



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

КАФЕДРА СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ И УПРАВЛЕНИЯ

Лабораторная работа №1

по дисциплине: «Методы машинного обучения»

Студент	Ваганов Даниил Дмитриевич
Группа	ИУ5-24М
Название	Создание "истории о данных"

Студент	<hr/>	Ваганов Д.Д. <i>подпись, дата</i>	Ваганов Д.Д. <i>фамилия, и.о.</i>
Преподаватель	<hr/>	Гапанюк Ю.Е. <i>подпись, дата</i>	Гапанюк Ю.Е. <i>фамилия, и.о.</i>

Оценка _____

Москва, 2024 г.

ЗАДАНИЕ

- Выбрать набор данных (датасет).
- Создать "историю о данных" в виде юпитер-ноутбука, с учетом следующих требований:
 1. История должна содержать не менее 5 шагов (где 5 - рекомендуемое количество шагов). Каждый шаг содержит график и его текстовую интерпретацию.
 2. На каждом шаге наряду с удачным итоговым графиком рекомендуется в юпитер-ноутбуке оставлять результаты предварительных "неудачных" графиков.
 3. Не рекомендуется повторять виды графиков, желательно создать 5 графиков различных видов.
 4. Выбор графиков должен быть обоснован использованием методологии data-to-viz. Рекомендуется учитывать типичные ошибки построения выбранного вида графика по методологии data-to-viz. Если методология Вами отвергается, то просьба обосновать Ваше решение по выбору графика.
 5. История должна содержать итоговые выводы. В реальных "историях о данных" именно эти выводы представляют собой основную ценность для предприятия.
- Сформировать отчет и разместить его в своем репозитории на github.

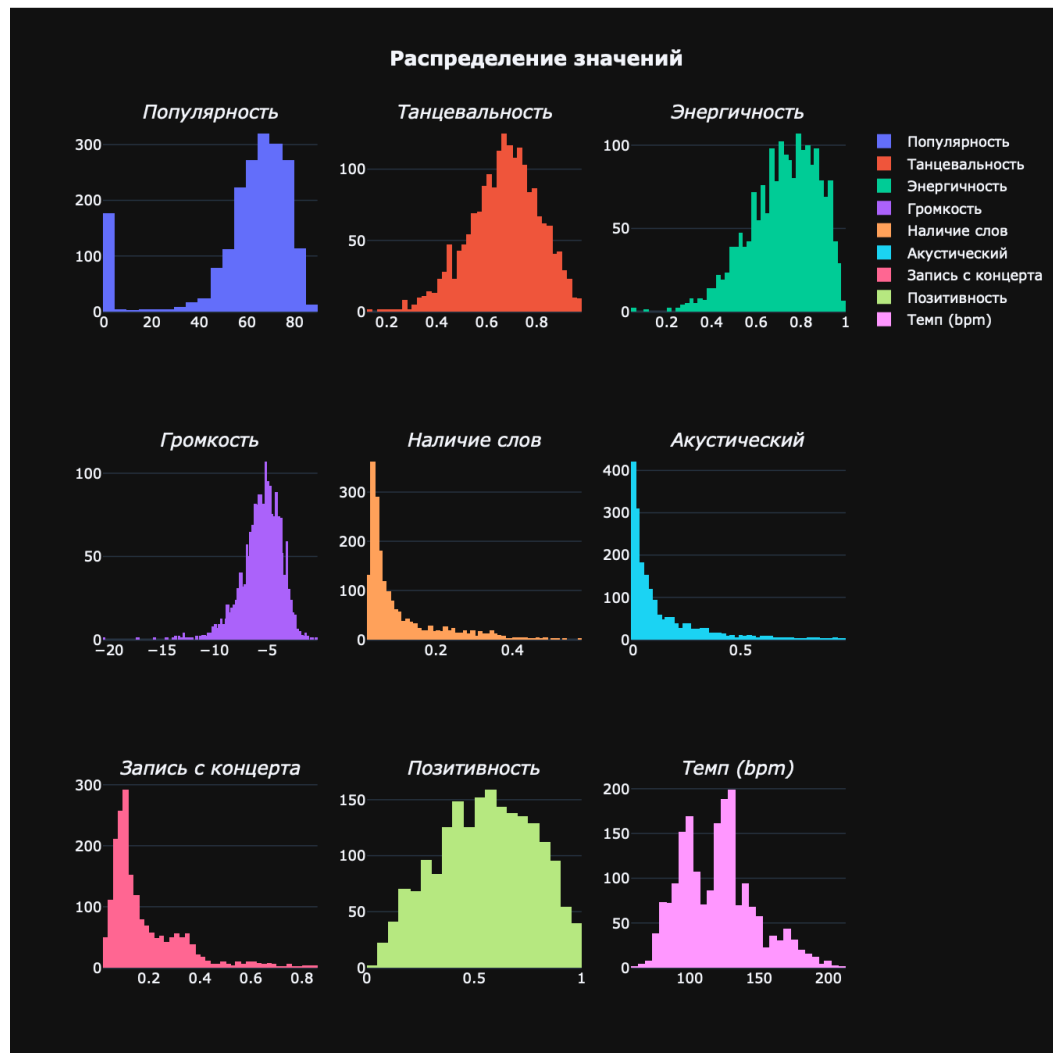
Выполнение лабораторной работы

Набор данных содержит информацию о различных песнях, взятых с платформы потоковой передачи музыки Spotify. Ниже приведены столбцы, содержащиеся в датасете.

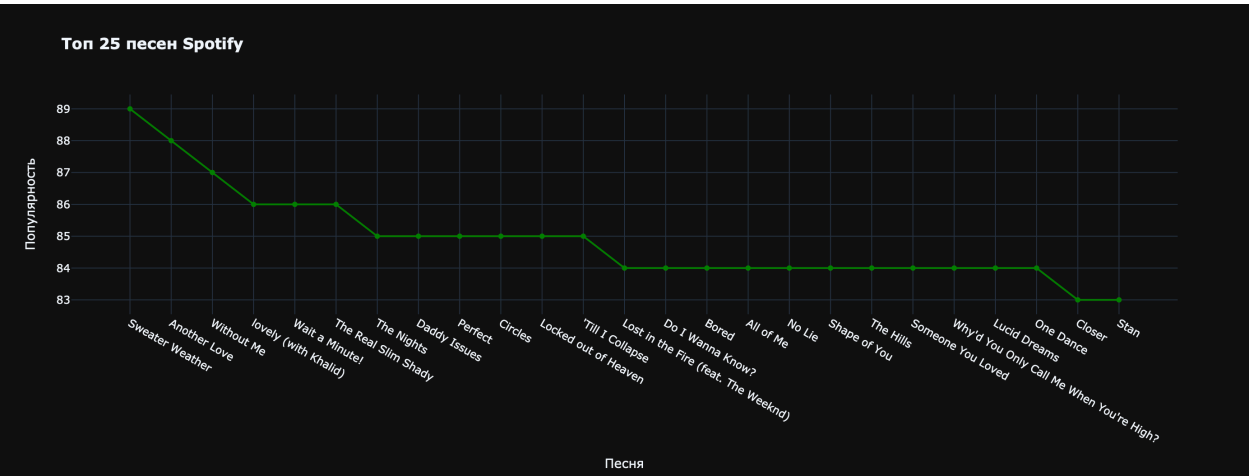
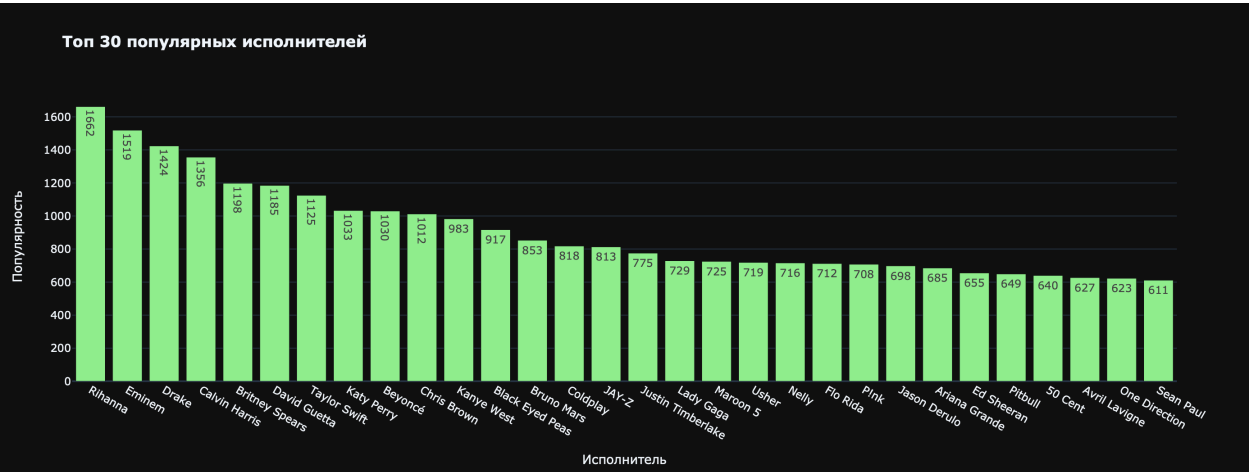
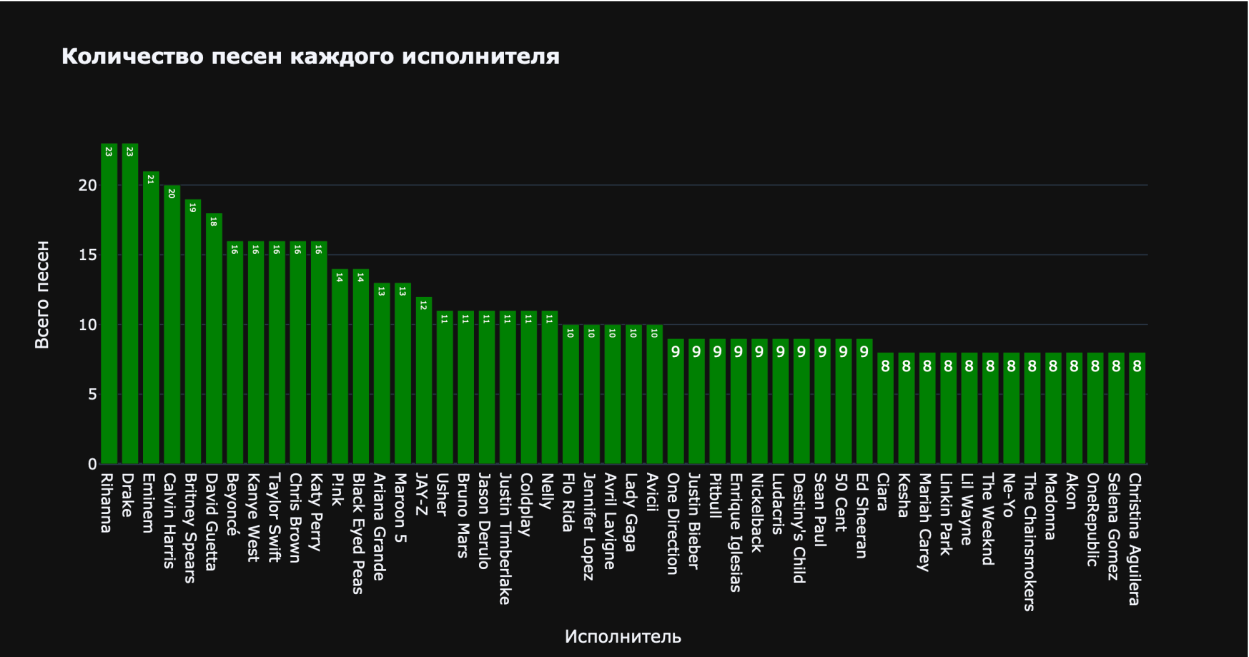
```
[ ] df.head()
```

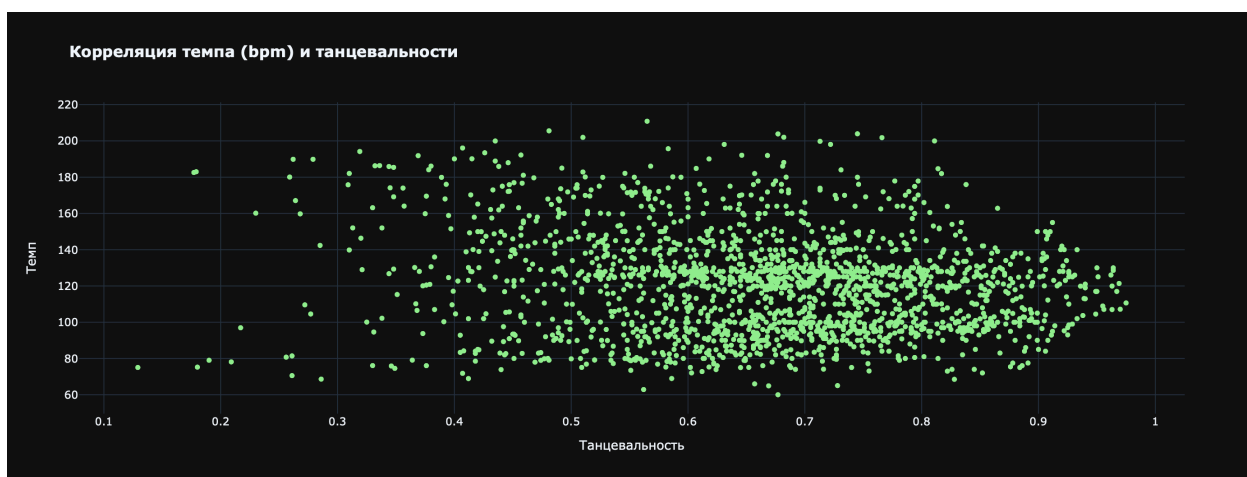
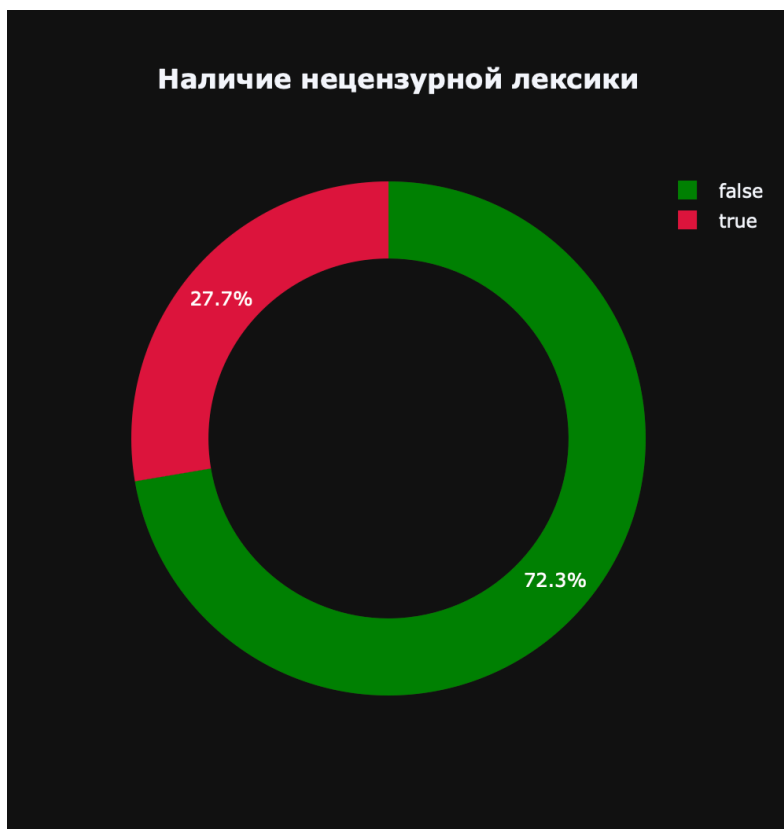
	artist	song	duration_ms	explicit	year	popularity	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence
0	Britney Spears	Oops!...I Did It Again	211160	False	2000	77	0.751	0.834	1	-5.444	0	0.0437	0.3000	0.000018	0.3550	0.894
1	blink-182	All The Small Things	167066	False	1999	79	0.434	0.897	0	-4.918	1	0.0488	0.0103	0.000000	0.6120	0.684
2	Faith Hill	Breathe	250546	False	1999	66	0.529	0.496	7	-9.007	1	0.0290	0.1730	0.000000	0.2510	0.278
3	Bon Jovi	It's My Life	224493	False	2000	78	0.551	0.913	0	-4.063	0	0.0466	0.0263	0.000013	0.3470	0.544
4	*NSYNC	Bye Bye Bye	200560	False	2000	65	0.614	0.928	8	-4.806	0	0.0516	0.0408	0.001040	0.0845	0.879

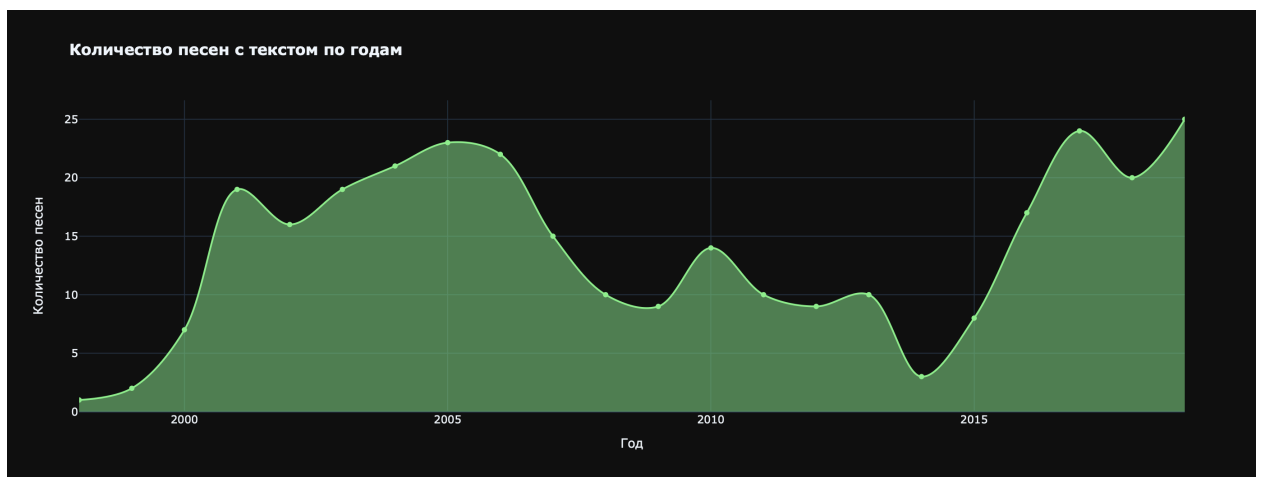
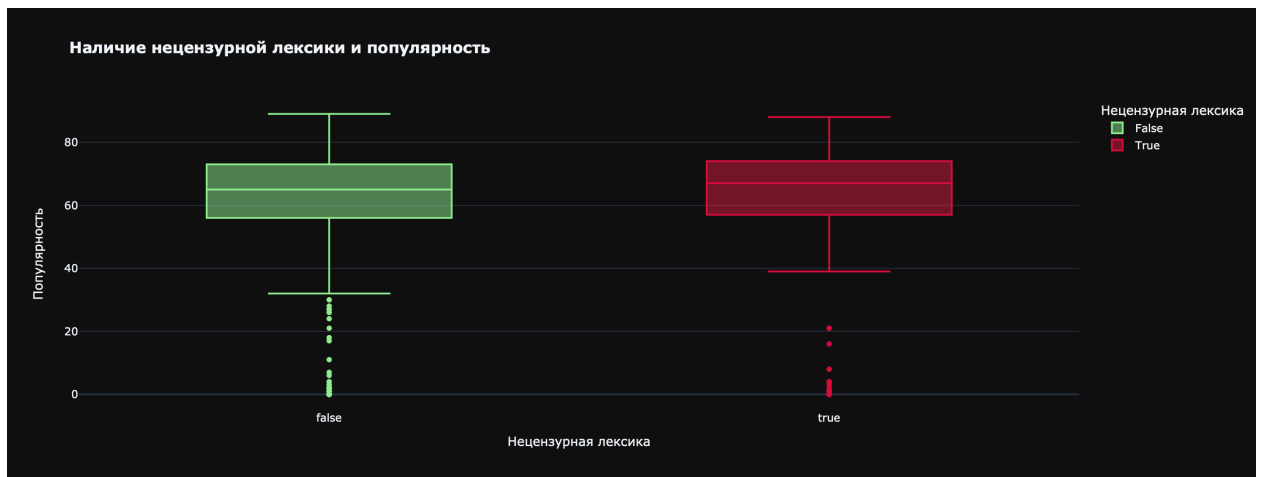
Распределение значений датасета



Визуализация данных







Выводы

В результате выполнения лабораторной работы нами были создана "история о данных" в виде юпитер-ноутбука.

Из анализа можно сделать вывод, что самым популярным жанром на музыкальном рынке является поп, а непопулярным жанром оказался кантри.

Больше всего песен выпустили Rihanna, Drake и Eminem, они же и возглавили топ 3 исполнителей по популярности.

Топ 4 песни на Spotify по популярности: The Neighbourhood – Sweater Weather, Tom Odell – Another Love, Eminem – Without Me, Billie Eilish & Khalid – lovely.

Почти 28% всех песен содержит нецензурный контент. Можно отметить, что медианная популярность песен с таким контентом выше, чем песни без нецензурной лексики.

Для танцевальных песен самый встречающийся темп равняется примерно 130bpm.

С 2006 по 2016 год наблюдается уменьшение количества слов в песнях, одной из возможных причин может быть рост популярности жанра электронной музыки, такой как транс, техно и хаус, где важную роль играет мелодия, ритм и атмосфера, а не наличие слов в песне.