

АНАЛИЗ РЕКЛАМНЫХ УПОМИНАНИЙ БРЕНДОВ В РОССИЙСКОМ YOUTUBE

Велесевич Надежда

НИУ ВШЭ ДПО

Компьютерная лингвистика

СОДЕРЖАНИЕ

- Актуальность проекта
- Цели проекта
- Задачи проекта
- Сбор корпуса
- Алгоритм поиска брендов
- Первые итоги
- Углубленный анализ полученных данных и их визуальное представление
- Результаты
- Инструменты
- Трудности
- Выводы

АКТУАЛЬНОСТЬ ПРОЕКТА

Рост влияния YouTube

- Увеличение роли блогеров в формировании потребительского поведения
- Рост инвестиций брендов в YouTube-маркетинг

Необходимость автоматизации

- Ручной анализ контента требует больших временных затрат
- Сложность отслеживания упоминаний брендов в большом объеме контента

Бизнес-потребности

- Необходимость оценки эффективности рекламных кампаний
- Важность мониторинга конкурентов

ЦЕЛИ ПРОЕКТА

- Анализ присутствия и упоминания брендов в русскоязычном YouTube-контенте
- Выявление паттернов рекламных интеграций
- Создание автоматизированной системы мониторинга упоминаний брендов
- Разработка инструментов анализа контента на русском языке

ЗАДАЧИ ПРОЕКТА

Сбор данных

- Получение видео с популярных российских YouTube-каналов
- Извлечение субтитров и транскриптов
- Сохранение метаданных (название, дата публикации, канал)

Обработка и анализ

- Обработка русскоязычных текстов
- Распознавание брендов с учетом различных вариаций написания
- Определение контекста упоминаний
- Выявление рекламных паттернов

Визуализация результатов

- Создание наглядных графиков и диаграмм
- Визуализация трендов и паттернов

СБОР КОРПУСА

Для анализа были выбраны 10 популярных русскоязычных YouTube-каналов различной направленности.

Критериями для отбора списка каналов были:

- Высокая популярность (большое количество подписчиков)
- Регулярная публикация контента
- Наличие рекламных интеграций

С каждого канала было отобрано по 10 последних опубликованных видео длительностью более 20 минут с доступными русскими субтитрами. Общий объем: 100 видео. Для сбора корпуса использовался YouTube API.

```
[9] # API ключ для доступа к YouTube Data API
API_KEY = 'AIzaSyCcnZo9KoApF5FrddR80vYq5T5V1vPXf_c' #наш youtube api
youtube = build('youtube', 'v3', developerKey=API_KEY) # создаем объект для работы с API
```

```
[10] # Словарь каналов и их ID
# Ключ - название канала, значение - его уникальный идентификатор на YouTube
CHANNELS = {
    'ЛИТВИН': 'UCy6p8krDY33mDRMt19Yr1Xg',
    'Маруся Чернишкина': 'UCMiWy2TJ3MckIwuzuyleRcw',
    'Exile Show': 'UC6JRrn_7Qe1CZBcQDMieadw',
    'Agatha Christie': 'UCRv1uDv7I2bv7aavxmkRTtw',
    'Калинкин!': 'UC7AfybgBq2g_c9C9Ynyru8Q',
    'парадеевич': 'UC8G2LhxAOI7n_lVQDUM6FVg',
    'FAMETIME TV': 'UCop3kn61J_3JhrYp7Jddkgw',
    'АЙДЕН': 'UCCPpKYz2F-73OuaMmLUvg1g',
    'Жекич Дубровский': 'UC9XJvt80T-9_8QHDBdqocaw',
    'ТОПЛЕС': 'UC2Ru64PHqW4FxoP0xhQRvJg'
}
```


Получаем 10 последних видео для канала: ЛИТВИН

Добавлено видео: Охота На Звезд 2 ! Литвин и Равшан vs Сэм и Егорик !

Добавлено видео: Кто Последний Уснёт Забирает 1 000 000 !

Добавлено видео: Украли Машину за 10 Миллионов ! Наказали Вора !!

Добавлено видео: Прошлась Голой За 1 000 000 ! На Что Люди Готовы Ради Денег ?

Добавлено видео: Охота На Звезд ! Литвин и Равшан vs Венгалби и Стил !

Добавлено видео: Спаси Свой Подарок Чтобы Забрать Его! Главный Кошмар Венгалби !

Добавлено видео: Мне Сломал Ребро Боец UFC ! Литвин vs Венгалби vs Равшан !

Добавлено видео: Я Вышел Против Бойца UFC !! Литвин vs Венгалби vs Равшан !!

Добавлено видео: НА ЧТО ДЕВУШКИ ГОТОВЫ РАДИ ДЕНЕГ ?! ПОЛУЧИ 100к РУБЛЕЙ ЗА УДАР ПО ЛИЦУ !

Добавлено видео: ЛЮДИ vs ЛАМБОРГИНИ !! КТО СИЛЬНЕЕ ?! Ахмед vs Равшан vs Леон !!

Получаем 10 последних видео для канала: Маруся Черничкина

Добавлено видео: ТИХИЙ ПСИХОПАТ: Как заботливый соцработник стал массовым убийцей | Сатоши Уэмацу

Добавлено видео: ВСЕГДА — НЕВЕСТА, НИКОГДА — ЖЕНА | Страшные истории подписчиков с @abridge

Добавлено видео: ДОМИНА ЧЕРРИ. Девушка из Тинд

Добавлено видео: МОЯ ЖЕНА - УБИЙЦА | Эрик и Ро

Добавлено видео: Я НЕ ГОТОВ УЗНАТЬ, БЫЛО ЛИ ЭТО

Добавлено видео: ДЖОН. Подросток, который зака

Добавлено видео: Идеальный сосед оказался ОПАС

Добавлено видео: УЖАСНАЯ НОЧЬ НА ЗАПРАВКЕ | Ст

Добавлено видео: CRIME NEWS: Луиджи Манджоне -

Добавлено видео: БРИЛЛИАНТЫ НАВСЕГДА. Миллениу

Получаем 10 последних видео для канала: Exile :

Добавлено видео: Проверка Экстремальных Экспер

Добавлено видео: Кто Найдёт Больше Всего Дорог

Добавлено видео: 6 Комнат! Пройди Все Испытани

Добавлено видео: Пройди Все Испытания Чтобы Вы

Добавлено видео: Выполни Безумное Задание - По

Добавлено видео: Обидные Вопросы Хазяевам ! Чу

Добавлено видео: Угадай Долларового Миллионера

Добавлено видео: Открыли Мишлен Ресторан За 24

Добавлено видео: Проверка Лайфхаков, Экспериме

Добавлено видео: Ответь Чтобы Не Упасть С Вышк

Получаем 10 последних видео для канала: АЙДЕН

Добавлено видео: Матизы в аренду - проблемы не закончились. Сколько заработал на бизнесе Матизы, Гранты, Ларгус?

Добавлено видео: КОФЕЙНЯ МЕНЯ РАЗОРИЛА! Все ушло в ремонт! Сколько приносит кофейня в 2025?

Добавлено видео: Я дал 200к на бизнесы подписчиков. Сколько заработали за неделю в Питере с нуля?

Добавлено видео: Моя самая дорогая сборка. Заработать или потерять все в бизнесе на перепродаже ПК. Сколько заработал

Добавлено видео: СТУДИЯ ПОДКАСТОВ: вбухал миллионы, где прибыль? Бизнес на видеостудии в Питере - сколько заработал?

Добавлено видео: Я открыл ЦВЕТОЧНЫЙ МАГАЗИН! (опять). Есть шанс заработать?

Добавлено видео: БИЗНЕС НА АЙФОНАХ - Снова кража! Сколько заработал?

Добавлено видео: Зарабатываю на сборках ПК! Проблема с WB - сколько заработал?

Добавлено видео: КОПИЦЕНТР С АВИТО - конец бизнесу?

Добавлено видео: Я стал ПРОДЮСЕРОМ! Сколько заработал на рекламе с ЮТУБ?

Получаем 10 последних видео для канала: Жекич Дубровский

Добавлено видео: Поменяй ЖИГУ на СПОРТКАР: Кадрол, Янгер, Столяров, Ликс, Габар, Супер Стас

Добавлено видео: КУПИЛИ МАШИНУ СПРЯТАННУЮ В ОГОРОДЕ ПОД ЗЕМЛЁЙ

Добавлено видео: ВОССТАНОВИЛ ЕЁ ИЗ ПЕПЛА!

Добавлено видео: ТАЧКА СГОРЕЛА, НО ОСТАЛАСЬ НА ХОДУ!

Добавлено видео: Жекич VS Булкин VS Форсаж! Драка за Порше в мусорке. Что произошло на самом деле?

Добавлено видео: 10 КОНТЕЙНЕРОВ - Коплю на Феррари! Масленников, Жекич, Булкин, Миша, Форсаж

Добавлено видео: ВЫШИБАЛЫ - Поменял ЖИГУ на СПОРТКАР за 1 час

Добавлено видео: ОДНА НОЧЬ В АВТОСАЛОНЕ. Что за мистика здесь происходит?

Добавлено видео: УГАДАЙ КОНТЕЙНЕР - ЗАБЕРИ ТАЧКУ! Новогодний выпуск!

Добавлено видео: Я КУПИЛ СРАЗУ 2 ROLLS-ROYCE - ДРУЗЬЯ В ШОКЕ ОТ ЦЕНЫ!

Получаем 10 последних видео для канала: ТОПЛЕС

Добавлено видео: ИНФОУГРОЗЫ. ЭТО видео МЕНЯЕТ сознание (на 3 МЕСЯЦА) — ТОПЛЕС

Добавлено видео: ГЕНЕТИКА работает НЕ ТАК, КАК вы ДУМАЕТЕ! — ТОПЛЕС

Добавлено видео: Вы поймете БЕСКОНЕЧНОСТЬ за 33 МИНУТЫ — ТОПЛЕС

Добавлено видео: Главная ЛОТЕРЕЯ жизни. Как ПАСПОРТ РЕШАЕТ ВСЁ? — ТОПЛЕС

Добавлено видео: Эволюция разведки. Как за вами следят из космоса? — ТОПЛЕС

Добавлено видео: Почему ХАОС — это не СЛУЧАЙНОСТЬ, а ПОРЯДОК? — ТОПЛЕС

Добавлено видео: ВЫ НИЧЕГО НЕ ЗНАЕТЕ О ВАШЕМ ТЕЛЕ — ТОПЛЕС

Добавлено видео: Интеллект ИЗ НИЧЕГО. РАЗУМ РОЯ — ТОПЛЕС

Добавлено видео: ГЛАВНОЕ ЧИСЛО ВСЕЛЕННОЙ — ТОПЛЕС

Добавлено видео: ТРИ УРОВНЯ вашей ЛИЧНОСТИ. КТО ТЫ? — ТОПЛЕС

Собрано и сохранено 100 видео

Получаем 10 последних видео для канала: Agatha Christie

Добавлено видео: Зачем она это сделала?

Добавлено видео: Игра в "Кошки-мышки" с интернет-убийцей

Добавлено видео: Шокирующее убийство в пятницу 13-го

Добавлено видео: Стал любовником семейной пары, чтобы украсть их дочь... дважды

Добавлено видео: Биологическая мать ищет пропавшую дочь, которую отдала когда-то

Добавлено видео: Сумасшедшая история близнецов. Детство и юность, которых никогда не было

Добавлено видео: Девочка без ноги пропала. Отец две недели не замечал исчезновения

Добавлено видео: История девочки, которая стала пленницей больницы. "Я когда-нибудь увижу маму?"

Добавлено видео: Паранормальное явление... Семья переехала в дом с призраками?

Добавлено видео: "Он сам попросил об этом..." Шокирующая история учительницы

Получаем 10 последних видео для канала: Калинкин!

Добавлено видео: Скриптонит, соседи, богатство, Троцкий, the game и т.д. и т.п. (№8)

Добавлено видео: GTA, анекдоты, Хантер Томпсон, лес, отцы и т.д. и т.п. (№7)

Добавлено видео: КАК МЫ СНЯЛИ РЕКЛАМУ, КОТОРАЯ ДЛИТСЯ ЧАС

Добавлено видео: появление. свинка пеппа. певность. бабushка, babushka и т.д. и т.п.

-ые, кошки и т.д. и т.п. (№5)

емцелы и т.д. и т.п. (№4)

и т.д. и т.п.

ничная клоунада и т.д. и т.п. (№2)

ьштам и т.д. и т.п. (№1)

ции

ы. (Егорик, FrameTamer, Сатир, Sqwoz Bab)

А, Сатир, Шах, Раговский // ЧАСТЬ 3

. Exile, Дипинс, Братишкин, SQWOZBAB, Горилла, FrameTamer)

FrameTamer, Горилла, Сатир, Ликс, by_Owl, Постик, SASAVOT

p, nkeeei, VOSKRESENSKII, Tenderlybae, Сабина и др.
НК 1.000.000P

, предательстве учеников и пошлости Аноры

группе «Мираж» и работе на зоне

илисы Володиной с «Давай поженимся» и возвращении в кино

легам и последнем альбоме

зда» и «Самолеты»

романе с Сосо Павлиашвили и жизни сегодня

Чеботиной, Крутом и ордене за заслуги перед Отечеством

, похищении детей и лишении мужа родительских прав

АЛГОРИТМ ПОИСКА БРЕНДОВ

```
# Содержит официальные названия различных компаний, сервисов и брендов на русском и английском языках
BRANDS_LIST = [
    'Сбер', 'Тинькофф', 'Яндекс', 'ВКонтакте', 'Aviasales', 'Т-Банк', 'Телеграм', 'Золотое яблоко', 'Realme', 'Магнит', 'Jaguar', 'Republic', 'Genotek',
    'Лаборатория Касперского', 'Wildberries', 'Ozon', 'Пятёрочка', 'Периодика', 'Бирдок', 'Genotek', 'Центральный университет', 'Кардио М', 'Rainbow',
    'Перекрёсток', 'Вкусно и точка', 'Avito', 'Xiaomi', 'Вкусвилл', 'Алиса', 'Lamoda', 'Черный жемчуг', 'Wae Thunder', 'Axe', 'Fit Sevice', 'Bombbar',
    'Додо пицца', 'ТНТ', 'Honor', 'OneTwoTrip', 'Samsung', 'Мегафон', 'CoolCola', 'Philips', 'Yves Rocher', 'Bang Bang Education', 'Дром', 'Битрикс24',
    'Читай-город', 'Contented', 'Flowwow', 'Skyeng', 'LitEnergy', 'Эльдорадо', 'Majestic RP', 'Mozabrick', 'Bonnie&Slide', 'Selectel',
    'МТС', 'Wibes', 'Билайн', 'Авто.ру', 'Кавёр', 'Московский институт психологии', 'BR Group', 'Аксум', 'Wink', 'FitStars', 'Redmi', 'Foreo', 'Skypro',
    'Синхронизация', 'Нетология', 'Букмейт', 'Суточно.ру', 'Kari', 'Erborian', 'Ivi', 'Индилайт', 'ЛЭТУАЛЬ', 'Pinskiy&Co', 'Smart', 'Островок!',
    'RingString', 'Здравсити.ру', 'Primavera', 'Roborock', 'Модульбанк', 'Махеев', 'PREMIER', 'Профи.ру', 'Rox', 'Fresh'
]

# Словарь для нормализации названий брендов
# Ключ - официальное название бренда
# Значение - список возможных вариантов написания (с ошибками, в различных падежах и т.д.), поиск которых осуществлялся вручную
BRAND_NORMALIZATION = {
    'Сбер': ['сбербанк', 'сбером', 'сбере', 'сбера', 'збер', 'сберкарта', 'спер'],
    'Тинькофф': ['тинькоф', 'тинькоффом', 'тинькоффа', 'тинькофф'],
    'Т-Банк': ['т банк', 'т банком', 'т банке', 'т банка', 'тбанк', 'тбанке', 'т-банк', 'т-банке', 'тбнк', 'т-банка', 'т-', 't-банк'],
    'Лаборатория Касперского': ['лаборатории касперского', 'лабораторию касперского', 'лабораторией касперского'],
    'Яндекс': ['yandex', 'яндексом', 'яндекса', 'яндексе'],
    'ВКонтакте': ['вк', 'vk'],
    'Wildberries': ['вайлдберриз', 'wildberry', 'валдберис', 'вайлдберис', 'вайлдберрисе', 'wibberies'],
    'Aviasales': ['авиасейлс', 'авиа сейлс', 'авиас', 'aviaes', 'авиа'],
    'Пятёрочка': ['пятерочке', 'пятерочку', 'пятерочках', 'пятёрочке', 'пятёрочку', 'пятёрочках'],
    'Телеграм': ['телеграм', 'телега', 'телеграме', 'телеге', 'тгк', 'телеграмом'],
    'Золотое Яблоко': ['золотое яблоко', 'золотом яблоке', 'золотого яблока'],
    'Ozon': ['азон', 'озон', 'озоне', 'озоном', 'ozone'],
    'Периодика': ['периодике', 'периодику', 'периодики', 'периодика'],
    'Бирдок': ['бир док', 'бирдок'],
    'Вкусвилл': ['вкусвилле', 'вкус вилл', 'вкус вилле', 'вкус вил', 'вкусвил'],
    'Алиса': ['алисой', 'алисы', 'алиса'],
    'Додо пицца': ['додо пиццу', 'додо пицца'],
    'Мегафон': ['мегафоном', 'мегафон'],
    'Читай-город': ['читай городе', 'читай город', 'читай города'],
    'Нетология': ['нетологии', 'нетология'],
}
```



```
def clean_for_general_analysis(text):
    """Общая очистка текста для анализа"""
    text = text.lower() #приводим к нижнему регистру
    text = ' '.join(text.split()) #удаляем лишние пробелы
    return text

def clean_for_brand_detection(text):
    """Очистка текста для поиска брендов"""
    return f' {text.lower()} ' #приводим к нижнему регистру и добавляем пробелы
```

```
def lemmatize_text(text):
    """
    Лемматизация текста - приведение слов к их начальной форме
    Сохраняет оригинальное написание брендов
    """
    words = word_tokenize(text) #разбиваем текст на слова

    # Создаем множество всех возможных вариантов написания брендов
    brand_variations = set() #множество для всех возможных вариантов написания брендов
    for brand in BRANDS_LIST:
        brand_variations.add(brand.lower()) #добавляем все бренды в множество
        if brand in BRAND_NORMALIZATION:
            brand_variations.update(v.strip() for v in BRAND_NORMALIZATION[brand]) #добавляем все варианты

    # Лемматизируем слова, пропуская бренды
    lemmatized_words = []
    for word in words:
        word_lower = word.lower() #приводим слово к нижнему регистру
        if word_lower in brand_variations:
            lemmatized_words.append(word) # сохраняем оригинальное написание бренда
        else:
            lemmatized_words.append(morph.parse(word)[0].normal_form) #приводим слово к начальной форме

    return ' '.join(lemmatized_words)
```

```
def remove_stopwords(text):
    """Удаление стоп-слов из текста"""
    stop_words = set(stopwords.words('russian') + ['это', 'вот', 'ну', 'всё', 'все', 'ещё', 'кстати'])
    words = word_tokenize(text)
    return ' '.join([word for word in words if word not in stop_words])
```

1.Предварительная обработка текста:

Приведение текста к нижнему регистру

- Для общего анализа: удаление лишних пробелов
- Для поиска брендов: добавление пробелов для точного поиска

2. Лемматизация и обработка слов:

Токенизация (разбиение текста на слова)

Обработка каждого слова:

- Для брендов: сохранение оригинального написания
- Для обычных слов: лемматизация (приведение к начальной форме)

Удаление стоп-слов


```
def find_brand_mentions_with_source(transcript):
    """Поиск упоминаний брендов с указанием источника (прямое совпадение или нормализация)"""
    mentions = []
    seen = set()

    for entry in transcript:
        text = clean_for_brand_detection(entry['text']) #очищаем текст для поиска брендов
        time = entry['start'] #время начала упоминания бренда

        # Проверка основных названий брендов (прямые совпадения)
        for brand in BRANDS_LIST:
            brand_lower = f'{brand.lower()} ' #приводим бренд к нижнему регистру и добавляем пробелы
            if brand_lower in text:
                key = (brand, int(time)) #ключ для бренда и времени
                if key not in seen: #если ключ не встречался, добавляем в список упоминаний
                    mentions.append({
                        'brand': brand,
                        'time': time,
                        'context': entry['text'][:50] + '...', #контекст упоминания бренда
                        'detection_type': 'прямое_совпадение' #тип обнаружения бренда
                    })
                    seen.add(key)

        # Проверка нормализованных вариантов написания брендов
        for brand, variations in BRAND_NORMALIZATION.items(): #проверяем все варианты написания брендов
            for variant in variations: #проверяем все варианты написания бренда
                if variant in text: #если вариант написания бренда в тексте
                    key = (brand, int(time)) #ключ для бренда и времени
                    if key not in seen:
                        mentions.append({
                            'brand': brand,
                            'time': time,
                            'context': entry['text'][:50] + '...',
                            'detection_type': 'подлежащие_нормализации',
                            'matched_variant': variant
                        })
                        seen.add(key)

    return sorted(mentions, key=lambda x: x['time']) #сортируем упоминания по времени
```

3. Поиск брендов:

- Поиск прямых совпадений из списка брендов
- Поиск вариаций написания брендов (нормализация)
- Сохранение контекста и времени упоминания

```
def analyze_videos_with_detection_stats(videos_df):
    """
    Анализ видео с подробной статистикой по методам обнаружения брендов
    """

    results = []
    all_mentions = []

    for _, video in videos_df.iterrows():
        # Получаем упоминания с типом обнаружения
        mentions = find_brand_mentions_with_source(video['transcript'])
        brands = list(set(m['brand'] for m in mentions)) #список уникальных брендов

        # Подсчитываем упоминания по типу обнаружения
        direct_matches = [m for m in mentions if m['detection_type'] == 'прямое_совпадение']
        normalized_matches = [m for m in mentions if m['detection_type'] == 'подлежащие_нормализации']

        # Определяем, является ли видео рекламой
        is_ad = any(brand in video['title'].lower() for brand in brands) #проверяем, есть ли бренд в названии видео
        is_ad_pattern = any(marker in video['title'].lower() or marker in video['raw_text'].lower() for marker in AD_MARKERS) #проверяем, есть ли рекламные маркеры

        # Формируем результаты анализа для видео
        results.append({
            'channel': video['channel'],
            'video_id': video['video_id'],
            'title': video['title'],
            'brands': brands,
            'brand_mentions': mentions,
            'brand_count': len(brands),
            'direct_match_count': len(direct_matches),
            'normalized_match_count': len(normalized_matches),
            'is_ad': is_ad,
            'is_ad_pattern': is_ad_pattern,
            'raw_text': video['raw_text'],
            'cleaned_text': lemmatize_text(clean_for_general_analysis(video['raw_text'])), #очищаем текст для анализа
            'no_stopwords': remove_stopwords(lemmatize_text(clean_for_general_analysis(video['raw_text']))) #очищаем текст от стоп-слов
        })

    # Добавление к общему списку упоминаний для последующего анализа
    all_mentions.extend(mentions)

    return pd.DataFrame(results), pd.DataFrame(all_mentions) #возвращаем результаты анализа и список упоминаний
```

4. Анализ рекламного контента:

- Подсчет упоминаний
- Проверка наличия брендов в названии видео
- Поиск рекламных маркеров в названии и тексте

5. Статистический анализ:

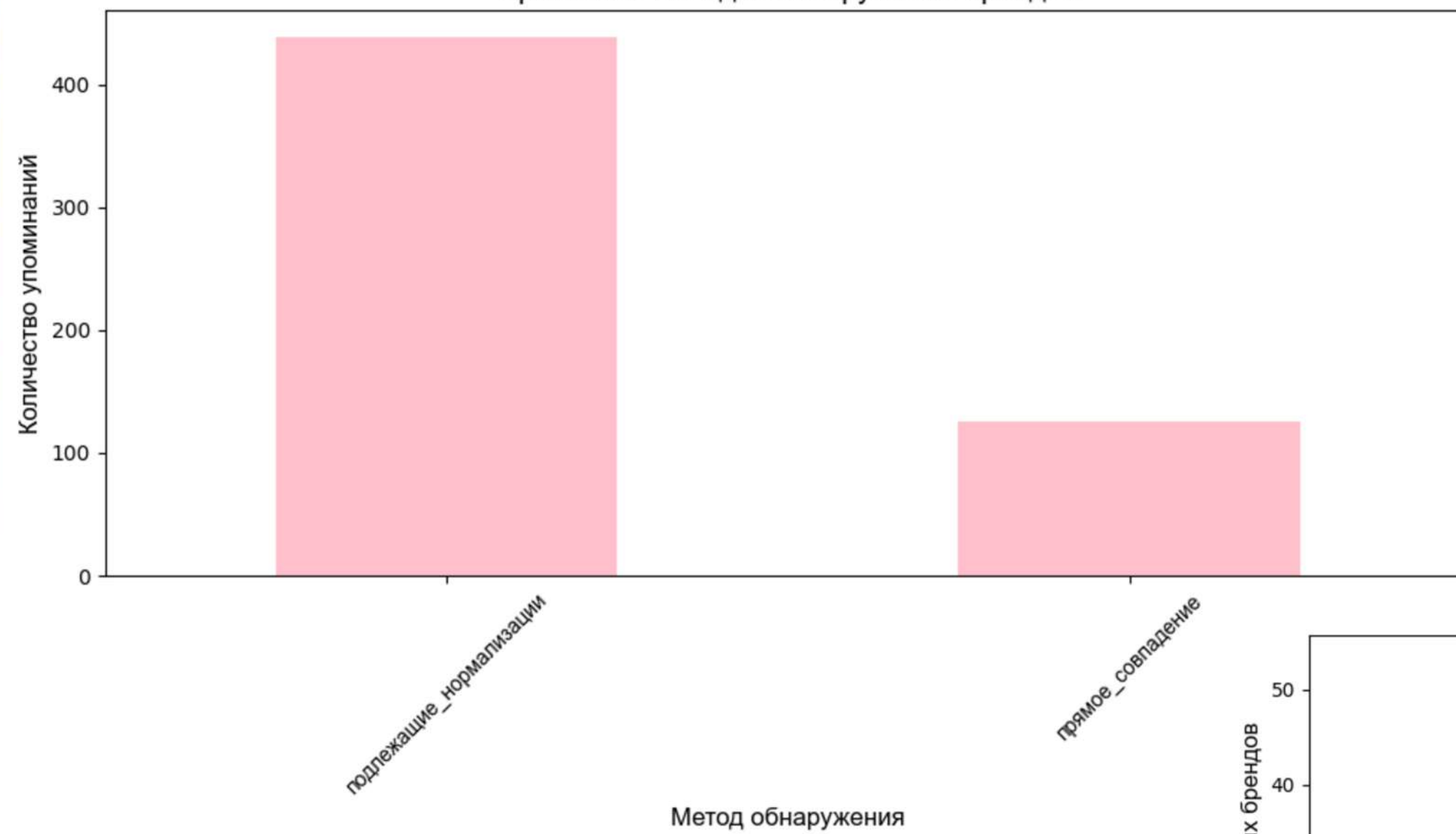
- Подсчет количества упоминаний по типам обнаружения
- Анализ уникальных брендов для каждого метода
- Анализ частоты нормализованных вариантов
- Определение наиболее часто нормализуемых брендов

```
def analyze_detection_methods(mentions_df):  
    """Анализ эффективности методов обнаружения брендов"""  
    # Подсчет упоминаний по типу обнаружения  
    detection_counts = mentions_df['detection_type'].value_counts()  
  
    # Подсчет уникальных брендов по типу обнаружения  
    unique_brands_by_type = mentions_df.groupby('detection_type')['brand'].nunique()  
  
    # Наиболее частые нормализованные варианты  
    normalized_variations = mentions_df[mentions_df['detection_type'] == 'подлежащие нормализации']['matched_variant'].value_counts()  
  
    # Бренды, чаще всего обнаруживаемые через нормализацию  
    normalized_brands = mentions_df[mentions_df['detection_type'] == 'подлежащие нормализации']['brand'].value_counts()  
  
    return {  
        'detection_counts': detection_counts,  
        'unique_brands_by_type': unique_brands_by_type,  
        'top_variations': normalized_variations.head(10), #топ 10 наиболее частых нормализованных вариантов  
        'top_normalized_brands': normalized_brands.head(10) #топ 10 брендов, чаще всего обнаруживаемых через нормализацию  
    }
```

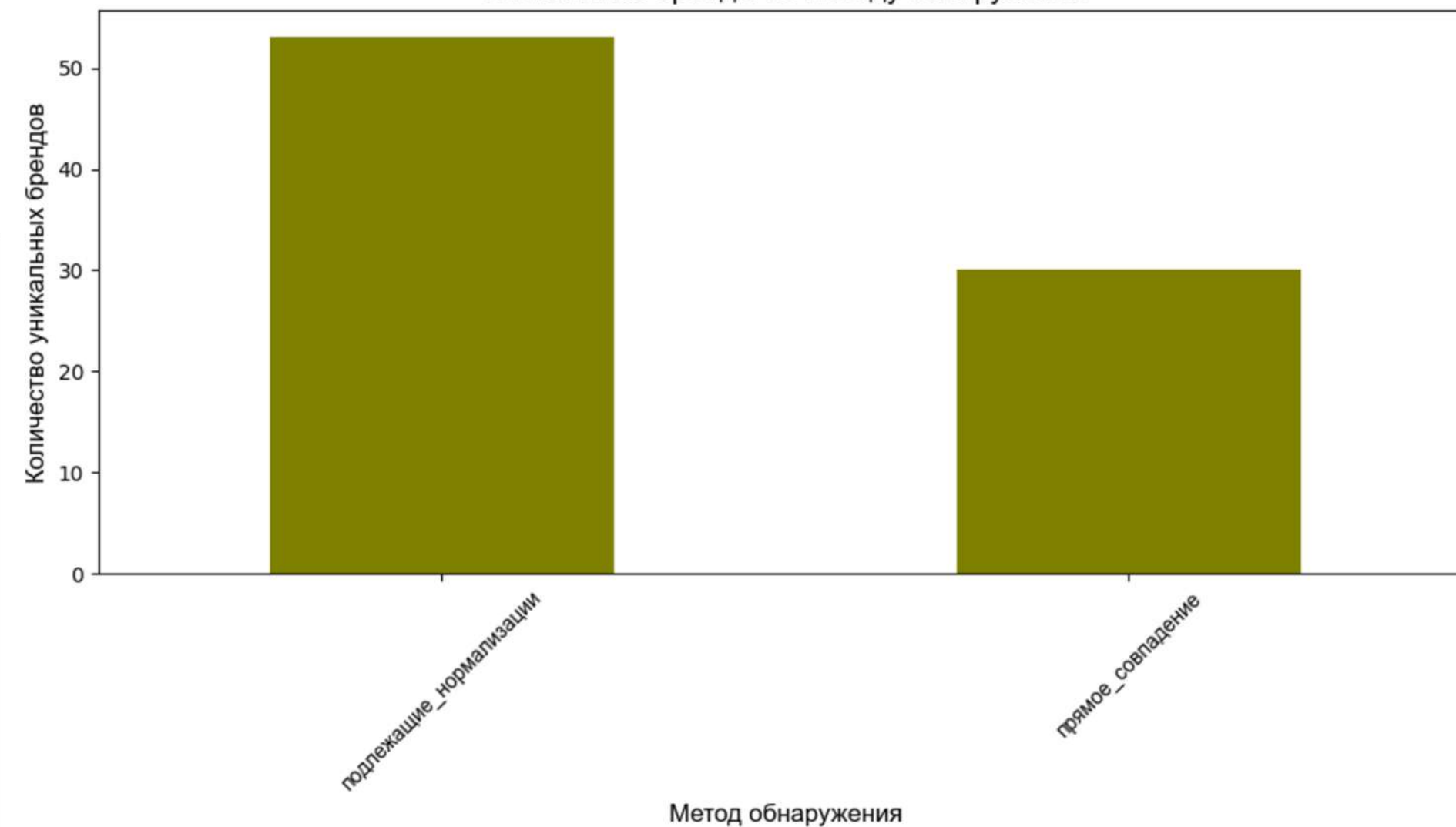
6. Визуализация результатов

ПЕРВЫЕ ИТОГИ

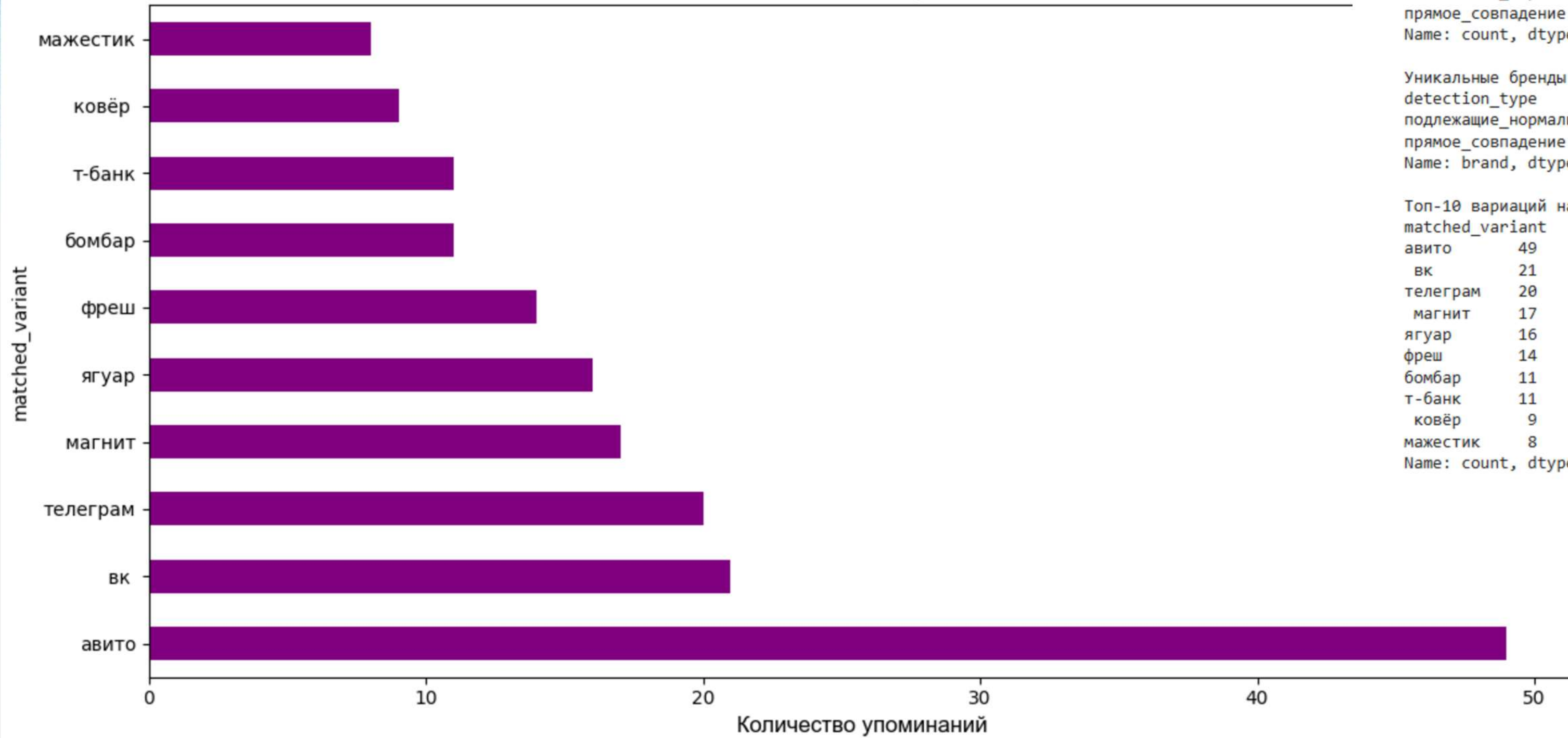
Сравнение методов обнаружения брендов



Уникальные бренды по методу обнаружения



Топ-10 наиболее частых вариаций написания



Статистика обнаружения брендов:
Всего упоминаний: 563

По методам обнаружения:
detection_type
подлежащие_нормализации 438
прямое_совпадение 125
Name: count, dtype: int64

Уникальные бренды по методам:
detection_type
подлежащие_нормализации 53
прямое_совпадение 30
Name: brand, dtype: int64

Топ-10 вариаций написания:
matched_variant
авито 49
вк 21
телеграм 20
магнит 17
ягуар 16
фреш 14
бомбар 11
т-банк 11
ковёр 9
мажестик 8
Name: count, dtype: int64

УГЛУБЛЕННЫЙ АНАЛИЗ ПОЛУЧЕННЫХ ДАННЫХ И ИХ ВИЗУАЛЬНОЕ ПРЕДСТАВЛЕНИЕ

- Вывод временных меток упоминаний брендов, их названия и контекст упоминания
- Визуализация результатов

```
def show_brand_timeline(df):  
    """Показывает временные метки упоминаний брендов"""  
    for _, row in df.iterrows(): #проходим по всем строкам dataframe  
        if not row['brand_mentions']: #если нет упоминаний брендов, пропускаем  
            continue  
  
        print(f"\nВременные метки брендов в видео: {row['title']}") #выводим название видео  
        for mention in sorted(row['brand_mentions'], key=lambda x: x['time']): #проходим по всем упоминаниям брендов  
            mins, secs = divmod(int(mention['time']), 60) #переводим время в минуты и секунды  
            print(f"{mins:02d}:{secs:02d} - {mention['brand']}: {mention['context']}") #выводим время, бренд и контекст упоминания
```


Временные метки брендов в видео: Охота На Звезд 2 ! Литвин и Равшан vs Сэм и Егорик !

12:51 - Majestic RP: зайти на любой серверов маес потому что...

13:13 - Majestic RP: поверь это не главная фишка маджестика...

76:04 - LitEnergy: чего Ладно я делаю глоток L Energy И...

Временные метки брендов в видео: Кто Последний Уснёт Забирает 1 000 000 !

12:09 - Majestic RP: Добро пожаловать на majestic RP...

12:50 - Majestic RP: Majestic RP Будьте готовы равшик...

12:56 - LitEnergy: баночку Lit Energy пока ты её держишь...

39:16 - Алиса: Алиса записала Алиса есть у тебя Ну чаме...

Временные метки брендов в видео: Украли Машину за 10 Миллионов ! Наказали Вора !!

01:20 - LitEnergy: именно За регистрацию банок L Energy в...

10:02 - Сбер: были встретиться у Сбербанка зачем вы...

10:04 - Сбер: хотели Сбербанк найти посчитать деньги...

11:10 - Сбер: быстрее Ну всё произошло в Сбербанке...

13:12 - Сбер: около сбера зайти внутрь сделать...

14:07 - Majestic RP: пользователей недавно на majestic RP...

15:06 - Majestic RP: мажестик РП обязательно вводите мой...

15:15 - Majestic RP: дней подписки majestic Premium Играйте и...

18:05 - Сбер: Сбербанке посчитать деньги отдам Мне...

23:12 - LitEnergy: бодрящий праздник линер...

23:48 - LitEnergy: предоставляю вам дорогие друзья тенер же...

24:20 - LitEnergy: любым вкусом L Energy все условия чтобы...

40:10 - Сбер: Сбербанке мы напишем сумму и подпишем...

44:17 - Contented: на кон не буду ставить но я готов здесь...

Временные метки брендов в видео: Прошлась Голой За 1 000 000 ! На Что Люди Готовы Ради Денег ?

01:41 - LitEnergy: L Energy и хотелось бы вас искренне...

03:47 - LitEnergy: запускал бизнес Le Energy для нас с...

44:17 - LitEnergy: на energ это жёстко energ...

44:21 - LitEnergy: доске ты голо дома не пол energ почему...

45:58 - LitEnergy: фамилия Energy сделай бра бомба же Ну...

46:25 - LitEnergy: Ты что сделаем сделаем Давай Lit Energy...

46:27 - LitEnergy: к доске L Energy там Выйди там Расскажи...

46:30 - LitEnergy: там T Energy Почему опоздал если ты это...

46:54 - LitEnergy: а т Energy Всё молодец молодец всё И...

47:01 - Телеграм: месяца одобряется у себя в телеграме я...

53:15 - LitEnergy: увидеть Наш новый продукт от Energy...

54:26 - LitEnergy: персиковый Lit Energy теперь я вам...

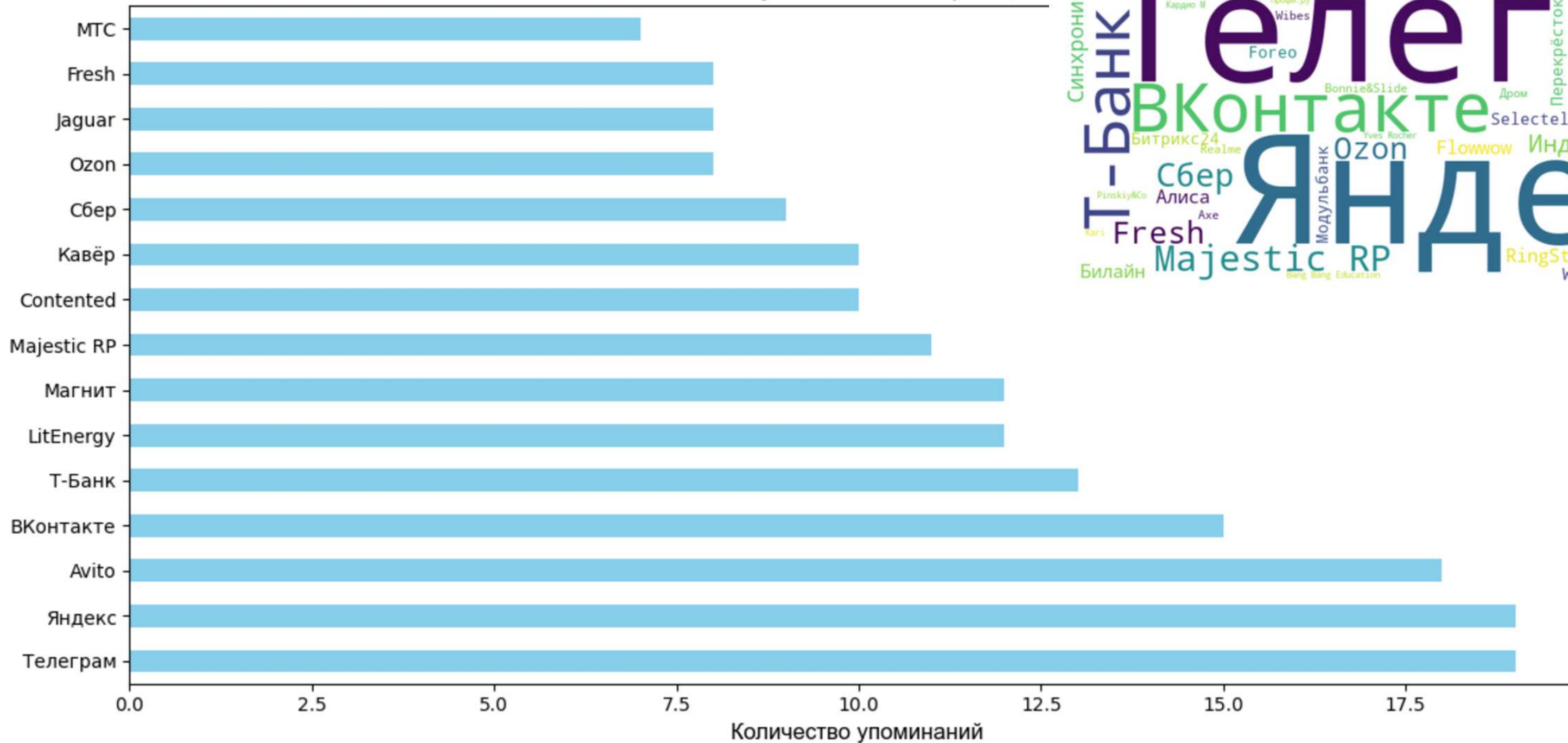
54:34 - LitEnergy: нашей компании Lit Energy и будем работа...

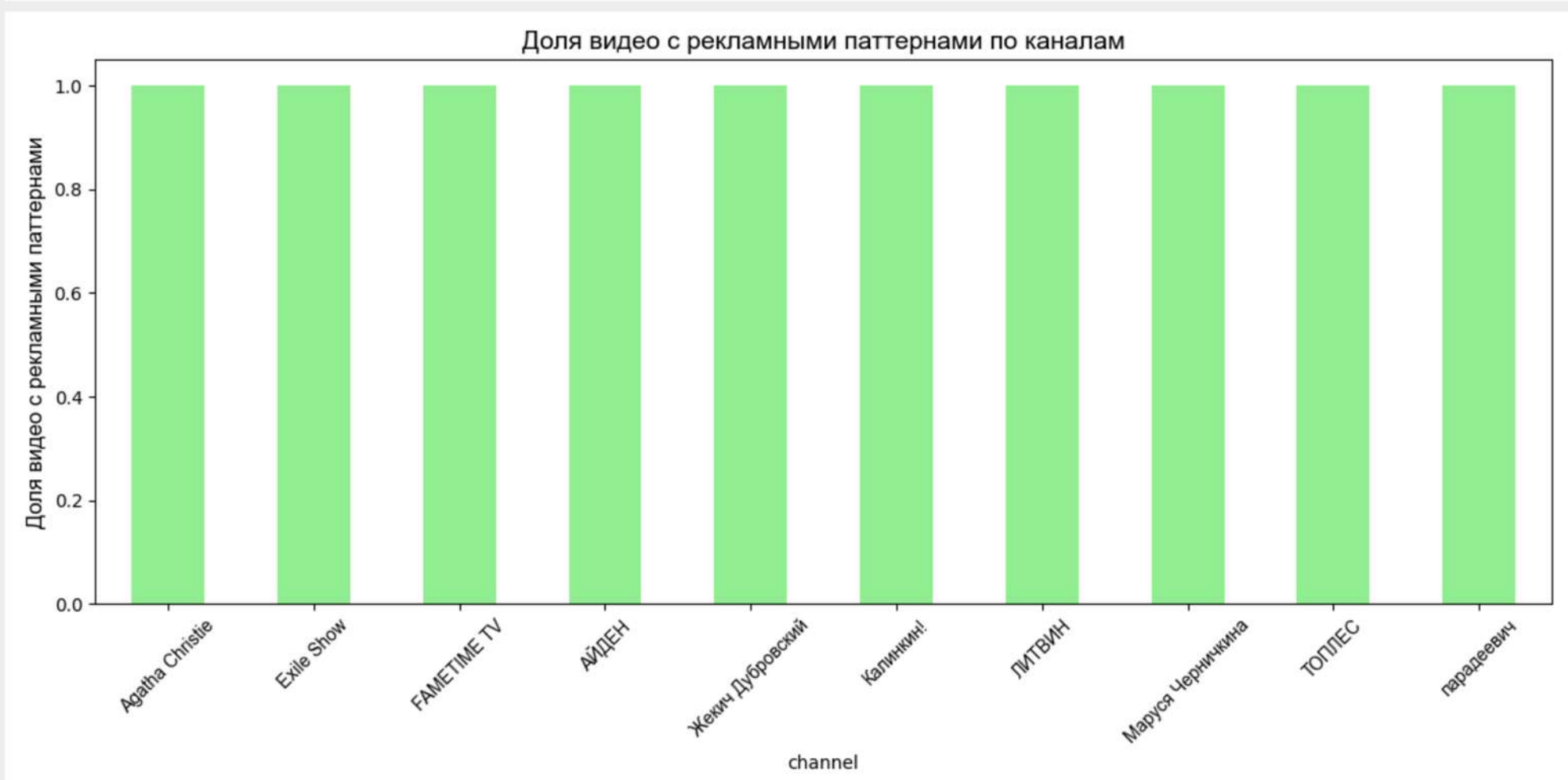
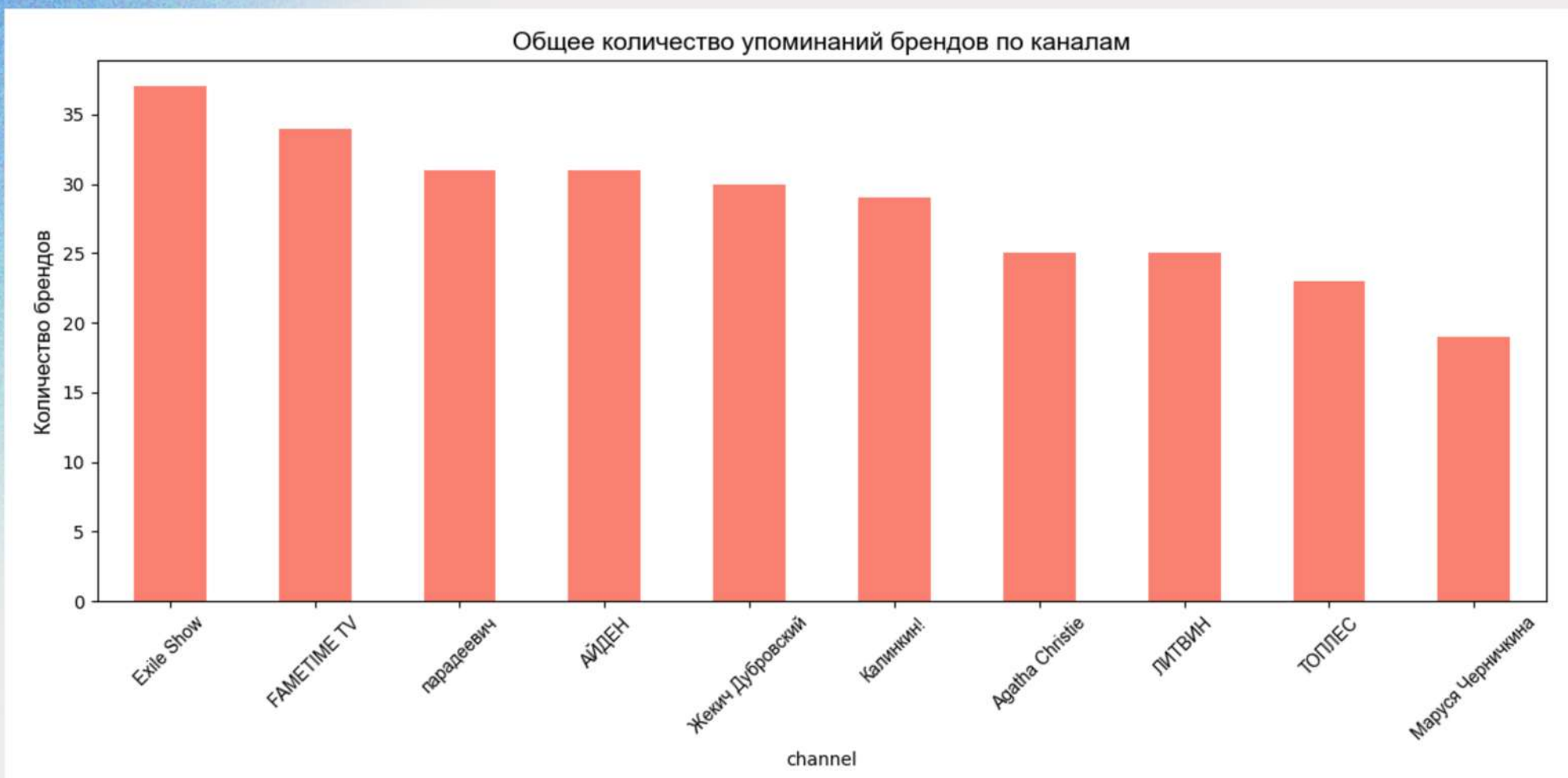
Топ брендов по упоминаниям:

Бренд	Количество
Телеграм	19
Яндекс	19
Avito	18
ВКонтакте	15
T-Банк	13
LitEnergy	12
Магнит	12
Majestic RP	11
Contented	10
Кавёр	10
Сбер	9
Ozon	8
Jaguar	8
Fresh	8
MTC	7

РЕЗУЛЬТАТЫ

Топ 15 упоминаемых брендов





Статистика рекламных паттернов по каналам:
Agatha Christie: 10/10 видео с рекламными паттернами (100.0%)
Exile Show: 10/10 видео с рекламными паттернами (100.0%)
FAMETIME TV: 10/10 видео с рекламными паттернами (100.0%)
АЙДЕН: 10/10 видео с рекламными паттернами (100.0%)
Жекич Дубровский: 10/10 видео с рекламными паттернами (100.0%)
Калинкин!: 10/10 видео с рекламными паттернами (100.0%)
ЛИТВИН: 10/10 видео с рекламными паттернами (100.0%)
Маруся Чернишкина: 10/10 видео с рекламными паттернами (100.0%)
ТОПЛЕС: 10/10 видео с рекламными паттернами (100.0%)
парадеевич: 10/10 видео с рекламными паттернами (100.0%)

Анализ завершен. Результаты сохранены в brand_analysis_detailed.xlsx

ИНСТРУМЕНТЫ

Для сбора данных:

- youtube_transcript_api
- googleapiclient.discovery
- pandas

Для обнаружения брендов:

- nltk (word_tokenize + stopwords)
- pymorphy3
- matplotlib
- pandas

Для анализа и визуализации :

- pandas
- matplotlib
- wordcloud

ТРУДНОСТИ

- 1) Некоторые бренды (например, сервис “Ясно”) имеют названия, совпадающие с общеупотребительными словами. Это приводило к попаданию в выборку большого количества нерелевантных упоминаний.
- 2) Ограничения Youtube API: 100 запросов в день + при частом обращении может перестать отдавать какие-либо результаты (пришлось сначала прибегнуть к `time.sleep(5)`)
- 3) Субтитры обнаруживались не во всех видео, хотя они там были. Пришлось их пропускать и продолжать поиск дальше.

ВЫВОДЫ

Всего обнаружено 563 упоминания брендов, большинство из них (438) требовали нормализации, случаев прямого совпадения было 125.

Топ-брендами по упоминаниям стали:

- Телеграм (19 упоминаний)
- Яндекс (19 упоминаний)
- Авито (18 упоминаний)
- ВКонтакте (15 упоминаний)

Exile Show лидирует по количеству упоминаний брендов (около 35). FAMETIME TV на втором месте (около 33). Наименьшее количество упоминаний у канала "Маруся Черничкина" (около 19).

100% видео на всех каналах содержат рекламные паттерны. Каждый канал имеет 10/10 видео с рекламными интеграциями

Авито имеет наибольшее количество вариаций написания - 49, ВК - 21, Телеграм - 20.

СПАСИБО ЗА ВНИМАНИЕ!

Проект на GitHub:

https://github.com/NadiaVeles/Subtitles_Analysis_on_Youtube