

**Яндекс Афиша**

# **Аналитика Яндекс Афиши**

Веденеева Екатерина  
Ковалёва Ольга  
Группа 5140904/30201

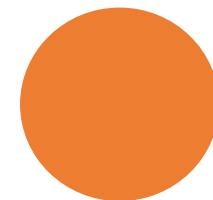
15.12.2023

# Предметная область

Яндекс Афиша – сервис для поиска мероприятий и продажи билетов.

Один из самых популярных сервисов, позволяющих отслеживать предстоящие события, подбирать мероприятия на любой вкус и относящиеся к любой ценовой категории.

Исследование выбранной предметной области вызывает большой интерес как у пользователей так и организаторов мероприятий. Анализ собранных данных может помочь в вопросах ценообразования, выбора места проведения мероприятий, целевой аудитории и направленности мероприятия.



# Сторонние исследования предметной области

- Работа с той же предметной областью была проведена в рамках одного из Проектов Яндекс.Практикума, задачей которого была помощь маркетологам с оптимизацией маркетинговых затрат и собирались данные о посещениях сайта Яндекс.Афиши, выполненных заказах за определенный период, статистике рекламных расходов.
- Чаще сбор данных по выбранной нами предметной области осуществляется с целью систематизировать и представить в более полном виде картину актуальных культурных событий по как можно большему числу населенных пунктов.



→ Агрегатор событий культурной жизни страны ←

→ Онлайн продажи билетов в кино, концерты и театры ←



**AFISHA7.RU**

# Цель и задачи проекта

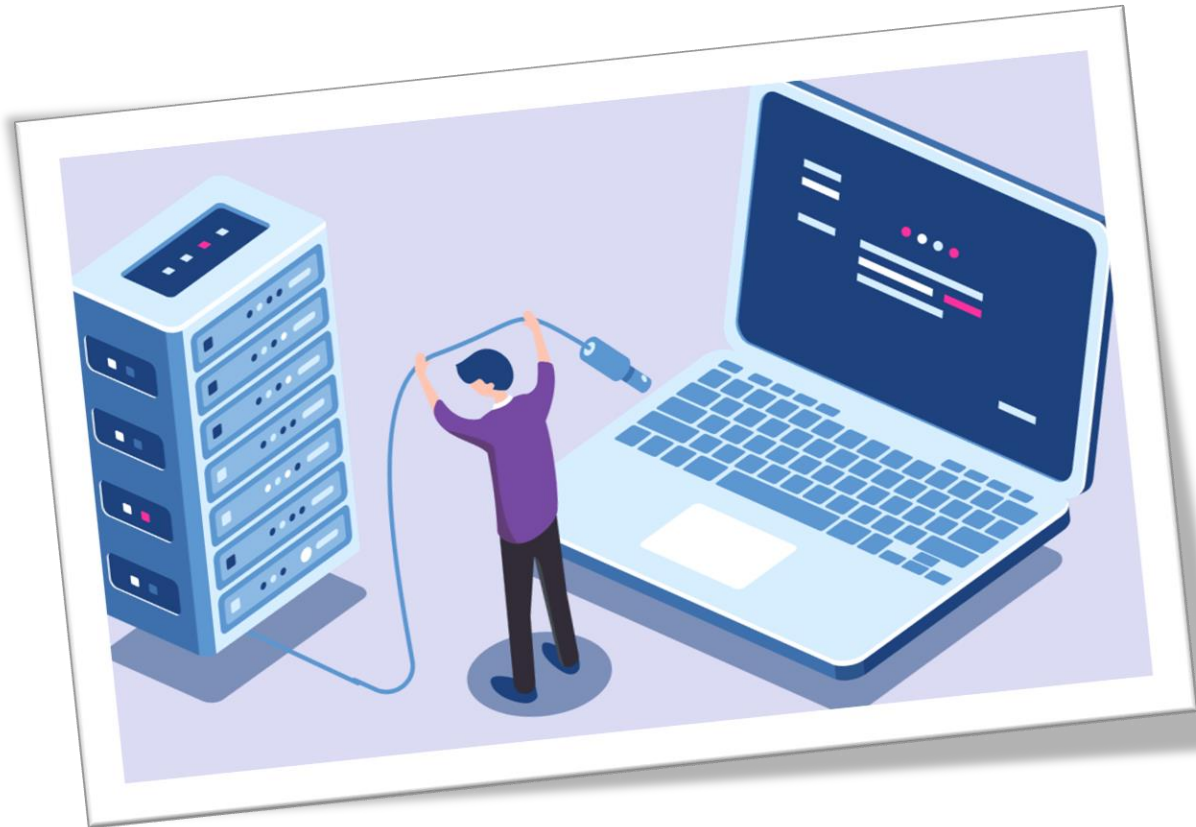
Разработка системы анализа данных о культурных событиях за период с 1 января 2024 года по 31 декабря 2024, опубликованных на начало декабря 2023.

## **Задачи:**

- Анализ географического распределения событий, представленных в агрегаторе Яндекс Афиша.
- Анализ распределения числа событий по видам и возрастным ограничениям.
- Анализ доступности «Пушкинской карты» и распределения событий с возможностью оплаты «Пушкинской картой» по видам.
- Выявления самого дешевого и самого дорогого билета.

# Формирование набора данных

- Доступ к API возможен только для партнеров Яндекс Афиши на коммерческой основе.
- URL запроса взят из XMLHttpRequest-запроса, выполняемого браузером при динамической загрузке страницы событий. Ответом на такой запрос является JSON-файл, содержащий список просматриваемых событий
- Сбор данных проводился в течение одного дня на двух машинах посредством непересекающихся запросов.
- В результате было собрано 1,22 ГБ данных подневно для каждого из 681 города, большая часть из которых содержала пустые списки событий.



# Ограничения при работе с источником

- Для создания реалистичности обращения к серверу запросы при переходе к новому дню отправлялись с интервалом в 3 секунды.
- При сборе данных были выявлены проблемы с получением ответа от сервера Яндекс Афиши в определенные часы: 00:00, 08:00, 16:00. Искусственным образом было создано “пережидание” предполагаемых технических работ или обновлений.



- Максимальное число событий, получаемых с одного запроса – 20.
- Недоступны архивные события.

# Предобработка данных

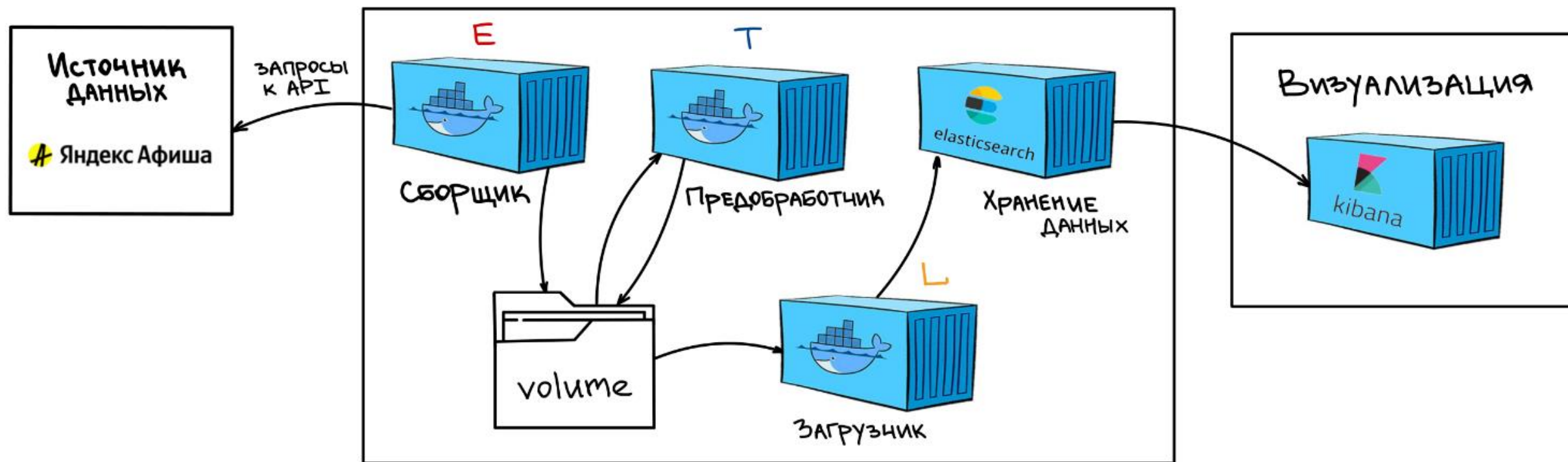
- С помощью предобработки удалось уменьшить размер полного набора данных до 57,7 МБ.
- Данные хранятся в виде JSON-файлов для каждого события за 2024 год. Всего – 51512 файлов.
- Итоговый файл содержит информацию о мероприятии, включающую рейтинг среди пользователей, минимальную и максимальную стоимости, информацию о месте проведения, в том числе ближайшие станции метро и географические координаты.

# Пример полученного файла

```
{
  "city_id": "moscow",
  "date": "01_03_2024",
  "title": "Вий",
  "contentRating": "12+",
  "userRating": {
    "value": 9,
    "count": 49
  },
  "type": {
    "name": "Театр",
    "type": "format"
  },
  "tags": [
    {
      "name": "Мюзикл",
      "type": "format"
    }
  ],
  "price": {
    "min": 1000,
    "max": 5000
  },
  "place": {
    "tags": [
      {
        "name": "Театр"
      }
    ],
    "title": "Театр Гоголя",
    "address": "ул. Казакова, 8",
    "city": "Москва",
    "location": {
      "lon": 37.662012,
      "lat": 55.762181
    },
    "metro": [
      "Курская",
      "Чкаловская"
    ]
  },
  "pushkinCardAllowed": false
}
```

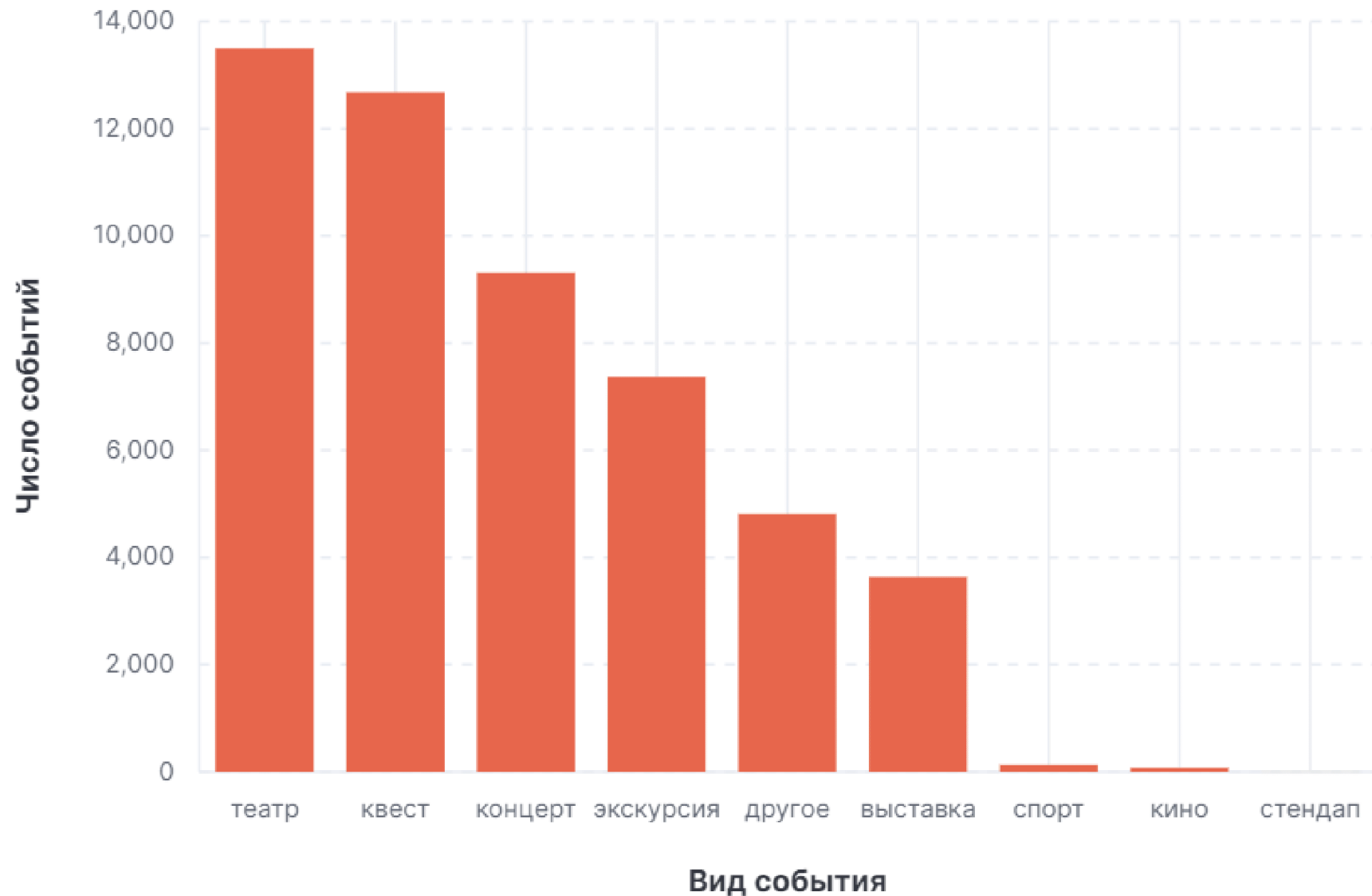


# Архитектура программного решения

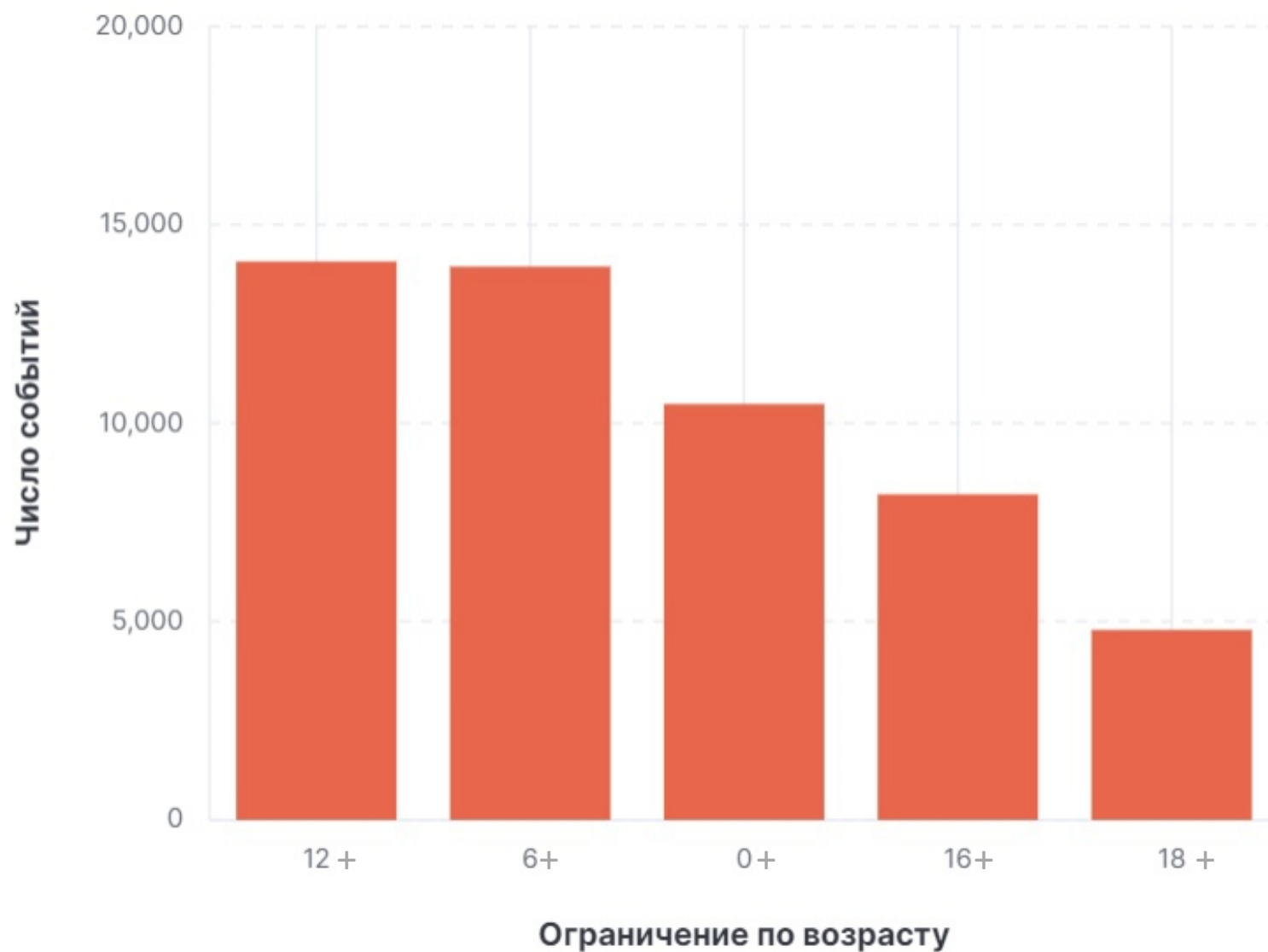




# Распределение по видам событий



# Распределение по возрастным ограничениям



# Максимальная и минимальная стоимости

Максимальная цена билета

**600,000**

Вип ложа на концерт рок группы  
Парк Горького в Крокус Сити Холл

Минимальная цена билета

**0**

- Около 2500 событий
- Билеты для детей





# События, одобренные Пушкиным Минкультом

