изменение настроек индикатора

Возвратимся в дашборд (рис. 2.28) и проведем аналогичные изменения настроек для всех объектов.

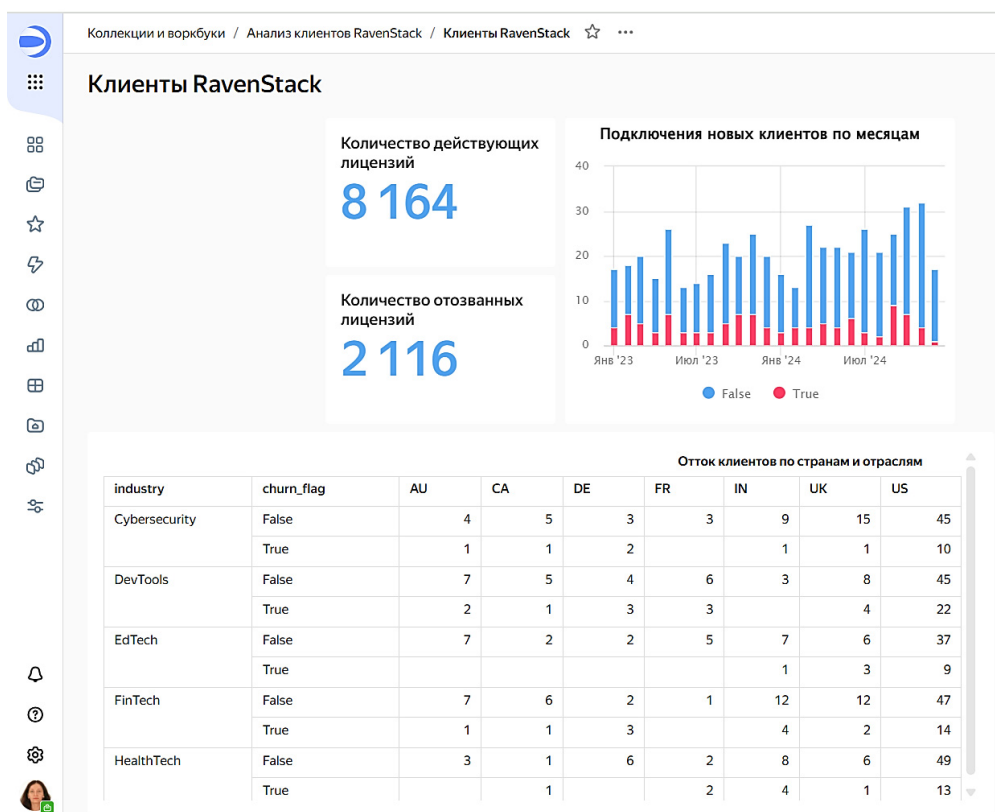


Рис. 2.28. Отображение дашборда с измененными настройками размещенных на нем объектов

В режиме редактирования добавим два селектора. Первый селектор по полю country (рис. 2.29, 2.30).

Селекторы

Селектор 1

+ Добавить селектор

Фильтр по имени

Аккаунты 16.08.2025

Настройки селектора

Данные

Источник: На основе датасета

Выбрать

Указать ссылку

Выбрать

Список

Не определено

Селектор 1

Отменить Сохранить

Рис. 2.29. Добавление селектора (выбор датасета)

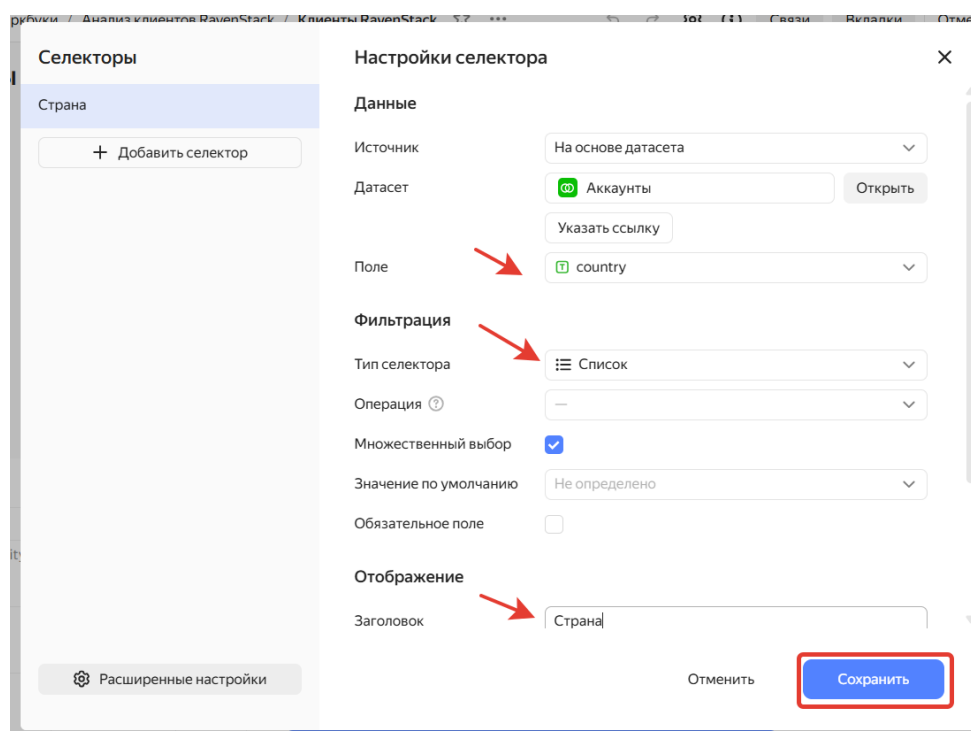


Рис. 2.30. Настройки для селектора по полю country

Аналогично проведем добавление селектора по полю industry. Также создадим селектор на основе поля signup_date (рис. 2.31), сохраним дашборд и в режиме просмотра выберем значения для селекторов (рис. 2.32).

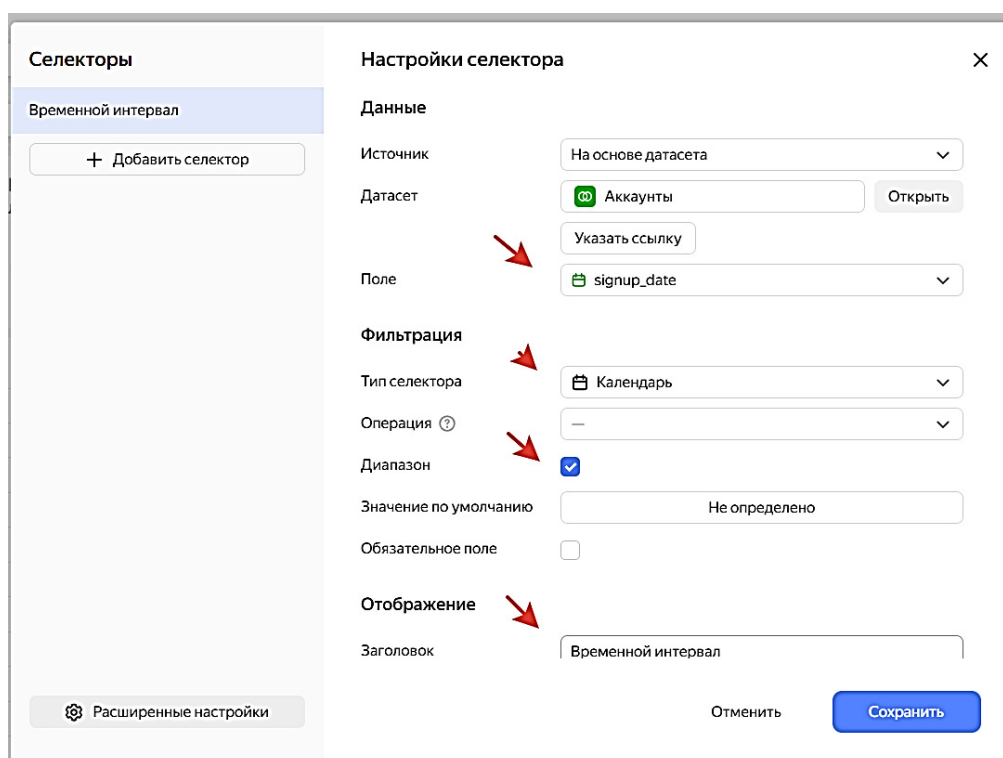


Рис. 2.31. Настройки для селектора по полю signup_date

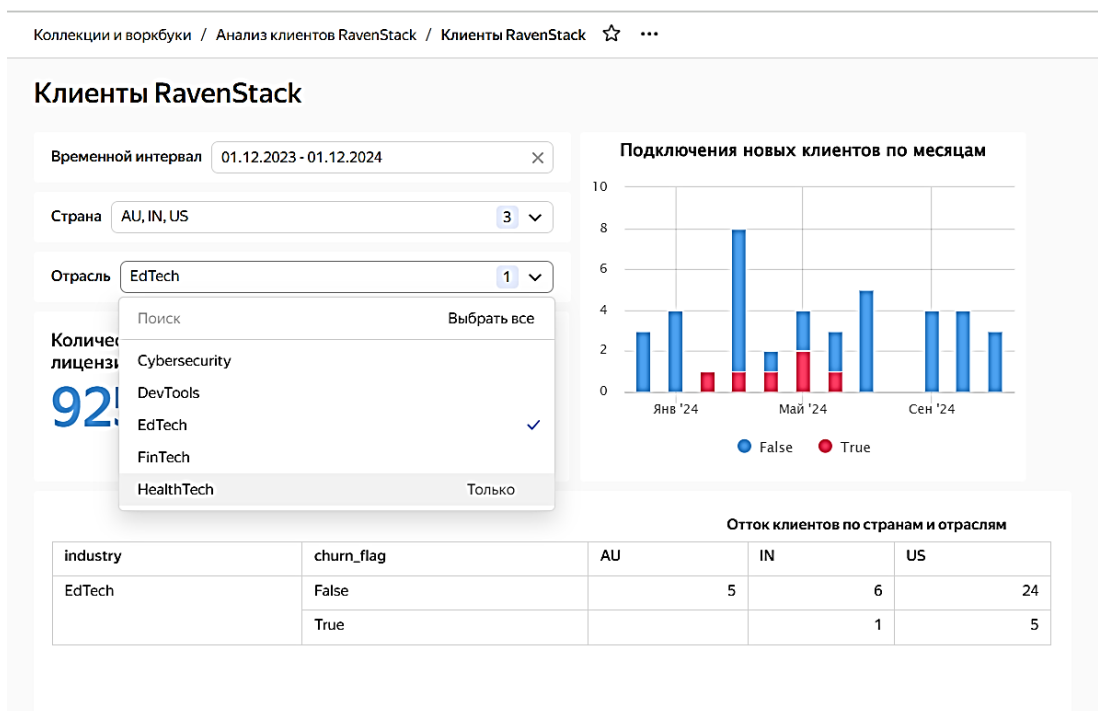


Рис. 2.32. Выбор значений селекторов в режиме просмотра дашборда

Для дальнейшего совершенствования навыков работы в Yandex DataLens можно воспользоваться справочной документацией [2] и пособием [3].

Самостоятельное задание:

1. Создайте новое подключение, загрузив еще один файл набора. Для датасета на основе этого подключения создайте несколько чартов и разместите их на том же дашборде Клиенты RavenStack, организовав дополнительную вкладку.

2. Создайте подключение, загрузив два файла, которые можно соединить (см. табл. 1.3). Для датасета настройте соединение файлов без потери информации, скройте дубликат поля, по которому произошло соединение таблиц. Для объединенной таблицы данных создайте несколько чартов и разместите их на новом дашборде.

2.2. Аналитика с Visiology

Рассмотрим базовые возможности платформы Visiology (<https://ru.visiology.su/>) для аналитики (рис. 2.33).

На сайте Visiology можно получить демодоступ (<https://ru.visiology.su/why/demo>) для знакомства с возможностями BI-системы (рис. 2.34).

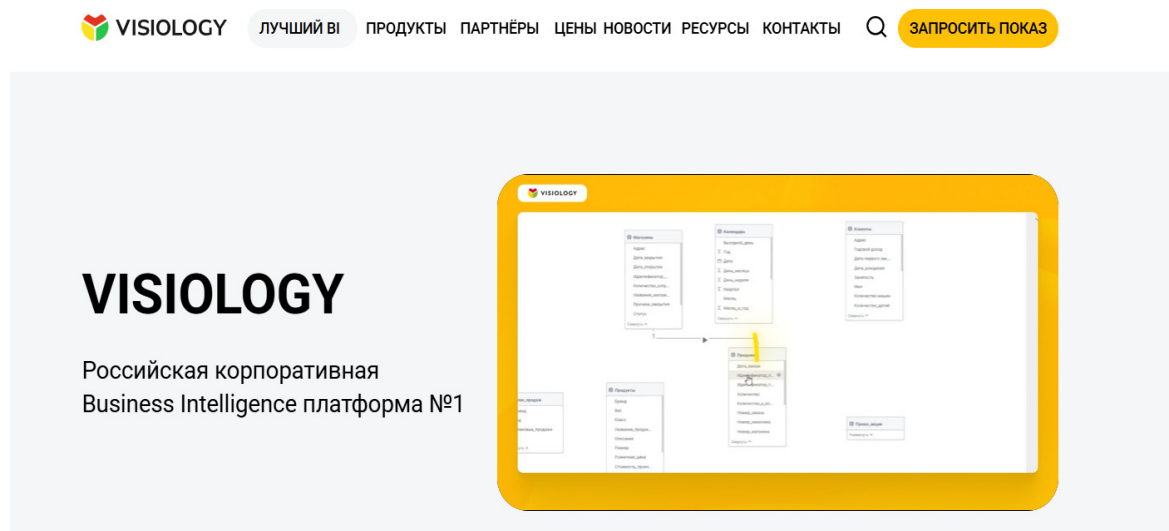


Рис. 2.33. Официальный сайт Visiology

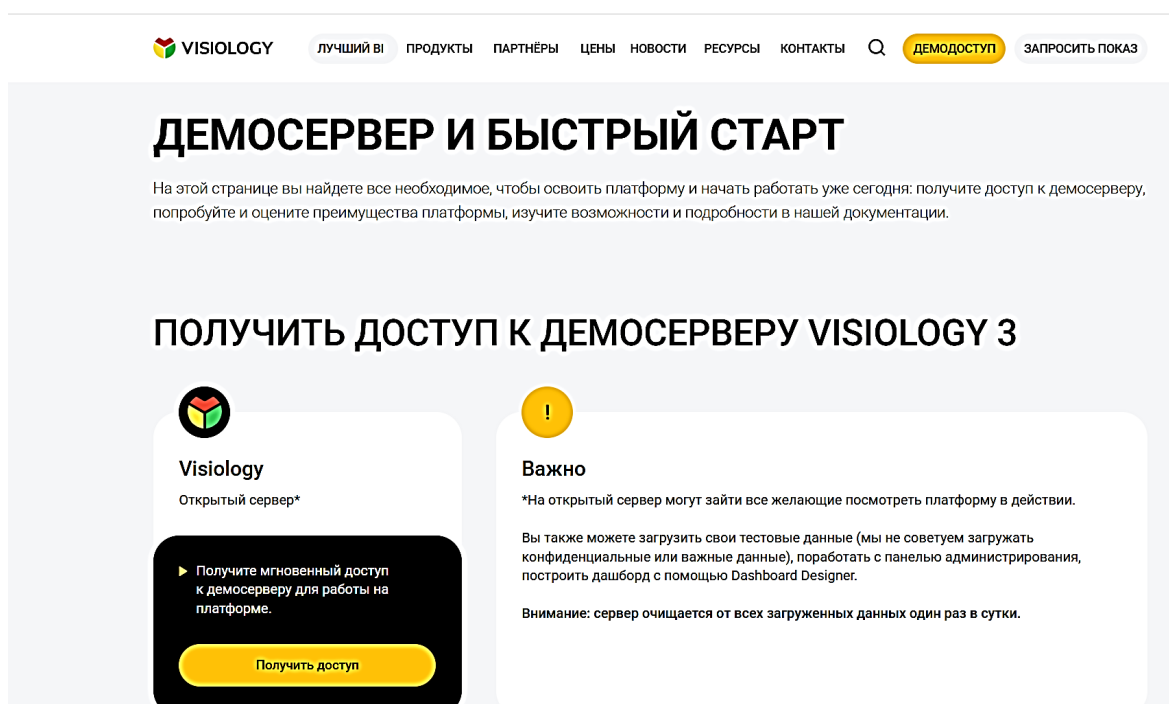


Рис. 2.34. Доступ к демосерверу Visiology

В рамках сотрудничества с образовательными учреждениями по академической программе предоставляется программное обеспечение для развертывания платформы на сервере под управлением операционной системы Linux. Клиентская часть платформы – Visiology Designer – реализована как веб-приложение, поддерживающее загрузку данных, создание и настройку дашбордов.

Платформа Visiology внесена в реестр отечественного программного обеспечения, а по результатам исследования вакансий на рекрутинговой платформе hh.ru признана самой популярной BI-платформой в России (в 2024 г.).

Базовые возможности Visiology заключаются в создании дашбордов, в этом платформа похожа на другие системы этого класса программных решений. Системы конкурируют по таким критериям, как удобство использования, безопасность данных, требования к программному и аппаратному окружению и т.п.

Visiology можно развернуть на своей инфраструктуре, что является привлекательной характеристикой для компаний, которые заинтересованы в повышенных мерах безопасности цифрового пространства и усиленного контроля над данными.

Платформа Visiology включает уникальные компоненты, составляющие ее конкурентное преимущество:

- аналитический движок ДанКо, который обеспечивает бесперебойную загрузку данных из множества источников, быструю обработку больших объемов информации, разделение доступа, представление аналитики в любых форматах;

- виртуальный аналитик ViTalk GPT. В основе ViTalk GPT лежат технологии NLP, машинного обучения и математической лингвистики. ViTalk GPT не ищет ответы в уже созданных отчетах, а самостоятельно строит простые отчеты, соответствующие запросу;

- встроенный ETL-инструмент, который позволяет обрабатывать и подготавливать данные для загрузки на платформу и проводить обработку данных, использует PostgreSQL;

- система сбора и обогащения данных Smart Forms, которая позволяет настраивать сбор и консолидирования данных ручного ввода.

Рассмотрим базовые возможности Visiology в построении дашбордов на примере демодоступа на онлайн-платформе (рис. 2.35).

Стандартный аналитический процесс на платформе включает в себя:

- 1) создание набора данных:
 - подключение к источнику данных или загрузку файлов с данными;
 - создание модели данных (при необходимости);
- 2) создание дашборда – визуализаций на основе набора данных.

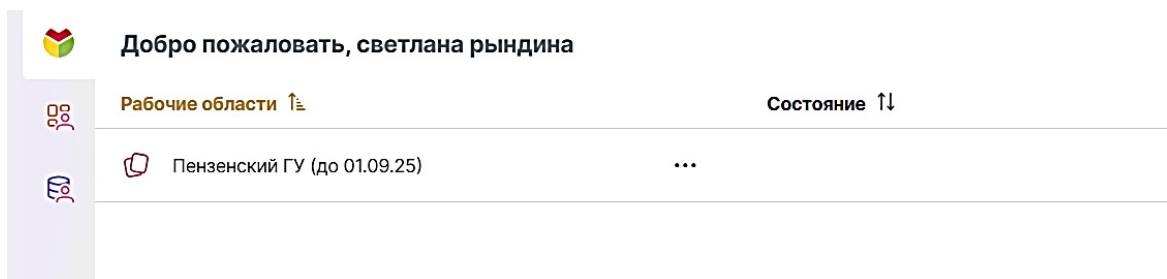


Рис. 2.35. Рабочая область для выполнения аналитических процессов пользователем

Для доступа к платформе используется ролевая модель – одна из важных особенностей платформы, позволяющая реализовать контроль над аналитическими процессами и усилить безопасность: каждый пользователь в системе обладает только теми доступами/правами, которые необходимы и определены администратором в рамках его рабочих обязанностей, что снижает вероятность выполнения ошибочных операций из-за недостаточной компетенции пользователей и ограничивает контакты с чувствительными для компании данными.

На главной странице платформы администратором настроен доступ к рабочей области (см. рис. 2.35).

В рабочем пространстве доступны для создания три типа объектов: наборы данных, дашборды, регламентные отчеты (рис. 2.36).

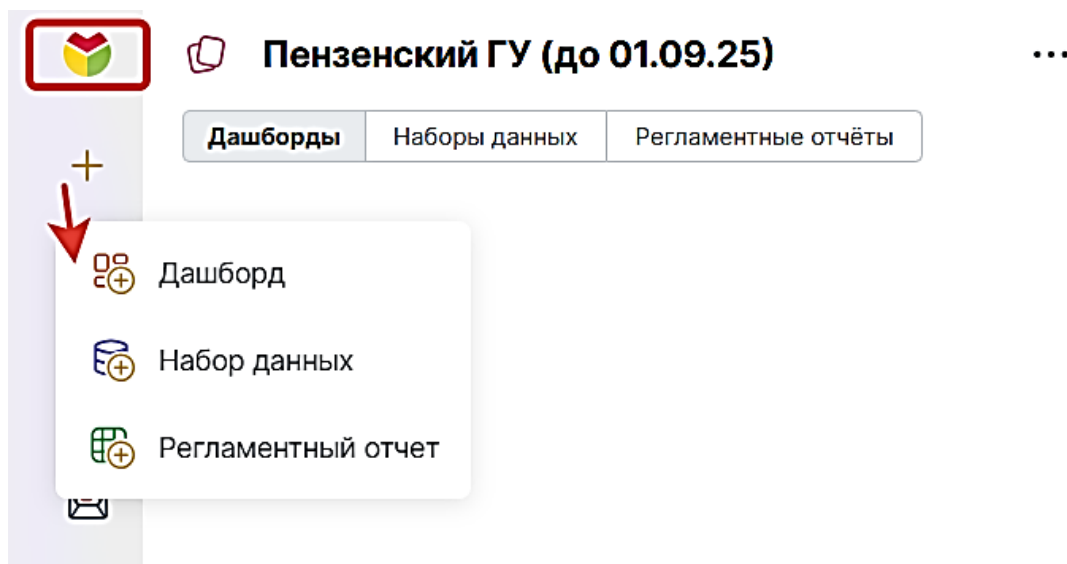


Рис. 2.36. Вкладки для доступа к объектам в рабочей области

Рассмотрим, как создаются первые два типа объектов. Начнем с создания набора данных (рис. 2.37). Для этого на сайтбаре (панель

навигации слева) нажимаем иконку «+», в выпадающем меню выбираем Набор данных и даем имя (название) набору.

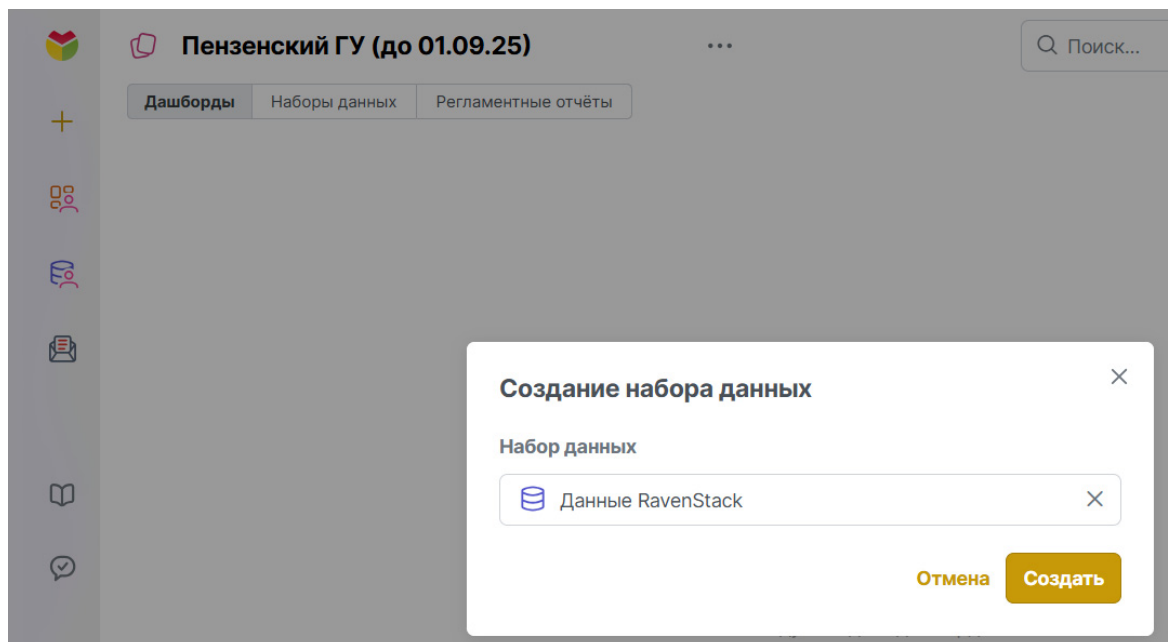


Рис. 2.37. Создание набора данных

Созданный набор данных (пока пустой) отобразится во вкладке Наборы данных в рабочей области. Для любого набора данных доступны различные операции и настройки в контекстном меню «...», а также отображается актуальная информация об изменениях и обновлениях (рис. 2.38).

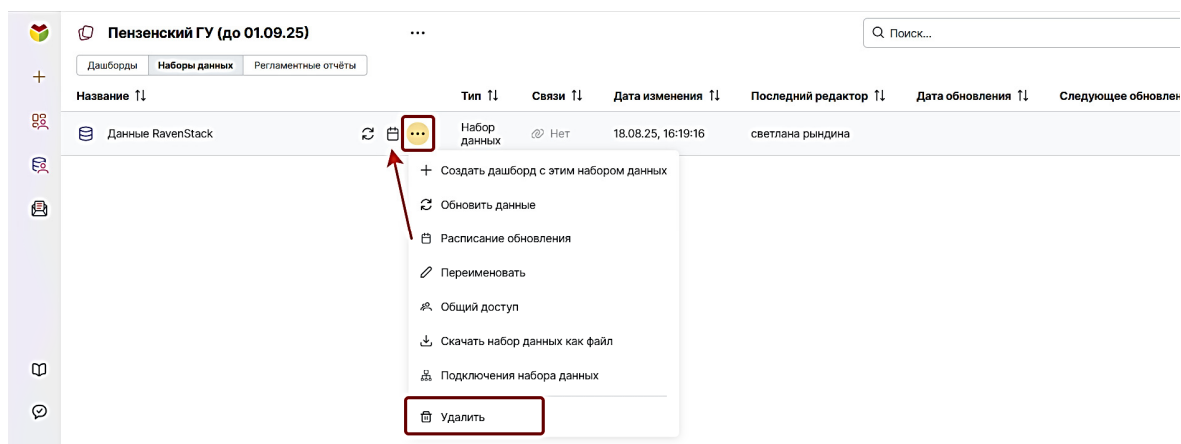


Рис. 2.38. Вкладка Наборы данных

Откроем страницу с созданным набором данных, а точнее перейдем в интерфейс настройки модели данных и загрузим на платформу данные из файла (рис. 2.39, 2.40).

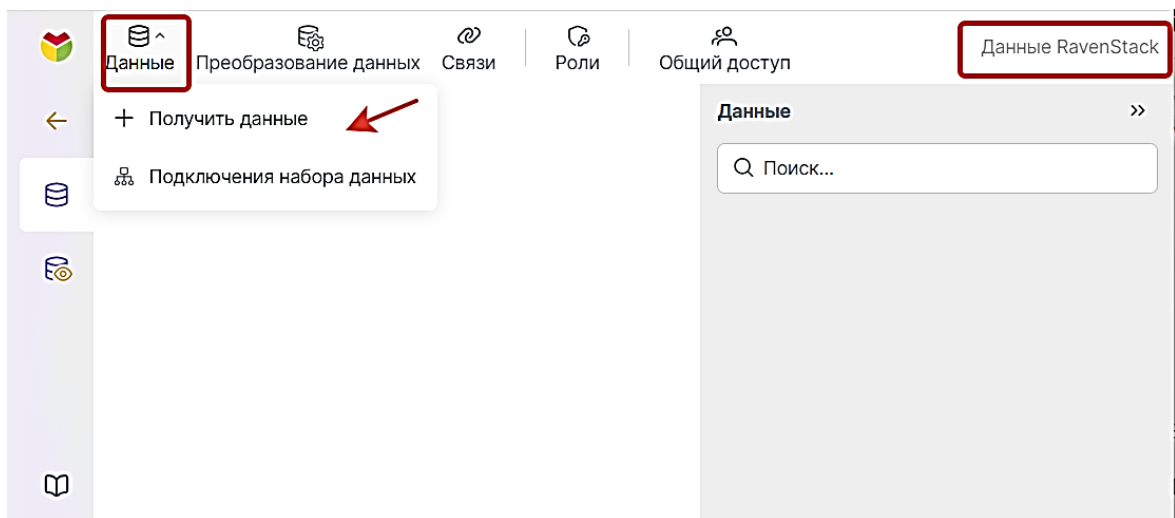


Рис. 2.39. Получение данных

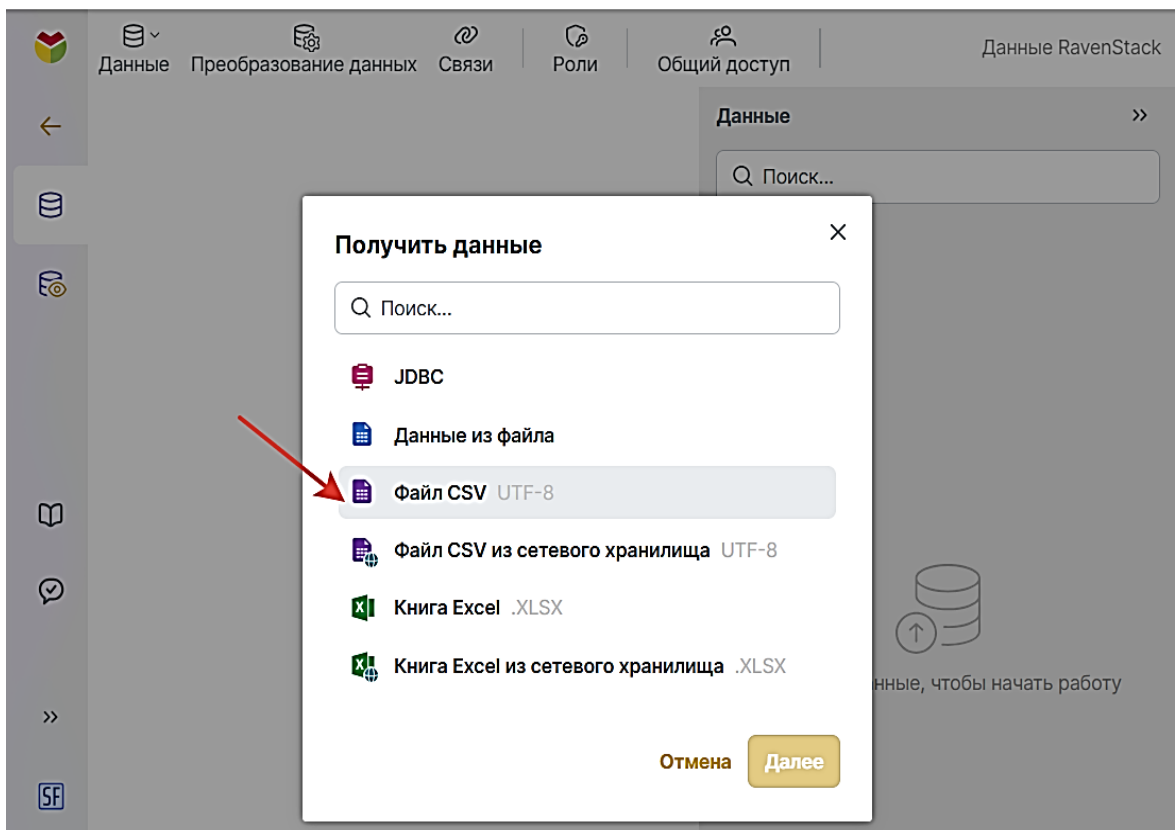


Рис. 2.40. Выбор типа источника данных (файл с расширением csv)

Загрузка происходит пофайлово. Загрузим два файла с данными (см. раздел 1.3): `ravenstack_accounts.csv` и `ravenstack_subscriptions.csv`, первым выберем `ravenstack_accounts.csv` (рис. 2.41). Слева от имени поля отображается тип данных.

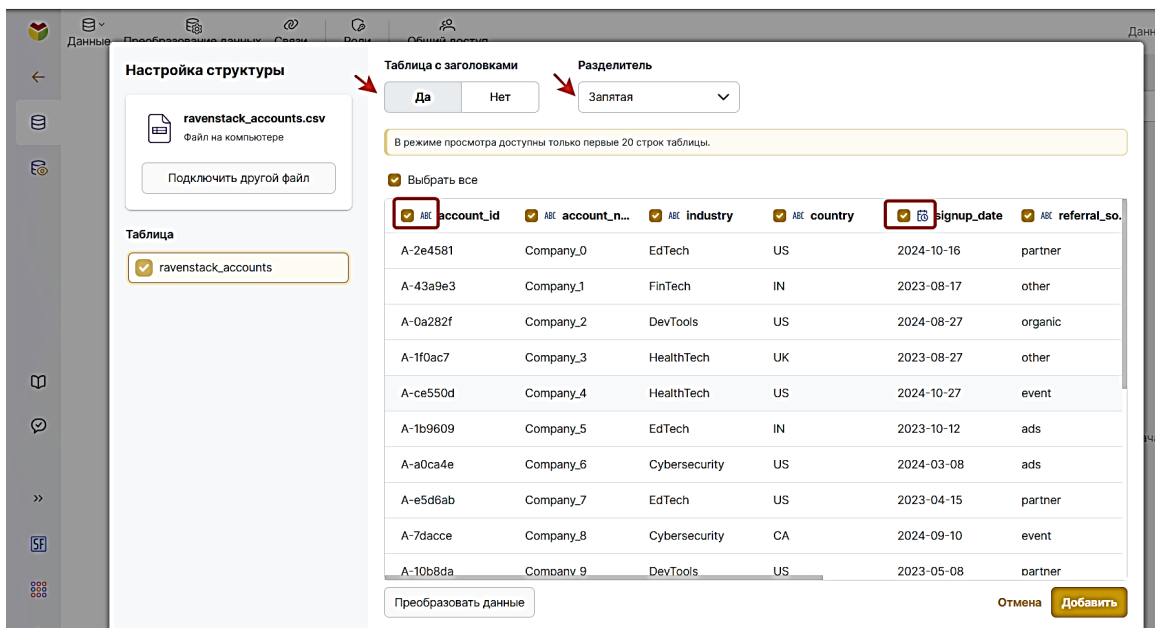


Рис. 2.41. Предварительный просмотр содержимого загружаемого файла с данными

Платформа поддерживает несколько типов данных, представленных в табл. 2.1.

Таблица 2.1

Типы данных, поддерживаемые платформой

Маркер	Тип	Описание
ABC	Текст	Строковый тип данных
123	Целое число	64-разрядное целочисленное значение
1.2	Десятичное число с плавающей запятой	64-разрядное число с плавающей запятой
\$	Десятичное число с фиксированной запятой	Десятичный тип данных (Decimal128), обладающий повышенной точностью при вычислениях, округляемый до 4-го знака после запятой и допускающий 34 цифры значения в целой части
📅	Дата и время	Данные в формате дата-время в виде ГГГГ-ММ-ДД ЧЧ:ММ:СС

В интерфейсе слева (по центру) отображается модель данных с таблицами и колонками (свернуты), справа список полей таблицы (для числовых полей маркер Σ , указывающий на возможность агрегации, для полей типа Дата и время маркер календаря) (рис. 2.42).

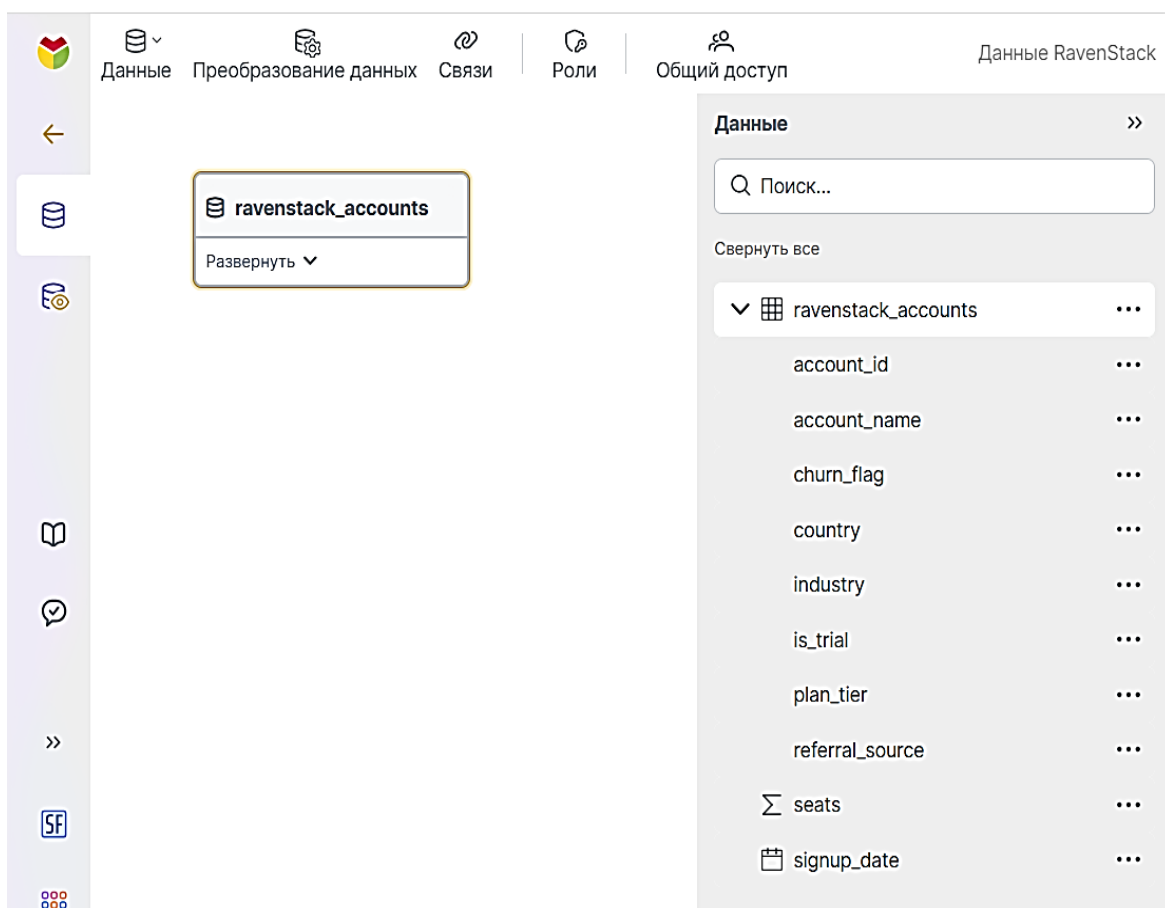


Рис. 2.42. Вид интерфейса настройки модели данных
Данные RavenStack после загрузки файла

Аналогично проведем загрузку данных из второго файла `ravenstack_subscriptions.csv` (если схема второй таблицы незаметна в рабочей области, а панель справа отображает вторую таблицу и состав ее полей, то возможно, что схемы наложились друг на друга и нужно перетащить мышкой одну из них в сторону). Установим связь между датасетами (см. табл. 1.3), перетащив поле из родительской таблицы (в которой это поле является первичным ключом) в связанную таблицу (в которой то же самое поле – внешний ключ). Контекстное меню позволяет удалить связь, если она установлена неверно (рис. 2.43).

Таким образом была создана модель данных. Остальные файлы этого набора загружать не будем.

Перейдем к созданию дашборда (рис. 2.44).

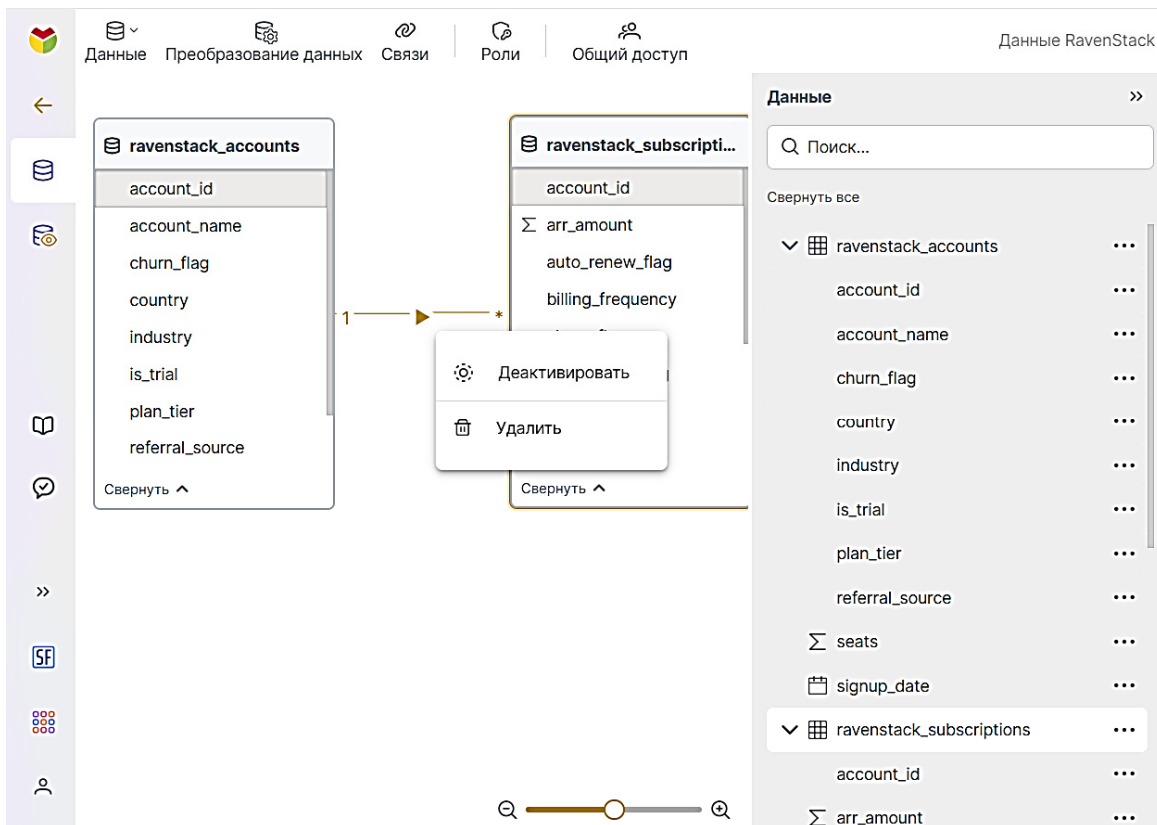


Рис. 2.43. Настройка связи между таблицами

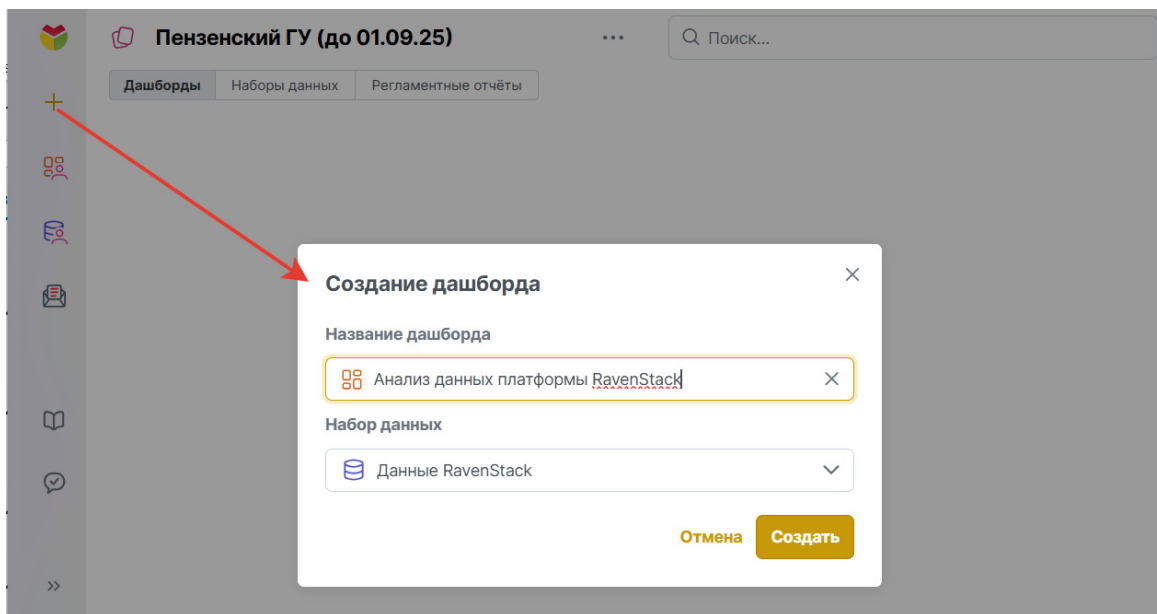


Рис. 2.44. Создание дашборда Анализ данных платформы RavenStack

Дашборд – зависимый объект, он создается на основе набора данных. Создать дашборд при отсутствии хотя бы одного доступного набора данных не получится.

В отличие от Yandex DataLens, где чарты создаются как отдельные объекты, все виджеты (чарты) в Visiology принадлежат определенному дашборду. Страница дашборда разделена на три области: рабочее пространство, в котором располагаются виджеты и происходит их настройка, внизу рабочей области есть возможность добавлять новые страницы (вкладки) дашборда, в центре располагаются панели настройки фильтров и виджетов, а справа – область доступа к данным: список всех таблиц и их полей, входящих в набор данных, на основе которого создается дашборд (рис. 2.45).

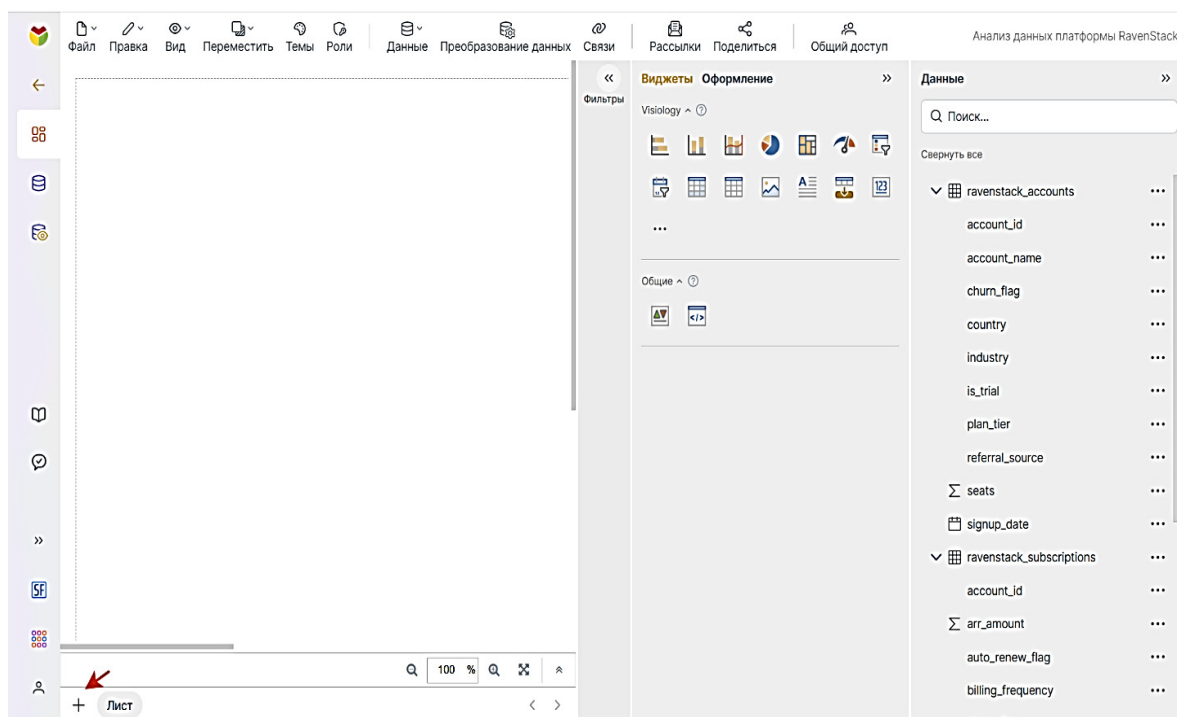


Рис. 2.45. Страница дашборда

Создадим два индикатора по количеству действующих и отозванных лицензий. Выберем виджет Индикатор, перетащим в поле значений Виджета показатель seats (рис. 2.46), для которого можно изменить в агрегирующую функцию (по умолчанию это Сумма, но можно выбрать и другие варианты).

Далее перейдем на вкладку Оформление (рис. 2.47), на ней находятся визуальные настройки виджета, такие как цвет, фон, рамка, прозрачность и т.д. Дадим название индикатору, это делается в разделе Заголовков.

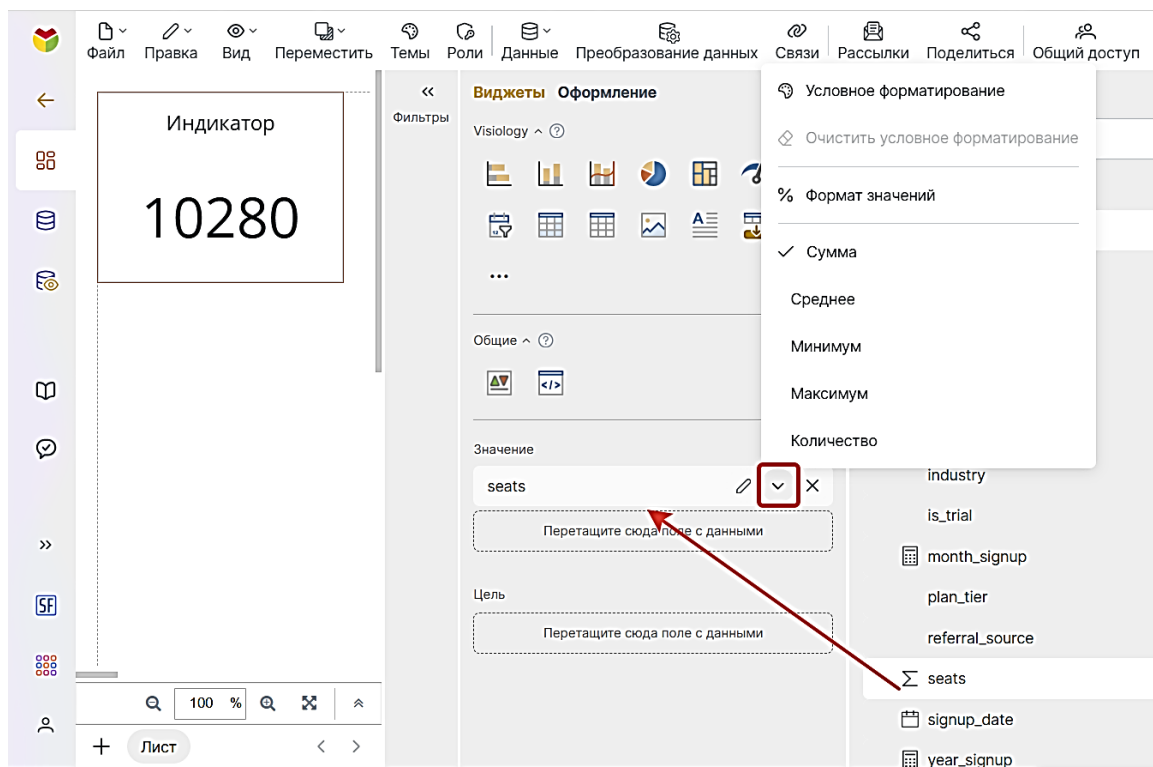


Рис. 2.46. Добавление виджета Индикатор

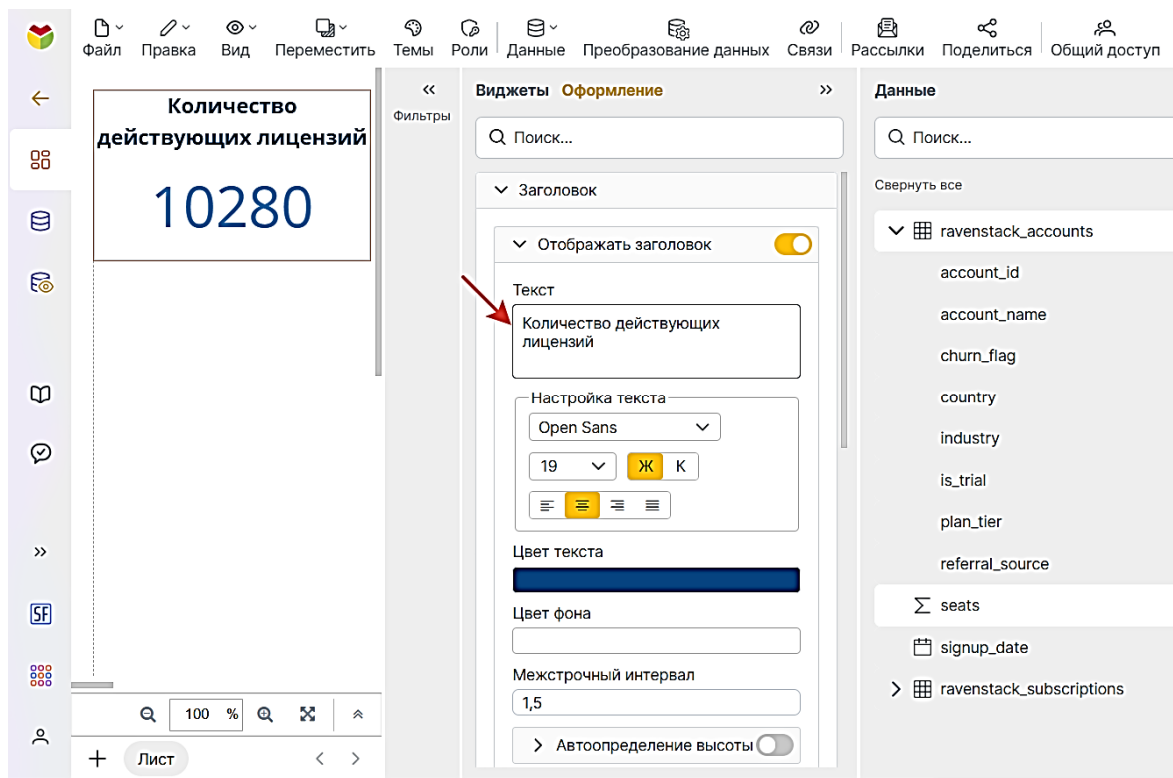


Рис. 2.47. Настройки оформления виджета

Для каждого виджета в Visiology мы также можем настроить его внутреннюю фильтрацию, т.е. установить правила фильтрации для конкретного виджета, причем бизнес-пользователь при работе с дашбордом этот фильтр никак изменить не сможет. Например, добавим к настройкам виджета фильтрацию по показателю `churn_flag`, который определяет действующий аккаунт у компании или то, что она перестала пользоваться платформой (рис. 2.48).

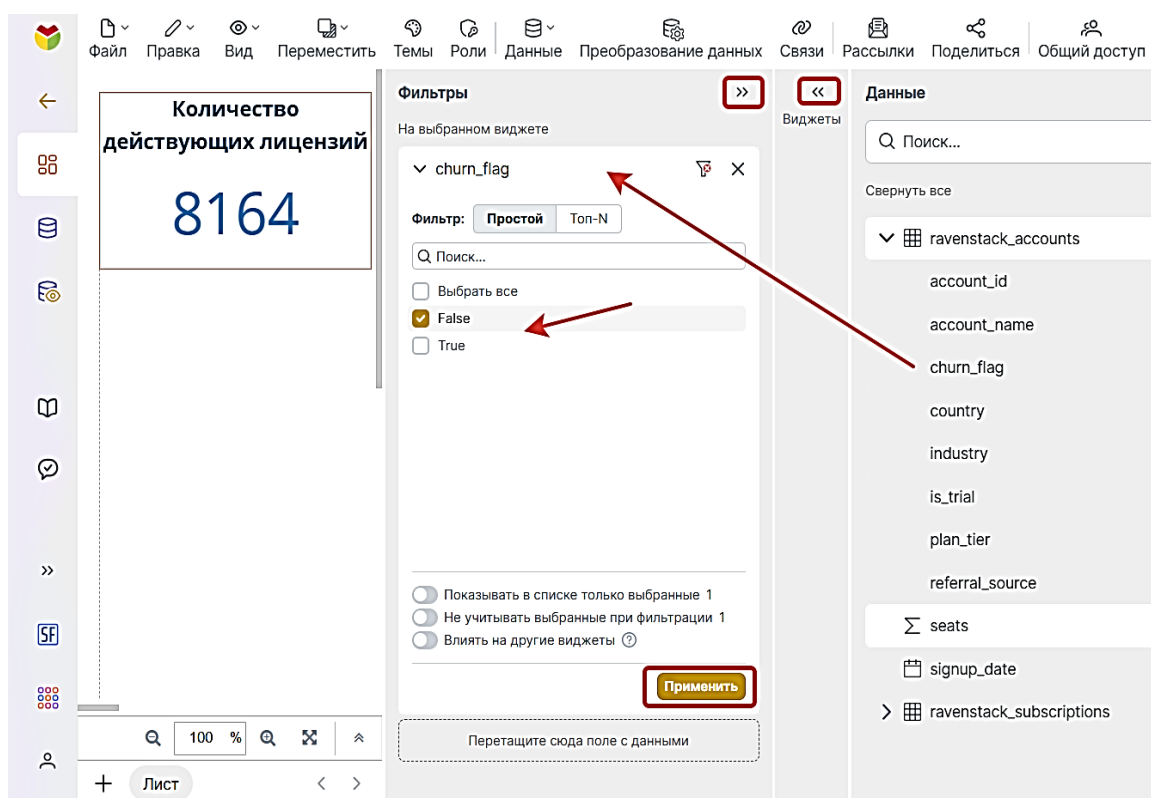


Рис. 2.48. Настройка фильтра для виджета

Аналогично добавляется второй индикатор, а можно воспользоваться копированием уже настроенного виджета и скорректировать его заголовки и настройки для фильтра (рис. 2.49).

Добавим также два однотипных динамических фильтра: фильтр по странам (рис. 2.50) и фильтр по отраслям. Значения таких фильтров можно устанавливать по выбору пользователя. Важно, чтобы в момент добавления фильтра ни один виджет не был выделен, иначе он добавится в активный виджет и будет влиять на него.

Изменим оформление виджета и выберем несколько значений из списка (AU, IN, FR) (рис. 2.51).

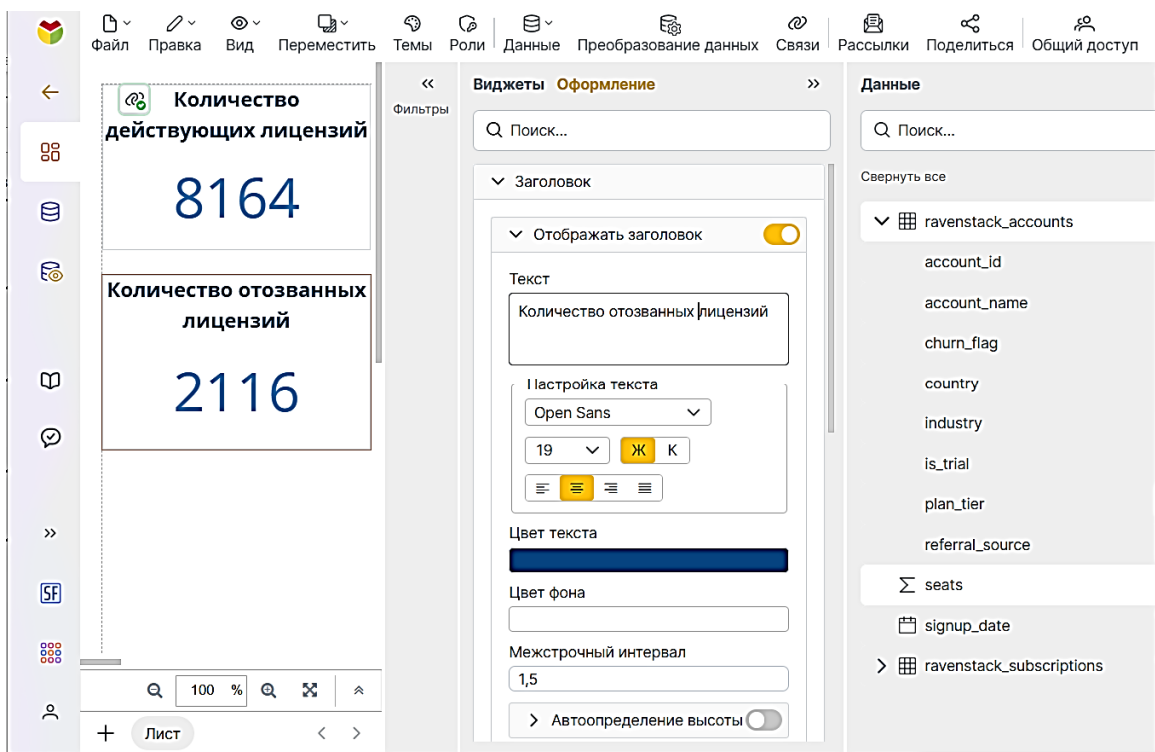


Рис. 2.49. Добавление второго виджета

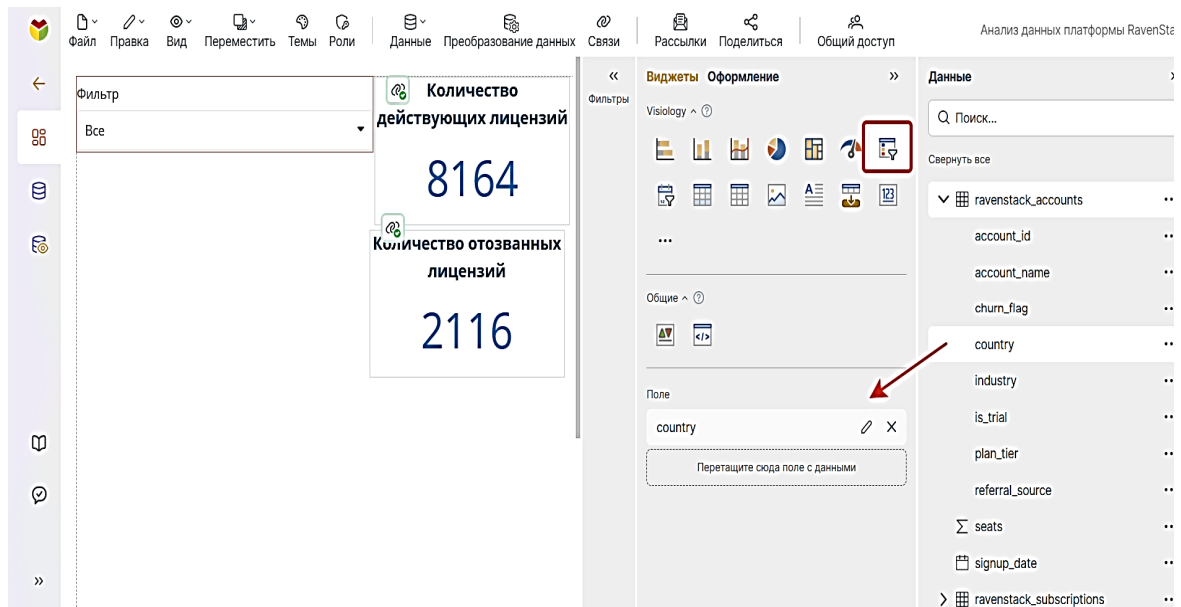


Рис. 2.50. Добавление фильтра по показателю country

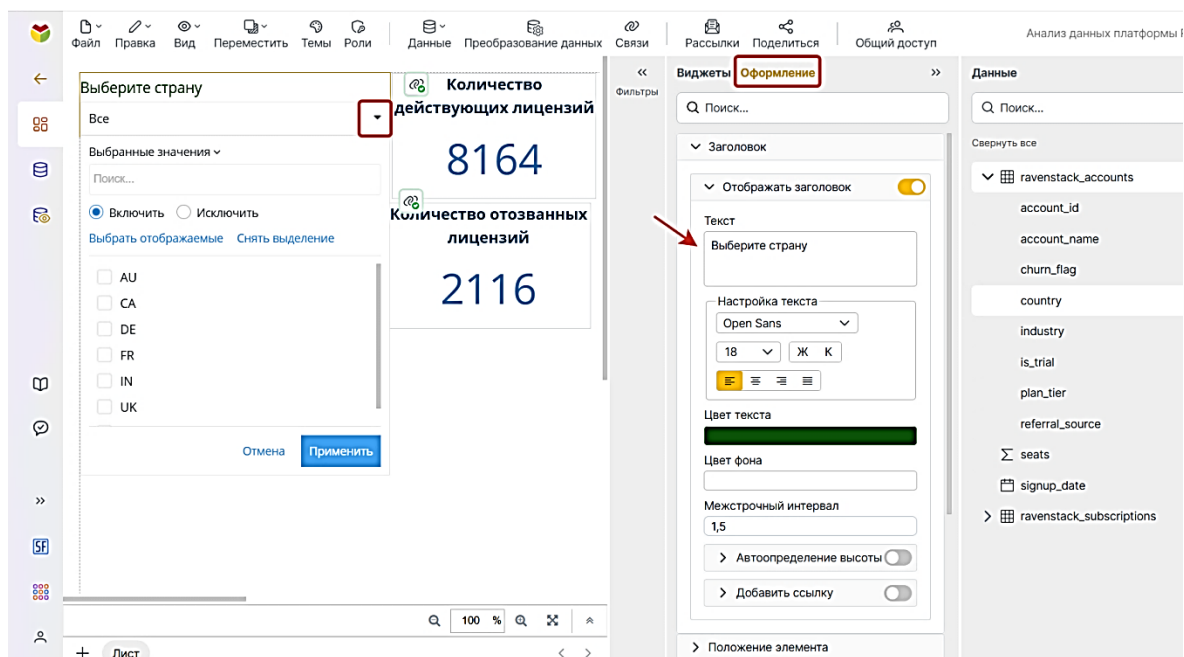


Рис. 2.51. Оформление и выбор значений для фильтра по странам

После выбора значений в фильтре по странам изменятся значения в виджетах Индикаторы. Добавим второй фильтр по отраслям аналогичным образом и еще один фильтр по датам (рис. 2.52), для которого используем показатель `signup_date`.

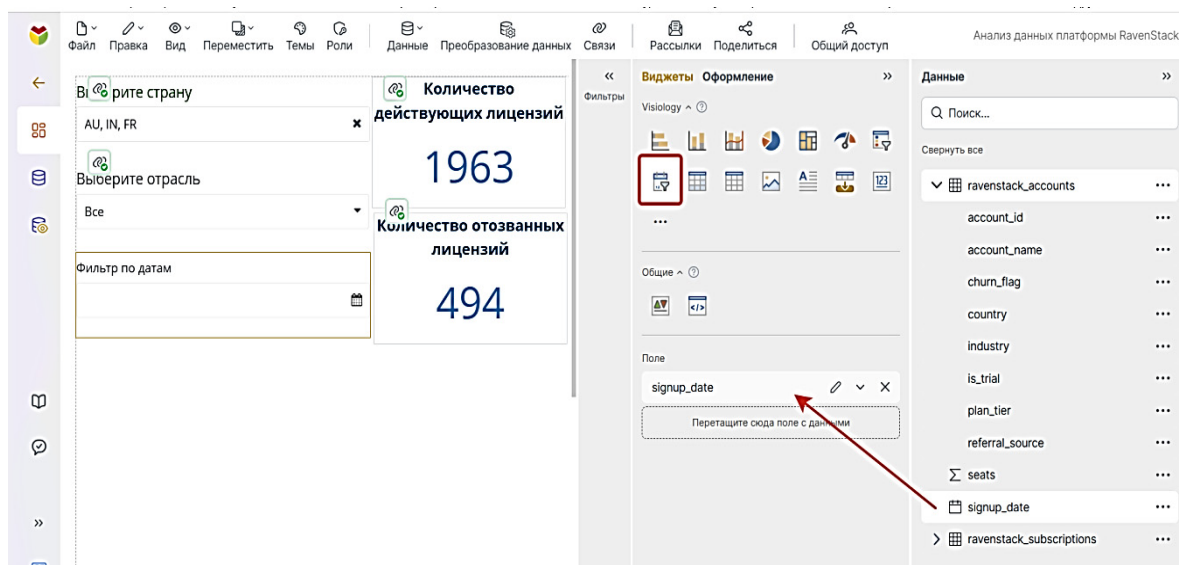


Рис. 2.52. Добавление фильтра по датам

Добавим также виджет Сводная таблица с настройками как на рис. 2.53. В качестве показателей строк выбраны `industry` и `churn_flag`, а для столбцов – показатель `country`.

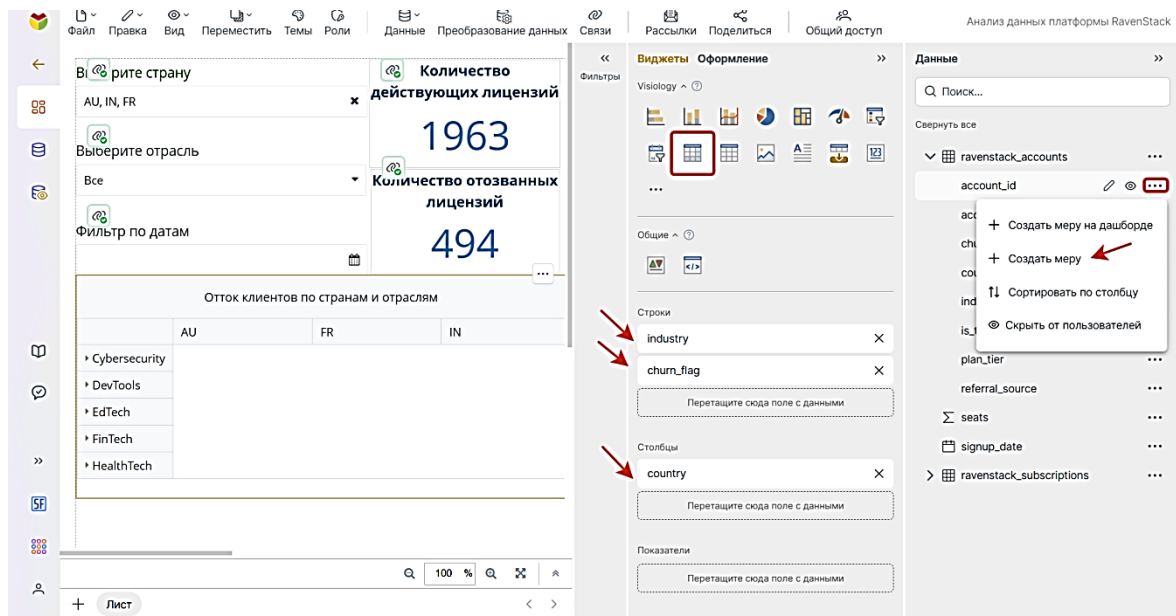


Рис. 2.53. Настройки для виджета Сводная таблица

Отображаются только три страны, те, что выбраны в фильтре по странам (фильтр применяется не только к уже созданным виджетам, но и к тем, которые добавляются на дашборд).

В качестве показателя значений в таблице планируется число компаний, которые пользуются платформой или отказались от нее. Можно использовать текстовый показатель account_id, для которого автоматически сформируется метрика Количество (или Количество уникальных) (рис. 2.54).

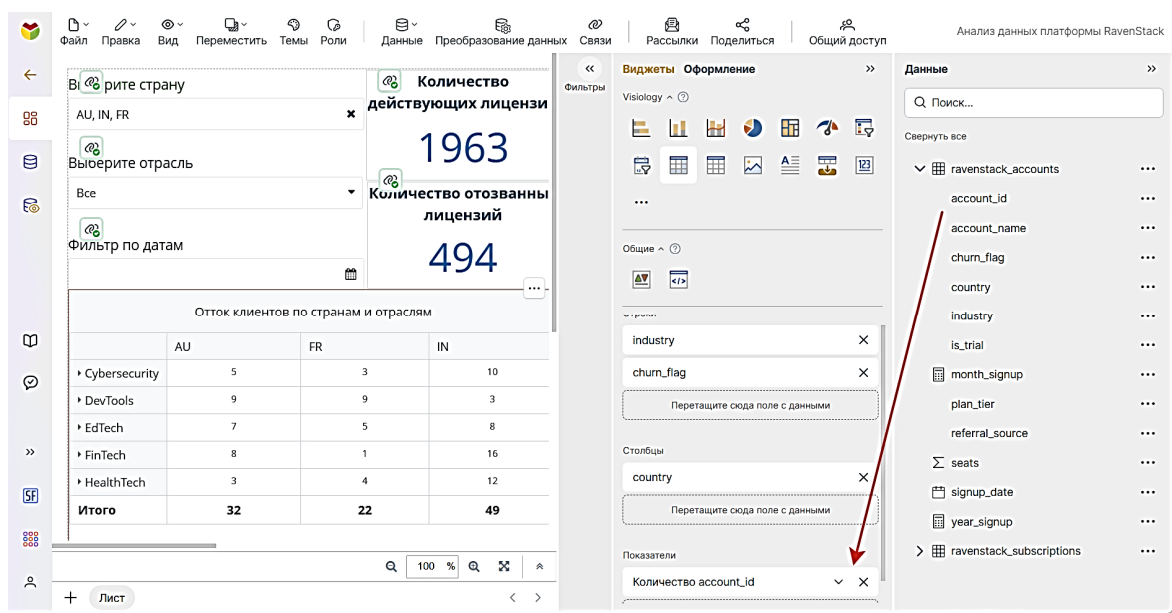


Рис. 2.54. Добавление показателя к сводной таблице

Очистим поле Показатели (Количество account_id) для сводной таблицы и покажем, как использовать DAX (Data Analysis eXpressions), чтобы создавать новую информацию из данных, уже имеющихся в модели. В дополнительном меню рядом с показателями таблицы выберем опцию Создать меру (см. рис. 2.53). И в специальном окошке (вверху слева под командным меню) введем формулу, используя подсказки (рис. 2.55). Назовем меру count_company и присвоим ей значение функции COUNT() – количество записей в показателе, которые отвечают определенным условиям.

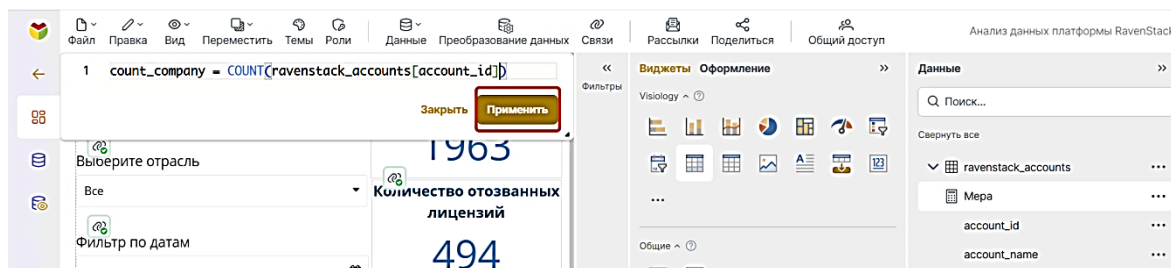


Рис. 2.55. Создание меры

После того как будет нажата кнопка Применить, значение показателя нового показателя count_company можно разместить в сводной таблице (рис. 2.56). Обратите внимание, что все фильтры на рисунке уже свободны от каких-либо значений.

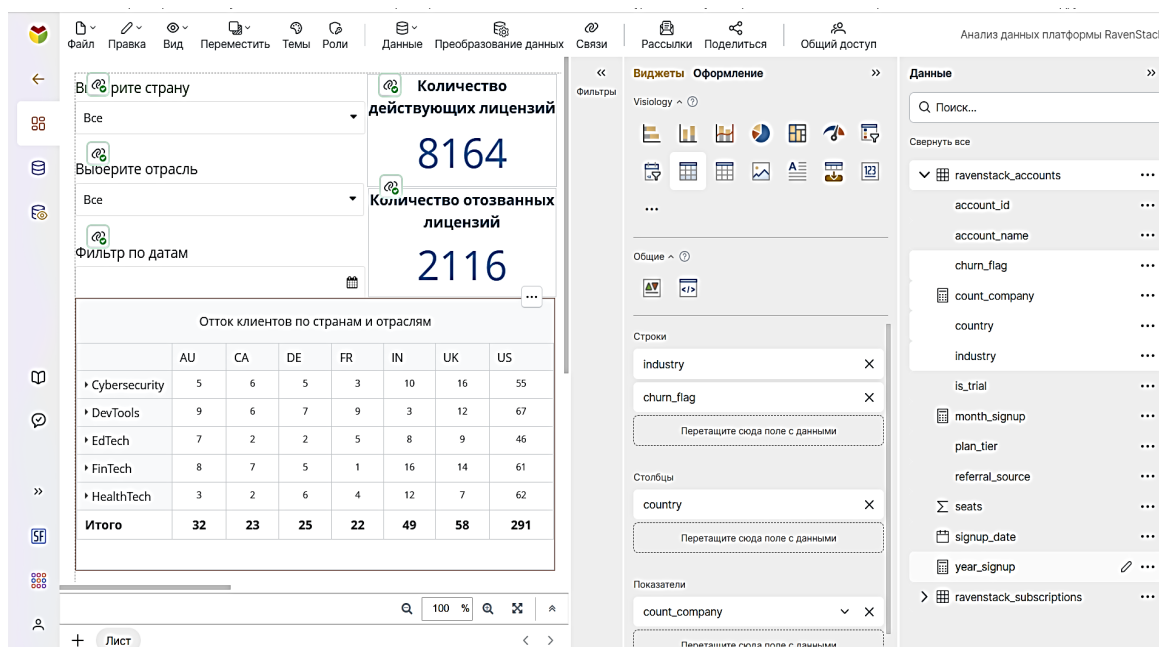


Рис. 2.56. Добавление созданной меры в качестве показателя к Сводной таблице

Выделение строк/столбцов в таблице срабатывает как фильтр (рис. 2.57).

Ваша страна: Выберите страну

Ваша отрасль: Выберите отрасль

Фильтр по датам

Количество действующих лицензий: 1614

Количество отозванных лицензий: 362

Отток клиентов по странам и отраслям

		AU	CA	DE	FR	IN	UK	US
▼ Cybersecurity	False	4	5	3	3	9	15	45
	True	1	1	2		1	1	10
Cybersecurity Всего		5	6	5	3	10	16	55
▸ DevTools		9	6	7	9	3	12	67
▸ EdTech		7	2	2	5	8	9	46
▸ FinTech		8	7	5	1	16	14	61
▸ HealthTech		3	2	6	4	12	7	62

Лист

Рис. 2.57. Фильтрация с использованием выделения строк/столбцов в Сводной таблице

Также можно настроить взаимную фильтрацию виджетов. Щелчком на соответствующем маркере можно исключить влияние выделения элементов Сводной таблицы на другие виджеты (рис. 2.58). К индикатору Количество действующих лицензий при переходе кнопки связи в неактивное состояние выделенные строки в Сводной таблице перестают применяться.

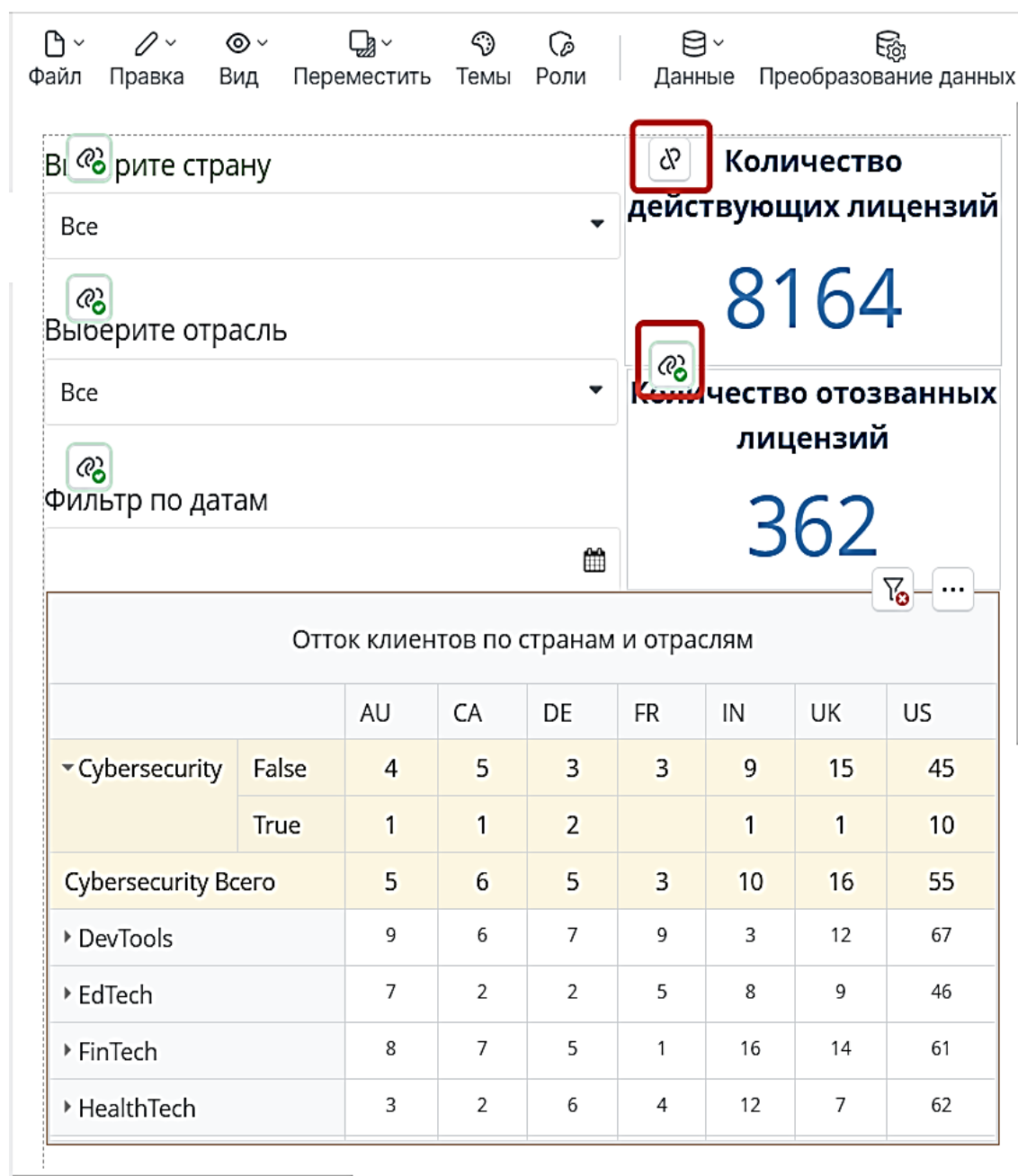


Рис. 2.58. Разница в применении фильтров на основе Сводной таблицы при активном/неактивном влиянии

Добавим фильтр по датам и вернем кнопку влияния для индикатора Количество действующих лицензий в активное состояние (рис. 2.59). А затем очистим фильтр по датам и снимем все выделения.

Сохраним дашборд и при необходимости экспортируем его в виде рисунка, документа или презентации (рис. 2.60).

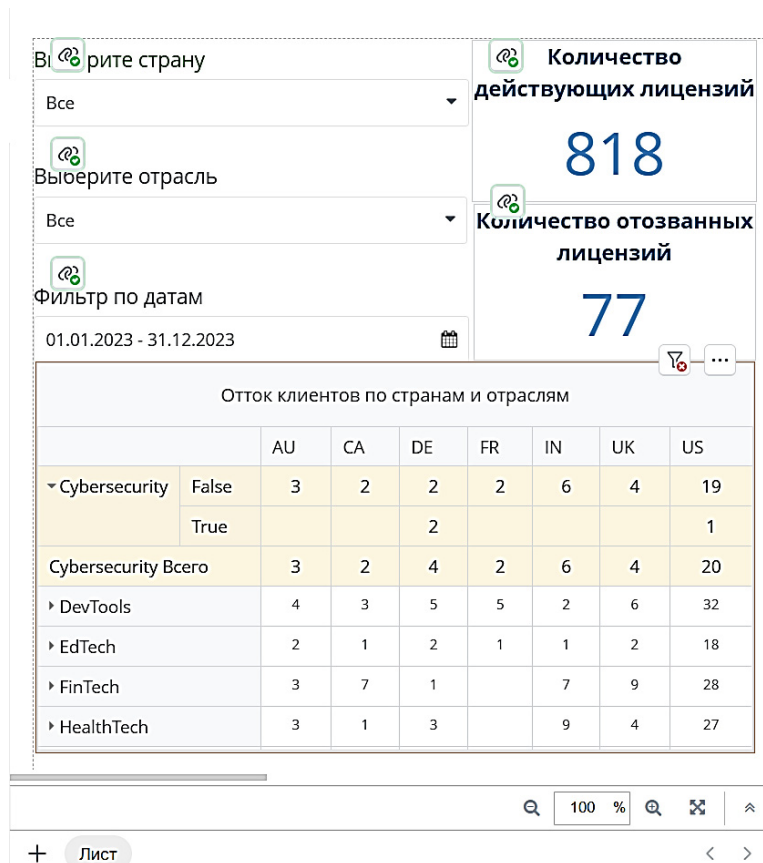


Рис. 2.59. Включение/отключение влияния фильтрации с виджета для индикатора Количество действующих лицензий

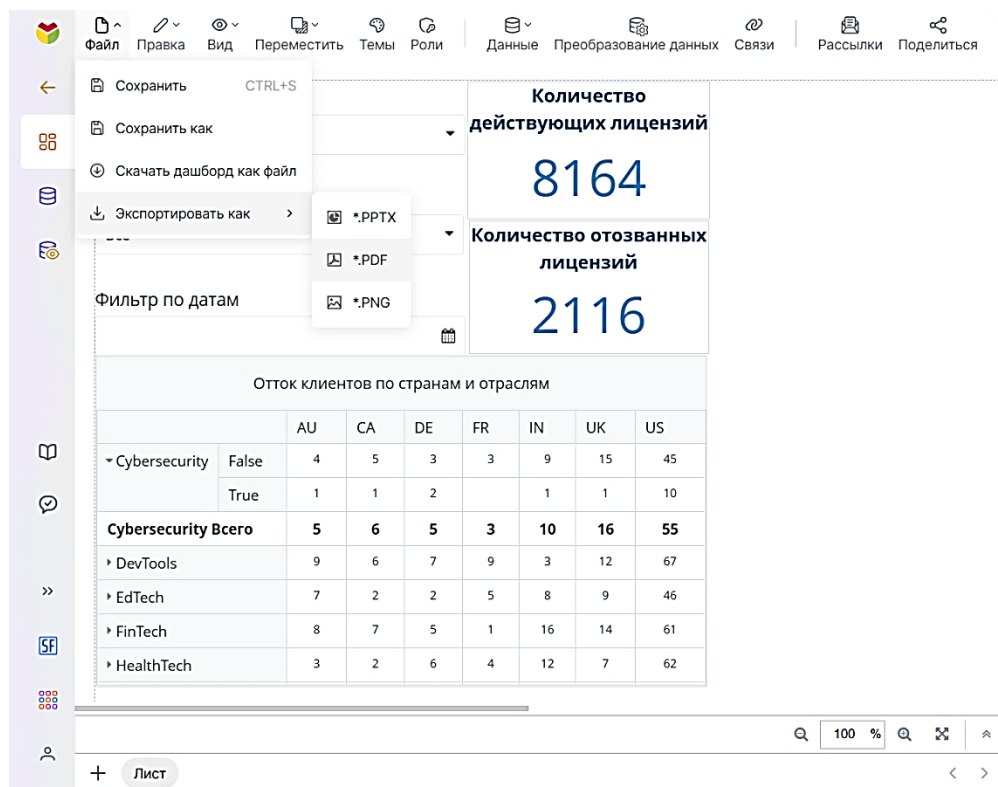


Рис. 2.60. Сохранение дашборда

Для дальнейшего совершенствования навыков работы на платформе Visiology можно воспользоваться справочной документацией [4, 5].

Самостоятельное задание:

1. Добавьте еще один лист и создайте на нем виджеты по второй таблице данных (файл ravenstack_subscriptions.csv).
2. Экспортируйте обновленный дашборд в презентацию.

2.3. Аналитика с Glarus BI

Glarus BI – это российская BI-платформа уровня Enterprise для бизнес-аналитики [6]. 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 состоит из нескольких сущностей, которые обеспечивают полный цикл работы с данными (рис. 2.61):

- запроса (SQL-запрос);
- дашборда (dashboard, приборная панель) – набор визуализаций по запросам, фильтров, текстовых блоков, ссылок и других элементов;
- коллекции;
- модели;
- действия;
- метрики.

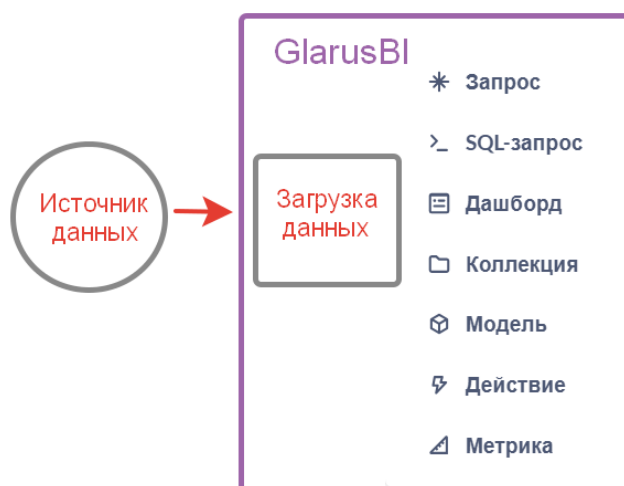


Рис. 2.61. Сущности Glarus BI

Коллекция – папка, в которой хранятся запросы, дашборды, модели. Бесплатный вариант работы в демоверсии возможен

в облаке <https://demo.glarus-analytics.ru>, доступ к которому можно получить автоматически, заполнив на сайте <https://glarus-bi.online> регистрационную форму (рис. 2.62) .

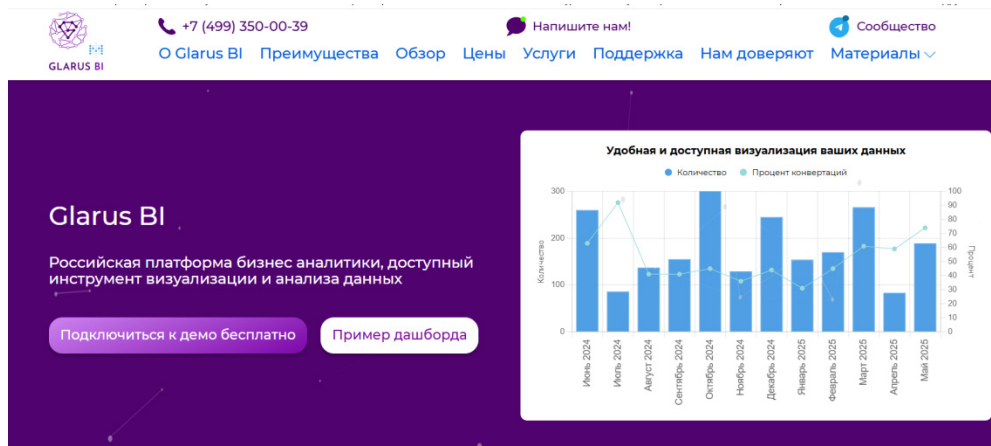


Рис. 2.62. Облачная демоверсия для одного пользователя

Будем использовать набор данных о платформе RavenStack (см. раздел 1.3). Переведем файл с данными `ravenstack_accounts.csv` в формат `.xlsx`. Для этого используем опцию Из текстового файла на вкладке главного меню Данные (рис. 2.63).

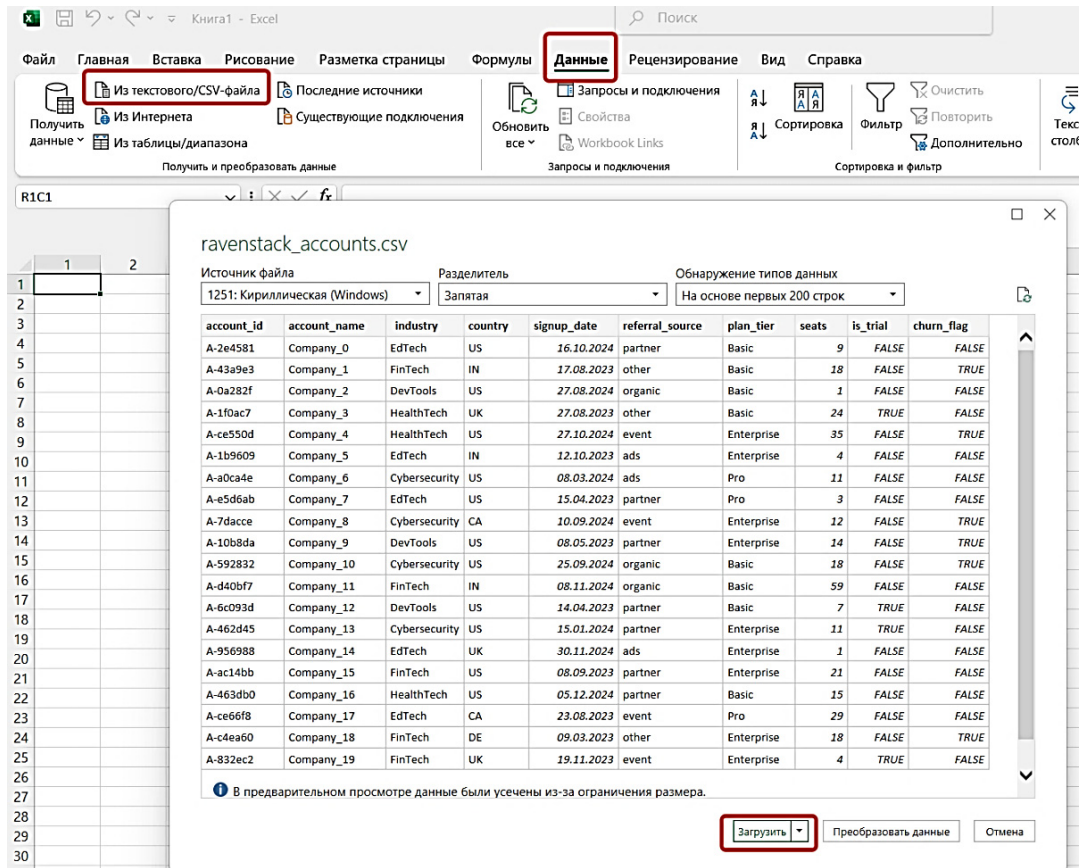


Рис. 2.63. Импорт данных из текстового файла в таблицу Excel

Удалим при необходимости все пустые листы книги и сохраним файл с именем `ravenstack_accounts.xlsx`.

В облачном сервисе Glarus BI в демоверсии доступна загрузка только файлов с расширением `.xlsx` (рис. 2.64).

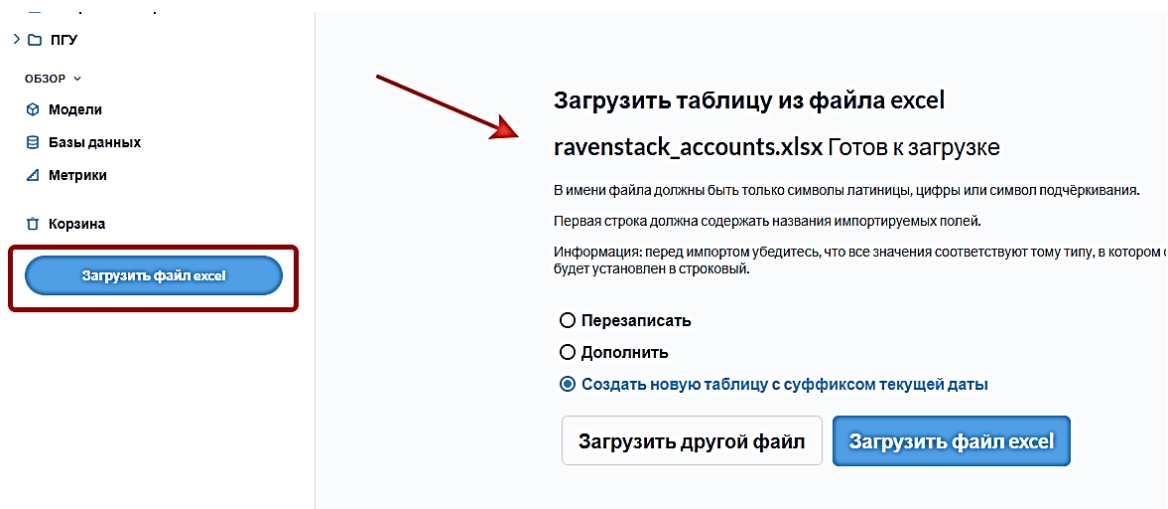


Рис. 2.64. Загрузка файла с данными

После загрузки данные будут доступны для просмотра в Базе данных (рис. 2.65).

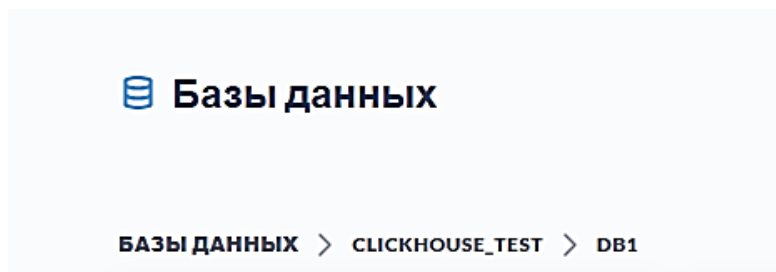


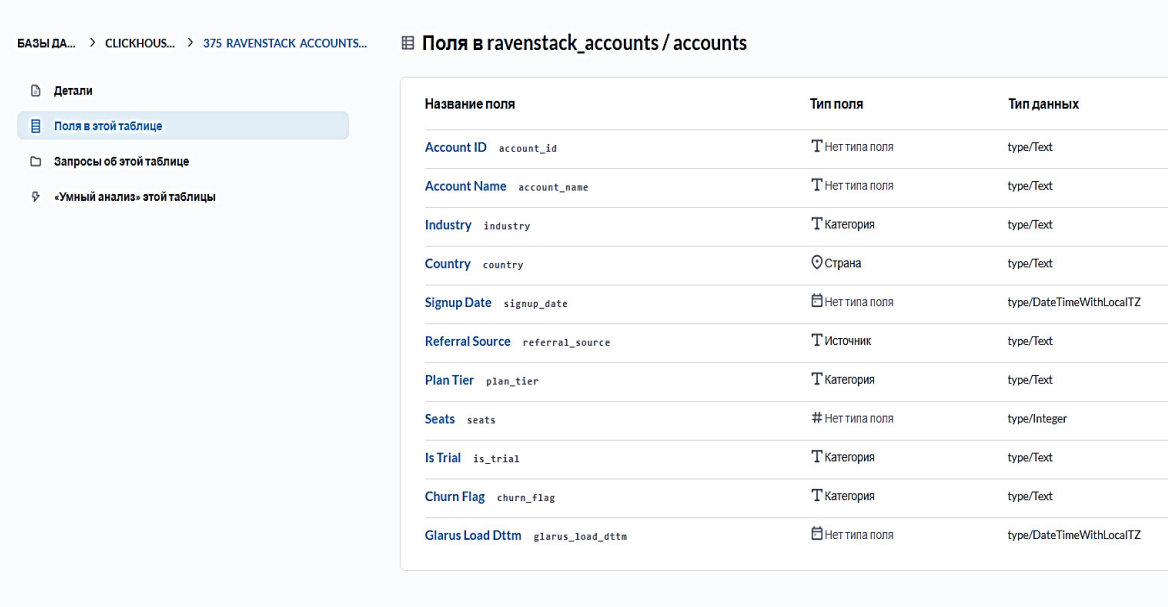
Рис. 2.65. Иерархия для загруженных таблиц в базу данных

Загруженные данные можно удалить, получить информацию о наборе данных (рис. 2.66), а также просмотреть предлагаемые визуализации (значки справа от выделенного набора позволяют это сделать).



Рис. 2.66. Загруженный набор данных в Базе данных

Перейдем к информации о загруженной таблице данных, щелкнув по значку  рядом с именем таблицы. На вкладке Поля в этой таблице (рис. 2.67) увидим значки, схожие с маркерами типов данных в Yandex DataLens.



Название поля	Тип поля	Тип данных
Account ID account_id	T Нет типа поля	type/Text
Account Name account_name	T Нет типа поля	type/Text
Industry industry	T Категория	type/Text
Country country	📍 Страна	type/Text
Signup Date signup_date	📅 Нет типа поля	type/DateTimeWithLocalTZ
Referral Source referral_source	T Источник	type/Text
Plan Tier plan_tier	T Категория	type/Text
Seats seats	# Нет типа поля	type/Integer
Is Trial is_trial	T Категория	type/Text
Churn Flag churn_flag	T Категория	type/Text
Glarus Load Dttm glarus_load_dttm	📅 Нет типа поля	type/DateTimeWithLocalTZ

Рис. 2.67. Поля в ravenstack_accounts

Несмотря на то что при загрузке произошло изменение названий полей, далее в тексте будем использовать исходные варианты обозначений для полей – account_id, хотя он преобразовался в вариант Account ID.

Для поля country определена метка геолокации (расположение), так как внутри аналитической системы есть геопозиционирование для стран по координатам, хранящимся в справочнике, благодаря чему есть возможность строить визуализации с привязкой к карте мира при использовании этого показателя. Текстовые поля, значения которых уникальны для каждого наблюдения (account_id, account_name), не определяются как категории, а вот текстовые поля, для которых количество уникальных значений существенно меньше числа наблюдений (industry), маркируются как категории.

«Умный анализ» таблицы дает представление о том, какие визуализации можно построить на основе имеющихся данных (рис. 2.68–2.70). Начало дашборда – фильтры по категориальным показателям (рис. 2.68).

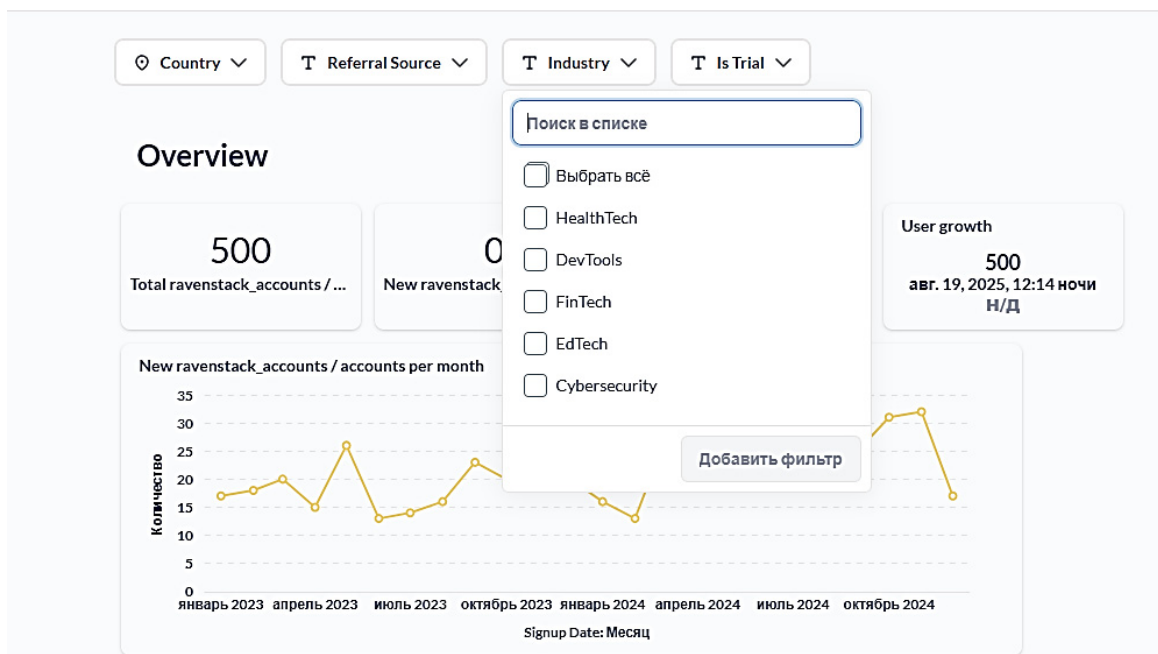


Рис. 2.68. Строка фильтров по категориальным показателям.

Далее располагаются индикаторы (рис. 2.69), а также график распределения подключений к платформе новых аккаунтов (клиентов) на временной шкале (с группировкой по месяцам).

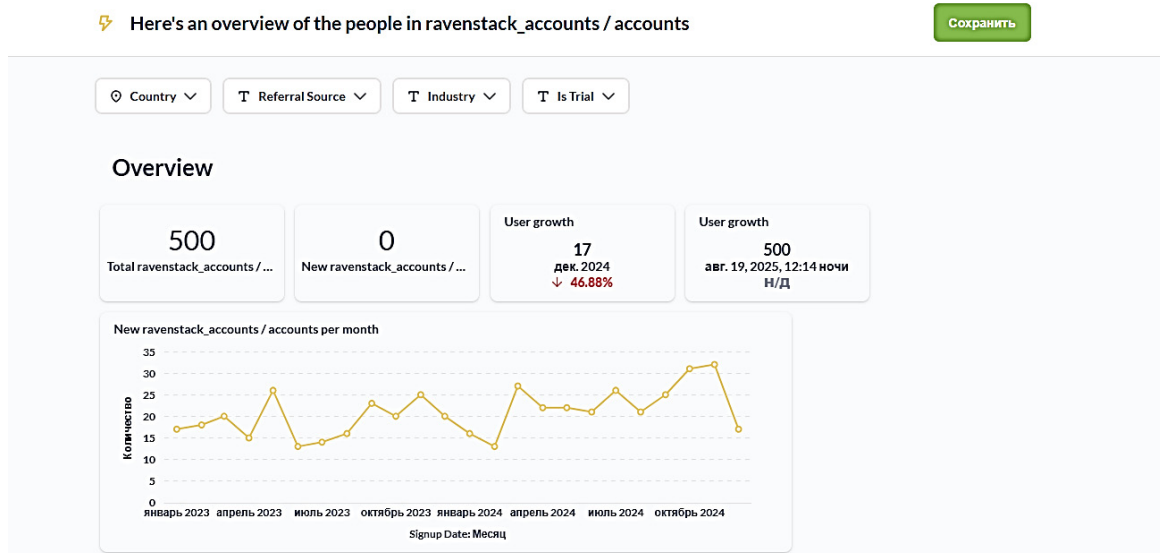


Рис. 2.69. Дашборд, построенный с помощью «умного анализа». Первая часть

Ниже на дашборде располагаются столбчатые диаграммы, гистограмма и карта (рис. 2.70). На карте выделены территории стран, в которых находится юрисдикция компаний-клиентов

платформы, чем интенсивнее цвет, тем больше клиентов из этой страны. Также интересна гистограмма, которая отображает распределение компаний по числу активных лицензий: до 25 лицензий активировало большинство компаний (352), а с большим числом активных лицензий у платформы мало компаний (в интервале от 125 до 150 лицензий и в интервале от 150 до 175 лицензий присутствует по одной компании).

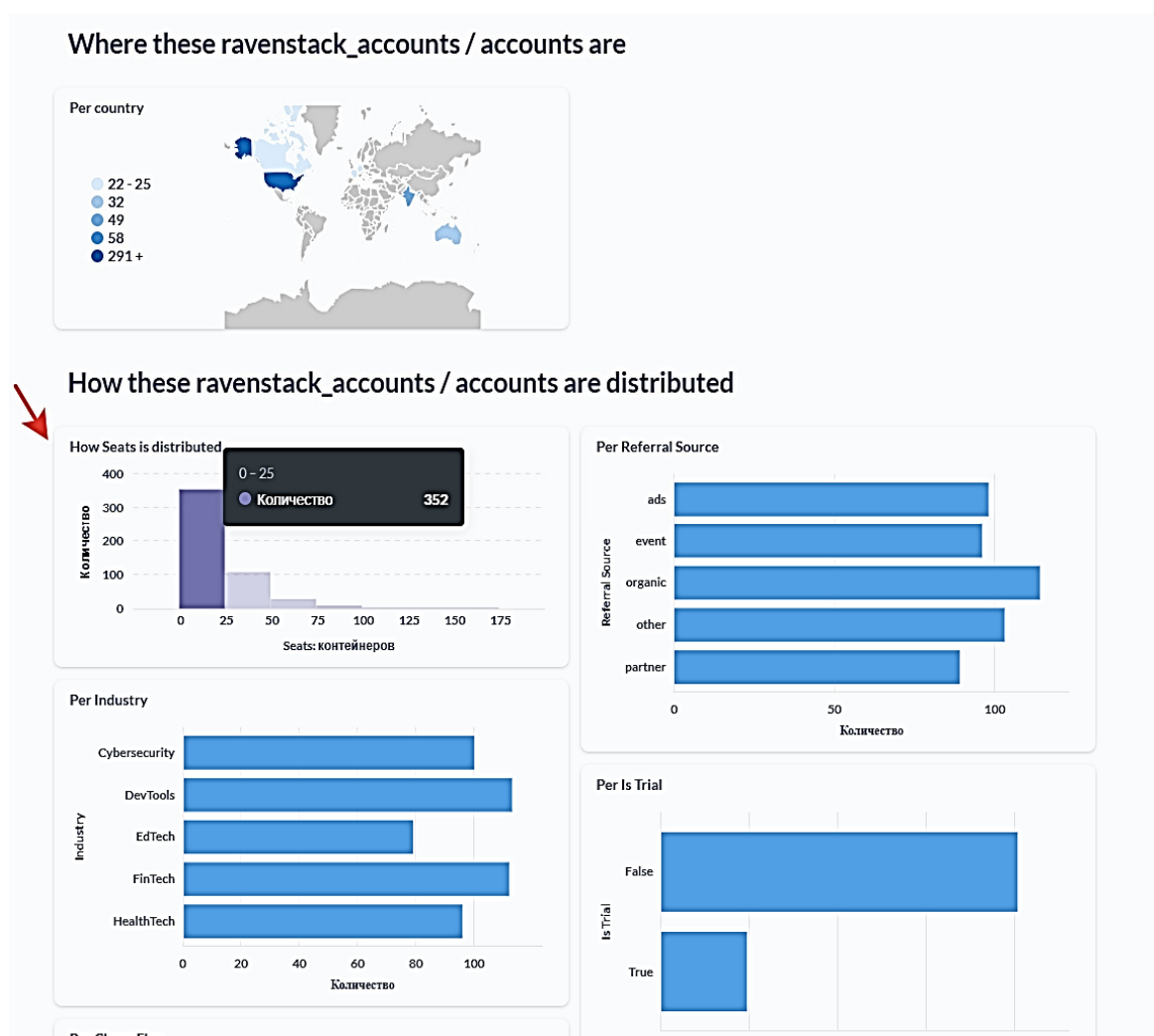


Рис. 2.70. Дашборд, построенный с помощью «умного анализа». Визуализации по показателям таблицы

«Умный анализ» — одна из сильных опций в аналитике с помощью платформы Glarus BI.

Сохраним результаты «умного анализа» и создадим в своей личной коллекции новую коллекцию с названием RavenStack (рис. 2.71). Здесь будут располагаться все создаваемые объекты по загруженному набору данных (запросы и дашборд).

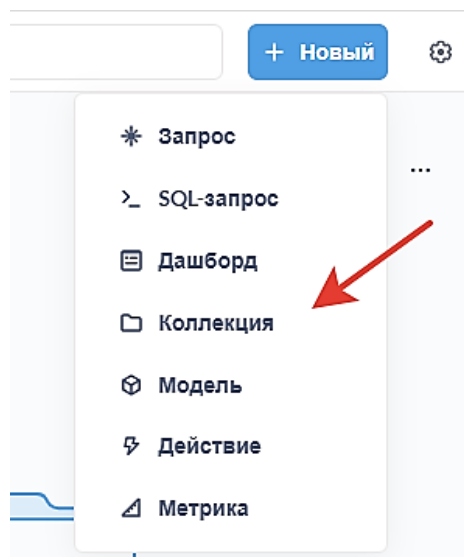


Рис. 2.71. Создание коллекции

Коллекция играет роль папки, в которой хранятся ссылки на объекты и/или другие коллекции.

Для результатов «умного анализа» была создана специальная коллекция, в которой присутствуют все визуализации дашборда в виде отдельных запросов (рис. 2.72).

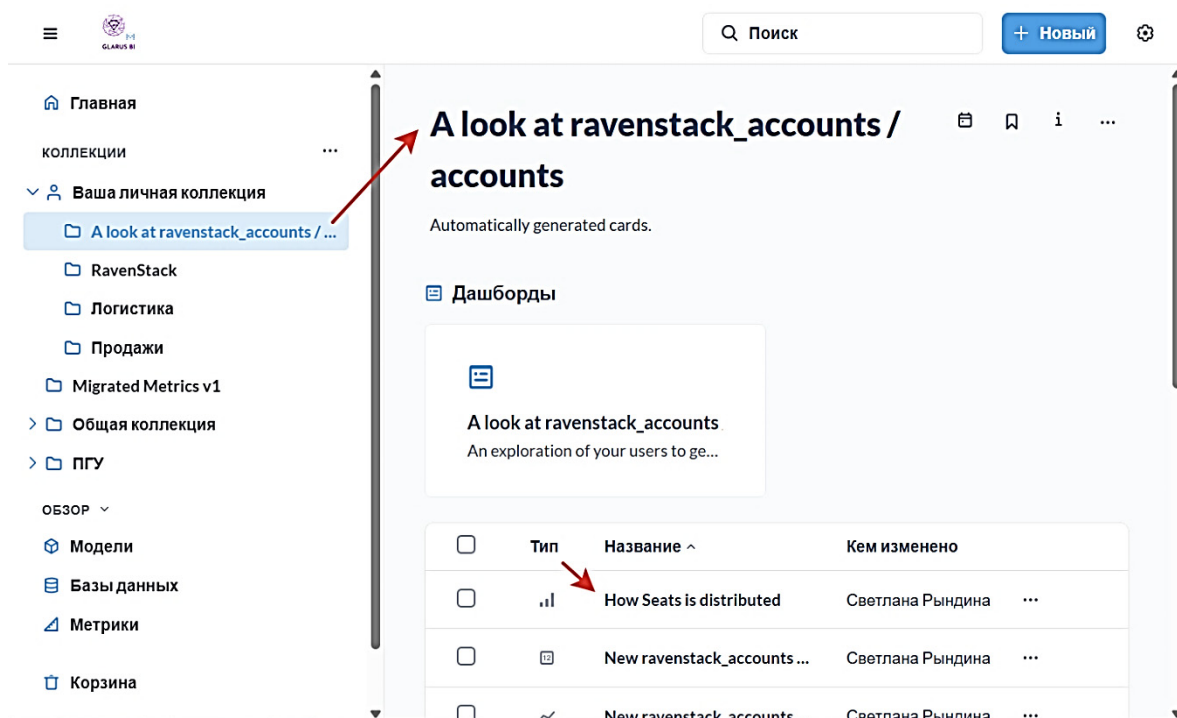


Рис. 2.72. Сохраненные результаты «умного анализа»

Эти запросы можно не только просматривать, но и редактировать (рис. 2.73).

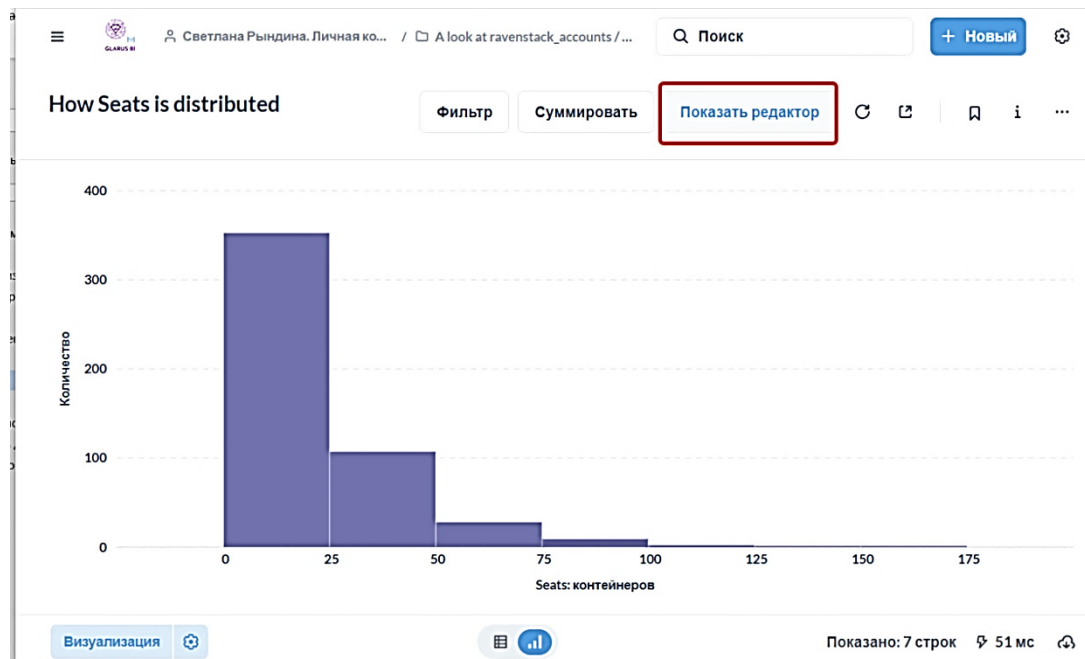


Рис. 2.73. Просмотр одного из сохраненных запросов

Однако вернемся в созданную коллекцию и выберем опцию добавления нового запроса.

В Glarus BI, как и в DataLens, построение визуализаций (запросов) не привязано к дашбордам. Запросы – гибкий инструмент построения датасетов по исходным таблицам и другим запросам, а результат запроса может быть представлен в виде базовой визуализации – таблицы или какого-либо другого подходящего виджета.

Для нового запроса необходимо выбрать данные (рис. 2.74).

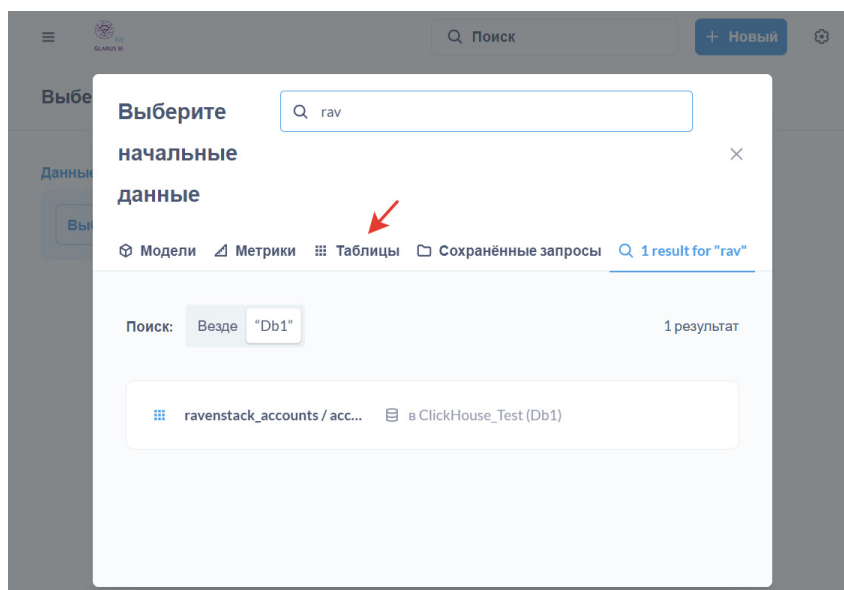


Рис. 2.74. Выбор данных для запроса

Можно перемещаться по иерархии сохраненных запросов в коллекциях или найти таблицу с исходными данными в базе данных. Однако при большом количестве объектов, которые подходят на роль данных в запросе, проще воспользоваться поисковой строкой. В качестве данных выберем только что загруженный набор `ravenstack_accounts`. Для набора данных можно изменить состав столбцов в рассмотрении (рис. 2.75) и вывести начало таблицы на просмотр (рис. 2.76). Так как поле `Glarus Load Dttm` – служебное, исключим его из рассмотрения.

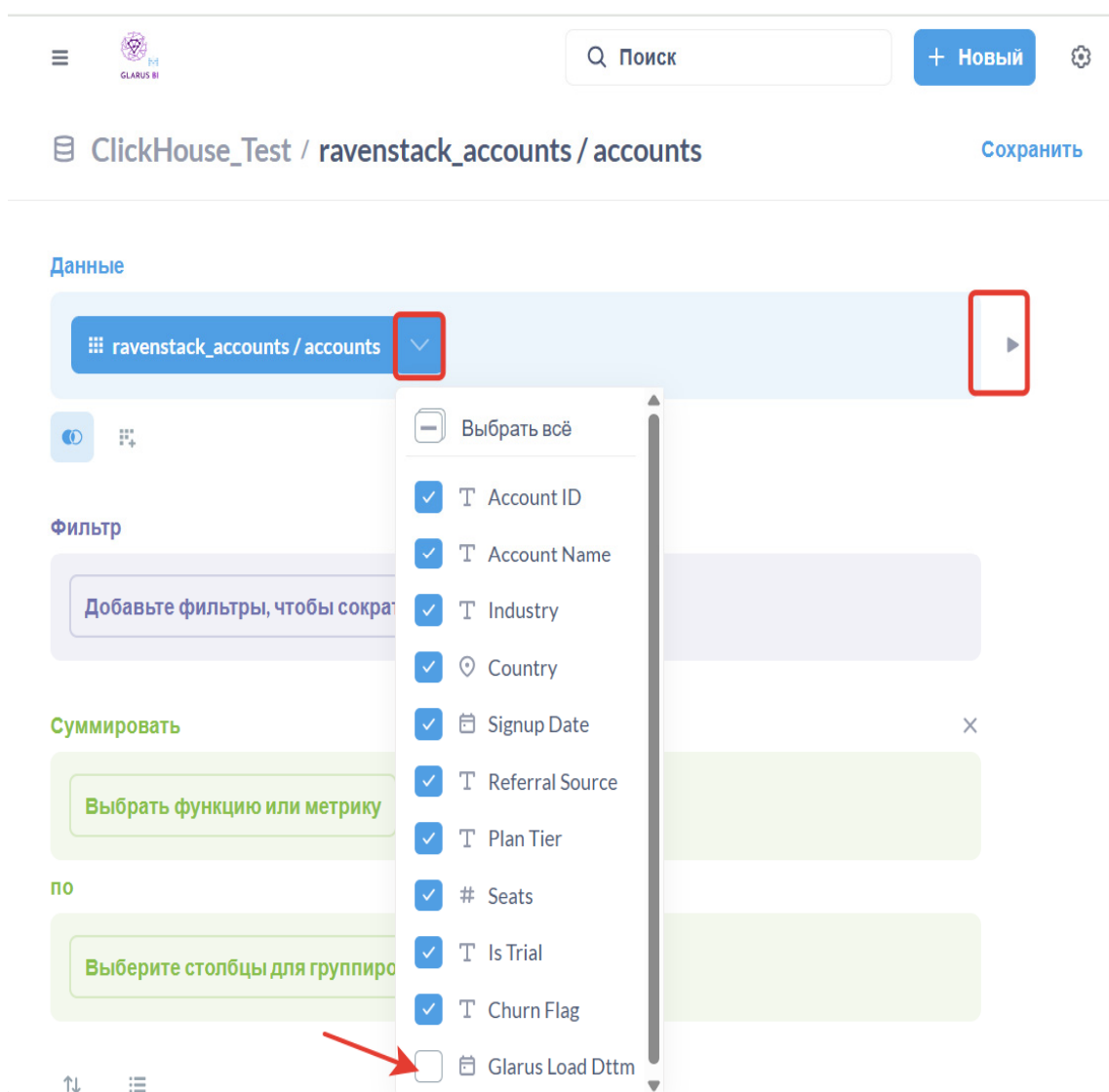


Рис. 2.75. Исключение из запроса показателей

Данные

ravenstack_accounts / accounts

Предварительный просмотр

✕

Account ID	Account Name	Industry	Country	Signup Date	Referral Source	Plan Tier	Size
A-2e4581	Company_0	EdTech	US	16 октября, 2024, 03:00	partner	Basic	9
A-43a9e3	Company_1	FinTech	IN	17 августа, 2023, 03:00	other	Basic	18
A-0a282f	Company_2	DevTools	US	27 августа, 2024, 03:00	organic	Basic	1
A-1f0ac7	Company_3	HealthTech	UK	27 августа, 2023, 03:00	other	Basic	24
A-ce550d	Company_4	HealthTech	US	27 октября, 2024, 03:00	event	Enterprise	35
A-1b9609	Company_5	EdTech	IN	12 октября, 2023, 03:00	ads	Enterprise	4
A-a0ca4e	Company_6	Cybersecurity	US	8 марта, 2024, 03:00	ads	Pro	11
A-e5d6ab	Company_7	EdTech	US	15 апреля, 2023, 03:00	partner	Pro	3
A-7adace	Company_8	Cybersecurity	CA	10 сентября, 2024, 03:00	event	Enterprise	12
A-10b8da	Company_9	DevTools	US	8 мая, 2023, 03:00	partner	Enterprise	14

Присоединить данные



Рис. 2.76. Вывод данных на просмотр

Для присоединения новых данных выбирается объект с данными (таблица, запрос и т.п.), а соединение осуществляется через поля одного типа данных с данными, принадлежащими одному множеству (идентификаторы аккаунтов и т.п.).

Создадим запрос для индикатора. Добавим фильтр по показателю `churn_flag` (рис. 2.77).

Данные

ravenstack_accounts / accounts



Фильтр

Добавьте фильтры, чтобы сократить результаты запроса

< Churn Flag

Является ▾

Поиск в списке

Выбрать всё

☒ False☐ True

Добавить фильтр

Рис. 2.77. Добавление фильтра к запросу

А в качестве показателя для агрегации выберем seats. Этот показатель числовой, и для него можно рассчитать несколько агрегирующих функций, выберем сумму, среднее и медиану (рис. 2.78). Для каждого шага настройки запроса можно вывести на предварительный просмотр промежуточный результат.

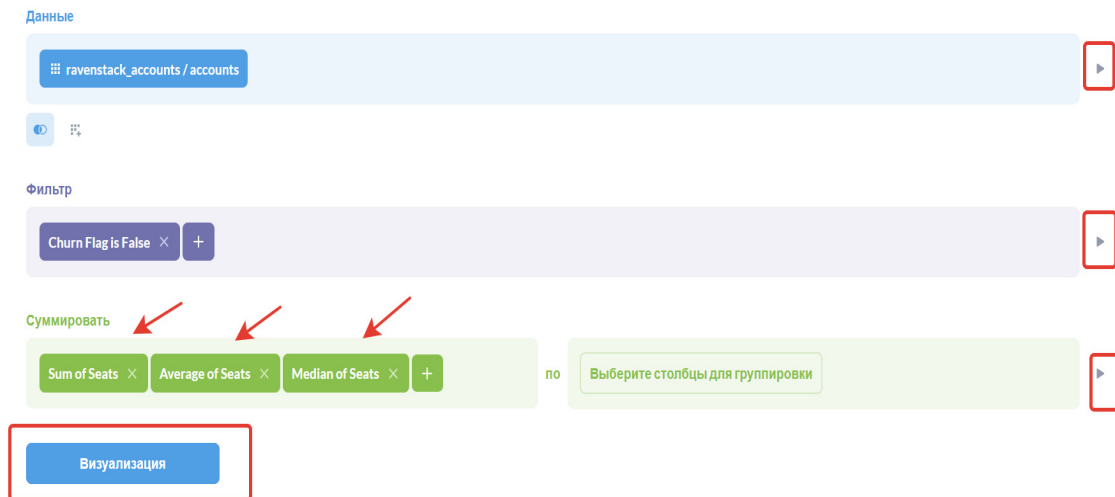


Рис. 2.78. Добавление агрегирующих функций для показателя seats

Построим по запросу визуализацию (рис. 2.79), вначале выберем ее вид – Детали (рис. 2.80), а потом в настройках задав заголовки столбцов (рис. 2.81).

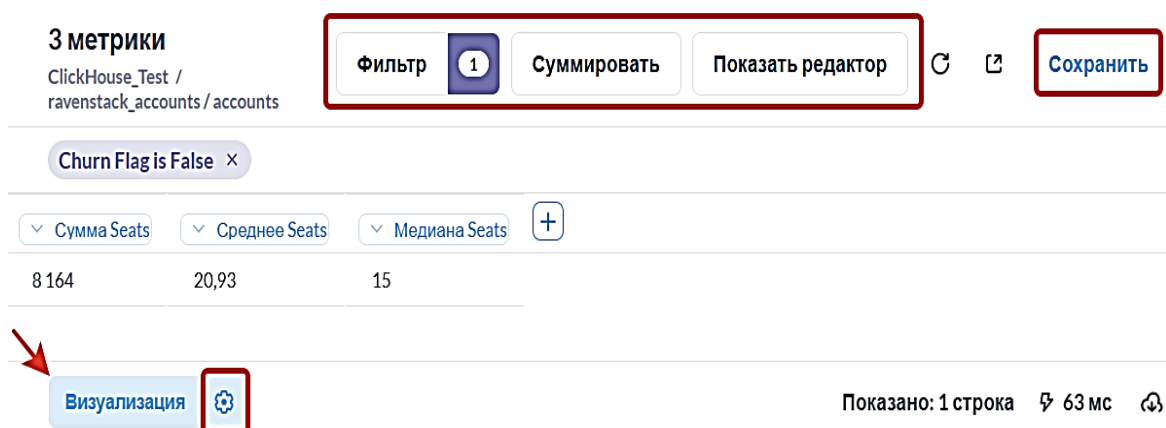


Рис. 2.79. Результат настроенного запроса

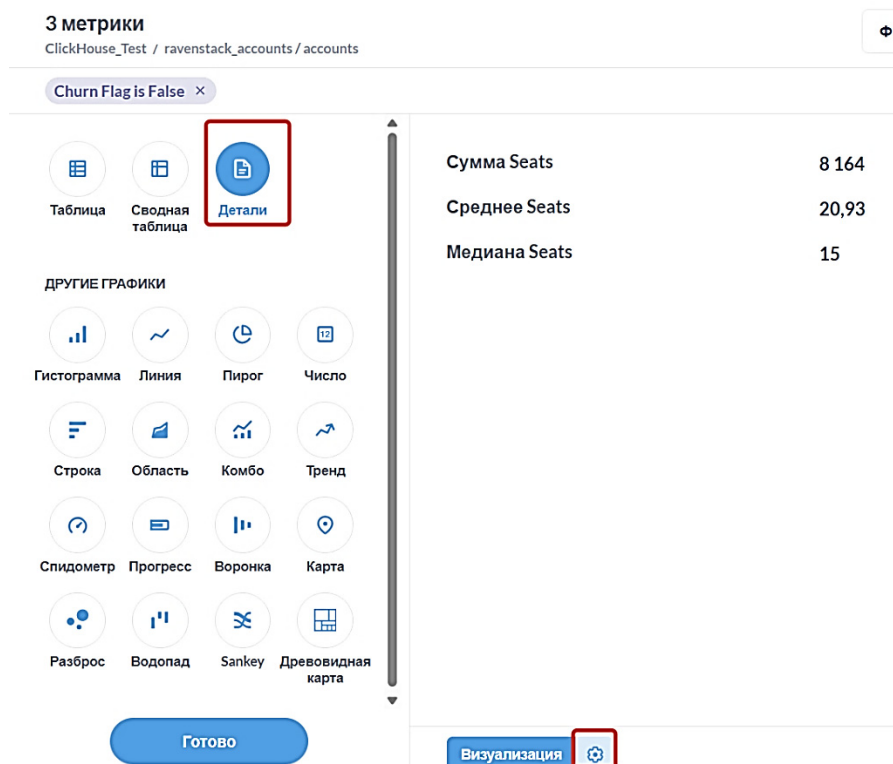


Рис. 2.80. Выбор вида визуализации

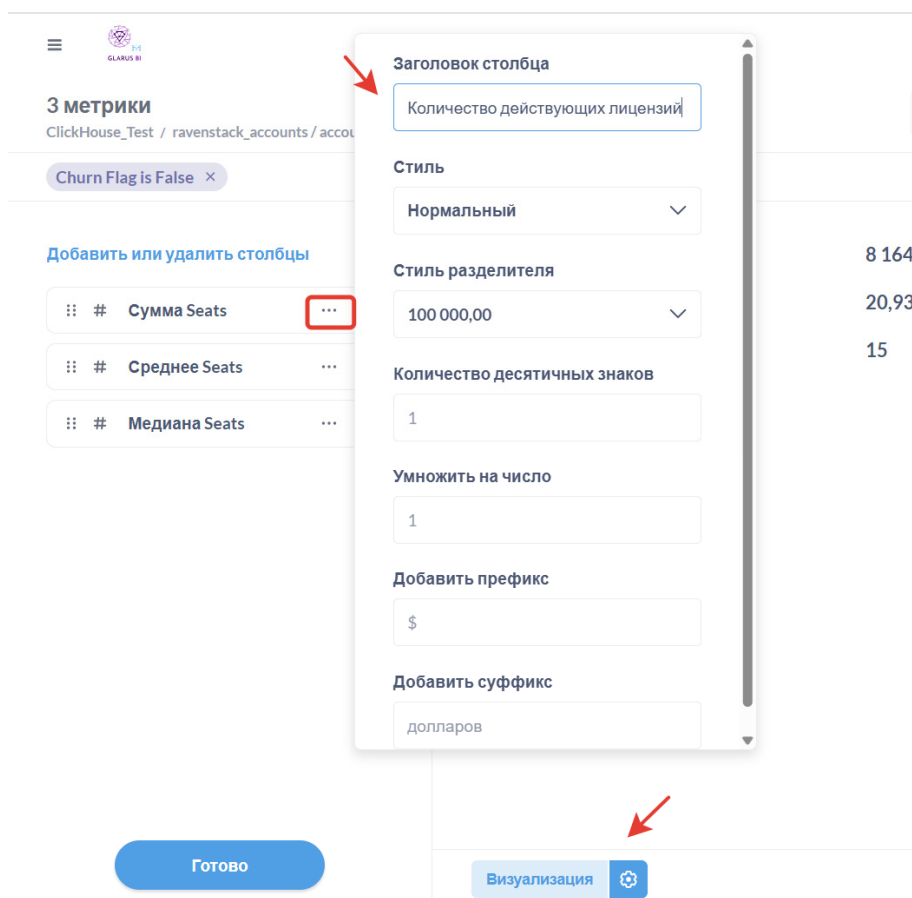


Рис. 2.81. Настройки для визуализации

Сохраним запрос вместе с построенной на его основе визуализацией в коллекцию RavenStack.

Создадим новый запрос, но так как он с небольшими изменениями повторяет предыдущий, то дублируем уже созданный запрос (рис. 2.82) и внесем в него коррективы, зайдя в Фильтр (рис. 2.83), а затем в настройки визуализации (рис. 2.84). При создании дубликата измените название и проконтролируйте, что оно соответствует лимиту на количество символов.

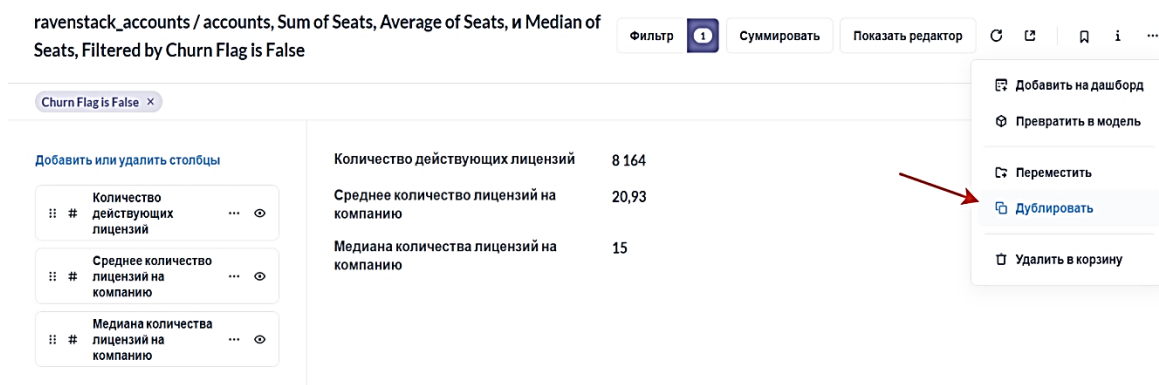


Рис. 2.82. Создание дубликата запроса

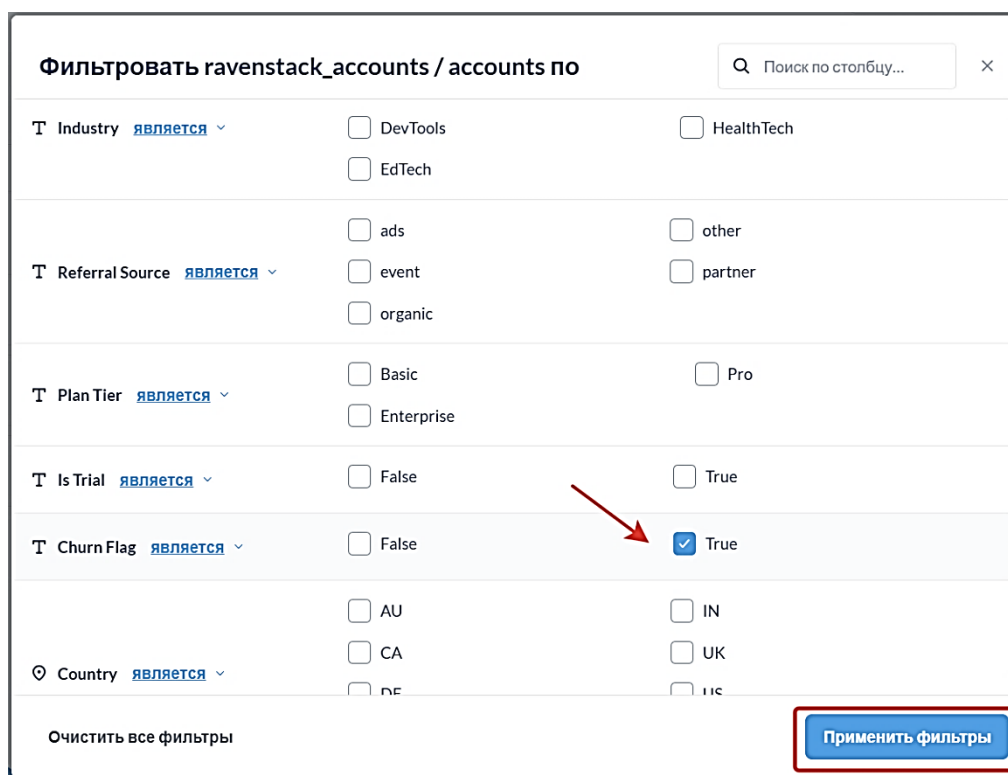


Рис. 2.83. Изменение настроек фильтра (кнопка Фильтр)

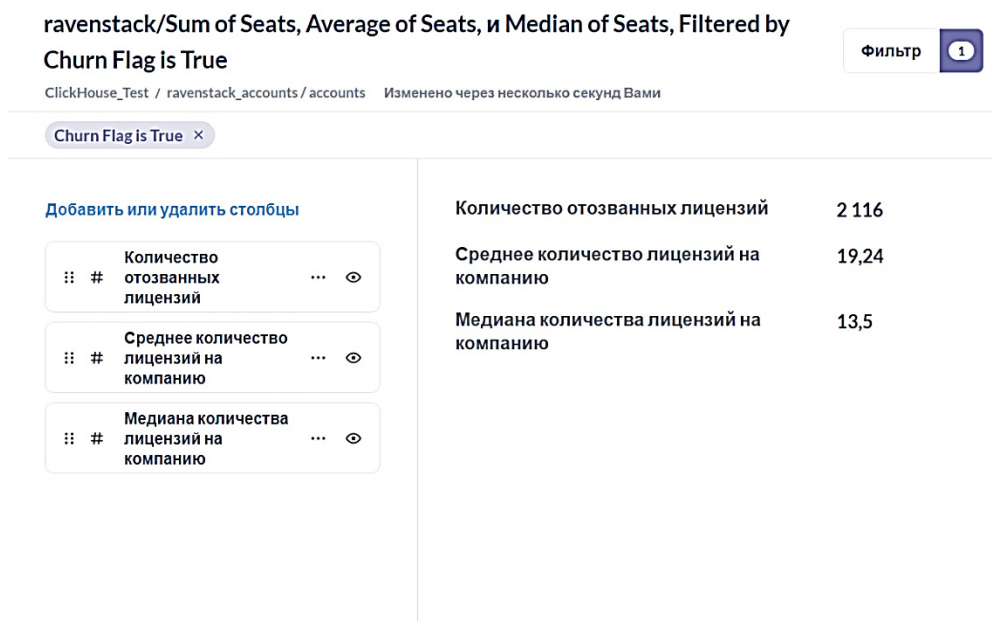


Рис. 2.84. Изменение настроек визуализации

Сохраним все изменения в запросе. И перейдем к созданию следующего запроса, в котором проанализируем новые подключения по месяцам и уточним, кто из подключившихся клиентов уже перестал работать на платформе. Для нового запроса оставим в рассмотрении только три столбца из исходного набора данных (рис. 2.85).

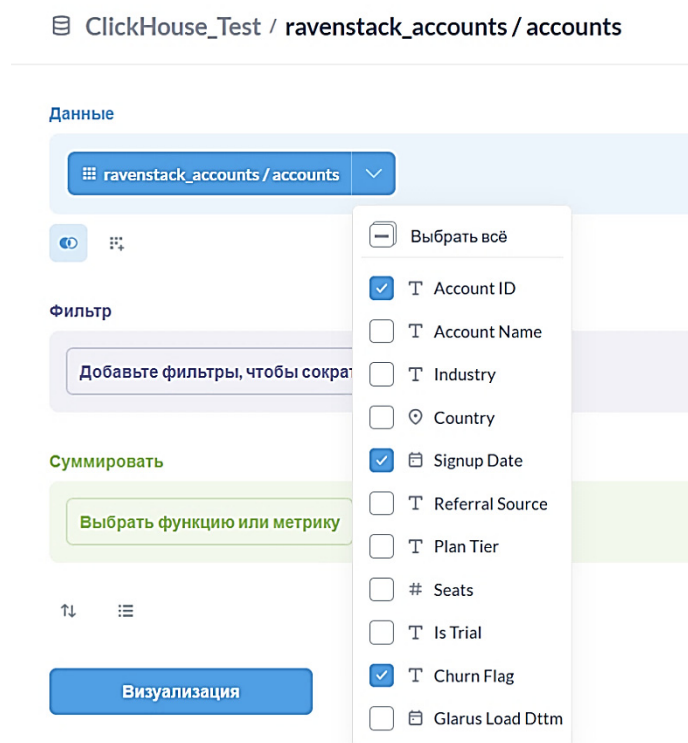


Рис. 2.85. Выбор полей для запроса

Так как в этом запросе используется текстовый показатель `account_id`, то функция агрегации может быть связана только с количеством записей, все остальные варианты для нечислового показателя не определяются. Группировку при агрегации числа записей будем осуществлять по показателю `signup_date`, по умолчанию будет предложен наиболее часто выбираемый вариант – ежемесячные данные, но эту настройку можно и изменить, а также `churn_flag` – показатель, отвечающий за событие: ушел клиент или продолжает использовать платформу (рис. 2.86). И выведем результаты настроек на предварительный просмотр.

Количество по Signup Date: Month и Churn Flag
ClickHouse_Test / ravenstack_accounts / accounts

Данные

ravenstack_accounts / accounts

Фильтр

Добавьте фильтры, чтобы сократить результаты запроса

Суммировать

Количество × + по Signup Date: Month × Churn Flag × +

Предварительный просмотр

Signup Date: Месяц	Churn Flag	Количество
январь 2023	False	13
январь 2023	True	4
февраль 2023	False	11
февраль 2023	True	7

Рис. 2.86. Настройки для агрегации в запросе

Построим визуализацию (рис. 2.87), изменим ее вид на Гистограмму и перейдем к более компактному объединенному представлению (рис. 2.88).

Сохраним запрос.

Построим еще один запрос по сгруппированным наблюдениям, выбрав в качестве метрики для агрегации Количество, а в качестве показателей группировки – `country`, `industry`, `churn_flag` (рис. 2.89, 2.90).

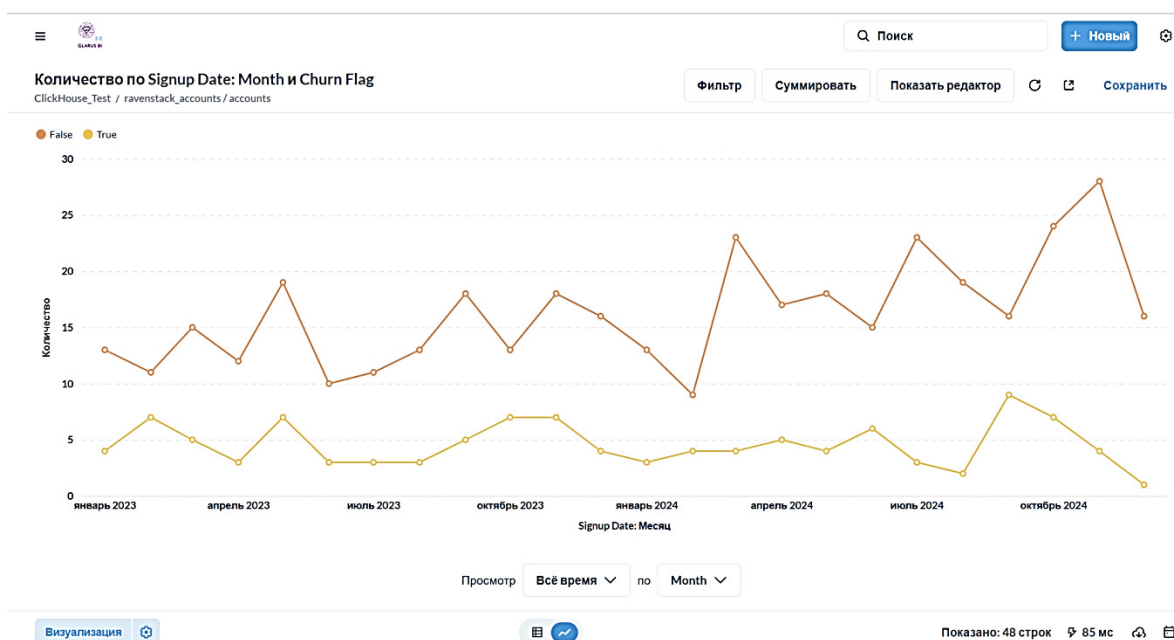


Рис. 2.87. Построенная визуализация

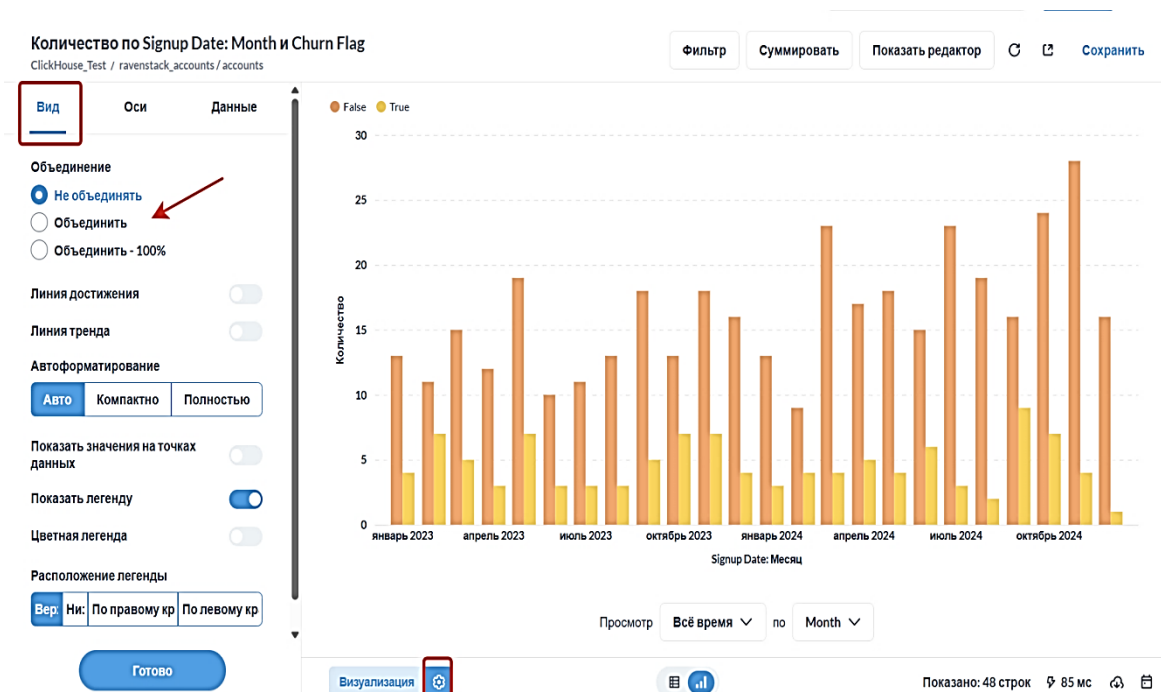


Рис. 2.88. Изменение визуализации

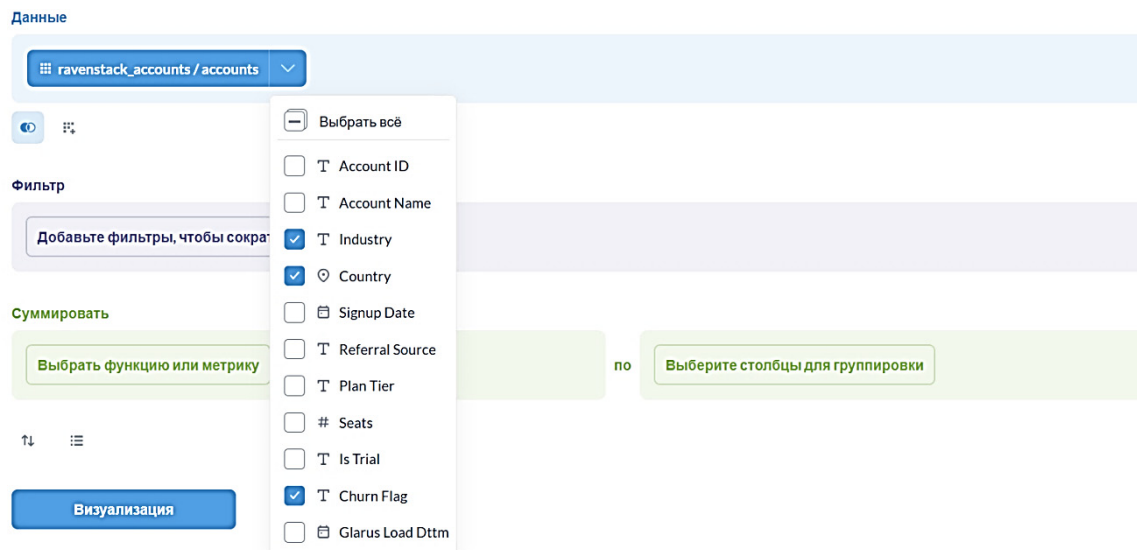


Рис. 2.89. Выбор полей для запроса

Количество по секции

ClickHouse_Test / ravenstack_accounts / accounts

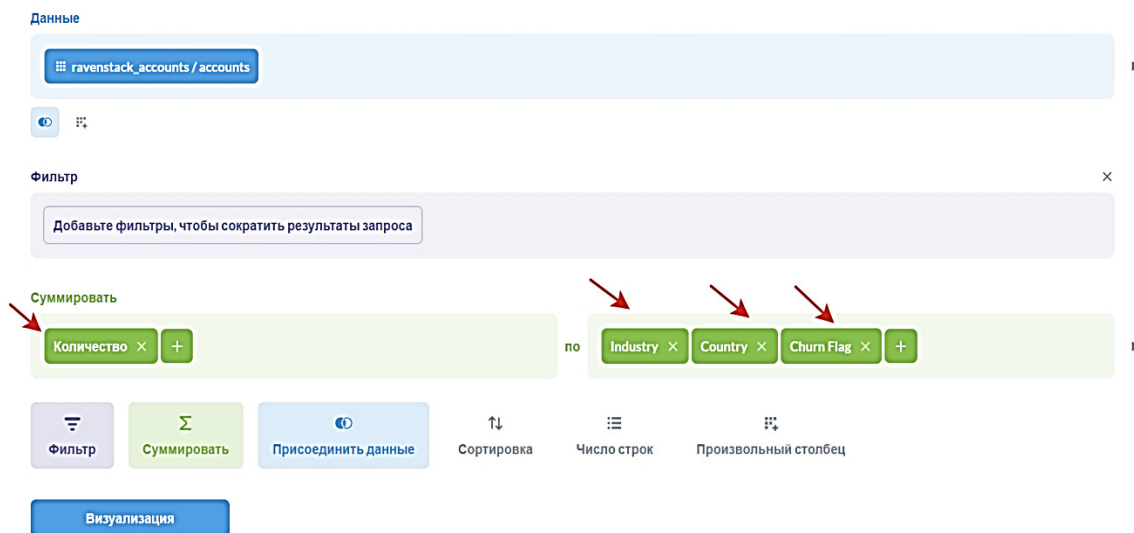


Рис. 2.90. Настройки для запроса

Вид визуализации – Сводная таблица, а настройки представлены на рис. 2.91.

Количество по секции
ClickHouse_Test / ravenstack_accounts / accounts

Фильтр Суммировать Показать редактор Сохранить

Показать в виде иерархии Показывать итоги строки Показывать итоги столбца

Строки

- Industry
- Churn Flag

Столбцы

- Country

Измерения

- Количество

Industry	Churn Flag	AU	CA	DE	FR	IN	UK	US	Итого
Cybersecurity	False	4	5	3	3	9	15	45	
	True	1	1	2		1	1	10	
Всего для Cybersecurity		5	6	5	3	10	16	55	
DevTools	False	7	5	4	6	3	8	45	
	True	2	1	3	3		4	22	
Всего для DevTools		9	6	7	9	3	12	67	
EdTech	False	7	2	2	5	7	6	37	
	True					1	3	9	
Всего для EdTech		7	2	2	5	8	9	46	
FinTech	False	7	6	2	1	12	12	47	
	True	1	1	3		4	2	14	
Всего для FinTech		8	7	5	1	16	14	61	
HealthTech	False	3	1	6	2	8	6	49	
	True		1		2	4	1	13	
Всего для HealthTech		3	2	6	4	12	7	62	
Итого		32	23	25	22	49	58	291	

Готово

Визуализация

Показано: 119 строк 51 мс

Рис. 2.91. Настройки для сводной таблицы по запросу

Создадим дашборд Клиенты RavenStack (рис. 2.92) в коллекции RavenStack. Дашборд может иметь несколько вкладок. Но сначала добавим на него все созданные визуализации (рис. 2.93, 2.94).

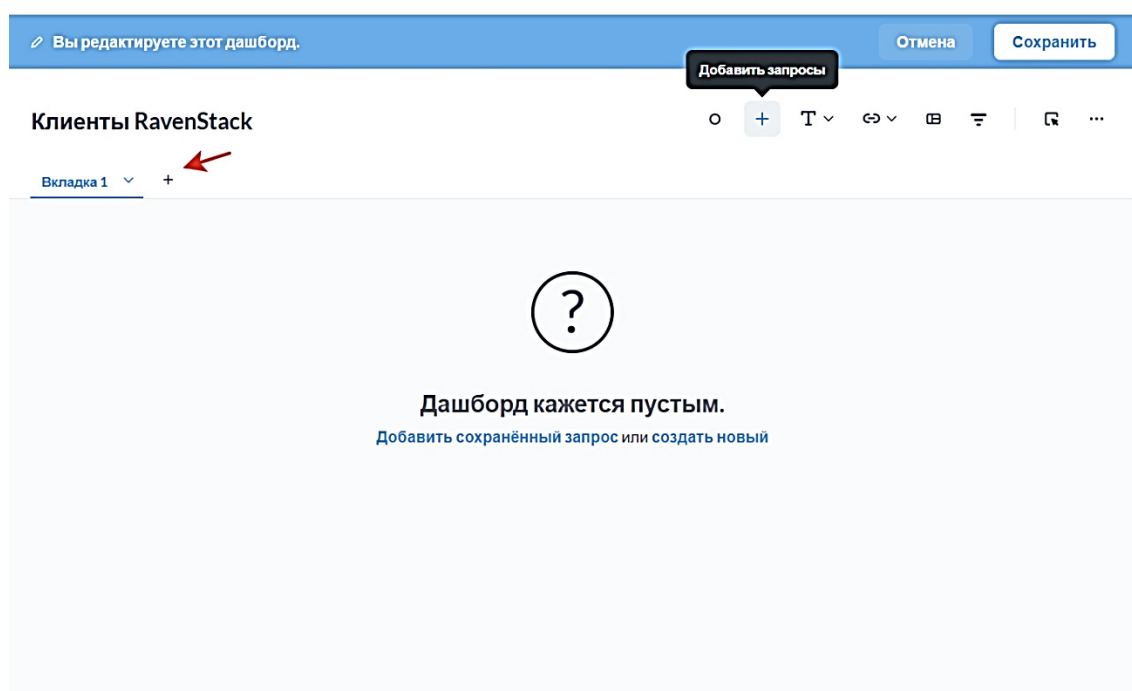


Рис. 2.92. Рабочая область дашборда

Откроем панель для добавления запросов на дашборд (рис. 2.93).

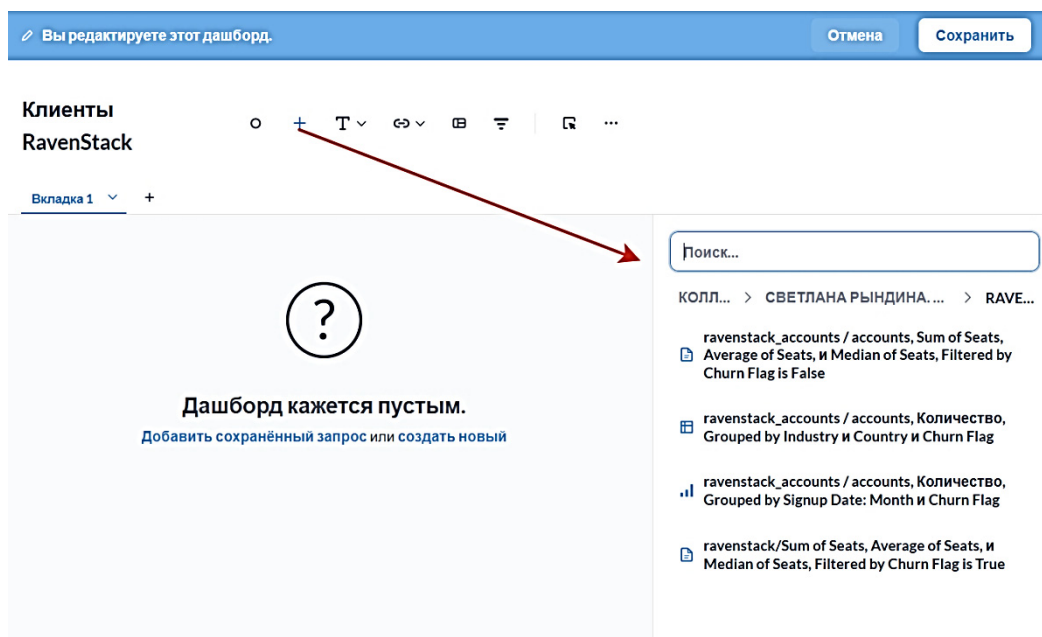


Рис. 2.93. Выбор запросов для размещения на дашборде

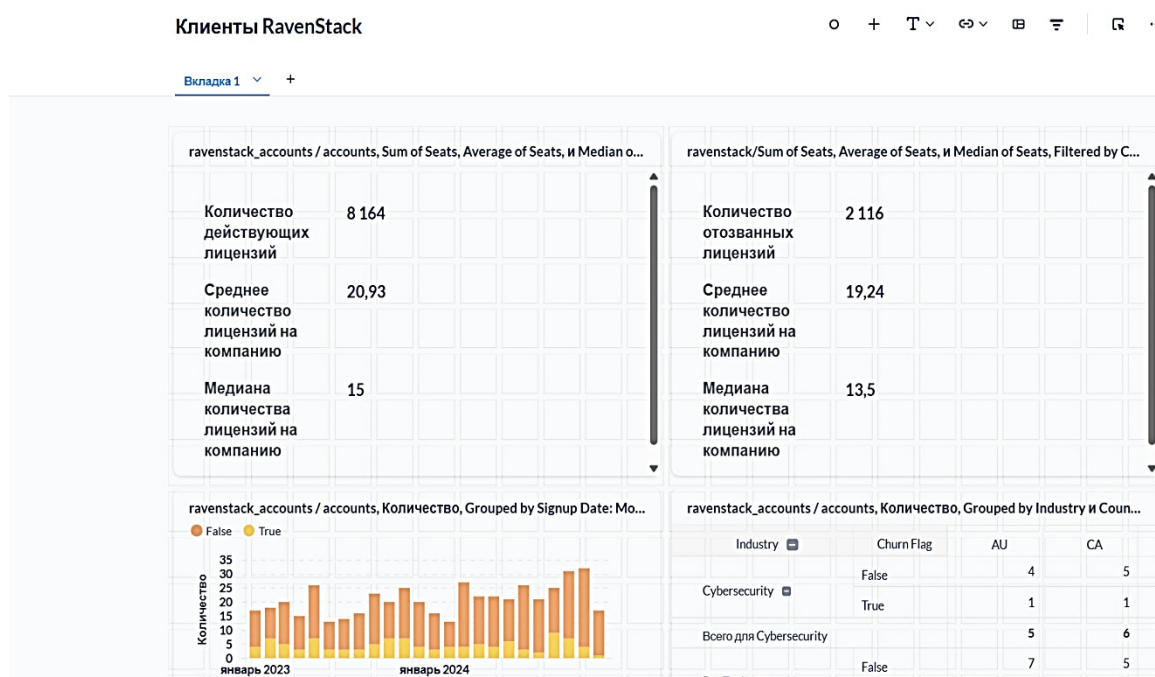


Рис. 2.94. Дашборд с размещенными запросами

Добавим все построенные запросы на дашборд и перейдем в настройки визуализации одного из них (см. рис. 2.94). Пиктографическое меню в правом верхнем углу доступно для каждой активной визуализации на дашборде (рис. 2.95).

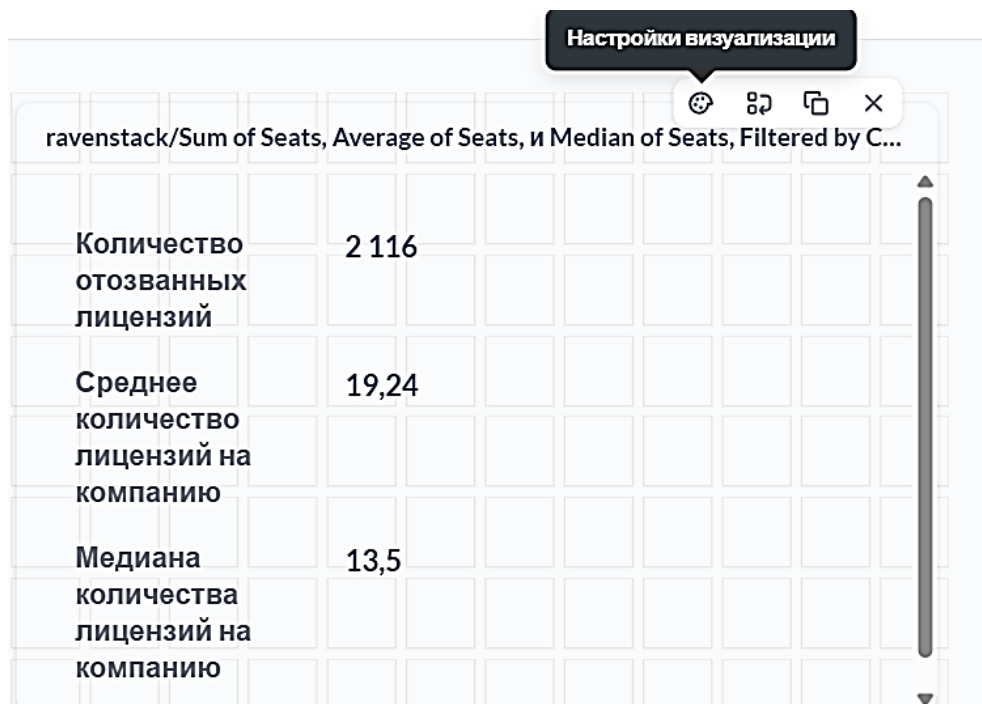


Рис. 2.95. Переход в настройки отдельной визуализации для редактирования

Добавим на дашборд фильтры (рис. 2.96).

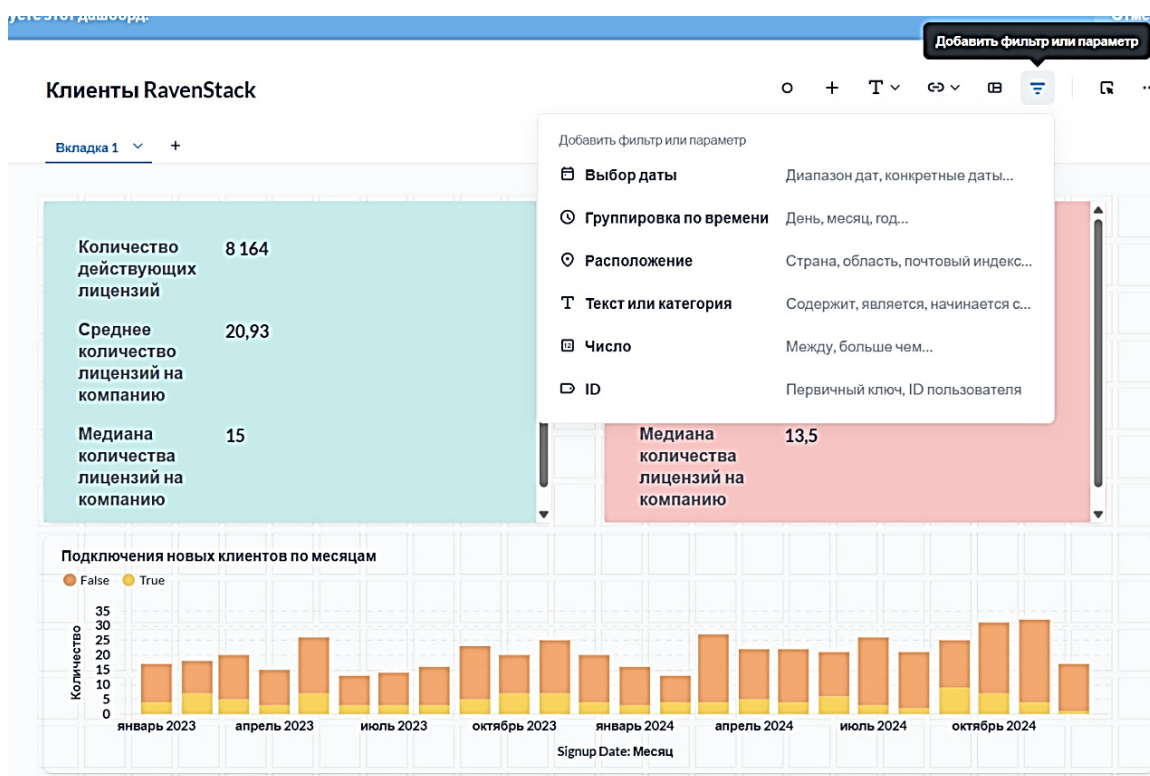


Рис. 2.96. Добавление фильтров

При выборе типа фильтра все визуализации, для которых он может быть применен, выделяются для выбора показателя фильтра. После выбора на карточке одного из запросов показателя с подходящим под этот фильтр типом данных для остальных эту настройку можно активировать автоматически (рис. 2.97).

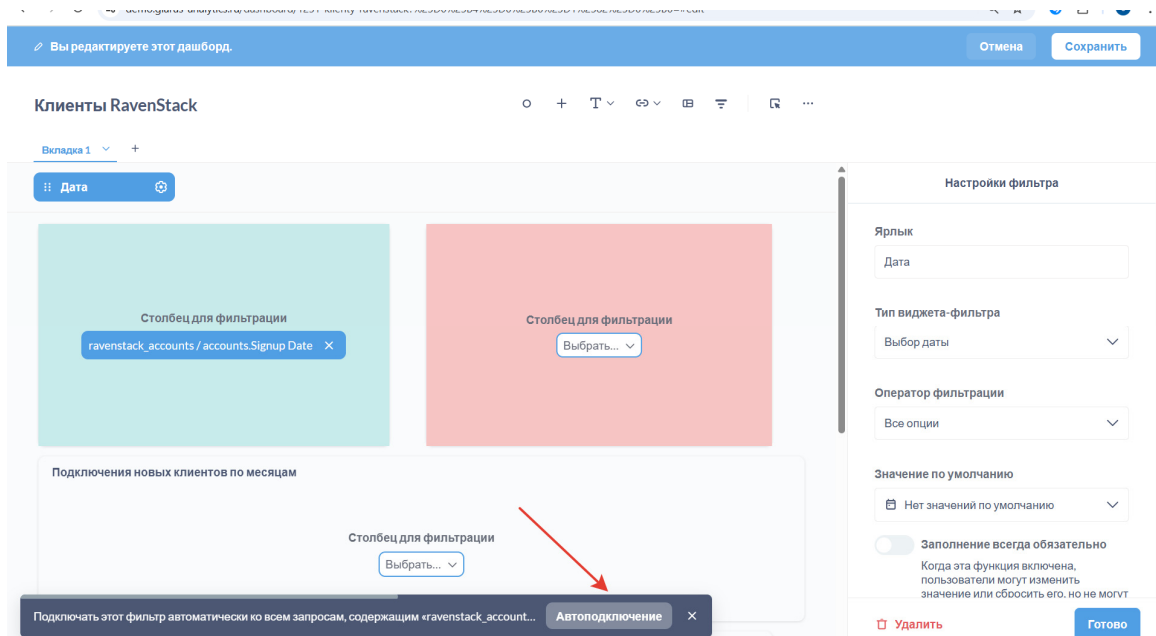


Рис. 2.97. Выбор показателя для фильтрации

Остальные настройки фильтра доступны справа, например, можно изменить ярлык. После добавления фильтров по датам, странам и отраслям сохраним дашборд (рис. 2.98).

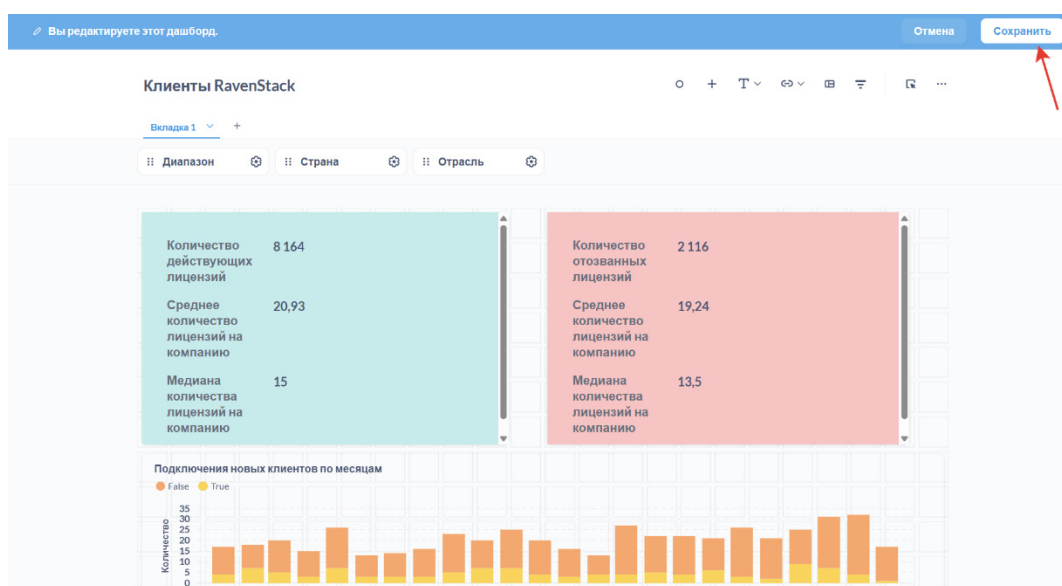


Рис. 2.98. Сохранение дашборда

После сохранения дашборд перейдет в режим демонстрации, в котором можно динамически менять значения фильтров (рис. 2.99). Для внесения изменений требуется перейти в режим редактирования.

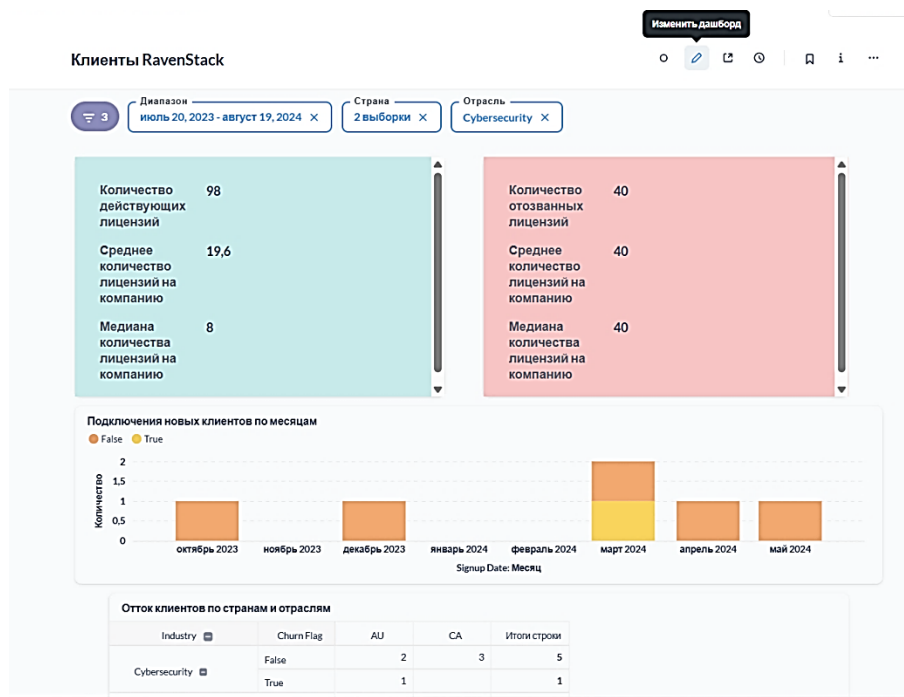


Рис. 2.99. Просмотр дашборда

Для созданного дашборда доступны многочисленные опции (рис. 2.100), в том числе экспорт.

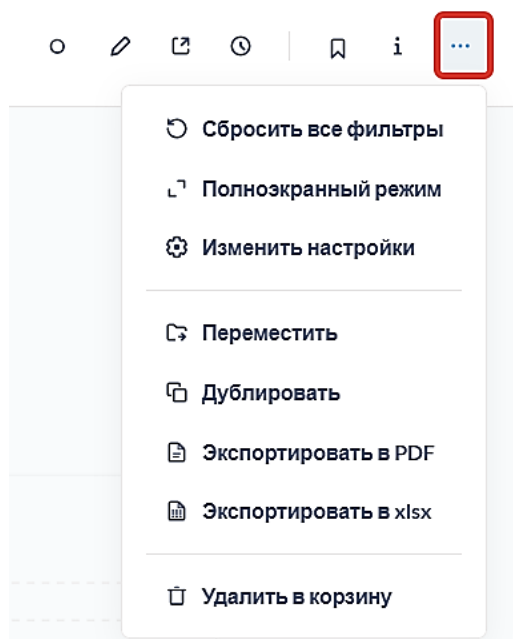


Рис. 2.100. Дополнительное меню для созданного дашборда

Для дальнейшего совершенствования навыков работы на платформе Glarus BI можно воспользоваться справочной документацией [7] и материалами пособия [8].

Самостоятельное задание:

1. Добавьте еще одну вкладку на дашборд и разместите на второй вкладке визуализации по запросам для набора данных по подпискам (файл `ravenstack_subscriptions.csv` нужно перевести в требуемый формат и загрузить на платформу).

2. Экспортируйте обновленный дашборд в документ pdf.

3. L-ПЛАТФОРМА LOGINOM

Аналитическая low-code-платформа Loginom [9] предоставляет инструменты для реализации всех аналитических процессов: от интеграции и подготовки данных до моделирования, развертывания и визуализации (рис. 3.1).

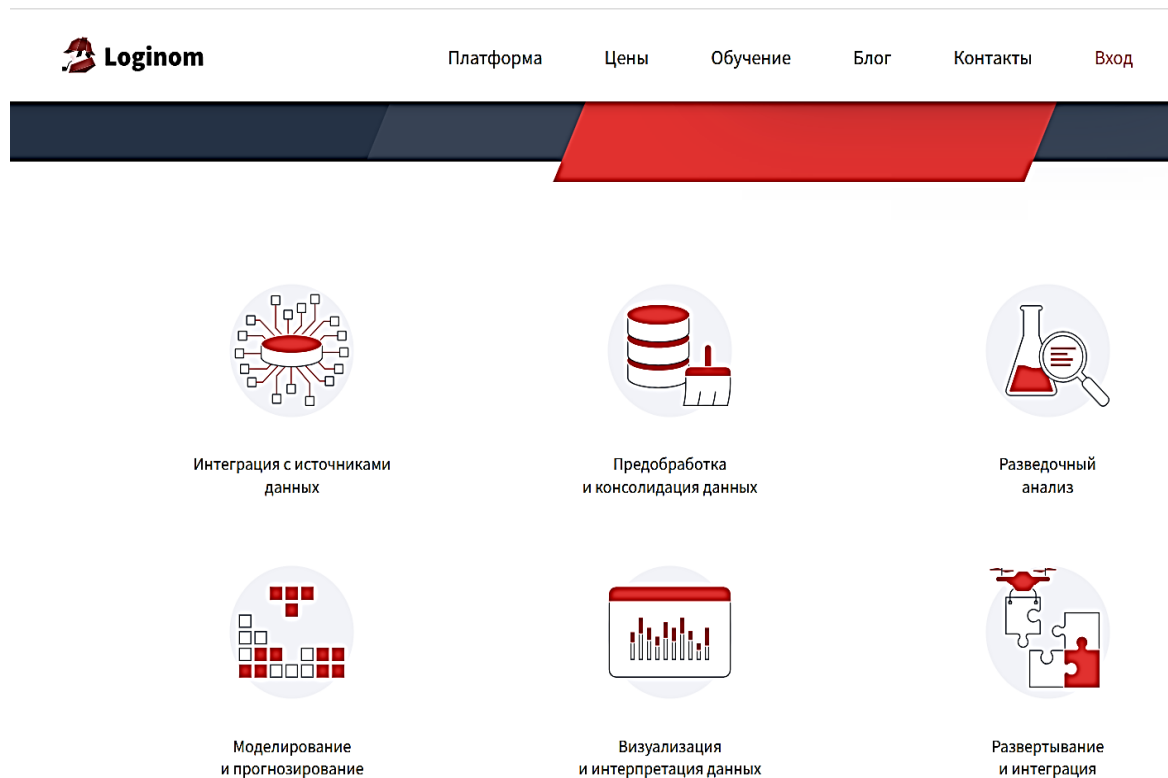


Рис. 3.1. Процессы аналитики платформы Loginom

Для некоммерческого использования есть бесплатная клиентская версия Loginom Community Edition (для профессиональной бизнес-аналитики версия Loginom Personal). После заполнения анкеты (<https://loginom.ru/download>) на указанный email придет ссылка на скачивание продукта.

В момент написания пособия актуальна версия Loginom Community 7.2.8.

В начальном окне можно создать первый пакет с названием RavenStack (рис. 3.2). Все пакеты Loginom имеют расширение .lgr. Расширение для файлов данных в хранилище данных платформы Loginom – .lgd.

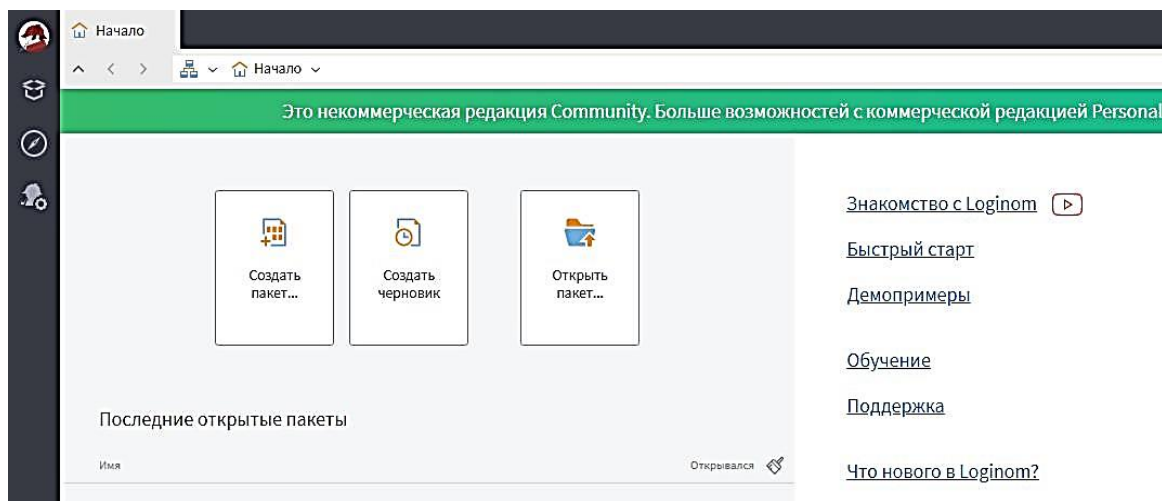


Рис. 3.2 Начальное окно Loginom

Слева располагается удобная вертикальная панель управления (рис. 3.3–3.5).

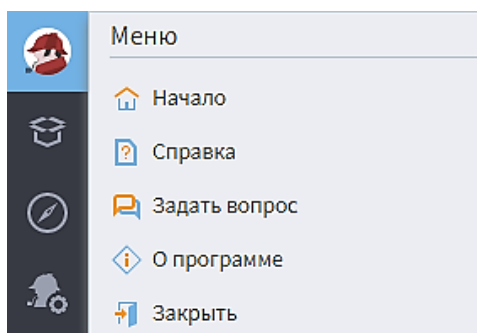


Рис. 3.3. Главное меню платформы

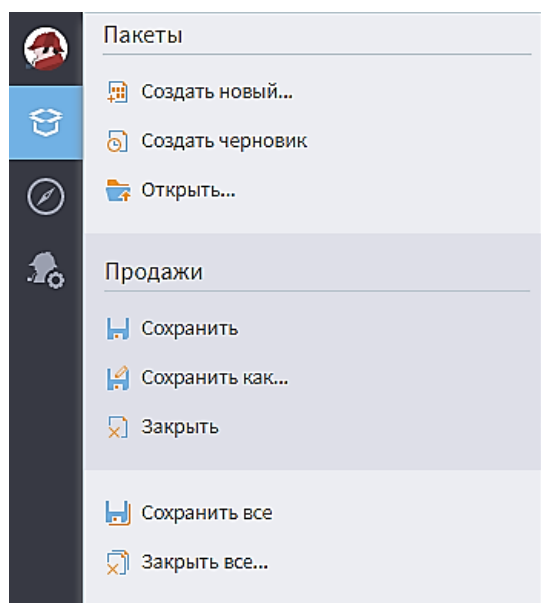


Рис. 3.4. Работа с пакетами

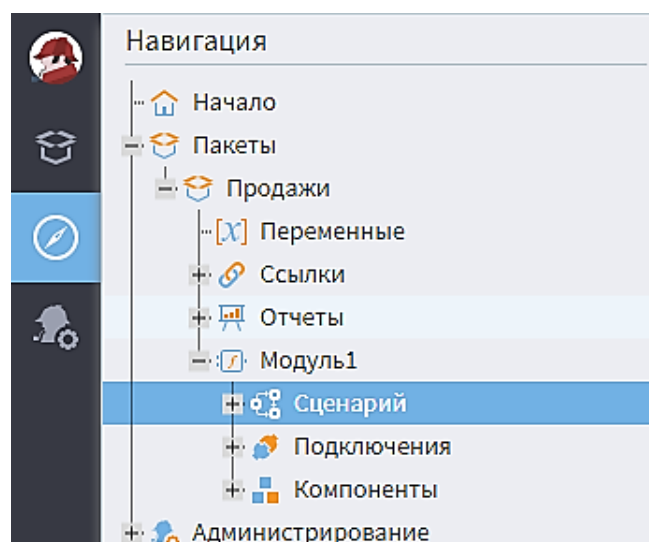


Рис. 3.5. Навигация по структуре пакетов

Перед закрытием программы необходимо сохранить текущий пакет или все открытые во время текущей сессии пакеты, так как в бесплатной версии автосохранение не предусмотрено.

Пакет – это контейнер, который содержит в себе настраиваемые элементы, необходимые для работы с данными в проекте аналитики: переменные, ссылки, отчеты и модули.

По умолчанию в создаваемом пакете обязательно будет Модуль1, содержащий пустой Сценарий.

В рабочей области сценария можно разместить экземпляры компонентов обработки данных, которые называются узлами (рис. 3.6). Узлы добавляются в сценарий перетаскиванием компонентов из панели в рабочую область.

Адресная строка (рис. 3.6) – альтернативный способ перемещения по иерархии объектов.

Для узла доступно пиктографическое меню (рис. 3.7).

В зависимости от назначения узел может иметь входные и выходные порты (рис. 3.8).

Слева располагаются входные порты, справа – выходные. Необязательный порт обозначается пунктирной линией, обязательный – сплошной. Несконфигурированный порт выделяется красным, а сконфигурированный – зеленым. В сценарии соединять можно только порты одинакового вида (табл. 3.1).

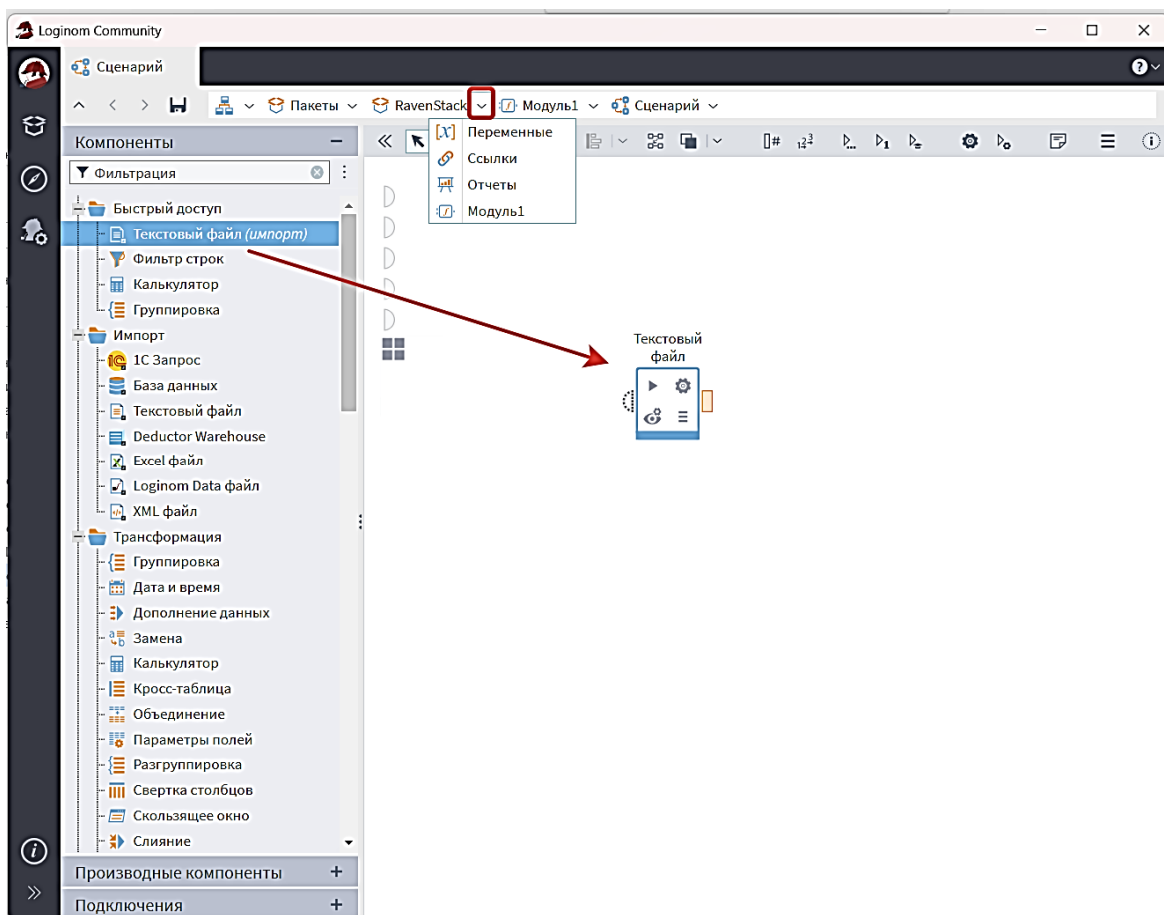


Рис. 3.6. Сценарий с узлом обработки данных Текстовый файл

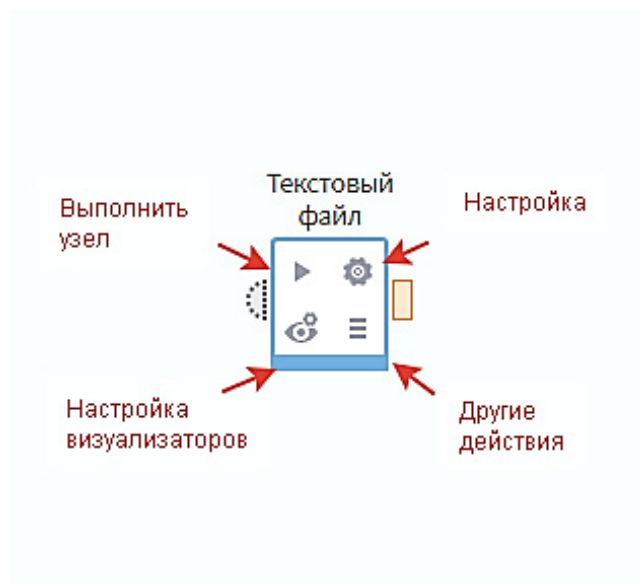


Рис. 3.7. Пиктографическое меню для узла обработки данных

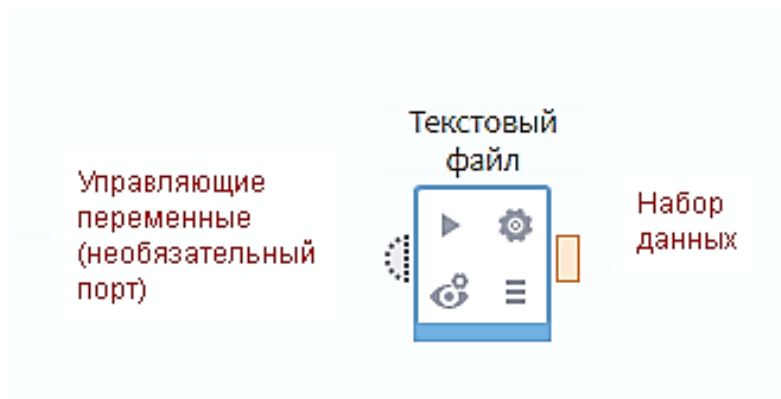


Рис. 3.8. Входной и выходной порты для узла Текстовый файл

Таблица 3.1

Виды портов узлов сценария

Порт	Значок	Описание
Таблица		Представляет собой структурированный набор данных, где все данные упорядочены в двумерную структуру, состоящую из столбцов и строк. В ячейках такой таблицы содержатся элементы данных: строки, числа, даты, логические значения
Переменные		Представляют собой объекты, содержащие только одно значение. С помощью специальных обработчиков имеется возможность преобразовать данные из таблиц в переменные и обратно
Дерево		Представляет собой набор данных в иерархической древовидной структуре
Подключения		Определяют настройки для работы с внешними источниками и приемниками данных

Обработка данных начинается с загрузки данных, возможно из различных источников. Наиболее частыми источниками данных являются файлы или базы данных.

Для размещенного в сценарии пакета RavenStack.lgr узла Текстовый файл выполним настройку (рис. 3.9). Выберем в Проводнике файл ravenstack_accounts.csv с данными для импорта.

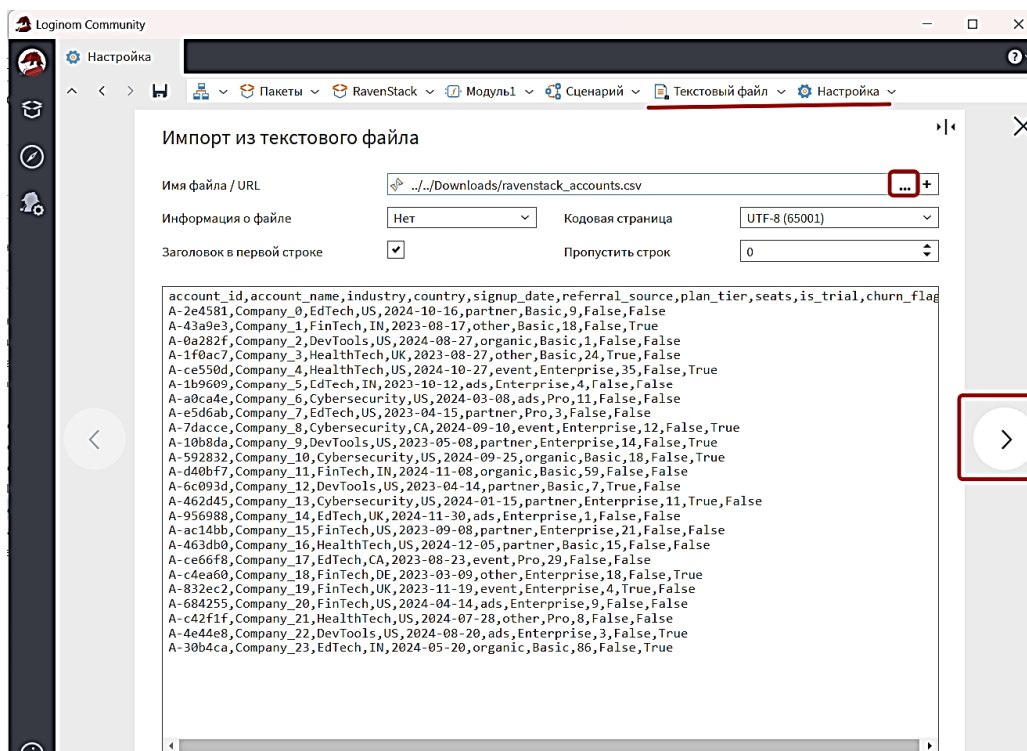


Рис. 3.9. Выбор файла

Если ожидаемый разделитель между столбцами не совпадает с тем, что в файле, то данные в следующем окне попадут в один столбец. Изменим этот разделитель с помощью кнопки Определить автоматически (рис. 3.10).

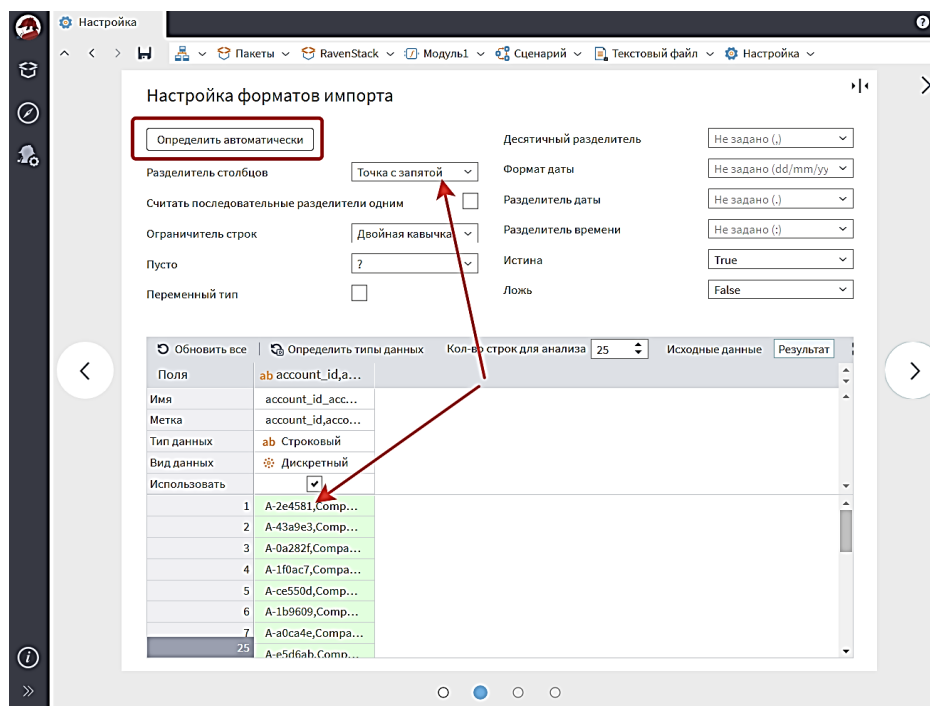


Рис. 3.10. Изменение разделителя для столбцов

Проверим правильность распознавания для каждого столбца типа данных и вида данных (рис. 3.11).

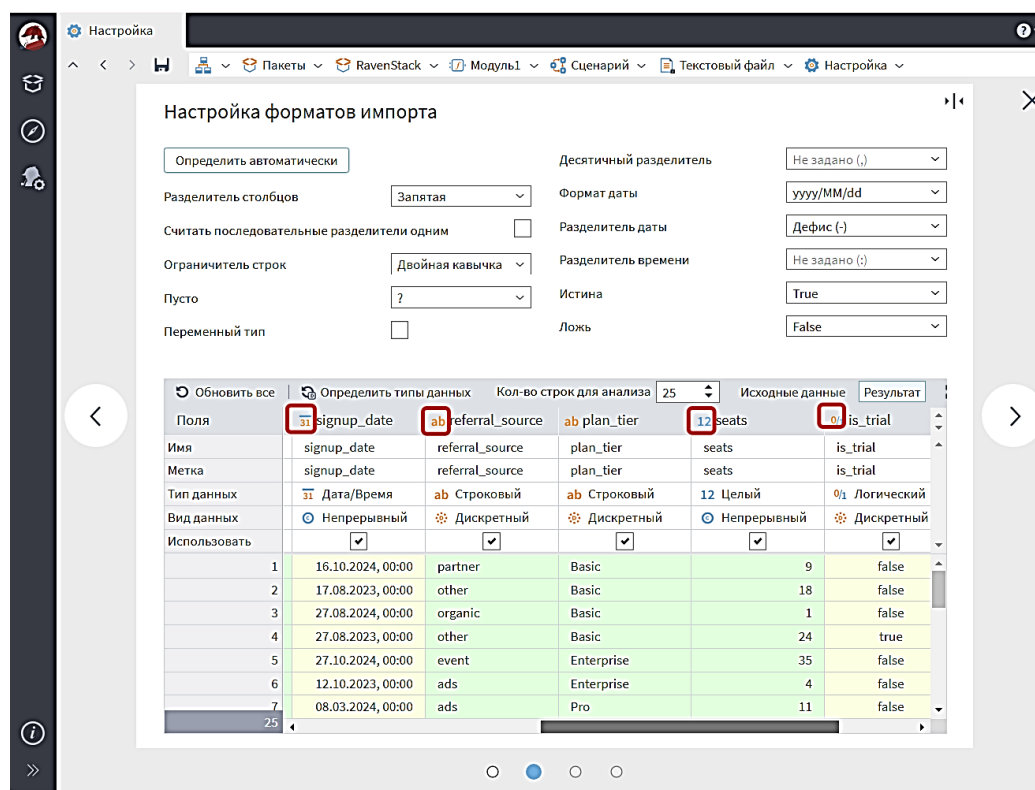


Рис. 3.11. Проверка типа и вида данных для каждого показателя

В следующем окне выводится сводка по показателям (рис. 3.12), а в последнем можно изменить название узла и/или добавить комментарий и сохранить настройки для узла Текстовый файл (рис. 3.13).

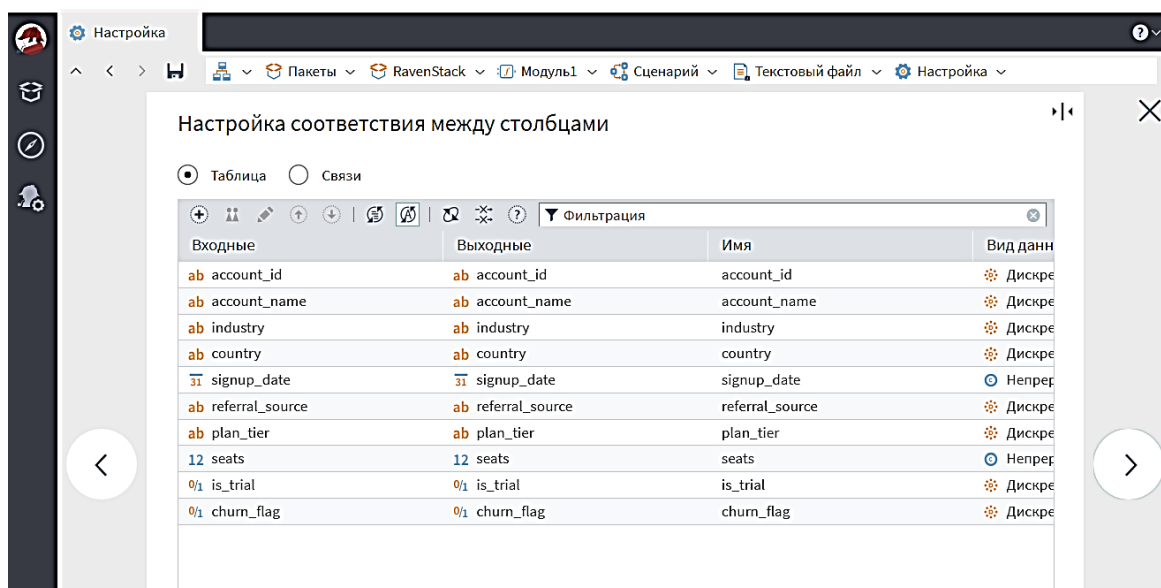


Рис. 3.12. Сводка по показателям файла с данными

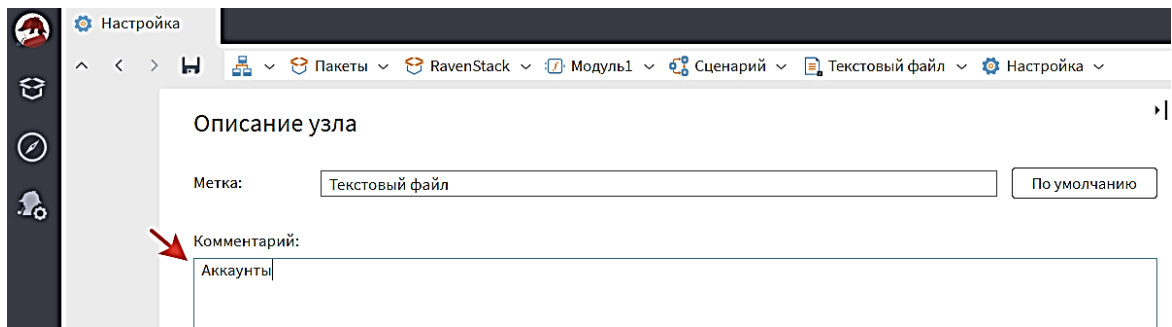


Рис. 3.13. Добавление комментария перед сохранением настроек узла

Выполним узел и выведем содержимое выходного порта на просмотр (рис. 3.14).

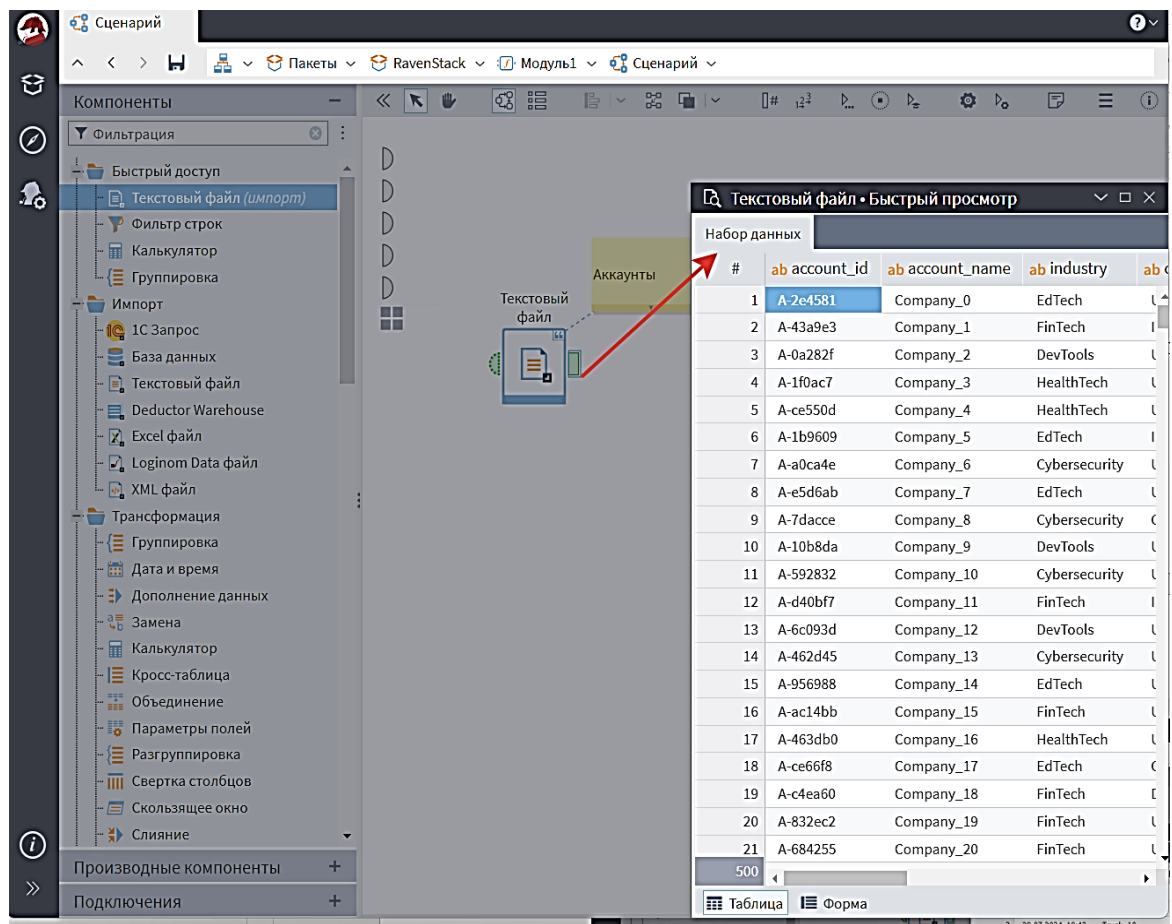


Рис. 3.14. Просмотр содержимого выходного порта узла Текстовый файл

Перейдем в визуализации узла Текстовый файл. Добавим визуализатор Качество данных (рис. 3.15), чтобы убедиться, что пропусков в данных нет и все показатели пригодны для анализа (рис. 3.16).

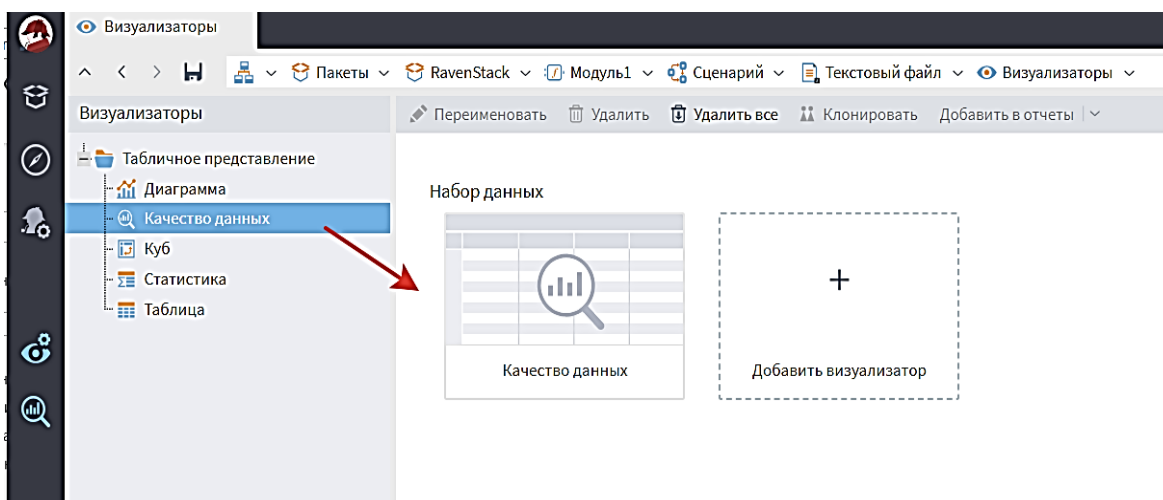


Рис. 3.15. Добавление визуализатора Качество данных

№	Тип	Метка	Вид	Проблемы	Виды проблем	Характеристика набора данных	Значение
7	12	seats		2,20%	Экстремальные - 0,40% (2) Выбросы	Метод определения нетипичных значений	Стандартное откл...
0	ab	account_id		0,00%	Уникальное	Столбцов	8
1	ab	account_name		0,00%	Уникальное	Строк	500
2	ab	industry				Заполненных полей	100,00%
3	ab	country				Полных записей	100,00%
4	31	signup_date				Пригодных столбцов	8 из 8
5	ab	referral_source				Индекс EPV	62,50
6	ab	plan_tier					
8	0/1	is_trial					
9	0/1	churn_flag					

Рис. 3.16. Анализ качества данных текущего датасета

Пропусков действительно нет, account_id и account_name содержат уникальные значения, а наличие выбросов в показателе seats говорит о том, что есть два клиента с очень большим числом лицензий, что сильно отличается от потребления лицензий остальными клиентами.

Визуализатор Статистика (рис. 3.17–3.19) позволяет для категориальных показателей (тип данных – Строковый) установить число уникальных значений и визуализировать распределение наблюдений по этим значениям, а для числовых показателей (тип данных – Вещественный или Целый), а также для показателей даты/времени построить диаграмму размаха. Чтобы показатель с целыми значениями рассматривался как числовой, необходим вид данных Непрерывный.

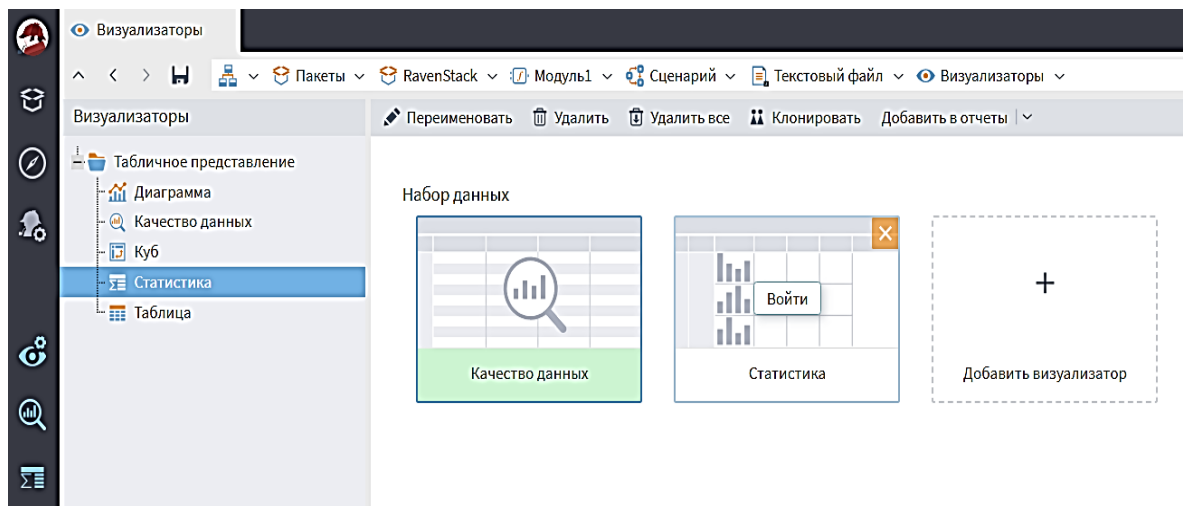


Рис. 3.17. Добавление визуализатора Статистика

Статистика

Пакеты

RavenStack

Модуль1

Сценарий

Текстовый файл

Визуализаторы

Статистика

Уникальные значения

Порядковые статистики

Кол-во строк данных: 500

Гистограмма

№	Метка	Вид	Гистограмма	Диаграмма размаха	Минимум	Максимум	Среднее	Стандарт...	Пропу...	Уникальн...
1	ab account_id		Число значений - 500	Недоступно	8	8	8	0	0	500
2	ab account_name		Число значений - 500	Недоступно	9	11	10,78	0,4604606...	0	500
3	ab industry			Недоступно	6	13	8,844	2,4338264...	0	5
4	ab country			Недоступно	2	2	2	0	0	7
5	31 signup_date				02.01.2023	31.12.2024	27.01.2024...	212,55789...	0	
6	ab referral_source			Недоступно	3	7	5,42	1,4953501...	0	5
7	ab plan_tier			Недоступно	3	10	5,828	2,9078044...	0	3
8	12 seats				1	163	21	21,044718...	0	
9	01 is_trial			Недоступно	Ложь	Истина	Истина	0,3958249...	0	2
10	01 churn_flag			Недоступно	Ложь	Истина	Истина	0,4146611...	0	2

Рис. 3.18. Описательные статистики для показателей набора данных

Гистограмму можно построить для выделенного показателя, сделаем это для показателя seats (рис. 3.19).

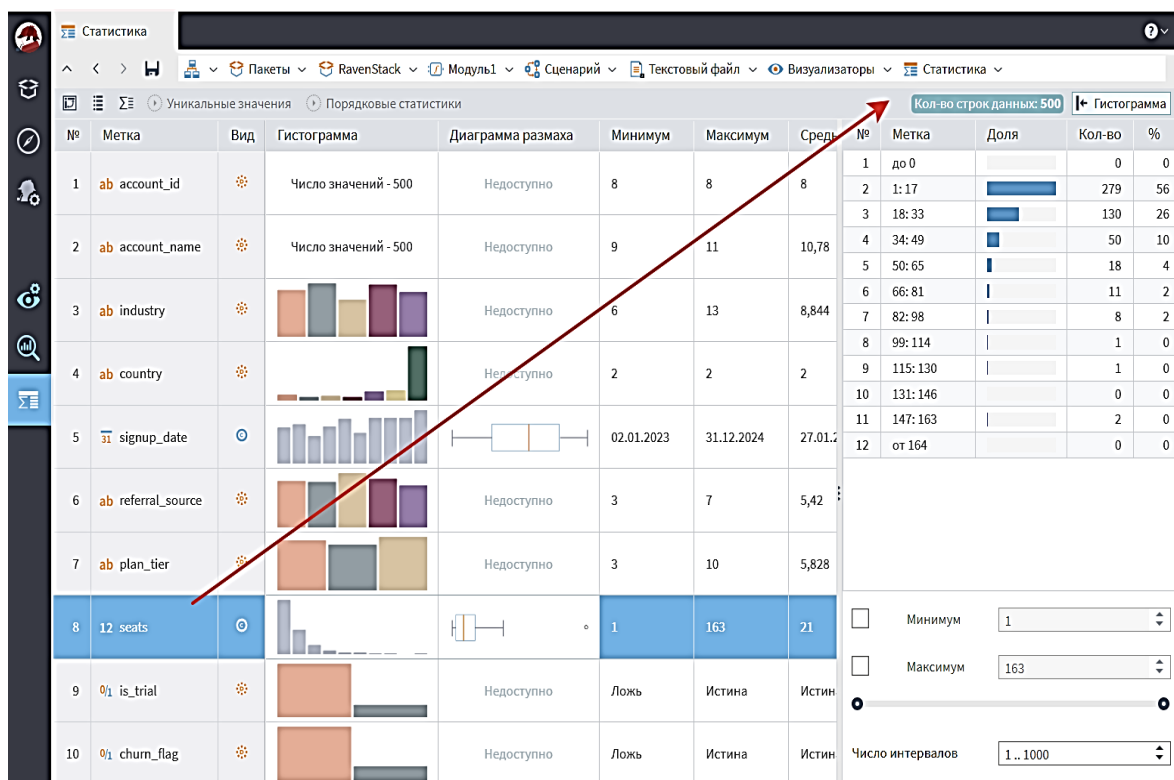


Рис. 3.19. Гистограмма для числового показателя seats

Добавим визуализатор Куб и проведем его настройку. Визуализатор Куб позволяет построить сводную таблицу, в которой в роли группирующих/агрегирующих показателей выступают категориальные переменные (называются в сервисе измерениями), а в роли агрегируемых – числовые показатели (называются в сервисе фактами). Для агрегируемых показателей можно выбрать метрику, определяющую, как находится сводное значение показателя для группы (сумма, среднее, минимальное значение и т.п.). Агрегирующие показатели можно добавлять и для строк (industry, churn_flag), и для столбцов (country). Измерения добавляются перетаскиванием показателей в соответствующую область или выбором из списка по кнопке «+» (рис. 3.20). А факты настраиваются только для столбцов.

Можно добавить факт с помощью пиктограммы в диалоговом окне (рис. 3.21) и выбрать нужную агрегацию из доступных. Но в случае той таблицы, которую строим, достаточно выбрать Количество (это число записей, удовлетворяющих каждой комбинации значений измерений).

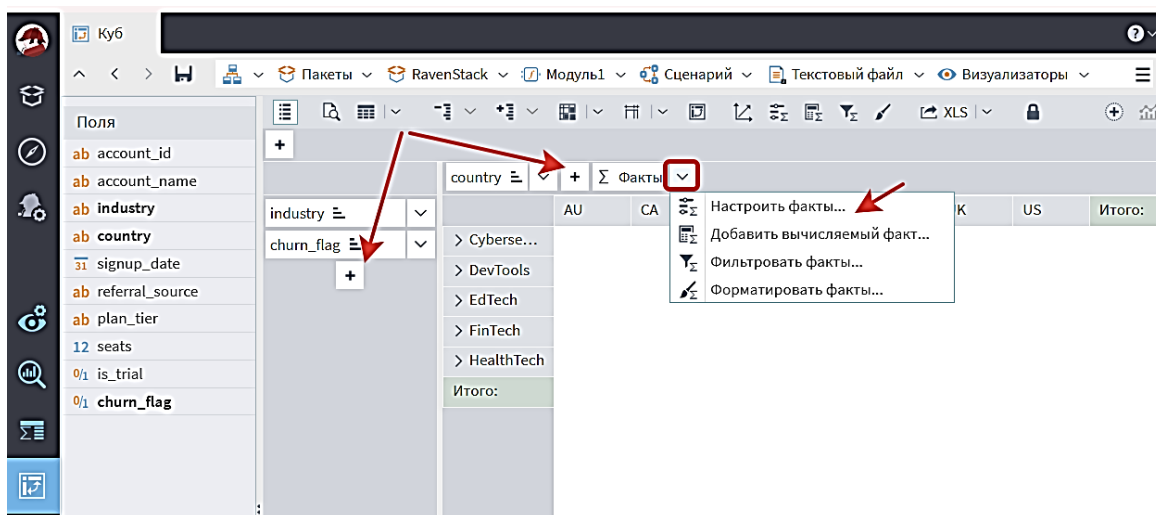


Рис. 3.20. Настройка фактов и измерений для визуализатора Куб

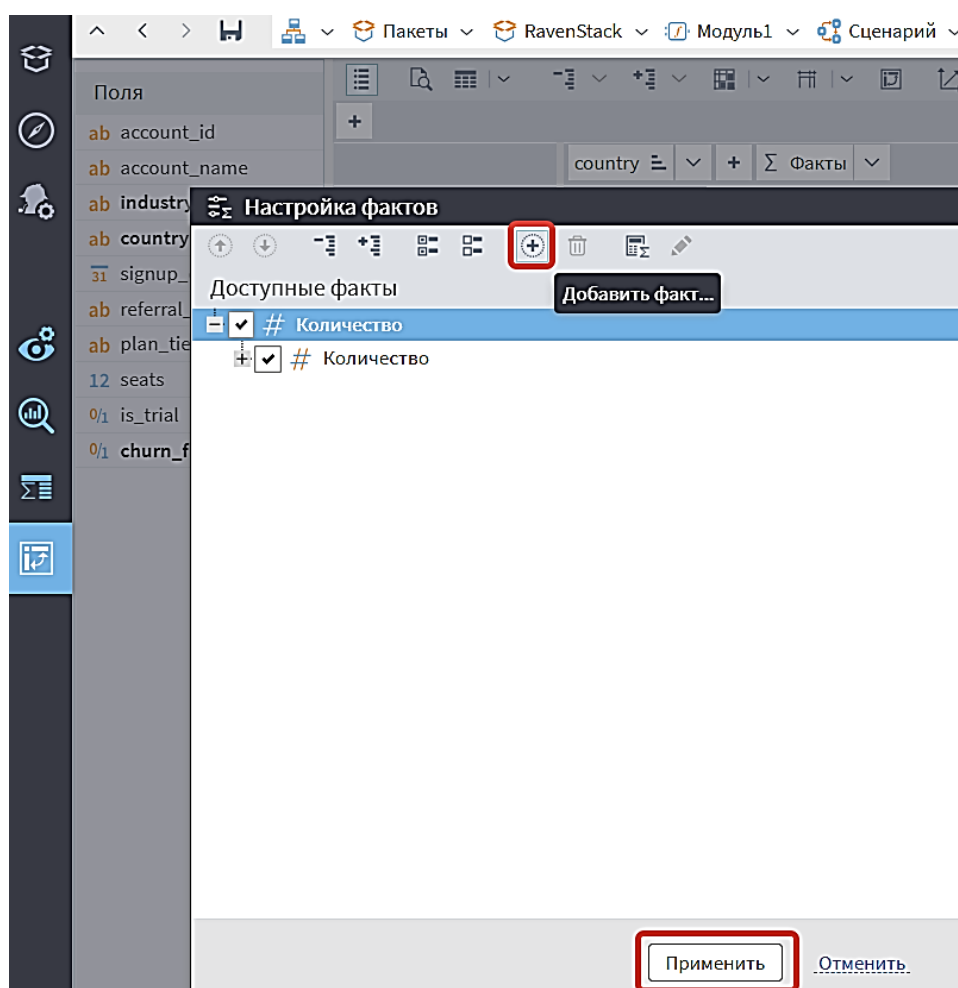


Рис. 3.21. Добавление факта

В результате получим сводную таблицу (рис. 3.22).

country	AU	CA	DE	FR	IN	UK	US	Итого:
Ложь	4	5	3	3	9	15	45	84
Истина	1	1	2	1	1	1	10	16
Итого:	5	6	5	3	10	16	55	100
> DevTools	9	6	7	9	3	12	67	113
> EdTech	7	2	2	5	8	9	46	79
> FinTech	8	7	5	1	16	14	61	112
> HealthTech	3	2	6	4	12	7	62	96
Итого:	32	23	25	22	49	58	291	500

Рис. 3.22. Сводная таблица

Для построения столбчатой диаграммы, анализирующей подключения новых клиентов по месяцам, необходимо выполнить преобразование значений для столбца `signup_date`, организовать новый столбец с усеченной датой. Вернемся в Сценарий и поместим компонент Дата и время из категории Трансформация. Если перемещать компонент в рабочую область, захватывая уже настроенный узел Текстовый файл, то связь между выходным портом предыдущего узла и входным добавленного образуется сама (рис. 3.23), иначе нужно, зажав клавишу мыши, перевести курсор с одного порта на другой.

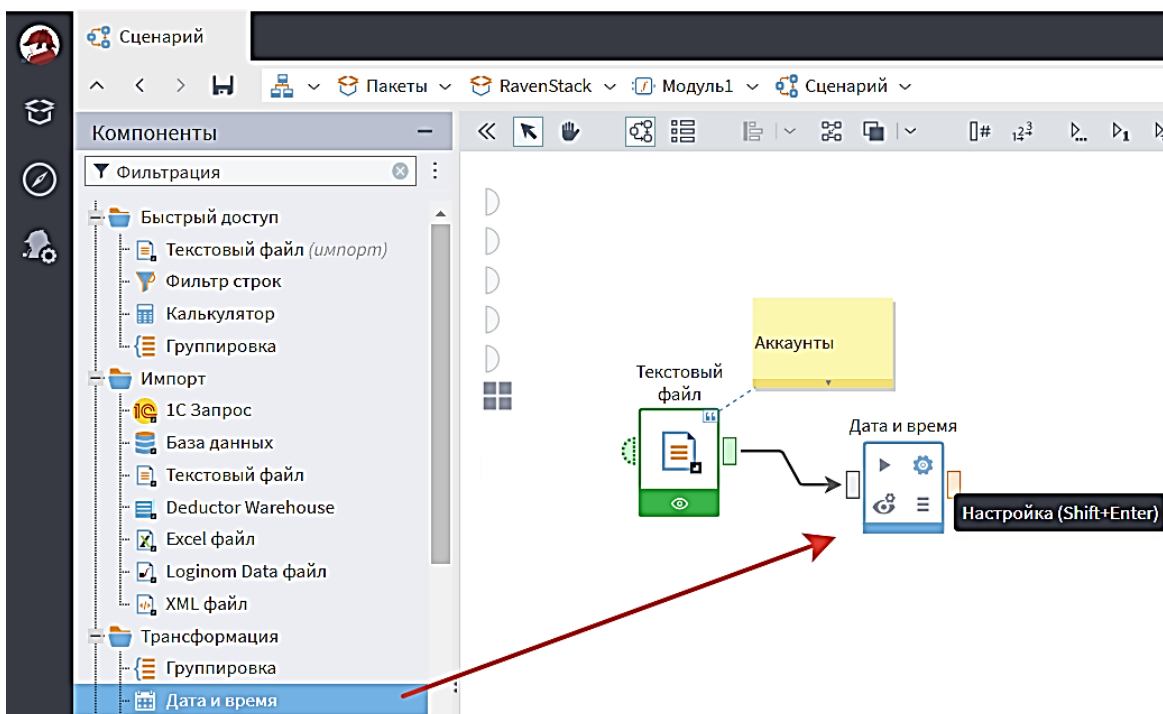


Рис. 3.23. Добавление нового узла и настройка связи между портами узлов

Перейдем в настройки узла Дата и время (рис. 3.24). Узел работает только с одним типом данных (дата/время), если показателя с таким типом данных в наборе, переданном на входной порт узла, нет, то и настроить узел не получится. Выберем нужный формат усеченной даты, оставив в рассмотрении только месяц и год. Заодно можно познакомиться с многообразием возможностей извлечения дополнительной информации из даты. Это одно из очень существенных преимуществ платформы, так как большинство бизнес-данных – это данные с привязкой ко времени: заказы, продажи, поставки и т.д. имеют временную метку. И возможность агрегировать эти данные с любой гранулярностью по времени, вплоть до дня недели, – это очень мощная часть бизнес-аналитики.

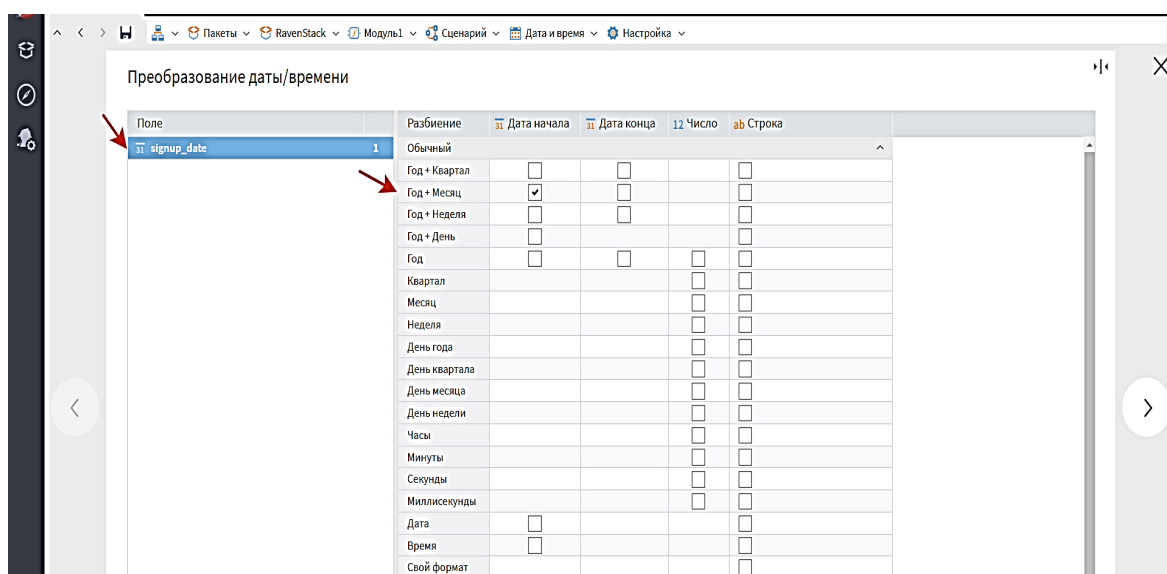


Рис. 3.24. Настройки узла Дата и время

Сохраним настройки, активируем узел и добавим визуализатор Куб для построения новой сводной таблицы уже по данным нового узла (рис. 3.25). В качестве измерений выберем новый показатель `signup_date(год+месяц)` и `churn_flag`, а в роли факта выступит количество (число записей, удовлетворяющих каждому сочетанию значений измерений). Скроем отображение полей набора (левая верхняя пиктограмма), создадим диаграмму (правая верхняя пиктограмма) (рис. 3.26).

Несмотря на замечательные возможности визуализации, основное назначение платформы связано с интеллектуальным анализом данных и теми видами аналитики, которые отвечают более глубокому пониманию данных и тех бизнес-ситуаций, которые эти данные описывают.

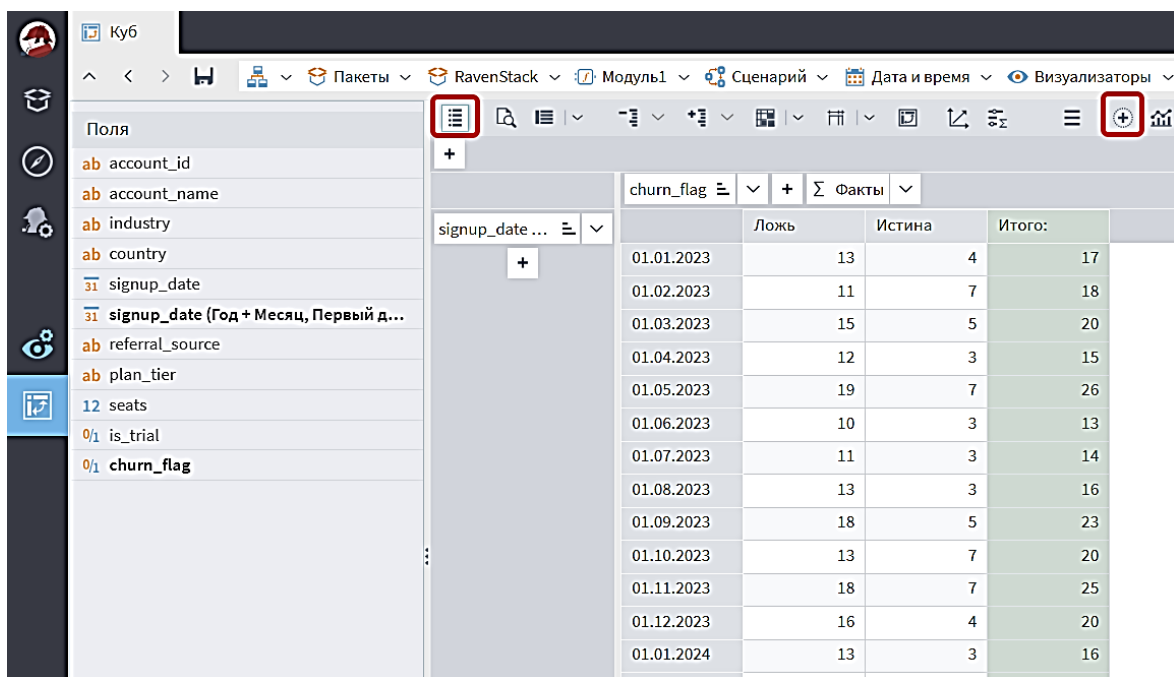


Рис. 3.25. Настройки для Куба

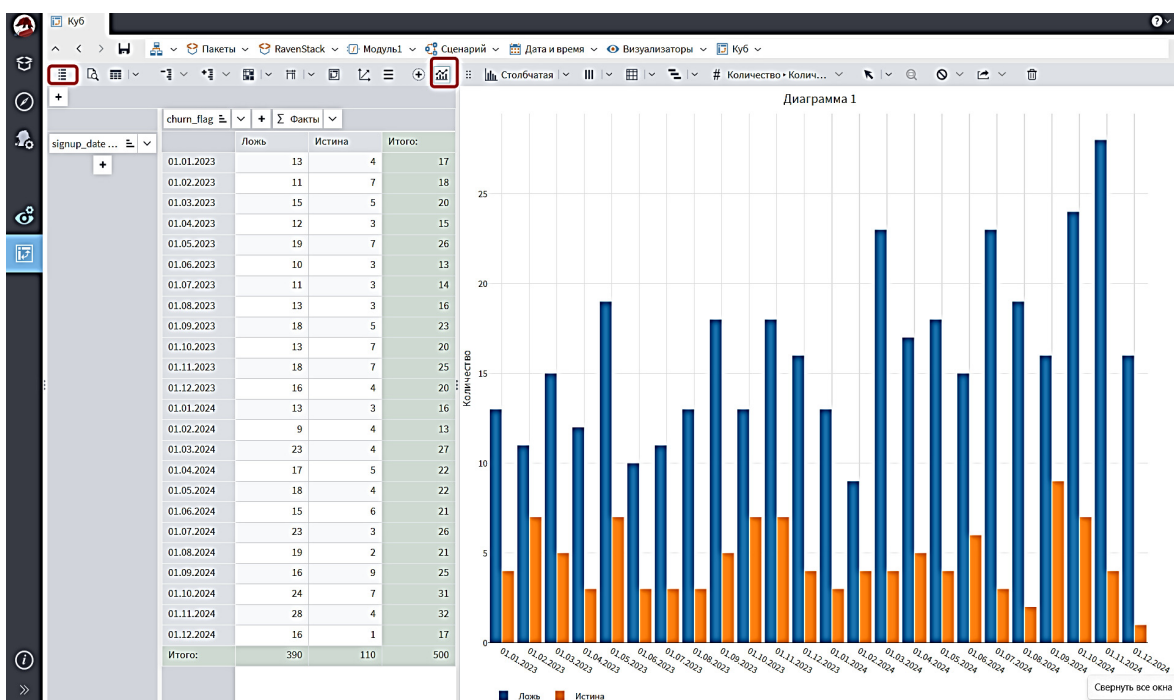


Рис. 3.26. Создание диаграммы по данным сводной таблицы

Это диагностическая аналитика и прогнозная аналитика. Погружение в эти виды аналитики не является целью данного пособия.

Самостоятельное задание:

Добавьте еще один модуль в пакет RavenStack.lgp. Разместите в сценарии этого модуля узлы для обработки данных еще одного файла из набора, постройте подходящие визуализации.

Больше информации по аналитической платформе Loginom и ее использованию для аналитики различных направлений бизнес-деятельности можно в официальном блоге компании [10] и пособиях [11,12].

СПИСОК ЛИТЕРАТУРЫ

1. Kaggle. Платформа для анализа данных и машинного обучения. URL: <https://www.kaggle.com/>
2. Yandex DataLens. URL: <https://yandex.cloud/ru/docs/datalens/>
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. Visiology Platform 3.0. URL: <https://visiology-doc.atlassian.net/wiki/spaces/v3/pages/2393060/Visiology+Platform+3.0>
5. Visiology Platform 3.13. URL: <https://visiology-doc.atlassian.net/wiki/spaces/3v13/overview>
6. Glarus BI. URL: <https://glarus-bi.ru/>
7. Документация Glarus BI. URL: <https://glarus-bi.ru/documentation/index.html>
8. Рындина С. В. Интеллектуальные информационные системы и технологии: системы Business Intelligence (Glarus BI). Пенза : Изд-во ПГУ, 2025. 64 с.
9. Loginom. URL: <https://loginom.ru/>
10. Loginom. URL: <https://loginom.ru/blog>
11. Рындина С. В. Интеллектуальные информационные системы и технологии: системы Business Intelligence (ML-платформа Loginom) : учеб.-метод. пособие. Пенза : Изд-во ПГУ, 2022. 68 с.
12. Рындина С. В. Практические вопросы управления цифровой логистикой : учеб. пособие. Пенза : Изд-во ПГУ, 2025. 118 с.

Учебное издание

Рындина Светлана Валентиновна

**Интеллектуальные информационные
системы и технологии. Аналитические
платформы для работы с данными**

Редактор *М. М. Кузнецова*
Технический редактор *Н. В. Иванова*
Компьютерная верстка *Н. В. Ивановой*
Дизайн обложки *О. И. Жаровой*

Подписано в печать 25.11.2025.
Формат 60×84¹/₁₆. Усл. печ. л. 6,28.
Тираж 21. Заказ № 643.

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

