**МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ**

**УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»**

**(РУТ (МИИТ)**

Институт/факультет «Академия “Высшая Инженерная Школа”, АВИШ»

Специальность/Направление подготовки Информатика и вычислительная техника

**КУРСОВАЯ РАБОТА**

**по дисциплине:** Технологии хранения больших данных

|  |  |
| --- | --- |
| **на тему:** | Аналитика чарта топ-100 песен на сайте Яндекс.Музыка на сентябрь 2023 года |

|  |  |
| --- | --- |
| **Студент группы ШАД 211** | **(Рябов С. В.)** |
| **Научный руководитель** | **(Зуев И. А.)** |

**Москва 2023 г.**

**Оглавление**

Введение3

**1. Описание дата-сета4**

**2. Аналитика характеристик фильмов, представленных в рейтинге6**

Заключение11

Список использованного ПО12

Список использованной литературы13

Приложение14

**ВВЕДЕНИЕ**

**Актуальность темы исследования** обусловлена необходимостью предсказания успешности той или иной песни и/или альбома в будущем, основываясь на таких входных данных, как жанр песни, её длительность, принадлежность к тому или иному лейблу, наличию метки «18+», на основании текста/темы продукта. Это позволит оценить возможную выручку, прослушивания и популярность трека в сервисах стриминга в принципе.

**Объект исследования –** песни, представленные в чарте ТОП-100 Яндекс.Музыки.

**Предмет исследования –** особенности представителей чарта в рейтинге ТОП-100 Яндекс.Музыки.

**Целью работы** является анализ песен, представленных в рейтинге ТОП-100 Яндекс.Музыки, по состоянию на сентябрь 2023 года.

В соответствии с этой целью работы решаются следующие **задачи:**

**-** рассмотреть чарт ТОП-100 песен на сервисе Яндекс.Музыка;

**-** проанализировать зависимость расположения песен в рейтинге, в зависимости от характеристик, которыми они обладают;

**-** сделать вывод о корреляции характеристик песен и их влиянии на положение песни в рейтинге;

1. **Описание дата-сета.**

Для исследования был взят дата-сет из песен, входящих в чарт ТОП-100 песен на сайте «Яндекс.Музыка» по состоянию на сентябрь 2023 года. Датасет имеет четкую структуру: первый его столбец (name) хранит название песни (тип данных: string); второй (track\_len) – содержит информацию о длительности песни в формате минуты:секунды (тип данных: string); третий (link) – ссылка на карточку песни с сайта Яндекс.Музыка (тип данных: string); четвертый (genre) – жанр, в котором представлена песня, если в карточке песни указано несколько жанров, то выбирается первый в списке (тип данных: string); пятый (artist(s)) – содержит информацию об артистах, участвовавших в создании трека. (тип данных: string); шестой (chart) – позиция в рейтинге чарта (тип данных: int64); седьмой (Explicit\_content) – указывает на наличие у песни метки «18+», 1 если метка имеется, 0 – если нет (тип данных: int64); восьмой (monthly\_listens\_total) – количество слушателей за месяц (тип данных: int64); десятый столбец (artists\_likes\_total) показывает количество лайков у артистов (кол-во их слушателей, суммарно) (тип int64).

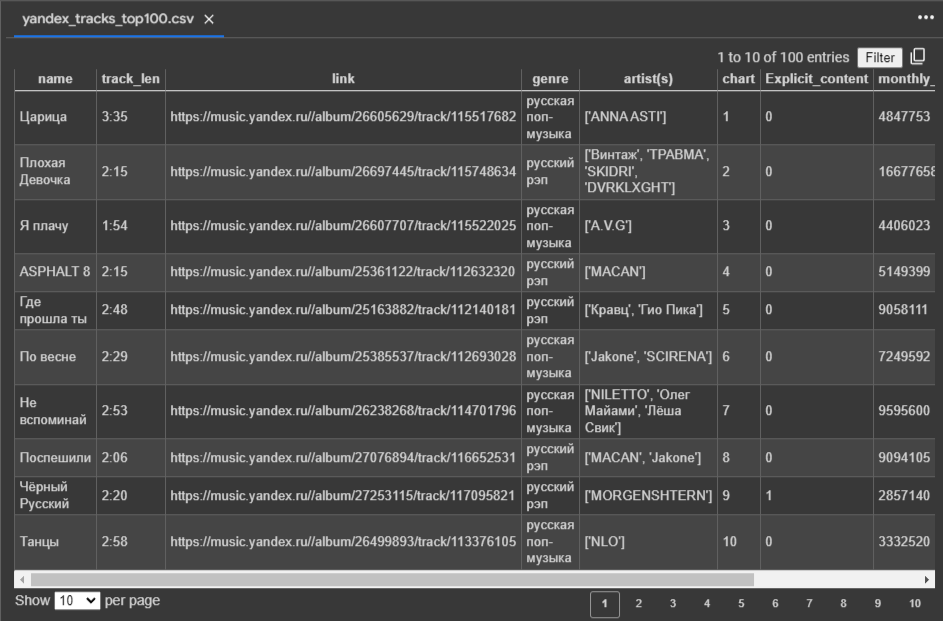


Рис. 1 Фрагмент дата-сета, представленный в Google Colab

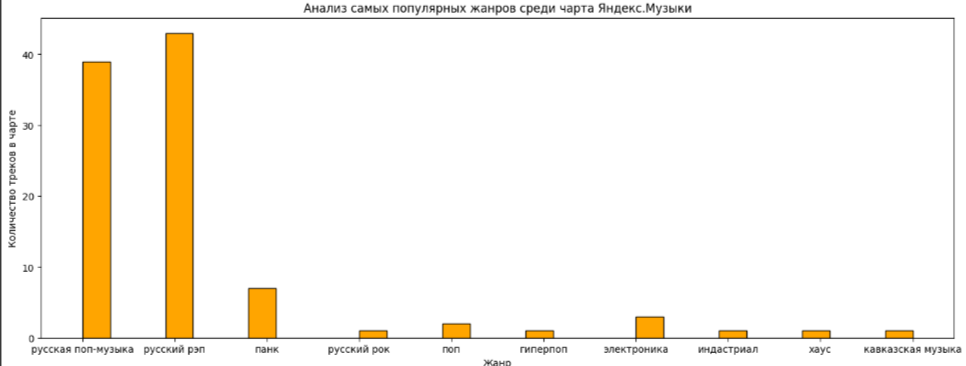
Главным условием для попадания в чарт ТОП-100 Яндекс.Музыки является наибольшее количество слушателей за последние 24 часа на песне. Благодаря данному механизму чарт может обновляться ежедневно.

1. **Аналитика характеристик песен, представленных в рейтинге.**

В рейтинге представлены песни, длительность которых варьируется 96 секунд (1 минута 36 секунд) до 408 секунд (6 минут 48 секунд).

«Типичная» песня в ТОП-100 имеет следующие характеристики: выпущена в жанре русский рэп/русская поп-музыка, имеет длительность в 150 секунд (2 минуты 30 секунд), имеет 4095789 слушателей в месяц, 1677281 лайков за все время. Не имеет возрастное ограничение.

Большинство фильмов в ТОП-100 исполнены в жанрах «русский рэп» (43 песни), «русская поп-музыка» (39 песен), «панк» (7 песен), «электроника» (3 песни) и «поп» (2 песни).

Рис. 2 Распределение песен по жанрам

Что интересно, пользователи чаще слушают русский рэп, но добавляют в плейлист русскую поп-музыку гораздо в большем количестве.



Рис. 8 Диаграмма, показывающая долю ограниченных по возрасту

треков в чарте.

Диаграмма показывает, что 77.8% треков в чарте Яндекс.Музыки не имеют метки «18+». Это значит, что они будут попадаются в волне куда большему количеству людей благодаря алгоритмам.

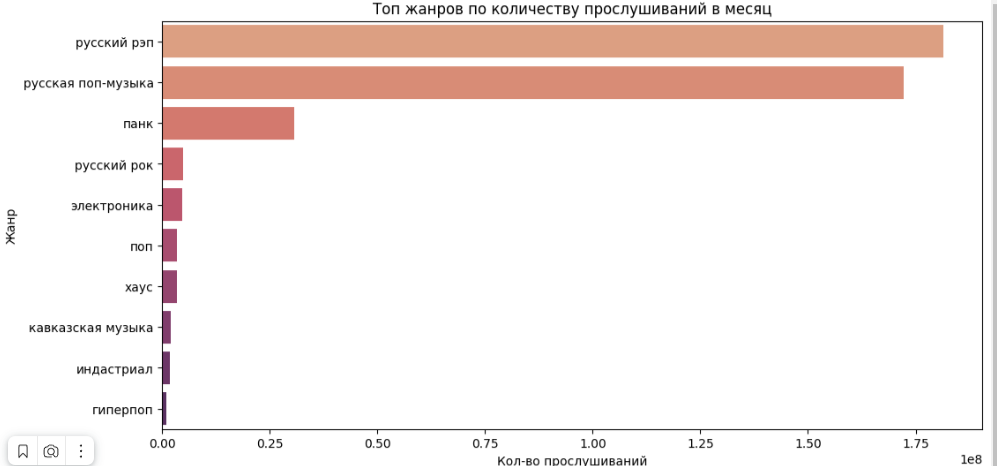


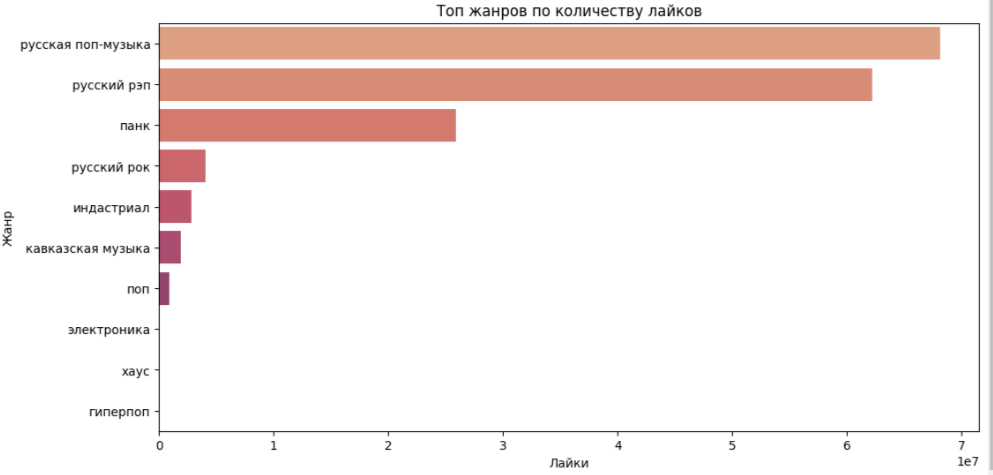
Рис. 3 рейтинг жанров по слушателям

Рис. 4 Рейтинг жанров по лайкам

На представленной выше диаграммах (3) и (4) видно, что музыка в жанре «панк» имеет наибольшую процентную разницу между прослушиваниями и лайками. Это говорит о более узкой, но достаточно активной аудитории.

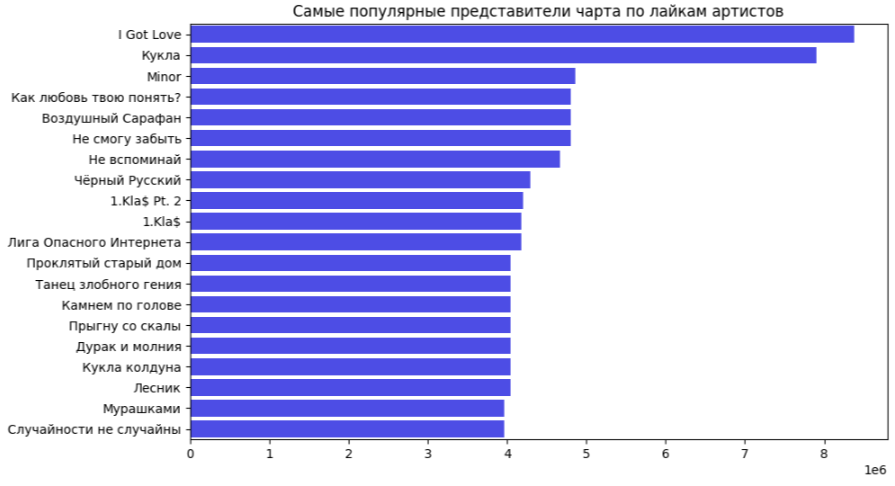


Рис. 5 Диаграмма, показывающая топ-20 треков среди чарта по лайкам артистов.

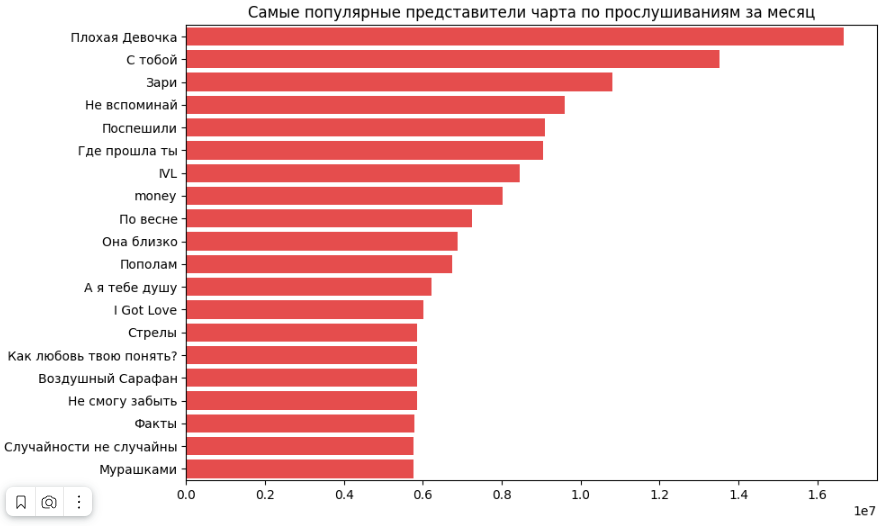


Рис. 6 Диаграммы, показывающая топ-20 треков среди прослушиваний за месяц

Интересно, что среди представленных топ-20, всего три песни входят в оба топа: «Мурашками», «Случайности не случайны» и «Не вспоминай». Остальные кардинально отличаются.

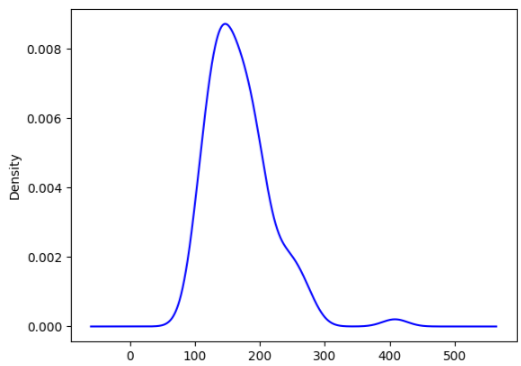


Рис. 7 Диаграмма плотности распределения длительности песен

в чарте.

На данной диаграмме можно заметить, что распределение близко к нормальному, вершина находится на отметке в 150 секунд (2 минуты 30 секунд).

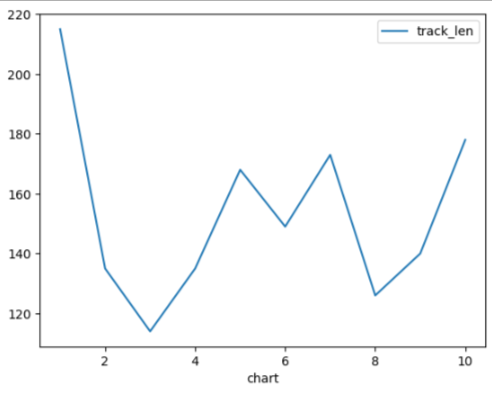


Рис. 7 Диаграмма, показывающая длительность песен в топ-10 чарта.

На диаграмме можно увидеть, что самый длинный трек из десятки – первый, а самый короткий – третий. Все остальные треки имеют длительность от 130 до 180 секунд.

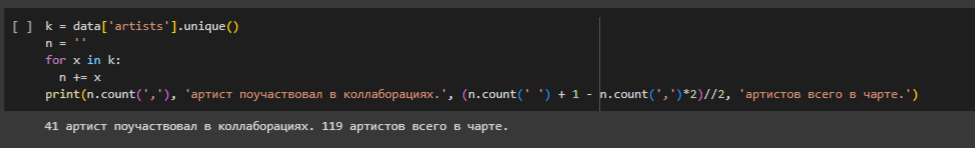
Помимо всего был сделан анализ-подсчёт коллабораций среди песен в чарте, и среди 119 артистов, 41 поучаствовали в песне-коллаборации с другими исполнителями. 

Рис. 10 Поиск количества коллабораций

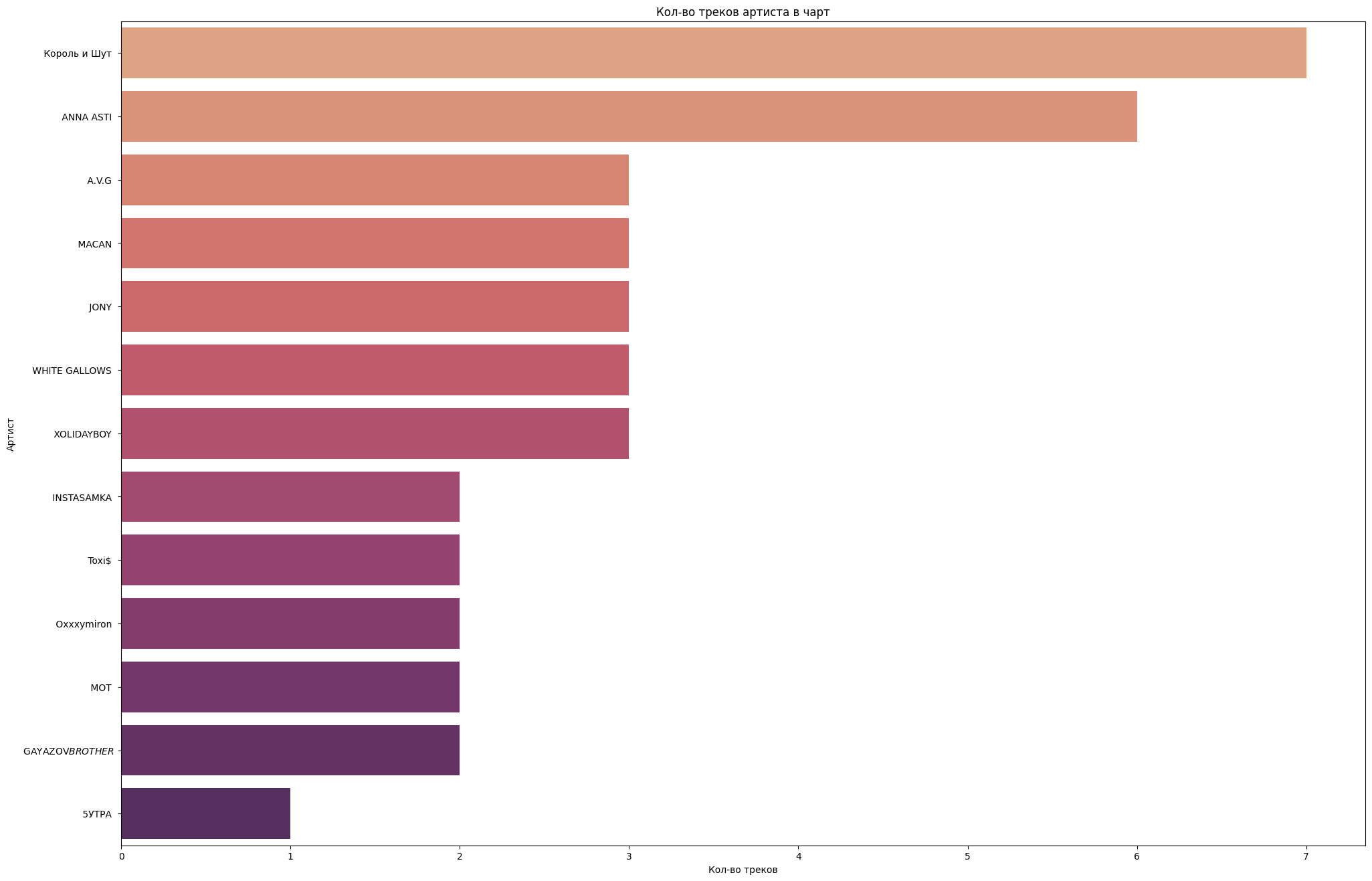


Рис. 11 Самые часто встречающиеся артисты в чарте

Данный анализ показал, что Король и Шут – лидеры и среди 100 песен имеют 7, которые и являются представителями жанра в чарте. Но, здесь не учитываются коллаборации, потому при более детальном анализе их могут обогнать другие артисты, но среди артистов без коллабораций они – лидеры.

**Заключение:**

Задачами данной курсовой работы были:

**-** рассмотреть чарт ТОП-100 песен Яндекс Музыки;

**-** проанализировать зависимость расположения песен в рейтинге, в зависимости от характеристик, которыми они обладают;

**-** сделать вывод о корреляции характеристик песен и их влиянии на положение песни в рейтинге;

Были проанализированы данные о песня, входящих в рейтинг ТОП-100 на сайте Яндекс.Музыка за сентябрь 2023 года. В рамках работы было установлено, что большинство успешных песен записаны в жанре русского рэпа, с небольшим отрывом от него идет русская поп-музыка, но если учитывать количество лайков, то поп-музыка является лидером. Большинство песен в рейтинге записаны в выше представленных жанрах. В рейтинге отсутствуют песни в жанрах «шансон», «инди», «R&B», «кантри», «детская» и др. Большое количество песен, представленных в чарте принадлежат одним и тем же артистам. Также, в топе музыки достаточно большое количество коллабораций популярных артистов, что помогает им получать в среднем больше прослушиваний. Не все песни в чарте выпущены в последний год. Присутствуют и уже несуществующие артисты и коллективы.

Таким образом можно заключить, что наиболее успешные песни имеют длительность около 150 секунд (2,5 минуты), исполнены в жанрах «русский рэп», «русская поп-музыка» и «панк», а также не имеют метки 18+ – это обусловлено разновозрастной аудиторией. Самым часто встречающимся артистом в чарте является группа «Король и Шут», а также исполнители «ANNA ASTI» и «MACAN»

**Список использованного ПО**

1. Google Colab.
2. matplotlib.
3. Microsoft Excel.
4. Microsoft Word.
5. Numpy.
6. Pandas.
7. Python 3.
8. seaborn.

**Список литературы**

1. Рейтинг – Яндекс Музыка Изнутри. [Официальный подкаст] // URL: https://music.yandex.ru/album/11904129 (дата обращения: 10.12.2023)
2. Документация к методам Pandas [Электронный ресурс] // URL: https://pandas.pydata.org/docs/ (дата обращения: 08.12.2022)
3. Pyplot tutorial – Matplotlib 3.8.2 documentation [Электронный ресурс] // Системные требования: Google Chrome версии 108 или выше. URL: https://matplotlib.org/stable/tutorials/pyplot.html (дата обращения: 09.12.2022)

**Приложение**

Исполняемый код представлен в Google Colab по ссылке:

https://colab.research.google.com/drive/1dgJmtMI7\_Sk4E3u\_h6W6j3ZK7aozJ8KS#scrollTo=piiin3GxeSeO

Дата-сет доступен по ссылке: https://docs.google.com/spreadsheets/d/1HQEpRGSrn4\_crPygNs1BUG7jYnOMxieXWGq5O4cWcks/edit#gid=1156984846