BANA 277 CUSTOMER & SOCIAL ANALYTICS PROJECT

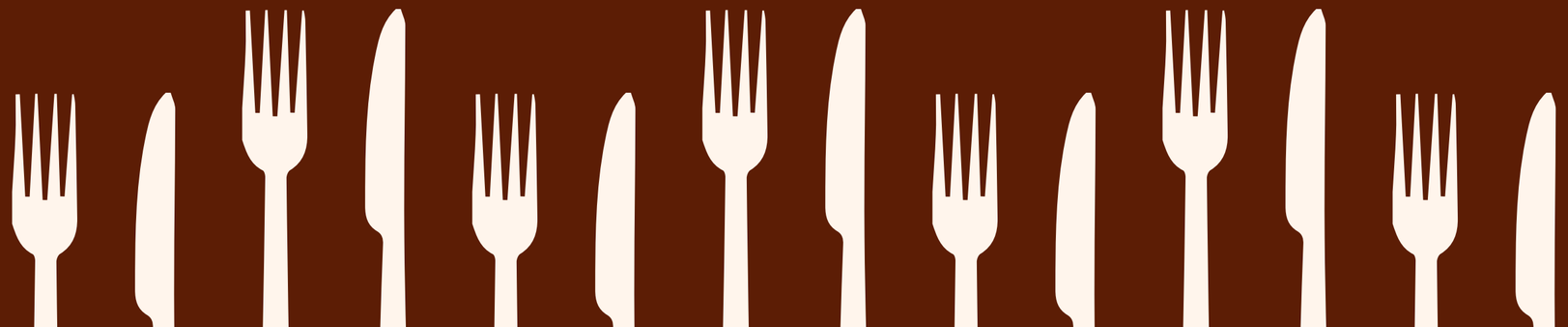
ANALYZING RECIPE SUCCESS

A Deep Dive into User Ratings

TEAM TRACE3B:

Vy Nguyen, Kento Morita, Qui Nguyen, Aria Zhou

Sahil Chennadi, Yuheshwar Kamakkapalayam Subramani



Introduction & Data Overview

Purpose Analyze recipe data on Kaggle to gain insights that can help recipe creators and platforms optimize recipes for higher ratings.

What recipe attributes influence ratings?

RAW_recipes.csv

- 231,637 rows, 12 columns
- Unique ids: user & recipe

How do user sentiment impact ratings?

RAW_interactions.csv

- 1,158,039 rows, 5 columns
- Unique id: user & recipe

Sentiment	Description
rating	User rating from 0 to 5.
review	User review in text.

Attribute	Description
name	The name of the recipe.
minutes	Minutes required to prepare the recipe.
tags	Food.com tags associated with the recipe (e.g. ‘course’)
nutrition	calories (#), total fat (PDV), sugar (PDV) , sodium (PDV) , protein (PDV) , saturated fat
n_steps	Number of steps required to prepare the recipe.
steps	The text description of each step in the recipe.
description	User-provided description for the recipe.
ingredients	A list of ingredients required for the recipe.
n_ingredients	The total number of ingredients in the recipe.

Recipe Attributes Analysis PROCESS

1. WRANGLING

- Outliers (z-score)
- Missing values

2. DATA TRANSFORMATION

- Submitted Date: Recalculate years since recipe publication (from 2019).
- Tag: Count the number of tags per recipe.
- Nutrition: Sum up the total nutrition values per recipe.
- Ingredients: Count the number of ingredients per recipe.

3. DATA PREPROCESSING

- Standardization
- Oversampling

4. ML MODELING

- Y => rating (0-5)
- Linear Regression, GLM, Neural Network, Random Forest

Recipe Attributes Analysis

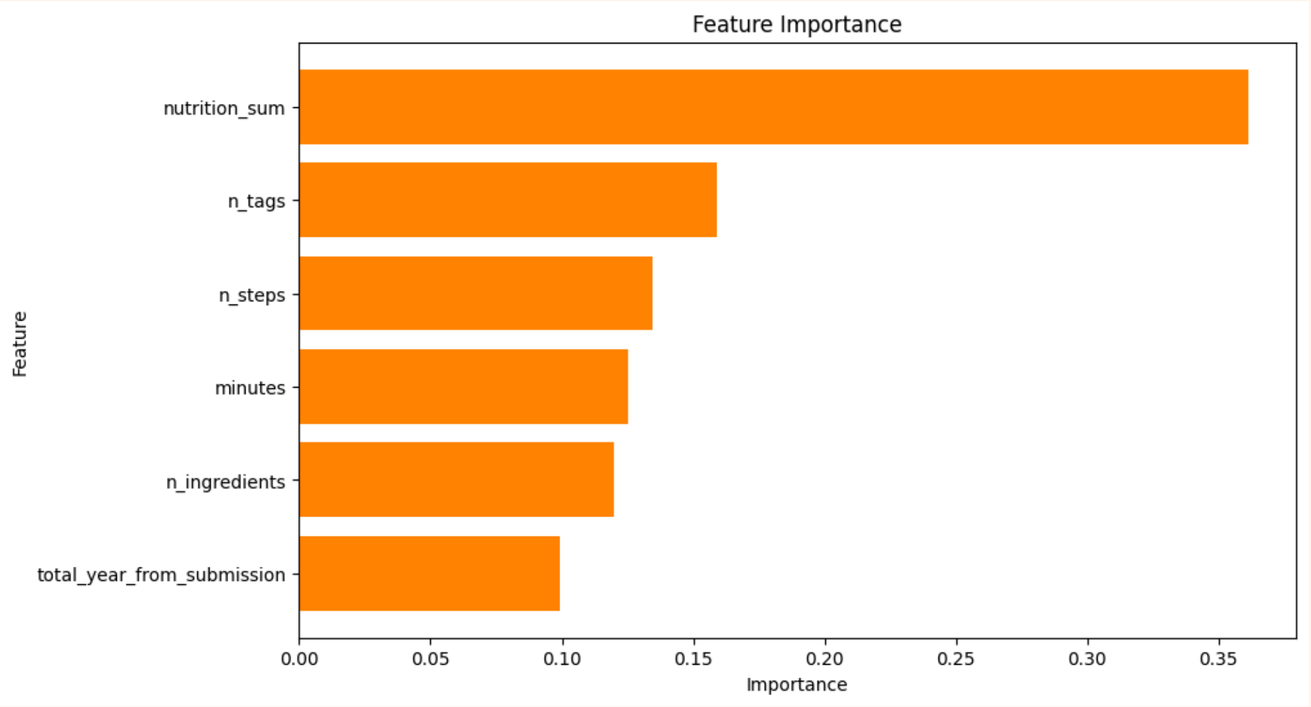
KEY FINDINGS

Models	MSE
Linear Regression	127.941
Generalized Linear Model	0.980
Neural Network	0.977
Random Forest	0.145

- Best model: Random Forest
- Stronger nonlinear relationship

Data was probably easy

BUT we are talented!



- Most important factors for rating: nutrition and number of tags and steps
- User loves easy, healthy food=> create nutrition focus tags

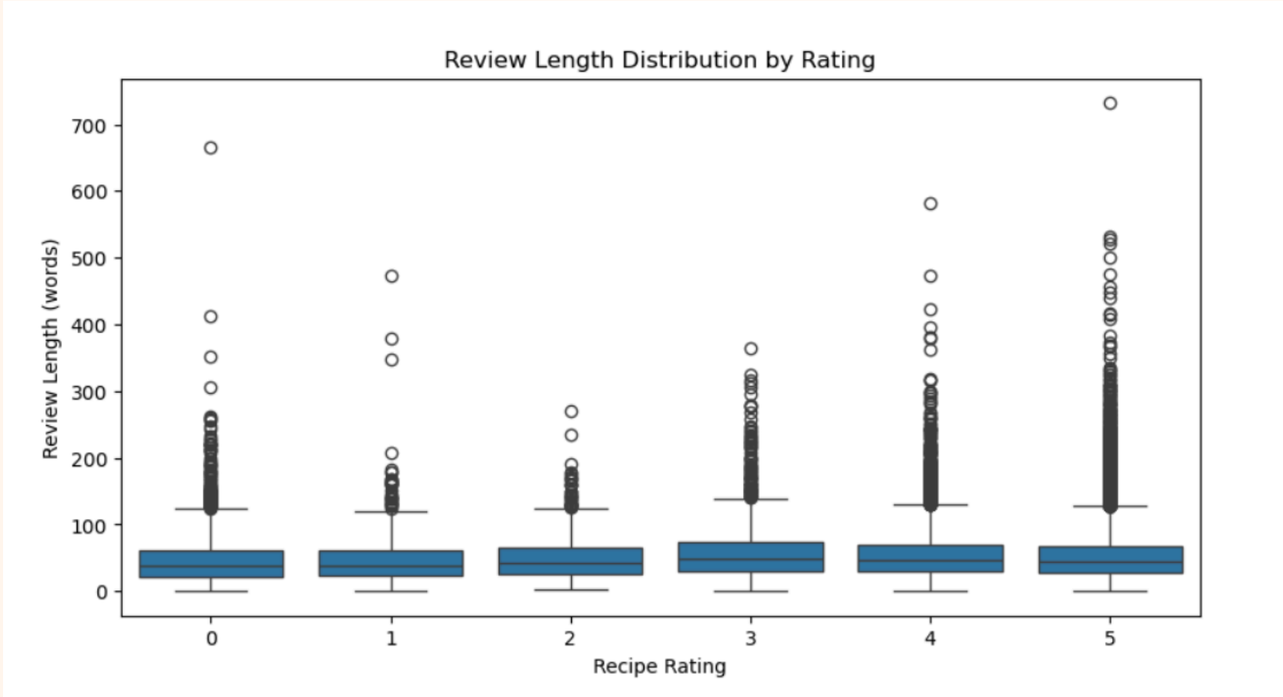
Review Sentiment Analysis MODELS

Random Forest Classifcation Report Across Ratings 1-5

Total Accuracy	0.93
Macro Average	0.96
Weighted Average	0.93
Precision Average	0.96

- The model is good at correctly predicting ratings with minimal misclassifications
- Best-performing attributes consistently correlate with positive sentiment and high-rated reviews.

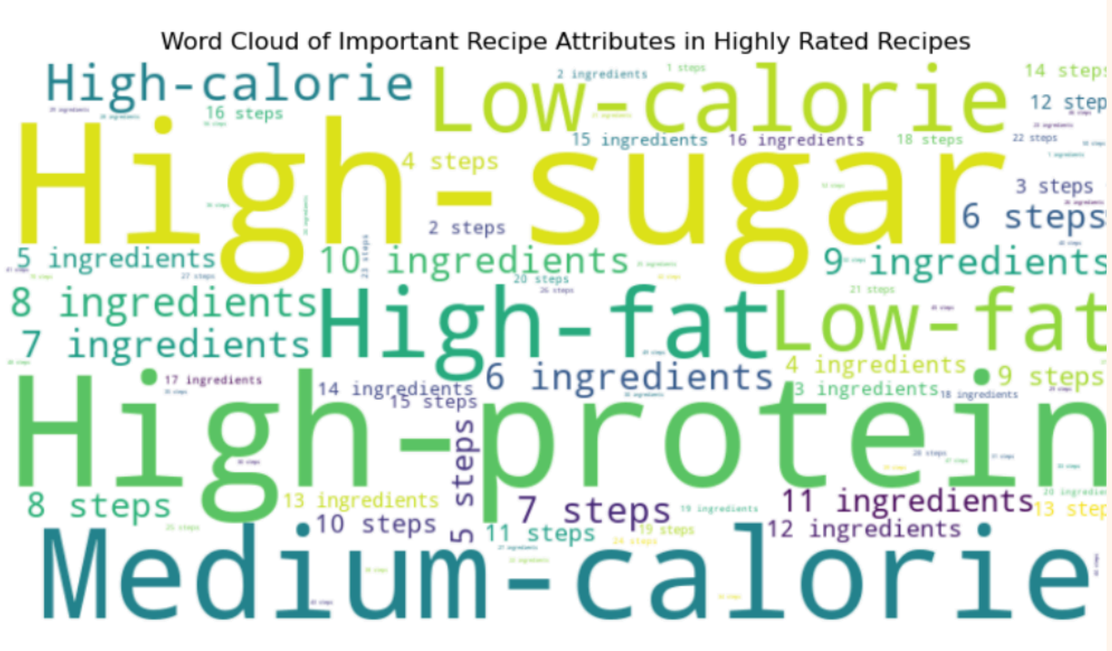
Review Length Distribution by Rating



- This analysis highlights the importance of encouraging user feedback to better understand their preferences
- Ratings in the 4-5 range have more outliers => Longer reviews may be associated with higher ratings

Review Sentiment Analysis

KEY FINDINGS



- High-protein recipes are commonly found among highly rated dishes
- Moderate ingredient and step counts are common among highly rated recipes

- "recipe," "made," "used," "added," "thank," "easy," "good," "great," "delicious," "flavor"
- Taste and ease of preparation are the biggest factors influencing positive feedback.

Recommendations

Recipe Creators and Platforms should:

Focus on Taste and Ease

- Positive language: "delicious," "easy," and "flavorful" in description
- 7-9 ingredients and 5-8 steps of preparation

Leverage Nutritional Appeal

- Highlight moderate caloric and fat content
- Avoid overemphasizing sodium or sugar levels

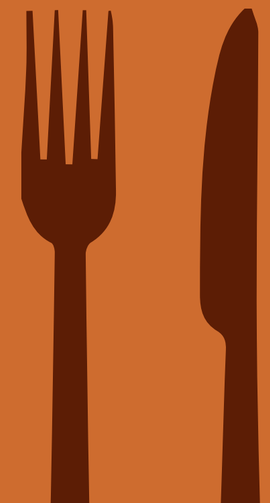
Encourage Detailed Reviews

- Prompt users to leave detailed feedback, as longer reviews often indicate stronger opinions
- Identify and address common pain points in low-rated recipes

Optimize Recipe Presentation

- Use tags effectively to categorize recipes (e.g. “low-carb”, “30-minute meal”)
- Ensure recipe steps are clear and easy to follow

Conclusion & Next Steps



Taste, ease of preparation, and user experience drive high ratings more than recipe complexity or nutritional content

Scaling our insights, futher analysis can be conducted:

Incorporate User Demographics

- How age, location, or dietary preferences influence ratings?

Integrate Advanced NLP Techniques

- Extract deeper insights from user reviews

Conduct A/B Testing

- Test different recipe formats (e.g, varying ingredient counts, step complexity) to validate findings.