

федеральное государственное автономное образовательное учреждение высшего
образования «Национальный исследовательский университет ИТМО»

ОТЧЁТ
по исследовательскому проекту
(Анализ отзывов на Яндекс.Картах; Торговые центры Москвы)

Проект выполнили студенты 1 курса:

Дубовицкий Александр

Кириллов Георгий

Агеев Никита

Лебедев Александр

Ким Артём

Содержание:

1. Описание проекта
2. Содержательная часть
 - 2.1. Сбор данных и их обработка
 - 2.2. Проектирование структуры данных
 - 2.3. Анализ полученных данных
 - 2.4. Полученные выводы
3. Заключение
4. Результат проекта

1. Описание проекта

Данный проект был создан в рамках учебной практики в университете ИТМО, ориентированной на исследовательскую деятельность. Проект представляет из себя исследование зависимостей оценок пользователей от различных факторов: количество отзывов, расположение торговых центров и т.д.

Конечная цель проекта - применение анализа данных и текстовой обработки для изучения пользовательских оценок и откликов на торговые центры.

Задачи проекта:

1. Сбор отзывов на торговые центры с сайта [Яндекс.Карты](#).
2. Создание способа хранения отзывов.
3. Анализ полученных метрик и их визуализация.
4. Подведение итогов проделанной работы.

2. Содержательная часть

2.1 Сбор данных и их обработка.

Сбор данных наша команда производила вручную с помощью парсера, написанного на языке программирования Python и библиотеки Selenium. Сначала мы получили json-файл с идентификационными номерами торговых центров и их названиями, благодаря которому получили json-файл с отзывами на эти самые торговые центры и датами их добавления.

2.2 Проектирование структуры данных

Мы представили данные в виде базы данных, показанной на картинке ниже:

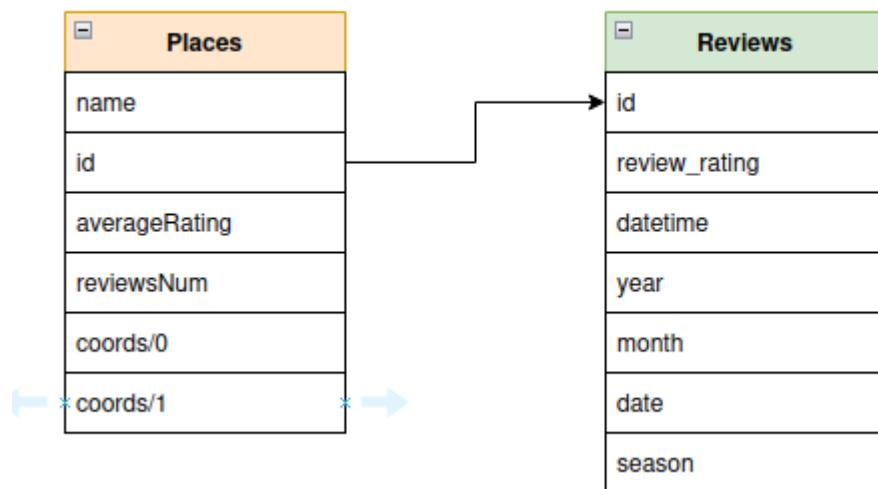


Рис.1 Схема базы данных

2.3 Анализ полученных данных

Получив данные в виде CSV-таблиц, мы создали несколько чартов, построили зависимости среднего рейтинга от количества отзывов и зависимость среднего рейтинга от сезонов. Самым худшим торговым центром по среднему рейтингу оказался “Скандинавский пассаж” с рейтингом 2.0.

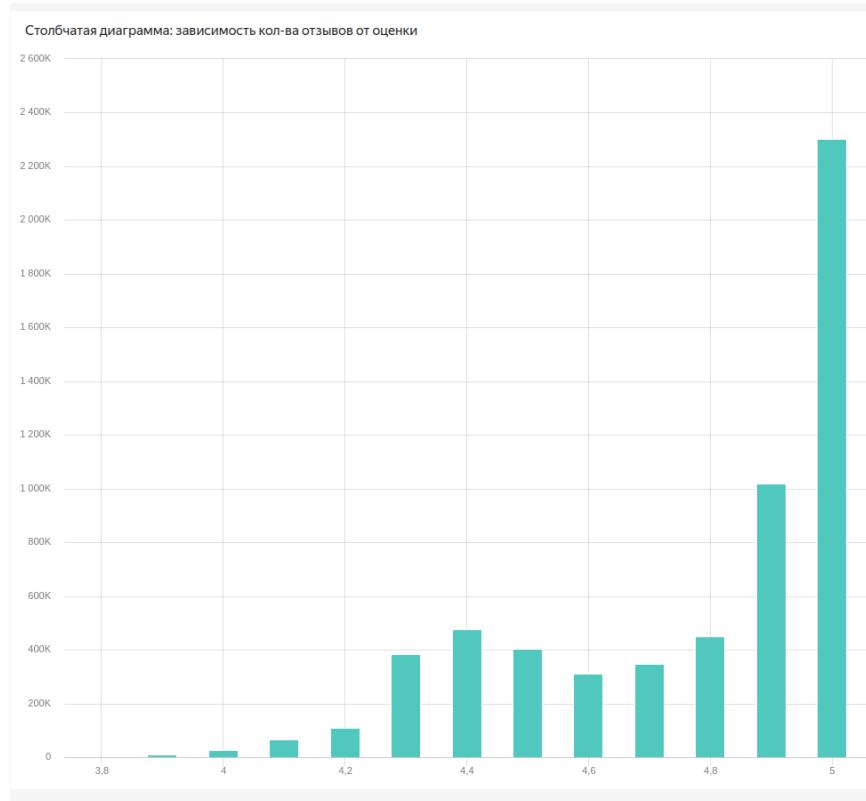


Рис.2 Зависимость количества отзывов от оценки

Данный график иллюстрирует распределение кол-ва отзывов от средней оценки торгового центра. Мы наблюдаем глобальный максимум при оценке «5»(38,39% всех оценок) и локальный при оценке «4,4»(8% всех оценок).

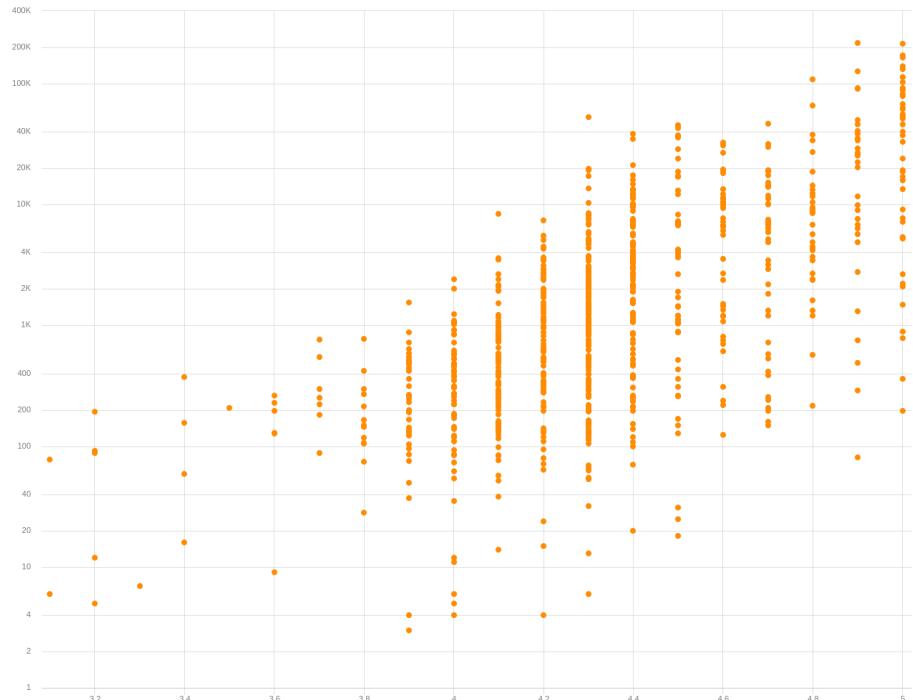


Рис.3 Точечная диаграмма: Зависимость кол-ва отзывов от средней оценки

На данном графике мы можем проследить, как в целом меняется динамика количества отзывов от средней оценки. Отмечаем положительную корреляцию кол-ва отзывов от оценки. Также заметны сгущения точек при значениях «4,3-4,4».

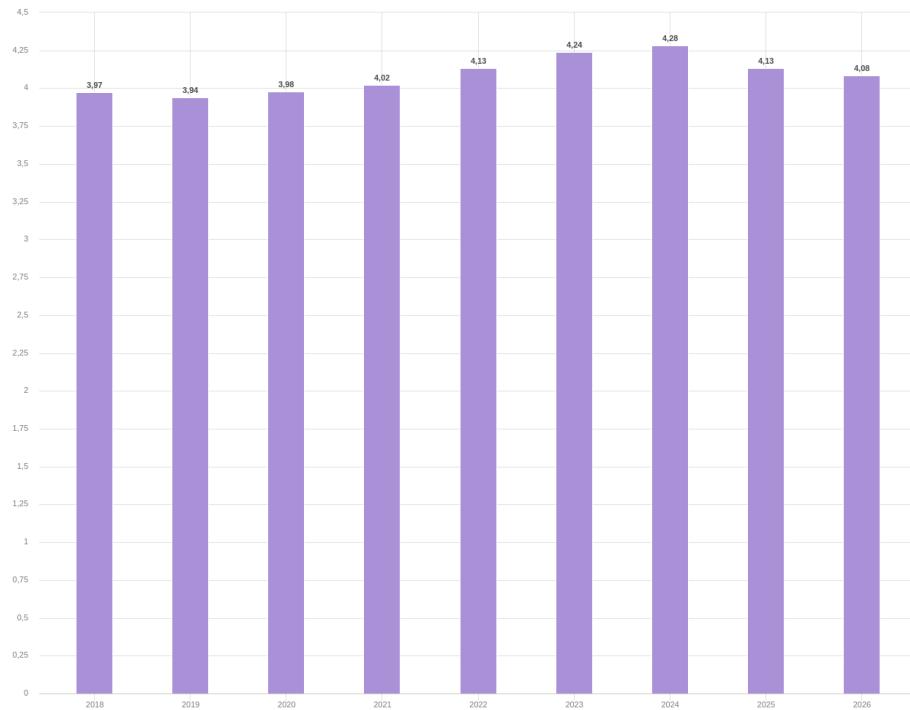


Рис.3 Зависимость средней оценки отзывов от года их публикации.

Следующий график иллюстрирует зависимость средней оценки от года публикации. Заметна положительная тенденция вплоть до 2024 года.(Увеличение на 8%, 0,3 рейтинга). Начиная с 2025 года оценки стали стремительно ухудшаться, хоть и до результатов 2018 года и не дошли.(Уменьшение на 5%, 0,2 рейтинга)

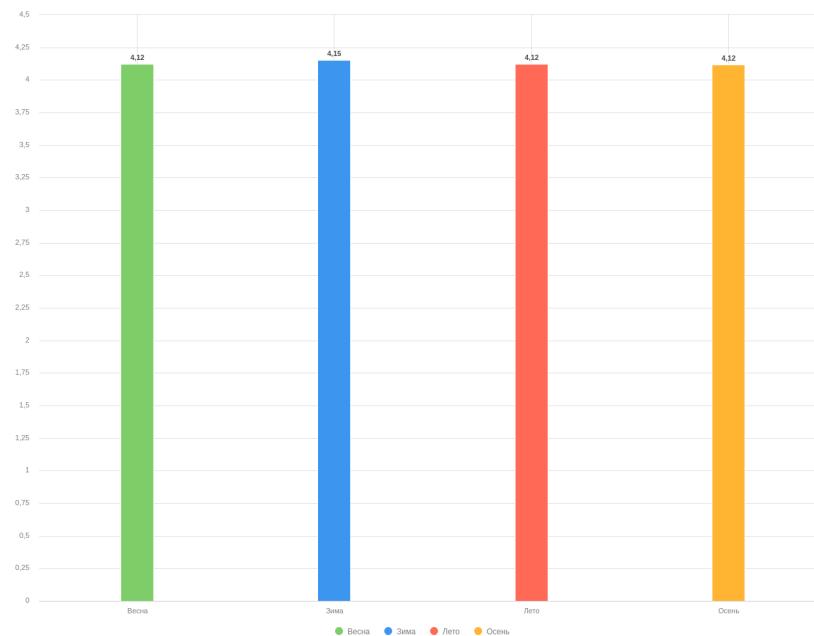


Рис.4 Зависимость средней оценки отзывов от сезона публикации.

Данная диаграмма отображает значения средней оценки отзывов относительно сезона публикации. Никакой зависимости мы не обнаружили. Некритичный рост средней оценки в зимнее время года можно связать с новогодними праздниками и распродажами.(0,7% от среднего).

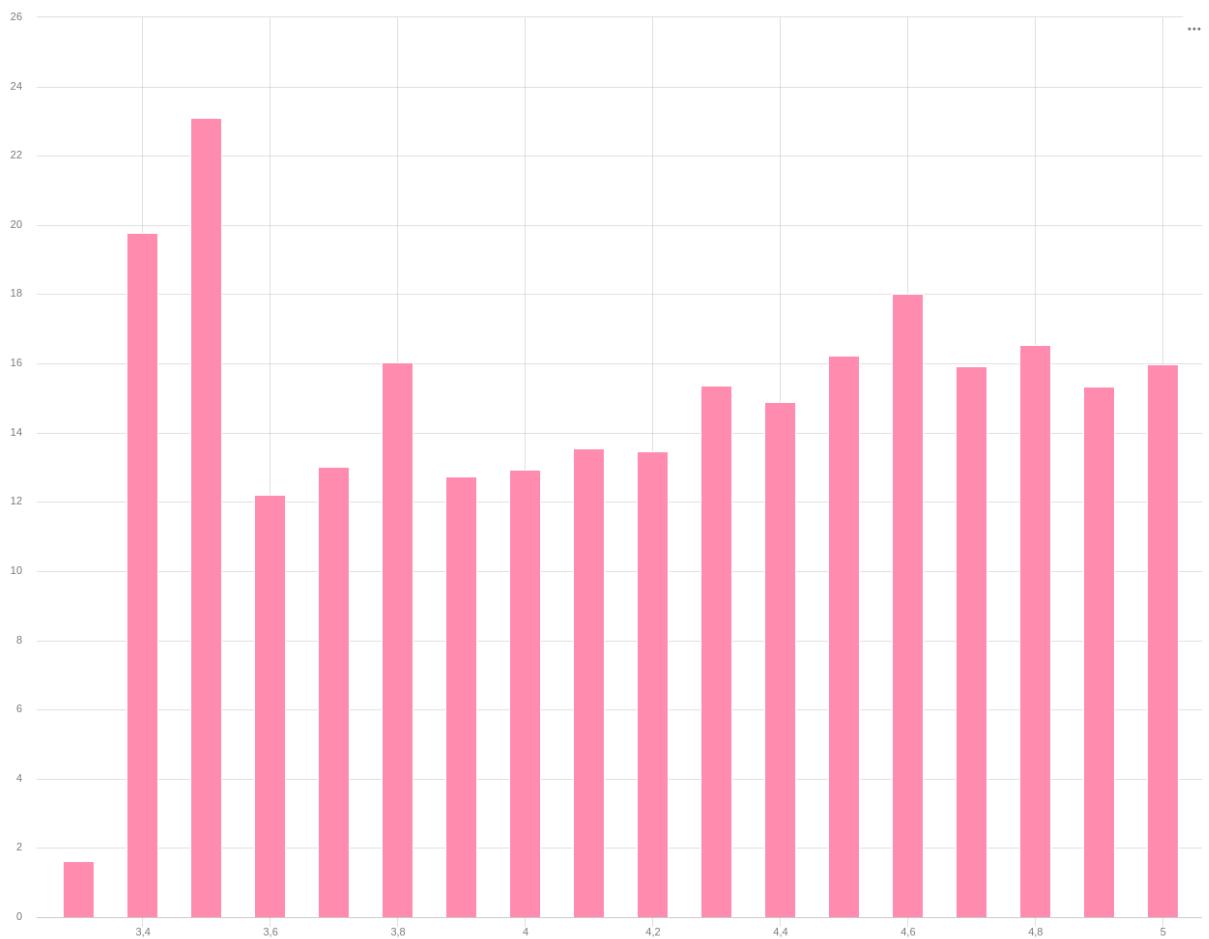


Рис.5 Зависимость расстояния от центра Москвы относительно средней оценки.

Данный график иллюстрирует зависимость расстояния от центра Москвы относительно средней оценки. Можно заметить сильный выброс(30% относительно среднего) при оценках «3,4-3,5», что может свидетельствовать о более плохом качестве торговых центров в отдаленности от центра Москвы.

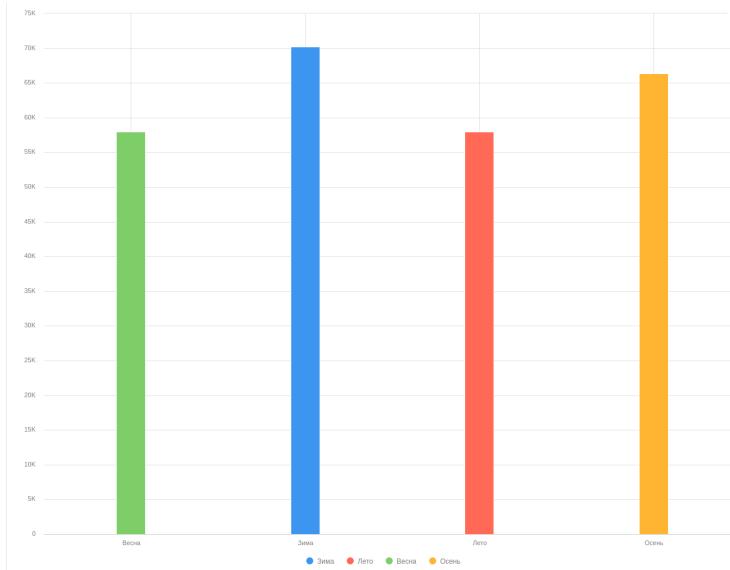


Рис.6 Зависимость количества отзывов от сезона.

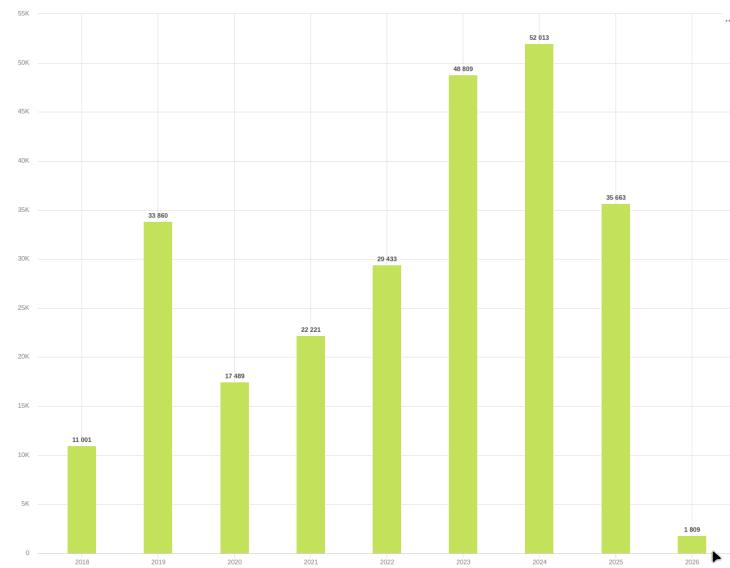


Рис.7 Зависимость количества отзывов от года.

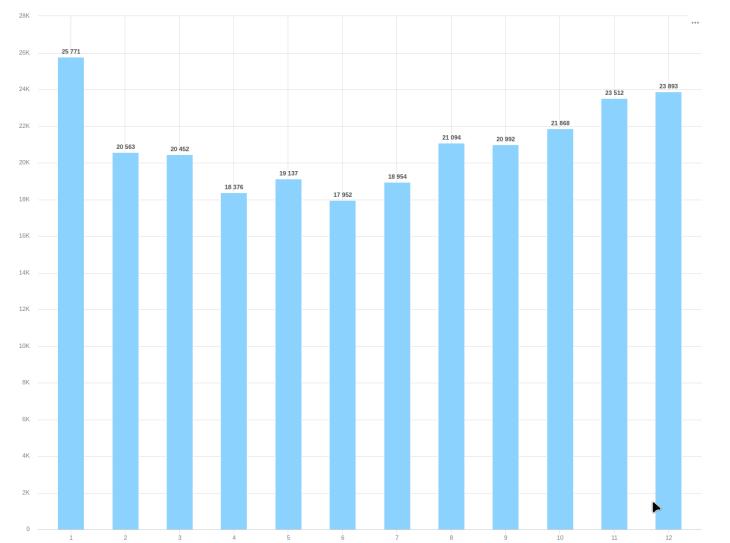


Рис.8 Зависимость количества отзывов от месяца.

По этим графикам мы сделали следующие выводы:

2024 год был самым активным относительно отзывов пользователей(52013 отзывов, 20,61% от всех отзывов); Чаще всего пользователи писали отзывы в Январе(25771, 10,21% от всех отзывов); Соответственно самый популярный сезон - Зима(70227 отзывов, 27,84% от всех отзывов).

2.4 Результаты

По итогам нашего исследования, мы сделали следующие выводы:

1. На протяжении последних 7 лет мы можем наблюдать рост средних оценок торговых центров(на 8,6% или на 0,26). Мы можем предположить, что это связано с улучшением качества обслуживания в торговых центрах. Но за последний год наблюдается сильное снижение оценки(на 4% или 0,15 рейтинга).
2. В целом по большинству показателей для торговых центров корреляция хорошая: 38,39% всех оценок - 5. Кол-во отзывов торговых центров прямо пропорционально их средним рейтингам. Среднее количество отзывов не сильно меняется в течение сезона(63074 отзыва в среднем, отклонение 11%).
3. Наблюдается аномально большое кол-во рейтингов 5.0. Их количество превышает ближайшее значение в 2,26 раз. Скорее всего, это связано с тем, что люди обычно ставят самый положительный отзыв так как их всё устраивает, чем высказывают какое-либо недовольство.

3. Заключение

В данном проекте были достигнуты поставленные цели. В ходе работы были выполнены все задачи. Все участники команды узнали много нового об анализе данных, попробовали себя в создании структуры данных, сборе информации и в визуализации данных. Итоговым результатом стал дашборд, созданный в Yandex Datalen, и репозиторий со всеми данными.

4. Результат проекта

Результаты проекта проявляются:

1. Созданная база данных
2. Парсеры сайта [Яндекс.Карты](#).
3. Репозиторий с проектом
4. Чарты и дашборды, составленные в Yandex DataLens.