RECIPE RECOMMENDER ASSIGNMENT EDA

SUBMITTED BY:

PRIYA

PRIYANKA SORATE

SIMRANJEET KAUR

Problem Statement

- As an ML engineer at food.com, our objective is to develop a recommender system that offers recipe suggestions to users by considering their preferences and the recipe they are currently viewing.
- A well-functioning recommender system holds the potential to boost user engagement and create additional business prospects. The efficiency of the recommendation engine directly influences the revenue generated by the website.
- However, constructing a recommender system from the ground up requires a significant amount of time and effort. In this assignment, our task is to analyze the available data and generate relevant features to construct the recommender system.

TASK LIST

- Task 1: Read the data
- Task 2: Extract individual features from the nutrition column.
- Task 3: Standardize the nutrition values.
- Task 4: Convert the tags column from a string to an array of strings.
- Task 5: Read the second data file
- Task 6: Create time-based features.

Feature Extraction:

- The Assignment is focused on extracting features from the data to build a recipe recommender system.
- As we have observed, when the Spark compiler reads the nutrition column from the raw_recipes_df DataFrame, it is treated as a string column instead of an array of float values. However, each row in the nutrition column contains seven values representing different nutrition information. Our task is to extract these individual values and create seven separate columns named calories, total fat (PDV), sugar (PDV), sodium (PDV), protein (PDV), saturated fat (PDV), and carbohydrates (PDV).
- Make nutrition-per-100 calorie columns
- Creating nutrition-per-100 calorie columns: Like putting nutrition values on a level playing field, we standardize them to relative terms by considering 100 calories as the reference point.
- Naming convention: Original column name "total fat (PDV)", column name after conversion "total_fat_per_100_cal"

Exploratory Data Analysis:

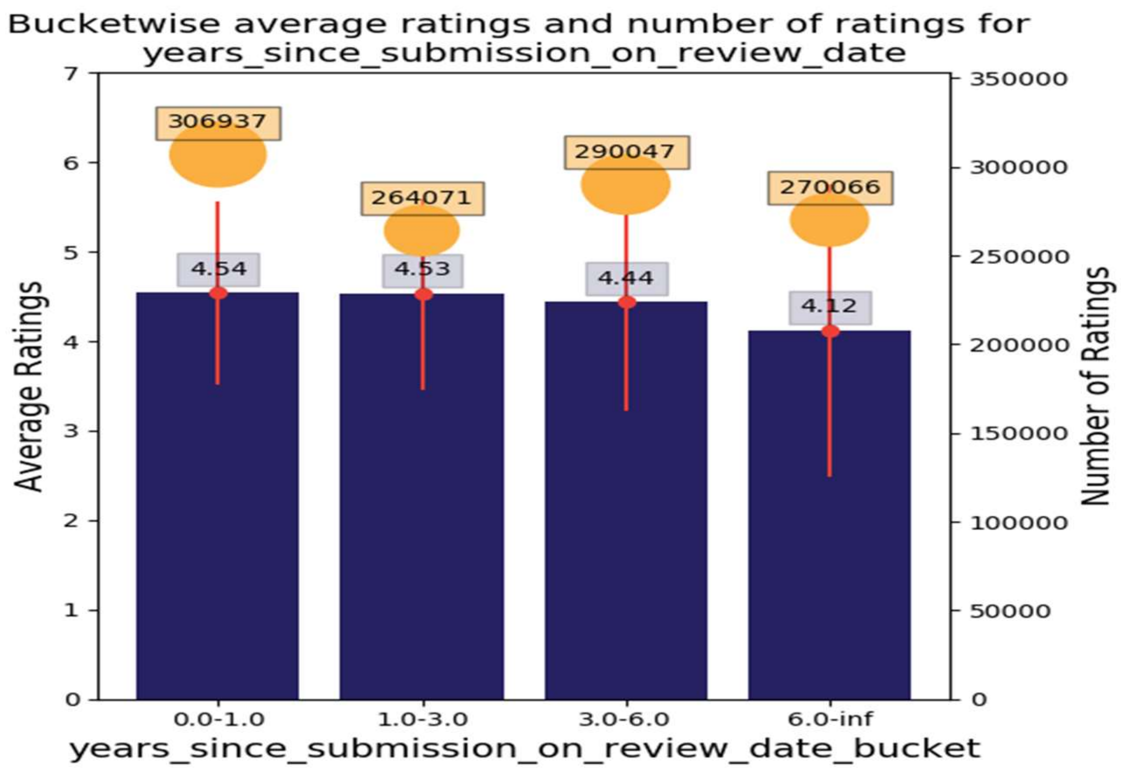
The first approach to EDA is to cover all possibilities for all columns.

Bucketing and Cleaning Numerical Features

1. years_since_submission_on_review_date

[Review Time Since Submission]

Recipes more than 6 years old are rated low

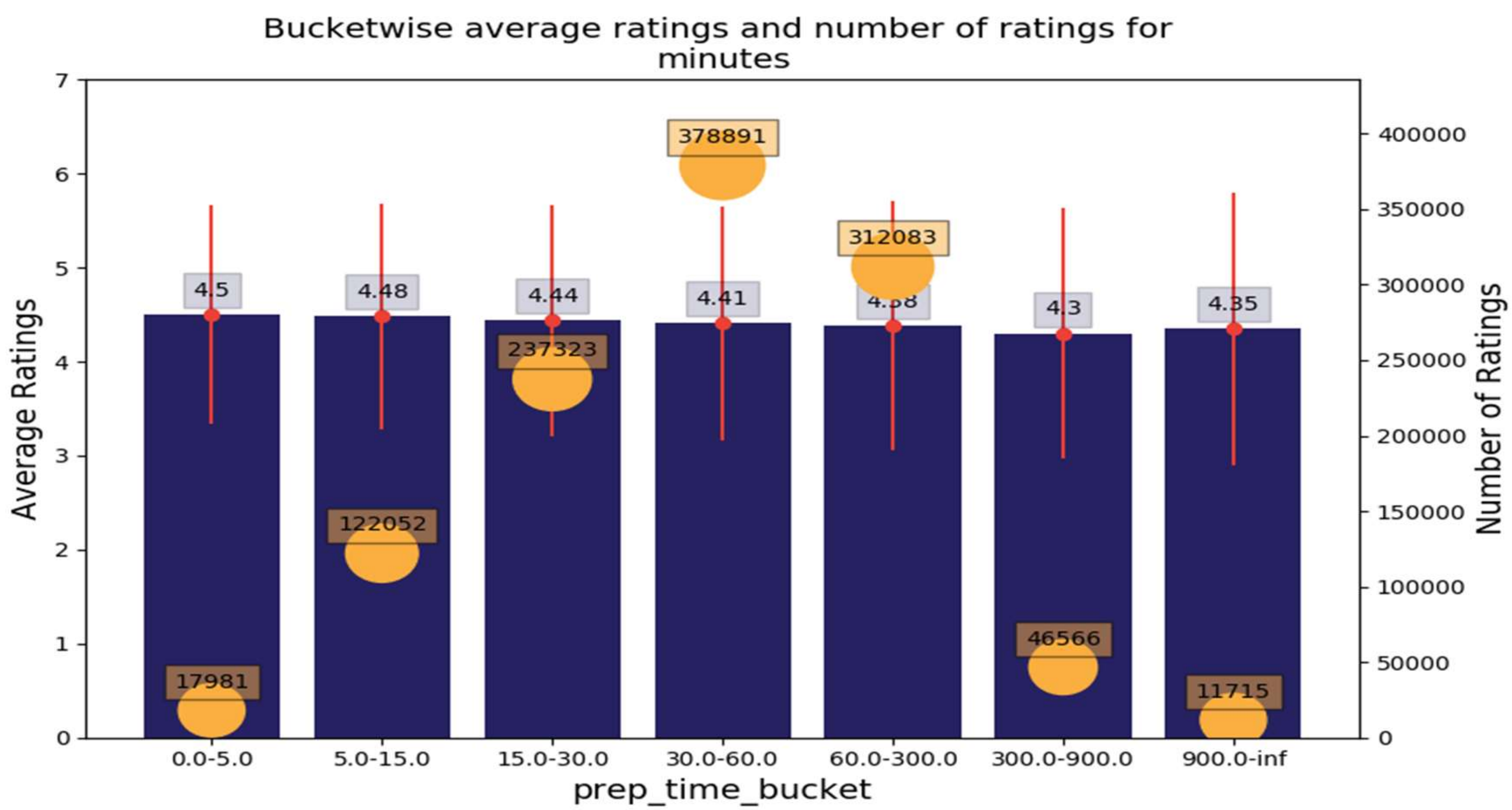


2. minutes

[prep time]

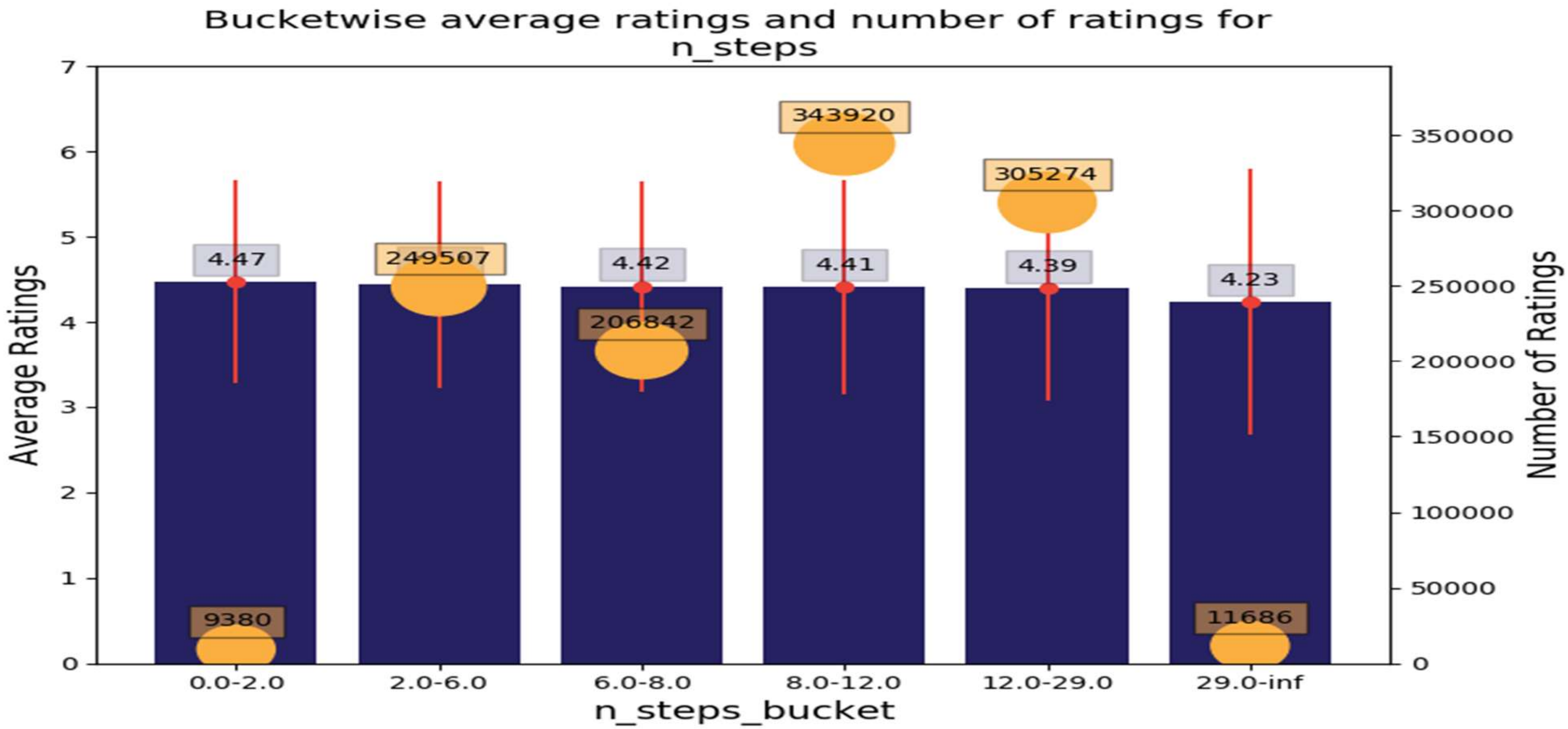
•Somewhat relevant

•Low prep time is preferred



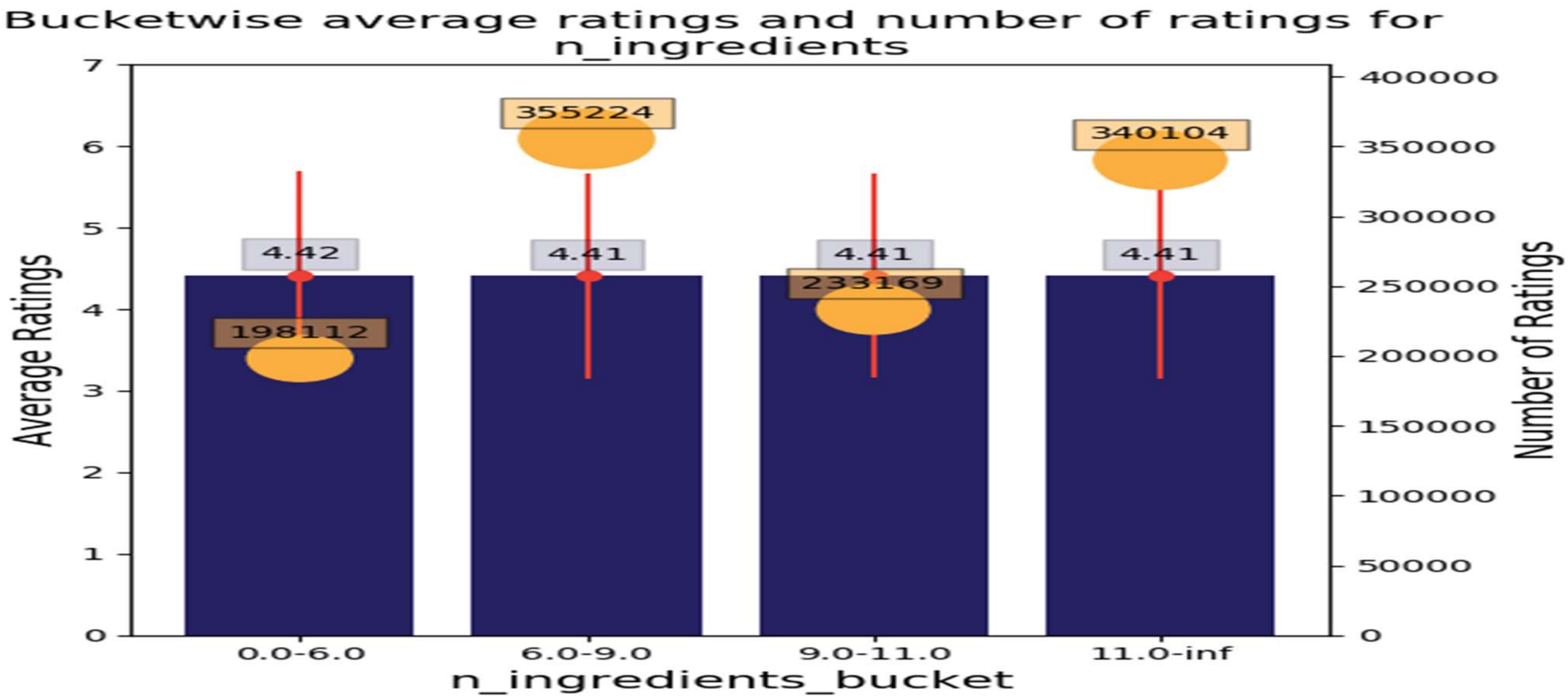
3. `n_steps`

•Clearly relevant
•Recipes with less than 2 steps are rated high
•Recipes with more than 29 steps are rated very low



4. n_ingredients

- Not relevant



5. NUTRITION COLUMNS

- ✓ calories - Calories per serving seems irrelevant
- ✓ fat (per 100 cal) - Calories per serving seems irrelevant
- ✓ sugar (per 100 cal) - Calories per serving seems irrelevant
- ✓ sodium (per 100 cal) - Calories per serving seems irrelevant
- ✓ protein (per 100 cal) - Calories per serving seems irrelevant
- ✓ sat. fat (per 100 cal) - Calories per serving seems irrelevant
- ✓ carbs (per 100 cal) - Calories per serving seems irrelevant

Adding user level average features

More Features:

high_ratings = 5 rating

- `user_avg_years_betwn_review_and_submission_high_ratings`
- `user_avg_prep_time_recipes_reviewed_high_ratings`
- `user_avg_n_steps_recipes_reviewed_high_ratings`
- `user_avg_n_ingredients_recipes_reviewed_high_ratings`

END CONCLUSION END

- In these notebooks we have performed EDA and feature extraction for the recipe recommendation.
- Designing an effective recommender system is vital for food.com to increase user engagement and revenue.
- Personalized recipe recommendations based on user preferences and the current recipe will keep users engaged on the website.
- Higher user engagement opens up opportunities for collaborations and promotions, leading to increased revenue.
- Exploring data and creating meaningful features lay the foundation for a successful recommendation engine.
- Algorithms like collaborative filtering and content-based filtering will be employed for accurate recommendations.
- Evaluating performance using metrics like precision, recall, and accuracy will guide improvements.
- Implementing a reliable recommender system benefits users and contributes to food.com's success.
- As an ML engineer, you have the opportunity to impact user engagement and revenue generation through this system.