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

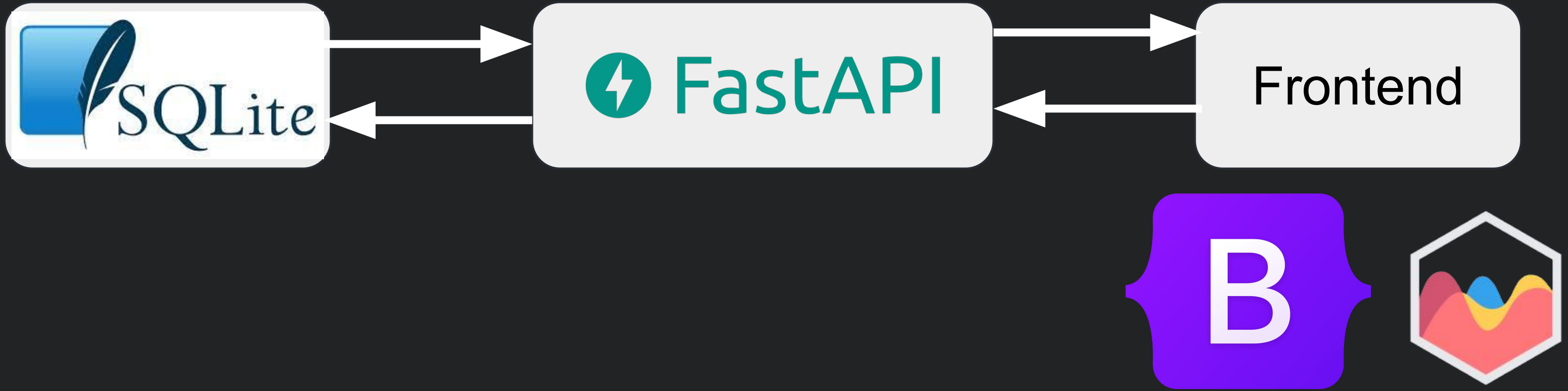
Pain Points

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

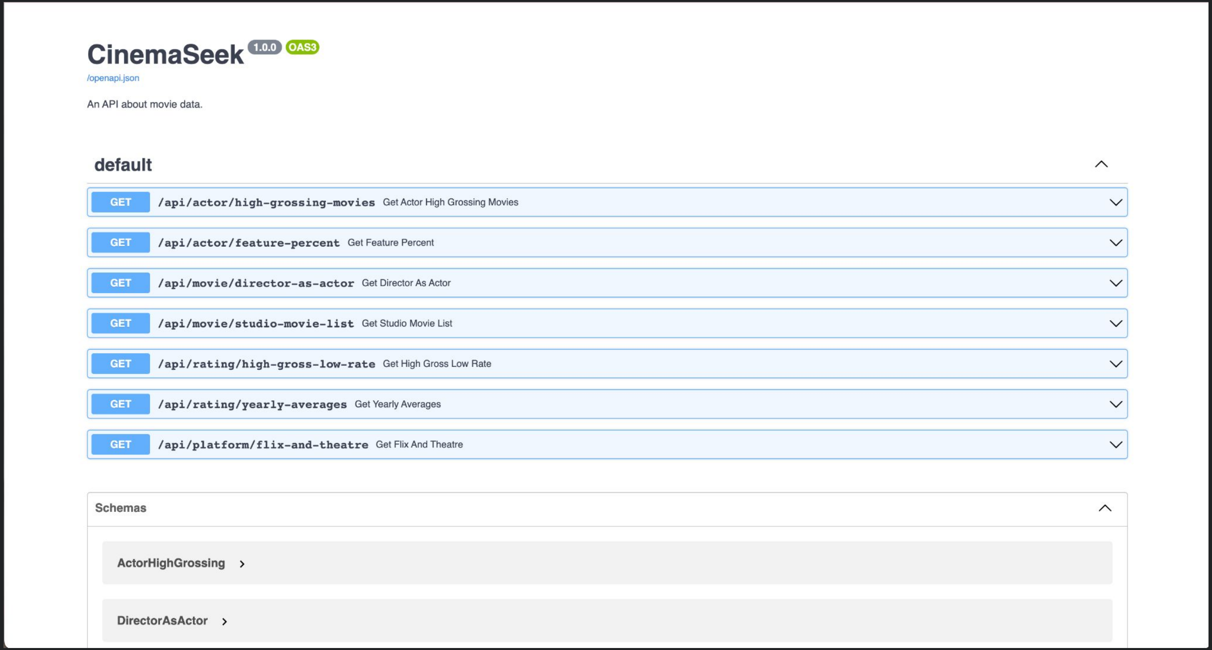
Overall architecture



OVERVIEW

API Documentation

Swagger documentation available at ‘/docs’



CinemaSeek 1.0.0 OAS3

/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

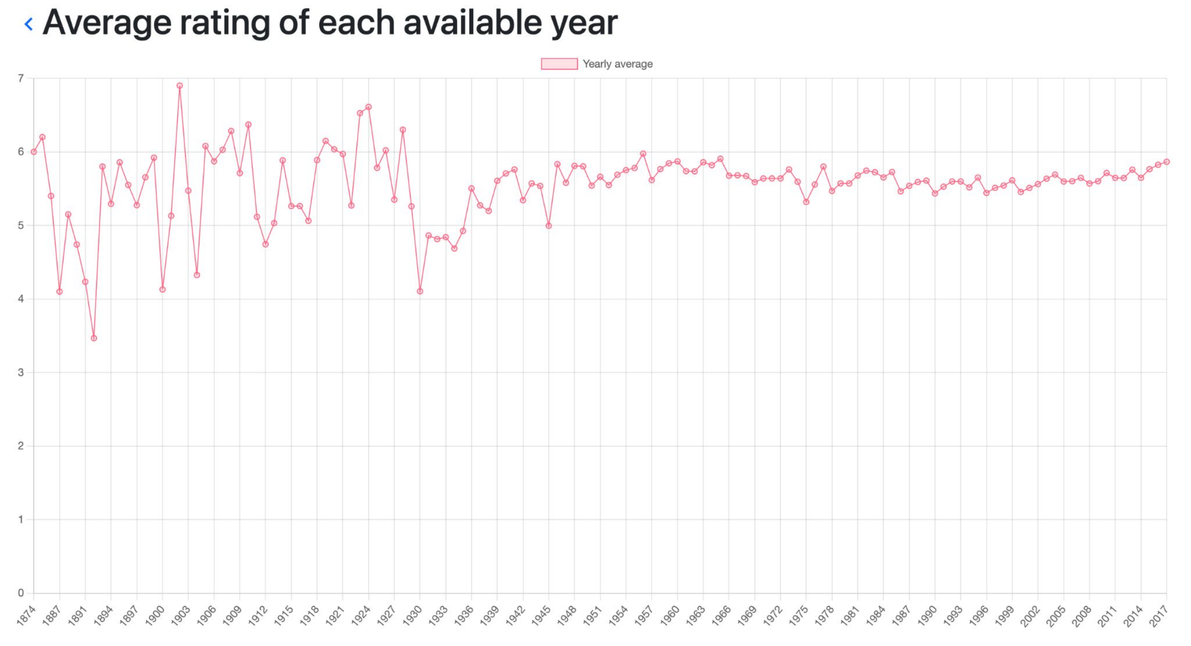
Schemas

- ActorHighGrossing ›
- DirectorAsActor ›

OVERVIEW

Data Visualization

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



‹ Average rating of each available year

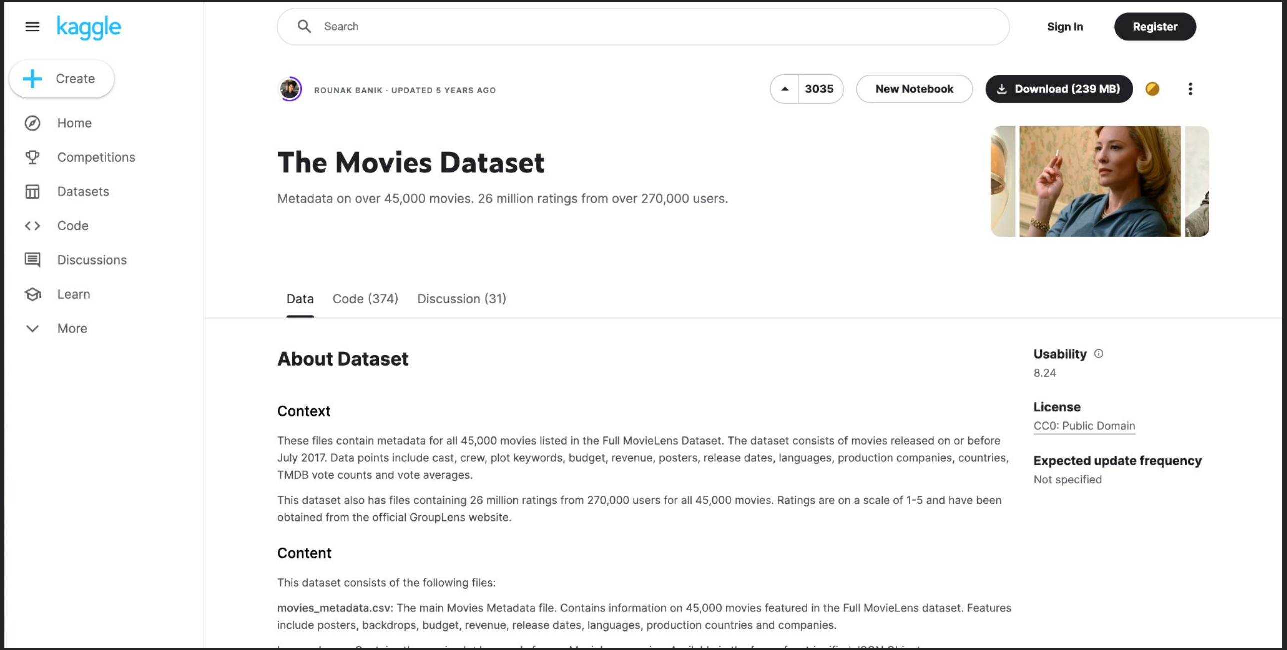
Data Sources

Where we get the data from

DATA SOURCE

Primary Datasource

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



ROUNAK BANIK · UPDATED 5 YEARS AGO

The Movies Dataset

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

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

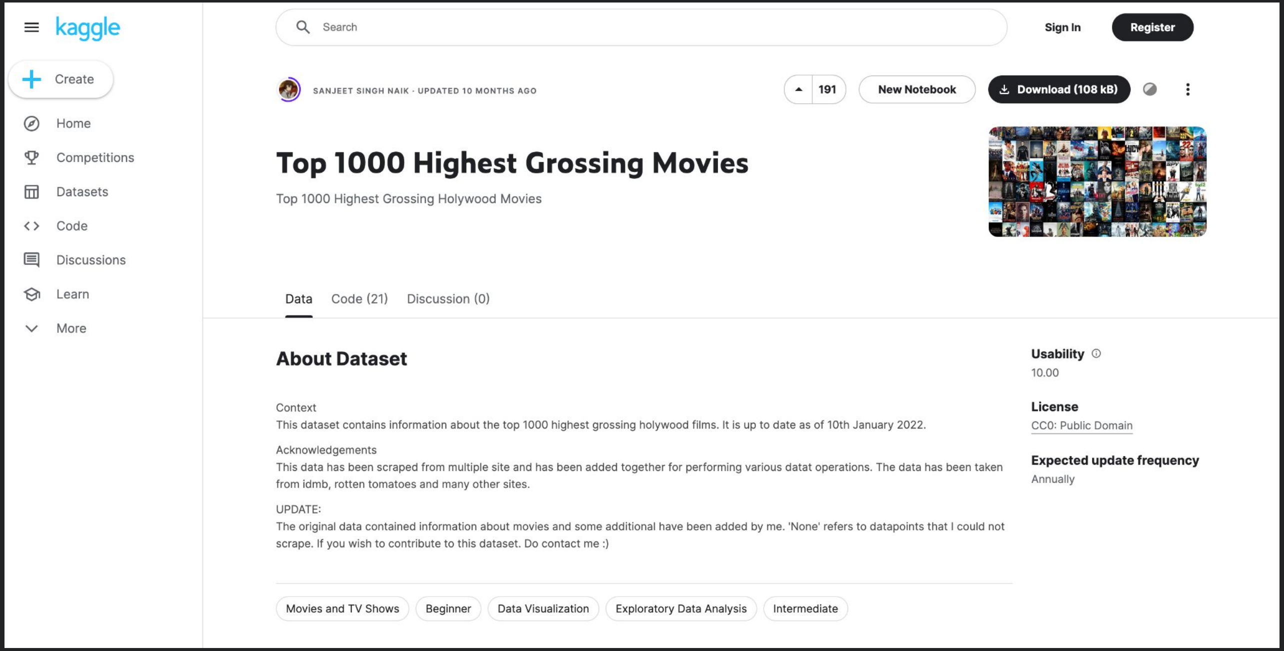
License
CC0: Public Domain

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)



SANJEET SINGH NAIK · UPDATED 10 MONTHS AGO

Top 1000 Highest Grossing Movies

Top 1000 Highest Grossing Holywood Movies

About Dataset

Context
This dataset contains information about the top 1000 highest grossing holywood 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

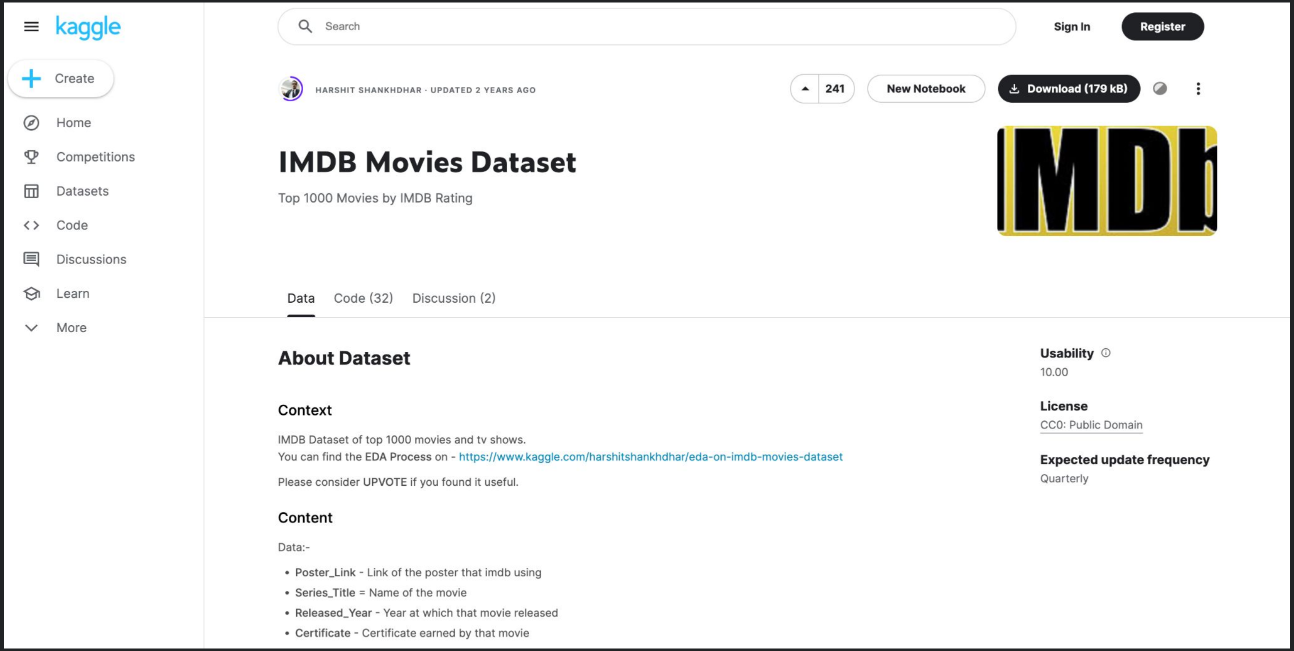
Expected update frequency
Annually

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)



HARSHIT SHANKHDHAR · UPDATED 2 YEARS AGO

IMDB Movies Dataset

Top 1000 Movies by IMDB Rating

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

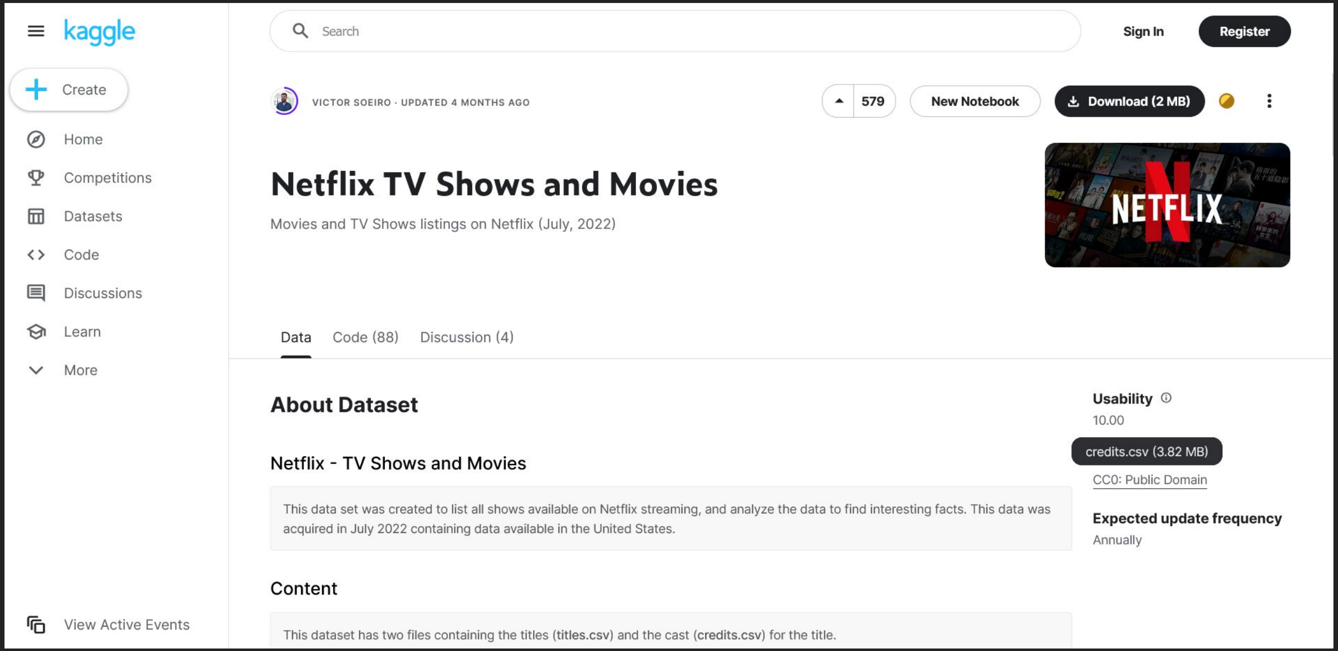
License
CC0: Public Domain

Expected update frequency
Quarterly

DATA SOURCE

Supporting Datasource

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



VICTOR SOEIRO · UPDATED 4 MONTHS AGO

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.

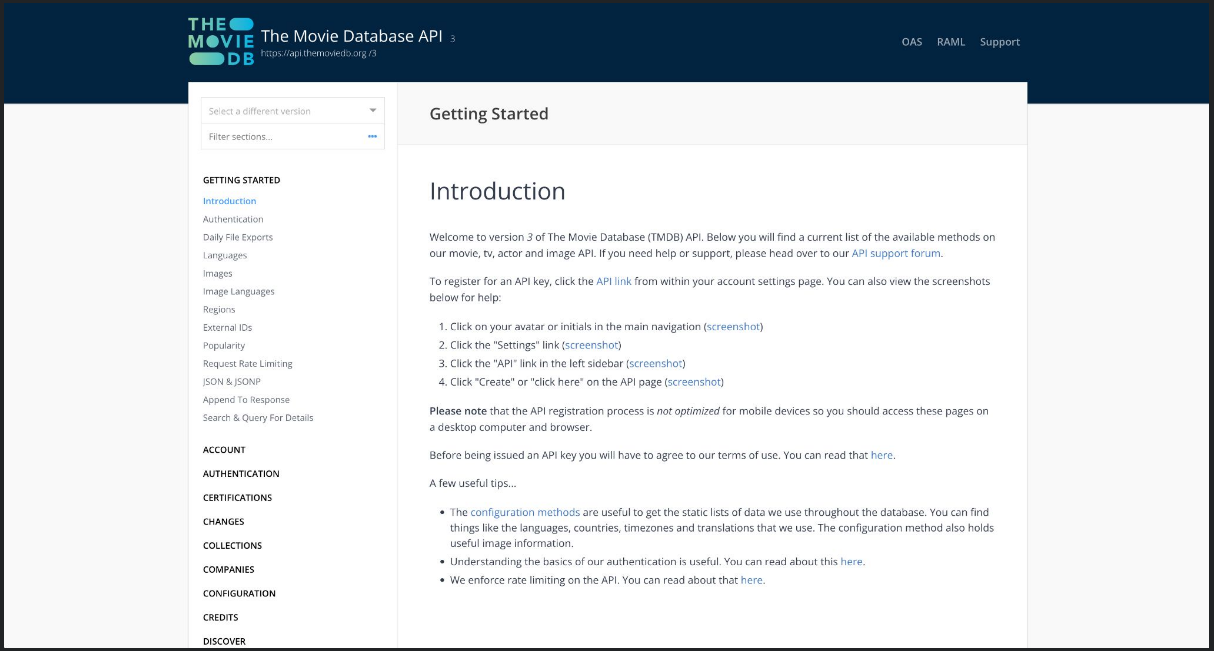
Content

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

DATA SOURCE

Back-up API

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



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 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_grossong _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!