

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 you will be working with

You will be working with the two CSV files linked below.

```
# raw recipe data
s3a://raw-recipes-clean-upgrad/RAW_recipes_cleaned.csv

# raw ratings data
s3a://raw-interactions-upgrad/RAW_interactions_cleaned.csv
```

Alternatively, you can download the files from the links below:

```
# raw recipe data

https://raw-recipes-clean-upgrad.s3.amazonaws.com/RAW_recipes_cleaned.csv

# raw ratings data
https://raw-interactions-upgrad.s3.amazonaws.com/RAW_interactions_cleaned.csv
```

The first file is the Raw_recipes.csv file. It contains all the recipe-related information. Each row in this file describes a recipe. The fields available in the data are described below.

Field Name	Field Description	Data type
name	Recipe name	String
id	Recipe ID	Integer
minutes	Time (in minutes) it takes to prepare recipe	Integer
contributor_id	Unique identifier for the user who submitted this recipe	Integer
submitted	Date recipe was submitted	Date
tags	Food.com tags for recipe	Array of strings
nutrition	An array containing seven nutrition values, appearing in the same order in each row. The values in the array represent the following nutrition information. (calories (#), total fat (PDV), sugar (PDV) , sodium (PDV) , protein (PDV) , saturated fat (PDV) , and carbohydrates (PDV))	Array of floats
n_steps	Number of steps in the recipe	integers
steps	Text for recipe steps, in order	String
description	Description of the recipe written by the person who submitted the recipe.	String

ingredients	List of ingredient names	Array of strings
n_ingredients	Number of ingredients	Integer

A few sample data points are shown in the image below.

name	id	minutes	contributor_id	submitted	tags	nutrition	n_steps	steps	description	ingredients	n_ingredients
arriba baked winter	137739	55	47892	9/16/05	['60-minutes-o	[51.5, 0.0, 13.0, 0.0, 2.0, 0.0, 4.0]	11	['make a choice a	autumn is my fa	['winter squash	7
a bit different breakf	31490	30	26278	6/17/02	['30-minutes-o	[173.4, 18.0, 0.0, 17.0, 22.0, 35.0, 1.4]	9	['preheat oven to	this recipe calls	['prepared pizz	6
all in the kitchen chil	112140	130	196586	2/25/05	['time-to-make	[269.8, 22.0, 32.0, 48.0, 39.0, 27.0, 5.0]	6	['brown ground b	this modified ve	['ground beef',	13
alouette potatoes	59389	45	68585	4/14/03	['60-minutes-o	[368.1, 17.0, 10.0, 2.0, 14.0, 8.0, 20.4]	11	['place potatoes	this is a super e	['spreadable ch	11
amish tomato ketchu	44061	190	41706	10/25/02	['weeknight', 't	[352.9, 1.0, 337.0, 23.0, 3.0, 0.0, 28.4]	5	['mix all ingredie	my dh's amish r	['tomato juice'	8
apple a day milk sha	5289	0	1533	12/6/99	['15-minutes-o	[160.2, 10.0, 55.0, 3.0, 9.0, 20.0, 7.0]	4	['combine ingredie	nts in blender'	['milk', 'vanilla	4
aww marinated olive	25274	15	21730	4/14/02	['15-minutes-o	[380.7, 53.0, 7.0, 24.0, 6.0, 24.0, 6.0]	4	['toast the fennel	my italian mil w	['fennel seeds'	9
backyard style barbe	67888	120	10404	7/30/03	['weeknight', 't	[1109.5, 83.0, 378.0, 275.0, 96.0, 86.0]	10	['in a medium sa	this recipe is po	['pork sparerib	22
bananas 4 ice cream	70971	180	102353	9/10/03	['weeknight', 't	[4270.8, 254.0, 1306.0, 111.0, 127.0]	8	['crumble cookies	into a 9-inch pi	['chocolate san	6
beat this banana bre	75452	70	15892	11/4/03	['weeknight', 't	[2669.3, 160.0, 976.0, 107.0, 62.0, 3.0]	12	['preheat oven to	from ann hodgr	['sugar', 'unsalt	9
berry good sandwich	109439	5	49168	1/25/05	['15-minutes-o	[79.2, 3.0, 58.0, 0.0, 0.0, 6.0, 5.0]	2	['in medium size	horseradish is o	['whole berry d	3
better than sex strav	42198	1460	41531	10/3/02	['weeknight', 't	[734.1, 66.0, 199.0, 10.0, 10.0, 117.0]	8	['crush vanilla wa	simple but sexy'	['vanilla wafers	7

The second file we will be using is the RAW_interactions.csv. Each row in this data file is one user reviewing one recipe. One user can review more than one recipe, and each recipe can be reviewed by more than one user, so there is a many-to-many relationship between users and recipes, but the combination of user_id and reviewer_id in each row will be unique. The rest of the fields in RAW_interactions.csv are given below.

Field Name	Field Description	Data type
user_id	User ID	Datatype
recipe_id	Recipe ID	Integer
date	Date of interaction	Date
rating	Rating given	Float
review	Review text	String

A few sample data points are shown in the image below.

user_id	recipe_id	date	rating	review
38094	40893	2/17/03	4	Great with a salad. Cooked on top of stove for 15 minutes. Added a shake of cayenne and a pinch of salt. Used low fat sour cream. Thanks.
1293707	40893	12/21/11	5	So simple so delicious! Great for chilly fall evening. Should have doubled it ;) Second time around forgot the remaining cumin. We usually love cumin but didn't notice the missing 1/2 teaspoon!
8937	44394	12/1/02	4	This worked very well and is EASY. I used not quite a whole package (10oz) of white chips. Great!
126440	85009	2/27/10	5	I made the Mexican topping and took it to bunko. Everyone loved it.
57222	85009	10/1/11	5	Made the cheddar bacon topping adding a sprinkling of black pepper. Yum!
52282	120345	5/21/05	4	very very sweet. after i waited the 2 days i bought 2 more pints of raspberries and added them to the mix. i'm going to add some as a cake filling today and will take a photo.