Национальный исследовательский ядерный университет «МИФИ»



Институт интеллектуальных кибернетических систем

Кафедра кибернетики (№ 22)

Отчёт о работе по курсу «Базы данных (теоретические основы баз данных)» Вариант «Яндекс Музыка»

Выполнил	Федоров Е.А.
Группа	Б19-501
Вариант	Яндекс Музыка
Преподаватель	Петровская А.В.
Проверяющий	Старков С.В
Оценка	

Оглавление

1. Формулировка задания	3	
2. Концептуальная модель	ь базы данных	3
3. Логическое проектиров	зание. 7	
4. Физическое проектиров	вание 8	
5. Выполнение запросов	21	

1. Формулировка задания

Спроектировать базу данных для интернет-сервиса прослушивания музыки «Яндекс Музыка». База должна содержать информацию о пользователях, пользующихся сервисом, о треках и сборниках треков и их исполнителях, музыкальных предпочтениях пользователей.

2. Концептуальная модель базы данных

В ходе работы над заданием была проанализирована предметная область, особенности которой отражает построенная концептуальная модель:

Концептуальная модель базы данных для предметной области «Яндекс Музыка» (Рис. 1).

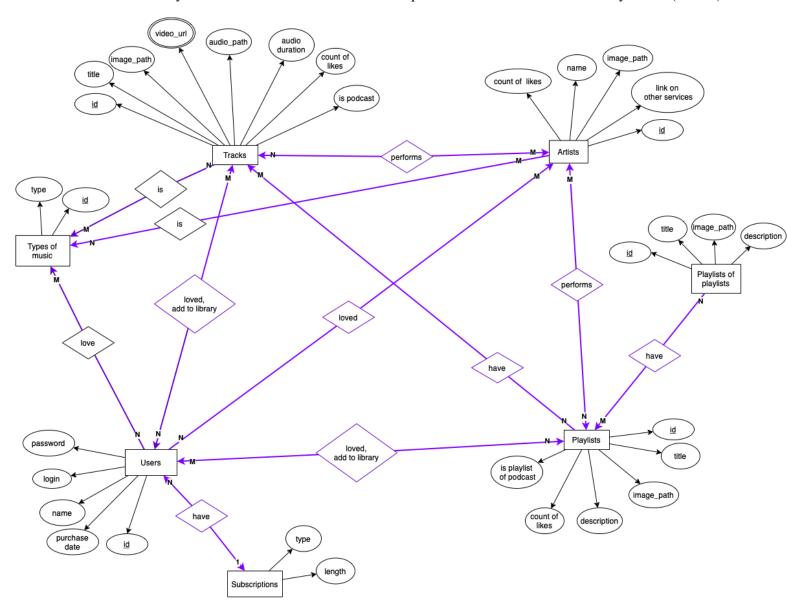


Рис. 1 - Концптуальная модель

2.1. Конкретизация предметной области

Необходимо создать систему, которая отражает основную информацию о музыке, пользователях и артистах, находящихся в интернет сервисе "Яндекс Музыка". Система также должна хранить информацию о музыкальных предпочтениях пользователей, информацию о подписке пользователя, информацию о треках и сборниках треков и их исполнителях.

2.2. Описание предметной области

Музыкальная площадка «Яндекс Музыка» ориентирована на зарегистрированных пользователей и не предоставляет незарегестрированным пользователям возможность прослушивания музыки.

Рассмотрим далее возможности, которыми обладают пользователи:

- 1. Пользователь может добавить трек в избранное, так как "Яндекс Музыка" это платформа для прослушивания музыки, то этот функционал является обязательным.
- 2. Пользователь может добавить исполнителя в избранное, этот функционал также является очень нужным, по тем же причинам что описаны в предыдущем пункте.
- 3. Аналогично первым двум пунктам, пользователь может добавить в избранное плейлист.
- 4. Также пользователь может установить жанры музыки которые он предпочитает, это также является обязательным, так как пользователи хотят видеть в рекомендациях именно ту музыку которая им потенциально может понравится.

Исходя из описания возможностей, предоставляемых интернет-сервисом «Яндекс Музыка», были выделены следующие сущности:

- 1. Пользователь (User);
- 2. Трек (Track);
- 3. Плейлист (Playlist);
- 4. Плейлист плейлистов (Playlist of playlist);
- 5. Артист (Artist);
- 6. Подписка (Subscription)
- 7. Жанр музыки (Type of music);

Для сущности пользователь (User):

Атрибут	Расшифровка	
id	Уникальный идентификатор пользователя.	
name	Имя пользователя	
login	Логин пользователя	
password	Пароль пользователя	
type of subscription	Тип подписки пользователя	
purchase date	Дата покупки подписки	

Для сущности трек (Track):

Атрибут	Расшифровка	
id	Уникальный идентификатор трека.	
title	Название трека.	
image_path	Ссылка на обложку трека	
video path	Ссылка на видео(у некоторых треков есть видео клипы)	
audio path	Ссылка на аудио файл трека	
audio duration	Продолжительность трека	
count of likes	Количество лайков на треке	
is podcast	Свойство которое указывает на то, подкаст это или нет	

Для сущности подписка (Subscription):

Атрибут	Расшифровка	
type	Тип подписки	
length	Длина подписки в днях	

Для сущности жанра музыки (Type of music)

Атрибут	Расшифровка	
id	Уникальный идентификатор жанра.	
type	Жанр музыки	

Атрибут	Расшифровка	
id	Уникальный идентификатор артиста.	
name	Имя артиста	
image_path	Ссылка на обложку артиста	
count of likes	Количество лайков у артиста	
link on other services	Ссылка на любой другой сервис исполнителя(один сервис)	

Для сущности плейлист (Playlist):

Атрибут	Расшифровка	
id	Уникальный идентификатор плейлиста.	
title	Название плейлиста	
image_path	Ссылка на обложку плейлиста	
description	Описание плейлиста	
count of likes	Количество лайков на плейлисте	
is playlist of podcast	Логическое значение указывающее на то, плейлист подкастов это или же нет	

Для сущности плейлист плейлистов (Playlist of playlist):

Атрибут	Расшифровка	
id	Уникальный идентификатор плейлиста плейлистов	
title	Название плейлиста плейлистов	
image_path	Ссылка на обложку плейлиста плейлистов	
description	Описание плейлиста плейлистов	

3. Логическое проектирование.

Далее, на основе КМПО была разработана логическая модель базы данных, все связи «многие ко многим» были реализованы в виде отдельных таблиц (Рис. 2). Стоить отметить, что данная база данных приведена к 3 нормальной форме, соответственно к 1 и 2 тоже.

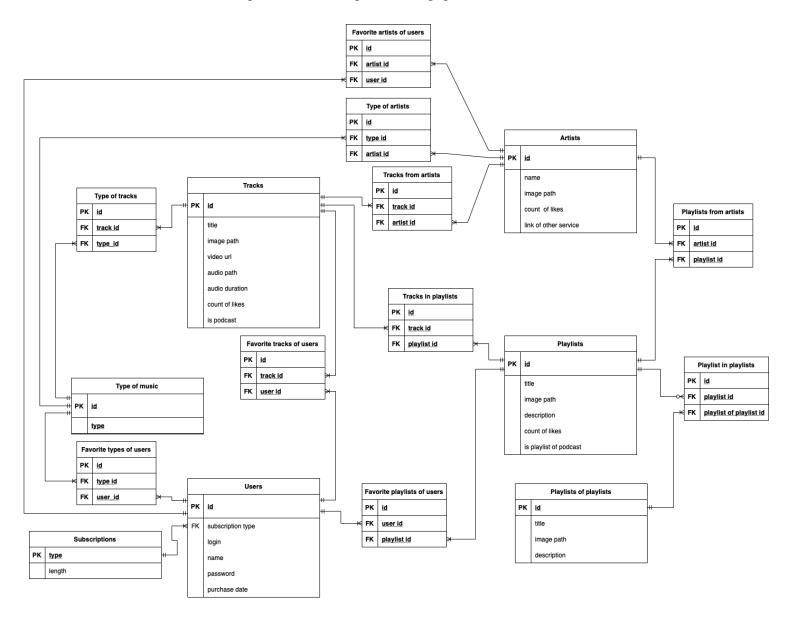


Рис. 2 - Логическая модель

4. Физическое проектирование

В качестве СУБД для реализации разработанной базы данных был выбран DataGrip. Исходя из приведенной логической модели была построена следующая физическая модель базы данных (Рис. 3):

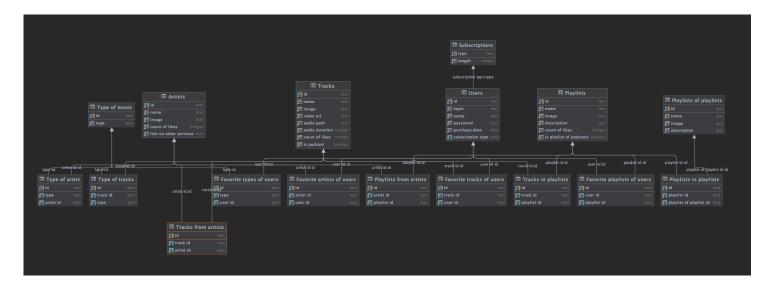


Рис. 3 - Физическая модель

На приведенном рисунке типы данных обозначаются следующими значениями:

- · «integer» целое число;
- · «text» строка;
- · «boolean» логическое значение;
- · «date» -дата;
- · Значок с желтым ключом primary key;
- · Значок с голубым ключом foreign key;

4.1. Создание таблиц

Создание таблицы артистов (Artists):

```
CREATE TABLE "Artists"

(
   id INT,
   name TEXT NOT NULL,
   "image path" TEXT NOT NULL,
   "count of likes" INT NOT NULL DEFAULT

0,
   "link on other services" TEXT NOT
NULL,
   PRIMARY KEY (id)
);
```

Создание таблицы плейлистов (Playlists):

```
CREATE TABLE "Playlists"

(

id INT,

title TEXT NOT NULL,

"image path" TEXT NOT NULL,

description TEXT NOT NULL,

"count of likes" INT NOT NULL DEFAULT

0,

"is playlist of podcasts" BOOl NOT

NULL DEFAULT FALSE,

PRIMARY KEY (id)

);
```

Создание таблицы плейлиста плейлистов (Playlist of playlists):

```
CREATE TABLE "Playlists of playlists"

(
   id INT,
   title TEXT NOT NULL,
   "image path" TEXT NOT NULL,
   description TEXT NOT NULL,
   PRIMARY KEY (id)
);
```

Создание таблицы подписок (Subscriptions):

```
CREATE TABLE "Subscriptions"

(
type TEXT,
length INT NOT NULL,
PRIMARY KEY (type)
);
```

Создание таблицы треков (Tracks):

```
CREATE TABLE "Tracks"

(

id INT,

title TEXT NOT NULL,

"image path" TEXT NOT NULL,

"video url" TEXT,

"audio path" TEXT NOT NULL,

"audio duration" INT NOT NULL,

"count of likes" INT NOT NULL DEFAULT 0,

"is podcast" BOOl NOT NULL DEFAULT FALSE,

PRIMARY KEY (id)

);
```

Создание таблицы типа музыки (Type of music):

```
CREATE TABLE "Type of music"

(
   id INT,
   type TEXT NOT NULL,
   PRIMARY KEY (id)
);
```

Создание таблицы пользователей (Users):

```
CREATE TABLE "Users"

(

id INT,

login TEXT NOT NULL,

name TEXT NOT NULL,

password TEXT NOT NULL,

"purchase date" DATE NOT NULL,

"subscription type" TEXT NOT NULL,

PRIMARY KEY (id),

FOREIGN KEY ("subscription type") REFERENCES "Subscriptions"(type)

);
```

Создание таблицы, связывающей артистов и юзеров(Артисты которых слушает юзер) (Favorite artists of users):

```
CREATE TABLE "Favorite artists of users"

(
   id INT,
   "artist id" INT NOT NULL,
   "user id" INT NOT NULL,
   PRIMARY KEY (id),
   FOREIGN KEY ("artist id") REFERENCES "Artists"(id),
   FOREIGN KEY ("user id") REFERENCES "Users"(id)
);
```

Создание таблицы, связывающей плейлисты и юзеров (Favorite playlists of users):

```
CREATE TABLE "Favorite playlists of users"

(
   id INT,
   "user id" INT NOT NULL,
   "playlist id" INT NOT NULL,

PRIMARY KEY (id),

FOREIGN KEY ("user id") REFERENCES "Users"(id),

FOREIGN KEY ("playlist id") REFERENCES "Playlists"(id)
);
```

Создание таблицы, связывающей треки и юзеров (Favorite tracks of users):

```
CREATE TABLE "Favorite tracks of users"

(

id INT,

"track id" INT NOT NULL,

"user id" INT NOT NULL,

PRIMARY KEY (id),

FOREIGN KEY ("track id") REFERENCES "Tracks"(id),

FOREIGN KEY ("user id") REFERENCES "Users"(id)

);
```

Создание таблицы, связывающей тип музыки с юзером (Favorite types of users):

```
CREATE TABLE "Favorite types of users"

(
   id INT,
   "type id" INT NOT NULL,
   "user id" INT NOT NULL,
   PRIMARY KEY (id),
   FOREIGN KEY ("type id") REFERENCES "Type of music"(id),
   FOREIGN KEY ("user id") REFERENCES "Users"(id)
);
```

Создание таблицы, связывающей плейлист с артистом (Playlists from artists):

```
CREATE TABLE "Playlists from artists"

(
   id INT,
   "artist id" INT NOT NULL,
   "playlist id" INT NOT NULL,

PRIMARY KEY (id),

FOREIGN KEY ("artist id") REFERENCES "Artists"(id),

FOREIGN KEY ("playlist id") REFERENCES "Playlists"(id)
);
```

Создание таблицы, связывающей плейлист плейлистов с плейлистом(Playlists in playlists):

```
CREATE TABLE "Playlists in playlists"

(
    id INT,
    "playlist id" INT NOT NULL,
    "playlist of playlist id" INT NOT NULL,

PRIMARY KEY (id),

FOREIGN KEY ("playlist id") REFERENCES "Playlists"(id),

FOREIGN KEY ("playlist of playlist id") REFERENCES "Playlists of playlists"(id)
);
```

Создание таблицы, связывающей треки с артистами(Tracks from artists):

```
CREATE TABLE "Tracks from artists"

(

id INT,

"track id" INT NOT NULL,

"artist id" INT NOT NULL,

PRIMARY KEY (id),

FOREIGN KEY ("track id") REFERENCES "Tracks"(id),

FOREIGN KEY ("artist id") REFERENCES "Artists"(id)

);
```

Создание таблицы, связывающей треки с плейлистами(Tracks in playlists):

```
CREATE TABLE "Tracks in playlists"

(

id INT,

"track id" INT NOT NULL,

"playlist id" INT NOT NULL,

PRIMARY KEY (id),

FOREIGN KEY ("track id") REFERENCES "Tracks"(id),

FOREIGN KEY ("playlist id") REFERENCES "Playlists"(id)

);
```

Создание таблицы, связывающей тип музыки с артистами(Type of artists):

```
CREATE TABLE "Type of artist"

(

id INT,

"type id" INT NOT NULL,

"artist id" INT NOT NULL,

PRIMARY KEY (id),

FOREIGN KEY ("type id") REFERENCES "Type of music"(id),

FOREIGN KEY ("artist id") REFERENCES "Artists"(id)

);
```

Создание таблицы, связывающей тип музыки с артистами(Type of tracks):

```
CREATE TABLE "Type of tracks"

(
  id INT,
  "track id" INT NOT NULL,
  "type id" INT NOT NULL,
  PRIMARY KEY (id),
  FOREIGN KEY ("track id") REFERENCES "Tracks"(id),
  FOREIGN KEY ("type id") REFERENCES "Type of music"(id)
);
```

4.2. Заполнение базы данных

Все базы заполнялись скриптом на языке Swift, представленном в приложении

4.3. Результаты заполнения

Таблица артистов(Artists) в ней 150 записей:

	. ∰id ≎	■ name ÷	.⊞ image path	. count of likes ≎	III link on other services ≎
1	0	John Davis	https:\B1mPY0r	249720	https:\KWbypb8Mn4
2	1	Teresa Hill	https:\N3R3jrI	639879	https:\rHBuligG1U
3	2	Louise Robinson	https:\aVQU5Kz	717934	https:\ExG9S1hMk3
4	3	Justin Webster	https:\2ubnXbG	510269	https:\eS6bc9c8QT
5	4	Walter Washington	https:\D6lX2l4	387007	https:\AYumqgNSfN
6	5	Frances Alexander	https:\YQkjKm3	494517	https:\m0Yf81uG4A
7	6	William Colon	https:\nynqF8d	188848	https:\fD68HSJYPh
_		M2 - b 2 U222	h + + + \ - - MO	05/07	h + + \ \ \ T \ \ 0.4 h F

Таблица любимых артистов(Favorite artists of users) в ней 423 записи:

	.id ÷	📭 artist id 🕏	. user id ≎
1	0	51	0
2	1	116	0
3	2	92	1
4	3	39	1
_	,		4

Таблица любимых плейлистов(Favorite playlists of users) в ней 444 записи:

	驔 id 🕏	🌠 user id 🕏	📭 playlist id 🕏
1	0	0	100
2	1	0	111
3	2	1	134
4	3	1	29
5	4	1	70
,	E	2	0.4

Таблица любимых треков(Favorite tracks of users) в ней 470 записей:

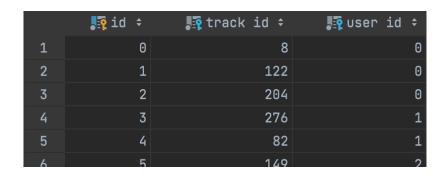


Таблица любимых типов музыки(Favorite types of users) в ней 426 записей:

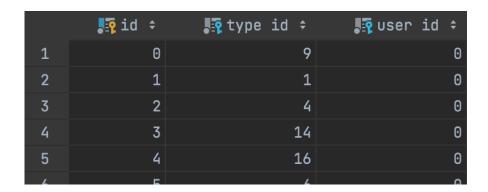


Таблица плейлистов(Playlists) в ней 150 записей:

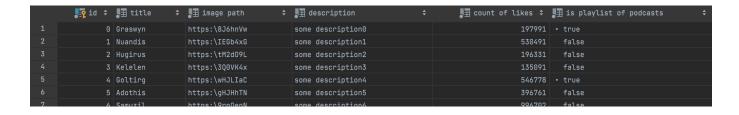


Таблица плейлистов от артистов(Playlists from artrists) в ней 500 записей:

	. id ≎	📭 artist id 🕏	驔 playlist id 🕏
1	0	0	47
2	1	0	65
3	2	0	93
4	3	0	25
5	4	1	82
6	5	1	149

Таблица плейлистов в плейлисте(Playlists in playlists) в ней 150 записей

	id ÷	📭 playlist id 🕏	📭 playlist of playlist id 🕏
1	0	139	0
2	1	56	1
3	2	101	1
4	3	1	2
5	4	82	3

Таблица плейлистов плейлистов(Playlists of playlist) в ней 150 записей

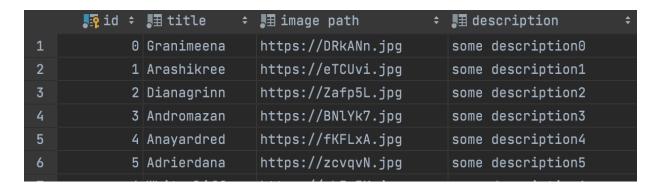
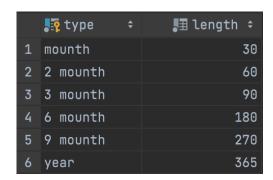


Таблица подписок Subscriptions(представлена полностью)



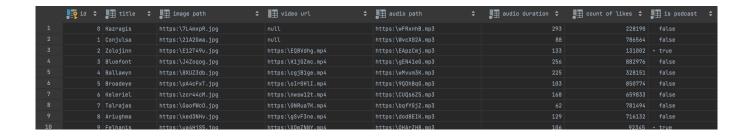


Таблица треков от артистов Tracks from artists 500 записей

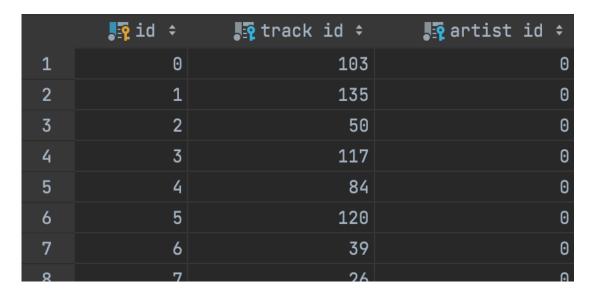
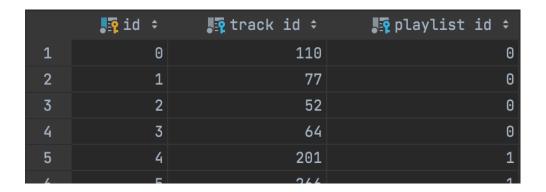


Таблица треков в плейлисте Tracks in playlists 500 значений



. id ≎	. ■ name ÷	.⊞ image path ≎	. count of likes ≎	■ link on other services ÷
132	John King	https:\wbAJZ6I	268457	https:\TQ1uMzR24d
133	John Miller	https:\ScNEYp2	218262	https:\H8y0IQuzn0
134	Kim McGuire	https:\okfzxGT	208256	https:\FDYCGofawp
135	Matthew Wright	https:\GGDKN8P	676354	https:\EoKV8wIOBN
136	Glen Diaz	https:\1IWjlXz	368093	https:\c0fy14AUqw
137	Julia Phillips	https:\9UPEQWY	216413	https:\pIooPuS4JS
130	Michalla Schnaidan	https:\/nWOTip	990761	h++ns·\i2nZN@H+MS

Таблица

типов музыки Type of music представлена полностью

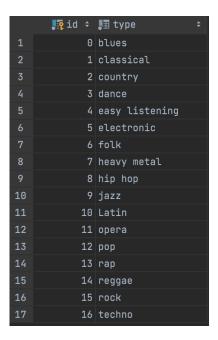
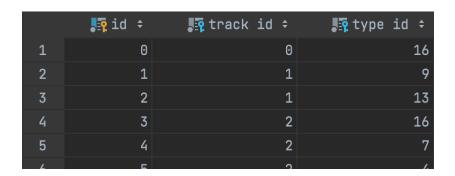
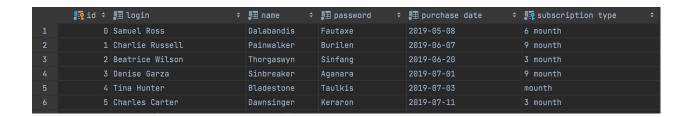


Таблица типа треков Type of tracks 500 значений





5. Выполнение запросов

В этом разделе приведены запросы, которые были написаны к разработанной базе данных в рамках курса.

1) Вывести id 5 треков у которых нет видео

```
SELECT id FROM "Tracks"

WHERE "video url" IS NOT NULL

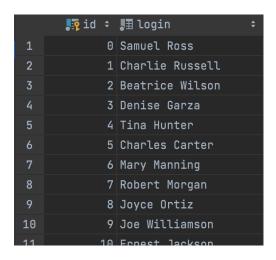
ORDER BY "count of likes" DESC LIMIT(5);
```

1 182 2 284 3 168 4 66 5 287		驔 id 🕏
3 168 4 66	1	182
4 66	2	284
	3	168
5 287	4	66
	5	287

2) Вывести id и login тех пользователей у кого подписка куплена раньше 2019-10-10

```
SELECT id, login FROM "Users"
WHERE "purchase date" <= '2019-10-10';</pre>
```

Результат(показана не вся таблица, потому что она весьма большая)



3) Вывести id login name у тех людей которые слушают подкасты длинней 200 секунд

```
SELECT id, login, name FROM "Users"

WHERE id in (
    SELECT "user id" FROM "Favorite tracks of users"

WHERE "track id" in (
    SELECT id FROM "Tracks"

WHERE "is podcast" = true AND "audio duration" >= 600
)
);
```

Результат(показана не вся таблица)



4) Вывести сколько треков каждого исполнителя есть у пользователя в порядке возрастания

Результат(только часть таблицы)

	■ artist id ‡	■ countoftracks ≎
20	125	1
21	26	1
22	49	2
23	2	2

5) Обновить количество лайков у исполнителей(Пробежались по всем трекам и посчитали сумму)

```
CREATE TEMP TABLE tempTable AS SELECT "artist id", SUM("Tracks"."count of likes")
AS countOfLikes FROM

"Tracks from artists" JOIN "Tracks" ON "Tracks from artists"."track id" =
"Tracks".id

GROUP BY "artist id";

UPDATE "Artists"

SET "count of likes" = tempTable.countOfLikes

FROM tempTable

WHERE "Artists".id = tempTable."artist id";
```

6) Смотрим самого прослушиваемого артиста и смотрим сколько его треки слушают без добавления (На след. странице)

```
WITH MostLikedArtists AS (
   SELECT id as "artist id" FROM "Artists"
       SELECT
              MAX("count of likes")
       FROM "Artists"
), TrackOfMostLikedArtistsStatistics AS (
          COUNT("user id") AS count of love users
   FROM "Favorite tracks of users"
       FROM MostLikedArtists JOIN "Tracks from artists" USING("artist id")
   ORDER BY count_of_love_users
), TracksOfMostLikedFullInfoStatistics AS (
   SELECT
       "count of likes" - count_of_love_users as likes_without_add_track_in_library
   FROM TrackOfMostLikedArtistsStatistics JOIN "Tracks"
       ON TrackOfMostLikedArtistsStatistics."track id" = "Tracks".id
   ORDER BY likes_without_add_track_in_library DESC
```

```
TracksFullInfoStatisticsWithPlaylist AS (
    SELECT
        "track id",
        TracksOfMostLikedFullInfoStatistics.title,
        likes_without_add_track_in_library,
        "playlist id"
    FROM TracksOfMostLikedFullInfoStatistics
        JOIN "Tracks in playlists" USING("track id")
), FinalStat AS (
    SELECT
        "track id",
        TracksFullInfoStatisticsWithPlaylist.title,
        likes_without_add_track_in_library,
        "playlist id",
        "Playlists".title as playlist_title
    FROM TracksFullInfoStatisticsWithPlaylist JOIN "Playlists" ON
TracksFullInfoStatisticsWithPlaylist id" = "Playlists".id
)
SELECT * FROM FinalStat
```

Результат(и исполнителя один трек)

- 7 Запрос
- 1) Берем юзера у которого больше всего жанров музыки лайкнуто
- 2) Смотрим все треки которые ему понравились и смотрим на их жанр, выявляем жанр который чаще всего встречается в треках
- 3) Берем понравившиеся ему плейлисты и смотрим треки в них и выявляем опять самый частовстречающийся жанр
- 4) Таким образом мы имеем два жанра и количество треков в этом жанре
- 5) И так мы понимаем тот жанр который больше всего нравится пользователю

```
WITH MostLikedTypedUser as (
          COUNT("user id") as count_of_liked_types
   FROM "Favorite types of users"
   ORDER BY count_of_liked_types DESC
), MostLikedTypedUserIds AS (
   SELECT "user id" FROM MostLikedTypedUser
), TracksOfUser AS (
   FROM "Favorite tracks of users"
       SELECT * FROM MostLikedTypedUserIds
), CountOfTracksWithType AS (
   SELECT "type id", COUNT("type id") as count_of_tracks_with_type
   FROM TracksOfUser
            JOIN "Type of tracks" USING ("track id")
   ORDER BY count_of_tracks_with_type DESC
), PlaylistOfUser AS (
   FROM "Favorite playlists of users"
       SELECT * FROM MostLikedTypedUserIds
```

```
CountOfPlaylistWithType AS (
    SELECT "track id" FROM "Tracks in playlists"
       SELECT * FROM PlaylistOfUser
), CountOfTracksInPLaylistWithType AS (
    SELECT "type id", COUNT("type id") as count_of_tracks_with_type
   FROM CountOfPlaylistWithType
             JOIN "Type of tracks" USING ("track id")
   GROUP BY "type id"
   ORDER BY count_of_tracks_with_type DESC
   LIMIT(1)
SELECT * FROM CountOfTracksWithType
UNION ALL
SELECT * FROM CountOfTracksInPLaylistWithType
ORDER BY count_of_tracks_with_type DESC
```

Результат(тип который больше всего нравится пользователю)

```
I⊞ "type id" † I⊞ count_of_tracks_with_type †

2 2
```

8) Мы выбираем плейлист плейлистов смотрим в его плейлисты, а затем в их треки и понимаем сколько лайков в сумме на этом плейлисте, сколько его суммарная длительность, и сколько людей слушало хотя бы один трек

```
WITH TracksInPLaylistOfPLaylist AS (
   FROM "Tracks in playlists"
            FROM "Playlists in playlists"
), TracksInPLaylistOfPLaylistWithListeners AS (
           TracksInPLaylistOfPLaylist. "track id",
   FROM TracksInPLaylistOfPLaylist JOIN "Favorite tracks of users"
           SUM("audio duration") AS total_duration,
           SUM("count of likes") AS total_count_of_likes,
           COUNT("user id") AS count_of_people
   FROM TracksInPLaylistOfPLaylistWithListeners JOIN "Tracks" ON
```

```
■ total_duration  
■ total_count_of_likes  
■ count_of_people  
1 5203 11695123 25
```

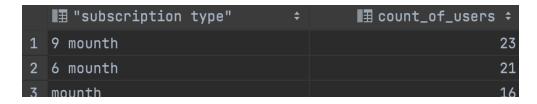
9) Сколько юзеров используют каждую из типов подписок

```
SELECT "subscription type", COUNT("subscription type") as count_of_users FROM
"Users"

GROUP BY "subscription type"

ORDER BY count_of_users DESC
```

Результат



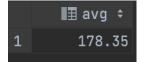
10) Средняя длина подписки

```
SELECT

AVG(length)

FROM "Users" JOIN "Subscriptions"

ON "subscription type" = "Subscriptions".type
```

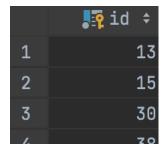


11) Артисты которых никто не слушает

```
SELECT id FROM "Artists"

WHERE id NOT IN (
    SELECT DISTINCT "artist id" FROM "Favorite artists of users"
    )
```

Результат



12) Выведем те треки у которых больше 5 исполнителей

```
"track id",

COUNT("track id") as count_of_artists

FROM "Tracks from artists"

GROUP BY "track id" HAVING COUNT("track id") > 5
```

