

I'll make you a movie analysis you can't refuse!

M20190904 – Ricardo Francisco M20190792 – Miguel Pereira M20190787 – João Malcata **Locations and Companies**

Financial Analysis

Cinema Evolution

Movie Details

Insights



Overview

Objectives (

- ✓ Develop insights from a large dataset in Cloudera VM
- ✓ Build and optimize powerful queries that run on large CSV files
- ✓ Merge the basis MovieLens database with TMDB's API to empower our insights
- ✓ Draw several conclusions about the Cinema Industry Environment



- 25M Ratings
- > 1M Tags
- ▶ 62K Movies
- ► 162K Users
- Ratings from 1995 to 2019
- Merged with TMDB dataset

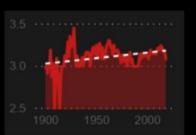
Workflow





ODBC



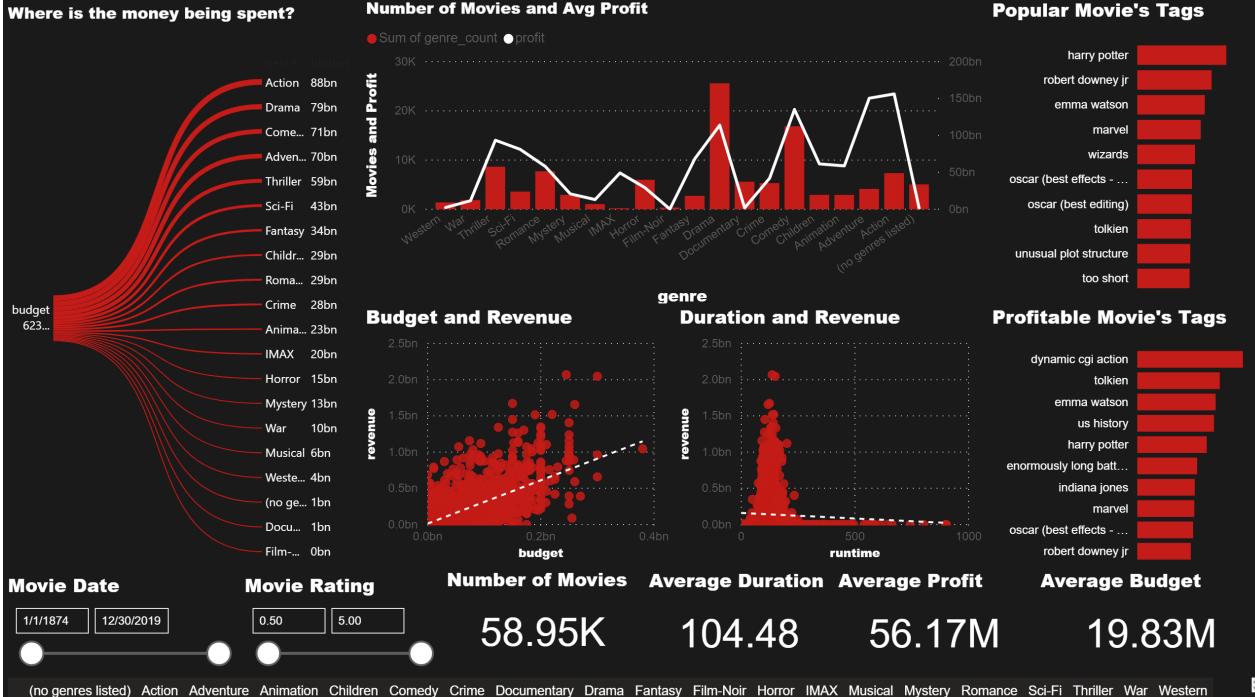


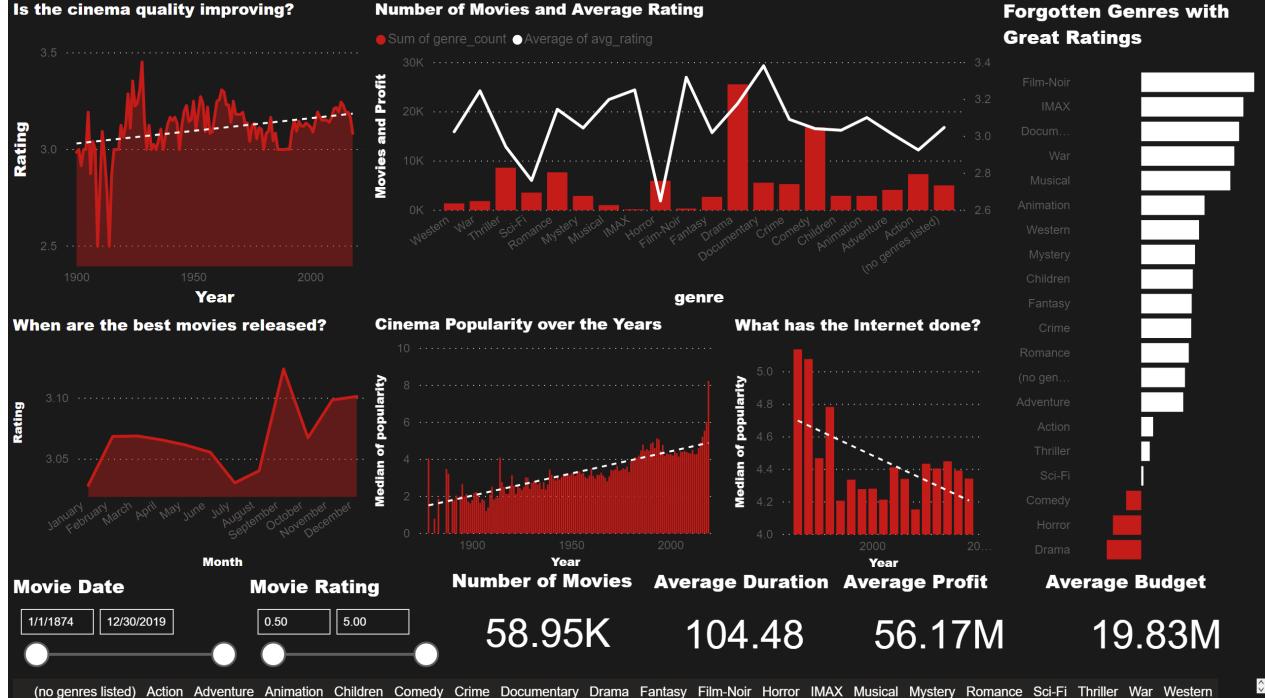




Star Schema tmdb_companies_exp... : movieid production company Collapse ^ dim_movies tmdb_countries_expl... movieid fact_movies movieid poster url m production country (M) title ∑ avg rating OF avg rating n ∑ budget Collapse ^ ∑ count genre ∑ count produc companies ∑ count produc countries dim_date 5 count rating movie date movieid genre_vs_rating Σ month ∑ avg rating movie date B avg rating n Collapse ^ fact_tags dim_genre movieid movieid (B) genre ∑ tag relevance movieid ⊞ tagid Collapse ^ dim_genre_groups dim_tags genre genre count ☐ tagid Collapse ^ Collapse ^







Top Movies



The Shawshank Redemption



The Godfather



The Usual Suspects



The Godfather: Part II



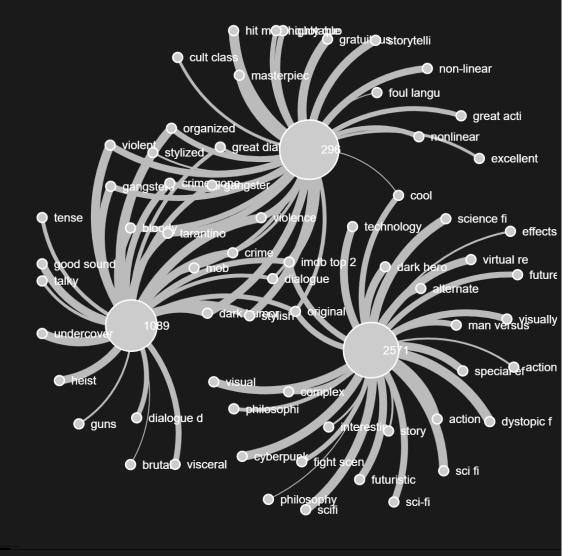
Seven Samurai (Shichinin no samurai)



Schindler's List

title

What people Say?



Movie Date

6/1/1940 12/31/2021

Movie Rating

0.50 5.00

Search for Movie

All

Relevance Score

Average Popularity



17.86

Rating explained by Genre and Tags **Rating influencers** Key influencers Top segments Genre Tag Film-Noirthe average of Median of When... ← Median of avg_rating is more likely to depressing avg_rating increases by increase when budget is 165000000 -217000000 than otherwise (on average). budget is 165000000 -0.31 217000000 Documentary suicide attempt genre is Documentary Film-Noir vienna Average of avg_rating **IMAX** sexy genre is War War talky genre is Drama 165000000 25000000 300000. more than 30000000 Musical fbi runtime goes up 280.23 rio de janeiro V Only show values that are influencers

