



Cut choice overload

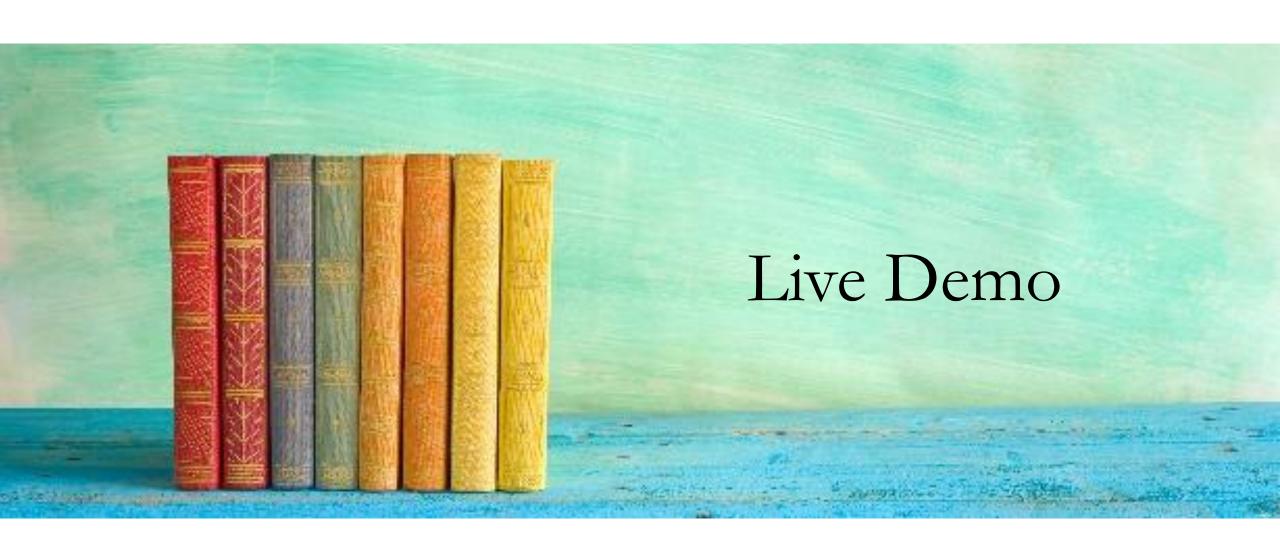
Mission



Save time



Save money



Data

Data from https://github.com/zygmuntz/goodbooks-10k

Originally from goodreads.com

This dataset contains six million ratings for ten thousand most popular (with most ratings) books

Humans tagged each book when they rated them, which is how I created the genres

Model

- Used a basic KNN model
- Predictions were computed offline
 - Inputs were limited to three choices from the top twenty books for each genre
- Used precision as the success metric
 - Goal is to show only books that the user will actually enjoy
 - Cost of showing false positives is high

Accuracy Metrics	
Genre	Precision
History	93%
Chick-Lit	97%

Insights

Some genres had too few books for meaningful recommendations

The Humor genre predicted Calvin and Hobbes to almost everyone (although, this is a great book, so maybe not a problem)



