

Выявление трендов в сфере бухгалтерского учета, поиск «болей» бухгалтера

Кейсодержатель

<https://buhexpert8.ru>

Организатор

<https://фиц2024.рф/>

Команда

Бухэксперт

ФИЦ

ФОРУМ
ИННОВАЦИОННЫХ
ЦЕНТРОВ

Скрипка



Задачи



01

Алгоритм поиска и выявления трендов

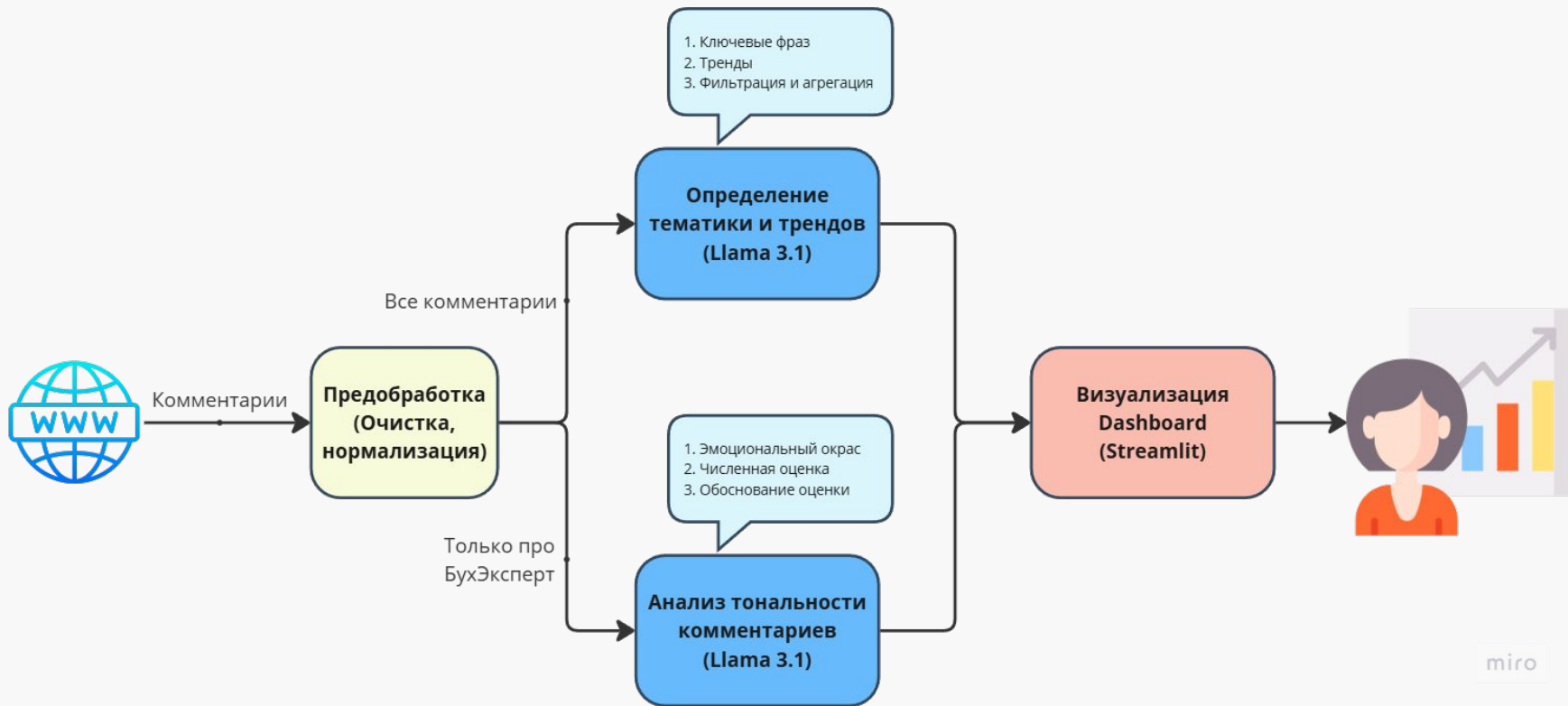
02

Анализ мнений о компании «БухЭксперт8»: положительно, нейтрально, отрицательно

03

Визуализация результатов анализа (дашборд)

Архитектура решения



miro

Исходные данные

Источники данных

Комментарии из 12 популярных форумов и чатов бухгалтеров за 2024 год

Объем данных

323,603 комментария, из них:

- **29 тыс.** осталось после фильтрации
- **164 шт.** связаны с сервисом "БухЭксперт8"



Предварительная обработка



- Исключение коротких сообщений
- Удаление дубликатов
- Очистка текста от спецсимволов
- Отбор по ключевым словам: “подскажите”, “вопрос”, “вопросы”
- Выделение комментариев о БухЭксперт8: расстояние Левенштейна до “бухэксперт”, “buhexpert”

Алгоритм поиска и выявления трендов



Извлечение тем

Выделение ключевых тем по бухгалтерии, налогам и другим аспектам бизнеса.
Модель Llama-3.1-8B (квантование 4-бит)

Категоризация

Классификация комментариев (с использованием извлеченных тем) на заранее определенные категории.
Модель Llama-3.1-8B (квантование 8-бит)

Алгоритм поиска и выявления трендов

Классификация комментариев по категориям:

- Налоги, налогообложение и налоговое планирование
- Бухгалтерский учет и финансовая отчетность
- Внутренний контроль и аудит
- Финансовое планирование и прогнозирование
- Учет заработной платы и кадровое делопроизводство
- Программное обеспечение для бухгалтерии и автоматизация
- Взаимодействие с государственными органами
- Правовое регулирование бухгалтерии, налогообложения и трудового законодательства



Анализ мнений о компании «БухЭксперт8»

Особенности комментариев



- 01** Комментарий не информативен, например просто ссылка на сервис
- 02** Негатив автора к оппоненту, а не к сервису
- 03** Комментарии без контекста (ответ на другой комментарий)
- 04** Названия статей содержат негативный окрас, например: “Малоценное оборудование и запасы”.

Выбор подхода к анализу эмоций



Готовые LLM

Комментарию ставится негатив, даже если статья содержит негатив на оппонента, а сервис БухЭксперт8 используется как доказательная база



Средние эмбединги

Сходство комментария к категории (positive, neutral, negative), но на внешних данных пользователи редко оставляют комментарии про конкретную работу сервиса



Модель на базе Llama-3.1-8B

Определялся процент окраса (0 - негатив, 1 - позитив), а также причина почему принято решение, что комментарий позитивный или негативный

Результаты анализа комментариев

**Классификация
эмоций из 164:**

Положительные: 139

Нейтральные: 10

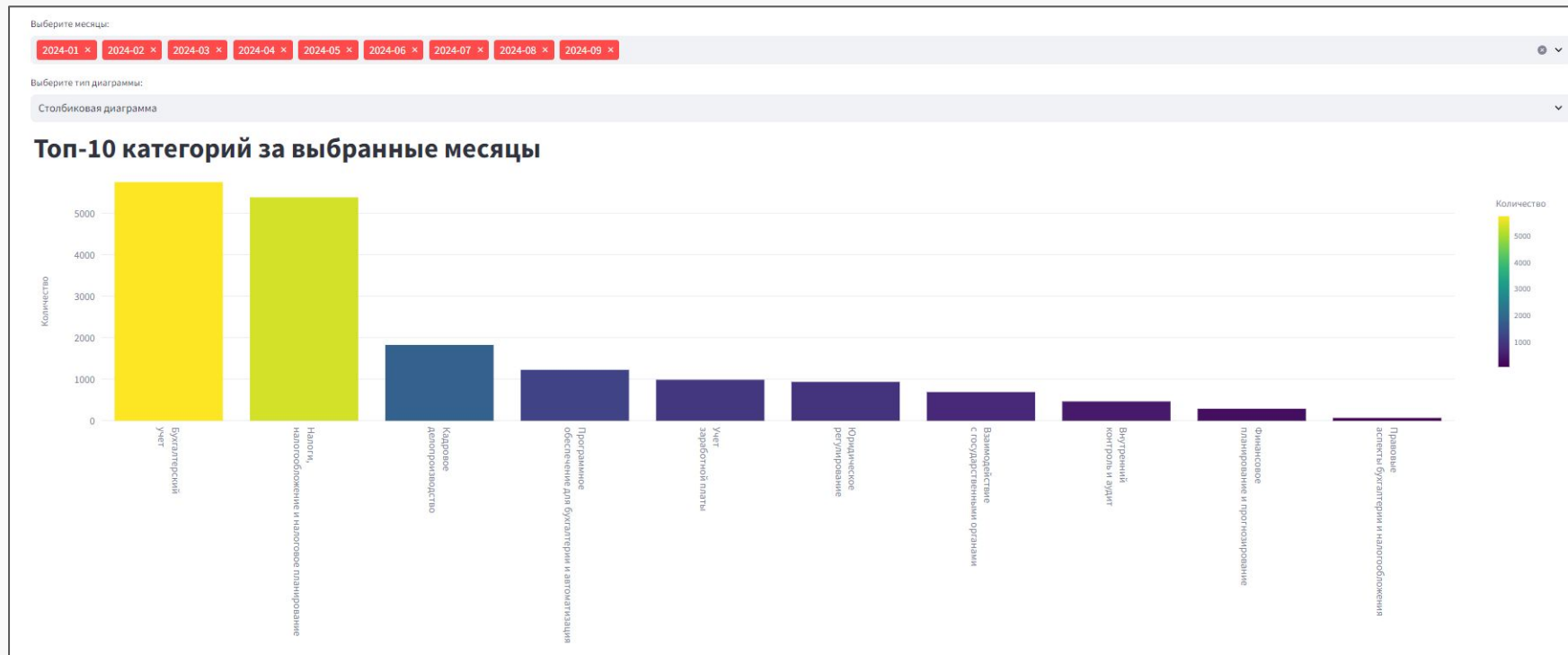
Негативные: 15

Примеры негативных комментариев:

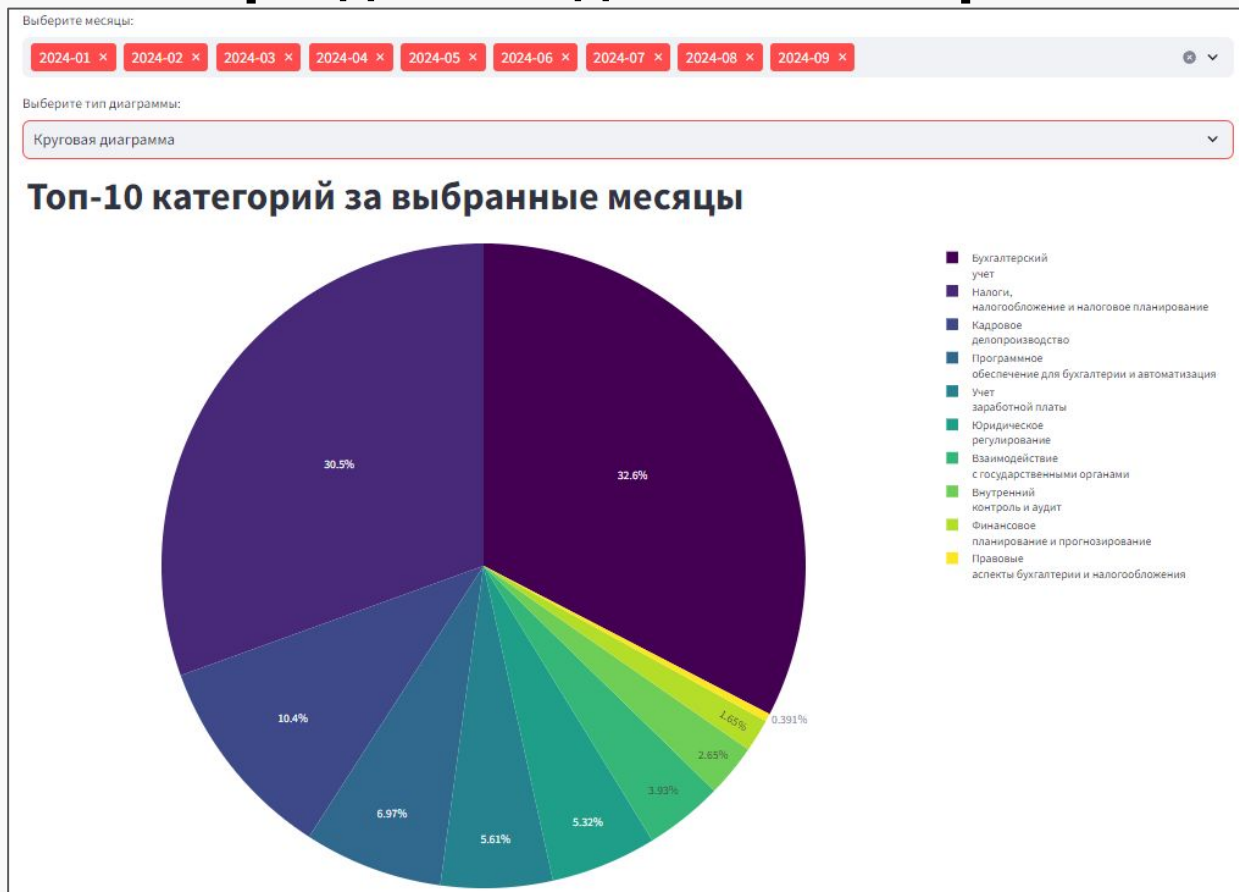
- 'не нашла подходящего ответа на Бухэксперте',
- 'Есть статья в БухЭксперт. Но ее можно прочитать **при подписке**',
- 'Добрый день. Хочу на 6 месяцев на бухэксперт 8 подписаться, **для нас дорого**. Есть желающие 50 на 50 поделить сумму?',
- 'Хочу добавить по поводу зарабатывания на записях, сейчас зашла на Бухэксперт, там запись вебинара **для не подписчиков 9900Р**'

Дашборд для визуализации результатов

Для визуализации результатов разработан интерактивный дашборд (с использованием streamlit). Ссылка: <https://skripka.ru.tuna.am/>



Тренды по задаваемым вопросам

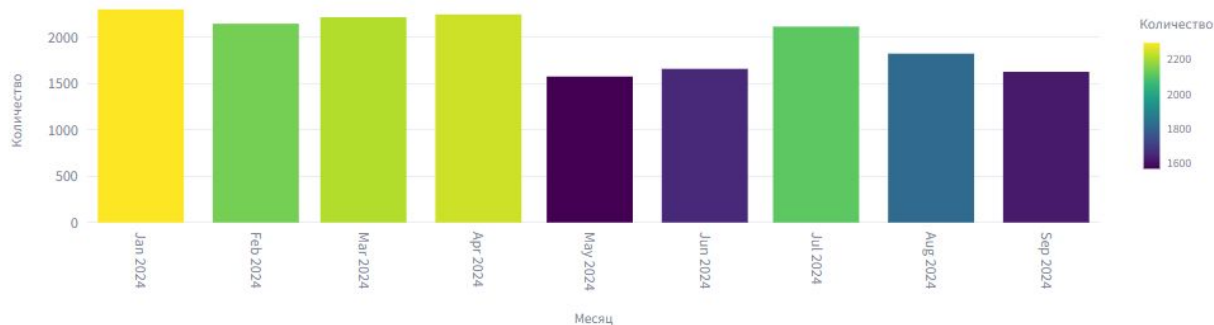


Анализ вопросов по месяцам

Выберите вид графика:

Отдельные графики

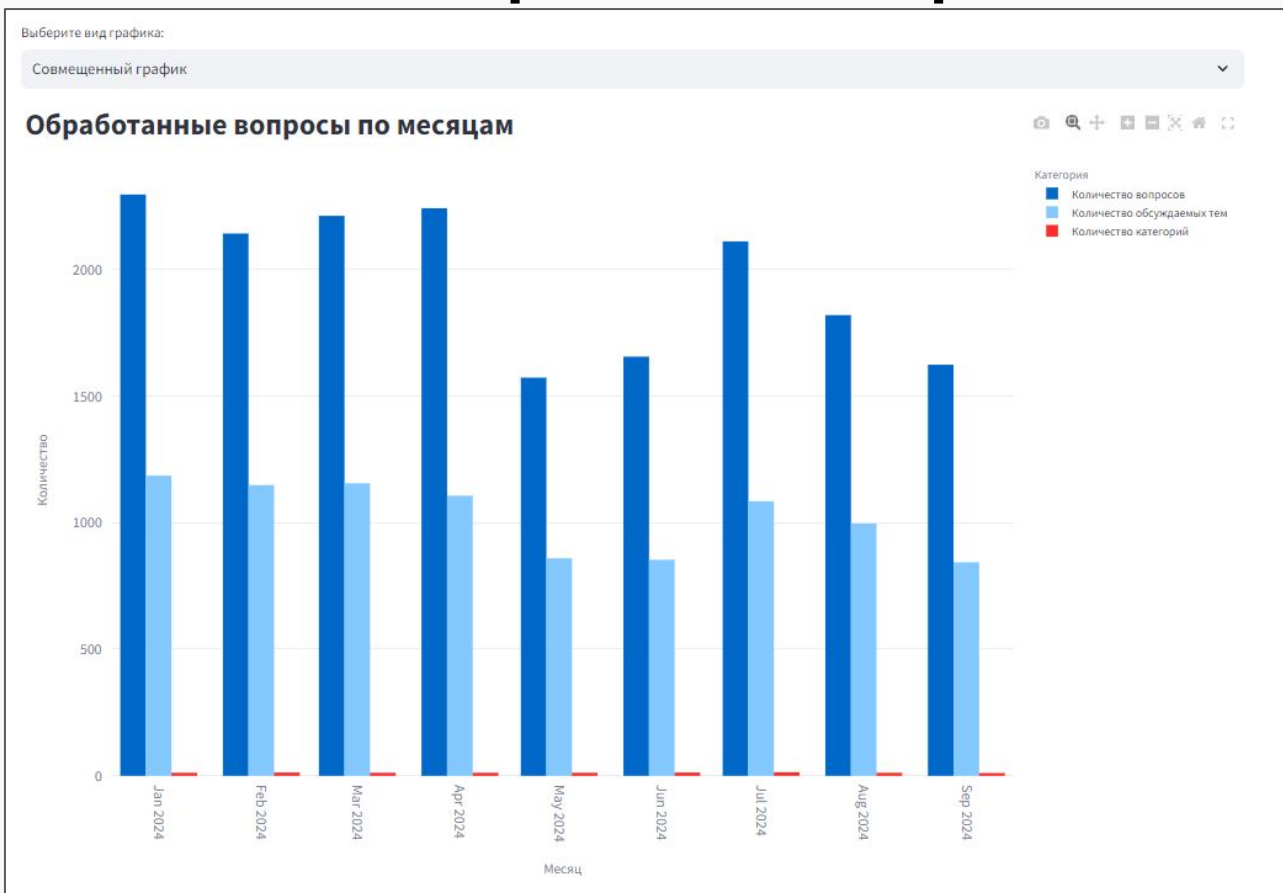
Количество вопросов по месяцам



Количество обсуждаемых тем по месяцам



Анализ обработанных вопросов

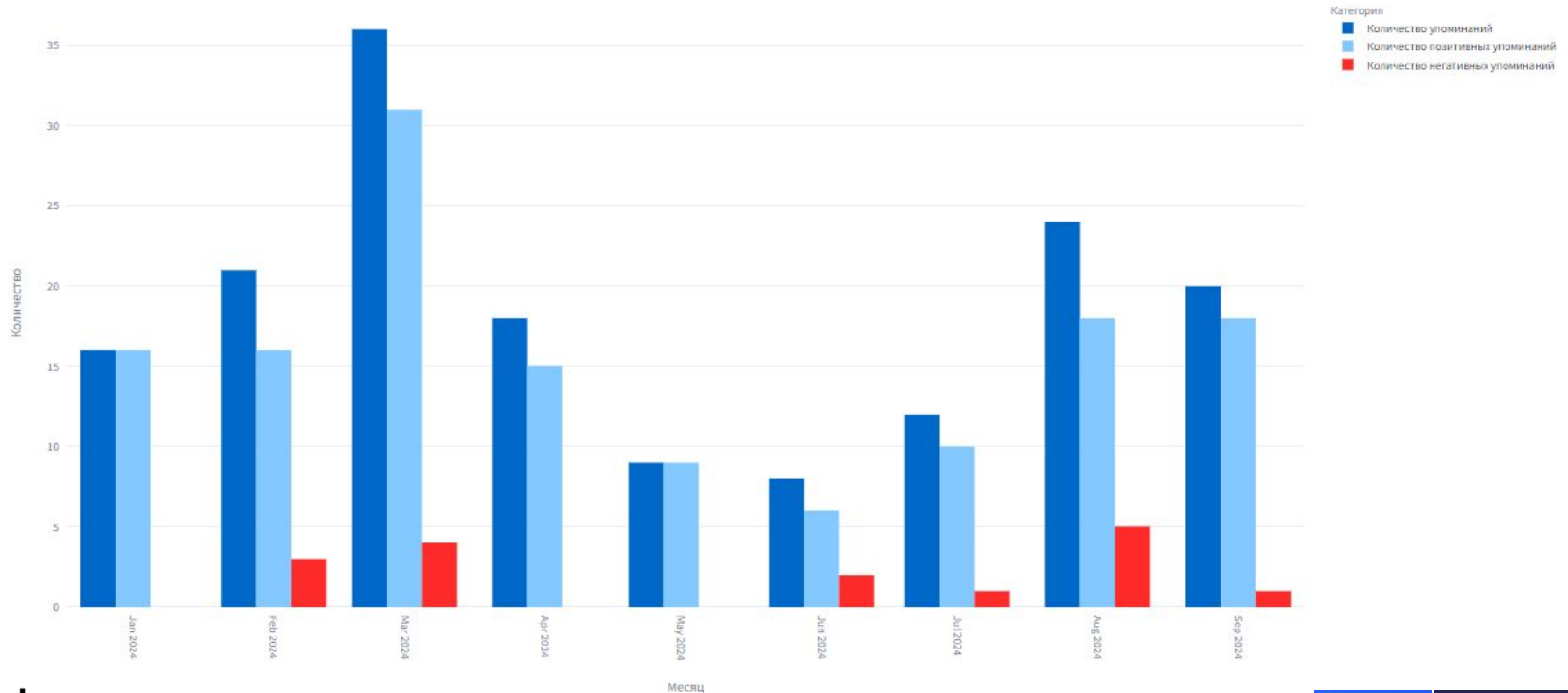


Анализ упоминаний о "БухЭксперт8"

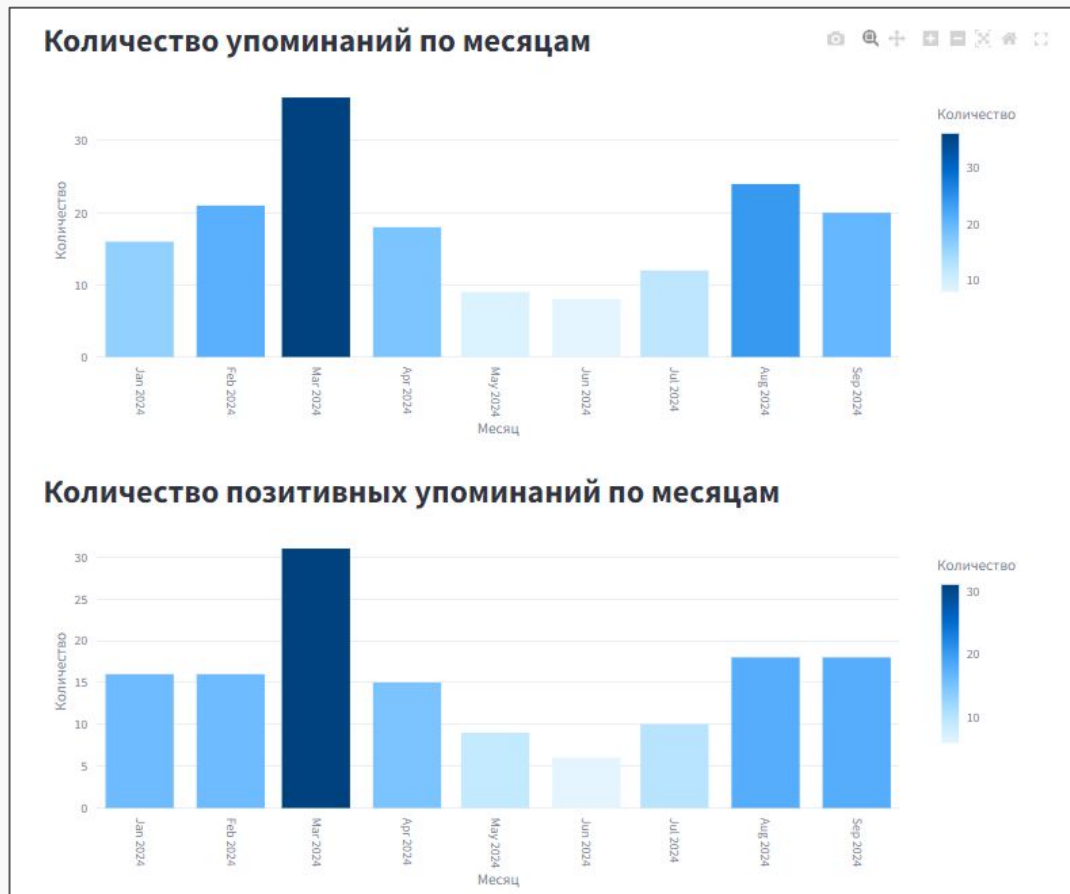
Выберите вид графика:

Совмещенный график

Обработанные эмоции по месяцам



Анализ упоминаний о "БухЭксперт8"



Детализация упоминаний о "БухЭксперт8"

	id	date	user_id	text	message	contains	emotion	emotion_score	reason
0	4,835	2024-01-12	128	Это с сайта бухэксперт приведен пример	message	<input checked="" type="checkbox"/>	positive	0.75	Пользователь ссылается на сервис бухэксперт как на источник достоверной информации, что считается хорошим примером для сервиса.
1	10,826	2024-01-19	31	Бухэксперт - подробно целая тема...)))	message	<input checked="" type="checkbox"/>	positive	0.7	Пользователь ссылается на бухгалтерский эксперта как на источник достоверной информации, что говорит о позитивном отношении к сервису.
2	10,884	2024-01-19	547	В Бухэксперте есть	message	<input checked="" type="checkbox"/>	positive	0.5	Пользователь ссылается на сервис Buhexpert как источник информации, но не выражает особую эмоцию по отношению к сервису. Поэтому классифицируется как нейтральное.
3	10,938	2024-01-20	13	Ну есть стаж, а там о выплатах фсс инфа, это что-то новое...я в пн буду искать еще	message	<input checked="" type="checkbox"/>	positive	0.5	Пользователь не доволен именно сервисом, но ссылается на него как на источник информации, делая комментарий положительным.
4	11,172	2024-01-20	13	Если вам это поможет https://buhexpert8.ru/1s-buhgalteriya/posredniceskie-operezhdeniya	message	<input type="checkbox"/>	positive	0.7	Пользователь ссылается на сервис бухэксперт 8 как на полезный источник информации, что считается положительным отзывом.
5	12,340	2024-01-22	13	https://buhexpert8.ru/1s-zup/etk-i-uchet-dlya-pfr/personifitsirovannyj-uchet/podrazbitye	message	<input type="checkbox"/>	positive	0.7	Пользователь ссылается на сервис Buhexpert 8 как на источник достоверной информации, что делает комментарий положительным. Хотя он не выражает особой эмоциональной оценки по сравнению с консультацией.
6	12,359	2024-01-22	13	Вот я ж написала, например в консультанте я про этот отчет информации еще не	message	<input checked="" type="checkbox"/>	positive	0.5	Пользователь ссылается на сервис как на источник информации, но не выражает особой эмоциональной оценки по сравнению с консультацией.
7	13,271	2024-01-24	249	Посмотрите здесь. В примере как раз превышение по итогу 4 кв https://buhexpert8.ru/voprosy/voprosy-1s-buhgalteriya/usn-dohody-rashody.html	message	<input type="checkbox"/>	positive	0.7	Пользователь ссылается на сервис как на источник информации, несмотря на то что выражение его мнения является негативным. Это считается нейтральным.
8	14,933	2024-01-26	13	https://buhexpert8.ru/voprosy/voprosy-1s-buhgalteriya/usn-dohody-rashody.html	message	<input type="checkbox"/>	positive	0.7	Пользователь ссылается на сервис бухэксперт 8 как на источник информации, что считается хорошим примером для сервиса. Хотя комментарий не выражает особой эмоциональной оценки по сравнению с консультацией.
9	15,906	2024-01-29	249	Возможно нужно открыть период, НДС все равно будет корректироваться по да	message	<input type="checkbox"/>	positive	0.5	Пользователь ссылается на сервис как на источник информации, но высказывается нейтрально о нем. Основная причина - сервис упоминается в контексте консультации.
10	17,999	2024-02-02	1,505	Отображаем удержанный НДФЛ, с помощью установленного патча в зуп. будут	message	<input type="checkbox"/>	positive	0.5	Пользователь ссылается на сервис как на источник информации, но выражение его мнения нейтральное. Поэтому классифицируется как нейтральное.
11	19,359	2024-02-06	31	Фразу " как оформить в Бухгалтерии предприятия возврат сотрудником излиш	message	<input checked="" type="checkbox"/>	positive	0.6	Пользователь ссылается на рекомендации бухгалтерского эксперта как на источник полезной информации, хотя сам комментарий не связан напрямую с сервисом.
12	19,360	2024-02-06	31	Для особо ленивых - https://buhexpert8.ru/1s-buhgalteriya/kadry-i-zarabotnaya-plata	message	<input type="checkbox"/>	positive	0.5	Пользователь упоминает бухэксперт8 как источник информации, но не выражает особую эмоцию по отношению к сервису. Это делает комментарий нейтральным.
13	19,369	2024-02-06	75	Доброе! По кассам не подскажу, но скорее всего да, нужно. А по освобождению	message	<input type="checkbox"/>	positive	0.5	Пользователь ссылается на сервис бухэксперт 8 как источник информации, но не выражает явное удовлетворение или недовольство. Поэтому классифицируется как нейтральное.
14	21,974	2024-02-14	31	А на фую такие деньги тратить? Что там есть? Не работать - сервисы читать? Воо	message	<input checked="" type="checkbox"/>	negative	0.8	Дороговизна сервиса
15	24,328	2024-02-20	1,308	Бухэксперт говорит что есть	message	<input checked="" type="checkbox"/>	positive	0.5	Пользователь ссылается на мнение бухгалтера как на достоверную информацию, но комментарий не содержит деталей о том, насколько хорошо сервис справляется с задачей.
16	25,224	2024-02-22	31	Итс в помощь.. Бух. эксперт. Гугл. Информации море. По шагам. С картинками. Л	message	<input type="checkbox"/>	positive	0.7	Пользователь ссылается на сервис как на источник информации, что делает комментарий положительным. Хотя он упоминает, что это платно.
17	28,058	2024-02-29	833	Татьяна, здесь всё-всё, что нужно, как настроить учет командировочных в 1с бу	message	<input type="checkbox"/>	positive	0.75	Пользователь ссылается на сервис как на источник полезной информации для настройки командировочных в 1с бухгалтерии, что считается хорошим отзывом.
18	28,089	2024-02-29	1,649	Вот тут посмотрите, по ссылке пройдите, там написано как создать начисление	message	<input type="checkbox"/>	positive	0.6	Пользователь ссылается на сервис бухэксперт 8 как на источник полезной информации, хотя высказывается о нем негативно, указывая, что информация не помогла решить проблему.
19	28,098	2024-02-29	31	Ну вот смотрит она в программу... И что?? Пока результатом не обрадовала. А в	message	<input checked="" type="checkbox"/>	negative	0.5	Пользователь не доволен результатом сервиса, хотя и ссылается на него как на источник информации. Причина недовольства - ожидание результата.



Инструменты

LLAMA - 3.1



Streamlit



HUGGING FACE

spaCy



Transformers

Скрипка

ФИЦ

Бухэксперт

Команда Скрипка



Капитан
DS developer


Кожемяко
Дмитрий

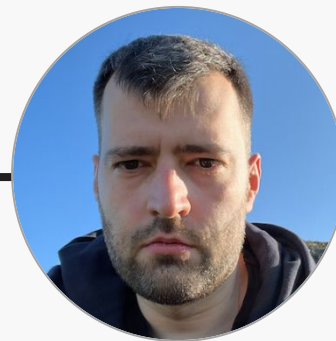
 @Dimk_88



DS developer

Черных
Иван

 @iceman_o_0




Fullstack developer
+ Analyst

Борисовский
Сергей

 @CodeDrivenAI

СПАСИБО!

по всем вопросам обращайтесь к капитану команды
Дмитрий Кожемяко  @Dimk_88



GitHub

<https://github.com/DimkKozhem/HACK>

Сервис

<https://skripka.ru.tuna.am/>

Документация *

<https://docs.google.com/document/d/1p5WZpnr5TvgkfSmF9k3JOKWfQHqqDNbWfoD-BI2ZtcA/edit?tab=t.0>