

Лучшие хиты Spotify с 2000 по 2019 год

курсовой проект Кречко Татьяны
01.08.2024





Сегодня Spotify является самым популярным глобальным сервисом потокового аудио с 365 миллионами пользователей, в том числе 165 миллионами подписчиков на 178 рынках. На сегодняшний день они являются крупнейшим источником дохода для музыкального бизнеса.

Цели и задачи проекта:

Изучить датасет и выявить зависимость популярности:

- от года, от количества написанных песен одним исполнителем, от показателей, указанных в датасете таких как акустичность, танцевальность, жанр, инструментальность и другие, тем самым определить – есть ли зависимость популярности от других характеристик.
- определить самые популярные песни, самых популярных исполнителей.

СОДЕРЖАНИЕ

1. ОБНОВЛЕНИЕ БИБЛИОТЕК, ЗАГРУЗКА ДАННЫХ, ЧИТСКА ДАННЫХ в Jupyter Notebook

2. ИНФОРМАЦИЯ О ДАННЫХ в Jupyter Notebook

3. ИССЛЕДОВАТЕЛЬСКИЙ АНАЛИЗ ДАННЫХ, Построение графиков в Jupyter Notebook

4. ВИЗУАЛИЗАЦИЯ ДАННЫХ В POWER BI

5. ВЫВОДЫ

Описание данных:

- artist: Имя исполнителя.
- song: (песня): Название трека.
- duration_ms: Длительность трека в миллисекундах.
- explicit: (явный, слишком откровенный): Текст или содержание песни или музыкального видео содержат один или несколько критериев, которые могут быть сочтены оскорбительными или неподходящими для детей.
- year: Год выпуска трека.
- popularity (популярность): Чем выше значение, тем популярнее песня.
- danceability (Танцевальность): Танцевальность описывает, насколько трек подходит для танцев на основе комбинации музыкальных элементов, включая темп, стабильность ритма, силу бита и общую регулярность. Значение 0,0 — наименее танцевальное, а 1,0 — наиболее танцевальное.
- energy (энергия): Энергия является мерой от 0,0 до 1,0 и представляет собой перцептивную (процесс отражения событий) меру интенсивности и активности.
- key: (ключ) Тональность, в которой находится трек.. Если ключ не обнаружен, значение равно -1.
- loudness: (громкость) Общая громкость трека в децибелах (дБ). Значения громкости усредняются по всей дорожке и полезны для сравнения относительной громкости дорожек. Громкость – это качество звука, которое является первичным психологическим коррелятом физической силы (амплитуды). Значения обычно находятся в диапазоне от -60 до 0 дБ.
- mode: (модальность) Mode указывает модальность (мажорную или минорную) трека, тип гаммы, из которой получено его мелодическое содержание. Мажор обозначается цифрой 1, а минор — 0.

- **Speechiness**: (определяет наличие произнесенных слов в треке). Чем больше запись похожа исключительно на речь (например, ток-шоу, аудиокнига, поэзия), тем ближе к 1.0 значение атрибута. Значения выше 0,66 описывают треки, которые, вероятно, полностью состоят из произнесенных слов. Значения от 0,33 до 0,66 описывают треки, которые могут содержать как музыку, так и речь, как в разделах, так и в слоях, включая такие случаи, как рэп-музыка. Значения ниже 0,33, скорее всего, представляют музыку и другие треки, не похожие на речь.
- **acousticness** (Акустическость) Достоверная мера от 0.0 до 1.0 того, является ли трек акустическим. 1.0 означает высокую уверенность в том, что трек является акустическим.
- **instrumentalness** (инструментальность): предсказывает, содержит ли трек вокал. Звуки «ох» и «ааа» в этом контексте рассматриваются как инструментальные. Рэп или треки с устной речью явно «вокальные». Чем ближе значение инструментальности к 1.0, тем больше вероятность того, что трек не содержит вокального контента. Значения выше 0,5 предназначены для представления инструментальных треков, но достоверность тем выше, чем ближе значение к 1,0.
- **liveness**: Определяет присутствие аудитории в записи. Более высокие значения живости означают повышенную вероятность того, что трек был исполнен в прямом эфире. Значение выше 0,8 обеспечивает высокую вероятность того, что трек находится в прямом эфире.
- **valence** (валентность): Мера от 0,0 до 1,0, описывающая музыкальную позитивность, передаваемую треком. Треки с высокой валентностью звучат более позитивно (например, счастливый, веселый, эйфорический), в то время как треки с низкой валентностью звучат более негативно (например, грустный, подавленный, сердитый). эйфорический), в то время как треки с низкой валентностью звучат более негативно (например, грустный, подавленный, сердитый). грустный, подавленный, сердитый).
- **tempo**: Общий расчетный темп трека в ударах в минуту (BPM). В музыкальной терминологии темп — это скорость или темп данного произведения, который напрямую зависит от средней длительности удара.
- **genre**: Жанр трека.

Проанализируем топ-5 самых популярных исполнителей (по частоте встречаемости)

- вывести топ-5 (по распространенности) исполнителей
- среди топ 5 исполнителей выбрать самое большое значение популярности их песни

```
In [20]: # вывести топ-5 (по распространенности) исполнителей
df1.artist.value_counts()
```

```
Out[20]: artist
Rihanna      22
Drake         20
Calvin Harris 20
Eminem        19
Britney Spears 19
```

```
In [22]: #среди топ 5 самых распространенных исполнителей выбрать самое большое значение популярности их песни
top5_popular = allsong \
    .groupby(['artist'], as_index=False) \
    .aggregate({'popularity': 'max'}) \
    .query("artist == 'Rihanna' or artist == 'Drake' or artist == 'Eminem' or artist == 'Calvin Harris' or artist == 'Britney Spears'") \
    .sort_values('popularity', ascending=False)
top5_popular
```

```
Out[22]:
```

	artist	popularity
3	Eminem	87
2	Drake	84
0	Britney Spears	81
1	Calvin Harris	81
4	Rihanna	81

Выведем таблицу с названиями самых популярных песен у топ-5 самых распространённых исполнителей

Out[30]:

	artist	song	duration_ms	explicit	year	popularity	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness
201	Eminem	Without Me	290320	True	2002	87	0.908	0.669	7	-2.827	1	0.0738	0.00286	0.000000	0.2
1601	Drake	One Dance	173986	False	2016	84	0.792	0.625	1	-5.609	1	0.0536	0.00776	0.001800	0.3
402	Britney Spears	Toxic	198800	False	2003	81	0.774	0.838	5	-3.914	0	0.1140	0.02490	0.025000	0.2
1812	Calvin Harris	One Kiss (with Dua Lipa)	214846	False	2018	81	0.791	0.862	9	-3.240	0	0.1100	0.03700	0.000022	0.0
701	Rihanna	Umbrella	275986	False	2008	81	0.583	0.829	1	-4.603	1	0.1340	0.00864	0.000000	0.0



Выведем «формулу идеального трека» объединив две таблицы по Топ-5 самых распространённых артистов и топ-5 самых популярных песен

top5_artists																
Out[30]:																
	artist	song	duration_ms	explicit	year	popularity	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	
201	Eminem	Without Me	290320	True	2002	87	0.908	0.669	7	-2.827	1	0.0738	0.00286	0.000000	0.2	
1601	Drake	One Dance	173986	False	2016	84	0.792	0.625	1	-5.609	1	0.0536	0.00776	0.001800	0.3	
402	Britney Spears	Toxic	198800	False	2003	81	0.774	0.838	5	-3.914	0	0.1140	0.02490	0.025000	0.2	
1812	Calvin Harris	One Kiss (with Dua Lipa)	214846	False	2018	81	0.791	0.862	9	-3.240	0	0.1100	0.03700	0.000022	0.0	
701	Rihanna	Umbrella	275986	False	2008	81	0.583	0.829	1	-4.603	1	0.1340	0.00864	0.000000	0.0	

top5_song																
Out[38]:																
	artist	song	duration_ms	explicit	year	popularity	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	
1322	The Neighbourhood	Sweater Weather	240400	False	2013	89	0.612	0.807	10	-2.810	1	0.0336	0.04950	0.017700	0.0	
1311	Tom Odell	Another Love	244360	True	2013	88	0.445	0.537	4	-8.532	0	0.0400	0.69500	0.000017	0.0	
201	Eminem	Without Me	290320	True	2002	87	0.908	0.669	7	-2.827	1	0.0738	0.00286	0.000000	0.2	
6	Eminem	The Real Slim Shady	284200	True	2000	86	0.949	0.661	5	-4.244	0	0.0572	0.03020	0.000000	0.0	
1819	Billie Eilish	lovely (with Khalid)	200185	False	2018	86	0.351	0.296	4	-10.109	0	0.0333	0.93400	0.000000	0.0	

```
merged_df = pd.merge(top5_artists, top5_song, on= 'artist')
```

merged_df

	artist	song_x	duration_ms_x	explicit_x	year_x	popularity_x	danceability_x	energy_x	key_x	loudness_x	mode_x	speechiness_x	acousticness_x	ins
0	Eminem	Without Me	290320	True	2002	87	0.908	0.669	7	-2.827	1	0.0738	0.00286	
1	Eminem	Without Me	290320	True	2002	87	0.908	0.669	7	-2.827	1	0.0738	0.00286	

Формула идеального трека: Самый часто встречающийся исполнитель с самой популярной песней - Eminem "Without Me"

Eminem
without me



Анализ Топ-10 песен по популярности (наиболее популярные)

Out[57]:

	artist	song	duration_ms	explicit	year	popularity	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness
1322	The Neighbourhood	Sweater Weather	240400	False	2013	89	0.612	0.807	10	-2.810	1	0.0336	0.04950	0.017700
1311	Tom Odell	Another Love	244360	True	2013	88	0.445	0.537	4	-8.532	0	0.0400	0.69500	0.000017
201	Eminem	Without Me	290320	True	2002	87	0.908	0.669	7	-2.827	1	0.0738	0.00286	0.000000
1819	Billie Eilish	lovely (with Khalid)	200185	False	2018	86	0.351	0.296	4	-10.109	0	0.0333	0.93400	0.000000
1613	WILLOW	Wait a Minute!	196520	False	2015	86	0.764	0.705	3	-5.279	0	0.0278	0.03710	0.000019
6	Eminem	The Real Slim Shady	284200	True	2000	86	0.949	0.661	5	-4.244	0	0.0572	0.03020	0.000000
1514	Avicii	The Nights	176658	False	2014	85	0.527	0.835	6	-5.298	1	0.0433	0.01660	0.000000
1512	The Neighbourhood	Daddy Issues	260173	False	2015	85	0.588	0.521	10	-9.461	1	0.0329	0.06780	0.149000
1227	Bruno Mars	Locked out of Heaven	233478	False	2012	85	0.726	0.698	5	-4.165	1	0.0431	0.04900	0.000000
1999	Post Malone	Circles	215280	False	2019	85	0.695	0.762	0	-3.497	1	0.0395	0.19200	0.002440

Проведем сравнительный анализ средних значений показателей топ-10 песен и средних значений показателей по всем песням в целом по датасету.

- Создадим новую таблицу со средними показателями, где
- столбец «0» - среднее значение показателя по топ-10 песен,
 - столбец «1» – среднее значение по всему датасету,
 - столбец «%» - на сколько % увеличилось/уменьшилось значение

Out[90]:

	0	1	%
duration_ms	234157.400000	228647.375200	2.0
popularity	86.200000	59.517829	45.0
popularity	86.200000	59.517829	45.0
danceability	0.656500	0.668402	-2.0
energy	0.649100	0.722988	-10.0
key	5.400000	5.351783	1.0
mode	0.600000	0.554018	8.0
speechiness	0.042450	0.104263	-59.0
acousticness	0.207406	0.128630	61.0
instrumentalness	0.016918	0.015738	7.0
liveness	0.143440	0.182079	-21.0
valence	0.515400	0.554258	-7.0
tempo	115.488200	120.201733	-4.0

Вывод средние значения **популярности** на 45 % выше у топ-10 песен, чем по всему датасету; значение **аккустичности** на 61 % выше у топ-10 песен, чем по всему датасету; И, наоборот, среднее значение **speechiness** (определяет наличие среднее произнесенных слов в треке) выше в среднем по всему датасету чем по топ-10 песен на 59%, т.е. данный показатель говорит нам о том, что чем меньше произнесенных слов в треке – тем выше популярность.

Исследовательский анализ данных, графики

#Построение корреляционной матрицы

Out[249]:

	duration_ms	popularity	duration_ms	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness
duration_ms	1.000	0.051	1.000	-0.060	-0.079	-0.003	-0.080	-0.004	0.067	0.011	-0.004	0.025
popularity	0.051	1.000	0.051	-0.004	-0.014	0.015	0.031	-0.021	0.021	0.025	-0.048	-0.010
duration_ms	1.000	0.051	1.000	-0.060	-0.079	-0.003	-0.080	-0.004	0.067	0.011	-0.004	0.025
danceability	-0.060	-0.004	-0.060	1.000	-0.104	0.033	-0.033	-0.068	0.146	-0.065	0.023	-0.126
energy	-0.079	-0.014	-0.079	-0.104	1.000	-0.003	0.651	-0.041	-0.057	-0.445	0.038	0.157
key	-0.003	0.015	-0.003	0.033	-0.003	1.000	-0.007	-0.153	0.007	0.002	-0.008	-0.033
loudness	-0.080	0.031	-0.080	-0.033	0.651	-0.007	1.000	-0.028	-0.076	-0.310	-0.105	0.102
mode	-0.004	-0.021	-0.004	-0.068	-0.041	-0.153	-0.028	1.000	-0.000	0.006	-0.039	0.025
speechiness	0.067	0.021	0.067	0.146	-0.057	0.007	-0.076	-0.000	1.000	0.000	-0.063	0.061
acousticness	0.011	0.025	0.011	-0.065	-0.445	0.002	-0.310	0.006	0.000	1.000	-0.005	-0.110
instrumentalness	-0.004	-0.048	-0.004	0.023	0.038	-0.008	-0.105	-0.039	-0.063	-0.005	1.000	-0.035
liveness	0.025	-0.010	0.025	-0.126	0.157	-0.033	0.102	0.025	0.061	-0.110	-0.035	1.000
valence	-0.117	-0.016	-0.117	0.403	0.334	0.037	0.232	-0.075	0.074	-0.128	-0.015	0.019
tempo	-0.029	0.014	-0.029	-0.173	0.154	-0.001	0.081	0.048	0.058	-0.104	0.035	0.029

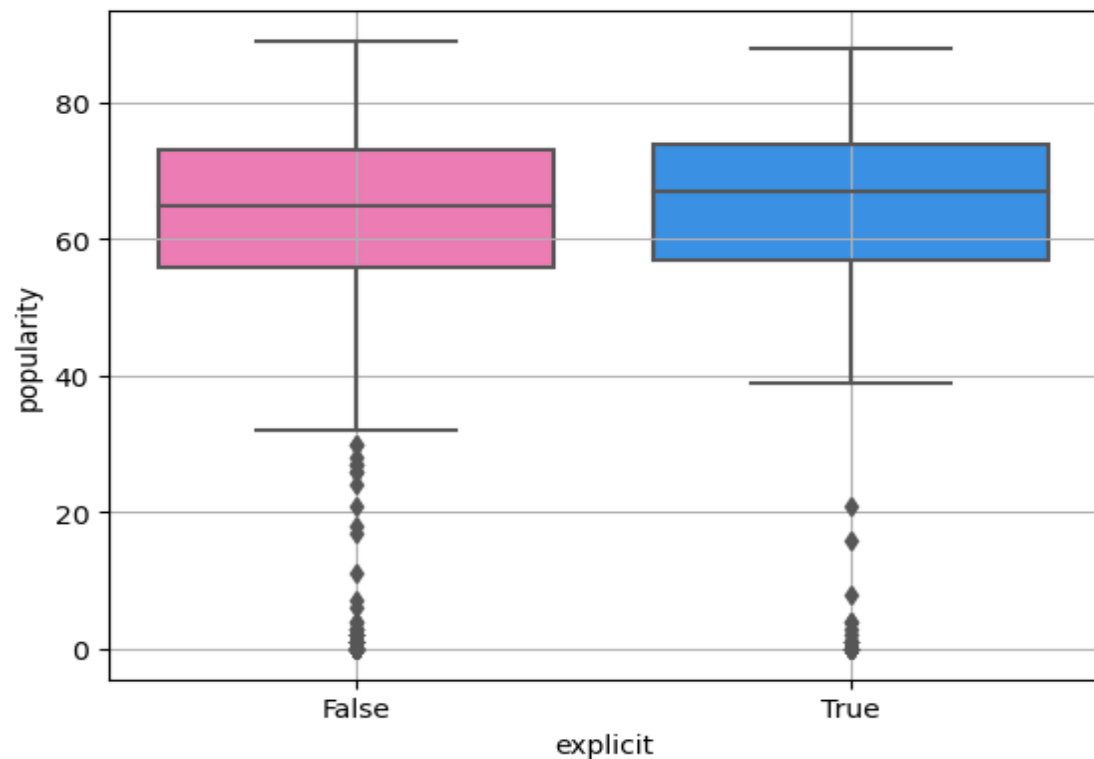
Все коэффициенты корреляции по диагонали таблицы равны 1, потому что каждая переменная совершенна и коррелирует сам с собой. Все остальные коэффициенты корреляции указывают на корреляцию между различными попарными комбинациями переменных.

Положительный коэффициент указывает на прямую связь, когда значения двух переменных увеличиваются вместе. Отрицательный коэффициент указывает на обратную связь, когда одно значение увеличивается, а другое уменьшается.

Линейной связи в указанном датасете между переменными не обнаружено.

Графики

Зависимость популярности от explicit (слишком откровенный)



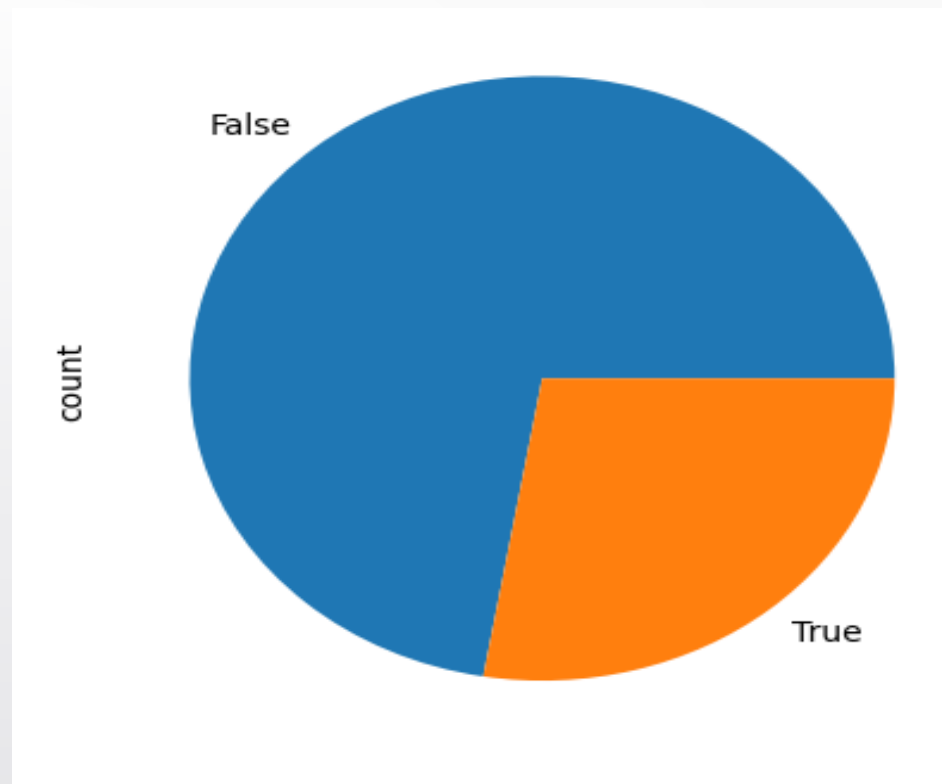
```
In [264]: df.groupby('explicit').agg({'popularity': ['mean', 'var', 'std', 'median']}).sort_values('explicit')
```

Out[264]:

explicit	popularity			
	mean	var	std	median
False	59.259489	465.217149	21.568893	65.0
True	61.484574	426.086580	20.641865	67.0

Среднее значение откровенности ниже у популярных песен 59), у не популярных 61.

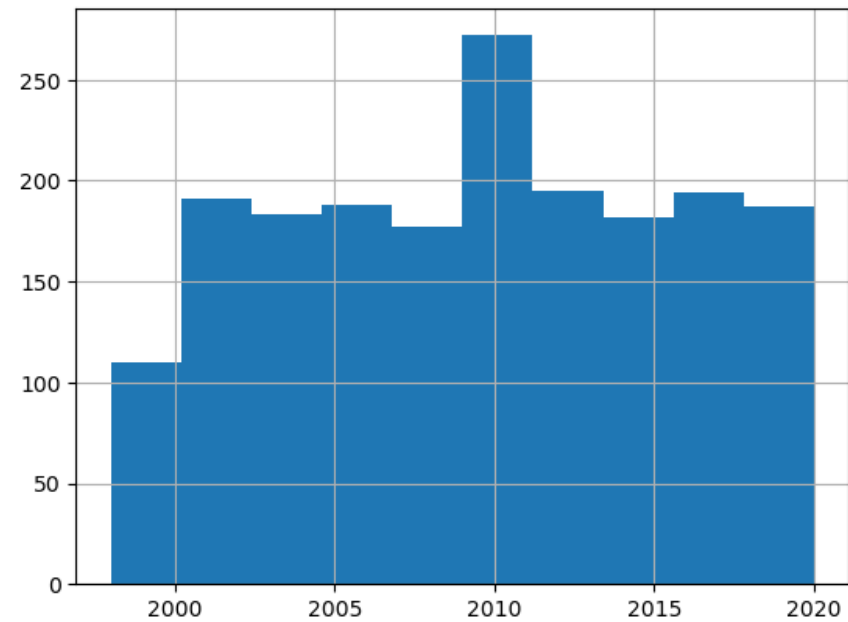
Бар чат по explicit



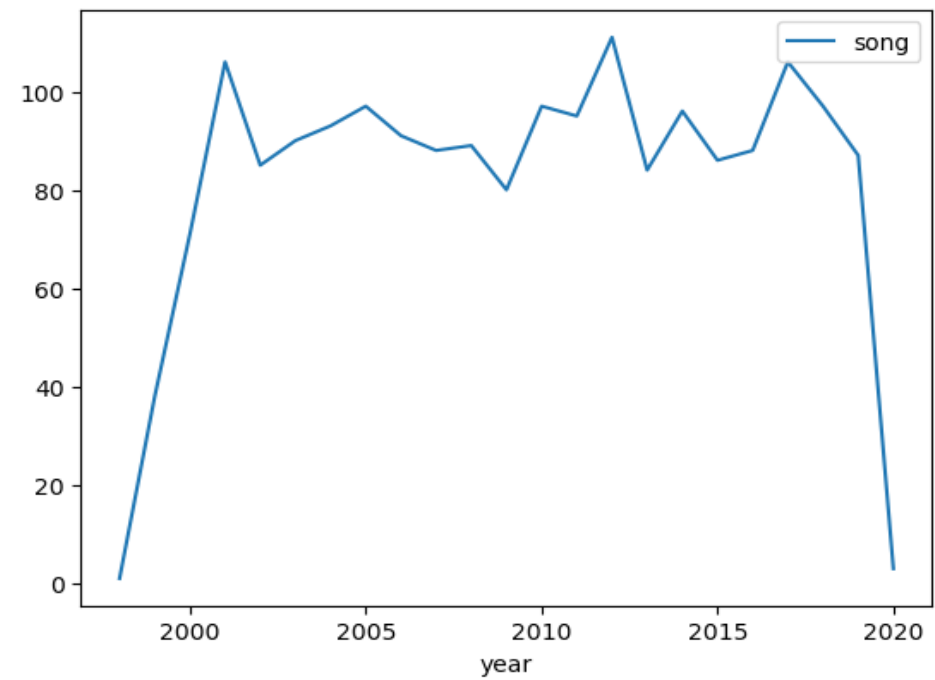
Явный (слишком откровенный): Текст или содержание песни или музыкального видео содержат один или несколько критериев, которые могут быть сочтены оскорбительными или неподходящими для детей. Количество не откровенных песен значительно больше.

Графики

Histogram частота встречаемости песен в разрезе годов

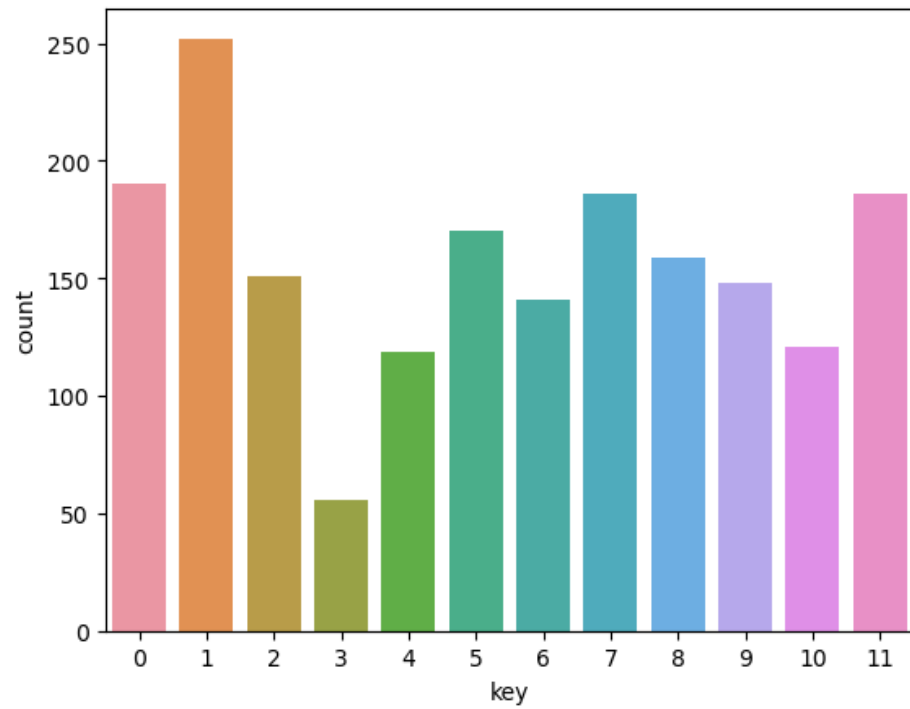


Line Chart «Количество песен по годам»

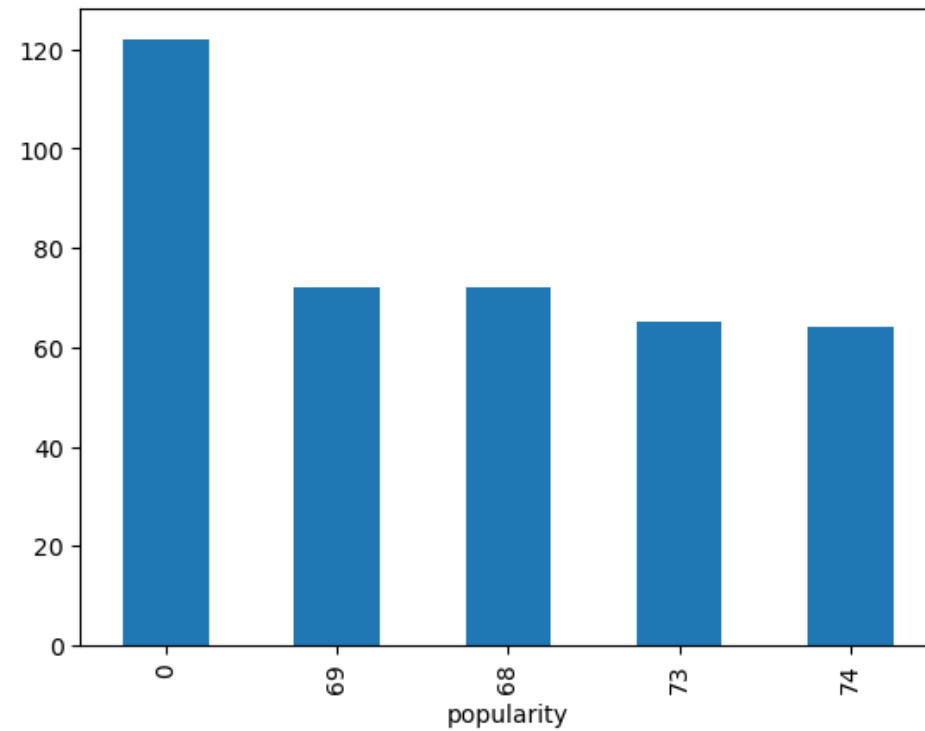


Графики

График подсчета тональности, в которой находятся все треки

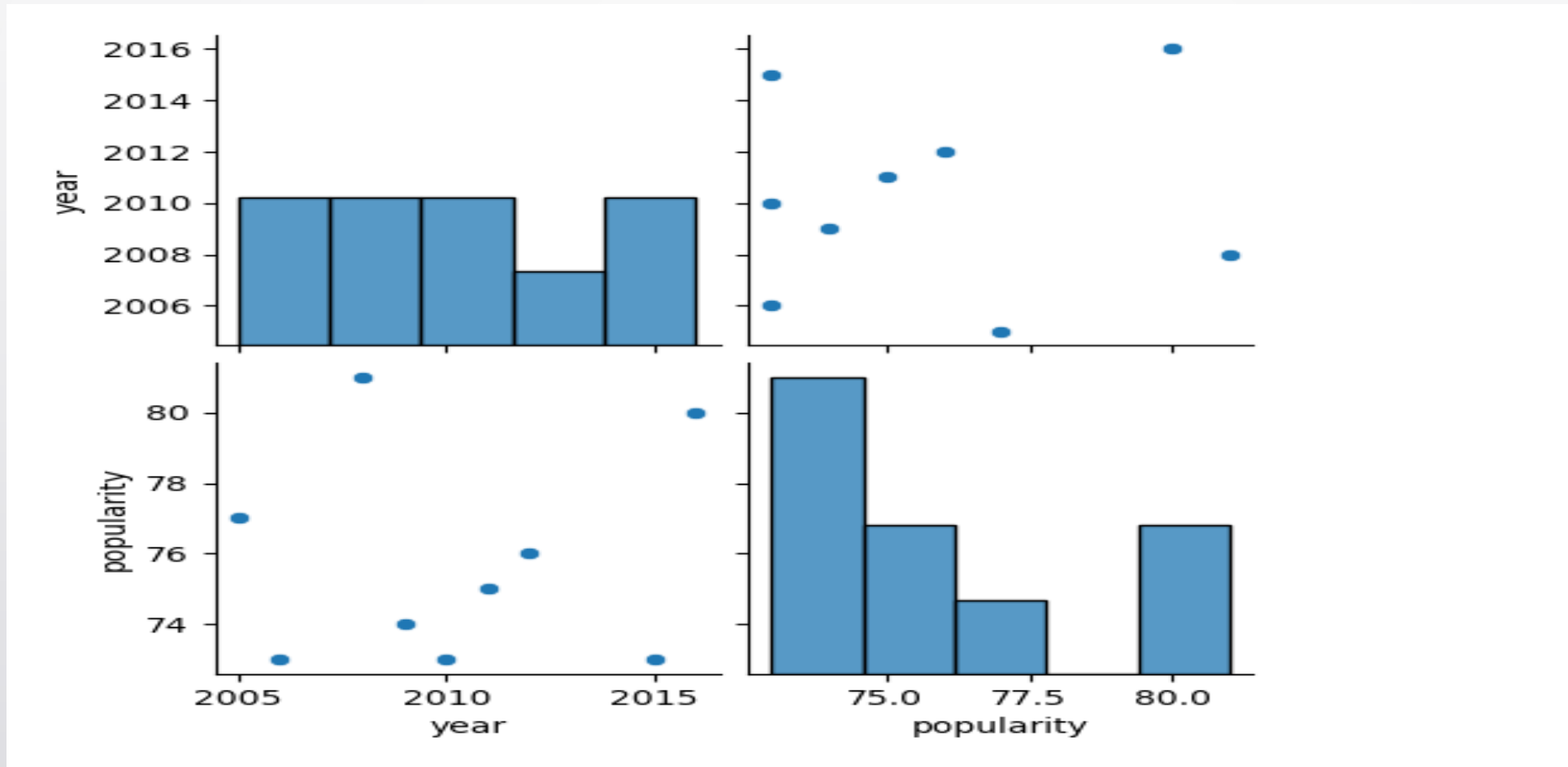


Значения популярности всех треков



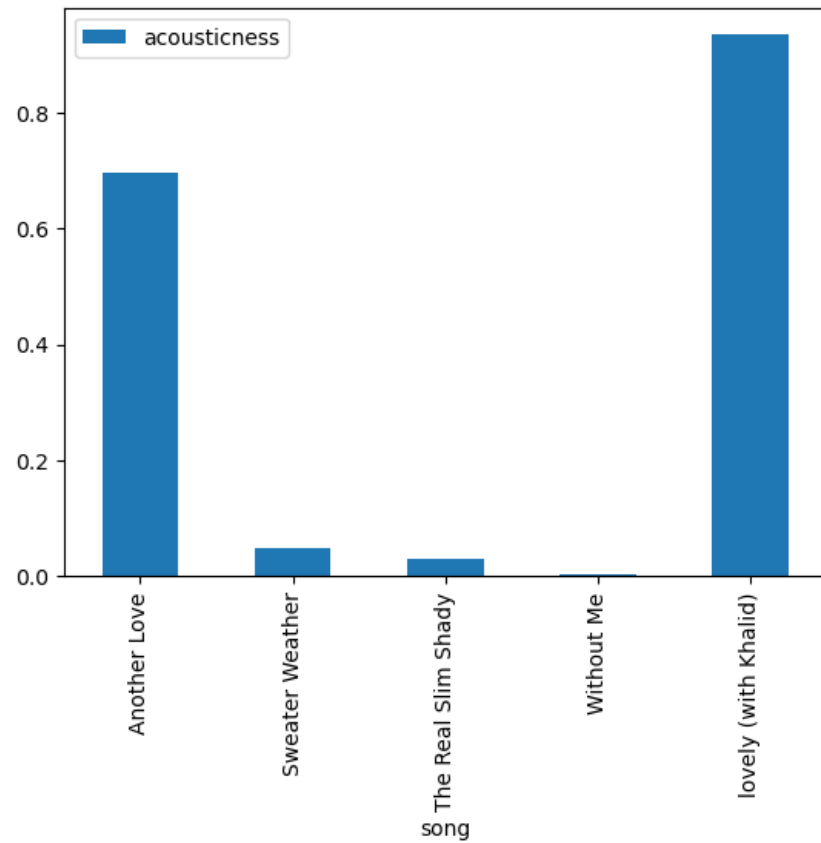
Графики

Многомерный анализ популярности песен Рианны по годам

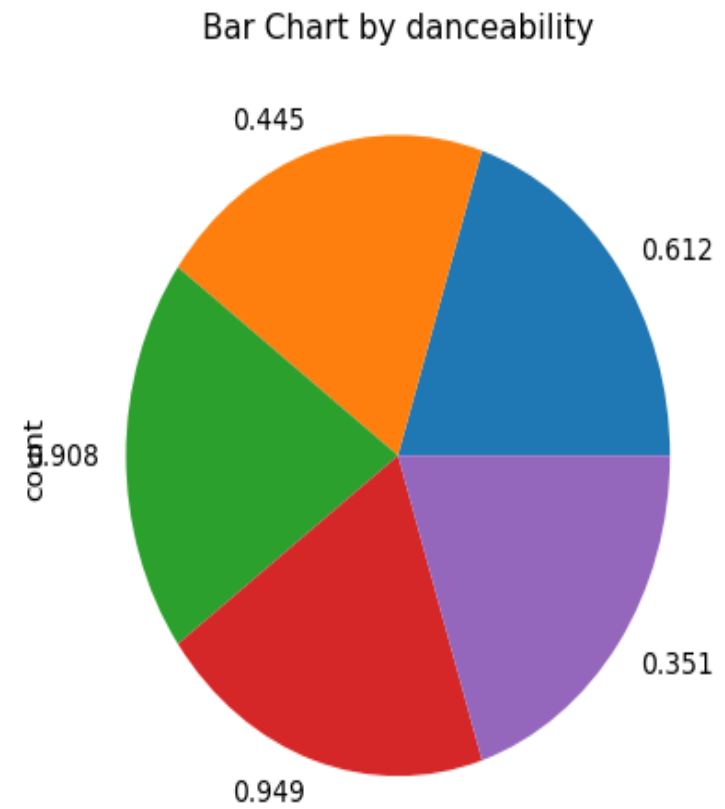


Графики

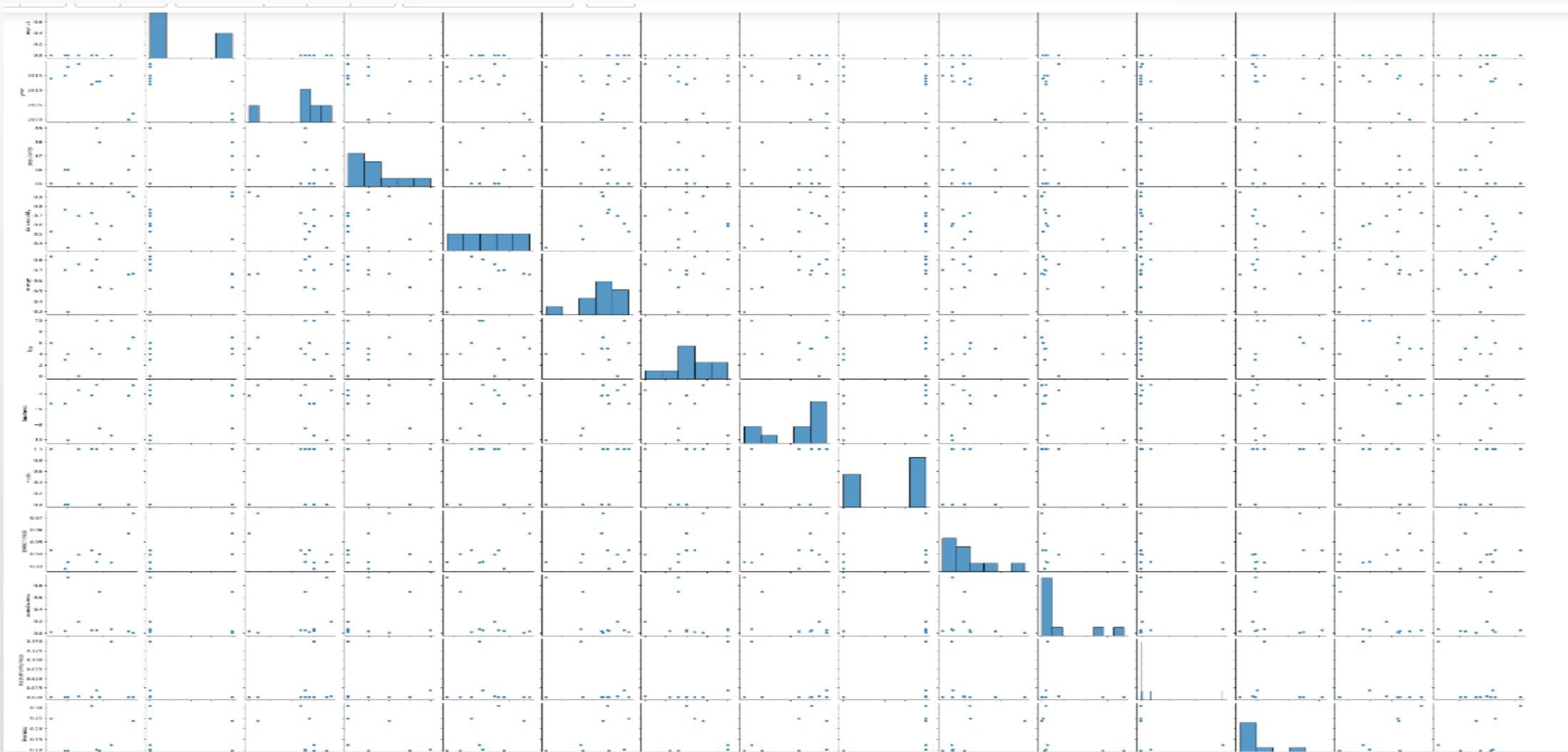
Уровень акустичности топ-5 песен



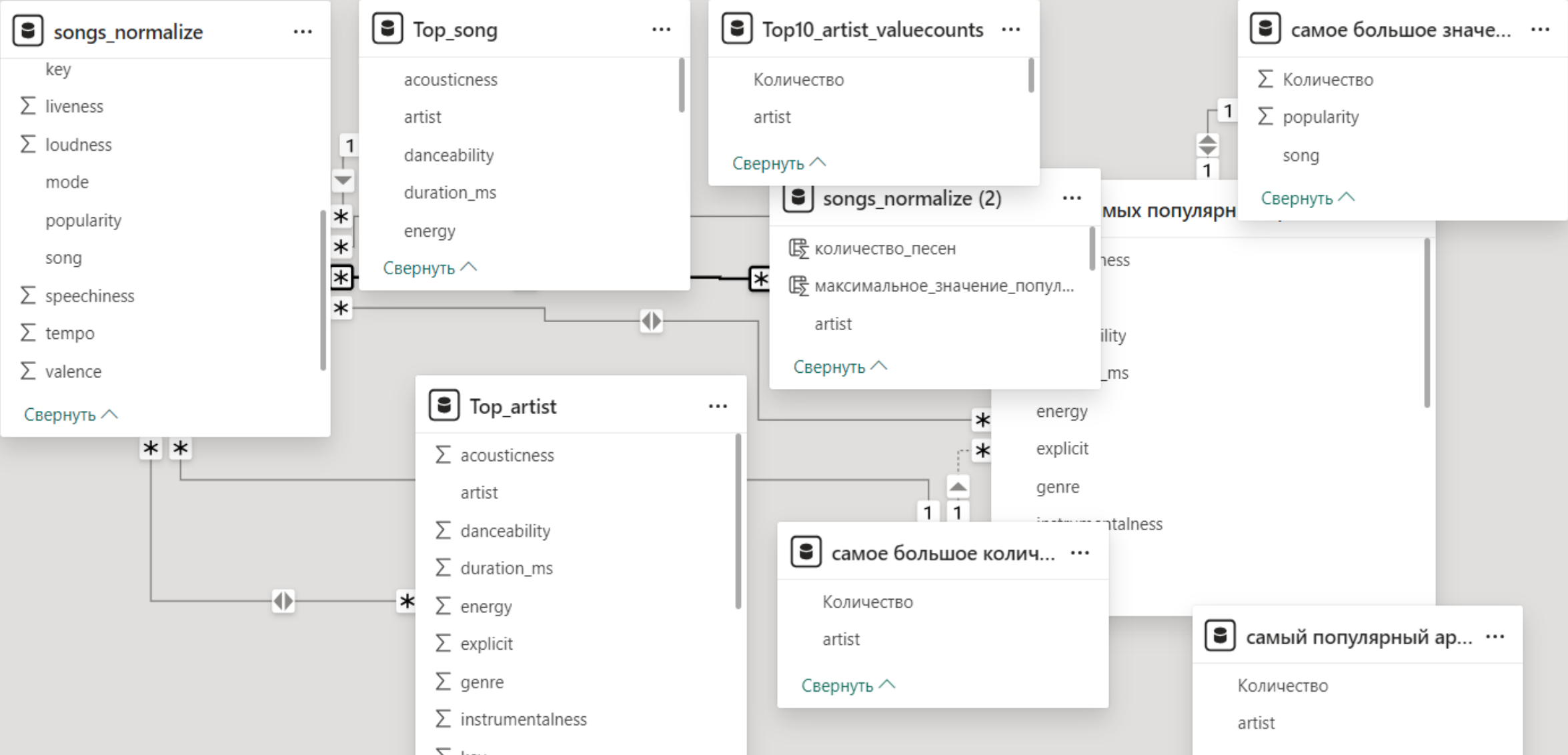
Значение танцевальности топ-5 песен



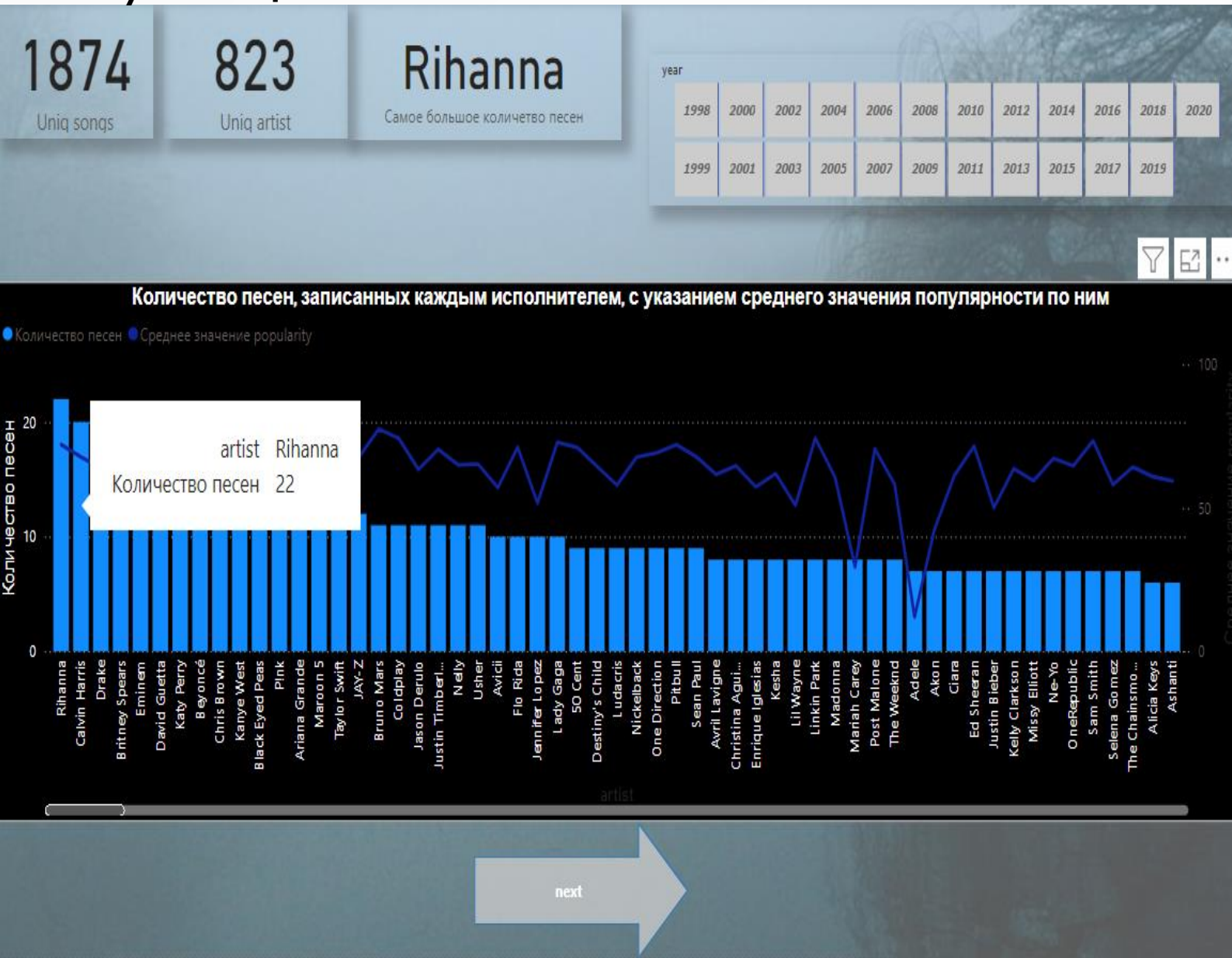
Многомерный анализ по топ-10 самым популярным песням



ПОСТРОЕНИЕ МОДЕЛИ



Визуализация



Созданы таблицы:

Top10_artist_valuecou...

- Количество
- Мера
- artist
- Total_artist
- Total_songs

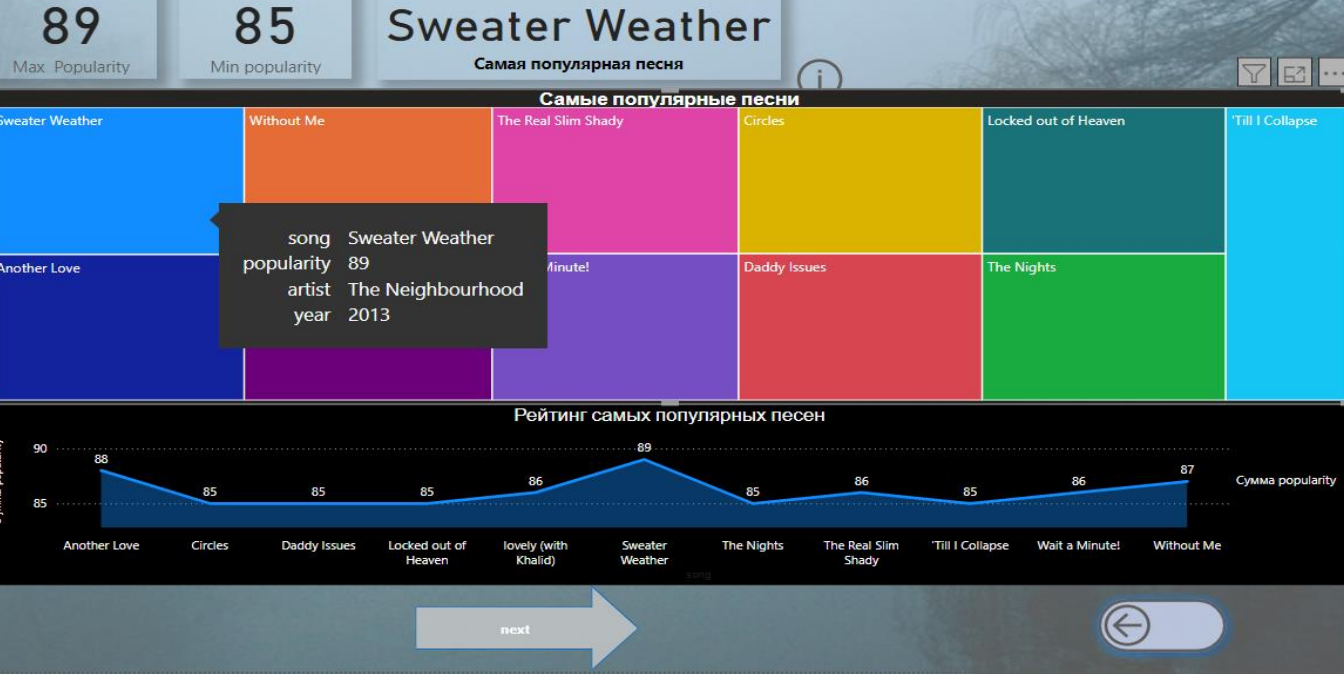
= Table.FirstN("#Строки с примененным фильтром",2)

ABc artist	123 Количество
1 Rihanna	22

Созданы меры:

Total_songs = DISTINCTCOUNT (songs_normalize[song])

Total_artist = DISTINCTCOUNT (songs_normalize[artist])



Созданы меры:

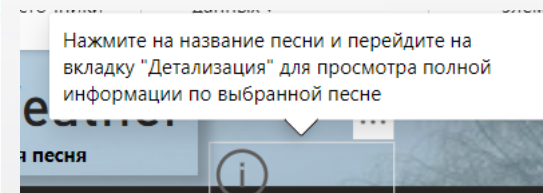
```
Top_song = TOPN(10, songs_normalize, songs_normalize[popularity])
```

Созданы таблицы:

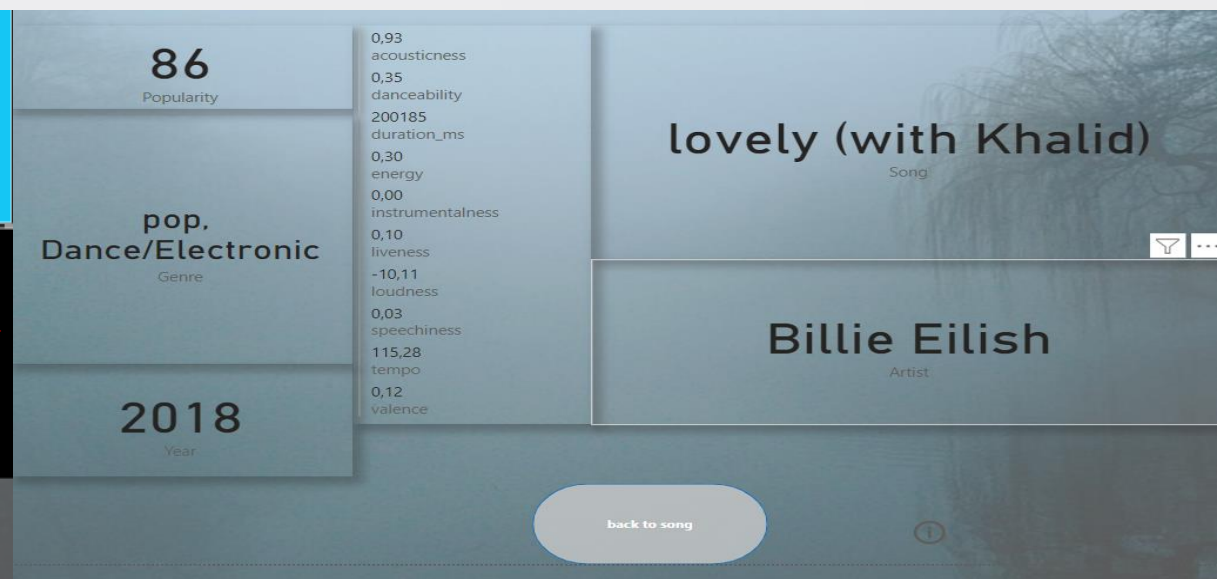
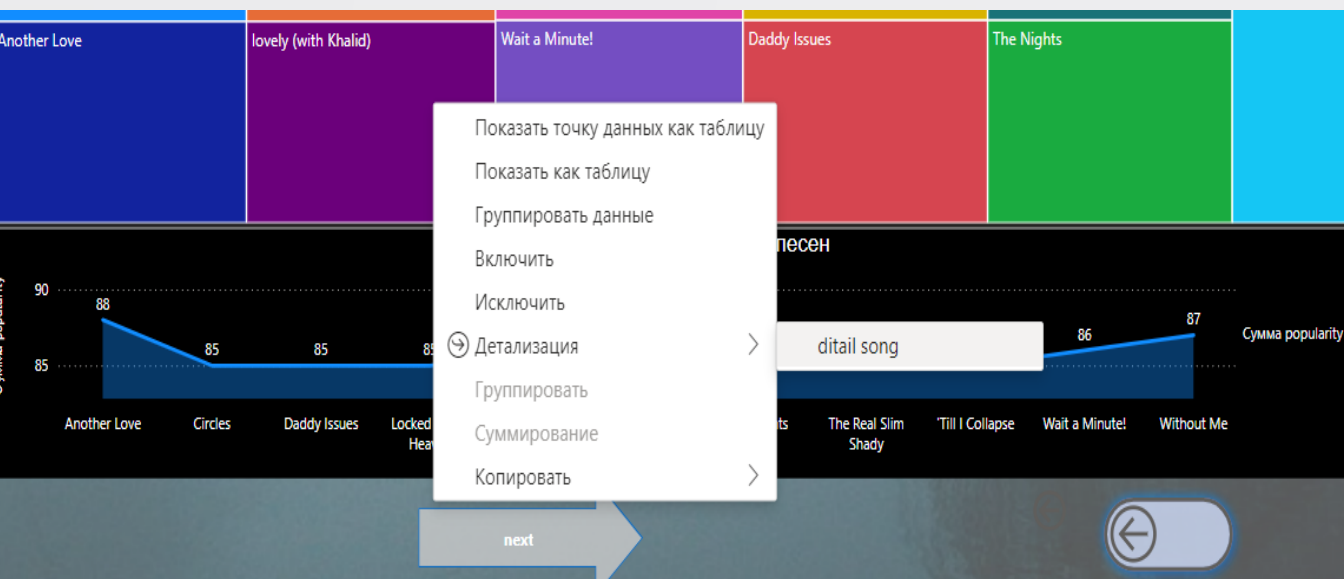
☐ ☒ ☐ ☐ = Table.FirstN("#Сортированные строки",1)

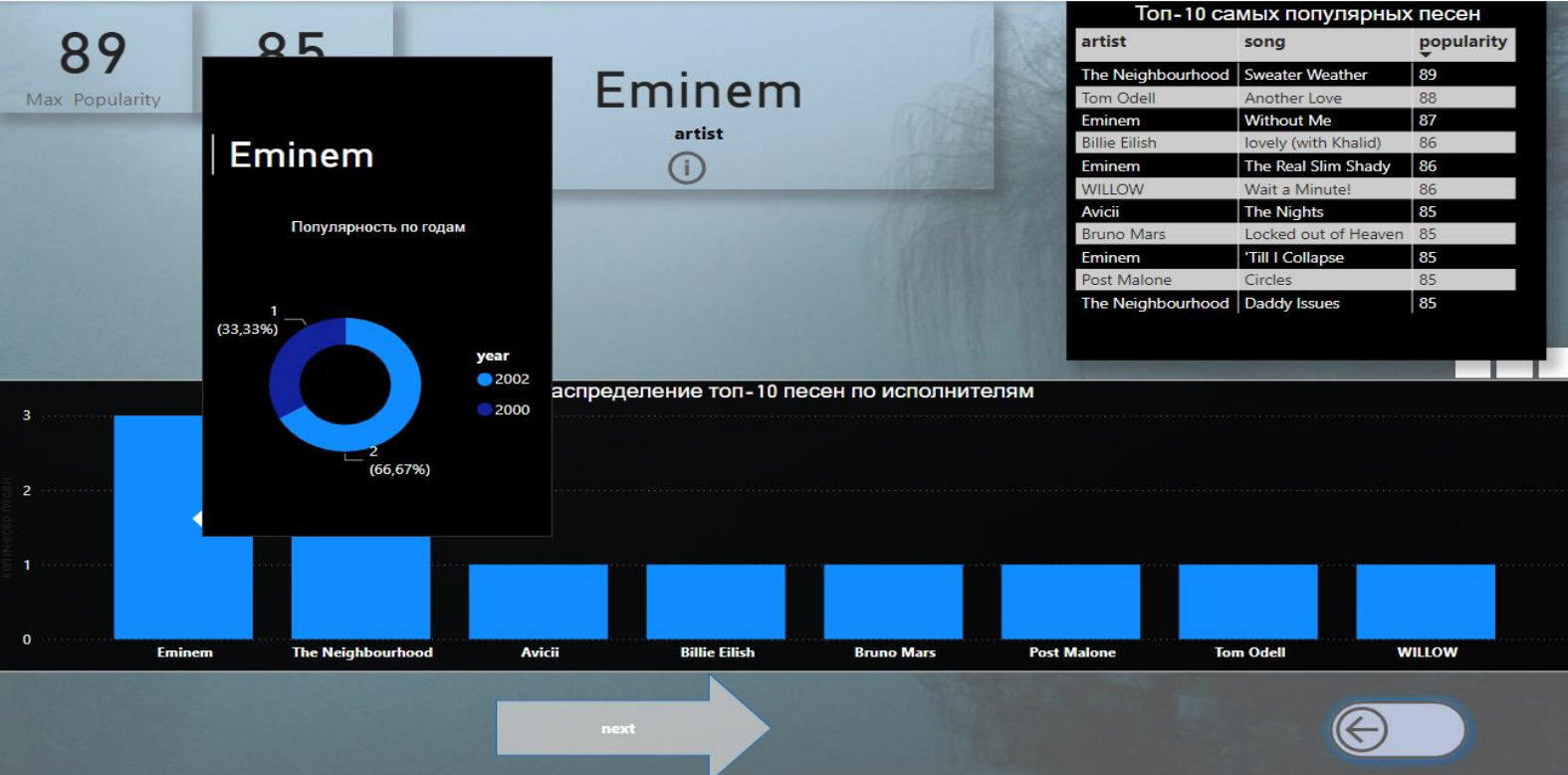
song	popularity	Количество
1 Sweater Weather	89	1

Добавлена кнопка информации:



Создана детализация





Топ-10 самых популярных песен		
artist	song	popularity
The Neighbourhood	Sweater Weather	89
Tom Odell	Another Love	88
Eminem	Without Me	87
Billie Eilish	lovely (with Khalid)	86
Eminem	The Real Slim Shady	86
WILLOW	Wait a Minute!	86
Avicii	The Nights	85
Bruno Mars	Locked out of Heaven	85
Eminem	'Till I Collapse	85
Post Malone	Circles	85
The Neighbourhood	Daddy Issues	85



Созданы таблицы:

fx

= Table.FirstN("#Сгруппированные строки",1)

AB artist

123 Количество

1

Eminem

3

fx

= Table.Sort("#Строки с примененным фильтром",{{"popularity", Order.Descending}})

AB artist	AB song	AB duration_ms	AB explicit	AB year	
1	The Neighbourhood	Sweater Weather	240400	False	2013
2	Tom Odell	Another Love	244360	True	2013
3	Eminem	Without Me	290320	True	2002
4	Eminem	The Real Slim Shady	284200	True	2000
5	Billie Eilish	lovely (with Khalid)	200185	False	2018
6	WILLOW	Wait a Minute!	196520	False	2015
7	The Neighbourhood	Daddy Issues	260173	False	2015
8	Post Malone	Circles	215280	False	2019
9	Avicii	The Nights	176658	False	2014
10	Bruno Mars	Locked out of Heaven	233478	False	2012
11	Eminem	'Till I Collapse	297786	True	2002

Добавлена кнопка информации



Добавлена кнопка информации
Создана мера:

Мера 3 = `CALCULATE([MAX([popularity])])`

Создана таблица:

`= Table.RemoveColumns("#Измененный тип",{ "genre", "tempo", "valence", "liveness",`

	AB_C artist	AB_C song	12_3 popularity
1	Britney Spears	Oops!...I Did It Again	77
2	blink-182	All The Small Things	79
3	Faith Hill	Breathe	66
4	Bon Jovi	It's My Life	78
5	*NSYNC	Bye Bye Bye	65
6	Sisqo	Thong Song	69
7	Eminem	The Real Slim Shady	86
8	Robbie Williams	Rock DJ	68
9	Destiny's Child	Say My Name	75
10	Modjo	Lady - Hear Me Tonight	77
11	Gigi D'Agostino	L'Amour Toujours	1
12	Eiffel 65	Move Your Body - Gabry Ponte Original Radio Edit	56
13	Bomfunk MC's	Freestyler	55
14	Sting	Desert Rose	62
15	Melanie C	Never Be The Same Again	61
16	Aaliyah	Try Again	53
17	Anastacia	I'm Outta Love - Radio Edit	64
18	Alice DeeJay	Better Off Alone	73
19	Gigi D'Agostino	The Riddle	64
20	Dr. Dre	The Next Episode	82
21	Linkin Park	In the End	83
22	Tom Jones	Sexbomb	65

На каждой странице добавлены:

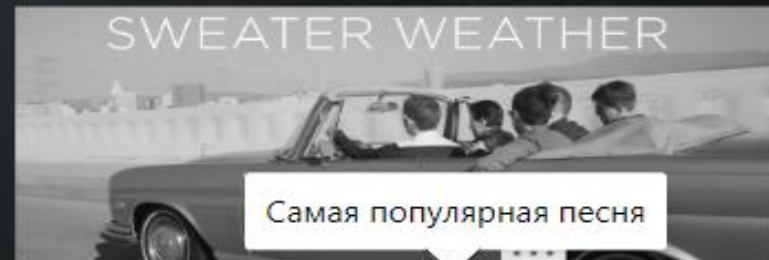
- кнопка «next» с настройкой перехода на следующую страницу через панель закладок
- кнопка «назад» в виде стрелки для перехода на страницу назад с автоматическим переходом назад



Популярный исполнитель с самым большим количеством песен



Исполнитель, у которого самое большое количество песен среди топ-10 самых популярных исполнителей



Самая популярная песня



РЕЗУЛЬТАТЫ

Проведенный анализ и исследование данных позволило нам сделать следующие выводы:
В исследовании учувствовали 1879 песен 823 исполнителей музыкального сервиса Spotify за период 2000 по 2019 гг.

Популярность – основной показатель, который необходимо исследовать в данном конкретном случае при анализе треков. При построении корреляционной матрицы видно, что между показателями нет линейной связи, а если есть, то она незначительная, соответственно, остальные показатели, указанные в датасете, не влияют на значение популярности.

Больше всего количество песен за исследуемый период у Рианны – 22 песни, однако ее песни не вошли в топ-10 самых популярных, значение популярности ее самой популярной песни «Umbrella», выпущенной в 2008 году составило 81, в то время как значение самой популярной песни по датасету «Sweater weather», выпущенной группой The Neighbourhood в 2013 году, составило 89.

В тоже время Eminem вошел вместе с Рианной в топ-5 исполнителей по количеству песен, их у него 19, а максимальное значение популярности его песен составило 87, с тремя своими песнями он вошел в топ-10 исполнителей с самыми популярными песнями, значения популярности их составило соответственно 87, 86 и 85.

Группу The Neighbourhood в данном датасете представляют всего 2 песни, что гораздо меньше чем у Рианны (22) и Эминема (19). Однако, две песни данной группы «Sweater weather» и «Daddy Issues» имеют значения популярности 89 и 85 соответственно и входят в топ-10 самых популярных песен, а песня «Sweater weather» - номер один по популярности по всему датасету.

На сновании всего изложенного, можно сделать следующей вывод, что **Первое место** среди проведенного анализа занимает исполнитель **Eminem с песней "Without Me"- формула идеального трека****

** согласно проведенного анализа Eminem самый часто встречающийся исполнитель с самыми популярными песнями, по данным топ-5 самых часто встречающихся исполнителей и данным топ-10 самых популярных песен.

WITHOUT ME

EMINEM

Qlibe Music



4:52

Eminem - Without Me (Lyrics)

Спасибо за внимание