

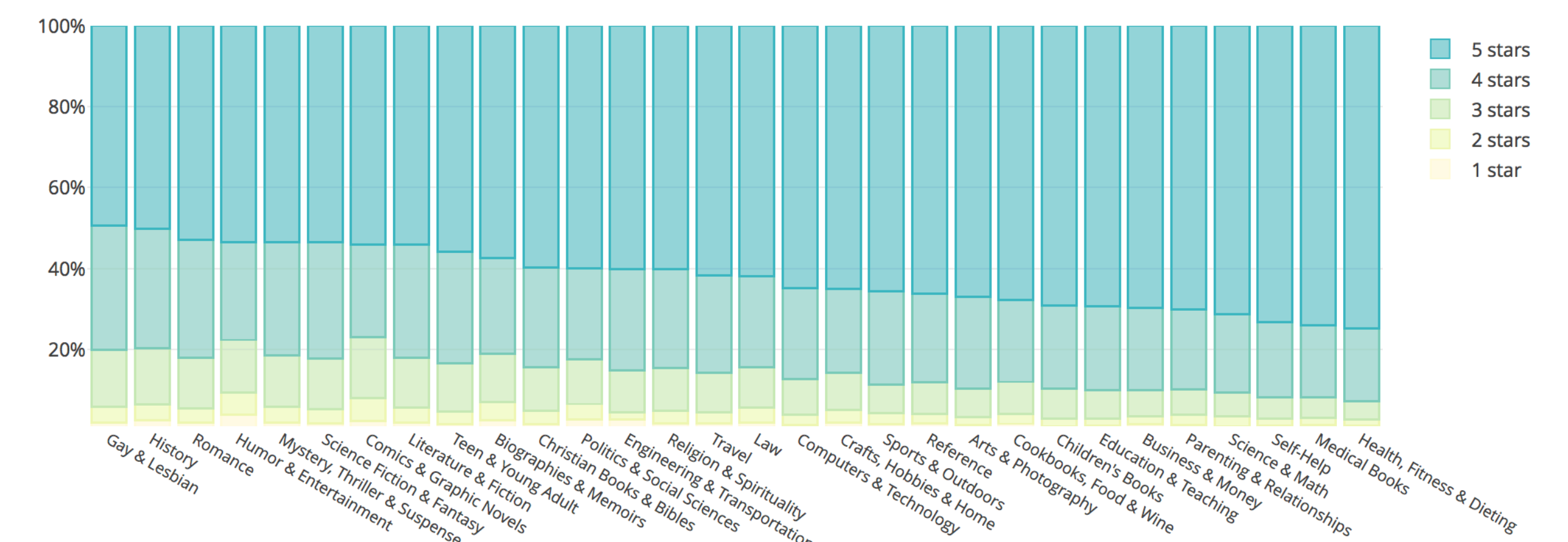
Introduction



- Reviews and ratings are everything for Amazon
- Based on the Amazon kindle store product and review datasets
- Mainly focus on the basic factors and relation between reviews and rating scores
- Data Analysis, machine learning models and natural language processing are utilised

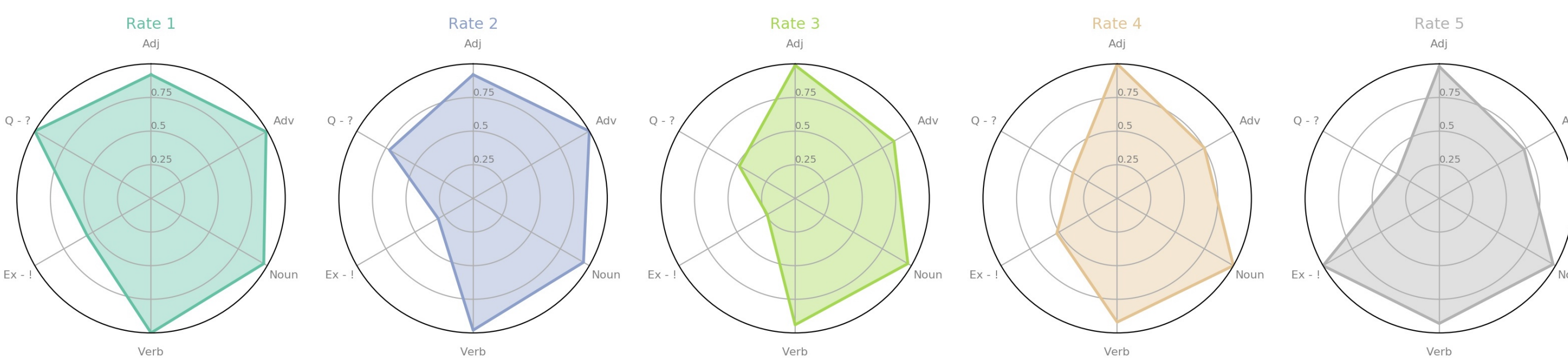
Rating for Category

- Scientific or health-related category wins
- Personal topics are not universally welcomed
- People love knowledge good for themselves
- Everyone define life and love differently
- History is artificial and not purely objective

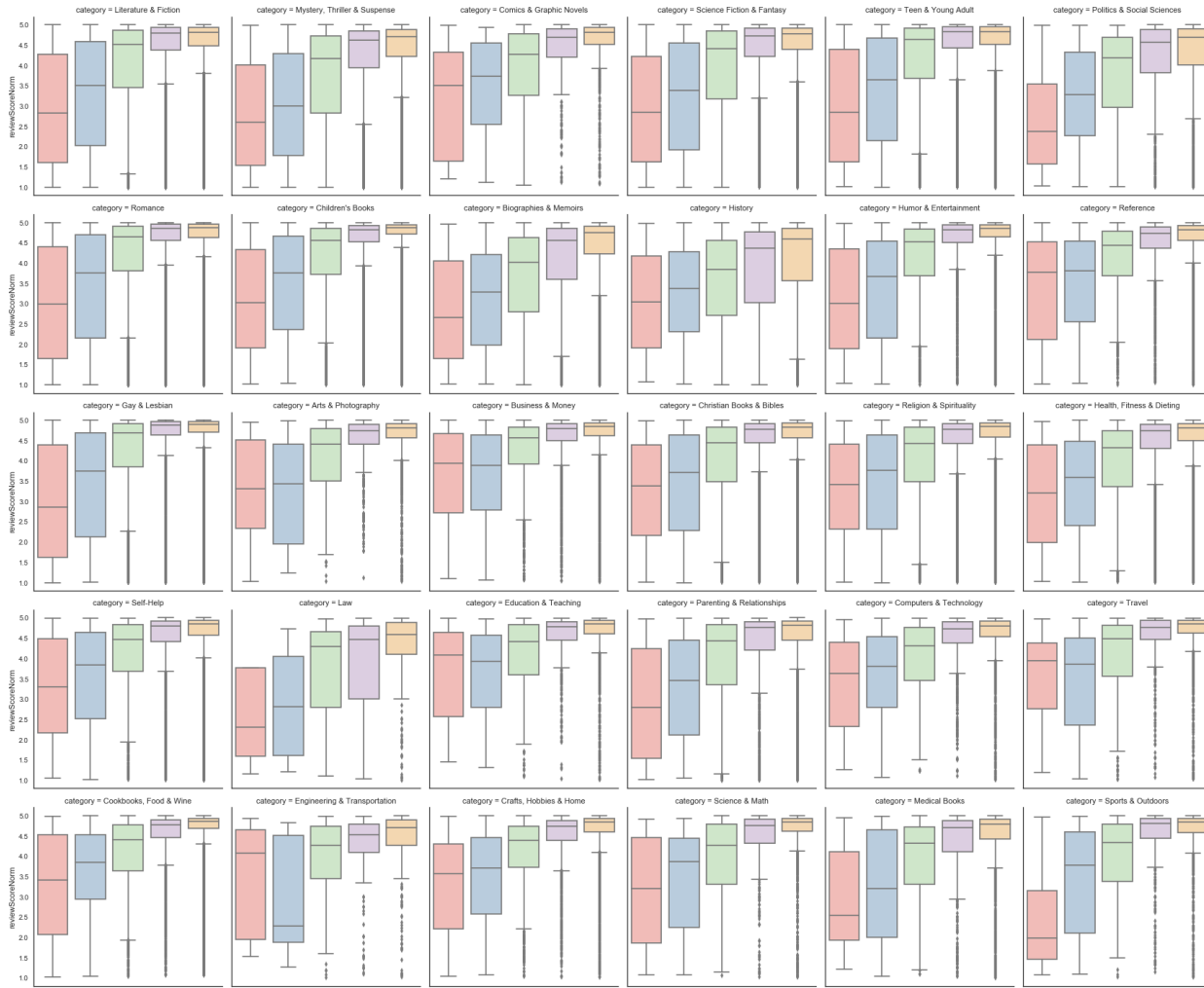


Look Deep into Review

Statistics of Punctuation and Part Of Speech

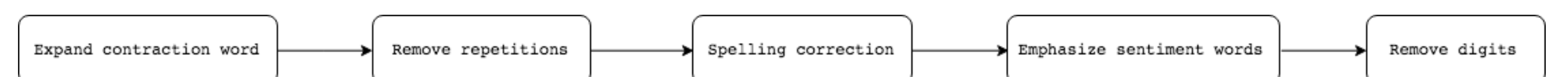


Sentimental score



- High score reviewers tend to write nice review
- One-star readers don't always say something mean or negative but maybe productive suggestions.

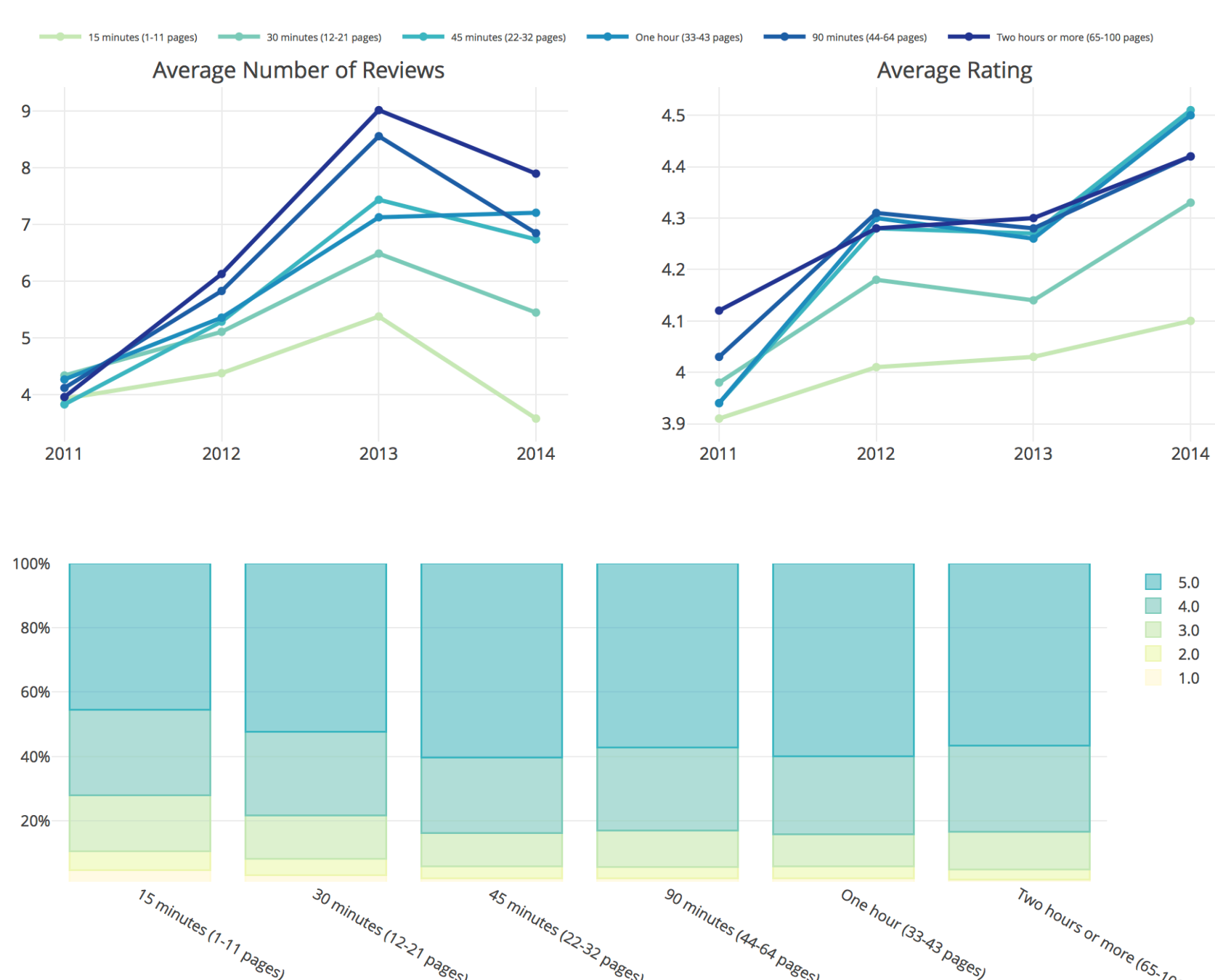
Build Rating Predicting Model



- Use fastText text classification
- Preprocess text with pipeline above
- Positive Group (Rate 5) v.s. Negative Group (Rate 1-4)
- Accuracy only reaches 75%
- Why? Low rating doesn't imply negative review

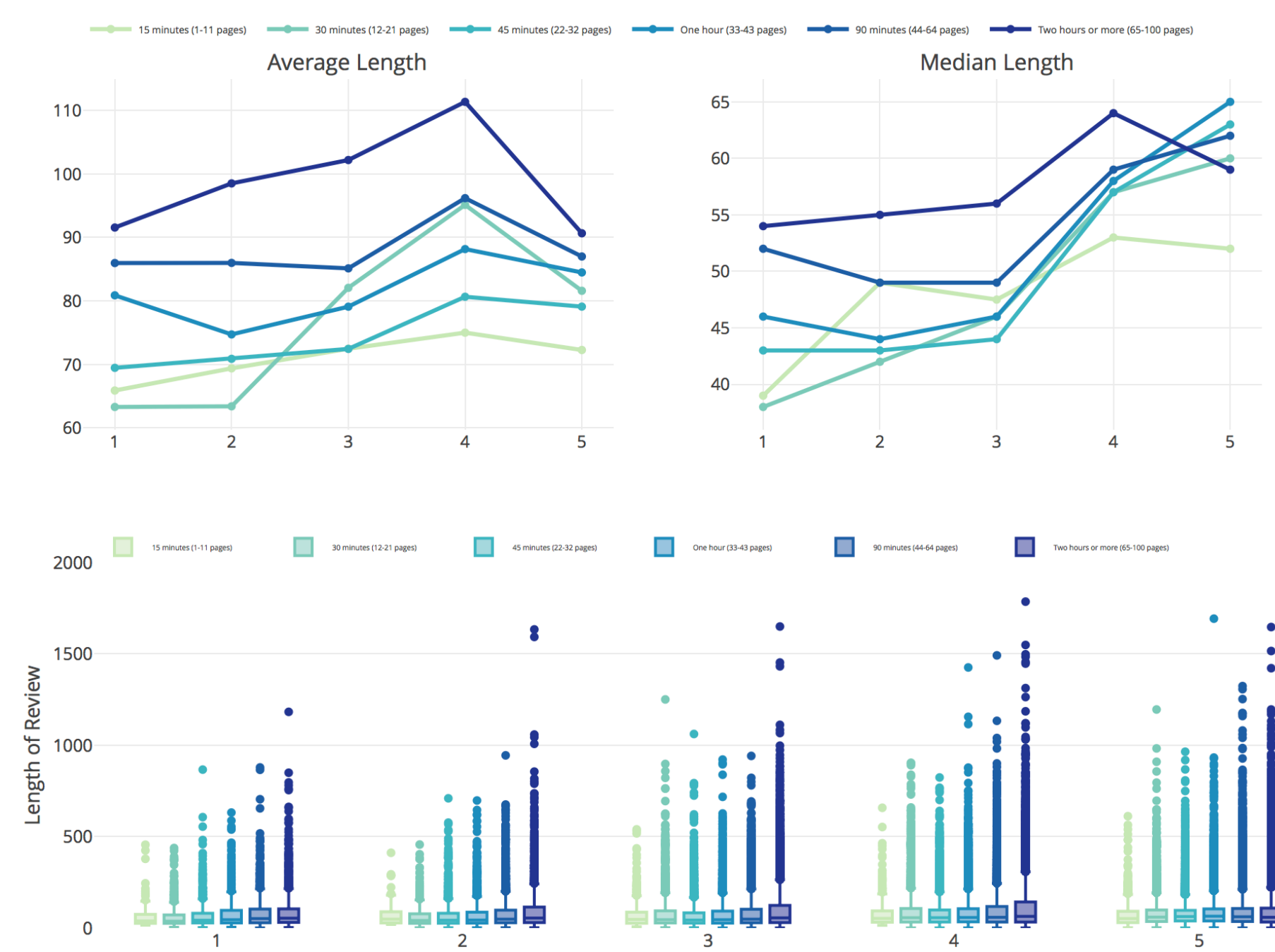
Short Reads Q&A

Reader loves short reads?



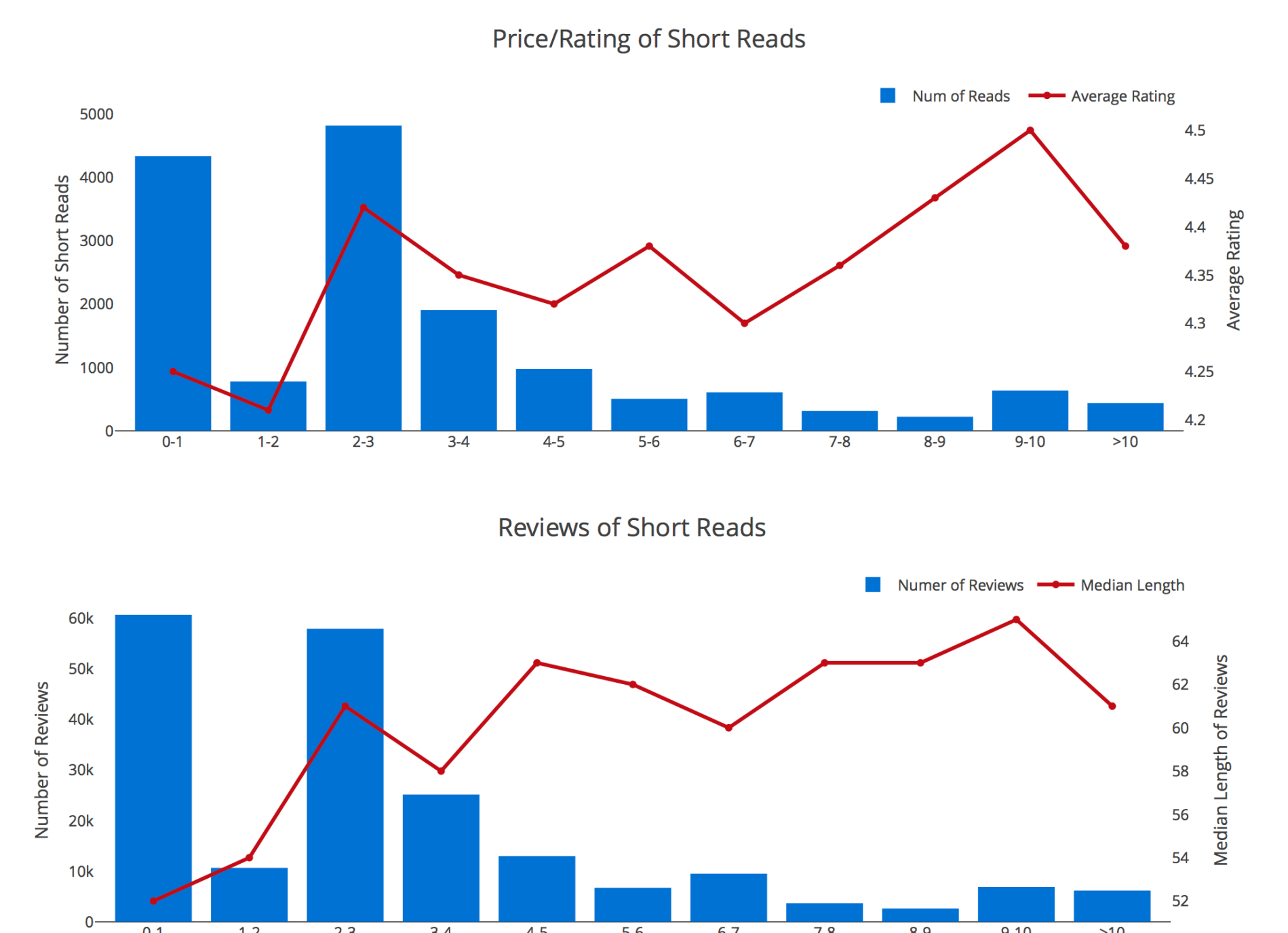
Not really!

Short reads, short review?



Sounds right!

Price plays an important role?



Tricky to answer!

Technology