

Анализ и визуализация данных о фильмах с использованием платформы Yandex DataLens

Саухин Сергей Иванович

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

В статье рассматриваются возможности платформы Yandex DataLens как инструмента анализа и визуализации данных на примере набора данных о фильмах, включающего рейтинги Кинопоиска и IMDb. Описан поэтапный процесс работы с сервисом, включающий создание рабочего пространства, формирование датасета и построение аналитических визуализаций. В практической части реализованы различные типы графиков и индикаторов, позволяющие проанализировать распределение и обобщённые характеристики показателей. Полученные результаты подтверждают целесообразность использования Yandex DataLens для анализа тематических наборов данных и демонстрируют возможности платформы при решении учебных и исследовательских задач.

Ключевые слова: анализ данных, визуализация данных, Yandex DataLens, фильмы, чарты.

Analysis and visualization of film data using the Yandex DataLens platform

Saukhin Sergey Ivanovich

Sholom-Aleichem Priamursky State University

Student

Abstract

This article explores the capabilities of the Yandex DataLens platform as a tool for data analysis and visualization using a movie dataset that includes Kinopoisk and IMDb ratings. The study describes the step-by-step workflow with the service, including workspace creation, dataset preparation, and the construction of analytical visualizations. In the practical part, various types of charts and indicators are implemented to analyze the distribution and aggregated characteristics of the data. The results demonstrate the effectiveness of Yandex DataLens for analyzing thematic datasets and highlight its applicability in educational and research projects.

Keywords: data analysis, data visualization, Yandex DataLens, films, charts.

1 Введение

1.1 Актуальность

В современных условиях стремительного роста объёмов данных возрастает потребность в эффективных инструментах их анализа и

визуализации. Тематические наборы данных, в том числе данные о фильмах, широко применяются в учебных и исследовательских проектах для изучения методов обработки информации, выявления статистических закономерностей и формирования навыков аналитического мышления. Однако без специализированных средств визуального анализа такие данные трудно интерпретировать и использовать для получения обоснованных выводов.

Использование систем бизнес-аналитики позволяет упростить процесс анализа данных за счёт наглядного представления информации в виде диаграмм и интерактивных отчётов [1]. Визуализация данных способствует более быстрому восприятию результатов анализа, выявлению структурных особенностей и сравнительному анализу показателей, что подтверждается исследованиями в области аналитических платформ и дашбордов [2].

Современные BI-инструменты обеспечивают интеграцию различных источников данных, их предварительную подготовку и последующую визуализацию без необходимости разработки специализированных программных решений [3]. Такие платформы особенно актуальными для использования в образовательной и исследовательской среде, где важна доступность инструментов и простота их освоения.

Платформа Yandex DataLens представляет собой облачное решение для анализа и визуализации данных, позволяющее создавать аналитические отчёты и карты на основе подготовленных датасетов [4]. Возможность работы через веб-интерфейс, поддержка различных типов визуализации и интерактивных элементов определяют актуальность использования данного сервиса при анализе тематических наборов данных, включая данные о фильмах.

Таким образом, исследование возможностей платформы Yandex DataLens и её применение для визуализации данных о фильмах является актуальной задачей, имеющей практическую значимость для учебных и аналитических проектов [5].

1.2 Обзор исследований

В научных и учебных публикациях анализ данных и их визуализация рассматриваются как ключевые элементы современного аналитического процесса. Отмечается, что применение визуальных средств представления информации позволяет повысить наглядность результатов анализа и упростить интерпретацию сложных наборов данных [1]. Использование диаграмм и дашбордов способствует более эффективному выявлению закономерностей и сравнительному анализу показателей.

Ряд исследований посвящён практическому использованию BI-платформ для анализа данных в различных предметных областях. В работах подчёркивается, что современные аналитические сервисы обеспечивают полный цикл работы с данными, включая их подготовку, агрегацию и визуализацию, что позволяет снизить зависимость от программной разработки и упростить внедрение аналитических решений [2]. Особое внимание

уделяется облачным инструментам, обеспечивающим доступность аналитики и масштабируемость вычислительных ресурсов.

В научных публикациях также рассматриваются вопросы сравнительного анализа инструментов визуализации данных. Отмечается, что платформы класса BI отличаются по функциональным возможностям, однако ключевыми критериями эффективности являются удобство интерфейса, поддержка различных типов визуализации и возможности интерактивного анализа [3]. Результаты таких исследований подтверждают целесообразность использования облачных сервисов визуальной аналитики при работе с большими и тематическими наборами данных.

Отдельное направление исследований связано с применением систем визуализации данных для анализа тематических информационных массивов, включая данные в сфере культуры и медиа. Подчёркивается, что анализ данных о фильмах позволяет изучать структуру и распределение характеристик, таких как жанры, рейтинги и количество произведений, а визуализация способствует более глубокому пониманию предметной области [4].

Анализ существующих публикаций показывает, что платформа Yandex DataLens рассматривается как перспективный инструмент для визуального анализа данных, сочетающий функциональные возможности BI-систем с доступностью и простотой использования [5]. Исследования подтверждают эффективность применения данного сервиса в учебных и исследовательских проектах, а также обосновывают целесообразность его использования для визуализации тематических наборов данных.

1.3 Цель исследования

Целью данного исследования является анализ возможностей платформы Yandex DataLens и демонстрация процесса визуализации данных о фильмах с использованием инструментов аналитики и визуального представления информации.

Для достижения поставленной цели в работе сформулированы следующие задачи:

- изучить функциональные возможности платформы Yandex DataLens;
- выполнить настройку аналитического рабочего пространства;
- сформировать датасет на основе данных о фильмах;
- реализовать визуализацию данных с использованием столбчатых и круговых диаграмм;
- проанализировать полученные результаты и оценить их практическую значимость.

2 Материалы и методы

В ходе проведения исследования использовался комплекс теоретических и практических методов анализа данных и визуализации информации. Методологической основой работы послужили научные и учебные публикации, посвящённые вопросам бизнес-аналитики, построения

дашбордов и применения BI-систем для анализа данных [1–3]. В данных источниках рассматриваются принципы визуального представления информации, этапы аналитического процесса и требования к интерпретируемости результатов анализа.

В качестве теоретических материалов исследования также использовались публикации, посвящённые сравнительному анализу инструментов визуализации данных и обзору систем для создания аналитических отчётов [4, 5]. Эти работы позволили определить основные подходы к выбору типов диаграмм и методов визуального представления данных в зависимости от характера анализируемой информации.

Практическая часть исследования реализована с использованием облачной платформы Yandex DataLens, предназначенной для анализа и визуализации данных [6, 7]. В процессе работы применялись стандартные инструменты сервиса, включая создание рабочих пространств, формирование датасетов и построение визуальных представлений данных.

Методы исследования включали поэтапную настройку аналитического пространства, подготовку и структурирование данных о фильмах [8], а также визуальный анализ полученных результатов. Для представления данных использовались столбчатые и круговые диаграммы, выбранные с учётом рекомендаций по визуализации количественных и структурных показателей [2, 4].

Для анализа результатов применялись методы интерпретации визуальных данных, направленные на выявление распределения показателей и сравнительный анализ категорий. Использование визуальных аналитических инструментов позволило повысить наглядность результатов исследования и обеспечить удобство их интерпретации без применения специализированных программных средств.

Применение совокупности указанных материалов и методов позволило реализовать полный цикл визуального анализа данных о фильмах и сформировать обоснованные выводы о возможностях использования платформы Yandex DataLens в учебных и исследовательских проектах.

3 Результаты и обсуждение

В рамках практической части исследования выполнена последовательная настройка аналитического пространства в Yandex DataLens и построены визуальные отчёты на основе созданного датасета. Реализация включала: вход в сервис, создание workbook, подготовку датасета и разработку чартов с применением различных типов визуализации.

3.1 Доступ к сервису Yandex DataLens

На первом этапе выполнена работа с веб-интерфейсом платформы Yandex DataLens. Доступ к инструменту осуществляется через браузер, что позволяет использовать сервис без локальной установки программного обеспечения (рис. 1).

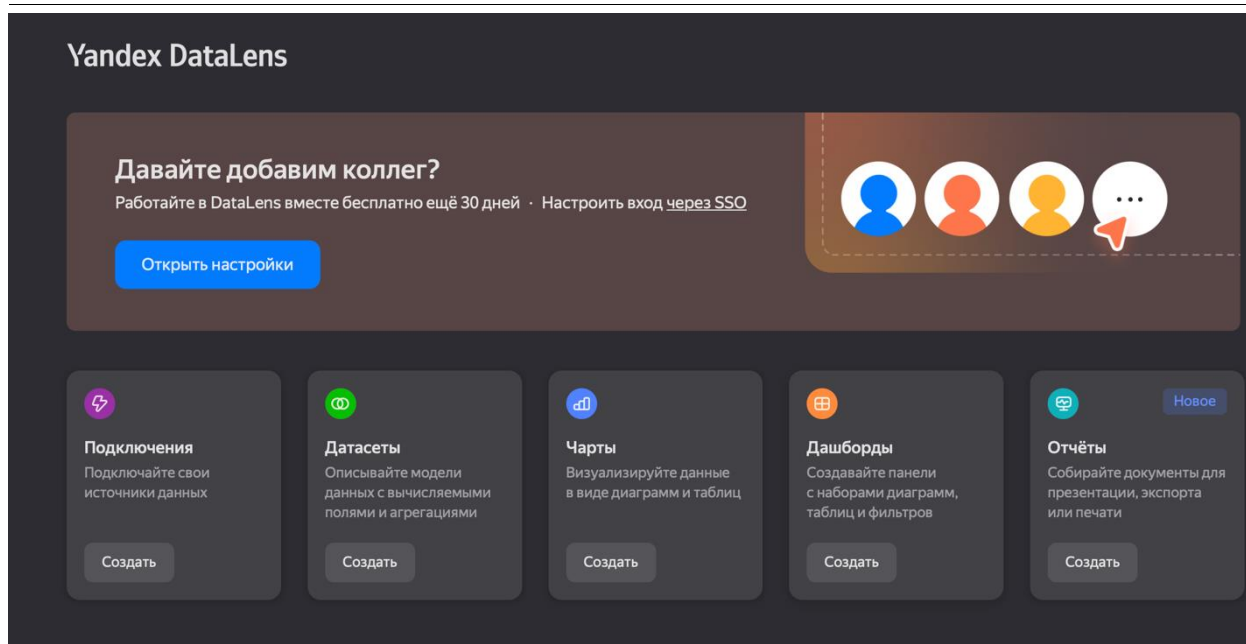


Рис. 1. Сайт Yandex Datalens

Облачный подход снижает требования к инфраструктуре и упрощает подключение пользователей, что особенно важно в корпоративной среде, где аналитические отчёты должны быть доступны различным подразделениям с минимальными затратами на развёртывание.

3.2 Создание рабочего пространства (workbook)

На втором этапе создан workbook, предназначенный для логического объединения аналитических объектов (датасетов, чартов и дашбордов). Workbook создан и назван WorkBookForFilms (рис. 2,3).

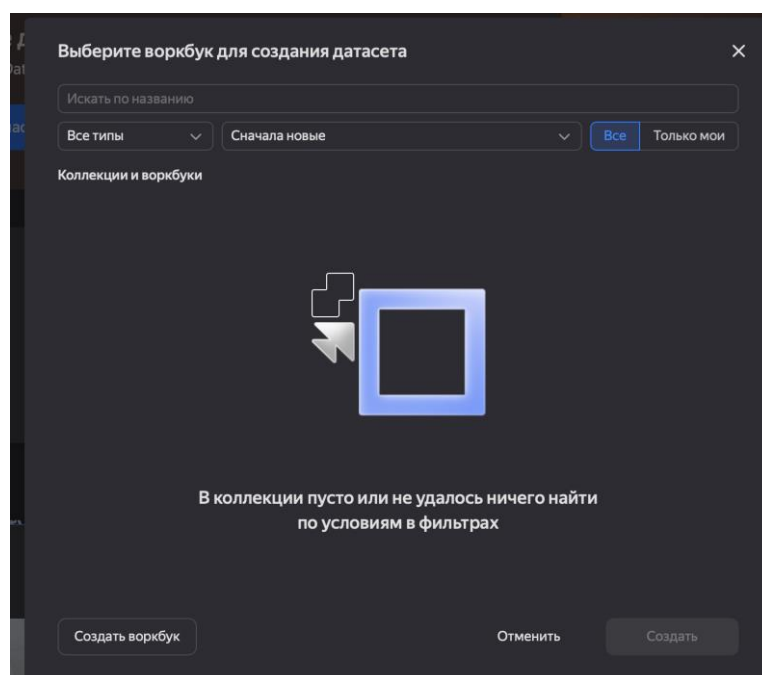


Рис. 2. Меню воркбуков

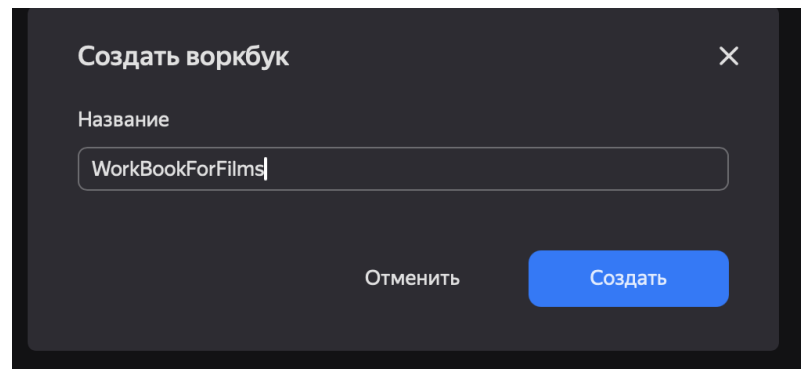


Рис. 3. Создание воркбука «WorkBookForFimls»

Использование workbook повышает структурированность проекта: аналитические сущности размещаются в едином пространстве, что облегчает сопровождение, повторное использование компонентов и командную работу. Кроме того, формирование workbook является базовым этапом подготовки аналитической системы, поскольку определяет контекст хранения и дальнейшей организации объектов.

3.3 Добавление и настройка датасета

На третьем этапе в workbook добавлен датасет, содержащий данные, необходимые для дальнейшего анализа. Выполняется выбор полей и определение показателей, которые использованы в визуализациях (рис. 4, 5).

Title	kinopoiskid	imdbid	Year	Rating Kinopoisk	Rating Imdb	Age Limit	Genres	Country
Реальная любовь	6144	tt0314331	2003.0	8.0	7.5	18+	драма, мелодрама, комедия	США, Фр
Заклятие	468994	tt1457767	2013.0	7.3	7.5	18+	детектив, ужасы	США
Чудо	722951	tt2543472	2017.0	8.2	7.9	12+	драма, семейный	США, Гон
Такси	14349	tt0152930	1998.0	8.0	7.0	18+	криминал, боевик, комедия	Франция
Диспетчер	5435407	tt11539996	2024.0	7.2	7.0	18+	триллер, боевик	США
Папины дочки. Новые	5249042	tt28751684	2023.0	8.3	5.2	12+	комедия	Россия
Детство Шелдона	1040419	tt6226232	2017.0	8.7	7.7	18+	комедия	США
Брестская крепость	436263	tt1343703	2010.0	8.1	7.4	18+	драма, боевик, военный, история	Россия, Б
Ковбой против пришельцев	81961	tt0409847	2011.0	5.9	6.0	12+	триллер, драма, фантастика, вестерн, боевик	США, Ин
Отец Мумбака: Противостояние	958519	tt5461944	2018.0	8.0	7.6	18+	триллер, драма, боевик, история	США, Авс
Далласский клуб покупателей	260162	tt0790636	2013.0	7.9	7.9	18+	драма, биография	США
Мир в огне	462921	tt1210027	2025.0		4.5	18+	фантастика, боевик, комедия	США
Ворон	12362	tt0109506	1994.0	7.9	7.5	18+	драма, криминал, боевик, фантази	США
Слово пацана. Кровь на асфальте	5304403	tt28429213	2023.0	8.1	7.6	18+	драма, криминал	Россия
Гарри Поттер и Дары Смерти: Часть II	407636	tt1201607	2011.0	8.2	8.1	12+	детектив, приключения, фантази, семейный	США, Вел
Мой парень – псих	462938	tt1045658	2012.0	7.2	7.7	18+	драма, мелодрама, комедия	США
Олдбой	75871	tt0364569	2003.0	8.1	8.3	18+	триллер, драма, криминал, детектив	Корея Юж
Репортаж из преисподней	417895	tt1245112	2009.0	6.6	6.5	18+	триллер, ужасы	Испания
Обитель зла 3	88172	tt0432021	2007.0	6.6	6.2	18+	триллер, фантастика, боевик, ужасы	США, Фр
Бэтмен	4205	tt0096895	1989.0	7.5	7.5	12+	триллер, криминал, фантастика, боевик	США, Вел
Как прогулять школу с пользой	1032308	tt6330246	2017.0	8.3	6.9	6+	драма, комедия, семейный	Франция

Рис. 4. Импортирован файл data_films.csv

Датасет выполняет роль промежуточного логического слоя между источником данных и визуализациями. Позволяет:

- унифицировать названия и типы данных;

- сократить риск ошибок при построении графиков;
- подготовить данные под конкретные аналитические задачи без изменения исходного хранилища.

Таким образом, корректная настройка датасета повышает качество аналитической модели и обеспечивает воспроизводимость результатов.

3.4 Создание чарта и визуализация данных в виде столбчатой диаграммы

На четвёртом этапе на основе созданного датасета разработан чарт, реализованный в виде столбчатой диаграммы. Данный тип визуализации используется для сравнения величин между категориями и оценки распределения показателей (рис. 5).

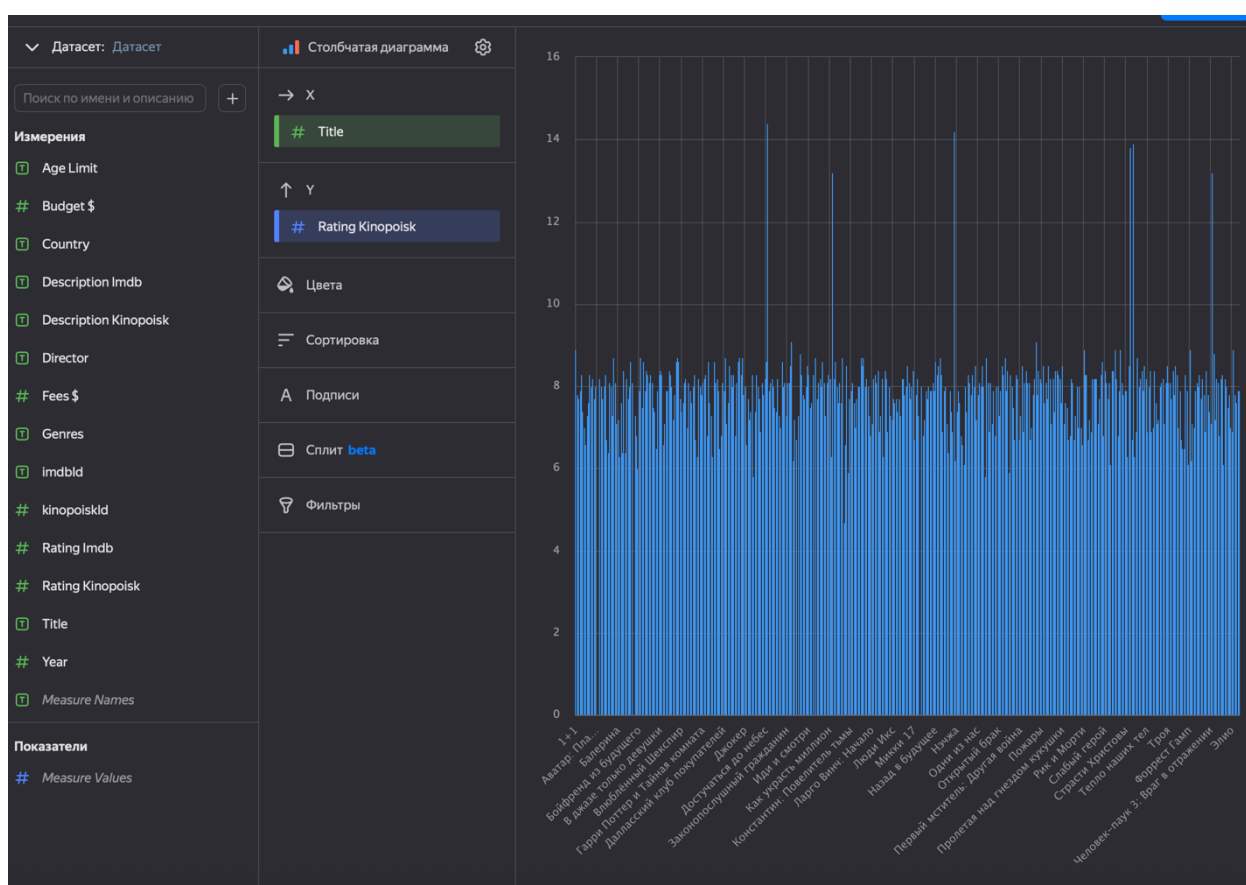


Рис. 5. Столбчатая диаграмма

Визуализация представляет значения рейтинга Кинопоиска для отдельных фильмов, где по оси X отложены названия фильмов (Title), а по оси Y — соответствующие значения Rating Kinopoisk. Столбчатая диаграмма позволяет оценить общее распределение рейтингов в выборке, однако из-за большого количества категорий не предназначена для точного сравнения отдельных фильмов между собой.

Анализ графика показывает, что подавляющее большинство значений рейтинга сосредоточено в узком диапазоне, приблизительно от 6 до 8 баллов. Высота столбцов для большинства фильмов визуально близка, что указывает

на относительно однородный уровень зрительских оценок в выборке. Явно выраженных экстремальных значений, значительно отличающихся от основной массы данных, на диаграмме не наблюдается.

Таким образом, данная визуализация позволяет сделать вывод не о лидерах и аутсайдерах, а о характере распределения рейтингов: оценки фильмов в выборке в основном сгруппированы вокруг среднего значения и не демонстрируют резкой поляризации. Это подтверждает, что рейтинг Кинопоиска в данном наборе данных характеризуется умеренным разбросом и отражает относительно схожий уровень восприятия фильмов аудиторией.

3.5 Проверка чарта на других параметрах

На следующем этапе был построен чарт, отображающий зависимость среднего рейтинга фильмов от года их выпуска. По оси X отложен показатель Year, по оси Y — среднее значение Rating IMDb (рис. 6). Столбчатая диаграмма позволяет сравнить значения рейтинга между отдельными годами и выявить периоды с более высокими и более низкими средними оценками.

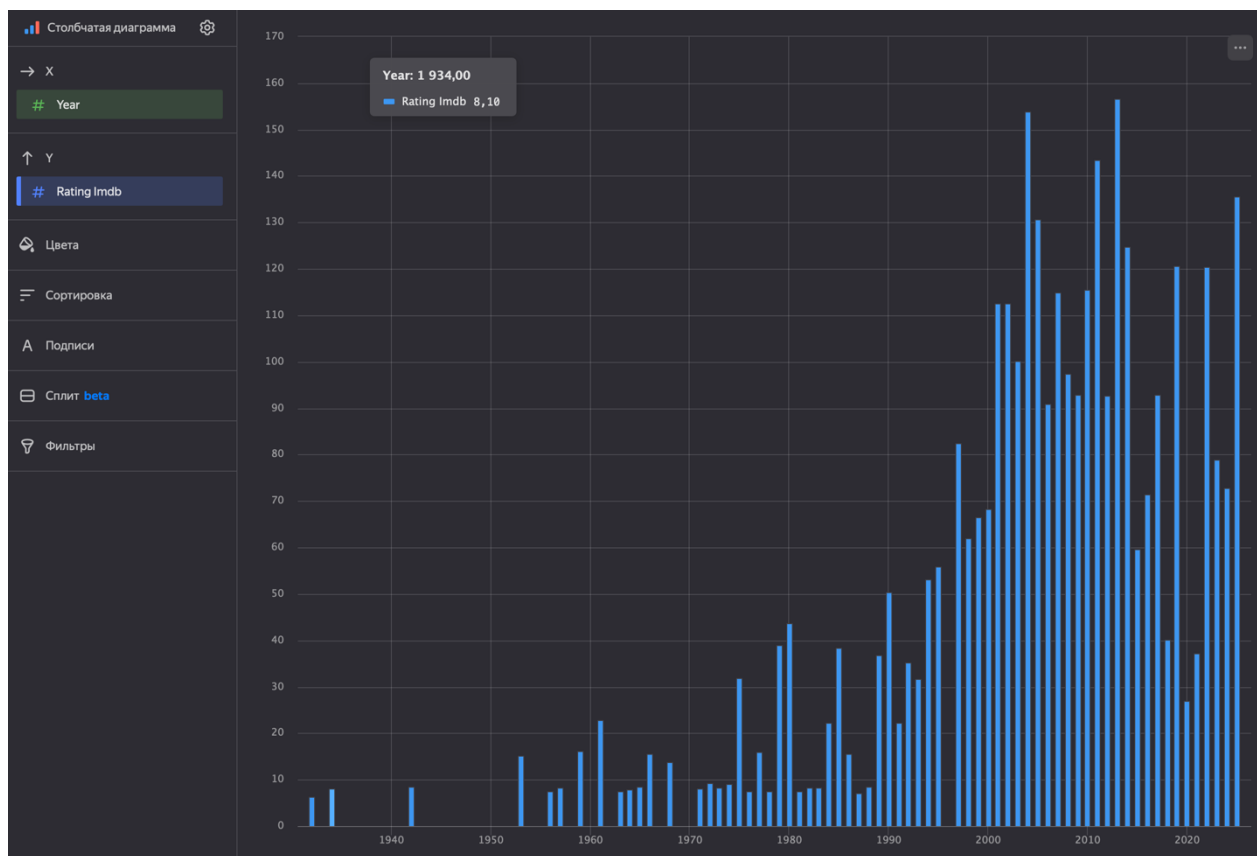


Рис. 6. Диаграмма «в какие годы снимали лучшие фильмы?»

Анализ визуализации показывает, что в ранние годы выборки (до середины XX века) значения среднего рейтинга представлены единичными столбцами с относительно низкими значениями, что указывает на ограниченное количество фильмов в данных за этот период. Начиная примерно с 1980–1990-х годов наблюдается увеличение как количества представленных лет, так и диапазона значений среднего рейтинга.

В период с конца 1990-х до начала 2010-х годов на графике фиксируется наибольшее количество лет с высокими значениями среднего рейтинга IMDb, что выражается в более высоких столбцах по сравнению с предыдущими периодами. В то же время в последующие годы значения среднего рейтинга демонстрируют заметный разброс: рядом с годами с высокими средними оценками присутствуют годы с более низкими значениями.

Отсутствие равномерного роста или снижения высоты столбцов по мере увеличения года выпуска указывает на то, что средний рейтинг фильмов изменяется не линейно. Значительные различия между соседними годами свидетельствуют о неоднородности данных и возможном влиянии состава фильмов, вошедших в выборку за конкретный год. Таким образом, визуализация показывает, что год выпуска сам по себе не определяет уровень среднего рейтинга фильмов.

3.6 Проверка чарта на других параметрах

На следующем этапе построена линейная диаграмма, отображающая соотношение рейтингов фильмов на платформах Кинопоиск и IMDb. В качестве осей использованы показатели Rating Kinopoisk (ось X) и Rating IMDb (ось Y), что позволяет напрямую сопоставить оценки одних и тех же фильмов на двух независимых платформах (рис. 7).

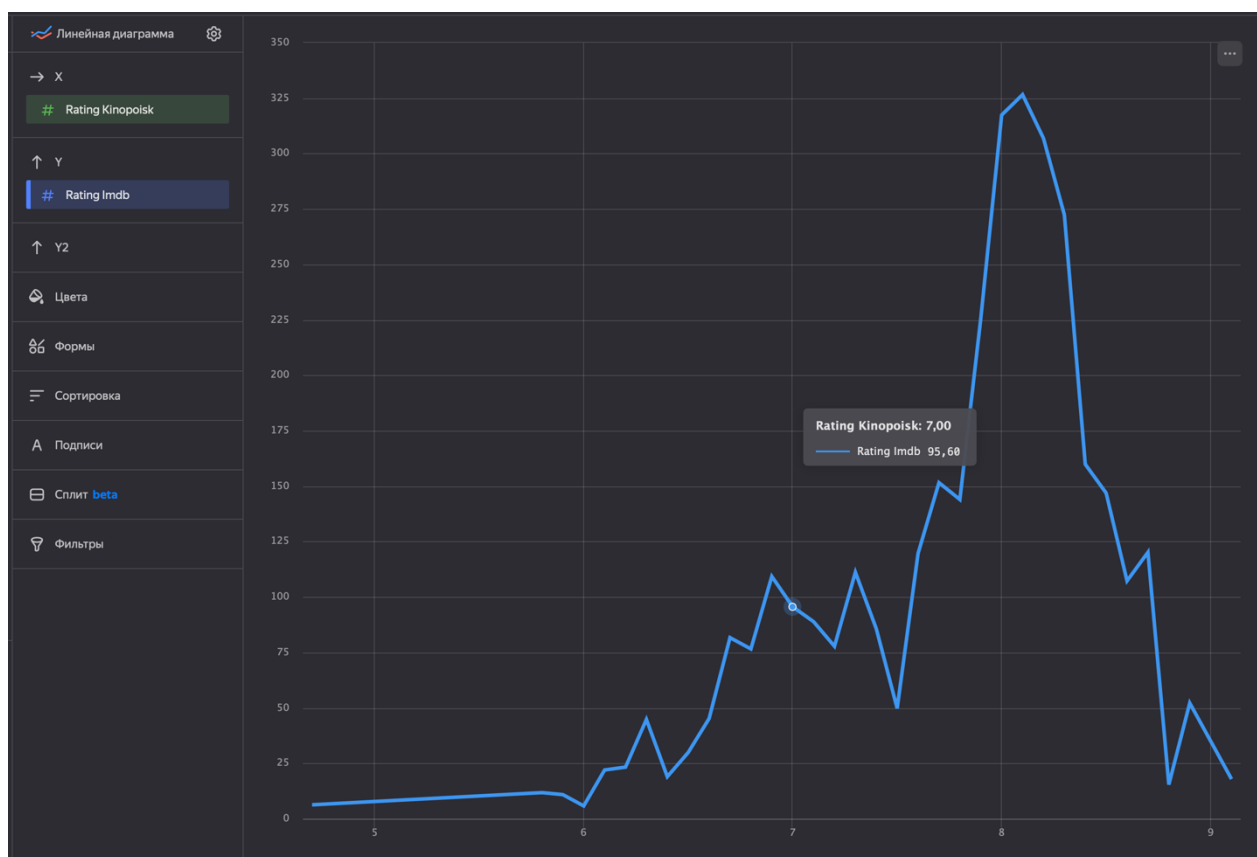


Рис. 7. Линейная диаграмма «Кинопоиск vs IMDb»

Анализ распределения значений на диаграмме показывает, что при увеличении рейтинга Кинопоиска в диапазоне примерно от 6,0 до 8,5

наблюдается соответствующий рост рейтинга IMDb. Так, фильмы с рейтингом Кинопоиска около 7,0 в большинстве случаев имеют рейтинг IMDb в интервале 80–100, а при значениях Кинопоиска выше 8,0 рейтинги IMDb достигают максимальных значений на графике. Это свидетельствует о наличии положительной зависимости между рассматриваемыми показателями.

Одновременно с этим визуализация демонстрирует заметный разброс значений рейтинга IMDb при одинаковых или близких значениях рейтинга Кинопоиска. Например, при рейтинге Кинопоиска около 7,0 наблюдаются фильмы с рейтингом IMDb как ниже среднего, так и значительно выше его. Наличие таких отклонений указывает на различия в оценке отдельных фильмов аудиториями двух платформ.

Таким образом, представленные данные наглядно показывают, что рейтинги Кинопоиска и IMDb в целом согласованы и изменяются в одном направлении, однако для части фильмов наблюдаются систематические расхождения, отражающие особенности восприятия и предпочтений различных аудиторий.

3.7 Добавление индикатора

На данном этапе создан индикатор, отображающий среднее значение рейтинга фильмов на платформе Кинопоиск. Использование индикатора позволяет получить обобщённую оценку качества фильмов в выборке без детализации по отдельным объектам (рис. 8,9).

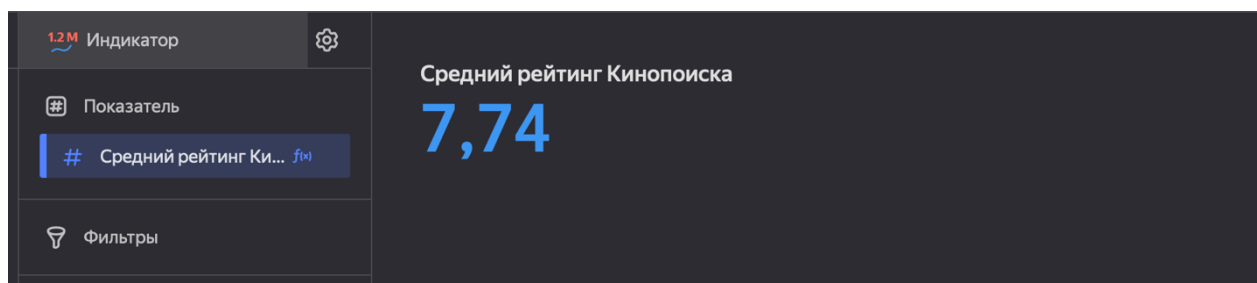


Рис. 8. Индикатор «Средний рейтинг Кинопоиска»

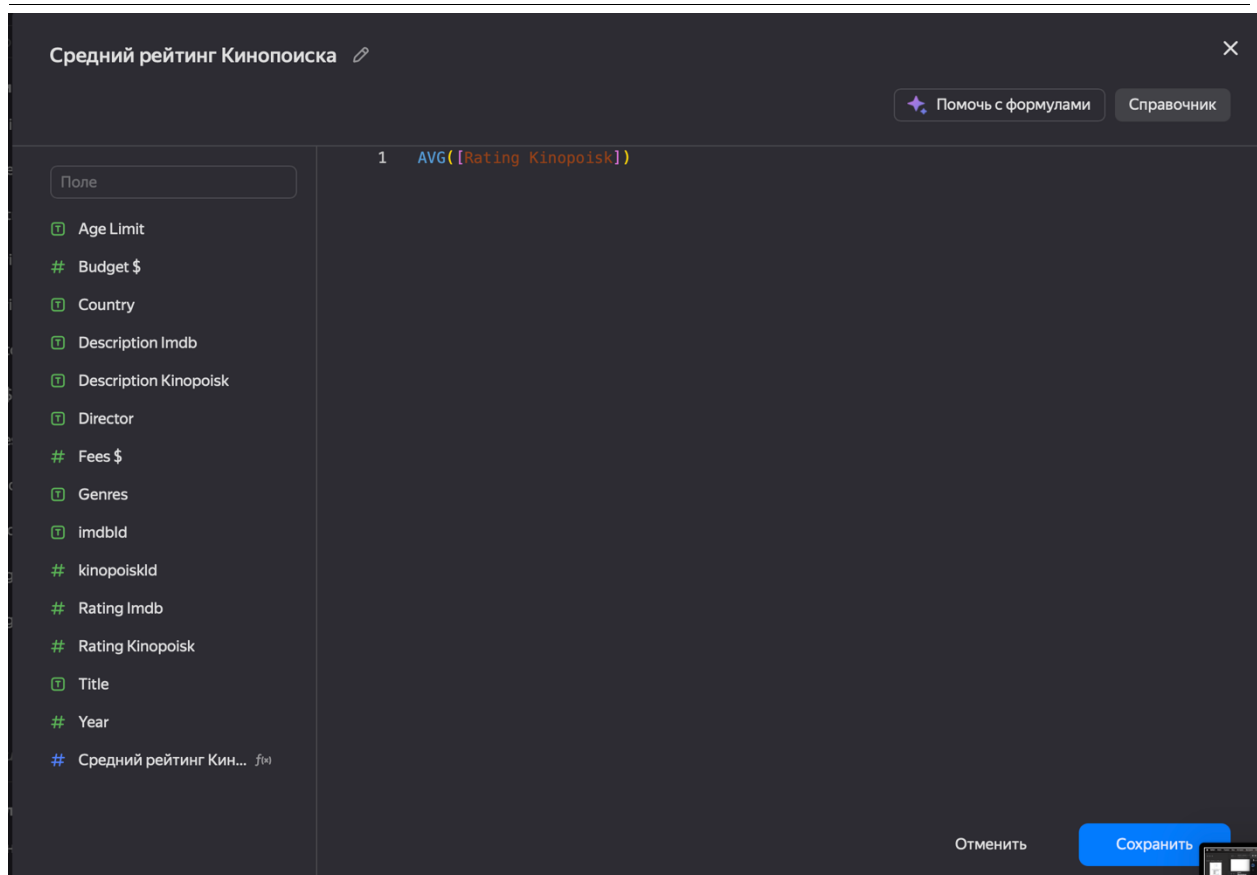


Рис. 9. Формула индикатор «Средний рейтинг Кинопоиска»

Значение среднего рейтинга характеризует общий уровень зрительской оценки и может быть использовано в качестве базового ориентира при сравнении с другими показателями или подвыборками данных. Данный индикатор служит отправной точкой для дальнейшего анализа и позволяет оценить, насколько высоко в целом аудитория оценивает представленные фильмы.

Таким образом, индикатор «Средний рейтинг Кинопоиска» демонстрирует обобщённый уровень качества фильмов в выборке и подтверждает целесообразность использования агрегированных показателей для первичной интерпретации данных.

3.8 Интерпретация вычислимого поля

В рамках анализа было создано вычисляемое поле Rating Difference, отражающее разницу между рейтингами фильмов на платформах Кинопоиск и IMDb. Данный показатель позволяет количественно оценить степень расхождения в восприятии фильмов аудиториями двух ресурсов.

Построенная визуализация показывает, что значения разницы рейтингов в большинстве случаев сосредоточены вблизи нуля, что указывает на общую согласованность оценок. Вместе с тем на графике присутствуют как положительные, так и отрицательные значения, свидетельствующие о фильмах, которые заметно выше оцениваются на одной из платформ.

Таким образом, использование вычислимого поля позволяет не только подтвердить наличие общей корреляции между рейтингами, но и выявить систематические различия в оценке отдельных фильмов. Это демонстрирует аналитическую ценность созданного показателя и обосновывает его применение для сравнительного анализа (рис. 10,11).

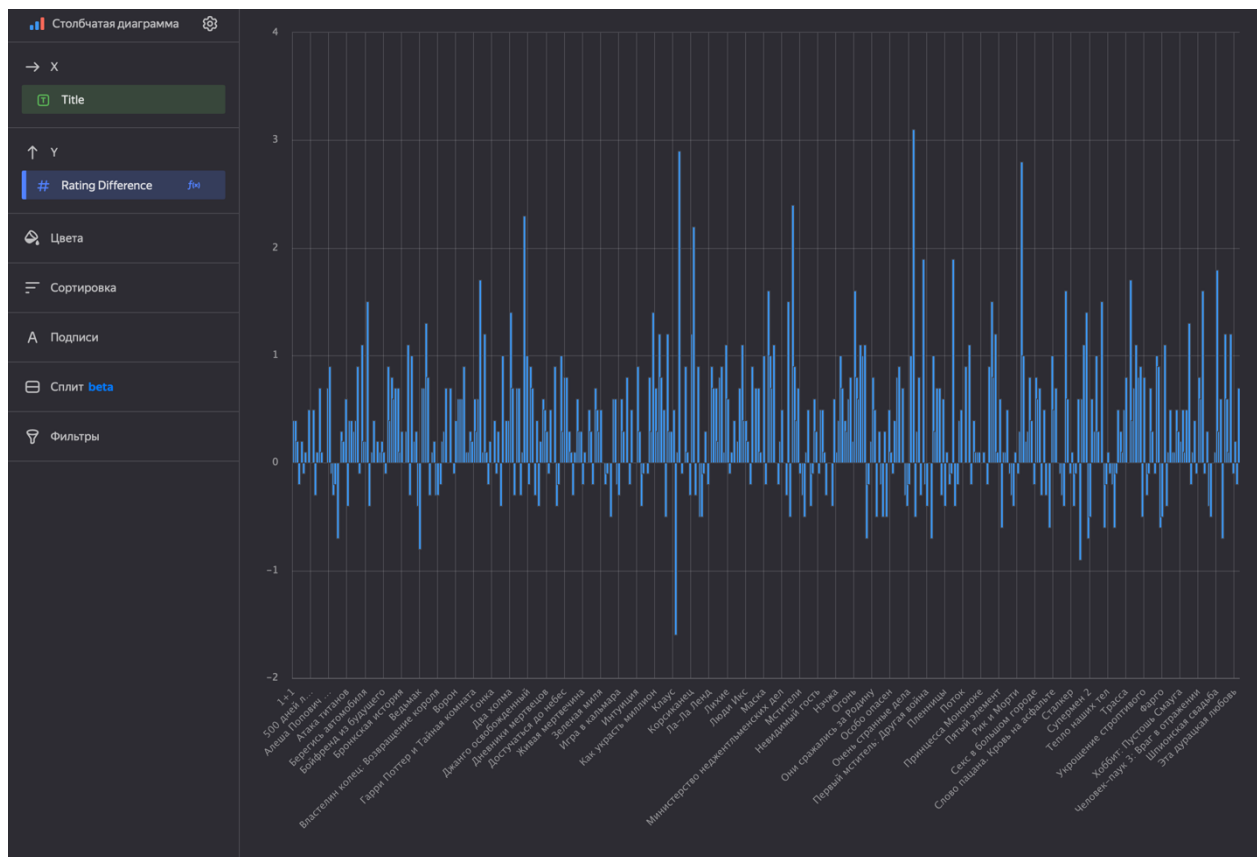


Рис. 10. Диаграмма «Рейтинг разницы IMDb и Кинопоиска»

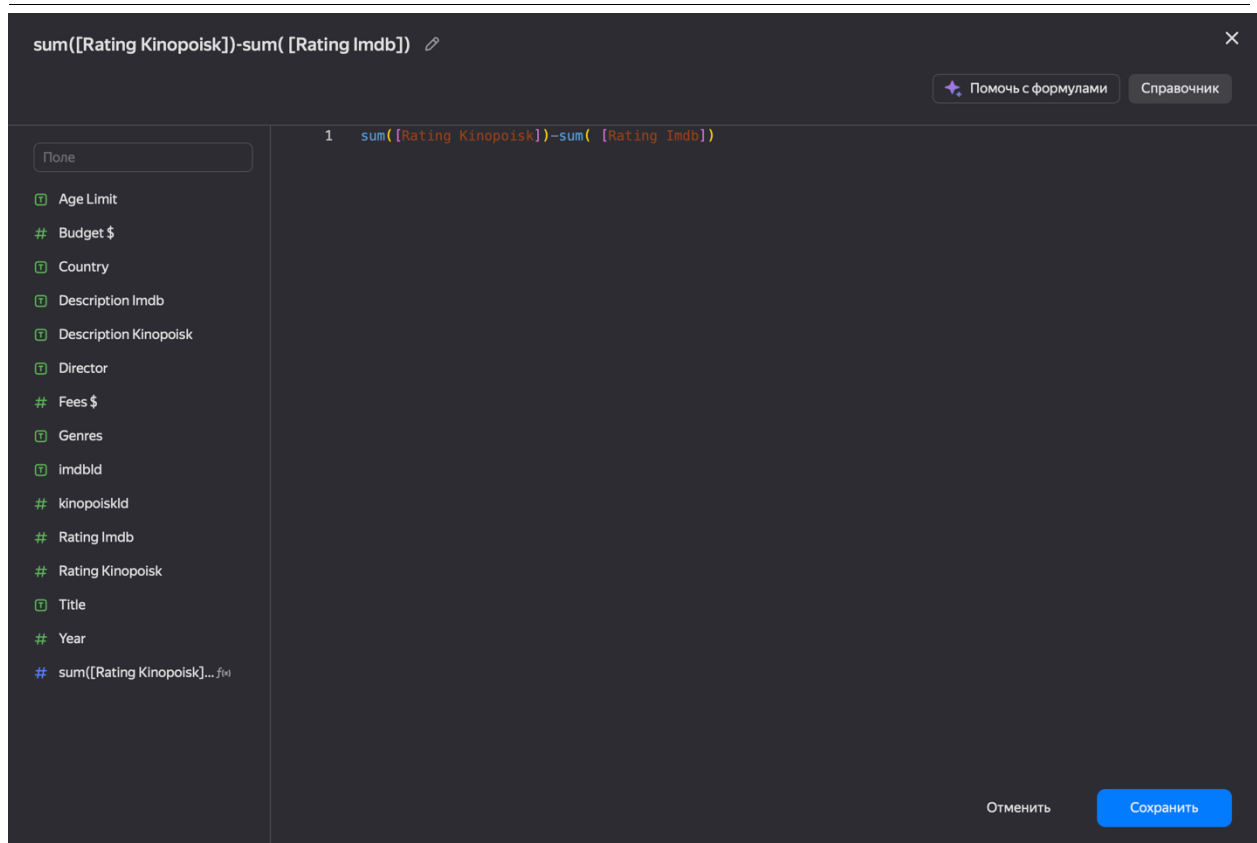


Рис. 7. Формула разности рейтингов

4 Выводы

В результате проведенного исследования рассмотрены возможности платформы Yandex DataLens для анализа и визуализации данных на примере набора данных о фильмах. В ходе работы реализован поэтапный процесс визуального анализа, включающий создание рабочего пространства, формирование датасета и построение аналитических чартов с использованием различных типов диаграмм.

Практическая реализация показала, что платформа Yandex DataLens обеспечивает удобные инструменты для визуализации данных и позволяет представлять аналитическую информацию в наглядной и интерпретируемой форме. Использование столбчатых диаграмм позволило эффективно отобразить распределение и структуру данных, а также выполнить сравнительный анализ показателей.

Полученные результаты подтверждают целесообразность применения Yandex DataLens для работы с тематическими наборами данных в учебных и исследовательских проектах. Облачная архитектура сервиса, доступность через веб-интерфейс и поддержка различных типов визуализации способствуют упрощению аналитического процесса и повышению наглядности результатов анализа.

В дальнейшем направлениями развития исследования могут являться расширение набора используемых визуализаций, применение интерактивных фильтров и параметров, а также построение комплексных дашбордов для более глубокого анализа данных.

Библиографический список

1. Березовская Е. А., Крюков С. В. Работа с сервисом бизнес-аналитики Yandex DataLens: учебное пособие. Ростов-на-Дону; Таганрог: Издательство Южного федерального университета, 2022. 92 с. URL: <https://books.google.com/books?hl=ru&lr=&id=-2KnEAAAQBAJ> (дата обращения: 11.01.2026).
2. Грищенкова Е. Н., Кутепов Ю. И., Кутепова Н. А. Разработка информационной системы онлайн-мониторинга на базе Yandex DataLens // Горный информационно-аналитический бюллетень. 2024. № 11–1. С. 5-24. URL: https://giab-online.ru/files/Data/2024/11/11-1_2024_5-24.pdf (дата обращения: 11.01.2026).
3. Хадорич Д. Д., Шелепина О. Д. Сравнительный анализ инструментов визуализации больших объемов данных на примере Power BI и Yandex DataLens // Вызовы глобализации и развитие цифрового общества в условиях новой реальности: сборник материалов IV Международной научно-практической конференции, Москва, 19 декабря 2022 года. М.: Алеф, 2022. С. 133-136. URL: <https://elibrary.ru/item.asp?id=50138371> (дата обращения: 11.01.2026).
4. Сачкова Г. Г. Анализ систем для создания дашбордов // Сборник тезисов студенческой открытой конференции : Теоретические и практические результаты исследования бакалавров, магистрантов и аспирантов института цифрового образования Московского городского педагогического университета, Москва, 21–25 ноября 2022 года / Сост. Н.В. Вознесенская. М.: Издательство ПАРАДИГМА, 2022. С. 139-142 URL: <https://elibrary.ru/item.asp?id=53324545> (дата обращения: 11.01.2026).
5. Котова М. Ю. Применение BI-систем для анализа эффективности маркетинговых инструментов // Задачи и возможности международного трансфера инновационных технологий: сборник статей по итогам Международной научно-практической конференции (г. Калуга, 30 декабря 2023 г.). Стерлитамак: АМИ, 2023. С. 97–98. URL: <https://ami.im/sbornik/MNPK-559.pdf#page=97> (дата обращения: 11.01.2026).
6. Yandex DataLens Documentation. Официальная документация сервиса Yandex DataLens // Yandex Cloud. 2026. URL: <https://cloud.yandex.ru/docs/datalens> (дата обращения: 11.01.2026).
7. Yandex Cloud. Платформа аналитики и визуализации данных Yandex DataLens. 2026. URL: <https://cloud.yandex.ru/services/datalens> (дата обращения: 11.01.2026).
8. Movie and TV Series Data from Kinopoisk and IMDb. Kaggle, 2024. URL: <https://www.kaggle.com/datasets/mostov/movie-and-tv-series-data-from-kinopoisk-and-imdb> (дата обращения: 11.01.2026).