

СИРИУС.ИИ



# Tinkoff AFASstt

A large yellow lightning bolt graphic is positioned between the two main sections of text. It originates from the bottom of the 'A' in 'AFASstt' and points upwards towards the top of the 'T' in 'Tinkoff'. The bolt has a bright yellow core and a darker yellow base.

AUTOMATED FEEDBACK ANALYSIS SYSTEM FOR TINKOFF

ИНСТРУМЕНТ ДЛЯ АНАЛИЗА КЛИЕНТСКИХ ОТЗЫВОВ

2024 ГОД



# ЗНАКОМЬТЕСЬ С КОМАНДОЙ

КАПИТАН



**Воропаев Ярослав  
Александрович**

Орловская область,  
15 лет,  
9 класс

УЧАСТНИК



**Логинова Светлана  
Ильинична**

Орловская область,  
15 лет,  
9 класс

УЧАСТНИК



**Борисенко Даниил  
Вячеславович**

Алтайский край,  
15 лет,  
9 класс

УЧАСТНИК



**Ясько  
Полина  
Юрьевна**

Алтайский край,  
16 лет,  
10 класс

# АКТУАЛЬНОСТЬ

**ОТЗЫВЫ** - это важная часть любого продукта, которая помогает в его улучшении и доработке. Важно уметь их анализировать, выделять положительные и отрицательные, а также выявлять достоинства и недостатки продукта. Стоит обратить внимание на дату написания отзыва, так как он может быть связан с временными проблемами, которые уже были решены. Поэтому важно не только фиксировать отзывы, но и анализировать их динамику и тренды, чтобы понять общую картину и выявить наиболее значимые аспекты, требующие внимания.

## Приятные впечатления

Нижний Новгород



Добрый день! Пришел по ссылке товарища и мне бесплатное обслуживание дали. Удивило что тут все переводы между тинькофскими клиентами без комиссий без разницы с какого города. Еще удивило что в Кыргызстан отправляю деньги племеннику по номеру телефона без комиссии 5-10 тыс руб обычно в мес. Удобно! Вообще приложение, картхолдер, сама карта, все приятно и современно. Думаю этот банк и я — это надолго. В чем подвох дебетовой карты тинькофф не знаю, мой отзыв такой что все устраивает.

[Оформить дебетовую карту](#)

## Пример отзыва

# ПРОБЛЕМА

К сожалению, проводить анализ вручную занимает очень много времени и сил, поэтому этот процесс необходимо оптимизировать. Отзывы бывают самыми разными - просто оценки или большие тексты, содержательные или нет. Вдобавок, они могут находиться на разных сайтах.



TINKOFF  
OFF TINKOFF  
TINKOFF  
TIN



# ЦЕЛЬ ПРОЕКТА:

Разработать инструмент, который **анализирует данные** из открытых источников и **помогает сотрудникам** Тинькофф улучшить клиентский опыт

# АНАЛИЗ



**1) С ПОМОЩЬЮ  
ЧЕГО**

**АНАЛИЗИРУЮТ  
ОТЗЫВЫ**

**2) КАК  
СОРТИРУЮТ И  
АНАЛИЗИРУЮТ  
ОТЗЫВЫ**

## HINTS

- 1) CHATGPT
- 2) СОРТИРОВКА ПО КАТЕГОРИЯМ, АНАЛИЗ СМЫСЛА И ОЦЕНКА ЭМОЦИОНАЛЬНОЙ ОКРАСКИ

## ROCKETDATA

- 1) МАШИННОЕ ОБУЧЕНИЕ
- 2) СОРТИРОВКА ПО КАТЕГОРИЯМ И РЕЙТИНГУ, АНАЛИЗ СМЫСЛА И ОЦЕНКА ЭМОЦИОНАЛЬНОЙ ОКРАСКИ

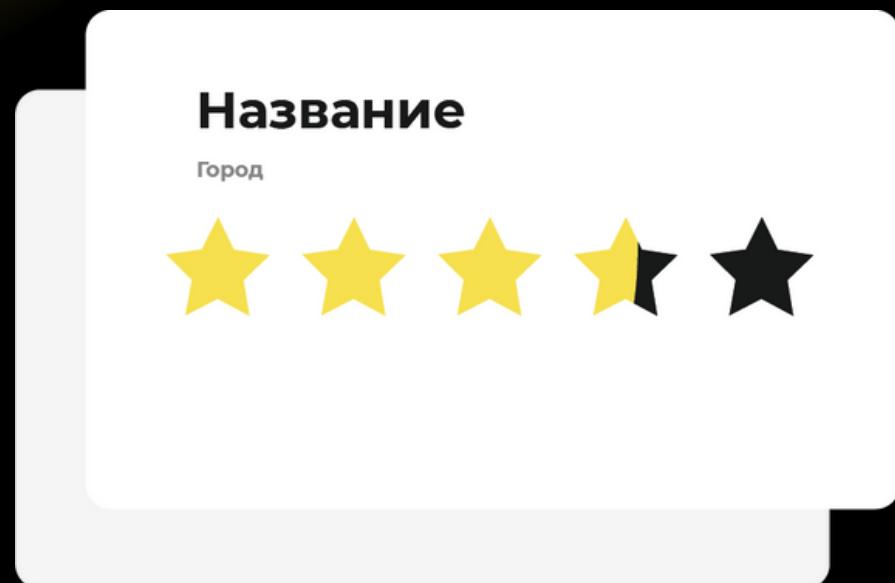
## МАЙКРОСОФТ

- 1) МАШИННОЕ ОБУЧЕНИЕ
- 2) ОЦЕНКА ЭМОЦИОНАЛЬНОЙ ОКРАСКИ, АНАЛИЗ СМЫСЛА И КЛАССИФИКАЦИЯ ПРОБЛЕМЫ

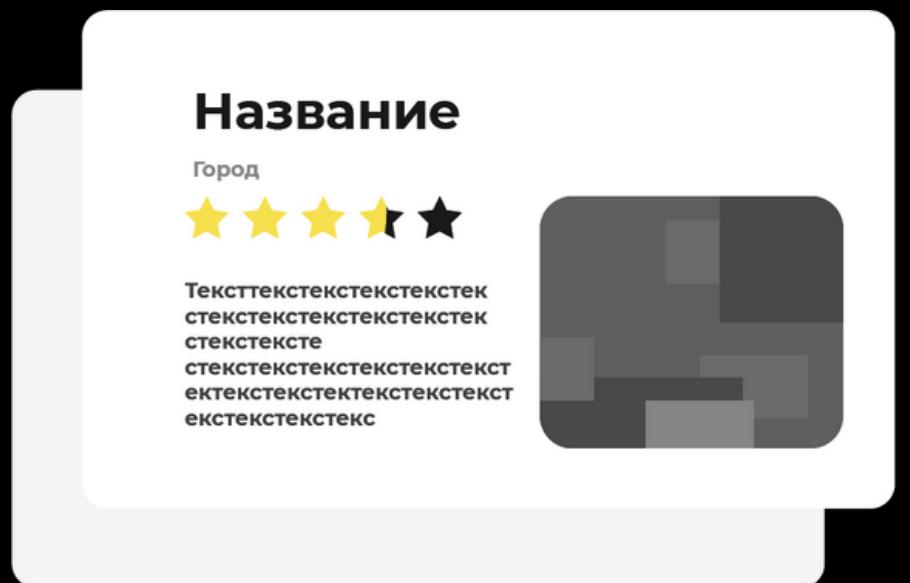
# ВИДЫ ОТЗЫВОВ

Анализ области

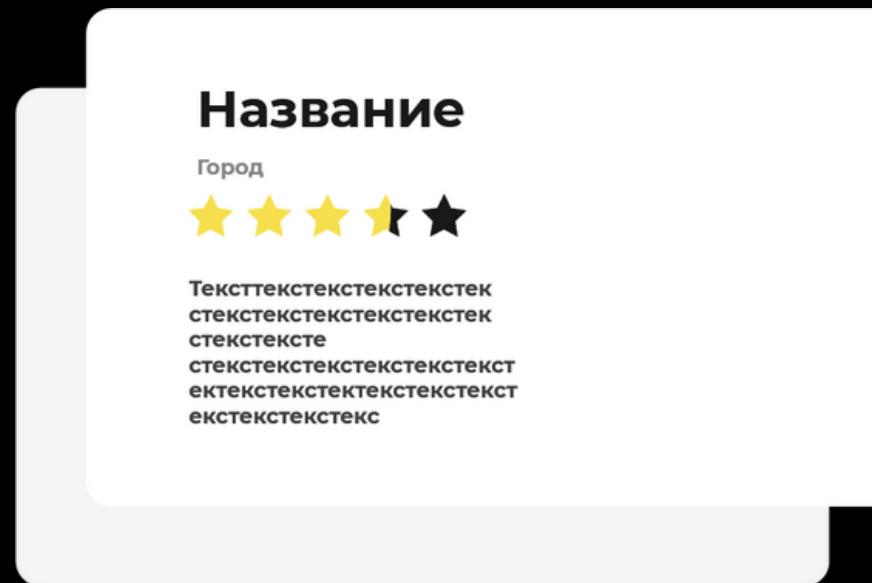
1. Только оценка



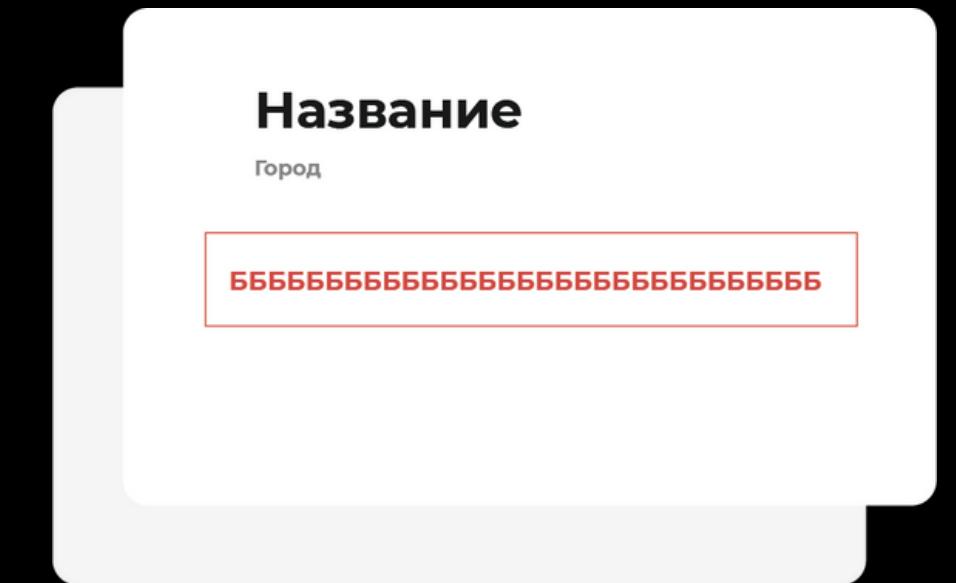
3. Оценка + текст + фото/видео



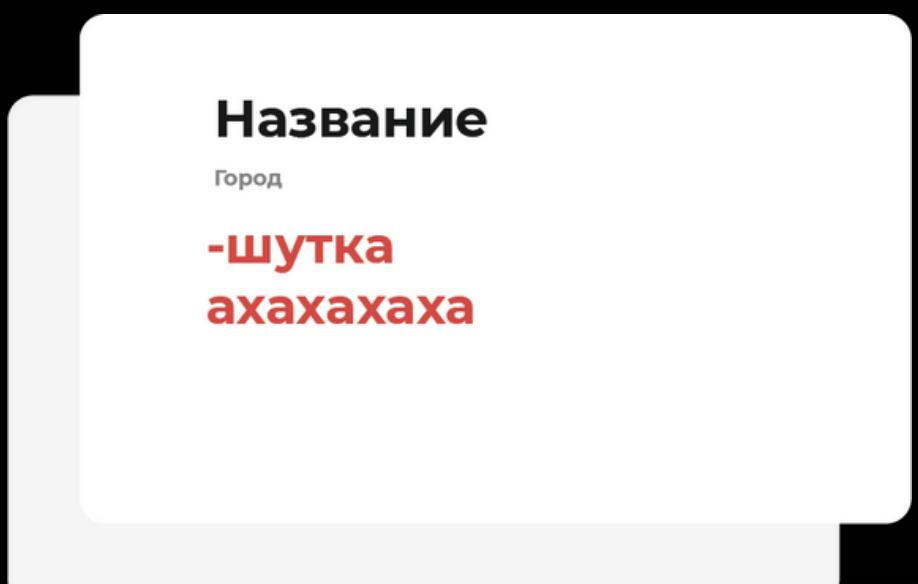
2. Оценка + текст



4. Неконструктивные отзывы



5. Шуточные отзывы



# ЧТО МОЖНО ВЫДЕЛИТЬ ИЗ ОТЗЫВОВ?

Отзывы можно анализировать с помощью LLM

Выводы, которые мы можем составить на основе ответа нейросети:

- 1) "Реальная" оценка\*
- 2) Средняя "реальная" оценка
- 3) Плюсы и минусы продукта, которые встречаются чаще всего

Причем мы можем посмотреть тренды первых двух пунктов не только за все время, но и за промежуток времени, который выберет специалист.

\*оценка, определенная нейросетью на основе отзыва

# АНАЛИЗ ДАННЫХ И ИСПОЛЬЗОВАНИЕ ПРИЛОЖЕНИЯ

**Большинство компаний при анализе отзывов используют следующий алгоритм:**

1. ИИ приводит отзыв к стандартному виду (обрабатывает их, чтобы удалить лишние символы, разделить на отдельные предложения, слова, фразы)
2. С помощью нейронных сетей анализируется тональность
3. Извлекаются наиболее частые фразы для выделения основных тем
5. Группировка данных по темам обсуждений и продуктам
6. Распределение отзыва в ту или иную тему
7. Определение главной темы отзыва
8. Ответ на отзыв

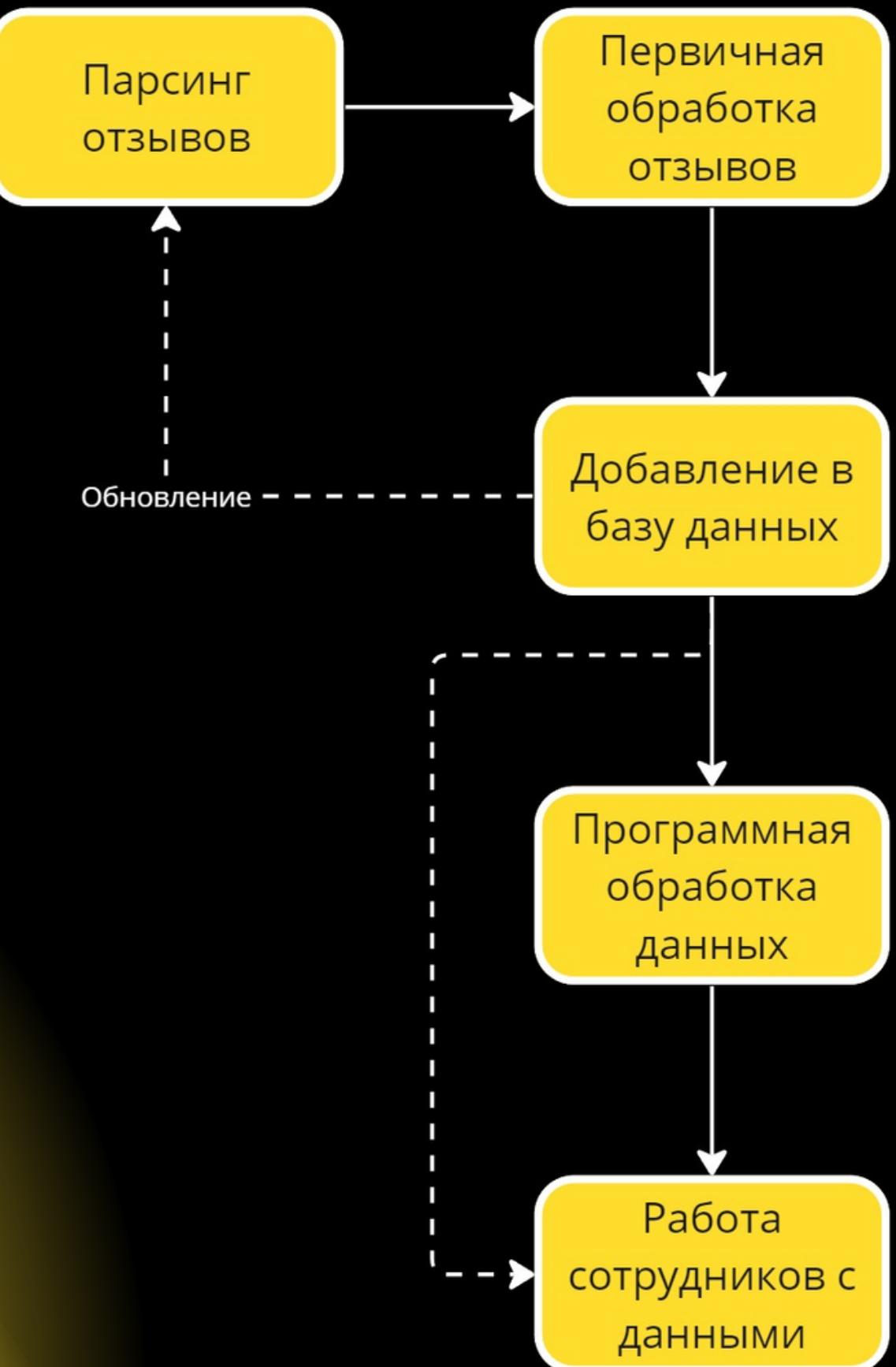
**По итогам анализа было решено добавить большее количество полезных функций:** сортировка отзывов по тональности, по времени, по оценке и возможность определить ключевые минусы и плюсы продукта. Также в нашей программе можно будет отследить динамику изменения минусов, плюсов и оценок отзывов за определенный промежуток времени, что очень поможет как самой компании, так и ее клиентам.

# НАШЕ РЕШЕНИЕ ПРОБЛЕМЫ- СОЗДАНИЕ САЙТА



Использование современных  
экономичных технологий

Сбор из разных источников,  
умное распределение отзывов  
по категориям, анализ  
динамики, определение  
тональности



# ОПИСАНИЕ ПРЕДЛАГАЕМОГО РЕШЕНИЯ

Программа собирает данные с нескольких сервисов с отзывами, затем отзывы проходят первичную обработку, оцениваются нейросетью и добавляются в базу данных.

Раз в некоторое время или по запросу сотрудника запускается скрипт, который будет добавлять новые обработанные отзывы в БД.

# МЕТОДЫ РАЗРАБОТКИ

LLM, предпочтительно русскоязычная (например, Yandex GPT)

Парсинг с помощью API и Python-скрипта

База данных S3 Object Storage

Сайт на React.js для взаимодействия с сервисом

# ВЫСОКОУРОВНЕВЫЙ ПЛАН РЕАЛИЗАЦИИ РЕШЕНИЯ

Сбор данных осуществляется **парсингом** отзывов со специальных сервисов. После этого из отзывов удаляются **стоп-слова и ненужные символы**, распознаются шуточные отзывы. Затем отзыв оценивается **нейросетью**. Все обработанные отзывы попадают в **БД**, которая обновляется раз в некоторое время или по запросу сотрудника.

После этого **аналитики могут как воспользоваться нашим приложением** (посмотреть плюсы и минусы, среднюю оценку за определённый период, динамику и причины её изменения, отфильтровать отзывы по конкретным минусам и тд), так и обработать "сырую" БД.

# ОПИСАНИЕ ОТДЕЛЬНЫХ ШАГОВ ПЛАНА

- Парсинг отзывов
- Предварительная обработка отзывов
- Распознавание шуточных отзывов
- Выявление плюсов и минусов продукта
- Оценка отзывов
- Формирование/обновление БД
- Анализ данных и использование приложения

# ПАРСИНГ ОТЗЫВОВ

Отзывы загружаются со специализированных сервисов: Тинькофф Отзывы, [banki.ru](#), [sravni.ru](#), [otzovik.ru](#) и тд. Возможен также сбор отзывов с соц.сетей и форумов. Это возможно при помощи использования **API сервисов или парсинга HTML-страниц**, например через библиотеку BeautifulSoup для Python

# ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ОТЗЫВОВ

После сбора отзывов происходит их **предварительная обработка**, включающая удаление стоп-слов (часто употребляемых слов, которые не несут смысловой нагрузки) и ненужных символов (например, пунктуации). Также происходит нормализация текста, такая как приведение всех слов к нижнему регистру.

# РАСПОЗНАВАНИЕ ШУТОЧНЫХ ОТЗЫВОВ

Некоторые отзывы могут содержать шутливые или несерьезные комментарии, которые не отражают реальное мнение о продукте или услуге. Такие отзывы не несут полезной информации и портят реальную оценку. **Эти отзывы помечаются в БД и могут быть использованы для более чистого анализа.**

# ВЫЯВЛЕНИЕ ПЛЮСОВ И МИНУСОВ ПРОДУКТА

Перед тем, как данные попадут в базу данных, происходит выделение ключевых аспектов или признаков, определяющих плюсы и минусы продукта. Это будет происходить на **основе анализа содержания отзывов**, выявления наиболее упоминаемых положительных ("красивый дизайн", "высокий кешбэк") и отрицательных ("плохая поддержка", "нестабильная работа приложения") аспектов продукта или услуги.



# ОЦЕНКА ОТЗЫВОВ

После **предварительной обработки и выявления плюсов и минусов** (для повышения точности) данные подвергаются оценке с помощью LLM. Нейросеть может быть обучена на размеченном наборе данных, чтобы присваивать оценку (например, от 1 до 5) каждому отзыву в зависимости от тональности и содержания. Это нужно для выделения **реальной оценки**, при этом оригинальные оценки также входят в базу данных.

# ФОРМИРОВАНИЕ/ ОБНОВЛЕНИЕ БД

База данных обновляется раз в некоторое время, чтобы добавить новые отзывы и обновить существующие. Это также происходит по запросу сотрудника. Хорошим вариантом БД может стать S3 хранилище.

Формат БД:

Отзывы	Оценка	Оценка нейросети	Плюсы	Минусы	Дата	Шуточный или нет

# АНАЛИЗ ДАННЫХ И ИСПОЛЬЗОВАНИЕ ПРИЛОЖЕНИЯ

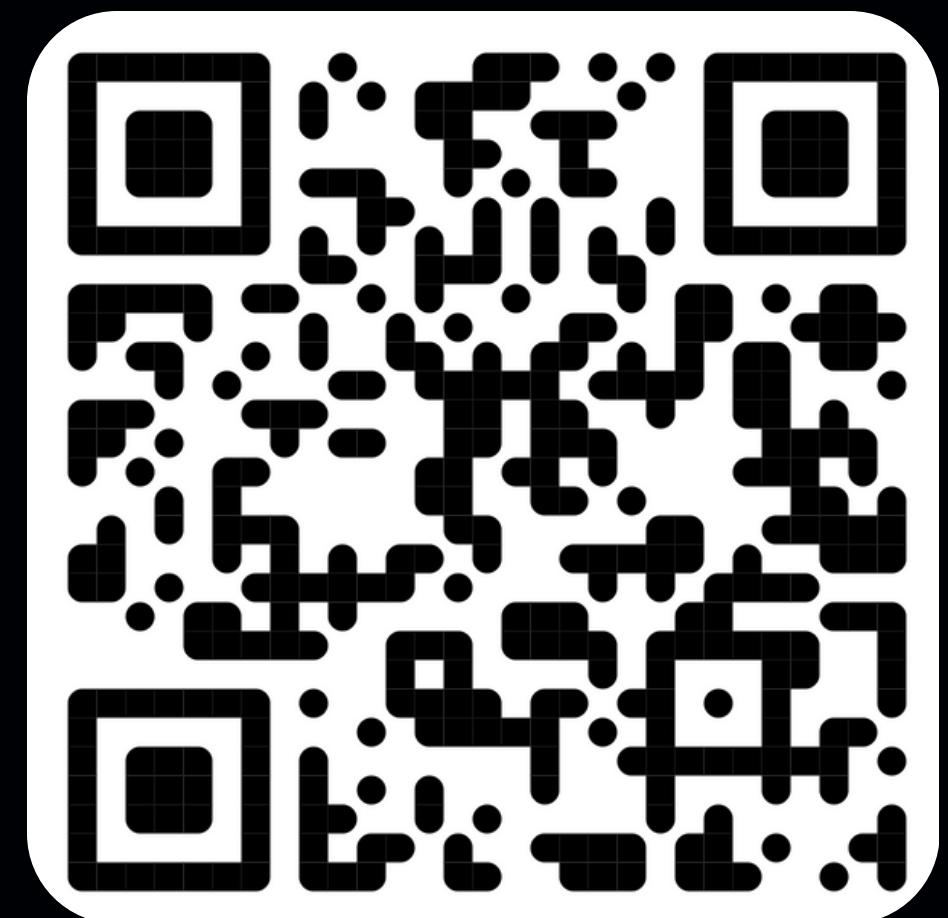
После того как данные обработаны и сохранены в БД, аналитики могут использовать различные инструменты для анализа данных: сортировка, фильтрация по различным значениям, построение графиков и диаграмм. Помимо обработки данных напрямую, аналитики и другие пользователи могут использовать специальное приложение, которое позволяет просматривать и анализировать отзывы.

## ФУНКЦИИ ПРИЛОЖЕНИЯ:

- большинство методов классического анализа,
- вычисление средних значений за определенный период времени,
- анализ динамики изменения оценок и выявление причин этих изменений,
- возможность отфильтровать отзывы по конкретным плюсам и минусам.

# ИТОГИ РАБОТЫ:

<https://github.com/zabojeb/tinkoff-feedback-analysis>



# ВЫБИРАЙТЕ НАС

- Сбор отзывов с разных платформ
- Выявление потребностей клиентов
- Отслеживание динамики
- Возможность выявления причин изменения оценки
- Выявление "реальной" оценки