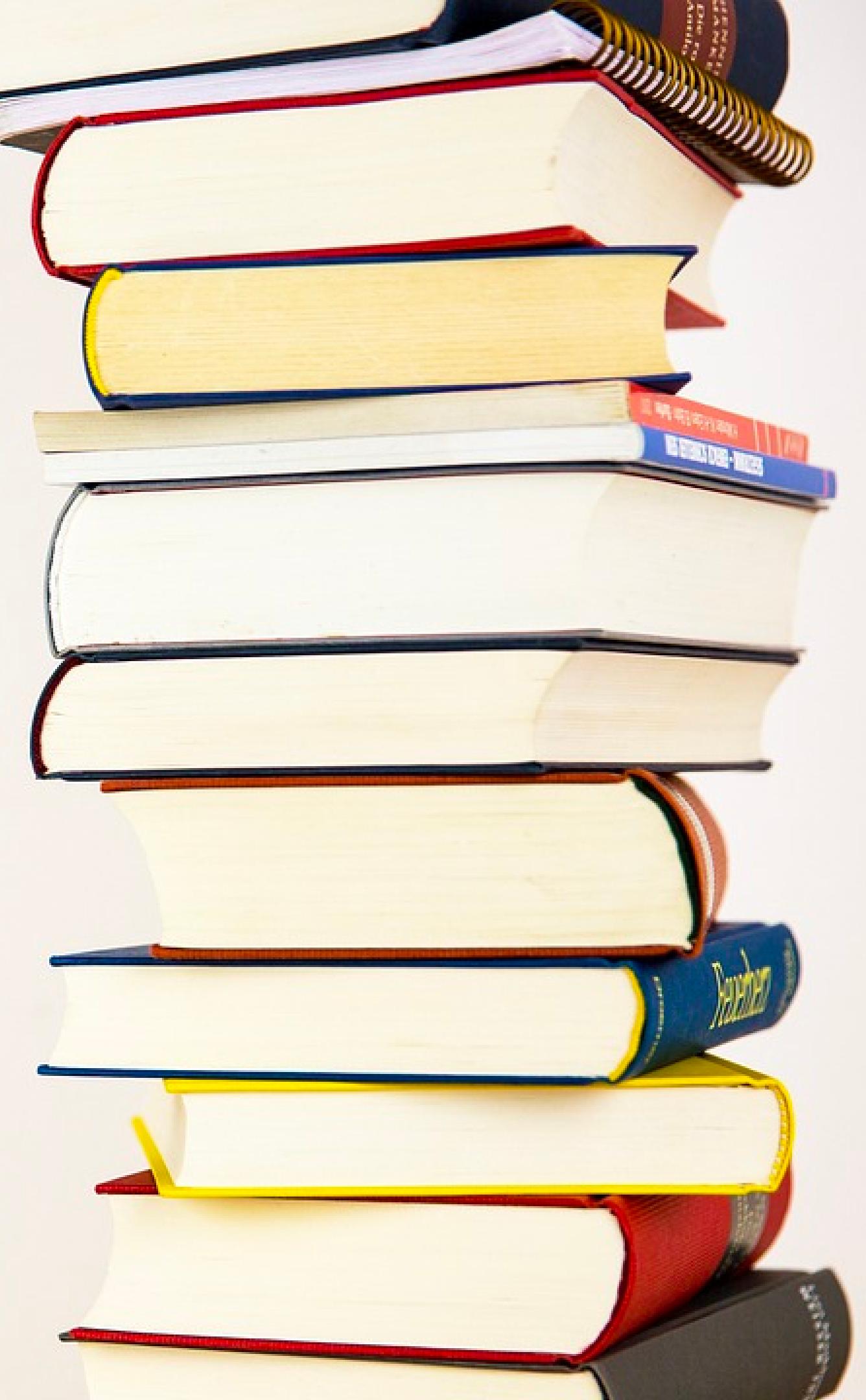


# GOODREADS DATASET PRELIMINARY ANALYSIS

BY QUINCI BIRKER



# Table of Contents

- Introduction
- Problem & Solution
- Impact of the Solution
- Dataset Introduction
- Next Steps...





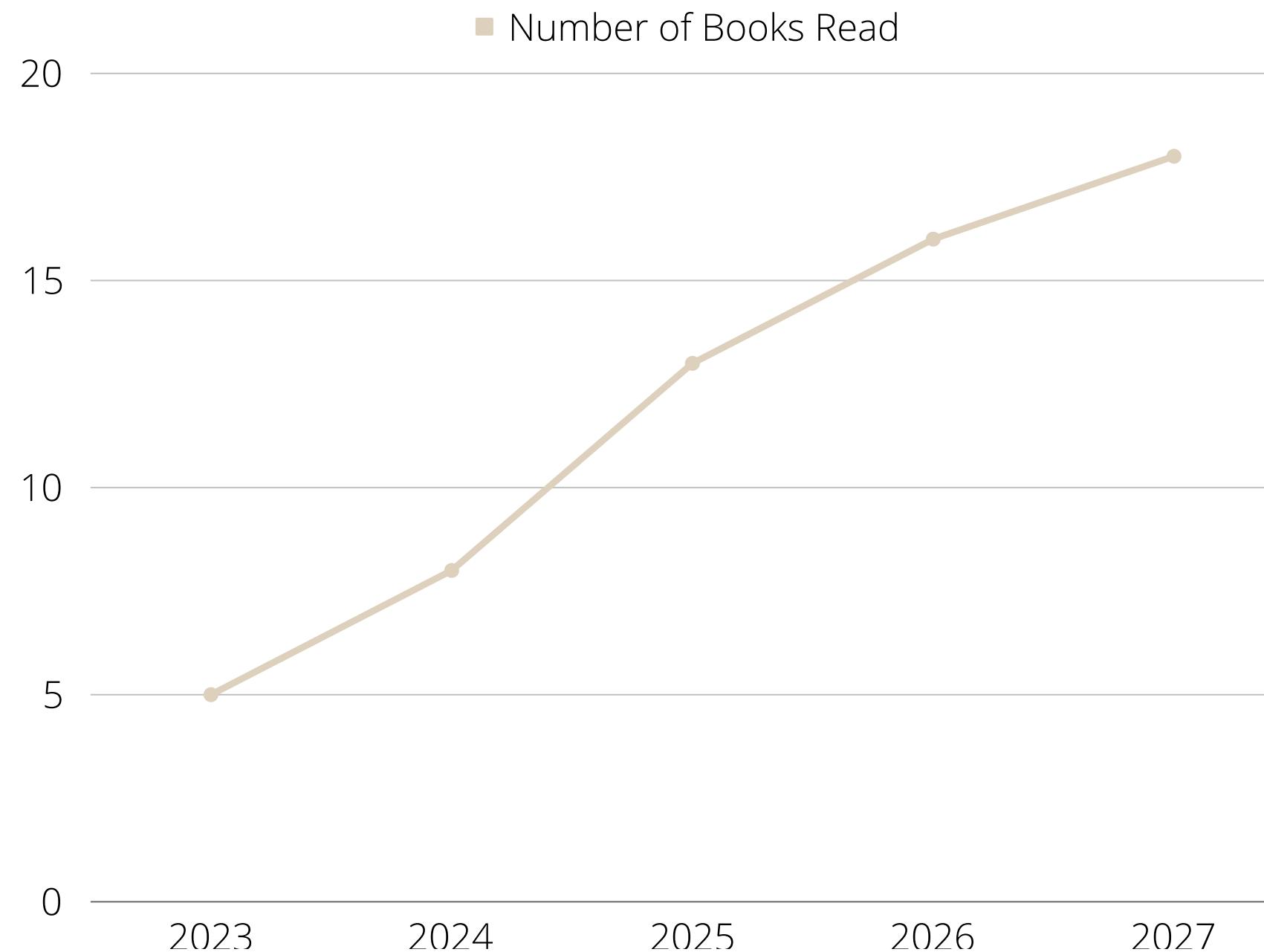
# Introduction

- Goodreads is one of the most popular book catalogue websites in North America.
- Updates:
  - Explored the dataset
  - Begun cleaning the data
  - Created visuals
  - Formulated ideas for my project



# Problem & Solution

- Problem: If you pick up a book you don't like, it can put you off reading for a long time.
- Solution: personalized book recommendations based on what you've enjoyed before.

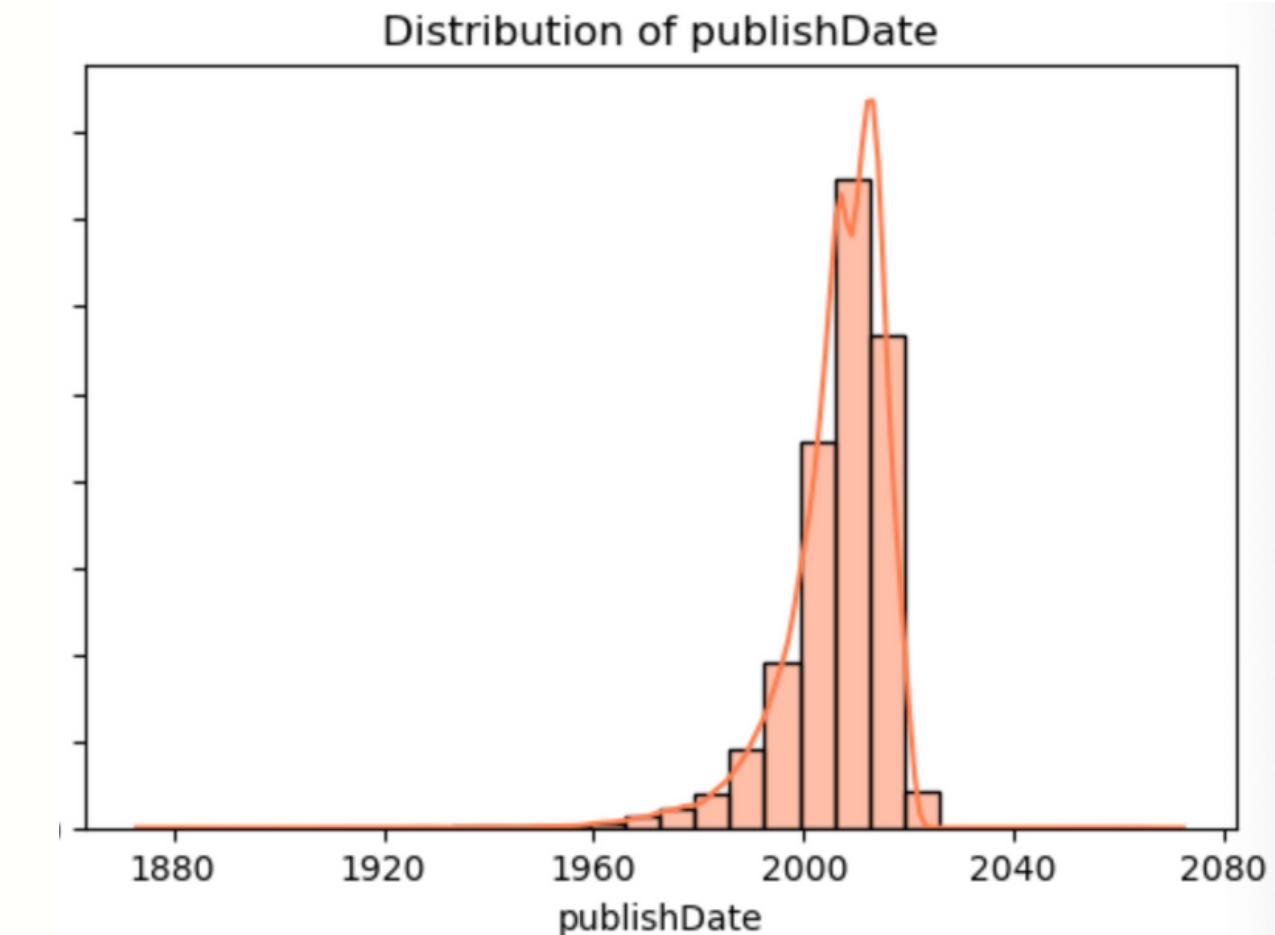
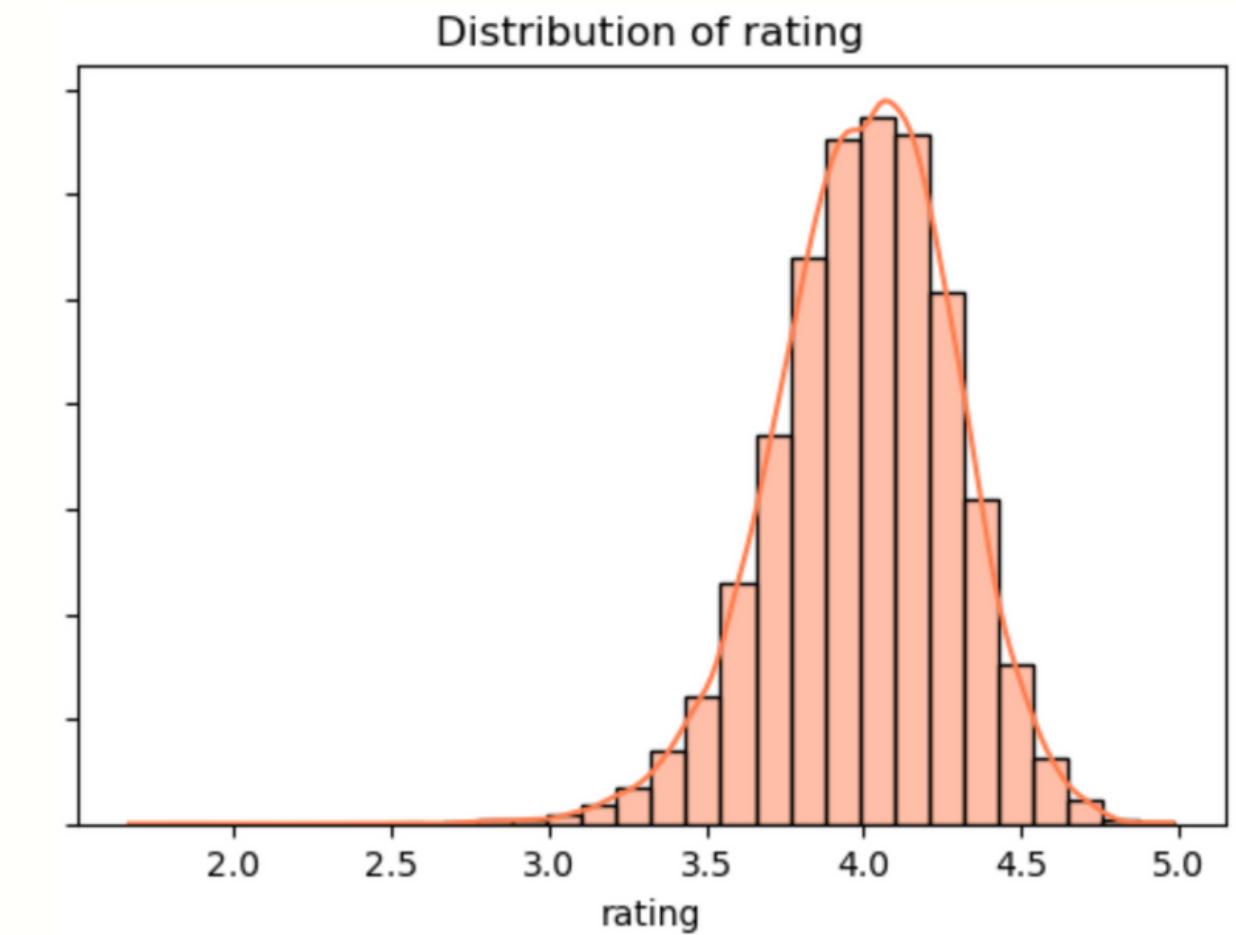


# Project Impact

- Make reading more popular
- Increase sales for bookstores
- Decrease screen time and increase reading time

# Data Introduction

- 52,400 rows and 25 columns
- Last updated Nov 2020
- Target variable: Rating Score (1 to 5)
- Useful variables:
  - Series
  - Author
  - Description
  - Genres
  - Publish Date
  - Number of Ratings



# Next Steps...

**Cleaning the Data**

**Formulating Hypotheses**

**Finding models suitable this project**

