

20211027 Lecture

Wednesday, October 27, 2021 4:28 PM

- Midterm

- Part of the goal was to show you a realistic experience working with data
- You can use a model to learn features
- Kaggle is allowing non-integer values
- Look at what products have way more ratings than other
- What are the least rated products
 - If there is only 5 ratings, there is not much there for you
- Kindest reviewers...
- Harshest reviewers...
- The way people write reviews is enough to identify the person
- Average of the 25 most rated products
- Average of the 25 least rated products
- Average reviews per hour isn't available
 - You have day information
- What is the average per month
- Are there years or release more?
- Length of the review
- Based on this exploration
- Added some of this data I thought was relevant
- `swifter.apply()`
 - Doesn't work on strings, text
 - Has a nice completion in his score?

- Percentage of which its done
- Summary length
- Top words that were unique per score group
- Review length per score
-
- Couple things that he tried: