

Read it or Watch it?

Lauren Hirs

Samantha Jaeger

Lauren Liou

Dylan Peters

Claudia Scott

Kelly Tallau

Motivation

Given the amount of media currently available, it is common for readers/movie viewers to ask: "Should I read the book, or should I watch the movie?". Today, a viewer must consult multiple sources to answer this question, whether that be Goodreads, IMDB, RottenTomatoes, etc. This process can be time consuming and biased.

We aim to synthesize this data to aid **potential audiences** deciding whether to read it / watch it and **movie producers** who are trying to make the most successful book adaptation possible.

Approaches

- 1) Data collection + cleaning
- 2) Latent Dirichlet Allocation (LDA) text analysis for topics/sentiments from book/movie reviews
- 3) User friendly + informational visualization

Approach success evaluated by user testing feedback

Innovations

- LDA – explore discussions around **book-movie topic similarity**, split by positive + negative reviews
- Synthesize significant information on book/movie pairs obtained in an **interactive visualization** tool for ease of use
- **"Read It or Watch It" gauge** – allows users quickly identify suggestions from us based upon the normalized book and movie rating

Experiments & Results

Text Analysis

1. Pre-processing collected book/movie reviews:

- remove foreign languages, symbols and any punctuations
- combined homonyms
- removed book/movie titles + stopword "noise"
- standardize book/movie rating

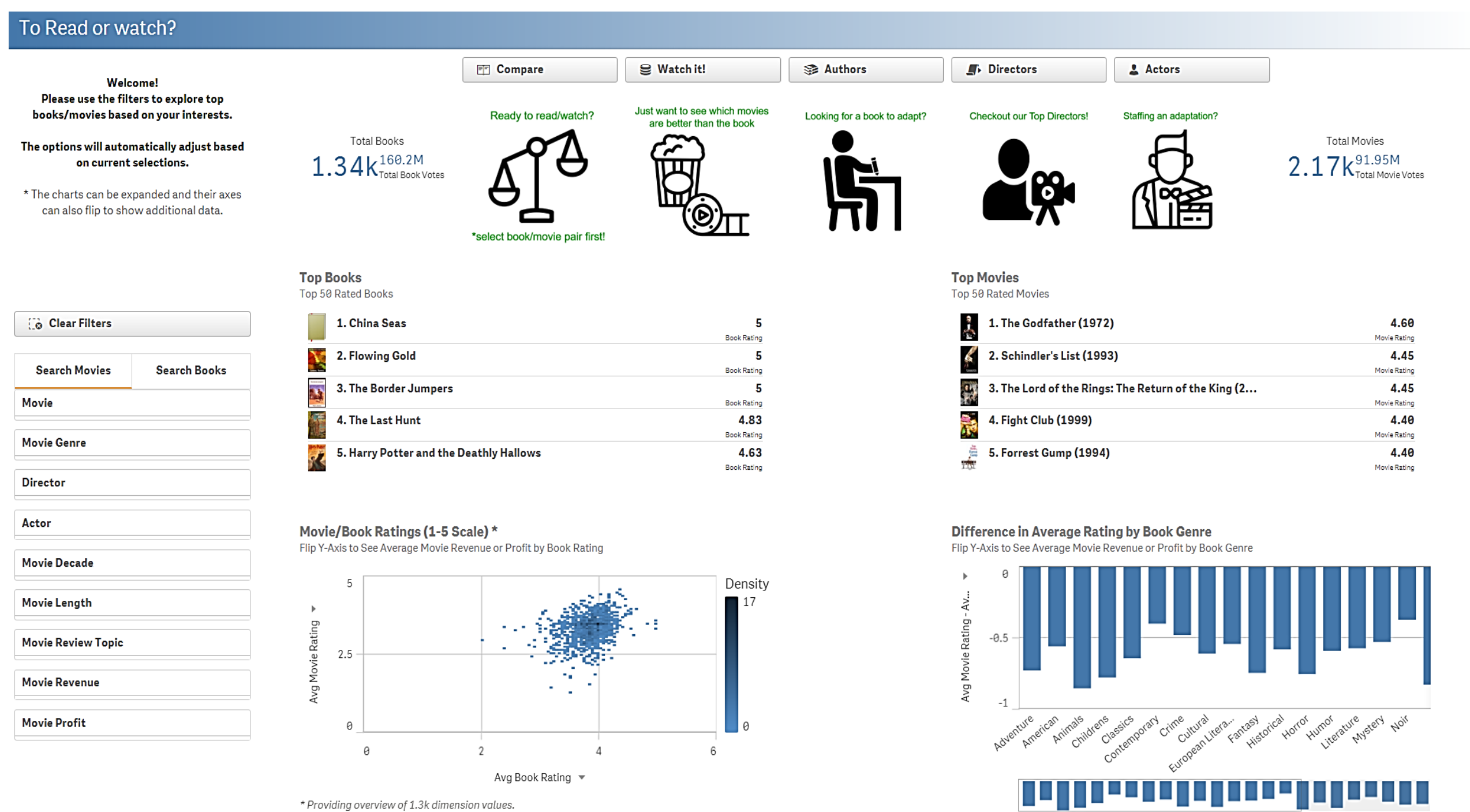
2. LDA model building:

- ✓ coherence found optimal # topics = 30
- ✓ identified a topic in movie reviews which mentioned "book", "read", "good" and vice versa for book reviews -> similarity topics
- ✓ similarity topics used to examine relationships between book/movie reviews vs. ratings

Visualization

Landing page:

- Top ranked books and movies that update with filtering by user interests
- Interactive plots: book/movie rating scatterplot & genre bar chart by rating difference
- Links out to feature pages



Additional interactive pages:

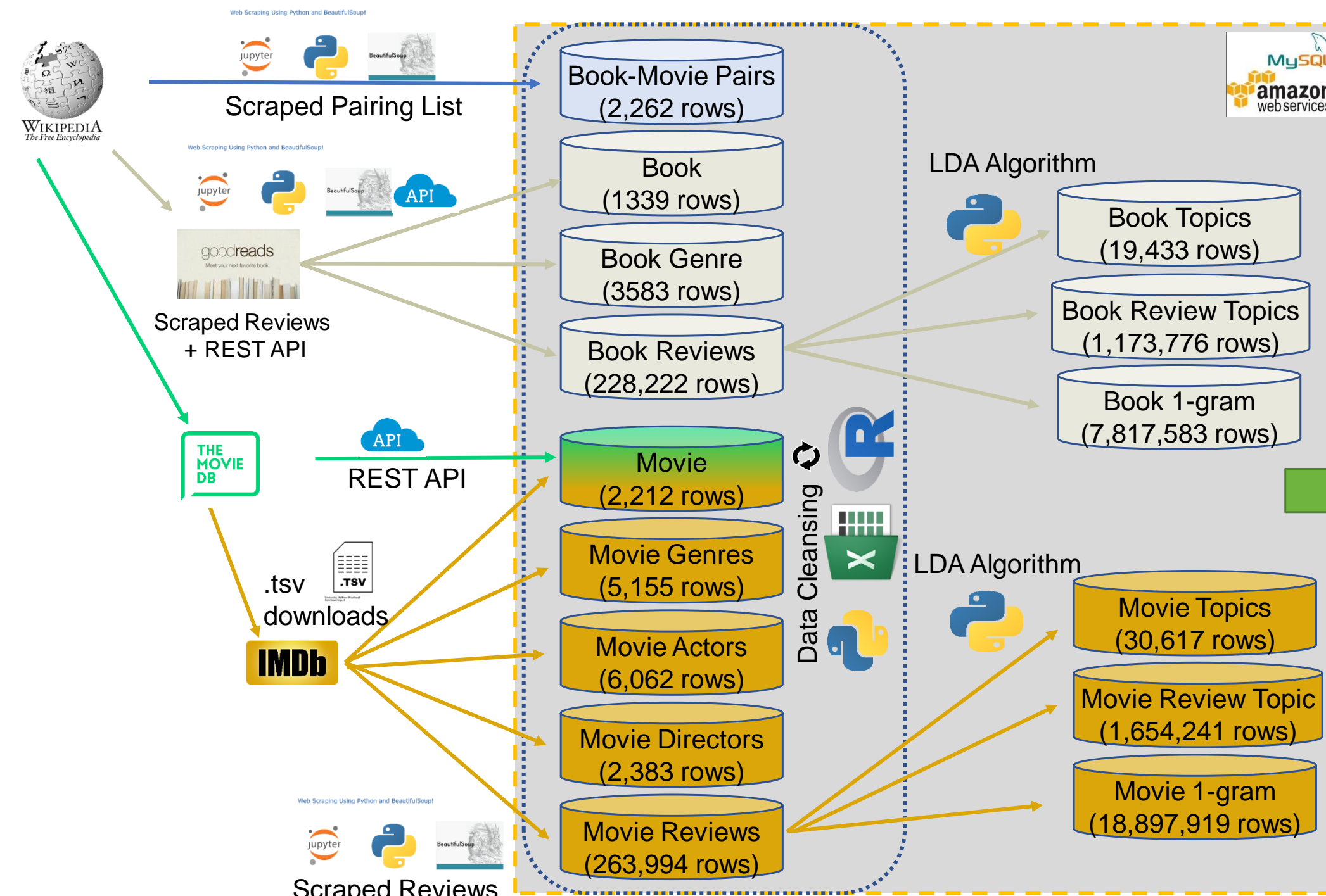
- "Watch it!" page – only movies rated higher than books, bar chart with % better & movie length, treemap of directors and production time period
- "Top Book Authors" page – authors ranked by movie ratings & profits, book genre treemap, author works -> movies comparison dendrogram
- "Top Directors" page – directors ranked by movie ratings & profits, bar chart of those whose movie ratings > book, dendrograms to compare filmographies
- "Top Actors" page – Actor listing by movie ratings & profits, bar chart of those whose movie ratings > book, and dendrograms to compare filmographies

Conclusion

This project exploited **LDA topic analysis** and **interactive visualization** to provide more accurate **book-movie comparisons** to our users. We leveraged Qlik for visualization and python scripts for LDA. The user testing survey revealed that most users can find 80% of the information in the applications within 5 min., 90% of user need at maximum 3 min. to locate and compare a book-movie pair. 85% users found the book-movie comparison information helpful.

Overall, the user survey results align with our project success metric and we have accomplished the objectives we set for this project in the given timeframe. Moreover, we are happy and proud to have developed an appealing application that users enjoy using to quickly learn about book-movie pairs, including features that cannot be found anywhere else.

Data Architecture



Static source data was collected in Feb 2019.

Combined tables used in Qlik visuals have **over 30 million rows**.



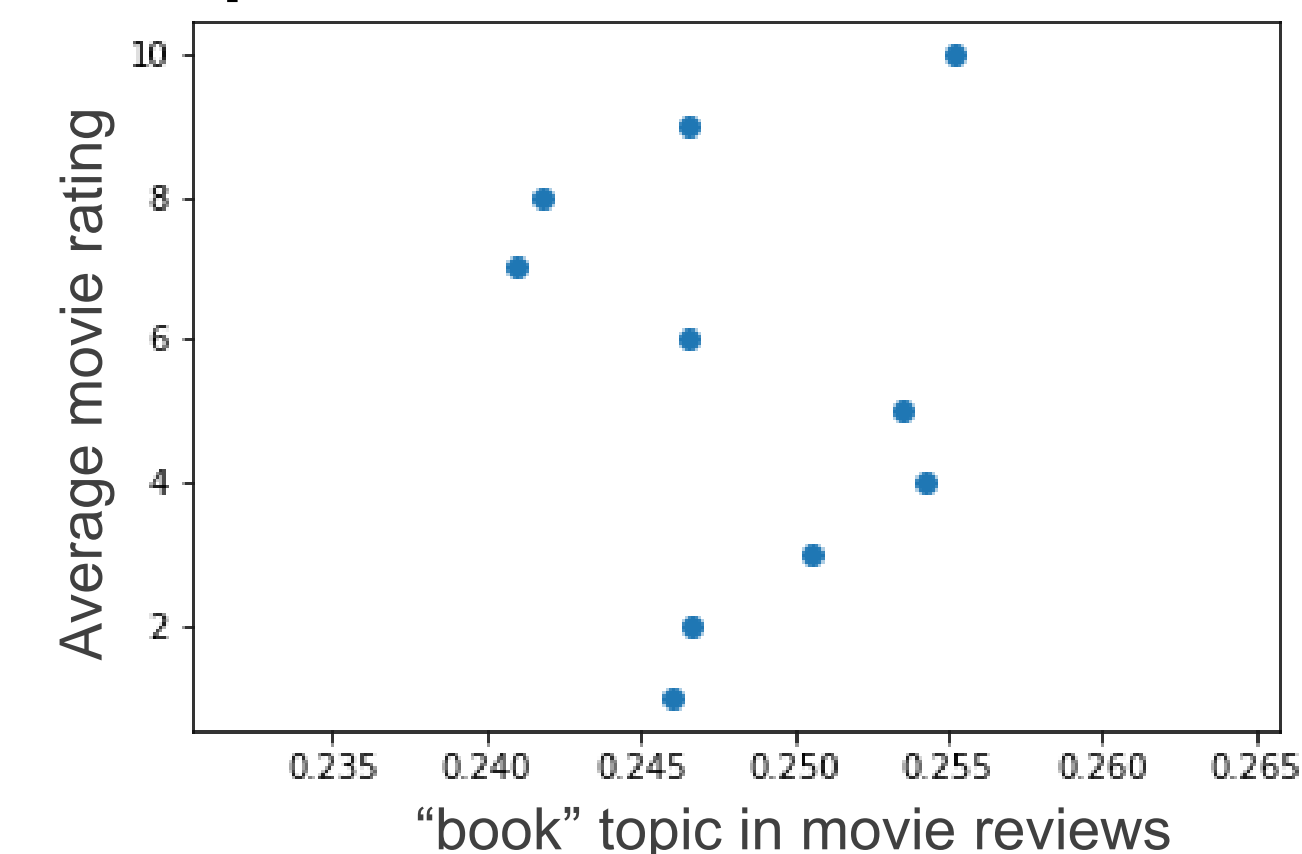
Similarity Topic (top 6) for Book Reviews

Word	Probability (%)
movie	0.1209
read	0.0327
watch	0.0289
version	0.0279
story	0.0191
based	0.0167

Similarity Topic (top 6) for Movie Reviews

Word	Probability (%)
book	0.0195
read	0.0184
good	0.0182
watch	0.0168
time	0.0136
story	0.0135

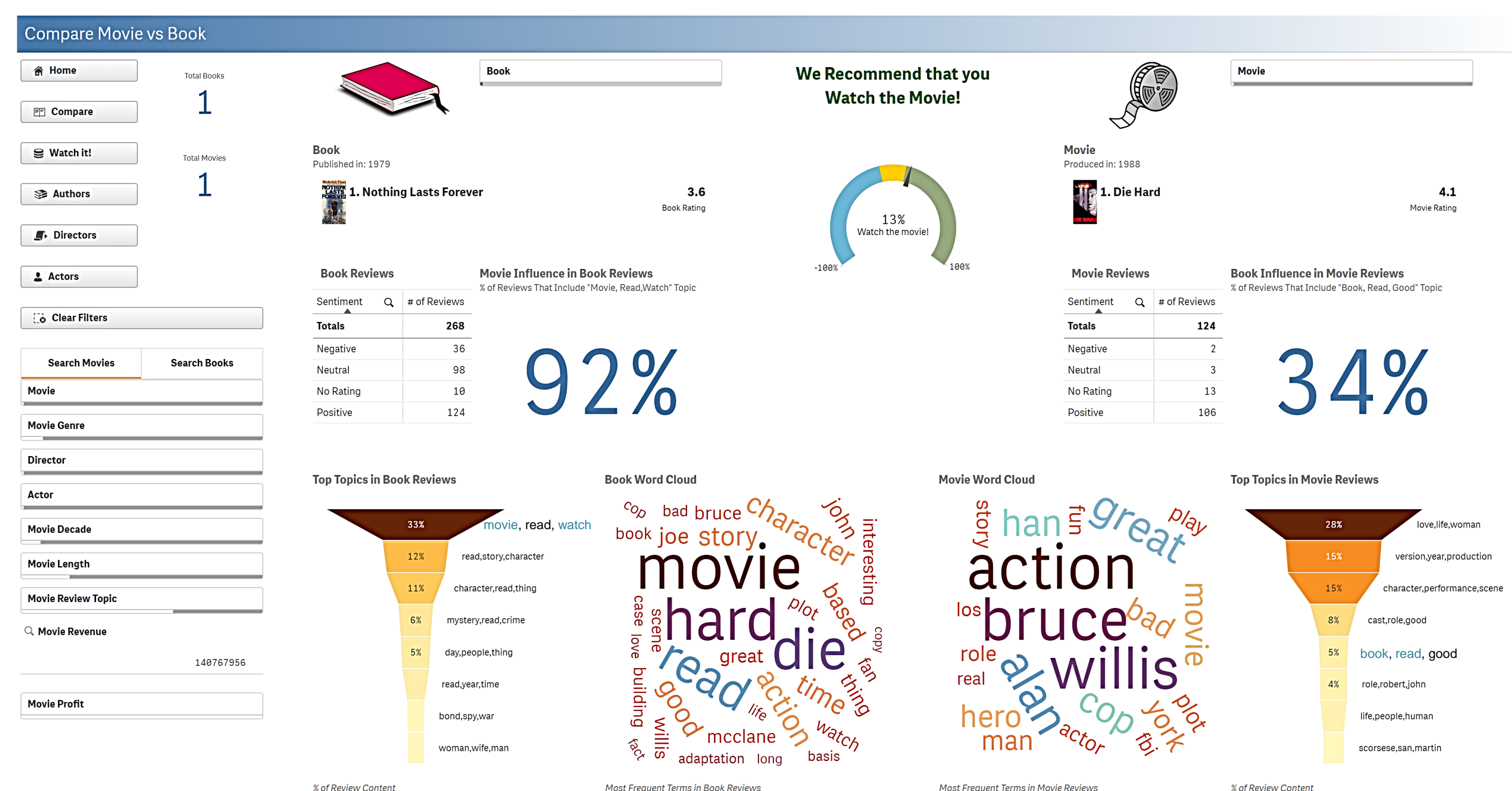
Average Movie Rating vs. "Book" Topic Discussion in Movie Reviews



"Compare" page:

- Evaluate individual book-movie pair
- Analyze similarities and differences in topics discussed in book + movie reviews, with option to filter by review rating sentiment
- Read it / Watch it Recommendation gauge calculation:

$$\frac{\left(\frac{m_{AR}}{2} - b_{AR}\right)}{\min\left\{\frac{m_{AR}}{2}, b_{AR}\right\}} \times 100\% ; m_{AR} : \text{average movie rating}; b_{AR} : \text{average book rating}$$



Results – User Testing

- Conducted 6 rounds of visualization design iteration
- 20 users from different age groups, backgrounds and interests
- 80% of users highlighted an application feature they found interesting, 50% of them mentioned features from our unique comparison page

Task performed	Result	Project goal
Locate a specified book-movie pair + compare them using the application	90% of users took ≤ 3 minutes	👍
Is "Compare" page info helpful in deciding to read it / watch it	85% of users think info is helpful	👍
Take 5 minutes to navigate and explore the application	≥ 65% users can identify 80% of main features* in application	👍

* main features tested (in order by % found) – rating + reviews for book-movie pairs, word clouds, genre filter in "Watch it" / "Top Book Authors" page, analysis of similarity (topic comparison)