Retomando conceitos iniciais, a ALV é um serviço de streaming desenvolvido por três amigos que abriga diversas produções adquiridas por meio de pagamento de royalties. Como em qualquer serviço de streaming, é possível analisar diversas características dos filmes nele presentes, como seus elencos, gêneros, produtoras, etc. Além disso, está presente um sistema de avaliações, em que um usuário, após assistir a um filme, pode dar uma nota (de 0 a 5 estrelas) que reflete sua opinião em relação à obra.

Observando nosso minimundo e comparando com os serviços de streaming que consumimos, vimos que um aspecto comum entre esses serviços que não está na ALV é a avaliação de fontes externas. Esse detalhe, apesar de simples, é capaz de mudar todo o fluxo de consumo dos conteúdos da plataforma, e, juntamente com a avaliação dos próprios usuários, conseguiríamos fazer análises interessantes.

Muitos usuários, por exemplo, depositariam sua fé no novo parâmetro, já que, por se tratar de uma base de dados maior, espera-se que uma boa nota reflita em um bom filme com maior frequência. Em adição a isso, podemos assumir que essa fonte de dados é munida de avaliações feitas por pessoas já mergulhadas no mundo cinematográfico, fazendo com que filmes classificados como “cult” tenham classificação superior. Nesse contexto, filmes como “O Farol” atrairiam a atenção de novos usuários e outros como “Velozes e Furiosos 10” veriam possivelmente um contraste entre suas notas.

Do ponto de vista da produção do trabalho, essa nova fonte de dados resultaria em análises bastante curiosas, principalmente em relação a comparações com as já existentes notas obtidas dos usuários. Podemos, por exemplo, comparar os filmes com as melhores avaliações, os gêneros mais amados, as produtoras mais renomadas, ressaltando as diferenças entre os dois dados observados.

Desse modo, visando uma boa fonte para avaliações, optamos pelo IMDb, uma das bases de dados relacionada a filmes mais conceituadas e respeitadas atualmente. Ela contém diversas informações como nota segundo usuários, popularidade da produção, elenco, diretor, etc. acerca de filmes, séries, novelas, jogos, entre outros tipos de produções audiovisuais.

Os dados foram extraídos dos [IMDb Data Files](https://datasets.imdbws.com/) que contêm milhões de dados disponíveis para download de forma gratuita. Devido aos dados estarem organizados em arquivos muito pesados, foi utilizado o DuckDB, ferramenta de código aberto para gerenciamento de dados relacionais desenhada para ser eficiente em operações locais, para que os dados pudessem ser filtrados de forma a não ultrapassar os limites computacionais de máquinas locais.