# Практическая работа №7-8. Визуализация данных

## Введение

Практическая работа заключается в знакомстве с бесплатной платформой Yandex DataLens <https://datalens.yandex.ru> для визуализации данных в формате построения динамических отчетов. В данной работе студенту предлагается ознакомиться с базовыми возможностями платформы для визуализации данных на основе прохождения пошаговой демонстрации с построением визуализаций. Также для студентов представлено небольшое самостоятельное задание в конце настоящего материала.

Yandex DataLens – это инструмент, разработанный компанией Яндекс, который предоставляет возможность анализировать и визуализировать большие объемы данных.

DataLens позволяет пользователям создавать интерактивные дашборды и отчеты, которые помогают визуализировать и понять сложные данные. Он предоставляет различные инструменты для фильтрации, сортировки и анализа данных, а также функции прогнозирования и моделирования.

DataLens может использоваться в различных областях, включая бизнес-аналитику, маркетинг, финансы, науку и другие. Он помогает пользователям принимать более информированные решения на основе данных и выявлять скрытые закономерности и тренды.

Для запуска службы DataLens на настроенной гостевой ОС необходимо перейти в директорию datalens и запустить docker container:

**> cd ~/datalens**

**> sudo HC=1 docker compose up**

После включения, в браузере пользовательской ОС можно запустить графический интерфейс службы и ознакомиться с базовым UI функционалом инструмента, введя в строку адреса:

**http://<guest-ip>:8080/**

В результате после перехода вы увидите экран работы с воркбуками.

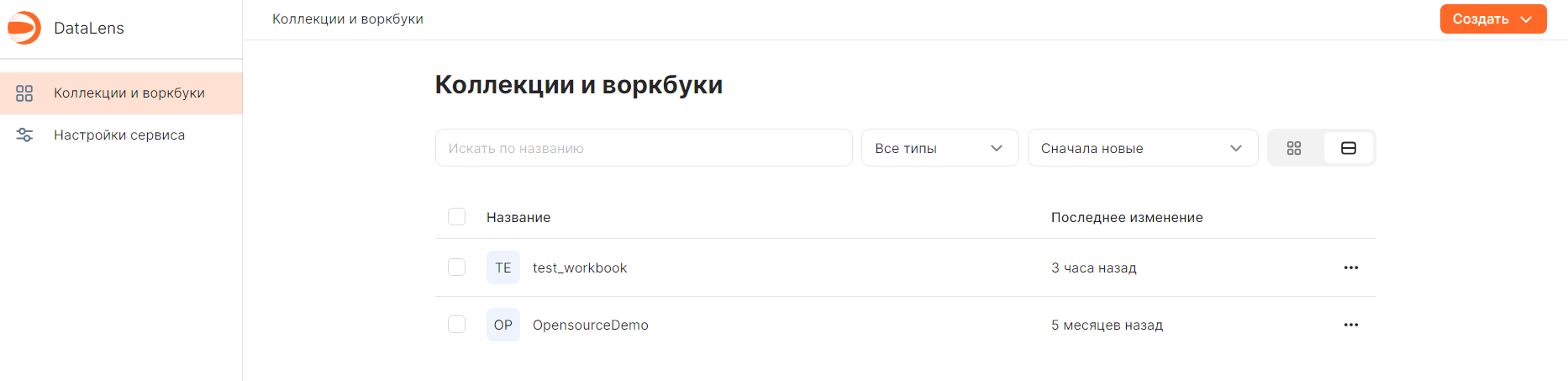


Рисунок 1. Приветственный экран

Настройка интерфейса не имеет широкой вариативности, и доступно лишь несколько функций.

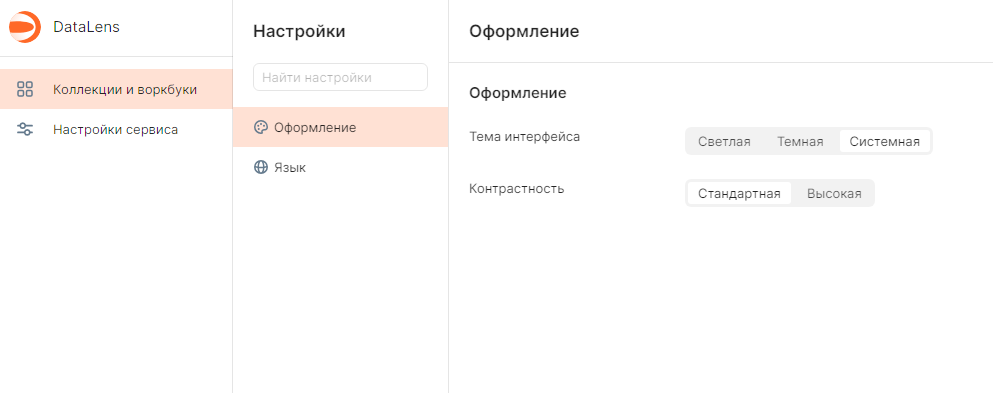


Рисунок 2. Настройки интерфейса и языка

В окне выбора воркбуков сразу же имеется возможность просмотреть демо-дашборд и чарты для него. Перейдем в него и увидим краткую документацию к построению чартов на различных вкладках демонстрационного дашборда.

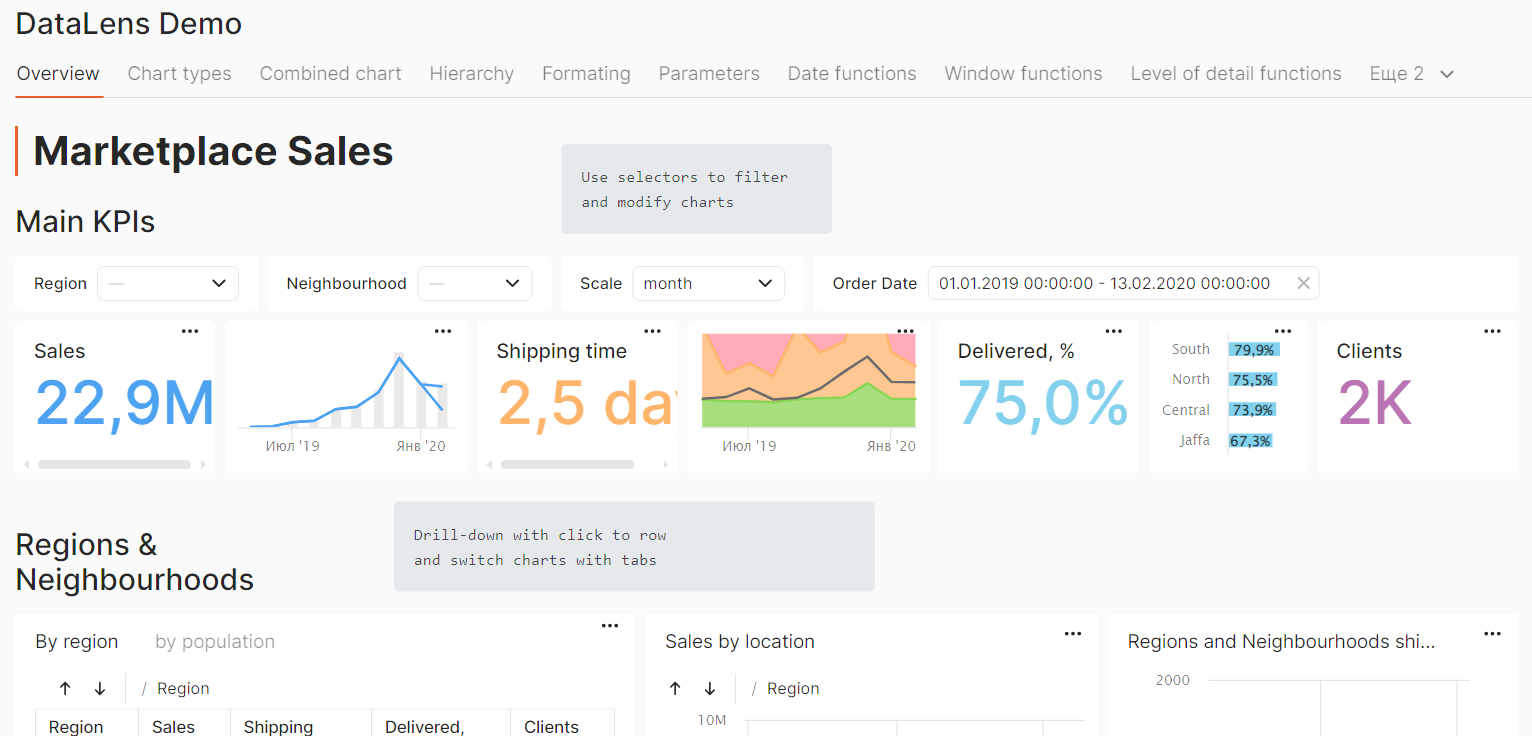


Рисунок 3. Демонстрационный дашборд

Перед выполнением данной практической работы настоятельно рекомендуется ознакомиться с возможностями DataLens через демонстрационный дашборд.

## Создание дашборда и чартов

По кнопке «Создать» мы можем создать коллекцию воркбуков и сам воркбук. Воркбук представляет собой рабочую область с подключениями к данным, наборами данных из подключений, чартами и дашбордами.

Создадим воркбук и назовем его «test\_workbook».

После создания воркбука нам доступны вкладки создания подключений, датасетов, чартов, дашбордов и списка всех объектов внутри воркбука (Рисунок 2).

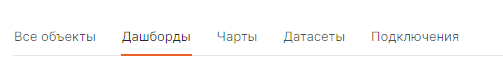


Рисунок 4. Вкладки объектов воркбука

Пайплайн работы с воркбуками состоит в следующем:

1. Настройка подключения. Необходимо соединиться с хранилищем данных, содержащим данные для визуализации. Это могут быть как оперативные источники данных, так и долговременные хранилища в зависимости от задач визуализации. В DataLens доступно только 4 драйвера подключения для следующих систем хранения данных: YDB, ClickHouse, PostgreSQL, YTsaurus CHYT (Рисунок 3).

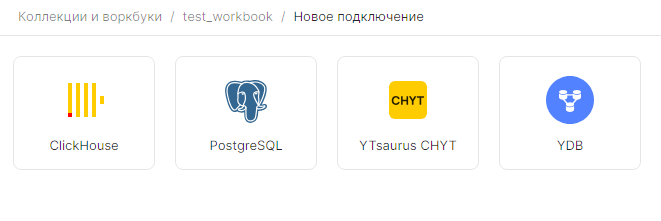


Рисунок 5. Доступные подключения

В облачной версии приложения DataLens доступно куда большее количество источников данных.

2. Настройка датасетов. После настройки подключения с хранилищем необходимо выбрать таблицы или данные, которые мы хотим визуализировать в формате структурированных данных. На данном этапе происходят фильтрация данных, вычисление столбцов, объединение данных нескольких таблиц и т.д. На основе датасетов в дальнейшем происходит сама визуализация данных.

3. Настройка чартов. Визуализация отдельных графиков для итогового дашборда. Доступен широкий, но неполный перечень графиков, среди которых: метрики, круговые диаграммы, столбчатые, линейные, точечные графики и карты.

4. Настройка дашборда с динамической фильтрацией записей для динамического изменения срезов и графиков на панели.

Для нашего нового созданного воркбука настроим подключение к PostgreSQL к схеме данных super\_market\_pg, которая содержит таблицы и записи операционных данных чеков покупок в магазине (Рисунок 4).

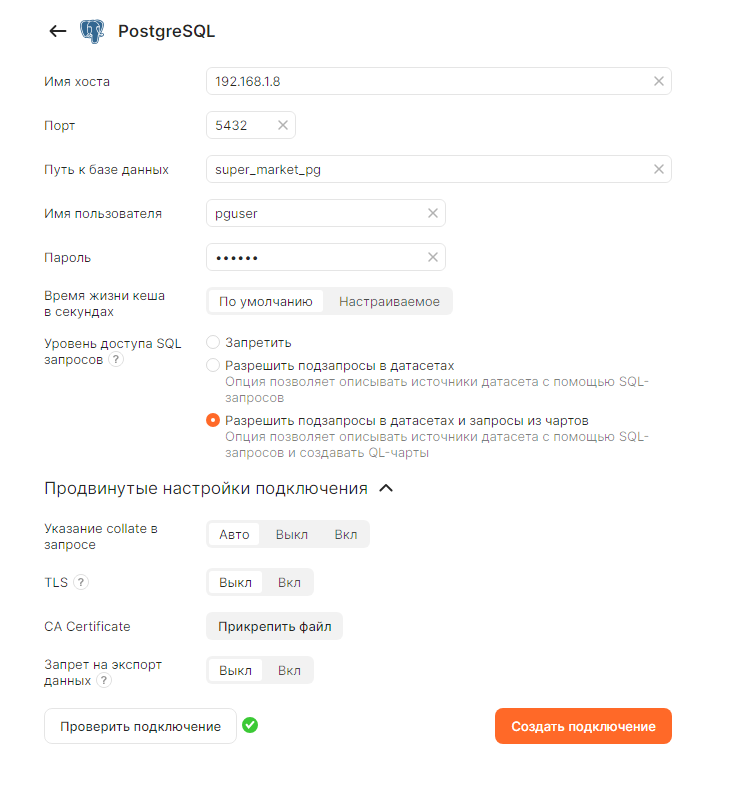


Рисунок 6. Настройка тестового подключения к БД транзакций супермаркета

Далее перейдем к стадии создания датасета на основе полученного подключения. Нажав на кнопку «Создать датасет» создадим набор данных для построения визуализации на основе настроенного подключения.

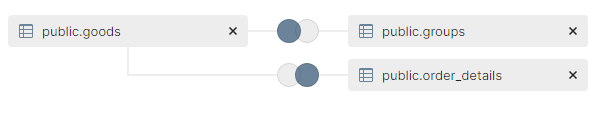


Рисунок 7. Соединение таблиц базы данных

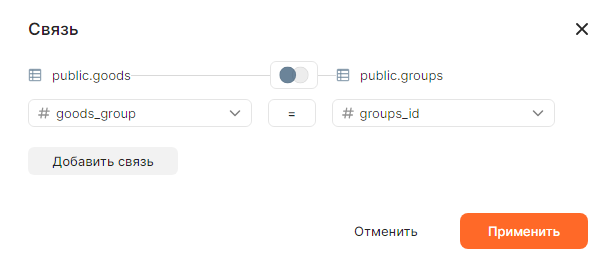


Рисунок 8. Настройка соединения таблиц goods и groups

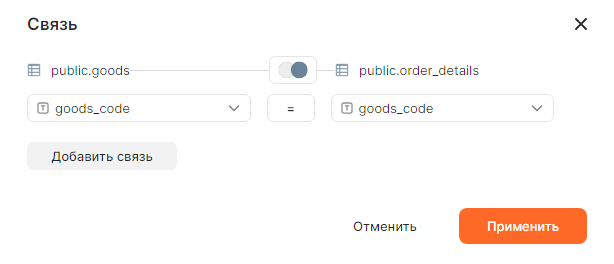


Рисунок 9. Настройка соединения таблиц order\_details и goods

Во вкладке «Поля» настроим поля набора данных, удалив лишние поля goods\_group и groups\_id.

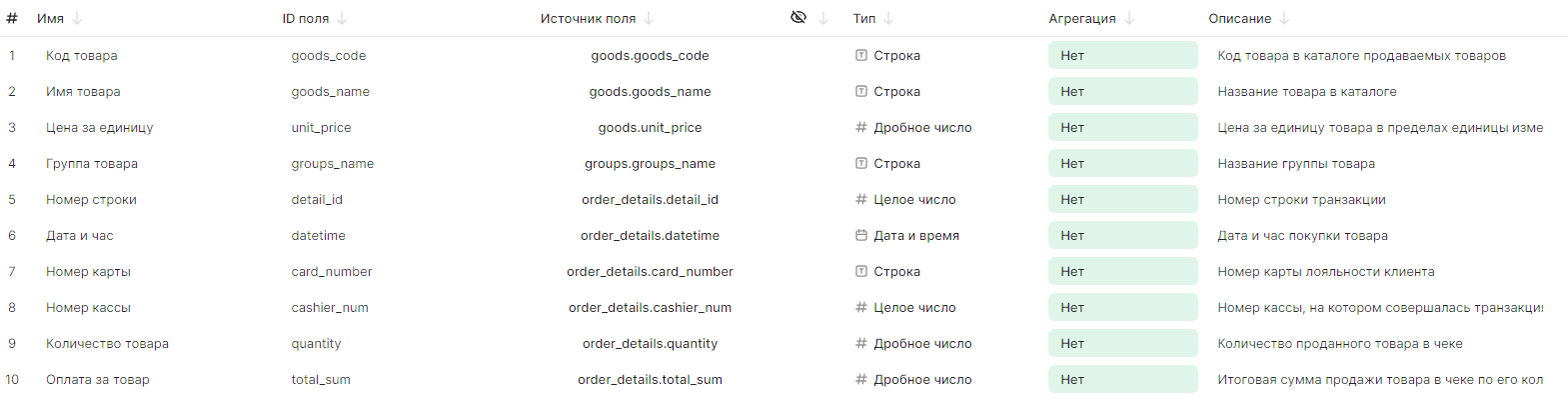


Рисунок 10. Настройка полей, удаление лишних столбцов.

Назовем данный датасет «Детализация чеков». С его помощью далее будем строить чарты.

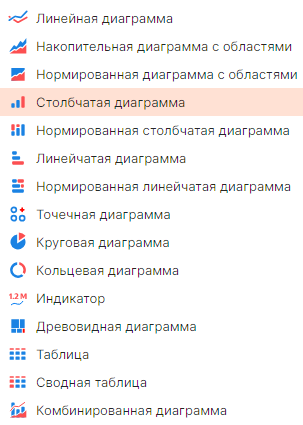


Рисунок 11. Список доступных чартов для построения

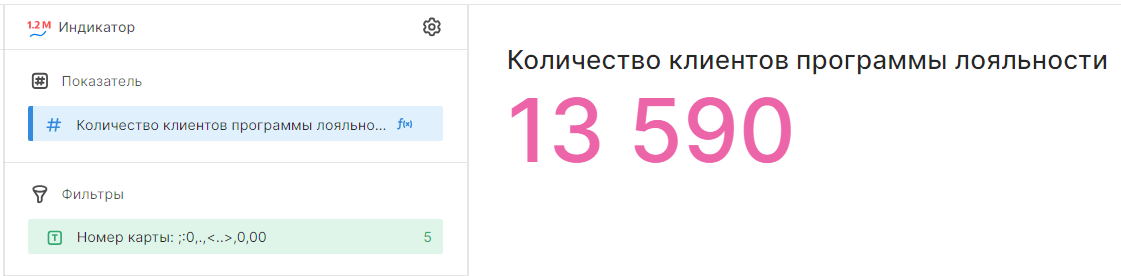


Рисунок 12. Индикатор по целевому признаку

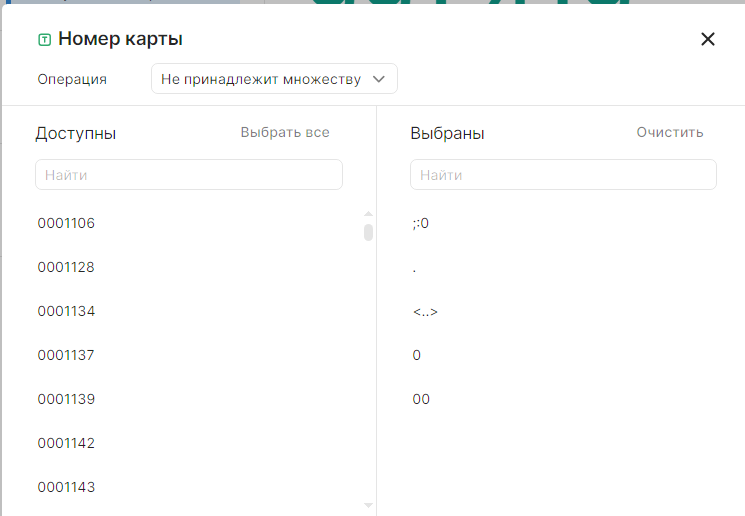


Рисунок 13. Настройка фильтра для визуализации количества купленных товаров с картой программы лояльности

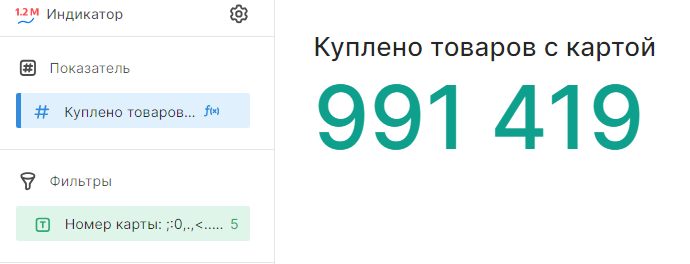


Рисунок 14. Индикатор количества купленных товаров с фильтром

Настроим визуализацию линейной диаграммы:

* X – дата продажи в отрыве от времени (вычисление через f(x) для поля Дата и час через выражение **DATE([Дата и час])**);
* Y – оплата за товар;
* цвета – номер кассы;
* фильтр – дата продажи.

Настройка фильтров по времени транзакций не обязательна в случае, если фильтр планируется быть на самом дашборде.

Результат настройки визуализации продемонстрирован на рисунке далее по тексту (Рисунок 13).

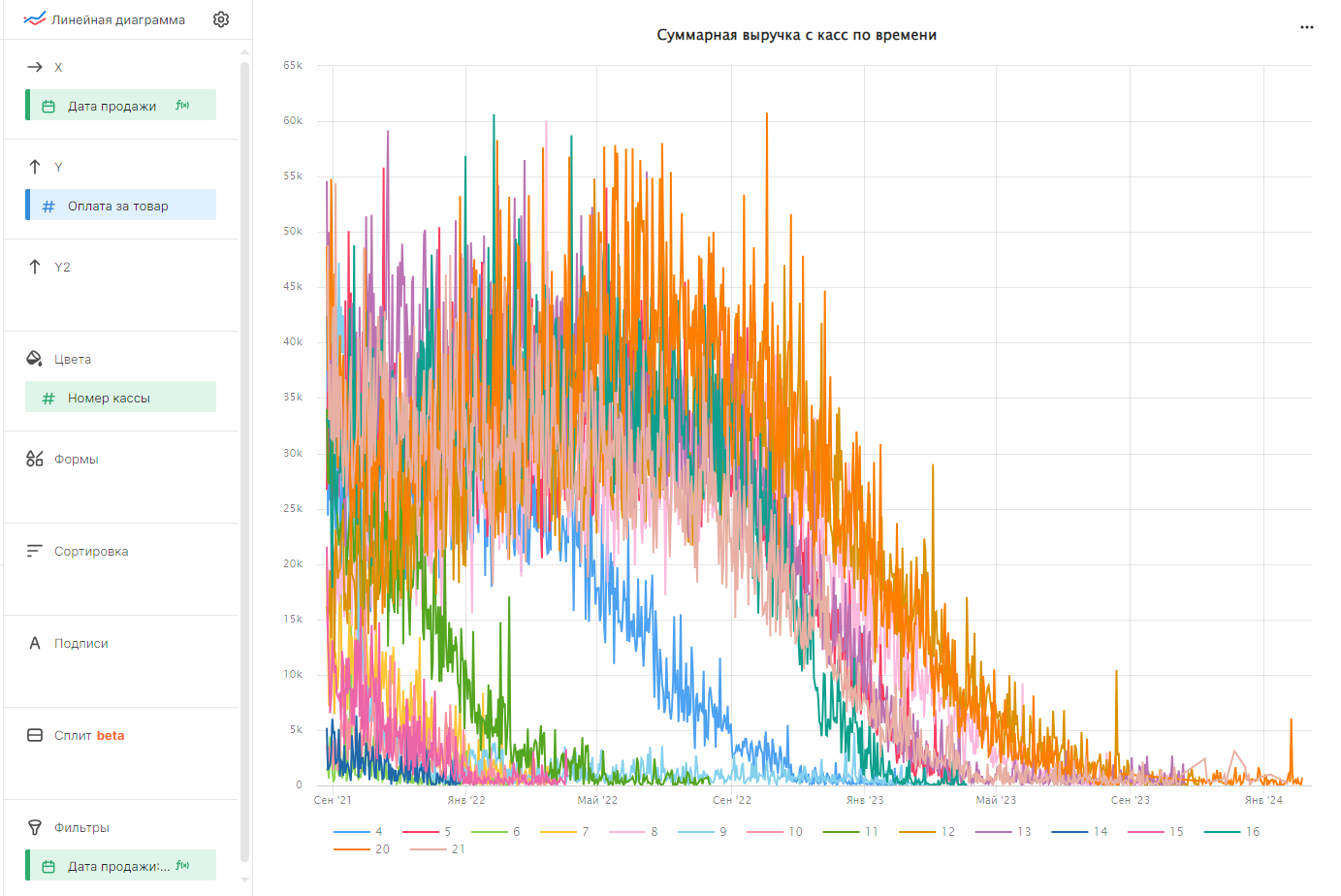


Рисунок 15. Суммарная выручка с касс по дням за весь период наблюдения

Настроим визуализацию сводной таблицы диаграммы:

* цвет – день недели (вычисление через f(x) для поля Дата и час через выражение **DAYOFWEEK([Дата и час])**);
* показатели – оплата за товар;
* сортировка – день недели (такое же вычисление);
* подписи – оплата за товар.

Настройка фильтров по времени транзакций не обязательна в случае, если фильтр планируется быть на самом дашборде.

Результат настройки визуализации продемонстрирован на рисунке далее по тексту (Рисунок 14).

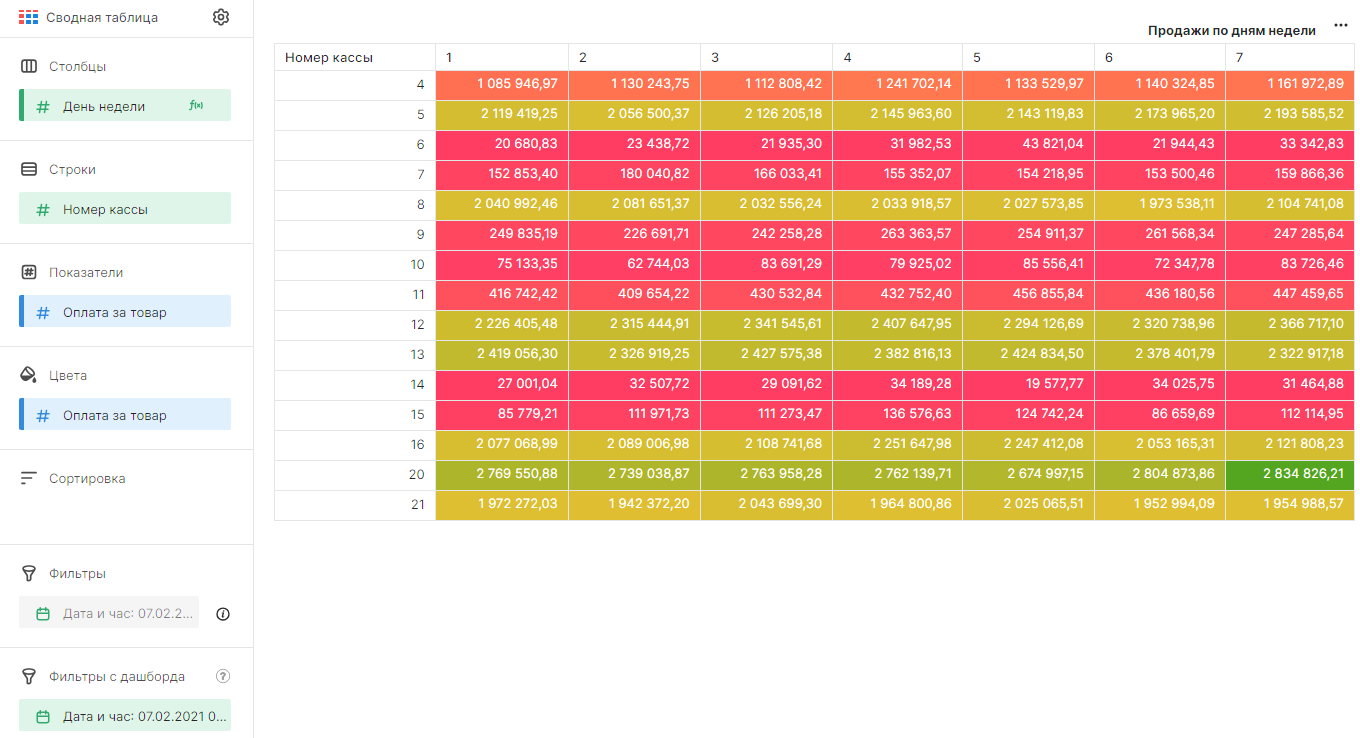


Рисунок 16. Настройка визуализации сводной таблицы

Настроим визуализацию круговой диаграммы:

* цвет – день недели (вычисление через f(x) для поля Дата и час через выражение **DAYOFWEEK([Дата и час])**);
* показатели – оплата за товар;
* сортировка – день недели (такое же вычисление);
* подписи – оплата за товар.

Настройка фильтров по времени транзакций не обязательна в случае, если фильтр планируется быть на самом дашборде.

Результат настройки визуализации продемонстрирован на рисунке далее по тексту (Рисунок 15).

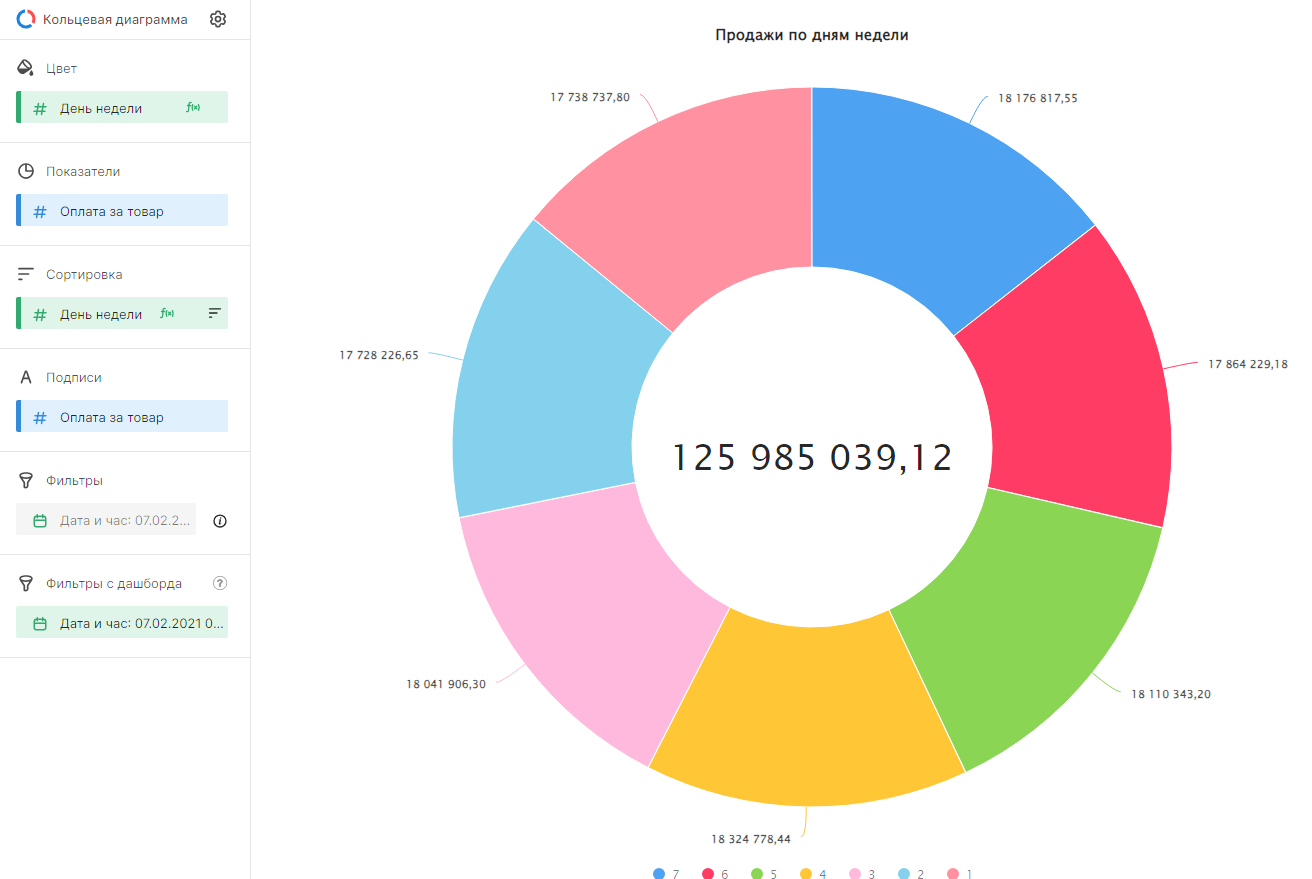


Рисунок 17. Настройка визуализации кольцевой диаграммы

Настроим визуализацию древовидной диаграммы:

* измерения – номер кассы;
* цвета – оплата за товар;
* размер – количество товара (цвет и размер именно в таком порядке);
* фильтры – дата и час (диапазон дат на выбор).

Результат настройки визуализации продемонстрирован на рисунке далее по тексту (Рисунок 16). Подписи на столбцах диаграммы на самом деле достаточно мелкие, их детальная настройка в DataLens невозможна.

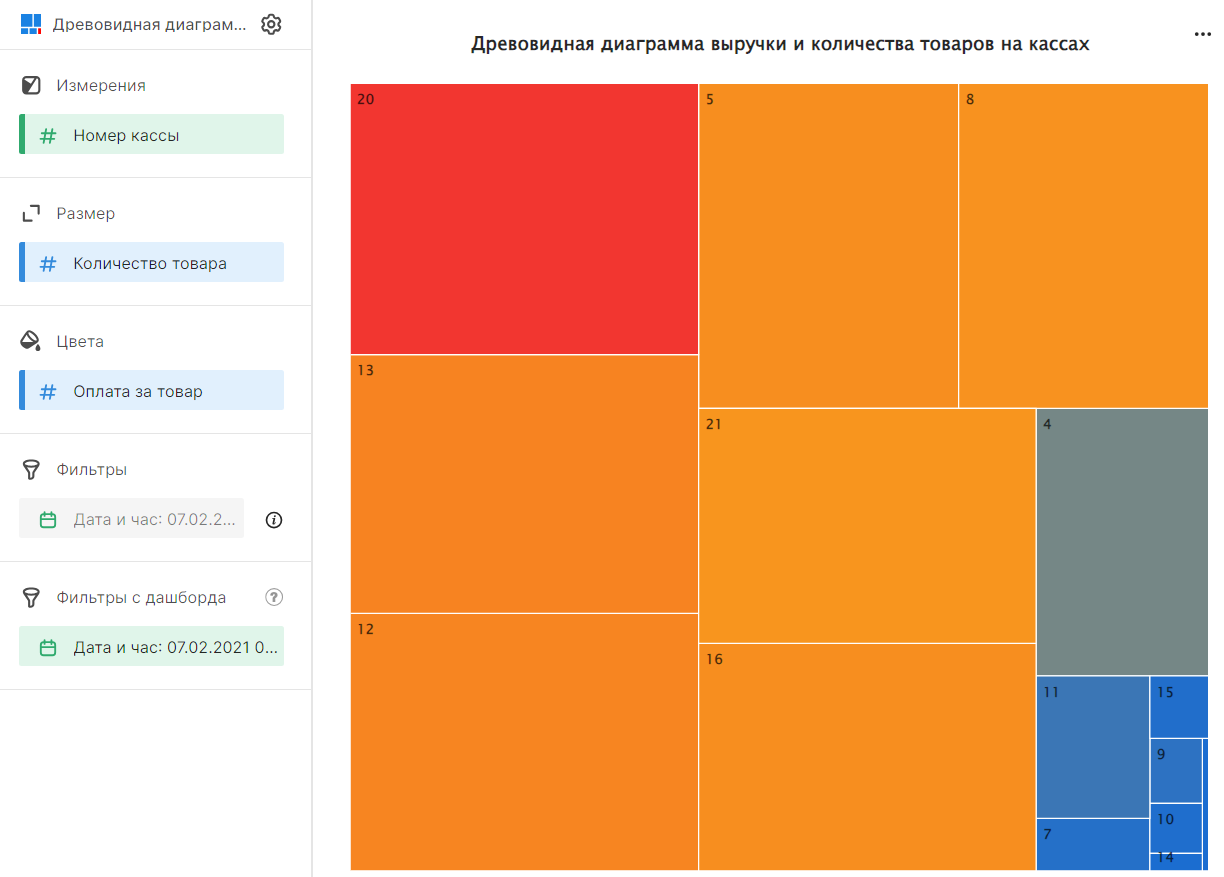


Рисунок 18. Настройка визуализации древовидной диаграммы

Составив несколько графиков, перейдем к созданию дашборда. На дашборд можно добавлять как ранее созданные отдельные чарты, так и селекторы для фильтрации значений по выбранному столбцу выбранного набора данных, так и поясняющий текст для пояснения визуализации или блока визуализаций чартами.

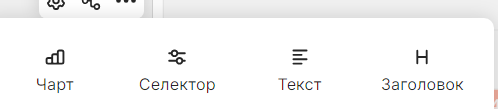


Рисунок 19. Выбор функциональных частей дашборда

Добавим на дашборд селектор по диапазону даты и времени, список, исключающий в попадание во множество для номера кассы и для группы товара. Данные селекторы расположим на одном уровне дашборда на самом его верху.

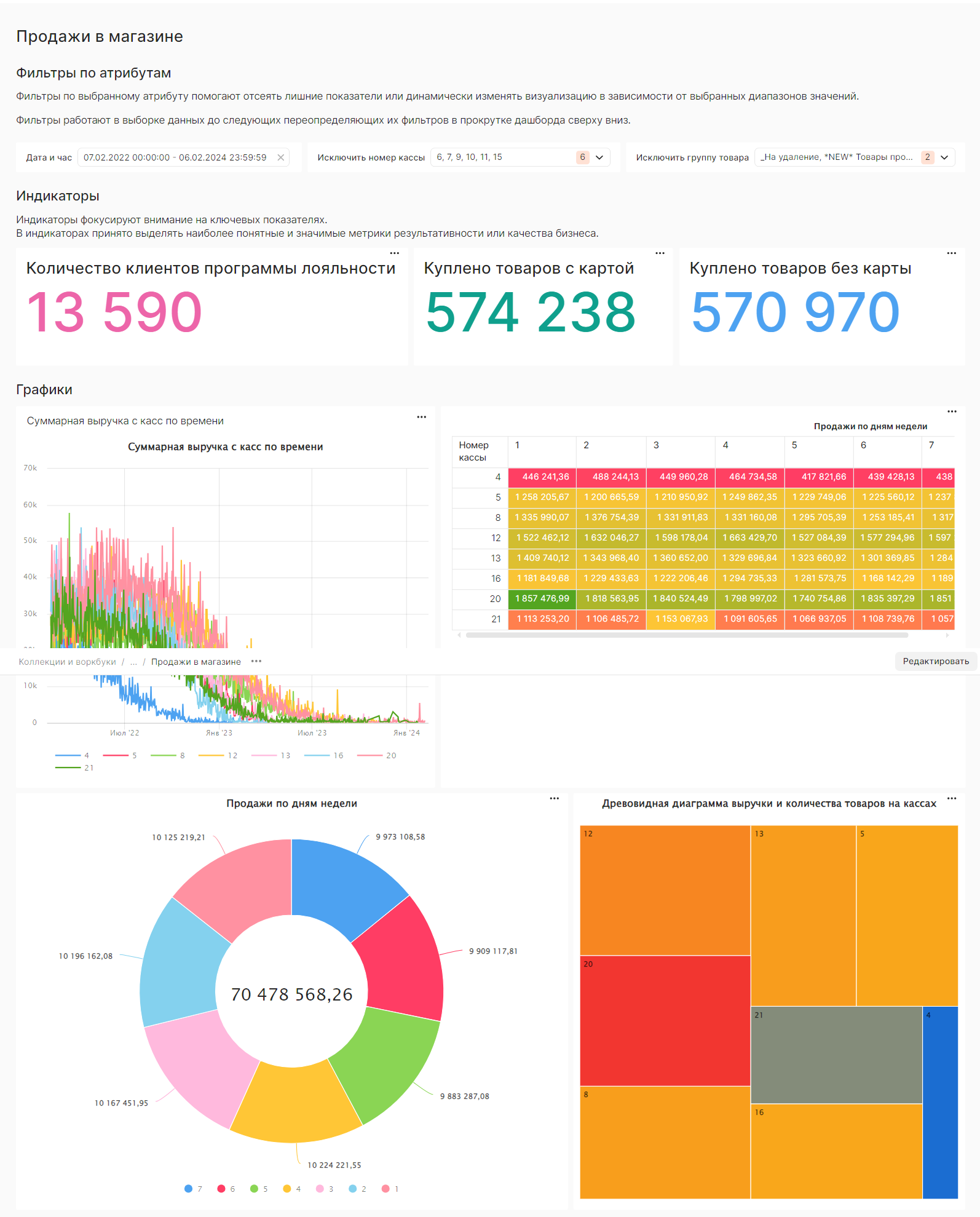


Рисунок 20. Селекторы и текст для обозначения функциональных единиц дашборда

После селекторов добавим главные метрики в формате индикаторов, которые сразу бросаются в глаза при первом просмотре визуализации.

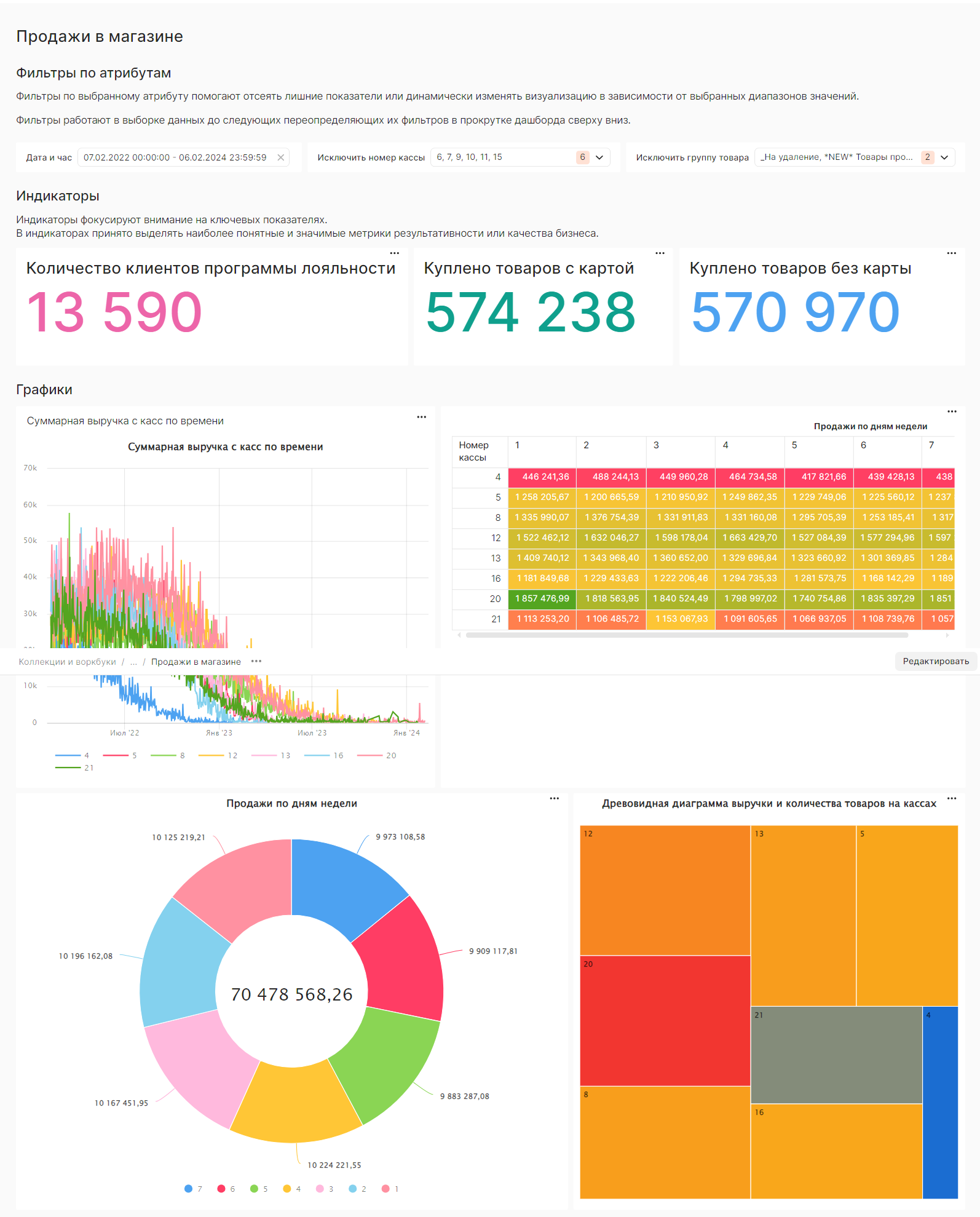


Рисунок 21. Строка индикаторов на дашборде

Отдельно ниже под основными метриками и селекторами размещаем основные визуализации набора данных. Графики желательно размещать так, чтобы графики были полностью видны и реализованы, а также сетка визуализаций была без пропусков.

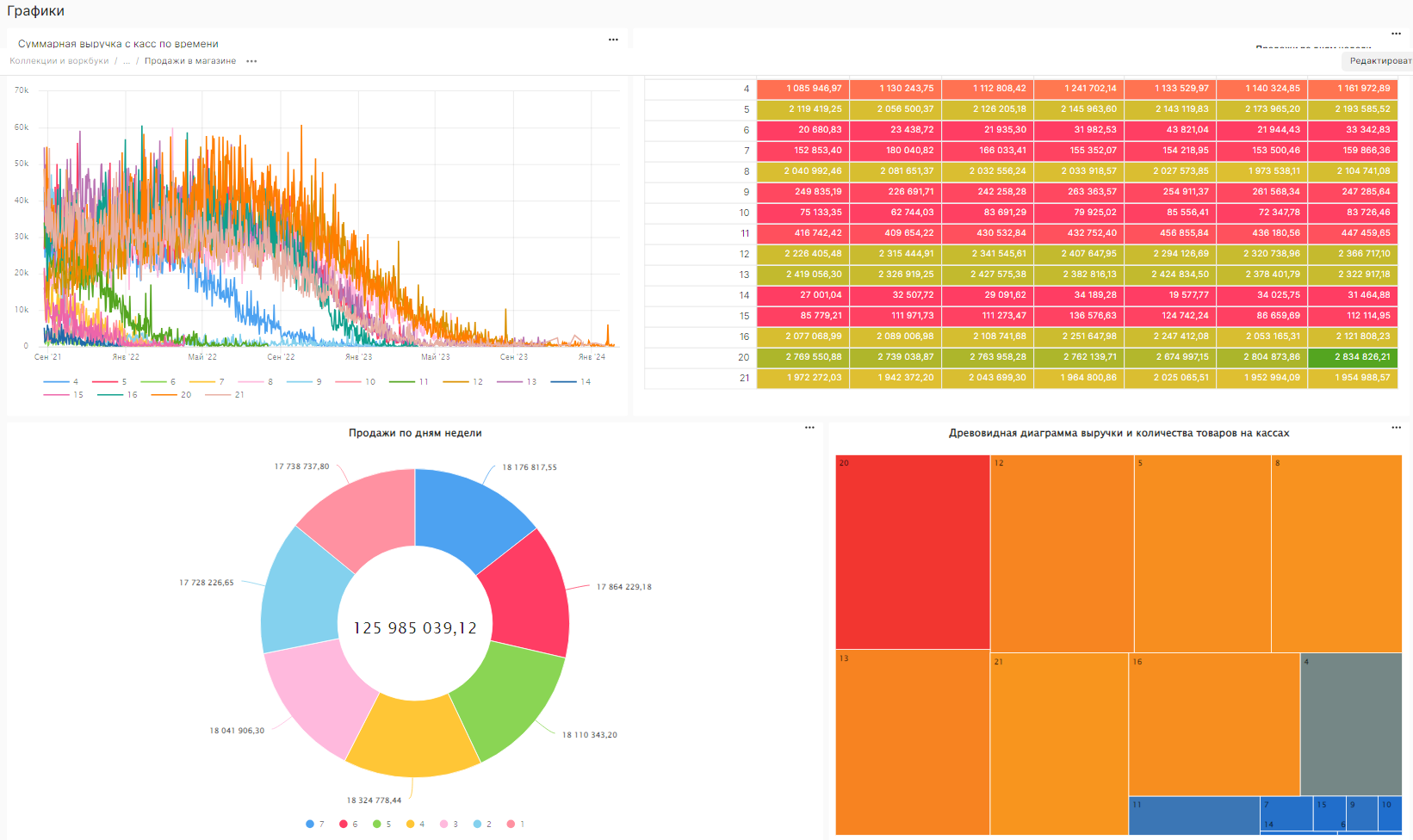


Рисунок 22. Матрица чартов как основная часть визуализации

Итоговый дашборд выглядит следующим образом:

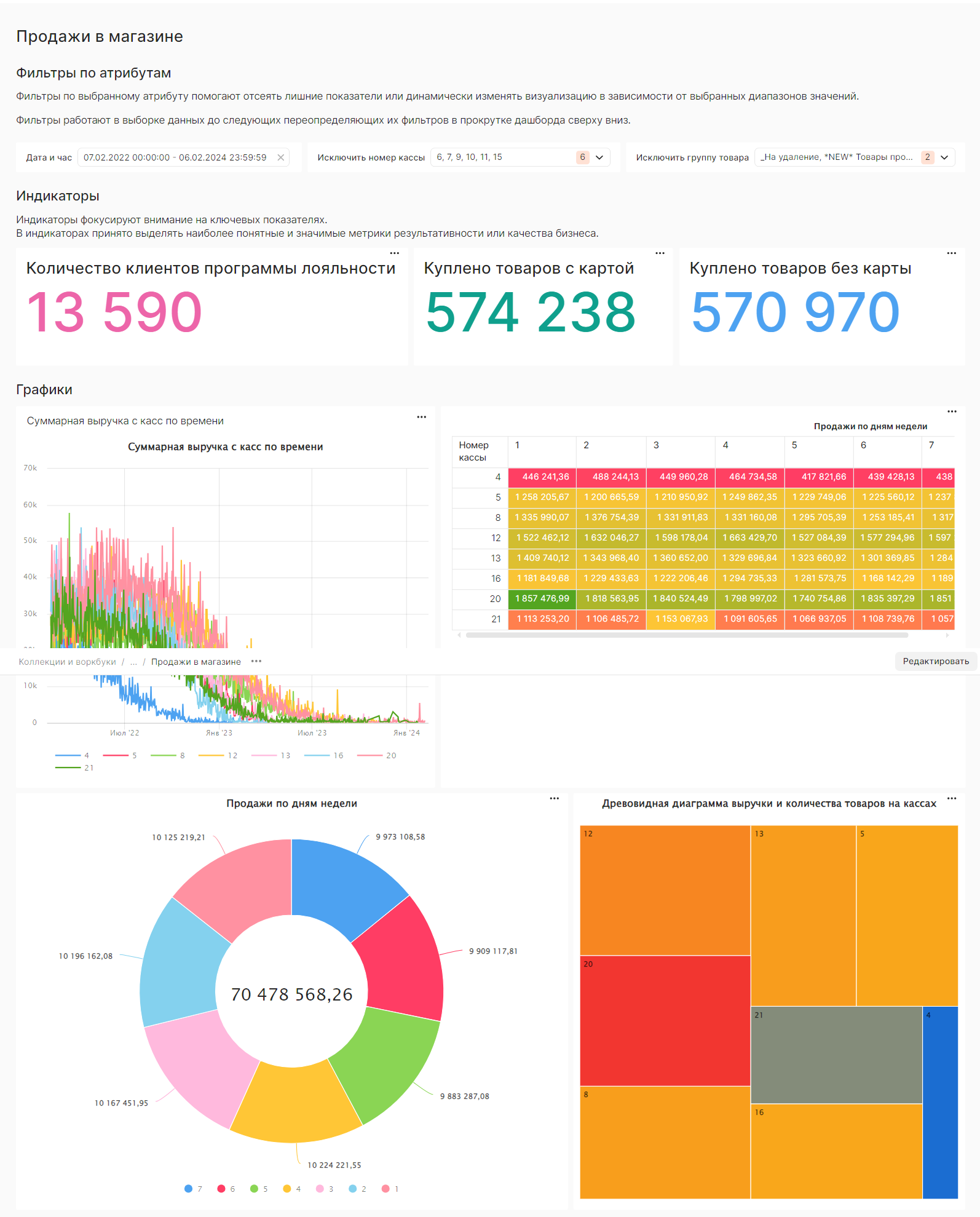


Рисунок 23. Итоговый дашборд на наборе данных продаж в магазин

## Самостоятельное задание

Предоставляемый для данной работе набор данных diy\_transactions.details находится в PostgreSQL на образе виртуальной машины. Схема набора данных выглядит следующим образом:

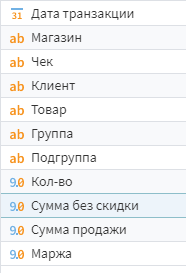


Рисунок 24. Схема набора данных diy\_transactions.details

1. Создайте подключение к базе данных diy\_transactions.
2. Создайте датасет на основе OLAP таблицы details**.**
3. Создайте чарт индикатора, содержащий общее число клиентов магазина **(1 балл).**
4. Создайте чарт индикатора, содержащего количество одновизитников (клиентов, которые покупали товары в магазине всего единожды) **(1 балл)**.
5. Создайте линейный чарт, содержащий график прибыли от времени **(1 балл)**.
6. Создайте столбчатую диаграмму количества чеков в зависимости от количества товаров в чеке **(2 балла)**.
7. Создайте ещё 4 различных видов чартов для заданного набора данных (**1 балл** за каждую визуализацию), с помощью текста к графикам объяснить смысл построенной визуализации.
8. Построить визуализацию в виде дашборда на основе графиков, построенных выше. Дашборд не должен иметь пустот и неровного размещения чартов, нечитаемых подписей или «зажевывания» части графика **(1 балл)**.

В отчёт по работе включить чарты и настройки чартов, а также итоговый дашборд.