

Описание плана работы

1. Подводка к проблеме.

Исследование музыкальных предпочтений и их изменений в контексте социальных кризисов и, в частности, таких событий, как специальная военная операция, представляет собой значимую область анализа. Оно позволяет выявить, как изменения в обществе отражаются на культурных предпочтениях, особенно у молодых людей, основной аудитории музыкального контента. В возрастной категории 18-35 лет наблюдается наибольшее потребление музыки, которое снижается после 35 лет, когда происходит переосмысление музыкальных вкусов¹.

Российский рынок характеризуется предпочтением отечественных исполнителей, что подтверждается мнениями экспертов и аналитиков рынка: около 70% потребляемой музыки приходится на российских артистов, с популярностью таких жанров как поп-музыка, рок и хип-хоп. Однако в 2022 году рынок музыкального стриминга в России испытал сокращение минимум на 30% из-за ухода мейджор-лейблов (Warner, Sony и Universal) и стриминговых сервисов (Spotify, YouTube Music и Apple Music (частично)), что вкуче с санкциями привело к сокращению музыкального разнообразия и финансовых возможностей артистов².

Существенные изменения коснулись не только дистрибуции контента, но и концертной активности (Россию покинула почти треть крупных артистов). Также стало заметно уменьшение новых имен на сцене и изменения в творческом подходе существующих исполнителей. “Раньше выходили какие-то сумасшедшие клипы Александра Гудкова, Егора Крида, Тимати. В 2022-м такого нет. Год прошел под знаменем певца Shaman, у которого дешевый клип и много концертов”³. Таким образом, музыкальный ландшафт выглядит следующим образом: знакомые исполнители продолжают выпускать музыку, но

¹ Исследование музыкального рынка Москвы:

<https://data.tedo.ru/publications/moscow-music-market.pdf>

² <https://www.setters.media/post/kak-perestroilsya-muzykalnyy-rynok-v-rossii-posle-24-fevralya>

³ https://dzen.ru/a/Y7KLQ6q_PRqahBZO

без значительных успехов и новых интересных решений, в то время как инди-музыканты стараются занять свободные ниши на стриминговых платформах, а люди больше слушают музыку в фоновом режиме.

Такие тенденции могут вести к усилению практик поиска и прослушивания контента вне официальных платформ. Кроме того, на фоне этих событий, тексты песен становятся более радикализированными и упрощенными (это своим творчеством доказывает Shaman), что говорит о смене акцентов в музыкальной культуре.

Таким образом, наш проект, направленный на анализ российских музыкальных топов до и после военных событий, позволит глубже понять изменения в музыкальной среде, выявив тенденции и предпочтения слушателей, что важно для понимания общих культурных сдвигов в обществе в ответ на кризисные явления.

2. Цель проекта.

Наш проект заключается в составлении базы данных и первичном анализе российских музыкальных рейтингов в период 2020-2023 гг. *с целью выявления изменений в музыкальных предпочтениях российского слушателя на фоне социально-политических тенденций данного временного периода.* Мы предполагаем, что такие события, как военные действия на Украине, ужесточение санкционного режима в отношении России, расширение списка лиц-”иностранных агентов” и нежелательных организаций, рост миграционных настроений и прочие явления последних лет, с которыми столкнулось российское общество, существенно повлияли на характер потребления многих культурных продуктов, которые, в свою очередь, стали “отражением” новых социальных потребностей. Музыкальные практики и предпочтения не являются в этом плане исключением и позволяют изучить общественный запрос на формы и содержание рекреационного творчества.

3. Мини-обзор на литературу по взаимосвязи музыки и политической мобилизации.

Музыка давно признана мощным средством коммуникации, способным влиять на социальные и политические процессы. В контексте политической мобилизации музыка может служить как инструментом укрепления идентичности и солидарности, так и средством протеста и сопротивления.

Например, в работе Джона Стрита “Музыка и политика” анализируется, как музыка используется в политических процессах, обсуждается её роль в политических кампаниях, манифестах и революциях⁴. Автор рассматривает музыку как форму политического дискурса, которая может как отражать, так и формировать политические идеи и идентичности, иными словами формироваться социальной реальностью и формировать ее одновременно. Также достаточно большое количество исследований написано о взаимосвязи музыки и социальных движений: Эйерман и Джеймисон рассматривают случаи, когда музыка становится центральной в мобилизации общественного мнения и действий, предлагая теоретический анализ роли музыки в социальных изменениях⁵. Розенталь и др. исследовали истории музыкантов и песен, которые играли ключевую роль в различных социальных движениях. Они проанализировали, как музыкальное искусство может вдохновлять на изменения и способствовать мобилизации и активизму⁶.

В итоге, несмотря на кажущуюся неполитизированность музыкальные практики участвуют в формировании политического мифа и часто включают значимые социальные и политические элементы. Эти элементы могут отражать как классовую принадлежность, так и идеологическую направленность, определяя музыку как средство мобилизации населения. Важность анализа музыкальных практик подчеркивается через их роль в социальных ритуалах, которые помогают группам манифестировать свою идентичность⁷.

⁴ Street J. Music and politics. – John Wiley & Sons, 2013.

⁵ Eyerman R., Jamison A. Music and social movements: Mobilizing traditions in the twentieth century. – Cambridge University Press, 1998.

⁶ Rosenthal R., Flacks R. Playing for change: Music and musicians in the service of social movements. – Routledge, 2015.

⁷ Симон М.Е. Голос субалтернов: репрезентации инакости в музыкальных практиках «черной Атлантики» // Логос. – 2016. – Том 26, № 4. – С. 63-94.

Принятие этого подхода из области культурных исследований позволяет глубже понять связь между музыкой, политикой и ритуальностью.

Анализ "внутренних кодов" музыки показывает, как музыка может передавать скрытые сообщения, не всегда понятные для внешних наблюдателей или цензоров. Исторические и современные примеры показывают, как музыкальные практики участвуют в формировании политических мифов и идентичностей, не являясь чем-то заранее заданным, но формируемым через творческий и социальный процессы⁸.

Критическая перспектива, например, в духе Теодора Адорна, поднимает вопрос о том, не потеряли ли музыкальные практики свой эмансипаторный характер, превратившись в инструмент манипуляции. Однако, некоторые исследования часто демонстрируют обратное, показывая, как поп-музыка может служить средством самовыражения и мобилизации "снизу"⁹. При этом политический контекст играет значительную роль в том, как формируются музыкальные практики и политические мифы, влияя на выбор акторов, актуализацию исторических и культурных традиций, формы музыкальных практик и объединяющие их социальные группы.

В заключение, данный небольшой обзор подчеркивает многослойность и глубину взаимосвязи между музыкой и политической мобилизацией, подтверждая значимость музыкальных практик в социально-политических процессах. Музыка не только отражает, но и активно формирует социальную и политическую реальность, служа средством для выражения идентичности, мобилизации сообществ или сопротивления. В рамках нашего проекта, направленного на изучение изменений в музыкальных предпочтениях российских слушателей, эти выводы подчеркивают важность анализа музыкального контента и практик в их способности реагировать на социально-политические изменения. Осознание того, что выбор и прослушивание музыки и какое-то участие в музыкальной среде могут служить не просто формой развлечения, но и активным участием в политической жизни, добавляет дополнительный уровень значимости в наш проект, позволяя в дальнейшем

⁸ Там же.

⁹ Gilroy P. *The Black Atlantic: Modernity and Double Consciousness*. – L.: Verso, 1993. – 280 p.

(при расширении выборки и методологии исследования) не только отследить изменения в предпочтениях, но и понять их более глубокие социальные и политические импликации.

4. Выбор источников. Обоснование выбора источника и технического решения.

Альтернативные технические решения для источника.

В качестве источников данных мы выбрали сайты: Apple music (<https://music.apple.com/ru/browse>) и Genius (<https://genius.com/>). В соответствии с целью проекта нам нужно было проанализировать музыкальные топы, для этого был выбран сайт Apple music, который отвечает основным критериям: во-первых, это единственный иностранный музыкальный стриминговый сервис, который все еще официально присутствует на территории РФ; во-вторых, это один из крупнейших музыкальных сервисов в мире; в-третьих, на сайте опубликованы российские музыкальные топы за последние 4 года, что позволяет составить релевантную выборку для анализа. Для того, чтобы проанализировать содержание отобранных треков и выявить в них закономерности, необходимо было включить в наш набор данных сами тексты песен, которые были собраны со специального сайта с текстами – Genius.com. Этот сайт является крупнейшей и самой популярной платформой, содержащей полноценные тексты практически ко всем трекам.

В качестве технического решения мы использовали Selenium по следующим причинам. Selenium позволяет выполнять сложные действия в браузере, такие как клики по элементам, прокрутка страницы и ввод текста в формы поиска, что необходимо для автоматизации процесса сбора данных. Например, на сайте Apple Music нужно переходить по плейлистам и собирать информацию о треках, а на Genius — вводить названия треков в поисковую строку и извлекать тексты песен. Selenium отлично подходит для этих задач, так как предоставляет возможность имитировать пользовательские действия в браузере.

В качестве альтернативного технического решения мы могли бы использовать API Apple Music и Genius, однако, у Apple Music API является платным, а у Genius API имеет ограничения в бесплатном доступе, в частности, ограничение по количеству запросов, что

тормозило бы процесс сбора данных. Также мы бы могли использовать библиотеки Beautiful Soup и Requests, но они все равно больше подходят для парсинга статических веб-страниц и будут неэффективными в условиях динамического контента и асинхронно загружаемых данных.

5. Карта проекта. Какие данные собираем?

На первом этапе мы собираем данные с сайта <https://music.apple.com/ru/browse>. Внутри каждого плейлиста за каждый год с 2020 по 2023 включительно мы собираем: “Номер в рейтинге”, “Название”, “Исполнитель”, “Альбом”, “Длительность”, “Ненормативный контент”, “Год вхождения в топ” и “Ссылка на прослушивание” (в качестве элемента метаданных). На втором этапе с сайта <https://genius.com/> мы собираем недостающие данные, а именно, “Тексты треков”, “Дата выхода” и “Ссылка на страницу”.

6. Этапы сбора данных.

1. Сбор данных с сайта <https://music.apple.com/ru/browse>

1.1 Для начала мы собираем ссылки на нужные нам плейлисты (то есть на плейлисты с российскими топами музыки за 2020, 2021, 2022 и 2023)

1.2 Далее мы применяем функцию, которая обрабатывает заданные веб-страницы, то есть страницы с плейлистами. Сначала функция находит нужные элементы на странице: место в рейтинге, название трека, исполнителей, альбом, длительность через поиск по названию класса (songs-list-row); содержание ненормативного контента через XPath; год через название тега (h1) и ссылки с атрибутом data-testid, равным track-seo-link. Затем функция обрабатывает полученные значения: извлекает текст, заменяет символы, переводит значение ненормативного контента в булеву переменную, переводит год в числовую переменную. И в конце формируется датафрейм с соответствующими столбцами. Функция возвращает датафрейм df, который содержит структурированные данные, собранные с каждой веб-страницы.

1.3 В заключении мы применяем функцию process_data к списку ссылок на плейлисты, собираем все нужные описанные выше переменные с каждой страницы и формируем единый датафрейм за 4 года.

2. Сбор данных с сайта <https://genius.com/>

2.1 Для начала мы собираем ссылки на нужные нам страницы с текстами песен. С помощью функции `get_lyrics_links()`, которая принимает на вход список названий песен, список имен исполнителей и url-ссылку, затем производит поисковой запрос при помощи Selenium на основании элементов списка, пробует забрать первую ссылку из выдачи по поисковому запросу в категории Songs и возвращает list со списком ссылок, по которым лежат тексты песен. В конце данного этапа у нас получается готовый список с ссылками lyrics links на нужные нам страницы с песнями.

2.2 Далее мы поочередно отправляем запросы через вебдрайвер к этим страницам и собираем интересующую нас информацию (текст и дата публикации песни). Это действие производится с помощью функции `scrape_lyrics()`, которая принимает на вход url-ссылку на индивидуальную страницу песни, собирает данные со страницы и возвращает два списка с данными соответственно (lyrics и date_ugly_list).

2.3 В заключении мы применяем функцию `scrape_lyrics()` ко всем ссылкам на страницы песен и обрабатываем исключения в случае, если какие-то данные отсутствуют на странице

7. Как эти данные могут быть использованы в целях соц.наук

Наш проект предполагает несколько способов, как можно использовать данные в сфере социальных наук (в особенности при дальнейшей его доработке, расширении выборки самих данных/источников и включения дополнительной методологии).

-Изучение культурных трендов: Данные о музыкальных предпочтениях могут служить индикатором культурных тенденций и изменений в обществе. Анализируя, какие исполнители, жанры, язык, смысловые элементы в песнях и их длительность становятся популярными в определенные периоды времени, исследователи в дальнейшем смогут сделать выводы о социальных и культурных сдвигах.

-Связь музыки с социальными и политическими событиями: Изучение корреляции между популярностью определенных песен, исполнителей, жанров и прочих элементов со значимыми социальными или политическими событиями, что может помочь выявить, как музыка отражает или реагирует на общественные изменения.

-*Анализ изменений языка в текстах песен*: изучение лексики, эмоциональной окраски и тематической составляющей песен может показать, как увеличивается или уменьшается количество конкретных использованных слов или фраз; как изменяется эмоциональная составляющая песен, их тональность; какие социальные вопросы становятся более актуальными.

-*Связь музыкальных предпочтений с психологической составляющей*: Исследования в области психологии могут использовать данные о музыкальных предпочтениях для изучения связей между музыкальным вкусом/потреблением и психологическими или когнитивными характеристиками (в этой области популярны исследования о связи депрессии с музыкальными вкусами или исследование о влиянии музыки на работу мозга).

8. Рефлексия

1. Дополнительный сбор данных:

-Для углубления анализа можно было бы собрать данные не только о топ-100 песнях, но и о более широком спектре, включая менее популярные песни, чтобы получить более полное представление о музыкальных тенденциях. Также можно было не ограничиваться топами от Apple music и дополнить их плейлистами из других стриминговых сервисов, например, Яндекс музыка или VK, которые на данный момент являются лидерами музыкального рынка в России.

-С третьего стороннего ресурса можно было бы собрать информацию о жанрах анализируемых нами треков (так как на сайтах Apple music и Genius эта информация не указана).

-Сбор данных о реакциях пользователей (например, комментарии, обсуждения) мог бы предоставить дополнительный контекст к популярности и восприятию песен в топ чартах.

-Кроме автоматизированного сбора реакций и обсуждений, для полноценного исследования можно было бы расширить методологию и провести опросы и интервью российских слушателей в разных возрастных категориях, что позволит более детально рассмотреть музыкальные предпочтения российских слушателей в контексте социально-политических кризисов и экстраполировать выводы на расширенную выборку.

2. Сложности и решения:

На первом этапе сложность возникла с первоначальным выбором стратегии сбора данных из плейлистов apple music. Сначала мы собирали отдельно все интересующие нас данные (то есть для каждой колонки собирали данные по отдельным элементам, это были отдельные CSS selectors): названия, исполнитель, длительность и альбомы. Такой вариант, как минимум, был неэффективен, и, как максимум, некорректно собирал столбец с исполнителями. Тогда мы поменяли подход, и решили написать функцию, которая бы проходила по всем трекам, нажимала на них, открывала страницу с каждым треком и собирала информацию со страницы. Но и этот подход не сработал, так как возникали сложности с переходом по кнопке “смотреть сведения об авторах” и возвращением обратно на страницу с плейлистом. Также этот вариант занимал бы много времени. В итоге мы нашли элемент, который для каждого трека в плейлисте сразу содержит практически всю нужную информацию (CLASS_NAME: songs-list-row') и включили его поиск и обработку в функцию process_data. Остальные данные (год, ненормативный контент и скрытую ссылку на песню) мы дособирали внутри функции с помощью обращения к различным элементам (для ссылок по CSS_SELECTOR, для года по TAG_NAME, для ненормативного контента по XPATH).

На втором этапе основная сложность заключалась в процессе сбора элементов с сайта Genius с помощью одной функции, которая проходила по списку собранных ранее названий треков и исполнителей, и по каждому треку вводила текстовый запрос (название песни + исполнитель), находила нужный трек в результатах поиска, кликала на него (при условии что таковой есть на сайте), переходила на страницу трека, собирала оттуда все нужные данные: текст, дату релиза, ссылку на страницу и возвращалась на предыдущую страницу браузера, чтобы повторить итерации для каждого наблюдения. Функция должна была возвращать кортеж, содержащий текст песни, дату и ссылки, и она работала исправно, однако, очень медленно. Один элемент собирался около 2-3 минут, то есть для 400 треков информация бы собиралась дольше 16 часов. Поэтому было принято решение разделить работу функции на 2 этапа: сначала отдельно собрать ссылки на страницы с песнями (то есть пройтись по спискам название треков+исполнитель, найти на странице нужный трек и поместить ссылку на него в список lyrics_link), а затем пройтись по списку ссылок и собрать нужные данные со страниц. Сбор ссылок занял примерно 4 часа,

обработка страниц примерно столько-же. Таким образом, удалось ускорить процесс сбора данных на этом этапе в 2 раза.

3. Будущее развитие проекта:

В будущем проект мог бы быть расширен за счет интеграции дополнительных данных (и источников) из первого пункта, что позволит провести полноценное исследование об изменении музыкальной сферы и предпочтений российских пользователей. Также с помощью включения моделей и методов машинного обучения можно добавить в наш проект прогностическую часть. Например, предсказания музыкальных тенденций, популярности определенных жанров и популярности определенных смысловых элементов в песнях.

4. Самый трудный раздел datasheet:

Наверное, самым трудным был раздел с описанием процесса сбора данных, так как этот процесс является особенно важным, поскольку он предоставляет полное понимание того, как данные были получены, и позволяет на первичном этапе оценить их качество, надежность и применимость. Также этот раздел является одним из самых объемных, так как включает в себя несколько содержательных подпунктов.

5. Этические дискуссии:

Персональные рекомендации многих сервисов вызывают вопросы о допустимости такого объема сбора данных и публикации результатов в целом. В качестве примера вспомним некоторые рекламные кампании музыкального стримингового сервиса Spotify. В 2017 году Spotify запустил рекламную кампанию, в которой использовал анонимизированные данные пользовательских плейлистов и прослушиваний. Один из рекламных слоганов гласил: "Dear person who played 'Sorry' 42 times on Valentine's Day, what did you do?"¹⁰. Хотя данные были анонимизированы, это вызвало дискуссию о том, где проходит граница между использованием агрегированных данных для маркетинга и нарушением частной жизни пользователей. Также можно вспомнить рекламную кампанию 2016 года, которая была запущена во время Брексита: "Dear 3,749 people who streamed 'It's The End Of The

¹⁰ <https://www.mirror.co.uk/news/weird-news/spotify-revealing-users-most-embarrassing-9378147>

World As We Know It ' the day of the Brexit Vote. Hang in There." Также в 2016 году людей по имени Пол ждал сюрприз от Spotify, который в рекомендациях предлагал плейлист: "Paul, You're Fired.", за которым следовало дополнительное шутовское описание на рекламных баннерах: Sorry you had to find out this way, Paul¹¹. Такими способами сервис пытается персонализировать пользовательский опыт прослушивания с помощью создания плейлистов и рекомендаций, в сочетании с использованием неформального тона высказываний и маркетинговых акций, однако, тут же и поднимаются вопросы о приватности и том, как данные могут быть использованы или потенциально раскрыты третьим сторонам.

Также в контексте этических дебатов можно рефлексировать по предмету публикации данных музыкальной платформы Last.fm. Публикуемые данные включают: всю информацию о треках, количество прослушиваний и информацию о социальном взаимодействии пользователей на платформе. На основе этих данных периодически проводятся исследования о музыкальных тенденциях, паттернах потребления музыки, социальных взаимодействиях в музыкальной среде. Конкретно никаких кейсов, связанных с некорректным сбором и использованием данных не наблюдалось, однако, в данном контексте можно задаться этическими вопросами о конфиденциальности данных, согласии пользователей, целях использования таких данных и возможного потенциального вреда.

¹¹ <https://www.linkedin.com/pulse/dear-person-who-played-sorry-42-times-valentines-day-spotify-young>