

CinemaSeek

An amazing movies API





Overview

Our motivation and what technology we use





OVERVIEW

Motivation

We love movies, and we know a substantial number of people do so as well. Which is why we want to provide basic to advanced information about movies that are easily accessible and comprehensive to the mass. Be it a simple query, or graphically plotting a statistic.



OVERVIEW

Pain Points

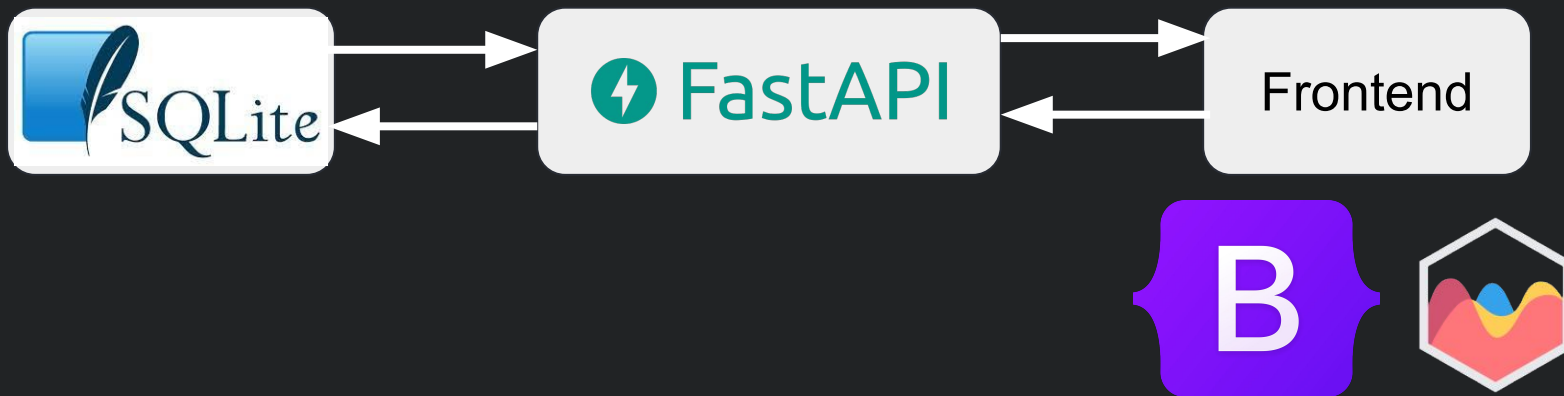
We are certain that our service can combat these pain points:

- Intimidating query services. As in the services are hard to comprehend and use.
- Messy API responses.
- Difficulty in checking a movie's availability on a streaming platform.
- Insufficient data visualization.



OVERVIEW

Overall architecture





OVERVIEW

API Documentation

Swagger documentation available at '/docs'

CinemaSeek 1.0.0 QASt3
[/openapi.json](#)
An API about movie data.

default ^

GET

/api/actor/high-grossing-movies

Get Actor High Grossing Movies

▼

GET

/api/actor/feature-percent

Get Feature Percent

▼

GET

/api/movie/director-as-actor

Get Director As Actor

▼

GET

/api/movie/studio-movie-list

Get Studio Movie List

▼

GET

/api/rating/high-gross-low-rate

Get High Gross Low Rate

▼

GET

/api/rating/yearly-averages

Get Yearly Averages

▼

GET

/api/platform/flix-and-theatre

Get Flix And Theatre

▼

Schemas ^

ActorHighGrossing >

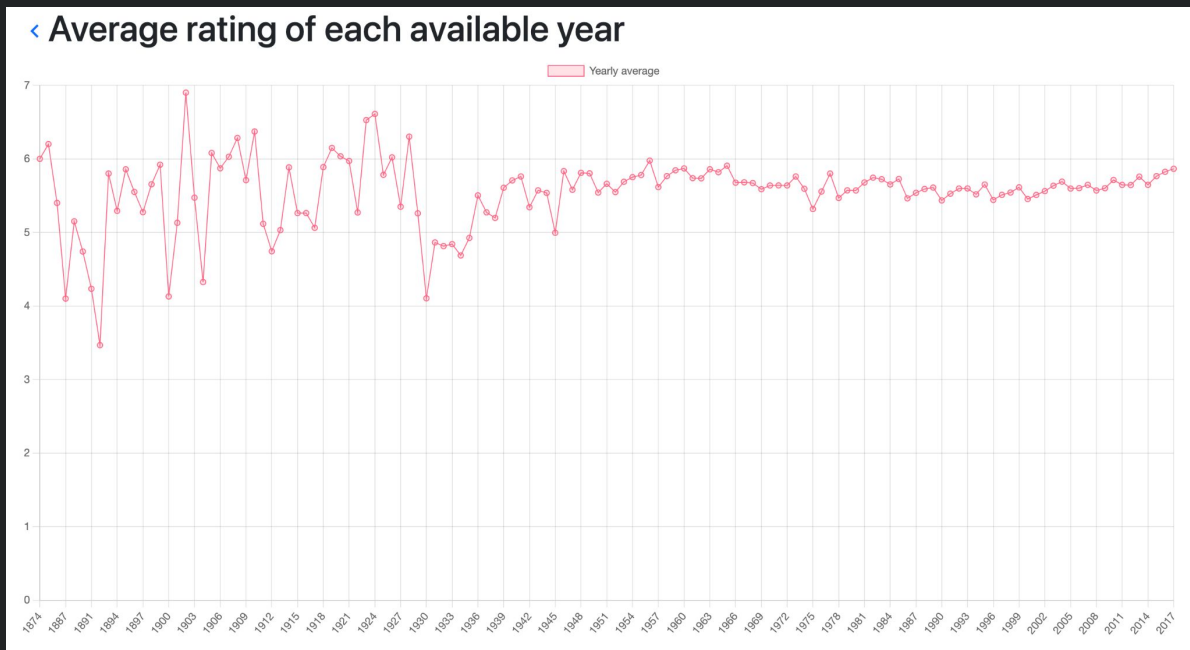
DirectorAsActor >



OVERVIEW

Data Visualization

Use Chart.js's line and bar chart for visualization





Data Sources

Where we get the data from





DATA SOURCE

Primary Datasource

<https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>

The screenshot shows the Kaggle dataset page for 'The Movies Dataset' by Rounak Banik. The page layout includes a left sidebar with navigation links (Home, Competitions, Datasets, Code, Discussions, Learn, More) and a top navigation bar with a search bar, 'Sign In', and 'Register' buttons. The dataset title 'The Movies Dataset' is prominently displayed, along with its description: 'Metadata on over 45,000 movies. 26 million ratings from over 270,000 users.' Below the title, there are tabs for 'Data', 'Code (374)', and 'Discussion (31)'. The 'About Dataset' section provides context and content details. On the right, there are metrics for 'Usability' (8.24), 'License' (CC0: Public Domain), and 'Expected update frequency' (Not specified). A download button indicates the dataset is 239 MB.

Kaggle

Search

Sign In Register

ROUNAK BANIK · UPDATED 5 YEARS AGO

3035 New Notebook Download (239 MB)

The Movies Dataset

Metadata on over 45,000 movies. 26 million ratings from over 270,000 users.

Data Code (374) Discussion (31)

About Dataset

Context

These files contain metadata for all 45,000 movies listed in the Full MovieLens Dataset. The dataset consists of movies released on or before July 2017. Data points include cast, crew, plot keywords, budget, revenue, posters, release dates, languages, production companies, countries, TMDb vote counts and vote averages.

This dataset also has files containing 26 million ratings from 270,000 users for all 45,000 movies. Ratings are on a scale of 1-5 and have been obtained from the official GroupLens website.

Content

This dataset consists of the following files:

movies_metadata.csv: The main Movies Metadata file. Contains information on 45,000 movies featured in the Full MovieLens dataset. Features include posters, backdrops, budget, revenue, release dates, languages, production countries and companies.

Usability

8.24

License

CC0: Public Domain

Expected update frequency

Not specified



DATA SOURCE

Support Datasource

<https://www.kaggle.com/datasets/sanjeetsinghnaik/top-1000-highest-grossing-movies> (Top 1000 Highest Grossing Movies)

The screenshot shows the Kaggle interface for a dataset titled "Top 1000 Highest Grossing Movies" by Sanjeet Singh Naik. The page includes a sidebar with navigation options like Home, Competitions, Datasets, Code, Discussions, Learn, and More. The main content area displays the dataset title, a subtitle "Top 1000 Highest Grossing Hollywood Movies", and a grid of movie posters. Below the title, there are tabs for Data, Code (21), and Discussion (0). The "About Dataset" section provides context, acknowledgements, and an update note. On the right, there are metrics for Usability (10.00), License (CC0: Public Domain), and Expected update frequency (Annually). At the bottom, there are tags for "Movies and TV Shows", "Beginner", "Data Visualization", "Exploratory Data Analysis", and "Intermediate".

Kaggle

Search

Sign In Register

Sanjeet Singh Naik - Updated 10 months ago

191 New Notebook Download (108 kB)

Top 1000 Highest Grossing Movies

Top 1000 Highest Grossing Hollywood Movies

Data Code (21) Discussion (0)

About Dataset

Context
This dataset contains information about the top 1000 highest grossing hollywood films. It is up to date as of 10th January 2022.

Acknowledgements
This data has been scraped from multiple site and has been added together for performing various datat operations. The data has been taken from idmb, rotten tomatoes and many other sites.

UPDATE:
The original data contained information about movies and some additional have been added by me. 'None' refers to datapoints that I could not scrape. If you wish to contribute to this dataset. Do contact me :)

Usability 10.00

License
CC0: Public Domain

Expected update frequency
Annually

Movies and TV Shows Beginner Data Visualization Exploratory Data Analysis Intermediate



DATA SOURCE

Supporting Datasource

<https://www.kaggle.com/datasets/harshitshankhdhar/imdb-dataset-of-top-1000-movies-and-tv-shows> (Unused)

The screenshot shows the Kaggle interface for the 'IMDB Movies Dataset'. On the left is a navigation sidebar with links to Home, Competitions, Datasets, Code, Discussions, Learn, and More. The main content area features a search bar, a user profile for 'HARSHIT SHANKHDHAR' (updated 2 years ago), and a '241' badge. The dataset title 'IMDB Movies Dataset' is prominently displayed, followed by the subtitle 'Top 1000 Movies by IMDB Rating'. Below this, there are tabs for 'Data', 'Code (32)', and 'Discussion (2)'. The 'About Dataset' section includes a 'Context' paragraph, a link to the EDA process, and a 'Content' section listing data fields: Poster_Link, Series_Title, Released_Year, and Certificate. On the right side of the dataset page, there are sections for 'Usability' (10.00), 'License' (CC0: Public Domain), and 'Expected update frequency' (Quarterly). A large IMDB logo is also visible on the right.

Kaggle

Search

Sign In Register

241 New Notebook Download (179 kB)

IMDB Movies Dataset

Top 1000 Movies by IMDB Rating

Data Code (32) Discussion (2)

About Dataset

Context

IMDB Dataset of top 1000 movies and tv shows.
You can find the EDA Process on - <https://www.kaggle.com/harshitshankhdhar/eda-on-imdb-movies-dataset>
Please consider UPVOTE if you found it useful.

Content

Data:-

- Poster_Link - Link of the poster that imdb using
- Series_Title = Name of the movie
- Released_Year - Year at which that movie released
- Certificate - Certificate earned by that movie

Usability 10.00

License
CC0: Public Domain

Expected update frequency
Quarterly



DATA SOURCE

Supporting Datasource

<https://www.kaggle.com/datasets/victorsoeiro/netflix-tv-shows-and-movies> (Netflix)

The screenshot shows the Kaggle dataset page for 'Netflix TV Shows and Movies'. The page layout includes a left sidebar with navigation links (Home, Competitions, Datasets, Code, Discussions, Learn, More) and a top navigation bar with a search bar, 'Sign In', and 'Register' buttons. The dataset page header shows the creator 'VICTOR SOEIRO - UPDATED 4 MONTHS AGO', a view count of '579', and buttons for 'New Notebook', 'Download (2 MB)', and a share icon. The main title 'Netflix TV Shows and Movies' is prominently displayed, followed by a subtitle 'Movies and TV Shows listings on Netflix (July, 2022)'. A thumbnail image of the Netflix logo is shown. Below the title, there are tabs for 'Data', 'Code (88)', and 'Discussion (4)'. The 'About Dataset' section provides a description: 'This data set was created to list all shows available on Netflix streaming, and analyze the data to find interesting facts. This data was acquired in July 2022 containing data available in the United States.' The 'Content' section states: 'This dataset has two files containing the titles (titles.csv) and the cast (credits.csv) for the title.' On the right side, the 'Usability' is rated '10.00', and the file 'credits.csv (3.82 MB)' is listed with a 'CC0: Public Domain' license. The 'Expected update frequency' is 'Annually'.

Netflix TV Shows and Movies
Movies and TV Shows listings on Netflix (July, 2022)

About Dataset

Netflix - TV Shows and Movies

This data set was created to list all shows available on Netflix streaming, and analyze the data to find interesting facts. This data was acquired in July 2022 containing data available in the United States.

Content

This dataset has two files containing the titles (titles.csv) and the cast (credits.csv) for the title.

Usability
10.00

credits.csv (3.82 MB)
CC0: Public Domain

Expected update frequency
Annually



DATA SOURCE

Back-up API

<https://developers.themoviedb.org/> (Unused)

THE
MOVIE
DB

The Movie Database API 3
<https://api.themoviedb.org/3>

OAS

RAML

Support

Select a different version

Filter sections...

GETTING STARTED

Introduction

Authentication

Daily File Exports

Languages

Images

Image Languages

Regions

External IDs

Popularity

Request Rate Limiting

JSON & JSONP

Append To Response

Search & Query For Details

ACCOUNT

AUTHENTICATION

CERTIFICATIONS

CHANGES

COLLECTIONS

COMPANIES

CONFIGURATION

CREDITS

DISCOVER

Getting Started

Introduction

Welcome to version 3 of The Movie Database (TMDB) API. Below you will find a current list of the available methods on our movie, tv, actor and image API. If you need help or support, please head over to our [API support forum](#).

To register for an API key, click the [API link](#) from within your account settings page. You can also view the screenshots below for help:

1. Click on your avatar or initials in the main navigation (screenshot)

2. Click the "Settings" link (screenshot)

3. Click the "API" link in the left sidebar (screenshot)

4. Click "Create" or "click here" on the API page (screenshot)

Please note

that the API registration process is *not optimized* for mobile devices so you should access these pages on a desktop computer and browser.

Before being issued an API key you will have to agree to our terms of use. You can read that [here](#).

A few useful tips...

The [configuration methods](#) are useful to get the static lists of data we use throughout the database. You can find things like the languages, countries, timezones and translations that we use. The configuration method also holds useful image information.

Understanding the basics of our authentication is useful. You can read about this [here](#).

We enforce rate limiting on the API. You can read about that [here](#).



DATA SOURCE

Collection Mechanism

- We gather various data from freely available data sources in CSV format. Mainly from Kaggle as seen in previous slides.
- The collected CSV data is then imported into an SQLite database table with accommodating schema.
- This lets our API utilise SQL queries.
- We first tried querying directly from CSV and realized that it is way too unstructured and leads to spaghetti code.



Database

We use it to not deal with 1GB of CSV manually :)





DATABASE

Database schema

We define the database schema based on the CSV data source, which is imported into our database. Here are some examples.

(Full schema available in 'table.sql')

Schema example

highest_grossing_movies

```
create table highest_grossing_movies
(
    id                INTEGER,
    Title             TEXT,
    "Movie Info"      TEXT,
    Distributor        TEXT,
    "Release Date"    TEXT,
    "Domestic Sales (in $)"  INTEGER,
    "International Sales (in $)"  INTEGER,
    "World Sales (in $)"    REAL,
    Genre             TEXT,
    "Movie Runtime"    TEXT,
    License            TEXT
);
```

Schema example

movies_credit

```
create table movies_credits
(
    cast TEXT,
    crew TEXT,
    id    TEXT
);
```



Demonstration time!

Let's see how it works!

