5.0 ★★★★★

SPRINT 1

PREDICTING BOOK RATINGS

SABRINA RUIZ

OVERVIEW & PROBLEM STATEMENT

1. I work at an elementary school, many kids do not like reading
2. there are so many books in the world and online!
3. How to choose? RATINGS
4. Amazon Book reviews have lots of data and “words”
5. NLP fun!

“Can we use customer book review comments to predict the book’s rating score?”

PROPOSED VISION

1. **EDA**
   - clean & merge two files

2. **NLP**
   - text embedding
   - sediment classifying

3. **MODEL**
   - KNN?
   - train model

IMPACT

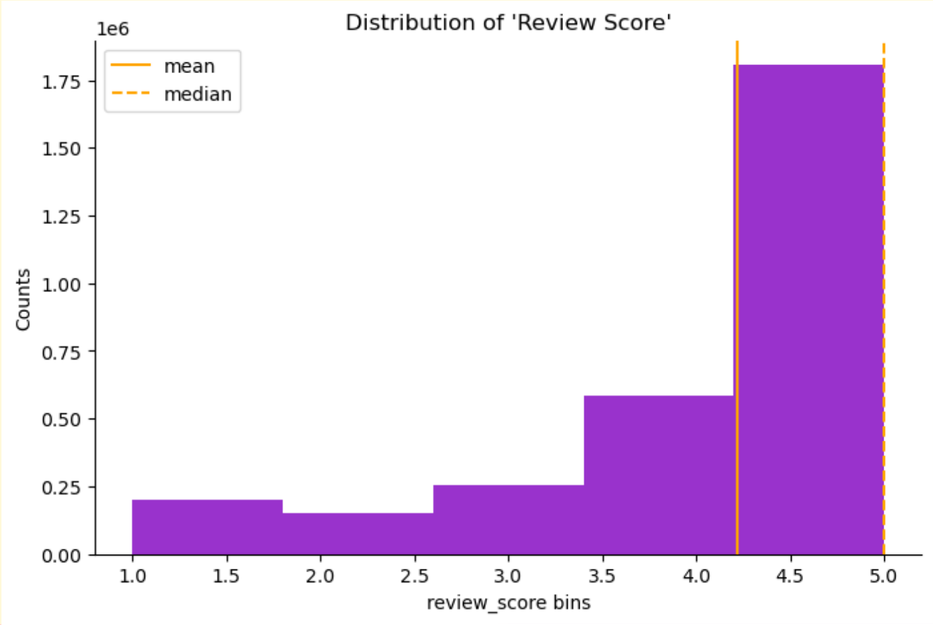
1. Enhanced book recommendations

2. Time Savings and Improved Reading Experience:

3. Universal Design and Educational Application

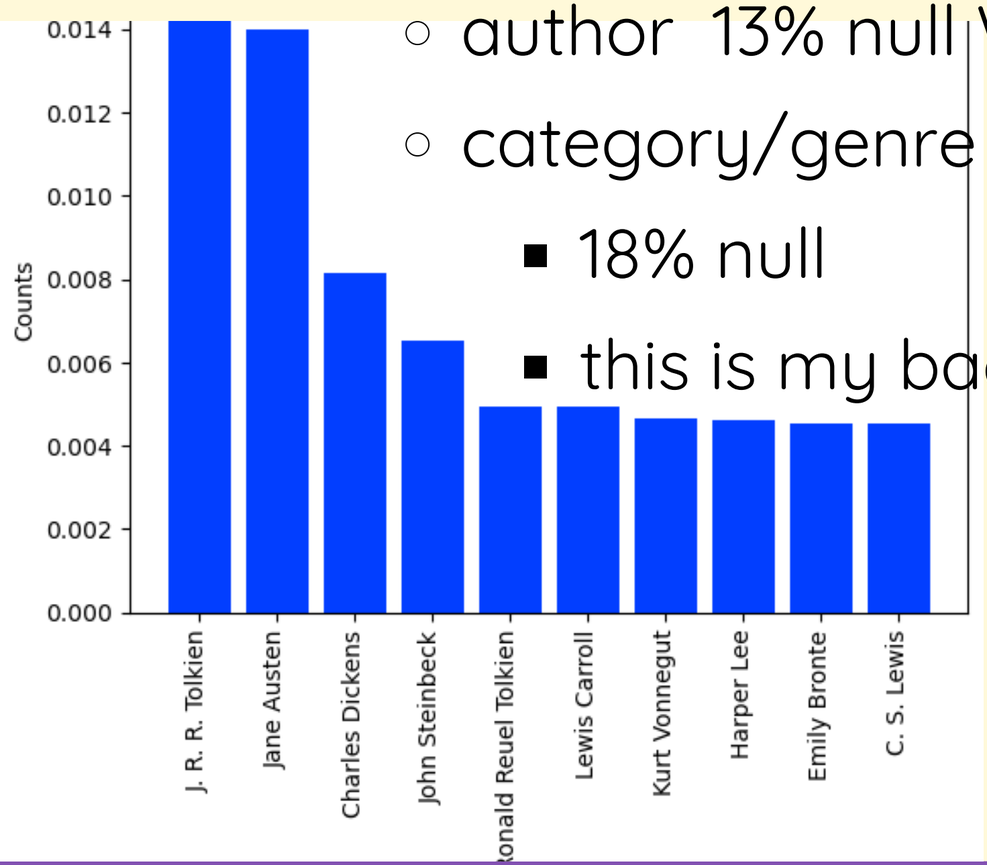
“BOOKS REVIEW”

- **3 mil. (reviews) x 4(10) features**
  - book title (212,404)
  - review_score (mean = 4.2)
  - review (obj) 0.000267 null



“BOOK DATA”

- **212,404 (books) x 2(10) features**
  - book title (212,404)
  - author 13% null WIP
  - category/genre (10,883)
    - 18% null
    - this is my backup



NEXT STEPS

1. RE-EVALUATE DATA

go back and add, move forward and delete, featureengineering?

2. GETTING STARTED ON TEXT EMBEDDING

estimated 90% of my capstone work going forward will be on this step and the next

3. SENTIMENT CLASSIFYING

along with text embedding, this is time consuming and iterative