RECIPE RECOMMENDER ASSIGNMENT EDA

By Binod Kumar Nahak

Problem Statement

Step into the shoes of an ML engineer working at food.com. Your job is to design a recommender system to recommend recipes to users based on their choice and the current recipe they are looking at.

The recommendation engine is a way to increase the website's user engagement. If a user is shown relevant recipes, they are more likely to spend more time on your site reading about recipes. Higher user engagement will likely result in more business opportunities like collaborations, promotions, etc.

The performance of a recommendation engine will significantly impact the revenue your recipe site can generate.

Designing a recommender from scratch is a time-consuming task. In this assignment, you are expected to explore the data and create features that will be used to build the recommender.

Data Available

Raw_Recipes file

1. It contain all the details of recipe.. Each row describes recipe.

2. It contain details like name, minutes to prepare the recipe, contributor_id, date when recipe submitted, tags for recipe, nutrition values, steps to prepare, ingredients and description.

Raw_Interaction file

1. It contain all the details of rating and review.

2. Rating includes individual recipe ratings and individual user details and user review.

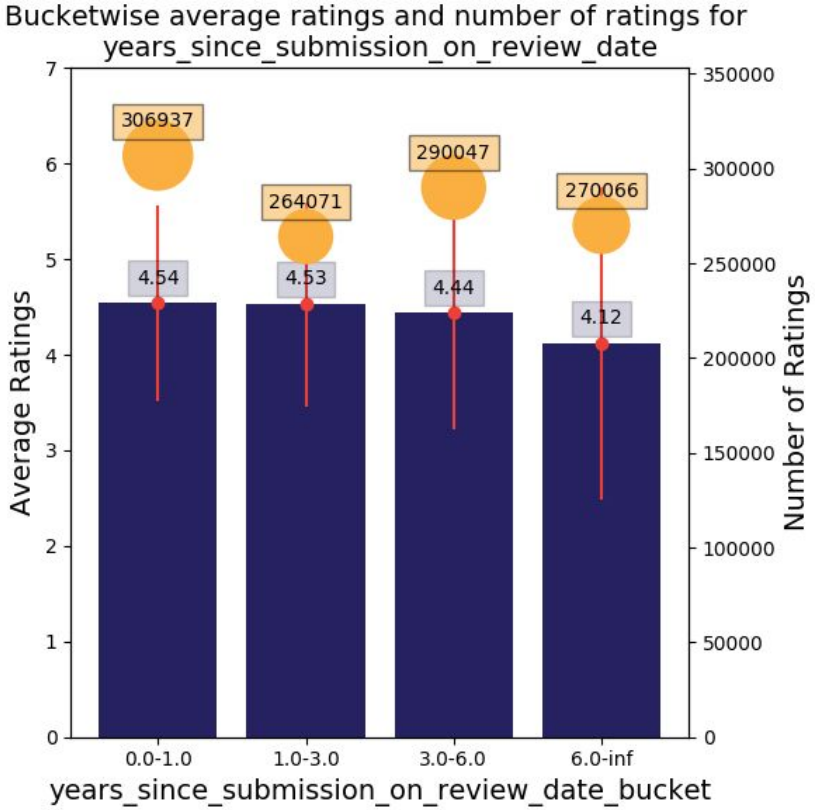
EDA

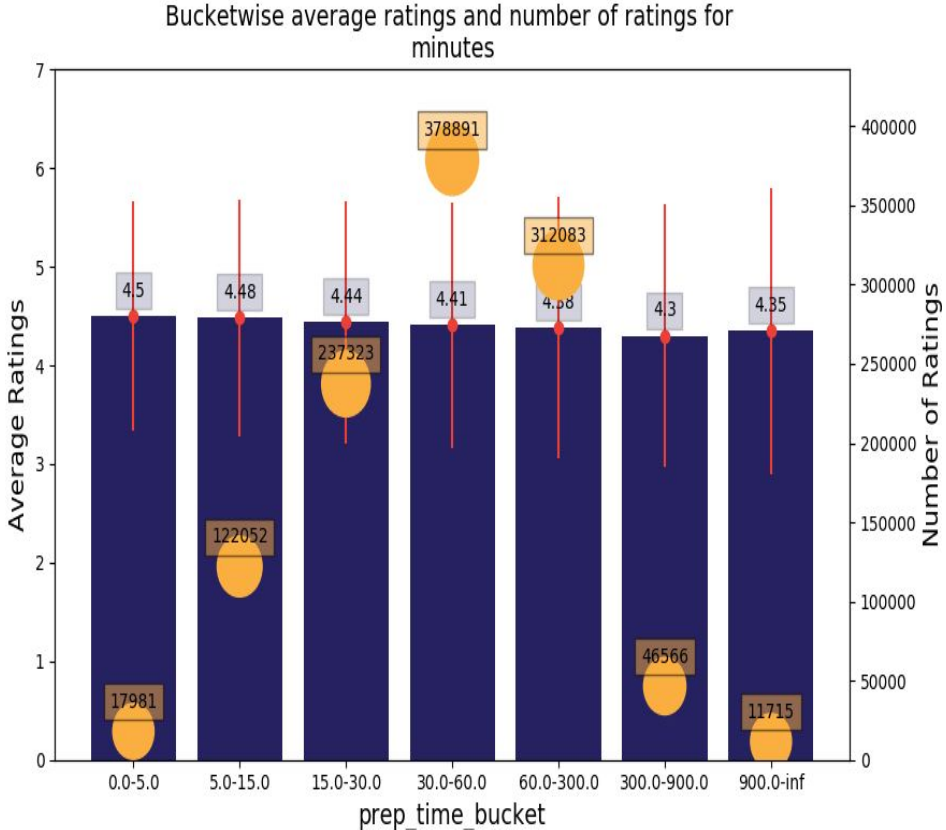
1. Since we have received clean file. There is no requirement of missing values and outlier treatment.
2. Special attention provided for nutrition and tags column for extracting more information.
3. Extracting individual features from the nutrition column.
4. Standardize the nutrition values.
5. Converting the tags column from a string to an array of string.

Feature Extraction

1. Created time based feature on review date and submitted date.
2. Numerical column processing like bucketing for minute column, calories other nutrition values.
3. Created user level features Like user ratings , user avg ratings,user avg prep times, user avg sugar per 100 cal .
4. Tag level feature like top 5 percentile frequent tags, top 5 highest rated tags, bottom 5 percentile highest rated tags.

Bucketing


Bucketwise average ratings and number of ratings for years_since_submission_on_review_date
306937
264071
290047
270066
4.54
4.53
4.44
4.12
Average Ratings
Number of Ratings
0.0-1.0
1.0-3.0
3.0-6.0
6.0-inf
years_since_submission_on_review_date_bucket


Bucketwise average ratings and number of ratings for minutes
17981
122052
237323
378891
312083
46566
11715
4.5
4.48
4.44
4.41
4.3
4.35
Average Ratings
Number of Ratings
0.0-5.0
5.0-15.0
15.0-30.0
30.0-60.0
60.0-300.0
300.0-900.0
900.0-inf
prep_time_bucket

Thank You