Open-Source Data

BOOK RECS ANALYSIS

Greta Lawani

CONTEXT

In my role as a newly hired Data Analyst at the fictitious Data Analytics company BookWise Metrics, our focus involves analyzing and strategizing for the integration of a new user feedback system.

The last few decades have witnessed the ascendance of platforms such as YouTube, Amazon, Netflix, and various other web services, solidifying the pervasive role of recommender systems in our daily lives. From shaping the landscape of e-commerce by suggesting items that align with buyer preferences to refining online advertising by delivering content tailored to individual user tastes, recommender systems have become an indispensable element of our digital experiences.

PROJECT OVERVIEW

Provide traditional book publishing companies with a recommender system to effectively compete and maintain relevance in today's dynamic book market, by leveraging data-driven personalization

Data

Kaggle Dataset from **Möbius** since February 9th 2024, compiled by **Cai-Nicolas Ziegler** from the **Book-Crossing community**

EDA summarizing the dataset

Data Limitations

The data is **static** and won't undergo updates. No revenue data is available

278,858 anonymized users providing **1,149,780** ratings about **271,379** books

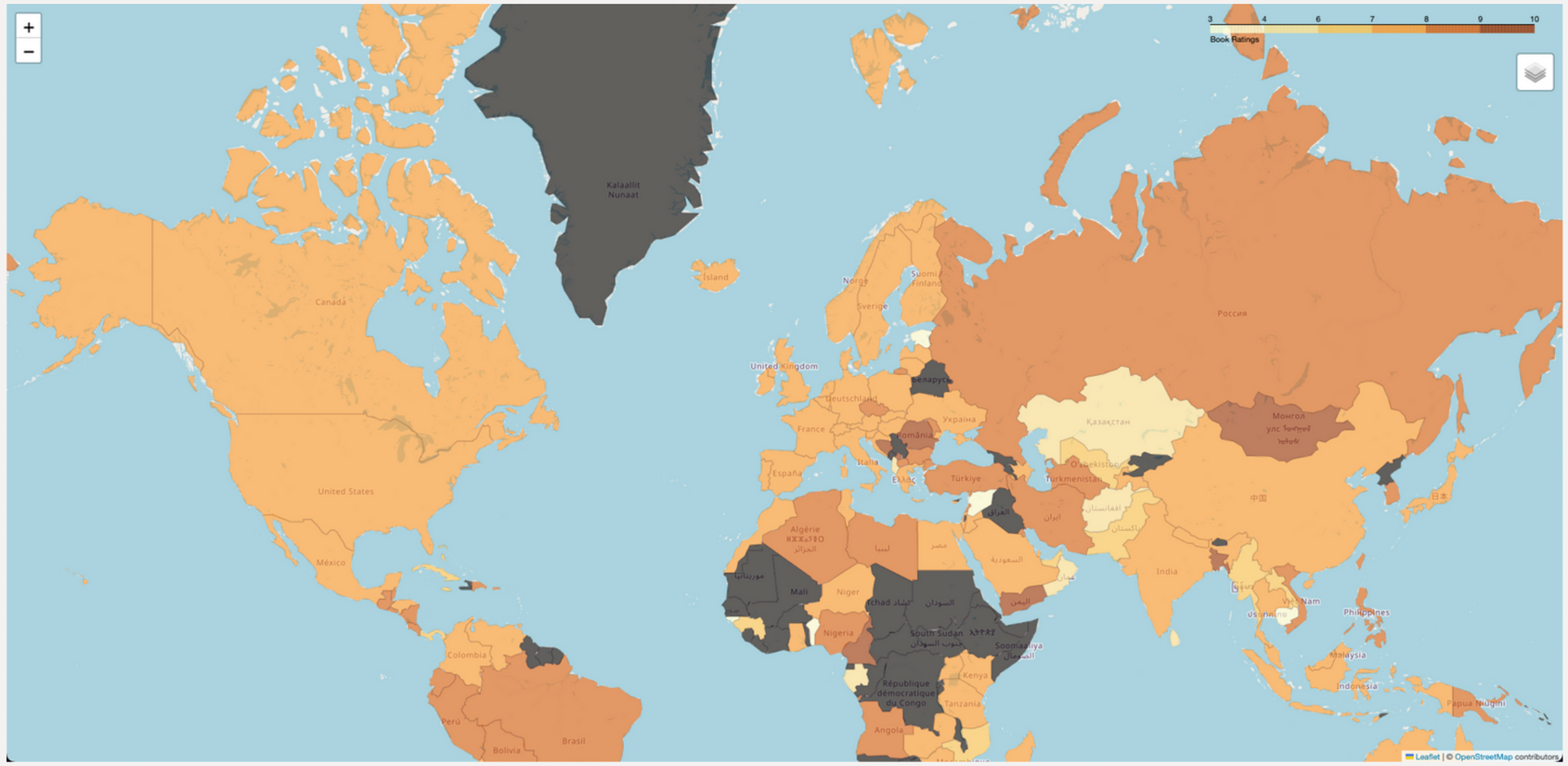
Skills

Data wrangling
Geo-Visualization
Supervised ML
Regression
Unsupervised ML
Clustering
Sourcing and Analysing
Time-series Data

Tools

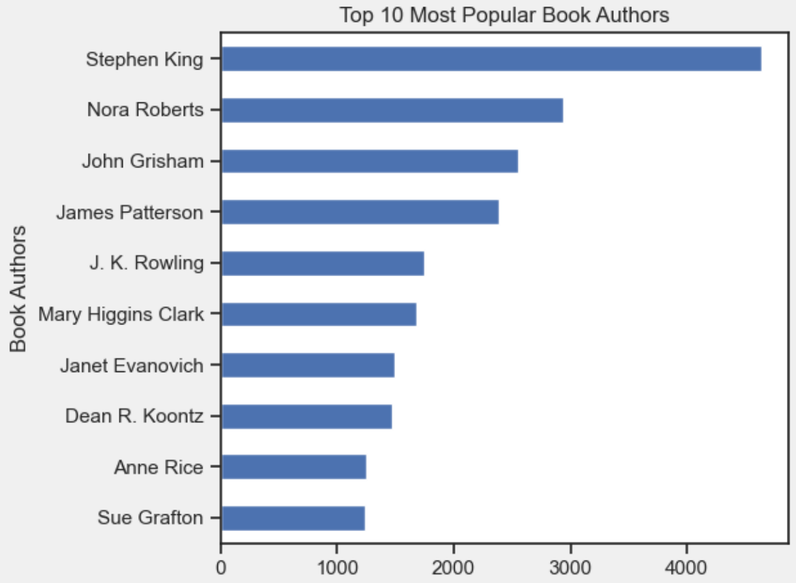
Python
Jupyter Notebooks
GitHub
Tableau
Canva

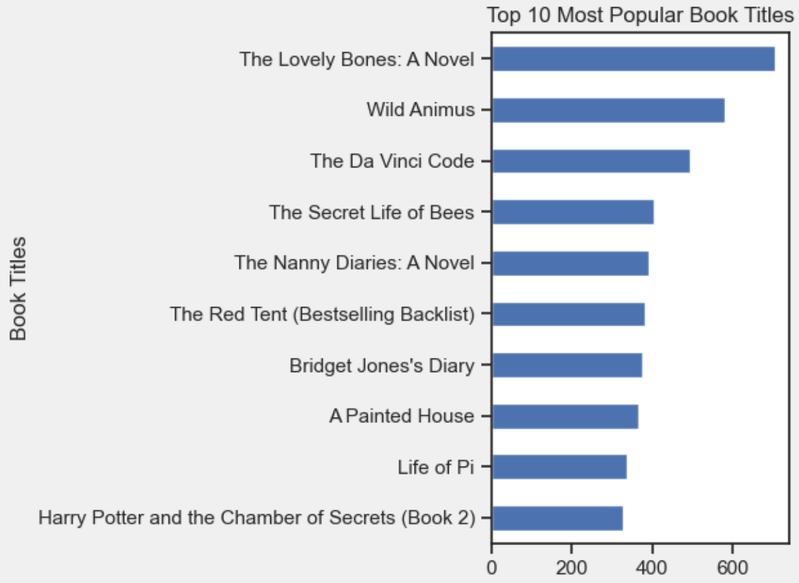
GEO VISUALIZATION



The book ratings per country are showing i.e. an average rating of **7.6** in the United States, the **number one country** in terms of users and ratings

POPULARITY ANALYSIS

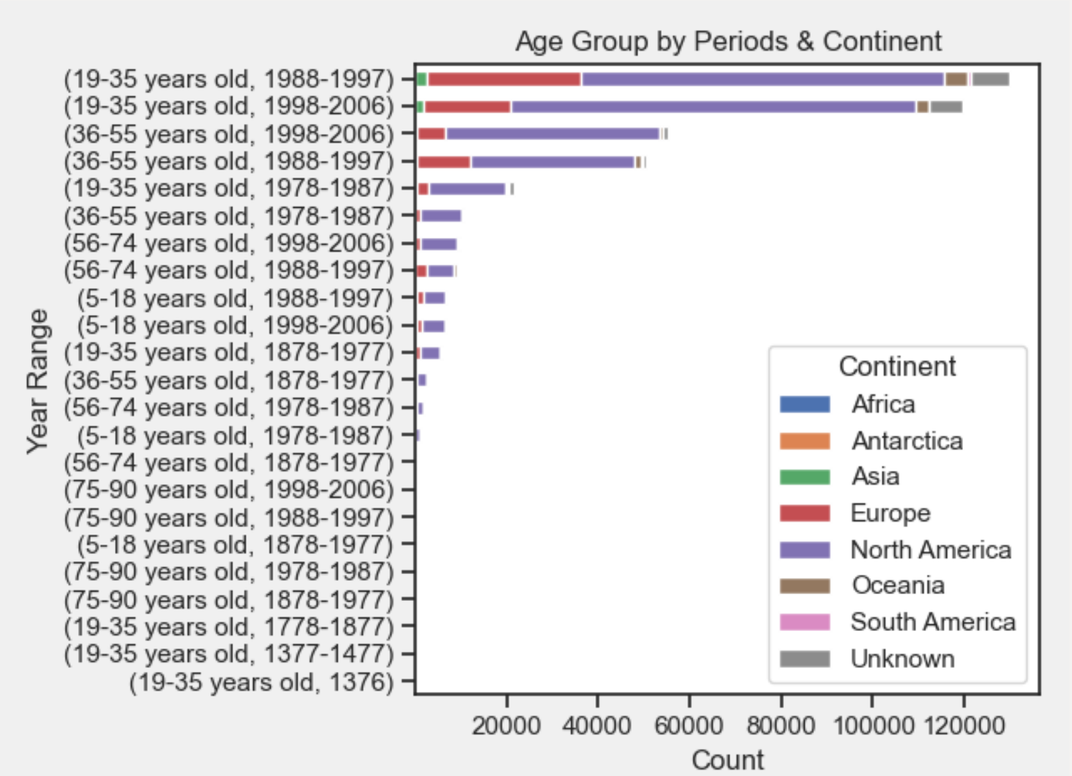




Top 3 authors of all time: Stephen King, followed by Nora Roberts and Josh Grisham are the most-read books

Top 3 books of all time: The Lovely Bones: A Novel, Wild Animus and The Da Vinci Code

GEOGRAPHIC ANALYSIS



Age Groups categorised by *year range of publications* and *continents*:

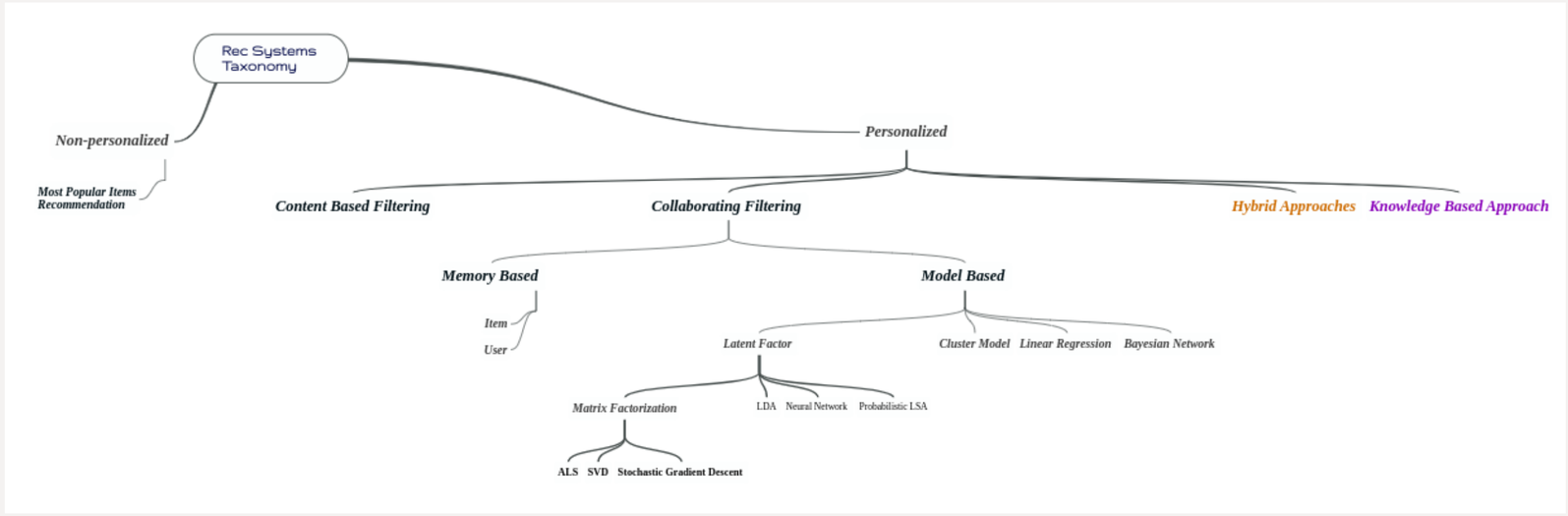
- **North America** with the **19-35 years old group** is leading for published books between ***1988 and 1997***, then from ***1998 to 2006***

RECOMMENDATIONS AND NEXT STEPS

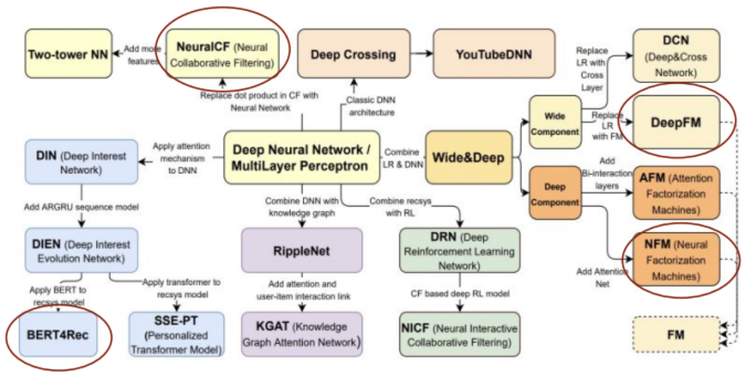
Consider the Key Insights	The most popular books, titles, and ratings for all the markets are now defined. **North American market**, with the **19-35 age group** is the most active group and region.
Consider the Data Limitations	With a dynamic data update, the next steps in the analysis would have involved recalculating the most popular books, and genres for accurate and personalized recommendations based on users' reading history.
Consider the Next Action	Creating a new column that automatically assigns specific categories to different genres will facilitate a more thorough clustering analysis, better-accommodating people's diverse topic interests.

GitHub, Tableau Deck, Tableau Dashboard

APPENDIX

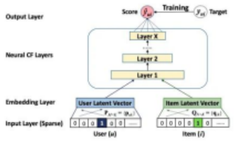


Deep Learning Recommendation Algorithms



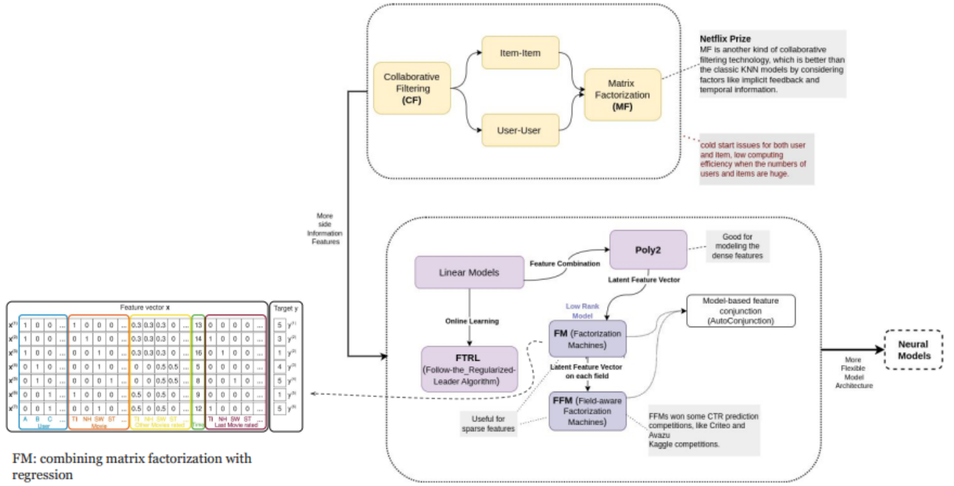
Evolution of deep learning recommendation model

- Classic deep neural network
- Shallow and deep model
- Natural Language Processing (NLP) inspired neural model
- Deep Reinforcement learning
- Graph neural network



"Neural Collaborative Filtering" by X He, L Liao, H Zhang, L Nie, X Hu, TS Chua the 26th international conference on world wide web, 2017

Classical Recommendation Algorithms



FM: combining matrix factorization with regression

THANK YOU

Greta Lawani