

Визуализация и Лексический Анализ Отзывов

Аптечные сети Санкт-Петербурга на Яндекс Картах

Обзор Проекта: Изучение Отзывов Пользователей

Данный проект направлен на глубокий анализ отзывов пользователей Яндекс Карт о сетях аптек. Используя методы обработки естественного языка (NLP), мы выделяем ключевые тематические акценты, оцениваем тональность высказываний и визуализируем результаты для идентификации "болевых точек" бизнеса.

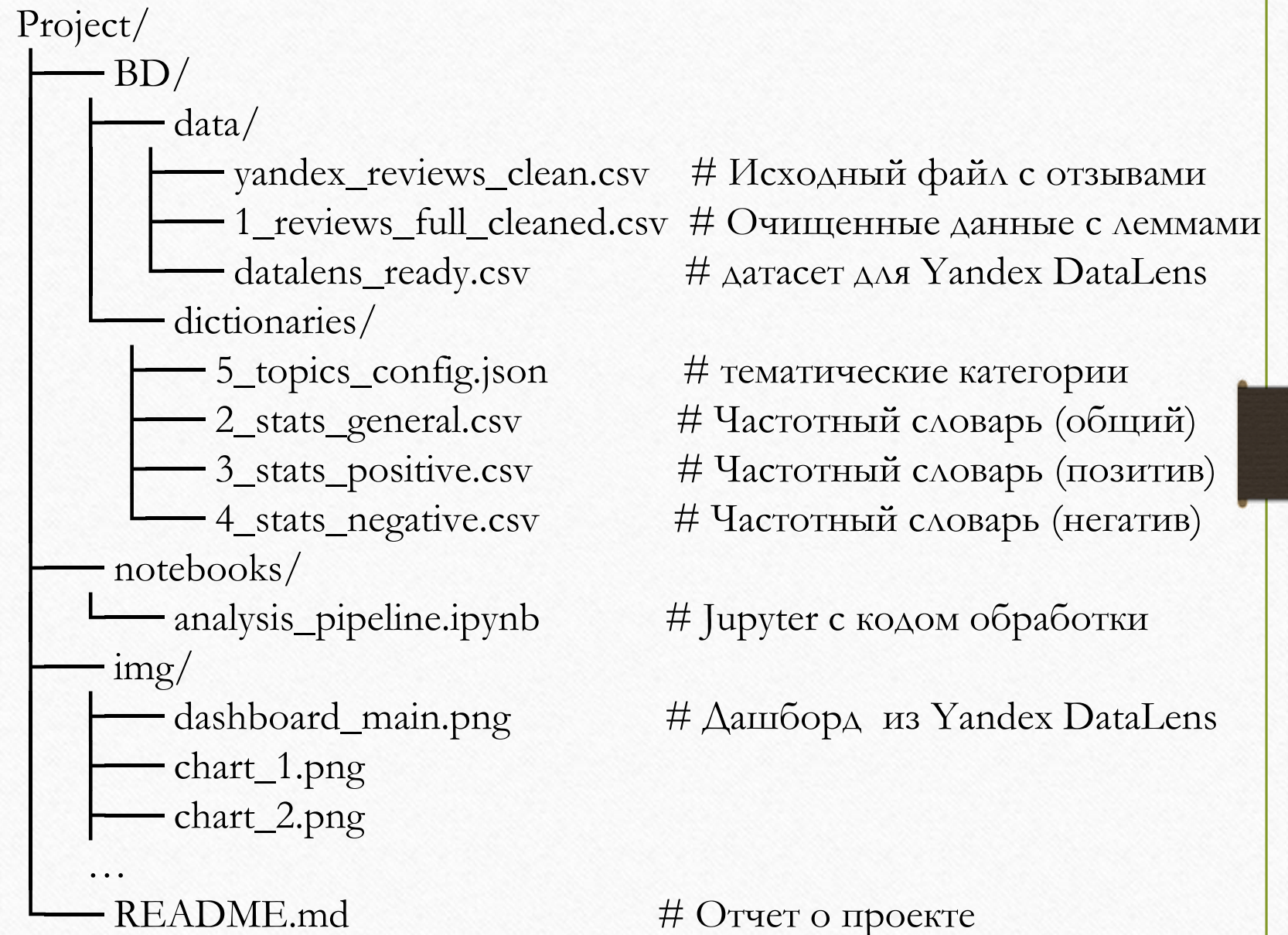
Для визуализации и построения дашборда использовался инструмент **Yandex DataLens**, предоставляющий интерактивную панель для фильтрации данных по сезонам.



Структура Проекта

Для обеспечения ясности и воспроизводимости анализа проект организован в следующую структуру:

- **BD/**: Содержит сырые и обработанные данные, а также словари для тематической категоризации.
- **notebooks/**: Включает Jupyter Notebook с кодом обработки данных.
- **img/**: Директория для всех графических материалов и визуализаций.
- **README.md**: Подробный отчет о проекте и его использовании.



Сбор Данных: Техническая Реализация Парсера

Исходный набор данных был сформирован путем автоматизированного сбора (парсинга) общедоступной информации с платформы Яндекс Карты.

Данный подход позволяет корректно обрабатывать динамический контент, подгружаемый через JavaScript, обеспечивая полное извлечение отзывов.

Техническая реализация

- **Язык:** Python
- **Библиотека:** Playwright (для автоматизации браузера)

Алгоритм работы парсера:

1. Инициализация

Запуск браузера в headless-режиме.

2. Навигация

Переход на страницу отзывов организации.

3. Бесконечная Прокрутка

Эмуляция прокрутки для загрузки всех отзывов.

4. Извлечение Данных

Парсинг автора, даты, рейтинга, текста отзыва.

5. Экспорт

Сохранение собранных объектов в структурированный файл CSV

Процесс Обработки Данных

Предобработка Данных

- Парсинг дат в datetime.
- Определение сезона (Зима/Весна/Лето/Осень).
- Очистка от пустых строк и дубликатов.

NLP

- Токенизация текста.
- Лемматизация с rymorphy2.
- Фильтрация стоп-слов и мусора.

Категоризация

- Выделение 4 ключевых тем
- (Доставка/Заказ, Цена/Выгода, Лекарства/Наличие, Сервис/Общение).

Визуализация

- Подготовка данных для Yandex DataLens.
- Построение интерактивного дашборда.

Реализованные Структуры Данных

В ходе работы были спроектированы и реализованы специальные структуры данных для хранения и агрегации текстовой информации, сохраняемые в виде CSV/JSON.

1 Обогащенная Таблица

Основная база данных, каждая строка - уникальный отзыв. Содержит: **date_clean, season, lemmas**.

BD/data/1_reviews_full_cleaned.csv

2 Частотные Словари

Три базы данных (**общая, позитивная, негативная**) для быстрого анализа топов слов.

BD/dictionaries/*.csv

3 Тематическая База Знаний

JSON-объект, сопоставляющий темы со списками ключевых слов.

BD/dictionaries/5_topics_config.json

4 Денормализованный Слой

Специализированная структура для Yandex DataLens, формат "одна строка — одно слово".

BD/data/datalens_ready.csv

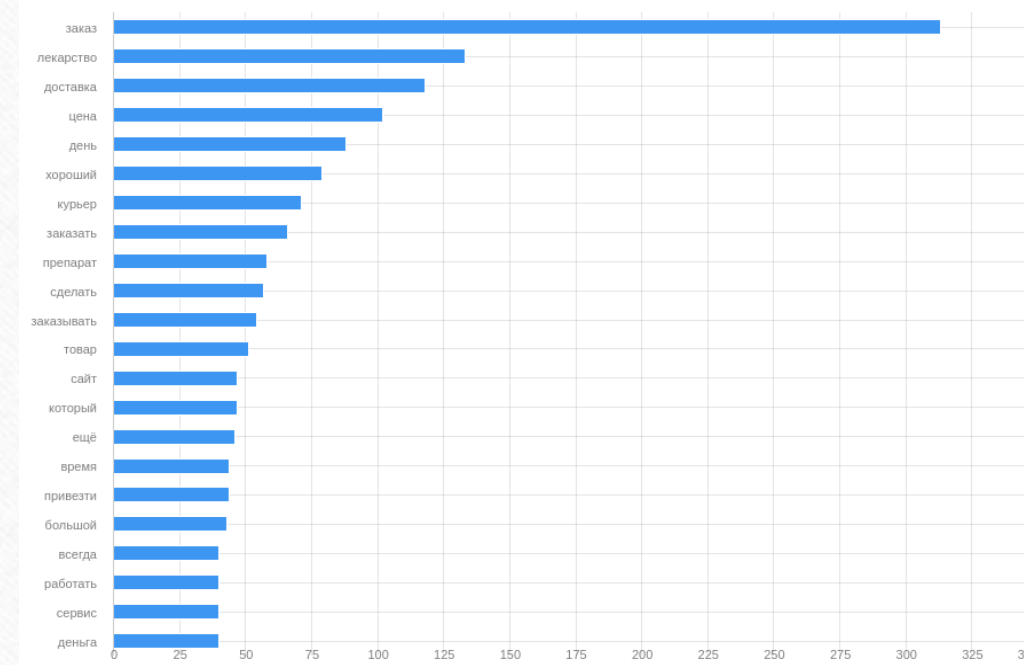
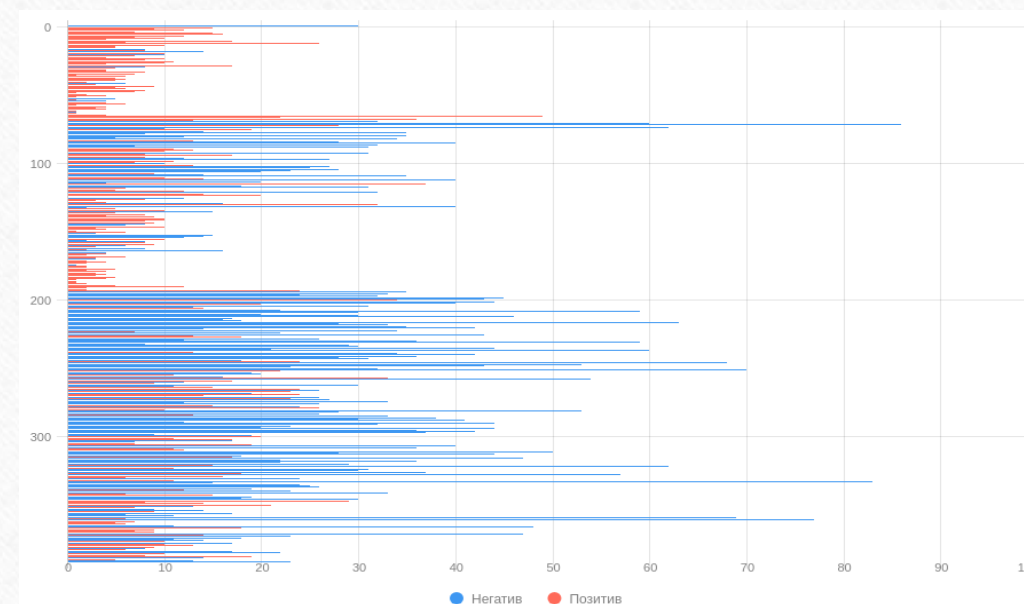
Результаты Анализа: Частотный Анализ Ключевых Слов

Лидер Упоминаний

Слово "заказ" встречается чаще всего (более 300 раз), за ним следуют "лекарство", "доставка" и "цена".

Смещение Фокуса

Аптека воспринимается как **сервис доставки**. Лексика онлайн-коммерции ("сайт", "приложение", "курьер") доминирует над традиционной ("зал", "чистота", "фармацевт").



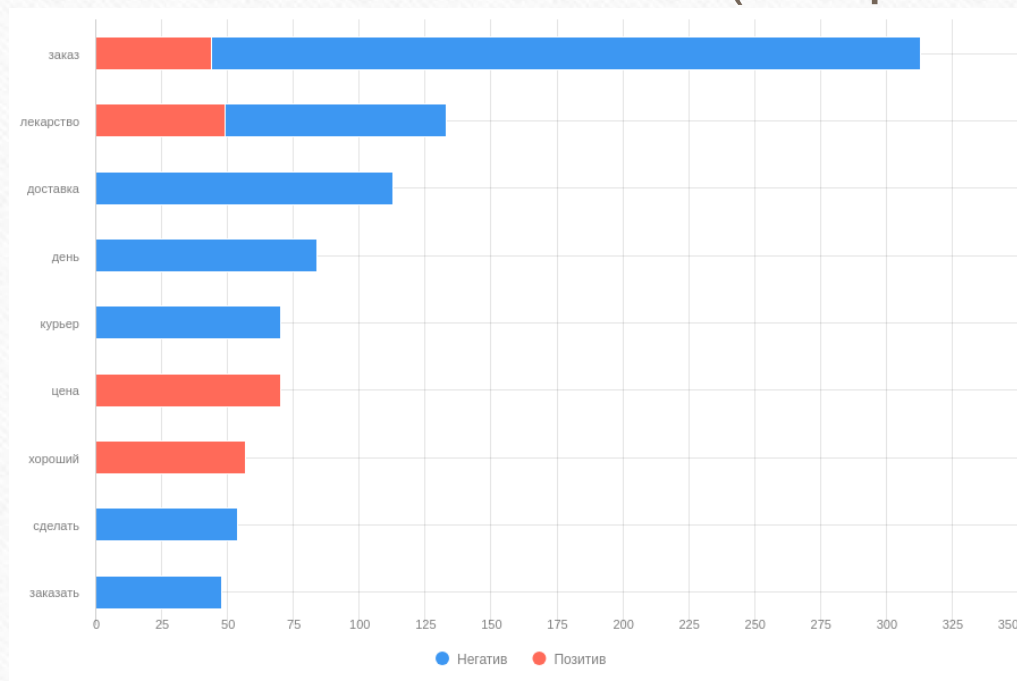
Сравнение Позитива и Негатива в Отзывах

Драйверы Негатива (Синие)

Основные жалобы связаны со словами "заказ", "доставка", "курьер".

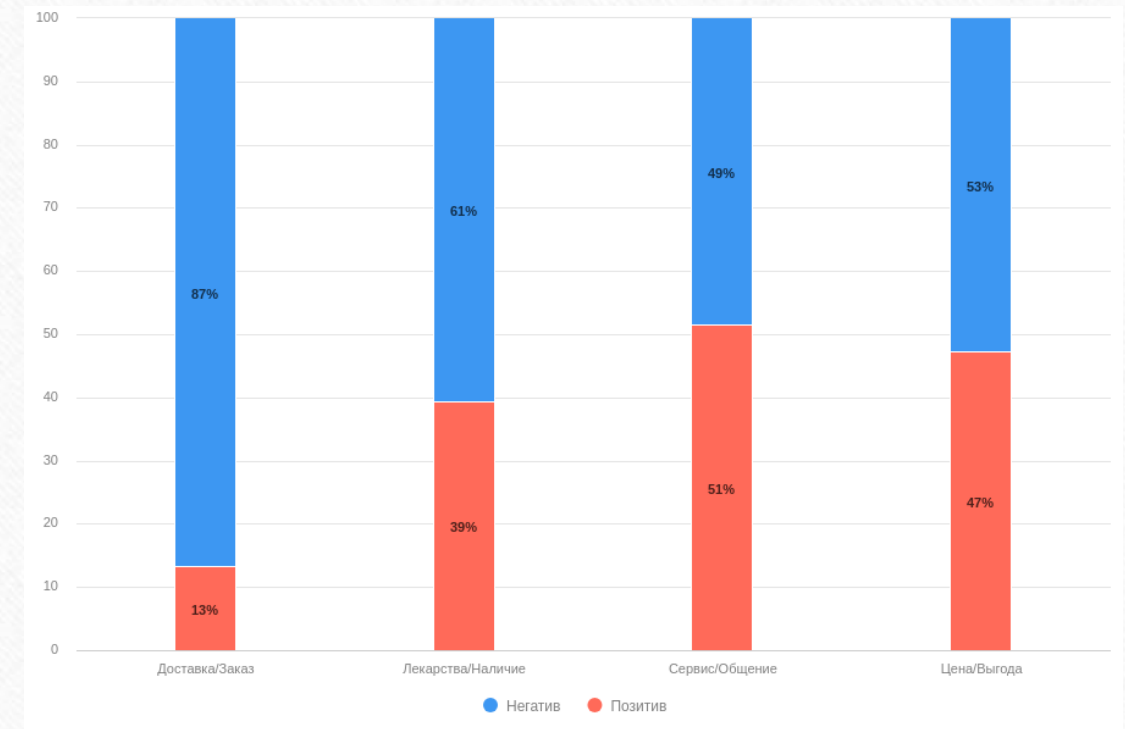
Пользователи недовольны срывами сроков и логистикой.

Слово "лекарство" встречается как в позитивном (нашли редкое), так и в негативном контексте (не привезли).



Драйверы Позитива (Красные)

Клиенты хвалят "цену" и "ассортимент" (наличие редких лекарств). В топе позитива также слова "хороший" и "вежливый".



Анализ Тематик: Выявление "Узких Горлышек" Бизнеса

Сервис/Общение

~10% упоминаний

51% позитива. Нейтрально-положительная оценка.

Лекарства/Наличие

~20% упоминаний

61% негатива. Часто из-за отсутствия товаров.

Цена/Выгода

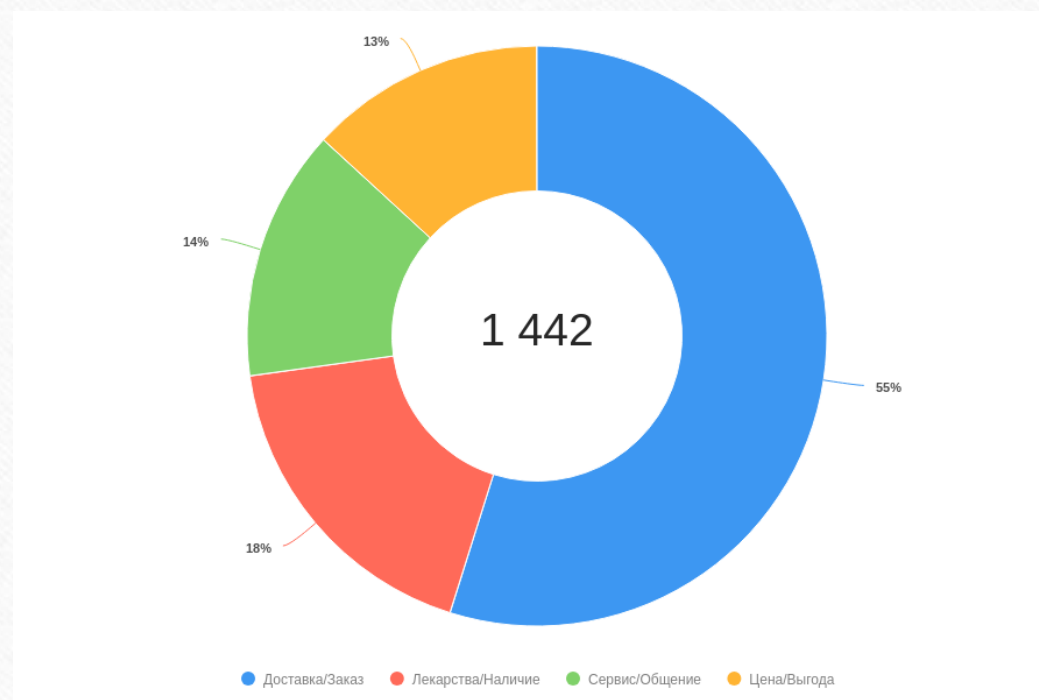
~15% упоминаний

47% позитива. Конкурентное преимущество.

Доставка/Заказ

55% упоминаний

87% негативных отзывов. Самая проблемная зона.



Сезонный анализ показывает, что проблемы с доставкой носят **системный**, а не временный характер, оставаясь стабильно высокими в любое время года.

Ключевые Выводы

Основные проблемы

Логистика.

- Тема «Доставка/Заказ» занимает 55% всех упоминаний
- 87% отзывов в этой категории — негативные

Недостаток товаров.

- 61% негатива по теме «Лекарства/Наличие»
- Клиенты часто не находят нужные препараты при оформлении заказа

Сильные Стороны

Ценовое преимущество.

- Тема «Цена/Выгода» имеет 47% позитивных отзывов
- Конкурентное преимущество, которое нужно усиливать

Качество сервиса.

- Персонал и общение оцениваются положительно (51% позитива)

Команда Проекта

- Берсанов Тимур
- Забунков Леонид
- Бессонов Александр
- Злотин Ефим