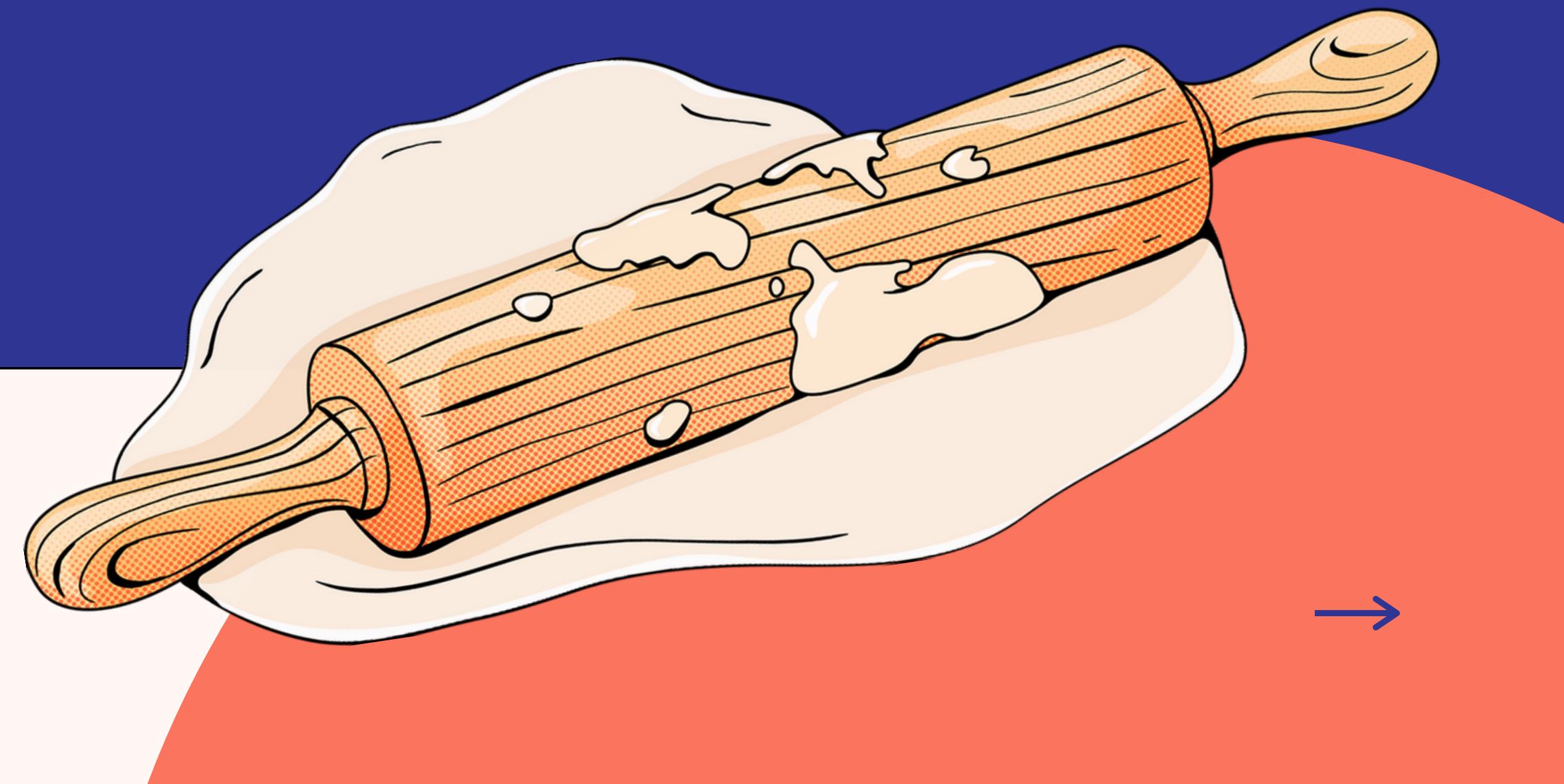


HOMECHEF

DIGITIZING THE FOOD CHAIN



DSI CAPSTONE PROJECT
MUBINA KAPASI



MEAL KIT SERVICE: AN OVERVIEW

- Intersection between meal delivery services like FoodPanda and Deliveroo, and grocery delivery services like RedMart
- Help consumers enjoy the benefits of a home cooked meal without the hassle of planning meals, buying groceries, or worrying about food wastage



BUSINESS MODEL

BASIC PLAN

- Novice home cooks can select recipes every week and ingredients will be provided to their door

FLEX PLAN

- Home cooks looking for a challenge can purchase their own ingredients a la carte from HomeChef's marketplace
- HomeChef will offer suggestions for cuisines, recipes, and other ingredients users may want to purchase based on their current selection



PROBLEM STATEMENT



1. Gain insights on the best cuisines, ingredients, and recipes to provide consumers
2. Use multiclass classification algorithms to predict the cuisine of a dish based on available ingredients
3. Provide recommendations for recipes based on available ingredients



Yummly:
15K+ recipes
from 18 different
cuisines

Removed non-
English recipes,
and recipes with
<2 ingredients

Ingredient parser
to extract list of
ingredients from
recipes

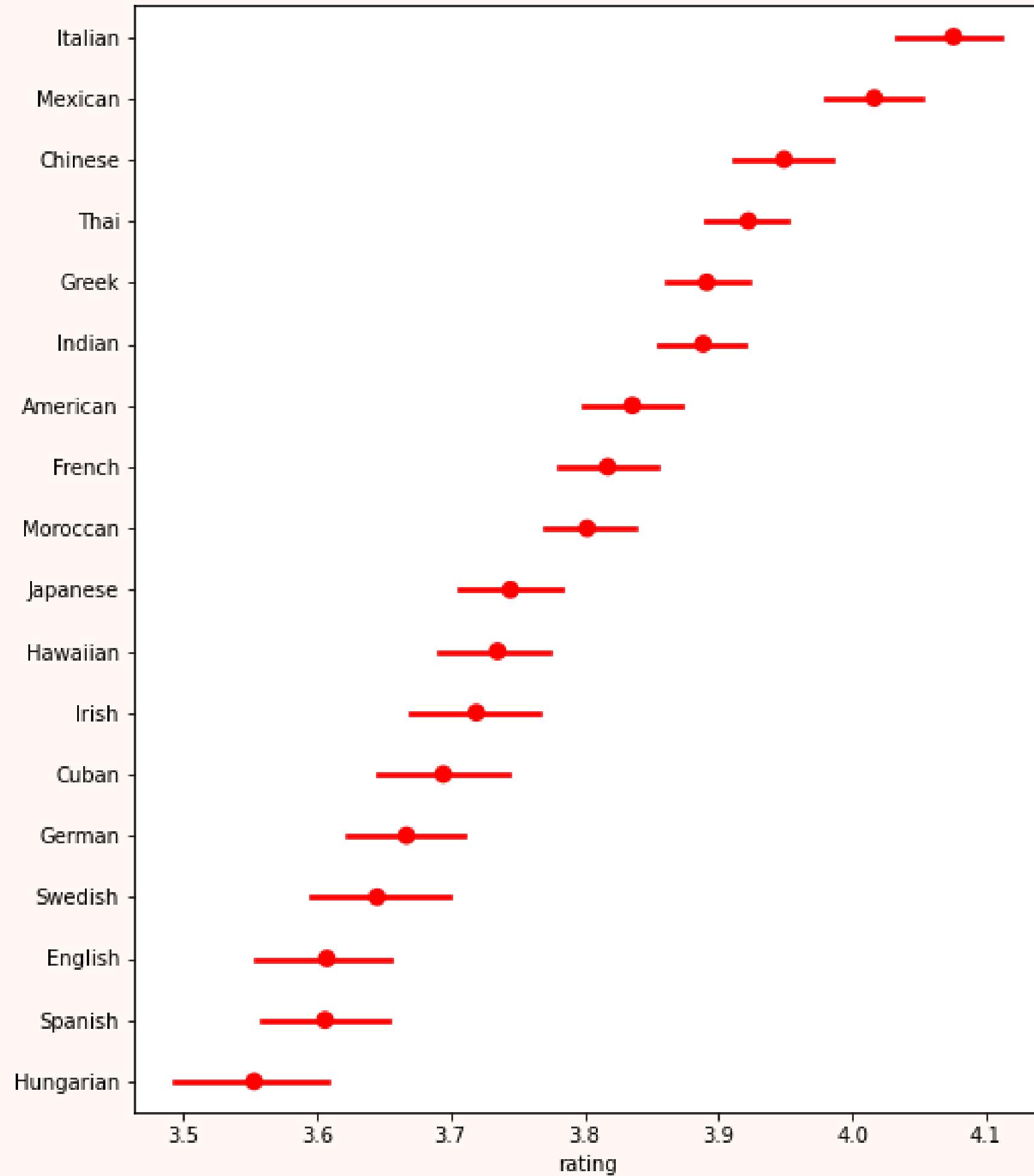
DATA COLLECTION

COMMON INGREDIENTS



clove thyme rice fennel pork tomato paste corn
sauerkraut flatleaf parsley
canola oil sea salt
saffron mustard butter fish sauce
mustard
beef
onion
cinnamon
brown sugar
chicken
almond
shrimp
nutmeg
cucumber
raisin
leek
spice
carrot soy sauce
flour
carrot soy sauce
beef
beef
beef
beef
black pepper
salt
garlic
tomato

BEST RATED CUISINES

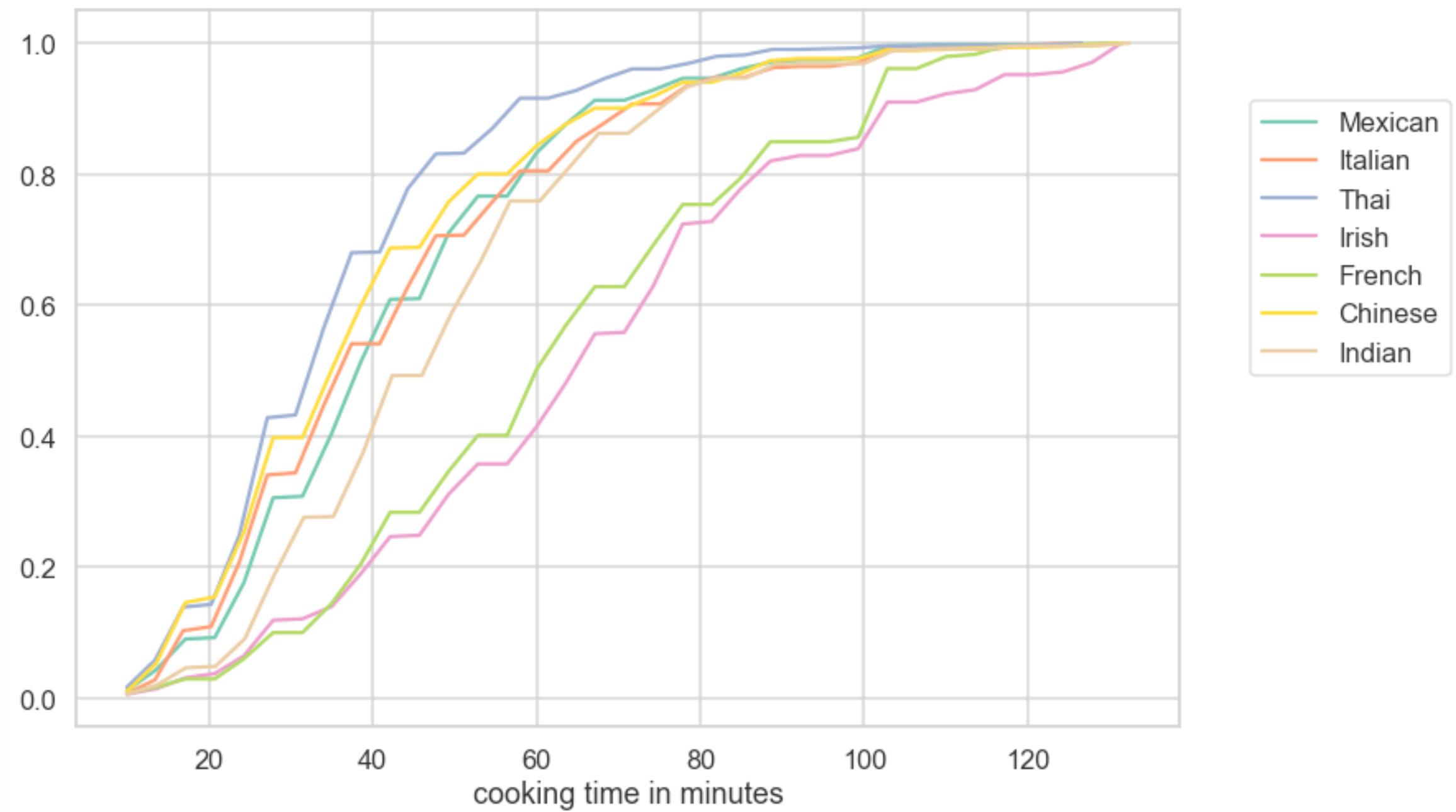


Italian and Mexican recipes have the highest rating, while Hungarian has the lowest rating



CUMULATIVE COOKING TIME

90% of Thai recipes take less than 60 minutes to make, but only 40% of Irish recipes take less than 60 minutes

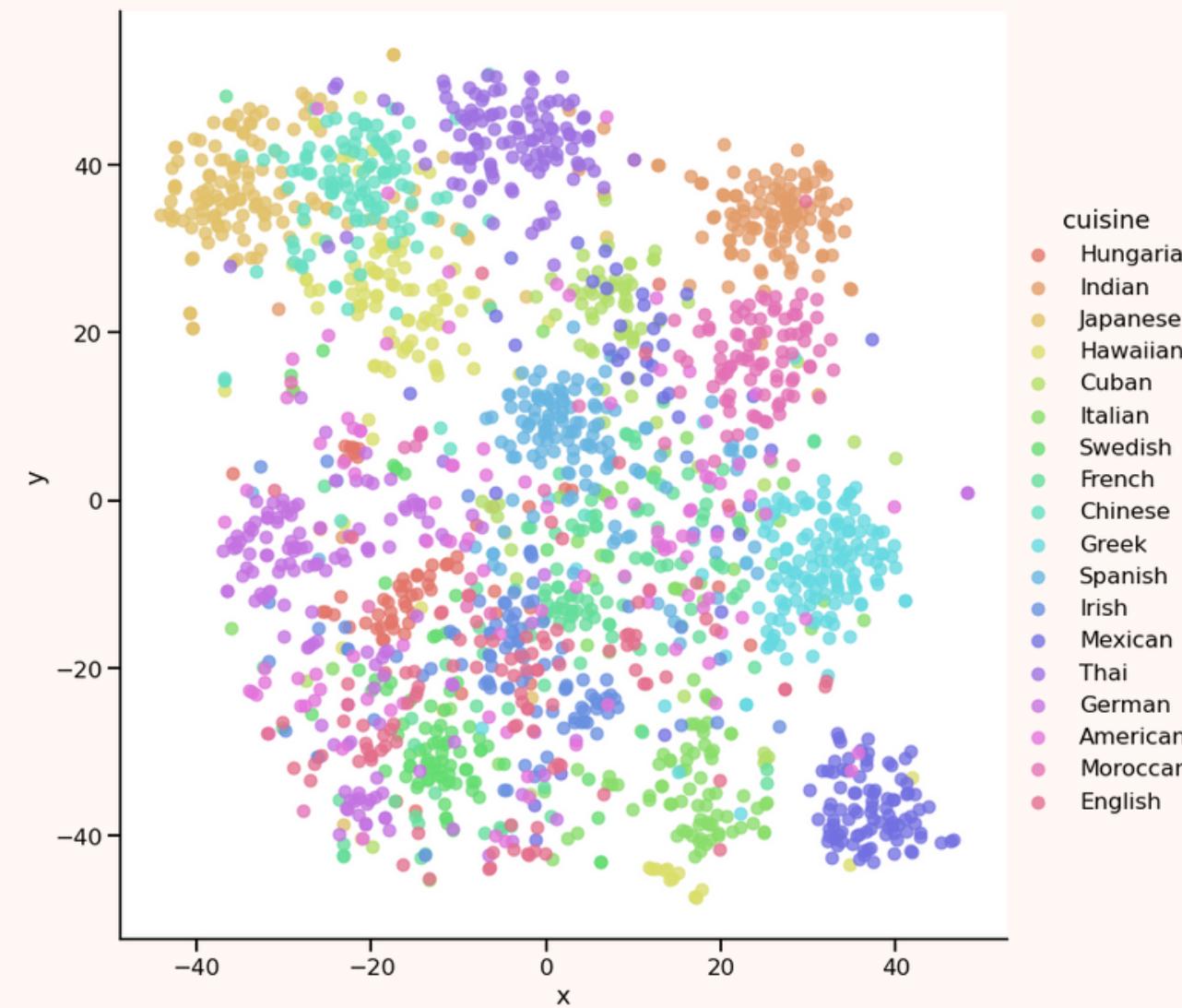


CLUSTERING ANALYSIS

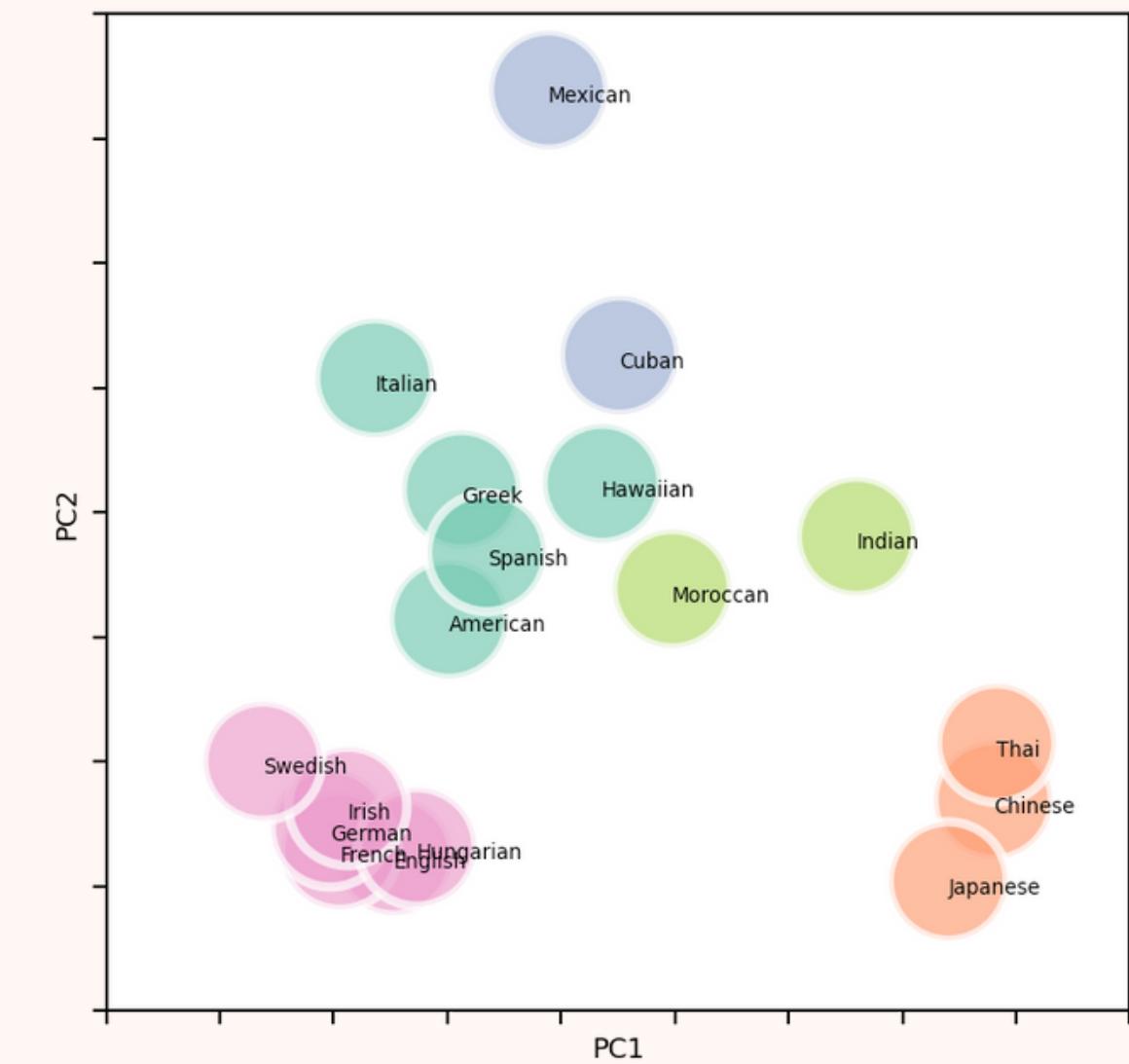
T-SNE AND KMEANS



t-stochastic neighborhood embedding for all recipes



K-Means clustering by cuiisine, k = 5



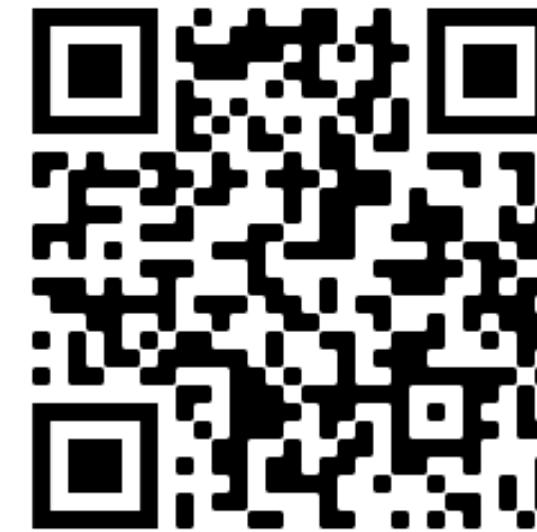
Some cuisines are more easy to distinguish than others: Mexican, Indian, and Thai recipes do not overlap that much with other cuisines, while English and American are scattered

With k=5, clusters seem to correspond to different geographic regions





explore cuisines
interactively!



<https://homechefexplore.herokuapp.com/>



MODELING: CUISINE CLASSIFICATION

1. TF-IDF Vectorizer

- maximum frequency of an ingredient occurring in a recipe is 1
- maximum document frequency is 95%

2. Multinomial logistic regression

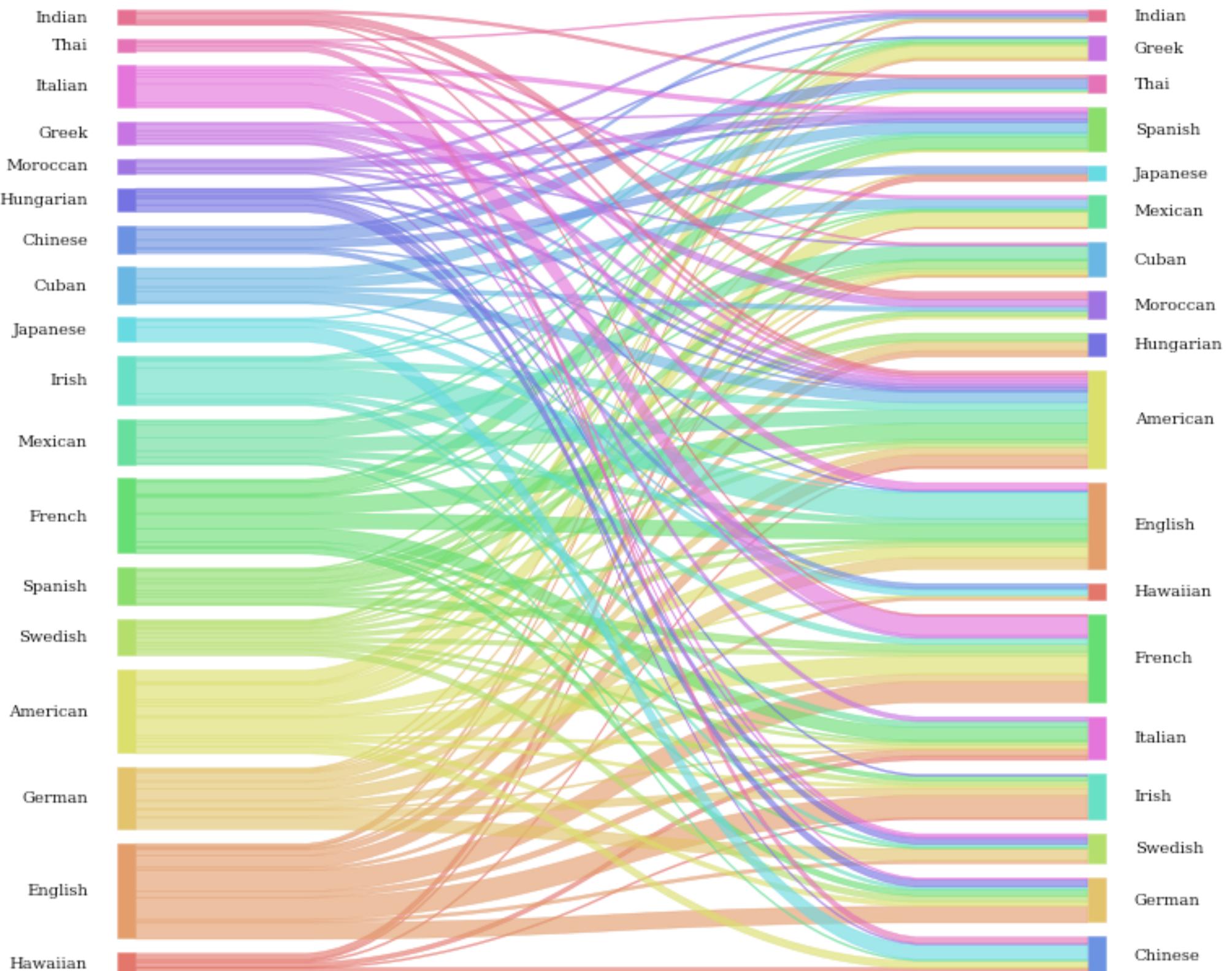
KAPPA SCORE

0.84



HOW DO CUISINES INFLUENCE EACH OTHER?

Most of the cuisines that are misclassified
are geographically close





CHINESE

soy sauce



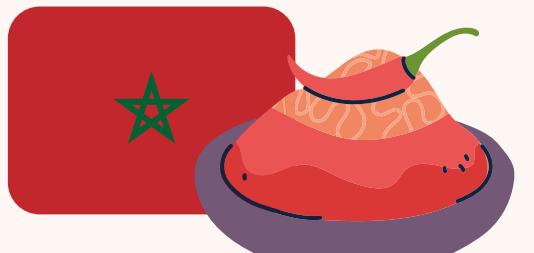
GERMAN

sauerkraut



FRENCH

thyme



MOROCCAN

harissa



IRISH

potato



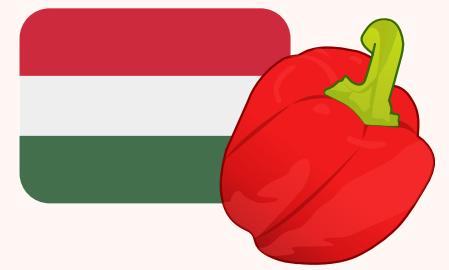
SPANISH

chorizo



MEXICAN

tortilla



HUNGARIAN

paprika



ITALIAN

pasta



AMERICAN

pecan



CUBAN

plantain



ENGLISH

pudding



GREEK

feta cheese



HAWAIIAN

pineapple



INDIAN

turmeric



JAPANESE

mirin



SWEDISH

meatball



THAI

fish sauce

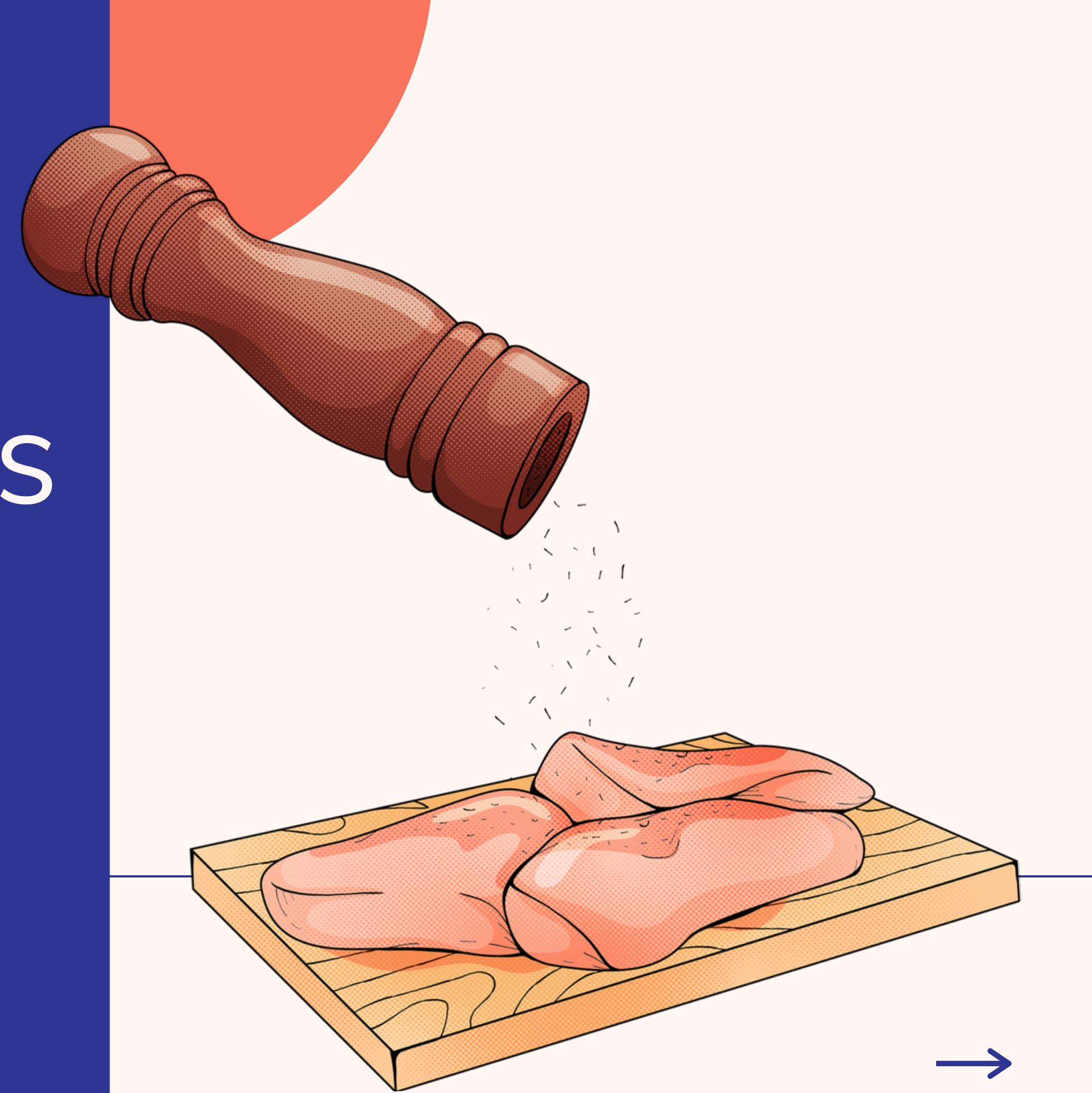
**MOST DISTINCTIVE
INGREDIENTS**

MODELING: RECIPE RECOMMENDATIONS

- 1.TF-IDF vectorizer
- 2.Cosine Similarity



Generate your own recipe
recommendations on telegram!



BUSINESS RECOMMENDATIONS

Focus on providing recipes from the following 5 cuisines:

1. Mexican
2. Italian
3. Chinese
4. Indian
5. French

In general, these cuisines have:

1. High recipe ratings
2. Manageable cooking times (except French)
3. Belong to different k-Means clusters



FURTHER STEPS

1. Collect recipes from more sources for full scope of taste preferences and variations
2. Sentiment analysis on target audience to better understand how consumers will react to certain cuisines and recipes
3. Collect user feedback to evaluate recipe recommendation system effectively
4. Build a personalized recommendation system



Thank you!

[click here to view my code for this project](#)

