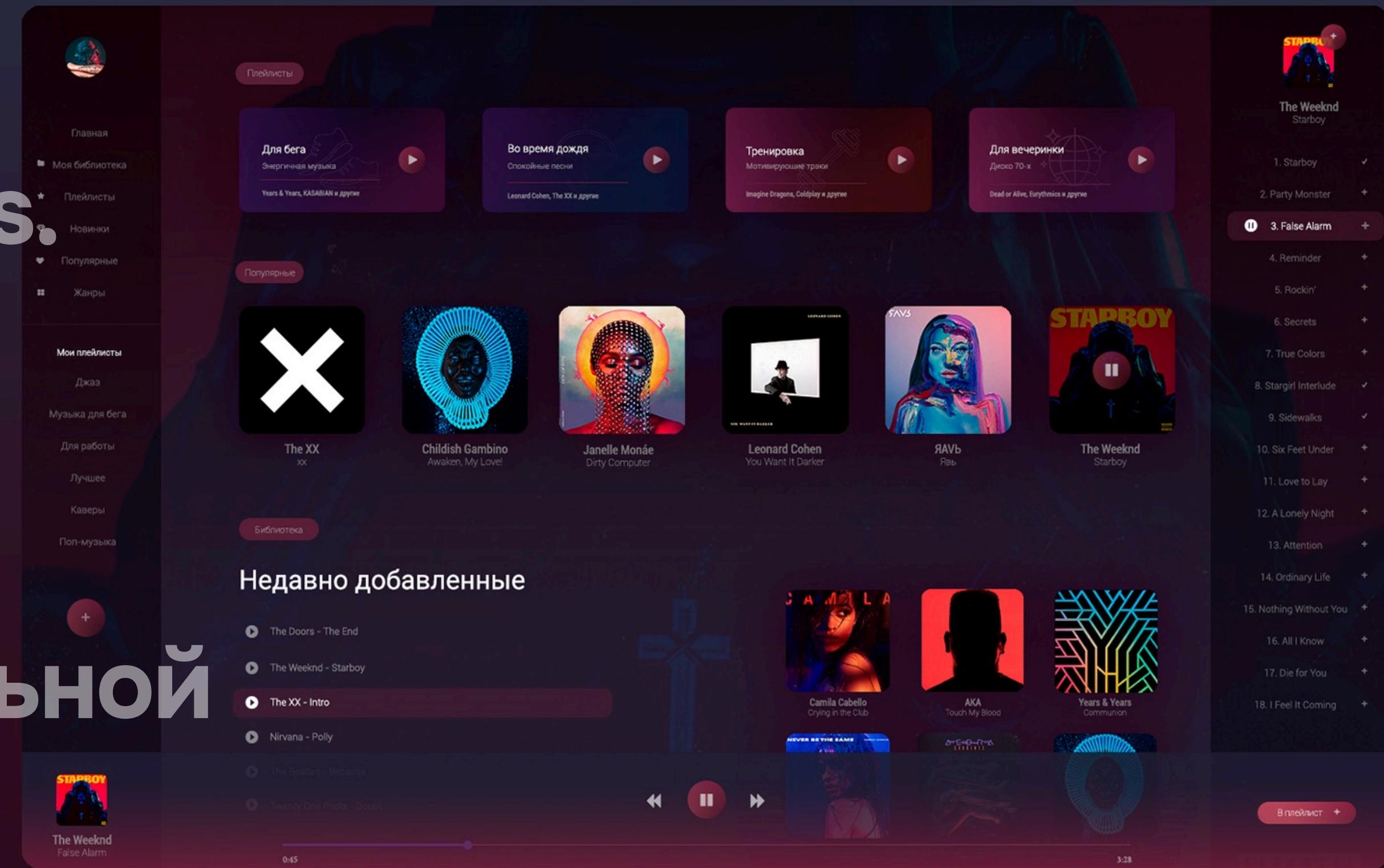


Групповой проект №2. Exploratory Data Analysis: Project Stardust: Data- Driven Подход К Поиску И Развитию Новой Глобальной Музыкальной Звезды



Команда 15: Хаявина Анна, Химич Анастасия, Бегаева Мария, Варнавская Анна, Пономарева Елизавета

В процессе сборания и обработки датасетов были добавлены 13 категорий для дальнейшего анализа

Merge и заполнение

Скрэпингом был собран файл **spotify_artists.csv**

На его основе были собраны дополнительные данные:
genius.csv - через API Genius

hitmotop_artist_tracks.csv - через скрэпинг Hitmotop
art_labels.csv и **spotify_artists.csv** - через API

Затем все данные были объединены в финальный файл **final_merged_artists_with_labels.csv**

hitmotop_artist_tracks.csv

Drake	1053	48128326	What's My Name?	4:24	https://rus.hitmotop.com/song/48128326
Drake	1053	48112888	Work	3:39	https://rus.hitmotop.com/song/48112888
Drake	1053	65044082	No Guidance	4:21	https://rus.hitmotop.com/song/65044082
Drake	1053	47965089	Hotline Bling	4:27	https://rus.hitmotop.com/song/47965089
Drake	1053	47965079	Too Good	4:23	https://rus.hitmotop.com/song/47965079

artist	- имя исполнителя
artist_id	- уникальный идентификатор артиста
followers	- количество подписчиков (в формате с десятичными разделителями)
genres	- музыкальные жанры (строкой или NaN)
popularity	- показатель популярности (0–100)
type	- тип записи (все "artist")

daily_streams	- ежедневные стримы
peak_daily	- пиковое значение дневных стримов
monthly_listeners	- месячные слушатели
total_page_views	- общее количество просмотров страницы
total_songs_count	- общее количество песен
songs_with_duration_count	- количество песен с указанной длительностью
labels	- лейблы звукозаписи

spotify_artists.csv

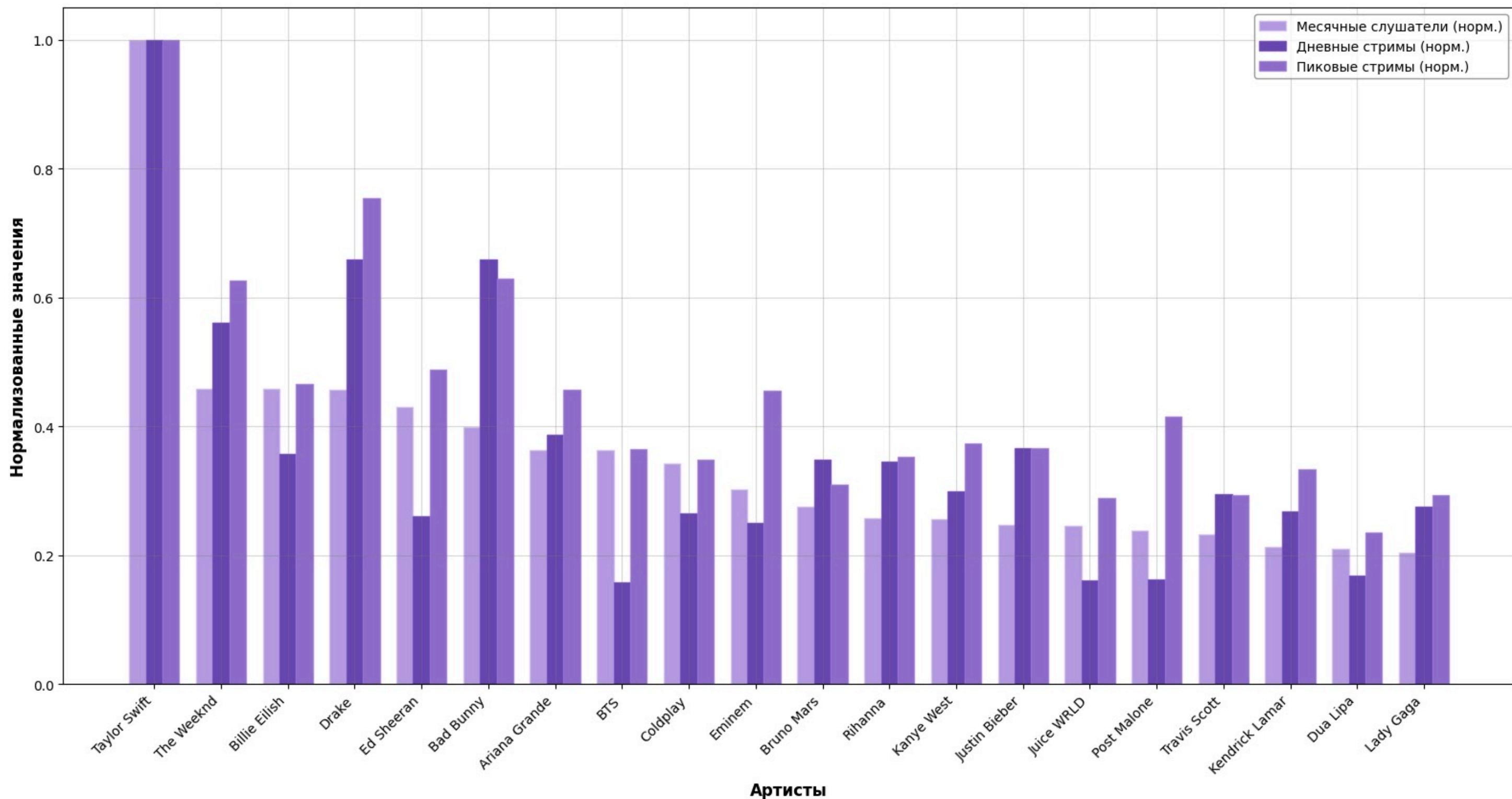
Drake	122,491.8	48,224	83,627.9	46,307.8
Taylor Swift	114,464.0	73,087	110,924.1	101,320.6
Bad Bunny	108,856.3	48,179	69,861.9	40,392.9
The Weeknd	86,408.1	40,980	69,564.2	46,496.7
Justin Bieber	65,506.9	26,747	40,668.3	24,976.2
Ariana Grande	60,848.3	28,300	50,709.6	36,843.6
Ed Sheeran	60,704.6	19,026	54,089.1	43,648.0
Travis Scott	60,193.5	21,593	32,599.6	23,495.8

genius.csv

10	58325	The Motto	Drake	3885094	https://genius.com/Drake-the-motto-lyrics
11	218330	Pound Cake / Paris Morton	Drake	3851674	https://genius.com/Drake-pound-cake-paris-morton-music-2-lyrics
12	156640	0 to 100 / The Catch Up	Drake	3833648	https://genius.com/Drake-0-to-100-the-catch-up-lyrics
13	2890553	Fake Love	Drake	3819887	https://genius.com/Drake-fake-love-lyrics
14	50895	Marvins Room	Drake	3614780	https://genius.com/Drake-marvins-room-lyrics
15	7076626	All Too Well (10 Minute Ver	Taylor Swift	6135146	https://genius.com/Taylor-swift-all-too-well-10-minute-version-taylors-version-from-the-vault-lyrics
16	12274793	Wood	Taylor Swift	4440731	https://genius.com/Taylor-swift-wood-lyrics
17	7394358	All Too Well (10 Minute Ver	Taylor Swift	4281449	https://genius.com/Taylor-swift-all-too-well-10-minute-version-taylors-version-live-acoustic-lyrics
18	10024009	Fortnight	Taylor Swift	4157873	https://genius.com/Taylor-swift-fortnight-lyrics

Анализ метрик лидеров рынка для формирования портрета суперзвезды

Топ-20 артистов: сравнение слушателей и стримов



Базовые критерии супер артиста

column	mean_value
followers	39574684.06
popularity	81.58
daily streams	14.40
monthly listeners	16907.65
total_page_views	36371280.28

На графике две разные модели успеха в современной музыкальной индустрии:

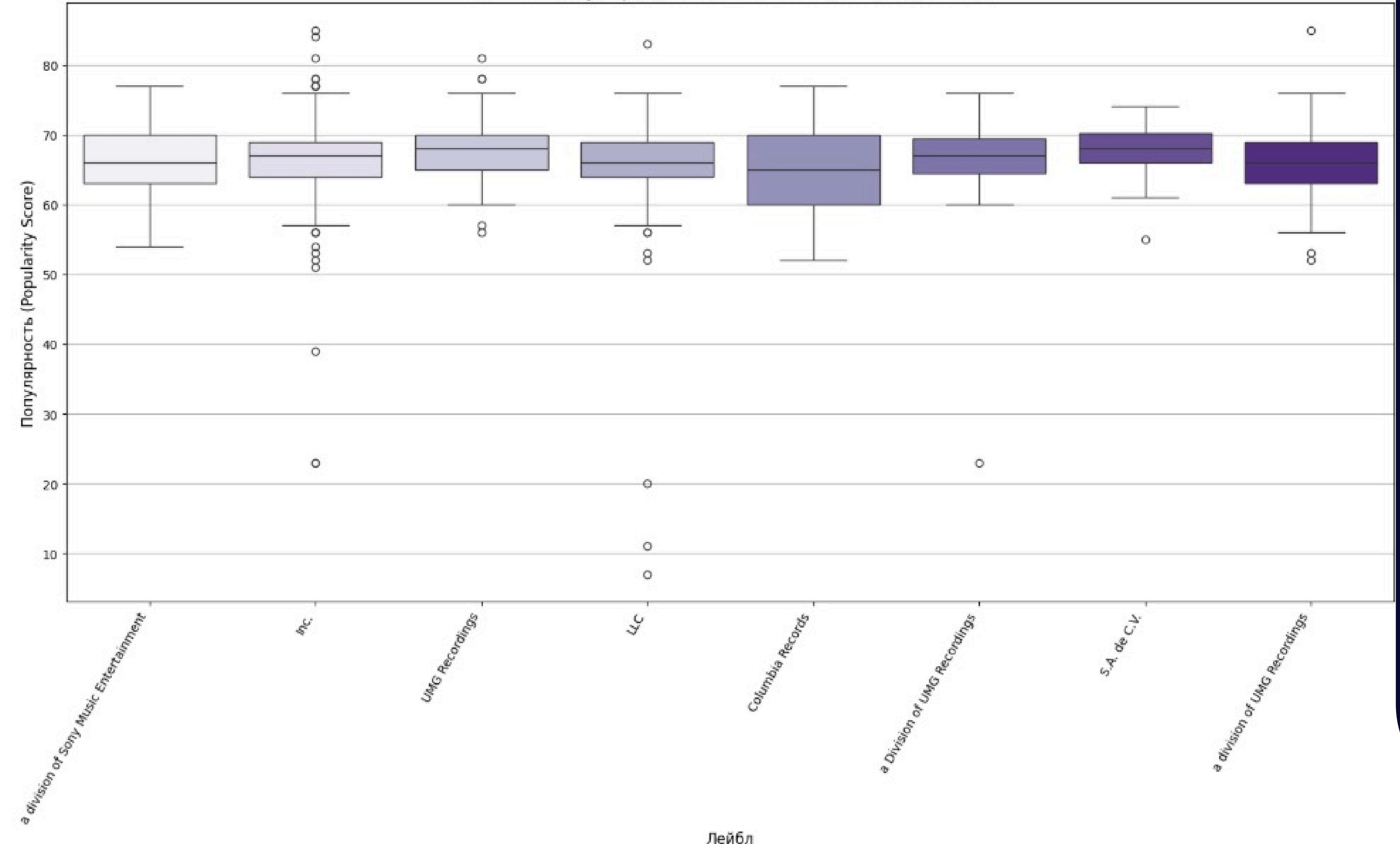
1. Широкая популярность: Большая, но менее вовлеченная аудитория.

2. Глубокая преданность: Меньшая по размеру, но гиперактивная и преданная фанбаза, способная генерировать рекордные всплески активности.

В большинстве случаев, если дневные стримы меньше месячных слушающих, то пиковые стримы выше по нормировочному значению, чем месячные слушатели. Такая тенденция у 8 артистов из 20.

Анализ мировой музыкальной индустрии

Анализ популярности по ТОП-8 Лейблам (Box Plot)



Эффективность Лейблов (Popularity)

1. 🏆 Лидеры по Стабильности (Медиана)

- Columbia Records и Capitol Records демонстрируют самую высокую медиану популярности.
- Вывод для слайда: Эти лейблы являются "тяжеловесами", обеспечивая наиболее высокий и стабильный успех своих артистов.

2. ⚡ Максимальный Потенциал

"Суперзвезд" (Максимум)

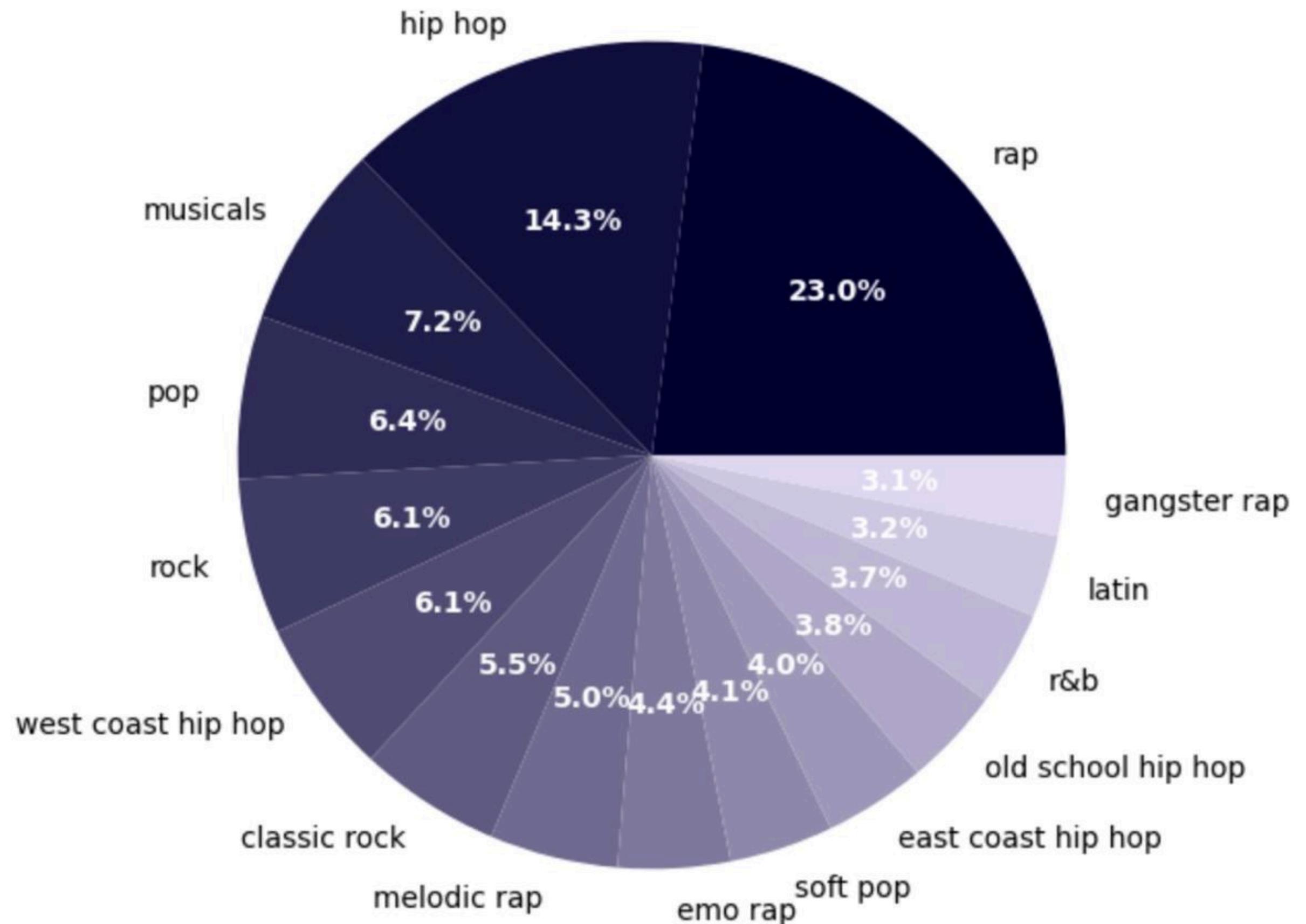
- Interscope Records, RCA Records и Atlantic Records имеют артистов с максимальным рейтингом популярности (≈ 100).
- Вывод для слайда: Это ключевые "производители" мировых суперзвезд с потенциалом достижения максимальной вовлеченности.

3. 💸 Фокус для Инвестиций

- Лейблы, сочетающие высокую медиану и высокий максимум (например, Columbia и Interscope), предлагают наилучший баланс

Анализ мировой музыкальной индустрии в топах жанров

Топ-15 жанров по общему количеству просмотров



Rauw Alejandro
reggaeton



Fuerza Regida
banda



Anuel AA
trap latino



KPop Demon Hunters Cast
k-pop



Lady Gaga
art pop



Morgan Wallen
country



Arijit Singh
bollywood



Sabrina Carpenter
pop



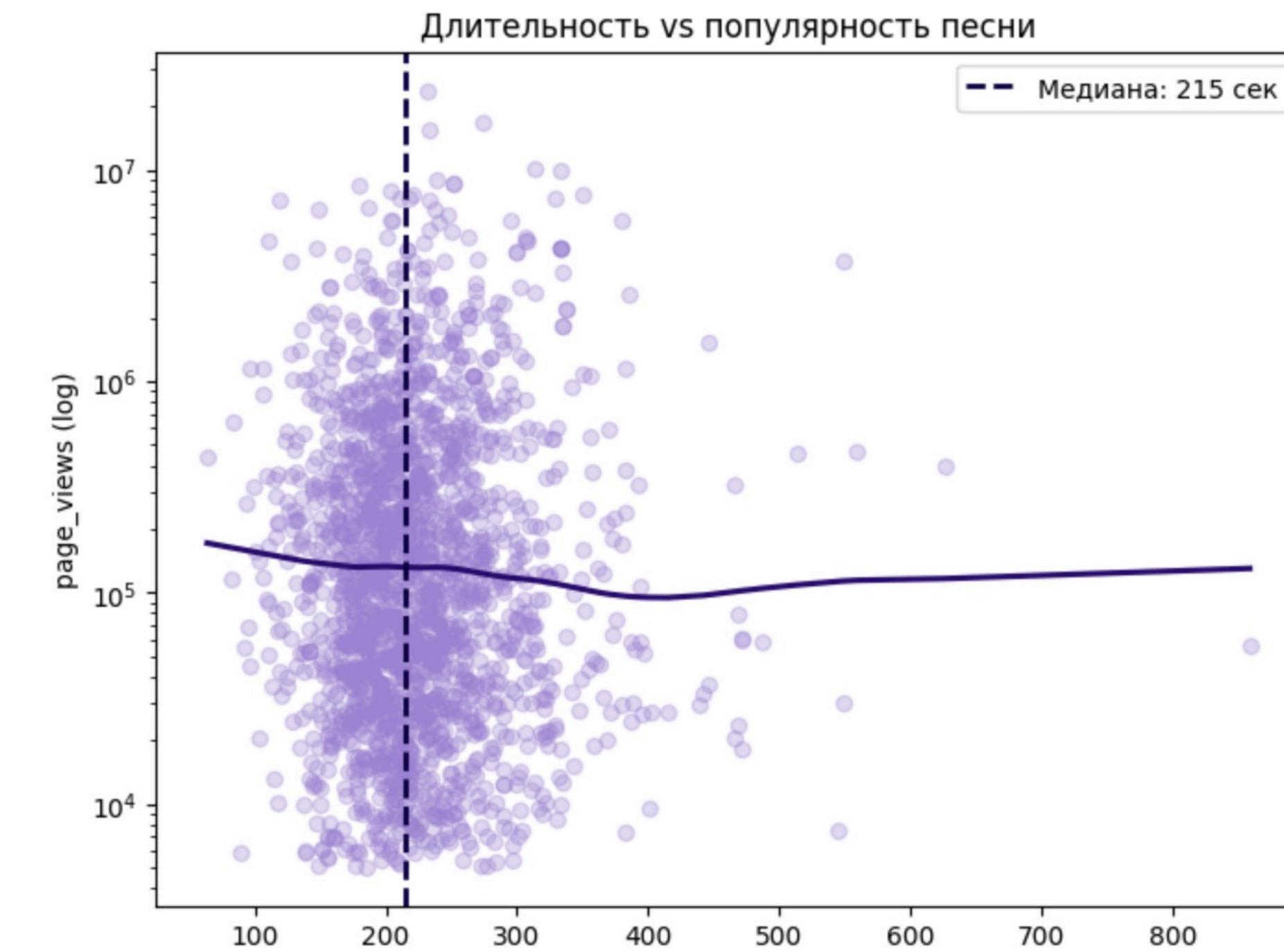
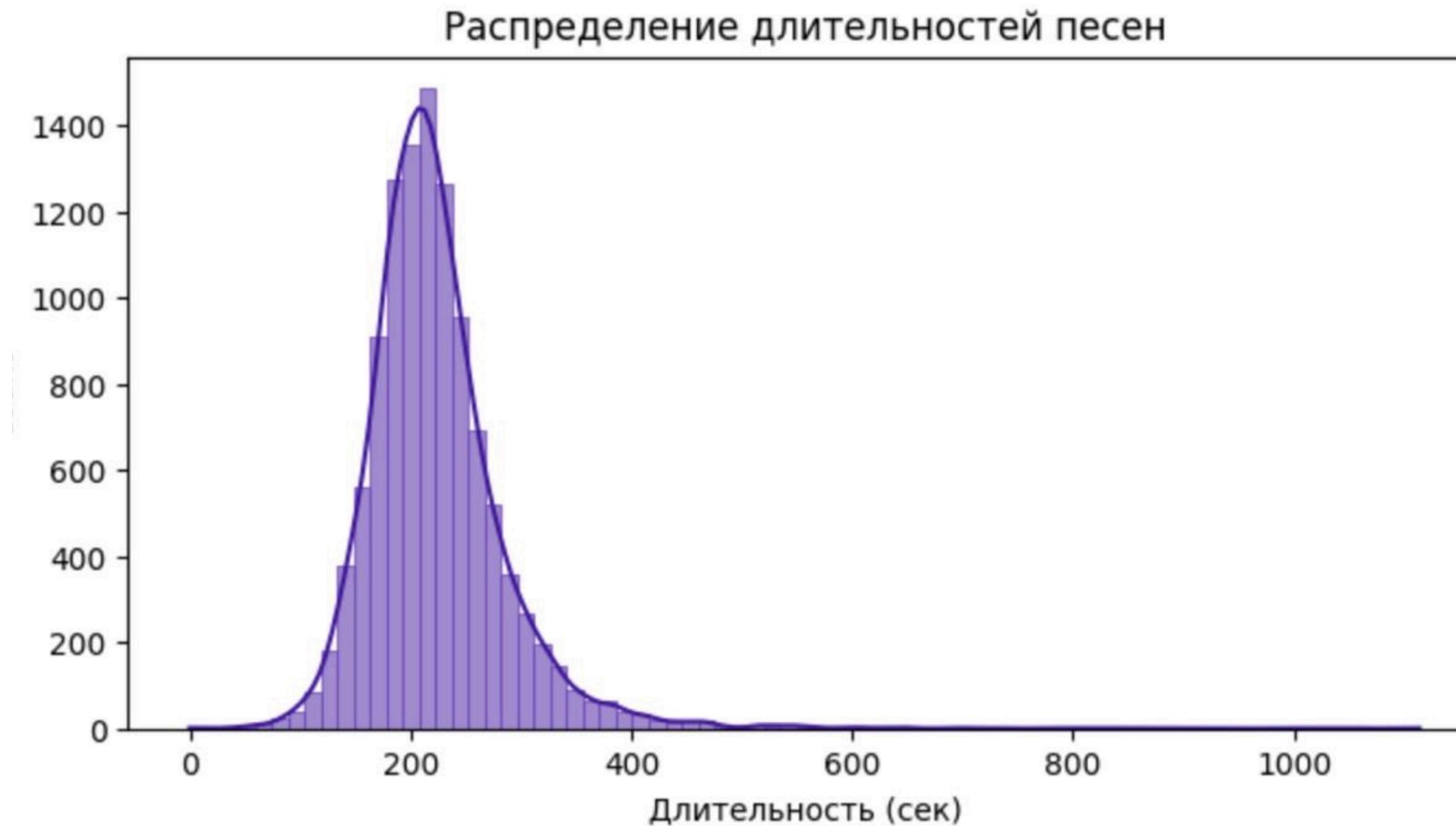
Bad Bunny
latin



Eminem
hip hop

Топ-15 самых популярных жанров:
Топ начинается с классического рэпа
Дальше идут представители классических поджанров: hip hop, musicals, pop, rock
Далее следуют нишевые поджанры.
В числе последних исключительные жанры: latin и r&b.

Анализ мировой музыкальной индустрии в длительности песен

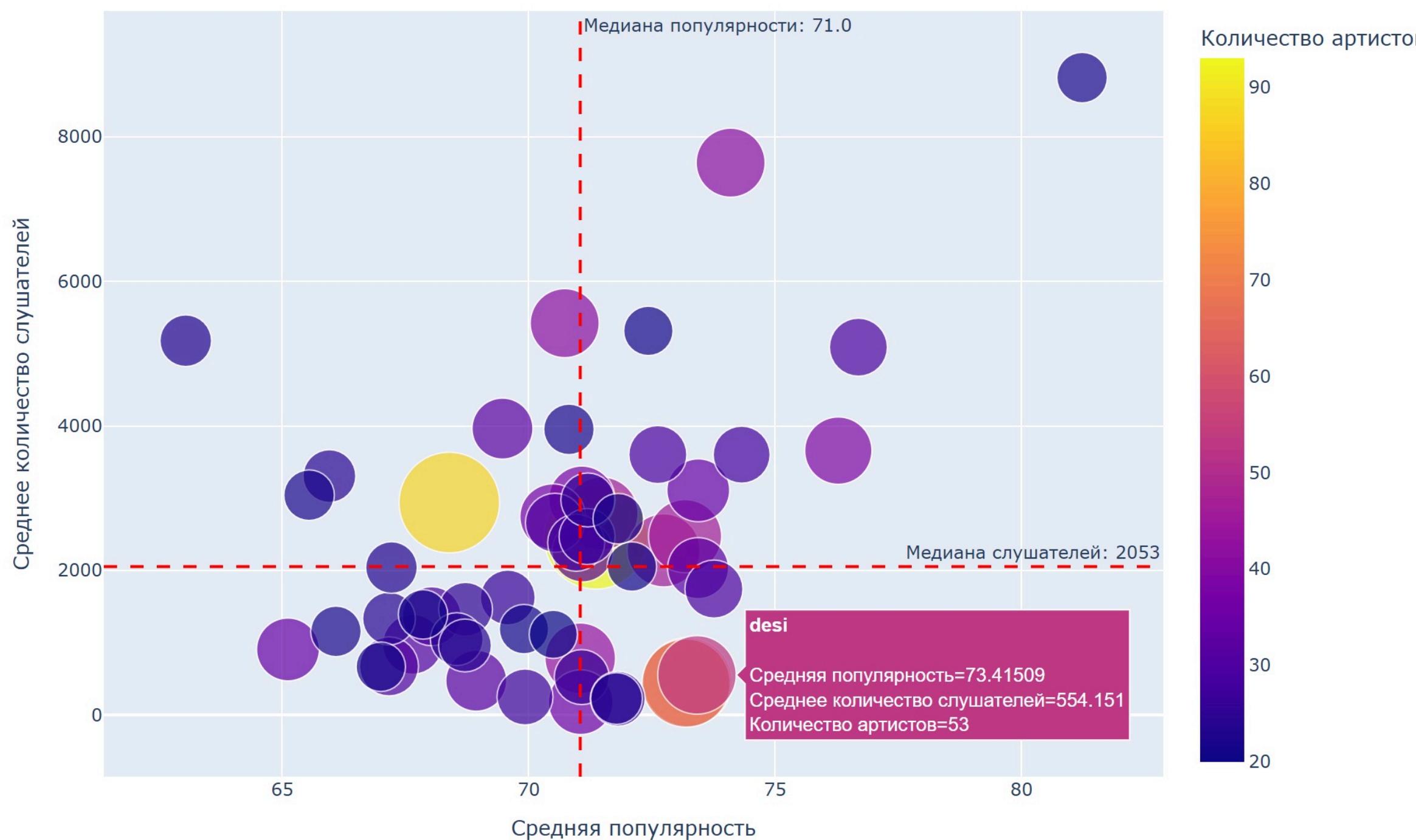


Поскольку наибольшее количество треков имеют продолжительность в окрестности 200 секунд на 1 графике, концентрация на втором графике идёт вдоль средней линии по длительности. Но как показано на графике, здесь отсутствует корреляция, т.е. отсутствует тенденция зависимости популярности от длительности.

Анализ мировой музыкальной индустрии

Выбранный квадрант (высокая популярность >71, малая аудитория <2000) — это зона незанятых ниш с высоким потенциалом.

Соотношение популярности и слушателей по жанрам (минимум 20 артистов)
Размер пузырька - количество артистов в жанре



Низкая конкуренция: Мало слушателей = мало артистов, которые доминируют в этом жанре.
Высокий спрос: Популярность жанра >75 означает, что музыка востребована широкой аудиторией, но сама ниша не перенасыщена контентом.
Эффект новизны: Выход сильного артиста в таком жанре вызовет взрывной интерес.
Целевой квадрат: правый верхний (слушателей больше среднего, популярность выше среднего)

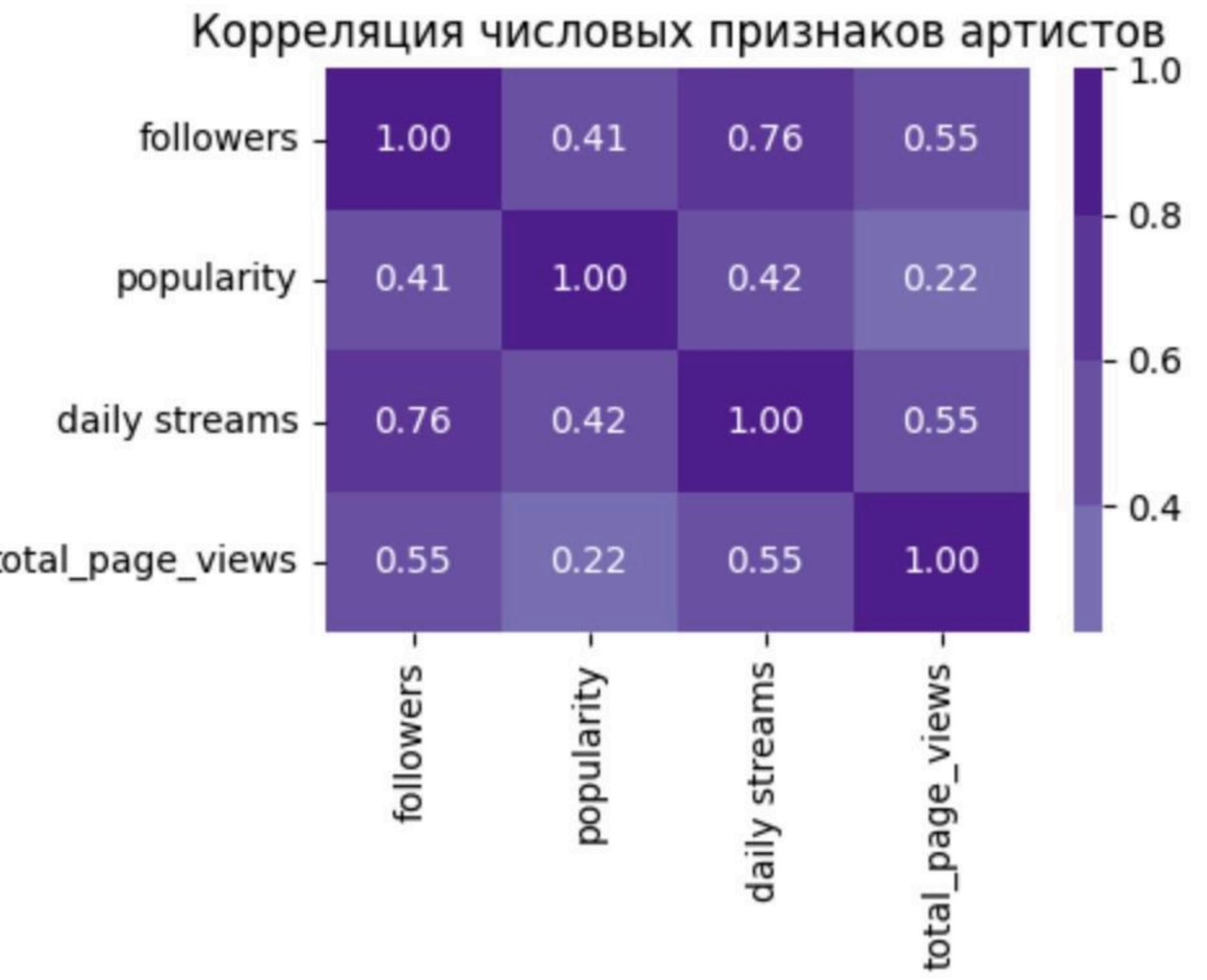
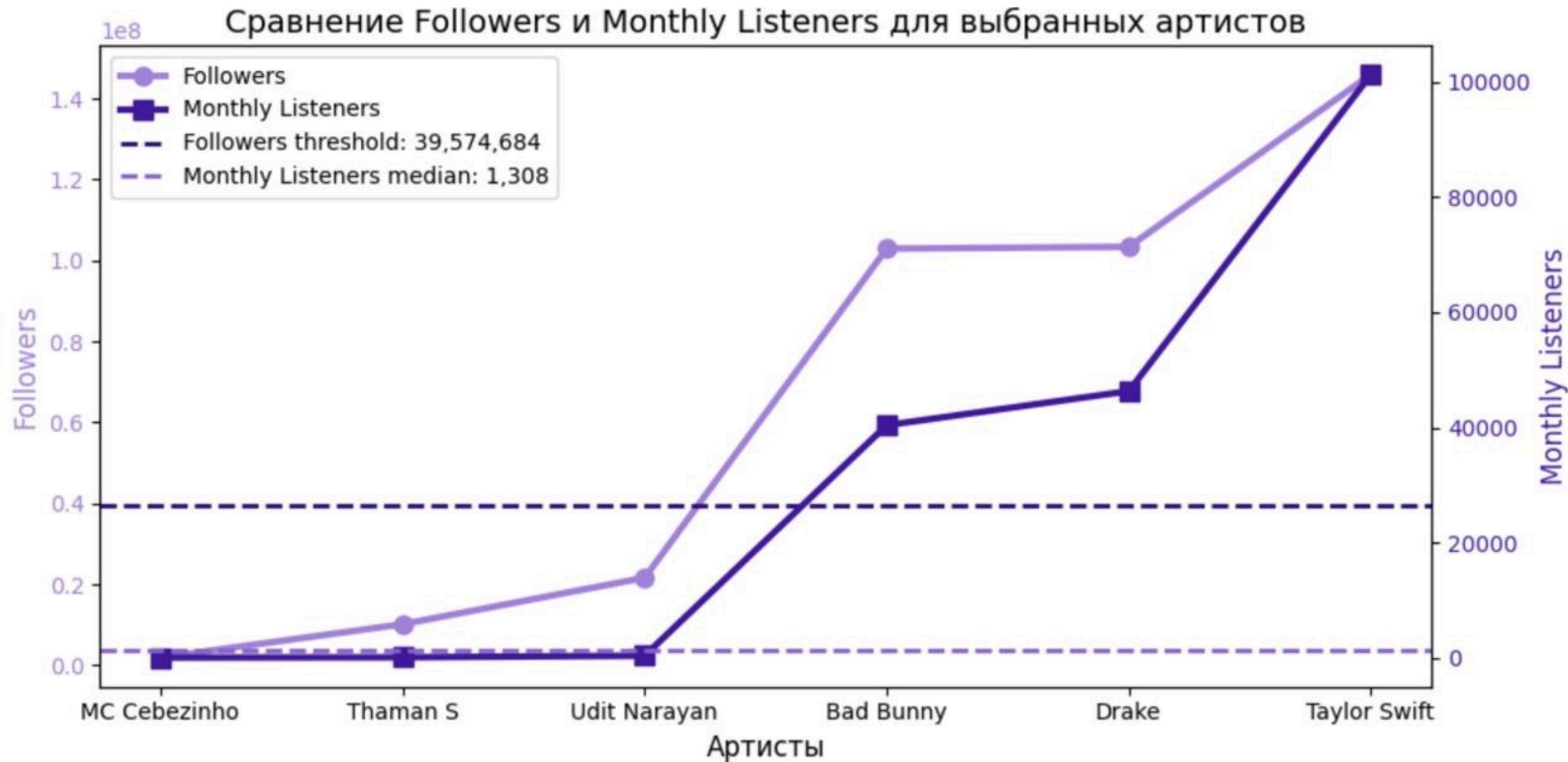
Топ-артисты по потенциальному
найти "недооцененных" артистов (высокая популярность при малом числе подписчиков)

	artist	genres	popularity	followers	monthly listeners
2361	Victor Mendivil	corridos bélicos	84.0	2481445.0	606.6
2337	Udit Narayan	bollywood	81.0	21661647.0	342.2
2337	Udit Narayan	hindi pop	81.0	21661647.0	342.2
1961	Sachin-Jigar	desi	81.0	7737399.0	17.7
2181	Thaman S	tamil pop	74.0	10187443.0	62.9
2181	Thaman S	kollywood	74.0	10187443.0	62.9
1429	MC Hariel	trap funk	73.0	7833402.0	485.8
1425	MC Cebezinho	brazilian funk	72.0	1946581.0	18.6
554	Darshan Raval	desi pop	72.0	26523343.0	2363.7
767	FiNCH	desi hip hop	12.0	5388.0	494.3

1. Аномально высокий коэффициент популярности к слушателям
Это ядро отбора. (например, >70) при относительно низком monthly listeners (например, <2500). Явные примеры из таблицы:
Sachin-Jigar: Популярность 81 при всего 17.7 тыс. слушателей — идеальный пример.
MC Cebezinho: 72 популярности и 18.6 тыс. слушателей.
Victor Mendivil: 84 популярности (очень высоко) и 606 тыс. слушателей (мало для такого уровня популярности).

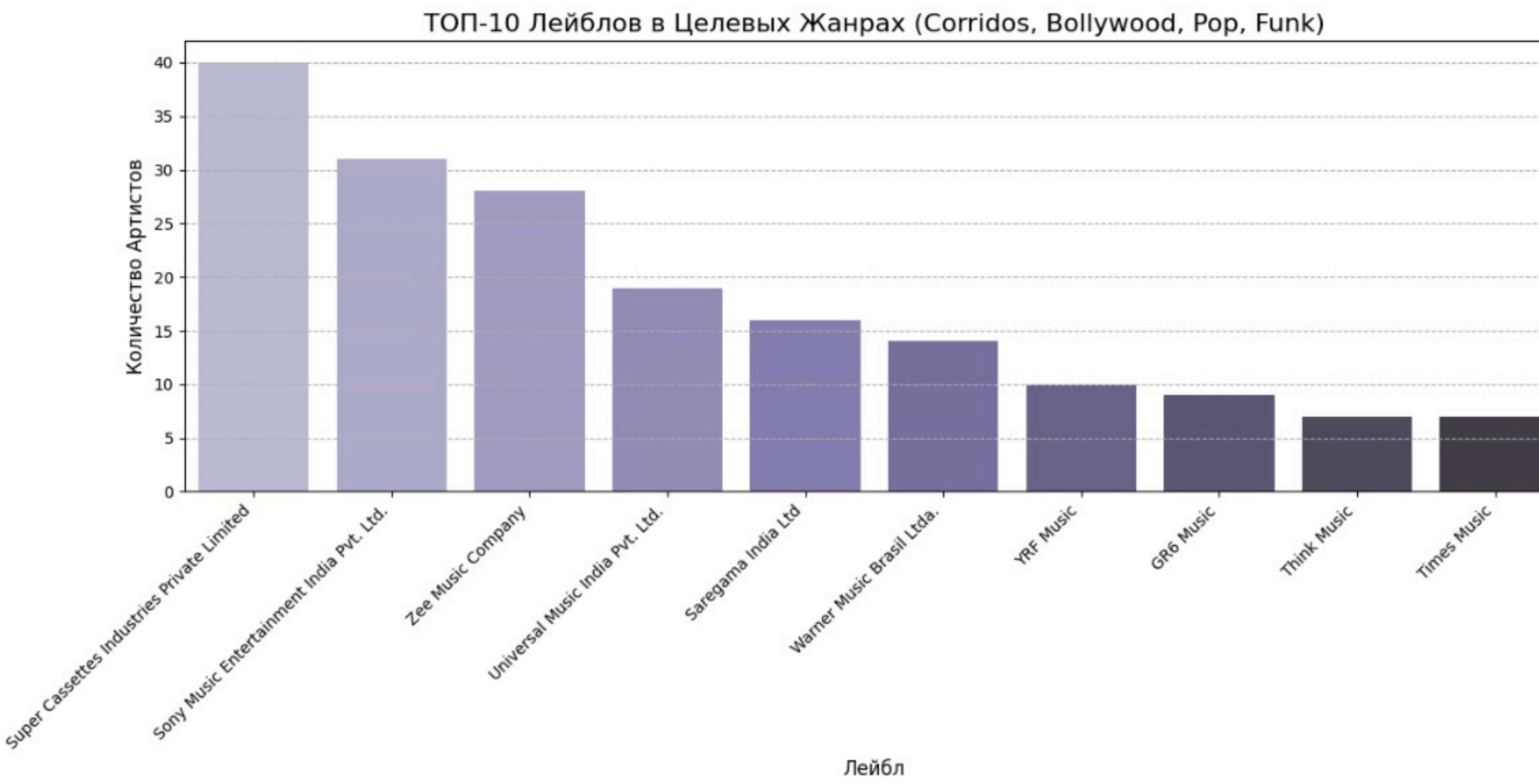
2. Анализ на абсолютных цифрах
Артисты с очень высоким абсолютным числом подписчиков (followers), но низким количеством текущих слушателей, могут считаться "легендами", чья текущая активность низка, но популярность остается на высоком уровне.
Пример:
Udit Narayan: 21 млн подписчиков, но лишь 342 тыс. текущих слушателей при популярности 81. Это артист с огромной базой, но не самой активной на текущий момент.

Анализ разницы потенциалов



Для попадания в суперартисты необходимо пройти 2 границы: по подписчикам и месячным слушателям, при этом по графику видно, что границу подписчиков пройти проще. Корреляции показывают, что сильнее всего с подписчиками связаны ежедневные стримы (≈ 0.76), тогда как «популярность» коррелирует с подписчиками умеренно (≈ 0.41), поэтому приоритет - daily streams, но также важно параллельно конвертировать охват в фолловеров . Такой фокус ускорит рост Monthly Listeners и повысит заметность в алгоритмах, что быстрее приблизит к статусу суперартиста

Топ лейблов на потенциальный контракт



	label	super artist	popularity
0	Spooniversal Studios	Lady A (89)	89
1	Universal Music Operations Limited	Olivia Dean (89)	89
2	ChrisRodMusic	Lady A (89)	89
3	under exclusive licence to Universal Music Ope...	Olivia Dean (89)	89
4	Sam Fender	Olivia Dean (89)	89
5	Roc Nation Distribution	sombr (85)	85
6	Jack & Jack LLC	Jack & Jack (85)	85
7	a division of UMG Recordings	Jack & Jack (85)	85
8	Dynamic Records / EMPIRE	sombr (85)	85
9	CORILLO LLC	sombr (85)	85
10	Island Records	Jack & Jack (85)	85
11	Atlantic Recording Corporation	Alex Warren (85)	85
12	La Funda	sombr (85)	85
13	LLC4 Records	sombr (85)	85
14	Jack & Jack LLC / Create Music Group Inc.	Jack & Jack (85)	85

Введение

Анализ

Инсайты

Рекомендации

Команда

Ключевые инсайты

Введение

Анализ

Инсайты

Рекомендации

Команда

Развития новой глобальной звезды

Команда



Анастасия Химич



Мария Бегаева



Елизавета Пономарева



Анна Варнавская



Анна Халявина