

EDA Findings and Approach

1) Bucketwise average ratings and number of ratings for years_since_submission_on_review_date

Its observed that maximum ratings and number of reviews are logged in the first year of the recipe submission. The ratings and number of reviews decreases as the recipe gets old

2) Bucketwise average ratings and number of ratings for minutes

Majority of the reviews and ratings are received for the video's starting from 15 mins to 300 mins. It can be assumed that people prefer descriptive but not too lengthy videos. The reviews has drastically reduced for video's more than 300 mins.

3) Bucketwise average ratings and number of ratings for n-steps

Number of reviews are high for recipe which has 8 to 29 steps. But the ratings are more for recipe with less than 6 steps

4) Bucketwise average ratings and number of ratings for n_ingredients

Number of reviews are higher for 6 to 9 ingredients. However, the ratings are not much effected by the ingredients

5) Preparation, time-to-make, course and dietary are few of the most rated tags and with maximum user ratings

6) Lamb-sheep-main-dish is the dish with no tags or ratings

What can be done to improve the recommendation?

- The count and percentage of positive, negative reviews can be used for better recommendation results.
- A more sophisticated formula like the Bayesian average can be used for calculating the popularity score, as the used formula is biased towards the number of votes.