

# База Данных для хранилища информации о фильмах (информация о фильме, рейтингах и сборах)

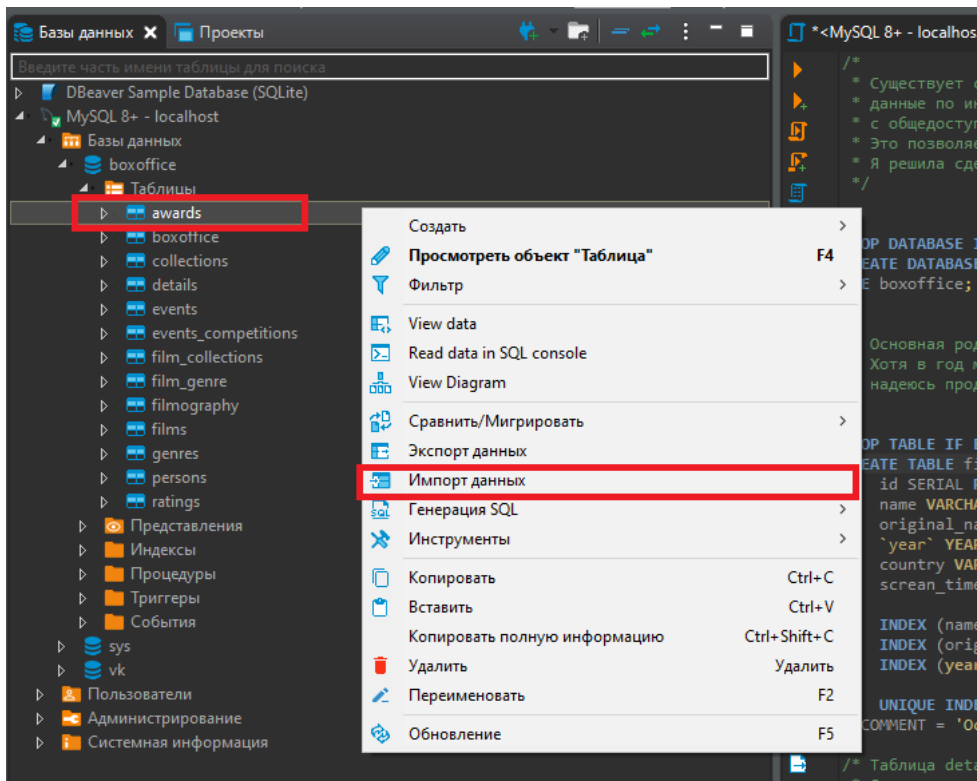
Существует сайт IMDB Pro для специалистов киноиндустрии. Он позволяет совместить данные по индустрии (детальные данные о сборах фильма в кино - ресурс "бывшего" BoxOfficeMojo) с общедоступной информацией о фильмах (общая информация о фильмах, рейтинги открытый сайт IMDB). Это позволяет расширить возможности анализа рынка.

Я решила сделать для себя такую же базу, чтобы использовать её для своих студенческих проектов по анализу рынка кино.

Внимание, в базе прописаны триггеры на BEFORE INSERT. Прежде чем загружать данные из скрипта «INSERT», пожалуйста, загрузите скрипт с триггерами или упрётесь в ошибку.

При подгрузке данных я решила воспользоваться SQL «напрямую», в скрипте **Anna\_By\_Awards\_CSV\_upload** я описываю, как переносила данные из скаченного с Kaggle датасета, в две таблицы своей базы данных по необходимым мне параметрам. CSV грузила через импорт данных внутри DBeaver. CSV файл, который я грузила, прилагаю. (П.С. на подгрузку CVS уходит от 2 до 5 минут, ставьте Ignore на последний столбец notes, если берете файл с Каггл).

Если не хочется грузить CSV, я сделала отдельный скрипт с INSERT INTO с урезанной выгрузкой из датасета (Anna\_By\_FIN\_Project\_Additional part- insert into awards).



Для работы были использованы данные из следующих источников:

<https://www.boxofficemojo.com/>

<https://www.kinopoisk.ru/>

А так же готовые датасеты с Kaggle.com

<https://www.kaggle.com/satkarjain/imdb-movie-19722019>

<https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset>

<https://www.kaggle.com/yjeong5126/box-office-data-20172019>

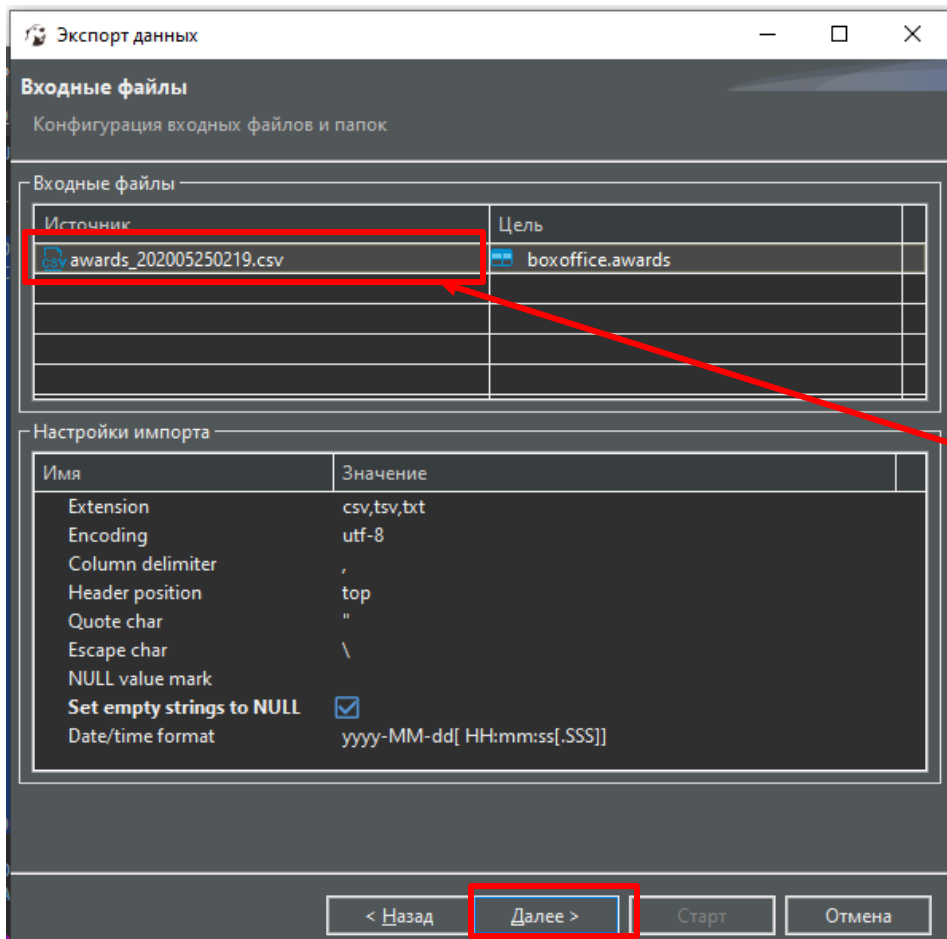
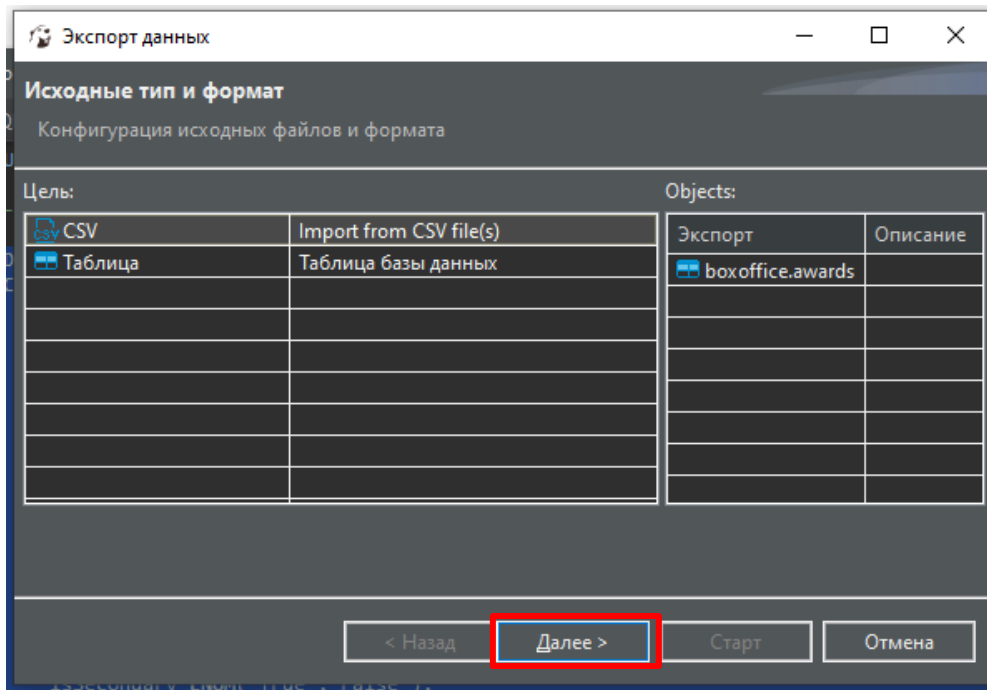
<https://www.kaggle.com/heyueyuan/rottentomatoesmoviesandcriticsdatasets>

И ссылка на Датасет, который я грузила в базу через CSV

<https://www.kaggle.com/iwooloowi/film-awards-imdb>

## Импорт CSV в DBeaver

Таблица и CSV файл подготовлены для стандартных параметров загрузки, нужно только указать путь к файлу. Загрузка займет около 2 минут.



Необходимо нажать на это поле и выбрать путь к файлу.

Экспорт данных

Настройки (Files to База данных, CSV)  
Set data transfer settings

ⓘ Data load settings

Data load

☒ Transfer auto-generated columns

☐ Truncate target table(s) before load

Performance

☒ Открыть новое соединение

☒ Use transactions

Commit after insert of : 10000

General

☒ Open table editor on finish

☒ Show finish message

< Назад

Далее >

Старт

Отмена

Экспорт данных

Просмотр импорта данных  
посмотреть как данные будут импортированы в таблицу

Соответствие колонок:

Таблица:

boxoffice.awards

Колонки:

Цель	Источник	Тип соответствия
ABC eventId	eventId	import
ABC eventName	eventName	import
ABC awardName	awardName	import
year	year	import
123 occurrence	occurrence	import

Предпросмотр:

eventId	eventName	awardName
ev0000837	Amnesty International Film Festival	Media Award
ev0000837	Amnesty International Film Festival	Media Award
ev0003078	Festival International Du Film De Vebron	Jury Prize
ev0003078	Festival International Du Film De Vebron	Jury Prize
ev0003078	Festival International Du Film De Vebron	Jury Prize
ev0003078	Festival International Du Film De Vebron	Jury Prize

< Назад

Далее >

Старт

Отмена

Экспорт данных

Подтвердить

Проверить настройки

Objects

Source Container	Источник	Target Container	Цель
C:\Users\Anna\Do...	awards_202005250...	MySQL 8+ - local...	boxoffice.awards

Source settings

Files settings:

CSV settings:

Extension: csv,tsv,txt

Encoding: utf-8

Column delimiter: ,

Header position: top

Quote char: "

Escape char: \

Set empty strings to NULL: true

Target settings

База данных settings:

Открыть новое соединение: Yes

Use transactions: Yes

Commit after: 10000

Transfer auto-generated columns: Yes

Truncate before load: No

Сохранить задачу

Открыть список задач

< Назад

Далее >

Старт

Отмена