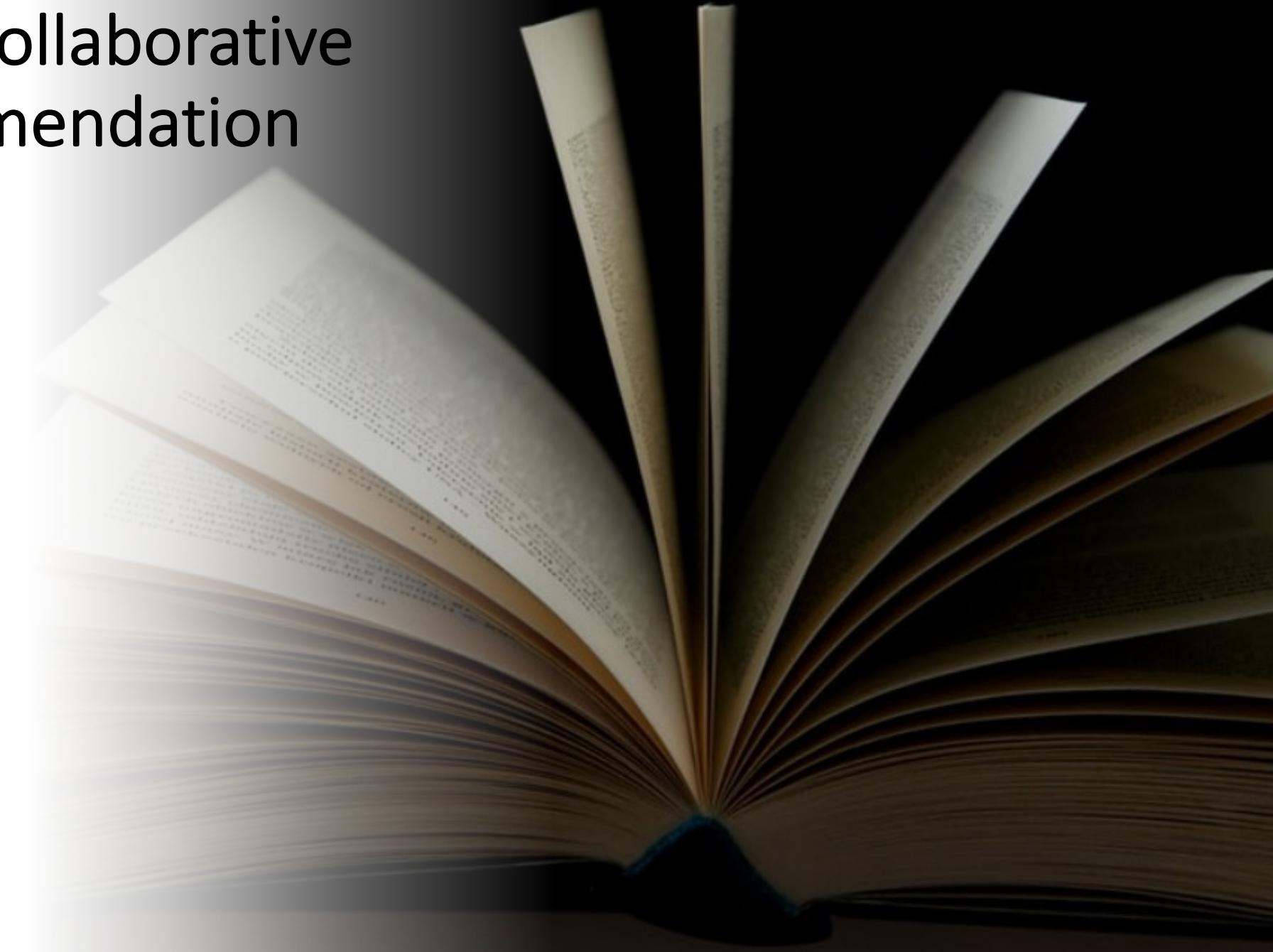


User-Based Collaborative Book Recommendation System

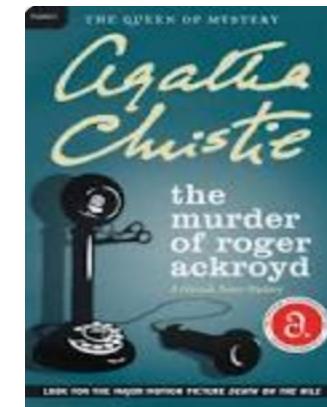
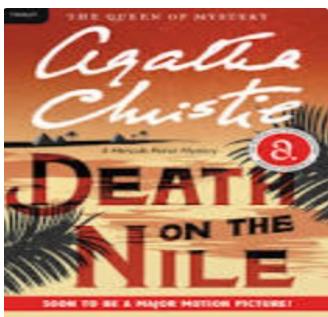
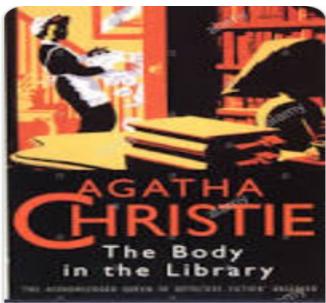
Sweta Prabha



Collaborative Recommendation System

- **Recommender systems** that recommend items through **user collaborations**
 - most widely used and proven method of providing recommendations
 - Two Types:
 - USER-BASED
 - ITEM BASED

Item-based Collaborative Recommendation



SIMILAR RATINGS & PURCHASED TOGETHER
THEY BELONG TO THE SAME CLASS !!!

User-Based Collaborative Recommendation



Data

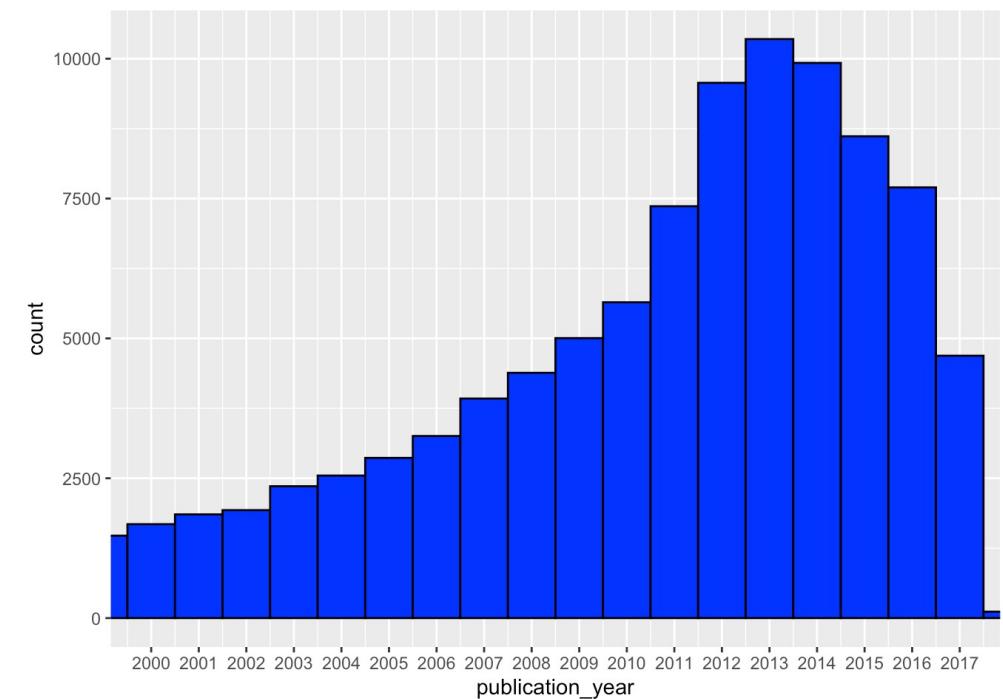
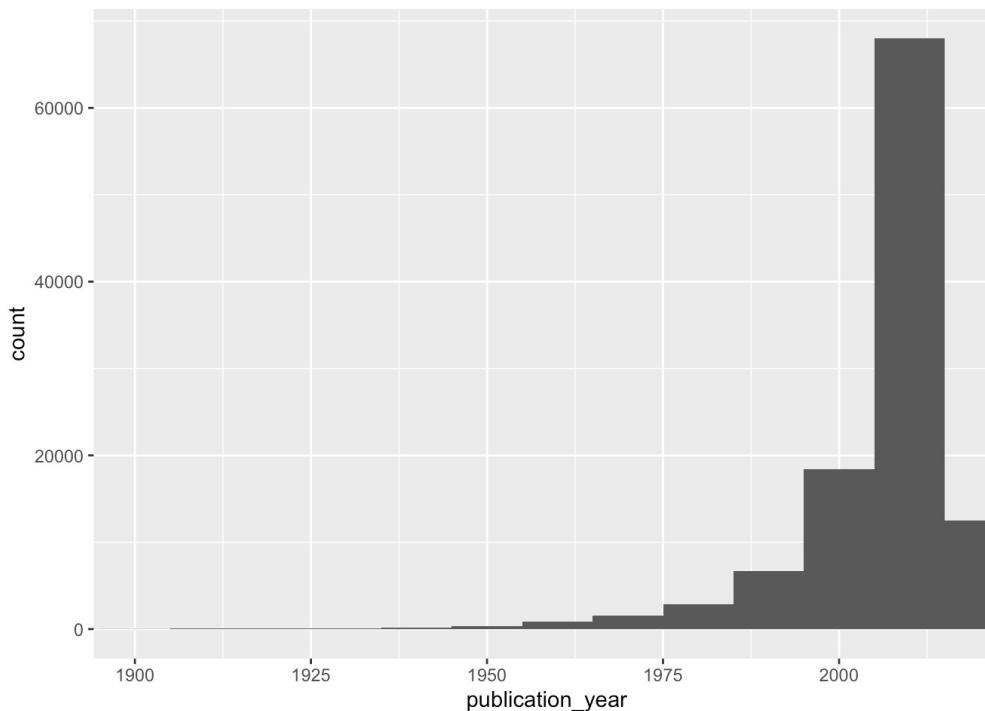
- Goodreads book dataset from UCSD Repository
 - Book Detail Data with 27 features for 150k rows
 - Important features: Title, publishers, ratings, image_url, language
 - User Rating Dataset with 10502 rows and 6 features
 - Important features: user_id, book_id, user_rating

Exploratory Data Analysis

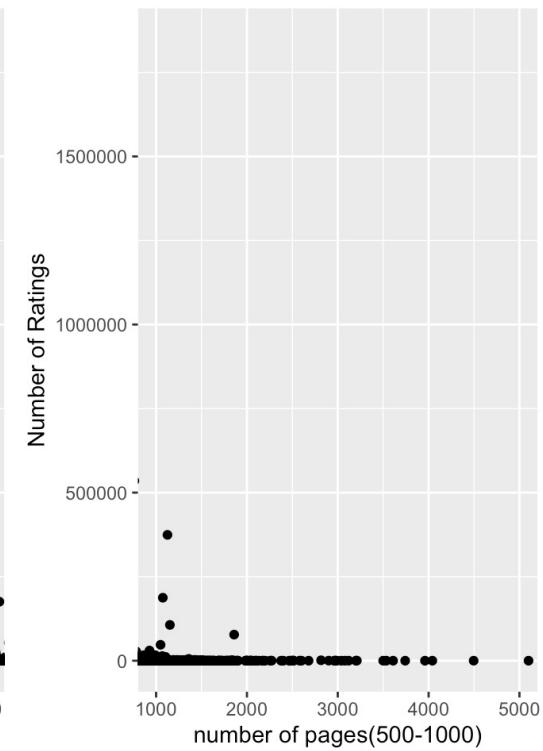
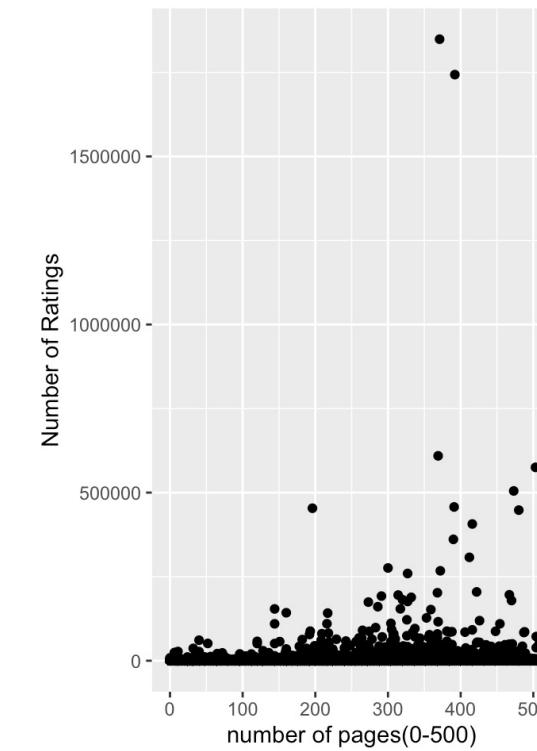
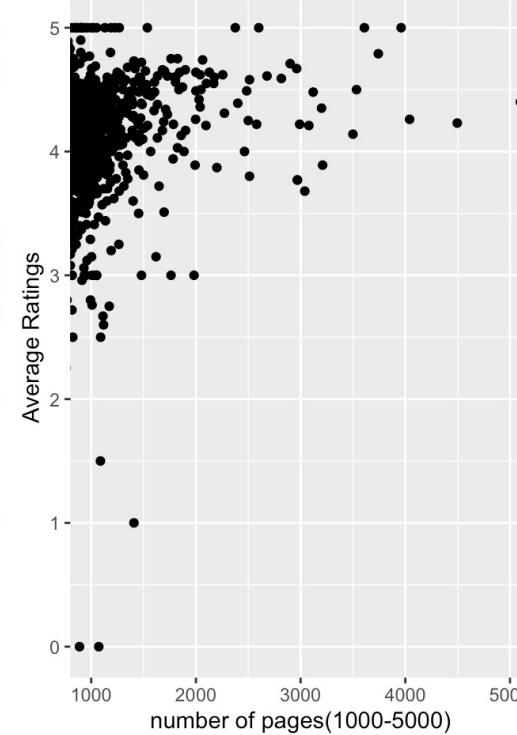
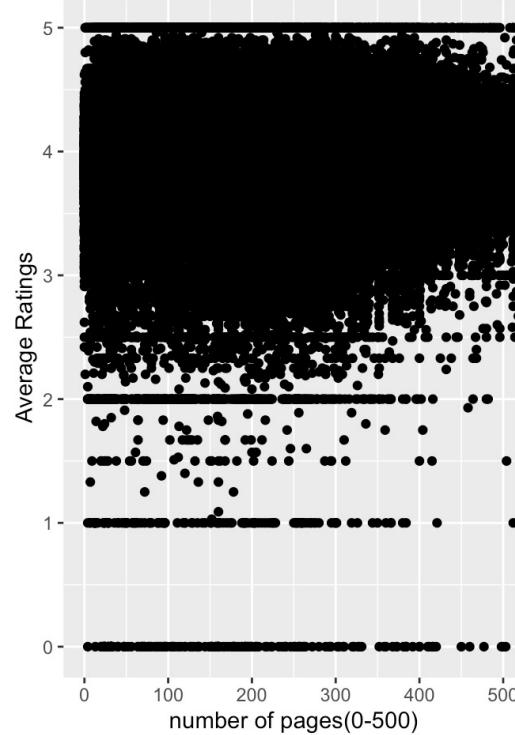
- Number of books published each year
- How does number of pages affect popularity?
- Which books have the maximum versions?
- What are the most popular languages?
- How did popularity of eBooks change in last decade?
- Which format is most popular?
- How are average rating and popularity related?
- What ratings are more likely?
- Top 10 publishers and their popularity?

Number of books Published Each Year

- More books were published after 2007. This is when kindle was launched, and eBooks gained popularity



Number of Pages

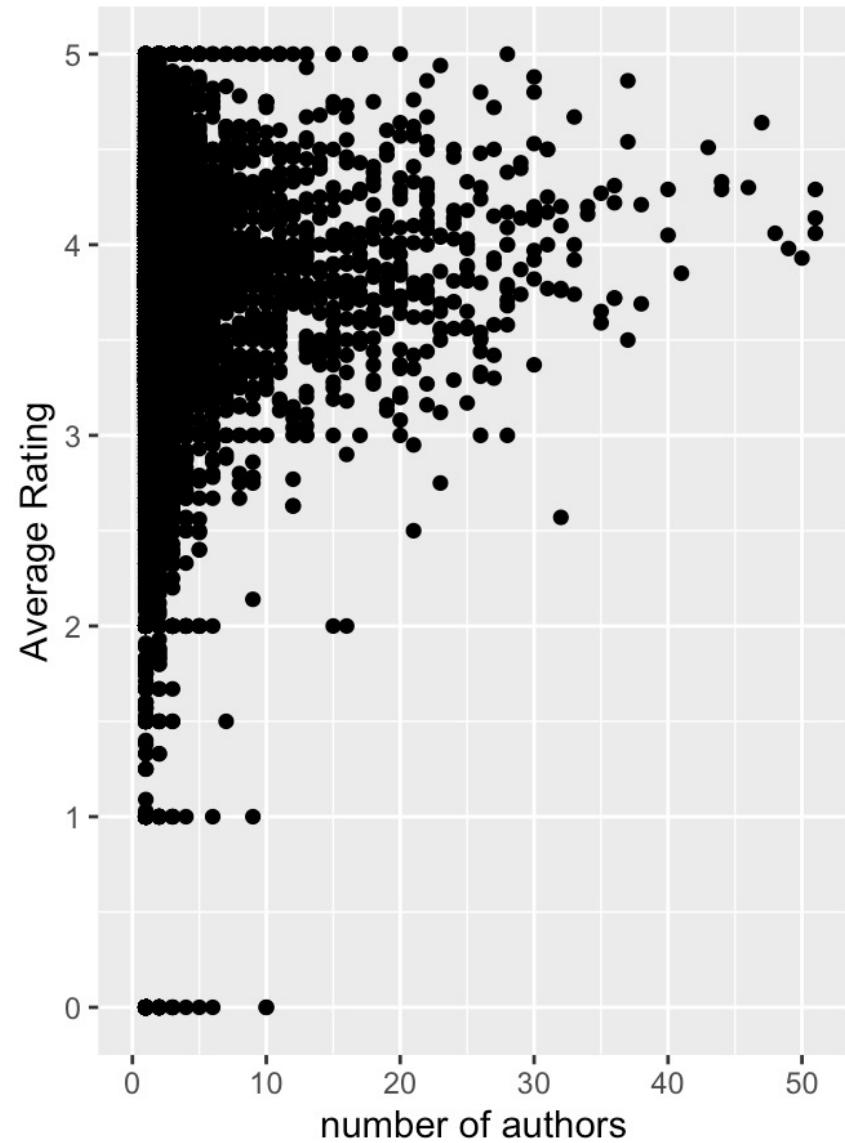


Less people buy THICK BOOKS! but THICK BOOKS are never rated low!

Most popular books(maximum number of ratings) are between 200-500 pages

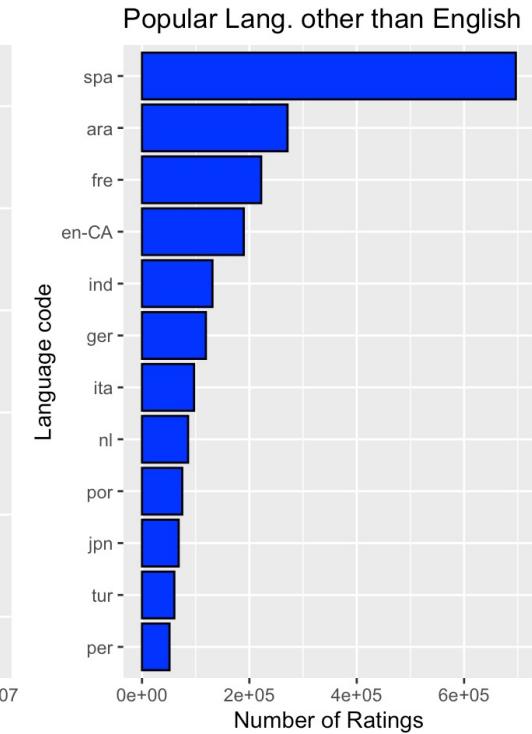
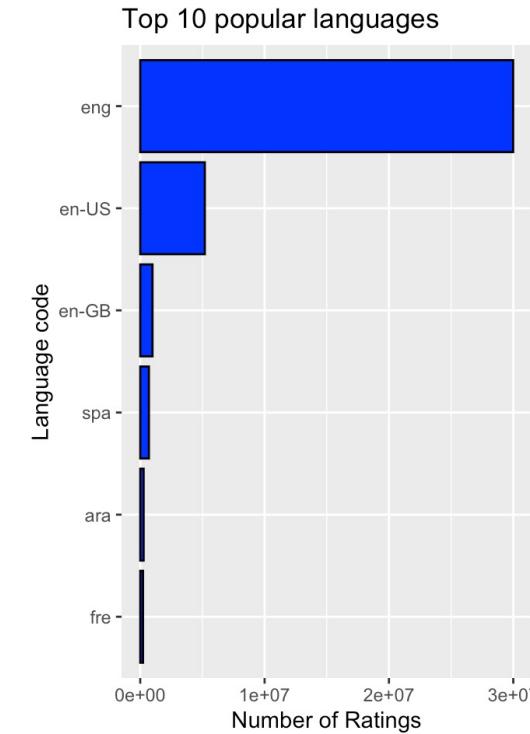
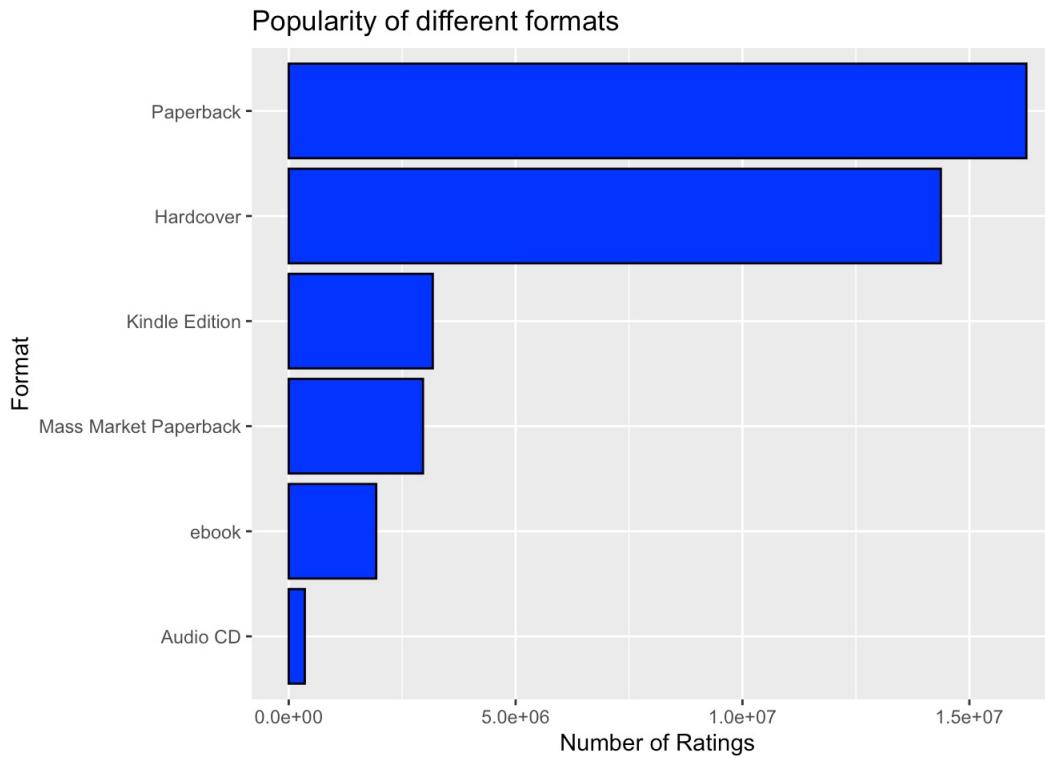
Number of Authors

Books with more than 10 authors are highly rated



Language & Format

- Kindle edition is more popular than other forms of eBooks
- English is the most popular language followed by Spanish, Arabic and French

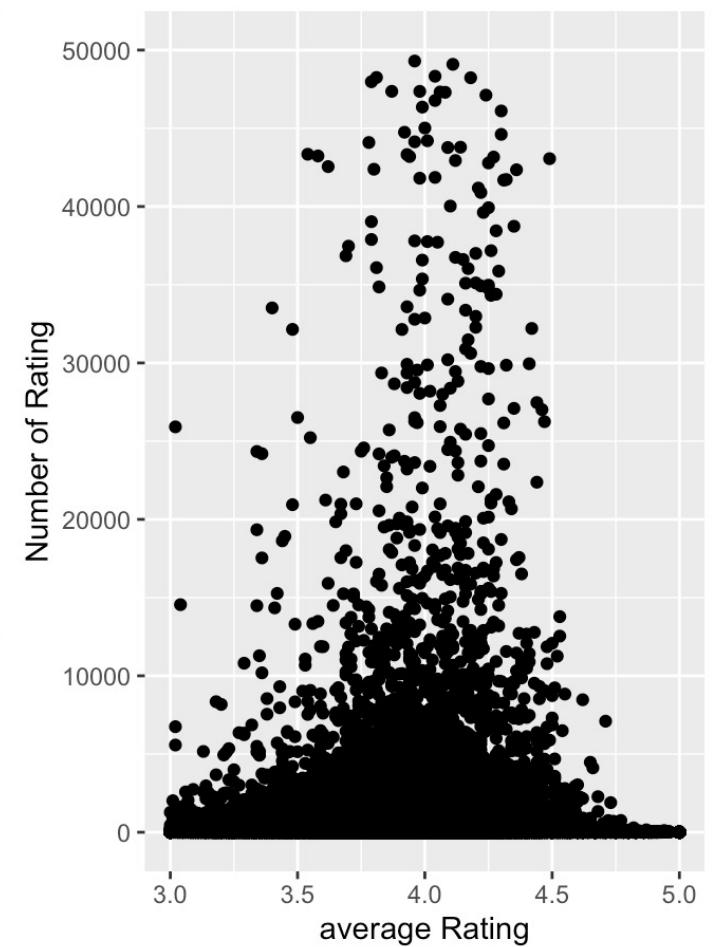
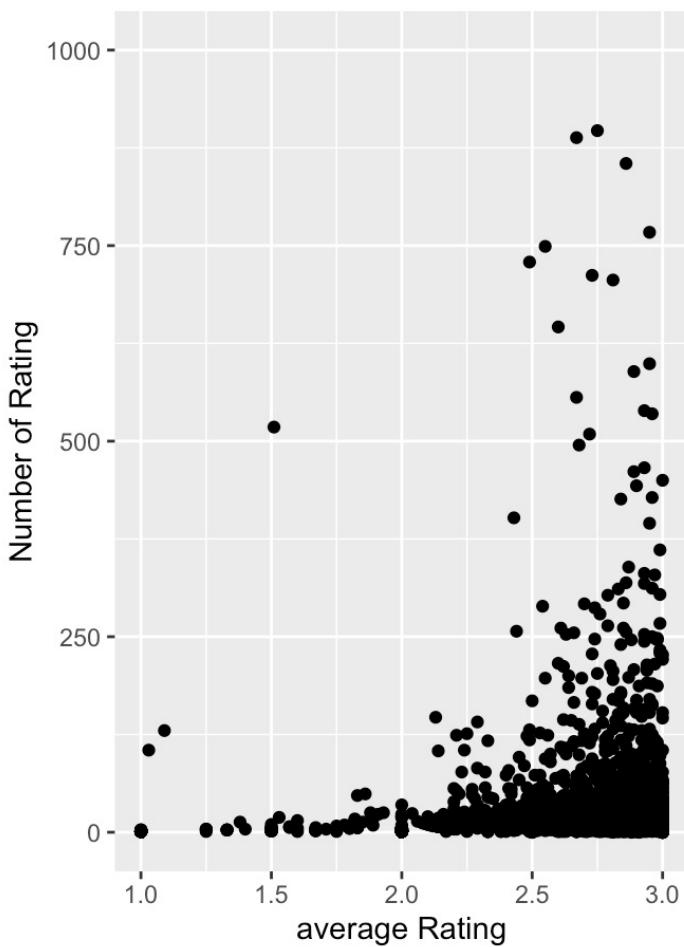


Average Rating & Popularity

No books with average rating 5
that has more than 500 reviews

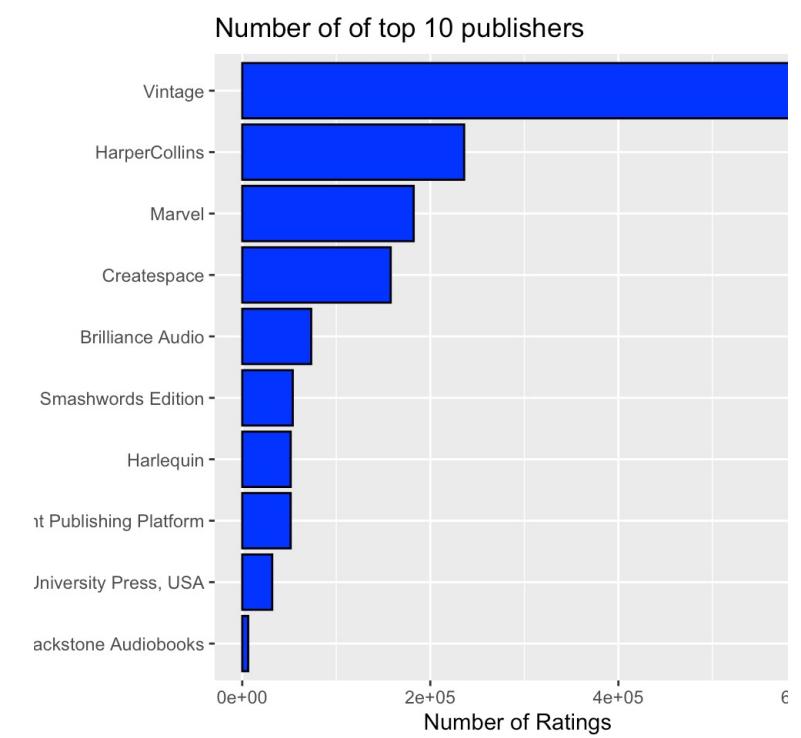
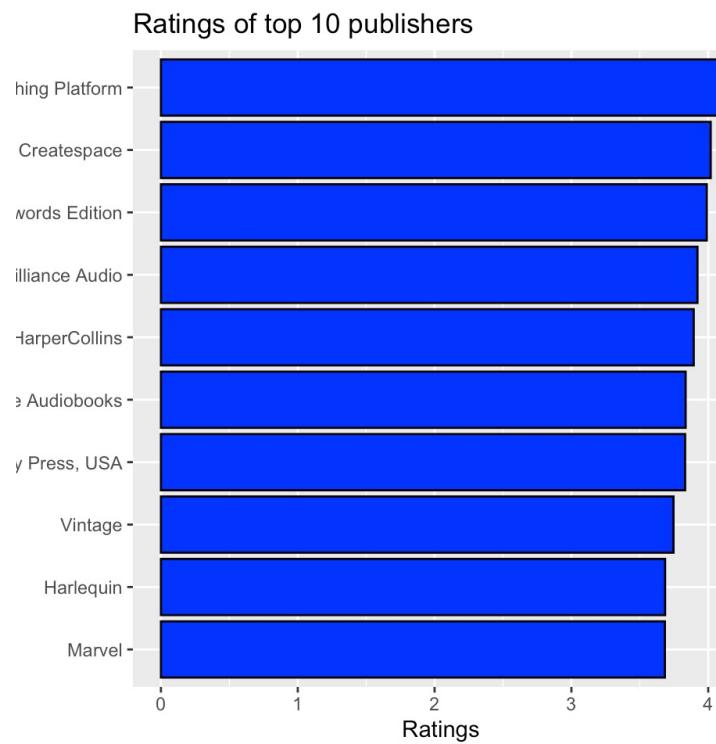
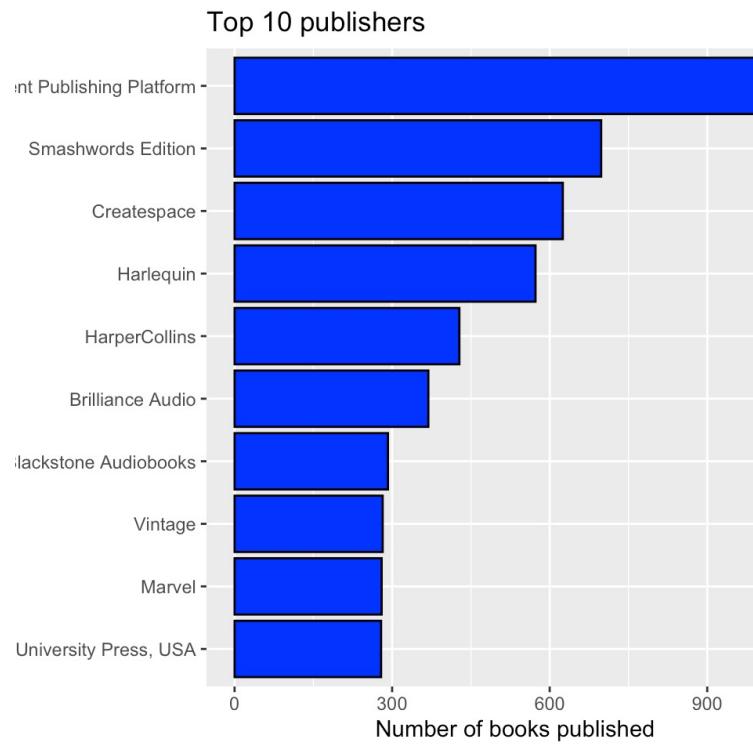
THERE ARE CRITICS
EVERYWHERE!!! ITS DIFFICULT
TO KEEP EVERYONE HAPPY

Average rating above 3.5 is safe



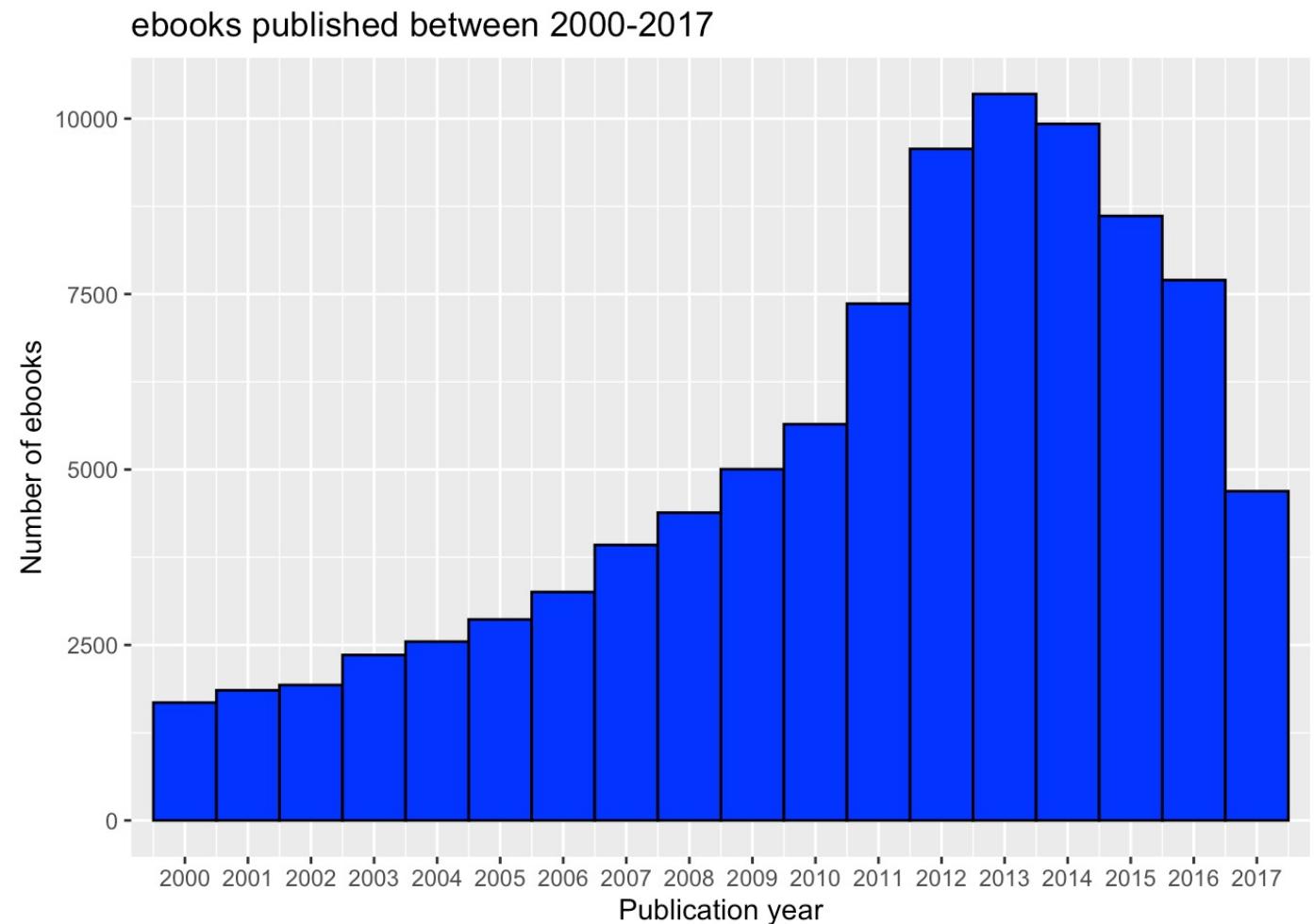
Publishers

- CreateSpace , self publishing platform owned by amazon publishes maximum number of books, highly rated and among top 10 in popularity
- Vintage is the most popular publishing house



Popularity of eBooks

- Kindle Revolutionized people's reading behavior
- After its release in 2007, the number of eBooks have grown tremendously
- In 2011, kindle helped amazon touch \$10Bn revenue



Recommendation System

- Similarity between users was derived from similarity of the recommended books

Creating the user-item matrix based on recommendation

```
user_item = Rating %>% pivot_wider(names_from=book_id, values_from = rating) %>%
  as.data.frame()
row.names(user_item)=user_item$user_id
user_item$user_id = NULL
user_item = as.matrix(user_item)
user_item[1:5,1:5]
```

```
##      741840 1890038 11295329 9264003 7920056
## 3717      4     NA     NA     NA     NA
## 2729     NA      4     NA     NA     NA
## 2955     NA     NA      5      5      3
## 2341     NA     NA     NA     NA     NA
## 1858     NA     NA     NA     NA     NA
```

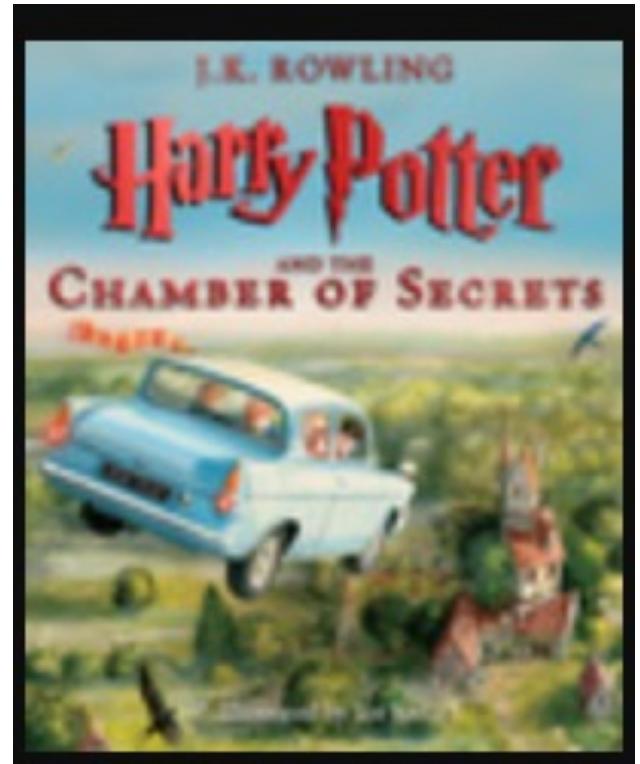
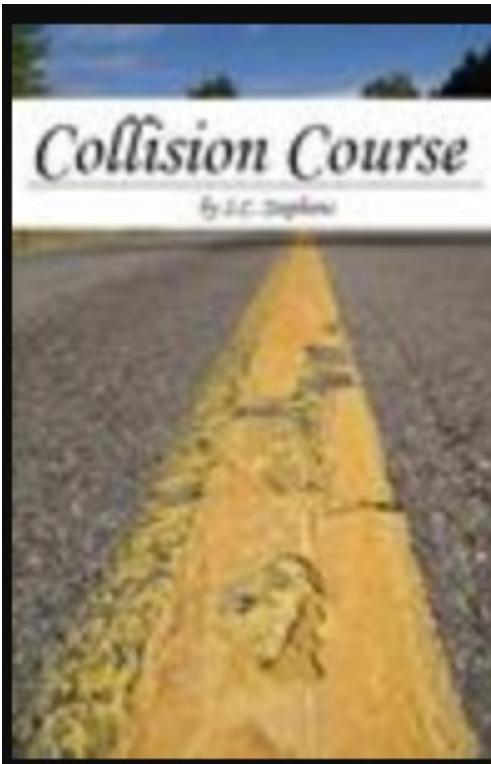
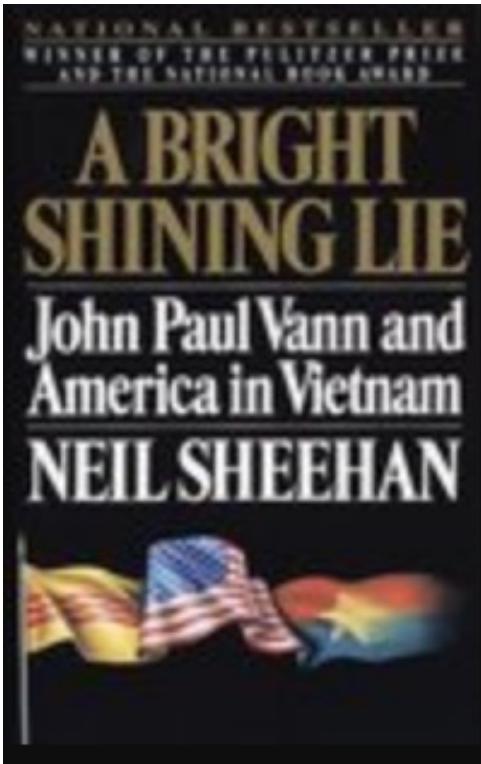
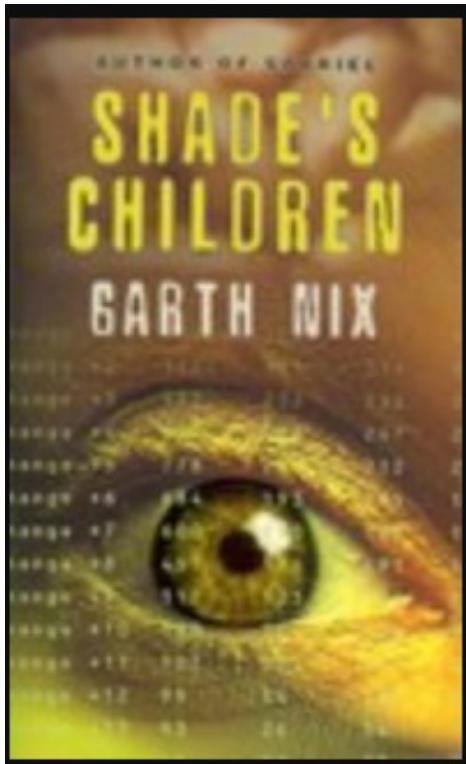
Top 5 Recommended books for each user based on maximum similarity

Retrieving the top 5 recommended books for user 2955

```
recom_cf_user = user_recommendation("2955", n_recommendation = 5)  
recom_cf_user
```

```
## # A tibble: 5 × 3  
##   book_id count rating  
##       <int>  <int>  <dbl>  
## 1     6708      4      5  
## 2   9879625      4      5  
## 3  29241319      4      5  
## 4     47626      3      5  
## 5   417640      3      5
```

User 2955:



FUTURE WORKS

- Hybrid model using User-based and Item-Based Models
- Dashboard with scalable code
- A/B Testing of the final Model



THANK YOU!

