

T3 SQL

Структура синтетически созданного датасета:

row_id - уникальный идентификатор записи

user_id - идентификатор пользователя

song - название композиции

artist - исполнитель композиции

genre - музыкальный жанр

region - регион прослушивания

day - день недели прослушивания

duration - длительность прослушивания в секундах

share - доля прослушивания трека

1. Вывести первые 10 строк, чтобы изучить ее, при этом не перегружать БД
2. Вывести первые 15 строк с информацией об исполнителе композиции, названии и музыкальном жанре
3. Выгрузите все уникальные жанры музыки и отсортируйте их
4. Определить, какие жанры наиболее популярны среди пользователей: вывести топ пяти жанров по количеству прослушиваний
5. Коллеги хотели бы визуализировать длительность прослушивания треков разных исполнителей: необходимо вывести название трека, исполнителя, длительность прослушивания трека в минутах; возможно в выборке у исполнителя окажутся одинаковые треки разной длительности - выведем каждую
6. Для маркетинговой кампании нужно отобрать жанры, у которых суммарная длительность прослушиваний треков превышает 1870 минут
7. Собрать статистику прослушиваний музыки по жанрам: для каждого жанра - общее количество и средняя длительность прослушиваний композиций; суммарная, минимальная, максимальная длительность прослушивания композиций; прослушивания треков.
8. Выгрузить все прослушивания треков жанра "Reggae" в северо-западном регионе
9. Сравнить музыкальные вкусы регионов в одной выборке: выведем информацию о жанре, регионе прослушивания, количестве композиций жанра и суммарной длительности их прослушивания
10. Изучить, сколько раз треки прослушали менее чем наполовину - отразим общее количество таких прослушиваний и их среднюю длительность в минутах
11. Для анализа прослушивания треков в понедельник и среду необходима выборка, которая отразит день недели прослушивания, композицию, ее исполнителя, длительность и долю прослушивания для всех воспроизведений в понедельник и среду

12. Изучить, как слушают треки определённых жанров по регионам: выведем информацию о регионе, жанре, названии композиции, об исполнителе, длительности и доли прослушивания. Ограничиться первыми двадцатью строками
13. Отобразить уникальные пары исполнителя и жанра, где исполнитель имеет более одной композиции в этом жанре, из выборки исключить все строки где (если) исполнитель неизвестен
14. Вычислить длину композиции, используя длительность в секундах и долю прослушивания. Исходную длину вычислять только для прослушиваний со значением доли больше 0
15. Провести классификацию композиций по жанрам со значениями: Pop - категория 'Popular Music'; Rock - категория 'Rock Music'; Jazz - категория 'Jazz Music'. Если жанр не указан или не входит в эти категории, то 'Other'
16. Разделить прослушивания на категории по их длительности и посчитать количество в каждой категории: меньше 180 секунд - 'Short'; от 180 до 300 секунд - 'Medium'; более 300 секунд - 'Long'