

# Анализ музыкальных трендов Spotify за последние 100 лет



100 лет  
ЮФУ



Лебедев  
Виктор Борисович

Мишин  
Кирилл Игоревич

27.06.2025

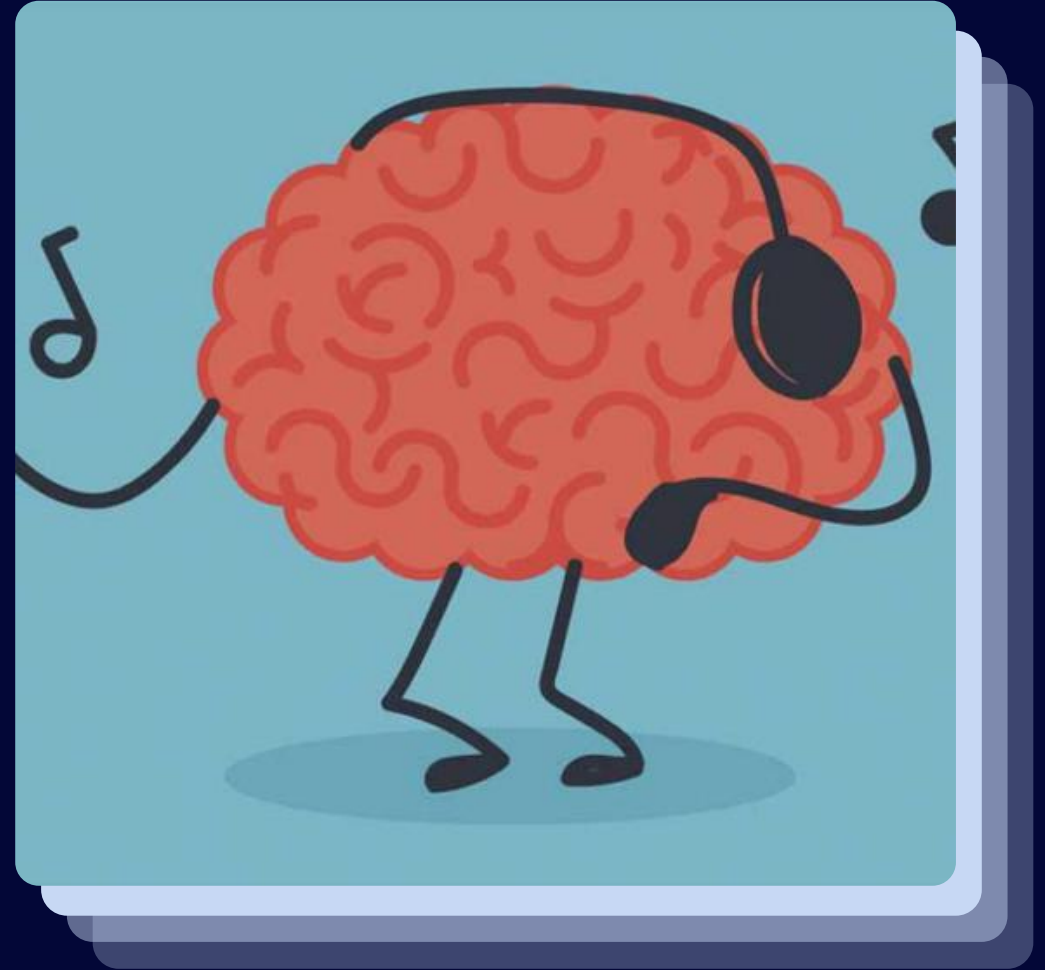
# Цель проекта и актуальность

Целью проекта является исследование изменений музыкальных характеристик треков за последние 100 лет с использованием инструментов Big Data.

Музыка отражает общественные и культурные изменения.

Понимание музыкальных трендов позволяет:

- анализировать вкусы слушателей
- прогнозировать развитие музыкальной индустрии
- использовать данные в рекомендационных системах



- Музыка со временем становится более ритмичной и энергичной.  
Современные треки предполагаются более бодрыми, чем в прошлом.
- Музыкальные предпочтения общества изменяются каждые десятилетия.  
Эти изменения могут отражать культурные и технологические сдвиги.
- Глобальные события влияют на стиль и настроение музыки.  
В периоды войн или кризисов музыка может становиться более мрачной или меланхоличной.
- Технологический прогресс меняет способы создания музыки.  
С распространением цифровых инструментов музыка всё чаще создаётся без живых инструментов.
- Развитие платформ (радио, интернет, стриминг) влияет на формат и структуру треков.  
Музыка адаптируется под новые форматы прослушивания и восприятия.

В работе использован открытый датасет **Spotify Historical Data 1921–2020** с платформы *Kaggle*. Он содержит информацию о треках, выпущенных за последние 100 лет, включая аудиофичи, автоматически рассчитанные Spotify.

## 🔍 Основные параметры датасета:

- **danceability** – насколько удобно под трек танцевать (0–1).
- **energy** – «энергичность» трека, уровень громкости и интенсивности.
- **loudness** – средняя громкость трека в децибелах.
- **speechiness** – наличие речевых компонентов (чем выше, тем ближе к подкасту).
- **acousticness** – вероятность того, что трек акустический.
- **instrumentalness** – вероятность, что трек без вокала.
- **liveness** – показатель «живого» исполнения (например, концерт).
- **valence** – эмоциональная окраска (0 – грустно, 1 – весело).
- **tempo** – темп музыки в ударах в минуту (BPM).
- **release\_date** – дата выпуска трека (год, месяц или точная дата).

Всего в датасете ~600 000 записей, отобраны треки с валидной датой и параметрами.

<https://www.kaggle.com/datasets/yamaerenay/spotify-dataset-19212020-600k-tracks>

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	id	name	popularity	duration_ms	explicit	artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalne	liveness	valence	tempo	time_signature	
2	35iwgR4jX	Carve	6	126903	0	['Uli']	['45tlt06Xc	22/02/1922	0.645	0.445	0	-13.338	1	0.451	0.674	0.744	0.151	0.127	104.851	3	
3	021ht4sdg	CapF-tulo 2.16 - Banquero Anarquista	0	98200	0	['Fernando Pessoa']	['14jtPCOc	01/06/1922	0.695	0.263	0	-22.136	1	0.957	0.797	0	0.148	0.655	102.009	1	
4	07A5yeht	Vivo para Quererte - Remasterizado	0	181640	0	['Ignacio Corsini']	['5LiOoJbx	21/03/1922	0.434	0.177	1	-21.18	1	0.0512	0.994	0.0218	0.212	0.457	130.418	5	
5	08FmqUh;	El Prisionero - Remasterizado	0	176907	0	['Ignacio Corsini']	['5LiOoJbx	21/03/1922	0.321	0.0946	7	-27.961	1	0.0504	0.995	0.918	0.104	0.397	169.98	3	
6	08y9Gfoq	Lady of the Evening	0	163080	0	['Dick Haymes']	['3BiJGZsy	1922	0.402	0.158	3	-16.9	0	0.039	0.989	0.13	0.311	0.196	103.22	4	
7	0BRXJHRN	Ave Maria	0	178933	0	['Dick Haymes']	['3BiJGZsy	1922	0.227	0.261	5	-12.343	1	0.0382	0.994	0.247	0.0977	0.0539	118.891	4	
8	0Dd9ImXt	La Butte Rouge	0	134467	0	['Francis Marty']	['2nuMRGz	1922	0.51	0.355	4	-12.833	1	0.124	0.965	0	0.155	0.727	85.754	5	
9	0IA0Hju8C	La Java	0	161427	0	['Mistinguett']	['4AxxgXfD7	1922	0.563	0.184	4	-13.757	1	0.0512	0.993	1.55e-05	0.325	0.654	133.088	3	
10	0lgl1UCz8	Old Fashioned Girl	0	310073	0	['Greg Fieger']	['5nWlsH5	1922	0.488	0.475	0	-16.222	0	0.0399	0.62	0.00645	0.107	0.544	139.952	4	
11	0JV4iqw2l	MartF-n Fierro - Remasterizado	0	181173	0	['Ignacio Corsini']	['5LiOoJbx	29/03/1922	0.548	0.0391	6	-23.228	1	0.153	0.996	0.933	0.148	0.612	75.595	3	
12	0OYGe21c	CapF-tulo 2.8 - Banquero Anarquista	0	99100	0	['Fernando Pessoa']	['14jtPCOc	01/06/1922	0.676	0.235	11	-22.447	0	0.96	0.794	0	0.21	0.724	96.777	3	
13	0PE42H6t	CapF-tulo 2.25 - Banquero Anarquista	0	132700	0	['Fernando Pessoa']	['14jtPCOc	01/06/1922	0.75	0.229	2	-22.077	1	0.955	0.578	0	0.314	0.531	102.629	3	
14	0PH9AAC;	Lazy Boi	0	157333	0	['Uli']	['45tlt06Xc	22/02/1922	0.298	0.46	1	-18.645	1	0.453	0.521	0.856	0.436	0.402	87.921	4	
15	0QiT0Oo5	Tu Verras Montmartre	1	186800	0	['Lucien Boyer']	['4mSouLp	1922	0.703	0.28	0	-15.39	1	0.174	0.995	6.84e-05	0.163	0.897	127.531	4	
16	0TWsNj5i	Elle Prend L'boulevard Magenta	0	172027	0	['Fr@lix Mayol']	['7DIIOK9L	1922	0.709	0.289	2	-14.978	1	0.18	0.967	0	0.119	0.84	107.515	4	
17	0cC9CYjL	CapF-tulo 1.23 - Banquero Anarquista	0	96600	0	['Fernando Pessoa']	['14jtPCOc	01/06/1922	0.687	0.198	4	-24.264	0	0.962	0.754	0	0.197	0.478	78.453	1	
18	0eb1PfHx	CapF-tulo 1.18 - Banquero Anarquista	0	103200	0	['Fernando Pessoa']	['14jtPCOc	01/06/1922	0.8	0.171	8	-24.384	1	0.953	0.67	0	0.123	0.693	59.613	3	
19	0grXU6GK	CapF-tulo 1.10 - Banquero Anarquista	0	95800	0	['Fernando Pessoa']	['14jtPCOc	01/06/1922	0.7	0.208	2	-23.874	1	0.956	0.691	0	0.441	0.613	85.739	1	
20	0kCB1bD\	Ca C'est Une Chose	0	188000	0	['Victor Boucher']	['7vVR02IJ	1922	0.352	0.334	5	-13.038	1	0.0594	0.996	0.00746	0.36	0.414	76.403	4	
21	0l3BQsVJ	El Vendaval - Remasterizado	0	153533	0	['Ignacio Corsini']	['5LiOoJbx	21/03/1922	0.37	0.372	2	-17.138	1	0.0865	0.985	0.000681	0.929	0.753	159.669	4	
22	0pXUGQc	Ta Bouche	0	180933	0	['Jeanne Saint Bonnet']	['4ZACxcK	1922	0.335	0.304	7	-14.062	1	0.0506	0.996	0.0832	0.102	0.215	76.576	3	
23	0wA4YFw	CapF-tulo 1.22 - Banquero Anarquista	0	109100	0	['Fernando Pessoa']	['14jtPCOc	01/06/1922	0.704	0.244	0	-25.353	1	0.959	0.728	0	0.168	0.644	84.489	3	
24	0xJCJ9XS	La Maleva - Remasterizado	0	181440	0	['Ignacio Corsini']	['5LiOoJbx	29/03/1922	0.339	0.0958	8	-26.944	1	0.053	0.994	0.969	0.123	0.219	86.279	3	
25	0zyKYuXw	CapF-tulo 2.3 - Banquero Anarquista	0	113200	0	['Fernando Pessoa']	['14jtPCOc	01/06/1922	0.655	0.214	0	-23.337	1	0.962	0.665	0	0.279	0.761	76.087	1	
26	10XXG7gG	CapF-tulo 1.2 - Banquero Anarquista	0	109500	0	['Fernando Pessoa']	['14jtPCOc	01/06/1922	0.673	0.239	5	-22.552	1	0.96	0.539	0	0.633	0.644	83.532	4	
27	10u9ZWsc	J'ai Vendu Mon F,me	0	168000	0	['Adolphe BF@rard']	['4avV2NA	1922	0.582	0.152	10	-13.742	1	0.0914	0.972	0	0.31	0.546	95.779	3	
28	112adU33	Nuits De Chine	4	180800	0	['Louis Lynel']	['28pbliOo	1922	0.42	0.288	7	-14.005	1	0.0701	0.994	0.000216	0.196	0.4	139.575	3	
29	15BEv0ml	En la Huella del Querer - Remasterizado	0	150160	0	['Ignacio Corsini']	['5LiOoJbx	29/03/1922	0.373	0.176	4	-22.045	0	0.0703	0.994	0.866	0.157	0.383	106.595	1	

## 1. Подключение хранилища

Использовали Google Drive для хранения CSV-файла с треками.

## 2. Инициализация Spark

Создали сессию Apache Spark для эффективной работы с большими данными.

## 3. Загрузка данных

Считали CSV-файл с автоматическим определением типов колонок.

## 4. Обработка даты релиза

Извлекли год из разных форматов даты (уууу, уууу-ММ, dd/ММ/уууу, уууу-ММ-dd).


## 5. Обработка десятилетий

На основе года рассчитали десятилетия для анализа трендов.

## 6. Группировка и сохранение результатов

Отфильтровали некорректные данные, вычислили средние значения параметров по десятилетиям, сохранили результат в CSV.





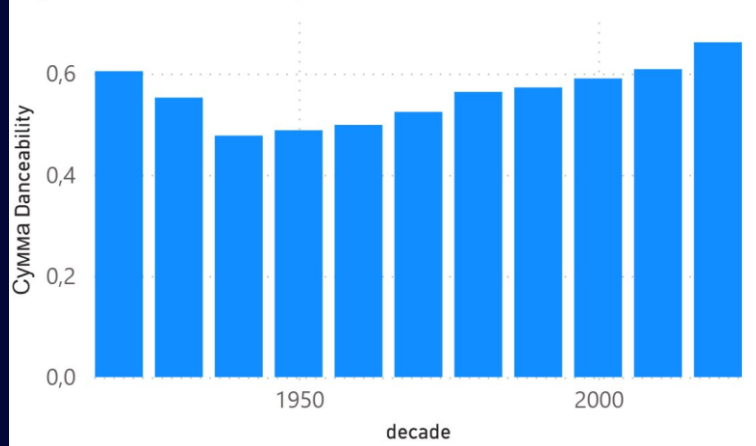
decade	avg(danceability)	avg(energy)	avg(loudness)	avg(speechiness)	avg(acousticness)	avg(instrumentalness)	avg(liveness)	avg(valence)	avg(tempo)
1920	0.6044980792001055	0.28125600052624633	-14.789149190895909	0.29171081436653223	0.8901129887383201	0.32611214068148864	0.2046233258781741	0.5999094605972903	113.12064886199191
1930	0.5522532350216158	0.30676277509265004	-13.522755867819649	0.19452228227300855	0.8693909407273038	0.27482508944101464	0.21714564546016088	0.572560382952439	112.83435013897511
1940	0.47718324893090214	0.2685315676344811	-14.196273970290271	0.10001664415935181	0.9146834272563601	0.3819400739984241	0.21389444069322566	0.5043820751744332	107.11015653837478
1950	0.48777139535568315	0.30281471289401696	-14.272708744150389	0.10806722488742378	0.8364244313797593	0.248301429149425	0.21083357859728677	0.5054985303028506	111.08317494187209
1960	0.49832130392115515	0.40633874567990913	-12.468499146648387	0.07663447326876341	0.6800565838548043	0.1595737188411493	0.2132387741605147	0.5639647559414579	114.20092215300626
1970	0.5240569810095742	0.5026020538548954	-11.427623145593232	0.103561257912677	0.4993326471702615	0.10073820331293694	0.22255466028242124	0.5808344534978104	117.4157947898074
1980	0.5635851921297385	0.5506233901646106	-11.305605634831725	0.13873560000973845	0.40595191095736044	0.0808659046232898	0.2257954519553894	0.5806300652607977	118.87677536161353
1990	0.5722725236500189	0.5746292606268649	-10.318843495785258	0.09942431810652587	0.3662772659834518	0.07329183580465275	0.21661842759597902	0.5697618027386174	119.69521585011297
2000	0.5900959242329817	0.6492454836530688	-7.555116409261975	0.08037426183715417	0.3129212164921375	0.0558838711614383	0.21077038980076993	0.5640294649906498	121.43506035736478
2010	0.6084883570714038	0.6584178113237791	-7.283913180583996	0.09072811318058385	0.2877811493377789	0.08885333605573306	0.20557987233839933	0.5142258487400535	122.08340345829554
2020	0.6615824750954046	0.6328855980373688	-7.647281310402949	0.11788785746146596	0.27387514179213956	0.11514137585171254	0.18660253754274736	0.503013339941517	121.86342414630492

Полученная таблица с данными, отфильтрованными по десятилетиям

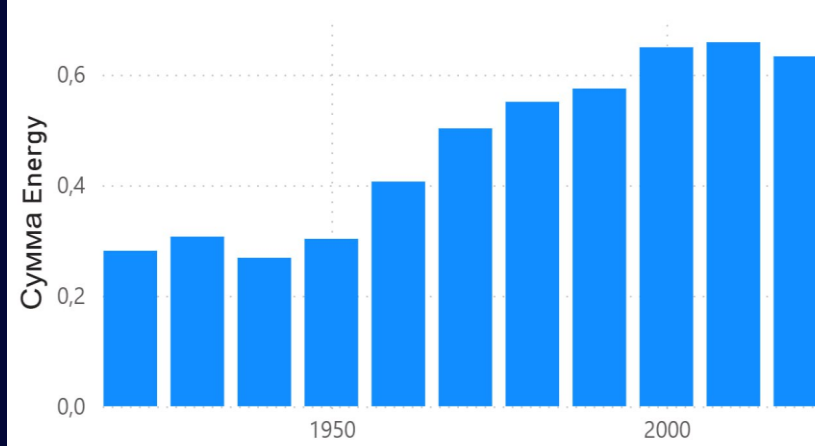


# Визуализация в Power BI

Сумма Danceability по decade



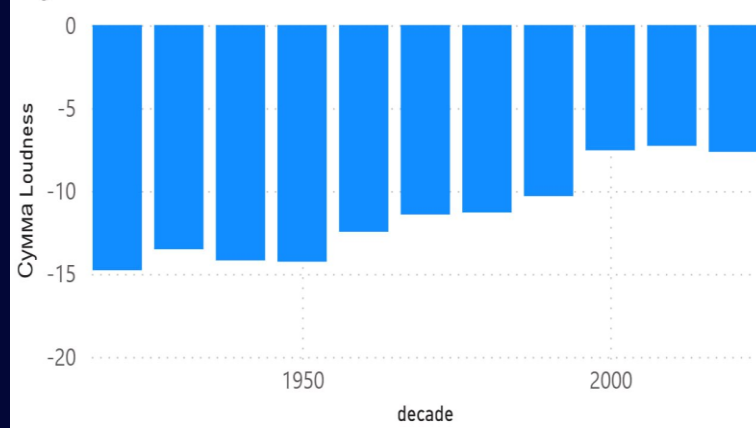
Сумма Energy по decade



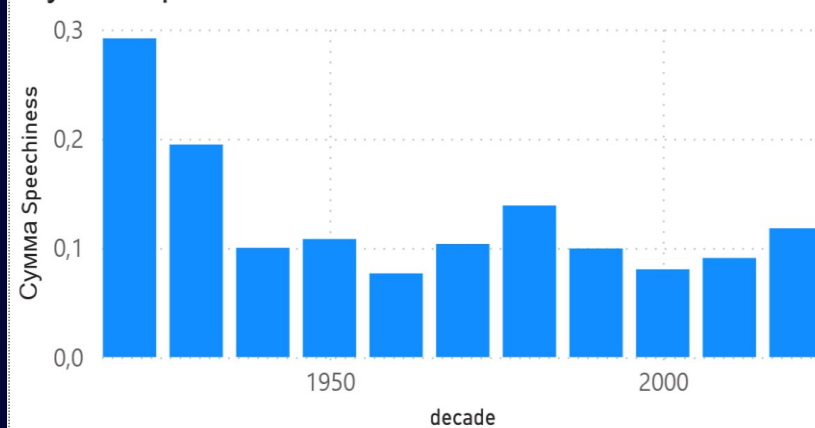
Danceability постепенно растёт с 1940-х до 2020-х, отражая увеличение доли танцевальной поп-музыки и упрощённых ритмов. Особенно заметен скачок после 2000-х.

Energy также демонстрирует рост: от спокойных композиций прошлого к более «заряженным» и динамичным трекам. Пик достигается в 2000–2010-х.

Сумма Loudness по decade



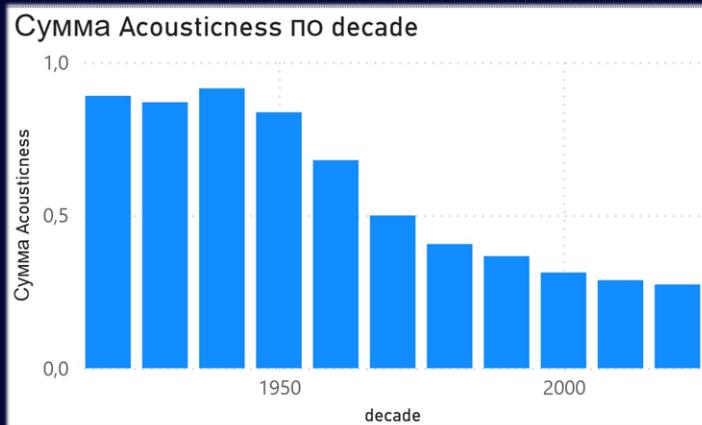
Сумма Speechiness по decade



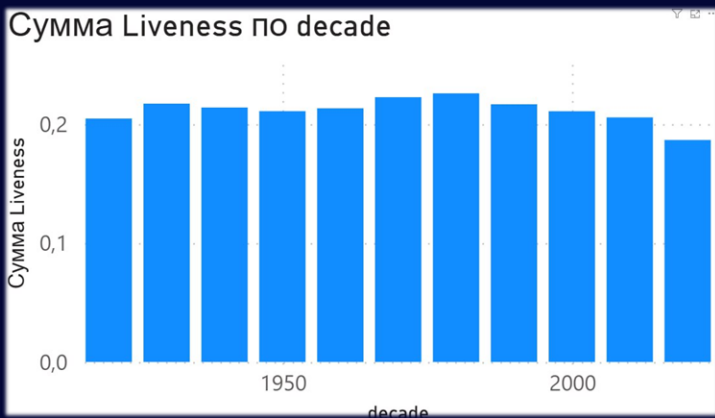
Loudness (громкость, в децибелах) повышается с десятилетиями, что связано с ростом уровня громкости в поп-индустрии («loudness war»). После 1990-х средняя громкость стабилизируется на высоком уровне.

Speechiness указывает на наличие речи: после высоких значений в 1920-х (возможно, из-за джазовых записей с речевыми вставками) падает, а в 2000-х и 2020-х снова растёт — из-за рэпа и трэп-музыки.

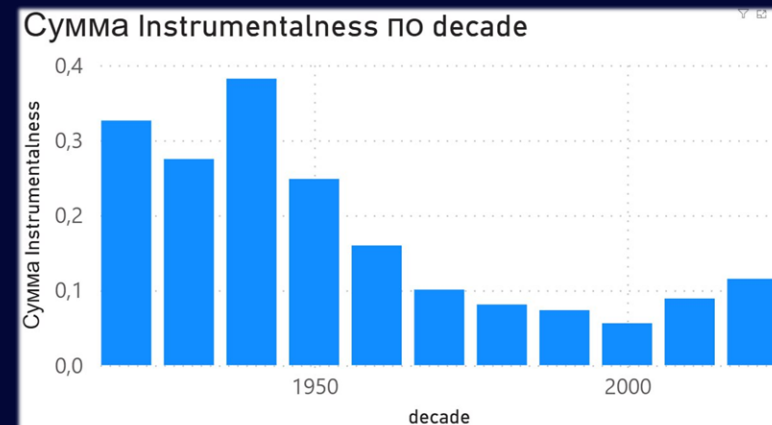
Liveness отражает «живость» записи — высокие значения характерны для концертных треков. Показатель остаётся примерно стабильным, с небольшим снижением к 2020-м, что говорит о росте студийных записей и постпродакшна.

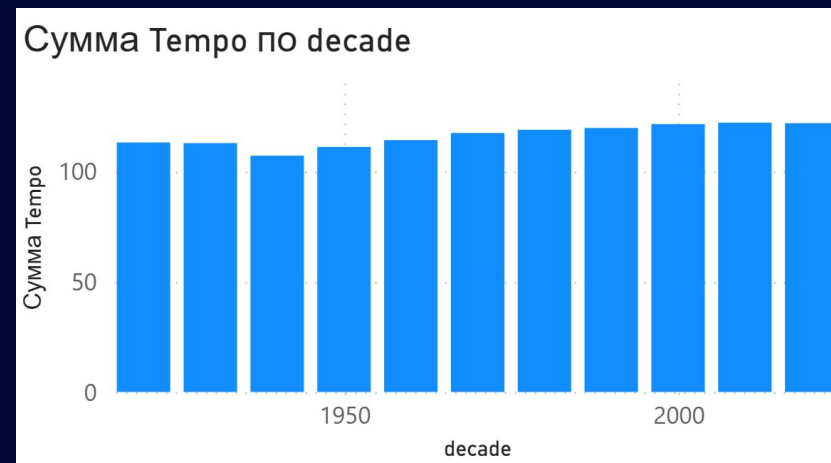
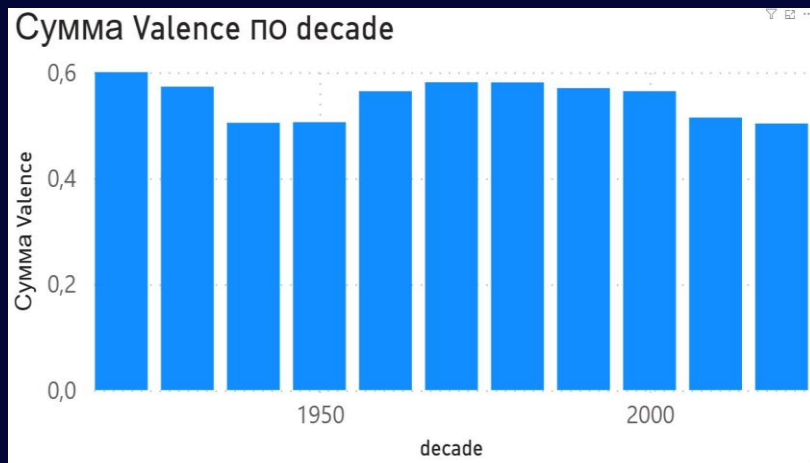


**Instrumentalness** тоже снижается — инструментальные композиции уступают место песням с вокалом. Особенно низкие значения после 2000-х говорят о том, что вокал стал почти обязательным элементом треков.



**Acousticness** заметно падает со временем: от доминирования акустических инструментов в 1920–50-х к более электронному звучанию в XXI веке. Это отражает общее снижение доли живой инструментальной музыки.





**Valence** — мера «позитивности» или эмоциональной окраски трека. С 1920-х по 1980-е значения были высокими, отражая более радостное звучание музыки. Начиная с 1990-х, наблюдается постепенное снижение, что может указывать на популяризацию более меланхоличных или серьёзных жанров в современной музыке.

**Tempo** — скорость трека в ударах в минуту. С 1940-х по 2010-е наблюдается плавный рост темпа: музыка становится чуть быстрее. Это связано с развитием танцевальной и электронной музыки, а также изменениями в предпочтениях слушателей.

Анализ параметров треков по десятилетиям показывает, как музыка отражает эпоху — как с точки зрения эмоций, так и технологий:

В первой половине XX века музыка была более акустической, живой и оптимистичной. Высокие значения *valence*, *acousticness* и низкие значения *loudness* указывают на доминирование инструментов и живого исполнения.

В 60–70-х начинается переход к экспериментам — появляются синтезаторы, усиливается *energy*, снижается *acousticness*.

80–90-е демонстрируют рост *danceability*, *tempo* и *instrumentalness* — влияние клубной культуры, диско и электронной музыки.

С 2000-х наблюдается уверенный рост громкости (*loudness*) и плотности аранжировок. Музыка становится громче, проще по структуре, но насыщеннее по обработке.

В 2010–2020-х популярность набирает интроспективный поп, хип-хоп и lo-fi, что отражается в снижении *valence*, но росте *energy* и *danceability*. Музыка становится менее радостной, но более ритмичной и танцевальной.

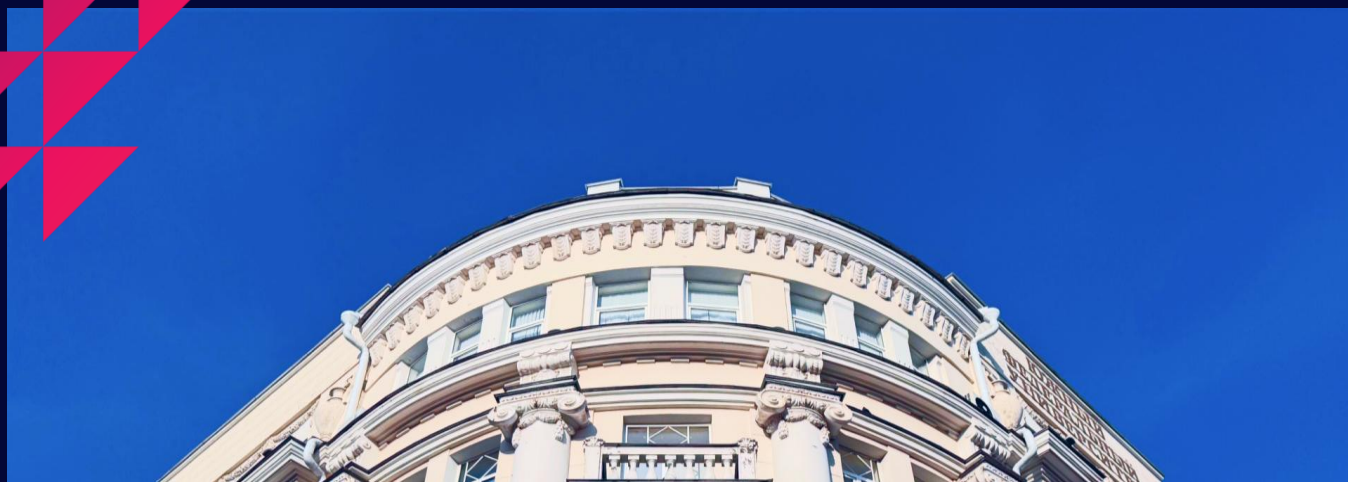
Общая тенденция — цифровизация, уход от живых инструментов и рост производственного качества при потере эмоциональной простоты.

🎵 Музыка XXI века — это смесь технологий, грусти и танца.

Спасибо за внимание!



110 лет  
ЮФУ



Лебедев  
Виктор Борисович  
Мишин  
Кирилл Игоревич



[www.sfedu.ru](http://www.sfedu.ru)