

Разведочный анализ данных в Graphext

Бородулин Андрей Вадимович

Приамурский государственный университет им. Шолом-Алейхема

Студент

Аннотация

С каждым днем количество данных и информации растет, обрабатывать эту информацию становится все сложнее, и для решения этой проблемы существует различные сервисы анализа данных, одним из которых является Graphext. Цель исследования – изучение функциональных возможностей сервиса, его интерфейса и области применения. В результате был проведен анализ большого массива данных. Сервис доказал свою эффективность и удобство использования.

Ключевые слова: Данные, таблицы, статистика, графики, анализ.

Exploratory data analysis in Graphext

Borodulin Andrei Vadimovich

Sholom-Aleichem Priamursky State University

Student

Abstract

Every day the amount of data and information grows, it becomes more and more difficult to process this information, and to solve this problem there are various data analysis services, one of which is Graphext. The purpose of the study is to study the functionality of the service, its interface and scope of application. As a result, a large array of data was analyzed. The service has proven its effectiveness and ease of use.

Key words: Data, tables, statistics, graphs, analysis.

1. Введение

1.1 Актуальность исследования

С каждым днем люди создают все больше сервисов для упрощения обработки информации, так как поток её становится все больше. Среди всего выбора сервисов для обработки данных сложно найти тот, что полностью удовлетворит все запросы пользователя, но если необходимо получить данные по уже готовой таблице данных или узнать статистику, то для этого подойдет сервис Graphext. Спектр его возможностей позволяет анализировать большие массивы данных не задействуя ресурсы компьютера.

1.2 Обзор исследований

Graphext - это платформа, которая служит для создания, анализа и контекстуализации информации из различных цифровых "каналов". Его

аналитика делает его инструментом, который очень помогает в поиске релевантных данных из этих источников, пишет J.P. Rey [2]. В статье J. Kubalik представлен новый подход к решению задач оптимизации сетевого потока. Целью является оптимизация трафика в сетевых структурированных системах, управляемых событиями [3]. В исследовании N. T Top излагаются стратегии использования инструмента анализа больших данных Graphext с использованием материалов, собранных из источника базы данных Scopus. Автор обнаруживает тенденцию роста количества научных работ опубликованных жителями Вьетнама [4].

1.3 Цель исследования

Целью исследования является изучить сервис Graphext, интерфейс, функциональных возможностей.

2. Материалы и методы

Анализ системы анализа данных Graphext с использованием данных о продажах определенного набора товаров в различных регионах России URL: https://docs.google.com/spreadsheets/d/1oTPVDMav_HBvLMGuSMYmiiGxPKNqB0Tr/edit?usp=share_link&oid=100275693769304001469&rtpof=true&sd=true

3. Результат и обсуждение

В данной статье было решено затронуть сервис по анализу данных Graphext, который позволяет произвести анализ и прогнозное моделирование данных не используя каких либо формул и самое главное без кода. Данный сервис предлагает простой и интуитивно понятный интерфейс. Для начала работы на официальном сайте понадобится пройти регистрацию (рис.1).

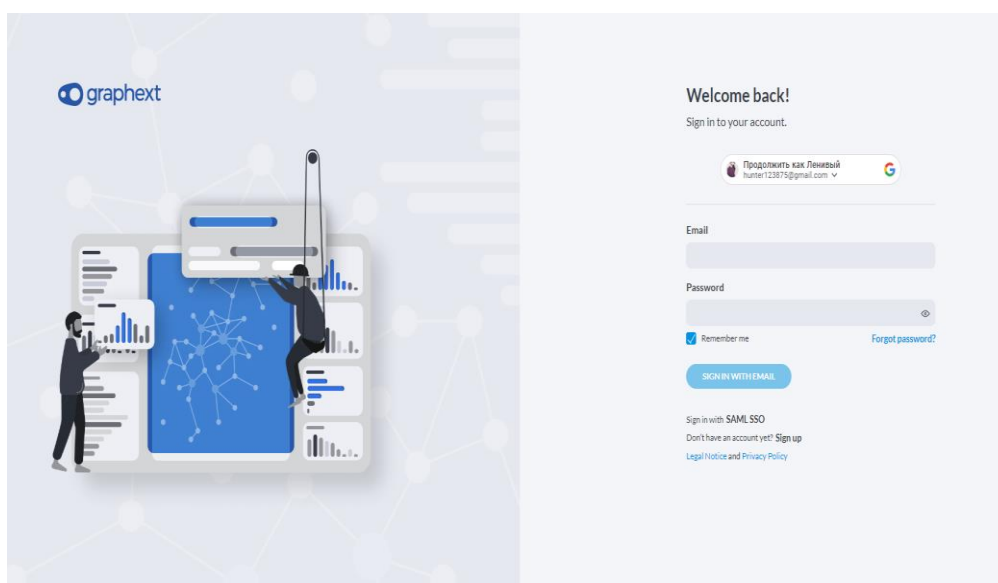


Рис. 1. Регистрация

Далее перед пользователем открывается список уже готовых графиков в качестве примера для изучения функционала Graphext (рис. 2).

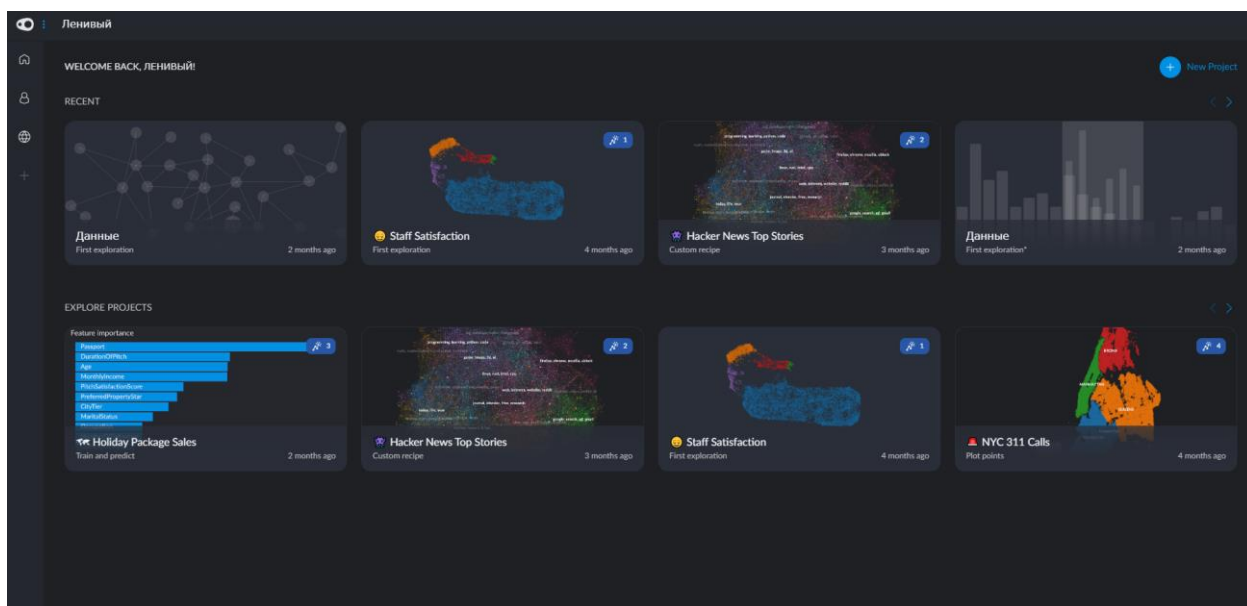


Рис. 2. Начало работы

Для дальнейшей работы с данными потребуется перенести данные в Graphext. Данный сервис может обрабатывать данные из различных баз данных (рис.3).

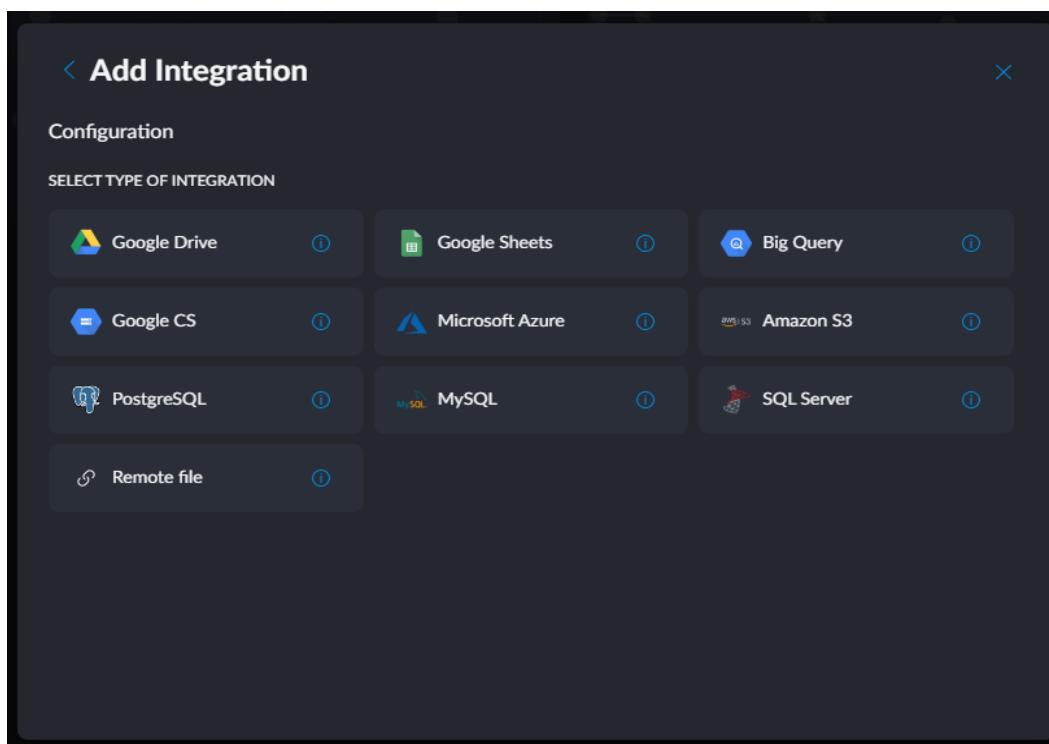


Рис. 3. Импорт баз данных

После загрузки данных в Graphext автоматически определяет типы данных и делит переменные на количественные, категориальные и текстовые. Graphext делает это практически без ошибок, но если надо подкорректировать то это можно сделать вручную (рис. 4).

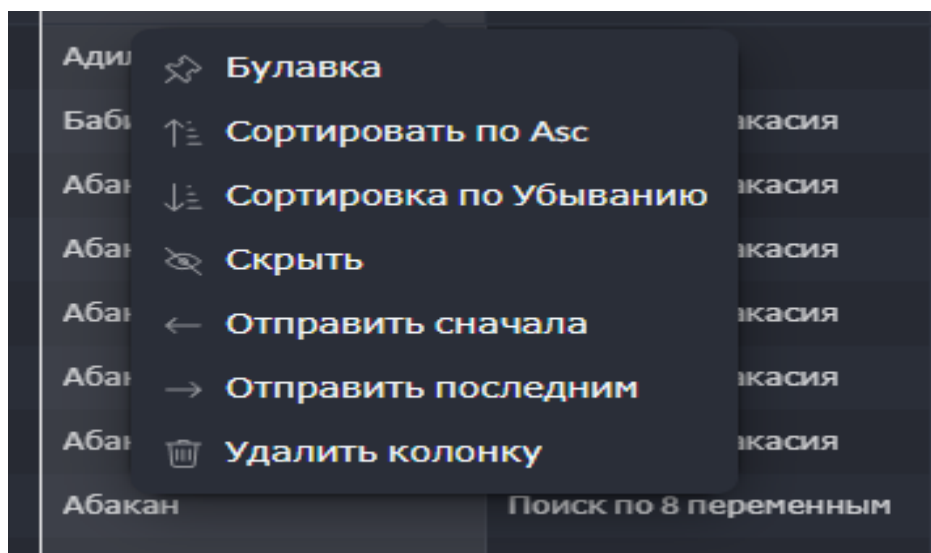


Рис. 4

При экспорте могут возникнуть трудности с Русской локализацией (рис. 5).

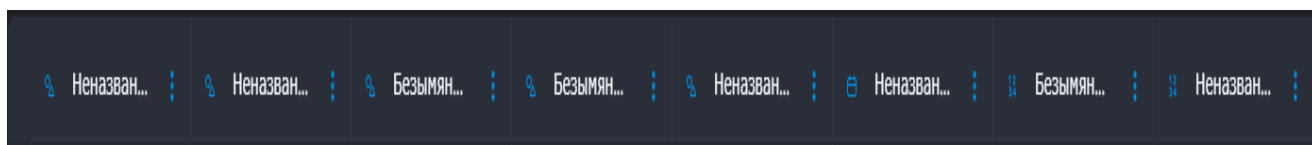


Рис. 5. Корректирование данных-1

Как видно на рис. 5-1 не подхватились названия столбцов и понадобится вручную их подкорректировать до исходного состояния как в Excel (рис. 6).

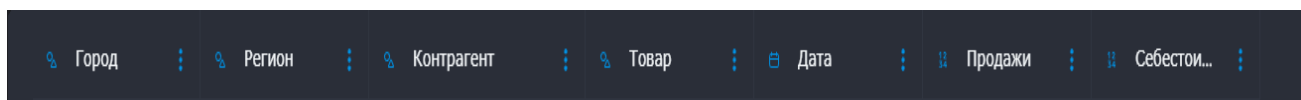


Рис. 6. Корректирование данных-2

Далее рассмотрим функции Graphext и начнем со сравнения переменных. Для этого перейдем в раздел Compare расположенный в верхней панели (рис. 7).

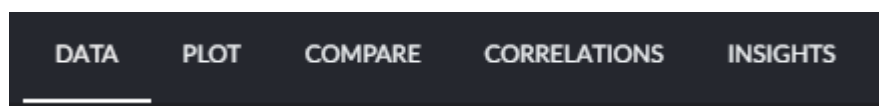


Рис. 7. Верхняя панель

Далее выбираем переменную для дальнейшего изучения функционала (рис. 8).

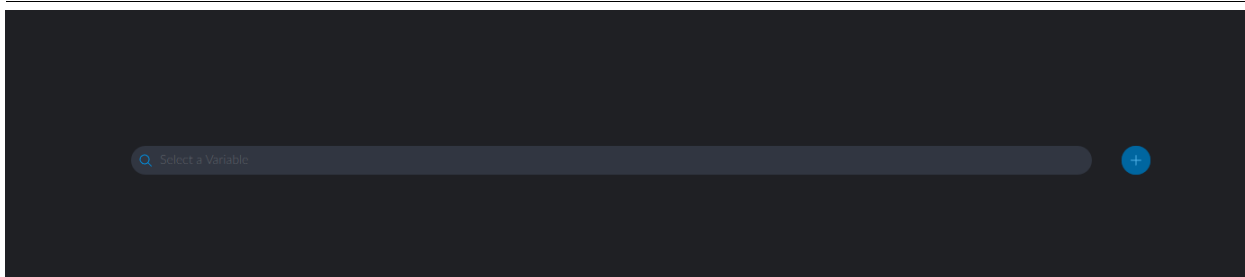


Рис. 8

Выбираем переменную «Город», после чего Graphext предоставляет уже готовые по данной переменной графики(рис. 9).

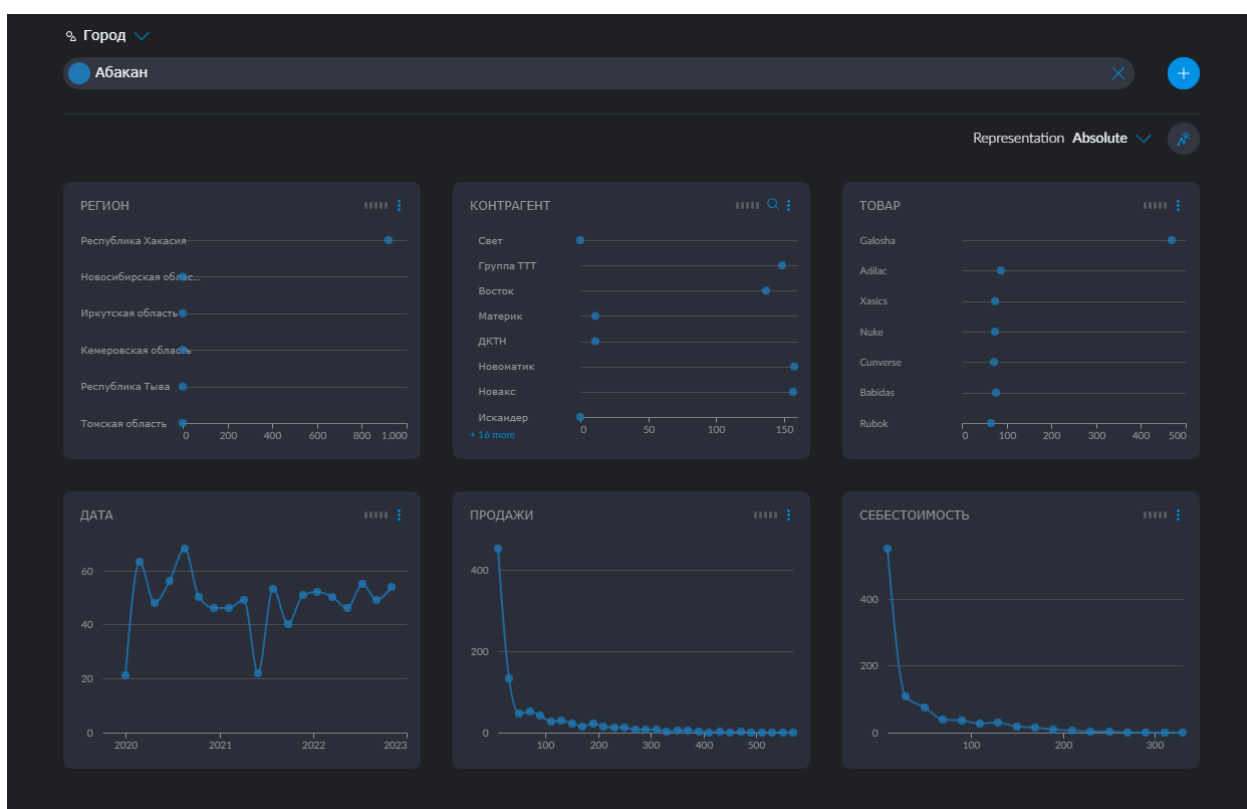


Рис. 9. Начальный анализ

На рисунке 9 показаны графики различных видов, таких как: диаграмма рассеяния, столбчатая диаграмма и линейная диаграмма. Далее рассмотрим следующую функцию Graphext и выберем в верхней панели функцию «Plot» (рис. 7). Graphext предлагает выбрать два значения для проведения анализа с помощью графиков и для этого было выбрано два значения «Регион» и «Продажи» (рис. 10).

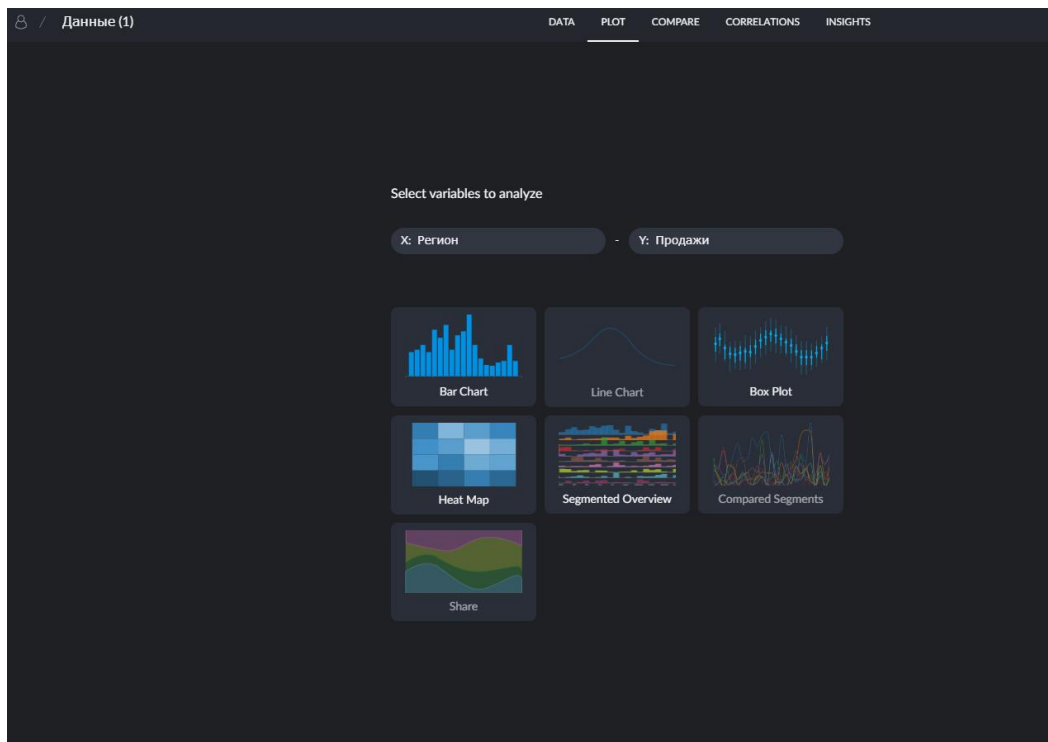


Рис. 10. Сравнительный анализ

При выборе этих переменных Graphext предлагает графики, которые может предоставить исходя из данных в этих переменных (рис. 11).



Рис. 11. Виды графиков

Для дальнейшего изучения была выбрана столбчатая диаграмма, на которой Graphext показал соотношение регионов по продажам (рис. 12).

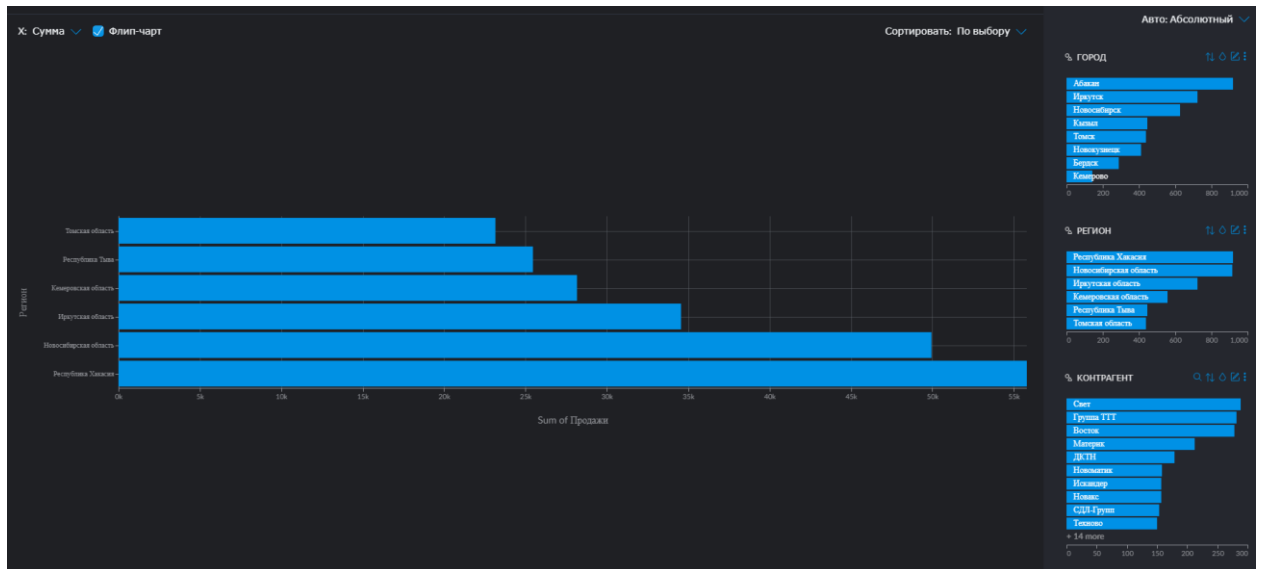


Рис. 12. Анализ данных

На рис. 12 показан график соотношения продаж и регионов, на котором видно сколько произведено продаж в каждом регионе и где сразу видно лидера продаж среди всех регионов Республику Хакасия. Справа показаны аналогичные графики, отношению продаж к другим переменным. Которые так же можно настроить, перейдя в настройки (рис. 13).



Рис. 13. Дополнительные возможности

При нажатии высветится окно, в котором появятся некоторые возможности. Таки как: Управление переменной, сохранить, экспорт и три варианта для дальнейшей работы с графиками – это открыть в сюжете, сравнении или корреляции (рис. 14).

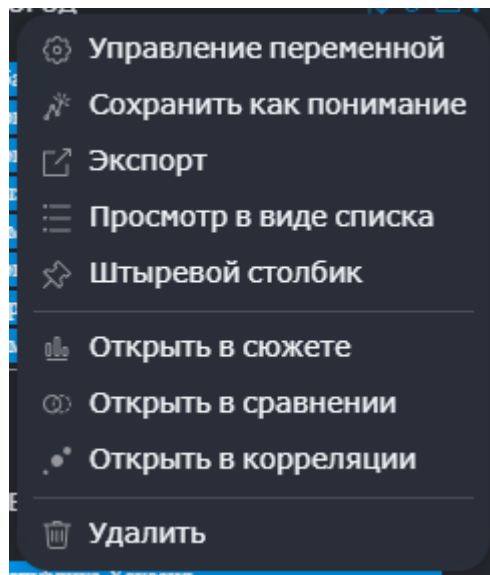


Рис. 14 Возможности обработки данных

Перейдя в раздел сравнения будет выбрано пять случайных переменных из первого столбца(рис. 15).

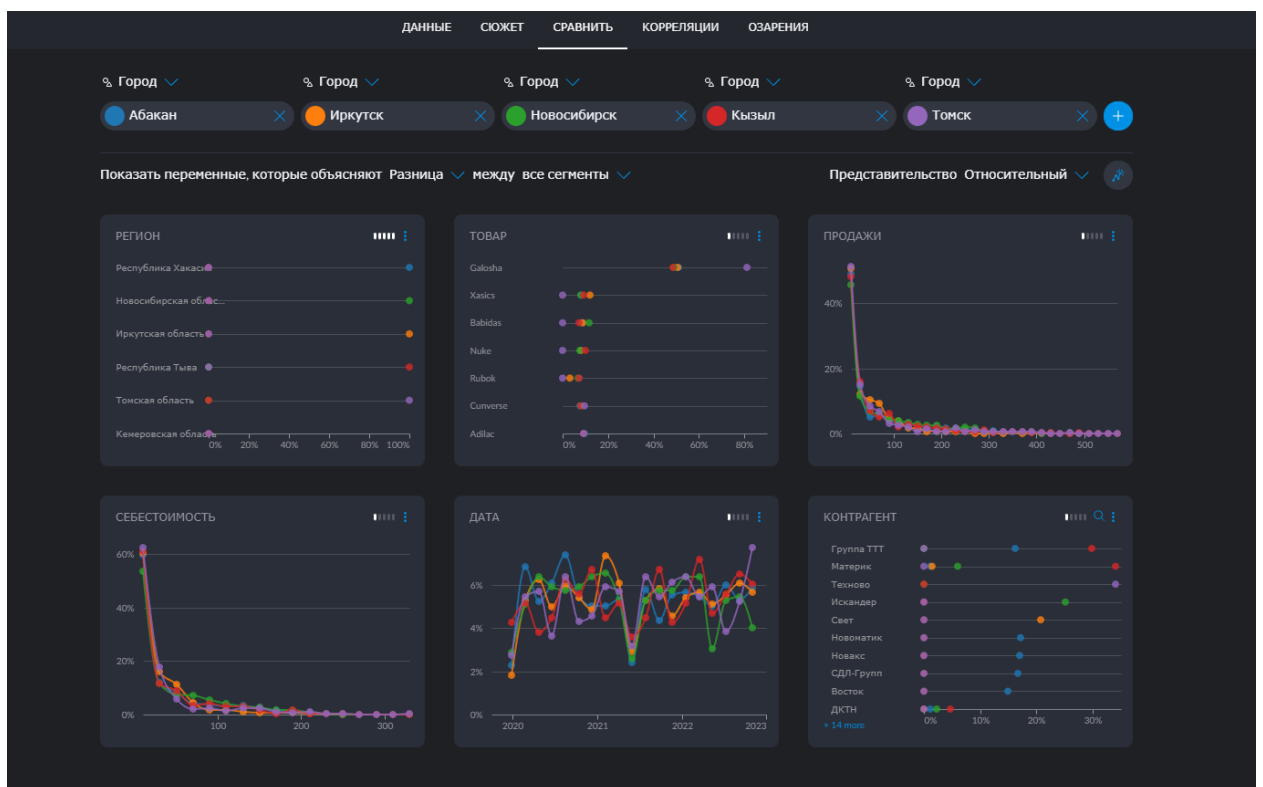


Рис. 15. Сравнительный анализ

Из данного списка графиков обратим внимание на линейный график под названием «Дата». Перейдя в параметры развернем для более подробного изучения (рис. 16).

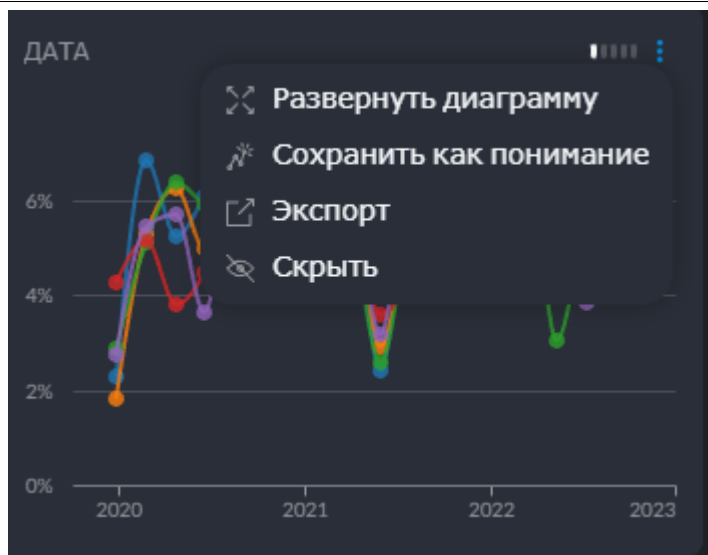


Рис. 16. Изучение графика

После проделанного действия можно увидеть какой город и в какое время лидировала либо же отставал по продажам (рис. 17).

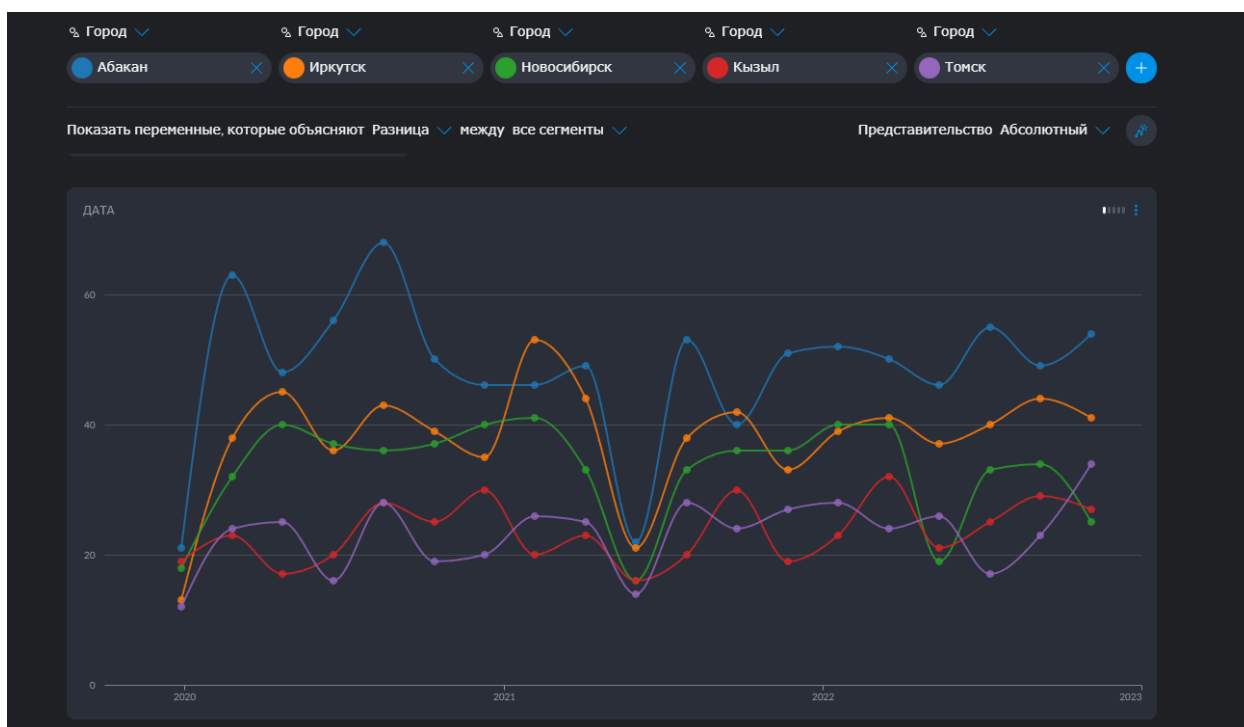


Рис. 17. График по датам

Если сможем сопоставить типы данных и имена переменных с проектами, которые выполняли ранее, Graphext автоматически предложит скрипт. Скрипт – это заранее определенный рабочий процесс уже содержащий инструкции по очистке, подготовки данных и применению алгоритмов по созданию проектов и анализа. Иногда велика вероятность что придется создать индивидуальный скрипт. В случае данных по продажам различных товаров в различных регионах России Graphext порекомендует несколько вариантов. Больше возможностей можно рассмотреть в платной версии

Graphext, но и в бесплатной версии достаточно вариантов для анализа данных для начинающих.

4. Выводы

Данная система отлично подходит для начинающих пользователей и среди аналогов предоставляет достойную конкуренцию. Единственным минусом является то что в данной системе нет возможности редактировать данные и добавлять новые переменные. Для использования данной системы вам не потребуются знания формул и языков программирования.

Библиографический список

1. Веб-платформа Graphext URL: <https://www.graphext.com>
2. Rey J. P., Izquierdo V. Starting up, Graphext: nunca más" lost in translation" en la gran conversación // Ctrl: control & estrategias. 2016. №. 632. С. 30-31.
3. Kubalik J., Mordinyi R. Optimizing events traffic in event-based systems by means of evolutionary algorithms // The Second International Conference on Availability, Reliability and Security (ARES'07). IEEE, 2007. №. 2. С. 1101-1107.
4. Ton N. T. H. Quality management issues in Vietnam: A systematic mapping analysis // VNUHCM Journal of Economics, Business and Law. 2022. №. 4. Т. 6. С. 857-872.