Book Recommendation System

Team Members

Daria Zinchenko

Neha Sharma

Seeke Olu-Daniel

Valentyna Kravets

Outline

Project Goals

Methodology

Recommendation Algorithm

Website Demo & Data visualization

Insights, Conclusion & Next Steps

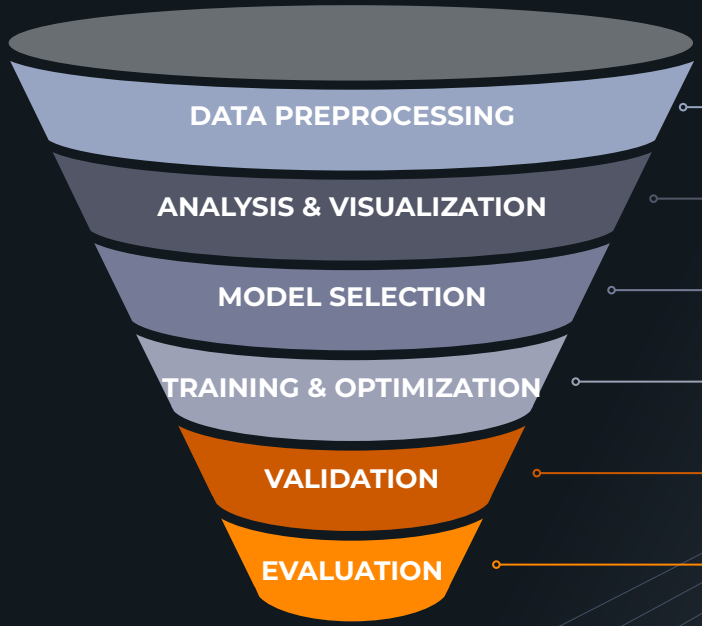
“Books are the plane, and the train, and the road. They are the destination, and the journey. They are home.”

– Anna Quindlen

Project Goals

- Increased user engagement by providing personalized recommendations aligned with their interests and preferences.
- Boost book sales.
- Maximize inventory utilization by promoting niche books to users likely to be interested in such.
- Enhance user experience.
- Optimize recommendation algorithms.

Methodology



Define problem, collect, clean, explore and split data

Identify trends and generate insights

Evaluate and compare several models

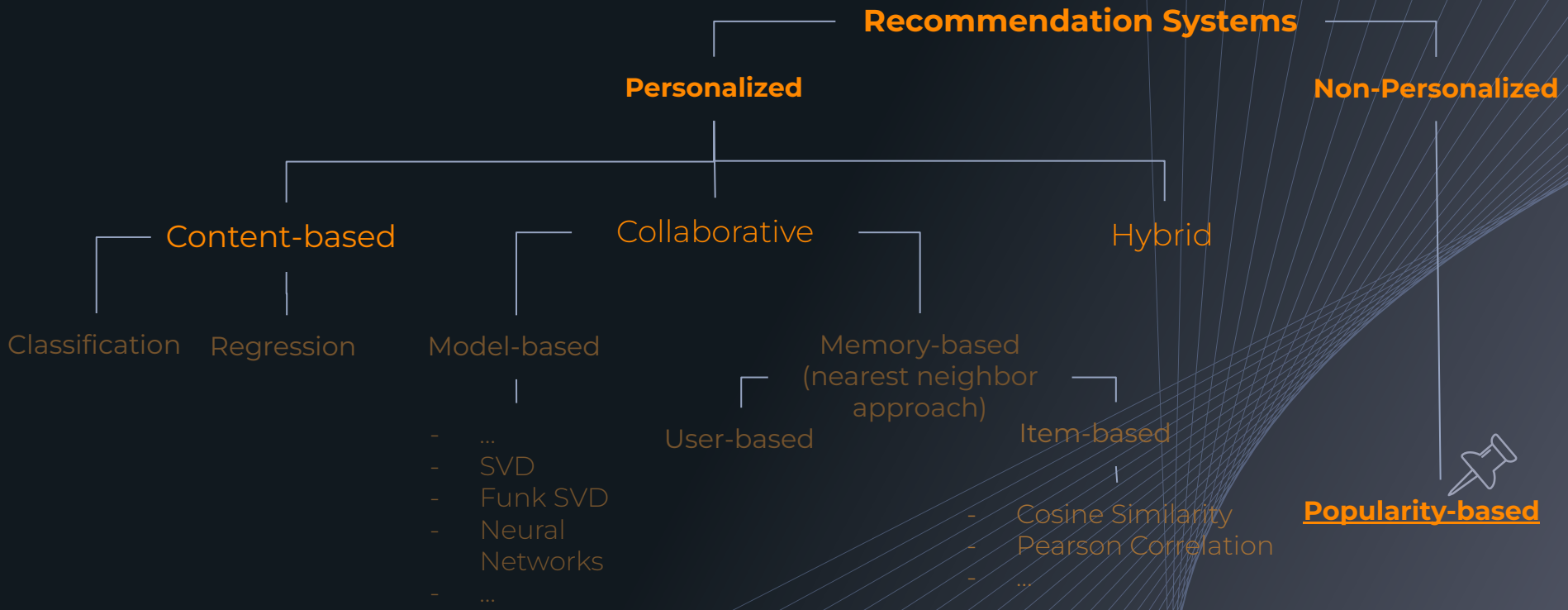
Establish performance baseline, hyperparameter tuning

Assess generalization performance

Compare performance and select best model

Recommendation Algorithm

Recommendation Algorithm



Recommendation Algorithm

Behavior-driven

Serendipity

Cold Start

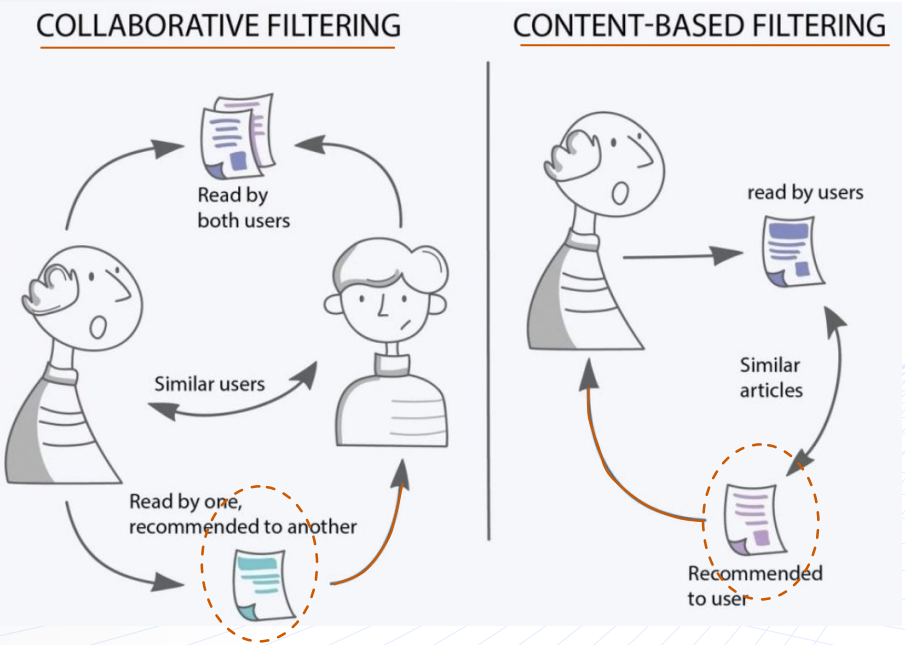
Popularity Bias

New Items among Recommendations

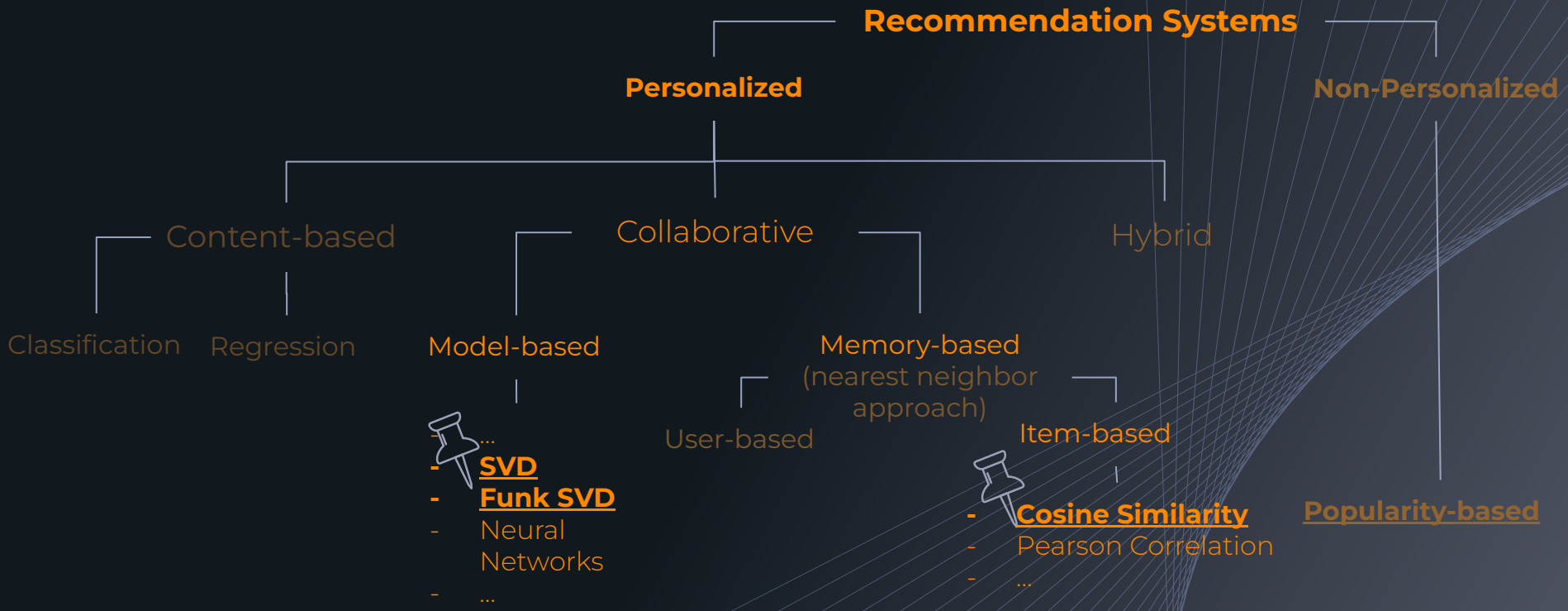
Transparency

Limited Analysis

Over-specialization



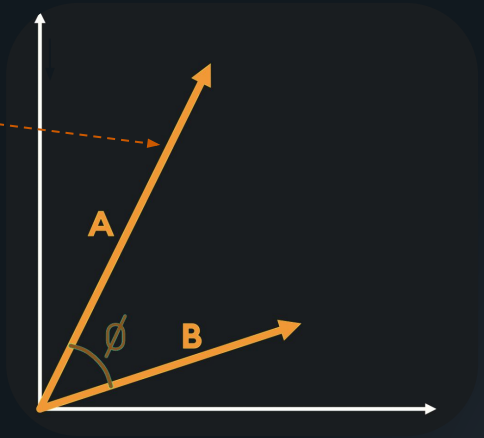
Recommendation Algorithm



Cosine Similarity

	User_1	...	User_n
Book_1	a_1	...	a_n
...	...	...	...
Book_m	b_1	...	b_n

a_i, b_i ∈ [1,10]



$$\cos(\theta) = \frac{a \bullet b}{\|a\|\|b\|} = \frac{(a_1 * b_1) + (a_2 * b_2) + \ldots + (a_n * b_n)}{\sqrt{(a_1^2 + a_2^2 + \ldots + a_n^2) * (b_1^2 + b_2^2 + \ldots + b_n^2)}}$$

Why?

Since overall ratings and popularity can create *different distances*, but the *direction of similar vectors remains close.*

Singular Value Decomposition

	Book_1	...	Book_n
User_1	x_{11}	...	x_{1n}
...	...	...	...
User_m	x_{m1}	...	x_{mn}

Data normalization

Matrix Factorization (Decomposition)

*Extracted top **r** latent factors*

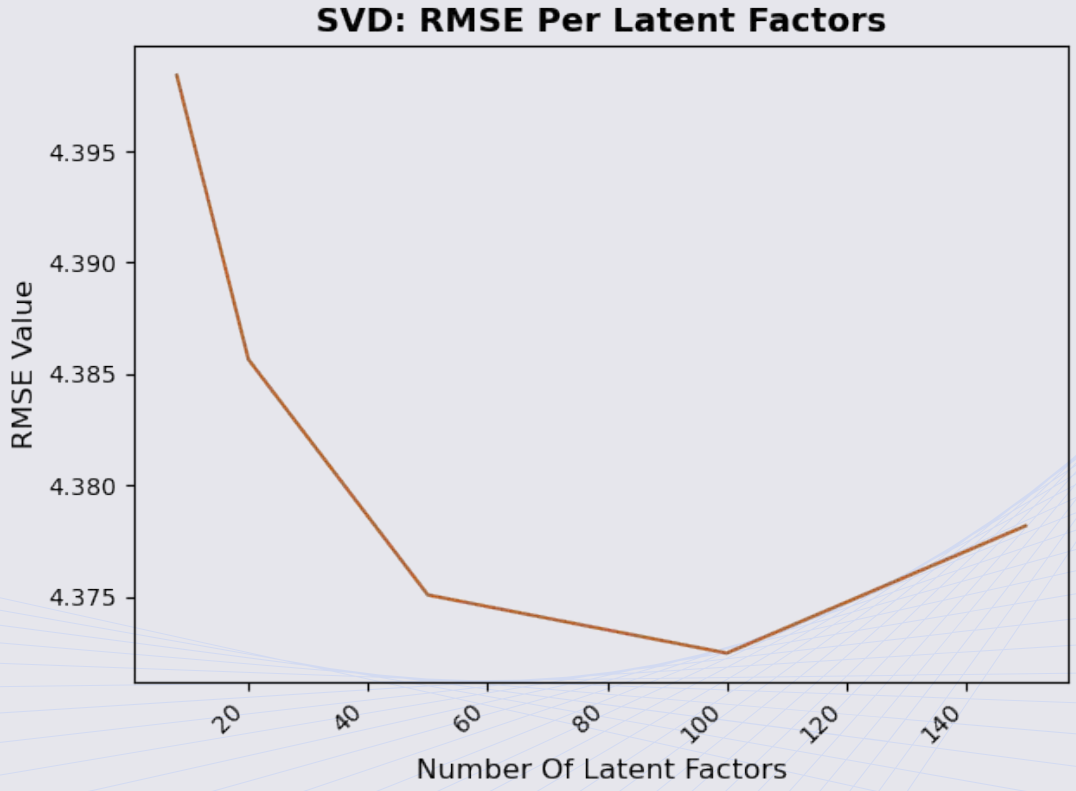
Predictions (Dot product)

De-normalization

Recommendations (Top rated titles, which were not read)

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & \\ \vdots & \vdots & \ddots & \\ x_{m1} & & & x_{mn} \end{pmatrix}_{m \times n} = \underset{m \times r}{\begin{pmatrix} u_{11} & \cdots & u_{1r} \\ \vdots & \ddots & \\ u_{m1} & & u_{mr} \end{pmatrix}} \underset{r \times r}{\begin{pmatrix} s_{11} & 0 & \cdots \\ 0 & \ddots & \\ \vdots & & s_{rr} \end{pmatrix}} \underset{r \times n}{\begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \\ v_{r1} & & v_{rn} \end{pmatrix}}$$

Model Evaluation



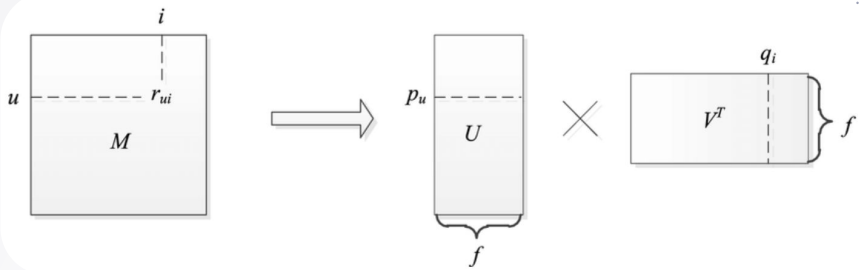
$RMSE=\sqrt{(\sum(Predicted\_i-Actual\_i)^2/N)}$

SVD Funk

NETFLIX

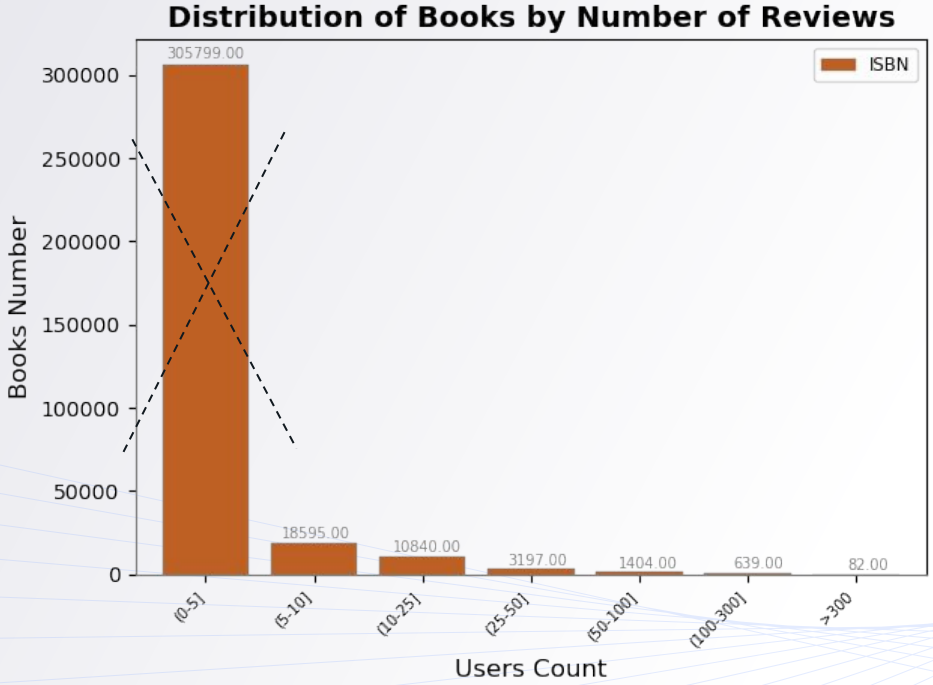
surprise

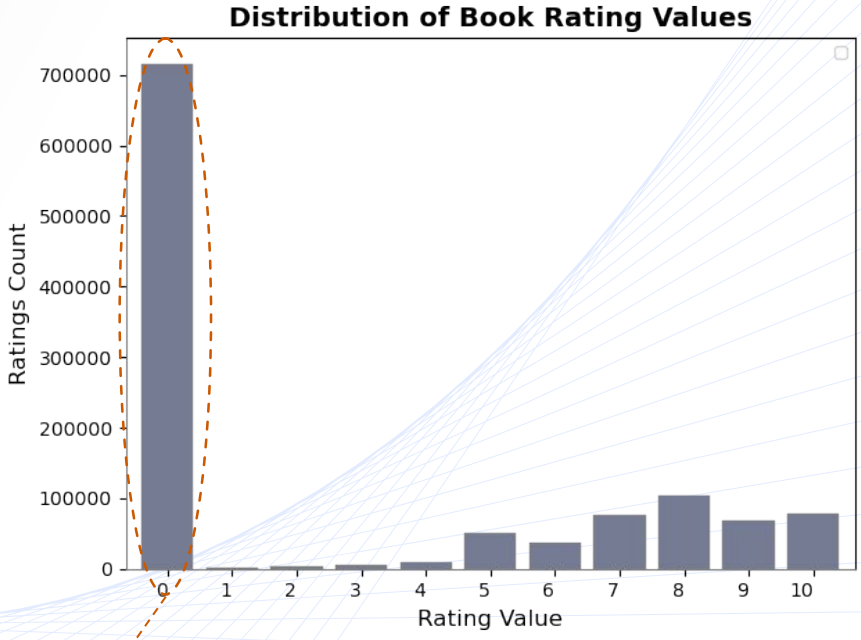
A Python scikit for recommender systems.



- Well-suited for Sparse Matrix
- Focuses on Accuracy Improvement

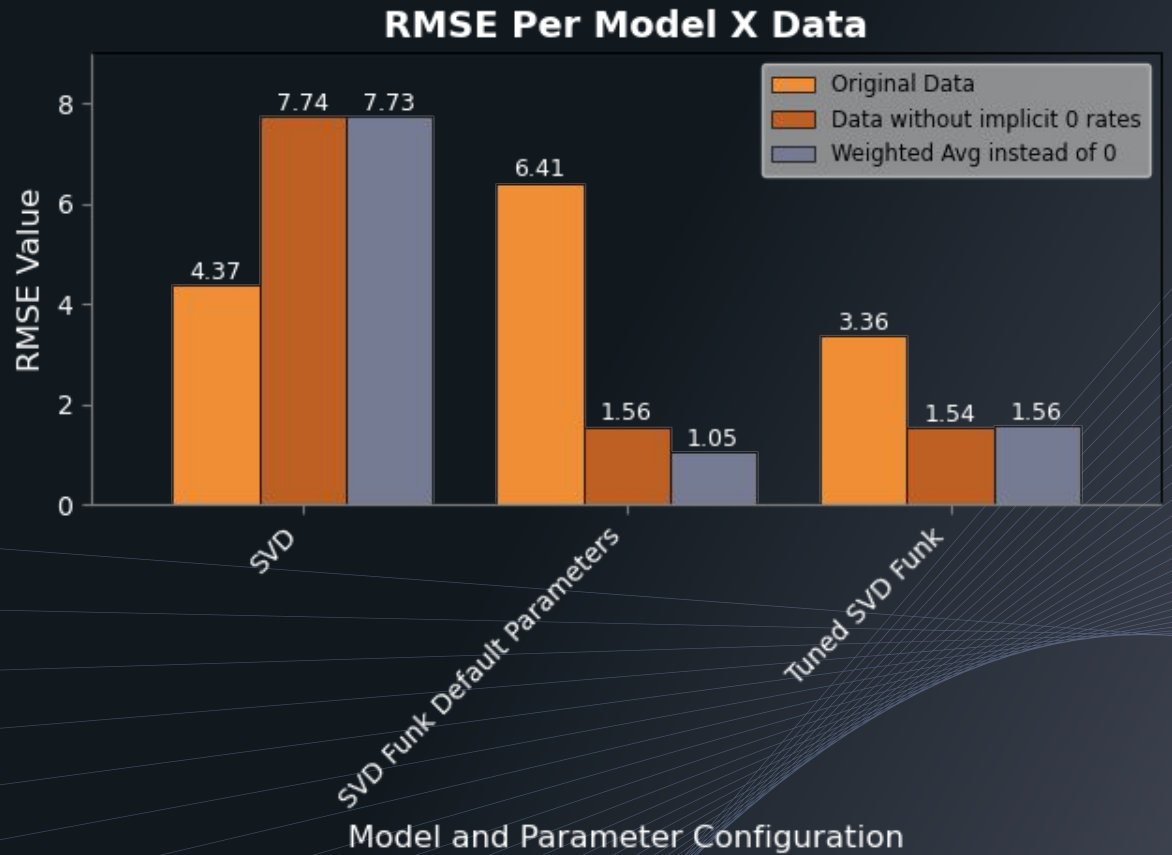
Algo Input Data





$$\frac{\text{Book\_Average\_Rating} * \text{Ratings\_Count} + \text{Min\_Ratings\_Count} * \text{Neutral\_Score}}{\text{Ratings\_Count} + \text{Min\_Ratings\_Count}}$$

Model Evaluation



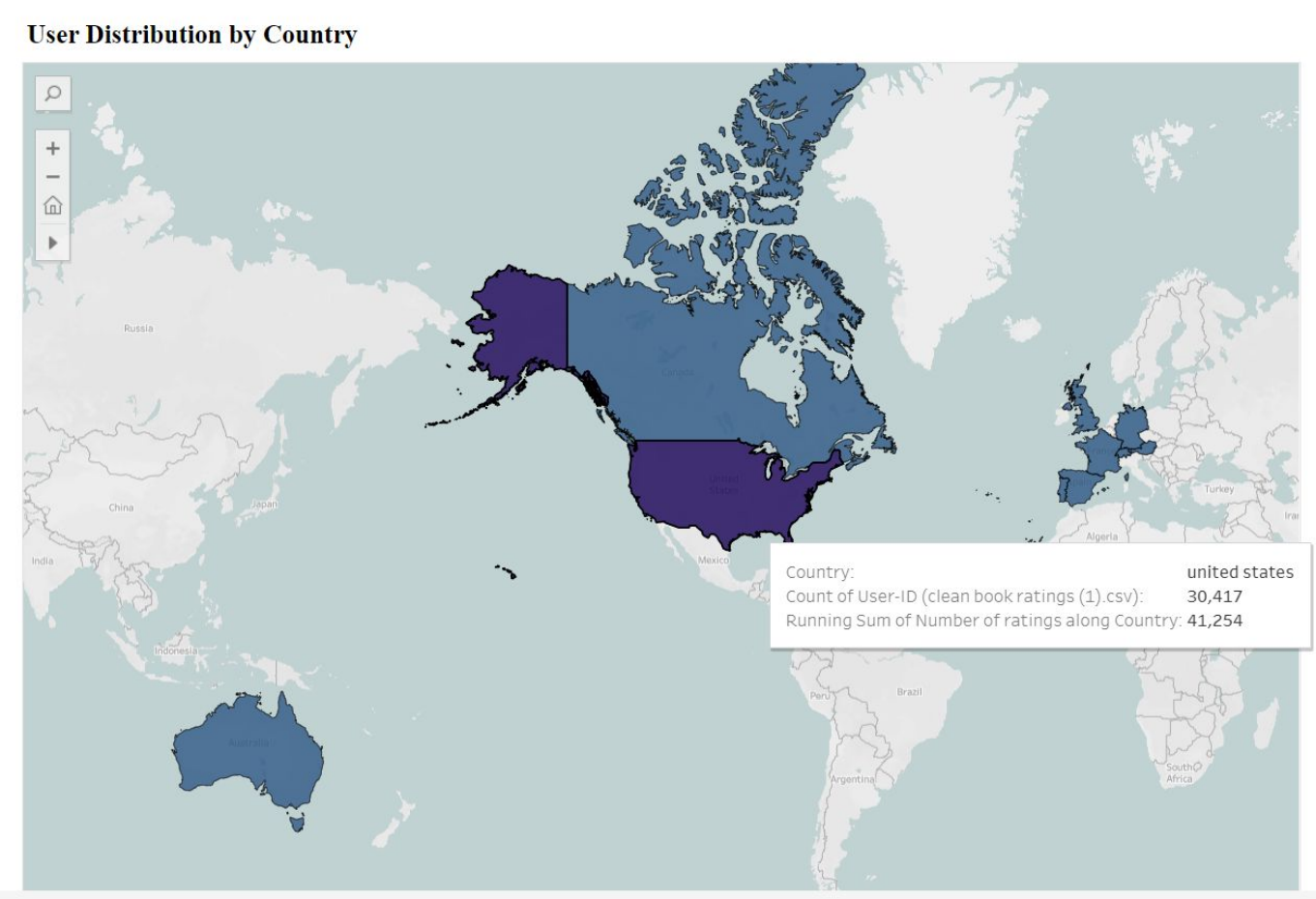
Data Visualization

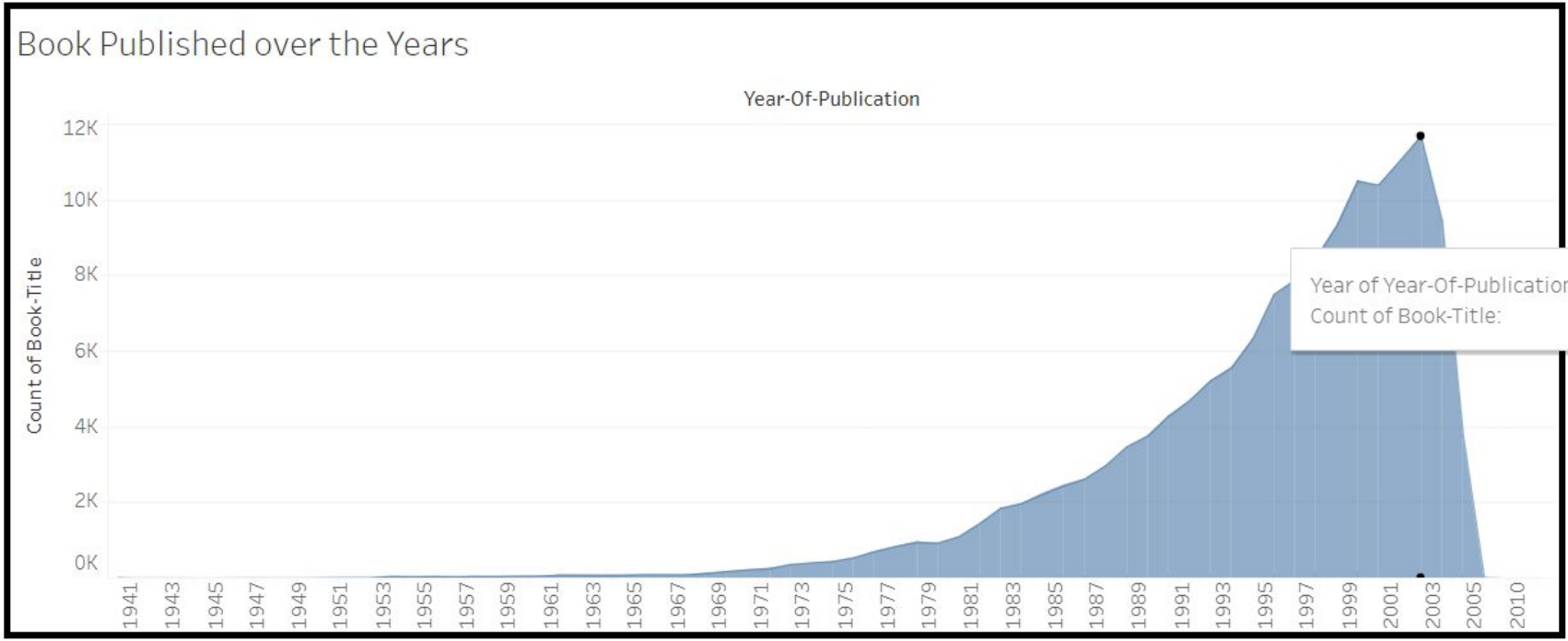
Demo

The Top 10


The Most Rated Authors
Stephen King
Agatha Christie
The Most Rated Publishers
Harlequin
Penguin Books
Bantam Books
Berkley Publishing Group
Publisher: Harlequin
Count of Book-Rating: 3,007
Ballantine Books
Warner Books
Signet Book
Penguin USA
Pocket
Scholastic

User Distribution by Country




Book Published over the Years
Year-Of-Publication
Count of Book-Title
12K
10K
8K
6K
4K
2K
0K
1941
1943
1945
1947
1949
1951
1953
1955
1957
1959
1961
1963
1965
1967
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2010
Year of Year-Of-Publication: 2002
Count of Book-Title: 11,686

Insights & Conclusion

Insights from the Data

- Adding reviews
- Incorporate Additional User Data(age, location, etc)
- Feature Engineering
- Continuous Refinement

Conclusions

Advantages to business:

- Improve Customer Retention
- Identify emerging trends and changes in user preferences
- Easily Analyze the Market
- Increase Sales
- Enhance The User Experience

Future Steps and Improvements

- ❑ Incorporate newly rated books
- ❑ Sentiment Analysis
- ❑ Tags and genre-based search functionalities
- ❑ Expand database to include more languages
- ❑ Feedback loop
- ❑ Predictive analytics
- ❑ Prescriptive analysis

Thanks!

Questions are now welcome.