RECIPE RECOMMENDATION – SUMMARY

This analysis is done to recommend recipes to users based on their choices and ratings provided by them for the current recipe they are looking at.

We have to perform EDA, including text analysis, on recipe dataset and interaction dataset.

GOAL: To know the factors that leads to both successful and unsuccessful recipes and then create a bucket of the categories required.

THE FOLLOWING STEPS ARE USED:

(i) Reading the recipe dataset and ensuring that all the categories are of appropriate data type.

(ii) Extraction of individual features into seven columns.

(iii) Standardize the nutrition column and convert the nutritional values to per 100 calories.

(iv)Converting tags column from a string to an array of strings.

(v) Reading the second file and doing feature selection to capture the time passed between one review and the date on which the recipe was submitted.

(vi)Converting numerical columns into categorical columns.

(vii) After creating buckets, check the average ratings of each bucket.

The following are the conclusions drawn from the analysis:

Years since submission on review date
[Review Time Since Submission]

- Recipes more than 6 years old are rated low
- The recent recipes are rated more

Minutes
[Preparation Time]

- Somewhat relevant
- Low prep time is preferred

Steps

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

Ingredients

- Not relevant

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

High Ratings

- ***5 rating***