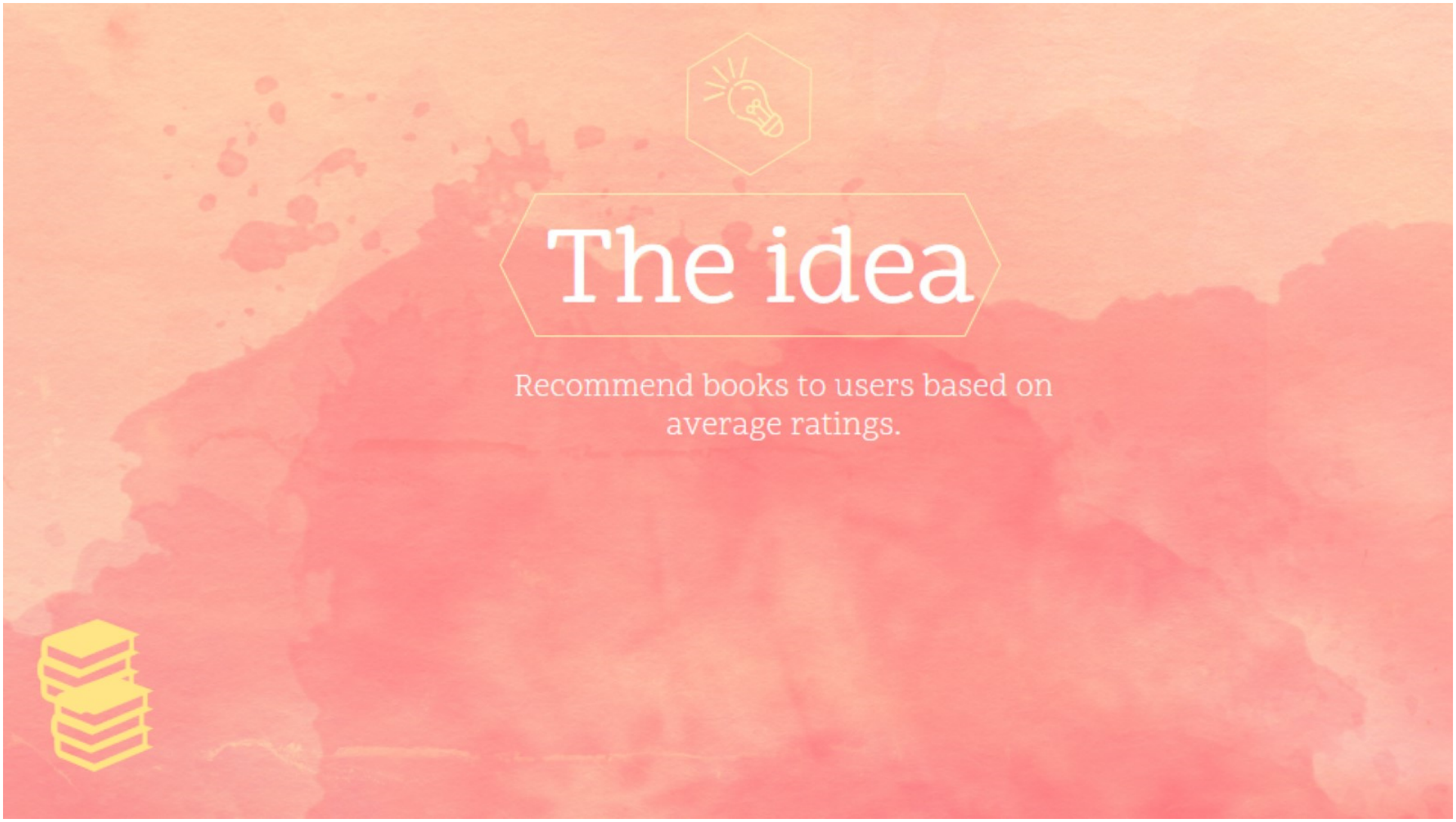
# How to recommend books with Data Analytics

Data Analytics Bootcamp
Ironhack, Berlin
Final Project, March 2021
Sybille Kiziltan

"'Classic' – a book which people praise and don't read."

—Mark Twain


The idea
Recommend books to users based on average ratings.

# Workflow

Data Gathering

↓

Data Exploration

↓

Data Cleaning

↓

Clustering the Data

↓

Fitting the Model

# The Data

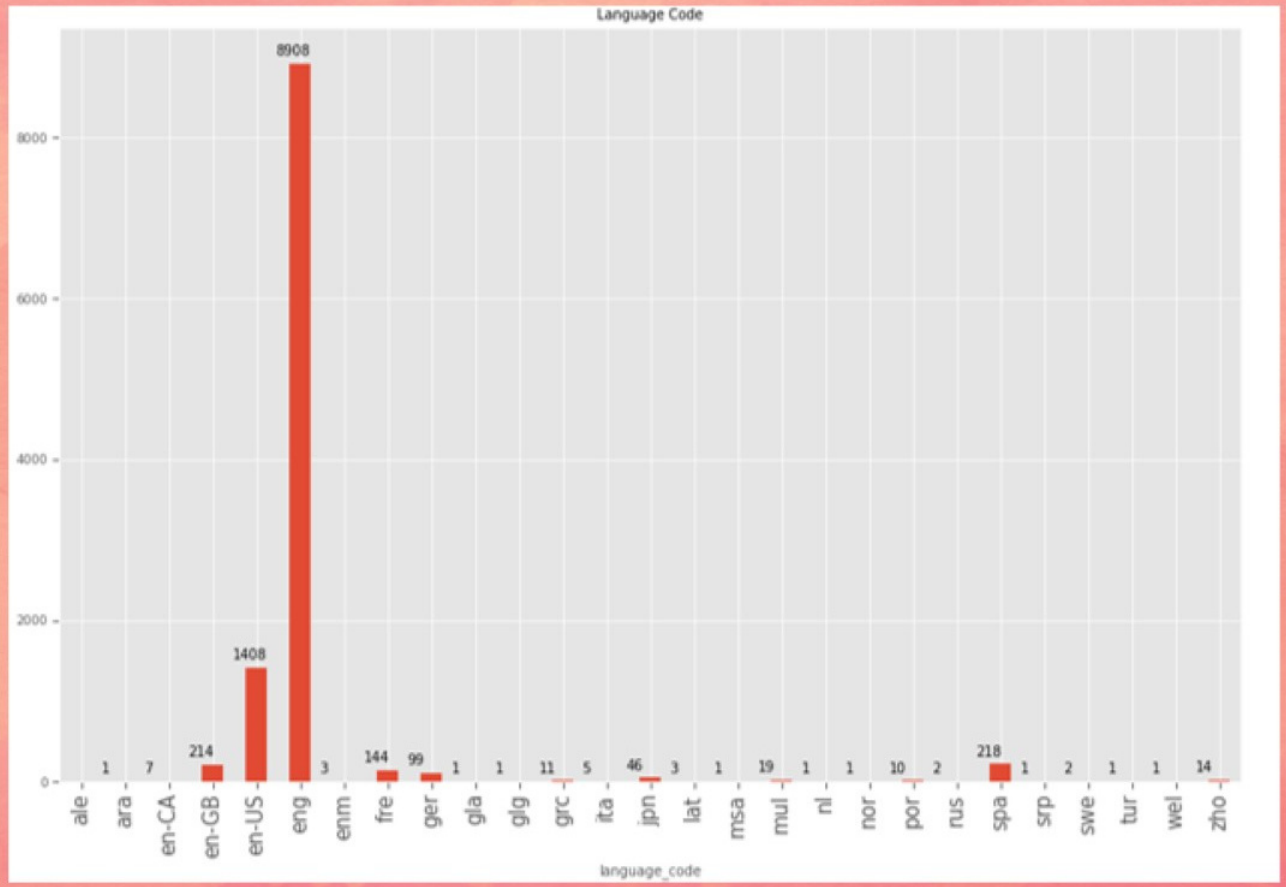
```
df.head()
```

| bookID | bookID | title | authors | average_rating | isbn | isbn13 | language_code | total_pages | ratings_count | text_reviews_count | publication_date | publisher |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Harry Potter and the Half-Blood Prince (Harry ... | J.K. Rowling | 4.57 | 0439785960 | 9780439785969 | eng | 652 | 2095690 | 27591 | 9/16/2006 | Scholastic Inc. |
| 2 | 2 | Harry Potter and the Order of the Phoenix (Har... | J.K. Rowling | 4.49 | 0439358078 | 9780439358071 | eng | 870 | 2153167 | 29221 | 9/1/2004 | Scholastic Inc. |
| 4 | 4 | Harry Potter and the Chamber of Secrets (Harry... | J.K. Rowling | 4.42 | 0439554896 | 9780439554893 | eng | 352 | 6333 | 244 | 11/1/2003 | Scholastic |
| 5 | 5 | Harry Potter and the Prisoner of Azkaban (Harr... | J.K. Rowling | 4.56 | 043965548X | 9780439655484 | eng | 435 | 2339585 | 36325 | 5/1/2004 | Scholastic Inc. |
| 8 | 8 | Harry Potter Boxed Set Books 1-5 (Harry Potte... | J.K. Rowling | 4.78 | 0439682584 | 9780439682589 | eng | 2690 | 41428 | 164 | 9/13/2004 | Scholastic |

- 11.123 books
- Features:
  - title, isbn, author, rating, average rating, language, number of pages, publication data, publisher

# Exploratory Data Analysis

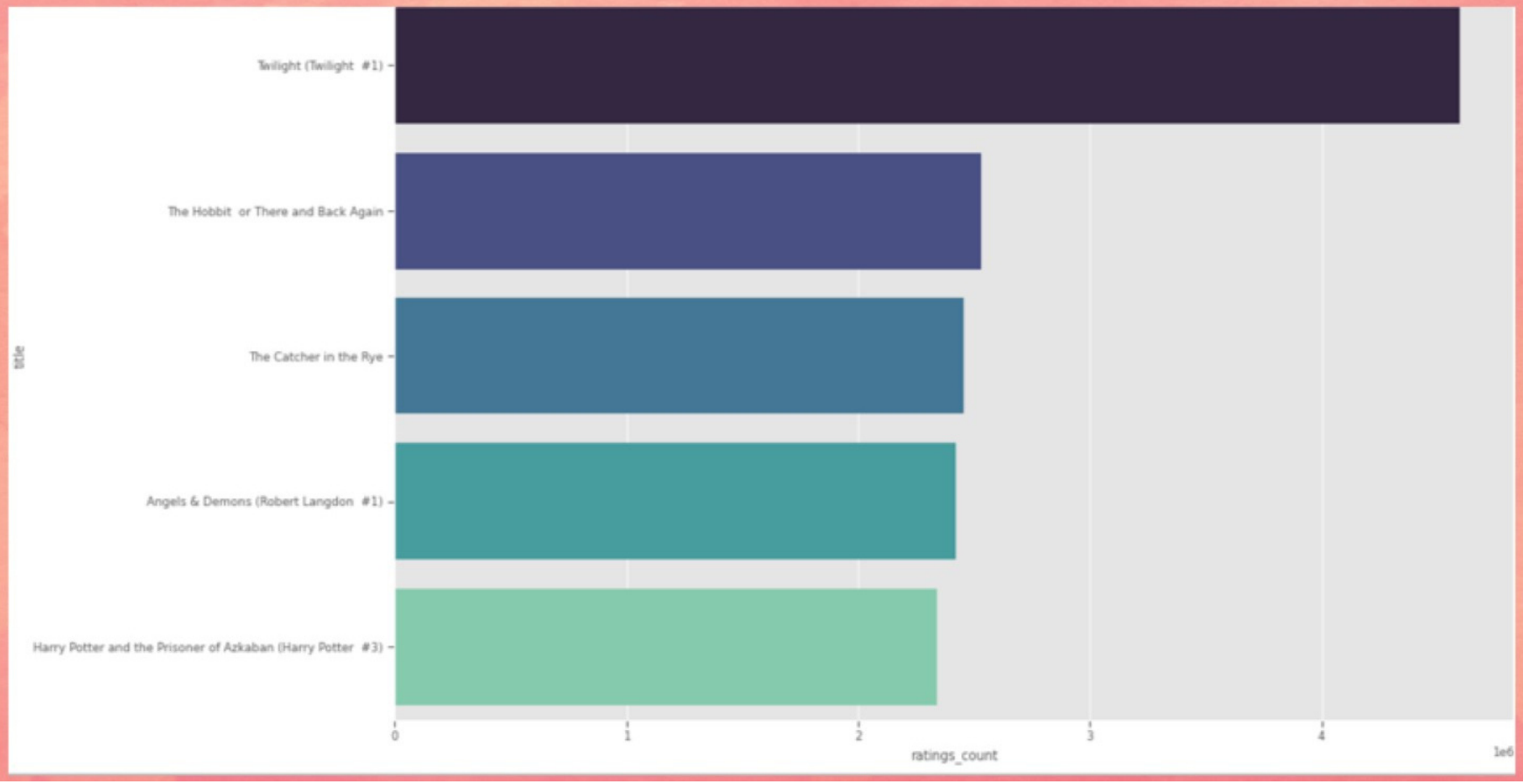
- Most of the books in the dataset written in english




Recommendation system makes mostly sense for english books

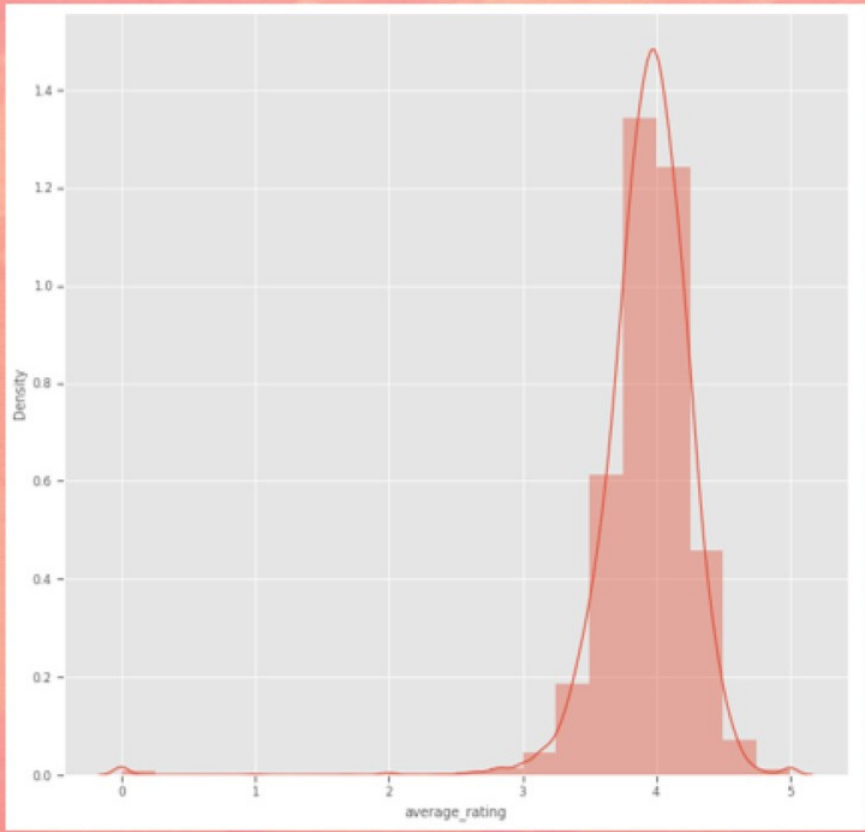
# Exploratory Data Analysis

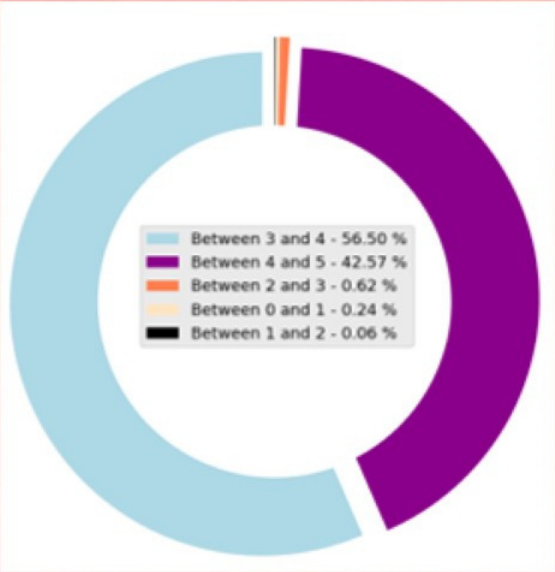
- most rated book: „Twilight“, Stephenie Meyer



# Exploratory Data Analysis

- How are the ratings distributed?




- Most of the ratings scatter around 4
- 56.5% are between 3 and 4
- 42.57% are between 4 and 5

# The Model

## KMeans clustering

Finding groups in the data, to find relationships between rating counts and average rating

Clusters

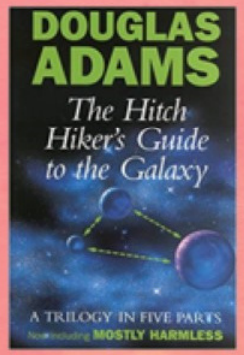
7

## Recommendation Engine

The recommendation system works with the algortihm of K Nearest Neighbors (KNN)

# How does it work?

Type in the book you like and get recommendations of five other books you may like.




```
[318]: book_rec("The Hitchhiker's Guide to the Galaxy (Hitchhiker's Guide  #1)")

       If you like:

       The Hitchhiker's Guide to the Galaxy (Hitchhiker's Guide  #1)

       we recommend you also the following five books:

       Cat Breaking Free (Joe Grey  #11)
       A Life in Letters
       Nabokov's Pale Fire: The Magic of Artistic Discovery
       Selected Speeches and Writings
       Return of the Straight Dope
       The Beleaguered City: The Vicksburg Campaign
```

# Thanks!

Special thanks to Jose for sending me the article which
made me do this project, to Raul for his help during this week and to Marcus for sharing the breakoutroom and music with me 😊
Thanks to Flo and the rest of teaching team for the last 9 weeks.